



Comparing methods for deriving the auditory brainstem response to continuous speech in human listeners

Tong Shan^a, Ross K. Maddox^{a,b,c}

^aDepartment of Biomedical Engineering, University of Rochester, Rochester, NY, United States

^bDepartment of Neuroscience, University of Rochester, Rochester, NY, United States

^cKresge Hearing Research Institute, Department of Otolaryngology-Head & Neck Surgery, University of Michigan, Ann Arbor, MI, United States

Corresponding Author: Ross K. Maddox (rkmaddox@med.umich.edu)

ABSTRACT

Several methods have recently been developed to derive the auditory brainstem response (ABR) from continuous natural speech, facilitating investigation into subcortical encoding of speech. These tools rely on deconvolution to compute the temporal response function (TRF), which models the subcortical auditory pathway as a linear system, where a nonlinearly processed stimulus is taken as the input (i.e., regressor), the electroencephalogram (EEG) data as the output, and the ABR as the impulse response deconvolved from the recorded EEG and the regressor. In this study, we analyzed EEG recordings from subjects listening to both unaltered natural speech and synthesized “peaky speech.” We compared the derived ABR TRFs using three regressors: the half-wave rectified stimulus (HWR) from [Maddox and Lee \(2018\)](#), the glottal pulse train (GP) from [Polonenko and Maddox \(2021\)](#), and the auditory nerve modeled response (ANM; [Zilany et al. \(2014\)](#); [\(2009\)](#)) used in [Shan et al. \(2024\)](#). Our evaluation focused on the signal-to-noise ratio, prediction accuracy, efficiency, and practicality of applying each regressor in both unaltered and peaky speech. The results indicate that the ANM regressor with peaky speech provides the best performance, with the ANM for unaltered speech and the GP regressor for peaky speech close behind, whereas the HWR regressor demonstrated relatively poorer performance. There are, thus, multiple stimulus and analysis tools that can provide high-quality subcortical TRFs, with the choices for which to use dictated by experimental needs. The findings in this study will guide future research and clinical use in selecting the most appropriate paradigm for ABR derivation from continuous, naturalistic speech.

Keywords: auditory brainstem response, natural stimuli, speech, EEG, temporal response function

1. INTRODUCTION

Speech is a complex sound encountered daily and plays a fundamental role in human communication. It is, thus, essential to understand the process through which the human brain translates speech from its basic encoding by the auditory periphery to higher level processing in the cortex. Subcortical structures have been proven to be critical in this auditory processing chain, notably in the encoding of vowels and processing speech in noisy envi-

ronments ([Carney et al., 2015](#)). The auditory brainstem response (ABR) serves as a key metric for subcortical auditory neuroscience research as well as clinical audiology. Traditionally, the ABR is characterized by a stereotypical evoked potential elicited by brief stimuli such as clicks, tones, or chirps ([R. F. Burkard et al., 2007](#); [Picton et al., 1974](#)) through electroencephalography (EEG) recording. This evoked potential is observed in the first ~10 ms post-stimulus, consisting of components that

Received: 3 June 2024 Revision: 21 April 2025 Accepted: 25 April 2025 Available Online: 14 May 2025



The MIT Press

© 2025 The Authors. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

Imaging Neuroscience, Volume 3, 2025
<https://doi.org/10.1162/IMAG.a.19>

reflect different stages of the auditory pathway according to their latency. Specifically, Waves I, III, and V are of particular interest, corresponding to the responses of the auditory nerve, cochlear nucleus, and inferior colliculus and lateral lemniscus, respectively (Picton et al., 1974).

Expanding upon this foundation, investigations into the brainstem's response to speech via complex ABR (cABR) have been undertaken (Krizman et al., 2010; Musacchia et al., 2007; Skoe & Kraus, 2010). These studies demonstrate that short speech vowels elicit a transient onset and a frequency following response (FFR) corresponding to the voiced part. However, the cABR method has limitations in its controversial neural sources (Coffey et al., 2016) and potential neural adaptation due to the repetitiveness of the speech stimuli used (i.e., repeated tokens of vowels or syllables).

Recently, studies have developed several methods for detecting the brainstem response to continuous, non-repetitive speech, thus offering a more ecologically valid approach (at the expense of some experimental control over the stimuli) and potential clinical use (Bachmann et al., 2021; Forte et al., 2017; Kulasingham, Bachmann, et al., 2024; Maddox & Lee, 2018; Polonenko & Maddox, 2021; Shan et al., 2024). One such technique involves extracting the fundamental waveform from the speech and cross-correlating the waveform with the EEG signal (Forte et al., 2017). This method yields a broad peak around 9 ms primarily originating from the inferior colliculus but lacking finer components showing distinct activity from earlier auditory stages. Another set of studies are based on a deconvolution method that was proposed by Maddox and Lee (2018). The result of this deconvolution is a temporal response function (TRF), which has been used extensively to study cortical responses to natural stimuli (Di Liberto et al., 2015; Ding & Simon, 2012; Lalor & Foxe, 2010; Lalor et al., 2009) and provides superior responses to the fundamental waveform-based methods (Bachmann et al., 2021). An encoding model was proposed as depicted in Figure 1a: the stimulus (more specifically, an acoustical feature derived from the stimulus) acted as the input (i.e., regressor, x in Figure 1a), the recorded EEG signal as the output y , and the ABR as the impulse response of a linear system that transforms x into y .

A subcortical TRF whose morphology matches the click-evoked response is important for two reasons. First and most important is that the morphology of the click ABR can be linked to specific subcortical nuclei. This means that the TRF waveforms can be interpreted using the same framework. Second is that the TRF weights are a model of the impulse response of a system. A click is a real-world implementation of an impulse, so an accurate model of the subcortical system should respond with an ABR to an impulse input. Such a response indicates the

model is working in a more intuitive way than comparing correlation coefficients.

A series of studies have offered improvements for deconvolution methods to compute the ABR (i.e., the subcortical TRF). The initial study by Maddox and Lee (2018) utilized half-wave rectification of the stimulus as the regressor (HWR, Figure 1b) as a simple simulation of cochlear nonlinearity. It was able to derive the speech ABR with a distinct Wave V that is highly correlated with the click-evoked ABR. Following this, Polonenko and Maddox (2021) proposed using “peaky speech,” a resynthesized speech stimulus that was made impulse-like by aligning the phase of the speech harmonics at the time of glottal pulses. The regressor used was a train of impulses placed at the times of the glottal pulses (GP regressor, Figure 1c). This method provided distinct earlier ABR components waves I and III in addition to wave V and enabled simultaneous ABR measurements from separate frequency bands. Shan et al. (2024) further extended deconvolution methods by incorporating a detailed computational model (Zilany et al., 2009, 2014) that simulates the neural representation of the auditory periphery, converting the stimulus waveform into an auditory nerve modeled response to be used as the regressor (ANM, Figure 1d). Kulasingham, Bachmann, et al. (2024) compared the ANM to several simpler regressors, finding that a more efficient model provides good responses as long as it recapitulates the adaptation present in the auditory nerve. The ANM method, like the peaky speech with the GP regressor, also yields ABRs with early wave components, and improves the speech ABR's signal-to-noise ratio (SNR) over the HWR regressor. Moreover, this ANM method is generalizable to other natural sounds, including music.

In this study, we aimed to compare ABR deconvolution using the HWR, GP, and ANM regressors. By examining the signal-to-noise ratio (SNR), prediction accuracy, efficiency, and practicality of each method in different scenarios, we hope to offer guidance on determining the most appropriate approach for deriving ABRs from natural speech as well as other complex sounds for a variety of experimental or clinical uses.

2. MATERIALS AND METHODS

2.1. EEG dataset

The data analyzed in this study were obtained from a broadband peaky speech experiment previously conducted by Polonenko and Maddox (2021). In that experiment, EEG was recorded from 22 normal hearing subjects (aged 18–32 years, mean \pm SD of 23.0 ± 3.6 years) while they listened to the audiobook *The Alchemyst* (Scott,

