Speech onsets and sustained speech contribute differentially to delta and theta speech tracking in auditory cortex

Nikos Chalas[1,2,4]*, Christoph Daube[3], Daniel S. Kluger[1,2], Omid Abbasi[1], Robert Nitsch[4], Joachim Gross[1,2]

[1] Institute for Biomagnetism and Biosignal Analysis, University of Münster, Münster, Germany

[2] Otto-Creutzfeldt-Center for Cognitive and Behavioral Neuroscience, University of Münster, Münster, Germany

[3] Centre for Cognitive Neuroimaging, University of Glasgow, Glasgow, UK

[4] Institute for Translational Neuroscience, University of Münster, Münster, Germany

* *Correspondence: Nikos Chalas, University of Münster, Institute for Biomagnetism and Biosignal Analysis, Malmedyweg 15, 48149 Münster, Germany.*

Abbreviated title

Speech onsets and sustained speech tracking

Abstract

When we attentively listen to an individual's speech, our brain activity dynamically aligns to the incoming acoustic input at multiple timescales. Although this systematic alignment between ongoing brain activity and speech in auditory brain areas is well established, the acoustic events that drive this phase-locking are not fully understood. Here, we use magnetoencephalographic (MEG) recordings of 24 human participants (12 females) while they were listening to a one hour story. We show that whereas speech-brain coupling is associated with sustained acoustic fluctuations in the speech envelope in the theta frequency range (4 - 7 Hz), speech tracking in the low frequency delta (below 1 Hz) was strongest around onsets of speech, like the beginning of a sentence. Crucially, delta tracking in bilateral auditory areas was not sustained after onsets, proposing a delta tracking during continuous speech perception that is driven by speech onsets. We conclude that both onsets and sustained components of speech contribute differentially to speech tracking in delta and theta frequency bands, orchestrating sampling of continuous speech. Thus, our results suggest a temporal dissociation of acoustically driven oscillatory activity in auditory areas during speech tracking, providing valuable implications for orchestration of speech tracking at multiple time scales

Keywords

auditory cortex; magnetoencephalography; mutual-information; speech perception; speech-brain coupling;

Introduction

Our perceptual system benefits from environmental regularities (Lakatos et al. 2019). When we listen to an individual's speech, amplitude fluctuations of the speech waveform – the so-called *speech envelope* – exhibits temporal regularity at around 5 Hz (Giraud and Poeppel 2012; Ding et al. 2017) (Giraud and Poeppel 2012; Ding et al. 2017; Poeppel and Assaneo 2020). A line of research has shown that the auditory cortex dynamically tracks these fluctuations through the phase of low-frequency oscillations (Luo and Poeppel 2007), a process which is thought to facilitate the parsing and grouping of continuous speech (Giraud and Poeppel 2012) with perceptual relevance (Keitel et al. 2018) and which is conceptually motivated by observations of rhythmic neural activity at different frequencies along the auditory pathway (Giraud and Poeppel 2012; Keitel and Gross 2016; Obleser and Kayser 2019; Brodbeck and Simon 2020; Meyer et al. 2020).

In this line, associations between speech and the brain have been prominently observed in the delta and theta frequency range (Ahissar et al. 2001; Luo and Poeppel 2007; Gross et al. 2013, 2013; Ding and Simon 2014, 2014; Kayser, Ince, et al. 2015; Ding et al. 2017; Keitel et al. 2018; Jin et al. 2020); (Ahissar et al. 2001; Luo and Poeppel 2007; Gross et al. 2013; Ding and Simon 2014; Kayser, Ince, et al. 2015; Jin et al. 2020), depicted as peaks in the frequency-resolved speech-brain coupling analysis (Gross et al. 2013; Ding and Simon 2014; Ding et al. 2017; Keitel et al. 2018). Speech tracking postulates a simultaneous phase-resetting of ongoing oscillations driven by acoustic landmarks in the speech envelope which in the temporal domain can include acoustic landmarks such as amplitude peaks (Doelling et al. 2014) or edges [peaks in the amplitude's rate of change (Oganian and Chang 2019)]. Crucially, speech tracking analysis benefits from the inclusion of the envelope's rate of change in low delta frequencies (0.6 - 0.8 Hz), but not in the typical theta-band peak at 5 Hz (Chalas et al. 2022), suggesting that speech-brain coupling is differentially driven by distinct temporal landmarks in the delta and theta bands.

Speech tracking in the theta band has been interpreted as an index of flexible theta oscillations aligning their phase and marking perceptual segments for further processing (Luo and Poeppel 2007; Giraud and Poeppel 2012; Ding and Simon 2014; Doelling et al. 2014) whereas the delta band has been suggested to represent both segmentation of speech components that are closely related to the acoustic structure of speech and linguistic elements (Giraud and Poeppel 2012; Ding et al. 2016; Meyer et al. 2017; Keitel et al. 2018; Boucher et al. 2019; Rimmele et al. 2021). Nevertheless, consensus has not yet been reached regarding the specific role of delta speech tracking and its relationship to theta speech tracking.

In natural settings, ongoing sustained speech is interleaved by subsequent periods of silences and speech onsets (Rosen 1992; Zellner 1994), which are crucial for speech segmentation (Dilley and Pitt 2010) and intelligibility (Koning and Wouters 2012). Importantly, neural responses at the beginning of a phrase or a sentence are spatially segregated from sustained responses within superior temporal gyrus (Hamilton et al. 2018). However, different speech components in the time domain (i.e. onsets/offsets and sustained speech) have not been systematically related to neural tracking of speech in the auditory cortex thus far. Therefore, it is yet unclear whether neural tracking of speech at the auditory areas aligns to amplitude fluctuations in ongoing sustained speech or reflects also different temporal landmarks such as speech onsets/offsets at the beginning of a phrase or a sentence.

Thus, here we aim to address whether different temporal events in the speech envelope can give rise to differential speech tracking in delta and theta bands in bilateral auditory areas. We focused on 0.6 Hz (delta band) and 5 Hz (theta band) as both frequencies occur as peaks in a recent speech-tracking analysis while only speech tracking in 0.6 Hz was benefited by the inclusion of the envelope's derivative (Chalas et al. 2022). While we take 0.6 Hz and 5 Hz as representative frequencies of the delta and theta band respectively, we also show that results generalize within each respective band. We capitalize on the high signal-to-noise ratio of a one hour-long speech

listening MEG dataset (Daube et al. 2019) and focus on the temporal segments that drive speech-tracking in delta (below 1 Hz) and theta (5 Hz). We find that events with high amplitude in low-delta and theta bands in the speech envelope correspond to speech onsets/offsets and ongoing sustained speech, respectively. Overall, our results indicate that both frequency bands contribute to speech tracking but that they do so at different times.

