Trends in Hearing
Volume 28: 1–15

DOI: 10.1177/23312165241246596
journals.sagepub.com/home/tia

Extending Subcortical EEG Responses to Continuous Speech to the Sound-Field

Florine L. Bachmann[1,*], Joshua P. Kulasingham[2,*], Kasper Eskelund[3], Martin Enqvist[2], Emina Alickovic[1,2,•] and Hamish Innes-Brown[1,4,•]

Abstract

The auditory brainstem response (ABR) is a valuable clinical tool for objective hearing assessment, which is conventionally detected by averaging neural responses to thousands of short stimuli. Progressing beyond these unnatural stimuli, brainstem responses to continuous speech presented via earphones have been recently detected using linear temporal response functions (TRFs). Here, we extend earlier studies by measuring subcortical responses to continuous speech presented in the sound-field, and assess the amount of data needed to estimate brainstem TRFs. Electroencephalography (EEG) was recorded from 24 normal hearing participants while they listened to clicks and stories presented via earphones and loudspeakers. Subcortical TRFs were computed after accounting for non-linear processing in the auditory periphery by either stimulus rectification or an auditory nerve model. Our results demonstrated that subcortical responses to continuous speech could be reliably measured in the sound-field. TRFs estimated using auditory nerve models outperformed simple rectification, and 16 minutes of data was sufficient for the TRFs of all participants to show clear wave V peaks for both earphones and sound-field stimuli. Subcortical TRFs to continuous speech were highly consistent in both earphone and sound-field conditions, and with click ABRs. However, sound-field TRFs required slightly more data (16 minutes) to achieve clear wave V peaks compared to earphone TRFs (12 minutes), possibly due to effects of room acoustics. By investigating subcortical responses to sound-field speech stimuli, this study lays the groundwork for bringing objective hearing assessment closer to real-life conditions, which may lead to improved hearing evaluations and smart hearing technologies.

Keywords

electroencephalography, temporal response function, speech ABR, neural speech processing, auditory brainstem response

Received 14 November 2023; Revised received 8 March 2024; accepted 26 March 2024

Introduction

The auditory brainstem response (ABR) is an electrophysiological measure of the subcortical neural activity of the auditory system in response to sound. It is widely used for hearing screening, especially in populations for which behavioral measures of hearing are not feasible, such as newborns (Galambos & Despland, 1980; Joint Committee on Infant Hearing, 2007) and coma patients (Guérit, 1999). The aggregate activity of several subcortical stages of the auditory pathway gives rise to the ABR event-related potential (ERP) which can be detected using electroencephalography (EEG) (Picton, 2010). Since the electric potentials generated by this neural activity are relatively weak when measured using scalp-EEG, conventional ERPs are computed using signal averaging methods, often with thousands of repetitive short stimuli, such as clicks (Picton, 2010), chirps (Rodrigues et al., 2013), or short speech syllables (Chandrasekaran & Kraus, 2010). Although such methods provide robust

[1]Eriksholm Research Centre, Snekkersten, Denmark
[2]Automatic Control, Department of Electrical Engineering, Linköping University, Linköping, Sweden
[3]Oticon A/S, Smørum, Denmark
[4]Department of Health Technology, Technical University of Denmark, Lyngby, Denmark

[*]Equally contributed as first authors.
[•]Equally contributed as last authors.

Corresponding Author:
Florine L. Bachmann, Eriksholm Research Centre, Snekkersten, Denmark.
Email: flln@eriksholm.com

ABRs, they are far removed from ecologically relevant stimuli, and recent years have seen a shift toward more complex stimuli relevant to daily life such as continuous speech in several fields of auditory and speech neuroscience (Brodbeck & Simon, 2020; Hamilton & Huth, 2020; Lunner et al., 2020).

Recent studies have found that subcortical responses to continuous natural speech can be reliably detected using deconvolution methods which produce a temporal response function (TRF) (Bachmann et al., 2019, 2021; Etard et al., 2019; Maddox & Lee, 2018; Van Canneyt et al., 2021a, 2021b, 2021c). While work that focused on the neural response to the fundamental frequency finds broader and later subcortical response peaks (Etard et al., 2019; Van Canneyt et al., 2021a, 2021b, 2021c), subcortical responses to more broadband speech information show similar waveform morphologies to the conventional click ABR (Maddox & Lee, 2018; Shan et al., 2024—see Bachmann et al., 2021 for a comparison of the two approaches). Specifically, these TRFs estimated with broadband speech information predictors show a prominent wave V peak, which is thought to arise from a mixture of subcortical structures, including the inferior colliculus (Møller & Jannetta, 1983; Moore, 1987; Starr, 1976). Although TRFs have been widely used in recent years to study *cortical* responses to continuous speech (Brodbeck & Simon, 2020; Di Liberto et al., 2015; Alickovic et al., 2019; Kulasingham et al., 2020), investigating *subcortical* TRFs to continuous speech is still a new and emerging field (Bachmann et al., 2019, 2021; Etard et al., 2019; Maddox & Lee, 2018; Polonenko & Maddox, 2021; Shan et al., 2024; Van Canneyt et al., 2021a, 2021b, 2021c).

In accordance with conventional ABR measurements, these early studies computing subcortical responses to continuous speech involved stimuli presented via insert earphones. Although this allows good control over the exact stimulus presented at the ear, measurements of subcortical responses to speech in sound-field conditions could enable several other use-cases. Subcortical responses to speech presented in the sound-field could enable clinical tests of hearing impairment with natural conversational stimuli in a realistic setting, leading to increased levels of relevance and patient comfort. Critically, since hearing aids incorporate speech-specific signal processing and noise cancellation stages that may attenuate the short transient stimuli that are typically used for evoked ABR paradigms (Garnham et al., 2000), the use of sound-field speech stimuli could lead to more reliable measurements of subcortical responses from listeners wearing hearing aids. Indeed, sound-field speech stimuli have already been successfully used to measure *cortical* responses in hearing aid users, and the impact of various hearing aid settings on neural speech tracking (Alickovic et al., 2020, 2021; Carta et al., 2023).

Although conventional averaged ABRs using sound-field stimuli have been studied in animal models (Kim et al., 2022; Land et al., 2016; Willott, 2006), there have been fewer studies in humans (Jarollahi et al., 2020; Schebsdat et al., 2018). Sound-field auditory steady state responses (ASSRs) have been more widely studied in humans (Hernández-Pérez & Torres-Fortuny, 2013; Stroebel et al., 2007; Zapata-Rodriguez et al., 2021), with a view toward similar applications in clinical diagnosis and hearing aid fitting (Damarla & Manjula, 2007; Shemesh et al., 2012). These studies highlight potential challenges when measuring sound-field responses, due to delays, reverberation, binaural interactions and other effects of room acoustics (Zapata-Rodríguez, 2020). Considering that the TRF is estimated using a predictor based on the stimulus, distortions of the stimuli that reach the ears due to propagation through the sound-field might either result in a distorted TRF estimate or increase the amount of data needed to estimate the TRF. The latter may be partly counteracted by employing predictors generated using an auditory nerve model (Zilany et al., 2009, 2014) instead of simple rectification, which has shown to substantially improve estimated TRFs (Kulasingham et al., 2024; Shan et al., 2024). Still, it remains unclear whether subcortical TRFs, which require temporally precise measurements, can be detected using continuous speech in the sound-field.

In this work, we investigated subcortical EEG responses to continuous speech in both insert earphone and sound-field conditions. We analyzed EEG data collected from 24 participants with normal hearing and used TRF methods to estimate subcortical responses. First, we explored morphological differences in subcortical responses to sound presented in the sound-field versus via insert earphones in the same participants, for both subcortical TRFs to speech stimuli and conventional click ABRs. Specifically, we examined whether after accounting for the speaker delay, the wave V peak in the sound-field condition occurred at similar latencies as in the insert earphone condition (RQ1). Second, we compared subcortical TRFs estimated using predictors generated by simple rectification or an auditory nerve model (ANM) (Zilany et al., 2009, 2014) (RQ2) and validated prior work showing that the ANM substantially improves estimation of wave V peaks. Third, we studied whether distortions of the stimulus in the sound-field leads to a need for more data (longer recording time) to estimate TRFs with response peaks above the noise floor (RQ3). Finally, we investigated whether subcortical responses to speech in the sound-field could be detected in all participants (RQ4).

Our work may lead to reliable objective measures of subcortical auditory activity using ecologically relevant stimuli presented in the sound-field, and lays the groundwork for future investigations into utilizing subcortical responses for clinical evaluations of hearing impairment and hearing aid fitting in more realistic environments.

Materials and Methods

Experimental Setup

Procedure. EEG data were collected from 24 participants (14 males, mean age = 37.1 years, standard deviation (SD) = 10.0

years) with clinically normal and symmetric hearing (all pure-tone thresholds in octave steps from 250 to 4000 Hz ≤ 20 dB HL, no more than 15 dB HL difference between ears at each frequency). All participants were native Danish speakers, right-handed, and provided written informed consent. The study was approved by the ethics committee for the capital region of Denmark (journal number 22010204). The experiment consisted of speech and clicks presented in both an insert and a sound-field condition block, and the block order was balanced across participants. The speech part consisted of four stories; story order was pseudo-randomized, presenting stories in the same order to two age-matched participants with different starting conditions. The insert earphone condition consisted of the following stimuli: speech (47 minutes 56 seconds, divided in 8 audiobook segments), clicks (5 minutes), control condition with clicks and insert eartips outside the ears (5 minutes). The sound-field condition consisted of the same speech (47 minutes 56 seconds) and click material (5 minutes; see Figure 1).

