# Problem Statement and Goals
# Model–to–Brain Alignment for Speech Recognition

Xiao Shao

April 22, 2026

Table 1: Revision History

| Date | Developer(s) | Change |
|---|---|---|
| 14 January 2026 | Xiao Shao | Initial release of document |
| 16 January 2026 | Xiao Shao | Refined based on feedback |
| ... | ... | ... |

## 1 Problem Statement

Understanding how the brain recognizes continuous speech remains an open scientific problem. Modern computational speech models can produce rich internal representations of speech, but it is unclear which representations best align with brain neural responses measured during speech recognition.

Magnetoencephalography (MEG) provides millisecond-resolution neural signals that capture the time course of speech processing. Encoding models such as multivariate Temporal Response Functions (mTRFs) enable a principled comparison between candidate predictors (e.g., acoustic features, model-derived representations) and MEG responses. However, performing these comparisons reproducibly is difficult due to the practical challenges of (i) extracting and time-aligning model representations to the stimulus, (ii) fitting and evaluating mTRFs consistently across different predictors, and (iii) generating clear visualizations and summaries for scientific interpretation.

This project develops a reusable scientific software pipeline that takes speech stimuli and MEG recordings as inputs, constructs time-aligned predictors from multiple speech model representations, fits mTRF encoding models, and produces quantitative and visual comparisons of how well each predictor explains MEG responses.

### 1.1 Problem

### 1.2 Inputs and Outputs

**Inputs:**

- **Speech stimuli:** audio waveforms (and associated metadata such as sampling rate and time alignment information). Optionally, a cochlea-inspired time–frequency decomposition method, gammatone filterbank, is adopted to increase the model's prediction performance.
- **Neural data:** MEG recordings corresponding to the same stimuli, assumed to be available in a standard research format and preprocessed to a usable state for modeling.

**Outputs:**

- **Time-aligned predictor matrices** for each candidate predictor, stored in a reproducible format for reuse.
- **Fitted encoding models** (mTRF weights and hyperparameters) for each predictor and target signal. The software will support linear TRF-style encoding models with a configurable time-lag window and regularization. The default will be a ridge-regularized mTRF formulation, which regularized linear regression over time-lagged predictors, following standard practice in neural speech encoding through Eelbrain.
- **Evaluation summaries** with clearly defined metrics (e.g., Word error rate, prediction correlation) and cross-validation results.
- **Figures and reports** that visualize predictor performance across time lags through Eelbrain.

### 1.3 Stakeholders

- The companies that will use the software to run experiments, compare predictors, and extend the pipeline.
- Researchers or students who wish to benchmark computational speech representations against neural data using a transparent, reusable workflow.

### 1.4 Environment

This software is intended to run on any modern operating system, including Windows 10, macOS, and Linux. Users require a standard scientific computing environment capable of running numerical workflows, reading common audio formats, and handling MEG research data formats. The workflow assumes access to sufficient computing resources for feature extraction and model fitting. A GPU with 8 GB memory is required to accelerate model training or representation extraction, but is not strictly required for the encoding analysis.

Visualization and inspection of results will be supported through a Eelbrain-based neuroscience visualization environment.

## 2 Goals

- **End-to-end reproducibility:** Provide a configuration-driven workflow that can run from input speech stimuli and MEG recordings to final evaluation summaries and figures with a documented command sequence.
- **Predictor modularity:** Support multiple predictor families in a consistent interface, including at least one acoustic baseline predictor and at least two model-derived representation predictors (e.g., hidden-state based).
- **Fair model comparison:** Ensure that all predictors are evaluated under the same cross-validation procedure to enable valid comparisons.
- **Scientific visualization:** Generate high-quality figures that summarizes predictor performance, using an Eelbrain-based visualization workflow.

## 3 Stretch Goals

- Add additional representation predictors (e.g., different models or multiple layers) and enable systematic layer-by-layer benchmarking.
- Extend the evaluation to support comparisons between different regions of interest.

## 4 Extras

This is a research project. The software will support comparisons between computational speech representations and MEG responses using encoding models. The project will include two extras (subject to instructor approval):

- **User Manual:** Provide user-facing documentation (README + quick-start).
- **Code walkthrough:** Conduct a structured walkthrough of the codebase with at least one reviewer.