ELSC
The Edmond & Lily Safra Center for Brain Sciences

Analysis and theory of perceptual learning in auditory cortex

Ravid Shwartz - Ziv

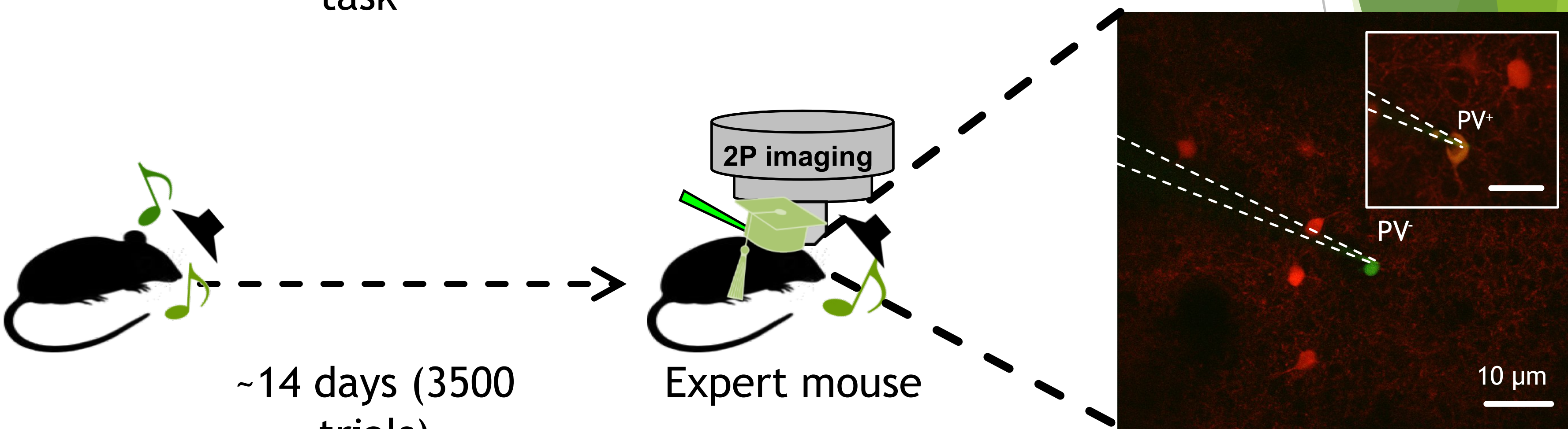
Advisor: Haim Sompolinsky

Experiment: Ido Maor and Adi Mizrahi

Experimental design

1. Perceptual Training
Two tones discrimination task

2. *In-vivo* imaging and electrophysiology of inhibitory (PV^+) and excitatory (PV^-) neurons



~14 days (3500 trials)

Expert mouse

Naïve mouse

The goals

- Building statistic modeling of neural coding in A1.
- Quantify the changes that accrued in the neural code due to perceptual learning.
- Learning model of perceptual learning with Fisher Information

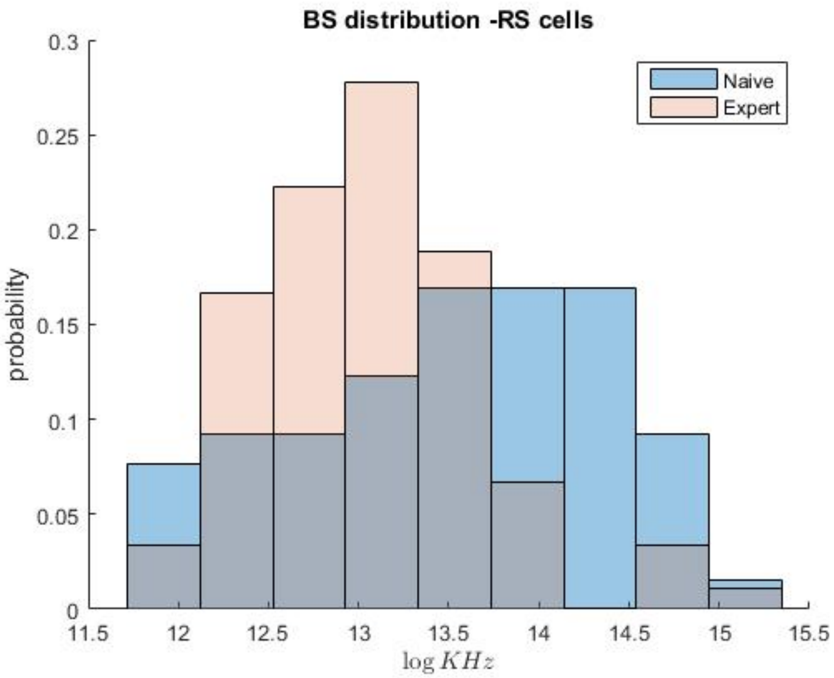
Statistical models - Gaussian

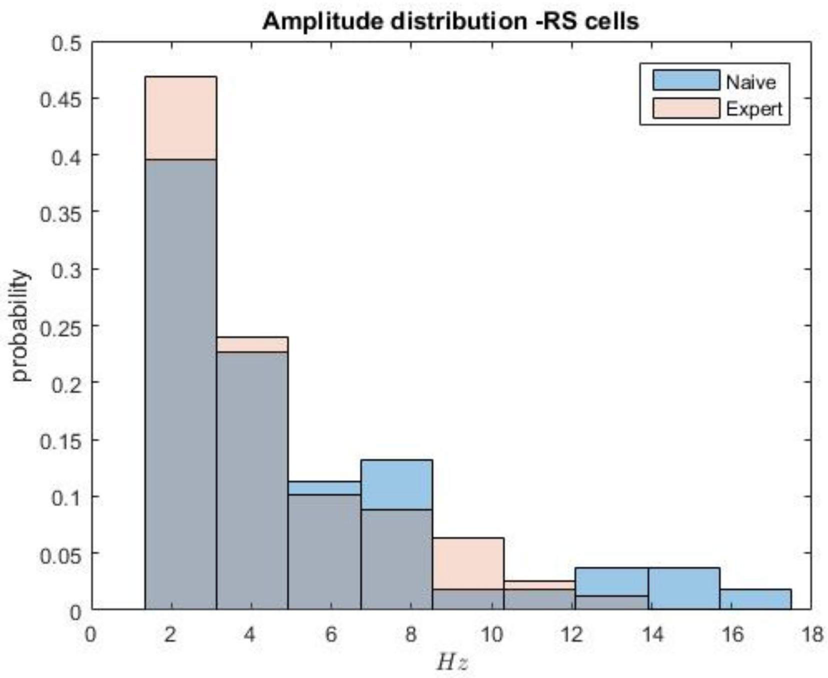
- Fitting the Firing Rate of the cells with Gaussian

$$f(x) = a \exp\left(-\left(\frac{x-b}{c}\right)^2\right) + d$$

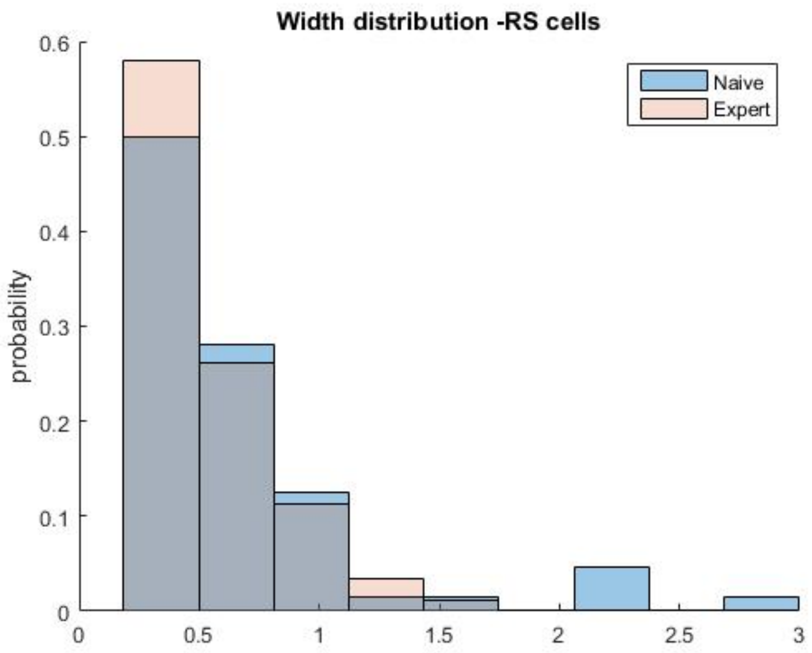
- Remove all the cells that don't fit ($R^2 \leq 0.6$)

