

Motivic analysis of neuronal responses to visual stimuli

A Project Report

submitted by

ATHUL VIJAYAN

*in partial fulfilment of the requirements
for the award of the degree of*

MASTER OF TECHNOLOGY



**DEPARTMENT OF ENGINEERING DESIGN
INDIAN INSTITUTE OF TECHNOLOGY, MADRAS.**

April 2016

THESIS CERTIFICATE

This is to certify that the thesis entitled **Motivic analysis of neuronal responses to visual stimuli**, submitted by **Athul Vijayan**, to the Indian Institute of Technology, Madras, for the award of the degree of **Master of Technology**, is a bona fide record of the research work carried out by him under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. Hema A. Murthy
Research Guide
Professor
Dept. of Computer Science
IIT-Madras, 600 036

Dr. Srikanth Vedantam
Research Guide
Professor
Dept. of Engineering Design
IIT-Madras, 600 036

Place: Chennai

Place: Chennai

Date:

Date:

ACKNOWLEDGEMENTS

Foremost, I would like to express my sincere gratitude to my guide Dr. Hema A. Murthy for the continuous support in the work, for her patience, motivation, enthusiasm, and immense knowledge in the field. I thank my committee members, especially Prof Srikanth Vedantam, who gave the liberty to pursue a project of my choice.

I would like to thank Dr. Mriganka Sur, MIT for inspiring me to the field of Neuroscience and for the opportunity I received to work in collaboration with Sur's lab, MIT. I thank Rajeev Rikhye for sharing experimental data and his valuable guidelines.

I am thankful to DON Lab family for a wonderful time and great support. I thank my parents for supporting throughout my life. I would like to thank Midhila Mohan for keeping me motivated in pursuing my dreams.

I would like to thank everyone I met in my life.

ABSTRACT

KEYWORDS: Motif, Neuronal data analysis, Neuronal correlations, Orientation and directional selectivity, Primary visual cortex, RLCS, ACFGram, Reliability of neurons

Neurons in primary visual cortex (V1) are known to be orientation and direction selective. Even with the understanding of the characteristics of neurons, a reliable information encoding mechanism in V1 is not well understood. Individual neuronal responses are known to be less reliable. Reliability is believed to exist in population responses rather than an individual neuron. Inter-neuronal correlations are great way to study information coding in a population of neurons.

Motifs are well studied in the context of music. Motifs in Carnatic music is based on the concept of *raga*. Melodic motifs are signature phrases that give *raga* its identity [Ishwar *et al.* [2013]]. We define motif in neuronal responses as subsequence of a response signal having significant length and repeats itself in other neuronal responses.

In this study, we analyses neuronal responses for the existence of motifs. The best motifs are then extracted with a signal matching algorithm - RLCS. The study showed presence of motifs across trial and across neurons in a mice. Presence of repeating subsequences with significant length provides more inferences on individual neuron reliability and inter-neuronal correlations. A new correlation measure for studying the Inter-neuronal similarity is proposed.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	i
ABSTRACT	ii
LIST OF TABLES	v
LIST OF FIGURES	viii
ABBREVIATIONS	ix
NOTATION	x
1 Introduction	1
2 Background and Previous work	3
2.1 Visual pathway in brain	3
2.1.1 Neurons	4
2.1.2 Visual pathway	5
2.2 Experiment setup	7
2.3 Previous works	9
3 Analyzing neuronal properties	11
3.1 Quantifying Orientation and Directional selectivity	11
3.2 Modeling neuronal response	14
3.3 Finding similarly tuned neurons	15
3.4 Study of principal components	17
3.5 Study of Reliability	19
4 Searching for Motifs	21

4.1	Cross-correlation Function	22
4.2	ACFgram	24
5	Rough Longest Common Subsequence (RLCS)	27
5.1	Longest Common Subsequence (LCS)	27
5.2	RLCS	28
5.2.1	Motivic analysis of Neuronal signals with RLCS	32
5.2.2	Normalized RLCS score as a correlation measure	35
6	Inferences and Future work	40
6.1	Inferences	40
6.2	Future work	41
A	Additional results from RLCS	42
A.1	RLCS of intertrial responses	42
A.2	RLCS of responses from two neurons in a mouse	44
A.3	RLCS of neuronal responses from two different mice	45

LIST OF TABLES

5.1 Some examples of LCS in string comparison	28
---	----

LIST OF FIGURES

2.1 Spike train response of an orientation selective neuron. The cell has a preferred orientation which is shown by dashed line. Response reduce exponentially when the stimuli orientation deviates from preferred orientation	7
2.2 With a stimuli aligned with preferred orientation, direction of motion of stimuli also affect the response in some cells. Complex neurons have stronger response to one direction than other. Neurons which respond similar to both directions are simple cells.	8
 3.1 Responses $R(\theta_k)$ of a simple neuron for each θ_k plotted in both orientation and direction space. Equal lobes in direction space shows direction is irrelevant to response.	12
3.2 Responses $R(\theta_k)$ of a complex neuron for each θ_k plotted in both orientation and direction space. Unequal lobes in direction space shows one direction is preferred than its opposite.	13
3.3 Responses $R(\theta_k)$ of a an orientation insensitive neuron for each θ_k plotted in both orientation and direction space. Similar responses to all orientations shows absence of selectivity.	13
3.4 Fit of orientation and direction tuning curves of a neuron. The distinct peak in the orientation tuning curve shows selectivity to orientation θ_{pref} . Different σ_1 and σ_2 in direction tuning curve shows direction sensitive cell. - Thus a complex cell	15
3.5 Fit of orientation and direction tuning curves of a neuron. The distinct peak in the orientation tuning curve shows selectivity to orientation θ_{pref} . Similar σ_1 and σ_2 in direction tuning curve direction of stimuli is irrelevant. - Thus a simple cell	16
3.6 Pairwise Pearson correlation of neurons. Both axes are sorted in decreasing order of OSI.	17
3.7 Orientation tuning curves of some neurons retrieved from correlation study. All of them have a same preferred orientation.	18
3.8 Principal Component Analysis of response of neuron to drifting sinusoidal grating stimuli.	19
3.9 Reliability measure R_A for different neurons	20

4.1	Cross-Correlation function between a target neuron response and a manually chosen subsequence from template neuron.	23
4.2	Extracting frame from template based on frame width and ending.	24
4.3	ACFGram plots for template and target responses from same neuron.	25
4.4	ACFGram plots for template and target responses from different neurons.	26
5.1	Score matrix from RLCS of responses a neuron to different trials.	34
5.2	Extracted motif (longest) from responses a neuron to different trials.	35
5.3	Score matrix from RLCS of responses of two neurons in a mouse.	36
5.4	Extracted motif (longest) from responses of two neurons in a mouse.	36
5.5	Score matrix from RLCS of responses of two neurons in a mouse.	37
5.6	Extracted motif (longest) from responses of two neurons in a mouse.	37
5.7	Demonstration of RLCS score's dominance over Pearson correlation as a metric of similarity. RLCS analysis of two signals having low Pearson correlation. Extracted common subsequence was not detected by Pearson correlation.	39
A.1	RLCS performed on intertrial responses of a neuron. Note the presence of ~ 150 sample subsequence which is undetectable in Pearson correlation test.	42
A.2	RLCS performed on intertrial responses of a neuron. Note the presence of ~ 150 sample subsequence which is undetectable in Pearson correlation test.	43
A.3	RLCS performed on intertrial responses of a neuron. Note the presence of ~ 150 sample subsequence which is undetectable in Pearson correlation test.	43
A.4	RLCS performed on responses of two neurons within a mouse. The two signals have a pearson correlation of 0.18	44
A.5	RLCS performed on responses of two neurons within a mouse. The two signals have a pearson correlation of 0.06	44
A.6	RLCS performed on responses of two neurons within a mouse. The two signals have a pearson correlation of 0.02	45
A.7	RLCS performed on responses of two neurons from different mouse. Note that length of subsequence is smaller than other in cases.	45

A.8	RLCS performed on responses of two neurons from different mouse. This motif <i>across mice</i> last for 3 seconds. (This data is from drifting sinusoidal grating experiment, in which sample rate is 20Hz) . . .	46
A.9	RLCS performed on responses of two neurons from different mouse. Note that length of subsequence is smaller than other in cases. . .	46

ABBREVIATIONS

RLCS	Rough Longest Common Subsequence
LCSS	Longest Common Segment Set
ANN	Artificial Neural Networks
OSI	Orientation Selectivity Index
DSI	Direction Selectivity Index
PCA	Principal Component Analysis
ACF	Auto-Correlation Function
V1	Primary Visual Cortex

NOTATION

$\rho(a, b)$	Pearson correlation between a and b
$R(\theta_k)$	The average response of a neuron to orientation angle angle θ_k .
R_A	Reliability measure of a neuron to visual stimuli A
X	Discrete signal $(x_1, x_2 \dots x_m)$
X_i	Discrete signal $(x_1, x_2 \dots x_i)$
δ	Constant penalty for sample mismatch
τ_{dist}	Distance threshold below which samples are considered as a match.

CHAPTER 1

Introduction

Neuroscience in modern world is an interdisciplinary science collaborating with many other fields like cognitive science, computer science, linguistics, medicine and philosophy. To understand a sophisticated system as the brain requires an interdisciplinary approach. Advances in modern imaging tools and machine learning models like Artificial Neural Network (ANN) enables us to study how the brain performs complex functions like vision. Understanding of the brain will help us design systems that are as powerful as the brain itself. Artificial Neural Networks (ANN) are remarkable machine learning models inspired from the functioning of brain.

Machine learning and signal processing work harmoniously to find patterns and their meaningful inferences from experimental data. The ability of algorithms to learn like the brain is intriguing. In this regard, a better understanding of the brain will help us create complex Artificial Intelligence systems.

Visual perception is a complex function performed by the brain. Visual pathway in brain starts from retina in the eye. Rods and cones are photo-receptor cells in the eye that are connected to the brain through neurons. Understanding information coding in brain will provide insights on neural circuitry in the brain. A neuron is reliable if it produces similar response every time a stimuli is presented to it. If we can perceive a same input correctly each time, there must be some form of information coding. Individual neuronal responses are not reliable due to intrinsic

noise. It is believed that primary visual cortex (V1) codes information in population with inter-neuronal correlations.

Motifs are well studied in the context of music. Motifs in Carnatic music is based on the concept of *raga*. Melodic motifs are signature phrases that give *raga* its identity [Ishwar *et al.* [2013]]. In Genetics, motif is a sequence pattern of nucleotides in a longer DNA sequence. In neuronal responses we define motif as subsequence of a response signal having significant length and repeats itself in other neuronal responses.

This work use signal processing and machine learning approaches to recognize patterns in responses of neurons in primary visual cortex to various visual stimuli. An understanding of the characteristics of neurons in V1 are given, and further the properties are quantified from experimental data. This work study the presence of motifs in neuronal signals. Inter-trial motifs are analyses in search of reliable subsequences in an individual neuron. Inter-neuronal motifs are used in analyzing the correlation between neurons. A better correlation measure is proposed to determine similarity between two neurons. Neuronal signals from two different mice were searched for motifs. These motifs were found to be small in length, yet their presence could be due to similar activation mechanism of neurons.

CHAPTER 2

Background and Previous work

Signal processing and machine learning methods are great tools to bring out valuable inferences from data. The conclusions we draw from statistical methods will have meaning only in the context of the problem. To analyze and draw conclusions from neuronal data, we need to have a good understanding of the background. This chapter provides background in basic Neuroscience and previous works in the similar problem.

2.1 Visual pathway in brain

Brain is the most sophisticated organ nature has devised. The brain is mainly made up of orderly arrays of cells called neurons (also called nerve cells). Neurons are electrically excitable cells which process information through electrical and chemical processes. Neurons are also information carriers that connect the whole body to the brain. Visual perception is a computationally complex function yet the brain performs it very efficiently. There are over 100 million photo-receptors in a human retina, but only around 1 million axons carry information to the brain [Bear *et al.* [2007]]. Understanding how the brain represents and processes visual information can help us create artificially intelligent machines with vision abilities.

This section describes functioning of neurons and background on visual pathway of brain.

2.1.1 Neurons

A neuron consists of three parts - soma, dendrites and axons. Soma is the cell body of neuron housing many organelles including Nissl granules and Nucleus. Axon is the signal transmitting nerve fiber that emerges from the soma. Dendrites are branched fibers that receive information from other nerve cells. The axon can be anywhere from less than a millimeter to a meter or more in length; the dendrites are mostly in the millimeter range. Near the point where it ends, an axon usually splits into many branches, whose terminal parts come very close to but do not quite touch the cell bodies or dendrites of other nerve cells. At these regions, called synapses, information is conveyed from one nerve cell, the presynaptic cell, to the next, the postsynaptic cell [Hubel [1995]]. The function of a nerve cell is to take in information from the cells that feed into it, process that information, and to deliver the integrated information to other cells. The information is usually conveyed in the form of brief events called nerve impulses. The nerve cell is bathed in and contains salt water. Because most of the salt molecules are ionized, the fluids both inside and outside the cell will contain chloride, potassium, sodium, and calcium ions (Cl^- , K^+ , Na^+ and Ca^{2+}). In the resting state, the inside and outside of the cell differ in electrical potential by 0.07 volts. The signals that the nerve conveys consist of transient changes in this resting potential, which travel along the fiber from the cell body to the axon endings. During an impulse, a large number of sodium pores in a short length of the nerve fiber suddenly open, so that briefly the sodium ions dominate and that part of the nerve suddenly becomes negative outside, relative to inside. The sodium pores then re-close, and meanwhile even more potassium pores have opened than are open in the resting state. Both events - the sodium pores re-closing and additional potassium pores opening - lead to the

rapid restoration of the resting state with positive potential outside . The whole sequence lasts about one-thousandth of a second. [Hubel [1995]]

Neuronal responses are quantified by measuring the number of spikes per second - called spike rate. In this work, we measure Ca^{2+} concentration as the response of neuron. As the spikes are caused by the ion concentration changes, Ca^{2+} ion concentration is a good proxy for spike rate. The Vogelstein deconvolution algorithm [Vogelstein *et al.* [2010]] extracts the spike train of each neuron from a raw fluorescence movie. We use the inferred spike rate from the Vogelstein deconvolution algorithm as the neuron's response.