Stimuli. EEG data were collected while participants listened to clicks and four stories from adventures by H.C. Anderson read by a male speaker. The 5 minutes of click trains were presented both via insert earphones and in the sound-field. Click trains consisted of 44.1 clicks of alternating polarity per second. The presentation time of each click was randomly distributed according to a pseudo-Poisson distribution, limiting the minimum inter-click-interval to 15 ms. Clicks were rectangular and of 91 µs duration (four samples, the closest approximation to 100 µs at a 44.1-kHz sampling rate). Click stimuli were presented at 72 dB peak-to-peak equivalent SPL, matching their peak-to-peak voltage to that of a 1-kHz pure-tone presented at 72 dB SPL.

Stories were divided into two segments, resulting in eight trials (average duration = 6 minutes 0 seconds, SD = 55 seconds). All trials were presented in two conditions, through insert earphones or through a loudspeaker in the sound-field. Before presentation, the original two-channel audiobook with a sampling frequency of 44.1 kHz was averaged to construct a mono signal, and high-pass filtered at 1 kHz using a first-order Butterworth filter because the neural response from the brainstem is mainly driven by high frequencies (Abdala & Folsom, 1995). Each of the eight trials were scaled to have the same root-mean-square (RMS) value as a 1-kHz pure tone at 72 dB sound pressure level (SPL).

Setup. Participants were seated facing the loudspeaker at ear height and a distance of 1.5 m (approximately 4.4 ms sound travelling delay) and instructed to relax and listen to the audio, while looking at a fixation cross in front of them. The room was quiet and insulated from outside noise, with dimensions of $4.5 \times 3.2 \times 2.5$ m and a reverberation time (RT_{60} estimated with T_{30}) of 0.32 seconds. All audio was replayed with a sampling rate of 44.1 kHz using an RME Fireface UCX audio interface (RME Audio, Haimhausen, Germany). In the insert condition, sound was delivered via Etymotic ER-2 (Etymotic Research, Illinois, USA) insert earphones (approximately 0.9 ms sound travelling delay), which were shielded using a grounded metal box to avoid direct stimulus artifacts on the EEG. Additionally, a control condition where the clicks were played while the insert earphones were not placed in the ear (but with the earphone wires in the same locations near the participant) was recorded to check whether there were electrical stimulus artifacts in the EEG. Visual inspection confirmed that stimulus artifacts were not

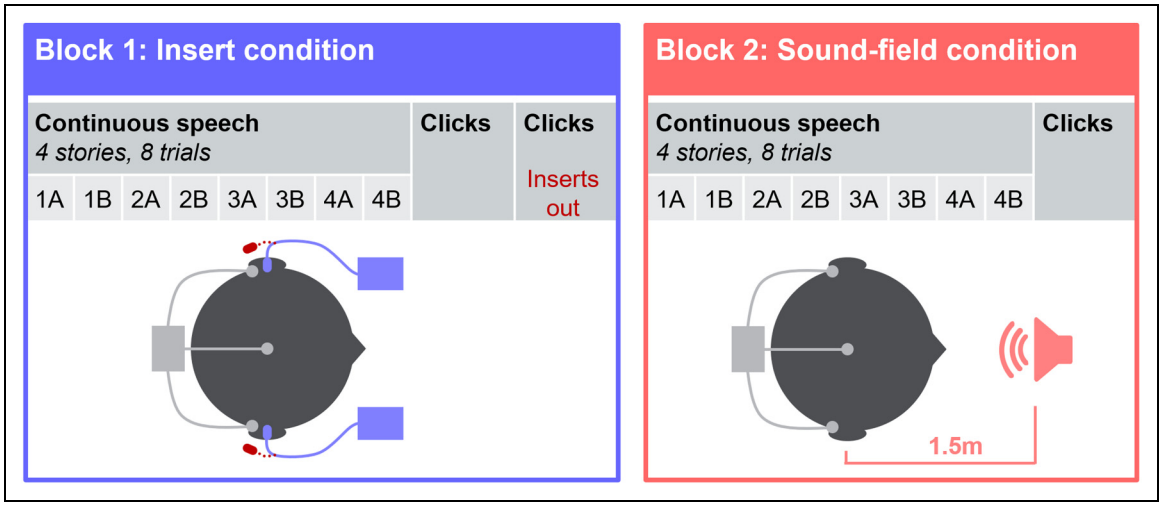

Figure 1. Experiment overview. Continuous speech and click stimuli were presented in an insert earphone condition block, and a sound-field condition block, balancing the starting condition across participants. The speech consisted of four stories, each split into two parts (A and B). Story presentation order was pseudo-randomized across participants and identical across conditions for a given participant. In the insert condition, clicks were also presented with earphone tips outside the ear, serving as a control condition.

present in the ERPs estimated from the control condition, or in any other conditions. For the sound-field condition, the stimuli were presented using a Genelec 8040A loudspeaker (Genelec Oy, Iisalmi, Finland).

EEG Preprocessing

EEG data were collected using a Biosemi Active 2 system (BioSemi, Amsterdam, The Netherlands) at a sampling frequency of 16,384 Hz. A 32-channel cap with electrodes in standard 10–20 positions was used. In addition, electrodes were placed on the left and right mastoids, left and right earlobes and above and below the right eye. Data analysis was conducted in Eelbrain Python toolbox (version 0.38.1) (Brodbeck et al., 2021), and the code is openly available under https://github.com/Eriksholm-Research-Centre/Sound-field-speech-ABR. Only the Cz channel was used for further analysis, referenced to the average of the two mastoid channels. A delay compensated 1 Hz highpass FIR filter was then applied. Next, to remove power line noise, FIR notch filters with widths of 5 Hz and center frequencies at all multiples of 50 Hz up to 1000 Hz were applied. High amplitude sections of the EEG data that were five SDs above the mean were assumed to be contaminated with artifacts, and 1 second segments around these points were set to zero in both the EEG data and the predictors used for later TRF analysis (mean percentage of data set to zero = 2.7%, SD = 1.7%). Only the data from 2 to 242 seconds in each of the eight trials was used for further analysis to avoid possible onset and offset artifacts or responses.

Since subcortical responses consist of peaks in the waveform that are only a few milliseconds in duration, it is crucial that the EEG data and the presented stimuli are synchronized with millisecond precision. This synchronization could be impacted by timing jitters in the trigger or clock drifts between the stimulus presentation system and the EEG recording system. To avoid these issues and ensure synchronization, the audio output of the RME soundcard was fed into the Biosemi system as an external sensor on the Erg1 channel via an optical isolator (StimTrak, BrainProducts, GmbH, Gilching, Germany) to maintain electrical separation. The recorded signal on the Erg1 channel of the EEG system was subsequently used to detect click onsets for ERPs and to generate speech predictors for TRF analysis.

Click ERP Estimation

The click-evoked ABR was estimated on EEG data that was further filtered between 30 and 500 Hz using a delay compensated FIR filter. Click onsets were detected using the Erg1 channel and EEG epochs −10 to 30 ms around the click onset were extracted. This resulted in 13,185 epochs that were averaged to estimate the ERP for each condition (inserts, sound-field and control conditions). The standard procedure of averaging both condensation and rarefaction click responses together was followed. The ERPs were then baseline corrected by subtracting the average of the pre-stimulus activity from −10 to 0 ms.

Speech Predictors

The speech stimuli were used to construct two types of predictors for the TRF model: rectified speech (RS) predictor and auditory nerve model (ANM) predictor. These predictors approximate the processing stages of the peripheral auditory system, and thereby account for non-linearities that cannot be fit using linear TRF models. Since the brainstem response is largely unaffected by the polarity of the auditory input, two predictors were estimated for each case using the speech signal recorded on the Erg1 channel and its sign-flipped (inverted) version. The RS predictors were constructed by rectifying the speech signals, leading to predictors that only kept either the positive or the negative peaks, in accordance with prior work (Maddox & Lee, 2018). This procedure serves as a very coarse approximation of the rectification properties of the peripheral auditory system.

The ANM predictor was formed using a model of the auditory periphery (Zilany et al., 2014), in line with recent work that showed this method improved TRF estimates (Kulasingham et al., 2024; Shan et al., 2024). In brief, the speech stimuli (and their polarity inverted version) were used as inputs to this model, which modelled 43 high spontaneous rate auditory nerve fibers with center frequencies logarithmically spaced between 125 Hz and 16 kHz. The outputs of this model were 43 mean firing rate signals, which were then averaged to form the final predictor pair (i.e., for the original speech signal and its polarity inverted version). The implementation in the cochlea python toolbox (Rudnicki et al., 2015: https://github.com/mrkrd/cochlea) was used to generate these predictors. To ensure that the estimated TRF peak latencies were not affected by inherent lags in the auditory model, the ANM predictors were delayed by 1.1 ms in order to have the maximum correlation with the RS predictor. For further comments on the suitability of this method of accounting for inherent lags in the auditory model, please see the section "Discussion."

