Exploratory Analysis on Allen Institute Visual Cortex Recordings

Arshia Razavi*

(Dated: December 16, 2023)

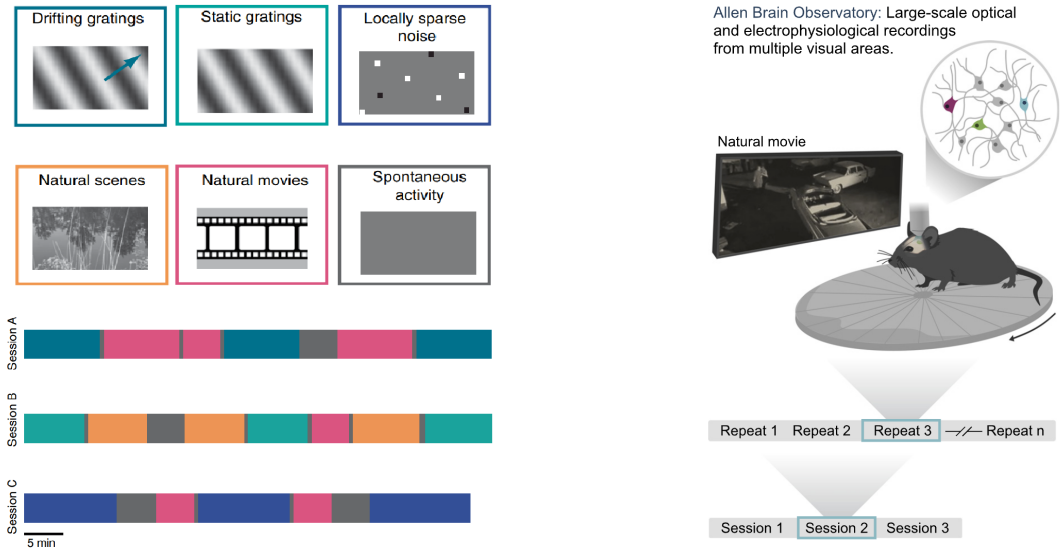
ABSTRACT

The project focuses on investigating neuronal images of the mouse visual cortex recorded at the Allen Institute using the calcium imaging technique. The mice were exposed to a visual stimulus in the form of a movie without any specific task. The primary objective of this study is to enhance our understanding of these brain signals through various data analysis processes, including data retrieval, data cleaning, Pearson correlation calculation, and specifically, principal component analysis (PCA). Our analysis based on PCA analysis led to the conclusion that the VISpm and VISrl cortex regions exhibit similar behavior to each other and effectively encapsulate more information from the stimulus movie within the PCA space.

* M.Sc. Physics and Astronomy department, University of Calgary

I. INTRODUCTION & BACKGROUND

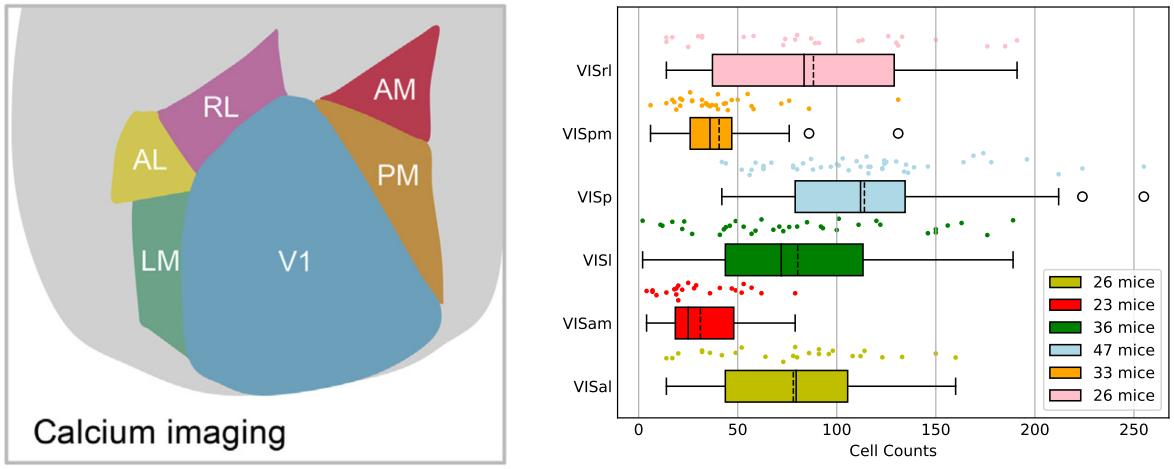
This project aims to explore and investigate neuronal recordings of mouse visual cortex dataset from Allen institute via different data analysis methods. The data analysis component involves working with Allen Institute recordings. These recordings has observations of neuronal population of the mouse brain's visual cortex under various visual stimuli conditions (Fig. 1a) with resolution of detecting single cells activity across time. The two used measurement techniques are Neuropixels and two-photon $Ca2^{+}$ imaging which have different structures. The data we chose to work with is two-photon imaging data-set [1].



