

## Lecture 56

Students are recommended to read this paper -

<http://www.cns.nyu.edu/~tony/vns/readings/olshausen-field-1996.pdf> (Olshausen & Field, 1996).

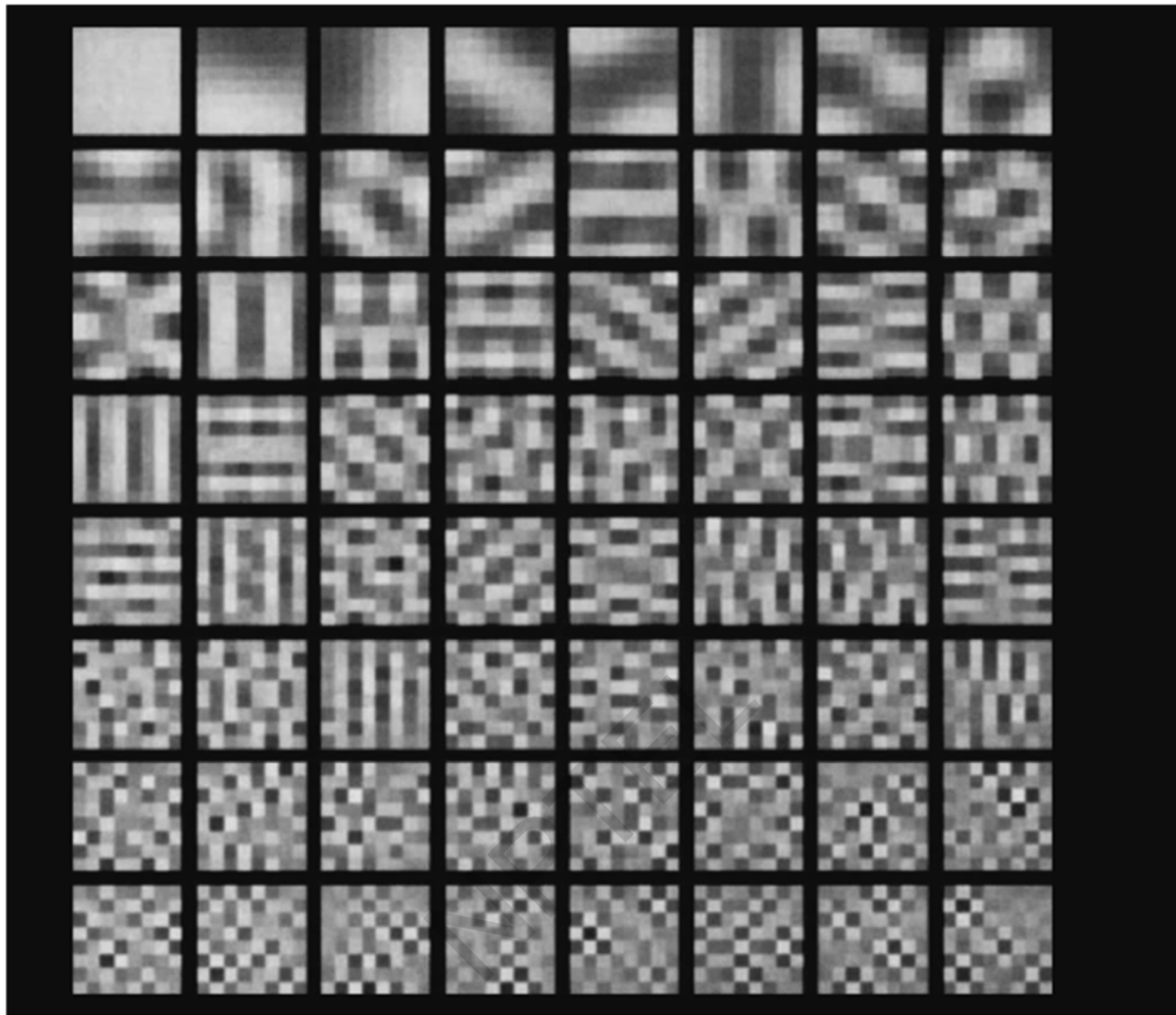
It is only of 3 pages. By reading the text, one can get intuition about basics ideas of efficient coding.

