Brain-to-Brain Linguistic Coupling in Natural Conversations

Zaid Zada (zaid@princeton.edu)
Department of Psychology, Washington Road
Princeton, NJ 08544 USA

Samuel Nastase (snastase@princeton.edu)
Princeton Neuroscience Institute, Washington Road
Princeton, NJ 08544 USA

Ariel Goldstein (ariel.y.goldstein@gmail.com)
Princeton Neuroscience Institute, Washington Road
Princeton, NJ 08544 USA

Uri Hasson (hasson@princeton.edu)
Princeton Neuroscience Institute, Washington Road
Princeton, NJ 08544 USA

Abstract

How does language mediate brain-to-brain coupling during face-to-face communication between speakers and listeners? Here, we explore whether the embedding space learned by deep language models can serve as a common linguistic intermediary for aligning brains during communication. We recorded real-world, face-to-face conversations in five dyadic pairs of electrocorticography (ECoG) patients. This unique setup allows us to model how a speaker encodes and transmits their thoughts to a listener during free-form conversations. Our findings reveal the temporal profile of information flow across brains during conversation: linguistic content emerges in the speaker's brain before word onset and is recapitulated in the listener's brain rapidly following word onset. This transmission process relies on a shared linguistic embedding space for translating internal states from one brain to another.

Keywords: brain-to-brain coupling; hyperscanning; language models; ECoG; speech

Introduction

Language allows us to describe our experience of the world and share our thoughts with others. Communication, however, relies on a shared agreement as to the meaning of words in context. This shared agreement varies across cultures and contexts; for example, in the context of chess, the word *game* is associated with a particular set of rules; in the context of a playground sandbox, the word *game* is associated with open-ended creative play. The way we use words is grounded in a shared code that all speakers in a community participate in.

Despite the importance of this agreement on shared contextual meaning, most studies of the neural basis of language processing have been constrained to studying single speakers in isolation (Pickering & Garrod, 2004; Hasson & Honey, 2012). The vast majority of these studies cannot speak to the spontaneous, contextual, and communicative nature of real-world dialogue. These limitations have led the community to push for the use of naturalistic language stimuli (Hasson, Ghazanfar, Galantucci, Garrod, & Keysers, 2012; Hamilton & Huth, 2020) and interactive "brain-to-brain" paradigms (Redcay & Schilbach, 2019).

The lack of shared, context-sensitive language models has hindered attempts to formally model the neural coupling between interlocutors. To overcome this limitation, intersubject correlation (ISC) analyses were developed to model the neural activity of one brain (e.g., a listener) based on the neural activity in another brain (e.g., a speaker) (Stephens, Silbert, & Hasson, 2010; Bevilacqua et al., 2019). ISC analyses bypass the shared meaning between interlocutors and cannot capture the linguistic content and context that drives brain-to-brain coupling in particular conversations.

In this paper, we tested whether we can use the embedding space derived from a recently-developed deep language model (DLM) as a shared linguistic intermediary for brain-to-brain coupling during natural conversations. DLMs learn the statistical structure of language from the way humans use language in real-world contexts. These models encode words in a high-dimensional embedding space that captures the shared, context-sensitive structure of language (Linzen & Baroni, 2021) and share important computational principles with human language processing (Schrimpf et al., 2021; Goldstein et al., 2022).

Results

We collected ECoG data in five dyads during real-time, free-form conversations. For each dyad, we spliced the neural data into word-level epochs, collated these epochs according to speaker and listener roles, and split the collated data for 10-fold cross-validation. We used time-resolved transcriptions of each conversation to extract embeddings for each word from the large language model GPT-2 (Radford et al., 2019). We estimated encoding models for the speaker and listener to predict the neural activity for each word using the embeddings from GPT-2 (Fig. 1A).

