Correlates of Syntax in EEG Responses to Naturalistic Speech

Ole Bialas (obialas@ur.rochester.edu)
Department of Biomedical Engineering, 204 Robert B. Goergen Hall
Rochester, NY 14627 United States of America

Sophea Urbi Biswas (sbiswas8@u.rochester.edu)
Department of Biomedical Engineering, 204 Robert B. Goergen Hall
Rochester, NY 14627 United States of America

Edmund C. Lalor (elalor@ur.rochester.edu)
Department of Biomedical Engineering, 204 Robert B. Goergen Hall
Rochester, NY 14627 United States of America

Abstract

Research on the cortical processing of continuous naturalistic speech has shown that the brain represents speech on different levels of abstraction. Speech is represented as a continuous acoustic signal, in discrete phoneme categories or as semantic information on the level of words. However, it remains unclear if and how the brain represents the syntax to combine words into larger meaningful structures. Here, we use multivariable models to predict the EEG recordings of participants listening to naturalistic speech. We show that syntactic features, derived from automated dependency parsing, improve the prediction accuracy of a baseline model using acoustic and semantic speech features. While our findings suggests that correlates of syntax can be found in EEG responses to continuous speech, they also highlight the need for further research to disentangle syntax from possible correlated acoustic speech cues.

Keywords: EEG; naturalistic speech; syntax; temporal response function, auditory

Introduction

Over the last decade, numerous studies provided new insights into how the brain processes speech by presenting continuous speech in its natural complexity and using computational models to dissect different components of the recorded neural signal (Hamilton & Huth, 2020). This revealed patterns of activity compatible with many different representations of speech signals such as their acoustic properties, the phonemes they are composed of, and their meaning (Aiken & Picton, 2008; Di Liberto, O'Sullivan, & Lalor, 2015; Broderick, Anderson, Di Liberto, Crosse, & Lalor, 2018). Despite this success, it is unclear if and how the brain represents the syntax to combine words into larger units like phrases and sentences. One seminal study presented four-word sentences consisting of two two-word phrases and found that the recorded brain response's spectrum resembled sentence and phrase frequency, independent of acoustic content (Ding, Melloni, Zhang, Tian, & Poeppel, 2016). While this study showed neural correlates of syntax, it used synthesized speech presented at a steady and invariable frequency, making it unclear whether this generalizes to naturalistic scenarios. A recent study found syntax features that correlate with activity in specific frequency bands of the MEG signal, recorded while participants listened to spoken stories (Zioga, Weissbart, Lewis, Haegens, & Martin, 2023). Here, we are using the same approach but apply it to EEG, which has a lower signal-to-noise ration but can be used in a wider range of contexts. Instead of focusing on specific frequency bands, we model the full broad-band EEG signal.

Methods

EEG Data

We analyzed EEG data that were acquired in previous studies (Di Liberto et al., 2015; Broderick et al., 2018) and contained neural recordings of 19 participants (age 19-38, 13 male) listening to 20 segments of an audiobook, each lasting about 3 minutes [1]. Data were recorded using a 128-channel EEG system (Biosemi) at a 512 Hz sampling rate. Offline, the recordings were resampled to 128 Hz, band-pass filtered between 1 and 20 Hz and re-referenced to the global average. Each channel was standardized by subtracting its mean and dividing by its standard deviation.

Features

We computed the speech signal's envelope as the absolute Hilbert transform, band-pass filtered between 1 and 20 Hz. We also computed the half-wave rectified derivative of the envelope to represent onsets in the acoustic energy of the signal. We estimated the semantic dissimilarity of each word given its context using a word2vec model (Broderick et al., 2018) and represented this feature as impulses at the onset of each word. We included a similar feature, where each impulse had the value one, to index the onset of words. Finally, we derived syntactic features using a dependency parsing algorithm that generates a description of relationships between words in a sentence. We counted the number of relationships that are opened, remain open and are closed, as each word is uttered and represented them as vectors similar to the semantic features (Zioga et al., 2023).

[1] The data are available on OpenNeuro.org (data set ds004408).

Modeling

We predicted EEG recordings at each channel using forward temporal response functions or TRFs (Crosse, Di Liberto, Bednar, & Lalor, 2016) and used Pearson's correlation between the predicted and recorded EEG as an estimate of model accuracy. We compared a baseline model using envelope, acoustic onsets, word onsets and semantic dissimilarity, to a model that also included syntactic features. Each model was optimized and evaluated using nested cross-validation loops. In the outer loop one segment was selected for testing and, in the inner loop, the remaining segments were used to optimize the regularization coefficient λ by leave-one-out cross-validation. We averaged across all test segments to obtain an unbiased estimate of model accuracy. For visualization, we computed the models with the same value of $\lambda = 10000$ across all participants and averaged the resulting syntactic TRFs.

Results

The syntactic TRF consistently predicted brain responses more accurately compared to the acoustic and semantic baseline model (one-tailed Wilcoxon signed-rank test for paired data, W=157, p=0.005). Inspecting the distribution of differences in accuracy across the scalp (Fig.1) revealed that the improvements were constrained to central and temporal regions, suggesting that the syntactic features index activity in auditory regions.

