Estimation of the Neural Sources of the EEG-measured Speech-FFR using a Phenomenological Model of Auditory-nerve Fiber Responses

1st Jonas Auernheimer
Department Artificial Intelligence in Biomedical Engineering (AIBE)
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
Erlangen, Germany
jonas.auernheimer@fau.de

2nd Tobias Reichenbach
Department Artificial Intelligence in Biomedical Engineering (AIBE)
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)
Erlangen, Germany
tobias.j.reichenbach@fau.de

***Abstract*—The frequency-following-response to continuous speech (speech-FFR) is a characteristic neural activity that emerges at the fundamental frequency of a speaker's voice and can be measured by electroencephalography (EEG) and magnetoencephalography (MEG). Its spectrotemporal dynamics and neural sources encode important pitch information critical for reliable speech processing. EEG studies have found a peak latency of the speech-FFR at around 10 ms with putative subcortical origin while recent MEG studies identified additional cortical contributions to the speech-FFR driven by the carrier and the envelope modulations at higher frequencies. In this study, we examined the spatiotemporal dynamics of the speech-FFR at the fundamental frequency using linear modelling and inverse source localization. We found that the response mainly originates from subcortical sources with major contributions from the midbrain for the two implemented acoustic features, the fundamental waveform and higher frequency envelope modulations. Interestingly, the latter evoked some additional brainstem activity at latencies significantly later than the typical subcortical latency range. Our results confirm the subcortically-dominated nature of the EEG-measured speech-FFR while additional top-down modulation might be evident by recurring brainstem activity.**

***Index Terms*—Frequency-following-response (FFR), EEG, Fundamental frequency, Speech-FFR, Neural source localization**

I. Introduction

The human auditory system has the remarkable capability to track complex acoustic stimuli at finely tuned processing rates and stages. In particular, slow responses following the envelope of speech and synchronizing to the rate of syllables are typically attributed to the auditory cortex [1], [2], as well as the tracking of higher-level linguistic fragments such as phonemes and word boundaries [3].

Early neural processing at subcortical stages of the auditory pathway can, however, capture the temporal fine-structure of speech such as pitch information at processing rates up to a few hundred Hz [4]. These processing pathways can furthermore be influenced by higher-level processes, such as attention, through cortigofugal feedback loops [5].

An important neural response that reflects rapid phase-locked neural processing and has been shown to receive higher-level cognitive feedback, is the frequency-following-response (FFR). For instance, the FFR can be shaped by long- and short-term auditory training [6] as well as musical experience [7] and can even serve as a marker for brainstem plasticity [8]. Its neural sources are thought to emerge mainly from subcortical sources such as the inferior colliculus and the midbrain [9]–[11], although there is rising evidence that cortical generators may contribute to the FFR as well [12], [13].

Recent studies have focused on relating the FFR to more natural stimuli such as continuous speech [14]. The neural response at the fundamental frequency of such continuous speech (speech-FFR) opens up a more comprehensive view on natural speech processing and has been shown to be modulated by attention [15], [16]. Further EEG and MEG studies on the speech-FFR at the fundamental frequency included both the carrier and high frequency envelope modulations and found cortical as well as subcortical contributors to the response [17]–[19].

Here, we follow up on the recent findings of the neural generators to the speech-FFR. We therefore examined the spatiotemporal fine-structure of the speech-FFR captured by EEG and applied inverse source localization to disentangle its neural contributors.

II. Materials and Methods

A. Dataset

The dataset consisted of EEG recordings from 13 young and healthy participants (mean age: 25 years, standard deviation: 3 years) who listened to continuous speech from audiobooks of approximately 40 min in duration read by a single male speaker. EEG was recorded using a 64-channel active-electrode system. The audio was presented in 15 parts with a mean length of 2.6 min. For further details about data aquisition and experimental setup, we refer to [19], [20].