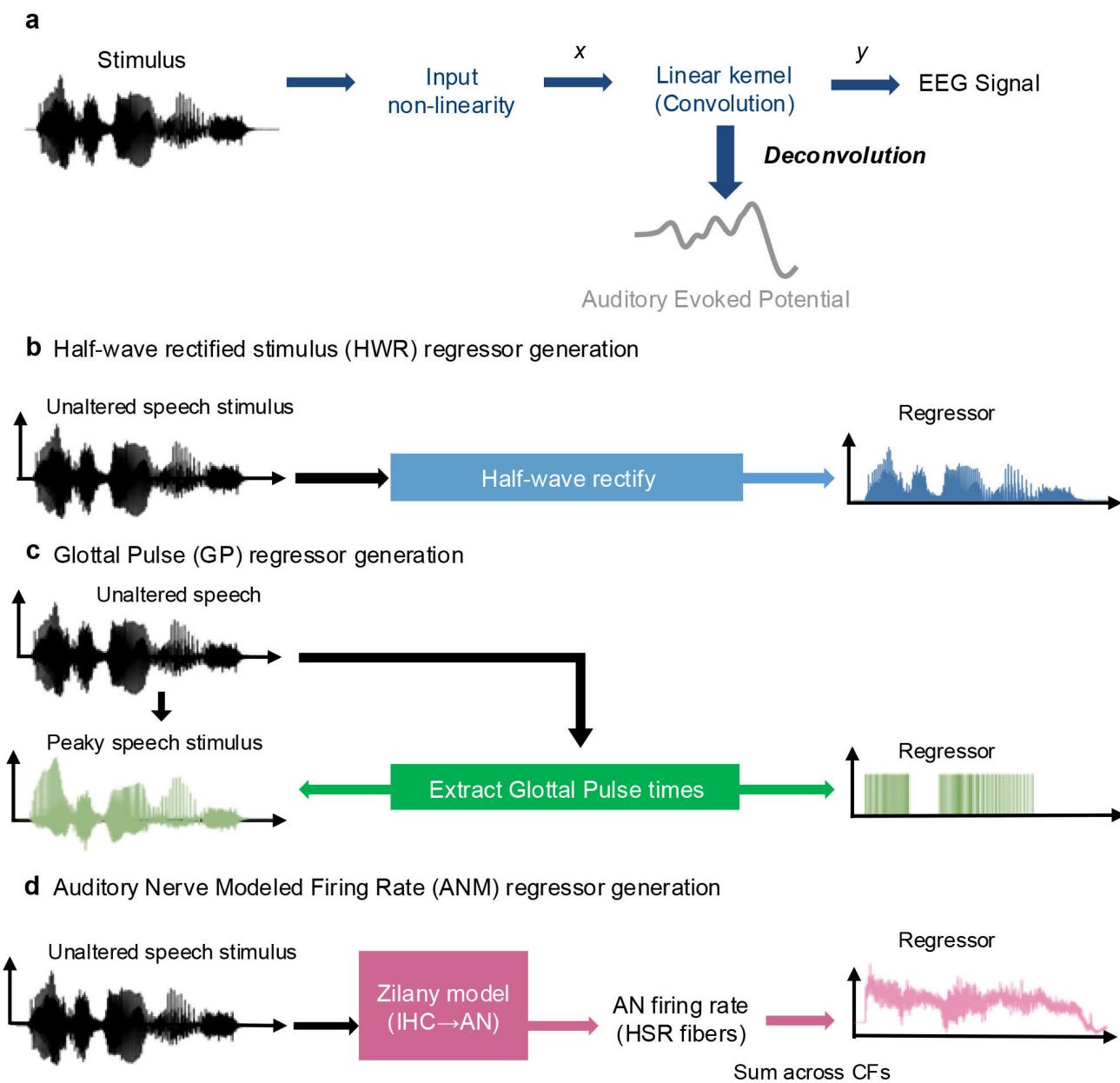


Fig. 1. The encoding model using deconvolution method and the regressors that were used. (a) The deconvolution paradigm for computing the TRF. (b) The half-wave rectified stimulus regressor (HWR). (c) The peaky speech waveform and the Glottal Pulse train regressor (GP). (d) The auditory Nerve Modeled firing rate regressor (ANM). IHC = inner hair cell, AN = auditory nerve, HSR = high spontaneous rate, CF = characteristic frequency.

2008), which was narrated by a male voice, as detailed in Polonenko and Maddox (2021). The silent pauses exceeding 0.5 s in the audiobook had been truncated, and the audiobook was segmented to 40 excerpts, each lasting 64 s. The recording time was 42 min and 40 s for each stimulus condition. During the experiment, subjects passively listened to the speech stimuli over ER-2 insert earphones at an average sound pressure level of 65 dB.

The EEG signal capturing subcortical activity (used to compute the ABR) was recorded using BrainVision's passive Ag/AgCl electrodes. These electrodes were placed

at the frontocentral position (FCz in the 10-20 system, active non-inverting), on the left and right earlobes (inverting references), and at the frontal pole (Fpz, ground). The electrodes were connected to an ActiCHamp system with the signal sampled at 10 kHz and high-pass filtered at 0.1 Hz. The recording process also applied a causal, fourth-order lowpass filter at 1/3 Nyquist (1667 Hz). Subsequent offline preprocessing included applying a high-pass filter at 1 Hz to remove any slow drifts, and a notch filter at 60 Hz along with its first three odd harmonics to reduce power line noise.

2.2. Stimuli

In the original dataset, subjects listened to three stimulus conditions. However, for the purpose of this study, we focused on analyzing only two of those conditions: 1) unaltered speech, 2) re-synthesized broadband peaky speech. The re-synthesized peaky speech was designed to make the speech audio impulse-like by aligning the phase of the harmonics at the time of glottal pulses. This design aimed to elicit brainstem responses similar to those elicited by clicks, thereby evoking canonical ABRs while still preserving the intelligibility of the speech with minimal perceptible differences from the unaltered version. For a detailed explanation of the peaky speech synthesis process and audio examples, see [Polonenko and Maddox \(2021\)](#).

2.3. ABR derivation

2.3.1. Deconvolution model for ABR

As described in [Maddox and Lee \(2018\)](#), [Polonenko and Maddox \(2021\)](#), and [Shan et al. \(2024\)](#), an encoding model of the ABR was defined as shown in [Figure 1a](#). The speech stimuli were processed differently to isolate a given stimulus feature (i.e., regressor) to be used as the input x , while the EEG signal was the output y , and the ABR was the impulse response of a linear system and determined through deconvolution. The computation was performed in the frequency domain for efficiency:

$$\text{response} = F^{-1} \left\{ \frac{\sum_n b_n X_n^* Y_n}{\frac{1}{N} \sum_n X_n^* X_n} \right\},$$

where *response* denotes the derived impulse response (i.e., the ABR), X_n the Fast Fourier transform (FFT) of the regressor for trial n , Y_n the FFT of EEG signal for trial n , $*$ the complex conjugate, F^{-1} the inverse FFT, b_n the weight for trial n (see below), N the total number of trials, and n the trial index.

When computing the average response, a Bayesian-like process ([Elberling & Wahlgreen, 1985](#)) was used to account for variations in noise level, so that noisier trials were weighted less. The EEG recording from each trial was weighted by its inverse variance, $\frac{1}{\sigma_n^2}$, relative to the sum of the inverse variances of all trials:

$$b_n = \frac{\frac{1}{\sigma_n^2}}{\sum_m \frac{1}{\sigma_m^2}}.$$

2.3.2. Three regressors

We compared the three regressors from previous three studies:

1) Half-wave rectified stimulus (HWR; [Fig. 1b](#))

The half-wave rectified stimulus regressor was generated by first taking the positive values of the stimulus waveform and downsampling it to 10 kHz. This positive component of the stimulus was then used as the input to the encoding model (i.e., x), denoted as HWR. Then, the same process was applied, but with the original stimulus inverted so that the negative values (now positive) were used, and downsampled as before. Deconvolution was performed independently using both the positive and negative components as inputs. The final ABR response for each epoch and each subject was computed by averaging the responses to the positive and negative components.

2) Glottal Pulse (GP; [Fig. 1c](#))

The glottal pulse times were initially extracted from the speech stimuli using speech processing software, *PRAAT* ([Boersma, 2011](#)) when the peaky speech stimuli were constructed. The sequence of impulses that occurred at the glottal pulse times in the peaky speech stimuli was then used as the input to the encoding model, denoted as GP.

3) Auditory Nerve Model firing rate (ANM; [Fig. 1d](#))

A computational auditory periphery model created by [Zilany et al. \(2009\)](#), updated in [Zilany et al. \(2014\)](#), and adapted for Python ([Rudnicki et al., 2015](#)) was utilized to generate simulated auditory neural responses. It was previously shown to be able to account for the peripheral nonlinearity effects ([Kulasingham, Bachmann, et al., 2024](#); [Shan et al., 2024](#)). The speech stimuli were upsampled to 100 kHz according to the model's requirement and converted to a pressure waveform (measured in pascals) at 65 dB SPL and used as inputs to the ANM model. We set the characteristic frequency (CF) ranging from 125 Hz to 16 kHz spaced at 1/6 octave intervals. The auditory nerve firing rate was then summed across all CFs of high spontaneous rate fibers and downsampled to match the EEG sampling rate of 10 kHz so it could be utilized as the regressor, denoted as ANM. Positive and negative polarities of the speech stimuli were used to derive two responses that were averaged to get the final ABR as described in [Shan et al. \(2024\)](#).

The three regressors were used for both the unaltered and the peaky speech conditions. Although the GP regressors were intended for use with peaky speech and

have limited effectiveness as representative features for unaltered speech, they still capture some acoustic information on the timing of glottal pulses in unaltered speech.