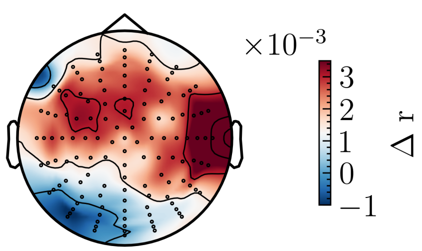

Figure 1: Difference in prediction accuracy between syntax and baseline model at each channel.

Because forward TRF weights can be interpreted as expected changes in the response following changes in the predictor (Haufe et al., 2014), visualizing their time course and distribution can be indicative of underlying neural processes. Thus, we averaged the TRFs across all subjects to obtain time courses for the open, remain and resolve feature. All TRFs showed strong oscillatory components consistent with the finding that these features are track activity in the α and β band (Zioga et al., 2023). To see if the TRFs contained slower components resembling evoked response potentials (ERPs), we low-pass filtered them at 5 Hz. Figure 2 shows the TRFs at a central electrode for the open, remain and resolve feature. The diverging distributions at a time lag of 260 ms shows that the features are associated with different parts of the neural response. However, there is substantial overlap and the positive peaks in the open and remain TRF at time lags around 100 ms suggest that the models also index activity related to low-level acoustic processing.

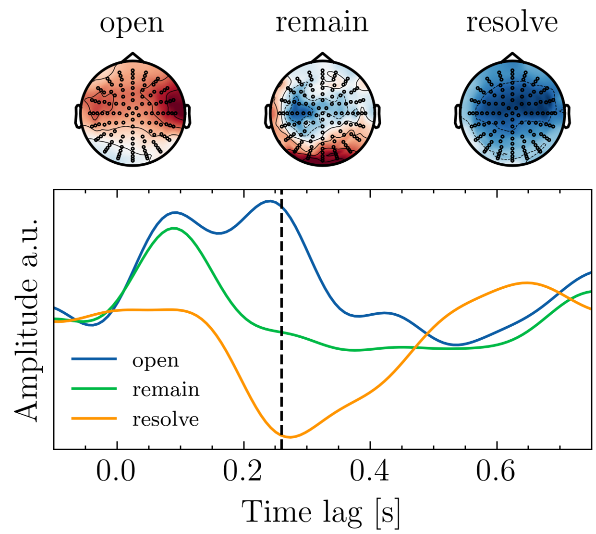

Figure 2: Group average TRFs for the syntactic features at a central electrode. Dashed line marks the time point where the topographical distribution of each model is shown.

Discussion

We showed that incorporating syntactic features improves the TRF's ability to predict EEG responses to naturalistic speech. This suggests that correlates of syntax, previously found in MEG data (Ding et al., 2016; Zioga et al., 2023) are also present in EEG. In contrast to previous work that tracked changes in α and β power (Zioga et al., 2023) we predicted the full broadband EEG signal and found that the syntactic features are also associated with slower, ERP-like, components. The TRFs for the open and remain feature showed early peaks indicating activity related to low-level acoustic processing. While this could be due to leakage between model features it may also arise from acoustic features that are inherently correlated with syntax. It is well known that prosody, the patterns of stress and intonation, contains cues about the syntactic structures of speech. For example, speakers tend to raise their pitch at the end of a question and such changes in pitch are known to affect EEG responses to naturalistic speech (Teoh, Cappelloni, & Lalor, 2019). Prosodic representation may even arise covertly, in absence of any physical cues, making it hard to distinguish them from genuine syntax processing (Glushko, Poeppel, & Steinhauer, 2022). It may even be that there is no representation of syntax that is entirely separate from the rest of language. While linguists have long postulated that an infinitely generative system like human language requires a distinct set of discrete syntactic rules, the recent success of large language models, which represent syntax and semantics in a joint embedding space, calls this dogma into question (Piantadosi, 2023). Future studies should try to disentangle syntactic and prosodic representations and test different feature spaces - after all, counting the number of relationships is a rather crude way of operationalizing syntax.

References

Aiken, S. J., & Picton, T. W. (2008). Human cortical responses to the speech envelope. *Ear and hearing*, *29*.

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*.

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*.

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).

Ding, Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature neuroscience*, *19*.

Glushko, A., Poeppel, D., & Steinhauer, K. (2022). Overt and implicit prosody contribute to neurophysiological responses previously attributed to grammatical processing. *Scientific Reports*, *12*.

Hamilton, L. S., & Huth, A. G. (2020). The revolution will not be controlled: natural stimuli in speech neuroscience. *Language, cognition and neuroscience*, *35*.

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage*, *87*.

Piantadosi, S. (2023). Modern language models refute chomsky's approach to language. *Lingbuzz Preprint*.

Teoh, E. S., Cappelloni, M. S., & Lalor, E. C. (2019). Prosodic pitch processing is represented in delta-band eeg and is dissociable from the cortical tracking of other acoustic and phonetic features. *European Journal of Neuroscience*, *50*.

Zioga, I., Weissbart, H., Lewis, A. G., Haegens, S., & Martin, A. E. (2023). Naturalistic spoken language comprehension is supported by alpha and beta oscillations. *Journal of Neuroscience*, *43*.