Hierarchical Latent Variable Models for Neural Data Analysis

Vaibhav Bommisetty
vbommisetty@ucsd.edu

Beomseuk Seo
bseo@ucsd.edu

Mentor: Mikio Aoi
maoi@ucsd.edu

UC San Diego
HALICIOĞLU DATA SCIENCE INSTITUTE

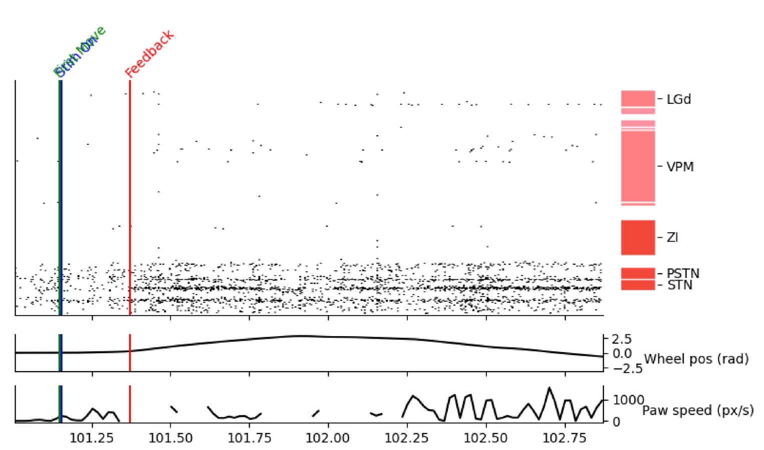
Objective

The objective of this project is to identify sensitive neural clusters in sensorimotor integration brain regions and analyze their hidden dynamics using hierarchical latent variable models to uncover underlying patterns in neural activity.

Introduction

Understanding how neural activity encodes behavior requires identifying sensitive clusters and their latent dynamics. This study focuses on the Superior Colliculus dorsal gray (SCdg) and Superior Colliculus intermediate white (SCiw) regions, which are implicated in sensory-motor integration and decision-making.

The horizontal lines mark key behavioral events (e.g., stimulus, feedback) during the trial, showing how neural activity in regions like LGd and VPM aligns with behavioral measures like paw speed, revealing the timing of neural and behavioral responses.

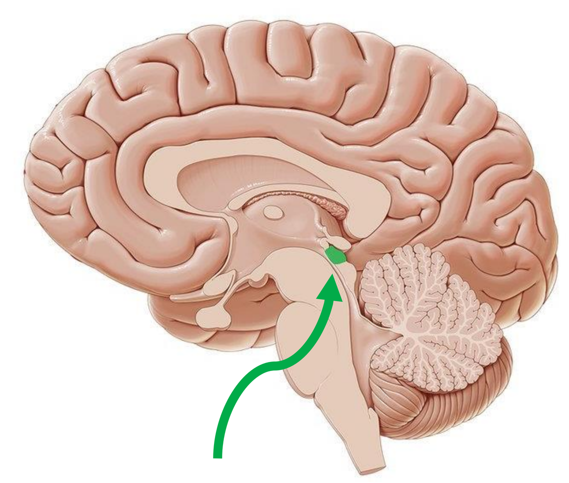


Background

Mice were trained to perform a decision-making task where they had to make choices based on visual stimuli. During the task, neural activity was recorded using Neuropixels probes as part of the International Brain Laboratory (IBL) initiative, a collaborative effort to study brain-wide neural circuits underlying decision-making. The mice were presented with sensory cues and had to turn a wheel to indicate their choice.

INTERNATIONAL BRAIN LABORATORY

Methodology



Superior Colliculus

The superior colliculus is a midbrain structure involved in sensory processing, motor control, and decision-making. Hierarchical latent variable models, such as PCCA, are powerful tools for uncovering hidden structure in high-dimensional neural data.

Data Analysis Pipeline:

IBL Data Collection → Pre-processing → Sensitive Cluster Identification → Latent Variable Modeling

IBL Data Collection

- Neural activity was recorded using probes, that captured spike times and waveforms, from neurons.
- Clusters are groups of spikes sorted based on waveform similarity, with only high-quality clusters used for analysis.

