# High Frequency Cortical Processing of Continuous Speech in Younger and Older Listeners

Joshua P. Kulasingham[1], Christian Brodbeck[2], Alessandro Presacco[2], Stefanie E. Kuchinsky[3], Samira Anderson[4] & Jonathan Z. Simon[1,2,5]

[1]Department of Electrical and Computer Engineering, University of Maryland, College Park, Maryland, USA

[2]Institute for Systems Research, University of Maryland, College Park, Maryland, USA

[3]Audiology and Speech Pathology Center, Walter Reed National Military Medical Center, Bethesda, Maryland, USA

[4]Department of Hearing and Speech Sciences, University of Maryland, College Park, Maryland, USA

[5]Department of Biology, University of Maryland, College Park, Maryland, USA

Author Contact Details:

Joshua P. Kulasingham: joshuapk@umd.edu

Christian Brodbeck: christianbrodbeck@me.com

Alessandro Presacco: apresacc@umd.edu

Stefanie E. Kuchinsky: skuchins@umd.edu

Samira Anderson: sander22@umd.edu

Jonathan Z. Simon: jzsimon@umd.edu

Corresponding Author:

Joshua P. Kulasingham: joshuapk@umd.edu

## Abstract

Neural processing along the ascending auditory pathway is often associated with a progressive reduction in characteristic processing rates. For instance, the well-known frequency-following response (FFR) of the auditory midbrain, as measured with electroencephalography (EEG), is dominated by frequencies from ~100 Hz to several hundred Hz, phase-locking to the stimulus waveform at those frequencies. In contrast, cortical responses, whether measured by EEG or magnetoencephalography (MEG), are typically characterized by frequencies of a few Hz to a few tens of Hz, time-locking to acoustic envelope features. In this study we investigated a crossover, cortically generated responses time-locked to continuous speech features at FFR-like rates. Using MEG, we analyzed high-frequency responses (70-300 Hz) to continuous speech using neural source-localized reverse correlation and its corresponding temporal response functions (TRFs). Continuous speech stimuli were presented to 40 subjects (17 younger, 23 older adults) with clinically normal hearing and their MEG responses were analyzed in the 70-300 Hz band. Consistent with the insensitivity of MEG to many subcortical structures, the spatiotemporal profile of these response components indicated a purely cortical origin with ~40 ms peak latency and a right hemisphere bias. TRF analysis was performed using two separate aspects of the speech stimuli: a) the 70-300 Hz band of the speech waveform itself, and b) the 70-300 Hz temporal modulations in the high frequency envelope (300-4000 Hz) of the speech stimulus. The response was dominantly driven by the high frequency envelope, with a much weaker contribution from the waveform (carrier) itself. Age-related differences were also analyzed to investigate a reversal previously seen along the ascending auditory pathway, whereby older listeners show weaker midbrain FFR responses than younger listeners, but, paradoxically, have stronger cortical low frequency responses. In contrast to both these earlier results, this study does not find clear age-related differences in high frequency cortical responses. Finally, these results suggest that EEG high (FFR-like) frequency responses have distinct and separable contributions from both subcortical and cortical sources. Cortical responses at FFR-like frequencies share some properties with midbrain responses at the same frequencies and with cortical responses at much lower frequencies.

## Keywords

auditory spectrogram, spectrotemporal, volume source localization, aging, high gamma, speech encoding,

## Highlights

- Cortical MEG responses time-lock at 80-90 Hz to continuous speech.
- Responses are predominantly time-locked to the high frequency envelope of speech.
- Response strength and latency are similar for younger and older subjects.

## 1. Introduction

The human auditory system time-locks to acoustic features of complex sounds, such as speech, as it extracts and encodes relevant information. The characteristic frequency of such time-locked activity is generally thought to decrease along the ascending auditory pathway. For example, subcortical high frequency activity may encode the pitch information of voiced speech (Krishnan et al., 2004), while cortical low frequency activity, which time-locks to the slowly varying envelope of speech, may represent higher level features of language such as phoneme and word boundaries (Brodbeck et al., 2018a). Prior research has also shown that age-related auditory temporal processing deficits result in differences in both subcortical and cortical processing for older and younger listeners (Anderson et al., 2012; Presacco et al., 2016a, 2016b). These effects have been investigated in human subjects using the complementary non-invasive neural recording techniques of electroencephalography (EEG) and magnetoencephalography (MEG).

The well-known frequency following response (FFR) is one such phase-locked response (Kraus et al., 2017), most commonly measured using EEG, and is believed to originate predominantly from the auditory midbrain (Bidelman, 2015; Smith et al., 1975). The FFR measures the phase-locked response to the fast steady state oscillation of a stimulus, such as a repeated speech syllable. The FFR provides insight into the peripheral representation of speech and is a useful tool for investigating temporal processing deficits (Basu et al., 2010; Hornickel et al., 2012; Kraus et al., 2017). In addition, the FFR may be used to investigate the robustness of speech representations in noise or a dual stream paradigm (Yellamsetty and Bidelman, 2019). Much FFR analysis is done

by averaging responses to stimuli of opposite polarity, in which case the response is actually to the envelope of the stimulus. However the FFR to the stimulus spectrum itself (also called fine-structure) can also be analyzed, e.g., by subtracting responses to stimuli of opposite polarity (Aiken and Picton, 2008; Hornickel et al., 2012). These studies show that high frequency neural responses are sensitive to both the high frequency envelope of the stimulus, and to its carrier.

The neural origins of the FFR have historically been thought to be mainly subcortical areas such as the inferior colliculus (Smith et al., 1975). But recent studies with MEG and EEG have shown that the FFR is not purely generated by subcortical areas, but has contributions from the auditory cortex as well (Bidelman, 2018; Coffey et al., 2017b, 2017a, 2016; Hartmann and Weisz, 2019; Puschmann et al., 2019). The dominantly cortical role in the MEG response follows from MEG's reduced sensitivity to deep structures such as the auditory midbrain (Baillet, 2017).

However, the repeated speech syllables commonly used to generate the FFR cannot capture the complexities of natural continuous speech. To understand how the brain represents speech in naturalistic environments, cortical low frequency responses to continuous speech have been widely studied (Peelle et al., 2013). The MEG and EEG response to continuous speech can be represented using Temporal Response Functions (TRFs) (Ding and Simon, 2012; Lalor et al., 2009) which are linear estimates of time-locked responses to time varying features of the auditory stimulus. The conventional low-frequency TRF time-locks to the slow (below 10 Hz) envelope of continuous speech, though the spectrotemporal fine structure of speech can also modulate these cortical low frequency responses (Ding et al., 2014; Ding and Simon, 2012).

Recently, short latency subcortical EEG responses to continuous speech have been found using TRF analysis (Maddox and Lee, 2018), proving that it is possible to detect midbrain responses to continuous speech. High frequency early latency responses that phase lock to the fundamental frequency of speech have also been found to be modulated by attention (Forte et al., 2017). One study has also found cortical high frequency MEG responses to speech stimuli, with latencies near 30 ms, that are time-locked to the ~100 Hz variations in the envelope of the even higher frequency speech spectrum (up to 2 kHz) (Hertrich et al., 2012). This can be seen as analogous to traditional FFR which follows the envelope of the fundamental frequency. As discussed above, EEG FFR has

also been found for the spectral carrier. Whether auditory cortex time-locks to the carrier as well as to the high frequency envelope (HFE) of continuous speech remains unclear.

Further complicating our understanding of the contributions of subcortical and cortical sources to the MEG response is the impact of age-related changes in the auditory pathway (Peelle and Wingfield, 2016). The temporal processing of speech can degrade with age, especially in noisy conditions (Gordon-Salant et al., 2006; He et al., 2008; Hopkins and Moore, 2011). Age-related differences have been found in both the EEG FFR and the MEG low frequency TRF to speech. Older adults have weaker, delayed FFRs with lower phase coherence when compared with younger adults (Anderson et al., 2012; Presacco et al., 2015; Zan et al., 2019). Possible causes include age-related inhibition-excitation imbalance (Caspary et al., 2008) resulting in a loss of temporal precision (Anderson et al., 2012). In surprising contrast, older adults' cortex exhibits *exaggerated* low frequency responses (Brodbeck et al., 2018b), even to the point of allowing better stimulus reconstruction via these low frequency cortical responses than in younger adults (Presacco et al., 2016a, 2016b). Several possible explanations, not necessarily exclusive, have been advanced to account for this surprising result, including decrease in inhibition, recruitment of additional brain regions and central compensatory mechanisms (Brodbeck et al., 2018b; Decruy et al., 2019). The fact that fast midbrain responses reduce with age while slow cortical responses are enhanced might indeed be due to anatomical and physiological differences between midbrain and cortex, but a fair comparison is complicated by the fact that the responses occur at vastly different frequencies. Hence it is entirely unknown whether high frequency cortical responses would show age-related reduction or enhancement.

In this study we focused on three main questions. Firstly, do high frequency time-locked MEG responses to continuous speech in the human auditory system originate from cortical or subcortical areas? Secondly, are these responses time-locked to the carrier or to the HFE of the speech stimulus? Thirdly, are there any age-related differences in these responses, and if so, do they show age-related decrease, like the EEG FFR, or show age-related increase, like the low frequency TRF? To this end, we analyzed source localized MEG recordings of younger and older subjects listening to continuous speech (narration by a male speaker). We estimated TRF models, for both the HFE

and the carrier, and used neural volume source localization to investigate the nature and origin of these high frequency responses.

# 2. Methods

### 2.1 Experiment dataset

The experimental dataset used for this study is described in detail in Presacco et al., 2016a, 2016b, supplemented with eight additional older adults with clinically normal hearing. The combined dataset consisted of MEG responses recorded from 17 younger adults (age 18-27, mean 22.3, 3 male) and 23 older adults (age 61-78, mean 67.2, 8 male) with clinically normal hearing while they listened to 60 second portions of an audiobook recording of The Legend of Sleepy Hollow by Washington Irving (https://librivox.org/the-legend-of-sleepy-hollow-by-washington-irving). All participants gave informed consent and were paid for their time. Experimental procedures were reviewed and approved the Institutional Review Board of the University of Maryland. The audio was delivered diotically through 50 Ω sound tubing (E-A-RTONE 3A) attached to E-A-RLINK foam earphones inserted into the ear canal at ~70 dB sound pressure level via a sound system with flat transfer function from 40 to 3000 Hz. The conditions analyzed in this study consist of two passages of 60 seconds duration presented in quiet (i.e., solo speaker), each of which was repeated three times, for a total of six minutes of MEG data per subject. Subjects were asked beforehand to silently count the number of occurrences of a particular word and report it to the experimenter at the conclusion of each trial, in order to encourage attention to the auditory stimuli. Handedness of the participants was assessed with the Edinburgh handedness scale (Oldfield, 1971), which can range from –1 (complete left-dominance) to 1 (complete right-dominance). To exclude lateralization bias due to handedness, all analyses were rerun excluding the 9 subjects scoring below 0.5. The only change in the results was a reduction in the number of neural sources showing right hemispheric dominance in younger subjects (discussed below).

