Decoding Behaviour from neural data

Neuromatch Academy Project - July 2020

The Neuromatch Academy[6] was a summer school that happened in the summer of 2020. In the midst of a pandemic, the school emerged on an online platform reaching out to everyone across the world. The academy served the purpose of training neuroscientists to learn computational tools, make connections to real world neuroscience problems, and promote networking with researchers.

The academy also provided an opportunity for the students to carry out their own mini-project that would last three weeks. Several datasets were provided for exploration. Our group Les Souris Grises, from the pod 063-Gay-Ladybug have decided to go with the Steinmetz data set. Below us is a description of the Steinmetz data set that we chose.

Dataset chosen:

The dataset used for this purpose was the dataset procured by Steinmetz et al. 2019[1].

The Steinmetz dataset is an **electrophysiological** recording from multiple regions of the mouse brain during a 2-Alternative Forced Choice Task paradigm. **Neuropixel** probes were used to record from approx. **30,000 neurons** from **42 regions**, while the mouse performed a visual discrimination task. In each trial (multiple trials conducted over each session; and a total of 39 sessions), a mouse was placed on a wheel with its **head fixed**, surrounded by 3 screens (left, right and in front). Images of differential contrast were presented to either the left, right or both the screens and the mouse had to turn the wheel in the correct direction in order to bring the greater-contrast image to the front-screen. If there was no image presented on either side, the correct response was to hold the wheel steady for 1.5s. Neural activity was continuously recorded for the entire duration of the task.

The locations in the dataset have been mapped according to the Allen Mouse Brain Atlas[4]

Code for the analysis was written in **Python**, with the help of scientific packages; **Numpy, Scipy, Matplotlib, Neural_Decoder, Sklearn.**

To load the data into our notebooks for further analysis, we used some code[2] provided by **Dr. Marius Pachitariu**.

For a detailed description of the dataset see this[7] document by Dr. Nick Steinmetz and this[8] writeup by us.

Questions:

Can we predict the rodent's movement based on activity from different brain regions before or after the movement has occurred?

Our questions were inspired by some of the original questions from some exemplary projects[3].

Scope:

Upon discussion with our mentor, Dr. Nathalie Rochefort, we came to the conclusion that for the short time span, our original question was **too ambitious**. So, we narrowed down our question to looking at only **three regions** and trying to establish how these two interact with each other, in terms of similarities or dissimilarities in their temporal activity.

We decided to focus on one rodent's data, from one session (Session no. 11). Our chosen regions were the **visual cortex area** , **thalamus region**(with or without visual thalamus area(LDg, LP) and the **motor areas** (MOs, MOp). These two are **anatomically and functionally distinct**, hence they were our best shot at investigating differences in activity.

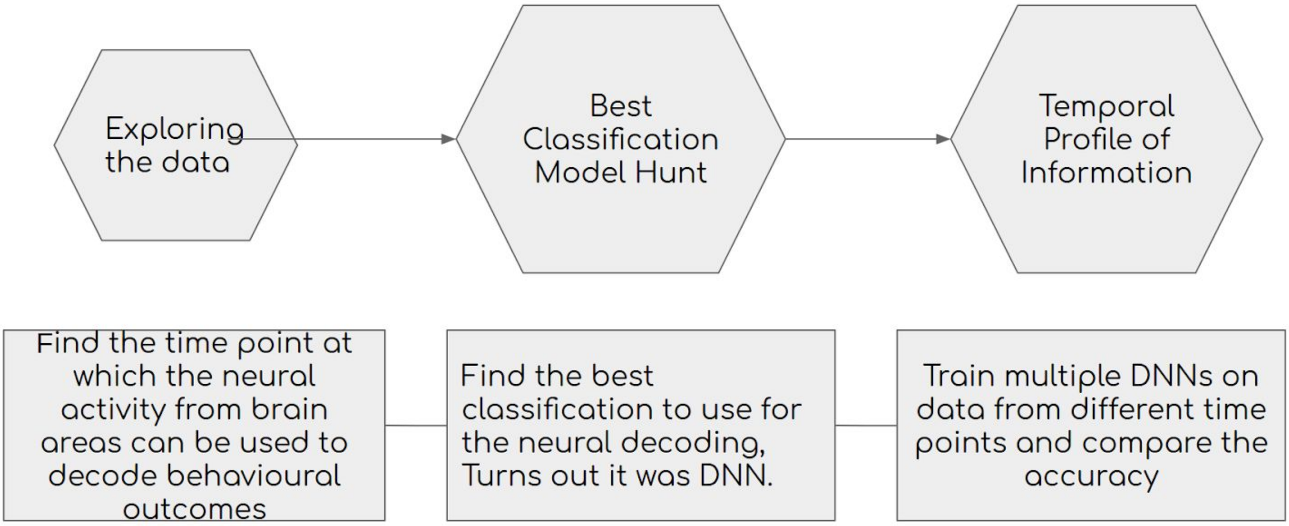
We believe that the dataset has enough potential, and given enough time we can build **better models** to represent the **temporal profile** of regions. However for the time being we shall settle on the simpler task.

