NeuroImage 258 (2022) 119395

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Multivariate analysis of speech envelope tracking reveals coupling beyond auditory cortex

Nikos Chalas[a,b,*], Christoph Daube[c], Daniel S. Kluger[a,b], Omid Abbasi[a], Robert Nitsch[d], Joachim Gross[a,b]

[a] Institute for Biomagnetism and Biosignal Analysis, University of Münster, Münster, Germany
[b] Otto-Creutzfeldt-Center for Cognitive and Behavioral Neuroscience, University of Münster, Münster, Germany
[c] Centre for Cognitive Neuroimaging, University of Glasgow, Glasgow, UK
[d] Institute for Translational Neuroscience, University of Münster, Münster, Germany

ARTICLE INFO

Keywords:
Speech–brain coupling
Mutual-information
Magnetoencephalography
Speech Perception

ABSTRACT

The systematic alignment of low-frequency brain oscillations with the acoustic speech envelope signal is well established and has been proposed to be crucial for actively perceiving speech. Previous studies investigating speech-brain coupling in source space are restricted to univariate pairwise approaches between brain and speech signals, and therefore speech tracking information in frequency-specific communication channels might be lacking. To address this, we propose a novel multivariate framework for estimating speech-brain coupling where neural variability from source-derived activity is taken into account along with the rate of envelope's amplitude change (derivative). We applied it in magnetoencephalographic (MEG) recordings while human participants (male and female) listened to one hour of continuous naturalistic speech, showing that a multivariate approach outperforms the corresponding univariate method in low- and high frequencies across frontal, motor, and temporal areas. Systematic comparisons revealed that the gain in low frequencies (0.6 - 0.8 Hz) was related to the envelope's rate of change whereas in higher frequencies (from 0.8 to 10 Hz) it was mostly related to the increased neural variability from source-derived cortical areas. Furthermore, following a non-negative matrix factorization approach we found distinct speech-brain components across time and cortical space related to speech processing. We confirm that speech envelope tracking operates mainly in two timescales (δ and θ frequency bands) and we extend those findings showing shorter coupling delays in auditory-related components and longer delays in higher-association frontal and motor components, indicating temporal differences of speech tracking and providing implications for hierarchical stimulus-driven speech processing.

1. Introduction

Our senses are confronted with signals often exhibiting regular or semi-regular patterns over time. Similarly, fluctuations of synchronized excitatory and inhibitory cortical activity drive rhythmic patterns of brain activity (Bishop, 1932; Buzsáki and Draguhn, 2004) which modulate the processing of incoming sensory signals (Arieli et al., 1996; Romei et al., 2010). Brain dynamics can temporally align to the rhythmic structure of sensory signals (Henry and Obleser, 2012; Obleser and Kayser, 2019; Schroeder and Lakatos, 2009), a process that is considered to facilitate structuring and gating of incoming information (Arieli et al., 1996; Lakatos et al., 2019; Romei et al., 2010) while forming predictions about future incoming events in space and time (Arnal and Giraud, 2012). In the case of audition, rhythmic auditory input can coordinate the phase of ongoing oscillations in the auditory cortex (Lakatos et al., 2005) at multiple timescales (Panzeri et al., 2010), reflecting functionally-distinct mechanisms (Ding and Simon, 2014). Importantly, 'neural tracking' of stimuli shows regionally specific delays, providing evidence for the timing and neural processing of sensory events (Brasselet et al., 2012; Johnston and Nishida, 2001; Zeki and Bartels, 1998).

Similarly, during continuous speech comprehension, brain activity dynamically aligns to quasi-rhythmic acoustic fluctuations (Poeppel and Assaneo, 2020) through the phase of low-frequency oscillations (Luo and Poeppel, 2007). Speech-brain coupling has been prominently observed in the δ (below 4 Hz) and θ (4 - 7 Hz) frequency range (Ahissar et al., 2001; Ding and Simon, 2014; Jin et al., 2020; Kayser et al., 2015; Luo and Poeppel, 2007) and it has been proposed to serve critical computations for speech comprehension including segmenting and decoding of acoustic features (Giraud and Poeppel, 2012;

[*] Correspondence author at: Nikos Chalas, University of Münster, Institute for Biomagnetism and Biosignal Analysis, Malmedyweg 15, 48149 Münster, Germany.
E-mail address: nchalas@uni-muenster.de (N. Chalas).

https://doi.org/10.1016/j.neuroimage.2022.119395.
Received 8 March 2022; Received in revised form 16 May 2022; Accepted 14 June 2022
Available online 16 June 2022.

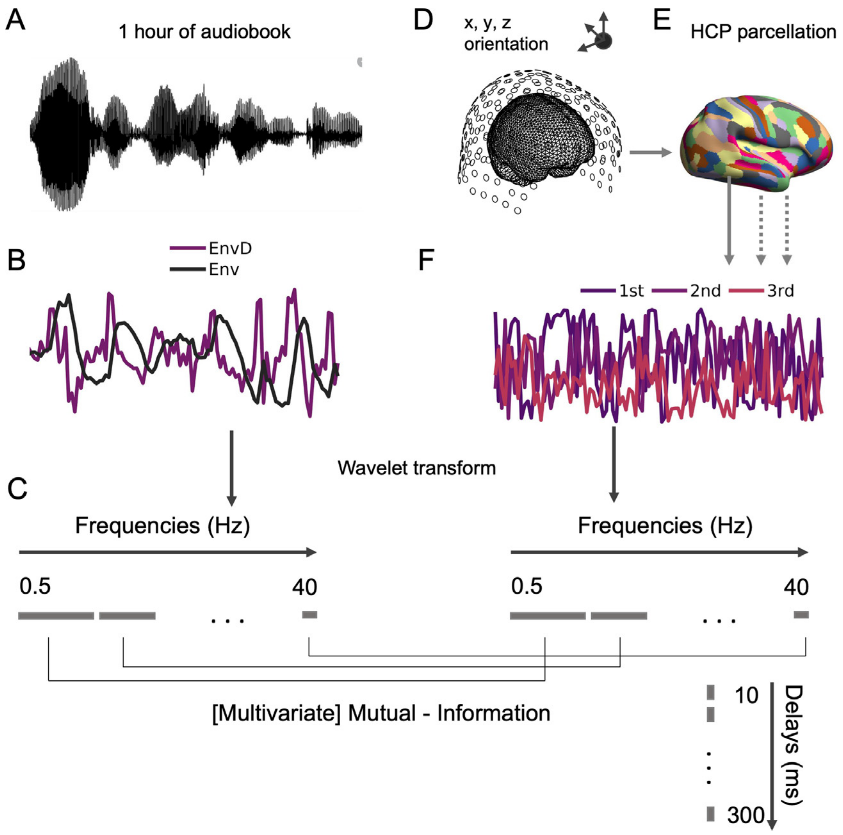

Figure 1. Multivariate speech tracking pipeline. (A) Participants (n = 24) listened to ~1 hour of an audiobook, divided into 6 blocks, while Magnetoencephalographic (MEG) measurements were acquired. (B) Amplitude envelope (black) and its derivative (purple) were extracted from the continuous speech signal (see Methods). (C) Individual MRIs were used to estimate source models per participant which were interpolated to a template volumetric grid. (D) Cortical areas were divided into 362 anatomical parcels according to the parcellation from the Human Connectome Project (Glasser et al., 2016) (E) For each parcel, estimated source time-series were extracted. (F) Acoustic signals (amplitude envelope and its derivative, see (B)) and source activity (three time-series per parcel, see Methods) were subjected to a continuous Morlet-transformation (0.5 - 40 Hz). After the transformation to frequency space, we computed Mutual Information between multivariate time-series (acoustic signals and source-time series per parcel; see Methods for details).

Peelle and Davis, 2012; Rimmele et al., 2021). Notably, while the spectral content of speech is crucial for comprehension (Lorenzi et al., 2006; Obleser et al., 2012; Scott and McGettigan, 2012), recent work proposed that disrupting entrainment through electrical stimulation leads to compromised speech intelligibility (Asamoah et al., 2019; Riecke et al., 2018; Wilsch et al., 2018; Zoefel et al., 2018). Speech tracking is also modulated by attention (Obleser and Kayser, 2019; Zion Golumbic et al., 2013) and correlated with semantic context (Broderick et al., 2019; Koskinen et al., 2020), indicating an interaction between bottom-up (that is, feedforward) and top-down (that is, feedback) processes (Assaneo et al., 2019; Barczak et al., 2018). Still, a complete whole brain characterization of speech tracking across relevant frequencies, delays, and cortical areas is lacking.

To date, studies investigating speech-brain coupling in cortical space have used a univariate pairwise approach between source estimates and speech signals. Voxel-based studies typically use the dominant source direction for each voxel and atlas-based studies often use the first component of a principal component analysis (PCA) of all voxel time series for each atlas parcel. Both approaches can miss relevant information as there is no fundamental justification to the assumption that the strongest source orientation (which is based on power) or the strongest PCA component (again based on power) captures the brain activity with the strongest coupling to the speech envelope (Jaworska et al., 2022). Here, we aimed to alleviate this limitation by using a novel atlas-based multivariate approach.