(a) (b)

FIG. 1: **(a)** Mice experienced six distinct stimulus blocks, divided into three separate imaging sessions labeled as session A,B and C, conducted across different days[1] **(b)** Schematic structure of experiments setup for selected stimulus (natural movie) [2]

In order to investigate my research's problem statement which relates to dynamics of brain across time, we needed an experiment which were repeated across different days and the only repeated stimulus in this experimental setup was natural movie (Fig. 1a). Natural movie stimulus which is 30 seconds long has been shown to the animal for 3 different days (session 1, 2, and 3) and each day has been repeated 10 times (Fig. 1b). Each mouse is imaged just for one cortical area (Fig. 2a) and the measurement contains calcium fluorescence signals for each neuron within that area (Fig. 2b).



(a) (b)

FIG. 2: **(a)** Visual cortex areas map : V1 (primary visual area), LM (lateral-medial visual area), AL (anterolateral visual area), PM (posteromedial visual area), RL (rostrolateral visual area), and AM (anteromedial visual area)[2] **(b)** distribution of experiments across different brain areas with different observed neurons (solid line : median, dashed line: mean)

In Figure 3, we present three instances of neurons that exhibit consistent and strong responses when exposed to the natural movie across 30 trials. These neurons seem to encode distinct time frames, capturing various features and information from the movie. Although there are also other neurons that do not consistently maintain their robustness across different trials. These neurons might be involved in more complex information processing or not responsive to the movie at all.

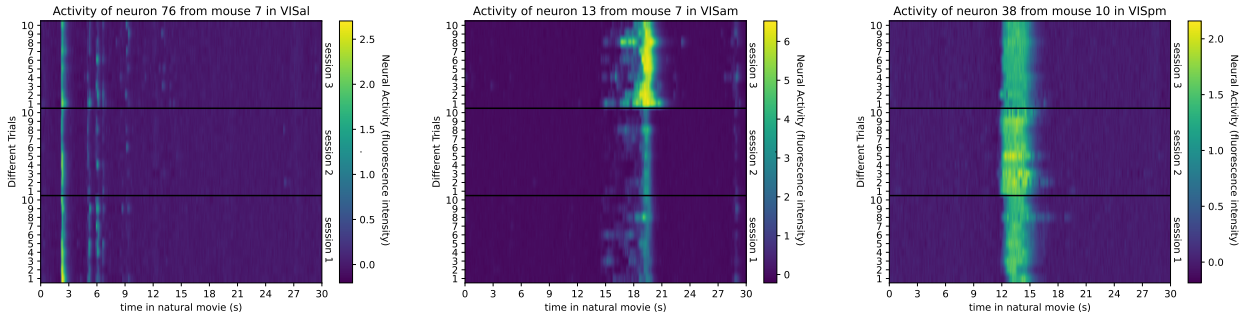


FIG. 3: Three example neurons which have reliable response to the natural movie stimulus.

II. METHOD

In this section, I will detail the data collection process and outline the methods used for analyzing these brain signals.

A. Data Retrieval and Cleaning

For collecting raw data from Allen institute there is an API which I used in python []. I downloaded all experiments available for two-photon calcium imaging and filtered them by the stimulus we are interested in (natural movie one). I created a general class of mouse which includes attributes: id, signals (Called dff traces) and etc. Then I created different instances of mouse class and write data from API to the attributes and finally stored them as .npz (from numpy package) files to begin analysis.

B. Data Statistics & Visualization

To get some primary insight of the dataset, I plotted a boxplot for the variability of cell counts per mouse in different brain regions (see Fig.2b) by reading the cleaned dataset and using matplotlib and numpy packages. I explored neuronal signals of different neurons by plotting imageshow (see Fig.3) and saw different responses across neurons. To find well behaved neurons, I calculated average correlation of each neuron with itself (see Fig.5) and picked the most correlated neurons $(\text{Correlation} > \text{threshold}\,;\text{threshold} = 0.8)$.