Here is a brief summary of the paper

- The paper tries to explain what fundamental principles of neural coding may lead to simple cell like properties in Visual system. Spoiler - They propose that it is preserving information (minimizing error) and sparse representation.
- If an image is to be encoded, it is assumed that it needs to be represented as linear sum of a basis functions.

$$I(x, y) = \sum_i a_i \phi_i(x, y)$$

- The goal is to find  $\phi_i$  s such that image  $I$  is efficiently encoded.
- The authors first try with PCA and don't get anything close to simple cells



- They postulate that PCA can work well only in gaussian properties but most of the images are not simple gaussian. Hence PCA fails to efficiently find basis encodings for images.
- So, they hypothesize that simple cell like receptive fields can be obtained if one takes into account 2 things - 1. Minimizing error. 2. Maximising sparseness.

$$E = -[\text{preserve information}] - \lambda[\text{sparseness of } a_i] \quad (2)$$

- Each term in cost function is explained as follows:

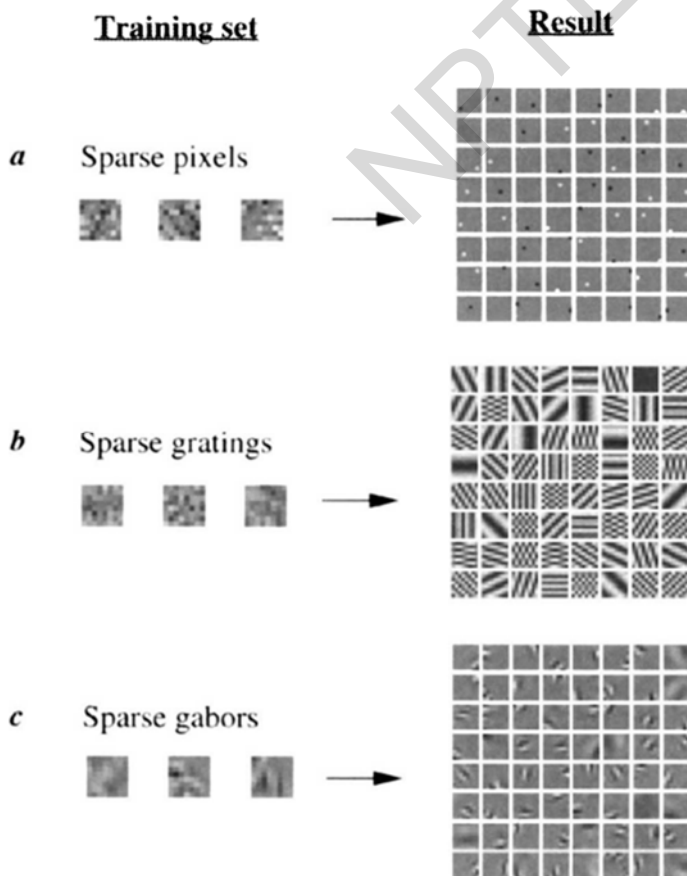
$$[\text{preserve information}] = - \sum_{x,y} \left[ I(x,y) - \sum_i a_i \phi_i(x,y) \right]^2 \quad (3)$$

$$[\text{sparseness of } a_i] = - \sum_i S\left(\frac{a_i}{\sigma}\right) \quad (4)$$

where  $\sigma$  is a scaling constant. The choices for  $S(x)$  that we have experimented with include  $-e^{-x^2}$ ,  $\log(1+x^2)$  and  $|x|$ , and all yield qualitatively similar results (described below). The reasoning behind these choices is that they will favour among activity states with equal variance those with the fewest number of non-zero coefficients. This is illustrated in geometric terms in Fig. 2.