B. Auditory feature processing

Since the speech-FFR is elicited both by the fundamental frequency f_0 as well as its higher harmonics, we computed two corresponding features from the speech signal and each speech part separately (Fig. 1). First, we extracted the fundamental waveform, a time-varying signal oscillating at f_0. The fundamental waveform was computed by applying a band-pass filter (4th order Butterworth filter, zero-phase, non-causal) to the audio signal, with filter boundaries determined by the probabilistic YIN (pYIN) algorithm [21] for estimating f_0 for each of the speech segments. In particular, the lower and upper filter boundaries were defined by the 5th and 95th percentile of the estimated f_0 values (mean lower bound: 76 Hz, mean upper bound: 151 Hz).

The second feature represented the envelope modulation in the speech signal at higher frequencies above f_0. For its computation, we utilized a phenomenological model of the auditory periphery introduced by Tan and Carney [22] to estimate the neural signal in response to the high-frequency speech content. We then filtered and averaged the amplitude of the signal in the frequency bins in the range between 200 and 4,000 Hz using the same band-pass filter as for the fundamental waveform to obtain the high-frequency envelope modulation. Finally, the processed features were concatenated to form one array for each of the two speech features.

C. Linear modelling

To prepare the EEG data for subsequent analysis, we first applied Independent Component Analysis (ICA) to remove artifacts. We then filtered the data in the range of the fundamental frequency between 75 Hz and 150 Hz using the same bandpass filter properties as above.

To assess the neural response to both speech features, we trained a two-feature linear forward model using multivariate regression with ridge regularization. Time lags were chosen between -50 ms (pre-stimulus) and 250 ms (post-stimulus) and the regularization parameter was set to 4.64. The fitted model coefficients are referred to as temporal response functions (TRF). They were averaged over subjects and the mean magnitude response over all channels was computed for each feature separately to assess the corresponding latencies of the evoked activity.

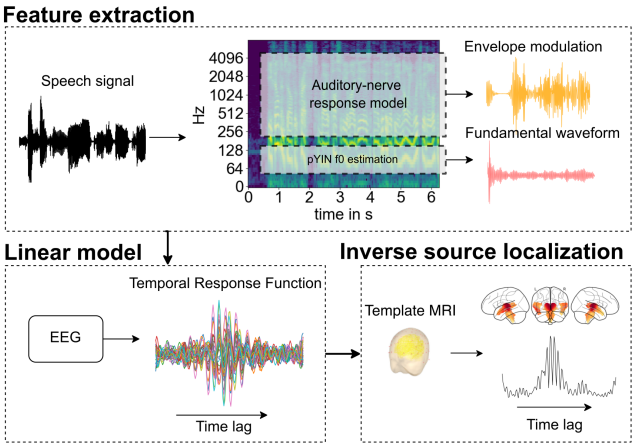

Fig. 1. Acoustic feature processing and source modelling pipeline. The fundamental waveform and envelope modulation features were extracted from the speech signal and used as predictors in a linear model with EEG. Inverse source localization was then applied on the model coefficients (temporal response functions (TRF)) to estimate the sources using a template MRI.

D. Neural source localization

To identify putative sources of the modelled sensor-level activity, we applied inverse source localization on the model coefficients using algorithms from the MNE-Python library [23]. We first computed a discrete volumetric source space based on the 'fsaverage' MRI template from Freesurfer [24] with 2 mm spacing between neighbouring vertices. We constrained the source space to three regions of interest (ROIs) with labels corresponding to the 'aseg' subcortical segmentation: the brainstem ('Brain-Stem', '4th Ventricle'), midbrain ('Right-Thalamus-Proper', 'Left-Thalamus-Proper', 'Left-VentralDC', 'Right-VentralDC', '3rd-Ventricle') and auditory cortex ('ctx-lh-bankssts', 'ctx-lh-superiortemporal', 'ctx-lh-transversetemporal', 'ctx-rh-bankssts', 'ctx-rh-superiortemporal', 'ctx-rh-transversetemporal'). We then estimated the electrical leadfield using a pre-computed volume conductor model and applied the inverse of the leadfield on the subject-averaged channel coefficients with *dynamical Statistical Parametric Mapping* (dSPM) [25] to obtain the estimated source activity for each time lag at each source point. We finally extracted the mean activity from each of the three ROIs yielding the label-specific activation time series.

