Improved Modeling of EEG Responses to Natural Speech Acoustics Using Dynamic Temporal Response Functions

Jin Dou (jdou3@ur.rochester.edu)
Department of Biomedical Engineering, University of Rochester
Rochester, NY 14627 United States of America

Andrew J. Anderson (andanderson@mcw.edu)
Department of Neurology, Medical College of Wisconsin
Department of Biomedical Engineering, Medical College of Wisconsin
Department of Neurosurgery, Medical College of Wisconsin
Milwaukee, WI 53226 United States of America

Samuel V. Norman-Haignere (samuel_norman-haignere@urmc.rochester.edu)
Department of Biostatistics and Computational Biology, University of Rochester Medical Center
Department of Biomedical Engineering, University of Rochester
Department of Brain and Cognitive Sciences, University of Rochester
Department of Neuroscience and Del Monte Institute for Neuroscience, University of Rochester Medical Center
Rochester, NY 14627 United States of America

Edmund C. Lalor (elalor@ur.rochester.edu)
Department of Biomedical Engineering, University of Rochester
Department of Neuroscience and Del Monte Institute for Neuroscience, University of Rochester Medical Center
Center for Visual Science, University of Rochester
Rochester, NY 14627 United States of America

Abstract

Speech is central to human life. However, how the human brain converts acoustic speech into language remains incompletely understood. One common way to study this process is by deriving models that map between speech stimuli and the resulting brain responses. The temporal response function (TRF) is one such model that assumes that responses to speech are time-invariant with magnitudes that are linearly related to the amplitude of various speech features. However, such linear time-invariant assumptions are sure to be suboptimal given what is known about the brain. Here, we relax the linear time-invariant assumptions using a recently proposed dynamically warped TRF model that can modulate both the amplitude and timing of the TRF based on the current and previous values of the stimulus feature of interest. Doing this improved the ability to model EEG responses to natural speech. This improvement was driven by the dynamic TRF's ability to account for the fact that larger acoustic onset values tended to evoke larger and earlier responses, a finding that is consistent with previous research. This study validated the efficacy of the dynamically warped TRF model and emphasizes the importance of considering the timing of brain responses to natural stimuli.

Keywords: EEG; speech; computational modeling; temporal response function; envelope acoustic onset

Introduction

Speech is one of the most important signals in human life. One way that researchers have attempted to understand how the human brain processes speech is by modeling neurophysiological responses to natural speech. One common approach involves fitting linear temporal response functions (TRFs) that describe a mapping between speech features (e.g. an envelope) and brain responses. However, this approach assumes that only the amplitude but not timing of brain responses will be modulated by the intensity of speech features. Although this assumption may be applicable for responses at certain stages along the neural pathway over certain time scales (Boynton, Engel, Glover, & Heeger, 1996), there is evidence indicating this assumption is suboptimal for modelling EEG responses. One such piece of evidence is the finding that auditory stimuli with higher intensity evoked responses with shorter latencies (Beagley & Knight, 1967).

Recently, we showed that allowing a linear model to vary as a function of the stimulus feature's intensity leads to improved prediction of unseen neural data (Drennan & Lalor, 2019). To do this, we divided the speech envelope into different intensity bins and fit separate TRFs for each intensity. However, this approach was somewhat ad hoc given that it involved the selection of a certain number of intensity bins, each with a width that was not determined based on any particular principle.

An alternative approach would be to determine how the TRF should be reshaped across stimulus feature intensity to optimize EEG prediction accuracy. Indeed, a recent study showed that taking such an optimization approach was useful in dynamically warping TRFs based on the lexical surprisal of words (Dou, Anderson, White, Norman-Haignere, & Lalor, 2024). In particular, a data-driven approach was used to fit TRFs for individual words, with its amplitude and time latency being modulated by word predictability. The results indicated that words that are easier to predict tend to evoke earlier and smaller EEG responses.

Here, we apply a similar approach to the speech acoustics. In particular, we focus on acoustic onsets – defined as the derivative of the broadband speech envelope, a feature that has been used in several previous studies (Drennan & Lalor, 2019; Brodbeck, Hong, & Simon, 2018; Synigal, Anderson, & Lalor, 2023). The idea is to model dynamic changes in the amplitude and timing of the temporal response function with the intensity of the acoustic onsets and, thus, to produce a more accurate model of EEG responses to natural speech.