Temporal Response Function Estimation

The RS and ANM predictors were used to estimate TRFs for each experimental condition. The TRF is a linear model that represents the time-locked activity of the neural system to the given predictor. The frequency domain method given in previous studies (Maddox & Lee, 2018; Polonenko & Maddox, 2021) was used to estimate the TRF:

$$\mathrm{TRF} = F^{-1}\left\{\frac{\sum_{i=1}^{N} w_i F\{x_i\} * F\{y_i\}}{\sum_{i=1}^{N} (\frac{1}{N}) F\{x_i\} * F\{x_i\}}\right\} \quad (1)$$

Here, F denotes the Fourier transform, N is the number of trials, x_i, y_i and w_i are the predictor, EEG signal and weight for trial i, and * denotes the complex conjugate. The trial weights w_i were set to be the reciprocal of the variance of the EEG data of trial i normalized to sum to 1 across trials, in line with prior work (Polonenko & Maddox, 2021). This was done to down-weight noisy (high variance) EEG trials. The resulting TRF has lags ranging from $-T/2$ to $T/2$ where T is the data length.

Two TRFs were estimated separately for each predictor pair (i.e., generated from the speech signal and its polarity-inverted version), and then averaged together. These TRFs were then bandpass filtered between 30 and 500 Hz using a delay compensated FIR filter, which helped to eliminate high frequency noise in the TRF and led to cleaner wave V estimates. After this, the TRF segment from −10 to 30 ms was extracted for further analysis and the mean baseline activity from −10 to 0 ms was subtracted from each TRF. Finally, the TRFs were scaled to have the same RMS as the click ERPs averaged across all participants, for enabling morphology comparison with the click ERPs.

To investigate the effect of data length on TRF estimation, TRFs were fit on a consecutively increasing number of trials (i.e., 2, 3, …, 8 trials, corresponding to 8, 12, …, 32 minutes of data) in the order that they were presented in the experiment. This simulates TRF estimation as if the experiment had been terminated after a variable number of trials. For each data length, a leave-one-out cross-validation approach was followed, with one trial being used as test data to estimate model fits and the other trials being used to fit the TRF. The TRFs for each cross-validation fold were averaged together to form the final TRF for that data length. A null model was formed by averaging the TRFs that were fit on circularly shifted predictors (shifts of 30, 60, and 90 seconds), similar to typical null models used in prior work with cortical TRFs (Kulasingham et al., 2020). This method preserved the temporal characteristics of the predictor, while destroying the alignment between the predictor and the EEG signals. The same leave-one-out cross-validation approach at each data length was followed for the null models.

Performance Metrics

The Pearson correlation between the predicted EEG from the TRF model and the actual EEG was used as an estimate of the model fit, in line with previous TRF studies (Crosse et al., 2016; Kulasingham & Simon, 2023). The TRF estimated on the training trials was used to predict the EEG on the test trial using the leave-one-out procedure, and the average of the prediction correlations across folds was used as the model fit. The null model fits were also estimated in the same manner using the null model TRFs.

The wave V peak of the subcortical TRFs was extracted by detecting the largest peak between 4 and 9 ms. The SNR of the peak amplitude was calculated as $\mathrm{SNR} = 10\log_{10}[(\sigma^2_{S+N} - \sigma^2_N)/\sigma^2_N]$ where the signal + noise variance σ^2_{S+N} was calculated as the variance in a 5-ms window around the wave V peak and the noise variance σ^2_N was estimated as the average variance in 5 ms windows of the TRF in the range −500 to −20 ms. The SNR was set to have a minimum value of −5 dB since lower SNR values lack meaningful interpretation. The threshold for a meaningful wave V peak was considered to be 0 dB (signal is equal to the noise floor), and corresponded with visually distinct peaks.

Statistical Analysis

The model fits for the full data length ANM and RS TRFs in the inserts and sound-field conditions were compared using related measures two-tailed t-tests with Holm–Bonferroni multiple comparisons correction. The null model fits for each participant were subtracted before the comparison. Cohen's d effect sizes, t-values and p-values are reported. For testing differences in wave V SNR, non-parametric statistical tests were employed, since the wave V SNRs had a skewed distribution with a minimum of −5 dB for participants without clear wave V peaks. Wave V SNRs for the ANM and RS TRFs at the full data length were compared between selected conditions using paired two-tailed small sample Wilcoxon signed rank tests with Holm–Bonferroni multiple comparisons corrections. Next, the same tests were used to examine differences in ANM TRF wave V SNR between insert earphone versus sound-field conditions at different data lengths. The group medians, test statistics (rank sums above zero) and p-values are reported. Two participants were excluded from the statistical tests since they did not have data for the full 32 minutes.

Results

Subcortical Responses for Speech Presented in the Sound-Field Versus Through Insert Earphones

The EEG data from the midline central (Cz) channel referenced to the average of the two mastoid channels was used to estimate click-evoked ERPs and subcortical TRFs to continuous speech. The TRFs were estimated using either the rectified speech (RS) or the auditory nerve model (ANM) predictors. The average click ERPs and speech TRFs for both speech presented via insert earphones and in the sound-field are shown in Figure 2 (top). Clear wave V peaks were seen on group-level for both the click ERPs and the speech TRFs. The pre-stimulus period (−10 to 0 ms) for the click ERP shows activity from the previous click response, since the click train had inter-click spacing mostly ranging from 15 to 25 ms. Even though the inter-click latency was randomly jittered (see section "Methods") there is still a smeared response to the previous click in Figure 2. However, this does not affect any further conclusions since the click ERP was investigated merely as a reference stimulus for inserts and speaker conditions, and the wave V peaks are

