



OPEN

DATA DESCRIPTOR

# The “Podcast” ECoG dataset for modeling neural activity during natural language comprehension

Zaid Zada<sup>1</sup>✉, Samuel A. Nastase<sup>1</sup>, Bobbi Aubrey<sup>1</sup>, Itamar Jalon<sup>1</sup>, Sebastian Michelmann<sup>1</sup>, Haocheng Wang<sup>1</sup>, Liat Hasenfratz<sup>1</sup>, Werner Doyle<sup>3</sup>, Daniel Friedman<sup>3</sup>, Patricia Dugan<sup>3</sup>, Lucia Melloni<sup>3</sup>, Sasha Devore<sup>3</sup>, Adeen Flinker<sup>3,4</sup>, Orrin Devinsky<sup>3</sup>, Ariel Goldstein<sup>2</sup> & Uri Hasson<sup>1</sup>

Naturalistic electrocorticography (ECoG) data are a rare but essential resource for studying the brain’s linguistic capabilities. ECoG offers high temporal resolution suitable for investigating processes at multiple temporal timescales and frequency bands. It also provides broad spatial coverage, often along critical language areas. Here, we share a dataset of nine ECoG participants with 1,330 electrodes listening to a 30-minute audio podcast. The richness of this naturalistic stimulus can be used for various research questions, from auditory perception to narrative integration. In addition to the neural data, we extracted linguistic features of the stimulus ranging from phonetic information to large language model word embeddings. We use these linguistic features in encoding models that relate stimulus properties to neural activity. Finally, we provide detailed tutorials for preprocessing raw data, extracting stimulus features, and running encoding analyses that can serve as a pedagogical resource or a springboard for new research.

## Background & Summary

We introduce the “Podcast” electrocorticography (ECoG) dataset for modeling neural activity supporting natural spoken narrative comprehension. This dataset combines the exceptional spatiotemporal resolution of human intracranial electrophysiology with a naturalistic experimental paradigm for language comprehension. In addition to the raw data, we provide a minimally preprocessed version in the high-gamma spectral band to showcase a simple pipeline and to make the dataset easier to use. Furthermore, we include the auditory stimuli, an aligned word-level transcript, and linguistic features ranging from low-level acoustic properties to large language model (LLM) embeddings. We also include tutorials replicating previous findings to serve as a pedagogical resource and a springboard for new research. The dataset comprises 9 participants with 1,330 electrodes, including grid, depth, and strip electrodes. The participants listened to a 30-minute story with over 5,000 words. By using a natural story with high-fidelity, invasive neural recordings, this dataset offers a unique opportunity to investigate language comprehension.

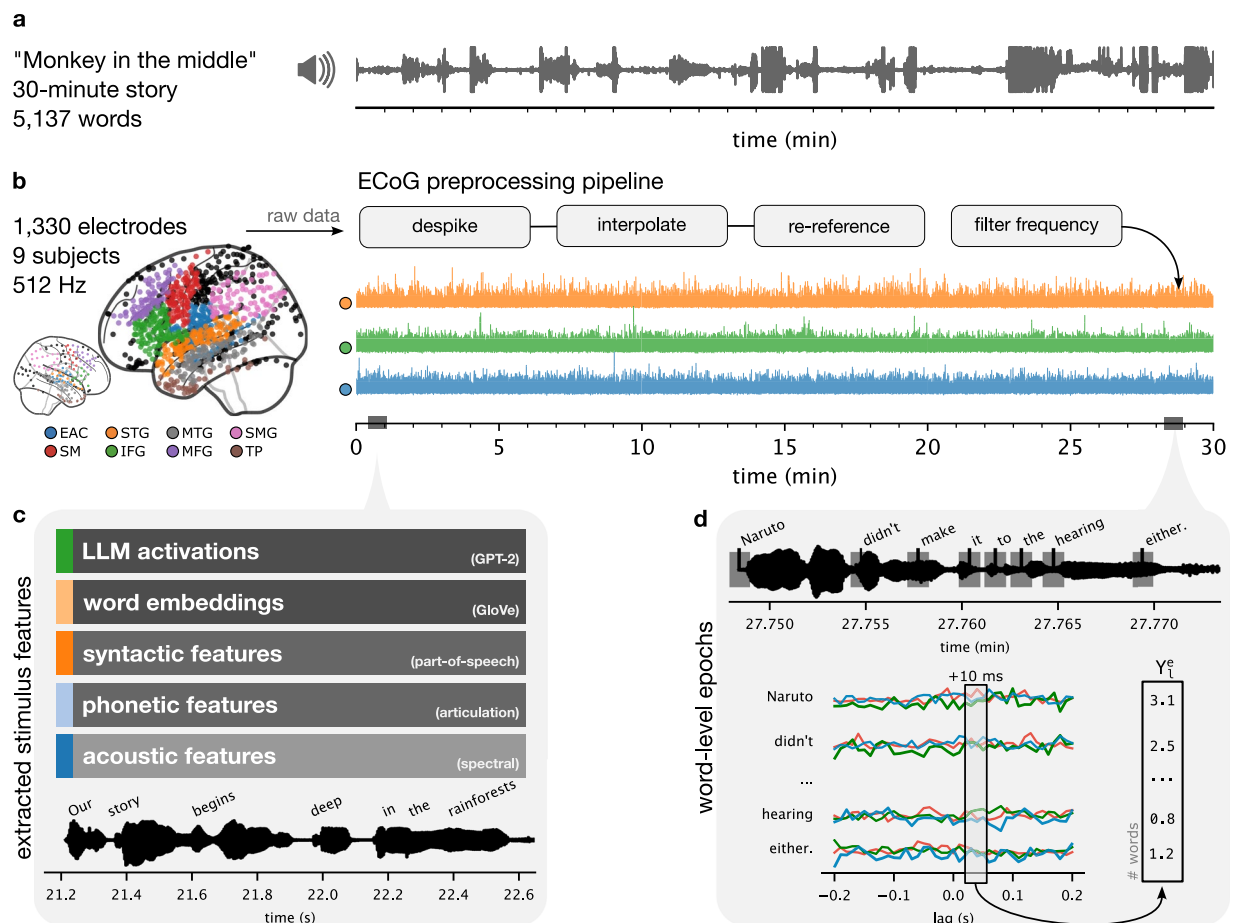
Language has a rich history of study in many scientific fields. Historically, research on the neural basis of language used highly controlled experimental stimuli to target particular linguistic phenomena, e.g., isolated words and sentences varying along a specific linguistic feature<sup>1,2</sup>. The past decade has seen a push for a more ecological, holistic account of the neural basis of language processing<sup>3–5</sup>. Yet naturalistic neural data is more challenging to model and understand than, for example, contrast-based experimental paradigms. Researchers have used data-driven methods like intersubject correlation (ISC) to isolate stimulus-driven processing in naturalistic contexts, sometimes under different conditions<sup>6,7</sup>. However, data-driven methods like ISC do not allow us to test various models for the neural computations driving language processing<sup>8–10</sup>.

