
Neural Data Analysis

DONlab

Athul Vijayan(ED11B004)

Contents

1	Introduction to motif analysis in neurons	2
2	Experiment	2
2.1	Setup	2
2.2	Data	2
3	Study of Reliability	3
4	Study of principal components	4
4.1	Reconstruction	5
4.2	OSI analysis on data in Eigen space	5
5	Searching for subsequences	6
5.1	ACF Gram	8
5.2	Longest common subsequences	8
5.2.1	Longest common segment set for neuron responses	8

1 Introduction to motif analysis in neurons

We aim to study how neurons encode the information in V1. Motifs are recurring patterns in the response of a single neuron or a population of neurons. If we take several trials of same stimuli, we try to extract what is invariant in these. The simplest example is getting similar spike rate in every trials for a neuron. This is called reliability of a neuron or a population. A neuron is reliable implies it produces similar responses to a stimuli every time it is presented.

As in the next section will show, The responses of single neurons are not very correlated; or the time series responses of single neurons are reliable. Looking further, we search for recurring subsequences/ motifs. In a population of neurons, It will be a multidimensional subsequence. Each section will dig deeper into topics.

2 Experiment

2.1 Setup

Calcium images were scanned using a Two-photon calcium imaging device on mice injected with GCaMP6f. Firing rates of these cells were then inferred using a fast nonnegative deconvolution algorithm on the Calcium imaging data.

Visual stimuli

Creation of noise movies.

“We developed an algorithm that allowed us to create noise images with a user-defined spectral slope. To do so, we took advantage of the inverse-square law: $P \sim k^{-\alpha}$, which translates to a circle with radius α in two-dimensional Fourier space. Thus, we constructed all noise movies in the Fourier domain. We first defined a matrix of the same size as the original image (256×256 pixels) and then created a noise amplitude spectrum as a 2D circle of radius α , with α taking values from 0 (K0 movie) to $\sqrt{2}$ (K2 movie). This was due to the squared relationship between the amplitude spectrum and the power spectrum. To create the final noise image, we combined this noise amplitude spectrum with a random phase spectrum, where phase values were randomly sampled from the range $0 - 2\pi$. The final noise images were visualized by computing its 2D inverse Fourier transform. Each frame of the noise movie was created using a new random seed, and as a result, the raw noise movies had no temporal correlations between frames. Noise-masking procedure. Figure 1 provides a schematic of the noise- masking procedure. First, each frame of a natural movie was decomposed into its Fourier components (phase and amplitude) via a 2D fast Fourier transform implemented in MATLAB. Next, a noise image was created as described above. The phase spectrum of the original movie was then combined with the amplitude spectrum of the noise movie. The resulting image was then inverse Fourier transformed to yield a noise-masked movie frame. This procedure was repeated for all frames. We used a total of five different natural movies, each 4 s in duration, from the van Hat- eren movie database.” [Sur and Rikhye 2015]

2.2 Data

MATLAB datafile AmpMov.mat contains the following fields.

Experiments are done various days, the data corresponding to each day are in each folder.

Subscript _nat corresponds to responses to video stimuli for original natural scenes video. similarly subscripts _K0, _K_1, _K1_5, _K_2 denote responses to manipulated natural scenes video stimuli.