First, we aimed to provide a comprehensive characterisation of whole-brain speech tracking in data with high signal-to-noise ratios based on a novel multivariate mutual information analysis (Ince et al., 2017). To this end, we developed an analytical multivariate approach in which neural variability across and within parcels was taken into account for estimation of speech-brain coupling along with information from speech envelope (Env) and its rate of change (derivative; EnvD, see Figure 1). We show that multivariate speech tracking is superior to univariate speech tracking in all brain areas and helps to uncover the diverse spectro-temporal structure of speech tracking across cortical areas. Then, we followed an unsupervised non-negative matrix factorization (NMF) approach to uncover spectral components across cortical areas and temporal delays coupled to the acoustic envelope in passive listening using MEG (Baillet, 2017; Gross, 2019).

2. Methods

2.1. Participants and data acquisition

A total of 24 volunteers (12 females; mean age = 24.0 years, age range 18-35 years) participated in this study. The study was approved by the College of Science and Engineering Ethics Committee at the University of Glasgow (application number: 300170024). While participants listened to a 55 minute duration audiobook, brain activity was monitored with a 248-magnetometer whole-head MEG system (MAGNES 3600, 4-D Neuroimaging) in a magnetically-shielded room. 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 six blocks of ~9.16 mins. To allow participants to better comprehend the story, the last 10 seconds of each block were repeated in the following block. In the case that head movement exceeded 5mm for a block, measurement was repeated. 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 had to answer 18 multiple choice questions (with three response options each) with the number of correct options varying between 1-3 per question (mean performance 0.95; SD 0.05; range 0.78-1). A different analysis of this dataset has been reported elsewhere (Daube et al., 2019).

2.2. Speech envelope extraction

We extracted the amplitude envelope from the continuous speech signal. To this end, 31-channel Log-Mel-Spectograms (124.1 Hz - 7284.1

Hz) were computed and absolute values were summed across bands to obtain a wideband speech envelope (Schädler et al., 2012).

2.3. Data preprocessing

MEG data were processed using the FieldTrip toolbox (Oostenveld et al., 2011) for MATLAB 2021a (The MathWorks, Inc.) and in-house MATLAB routines. We note that data were preprocessed again using the same scripts used for Daube et al., (2019). We briefly mention here that for each block, continuous data starting at the onset of the story were denoised using the denoise_pca function of FieldTrip where bad channels were manually detected and spherical-spline interpolated from neighboring channels (mean number of rejected channels per block M = 3.07; SD = 3.64). Squid jumps were replaced with DC patches. Continuous data were filtered offline with a fourth-order forward-reverse zero-phase Butterworth high-pass filter with a cutoff-frequency of 0.5 Hz and downsampled to 100 Hz for computational efficiency. Independent components (mean number of rejected components per block M = 5; SD = 5.3) arising from heartbeats and eye movements were visually isolated and removed using the runica ICA algorithm.

2.4. Source localization

Individual T1-weighted MRIs were coregistered in the MEG coordinate system, aligned with the digitized head shapes using the iterative closest point algorithm (Besl and McKay, 1992), and segmented (into white matter, gray matter, and cerebrospinal fluid) for generating single-shell volume conductor models (Nolte, 2003). For group analyses, individual MRIs were linearly transformed to a MNI template provided by FieldTrip. Source activity was estimated computing LCMV beamformer coefficients from the MEG time-series for each voxel on a 5mm grid (Van Veen et al., 1997). The sensor covariance matrix used was computed across all trials. The lambda regularization parameter was set to 0% and time series were extracted for each dipole orientation, resulting in three time-series per voxel. To reduce the dimensionality of the data, we applied an atlas-based parcellation of cortical space, resulting in 181 ROIs per hemisphere (Glasser et al., 2016). Source time-series for each parcel were concatenated across voxels and orientations and we extracted the principal components, along with their explained variance.

2.5. Multivariate Mutual Information

Statistical dependencies between the speech envelope together with its first derivative (Ince et al., 2017) and the source space parcels were computed on the basis of information theory (Shannon, 1948). We estimated mutual information (MI) using Gaussian Copula MI (GCMI) between multivariate speech signals and source time-series (Ince et al., 2017). GCMI was estimated for various delays (-300ms to 300ms, steps of 10ms). To identify frequency-specific interactions, we applied a continuous wavelet transformation (CWT; cwtfilterbank.m in MATLAB; wt.m performs the actual transformation into the frequency domain) for 64 frequencies (from 0.1 Hz to 40 Hz). With L1-normalization implemented in the algorithm, equal amplitude oscillatory components across different scales have equal magnitude in the CWT, providing a more accurate depiction of the signal.

Our multivariate analysis capitalized on the inherent ability of GCMI to estimate dependencies between multidimensional data (Ince et al., 2017). We included three source time-series per parcel and two speech signals (envelope and derivative), making the estimation multivariate (3×2 analytical framework) for each parcel and frequency. While we are not aware of any multivariate methods quantifying speech-brain coupling in the frequency domain, multivariate methods based on regression have been used in the time domain for this type of analysis (Crosse et al., 2016; Daube et al., 2019). Our multivariate framework differs from these methods in three important ways: First, as mentioned, our method operates in the frequency domain and not the time domain. Second, our analysis is based on mutual information and therefore sensitive to nonlinear dependencies between speech and brain signals in contrast to the linear regression models. Third, regression models work on n x 1 data meaning that decoding models reconstruct 1-dimensional stimulus data from n-dimensional brain data or encoding models predict 1-dimensional brain data from n-dimensional stimulus data. Our framework operates on n x m data and allows the quantification of dependencies between n-dimensional stimulus data and m-dimensional brain data.

For the GCMI estimation we additionally computed 500 surrogate MI computations on the basis of random temporal shifting of the speech signal with respect to the source time-series via a circular wrapping around the edges as proposed in (Andrzejak et al., 2003). This way, we created a distribution of 500 surrogate MI values for each frequency, delay, and parcel. From the surrogate distribution we obtained normalized MI values for each frequency and parcel by subtracting the mean of the distribution and dividing it by the standard deviation, thus correcting for auto-correlation across frequencies. We obtained normalized MI values for each participant, frequency, delay, and parcel. Significance of normalized MI values at the group level was determined with cluster-based permutation tests (Maris and Oostenveld, 2007) comparing empirical MI values with the 95th percentile of the 500 surrogate distribution. We note that in statistical comparisons maximum MI values across delays were used. This includes multiple steps: a series of one-tailed t-tests of individual MI per parcel and frequency were conducted and thresholded at $p = 0.05$. This included 64×61×362 comparisons (Frequency x Delays x Parcels). To control for multiple testing a non-parametric cluster-based permutation test was applied. For that, spectro- and spatial- adjacent data were clustered together and assigned a cluster-level statistic depicting the sum of t-values within each cluster. Then, each cluster was subjected to significance testing through Monte-Carlo approximation. For that, individual MI spectra were randomly interchanged with the 95th percentile of the initial surrogate distribution and t-tests were recomputed in the cluster-level. This procedure was applied 5000 times and then the original cluster-statistics were compared with the distribution of the 5000 randomized null-statistics distribution. Significance for the original clusters was reached when they exhibited a higher test statistic than 95% of the randomized null data.

2.6. Non-negative Matrix Factorization

As MI values are inherently positive, we sought to further describe the spatial, spectral, and temporal characteristics of cortical responses to continuous speech. To this end, we applied non-negative matrix factorization (NMF) to mutual information values across parcels, frequencies, and delays. NMF factors a n-by-m matrix A into W (n-by-k) and H (k-by-m) factors, such that the root mean square between A and WxH is minimized (Berry et al., 2007). This method has been used previously in speech processing (Hamilton et al., 2018) and for identifying distinct spectral and spatial patterns of source-reconstructed data (Ince et al., 2015; Kluger and Gross, 2020; Schoffelen et al., 2017) as it extracts features to reduce the dimensionality of the data, while preserving distinct profiles across a predefined number of components. The optimal number of components (k) to be extracted from the NMF was determined with a rank optimization using singular value decomposition (Qiao, 2015), resulting in 15 components.

