High Frequency Time Locking in Auditory Cortex to Continuous Speech

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

[1]Department of Electrical and Computer Engineering, [2]Institute for Systems Research, [3]Department of Hearing and Speech Sciences, [4]Department of Biology, University of Maryland, College Park, Maryland,
[5]Audiology and Speech Pathology Center, Water Reed National Military Medical Center

Introduction

The neural processing of natural sounds, such as speech, changes along the ascending auditory pathway, and is often characterized by a progressive reduction in representative frequencies. This is observed in two well known neural responses:

1. The Frequency Following Response (FFR):
   - Thought to originate in the auditory brainstem.
   - Frequencies of ~100 Hz to several hundred Hz.
   - Time-locks to fast acoustic envelope and waveform.
2. The cortical low frequency Temporal Response Function (TRF):
   - Frequencies of around 1-20 Hz.
   - Time-locks to slow acoustic envelope.

Recent studies have shown that the FFR may not have a purely subcortical origin (Coffey et al. 2016). Recent studies have also shown a continuous speech generated FFR-like response in brainstem (Maddox and Lee 2018) and that brainstem responses are modulated by attention (Forte et al. 2017).

Age-related differences, of opposite direction, have been observed for FFR and the slow cortical TRF:
- FFR response is stronger in younger subjects (Anderson et al. 2012).
- cortical low frequency TRF response is stronger in older subjects (Brodbeck et al. 2018).

Does the cortex time-lock to high frequency components of connected speech stimuli?

Are there age-related differences in high frequency responses in the cortex?

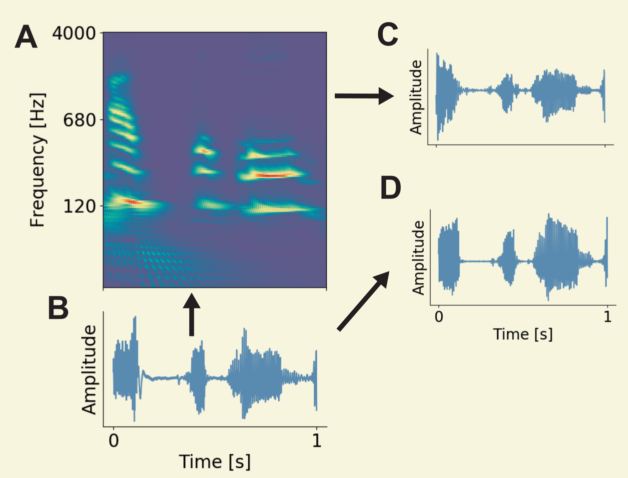
Methods

MEG data was collected from 17 younger and 23 older subjects while they listened to 6 minutes of speech from an audiobook. Earlier analysis of dataset published as Presacco et. al. 2016a, 2016b.

Statistical significance was tested using permutation tests on the model prediction accuracy and the spatially and temporally smoothed rectified TRFs.

Stimulus Representation

Two separate high frequency aspects of the speech stimuli were considered.
1. **The waveform** of the speech bandpassed 80-300 Hz.
2. **The envelope** of the high frequency (300-4000 Hz) components of the speech stimuli, bandpassed 80-300 Hz.
   - Compute the envelope of the auditory spectrogram (Yang et al. 1991).
   - Filter the 300-4000 Hz components of the spectrogram envelope at 80-300 Hz.
   - Average those filtereed spectrogram envelope components across frequency.



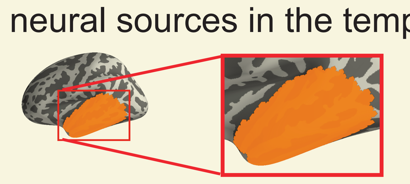
(A) Spectrogram representation (B) Speech stimulus waveform (C) Envelope of high frequencies (D) Waveform bandpassed 80-300 Hz

Source Localization

The 157 MEG sensors were transformed to current dipole sources distributed on the brain using MNE source localization (Gramfort et al. 2014).

1. Volume source space
- The brain volume is divided into 10mm³ voxels.
- Current dipole vector with 3D orientation and magnitude placed at each voxel.
- ROI: voxels closest to a) the temporal lobe and b) the brainstem.