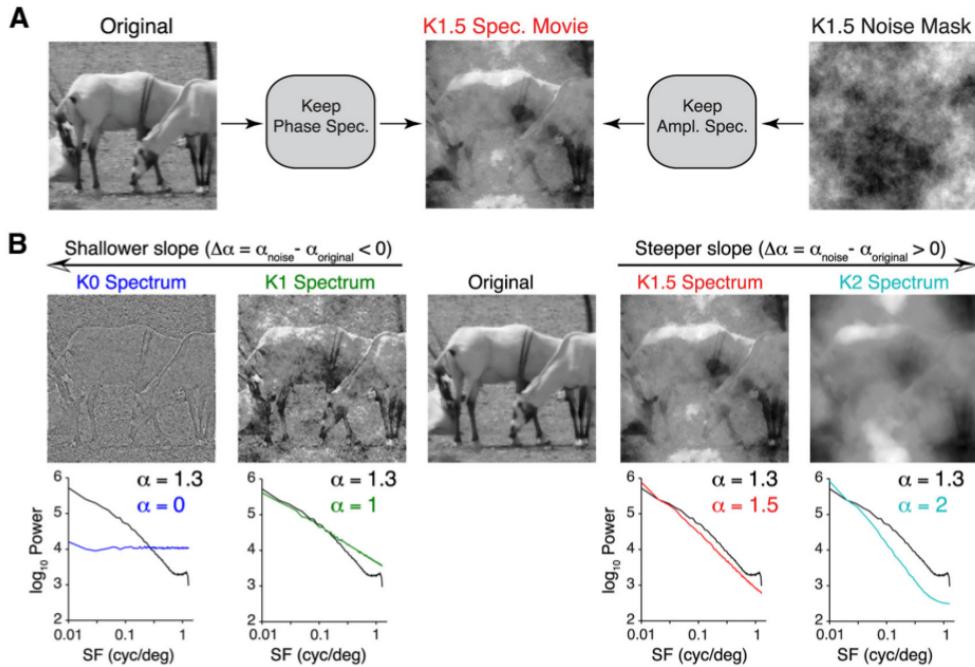


Figure 1: Process of making noise movies

There are 6 different video stimuli. Each of them have 5 trials of experiment. Each trial is presenting a video sandwiched in gray non-stimuli. In each of this trial, every neuron sampled produced a time-series of 200 samples as response.

Field	Description
Sorted.SpikeRate	No doc
Blank	Dimension 47 x 16800
NumNeurons	Number of neurons sampled
NumMovies	Number of movies used as stimulus
M_nat	Dimension 5 x 47 x 1200
MT_nat	Dimension 5 x 47 x 200 x 6
MTA_nat	Dimension 5 x 47 x 200
MTNA_nat	Dimension 5 x 47 x 1200

3 Study of Reliability

Here we analyze how reliable are the time series responses produced by a neuron to various trials of same stimuli. We would like to bring out a metric for quantifying reliability of a neuron. Also we analyze the orientation selectivity properties of these ‘reliable’ neurons.

Plots of responses to various trials of same stimuli are given in Figure . The responses are not similar/ have less correlation. This demands further study to find what is conserved between trials. To quantify a neurons reliability, we use following reliability measure. Response reliability to movie

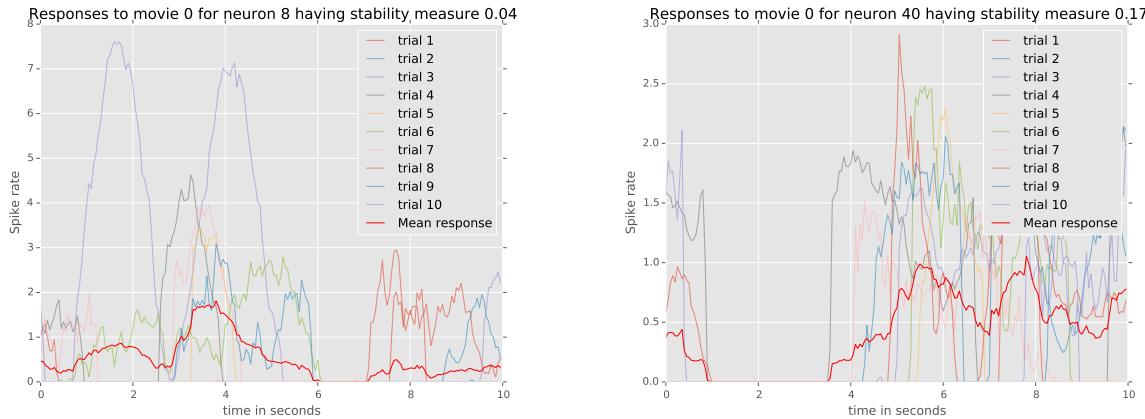


Figure 2: Responses of two neurons to various trials of same stimuli.