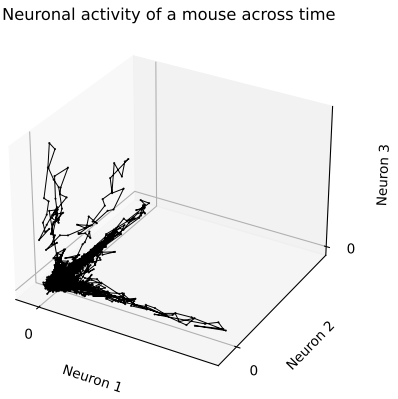


FIG. 4: Activity of a single mouse in high dimensional space

C. PCA Analysis

You can think of the activity of all neurons (for a single animal) as a trajectory in a high dimensional space (Fig. 4) where each point corresponds to a time frame of movie with N (total number of neurons) components in a sense that i^{th} component shows activity of i^{th} neuron at time t. Ideally this trajectory should contain the information regarding the movie and one might think of reducing this trajectory into an smaller subspace. The first approach that comes to mind is PCA (principal component analysis)[3]. I performed PCA on each mouse using sklearn.decomposition

module and categorized them into different brain areas, then plotted the explained variance ratio for each area (see Fig. 6).

III. RESULTS & DISCUSSION

Statistics of cells : According to the method for selecting robust cells (II B) we obtain a scatter plot of correlations of all cells of all mice in different brain regions, for instance we only presented it for the VISp (V1) area (see Fig. 5). The blue error bars are 95% confidence interval for mean correlation of each mouse and population mean (across mice) of correlations is between 0.078 and 0.105 (green box) for VISp with 95% confidence interval. By choosing a threshold on correlation, you can find robust cells (as shown in Fig.3). The other areas have more or less the same statistics and you can find their plots in Github.

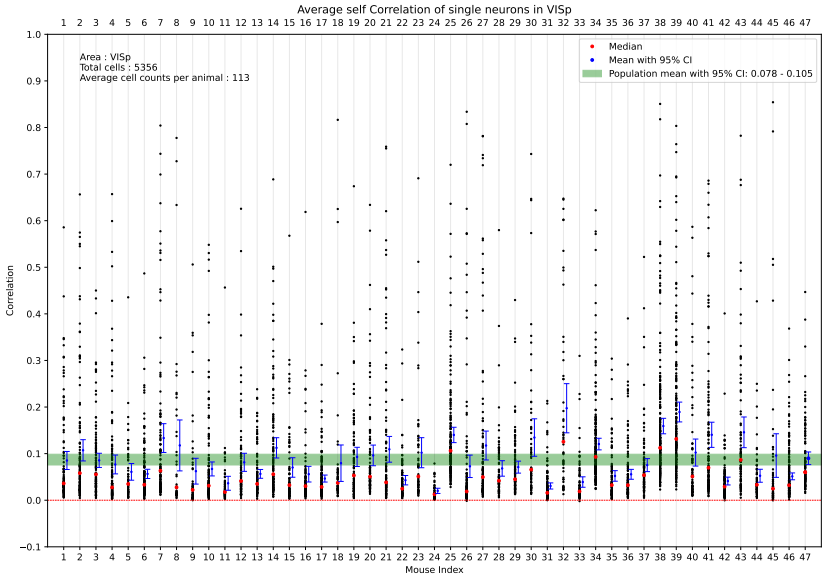


FIG. 5: Cell correlation with itself. Each point corresponds to a neuron and higher correlation means more robustness in activity for the cell.

PCA results:

After performing PCA on each mouse, we end up with a list of variance ratios corresponding to different principal coordinates which are in descending order by default (Fig. 6).