2.1.2 Visual pathway

The brain performs the enormous task of translating light that falls on retina into a meaningful visual scene. The visual pathway starts with eye, of which most important part is the retina. Retina contains light receptors called rods and cones. Rods are far more numerous than cones and are responsible for our vision in dim light but do not respond to bright light. Cones do not respond to dim light but are responsible for our ability to see fine detail and for our color vision. The layer of nerve cells at the back of the retina consist of retinal ganglion cells which connect eye to the brain through optic nerve.

Receptive field refers to the specific receptors that input information into a given cell in the nervous system. In visual pathway, it refers to a region on the retina from which information goes to a particular neuron. Receptive fields of adjacent neurons are usually overlapping. As we go into higher levels of visual pathway, the receptive field gets complicated.

The optical nerve reaches lateral geniculate cells before it joins the primary visual cortex in the brain. About half the fibers coming from a particular eye cross to the side of the brain opposite the eye of origin, and half stay on the same side. Lateral geniculate cells receive fibers from the cerebral cortex also, which plays some role in attention. The primary visual cortex (V1) is a plate of cells 2 millimeters thick, with a surface area of a few square inches. V1 contains three different kinds of cells - simple cells, complex cells and orientation insensitive cells. Neurons in V1 are orientation selective cells which can represent visual information in terms of edges as visual features. [Hubel [1995]]

Neurons in V1 are selective to orientation and direction of a visually contrasting pattern in the receptive field. The orientation is the angle to which the patterns is aligned, direction is the angle to which the pattern is moving in the next frame. Figure 2.1 shows the behavior of orientation selective neurons. They have a preferred orientation to which they produce an optimum response. As the orientation deviates from preferred orientation, the response reduces exponentially. Within orientation selective cells, there are two kinds of cells - simple and complex cells. Consider moving the oriented stimuli perpendicular to the angle of orientation. Among the two perpendicular directions, some neurons respond to one perpendicular direction more than the other. These cells are called complex cells. Neurons which respond similar to both directions of motion is called simple cells. The Figure explains the directional sensitivity of simple and complex cells.

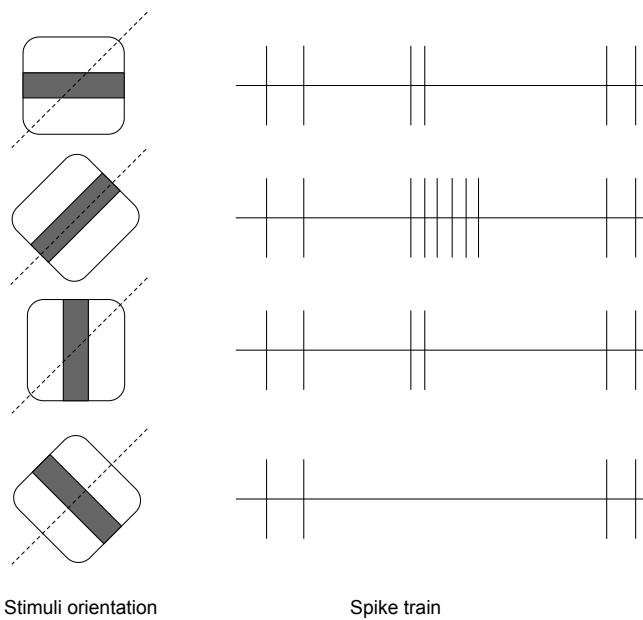


Figure 2.1: Spike train response of an orientation selective neuron. The cell has a preferred orientation which is shown by dashed line. Response reduce exponentially when the stimuli orientation deviates from preferred orientation

2.2 Experiment setup

The data for this work was obtained from experiments conducted in Sur's lab of Neuroscience, MIT. A brief overview of the experiments and data is required to follow the upcoming chapters.

"Data in this study were collected from adult (> 8 weeks old) mice of either sex. Mice were anesthetized using isoflurane (3% induction, 1.52% during surgery). A custom-built metal head post was attached to the skull using dental cement (C&B-Metabond, Parkell), and a 3-mm-diameter craniotomy was performed over binocular V1 (~ 23 mm lateral and 0.5 mm anterior to lambda). Care was taken not to rupture the dura mater. The core body temperature was maintained at 37.5C using a heating blanket (Harvard Apparatus).

For awake experiments, mice were first habituated for 5 d to head fixation on a

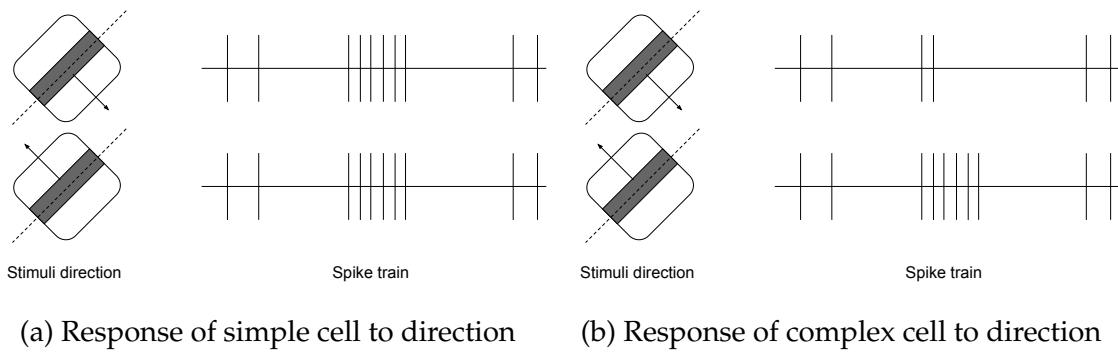


Figure 2.2: With a stimuli aligned with preferred orientation, direction of motion of stimuli also affect the response in some cells. Complex neurons have stronger response to one direction than other. Neurons which respond similar to both directions are simple cells.

custom-built stage. Once habituated, the mice received a microinjection of 100-200 nl of AAV1.Syn.GCaMP6f.WPRE.SV40 (University of Pennsylvania Vector Core, diluted to a titer of 10^{12} genomes ml⁻¹), following which a cranial window was implanted over the craniotomy and sealed. Mice were allowed to recover for 23 weeks to allow for adequate expression of the virus before imaging commenced. Imaging was performed using a Prairie Ultima two-photon system. The imaging was done at a frequency of 20Hz. The imaging was done while mice were presented with a visual stimuli on a 23 inch gamma-corrected LCD monitor (Dell) covering a visual space of $\sim 96 \times 54 \text{ deg}^2.$ [Rikhye and Sur [2015]]

The experiment was performed for the following two sets of visual stimuli.

1. Drifting sinusoidal gratings

A drifting sinusoidal grating pattern with two parameters - orientation and direction is used as visual stimuli

- Sinusoidal drifting gratings at 16 different directions (0:22.5:337.5) were used. Spatial frequency was fixed at 0.03 cycles per degree.
- Each direction was repeated 10 times. Directions are presented in a randomized order.
- Each direction was presented for 2 seconds and was always preceded by a 4 seconds gray screen. Therefore the total duration of the stimulus is 6 seconds.

- Calcium signals (GCaMP6f in awake mice) were acquired from awake mice at 20 frames per seconds. Thus, sampling rate is 20Hz.
2. **Natural movies** Five different natural movies from the the van Hateren movie database are used as visual stimuli.
- Each video has a duration of 4 seconds.
 - Each video was repeated 10 times in a randomized order.
 - A gray screen was presented for 2 seconds before and after every movie.
 - Calcium transients were sampled from mice at 50 frames per seconds.

2.3 Previous works

Kuffler (1953) showed retinal ganglion cells have a concentric receptive field with an ‘On-center’ and ‘Off-center’ periphery. The on and off areas within a receptive field were found to be mutually antagonistic, and a spot restricted to the centre of the field was more effective than one covering the whole receptive field [Barlow *et al.* [1957]]. David Hubel and Torsten Wiesel made a quantum leap in understanding of the visual perception in brain. The central point of their initial finding in 1959 [Hubel and Wiesel [1959]] was that oriented slits of light were the most effective stimuli for activating striate cortex neurons. In the light of new imaging technologies, the problem is studied with clarity and robust measures of selectivity are introduced in [Mazurek *et al.* [2015]].

Even with the knowledge of properties of neurons in V1, Information representation in V1 is still an unclear topic. The significance of individual action potentials has been difficult to determine, particularly in light of the considerable trial-to-trial variability of responses to visual stimuli [Reich *et al.* [2000]]. A study of relation between the stimulus and the population response by computing the spatial correlation between the maps of the expected and observed response revealed

continuous spatial encoding of low and high level features of natural images in V1 [Ayzenshtat *et al.* [2012]]. A study to find role of spatial correlations in natural scenes in reliable coding revealed responses in V1 were much less reliable at both the single neuron and population level when spatial correlations were removed from the image. This change in reliability was due to a reorganization of between-neuron correlations. Strongly correlated neurons formed ensembles that reliably and accurately encoded visual stimuli, whereas reducing spatial correlations reduced the activation of these ensembles, leading to an unreliable code [Rikhye and Sur [2015]]. This motivates us to examine interneuronal correlations as the reliable coding mechanism in V1.

Motifs are well studied in the context of music. Motifs in Carnatic music is based on the concept of *raga*. Melodic motifs are signature phrases that give *raga* its identity [Ishwar *et al.* [2013]]. Motif spotting was used to verify *raga* by roughly matching pitch contours of the reference motif and the musical piece [Dutta *et al.* [2013]]. Rough longest common subsequence (RLCS) is used in music for matching a query in the reference signal [Lin *et al.* [2011]].

CHAPTER 3

Analyzing neuronal properties

We have discussed characteristics of neurons of V1 in the background section 2.3. In this chapter, we use experimental data to demonstrate the claimed properties of neurons in V1. In the experiment, a drifting sinusoidal grating video is shown to awake mice and neuronal responses were recorded. See Section 2.2 for detailed experiment setup.

Average response of a neuron is modeled as a function of orientation angle using a Gaussian function. Similarly, average response to various directions are modeled using a mixture of Gaussian functions. Root mean square of residuals are used as a goodness of fit measure.

Tuned cells are neurons having similar preferred orientation and an ‘acceptable’ degree of selectivity. Similarity in function of neurons can be examined by finding similarly tuned neurons.

3.1 Quantifying Orientation and Directional selectivity

Modern imaging techniques allow us to examine responses of neurons with clarity. Even though it was known that neurons in V1 are selective to orientation, we require robust metrics to quantify degree of selectivity and preferred orientation.

Plotting response to each stimuli direction in a vector space provides an intuition about the characteristics of the cell. In orientation space, responses of two opposite directions are averaged. Angle varies from 0° to 180° in orientation space while angle it changes from 0° to 360° in direction space. Length and direction of the vector sum is a good metric for degree of selectivity and preferred orientation. Normalized length of vector sum in orientation space is defined as OSI(Orientation Selectivity Index).

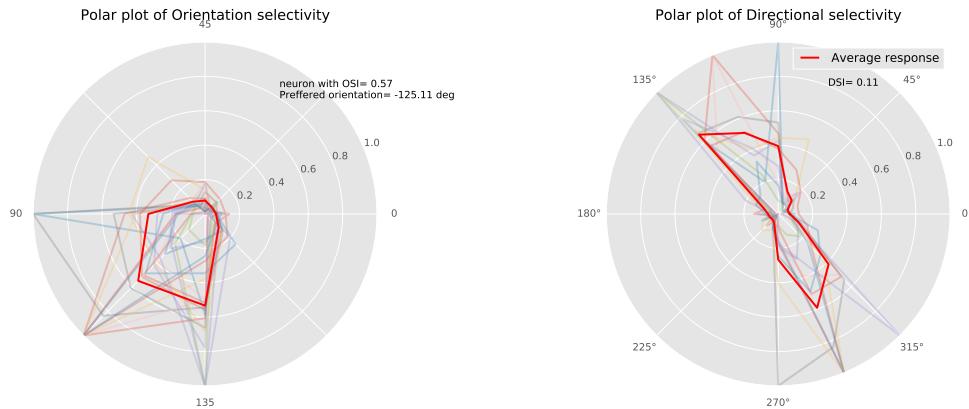
$$OSI = \left| \frac{\sum_k R(\theta_k) \exp(2i\theta_k)}{\sum_k R(\theta_k)} \right|$$

Similarly, Normalized length of vector sum in direction space is defined as DSI(Direction Selectivity Index).

$$DSI = \left| \frac{\sum_k R(\theta_k) \exp(i\theta_k)}{\sum_k R(\theta_k)} \right|$$

Where $R(\theta_k)$ is the average response to angle θ_k .

A simple neuron is expected to have high OSI but low DSI. Responses of a simple cell are plotted in orientation and direction space in Figure 5.7.