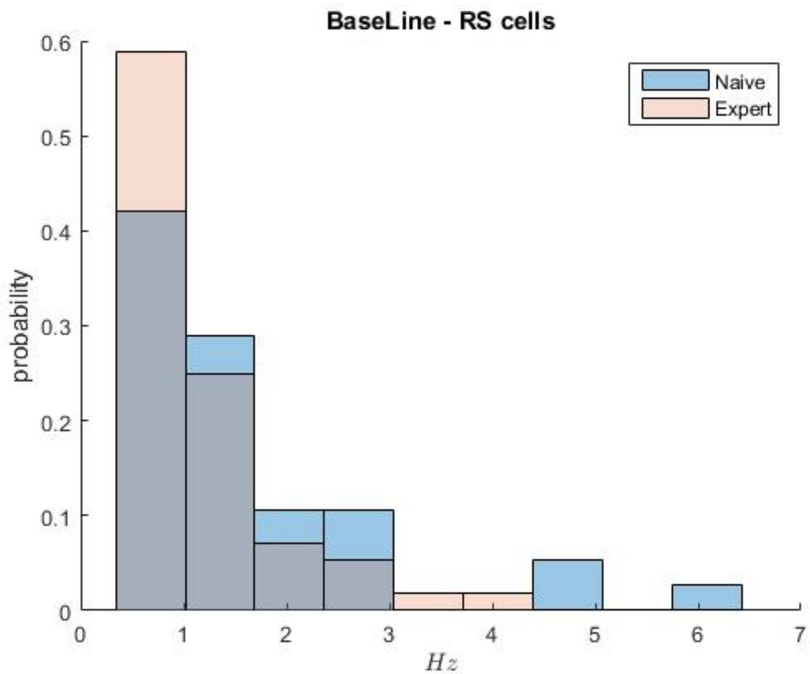
Parameters distribution





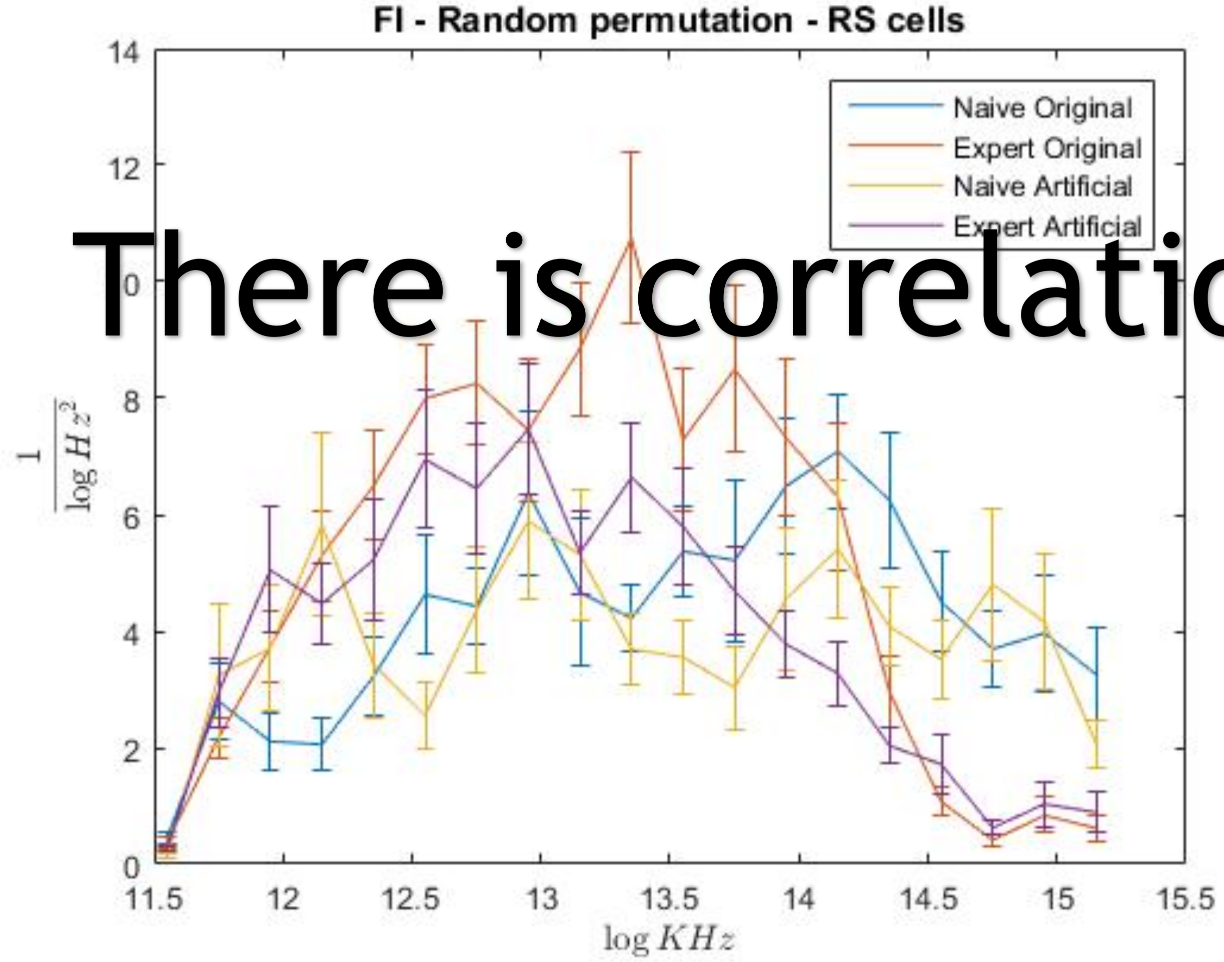
Parameters distribution





Is there correlation between the parameters?