Approaching the problem:

Dataset exploration and explanation:

The neural data we had was a time series data of **spike waveforms** for all the neurons from each of the 42 regions. Since we wanted to look at the activity of a population of neurons from 2 regions, we needed to obtain the average activity of all neurons.

METHODOLOGY

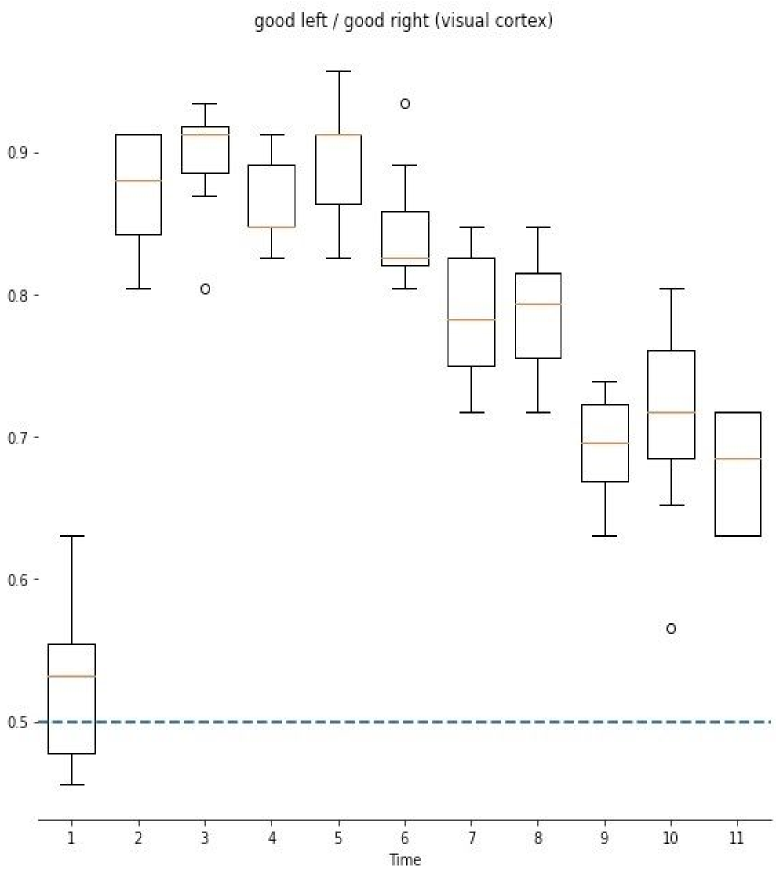


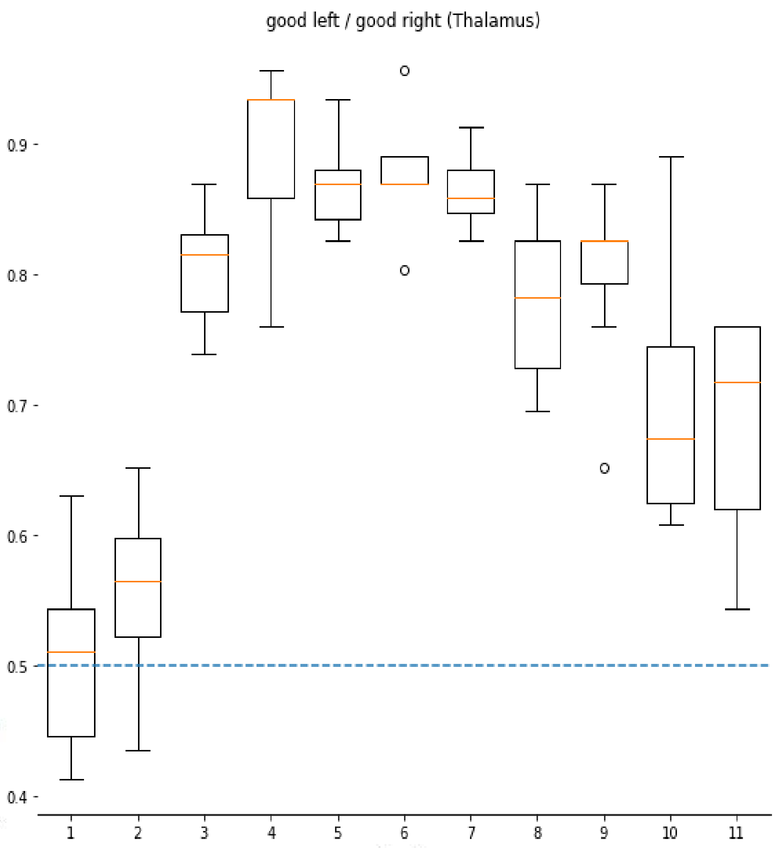
Analysis Pipeline:

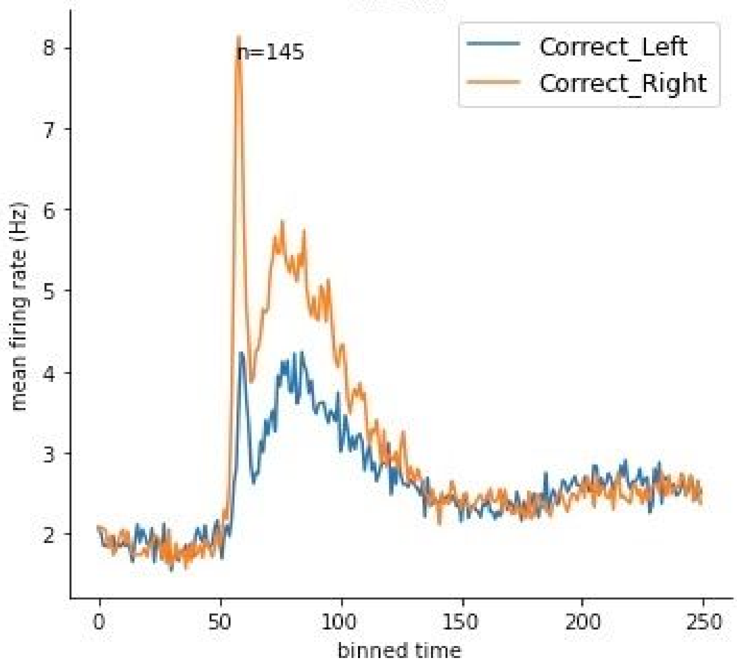
- Find the best classification model for neural decoding.
- Obtain trails where there only has been right/left movement.
- Obtain trails where the rodent has chosen the correct choice.
- Obtain trails where the rodent made a motion and no-motion.
- Obtaining the **box plots** showing the time course of the average activity of the 3 regions of our interest.
- **Binning** the data into 20ms bins and plotting the cross validated data over time