E. Statistics

Null models were computed for both sensor and source level analysis. To this end, we computed nonsense auditory features by randomly sampling from the actual model features n times (where n is the length of the actual feature). We then computed linear models to relate the EEG data to the nonsense auditory features as described above, as well as source reconstruction of the resulting TRFs. We thus obtained the null magnitudes from the population-average null model as well as a null source activation time series for each ROI label.

To test for statistical significance, we compared the amplitude values from the actual model with 10000 randomly sampled values from the null model at each time lag and counted the number of times the null value exceeded the corresponding model value. The resulting proportion yielded a p-value at each time lag which was then corrected for multiple comparisons using the Bonferroni method. Lags at which the corrected p-value was below 0.05 were deemed statistically significant.

III. Results

We analyzed the speech-FFR through two features of the speech stimuli and as measured from EEG recordings from 13 subjects who listened to continuous speech from audiobooks.

For the fundamental waveform, we found an early neural activity with a broad significant range between -11 ms and 44 ms centered at around 18 ms (Fig. 2A). An additional later contribution could be identified between 64 ms and 74 ms. Topographic analysis of the model coefficients revealed major central-frontal activation at the peak latency of 18 ms and a slightly right-lateralized weaker frontal activity in the later significant contribution at 73 ms.

The envelope modulation feature elicited a slightly later response compared to the fundamental waveform at a latency of 25 ms with significant lags ranging from -1 ms to 49 ms and a central topographic activation pattern (Fig. 2B). A further significant plateau was found between 54 ms and 64 ms indicating right-lateralized temporal activity at 55 ms latency.

To elucidate the neural generators of the evoked activity on source level, we performed source localization on the subject-averaged model coefficients which yielded a source estimate at each discrete point in the constrained source space. We then computed the mean activation time course for each ROI (brainstem, midbrain and auditory cortex) and tested the significance at each time lag against the corresponding bootstrapped null source time course.

For both the fundamental waveform and the envelope modulation, the major activity emerged from sources in the midbrain with peak latencies at 17 ms and 19 ms, respectively (Fig. 3B). In the cortical ROI, smaller activity occurred for both features at time lags which temporally overlapped with major responses in the brainstem and midbrain. These were likely due to leakage as mainly midbrain sources were activated at the corresponding peak latencies (Fig. 3C). In the brainstem ROI, the fundamental waveform elicited an early activity with a significant range from -1 ms to 39 ms and two peak latencies at 5 ms and 23 ms (Fig. 3A, left). Similarly, we found significant activity driven by the envelope modulation from -4 ms to 34 ms peaking at 24 ms. However, specifically for the envelope modulation feature, we identified additional later activity in the brainstem between 50 ms to 56 ms as well as from 73 ms to 78 ms (Fig. 3A, right).