### 2.2 MEG data collection and preprocessing

MEG data was recorded from a 157 axial gradiometer whole head KIT MEG system while subjects were resting in supine position in a magnetically shielded room. The data was recorded at a sampling rate of 1 kHz with an online 200 Hz low pass filter with a wide transition band, and a 60

Hz notch filter. Data was preprocessed in MATLAB by first automatically excluding saturating channels and then applying time-shift principal component analysis (de Cheveigné and Simon, 2007) to remove external noise, and sensor noise suppression (de Cheveigné and Simon, 2008) to suppress channel artifacts. On average, two MEG channels were excluded during these stages. All subsequent analysis was performed in mne-python 0.17.0 (Gramfort, 2013; Gramfort et al., 2014) and eelbrain 0.30 (Brodbeck et al., 2019). The MEG data was filtered in the band 70-300 Hz (high gamma band) using an FIR filter, and six 60 second epochs during which the stimulus was presented were extracted for analysis.

**2.3 Stimulus Representation**

As discussed above, prior work on the FFR has shown that time-locked neural responses are sensitive to both the carrier and the envelope of an auditory stimulus. Similarly, time-locked responses to speech in the 100 Hz range may be driven either by ~100 Hz modulations in the waveform (carrier), or by ~100 Hz modulations in the envelope of even higher frequencies. Accordingly, two distinct representations of the speech stimulus were used as predictors for the TRF model (see Fig. 1A). For the former case, the carrier predictor was constructed by downsampling the speech waveform to 1 kHz and bandpass filtering from 70-300 Hz. This carrier predictor captures the 70-300 Hz modulations in the speech waveform itself. For the latter case, the HFE predictor was constructed from the ~100 Hz modulations in the envelope of the highpassed stimulus waveform (envelopes are only well-defined when they modulate carriers of much higher frequencies than those of the modulations themselves; Rosen, 1992). Specifically, first the speech was transformed into an auditory spectrogram representation by computing the acoustic energy in the speech waveform for each frequency bin in the range 300-4000 Hz at millisecond resolution using a model of the auditory periphery (Yang et al., 1992). This auditory spectrogram is a 2-dimensional matrix representation of the acoustic envelope over time for different frequency bins. Each frequency bin component of this spectrogram was then filtered using the same 70-300 Hz bandpass filter, producing a 70-300 Hz band limited envelope for each bin. Finally the resulting 2-dimensional matrix was averaged across frequency bins to provide a single envelope, resulting in the HFE predictor. Thus, this predictor captures the 70-300 Hz modulations in the 300-4000 Hz envelope of the speech waveform. These two predictors were used for all further TRF analysis.

The 70-300 Hz bandpass filter was formed using the default FIR filter in mne-python with an upper and lower transition bandwidth of 10 Hz, at 1 kHz sampling frequency, but applied twice in a forward fashion to the data. This resulted in a combined filter of length 661 with a phase delay of 330 ms. Other bandpass filters were also employed as alternatives, including IIR minimum-phase-delay Bessel filters (results not shown); no results depended critically on the filters used.

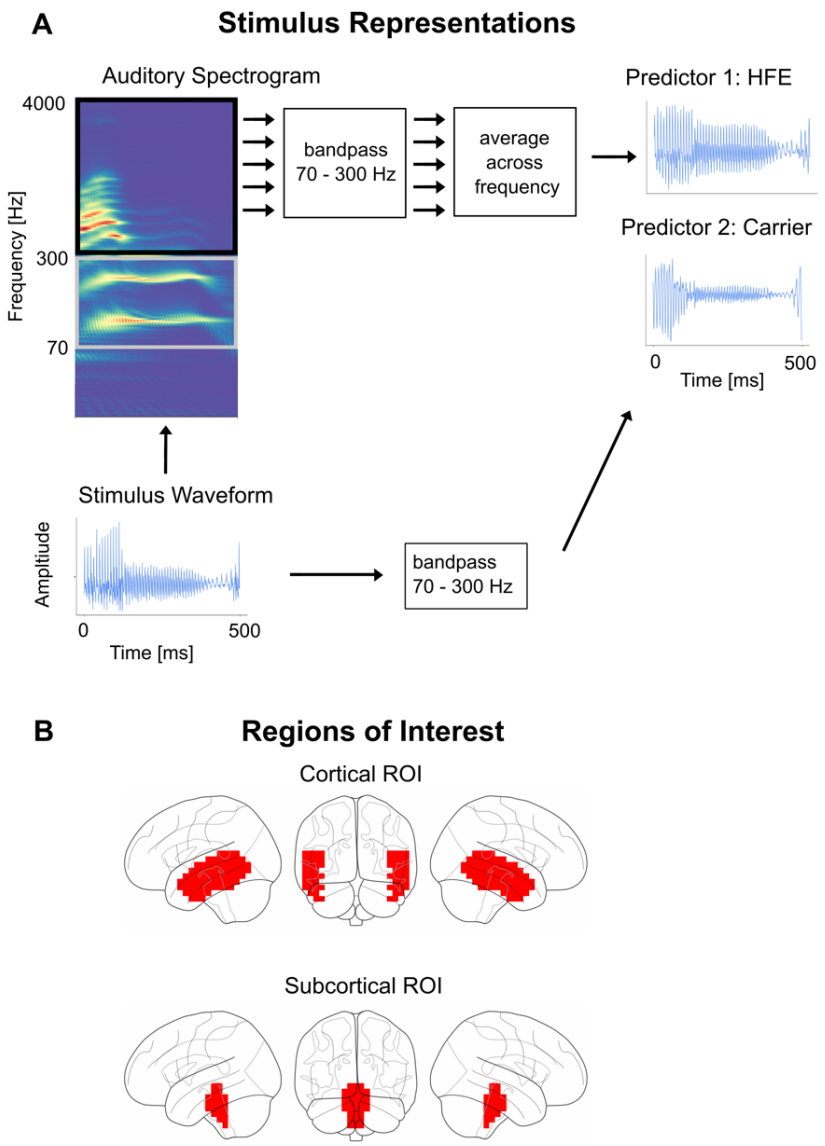


**Fig. 1.** Stimulus Representations and ROIs. **A.** The stimulus waveform for a representative 500 ms speech segment is shown along with its auditory spectrogram and the two predictors: carrier and HFE. The predictors are correlated (Pearson's $r$ = -0.16) but have noticeably distinct waveforms. **B.** Regions of Interest. The volume source space voxels for the cortical and subcortical ROIs are shown.

### 2.4 Neural Source localization

Before each MEG recording, the head shape of each subject was digitized using a Polhemus 3SPACE FASTRAK system, after which five marker coils were attached. The marker coil locations were measured while the subject's head was positioned in the MEG scanner before and after the experiment, in order to determine the position of the head with respect to the MEG sensors. Source localization was performed using the mne-python software package. The marker coil locations and the digitized head shape were used to coregister the template Freesurfer 'fsaverage' brain (Fischl, 2012) using rotation, translation and uniform scaling. A volume source space was formed by dividing the brain volume into a grid of 7 mm sized voxels. This source space was used to compute an inverse operator using minimum norm estimation (MNE) (Gramfort et al., 2014) with a noise covariance matrix estimated from empty room data. This method results in a 3-dimensional current dipole vector with magnitude and direction at each voxel. The *'aparc+aseg'* parcellation was used to define cortical and subcortical regions of interest (ROIs) (see Fig. 1B). The cortical ROI consisted of voxels in the gray and white matter of the brain that were closest to the temporal lobe parcellations. The subcortical ROI was selected to consist of voxels that were closest to the '*Brain-Stem*' parcellation. All brain plots show the maximum intensity projection of the voxels onto a 2-dimensional plane, with an overlaid average brain outline (implemented in eelbrain). Minimum norm estimation in volume source space may lead to spatial leakage from the true neural source to neighboring voxels. In order to characterize this artifactual spatial leakage, current dipoles in Heschl's gyrus were simulated using the surface source space, projected into sensor space, and then projected into volume source space. Additionally, a separate cortical surface source space model was also used; results obtained using this method were not qualitatively different than those of the volume space model (see Appendix).

### 2.5 Temporal Response Functions

The simplest linear model used to estimate the TRF is given by

$$y_t = \sum_d (\tau_d x_{t-d}) + n_t \tag{1}$$

where $y_t$ is the response at a neural source for time $t$, $x_{t-d}$ is the time shifted predictor with a time lag of $d$, $\tau_d$ is the TRF value at lag $d$ and $n_t$ is the residual noise. The TRF is the time-dependent

weight, of a linear combination of current and past samples of the predictor, that best predicts the current neural response at that neural source (Lalor et al., 2009). Hence the TRF can also be interpreted as the average time-locked response to a predictor impulse. In this work, a TRF model with two predictors, HFE and carrier, was used.

$$y_t = \sum_d (\tau_{e,d} e_{t-d} + \tau_{c,d} c_{t-d}) + n_t \quad (2)$$

Where $e_{t-d}$ is the delayed HFE (envelope) predictor and $\tau_{e,d}$ the corresponding HFE TRF, $c_{t-d}$ is the delayed carrier predictor and $\tau_{c,d}$ the corresponding carrier TRF. In this model, the two predictors compete against each other to explain response variance, which results in larger TRFs for predictors that contribute more to the neural response. TRF estimation, for lags from -40 to 200 ms, was performed with the boosting algorithm and early stopping based on cross validation (David et al., 2007) as implemented in eelbrain. The boosting algorithm may result in overly sparse TRFs, and hence an overlapping basis of 4 ms width Hamming windows (spaced 1 ms apart) was used in order to generate smoothly varying responses. For the volume source space, the neural response at each voxel is a 3-dimensional current vector. Accordingly, for each voxel, a TRF vector was computed using the boosting algorithm and was used to predict the neural response vector. For each voxel, the prediction accuracy was assessed through the average dot product between the normalized predicted and true response, which varies between zero and one in analogy to the Pearson correlation coefficient.