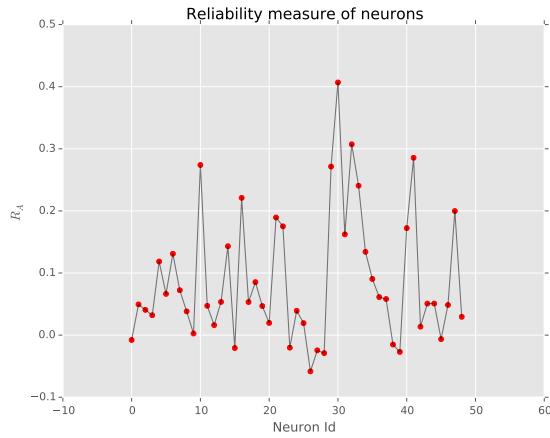


Figure 3: Reliability measure R_A for different neurons

A (R_A) is calculated using the following equation:

$$R_A = \frac{2}{T^2 - T} \sum_{i=1}^T \sum_{j=i+1}^T \rho(f_{i,A}, f_{j,A})$$

where $f_{i,A}$ is the response of neuron to i^{th} trial of movie A and ρ is the Pearson correlation.

From figure 3 is seen that the above measure is very low (< 0.4) for most of the neurons. Which implies that the time series response is not the invariant structure we are looking for.

4 Study of principal components

It is known that there are three kinds of cells in V1 - simple, complex and orientation insensitive cells. Are there any correlation within these neurons? Is there redundancy in the code? We do an eigen analysis on the responses of all neurons in a mice to find how many uncorrelated dimensions

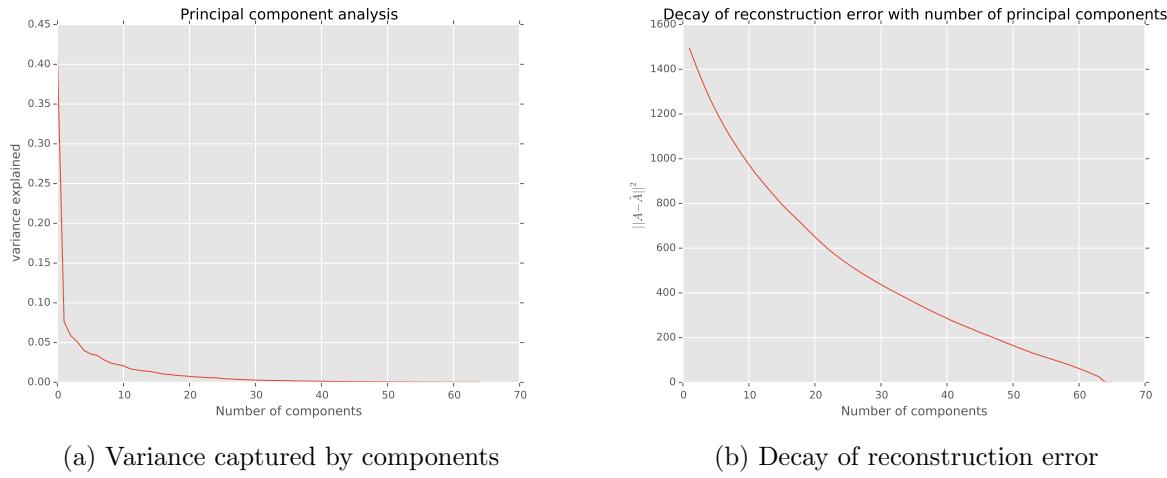


Figure 4: Principal Components Analysis.

are required to capture most of the variance.

Here we take number of neurons as feature dimension. The responses are averages across trials to form each point in the feature space. Every such time series responses were stacked to form the rows. Then PCA is done on the data to find number of components to variance explained ratio. Figure 4a shows the result for a mouse towards a sinusoidal grating stimuli.

It is evident that original basis contain correlation/have features that doesn't contribute to variance in data. We project the original data to a smaller Eigen basis which capture most of the variance. Two studies are possible with this:

1. Recreate the initial data from subset of principal components- Check how many principal components it takes for reasonable reconstruction.
2. Do Further analysis like Orientation selectivity analysis with these data in Eigen space.

4.1 Reconstruction

We study the effect of number of principal components on information capacity on Figure 4b. As we increase principal components, the error should decrease. And finally as the number components equal the original feature dimension, reconstruction error should vanish.

4.2 OSI analysis on data in Eigen space

The transformed data - the data with uncorrelated basis is studied here as if they are produced by neurons in V1. We call it 'Eigen neurons' here. Will we be able to find an uncorrelated set of neurons?

