Neural tracking of speech envelope does not unequivocally reflect intelligibility

Anne Kösem, Bohan Dai, James M Mcqueen, Peter Hagoort

NeuroImage 272 (2023) 120040

Contents lists available at ScienceDirect

NeuroImage

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

Neural tracking of speech envelope does not unequivocally reflect intelligibility

Anne Kösem[a,b,c,1,*], Bohan Dai[a,b,1], James M. McQueen[a,b], Peter Hagoort[a,b]

[a] Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands
[b] Donders Institute for Brain, Cognition and Behaviour, Radboud University, 6500 HB Nijmegen, The Netherlands
[c] Lyon Neuroscience Research Center (CRNL), CoPhy Team, INSERM U1028, Bron 69500, France

ARTICLE INFO

Keywords:
Speech
Language
Entrainment
Neural oscillations
MEG

ABSTRACT

During listening, brain activity tracks the rhythmic structures of speech signals. Here, we directly dissociated the contribution of neural envelope tracking in the processing of speech acoustic cues from that related to linguistic processing. We examined the neural changes associated with the comprehension of Noise-Vocoded (NV) speech using magnetoencephalography (MEG). Participants listened to NV sentences in a 3-phase training paradigm: (1) pre-training, where NV stimuli were barely comprehended, (2) training with exposure of the original clear version of speech stimulus, and (3) post-training, where the same stimuli gained intelligibility from the training phase. Using this paradigm, we tested if the neural responses of a speech signal was modulated by its intelligibility without any change in its acoustic structure. To test the influence of spectral degradation on neural envelope tracking independently of training, participants listened to two types of NV sentences (4-band and 2-band NV speech), but were only trained to understand 4-band NV speech. Significant changes in neural tracking were observed in the delta range in relation to the acoustic degradation of speech. However, we failed to find a direct effect of intelligibility on the neural tracking of speech envelope in both theta and delta ranges, in both auditory regions-of-interest and whole-brain sensor-space analyses. This suggests that acoustics greatly influence the neural tracking response to speech envelope, and that caution needs to be taken when choosing the control signals for speech-brain tracking analyses, considering that a slight change in acoustic parameters can have strong effects on the neural tracking response.

1. Introduction

Speech presents inherent rhythmic dynamics (Ding et al., 2017; Greenberg et al., 2003) to which brain activity synchronizes. Neural dynamics in the delta range (1–4 Hz) and theta range (4–8 Hz) in particular follow the slow temporal structure of speech (Ahissar et al., 2001; Gross et al., 2013; Luo and Poeppel, 2007). This neural tracking of speech envelope is thought to be an important mechanism that would contribute to syllabic and phrasal level segmentation (Greenberg et al., 2003) therefore influencing speech perception (Giraud and Poeppel, 2012; Peelle and Davis, 2012). Yet, a longstanding debate resides on the exact mechanistic role of neural tracking in speech processing (Ding and Simon, 2014; Doelling and Assaneo, 2021; Kösem and van Wassenhove, 2017; Lakatos et al., 2019; Obleser and Kayser, 2019). Speech comprehension requires a complex series of processing stages to extract meaning from sound. Therefore, neural envelope tracking could affect speech comprehension by modulating early auditory analysis, and/or later abstract linguistic processing. In the present paper, we asked which processing stages neural envelope tracking is involved in.

To test if neural envelope tracking has a specific role in speech processing, experimental designs usually contrast the neural response to speech with the response to an unintelligible “control" signal. The control often results from a modulation of the clear speech’s acoustics, for instance by reversing temporally the speech signal or varying its temporal properties (Ahissar et al., 2001; Broderick et al., 2018; Di Liberto et al., 2015; Doelling et al., 2014; Gross et al., 2013; Howard and Poeppel, 2010; Hincapié-Casas et al., 2021; Kayser et al., 2015; Pefkou et al., 2017), by degrading the spectral resolution (Chen et al., 2022; Hincapié-Casas et al., 2021; Meng et al., 2021; Molinaro and Lizarazu, 2018; Peelle et al., 2013), or by changing the auditory background (Ding and Simon, 2013; Rimmele et al., 2015; Zion Golumbic et al., 2013; Zoefel and VanRullen, 2015b). These studies mostly report

* Corresponding author at: Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands.
E-mail address: anne.kosem@inserm.fr (A. Kösem).
[1] Authors contributed equally.

https://doi.org/10.1016/j.neuroimage.2023.120040.
Received 6 September 2022; Received in revised form 13 March 2023; Accepted 15 March 2023
Available online 17 March 2023.

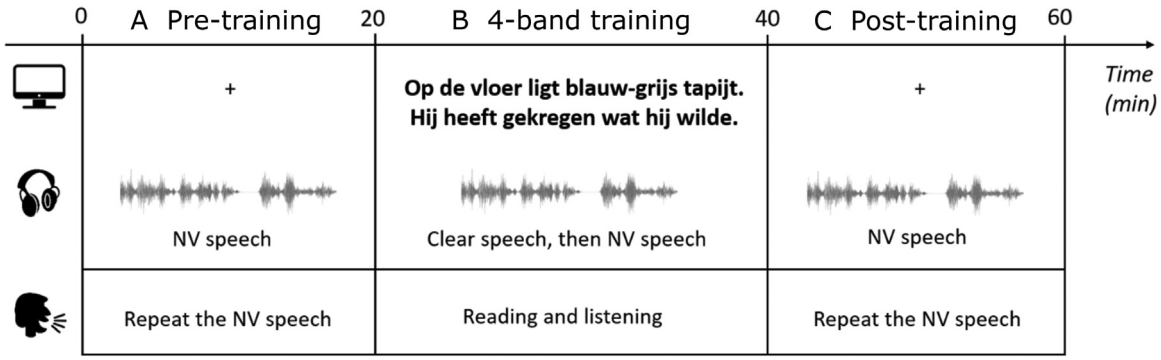

Fig. 1. Experimental design. The experiment consisted of three phases: pre-training (A), 4-band training (B), and post-training (C). In the pre- and post-training phases, the participants were tested on their ability to understand the 4-band and 2-band vocoded speech stimuli. They were presented with the speech signal binaurally and were asked to report the sentences afterwards. During the training phase, participants listened to clear-speech versions of the 4-band pre-training sentences followed by the NV versions. At the same time, they read the text of the sentences on the screen. The experiment's duration was of 60–70 min approximately, it could slightly change depending on the how fast the participants repeated the stimuli.

that neural envelope tracking is stronger when listening to intelligible speech as compared to unintelligible signals in both theta (Ahissar et al., 2001; Doelling et al., 2014; Peelle et al., 2013) and delta ranges (Di Liberto et al., 2015; Ding and Simon, 2013; Doelling et al., 2014), though some failed to observe a direct link between neural envelope tracking strength and intelligibility (Hincapié-Casas et al., 2021; Howard and Poeppel, 2010; Pefkou et al., 2017; Zoefel and VanRullen, 2015c). Yet, as speech's intelligibility covaries with acoustical changes, it is unclear from these findings whether changes in neural envelope tracking reflect linguistic processing, or whether changes in acoustics alone can modulate neural envelope tracking (Ding et al., 2013; Kösem and van Wassenhove, 2017; Meng et al., 2021; Pinto et al., 2022; Glushko et al., 2022; Chalas et al., 2023).

