Predicting Neural Responses to Multimodal Stimuli with Machine Learning Models

Bertha Shipper Vassar College
Olivia Sopala Barnard College
Qianyi He University of Chicago
Monica Rosenberg University of Chicago
Yuan Chang Leong University of Chicago

Introduction

Why Model the Brain?

- **To understand how our brains make sense of the world.**
- **Movies** are immersive, realistic stimuli that engage vision, language, emotion, and memory all at once.
- **fMRI** captures brain activity over time from thousands of regions.
- Can we **decode** what people see or think just from this activity?
- Machine learning lets us link brain patterns to complex experiences, helping us learn how the mind interprets the world.

What is the Algonauts Challenge?

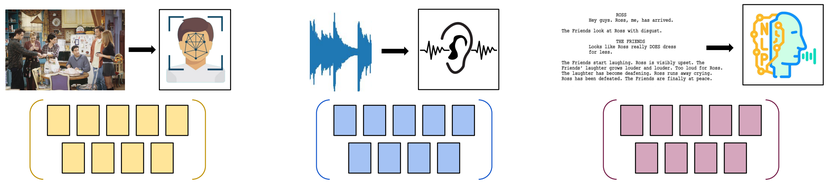
- International competition to **model brain responses** to movies.
- Uses fMRI data from people watching the show *Friends* and films.
- **Goal:** Build encoding models that predict brain activity from movie features.
- Models must **generalize** — accurately predict responses to unseen movies.

The Project Pipeline

1) **Extract** multimodal features from stimuli (visual, audio, language)
2) **Align** features with fMRI timepoints
3) Dimensionality **reduction**
4) **Train** & **validate** encoding models
5) Evaluate brain prediction **accuracy**

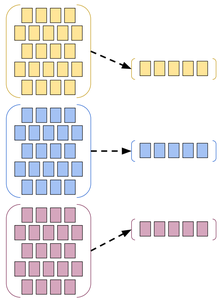
Feature Extraction

Visual	Audio	Language
AlexNet is a CNN that processes frames like a brain does — detecting edges, objects, and scenes.	**HuBERT** is a transformer-based model that takes in raw audio and extracts patterns in speech and sound.	**RoBERTa** is a transformer-based model that uses NLP to capture semantic content in transcripts.



Extracted feature vectors represent **high-dimensional outputs** from **deep neural networks**.

Dimensionality Reduction



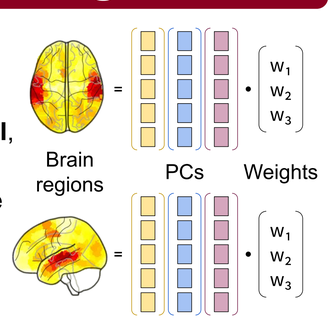
To improve model efficiency and reduce overfitting, we apply Principal Component Analysis (PCA) separately to each modality.

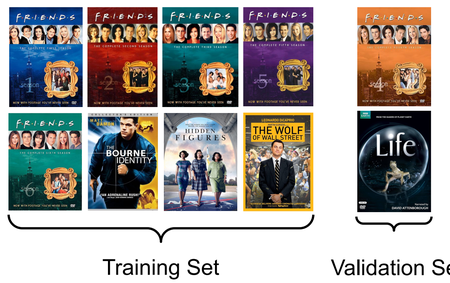
Through hyperparameter tuning, we found that **optimal dimensionality** varies by modality; we retain the top *n* principal components that capture the most informative variance.

Train & Validate Encoding Model

We then train our model to predict fMRI responses to movie stimuli.

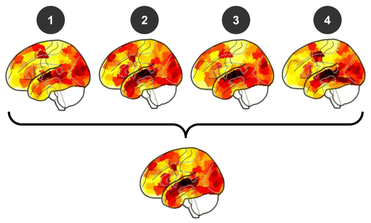
Ridge regression, a **regularized linear model**, learns a set of weights that **map** extracted features to fMRI activity while penalizing large coefficients to **reduce overfitting**.

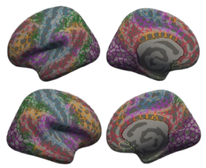




Model performance is **validated** on **held-out** data (unseen movie clips and seasons), assessing the model's ability to **generalize** across **new stimuli**.

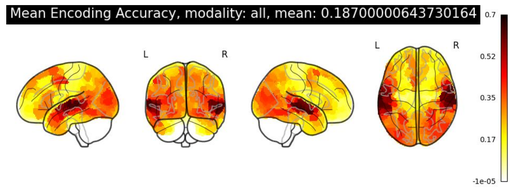
Accuracy is assessed using **Pearson's r** between *predicted* and *actual* fMRI responses across each **region of interest** (ROI).




Schaefer 1000 ROIs

Per-subject accuracy is computed for all **4** subjects, and then mean accuracy is taken across subjects to produce the reported values.

Results: Final Model



Ranked **27th** out of 263 participants in the final phase of the challenge.

Achieved a validation-set accuracy of **r = 0.187** and a test-set accuracy of **r = 0.13**.

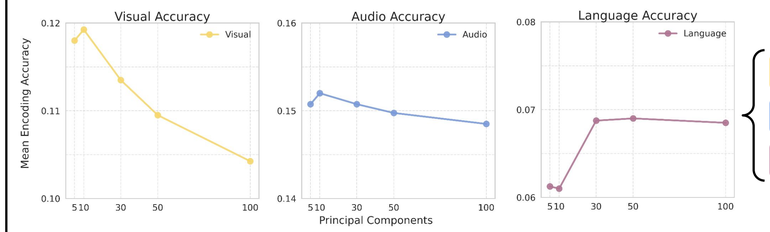
Three Main Knobs for Model Tuning

Features
- Number of **PCs per modality**
- **Neural network layers** for feature extraction

Model Type
- **Linear** vs. **nonlinear** training models (ridge, MLP)

Time
- **Stimulus window** size (temporal context)
- **HRF delay** (hemodynamic lag)

Example: Systematically evaluate **accuracy across PCs** to identify optimal representation for each **modality**.



Conclusions

- Our model predicts **brain activity**, with highest accuracy in **auditory and visual cortices**.
- By combining **sensory modalities**, we advance prior approaches, achieving more comprehensive predictions of neural responses.
- Decoding brain responses from multimodal stimuli is feasible and increasingly effective.

Future Interests

- Explore **temporal integration**: How prior context from language features and stimulus history shape neural responses.

Acknowledgement

We are grateful for the guidance and support of our mentors, YC and Monica, throughout this project, as well as the DSI Summer Lab coordinators. We also thank Qianyi and the members of the CASNL Lab for their feedback and collaboration. Finally, we appreciate the organizers of the Algonauts 2025 Challenge for providing the data and framework that inspired this work.