NMF is implemented as an iterative optimisation, where convergence may vary according to random initial values. To ensure reproducibility, we repeated the NMF estimation 300 times. Each time, the NMF was initiated using the multiplicative algorithm with 10 iterations and the best solutions based on residuals were used as starting points for 1000 more NMF iterations using the alternating least squares algorithm. This combination of algorithms was applied to make use of the computational efficiency of the multiplicative algorithm and stability in convergence

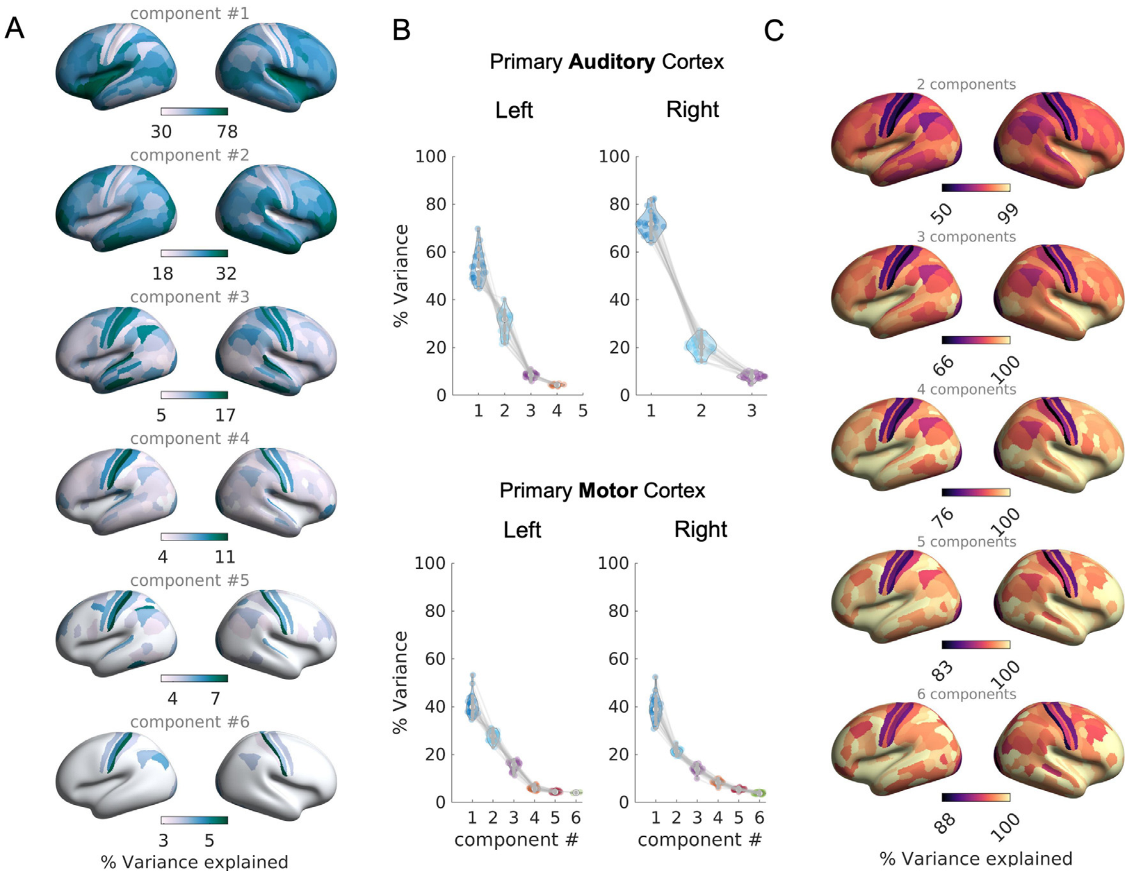

Figure 2. Explained variance of principal components in cortical space. (A) Grand-average of cortical maps depicting the amount of variance (%) explained for components 1 - 6 after principal component analysis of source time-series of all voxels within a parcel (B) Amount of total variance (%) explained per participant in Left Primary Auditory Cortex (*Upper Left*), Right Primary Auditory (*Upper Right*), Left Primary Motor Cortex (*Bottom Left*), Right Motor Auditory (*Bottom Right*) (C) Cortical maps of grand-average of cumulative variance explained per parcel with 2 to 6 components (from top to bottom)

of the alternating least squares algorithm. From this procedure we chose the best solution based on the residuals. Visual inspection of the first 10 best solutions confirmed the similarity between solutions and thus the reproducibility of the method applied.

NMF components of Frequency x Delays x Parcels were estimated per subject. For identifying group-level consistent effects we applied a cluster-based permutation statistical analysis of the spectral profiles for each component. Statistical significance was determined with one-sample t-tests, after permuting 5000 times the Frequency x Delays x Parcels matrix and thresholding it at $p < 0.05$ (FDR – corrected).

3. Results

3.1. *Multivariate speech tracking: Differential modulation across cortical areas*

Before estimating speech-brain tracking, we wanted to validate the amount of variance that the principal components time-series capture after the PCA of source-time series within each parcel. Specifically, in each parcel we stacked the time series (corresponding to the three source orientations [x,y,z]) across all voxels in this parcel and computed PCA components of this matrix. We used the grand average of variance per components, parcels, and participants. In Figure 2 we summarize the main findings. It is evident that the first component captures maximum variance (up to 80%) in early auditory areas but only around 40% of the total variance for higher association, frontal, and motor areas (Figure 2a). Further investigation of the variance explained for the primary auditory and motor cortex reveals that at least the first three components contribute non-negligible variance to the total signal (Figure 2b). More importantly, when the first three components are considered, it is sufficient to capture 66-100% of total variance in cortical space (Figure 3c).

Thus, we simultaneously used three time series to represent each brain area. The resulting time series represent the optimal (in the sense of explained variance) three-dimensional representation of brain activity in a given parcel resulting from a weighted mixing of time series across different voxels and orientations For the speech signal we used the speech envelope and its derivative. Although the speech envelope is critical for comprehension, recent evidence suggests that acoustic landmarks (local maxima in the envelope rate of change) are encoded in the human superior temporal gyrus (Hertrich et al., 2012; Oganian and Chang, 2019). Thus, along with the speech envelope, we included the rate of amplitude change, estimated by the first derivative. In summary, we followed an information-theoretic approach following previously validated and systematically compared approaches (Gross et al., 2021), quantifying multivariate mutual-information (MI) between speech signals (Env - EnvRate) and three source estimates per parcel (3 PCA components from 362 parcels; parcellation by (Glasser et al. 2016)).

As we were interested in frequency-specific speech tracking, we transformed source estimates and speech signals with a continuous wavelet transform from 0.1-40 Hz before computing MI. We tested phase alignment of cortical areas to speech signals and compared it to the 95th percentile of a surrogate distribution (see Methods).

Figure 3 summarizes the group level results obtained with our multivariate MI analysis. First, areas that are phase-aligned to speech signals (0.1-15 Hz) were localized to broad clusters including areas in temporal, parietal, frontal, and motor cortex in both hemispheres (shown as t-values, $p < .05$, cluster corrected, see Figure 3c and Methods section). MI spectra show stronger phase alignment in the right hemisphere, while

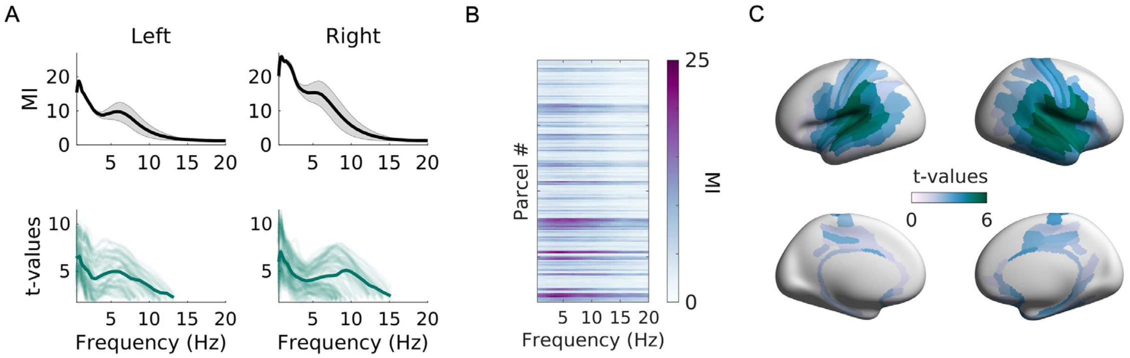

Figure 3. Significant speech tracking. (A) Upper graphs show the averaged mutual information spectra of significant parcels for the left *(Left)* and right (*Right)* hemisphere after a non-parametric cluster permutation test, comparing normalized mutual information values with the 95th percentile of the surrogate distribution. For the comparison, maximum MI values across computed delays were selected. Shaded area depicts the bootstrapped standard deviation around the mean. Accordingly, lower graphs show t-values belonging to significant clusters (see Methods; n = 184, p < .05). Lines depict t-values for each parcel which are statistically significant (p < .05, cluster corrected) and bold lines depict the average (B) Grand-average of MI across significant parcels from panel (A) and frequencies at a fixed delay of 100 ms. (C) Cortical maps of t-values averaged across significant frequencies.