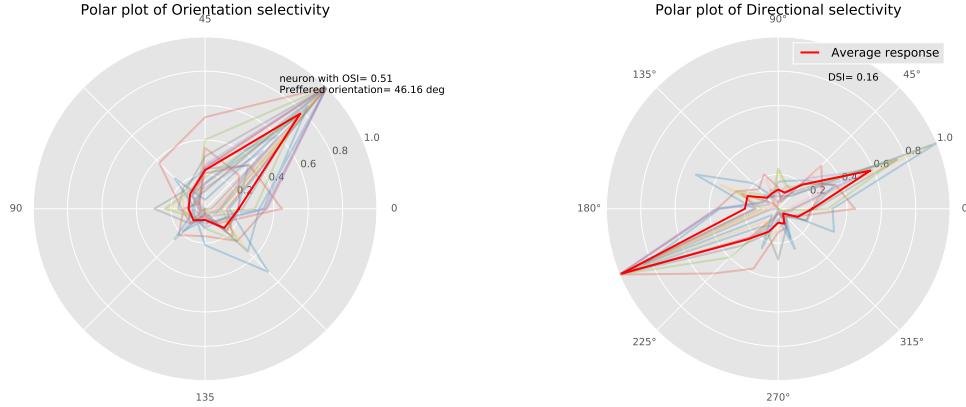


(a) Responses plotted in orientation space (b) Responses plotted in direction space

Figure 3.1: Responses $R(\theta_k)$ of a simple neuron for each θ_k plotted in both orientation and direction space. Equal lobes in direction space shows direction is irrelevant to response.

A complex neuron is expected to have high OSI and high DSI. Responses of a

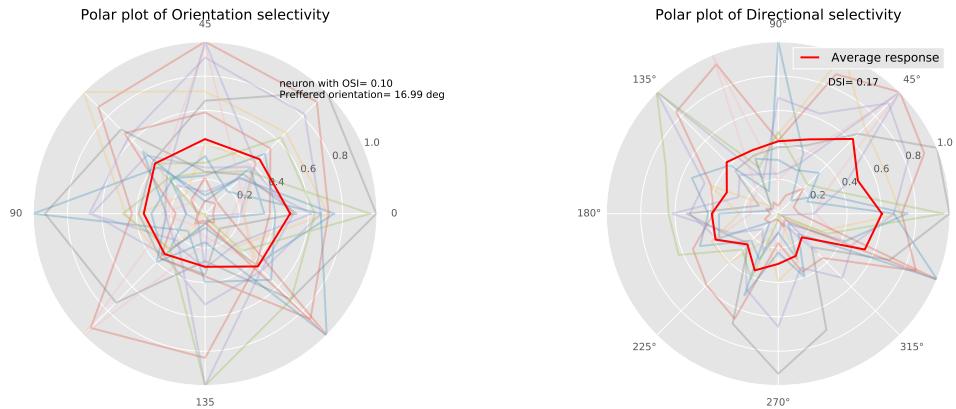
complex cell are plotted in orientation and direction space in Figure 3.2.



(a) Responses plotted in orientation space (b) Responses plotted in direction space

Figure 3.2: Responses $R(\theta_k)$ of a complex neuron for each θ_k plotted in both orientation and direction space. Unequal lobes in direction space shows one direction is preferred than its opposite.

An orientation insensitive neuron is expected to have low values for both OSI and DSI. Polar plots of response from an unselective cell in orientation and direction space is shown in Figure 3.3.



(a) Responses plotted in orientation space (b) Responses plotted in direction space

Figure 3.3: Responses $R(\theta_k)$ of a an orientation insensitive neuron for each θ_k plotted in both orientation and direction space. Similar responses to all orientations shows absence of selectivity.

3.2 Modeling neuronal response

Modeling the response of neuron to various orientations and visualizing is a great way to study characteristics of neurons like orientation selectivity. We can also characterize the degree of selectivity and preferred orientation from model parameters. Orientation tuning curve is modeled using a Gaussian function with constant offset. The empirical form of the orientation tuning curve is,

$$R_o(\theta) = C + R_p \exp \left\{ \frac{-\|\theta - \theta_{pref}\|^2}{2\sigma^2} \right\}$$

Where $R_o(\theta)$ is the time-averaged response of a neuron to angle of orientation θ . Parameter θ_{pref} is the preferred orientation of the neuron. Tuning width σ tell us how much the cell is selective. C is a constant offset.

Similarly, we can model direction tuning curve using a mixture of two Gaussian functions with a constant offset.

$$R_d(\theta) = C + R_p \exp \left\{ \frac{-\|\theta - \theta_{pref}\|^2}{2\sigma_1^2} \right\} + R_n \exp \left\{ \frac{-\|\theta - \theta_{null}\|^2}{2\sigma_2^2} \right\}$$

Where $R_o(\theta)$ is the time-averaged response of neuron to angle of direction θ . Relative magnitude of tuning widths, σ_1 and σ_2 denote the amount of directional selectivity. C is a constant offset.

Parameters are estimated by minimizing squared sum of error. Sum of squared error is defined as:

$$SSE = \sum_{i=1}^N \|R(\theta_i) - R_o(\theta_i)\|^2$$

A gradient descent algorithm finds the optimum parameters by minimizing SSE. In Figure 3.4 , tuning curves of a complex cell is modeled. The distinct peak in

the orientation tuning curve shows selectivity to orientation θ_{pref} . In the direction tuning curve, peaks of different magnitude shows one direction is more preferred than other. In Figure 3.5 , tuning curves of a complex cell is modeled. The distinct

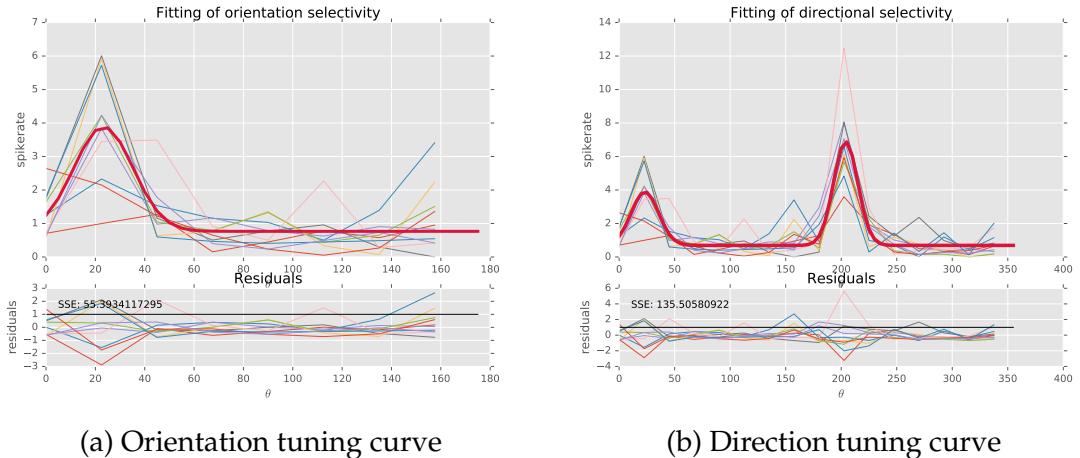


Figure 3.4: Fit of orientation and direction tuning curves of a neuron. The distinct peak in the orientation tuning curve shows selectivity to orientation θ_{pref} . Different σ_1 and σ_2 in direction tuning curve shows direction sensitive cell. - Thus a complex cell

peak in the orientation tuning curve shows selectivity to orientation θ_{pref} . In the direction tuning curve, peaks of same magnitude shows of stimuli direction is irrelevant.

3.3 Finding similarly tuned neurons

Receptive field of a neuron in V1 consists of a subset of neurons in one layer below it. Those neurons in turn have a receptive field. By following down layers by layer, we can find a visual field for each neuron in V1. Orientation selectivity of neurons in V1 detects edges and their orientation in the visual field. By studying similarly tuned neurons, we can get an insight to redundancy of coding and distribution of functionally similar cells in V1.

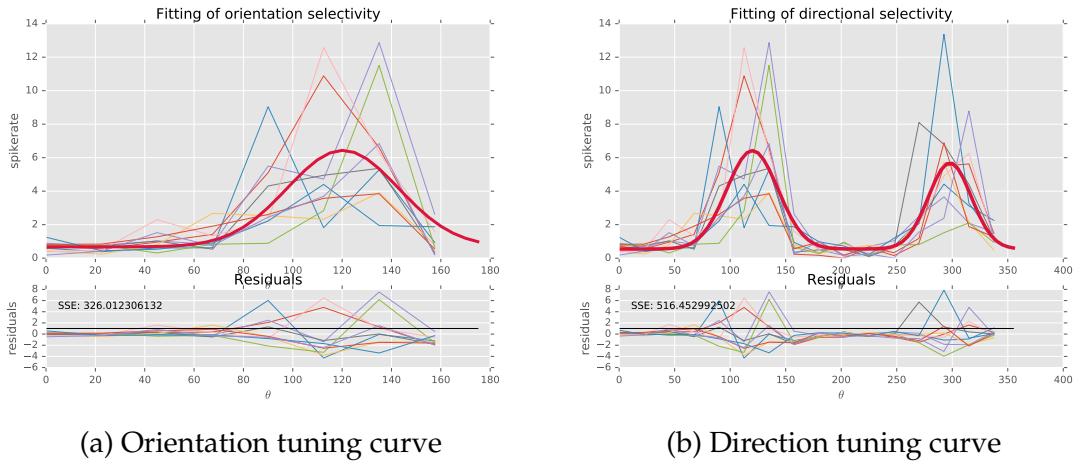


Figure 3.5: Fit of orientation and direction tuning curves of a neuron. The distinct peak in the orientation tuning curve shows selectivity to orientation θ_{pref} . Similar σ_1 and σ_2 in direction tuning curve direction of stimuli is irrelevant. - Thus a simple cell

Similarly tuned cells are expected to have similar response to different stimuli orientations. A good correlation between responses $R(\theta_k)$ of two neurons represents similarly tuned cells. Even after having same preferred orientations, the neurons could have different degree of selectivity. We would like to find similarly tuned neurons which have an ‘acceptable’ OSI.

Pearson correlation between each pair of neurons in a mice are computed and plotted in Figure 3.6. The neurons in both axes are sorted in order of decreasing OSI values. The neuron pairs that occupy bottom left of the Figure 3.6 are the ones we are interested in. Finally to find neurons similar to a particular neuron, choose the row corresponding to the neuron and all the neurons in that row having good correlation value and an acceptable OSI. Figure 3.7 shows tuning curves of some neurons retrieved from correlation study.

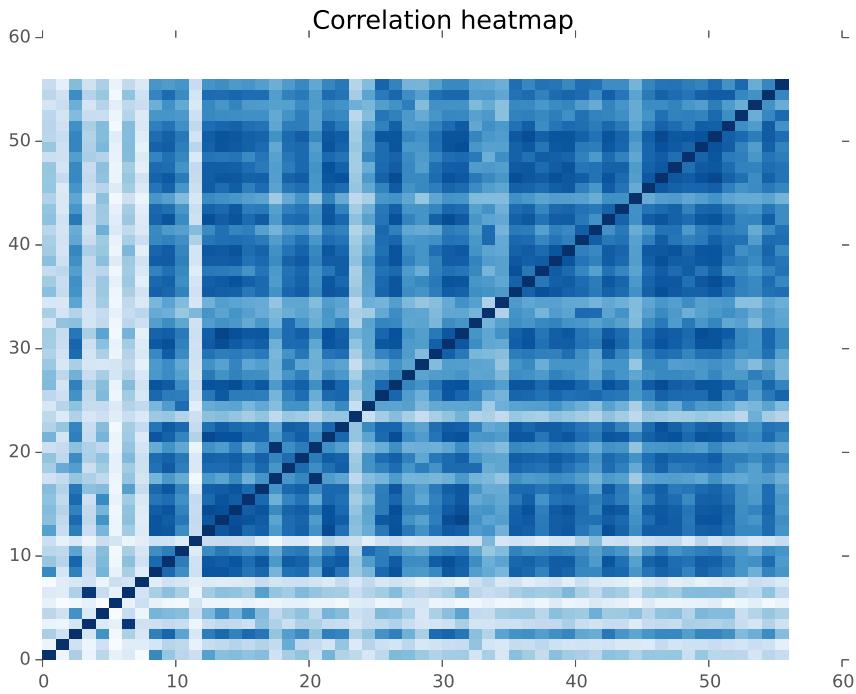


Figure 3.6: Pairwise Pearson correlation of neurons. Both axes are sorted in decreasing order of OSI.

3.4 Study of principal components

Information contained in the visual stimuli reaches V1 after it pass through lower layers. Basic features like intensity, simple gradient of the visual scene are captured by neurons in the lower layers. Edges and motion of edges in the stimuli are represented by orientation and direction sensitive neurons in V1. With Principal Component Analysis (PCA) we aim to study the tightness of information representation. A subset of neurons (~ 60) in V1 are sampled at 20Hz for 6 seconds during the experiment. We analyze principal components of the responses to find redundancy in these responses. The number of uncorrelated principal components which capture ‘most’ of the variance in data can determine amount of redundancy.

