bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Delta/theta band EEG differentially tracks low and high frequency speech envelopes

Felix Bröhl[1*] and Christoph Kayser[1]

[1]*Department for Cognitive Neuroscience, Faculty of Biology, Bielefeld University, Universitätsstr. 25, 33615, Bielefeld, Germany*

[*] **Correspondence to:** Felix Bröhl (felix.broehl@uni-bielefeld.de)

Highlights

- Delta/theta band EEG tracks band-limited speech envelopes similar to real speech
- Low and high frequency derived speech envelopes are represented differentially
- High-frequency derived envelopes are more susceptible to attentional and contextual manipulations
- Delta band tracking shifts towards low frequency derived envelopes with more acoustic detail

Abstract

The representation of speech in the brain is often examined by measuring the alignment of rhythmic brain activity to the speech envelope. To conveniently quantify this alignment (termed 'speech tracking') many studies consider the overall speech envelope, which combines acoustic fluctuations across the spectral range. Using EEG recordings, we show that using this overall envelope provides a distorted picture on speech encoding. We systematically investigated the encoding of spectrally-limited speech envelopes presented by individual and multiple noise carriers in the human brain. Tracking in the 1 to 6 Hz EEG bands differentially reflected low (0.2 - 0.83 kHz) and high (2.66 - 8 kHz) frequency envelopes. This was independent of the specific carrier frequency but sensitive to attentional manipulations, and reflects the context-dependent emphasis of information from distinct spectral ranges of the speech envelope in low frequency brain activity. As low and high frequency speech envelopes relate to distinct phonemic features, our results suggest that functionally distinct processes contribute to speech tracking in the same EEG bands, and are easily confounded when considering the overall speech envelope.

Keywords

speech entrainment, brain rhythms, hearing, EEG

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1. Introduction

A central notion governing the neurobiology of speech has been that rhythmic brain activity aligns to the regularities of speech, particularly in the delta (1 - 4 Hz) and theta (4 - 8 Hz) bands. This alignment has been termed "speech entrainment" or "speech tracking" (Obleser and Kayser, 2019) and both, serves as a powerful tool to study speech encoding, but also promotes a fundamental role of rhythmic brain activity for speech comprehension and production (Giraud and Poeppel, 2012; Obleser and Kayser, 2019; Poeppel and Assaneo, 2020). Experimental signatures of speech tracking in M/EEG or ECoG recordings are often obtained by relating the neural signal to a simplified signature of the acoustic energy: the latter is often defined as the wide-band speech envelope and is computed by averaging individual band-limited envelopes along the spectral dimension (Ahissar et al., 2001; Aiken and Picton, 2008; Doelling et al., 2014; Kayser et al., 2015; Keitel et al., 2018; Meyer, 2018; Nourski et al., 2009). However, relevant information is not distributed homogeneously along this spectral dimension: lower frequencies often carry more information about the majority of phonemes (Fig. 1) and are more relevant for comprehension (Elliott and Theunissen, 2009; Erb et al., 2020; Monson et al., 2014). Also, the auditory system can selectively track spectrally-defined acoustic streams (Broderick et al., 2019; Mesgarani and Chang, 2012; O'Sullivan et al., 2019) and many auditory cortices segregate sounds along a tonotopic organization. If the experimentally observed speech tracking signatures arise from such processes, combining individual speech envelopes along the spectral dimension would effectively blend distinct neural representations into a single experimental measure. Hence, we conjecture that the use of the wide-band envelope to study speech encoding may obscure possibly distinct underlying neural processes and provide a distorted picture.

To investigate this hypothesis, we studied the representation of synthetic and speech-like stimuli derived from natural German speech in the human EEG. The stimuli were created by applying amplitude-envelopes derived from limited ranges of the spectral dimension of natural speech to band-limited noise carriers. They hence captured essential aspects of the envelope dynamics of natural speech. Quantifying the alignment of rhythmic components of the EEG to the acoustic envelope (Giordano et al., 2017; Kayser et al., 2015; Park et al., 2016) we then compared the cerebral tracking of such individual band-limited speech envelopes. Across experimental conditions we manipulated the spectral match between envelope and carrier, participants' attention to specific envelope-carrier pairs, and varied the number of simultaneously presented envelope-carrier bands to probe a number of specific questions as outlined next.

First, we asked whether envelopes extracted from lower and higher spectral bands (c.f. Fig. 1) are tracked similarly in the EEG, i.e. over the same electrodes and in the same EEG bands. Second, we asked whether this tracking is sensitive to the match between carrier frequency and the spectral range covered by the envelope, motivated by relevance of the natural match between carrier frequency and

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

envelope band for comprehension as shown by spectrally rotated speech (Molinaro and Lizarazu, 2018; Peelle et al., 2013). Third, we probed whether and for which envelope bands tracking is affected by attention, similar as the known attentional effects for the tracking of natural speech (Cusack et al., 2004; Kerlin et al., 2010; Lakatos et al., 2008; Rimmele et al., 2015; Teoh and Lalor, 2020). Fourth, we investigated the tracking of spectrally-limited envelopes for individual stimuli and the superposition of multiple envelope-noise pairs (i.e. noise vocoded speech) to probe which tracking signatures covary with the signal's complexity and participant's comprehension. And lastly, we investigated the statistical regularities of acoustic landmarks that supposedly give rise to speech tracking in order to illustrate the necessity of studying the tracking of spectrally limited envelopes.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2. Materials and Methods

2.1 Participants and data acquisition

We collected data from 24 native German speaking participants (19 females, mean age = 23.4 ± 3.6 SD). All participants provided written informed consent and were screened for hearing impairment with the speech, spatial and qualities of hearing scale (SSQ) prior to the study (Gatehouse and Noble, 2004). We excluded participants with scores two standard deviations below the mean of a young control group reported by Demeester et al. (Demeester et al., 2012). Participants received monetary compensation of 10 € per hour. The experiment was approved by the local ethics committee of the Bielefeld University, Germany, and was conducted in compliance with the Declaration of Helsinki. Participants were seated in a dimly lit, acoustically and electrically shielded room (Desone Ebox, Germany). Auditory stimuli were presented binaurally using Sennheiser headphones (Model HD200) from a Creative sound blaster z Soundcard at an average intensity of 68 dB SPL. EEG signals were continuously recorded using an active 128 channel BioSemi system using Ag-AgCl electrodes mounted on an elastic cap (BioSemi). Four additional electrodes were placed at the outer canthi and below the eyes to obtain the electrooculogram. Electrode impedance was kept at <40 kΩ. Data were acquired at a sampling rate of 1024 Hz using a lowpass filter of 400 Hz.

2.2 Speech material and stimuli

The original material based on which the experimental stimuli were modelled comprised 39 sentences extracted from German radio interviews conducted by the Deutschlandfunk on various topics. The sentences were spoken by nine different professional speakers (2 female, 7 male) and lasted on average 8.65 ± 2.3 s (mean ± SD, ranging from 4.13 s to 13.67 s). The sentences consisted of 22.1 ± 7.2 words. All recordings are publicly available online and were downloaded and converted to .wav files with a sampling rate of 44.1 kHz (https://www.deutschlandradio.de/audio-archiv.260.de.html).

To create the experimental stimuli, we first extracted amplitude-envelopes from the spoken sentences and then applied these to bandpass filtered Gaussian white noise with varying center frequencies as described below. The band-limited envelopes were obtained by bandpass filtering each sentence into logarithmically spaced frequency bands covering the range from 0.2 to 8 kHz. Depending on the specific analysis or experimental condition, we used either three frequency bands (boundaries: 0.2, 0.83, 2.66, 8 kHz), six bands (boundaries: 0.2, 0.43, 0.83, 1.5, 2.66, 4.6, 8 kHz) or twelve bands (boundaries: 0.2, 0.3, 0.43, 0.6, 0.83, 1.1, 1.5, 2, 2.66, 3.5, 4.6, 6.1, 8 kHz). In the following we refer to the bands as followed: Env_{low}: 0.2 - 0.83 kHz, Env_{mid}: 0.83 - 2.66 kHz and Env_{high}: 2.66 - 8 kHz. For filtering we applied a zero-phase 4th order Butterworth IIR filter using the Matlab filtfilt function. The band-limited envelope was then computed by taking the absolute of the Hilbert transform (Fig. 1A). Next, we bandpass filtered Gaussian white noise into the same logarithmically spaced frequency bands and applied the envelopes extracted from the speech material to create amplitude-modulated band-limited

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

noises. Similar to the envelopes, carriers are thus defined by their respective frequency bands: Car_{low}: 0.2 - 0.83 kHz, Car_{mid}: 0.83 - 2.66 kHz and Car_{high}: 2.66 - 8 kHz. To create different experimental conditions, we constructed stimuli by combining these band-limited envelopes not only with a noise carrier at the same (i.e. 'natural', Fig. 1B, signals 1, 5) frequency, but also by applying the envelope of one frequency band to a different carrier band (i.e. 'unnatural', Fig. 1B, signals 2, 4). We also constructed three noise-vocoded versions of the same sentence by averaging 'naturally' paired amplitude-modulated band-limited noises over 3, 6 or 12 bands, as described further below.