### 2.6 Statistical tests

Statistical tests were performed across subjects by comparing the TRF model to a noise model. The predictor was circularly shifted in time and TRFs were estimated using this time-shifted predictor as noise models (Brodbeck et al., 2018a, 2018b). This preserves the local temporal structure of the predictor while removing the temporal relationship between the predictor and the response. Circular shifts of duration 15, 30 and 45 seconds were used to form three noise models. For each voxel, the prediction accuracies of the true model were compared to the average prediction accuracies of the three noise models as a measure of model fit.

To account for variability in neural source locations due to mapping the responses of individual subjects onto the 'fsaverage' brain, these coefficients were spatially smoothed using a Gaussian

window with 5 mm standard deviation. Nonparametric permutation tests (Nichols and Holmes, 2002) and Threshold Free Cluster Enhancement (TFCE) (Smith and Nichols, 2009) were used to control for multiple comparisons. This method, as outlined in full in Brodbeck et al. (2018c, 2018a), is implemented in eelbrain, and is briefly recounted here. Firstly, a paired sample *t*-value was evaluated for each neural source, across subjects, from the difference of the prediction accuracies of the true model and the average of the three noise models after rescaling using Fisher's z-transform. Then the TFCE algorithm was applied to those *t*-values, which enhanced continuous clusters of large values, based on the assumption that significant neural activity would have a larger spatial spread than spurious noise peaks. This procedure was repeated 10,000 times with random permutations of the data where the labels of the condition were flipped on a randomly selected subset of the subjects. A distribution of TFCE values was formed using the maximum TFCE value of each permutation to correct for multiple comparisons across the brain volume. Any value of the original TFCE map that exceeded the 95th percentile of the distribution was considered as significant at the 5% significance level.

The TRF itself was also tested for significance against the noise model in a similar manner. In the volume source space, a TRF that consists of a 3-dimensional vector which varies with time was estimated for each voxel, representing the estimated current dipole amplitude and direction at that voxel. The amplitudes of these TRF vectors for the true model and the average noise model were used for significance testing. The TRF amplitudes were spatially smoothed using the same Gaussian window before performing the tests. A one-tailed test was done with paired sample *t*-values and TFCE, and the procedure is identical to that outlined previously, with the added dimension of time (Brodbeck et al., 2018a).

Lateralization tests were performed to check for hemispheric asymmetry. The volume source space estimates in the cortical ROI were separated into left and right hemispheres and, as above, the prediction accuracies were spatially smoothed with the same Gaussian window. The prediction accuracies of the average noise model were subtracted from that of the true model and paired sample *t*-values with TFCE in a two-tailed test were used to test for significant differences between each of the corresponding left and right voxels.

Age-related differences were assessed between the younger and older groups. The difference of prediction accuracies between the true TRF model and the average of the noise TRF models were used to form independent sample *t*-values for each source across age groups after which a two-tailed test was performed with TFCE. Significant differences in lateralization across age groups were assessed by subtracting the prediction accuracies of the left hemisphere from the right hemisphere and then conducting independent samples tests across age groups as described above. The peak latency of the TRFs was also tested for significant differences across age groups. The latency of the maximum value of the norm of the TRF vectors in the time range of significant responses (20-70 ms) was used to test for peak latency differences across age groups using a two-tailed test with independent sample *t*-values and TFCE.

In addition, to investigate possible evidence in favor of the null hypothesis (i.e., in support of the hypothesis that there are no age-related differences), Bayes factor analysis was performed. The Bayes factor $BF_{01}$ is the ratio of the likelihood of the observations given the null and the alternative hypothesis; $BF_{01} = \frac{P(D|H_0)}{P(D|H_1)}$ where $D$ is the set of observations, and $H_0, H_1$ are the null and alternative hypotheses. Assuming that $H_0$ and $H_1$ are equiprobable ($P(H_0) = P(H_1)$), the posterior odds reduce to the Bayes factor:

$$\frac{P(H_0|D)}{P(H_1|D)} = \frac{P(D|H_0)P(H_0)}{P(D|H_1)P(H_1)} = BF_{01}\frac{P(H_0)}{P(H_1)} = BF_{01} \quad (3)$$

For example, $BF_{01} = 10$ would indicate that, given the data, the null hypothesis is 10 times more probable than the alternative, whereas $BF_{01} = 0.1$ would indicate that the alternative is 10 times more probable than the null. $BF_{01} > 3$ is considered to indicate moderate support for the null hypothesis (Jarosz and Wiley, 2014). The model prediction accuracy for the competing model was averaged across sources for each subject and a two-tailed independent sample *t*-value was calculated for the older and younger subjects. The method described in Rouder et al., 2009 was used to transform this t-value into a Bayes factor. To further test for significant differences by age across both low frequency and high frequency responses, two further models were analyzed; a low frequency TRF and a high frequency TRF using surface source space. An ANOVA was performed on the prediction accuracies of these two models with factors TRF frequency (high or low) and age (young or old) (methods and results in Appendix).

# 3. Results

### 3.1 Cortical origins of high frequency responses to continuous speech

High frequency responses were used to estimate per-voxel TRFs in volume source space for the two ROIs: the temporal lobes, and the brainstem (plus its surrounding volume). The prediction accuracies of high-frequency responses (mean = 0.02, std = 0.005) were much smaller (factor of 3) than those resulting from low frequency cortical TRFs (Brodbeck et al., 2018a), indicating that these responses are weaker than slow cortical responses. This is not surprising, as the spectrum of the MEG response decays with frequency. Noise floor models, used to test for significant responses, generated corresponding noise model prediction accuracies (mean = 0.0138, std = 0.0016). For each voxel, a one-tailed test with paired sample *t*-values and TFCE (to account for multiple comparisons) was used to test for significant increases in the prediction accuracies of the true model against the noise model across subjects. Almost all of the voxels showed a significant increase in prediction accuracy (younger subjects $t_{max} = 6.95$, $p < 0.001$; older subjects $t_{max} = 6.35$, $p < 0.001$; see Fig. 2A). The disproportionate success of the extent of this result is not unexpected, however, due to the large spatial spread of MNE volume source space estimates. The prediction accuracy over the noise model within the cortical ROI was significantly larger than that within the subcortical ROI for younger subjects (one-tailed paired sample *t*-test; younger subjects $t = 2.73$, $p = 0.007$; older subjects $t = 1.47$, $p = 0.078$; difference across age not significant).

Lateralization differences were tested using the prediction accuracy at each voxel. The prediction accuracy of the average noise model was subtracted from that of the true model and a two-tailed test with paired sample *t*-values and TFCE was performed for significant differences in the left and right hemispheres. The tests revealed significantly higher prediction accuracies for younger subjects in the right hemisphere than in the left ($t_{max} = 6.28$, $p < 0.01$), but only for a subset of voxels (24%) in the temporal lobe closest to auditory areas (see Fig. 2B). No significant differences in lateralization were seen for older subjects ($t_{min} = -2.01$, $t_{max} = 3.34$, $p > 0.097$), nor was lateralization significantly different across age groups (independent samples test; $t_{min} = -2.77$, $t_{max} = 1.7$, $p > 0.33$). When the analysis was constrained to only right-handed subjects (13 younger, 18 older; see Methods for details), the only resulting change was that a lower number of voxels were significantly right lateralized in younger subjects.

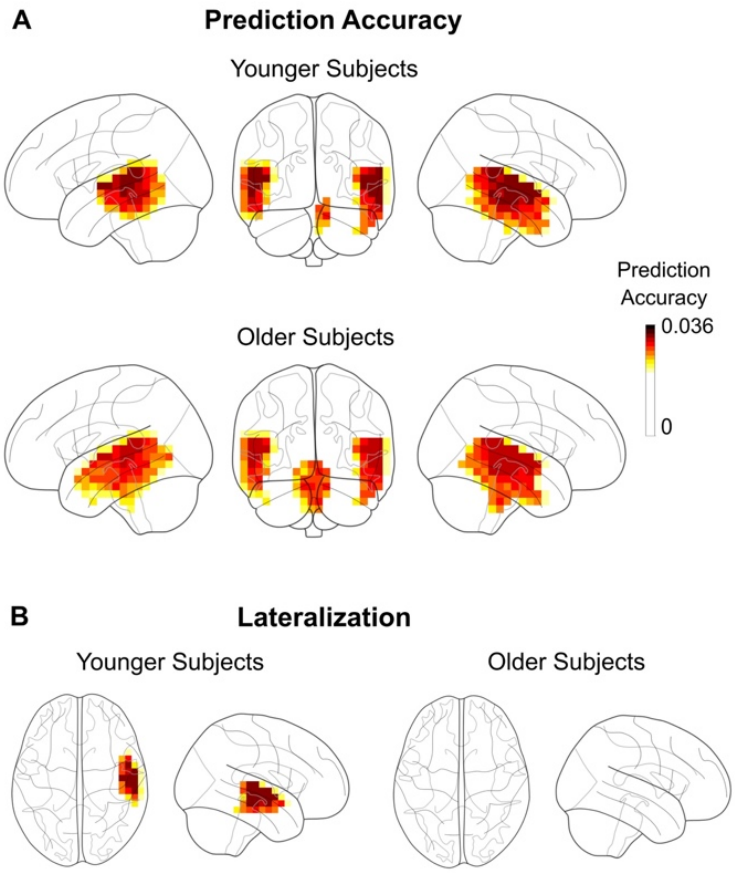


**Fig. 2.** Prediction Accuracy of Volume Source Localized TRFs. **A.** Prediction accuracy using the TRF model for each voxel in the volume source space ROIs averaged across subjects. Only ROI voxels for which model prediction accuracy significantly increased over the noise model are plotted ($p < 0.05$). The prediction accuracy is significantly larger in cortical areas than in subcortical areas for the younger subjects. Plots are of the maximum intensity projection, with an overlay of the brain. **B.** An area in the right hemisphere near the auditory cortex is significantly more predictive than the left hemisphere, but only in the younger subjects. When taking into account expected MEG volume source localization leakage, these results are consistent with the response originating solely from cortical areas and with a right hemispheric bias.