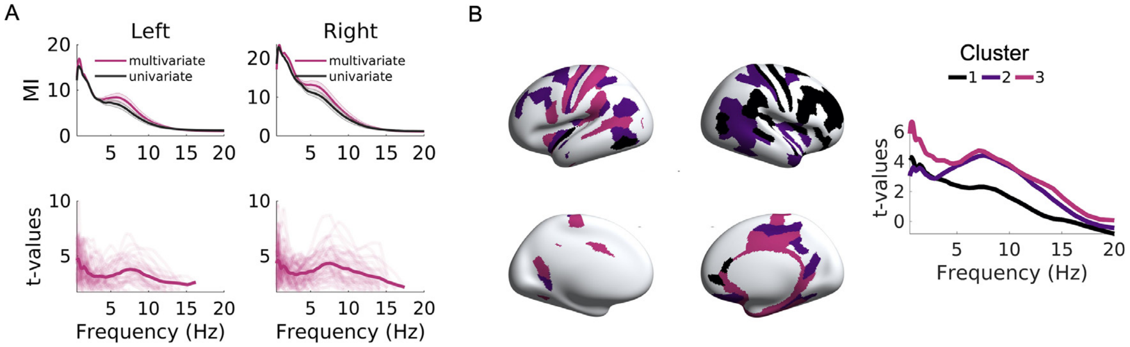

Figure 4. Multivariate versus univariate speech-brain coupling. (A) Mutual information spectra of significant parcels with one PCA component (univariate; black) and three PCA components (multivariate; red), averaged across participants. Here, multivariate refers to the neural time-course at each parcel [three PCA components (multivariate); 1 PCA component (univariate)]. Shaded area depicts the bootstrapped standard deviation around the mean. On the bottom, t-values of significant clusters for the left and right hemisphere (*Upper)*. Thin lines depict t-values that belong to significant clusters (see Methods; n = 122, p < .05) for each parcel and thick lines represent the average. Lines depict t-values for each parcel which are statistically significant, (p < .05, cluster corrected). For the comparison, maximum MI values across computed delays were selected (B) Cortical maps of averaged t-values across frequencies. Significant areas and spectra were clustered with k-means (k = 3; see Methods). Colormap indicates the clusters in the surface area. On the right, clustered mutual information spectra with k-means, extracted from significant t-values.

both hemispheres exhibit spectral peaks at around 1 and 6 Hz (Figure 3a, top and Figure 4b). Additionally, spectra of t-values in significant brain areas confirm the phase synchronization in low δ and θ frequency bands, while visual inspection of single-parcel spectra show different spectral peaks, indicating differential modulations of cortical oscillations across areas (Figure 3a, bottom).

Our approach replicates previous findings showing phase alignment of cortical oscillations in low δ and θ bands (1 and 6 Hz), lateralized to the right hemisphere (Boemio et al., 2005; Gross et al., 2013; Luo and Poeppel, 2007) and extends those findings by revealing differential modulations of cortical oscillations across frequencies and areas. Before we describe these differential modulations in more detail, we first compare the multivariate approach to a standard univariate approach for a better understanding of the information we gain with multivariate data.

3.2. *Neural variability in cortical components captures speech-tracking*

We hypothesized that accounting for higher-dimensional representations in each brain area would improve our estimate of coupling between neural and speech signals in frequencies and areas relevant to speech tracking compared to univariate analyses. We further hypothesized that this improvement is not uniform across frequencies and cortical areas. Instead, we expected that brain areas showing more complex brain activity patterns will benefit more from the higher dimensionality used in the multivariate analysis. Therefore, we computed cortical maps quantifying the difference between multivariate and univariate analysis. The analysis is based on a statistical comparison on the multivariate MI maps used in the previous section (based on three PCA components) and univariate results where only the first PCA component is used for each anatomical parcel.

In Figure 4a, we show the normalized MI values for the multivariate and univariate approach (top panels) and the corresponding t-spectra from parcels where coupling with the speech envelope is significantly higher in multivariate compared to univariate analysis (left hemisphere: 0.1-12 Hz, 52 parcels; right hemisphere: 0.1-14.1 Hz, 70 parcels, group statistics; p < 0.05 cluster-corrected; bottom panels). Visual inspection of t-value spectra indicates that the benefit of our multivariate approach differed across parcels. As we wanted to further unravel their spectral characteristics, we proceeded with an unsupervised clustering of spectra to an optimal number of clusters (k = 3; see Methods). In Figure 4b we show spectra of three corresponding clusters (bottom) and their rendering on the cortical surface (top). We find that bilateral auditory and left motor areas (Cluster #1) show increased MI values both for low frequencies (peaks at 0.8 and 1.2 Hz) and higher frequencies (peak at 7

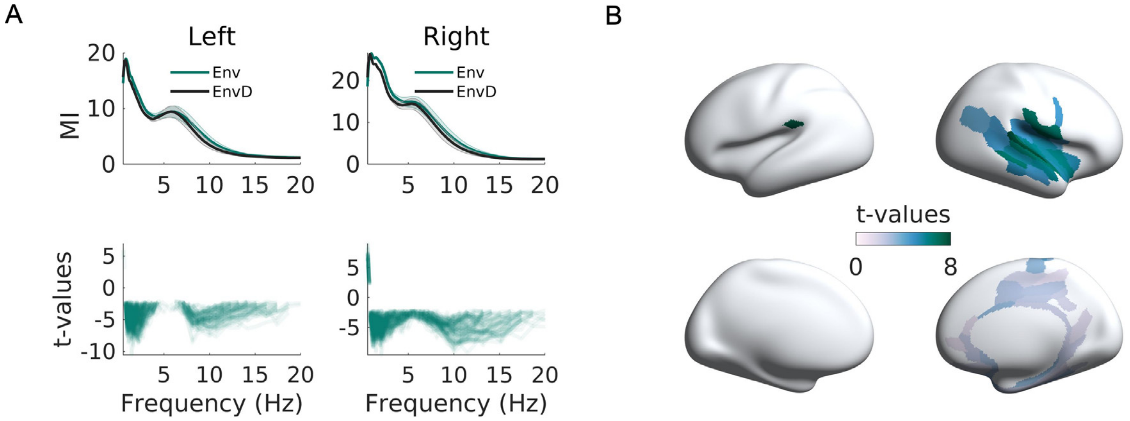

Figure 5. Comparing speech tracking using the derivative vs the envelope. (A) At the top row, averaged mutual information spectra of significant parcels with the envelope (green) and the derivative of the envelope (black). Shaded area depicts the bootstrapped standard deviation around the mean. At the bottom, t-value spectra for the left and right hemisphere. Only significant t-values that reached significance at the cluster level (n = 33, cluster p < .05; see Methods) are shown (p < .05, cluster corrected). For the comparison, maximum MI values across computed delays were selected (B) Cortical maps of t-values averaged across frequencies where derivative > envelope.

Hz). Bilateral motor, temporal, and frontal areas exhibit a spectral peak at 7 Hz (Cluster #2), whereas prefrontal and visual areas show no peak within the significant frequencies (Cluster #3).

Thus, we find that the investigation of speech tracking benefits from the inclusion of more components in the estimation of speech-tracking, as it includes more neuronal activity relevant for speech tracking. At the same time, there is no disadvantage of applying multivariate speech tracking analysis since –as expected– we did not find significantly reduced speech-tracking for multivariate compared to univariate coupling in any brain parcel.

3.3. Rate of amplitude change in the envelope drives low-frequency oscillations

Having established the benefit of using multivariate brain signals for speech tracking analysis, we proceeded with a similar analysis for the speech signal. Figure 1 illustrates that our multivariate approach represents the speech signal with both the speech envelope (Env) and the derivative of the envelope (EnvD). Here, we aimed to contrast the individual contributions of both signals (Env and EnvD) to speech tracking. As rapid changes in the speech envelope (corresponding to peaks in EnvD) provide important temporal cues, we hypothesized that these acoustic landmarks will coordinate excitatory states of ongoing neural activity by re-aligning the phase of low-frequency oscillations. This realignment would not be evident considering only the amplitude modulation of speech represented in Env. Thus, we expected to find increased phase alignment of slow oscillatory activity when using EnvD compared to Env.

Consequently, we statistically compared normalized MI values based on Env to those based on EnvD using a two-tailed nonparametric cluster test (see Methods). Results of this analysis are summarized in Figure 5. We found acoustic landmarks (EnvD) to modulate the phase of low-frequency oscillations in temporal areas of the right hemisphere significantly stronger than the speech envelope (Env; 0.1-0.8 Hz; group statistics; p < 0.05 cluster-corrected; see Figure 4b). Interestingly, this effect is localized to these low frequencies whereas higher frequencies (0.8-20 Hz) show the opposite effect. Here, brain activity shows higher phase synchronization with speech envelope compared to its derivative.

In sum, we find evidence that acoustic landmarks in the broadband speech signal provide temporal evidence to ongoing low-frequency oscillations (below 1 Hz) in the right auditory areas, indicating a phase realignment. For the rest of the spectra (0.8-20 Hz) we find that the phase of cortical oscillations is modulated to a stronger extent by the speech envelope.

3.4. Cortical coupling to speech consists of multiple spectral modes across delays

Speech tracking has been extensively studied with respect to the temporal cortex (Ding and Simon, 2014; Hamilton et al., 2018; Oganian and Chang, 2019; Yi et al., 2019). However, other brain areas in the frontal, parietal, and motor cortex have been implied as well (Assaneo and Poeppel, 2018; Park et al., 2015). Different areas are likely engaged in different frequencies and at different delays relative to the incoming speech stream, reflecting bottom-up and top-down processes at different timescales.

We aimed to utilize the superior performance of our multivariate approach (compared to univariate analysis) to comprehensively characterize the spatial and spectral structure of speech tracking across delays. We applied our multivariate MI approach (using three PCA components for each parcel and Env and EnvD for speech, see Figure 1) for 61 delays between brain and speech signals (brain signal following speech signals with a delay of -300 to 300ms in steps of 10ms). As MI values are inherently positive, we applied a non-negative matrix factorization (NMF) to mutual-information spectra across participants. This way, we sought to reduce the dimensionality of our data (Parcels x Frequencies x Delays) into a fixed number of spatial components, with each one exhibiting distinct neural tracking for each time-lag. This approach is motivated by NMF's inherent clustering property (Ding et al., 2005; Lee and Seung, 1999) without a priori assumptions of the spectral or topological properties of the components. Using rank optimization, we found that our speech-brain mutual-information spectra can be divided into an optimal number of 15 anatomical components (see Methods). We used one-sample non-parametric statistical testing to emphasize consistent effects at the group level. The results of this analysis are summarized in Figure 6 (p < .05, FDR corrected). We note that in Supplementary figure 2 we show an identical figure but with constant y-axis from the t-value spectra across all components.

