Robust neuronal oscillatory entrainment to speech displays individual variation in lateralisation

Nietzsche H. L. Lam, Annika Hultén, Peter Hagoort & Jan-Mathijs Schoffelen

Published online: 07 Feb 2018.

Submit your article to this journal

Article views: 665

View Crossmark data

Routledge
Taylor & Francis Group

REGULAR ARTICLE

OPEN ACCESS
Check for updates

Robust neuronal oscillatory entrainment to speech displays individual variation in lateralisation

Nietzsche H. L. Lam[a,b], Annika Hultén[a,b,c], Peter Hagoort[a,b] and Jan-Mathijs Schoffelen [a,b]

[a]Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands; [b]Radboud University, Donders Institute for Brain, Cognition and Behaviour, Nijmegen, Netherlands; [c]Department of Neuroscience and Biomedical Engineering, Aalto University, Helsinki, Finland

ABSTRACT

Neural oscillations may be instrumental for the tracking and segmentation of continuous speech. Earlier work has suggested that delta, theta and gamma oscillations entrain to the speech rhythm. We used magnetoencephalography and a large sample of 102 participants to investigate oscillatory entrainment to speech, and observed robust entrainment of delta and theta activity, and weak group-level gamma entrainment. We show that the peak frequency and the hemispheric lateralisation of the entrainment are subject to considerable individual variability. The first finding may support the involvement of intrinsic oscillations in entrainment, and the second finding suggests that there is no systematic default right-hemispheric bias for processing acoustic signals on a slow time scale. Although low frequency entrainment to speech is a robust phenomenon, the characteristics of entrainment vary across individuals, and this variation is important for understanding the underlying neural mechanisms of entrainment, as well as its functional significance.

Introduction

Human speech represents one of the most complex auditory signals that are perceived, containing information at multiple temporal scales that needs to be processed and integrated for adequate comprehension. Yet, we listen to speech with great ease. How does the brain keep up with this task? Focusing on the early stages of auditory processing, the input needs to be parsed into relevant temporal segments, which can then be further processed by the brain system for language, and integrated into a meaningful linguistic context. Ultimately, this cascade of processing operations results in comprehension. A popular perspective on early auditory processing of speech is that neuronal oscillations play an important mechanistic role in the processing and prediction of temporally structured perceptual information.

Neural oscillations, as picked up for instance in the magnetoencephalogram (MEG) reflect cyclic fluctuations in the excitability of neuronal populations, and certain phases within each cycle are considered optimal for processing input from the environment (Buzsáki & Draguhn, 2004; Schroeder & Lakatos, 2009; Schroeder, Wilson, Radman, Scharfman, & Lakatos, 2010; Van Rullen & Koch, 2003). Upon presentation of a periodic external signal, alignment of neural oscillatory activity to the

signals' rhythm allows for these periodic occurrences of high levels of neuronal excitability to align with certain periodic events in the signal. This synchronisation of rhythms is referred to as entrainment, and facilitates optimal sampling in discrete time windows. It allows the information in the signal to be divided into meaningful chunks, which can then be processed and understood. Subsequently, this information may be used to predict the upcoming signal, whereby higher order regions may provide top-down feedback to modulate the entrainment of oscillations to the signal (Park, Ince, Thut, Gross, & Schyns, 2015). Direct evidence for these entrainment mechanisms has been reported in both monkeys and humans. In these studies entrainment of neural oscillations with rhythmic auditory or visual stimuli have been shown to shape perception (Busch, Dubois, & VanRullen, 2009; de Graaf et al., 2013; Lakatos et al., 2005; Lakatos, Chen, O'Connell, Mills, & Schroeder, 2007; Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008; Romei, Gross, & Thut, 2010; Spaak, de Lange, & Jensen, 2014).

Evidence for an instrumental role of oscillatory entrainment to speech processing is more tentative. This is partially because speech is complex, and difficult to manipulate in comparison to the basic stimuli in the

This article was originally published with errors. This version has been corrected. Please see Corrigendum (http://dx.doi.org/10.1080/23273798.2018.1441682).

aforementioned studies (e.g. circular sine-wave gratings presented at a specific frequency, flashes of light, or Gaussian noise bursts). Another reason is that speech is quasi-periodic (i.e. less regular than strictly periodic signals), which challenges the exact temporal predictability of novel input, and thus the interpretation of the neural response to the speech signal in terms of rhythmic entrainment. To rise to the challenge of understanding speech perception, several theories have been proposed on the role of neural oscillations in speech perception (e.g. Ghitza, 2011; Ghitza & Greenberg, 2009; Giraud & Poeppel, 2012; Howard & Poeppel, 2010; Peelle, 2012; Poeppel, 2003; Shamma, Elhilali, & Micheyl, 2011). For instance, Giraud and Poeppel (2012) proposed that the spike train input to auditory cortex (which captures the energy fluctuations of the speech signal) influences neuronal excitability: the neurons in primary auditory cortex would adjust (reset) the phase of their excitability rhythm thereby allowing them to entrain to the rhythmic regularities of the speech signal.

As the linguistic information in speech occurs at different rates in a quasi-regular manner, these theories assume that neural oscillations at frequencies that roughly correspond to these rates are suited for parsing and decoding speech. On average, prosody occurs at a rate of about 1–3 Hz, syllables at about 4–7 Hz, and phonemes at about 30–50 Hz. Accordingly, slow oscillations, in the delta and theta frequency range, sample speech at the prosodic and syllabic rate, respectively. Fast oscillations in the gamma band (around 30 Hz and beyond) facilitate the sampling of phonemic information. In addition, it has been suggested that sampling rates are thought to be hierarchically embedded, with theta as the dominant sampling rhythm, and with coordinated sampling at the delta and gamma frequencies (Giraud & Poeppel, 2012; Gross et al., 2013). A related issue to sampling speech at multiple time scales is whether it is supported by a division of labour between the hemispheres as proposed in the Asymmetric Sampling Time (AST) model by Poeppel (2003) and further elaborated by Giraud and Poeppel (2012). In this model, the left auditory cortex is biased towards sampling signals at fast time scales, while the right auditory cortex is biased to sample at slower time scales (for support see Boemio, Fromm, Braun, & Poeppel, 2005; Giraud et al., 2007; Morillon et al., 2010; Shtyrov, Kujala, Palva, Ilmoniemi, & Näätänen, 2000; for a counter argument see Mcgettigan & Scott, 2012).

