The Early Subcortical Response at the Fundamental Frequency of Speech Is Temporally Separated from Later Cortical Contributions

Alina Schüller[1], Achim Schilling[2], Patrick Krauss[1,2], and Tobias Reichenbach[1]

Abstract

■ Most parts of speech are voiced, exhibiting a degree of periodicity with a fundamental frequency and many higher harmonics. Some neural populations respond to this temporal fine structure, in particular at the fundamental frequency. This frequency-following response to speech consists of both subcortical and cortical contributions and can be measured through EEG as well as through magnetoencephalography (MEG), although both differ in the aspects of neural activity that they capture: EEG is sensitive to both radial and tangential sources as well as to deep sources, whereas MEG is more restrained to the measurement of tangential and superficial neural activity. EEG responses to continuous speech have shown an early subcortical contribution, at a latency of around 9 msec, in agreement with MEG measurements in response to short speech tokens, whereas MEG responses to continuous speech have not yet revealed such an early component. Here, we analyze MEG responses to long segments of continuous speech. We find an early subcortical response at latencies of 4–11 msec, followed by later right-lateralized cortical activities at delays of 20–58 msec as well as potential subcortical activities. Our results show that the early subcortical component of the FFR to continuous speech can be measured from MEG in populations of participants and that its latency agrees with that measured with EEG. They furthermore show that the early subcortical component is temporally well separated from later cortical contributions, enabling an independent assessment of both components toward further aspects of speech processing. ■

INTRODUCTION

Speech is a highly complex acoustic signal that needs to be processed in the brain in real time for comprehension. Investigations of the neural mechanisms that yield such rapid processing are increasingly employing more natural stimuli, from individual syllables and words to sentences and even entire stories (Brodbeck & Simon, 2020; Hickok & Poeppel, 2007). Such studies have, for instance, revealed cortical tracking of characteristic, slow rhythms in speech set by the rates of phonemes, syllables, and words (Brodbeck & Simon, 2020; Weissbart, Kandylaki, & Reichenbach, 2020; Etard & Reichenbach, 2019; Di Liberto, O'Sullivan, & Lalor, 2015; Ding & Simon, 2014).

Faster neural activity reflects the temporal fine structure of voiced speech such as vowels or voiced consonants. During the production of these speech parts, the vocal chords vibrate at a certain fundamental frequency (f_0), typically between 100 and 300 Hz, resulting in a periodic signal (Benesty, Sondhi, & Huang, 2008). The f_0 and its higher harmonics constitute the signal's temporal fine structure (Drullman, 1995; Rosen, 1992).

A subcortical response to the temporal fine structure of speech can be measured noninvasively in humans using EEG as well as magnetoencephalography (MEG) (Gorina-Careta, Kurkela, Hämäläinen, Astikainen, & Escera, 2021; Coffey et al., 2019; Bidelman, 2018; Coffey, Herholz, Chepesiuk, Baillet, & Zatorre, 2016). Moreover, such measurements of the frequency-following response to speech (speech-FFR) through EEG or MEG with source reconstruction have recently identified contributions from the auditory cortex (Kulasingham et al., 2020; Hartmann & Weisz, 2019; Bidelman, 2018; Coffey, Chepesiuk, Herholz, Baillet, & Zatorre, 2017; Coffey, Musacchia, & Zatorre, 2017; Coffey et al., 2016).

Although the spatial origins of the different neural contributions to the speech-FFR have been increasingly clarified, the temporal aspects remain less clear. Neural activity in the inferior colliculus is assumed to occur at a delay of 5–7 msec (Moore, 1987). However, EEG measurements of FFRs often find somewhat longer latencies of around 9 msec and up to 14 msec (Forte, Etard, & Reichenbach, 2017; Kraus, Anderson, & White-Schwoch, 2017; King, Hopkins, & Plack, 2016; Bidelman, 2015). MEG measurements of cortical contributions to the speech-FFR have an even larger uncertainty around the timing, pinning the response between 12 and 60 msec (Kulasingham et al., 2020; Coffey, Chepesiuk et al., 2017). Because the earliest sound-evoked neuronal activities in the auditory cortex can occur already 9 msec after a stimulus onset, the cortical

[1]Friedrich-Alexander-Universität Erlangen-Nürnberg, [2]Universitätsklinikum Erlangen

Journal of Cognitive Neuroscience 36:3, pp. 475–491
https://doi.org/10.1162/jocn_a_02103

contributions to the speech-FFR might overlap in time with subcortical contributions.

A better understanding of the timing and source of the different components of the speech-FFR matters for investigating the role of these different components for speech processing. The speech-FFR has been found to be related to frequency discrimination, to language experience, and to musical expertise (Marmel et al., 2013; Krishnan, Gandour, & Bidelman, 2010; Musacchia, Sams, Skoe, & Kraus, 2007; Wong, Skoe, Russo, Dees, & Kraus, 2007; Krishnan, Xu, Gandour, & Cariani, 2005). Moreover, we demonstrated that the EEG-measured response is modulated by selective attention to one of two competing speakers, presumably because of top–down feedback from higher cortical areas (Etard, Kegler, Braiman, Forte, & Reichenbach, 2019; Forte et al., 2017). However, a more precise understanding of the involved neural feedback loops requires a better spatiotemporal segregation of the different neural activities.

MEG and EEG measurements of the speech-FFR have predominantly employed short speech tokens such as single vowels or syllables and achieved high accuracy of spatial source localization because of the use of a high number of repetitions (Bidelman, 2018; Coffey, Chepesiuk et al., 2017; Coffey, Musacchia, et al., 2017; Coffey et al., 2016). However, the temporal spread in the autocorrelation of the voiced parts of these short speech signals limited the temporal resolution of the neural activities.

In particular, a recent MEG study estimated the response delay to a single repeated syllable through computing the explanatory power over successive windows of 12 msec in duration (Coffey et al., 2016). This power was found to increase between 0 and 36 msec for the subcortical structures, and between 0 and 48 msec for the auditory cortex. This comparatively coarse estimate of the different delays left unclear to which degree subcortical and cortical activities might temporally overlap. A recent EEG investigation into the speech-FFR elicited by repeated presentation of a short speech token did not determine the delays of the responses from the subcortical and cortical sources, but their relative delays, obtaining significant spread in these latencies (Bidelman, 2018). Another recent MEG experiment investigated the neural responses to pure tones of different frequency (Gorina-Careta et al., 2021). Although it could successfully discriminate between different subcortical and cortical contributions to the FFR, it did not allow to obtain timing estimates of the different sources.

As another approach, we recently employed continuous speech to measure the speech-FFR with EEG (Etard et al., 2019; Forte et al., 2017). We therefore extracted a fundamental waveform from the speech signal that, at each time instance, oscillates at f_0. This waveform could then be related to the EEG recordings through regularized regression, yielding temporal response functions (TRFs) that show the contribution of different scalp electrodes at different delays. This statistical approach allowed for an estimation of the delay of the response at about 9 msec, indicating that only subcortical contributions were measured.

A similar study analyzed MEG responses to continuous speech and found neural sources between 23 and 63 msec, indicating that mostly cortical contributions were recorded (Kulasingham et al., 2020). The lack of subcortical activity was presumably because of the low sensitivity of MEG for deeper sources and the relatively short speech material of 6 min per participant.

Here, we sought to measure and spatiotemporally localize both subcortical and cortical contributions to the speech-FFR using continuous speech as measured from MEG. To compensate for the poor sensitivity of MEG to subcortical activity, we employed comparatively long MEG measurements of 17 min per participant (Figure 1). We then performed a spatio-temporal source reconstruction to differentiate and locate the neural activities.

METHODS

Experimental Design

We employed an existing experimental data set that was collected for the study of continuous neuronal activity evoked by natural speech (Schilling et al., 2021). MEG recordings were obtained from 15 healthy right-handed monolingual native German speakers (20–42 years, 8 women, 7 men). Participants had no history of neurological illness, drug abuse, or hearing impairment. The study was granted permission by the ethics board of the University Hospital Erlangen. The number of participants was chosen based on previous studies on neural responses to speech (Brodbeck & Simon, 2020; Kulasingham et al., 2020; Etard et al., 2019).

The participants listened to continuous speech, presented in the form of an audio book, which was based on the German novel *Gut gegen Nordwind*, written by Daniel Glattauer and published by Hörbuch Hamburg. The audio book is available in stores, and permission to use it for current and future studies has been granted by the publisher. The audio book has a total duration of 4.5 hr. One female and one male speaker narrate alternately without a competing talker or other background noise. The auditory stimuli were presented at approximately 50 dB SPL. Small adjustments to the sound pressure level were carried out for a few participants to ensure they could hear the sound comfortably.

The first 40 min of the audio book were presented diotically to the participants in 10 parts. After each part, three multiple-choice comprehension questions had to be answered on a monitor, to test attention. The MEG recording during these breaks was eliminated from the analysis. Furthermore, the acoustic stimulation was stopped 2 times for a 5-min break. This resulted in a total experimental duration of approximately 1 hr. For this study, we only considered the MEG responses to the male