In this current study, therefore, we directly dissociated the contribution of neural envelope tracking in the processing of speech acoustic cues from those related to linguistic processing. To achieve this, we examined the neural changes associated with the comprehension of Noise-Vocoded (NV) speech (Davis et al., 2005; Shannon et al., 1995). The intelligibility of a NV sentence is dependent on its amount of spectral degradation, as directly indexed by the number of frequency bands used in the noise-vocoding procedure (Davis et al., 2005). However, the intelligibility of initially unintelligible NV speech can be recovered by training (specifically via exposure to the original clear version of the sentence) (Dahan and Mead, 2010; Sohoglu and Davis, 2016). We recorded the cortical activity using magnetoencephalography (MEG) while participants listened to NV sentences in a 3-phase training paradigm: (1) pre-training, where the NV stimulus was barely comprehended, (2) training with exposure of the original clear version of speech stimulus, and (3) post-training, where the same stimulus was more intelligible after the training phase (Fig. 1). Using this paradigm, we tested if the neural responses to a speech signal were modulated by its intelligibility without changing its acoustic structure. To test the influence of spectral degradation on neural envelope tracking independently of training, participants were listening to two NV type of sentences (4-band and 2-band NV speech), but were trained to understand only 4-band NV speech.

2. Materials and methods

2.1. Participants

Thirty-two participants were recruited. The experimental procedure was approved by the local ethics committee (CMO region Arnhem-Nijmegen), and all participants gave informed consent in accordance with the Declaration of Helsinki. All participants were right-handed native Dutch speakers, and had no known history of neurological, language, or hearing problems. One participant was excluded because she was unable to finish the experiment; leaving thirty-one participants (15 females; mean ± SD, 23 ± 3.1 years) in the analysis.

2.2. Stimuli

We used the same NV speech stimuli as in previous behavioral and MEG studies (Dai et al., 2017, 2022). The original speech were selected from a corpus with daily conversational Dutch sentences, digitized at a 44,100 Hz sampling rate and recorded either by a native male or a native female speaker (Versfeld et al., 2000). The stimulus-set was created and validated as efficient measurement of the speech reception threshold (Versfeld et al., 2000), so that sentences from this set were equally intelligible in adverse listening conditions. Each sentence consisted of 5–8 words (e.g., 'Mijn handen en voeten zijn ijskoud', in English: 'My hands and feet are freezing'). Two semantically independent sentences recorded by the same speaker were combined into one stimulus, separated by a 300-ms silence gap (average duration = 4.2 s, min = 4.0 s, max = 4.5 s). In total, 160 stimuli were constructed, half of them were spoken by the male speaker and half by the female speaker. The two-sentence stimuli were then manipulated by noise-vocoding (Shannon et al., 1995) with Praat software (Version: 6.0.39 from http://www.praat.org), using either 4 or 2 frequency bands logarithmically spaced between 50 and 8000 Hz, resulting in 80 trials per noise vocoding condition. The same 2-band and 4-band NV stimuli were presented to all participants. As the 2-band and 4-band NV stimuli were generated with distinct spoken segments their temporal envelope was uncorrelated. The noise-vocoding technique degrades the spectral content of the acoustic signal (i.e., the fine structure) but keeps the temporal information (i.e., speech envelope) largely intact (Fig. S1 describes power and modulation spectra (Ding et al., 2017) of the speech materials). All stimuli were presented at ~70 dB SPL.

2.3. Procedure

The training used in this MEG experiment was similar to our previous studies (Dai et al., 2017, 2022), but combined with more testing trials. The experiment included three phases: pre-training, training, and post-training. In the pre-training and post-training phases, the participants were tested on their ability to understand the 4-band and 2-band vocoded speech stimuli. For each trial, participants heard a speech stimulus binaurally and were asked to repeat the sentences afterwards. Participants' responses were recorded by a digital microphone with a sampling rate of 44,100 Hz. In both pre-training and post-training phases, participants were exposed to the same 160 trials, with the order of

presentation fully randomized in each phase. In between pre-training and post-training, participants performed a training session to improve the intelligibility of the 4-band vocoded speech stimuli. For this, they were presented one time to the clear version of a trial, followed by the vocoded version of that trial; simultaneously, to enhance the training effect, they could read the written version of the trial on a computer screen. 2-band vocoded speech was not trained in this phase. The participant remained in the MEG during training session, which lasted for about 20 min. The experiment was implemented using Presentation software (Version 16.2, www.neurobs.com), and took about 70 min in total.

2.4. Behavioral analysis

The intelligibility of vocoded speech was measured by calculating the percentage of correct content words (excluding function words) in participants' reports for each trial. Words were regarded as correct if there was a perfect match (correct word without any tense errors, singular/plural form changes, or changes in sentential position). The percentage of correct content words was chosen as a more accurate measure of intelligibility based on acoustic cues than percentage correct of all words, considering that function words can be guessed based on the content words (Brouwer et al., 2012). A two-way repeated-measures ANOVA was performed with factors of NV band (trained 4-band and untrained 2-band) and Time (pre- and post-training). As the data violated the assumption of homogeneity of variance (Levene's statistic (absolute) 30.6, $p < 0.001$), we also performed non-parametric statistical testing on the interaction effect using Wilcoxon Signed-rank test statistic.

2.5. MEG measurement

MEG data were recorded with a 275-channel whole-head system (CTF Systems Inc., Port Coquitlam, Canada) at a sampling rate of 1200 Hz (with anti-aliasing low-pass filter at 300 Hz) in a magnetically shielded room. Data of four channels (MLC11, MLC32, MLF62, MRF66) were not recorded due to channel malfunctioning. Participants were seated in an upright position. Head location was measured with two coils in the ears (fixed to anatomical landmarks) and one on the nasion. To reduce head motion, a neck brace was used to stabilize the head. Head motion was monitored online throughout the experiment with a real-time head localizer and if necessary corrected between the experimental blocks. The speech signal was delivered through plastic air tubes connected to foam earpieces in the MEG scanner.

