Application of data science to multi-brain area simultaneous recordings in the mouse during movement planning

Smit Patel

MSc in Applied Data Science

Georg-August-Universität Göttingen, Germany

E-mail: smit.patel@stud.uni-goettingen.de

Student ID: 29722852

January 15, 2024

Abstract

The study delves into the application of data science methodologies to multi-brain area simultaneous recordings in mice during movement planning. Our focus revolves around determining the significance of the Anteromedial Prefrontal Cortex (ALM) in decision-making and assessing the Thalamus's responsiveness to ALM neuron firing using Cross-Correlograms analysis. The experiments also yield insights into selectivity of left and right lick trials with the use of Coding direction.

1. Introduction:

In recent years, there has been notable progress in the field of neuroscience with an escalating utilization of technologies designed for the recording of extensive neuronal activities, either sequentially or simultaneously. Despite this advancement, the direct observation of functional interactions at the cellular level throughout the neural hierarchy remains a formidable challenge, primarily due to the complexity of simultaneously recording activities across numerous regions (Siegle 2021). The current array of multi-electrode and optical recording technologies has enabled the simultaneous monitoring of neuronal activities in the cortex and, in certain instances, in deeper anatomical structures. Ongoing advancements in recording technologies hold the promise of significantly amplifying the number of neurons that can be simultaneously recorded, potentially by orders of magnitude (Cunningham 2014).

The primary objective of this project is to unravel the underlying mechanism governing the planning processes associated in memory-guided tasks. The focus is on discerning the potential decision-making role of the Anteromedial Prefrontal Cortex (ALM) in the execution of these tasks in mice. Furthermore, the project aims to investigate and elucidate the involvement of other regions within the brain, seeking to understand how these regions interact with the ALM during the decision-making processes.

2. Data Source

The Allen Institute for Neural Dynamics and Janelia Research Campus has been actively engaged in the field of large-scale neural recordings, achieving a significant milestone in 2023 by successfully capturing neural activity across multiple regions of the brain simultaneously. This accomplishment is underpinned by an extensive analysis of data tables derived from recordings conducted on 28 mice (Mus musculus - House mouse) during memory-guided tasks.

The dataset encompasses anatomy-guided recordings spanning various connected brain regions, extending from the anterior lateral motor cortex to the medulla. The data collection involved the acquisition of electrophysiological and behavioral data, facilitated by the creation of three to four small craniotomies (diameter, 1 ~ 1.5 mm) over the subjects. Multiple Neuropixels probes were strategically inserted at different depths within various areas of the brain.

Behavioral sessions, lasting 1 to 2 hours, served as the context for data collection, with all recordings stored as time-series data in formats such as BehavioralEvents, BehavioralTimeSeries, Unit, and ElectrodeGroup. Of particular interest within this dataset is the spike times of individual neurons, which provide crucial insights into the neural dynamics associated with the observed behaviors. The dataset comprises stimulus epochs lasting 0.65 seconds, delay epochs lasting 1.2 seconds, and go epochs spanning 1.5 seconds.

All the Data has open access and available on DANDI in the form of Neurodata Without Borders (NWB).

3. Data Engineering:

3.1 Data Curating:

To obtain the dataset from the DANDI repository, one can utilize either the Python CLI Client or the DANDI Web application. The process involves downloading data for specific subjects, the names of which can be identified on the DANDI web application or by executing commands through the DANDI CLI. Once the subject ID is determined, a specific file or session link can be obtained through the DANDI CLI for further data retrieval.

Subsequently, the curated data must be organized, consolidating information from different sessions under the main Data Table for each mouse while still preserving individual session details. This curation process ensures a comprehensive yet structured representation of the dataset, facilitating subsequent analyses and interpretations.

3.2 Data Cleaning:

Data Contains 100’s of Time series data and two tables called Units and Trials. Units contains 39 columns. It has categorical, numeric and multiple time series Data as a columns in it. For the purpose of our data analysis, the BehavioralTimeSeries data presents start and stop timing for each epoch. Ideally, the stop time of one epoch and the start time of the next epoch should align. To ensure data integrity, we intend to discard multiple epochs within a trial where timing misalignment occurs. Specifically, instances with more start times than stop times will be excluded from the analysis.