Learning is accomplished by minimizing the total cost functional,  $E$  (equation (2)). For each image presentation,  $E$  is minimized with respect to the  $a_i$ . The  $\phi_i$  then evolve by gradient descent on  $E$  averaged over many image presentations. Thus for a given image, the  $a_i$  are determined from the equilibrium solution to the differential equation:

- On minimizing this cost function, they get receptive fields, that are similar to that of simple cells.



- To conclude, we say that in visual system, simple cell like receptive fields are formed as a result of minimizing error and maximizing sparseness.

## Lecture 57

Students are recommended to read this paper -

<https://www.nature.com/articles/nature04485#Abs1> (Smith & Lewicki, 2006). This paper is relatively hard to read due to the maths and signal processing involved. Hence, trying to get the basic ideas would suffice. *All images and text below are taken from the [Smith & Lewicki, 2006 Nature paper](#)*

Here is a brief summary of the paper:

- The auditory neural code is essential for various auditory tasks, requiring sensitivity in time and frequency amidst diverse natural sounds.
- Efficient coding strategies in sensory systems may have evolved to optimize information conveyed to the brain while **minimizing energy** and **resource use**.
- Optimizing features for coding natural sounds or speech reveals similarities to time-domain cochlear filter estimates and a frequency-bandwidth dependence like auditory nerve fibers.
- Findings suggest that the auditory code may be **near an information-theoretic optimum**, and speech's acoustic **structure might be adapted to the mammalian auditory system's coding capacity**.

Brief steps:

- To start with just like in previous paper, the input signal is assumed as linear sum of different basis functions (here kernel functions)

To test this hypothesis, we must start with a mathematical description of an acoustic waveform that can then be used to derive theoretically optimal codes. Here we use a model in which sounds are encoded as a pattern of spikes<sup>8,9,10</sup>. The signal,  $x(t)$ , is encoded with a set of kernel functions,  $\varphi_1, \dots, \varphi_m$ , that can be positioned arbitrarily and independently in time. The mathematical form of the representation with additive noise is

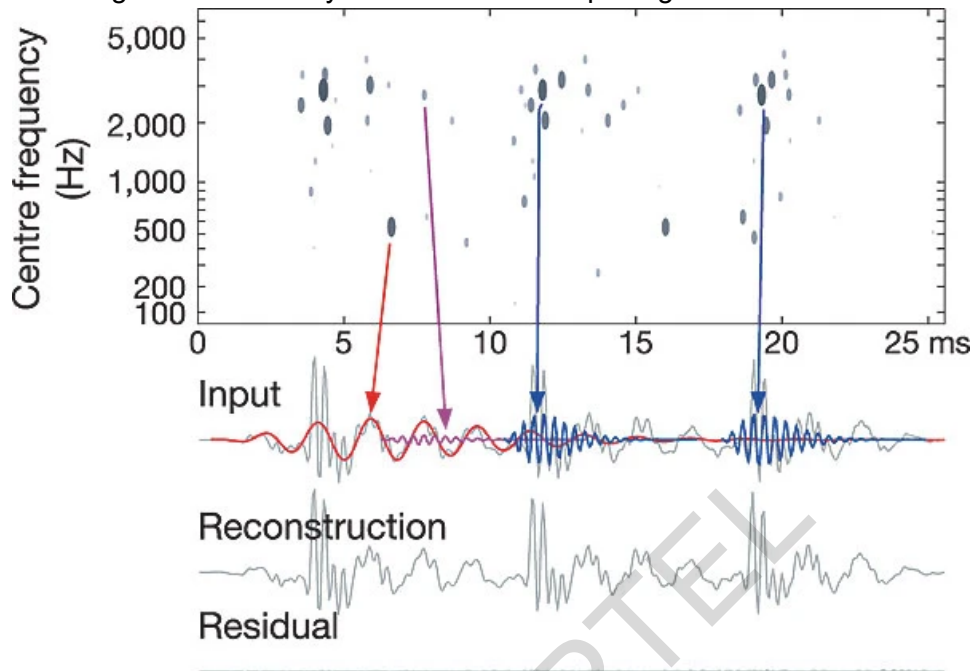
$$x(t) = \sum_m \sum_i s_i^m \phi_m(t - \tau_i^m) + \varepsilon(t) \quad (1)$$