Methods

Participants and study design

Twenty-four participants (12 female; mean age = 24.0 years, age range 18 - 35 years) participated in this study. All participants provided written consent prior to the experiment and received a monetary compensation of £9 per hour. The study was approved by the College of Science and Engineering Ethics Committee at the University of Glasgow (application number: 300170024). Participants were asked to listen attentively to a 55 minute long audiobook, and they were informed that they would have to answer questions relative to the story at the end of the session. Brain activity was monitored with a 248-magnetometer whole-head MEG system (MAGNES 3600, 4-D Neuroimaging) in a magnetically shielded room. Weighted T1-MRI images were obtained prior to the experiment from each individual. MEG Data were acquired at a sampling rate of 1017.25 Hz for 10 participants and 2035.51 Hz for 14 participants. Individual head shapes were digitized before each recording via five coils attached to the head. Each MEG session was separated into 6 blocks of ~9.16 minutes. The last 10 seconds of each block were repeated in the following block to allow participants to more easily follow the story. The stimulus was delivered using PsychToolBox (Brainard 1997) with two Etymotic ER-30 insert earphones. To assess whether participants paid attention to the story, they answered 18 multiple choice questions at the end of the recording (3 response options each) with the number of correct options varying

between 1-3 per question (mean accuracy 0.95; SD 0.05; range 0.78-1). A different analysis of this dataset has been reported elsewhere (Daube et al. 2019; Chalas et al. 2022).

Data preprocessing

MEG data were processed with the Fieldtrip toolbox (Oostenveld et al. 2011) and in-house Matlab scripts [Matlab 2022a (The MathWorks Inc)]. Bad channels were manually detected and interpolated (spherical-spline) from neighboring channels (mean number of rejected channels per block M = 3.07; SD = 3.64). Squid jumps were replaced with DC patches and continuous data from the onset of the story were denoised by subtracting the projection of the data on orthogonal basis of the reference channels (using *ft_denoise_pca*). Continuous data were further filtered offline with a high-pass filter of 0.5 Hz (fourth-order forward-reverse zero-phase Butterworth) and downsampled to 100 Hz. Independent components from heartbeats and eye movements were visually isolated and removed using the *runica* function of FieldTrip (mean number of rejected components per block M = 5; SD = 5.3)

Frequency-specific speech envelope components

First, the amplitude-modulated speech envelope was extracted from the continuous speech signal for each experimental block. For this, 31-channel Log-Mel-Spectograms (124.1 Hz - 7284.1 Hz) were computed and the wide-band speech envelope was estimated by the sum of the absolute values across bands (Schädler et al. 2012) and further resampled at 100 Hz. We were further interested in frequency-specific components of the speech envelope. For that, we transformed the speech envelope to the frequency domain using a continuous morlet wavelet transform (CWT) for 64 frequencies (0.5 to 40 Hz; cwtfilterbank.m in Matlab - wt.m performs the actual transformation into the frequency domain; the Morse wavelet had symmetry parameter gamma = 3 and time-bandwidth = 10). We selected bands centered on 0.6 Hz and 5 Hz as representative of the delta and theta frequency bands (see Introduction) respectively, and we extracted the

absolute values of the complex data as a measure of amplitude along with the phase angles for the frequencies of interest (Figure 1A). While we focused on two frequencies (0.6 and 5 Hz), we show that the results presented here generalize across neighboring frequencies in both bands.

We aimed to identify time segments in the speech envelope that show strong power in the delta or theta band. Therefore, we extracted 4-s segments centered on the peaks of the cosine of the phase from the speech envelope (using findpeaks.m in Matlab) for frequency bins centered on 0.6 Hz and 5 Hz (not wavelet-transformed; Figure 1A; *top panel*). Then we sorted these segments by their mean amplitude (0.6 Hz or 5 Hz). This resulted in data segments with high and low power for 0.6 Hz or 5 Hz. We extracted an equal number of segments for each category (n=886; high and low power for delta and theta). Next, we replaced the data in each segment with the original speech envelope. As a result, we obtained segments of the speech envelope corresponding to the four categories (high and low power for delta and theta; see Figure 1B and 1D for segments for delta and theta, respectively). To characterize the temporal structure of the extracted segments, we clustered those exhibiting high power in the delta and theta frequency band using a k-means algorithm. We estimated the optimal number of clusters using a Silhouette evaluation (function *evalclusters* in matlab) which resulted in k = 2 for delta (0.6 Hz; Figure 2A) and k=4 for theta (5 Hz; Figure 2D). The resulting temporal patterns correspond to prototypical waveforms of the speech envelope when power in the delta or theta frequency band is high.

Regions of interest

For source analysis, individual T1-weighted MRIs were used to estimate the individuals' head models. MRIs were coregistered to the MEG coordinate system, using the digitized head shapes and an iterative closest point algorithm (Besl and McKay 1992). Single-shell volume conductor models were generated from individual MRI after segmentation to white matter, gray matter, and cerebrospinal fluid (Nolte 2003). To estimate source activity, LCMV Beamformer was used to

estimate coefficients from the MEG time series for each voxel on a 5 mm grid (Van Veen et al. 1997). The sensor covariance matrix used was computed across all blocks and the λ regularization parameter was set to 0%. Time series were extracted for each dipole orientation, resulting in three time series per voxel. We applied an atlas-based parcellation of cortical space, resulting in 181 ROIs per hemisphere (Glasser et al. 2016). We constrained our analysis to auditory responses so we combined bilateral A1, Mbelt, LBelt, PBelt, and RI parcels, resulting in a combined parcel that is referred to as Early Auditory Cortex. Source time series of these parcels were concatenated across voxels and orientations and we extracted the first three principal components for the speech tracking analysis.

Speech tracking and statistical analysis

We computed statistical dependencies between the speech envelope and the bilateral early auditory cortex on the basis of information theory (Shannon 1948). We estimated mutual information (MI) using Gaussian Copula MI (GCMI) between speech signals and the first three principal components from the regions of interest (Ince et al. 2017). GCMI performs analytical computation of MI between two signals after gaussian copula transformation. This approach has the advantage of being computationally efficient, avoiding the estimation of probability distributions from discrete variables which requires an excessive number of data samples, while it doesn't require an a-priori assumption of the marginal distributions.

To identify frequency specific interactions in speech and brain signals, we also applied a continuous wavelet transformation (CWT) for 64 frequencies (from 0.5 Hz to 40 Hz) to each time series extracted from bilateral early auditory cortex (three principal components per ROI; see above). The estimation resulted in GCMI spectra indicating phase alignment between speech and brain at zero delay. Speech-tracking (as depicted in GCMI values) can vary in temporal delays for different frequencies, but we did not have a specific hypothesis for temporal delays operating

for delta and theta frequency bands. Thus, we decided to report speech-brain coupling at zero delay, but we replicated the same analysis for various positive delays (from 10 to 300 ms with steps of 10 ms) and the results were qualitatively similar (see Supplementary Figure 2). Then, significance of GCMI values at the group level were determined with a series of two-tailed permutation tests (Maris and Oostenveld 2007) comparing individual GCMI per parcels, frequencies and conditions (high and low power) and thresholded at p=0.05 and corrected for multiple comparisons with False Discovery Rate (Benjamini 2010).