Moreover, it is imperative to consider the duration of each epoch. Epochs with stop-start pairs whose duration deviates from the norm will be omitted from the dataset. The distribution of epoch durations exhibits variability, and thus, extreme outliers will be removed to maintain data consistency.

Notably, the duration of delay epochs varies significantly. To address this variability, we will focus on the most frequently occurring duration, which falls within the range of [1, 1.3] seconds. Additionally, to ensure overall trial length consistency, trials exceeding a length of 5.3 seconds will be excluded from the analysis. This filtering process aims to enhance the quality and reliability of the data for subsequent analysis.

To make progress quickly in our project, we decided to use some MATLAB code that was already available instead of starting from scratch in Python. This is because the data we are working with required careful preprocessing, and using existing code in MATLAB helped us speed up our work and meet milestones faster. By doing this, we could take advantage of the work that had already been done in MATLAB, saving time and ensuring consistency in our analysis. This approach allowed us to be more efficient and practical in handling the complexities of the dataset and the analyses we needed to perform.

4. Data Science

4.1 Introduction: Tasks and relevant questions:

The intricate organ that is the brain engages in diverse computations essential for our survival, one of which involves the planning of movements. This cognitive process has been the subject of investigation across various animal models, with a recent focus on rodents. Technological advancements now enable simultaneous recordings of neuronal activity across different areas in mice, holding the promise to unravel the underlying mechanisms of movement planning and related functions.

However, the interpretation of this high-dimensional neural data poses a challenge. To unravel the intricacies of planning mechanisms, we aim to employ various dimensionality reduction techniques, such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) in the form of Coding Direction (CD). Our investigation extends beyond these conventional techniques to explore additional methods addressing the debate over weak versus strong arguments for dimensionality reduction.

The weak argument posits that dimensionality reduction serves as a convenient tool for interpreting high-dimensional datasets. Conversely, the strong argument asserts that the brain actively utilizes dimensionality reduction as a strategy to encode and communicate information effectively. Our exploration seeks to discern whether dimensionality reduction merely aids in data interpretation or if it constitutes a fundamental aspect of how the brain encodes and communicates information, thereby uncovering the neural code explicitly.

In another aspect of the project, we are focused on understanding how different parts of the mouse brain communicate and identifying multi-regional loops. We utilize Canonical Correlation Analysis (CCA) to reveal relationships between neural activities in different brain regions, and Cross-Correlograms (CCG) for a temporal analysis of spike train correlations. These methods aim to characterize multi-regional computations, providing insights into the intricate functional architecture and communication patterns within the mouse brain.

4.2 Firing Rate:

Information in the brain is represented as action potentials (neuron spikes), which may be grouped into spike trains or even coordinated waves of brain activity. Action potentials convey information through their timing (Larry Abbott 2001), and to use it in meaningful way, other spike statistic like spike rate and firing rate are used in our analysis.

Neuronal firing rate, representing the frequency of action potentials, is pivotal in decoding neural information processing.

$$FR=T/S$$

Where:

FR is the firing rate,

S is the total number of action potentials (spikes) fired by the neuron,

T is the total time period over which the spikes are counted.