where  $\tau_i^m$  and  $s_i^m$  are the temporal position and coefficient of the  $i$ th instance of kernel  $\varphi_m$ , respectively. Note that the number of instances of  $\varphi_m$  need not be the same across kernels.

- Now, the goal is to find  $\tau_i$  s and  $s_i$  s. They do it using an algorithm called matching pursuit algorithm

Here we compute the values of for a given signal by using a **matching pursuit algorithm** which iteratively approximates the input signal and has been shown to yield highly efficient representations for a broad range of sounds.

- The algorithm efficiently reconstructs the input signals as shown in below image



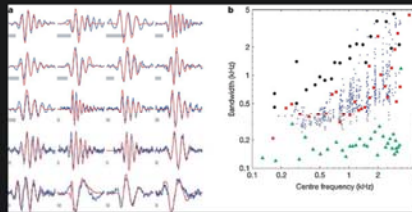
#### Kernel functions and Revcor filters:

All kernels are time-localized, have a narrow spectral bandwidth and show a strong temporal asymmetry not predicted by previous theoretical models. The sharp attack and gradual decay of the envelope match the physiological filtering properties of auditory nerves as characterized by reverse-correlation, which estimates the **impulse response function of individual auditory nerve fibres, the so-called 'revcor' filters**. Kernel functions in Fig. 2a are overlaid on revcor filters obtained from cat auditory nerves,

- The authors also show that **efficient codes for speech** immediately yielded **kernels that closely matched the auditory revcor filters**.

**Figure 2: Efficient codes for natural sounds predict revcor filter shapes and population characteristics.**

From: [Efficient auditory coding](#)



**a.** When optimized to encode an ensemble of natural sounds, kernel functions become asymmetric sinusoids (smooth curves in red, with padding removed) with sharp attacks and gradual decays. They also adapt in temporal extent, with longer and shorter functions emerging from the same initial length (grey scale bars, 5 ms). Each kernel function is overlaid on a revcor function obtained from cat auditory nerve fibres (noisy curves in blue). **b.** The bandwidth-centre-frequency distribution of learned kernel functions (red squares) is plotted together with cat physiological data (small blue dots) and with kernel functions trained on environmental sounds alone (black circles) or animal vocalizations alone (green triangles).

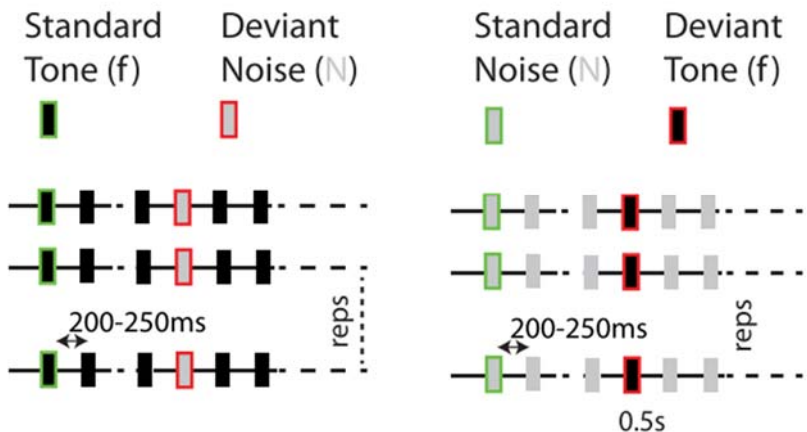
Source - <https://www.nature.com/articles/nature04485/figures/2>