The TRFs at each source voxel are represented by a 3-dimensional current vector that varies over the time lags. Hence for each voxel and time lag, the amplitude of the TRF vector for the true model was tested for significance against the average of the noise models across subjects using a one-tailed test with paired sample *t*-values and TFCE. The TRFs for the HFE predictor in the cortical ROI were significant (younger $t_{max} = 4.89$, $p < 0.001$; older $t_{max} = 4.54$, $p < 0.001$) starting

at a time lag of 24 ms, and ending at 62 ms, with an average peak latency of 40 ms (see Fig. 3A). The subcortical ROI was also analyzed in a similar manner and the TRF showed significance in a much smaller time range of 36-45 ms (younger $t_{max} = 3.68$, $p < 0.01$; older $t_{max} = 3.43$, $p < 0.01$) (see Fig 3B). Although the average TRF of the younger subjects has larger amplitude than the older subjects, this difference is not significant (cortical ROI $t_{max} = 3.81$, $t_{min} = -3.87$, $p > 0.56$; subcortical ROI $t_{max} = 3.71$, $t_{min} = -3.24$, $p > 0.56$), and is mainly driven by two individuals in the younger age group. The TRF responses oscillate at a frequency of 80-90 Hz (see below for a more detailed spectral analysis). The amplitude of these TRFs was significantly larger in the cortical ROI than the subcortical ROI (two-tailed test with paired sample t-values on the maximum amplitudes of the TRFs across subjects: younger $t = 5.47$, $p < 0.001$; older $t = 6.21$, $p < 0.001$). Since the subcortical TRFs also have a similar latency and shape to the cortical TRFs, and because a latency of 40 ms is late for a subcortical response, these subcortical TRFs are consistent with artifactual leakage from the cortical TRFs due to the spatial spread of MNE source localization. Simulated volume source estimates for current dipoles originating only in Heschl's gyrus generated a spatial distribution of TRF directions consistent with the experimental data (see Fig. 3), i.e. the spatial spread of MNE localized cortical responses resulted in apparent TRF vectors even in the subcortical ROI. These results seem to indicate that the response originates predominantly from cortical regions.

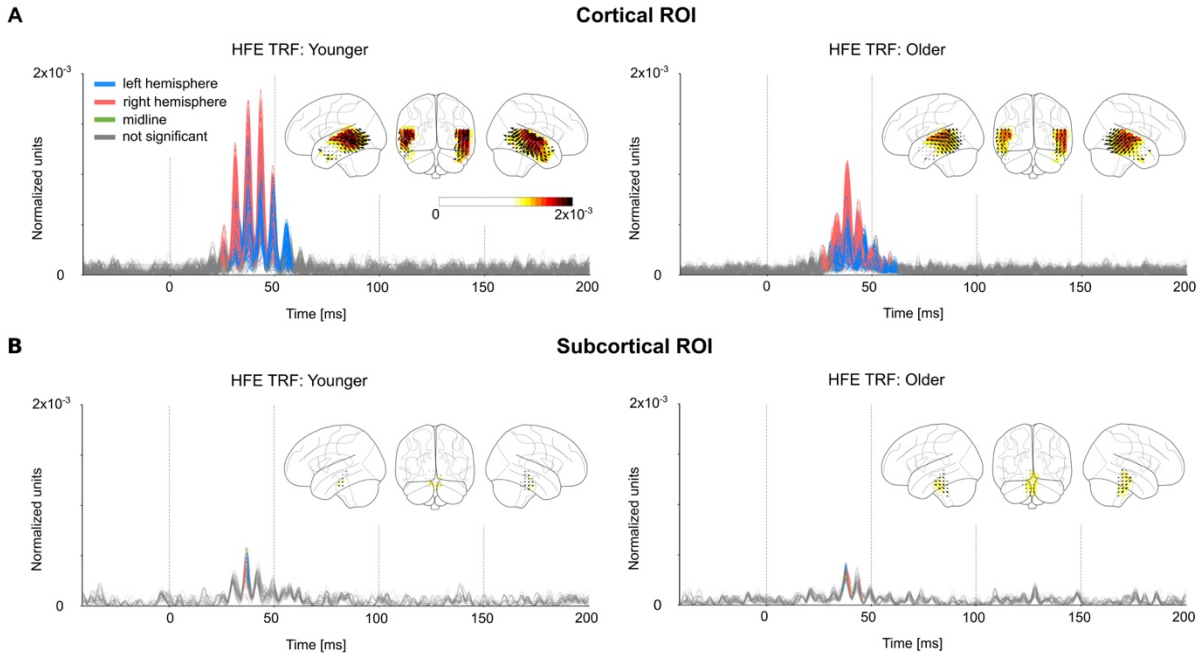

**Fig. 3.** Volume Source Localized HFE TRFs. The amplitude of the TRF vectors for the HFE predictor averaged across subjects. Each individual line in the plots is a TRF for a particular voxel. Color encodes the time points when the TRF showed a significant increase in amplitude over noise (blue: significant left hemisphere, red: significant right hemisphere, green: significant midline, gray: not significant). The TRF shows a clear response with a peak latency of 40 ms. The distribution of TRF vectors in the brain at each voxel at the time with the maximum response are plotted as an inset for each TRF, with color representing response strength and the arrows representing the TRF directions. The color bar represents the response strength for all 4 brain insets. The response oscillates around a frequency of 80-90 Hz and is much stronger in the cortical ROI compared to the subcortical ROI. Note that since only the TRF amplitude is shown, and not signed current values, signal troughs and peaks both appear as peaks. Although the younger subjects have a stronger average response in the cortical ROI, it is not significantly larger than the older subjects ($p = 0.56$). The latency and amplitude of the response suggests a predominantly cortical origin.

### 3.2 Responses to the HFE and the carrier

Next, the neural response to the carrier was compared with that to the HFE. The carrier TRF was also tested for significance using a corresponding noise model (as employed above). The carrier TRF showed weak responses that were only significant for older subjects in the cortical ROI between 31-51 ms ($t_{max} = 3.86$, $p = 0.0014$) (see Fig. 4A,B). Two-tailed paired sample *t*-values and TFCE were used to test for a significant increase of the norm of the HFE TRF when compared to the carrier TRF in a time window of 20-70 ms in the cortical ROI (see Fig. 5A). This test was significant for both younger ($t_{max} = 6.97$, $p < 0.001$) and older ($t_{max} = 4.17$, $p < 0.006$) subjects. However, this test did not find a significant increase in the HFE TRF over the carrier TRF in the subcortical ROI for either younger ($t_{max} = 1.43$, $p > 0.46$) or older subjects ($t_{max} = 2.66$, $p > 0.13$). Since the TRF analysis allows both stimulus predictors to directly compete for explaining response variance, the results strongly indicate that the response is primarily due to the HFE over the carrier.

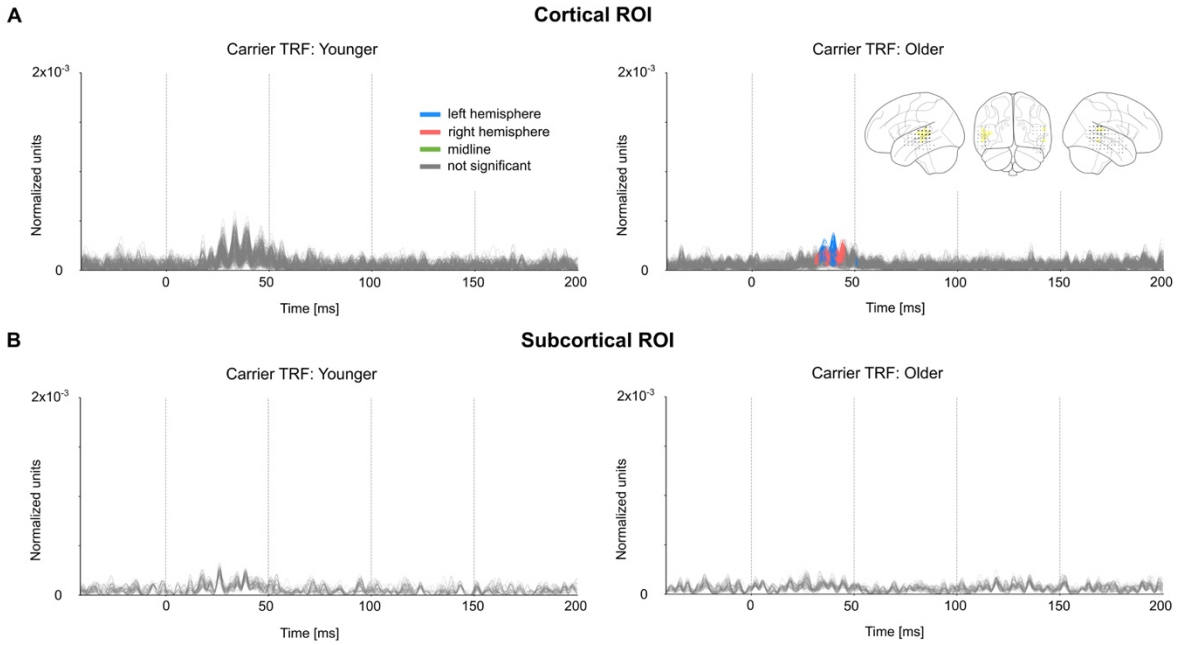


**Fig. 4.** Volume Source Localized Carrier TRFs. The amplitude of the TRF vectors for the carrier predictor averaged across subjects, analogous to Fig. 3. For comparison, the axis limits and color scale are identical to that in Fig. 3. The TRF shows a weak response, with a peak latency of 40 ms, that is only significant in the cortical ROI for older subjects. Comparison with Fig. 3 suggests that the high frequency response is dominated by the HFE over the carrier.

To further understand the contributions of these predictors to the TRF, the frequency spectrum of the TRFs and the predictors were compared (see Fig. 5B). The frequency spectrum of the average TRFs showed a broad peak centered near 80-90 Hz for both predictors and both age groups. In contrast, the spectral peak of the predictor variables was near 110-120 Hz for the carrier, and near 70-75 Hz for the HFE. Since the TRF peak frequency did not match the peak power in either of the predictors, this suggests that the TRF oscillatory rate is not directly determined by the stimulus spectrum.