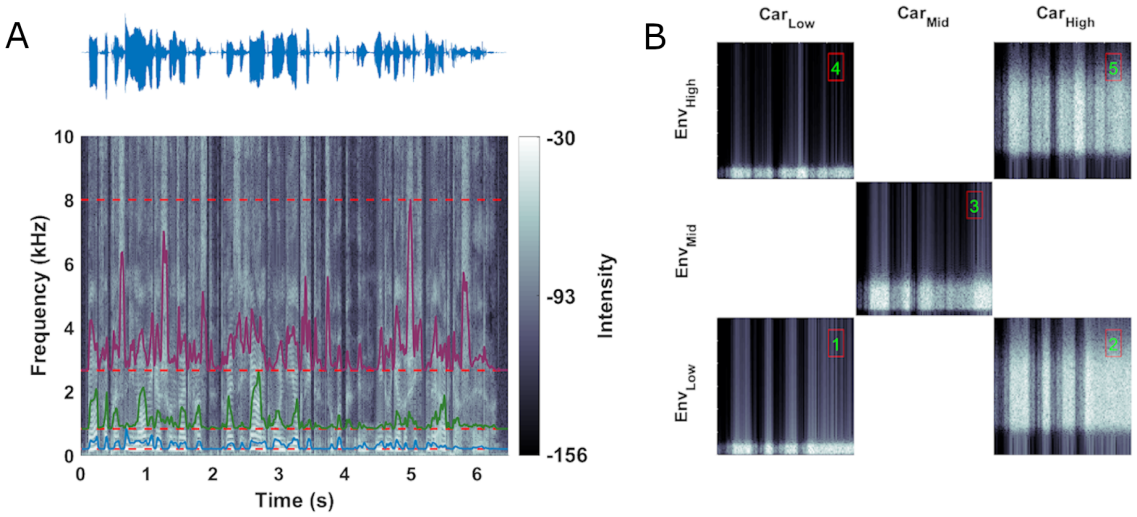

Fig. 1. Stimulus design. (**A**) Acoustic waveform (top) and spectrogram (bottom) of one example sentence used in the study. Colored outlines show amplitude envelopes extracted from (here) three logarithmically spaced frequency ranges between 0.2 and 8 kHz (red dashed lines). (**B**) Spectrograms of 1 second snippets taken from the different stimulus conditions (Block 1). Stimuli were envelope-modulated band-limited noises. The spectral ranges covered by the carrier noise and the spectral range from which the envelope was extracted were paired across conditions as shown by the matrix (low: 0.2 - 0.83 kHz, mid: 0.83 - 2.66 kHz and high: 2.66 - 8 kHz). Conditions 1, 3 and 5 reflect natural pairings where the frequency band of the carrier was modulated with the envelope derived from the same band, while conditions 2 and 4 reflect unnatural pairings.

2.3 Experimental design

The experiment consisted of five blocks, in which participants were presented with different combinations of amplitude-modulated noises. All blocks were presented within the same session in a fixed order.

The first block comprised five conditions, in each of which we presented one carrier modulated with one envelope, as depicted in figure 1B. Three conditions presented natural combinations of carrier frequency and band-limited envelope, e.g. Env_{low} with Car_{low} (Fig. 1B, signals 1, 3, 5), two conditions presented unnatural pairings, i.e. Env_{low} with Car_{high} and Env_{high} with Car_{low} (signals 2, 4).

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

In the second block we presented a mixture of two amplitude-modulated noises, whereby two carriers at different frequencies were modulated with the same band-limited envelope (one natural, one unnatural pair). This resulted in one condition for Env_{low} (i.e. averaged signals 1 and 2; Fig. 1B) and one condition for Env_{high} (i.e. averaged signals 4 and 5).

In the third and fourth blocks we presented a mixture of two amplitude-modulated noises, whereby each noise was modulated by its natural band-limited envelope (i.e. signals 1 and 5; Fig. 1B), while also manipulating participants' attentional demands. Hence, block three comprised two conditions for natural and unnatural pairings and block four comprised four conditions for natural and unnatural pairings, attending either the low or the high carrier.

To ensure attentive engagement in blocks 1 to 4, participants performed a dummy task in which they had to report the occurrence of a silent gap within the stimulus. These gaps were randomly distributed over all trials and occurred in 20 % of trials. Each gap lasted between 450 and 600 ms and randomly occurred within the third quarter of the stimulus. In block three, participants were tasked to detect a gap which was implemented to appear simultaneously in both streams, whereas in block four participants were tasked to attend either the low or the high carrier and only respond to a gap occurrence in that attended carrier. Hence, between blocks 3 and 4 we manipulated the locus of attention.

In the fifth block participants were presented with noise-vocoded speech. Each sentence was repeated in each of three conditions. First speech was vocoded by summing 3 bands, followed by summing 6 bands, and lastly by summing 12 bands. Participants were asked to pay attention in order to perform a comprehension task. Subsequent to each sentence they had to report which out of three candidate words they had heard by pressing one of three buttons. Additionally, they were asked to rate their confidence in their decision on a scale from 1 (low) over 2 (medium) to 3 (high). As each sentence was repeated in each sub-block, this introduced a potential confound of memory and recognition. Given that performance was already high for 6 envelopes, we did not investigate the subsequently presented 12 envelope condition any further, as the cerebral encoding and perceptual reports may be confounded by memory-related effects.

During the entire experiment, participants were instructed to fixate a central fixation cross and to blink as little as possible. All 39 sentences were repeated in a pseudo-random sequence in each condition.

2.4 Analysis of stimulus material

We analyzed statistical regularities of the acoustic onsets in the speech material as these provide prominent landmarks that drive the neural entrainment to speech (Drennan and Lalor, 2019; Khalighinejad et al., 2017; Oganian and Chang, 2019). For each band-limited envelope we computed its derivative and determined local peaks using Matlab's 'findpeaks' function with a minimum distance of 125 ms between neighboring peaks. For each sentence we then computed the onset frequency by dividing the number of onsets by the length of the given sentence. This was done for each envelope band separately and subsequently averaged over sentences. Additionally, we computed the correlation

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

between each band-limited envelope band and the overall envelope, which was defined as the average of the three envelopes derived from band-limited speech (Fig. 2A - B). To determine whether onsets in different bands were temporally structured we determined their joint probability: for each onset in a given envelope band we calculated the probability of onsets in all remaining bands that fell within a 400 ms window using temporal bins of 40 ms length (Fig. 2C).

To link the acoustic onsets with their phonemic content, we examined how these onsets coincide with phonemes. For this we used the web-based WebMAUS Basic tool by the Bavarian Archive for Speech Signals to force-align written transcripts to the acoustic waveform (Kisler et al., 2017). Phonemic onsets were stored as TextGrid files and subsequently adjusted by hand using Praat (Boersma and van Heuven, 2001). For each envelope band we determined the closest phoneme onset to each acoustic onset within a temporal distance of 40 ms. Phoneme counts were grouped by their manner of articulation (fricatives, plosives, vowels, pure vowels, nasals) (Fig. 2D). Phoneme counts for each envelope were then divided into two subgroups: 'unique' phonemes that coincide with acoustic onsets in only one band, and 'common' phonemes coinciding with acoustic onsets in more than one envelope band. To compute the probability of phoneme groups given any acoustic onset, we normalized the count of each phoneme group to the total number of phonemes in all sentences.

2.5 EEG preprocessing

Data analysis was performed offline using the FieldTrip toolbox (Oostenveld et al., 2011) Version 20190905 and custom written MATLAB scripts. Data and code used in this study will be made available on the Data Server of the University of Bielefeld upon acceptance of the article. The EEG data were bandpass filtered between 0.2 and 60 Hz using a zero-phase 4th order Butterworth IIR filter and resampled to 150 Hz. To detect channels with excessive noise we selected channels with an amplitude above 3.5 standard deviations from the mean of all channels and interpolated these using a weighted neighbor approach: on average we interpolated 13.55 ± 6.73 channels across subjects (mean ± SD). Noise cleaning was performed using independent component analysis based on 40 components across all blocks simultaneously. Artifacts were detected automatically using predefined noise templates capturing typical artifacts related to eye movements, blinks or localized muscle activity based on our previous work (Kayser et al., 2016; McNair et al., 2019). On average 12.16 ± 4.76 components were removed (mean ± SD). Finally, all channels were referenced to the mean of eight mastoid electrodes (B14, B25, B26, B27, D8, D22, D23, D24).

2.6 EEG analysis

To quantify the relationship between the EEG and the stimulus envelope we used a mutual information (MI) approach with a frequency-specific time lag between the amplitude envelope and the electrode-wise EEG signals (Kayser et al., 2015; Keitel et al., 2017). Time lags were determined for each

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

frequency band across all electrodes and participants by computing the MI between envelope and EEG over a 500 ms range with 40 ms steps. The lags with the highest average MI across electrodes was used for the subsequent analysis.

The single trial EEG data was cut to the same length as the corresponding acoustic envelope. Then, we appended all trial-wise envelopes and EEG data from the same condition and filtered these vectors into partly overlapping frequency bands: delta1 (0.5 - 2 Hz), delta2 (1 - 4 Hz), theta1 (2 - 6 Hz), theta2 (4 - 8 Hz) and alpha (8 - 12 Hz), respectively using a zero-phase 4th order Butterworth IIR filter. To compute the mutual information between EEG and stimulus vectors we derived the analytic signal using the Hilbert transform and calculated the MI using the Gaussian copula approach (Ince et al., 2017). By using the complex signals of the Hilbert transform we effectively quantified the relationship between two signals using both power and phase information. Analysis in block 5 of the experiment was conducted by computing the MI between the EEG response and the three envelopes (i.e. Env_{low}, Env_{mid}, Env_{high}) for both the 3 and 6 envelopes condition. This enabled us to compare tracking of equal spectral distances across conditions with varying detail.