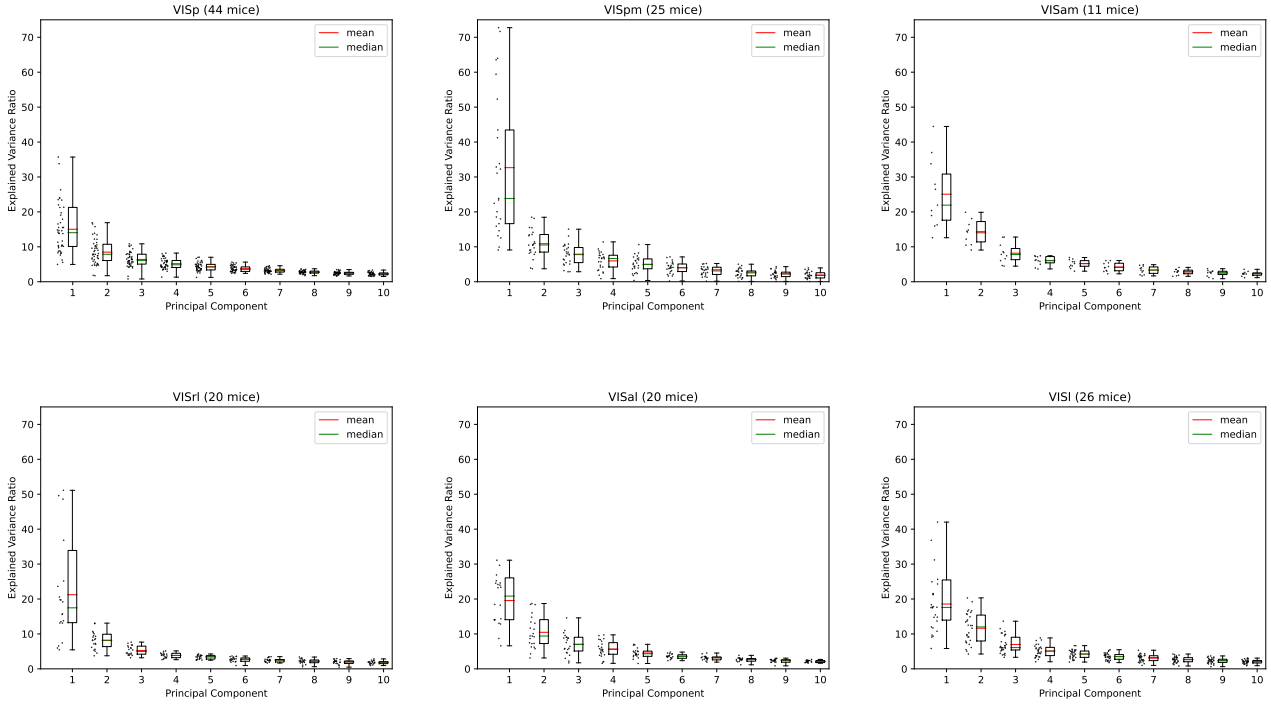


FIG. 6: Boxplots of variance ratio of principal components for mice in different brain regions. Each dot shows one mouse and the red solid lines are mean and yellows are medians of variance distributions

In order to get a reliable and statistical variance ratio curve (Fig.6) we have to take average across all mice and calculate the confidence intervals which brings us to Figure 7.

IV. CONCLUSION

Firstly, Cumulative variance curves of VISam and VISpm reach to 100% sooner and with less principal components explaning them (Fig.7b). On the other hand, The VISpm and VISrl are known to be highly connected to RSP (retrosplenial areas and postsubiculum) cortex and this connectivity make it likely to be more responsible for processing visual information (such as global motion signals) that might be used for spatial navigation[4]. As our visual stimulus is a movie, it is more likely to have more information encapsualtion in the VISpm and the VISrl rather than other areas. Secondly, PCA1 variance of Visual primary (VISp or V1) is lower than other visual areas which needs more research to understand the reason. And finally, we understood all cells are not robust across sessions and this variability can affect the reliability of the whole PCA analysis.

Future Work

a. Categorization of cells into robust across sessions, robust per session, not responsive and drifting cells and after that Computation of ratio of size of each category to the total cells to see if it is consistent across statistics of mice.

b. Exploring what features of movie are encoded in PCA components and neuronal activities in both cell level and population level.

Data & Codes Availability

Cleaned data sets and the applied methods in jupyter notebook format are accessible in Github and there is a readme file to proceed with files.