To know if such a study makes sense, we plot average response with orientation. If it forms a double Gaussian, we can go further. Figure shows response to orientation and clearly it does not form double Gaussian. Thus, further study in this regard is not proceeded.

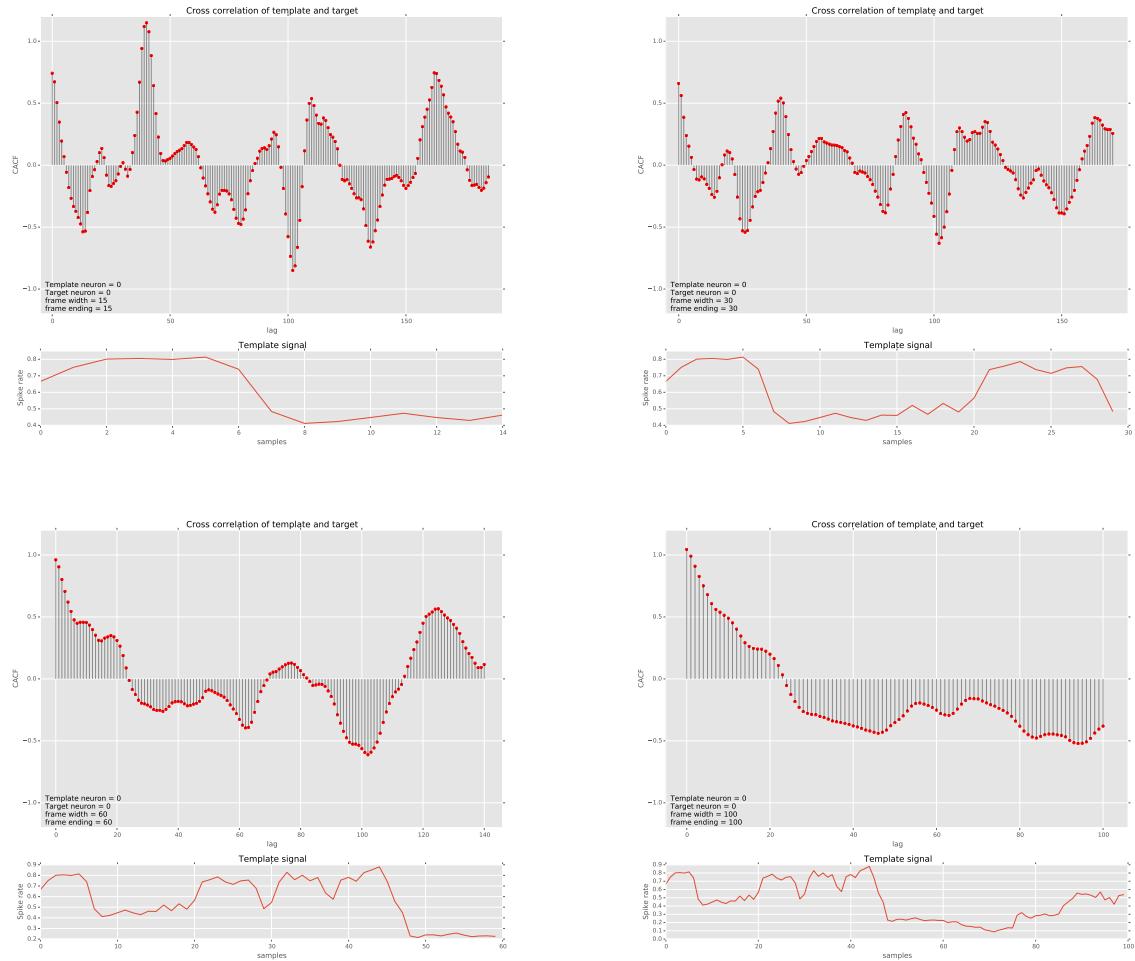


Figure 5: ACF plots for template and target neurons as same.

5 Searching for subsequences

If the whole sequences are not reliable, could there be subsequences that are reliable? These motifs can repeat within response of the same neuron, or in a different trial, or in a different neuron. In this section, we search for such segments using ACVF.

For the scope of this section, ‘template’ neuron is the neuron from which we take a subsequence to check for its other occurrences in the response of a ‘target’ neuron. When template and target is same neuron, we study repeated occurrences of a subsequence in a neuron’s response.