2.4. Performance metrics and statistical analysis

2.4.1. Response signal-to-noise ratio (SNR)

To evaluate the quality of the derived ABRs, we estimated the broadband SNRs of each waveform as described in previous studies (Maddox & Lee, 2018; Polonenko & Maddox, 2021; Shan et al., 2024) using the following equation

$$SNR = 10 \left[\frac{\sigma_{S+N}^2 - \sigma_N^2}{\sigma_N^2} \right],$$

where σ_{S+N}^2 is the variance of the ABR waveform measured within the time interval of 0 to 15 ms, and σ_N^2 is the noise variance computed by averaging the variance across each non-overlapping 15 ms segment within the pre-stimulus baseline period, spanning from -1000 to -500 ms. Therefore, subtracting σ_N^2 from σ_{S+N}^2 in the numerator offers an estimate of the signal variance σ_S^2 , which is then divided by the noise variance and log transformed and scaled to estimate SNR in decibels.

SNR was analyzed through a repeated-measures ANOVA followed by a post-hoc pairwise t-test to compare the three regressors.

We also estimate the SNR per frequency. This was done similarly as was done in broadband SNR. But here the power spectral density of the response was used instead of the variance for each frequency bin (f).

$$SNR(f) = 10 \left[\frac{P_{S+N}(f) - P_N(f)}{P_N(f)} \right],$$

where $P_{S+N}(f)$ is the power of the ABR waveform measured within the time interval of 0 to 15 ms in frequency bin f , and $P_N(f)$ is the noise variance computed by averaging the variance across each non-overlapping 15 ms segment within the pre-stimulus baseline period, spanning from -1000 to -500 ms in frequency bin f . The power spectral density of the signals (P) was computed using `psd_array_multitaper` function in `mne` package (Larson et al., 2023) with the first Slepian window with a bandwidth of 67 Hz (Slepian, 1978). At some frequency, bin $P_N(f)$ was higher than $P_{S+N}(f)$, making the result undefined (i.e., the log of a negative number). That is why some lines are broken when plotted.

2.4.2. Time required to obtain robust responses

We were interested in how long it took to record data in order to get a robust ABR using each of the three regres-

sors. For each subject, we calculated the broadband SNRs of ABRs, using the previously mentioned formula, across a recording duration ranging from 1 to 42 min. We then reported the cumulative proportion of subjects who achieved an ABR with an SNR of at least 0 dB throughout the recording process, as in the original peaky speech study (Polonenko & Maddox, 2021).

2.4.3. Correlation between the predicted and the real EEG

To compare the power of the regressors to predict EEG, we used the responses to predict the EEG and calculated the correlation coefficient between the predicted EEG and the real EEG data, as in our previous study (Shan et al., 2024). The predicted EEG were generated by utilizing the ABRs from each regressor as kernels (full kernel: [0, 200] ms time range; subcortical kernel: [0, 15] ms time range), which were then convolved with the corresponding stimulus's regressors. We then calculated the Pearson correlation coefficient between the predicted and real EEG data as a performance metric for each regressor.

2.4.4. Spectral coherence

The ability of the regressors to predict EEG across different frequencies was evaluated using spectral coherence analysis, as outlined in Shan et al. (2024). This approach served as a normalized correlation between the predicted EEG and the real EEG data but is split across various frequency bins. To determine spectral coherence, the predicted EEG and the real EEG data were sliced into segments of specific window sizes (0.2 s in this study), which then determined the frequency bins. The coherence of each of these frequency bins was computed as the following equation

$$C_{xy}(f) = \frac{E[X_i^*(f) Y_i(f)]}{\sqrt{E[X_i^*(f) X_i(f)] E[Y_i^*(f) Y_i(f)]}},$$

where $C_{xy}(f)$ denotes the coherence between signal x and y at frequency bin f , $E[\cdot]$ is the expected value across slices, * the complex conjugate, $X_i(f)$ the FFT for predicted EEG slice i in frequency bin f , and $Y_i(f)$ the FFT for real EEG data slice i in frequency bin f .

To estimate the noise floor of the spectral coherence, we shuffled the order of the predicted EEG and real EEG data and calculated the spectral coherence for these mismatched trials. The median coherence value from these mismatched trials served as the noise floor.

To compare the performance of the three regressors in spectral coherence analysis, we computed the mean of

the absolute value of the spectral coherence across three specific frequency bands for each regressor. These three frequency bands—[0, 25] Hz, [25, 85] Hz, and [85, 135] Hz—were selected based on findings from a previous study indicating superior performance of ANM in these ranges compared to HWR (Shan et al., 2024).

2.4.5. Statistical test

Data were checked and confirmed for normality using the Shapiro–Wilk test for any following parametric test. To compare the performance metrics of the three regressors, mixed-effects linear regression models were constructed (formula below) in python, using stimulus condition, regressor and their interaction as the fixed effects and subject as random effect.

$$\begin{aligned} \text{Performance metric} \sim & \text{ condition} + \text{regressor} \\ & + \text{condition} \times \text{regressor} + (1 | \text{subject}) \end{aligned}$$

To compare the metrics within each stimulus condition, repeated-measure ANOVAs followed by a pairwise post-hoc paired t-test with Holm-Bonferroni correction were used. The performance metrics used in these statistical tests were SNR analysis, broadband prediction accuracy (Pearson correlation), and the mean absolute value of the spectral coherence from the three frequency bands, as described above.

3. RESULTS

The data analyzed in this study were obtained from a peaky speech experiment previously conducted by Polonenko and Maddox (2021). Data were collected under a protocol approved by the University of Rochester Research Subjects Review Board (#1227). This dataset includes EEG recordings from 22 subjects with normal hearing who were passively listening to an English audio-

book under two conditions: unaltered and peaky speech. In the Results section, we present the ABRs derived from the three different regressors and assess their quality using various quantitative metrics. Following this, we introduce a novel approach designed to enable a more equitable comparison of time-domain responses derived from these spectrally different regressors.

3.1. GP and ANM regressors yield quicker and more robust ABR

Using the deconvolution method with the regressors depicted in Figure 1, we obtained the ABRs for both unaltered and peaky speech from the three regressors. Illustrated in Figure 2 are the responses derived from HWR (Fig. 2a), GP (Fig. 2b) and ANM (Fig. 2c). By looking at the general waveforms, it is apparent that the GP for peaky speech condition and the ANM for both conditions exhibit better ABR morphology.

The ABR derived from HWR shows a broad wave V for both stimulus conditions (Fig. 2a; see Fig. S1 for individual responses), consistent with findings from previous studies (Maddox & Lee, 2018; Polonenko & Maddox, 2021). The ABR derived from GP for peaky speech has a distinct and narrow wave V at around 7.2 ms along with an early component (Wave I) at around 3.2 ms. The GP regressor is not designed for unaltered speech, but since it captures limited acoustical representation for the speech at the glottal pulse time, a much smaller wave V is still observable (Fig 2b; see Fig. S2 for individual responses). Note that we ran this regressor-stimulus combination for completeness, but we did not expect high-quality responses from it. The ANM regressor yields clear ABRs for both unaltered and peaky speech with very similar waveforms and high consistency across subjects. Early components (Wave I and Wave III) were present in the waveforms, in addition to Wave V (Fig. 2c; see Fig. S3 for individual responses). A broader time window,

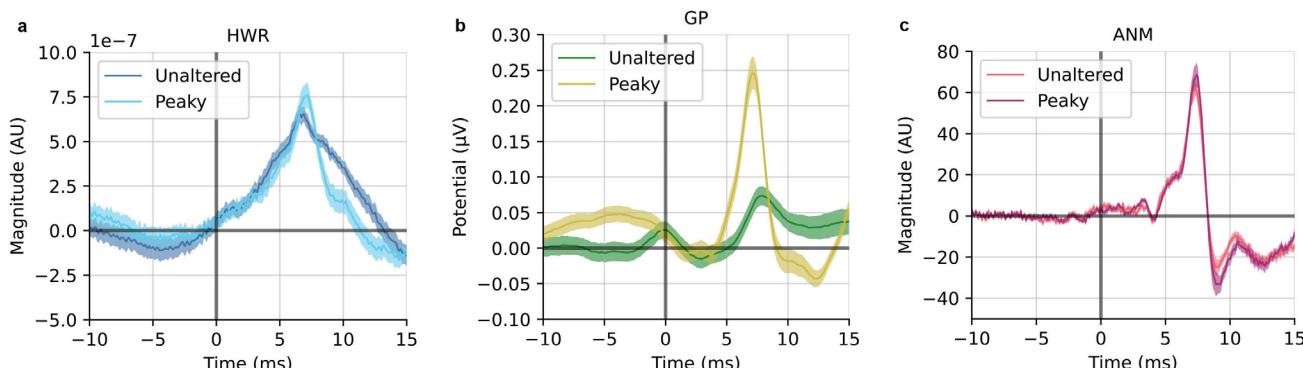
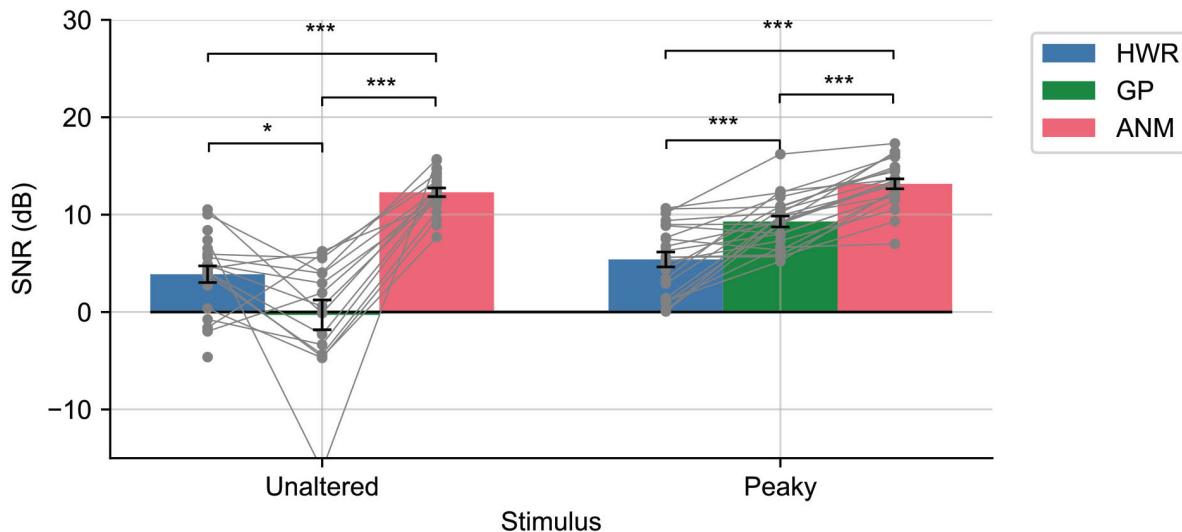