Principal Component Analysis is an orthogonal transformation of possibly cor-

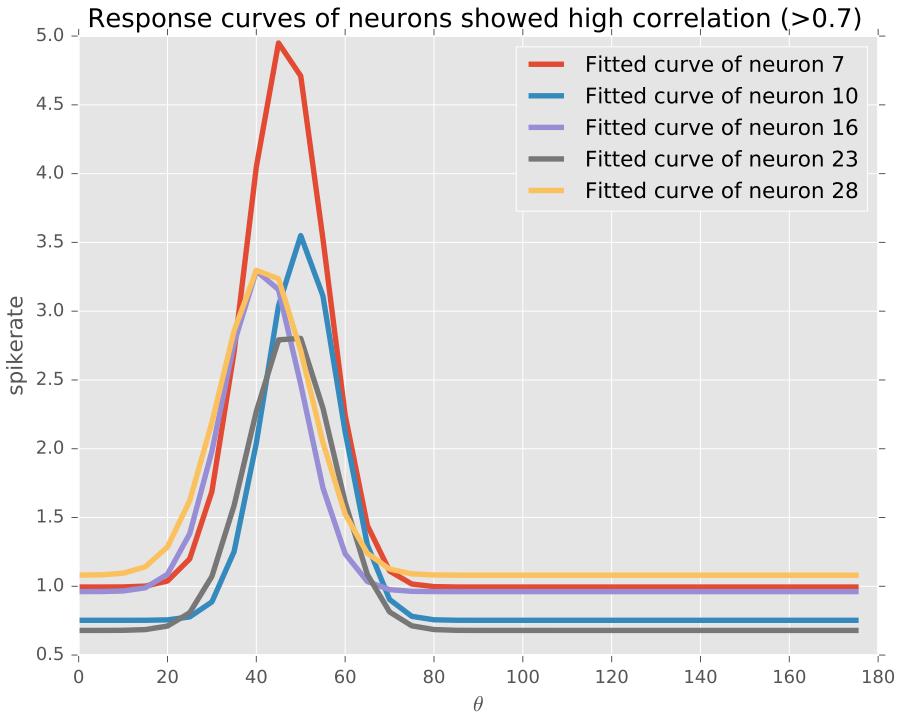


Figure 3.7: Orientation tuning curves of some neurons retrieved from correlation study. All of them have a same preferred orientation.

related observations to linearly uncorrelated components called principal components. The principal components are orthogonal because they are the eigenvectors of the covariance matrix, which is symmetric. The first principal component captures largest variance and decreases further on. By observing number of components that needs to capture a desired variance, we can tell correlations in the data. Transforming original data to new orthogonal basis gives a representation which has dimension less than or equal to the number of original dimension. Reconstructing original data from transformed data also indicates amount of correlation in original data.

Here we take number of neurons as feature dimension. Possibility of a valid reduction will imply that same information can be represented with less number of neurons. The responses for each neuron is averaged across trials. PCA is done

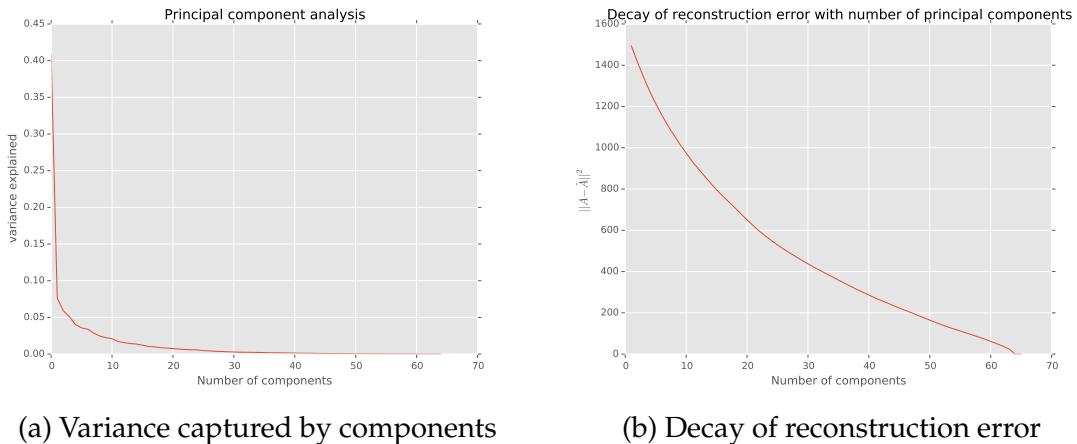


Figure 3.8: Principal Component Analysis of response of neuron to drifting sinusoidal grating stimuli.

on the data to find ratio between number of principal components and variance explained. Figure 3.8a shows variance explained ratio for signals from a mouse presented with a drifting sinusoidal grating visual stimuli. The original data was transformed to principal components basis with a subset of principal components. Attempting to reconstructing original data from transformed data produces an error. The reconstruction error depends on number of chosen subset of principal components. Figure shows reconstruction error for different number of principal components. As we increase number of principal components, the error decreases. And finally as the number components equals the original feature dimension, reconstruction error vanish.

3.5 Study of Reliability

Information encoding in primary visual cortex is a complex process due to intrinsic neuronal variability. Analysis of intertrial reliability of information encoding in neurons is the first step into understanding coding mechanism of V1. The degree of trial-to-trial variability in a response is commonly measured in terms of reliability.

A neuron is said to be reliable if it fires the same number of precisely timed spikes on every repetition of a stimulus [Tiesinga *et al.* [2008]].

The devised experiment measures calcium concentration in neurons rather than spike rate. Trial to trial reliability can be computed using correlation between calcium responses of each trial. When a visual stimuli is presented T times to a mouse, we can compute response reliability R_A of a neuron to the stimuli A as

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

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

The study was done for 48 neurons in a mice and the response reliability is plotted in Figure 3.9 for each neuron. It is evident that response reliability for most of neurons are small. The study shows responses of neurons in V1 are not robust because of intrinsic noise. If the whole sequences is not reliable, could there be subsequences that are reliable? We will explore it in coming chapters by searching for common subsequences.

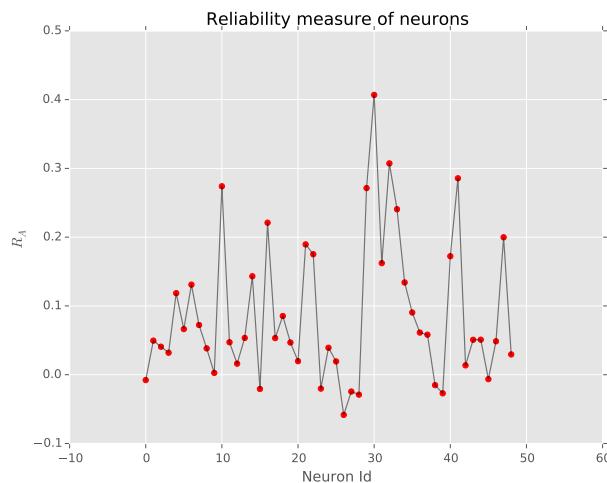


Figure 3.9: Reliability measure R_A for different neurons

CHAPTER 4

Searching for Motifs

In Music, motif is a perceivable recurring fragment. Detecting such motifs in a song tells us more about the song like its melody. In Genetics, motif is a sequence pattern of nucleotides in a longer DNA sequence. From a signal processing perspective, motifs are signal segments that recurs in a longer signal. A valid motif should have an ‘acceptable’ length such that it has a significance. Motif analysis is important since in every domain, occurrence of motifs shows an underlying pattern and has a meaning to it.

In the experiments performed, each neuron responses are captured as a time series sampled at 50Hz for 4 seconds. In neuron responses we define a motif as subsequence of response signal having significant length and repeats itself either in

- the same response signal but in a different part **or**
- the response of the same neuron to a different trial **or**
- the response of another neuron in the same mouse **or**
- the response of neuron in a different mouse **and**

Analysis of motifs in neuronal signals will help us understand reliable information representation. Even if two responses of same trial has a long subsequence but not time synchronized, Study of reliability in Section 3.5 will fail as the Pearson correlation coefficient will output a small value. Motivic analysis can detect such subsequences even if they are time shifted.

Study of motifs across two different neurons within a mouse can explain correlation between two neurons. Correlated neurons can represent neuronal interconnections. Correlated and synchronous activity in populations of neurons has been observed in many brain regions and has been shown to play a crucial role in cortical coding, attention, and network dynamics [Rosenbaum *et al.* [2014]]. Studying correlations between neurons is again not effective if their responses are time separated as Pearson correlation will fail. Analyzing motifs across neurons will be a better way to study neuronal correlations.

Motifs found in neurons from two *different* mice will represent similar biological process both the neurons share. As the motifs found cannot be due to interconnections, they are due to similar activation mechanism of neurons. In this case, we would expect a small motif compared to motifs found between neurons within a mouse.

In this chapter, we will analyze the presence of motifs in neuronal signals. The presence of motifs will motivate us to extract the motifs and infer their significance.

4.1 Cross-correlation Function

Cross-correlation is a measure of similarity between two signals as a function of lag of one signal relative to the other. It is commonly used for searching a query sequence in a large reference sequence. The function calculates sliding dot product at various lags. For discrete signals, cross correlation is defined as:

$$(f \star g)[n] \stackrel{\text{def}}{=} \sum_{m=-\infty}^{\infty} f^*[m] g[m+n].$$

Searching for a motif using cross-correlation function is possible if we already have a subsequence that we suspect as a motif. To visualize correlation function in query search, we will manually select a query sequence and search for its occurrences in another neuronal response. For the scope of this chapter, we take a subsequence from the response of a ‘template neuron’ and search for its occurrences in response of a ‘target neuron’.

Trial averaged responses of template and target neurons are chosen and a subset frame of former is extracted using frame ending and frame width as parameters. This query sequence is compared with target sequence as a function of lag. Figure 4.1 shows Cross-Correlation function between a target neuron response and a manually chosen subsequence from template neuron. This study is not effective as

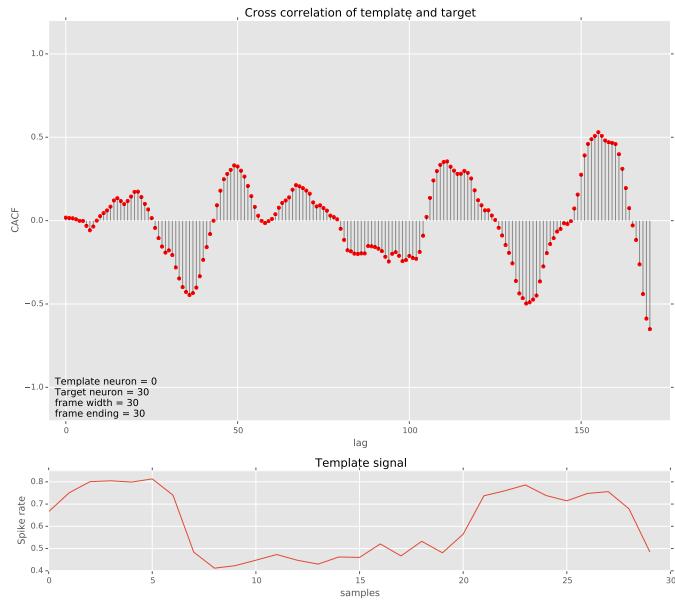


Figure 4.1: Cross-Correlation function between a target neuron response and a manually chosen subsequence from template neuron.

we do not know what to search for. Manually selecting a subsequence is inefficient and selected subsequence does not guarantee to be a valid motif.

4.2 ACFGram

A spectrum describes a signal in terms of energy spread over its frequency components. A spectrogram does exactly the same but also takes time into consideration. Computing spectrogram is done by first making chunks/frames of time-domain signal which usually overlap; Magnitude of Fourier transform of each frame is computed to form magnitude spectrum of that frame. These frame spectra are stacked horizontally in increasing order of frame ending time to form a temporal dimension. The resulting frequency vs time vs magnitude representation is called spectrogram. Spectrogram enables us to study the change of spectra over time.

We formulate an analogous visualization of cross-Correlation function where the template signal changes in each frame. A frame is a subsequence of original signal with a width and frame ending parameter. Changing the frame ending successively will return overlapping frames having same width. Figure 4.2 shows how a frame from template signal is selected. Cross-Correlation of extracted

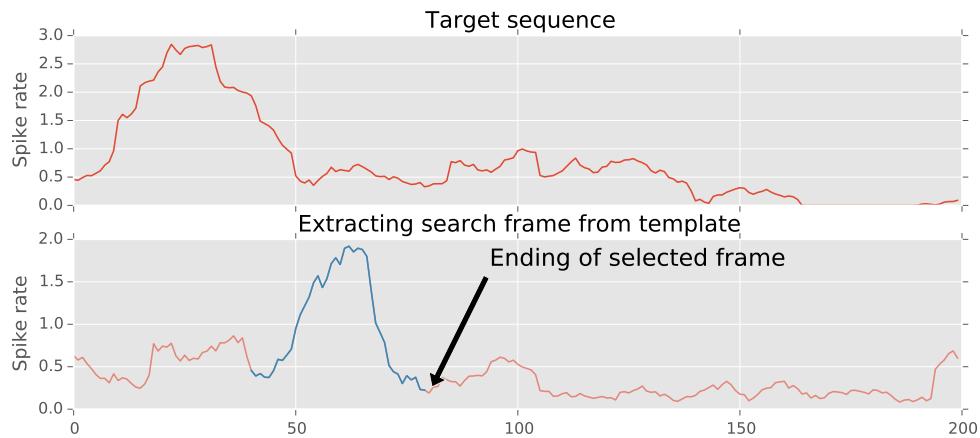


Figure 4.2: Extracting frame from template based on frame width and ending.

frame and target signal at various lags are computed. Next frame is extracted by incrementing frame ending parameter by another parameter called frame shift.