Sampling new cells from the Gaussian's parameters

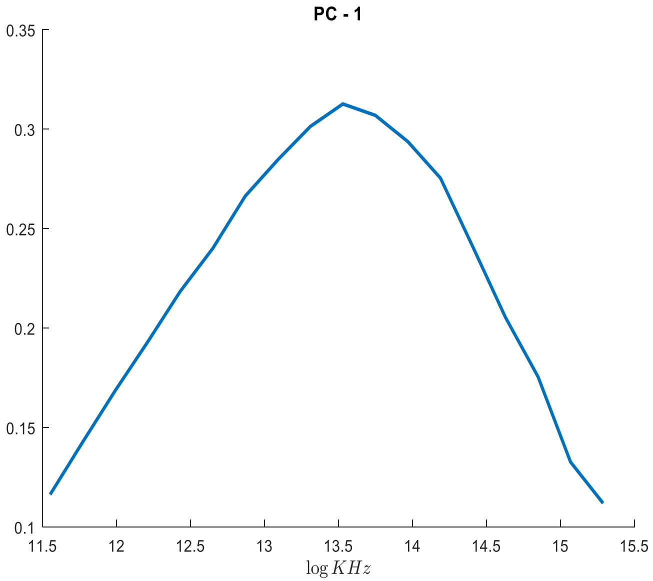


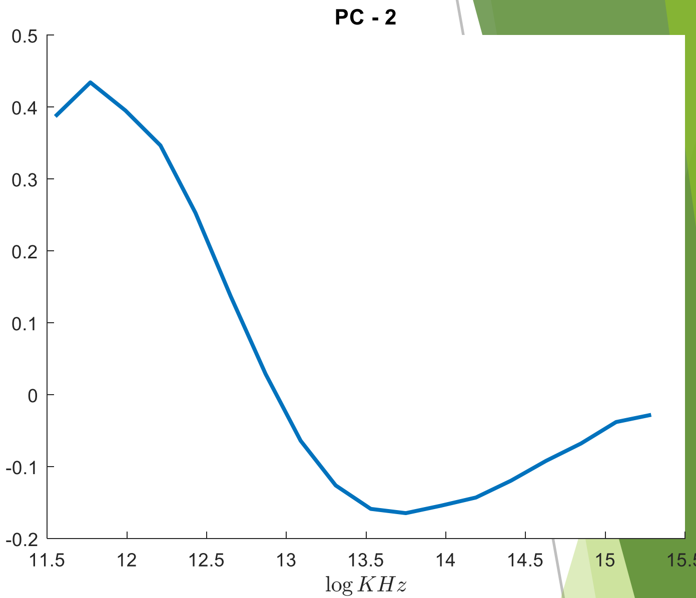
There is correlation

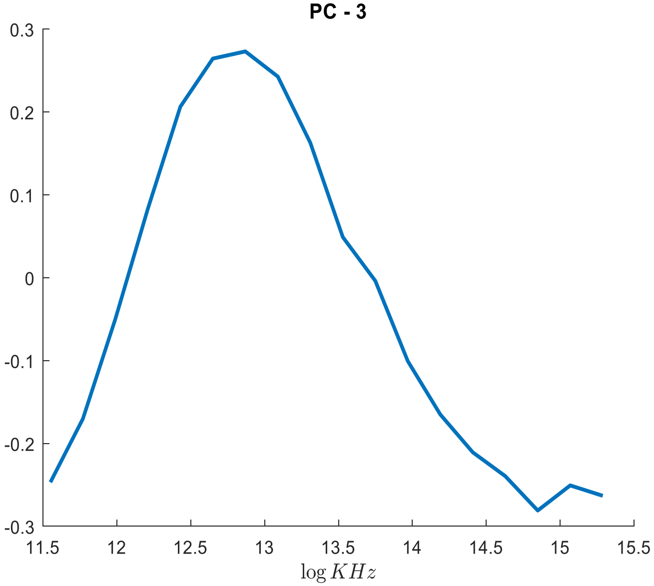
Principal component analysis (PCA) based model

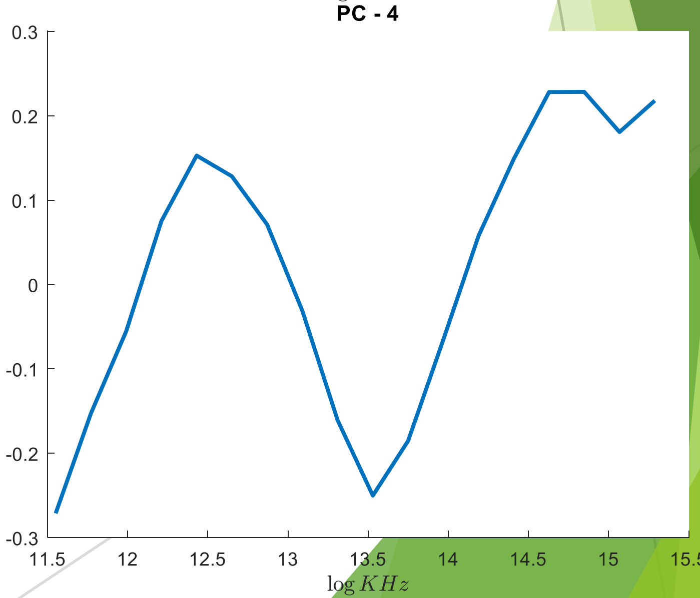
- PCA is a statistical method that does an orthogonal transformation.
- Convert a set of correlated variables into a set of linearly uncorrelated variables.
- The first PC has the largest possible variance.
- Each succeeding component has the highest variance under the constraint that it is orthogonal to the preceding components.