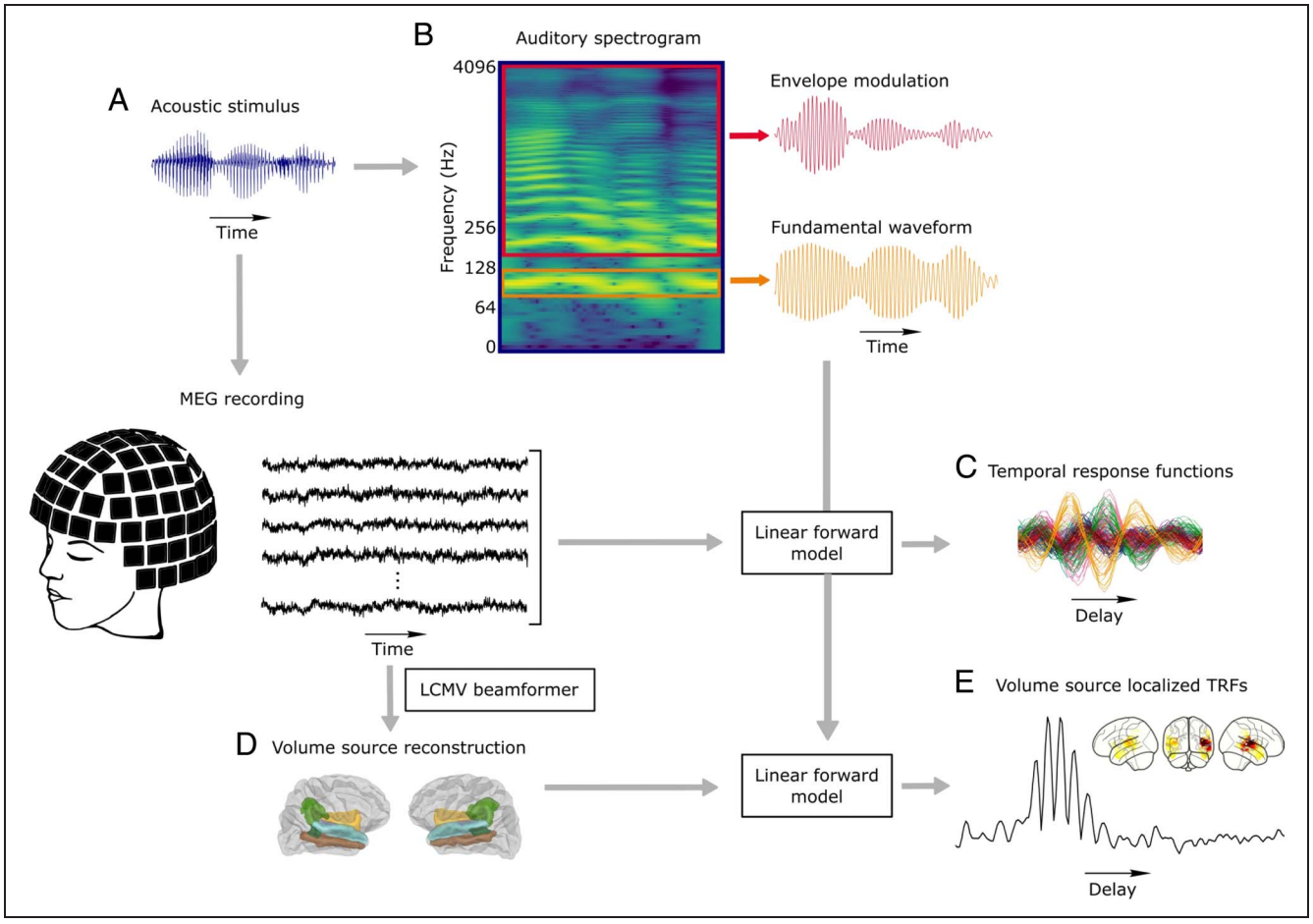

Figure 1. Overview of the experimental setup and data analysis. (A) We measured MEG (black) in response to continuous speech (blue). (B) A spectrogram was computed from the audio input to extract the fundamental waveform (orange) and the envelope modulation (red). (C) Sensor-level TRFs were calculated for both audio features through a linear forward model, which estimates the neural response from the speech features. (D) Volume source reconstruction was performed on an average MRI brain template for two ROIs by applying a LCMV beamformer to the preprocessed MEG data. (E) Volume source localized, that is, source-level, TRFs were calculated for both audio features through the linear forward model.

speaker (17 min of the 40-min audio stimulus) because of its lower f_0, leading to larger neural responses (Van Canneyt, Wouters, & Francart, 2021).

MEG data (248 magnetometer, 4D Neuroimaging) were recorded with a sampling frequency of 1017.25 Hz (supine position, eyes open, analogue band-pass filtering between 0.1 and 200 Hz). By the use of an integrated digitizer (Polhemus), five landmark positions were recorded and a calibrated linear weighting of 23 reference sensors (manufacturing algorithm, 4D Neuroimaging) was used to correct for environmental noise. The collected data were further processed by applying a digital band-pass filter (70–130 Hz) offline for speech feature analysis, as well as a 50-Hz notch filter. The data were furthermore downsampled to a sampling frequency of 1000 Hz.

The speech signal was presented simultaneously to the MEG recording (Figure 1A and B) through a custom-made setup that is described in detail by Schilling and colleagues (2021). A stimulation computer was connected to an external USB sound device, which provided five analogue outputs. The first two of these outputs were connected to an audio amplifier, of which the first output was connected in parallel to an analogue input channel of the MEG data logger.

An alignment of the speech stimulus to the MEG recording with an accuracy of 1 msec could be achieved through cross-correlating the speech stimulus with the audio reference recording obtained by the analogue input channel of the MEG data logger. A potential drift between the MEG recording system and the sound card because of the different clock speeds was found to be less than 1 msec within a 4-min part, so that we did not need to correct for such drift.

Data Analysis

Acoustic Stimulus Representations

We used two speech features to investigate the speech-FFR from MEG recordings, the fundamental waveform, as well as the high-mode envelope modulation (Figure 1B).

The first feature, the fundamental waveform, was computed through applying a bandpass-filter to the speech signal between 70 and 130 Hz, that is, around the f_0 that was, on average, 95 Hz. The so-obtained fundamental

waveform is very similar to that obtained from empirical mode decomposition (Etard et al., 2019; Forte et al., 2017; Huang & Pan, 2006). The neural response to the fundamental waveform captures the neural activity that emerges directly in response to the f_0 and is sometimes referred to as spectral FFR (Aiken & Picton, 2008).

Previous studies showed that the neural response at the f_0 of speech is also driven by the envelope modulation of the higher harmonics (Kegler, Weissbart, & Reichenbach, 2022; Kulasingham et al., 2020). We extracted these higher modes of the speech signal through an IIR filterbank that was inspired by the tonotopic organization of the cochlea. Different locations of the inner ear respond best to particular frequencies, with bandwidths that increase with the best frequency. Our IIR filterbank employed filters that were centered at multiples of the mean f_0 and whose bandwidth was proportional to the center frequency. The incorporation of overlapping filters ensured a seamless transition between neighboring frequency bands.

In detail, we extracted the mean $f_{0,mean}$ of the speech signal. We then defined the cutoff frequencies as $[n \cdot f_{0,mean} - n \cdot std(f_0),\ n \cdot f_{0,mean} + n \cdot std(f_0)]$, where $std(f_0)$ is the standard deviation of the f_0 and n represents the index of the higher mode. The applied band-pass filters were implemented using Python's *Scipy* library (Virtanen et al., 2020). A linear fourth order digital IIR-filter (critical frequencies obtained by dividing the lower and upper cutoff frequency by the Nyquist frequency) was applied twice, once forward and once backward, to prevent phase delays.

We subsequently applied a Hilbert transform to each mode, yielding an analytic signal, the magnitude of which served as envelope. All so-obtained higher-mode envelopes were then averaged across the envelopes and subsequently band-pass filtered between 70 and 130 Hz, that is, in the same range as the fundamental waveform. This yielded an acoustic feature that effectively captured the temporal modulation in the envelopes of the higher harmonics, within the 70- to 130-Hz range. The corresponding neural response has previously also been referred to as envelope-FFR (Aiken & Picton, 2008).

Temporal Response Functions

To investigate the origin of the neural response to continuous speech measured with MEG, we computed TRFs for the MEG channels as well as for the estimated vertices in the source space. We therefore applied a linear forward model that reconstructed the multichannel MEG response $y_t^{(c)}$ at each MEG channel (or source voxel) c and at time t from a linear combination of acoustic stimulus samples, shifted by time delays τ that ranged from a minimal value $\tau_{\min}$ to a maximal value $\tau_{\max}$:

$$y_t^{(c)} = \sum_{\tau=\tau_{\min}}^{\tau_{\max}} \left(\alpha_\tau^{(c)} e_{t-\tau} + \beta_\tau^{(c)} f_{t-\tau} \right) \quad (1)$$

where $e_{t-\tau}$ and $f_{t-\tau}$ describe the time-delayed envelope modulation and fundamental waveform, respectively. The weights $\alpha_\tau^{(c)}$ and $\beta_\tau^{(c)}$ of this linear combination are referred to as the TRFs of the two acoustic stimulus features (Figure 1C). A TRF can be viewed as the set of weights that best describe the time course of the neural response $y_t^{(c)}$ to a feature at each channel (or source voxel) c and therefore gives rise to the neural response to each acoustic feature across the different time lags τ.

The TRFs were computed for time lags ranging from $\tau_{\min} = -20$ msec to $\tau_{\max} = 140$ msec, with an increment of 1 msec, corresponding to a sampling frequency of 1000 Hz. This resulted in 161 time lags in total. Although we did not expect any neural response to occur at negative time lags, where the acoustic stimulus lagged behind the neural response, or for time lags larger than 100 msec, we still took both temporal ranges into account to control for the absence of significant responses there.

The TRF coefficients were estimated using regularized ridge regression (Hastie, Tibshirani, & Friedman, 2009). The regularization parameter λ can thereby be expressed as $\lambda = \lambda_n \cdot e_m$, in which e_m is the mean eigenvalue of the covariance matrix and λ_n is the normalized regularization parameter. For the TRF estimations in this study, we used a fixed normalized regularization parameter $\lambda_n = 0.1$ across all participants. This value was chosen based on cross-validation, which was applied to all participants individually and yielded an optimum around $\lambda_n = 0.1$ for each participant. The implementation of the forward model and the TRF estimation employed the algorithms developed by Etard and colleagues (2019) and Kegler and colleagues (2022).

Because of an incomplete data set for two of the 15 participants, we excluded those from the further analysis. The TRFs were estimated for each of the 13 participants (20–42 years, 7 females, 6 males) individually. Before calculating the TRFs, all acoustic features as well as the source-reconstructed MEG data underwent z-scoring. The TRF magnitudes were then averaged across participants, yielding population-average models. To obtain one TRF magnitude value for each time lag, the average of the magnitudes across MEG channels (or across vertices) was taken.

Neural Source Estimation

The neural sources of the MEG signals were computed using the MNE-Python software package (Gramfort et al., 2014). Because no subject-specific MR-scans were available, we used the Freesurfer template MRI *fsaverage* (Fischl, 2012). It is worth noting that the use of an average brain template can provide comparable results to individual MR scans in source localization analyses (Douw, Nieboer, Stam, Tewarie, & Hillebrand, 2018; Holliday, Barnes, Hillebrand, & Singh, 2003) and has been validated in previous research on neural mechanisms of speech

processing (Kulasingham et al., 2020). The head position of each participant with respect to the MEG scanner was recorded in the beginning and at the end of each measurement with five marker coils. Moreover, the head shape was digitized (Polhemus). The so-obtained subject-individual information were used to coregister the *fsaverage* brain template using rotation, translation, and uniform scaling. For one participant, there was no head digitization recorded and we therefore excluded this participant from the source reconstruction analysis, in addition to the two prior excluded participants, resulting in 12 participants (20–42 years, 6 women, 6 men) for the source analysis.

We created a volumetric source space for the average brain. The volume source space was defined on a regular grid of 5-mm spacing between neighboring grid points and the Freesurfer *aparc* + *aseg* parcellation was applied to define regions on which the sources were estimated. The so obtained source space calculated on the whole brain contained 14,629 source locations with arbitrary orientations.

For a region-specific analysis, we divided the volume source space into a cortical and a subcortical portion. The cortical space included the auditory cortex as well as Wernicke's area, which is also known as speech area, on both the right and left hemispheres (*aparc* labels: *middletemporal*, *transversetemporal*, *superiortemporal*, *bankssts*, *supramarginal*, and *insula*), leading to 525 source locations with arbitrary orientations. The subcortical region contained the brainstem (*aseg* labels: *Brain-Stem*), resulting in 207 source locations with arbitrary orientations.

To obtain a realistically shaped volume conductor model for source reconstruction, despite the lack of subject-specific MR scans, we used the boundary element model for the fsaverage brain template provided by Freesurfer. On the basis of the volume source space and the lead-field matrix computed in the forward solution, we then computed a linearly constrained minimum variance (LCMV) beamformer (Bourgeois & Minker, 2009), that is, a spatial filter that is scanned, with a set of weights, for each source location through the predefined source space grid and estimated the MEG activity at each source point independently. We thereby used a data covariance matrix estimated from a 1-min MEG data segment and a noise covariance matrix estimated from 3-min prestimulus empty room recordings. The beamformer was applied to the raw MEG data of each participant, leading to an estimation of a 3-D current dipole vector with a certain magnitude and direction at each of the source locations.

All brain plots show the 2-D projection of the maximum magnitude of activated voxels either on the whole brain or in the described ROI, with the average brain template as overlay.