Each subpanel corresponds to one component. The localisation of each component is displayed in the cortical rendering together with a frequency-delay map of group-level t-values. To illustrate how speech-tracking spectra change with delay, we plot t-values spectra for selected delays (10, 100, 200, and 300 ms; see Supplementary figure 1 for GCMI spectra). A number of observations can be made.

First, NMF components show a striking variability of spectral patterns and delay dependencies. This variability suggests the existence of multiple, partly independent processes that go well beyond two speech

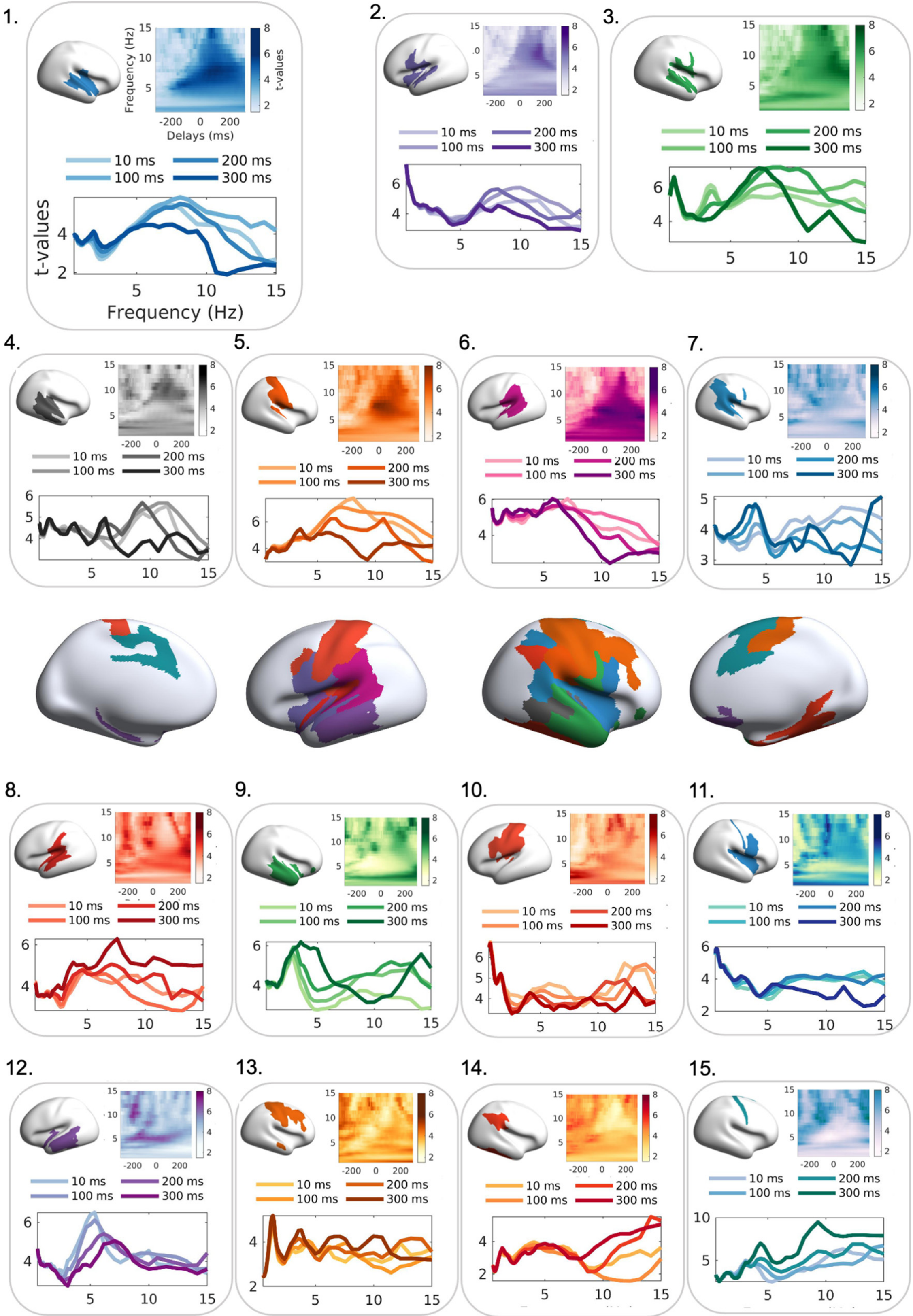

Figure 6. Spectral modes of speech coupling across delays: MI spectra clustered into 15 components as estimated from non-negative matrix factorization (NMF). Each subpanel (1-15) represents a component of mutual information spectra across delays. For each component, we illustrate the cortical rendering of the component (top left), the frequency-delays plot (top right)**,** and t-values spectra for 4 delays (10, 100, 200, 300 ms; bottom)**.** In the center of the figure, we plot the cortical rendering of all estimated components with the same color code as in subpanels (1-15).

tracking channels at δ and θ frequencies predominantly described in the literature (Ding and Simon, 2014).

Second, t-values at low frequencies show a consistent profile independent of delays. This is expected, because our speech-tracking captures phase synchronization at any given delay [e.g., 50ms leads to very small changes in phase difference for low frequency signals (such as 1 Hz) compared to higher frequency signals (such as 5 Hz)]. Still, low frequencies are informative and show three different patterns across components. Components localized partly in parietal and temporal areas show no consistent peak at frequencies below about 2 Hz (see Components 6, 7, 8, 9, and 12). In contrast, components with contributions from motor areas show strong speech-tracking at these frequencies peaking at the lowest computed frequency (Components 10, 11), even at negative delays (i.e Component 10; see the frequency-delay map). Still other components show a distinct peak at frequencies below 3 Hz and are mostly localized in the temporal cortex and around the lateral fissure (Components 2, 3, 4, 13, and 14).

Third, two types of delay dependencies can be seen mostly at frequencies of about 5 Hz and above and are clearly evident in Components 2 and 8. While Component 8 shows highest t-values for longer delays (300 ms) at frequencies between about 5-10 Hz, Component 2 shows highest t-values for short delays. More generally, components reflecting the strongest speech tracking at short delays are localized near early auditory areas (Components 1, 2, 3, and 12) also extending to motor areas (Component 5). Components with maxima at longer delays include higher-order areas in the left frontal and bilateral parietal cortex (Components 7, 15). However, NMF also identifies components with the preference for long delays in the temporal cortex (Components 3, 4, 8, and 9).

All in all, we report that speech tracking is associated with partly-distinct spectral components operating in lower- and higher-association areas with distinct faster- and slower–temporal modulations.

4. Discussion

“Speech tracking” operates at multiple timescales and arguably reflects both feed-forward and top-down processing, while orchestrating computations towards understanding. This study aimed to provide a novel multivariate framework, in which neural variability from multiple cortical areas along with the rate of speech signal amplitude change are also utilized for the assessment of speech-brain coupling. After establishing the gain of information coupling with the multivariate approach in speech-relevant timescales in auditory, frontal, and motor areas, we proceeded with a data-driven spectro-temporal characterization of cortical components coupled to the acoustic envelope. In summary, consistent with previous reports (Ding and Simon, 2014; Gross et al., 2013; Poeppel and Assaneo, 2020), we find auditory components in two distinct timescales in low frequencies [δ (0.6-3 Hz) and θ (4-7 Hz)], and we extend those findings providing a characterisation of speech-brain coupling across the cortex exhibiting higher-association bilateral motor, frontal, and temporal components operating in distinct temporal and frequency channels.

Our novel multivariate analysis was applied to investigate speech-brain coupling, but we propose that it is of relevance to any study using atlas-based source localisation. We find that representing activity of a brain area with a single time series derived from dimension-reduction techniques (such as PCA) is generally not sufficient as the first principal component is computed to provide the linear combination of the original time series (all time series along all three orientations of all voxels/vertices in a parcel) that explains most of the total variance. Since signal amplitude in MEG/EEG signals is strongest in low frequencies compared to higher frequencies, the first SVD component mostly accounts for low frequency activity at the expense of higher frequencies. Therefore, using multivariate analysis can preserve higher frequency components that are lost in the univariate case. Indeed, Cluster 2 in Figure 4b shows a distinct boost of MI in frequencies between 5-10 Hz. However, even frequencies below 5 Hz benefit significantly from the multivariate approach indicating the existence of several independent components in the data. Here we chose a three-dimensional representation per brain area but future studies will need to assess the optimal dimensionality for a multivariate approach and its dependence on the size and location of the respective brain area. Similarly, it will be interesting to compare multivariate mutual information to other multivariate methods such as multivariate pattern dependence, distance correlation, representational connectivity analysis, or canonical component analysis (Basti et al., 2020; de Cheveigné et al., 2018; Lankinen et al., 2014). We report significantly higher stimulus-brain coupling (as measured with GCMI) compared to univariate analysis across 122 out of 362 brain areas and as we expected there was no significantly reduced coupling in any brain area (Figure 4). Therefore, we can confidently posit that there is no drawback in using multivariate analysis for studying speech tracking — a finding that awaits replication in other MEG and EEG data. Instead, significantly increased performance can be seen for all frequencies up to 15 Hz and in temporal, parietal, and frontal brain areas. Regarding superior temporal areas, we note that, as shown in Figure 3b, a univariate analysis might still be sufficient, as we did not find significant improvement with the multivariate approach. Higher-order areas are those which might benefit more, as seen in lower captured variance in their first principal component (Figure 2, Panel A), possibly related to higher gradients of myelin density reflected in hierarchical gradients of timescales (Chien and Honey, 2020; Gao et al., 2020; Glasser and Van Essen, 2011) and thus indicative of more multifold relations to speech processing.