Linear regression encoding models serve as a model-based analysis framework that maps linguistic and stimulus features to neural data<sup>11–13</sup>. These models learn a direct mapping from stimuli properties (e.g., word embeddings) to neural activity. With hundreds of regressors, regularization is often employed through PCA or a ridge

<sup>1</sup>Princeton Neuroscience Institute and Department of Psychology, Princeton University, New Jersey, 08544, USA.

<sup>2</sup>Department of Cognitive and Brain Sciences and Business School, Hebrew University, Jerusalem, 9190501, Israel.

<sup>3</sup>Grossman School of Medicine, New York University, New York, 10016, USA. <sup>4</sup>Tandon School of Engineering, New York University, New York, 10016, USA. ✉e-mail: [zzada@princeton.edu](mailto:zzada@princeton.edu)



**Fig. 1** Experiment setup and dataset components. **(a)** A 30-minute audio story (podcast) was played to **(b)** nine participants while undergoing electrocorticographic monitoring for epilepsy. We implement a simple preprocessing pipeline to extract the high-gamma band power per electrode. **(c)** We manually transcribed the story and timestamped words at high temporal resolution. We extract five linguistic features from the audio waveform and transcript. **(d)** Some analyses depend on word-level epochs, where a window relative to word onset at each word is then stacked across words. This results in a predictor matrix of words by linguistic features and a target vector across words for each lag and each electrode. In electrode-wise encoding analysis, we use linear regression to predict the neural activity at a particular lag across words, separately for each electrode, depicted as a Y vector in **(d)**, from the stimulus features defined in **(c)**.

penalty<sup>14</sup>. To evaluate their performance, encoding models are trained on a subset of data, then model-based predictions are correlated with actual neural data for a held-out test set of data. If the prediction accuracy of the held-out test is high, we conclude that the neural activity encodes stimulus features represented by the model features. The encoding framework allows us to compare different stimulus properties, where different stimulus features act as hypotheses about the neural activity's underlying function or representation.

With the current dataset, we offer five distinct models for speech and language features to evaluate on the ECoG dataset: (1) Low-level sensory features are expressed as 80 mel-spectral power bins derived from the raw audio waveform. (2) Phoneme-level speech units are defined by decomposing each word into its constituent phonemes and represented using a 44-dimensional binary vector indicating the presence of a phoneme. In addition, we group consonants into the manner of articulation (e.g., nasal, plosive) and place of articulation (e.g., dental, palatal), based on de Heer and colleagues<sup>15</sup>. (3) Syntactic properties constituting 50 part-of-speech tags (e.g., nouns, verbs, adjectives, etc.), and 45 syntactic dependencies based on sentence parse trees codified as binary vectors. (4) Non-contextual word embeddings from pre-trained language models (e.g., GloVe<sup>16</sup>) representing each word as a 300-dimensional vector representation learned from large corpora of natural text. Notably, each word is assigned the same vector regardless of its context. (5) Contextual word embeddings from a large language model (LLM) where each word vector is influenced by all preceding words within its contextual window. For example, we used GPT-2 XL<sup>17</sup>, where word embeddings are 1,600-dimensional vectors. This tutorial demonstrates how to use these five models within a model comparison framework to predict neural activity, testing the effectiveness of competing models in capturing neural processes during natural language processing across different electrodes and brain areas.

The model comparison analyses showed that contextual word embeddings extracted from GPT-2 XL accounted for most of the variance across nearly all the electrodes tested in this dataset. These findings align with recent work from our lab and others that find alignment between the neural activity in the human brain and the internal activity



in LLMs during the processing of natural language<sup>18,19</sup>. For example, researchers have used encoding models to investigate predictive<sup>10,20</sup> and hierarchical linguistic processing<sup>21,22</sup>; the shared underlying geometry<sup>23,24</sup>; the similarity between reading and listening<sup>25</sup>; the role of speech in language comprehension<sup>26</sup>; and production-comprehension coupling between speakers<sup>9</sup>. Recent reviews provide a more comprehensive view of this growing body of work<sup>27–30</sup>.

The following describes the “Podcast” ECoG dataset, including data acquisition and preprocessing details (Fig. 1). Next, we provide two quality control analyses to ensure the neural activity is precisely aligned with the onset of words in the podcast stimulus. Finally, we use electrode-wise encoding analysis to test several linguistic features against the neural data.

## Methods

**Participants.** Nine participants underwent clinical intracranial monitoring in the New York University School of Medicine’s Comprehensive Epilepsy Center. Participants consented to be included in our study following approval by the New York University Langone Medical Center’s Institutional Review Board (IRB committee number s14-02101). Clinicians determined electrode location and type per participant, based solely on the clinical need of the participant—without regard to any research study. Several standard behavioral tests were conducted to assess the patient’s memory and linguistic abilities. These include the Verbal Comprehension Index (VCI), Perceptual Organization Index (POI), Processing Speed Index (PSI), and Working Memory Index (WMI). Participant data is detailed in Table 1.