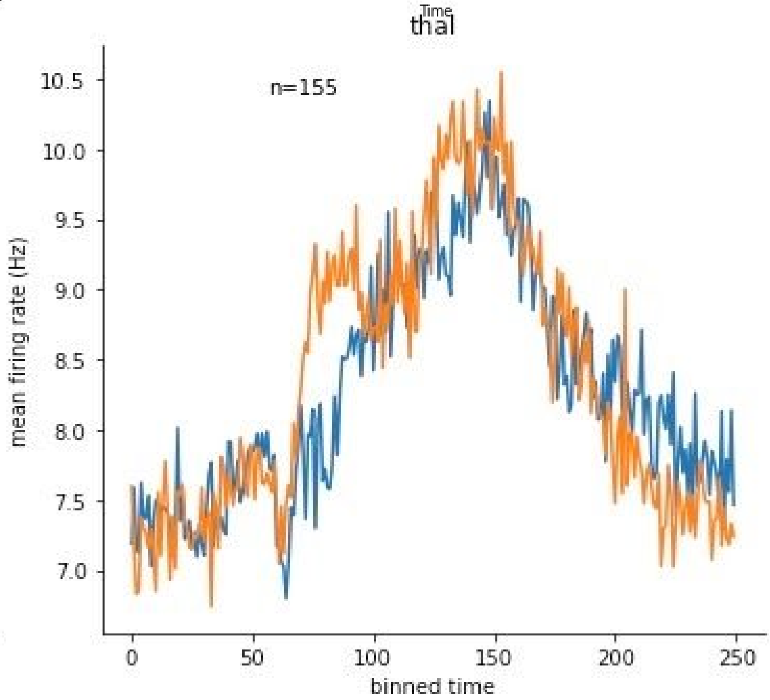
Results:

All our results and processes involved can be found in our GitHub repo[5]
Our initial tasks were to extract the data into a more interpretable form. We successfully did that with some help from some helper code and then we had to find the best classification model and it turns out **DenseNN** is the best bit for that(out of Simple RNN, Logistics Regression, LSTM Classification, Wiener Filter Classification, SV Classification, GRUClassification, XGBoost Classification.





