Predictors for estimating subcortical EEG responses to continuous speech

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

1 Automatic Control, Department of Electrical Engineering, Linköping University, Sweden
2 Eriksholm Research Centre, Snekkersten, Denmark
3 Oticon A/S, Smørum, Denmark
4 Department of Health Technology, Technical University of Denmark, Lyngby, Denmark
¶ Equally contributed as senior authors

* joshua.kulasingham@liu.se

Abstract

Perception of sounds and speech involves structures in the auditory brainstem that rapidly process ongoing auditory stimuli. The role of these structures in speech processing can be investigated by measuring their electrical activity using scalp-mounted electrodes. However, typical analysis methods involve averaging neural responses to many short repetitive stimuli that bear little relevance to daily listening environments. Recently, subcortical responses to more ecologically relevant continuous speech were detected using linear encoding models. These methods estimate the temporal response function (TRF), which is a regression model that minimises the error between the measured neural signal and a predictor derived from the stimulus. Using predictors that model the highly non-linear peripheral auditory system may improve linear TRF estimation accuracy and peak detection. Here, we compare predictors from both simple and complex peripheral auditory models for estimating brainstem TRFs on electroencephalography (EEG) data from 24 participants listening to continuous speech. We also discuss the data length required for estimating subcortical TRFs with clear peaks. Interestingly, predictors from simple models resulted in TRFs that were similar to those estimated using complex models, and were much faster to compute. This work paves the way for efficient modelling and detection of subcortical processing of continuous speech, which may lead to improved diagnosis metrics for hearing impairment and assistive hearing technology.

Introduction

The human auditory system consists of several subcortical and cortical structures that rapidly process incoming sound signals such as speech. Electroencephalography (EEG) measurements of the aggregate activity of these neural structures have been instrumental in understanding the mechanisms underlying normal hearing and hearing impairments [1, 2]. One important measure is the morphology of the auditory brainstem response (ABR), and the amplitude and latency of ABR peaks have been widely used in many clinical settings such as neonatal hearing screening [3]. Conventional methods to

detect the ABR rely on averaging responses over multiple trials of non-natural, short stimuli such as clicks, chirps or speech syllables [4].

Recently, ABR-like responses to continuous, ongoing speech were detected [5, 6], allowing for the exploration of subcortical processing of ecologically relevant speech stimuli. One method to estimate these subcortical responses is the temporal response function (TRF), a linear encoding model of time-locked neural responses to continuous stimuli [7]. TRFs have been widely used for estimating *cortical* responses to speech [8–12], but fewer studies have investigated *subcortical* TRFs [5, 13–16].

Electrical responses that are generated in the brainstem and measured at the scalp are small compared to the amplitude of the on-going EEG. They are subsequently difficult to detect, and a large amount of data is required for reliable TRF estimation [5, 13]. Additionally, subcortical neural processes rapidly time-lock to fast stimulus fluctuations, and a measurement system with precise synchronization (sub-millisecond) between the stimulus and the EEG is essential in order to extract these responses. Another concern is that linear models may ignore several highly non-linear and adaptive processing stages in the auditory periphery and brainstem [17–19]. The TRF is a linear model that relates the EEG signal to a stimulus-derived predictor, and therefore cannot capture the non-linear processing stages of the peripheral auditory system. However, including these non-linearities in the predictor could lead to improved linear TRF models. Previous work has used the rectified speech waveform as a predictor, which is a coarse approximation of the initial rectifying non-linearity in the cochlea [5]. However, a recent study has shown that predictors derived from a complex model of the auditory periphery [20] that incorporates non-linear stages can lead to improved subcortical TRFs [21]. Another recent study showed that auditory-model-derived predictors outperform previously used envelope predictors even for *cortical* TRFs [22].

It is essential to determine computationally efficient predictors that result in clear TRF peaks for clinical applications involving realistic speech stimuli, or for future assistive hearing technology. Previous work has compared different methods to compute envelope predictors for investigating *cortical* responses to continuous speech [23]. In this work, we compared the rectified speech waveform with predictors derived from various auditory models in terms of their suitability for estimating subcortical TRFs. We computed predictors from filterbank models [17], with or without adaptation [24], and compared them to a more complex auditory nerve model [20, 25] that has been previously used to fit subcortical TRFs [21]. TRFs were estimated from EEG data recorded from 24 participants listening to continuous speech. Prior work indicates that the most prominent feature of subcortical TRFs is the wave V peak [5, 15], which was used as a performance measure in our study. Additional measures such as the computational time taken to generate predictors and the amount of data required for fitting TRFs for each predictor type are also reported.

We corroborate recent findings [21] by confirming that the predictor derived from a complex model [20] of the auditory nerve outperforms the rectified speech predictor. Interestingly, our results indicate that predictors from simpler models [24] can reach similar performance for estimating wave V peaks as complex models, with the added advantage of being more than 50 times faster to compute. These simpler models, combined with TRF analysis, could lead to efficient algorithms for future assistive hearing technology [26], and encourage the use of more ecologically relevant continuous speech stimuli in clinical applications.

Materials and methods

Experimental setup

EEG data was collected from 24 participants with clinically normal hearing thresholds (14 males, M_{age}=37.07, SD_{age}=10.02 years). All participants provided written informed consent and the study was approved by the ethics committee for the capital region of Denmark (journal number 22010204). EEG data was recorded while participants were seated listening to continuous segments from a Danish audiobook of H.C. Andersen adventures, read by Jens Okking. The 2-channel audio was averaged to form a mono audio channel, which was then highpass filtered at 1kHz using a first order Butterworth filter, since the brainstem response is more strongly driven by high frequencies. The participants were instructed to relax and listen to the story. There were four audiobook stories which were presented in randomized order. Stories were split into 4 to 5 minute segments each, leading to a total of 8 trials. The single channel speech segments were scaled to have the same root mean square (r.m.s.) value as a 1 kHz pure tone at 72 dB SPL, and were presented using an RME Fireface UCX soundcard (RME Audio, Haimhausen Germany) and Etymotic ER-2 (Etymotic Research, Illinois, USA) insert earphones, which were shielded using a grounded metal box to avoid direct stimulus artifacts on the EEG. Later visual inspection of TRFs confirmed that stimulus artifacts were not present in the estimated TRFs.

EEG data collection and preprocessing

A Biosemi 32-channel EEG system was used with a sampling frequency of 16,384 Hz and a fifth order cascaded integrator-comb anti-aliasing filter with a -3 dB point at 3276.8 Hz. Electrodes were placed on the mastoids and earlobes, as well as above and below the right eye. Scalp electrodes were placed according to the 10-20 system. Data analysis was conducted in MATLAB (version R2021a) and the Eelbrain Python toolbox (version 0.38.1) [27] using only the Cz electrode referenced to the average of the two mastoid electrodes. The EEG data was highpass filtered using a first order Butterworth filter with cutoff frequency of 1 Hz. To remove power line noise, the signal was passed through FIR notch filters at all multiples of 50 Hz until 1000 Hz, with widths of 5 Hz. The data was then downsampled to 4096 Hz (after passing through an anti-aliasing lowpass filter) to speed up computation. Simple artifact removal was performed by zeroing out 1 second segments around parts of the EEG data that had amplitudes larger than 5 standard deviations above the mean, similar to prior work [5]. Finally, only the data from 2 to 242 seconds of each trial was used for further analysis to avoid onset effects and to have the same amount of data in each trial.