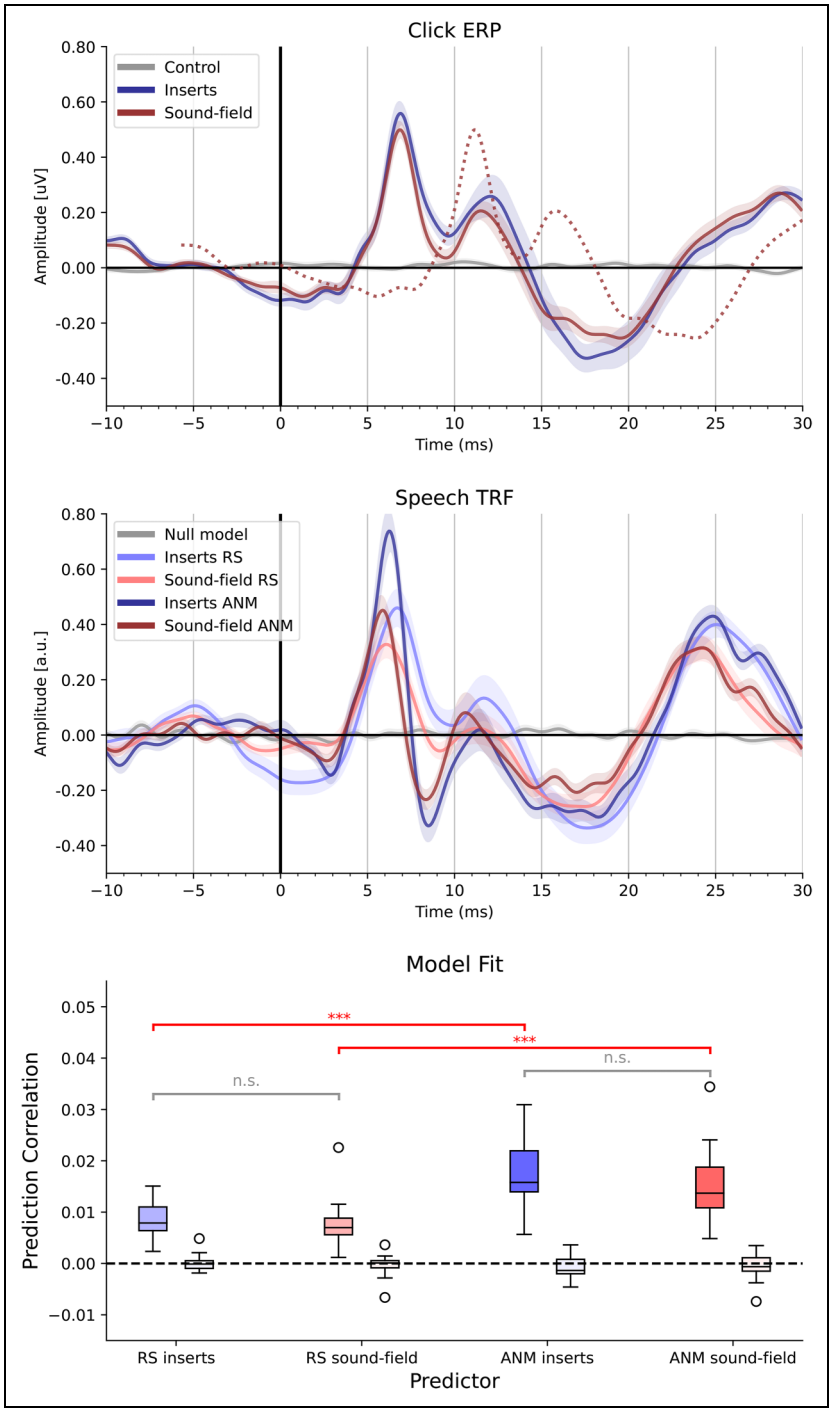

Figure 2. Group-level click ERPs and speech TRFs in inserts and speaker conditions. *Top:* The group-level click ERP across 24 participants is shown for both inserts (blue) and sound-field conditions corrected (red solid) and uncorrected (red dotted) for speaker delay of 4.4 ms. A clear wave V peak is seen that is consistent for both conditions. The pre-stimulus activity may be due to the fast repetition rate of the click stimuli. The control condition with insert eartips outside the ears is depicted in grey and has very low amplitude. *Middle:* The group-level speech TRF across 24 participants is shown for both inserts and speaker conditions for the RS and ANM predictors. Clear wave Vs are seen in all cases, but the ANM predictor results in narrower and larger wave Vs. The amplitude of the wave V is reduced for all speaker conditions, possibly due to effects of room acoustics. The null model is also shown in grey for visual comparison with the noise floor. *Bottom:* The model fit prediction correlations for the estimated TRFs. Boxplots are shown across participants for each condition and predictor, and the null model fits are indicated by the lighter colored boxes next to them. All TRF models significantly outperform the null models. The ANM TRF model fits are significantly higher than the RS TRF model fits in both conditions (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

still clearly visible in both cases at the appropriate latencies within 5 to 10 ms.

The general response morphologies across sound-field and insert earphone condition showed close resemblance. For illustrative purposes, the sound-field response without correcting for the travelling time of sound from the speaker to the eardrum (4.4 ms) is also indicated for the click stimulus (dashed line in Figure 2 top). Henceforth, all results will be reported after accounting for this delay, as well as an earphone presentation delay of 0.9 ms. After accounting for both these sound travelling delays, the average latencies of wave V peaks were similar across inserts and sound-field conditions for both click ERPs (mean [SD] of click ERP wave V latency for inserts = 7.0 [0.4] ms, sound-field = 6.9 [0.4] ms) and speech TRFs (RS inserts = 6.7 [0.6] ms, RS sound-field = 6.2 [0.6], ANM inserts = 6.4 [0.4], ANM sound-field = 6.1 [0.4]). As in previous work, response peaks for the speech occurred slightly earlier than for clicks (Maddox & Lee, 2018). This difference was more pronounced in the sound-field than the insert earphone condition, which will be addressed in the Discussion. On group level, the sound-field conditions resulted in lower amplitudes compared to the insert condition for all cases (percentage reduction in average wave V peak amplitude for sound-field compared to inserts: click ERP = 11.9%, speech TRF RS = 29.2%, speech TRF ANM = 35.6%). An exploration of possible challenges with subcortical responses to sound-field stimuli is provided in the section "Discussion."

Responses Computed to Predictors Obtained Using Simple Rectification Versus an Auditory Nerve Model

The average wave V latency for the ANM predictor was slightly earlier than that for the RS predictor, even though the model delays in the ANM predictor were compensated for by shifting the ANM predictor to have the maximum correlation with the RS predictor (see section "Discussion"). Qualitatively, responses obtained with the RS predictor show a broader peak than responses computed with the ANM predictor. The amplitudes of the TRF wave V were scaled to be comparable to the amplitudes of the click ERP wave V using a common scaling factor for all participants. The individual speech TRFs for the ANM predictor also show clear wave V peaks and consistent TRF waveforms for all participants (see Figure 4).

The model fits of the estimated speech TRF models are shown in the bottom row of Figure 2. All the estimated model fits are well above the null model fits. The model fits for each case were compared after subtracting the individual null model fits using paired *t*-tests with Holm-Bonferroni correction. The ANM model fits were significantly higher than the RS model fits for both inserts ($d = 1.81$, $t_{21} = 8.29$, $p < 0.0001$) and sound-field conditions ($d = 1.85$, $t_{21} = 8.49$, $p < 0.0001$). These results indicate that the ANM leads to a better TRF estimate than the RS. Model fits were not significantly different across inserts and sound-field conditions for the RS TRFs ($d = 0.04$, $t_{21} = 0.17$, $p = 0.8$) or the ANM TRFs ($d = 0.46$, $t_{21} = 2.09$, $p = 0.097$). However, a trend was observed with slightly lower model fits for the sound-field compared to the insert condition, possibly due to effects of room acoustics.

Amount of Data Needed for Subcortical Responses to Continuous Speech

Next, we investigated the amount of data required to estimate reliable TRFs using two metrics; the model fit prediction correlations and the wave V SNR. The latter is a measure of the wave V amplitude relative to the noise floor (see section "Methods"), and has been previously used to evaluate subcortical TRFs (Polonenko & Maddox, 2021; Shan et al., 2024). TRFs were fit on a sequentially increasing number of 4 minutes trials, and both model fit and wave V SNR were calculated (see Figure 3). The ANM predictor outperformed the RS predictor in all cases. The sound-field condition had lower wave V SNR than the inserts conditions, in line with the reduced peak amplitudes seen in the TRFs. ANM TRF model fits and wave V SNRs for the insert condition were above the noise floor for all participants with only 12 minutes of data. For ANM TRFs when speech was presented in the sound-field, model fits were above zero and wave V SNR was above the noise floor in all participants after an EEG recording time of 16 minutes and more.

To further investigate these trends, we used non-parametric pairwise Wilcoxon signed rank tests with Holm-Bonferroni multiple comparisons corrections on the wave V SNRs (see Methods). The ANM TRF SNRs were significantly higher than the RS TRF SNRs for the full 32 minutes of data (inserts RS median = 7.4 dB, ANM median = 14 dB, $T = 0$, $p < 0.001$; sound-field RS median = 4.9 dB, ANM median = 10.9 dB, $T = 0$, $p < 0.001$). We next restricted our statistical tests to only the ANM TRF SNRs and investigated the difference between the insert earphone SNRs and the sound-field SNRs. The tests were not conducted on 8 minutes of data, since several participants did not have clear wave V peaks (≤0 dB SNR). For 12 minutes of data, all participants were above the noise floor for the insert earphone condition, and the SNRs for the insert earphone condition (median = 7.4 dB) were significantly higher ($T = 41$, $p = 0.012$) than the SNRs for the sound-field condition (median = 5.8 dB). This indicated that the sound-field TRFs did not have wave V peaks that were as clear as those for the insert TRFs. We then investigated the amount of data required to reach similar SNRs for the sound-field ANM TRFs. There was no significant difference between the sound-field ANM TRF SNRs at 16 minutes (median = 8.7 dB) and the insert ANM TRF SNRs at 12 minutes ($T = 111$, $p = 0.63$). Furthermore, the sound-field ANM TRF SNRs at 20 minutes (median = 9.4 dB) were significantly higher than insert ANM TRF SNRs at 12 minutes ($T = 42$, $p = 0.012$). These tests indicate that the sound-field ANM TRF wave Vs show comparable SNR to the insert earphone ANM TRF wave Vs after an additional 4 minutes of data.