Results

Speech envelope related to onsets and sustained speech

We hypothesized that not all speech segments contribute equally in frequency-resolved speech-brain coupling. To identify temporal segments that differentially lead to delta and theta coupling, first we aimed to characterize temporal segments of the acoustic speech that give rise to high and low amplitudes in delta and theta frequency bands in the speech envelope.

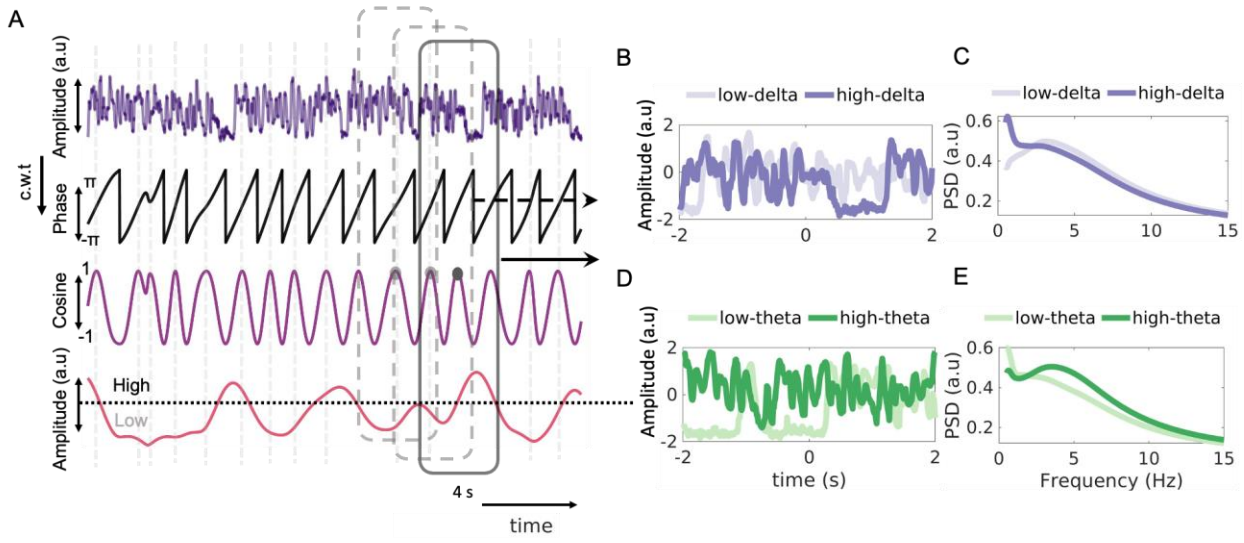

Figure 1: **Procedure of identifying segments of high/low power in delta and theta frequency bands.** (A) Speech envelope (top panel) was transformed in the frequency domain with a continuous Morlet wavelet transform (c.w.t). We identified peaks of the cosine of the phase (third row) exhibiting high/low amplitude (fourth row) and we sorted them according to their amplitude to high and low. (B) Examples of single-segments for high (dark purple) and low power at 0.6 Hz (light purple). (C) Power spectral densities for segments with high power at 0.6 Hz (high delta; dark purple) and low power at 0.6 Hz (low theta; light green). (D) Examples of single-segments for high (dark green) and low power in 5 Hz (light green). (E) Power spectral densities for segments with high power in 5 Hz (high theta; dark green) and low power in 5 Hz (low theta; light green).

In Figure 1B, we plot representative single-trial segments for high and low power in 0.6 Hz. We observed that the high-delta segment (dark purple) contains a gap in the speech signal near time 0s, which was absent in a segment with low delta power (light purple). Figure 1C shows the power spectral densities of segments with high and low power at 0.6 Hz (high delta, low delta). It is evident that there is a broad 1/f profile, with increased power at progressively lower frequencies with no clear deviating peak, corroborating that segments with high delta power do not exhibit sustained periodic activity. In contrast, with a similar approach power spectral density of high theta shows a clear peak in 5 Hz (Figure 1E) suggesting sustained speech activity, evidence in single trial segments (Figure 1D; see below for a more detailed description).

To identify the temporal structure giving rise to high delta power, high-delta segments were subjected to k-means clustering (k=2; see Methods). Figure 2A shows the two resulting clusters and Figure 2C all the high delta trials sorted according to the k-means clustering. In figure 2B we plot 25 representative segments with high-delta power at 0.6 Hz. It is evident that segments with high delta power at 0.6 Hz correspond to trials containing speech onsets in the envelope, such the ones at the beginning of a sentence (Supplementary figure 1B).

We employed a similar approach for the theta band and extracted data segments with high theta amplitude in 5Hz (See Methods). We note that the modulation spectrum of our stimulus material peaks at 5 Hz (Supplementary figure 1C), thus we were expecting sustained periodic activity in the speech envelope segment's exhibiting high theta power in 5 Hz. We identified 858 segments with high theta power (Figure 2E). In Figure 1D, we plot representative single-trial

segments for high theta power. As expected, the segment with high theta power exhibited sustained speech activity, opposed to the low theta segment. We wanted to see if this generalizes to all segments with high theta power. In this case, k-means clustering (k=4; see Methods) did not yield interpretable results as the higher frequency of the theta band results in higher temporal variability and thus precludes the identification of clear temporal structure (Figure 2D). We illustrate the temporal structure giving rise to high theta amplitude by plotting the speech envelope segments with highest theta amplitude. As expected, these speech envelope segments reveal sustained rhythmic structure in the theta frequency band (Figure 2F).