2.7 Statistical analysis

For statistical analysis of the EEG data we relied on a two-step cluster-based permutation approach (Maris and Oostenveld, 2007). To compare the speech MI between two conditions, we first computed electrode- and frequency-wise paired two-tailed t-tests across participants. Electrodes were considered significant if their t-value exceeded a critical level of $p<0.01$. Then, to create a null distribution under the hypothesis of no difference, we randomly flipped the sign of each t-value 2000 times. To control for multiple comparisons across electrodes and EEG frequency bands the maximum statistics was used. For clustering we required a minimum of three significant neighboring electrodes and used the 'cluster-mass' as cluster-forming criterion (Maris and Oostenveld, 2007). Clusters were considered significant if they exceeded a two-sided alpha level of 0.01. For each significant cluster the electrode with the maximum (for positive or minimum for negative clusters) test statistic was selected and subsequently used to compute cohen's d as an indicator of effect size. In cases where nearby clusters reflected a similar effect but were separated by at most one electrode, we conjugated these into one cluster. Instances of such conjugations are indicated in the text.

To investigate changes in speech tracking with different numbers of envelope bands (Fig. 7), we regressed the MI values for each electrode and frequency band against the number of envelopes. We applied the cluster-based permutation approach to test the regression betas against zero.

Spatial similarity of MI topographies was obtained as the correlation coefficient between the group-averaged topographies in each frequency band, averaged across envelopes and carriers. Statistical significance of these correlation values was obtained with a bootstrap procedure, by randomly sampling participants with replacement 2000 times. Correlations excluding zero at an alpha level of 0.01 were

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

considered significantly positive. Then for each significant results we additionally computed cohen's d as an indicator of effect size.

To test for differences in the frequency of acoustic onsets between bands we compared their onset frequency with a one-way repeated measures ANOVA followed by a post hoc Tukey-Kramer multiple comparison test. A threshold of 0.01 was considered statistically significant. Additionally, to test the joint onset probability for significance we created a null distribution by randomly shifting each envelope separately in time and computing the randomized likelihood 2000 times. For all statistical tests we provide exact p-values, except where these are below 10^{-3}.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3. Results

Across five blocks participants listened to acoustic stimuli presenting band-limited noises modulated by speech-derived band-limited envelopes. Within and between blocks we manipulated the spectral match between envelope (i.e. the frequency range in the natural speech over which the envelope was extracted) and carrier (i.e. the spectral band of the respective noise), participants' attention to specific envelope-carrier pairs, and varied the number of simultaneously presented envelope modulated carriers.

3.1 Behavioural Performance

In the first three blocks participants were asked to detect a silent gap to ensure they paid attention to the stimuli. Gaps were reported with an accuracy of 95.4 ± 0.4 % (mean ± SD; averaged over participants and blocks 1-3). In the fourth block, participants were presented with two carriers but had to detect a gap in one 'attended' stimulus: they did so correctly in 79.3 ± 9.1% of trials. In the fifth block, participants listened to noise-vocoded speech with varying spectral detail (3, 6 and 12 envelopes) and were asked to perform a 3-choice word recognition task. Accuracy increased from 45.1 ± 11.4 % to 91.8 ± 6.7 % and 97.9 ± 4.0 % across conditions. In this block they also indicated their confidence on an interval scale from 1 to 3, which increased over the three conditions from 1.34 ± 0.43 to 2.65 ± 0.21 and 2.93 ± 0.08.

3.2 Statistical properties of band-limited envelopes

To understand the temporal structure of the band-limited speech envelopes we investigated their statistical regularities (Fig. 2A). First, we compared the similarity (Pearson correlation) between each band-limited envelope and the overall envelope (Fig. 2B). While Env_{low} (0.2 - 0.83 kHz) was strongly correlated with the overall envelope (R^2=0.89 ± 0.01, mean ± s.e.m. across sentences), the similarities between Env_{mid} (0.83 - 2.66 kHz) and Env_{high} (2.66 - 8 kHz) and the overall envelope were much reduced: a one-way ANOVA revealed a significant effect of band ($F(2,114)=6.57$, $p<10^{-4}$, $\eta^2=0.94$) and a Tukey-Kramer post hoc comparison revealed all pairwise comparisons as significant (each $p<10^{-4}$). Second, we investigated the structure of acoustic onsets (Fig. 2A, dots). Focusing on these three spectral bands we found that Env_{low} featured the highest onset frequency (3.57 ± 0.44 Hz), followed by Env_{mid} (2.94 ± 0.73 Hz) and Env_{high} (2.46 ± 0.83 Hz; Fig 2B). A one-way ANOVA revealed a significant difference between bands ($F(2,114)=26.32$, $p<10^{-4}$, $\eta^2=0.32$) and a Tukey-Kramer post hoc comparison revealed all pairwise comparisons as significant (each $p<0.0065$). We then asked if acoustic onsets are temporally co-structured between bands. For this, we computed their joint probability across frequency bands and time-lags, using a division into 12 bands (Fig. 2C). This revealed that onsets in neighboring bands occur at the same time, but this dependency decreases distinctively with spectral distance: acoustic onsets in bands approximately >3kHz were largely independent from onsets in bands below 3 kHz.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Last, we asked whether the acoustic onsets are related to the phonemic content of each envelope. For this we computed the probability of occurrence of individual phonemes near the acoustic onsets and divided these occurrences into those 'unique' and 'common' between bands (Fig. 2D). Acoustic onsets in Env_{high} coincided most with fricatives and plosives and were largely unique to these. On the contrary, acoustic onsets from Env_{mid} and Env_{low} were more likely to coincide with vowels, pure vowels and nasals. Together these results show that speech envelopes derived from distinct frequency ranges are temporally independent and carry partially complementary information about the occurrence of distinct types of phonemes.

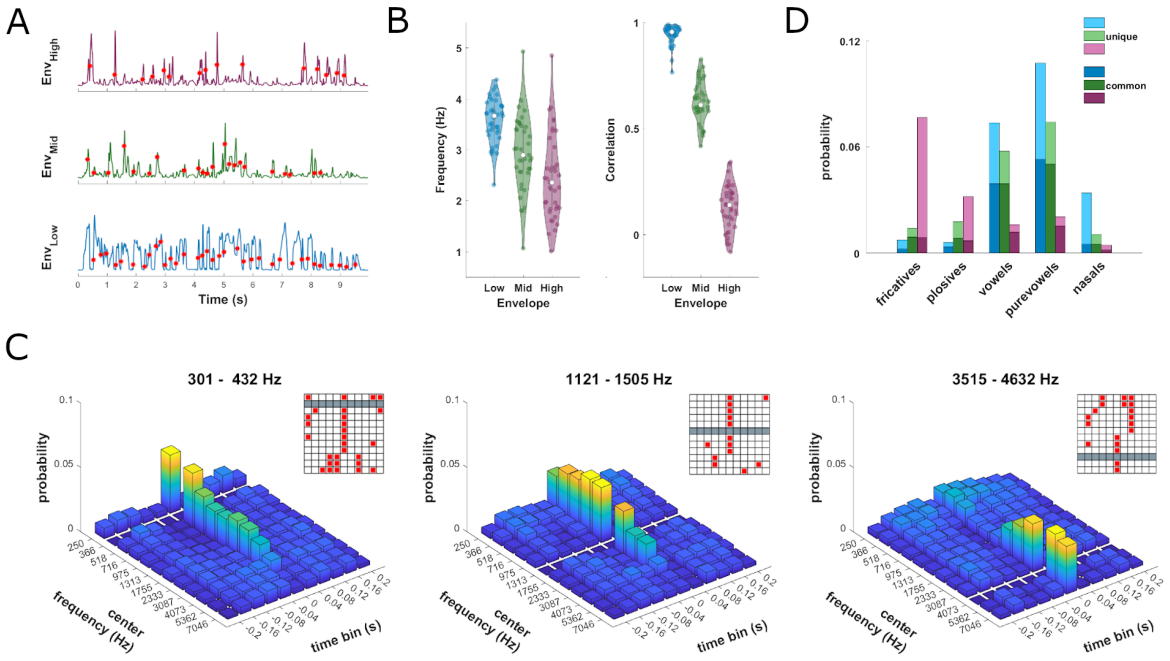

Fig. 2. Acoustic features in band-limited speech envelopes. (**A**) Time course of three band-limited envelopes (boundaries between low, mid and high: 0.2, 0.83, 2.66, 8 kHz) from a single sentence. For each envelope band we detected acoustic onsets (red dots) as peaks in the rate of change (first derivative). (**B**) Left: Frequency of onsets in each envelope band across sentences (dots). The lowest band exhibits the highest average onset frequency and is most consistent across all sentences, followed by the middle and the high band. Right: Correlation of band-limited envelopes from three frequency ranges with the overall envelope (see Methods for definition). (**C**) Temporal structure of envelope onsets computed for a finer spectral resolution using 12 bands. For each band-limited envelope we determined the joint probability of onsets across bands that fell within a 400 ms window (divided into 40 ms bins). Envelopes are referred to in terms of their center frequency with the respective reference envelope being cleared out (gray bars). Inlays show temporal bins with significant likelihoods (red boxes). (**D**) Phoneme probability near acoustic onsets (+/-40ms), divided by envelope and subgroup. Subgroups are defined as 'unique' and 'common' phonemes occurring across frequency bands.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

3.3 Tracking of spectrally limited speech envelopes

In block 1 we presented five conditions consisting of different combinations of envelope-modulated band-limited noises. Three conditions presented natural pairings, i.e. the envelope and carrier covered the same frequency range, while two conditions presented unnatural pairings (Fig. 1B). For each condition we quantified the envelope tracking as the mutual information between the complex-valued representation of the envelope and the EEG activity (Ince et al., 2017; Keitel et al., 2017).