We have similarly extended the multivariate approach for the speech signal by including the speech envelope and its derivative. Recent studies have demonstrated that the speech envelope and derivative account for partly different components of speech signal and neural activity during listening (Brodbeck et al., 2018; Daube et al., 2019). Whereas the envelope quantifies the instantaneous overall amplitude of speech, the derivative quantifies the rate of change of speech amplitude. The derivative is potentially informative since acoustic onset edges cue syllabic nucleus onsets and their slopes are related to syllabic stress (Oganian and Chang, 2019) and are represented in the auditory cortex (Brodbeck et al., 2020). Nevertheless, sharp amplitude modulations of the speech signal provide acoustic indices for segmentation and encoding of continuous speech in faster timescales (Doelling et al., 2014; Oganian and Chang, 2019). We speculate that acoustic edges coded in the envelope's derivative could serve as an update of sensory gain for the attended speech (Obleser and Kayser, 2019). Supporting this notion, we found that the derivative was more informative compared to the envelope at around 0.6 Hz. Low-frequency oscillations have been traditionally associated with sensory selection (Schroeder and Lakatos, 2009), attentional orientation (Lakatos et al., 2013, 2008), and temporal anticipations (Herbst and Obleser, 2019; Stefanics et al., 2010), which —in line with the active sensing framework (Bajcsy, 1988; Bajcsy et al., 2018; Prescott et al., 2011; Schroeder et al., 2010)— can facilitate language processing (Giraud, 2020; Kandylaki and Kotz, 2020; Meyer et al., 2020). At this frequency (around 0.5 Hz), neural tracking was also found to be disrupted after altered temporal distribution of speech pauses (Kayser et al., 2015). Future studies would need to assess the extent to which endogenous δ-band activity is associated with acoustic fluctuations of speech for sensory (Boucher et al., 2019) and linguistic (Bourguignon et al., 2013; Keitel et al., 2018) related chunking of continuous speech.

How complex sounds such as speech are integrated and processed in the human brain remains a central question to neuroscientific research (Brodbeck et al., 2018; Ding et al., 2016; Jin et al., 2020). For that, both the analysis of acoustic [i.e spectrotemporal modulations; (Daube et al., 2019; Hullett et al., 2016)] and category-specific [i.e linguistic elements; (Davis and Johnsrude, 2003; Di Liberto et al., 2019, 2015; Gwilliams et al., 2020)] computations are potentially informative. How these —presumably parallel (Hamilton et al., 2021) — computational streams interact remains unclear. Our multivariate characterization of speech-brain coupling across temporal delays, cortical areas, and frequencies adds to the understanding of the stimulus-driven hierarchical organization of speech processing. We show varying coupling delays ranging from short negative to longer positive going beyond a fixed delay of around 100 ms previously reported in similar studies (Brodbeck et al., 2018; Broderick et al., 2019; Ding and Simon, 2014; Gross et al., 2013). We speculate that short negative delays found in motor areas (see Components 10, 11) might be indicative of predictive processes that serve as a top-down facilitation of early auditory processing during speech comprehension (Park et al., 2015). We have to note that while the GCMI used here for assessing speech tracking preserves sensitivity despite changes in phase distributions (Gross et al., 2021; Ince et al., 2017), a natural concern arises whether the degree of temporal smoothing (and hence precision of delay estimates) will vary as a function of frequency and consequently confound our findings. For instance, it is apparent in Figure 5 that at lower frequencies, analyses appear to be relatively insensitive to changes in delays, but these become more sensitive at higher frequencies. Considering this, delayed GCMI was previously successful in recovering ground truth in a broadband filtered time series in various frequencies (Daube et al., 2022). As previously reported, we observe two main temporal timescales (δ and θ) in auditory areas (Donhauser and Baillet, 2020; Teng and Poeppel, 2020) extending to frontal and motor areas, showing distinct computations across cortical areas, previously reported for primary and nonprimary auditory areas (Norman-Haignere and McDermott, 2018). In a similar vein, substantially longer integration windows (longer than 200ms) were found in non-primary compared to primary auditory areas. Interestingly, we observed low-α tracking of the acoustic envelope in motor areas (see Components 5, 13, and 15), possibly reflecting the excitability-related activity during segments of high sensory gain (Kayser et al., 2015). Here, speech-brain coupling was assessed during continuous speech, without dissociating temporal segments of speech. For example, spatial dissociation in the STG has been reported between speech onsets (i.e the starting of a sentence) and sustained speech (Hamilton et al., 2018). In the future, it would be interesting to investigate whether such a distinction is reflected in excitability-related responses within the δ, θ, and α frequency bands during attentive speech.

Declaration of Competing Interest

The authors declare no conflict of interest.

Credit authorship contribution statement

Nikos Chalas: Conceptualization, Investigation, Methodology, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Christoph Daube:** Data curation, Writing – review & editing. **Daniel S. Kluger:** Writing – original draft, Writing – review & editing, Visualization. **Omid Abbasi:** Methodology, Writing – original draft, Writing – review & editing. **Robert Nitsch:** Writing – review & editing. **Joachim Gross:** Conceptualization, Methodology, Formal analysis, Resources, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Acknowledgments

We acknowledge support by the Interdisciplinary Center for Clinical Research (IZKF) of the medical faculty of Münster (grant number Gro3/001/19). JG was further supported by the DFG (GR 2024/5-1).

Data availability

All Matlab code and data supporting the findings will be publicly accessible in full through GitHub (https://github.com/Nichalas) upon acceptance.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2022.119395.

Bibliography

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., Merzenich, M.M., 2001. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc Natl Acad Sci USA 98, 13367–13372. doi:10.1073/pnas.201400998.

Andrzejak, R.G., Kraskov, A., Stögbauer, H., Mormann, F., Kreuz, T., 2003. Bivariate surrogate techniques: necessity, strengths, and caveats. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 68, 066202. doi:10.1103/PhysRevE.68.066202.

Arieli, A., Sterkin, A., Grinvald, A., Aertsen, A., 1996. Dynamics of ongoing activity: explanation of the large variability in evoked cortical responses. Science 273, 1868–1871. doi:10.1126/science.273.5283.1868.

Arnal, L.H., Giraud, A.-L., 2012. Cortical oscillations and sensory predictions. Trends Cogn Sci (Regul Ed) 16, 390–398. doi:10.1016/j.tics.2012.05.003.

Asamoah, B., Khatoun, A., Mc Laughlin, M., 2019. Analytical bias accounts for some of the reported effects of tACS on auditory perception. Brain Stimulat 12, 1001–1009. doi:10.1016/j.brs.2019.03.011.

Assaneo, M.F., Poeppel, D., 2018. The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. Sci. Adv. 4, eaao3842. doi:10.1126/sciadv.aao3842.

Assaneo, M.F., 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. Front. Integr. Neurosci. 13, 28. doi:10.3389/fnint.2019.00028.

Baillet, S., 2017. Magnetoencephalography for brain electrophysiology and imaging. Nat. Neurosci. 20, 327–339. doi:10.1038/nn.4504.

Bajcsy, R., Aloimonos, Y., Tsotsos, J.K., 2018. Revisiting active perception. Auton. Robots 42, 177–196. doi:10.1007/s10514-017-9615-3.

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

Barczak, A., O'Connell, M.N., McGinnis, T., Ross, D., Mowery, T., Falchier, A., Lakatos, P., 2018. Top-down, contextual entrainment of neuronal oscillations in the auditory thalamocortical circuit. Proc Natl Acad Sci USA 115, E7605–E7614. doi:10.1073/pnas.1714684115.

Basti, A., Nili, H., Hauk, O., Marzetti, L., Henson, R.N., 2020. Multi-dimensional connectivity: a conceptual and mathematical review. Neuroimage 221, 117179. doi:10.1016/j.neuroimage.2020.117179.

Berry, M.W., Browne, M., Langville, A.N., Pauca, V.P., Plemmons, R.J., 2007. Algorithms and applications for approximate nonnegative matrix factorization. Comput. Stat. Data Anal. 52, 155–173. doi:10.1016/j.csda.2006.11.006.

Besl, P.J., McKay, H.D., 1992. A method for registration of 3-D shapes. IEEE Trans. Pattern Anal. Mach. Intell. 14, 239–256. doi:10.1109/34.121791.