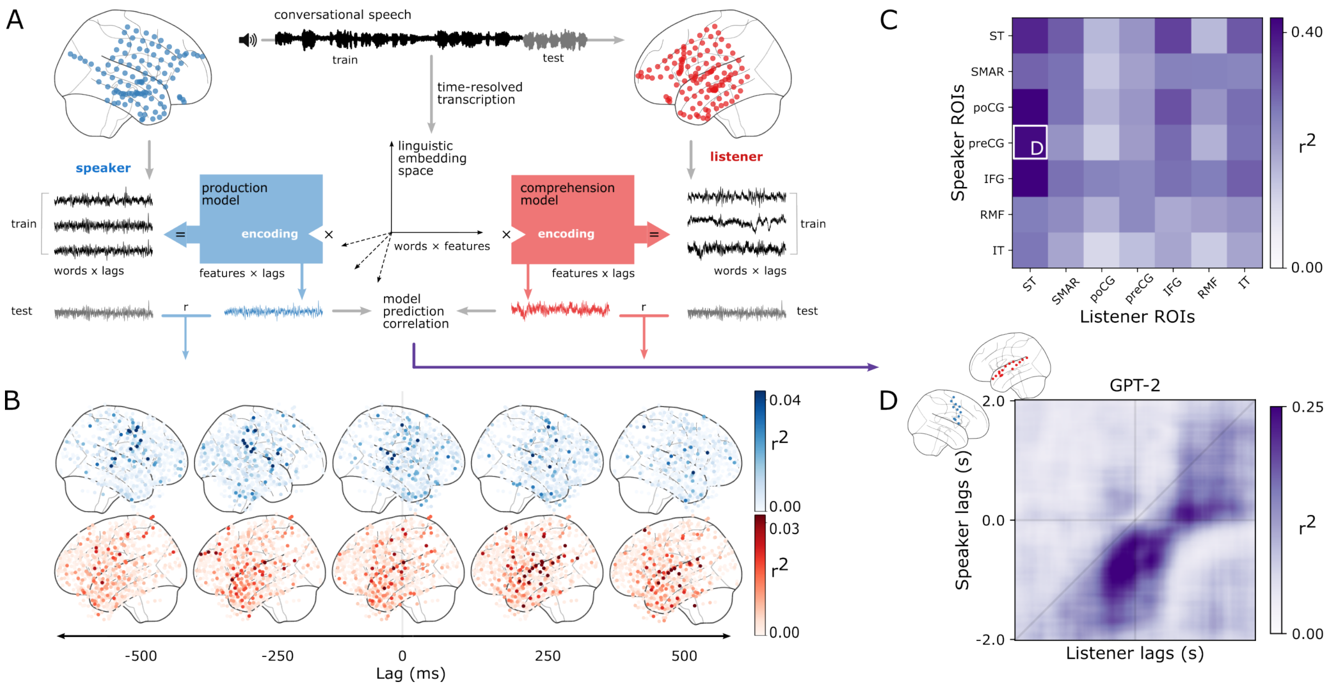

Figure 1: (A) Modeling speaker-listener linguistic coupling using a shared language model. (B) Speaker (blue) and listener (red) encoding performance at five lags relative to word onset for all electrodes and subjects. (C) Linguistic coupling between speaker and listener across multiple regions of interest. (D) Time-resolved linguistic coupling between somatomotor electrodes in the speaker and superior temporal electrodes in the listener.

Word Embeddings Predict Brain Activity in Both Speaker and Listener We first assessed whether the linguistic embedding space can capture time-resolved, word-related neural activity in both speaker and listener. We trained encoding models with 10-fold consecutive cross-validation and measured the correlation between actual and predicted word-related activity in each test fold at lags ranging from -2000 ms to +2000 ms relative to word onset (Fig. 1A). For the speaker, we found that encoding performance peaks before word onset in somatomotor and inferior frontal electrodes and quickly decreases after 500 ms post-word-onset (Fig. 1B, blue). For the listener, encoding performance increases gradually at word onset and peaks at roughly 250 ms post-word-onset in superior and anterior temporal electrodes (Fig. 1B, red). These results demonstrate that the linguistic embedding space learned by GPT-2 captures relevant features for predicting neural activity during both language production and comprehension.

Brain-to-Brain Linguistic Coupling How are the speaker and listener's brains aligned during the conversation? The previous analysis used the encoding models to predict the neural signal from word embeddings. This "encoded" signal—i.e. the model-based predictions of neural activity—captures linguistic features of the neural signal within each brain. To assess linguistic coupling across brains, we directly correlated the predictions of the speaker's encoding model to the listener's encoding model at varying regions (Fig. 1A, purple).

We found widespread, asymmetric inter-regional speaker-listener coupling between the speaker's articulatory system, including the inferior frontal gyrus (IFG), precentral gyrus (preCG), and postcentral gyrus (poCG), and the listener's superior temporal (ST) language areas, as well as higher-level supramarginal and inferior frontal areas (Fig. 1C).

We further investigated the temporal profile of the coupling between the speaker's articulatory regions in somatomotor cortex and the listener's superior temporal language areas. We repeated the same encoding analysis and correlated the predictions of each encoding model for every pair of lags to obtain a time-resolved, lag-by-lag correlation matrix. We found that the linguistic features of the speaker's pre-word-onset activity in somatomotor cortex best modeled the linguistic features of the post-word-onset responses in the listener's superior temporal cortex (Fig. 1D).