Methods

Our hypothesis is that larger acoustic onsets will evoke not only larger but earlier responses (TRFs). We will test our hypothesis by comparing the prediction accuracy between the static (standard) TRF and the dynamically warped TRFs, and by inspecting specific parameters in the dynamic TRF model.

Data

EEG Dataset We used a publicly available natural speech EEG dataset to fit and test our model (Broderick, Anderson, Di Liberto, Crosse, & Lalor, 2018). This dataset contains EEG collected from 19 participants listening to 20 continuous pieces of a narrative audiobook with each piece lasting about 3 minutes. The Biosemi 128-channel EEG recordings were rereferenced to the average data from two mastoid electrodes, band-pass filtered (0.5 to 8 Hz), interpolated (if it was too noisy compared with the surrounding channels, see (Dou et al., 2024) for details), downsampled to 64 Hz, and z-scored.

Stimulus Feature The acoustic onset we used here was calculated as the difference between adjacent samples along the time axis of the broadband speech envelope, a feature that has been used in several previous studies (Drennan & Lalor, 2019; Brodbeck et al., 2018; Synigal et al., 2023). Acoustic onset values smaller than 1e-4 were discarded in the analysis. The acoustic onset time series were z-scored.

Dynamically warped TRF

The static TRF (Crosse, Di Liberto, Bednar, & Lalor, 2016), which can be formulated as $r(t) = s(t) * h(\tau)$, assumes the amplitude of the response, $r(t)$, is linearly modulated by the speech feature intensity s. In other words, $h(\tau)$ is a fixed linear kernel. The dynamically warped TRF, on the other hand, transforms the TRF kernel along both amplitude and time axes for each input. To do this, we first fitted a static TRF for the acoustic onset time series $s(t)$ with a time lag of 0-300 ms. We then obtained its functional representation $h(t)$ by repre-

senting the TRF weights as a weighted sum of Fourier bases, as described in (Dou et al., 2024). This approach provides the benefit of differentiable transformation on the TRF weights. Then, we estimated transformation parameters of amplitude scaling a_i and time shifting b_i for the ith input by convolving two learnable kernels β_1 and β_2 with $s(t)$, respectively:

$$a_i = |\sum_{\mu} \beta_1(\mu) s(t_i - \mu) + \gamma_1| \quad (1)$$

$$b_i = min(max(\sum_{\mu} \beta_2(\mu) s(t_i - \mu) + \gamma_2, -0.1), 0.1) \quad (2)$$

where γ represents the y-intercept, a_i was forced to be non-negative and b_i was limited between -0.1 and 0.1s. The process of obtaining the final predicted responses $r(t)$ can be formulated as:

$$r_i(t) = \begin{cases} a_i h(t - t_i - b_i) & t_i < t < t_i + \tau_{max} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

$$r(t) = \sum_i r_i(t) \quad (4)$$

where t_i indicates the onset time of the ith non-zero value of the acoustic onset and τ_{max} indicates the maximum time lag relative to time t_i. The two convolution kernels had a length of 10, and were optimized using gradient descent provided by pyTorch (Paszke et al., 2019). τ_{max} was set as 300 ms.

Statistics

The performance of the static and dynamic TRFs were compared by assessing how accurately they could predict EEG to unseen stimuli – where prediction accuracy was quantified using Pearson's correlation between the predicted and true EEG responses for each channel. To correct for multiple comparisons when comparing the prediction accuracy of dynamic and static TRF, we used a non-parametric cluster-level paired t-test (right tailed), with 2048 permutations.

Result

As shown in Figure 1, prediction accuracy was significantly improved around central scalp regions. On average across the significant channels, the dynamic TRF improved prediction accuracy significantly from 0.0284 to 0.0295. Please note: these numbers reflect the prediction accuracy of unaveraged EEG. As such, they are small in absolute terms, but highly significant (Di Liberto, O'Sullivan, & Lalor, 2015).

