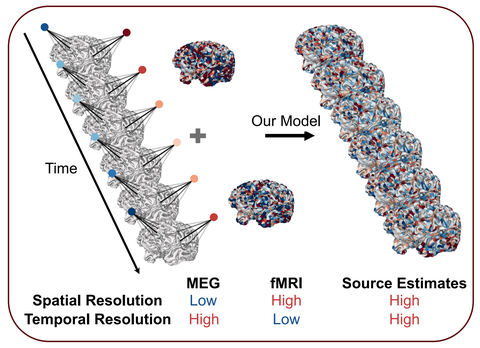
MEG–fMRI Integration in Naturalistic Speech Comprehension for High Spatial and Temporal Resolution

Beige Jerry Jin & Leila Wehbe
Carnegie Mellon University

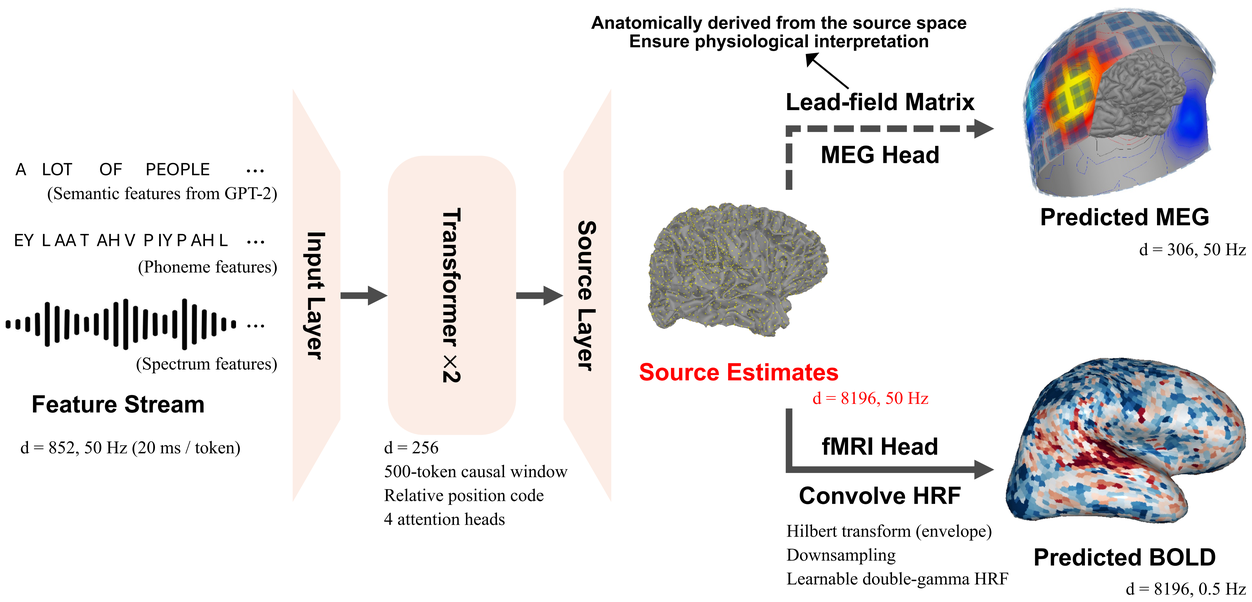
INTRODUCTION

- **The time-space trade-off**: MEG can capture rapid neural dynamics but lacks spatial detail, while fMRI can spatially localize brain activity but lags behind neural activity. A unified picture that preserves *both* high resolutions remains elusive with existing source localization or MEG-fMRI fusion methods, especially for **single-trial naturalistic data**.
- We collect MEG when participants listened passively to **27 narrative stories** (more than 7 hours in total); each also completed a fMRI scan on repeated presentations of one anchor story. For the remaining stories, we design a pipeline to project an open fMRI dataset collected on identical stimuli (LeBel et al., 2023) onto each participant's cortical surface.



- We propose **a transformer-based encoding model** that predicts MEG and fMRI signals as a function of the stimulus features but constrains these signals to be simultaneously generated by **source estimates** in a latent source space, thus effectively estimating the source activity that is high-resolution in both time and space.