Here, the search space is exponential in length of response. So before arriving at a method to reduce the search space, we analyze ACF with manually set template subsequence. Trial averaged responses of two neurons- template and target are taken, then a subset of former is extracted by frame ending and frame width. This is compared with target sequence as a function of lag. The results for this study with a subsequence starting from beginning of signal and of various widths are in Figure 5 for same neuron and figure 6 for different neurons.

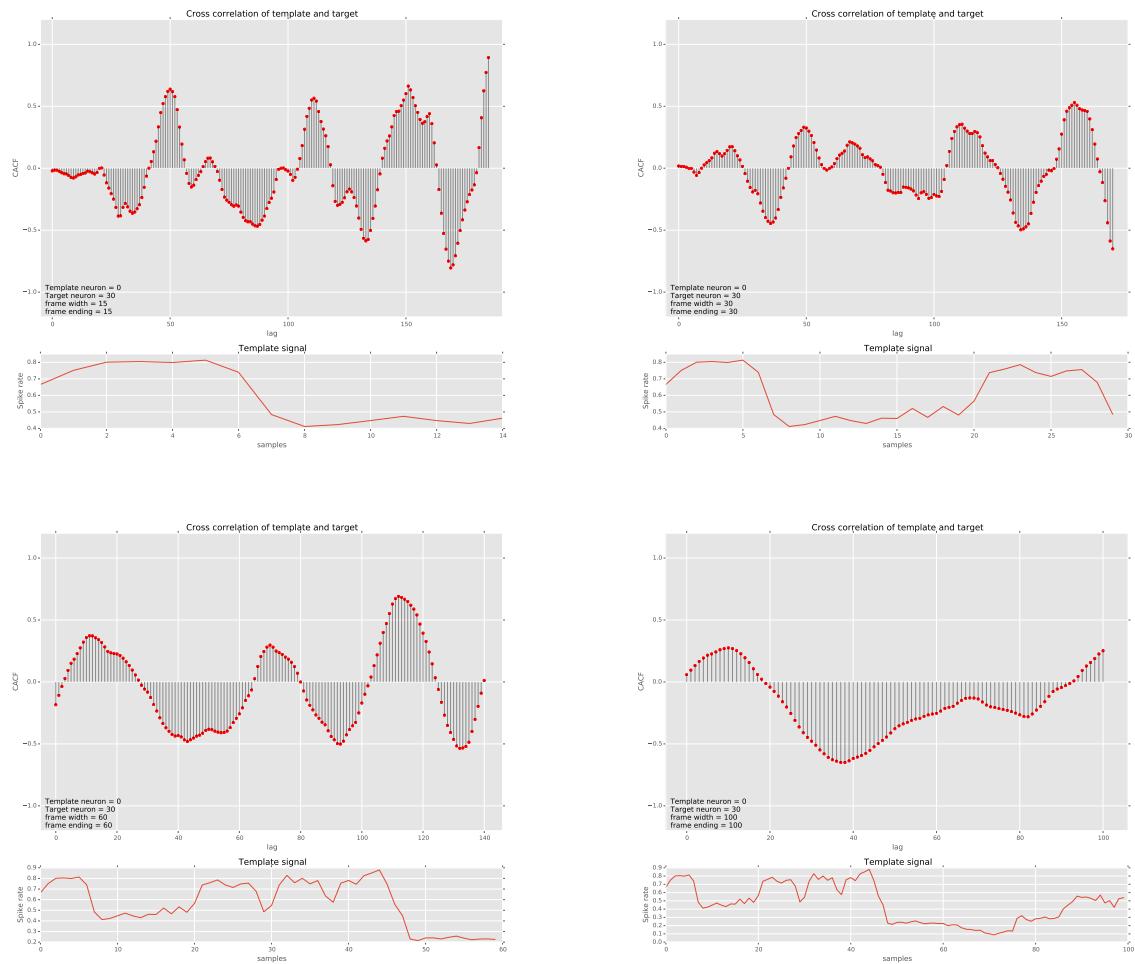


Figure 6: ACF plots for template and target neurons as different .

5.1 ACF Gram

In the previous study, we fixed the beginning of the frame in template and changed the width of frame. Here we will slide the frame also to create a temporal dimension too. Heatmaps of the ACF gram - a time vs lag vs ACVF plot are given in figure 7 for same neuron and figure 8 for different neurons.