Empirical studies have provided support for the involvement of the delta and theta bands in entrainment (e.g. Abrams, Nicol, Zecker, & Kraus, 2008; Bourguignon et al., 2013; Cogan & Poeppel, 2011; Gross et al., 2013; Luo & Poeppel, 2007; Luo, Liu, & Poeppel, 2010; Molinaro, Barraza, & Carreiras, 2013; Peelle, Gross, & Davis, 2013), and mainly theta entrainment has been associated with speech intelligibility (Ahissar et al., 2001; Luo & Poeppel, 2007; Peelle et al., 2013). To date, only one study has demonstrated involvement of the gamma band (Gross et al., 2013), reporting theta to gamma phase-amplitude coupling to the speech envelope, alongside with phase-phase coupling of delta and theta oscillations to speech.

The conclusions from these earlier empirical studies have strong implications for theories on speech perception mechanisms. To corroborate earlier findings, we sought to confirm previous findings of entrainment. We used data from an unprecedented sample of 102 participants to address neural oscillatory entrainment to the speech envelope. We measured brain activity with magnetoencephalography (MEG) while participants listened to sentences and quantified the relationship between neuronal oscillations and the envelope of the speech signal using coherence.

In addition to corroborating previous work, a second goal of our study was to seize the opportunity to quantify individual differences in neural oscillatory entrainment of speech. We found support for entrainment at low but not high frequencies. Moreover, we were able to quantify and observe a rather large individual variability in entrainment, both in terms of strength and in terms of lateralisation.

Methods

Participants

A total of 102 native Dutch speakers (51 males), with an age range of 18–33 years (mean of 22 years), participated in the experiment. These participants formed part of the MOUS study (Mother of all Unification Studies; N = 204), and all participated in an fMRI and a MEG session. 102 participants completed both sessions where they read the stimuli, and the other half listened to recordings of the stimuli. The current study pertains to participants from the MEG session in the auditory modality. A more in depth description of the experimental protocol has been provided elsewhere (Lam, Schoffelen, Udden, Hulten, & Hagoort, 2016; Schoffelen et al., 2017). All participants were right-handed, reported normal hearing, had normal or corrected-to-normal vision, and had no known history of neurological, developmental or language deficits. The study was approved by the local ethics committee (CMO – the local "Committee on Research Involving Human Participants" in the Arnhem-Nijmegen region) and followed the guidelines of the Helsinki declaration.

Language stimuli

The full stimulus set consisted of 360 custom-made sentences, adjusted from newspaper clippings, book fragments, and social media content (for example: "Bij de opening van de nieuwe sporthal kregen de talrijke bezoekers een consumptie." (literal translation: At the opening of the new sports hall received the many visitors a drink)), and their word list counterparts (constructed by random shuffling of the words within the sentences). The stimuli varied in length between 9 and 15 words, of which half contained an embedded clause and half did not. The stimulus material was recorded by a native female Dutch speaker in a sound-proof recording booth. The speaker read the stimuli in a natural manner: the sentences were read at a regular pace with an average duration of 4.2 s (min: 2.8 s, max: 6.0 s), and the word lists were read with a brief pause between words, averaging 7.7 s (min: 5.5 s, max: 11.1 s). Subsequently, all stimuli were equalised to the same amplitude, and an onset and offset ramp of 10 ms was applied. In the current study, we only analysed the sentence trials, in order to be able to compare our results to the literature that use more naturalistic stimuli. Word list trials had temporal gaps between each of the words, whereas the sentence stimuli consisted of co-articulated speech. As a consequence, there were salient low-level differences in the temporal properties and rhythmic structure of the stimulus envelope signals, which precludes a fair interpretation of observed differences in entrainment across conditions.

Task and procedure

Experimental design

Using a Latin square design we created 6 sets of 240 stimuli, of which 120 consisted of proper sentences, and 120 were word lists. Participants were exposed to one of the 6 sets. Participants assigned the same set had sentences presented in a different (randomised) order. In the experiment, the stimuli were presented in a mini block design, and alternated between a sentence block (containing 5 sentences) and a word list block (containing 5 word lists, that were constructed). The first mini block (sentences or word lists) was randomised across participants.

At the beginning of each block, the block type was announced for 1500 ms: *zinnen* (sentences) or *woorden* (words), followed by a 2000 ms blank screen. At the beginning of each trial a fixation cross was presented for a jittered duration between 1200–2200 ms. Subsequently, the speech signal was presented for each trial (sentence or word list), and the fixation cross remained on the screen until the auditory signal was completed. Within each block, the inter-trial interval was a blank screen with a jittered duration between 1200–2200 ms.

In order to check for compliance, 10% of the trials were randomly followed by a yes/no question about the content of the previous sentence/word list. Half of the questions on the sentences addressed sentence comprehension (e.g. *Did grandma eat a pancake?).* The other half of the sentences, and the questions following the word lists addressed a content word (e.g. *Was a music instrument named?*). Participants answered the question by pressing a button for "Yes"/ "No" with their left index and middle fingers, respectively. For both question types, half of the trials had a yes-response as the correct answer.

All stimuli were presented using Presentation software Version 16.0, Neurobehavioral Systems, Inc). Speech stimuli were presented binaurally via MEG-compatible tubes. The questions were presented in black mono-spaced font, on a grey background. To reduce eye movements during listening, subjects were instructed to focus on a fixation cross. These visual stimuli were presented with an LCD projector (with a vertical refresh rate of 60 Hz) situated outside the MEG, and projected via mirrors onto the centre of the screen inside the MEG room, within a visual angle of 4 degrees.

Prior to performing the sentence listening task, we adjusted the hearing level for each subject. To ensure a sufficient cortical auditory response, the minimal auditory threshold was determined, and subsequently all auditory stimuli were presented at 50 dB above the minimum threshold. For task familiarisation purposes participants completed a practice task (using a separate set of stimuli from the actual task).