Statistical Analysis

Statistical tests of the significance of neural responses on the population level were done by comparing the calculated TRFs to noise models. For each participant, the noise models were created by time-reversing the acoustic features. Because of the resulting dissimilarity between the time-reversed speech features and the MEG signal, the noise models could not contain any actual brain response at any of the considered time lags. This was done for each participant, resulting in two noise models (one for each audio feature) for each participant. Although the temporal relation between the audio stimulus and the MEG response is destroyed this way, the local temporal structure of the audio signal is preserved.

To assess statistical significance in the sensor-level TRFs, we bootstrapped the single-subject noise models for each audio feature. We therefore performed 10,000 permutations of the noise models, where we resampled the noise models across participants and time lags to obtain a distribution of noise model magnitudes. Subsequently, we estimated an empirical p value as the proportion of values from the noise distribution that exceeded the actual TRF. We thus evaluated the TRF at each specific time lag against the distribution of noise models. The p values were then corrected for multiple comparisons using the Bonferroni method and considering each time lag separately for comparison.

The statistical testing for the source-level TRFs was done analogously. For each participant and for each of the reversed audio features, we calculated noise source-level TRFs, using the source-reconstructed MEG data. We then applied the same bootstrap algorithm as described for the sensor-level TRFs.

The statistical analysis on single participants was carried out by only resampling the data 10,000 times from the noise model of the corresponding participant across time lags, but not from other participants.

In the following figures, we show the mean across the 10,000 permutations of the respective noise model, as well as the corresponding standard deviation.

To assess putative laterization of cortical activity in time regions where significant responses emerged, we applied a two-tailed Wilcoxon signed-rank test on the magnitude differences of the TRFs in these time windows between the right and the left cortical ROI.

RESULTS

We assessed the speech-FFR through two speech features. First, as already employed in our previous EEG studies, we computed a fundamental waveform from the speech signal that, at each time instance, oscillated at f_0 (Etard et al., 2019; Forte et al., 2017). Second, nonlinearities present not only within the cochlea but also in subsequent stages of auditory processing can yield neural responses at f_0 from the higher harmonics contained in the stimulus. We therefore considered envelope modulations at f_0 in the higher frequency bands as a speech feature, as has indeed previously been shown to elicit strong MEG and EEG signals (Kegler et al., 2022; Kulasingham et al., 2020).

Temporal Aspects of the Neural Responses on the Sensor Level

As a first approach, we computed TRFs that related the two speech features, the fundamental waveform and the envelope modulation, at several temporal lags, to the MEG signals at the different sensors (Figure 2). To assess at which latencies significant neural responses emerged, we compared, for the different latencies, the amplitude of the TRFs averaged over all MEG sensors to the amplitudes of noise models, with a Bonferroni correction for multiple comparisons (see Methods section). Averaging across all MEG sensors may lead to a more noisy signal than when selecting particular regions of interest. However, this conservative approach avoids bias toward particular MEG channels and captures all measured activities.

The TRFs for the neural response to the fundamental waveform of speech yielded significant activities at time lags that ranged from 33 to 56 msec, with a peak activity at 44 msec (Figure 2A and B). The topographic plot at the delay of the peak showed that the highest magnitudes occurred for MEG sensors in the right hemisphere.

Regarding the neural response to the envelope modulation, the corresponding TRF showed significant activity between 8 and 15 msec, with peak activity at 9 msec and the highest magnitudes equally distributed in the right and left regions. The neural activity then peaked a second time at 27 msec, with significant activity between 20 and 47 msec, and was shaped by the left frontal, right frontal, and right temporal regions (Figure 2C and D).

Cortical Component of the Speech-FFR

In addition to the temporal characteristics, we aimed to investigate the neural origins of the measured MEG signals. We therefore performed a subject-wise source estimation on the preprocessed MEG data to localize the origin of the measured neural responses. For the cortical

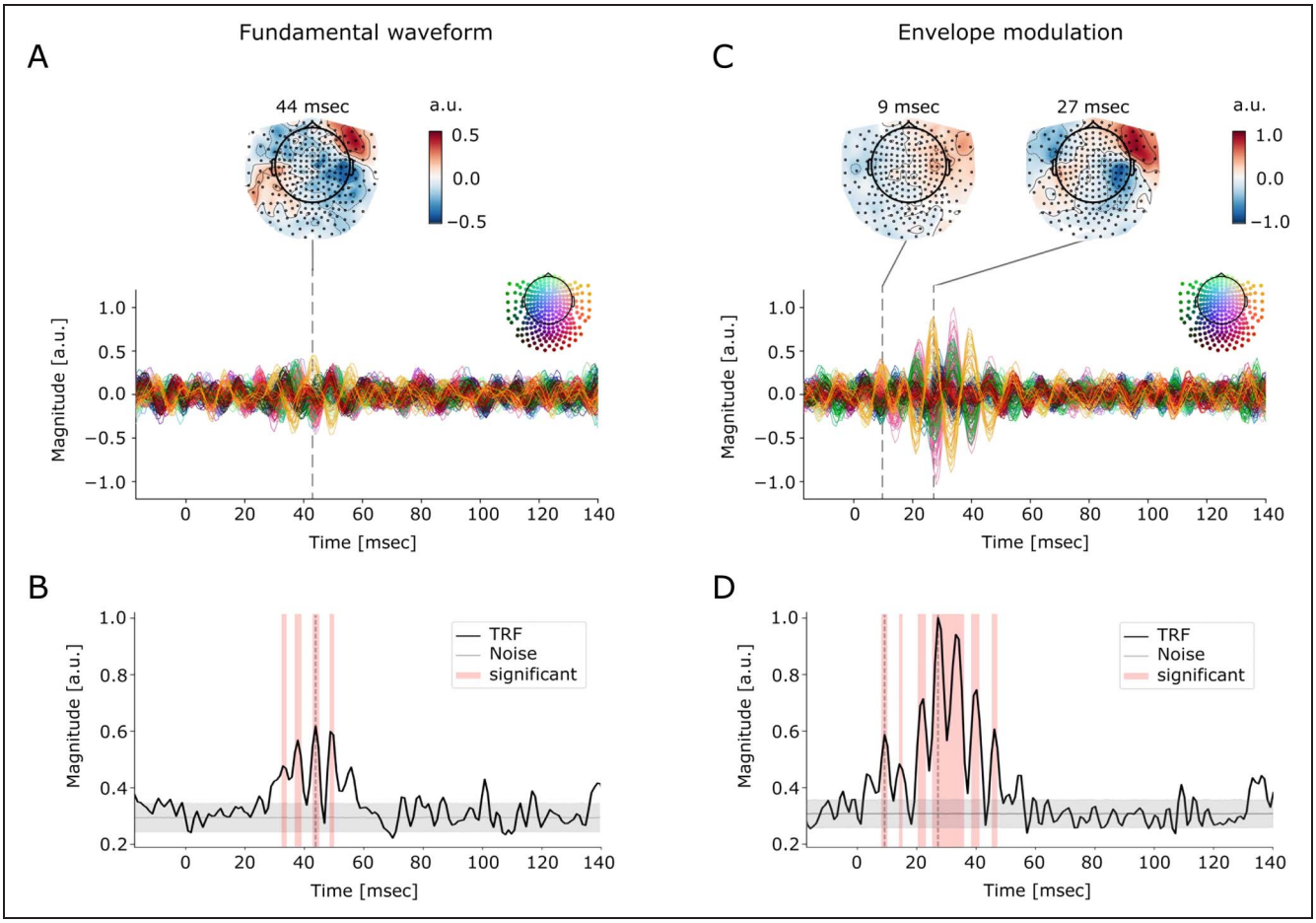

Figure 2. Sensor-level TRFs for the acoustic stimuli. (A, C) The normalized sensor-level TRF for each MEG sensor for time lags between −20 and 140 msec, for the fundamental waveform (A) and the envelope modulation (C). The delays at which the amplitude of the sensor-level TRFs peak are indicated by dashed lines; the topographic plots show the corresponding sensor activations. (B, D) The normalized absolute values of the TRFs averaged across the different sensors. The comparison of the TRF magnitudes to those of the noise models (mean ± standard deviation, gray line and shading) showed that significant responses emerged around certain peak latencies (red shaded area, thicker black line, $p < .05$, corrected for multiple comparisons). (B) The absolute value of the TRF for the fundamental waveform displays the largest peak at a delay of 44 msec (dashed line). (D) The absolute value of the TRF for the envelope modulation peaks at delays of 9 and 27 msec (dashed lines).

contribution, we determined a volume source space for the whole brain and estimated the preprocessed MEG signal at each voxel with a LCMV beamformer. We subsequently extracted the estimated voxel activities that were located in the predefined cortical ROI (see Methods section).

We then computed TRFs on the source-level MEG data to relate them to both speech features at the different time lags. As already done for the sensor-level TRFs, we computed the amplitude of the TRFs, averaged over all participants and all vertices in the cortical ROI (Figure 3A).

The average amplitudes of the source-level TRFs were tested for significance against a noise model using a bootstrap statistic for each time lag, with Bonferroni correction for multiple comparisons. The TRF for the fundamental waveform showed significant time lags in the range of 28 to 58 msec, significantly peaking at 44 msec as well as at 80 msec (Figure 3B). To further analyze the origin of the peak signals at 44 msec and at 80 msec, we projected the magnitudes of the subject-averaged voxel TRFs to the cortical ROI of the fsaverage brain template (Figure 3C). For the earlier peak, we found the highest magnitude in the right transverse-temporal part of the cortical ROI. The neural activity in this latency range and in the right hemisphere was indeed significantly higher than that in the left hemisphere (Wilcoxon signed-rank test, $p = 2.2 \times 10^{-6}$). The significant peak around 80 msec in contrast showed no significant right laterization (Wilcoxon signed-rank test, $p = .5$).

