Linguistic coupling between neural systems for speech production and comprehension during real-time dyadic conversations

Zaid Zada*[1], Samuel A. Nastase[1], Sebastian Speer[1], Laetitia Mwilambwe-Tshilobo[1], Lily Tsoi[1,2], Shannon Burns[1,3], Emily Falk[4], Uri Hasson[1], Diana Tamir[1]

[1] Neuroscience Institute and Psychology Department, Princeton University, Princeton NJ
[2] Department of Psychology, Caldwell University, Caldwell NJ
[3] Psychological Science and Neuroscience, Pomona College, Claremont CA
[4] Department of Psychology, University of Pennsylvania, Philadelphia PA

*Corresponding author. Email: zzada@princeton.edu

Keywords: brain-to-brain coupling; speech production; speech comprehension; communication; conversations; encoding models; naturalistic paradigm; large language models (LLMs); hyperscanning.

Abstract

The core use of human language is communicating complex ideas from one mind to another in everyday conversations. In conversations, comprehension and production processes are intertwined, as speakers soon become listeners, and listeners become speakers. Nonetheless, the neural systems underlying these faculties are typically studied in isolation using paradigms that cannot fully engage our capacity for interactive communication. Here, we used an fMRI hyperscanning paradigm to measure neural activity simultaneously in pairs of subjects engaged in real-time, interactive conversations. We used contextual word embeddings from a large language model to quantify the linguistic coupling between production and comprehension systems within and across individual brains. We found a highly overlapping network of regions involved in both production and comprehension spanning much of the cortical language network. Our findings reveal that shared representations for both processes extend beyond the language network into areas associated with social cognition. Together, these results suggest that the specialized neural systems for speech perception and production align on a common set of linguistic features encoded in a broad cortical network for language and communication.

Introduction

Conversational language is a fundamentally social process: a tool for communication (Fedorenko, Piantadosi, et al., 2024; Hasson et al., 2012) involving at least two people. It allows us to transmit one's internal states—encoding personal experiences and complex thoughts—to another person's mind. Social interaction, especially conversation, requires coordination in the form of interactive alignment (Pickering & Garrod, 2004), and agreement on meaning through common ground (Brennan & Clark, 1996; Clark & Brennan, 1991; Wilkes-Gibbs & Clark, 1992). Conversation is arguably the most *fundamental* setting of language use. It is universal to human societies, does not require specialized skills (e.g., literacy) or technologies (e.g., telephones) (Clark, 1996), and allows people to go well beyond simple stimulus-response signaling to share and shape each others' representational thought through language. Here we interrogate the mechanism that underlies the transmission of these signals from one brain to the next. To do so, we capture real-time, conversational comprehension and production using fMRI hyperscanning and quantify *what* is shared using large language models (LLMs).

Participants in a dyadic conversation alternate between speaking (production) and listening (comprehension) roles, responding to each other's utterances and redirecting the conversation in new directions. Classic neurobiological language models are modular and descriptive, aiming to label particular brain areas with specific cognitive functions, such as Broca's area for speech production and Wernicke's area for speech comprehension. This has led the field to treat language production and comprehension as separate and unrelated research endeavors (Friederici, 2011; Hickok & Poeppel, 2007; Price, 2010). More recently, researchers have promoted a more integrated view of production and comprehension, prompting further definition of their similarity and interaction (Liberman & Whalen, 2000; Menenti et al., 2012; Pickering & Garrod, 2013; Pulvermüller, 2018; Pulvermüller & Fadiga, 2016).

Naturalistic experimental paradigms are ideally suited to shed further light on the neural processes and representations used to comprehend and produce natural language (Hasson et al., 2012). One such paradigm uses a sequential, asynchronous protocol: first recording a subject speaking and then playing the speech back to multiple listeners at a later time (Chang et al., 2023; Jiang et al., 2012; Kinreich et al., 2017; Liu et al., 2022; Nguyen et al., 2022; Silbert et al., 2014; Stephens et al., 2010; Zadbood et al., 2017). These studies find that during communication, the speaker's neural activity is coupled to that of the listeners in regions associated with both comprehension and production. Moreover, the strength of speaker-listener coupling is related to outcome measures of comprehension. However, these paradigms often use separate stimuli for production or comprehension, use controlled paradigms (e.g., rehearsed speech, covert production), and isolated linguistic contexts. This raises the question of whether paradigms for asynchronous speech and passive comprehension fully capture the neural systems for real-time, interactive conversation. It remains unclear how the brain's

production and comprehension neural systems are related during real-time conversations–both within and between subjects.

In real-time dyadic conversations, these processes are contemporaneous and interleaved. Production is often spontaneous rather than read or rehearsed, and comprehension must be *proactive* as listeners must be ready to respond in a relevant way as they process the incoming speech (Grice, 1975; Redcay & Schilbach, 2019). Hyperscanning paradigms, where researchers simultaneously measure neural processes during dyadic social interactions using two MRI scanners, are uniquely suited to studying two interacting participants as they engage in free conversations (Babiloni & Astolfi, 2014; Czeszumski et al., 2020; Montague et al., 2002; Nam et al., 2020; Redcay & Schilbach, 2019; Speer et al., 2024; Tsoi et al., 2022; Wheatley et al., 2019). Paradigms of this kind can extend asynchronous work on coupling neural systems for production and comprehension to language use in conversations where production is spontaneous and comprehension is intertwined with production. It also allows the investigation and comparison of both processes within the same subject. We hypothesize that despite their differences, speech production and comprehension rely on linguistic representations that are shared within subjects and contemporaneously coupled across subjects.

To quantify linguistic representations shared between communicators, researchers need to go beyond the metrics of brain-to-brain coupling and synchrony that dominate previous research. Data-driven metrics such as intersubject correlation (Hasson et al., 2004; Nastase et al., 2019), phase locking value (e.g., Tognoli et al., 2007), and Granger causality (e.g., Schippers et al., 2010) use one subject's neural activity to model another's in the absence of an explicit model of linguistic processing. They have yielded insights into the "where" and "how much" of brain-to-brain coupling during naturalistic narrative comprehension and language processing (Dikker et al., 2014; Liu et al., 2022; Silbert et al., 2014; Stephens et al., 2010); memory recall (Chen et al., 2017; Zadbood et al., 2017, 2022); and teacher-student coupling and student outcomes (Bevilacqua et al., 2019; Davidesco et al., 2023; Dikker et al., 2017; Meshulam et al., 2021; Nguyen et al., 2022). However, these coupling metrics are fundamentally content-agnostic—they cannot tell us "what" is shared between brains (Zada et al., 2024).

Here we leverage a new framework for *model-based coupling* that allows us to identify *what* features are shared between brains and test different models of the features along which the speaker and listener's brains align. Any number of features of a conversation (e.g., acoustic features of speech like intonation and prosody) can induce synchrony or correlation across subjects. Encoding models, in particular, can quantify the linguistic features (e.g., using word embeddings) encoded in neural activity during passive language comprehension (de Heer et al., 2017; Huth et al., 2016; Wehbe et al., 2014). By leveraging the rich linguistic representations from large language models (LLMs), encoding models have yielded insights into neural systems for the comprehension of

natural language: for example, that representations for a single word depend on context (Goldstein et al., 2022) and that language processing involves prediction (Caucheteux et al., 2023; Goldstein et al., 2022; Heilbron et al., 2022; Schrimpf et al., 2021). Notably, these methods have begun to lend further support for the integrated view of neural representations for production and comprehension both within subjects (Cai et al., 2023; Goldstein et al., 2023; Yamashita et al., 2023), and across subjects (Zada et al., 2024) during real-time dialogue. For example, Zada et al. (2024) used contextual embeddings from an LLM as an explicit, shared model mediating speaker-listener neural coupling during real-time conversations in electrocorticography. However, electrocorticography is limited to a small number of participants and only brain regions in the left temporal and frontal cortical areas.

In this paper, we developed an fMRI hyperscanning paradigm to simultaneously measure whole brain activity in dyadic pairs of subjects engaged in free-form, interactive conversations across a range of prompted topics. We used encoding models to map the cortical areas representing linguistic content during spontaneous speech production and comprehension. Within subjects, we found overlapping cortical systems for production and comprehension: both systems depend strongly on common representations and only partially on specialized representations. Then, using the real-time, dyadic conversations paradigm, we compute model-based coupling across the speaker's and listener's brains via the LLM embeddings. We found that model-based speaker-listener coupling engages areas associated with social cognition. Through model-based LLM encoding analysis of whole brain fMRI neural signals, we gain valuable insights into how successful conversations depend on shared language representations between production and comprehension across various cortical regions.

Results

We aimed to model linguistic processing within and between brains during free-form, turn-based conversations. We used hyperscanning to collect simultaneous fMRI data in 30 dyads (60 subjects) as they freely discussed ten topics across five ~6 min runs (Figure 1A) (Speer et al., 2024). Topic prompts were presented as a starting point, but each dyad was free to pursue the discussion differently, resulting in 30 unique conversations (Table S1). To characterize the linguistic content in the BOLD signal, we explicitly represented the language stimuli with several different feature spaces: confound variables (e.g., word rate), spectral acoustic features, phonemic articulatory features, and word embeddings extracted from GPT-2 (Radford et al., 2019) (Figure 1C). Then, we used banded ridge regression to estimate a linear mapping from the model features onto the BOLD activity at each vertex (Dupré La Tour et al., 2022; Huth et al., 2016; Naselaris et al., 2011; Nunez-Elizalde et al., 2019) (Figure 1D). To evaluate the models, we correlated the model-predicted and actual BOLD time series for left-out runs for each feature space and for production or comprehension time points separately

(Figure 1E). Finally, we averaged the model performance correlations across subjects for all analyses. Statistically, we evaluated the average using a one-sample t-test, correcting for multiple comparisons over all ~75k cortical vertices. To summarize our results, we averaged the encoding performance across vertices within 11 regions of interest (ROIs) spanning an extended language network, from low-level auditory and articulatory areas to high-level semantic areas: early auditory cortex (EAC), posterior and anterior superior temporal gyrus (pSTG, aSTG), inferior and middle frontal gyri (IFG, MFG), somatomotor cortex (SM), supplementary motor area (SMA), frontal opercular (FOP), intraparietal sulcus (IPS), temporoparietal junction (TPJ), and posterior medial cortex (PMC).

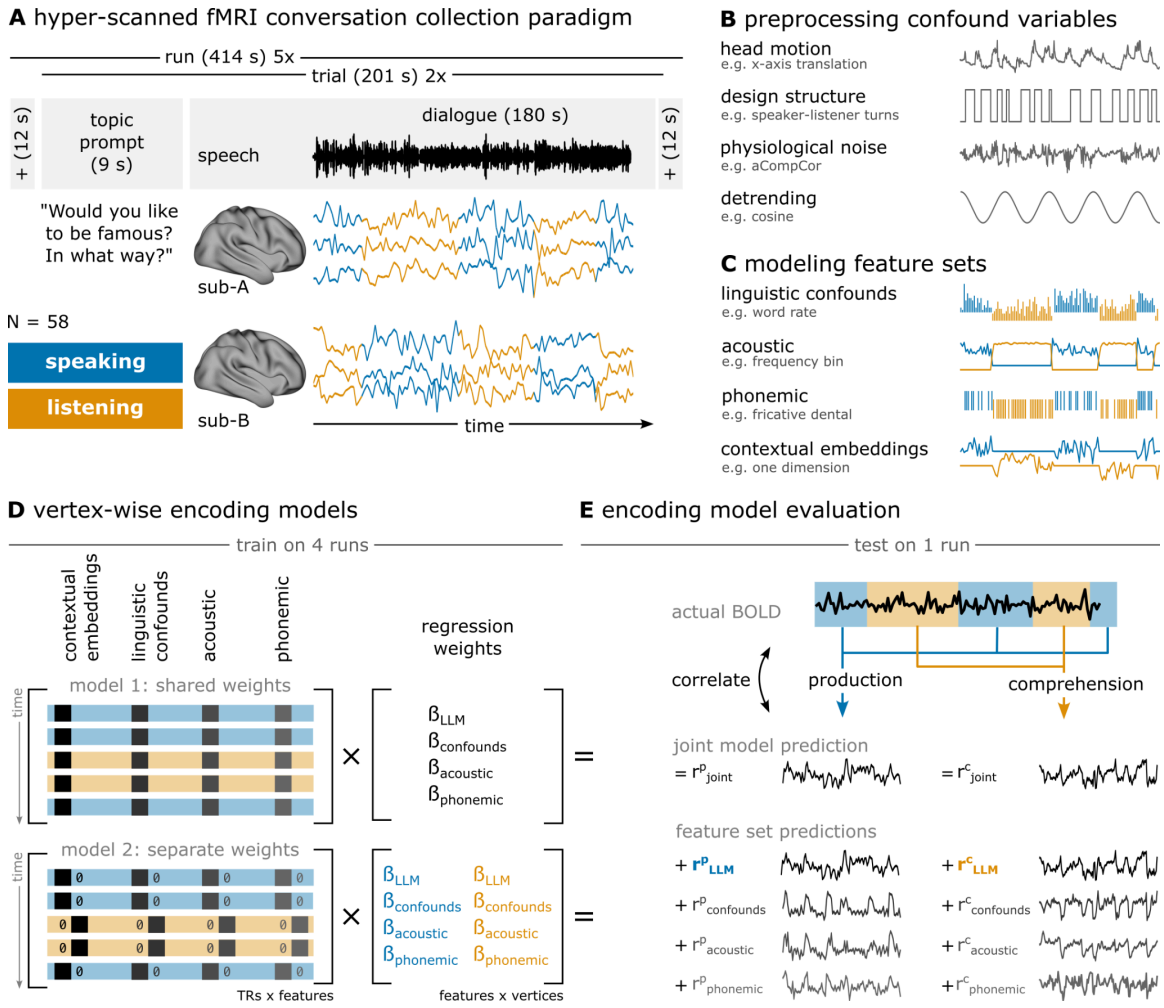

Figure 1. Data collection and modeling framework. (**A**) We collected fMRI data simultaneously from pairs of subjects as they engaged in interactive, prompted conversations. (**B**) Typical confound variables were regressed from the BOLD signal, including head motion, physiological noise, and drift. In addition, we added several structural box-car and impulse regressors to account for the turn-taking nature of the paradigm. (**C**) We quantified several feature spaces for each conversation, including nuisance regressors (e.g., word rate), acoustic features, articulatory features, and word embeddings from a large language model. We split the regressors into separate time series for production (blue) and comprehension (orange). (**D**) We fit vertex-wise

encoding models with banded ridge regression to simultaneously predict vertex-wise BOLD activity from all four feature spaces. (**E**) This allowed us to evaluate the relative performance of each feature space separately in a held-out test run of different conversations.