MEG data acquisition

MEG data were collected with a 275 axial gradiometer system (CTF). The signals were digitised at a sampling frequency of 1200 Hz (the cutoff frequency of the analogue anti-aliasing low pass filter was 300 Hz). Three coils were placed on anatomical landmarks of the participant's head (nasion, orifice of left and right ear canals) to determine the position of the head relative to the MEG-sensors. Throughout the measurement the head position was continuously monitored using custom software (Stolk, Todorovic, Schoffelen, & Oostenveld, 2013). During breaks the participant was allowed to reposition if needed. Participants were able to maintain a head position within 5 mm of their original position. Three bipolar Ag/AgCl electrode pairs were used to measure the horizontal and vertical electro-oculogram, and the electro-cardiogram.

Data processing

All analyses were done with custom written Matlab scripts and FieldTrip (Oostenveld, Fries, Maris, & Schoffelen, 2011).

Artifact detection

Physiological artifacts (eye movements and muscle contractions) and superconducting quantum interference device (SQUID) jumps were identified using a semi-automatic artifact identification procedure (http://www.fieldtriptoolbox.org/tutorial/automatic_artifact_rejection), followed by visual inspection. Data segments that contained artifacts were not subjected to further analysis. Across subjects, an average of 80% (standard deviation 10%) of the data was retained after rejection.

Preprocessing

The envelope of each speech signal was constructed as the sum of the Hilbert envelopes of 10 distinct bandpass filtered frequency bands of the original auditory signal (as per Gross et al., 2013). Subsequently, each envelope signal was downsampled to 1200 Hz, and temporally aligned to the corresponding MEG data. The MEG signal was initially epoched into the individual sentences, demeaned, and the power line interference was removed using a band stop filter (finite impulse response window sinc filter) between 49 and 51 Hz. Subsequently, the MEG signal and speech signal were downsampled to 300 Hz, and cut into 2 s long epochs with a 50% overlap, allowing for spectral estimates at a frequency resolution of 0.5 Hz. For the sensor level analysis, to facilitate the combination of MEG topographies across subjects, the data was transformed to a synthetic horizontal and vertical planar gradient representation using interpolation.

Sensor level coherence analysis and peak frequency selection

To quantify entrainment of the MEG signal to the speech envelope we calculated coherence, using multitaper spectral analysis, with a smoothing parameter of ±2 Hz. In addition, to quantify phase-to-amplitude cross-frequency entrainment, we computed the amplitude envelope of the gamma frequency band (30–50 Hz) in the MEG signals, using a bandpass filter followed by a Hilbert transform, and used these envelope signals for a second coherence analysis with the speech envelope.

Visual inspection of the sensor level data indicated that there was considerable variability across participants for the frequencies at which coherence peaked, see Figure 1. For further analysis, we defined individual peak frequencies for each participant at the sensor level. These individually selected peaks were subsequently used for source level analysis and for further evaluation.

For each participant, a peak detection algorithm was used to identify, for each channel, the frequency bins that showed a distinct peak in the coherence spectrum with a coherence value larger than 0.02. This yielded a binary vector (as a function of frequency), with a 1 indicating a peak, for each channel. Next, these binary vectors were summed across channels yielding a spectrum of peak counts across channels. The assumption here was the higher the peak count, the more reliable the peak. This vector was then multiplied to a second vector containing the standardised coherence across frequencies. This accounted for variance in coherence strength. The weighted vector was then smoothed (boxcar of 2 samples) and a second peak detection with a threshold of 2 was performed to identify the peak frequencies across the sensor array. Comparison of the estimated peak frequency with visual identification of the peak frequency for 10 subjects determined this peak detection process to be adequate.

Processing of the anatomical MRI and digitised headshape for MEG source reconstruction

We coregistered the anatomical MRI to the MEG-sensors, by aligning the scalp surface obtained from the MRI image with a digitised head shape, consisting of approximately 500 points across the scalp. The latter was obtained with a Polhemus device (Fastrak, Polhemus Inc. Colchester, VA, USA). Subsequently, the aligned anatomical image was used to create a volume conduction model based on a single shell description (Nolte, 2003) of the inner surface of the skull, using the segmentation function in SPM8. Source reconstruction was performed on a set of 8196 dipole locations distributed across the cortical sheet, which was extracted from the anatomical images, using Freesurfer 5.1 (Dale, Fischl, & Sereno, 1999). Next, these cortical surfaces were surface-registered to a template mesh using the Caret Software package (Van Essen et al., 2001), and subsequently downsampled from 168,342 dipoles per hemisphere to 4098 dipoles. The surface registration procedure resulted in individual cortical sheets that are topologically equivalent across participants (i.e. a particular topological point in the cortical sheet of one participant correspond to the same particular point in all other participants).