Regarding the TRF for the envelope modulation, we found significant time lags in the range of 19–47 msec, peaking at 27 msec (Figure 3D). As for the fundamental

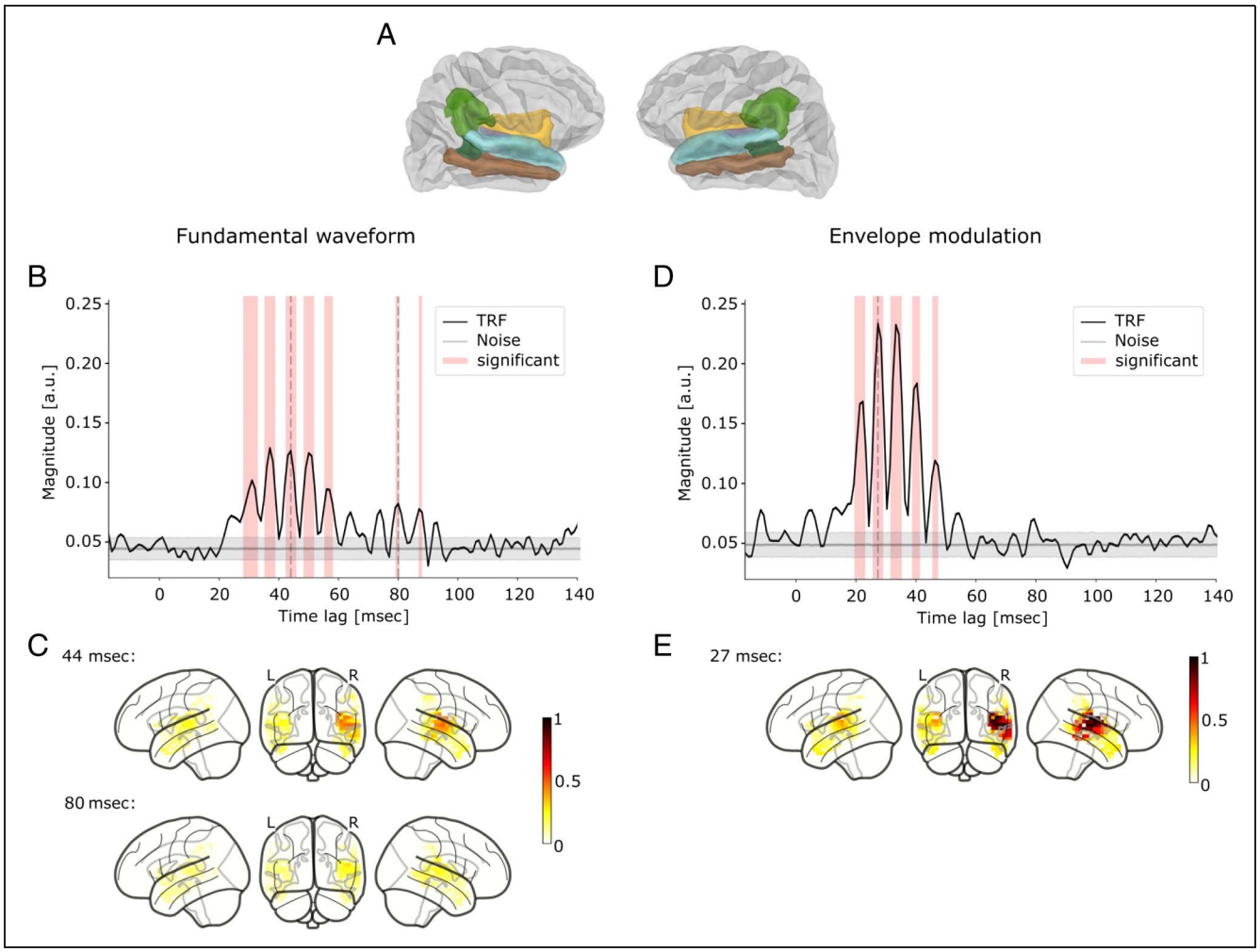

Figure 3. Volume source localization and source-level TRFs for the cortical ROI. (A) The cortical ROI consisted of 12 subregions of the Freesurfer *aparc + aseg* parcellation (middle-temporal in brown, transverse-temporal in purple, superior-temporal in turquoise, banks in dark green, supramarginal in yellow, and insula in light green, right and left each). (B, D) The normalized amplitude of the source-level TRF for the fundamental waveform (B) and the envelope modulation (D), averaged across participants and vertices in the cortical ROI. The significant time lags (red background and thicker black line, $p < .05$, corrected for multiple comparisons) peak at 44 msec for the fundamental waveform and at 27 msec for the envelope modulation (dashed lines). The results from the noise models are shown through the mean (gray line) ± the standard deviation (gray shading). (C) The projection of the magnitudes of the voxel TRFs in the cortical ROI to the average brain template at the peak latency of 44 msec showed a dominant contribution from the right Heschl's gyrus. (E) The highest magnitudes of the voxel TRFs in the cortical ROI at the latency of 33 msec occurred again in the right transverse temporal gyrus.

waveform TRF, the average amplitudes of the source-level TRFs were tested for significance against a noise model using a bootstrap statistic for each time lag, with Bonferroni correction for multiple comparisons. The projection of the estimated amplitudes of the subject-averaged vertex TRFs on the cortical ROI of the fsaverage brain template at the peak time lag of 27 msec (Figure 3D) showed that the highest magnitude occurred again in the right-hemispheric transverse-temporal region. The TRF amplitudes obtained in the right hemisphere were indeed significantly higher than those in the left hemisphere (Wilcoxon signed-rank test, $p = 9.3 \times 10^{-8}$).

Comparing the source-level TRFs of the fundamental waveform to those of the envelope modulation, we found that the amplitude of the latter was approximately twice as large as the amplitude of the former.

In addition to the population-averaged TRFs, we also assessed subject-specific TRFs for both speech features in the cortical ROI (Figure 4). The subject-specific TRFs were computed for each individual participant. They were determined to assess how reliable the neural responses could be detected at the level of individual participants, and to assess the subject-to-subject variability.

As for the population-averaged TRFs, the average amplitudes of the subject-specific TRFs were tested for statistical significance against a subject-specific noise model using a bootstrap statistic for each time lag, with Bonferroni correction for multiple comparisons.

The source-level, subject-specific TRFs in the cortical ROI for the envelope modulation feature showed significant peaks between 19 and 49 msec for 7 out of the 12 participants. For the fundamental waveform TRFs, 6 out of 12 participants revealed significant responses at time lags between 24 and 66 msec.

Subcortical Contribution to the Speech-FFR

In addition to the cortical investigation, we estimated the preprocessed MEG data on vertices of a subcortical ROI (Figure 5A), including the brainstem, to analyze possible subcortical contributions to the neural response. As for the cortical ROI, we computed TRFs on the source-level

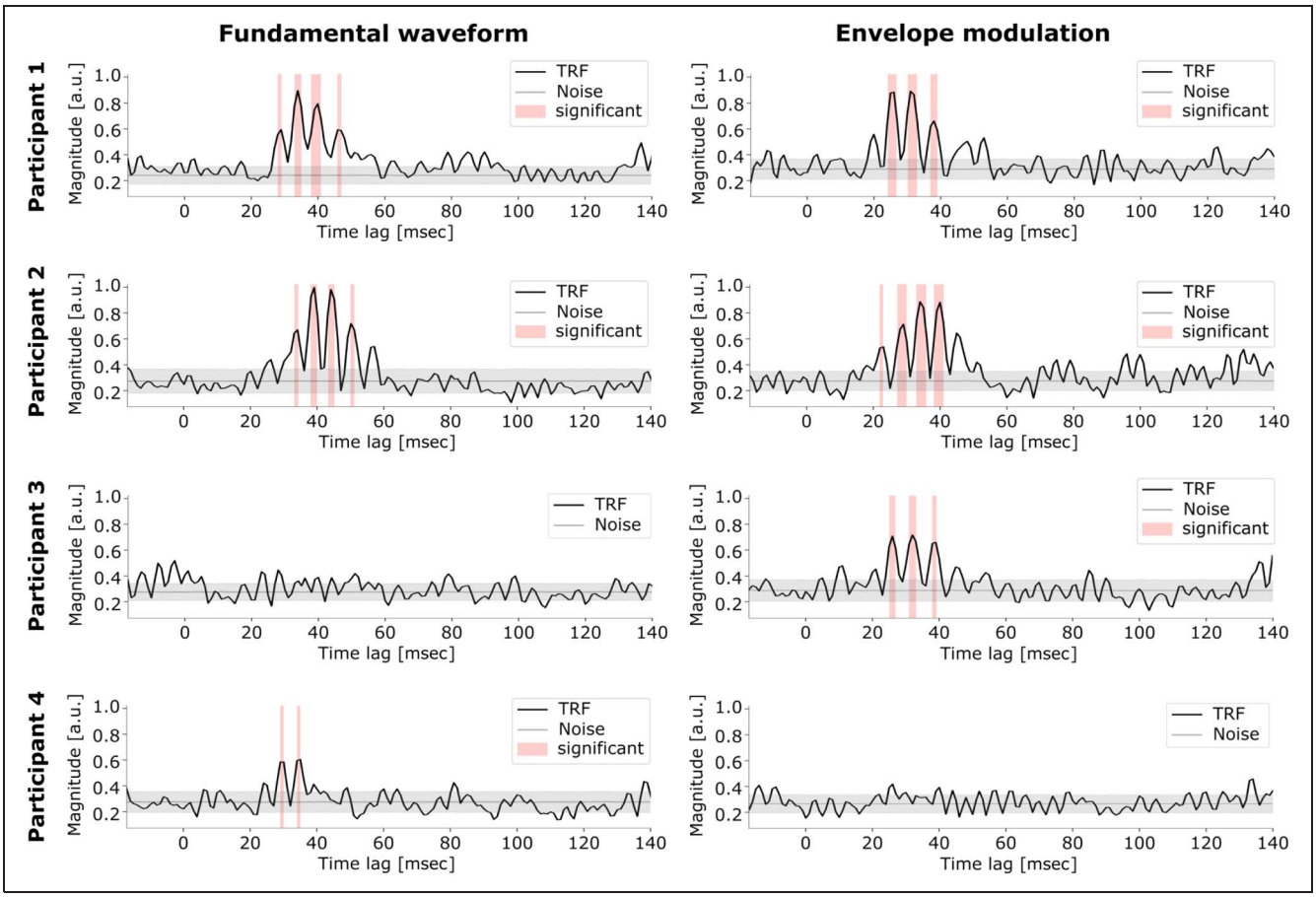

Figure 4. Source-level, subject-specific TRFs for the cortical ROI. The normalized amplitude of the source-level TRF for the fundamental waveform (left) and envelope modulation (right) is shown averaged across vertices in the cortical ROI for four participants. Time lags at which significant neural responses occur are highlighted through a red background ($p < .05$, corrected for multiple comparisons). The results from the noise models are shown through the mean (gray line) ± the standard deviation (gray shading). Participants 1 and 2 are examples where significant neural responses to both speech features occurred. For Participant 3, only the envelope modulation feature yielded a significant response. Participant 4, in contrast, shows examples where little or no significant neural activity is detected.