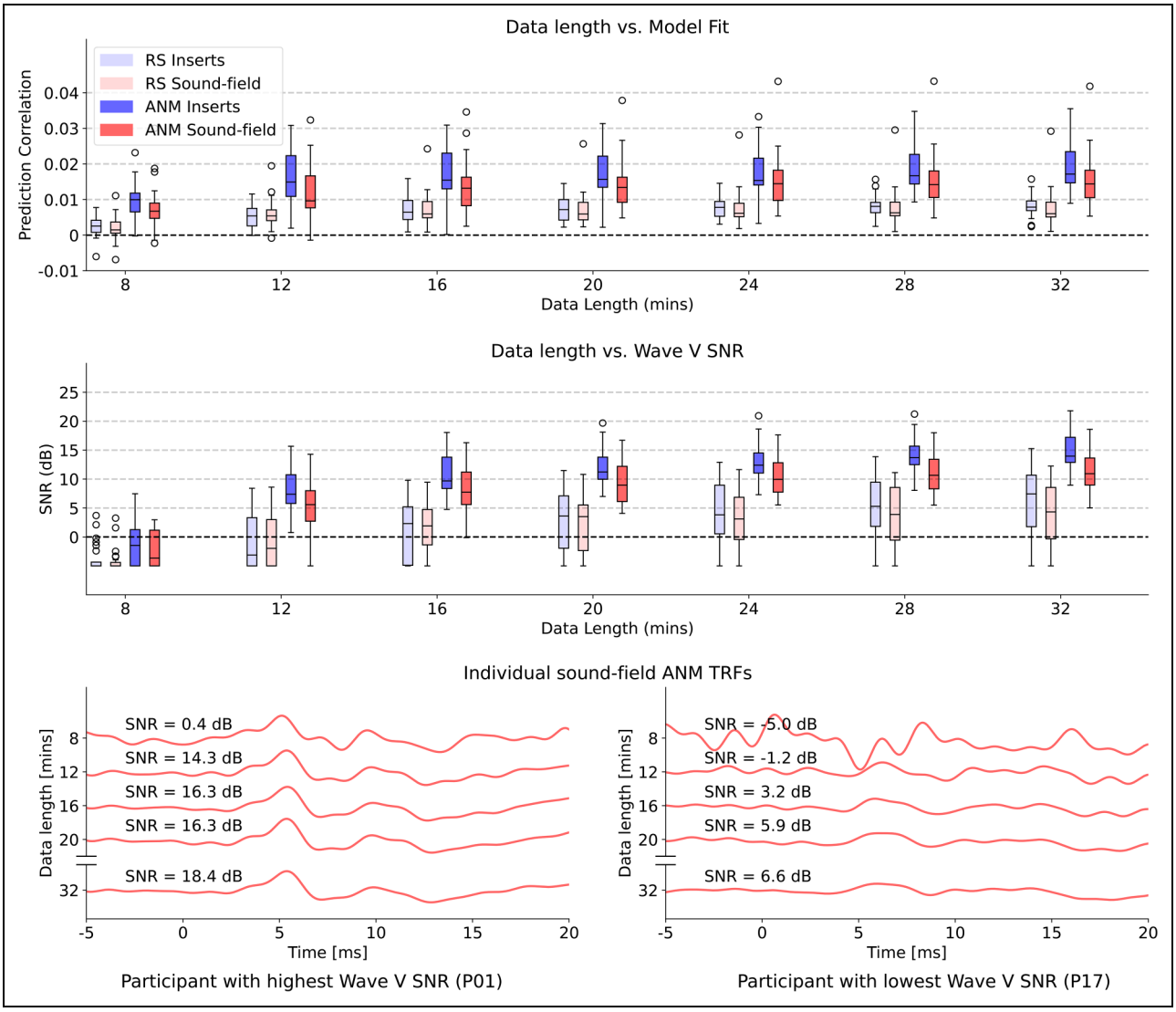

Figure 3. Impact of data length and predictor type. All boxplots are shown across participants. *Top:* Impact of data length on model fit prediction correlations. *Middle:* Impact of data length on wave V SNR. There is a clear increase in both model fits and wave V SNRs with increasing data length. The ANM model outperforms the RS model in all conditions, and the inserts condition seems to have both larger model fits and wave V SNRs than the sound-field condition for most participants. Interestingly, ANM TRFs with model fits above zero and wave V SNR above the noise floor for all participants were obtained with only 12 and 16 minutes of data for inserts and sound-field conditions, respectively. *Bottom:* Representative sound-field ANM TRFs of different data lengths for two participants.

Consistency of Individual Responses to Clicks and Speech in Insert and Sound-Field Conditions

Finally, the individual responses as well as the consistency of wave V peaks across individuals were investigated. Clear wave V peaks could be seen for all participants for the click ERPs and the speech ANM TRFs (see Figure 4 and Supplementary Figures S1, S2 for individual RS TRFs and click ERPs). As discussed above, average wave V latencies and amplitudes show some differences across insert and sound-field conditions (see Figure 2), possibly due to room acoustics. However, these effects should primarily result in a systematic shift and the distribution of individual responses should be largely consistent, as evidenced by the individual TRFs in Figure 4. To investigate this, we correlated individual peak V amplitudes and latencies across insert and sound-field conditions for both click ERPs and speech TRFs (see Figure 5). The resulting Pearson correlation values were significant in all cases after correcting for multiple comparisons using the Holm–Bonferroni method (correlations and p-values are provided in Figure 5 titles). This consistency could be due to a variety of individual factors such as signal quality, anatomical details (e.g., thickness of the scalp leading to larger responses for some participants),

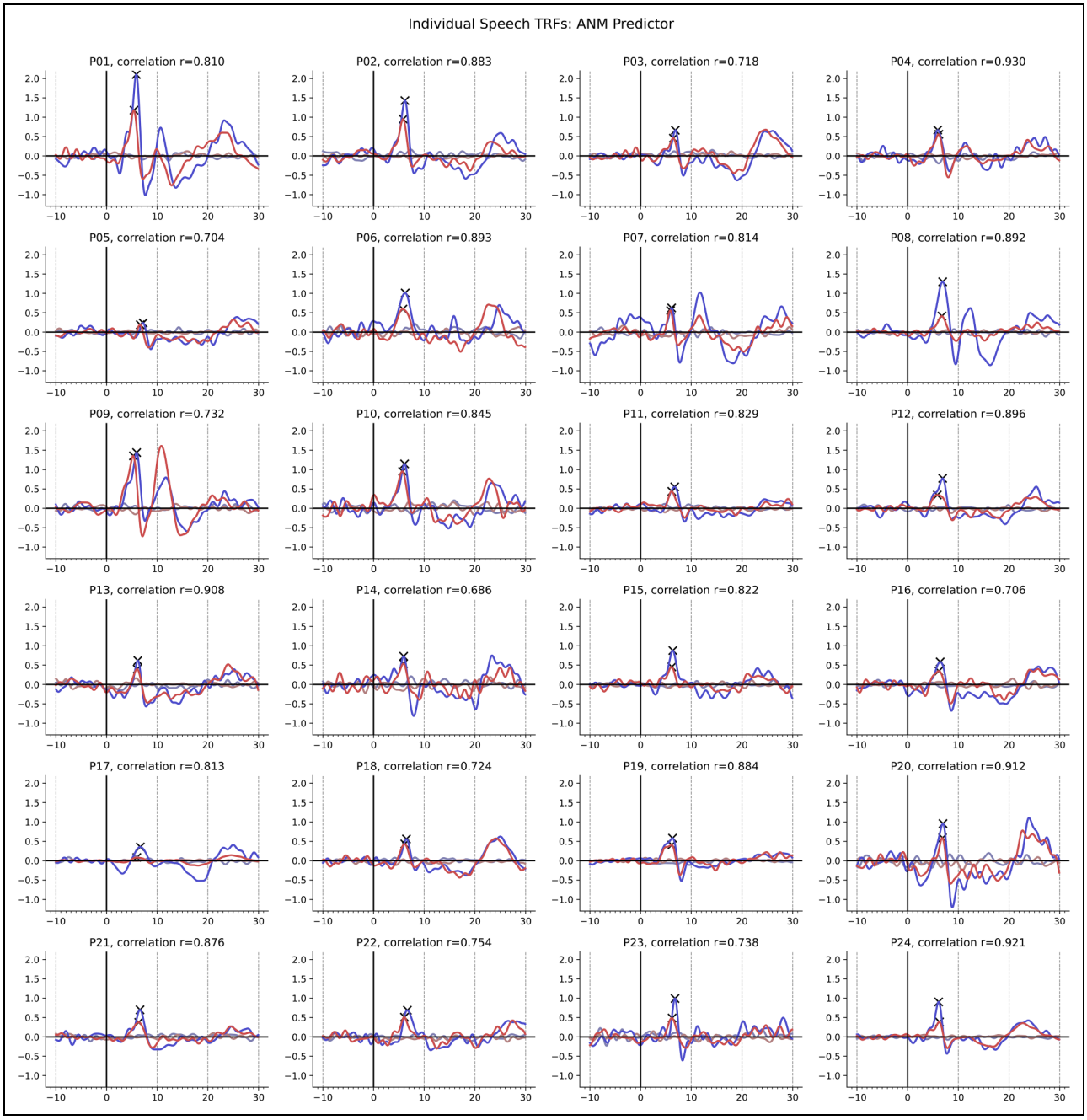

Figure 4. Individual speech TRFs for the ANM predictor. Individual participant TRFs are shown for both insert earphone (blue) and sound-field (red) conditions, along with corresponding null models (lighter colors). Clear wave V peaks are obtained in all participants and the TRFs for both conditions are similar on a single participant level (Pearson correlations between the TRF waveforms of both conditions are shown on top of each subplot).

and properties of the auditory system. Although further investigation is required to characterize systematic differences such as the reduced amplitude in sound-field conditions, it is clear that subcortical responses can be reliably detected in sound-field conditions that are mostly consistent with responses detected using insert earphones on an individual basis.