Source level coherence analysis

We used Dynamic Imaging of Coherent Sources (DICS; Gross et al., 2001) to compute coherence at the source
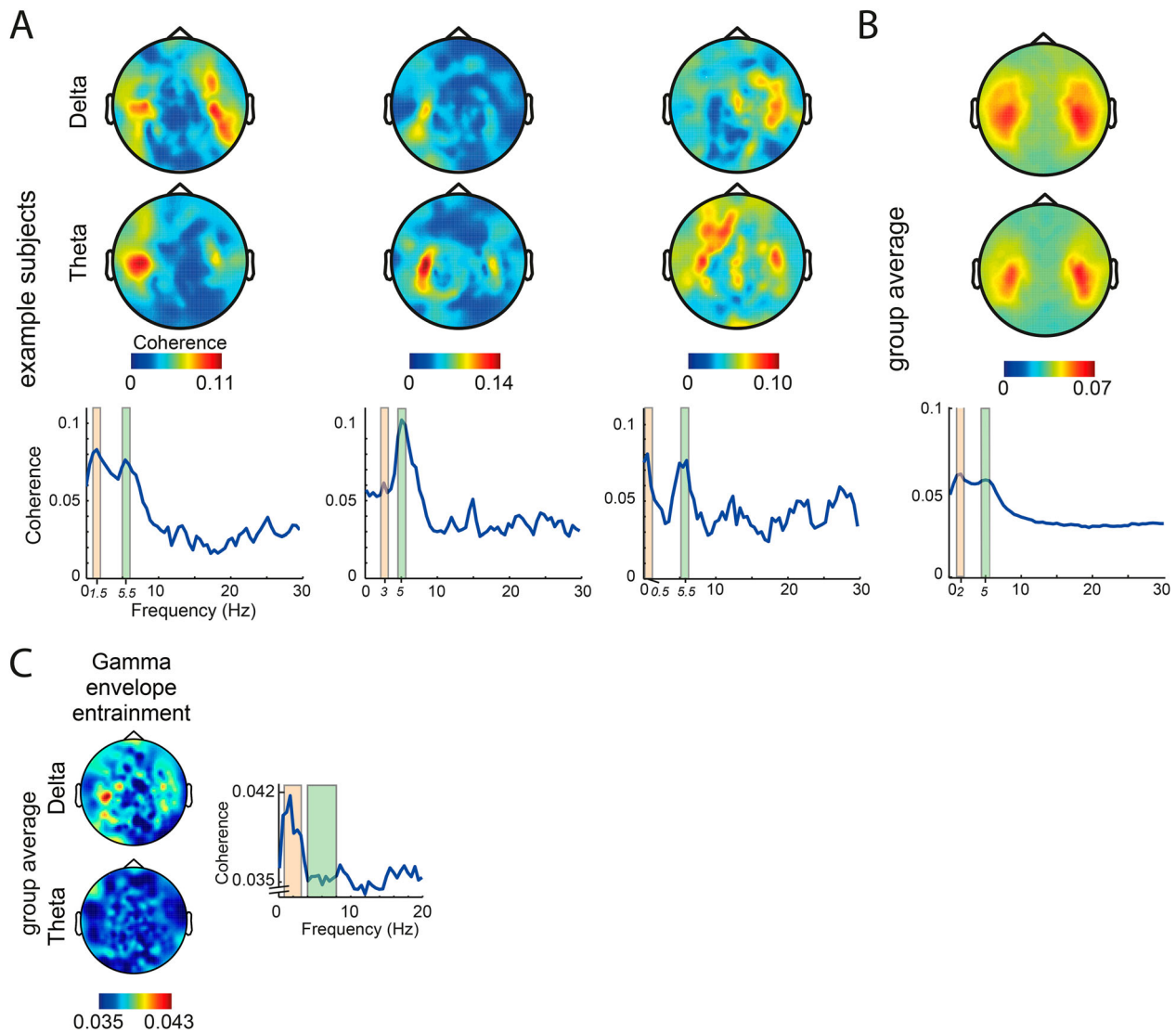
Figure 1. Sensor topographical distributions and spectra of neuronal entrainment to the speech envelope. (A): Individual subject topographies of entrainment in the delta band (top), theta band (middle), and corresponding frequency spectrum of entrainment (bottom: averaged across all MEG channels). Coloured bars reflect the frequency ranges used for the topographies displayed on the left. (B): Group average topography of delta entrainment (top), theta entrainment (middle), and corresponding frequency spectrum of entrainment (bottom: averaged across all MEG channels) (C): Group average topographies for MEG gamma band envelope entrainment by the audio envelope, for the delta band (upper panel) and theta band (lower panel), and corresponding spectrum averaged across 6 channels displaying the largest coherence in the delta band.

level. To obtain the envelope signal in the gamma band at the source level, for the entrainment analysis of the gamma band envelope to the speech input, we used Linearly Constrained Minimum Variance beamforming (LCMV; Van Veen, Van Drongelen, Yuchtman, & Suzuki, 1997), followed by a Hilbert transformation.

Statistical inference

For the theta and delta bands, we determined whether coherence was stronger in one hemisphere than the other, using a non-parametric permutation test together with a clustering method to control for the family wise error rate (Maris & Oostenveld, 2007). We quantified the difference in entrainment between each left hemispheric dipole location and its right-sided homologue as a dependent samples *t*-statistic. Samples that exceeded the uncorrected *p*-value of 1% were clustered according to adjacency (in space), and the cluster-based test statistic was defined by summing all suprathreshold *t*-values in a cluster. A reference distribution of cluster-based test statistics was created by permuting the labels of the hemispheres. For each permutation the maximal positive and negative cluster-level test statistic was computed. The observed test statistic was then tested against this reference distribution.

Results

Behavioural performance

The mean percentage of correct answers for the questions that proceeded a sentence was 86.8% (*SD* = 9.9%). This suggested that all participants were attentive and listened to the materials properly.

Sensor level coherence

We first computed coherence between neural activity and the speech envelope signal, focusing on the delta (0.5–3 Hz) and theta (4–7 Hz) frequency bands. Phase coupling was observed in both frequency ranges, and had a spatial maximum in temporal-parietal sensors. This topography was evident at both the group level and single subject level results (see Figure 1).

Using a peak detection algorithm, we identified the individual peak frequency for each frequency band, as displayed in Figure 2. We exploited this individual difference in the source analysis by only including participants with a distinct frequency peak in the sensor data, and using the peak frequency for source estimation. For the delta band, 88 participants showed a distinct coherence peak, and for the theta band, 91 participants were selected for further analysis. The main reason our algorithm did not detect a peak in a small number of the

participants was because there were too few channels (under 10%) with a clear peak in the theta band (see Peak frequency selection under Methods). This led to extremely low standardised peak values that did not surpass the peak detection threshold. Lowering the threshold revealed a peak for 3 out of 10 participants. Importantly, the lack of peaks was not because the peak fell on boundary of the defined theta window (4–7 Hz), since relaxing the boundaries to 3.5–7.5 Hz did not lead to the detection of peaks.