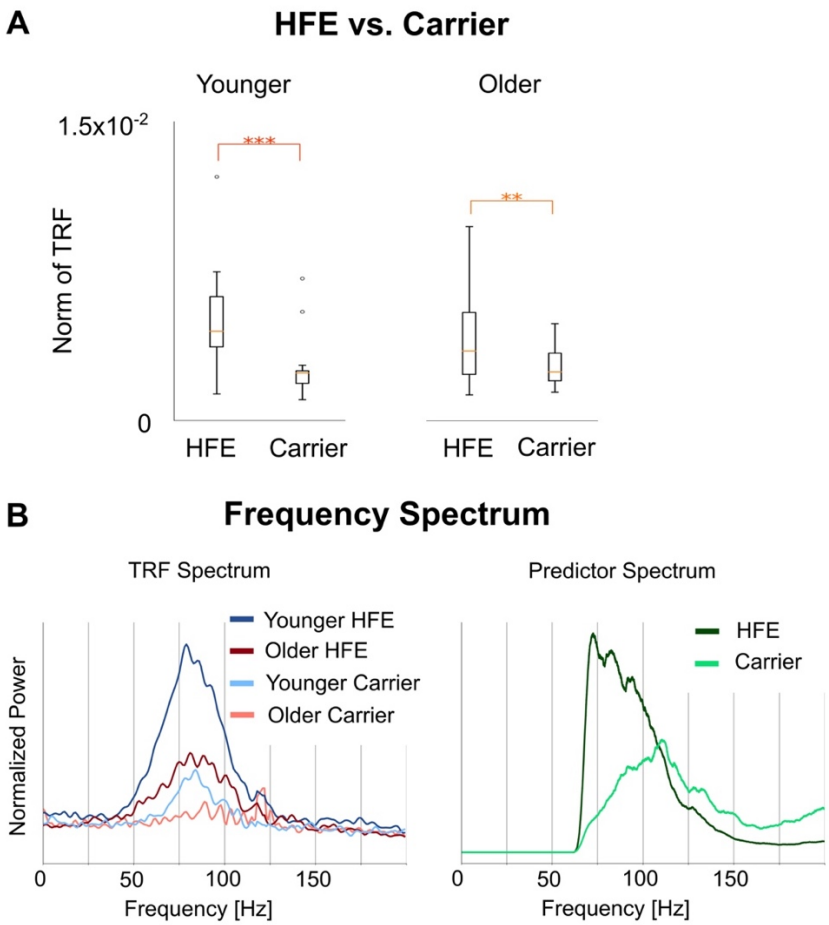


**Fig. 5.** Comparison of Responses to the HFE and to the Carrier. **A**. The norm of the TRF between 20 ms and 70 ms was larger in the HFE TRF than the carrier TRF for both age groups (*** $p < 0.001$, ** $p < 0.01$). **B**. The frequency spectrum of the TRF reveals that the oscillation has a broad peak around 80-90 Hz. However, the predictors have peaks at 70-75 Hz (for the HFE) and 110-120 Hz (for the carrier). Although the peak of the HFE spectrum is near 70-75 Hz, there is still substantial power in the HFE predictor near 80-90 Hz as well, and similarly for the carrier spectrum. This suggests that the TRF response is not locking to the peak frequency of the predictor, but to an intrinsic frequency determined internally (neurally) rather than externally.

### 3.3 Age-related differences

Statistical tests were performed for age-related differences between older and younger subjects on both the prediction accuracy and the TRFs. Two tailed tests of prediction accuracy with independent sample *t*-values and TFCE indicated no significant difference (cortical ROI $t_{max}$ = 1.64, $t_{min}$ = -1.8, $p > 0.78$; subcortical ROI $t_{max}$ = -0.6, $t_{min}$ = -1.33, $p > 0.42$). Similarly, no voxels or time points were significantly different in either the HFE TRF (cortical ROI $t_{max}$ = 3.81, $t_{min}$ = -3.87, $p > 0.56$; subcortical ROI $t_{max}$ = 3.24, $t_{min}$ = -3.71, $p > 0.56$) or the carrier TRF (cortical ROI $t_{max}$ = 3.6, $t_{min}$ = -3.53, $p > 0.9$; subcortical ROI $t_{max}$ = 2.4, $t_{min}$ = -3.38, $p > 0.33$). In addition, the cortical ROI TRFs showed no significant differences across age groups in peak latency either (HFE TRF $t_{max}$ = 2.23, $t_{min}$ = -1.86, $p > 0.74$; carrier TRF $t_{max}$ = 1.72, $t_{min}$ = -2.02, $p > 0.97$). In the above analysis, it is unclear if the absence of significance is due to older and younger subjects having similar responses, or merely due to a lack of statistical power. Hence a Bayes factor analysis was performed on the prediction accuracy by averaging across sources per subject and computing an independent sample *t*-value across age groups. This resulted in a Bayes factor of $BF_{01}$ = 4.21, and hence provides moderate evidence for the null hypothesis, that there are no differences between the age groups. An additional analysis was performed using surface source space TRFs as described in Appendix. Both high and low frequency TRFs were computed in surface source space, and model prediction accuracy was assessed with an ANOVA with factors TRF frequency and age. The ANOVA showed a significant frequency × age interaction, suggesting that age related differences are indeed not consistent across high and low frequency responses (detailed results in Appendix), i.e. present at low frequencies but not so for high.

## 4. Discussion

In this study, we investigated high frequency time-locked responses to continuous speech measured using MEG. Volume source localized TRFs provided evidence that these responses originated from cortical areas and have a peak response latency of around 40 ms. The responses showed a significant right hemispheric asymmetry, especially in core auditory areas. These responses oscillate with a frequency of 80-90 Hz, despite the stimulus predictors having maximum power outside this frequency range. We also showed that the response is significantly stronger to

the HFE than the carrier. Surprisingly, there were no significant age-related differences in response amplitude, latency, localization or predictive power. In fact, Bayes factor analysis provided moderate support for the null hypothesis, that there is no difference in these responses between younger and older adults; this is in contrast to age-related differences seen in both the subcortical EEG FFR (younger > older) and the cortical low frequency TRF (older > younger).

### 4.1 MEG sensitivity to deep structures

A primary motivation of this study was to investigate the relative contributions of cortical and subcortical sources to high frequency time-locked activity. Because of the sensitivity to cortical sources as opposed to subcortical, MEG is an ideal tool to investigate if there are cortical contributions to the well-known high frequency subcortical responses. MEG is physically constrained to be less sensitive to deep structures, typically resulting in subcortical MEG responses being up to 100 times weaker than cortical responses at equivalent current strengths (Attal et al., 2007; Hillebrand and Barnes, 2002). Several source localization techniques have been proposed to correct for this inherent bias towards cortical sources (Dale et al., 2000; Krishnaswamy et al., 2017; Pascual-Marqui, 2002). Some studies were able to resolve MEG responses to the hippocampus (Cornwell et al., 2012), insula (Park and Tallon-Baudry, 2014), amygdala (Balderston et al., 2014; Cornwell et al., 2008; Dumas et al., 2013) and thalamus (Roux et al., 2013). Prior work has also been done using MEG for measuring brainstem responses (Coffey et al., 2016; Parkkonen et al., 2009). Hence, MEG is technically able to record responses from deep structures, although simultaneous cortical responses may overpower them (e.g., in FFR studies). However subcortical responses to continuous speech using TRFs should have a much shorter latency (< 20 ms) (Maddox and Lee, 2018) than the cortical latencies seen in this work. Therefore, the amplitude and latency of our volume source localized TRFs suggest that the high frequency response to speech is primarily cortical.

### 4.2 Cortical FFRs and high frequency TRFs

Cortical FFRs to repeated single speech syllables have been measured in MEG (Coffey et al., 2016) and EEG (Bidelman, 2018; Coffey et al., 2017b). Our work shows that cortical TRFs contain significant responses up to 64 ms, comparable to the long-lasting explanatory power of the auditory cortex ROI in Coffey et al., 2016. These TRFs are also predominantly from auditory cortex,

centered around Heschl's gyrus, and right lateralized similar to the MEG FFR (Coffey et al., 2016). However, Bidelman (2018) demonstrated that the contribution of cortical sources to the FFR as measured with EEG is much weaker than when measured with MEG, and rapidly decreases for stimuli with fundamental frequencies above 100 Hz. In addition, while subcortical FFR is measurable with EEG for stimulus frequencies up to 1000 Hz, there were no cortical contributions to the FFR above 150 Hz (Bidelman, 2018). This strengthens the idea that MEG and EEG are sensitive to different structures in the ascending auditory pathway. Hence MEG can be used in conjunction with EEG to explore the hierarchical processing of sound.

### 4.3 Comparison of responses to the envelope vs. the carrier

Previous studies have shown that auditory cortex responds to the slow varying envelope of speech and is also modulated by the spectrotemporal fine structure of the stimulus (Ding et al., 2014). Similarly, the subcortical FFR is typically analyzed by averaging across stimulus presentations of opposite polarity, which results in phase-locked responses to the envelope of the fundamental frequency. However, subcortical responses to the carrier have also been found by studies that distinguish between the traditional envelope FFR and the carrier FFR (which is analyzed by subtracting the responses to stimulus presentations of opposite polarity) (Aiken and Picton, 2008). Although the EEG FFR follows both the envelope and the carrier, our results indicate that time-locked high frequency cortical responses to continuous speech are predominantly driven by the HFE.

### 4.4 TRF oscillation frequency not strongly driven by stimulus spectrum

The TRF response oscillates with a peak frequency around 80-90 Hz. Cortical auditory phase locked responses to simple sounds have been measured using MEG (Coffey et al., 2016; Hertrich et al., 2004; Schoonhoven et al., 2003) at frequencies of up to 111 Hz. For continuous speech stimuli, such phase locked responses could reflect a cortical mechanism that represents complex speech features such as modulations in vowel formants, using fluctuations in the fundamental frequency domain of natural speech. However, the frequency of oscillation of the TRF at 80-90 Hz does not directly reflect the fundamental frequency of the speech at 110-120 Hz, nor spectral peak of the HFE predictor around 70-75 Hz. Thus, the 80-90 Hz oscillation could be driven by properties of the auditory system rather than stimulus features. The EEG FFR to triangle waves

can show oscillatory peaks at 80-90 Hz (Tichko and Skoe, 2017). Auditory steady state response (ASSR) studies also provide evidence for strong phase locked responses at harmonics of 40 Hz (Ross et al., 2000; Schoonhoven et al., 2003), which could be a contributing factor in the bias in TRF responses towards 80 Hz.