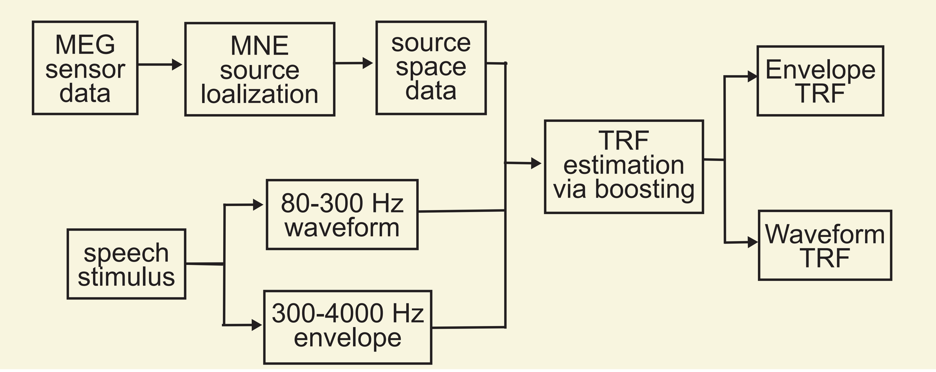
2. Surface source space
- Neural sources are distributed on the white matter surface of the brain at ~4mm².
- Current dipoles fixed at an orientation perpendicular to the surface of the brain.
- ROI: neural sources in the temporal lobe.



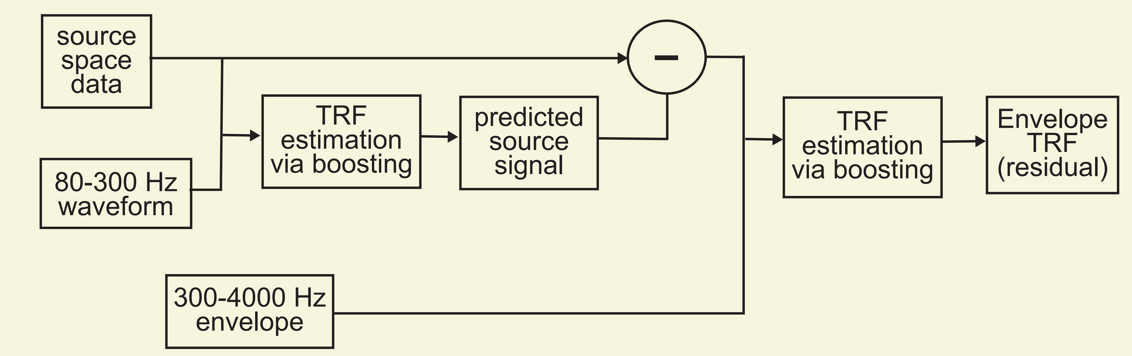
Linear Kernel Estimation Models

Model 1: Joint Estimation of Waveform and Envelope TRFs
Envelope and waveform TRFs were estimated with the boosting algorithm (David et al. 2007) using Eelbrain (Brodbeck 2018).

$$r^s = \tau_e^s * e + \tau_w^s * w$$



Model 2: Residual TRF Estimation
The waveform TRF was first computed alone, its prediction subtracted from the response, and only then the envelope TRF was computed on the residual. The same was done in reverse order.