MODEL



Training:

$$L = \alpha_1 \left(1 - \mathrm{corr}(\mathbf{m}_t, \hat{\mathbf{m}}_t)\right) + \alpha_2 \left(1 - \mathrm{corr}(\mathbf{y}_\tau, \hat{\mathbf{y}}_\tau)\right) + \alpha_3 L_{\text{smooth}}$$

true MEG (denoised) · pred MEG · true fMRI (denoised) · pred fMRI · smoothness of source estimates

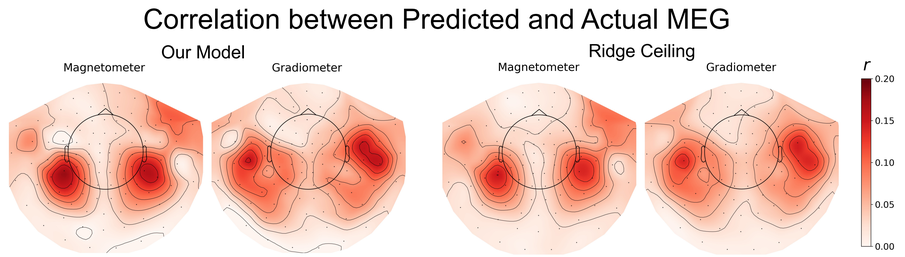
- MEG-first curriculum: $\alpha_1 = 1$, $\alpha_2 = 0$, $\alpha_3 = 0.0001$ for first 30 epochs; $\alpha_1 = 1$, $\alpha_2 = 1$, $\alpha_3 = 0.0001$ for later epochs
- 21 training stories, 1 validation story for early stopping
- Separate model for different subjects

SUMMARY

- We propose a transformer-based encoding model that recovers stimulus-relevant source estimates constrained jointly by MEG and fMRI, which matches single-modality regression-based methods in MEG and fMRI prediction, outperforms classic source localization methods, and generalizes to unseen ECoG data.

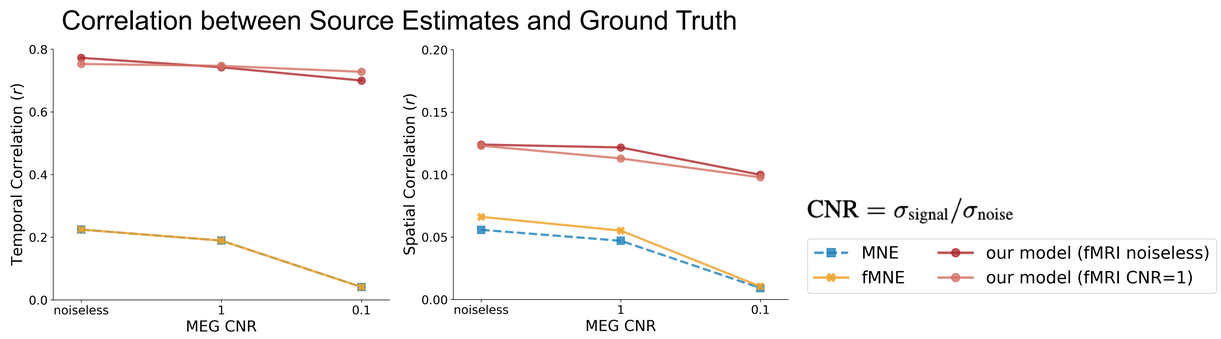
RESULTS

Predictive Performance



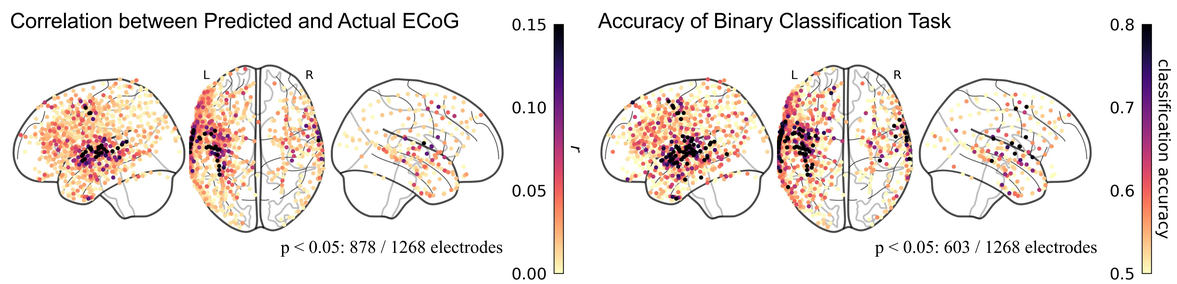
- Ridge Ceilings (**single-modality linear encoding models**) are less constrained, yet unable to estimate source activity.
- Our model predicts MEG and fMRI of 4 held-out test stories on par with Ridge Ceilings.

Simulation Experiments



- We generate simulated source activity and corresponding MEG and fMRI signal at different noise levels (CNR).
- Our model recovers source activity better than minimum norm estimate (MNE) and fMRI-weighted MNE (fMNE).

Zero-shot ECoG Prediction



- A **novel** ECoG dataset: 9 participants listened to a 30-minute audio podcast (Zada et al., 2025).
- We extract features from the audio and input them into our trained model to get source estimates.
- We get zero-shot predictions for ECoG signals by assigning the time series of the nearest source to each electrode.
- Our model generalizes effectively to new subjects and modalities.