Discussion

Subcortical TRFs Show Clear Wave V Peaks in Both Insert Earphone and Sound-Field Conditions

Brainstem responses to clicks and speech were detected using ERPs and TRFs respectively, with the most prominent feature being the wave V peak. Speech TRFs were estimated

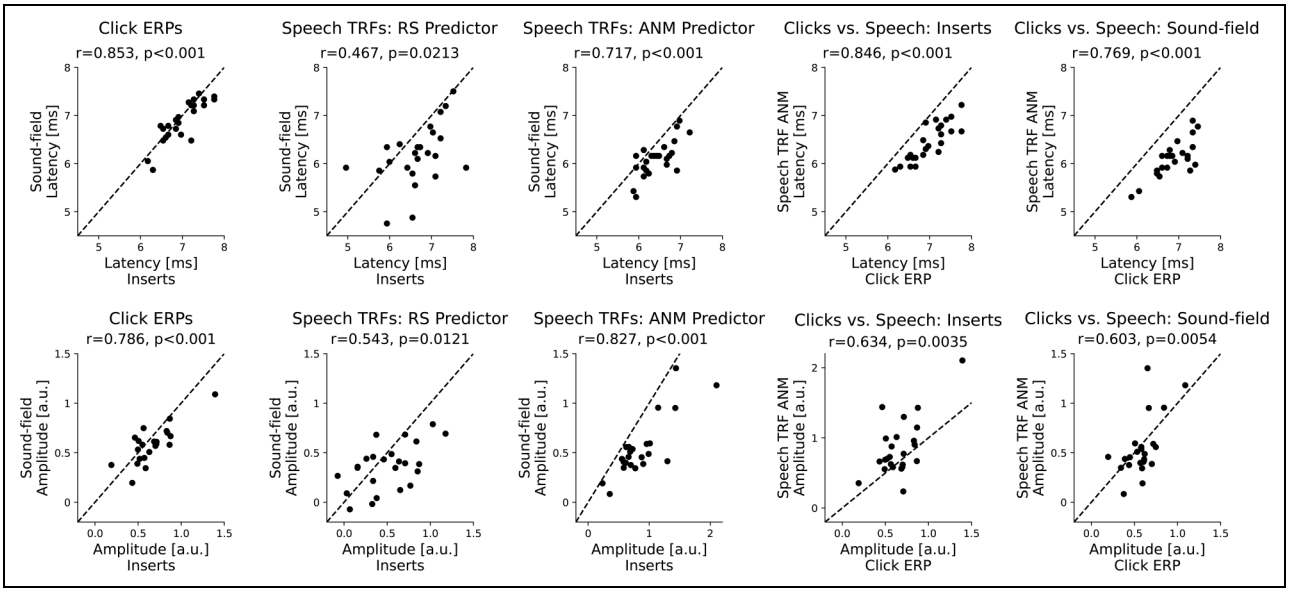

Figure 5. Individual wave V peak latencies and amplitudes. Scatterplots across participants are shown. *Top row:* wave V peak latencies, *Bottom row:* wave V peak amplitudes. The three leftmost columns show the comparison between insert and sound-field condition for click ERPs, Speech TRF with RS predictor and Speech TRFs with the ANM predictor. The Pearson correlations are shown in each title, along with the corresponding *p*-value (Holm–Bonferroni corrected). The wave Vs in the insert and sound-field show a high degree of correlation, especially for click ERPs and ANM TRFs. The two rightmost columns show the comparison between click ERPs and speech ANM TRFs for inserts and sound-field. There is a high degree of correlation for both conditions, indicating that click ERPs and speech ANM TRFs have similar wave Vs. The speech ANM TRF shows an earlier wave V latency compared to the click ERP for all participants (see section "Discussion").

using a simple rectified speech predictor (RS) and a complex auditory nerve model (ANM) predictor. The wave V was detected both when sound was presented through insert earphones and in the sound-field for click ERPs, RS TRFs and ANM TRFs. After accounting for sound propagation delays in the sound-field, response latencies for TRFs generated with the different predictors showed high similarity. TRF amplitudes were reduced when speech was presented in the sound-field versus via insert earphones. TRF estimations using an auditory nerve model consistently outperformed those obtained with simple rectification. Employing the auditory nerve model, slightly more data was required to obtain significant wave V responses from all participants when speech was presented in the sound-field (16 minutes) versus via insert earphones (12 minutes). Clear individual subcortical responses to continuous speech presented in the sound-field were observed for each individual participant.

Challenges When Presenting Stimuli in the Sound-Field

This experiment was conducted in a quiet room, though it was not anechoic or completely sound-proof. Distortions in the signal at the eardrum compared to the stimulus waveform may have occurred due to distortions in the transducers used, reverberation, binaural interactions, and other factors of room acoustics in the sound-field condition. Prior work has shown that room acoustics such as reverberations detrimentally impact both subcortical responses such as ASSRs (Zapata-Rodríguez, 2020) and behavioral measures such as speech intelligibility (Hodgson & Nosal, 2002). The effects of room acoustics lead to a mismatch between the heard signal and the stimulus used to generate the predictor. This may result in a "noisy" linear model with potential smearing which may explain the reduced peak amplitudes seen in our sound-field TRFs. Differences in peak latency should mostly be captured by accounting for the sound travelling delay from the loudspeaker. Indeed, our results confirm that the wave V latencies were quite consistent across both insert and sound-field conditions once the propagation delay was accounted for. However, the observation that responses to speech occurred earlier than to clicks, as previously described by (Maddox & Lee, 2018), was more pronounced in the sound-field (RS: 0.7 ms, ANM: 0.8 ms) compared to the insert earphone (RS: 0.3 ms, ANM: 0.5 ms) condition. Effects of room acoustics might also play a role here. Reverberation time in a given room varies for different frequency bands (Zapata-Rodriguez et al., 2021) and might thus differently affect the click and speech stimuli with their respective frequency content. This may be one underlying reason how effects of room acoustics also have a distinct influence on perceived loudness. Stimuli calibrated to the same intensity at the eardrum are perceived louder when presented through loudspeakers than with headphones, an effect which is more pronounced for lower frequencies (Denk et al., 2021). As louder perceived stimuli show shorter ABR wave V latencies (Serpanos et al., 1997) and speech

contains more energy at lower frequencies than the click sounds, this may lead to earlier ABR wave V peaks in the sound-field, especially for speech stimuli. The impact of room acoustics should thus be considered when using sound-field stimuli for clinical applications, perhaps by modelling the room acoustics and convolving the stimuli with the room impulse response (Zapata-Rodriguez et al., 2021) before generating the predictor. Another option would be to generate predictors from a recording of the audio signal heard at the ear, which may be suitable for applications using hearing aids with built-in microphones. However, our work shows that even without applying these corrections, it is possible to detect meaningful subcortical responses to sound-field speech stimuli, thereby laying the groundwork for applications of such techniques in settings that may not be as tightly controlled.

Predictors Derived from Auditory Models Improve TRF Estimates

Consistent with prior work (Kulasingham et al., 2024; Shan et al., 2024), we show that the ANM predictor greatly outperforms the RS predictor in terms of model fits, wave V SNRs, and amplitude of wave V peaks at an individual participant level. However, it should be noted that the ANM predictor also resulted in slightly earlier wave V peaks (around 0.2 ms) compared to both the RS predictor and the click ERPs. This difference in wave V peak latency was seen even after the inherent delay in the ANM predictor was compensated for by delaying the predictor by 1.1 ms. It is possible that this wave V peak latency shift is due to the fact that the ANM predictor represents the signal at a stage in the auditory pathway that is closer to the wave V generator, compared to the RS predictor. This could result in a smaller delay for the wave V in the ANM TRF and might explain that wave I is undetectable in our ANM TRFs, as the ANM accounts for auditory nerve processing. However, we cannot exclude that this delay may be due to poorer temporal resolution of the TRFs after lowpass filtering at 500 Hz, or due to a misaligned predictor. Accounting for the inherent delay is not straightforward, as the auditory model results in varying delays based on the input stimulus properties, to better simulate realistic auditory system behavior. Our correlation method might not optimally capture this, and further work is needed to determine a more accurate method to align the ANM predictor, and to investigate the interplay between AN modelling delays and the impact of using a predictor representation that is closer to wave V generators. While our present work focuses on stable, automatic wave V detection that aims to lay the groundwork for potential later clinical implementation, we have also explored the possibility of identifying earlier ABR waves. When applying a broader first-order Butterworth IIR filter between 150 and 500 Hz instead of our 30–500 Hz FIR filter, waves I and III (or possibly IV) became apparent in the group-level click ERPs. For the group-level speech ANM TRFs, there was a slight deflection that may indicate wave III (or IV), but wave I activity could not be identified (see Supplementary Figure S3). However, this filtering method resulted in noisier TRFs and complicated the automatic detection of wave V on individual level, and was thus not used in our primary analysis. Wave I might remain undetectable due to the noise floor, or because including the ANM already accounts for the auditory nerve processing generating wave I. Other types of auditory models could also be used to generate predictors (e.g., Dau et al., 1996a, 1996b; Osses Vecchi & Kohlrausch, 2021; Verhulst et al., 2015) and may result in improved TRFs, perhaps using only early stages of complex models might contribute to clearer observations of early peaks (waves I–IV). A comparison of several auditory peripheral models in terms of simulating the auditory system is provided in Vecchi et al. (2022). A more direct comparison of auditory model predictors for subcortical TRF estimation is provided in our recent work (Kulasingham et al., 2024).