Detecting subcortical responses requires precise synchronization between the EEG and the audio stimuli. Hence, to avoid trigger jitters and clock drifts, the output of the audio interface was also fed to the BioSemi Erg1 channel via an optical isolator to maintain electrical separation between the mains power and the data collection system (StrimTrak, BrainProducts GmbH, Gilching, Germany). The recorded signal from the StimTrak was used to generate predictors for the TRF analysis.

Auditory models

Predictors were computed using several auditory models, described below in order of increasing complexity. For all models, the input was the audio stimulus, as recorded by the StrimTrak system. The lags inherent in the output of each model were accounted for by shifting the generated predictors to maximize the correlation with the rectified speech predictor. Since brainstem responses are largely agnostic to stimulus polarities, a pair of

predictors were generated for each model, using an input stimulus pair with the original stimulus and the stimulus with opposite sign. In line with prior work [5, 21], TRFs were fit to both predictors separately and then averaged together. Although some auditory models account for peripheral rectification, predictor pairs were generated for all models in order to have the same preprocessing setup and since averaging the two TRFs fit to the predictor pair led to cleaner estimates. The overall execution times to generate each predictor and the Pearson correlations between predictors are also reported in Table 1.

Rectified speech (RS)

Previous studies have shown that the rectified speech signal can be used to estimate subcortical TRFs to continuous speech [5]. This method was used for the first predictor pair, termed RS, which was formed by rectifying the speech stimulus (and the stimulus with opposite sign).

Gammatone spectrogram predictor (GT)

Incoming sounds undergo several stages of non-linear processing in the human ear and cochlea. The gammatone filterbank is a simple approximation of this system [28]. A gammatone filterbank consisting of 31 filters from 80-8000 Hz with 1 equivalent rectangular bandwidth (ERB) spacing was applied to the stimulus pair and the resulting amplitude spectra were averaged over all bands to generate the second predictor pair, which was termed GT. The Auditory Modeling Toolbox (AMT) version 1.1.0 [29] (function auditoryfilterbank with default parameters) was used.

Simple model without adaptation (OSS)

The next predictor pair, termed OSS, was generated using the auditory model provided in [24], which is based on the model in [17]. The implementation in AMT (function osses2021) was used. This model consists of an initial headphone and outer ear pre-filter (stage 1), a gammatone filterbank (stage 2), and an approximation of inner hair cell transduction that includes rectification followed by lowpass filtering (stage 3). The next stage of the model consists of adaptation loops (stage 4), which approximate the adaptation properties of the auditory nerve. The initial prefilter was omitted since it is not required for stimuli presented with insert earphones. The adaptation stage was also omitted for this version of the model. Therefore only stages 2 and 3 were used, and the resulting signals with 31 center frequencies (similar to GT) were averaged together to form the predictor pair.

Simple model with adaptation (OSSA)

The adaptation loops (stage 4) of the previous auditory model [24] were now included (i.e., stages 2, 3 and 4 were used). The 31 channel output from the adaptation loops were averaged together to generate the pair of predictors, termed OSSA.

Complex model (ZIL)

Finally, a more complex auditory model [20] was used to generate predictors, and was termed ZIL. This model has been recently used to estimate subcortical TRFs [21] and consists of several stages approximating non-linear cochlear filters, inner and outer hair cell properties, auditory nerve synapses, and adaptation. The implementation in the Python cochlea package [30] was used with 43 auditory nerve fibers with high spontaneous firing rates and center frequencies logarithmically spaced between 125 Hz and 16 kHz, in line with previous work [21]. To speed up computation, an

approximation of the power-law adaptation was used [21]. The outputs of this model are the mean firing rates of the auditory nerves, which were averaged to form the final predictor pair.

Temporal response function estimation

TRFs were fit for each predictor using the frequency domain method outlined in previous studies [5, 13] and shown in eq. (1).

$$TRF = \mathcal{F}^{-1}\left\{\frac{\sum_{i=1}^{N} w_i \mathcal{F}\{x_i\}^* \mathcal{F}\{y_i\}}{\sum_{i=1}^{N} \frac{1}{N} \mathcal{F}\{x_i\}^* \mathcal{F}\{x_i\}}\right\} \tag{1}$$

Here, $\mathcal{F}$ 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 [13], this was done to down-weight noisy (high variance) EEG trials. This frequency domain method results in TRFs with lags from $-T/2$ to $T/2$ where T is the data length.

Two TRFs were estimated separately for each predictor pair, and then averaged together. These TRFs were then bandpass filtered between 30-1000 Hz using a delay compensated FIR filter and then smoothed using a Hamming window of width 2 ms. The smoothing step was necessary since this unregularized TRF approach resulted in noisy estimates for the OSS and OSSA models (see Discussion). Although smoothing could obscure early subcortical peaks, there were no clear early peaks detected visually in the TRFs without smoothing. Given that incorporating smoothing led to more distinct wave V peaks and cleaner TRFs (less noise in the baseline period), it was used for all further TRF analysis. The TRF segment from -10 to 30 ms was extracted for further analysis. Finally, the baseline activity (mean of the TRF segment from -10 to 0 ms) was subtracted from each TRF.

To investigate the effect of data length, TRFs were estimated 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 few 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. This resulted in 7 TRFs for each predictor that allowed for quantifying the improvement of TRF estimation with increasing data length.

Performance metrics and statistical tests

The goodness of fits of the TRF models were evaluated using prediction correlations. The average TRF across positive and negative predictors fit on the training dataset was used to predict the EEG signal of the test trial by convolving it with the appropriate predictors, and the Pearson correlation between the predicted EEG and the actual EEG signal was calculated. The prediction correlations across all cross-validation folds were averaged together to form an estimate of the model fit. To estimate the noise floor, a null model was formed by averaging the prediction correlations from 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 [12]. The same leave-one-out cross-validation approach at each data length was followed for the null models.

The most prominent feature of ABR TRFs is the wave V peak that occurs around 5-10 ms [5, 14, 15]. The amplitude of this wave V peak was used as the primary metric for comparing TRFs from each predictor type. The SNR of the wave V peak was computed, similar to prior work [5]. First, the TRF peak between 5-10 ms was automatically detected, and the power in a 5 ms window around the peak was computed as a measure of the signal power S. Next, the noise power N was estimated as the average TRF power in 5 ms windows in the range -500 to -20 ms. Finally, the wave V SNR was computed as $SNR = 10\log_{10}(S/N)$. Since the signal power cannot theoretically be lower than the noise floor (i.e., 0 dB SNR), negative SNRs were assumed to be meaningless and were set to be 0 dB. The threshold for detecting meaningful wave V peaks was considered to be 3 dB (signal power is twice the noise power).

The amplitudes and latencies of the TRF wave V for each predictor for each participant were also extracted. The consistency of individual wave V was investigated using correlations of wave V amplitudes and latencies across the different predictors.

Statistical analysis was performed using non-parametric tests since the wave V SNRs have a skewed distribution with some TRFs having 0 dB SNRs (i.e., no clear wave V peaks for RS predictor). Non-parametric small sample two-tailed Wilcoxon signed rank tests with Holm Bonferroni multiple comparisons correction were used to test pairwise differences in wave V SNR across predictors. Two participants were excluded from the statistical tests since they did not have data for the full 32 minutes. The group medians, test statistics (rank sums above zero) and p-values are reported.