First, we quantified envelope tracking separately for carrier noises and for envelopes derived from the low and high frequency regimes (Car_{low}, Car_{high}, Env_{low}, Env_{high} Fig. 3). For each of these, tracking prevailed at low EEG frequencies and over central and temporal electrodes, as expected from previous work (Giordano et al., 2017; Kayser et al., 2015; Keitel et al., 2018). We then asked whether the neurophysiological processes giving rise to these MI topographies share a similar spatial layout, or rather reflect spatially distinct generators. The similarities between condition-averaged MI topographies were significantly correlated across neighboring EEG bands, effectively revealing four clusters: delta1 (0.5 - 2 Hz), delta2 & theta1 (1 - 6 Hz), theta1 & theta2 (2 - 8 Hz), and alpha (8 - 12 Hz; Fig. 3B). Concerning the role of carrier or envelope bands, we found that the topographies obtained using different carriers were highly similar (averaged across EEG bands; $r=0.887$, $p<10^{-3}$, cohen's $d=19.84$), while the topographies obtained using low and high envelope rages were only moderately correlated ($r=0.327$, $p<10^{-3}$, cohen's $d=4.46$). These results suggest that envelope tracking in different bands of the EEG likely reflects distinct neurophysiological processes. These tracking signatures are not influenced by the choice of carrier frequency but speech envelopes extracted from low (0.2 - 0.83 kHz) and high (2.66 - 8 kHz) spectral ranges seem to be reflected by distinct neurophysiological signatures.

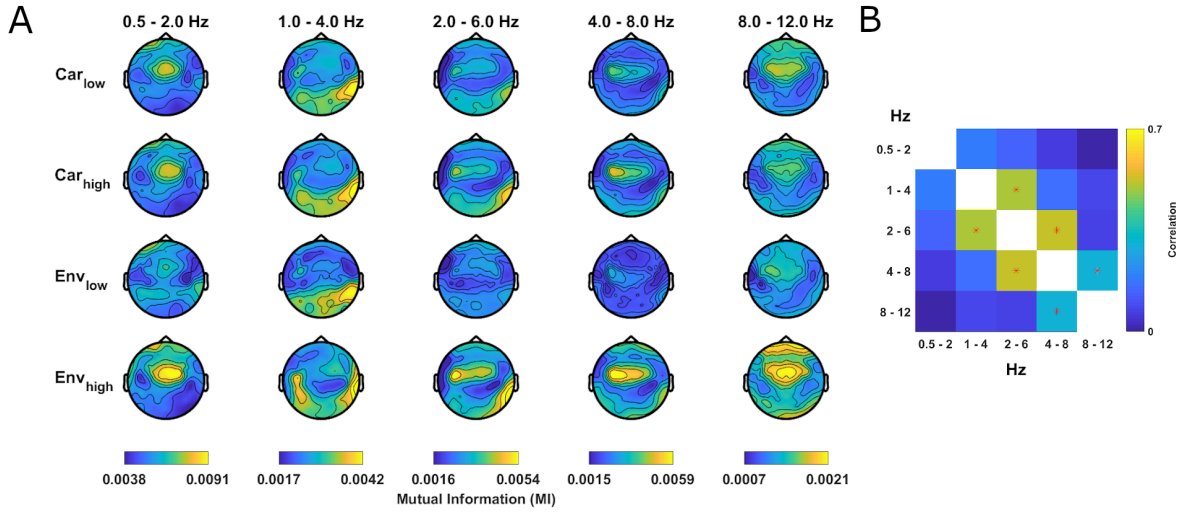

Fig. 3. Specificity and spatial topographies of envelope tracking. (**A**) Group-average topographies of envelope tracking, quantified as mutual information (MI), separately for low/high carrier bands (Car_{low}, Car_{high}) and low/high envelope ranges (Env_{low}: 0.2 - 0.83 kHz, Env_{high}: 2.66 - 8 kHz) in block 1. Here, the MI values were averaged across the respective other dimension (e.g. over envelope bands for Car_{low}, Car_{high}). Color ranges are

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

fixed within each EEG band. (**B**) Spatial similarity (Pearson correlation) of group-level MI topographies between EEG bands (averaged over conditions). Red stars: $p < 0.01$ for a bootstrap test against zero.

3.4 High-frequency envelopes are reflected more strongly in the EEG

Addressing the first of our main questions, we asked whether these signatures of envelope tracking are specific to the match between the carrier frequency and the spectral range from which the envelope was extracted. A direct contrast of MI values between conditions in which Car_{low} or Car_{high} were presented (Fig. 4, ΔCar) revealed no significant effects except in the alpha band, where MI values were higher for Car_{high} ($p<10^{-3}$, cstat=-76.08, cohen's d=-0.78). A direct contrast between conditions in which either Env_{low} or Env_{high} were presented (Fig. 4, ΔEnv) revealed higher MI for Env_{high} over central channels in delta1 (0.5 - 2 Hz; $p<10^{-3}$, cstat=23.86, cohen's d=0.68), theta1 and theta2 (2 - 8 Hz; $p<10^{-3}$, cstat=177.83, cohen's d=0.96), being conjugated from two separate clusters. Finally, we asked whether the interaction between envelope and carriers matters. For this we compared unnatural (Fig 1B, signals 2,4) and natural pairings (Fig 1B, signals 1, 5) (Fig. 4, Inter), and found higher MI values for unnatural pairings in the lowest EEG band (delta 1) over right posterior electrodes (0.5 - 2 Hz; $p<10^{-3}$, cstat=17.84, cohen's d=0.71). Together with the topographical differences in the tracking of Env_{low} and Env_{high} (c.f. Fig. 3), these results suggest that speech envelopes derived from lower and higher spectral bands are reflected differently in the EEG.

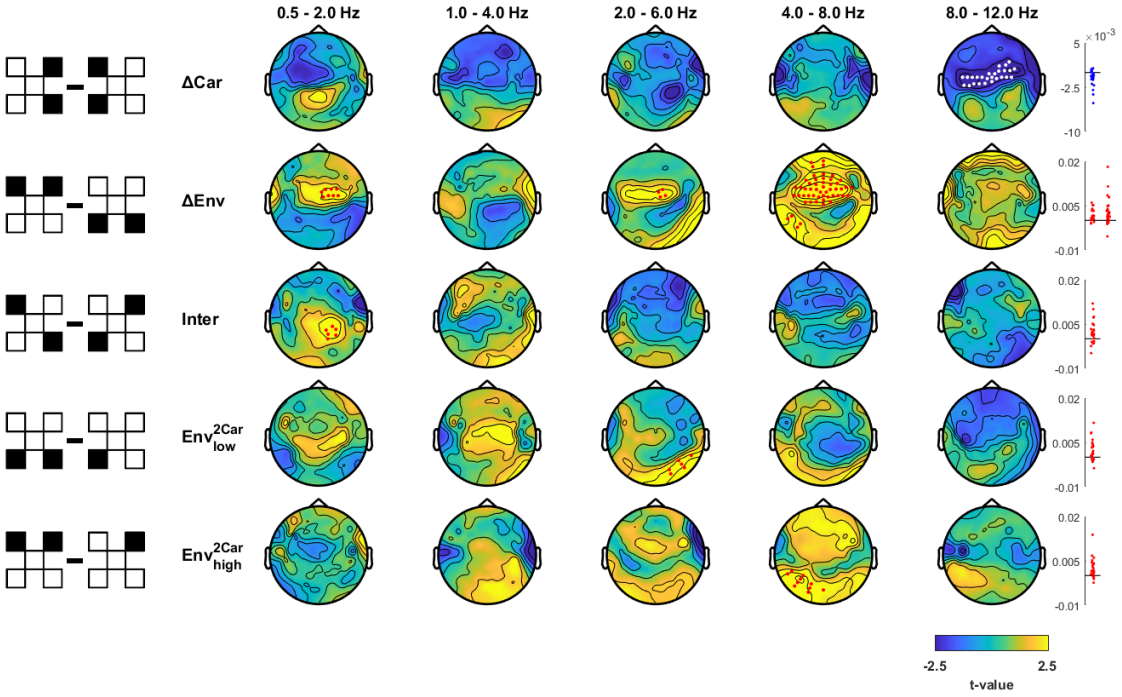

Fig. 4. Effects of carrier frequency and envelope band on envelope tracking. Differences in MI values between Car_{high} minus Car_{low} (ΔCar), Env_{high} minus Env_{low} (ΔEnv), unnatural minus natural pairings (Inter), and the

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

enhancement of envelope tracking when presented with two minus one carrier (Env_{low}^{2Car}, Env_{high}^{2Car}). Topographies show the electrode-wise t-values for the group-level differences between conditions of interest. Red and white dots mark positive and negative clusters (derived using cluster-based permutation statistics corrected across electrodes and EEG bands, $p<0.01$). Panels on the right show differences in MI values between conditions for individual participants within each cluster (red for positive and blue for negative clusters). Pictograms on the left indicate which conditions (black filled squares) were summed and contrasted (c.f. Fig. 1B).