Images show results for for different manually fixed window widths. The following inferences can be taken:

- For small window length (≈ 15), the template is small. Such a small template does not satisfy as a motif. Also as the samples are less, estimate of sample correlation will be poor.
- For window lengths ≈ 30 we see patterns. These shows that there are repeating subsequences in the response.
- The repeating patterns seem to be reappearing periodically in lag.
- For window widths ≈ 60 , there are two subsequences that match. Above this, Only match is for lag 0.
- There are common subsequences across different neurons.

5.2 Longest common subsequences

In previous study, we fixed the window widths and saw that there are in fact repeating patterns. Here we try to find set of common subsequences in neurons. Also, till now we used Pearson correlation to compare two signals. We look into other ways of comparing two responses too.

5.2.1 Longest common segment set for neuron responses

Here we compare responses of two neurons- template and target neurons. Both signals are time series signals obtained by averaging responses to each trial of the same stimuli. Subsequence matching needs to be performed between these two signals. The number of similar portions could be more than one and spread across the entire signal.

Let $\mathbf{X} = \langle x_1, x_2, \dots, x_n; x_i \in R \rangle$ be template signal and $\mathbf{Y} = \langle y_1, y_2, \dots, y_m; y_m \in R \rangle$ be the target signal. Now similarity between two points has to be decided. Following options of similarity measures are considered.