Results

Subcortical TRFs for predictors derived from auditory models

A comparison of the computational time required to generate each predictor and their correlations with the simplest (RS) and the most complex (ZIL) models are provided in Table 1. The computations were performed on an AMD Ryzen 7 PRO 5850U 1.9 GHz CPU with 32 GB RAM. Note that even the approximate ZIL model is more than 50 times slower than the others.

Table 1. Predictor comparison. The ZIL model is more than 50 times slower to compute than the other models.

Predictor	**Computation Time** (1 s input)	**Correlation with RS**	**Correlation with ZIL**
RS	-	-	0.316
GT	0.0521 s	0.461	0.550
OSS	0.0563 s	0.438	0.496
OSSA	0.0680 s	0.262	0.577
ZIL	4.1208 s	0.316	-

The grand average TRFs for the five predictors over all 24 participants are shown in Fig. 1 on the left panel. The TRFs for all predictors show clear wave V peaks. The wave V peak latency slightly varies across the predictor types, even after removing lags arising from the models themselves by shifting each predictor to have the maximum correlation with RS (see Discussion). The right panel shows the model fits for each predictor as well as the corresponding null model fits. Both OSSA and ZIL show an improvement in model fits over the other 3 models. All the individual TRFs are also shown in Fig. 2, highlighting the consistent wave V peak across all participants.

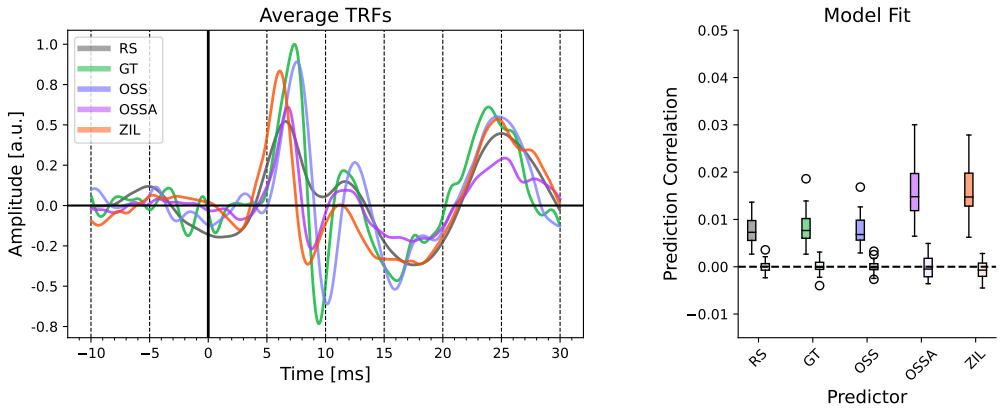


Fig 1. Estimated TRFs for each predictor. Left: The mean TRFs across 24 participants are shown. Clear wave V peaks are seen for all TRFs. **Right**: The model fit prediction correlations across participants are shown for each predictor. The prediction correlations for the null models are shown by the lighter colored boxplots next to each predictor. The OSSA and ZIL models have noticeably better model fits compared to the other models.

Interaction of data length and predictor type on subcortical TRFs

The amount of data required for estimating the subcortical TRFs was investigated by fitting TRFs on an increasing number of 4 minute trials. Two metrics, the model fit and the wave V SNR, were used to compare TRFs across predictors and data lengths as shown in Fig. 3. Almost all participants reached above zero prediction correlation and above 3 dB wave V SNR with 12 minutes of data for the OSSA and ZIL models. Two trends can be observed in Fig. 3; 1) models with filterbanks (GT, OSS) produce wave V estimates with higher SNR compared to RS, 2) models with adaptation and level dependency (OSSA, ZIL) have higher wave V SNR compared to models with filterbanks. Interestingly, wave V SNR and prediction correlation of the simpler OSSA model was comparable to the more complex ZIL model. Statistical tests were performed on the wave V SNR for 32 minutes of data using pairwise non-parametric two-tailed Wilcoxon signed rank tests with Holm-Bonferroni correction. RS had significantly lower wave V SNRs (median 6.55 dB) than all other predictors (GT vs. RS $T = 32$, $p = 0.005$; OSS vs. RS $T = 49$, $p = 0.031$; OSSA vs. RS $T = 0$, $p < 0.001$; ZIL vs. RS $T = 0$, $p < 0.001$). Wave V SNRs for GT (median 9.38 dB) was not significantly different than for OSS (median 9.14 dB) ($T = 59$, $p = 0.055$). The wave V SNRs for OSSA (median 13.58 dB) and ZIL (median 13.99 dB) were larger than for OSS (OSSA vs. OSS $T = 0$, $p < 0.001$; ZIL vs. OSS $T = 2$, $p < 0.001$) or GT (OSSA vs. GT $T = 0$, $p < 0.001$; ZIL vs. OSS $T = 2$, $p < 0.001$). Critically, there was no significant difference in wave V SNRs between OSSA and ZIL ($T = 126$, $p > 0.5$), indicating that the simpler OSSA model provided comparable wave V peak amplitudes to the complex ZIL model.

Individual amplitudes and latencies of wave V

Finally, the TRFs for the OSSA and ZIL predictors were compared as shown in Fig. 4, to further investigate their similarity. The OSSA model showed a high degree of correlation with the ZIL model on a single participant level (Pearson correlation of OSSA vs. ZIL: r=0.865 for wave V SNR, r=0.913 for the peak latencies, r=0.934 for the peak amplitudes, and r=0.852 for model fits). This confirms that both models provide