Bishop, Geo.H., 1932. Cyclic changes in excitability of the optic pathway of the rabbit. American Journal of Physiology-Legacy Content 103, 213–224. doi:10.1152/ajplegacy.1932.103.1.213.

Boemio, A., Fromm, S., Braun, A., Poeppel, D., 2005. Hierarchical and asymmetric temporal sensitivity in human auditory cortices. Nat. Neurosci. 8, 389–395. doi:10.1038/nn1409.

Boucher, V.J., Gilbert, A.C., Jemel, B., 2019. The Role of Low-frequency Neural Oscillations in Speech Processing: Revisiting Delta Entrainment. J. Cogn. Neurosci. 31, 1205–1215. doi:10.1162/jocn_a_01410.

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. Hum. Brain Mapp 34, 314–326. doi:10.1002/hbm.21442.

Brainard, D.H., 1997. The Psychophysics Toolbox. Spat. Vis. 10, 433–436. doi:10.1163/156856897×00357.

Brasselet, R., Panzeri, S., Logothetis, N.K., Kayser, C., 2012. Neurons with stereotyped and rapid responses provide a reference frame for relative temporal coding in primate auditory cortex. J. Neurosci. 32, 2998–3008. doi:10.1523/JNEUROSCI.5435-11.2012.

Brodbeck, C., Hong, L.E., Simon, J.Z., 2018. Rapid Transformation from Auditory to Linguistic Representations of Continuous Speech. Curr. Biol. 28, 3976–3983. doi:10.1016/j.cub.2018.10.042, e5. doi:.

Brodbeck, C., Jiao, A., Hong, L.E., Simon, J.Z., 2020. Neural speech restoration at the cocktail party: Auditory cortex recovers masked speech of both attended and ignored speakers. PLoS Biol 18, e3000883. doi:10.1371/journal.pbio.3000883.

Broderick, M.P., Anderson, A.J., Lalor, E.C., 2019. Semantic context enhances the early auditory encoding of natural speech. J. Neurosci. 39, 7564–7575. doi:10.1523/JNEUROSCI.0584-19.2019.

Buzsáki, G., Draguhn, A., 2004. Neuronal oscillations in cortical networks. Science 304, 1926–1929. doi:10.1126/science.1099745.

Chien, H.-Y.S., Honey, C.J., 2020. Constructing and forgetting temporal context in the human cerebral cortex. Neuron 106, 675–686. doi:10.1016/j.neuron.2020.02.013, e11. doi.

Crosse, M.J., Di Liberto, G.M., Bednar, A., Lalor, E.C., 2016. The multivariate temporal response function (mtrf) toolbox: A MATLAB toolbox for relating neural signals to continuous stimuli. Front. Hum. Neurosci. 10, 604. doi:10.3389/fnhum.2016.00604.

Daube, C., Gross, J. and Ince, R.A., 2022. A whitening approach for Transfer Entropy permits the application to narrow-band signals. arXiv preprint arXiv:2201.02461.

Daube, C., Ince, R.A.A., Gross, J., 2019. Simple Acoustic Features Can Explain Phoneme-Based Predictions of Cortical Responses to Speech. Curr. Biol. 29, 1924–1937. doi:10.1016/j.cub.2019.04.067, .e9.

Davis, M.H., Johnsrude, I.S., 2003. Hierarchical processing in spoken language comprehension. J. Neurosci. 23, 3423–3431. doi:10.1523/JNEUROSCI.23-08-03423.2003.

de Cheveigné, A., Wong, D.D.E., Di Liberto, G.M., Hjortkjær, J., Slaney, M., Lalor, E., 2018. Decoding the auditory brain with canonical component analysis. Neuroimage 172, 206–216. doi:10.1016/j.neuroimage.2018.01.033.

Di Liberto, G.M., O'Sullivan, J.A., Lalor, E.C., 2015. Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. Curr. Biol. 25, 2457–2465. doi:10.1016/j.cub.2015.08.030.

Di Liberto, G.M., Wong, D., Melnik, G.A., de Cheveigné, A., 2019. Low-frequency cortical responses to natural speech reflect probabilistic phonotactics. Neuroimage 196, 237–247. doi:10.1016/j.neuroimage.2019.04.037.

Ding, C., He, X., Simon, H.D., 2005. On the equivalence of nonnegative matrix factorization and spectral clustering. In: Kargupta, H., Srivastava, J., Kamath, C., Goodman, A. (Eds.), Proceedings of the 2005 SIAM International Conference on Data Mining. Presented at the Proceedings of the 2005 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, Philadelphia, PA, pp. 606–610. doi:10.1137/1.9781611972757.70.

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. doi:10.1038/nn.4186.

Ding, N., Simon, J.Z., 2014. Cortical entrainment to continuous speech: functional roles and interpretations. Front. Hum. Neurosci. 8, 311. doi:10.3389/fnhum.2014.00311.

Doelling, K.B., Arnal, L.H., Ghitza, O., Poeppel, D., 2014. Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. Neuroimage 85, 761–768. doi:10.1016/j.neuroimage.2013.06.035, Pt 2.

Donhauser, P.W., Baillet, S., 2020. Two distinct neural timescales for predictive speech processing. Neuron 105, 385–393. doi:10.1016/j.neuron.2019.10.019, e9.

Gao, R., van den Brink, R.L., Pfeffer, T., Voytek, B., 2020. Neuronal timescales are functionally dynamic and shaped by cortical microarchitecture. eLife 9. doi:10.7554/eLife.61277.

Giraud, A.-L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. 15, 511–517. doi:10.1038/nn.3063.

Giraud, A.-L., 2020. Oscillations for all A commentary on Meyer, Sun & Martin (2020). Lang. Cogn. Neurosci. 1–8. doi:10.1080/23273798.2020.1764990.

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, 171–178. doi:10.1038/nature18933.

Glasser, M.F., Van Essen, D.C., 2011. Mapping human cortical areas in vivo based on myelin content as revealed by T1- and T2-weighted MRI. J. Neurosci. 31, 11597–11616. doi:10.1523/JNEUROSCI.2180-11.2011.

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. doi:10.1371/journal.pbio.1001752.

Gross, J., Kluger, D.S., 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. doi:10.1016/j.neuroimage.2021.118660.

Gross, J., 2019. Magnetoencephalography in cognitive neuroscience: A primer. Neuron 104, 189–204. doi:10.1016/j.neuron.2019.07.001.

Gwilliams, L., King, J.-R., Marantz, A., Poeppel, D., 2020. Neural dynamics of phoneme sequencing in real speech jointly encode order and invariant content. BioRxiv doi:10.1101/2020.04.04.025684.

Hamilton, L.S., Edwards, E., Chang, E.F., 2018. A spatial map of onset and sustained responses to speech in the human superior temporal gyrus. Curr. Biol. 28, 1860–1871. doi:10.1016/j.cub.2018.04.033, e4. doi.

Hamilton, L.S., Oganian, Y., Hall, J., Chang, E.F., 2021. Parallel and distributed encoding of speech across human auditory cortex. Cell 184, 4626–4639. doi:10.1016/j.cell.2021.07.019, e13.

Henry, M.J., Obleser, J., 2012. Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. Proc Natl Acad Sci USA 109, 20095–20100. doi:10.1073/pnas.1213390109.

Herbst, S.K., Obleser, J., 2019. Implicit temporal predictability enhances pitch discrimination sensitivity and biases the phase of delta oscillations in auditory cortex. Neuroimage 203, 116198. doi:10.1016/j.neuroimage.2019.116198.

Hertrich, I., Dietrich, S., Trouvain, J., Moos, A., Ackermann, H., 2012. Magnetic brain activity phase-locked to the envelope, the syllable onsets, and the fundamental frequency of a perceived speech signal. Psychophysiology 49, 322–334. doi:10.1111/j.1469-8986.2011.01314.x.

Hullett, P.W., Hamilton, L.S., Mesgarani, N., Schreiner, C.E., Chang, E.F., 2016. Human Superior Temporal Gyrus Organization of Spectrotemporal Modulation Tuning Derived from Speech Stimuli. J. Neurosci. 36, 2014–2026. doi:10.1523/JNEUROSCI.1779-15.2016.

Ince, R.A.A., Giordano, B.L., Kayser, C., Rousselet, G.A., Gross, J., Schyns, P.G., 2017. A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula. Hum. Brain Mapp. 38, 1541–1573. doi:10.1002/hbm.23471.

Ince, R.A.A., van Rijsbergen, N.J., Thut, G., Rousselet, G.A., Gross, J., Panzeri, S., Schyns, P.G., 2015. Tracing the flow of perceptual features in an algorithmic brain network. Sci. Rep. 5, 17681. doi:10.1038/srep17681.

Jaworska, K., Yan, Y., van Rijsbergen, N.J., Ince, R.A.A., Schyns, P.G., 2022. Different computations over the same inputs produce selective behavior in algorithmic brain networks. eLife 11. doi:10.7554/eLife.73651.

Jin, P., Lu, Y., Ding, N., 2020. Low-frequency neural activity reflects rule-based chunking during speech listening. eLife 9. doi:10.7554/eLife.55613.