2.6. MEG data preprocessing

MEG Data analysis was conducted in MATLAB using the FieldTrip toolbox (fieldtrip-20,190,327) (Oostenveld et al., 2011) during pre-training and post-training sessions. Trials were defined as data between 500 ms before the onset of sound signal and 4000 ms thereafter. Three steps were taken to remove artifacts. Firstly, trials were rejected if the range and variance of the MEG signal differed, on visual inspection, by at least an order of magnitude from the other trials of the same participant. Secondly, independent component analysis (ICA) was performed. Data was decomposed into 270 independent components. Based on visual inspection of the ICA components' time courses and scalp topographies, components showing clear signature of eye blinks, eye movement, heartbeat and noise were identified and removed from the data. On average 7 (SD = 2) independent components were removed with this procedure. Data was back-projected to sensor space after removal of the bad ICA components. Visual inspection of trials was performed again after ICA component rejection, and trials were rejected based on the range and variance. In total, 16 trials (5% of total trials, SD = 8) were removed, resulting in an average of 304 included trials per participant (average number of trials in condition pre-training 4-band: 76 (SD = 3), post-training 4-band: 77 (SD = 2), pre-training 2-band: 75 (SD = 3), post-training 2-band: 76 (SD = 2)).

2.7. MEG analysis

Region of Interest: A data-driven approach was first performed to identify the reactive channels for sound processing. Event-related fields were computed between (-300, 400 ms) relative to sentence onset. For ERF analyses, epoched data was low-pass filtered at 35 Hz, and baseline-corrected using a baseline window (-300, 0 ms) relative to sentence onset. The M100 (within the time window between 80 and 120 ms after the first word were presented) response was measured on the data over all experimental conditions, after planar gradient transformation. We selected the 6 channels with the relatively strongest response at the group level on each hemisphere, and the averages of these channels were used for all subsequent analysis. The locations of the identified channels cover the classic auditory areas (Fig. 3A). Description of M100 responses per condition is provided in supplementary Fig. S2.

Speech-brain coherence: Magnitude-squared coherence between the broadband envelope of the speech signal (*env*) and MEG activity for each sensor (*brain*) for each frequency *f*, following the formula:

$$Coh_{speech-brain}(f) = \frac{\left|CSD_{speech-brain}(f)\right|^2}{ASD_{speech}(f)\ ASD_{brain}(f)}$$

Where $CSD_{speech-brain}$ represents the cross-spectral density between speech and brain signals, and ASD_{speech} ASD_{brain} the auto-spectral densities of speech and brain signals respectively. Broad-band speech envelopes were computed by band-pass filtering the acoustic waveforms (fourth-order Butterworth filter with [250–4000 Hz] cut-off frequencies), and by computing the absolute value of the Hilbert transform of the filtered signal. Cross- and auto- spectral density analysis of MEG signals was performed using discrete prolate spheroidal sequence (dpss) multi-tapers with $a \pm 1$ Hz smoothing window of the speech envelopes. Epochs were redefined for speech-brain coherence analysis: the first 500 ms of each epoch were removed to exclude the evoked response to the onset of the sentence. The speech-brain coherence was measured at different frequencies (1 to 30 Hz, 1 Hz step). Finally, the coherence data were projected into planar gradient representations. We repeated the same analysis described above to quantify the speech-brain coherence for each condition. For the investigation of our main hypotheses, we restricted the speech-brain coherence analyses to delta band (1–4 Hz) and theta band (4–8 Hz) activity and for this we averaged speech-brain coherence within the two frequency ranges of interest. These frequency bands were chosen based on the previous literature (Ding and Simon, 2014; Kösem and van Wassenhove, 2017). For supplementary analyses, we also explored speech-brain coherence within other definitions of the delta frequency range: (0.5–4 Hz), (0.5–1.5 Hz), and (2.5–3.5 Hz) (Fig. S5).

ROI analyses: The speech-brain coherence was averaged within the strongest 6 channels on each hemisphere. We tested the speech-brain coherence in the delta and theta range using a three-way repeated measure ANOVA with factors NV band (4-band, 2-band), Time (pre-training, post-training) and Hemisphere (left, right). Data verified the assumptions of the ANOVA, as homogeneity of variance between conditions was not violated (Levene's statistic (absolute), delta: 1.41, $p = 0.20$, theta: 0.65, $p = 0.71$) and residuals followed a normal distribution (Kolmogorov–Smirnov limiting form's statistic, delta: 0.69, $p = 0.71$, theta: 0.92, $p = 0.36$).

Whole sensor space analysis: We performed cluster-based permutation statistics across subjects (Oostenveld et al., 2011) to test whether we could observe a main effect of NV-band across sensors on speech-brain coherence *Coh* (by contrasting between $Coh_{4\text{-band}}$ and $Coh_{2\text{-band}}$, averaged across pre- and post- training sessions) and an interaction effect between NV-band and Time (by contrasting between ($Coh_{4\text{-band, post}}$ - $Coh_{4\text{-band, pre}}$) and ($Coh_{2\text{-band, post}}$ - $Coh_{2\text{-band, pre}}$) in both delta and theta frequency ranges. Pairwise t-tests were then computed for each sensor between the two conditions. Sensors with a p-value associated to the *t*-test of 5% or lower were selected as cluster candidates (a minimum

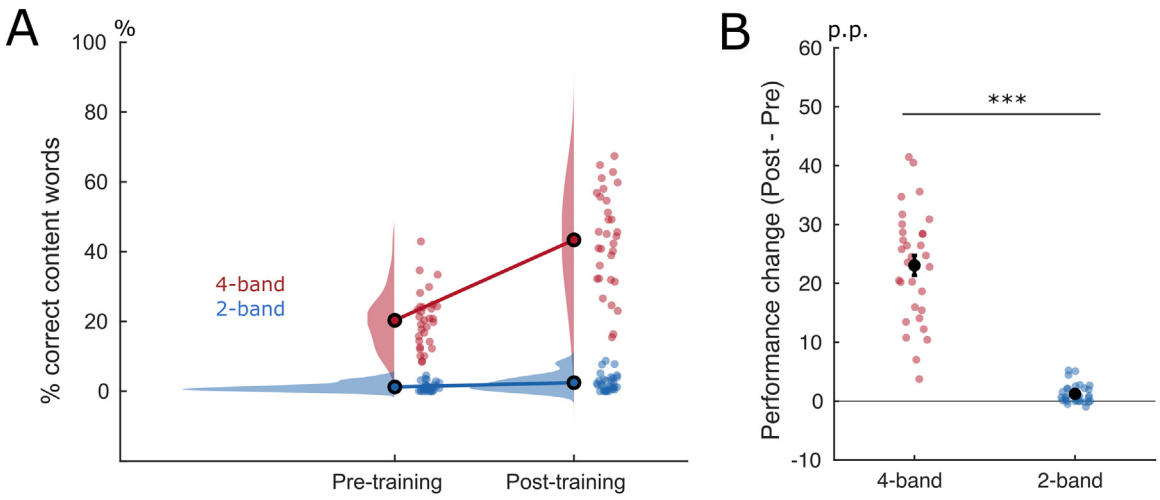

Fig. 2. Behavioral results. (A) Proportion of corrected reported content words pre- and post-training for both NV-speech conditions. (B) Performance change Post – Pre training. The intelligibility of trained 4-band (red) significantly improved by 23% on average with training, while untrained 2-band (blue) NV speech remained mostly unintelligible post-training. The open dots connected by lines (panel A) and the large black dots (panel B) indicate the grand average performance in each condition. The raincloud indicated the distributions of individual data, and each small dot corresponds to one participant.

of two significant adjacent sensors was required to form a cluster). The sum of the t-values within a cluster was used as the cluster-level statistic. The reference distribution for cluster-level statistics was computed by performing 1000 permutations between the two conditions. The contrast was considered significant if the probability of observing a cluster test statistic of that size in the reference distribution was 0.025 or lower (two-tailed test).