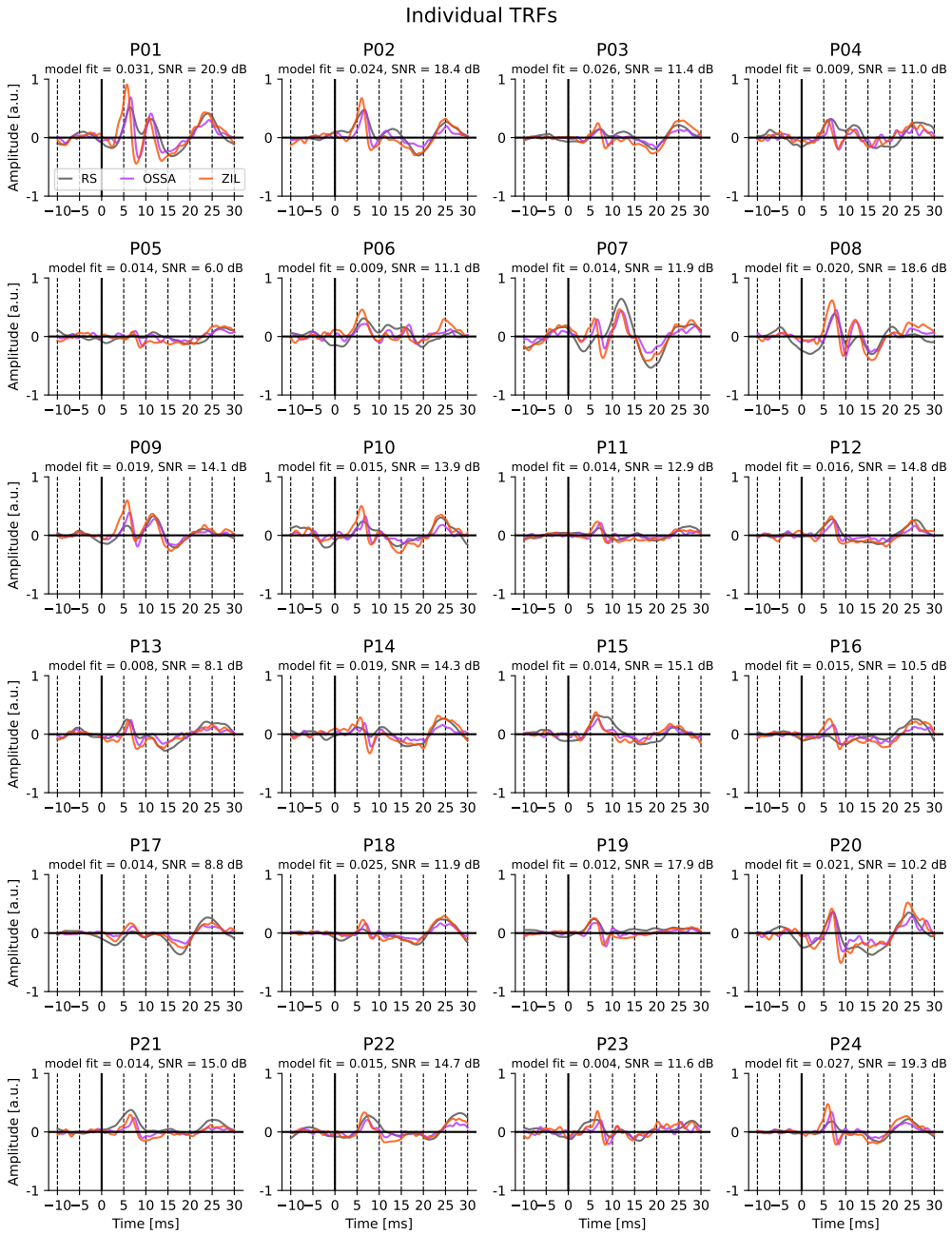


Fig 2. Individual TRFs for all 24 participants. For visual clarity, only TRFs for RS (gray), OSSA (magenta) and ZIL (red) predictors are shown. The model fit prediction correlation and wave V SNR for the ZIL TRF is shown above each subplot. Wave V peaks can be seen for all participants.

TRF wave V estimates that are consistent for each participant. However, the ZIL model has a shorter mean latency, also seen in Fig. 1 and Fig. 2 (see Discussion). Additionally, the OSSA model seems to have slightly smaller wave V peaks than the ZIL model. Nevertheless, this correlation analysis indicates that the simpler OSSA model may provide a good trade-off between computational efficiency and reliable wave V peaks. Additionally, individual wave V SNR and model fits show moderate correlation as seen

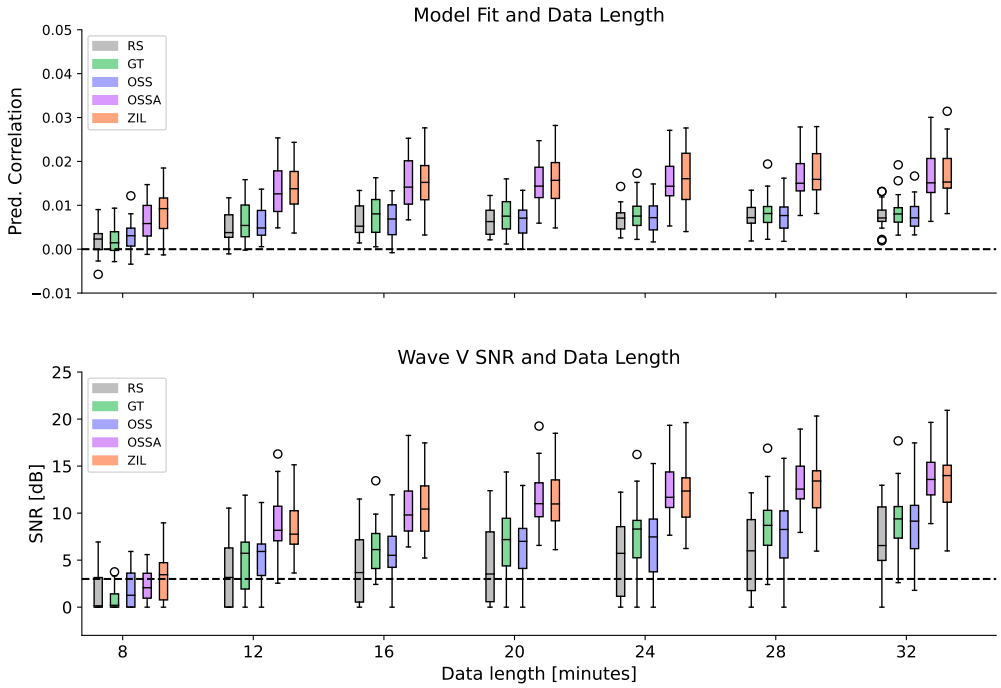


Fig 3. Effect of predictor type and data length. Top: Change in model fit prediction correlation with data length. Prediction correlations are shown after subtracting the corresponding null models. **Bottom**: Change in wave V SNR with data length. The threshold of 3 dB SNR is shown as a dashed line. All boxplots are shown across participants. Model fits are above zero for OSSA and ZIL for all participants after 12 minutes of data. Wave V SNRs are above 3 dB for almost all participants for OSSA and ZIL after 12 minutes of data

in Fig. 4 (OSSA r=0.485, ZIL r=0.538), indicating that higher model fits may not always lead to higher wave V peaks. Therefore the appropriate metric should be considered based on whether the goal is to detect wave Vs or to evaluate model estimation quality.

Discussion

In this work, we compared the suitability of several predictors for estimating subcortical TRFs to continuous speech. We replicated prior work and showed that including non-linearities in the predictor using auditory models leads to improved linear TRF estimates. Our results indicate that the addition of filterbanks and adaptation stages to the predictor models greatly improves estimation of wave V in the TRFs over the rectified speech predictor. Critically, we show that even simpler models may allow for robust model fits and wave V peaks using around 12 minutes of data. These simple models give TRFs that are comparable to a more complex model, even though the complex model is more than 50 times slower to compute. However, it must be noted that OSSA wave V SNRs were comparable to ZIL only after smoothing the TRFs using a 2 ms Hamming window (see Methods), perhaps because the OSSA TRFs were noisier. Other methods such as regularized regression, which is widely used for cortical TRFs [9, 31, 32], or direct estimation of TRF peaks [33] may be able to overcome this issue. Nevertheless, our correlation analysis revealed that these smoothed TRFs resulted