Contextual embeddings capture both production and comprehension.

We first validated that we can successfully model brain activity during spontaneous production and comprehension in our hyperscanning paradigm. To do so, we built two models to quantify linguistic processing and to measure the cortical overlap between production and comprehension. In one, we constrained the model to learn one set of shared weights for production and comprehension for all feature spaces. In this model, a vertex must code for both processes with the same functional tuning (i.e., shared weights) to be well predicted. In the second model, we split all regressors into separate sets for production and comprehension, allowing the model to learn separate weights for each process (Figure 1D). We treat the confound, acoustic, and phonemic feature sets as nuisance variables and report only the LLM contextual embedding performance. We first inspect the performance of the second, more flexible model, which we expect to outperform the unified constrained model.

Using the more flexible model with separate weights for production and comprehension, we found significant within-subject encoding performance throughout the core language network: STG, IFG, and MFG for speech production and speech comprehension (Figure 2A). Moreover, encoding performance extended bilaterally to higher-level regions like TPJ and PMC. We found considerable spatial overlap between encoding performance for production and comprehension—i.e., vertices well predicted during production are also likely to be well predicted during comprehension ($r = 0.712$, $p < 1e\text{-}5$). To quantify whether production and comprehension encoding rely on shared or divergent weights, we compared the performance of the shared-weights and separate-weights models. We found that across all 11 ROIs, a large proportion of encoding performance can be attributed to shared functional tuning rather than idiosyncratic production- or comprehension-specific variance (Figure 2B). Peripheral regions for speech perception (EAC) and speech articulation (SM) showed the largest divergence, but the shared-weights model still recovered over half the performance of the separate-weights model. These results suggest that cortical activity during both production and comprehension keys to similar features captured by the LLM embeddings.

We observed several qualitative differences across tasks, regions, and hemispheres. First, overall encoding performance appears higher in right STG and TPJ than in the left-hemisphere homologs. Second, overall encoding performance appears stronger for production, especially in bilateral PMC and right TPJ. Third, encoding performance for comprehension appears stronger and more bilateral in STG than in production. Despite these differences, the overall encoding performance suggests that LLM embeddings provide a rich basis for modeling linguistic encoding throughout much of the cortex.

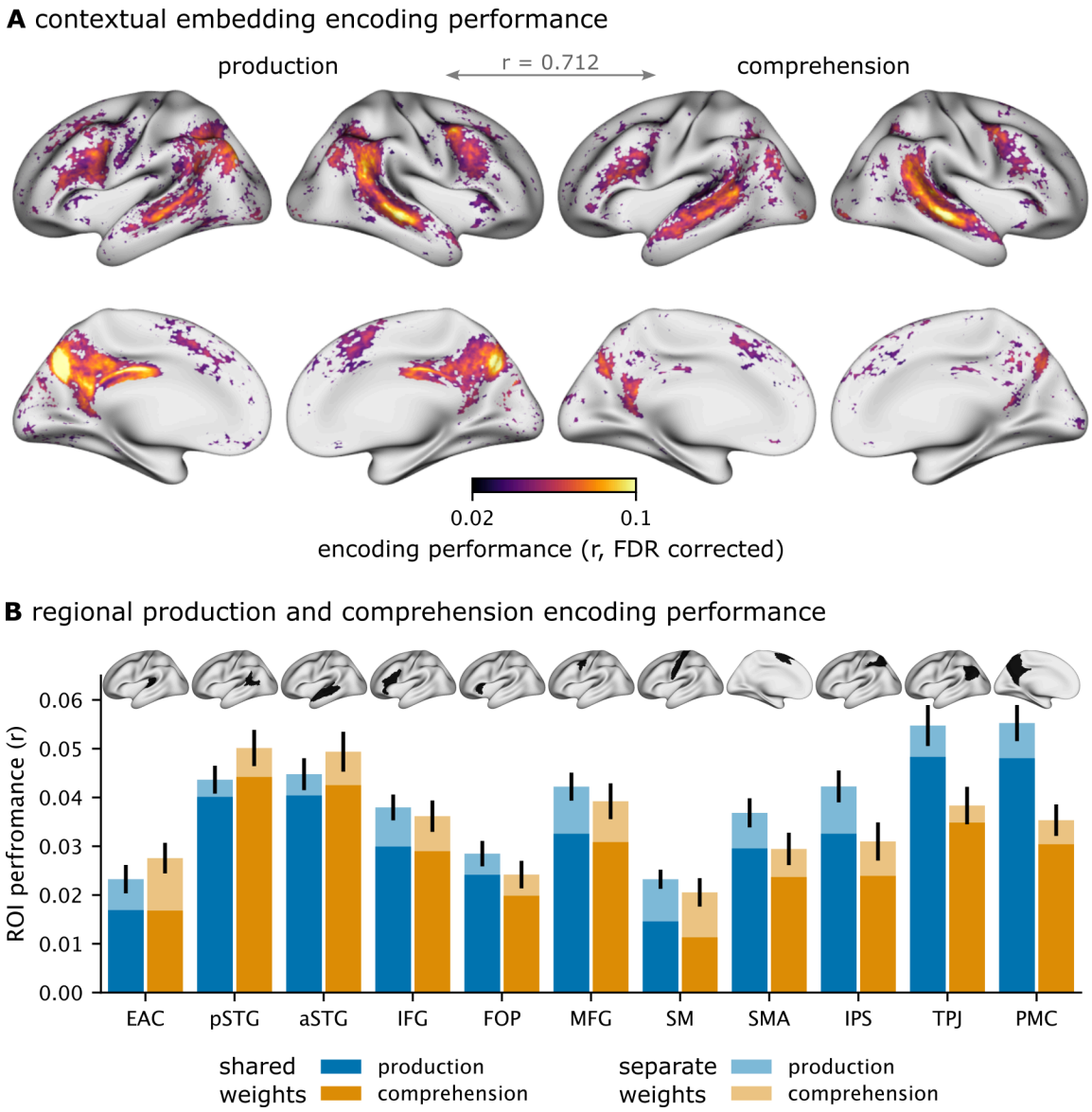

Figure 2. Within-subject speaking and listening encoding performance. (**A**) Encoding model performance of the separate-weights model for production and comprehension relative to the control feature spaces. (**B**) We summarized the un-thresholded encoding performance of the shared- and separate-weights models in 11 ROIs spanning the extended language network, averaged across left and right hemispheres (see Methods).

Story-listening comprehension shares a subset of linguistic features with interactive production and comprehension neural systems

In addition to the hyperscanning conversations paradigm, we recorded participants as they listened to a 13-minute story in a separate scanning session. This presented a unique opportunity to compare linguistic processing during spontaneous production, (inter)active comprehension, and non-interactive comprehension. Specifically, we aimed to test the shared processing between passive listening and active comprehension and production. To do so, for each subject, we estimated a comprehension encoding model using the story data only and then evaluated the fitted model on the subject's conversation data. We extracted the same four feature spaces from the story (Figure 1C), and evaluated the model performance similarly to the conversation models (Figure 1E).

We found significant within-subject generalization performance from the passive listening paradigm to the conversational paradigm for production and comprehension (Figure 3). Generalization to conversational comprehension was stronger than to production. However, both were lower than when training on conversational data, capturing only a portion of the variance as training on conversations—even when equating their training data (Figure S1). Training on conversation data resulted in an increase of +41% in average encoding performance for comprehension and an increase of +49% for production. A paired t-test found a significant difference ($p < 0.00212$) between subjects' average vertex encoding performance when training on conversation or story. Generalization performance was more bilateral than performance based on the conversational paradigm. An overlapping set of regions was well predicted, particularly STG during comprehension and PMC during production. Notably, generalization was poorer for IFG and MFG compared to temporal regions. Though incomplete, generalization from passive comprehension to both production and comprehension in a conversation context provides further evidence for a common subset of linguistic features that span both processes, while still highlighting the boost in these systems during active, naturalistic communication.

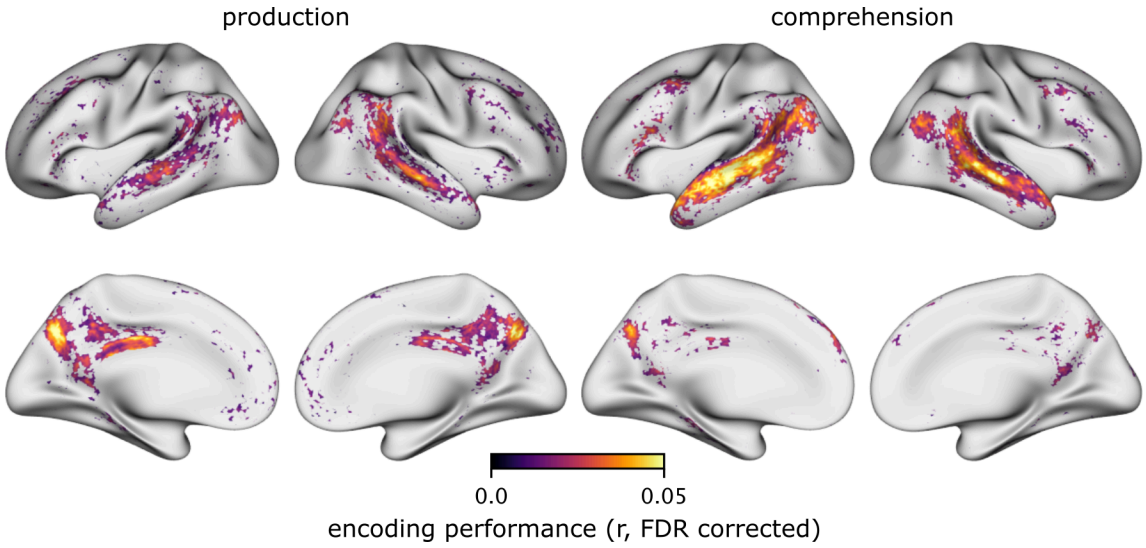

Figure 3. Encoding models trained on passive listening partially generalize to neural responses during conversations. Participants passively listened to a 13-minute story in a separate scanning session before the hyperscanning procedure. We estimated encoding models using the same four feature spaces from this passive listening-only dataset. Then we evaluated how well they generalize to data acquired during conversational production and comprehension conversations. Here we present only the performance of the contextual embedding feature space, after testing for significance and correcting for multiple comparisons. We found significant encoding performance in STG and PMC, which is significantly weaker than when training on conversations.

Contextual embeddings outperform other features of speech and language

Our modeling framework allows us to test different hypotheses about features of brain activity during production and comprehension by comparing the performance of different models. So far, we have only reported the performance of the contextual word embeddings from a pre-trained language model. However, we can decompose the joint model performance into the relative contribution from each feature space (Figure 1E). Here, we report the performance of each feature space and then use a variance partitioning analysis to compute the unique variance predicted by the contextual LLM embeddings.

We found that during both production and comprehension, the contextual LLM embeddings outperformed all other feature spaces regarding correlation strength and cortical coverage (Figure 4A, Figure S2). Among the lower-level control feature spaces, we observed that the acoustic features were the most predictive, especially in EAC and STG. In contrast, the articulation band was least predictive throughout all regions (likely due to collinearity with the better-fitting acoustic space). Moreover, the confound variables were most predictive in SM, EAC, and aSTG. These regions are likely to exhibit large signal fluctuations between speech production or comprehension and are more susceptible to regressors such as word rate.