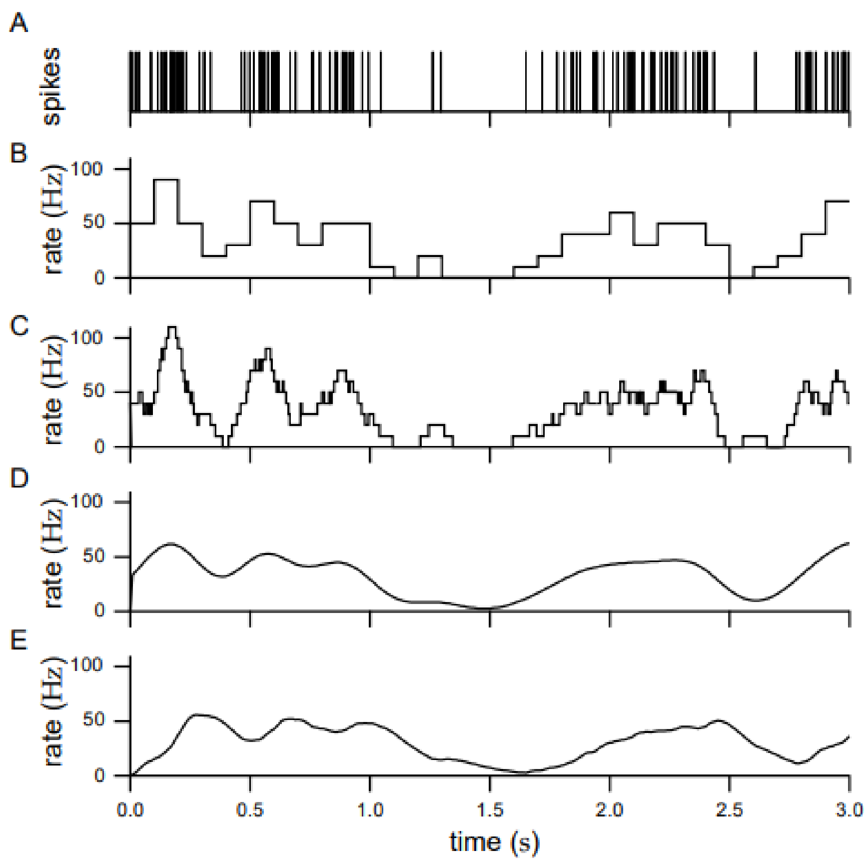


Figure 1 (Larry Abbott 2001) A spike train from a neuron in the inferotemporal cortex of a monkey recorded while that animal watched a video on a monitor under free viewing conditions. (B) Discrete time firing rate obtained by binning time and counting spikes with 1t = 100 ms. (C) Approximate firing rate determined by sliding a rectangular window function along the spike train with 1t = 100 ms. (D) Approximate firing rate computed using a Gaussian window function with σt = 100 ms. (E) Approximate firing rate using the causal window function with 1/α = 100 ms.

In our analysis, we employed histograms and window functions to effectively quantify and visually represent the firing rate for a more comprehensive insight into the data. This approach yielded a robust approximation by applying a Gaussian kernel with a sigma of 0.1.

4.3 Population Analysis:

4.3.1 Coding Direction:

A set of orthogonal directions in the activity space (n x 1 vectors) is established for a population of n neurons. These vectors, referred to as coding directions (CD), are designed to maximize the separation between response vectors during lick-left and lick-right trials at specific task-related times within the n-dimensional activity space (Chen 2023).

Method: In our analysis for each lick direction, we calculated the mean spike counts separately for$\vec{x}_{\text{lick left}}$ and $\vec{x}_{\text{lick right}}$. Subsequently, we determined the direction of the difference in the mean response vectors at each time point $\vec{w_t}$ =$\vec{x}_{\text{lick left}}$ and $\vec{x}_{\text{lick right}}$. To obtain a comprehensive measure, we averaged the $\vec{w}_t$ values over the specified epoch. Multiplying this result with the corresponding lick firing vector allowed us to project the values onto the coding direction (CD) vector (Chen 2023). This process provided a meaningful representation of the relationship between lick direction and neural responses.

The investigation involved the analysis of the activity within each brain region, utilizing an activity space where individual dimensions represent the activity levels of distinct neurons. Over time, the population activity traces a trajectory within this multidimensional activity space. Through an examination of correct trials, a specific direction in the activity space was defined. This direction maximally discriminated upcoming choices, distinguishing between lick-left and lick-right trials, particularly during the late delay epoch (0.6 sec) (Chen 2023).