The shape of the PCs

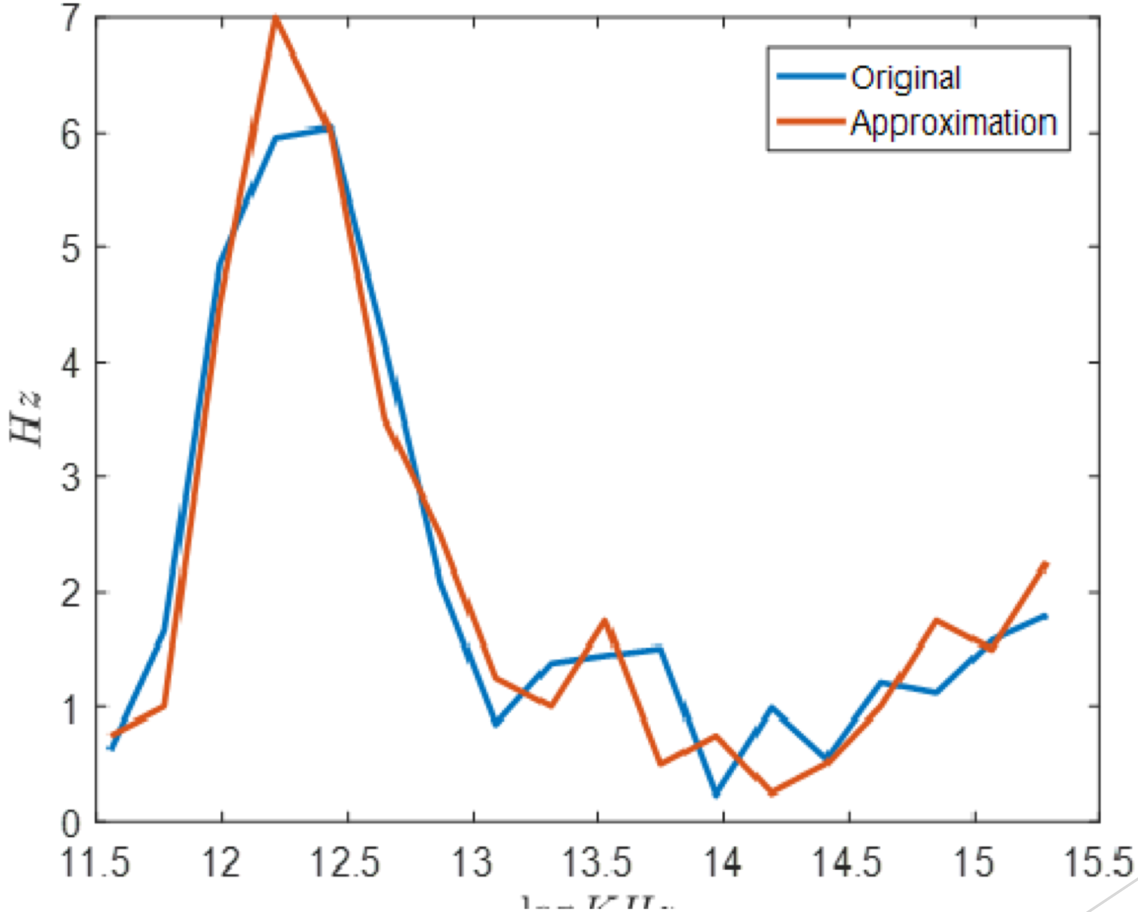








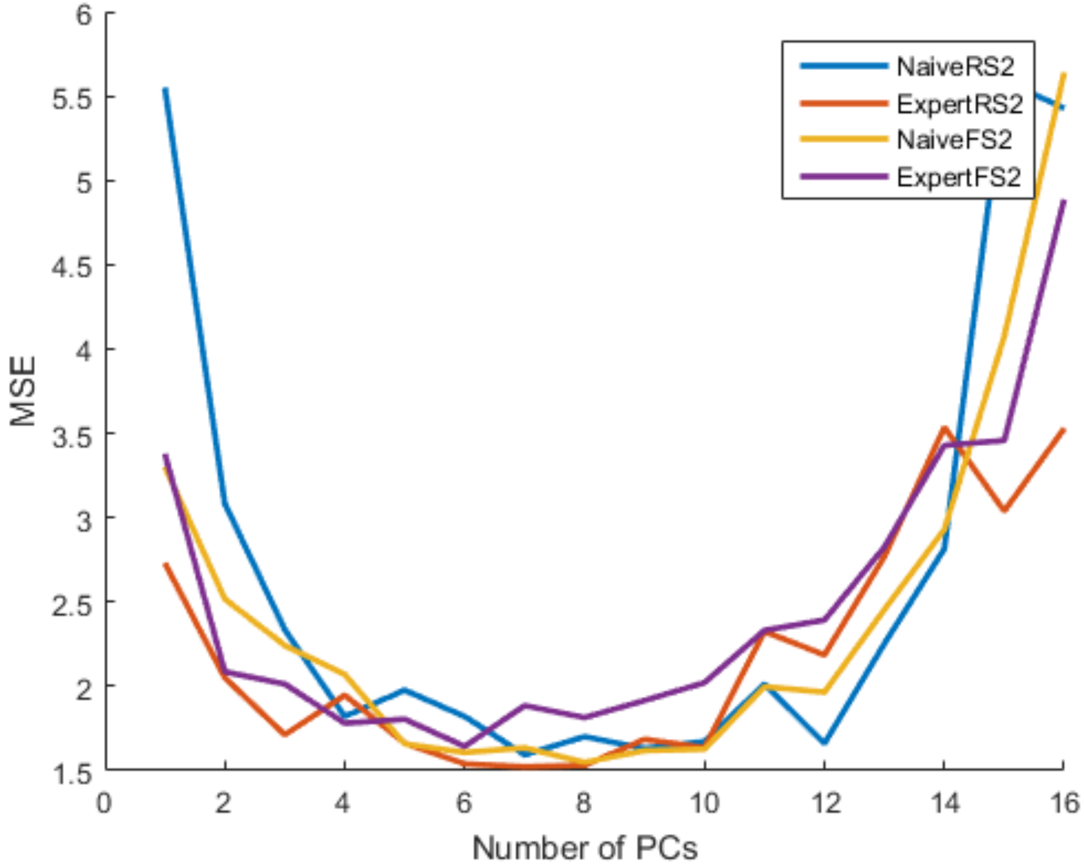
Reconstruction



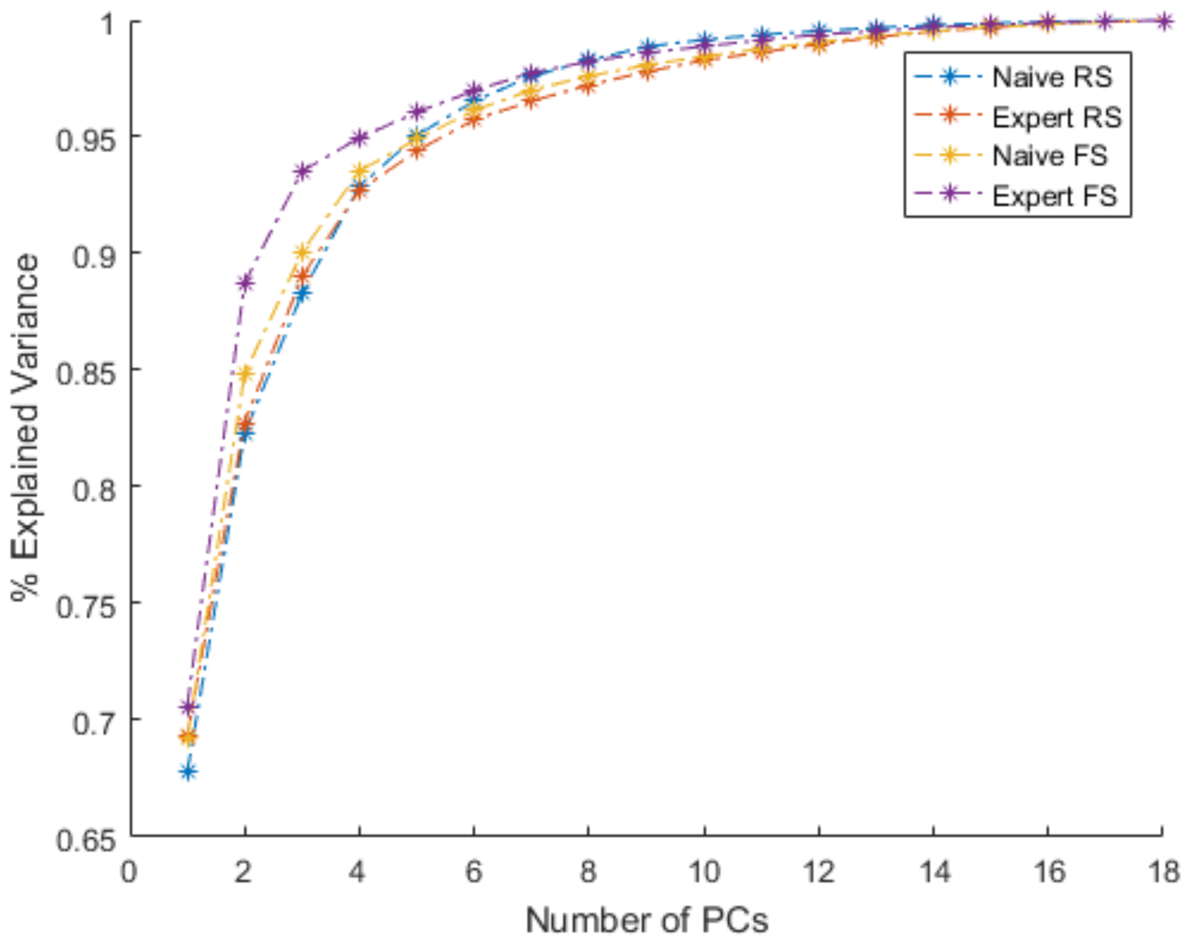
The optimal number of PCs

- How can we decide what is the optimal number of modes?
- High number of modes decrease the reconstruction error
- Low number of modes decrease the noise.
- Solution: Create the tuning curve based on part of the trails and check the MSE on the other part.

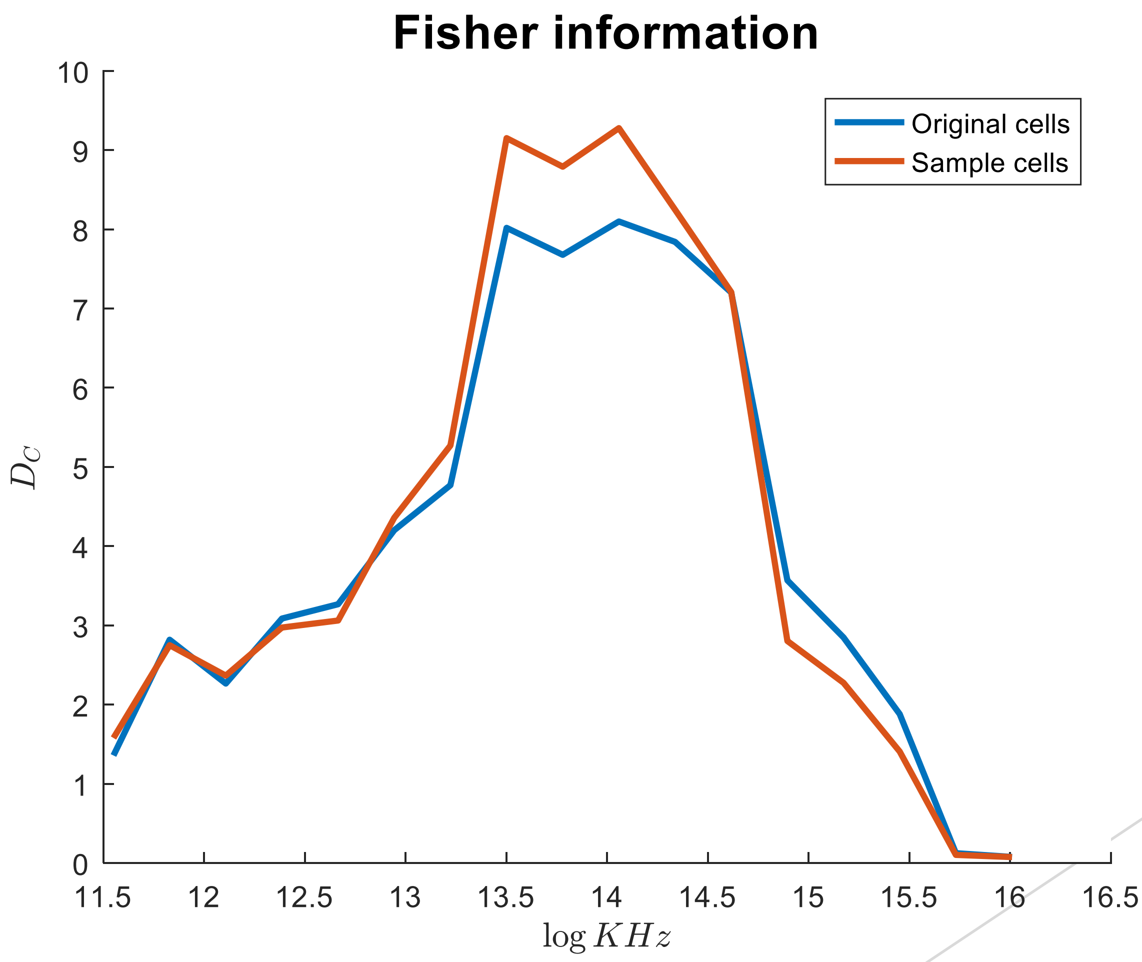
The optimal number of PCs



Explained variance



Sampling new cells



Summary

- PCA based model success to describe the

Changes due to perceptual learning

Neuronal coding efficiency

- Fisher Information
- Chernoff Distance
- Maximum likelihood discrimination with SVM
- Maximum likelihood estimation
- Optimal linear estimation
- Optimal linear discrimination