Electrodes were made of platinum-iridium and were embedded in silastic sheets, with 2.3 mm diameter contacts (Ad-Tech Medical Instrument). Electrodes were grouped into three types: grid arrays, linear strips, or depth electrodes. Grid arrays consisted of  $8 \times 8$  contacts with 5 or 10 mm center-to-center spacing. Three participants (sub-03, sub-06, sub-09) consented to a hybrid grid: a standard electrode grid augmented with additional electrodes between the standard contacts (i.e., a U.S. Food and Drug Administration-approved hybrid clinical research grid). In addition to the 64 standard contacts (2 mm diameter), the hybrid grid interleaved 64 additional contacts with 1 mm in diameter (these electrodes are labeled as the type “EG” in the dataset).

**Electrode localization.** Each participant underwent a pre-surgical and post-surgical T1-weighted MRI scan. These two images were aligned following published methods<sup>31</sup>. All accompanying anatomical scans are anonymized with *pydeface*<sup>32</sup>. Electrodes were then localized on the cortical surface from a co-registered MRI or computed tomography scan. Montreal Neurological Institute (MNI) electrode coordinates were obtained by nonlinearly registering the skull-stripped T1 images to the MNI152 template. Not all electrodes were on the brain; for example, some electrodes may be on the skull. We identified 80 electrodes that were not localized and were marked accordingly in the dataset.

**ECoG data acquisition.** Electrodes were connected to one of two amplifier types: a NicoletOne C64 clinical amplifier (Natus Neurologics) that is band-pass filtered from 0.16–250 Hz, and digitized at 512 Hz; or a NeuroWorks Quantum Amplifier that is high-pass filtered at 0.01 Hz and recorded at 2,048 Hz. All signals were referenced to a two-contact subdural strip facing toward the skull near the craniotomy site.

**Stimuli.** The stimulus presented to participants was a segment from the podcast “This American Life” entitled “So a Monkey and a Horse Walk Into a Bar: Act One, Monkey in the Middle” released on November 10, 2017. The original audio and transcript are freely available online (<https://www.thisamericanlife.org/631/transcript>). We used this transcript as initial input to the Penn Phonetics Lab Forced Aligner ([https://babel.ling.upenn.edu/phonetics/old\\_website\\_2015/p2fa/index.html](https://babel.ling.upenn.edu/phonetics/old_website_2015/p2fa/index.html)) to estimate word onsets and offsets. Next, we manually verified and adjusted the timing for each word using a custom GUI interface (however, this step can also be performed with Praat or Audacity).

**ECoG preprocessing.** Raw electrode data underwent the following preprocessing pipeline: removing bad electrodes, downsampling, despiking and interpolating high-amplitude spikes, common average re-referencing, and notch filtering. First, we visualized the power spectrum density of each electrode per subject. From this, we were able to annotate unusual electrodes that did not conform to the expected  $1/f$  pattern, had a consistent oscillatory pattern, or showed other unusual artifacts. We found 31 such electrodes and marked them as “bad” (identified in the accompanying metadata). The source of these artifacts may be due to several factors, including excessive noise, epileptic activity, or poor contact. For data acquired with a sampling rate greater than 512 Hz, we downsampled to 512 Hz to match the sampling rate across subjects. We then applied a despiking and interpolation procedure to remove time points that exceeded four quartiles above the median of the signal and refill it using *pchip* interpolation. For re-referencing, we subtracted the mean signal across all electrodes per subject from each of their individual electrode time series. Finally, we used notch filters at 60, 120, 180, and 240 Hz to remove power line noise. This pipeline produces a “cleaned” version of the raw signal.

Linguistic processing may span multiple frequency bands and rely on cross-frequency coupling<sup>33,34</sup>. For the sake of simplicity, we focus on high-gamma band power as an index of local, stimulus-driven neuronal activity<sup>35–37</sup>. We extracted the high-gamma band by applying a Butterworth band-pass infinite impulse response (IIR) filter at 70–200 Hz. We extract the broadband power by computing the envelope of the Hilbert transform.

## Data Records

The dataset is freely available on OpenNeuro: <https://openneuro.org/datasets/ds005574><sup>38</sup>. Tutorials for usage are available at: <https://hassonlab.github.io/podcast-ecog-tutorials>. We followed the BIDS-iEEG<sup>39,40</sup> standard for file structures and naming conventions. The main directory contains the raw ECoG data for each subject in EDF format under the *ieeg* datatype directory. In addition, channel information and electrode localization for MNI