### 4.5 Right lateralization of responses

The TRF model prediction accuracy was significantly right lateralized in younger subjects. Lack of significant right lateralization among older subjects may not indicate an age related lateralization difference, but rather a lack of statistical power, since the lateralization was not significantly different across age groups. Stronger responses in the right auditory cortex have been observed for ASSR using EEG (Ross et al., 2005) and MEG (Hertrich et al., 2004) as well as in cortical FFRs using MEG (Coffey et al., 2016). This agrees with prior studies showing that right auditory cortex is specialized for tonal processing and pitch resolution (Cha et al., 2016; Hyde et al., 2008; Zatorre, 1988).

### 4.6 Absence of age-related differences

Temporal precision and synchronized activity decreases in the auditory system with age and is characterized by age related differences in both subcortical and cortical responses. Older adults have subcortical FFR responses with smaller amplitudes, longer latencies and reduced phase coherence, which could be due to an excitation-inhibition imbalance or a lack of neural synchrony (Hornickel et al., 2012). In contrast, MEG studies have revealed that older adults have larger low frequency cortical responses that result in better prediction accuracy for reverse correlation methods (Decruy et al., 2019; Presacco et al., 2016a, 2016b). This opposite effect could be due to cortical compensatory central mechanisms, lack of inhibition or recruitment of additional neural areas for redundant processing (Brodbeck et al., 2018b). Contrary to both these cases, we found no significant age-related differences in high frequency cortical responses. In fact, both Bayes factor analysis and an ANOVA with factors TRF frequency and age (see Appendix) provided evidence that this was not merely due to a lack of statistical power, suggesting that older and younger adults have similar high frequency responses. Although the TRFs averaged across subjects seemed larger for younger adults on visual inspection, this effect is largely driven by two outlier younger subjects with very large responses, and does not reflect the population as a whole.

The high frequency TRF reflects fine-grained time-locked neural activity, similar to the subcortical FFR, but arising from cortical areas. It is possible that the exaggerated responses in cortical areas and the lack of neural synchrony at high frequencies (found in subcortical FFRs) in older adults affect their high frequency responses in opposite directions and obscure otherwise detectable age-related differences.

### 4.7 Neural mechanisms for the MEG high frequency response

Given that MEG records the aggregate response over a large population of neurons, the specific origins of high frequency time-locked responses are not readily apparent. It is possible that the high frequency TRF reflects the effects of several processing stages along the auditory pathway, similar to the FFR (Coffey et al., 2019). Electrocorticography (ECoG) studies have seen cortical phase-locked activity at these high rates (Nourski et al., 2014; Steinschneider et al., 2013). However, cortical phase-locking at the individual neuron level drastically reduces with increasing frequency (Lu et al., 2001), and hence cortical neurons may not be the sole contributor to these high frequency responses.

Such phase locked auditory activity is also compatible with the spiking output of the Medial Geniculate Body (MGB) (Miller et al., 2002), which are inputs to early auditory cortical areas. The MEG signal is dominantly driven by dendritic currents (similar to the Local Field Potential) (Hämäläinen et al., 1993), and hence these high frequency responses may be due to the inputs from the MGB into auditory cortex. Prior work has shown that auditory cortex is able to transiently time-lock to continuous acoustic features with surprisingly high temporal precision (Elhilali, 2004). Time-locked inputs from MGB may provide a neural substrate for such precise transient temporal locking to stimulus features. Direct correspondences with age-related changes in thalamus from animal work are limited (Caspary and Llano, 2019), and hence it is unclear if time-locked high frequency spiking activity in MGB animal models would be similar across age. However, invasive neural recordings could help to disentangle the opposite effects of aging in the brainstem and the cortex seen with M/EEG, leading to a better understanding of time-locked responses in the aging auditory pathway.

## 5. Conclusion

In this study, we found time-locked responses to continuous speech, using MEG, that localize to auditory cortex, occur with a peak latency of approximately 40 ms, and are stronger in the right hemisphere. The response function oscillates at approximately 80-90 Hz, which is well below the fundamental frequency of the speech stimulus, and is predominantly driven by the variations in the high frequency envelope of the stimulus. Such high frequency time-locked responses may possibly be due to thalamic inputs to cortical neurons. Because MEG is more sensitive to cortical sources, it may be able to cleanly detect cortical responses that, in EEG, are strongly mixed with subcortical responses. Furthermore, there were no significant age-related differences in these high frequency responses, unlike in both the low frequency cortical TRFs or the subcortical FFRs. Hence both the neural origin and the frequency domain must be considered when investigating age-related changes in the auditory system.

## Acknowledgements

Funding: This work was supported by the National Institute on Deafness and Other Communication Disorders (R01 DC-014085); the National Institute of Aging (P01 AG-055365); and the National Science Foundation (SMA1734892).

The identification of specific products or scientific instrumentation is considered an integral part of the scientific endeavor and does not constitute endorsement or implied endorsement on the part of the authors, DoD, or any component agency. The views expressed in this article are those of the authors and do not reflect the official policy of the Department of Army/Navy/Air Force, Department of Defense, or U.S. Government.

## Disclosures

The authors declare that there are no conflicts of interest.

# Appendix

**Surface source space TRF methods and results**

Surface source space estimation was performed using the 'ico-4' source space, which consists of a fourfold icosahedral subdivision of the white matter surface of the brain with dipoles oriented normal to the surface. The 'aparc' parcellation was used to select dipoles in the temporal lobe for further analysis. In this surface source space analysis, current dipoles have a fixed orientation normal to the surface, and hence the TRF consists only of signed scalar amplitude variations with time. The Pearson correlation between the actual and predicted neural response was used as a measure of prediction accuracy for each neural source. For statistical tests, the TRFs and the correlation values were first rectified and then spatially smoothed using a Gaussian window with a standard deviation of 5 mm. The rectified, smoothed TRF of the true model was compared to the average of that of the three noise models using the same one tailed test with paired sample *t*-values and the TFCE procedure outlined in Methods.

Lateralization tests were performed to check for hemispheric asymmetry. The correlation values at each neural source in both left and right hemisphere were morphed onto the right hemisphere of the 'fsaverage_sym' brain as described in (Brodbeck et al., 2018a). This brain model is symmetric in left and right hemispheres, allowing for comparisons between corresponding neural sources in both hemispheres. As before, these correlation coefficients were spatially smoothed using the same Gaussian window. After morphing, the correlation values of the average noise model were subtracted from that of the true model and a two-tailed test with paired sample *t*-values and TFCE was used to assess for significant differences in each of the corresponding left and right current dipoles.

TRFs were estimated using the cortical surface source space for neural sources in the temporal lobe, using both the HFE and the carrier predictors in a competing model. Both predictors were time-shifted to generate noise models. All surface space results were similar to volume source space results. The prediction accuracies and TRFs are shown in Fig. A1, Fig. A2. The prediction accuracies were right lateralized but only in younger subjects ($t_{max} = 4.12$, $p = 0.023$). The TRFs showed a significant response in the range of 24-62 ms for the HFE and 34-47 ms for the carrier. The HFE TRF was stronger than the carrier TRF using the same tests as in the volume source

space (younger $t_{max}$ = 8.63, $p < 0.001$; older $t_{max}$ = 4.94, $p < 0.001$). There were no age-related differences in surface source space analyses (prediction accuracy $t_{max}$ = 2.59, $t_{min}$ = -1.66, $p > 0.35$; maximum amplitude of HFE TRF $t_{max}$ = 3.53, $t_{min}$ = -3.71, $p > 0.69$; maximum amplitude of carrier TRF $t_{max}$ = 3.71, $t_{min}$ = -3.64, $p > 0.18$).

In addition, low frequency TRFs were also estimated to compare age-related differences in both frequency domains. The stimulus representation for this model was the Hilbert envelope of the speech waveform filtered at 1-10 Hz with a logarithmic nonlinearity applied. The MEG data was also filtered at 1-10 Hz and TRFs were estimated using the surface source space. The resulting TRFs were as expected from prior work (Brodbeck et al., 2018b), with older subjects showing significantly higher reconstruction accuracies ($t_{max}$ = 0.93, $t_{min}$ = -3.45, $p = 0.022$). The increase in model prediction accuracies above the noise, for the high frequency TRF and the low frequency TRF were averaged across neural sources per subject, and a TRF frequency by age ANOVA was performed. Results indicated a significant main effect of TRF frequency ($F_{(1,38)}$ = 125.94, $p < 0.001$) and interaction of TRF frequency x age ($F_{(1,38)}$ = 7.52, $p = 0.009$) but no main effect of age ($F_{(1,38)}$ = 2.26, $p = 0.141$). This suggests that age-related changes are not consistent across low and high frequency responses, agreeing with all the above results in support of the null hypothesis that there are no age-related differences in high frequency cortical responses.

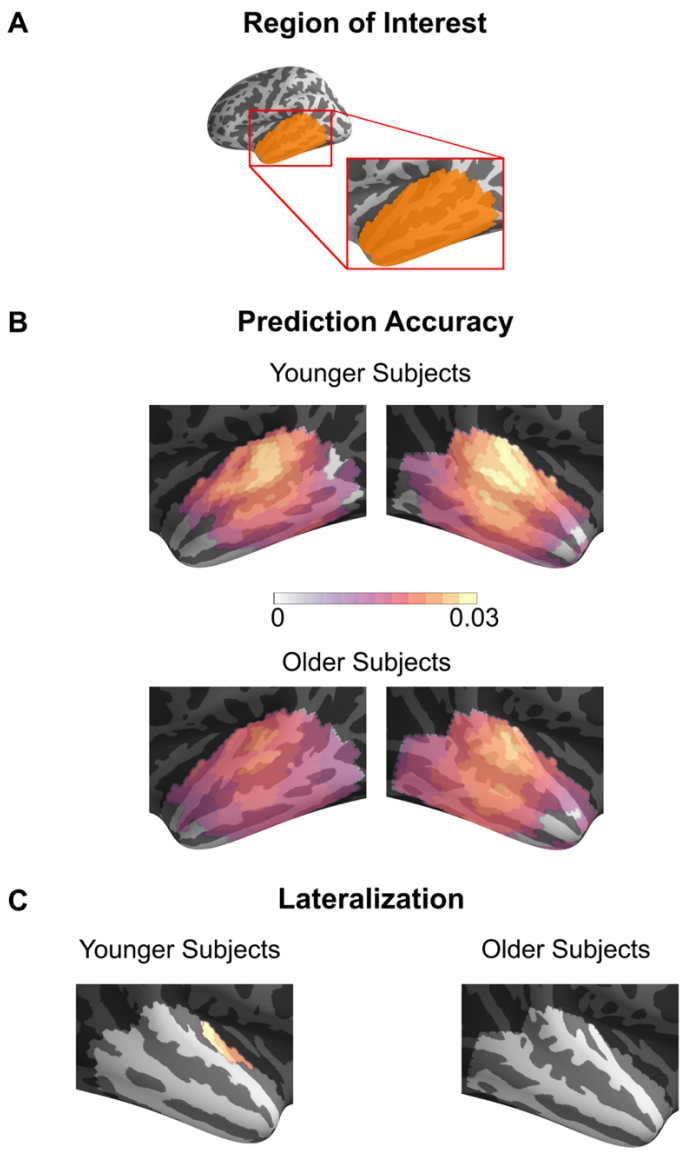


**Fig. A1.** Prediction Accuracy of Surface Source Space TRFs. Pearson correlation coefficients between the actual and predicted response using the TRF model for each source in the surface source space ROI averaged across subjects. Only the voxels showing a significant increase in prediction accuracy over the noise model are plotted. Although most neural sources are significantly predictive, the actual prediction accuracy values are larger in areas near core auditory cortex. One region in auditory cortex is significantly more predictive in the right hemisphere than the left, but only in younger subjects.