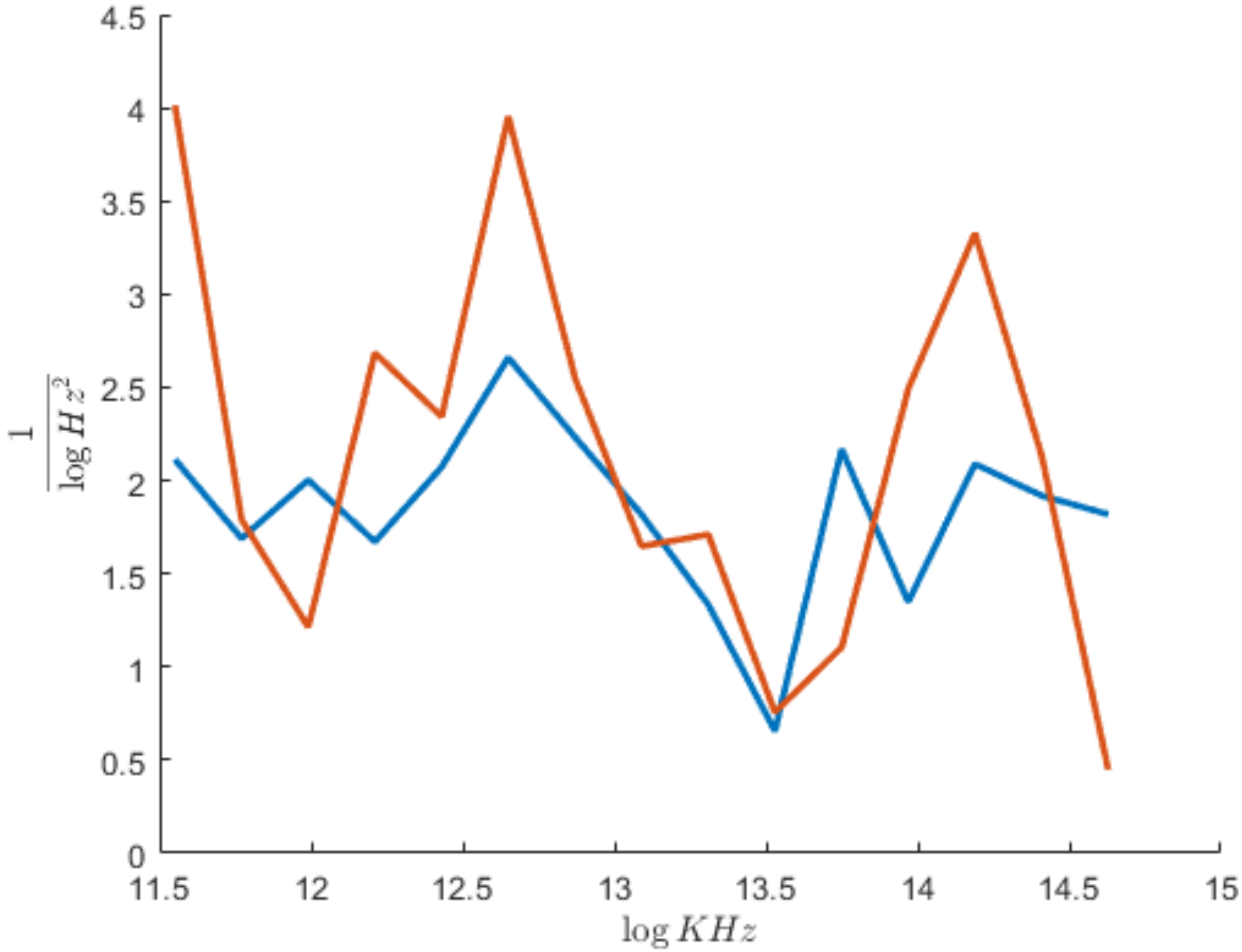
Fisher information

- Measure the amount of information that an observable random variable X carries about an unknown parameter θ.
- Gives the discrimination threshold that would be obtained by an optimal decoder.
- JND = threshold(θ) $\geq \frac{1}{\sqrt{J(\theta)}}$

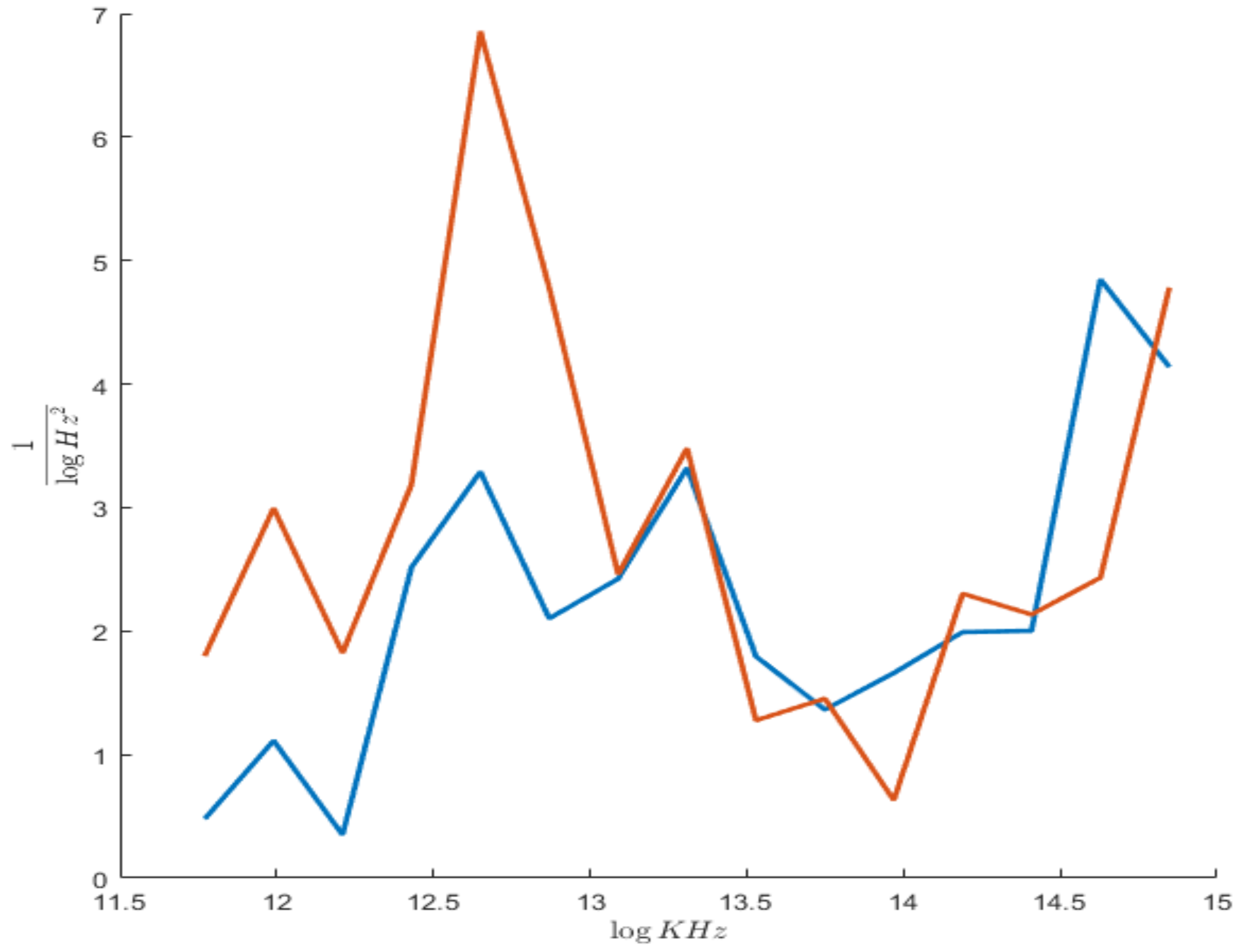
$$J(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log p\,(n|\theta)\right)^2 |\theta\right]$$