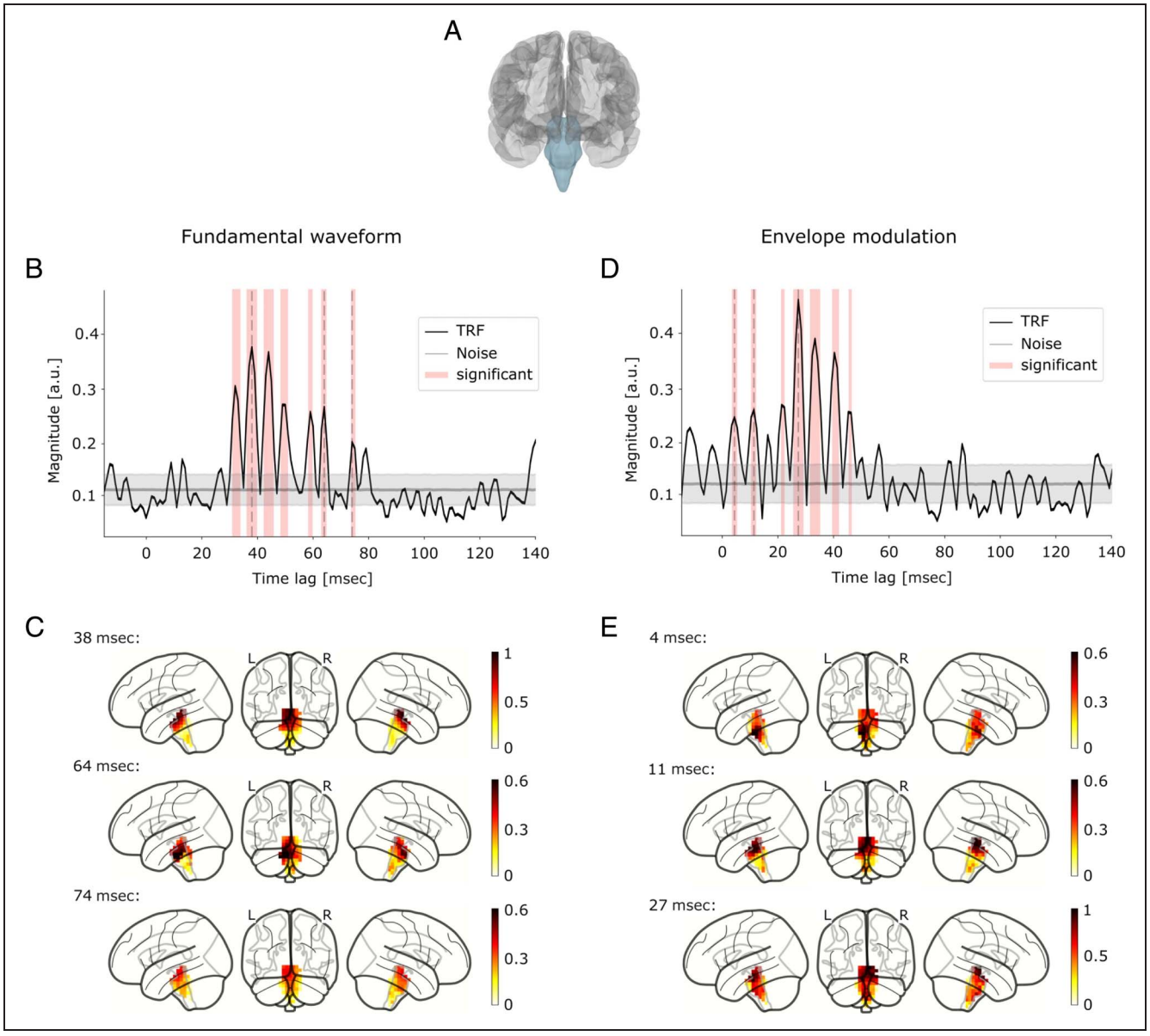

Figure 5. Volume source localization and source-level TRFs for the subcortical ROI. (A) The subcortical ROI consisted of the *Brain-Stem* region of the Freesurfer *aparc* + *aseg* parcellation. (B, D) The normalized amplitude of the source-level TRF for the fundamental waveform (B) and the envelope modulation (D) were averaged across participants and vertices in the subcortical ROI. Significant neural responses (red background and thicker black line, $p < .05$, corrected for multiple comparisons) at 38 msec to the fundamental waveform emerged, as well as later at 64 and 74 msec (dashed lines). For the envelope modulation, two peaks emerged at 4 and 11 msec and around 27 msec (dashed lines). The results from the noise models are shown through the mean (gray line) ± the standard deviation (gray shading). (C) This is the projection of the magnitudes of the source-level TRFs in the subcortical ROI to the average brain template at the peak latency of 38 msec (top), at 64 msec (middle), and at 74 msec (bottom). (e) Projection of the magnitudes of the source-level TRFs in the subcortical ROI to the average brain template at the latencies of 4, 11, and 27 msec.

MEG data, calculated on the whole brain and then extracted for the subcortical ROI for both audio features. The average amplitudes of the source-level TRFs were again tested for significance against a noise model using a bootstrap statistic for each time lag, with Bonferroni correction for multiple comparisons.

The normalized amplitude of the subject- and voxel-averaged TRF for the fundamental waveform showed significant time lags in the range of 31–51 msec and again from 58 to 65 msec and from 74 to 79 msec, peaking at 38 msec, 64 msec, and 74 msec (Figure 5B). The projection of the magnitudes of the subject-averaged, source-level TRFs at these peak latencies to the subcortical ROI, that is, the brainstem, of the average brain template indicated highest activation in the upper brainstem region at the earlier and later peak latencies, whereas at 64 msec, the activation was located predominantly in the mid to upper brainstem (Figure 5C).

For the envelope modulation feature, the normalized amplitude of the subject- and voxel-averaged TRF showed

significant neural activity at time lags from 3 to 12 msec, peaking at 4 and 11 msec, as well as between 21 and 46 msec, peaking at 27 msec (Figure 5D). We projected the source-level TRF magnitudes of both peaks to the subcortical ROI of the average brain template. At 4 msec, the highest activation emerged in the mid brainstem region, whereas at 11 msec and at 27 msec, the highest activation could be observed in the upper brainstem (Figure 5E).

As for the cortical ROI, we also computed subject-specific TRFs for both speech features in the subcortical ROI (Figure 6). The average amplitudes of the TRFs were again tested for significance against a noise model using a bootstrap statistic for each time lag, with Bonferroni correction for multiple comparisons.

The source-level, subject-specific TRFs in the subcortical ROI for the envelope modulation feature showed significant peaks between 20 and 45 msec for two out of the 12 participants, whereas one participant showed significant responses at time lags around 80 msec. For the fundamental waveform TRFs, two out of the 12 participants revealed significant responses at time lags between 33 and 51 msec.

Whole-brain Source Activation to the Speech-FFR

We estimated the source reconstruction of the preprocessed MEG data on a source space that has previously been calculated on the whole *fsaverage* brain volume. This resulted in 14,629 source locations with arbitrary orientations. For the cortical and subcortical analysis, we extracted from this whole-brain source reconstruction the portion of neural signal that arose in those vertices located in the respective ROI. However, to get an overview of the overall activity distribution in the brain, we also analyzed the TRFs that contained the activity of all 14,629 source points. Figure 7 shows the normalized amplitude of the subject- and voxel-averaged TRF for the fundamental waveform (Figure 7A, left) and the envelope modulation (Figure 7A, right).

For both acoustic features, the subject- and voxel-averaged TRF yielded strong neural responses for time lags that matched the ones that already arose in the cortical ROI. The corresponding projection of the voxel activation to the brain template confirmed the cortical origin of these strong responses (fundamental waveform: 44 msec, Figure 7B, left; envelope modulation: 27 msec, Figure 7B, right).

We were furthermore interested in whether the observed significant responses in the ROI analysis can already be detected in the whole-brain voxel activation. Therefore, we estimated projections on the brain template of the voxel activation for the time lags at which we found significant responses in the ROI analysis. For the fundamental waveform, the strong peak predominantly emerged in the cortical area. However, at 64 msec, the whole-brain voxel activation was weakly centered in subcortical and midbrain regions, whereas the latter that peaked at 74 and 80 msec showed an activation driven by the cortical region (Figure 7B, left).

Although the response at 27 msec for the envelope modulation revealed a strong cortical source activation, the whole-brain voxel activation for the envelope modulation feature showed a diffuse pattern at 5 and 11 msec (Figure 7B, right).

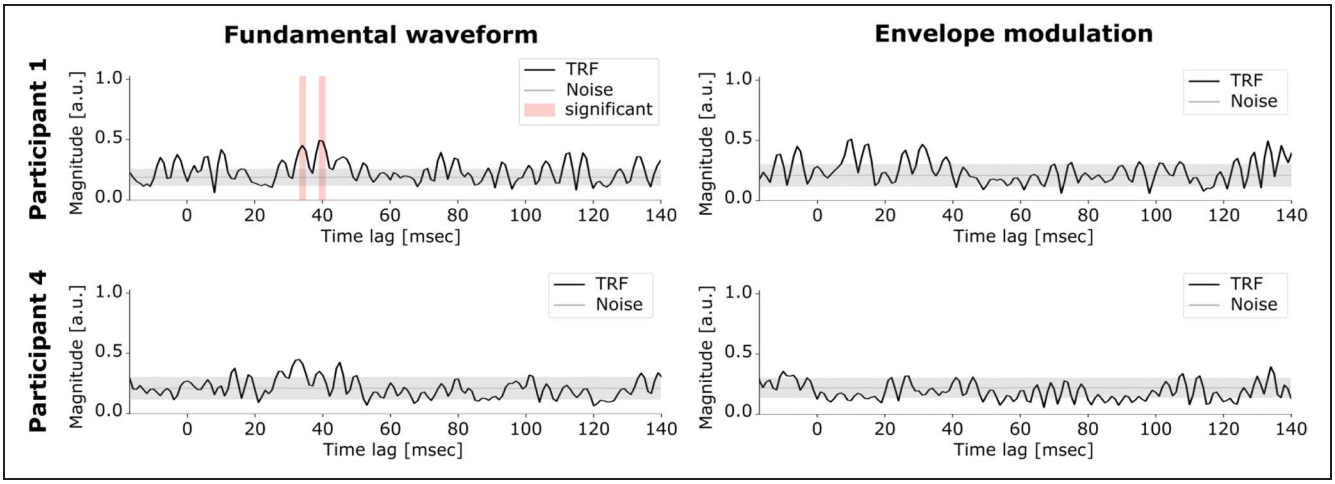

Figure 6. Source-level, subject-specific TRFs for the subcortical ROI, presented for two typical participants. The normalized amplitudes of the source-level TRFs for the fundamental waveform (left) and for the envelope modulation (right) are shown averaged across vertices in the subcortical ROI. Time lags at which significant neural responses occur are highlighted through a red background ($p < .05$, corrected for multiple comparisons). The results from the noise models are shown through the mean (gray line) ± the standard deviation (gray shading). The TRF for the fundamental waveform for Participant 1 revealed a significant peak between 33 and 41 msec. The TRF for the envelope modulation for Participant 1 revealed no significant peaks. The TRFs for Participant 4 showed no significant peak for both features.

