Title: Characterizing cue related response activity in the rodent premotor cortex

- **Troy Schouten (tdschout)**
- **Francisco Aparicio (fapar)**
- **Connor Ding (czsding)**

The brain, with over 80 billion neurons, produces complex signals to drive a wide variety of behaviors. Neuroscientists study brain activity to understand how these high-dimensional signals code for specific behaviors. A common approach in neuroscience is to record neural activity during a behavior task to determine what activity drives particular behaviors. *However, since neural data is high dimensional, extracting meaningful information proves to be challenging*. To overcome this challenge, machine learning techniques are often used to reduce the dimensionality of the data and extract key features in neural activity. It is often the case that neural activity becomes less variable and more structured in the later stages of learning. This is often represented as the increased synchronous firing of neurons or as the emergence of stereotyped activity patterns. **To explore this phenomenon, we will apply machine learning techniques to characterize how the neural activity of the mouse secondary motor cortex (M2) evolves during the course of learning**.

The animal motor system contains several specialized nodes that are engaged during different phases of a movement. For example, the premotor cortex (M2) is largely believed to be involved in action planning and in the integration of sensory input that elicits a motor action. Interestingly, recent work looking at the activity of M2 during a spatial recognition task found cue-related activity in M2 that was present not only during the task but was 'reactivated' during periods of rest[1]. This suggests that the cue-related activity could be undergoing reactivation-dependent consolidation after training - a phenomenon widely observed in the hippocampus. This phenomenon, however, has not been explored in non-spatial recognition tasks. Here we are interested in exploring the relationship between the M2 activity, reactivations, and behavior performance in a cued lever-pushing task. Therefore, we aim to: 1) **use dimensionality reduction techniques to find a low dimensional representation of the M2 cue-related activity in a skilled motor task (lever push)** and 2) **analyze 'offline' periods (between trial/ sleep) to detect reactivations of the cue response activity**.

Our dataset was collected, by a lab member in the Ganguly Lab at UCSF, from multiple mice across several days during the learning phase of the task. Mice were placed in a head-fixed experimental setup where they were presented with a cue that indicated the start of the trial. In each trial, the mice had a three-second window to push the lever across a threshold distance. Successful trials, were where the mice crossed the threshold, upon which they were rewarded with a drop of sugar water. The training session consisted of a series of blocks (5 trials) followed by offline blocks in between each training block. During the entire training session, a high-density probe (Neuropixels) was used to record neural activity (at 30khz) in M2. The activity (spikes) of single units (putative neurons) were obtained (by a group member) by running a widely used spike sorting algorithm (Kilosort) and curated using the accompanying program Phy. The resulting dataset contains cue timestamps, lever positions, and the spike times of each putative neuron across the whole training session (online and offline periods).

Using our dataset, we plan to characterize the evolution of cue-related neural activity using several measurements of neural engagement. We started by using a baseline model (logistic regression) to classify whether a neuron responds to the cue. Identifying the number of neurons that are “cue modulated” will allow us to determine whether there is a coupling between the onset tone and M2 activity. To determine this, we aligned the neural spiking data to specific cue events to understand the neural dynamics surrounding these stimuli. By visualizing (see Figure 1 below) and comparing the neural activity before and after the cue onset, we derived a `ratechange` vector, which captured changes in firing rates, spanning values from -2 to 19. To categorize these changes and provide a clear dichotomy, we defined a threshold: if the change in firing rate (`rateChange`) was two standard deviations above the mean or more, the outcome variable `y` was set to 1, and 0 otherwise. This combined rateChange and y vector is our training and testing dataset for our logistics regression model. We implemented a 10-fold cross-validation to ensure the robustness and generalizability of our model. The resulting visual representation (See Figure 2 below) highlighted the logistic curve, data points, and a decision boundary, indicating a strong association between significant rate changes and the binary outcome.