Fig. 2. Grand averaged ABR waveforms for unaltered and peaky speech derived from HWR (a), GP (b), and ANM (c) regressor. Shaded area shows ± 1 SEM ($n = 22$).

a



b

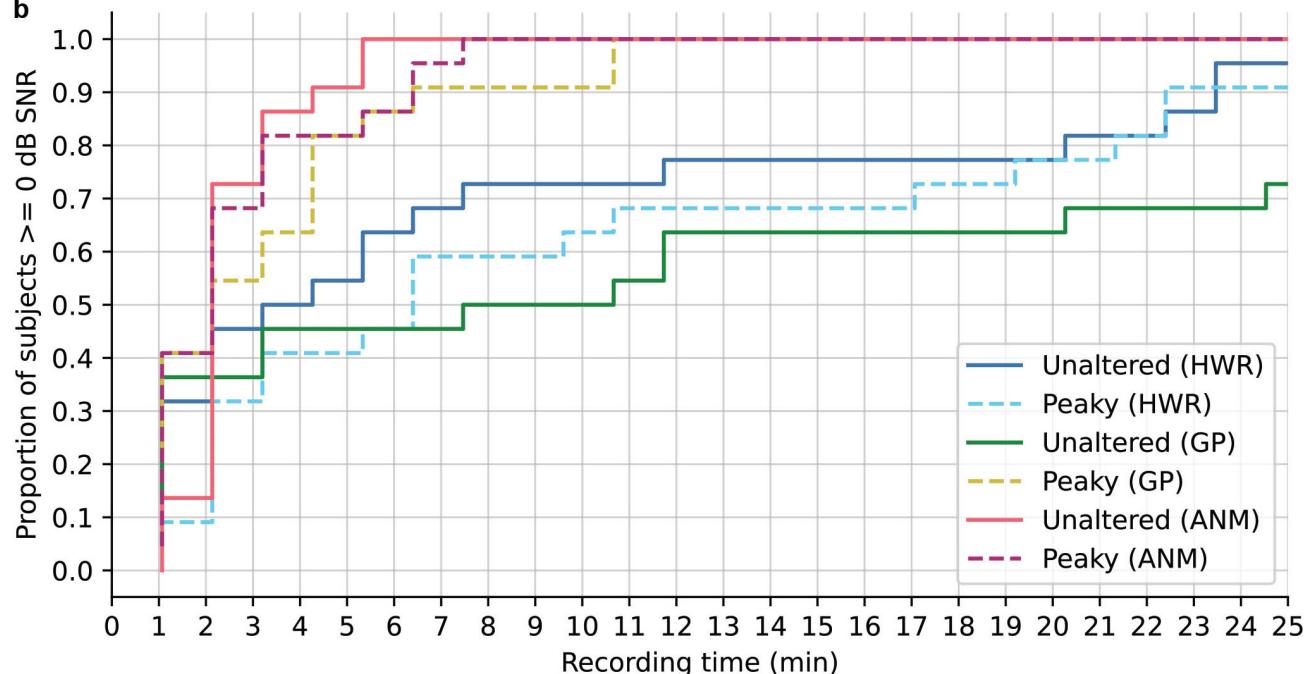


Fig. 3. SNR analysis for the derived ABRs. (a) The averaged SNR of the ABR for unaltered and peaky speech derived from the three regressors. The bar represents the averaged SNR across subjects, and the grey dots with lines are the SNRs for each individual subject. (b) The cumulative proportion of subjects that has ABR SNR ≥ 0 dB as a function of recording time.

including cortical responses ($[-50, 300]$ ms time window), was shown in Figure S4.

We then performed an analysis of the SNR for the ABR waveforms derived from the three regressors within the 0 to 15 ms time window (see Materials and Methods for details of SNR computation). The analysis revealed significant variability in SNR across the regressors for both unaltered and peaky speech ($p < 0.001$; repeated-measures ANOVA). The results are shown in Figure 3a.

For unaltered speech, the ANM regressor demonstrated the highest SNR of 12.29 ± 0.44 (mean \pm SEM), which was significantly better than the SNR obtained with HWR, which averaged 3.90 ± 0.86 ($p < 0.001$; two-tailed paired t-test, Holm-Bonferroni corrected). As expected, both ANM and HWR showed higher SNR than GP in this condition ($p < 0.05$; two-tailed paired t-test, Holm-Bonferroni corrected). For the peaky speech condition, the SNR was the greatest for ANM (13.17 ± 0.51), followed by GP

(9.25 ± 0.56) and HWR (5.31 ± 0.77) in that order. Post-hoc pairwise comparison further showed significant differences between each regressor ($p < 0.001$; two-tailed paired t-test, Holm-Bonferroni corrected). The effect was further confirmed by a mixed-effect linear model (see Table S1 for details). A similar trend was observed when extending the analysis to a time range of 0 to 30 ms for the derived waveforms (Fig. S5).

We were interested in how efficient each regressor is in deriving the ABR by measuring the time required for subjects to achieve a good response with SNR greater than 0 dB. As shown in Figure 3b, for unaltered speech, all subjects reach 0 dB SNR within 6.4 min when using the ANM regressor. Conversely, when using the HWR and GP regressors for the duration of the 42-min experiment, only 95% and 86% of subjects reached 0 dB, respectively. With peaky speech, it took 35.2 min for all subjects to reach 0 dB using HWR, while it only took 11.7 min with GP and 8.5 min with ANM. These indicate that the ANM is efficient in both conditions, and GP exhibited superior performance in peaky speech. Both the ANM and GP outperformed HWR.

3.2. ABR derived from GP and ANM regressor can better predict EEG

In line with common practices in cortical TRF studies (Crosse et al., 2016; David et al., 2007), we conducted a Pearson correlation analysis to evaluate the accuracy of EEG signal prediction against real EEG recordings utilizing the waveform derived by each regressor. As in the previous study (Shan et al., 2024), we used the derived waveforms from the short time range of [0, 15] ms as a kernel with emphasis on subcortical encoding, which was then convolved with the regressors to generate the predicted EEG. The prediction accuracies (i.e., the Pearson's r) were low since the later, slower cortical component of the EEG was not part of the model (but were still present in the signal). There were also no differences among the regressors ($p = 0.257$ and $p = 0.099$, respectively; repeated-measures ANOVA). However, a distinct divergence among regressors emerged upon applying a high-pass filter at 40 Hz to the EEG signals to de-emphasize slower cortical activity, significant for both speech conditions ($p < 0.001$; repeated-measures ANOVA; Fig. 4a). Specifically, in the unaltered speech condition, we again observed that both HWR and ANM demonstrated better accuracy compared to GP ($p < 0.001$; two-tailed paired t-test, Holm-Bonferroni corrected). Additionally, ANM exhibited an advantage over HWR ($p < 0.001$; two-tailed paired t-test, Holm-Bonferroni corrected). In the peaky speech condition, GP and ANM both outperformed HWR ($p < 0.001$; two-

tailed paired t-test, Holm-Bonferroni corrected), and ANM also showed significantly better accuracy than GP ($p < 0.001$; two-tailed paired t-test, Holm-Bonferroni corrected). Notably, mixed-effects linear regression showed that the correlation coefficients achieved from the two stimulus conditions were significantly different, with the peaky speech having higher coefficients ($p = 0.049$ for stimulus condition variable; see Table S2 for the detailed model results.).

We also assessed a broadband correlation coefficient using a [0, 200] ms kernel. However, there was no significant effect of regressor type in either speech condition ($p = 0.097$ for unaltered and $p = 0.44$ for peaky speech; repeated-measures ANOVA; Fig. S6). This broadband measure reflected a large portion of signals from cortical activity, indicating a consistent predictive performance across regressors in a later component of the auditory potentials.