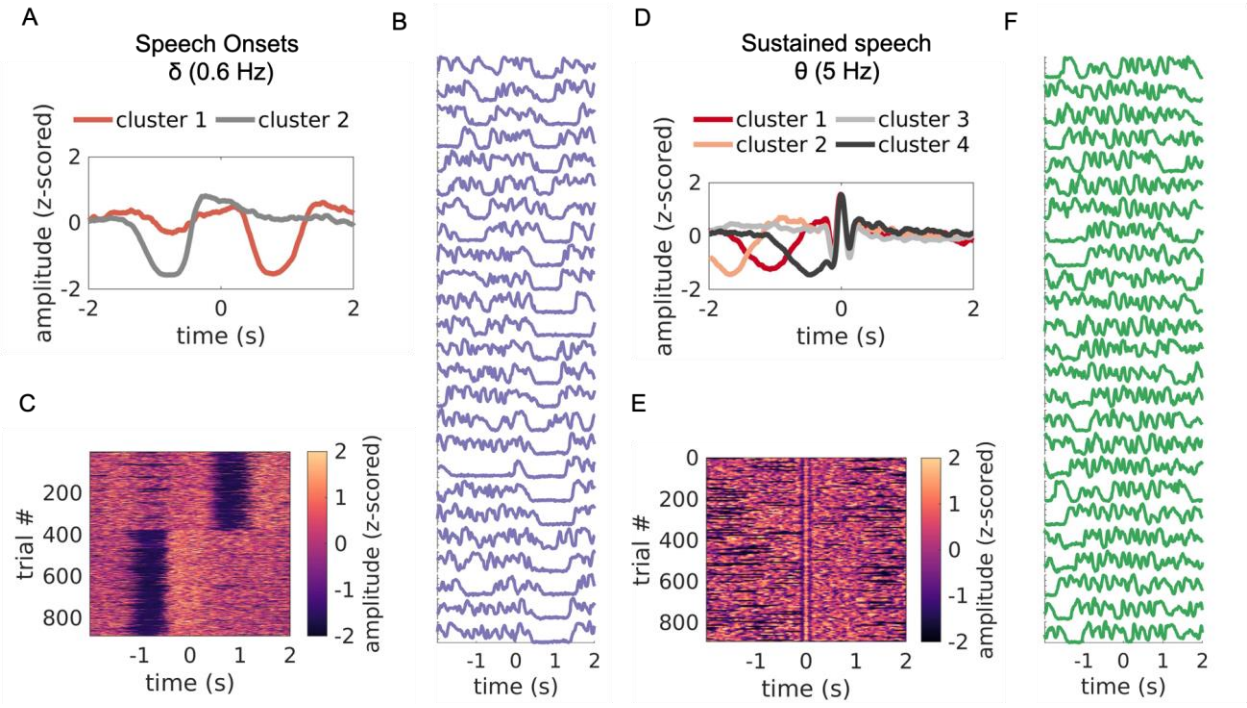

Figure 2: **Onsets and sustained speech components in high delta and theta speech envelope segments**: (A) Clustering of segments exhibiting high power in 0.6 Hz with k-means (k=2). (B) Example of 25 segments with high power in 0.6 Hz. (C) Segments of speech envelope with high delta power in 0.6 Hz sorted according to a k-means algorithm (k=2; see Methods). (D) Clustering of segments exhibiting high power in 5 Hz with k-means (k=4). (E) Segments of speech envelope with high theta power in 5 Hz. (F) Example of 25 segments with high power in 5 Hz.

In summary, we find that the two frequency bands most often discussed in the literature of speech tracking, namely delta and theta, correspond to different temporal phenomena in the speech envelope: While high delta power coincides with abrupt amplitude step changes in the speech

envelope (energy resets such as during speech onset), high theta power reflects sustained rhythmic temporal structure at 5 Hz, reflecting the peak at modulation spectra of the stimuli.

High delta and theta speech envelope segments are temporally dissociated

Since speech onsets (associated with high delta power) and sustained speech (related to high theta power) correspond to different time segments in the speech envelope, we hypothesized that the amplitude of both frequency bands is anti-correlated. Figure 3 shows that this is indeed the case. In fact, correlating 0.6 Hz amplitude with the amplitude of all frequencies up to 15 Hz (Figure 3A) reveals the strongest negative correlation in the theta band which is not evident for other frequencies ($r(32998) = -0.2$; $p << .001$; Figure 3B)

Surprisingly, we observed that the low theta segment (Figure 1D) is mostly characterized by periods of silence, dynamics that we found to characterize high delta segments. This suggested a phase-amplitude relationship between delta and theta segments. Indeed, plotting the grand average amplitude of speech segments locked to high delta events reveals that amplitude of theta follows phase of delta (Figure 3C), which is not the case for other frequencies in the delta band (see Supplementary Figure 2).

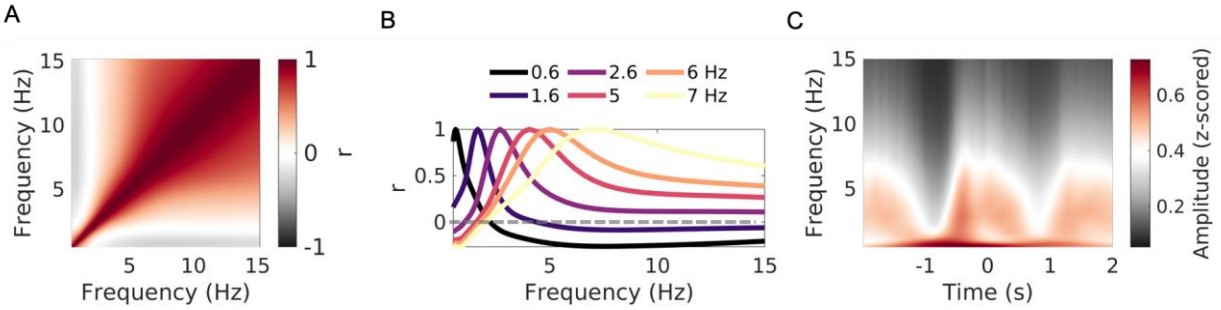

Figure 3: **Phase and amplitude relation of delta and theta band in the speech envelope**. (A) Correlation matrix for power of 0.5 - 15 Hz after continuous Morlet wavelet transformation of the speech envelope. (B) Individual frequency correlations of 0.6, 1.6, 2.6, 5, 6, and 7 Hz amplitude with all frequencies up to 15 Hz (*bottom*). (C) theta amplitude follows delta phase of 0.6 Hz segments in the speech envelope.

Differential high and low delta and theta speech tracking in auditory areas

We identified temporal segments in the speech envelope that exhibit differences in the amplitude power in the delta and theta frequency bands. Next, we wanted to test whether this high or low delta and theta power in the speech envelope is reflected in speech-brain tracking. To quantify speech tracking, we followed an information-theoretic approach and quantified the frequency-specific interactions of speech envelope segments with source-estimated auditory activity following previously validated measures (Gross et al. 2021). For estimating an individual's auditory activity, we combined activity from early auditory parcels of the HCP-MMP1 atlas, concatenated them, and extracted the first three principal components (see Methods). In Figure 4A, we show the cortical location of the parcels that were combined to extract auditory activity.

We computed GCMI (Ince et al. 2017) between speech envelope and brain activity from left and right early auditory cortex for time segments of high and low delta power in 0.6 Hz (Figure 4B, *top panel*) and statistically compared them to obtain t-value spectra at the group level. The black line in Figure 4C shows the corresponding t-values for high versus low power at 0.6 Hz. As expected, higher delta power at 0.6 Hz in the speech envelope leads to significantly higher GCMI values at this frequency (black line; group statistics; $p<0.05$; FDR corrected). Interestingly, the same contrast reveals significantly lower theta and higher beta (β) GCMI (group statistics; $p<0.05$; FDR corrected). This means that time segments with high delta power in the speech envelope (corresponding to speech onsets) are associated with high delta and beta but low theta speech tracking. This effect is significant in the left and right early auditory cortex. To check if this holds for more frequencies inside the delta band, we repeated this analysis for two more frequencies (1.6 and 2.6 Hz) and we found a similar but less prominent effect beyond the 0.6 Hz (Figure 4B, *top panel*). We also directly compared the difference in delta and theta tracking between left and right AC. For that we computed and we statistically compared the difference between high- and low- for delta and theta tracking in left and right AC. We found no significant difference between left and right AC (Supplementary Figure 2).