Next, we performed a variance partitioning analysis (de Heer et al., 2017; Lee Masson & Isik, 2021; Lescroart et al., 2015) to isolate the unique variance explained by the contextual LLM embeddings. We use hierarchical regression to compare a full model with all features and a nested model excluding the features of interest. In this analysis, the full model is composed of the LLM contextual embeddings (L), acoustic (A), and articulatory phonemic (P) features, resulting in encoding performance $R_{L,A,P}$. The nested model is the same, except that it excludes the LLM contextual embeddings from the predictors. Therefore, the unique contextual variance can be calculated as $U_L = R_{L,A,P} - R_{A,P}$. The contextual embeddings accounted for unique variance bilaterally across all previously reported brain regions (Figure 4B). Together, these results suggest that while part of the variability in brain activity can be predicted by acoustic speech features, the contextual word embeddings of LLMs provide unique predictive power, especially in higher-order regions.

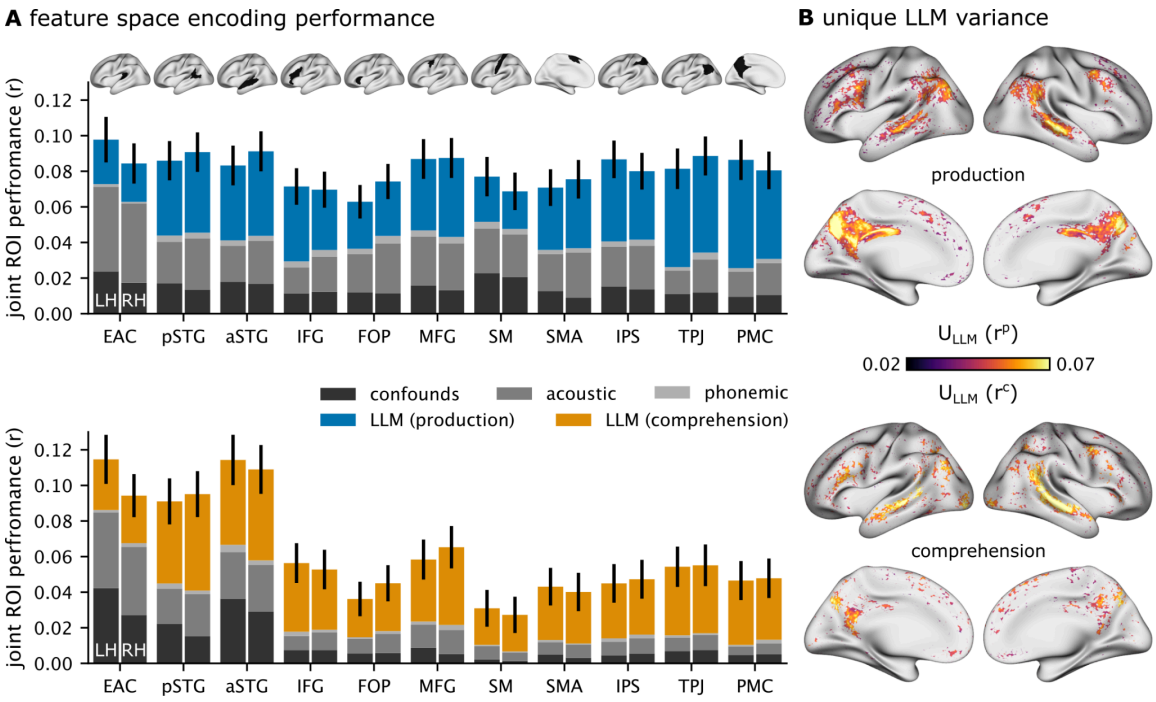

Figure 4. Model comparison and variance partitioning. We compared the variance explained by LLM embeddings with other linguistic feature spaces. (**A**) The joint encoding performance of the full model was decomposed into the contribution of each space separately for production and comprehension. (**B**) We performed a variance partitioning analysis within subjects to quantify the unique contribution of LLM word embeddings. We trained one full encoding model with all features and a nested model with all features, excluding the LLM word embeddings. Then, we subtracted the nested model performance from the full model to quantify the unique variance explained by the LLM embeddings.

Model-based brain-to-brain coupling between conversational partners

When two people converse, we expect their brain activity to align along certain shared features between speech production and comprehension (Hasson et al., 2012; Silbert et al., 2014; Stephens et al., 2010; Zada et al., 2024). Consider a face-to-face conversation: neural activity may align on linguistic features (e.g., the meaning of words) and non-linguistic features (e.g., gestures, facial expressions). To isolate *linguistic* features of shared brain activity across brains, we estimated encoding models from LLM embeddings (jointly with control features) and evaluated how well models trained on one subject generalize to their conversational partner. Specifically, given subject A and their conversational partner subject B, we correlated subject A's production model predictions with subject B's actual comprehension neural responses (Figure 5A). This analysis enabled us to test whether subject A's encoding models in one conversational role can generalize and predict their partner's neural responses in the other conversational role, vertex-by-vertex (Toneva et al., 2022; Zada et al., 2024). Our previous results showed that production and comprehension rely on similar brain regions and share similar linguistic features *within* subjects. This analysis reveals areas where production and comprehension are linguistically coupled *between* subjects.

We found significant model-based speaker-listener coupling for LLM embeddings in the right hemisphere along pSTG, extending into the TPJ, the MFG, and bilaterally in precuneus in PMC (Figure 5B). Because the trained encoding model has to generalize to another subject's brain performing a different process (production vs comprehension), the overall magnitude of the correlation is lower. Interestingly, this model-based linguistic coupling appears right-lateralized (in right-handed subjects). While relatively few vertices in left-hemisphere language areas were significant, we observed strong coupling in right-lateralized temporal areas and in bilateral PMC. For example, brain-to-brain coupling for LLM embeddings was found in the right TPJ, a structure commonly associated with mentalizing and social cognition (Frith & Frith, 1999, 2021). Thus, unlike within subjects where we find broad and bilateral model-based coupling (e.g., in STG, IFG, and PMC), model-based coupling between speaker and listener relies on right-lateralized pSTG and TPJ regions and bilateral precuneus, which are regarded as higher-order cognition areas.

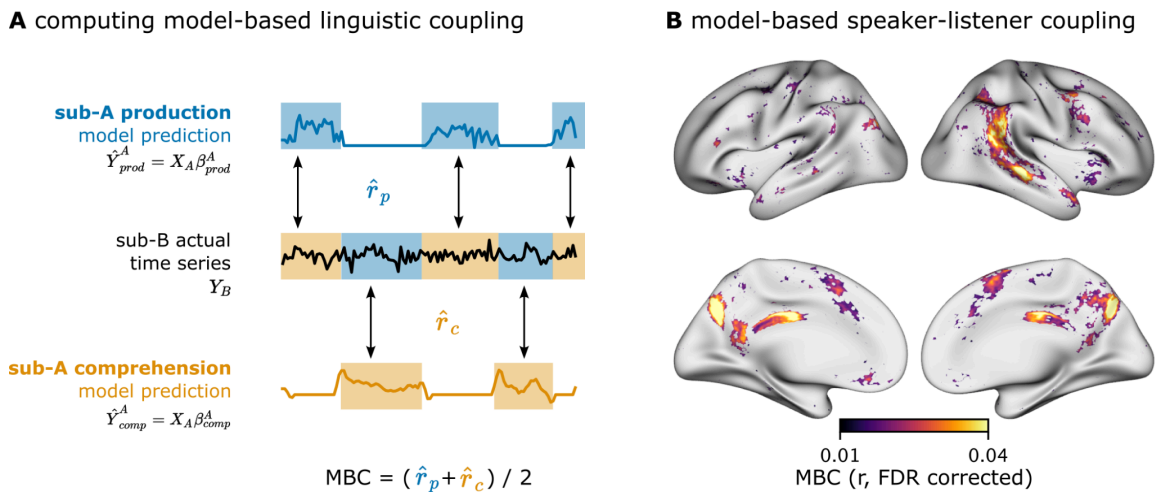

Figure 5. Model-based speaker-listener coupling. (**A**) Schematic of model-based coupling (MBC). We use the already-trained encoding models from subject A's production data to predict subject B's time series during comprehension. We correlate subject A's model predictions with subject B's time series separately for production and comprehension to obtain two correlations per subject per trial. (**B**) We average production and comprehension coupling correlations to obtain a group map of model-based coupling.

Spatial and temporal network structure in model-based conversational coupling

So far, we restricted the scope of speaker-listener coupling in both spatial and temporal dimensions for simplicity: we have only considered coupling between one brain area in the speaker and the homologous area in the listener, and we have only considered instantaneous, or "zero-lag" coupling between partners. The reality is much more complicated. For example, activity in some areas of the speaker's brain may be coupled to activity in *different* regions of the listener's brain, and in some cases, the speaker's brain may *precede* that of the listener (Stephens et al., 2010; Zada et al., 2024). Here, we briefly explore variations in coupling along both of these axes. Since vertices are plentiful, adding spatial and temporal dimensions would exponentially increase the number of comparisons. Thus, we constrain this exploratory analysis to the 11 discussed ROIs.

We first assessed how well a model trained on the speaker's brain activity in one ROI generalizes to the listener's brain activity across all other ROIs. We did this by averaging the predicted and actual time series within each ROI across its vertices. This generated an inter-regional generalization matrix that summarizes the speaker-listener coupling across all ROI pairs at lag 0 (Figure 6A). We observed that the right hemisphere is more connected between speaker and listener than the left hemisphere. Moreover, some areas are relatively uncoupled from other regions (e.g., SM), whereas others are coupled with multiple areas (e.g., pSTG). Interestingly, this matrix has no clear diagonal, meaning that speaker-listener coupling across areas is similarly strong (or weak) to coupling between homologs.

To investigate temporal variation in linguistic coupling, we cross-correlated the predicted time series in subject A with the actual time series of subject B at varying lags (building off of Figure 5A). This resulted in a temporal profile for each ROI pair (484). A carpet plot of these profiles sorted by their peak lag suggests different clusters of temporal coupling (Figure 6B). Most pairs of regions exhibit peak coupling at lag 0 ± 3 seconds (e.g., TPJ, PMC, FOP). For a subset of region pairs, the speaker's brain precedes the listener's brain. For example, the speaker's left MFG and SMA precede the listener's brain activity (e.g., the speaker's MFG precedes the listener's pSTG). In another subset of region pairs, the listener's brain appears to precede the speaker. For example, the speaker's right aSTG and pSTG tend to lag behind the listener's brain activity (e.g., the speaker's aSTG lags behind the listener's pSTG). These results suggest that linguistic coupling between conversation partners is spatially and temporally extended.

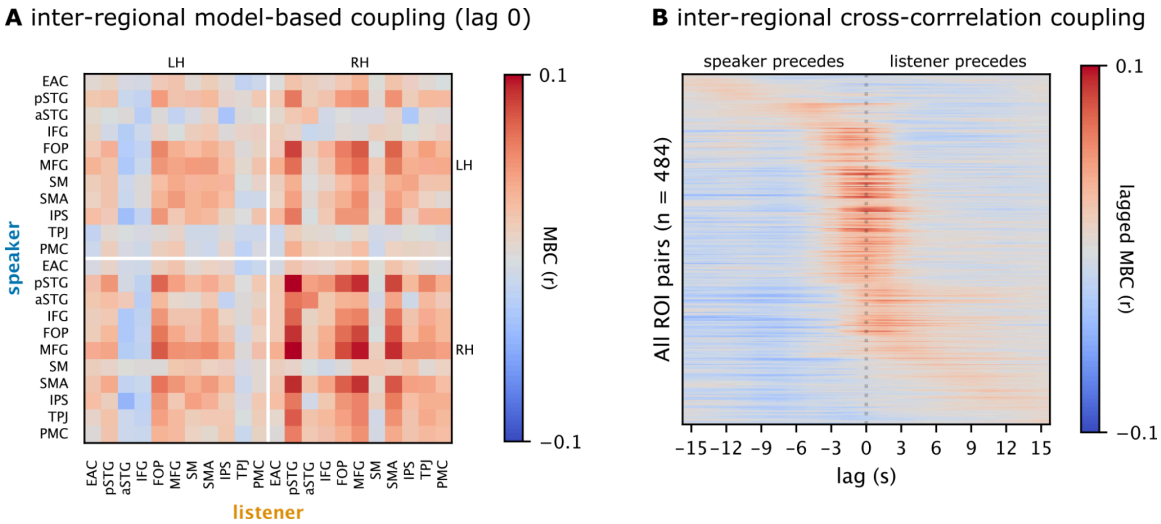

Figure 6. Inter-regional and cross-correlated model-based coupling. (**A**) We extend the speaker-listener model-based coupling results (Figure 5) along two dimensions. First, we correlate a speaker's model-based prediction (averaged across vertices within ROIs) to all ROIs in the listener. (**B**) Second, for each pair of ROIs (a total of 484 pairs across both hemispheres), we cross-correlate the speaker's predicted time series and the listener's actual time series to extend coupling temporally. Rows are ordered by the lag at which they achieve maximum encoding performance.