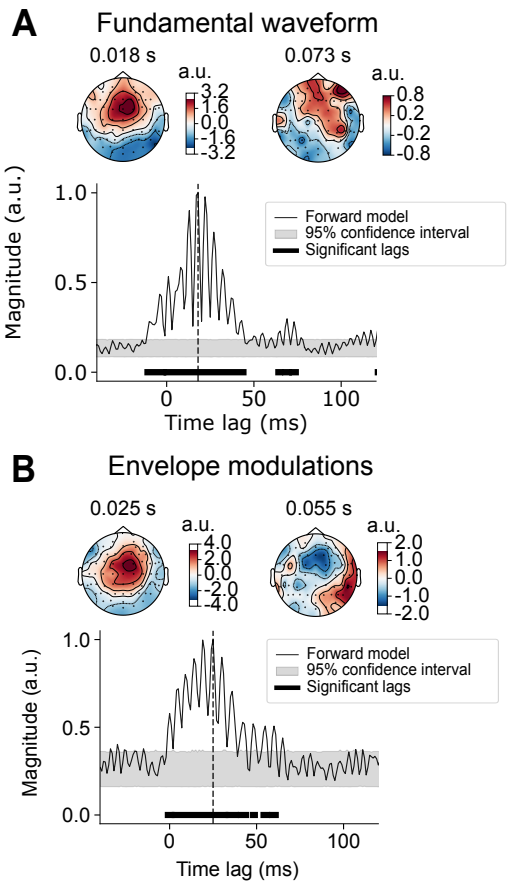

Fig. 2. Sensor-level neural response to the fundamental waveform (A) and envelope modulations (B). The normalised subject- and channel-averaged magnitudes of the linear model coefficients are shown with corresponding amplitude topographies at peak latencies (black dashed lines) and further significant time lags. Significance of neural activity was tested against bootstrapped samples from a null distribution at each time lag ($p < 0.05$, Bonferroni-corrected) yielding a 95% confidence interval (gray-shaded area) and signficant time lags (black bars) where the magnitude values exceeded the chance level.

IV. Discussion

We examined the neural contributors to the speech-FFR at the fundamental frequency using EEG-based inverse source localization combined with sensor-space TRF analysis. Neural activity at the fundamental frequency might be elicited by both the carrier of a speech signal and the envelope modulation at higher [18]. For both features, we found early neural activity dominated by the midbrain. However, we further identified later significant activity in the brainstem that might indicate a putative top-down modulation from higher-level to lower-level stages of the auditory pathway.

On the sensor level, the fundamental waveform elicited a broad neural activity with a peak latency at around 18 ms. Previous studies have found a slightly earlier response to the fundamental waveform of running speech at around 8 ms peak latency [15], [16]. This is likely due to differences in the fundamental frequency of the underlying speaker (female/male/both) and the chosen filter bandwidth applied on the speech waveform and EEG. Here, we used a male speaker with a relatively low mean fundamental frequency of 109.3 Hz and a bandpass filter between 75 Hz and 150 Hz which induced

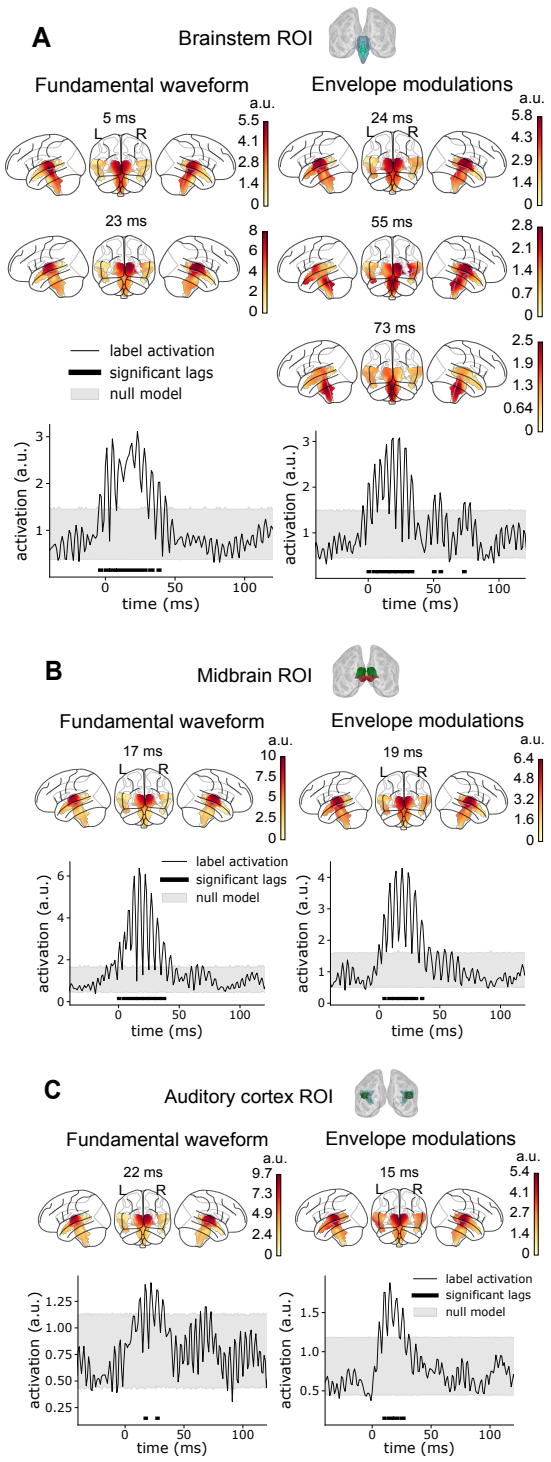

Fig. 3. Source-level neural activity to the fundamental waveform (left) and envelope modulations (right) for three ROIs, brainstem (A), midbrain (B) and auditory cortex (C). The extracted time courses from the source-localised subject-averaged TRF coefficients are shown for each ROI with the corresponding distributed source activity at selected time lags projected onto an average MRI template. Significant time lags (black bars) indicate where the source activity exceeded the 95% chance level (gray-shaded area, Bonferroni-corrected).