Source reconstruction analysis: Anatomical MRI scans were obtained after the MEG session using either a 1.5 T Siemens Magnetom Avanto system or a 3 T Siemens Skyra system for each participant (anatomical MRI was not recorded for two participants, their data were excluded for the source reconstruction analysis). The co-registration of MEG data with the individual anatomical MRI was performed via the realignment of the fiducial points (nasion, left and right pre-auricular points). Lead fields were constructed using a single shell head model based on the individual anatomical MRI. Each brain volume was divided into grid points of 1 cm voxel resolution, and warped to a template MNI brain. For each grid point the lead field matrix was calculated. The sources of the observed delta and theta speech-brain coherence were computed using beamforming analysis with the dynamic imaging of coherent sources (DICS) technique to the coherence data (Gross et al., 2001).

3. Results

3.1. Behavioral results

We compared the participants' comprehension of NV speech before and after training. Consistent with previous findings (Dai et al., 2017, 2022; Davis et al., 2005; Sohoglu and Davis, 2016), and as shown in Fig. 2, the training significantly improved the perception of 4-band NV speech. A two-way repeated-measure ANOVA showed that the main effects of noise vocoding (2-band vs. 4-band) and time (pre- vs. post-training) were significant (noise vocoding: ($F(1, 30) = 331.48$, $p < 0.001$, $\eta 2 = 0.92$; time: $F(1,30) = 183.46$, $p < 0.001$, $\eta 2 = 0.86$). Crucially, a significant interaction between noise vocoding and time was observed ($F(1,30) = 180.99$, $p < 0.001$, $\eta 2 = 0.86$), meaning that the intelligibility of 4-band NV speech was significantly improved compared to that of 2-band NV speech (4-band(post-pre) vs. 2-band(post-pre): Wilcoxon Signed rank test $Z = 4.89$, $p < 0.001$). After training, 4-band NV sentences had a score of 43.38 ± 2.53% recognition accuracy (23.06 ± 1.69% improvement during training; values here and below indicate mean ± SEM), while 2-band NV sentences remained mostly unintelligible with a score of 2.39 ± 0.43% recognition accuracy (1.58 ± 0.28% improvement during training).

3.2. MEG results

The behavioral results confirmed that intelligibility and spectral complexity could be dissociated in the present study. We then investigated how speech-brain coherence in auditory regions was impacted by the training session (Fig. 3). In line with previous studies (Meng et al., 2021; Peelle et al., 2013), we show that the neural envelope tracking of 4-band NV speech was stronger than that to 2-band NV speech (Fig. 3B–D). This was observed in the delta but not the theta frequency range (main effect of NV band, delta: $F(1, 30) = 25.95$, $p < .001$, $\eta 2 = 0.46$; theta: $F(1, 30) = 4.11$, $p = .052$, $\eta 2 = 0.12$, Fig. 3E–F).

However, the neural envelope tracking of NV speech was not significantly affected by training at delta frequencies (Fig. 3E, delta, main effect of time: ($F(1, 30) = .24$, $p = .63$, $\eta 2 = 0.008$; Fig. 3E and F). Theta neural envelope tracking overall reduced after training: ($F(1, 30) = 5.23$, $p = .030$, $\eta 2 = 0.15$)). But importantly, if neural envelope tracking reflected intelligibility, we specifically predicted that neural tracking to NV speech envelope would be stronger after training for the intelligible 4-band NV sentences. Yet, this was not observed (interaction between NV-band and Time, delta: $F(1, 30) = 4.16$, $p = .051$, $\eta 2 = 0.12$; theta: $F(1, 30) = .52$, $p = .48$, $\eta 2 = 0.02$). As such, the change in speech-brain coherence *post - pre* training was not significantly correlated with change in speech intelligibility (Fig. S2).

Looking at speech-brain coherence effects at each hemisphere (Fig. S3 A and B), theta neural tracking was significantly stronger in right than in left auditory ROIs ($F(1, 30) = 4.58$, $p = .041$, $\eta 2 = 0.13$), while delta neural envelope tracking was not significantly different across hemispheres ($F(1, 30) = 2.56$, $p = .12$, $\eta 2 = 0.08$). Effect of training and NV-band were not significantly different between the left and right ROIs (interaction between Hemisphere and NV-band, delta : $F(1, 30) = 0.01$, $p = .92$, $\eta 2 = 0.00$, theta: $F(1, 30) = 0.99$, $p = .37$, $\eta 2 = 0.03$; interaction between Hemisphere and Time, delta: : $F(1, 30) = 2.84$, $p = .10$, $\eta 2 = 0.09$, theta: $F(1, 30) = 0.08$, $p = .78$, $\eta 2 = 0.00$; interaction between Hemisphere, NV-band, and Time, delta: $F(1, 30) = 2.91$, $p = .10$, $\eta 2 = .09$, theta: $F(1, 30) = 1.81$, $p = .19$, $\eta 2 = .06$).

Further whole brain analysis showed a similar pattern of results (Figs. 4A, B and S4). Cluster-based permutation tests revealed a main effect of NV-band in the delta range (cluster $p < 0.001$), but not in the theta range. No significant interaction effects were observed.