Pre-processing

- Spike data was filtered to include only "good" clusters (75,708 neurons), ensuring high-quality signals.
- Spike times were binned into time windows aligned to behavioral events for analysis.

Sensitive Cluster Identification

- Permutation testing was used to identify clusters with significant responses to behavioral events.
- Clusters were tested for sensitivity to events like stimulus onset, first movement, and feedback.

Latent Variable Modeling

- PCA, CCA, and PCCA were applied to extract shared and dataset-specific variability in neural activity.
- PCCA was used to model hidden dynamics, capturing the relationship between neural activity and behavior.

CCA, PCCA:

Canonical Correlation Analysis (CCA)

CCA finds linear projections that maximize correlation between two datasets:

$$\rho = \max_{\mathbf{w}_1,\mathbf{w}_2} \frac{\mathbf{w}_1^T \mathbf{\Sigma}_{12}\mathbf{w}_2}{\sqrt{\mathbf{w}_1^T \mathbf{\Sigma}_{11}\mathbf{w}_1}\sqrt{\mathbf{w}_2^T \mathbf{\Sigma}_{22}\mathbf{w}_2}} \tag{3}$$

Where:

- $\mathbf{\Sigma}_{11} = E[\mathbf{X}_1\mathbf{X}_1^T]$ and $\mathbf{\Sigma}_{22} = E[\mathbf{X}_2\mathbf{X}_2^T]$ are the covariance matrices.
- $\mathbf{\Sigma}_{12} = E[\mathbf{X}_1\mathbf{X}_2^T]$ is the cross-covariance matrix.
- $\mathbf{w}_1, \mathbf{w}_2$ are the projection vectors.

Probabilistic Canonical Correlation Analysis (PCCA)

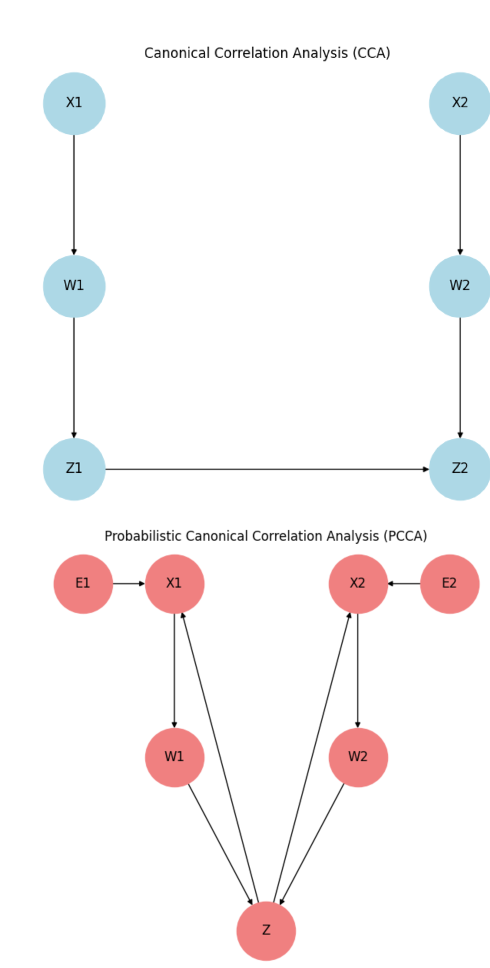
PCCA extends CCA to a probabilistic setting:

$$\mathbf{X}_1 = \mathbf{W}_1\mathbf{Z} + \epsilon_1, \quad \mathbf{X}_2 = \mathbf{W}_2\mathbf{Z} + \epsilon_2 \tag{4}$$