More Data Required for Estimating Subcortical TRFs in the Sound-Field Than with Insert Earphones

Incorporating the auditory nerve model reduced the amount of data needed to estimate a TRF with a positive SNR to only 12 to 16 minutes. However, slightly more data was still needed to obtain clear responses ($\geq$0 dB SNR) for all participants when speech was presented in the sound-field (16 minutes) compared to when speech was presented through insert earphones (12 minutes). This is likely due to room acoustics affecting speech propagating through the sound-field, resulting in differences between the employed predictors and the signal reaching the ear drum, and could be alleviated by basing predictors on sound recorded close to the ear. The TRF methods used in this work closely follow the techniques provided in previous investigations into subcortical responses to speech (Bachmann et al., 2019, 2021; Maddox & Lee, 2018; Polonenko & Maddox, 2021). However, there are several possible alternatives that could be explored that may optimize the estimation of subcortical TRFs, and reduce the amount of data required for clear subcortical responses. In this work, the Cz channel referenced to the linked mastoids was used for all analysis, but other EEG channels and reference schemes have also been used for ABR ERPs (Skoe & Kraus, 2010). Indeed a multichannel approach could allow for the incorporation of more advanced preprocessing and artifact removal methods using spatial filters such as ICA similar to what is commonly done for cortical ERPs or TRFs, albeit at the expense of increased sensors and more experimental burden. Another large difference between our method of estimating subcortical TRFs and the widely used *cortical* TRF methods is the

lack of regularization. Almost all methods for estimating cortical TRFs use some form of regularization (Alickovic et al., 2019; Brodbeck et al., 2021; Crosse et al., 2021; Kulasingham & Simon, 2023), in order to reduce noise and produce interpretable TRF peaks. However, previous work has shown that the need for regularization is greatly reduced for analyses using fast-changing predictors with low auto-correlation, as employed here (Bachmann et al., 2021). Furthermore, optimizing the regularization parameter could greatly increase the analysis time and the fine-tuning required for reasonable TRF results. We found no need to use regularization, since our goal in this work was to detect subcortical TRFs using simpler algorithms that are consistent with methods used in prior work without regularization (Maddox & Lee, 2018; Shan et al., 2024). Indeed, our work shows that it is possible to detect subcortical wave Vs using a simple artifact rejection procedure and a single-channel electrode configuration using unregularized TRFs and the ANM predictor. However, investigating more complex alternatives might provide more insights using less EEG data, albeit with larger computational costs.

Potential Audiological Applications

Clear subcortical responses to continuous speech presented in the sound-field were found for all participants on an individual level, which is crucial for potential audiological applications. Measuring subcortical neural responses to naturalistic sound-field stimuli offers the possibility of objective hearing assessment in a more life-like setting, with realistic testing conditions and using an ecologically relevant, complex stimuli, which might ultimately enable gaining a better representation of people's hearing ability in daily life. Objective hearing testing is of special relevance for populations such as newborns or people with neurodegenerative diseases who are unable to provide reliable feedback required for behavioral hearing testing (e.g., pure-tone thresholds). Especially in these populations, eliminating the need for headphones might alleviate agitation associated with the testing situation. Measuring in the sound-field additionally offers the option to include assistive hearing devices, and investigate their effect on subcortical neural processing. For example, the effect of directionality algorithms for noise suppression in hearing aids could be evaluated at the brainstem level, similar to recent work investigating such effects at the cortical level (Alickovic et al., 2020, 2021; Andersen et al., 2021). Even though we demonstrate that subcortical responses can be obtained to speech presented in the sound-field, our work shows that such responses may have been affected by room acoustics, and points toward the possibility of improved brainstem responses when estimated using sound recorded close to the ear. Combined with the information it may reveal about hearing status, this might pave the way toward smart assistive hearing technologies in the future.

Conclusion

Our work demonstrates that brainstem responses can be detected to continuous speech presented in the sound-field, which largely corresponded to those measured when presenting a click, and when speech was delivered via insert earphones. This brings the assessment of early objective sound processing closer to real-life conditions. We show that incorporating models of nonlinear neural processing along the auditory pathway up to the brainstem improves subcortical response estimation beyond simple halfwave rectification techniques, pointing toward the importance of predictors closely matched with the neural representation of sound at the processing stage under investigation. Comparing TRFs obtained using different auditory models in predictor estimation might provide valuable experimental feedback at the brainstem level for speech processing models. Furthermore, using tailored predictors might offer the possibility to distinctly target certain processing stages along the auditory pathway. Due to effects of room acoustics, response amplitudes were reduced when sound was presented in the sound-field instead of via insert earphones, and thus slightly longer EEG recording time was needed in the sound-field. Even so, our approach measuring subcortical neural responses to continuous speech presented in the sound-field and analyzed incorporating an auditory nerve model yields clear neural responses when computed on a 16-minute portion of the data. Crucially, clear wave V peaks were obtained for each participant, which is an essential prerequisite for potential clinical applications. These insights could pave the way toward objective hearing evaluation using ecologically relevant stimuli in a more realistic setting, and future smart hearing solutions.

Acknowledgments

This work was supported by the William Demant Foundation. The authors are also grateful to all participants for their participation in this study.

Data Availability

There are ethical restrictions on sharing the data set. The consent given by participants at the outset of this study did not explicitly detail sharing of the data in any format; this limitation is keeping with EU General Data Protection Regulation, and is imposed by the Research Ethics Committees of the Capital Region of Denmark. Due to this regulation and the way data was collected with low number of participants, it is not possible to fully anonymize the dataset and hence cannot be shared. As a non-author contact point, data requests can be sent to Claus Nielsen, Eriksholm research operations manager at mailto:clni@eriksholm.com.

Declaration of Conflicting Interests

The authors declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: The commercial affiliation of authors FLB, KE, EA, and

HI does not alter our adherence to Trends in Hearing policies on sharing data and materials.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the William Demant Fonden (grant number Case no. 20-0480).

ORCID iDs

Florine L. Bachmann https://orcid.org/0000-0001-6731-6032
Joshua P. Kulasingham https://orcid.org/0000-0003-3599-9160
Kasper Eskelund https://orcid.org/0000-0002-6576-1930
Hamish Innes-Brown https://orcid.org/0000-0002-1512-2823

Supplemental Material

Supplemental material for this article is available online.

References

Abdala, C., & Folsom, R. C. (1995). The development of frequency resolution in humans as revealed by the auditory brain-stem response recorded with notched-noise masking. *The Journal of the Acoustical Society of America*, *98*(2), 921–930. https://doi.org/10.1121/1.414350

Alickovic, E., Lunner, T., Gustafsson, F., & Ljung, L. (2019). A tutorial on auditory attention identification methods. *Frontiers in Neuroscience*, *13*, 153. https://doi.org/10.3389/fnins.2019.00153

Alickovic, E., Lunner, T., Wendt, D., Fiedler, L., Hietkamp, R., Ng, E. H. N., & Graversen, C. (2020). Neural representation enhanced for speech and reduced for background noise with a hearing aid noise reduction scheme during a selective attention task. *Frontiers in Neuroscience*, *14*, 846. https://doi.org/10.3389/fnins.2020.00846

Alickovic, E., Ng, E. H. N., Fiedler, L., Santurette, S., Innes-Brown, H., & Graversen, C. (2021). Effects of hearing aid noise reduction on early and late cortical representations of competing talkers in noise. *Frontiers in Neuroscience*, *15*, 636060. https://doi.org/10.3389/fnins.2021.636060

Andersen, A. H., Santurette, S., Pedersen, M. S., Alickovic, E., Fiedler, L., Jensen, J., & Behrens, T. (2021). Creating clarity in noisy environments by using deep learning in hearing aids. In *Seminars in Hearing 42*(3), 260–281). https://doi.org/10.1055/s-0041-1735134

Bachmann, F. L., MacDonald, E. N., & Hjortkjær, J. (2019). A comparison of two measures of subcortical responses to ongoing speech: Preliminary results. *Proceedings of the International Symposium on Auditory and Audiological Research*, *7*, 461–468. Retrieved from https://proceedings.isaar.eu/index.php/isaarproc/article/view/2019-54

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

Brodbeck, C., Das, P., Kulasingham, J. P., Bhattasali, S., Gaston, P., Resnik, P., & Simon, J. Z. (2021). Eelbrain: A python toolkit for time-continuous analysis with temporal response functions. https://doi.org/10.1101/2021.08.01.454687.

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

Carta, S., Alickovic, E., Zaar, J., Valdes, A. L., & Liberto, G. M. D. (2023). Cortical over-representation of phonetic onsets of ignored speech in hearing impaired individuals [preprint]. *bioRxiv*. https://doi.org/10.1101/2023.06.26.546549.

Chandrasekaran, B., & Kraus, N. (2010). The scalp-recorded brainstem response to speech: Neural origins and plasticity. *Psychophysiology*, *47*(2), 236–246. https://doi.org/10.1111/j.1469-8986.2009.00928.x

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

Crosse, M. J., Zuk, N. J., Di Liberto, G. M., Nidiffer, A. R., Molholm, S., & Lalor, E. C. (2021). Linear modeling of neurophysiological responses to speech and other continuous stimuli: Methodological considerations for applied research. *Frontiers in Neuroscience*, *15*, 705621. https://doi.org/10.3389/fnins.2021.705621

Damarla, V. K., & Manjula, P. (2007). Application of ASSR in the hearing aid selection process. *Australian and New Zealand Journal of Audiology*, *29*(2), 89–97. https://doi.org/10.1375/audi.29.2.89

Dau, T., Püschel, D., & Kohlrausch, A. (1996a). A quantitative model of the "effective" signal processing in the auditory system. I. Model structure. *The Journal of the Acoustical Society of America*, *99*(6), 3615–3622. https://doi.org/10.1121/1.414959

Dau, T., Püschel, D., & Kohlrausch, A. (1996b). A quantitative model of the "effective" signal processing in the auditory system. II. Simulations and measurements. *The Journal of the Acoustical Society of America*, *99*(6), 3623–3631. https://doi.org/10.1121/1.414960