## Lecture 58

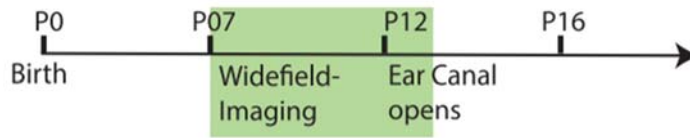
REF :- <https://www.jneurosci.org/content/42/8/1454>

- Subplate neuron (SPNs), the first layer of neurons plays key role in cortical responses
- They are found to be earliest responding layer in the ACx
- SPNs receive inputs from the thalamic axons projecting to L4, and drive L4 activity over a time window of development
- Before ECO (ear canal opening), SPNs have unique response properties. SPNs, unlike cortical plate neurons, show selectivity to oddball sounds in a stream of repeating sounds before ECO, while after ECO, cortical plate L4 neurons show stronger deviant detection than remaining SPNs
- Stimulus used (tone-noise (f-N) oddball stimulation protocol with the swap on the right.)

Swap of left



- Experimental timeline



- P1( postnatal day 1) , P16( post natal day 16)
- 
- It was found that the SPN s and neurons in the different cortical layers responded to sounds before ear canal opening as early as in P7 in mice
- 

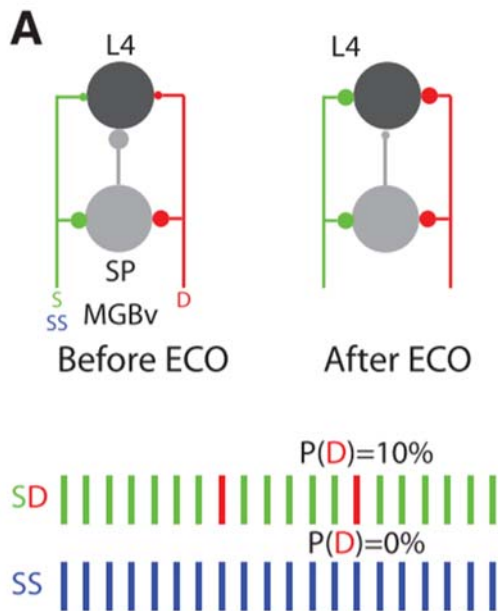
Response properties of SPNs and L4 neurons in the context of stimulus-specific adaptation (SSA) was observed:- common selectivity index (CSI) was used to quantify a neuron's deviant detection strength

$$CSI = \frac{RD'(f) + RD(N) - RS(f) - RS(N)}{RD(f) + RD(N) + RS(f) + RS(N)}$$

where RS(.) and RD(.) denote response to S and D, respectively.

- CSI metric finds this deviant preference with respect to an adapted standard stimulus and thus measures SSA.
- SPNs are stronger deviant detectors than L4, while the relative deviant detection strength switches after ECO.

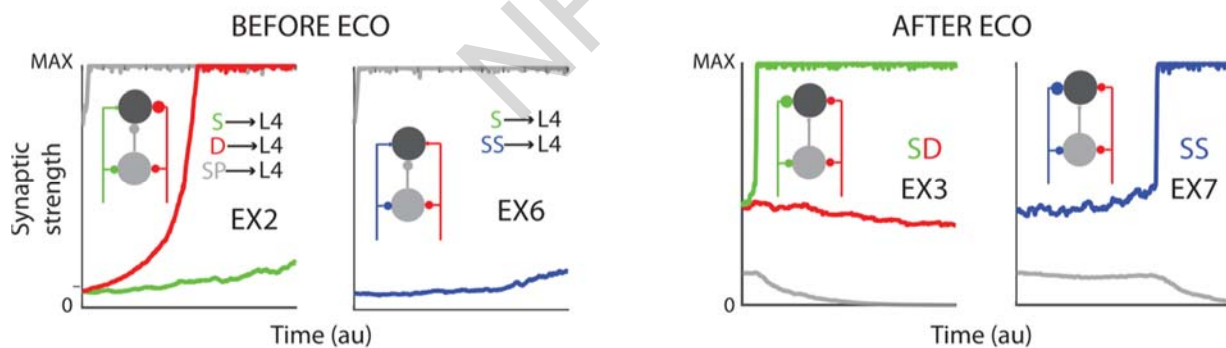
**Model binary network with observed oddball selectivity of SPNs and L4 shows a stronger representation of deviant frequency before ECO and standard frequency after ECO**