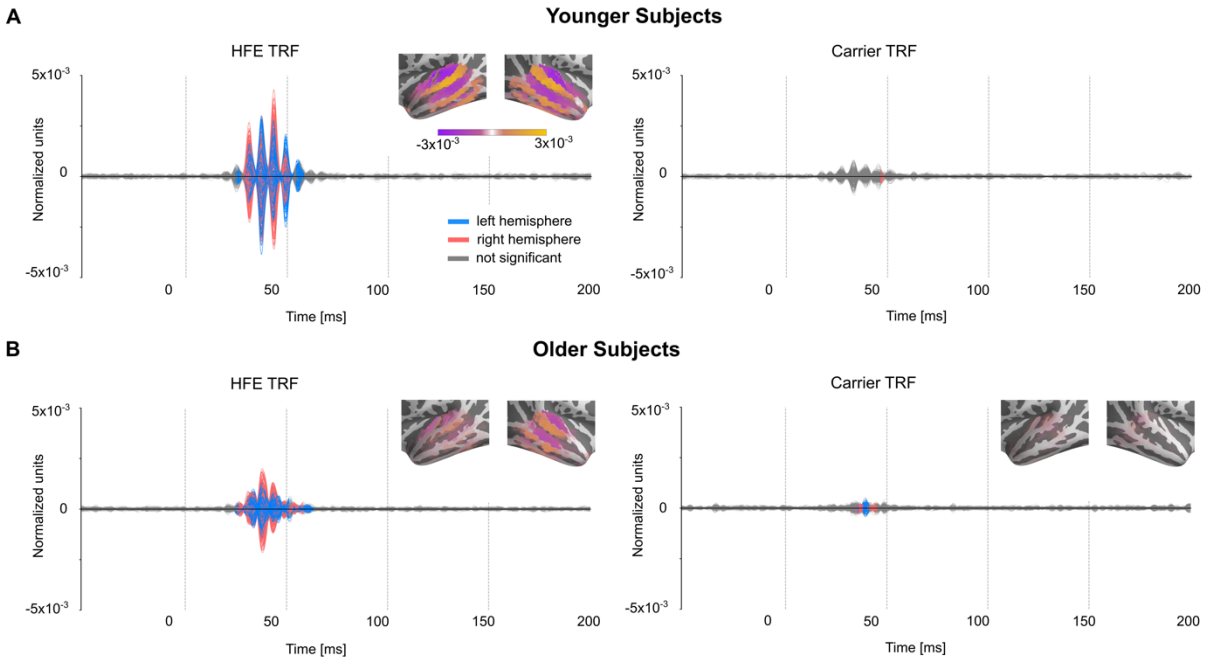


**Fig. A2.** Surface Space Source Localized TRFs. The TRFs for the competing model for both predictors averaged across subjects and masked by significance against the noise model. Unlike the volume source space, the surface source space comprises of current dipoles with fixed orientation normal to the cortical surface. The signed magnitudes of these fixed direction dipoles are plotted, allowing for positive and negative values for outward and inward directions. The distribution of current dipoles in the temporal lobe ROI at the peak of the response is shown as an inset (orange: positive, purple: negative).

# References


Aiken, S.J., Picton, T.W., 2008. Envelope and spectral frequency-following responses to vowel sounds. Hear. Res. 245, 35–47. https://doi.org/10.1016/j.heares.2008.08.004

Anderson, S., Parbery-Clark, A., White-Schwoch, T., Kraus, N., 2012. Aging Affects Neural Precision of Speech Encoding. J. Neurosci. 32, 14156–14164. https://doi.org/10.1523/JNEUROSCI.2176-12.2012

Attal, Y., Bhattacharjee, M., Yelnik, J., Cottereau, B., Lefevre, J., Okada, Y., Bardinet, E., Chupin, M., Baillet, S., 2007. Modeling and Detecting Deep Brain Activity with MEG EEG, in: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Presented at the 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 4937–4940. https://doi.org/10.1109/IEMBS.2007.4353448

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

Balderston, N.L., Schultz, D.H., Baillet, S., Helmstetter, F.J., 2014. Rapid Amygdala Responses during Trace Fear Conditioning without Awareness. PLOS ONE 9, e96803. https://doi.org/10.1371/journal.pone.0096803

Basu, M., Krishnan, A., Weber-Fox, C., 2010. Brainstem correlates of temporal auditory processing in children with specific language impairment: Brainstem correlates of temporal processing. Dev. Sci. 13, 77–91. https://doi.org/10.1111/j.1467-7687.2009.00849.x

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

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

Brodbeck, C., Brooks, T.L., Das, P., Reddigari, S., 2019. christianbrodbeck/Eelbrain: 0.30. Zenodo. https://doi.org/10.5281/zenodo.2653785

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

Brodbeck, C., Presacco, A., Anderson, S., Simon, J.Z., 2018b. Over-Representation of Speech in Older Adults Originates from Early Response in Higher Order Auditory Cortex. Acta Acust. United Acust. 104, 774–777. https://doi.org/10.3813/AAA.919221

Brodbeck, C., Presacco, A., Simon, J.Z., 2018c. Neural source dynamics of brain responses to continuous stimuli: Speech processing from acoustics to comprehension. NeuroImage 172, 162–174. https://doi.org/10.1016/j.neuroimage.2018.01.042

Caspary, D.M., Ling, L., Turner, J.G., Hughes, L.F., 2008. Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system. J. Exp. Biol. 211, 1781–1791. https://doi.org/10.1242/jeb.013581

Caspary, D.M., Llano, D.A., 2019. Aging Processes in the Subcortical Auditory System, in: Kandler, K. (Ed.), The Oxford Handbook of the Auditory Brainstem. Oxford University Press, pp. 638–680. https://doi.org/10.1093/oxfordhb/9780190849061.013.16

Cha, K., Zatorre, R.J., Schönwiesner, M., 2016. Frequency Selectivity of Voxel-by-Voxel Functional Connectivity in Human Auditory Cortex. Cereb. Cortex N. Y. NY 26, 211–224. https://doi.org/10.1093/cercor/bhu193

Coffey, E.B.J., Chepesiuk, A.M.P., Herholz, S.C., Baillet, S., Zatorre, R.J., 2017a. Neural Correlates of Early Sound Encoding and their Relationship to Speech-in-Noise Perception. Front. Neurosci. 11, 479. https://doi.org/10.3389/fnins.2017.00479

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. Nat. Commun. 7, 11070. https://doi.org/10.1038/ncomms11070

Coffey, E.B.J., Musacchia, G., Zatorre, R.J., 2017b. Cortical Correlates of the Auditory Frequency-Following and Onset Responses: EEG and fMRI Evidence. J. Neurosci. 37, 830–838. https://doi.org/10.1523/JNEUROSCI.1265-16.2016

Coffey, E.B.J., Nicol, T., White-Schwoch, T., Chandrasekaran, B., Krizman, J., Skoe, E., Zatorre, R.J., Kraus, N., 2019. Evolving perspectives on the sources of the frequency-following response. Nat. Commun. 10. https://doi.org/10.1038/s41467-019-13003-w

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

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 Res. 1244, 103–112. https://doi.org/10.1016/j.brainres.2008.09.068

Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine, J.D., Halgren, E., 2000. Dynamic Statistical Parametric Mapping: Combining fMRI and MEG for High-Resolution Imaging of Cortical Activity. Neuron 26, 55–67. https://doi.org/10.1016/S0896-6273(00)81138-1

David, S.V., Mesgarani, N., Shamma, S.A., 2007. Estimating sparse spectro-temporal receptive fields with natural stimuli. Netw. Bristol Engl. 18, 191–212. https://doi.org/10.1080/09548980701609235

de Cheveigné, A., Simon, J.Z., 2008. Sensor noise suppression. J. Neurosci. Methods 168, 195–202. https://doi.org/10.1016/j.jneumeth.2007.09.012

de Cheveigné, A., Simon, J.Z., 2007. Denoising based on Time-Shift PCA. J. Neurosci. Methods 165, 297–305. https://doi.org/10.1016/j.jneumeth.2007.06.003

Decruy, L., Vanthornhout, J., Francart, T., 2019. Evidence for enhanced neural tracking of the speech envelope underlying age-related speech-in-noise difficulties. J. Neurophysiol. 122, 601–615. https://doi.org/10.1152/jn.00687.2018

Ding, N., Chatterjee, M., Simon, J.Z., 2014. Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. NeuroImage 88, 41–46. https://doi.org/10.1016/j.neuroimage.2013.10.054

Ding, N., Simon, J.Z., 2012. Emergence of neural encoding of auditory objects while listening to competing speakers. Proc. Natl. Acad. Sci. 109, 11854–11859. https://doi.org/10.1073/pnas.1205381109

Dumas, T., Dubal, S., Attal, Y., Chupin, M., Jouvent, R., Morel, S., George, N., 2013. MEG Evidence for Dynamic Amygdala Modulations by Gaze and Facial Emotions. PLoS ONE 8. https://doi.org/10.1371/journal.pone.0074145