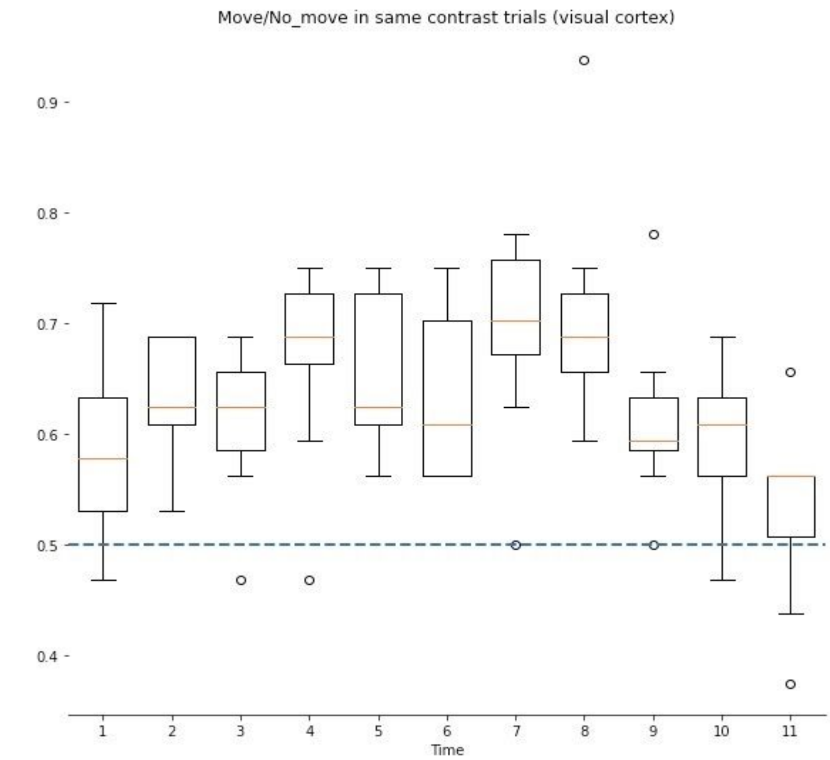


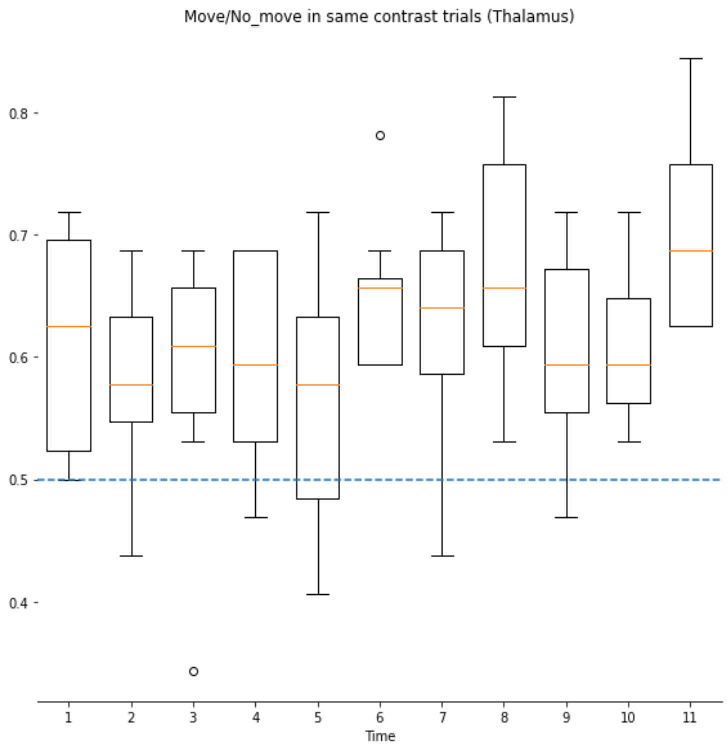


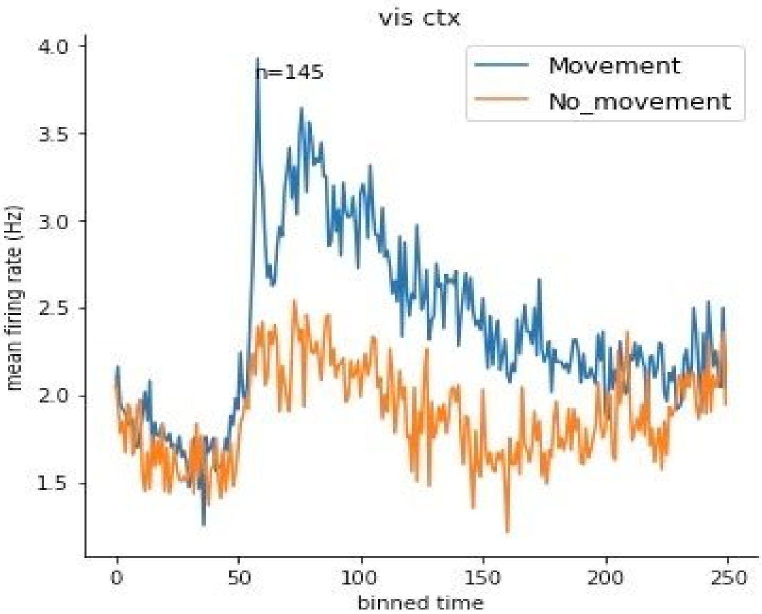
The above is one of our plots that we got. On the Y axis we have labeled the accuracy for a particular session and on the X axis we have the time bins (binned at 20ms width). For the above plot, we considered only trails in which the rodent has chosen the correct choice for training our DNN model.