As participants received one of the six sets of stimuli, we investigated whether differences in low-level stimulus properties between sets could cause this variation in peak frequency. First, we inspected the peak frequencies within each set, and noted that the variation was maintained in each set (Figure 2(A)). Second, we calculated the set-specific power spectra of the stimulus envelopes to inspect the dominant frequency of each set-specific speech envelopes. We observed a strong overlap in power between the speech envelopes of each set. As shown in the power spectra (Figure 2(B)), 5 out of the 6 sets had a peak at 3 Hz for the delta band and at 5 Hz for the theta band. Since these sets of stimuli had similar energy profiles (envelopes), this suggests that the variation in peak frequency across subjects is due to subject-specific factors, and not due to differences in the input signals.
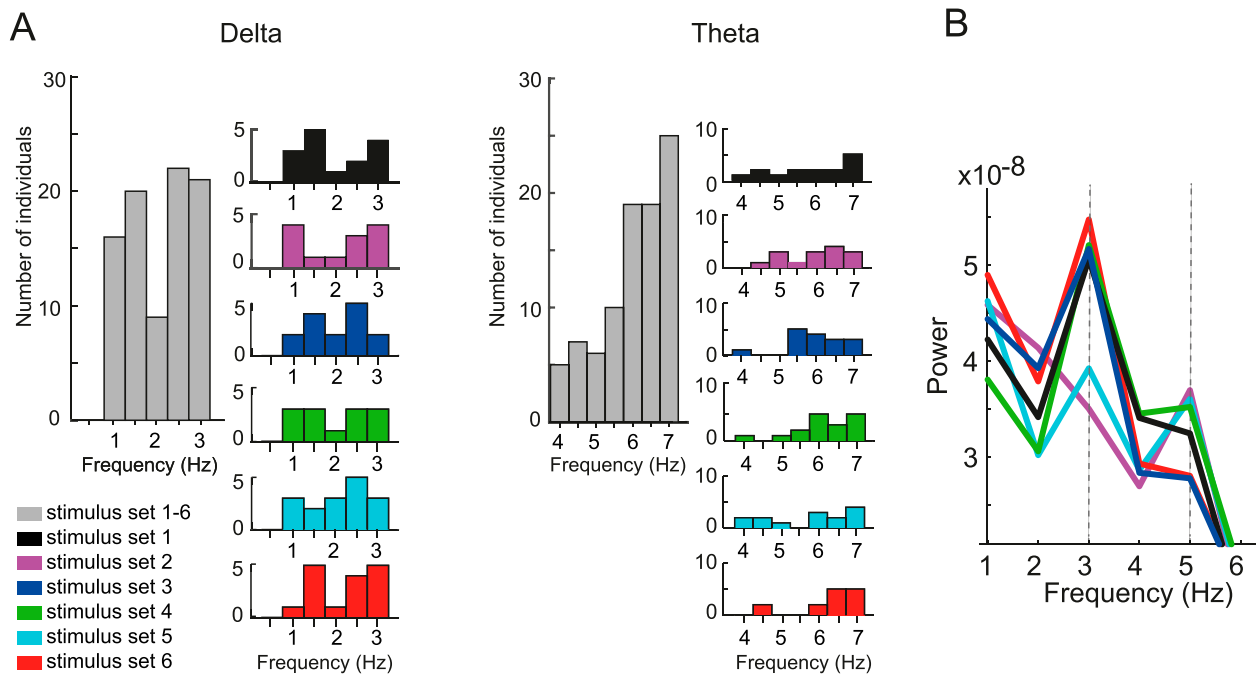

Figure 2. Individual peak frequencies of entrainment vary across subjects, but are independent on low-level characteristics of the input signal. (A): Histograms of the distribution of the peak frequencies of entrainment in the delta frequency band (left pair of columns), and the theta frequency band (right pair of columns). The coloured histograms reflect the distribution of peak frequencies of subsets of subjects, according to the exact set of stimuli they listened to. (B): Power spectra of the envelope of the speech signal for each set of stimuli.

Sensor level phase-amplitude coupling

There is some empirical evidence that the envelope of gamma oscillations entrains to the low-frequency phase of the speech envelope (Joachim Gross et al., 2013). We computed coherence between the gamma envelope (30–50 Hz) of the neural signals and the phase of the speech envelope. The average across subjects showed a small spatially (left lateralised) and spectrally selective effect in the delta band, but no effect in the theta band (Figure 1(C)). Note, in comparison to the data shown in Figure 1(B), that the amplitude of the spectral peak is approximately 5 times smaller than the average noise bias in the rest of the spectrum. At the single subject level, no convincing spectral peaks could be identified.

Source level low frequency phase-phase coupling

Figure 3(A) depicts the group level source coherence estimates and the statistical evaluation of lateralisation (Figure 3(B/C)), for delta (left panels) and theta (right panels). We did not source localise the alpha and beta band because there was no clear topography or peaks in the coherence spectrum at the sensor level. Coherence between speech and the delta and theta oscillations localised to superior temporal cortex (the activity in inferior motor cortex is due to spatial blur). This location is in line with early auditory areas. The spatial maps suggest that the peak location for delta entrainment may be slightly more posterior in bilateral temporal cortex, whereas it is more anterior for theta entrainment. This was statistically evaluated by comparing between frequencies the peak locations of the spatial maxima for the individuals that had both delta and theta entrainment. Specifically, we computed for each subject, hemisphere and frequency the average location of a set of dipoles, constrained to Brodmann Areas 22, 41 and 42. We selected the dipole locations at which the entrainment was ≥90% of the region-of-interest masked maximum value. The peak locations were compared between frequencies using Hotelling's T^2-statistic. Neither comparison (for left and right hemisphere separately) was significant (left hemisphere: $T^2 = 0.54$, p-value = 0.91, right hemisphere: $T^2 = 4.9$, p-value = 0.2).

Lateralisation of speech entrainment

According to the AST theory, (the auditory cortex in) each hemisphere has a bias in tracking a specific speech rhythm (Giraud & Poeppel, 2012; Poeppel, 2003), with a preferential role for the right hemisphere at the theta and delta frequencies. We performed a statistical comparison between the coherence values of the left and right homologous cortical regions, separately for the delta and theta bands. Only the theta band showed stronger entrainment on the right than left auditory cortex ($p = 0.0035$, corrected) (Figure 3(C)). Figure 3 (D) shows individual results for 4 example subjects, to illustrate the variability in lateralisation of entrainment. For instance, example subject 1 shows lateralised entrainment in both the delta and theta bands, whereas subject 2 is left lateralised for delta, and right lateralised for theta entrainment. Subjects 3 and 4 show a more (if not slightly left dominant) entrainment. To further explore the lack of right hemisphere dominance in the delta band, we computed a lateralisation index for each of the subjects. This was computed from the top 80 homologous dipole pairs in Brodmann areas 22, 41 and 42 (as reflected in a template anatomical atlas), which displayed the highest coherence at the group level. From these 80 dipole pairs, the lateralisation index was computed as (R-L)/(R + L), where R and L reflect the average coherence across the 80 right and left hemispheric dipoles, respectively. A similar number of left- and right-lateralised individuals were found for entrainment of both the delta and theta band (Figure 4). For the delta band, 55% (48/88) of the individuals were right-lateralised, and the mean right-, and left lateralised values were 0.36 and 0.36, respectively. For the theta band, 56% of individuals were right-lateralised (51/91) and there was a larger difference between the mean right-lateralised value (0.50) and left-lateralised value (0.32) which explains why only the theta but not delta band showed a significant bias for right hemisphere entrainment at the group level.