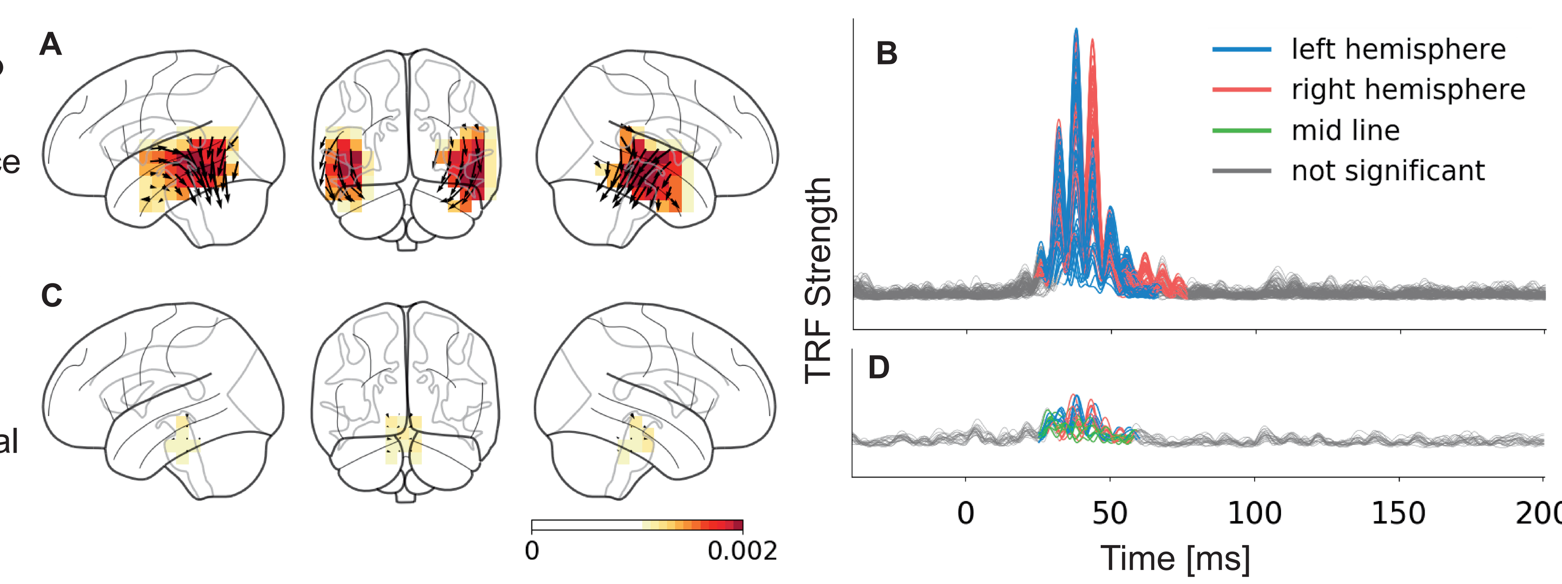
Results

Cortical or Subcortical?

TRF estimation using volume source space.

Significant TRFs were observed for all relevant voxels, but the subcortical responses appear as artifactual leakage from cortical sources: low amplitudes with cortical latencies.

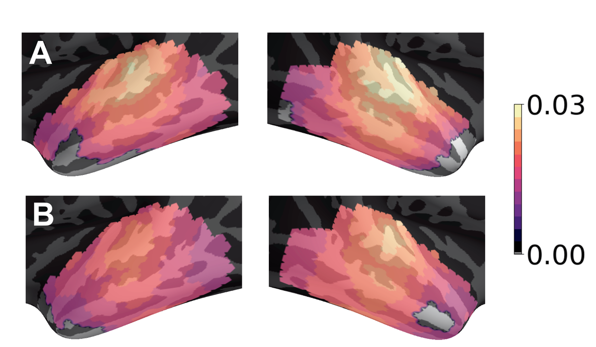
TRF peak latency of 40 ms and spatial location strongly indicate a purely cortical source



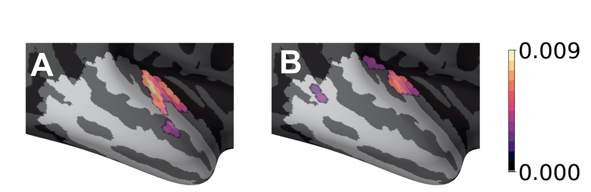
Volume Source Space TRFs: (A) Peak TRF vectors in voxels closest to temporal lobe. (B) TRFs of voxels closest to temporal lobe. (C) Peak TRF vectors of voxels closest to brainstem. (D) TRFs of voxels closest to brainstem. (for the envelope predictor)

Cortically Generated Responses

Surface source space data was analyzed using Model 1.