Next, we proceeded with the estimation of speech tracking for high and low theta power in 5 Hz (Figure 4B, *lower panel*). Again, we statistically contrasted them, resulting in t-value spectra at the group level. We find that high theta power in the speech envelope is associated with significantly higher GCMI in the theta band and significantly lower GCMI in the delta and alpha (α) frequency band (red line; group statistics; $p<0.05$; FDR corrected). To test if this effect generalized in the theta band we repeated the same procedure for different frequencies inside the theta band. As expected, the effect was robust across different frequencies inside the theta frequency range (6 and 7 Hz), indicating speech-tracking in the theta timescale through flexible oscillators, as previously proposed (Giraud and Poeppel 2012; Hyafil et al. 2015; Rimmele et al. 2018).

In summary, we observed high delta speech tracking to segments containing energy onsets, such as silent gaps before the start of the sentence. We hypothesized that the high GCMI observed is specific to the transition from silence to speech, and thus it would drop after speech unfolds after onset. To test this, we further characterized the time course of delta speech tracking relative to onsets. We computed delta MI between speech envelope and brain activity in left and right early auditory cortices in 2 second-long windows moving from -1 second to 1 second (steps of 50 ms) relative to the center of 450 onsets (Figure 4D). As expected, we observed MI increase around speech onsets. However, this effect was not sustained, but GCMI decreased already shortly after onsets. This further supports our finding that delta speech tracking is maximal near rapid speech envelope changes.

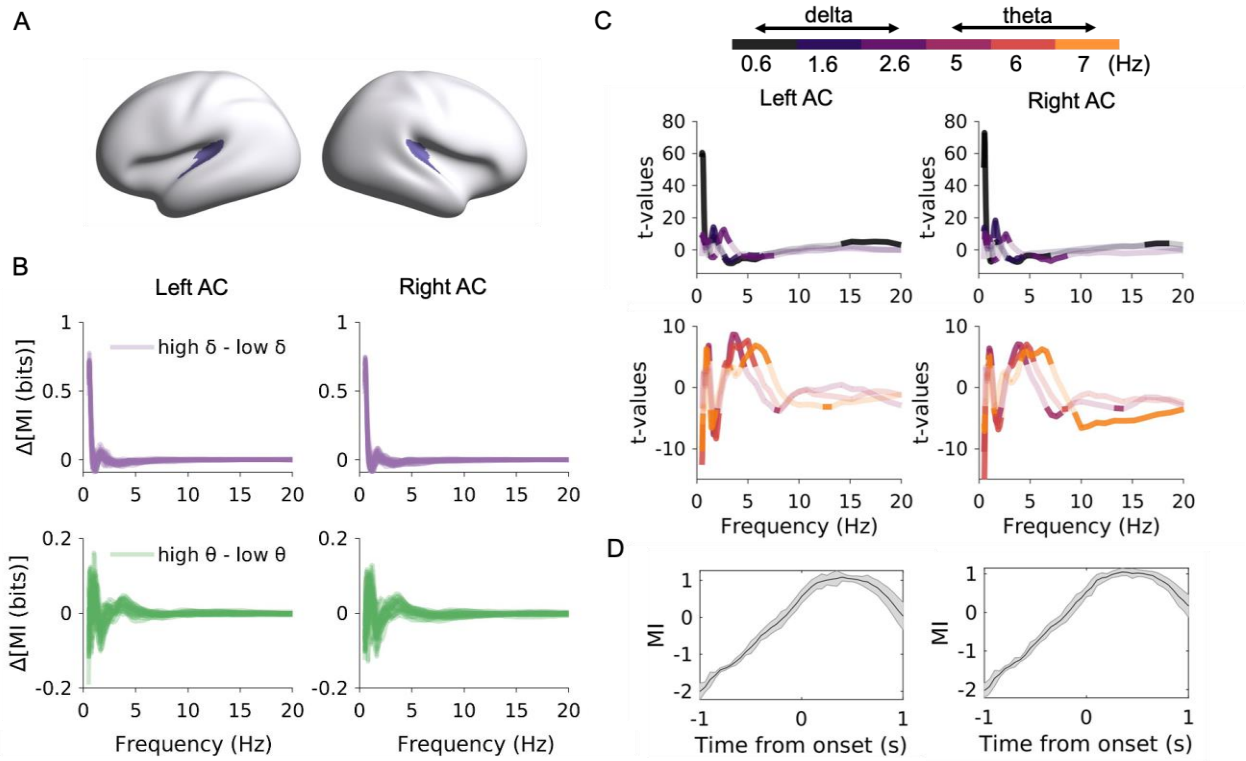

Figure 4: **High- and low- delta and theta speech components contributing to speech tracking**: (A) Early auditory areas where regions-of-interest source activity was extracted. Analysis included the first three principal components per region-of-interest (B) Individuals' speech tracking difference (GCMI values) for 4 second time segments between high and low power in 0.6 Hz (*top panel)* and 5 Hz (bottom panel) for left (*left panels*) and right (*right panels)* auditory early cortex. (C) Statistical comparison for high and low power in delta and theta speech envelope segments. A two-tailed non-parametric test was applied to compare high-delta vs low-delta (*top panels*) and high-theta vs low-theta (*bottom panels*) for left (*left panels)* and for right early auditory cortex (*right panels).* High opacity illustrates statistically significant tracking *(p<0.05,* FDR corrected). (D) Speech tracking (GCMI values, z-scored) at 0.6 Hz across delays (steps of 50 ms) for 2 second time segments with high power in 0.6 Hz centered on onsets (n=450).

Overall, our results indicate that the delta and theta speech tracking is at least partly caused by different temporal patterns in the speech envelope. The delta band in the speech envelope reflects temporal patterns in the speech envelope containing energy resets such as onsets of speech that lead to high phase coupling to brain activity in the auditory cortex. In contrast, theta band activity in the speech envelope corresponds to a sustained, rhythmic pattern and leads to high phase coupling to brain activity in the auditory cortex in the theta band.

Discussion

Systematic speech-brain alignment occurs in two distinct timescales in low frequencies, delta (delta: 0.5 - 3 Hz) and theta (theta: 4 - 7 Hz). Do all temporal segments of continuous speech contribute equally to speech tracking? Here, we find that neural tracking at delta and theta frequencies can be acoustically driven both by temporally dissociated onsets and sustained speech, respectively. Specifically, while tracking in the theta range reflects speech components with sustained periodic activity, neural tracking at the delta frequency range was related to speech segments containing onsets of speech with no sustained periodic activity.

