On the generative mechanisms underlying the cortical tracking of natural speech

Edmund C Lalor (elalor@ur.rochester.edu)
Department of Biomedical Engineering, Department of Neuroscience, Del Monte Institute for Neuroscience, Center for Visual Neuroscience, University of Rochester, Rochester, NY 14627, USA

Andre Palacios Duran (apalaci6@ur.rochester.edu)
Department of Biomedical Engineering, University of Rochester, Rochester, NY 14627, USA

Aaron R Nidiffer (anidiffe@ur.rochester.edu)
Department of Neuroscience, Del Monte Institute for Neuroscience, University of Rochester, Rochester, NY 14627, USA

Abstract:

Low frequency cortical activity tracks the dynamics of natural speech. However, the mechanisms that produce this tracking are debated. One theory proposes that intrinsic cortical oscillations entrain to the rhythms of speech in an anticipatory manner. Meanwhile, a second theory assumes that neural measures of speech processing reflect transient evoked responses. Here, we attempt to reconcile these theories. We leverage the fact that, when you regress neurophysiological data against (say) the amplitude envelope of speech, you obtain a temporal response function (TRF) that that can reliably predict responses to novel speech stimuli. We then ask: can the existence of TRFs be explained as deriving from the entrainment of an ongoing oscillation? We do this by driving two oscillatory models with speech stimuli, attempting to fit TRFs to the resulting simulated brain activity, and then assessing whether such simulated brain activity can be predicted using the resulting TRF. We find that both models could produce TRFs with predictive power. However, one model is biologically implausible, and the second model produces simulated neural activity and TRFs with highly atypical characteristics. Nonetheless, this study establishes a framework for resolving an important debate in the field of speech neurophysiology.

Keywords: speech; EEG; oscillations; modeling.

Introduction

Speech is central to human life. However, how our brains parse and process speech remains unclear. In recent years, much progress has been made by recognizing that the dynamics of cortical activity “track” the dynamics of natural speech (Ahissar et al., 2001). However, the generative mechanisms of this tracking remain unclear (Obleser & Kayser, 2019). In particular, two mechanistic theories of this tracking have emerged in largely separate literatures. The first posits that intrinsic oscillatory brain rhythms “entrain” to the dynamics of speech (Giraud & Poeppel, 2012). Meanwhile, the second centers on the idea that the neural tracking of speech (or any auditory stimulus) reflects stimulus-driven evoked responses in neuronal populations that are tuned to the features of that stimulus (Crosse et al., 2021). This is typically operationalized by fitting models between different features of a speech stimulus and the associated brain activity – with one such popular modeling framework being the temporal response function (TRF) approach (Crosse et al., 2016)). The goal of the present study is to attempt to reconcile these contrasting ideas.

We start with the established facts that: 1) when you regress EEG (or MEG or ECoG) against (say) the amplitude envelope of speech, you obtain a TRF that is limited in its temporal duration (to around 50-250 ms; Fig 1); and 2) such TRFs can be used to predict neural responses to novel stimuli (Crosse et al., 2016).

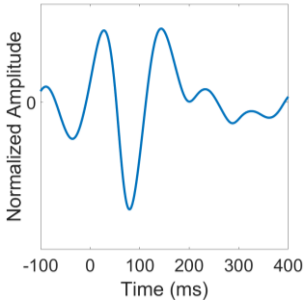

Figure 1: A temporal response function derived by regressing EEG (frontocentral channel Fz) against the amplitude envelope of an audiobook speech stimulus (~80 mins) for a single participant.

Explaining the existence of TRFs and their predictive ability under the assumption that neural responses to speech represent the concatenation of evoked responses to modulations of the stimulus is trivial. Here, we ask the novel question: can the existence of such TRFs be explained as deriving from the entrainment of an ongoing oscillation? Moreover, if so, can such a TRF succeed in modeling neural responses to novel stimuli?

Methods

We wish to assess whether the entrainment of an ongoing neural oscillation to speech can masquerade as an “evoked” TRF model. To do this, we simulate EEG responses to speech based on two models of oscillatory entrainment and then analyze that simulated data using linear regression.

Oscillatory Entrainment Models

Oscillatory Model 1 is very simple and, admittedly, physiologically implausible. It consists of a pure sinusoid whose phase is reset to 0 radians by an acoustic “edge”. This model is based on the idea that salient points (‘edges’) in speech cause “phase resetting” of ongoing low frequency oscillations, which aligns phases of high cortical excitability to features of continuous speech, thus parsing that speech into discrete linguistic units for further processing (Giraud & Poeppel, 2012). The specific acoustic edges used here are peaks in the derivative of the amplitude envelope of the speech stimuli (Oganian & Chang, 2019).

Oscillatory Model 2 aims to relate the dynamics of an ongoing oscillations to excitatory and inhibitory neural populations as described in the Wilson & Cowan model (Doelling et al., 2019; Wilson & Cowan, 1973).

$$\tau\frac{dE}{dt} = -E + S\big(\rho_E + cE - aI + \kappa A(t)\big)$$

$$\tau\frac{dI}{dt} = -I + S(\rho_I + bE - dI)$$

Where $S(z) = \frac{1}{1+e^{-z}}$ is a sigmoid function whose argument represents the activity of each neural population, E and I represent the activity of excitatory and inhibitory populations. The values of synaptic coefficients a and b, feedback connection parameters c and d were set to: a = b = c = 10 and d = -2. ρ is a constant reflecting input from other brain regions and was set to 2.3 for excitatory inputs, ρ_E, and -3.2 for inhibitory inputs, ρ_I. These parameter values were chosen based on the literature to be consistent with Hopf–Andronov bifurcation and the onset of spontaneous periodic activity. $A(t)$ is the acoustic input and, importantly, κ is the strength of coupling between that input and the neural oscillator. τ represents the membrane time constant, which influences the frequency of the oscillator. We chose a value of τ = 25 to obtain a spontaneous oscillation of 4 Hz – which is around the peak of the modulation spectrum for speech in different languages (Ding et al., 2017). Notably, oscillatory model 2 has been reported to outperform models of evoked responses in the context of rhythmic music (Doelling et al., 2019).