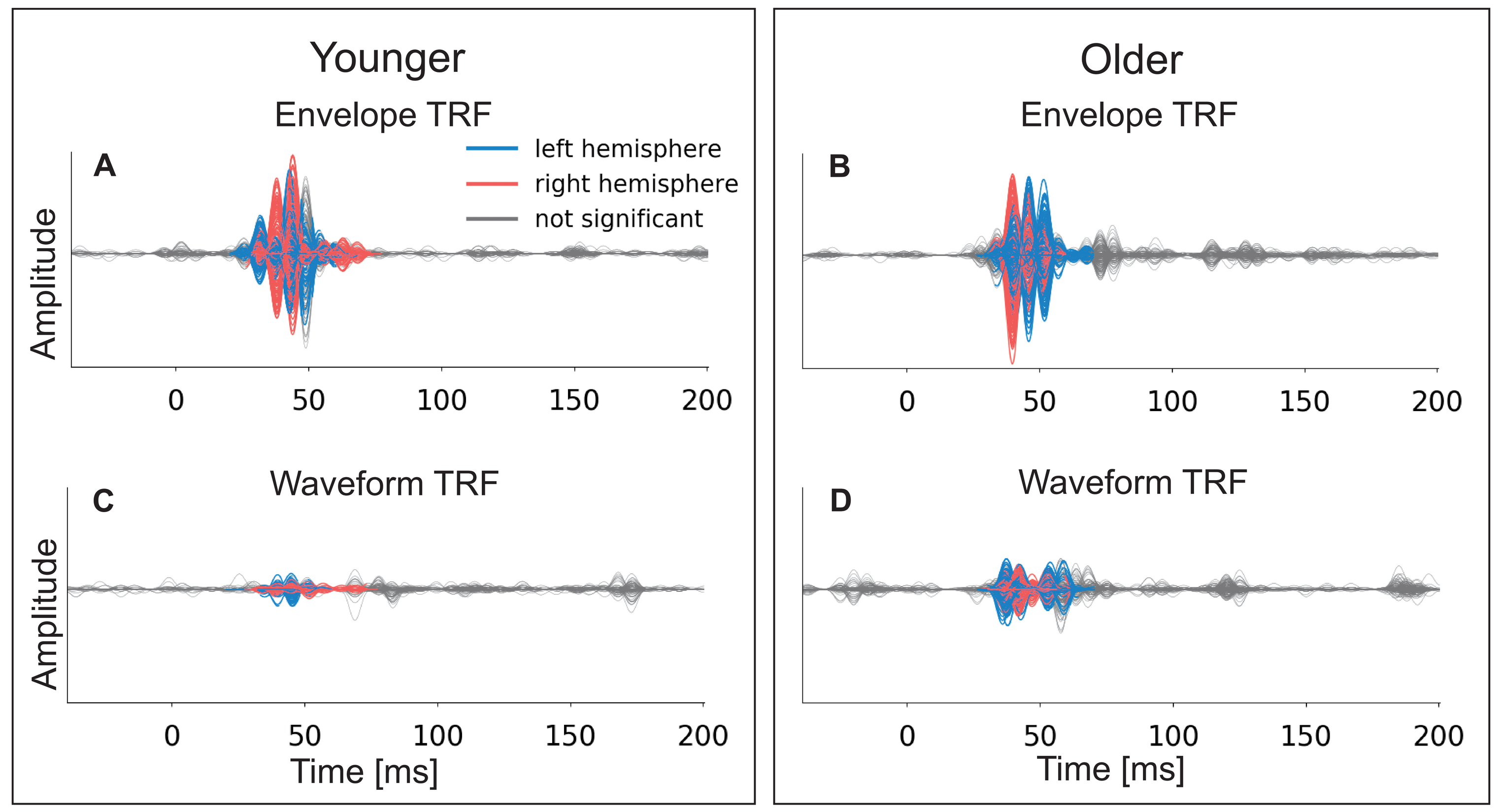


TRF analysis reveals significant predictive power in younger (A) and older (B) groups for neural sources in the temporal lobe.



Predictive power is right lateralized in both younger (A) and older (B) groups.

Significantly right lateralized in early auditory cortex for both older and younger subjects.



TRFs for a representative younger subject (A), (C) and for a representative older subject (B), (D).