To ensure that the model-based speaker-listener coupling was driven by the linguistic structure of the DLM embedding space, we performed a control analysis using "arbitrary" embeddings. We generated random embeddings for each word and reran the lag-by-lag inter-subject encoding analysis. The random embeddings only capture the occurrence of individual words and do not contain the context-sensitive linguistic structure of the DLM embeddings. These control embeddings explained only a small proportion of variance in comparison to the linguistic embeddings. We also found that shuffling model predictions between dyads attenuates coupling; i.e. the coupling of each dyadic conversation is unique.

Discussion

In the current work, we use a DLM embedding space to isolate shared linguistic features linking the brain activity between speakers and listeners in real-world conversations. This is one of the first attempts to model context-dependent, word-level neural activity during free, spontaneous conversations. While previous work has described speaker-listener neural coupling during storytelling (Silbert, Honey, Simony, Poeppel, & Hasson, 2014), the current work leverages a shared language model to capture linguistic coupling between brains. Our results reveal time-resolved linguistic coupling between speaker and listener: shared word- and conversation-specific linguistic features emerge in the speaker's language-production areas before word articulation, and later, post articulation, reemerge in the listener's comprehension areas.

The current conversational "hyperscanning" dataset allows us to model brain activity simultaneously in both the speaker and listener. Our comprehension result (Fig. 1B, red) replicates previous work examining the temporal dynamics of linguistic encoding using DLM embeddings in a naturalistic spoken narrative (Goldstein et al., 2022). Our results expand on this work in two ways. First, we demonstrate that DLM embeddings also capture linguistic features of neural activity in the speaker—prior to word onset in articulatory cortical areas (Fig. 1B, blue). Second, the dyadic, conversational nature of our dataset allows us to map from the linguistic features of neural activity in the speaker to linguistic features of neural activity in the listener on a time-resolved word-by-word basis.

Acronyms: Superior temporal (ST); Supramarginal (SMAR); Postcentral gyrus (poCG); Precentral gyrus (preCG); Inferior frontal gyrus (IFG); Rostral middle frontal (RMF); Inferior temporal (IT).

Acknowledgments

This work was supported by the National Institutes of Health under award numbers DP1HD091948 (Z.Z., A.G., and U.H.), and R01MH112566 (S.A.N.).

References

Bevilacqua, D., Davidesco, I., Wan, L., Chaloner, K., Rowland, J., Ding, M., ... Dikker, S. (2019, March). Brain-to-brain synchrony and learning outcomes vary by student–teacher dynamics: evidence from a real-world classroom electroencephalography study. *Journal of Cognitive Neuroscience*, *31*(3), 401–411. doi: 10.1162/jocn_a_01274

Goldstein, A., Zada, Z., Buchnik, E., Schain, M., Price, A., Aubrey, B., ... Hasson, U. (2022, March). Shared computational principles for language processing in humans and deep language models. *Nature Neuroscience*, *25*(3), 369–380. doi: 10.1038/s41593-022-01026-4

Hamilton, L. S., & Huth, A. G. (2020, June). The revolution will not be controlled: natural stimuli in speech neuroscience. *Language, Cognition and Neuroscience*, *35*(5), 573–582. doi: 10.1080/23273798.2018.1499946

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012, February). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends in Cognitive Sciences*, *16*(2), 114–121. doi: 10.1016/j.tics.2011.12.007

Hasson, U., & Honey, C. J. (2012, August). Future trends in neuroimaging: neural processes as expressed within real-life contexts. *NeuroImage*, *62*(2), 1272–1278. doi: 10.1016/j.neuroimage.2012.02.004

Linzen, T., & Baroni, M. (2021, January). Syntactic structure from deep learning. *Annual Review of Linguistics*, *7*(1), 195–212. doi: 10.1146/annurev-linguistics-032020-051035

Pickering, M. J., & Garrod, S. (2004, April). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, *27*(02). doi: 10.1017/S0140525X04000056

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners.

Redcay, E., & Schilbach, L. (2019, August). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience*, *20*(8), 495–505. doi: 10.1038/s41583-019-0179-4

Schrimpf, M., Blank, I. A., Tuckute, G., Kauf, C., Hosseini, E. A., Kanwisher, N., ... Fedorenko, E. (2021, November). The neural architecture of language: integrative modeling converges on predictive processing. *Proceedings of the National Academy of Sciences*, *118*(45), e2105646118. doi: 10.1073/pnas.2105646118

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014, October). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences*, *111*(43). doi: 10.1073/pnas.1323812111

Stephens, G. J., Silbert, L. J., & Hasson, U. (2010, August). Speaker–listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, *107*(32), 14425–14430. doi: 10.1073/pnas.1008662107