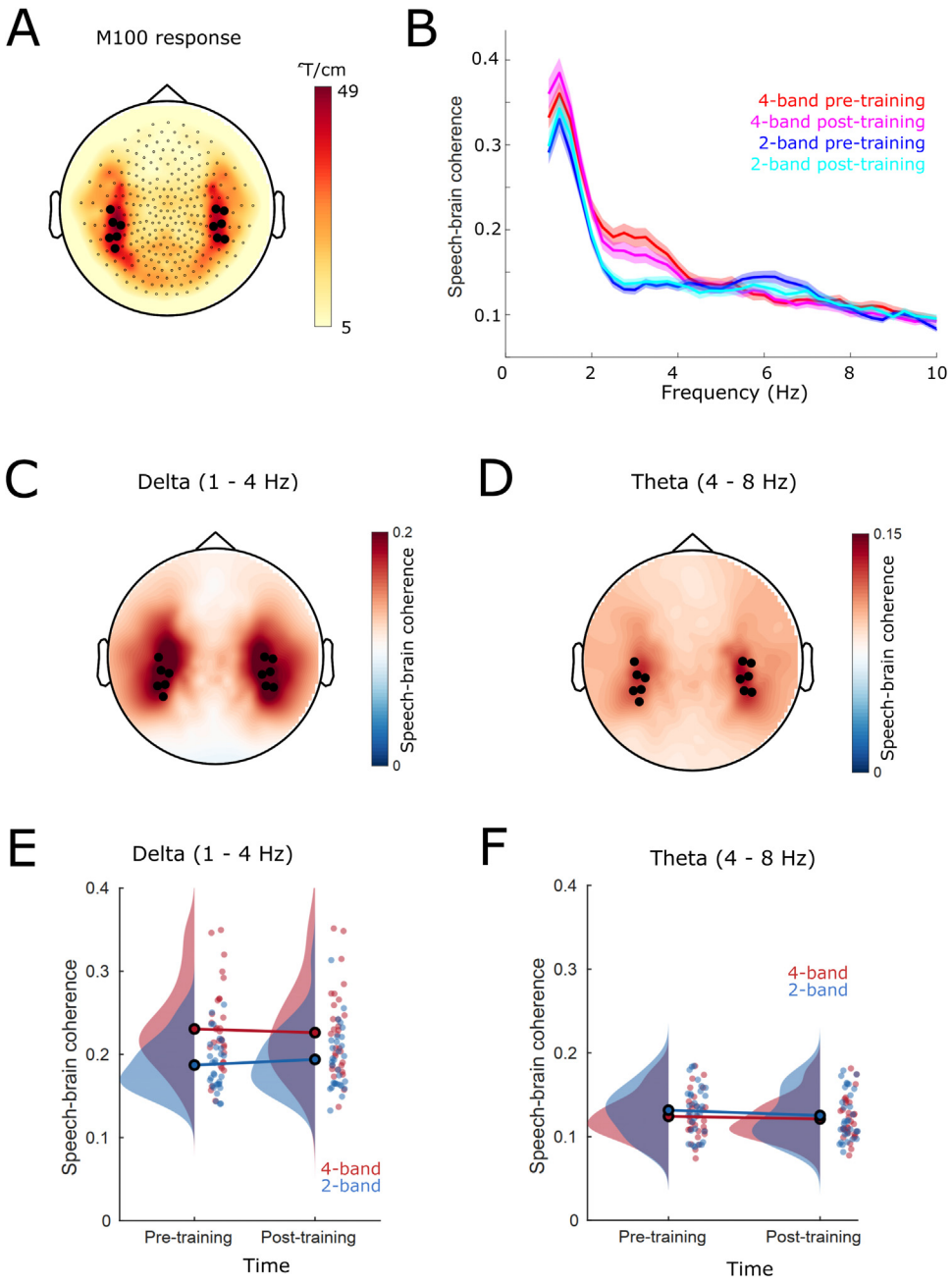

Fig. 3. Neural envelope tracking responses in auditory cortices as a function of intelligibility and acoustic spectral complexity. (A) Topography of the M100 response. The highlighted six channels showed the relatively strongest response at the group level on each hemisphere, and the average within these channels was used for all subsequent region-of-interest analysis. (B) Average speech-brain coherence between conditions across selected channels. Shaded areas denote standard error of the mean. (C) Topography of speech-brain coherence averaged across all conditions within the delta range (1–4 Hz). (D) Topography of average speech-brain coherence within the theta range (4–8 Hz). (E) between neural activity and 4-band NV speech (red) or 2-band NV speech (blue) in the delta (1–4 Hz) range averaged across selected channels. The open dots connected by lines indicate the grand average speech brain coherence in the pre- and post-training phases. The raincloud indicated the distribution of individual data, and each small dot corresponds to one participant. (F) Coherence between neural activity and 4-band NV speech (red) or 2-band NV speech (blue) in the theta (4–8 Hz) range averaged across selected channels.

We considered speech-brain coherence analysis in the delta frequency band within (1- 4 Hz) range, which was justified upon previous literature. Considering that the speech envelope contained distinct peak dynamics within this range, as well as strong power around 0.5 Hz (Fig. S1), we additionally performed exploratory speech-brain coherence analyses across different delta frequency ranges (Fig. S5). Widening the delta frequency range to (0.5–4 Hz) did not change the main patterns of results in ROI (Fig. S5A and B) and whole-brain results (Fig. S5E). There was no significant main effect of time ($F(1, 30) = 1.2$, $p = .28$, $\eta 2 = 0.04$), a main effect of NV-band ($F(1, 30) = 26.9$, $p < .001$, $\eta 2 = 0.47$), and no significant interaction between NV-band and Time, $F(1, 30) = 1.35$, $p = .25$, $\eta 2 = 0.04$). Restricting analyses to low-delta (0.5–1.5 Hz) (Fig. S3C), we observed, in addition to the significant NV-band effect band ($F(1, 30) = 18.4$, $p < .001$, $\eta 2 = 0.38$), a significant effect of time ($F(1, 30) = 7.1$, $p = .01$, $\eta 2 = 0.19$). This means that low-delta speech-brain coherence increased post-training compared to pre-training, irrespective of the NV speech condition. Important, there was no significant interaction effect in ROIs (NV-band * Time, $F(1, 30) = 1.14$, $p = .29$, $\eta 2 = 0.04$) and whole brain analyses (Fig. S5F). A second peak in speech envelope dynamics was observable around (2.5 – 3.5 Hz) (Fig. S1). Analyzing speech-brain coherence at this range, we did not observe a significant main effect of time ($F(1, 30) = 3.2$, $p = .08$, $\eta 2 = 0.10$), though the main effect of NV-band was observable ($F(1, 30) = 23.3$, $p < .001$, $\eta 2 = 0.44$). For this frequency range, we did observe a significant interaction effect ($F(1, 30) = 13.1$, $p = .001$, $\eta 2 = 0.30$) (Fig. S5D). However, importantly, this interaction is due to a reduction in neural-speech tracking for 4 band-NV speech after training compared to before training. This reduction in tracking strength post-training is in the opposite direction than we expected (neural tracking strength should increase with intelligibility, and here we observe a decrease). Furthermore, the interaction effect was not significant in whole-brain sensor analyses (Fig. S5G).