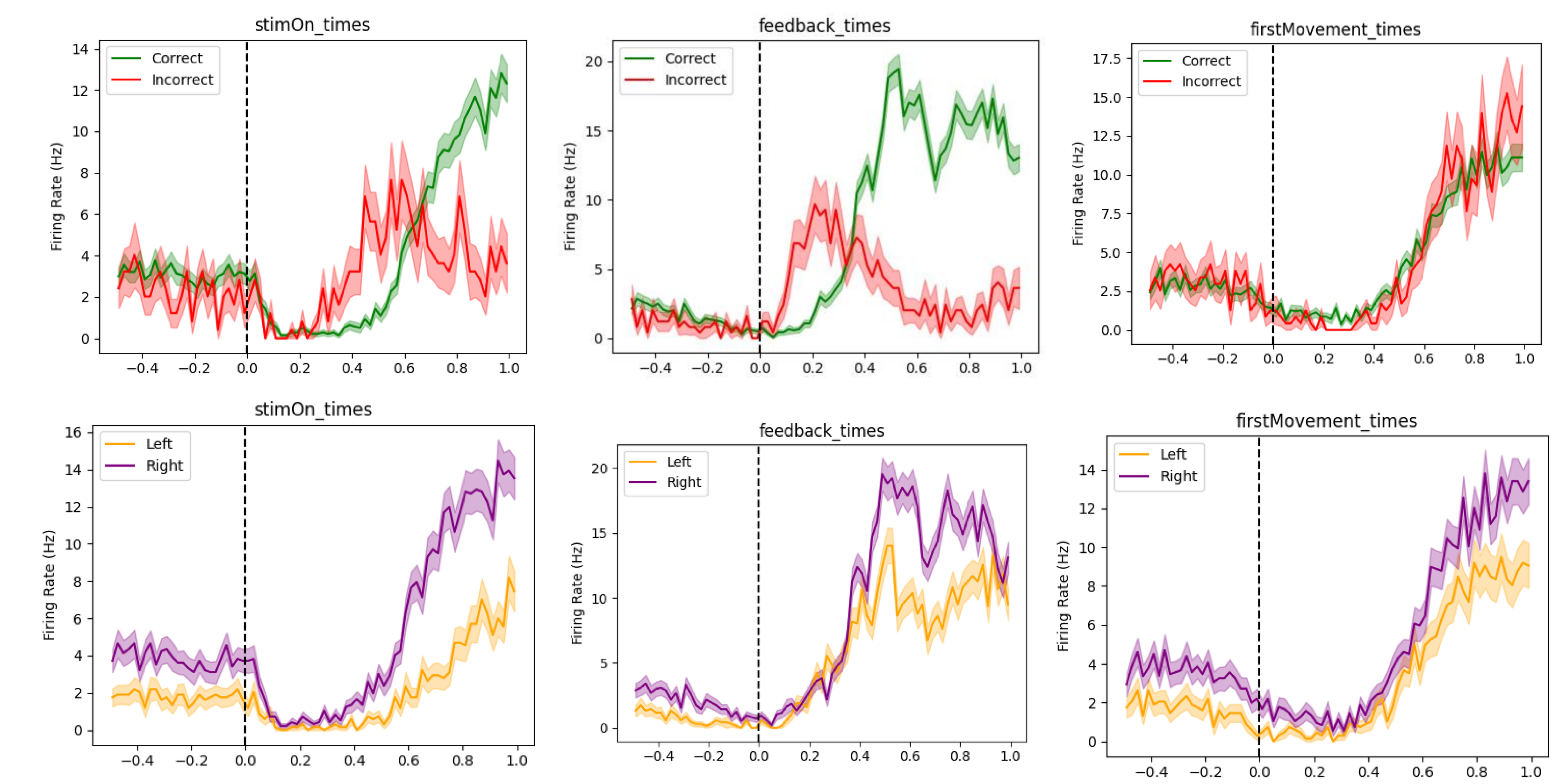
- $\mathbf{Z} \sim \mathcal{N}(0, \mathbf{I})$ is a latent variable.
- $\epsilon_1 \sim \mathcal{N}(0, \mathbf{\Psi}_1)$ and $\epsilon_2 \sim \mathcal{N}(0, \mathbf{\Psi}_2)$ are noise terms.
- $\mathbf{W}_1, \mathbf{W}_2$ are the factor loading matrices.