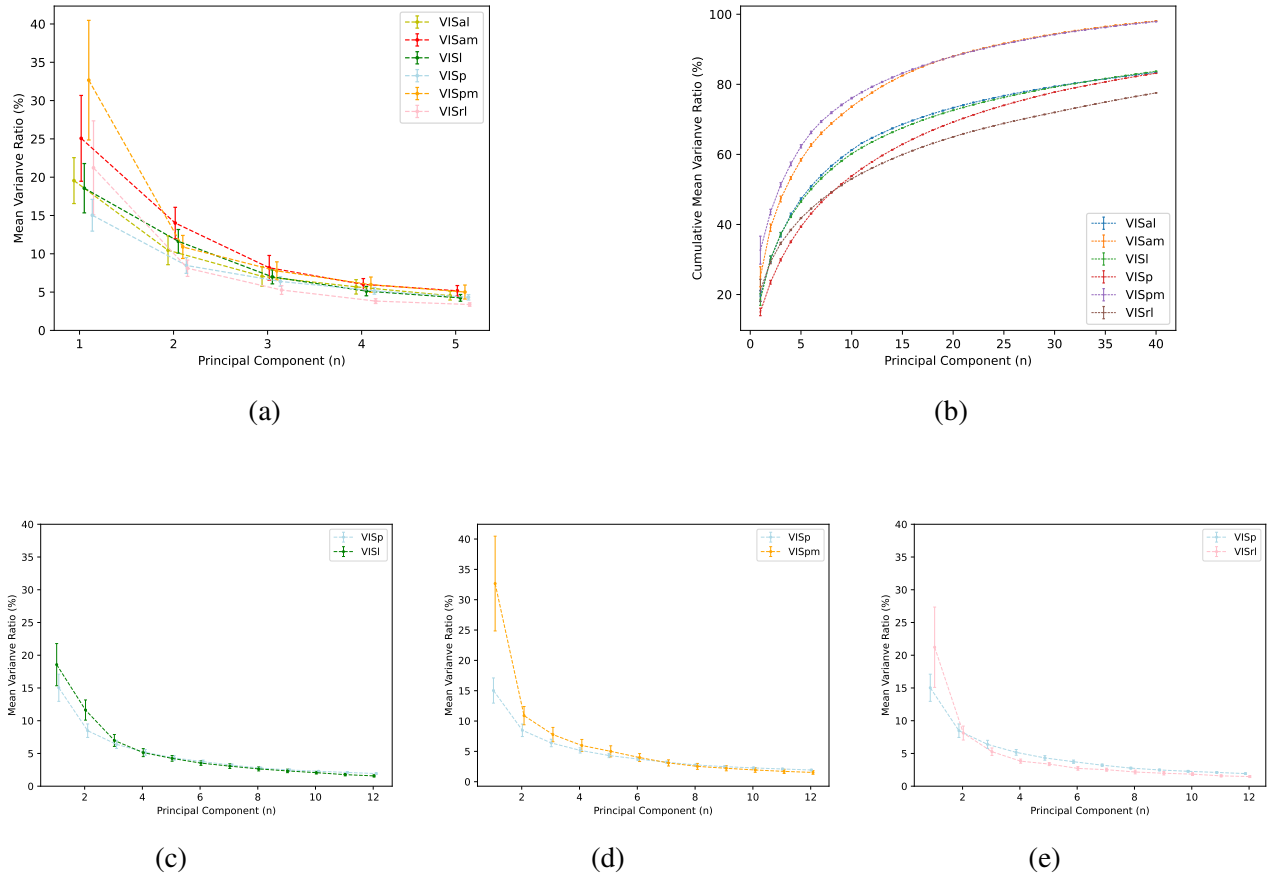


FIG. 7: (a) Population average of explained variance ratio (across) with 95% CI (b) Cumulative averaged variance ratio for 6 visual cortex regions (c) Comparison of VISp vs VISl variance profile(d) Comparison of VISp vs VISpm (e) Comparison of VISp vs VISrl

[1] S. E. de Vries, J. A. Lecoq, M. A. Buice, P. A. Groblewski, G. K. Ocker, M. Oliver, D. Feng, N. Cain, P. Ledochowitsch, D. Millman, *et al.*, A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex, Nature neuroscience **23**, 138 (2020).

[2] D. Deitch, A. Rubin, and Y. Ziv, Representational drift in the mouse visual cortex, Current biology **31**, 4327 (2021).

[3] Wikipedia, Principal Component Analysis, https://en.wikipedia.org/wiki/Principal_component_analysis.

[4] E. Froudarakis, P. G. Fahey, J. Reimer, S. M. Smirnakis, E. J. Tehovnik, and A. S. Tolias, The visual cortex in context, Annual review of vision science **5**, 317 (2019).