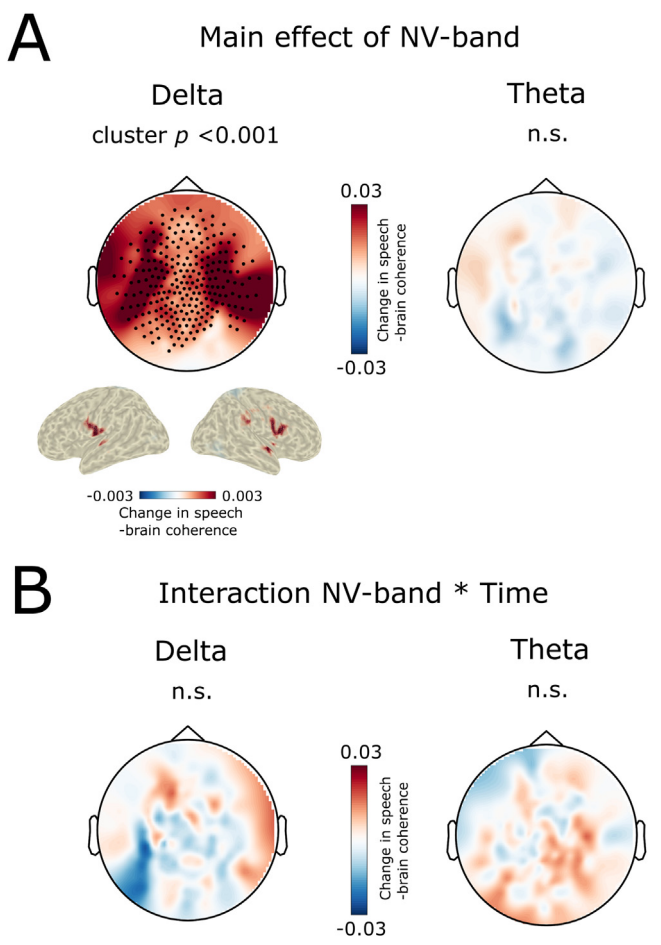

Fig. 4. Whole brain analysis. (A) Main effect of NV-band. Topographies and reconstructed source represent the contrast in neural tracking of 4-band vs 2-band NV speech envelope. Left panel: A main effect of NV-band is observed in the delta range (dots represent spatial topography of the significant cluster). Delta tracking is stronger for 4-band NV speech as compared to 2-band NV, irrespective of training. Right panel: No significant changes in neural tracking of speech envelope in the theta range irrespective of their spectral complexity. (B) Interaction effects NV-band * Time. No significant effect of training is observed, specifically the gain in intelligibility of the 4-band NV speech after training is not associated with a gain in neural envelope tracking.

Overall, the results suggest that neural envelope tracking is influenced by the acoustic structure of the speech signal (by its spectral degradation in particular), but we failed to find a positive correlation between strength in neural envelope tracking and speech intelligibility.

4. Discussion

In the present study, we tested the effect of intelligibility on the neural tracking of speech envelope. We used NV speech that could gain in intelligibility via training. With this manipulation we could dissociate gains in intelligibility linked to acoustic cues (spectral degradation), from those linked to linguistic processing of the speech signal. The training increased the intelligibility of NV speech but did not change its neural tracking response. In contrast, neural envelope tracking in the delta range was still modulated by the acoustic detail of the NV speech signal. These results are in line with previous reports showing that neural tracking of the speech envelope reduces with the amount of spectral degradation (Chen et al., 2022; Meng et al., 2021; Peelle et al., 2013), and others failing to find a correlation between the neural tracking of speech envelope in auditory cortex and speech intelligibility when acoustic details are controlled (Kösem et al., 2016; Millman et al., 2015; Peña and Melloni, 2012; Zoefel and VanRullen, 2015a; Baltzell et al., 2017). Therefore, the results suggest that brain-speech tracking in auditory areas reflects relevant neural mechanisms during the processing of speech acoustics, but does not unequivocally reflect the processing of more abstract linguistic information in speech.

This interpretation seems in apparent contradiction with other findings. Ding and colleagues (Ding et al., 2016) have found that neural oscillations in the delta range could track sentential and phrasal linguistic structures in speech in the absence of acoustic cues (although neural oscillatory peaks at constituent phrases could also partially reflect non-syntactic information (Kalenkovich et al., 2022), such as prosodic cues (Boucher et al., 2019; Glushko et al., 2022)). A recent study reanalyzing the data of (Millman et al., 2015) has found that delta tracking of speech is increased when the NV speech is intelligible as compared to when it is not understood (Di Liberto et al., 2018). In noisy environments, neural envelope tracking of the attended speech signal is stronger when the attended speech is fully understood (Dai et al., 2022; Keitel et al., 2018), or when the attended speech is in competition with unstructured speech (words were presented in random order) as compared to structured speech (speech with phrasal structure) (Har-Shai Yahav and Zion-Golumbic, 2021). The language proficiency of the listener also affects the neural envelope tracking of naturally spoken speech (Lizarazu et al., 2021).

One difference between these other studies and the present one concerns the intelligibility level of the stimuli. In the prior studies, intelligibility ratings were very high as compared to our design, where maximum intelligibility reached 40–60%. This means that our participants may have learned to extract some phonological and lexical cues from the speech, but may not have enough information to extract the full content of the sentences or be able to predict their linguistic structure. In contrast, in the prior studies, the intelligible stimuli were understood for the most part. Moreover, the syntactic structure of the stimuli was clearly predictable in some experimental designs: in Ding et al. (2016) sentences with similar phrasal and sentential structure were presented in blocks; in Di Liberto et al. (2018) the same sentence was repeated over and over. Sentence structure priming is known to increase the neural tracking of primed speech, this without correlating with intelligibility (Baltzell et al., 2017). Therefore, delta tracking may reflect the processing of intelligible and predictable linguistic information (as in the prior studies), but may not do so (as in the current study) when the speech signal is too noisy, does not have a predictable syntactic structure, and/or is not fully intelligible.

It is also important to point out that, in the prior studies mentioned above, the effect of intelligibility on brain-speech tracking seemed to be restricted to delta dynamics (< 4 Hz) and was less clearly observable for theta dynamics. These data supports the predominant role in delta tracking in the processing of linguistic structure, while theta tracking may affect the processing of acoustic and phonological information (Kösem and van Wassenhove, 2017). Still, we did not find an effect of intelligibility in delta dynamics, and we show that spectral degradation differently affected delta and theta neural tracking of speech envelope. The increased spectral degradation of speech was associated with decreased delta tracking in auditory areas, while theta tracking remained unaffected by the amount of noise vocoding. These results suggest that theta dynamics may primarily track broadband envelope temporal information (that is unaffected by the amount of vocoding), while neural tracking of speech envelope in the delta range may be impacted by the spectral complexity of the speech signal (Ding et al., 2013; Meng et al., 2021).