Denk, F., Kohnen, M., Llorca-Bofí, J., Vorländer, M., & Kollmeier, B. (2021). The "missing 6 db" revisited: Influence of room acoustics and binaural parameters on the loudness mismatch between headphones and loudspeakers. *Frontiers in Psychology*, *12*, 623670. https://doi.org/10.3389/fpsyg.2021.623670

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

Etard, O., Kegler, M., Braiman, C., Forte, A. E., & Reichenbach, T. (2019). Decoding of selective attention to continuous speech from the human auditory brainstem response. *NeuroImage*, *200*, 1–11. http://doi.org/10.1016/j.neuroimage.2019.06.029

Galambos, R., & Despland, P.-A. (1980). The auditory brainstem response (ABR) evaluates risk factors for hearing loss in the newborn. *Pediatric Research*, *14*(2), 159–163. https://doi.org/10.1203/00006450-198002000-00019

Garnham, J., Cope, Y., Durst, C., McCormick, B., & Mason, S. M. (2000). Assessment of aided ABR thresholds before cochlear implantation. *British Journal of Audiology*, *34*(5), 267–278. https://doi.org/10.3109/03005364000000138

Guérit, J. M. (1999). EEG and evoked potentials in the intensive care unit. *Neurophysiologie Clinique/Clinical Neurophysiology*, *29*(4), 301–317. https://doi.org/10.1016/S0987-7053(99)90044-8

Hamilton, L. S., & Huth, A. G. (2020). The revolution will not be controlled: Natural stimuli in speech neuroscience. *Language,*

Cognition and Neuroscience, *35*(5), 573–582. https://doi.org/10.1080/23273798.2018.1499946

Hernández-Pérez, H., & Torres-Fortuny, A. (2013). Auditory steady state response in sound field. *International Journal of Audiology*, *52*(2), 139–143. https://doi.org/10.3109/14992027.2012.727103

Hodgson, M., & Nosal, E.-M. (2002). Effect of noise and occupancy on optimal reverberation times for speech intelligibility in classrooms. *The Journal of the Acoustical Society of America*, *111*(2), 931–939. https://doi.org/10.1121/1.1428264

Jarollahi, F., Valadbeigi, A., Jalaei, B., Maarefvand, M., Zarandy, M. M., Haghani, H., & Shirzhiyan, Z. (2020). Sound-field speech evoked auditory brainstem response in cochlear-implant recipients. *Journal of Audiology & Otology*, *24*(2), 71–78. https://doi.org/10.7874/jao.2019.00353

Joint Committee on Infant Hearing (2007). Year 2007 position statement: Principles and guidelines for early hearing detection and intervention programs. (PS2007-00281), PS2007–00281. https://doi.org/10.1044/policy.PS2007-00281

Kim, Y.-H., Schrode, K. M., & Lauer, A. M. (2022). Auditory brainstem response (ABR) measurements in small mammals. In A. K. Groves (Ed.), *Developmental, physiological, and functional neurobiology of the inner ear. Neuromethods* (Vol. 176, pp. 357–375). Humana. https://doi.org/10.1007/978-1-0716-2022-9_16

Kulasingham, J. P., Bachmann, F. L., Eskelund, K., Enqvist, M., Alickovic, E., & Innes-Brown, H. (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., Brodbeck, C., Presacco, A., Kuchinsky, S. E., Anderson, S., & Simon, J. Z. (2020). High gamma cortical processing of continuous speech in younger and older listeners. *NeuroImage*, *222*, 117291. https://doi.org/10.1016/j.neuroimage.2020.117291

Kulasingham, J. P., & Simon, J. Z. (2023). Algorithms for estimating time-locked neural response components in cortical processing of continuous speech. *IEEE Transactions on Biomedical Engineering*, *70*(1), 88–96. https://doi.org/10.1109/TBME.2022.3185005

Land, R., Burghard, A., & Kral, A. (2016). The contribution of inferior colliculus activity to the auditory brainstem response (ABR) in mice. *Hearing Research*, *341*, 109–118. https://doi.org/10.1016/j.heares.2016.08.008

Lunner, T., Alickovic, E., Graversen, C., Ng, E. H. N., Wendt, D., & Keidser, G. (2020). Three new outcome measures that tap into cognitive processes required for real-life communication. *Ear and Hearing*, *41*, 39S–47S. https://doi.org/10.1097/AUD.0000000000000941

Maddox, R. K., & Lee, A. K. C. (2018). Auditory brainstem responses to continuous natural speech in human listeners. *eNeuro*, *5*(1), 1–13. https://doi.org/10.1523/ENEURO.0441-17.2018.

Møller, A. R., & Jannetta, P. J. (1983). Interpretation of brainstem auditory evoked potentials: Results from intracranial recordings in humans. *Scandinavian Audiology*, *12*(2), 125–133. https://doi.org/10.3109/01050398309076235

Moore, J. K. (1987). The human auditory brain stem as a generator of auditory evoked potentials. *Hearing Research*, *29*(1), 33–43. https://doi.org/10.1016/0378-5955(87)90203-6

Osses 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. https://doi.org/10.1121/10.0004818

Picton, T. W. (2010). *Human auditory evoked potentials*. Plural Publishing.

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

Rodrigues, G. R. I., Ramos, N., & Lewis, D. R. (2013). Comparing auditory brainstem responses (ABRs) to toneburst and narrow band CE-chirp® in young infants. *International Journal of Pediatric Otorhinolaryngology*, *77*(9), 1555–1560. https://doi.org/10.1016/j.ijporl.2013.07.003

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

Schebsdat, E., Kohl, M. C., Corona-Strauss, F. I., Seidler, H., & Strauss, D. J. (2018). Free-field evoked auditory brainstem responses in cochlear implant users. *Audiology Research*, *8*(2), 216. https://doi.org/10.4081/audiores.2018.216

Serpanos, Y. C., O'Malley, H., & Gravel, J. S. (1997). The relationship between loudness intensity functions and the click-ABR wave V latency. *Ear and Hearing*, *18*(5), 409–419. https://doi.org/10.1097/00003446-199710000-00006

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

Shemesh, R., Attias, J., Magdoub, H., & Nageris, B. I. (2012). Prediction of aided and unaided audiograms using sound-field auditory steady-state evoked responses. *International Journal of Audiology*, *51*(10), 746–753. https://doi.org/10.3109/14992027.2012.700771

Skoe, E., & Kraus, N. (2010). Auditory brainstem response to complex sounds: A tutorial. *Ear and Hearing*, *31*(3), 302–324. https://doi.org/10.1097/AUD.0b013e3181cdb272

Starr, A. (1976). Correlation between confirmed sites of neurological lesions and abnormalities of farfield auditory brainstem responses. *Electroencephalography and Clinical Neurophysiology*, *41*(6), 595–608. https://doi.org/10.1016/0013-4694(76)90005-5

Stroebel, D., Swanepoel, D., & Groenewald, E. (2007). Aided auditory steady-state responses in infants. *International Journal of Audiology*, *46*(6), 287–292. https://doi.org/10.1080/14992020701212630

Van Canneyt, J., Wouters, J., & Francart, T. (2021a). Cortical compensation for hearing loss, but not age, in neural tracking of the fundamental frequency of the voice. *Journal of Neurophysiology*, *126*(3), 791–802. http://doi.org/10.1152/jn.00156.2021

Van Canneyt, J., Wouters, J., & Francart, T. (2021b). Enhanced neural tracking of the fundamental frequency of the voice. *IEEE Transactions on Biomedical Engineering*, *68*(12), 3612–3619. http://doi.org/10.1109/TBME.2021.3080123

Van Canneyt, J., Wouters, J., Francart, T., & Lalor, E. (2021c). Neural tracking of the fundamental frequency of the voice: The effect of voice characteristics. *European Journal of Neuroscience*, *53*(11), 3640–3653. http://doi.org/10.1111/ejn.v53.11

Vecchi, A. O., Varnet, L., Carney, L. H., Dau, T., Bruce, I. C., Verhulst, S., & Majdak, P. (2022). A comparative study of eight human auditory models of monaural processing. *Acta Acustica*, *6*, 17. https://doi.org/10.1051/aacus/2022008

Verhulst, S., Bharadwaj, H. M., Mehraei, G., Shera, C. A., & Shinn-Cunningham, B. G. (2015). Functional modeling of the human auditory brainstem response to broadband stimulation. *The Journal of the Acoustical Society of America*, *138*(3), 1637–1659. https://doi.org/10.1121/1.4928305

Willott, J. F. (2006). Measurement of the auditory brainstem response (ABR) to study auditory sensitivity in mice. *Current Protocols in Neuroscience*, *34*(1), 8.21B.1–8.21B.12. https://doi.org/10.1002/0471142301.ns0821bs34

Zapata-Rodríguez, V. (2020). Effect of room acoustics on sound-field auditory steady-state response (ASSR) measurements [PhD Thesis]. *Technical University of Denmark*.

Zapata-Rodriguez, V., Laugesen, S., Jeong, C.-H., Brunskog, J., & Harte, J. (2021). Do room acoustics affect the amplitude of sound-field auditory steady-state responses? *Trends in Hearing*, *25*, 2331216520965029. https://doi.org/10.1177/2331216520965029

Zilany, M. S. A., 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. A., 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