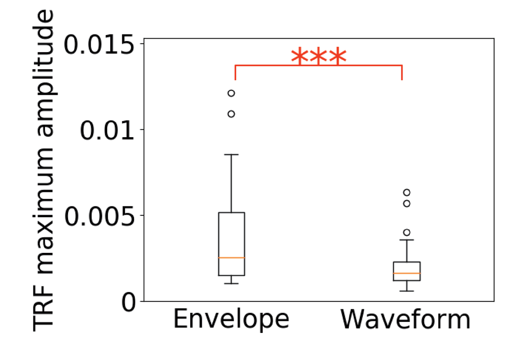
TRF peak latency of 40 ms and spatial location strongly indicate that the response originates in early auditory cortex.

No significant differences between age groups with these measures.

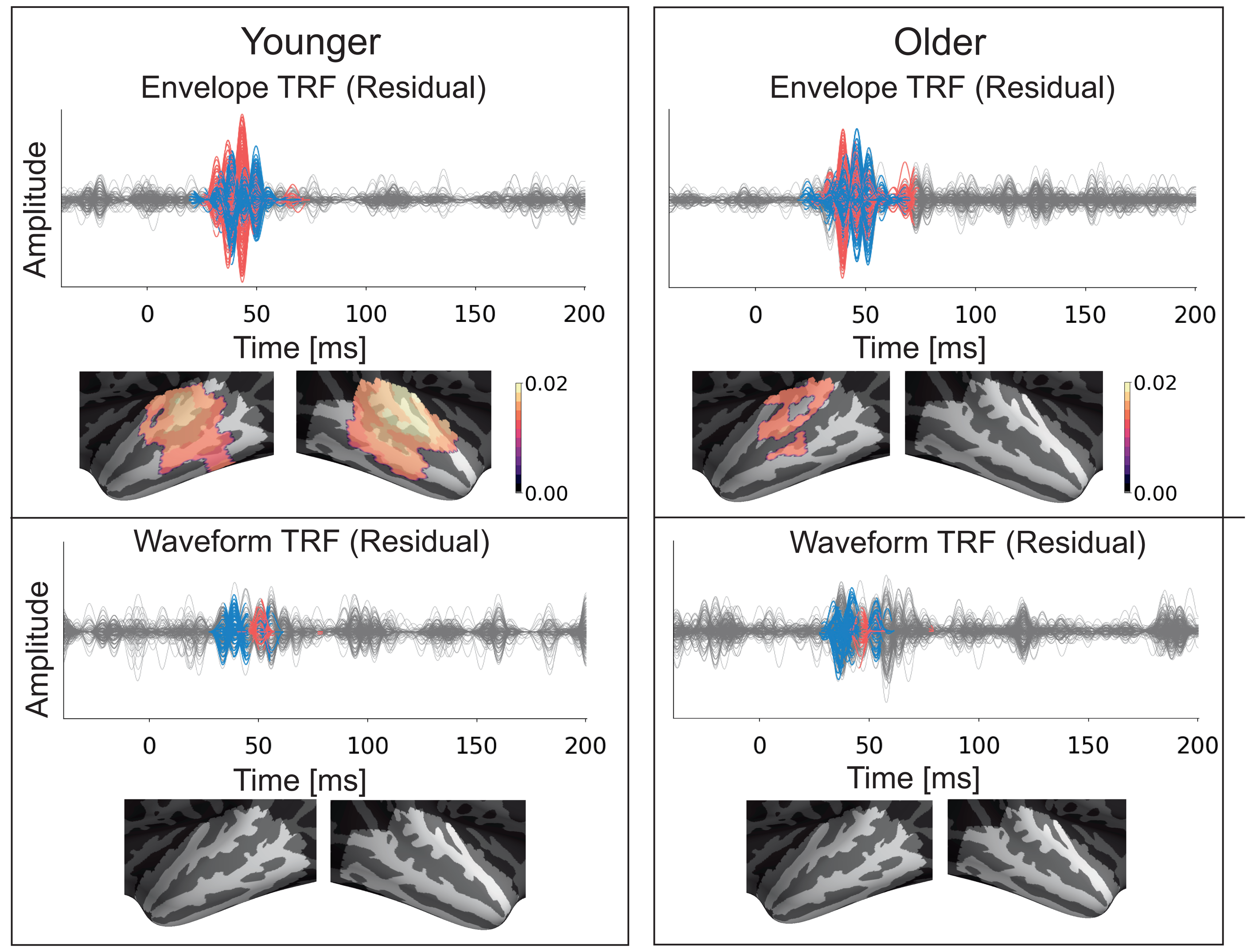
Envelope vs. Waveform Contributions

Since the 2 predictors are correlated, tests were performed to evaluate the conservative contribution of each predictor to the model.