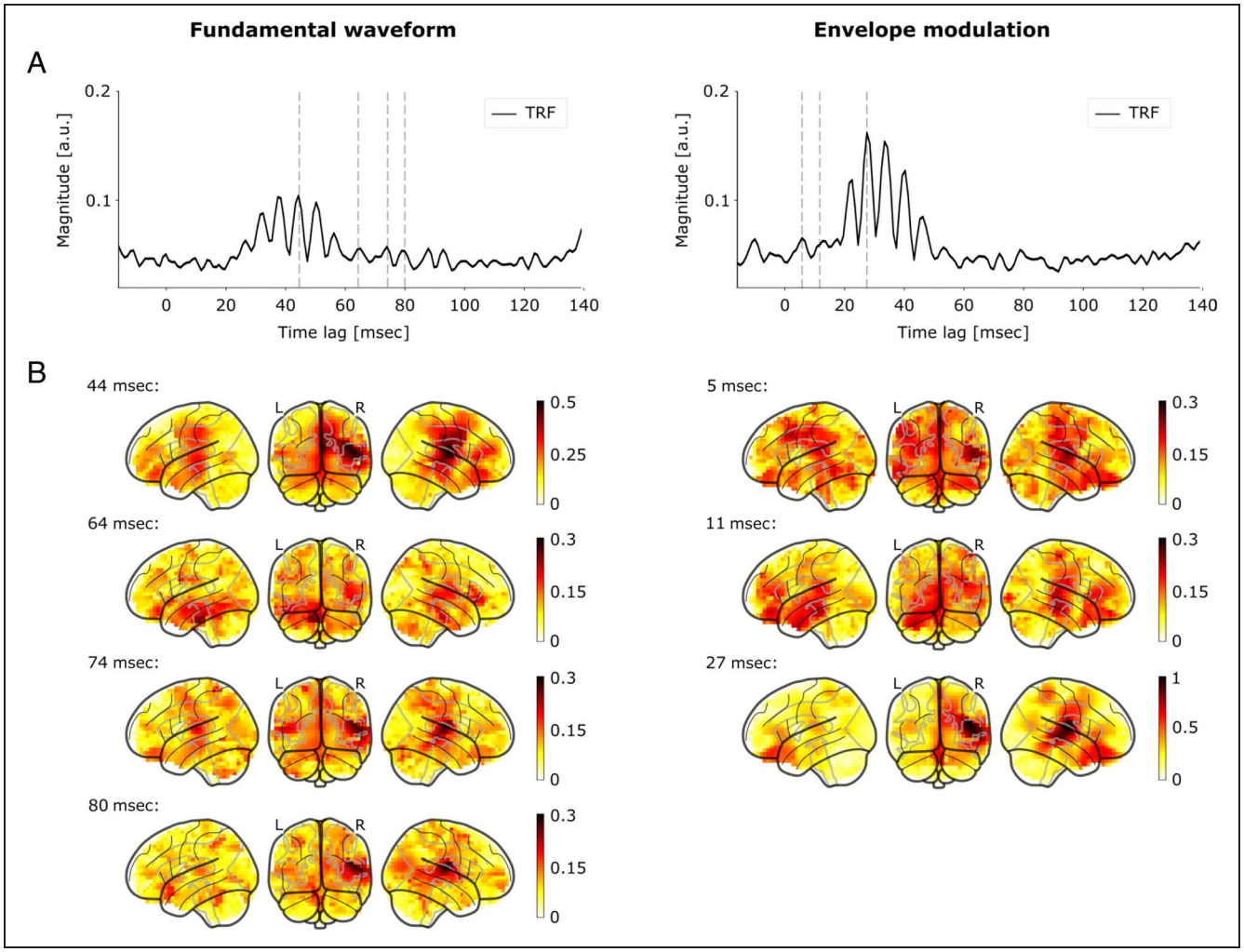

Figure 7. Volume source localization and source-level TRFs for the whole brain. (A) The normalized amplitudes of the source-level TRFs for the fundamental waveform (left) and for the envelope modulation (right) are shown, averaged across vertices. Dashed lines mark time lags that showed significant contributions in the previously performed ROI analysis. (B) Projection of the voxel magnitudes of the source-level TRFs for the fundamental waveform (left) and for the envelope modulation (right) to the average brain template at the time lags that showed significant contributions in the previously performed ROI analysis.

DISCUSSION

In this study, we explored subcortical and cortical contributions to the speech-FFR in response to continuous speech, based on MEG measurements. We showed that the cortical contribution can be measured in individual participants as well as on the population level, while the subcortical response was much weaker and emerged only reliably on the population level. We further showed that this early subcortical contribution is temporally separated from later cortical and putative subcortical activities.

Early Subcortical Neural Response on the Population Level

We considered two features of the speech stimulus: a fundamental waveform, oscillating at the speaker's f_0 of around 100 Hz, and the envelope modulation of the higher harmonics. Volume source reconstruction followed by the estimation of source-level TRFs allowed to trace the responses to these two speech features back to their neural origins in the brain. Because we performed the source reconstruction based on an average brain because of the absence of subject-specific MR scans, the interpretation of the obtained results needs to be done particularly carefully.

Importantly, we measured an early subcortical contribution to the speech-FFR. The early subcortical signal occurred in a temporal range between 3 and 15 msec, with a significant peak at a delay of 9 msec regarding the sensor-level analysis (Figure 2D) and at 4 msec as well as 11 msec regarding the source-level analysis (Figure 5D). Importantly, the peak subcortical activity occurred much before the first cortical activity peak, at 9 msec versus 27 msec.

The early subcortical activity was hence temporally well separated from the later cortical and putative subcortical responses.

Subcortical responses at the f_0 of continuous speech had been measured before in the same range of delays through EEG, confirming that our MEG measurements relate to the same neural activity (Etard et al., 2019; Forte et al., 2017). EEG is known to be more sensitive to deep subcortical sources, whereas MEG is known to have more sensitivity to cortical structures. This is because of the fact that MEG captures only the magnetic field generated by tangential dipoles in the cortex, rather than the one generated by the radial dipoles at the center of the head. The MEG recordings in this study were obtained from a magnetometer-based MEG. This type of MEG is more sensitive to deeper subcortical structures than gradiometer-based MEG (Baillet, 2017; Lopes da Silva & van Rotterdam, 2005). Prior studies on magnetometer-based MEG responses to clicks or short speech tokens accordingly identified subcortical contributions (Coffey et al., 2016; Parkkonen, Fujiki, & Mäkelä, 2009). Moreover, MEG recordings found contributions from the hippocampus (Cornwell, Arkin, Overstreet, Carver, & Grillon, 2012), the amygdala (Cornwell et al., 2008), and the thalamus (Roux, Wibral, Singer, Aru, & Uhlhaas, 2013) elicited by other types of stimuli.

The subcortical response that we measured between 4 and 11 msec presumably stems from multiple neural generators in the brainstem and midbrain, with the largest contribution expected from the inferior colliculus (Chandrasekaran & Kraus, 2010; Smith, Marsh, & Brown, 1975). We showed in a recent computational study that these different subcortical sources overlap considerably in time and cannot be distinguished for the speech-FFR, because of the significant spread in the autocorrelation function of the acoustic features (Saiz-Alía & Reichenbach, 2020). In our current study, we accordingly observed a peak of the subcortical contribution.

We note that the earliest significant activities that we obtain start at 3 msec (Figure 5D). However, this early activity might result from smearing of the response because of filtering. To not alter the latencies of the neural responses, we in fact employed forward–backward filters, which can lead to precausal artifacts and hence the early activity that we partly observe here.

Although we see early peaks emerging in the TRF for the fundamental waveform at delays of 9 and 15 msec (Figure 5B), these peaks are not statistically significant. However, its shape resembles that of the significant early peak at similar latencies in the TRF for the envelope modulation (Figure 5D), supporting the notion of the envelope modulation causing a greater neural response than the fundamental waveform, not only in cortical regions but also regarding subcortical activity. However, this result is contrary to a recent EEG study that found a response at a delay of 10 msec tracking the speaker's f_0 and a later response at a latency of 21 msec, tracking the envelope modulation (Kegler et al., 2022). These differences might once more reflect the different neural sources measured with MEG and EEG.

Late Subcortical Responses

The subject- and voxel-averaged, source-level TRFs yielded significant subcortical activity around the later time lags of 27, 38, 64, and 74 msec (Figure 5). These neural activities were mostly visible in the upper brainstem and less in the deeper brainstem structures, except for the peak at 64 msec, which might originate in the middle of the brainstem. Because of the comparatively late timing of the peaks at 27 and 38 msec, matching that of the stronger cortical responses, these neural activities may in fact stem from dispersed cortical activity and may have appeared in subcortical sources because of associated source reconstruction spread, that is, spatial smearing, or, to some extent, also because of volume conduction effects.

Volume conduction happens when electrical activity propagates from one part of the brain to another through the surrounding tissues and fluids, which may result in the appearance of activity in brain regions where it does not actually occur (Nunez & Srinivasan, 2006; Pascual-Marqui, 2002). However, this effect is less prominent in MEG responses than in EEG signals that are more sensitive to the electrical conductivity of head tissues (van den Broek, Reinders, Donderwinkel, & Peters, 1998). Source reconstruction spread refers to the spatial blurring or uncertainty associated with the estimated location of neural sources in the brain because of the inherent complexities of solving the underdetermined inverse problem. This effect is particularly not negligible when using an average brain template for source reconstruction, which can result in activity appearing in other brain regions (Grova et al., 2006; Darvas, Pantazis, Kucukaltun-Yildirim, & Leahy, 2004; Baillet, Mosher, & Leahy, 2001). These effects are indeed common and cannot be neglected when interpreting source reconstruction results that employ, in the lack of individual structural MRIs of the participants, an average brain MRI.

The even later peaks at 64 and 74 msec might reflect late brainstem activity that may emerge because of top–down interaction between the auditory cortex and the brainstem. Such top–down interaction is, for instance, assumed to cause the modulation of brainstem responses to speech by selective attention as well as linguistic factors (Kegler et al., 2022; Etard et al., 2019; Forte et al., 2017).

Subcortical Neural Response for Individual Participants

The assessment of the individual subject-specific TRFs for the subcortical ROI revealed no early significant responses for any of the 12 participants. This lack of a significant response in single participants presumably reflects the

lack of sensitivity of MEG to subcortical sources, hindering their reliable detection on the level of individual participants. It reinforces the need of a population average to robustly reveal the early subcortical response from MEG.

For five out of the 12 participants, we observed later responses around 40 msec. As already discussed above, these were also present in the population average and may stem from dispersed cortical activity.

Cortical Neural Response on the Population Level

The subject- and voxel-averaged, source-level TRF for the cortical ROI in response to the fundamental waveform exhibited at the peak latency time of 44 msec (Figure 3B) and revealed a dominant cortical origin in the transverse-temporal region. The source-localized cortical activity in response to the envelope modulation peaked at a latency of 27 msec. It had a greater amplitude than the response to the fundamental waveform, matching our findings in the sensor-level TRFs that suggested the envelope modulation feature to have a greater contribution to the neural response than the fundamental waveform. These results are also consistent with a recent MEG study that found a cortical response to the envelope modulation of the higher harmonics of continuous speech, emerging at a latency of about 40 msec (Kulasingham et al., 2020).

Despite the limitations posed by the use of an average MR scan and a 5-mm source grid, our results show the presence of cortical responses at the f_0 between 19 and 58 msec. Although natural continuous speech is much more complex than the repeated syllables used in a previous study on these neural responses, our observations match the long-lasting explanatory power of the neural responses in the auditory cortex described there (Coffey et al., 2016).

The neural responses in the cortical ROI to both speech features showed the highest magnitude in the transverse-temporal gyrus (Figure 3C and Figure 5E), also known as Heschl's gyrus, which defines the primary auditory cortex region. This is in line with previous studies indicating latencies around 40 msec to occur from Heschl's gyrus (Borgmann, Ross, Draganova, & Pantev, 2001; Yoshiura, Ueno, Iramina, & Masuda, 1995; Liégeois-Chauvel, Musolino, Badier, Marquis, & Chauvel, 1994).

We found a significant right laterization in the cortical ROI for the neural responses to both the f_0 and the envelope modulation feature. This phenomenon has already been shown in previous MEG studies on continuous speech (Kulasingham et al., 2020) as well as for short speech tokens (Coffey et al., 2016). Moreover, prior studies on neurophysiological processing of voice information with fMRI showed the right hemisphere to play a fundamental role in spoken language comprehension (Lattner, Meyer, & Friederici, 2005). Further studies on fMRI responses observed the right laterization using sung speech stimuli, indicating the presence of a brain asymmetry for speech and melody (Albouy, Benjamin, Morillon, & Zatorre, 2020). Motivated from prior studies that showed the right auditory cortex to be specialized for early tonal processing and pitch resolution, it has been suggested that the cortical responses occur as a consequence of early auditory processing of acoustic periodicity (Kulasingham et al., 2020; Cha, Zatorre, & Schönwiesner, 2016; Hyde, Peretz, & Zatorre, 2008).