3.5 Contextual information influences tracking of Env_{high} more than Env_{low}

Using the conditions in blocks 2 and 3 we examined whether and how the tracking of individual envelopes is influenced by concurrent information presented to the listener. First, we asked whether the tracking of a specific envelope is enhanced when the same temporal envelope is presented in a broader spectral range. In block 2, we presented carriers at two frequency ranges, each modulated by the same envelope (2Car condition; Fig 4). Comparing this 2-carrier condition to the respective 1-carrier condition (from block 1) revealed higher MI values for the 2-carrier condition: for Env_{low} over right posterior electrodes in theta1 (Fig. 4, 2 - 6 Hz; Env_{low}^{2Car}, $p<10^{-3}$, cstat=18.43, cohen's d=0.74); for Env_{high} over left posterior electrodes in theta2 (4 - 8 Hz; $p<10^{-3}$, cstat=31.45, cohen's d=0.97). This shows that envelope tracking is enhanced when the same information is present in a broader spectral range.

In block 3 we presented two carriers each modulated by their respective matching (natural) envelopes in order to probe whether the tracking of a specific envelope is enhanced when the complementary speech-like information is presented in another spectral range (i.e. presenting Env_{low} carried by Car_{low} and Env_{high} carried by Car_{high} together Fig. 5). MI values were lower in the 2-carrier condition compared to the respective values obtained from block 1: for Env_{low} MI was reduced over right central-posterior electrodes in theta2 (Fig. 5, 4 - 8 Hz; $p<10^{-3}$, cstat=-37.02, cohen's d=-0.74); for Env_{high} MI values were reduced over posterior electrodes in the theta1 to theta2 bands (2 - 8 Hz; $p<10^{-3}$, cstat=-149.67, cohen's d=-1.04) and over central electrodes in theta2 band (4 - 8 Hz; $p<10^{-3}$, cstat=-21.53, cohen's d=-0.58). A direct comparison between Env_{high} and Env_{low} revealed a stronger reduction of MI values for Env_{high} over left posterior sites in theta2 (4 - 8 Hz; $p<10^{-3}$, cstat=36.44, cohen's d=-0.97). Hence, the tracking of a specific envelope-carrier pair in the theta band becomes less prominent when an additional complementary speech envelope is presented and this effect is particularly prominent for Env_{high}.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

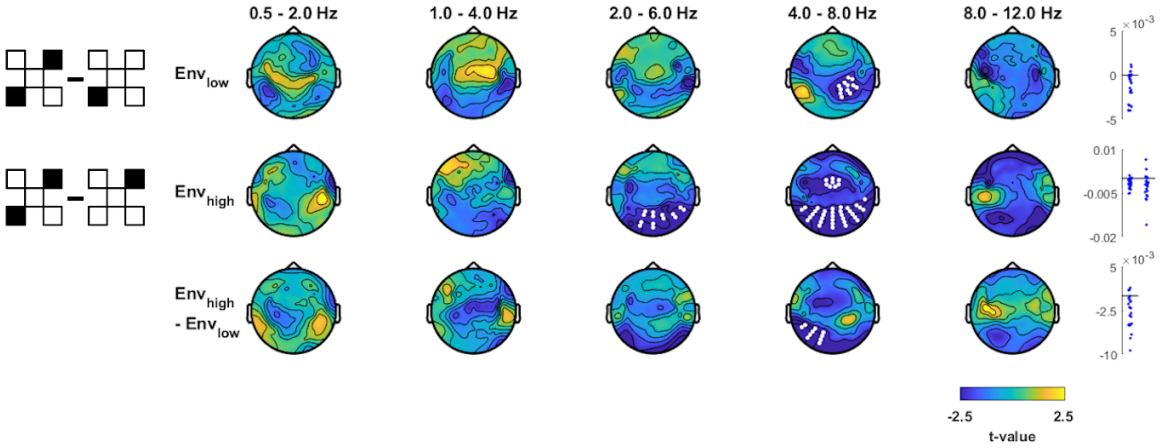

Fig. 5. Effects of contextual information. Tracking of one individual envelope-noise pair with a second envelope-noise pair presented (over a different frequency band) minus tracking of just the individual envelope. Topographies show the electrode-wise t-values for the group-level statistics. Red and white dots mark positive and negative clusters (derived using cluster-based permutation statistics, $p<0.01$). Panels on the right show differences in MI values between conditions for individual participants. Pictograms on the left indicate which conditions (black filled squares) were summed and contrasted.

3.6 Attention influences tracking of Env_{high} more than Env_{low}

We then asked how focused attention affects the tracking of a specific envelope. In block 4 we presented a mixture of two natural envelope-carrier pairs similar to block 3, but asked participants to attend to either Env_{low} or Env_{high}. The contrast between the unfocused attention in block 3 (target gap in both envelopes) and block 4 (focused attention) revealed no effect for Env_{low}, while the MI for Env_{high} was enhanced by attention over left central electrodes in delta1 (Fig. 6, 0.5 - 2 Hz; $p<10^{-3}$, cstat=65.91, cohen's d=0.84) and over posterior electrodes in theta1 and theta2 (2 - 8 Hz; $p<10^{-3}$, cstat=31.14, cohen's d=0.77), being conjugated from two separate clusters. To understand whether the attentional modulation differed between Env_{low} or Env_{high}, we directly contrasted the attention effect between these (Fig. 6, Env_{high} - Env_{low}). This revealed that attention had a larger influence on the tracking of Env_{high} than of Env_{low} over central posterior electrodes in theta2 (4 - 8 Hz; $p<10^{-3}$, cstat=11.3, cohen's d=0.63).

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

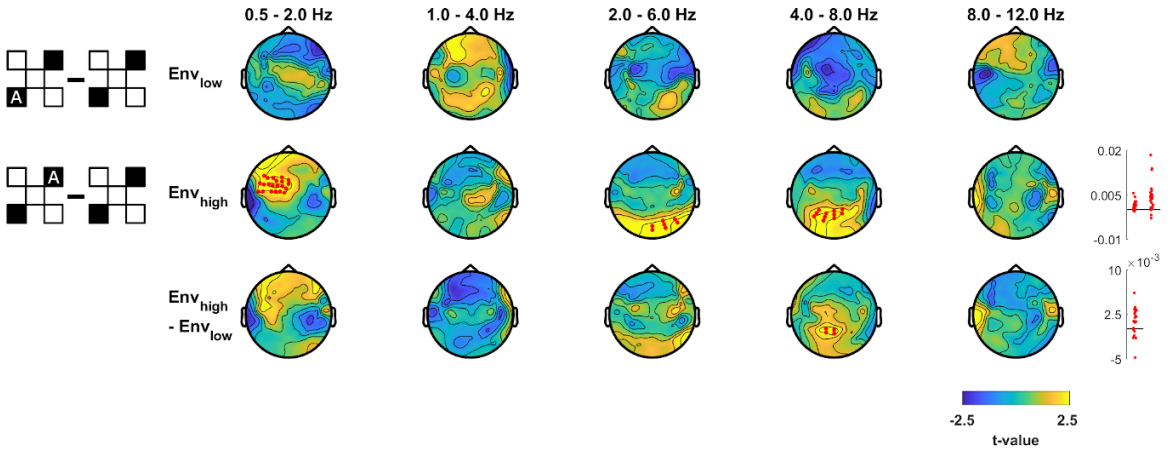

Fig. 6. Effect of attention on envelope tracking. Tracking of individual envelopes in conditions where one envelope was actively attended minus conditions in which attention was not focused on a specific envelope. Topographies show the electrode-wise t-values for the group-level statistics. Red and white dots mark positive and negative clusters (derived using cluster-based permutation statistics, $p<0.01$). Panels on the right show MI values for individual participants in each cluster (red for positive and blue for negative clusters). Pictograms on the left indicate which conditions (black filled squares) were summed and then contrasted (A indicating the attended signal).

3.7 Tracking of Env_{low} prevails when speech becomes intelligible

Combining the data from blocks 1, 2 and 5 we asked how the tracking of one specific band-limited envelope from among a mixture of multiple natural envelope-noise pairs is affected when the amount of spectral detail is increased. Block 5 presented noise-vocoded speech using 3, 6 and 12 bands, whereby the 12-band condition was not analyzed as participant's comprehension was near ceiling for 6 bands already (see above).

First, we investigated how the tracking of Env_{low} and Env_{high} scaled with the number of envelopes (Fig. 7A). The MI values for Env_{low} generally increased: in delta1 and theta1 over fronto-central electrodes (0.5 - 6 Hz; $p<10^{-3}$, cstat=125.04, cohen's d=0.77) and in theta1 over frontal electrodes (2 - 6 Hz; $p<10^{-3}$, cstat=11.63, cohen's d=0.79). In contrast, MI values for Env_{high} generally decreased: in theta1 over central (2 - 6 Hz; $p<10^{-3}$, cstat=-10.78, cohen's d=-0.57), left posterior ($p<10^{-3}$, cstat=-12.91, cohen's d=-0.71) and right posterior electrodes ($p<10^{-3}$, cstat=-10.74, cohen's d=-0.58), and additionally in the alpha band over posterior electrodes (8 - 12 Hz; $p<10^{-3}$, cstat=-40.23, cohen's d=-0.73). To confirm this differential effect, we performed a direct contrast between Env_{low} and Env_{high} for each level of spectral detail (Fig. 7B). For the 3 envelope condition, which was generally unintelligible (Fig. 7B, ΔEnv 3Bands), we found higher MI for Env_{high} in theta2 over central electrodes (4 - 8 Hz; $p<10^{-3}$, cstat=89.64, cohen's d=0.64) and higher MI for Env_{low} in delta2 over central electrodes (1 - 4 Hz; $p<10^{-3}$, cstat=-98.96, cohen's d=-0.79). For 6 envelopes (Fig. 7B, ΔEnv 6Bands), when participants could readily

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

understand the sentences, we found stronger tracking of Env_{high} in alpha over frontal electrodes (8 - 12 Hz; $p<10^{-3}$, cstat=33, cohen's d=0.83) and stronger tracking of Env_{low} in a cluster ranging from delta1 to theta1 over fronto-central electrodes (0.5 - 6 Hz; $p<10^{-3}$, cstat=-368.11, cohen's d=-0.88). Figure 7C directly illustrates this differential shift of envelope tracking in the delta2 & theta1 (1 - 6 Hz) EEG bands from Env_{high} to Env_{low} with increasing number of carriers.