- size of the blobs representing synaptic weights, which changes across time.

- How does thalamic input to L4 develop?

Before and after ECO



- Before ECO with higher oddball selectivity in SP than in L4, with an SD exposure, the D thalamic input to L4 stays after ECO, with the oddball selectivity
- After ECO, with the oddball selectivity reversed in SP and L4, D to L4 synapse weakens, and S to L4 strengthens, as L4 spikes more because of thalamic input spikes in that of S and not for that of D strengthens faster than the S thalamic input

— Sparse coding and information maximization principles imply a strengthening of response to low probability stimuli



- The objective function has a mutual information term between the response (R) and the stimuli token (T) that will increase as successive stimuli tokens come. A sparseness constraint term containing the mean rate limits the increase in mutual information. We call this metric as the Constrained Information (CI). This function was optimized in two separate regimens corresponding to the two separate regimens
- Gaussian probability distribution was assumed, with a constant variance but variable mean

$$\text{Constrained information (CI)} = \text{Mutual information (MI)} - \rho(\text{mean rate})$$

- sparseness in later stages (after ECO), when spike rates are higher, needs to be weighted more than earlier stages (before ECO) to explain the differential effect observed before and after ECO in response to the low probability stimulus.
- 

## Lecture 59

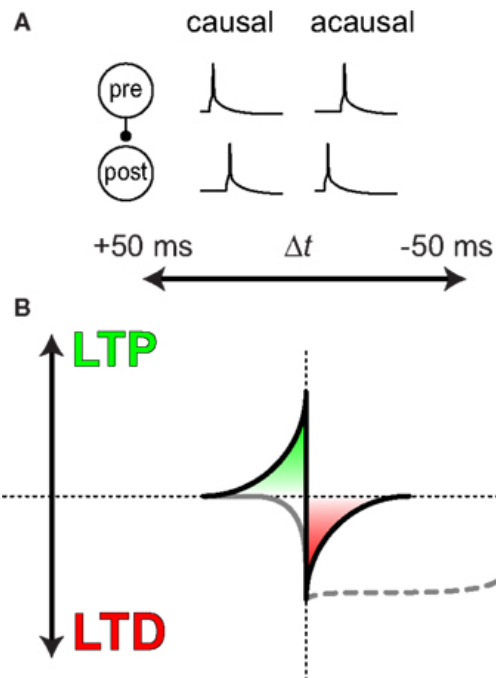
Students are recommended to go through this paper -

<https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=cf2f91325c9517ea8a97a3b05b3480989370d934>

All text and images (otherwise mentioned) are taken from [G. Chechik, 2003 paper](#)

Here is what the paper is briefly about:

We have studied Spike timing-dependent plasticity (STDP) and Hebbian Plasticity. If post-synaptic neuron is followed by a pre-synaptic spike, the weight increases and vice-versa. The following graph explains Hebbian STDP rule:



[Source](#)

The above is an experimental result. Gal Chechik in this work tries to produce this result theoretically with pen and paper.

If you have a post-synaptic neuron receiving a set of inputs, they will follow Hebbian STDP as above if the firing pattern of the neuron is aimed to maximize Mutual information between pre-synaptic and post-synaptic activity. And the maximization is done through a Gradient descent learning rule (See [what is gradient descent](#)).