Discussion

The neural systems involved in speech production and comprehension may require different processes, but they must converge on similar representations. After all, a shared linguistic space is necessary to align the linguistic information across the speaker's and listener's brains. This paper has sought to map the shared neural machinery mediating between speech comprehension and speech production in natural conversations. Our findings revealed that speech production and comprehension recruit

similar brain areas and shared linguistic representations when engaged in natural conversation (Figure 2). In a test of generalization, we found that the brain's linguistic processing during passive story listening is related to spontaneous speech production and comprehension during conversations (Figure 3). However, encoding performance was significantly weaker and missed key frontal language areas than when training on conversations. The model comparison analyses demonstrated that contextual embeddings from an LLM better capture the linguistic features shared between production and comprehension than other candidate models (Figure 4). Finally, our model-based coupling analysis revealed brain-to-brain production-comprehension coupling in high-level cortical areas, particularly right-hemisphere areas associated with social cognition (Figure 5). Overall, our results suggest that speech comprehension and speech production systems align on a set of shared, intermediate features, allowing the brain to translate between the two processes effectively.

We identified a unified language network with shared weights engaged during production and comprehension in real-time conversations. Encoding models have become essential for mapping linguistic features (e.g., acoustic, syntactic, and semantic features) to brain activity. Many recent studies have applied them during passive language comprehension (Caucheteux & King, 2022; de Heer et al., 2017; Deniz et al., 2019; Goldstein et al., 2022; Heilbron et al., 2022; Huth et al., 2016; Kumar et al., 2024; Schrimpf et al., 2021). However, only a handful of recent studies have begun leveraging encoding models for spontaneous language production and active comprehension (Cai et al., 2023; Goldstein et al., 2023; Yamashita et al., 2023), and even fewer have simultaneously recorded two participants engaged in dialogue (Spiegelhalder et al., 2014; Zada et al., 2024). Our encoding models were able to predict neural responses during spontaneous speech production and comprehension (Figure 2A). They also provide an elegant way of comparing these processes within subjects. By constraining the model to share weights, we found that most brain regions exhibited shared linguistic representations between production and comprehension (Figure 2B). Thus, providing evidence for an overlap between speech production and comprehension, which relies on a unified and shared language network. Part of this common network constitutes well-established language regions (Fedorenko, Ivanova, et al., 2024), and extends into general systems responsible for interactive, social cognition. We also found a production-comprehension overlap in low-level perceptual and motor areas (e.g., EAC, SM), suggesting that modality-specific areas may be more localized than previously thought. By using natural conversations, we were able to demonstrate how participants engage these neural processes in real-world, interactive communication (Hagoort, 2019; Wheatley et al., 2024) that embody the principles of ecological validity in social neuroscience (Hasson & Honey, 2012; Nastase et al., 2020; Zaki & Ochsner, 2009).

Advances in simultaneous neuroimaging allowed us to move beyond asynchronous protocols of speaker-listener coupling to real-time, turn-taking conversations. Our hyperscanning paradigm allowed us to simultaneously record brain activity during

speech production and speech comprehension in two interacting subjects. Whereas landmark studies were limited to relating a single subject's production to multiple subject's comprehension responses acquired at a later time (Silbert et al., 2014; Stephens et al., 2010), our paradigm engages each subject's production and comprehension processes in an (inter)active, real-time, turn-taking conversation. We found that conversations recruit more brain regions and different representations than non-interactive paradigms (Figures S1, S5). Similar to studies of asynchronous communication, we observed production-comprehension coupling in PMC, pSTG, and TPJ (Figure 5B). While these studies relied on a content-agnostic approach to speaker-listener coupling (e.g., using ISC, see Introduction), we used a model-based approach to quantify *linguistic* coupling across speakers. Rather than merely showing us *where* coupling occurs, this approach allows us to explicitly model *what* features are coupled across brains (Zada et al., 2024). Moreover, by explicitly representing different linguistic features in one model, we ensure that it is the linguistic information from the contextual LLM embeddings that we find coupled within and between the subject's production and comprehension processes (Figure 4). In doing so, this approach does not register coupling in EAC, as Stephens and colleagues (2010) reported, which likely stems from synchronous, low-level auditory speech features, rather than contentful representations.

We speculate that interactive communication, where partners must actively listen and figuratively "speak to" one another's thoughts and intentions, may engage the social brain in a way that traditional language paradigms do not. Historically, language processing—both comprehension and production processes—has been associated with the left hemisphere (Broca, 1865; Corballis, 2014; Dax, 1865; Knecht et al., 2000; Wernicke, 1874). On the other hand, both ISC analyses (e.g., Nastase et al., 2021) and encoding models (e.g., Huth et al., 2016) tend to yield largely bilateral maps during natural language comprehension. In the current study, we observed brain-to-brain linguistic coupling in the right-lateralized superior temporal cortex, TPJ, and prefrontal cortex, as well as bilateral precuneus and posterior cingulate. This result indicates that the same features that mediate between comprehension and production processes within a brain are also partly shared across individuals. However, these areas are not simply right-hemisphere homologs of typical language regions (Braga et al., 2020; Fedorenko, Ivanova, et al., 2024). In the neuropsychology literature, the right hemisphere has been associated with affective and other paralinguistic features of speech (Heilman et al., 1975; Lindell, 2006), as well as pragmatic and discourse-level processing (Beeman, 1993; Beeman & Chiarello, 1998; Kaplan et al., 1990). Neuroimaging work has generally corroborated these findings (Bottini et al., 1994; Gernsbacher & Kaschak, 2003; Robertson et al., 2000; Vigneau et al., 2011); for example, Yarkoni and colleagues (2008) reported a very similar set of regions to ours, including right TPJ, and bilateral posterior cingulate and precuneus, involved explicitly in tracking narrative comprehension across sentences. Interestingly, several of these

areas overlap with regions often associated with mentalizing and other aspects of social cognition (Frith & Frith, 2012; Saxe, 2006), highlighting the key role that the social brain may play in real-time, naturalistic social interactions.

Large language models are trained to predict the next word in large text corpora. After training, these models can generate increasingly fluent, surprisingly meaningful language, one word at a time, by sampling from a probability distribution of upcoming words. These models do not have dedicated systems for comprehension or production resembling anything like the human brain. Why do these models capture neural activity so well during language comprehension and production? Generative language models operate in a simple perception-action loop by mapping each current word to predict the upcoming word (Pulvermüller, 2018). We speculate that this constraint, which forces language models to learn *shared* representations that inform upcoming word predictions, may yield embeddings that can capture brain activity during both comprehension and production. While the brain has specialized systems for perception and production, our findings suggest that many of the brain's language machinery occupies a middle ground to LLM embeddings: multimodal, active representations with mixed features for both comprehension and production.

Methods

Participants

Thirty dyads (*N* = 60 participants) engaged in real-time conversations while they were simultaneously scanned with fMRI hyperscanning. These data are a subset of a larger dataset collected with additional conditions and participants (see Speer et al., 2024). Participants were recruited from Princeton University and received monetary compensation for their participation. Eligibility requirements included: must be 18 years or older, right-handed, and with normal or corrected vision. Of the 58 included participants, 41 were female, and the average age was 20.74 (minimum 18, maximum 36). One dyad was excluded due to an unexpected scanning issue that resulted in fewer conversations than others.

Design

Two participants at a time arrived at two fMRI scanners in adjacent rooms. The participants did not know each other before the experiment but briefly met before entering the scanners. Participants were instructed to engage in prompted conversations across five runs. Prompts were specifically designed to increase the level of intimacy of conversations across the runs, and are based on stimuli from Aron and colleagues (1997) (Table S1). Each run was 13:36 minutes long and consisted of four trials. We only used two trials of each run because the other two trials were not spontaneous conversations, and were used for a different experiment. Each trial was

03:21 minutes long and started with the topic prompt displayed on screen for 9 seconds, followed by the conversation for 180 seconds, and ended with 12 seconds of a fixation cross (Figure 1). The participant who would start speaking first was randomly assigned. Once a participant finished their utterance, they were instructed to press a button to "pass the virtual mic" to their conversational partner. When a participant had the virtual mic, the screen displayed the text "your turn to speak, when you want to pass the mic, press '1'", followed by a countdown timer displaying the number of seconds left. When listening, the screen showed "your turn to listen", followed by the same countdown timer. Participants were instructed to fill the entire three minutes. After all runs, participants filled out a survey answering questions about the level of enjoyment, similarity, and closeness they felt during their conversations.

MRI acquisition

We recorded neuroimaging data using 3T Siemens Skyra and 3T Siemens Prisma MRI systems. Both machines were configured using the same scanning parameters. Functional scans were acquired with whole brain coverage in interleaved order: 3.0 mm slice thickness, 3.0 × 3.0 mm in-plane resolution, flip angle = 80°, TE = 28 ms, TR = 1500 ms, multiband acceleration factor = 2. A T1-weighted image was acquired for anatomical reference: 1.0 × .0 × 1.0 mm resolution, 176 sagittal slices, flip angle = 9°, TE = 2.98 ms, TR = 2300 ms. To minimize head movement, the subjects' heads were stabilized with foam padding.

Conversation audio transcription

Each three-minute audio segment was transcribed, aligned, and diarized (assigned unique speaker labels) at the word level using WhisperX (Bain et al., 2023)—an automatic speech recognition tool. We used the *faster-whisper-large-v2* model and set the minimum and maximum speakers to two. Each resulting transcription consisted of each word spoken, its onset and duration, and the identity of the speaker.

fMRIPrep preprocessing

Results included in this manuscript come from preprocessing performed using *fMRIPrep* 20.2.0 (Esteban et al., 2018, 2019), which is based on *Nipype* 1.5.1 (K. Gorgolewski et al., 2011; K. J. Gorgolewski et al., 2018) and *Nilearn* 0.6.2 (Abraham et al., 2014).

T1-weighted images were corrected for intensity non-uniformity (INU) with *N4BiasFieldCorrection* (Tustison et al., 2010), distributed with ANTs 2.3.3 (Avants et al., 2008), and used as a reference throughout the workflow. The T1 reference was then skull-stripped with a *Nipype* implementation of the *antsBrainExtraction.sh* workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on

the brain-extracted T1 image using *fast* (FSL 5.0.9 Zhang et al., 2001). Brain surfaces were reconstructed using *recon-all (FreeSurfer 6.0.1 Dale et al., 1999)*, and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (Klein et al., 2017). Individual cortical surface reconstructions were aligned to the *fsaverage6* surface template (40,962 vertices per hemisphere) based on sulcal curvature (Fischl et al., 1999).

Functional data preprocessing

For each of the 6 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated. A deformation field to correct for susceptibility distortions was estimated based on *fMRIPrep*'s *fieldmap-less* approach. The deformation field is constructed by co-registering the BOLD reference to the same-subject T1 reference with inverted intensity (Huntenburg, 2014; Wang et al., 2017). Registration is performed with *antsRegistration* (ANTs 2.3.3), and the process is regularized by constraining deformation to be nonzero only along the phase-encoding direction, and modulated with an average fieldmap template (Treiber et al., 2016). Based on the estimated susceptibility distortion, a corrected BOLD reference was calculated for a more accurate co-registration with the anatomical reference.

The BOLD reference was then co-registered to the T1w reference using FreeSurfer's *bbregister*, which implements boundary-based registration (Greve & Fischl, 2009). Co-registration was configured with six degrees of freedom. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) were estimated before any spatiotemporal filtering using *mcflirt* (FSL 5.0.9 Jenkinson et al., 2002). BOLD runs were slice-time corrected using *3dTshift* from AFNI 20160207 (Cox & Hyde, 1997). The BOLD time series were ultimately resampled onto the *fsaverage6* surface template using FreeSurfer's *mri_vol2surf*. Resampling was performed with a single interpolation step by applying a single, composite transform to correct for head motion, slice-timing, susceptibility distortions, and normalization to the surface template. All subsequent analyses were applied to the vertex-level functional data in surface space; our use of the term "vertex" is otherwise synonymous with the use of "voxel" in volumetric analyses (e.g., "voxelwise encoding models").

Several confounding time series were calculated while preprocessing the BOLD data: six head motion parameters, framewise displacement (FD), and a set of physiological components. FD was estimated for each functional run by computing the absolute sum of relative motions (Power et al., 2014). FD was calculated for each functional run using the implementation in *Nipype* (following the definitions by Power et al., 2014). The three global signals are extracted within the CSF, the white matter, and the whole-brain

masks. Additionally, a set of physiological regressors were extracted to allow for anatomically constrained component-based noise correction (aCompCor Behzadi et al., 2007). Principal components are estimated after high-pass filtering the preprocessed BOLD time series using a discrete cosine filter with 128s cut-off. We retained 10 aCompCor components, five estimated from a white matter mask, and five from a CSF mask.

Confound regression and head motion correction

A typical fMRI signal cleaning pipeline involves regressing out nuisance variables from fMRIPrep's output from the BOLD signal across an entire run or scan (Ciric et al., 2017; e.g., Friston et al., 1996; Parkes et al., 2018; Satterthwaite et al., 2013). Nuisance variables include head motion (e.g., rigid-body motion parameters), physiological noise (e.g., cardiac fluctuations), and scanner noise (e.g., signal drift). However, our hyperscanning paradigm with freely alternating speech production and comprehension between subjects requires additional task-related nuisance variables.