Speech envelope tracking in delta and theta frequency bands

Human communication is rich in temporal complexity (Giraud and Poeppel 2012). Encoding speech with high fidelity requires the human brain to temporally sample speech signals at different scales (Poeppel 2003; Kiebel et al. 2008), resulting in the temporal multiplexing of neural information (Panzeri et al. 2010; Gross et al. 2013). Our analysis links speech tracking measures (in our case, GCMI) in the delta and theta frequency bands to the speech envelope signal. The results are relatively straightforward for the theta band: Time segments of the speech envelope with high theta power are characterized by sustained periodic components in the envelope, which is reflected by a peak in the modulation spectrum of the stimuli (Ding et al. 2017). Statistically comparing speech tracking of segments of high versus low theta power shows

311 significantly higher speech tracking (higher GCMI values) in the theta band for segments with high
312 compared to low theta power. This indicates that time segments in the envelope with highly
313 sustained speech at 5 Hz lead to strong theta speech tracking, consistent with previous reports
314 (Luo and Poeppel 2007; Gross et al. 2013; Zion Golumbic et al. 2013; Ding and Simon 2014).
315 However, we note that these results are based on a statistical contrast of speech tracking between
316 segments categorized by power. A more direct approach would contrast segments of high versus
317 low theta speech tracking, but this is highly non-trivial since speech tracking is quantified with
318 connectivity measures that are computed across many samples or segments. In summary, our
319 findings support the notion that theta speech tracking represents alignment of neural activity to
320 periodic and sustained patterns of speech at the modulation rate.

321 However, that's not the case for the delta band. Speech segments with high delta power coincide
322 with acoustic onsets rather than with low-frequency patterns during sustained speech. More
323 precisely, time points of strongest delta power mark the transition from acoustic gaps to a speech
324 onset. In addition, delta power scales with the magnitude of the acoustic step change where a
325 transition from silence to high acoustic amplitude leads to high delta power. Similar to the theta
326 band, higher delta power in the speech envelope is associated with higher delta speech tracking
327 (and reduced theta speech tracking). Our results show that speech onsets specifically drive delta
328 speech tracking in frequencies below 1 Hz, which in turn is anticorrelated with theta tracking. This
329 is evident both in speech envelope power (Figure 3B) and in speech tracking MI (Figure 4C). Our
330 observed anti-correlation suggests a partial temporal dissociation between both frequency bands.
331 Previously, it was shown that sustained speech (high theta power) and onsets (high delta power)
332 are dissociated spatially within superior temporal gyrus with high-frequency activity (above 70 Hz;
333 (Hamilton et al. 2018) and in the temporal lobe with fMRI (Davis et al. 2011). We find that time
334 periods of high delta power in the speech envelope coincide with periods of low theta power and
335 vice versa, consistent with the view that delta signals represent onsets, whereas theta represents

sustained periodicities at the modulation rate. We thus suggest that delta speech tracking – below 1 Hz – is not a marker representing continuous tracking of speech features, but rather aligns to salient energy resets during speech perception.

Acoustically driven delta speech tracking

We report acoustic-triggered delta speech tracking related to speech onsets. Previously, delta was found to track characteristic “temporal marks of articulated sounds” (Boucher et al. 2019). In a recent MEG study, speech tracking was higher in delta when compared to continuous white noise or spectrally inverted speech, while theta activity showed sustained coupling across varied conditions (Molinaro and Lizarazu 2018). Traditionally, low-frequency oscillations have been associated with sensory selection (Schroeder and Lakatos 2009). Energy resets (i.e., onsets of speech) could serve as an bottom-up update of the sensory gain to the attentive speech (Obleser and Kayser 2019) following the framework of active sensing in the auditory domain (Bajcsy 1988; Schroeder et al. 2010; Prescott et al. 2011; Bajcsy et al. 2018). Speech-tracking at energy onsets showed also higher coupling in the beta frequency range suggesting an interaction between bottom-up delta and beta, which previously was found to serve top-down predictive streams (Park et al. 2015). This seems to be specific for speech onsets as we observe low alpha and beta tracking during segments with high-theta power. We suggest that low-α during high-theta segments might reflect the cortical-excitability attenuation during high sensory gain, previously reported in an animal study (Kayser, Wilson, et al. 2015) and also related to attenuated temporal anticipation processes during sustained speech (Rohenkohl and Nobre 2011; Samaha et al. 2015).

For the latter, the high-beta tracking during high-delta segments suggests a delta-beta interaction which might underline temporal predictions during speech onsets (Arnal et al. 2015). In any case, speech perception entails temporal prediction processes (Engel et al. 2001; Arnal and Giraud

2012; Rimmele et al. 2018). As speech in natural settings does not show robust temporal rhythmicity in all timescales, acoustic edges at the delta timescale could provide the temporal coding for adjusting and re-aligning neural tracking in multiple timescales. Temporal expectations align delta oscillations in time in rhythmic (Stefanics et al. 2010; Herbst and Obleser 2019; Daume et al. 2021) and non-rhythmic tasks (Daume et al. 2021). During speech perception, delta tracking was previously disrupted when randomly altering the length of silences, yielding the speech rate temporally irregular (Kayser, Ince, et al. 2015). As temporal processing is relevant to the motor system (Chen et al. 2008), motor origins of temporal predictions in auditory tasks have also been previously described in delta band (Morillon and Baillet 2017) with relevance for speech perception (Morillon et al. 2019) suggesting a motor control over delta speech tracking, which remained out of the scope of this study and thus requires further investigation. In any case, we suggest that delta speech tracking during speech onsets employs temporal predictive processes.

Previously, delta tracking has also been implied to track linguistic elements of speech. Specifically, spectral peaks were observed corresponding to phrasal units (using isochronous syllable sequence = 1 Hz), indicating that delta oscillations may align to linguistic content, irrespective of acoustic evidence (Ding et al. 2016). Despite a lack of acoustic cues, participants were able to chunk syntactic phrases through the phase of delta oscillations (Meyer et al. 2017). delta-phase coupling to speech envelope has been previously reported (Bourguignon et al. 2013) and it was found to be stronger in forward compared to backward-played speech (Gross et al. 2013; Park et al. 2015), which also argues against purely acoustic-driven delta tracking. Though, while the energy modulations between forward and backward speech are equal, the listener's attention level – and thus speech tracking – is inherently affected (Zion Golumbic et al. 2013). Delta tracking has been associated with phrasal structure (Keitel et al. 2018), with syntactic phrases sharing common structure and timing, while being distinguishable due to prominent energy resets (i.e., pauses). Thus, it remains unclear to what extent delta activity in auditory areas

reflects knowledge-based linguistic units, or whether it is driven by certain events of the speech as stress, intonation, or pitch contour (Himmelmann et al. 2018). Disentangling purely syntactic markers from acoustic events remains challenging. Even if clear prosodic influences are absent, top-down implicit prosodic features may still influence processing (Glushko et al. 2020; Kalenkovich et al. 2022), thus interfering with bottom-up acoustic and top-down linguistic encoding of spoken narratives (Jin et al. 2020).