Figure 2 (top) shows that the improved prediction accuracy is driven by a negative correlation between amplitude scaling and time shift (data shown are from one round of 5-fold cross-validation). Figure 2 (bottom) visualizes the TRFs corresponding to the circled dots in Figure 2 (top). The averaged correlation across all rounds of cross-validation between amplitude scaling and time shift is -0.145.

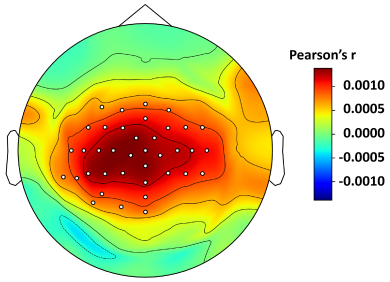

Figure 1: Channel-wise improvement of prediction accuracy.

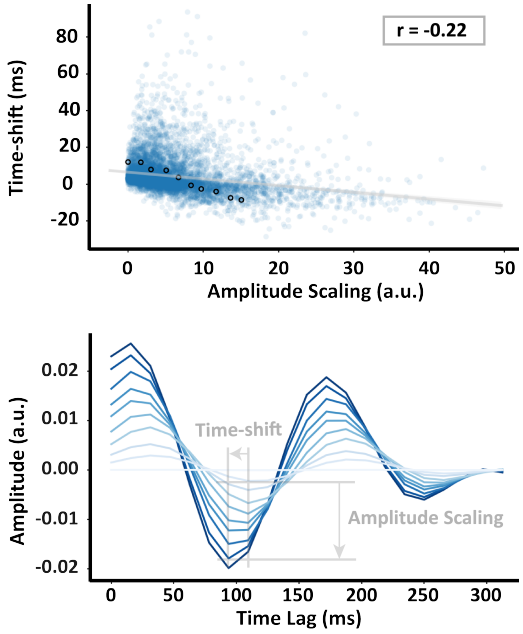

Figure 2: Correlation between amplitude scaling and time shift, and visualization of dynamic TRFs.

Discussion

In this study, we show that a dynamic TRF can improve the prediction accuracy of EEG when using acoustic onset as the predictor. Evidence was also found showing a negative monotonic relationship between the amplitude and time latency of the responses to acoustic onset. Given the challenging signal-to-noise ratio of EEG data, this improved prediction accuracy has important implications for future research on how the human brain processes speech, and how such processing might differ in certain populations (Federmeier, 2022).

Slightly different from what was shown previously (Drennan & Lalor, 2019), here the channels where improvement is significant concentrate more around central electrode sites instead of frontal scalp. One potential reason is the correlation between semantic features and acoustic onset, which may mean that the acoustic onset TRF also reflects brain activity related to the linguistic content of the speech. Future work will examine how to disentangle these contributions.

Acknowledgments

This work was supported by the Del Monte Institute for Neuroscience at the University of Rochester.

References

Beagley, H., & Knight, J. (1967). Changes in auditory evoked response with intensity. *The Journal of Laryngology & Otology*, *81*(8), 861–873.

Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human v1. *Journal of Neuroscience*, *16*(13), 4207–4221.

Brodbeck, C., Hong, L. E., & Simon, J. Z. (2018). Rapid transformation from auditory to linguistic representations of continuous speech. *Current Biology*, *28*(24), 3976–3983.

Broderick, M. P., Anderson, A. J., Di Liberto, G. M., Crosse, M. J., & Lalor, E. C. (2018). Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech. *Current Biology*, *28*(5), 803–809.

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.

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.

Dou, J., Anderson, A. J., White, A., Norman-Haignere, S., & Lalor, E. C. (2024). Dynamic modeling of eeg responses to natural speech reveals earlier processing of predictable words. *bioRxiv*, 2024–04.

Drennan, D. P., & Lalor, E. C. (2019). Cortical tracking of complex sound envelopes: modeling the changes in response with intensity. *eneuro*, *6*(3).

Federmeier, K. D. (2022). Connecting and considering: Electrophysiology provides insights into comprehension. *Psychophysiology*, *59*(1), e13940.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... others (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, *32*.

Synigal, S. R., Anderson, A. J., & Lalor, E. C. (2023). Electrophysiological indices of hierarchical speech processing differentially reflect the comprehension of speech in noise. *bioRxiv*, 2023–03.