Discussion

Frequency-specific phase synchronisation of cortical oscillatory activity to the envelope of the auditory input signal has been proposed to provide an initial temporal parsing mechanism of the relevant linguistic structures needed for speech processing. In the current study, using a larger population than before ($N = 102$), we observed robust entrainment of neural oscillations in the delta and theta frequency bands, which was localised to early auditory cortical areas bilaterally. We also present novel findings of considerable individual variability in the peak frequency for entrainment and the preferred laterality of the entrainment.

Our data show entrainment to the speech envelope for the delta (0.5–3 Hz) and theta (4–7 Hz) bands, localised to bilateral primary auditory cortex and posterior superior temporal regions. Moreover, the theta band entrainment was significantly right lateralised at the
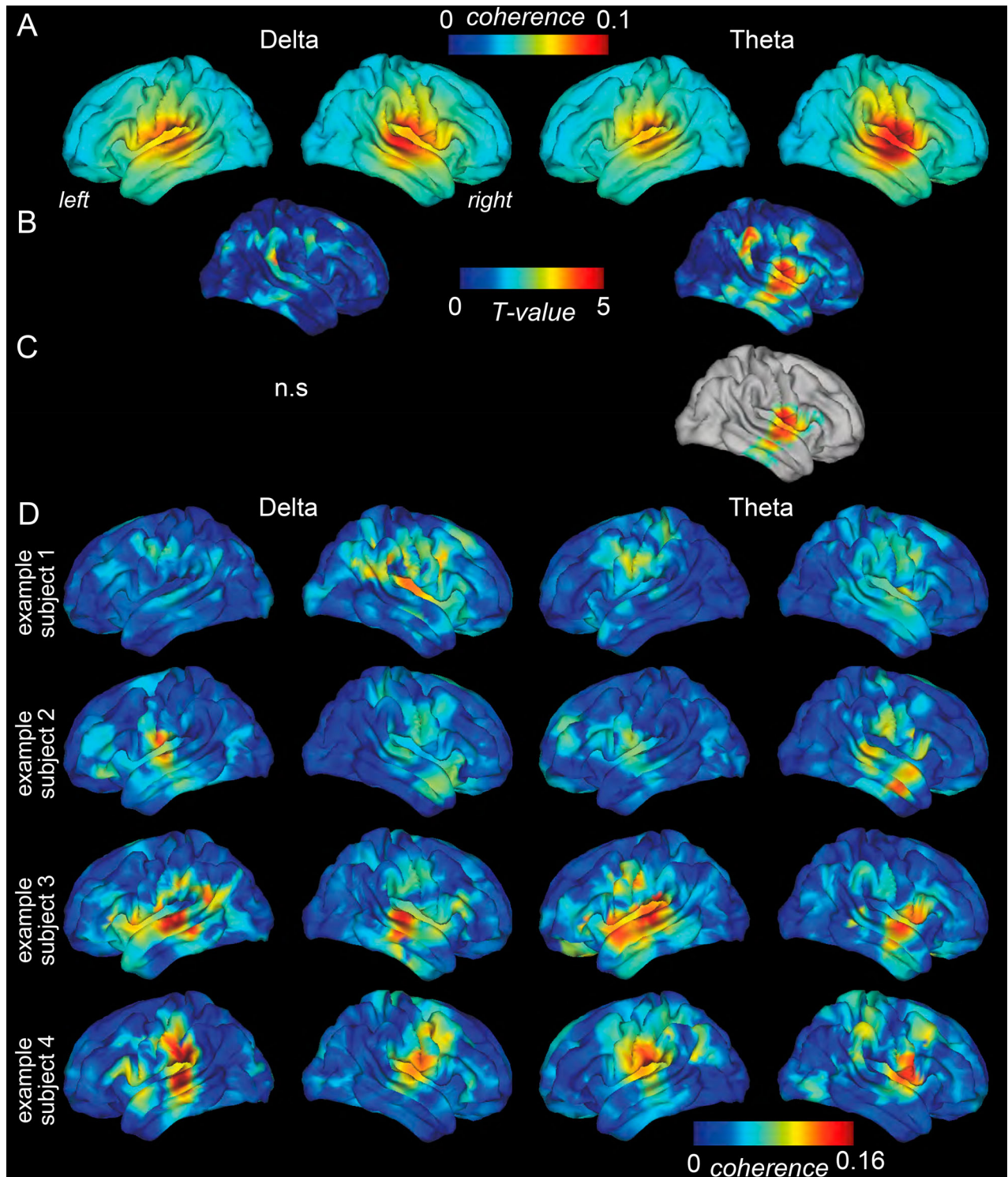
Figure 3. Source-level reconstruction of neuronal entrainment to the speech envelope. (A): Group average delta band (left panels) and theta band (right panels) entrainment, reconstructed at the cortical sheet. (B): Difference between right and left hemisphere homologous dipole locations, expressed as a T-statistic across subjects (right minus left), displayed on the right hemisphere. (C): Statistically masked difference between right and left hemisphere homologous dipole locations, displaying only the locations with a significant difference. (D): Single subject results for delta (left pair of columns) and theta (right pair of columns) entrainment for four subjects, illustrating the variability in lateralisation of the entrainment.
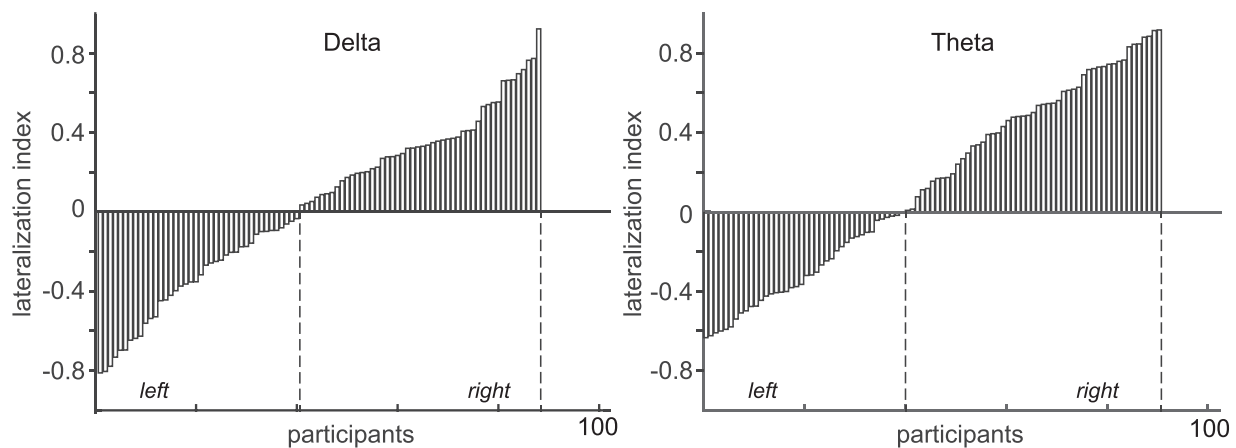
Figure 4. Distribution of single subject lateralisation indices for delta (left) and theta (right) entrainment.

