# Methods — EEG Finger-Print Pipeline for Predicting fMRI BOLD Activity

## ROI Signal Extraction and fMRI Response Variable Construction

The response variable for model training was derived from the fMRI BOLD signal of a region of interest (ROI). The ROI was defined as a 6 mm radius sphere centered on MNI coordinates corresponding to the amygdala. The extracted BOLD signal was upsampled from the native repetition time (TR) to 4 Hz using cubic-spline interpolation to match the temporal resolution of the EEG features. The signal was then normalized to zero mean and unit variance, as described in Meir-Hasson et al. (2014).

## EEG Preprocessing and Time–Frequency Representation

Raw EEG data were acquired simultaneously with fMRI at 250 Hz and preprocessed using a standard pipeline including artifact rejection and bandpass filtering (1–40 Hz). To obtain a time-frequency representation of the EEG, we applied a short-time Fourier transform (STFT) with 1 Hz frequency resolution and 250 ms temporal resolution, approximating the Stockwell transform (Stockwell et al., 1996). While Meir-Hasson et al. originally employed the Stockwell transform to leverage frequency-dependent temporal precision, our implementation uses STFT, which is computationally efficient and sufficiently accurate for our frequency range of interest (Podlipsky et al., 2012). The resulting spectrogram was averaged over 250 ms windows and downsampled to 4 Hz to align with the PDA signal. The frequency spectrum was then compressed into 10 energy-normalized frequency bands using a log-frequency binning approach, dividing the cumulative spectral energy into equal partitions (Atick & Redlich, 1992; Meir-Hasson et al., 2014). Each frequency band signal was z-scored across time to normalize the data and ensure equal contribution across features.

## Feature Space Construction Using Delay Embedding

Given the known variability of the hemodynamic response function (HRF) across regions, frequencies, and individuals (Aguirre et al., 1998), our model incorporated a temporal embedding strategy. Specifically, each EEG channel × frequency band pair was delay-embedded using a sliding window spanning 2 seconds in 250 ms steps (9 delays in total). This resulted in a multivariate feature matrix with dimensions: time × (channels × bands × delays), consistent with the predictive space defined in Meir-Hasson et al. (2014), where the signal at time T in fMRI is modeled as a weighted sum of delayed EEG activity.

## Ridge Regression Modeling and Nested Cross-Validation

To predict the PDA from the EEG features, we employed ridge regression, a regularized linear model that balances data fit and coefficient stability (Tikhonov, 1963; Hoerl & Kennard, 1970). Given the high dimensionality of the feature matrix, ridge regression is well-suited for EEG-fMRI integration due to its interpretability and robustness. Model selection was performed using nested cross-validation to avoid overfitting: An outer loop employed 2 × 5-fold cross-validation to split the data into training and testing sets (Efron & Tibshirani, 1993). Within each training set, an inner loop of 30 random 80/20 train-validation splits was used to select the optimal regularization parameter λ by minimizing the normalized mean squared error (NMSE). The range of λ values was defined by the singular value spectrum of the training data and sampled on a logarithmic scale for computational efficiency (Meir-Hasson et al., 2014).

## Model Evaluation and Statistical Correction

Model performance was assessed using NMSE and Pearson correlation between predicted and actual PDA time series. NMSE was computed as:  
  
NMSE = sum((x\_i - x̂\_i)^2) / sum((x\_i - x̄)^2)  
  
where x is the true PDA signal and x̂ the predicted signal. A value of NMSE < 1 indicates better-than-mean prediction performance. Pearson correlation was also computed to evaluate alignment in signal fluctuations between modalities. To assess the significance of model coefficients, t-tests were applied to weights across cross-validation folds. Multiple comparisons were corrected using the False Discovery Rate (FDR) method (Storey, 2002), retaining only statistically significant electrode-frequency-delay weights (p < 0.05, FDR-corrected). Predictors and electrodes not passing FDR correction were excluded from further analysis.

## Model Interpretation and Visualization

The final EEG Finger-Print (EFP) for the ROI was visualized through:  
1. Topographic maps showing the spatial distribution of predictive EEG sources for each frequency band.  
2. Time series plots comparing the smoothed EEG predictor from the most informative feature (electrode × band) to the PDA.  
3. A ranked list of electrode-frequency-delay contributions to the final prediction, revealing dominant channels and bands contributing to the ROI signal. These outputs reflect a data-driven estimation of the spatiotemporal EEG signature associated with deep fMRI activity and are consistent with the framework proposed by Meir-Hasson et al. (2014).

## References

* Meir-Hasson, Y., Kinreich, S., Podlipsky, I., Hendler, T., & Intrator, N. (2014). An EEG Finger-Print of fMRI deep regional activation. NeuroImage, 102(2), 128–141.
* Stockwell, R. G., Mansinha, L., & Lowe, R. P. (1996). Localization of the complex spectrum: the S transform. IEEE Transactions on Signal Processing, 44(4), 998–1001.
* Podlipsky, I., et al. (2012). fMRI and MEG study of frequency band-specific functional connectivity in the human brain. NeuroImage, 62(2), 932–940.
* Atick, J. J., & Redlich, A. N. (1992). What does the retina know about natural scenes? Neural Computation, 4(2), 196–210.
* Aguirre, G. K., Zarahn, E., & D’Esposito, M. (1998). The variability of human, BOLD hemodynamic responses. NeuroImage, 8(4), 360–369.
* Tikhonov, A. N. (1963). Solution of incorrectly formulated problems and the regularization method. Soviet Mathematics Doklady, 4, 1035–1038.
* Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55–67.
* Efron, B., & Tibshirani, R. (1993). An Introduction to the Bootstrap. Chapman & Hall/CRC.
* Storey, J. D. (2002). A direct approach to false discovery rates. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 64(3), 479–498.