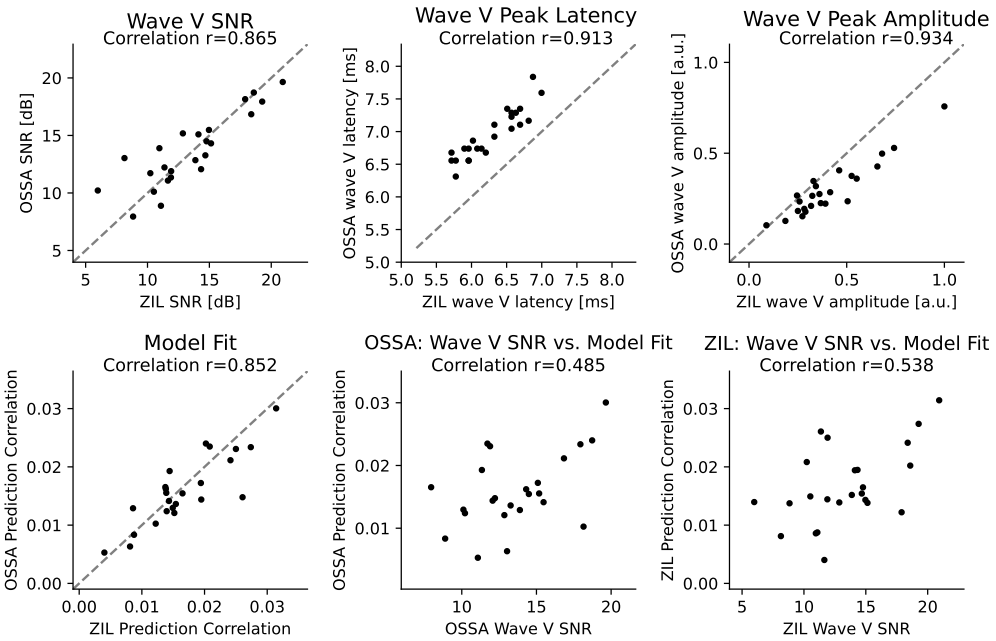


Fig 4. Comparison of ZIL and OSSA models. The ZIL and OSSA models are compared using several metrics and scatterplots across participants are shown. **Top row**: Wave V peak SNR (left), latency (middle) and amplitudes (right). A high degree of correlation is seen for all three metrics. **Bottom row**: Model fits (left), model fits vs. wave V SNR for OSSA (middle) and ZIL (right). Both OSSA and ZIL models give consistent results at an individual participant level. However, the latter two subplots do not show as high a correlation between wave V SNRs and model fits for either model, indicating that it is important to look at both metrics when drawing conclusions regarding wave V peaks or TRF estimation accuracy.

in wave V peak amplitudes and latencies for OSSA and ZIL that were consistent across participants.

The auditory models considered in this work can be categorized into three groups: rectification only (RS), models with filterbanks but without adaptation (GT and OSS), and models with adaptation (OSSA, ZIL). It should be noted that models with filterbanks provided an improvement in TRFs over rectification alone, and that models with adaptation provide the best TRFs. These results are as expected, since including these non-linearities of the peripheral auditory system in the predictor should lead to better linear TRF models.

However, it is surprising that the simpler OSSA model performs as well as the more complex ZIL model. The OSSA model is a functional model that simulates behavioral results while the ZIL model is a phenomenological model that simulates biophysical properties of the neural system [34]. Indeed the ZIL model has several stages that are absent in the OSSA model, such as adaptive filterbanks that simulate inner and outer hair cell activity, power-law adaptation, and models of the auditory synapse. However, our results indicate that perhaps such complex simulations are not necessary for estimating reliable TRF wave Vs. This does not indicate that the OSSA model simulates the auditory system as accurately as the ZIL model for other types of metrics (see [34] for a more detailed comparison of each model), but only that it may suffice for accounting for peripheral non-linearities in TRF estimation in a computationally efficient manner.

This work does not provide an exhaustive list of auditory models or predictors for estimating subcortical TRFs. We also do not directly compare the performance of the auditory models themselves (see [34]), but only evaluate their suitability to generate predictors for subcortical TRFs. Several other models (e.g., [35, 36]) could be utilized to generate predictors, although our work suggests that simple models are reliable enough to fit TRFs with clear wave V peaks.

It must be noted that although the wave V peak was used as the primary metric of performance, the conventional click ABR consists of several other morphological features [1]. The wave V peak was selected here to both be consistent with prior work [5, 15, 21], and because it was the only consistent feature that was visually detected in all participants (see Fig. 2). Conventional click-ABR studies show that early peaks of the ABR are weaker with increasing stimulus rate, and that the wave V is the most consistently detected for different stimulus rates and amplitudes [2]. Therefore, these early peaks may be more difficult to detect using a continuous stimulus like speech, although one study has shown that it may be possible for some participants [21]. Future work could explore if improved predictors or TRF methods could help detect these early subcortical peaks.

TRFs using the ZIL predictor had shorter wave V peak latencies (see Figures 1 and 4), even after accounting for modelling delays by shifting the ZIL predictor to have the maximum correlation with the RS predictor. It is possible that the wave V from the ZIL model is earlier since the ZIL model better incorporates peripheral non-linearities. This may provide a predictor that is similar to intermediate signal representations in the auditory pathway near the wave V generators, which could in turn result in an earlier estimated wave V. Further investigation is needed to disentangle the effects of lags introduced by the auditory peripheral models in order to ascertain whether these latency differences are meaningful properties of the ABR.

Finally, this work only analyses subcortical responses to *speech* stimuli. Recent work indicates that complex auditory model predictors (ZIL) provide significant advantages over rectified speech when estimating subcortical TRFs for *music* [21]. Future work could investigate the suitability of simpler auditory model predictors for estimating TRFs for non-speech stimuli.

Conclusion

This work provides a systematic comparison of predictors derived from auditory peripheral models for estimating subcortical TRFs to continuous speech. Our results indicate that simple models with filterbanks and adaptation loops may suffice to estimate reliable subcortical TRFs. Such efficient algorithms may pave the way toward the use of more ecologically relevant natural speech for investigating hearing impairment and for future assistive hearing technology.

Acknowledgments

The authors are grateful to all participants for their participation in this study.

References

1. Picton TW, Hillyard SA, Krausz HI, Galambos R. Human auditory evoked potentials. I: Evaluation of components. Electroencephalography and Clinical Neurophysiology. 1974;36:179–190.

2. Picton TW. Human Auditory Evoked Potentials. Plural Publishing, Inc.;.

3. Warren MP. The Auditory Brainstem Response in Pediatrics. Otolaryngologic Clinics of N America. 1989;22(3):473–500.

4. Skoe E, Kraus N. Auditory brainstem response to complex sounds: a tutorial. Ear and Hearing. 2010;31(3):302–324. doi:10.1097/AUD.0b013e3181cdb272.

5. Maddox RK, Lee AKC. Auditory Brainstem Responses to Continuous Natural Speech in Human Listeners. eNeuro. 2018;5(1). doi:10.1523/ENEURO.0441-17.2018.

6. Etard O, Kegler M, Braiman C, Forte AE, Reichenbach T. Decoding of selective attention to continuous speech from the human auditory brainstem response. NeuroImage. 2019;200:1–11. doi:10.1016/j.neuroimage.2019.06.029.

7. Lalor EC, Power AJ, Reilly RB, Foxe JJ. Resolving Precise Temporal Processing Properties of the Auditory System Using Continuous Stimuli. Journal of Neurophysiology. 2009;102(1):349–359. doi:10.1152/jn.90896.2008.

8. Di Liberto GM, O'Sullivan JA, Lalor EC. Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. Current Biology. 2015;25(19):2457–2465. doi:10.1016/j.cub.2015.08.030.

9. Alickovic E, Lunner T, Gustafsson F, Ljung L. A tutorial on auditory attention identification methods. Frontiers in Neuroscience. 2019; p. 153.

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

11. Kulasingham JP, Joshi NH, Rezaeizadeh M, Simon JZ. Cortical Processing of Arithmetic and Simple Sentences in an Auditory Attention Task. Journal of Neuroscience. 2021;41(38):8023–8039. doi:10.1523/JNEUROSCI.0269-21.2021.

12. Kulasingham JP, Brodbeck C, Presacco A, Kuchinsky SE, Anderson S, Simon JZ. High gamma cortical processing of continuous speech in younger and older listeners. NeuroImage. 2020;222:117291.