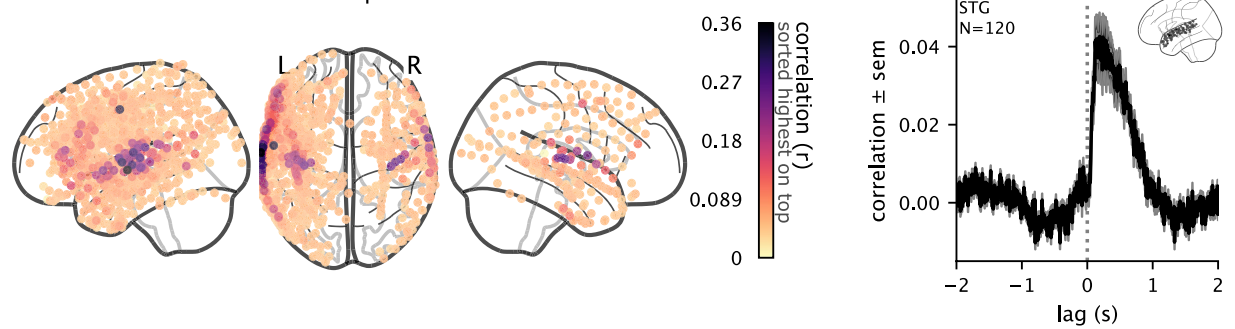
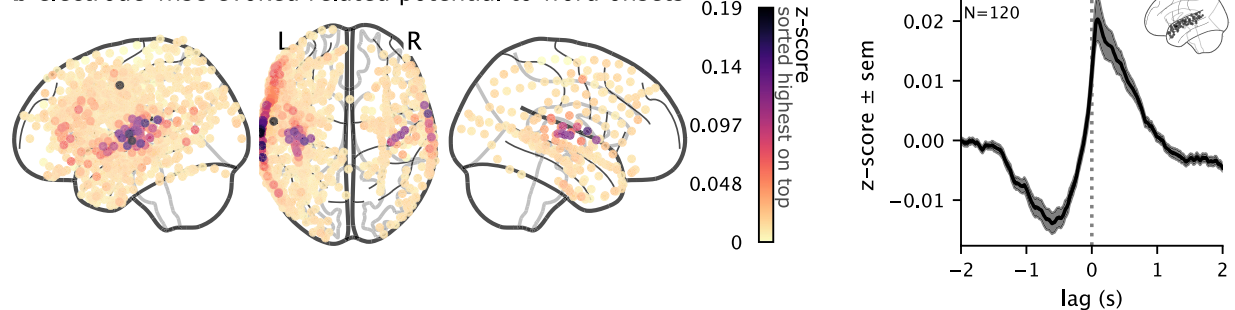
sub	demographic				behavioral tests				electrode coverage					
	sex	age	hand	race	VCI	POI	PSI	WMI	total	LH	RH	Grid	Strip	Depth
1	M	28	R	White	134	127	124	119	103	n/a	103	63	24	16
Diagnosis: Focal epilepsy arising from the right mid-temporal and to a lesser extent anterior temporal neocortical structures and right mesial temporal structures. There was also evidence of multifocal right hemispheric cortical hyperexcitability in the aforementioned regions as well as the right posterior temporal neocortex. Additionally, there is evidence of right temporal focal cerebral dysfunction.														
2	F	40	R	1+	63	79	84	89	95	95	n/a	n/a	43	52
Diagnosis: Hyperexcitability within the left insula (maximal in the anterior insula) with posterior insula and frontal lobe involvement. A single focal to bilateral tonic clonic seizure, consistent with the patient's targeted semiology. The first discernible electrographic changes were detected within the left anterior temporal resection cavity (DLT 1-2); this ictal pattern was most consistent with spread or volume conduction. There was then robust spread and evolution within the left insula, which also coincided with the first clear changes in clinical behavior.														
3	F	48	R	White	107	104	111	114	264	213	51	127	65	72
Diagnosis: Five focal unaware seizures were captured arising from the right temporal region, three from the right temporal pole and two from the right hippocampus. There were bilateral independent temporal epileptiform discharges, more over the left, at times in a lateralized periodic fashion on the left, indicating cortical hyperexcitability in these regions. These findings indicate right mesial and polar temporal seizure onsets, but does not exclude left temporal seizure onsets, especially with abundance of epileptiform discharges in left temporal region.														
4	M	25	R	White	n/a	n/a	n/a	n/a	155	77	78	n/a	79	76
Diagnosis: Focal cortical hyperexcitability in the bilateral mesial temporal regions independently. Habitual seizures were only captured in the setting of cortical stimulation; these were provoked by cortical stimulation of the right and left mesial temporal lobes independently.														
5	M	24	L > R	White	96	79	81	86	160	159	1	110	8	42
Diagnosis: Left hemispheric multilobar epilepsy.														
6	F	34	R	n/a	n/a	n/a	n/a	n/a	171	171	n/a	127	8	36
Diagnosis: This is a normal intracranial EEG study. No clear interictal epileptiform abnormalities were observed, and no electrographic seizures were recorded.														
7	F	28	R	White	100	109	92	86	119	119	n/a	63	32	24
Diagnosis: Focal epilepsy arising from the left posterior insula and peri-opercular region at the frontotemporoparietal junction. There is evidence of an extensive epileptic network throughout the neocortical frontal, temporal and parietal lobe, along with the deeper regions of at least the temporal lobe and insula. The location of the ictal onset zone raised concern for involvement of eloquent cortex for language, motor and sensory function.														
8	M	42	R	n/a	87	123	97	89	75	75	n/a	47	8	20
Diagnosis: Targeted seizures in the left temporal lobe perilesionally and left posterior mesial temporal structures. Focal cortical hyperexcitability was also noted broadly in the occipital, temporal, mesial temporal and insular cortex.														
9	M	24	R	White	145	96	86	95	188	188	n/a	124	40	24
Diagnosis: Focal epilepsy arising from the left superior temporal gyrus, middle temporal gyrus, inferior/lateral postcentral gyrus, and basal temporal region, at times also including the inferior parietal lobule, inferior frontal gyrus, superior occipital lobe, mesial temporal region, and insula. Multiple habitual focal aware seizures and subclinical seizures were captured with broad ictal onsets from these areas, as well as a minority of seizures with more focal onsets in the superior and middle temporal-occipital areas. There was evidence of focal cortical hyperexcitability in the left superior temporal gyrus and posterior mesial temporal region, often with a broad field including the left middle frontal gyrus, precentral gyrus, inferior frontal gyrus, superior and middle temporal gyri, basal temporal region, anterior mesial temporal region, and insula. Additionally, there was evidence of focal cortical hyperexcitability in the left basal temporal region.														

**Table 1.** Participant electrode, demographic, and behavioral information. Acronyms: left hemisphere (LH); right hemisphere (RH); verbal comprehension index (VCI); perceptual organization index (POI); processing speed index (PSI); working memory index (WMI); and ambidextrous (ambi.).

and T1 spaces are also located in the same folder, following BIDS guidelines. Notably, the channels tsv file contained an annotation for each electrode in the EDF that denoted whether it was a “good” or “bad” channel. Bad channels are those that are rejected either due to “no localization” or noisy power spectrum density (see “ECoG preprocessing” in the “Methods” section). Each subject also has a de-faced T1 anatomical MRI scan under their respective *anat* datatype directory.

The *derivatives* directory contains the preprocessed ECoG data in MNE .*fif* format within the “ecogprep” subdirectory (inspired by fMRIPrep<sup>41</sup>). There are two versions of the preprocessed data, one is unfiltered and the other is filtered to the high-gamma broadband power. In addition, the “ecogqc” subdirectory contains the output of a quality-check program in HTML that contains an interactive viewer for electrodes and the power spectrum density plots (inspired by MRIQC<sup>42</sup>). These were the primary sources used for determining the quality of an electrode's signal.