Model 1: Both predictors contributed to the model, with the envelope contributing more than the waveform.



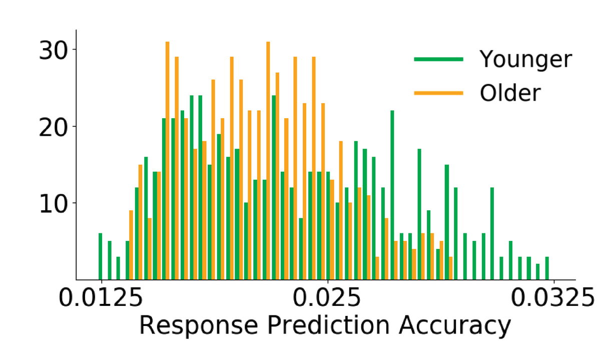
Model 2: The conservative contribution of the envelope was larger than the conservative contribution of the waveform.



No significant differences between age groups using these measures.

Subtle Age Differences

Significant differences were observed when averaging across subjects per neural source, rather than when averaging across neural sources per subject. The responses of younger subjects were more predictive than those of older subjects.



Conclusions

- This work reveals that there exist high frequency cortical responses to continuous speech that are measurable using MEG.
- Both peak latency (40ms) and source localization strongly indicate a purely cortical origin for this response.
- Two models for TRF estimation were used and both indicate that the high frequency envelope contributes more than the waveform itself to this response
- Age differences for high frequencies in cortex are subtle: standard analysis methods do not show significant differences, but other techniques do. This is unlike the FFR for which younger subjects have more reliable responses, or the cortical slow TRF for which older subjects have more reliable responses.
- Together, these results suggest that the traditional EEG-measured FFR may have distinct and separate contributions from both subcortical and cortical sources.

References

Coffey, E. B. J., et al. "Cortical contributions to the auditory frequency-following response revealed by MEG." Nature communications 7 (2016): 11070.
Maddox, R. K., & Lee, A. K. C. "Auditory brainstem responses to continuous natural speech in human listeners." ENeuro 5.1 (2018).
Forte, A. E., Etard, O., & Reichenbach, E.. "The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention." elife 6 (2017): e27203.
Brodbeck, C. et al. "Over-representation of speech in older adults originates from early response in higher order auditory cortex." Acta Acustica united with Acustica 104.5 (2018): 774-777.
Anderson, S. et al. "Aging affects neural precision of speech encoding." Journal of Neuroscience 32.41 (2012): 14156-14164.
Yang, X., Wang, K., & Shamma, S. A. Auditory representations of acoustic signals. 1991.
Presacco, A., Simon, J. Z., & Anderson, S. "Effect of informational content of noise on speech representation in the aging midbrain and cortex." Journal of neurophysiology 116.5 (2016): 2356-2367.
Presacco, A., Simon, J. Z., & Anderson, S. "Evidence of degraded representation of speech in noise, in the aging midbrain and cortex." Journal of neurophysiology 116.5 (2016): 2346-2355.
Gramfort, A., et al. "MNE software for processing MEG and EEG data." Neuroimage 86 (2014): 446-460.
David, S. V., Mesgarani, N., & Shamma, S. A. "Estimating sparse spectro-temporal receptive fields with natural stimuli." Network: Computation in Neural Systems 18.3 (2007): 191-212.
Brodbeck, C. : Eelbrain 0.30. Zenodo, 2019.
Kuchinsky, S., et al. "Auditory-Cognitive Training Alters the Neural Encoding of Speech for Older Adults with Normal Hearing" (ARO 2017)

Acknowledgments

This work was funded by the National Institutes of Health (NIH), grants R01-DC014085 and P01-AG055365