The process is then repeated for the next frames until the template sequence is exhausted. Cross-Correlation function of each frame is then stacked horizontally to form a temporal dimension. We call the resulting Time Vs Lag Vs Cross-Correlation function as ACFGram. In Section 4.1 we varied frame width and frame ending manually to extract a signal subsequence and then queried it in target signal. ACFGram removes frame ending parameter from manual configuration. Existence of motifs will be clear if we examine ACFGram for various frame widths. The study is conducted across trials of same neuron and across different neurons.

The Figure 4.3b shows ACFGram plots for template and target responses from same neuron. The study aim to find presence of repeating subsequence within a neuron's response. The length of suspected motif is varied as a parameter. For a small frame width (~ 15) the estimation of Cross-Correlation function is poor. Also such a small subsequence is not significant enough to be considered as a motif. As we Increase the frame width, repeating patterns are found which indicates the presence of repeating subsequence. The Figure 5.7b shows ACFGram

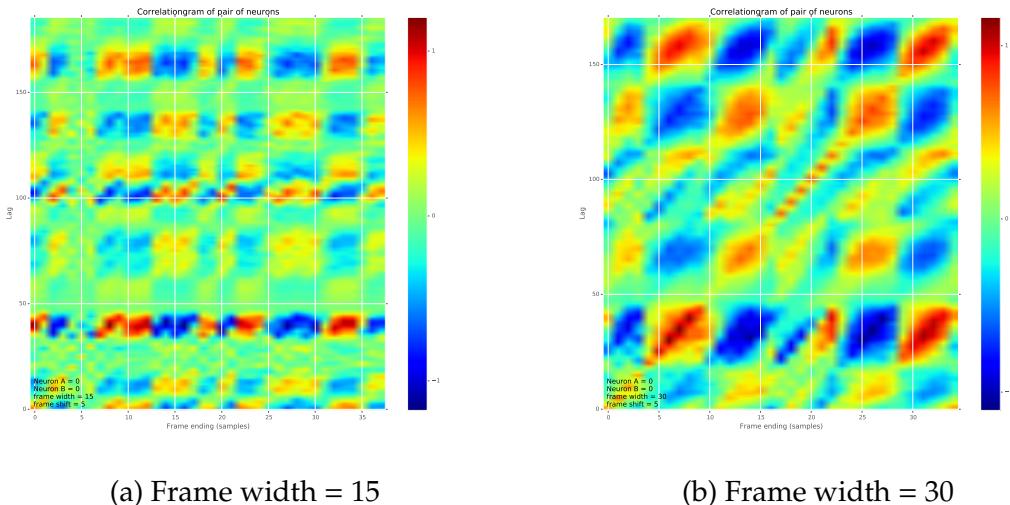


Figure 4.3: ACFGram plots for template and target responses from same neuron.

plots for two different neuron responses. Motifs across neurons are studied for

finding a better correlation metric. For different frame widths, Cross-Correlation function indicate presence of motifs across neurons. The approximate length of motifs is more than 45 samples, which is significant. Importantly, more or less any pair of neurons taken from the sampled set of 65 neurons from each mouse exhibited presence of motifs. This indicates existence of inter neuronal correlations. This supports the claim that information in V1 is coded in population by inter-neuronal correlations. The search for the presence of motifs in primary visual

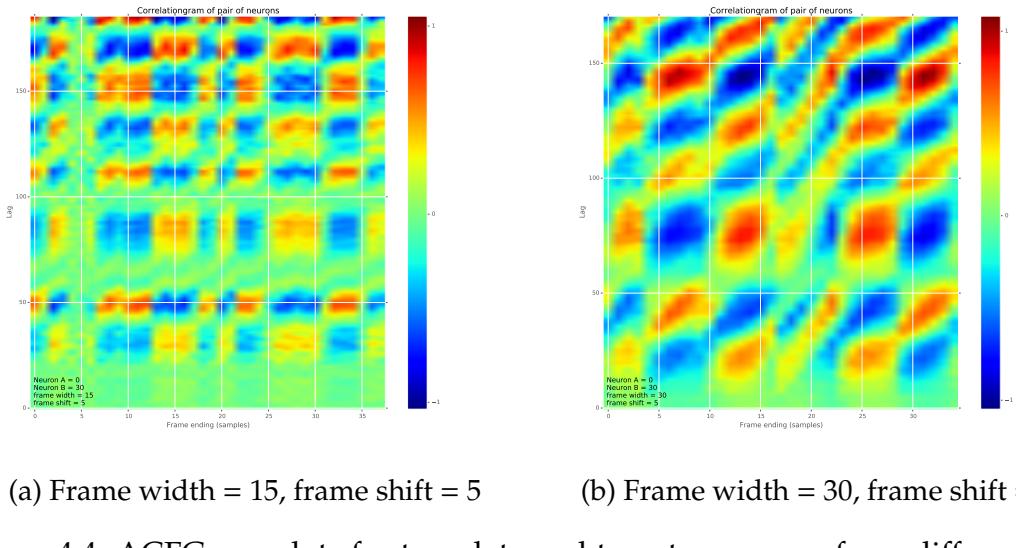


Figure 4.4: ACFGram plots for template and target responses from different neurons.

cortex succeeded with experimentally verifying rough common subsequences of significance (> 2 seconds) exists in neuronal responses. This motivates us to find the longest motifs you can find between two neurons. Also till now we used Pearson correlation as a measure of similarity. We look into rough measures of similarity and longest common subsequence in the next chapter.

CHAPTER 5

Rough Longest Common Subsequence (RLCS)

Having established existence of motifs in neuronal responses, A robust algorithm to extract best possible motifs from two signals is required. As a longer motif has more significance in the context, an algorithm which produce longest common subsequence is desirable. Longest Common Subsequence (LCS) is a classical problem in query string matching, data comparison and Bio-informatics. We will eventually look into an adapted LCS problem which can be used to match real signals called Rough Longest Common Subsequence (RLCS).

5.1 Longest Common Subsequence (LCS)

A subsequence of a string S is a set of characters that appear from left to right in original string S but not necessarily consecutive. There will be more than one subsequence for any string with length greater than unity. For two strings S_1 and S_2 , the common entries in sets of substrings of each strings are called common subsequences of S_1 and S_2 . Among the common subsequences, the subsequence with maximal length is called Longest Common Subsequence.

Some examples of LCS are given in Table 5.1. Note that there can be more than one LCS for a pair of strings.

LCS problem can be defined formally for two subsequences $X = (x_1, x_2 \dots x_m)$

String 1 - S_1	String 2 - S_2	$LCS(S_1, S_2)$
ABCDGH	AEDFHR	ADH
AGGTAB	GXTXAYB	GTAB
AGCAT	GAC	AC, GC and GA
BACDB	BDCB	BCB

Table 5.1: Some examples of LCS in string comparison

and $Y = (y_1, y_2 \dots y_n)$ as

$$LCS(X_i, Y_j) = \begin{cases} & \text{if } i = 0 \text{ or } j = 0 \\ LCS(X_{i-1}, Y_{j-1}) \cup x_i & \text{if } x_i = y_j \\ \text{longest}(LCS(X_i, Y_{j-1}), LCS(X_{i-1}, Y_j)) & \text{if } x_i \neq y_j \end{cases}$$

Where $X_i = (x_1, x_2 \dots x_i)$ and $Y_j = (y_1, y_2 \dots y_j)$.

LCS problem has an optimal substructure- the problem can be broken down into subproblems of similar kind. As the problem also has overlapping subproblems, thus a Dynamic Programming approach is used to find the LCS. Running time of the dynamic programming approach is $O(nm)$ where n and m are the lengths of strings.

5.2 RLCS

LCS problem can be adapted for signal matching. Instead of equality, a rough distance measure is used. If the distance between samples is less than a threshold, they are considered a match. It is not desirable just to optimize the length of subsequence in signals as the gaps between selected samples can cause false alarms. Dynamic Programming algorithm for rough signal matching has to be penalizing gaps and motivating matches. The rest of the problem and the solution remains

similar to LCS problem and similar Dynamic Programming approach can solve the problem efficiently.

Modified LCS problem with rough comparison is stated below. The RLCS of two discrete signal sequences $X = \langle x_1, x_2 \dots x_m \rangle$ and $Y = \langle y_1, y_2 \dots y_n \rangle$ is obtained by maximizing the score value S .

$$S(X_i, Y_j) = \begin{cases} 0 & \text{if } i = 0 \text{ or } j = 0 \\ S(X_{i-1}, Y_{j-1}) + x_i & \text{if } dist(x_i, y_j) < \tau_{dist} \\ \max(S(X_i, Y_{j-1}), S(X_{i-1}, Y_j)) & \text{if } dist(x_i, y_j) > \tau_{dist} \end{cases}$$

Where $X_i = \langle x_1, x_2 \dots x_i \rangle$ and $Y_j = \langle y_1, y_2 \dots y_j \rangle$.

Various context uses different distance measures for comparing samples. Manhattan distance (city block distance) of notes is used in work [Lin *et al.* [2011]] in the context of music matching. In a work on verification of Raga in Carnatic music [Dutta *et al.* [2013]], a domain specific similarity measure is used. RLCS on neuronal signals in this work uses normalized Euclidean distance for comparing similarity.

$$dist(x_i, y_i) = (x_i - y_i)^2$$

Distances are normalized from 0 to 1 by performing:

$$dist(x_i, y_i) = \frac{dist(x_i, y_i) - minDist}{maxDist - minDist}$$

Where $minDist$ is the minimum pairwise distance and $maxDist$ is the maximum pairwise distance.

Most important part of RLCS is the adapted score update rules. In the classical

LCS problem, the gaps between the a common subsequence is not penalized; Instead only the length of the subsequence was of consideration. But in signals, having a gap in between will give a poor match of signal. So it is important to penalize the gaps and reward matches. The following score update rules penalizes gaps in subsequence and awards positive score for matches.

1. Score update for a match. $dist(x_i, y_j) < \tau_{dist}$

$$score(i, j) = score(i - 1, j - 1) + 1 - \frac{dist(x_i, y_j)}{\tau_{dist}}$$

This case happens when a match is found. The score should be rewarded with a positive value. The maximum awarded score is 1 for an exact sample match and minimum score of 0 is given when the distance equals distance threshold ($dist(j, j) = \tau_{dist}$) threshold. The closer the samples, more the score; Lesser the gap between samples, more the score.

2. Score update for a mismatch. $dist(x_i, y_j) > \tau_{dist}$

$$score(i, j) = score(i - 1, j - 1) - \delta$$

We allowed any number of gaps in a string LCS problem. But in the case of signals, we cannot allow large gaps as it would produce false alarms. We penalize each mismatching samples with a constant score δ .

The problem is then recursively solved using Dynamic Programming algorithm which will provide us with a 2-D score matrix. The Dynamic programming algorithm for input sequences $X = < x_0, x_1, \dots, x_{m-1} >$ and $Y = < y_0, y_1, \dots, y_{n-1} >$ is explained in algorithm 1.

Backtracking through score matrix will gives the Longest Common Subsequence. As we have been penalizing gaps, there will be points in the track where score is zero. For each sample match increases and for every mismatch, score decreases. So the zero in backtracking path will mean that there are enough mismatches to balance the total positive score created by matches. This identifies the boundary of a common subsequence. A number of subsequences could be

Algorithm 1 Dynamic Programming algorithm for RLCS

```
1: function RLCS( $X, Y, \tau_{dist}, \delta$ )
2:    $p \leftarrow \min(m, n)$ 
3:   for  $i = 0; i < m; i++$  do
4:     for  $j = 0; j < n; j++$  do
5:        $dist(i, j) \leftarrow euclideanDistance(x_i, y_j)$ 
6:     end for
7:   end for
8:    $\text{minDist} \leftarrow \min(dist)$                                 ▷ For normalization
9:    $\text{maxDist} \leftarrow \max(dist)$                                 ▷ For normalization
10:
11:   $c \leftarrow zeros(m + 2, n + 2)$                                 ▷ For storing running score
12:   $s \leftarrow zeros(m + 2, n + 2)$                                 ▷ For storing score
13:   $d \leftarrow zeros(m + 2, n + 2)$                                 ▷ For backtracking
14:   $p \leftarrow zeros(m + 2, n + 2)$                                 ▷ For storing partial scores.
15:
16:  for  $i = 0; i < m; i++$  do
17:    for  $j = 0; j < n; j++$  do
18:       $d \leftarrow (dist(i, j) - \text{minDist}) / (\text{maxDist} - \text{minDist})$           ▷ Normalize
19:      if ( $d < \tau_{dist}$ ) then
20:         $d(i, j) \leftarrow '↖'$                       ▷ Path to travel while backtracking
21:         $c(i, j) \leftarrow c(i - 1, j - 1) + (1 - d/\tau_{dist})$ 
22:         $s(i, j) \leftarrow s(i - 1, j - 1)$ 
23:      else if  $c(i - 1, j) > c(i, j - 1)$  then
24:         $d(i, j) \leftarrow '↑'$ 
25:         $c(i, j) \leftarrow c(i - 1, j) - \delta$           ▷ Penalize
26:         $c(i, j) \leftarrow \max(0, c(i, j))$ 
27:        if ( $d(i - 1, j) == '↖'$ ) then
28:           $s(i, j) \leftarrow (p(i - 1, j) * p^2 + (c(i - 1, j)/p)^2$ 
29:        else
30:           $s(i, j) \leftarrow s(i - 1, j)$ 
31:        end if
32:      else
33:         $d(i, j) \leftarrow '←'$ 
34:         $c(i, j) \leftarrow c(i, j - 1) - \delta$           ▷ Penalize
35:         $c(i, j) \leftarrow \max(0, c(i, j))$ 
36:        if ( $d(i, j - 1) == '↖'$ ) then
37:           $s(i, j) \leftarrow (p(i, j - 1) * p^2 + (c(i, j - 1)/p)^2$ 
38:        else
39:           $s(i, j) \leftarrow s(i, j - 1)$ 
40:        end if
41:      end if
42:    end for
43:  end for
44:  return  $c, s, d, p$ 
45: end function
```

extracted by cutting backtrack path at points where score is zero. This set of subsequences are called **Longest Common Subsequence Set** (LCSS) [Dutta *et al.* [2013]]. Algorithm 2 explains backtracking to extract the LCSS.