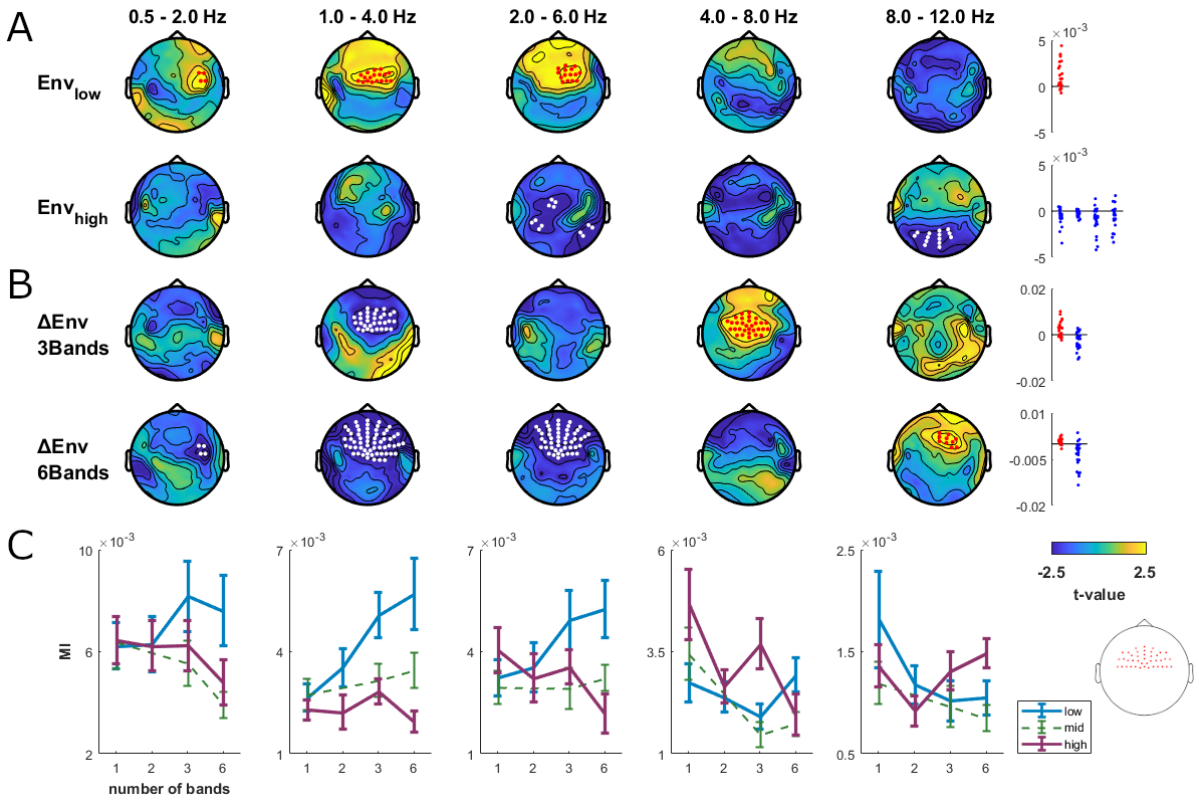

Fig. 7. Differential envelope tracking with increased spectral detail. (**A**) Change in envelope tracking in conditions with 1, 2, 3 and 6 concurrently presented envelope-carrier pairs, computed separately for Env_{low} and Env_{high}. MI values for Env_{low} are enhanced with more spectral detail in delta1 to theta1 (1 - 6 Hz), while MI values of Env_{high} decreases in theta1 (2 - 6 Hz). (**B**) MI for Env_{high} minus Env_{low} in conditions with 3 and 6 simultaneously presented envelopes. Topographies show the electrode-wise t-values for the group-level statistics (derived using cluster-based permutation statistics, $p<0.01$). Panels on the right show individual data of each cluster (red for positive and blue for negative clusters). (**C**) MI values averaged across fronto-central electrodes (inlay) for Env_{low}, Env_{mid}, and Env_{high}. Error bars indicate s.e.m. across subjects.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

4. Discussion

The representation of speech in the brain is often examined by measuring the alignment of rhythmic brain activity to the acoustic envelope of the signal (Ahissar et al., 2001; Ding et al., 2014; Giraud and Poeppel, 2012; Gross et al., 2013; Kayser et al., 2015; Oganian and Chang, 2019; Teng et al., 2019; Teoh et al., 2019). To quantify this alignment (here technically termed speech tracking) many studies rely on the overall envelope, which describes the amplitude fluctuation of speech across the spectral range and which provides a convenient and low-dimensional representation for data analysis. Our results show that this overall envelope combines speech signatures that are seemingly encoded by distinct neurophysiological processes giving rise to spatially distinct signatures of speech tracking in the human EEG. Because low and high frequency speech envelopes relate to separate acoustically and phonetically features (Fig. 2), the neural processes encoding these spectrally distinct envelopes likely carry complementary information for speech encoding.

4.1 The overall envelope provides a distorted picture on speech tracking

We probed whether synthetic and incomprehensible sounds carrying typical regularities of natural speech envelopes are reflected in low frequency EEG activity in a similar manner as real speech. In large our data show that this is indeed the case: The speech-derived amplitude envelopes were tracked over fronto-central electrodes at EEG frequencies below 8 Hz, consistent with previous studies using speech (Drennan and Lalor, 2019; Etard and Reichenbach, 2019; Kayser et al., 2015; Mai and Wang, 2019; Synigal et al., 2019).

However, the envelopes derived from a higher spectral range of natural speech (Env_{high} defined here as 2.66 - 8 kHz) were reflected in spatially distinct EEG topographies compared to the envelopes from the lower spectral range (Env_{low} here defined as 0.2 - 0.83 kHz), suggesting that acoustically segregated aspects of speech dynamics are reflected in distinct neurophysiological processes (Fig. 3). In addition, while we observed envelope tracking over fronto-central electrodes, effects of contextual information and attention were most prominent in occipital electrodes and more so for Env_{high} than for Env_{low} (Fig. 4, 5). As such, the observed effects are spatially distinct and not equally sensitive for both envelopes.

Interestingly, the differential tracking of Env_{high} and Env_{low} was not specific to the frequency of the respective carrier noise, indicating that the underlying processes are not selective to the natural spectral match of carrier and speech-like envelope (Fig. 3). Together with the differential sensitivity of the tracking of Env_{high} and Env_{low} to contextual and attention manipulations, this shows that combining band-limited speech envelopes into an overall speech envelope for data analysis provides a distorted picture on the encoding of speech in dynamic brain activity: both in terms of the topographical and spectral distribution of the respective signatures in the EEG, and in terms of their sensitivity to experimental manipulations. Given that the representation of Env_{low} prevailed for rich and comprehensible speech (6 bands noise-vocoding), and given the high correlation between Env_{low} and

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the overall envelope, we posit that studies focusing on the overall envelope of natural speech mostly capture the tracking of the speech envelope from the lower spectrum. Yet, when the acoustic signal is impoverished, this will confound distinct neurophysiological processes.

4.2 Differential encoding of low and high frequency speech envelopes

Our data show that the speech envelopes obtained from low and high frequency ranges are reflected differentially in the 1 - 6 Hz EEG. Speech tracking has been frequently related to the neural encoding of acoustic onsets, in particular as acoustic transients drive individual neurons and induce an alignment of rhythmic brain activity to the acoustic signal (Daube et al., 2019; Drennan and Lalor, 2019; Khalighinejad et al., 2017; Oganian and Chang, 2019). The analysis of the acoustic speech material showed that low and high frequency envelopes can reflect independent acoustic landmarks which relate, in part, to the unique signaling of the occurrence of specific phonemes. Our results hence suggest that the preferential tracking of the low frequency envelope for (near) intelligible speech can be functionally meaningful. The low frequency envelope in real speech carries more energy (Peelle et al., 2013), and lowpass filtered speech has been shown to be sufficient for comprehension (Elliott and Theunissen, 2009). Prominent acoustic landmarks in the low frequency envelope relate to the occurrence of a large group of phonemes, and hence the prominent tracking of the low frequency envelope may reflect an adaptation to the typically most robust aspect of the speech signal.

4.3 The neurophysiological processes underlying speech tracking

Our results corroborate the notion that the tracking of speech envelopes in delta and theta band brain signals represent functionally distinct processes. Delta-band tracking (below 4 Hz) has been shown to covary with experimental manipulations affecting intelligibility (Ding et al., 2016; Zion Golumbic et al., 2013; Zoefel et al., 2018) and has been related to temporally extended supra-syllabic features (Keitel et al., 2018; Mai and Wang, 2019). Nevertheless, it is still debated whether delta tracking is causally related to comprehension. While some studies advocate for such a role (Etard and Reichenbach, 2019; Peelle et al., 2013; Wilsch et al., 2018; Zoefel et al., 2018), experiments using pop-out stimuli did not find a clear relation (Baltzell et al., 2017; Millman et al., 2015) and a manipulation of speech rhythm resulted in reduced delta tracking while preserving intelligibility (Kayser et al., 2015). In the present data tracking of Env_{low} in the 1 - 4 Hz EEG increased monotonically with acoustic detail rather than correlating with comprehension performance. One possibility is that delta band tracking reflects processes that screen sounds for speech-specific regularities, such as sentence-level structures or prosody, to which early stage spectro-temporal filters could be tuned to (Schönwiesner and Zatorre, 2009). These processes then drive speech encoding prior to comprehension (Giraud and Poeppel, 2012; Keitel et al., 2018; Schroeder and Lakatos, 2009; Scott, 2019).