The current experimental design, while allowing us to change intelligibility levels for the same acoustic signal, presents limitations. It could be argued that the participants primarily relied on memory to perform the task: participants may have recognized the stimuli in the post-training phase and not listened to the stimuli anymore because they have memorized it. Therefore, neural data would not reflect speech processing but memory effects. We argue that the memory hypothesis can

unlikely account for the present results. A total of 160 sentential stimuli were presented to the participants, including 80 trials for the trained 4-band NV condition. The trials are composed of two semantically unrelated sentences of 5–8 words each, therefore a trial was 13 words on average. The task given to the participant in pre- and post-training sessions was to exactly repeat the trials. In each pre- and post- training sessions, the presentation of the trials was fully randomized and unpredictable. In this situation, it is unlikely that the participants relied on memory and stopped paying attention to the acoustic stimuli. Furthermore, if participants stopped paying attention to the 4-band NV condition after training, we would have then expected a severe drop in speech-brain coherence in the 4-band condition as we know that the tracking response is highly dependent on attention (Zion Golumbic et al., 2013), but this is not what we observed.

We have focused our investigation on the tracking of the acoustic temporal envelope, as this has been proposed to reflect relevant mechanisms involved in speech processing (Giraud and Poeppel, 2012; Peelle and Davis, 2012). We do not claim that neural tracking cannot reflet linguistic processing, as previous studies reported that neural tracking can track semantic and syntactic structures (Brodbeck et al., 2018; Ding et al., 2016; Verschueren et al., 2022). Additionally, while we failed to find significant effects outside auditory cortex, our study does not exclude that other brain areas could track linguistic structures in speech. Frontal motor and parietal regions in particular have previously been shown to be influenced by linguistic content, and to top-down modulate neural tracking in auditory cortex (Chalas et al., 2022; Hincapié-Casas et al., 2021; Keitel et al., 2018; Park et al., 2015).

In conclusion, we failed to find a direct effect of intelligibility on the neural tracking of speech envelope in both theta and delta ranges in auditory cortices. Significant changes in neural tracking were still observed in the delta range in relation to the acoustic degradation of speech. These findings suggest that acoustics greatly influence the neural tracking of speech envelope. They also suggest that caution is required when choosing the control condition for analyses of tracking responses because, as we have shown, a slight change in acoustic parameters can have strong effects on the neural tracking response. Finally, they suggest that neural envelope tracking is not necessarily modulated by the intelligibility of the speech signal.

Data and code availability statement

Stimuli, data, and scripts are available upon request from the Donders Repository (https://doi.org/10.34973/qksk-6x25), a data archive hosted by the Donders Institute for Brain, Cognition and Behaviour.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Credit authorship contribution statement

Anne Kösem: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing, Data curation. **Bohan Dai:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – original draft. **James M. McQueen:** Conceptualization, Methodology, Writing – review & editing. **Peter Hagoort:** Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition.

Data availability

Data will be made available on request.

Acknowledgments

This research was supported by a Spinoza award to P.H. and by a Marie Skłodowska-Curie Individual Fellowship (Grant No. 843088) and by an ANR Grant No. (ANR-21-CE37–0003) to A.K.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2023.120040.

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 (23), 13367–13372. doi:10.1073/pnas.201400998.

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 (6), 3144–3151.

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 (8), 1205–1215. doi:10.1162/jocn_a_01410.

Broderick, M.P., Anderson, A.J., Di Liberto, G.M., Crosse, M.J., Lalor, E.C., 2018. Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. Curr. Biol. 28 (5), 803–809. doi:10.1016/J.CUB.2018.01.080.

Brouwer, S., Van Engen, K.J., Calandruccio, L., Bradlow, A.R., 2012. Linguistic contributions to speech-on-speech masking for native and non-native listeners: language familiarity and semantic content. J. Acoust. Soc. Am. 131 (2), 1449–1464.

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

Chalas, N., Daube, C., Kluger, D.S., Abbasi, O., Nitsch, R., Gross, J., 2023. Speech onsets and sustainedspeech contribute differentially to delta and theta speech tracking in auditory cortex. CerebCortex doi:10.1093/cercor/bhac502.

Chen, Y., Schmidt, F., Keitel, A., Rosch, S., Hauswald, A., Weisz, N., 2022. Speech intelligibility changes the temporal evolution of neural speech tracking. Neuroimage doi:10.1016/j.neuroimage.2023.119894.

Dahan, D., Mead, R.L., 2010. Context-conditioned generalization in adaptation to distorted speech. J. Exp. Psychol. Hum. Percept. Perform. 36 (3), 704–728. doi:10.1037/a0017449.

Dai, B., McQueen, J.M., Hagoort, P., Kösem, A., 2017. Pure linguistic interference during comprehension of competing speech signals. J. Acoust. Soc. Am. 141 (3), EL249–EL254. doi:10.1121/1.4977590.

Dai, B., McQueen, J.M., Terporten, R., Hagoort, P., Kösem, A., 2022. Distracting linguistic information impairs neural tracking of attended speech. Curr. Res. Neurobiol. 3, 100043.

Davis, M.H., Johnsrude, I.S., Hervais-Adelman, A., Taylor, K., McGettigan, C., 2005. Lexical information drives perceptual learning of distorted speech: evidence from the comprehension of noise-vocoded sentences. J. Exp. Psychol. Gen. 134 (2), 222–241. doi:10.1037/0096-3445.134.2.222.

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 (19), 2457–2465. doi:10.1016/j.cub.2015.08.030.

Di Liberto, G.M., Lalor, E.C., Millman, R.E., 2018. Causal cortical dynamics of a predictive enhancement of speech intelligibility. Neuroimage 166, 247–258. doi:10.1016/J.NEUROIMAGE.2017.10.066.

Ding, N., Chatterjee, M., Simon, J.Z., 2013. Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. Neuroimage 88C, 41–46. doi:10.1016/j.neuroimage.2013.10.054.

Ding, N., Melloni, L., Zhang, H., Tian, X., Poeppel, D., 2016. Cortical tracking of hierarchical linguistic structures in connected speech. Nat. Neurosci. 19 (1), 158. doi:10.1038/nn.4186.

Ding, N., Patel, A.D., Chen, L., Butler, H., Luo, C., Poeppel, D., 2017. Temporal modulations in speech and music. Neurosci. Biobehav. Rev. 81, 181–187. doi:10.1016/J.NEUBIOREV.2017.02.011.

Ding, N., Simon, J.Z., 2013. Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. J. Neurosci. Off. J. Soc. Neurosci. 33 (13), 5728–5735. doi:10.1523/JNEUROSCI.5297-12.2013.

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.

Doelling, K.B., Assaneo, M.F., 2021. Neural oscillations are a start toward understanding brain activity rather than the end. PLOS Biol. 19 (5), e3001234. doi:10.1371/journal.pbio.3001234.

Giraud, A.L., Poeppel, D., 2012. Cortical oscillations and speech processing: emerging computational principles and operations. Nat. Neurosci. 15 (4), 511–517. doi:10.1038/nn.3063.