Algorithm 2 Backtracking to find the subsequences

```

1: function BACKTRACK(X, Y, c, s, d)
2:   maxScore  $\leftarrow \max(s(i, j))$ 
3:    $x, y \leftarrow \text{indexOf}(\text{maxScore})$ 
4:    $l \leftarrow 0$ 
5:   while ( $x \neq 0$ ) and ( $y \neq 0$ ) do
6:     segment( $l$ ) = [ $x - 1, y - 1, c(x, y)$ ]
7:      $l = l + 1$ 
8:     if ( $d(x, y) == ' \nearrow '$ ) then
9:        $x = x + 1$ 
10:       $y = y + 1$ 
11:      else if ( $d(x, y) == ' \uparrow '$ ) then
12:         $x = x + 1$ 
13:      else
14:         $y = y + 1$ 
15:      end if
16:    end while
17: end function
```

5.2.1 Motivic analysis of Neuronal signals with RLCS

Having established presence of motifs in neuronal responses, we use Rough Longest Common Subsequence (RLCS) algorithm to find the ‘best’ motifs from neuronal responses. Penalizing the gaps makes RLCS a good signal matching algorithm.

The experiments are conducted on responses of neurons from V1 of 11 mice to different visual stimuli. Each signal is sampled at 20 Hz and last for 10 seconds. The study is split into different sections based on which neurons the signals for analysis come from.

Motifs in responses of a neuron to different trials

The signals from a neuron's response to different trials of same stimuli are passed to RLCS algorithm to detect motifs. The Figure 5.1 shows score matrix obtained after RLCS. The effect of τ_{dist} was experimentally studied and fixed at 0.005. For τ_{dist} above 0.001, number of false alarms increased. Decreasing threshold further caused bad detection rate. The desired motif takes a path through the score matrix. The diagonal shifts in the score path are sample matches - as a result of that we can see the increase in score. Shifting to right or top denotes a gap in the subsequence. As mismatches cause penalty, the score will decrease. When enough mismatches to compensate the total positive score are encountered, The score will become zero. Such points on the backtrack path having zero score mark boundaries of subsequences.

The Figure 5.2 shows a motif from the responses of a neuron to different trials. The signal retrieved shows good similarity. At 50Hz sampling frequency, the motif lasts more than 1.5 seconds. Which is a significant time for two neurons to have very similar response. Also, similar phenomena occurs in each neuron we studied.

Motifs in responses of two neurons in a mice

Responses of two neurons to same visual stimuli are analyzed for motifs. Both neurons were sampled simultaneously during experiment. The Figure 5.3 shows score matrix and Figure 5.4 shows longest motif spotted. Note that in this case, motif has a duration of 2.5 seconds. In the score matrix, two disconnected paths can be found. Each one of these correspond to a common subsequence. Set of such subsequences is called LCSS. Motifs across mice have an average duration

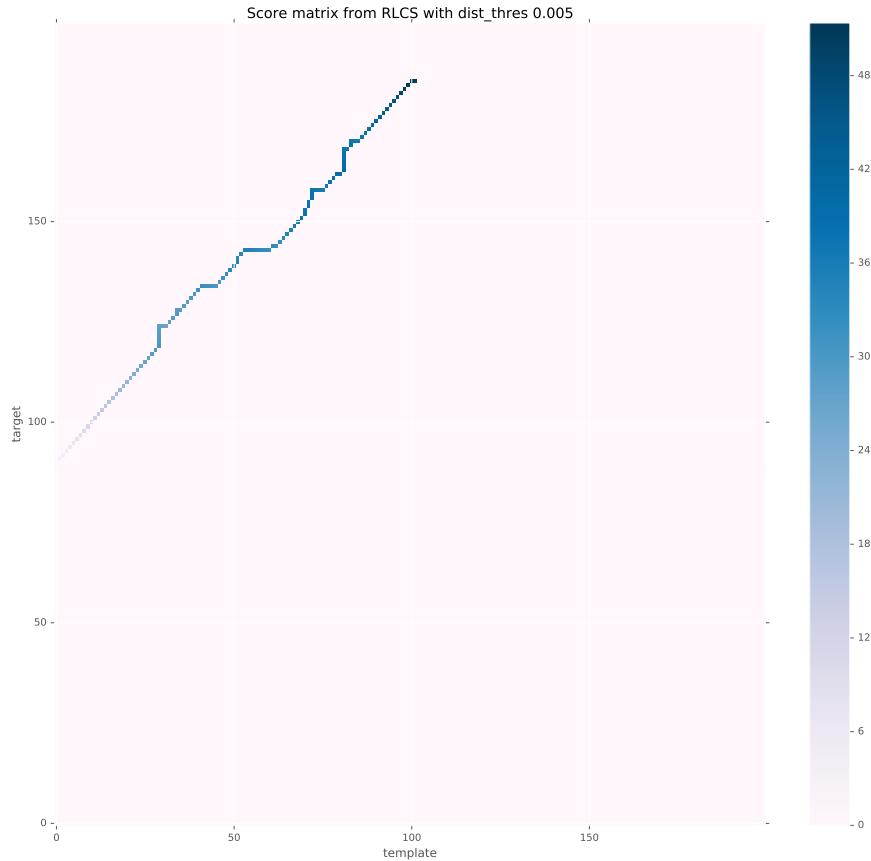


Figure 5.1: Score matrix from RLCS of responses a neuron to different trials.

about 1 seconds. More or less any two neurons taken for study, motifs of length 40 samples were spotted. This shows correlation in responses of neurons. In some cases, the length of motif exceeded 2 seconds.

Motifs in responses of neurons across different mice

Two neurons across two mice were taken for study. As these neurons does not have any interconnections, the motifs are expected to have less duration if at all they exists. The experimental study proved that motifs across mice occurs very less frequent than previous cases. Even when they exists, they have short duration

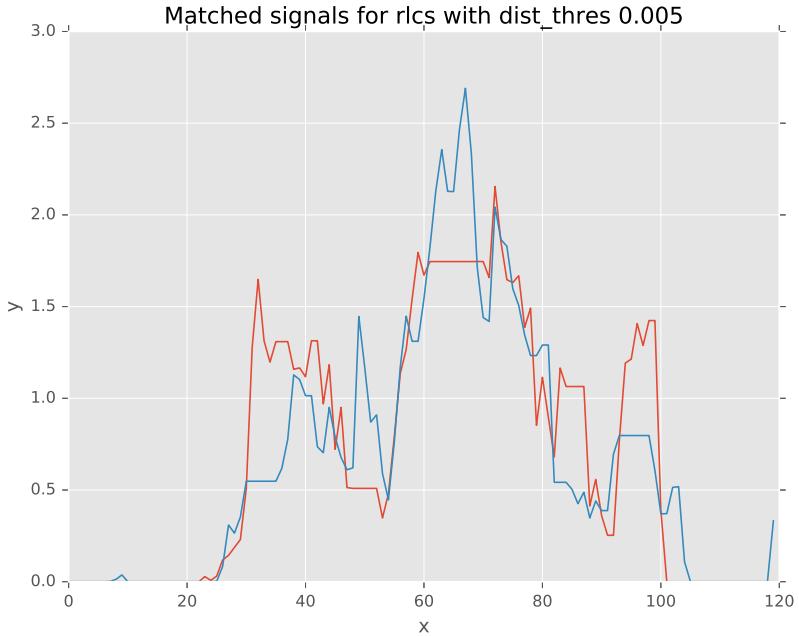


Figure 5.2: Extracted motif (longest) from responses a neuron to different trials.

~ 0.5 seconds. Motifs across mice could be caused by similar activation mechanism of neurons. This explains absence of long motifs across mice and presence of short subsequences. Exceptionally in some cases motifs of duration ~ 1.2 seconds were spotted across mice. Figure 5.5 and Figure 5.6 shows score matrix and extracted motif from responses of two neurons across mice. More plots of score matrix and motifs are in Appendix A.

5.2.2 Normalized RLCS score as a correlation measure

We have seen RLCS was very effective in query search, signal matching and correlation study even when classical correlations metric ‘Pearson correlation coefficient’ failed to detect correlation. As the information encoding in V1 largely depend on population of neurons and their inter-correlations, A robust measure of signal correlations needs to be used for analysis. We explore the possibility of normalized RLCS score between two signals as a robust correlation metric.

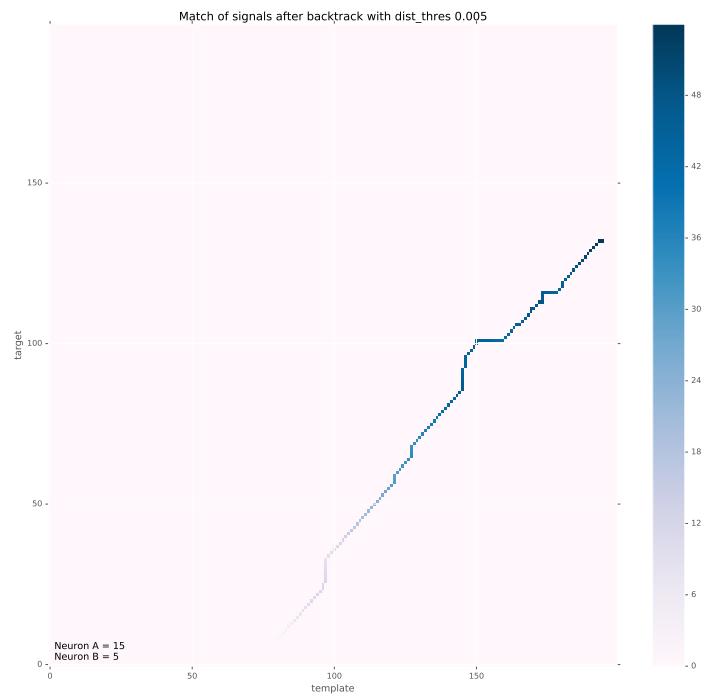


Figure 5.3: Score matrix from RLCS of responses of two neurons in a mouse.

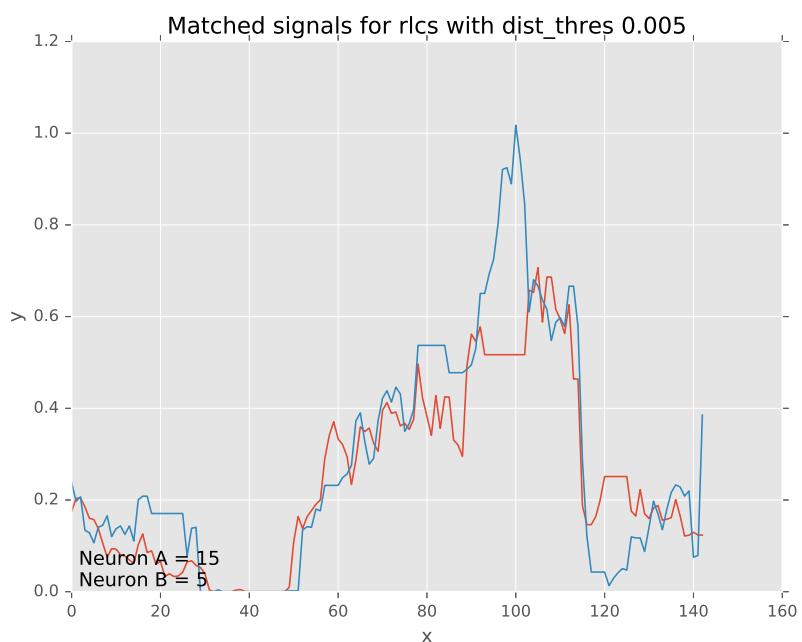


Figure 5.4: Extracted motif (longest) from responses of two neurons in a mouse.

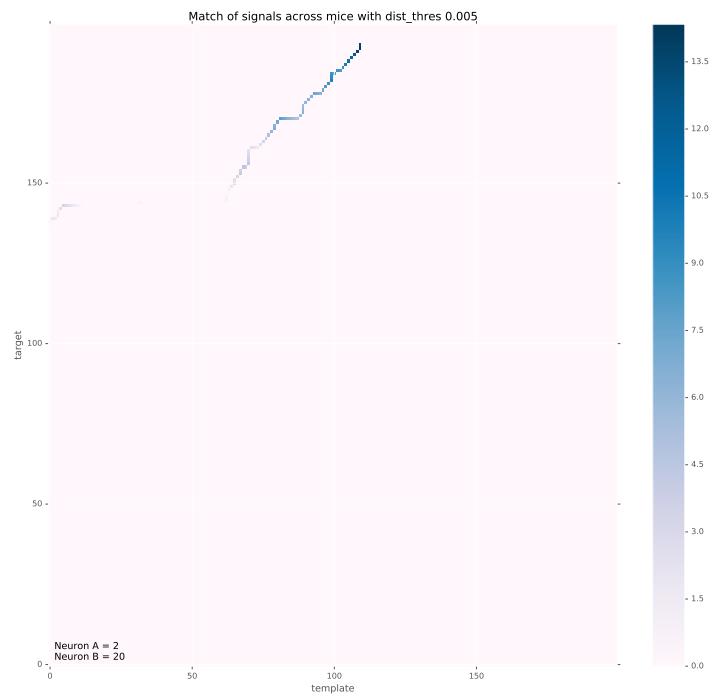


Figure 5.5: Score matrix from RLCS of responses of two neurons in a mouse.