group level, but the delta band entrainment was not. Furthermore, for these two frequency bands, we were able to identify clear spectral peaks at the single subject level, demonstrating the robustness and reliability of entrainment. We also observed weak left-lateralised phase-amplitude coupling between the signal envelope of the gamma band (30–50 Hz) and the speech envelope (Gross et al., 2013), although this could not robustly be identified in individual subjects.

Leveraging the large number of participants in our study, we explored individual differences in parameters of the entrainment. We show that both delta and theta entrainment vary across individuals in two aspects: the peak frequency of entrainment, as well as the degree and extent of lateralisation of entrainment. We observed individual variation in the peak frequency for both the delta and theta range, which could not be explained by the specific set of stimuli that the subjects listened to.

There has been a recent rise in popularity of the theory that the neural entrainment is a result of the phase-resetting of ongoing oscillations, where different frequencies are suited to track different rhythmicities in the speech input, and allow to segregate speech into smaller chunks for processing (Ghitza, 2011; Giraud et al., 2007; Giraud & Poeppel, 2012; Poeppel, 2003). As ongoing oscillations might vary across individual in their exact frequency content, this might explain the observed variability in the peak frequency of entrainment. Yet, in itself this variability in peak frequency does not prove that intrinsic oscillations are instrumental for the brain's response. Specifically, the observed brain response could well be the result of a series of overlapping transient responses to salient changes in the envelope of the speech signal. In such a scenario, intrinsic ongoing oscillations may be irrelevant, and the estimated entrainment would be just the frequency domain representation of a possibly non-linear cross-

correlation function between the brain response and the speech envelope. In other words, taking this line of reasoning even further, if the entrainment response is operationalised as a sequence of transient stereotypical responses of the auditory system to salient changes in the speech envelope signal, the differences in peak frequency may need to be explained in terms of subject specific latency differences in key components of these event-related responses.

To address which of these two explanations holds more merit, we propose further investigations that build on our current findings of peak frequency variability. For instance, one could seek evidence for oscillations underlying entrainment by showing that the individual frequency peaks of the on-going oscillatory activity in auditory regions, in the absence of auditory input, would correlate with the individual frequency peaks identified during auditory stimulation.

Next to observing substantial individual variability in the peak frequency of entrainment, we observed quite some variability with respect to the extent of hemispheric lateralisation of the entrainment. Specifically, we found an almost equal division of left- and right-lateralised individuals for delta and theta entrainment (Figure 4). Moreover, the degree of lateralisation varied across individuals. This suggests that theories on asymmetrical speech perception, which posit a functionally relevant lateralisation for slow frequencies (Giraud & Poeppel, 2012; Poeppel, 2003), may need revision. The fact that we found a significant right-lateralisation for theta entrainment at the group level was caused by the fact that right-lateralised subjects had a higher normalised lateralisation value (0.50) than left-lateralised individuals (0.31). In comparison, for the delta entrainment, the average normalised lateralisation values were similar between left-lateralised (0.36) and right-lateralised individuals (0.36). Taken together, these findings may

require revision of the notion that the right hemisphere by default is biased for the processing of slow time scales, as has been put forward by the AST-model.

Earlier work in support of this model have used hemodynamic signals (Boemio et al., 2005), and correlation estimates between hemodynamic signals and band-limited power fluctuations in the electrophysiological signal (Giraud et al., 2007). Only more recently, the use phase synchronisation measures has become the standard in order to evaluate the neuronal response to speech in the frequency domain (Bourguignon et al., 2013; Gross et al., 2013; Luo & Poeppel, 2007; Peelle et al., 2013). This is inspired by the putatively mechanistic role of oscillatory activity as a temporal parsing mechanism in speech processing. Our findings do not support a systematic right hemispheric lateralisation of the neural response to the slower time scales in the speech signal. In order to further investigate the variability in lateralisation, one line of enquiry might attempt to relate parameters of the functional response to structural parameters. For instance, anatomical variability in the shape (cortical folding), and extent of auditory cortical areas might affect the synchronicity of the electrophysiological response and/or measurability of the signal. Anatomical asymmetries across hemispheres might be related to the individual lateralisation of the response.

Given the large variability in both the frequency peak of the entrainment, and the extent of variability in hemispheric lateralisation one may wonder about the functional significance of the brain signals' entrainment to the audio envelope. Admittedly, the current work did not use an explicit experimental manipulation, yet all subjects were perfectly able to understand the stimuli. For this reason, one may argue that in normal, everyday use, the frequency specific entrainment response is not instrumental for the extraction of the linguistic content of speech signals.