13. Polonenko MJ, Maddox RK. Exposing distinct subcortical components of the auditory brainstem response evoked by continuous naturalistic speech. eLife. 2021;10:e62329. doi:10.7554/eLife.62329.

14. Bachmann FL, MacDonald E, Hjortkjær J. A comparison of two measures of subcortical responses to ongoing speech: Preliminary results. Proceedings of the International Symposium on Auditory and Audiological Research. 2019;7:461–468.

15. Bachmann FL, MacDonald EN, Hjortkjær J. Neural Measures of Pitch Processing in EEG Responses to Running Speech. Frontiers in Neuroscience. 2021;15:738408. doi:10.3389/fnins.2021.738408.

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

17. Dau T, Püschel D, Kohlrausch A. 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. doi:10.1121/1.414959.

18. Dau T, Püschel D, Kohlrausch A. 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. doi:10.1121/1.414960.

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

20. Zilany MSA, Bruce IC, Carney LH. Updated parameters and expanded simulation options for a model of the auditory periphery. The Journal of the Acoustical Society of America. 2014;135(1):283–286. doi:10.1121/1.4837815.

21. Shan T, Cappelloni MS, Maddox RK. Music and Speech Elicit Similar Subcortical Responses in Human Listeners. bioRxiv, preprint. 2022;.

22. Lindboom E, Nidiffer A, Carney LH, Lalor E. Incorporating models of subcortical processing improves the ability to predict EEG responses to natural speech. bioRxiv, preprint. 2023;doi:10.1101/2023.01.02.522438.

23. Biesmans W, Das N, Francart T, Bertrand A. Auditory-Inspired Speech Envelope Extraction Methods for Improved EEG-Based Auditory Attention Detection in a Cocktail Party Scenario;25(5):402–412. doi:10.1109/TNSRE.2016.2571900.

24. Osses Vecchi A, Kohlrausch A. Perceptual similarity between piano notes: Simulations with a template-based perception model. The Journal of the Acoustical Sociecty of America. 2021;149(5):3534. doi:10.1121/10.0004818.

25. Zilany MSA, Bruce IC, Nelson PC, Carney LH. 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). doi:10.1121/1.3238250.

26. Geirnaert S, Vandecappelle S, Alickovic E, de Cheveigne A, Lalor E, Meyer BT, et al. Electroencephalography-Based Auditory Attention Decoding: Toward Neurosteered Hearing Devices. IEEE Signal Processing Magazine. 2021;38(4):89–102. doi:10.1109/MSP.2021.3075932.

27. Brodbeck C, Das P, Kulasingham JP, Bhattasali S, Gaston P, Resnik P, et al. Eelbrain: A Python toolkit for time-continuous analysis with temporal response functions. bioRxiv, preprint. 2022; p. 2021.08.01.454687. doi:10.1101/2021.08.01.454687.

28. Patterson RD, Nimmo-Smith I, Holdsworth J, Rice P. An efficient auditory filterbank based on the gammatone function. In: a meeting of the IOC Speech Group on Auditory Modelling at RSRE. vol. 2; 1987.

29. Majdak P, Hollomey C, Baumgartner R. AMT 1.x: A toolbox for reproducible research in auditory modeling. Acta Acustica. 2022;6:19. doi:10.1051/aacus/2022011.

30. Rudnicki M, Schoppe O, Isik M, Völk F, Hemmert W. Modeling auditory coding: from sound to spikes. Cell and Tissue Research. 2015;361(1):159–175. doi:10.1007/s00441-015-2202-z.

31. Wong DD, Fuglsang SA, Hjortkjær J, Ceolini E, Slaney M, De Cheveigne A. A comparison of regularization methods in forward and backward models for auditory attention decoding. Frontiers in Neuroscience. 2018;12:531.

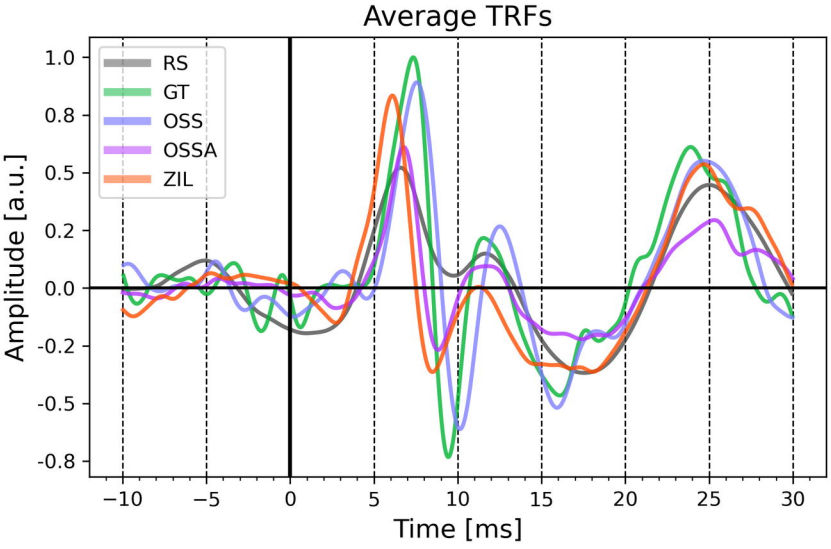
32. Crosse MJ, Zuk NJ, Di Liberto GM, Nidiffer AR, Molholm S, Lalor EC. Linear modeling of neurophysiological responses to speech and other continuous stimuli: methodological considerations for applied research. Frontiers in Neuroscience. 2021; p. 1350.

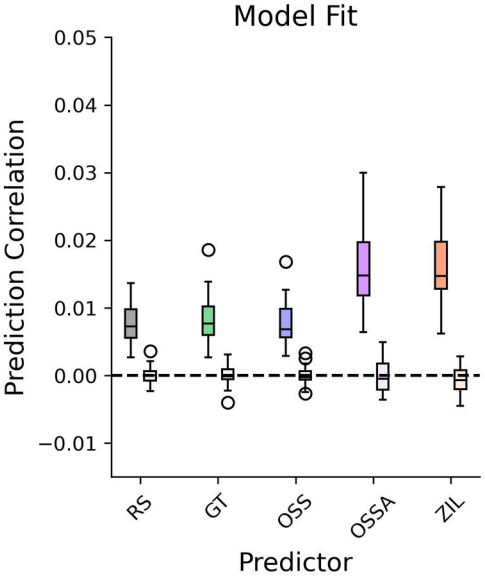
33. Kulasingham JP, Simon JZ. Algorithms for Estimating Time-Locked Neural Response Components in Cortical Processing of Continuous Speech. IEEE Transactions on Biomedical Engineering. 2023;70(1):88–96. doi:10.1109/TBME.2022.3185005.