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Theta band tracking has been linked to the bottom up processing of the acoustic input, in particular of syllabic and sub-syllabic features (Ding et al., 2014; Ding and Simon, 2013; Etard and Reichenbach, 2019; Keitel et al., 2018; Mai and Wang, 2019; Peelle et al., 2013; Rimmele et al., 2015; Scott, 2019). In contrast to some previous studies (Ding et al., 2014; Peelle et al., 2013; Rimmele et al., 2015), in the present data tracking in theta (4 - 8 Hz) and alpha bands (8 - 12 Hz) did not increase with detail. Based on a similar observation a recent study argued that stronger theta tracking of impoverished sounds may reflect effects of listening effort (Hauswald et al., 2019). However, apparent discrepancies between studies may also arise from the distinct approaches used to study speech tracking: while some studies used encoding or reconstruction methods (Daube et al., 2019; Hausfeld et al., 2018; Zion Golumbic et al., 2013), others quantified the temporal alignment of speech envelope and brain activity (Hauswald et al., 2019; Peelle et al., 2013), and yet others quantified the between-trial alignment of brain activity (Ding et al., 2014; Rimmele et al., 2015). Furthermore, for vocoded speech or speech-in-noise some studies considered the envelope of the original speech (Ding and Simon, 2013; Peelle et al., 2013) while others used the envelope of the actually presented sound (Hauswald et al., 2019; Millman et al., 2015). In the present study, we quantified tracking of the same band-limited speech envelopes across conditions. With this approach theta band tracking did not exhibit a systematic in- or decrease with the number of simultaneous envelopes presented, but tracking of Env_{high} was significantly reduced by attentional and contextual manipulations, corroborating the notion that theta tracking is sensitive to listening effort (Hauswald et al., 2019).

Where are the neurophysiological processes giving rise to these tracking signatures located? Early auditory regions entrain to acoustic regularities (Ding and Simon, 2014; Lakatos et al., 2019; Meyer, 2018; Obleser and Kayser, 2019) but do so spatially specific along the tonotopic axis (Lakatos et al., 2013; O'connell et al., 2014). If these regions gave rise to the observed tracking signatures, one should hence expect a sensitivity of these to the match between the carrier frequency and the spectral range at which the envelope is extracted. However, we found no evidence for this in the delta and theta bands typically associated with speech encoding. Given that early auditory cortical regions are sensitive to the natural speech-specific regularities (Edwards and Chang, 2013; Iverson et al., 2016; Scott, 2019) this could suggest that these tracking signatures arise from higher superior temporal regions that encode temporal and spectral information largely independently (Sohoglu et al., 2020). Indeed, previous neuroimaging studies showed that delta and theta tracking prevails in large parts of the temporal lobe (Gross et al., 2013; Keitel et al., 2018). Still, further work is required to directly determine where in the brain low and high frequency speech envelopes are reflected and where the underlying neural representations give rise to the differential pattern of envelope sensitivity observed here.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

5. Conclusion

Low and high frequency speech envelopes provide independent information about acoustic and phonemic features in speech. Our results show that these envelopes are reflected in spatially and functionally distinct processes in the delta and theta EEG bands. These processes are easily confounded when considering the overall speech envelope for data analysis, but may offer independent windows on the neurobiology of speech.

Conflicts of interest

We declare no conflict of interest.

Acknowledgement

Funding: This work was supported by the European Research Council (to C.K. ERC-2014-CoG; grant 618 No 646657).

References

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. U. S. A. 98, 13367–13372. https://doi.org/10.1073/pnas.201400998

Aiken, S.J., Picton, T.W., 2008. Human cortical responses to the speech envelope. Ear Hear. 29, 139–157. https://doi.org/10.1097/AUD.0b013e31816453dc

Baltzell, L.S., Srinivasan, R., Richards, V.M., 2017. The effect of prior knowledge and intelligibility on the cortical entrainment response to speech. J. Neurophysiol. 118, 3144–3151. https://doi.org/10.1152/jn.00023.2017

Boersma, P., van Heuven, V., 2001. PRAAT, a system for doing phonetics by computer. Glot Int. 5, 341–347.

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. https://doi.org/10.1523/JNEUROSCI.0584-19.2019

Cusack, R., Deeks, J., Aikman, G., Carlyon, R.P., 2004. Effects of location, frequency region, and time course of selective attention on auditory scene analysis. J. Exp. Psychol. Hum. Percept. Perform. 30, 643–656. https://doi.org/10.1037/0096-1523.30.4.643

Daube, C., Ince, R.A.A., Gross, J., 2019. Simple Acoustic Features Can Explain Phoneme-Based

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

611 Predictions of Cortical Responses to Speech. Curr. Biol. 29, 1924-1937.e9.
612 https://doi.org/10.1016/J.CUB.2019.04.067
613 Demeester, K., Topsakal, V., Hendrickx, J.J., Fransen, E., Van Laer, L., Van Camp, G., Van De
614 Heyning, P., Van Wieringen, A., 2012. Hearing disability measured by the speech, spatial, and
615 qualities of hearing scale in clinically normal-hearing and hearing-impaired middle-aged
616 persons, and disability screening by means of a reduced SSQ (the SSQ5). Ear Hear. 33, 615–
617 626. https://doi.org/10.1097/AUD.0b013e31824e0ba7
618 Ding, N., Chatterjee, M., Simon, J.Z., 2014. Robust cortical entrainment to the speech envelope relies
619 on the spectro-temporal fine structure. Neuroimage 88, 41–6.
620 https://doi.org/10.1016/j.neuroimage.2013.10.054
621 Ding, N., Melloni, L., Zhang, H., Tian, X., Poeppel, D., 2016. Cortical tracking of hierarchical
622 linguistic structures in connected speech. Nat. Neurosci. 19, 158–64.
623 https://doi.org/10.1038/nn.4186
624 Ding, N., Simon, J.Z., 2014. Cortical entrainment to continuous speech: functional roles and
625 interpretations. Front. Hum. Neurosci. 8, 311. https://doi.org/10.3389/fnhum.2014.00311
626 Ding, N., Simon, J.Z., 2013. Adaptive temporal encoding leads to a background-insensitive cortical
627 representation of speech. J. Neurosci. 33, 5728–35. https://doi.org/10.1523/JNEUROSCI.5297-
628 12.2013
629 Doelling, K.B., Arnal, L.H., Ghitza, O., Poeppel, D., 2014. Acoustic landmarks drive delta-theta
630 oscillations to enable speech comprehension by facilitating perceptual parsing. Neuroimage 85,
631 761–768. https://doi.org/10.1016/j.neuroimage.2013.06.035
632 Drennan, D.P., Lalor, E.C., 2019. Cortical tracking of complex sound envelopes: modeling the
633 changes in response with intensity. eNeuro ENEURO.0082-19.2019.
634 https://doi.org/10.1523/ENEURO.0082-19.2019
635 Edwards, E., Chang, E.F., 2013. Syllabic (~2-5 Hz) and fluctuation (~1-10 Hz) ranges in speech and
636 auditory processing. Hear. Res. 305, 113–34. https://doi.org/10.1016/j.heares.2013.08.017
637 Elliott, T.M., Theunissen, F.E., 2009. The Modulation Transfer Function for Speech Intelligibility.
638 PLoS Comput. Biol. 5, e1000302. https://doi.org/10.1371/journal.pcbi.1000302
639 Erb, J., Kreitewolf, J., Pinheiro, A.P., Obleser, J., 2020. Aberrant perceptual judgements on speech-
640 relevant acoustic features in hallucination-prone individuals 1–25.
641 Etard, O., Reichenbach, T., 2019. Neural Speech Tracking in the Theta and in the Delta Frequency
642 Band Differentially Encode Clarity and Comprehension of Speech in Noise. J. Neurosci. 39,
643 5750–5759. https://doi.org/10.1523/JNEUROSCI.1828-18.2019
644 Gatehouse, S., Noble, W., 2004. The Speech, Spatial and Qualities of Hearing Scale (SSQ). Int J
645 Audiol. 43, 85–99.
646 Giordano, B.L., Ince, R.A.A., Gross, J., Schyns, P.G., Panzeri, S., Kayser, C., 2017. Contributions of
647 local speech encoding and functional connectivity to audio-visual speech perception. Elife 6.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

https://doi.org/10.7554/eLife.24763

Giraud, A.-L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. 15, 511–7. https://doi.org/10.1038/nn.3063

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. https://doi.org/10.1371/journal.pbio.1001752

Hausfeld, L., Riecke, L., Valente, G., Formisano, E., 2018. Cortical tracking of multiple streams outside the focus of attention in naturalistic auditory scenes. Neuroimage 181, 617–626. https://doi.org/10.1016/j.neuroimage.2018.07.052

Hauswald, A., Keitel, A., Rösch, S., Weisz, N., Hauswaldsbgacat, A., 2019. Degraded auditory and visual speech affects theta synchronization and alpha power differently. Biorxiv 0–25.