From fMRIPrep confounds, we chose the six head motion variables, all available cosine variables, and the top five components from aCompCor for white matter and CSF masks, separately. This resulted in 26 nuisance regressors. Next, we added five regressors based on the task structure (see the previous Design section). Three boxcar regressors were initialized with zeros across the entire run and populated with ones for (1) indicating the two different trial types, (2) indicating turn to speak, and (3) indicating turn to listen. Two indicator regressors were initialized with zeros and filled with ones when either (1) the subject pressed the button to end their turn, or (2) their conversation partner pressed the button (the instructions on the screen switched each time a button was pressed). These regressors were convolved with an HRF to account for the hemodynamic response using Nilearn's *glm.first_level.glover_hrf* implementation. Finally, all confound variables were passed to Nilearn's *signal.clean* function to detrend, regress out the variables, and z-score the time series.

Defining cortical regions of interest

In order to summarize results across the cortex, we first aggregated the 40,962 vertices in each hemisphere into 180 parcels from a widely-used Glasser multimodal parcellation (Glasser et al., 2016). Then, we defined an extended parcel-level language network from four primary sources: a collection of functionally defined language regions (Fedorenko et al., 2010), a probabilistic atlas based on language localizer tasks in 806 subjects (Lipkin et al., 2022), an activation map corresponding to the "language" topic from NeuroSynth (Yarkoni et al., 2011), and an intersubject correlation map (ISC) based on 345 subjects listening to natural stories (Nastase et al., 2021). We thresholded the probabilistic atlas at p=0.10, the NeuroSynth map at t=0.10, and the intersubject map at

r=0.10. We overlaid these four maps to form an extended "meta" map of language areas (Figure S3A).

We grouped the 55 parcels within this final brain map into 11 regions of interest based on their spatial proximity and previously identified groupings (Figure S3B, Table S2). Specifically, following the networks identified by Glasser and colleagues (2016), we identified the following regions: early auditory cortex (EAC), posterior and anterior superior temporal gyrus (pSTG, aSTG), inferior and middle frontal gyri (IFG, MFG), somatomotor cortex (SM), supplementary motor area (SMA), frontal operculum (FOP), intraparietal sulcus (IPS), temporoparietal junction (TPJ), and posterior medial cortex (PMC). Finally, given that the maps derived from prior studies may be biased toward comprehension tasks, we defined a somatomotor region of interest we expect to be involved in language production (Silbert et al., 2014). Note that we are deliberately defining a more inclusive "language network" than prior work (Fedorenko et al., 2010) to explore both more peripheral perception (e.g., EAC) and production (e.g., SM) areas, as well as higher-level areas that may be involved in narrative and social cognition (e.g., TPJ, PMC).

Linguistic features for encoding analysis

In vertex-wise encoding analysis, we use ridge regression to learn a linear model mapping from a set of explicit features (i.e., design matrix) to the observed brain activity (Naselaris et al., 2011). We first re-represent the language task and stimulus in one or more feature spaces. We defined several feature spaces from the conversation stimuli to build these design matrices.

Task structure and nuisance variables. We computed four low-level variables from each transcript that could affect the BOLD signal (Huth et al., 2016). For each TR, we quantify the word rate (number of words in a TR), phoneme rate (number of phonemes in a TR), word occurrence (some TRs contained no words), and a variable indicating whether it was the subject's turn to speak or listen. The word and phoneme rates were continuous, while the word onset and indicator variables were binary.

Acoustic spectral features. For each pair of subjects, we had one audio recording of the entire conversation that was recorded from one mic at a time and switched upon button presses indicating the end of turn. We computed acoustic features from the speech audio files (de Heer et al., 2017). Specifically, we used the *WhisperFeatureExtractor* class from the *HuggingFace* (Wolf et al., 2020) library with the default settings to extract a spectral representation of the audio. This function uses a short-time Fourier transform to compute a mel-filter bank of 80 features that represent the spectral power density on a Mel log scale. Note that these features likely capture more than just acoustic features because they were recorded in MRI machines with different noise characteristics, and were saved into one file from two sources. Thus, at minimum, it also encodes information about the conversation turns.

Articulatory phonemic features. Following de Heer et al., (2017), we quantify the articulatory features of speech based on the phonemes in the transcript. Specifically, we used the CMU pronunciation dictionary (http://www.speech.cs.cmu.edu/cgi-bin/cmudict) to obtain the phonemes associated with each word in the transcript. We then constructed the articulatory features for each phoneme based on the place and manner of consonants, and voicing of vowels. This resulted in a binary vector of 22 features for each phoneme.

Large language model features. We extracted word embeddings from the large language model GPT-2 XL (Radford et al., 2019) using the *HuggingFace* library (Wolf et al., 2020). For each 3-minute conversation transcript, we first converted all words to GPT-2 tokens. We then passed these tokens as input to the LLM, where they were converted to 1,600-dimensional token embeddings and passed through the decoder layers. We extracted the activations from the middle (24th) layer to serve as contextual word embeddings.

Encoding model construction and evaluation

Encoding models were the core analytical approach we took to estimating linguistic content in the BOLD signal (Naselaris et al., 2011). For all analyses, we used kernel ridge regression to prevent overfitting, and banded ridge regression to find different regularization parameters for each feature space separately (Nunez-Elizalde et al., 2019). We used the *MultipleKernelRidgeCV* class from the *himalaya* library (Dupré La Tour et al., 2022) to perform cross-validation within the training set to select the best regularization parameter per feature space. All results we report on encoding performance were evaluated on a held-out test sample.

Design matrix construction. Each 3-minute conversation (trial) consisted of a 120-TR BOLD time series. With two trials per run (240 TRs) and five total runs, we had a total of 1,200 TRs per subject. Thus, our design matrix had 1,200 rows. The initial number of columns was based on the selected feature spaces for each analysis. For example, for the full joint model (Figure 1), we used five feature spaces: task (8 dimensions), acoustic (80), articulatory (22), and contextual embeddings (1,600). Stimuli features that were defined on the word or token level were averaged within TRs (e.g., LLM embeddings). Then, we split each feature space into two groups, for production or comprehension, and filled the gaps between one process and the other with zeros (see Figure 1C).

Model definition. We used a Scikit-learn (Pedregosa et al., 2011) pipeline to define the full encoding model. The pipeline consisted of three main steps before model fitting. First, the regressors were mean-centered using *StandardScaler*. Then, each feature space was duplicated and shifted by 2–5 TRs (3–7.5 s) to account for the hemodynamic lag in the BOLD signal (Huth et al., 2016). Finally, because the design matrix was wider than it is long, we used the kernel method to solve the ridge regression in its dual form

(Dupré La Tour et al., 2024). Specifically, we used a linear kernel for each feature space separately before fitting the model.

Model fitting and evaluation. We used cross-validation to evaluate each model on a held-out test sample. Specifically, we defined five folds, based on the five runs, to fit a model on four runs (960 TRs), and tested it on the held-out run (240 TRs). We repeated this procedure five times, testing each run in turn, and then averaging the encoding performance across the five runs. Each run contained unique conversations based on different prompts.

Banded ridge regression allows us to evaluate each feature space separately, relative to all the others. To do this, the joint predicted time series on the held-out run can be decomposed into one time series per feature space (Dupré La Tour et al., 2022). Similarly, the encoding model performance (i.e., the correlation between the predicted and actual time series) can be split into one correlation for each feature space (Figure 1C). Importantly, we segmented the actual and predicted time series into production and comprehension TRs to obtain their separate correlations for each process. Moreover, because of the hemodynamic response, some TRs may be affected by both processes. Thus, we selected the exclusive set of TRs where there is no overlap.

Finally, we confirmed that head motion degrades encoding performance and that there is considerably more head motion during speech production than comprehension (Figure S4).

Statistical significance. We tested whether a vertex's encoding performance correlation is statistically significant by using a two-sided, one-sample *t*-test, as implemented in SciPy (Virtanen et al., 2020). All p-values were corrected for multiple comparisons by controlling the false discovery rate (FDR; Benjamini & Hochberg, 1995).

Speaker–listener model-based coupling

We used the already-trained encoding models to evaluate the model-based coupling between conversation partners. The intuition behind this evaluation is to correlate one subject's model-predicted time series with their conversational partner's actual time series (as opposed to correlating it with their *own* actual time series). In effect, this simultaneously tests whether the model can generalize from one subject to another and from one process to another (e.g., production to comprehension) (Toneva et al., 2022). Thus, we use the same evaluation procedure as described before, except with one major change. For each voxel, we correlate a subject's predicted time series with their partner's actual time series for the same voxel. Critically, we use the predictions from all feature spaces and compute the relative encoding performance of the LLM contextual embedding feature space only. By applying the same evaluation procedure as within-subject, we control for variance that can be explained by the nuisance feature spaces. When testing model-based coupling across regions and time (Figure 6), we first

extract speaker turns that are at least 9 seconds long in order to exclude turns that are too short.

Story-listening task and analysis

Prior to hyperscanning acquisition, participants listened to a ~13-minute story ("I Knew You Were Black" by Carol Daniel). Three participants did not complete this task and were excluded from this particular analysis. We used the same procedures as described above for conversations for the story, including MRI acquisition parameters, BOLD preprocessing, confound regression, linguistic features, and encoding model construction, training, and evaluation. However, there were two differences. First, the confound regression did not include any design structure variables. Second, we did not split regressors because the story is only comprehension—thus this model corresponds to the shared-weights model for conversations.

Software resources

In addition to the software mentioned throughout the Methods, we used *Surfplot* (Gale et al., 2021) for visualizing brain maps.

Acknowledgments

We would like to extend thanks to Ahmad Samara and Itamar Jalon for helpful feedback on head motion correction and fMRI preprocessing details.

Funding:

National Institutes of Health grant R21MH127284 (DT)
National Institutes of Health grant DP1HD091948 (UH)

Author contributions:

Conceptualization: ZZ, SAN, UH, DT
Data curation: LT, SB, SS
Formal analysis: ZZ
Funding acquisition: DT
Investigation: ZZ
Methodology: ZZ
Project administration: DT
Software: ZZ
Supervision: SAN, DT
Visualization: ZZ
Writing – original draft: ZZ
Writing – review & editing: ZZ, SAN, LMT, UH, DT

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: Code for all results in this manuscript is publicly available on GitHub (https://github.com/zaidzada/fconv). Neural data and transcripts not currently available to protect participant privacy.

Supplementary materials

Includes supplementary figures S1–S4 and tables S1–S2.

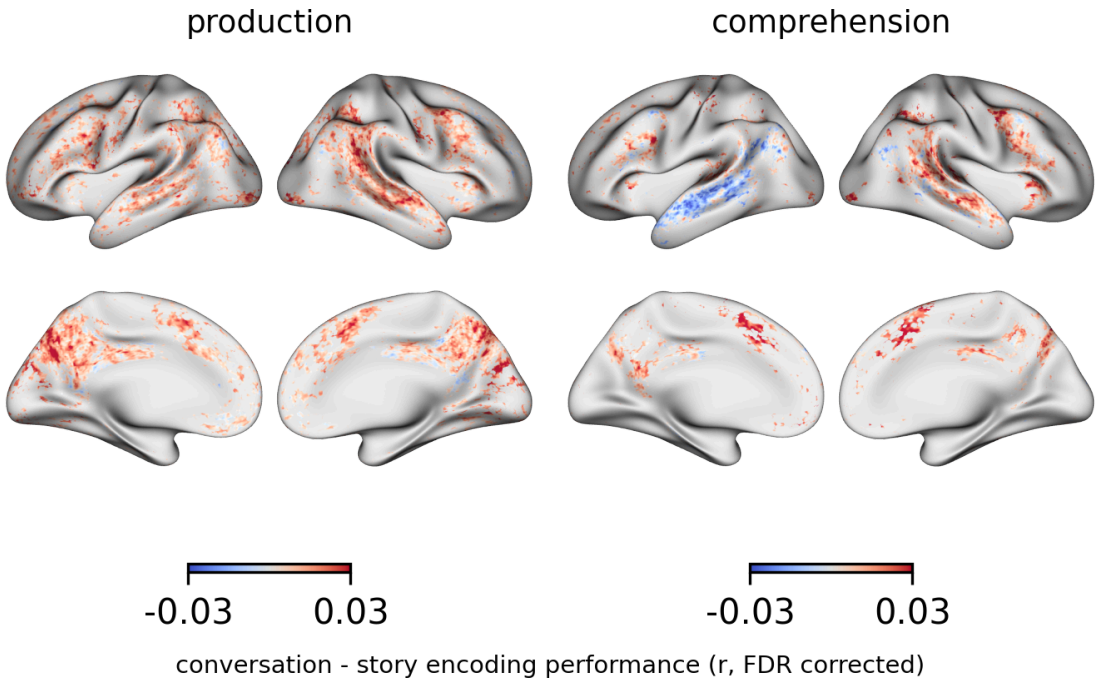

Figure S1. Within-subject contrast between training on conversations versus story. To fairly compare encoding performance between training encoding models on conversational data or story listening data, we performed an additional analysis. Instead of the 5-fold cross validation procedure used before, here we held out 3 conversation runs for testing for both story- and conversation-trained models. For the story, we trained encoding models on all 534 TRs of the story, while for the conversation, we trained encoding models on the first two runs only (480 TRs of both production and comprehension). For conversational encoding models, we used the shared weights model. Thus, this analysis ensures that both encoding models are tested on the same data and have roughly similar training set sizes. Evaluation of each model was performed *within-subjects* (i.e., training and testing on each subject's data separately). We plot the difference between conversation- and story- encoding model performance while thresholding for significantly predicted vertices only. Positive numbers (red) reflect vertices where training on conversational data performs better, and vice versa for negative numbers (blue).