$$J_a(\theta) = \frac{f_a'(\theta)^2}{f_a(\theta)}$$

Fisher information- RS cells



Fisher information – FS cells



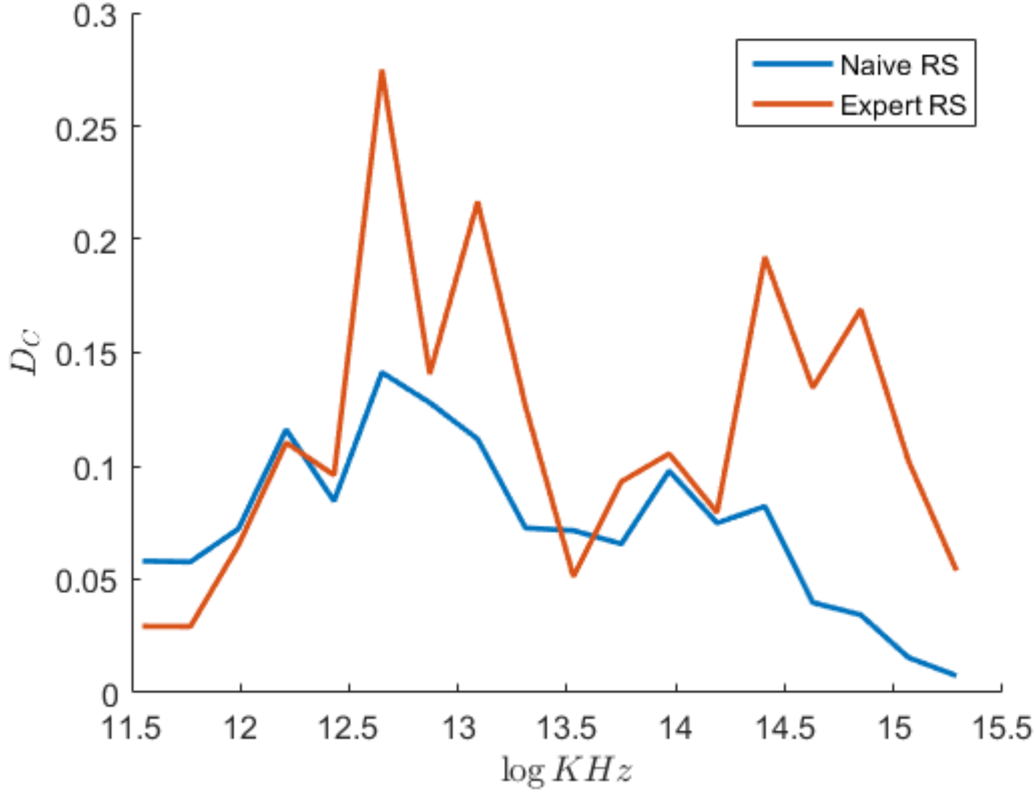
Chernoff distance

$$D_{\alpha}(f_1, f_2) = -\log Tr_{\vec{r}} P^{\alpha}(\vec{r}|f_1) P^{1-\alpha}(\vec{r}|f_2)$$

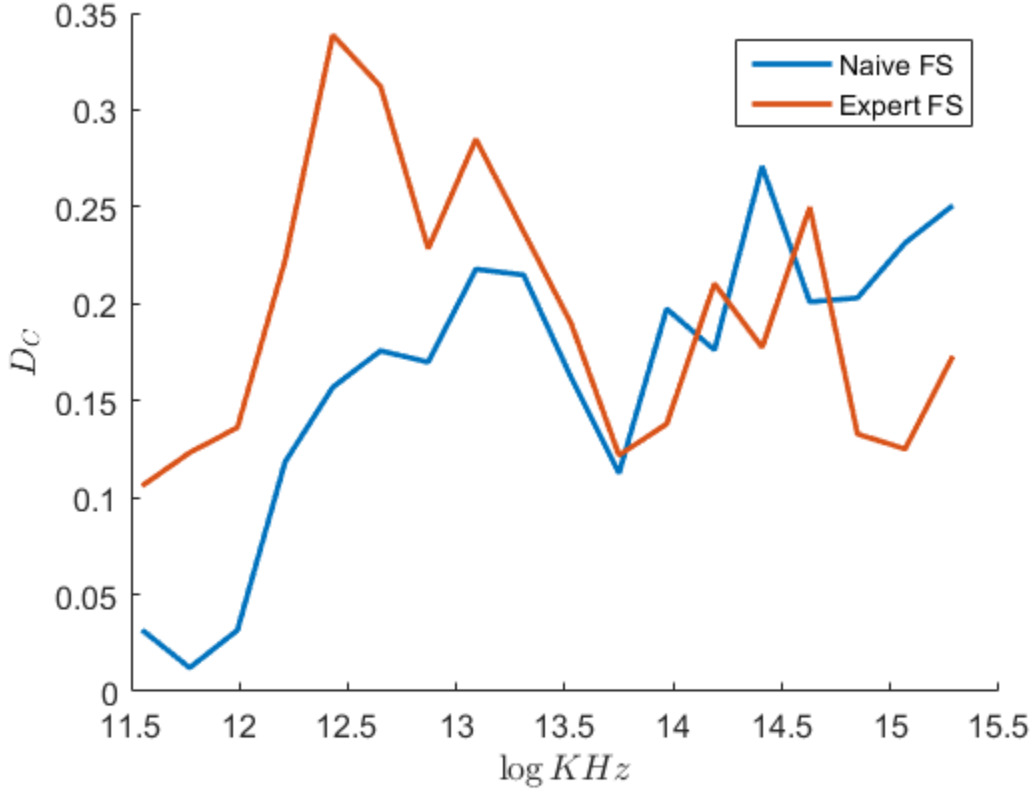
$$D_c(f_1, f_2) = \max_{\alpha} D_{\alpha}(f_1, f_2)$$

- f_1, f_2 are different frequencies and $\vec{r}$ is a vector of spike counts for a population of neurons.
- $P\ (\vec{r}|f_i)$ is the distribution of activity across the population $\vec{r}$ when the stimulus with the frequency f_i is presented.
- Relationships with Euclidean distance, error of maximum-likelihood discriminator, Fisher information and mutual information