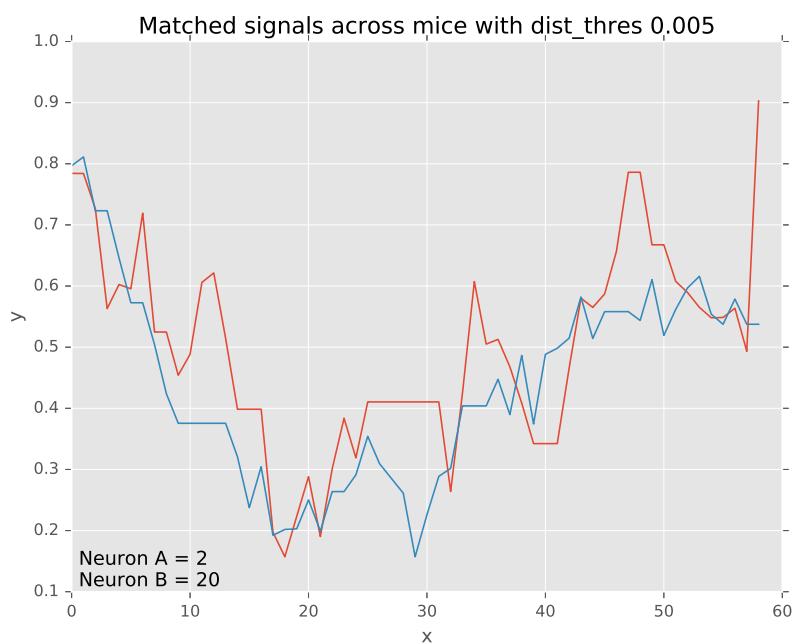


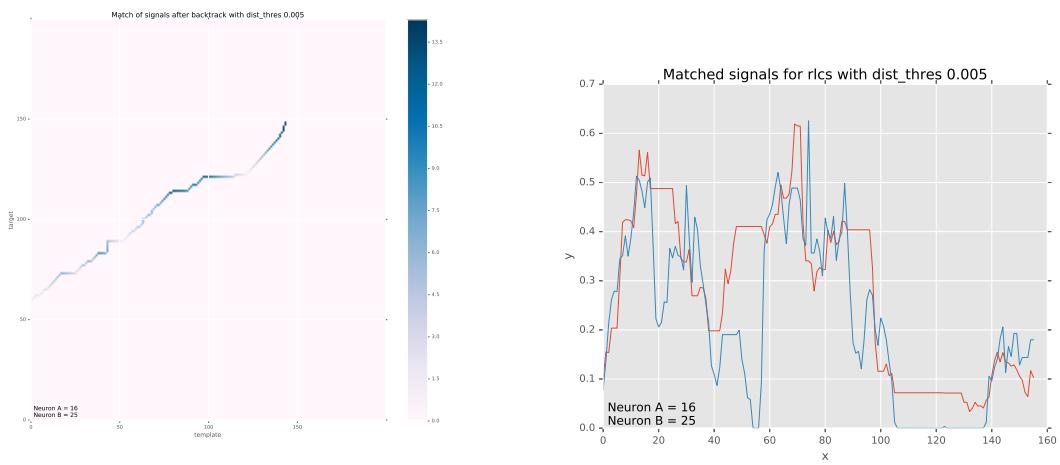
Figure 5.6: Extracted motif (longest) from responses of two neurons in a mouse.

The sum of RLCS scores contain information about how close the matched samples are, how long is the detected subsentence, how much gaps they have in between samples. This brings the consideration of using RLCS score as correlation measure. But to compare two scores, we need it to be normalized. We use standard z-normalization In which average score will be transformed to zero; all the scores above mean gets a positive score and every score below mean score gets a negative score. Please note that this is effective only when a large dataset is used as mean and standard deviation needs to be estimated. The standard z - score of a raw RLCS score s is

$$z = \frac{x - \mu}{\sigma}$$

Where $s = \sum_{i,j} \text{score}(i, j)$ is the raw RLCS score. μ is the mean and σ is the standard deviation of the population of scores.

As a demonstration, Figure shows two signals which have a low Pearson correlation $\rho = 0.18$ and good RLCS score. The extracted subsequence shows similarity of ~ 140 samples which was not detected by Pearson correlation. The score plot shows that many of the matches are not on the diagonal line. Pearson correlation is robust only to detect similarities in the diagonal line of score matrix.



(a) RLCS score matrix

(b) Extracted common subsequence

Figure 5.7: Demonstration of RLCS score's dominance over Pearson correlation as a metric of similarity. RLCS analysis of two signals having low Pearson correlation. Extracted common subsequence was not detected by Pearson correlation.

CHAPTER 6

Inferences and Future work

6.1 Inferences

Presence of motifs in neuronal responses is confirmed and it provided motivation to study motifs using RLCS algorithm. In the intertrial responses of a neuron, simple correlation measure failed to detect reliability of response. By extracting motifs from the same, we were able to obtain reliable subsequences of significant length ($\sim 60 - 120$ samples out from total of 200 samples) across trials. The similarity in signals could be attributed to reliability of information representation and similar activation mechanism of neurons.

As an effort to study response similarities of two neurons in a mice, common subsequence spotting is performed on time-series responses of two different neurons. The study resulted in detection of motifs of significant length ($\sim 40 - 60$ samples out from total of 200 samples) across neurons. Even after the receptive fields of two neurons being different, The presence of motifs could be due to information encoding by correlation of a population, neuronal interconnection and similar activation mechanism of neurons.

Responses of neurons across different mice are analyzed for the presence of motifs. The study is not expected to give long motifs as the two neurons are independent. The experimental results showed presence of common subsequences but as expected, the subsequences are small in length ($\sim 15 - 30$ samples out from

total of 200 samples). This could be attributed to similar activation mechanism of neurons. Exceptionally, few neurons across different mice produced longer subsequences (of length \sim 50 samples). Inferring meaning of this exception is left as future work.

6.2 Future work

The presence of significant motifs in neuronal signals is established in this work. But the impact the result has in the context of Neuroscience is yet to be studied. We have not studied the significance of signal extracted as a motif. We formulated a better measure for correlation between neurons. Interneuronal correlations play a great role in information encoding in brain. With the new correlation measure, information coding with inter-neuronal correlations can be studied further.

The current RLCS algorithm penalize a constant value for any mismatch. It does not take care how much the samples mismatch. The penalization can be selected as a function of distance between samples. This will create a better match of signals.

APPENDIX A

Additional results from RLCS

Chapter 5.2 presented only few examples of motif extraction for brevity. Additional results are provided here for reference. Each result consist of RLCS score matrix and longest motif extracted. Results are categorized based on neurons considered for the study.

A.1 RLCS of intertrial responses

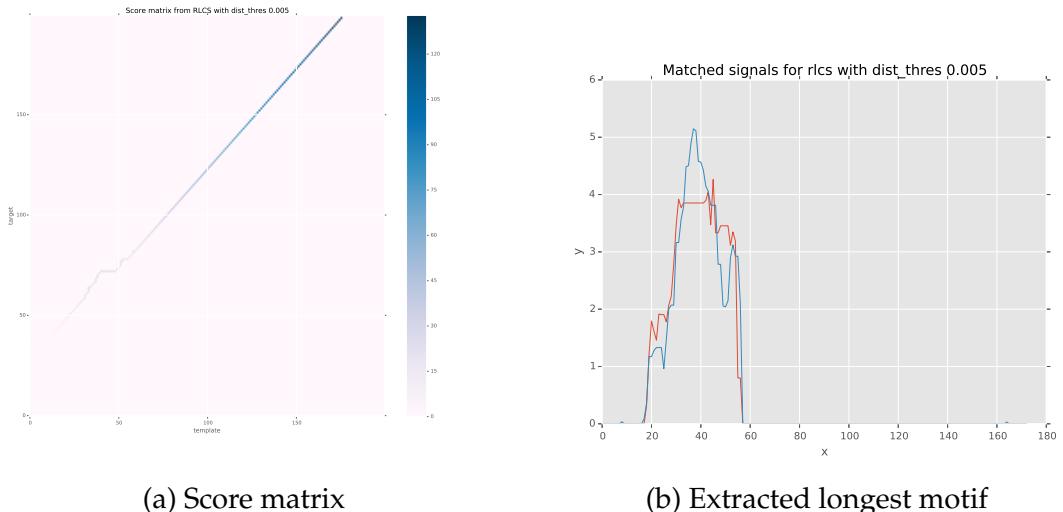
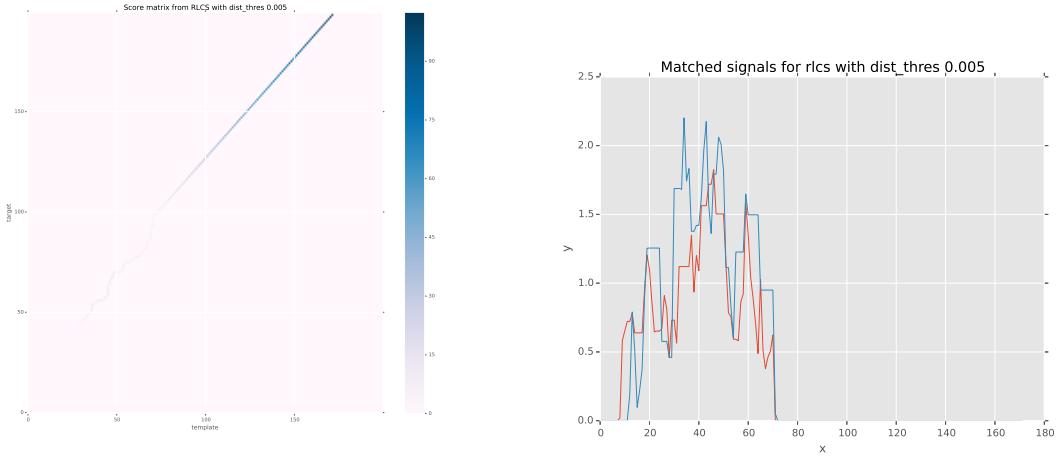


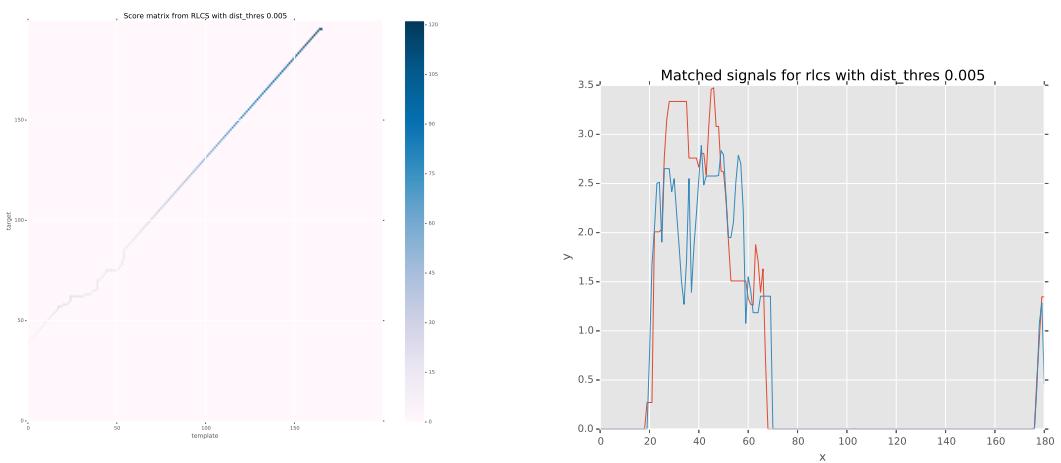
Figure A.1: RLCS performed on intertrial responses of a neuron. Note the presence of ~ 150 sample subsequence which is undetectable in Pearson correlation test.



(a) Score matrix

(b) Extracted longest motif

Figure A.2: RLCS performed on intertrial responses of a neuron. Note the presence of ~ 150 sample subsequence which is undetectable in Pearson correlation test.



(a) Score matrix

(b) Extracted longest motif

Figure A.3: RLCS performed on intertrial responses of a neuron. Note the presence of ~ 150 sample subsequence which is undetectable in Pearson correlation test.