A limitation of our study is that we proceeded with a description of the speech signal only in terms of acoustic dynamics. As mentioned, linguistic content is also known to influence speech tracking (Peelle et al. 2013; Haegens and Zion Golumbic 2018; Rimmele et al. 2018) which might interfere with onset-related delta tracking. Further research needs to address this potential confound. Relevant to that, another drawback is the subsequent interpretation of delta and theta tracking, after the identification of higher- and lower-power segments in the frequencies of interest (0.6 and 5 Hz). Inevitably, this approach does not allow the characterization of other – temporal – features in the speech signal (besides acoustic step functions) that give rise to delta speech tracking. Furthermore, while our dataset has the advantage of high signal-to-noise ratio, it lacks experimental conditions and measures of behavioral performance, in which the relevance of high power delta events could be directly tested at the neural and behavioral level.

Our analysis focused on speech envelope without considering linguistic annotations of sentence, phrase or word onsets. Thus, we find that sharp acoustic transitions from silence to energy onsets are adequate to result in delta speech tracking in bilateral auditory cortex, irrespective of high-level constructs such as sentence, phrase or word onsets. Although, inherently in a natural speech setting (i.e listening to an audiobook) the silent pauses most frequently mark the beginning of a phrase or a sentence. In supplementary figure 1B we show that the length of silences marking sentence onsets match the speech onsets that we find to contribute to delta speech tracking. This is not a direct comparison - which remained outside the scope of this study - but provides a

valuable indication that the speech onsets described here can mark the beginning of a sentence or a phrase. We note that in a natural speech setting, the annotation of ‘phrases’ is not trivial and the length and total number can vary according to dependency definition parameters, so we mainly focused on the sentences onsets. Importantly, we find that delta speech tracking is specific to pauses before- and drops after- an onset (analysis in figure 4D). That rules out the possibility that the delta speech tracking described in the study serves as a processing mechanism of a whole sentence, as we find it to be active and temporally specific during the pauses.

In summary, our results indicate that the strongest drivers in the speech envelope for speech-brain coupling differ for delta and theta. Theta effects are strongest during sustained speech while delta effects are strongest around gaps. While this does not rule out a role of delta rhythms in chunking it calls for caution when interpreting speech-brain coupling - especially in the delta band.

Acknowledgments

We acknowledge support by the Interdisciplinary Center for Clinical Research (IZKF) of the medical faculty of Münster (grant number Gro3/001/19). We thank Pascal Nicklas for critical remarks on a previous version of the manuscript. JG was further supported by the DFG (GR 2024/5-1) and RN was supported by a grant from the German Science Foundation (CRC 1451/A07).

Conflicts of Interest

The authors declare no conflict of interest.

Bibliography

Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, Merzenich MM. 2001. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc Natl Acad Sci USA. 98:13367–13372.

Arnal LH, Doelling KB, Poeppel D. 2015. Delta-Beta Coupled Oscillations Underlie Temporal Prediction Accuracy. Cereb Cortex. 25:3077–3085.

Arnal LH, Giraud A-L. 2012. Cortical oscillations and sensory predictions. Trends Cogn Sci (Regul Ed). 16:390–398.

Bajcsy R, Aloimonos Y, Tsotsos JK. 2018. Revisiting active perception. Auton Robots. 42:177–196.

Bajcsy R. 1988. Active Perception. Proc IEEE Inst Electr Electron Eng.

Benjamini Y. 2010. Discovering the false discovery rate. J Royal Statistical Soc B. 72:405–416.

Besl PJ, McKay HD. 1992. A method for registration of 3-D shapes. IEEE Trans Pattern Anal Mach Intell. 14:239–256.

Boucher VJ, Gilbert AC, Jemel B. 2019. The Role of Low-frequency Neural Oscillations in Speech Processing: Revisiting Delta Entrainment. J Cogn Neurosci. 31:1205–1215.

Bourguignon M, De Tiège X, de Beeck MO, 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. Hum Brain Mapp. 34:314–326.

Brainard DH. 1997. The Psychophysics Toolbox. Spat Vis. 10:433–436.

Brodbeck C, Simon JZ. 2020. Continuous speech processing. Curr Opin Physiol. 18:25–31.

Chalas N, Daube C, Kluger DS, Abbasi O, Nitsch R, Gross J. 2022. Multivariate analysis of speech envelope tracking reveals coupling beyond auditory cortex. Neuroimage. 258:119395.

Chen JL, Penhune VB, Zatorre RJ. 2008. Listening to musical rhythms recruits motor regions of the brain. Cereb Cortex. 18:2844–2854.

Daube C, Ince RAA, Gross J. 2019. Simple Acoustic Features Can Explain Phoneme-Based Predictions of Cortical Responses to Speech. Curr Biol. 29:1924-1937.e9.

Daume J, Wang P, Maye A, Zhang D, Engel AK. 2021. Non-rhythmic temporal prediction involves phase resets of low-frequency delta oscillations. Neuroimage. 224:117376.

Davis MH, Ford MA, Kherif F, Johnsrude IS. 2011. Does semantic context benefit speech understanding through “top-down” processes? Evidence from time-resolved sparse fMRI. J Cogn Neurosci. 23:3914–3932.

Dilley LC, Pitt MA. 2010. Altering context speech rate can cause words to appear or disappear. Psychol Sci. 21:1664–1670.

Ding N, Melloni L, Zhang H, Tian X, Poeppel D. 2016. Cortical tracking of hierarchical linguistic structures in connected speech. Nat Neurosci. 19:158–164.

Ding N, Patel AD, Chen L, Butler H, Luo C, Poeppel D. 2017. Temporal modulations in speech and music. Neurosci Biobehav Rev. 81:181–187.

Ding N, Simon JZ. 2014. Cortical entrainment to continuous speech: functional roles and interpretations. Front Hum Neurosci. 8:311.

Doelling KB, Arnal LH, Ghitza O, Poeppel D. 2014. Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. Neuroimage. 85 Pt 2:761–768.

Engel AK, Fries P, Singer W. 2001. Dynamic predictions: oscillations and synchrony in top-down processing. Nat Rev Neurosci. 2:704–716.

Giraud A-L, Poeppel D. 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci. 15:511–517.

Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, Ugurbil K, Andersson J, Beckmann CF, Jenkinson M, Smith SM, Van Essen DC. 2016. A multi-modal parcellation of human cerebral cortex. Nature. 536:171–178.

Glushko A, Poeppel D, Steinhauer K. 2020. Overt and covert prosody are reflected in

neurophysiological responses previously attributed to grammatical processing. BioRxiv.

Gross J, Hoogenboom N, Thut G, Schyns P, Panzeri S, Belin P, Garrod S. 2013. Speech rhythms and multiplexed oscillatory sensory coding in the human brain. PLoS Biol. 11:e1001752.

Gross J, Kluger DS, Abbasi O, Chalas N, Steingräber N, Daube C, Schoffelen J-M. 2021. Comparison of undirected frequency-domain connectivity measures for cerebro-peripheral analysis. Neuroimage. 245:118660.

Haegens S, Zion Golumbic E. 2018. Rhythmic facilitation of sensory processing: A critical review. Neurosci Biobehav Rev. 86:150–165.

Hamilton LS, Edwards E, Chang EF. 2018. A spatial map of onset and sustained responses to speech in the human superior temporal gyrus. Curr Biol. 28:1860-1871.e4.