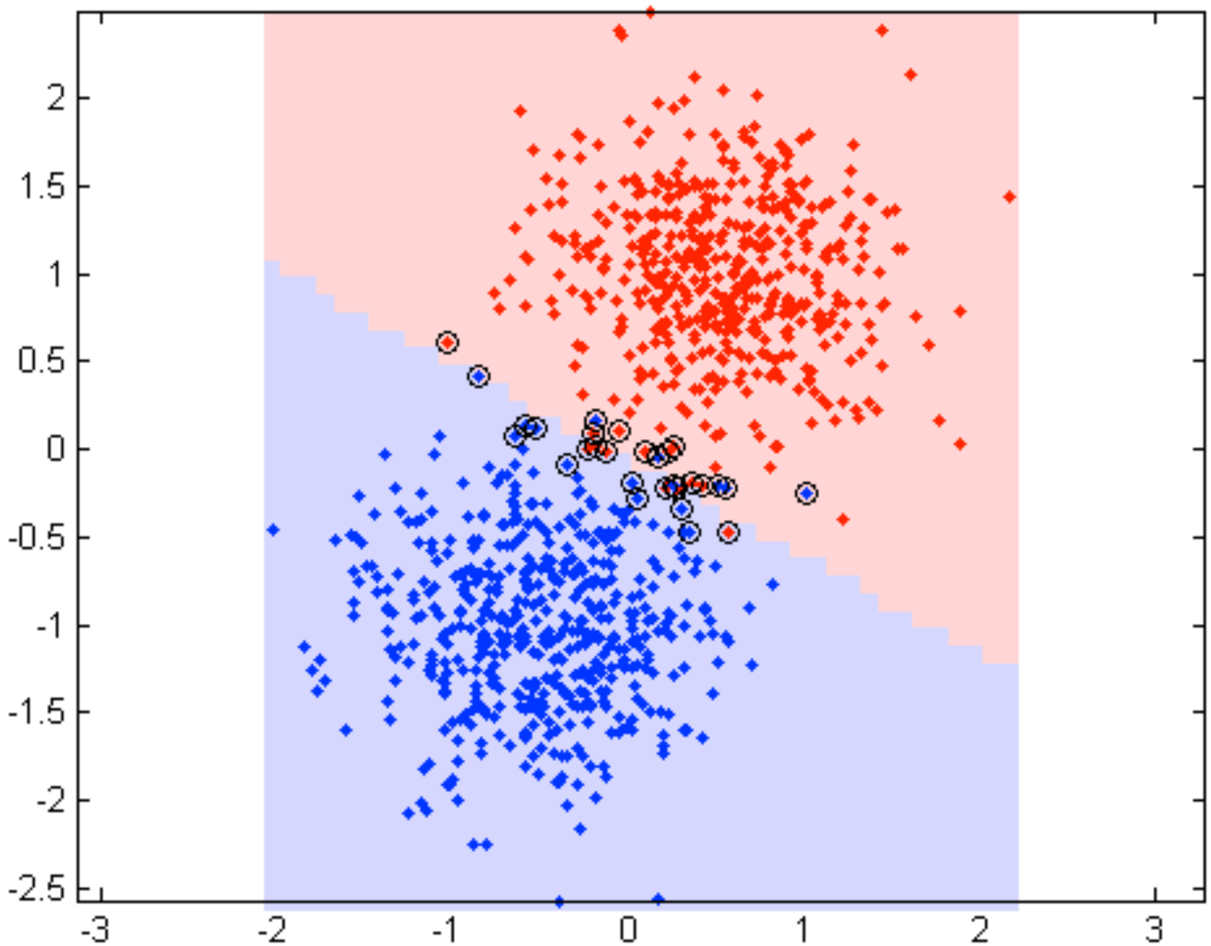
Chernoff distance – RS cells



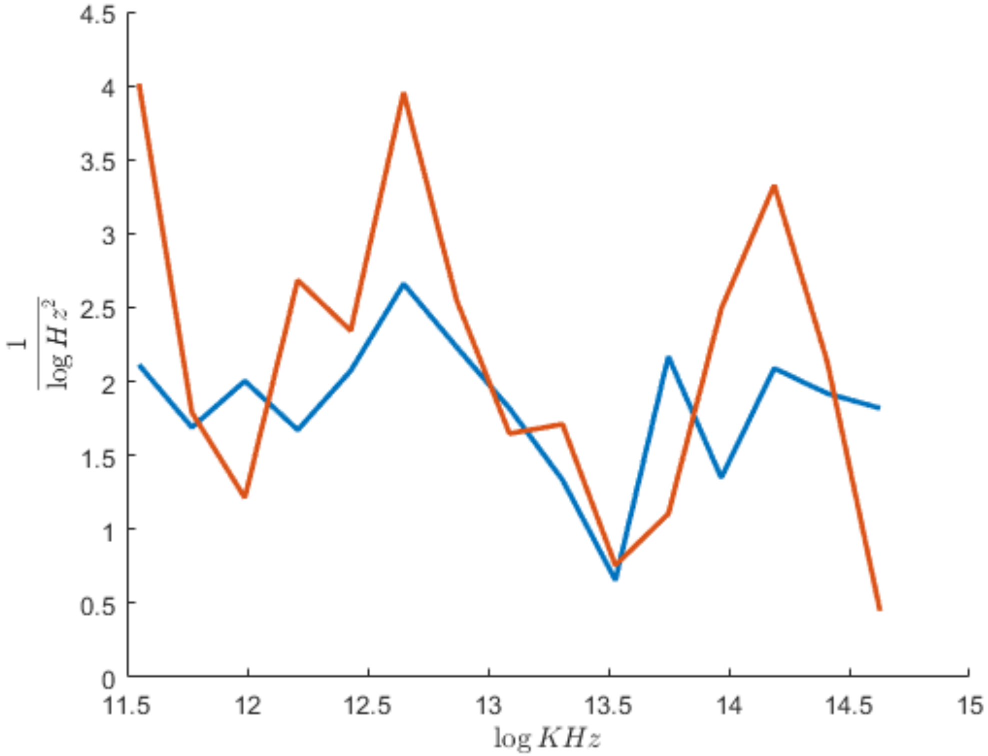
Chernoff distance – FS cells



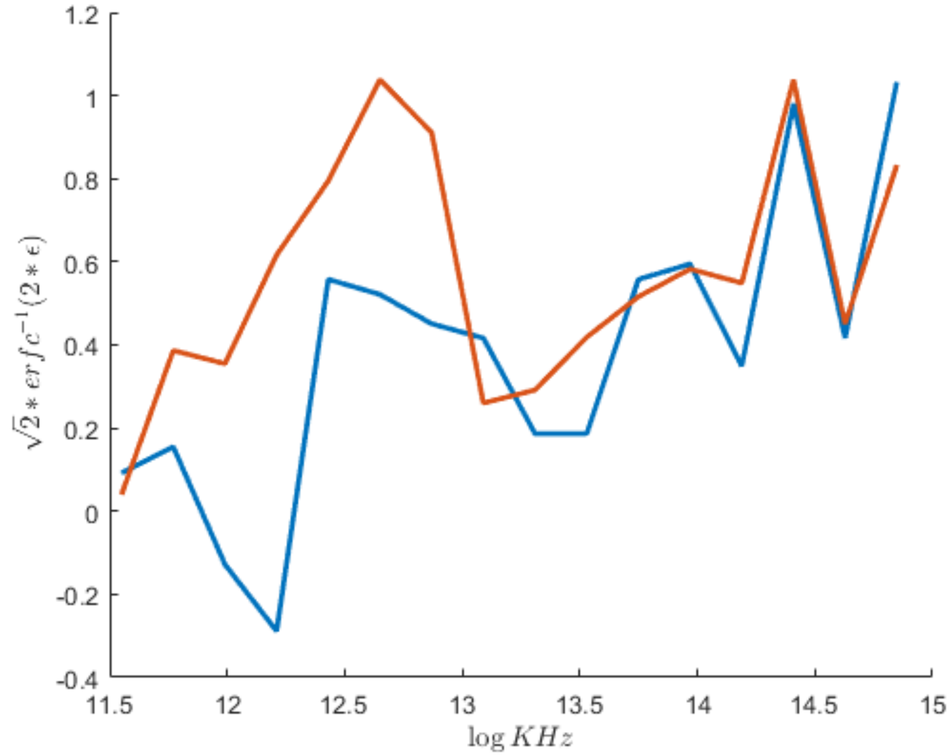
Maximum likelihood discrimination with SVM