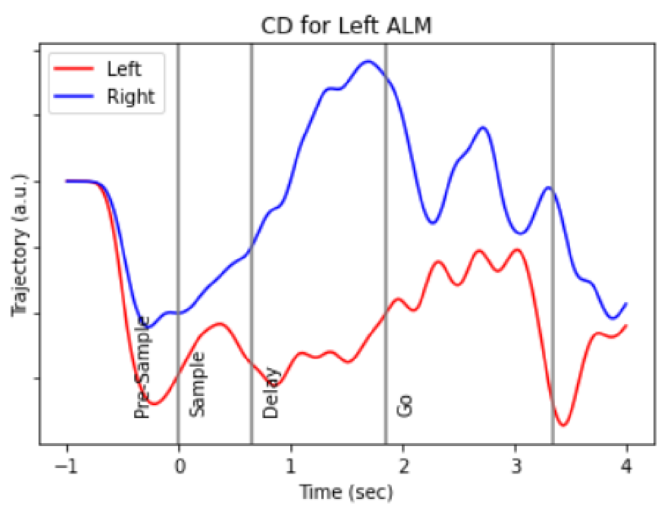
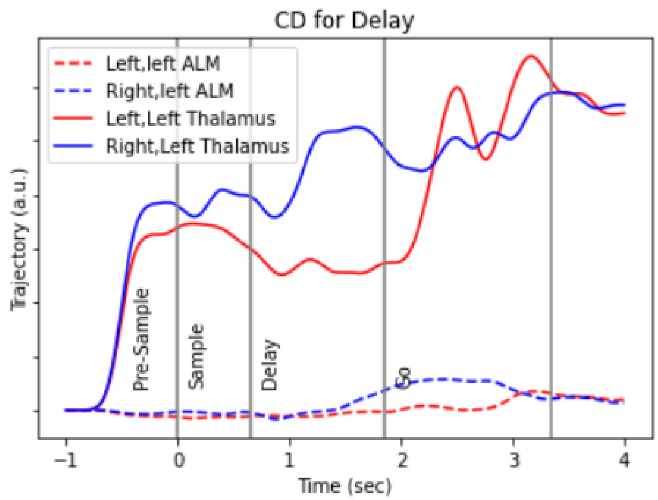


Figure 2 Left and right lick's firing rate Projection onto Coding Direction (Delay)

During our experimentation with the left anterior lateral motor cortex (ALM) in the coding direction, averaged over the late delay period (last 0.6 seconds), we established boundaries to distinguish between left and right trial coding directions (CD_{delay}). Subsequently, in our analysis, we extended this approach to both the Left ALM and Thalamus, considering both the late delay period and early Go period (first 0.6 seconds). Remarkably, we observed a distinct separation in coding direction for both regions during these specific time intervals.



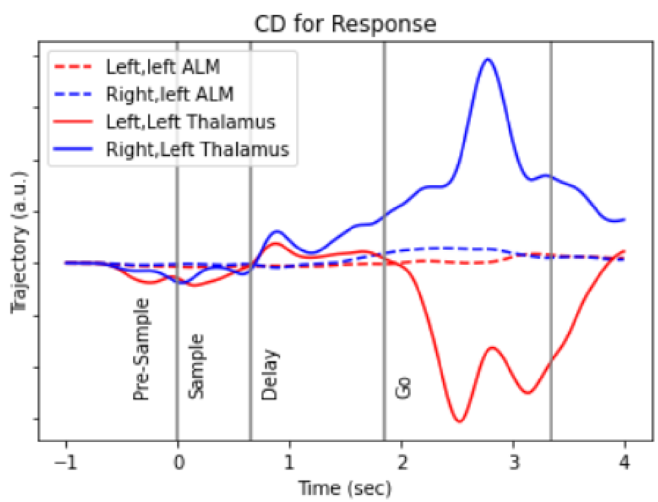


Figure 3 (1.) Projection onto coding direction for late delay. (2.) Projection onto Coding Direction for early go

4.3.2 PCA:

PCA identifies an ordered set of orthogonal directions that captures the greatest variance in the data. Here PCA is used to reduce the dimensionality of the data by identifying the principal components (PCs) that capture the most significant variance in the firing rate patterns. This reduction can help us to represent underlying patterns or trends in the neural data, aiding in the extraction of meaningful information.