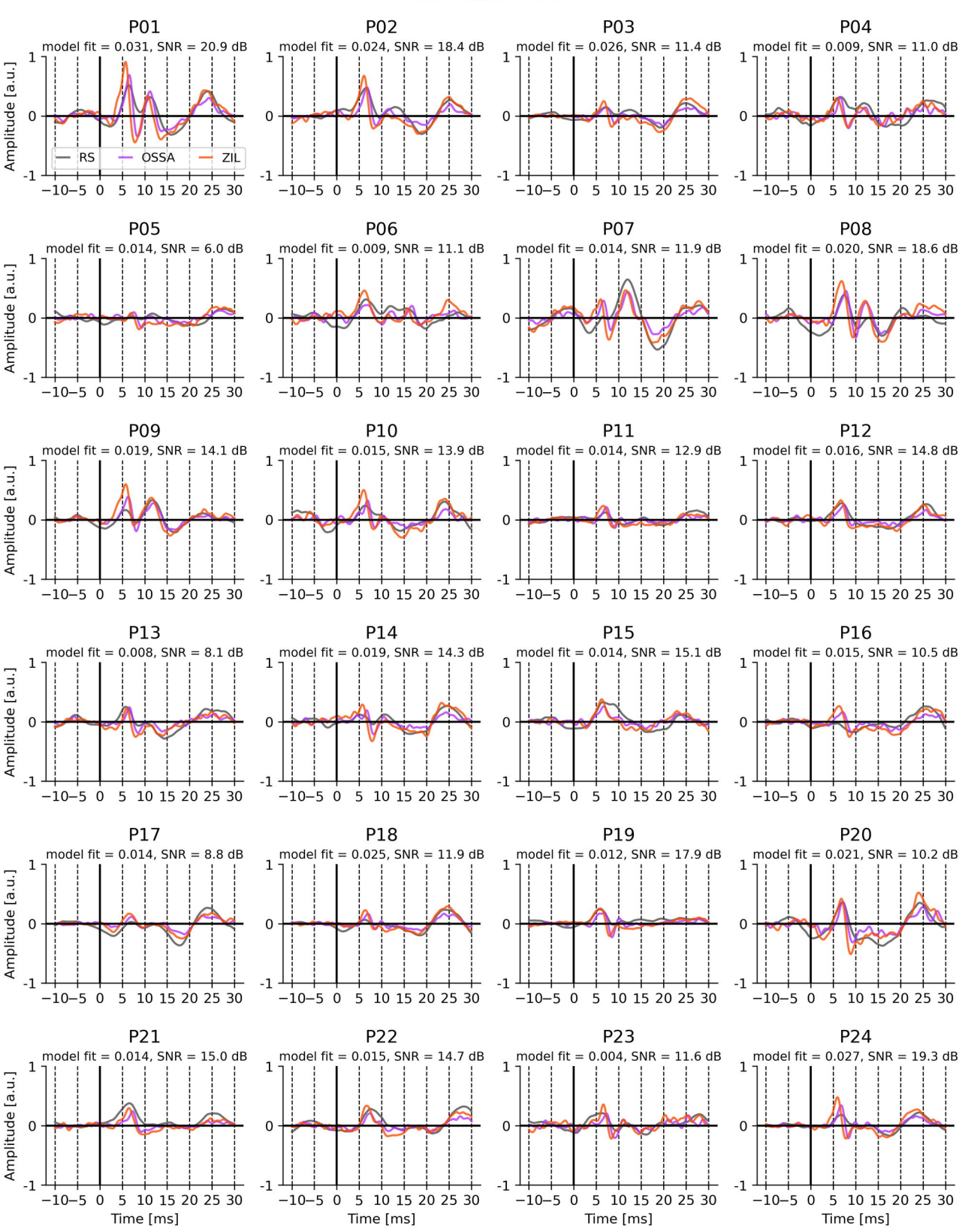
34. Vecchi AO, Varnet L, Carney LH, Dau T, Bruce IC, Verhulst S, et al. A comparative study of eight human auditory models of monaural processing. Acta Acustica. 2022;6:17. doi:10.1051/aacus/2022008.

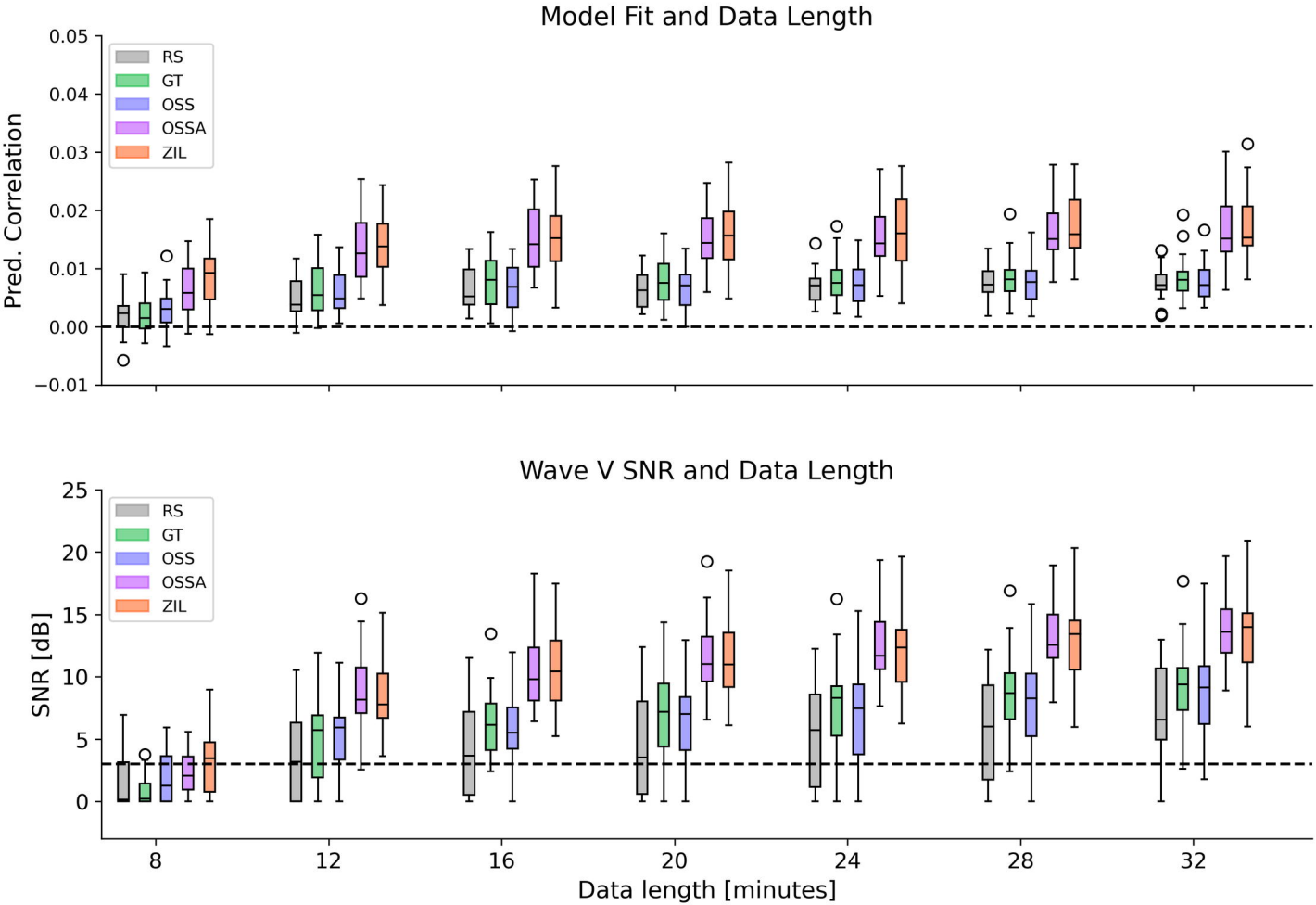
35. Relaño-Iborra H, Zaar J, Dau T. A speech-based computational auditory signal processing and perception model. The Journal of the Acoustical Society of America. 2019;146(5):3306–3317. doi:10.1121/1.5129114.