In conclusion, we show that neural entrainment to speech is a robust and reliably detectable phenomenon, but the characteristics of the entrainment vary across individuals. Acknowledgement of this variation is important for understanding the underlying neural mechanisms of entrainment, as well as its functional significance.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

ORCID

Jan-Mathijs Schoffelen http://orcid.org/0000-0003-0923-6610

References

Abrams, D. A., Nicol, T., Zecker, S., & Kraus, N. (2008). Right-hemisphere auditory cortex is dominant for coding syllable patterns in speech. *Journal of Neuroscience*, *28*(15), 3958–3965. doi:10.1523/JNEUROSCI.0187-08.2008

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. *Proceedings of the National Academy of Sciences*, *98*(23), 13367–13372. doi:10.1073/pnas.201400998

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, *8*(3), 389–395. doi:10.1038/nn1409

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

Busch, N. A., Dubois, J., & VanRullen, R. (2009). The phase of ongoing EEG oscillations predicts visual perception. *Journal of Neuroscience*, *29*(24), 7869–7876. doi:10.1523/JNEUROSCI.0113-09.2009

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

Cogan, G. B., & Poeppel, D. (2011). A mutual information analysis of neural coding of speech by low-frequency MEG phase information. *Journal of Neurophysiology*, *106*(2), 554–563. doi:10.1152/jn.00075.2011

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. *NeuroImage*, *9*, 179–194. doi:10.1006/nimg.1998.0395

de Graaf, T. A., Gross, J., Paterson, G., Rusch, T., Sack, A. T., & Thut, G. (2013). Alpha-band rhythms in visual task performance: Phase-locking by rhythmic sensory stimulation. *PLoS ONE*, *8*(3), e60035. doi:10.1371/journal.pone.0060035

Ghitza, O. (2011). Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology*, *2*, 1–13. doi:10.3389/fpsyg.2011.00130

Ghitza, O., & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, *66*, 113–126. doi:10.1159/000208934

Giraud, A. L., Kleinschmidt, A., Poeppel, D., Lund, T. E., Frackowiak, R. S. J., & Laufs, H. (2007). Endogenous cortical rhythms determine cerebral specialization for speech perception and production. *Neuron*, *56*(6), 1127–1134. doi:10.1016/j.neuron.2007.09.038

Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and

operations. *Nature Neuroscience*, *15*(4), 511–517. doi: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 Biology*, *11*(12), e1001752. doi:10.1371/journal.pbio.1001752

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

Howard, M. F., & Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *Journal of Neurophysiology*, *2007*, 2500–2511. doi:10.1152/jn.00251.2010

Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron*, *53*(2), 279–292. doi:10.1016/j.neuron.2006.12.011

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

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

Lam, N. H. L., Schoffelen, J.-M., Udden, J., Hulten, A., & Hagoort, P. (2016). Neural activity during sentence processing as reflected in theta, alpha, beta, and gamma oscillations. *Neuroimage*, *142*, 43–54. doi:10.1016/j.neuroimage.2016.03.007

Luo, H., Liu, Z., & Poeppel, D. (2010). Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biology*, *8*(8), e1000445. doi:10.1371/journal.pbio.1000445

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

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. doi:10.1016/j.jneumeth.2007.03.024

Mcgettigan, C., & Scott, S. K. (2012). Cortical asymmetries in speech perception: What's wrong, what's right and what's left? *Trends in Cognitive Sciences*, *16*(5), 269–276. doi:10.1016/j.tics.2012.04.006

Molinaro, N., Barraza, P., & Carreiras, M. (2013). Long-range neural synchronization supports fast and efficient reading: EEG correlates of processing expected words in sentences. *NeuroImage*, *72*, 120–132. doi:10.1016/j.neuroimage.2013.01.031

Morillon, B., Lehongre, K., Frackowiak, R. S. J., Ducorps, A., Kleinschmidt, A., Poeppel, D., & Giraud, A. L. (2010). Neurophysiological origin of human brain asymmetry for speech and language. *Proceedings of the National Academy of Sciences*, *107*(43), 18688–18693. doi:10.1073/pnas.1007189107/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1007189107

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

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). Fieldtrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. doi:10.1155/2011/156869

Park, H., Ince, R. A. A., Thut, G., Gross, J., & Schyns, P. G. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners report frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Current Biology*, *25*, 1649–1653.

Peelle, J. E. (2012). The hemispheric lateralization of speech processing depends on what "speech" is: A hierarchical perspective. *Frontiers in Human Neuroscience*, *6*, 309. doi:10.3389/fnhum.2012.00309

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

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as "asymmetric sampling in time". *Speech Communication*, *41*(1), 245–255. doi:10.1016/S0167-6393(02)00107-3

Romei, V., Gross, J., & Thut, G. (2010). On the role of prestimulus alpha rhythms over occipito-parietal areas in visual input regulation: Correlation or causation? *Journal of Neuroscience*, *30*(25), 8692–8697. doi:10.1523/JNEUROSCI.0160-10.2010

Schoffelen, J. M., Hulten, A., Lam, N., Marquand, A., Udden, J., & Hagoort, P. (2017). Frequency-specific directed interactions in the human brain network for language. *Proceedings of the National Academy of Sciences*, *11*, e1001752. doi:10.1073/pnas.1703155114

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

Schroeder, C. E., Wilson, D. a., Radman, T., Scharfman, H., & Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. *Current Opinion in Neurobiology*, *20*(2), 172–176. doi:10.1016/j.conb.2010.02.010

Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, *34*(3), 114–123. doi:10.1016/j.tins.2010.11.002

Shtyrov, Y., Kujala, T., Palva, S., Ilmoniemi, R. J., & Näätänen, R. (2000). Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. *NeuroImage*, *12*, 657–663. doi:10.1006/nimg.2000.0646

Spaak, E., de Lange, F. P., & Jensen, O. (2014). Local entrainment of alpha oscillations by visual stimuli causes cyclic modulation of perception. *Journal of Neuroscience*, *34*(10), 3536–3544. doi:10.1523/JNEUROSCI.4385-13.2014

Stolk, A., Todorovic, A., Schoffelen, J. M., & Oostenveld, R. (2013). Online and offline tools for head movement compensation in MEG. *NeuroImage*, *68*, 39–48. doi:10.1016/j.neuroimage.2012.11.047

Van Essen, D. C., Drury, H. A., Dickson, J., Harwell, J., Hanlon, D., & Anderson, C. H. (2001). An integrated software suite for surface-based analyses of cerebral cortex. *Journal of the*

American Medical Informatics Association, *8*(5), 443–459. doi:10.1136/jamia.2001.0080443

Van Rullen, R., & Koch, C. (2003). Is perception discrete or continuous? *Trends in Cognitive Sciences*, *7*(5), 207–213. doi:10.1016/S1364-6613(03)00095-0

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