While capturing the largest amount of variance is desirable in certain scenarios, it's important to note a caveat associated with Principal Component Analysis (PCA). The low-dimensional space identified by PCA captures variance of all types, encompassing both firing rate variability and spiking variability (Cunningham 2014).

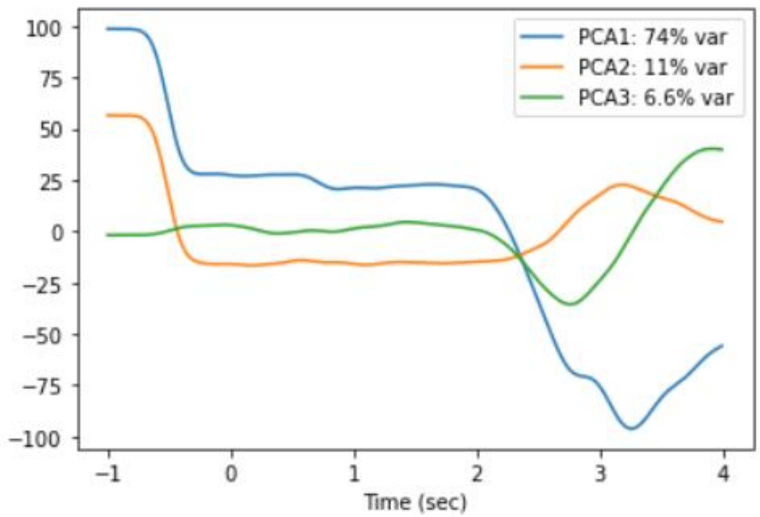


Figure 4 PCA components of firing rate vectors

In our experiments, the initial impression of neuron firing rates appeared to be highly random. However, upon conducting Principal Component Analysis (PCA), we discovered that the seemingly complex and erratic structure of the firing rates could be effectively distilled into a combination of a few key vectors, represented by the principal components.