It is possible that one regressor is a better predictor of slower response components, while another better explains higher-frequency portions of the EEG. These differences would be washed out in the broadband correlation analysis. To address this possibility, we conducted a spectral coherence analysis to evaluate the models' prediction accuracy across frequency, similar to the approach utilized in Shan et al. (2024). This analysis quantifies the normalized similarity between the predicted and actual EEG data at each frequency, providing detailed insights into model performance on a per-frequency basis (see Materials and Methods for details). Figure 4c and 4e highlights the superiority of the ANM regressor over GP and HWR in unaltered speech and the advantage of ANM and GP over HWR in the peaky speech condition. These coherence trends are consistent with the comparative superiority of ANM for unaltered speech and ANM and GP in peaky speech seen with other metrics.

Shan et al. (2024) identified significant advantages of the ANM regressor over HWR particularly in the frequency ranges centered around 50 Hz and 100 Hz. Therefore, we further break down the coherence comparison into three frequency bands: [0, 20] Hz, [40, 60] Hz, and [80, 120] Hz (Fig. 4d and 4f). We then conducted a statistical comparison of the mean coherence from the three frequency bands across the regressors. We found that the ANM regressor outperformed the other two regressors in all three bands for unaltered speech ($p < 0.01$; two-tailed paired t-test, Holm-Bonferroni corrected; Fig. 4d). In the peaky speech condition, both ANM and GP exhibited superior performance compared to HWR in [80, 120] Hz, and ANM was slightly superior compared to GP (ANM vs. HWR, $p < 0.001$; GP vs. HWR, $p < 0.001$; ANM vs. GP, $p = 0.02$; two-tailed paired t-test, Holm-Bonferroni corrected). The ANM regressor was also found to show

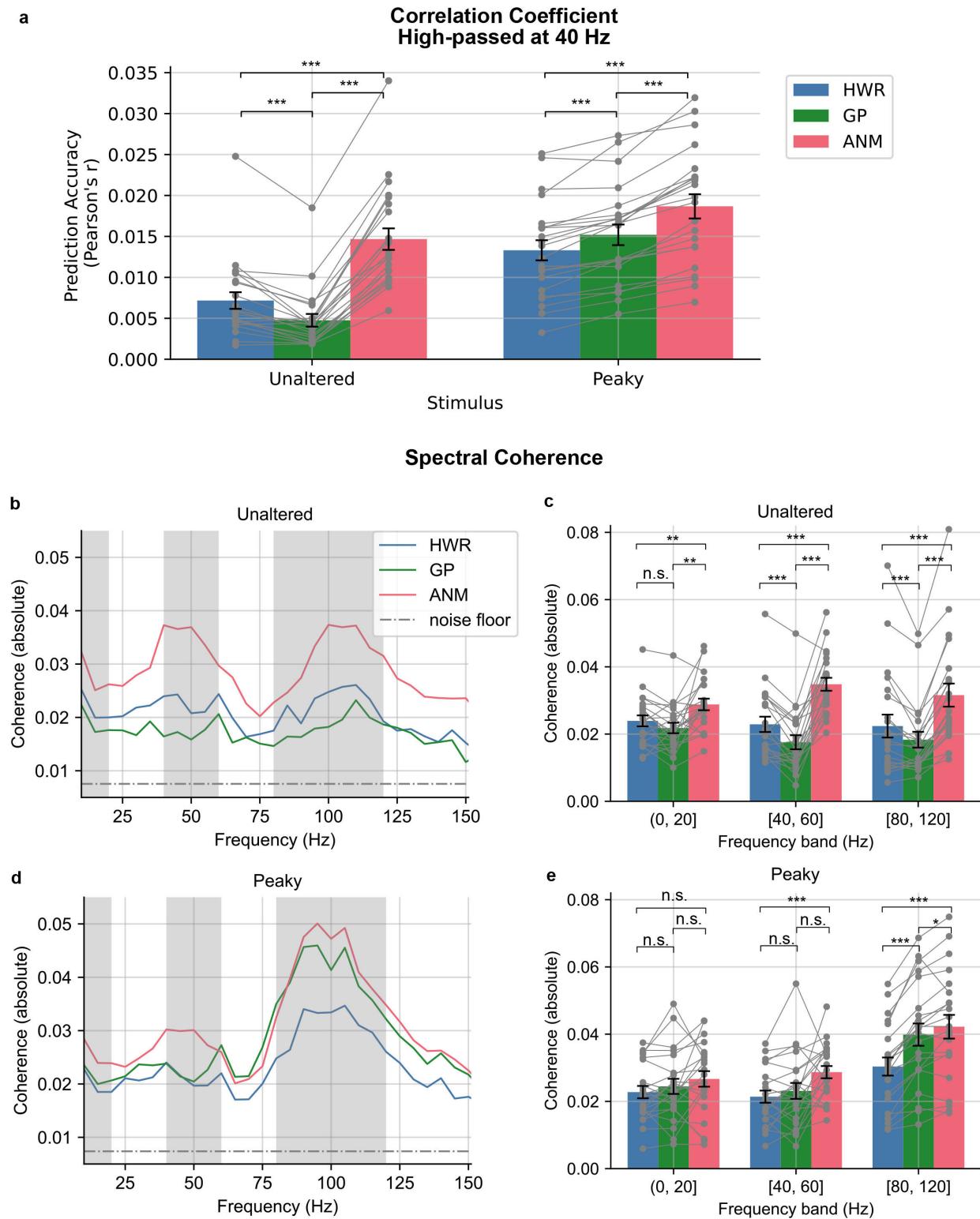


Fig. 4. Prediction accuracy as the correlation coefficient and spectral coherence between predicted and real EEG data. (a) Broadband Correlation coefficient of high-pass filtered EEG with subcortical kernel (0–15 ms). The bars are averaged accuracy across subjects with error bars showing ± 1 SEM, and the grey dots with lines are for each individual subject. (b) The mean absolute value of spectral coherence for unaltered speech. (c) The mean coherence in three frequency bands for unaltered speech across subjects. (d) The mean absolute value of spectral coherence for peaky speech. (e) The mean coherence in three frequency bands for peaky speech across subjects. The dash-dotted lines in (b) and (d) indicate the noise floor. The shaded grey areas indicate the frequency bands analyzed in (c) and (e). The bars in (c) and (e) are averaged coherence across subjects, with error bars showing ± 1 SEM and the grey dots with lines are for each individual subject. (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$)

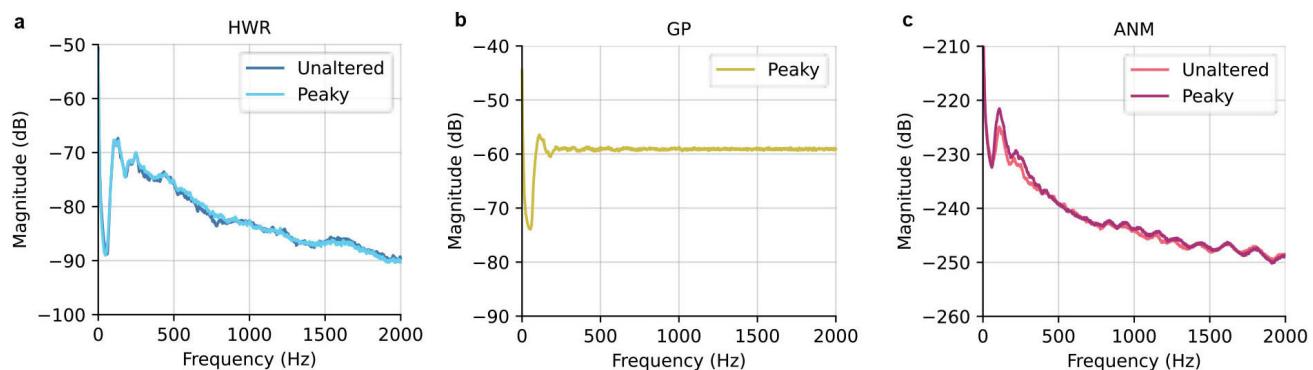


Fig. 5. Averaged power density spectrum with Welch estimate for HWR (a), GP (b), and ANM (c). (The GP regressor for peaky and unaltered is the same.)

higher coherence than HWR in [40, 60] Hz band, but not significantly higher than GP (ANM vs. HWR, $p < 0.001$; ANM vs. GP, $p = 0.06$; GP vs. HWR, $p = 0.22$ two-tailed paired t-test, Holm-Bonferroni corrected). However, no significant advantage was observed in the low-frequency band ($p = 0.53$; repeated-measures ANOVA; Fig. 4f).