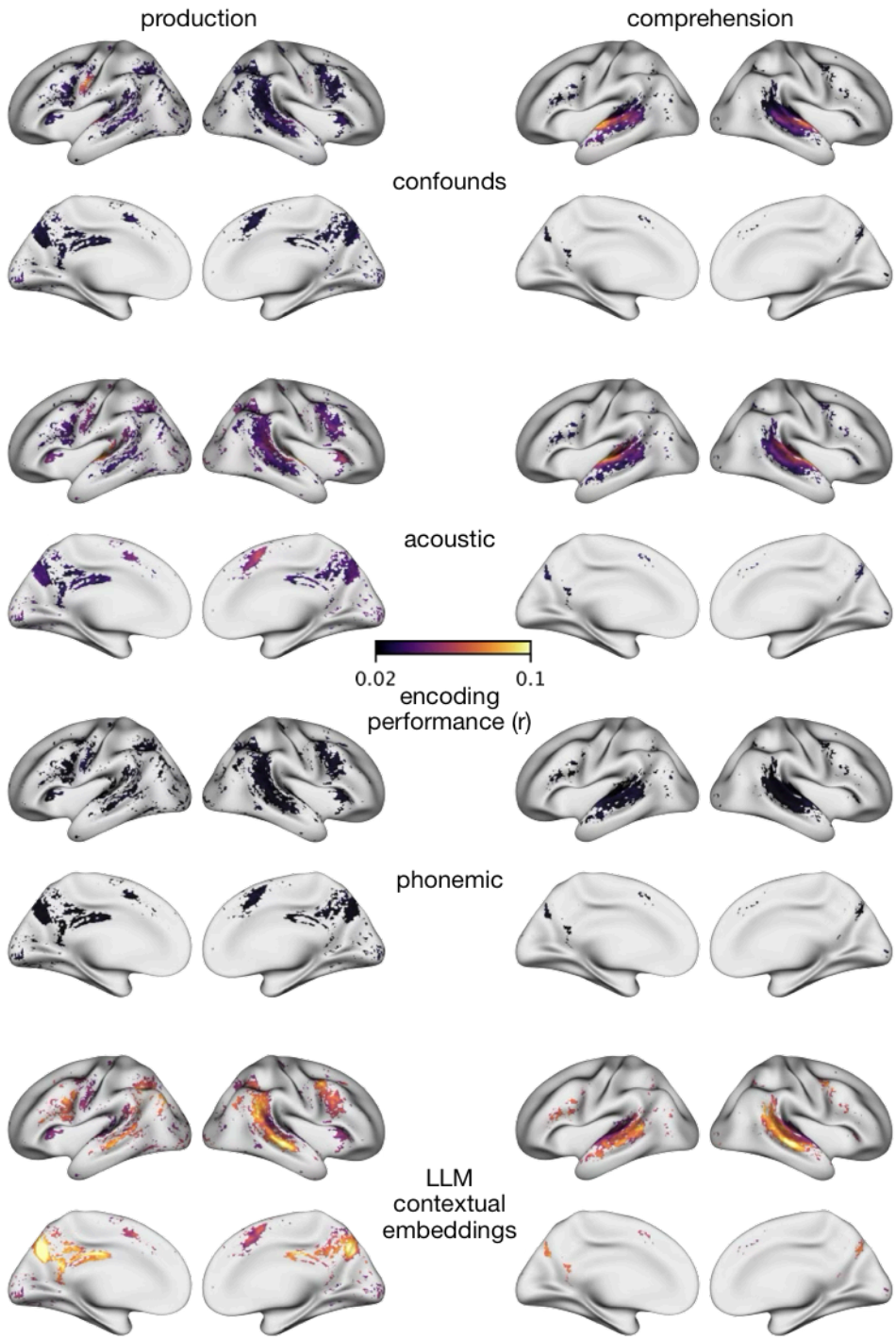

Figure S2. Within-subject encoding performance per band during production and comprehension. The joint model encoding performance can be decomposed into the relative contribution per feature space (Figure 1). Moreover, we evaluate production time points separately from comprehension time points. Here, we threshold the brain maps using a one-sample *t*-test based on the joint model performance, and then apply Bonferroni correction.

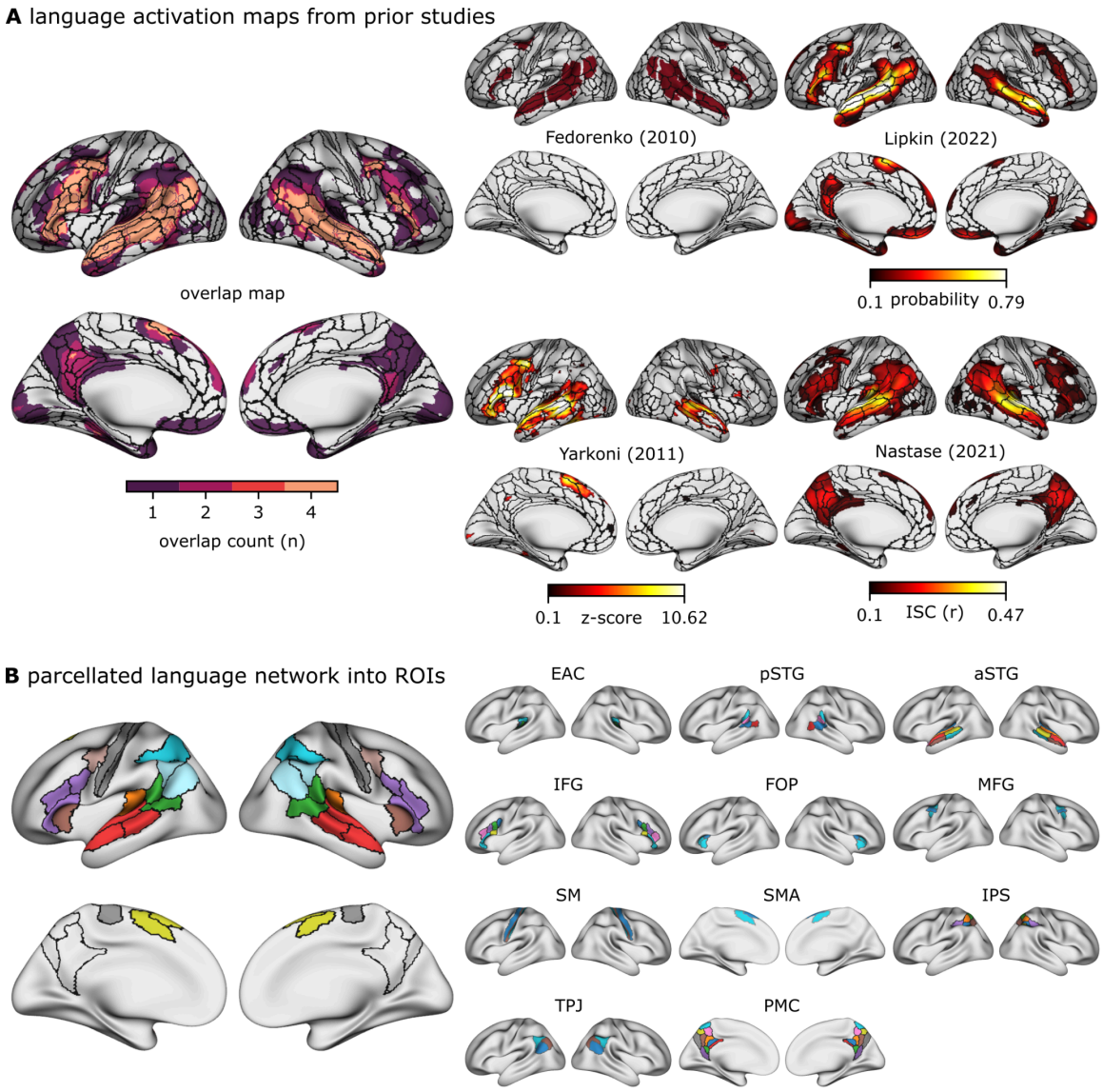

Figure S3. Regions of interest within the extended language network. (**A**) We define linguistic regions of interest based on the overlap of four primary sources of language-related brain maps. See Methods for details on thresholding. (**B**) Then, we select parcels in the Glasser atlas where overlap occurs, and group parcels into 11 regions per hemisphere.

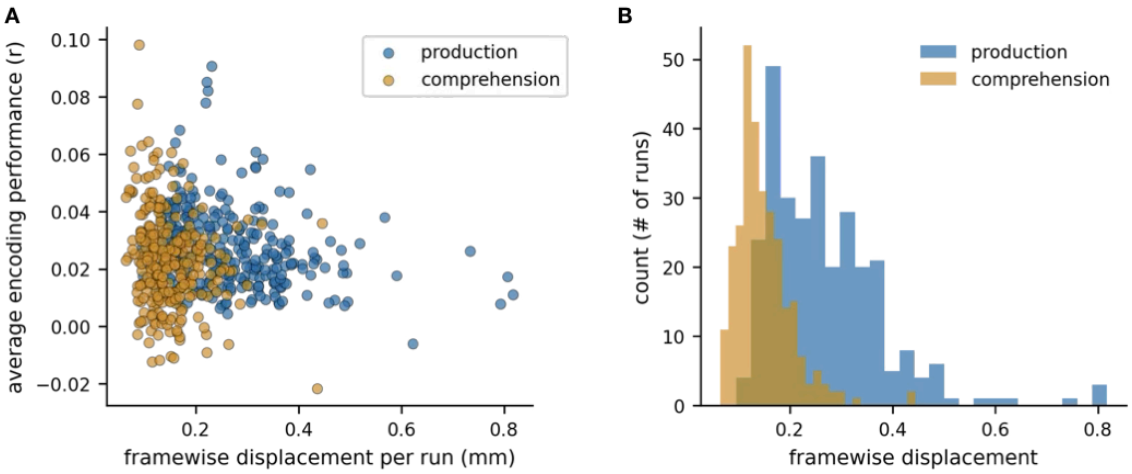

Figure S4. Head motion impact on model performance. (**A**) We found that head motion degrades the model performance for both production ($r = -0.291$, $p < 1e\text{-}07$) and comprehension ($r = -0.207$, $p < 0.00038$). (**B**) Production and comprehension histogram of the average framewise displacement per subject for each of their five runs. As expected, more head motion (as measured with framewise displacement) occurs during production than comprehension.

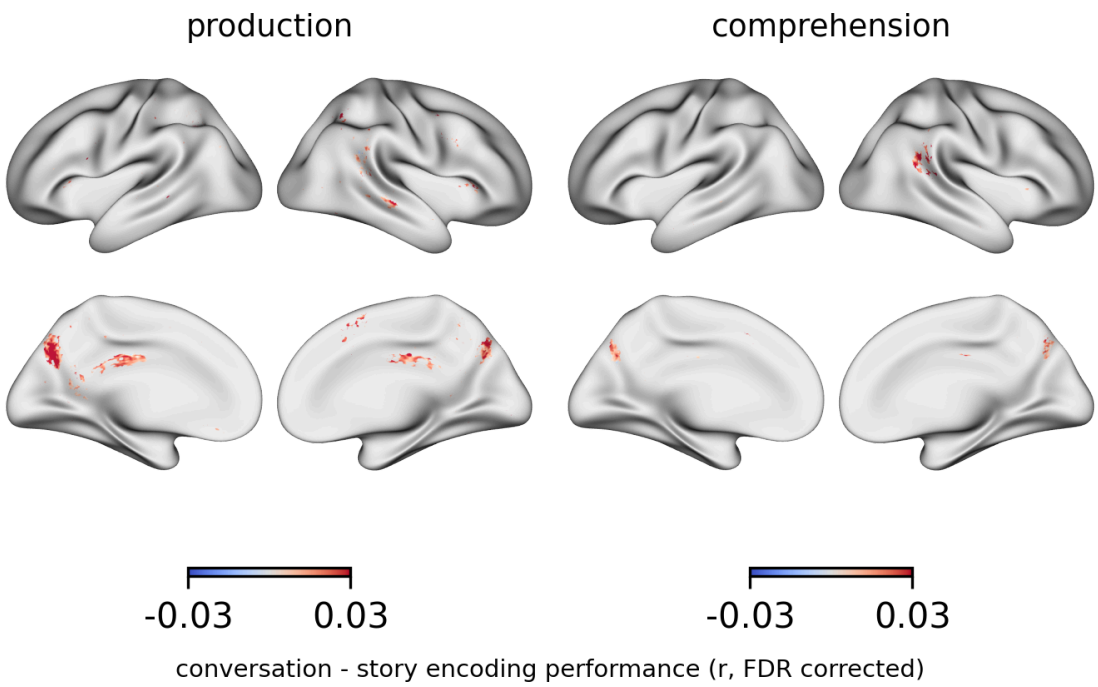

Figure S5. Across-subject contrast between training on conversations versus story. Here we test the difference in encoding performance *across subjects* when training on conversational or story data. The procedure is the same as Figure S1 but instead of evaluating each model on a subject's own data, we test it on their conversational partner's neural data for the three held-out conversation runs.

prompt	set	prompt text
1	1	Given the choice of anyone in the world, whom would you want as a dinner guest?
2	1	Would you like to be famous? In what way?
3	1	Before making a telephone call, do you ever rehearse what you are going to say? Why?
4	1	What would constitute a "perfect" day for you?
5	1	When did you last sing to yourself? To someone else?
6	1	If you were able to live to the age of 90 and retain either the mind or body of a 30-year-old for the last 60 years of your life, which would you want?
7	1	For what in your life do you feel most grateful?
8	1	If you could change anything about the way you were raised, what would it be?
9	2	If a crystal ball could tell you the truth about yourself, your life, the future, or anything else, what would you want to know?
10	2	Is there something that you've dreamed of doing for a long time? Why haven't you done it?
11	2	What is the greatest accomplishment of your life?
12	2	What do you value most in a friendship?
13	2	What is your most treasured memory?
14	2	How close and warm is your family? Do you feel your childhood was happier than most other people's?
15	3	Complete this sentence: I wish I had someone with whom I could share…
16	3	Please share what would be important for your study partner to know as your close friend.
17	3	Share with your partner an embarrassing moment in your life.
18	3	What, if anything, is too serious to be joked about?
19	3	Your house, containing everything you own, catches fire. After saving your loved ones and pets, you have time to safely make a final dash to save any one item. What would it be? Why?
20	3	Share a personal problem and ask your partner's advice on how he or she might handle it. Also, ask your partner to reflect back to you how you seem to be feeling about the problem you have chosen.