The top-level *stimuli* directory contains the original audio presented to participants in .*wav* format, along with a time-stamped transcript in .*csv* format. Extracted data from each of the five feature spaces is placed in a subdirectory of *stimuli* denoting its type. For example, the *stimuli/spectral/* directory contains spectrogram features of the audio waveform. Each feature directory contains at least two files: *features.hdf5* and *transcript.tsv*. The *transcript* file is a modified version of the original transcript (under *stimuli*) where a particular feature type may remove or add additional rows or columns (e.g., for tokenization). The *features* file contains the numerical vectors of the feature space as a matrix, where the rows correspond one-to-one to the rows in the *transcript* file. The number of columns varies depending on the dimensionality of the feature space. In some cases, such as large language models, this file may include the activations of all layers in separate HDF5 groups in the same file. Examples of loading and manipulating these files can be found in the accompanying code and tutorials.

**a** electrode-wise audio envelope cross-correlation**b** electrode-wise evoked related potential to word onsets

**Fig. 2** Validation of brain activity responses to auditory stimuli. **(a)** We cross-correlated the high-gamma band ECoG signal with the envelope of the audio waveform to validate the alignment between the audio and the brain data, especially in the early auditory cortex. **(b)** For each electrode, we computed the average neural responses across words to verify that electrodes in the early auditory cortex exhibit increased word-related activity.

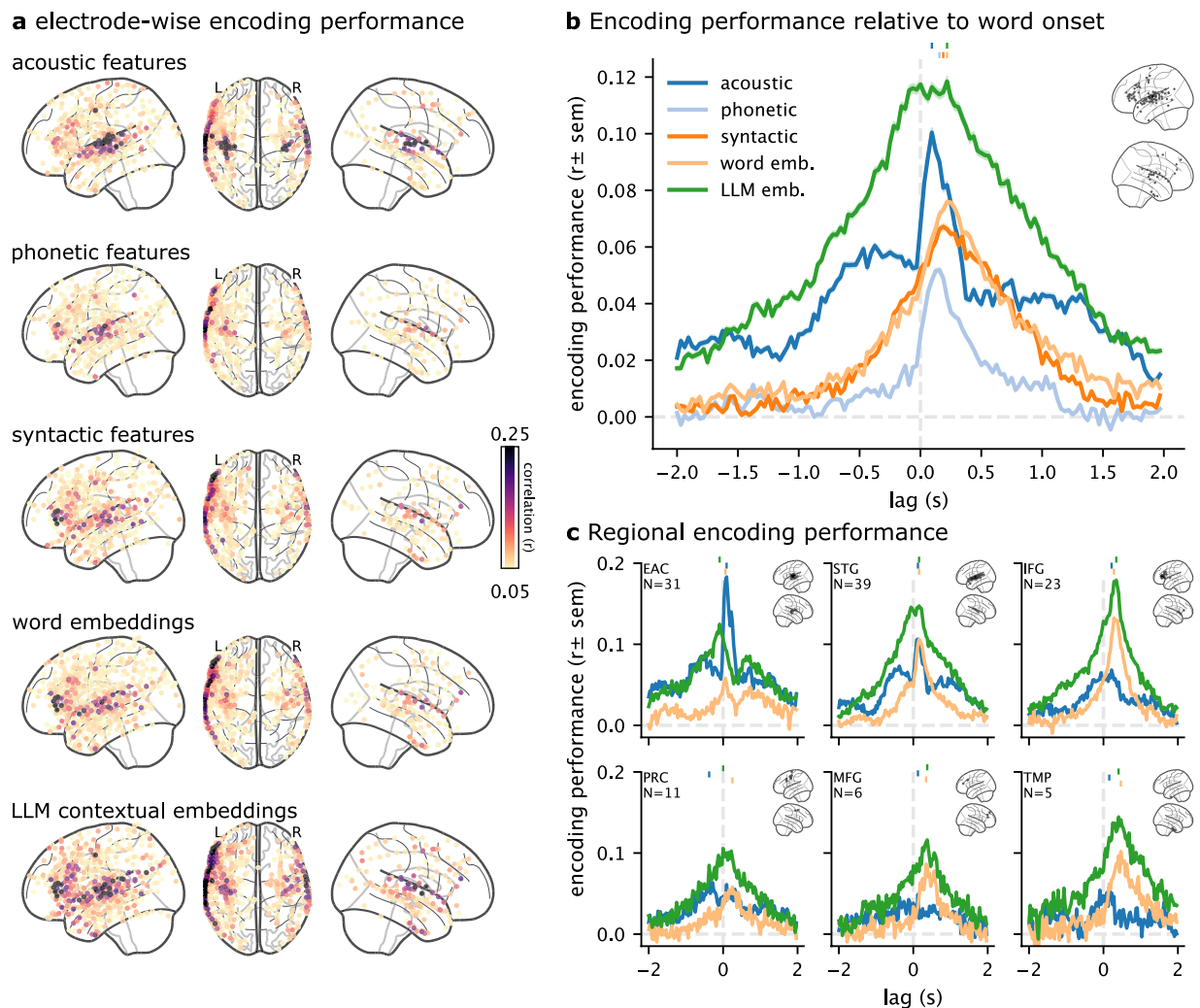
### Technical Validation

**Cross-correlation between electrode activity and auditory stimulus.** The first analysis tested electrodes that track the acoustic properties of the stimulus, roughly aligning the neural activity with the audio. We cross-correlated the stimulus audio envelope with each electrode's high-gamma broadband power. This analysis was used in Honey and colleagues<sup>43</sup> to find electrodes that tracked the acoustic properties of the stimulus. Similar to their results, we also found a gradient of audio correlations that peaks in electrodes closest to the early auditory cortex and decrements along the anterior-posterior axis (Fig. 2a). We also visualized the average cross-correlation of electrodes in the superior temporal cortex in a four-second window. This validation analysis reveals that brain activity lags just behind the auditory stimulus, demonstrating that the ECoG recordings are closely aligned with the stimulus.