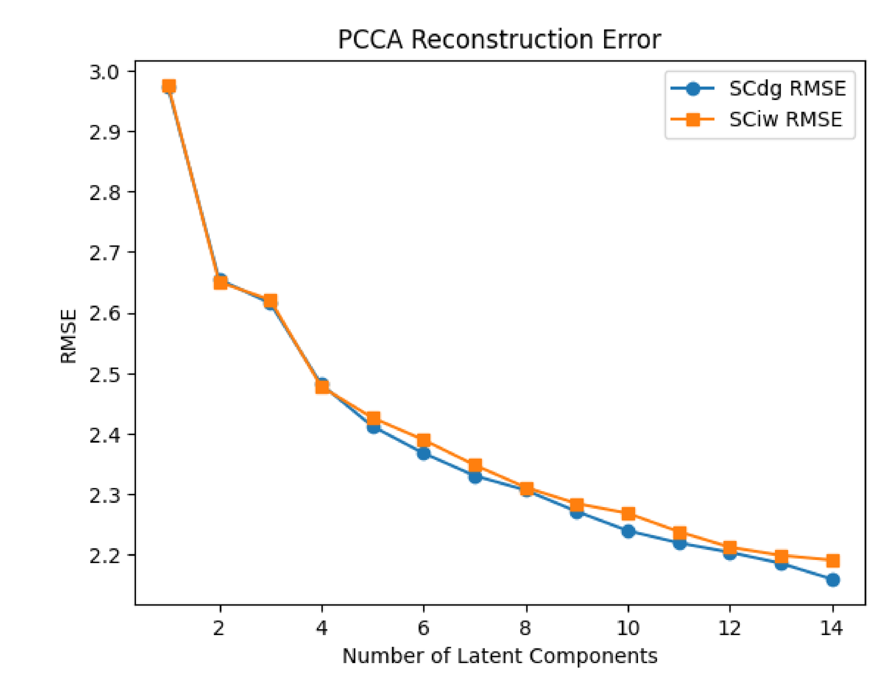
The parameters are learned using Expectation-Maximization (EM), and the marginal likelihood is given by:

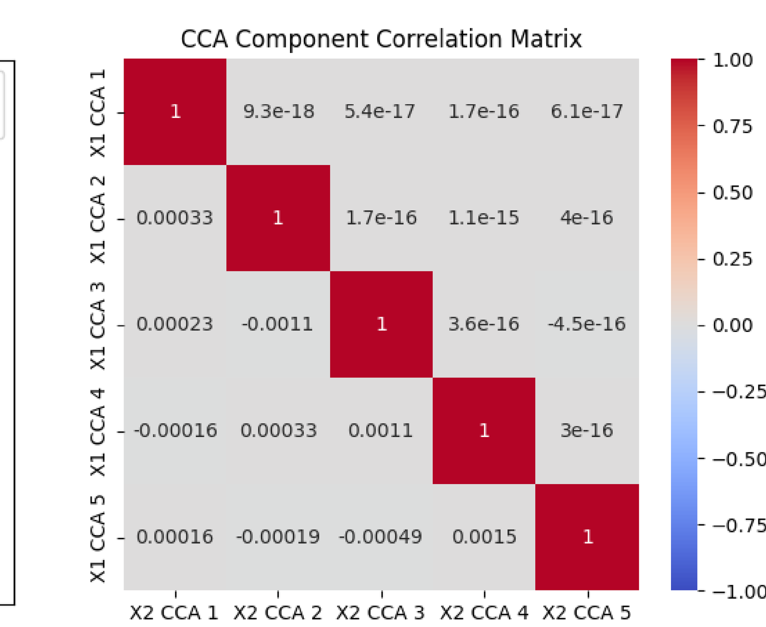
$$p(\mathbf{X}_1, \mathbf{X}_2) = \int p(\mathbf{X}_1|\mathbf{Z})p(\mathbf{X}_2|\mathbf{Z})p(\mathbf{Z})d\mathbf{Z} \tag{5}$$



Results







Conclusion:

Our analysis of SCdg and SCiw neural activity using PCA, CCA, and PCCA suggests that while some shared structure exists between the two regions, the degree of alignment is weaker than expected.

- Canonical Correlation Analysis (CCA) results indicate low cross-component correlations, suggesting that SCdg and SCiw may not share strongly correlated latent representations.

- Probabilistic Canonical Correlation Analysis (PCCA) showed gradual RMSE reduction with increasing latent dimensions, but reconstruction errors remain relatively high (~2.2-2.3). This suggests that PCCA is capturing some shared variance but struggles to fully align the datasets.

- Firing rate analyses reveal that feedback processing shows the most distinct activity differences between correct and incorrect trials, while movement-related activity is less differentiated. This could imply that these regions are more involved in feedback-based learning rather than motor execution.