1. Euclidean distance

$$\text{sim}(x_i, y_j) = (x_i - y_j)^2$$

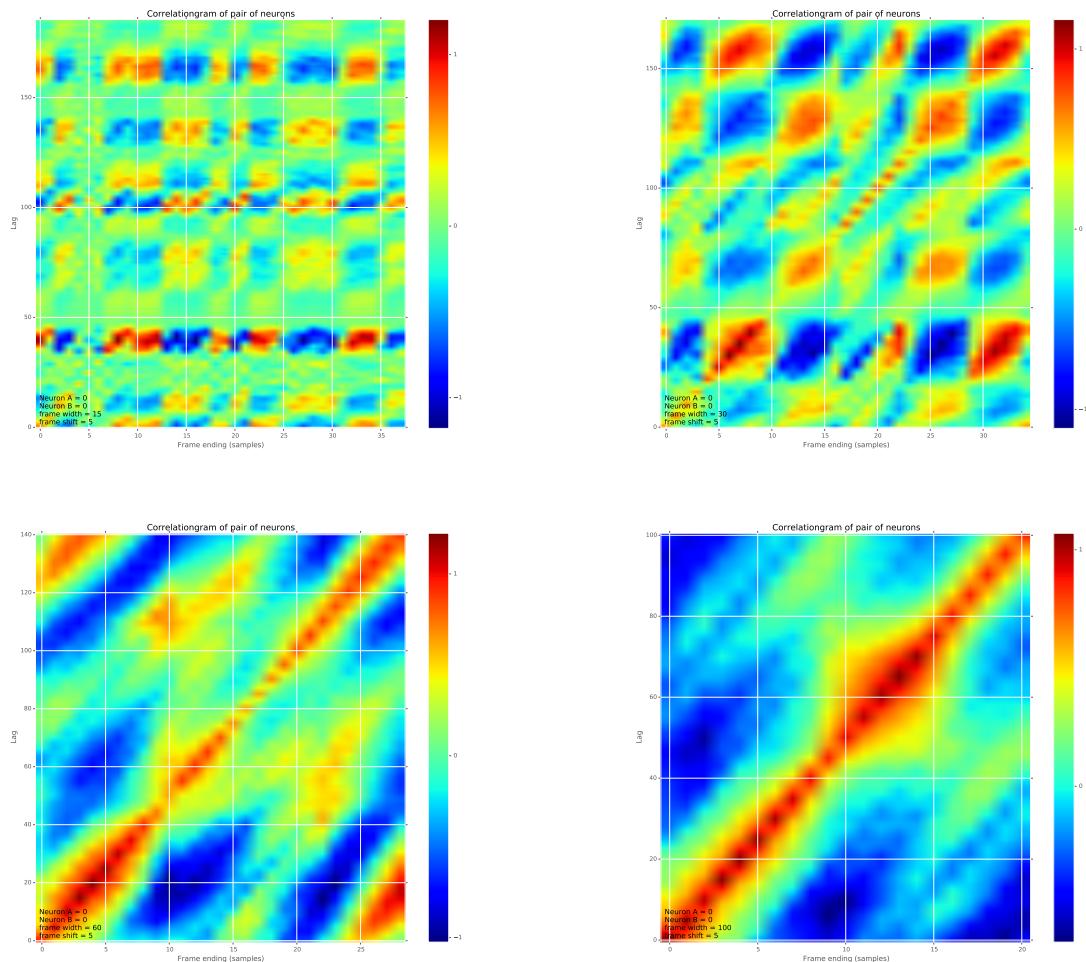


Figure 7: ACFgram for template and target neurons as same.

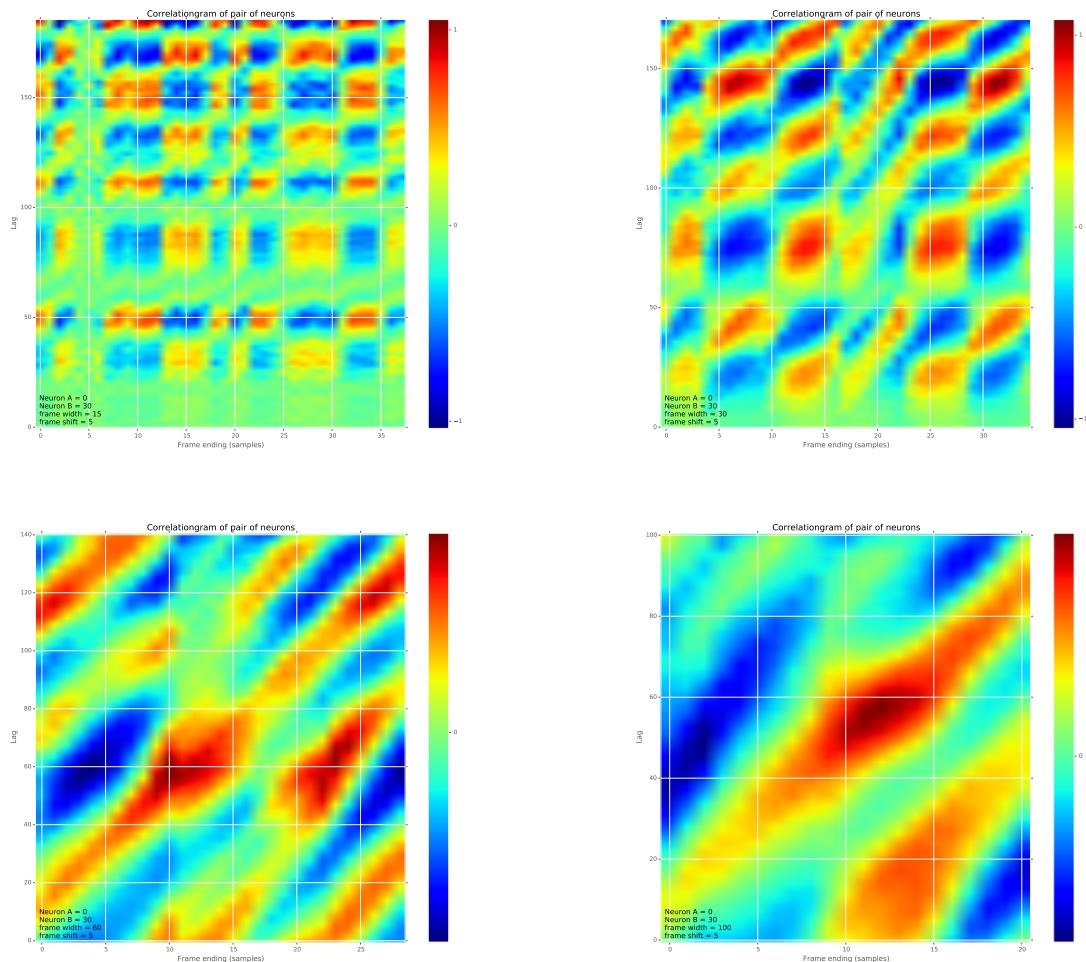


Figure 8: ACFgram for template and target neurons as different .