**Electrode-wise evoked response to word onsets.** We used an event-related potential (ERP) analysis as a complementary method to confirm the transcript word timings and validate the neural signal. We used the timed transcript to epoch the neural data (Fig. 1d) and then averaged the neural signal in a four-second window around word onset. As expected, electrodes that are sensitive to word occurrence increased their activity from baseline (Fig. 2b). These electrodes were localized in the vicinity of the early auditory cortex in the superior temporal cortex. The average evoked response across electrodes in the superior temporal cortex exhibited a sharp increase in activity around word onset.

**Electrode-wise encoding models for multiple feature spaces.** We used electrode-wise encoding models to measure the neural responses to specific features of the podcast stimulus. We extracted five sets of linguistic features from the stimulus (Fig. 1c). First, we computed a spectrogram from the audio waveform based on 80 log mel-spectral bins. We then epoched the spectrogram according to word onsets from lag 0 ms to 200 ms and downsampled each epoch to a sampling rate of 10 Hz. Thus, each word was represented by 10 frames of 80 bins, flattened into an 800-dimensional vector. Second, we defined a phoneme-based word representation using a word-to-phoneme dictionary to create a 40-dimensional binary vector of phonemes present in a word. Third, we extracted syntactic features from the transcript using spaCy<sup>44</sup>. These features include 50 part-of-speech tags and 45 dependency relations, resulting in a 95-dimensional binary vector. Fourth, we used spaCy's *en\_core\_web\_lg* model to extract 95-dimensional non-contextual word embeddings (where the same word is assigned the same embedding regardless of context). Fifth and finally, we used HuggingFace<sup>45</sup> to extract LLM contextual embeddings from the activations of the middle layer of GPT-2-XL<sup>17</sup>. These feature spaces form the regressors we used to predict the ECoG data.

We epoched the high-gamma band electrode time series per word from  $-2$  seconds to  $+2$  seconds (Fig. 1d). Thus, the linear regression model predicts the word-by-word fluctuations in neural activity at a specific lag and electrode. We used two-fold cross-validation to train separate encoding models for each feature space, electrode,



**Fig. 3** Encoding model performance for five feature spaces. **(a)** Maximum encoding performance (correlation between model-predicted and actual neural activity) across lags at each electrode for five different feature spaces. Electrodes with significant performance ( $p < 0.01$ , FDR corrected) for any of the feature spaces are visualized. **(b)** Comparing encoding performance across five different feature spaces and lags relative to word onset. The average over all selected electrodes per feature space is plotted. **(c)** Encoding performance for the acoustic, non-contextual word embeddings, and LLM embeddings are visualized for six regions of interest: early auditory cortex (EAC), superior temporal gyrus (STG), inferior frontal gyrus (IFG), precentral gyrus (PRC), middle frontal gyrus (MFG), and temporal pole (TMP). Electrodes are assigned regions based on the Destrieux atlas<sup>51</sup>.

and lag on one half of the podcast stimulus. Since different feature spaces are of different dimensionality, we used ridge regression where the penalty hyperparameter is learned within the training set using the *himalaya* python library<sup>46</sup>. We evaluated the performance of each encoding model on the held-out fold by correlating the model-predicted neural activity with the actual neural activity across words in the test fold. Finally, we averaged the two correlations for each fold to obtain one correlation value denoting the encoding performance for each feature space, electrode, and lag.

We visualize encoding model performance both spatially (Fig. 3a) and temporally (Fig. 3b). We found that the acoustic and phonetic feature spaces performed well in the early auditory cortex (EAC) and mid-superior temporal gyrus (STG). Syntactic features and non-contextual embeddings performed well along the STG and extended into the inferior frontal gyrus (IFG). LLM contextual embeddings achieved the highest encoding performance and for the largest number of electrodes. The different feature spaces also yielded different temporal profiles across lags relative to word onset. For example, LLM embeddings had high encoding performance across the four-second window, whereas other feature spaces peaked shortly after word onset (Fig. 3b). In particular, the acoustic and phonetic features peaked sharply after word onset. We also observed certain temporal differences in specific regions. For example, LLM embeddings peaked sharply after word onset in IFG, but more broadly in STG (Fig. 3c). Moreover, acoustic features performed better than LLM embeddings in EAC, at and after word onset. Across all regions, LLM embeddings exhibit a broader predictive window, including before word onset. This observation likely results from the fact that LLMs integrate context (linguistic features



of previous words) into the current word embedding. The predictability of different features can be affected by the degree of autocorrelation in regressors. As can be seen in Fig. 2b, the tails of LLM and acoustic features are higher than the other features. For analyses requiring comparable baselines, we recommend addressing any potential differences in the regression variable autocorrelation. Altogether, these results validate our procedures for stimulus feature extraction, ECoG preprocessing, and temporal alignment between the ECoG signals and stimulus. Our findings also suggest that this dataset can serve as a useful resource for testing different models of language processing.

## Usage Notes

We include tutorial notebooks to serve as pedagogical examples for using the data and replicating the results presented in this paper. These tutorials are available online at <https://hassonlab.github.io/podcast-ecog-tutorials> and in the supplementary materials. We encourage researchers to familiarize themselves with the BIDS-iEEG<sup>40</sup> modality-specific standards to best use the dataset. We also encourage the use of Python-based, open-source MNE tools<sup>47</sup> for reading the data (e.g., the preprocessed .fif files and to facilitate analyses). The tutorials cover the following topics: downloading data and installing the required libraries, preprocessing ECoG data, running quality control analyses, extracting stimuli features, and performing encoding analyses.