higher auto-correlation effects in the TRFs and thus a broader temporal spread of the observed activity [26]. However, despite small variations in the peak latency of the neural response, the identified frontral-central topographic activation pattern strongly corresponds to related studies [9], [19], [26] and points towards a centrally-located subcortical origin.

Indeed, we found major midbrain-driven subcortical activity at the TRF peak latencies when we applied source localization on the TRF amplitudes for both the fundamental waveform and the envelope modulations. This relates to the well-known notion that the scalp-recorded FFR to both short, repeated stimuli (FFR) and continuous natural speech (speech-ABR) is mainly shaped by subcortical generators such as cochlear nuclei, the inferior colliculus and medial geniculate body [11], [13], [27]. Although comparing neural effects from 'classical' FFR studies with synthesized speech tokens and speech-FFR measured from natural speech must still be done with caution, the underlying mechanisms and neural response characteristics tend to be similar [14].

Interestingly, two separate peaks ocurred in the brainstem ROI evoked by the fundamental waveform, an early one at 5 ms and the later main peak at 23 ms. While the second peak is fairly late for a purely brainstem-driven activity and furthermore coincides with spurious stronger activity from the midbrain (peaking at 17 ms), the timing of the first peak corresponds to the modelled latency of the speech-ABR between the auditory nerve and cochlear nucleus in a model of different brainstem stages [27] as well as the EEG-measured response to click and continuous speech stimuli [14].

Due to nonlinearities in the auditory pathway, the response at the fundamental frequency may also emerge from higher harmonics in the speech signal. To capture this effect, we employed a second feature that described the envelope modulation of higher frequencies at f_0. We found a slightly later peak latency at 25 ms compared to the fundamental waveform. We note that the two employed features are partly anti-correlated (Pearson's correlation coefficient: $r = -0.13$), but nonetheless capture mostly different aspects of the neural responses [28].

The topographic pattern at the peak latency as well as major midbrain-driven source activity evidence a mainly subcortical origin of the high-frequency envelope modulation-driven response. However, we identified a later right-lateralized temporal activity to the envelope modulation feature at 55 ms, a latency and corresponding topography typically attributed to cortical sources. Previous studies on the MEG-measured speech-FFR have indeed found cortical involvement at similar latencies elicited by the high-frequency envelope modulation of a speech stimulus [17], [18]. In contrast, source localization revealed recurring brainstem activity at this latency as well as another activation at around 73 ms. Similar findings have been reported in an MEG study on the speech-FFR [17] and might reflect top-down processing from the auditory cortex to subcortical structures underpinning attentional and linguistic modulation of the speech-FFR [15], [16], [19], [29].

Minor cortical activity emerged for both features at latencies

that largely overlapped with the main response from the midbrain. In general, the lack of isolated cortical activity – contrasting the findings from previous MEG measurements of the (speech)-FFR [17], [18], [30] – is presumably due to the intrinsic properties of the underlying measurement modality (EEG). In particular, EEG can capture radial and tangential sources and is thus biased towards activity at deeper sources compared to MEG [31], [32]. Furthermore, the underdetermined inverse problem, that is a disbalance of the number of estimated sources compared to the number of EEG sensors at the scalp, can lead to spurious source activity across the different ROIs and thus requires careful interpretation of the resulting neural activities [33]–[35].

V. Conclusion

Modelling the EEG-measured speech-FFR with TRF-based inverse source localization confirmed that the response is mainly shaped by subcortical sources with major activity in the midbrain. Later brainstem contributions might reflect top-down control driving the modulation of the speech-FFR by higher level processes such as attention and linguistic processing.