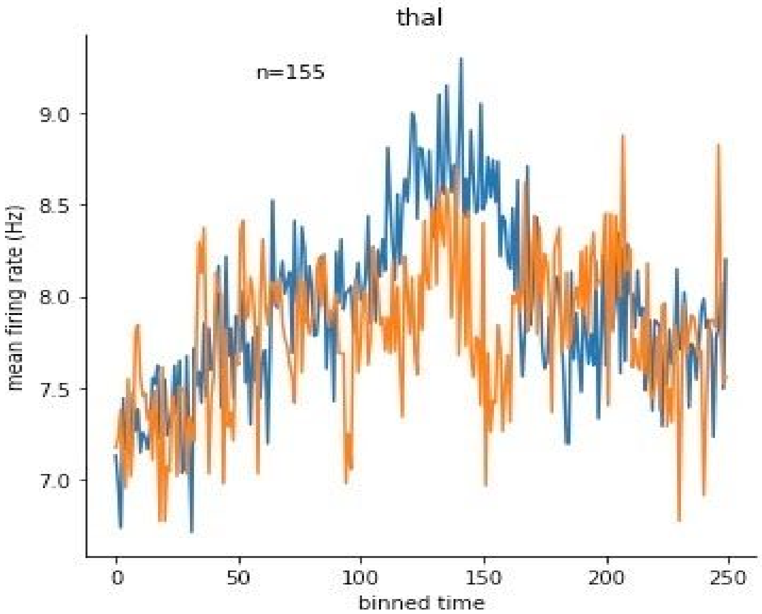
It turns by using visual cortex we can have a maximum median accuracy of around **90%** while in the case of thalamus it's around **95%.**

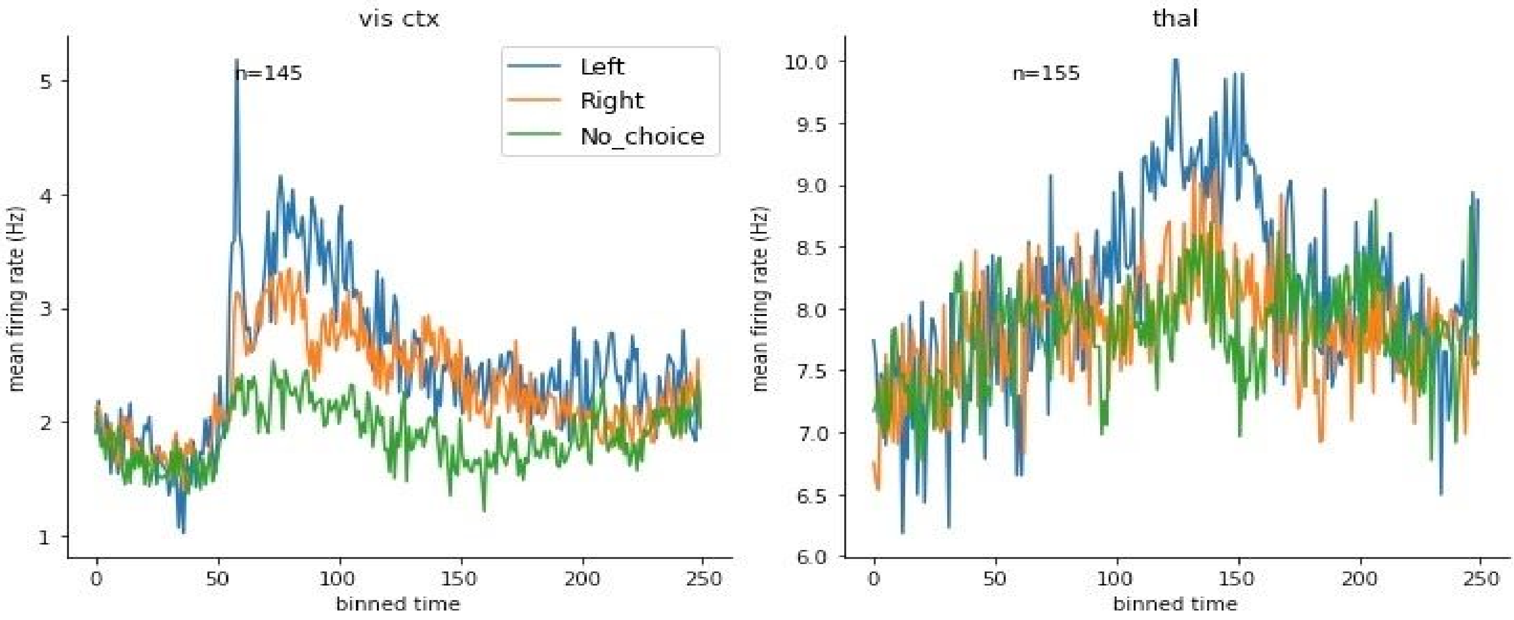
Apart from the other graphs we also decided to plot an average plot for a particular region over all Move trials / No Move plot is presented below.