In this paper, we introduced the “Podcast” ECoG dataset and replicated several previously published analyses. That said, many of these analyses include several simplifying choices and may overlook certain questions of interest. For example, we epoch the neural activity word-by-word and train separate models at different lags. Linguistic information may be encoded in finer-grained subword dynamics or dynamics evolving over the course of word articulation. Depending on specific research aims, one may bin the signal differently (here we use 200 ms bins) or explore a smaller or larger range of lags relative to word onset (here we analyzed windows ranging from −2 to +2 s). An alternative method may bypass epoching altogether and perform a time-resolved analysis similar to encoding models in fMRI<sup>48</sup>. Furthermore, we index our analyses to the onset of words. However, words have different lengths and internal structures; anchoring to the center of each word may provide a good compromise<sup>49</sup>. Models that aggregate features across words (e.g., sentence embeddings<sup>50</sup>) may also provide novel insights relative to our word-by-word analysis. In the technical validation section, we used a simple approach to aggregate electrodes into broader regions based on each electrode’s MNI coordinates and a widely used anatomical atlas<sup>51</sup>. For analyses with different requirements for localization, we recommend indexing electrode locations via the subject-specific electrode coordinates with the native-space T1w images. Our analyses use high-gamma band power as an index of neural activity. Linguistic information may also be encoded in oscillations at other frequency bands or in cross-frequency coupling<sup>52</sup>. Finally, we recommend using a variance partitioning analysis<sup>53</sup> to measure the unique contribution of a particular feature set, such as LLMs, relative to other features (see our tutorials for an example).

## Code availability

Data and code associated with data curation are available on OpenNeuro<sup>54</sup> (<https://openneuro.org/datasets/ds005574>)<sup>38</sup>. The code for the analyses in this paper is available on GitHub (<https://github.com/hassonlab/podcast-ecog-paper>). Tutorials are also available on GitHub (<https://hassonlab.github.io/podcast-ecog-tutorials>).

Received: 4 March 2025; Accepted: 24 June 2025;

Published online: 03 July 2025

## References

1. Friederici, A. D. The Brain Basis of Language Processing: From Structure to Function. *Physiol. Rev.* **91**, 1357–1392 (2011).
2. Price, C. J. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage* **62**, 816–847 (2012).
3. Hasson, U. & Honey, C. J. Future trends in Neuroimaging: Neural processes as expressed within real-life contexts. *Neuroimage* **62**, 1272–1278 (2012).
4. Hamilton, L. S. & Huth, A. G. The revolution will not be controlled: natural stimuli in speech neuroscience. *Lang. Cogn. Neurosci.* **35**, 573–582 (2020).
5. Nastase, S. A., Goldstein, A. & Hasson, U. Keep it real: rethinking the primacy of experimental control in cognitive neuroscience. *Neuroimage* **222**, 117254 (2020).
6. Hasson, U., Nir, Y., Levy, I., Fuhrmann, G. & Malach, R. Intersubject Synchronization of Cortical Activity During Natural Vision. *Science* **303**, 1634–1640 (2004).
7. Nastase, S. A., Gazzola, V., Hasson, U. & Keysers, C. Measuring shared responses across subjects using intersubject correlation. *Soc. Cogn. Affect. Neurosci.* nsz037 <https://doi.org/10.1093/scan/nsz037> (2019).
8. Yarkoni, T. & Westfall, J. Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspect. Psychol. Sci.* **12**, 1100–1122 (2017).
9. Zada, Z. *et al.* 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> (2024).
10. Goldstein, A. *et al.* Shared computational principles for language processing in humans and deep language models. *Nat. Neurosci.* **25**, 369–380 (2022).
11. Naselaris, T., Kay, K. N., Nishimoto, S. & Gallant, J. L. Encoding and decoding in fMRI. *Neuroimage* **56**, 400–410 (2011).
12. Holdgraf, C. *et al.* Encoding and Decoding Models in Cognitive Electrophysiology. *Front. Syst. Neurosci.* **11**, 61 (2017).
13. Dupré la Tour, T., Visconti di Oleggio Castello, M. & Gallant, J. L. The Voxelwise Encoding Model framework: a tutorial introduction to fitting encoding models to fMRI data. *Imaging Neuroscience* [https://doi.org/10.1162/imag\\_a\\_00575](https://doi.org/10.1162/imag_a_00575) (2025).
14. Nunez-Elizalde, A. O., Huth, A. G. & Gallant, J. L. Voxelwise encoding models with non-spherical multivariate normal priors. *Neuroimage* **197**, 482–492 (2019).
15. de Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L. & Theunissen, F. E. The Hierarchical Cortical Organization of Human Speech Processing. *J. Neurosci.* **37**, 6539–6557 (2017).
16. Pennington, J., Socher, R. & Manning, C. *Glove: Global Vectors for Word Representation*. <https://doi.org/10.3115/v1/D14-1162> (Association for Computational Linguistics, 2014).
17. Radford, A. *et al.* GPT-2 Language Models are Unsupervised Multitask Learners. 24 (2019).