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. https://doi.org/10.1002/hbm.23471

Iverson, P., Wagner, A., Rosen, S., 2016. Effects of language experience on pre-categorical perception: Distinguishing general from specialized processes in speech perception. J. Acoust. Soc. Am. 139, 1799–1809. https://doi.org/10.1121/1.4944755

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, 14691–701. https://doi.org/10.1523/JNEUROSCI.2243-15.2015

Kayser, S.J., McNair, S.W., Kayser, C., 2016. Prestimulus influences on auditory perception from sensory representations and decision processes. Proc. Natl. Acad. Sci. U. S. A. 113, 4842–4847. https://doi.org/10.1073/pnas.1524087113

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

Keitel, A., Ince, R.A.A., Gross, J., Kayser, C., 2017. Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. Neuroimage 147, 32–42. https://doi.org/10.1016/j.neuroimage.2016.11.062

Kerlin, J.R., Shahin, A.J., Miller, L.M., 2010. Attentional gain control of ongoing cortical speech representations in a “cocktail party.” J. Neurosci. 30, 620–628. https://doi.org/10.1523/JNEUROSCI.3631-09.2010

Khalighinejad, B., da Silva, G.C., Mesgarani, N., 2017. Dynamic encoding of acoustic features in neural responses to continuous speech. J. Neurosci. 37, 2176–2185. https://doi.org/10.1523/JNEUROSCI.2383-16.2017

Kisler, T., Reichel, U., Schiel, F., 2017. Multilingual processing of speech via web services. Comput.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

685 Speech Lang. 45, 326–347. https://doi.org/10.1016/j.csl.2017.01.005
686 Lakatos, P., Gross, J., Thut, G., 2019. A new unifying account of the roles of neuronal entrainment.
687 Curr. Biol. 1–16. https://doi.org/10.1016/j.cub.2019.07.075
688 Lakatos, P., Karmos, G., Mehta, A.D., Ulbert, I., Schroeder, C.E., 2008. Entrainment of neuronal
689 oscillations as a mechanism of attentional selection. Science 320, 110–3.
690 https://doi.org/10.1126/science.1154735
691 Lakatos, P., Musacchia, G., O'Connel, M.N., Falchier, A.Y., Javitt, D.C., Schroeder, C.E., 2013. The
692 Spectrotemporal Filter Mechanism of Auditory Selective Attention. Neuron 77, 750–761.
693 https://doi.org/10.1016/j.neuron.2012.11.034
694 Mai, G., Wang, W.S.-Y., 2019. Delta and theta neural entrainment during phonological and semantic
695 processing in speech perception. bioRxiv 556837. https://doi.org/10.1101/556837
696 Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J.
697 Neurosci. Methods 164, 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024
698 McNair, S.W., Kayser, S.J., Kayser, C., 2019. Consistent pre-stimulus influences on auditory
699 perception across the lifespan. Neuroimage 186, 22–32.
700 https://doi.org/10.1016/j.neuroimage.2018.10.085
701 Mesgarani, N., Chang, E.F., 2012. Selective cortical representation of attended speaker in multi-talker
702 speech perception. Nature. https://doi.org/10.1038/nature11020
703 Meyer, L., 2018. The neural oscillations of speech processing and language comprehension: state of
704 the art and emerging mechanisms. Eur. J. Neurosci. 48, 2609–2621.
705 https://doi.org/10.1111/ejn.13748
706 Millman, R.E., Johnson, S.R., Prendergast, G., 2015. The role of phase-locking to the temporal
707 envelope of speech in auditory perception and speech intelligibility. J. Cogn. Neurosci. 27.
708 https://doi.org/10.1162/jocn_a_00719
709 Molinaro, N., Lizarazu, M., 2018. Delta(but not theta)-band cortical entrainment involves speech-
710 specific processing. Eur. J. Neurosci. 48, 2642–2650. https://doi.org/10.1111/ejn.13811
711 Monson, B.B., Hunter, E.J., Lotto, A.J., Story, B.H., 2014. The perceptual significance of high-
712 frequency energy in the human voice. Front. Psychol. 5.
713 https://doi.org/10.3389/fpsyg.2014.00587
714 Nourski, K. V., Reale, R.A., Oya, H., Kawasaki, H., Kovach, C.K., Chen, H., Howard, M.A., Brugge,
715 J.F., 2009. Temporal envelope of time-compressed speech represented in the human auditory
716 cortex. J. Neurosci. 29, 15564–15574. https://doi.org/10.1523/JNEUROSCI.3065-09.2009
717 O'connell, M.N., Barczak, A., Schroeder, C.E., Lakatos, P., 2014. Layer specific sharpening of
718 frequency tuning by selective attention in primary auditory cortex. J. Neurosci. 34, 16496–
719 16508. https://doi.org/10.1523/JNEUROSCI.2055-14.2014
720 O'Sullivan, J., Herrero, J., Smith, E., Schevon, C., McKhann, G.M., Sheth, S.A., Mehta, A.D.,
721 Mesgarani, N., 2019. Hierarchical Encoding of Attended Auditory Objects in Multi-talker

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

722 Speech Perception. Neuron. https://doi.org/10.1016/j.neuron.2019.09.007
723 Obleser, J., Kayser, C., 2019. Neural Entrainment and Attentional Selection in the Listening Brain.
724 Trends Cogn. Sci. 23, 913–926. https://doi.org/10.1016/j.tics.2019.08.004
725 Oganian, Y., Chang, E.F., 2019. A speech envelope landmark for syllable encoding in human superior
726 temporal gyrus. Sci. Adv. 5, 1–14. https://doi.org/10.1126/sciadv.aay6279
727 Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.-M., 2011. FieldTrip: Open source software for
728 advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput. Intell.
729 Neurosci. 2011, 156869. https://doi.org/10.1155/2011/156869
730 Park, H., Kayser, C., Thut, G., Gross, J., 2016. Lip movements entrain the observers' low-frequency
731 brain oscillations to facilitate speech intelligibility. Elife 5. https://doi.org/10.7554/eLife.14521
732 Peelle, J.E., Gross, J., Davis, M.H., 2013. Phase-locked responses to speech in human auditory cortex
733 are enhanced during comprehension. Cereb. Cortex 23, 1378–1387.
734 https://doi.org/10.1093/cercor/bhs118
735 Poeppel, D., Assaneo, M.F., 2020. Speech rhythms and their neural foundations. Nat. Rev. Neurosci.
736 https://doi.org/10.1038/s41583-020-0304-4
737 Rimmele, J.M., Zion Golumbic, E., Schröger, E., Poeppel, D., 2015. The effects of selective attention
738 and speech acoustics on neural speech-tracking in a multi-talker scene. Cortex 68.
739 https://doi.org/10.1016/j.cortex.2014.12.014
740 Schönwiesner, M., Zatorre, R.J., 2009. Spectro-temporal modulation transfer function of single voxels
741 in the human auditory cortex measured with high-resolution fMRI. Proc. Natl. Acad. Sci. U. S.
742 A. 106, 14611–14616. https://doi.org/10.1073/pnas.0907682106
743 Schroeder, C.E., Lakatos, P., 2009. Low-frequency neuronal oscillations as instruments of sensory
744 selection. Trends Neurosci. 32, 9–18. https://doi.org/10.1016/j.tins.2008.09.012
745 Scott, S.K., 2019. From speech and talkers to the social world: The neural processing of human
746 spoken language. Science (80-.). https://doi.org/10.1126/science.aax0288
747 Sohoglu, E., Kumar, S., Chait, M., Griffiths, T.D., 2020. Multivoxel codes for representing and
748 integrating acoustic features in human cortex. Neuroimage 116661.
749 https://doi.org/10.1016/j.neuroimage.2020.116661
750 Synigal, S.R., Teoh, E.S., Lalor, E.C., 2019. Including measures of high gamma power can improve
751 the decoding of natural speech from EEG.
752 Teng, X., Cogan, G.B., Poeppel, D., 2019. Speech fine structure contains critical temporal cues to
753 support speech segmentation. Neuroimage 202, 116152.
754 https://doi.org/10.1016/j.neuroimage.2019.116152
755 Teoh, E.S., Cappelloni, M.S., Lalor, E.C., 2019. Prosodic pitch processing is represented in delta-
756 band EEG and is dissociable from the cortical tracking of other acoustic and phonetic features.
757 Eur. J. Neurosci. 50, 3831–3842. https://doi.org/10.1111/ejn.14510
758 Teoh, E.S., Lalor, E.C., 2020. Attention differentially affects acoustic and phonetic feature encoding

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.26.221838; this version posted July 26, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

759 in a multispeaker environment. bioRxiv 2020.06.08.141234.
760 https://doi.org/10.1101/2020.06.08.141234
761 Wilsch, A., Neuling, T., Obleser, J., Herrmann, C.S., 2018. Transcranial alternating current
762 stimulation with speech envelopes modulates speech comprehension. Neuroimage 172, 766–
763 774. https://doi.org/10.1016/j.neuroimage.2018.01.038
764 Zion Golumbic, E.M., Ding, N., Bickel, S., Lakatos, P., Schevon, C.A., McKhann, G.M., Goodman,
765 R.R., Emerson, R., Mehta, A.D., Simon, J.Z., Poeppel, D., Schroeder, C.E., 2013. Mechanisms
766 underlying selective neuronal tracking of attended speech at a “cocktail party”. Neuron 77, 980–
767 91. https://doi.org/10.1016/j.neuron.2012.12.037
768 Zoefel, B., Archer-Boyd, A., Davis, M.H., 2018. Phase Entrainment of Brain Oscillations Causally
769 Modulates Neural Responses to Intelligible Speech. Curr. Biol. 28, 401-408.e5.
770 https://doi.org/10.1016/j.cub.2017.11.071
771