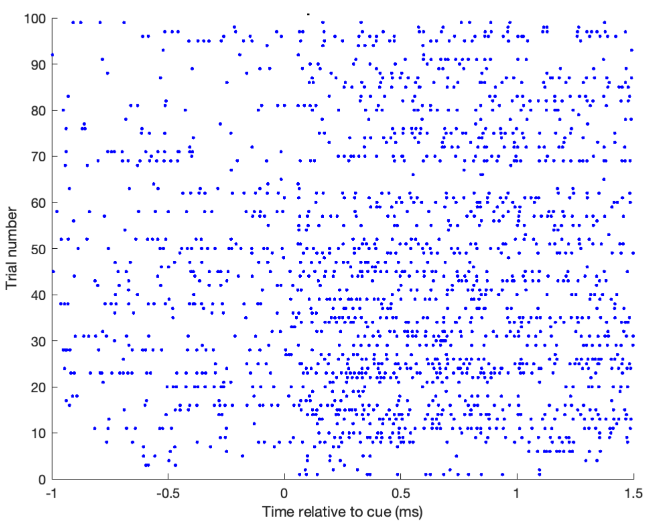

Figure 1: Raster plot for a sample unit (5).

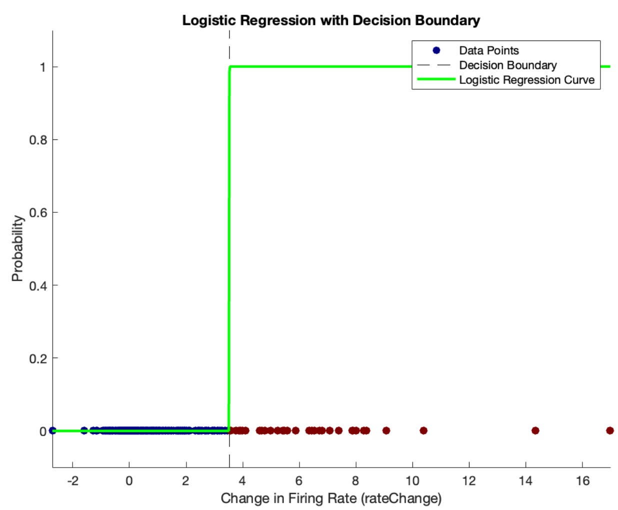

Figure 2: Preliminary Logistic Regression Results

Future directions: To characterize the emergence of stereotyped correlated activity, we will perform principal component analysis (PCA) to evaluate the low dimensional space of the neural activity. PCA is typically used on data where the variables in the data are correlated with each other, which is the case for neuronal data where individual neurons may be part of the same neural network. Specifically, each time window (e.g., 50 ms) can be treated as a point in n-dimensional space, where n is the number of neurons contributing to the signal being explored. Since PCA reduces dimensionality while preserving the most important information in the data, it provides a method for feature extraction of which parts of the data contribute to motor learning the most. To quantify increases in correlated neural activity, we will determine how much variance is explained by each component and how many components would be necessary to explain >95% of the variance. Since PCA focuses on linear relationships, we will standardize the data to a mean of 0 and a variance of 1, which will allow the PCA to respond appropriately to the scaling of the data. We expect that when specific network activity in M2 becomes coupled to the tone, neural activity will become more correlated, resulting in more prominent components that explain a larger percent of the variance. Ultimately, we will plot the first two principal components, allowing us to find clusters of related data points in 2-dimensional space.

In future analysis, we plan to further explore the relationship between the tone and M2 activity by looking at the latency between cue introduction and M2 neuron spiking. Here we expect to see a population of neurons that fire shortly after the tone and before the movement. Additionally, to explore neural reactivations following training we plan to combine Principal Component Analysis (PCA) with Recurrent Neural Networks (RNNs). First, we employ PCA on our dataset from 300+ neural channels in order to capture the most significant patterns of activity during the training ("encoding phase"). This dimensionality reduction method allows us to capture essential information on the neuronal activity. Next, these principal components become the input for the RNNs. We choose RNNs for their prowess in handling temporal data and they will be trained on the encoding patterns to predict subsequent reactivation sequences during the offline period. By pairing PCAs with RNNs, we aim to offer a robust methodology for investigating the nuances of neural reactivation during awake resting phases and its role in learning consolidation.