Simulating Neural Responses to Speech

We used each of the two oscillatory models to simulate how ongoing oscillations might entrain to a speech stimulus. In particular, we used ~3-minute-long segments of an audiobook read by an American male speaker. For Oscillatory Model 1, we identified peaks in the derivative of the amplitude envelope of the speech (following (Oganian & Chang, 2019)), and reset the phase of a 4 Hz sinusoid to zero at these "Peak Rate" timepoints with an 80 ms delay to approximate cochlea-to-cortex transmission (Fig 2, top left). For Oscillatory Model 2, we drove the Wilson & Cowan model with the amplitude envelope of the speech segments (i.e., we set $A(t)$ to be the envelope of the speech). Again, we incorporated an 80 ms delay and we used several different values of the coupling parameter κ: 2, 20, 200, 2000, 20000 (Fig 2, top right with κ = 200).

Having simulated the neural data, we then attempted to derive a TRF using linear (ridge) regression (Crosse et al., 2016). Oscillatory Model 1 produced a TRF (Fig 2, bottom left) with a timecourse that is comparable to that for TRFs derived from real EEG (Fig 1; (Di Liberto et al., 2015)). Moreover, when a train of impulses at the Peak Rate timepoints of a new speech segment was convolved with this TRF to predict the neural response, that response was correlated with the responses simulated via phase-reset (r = 0.646, p << 0.01).

Meanwhile, the TRF derived using the data from Oscillatory Model 2 (Fig 2, bottom right) displayed temporal/frequency characteristics that were not similar to those seen in TRFs derived using real EEG responses (Fig 1). However, it was able to predict simulated EEG responses to novel stimuli, although these prediction accuracies varied with the value of the coupling parameter (r = 0.0644; 0.3685; 0.5356; 0.5793; 0.5213; 0.5285 for κ = 2, 20, 200, 2000, 20000, respectively, all p < 0.01).

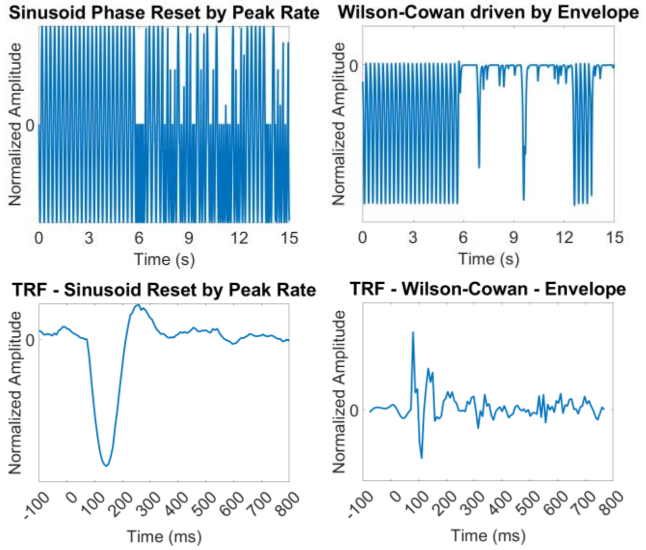

Figure 2: Simulated neural responses to speech (that begins slightly before 6 s) using Oscillatory Model 1 (top left); Oscillatory Model 2 (top right); TRFs derived from the simulated neural responses using linear (ridge) regression.

Discussion

In this preliminary study, we have shown that the established existence of TRFs may be, in principle, compatible with their arising from the entrainment of ongoing oscillations by speech stimuli. That said, the simulated neural activity in Fig 2 (top row) is decidedly unlike real neural activity, with the Oscillatory Model 1 being highly biologically implausible, and Oscillatory Model 2 displaying very unnatural temporal dynamics. Further work is required to determine whether entrainment models (with appropriate parameters) can produce both realistic EEG responses to speech stimuli *and* TRFs with realistic characteristics and significant predictive power. Of course, it may also be true that entrained oscillations occur, but that they do not contribute the TRFs seen in the literature. In that case, one might expect that such entrained oscillations would explain additional variance in EEG responses to speech beyond that explained by TRFs. Again, future work will explore this using a broad range of speech stimuli.

Acknowledgments

This study was supported by a grant from the National Institute on Deafness and Other Communication Disorders R01 DC021140. Thanks to Samuel Norman-Haignere and Nathaniel Zuk for useful discussions.

References

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences*, *98*(23), 13367–13372.

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*.

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*.

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.

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews*.

Doelling, K. B., Assaneo, M. F., Bevilacqua, D., Pesaran, B., & Poeppel, D. (2019). An oscillator model better predicts cortical entrainment to music. *Proceedings of the National Academy of Sciences*, *116*(20), 10113-10121.

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517.

Obleser, J., & Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. *Trends in Cognitive Sciences*, *23*(11), 913-926.

Oganian, Y., & Chang, E. F. (2019). A speech envelope landmark for syllable encoding in human superior temporal gyrus. *Science Advances*, *5*(11), eaay6279.

Wilson, H. R., & Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik*, *13*(2), 55-80.