References

[1] N. Ding and J. Z. Simon, "Cortical entrainment to continuous speech: functional roles and interpretations," *Frontiers in Human Neuroscience*, vol. 8, May 2014.

[2] G. Hickok and D. Poeppel, "The cortical organization of speech processing," *Nature Reviews Neuroscience*, vol. 8, pp. 393–402, May 2007.

[3] C. Brodbeck, A. Presacco, and J. Z. Simon, "Neural source dynamics of brain responses to continuous stimuli: Speech processing from acoustics to comprehension," *NeuroImage*, vol. 172, pp. 162–174, May 2018.

[4] A. Krishnan, Y. Xu, J. T. Gandour, and P. A. Cariani, "Human frequency-following response: representation of pitch contours in Chinese tones," *Hearing Research*, vol. 189, pp. 1–12, Mar. 2004.

[5] J. A. Winer, "Decoding the auditory corticofugal systems," *Hearing Research*, vol. 207, pp. 1–9, Sept. 2005.

[6] A. Krishnan, Y. Xu, J. Gandour, and P. Cariani, "Encoding of pitch in the human brainstem is sensitive to language experience," *Cognitive Brain Research*, vol. 25, pp. 161–168, Sept. 2005.

[7] P. C. M. Wong, E. Skoe, N. M. Russo, T. Dees, and N. Kraus, "Musical experience shapes human brainstem encoding of linguistic pitch patterns," *Nature Neuroscience*, vol. 10, pp. 420–422, Apr. 2007.

[8] A. Krishnan, J. Swaminathan, and J. T. Gandour, "Experience-dependent Enhancement of Linguistic Pitch Representation in the Brainstem Is Not Specific to a Speech Context," *Journal of Cognitive Neuroscience*, vol. 21, pp. 1092–1105, June 2009.

[9] G. M. Bidelman, "Multichannel recordings of the human brainstem frequency-following response: Scalp topography, source generators, and distinctions from the transient ABR," *Hearing Research*, vol. 323, pp. 68–80, May 2015.

[10] P. Tichko and E. Skoe, "Frequency-dependent fine structure in the frequency-following response: The byproduct of multiple generators," *Hearing Research*, vol. 348, pp. 1–15, May 2017.

[11] B. Chandrasekaran and N. Kraus, "The scalp-recorded brainstem response to speech: Neural origins and plasticity," *Psychophysiology*, vol. 47, pp. 236–246, Mar. 2010.

[12] E. B. J. Coffey, T. Nicol, T. White-Schwoch, B. Chandrasekaran, J. Krizman, E. Skoe, R. J. Zatorre, and N. Kraus, "Evolving perspectives on the sources of the frequency-following response," *Nature Communications*, vol. 10, p. 5036, Nov. 2019.

[13] G. M. Bidelman, "Subcortical sources dominate the neuroelectric auditory frequency-following response to speech," *NeuroImage*, vol. 175, pp. 56–69, July 2018.

[14] R. K. Maddox and A. K. C. Lee, "Auditory Brainstem Responses to Continuous Natural Speech in Human Listeners," *eneuro*, vol. 5, pp. ENEURO.0441–17.2018, Jan. 2018.

[15] A. E. Forte, O. Etard, and T. Reichenbach, "The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention," *eLife*, vol. 6, p. e27203, Oct. 2017.

[16] O. Etard, M. Kegler, C. Braiman, A. E. Forte, and T. Reichenbach, "Decoding of selective attention to continuous speech from the human auditory brainstem response," *NeuroImage*, vol. 200, pp. 1–11, Oct. 2019.

[17] A. Schüller, A. Schilling, P. Krauss, and T. Reichenbach, "The Early Subcortical Response at the Fundamental Frequency of Speech Is Temporally Separated from Later Cortical Contributions," *Journal of Cognitive Neuroscience*, pp. 1–17, Jan. 2024.

[18] J. P. Kulasingham, C. Brodbeck, A. Presacco, S. E. Kuchinsky, S. Anderson, and J. Z. Simon, "High gamma cortical processing of continuous speech in younger and older listeners," *NeuroImage*, vol. 222, p. 117291, Nov. 2020.