A.2 RLCS of responses from two neurons in a mouse

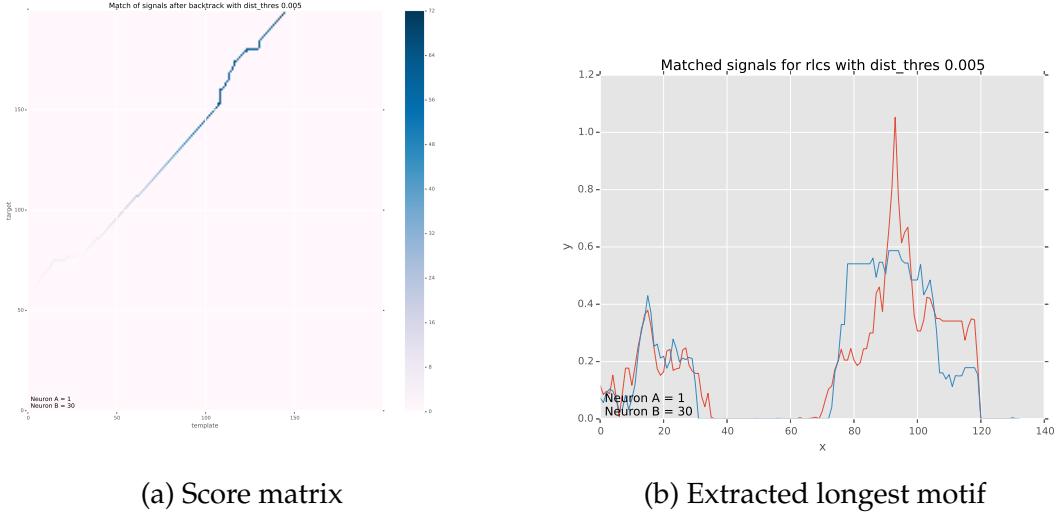


Figure A.4: RLCS performed on responses of two neurons within a mouse. The two signals have a pearson correlation of 0.18

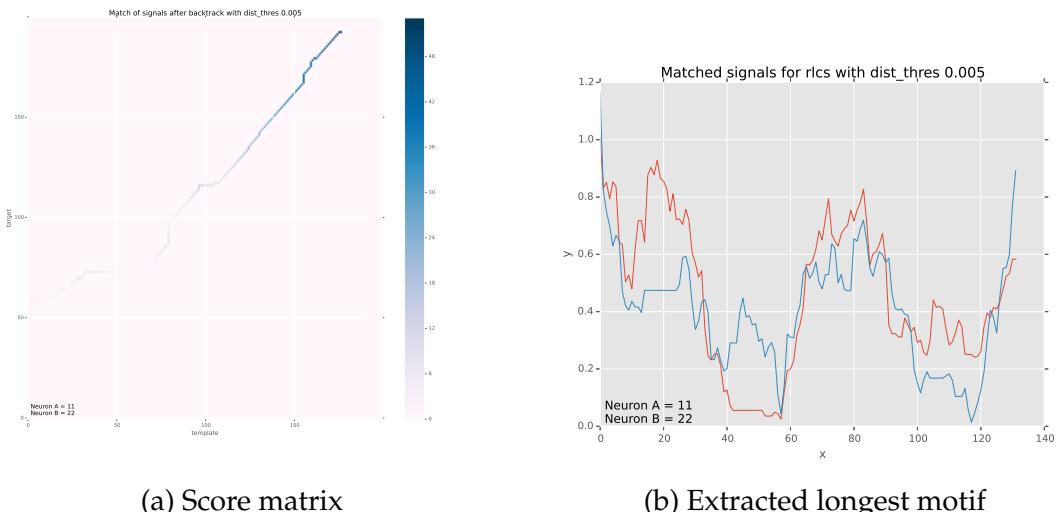


Figure A.5: RLCS performed on responses of two neurons within a mouse. The two signals have a pearson correlation of 0.06

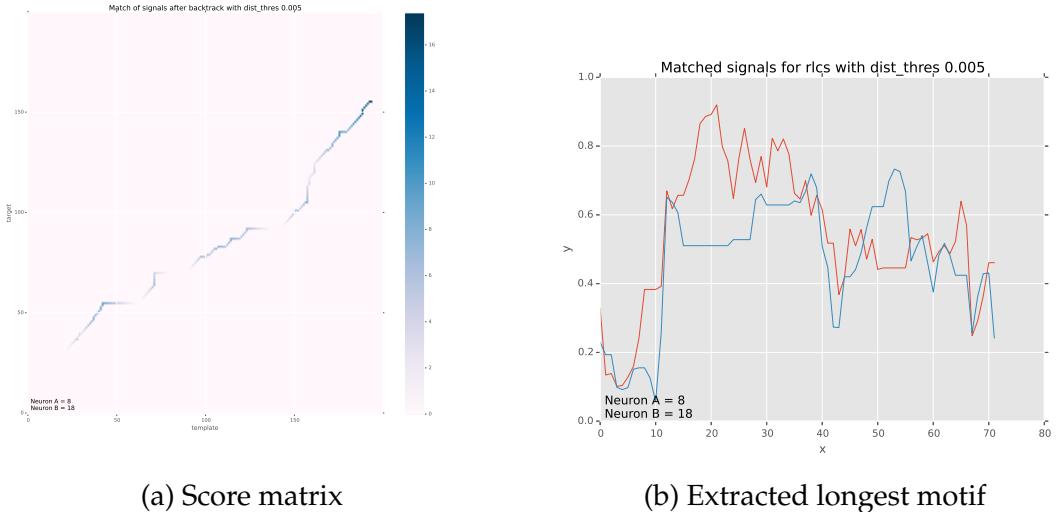


Figure A.6: RLCS performed on responses of two neurons within a mouse. The two signals have a pearson correlation of 0.02

A.3 RLCS of neuronal responses from two different mice

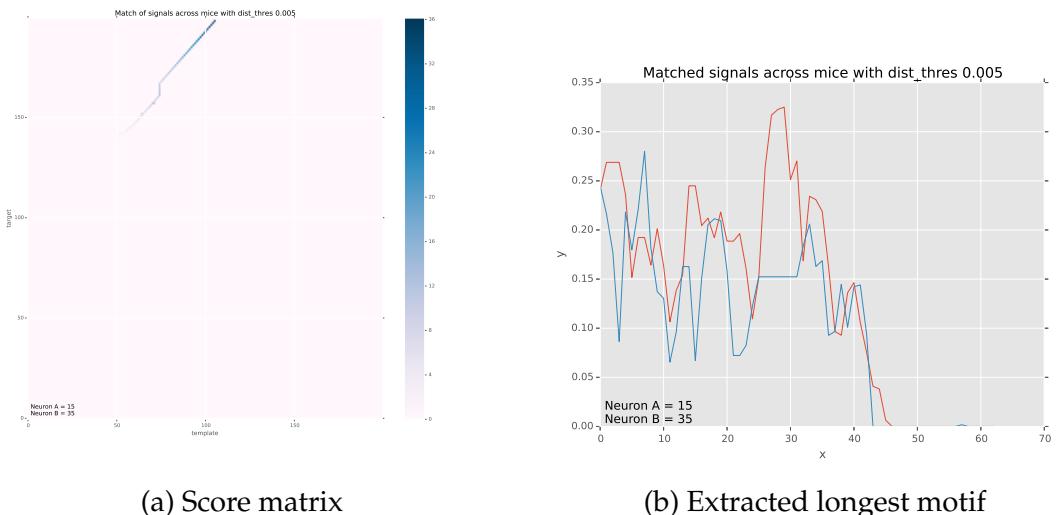


Figure A.7: RLCS performed on responses of two neurons from different mouse. Note that length of subsequence is smaller than other in cases.

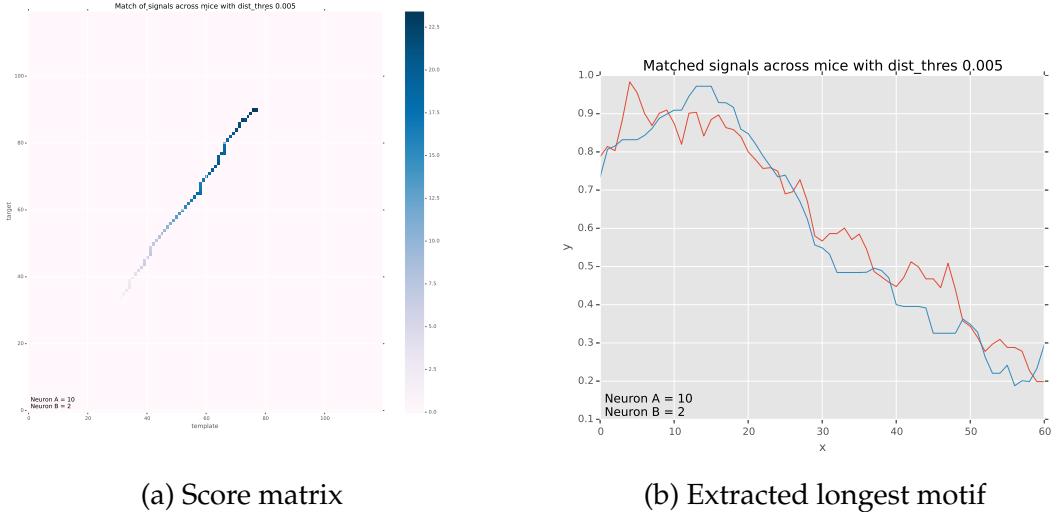


Figure A.8: RLCS performed on responses of two neurons from different mouse. This motif *across mice* last for 3 seconds. (This data is from drifting sinusoidal grating experiment, in which sample rate is 20Hz)

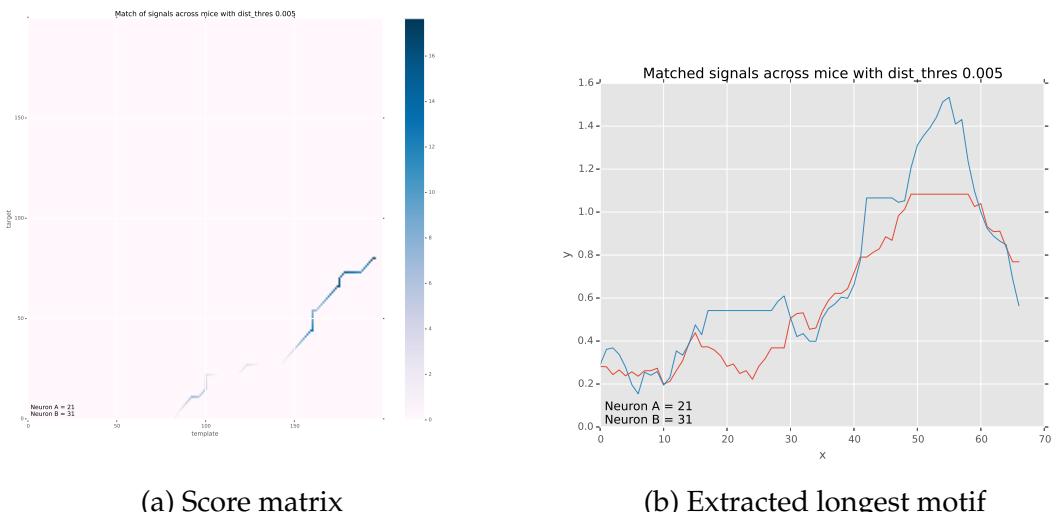


Figure A.9: RLCS performed on responses of two neurons from different mouse. Note that length of subsequence is smaller than other in cases.

REFERENCES

- Ayzenshtat, I., A. Gilad, G. Zurawel, and H. Slovin** (2012). Population response to natural images in the primary visual cortex encodes local stimulus attributes and perceptual processing. *The Journal of Neuroscience*, **32**(40), 13971–13986.
- Barlow, H., R. Fitzhugh, and S. Kuffler** (1957). Change of organization in the receptive fields of the cat's retina during dark adaptation. *The Journal of physiology*, **137**(3), 338.
- Bear, M., B. Connors, and M. Paradiso**, *Neuroscience: Neuroscience: Exploring the Brain*. Lippincott Williams & Wilkins, 2007. ISBN 9780781760034. URL <https://books.google.co.in/books?id=75NgwLzueikC>.
- Dutta, S., K. S. PV, and H. A. Murthy** (2013). Raga verification in carnatic music using longest common segment set.
- Hubel, D.**, *Eye, Brain, and Vision*. Scientific American Library Series. Henry Holt and Company, 1995. ISBN 9780716760092. URL <https://books.google.co.in/books?id=2Id9QgAACAAJ>.
- Hubel, D. H. and T. N. Wiesel** (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, **148**(3), 574–591.
- Ishwar, V., S. Dutta, A. Bellur, and H. A. Murthy**, Motif spotting in an alapana in carnatic music. 2013.
- Lin, H.-J., H.-H. Wu, and C.-W. Wang** (2011). Music matching based on rough longest common subsequence. *J. Inf. Sci. Eng.*, **27**(1), 95–110.
- Mazurek, M., M. Kager, and S. D. Van Hooser** (2015). Robust quantification of orientation selectivity and direction selectivity. *Neural Circuits Revealed*, 127.
- Reich, D. S., F. Mechler, K. P. Purpura, and J. D. Victor** (2000). Interspike intervals, receptive fields, and information encoding in primary visual cortex. *The Journal of neuroscience*, **20**(5), 1964–1974.
- Rikhye, R. V. and M. Sur** (2015). Spatial correlations in natural scenes modulate response reliability in mouse visual cortex. *The Journal of Neuroscience*, **35**(43), 14661–14680.
- Rosenbaum, R., T. Tchumatchenko, and R. Moreno-Bote**, *Correlated neuronal activity and its relationship to coding, dynamics and network architecture*. Frontiers Research Topics. Frontiers E-books, 2014. ISBN 9782889193578. URL <https://books.google.co.in/books?id=AQNdBgAAQBAJ>.
- Tiesinga, P., J.-M. Fellous, and T. J. Sejnowski** (2008). Regulation of spike timing in visual cortical circuits. *Nature reviews neuroscience*, **9**(2), 97–107.
- Vogelstein, J. T., A. M. Packer, T. A. Machado, T. Sippy, B. Babadi, R. Yuste, and L. Paninski** (2010). Fast nonnegative deconvolution for spike train inference from population calcium imaging. *Journal of neurophysiology*, **104**(6), 3691–3704.