Herbst SK, Obleser J. 2019. Implicit temporal predictability enhances pitch discrimination sensitivity and biases the phase of delta oscillations in auditory cortex. Neuroimage. 203:116198.

Himmelmann NP, Sandler M, Strunk J, Unterladstetter V. 2018. On the universality of intonational phrases: a cross-linguistic interrater study. Phonology. 35:207–245.

Hyafil A, Fontolan L, Kabdebon C, Gutkin B, Giraud A-L. 2015. Speech encoding by coupled cortical theta and gamma oscillations. eLife. 4:e06213.

Ince RAA, Giordano BL, Kayser C, Rousselet GA, Gross J, Schyns PG. 2017. A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula. Hum Brain Mapp. 38:1541–1573.

Jin P, Lu Y, Ding N. 2020. Low-frequency neural activity reflects rule-based chunking during speech listening. eLife. 9.

Kalenkovich E, Shestakova A, Kazanina N. 2022. Frequency tagging of syntactic structure or lexical properties; a registered MEG study. Cortex. 146:24–38.

Kayser C, Wilson C, Safaai H, Sakata S, Panzeri S. 2015. Rhythmic auditory cortex activity at multiple timescales shapes stimulus-response gain and background firing. J Neurosci.

35:7750–7762.

Kayser SJ, Ince RAA, Gross J, Kayser C. 2015. Irregular speech rate dissociates auditory cortical entrainment, evoked responses, and frontal alpha. J Neurosci. 35:14691–14701.

Keitel A, Gross J, Kayser C. 2018. Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. PLoS Biol. 16:e2004473.

Keitel A, Gross J. 2016. Individual Human Brain Areas Can Be Identified from Their Characteristic Spectral Activation Fingerprints. PLoS Biol. 14:e1002498.

Kiebel SJ, Daunizeau J, Friston KJ. 2008. A hierarchy of time-scales and the brain. PLoS Comput Biol. 4:e1000209.

Koning R, Wouters J. 2012. The potential of onset enhancement for increased speech intelligibility in auditory prostheses. J Acoust Soc Am. 132:2569–2581.

Lakatos P, Gross J, Thut G. 2019. A new unifying account of the roles of neuronal entrainment. Curr Biol. 29:R890–R905.

Luo H, Poeppel D. 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron. 54:1001–1010.

Maris E, Oostenveld R. 2007. Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods. 164:177–190.

Meyer L, Henry MJ, Gaston P, Schmuck N, Friederici AD. 2017. Linguistic Bias Modulates Interpretation of Speech via Neural Delta-Band Oscillations. Cereb Cortex. 27:4293–4302.

Meyer L, Sun Y, Martin AE. 2020. “Entraining” to speech, generating language? Lang Cogn Neurosci. 1–11.

Molinaro N, Lizarazu M. 2018. Delta(but not theta)-band cortical entrainment involves speech-specific processing. Eur J Neurosci. 48:2642–2650.

Morillon B, Arnal LH, Schroeder CE, Keitel A. 2019. Prominence of delta oscillatory rhythms in the motor cortex and their relevance for auditory and speech perception. Neurosci Biobehav Rev. 107:136–142.

Morillon B, Baillet S. 2017. Motor origin of temporal predictions in auditory attention. Proc Natl Acad Sci USA. 114:E8913–E8921.

Nolte G. 2003. The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. Phys Med Biol. 48:3637–3652.

Obleser J, Kayser C. 2019. Neural entrainment and attentional selection in the listening brain. Trends Cogn Sci (Regul Ed). 23:913–926.

Oganian Y, Chang EF. 2019. A speech envelope landmark for syllable encoding in human superior temporal gyrus. Sci Adv. 5:eaay6279.

Oostenveld R, Fries P, Maris E, Schoffelen J-M. 2011. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci. 2011:156869.

Panzeri S, Brunel N, Logothetis NK, Kayser C. 2010. Sensory neural codes using multiplexed temporal scales. Trends Neurosci. 33:111–120.

Park H, Ince RAA, Schyns PG, Thut G, Gross J. 2015. Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. Curr Biol. 25:1649–1653.

Peelle JE, Gross J, Davis MH. 2013. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cereb Cortex. 23:1378–1387.

Poeppel D, Assaneo MF. 2020. Speech rhythms and their neural foundations. Nat Rev Neurosci. 21:322–334.

Poeppel D. 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time.’ Speech Commun. 41:245–255.

Prescott TJ, Diamond ME, Wing AM. 2011. Active touch sensing. Philos Trans R Soc Lond B Biol Sci. 366:2989–2995.

Rimmele JM, Morillon B, Poeppel D, Arnal LH. 2018. Proactive sensing of periodic and aperiodic

559 auditory patterns. Trends Cogn Sci (Regul Ed). 22:870–882.
560 Rimmele JM, Poeppel D, Ghitza O. 2021. Acoustically driven cortical delta oscillations underpin
561 prosodic chunking. eNeuro.
562 Rohenkohl G, Nobre AC. 2011. α oscillations related to anticipatory attention follow temporal
563 expectations. J Neurosci. 31:14076–14084.
564 Rosen S. 1992. Temporal information in speech: acoustic, auditory and linguistic aspects. Philos
565 Trans R Soc Lond B Biol Sci. 336:367–373.
566 Samaha J, Bauer P, Cimaroli S, Postle B. 2015. Top-down control of the phase of alpha-band
567 oscillations as a mechanism for temporal prediction. Proc Natl Acad Sci USA. 112:E6410–
568 E6410.
569 Schädler MR, Meyer BT, Kollmeier B. 2012. Spectro-temporal modulation subspace-spanning
570 filter bank features for robust automatic speech recognition. J Acoust Soc Am. 131:4134–
571 4151.
572 Schroeder CE, Lakatos P. 2009. Low-frequency neuronal oscillations as instruments of sensory
573 selection. Trends Neurosci. 32:9–18.
574 Schroeder CE, Wilson DA, Radman T, Scharfman H, Lakatos P. 2010. Dynamics of Active
575 Sensing and perceptual selection. Curr Opin Neurobiol. 20:172–176.
576 Shannon CE. 1948. A mathematical theory of communication. Bell System Technical Journal.
577 27:379–423.
578 Stefanics G, Hangya B, Hernádi I, Winkler I, Lakatos P, Ulbert I. 2010. Phase entrainment of
579 human delta oscillations can mediate the effects of expectation on reaction speed. J
580 Neurosci. 30:13578–13585.
581 Van Veen BD, van Drongelen W, Yuchtman M, Suzuki A. 1997. Localization of brain electrical
582 activity via linearly constrained minimum variance spatial filtering. IEEE Trans Biomed Eng.
583 44:867–880.
584 Zellner B. 1994. Pauses and theTemporal Structureof Speech. In: Fundamentals of speech

synthesis and speech recognition. p. 41–62.

Zion Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, Goodman RR, Emerson R, Mehta AD, Simon JZ, Poeppel D, Schroeder CE. 2013. Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party”. Neuron. 77:980–991.