Johnston, A., Nishida, S., 2001. Time perception: brain time or event time? Curr. Biol. 11, R427–R430. doi:10.1016/s0960-9822(01)00252-4.

Kandylaki, K.D., Kotz, S.A., 2020. Distinct cortical rhythms in speech and language processing and some more: a commentary on Meyer, Sun, & Martin (2019). Lang. Cogn. Neurosci. 1–5. doi:10.1080/23273798.2020.1757729.

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. doi:10.1523/JNEUROSCI.0268-15.2015.

Kayser, S.J., Ince, R.A.A., Gross, J., Kayser, C., 2015. Irregular speech rate dissociates auditory cortical entrainment, evoked responses, and frontal alpha. J. Neurosci. 35, 14691–14701. doi:10.1523/JNEUROSCI.2243-15.2015.

Keitel, A., Gross, J., Kayser, C., 2018. Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. PLoS Biol 16, e2004473. doi:10.1371/journal.pbio.2004473.

Kluger, D.S., Gross, J., 2020. Respiration modulates oscillatory neural network activity at rest. BioRxiv doi:10.1101/2020.04.23.057216.

Koskinen, M., Kurimo, M., Gross, J., Hyvärinen, A., Hari, R., 2020. Brain activity reflects the predictability of word sequences in listened continuous speech. Neuroimage 219, 116936. doi:10.1016/j.neuroimage.2020.116936.

Lakatos, P., Gross, J., Thut, G., 2019. A new unifying account of the roles of neuronal entrainment. Curr. Biol. 29, R890–R905. doi:10.1016/j.cub.2019.07.075.

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, 110–113. doi:10.1126/science.1154735.

Lakatos, P., Musacchia, G., O'Connel, M.N., Falchier, A.Y., Javitt, D.C., Schroeder, C.E., 2013. The spectrotemporal filter mechanism of auditory selective attention. Neuron 77, 750–761. doi:10.1016/j.neuron.2012.11.034.

Lakatos, P., Shah, A.S., Knuth, K.H., Ulbert, I., Karmos, G., Schroeder, C.E., 2005. An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. J. Neurophysiol. 94, 1904–1911. doi:10.1152/jn.00263.2005.

Lankinen, K., Saari, J., Hari, R., Koskinen, M., 2014. Intersubject consistency of cortical MEG signals during movie viewing. Neuroimage 92, 217–224. doi:10.1016/j.neuroimage.2014.02.004.

Lee, D.D., Seung, H.S., 1999. Learning the parts of objects by non-negative matrix factorization. Nature 401, 788–791. doi:10.1038/44565.

Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., Moore, B.C.J., 2006. Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. Proc Natl Acad Sci USA 103, 18866–18869. doi:10.1073/pnas.0607364103.

Luo, H., Poeppel, D., 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54, 1001–1010. doi:10.1016/j.neuron.2007.06.004.

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164, 177–190. doi:10.1016/j.jneumeth.2007.03.024.

Meyer, L., Sun, Y., Martin, A.E., 2020. Entraining" to speech, generating language? Lang. Cogn. Neurosci. 1–11. doi:10.1080/23273798.2020.1827155.

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. doi:10.1088/0031-9155/48/22/002.

Norman-Haignere, S.V., McDermott, J.H., 2018. Neural responses to natural and model-matched stimuli reveal distinct computations in primary and nonprimary auditory cortex. PLoS Biol 16, e2005127. doi:10.1371/journal.pbio.2005127.

Obleser, J., Herrmann, B., Henry, M.J., 2012. Neural oscillations in speech: don't be enslaved by the envelope. Front. Hum. Neurosci. 6, 250. doi:10.3389/fnhum.2012.00250.

Obleser, J., Kayser, C., 2019. Neural entrainment and attentional selection in the listening brain. Trends Cogn Sci (Regul Ed) 23, 913–926. doi:10.1016/j.tics.2019.08.004.

Oganian, Y., Chang, E.F., 2019. A speech envelope landmark for syllable encoding in human superior temporal gyrus. Sci. Adv. 5, eaay6279. doi:10.1126/sciadv.aay6279.

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. doi:10.1155/2011/156869.

Panzeri, S., Brunel, N., Logothetis, N.K., Kayser, C., 2010. Sensory neural codes using multiplexed temporal scales. Trends Neurosci 33, 111–120. doi:10.1016/j.tins.2009.12.001.

Park, H., Ince, R.A.A., Schyns, P.G., 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. doi:10.1016/j.cub.2015.04.049.

Peelle, J.E., Davis, M.H., 2012. Neural Oscillations Carry Speech Rhythm through to Comprehension. Front. Psychol. 3, 320. doi:10.3389/fpsyg.2012.00320.

Poeppel, D., Assaneo, M.F., 2020. Speech rhythms and their neural foundations. Nat. Rev. Neurosci. 21, 322–334. doi:10.1038/s41583-020-0304-4.

Prescott, T.J., Diamond, M.E., Wing, A.M., 2011. Active touch sensing. Philos. Trans. R. Soc. Lond. B Biol. Sci. 366, 2989–2995. doi:10.1098/rstb.2011.0167.

Qiao, H., 2015. New SVD based initialization strategy for non-negative matrix factorization. Pattern Recognit. Lett. 63, 71–77. doi:10.1016/j.patrec.2015.05.019.

Riecke, L., Formisano, E., Sorger, B., Başkent, D., Gaudrain, E., 2018. Neural entrainment to speech modulates speech intelligibility. Curr. Biol. 28, 161–169. doi:10.1016/j.cub.2017.11.033, e5. doi:.

Rimmele, J.M., Poeppel, D., Ghitza, O., 2021. Acoustically driven cortical delta oscillations underpin prosodic chunking. eNeuro doi:10.1523/ENEURO.0562-20.2021.

Romei, V., Gross, J., Thut, G., 2010. On the role of prestimulus alpha rhythms over occipito-parietal areas in visual input regulation: correlation or causation? J. Neurosci. 30, 8692–8697. doi:10.1523/JNEUROSCI.0160-10.2010.

Schädler, M.R., Meyer, B.T., Kollmeier, B., 2012. Spectro-temporal modulation subspace-spanning filter bank features for robust automatic speech recognition. J. Acoust. Soc. Am. 131, 4134–4151. doi:10.1121/1.3699200.

Schoffelen, J.M., Hultén, A., Lam, N., Marquand, A.F., Uddén, J., Hagoort, P., 2017. Frequency-specific directed interactions in the human brain network for language. Proc Natl Acad Sci USA 114, 8083–8088. doi:10.1073/pnas.1703155114.

Schroeder, C.E., Lakatos, P., 2009. Low-frequency neuronal oscillations as instruments of sensory selection. Trends Neurosci 32, 9–18. doi:10.1016/j.tins.2008.09.012.

Schroeder, C.E., Wilson, D.A., Radman, T., Scharfman, H., Lakatos, P., 2010. Dynamics of Active Sensing and perceptual selection. Curr. Opin. Neurobiol. 20, 172–176. doi:10.1016/j.conb.2010.02.010.

Scott, S., McGettigan, C., 2012. Amplitude onsets and spectral energy in perceptual experience. Front. Psychol. 3, 80. doi:10.3389/fpsyg.2012.00080.

Shannon, C.E., 1948. A mathematical theory of communication. Bell System Technical Journal 27, 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.

Stefanics, G., Hangya, B., Hernádi, I., Winkler, I., Lakatos, P., Ulbert, I., 2010. Phase entrainment of human delta oscillations can mediate the effects of expectation on reaction speed. J. Neurosci. 30, 13578–13585. doi:10.1523/JNEUROSCI.0703-10.2010.

Teng, X., Poeppel, D., 2020. Theta and Gamma Bands Encode Acoustic Dynamics over Wide-Ranging Timescales. Cereb. Cortex 30, 2600–2614. doi:10.1093/cercor/bhz263.

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 Trans Biomed Eng 44, 867–880. doi:10.1109/10.623056.

Wilsch, A., Neuling, T., Obleser, J., Herrmann, C.S., 2018. Transcranial alternating current stimulation with speech envelopes modulates speech comprehension. Neuroimage 172, 766–774. doi:10.1016/j.neuroimage.2018.01.038.

Yi, H.G., Leonard, M.K., Chang, E.F., 2019. The encoding of speech sounds in the superior temporal gyrus. Neuron 102, 1096–1110. doi:10.1016/j.neuron.2019.04.023.

Zeki, S., Bartels, A., 1998. The asynchrony of consciousness. Proc. Biol. Sci. 265, 1583–1585. doi:10.1098/rspb.1998.0475.

Zion Golumbic, E.M., Ding, N., Bickel, S., Lakatos, P., Schevon, C.A., McKhann, G.M., Goodman, R.R., Emerson, R., Mehta, A.D., Simon, J.Z., Poeppel, D., Schroeder, C.E., 2013. Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party. Neuron 77, 980–991. doi:10.1016/j.neuron.2012.12.037.

Zoefel, B., Archer-Boyd, A., Davis, M.H., 2018. Phase entrainment of brain oscillations causally modulates neural responses to intelligible speech. Curr. Biol. 28, 401–408. doi:10.1016/j.cub.2017.11.071, e5.