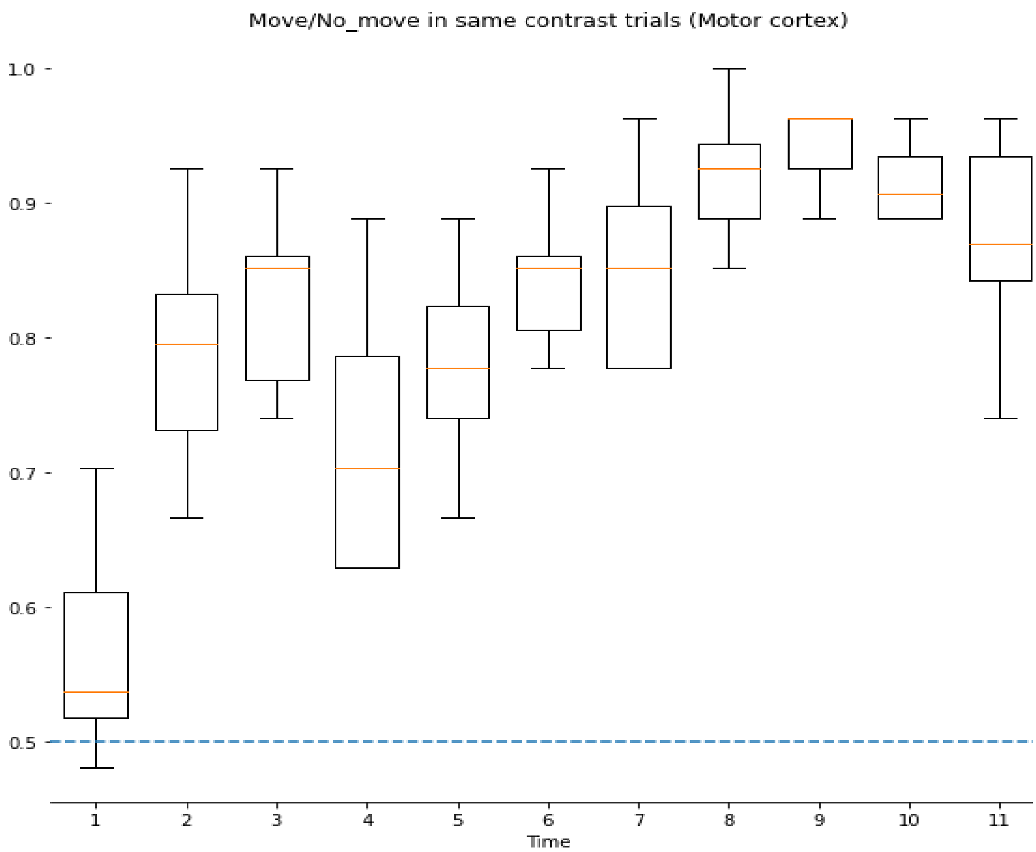


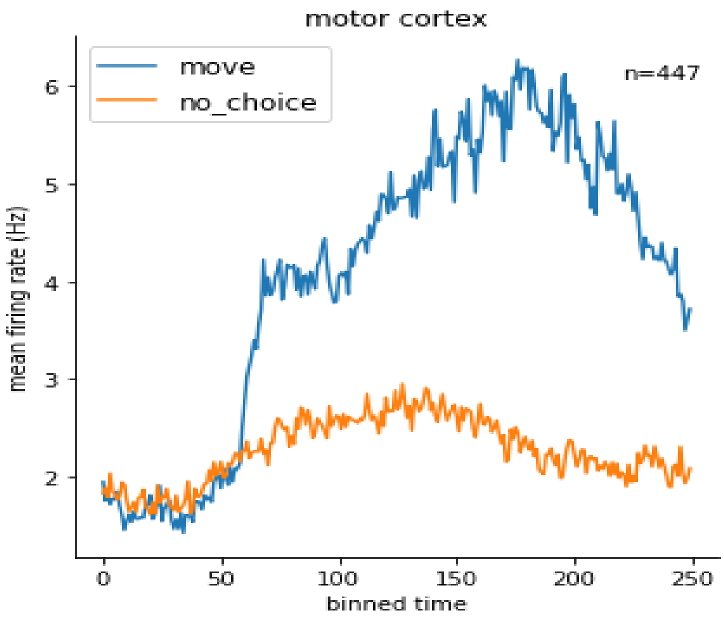


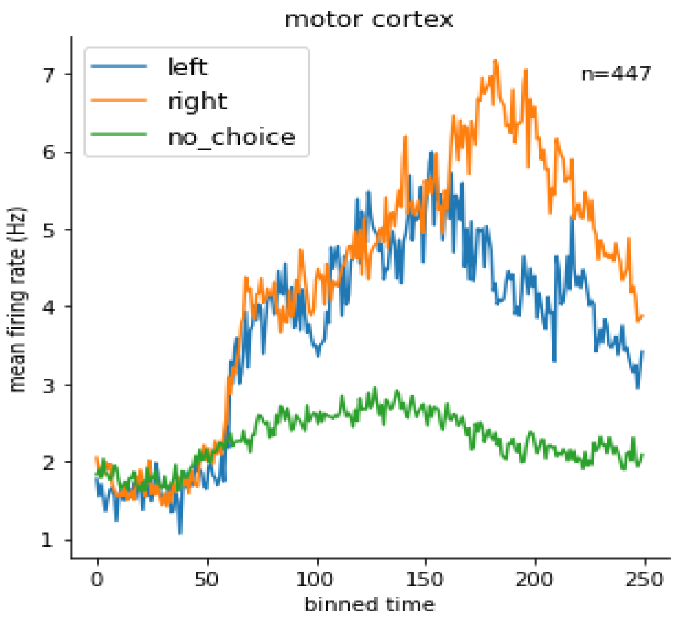


It can be seen from the above two plots that there is quite a varied **difference** between the responses in general from the regions during the two different outcomes. The median accuracy obtained from visual cortex is about **70%** while in the case of thalamus it's even lesser.

For obtaining more median in this case we should use motor cortex cells so the below plotted graphs is for the motor cortex cells (MOs and MOp).







It can be seen the accuracy has improved significantly in the case of Motor Cortex Region, It's around **95%.**

Conclusions:

- We conclude that the visual cortex contains information about the value of the stimulus presented (before movement onset).
- We conclude that there a temporal component to this value information, and that maximum value occurs after stimulus onset
- We conclude that there is a very good signature of movement in the motor cortex, and a weaker signature in other brain areas.

People:

- Nathalie Rochefort (Mentor)
- Zane Mitrevica (TA)
- Kun-lin
- Omika Wadhwa
- Piyush Chauhan

Resources:

1. Dataset:
   https://figshare.com/articles/steinmetz/9598406

2. Code by Dr. Pachitariu:
   https://github.com/MouseLand/steinmetz2019_NMA
3. Example Projects PPT:
   https://docs.google.com/presentation/d/1WAHfJcBPM4rmwwvreAAS92sRYtltJRwklxH-82NzCYo/preview?pru=AAABc3cRwPE*S0Y87T5BNFvf9wvSREdLUQ&slide=id.p
4. GitHub Repository:
   https://github.com/piyushchauhan1/Behaviour-Decoder
5. Neuromatch Academy
   https://www.neuromatchacademy.org/
6. Nick's Explanation of the data.
   https://github.com/nsteinme/steinmetz-et-al-2019/wiki/data-files
7. Our explanation of the data.
   https://docs.google.com/document/d/1iDkVdRwfNnwH7mg--xJijK5lbxgtE879HlcwWts3yGw/edit#heading=h.lf7ihnqn5bgz