Table S1. Conversation topic prompts. Participants were presented with 20 different prompts to inspire otherwise free-form conversations. The prompts were constructed to become increasingly personal over the course of the experiment.

ROI	sub-group	parcels from Glasser (2016)
EAC	EAC	[A1, LBelt, MBelt, PBelt, RI]
pSTG	AG	[TPOJ1, TPOJ2]
pSTG	SMG	[STV, PSL]
aSTG	STG	[A4, A5]
aSTG	aSTS	[STSda, STSva, STGa]
aSTG	pSTS	[STSdp, STSvp]
IFG	IFJ	[IFJp, IFJa]
IFG	IFS	[IFSp, IFSa]
IFG	IFG	[44, 45, 47l]
MFG	MFG	[55b, FEF, PEF]
SM	M1	[4]
SM	S1	[3a, 3b]
FOP	FOP	[FOP4, FOP5, AVI]
SMA	SFL1	[SFL]
SMA	SFL2	[SCEF]
IPS	SPC	[LIPd, LIPv, VIP, AIP, MIP]
IPS	dIPC	[IP0, IP1, IP2]
TPJ	IPC1	[PGi, PGs]
TPJ	IPC2	[PFm]
PMC	PCC1	[31pv, 31pd, v23ab, d23ab, POS1, 7m, PCV]
PMC	PCC3	[POS2]
PMC	SPC1	[7Pm, 7Am]

Table S2. Language network atlas constituents. We constructed 11 ROIs spanning an extended language network, including early auditory areas, language areas, higher-level areas associated with semantic representation and narrative processing, and somatomotor areas. The ROIs were selected based on prior work (Figure S3) and constructed by combining parcels from a multimodal atlas (Glasser et al., 2016).

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*. https://doi.org/10.3389/fninf.2014.00014

Aron, A., Melinat, E., Aron, E. N., Vallone, R. D., & Bator, R. J. (1997). The Experimental Generation of Interpersonal Closeness: A Procedure and Some Preliminary Findings. *Personality & Social Psychology Bulletin*, *23*(4), 363–377.

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26–41.

Babiloni, F., & Astolfi, L. (2014). Social neuroscience and hyperscanning techniques: Past, present and future. *Neuroscience and Biobehavioral Reviews*, *44*, 76–93.

Bain, M., Huh, J., Han, T., & Zisserman, A. (2023). WhisperX: Time-Accurate Speech Transcription of Long-Form Audio. *INTERSPEECH 2023*. https://github.com/m-bain/whisperX

Beeman, M. (1993). Semantic processing in the right hemisphere may contribute to drawing inferences from discourse. *Brain and Language*, *44*(1), 80–120.

Beeman, M., & Chiarello, C. (1998). Complementary Right- and Left-Hemisphere Language Comprehension. *Current Directions in Psychological Science*, *7*(1), 2–8.

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, *57*(1), 289–300.

Bevilacqua, D., Davidesco, I., Wan, L., Chaloner, K., Rowland, J., Ding, M., Poeppel, D., & Dikker, S. (2019). 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.

Bottini, G., Corcoran, R., Sterzi, R., Paulesu, E., Schenone, P., Scarpa, P., Frackowiak, R. S., & Frith, C. D. (1994). The role of the right hemisphere in the interpretation of figurative aspects of language. A positron emission tomography activation study. *Brain: A Journal of Neurology*, *117 (Pt 6)*(6), 1241–1253.

Braga, R. M., DiNicola, L. M., Becker, H. C., & Buckner, R. L. (2020). Situating the left-lateralized language network in the broader organization of multiple specialized large-scale distributed networks. *Journal of Neurophysiology*, *124*(5), 1415–1448.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *22*(6), 1482–1493.

Broca, P. (1865). Sur le siège de la faculté du langage articulé. *Bulletins de La Société D'anthropologie de Paris*, *6*(1), 377–393.

Cai, J., Hadjinicolaou, A. E., Paulk, A. C., Williams, Z. M., & Cash, S. S. (2023). *Natural language processing models reveal neural dynamics of human conversation*. Neuroscience. http://biorxiv.org/lookup/doi/10.1101/2023.03.10.531095

Caucheteux, C., Gramfort, A., & King, J.-R. (2023). Evidence of a predictive coding hierarchy in the human brain listening to speech. *Nature Human Behaviour*, *7*(3), 430–441.

Caucheteux, C., & King, J.-R. (2022). Brains and algorithms partially converge in natural language processing. *Communications Biology*, *5*(1), 134.

Chang, C. H. C., Nastase, S. A., & Hasson, U. (2023). *How a speaker herds the audience: Multi-brain neural convergence over time during naturalistic storytelling*. Neuroscience. http://biorxiv.org/lookup/doi/10.1101/2023.10.10.561803

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, *20*(1), 115–125.

Ciric, R., Wolf, D. H., Power, J. D., Roalf, D. R., Baum, G. L., Ruparel, K., Shinohara, R. T., Elliott, M. A., Eickhoff, S. B., Davatzikos, C., Gur, R. C., Gur, R. E., Bassett, D. S., & Satterthwaite,

T. D. (2017). Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *NeuroImage*, *154*, 174–187.

Clark, H. H. (1996). *Using language*. Cambridge university press.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In *Perspectives on socially shared cognition.* (pp. 127–149). American Psychological Association.

Corballis, M. C. (2014). Left brain, right brain: facts and fantasies. *PLoS Biology*, *12*(1), e1001767.

Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine*, *10*(4-5), 171–178.

Czeszumski, A., Eustergerling, S., Lang, A., Menrath, D., Gerstenberger, M., Schuberth, S., Schreiber, F., Rendon, Z. Z., & König, P. (2020). Hyperscanning: A Valid Method to Study Neural Inter-brain Underpinnings of Social Interaction. *Frontiers in Human Neuroscience*, *14*, 39.

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical Surface-Based Analysis: I. Segmentation and Surface Reconstruction. *NeuroImage*, *9*(2), 179–194.

Davidesco, I., Laurent, E., Valk, H., West, T., Milne, C., Poeppel, D., & Dikker, S. (2023). The Temporal Dynamics of Brain-to-Brain Synchrony Between Students and Teachers Predict Learning Outcomes. *Psychological Science*, *34*(5), 633–643.

Dax, M. (1865). Lesions de la motie gauche de l’encephale coincident avec l'oublie des signes de la pensee. *Gaz Hbd Med Chir*, *2*, 259–262.

de Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L., & Theunissen, F. E. (2017). The Hierarchical Cortical Organization of Human Speech Processing. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *37*(27), 6539–6557.

Deniz, F., Nunez-Elizalde, A. O., Huth, A. G., & Gallant, J. L. (2019). The Representation of Semantic Information Across Human Cerebral Cortex During Listening Versus Reading Is Invariant to Stimulus Modality. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *39*(39), 7722–7736.

Dikker, S., Silbert, L. J., Hasson, U., & Zevin, J. D. (2014). On the Same Wavelength: Predictable Language Enhances Speaker–Listener Brain-to-Brain Synchrony in Posterior Superior Temporal Gyrus. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(18), 6267–6272.

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., Rowland, J., Michalareas, G., Van Bavel, J. J., Ding, M., & Poeppel, D. (2017). Brain-to-Brain Synchrony Tracks Real-World Dynamic Group Interactions in the Classroom. *Current Biology: CB*, *27*(9), 1375–1380.

Dupré La Tour, T., Eickenberg, M., Nunez-Elizalde, A. O., & Gallant, J. L. (2022). Feature-space selection with banded ridge regression. *NeuroImage*, *264*, 119728.

Dupré La Tour, T., Visconti Di Oleggio Castello, M., & Gallant, J. L. (2024). *The Voxelwise Modeling framework: a tutorial introduction to fitting encoding models to fMRI data*. https://doi.org/10.31234/osf.io/t975e

Esteban, O., Blair, R., Markiewicz, C. J., Berleant, S. L., Moodie, C., Ma, F., Isik, A. I., Erramuzpe, A., Kent, M., James D. andGoncalves, DuPre, E., Sitek, K. R., Gomez, D. E. P., Lurie, D. J., Ye, Z., Poldrack, R. A., & Gorgolewski, K. J. (2018). fMRIPrep. *Software*. https://doi.org/10.5281/zenodo.852659

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2019). fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116.

Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New method for fMRI investigations of language: defining ROIs functionally in individual subjects. *Journal of Neurophysiology*, *104*(2), 1177–1194.

Fedorenko, E., Ivanova, A. A., & Regev, T. I. (2024). The language network as a natural kind within the broader landscape of the human brain. *Nature Reviews. Neuroscience*, *25*(5), 289–312.

Fedorenko, E., Piantadosi, S. T., & Gibson, E. A. F. (2024). Language is primarily a tool for communication rather than thought. *Nature*, *630*(8017), 575–586.

Fischl, B., Sereno, M. I., Tootell, R. B. H., & Dale, A. M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping*, *8*(4), 272–284.

Friederici, A. D. (2011). The Brain Basis of Language Processing: From Structure to Function. *Physiological Reviews*, *91*(4), 1357–1392.

Friston, K. J., Williams, S., Howard, R., Frackowiak, R. S. J., & Turner, R. (1996). Movement-Related effects in fMRI time-series. *Magnetic Resonance in Medicine*, *35*(3), 346–355.

Frith, C. D., & Frith, U. (1999). Interacting Minds--A Biological Basis. *Science (New York, N.Y.)*, *286*(5445), 1692–1695.

Frith, C. D., & Frith, U. (2012). Mechanisms of social cognition. *Annual Review of Psychology*, *63*(1), 287–313.

Frith, C. D., & Frith, U. (2021). Mapping Mentalising in the Brain. In M. Gilead & K. N. Ochsner (Eds.), *The Neural Basis of Mentalizing* (pp. 17–45). Springer International Publishing.

Gale, D. J., Vos de Wael., R., Benkarim, O., & Bernhardt, B. (2021). *Surfplot: Publication-ready brain surface figures*. Zenodo. https://zenodo.org/record/5567926

Gernsbacher, M. A., & Kaschak, M. P. (2003). Neuroimaging studies of language production and comprehension. *Annual Review of Psychology*, *54*(1), 91–114.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., Smith, S. M., & Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178.

Goldstein, A., Wang, H., Niekerken, L., Zada, Z., Aubrey, B., Sheffer, T., Nastase, S. A., Gazula, H., Schain, M., Singh, A., Rao, A., Choe, G., Kim, C., Doyle, W., Friedman, D., Devore, S., Dugan, P., Hassidim, A., Brenner, M., … Hasson, U. (2023). *Deep speech-to-text models capture the neural basis of spontaneous speech in everyday conversations*. Neuroscience.

http://biorxiv.org/lookup/doi/10.1101/2023.06.26.546557

Goldstein, A., Zada, Z., Buchnik, E., Schain, M., Price, A., Aubrey, B., Nastase, S. A., Feder, A., Emanuel, D., Cohen, A., Jansen, A., Gazula, H., Choe, G., Rao, A., Kim, C., Casto, C., Fanda, L., Doyle, W., Friedman, D., … Hasson, U. (2022). Shared computational principles for language processing in humans and deep language models. *Nature Neuroscience*, *25*(3), 369–380.

Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. (2011). Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. *Frontiers in Neuroinformatics*, *5*, 13.