Elhilali, M., 2004. Dynamics of Precise Spike Timing in Primary Auditory Cortex. J. Neurosci. 24, 1159–1172. https://doi.org/10.1523/JNEUROSCI.3825-03.2004

Fischl, B., 2012. FreeSurfer. NeuroImage 62, 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021
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
Gordon-Salant, S., Yeni-Komshian, G.H., Fitzgibbons, P.J., Barrett, J., 2006. Age-related differences in identification and discrimination of temporal cues in speech segments. J. Acoust. Soc. Am. 119, 2455–2466. https://doi.org/10.1121/1.2171527
Gramfort, A., 2013. MEG and EEG data analysis with MNE-Python. Front. Neurosci. 7. https://doi.org/10.3389/fnins.2013.00267
Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Parkkonen, L., Hämäläinen, M.S., 2014. MNE software for processing MEG and EEG data. NeuroImage 86, 446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027
Hämäläinen, M., Hari, R., Ilmoniemi, R.J., Knuutila, J., Lounasmaa, O.V., 1993. Magnetoencephalography---theory, instrumentation, and applications to noninvasive studies of the working human brain. Rev. Mod. Phys. 65, 413–497. https://doi.org/10.1103/RevModPhys.65.413
Hartmann, T., Weisz, N., 2019. Auditory cortical generators of the Frequency Following Response are modulated by intermodal attention. bioRxiv. https://doi.org/10.1101/633834
He, N., Mills, J.H., Ahlstrom, J.B., Dubno, J.R., 2008. Age-related differences in the temporal modulation transfer function with pure-tone carriers. J. Acoust. Soc. Am. 124, 3841–3849. https://doi.org/10.1121/1.2998779
Hertrich, I., Dietrich, S., Trouvain, J., Moos, A., Ackermann, H., 2012. Magnetic brain activity phase-locked to the envelope, the syllable onsets, and the fundamental frequency of a perceived speech signal. Psychophysiology 49, 322–334. https://doi.org/10.1111/j.1469-8986.2011.01314.x
Hertrich, I., Mathiak, K., Lutzenberger, W., Ackermann, H., 2004. Transient and phase-locked evoked magnetic fields in response to periodic acoustic signals. Neuroreport 15, 1687–1690. https://doi.org/10.1097/01.wnr.0000134930.04561.b2
Hillebrand, A., Barnes, G.R., 2002. A Quantitative Assessment of the Sensitivity of Whole-Head MEG to Activity in the Adult Human Cortex. NeuroImage 16, 638–650. https://doi.org/10.1006/nimg.2002.1102
Hopkins, K., Moore, B.C.J., 2011. The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise. J. Acoust. Soc. Am. 130, 334–349. https://doi.org/10.1121/1.3585848
Hornickel, J., Anderson, S., Skoe, E., Yi, H.-G., Kraus, N., 2012. Subcortical representation of speech fine structure relates to reading ability: NeuroReport 23, 6–9. https://doi.org/10.1097/WNR.0b013e32834d2ffd
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
Jarosz, A.F., Wiley, J., 2014. What Are the Odds? A Practical Guide to Computing and Reporting Bayes Factors. J. Probl. Solving 7. https://doi.org/10.7771/1932-6246.1167
Kraus, N., Anderson, S., White-Schwoch, T., Fay, R.R., Popper, A.N., 2017. The Frequency-Following Response: A Window into Human Communication. Springer.

Krishnaswamy, P., Obregon-Henao, G., Ahveninen, J., Khan, S., Babadi, B., Iglesias, J.E., Hämäläinen, M.S., Purdon, P.L., 2017. Sparsity enables estimation of both subcortical and cortical activity from MEG and EEG. Proc. Natl. Acad. Sci. U. S. A. 114, E10465–E10474. https://doi.org/10.1073/pnas.1705414114

Lalor, E.C., Power, A.J., Reilly, R.B., Foxe, J.J., 2009. Resolving Precise Temporal Processing Properties of the Auditory System Using Continuous Stimuli. J. Neurophysiol. 102, 349–359. https://doi.org/10.1152/jn.90896.2008

Lu, T., Liang, L., Wang, X., 2001. Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. Nat. Neurosci. 4, 1131–1138. https://doi.org/10.1038/nn737

Maddox, R.K., Lee, A.K.C., 2018. Auditory Brainstem Responses to Continuous Natural Speech in Human Listeners. eNeuro 5. https://doi.org/10.1523/ENEURO.0441-17.2018

Miller, L.M., Escabí, M.A., Read, H.L., Schreiner, C.E., 2002. Spectrotemporal Receptive Fields in the Lemniscal Auditory Thalamus and Cortex. J. Neurophysiol. 87, 516–527. https://doi.org/10.1152/jn.00395.2001

Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: A primer with examples. Hum. Brain Mapp. 15, 1–25. https://doi.org/10.1002/hbm.1058

Nourski, K.V., Steinschneider, M., McMurray, B., Kovach, C.K., Oya, H., Kawasaki, H., Howard, M.A., 2014. Functional organization of human auditory cortex: Investigation of response latencies through direct recordings. NeuroImage 101, 598–609. https://doi.org/10.1016/j.neuroimage.2014.07.004

Oldfield, R.C., 1971. The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 9, 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Park, H.-D., Tallon-Baudry, C., 2014. The neural subjective frame: from bodily signals to perceptual consciousness. Philos. Trans. R. Soc. B Biol. Sci. 369. https://doi.org/10.1098/rstb.2013.0208

Parkkonen, L., Fujiki, N., Mäkelä, J.P., 2009. Sources of auditory brainstem responses revisited: Contribution by magnetoencephalography. Hum. Brain Mapp. 30, 1772–1782. https://doi.org/10.1002/hbm.20788

Pascual-Marqui, R.D., 2002. Standardized low-resolution brain electromagnetic tomography (sLORETA): technical details. Methods Find. Exp. Clin. Pharmacol. 24 Suppl D, 5–12.

Peelle, J.E., Gross, J., Davis, M.H., 2013. Phase-Locked Responses to Speech in Human Auditory Cortex are Enhanced During Comprehension. Cereb. Cortex 23, 1378–1387. https://doi.org/10.1093/cercor/bhs118

Peelle, J.E., Wingfield, A., 2016. The Neural Consequences of Age-Related Hearing Loss. Trends Neurosci. 39, 486–497. https://doi.org/10.1016/j.tins.2016.05.001

Presacco, A., Jenkins, K., Lieberman, R., Anderson, S., 2015. Effects of Aging on the Encoding of Dynamic and Static Components of Speech. Ear Hear. 36, e352–e363. https://doi.org/10.1097/AUD.0000000000000193

Presacco, A., Simon, J.Z., Anderson, S., 2016a. Evidence of degraded representation of speech in noise, in the aging midbrain and cortex. J. Neurophysiol. 116, 2346–2355. https://doi.org/10.1152/jn.00372.2016

Presacco, A., Simon, J.Z., Anderson, S., 2016b. Effect of informational content of noise on speech representation in the aging midbrain and cortex. J. Neurophysiol. 116, 2356–2367. https://doi.org/10.1152/jn.00373.2016

Puschmann, S., Baillet, S., Zatorre, R.J., 2019. Musicians at the Cocktail Party: Neural Substrates of Musical Training During Selective Listening in Multispeaker Situations. Cereb. Cortex 29, 3253–3265. https://doi.org/10.1093/cercor/bhy193

Rosen, S., 1992. Temporal information in speech: acoustic, auditory and linguistic aspects. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 336, 367–373. https://doi.org/10.1098/rstb.1992.0070

Ross, B., Borgmann, C., Draganova, R., Roberts, L.E., Pantev, C., 2000. A high-precision magnetoencephalographic study of human auditory steady-state responses to amplitude-modulated tones. J. Acoust. Soc. Am. 108, 679–691. https://doi.org/10.1121/1.429600

Ross, B., Herdman, A.T., Pantev, C., 2005. Right Hemispheric Laterality of Human 40 Hz Auditory Steady-state Responses. Cereb. Cortex 15, 2029–2039. https://doi.org/10.1093/cercor/bhi078

Rouder, J.N., Speckman, P.L., Sun, D., Morey, R.D., Iverson, G., 2009. Bayesian t tests for accepting and rejecting the null hypothesis. Psychon. Bull. Rev. 16, 225–237. https://doi.org/10.3758/PBR.16.2.225

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. J. Neurosci. 33, 17827–17835. https://doi.org/10.1523/JNEUROSCI.5778-12.2013

Schoonhoven, R., Boden, C.J.R., Verbunt, J.P.A., de Munck, J.C., 2003. A whole head MEG study of the amplitude-modulation-following response: phase coherence, group delay and dipole source analysis. Clin. Neurophysiol. 114, 2096–2106. https://doi.org/10.1016/S1388-2457(03)00200-1

Smith, J.C., Marsh, J.T., Brown, W.S., 1975. Far-field recorded frequency-following responses: Evidence for the locus of brainstem sources. Electroencephalogr. Clin. Neurophysiol. 39, 465–472. https://doi.org/10.1016/0013-4694(75)90047-4

Smith, S.M., Nichols, T.E., 2009. Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. NeuroImage 44, 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061

Steinschneider, M., Nourski, K.V., Fishman, Y.I., 2013. Representation of speech in human auditory cortex: Is it special? Hear. Res. 305. https://doi.org/10.1016/j.heares.2013.05.013

Tichko, P., Skoe, E., 2017. Frequency-dependent fine structure in the frequency-following response: The byproduct of multiple generators. Hear. Res. 348, 1–15. https://doi.org/10.1016/j.heares.2017.01.014

Yang, X., Wang, K., Shamma, S.A., 1992. Auditory representations of acoustic signals. IEEE Trans. Inf. Theory 38, 824–839. https://doi.org/10.1109/18.119739

Yellamsetty, A., Bidelman, G.M., 2019. Brainstem correlates of concurrent speech identification in adverse listening conditions. Brain Res. 1714, 182–192. https://doi.org/10.1016/j.brainres.2019.02.025

Zan, P., Presacco, A., Anderson, S., Simon, J.Z., 2019. Mutual information analysis of neural representations of speech in noise in the aging midbrain. J. Neurophysiol. 122, 2372–2387. https://doi.org/10.1152/jn.00270.2019

Zatorre, R.J., 1988. Pitch perception of complex tones and human temporal-lobe function. J. Acoust. Soc. Am. 84, 566–572. https://doi.org/10.1121/1.396834