We found a late cortical peak around 80 msec for the fundamental waveform (Figure 3B), which showed a similar but weaker activation pattern as the earlier peak at 44 msec. We therefore interpret this late peak to be the flattened activity that stems from the peak at 44 msec.

Cortical Neural Response for Individual Participants

Assessing individual subject-specific TRFs allowed us to investigate the variability in the speech-FFR across individuals. Our findings demonstrate that the speech-FFR evoked by continuous speech in the cortical ROI can be detected not only at the population level but also at the level of individual participants, providing further evidence for the robustness of this response.

In particular, we found that the neural response to the envelope modulation showed significant peaks in the source-level, subject-wise TRFs in the cortical ROI (Figure 4) for the majority of the participants (7 out of 12).

Accordingly, the TRFs for the fundamental waveform showed significant peaks in half the number of participants (6 out of 12), underlining the prior findings that the envelope modulation leads to a greater neural signal than the fundamental waveform (Kulasingham et al., 2020).

Neural Responses Analyzed on the Whole Brain on the Population Level

We computed the source reconstruction of the preprocessed MEG signal on the whole brain, including 14,629 sources. However, we initially concentrated on ROI analyses, extracting neural signals from vertices located in specific ROIs, one subcortical and one representing the auditory cortex area. This approach provided insights into the localized responses within the cortex for both the fundamental waveform and envelope modulation, as well as within the brainstem for both acoustic features.

Next, we aimed to put these ROI-specific results in a broader context, that is, analyzing whether we can also find the observed activities in TRFs calculated and averaged on all vertices contained in the whole-brain source space. We found that, for both the fundamental waveform and envelope modulation, strong neural responses were observed, with highest source activation in the cortical area (Figure 7). The spatial and temporal patterns for this high activation between 20 and 60 msec was similar as it

was in the cortical ROI analysis for both acoustic features (Figure 3).

The late peaks at 64 and 74 msec that we observed for the fundamental waveform in the subcortical ROI (Figure 5) indicated, when looking at the whole-brain activation, both subcortical as well as cortical contributions. The early subcortical contribution that we observed for the envelope modulation at 11 msec seemed to mainly stem from upper brainstem and midbrain sources (Figure 7B). However, the overall activity at these early time lags appears weak and diffuse when looking at the whole-brain voxel activation.

Although the combination of both the analysis in the ROIs and in the whole brain enriches our understanding of the complex neural processes underlying auditory perception, the results once more reflect the sensitivity of MEG to cortical sources rather than subcortical sources.

Differences in Neural Responses to the Fundamental Waveform and to the Envelope Modulation

In this study, we used two speech features, the fundamental waveform and the envelope modulation, to investigate the speech-FFR elicited by continuous speech. Using TRFs, we examined how neural responses to these different stimulus features arise in subcortical as well as cortical areas. Our results showed that both the envelope modulation and the fundamental waveform drive significant cortical responses. However, the envelope modulation caused a larger neural response than the fundamental waveform.

Regarding subcortical activity, we only observed significant activity in response to the envelope modulation, and not in response to the fundamental waveform. The latter was presumably too small to be detected, reinforcing the notion that the envelope modulation drives a stronger neural response than the fundamental waveform, both on the subcortical and on the cortical level.

Previous studies have indeed shown the perceptual relevance of envelope modulation in speech understanding, particularly above 300 Hz, as well as its greater resistance to background noise as compared with the lower frequencies below 200 Hz (Assmann & Summerfield, 2004). This perceptual relevance might be reflected in the larger neural response to the envelope modulation as opposed to the fundamental waveform that we observed here.

Moreover, previous studies on FFRs to specially designed tones discovered that the response at the f_0 emerges even when that frequency itself is missing from the stimulus, as long as higher harmonics are present (Galbraith, 1994; Smith, Marsh, Greenberg, & Brown, 1978). The extensive nonlinearities in the auditory system, starting from the compressive nonlinearity in the inner ear and continuing through the nonlinearities associated to neural responses, can indeed extract the f_0 from the higher harmonics. This mechanism appears to dominate over the direct neural response to the f_0 itself.

In line with our findings, a previous MEG study on the speech-FFR elicited by continuous speech likewise obtained larger cortical responses to the envelope modulation as compared with the fundamental waveform (Kulasingham et al., 2020). In addition, one of our earlier EEG investigations into this issue found that the envelope modulation explained a larger variance of the neural data than the fundamental waveform, further supporting our findings on the subcortical level (Kegler et al., 2022). Taken together, these previous findings as well as our current ones suggest that the envelope modulation is the more important speech feature for assessing the speech-FFR to continuous speech than the fundamental waveform.

Conclusions

In summary, we simultaneously recorded early subcortical and cortical contributions to the speech-FFR using natural continuous speech and MEG. Employing TRF analysis and neural source estimation, we showed that it is possible to separate contributions to the neural response from cortical and subcortical sources and, moreover, to assign those signals to two different features of continuous speech. This provides the opportunity of TRF analysis applied on MEG recordings to investigate further research questions on auditory processing of continuous speech under several stimulus conditions. However, the subcortical signal measured with MEG is weak and could only be detected under constraints such as the limitation of the analysis to specific ROIs.

The temporal separation of cortical and subcortical neural signals may allow to investigate the involvement of early subcortical responses in higher cognitive aspects of speech processing, such as attention to one of several competing speakers. Such studies might allow to further clarify interactions between subcortical and cortical structures during auditory processing.

Acknowledgments

The authors are grateful to the publishers Deuticke Verlag and Hörbuch Hamburg for the permission to use the novel and corresponding audio book *Gut gegen Nordwind* by Daniel Glattauer for the present and future studies.

Corresponding author: Tobias Reichenbach, Department Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg, Konrad-Zuse-Strasse 3, Erlangen, Germany, or via e-mail: tobias.j.reichenbach@fau.de.

Data Availability Statement

The MEG data used in this study are available from the authors upon request.

Author Contributions

Alina Schüller: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Writing—Original

draft. Achim Schilling: Data curation, Formal analysis; Methodology; Writing—Review & editing. Patrick Krauss: Data curation; Formal analysis; Methodology; Writing—Review & editing. Tobias Reichenbach: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Supervision; Writing—Review & editing.

Funding Information

Patrick Krauss, Deutsche Forschungsgemeinschaft (https://dx.doi.org/10.13039/501100001659), grant number: KR 5148/2-1. Alina Schüller, Deutsche Forschungsgemeinschaft (https://dx.doi.org/10.13039/501100001659), grant number: SCHI 1482/3-1. Patrick Krauss, Emerging Talents Initiative of the University Erlangen-Nuremberg, grant number: 2019/2-Phil-01.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

REFERENCES

Aiken, S. J., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research*, *245*, 35–47. https://doi.org/10.1016/j.heares.2008.08.004, PubMed: 18765275

Albouy, P., Benjamin, L., Morillon, B., & Zatorre, R. J. (2020). Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. *Science*, *367*, 1043–1047. https://doi.org/10.1126/science.aaz3468, PubMed: 32108113

Assmann, P., & Summerfield, Q. (2004). The perception of speech under adverse conditions. In S. Greenberg, W. A. Ainsworth, A. N. Popper, & R. R. Fay (Eds.), *Speech processing in the auditory system* (pp. 231–308). New York, NY: Springer. https://doi.org/10.1007/0-387-21575-1_5

Baillet, S. (2017). Magnetoencephalography for brain electrophysiology and imaging. *Nature Neuroscience*, *20*, 327–339. https://doi.org/10.1038/nn.4504, PubMed: 28230841

Baillet, S., Mosher, J. C., & Leahy, R. M. (2001). Electromagnetic brain mapping. *IEEE Signal Processing Magazine*, *18*, 14–30. https://doi.org/10.1109/79.962275

Benesty, J., Sondhi, M. M., & Huang, Y. (Eds.). (2008). *Springer handbook of speech processing* (Vol. 1). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-49127-9

Bidelman, G. M. (2015). Multichannel recordings of the human brainstem frequency-following response: Scalp topography, source generators, and distinctions from the transient ABR. *Hearing Research*, *323*, 68–80. https://doi.org/10.1016/j.heares.2015.01.011, PubMed: 25660195

Bidelman, G. M. (2018). Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *Neuroimage*, *175*, 56–69. https://doi.org/10.1016/j.neuroimage.2018.03.060, PubMed: 29604459

Borgmann, C., Ross, B., Draganova, R., & Pantev, C. (2001). Human auditory middle latency responses: Influence of stimulus type and intensity. *Hearing Research*, *158*, 57–64. https://doi.org/10.1016/S0378-5955(01)00292-1, PubMed: 11506937

Bourgeois, J., & Minker, W. (2009). Linearly constrained minimum variance beamforming. In J. Bourgeois & W. Minker (Eds.), *Time-domain beamforming and blind source separation: Speech input in the car environment* (pp. 27–38). Boston, MA: Springer. https://doi.org/10.1007/978-0-387-68836-7_3

Brodbeck, C., & Simon, J. Z. (2020). Continuous speech processing. *Current Opinion in Physiology*, *18*, 25–31. https://doi.org/10.1016/j.cophys.2020.07.014, PubMed: 33225119

Cha, K., Zatorre, R. J., & Schönwiesner, M. (2016). Frequency selectivity of voxel-by-voxel functional connectivity in human auditory cortex. *Cerebral Cortex*, *26*, 211–224. https://doi.org/10.1093/cercor/bhu193, PubMed: 25183885

Chandrasekaran, B., & Kraus, N. (2010). The scalp-recorded brainstem response to speech: Neural origins and plasticity. *Psychophysiology*, *47*, 236–246. https://doi.org/10.1111/j.1469-8986.2009.00928.x, PubMed: 19824950

Coffey, E. B. J., Chepesiuk, A. M. P., Herholz, S. C., Baillet, S., & Zatorre, R. J. (2017). Neural correlates of early sound encoding and their relationship to speech-in-noise perception. *Frontiers in Neuroscience*, *11*, 479. https://doi.org/10.3389/fnins.2017.00479, PubMed: 28890684

Coffey, E. B. J., Herholz, S. C., Chepesiuk, A. M. P., Baillet, S., & Zatorre, R. J. (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nature Communications*, *7*, 11070. https://doi.org/10.1038/ncomms11070, PubMed: 27009409

Coffey, E. B. J., Musacchia, G., & Zatorre, R. J. (2017). Cortical correlates of the auditory frequency-following and onset responses: EEG and fMRI evidence. *Journal of Neuroscience*, *37*, 830–838. https://doi.org/10.1523/JNEUROSCI.1265-16.2016, PubMed: 28123019

Coffey, E. B. J., Nicol, T., White-Schwoch, T., Chandrasekaran, B., Krizman, J., Skoe, E., et al. (2019). Evolving perspectives on the sources of the frequency-following response. *Nature Communications*, *10*, 5036. https://doi.org/10.1038/s41467-019-13003-w, PubMed: 31695046

Cornwell, B. R., Arkin, N., Overstreet, C., Carver, F. W., & Grillon, C. (2012). Distinct contributions of human hippocampal theta to spatial cognition and anxiety. *Hippocampus*, *22*, 1848–1859. https://doi.org/10.1002/hipo.22019, PubMed: 22467298