[19] M. Kegler, H. Weissbart, and T. Reichenbach, "The neural response at the fundamental frequency of speech is modulated by word-level acoustic and linguistic information," *Frontiers in Neuroscience*, vol. 16, 2022.

[20] H. Weissbart, K. D. Kandylaki, and T. Reichenbach, "Cortical Tracking of Surprisal during Continuous Speech Comprehension," *Journal of Cognitive Neuroscience*, vol. 32, pp. 155–166, Jan. 2020.

[21] M. Mauch and S. Dixon, "PYIN: A fundamental frequency estimator using probabilistic threshold distributions," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Florence, Italy), pp. 659–663, IEEE, May 2014.

[22] Q. Tan and L. H. Carney, "A phenomenological model for the responses of auditory-nerve fibers. II. Nonlinear tuning with a frequency glide," *The Journal of the Acoustical Society of America*, vol. 114, pp. 2007–2020, Oct. 2003.

[23] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, L. Parkkonen, and M. S. Hämäläinen, "MNE software for processing MEG and EEG data," *NeuroImage*, vol. 86, pp. 446–460, Feb. 2014.

[24] B. Fischl, "FreeSurfer," *NeuroImage*, vol. 62, pp. 774–781, Aug. 2012.

[25] A. M. Dale, A. K. Liu, B. R. Fischl, R. L. Buckner, J. W. Belliveau, J. D. Lewine, and E. Halgren, "Dynamic Statistical Parametric Mapping: Combining fMRI and MEG for High-Resolution Imaging of Cortical Activity," *Neuron*, vol. 26, pp. 55–67, Apr. 2000.

[26] J. Van Canneyt, J. Wouters, and T. Francart, "Neural tracking of the fundamental frequency of the voice: The effect of voice characteristics," *European Journal of Neuroscience*, vol. 53, pp. 3640–3653, June 2021.

[27] M. Saiz-Alía and T. Reichenbach, "Computational modeling of the auditory brainstem response to continuous speech," *Journal of Neural Engineering*, vol. 17, p. 036035, June 2020.

[28] M. J. Crosse, G. M. Di Liberto, A. Bednar, and E. C. Lalor, "The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli," *Frontiers in Human Neuroscience*, vol. 10, Nov. 2016.

[29] A. Schüller, A. Schilling, P. Krauss, S. Rampp, and T. Reichenbach, "Attentional Modulation of the Cortical Contribution to the Frequency-Following Response Evoked by Continuous Speech," *The Journal of Neuroscience*, vol. 43, pp. 7429–7440, Nov. 2023.

[30] E. B. J. Coffey, S. C. Herholz, A. M. P. Chepesiuk, S. Baillet, and R. J. Zatorre, "Cortical contributions to the auditory frequency-following response revealed by MEG," *Nature Communications*, vol. 7, p. 11070, Mar. 2016.

[31] S. Baillet, J. Mosher, and R. Leahy, "Electromagnetic brain mapping," *IEEE Signal Processing Magazine*, vol. 18, pp. 14–30, Nov. 2001.

[32] S. Baillet, "Magnetoencephalography for brain electrophysiology and imaging," *Nature Neuroscience*, vol. 20, pp. 327–339, Mar. 2017.

[33] S. Haufe, F. Meinecke, K. Görgen, S. Dähne, J.-D. Haynes, B. Blankertz, and F. Bießmann, "On the interpretation of weight vectors of linear models in multivariate neuroimaging," *NeuroImage*, vol. 87, pp. 96–110, Feb. 2014.

[34] J. G. Samuelsson, N. Peled, F. Mamashli, J. Ahveninen, and M. S. Hämäläinen, "Spatial fidelity of MEG/EEG source estimates: A general evaluation approach," *NeuroImage*, vol. 224, p. 117430, Jan. 2021.

[35] R. Srinivasan, D. M. Tucker, and M. Murias, "Estimating the spatial Nyquist of the human EEG," *Behavior Research Methods, Instruments, & Computers*, vol. 30, pp. 8–19, Mar. 1998.