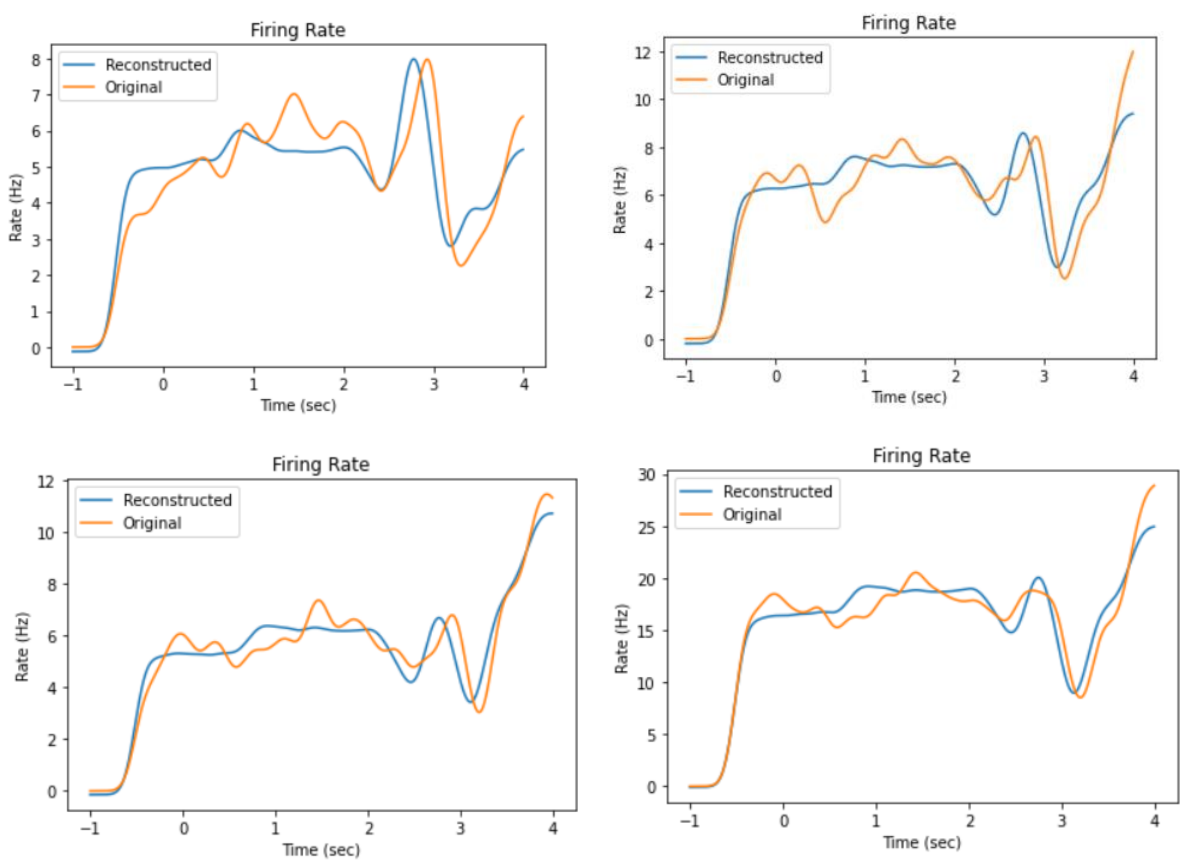


Figure 5 Reconstructed Firing rate of random neurons

Multimodal Analysis:

Cross correlation analyses:

Cross-Correlograms (CCG): In our study, we quantified functional interactions between pairs of neural units utilizing Cross-Correlograms (CCGs). Specifically, we employed CCG time-lag analysis to reveal a significant correspondence between the Anterior Lateral Motor cortex (ALM) and the Thalamus during the memory-guided task. The objective was to discern the dynamics of signal transmission between these brain regions, identifying which part plays the role of the sender and elucidating the associated time delays.

$$\mathrm{CCG}(\tau) = \frac{\frac{1}{M}\sum_{i=1}^{M}\sum_{t=1}^{N} x_1^i(t)x_2^i(t+\tau)}{\theta(\tau)\sqrt{\lambda_1\lambda_2}}$$

where M is the number of trials, N is the number of bins in the trial, x^i_1 and x^i_2 are the spike trains of the two units on trial i, τ is the time lag relative to reference spikes, and λ_1 and λ_2 are the mean firing rates of the two units, and the CCG is essentially a sliding dot product between two spike trains. $\theta(\tau)$ is the triangular function which corrects for the overlap time bins caused by the sliding window. In the results, a sharp peak was deemed significant if the maximum of raw CCG amplitude within a ±10ms window had a magnitude larger than sixth fold of the standard deviation of the CCG flanks (between ±50ms from zero). All subsequent analysis was based on significant CCG sharp peaks (Siegle 2021).

[Graphs]

[Explanation]

Discussion, Summary, Future Work.

References

Chen, Susu and Liu, Yi and Wang, Ziyue and Colonell, Jennifer and Liu, Liu D. and Hou, Han and Tien, Nai-Wen and Wang, Tim and Harris, Timothy and Druckmann, Shaul and Li, Nuo and Svoboda, Karel. 2023. "Brain-wide neural activity underlying memory-guided movement." Cold Spring Harbor Laboratory.

Cunningham, John P and Yu, Byron M. 2014. "Dimensionality reduction for large-scale neural recordings." *Nature Neuroscience* 17 (Springer Science and Business Media LLC): 1500-1509.

Larry Abbott, Peter Dayan. 2001 . *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems.* The MIT Press.

Manly, Bryan F.J. 2004. *Multivariate Statistical Model.* Third Editon. Chapman and Hall/CRC.

Siegle, Joshua H. and Jia, Xiaoxuan and Durand, Séverine and Gale, Sam and Bennett, Corbett and Graddis, Nile and Heller, Greggory and Ramirez, Tamina K. and Choi, Hannah and Luviano, Jennifer A. and Groblewski, Peter A. and Ahmed. 2021. "Survey of spiking in the mouse visual system reveals functional hierarchy." *Nature* 592 (Springer Science and Business Media LLC): 86-92.

Svoboda, Karel and Li, Nuo. 2018. "Neural mechanisms of movement planning: motor cortex and beyond." *Current Opinion in Neurobiology* 49 (Elsevier BV): 33-41.