Cornwell, B. R., Carver, F. W., Coppola, R., Johnson, L., Alvarez, R., & Grillon, C. (2008). Evoked amygdala responses to negative faces revealed by adaptive MEG beamformers. *Brain Research*, *1244*, 103–112. https://doi.org/10.1016/j.brainres.2008.09.068, PubMed: 18930036

Darvas, F., Pantazis, D., Kucukaltun-Yildirim, E., & Leahy, R. M. (2004). Mapping human brain function with MEG and EEG: Methods and validation. *Neuroimage*, *23*, S289–S299. https://doi.org/10.1016/j.neuroimage.2004.07.014, PubMed: 15501098

Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015). Low-frequency cortical entrainment to speech reflects

phoneme-level processing. *Current Biology*, *25*, 2457–2465. https://doi.org/10.1016/j.cub.2015.08.030, PubMed: 26412129
Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience*, *8*, 311. https://doi.org/10.3389/fnhum.2014.00311, PubMed: 24904354
Douw, L., Nieboer, D., Stam, C. J., Tewarie, P., & Hillebrand, A. (2018). Consistency of magnetoencephalographic functional connectivity and network reconstruction using a template versus native MRI for co-registration. *Human Brain Mapping*, *39*, 104–119. https://doi.org/10.1002/hbm.23827, PubMed: 28990264
Drullman, R. (1995). Temporal envelope and fine structure cues for speech intelligibility. *Journal of the Acoustical Society of America*, *97*, 585–592. https://doi.org/10.1121/1.413112, PubMed: 7860835
Etard, O., Kegler, M., Braiman, C., Forte, A. E., & Reichenbach, T. (2019). Decoding of selective attention to continuous speech from the human auditory brainstem response. *Neuroimage*, *200*, 1–11. https://doi.org/10.1016/j.neuroimage.2019.06.029, PubMed: 31212098
Etard, O., & Reichenbach, T. (2019). Neural speech tracking in the theta and in the delta frequency band differentially encode clarity and comprehension of speech in noise. *Journal of Neuroscience*, *39*, 5750–5759. https://doi.org/10.1523/JNEUROSCI.1828-18.2019, PubMed: 31109963
Fischl, B. (2012). FreeSurfer. *Neuroimage*, *62*, 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021, PubMed: 22248573
Forte, A. E., Etard, O., & Reichenbach, T. (2017). The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention. *eLife*, *6*, e27203. https://doi.org/10.7554/eLife.27203, PubMed: 28992445
Galbraith, G. C. (1994). Two-channel brain-stem frequency-following responses to pure tone and missing fundamental stimuli. *Electroencephalography and Clinical Neurophysiology*, *92*, 321–330. https://doi.org/10.1016/0168-5597(94)90100-7, PubMed: 7517854
Gorina-Careta, N., Kurkela, J. L. O., Hämäläinen, J., Astikainen, P., & Escera, C. (2021). Neural generators of the frequency-following response elicited to stimuli of low and high frequency: A magnetoencephalographic (MEG) study. *Neuroimage*, *231*, 117866. https://doi.org/10.1016/j.neuroimage.2021.117866, PubMed: 33592244
Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2014). MNE software for processing MEG and EEG data. *Neuroimage*, *86*, 446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027, PubMed: 24161808
Grova, C., Daunizeau, J., Lina, J.-M., Bénar, C. G., Benali, H., & Gotman, J. (2006). Evaluation of EEG localization methods using realistic simulations of interictal spikes. *Neuroimage*, *29*, 734–753. https://doi.org/10.1016/j.neuroimage.2005.08.053, PubMed: 16271483
Hartmann, T., & Weisz, N. (2019). Auditory cortical generators of the frequency following response are modulated by intermodal attention. *Neuroimage*, *203*, 116185. https://doi.org/10.1016/j.neuroimage.2019.116185, PubMed: 31520743
Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). New York, NY: Springer. https://doi.org/10.1007/978-0-387-84858-7
Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*, 393–402. https://doi.org/10.1038/nrn2113, PubMed: 17431404
Holliday, I. E., Barnes, G. R., Hillebrand, A., & Singh, K. D. (2003). Accuracy and applications of group MEG studies using cortical source locations estimated from participants' scalp surfaces. *Human Brain Mapping*, *20*, 142–147. https://doi.org/10.1002/hbm.10133, PubMed: 14601140
Huang, H., & Pan, J. (2006). Speech pitch determination based on Hilbert-Huang transform. *Signal Processing*, *86*, 792–803. https://doi.org/10.1016/j.sigpro.2005.06.011
Hyde, K. L., Peretz, I., & Zatorre, R. J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia*, *46*, 632–639. https://doi.org/10.1016/j.neuropsychologia.2007.09.004, PubMed: 17959204
Kegler, M., Weissbart, H., & Reichenbach, T. (2022). The neural response at the fundamental frequency of speech is modulated by word-level acoustic and linguistic information. *Frontiers in Neuroscience*, *16*, 915744. https://doi.org/10.3389/fnins.2022.915744, PubMed: 35942153
King, A., Hopkins, K., & Plack, C. J. (2016). Differential group delay of the frequency following response measured vertically and horizontally. *Journal of the Association for Research in Otolaryngology*, *17*, 133–143. https://doi.org/10.1007/s10162-016-0556-x, PubMed: 26920344
Kraus, N., Anderson, S., & White-Schwoch, T. (2017). The frequency-following response: A window into human communication. In N. Kraus, S. Anderson, T. White-Schwoch, R. R. Fay, & A. N. Popper (Eds.), *The frequency-following response: A window into human communication* (pp. 1–15). Springer. https://doi.org/10.1007/978-3-319-47944-6_1
Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010). The effects of tone language experience on pitch processing in the brainstem. *Journal of Neurolinguistics*, *23*, 81–95. https://doi.org/10.1016/j.jneuroling.2009.09.001, PubMed: 20161561
Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, *25*, 161–168. https://doi.org/10.1016/j.cogbrainres.2005.05.004, PubMed: 15935624
Kulasingham, J. P., Brodbeck, C., Presacco, A., Kuchinsky, S. E., Anderson, S., & Simon, J. Z. (2020). High gamma cortical processing of continuous speech in younger and older listeners. *Neuroimage*, *222*, 117291. https://doi.org/10.1016/j.neuroimage.2020.117291, PubMed: 32835821
Lattner, S., Meyer, M. E., & Friederici, A. D. (2005). Voice perception: Sex, pitch, and the right hemisphere. *Human Brain Mapping*, *24*, 11–20. https://doi.org/10.1002/hbm.20065, PubMed: 15593269
Liégeois-Chauvel, C., Musolino, A., Badier, J. M., Marquis, P., & Chauvel, P. (1994). Evoked potentials recorded from the auditory cortex in man: Evaluation and topography of the middle latency components. *Electroencephalography and Clinical Neurophysiology*, *92*, 204–214. https://doi.org/10.1016/0168-5597(94)90064-7, PubMed: 7514990
Lopes da Silva, F. H., & van Rotterdam, A. (2005). Biophysical aspects of EEG and magnetoencephalographic generation. In E. Niedermeyer & F. H. da Lopes, Silva (Eds.), *Electroencephalography: Basic principles, clinical applications and related fields* (5th ed.). New York: Lippincott Williams & Wilkins.
Marmel, F., Linley, D., Carlyon, R. P., Gockel, H. E., Hopkins, K., & Plack, C. J. (2013). Subcortical neural synchrony and absolute thresholds predict frequency discrimination independently. *Journal of the Association for Research in Otolaryngology*, *14*, 757–766. https://doi.org/10.1007/s10162-013-0402-3, PubMed: 23760984
Moore, J. K. (1987). The human auditory brain stem as a generator of auditory evoked potentials. *Hearing Research*, *29*, 33–43. https://doi.org/10.1016/0378-5955(87)90203-6, PubMed: 3654395

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences, U.S.A.*, *104*, 15894–15898. https://doi.org/10.1073/pnas.0701498104, PubMed: 17898180

Nunez, P. L., & Srinivasan, R. (2006). *Electric fields of the brain: The neurophysics of EEG* (2nd ed.). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195050387.001.0001

Parkkonen, L., Fujiki, N., & Mäkelä, J. P. (2009). Sources of auditory brainstem responses revisited: Contribution by magnetoencephalography. *Human Brain Mapping*, *30*, 1772–1782. https://doi.org/10.1002/hbm.20788, PubMed: 19378273

Pascual-Marqui, R. D. (2002). Standardized low resolution brain electromagnetic tomography (sLORETA): Technical details. *Methods and Findings in Experimental and Clinical Pharmacology*, *24(Suppl D)*, 5–12.

Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *336*, 367–373. https://doi.org/10.1098/rstb.1992.0070, PubMed: 1354376

Roux, F., Wibral, M., Singer, W., Aru, J., & Uhlhaas, P. J. (2013). The phase of thalamic alpha activity modulates cortical gamma-band activity: Evidence from resting-state MEG recordings. *Journal of Neuroscience*, *33*, 17827–17835. https://doi.org/10.1523/JNEUROSCI.5778-12.2013, PubMed: 24198372

Saiz-Alía, M., & Reichenbach, T. (2020). Computational modeling of the auditory brainstem response to continuous speech. *Journal of Neural Engineering*, *17*, 036035. https://doi.org/10.1088/1741-2552/ab970d, PubMed: 32460257

Schilling, A., Tomasello, R., Henningsen-Schomers, M. R., Zankl, A., Surendra, K., Haller, M., et al. (2021). Analysis of continuous neuronal activity evoked by natural speech with computational corpus linguistics methods. *Language, Cognition and Neuroscience*, *36*, 167–186. https://doi.org/10.1080/23273798.2020.1803375

Smith, J. C., Marsh, J. T., & Brown, W. S. (1975). Far-field recorded frequency-following responses: Evidence for the locus of brainstem sources. *Electroencephalography and Clinical Neurophysiology*, *39*, 465–472. https://doi.org/10.1016/0013-4694(75)90047-4, PubMed: 52439

Smith, J. C., Marsh, J. T., Greenberg, S., & Brown, W. S. (1978). Human auditory frequency-following responses to a missing fundamental. *Science*, *201*, 639–641. https://doi.org/10.1126/science.675250, PubMed: 675250

Van Canneyt, J., Wouters, J., & Francart, T. (2021). Neural tracking of the fundamental frequency of the voice: The effect of voice characteristics. *European Journal of Neuroscience*, *53*, 3640–3653. https://doi.org/10.1111/ejn.15229, PubMed: 33861480

van den Broek, S. P., Reinders, F., Donderwinkel, M., & Peters, M. J. (1998). Volume conduction effects in EEG and MEG. *Electroencephalography and Clinical Neurophysiology*, *106*, 522–534. https://doi.org/10.1016/S0013-4694(97)00147-8, PubMed: 9741752

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2, PubMed: 32015543

Weissbart, H., Kandylaki, K. D., & Reichenbach, T. (2020). Cortical tracking of surprisal during continuous speech comprehension. *Journal of Cognitive Neuroscience*, *32*, 155–166. https://doi.org/10.1162/jocn_a_01467, PubMed: 31479349

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*, 420–422. https://doi.org/10.1038/nn1872, PubMed: 17351633

Yoshiura, T., Ueno, S., Iramina, K., & Masuda, K. (1995). Source localization of middle latency auditory evoked magnetic fields. *Brain Research*, *703*, 139–144. https://doi.org/10.1016/0006-8993(95)01075-0, PubMed: 8719625