Gorgolewski, K. J., Esteban, O., Markiewicz, C. J., Ziegler, E., Ellis, D. G., Notter, M. P., Jarecka, D., Johnson, H., Burns, C., Manhães-Savio, A., Hamalainen, C., Yvernault, B., Salo, T., Jordan, K., Goncalves, M., Waskom, M., Clark, D., Wong, J., Loney, F., … Ghosh, S. (2018). Nipype. *Software*. https://doi.org/10.5281/zenodo.596855

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill.

Hagoort, P. (2019). The neurobiology of language beyond single-word processing. *Science*, *366*(6461), 55–58.

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

Hasson, U., & Honey, C. J. (2012). Future trends in Neuroimaging: Neural processes as expressed within real-life contexts. *NeuroImage*, *62*(2), 1272–1278.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., & Malach, R. (2004). Intersubject Synchronization of Cortical Activity During Natural Vision. *Science*, *303*(5664), 1634–1640.

Heilbron, M., Armeni, K., Schoffelen, J.-M., Hagoort, P., & De Lange, F. P. (2022). A hierarchy of linguistic predictions during natural language comprehension. *Proceedings of the National*

Academy of Sciences of the United States of America, *119*(32), e2201968119.

Heilman, K. M., Scholes, R., & Watson, R. T. (1975). Auditory affective agnosia. Disturbed comprehension of affective speech. *Journal of Neurology, Neurosurgery, and Psychiatry*, *38*(1), 69–72.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews. Neuroscience*, *8*(5), 393–402.

Huntenburg, J. M. (2014). *Evaluating nonlinear coregistration of BOLD EPI and T1w images* [Freie Universität]. http://hdl.handle.net/11858/00-001M-0000-002B-1CB5-A

Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage*, *17*(2), 825–841.

Jiang, J., Dai, B., Peng, D., Zhu, C., Liu, L., & Lu, C. (2012). Neural Synchronization during Face-to-Face Communication. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(45), 16064–16069.

Kaplan, J. A., Brownell, H. H., Jacobs, J. R., & Gardner, H. (1990). The effects of right hemisphere damage on the pragmatic interpretation of conversational remarks. *Brain and Language*, *38*(2), 315–333.

Kinreich, S., Djalovski, A., Kraus, L., Louzoun, Y., & Feldman, R. (2017). Brain-to-Brain Synchrony during Naturalistic Social Interactions. *Scientific Reports*, *7*(1), 17060.

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Neto, E. C., & Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLOS Computational Biology*, *13*(2), e1005350.

Knecht, S., Deppe, M., Dräger, B., Bobe, L., Lohmann, H., Ringelstein, E.-B., & Henningsen, H. (2000). Language lateralization in healthy right-handers. *Brain: A Journal of Neurology*,

123(1), 74–81.

Kumar, S., Sumers, T. R., Yamakoshi, T., Goldstein, A., Hasson, U., Norman, K. A., Griffiths, T. L., Hawkins, R. D., & Nastase, S. A. (2024). Shared functional specialization in transformer-based language models and the human brain. *Nature Communications*, *15*(1), 5523.

Lee Masson, H., & Isik, L. (2021). Functional selectivity for social interaction perception in the human superior temporal sulcus during natural viewing. *NeuroImage*, *245*, 118741.

Lescroart, M. D., Stansbury, D. E., & Gallant, J. L. (2015). Fourier power, subjective distance, and object categories all provide plausible models of BOLD responses in scene-selective visual areas. *Frontiers in Computational Neuroscience*, *9*, 135.

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, *4*(5), 187–196.

Lindell, A. K. (2006). In your right mind: right hemisphere contributions to language processing and production. *Neuropsychology Review*, *16*(3), 131–148.

Lipkin, B., Tuckute, G., Affourtit, J., Small, H., Mineroff, Z., Kean, H., Jouravlev, O., Rakocevic, L., Pritchett, B., Siegelman, M., Hoeflin, C., Pongos, A., Blank, I. A., Struhl, M. K., Ivanova, A., Shannon, S., Sathe, A., Hoffmann, M., Nieto-Castañón, A., & Fedorenko, E. (2022). Probabilistic atlas for the language network based on precision fMRI data from >800 individuals. *Scientific Data*, *9*(1), 529.

Liu, L., Li, H., Ren, Z., Zhou, Q., Zhang, Y., Lu, C., Qiu, J., Chen, H., & Ding, G. (2022). The "Two-Brain" Approach Reveals the Active Role of Task-Deactivated Default Mode Network in Speech Comprehension. *Cerebral Cortex (New York, N.Y.: 1991)*, bhab521.

Menenti, L., Pickering, M. J., & Garrod, S. C. (2012). Toward a neural basis of interactive alignment in conversation. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00185

Meshulam, M., Hasenfratz, L., Hillman, H., Liu, Y.-F., Nguyen, M., Norman, K. A., & Hasson, U. (2021). Neural alignment predicts learning outcomes in students taking an introduction to

computer science course. *Nature Communications*, *12*(1), 1922.

Montague, P. R., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., Wiest, M. C., Karpov, I., King, R. D., & Apple, N. (2002). Hyperscanning: simultaneous fMRI during linked social interactions. *Neuroimage*, *16*(4), 1159–1164.

Nam, C. S., Choo, S., Huang, J., & Park, J. (2020). Brain-to-Brain Neural Synchrony During Social Interactions: A Systematic Review on Hyperscanning Studies. *Applied Sciences (Basel, Switzerland)*, *10*(19), 6669.

Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *NeuroImage*, *56*(2), 400–410.

Nastase, S. A., Gazzola, V., Hasson, U., & Keysers, C. (2019). Measuring shared responses across subjects using intersubject correlation. *Social Cognitive and Affective Neuroscience*, nsz037.

Nastase, S. A., Goldstein, A., & Hasson, U. (2020). Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage*, *222*, 117254.

Nastase, S. A., Liu, Y.-F., Hillman, H., Zadbood, A., Hasenfratz, L., Keshavarzian, N., Chen, J., Honey, C. J., Yeshurun, Y., Regev, M., Nguyen, M., Chang, C. H. C., Baldassano, C., Lositsky, O., Simony, E., Chow, M. A., Leong, Y. C., Brooks, P. P., Micciche, E., … Hasson, U. (2021). The “Narratives” fMRI dataset for evaluating models of naturalistic language comprehension. *Scientific Data*, *8*(1), 250.

Nguyen, M., Chang, A., Micciche, E., Meshulam, M., Nastase, S. A., & Hasson, U. (2022). Teacher–student neural coupling during teaching and learning. *Social Cognitive and Affective Neuroscience*, *17*(4), 367–376.

Nunez-Elizalde, A. O., Huth, A. G., & Gallant, J. L. (2019). Voxelwise encoding models with non-spherical multivariate normal priors. *NeuroImage*, *197*, 482–492.

Parkes, L., Fulcher, B., Yücel, M., & Fornito, A. (2018). An evaluation of the efficacy, reliability, and sensitivity of motion correction strategies for resting-state functional MRI. *NeuroImage*, *171*, 415–436.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research: JMLR*, *12*, 2825–2830.

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

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *The Behavioral and Brain Sciences*, *36*(4), 329–347.

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, *84*(Supplement C), 320–341.

Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the New York Academy of Sciences*, *1191*(1), 62–88.

Pulvermüller, F. (2018). Neural reuse of action perception circuits for language, concepts and communication. *Progress in Neurobiology*, *160*, 1–44.

Pulvermüller, F., & Fadiga, L. (2016). Brain language mechanisms built on action and perception. *Neurobiology of Language*, 311–324.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *GPT-2 Language Models are Unsupervised Multitask Learners*. 24.

Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews. Neuroscience*, *20*(8), 495–505.

Robertson, D. A., Gernsbacher, M. A., Guidotti, S. J., Robertson, R. R., Irwin, W., Mock, B. J., & Campana, M. E. (2000). Functional neuroanatomy of the cognitive process of mapping during discourse comprehension. *Psychological Science*, *11*(3), 255–260.

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the

preprocessing of resting-state functional connectivity data. *NeuroImage*, *64*(1), 240–256.

Saxe, R. (2006). Uniquely human social cognition. *Current Opinion in Neurobiology*, *16*(2), 235–239.

Schippers, M. B., Roebroeck, A., Renken, R., Nanetti, L., & Keysers, C. (2010). Mapping the information flow from one brain to another during gestural communication. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(20), 9388–9393.

Schrimpf, M., Blank, I. A., Tuckute, G., Kauf, C., Hosseini, E. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2021). The neural architecture of language: Integrative modeling converges on predictive processing. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(45), e2105646118.

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

Speer, S. P. H., Mwilambwe-Tshilobo, L., Tsoi, L., Burns, S. M., Falk, E. B., & Tamir, D. I. (2024). Hyperscanning shows friends explore and strangers converge in conversation. *Nature Communications*, *15*(1), 7781.

Spiegelhalder, K., Ohlendorf, S., Regen, W., Feige, B., Tebartz Van Elst, L., Weiller, C., Hennig, J., Berger, M., & Tüscher, O. (2014). Interindividual synchronization of brain activity during live verbal communication. *Behavioural Brain Research*, *258*, 75–79.

Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(32), 14425–14430.

Tognoli, E., Lagarde, J., DeGuzman, G. C., & Kelso, J. A. S. (2007). The phi complex as a neuromarker of human social coordination. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(19), 8190–8195.

Toneva, M., Williams, J., Bollu, A., Dann, C., & Wehbe, L. (2022). Same cause; Different effects in

the brain. In *arXiv [q-bio.NC]*. arXiv. http://arxiv.org/abs/2202.10376

Treiber, J. M., White, N. S., Steed, T. C., Bartsch, H., Holland, D., Farid, N., McDonald, C. R., Carter, B. S., Dale, A. M., & Chen, C. C. (2016). Characterization and Correction of Geometric Distortions in 814 Diffusion Weighted Images. *PLOS ONE*, *11*(3), e0152472.

Tsoi, L., Burns, S. M., Falk, E. B., & Tamir, D. I. (2022). The promises and pitfalls of functional magnetic resonance imaging hyperscanning for social interaction research. *Social and Personality Psychology Compass*. https://doi.org/10.1111/spc3.12707

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 Bias Correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320.

Vigneau, M., Beaucousin, V., Hervé, P.-Y., Jobard, G., Petit, L., Crivello, F., Mellet, E., Zago, L., Mazoyer, B., & Tzourio-Mazoyer, N. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *NeuroImage*, *54*(1), 577–593.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272.

Wang, S., Peterson, D. J., Gatenby, J. C., Li, W., Grabowski, T. J., & Madhyastha, T. M. (2017). Evaluation of Field Map and Nonlinear Registration Methods for Correction of Susceptibility Artifacts in Diffusion MRI. *Frontiers in Neuroinformatics*, *11*. https://doi.org/10.3389/fninf.2017.00017

Wehbe, L., Murphy, B., Talukdar, P., Fyshe, A., Ramdas, A., & Mitchell, T. (2014). Simultaneously Uncovering the Patterns of Brain Regions Involved in Different Story Reading Subprocesses. *PloS One*, *9*(11), e112575.

Wernicke, C. (1874). *Der aphasische Symptomencomplex: eine psychologische Studie auf*

anatomischer Basis. Cohn & Weigert.

Wheatley, T., Boncz, A., Toni, I., & Stolk, A. (2019). Beyond the Isolated Brain: The Promise and Challenge of Interacting Minds. *Neuron*, *103*(2), 186–188.

Wheatley, T., Thornton, M. A., Stolk, A., & Chang, L. J. (2024). The Emerging Science of Interacting Minds. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, *19*(2), 355–373.

Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, *31*(2), 183–194.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., Platen, P. von, Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., … Rush, A. M. (2020). *Transformers: State-of-the-Art Natural Language Processing*. Association for Computational Linguistics.

Yamashita, M., Kubo, R., & Nishimoto, S. (2023). *Cortical representations of languages during natural dialogue*. Neuroscience. http://biorxiv.org/lookup/doi/10.1101/2023.08.21.553821

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods*, *8*(8), 665–670.

Yarkoni, T., Speer, N. K., & Zacks, J. M. (2008). Neural substrates of narrative comprehension and memory. *NeuroImage*, *41*(4), 1408–1425.

Zada, Z., Goldstein, A., Michelmann, S., Simony, E., Price, A., Hasenfratz, L., Barham, E., Zadbood, A., Doyle, W., Friedman, D., Dugan, P., Melloni, L., Devore, S., Flinker, A., Devinsky, O., Nastase, S. A., & Hasson, U. (2024). A shared model-based linguistic space for transmitting our thoughts from brain to brain in natural conversations. *Neuron*. https://doi.org/10.1016/j.neuron.2024.06.025

Zadbood, A., Chen, J., Leong, Y. C., Norman, K. A., & Hasson, U. (2017). How We Transmit Memories to Other Brains: Constructing Shared Neural Representations Via Communication. *Cerebral Cortex (New York, N.Y.: 1991)*, *27*(10), 4988–5000.

Zadbood, A., Nastase, S., Chen, J., Norman, K. A., & Hasson, U. (2022). Neural representations of naturalistic events are updated as our understanding of the past changes. *eLife*, *11*, e79045.

Zaki, J., & Ochsner, K. (2009). The need for a cognitive neuroscience of naturalistic social cognition. *Annals of the New York Academy of Sciences*, *1167*(1), 16–30.

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57.