Glushko, A., Poeppel, D., Steinhauer, K., 2022. Overt and implicit prosody contribute to neurophysiological responses previously attributed to grammatical processing. Sci. Rep. 12, 14759. doi:10.1038/s41598-022-18162-3.

Greenberg, S., Carvey, H., Hitchcock, L., Chang, S., 2003. Temporal properties of spontaneous speech—a syllable-centric perspective. J. Phon. 31 (3–4), 465–485. doi:10.1016/j.wocn.2003.09.005.

Gross, J., Kujala, J., Hämäläinen, M., Timmermann, L., Schnitzler, A., Salmelin, R., 2001. Dynamic imaging of coherent sources: studying neural interactions in the human brain. Proc. Natl. Acad. Sci. 98 (2), 694–699.

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 (12), e1001752. doi:10.1371/journal.pbio.1001752.

Har-shai, P., Zion Golumbic, E., 2021. Linguistic processing of task-irrelevant speech at a cocktail party. Elife 10, e65096.

Hincapié-Casas, A.S., Lajnef, T., Pascarella, A., Guiraud-Vinatea, H., Laaksonen, H., Bayle, D., Boulenger, V., 2021. Neural oscillations track natural but not artificial fast speech: novel insights from speech-brain coupling using MEG. Neuroimage 244, 118577.

Howard, M.F., Poeppel, D., 2010. Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. J. Neurophysiol. 104 (5), 2500–2511. doi:10.1152/jn.00251.2010.

Kalenkovich, E., Shestakova, A., Kazanina, N., 2022. Frequency tagging of syntactic structure or lexical properties; a registered MEG study. Cortex 146, 24–38. doi:10.1016/j.cortex.2021.09.012.

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. Off. J. Soc. Neurosci. 35 (44), 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 (3), e2004473. doi:10.1371/journal.pbio.2004473.

Kösem, A., Basirat, A., Azizi, L., van Wassenhove, V., 2016. High-frequency neural activity predicts word parsing in ambiguous speech streams. J. Neurophysiol. 116 (6), 2497–2512. doi:10.1152/jn.00074.2016.

Kösem, A., van Wassenhove, V., 2017. Distinct contributions of low- and high-frequency neural oscillations to speech comprehension. Lang. Cogn. Neurosci. 1–9. doi:10.1080/23273798.2016.1238495.

Lakatos, P., Gross, J., Thut, G., 2019. A new unifying account of the roles of neuronal entrainment. Curr. Biol. 29 (18), R890–R905. doi:10.1016/J.CUB.2019.07.075.

Lizarazu, M., Carreiras, M., Bourguignon, M., Zarraga, A., Molinaro, N., 2021. Language proficiency entails tuning cortical activity to second language speech. Cereb. Cortex 31 (8), 3820–3831. doi:10.1093/cercor/bhab051.

Luo, H., Poeppel, D., 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54 (6), 1001–1010. doi:10.1016/j.neuron.2007.06.004.

Meng, Q., Hegner, Y.L., Giblin, I., McMahon, C., Johnson, B.W., 2021. Lateralized cerebral processing of abstract linguistic structure in clear and degraded speech. Cereb. Cortex 31 (1), 591–602. doi:10.1093/cercor/bhaa245.

Millman, R.E., Johnson, S.R., Prendergast, G., 2015. The role of phase-locking to the temporal envelope of speech in auditory perception and speech intelligibility. J. Cogn. Neurosci. 27 (3), 533–545. doi:10.1162/jocn_a_00719.

Molinaro, N., Lizarazu, M., 2018. Delta(but not theta)-band cortical entrainment involves speech-specific processing. Eur. J. Neurosci. 48 (7), 2642–2650. doi:10.1111/ejn.13811.

Obleser, J., Kayser, C., 2019. Neural entrainment and attentional selection in the listening brain. Trends Cogn. Sci. 23 (11), 913–926. doi:10.1016/J.TICS.2019.08.004.

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, 1–9. doi:10.1155/2011/156869.

Park, H., Ince, R.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 (12), 1649–1653.

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.

Peelle, J.E., Gross, J., Davis, M.H., 2013. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cereb. Cortex 23 (6), 1378–1387. doi:10.1093/cercor/bhs118.

Peña, M., Melloni, L., 2012. Brain oscillations during spoken sentence processing. J. Cogn. Neurosci. 24 (5), 1149–1164. doi:10.1162/jocn_a_00144.

Pinto, D., Prior, A., Zion Golumbic, E., 2022. Assessing the sensitivity of EEG-based frequency-tagging as a metric for statistical learning. Neurobiol. Lang. 1–21. doi:10.1162/nol_a_00061.

Rimmele, J.M., Zion Golumbic, E., Schröger, E., Poeppel, D., 2015. The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. Cortex doi:10.1016/j.cortex.2014.12.014.

Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. Science 270 (5234), 303–304. doi:10.1126/science.270.5234.303.

Sohoglu, E., Davis, M.H., 2016. Perceptual learning of degraded speech by minimizing prediction error. Proc. Natl. Acad. Sci. U.S.A. 113 (12), E1747–E1756. doi:10.1073/pnas.1523266113.

Verschueren, E., Gillis, M., Decruy, L., Vanthornhout, J., Francart, T., 2022. Speech understanding oppositely affects acoustic and linguistic neural tracking in a speech rate manipulation paradigm. J. Neurosci. 42 (39), 7442–7453.

Versfeld, N.J., Daalder, L., Festen, J.M., Houtgast, T., 2000. Method for the selection of sentence materials for efficient measurement of the speech reception threshold. J. Acoust. Soc. Am. 107 (3), 1671–1684. doi:10.1121/1.428451.

Zion Golumbic, E.M., Ding, N., Bickel, S., Lakatos, P., Schevon, C.A., McKhann, G.M., Schroeder, C.E., 2013. Mechanisms underlying selective neuronal tracking of attended speech at a “cocktail party. Neuron 77 (5), 980–991. doi:10.1016/j.neuron.2012.12.037.

Zoefel, B., VanRullen, R., 2015a. EEG oscillations entrain their phase to high-level features of speech sound. Neuroimage 124, 16–23. doi:10.1016/j.neuroimage.2015.08.054.

Zoefel, B., VanRullen, R., 2015b. Selective perceptual phase entrainment to speech rhythm in the absence of spectral energy fluctuations. J. Neurosci. Off. J. Soc. Neurosci. 35 (5), 1954–1964. doi:10.1523/JNEUROSCI.3484-14.2015.

Zoefel, B., VanRullen, R., 2015c. The role of high-level processes for oscillatory phase entrainment to speech sound. Front. Hum. Neurosci. 9, 651. doi:10.3389/fnhum.2015.00651.