Papers like this give an idea that observed biological phenomena underlie some fundamental principles like maximizing *reduction in uncertainty* (Mutual information) play a key role in governing how weights between neurons should change.

Briefly the steps:

- **Theoretical Setup:**  $N$  input neurons affect the activity of a single output neuron  $Y$ . The input current of a pre-synaptic neuron is convolution of kernel with its spike train.

We study a generic learning task in a network with  $N$  input neurons  $S_1, \dots, S_N$  firing spike trains, and a single output (target) neuron  $Y$ . At any point in time, the target neuron integrates its inputs with some continuous temporal filter  $F$  due to voltage attenuation and a synaptic transfer function

$$Y(t) = \sum_{i=1}^N W_i X_i(t) \quad (1)$$

$$X_i(t) \equiv \int_{-\infty}^t F_\tau(t-t') S_i(t') dt' \quad ; \quad \int_0^\infty F_\tau(t) dt = 1$$

where  $W_i$  is the amplitude of synaptic efficacy between the  $i$ -th input neuron and the target neuron,  $S_i(t) = \sum_{t_{spike}} \delta(t - t_{spike})$  is the  $i$ th spike train. The filter  $F$  may be used to consider general synaptic transfer functions and voltage decay effects. For example, voltage attenuation due to leak currents in a passive membrane is realized as an exponential filter  $F_\tau(t) = \frac{1}{\tau} \exp(-t/\tau)$ ,  $\forall t > 0$  with  $\tau$  being the membrane time constant.

- Weights are assumed to change such that the changed weights try to **maximize Mutual information** through **Gradient Descent rule**.

### 3 Mutual Information Maximization

The goal of the current section is to derive a learning algorithm that maximizes the input-output mutual information of the above model by changing the synaptic weights. We first focus on the modification of the synaptic magnitudes  $W$ , while keeping the temporal filter  $F$  fixed. Learning the optimal temporal filter is discussed in section 6.3.

#### 3.1 Deriving a gradient ascent learning rule

Let us focus on a single presentation period and look at the value of  $Y = Y(T)$  at the end of this period. Assuming that the temporal filter is mostly concentrated in a time period of length  $T$  and omitting the notation of  $t$ , we obtain from Eq. 1

$$Y = \sum_{i=1}^N W_i X_i \quad ; \quad X_i = \int_{-T}^0 F_\tau(0-t') S_i(t') dt' \quad ; \quad (2)$$

-

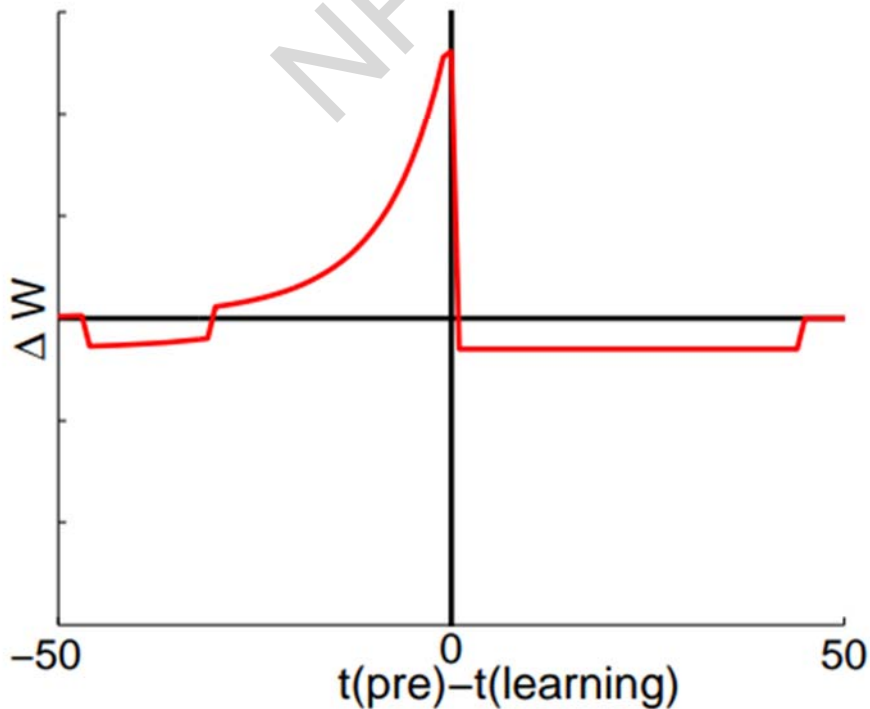
The mutual information (Shannon, 1948; Cover & Thomas, 1991) (and see also (Linsker, 1992)) between the output and pattern identity in this network is defined by