3.3. The relationship between regressor and TRF power spectra

While our analysis demonstrated that ANM regressor, especially when combined with peaky speech stimuli, offers an advantage across multiple metrics among the three regressors, it is crucial to acknowledge the inherent spectral differences among the regressors as illustrated in Figure 5. The deconvolution process, which includes dividing the Fourier transform of the EEG signal by the Fourier transform of the regressor, highlights the significance of the regressor's spectrum on the resulting TRF. The inverse regressor spectrum effectively acts as a filter, where frequencies with lower amplitude in the regressor are emphasized in the resulting TRF. (It should be noted that even if the analysis is done in the time-domain, the same still applies, as this process involves multiplying by the inverse of the autocorrelation matrix.) For example, the ANM regressor (Fig. 5c) has a decreasing magnitude in higher frequency regions compared to the GP (Fig. 5b), leading to the ABR TRF derived from ANM containing larger magnitudes in higher frequencies than that derived from the GP. These spectral differences have no effect on prediction accuracy, since the TRF as a convolution kernel compensates for the input spectrum, but have a large effect on the TRF waveform.

We can explore the effect of these spectral differences in two ways. The first is simply to apply filters that accentuate the standard ABR morphology. Up to this point, we have used broad filters, applying only a first-order 1 Hz high-pass filter to the raw EEG recordings to remove drift. It is more common to somewhat aggressively high-pass

filter ABRs. Figure 6 compares the ABR waveforms from this study and click responses recorded from recent study (Shan et al., 2024) without (Fig. 6a-c) and with (Fig. 6d-f) a 150 Hz third-order high-pass filter, as filtering this way can improve visibility of early ABR components (Polonenko & Maddox, 2021). This comparison demonstrates a few important points. The first is that the click response without the high-pass filter does not show the standard ABR waveform of distinct waves, with wave I followed by a broad wave V, with wave II “riding” on top of that. Applying the high-pass filter, however, makes waves I, III, and V distinct and obvious (waves II and IV are variable and rarely seen in grand averages, even with clicks). The same is true of the ANM speech ABR for both stimulus types. The GP ABR, due to the GP regressor's flat spectrum, is much more dominated by low-frequency energy, obscuring the individual waves, but the high-pass filter again reveals distinct components. Thus, when the responses are compared only over the most relevant ABR frequency range, they become much more similar, all showing clear waves I, III, and V.

The second consideration we can give to differing TRF spectra is by analyzing waveform SNR in a frequency-specific way, such that the effect of the overall spectral shape is minimized. Figure 7 shows the median SNR for each regressor-stimulus condition across frequency. There are useful SNRs up to about 500 Hz for peaky speech with all three regressors, and for the ANM regressor with unaltered speech. In that range, the ANM regressor with peaky speech offers the best SNR, consistent with its superiority from the previous sections. The peaky-GP and unaltered-ANM are very similar to each other, except for the lowest frequency bin, and are a few decibels below the peaky-ANM combination.

4. DISCUSSION

This study presents a comprehensive quantitative analysis and comparison of deconvolution using stimulus regres-

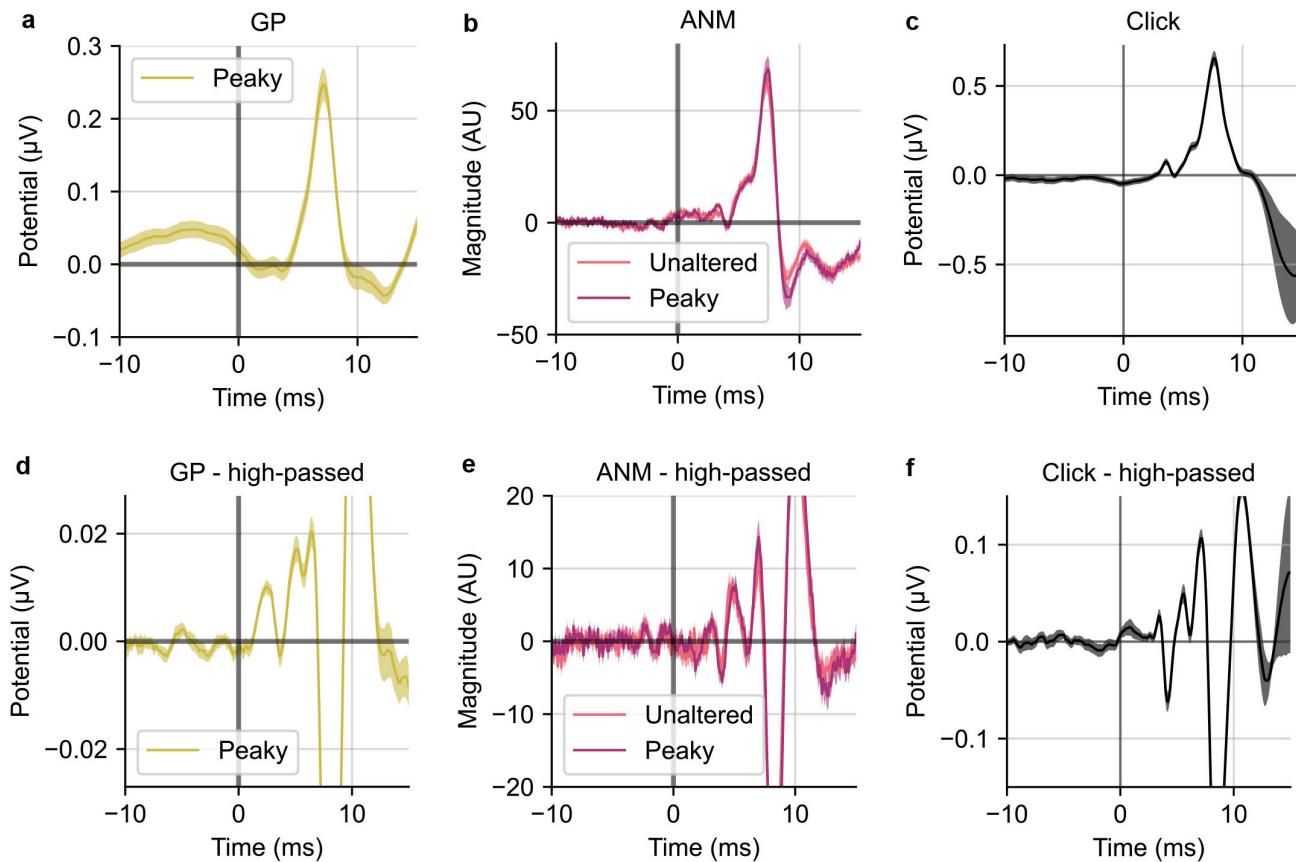


Fig. 6. The ABR derived from the three regressors: HWR (a), GP (b), ANM (c) and high-passed at 150 Hz (d), (e), and (f). The vertical axis ranges were adjusted to show the early waves for both original and high-passed ABRs in this figure.

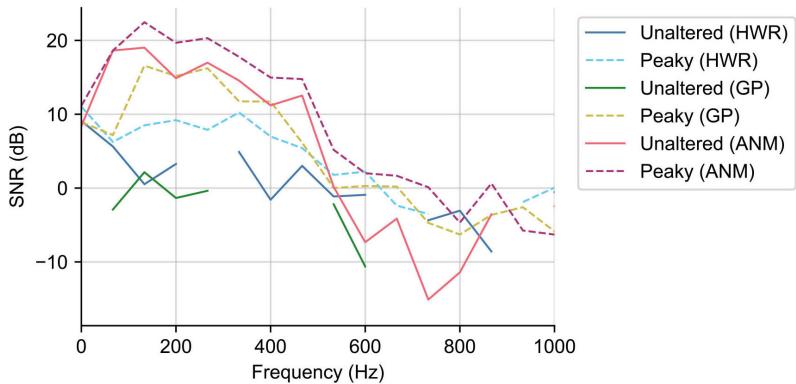


Fig. 7. The SNR analysis for the derived ABRs per frequency bin.

sors designed to derive the TRF corresponding to the human ABR from continuous naturalistic speech. We analyzed EEG recordings from subjects listening to both unaltered speech and modified peaky speech. We compared three regressors that were developed in recent studies: the HWR from [Maddox and Lee \(2018\)](#), the GP from [Polonenko and Maddox \(2021\)](#), and the ANM from [Shan et al. \(2024\)](#). Several metrics were conducted to compare these regres-

sors' performance, including the derived ABR waveform SNR, the time required for subjects to get robust ABRs, and the prediction accuracies of the ABR kernel with broadband (Pearson's correlation) and per-frequency (spectral coherence) approaches. The insights gained from these evaluations are intended to inform and guide future research in selecting the most appropriate regressors for ABR derivation from continuous, naturalistic speech.

4.1. Quantitatively comparing input regressors

To generalize our results: we found that the ANM regressor with peaky speech provided the best performance, with the ANM for unaltered speech and the GP regressor for peaky speech close behind. Some caveats and specific situations where one technique might be favored over another are discussed below. The HWR regressor provided relatively poor ABRs for both speech conditions. We also derived the response from natural speech using the GP regressor for completeness, but we do not recommend this combination for practical use. Even though it did yield an ABR, the quality was predictably bad, with responses showing small amplitude and broad Wave V. This combination is not discussed further.