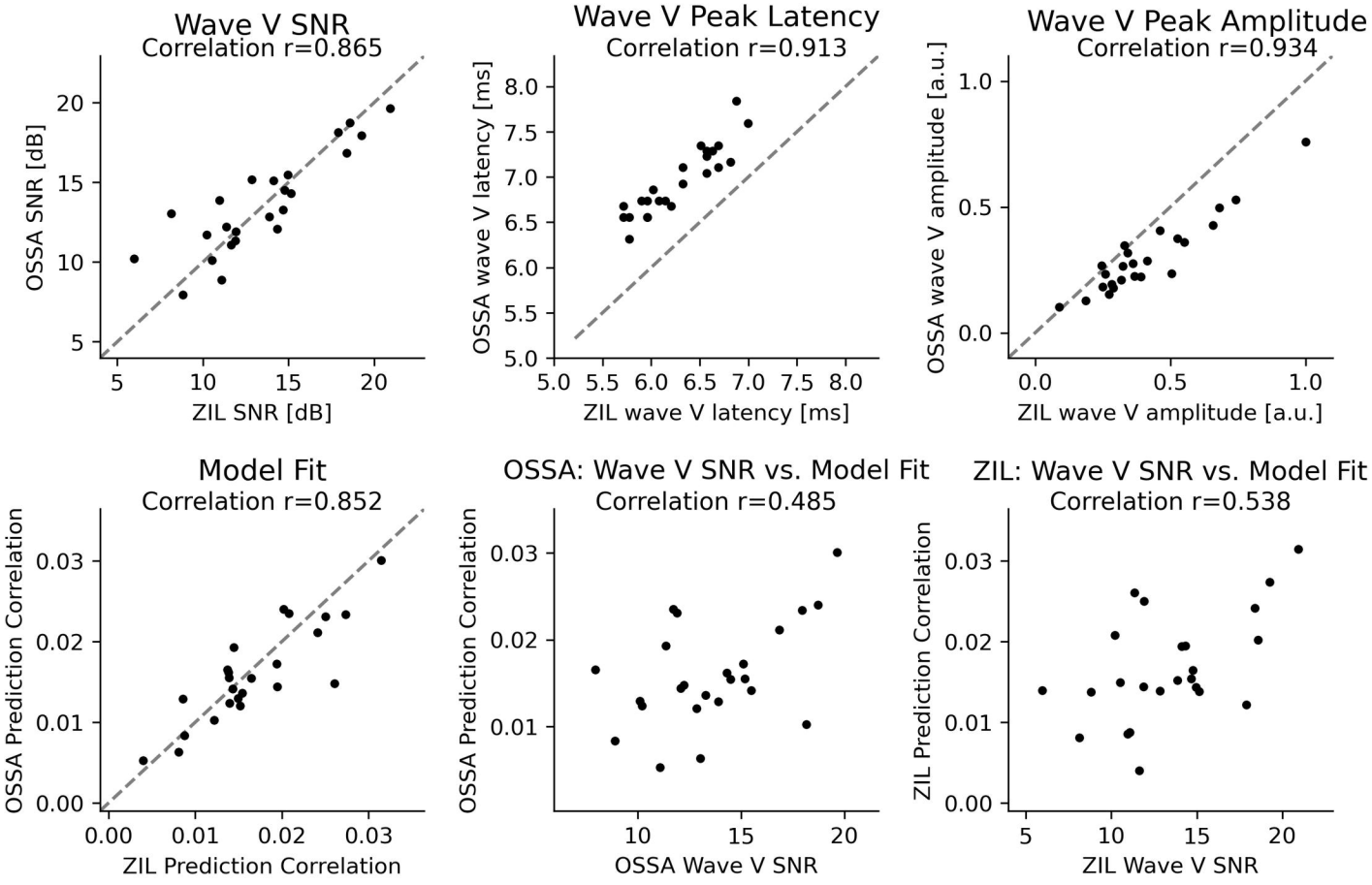
36. Verhulst S, Bharadwaj HM, Mehraei G, Shera CA, Shinn-Cunningham BG. Functional modeling of the human auditory brainstem response to broadband stimulation. The Journal of the Acoustical Society of America. 2015;138(3):1637–1659. doi:10.1121/1.4928305.


Average TRFs
RS
GT
OSS
OSSA
ZIL
Amplitude [a.u.]
1.0
0.8
0.5
0.2
0.0
-0.2
-0.5
-0.8
−10
−5
0
5
10
15
20
25
30
Time [ms]


Model Fit
Prediction Correlation
0.05
0.04
0.03
0.02
0.01
0.00
−0.01
RS
GT
OSS
OSSA
ZIL
Predictor


Individual TRFs
P01
model fit = 0.031, SNR = 20.9 dB
P02
model fit = 0.024, SNR = 18.4 dB
P03
model fit = 0.026, SNR = 11.4 dB
P04
model fit = 0.009, SNR = 11.0 dB
P05
model fit = 0.014, SNR = 6.0 dB
P06
model fit = 0.009, SNR = 11.1 dB
P07
model fit = 0.014, SNR = 11.9 dB
P08
model fit = 0.020, SNR = 18.6 dB
P09
model fit = 0.019, SNR = 14.1 dB
P10
model fit = 0.015, SNR = 13.9 dB
P11
model fit = 0.014, SNR = 12.9 dB
P12
model fit = 0.016, SNR = 14.8 dB
P13
model fit = 0.008, SNR = 8.1 dB
P14
model fit = 0.019, SNR = 14.3 dB
P15
model fit = 0.014, SNR = 15.1 dB
P16
model fit = 0.015, SNR = 10.5 dB
P17
model fit = 0.014, SNR = 8.8 dB
P18
model fit = 0.025, SNR = 11.9 dB
P19
model fit = 0.012, SNR = 17.9 dB
P20
model fit = 0.021, SNR = 10.2 dB
P21
model fit = 0.014, SNR = 15.0 dB
P22
model fit = 0.015, SNR = 14.7 dB
P23
model fit = 0.004, SNR = 11.6 dB
P24
model fit = 0.027, SNR = 19.3 dB
RS
OSSA
ZIL
Amplitude [a.u.]
Time [ms]


Model Fit and Data Length
Pred. Correlation
0.05
0.04
0.03
0.02
0.01
0.00
−0.01
RS
GT
OSS
OSSA
ZIL
Wave V SNR and Data Length
SNR [dB]
25
20
15
10
5
0
RS
GT
OSS
OSSA
ZIL
8
12
16
20
24
28
32
Data length [minutes]


Wave V SNR
Correlation r=0.865
OSSA SNR [dB]
ZIL SNR [dB]
Wave V Peak Latency
Correlation r=0.913
OSSA wave V latency [ms]
ZIL wave V latency [ms]
Wave V Peak Amplitude
Correlation r=0.934
OSSA wave V amplitude [a.u.]
ZIL wave V amplitude [a.u.]
Model Fit
Correlation r=0.852
OSSA Prediction Correlation
ZIL Prediction Correlation
OSSA: Wave V SNR vs. Model Fit
Correlation r=0.485
OSSA Prediction Correlation
OSSA Wave V SNR
ZIL: Wave V SNR vs. Model Fit
Correlation r=0.538
ZIL Prediction Correlation
ZIL Wave V SNR