$$I(Y; \eta) = h(Y) - h(Y|\eta) \quad ; \quad h(Y) = - \int_{-\infty}^{\infty} f(y) \log(f(y)) dy \quad (3)$$

where the  $h(Y)$  is the differential entropy (Cover & Thomas, 1991) of the distribution of  $Y$ , and  $h(Y|\eta)$  is the differential entropy of the  $Y$  distribution given that the network is presented with a known input pattern  $\eta$ .  $f(Y)$  is the p.d.f. of  $Y$ . This mutual information measures how easy it is to decide which input pattern  $\eta$  is presented to the network by observing the network's output  $Y$ .

- The result, we get a graph similar to what we see experimentally,

**B.**



- (Note that we have been used to seeing post spike time - pre spike time, here the graph is pre spike time - post spike time. The x-axis is inter-changed)

## Lecture 60

The Lecture discusses some important unsolved problems in neuroscience with an enlisting of topics we discussed

### Models of Neuron

We discussed that the neural systems can be modelled as one of three

1. Biophysical model: Hodgkin-Huxley neuron model
2. Leaky Integrate and Fire model: A simpler model that captures the outcome of the neurons but NOT the underlying physical process
3. Linear Time Invariant Model: A linear model that maps from input to spikes with unit-impulse response, static non-linearity and Poisson Point Process generator.

### Receptive Field:

We discussed biologically present receptive fields that are present in Visual System and Auditory system, and how a representative neuron responds to the receptive field.

We also discussed Maximally informative dimensions, a theoretical perspective of it.

### Decoding:

In neural decoding, we saw the following methods:

1. Discrimination: Using ROC-curve, trying to decode how animal discriminates from single cell responses
2. Statistical methods of Decoding: Using Mutual Information
3. Biologically plausible methods: Van Rossum Distance

### About Important problems of Neuroscience:

- Learning and Memory will always be an important problem as these aspects are key to intelligent behaviour. Understanding the detailed mechanisms and underlying principles behind STP/D and LTP/D is an important problem.

- Matching Sensory to Motor Output: With the increase of high throughput data, new methods have arisen that try to build models that map all of the sensory inputs of an animal to all of the motor outputs. However, these models might be very animal and task specific, and may not help to find general principles that guide neural activity and animal behaviour.

- State Dependence: Neural responses and behaviour of any animals depend on animal's internal state. Though in this course, we have not discussed anything about state, many studies focus on investigating neural mechanism that are responsible for an animal's internal state. (For

more discussion, read [https://www.cell.com/neuron/pdf/S0896-6273\(18\)30905-X.pdf](https://www.cell.com/neuron/pdf/S0896-6273(18)30905-X.pdf) - *New Methods for Assessing Internal State Reveal Its Impact on Neural Activity and Behavior* section)

- High Dimensionality Data: With increase in technology, new experimental methods collect a large volume of data. New methods have been continuously developed to analyse and visualize this kind of high dimensional data(For example, see UMAP- [UMAP paper](#) ,[video explaining UMAP](#), [A paper using UMAP to visualize Mouse vocalization data- Figure 3](#)).

NPTEL