The HWR regressor, which was the first of these techniques to be developed (Maddox & Lee, 2018), did not match the performance of other regressors in either speech condition. The HWR-derived ABR exhibited a relatively noisy waveform with a broad Wave V (Fig. 2a), requiring more than 42 min to acquire robust ABRs from all subjects (SNR ≥ 0 dB; Fig. 3b). The HWR ABR kernel resulted in low prediction accuracy because the kernel lacked the temporal detail of subcortical responses and had lower SNR. However, when the kernel time window was extended to incorporate the response with cortical responses, its performance was similar to the other two regressors (Fig. 4a).

The GP regressor coupled with peaky speech provided ABRs that showed early waves (Wave I) in the raw responses. When high-passed at 150 Hz, both Wave I and Wave III could be seen (Fig. 6d), allowing for examining the early generators of the auditory evoked potential. These responses may hold potential for clinical use, for example for fitting hearing aids using relevant sounds, rather than artificial ones like tones. The GP regressor was also more efficient than HWR, with all subjects reaching the 0 dB SNR criterion in only 12 min. This efficiency could be further enhanced, as the prior study has shown, with high-pass filtering at 150 Hz potentially reducing the time to around 5 min (Polonenko & Maddox, 2021). GP-derived kernels also provided better prediction than HWR.

The ANM regressor demonstrated superior performance in unaltered speech and comparable performance as GP in peaky speech conditions. This regressor did not only derive the best SNR ABR, but like the GP's ability in peaky speech, this regressor also has the benefit of showing early ABR components—Wave I and Wave III—for both speech conditions, even without the need for further filtering (Fig. 2c). The time required to get decent ABRs in both conditions was substantially reduced compared to HWR, and it was even faster than GP for peaky

speech (Fig. 3b). The best prediction accuracy was achieved using the ANM-derived kernels in both unaltered and peaky speech in correlation and spectral coherence analysis. The ANM's excellent performance stems from its biological fidelity, as it takes the auditory system's peripheral nonlinearities into account before linear deconvolution is performed, with the adaptation in the auditory nerve being particularly important (Kulasingham, Bachmann, et al., 2024; Shan et al., 2024).

Some of the metrics we tested, such as SNR and acquisition time (as well as general waveform morphology), are frequency dependent, and thus affected by the power spectrum of the regressors, which differed substantially (Fig. 5). Because deconvolution can be computed through frequency domain division, using spectrally different regressors is equivalent to applying different filtering to the EEG data (and equivalently, to the deconvolved response). These differences mean that direct comparison of the responses with different filtering might not be fair, because one regressor may accentuate noisier frequency bands than others. We attempted to address this issue in a few different ways. Most simply, we high-pass filtered ANM and GP TRFs from above 150 Hz (Fig. 6). Eliminating the lower frequencies, where the regressor spectra differed substantially, did two things. First, it made the ANM and GP TRF morphologies much more similar to each other. Second, it increased the similarity of both TRFs to the standard click-evoked ABR morphology in which waves I, III, and V can be clearly distinguished (waves II and IV are present under ideal circumstances, but are often missing in practice, even from high-quality ABR measurements). This result of filtering shows that the spectrum of the regressor (and thus the TRF) has a large effect on the way the waveforms look. We also computed the frequency-specific SNR for each regressor (Fig. 7), so that the overall effects of the regressors' spectral shape were minimized (as applying a gain to a signal has no effect on its SNR). We found that the GP and both ANM responses had high SNR up to about 500 Hz, with the peaky-ANM TRF slightly edging out the other two, as it did for the measures discussed above.

Finally, we found that, between the two stimulus types, peaky speech elicited subcortical EEG responses that could be predicted with higher accuracy than unaltered speech. When analyzing phase-only regressors, this trend holds true across all regressors, with peaky speech resulting in superior SNR regardless of the regressor employed. Even the HWR-derived ABR from peaky speech had better SNR than unaltered speech. Similar to the CHEECH (CHirp-spEECH) stimuli (Backer et al., 2019) that incorporated chirps into speech, peaky speech is designed to make the speech click-like, aligning neural responses across the tonotopic axis, thereby eliciting

stronger auditory evoked potentials (Polonenko & Maddox, 2021). Given that peaky speech hardly alters sound quality and does not impact intelligibility, it stands out as the preferable stimulus for deriving speech-evoked ABR, when experimental conditions allow.

4.2. Qualitatively comparing input regressors

The natural question following a comparison of two stimulus types and three regressors is what to use in future experiments. Since the ANM regressor (for both stimulus types) and GP regressor (for peaky speech) provided similar performance, the answer is nuanced and experiment-dependent (the HWR regressor was poorest by all metrics and is unlikely to be appropriate). Both of these regressors and both stimulus types have their strengths and weaknesses that will determine the best choice. Where the GP regressor is discussed below, it is on the assumption that peaky speech is used as the stimulus. The overall shape of the response is not considered a differentiating factor between the GP and ANM regressors because they can be made to be very similar through spectral manipulation (i.e., filtering).

We will discuss stimulus type first. Peaky speech's primary disadvantages are that it requires pre-processing and that it is not quite natural, although we consider the latter issue to be minor. It also cannot be broadly applied to arbitrary stimulus types, as it assumes a calculable fundamental frequency. Its advantages are that it can be used with either the GP or ANM regressor, affording greater flexibility for analysis, and provides slightly better responses than natural speech with both regressors. Natural speech, beyond the obvious benefit of its inherent ecological validity, has the advantage of needing no pre-processing, making it appropriate for real-time use where sound and EEG data are recorded at the same time. For example, one can directly use the speech sound from a TV show that the subject intends to watch in real life as the stimulus. However, natural speech cannot be used with the GP regressor, it so requires that the ANM be used for analysis (or similar methods, as described in Kulasingham, Bachmann, et al. (2024)).

A unique benefit of the GP regressor is that the impulses that make up the pulse train regressor are of unitary magnitude regardless of stimulus amplitude, meaning the deconvolved ABR can be expressed in simple and easily interpretable units of electrical potential. While TRFs computed with the other regressors also have units, they are more complicated ($\mu\text{V} / \text{Pa}$ for HWR, $\mu\text{V} / (\text{spikes} / \text{s})$ for ANM) and also imply a linear relationship with changes in regressor magnitude that is unlikely to be accurate—this is discussed more fully in the final paragraph of this section. While not explored in this

study, Polonenko and Maddox (2021) highlighted another benefit of using the GP regressor with multiband peaky speech, where the GP regressor can be extended to simultaneously investigate ABRs across different frequency regions, working on a similar principle to the parallel ABR (Polonenko & Maddox, 2019), offering a broader clinical application scope.

The ANM does not require pre-processed stimuli and is useful for studying a wide range of spectro-temporally rich natural stimuli, including music (Shan et al., 2024), making it versatile for various research purposes. However, compared to the GP, it has the limitation that the derived ABR is not expressed in meaningful units. Computing the ANM regressor takes considerable computation time, although this can be mitigated by using similar regressors that still include adaptation (Kulasingham, Bachmann, et al., 2024). Thus, while the GP requires significant stimulus pre-processing, use of the ANM regressor requires substantial processing at the analysis stage. In the majority of use cases, neither of these requirements poses a problem, as stimulus and regressor generation are both typically one-time offline procedures. A recent study by Kulasingham, Bachmann, et al. (2024) compared the ANM with regressors generated by other simpler auditory periphery models. They found that when using a more computationally efficient regressor that still includes nonlinear effect of adaptation (Osse Vecchi & Kohlrausch, 2021), the SNR of the derived ABR is similar to that of the more complicated ANM, despite the derived ABR's lack of early components (Kulasingham, Bachmann, et al., 2024).

A limitation of our study was its exclusive focus on a single speech stream narrated by a male speaker. Previous studies indicate that speech from a female speaker, characterized by a higher pitch, tends to reduce the amplitude of wave V (Polonenko & Maddox, 2021, 2024; Saiz-Alia & Reichenbach, 2020). This effect is particularly relevant for peaky speech, where a higher pitch correlates with a faster rate of glottal pulse, leading to neuronal adaptation and refractoriness (Burkard & Hecox, 1983; R. Burkard et al., 1990).

Finally, it is important to consider what the deconvolved response really represents. Calling it a response is a bit of a misnomer—it is a temporal kernel that relates a regressor to an EEG recording through convolution. This distinction is not pedantic. Consider an example experiment in which the same peaky speech stream is presented at a high- and low-level 20 dB apart. The subcortical response to the lower-level stimulus will be smaller and later. The GP regressor is the same for both stimulus levels, and the deconvolved ABR should be smaller and later, as expected. The ANM regressor, however, changes based on stimulus level. The regressor itself should be smaller and later at the lower level. If we

assume it perfectly estimates the change, then the deconvolved response will be the same for both stimulus levels. If the ANM overestimates the amplitude reduction and delay, then the deconvolved ABR could even be larger and earlier for the lower stimulus level, which would be a very strange result on its face. [Kulasingham, Innes-Brown, et al. \(2024\)](#) ran such an experiment using several regressors to estimate level effects on subcortical speech encoding. They, indeed, found that the ANM and other more complex regressors were inappropriate, and relied on simpler ones, even though it resulted in an SNR tradeoff. That experiment did not use peaky speech, but had it, the GP regressor would have allowed level effects to be observed with only a small decrease in SNR compared to the ANM. These results demonstrate that SNR is not the only important factor. Careful consideration must be given to the design, analysis, and interpretation of deconvolution studies.