18. Richards, B. A. *et al.* A deep learning framework for neuroscience. *Nat. Neurosci.* **22**, 1761–1770 (2019).
19. Hasson, U., Nastase, S. A. & Goldstein, A. Direct Fit to Nature: An Evolutionary Perspective on Biological and Artificial Neural Networks. *Neuron* **105**, 416–434 (2020).
20. Schrimpf, M. *et al.* The neural architecture of language: Integrative modeling converges on predictive processing. *Proc. Natl. Acad. Sci. USA* **118**, e2105646118 (2021).
21. Heilbron, M., Armeni, K., Schoffelen, J.-M., Hagoort, P. & De Lange, F. P. A hierarchy of linguistic predictions during natural language comprehension. *Proc. Natl. Acad. Sci. USA* **119**, e2201968119 (2022).
22. Caucheteux, C., Gramfort, A. & King, J.-R. Evidence of a predictive coding hierarchy in the human brain listening to speech. *Nat. Hum. Behav.* **7**, 430–441 (2023).
23. Goldstein, A. *et al.* Alignment of brain embeddings and artificial contextual embeddings in natural language points to common geometric patterns. *Nat. Commun.* **15**, 2768 (2024).
24. Bhattacharjee, A. *et al.* Aligning Brains into a Shared Space Improves their Alignment to Large Language Models. Preprint at <https://doi.org/10.1101/2024.06.04.597448> (2024).
25. Deniz, F., Nunez-Elizalde, A. O., Huth, A. G. & Gallant, J. L. The Representation of Semantic Information Across Human Cerebral Cortex During Listening Versus Reading Is Invariant to Stimulus Modality. *J. Neurosci.* **39**, 7722–7736 (2019).
26. Goldstein, A. *et al.* A unified acoustic-to-speech-to-language embedding space captures the neural basis of natural language processing in everyday conversations. *Nat. Hum. Behav.* <https://doi.org/10.1038/s41562-025-02105-9> (2025).
27. Arana, S., Pesnot Lerousseau, J. & Hagoort, P. Deep learning models to study sentence comprehension in the human brain. *Lang. Cogn. Neurosci.* **39**, 972–990 (2024).
28. Jain, S., Vo, V. A., Wehbe, L. & Huth, A. G. Computational language modeling and the promise of *in silico* experimentation. *Neurobiol. Lang. (Camb.)* **5**, 80–106 (2024).
29. Lopopolo, A., Fedorenko, E., Levy, R. & Rabovsky, M. Cognitive computational neuroscience of language: Using computational models to investigate language processing in the brain. *Neurobiol. Lang. (Camb.)* **5**, 1–6 (2024).
30. Tuckute, G., Kanwisher, N. & Fedorenko, E. Language in brains, minds, and machines. *Annu. Rev. Neurosci.* **47**, 277–301 (2024).
31. Yang, A. I. *et al.* Localization of dense intracranial electrode arrays using magnetic resonance imaging. *Neuroimage* **63**, 157–165 (2012).
32. Gulban, O. F. *et al.* Poldracklab/pydeface: PyDeface v2.0.2. <https://doi.org/10.5281/ZENODO.6856482> (Zenodo, 2022).
33. Martin, A. E. A compositional neural architecture for language. *J. Cogn. Neurosci.* **32**, 1407–1427 (2020).
34. Murphy, E. ROSE: A neurocomputational architecture for syntax. *J. Neurolinguistics* **70**, 101180 (2024).
35. Mukamel, R. *et al.* Coupling Between Neuronal Firing, Field Potentials, and fMRI in Human Auditory Cortex. *Science* **309**, 951–954 (2005).
36. Manning, J. R., Jacobs, J., Fried, I. & Kahana, M. J. Broadband Shifts in Local Field Potential Power Spectra Are Correlated with Single-Neuron Spiking in Humans. *J. Neurosci.* **29**, 13613–13620 (2009).
37. Haufe, S. *et al.* Elucidating relations between fMRI, ECoG, and EEG through a common natural stimulus. *Neuroimage* **179**, 79–91 (2018).
38. Zada, Z. *et al.* The ‘Podcast’ ECoG dataset for modeling neural activity during natural story listening. *Openneuro* <https://doi.org/10.18112/OPENNEURO.DS005574.V1.0.2> (2025).
39. Gorgolewski, K. J. *et al.* The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data* **3**, 160044 (2016).
40. Holdgraf, C. *et al.* iEEG-BIDS, extending the Brain Imaging Data Structure specification to human intracranial electrophysiology. *Sci. Data* **6**, 102 (2019).
41. Esteban, O. *et al.* fMRIPrep: a robust preprocessing pipeline for functional MRI. *Nat. Methods* **16**, 111–116 (2019).
42. Esteban, O. *et al.* MRIQC: Advancing the automatic prediction of image quality in MRI from unseen sites. *PLoS One* **12**, e0184661 (2017).
43. Honey, C. J. *et al.* Slow Cortical Dynamics and the Accumulation of Information over Long Timescales. *Neuron* **76**, 423–434 (2012).
44. Honnibal, M., Montani, I., Van Landeghem, S. & Boyd, A. spaCy: Industrial-strength Natural Language Processing in Python <https://doi.org/10.5281/zenodo.1212303> (2020).
45. Wolf, T. *et al.* Transformers: State-of-the-Art Natural Language Processing. <https://doi.org/10.18653/v1/2020.emnlp-demos.6> (Association for Computational Linguistics, 2020).
46. Dupré La Tour, T., Eickenberg, M., Nunez-Elizalde, A. O. & Gallant, J. L. Feature-space selection with banded ridge regression. *Neuroimage* **264**, 119728 (2022).
47. Gramfort, A. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience* **7** (2013).
48. Huth, A. G., de Heer, W. A., Griffiths, T. L., Theunissen, F. E. & Gallant, J. L. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature* **532**, 453–458 (2016).
49. Mischler, G., Li, Y. A., Bickel, S., Mehta, A. D. & Mesgarani, N. Contextual feature extraction hierarchies converge in large language models and the brain. *Nat. Mach. Intell.* **6**, 1467–1477 (2024).
50. Muennighoff, N., Tazi, N., Magne, L. & Reimers, N. MTEB: Massive Text Embedding Benchmark. *arXiv [cs.CL]* (2022).
51. Destrieux, C., Fischl, B., Dale, A. & Halgren, E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* **53**, 1–15 (2010).
52. Weissbart, H. & Martin, A. E. The structure and statistics of language jointly shape cross-frequency neural dynamics during spoken language comprehension. *Nat. Commun.* **15**, 8850 (2024).
53. Lescroart, M. D., Stansbury, D. E. & Gallant, J. L. Fourier power, subjective distance, and object categories all provide plausible models of BOLD responses in scene-selective visual areas. *Front. Comput. Neurosci.* **9**, 135 (2015).
54. Markiewicz, C. J. *et al.* The OpenNeuro resource for sharing of neuroscience data. *Elife* **10** (2021).

## Acknowledgements

We thank Catherine Chen and Amir Khalilian for the discussion regarding ECoG preprocessing. This work was supported by the National Institutes of Health grants R01DC022534 (U.H.) and DP1HD091948 (A.F.).

## Author contributions

Z.Z. curated, validated, visualized, and analyzed the data. Z.Z., A.G., I.J., S.M. designed methodologies used for the analyses. B.A., W.D., D.F., P.D., L.M., S.D., A.F. and O.D. provided resources and curated the initial data. U.H., L.H. and A.F. acquired funding. Z.Z. wrote the software for the analyses and tutorials; H.W. wrote software for the tutorials and validated them. S.A.N. and U.H. administered the project. Z.Z. wrote the original draft and S.A.N. and U.H. reviewed and edited the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to Z.Z.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025