DATA AND CODE AVAILABILITY

EEG recordings were from [Polonenko and Maddox \(2021\)](#) Dryad repository (<https://doi.org/10.5061/dryad.12jm63xwd>). The Python code for this study is available on GitHub repository (https://github.com/maddoxlab/peaky_vs_anm).

AUTHOR CONTRIBUTIONS

Study conception and design: T.S., R.K.M.; Analysis and interpretation of results: T.S., R.K.M.; Manuscript editing: T.S., R.K.M. All authors reviewed the results and approved the final version of the manuscript.

DECLARATION OF COMPETING INTEREST

The authors have declared no competing interest.

ACKNOWLEDGMENTS

Research reported in this publication was supported by NSF CAREER grant 2142612 awarded to R.K.M.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available with the online version here: <https://doi.org/10.1162/IMAG.a.19>

REFERENCES

Bachmann, F. L., MacDonald, E. N., & Hjortkjær, J. (2021). Neural measures of pitch processing in EEG responses to running speech. *Frontiers in Neuroscience*, 15, 738408. <https://doi.org/10.3389/fnins.2021.738408>

Backer, K. C., Kessler, A. S., Lawyer, L. A., Corina, D. P., & Miller, L. M. (2019). A novel EEG paradigm to simultaneously and rapidly assess the functioning of auditory and visual pathways. *Journal of Neurophysiology*, 122(4), 1312–1329. <https://doi.org/10.1152/jn.00868.2018>

Boersma, P. (2011). *Praat: Doing phonetics by computer* [Computer program]. <http://www.praat.org/>.

Burkard, R., & Hecox, K. (1983). The effect of broadband noise on the human brainstem auditory evoked response. I. Rate and intensity effects. *The Journal of the Acoustical Society of America*, 74(4), 1204–1213. <https://doi.org/10.1121/1.390024>

Burkard, R., Shi, Y., & Hecox, K. E. (1990). A comparison of maximum length and Legendre sequences for the derivation of brain-stem auditory-evoked responses at rapid rates of stimulation. *The Journal of the Acoustical Society of America*, 87(4), 1656–1664. <https://doi.org/10.1121/1.399413>

Burkard, R. F., Eggermont, J. J., & Don, M. (2007). *Auditory evoked potentials: Basic principles and clinical application*. Lippincott Williams & Wilkins. <https://doi.org/10.1097/aud.0b013e31816a0d7c>

Carney, L. H., Li, T., & McDonough, J. M. (2015). Speech coding in the brain: Representation of vowel formants by midbrain neurons tuned to sound fluctuations. *Environ*, 2(4). <https://doi.org/10.1523/eneuro.0004-15.2015>

Coffey, E. B., Herholz, S. C., Chepeliuk, A. M., Baillet, S., & Zatorre, R. J. (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nature Communications*, 7(1), 1–11. <https://doi.org/10.1038/ncomms11070>

Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The multivariate temporal response function (mTRF) toolbox: A MATLAB toolbox for relating neural signals to continuous stimuli. *Frontiers in Human Neuroscience*, 10, 604. <https://doi.org/10.3389/fnhum.2016.00604>

David, S. V., Mesgarani, N., & Shamma, S. A. (2007). Estimating sparse spectro-temporal receptive fields with natural stimuli. *Network: Computation in Neural Systems*, 18(3), 191–212. <https://doi.org/10.1080/09548980701609235>

Di Liberto, G. M., O'sullivan, J. A., & Lalor, E. C. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Current Biology*, 25(19), 2457–2465. <https://doi.org/10.1016/j.cub.2015.08.030>

Ding, N., & Simon, J. Z. (2012). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology*, 107(1), 78–89. <https://doi.org/10.1152/jn.00297.2011>

Elberling, C., & Wahlgreen, O. (1985). Estimation of auditory brainstem response, ABR, by means of Bayesian inference. *Scandinavian Audiology*, 14(2), 89–96. <https://doi.org/10.3109/01050398509045928>

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>

Krizman, J., Skoe, E., & Kraus, N. (2010). Stimulus rate and subcortical auditory processing of speech. *Audiology and Neurotology*, 15(5), 332–342. <https://doi.org/10.1159/000289572>

Kulasingham, J. P., Bachmann, F. L., Eskelund, K., Enqvist, M., Innes-Brown, H., & Alickovic, E. (2024). Predictors for estimating subcortical EEG responses to continuous speech. *PLoS One*, 19(2), e0297826. <https://doi.org/10.1371/journal.pone.0297826>

Kulasingham, J. P., Innes-Brown, H., Enqvist, M., & Alickovic, E. (2024). Level-dependent subcortical electroencephalography responses to continuous speech. *Eneuro*, 11(8). <https://doi.org/10.1523/eneuro.0135-24.2024>

Lalor, E. C., & Foxe, J. J. (2010). Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *European Journal of Neuroscience*, 31(1), 189–193. <https://doi.org/10.1111/j.1460-9568.2009.07055.x>

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. *Journal of Neurophysiology*, 102(1), 349–359. <https://doi.org/10.1152/jn.90896.2008>

Larson, E., Gramfort, A., Engemann, D. A., Leppakangas, J., Brodbeck, C., Jas, M., Brooks, T., Sassenhagen, J., Luessi, M., McCloy, D., King, J.-R., Höchenberger, R., Goj, R., Favelier, G., Brunner, C., van Vliet, M., Wronkiewicz, M., Holdgraf, C., Rockhill, A., ... Luzpaz. (2023). MNE-Python. In (Version v1.6.0) Zenodo. <https://doi.org/10.5281/zenodo.10161630>

Maddox, R. K., & Lee, A. K. (2018). Auditory brainstem responses to continuous natural speech in human listeners. *Eneuro*, 5(1). <https://doi.org/10.1523/eneuro.0441-17.2018>

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences*, 104(40), 15894–15898. <https://doi.org/10.1073/pnas.0701498104>

Osse Vecchi, A., & Kohlrausch, A. (2021). Perceptual similarity between piano notes: Simulations with a template-based perception model. *The Journal of the Acoustical Society of America*, 149(5), 3534–3552. <https://doi.org/10.1121/10.0004818>

Picton, T. W., Hillyard, S. A., Krausz, H. I., & Galambos, R. (1974). Human auditory evoked potentials. I: Evaluation of components. *Electroencephalography and Clinical Neurophysiology*, 36, 179–190. [https://doi.org/10.1016/0013-4694\(74\)90155-2](https://doi.org/10.1016/0013-4694(74)90155-2)

Polonenko, M. J., & Maddox, R. K. (2019). The parallel auditory brainstem response. *Trends in Hearing*, 23, 2331216519871395. <https://doi.org/10.1177/2331216519871395>

Polonenko, M. J., & Maddox, R. K. (2021). Exposing distinct subcortical components of the auditory brainstem response evoked by continuous naturalistic speech. *Elife*, 10, e62329. <https://doi.org/10.7554/elife.62329>

Polonenko, M. J., & Maddox, R. K. (2024). Fundamental frequency predominantly drives talker differences in auditory brainstem responses to continuous speech. *JASA Express Letters*, 4(11), 114401. <https://doi.org/10.1121/10.0034329>

Rudnicki, M., Schoppe, O., Isik, M., Völk, F., & Hemmert, W. (2015). Modeling auditory coding: From sound to spikes. *Cell and Tissue Research*, 361(1), 159–175. <https://doi.org/10.1007/s00441-015-2202-z>

Saiz-Alía, M., & Reichenbach, T. (2020). Computational modeling of the auditory brainstem response to continuous speech. *Journal of Neural Engineering*, 17(3), 036035. <https://doi.org/10.1088/1741-2552/ab970d>

Scott, M. (2008). *The alchemyst* (Vol. 1). Ember. <https://doi.org/10.5040/9798400605468.0079>

Shan, T., Cappelloni, M. S., & Maddox, R. K. (2024). Subcortical responses to music and speech are alike while cortical responses diverge. *Scientific Reports*, 14(1), 789. <https://doi.org/10.1038/s41598-023-50438-0>

Skoe, E., & Kraus, N. (2010). Auditory brainstem response to complex sounds: A tutorial. *Ear and Hearing*, 31(3), 302. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2868335/pdf/nihms-181042.pdf>

Slepian, D. (1978). Prolate spheroidal wave functions, Fourier analysis, and uncertainty—V: The discrete case. *Bell System Technical Journal*, 57(5), 1371–1430. <https://doi.org/10.1002/j.1538-7305.1978.tb02104.x>

Zilany, M. S., Bruce, I. C., & Carney, L. H. (2014). Updated parameters and expanded simulation options for a model of the auditory periphery. *The Journal of the Acoustical Society of America*, 135(1), 283–286. <https://doi.org/10.1121/1.4837815>

Zilany, M. S., Bruce, I. C., Nelson, P. C., & Carney, L. H. (2009). A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics. *The Journal of the Acoustical Society of America*, 126(5), 2390–2412. <https://doi.org/10.1121/1.3238250>