ML discrimination with SVM – RS cells



Linear estimation - FS

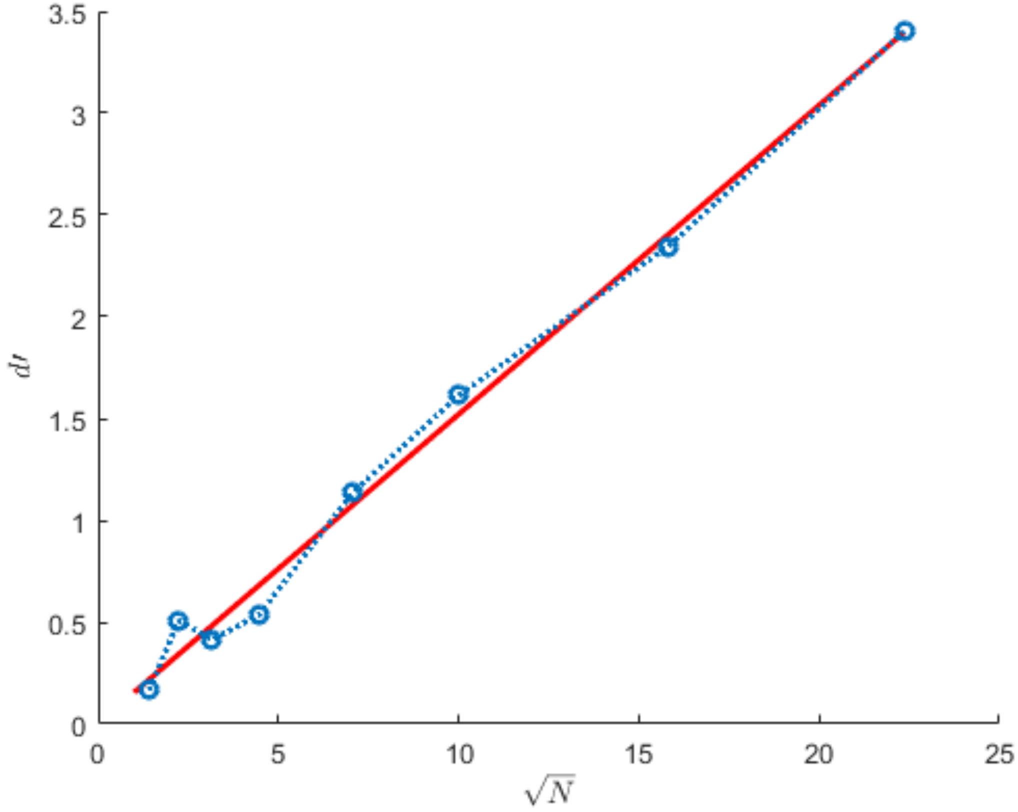


Estimation of the parameters

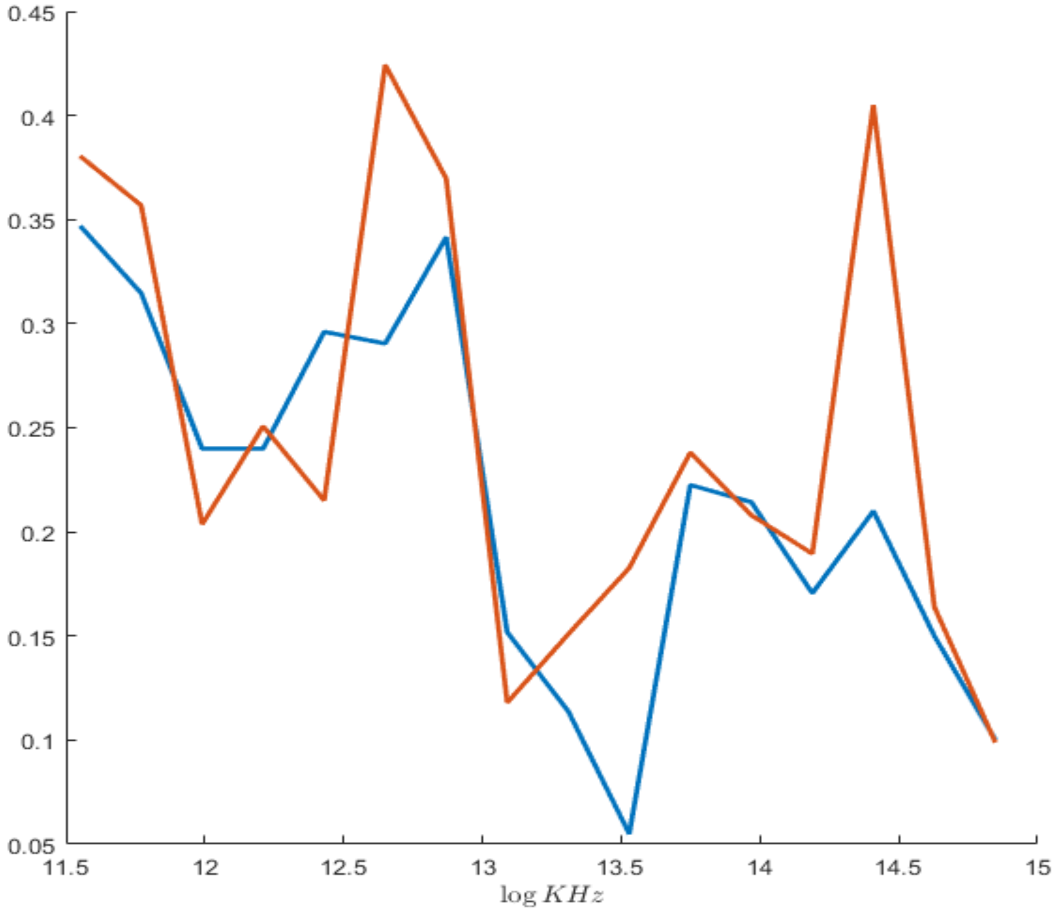
- For a large population of neurons the probabilities of error for Maximum Likelihood discrimination is $H\left(\frac{d'}{2}\right)$
- Where $H(x) = \left(\sqrt{2\pi}\right)\int_x^{\infty} e^{-\frac{x^2}{3}}$
- d’ is the discriminability of the two stimuli -

$$d' = |\delta\theta|\sqrt{J[r](\theta)}\sqrt{N}$$

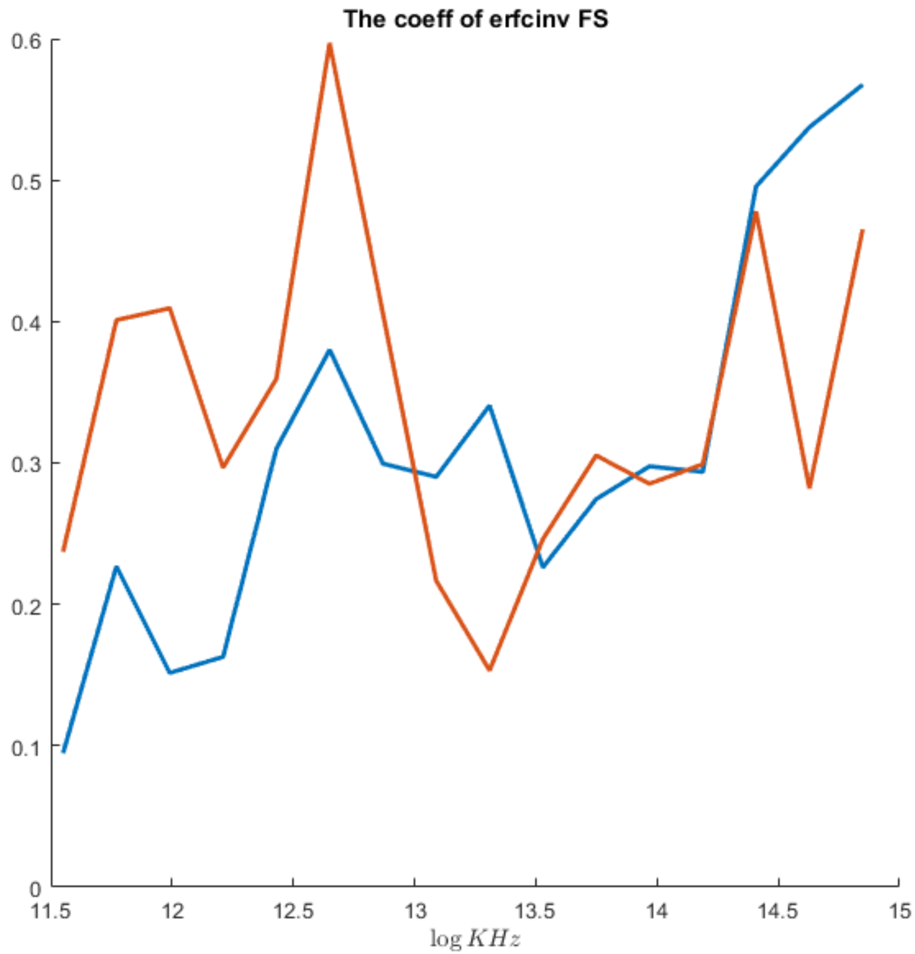
Estimate the parameters



Estimate the parameters - RS



Estimate the parameters - FS



Conclusions

Learning model of perceptual learning **with Fisher Information**

The model

- learning a discrimination task around the stimulus θ_{tr}
- training a readout from a population of tuned neurons

$$f_i(\theta_0) = h_i(\theta_0),\ i = 1, \ldots N$$

$$h_i(\theta_0) = \sum_{j=1}^{N_0} W_{ij} h_j^0(\theta_0)$$

- Assume the noise is Gaussian with stimulus independent σ

The goal

- Optimize the fisher information in the training stimulus

$$\sigma^2 I(\Delta W|\theta_{tr}) = \sum_i (\partial_{\theta_{tr}} f_i)^2 = \sum_i \left(\sum_j (W_{ij} + \Delta W_{ij}) \partial_{\theta_{tr}} h_j^0 \right)^2$$

-
- The cost function will

$$E(\Delta W) = -\sigma^2 I(\Delta W|\theta_{tr}) + \frac{1}{2}\lambda \Delta W^T \Delta W, \ \lambda > 0$$

Deep Network

- For each layer -

$$f_i^l(\theta_0) = f(h_i^l(\theta_0)), i = 1, ...N - 1$$

$$h_i^l(\theta_0) = \sum_{j=1}^{N_0} W_{ij}^l f(h_j^{l-1}(\theta_0))$$

Deep Network – general case

$$\partial_{\Delta W^l_{ij}} E = \sum_{i_0} \left(\partial_{\theta_{tr}} f^L_{i_0}\right) \partial_{\theta_{tr}} \left(\delta^l_{i,i_0} f^{l-1}_j\right) - \lambda \Delta W^l_{ij}$$

$$f^L_{i_0} = g\left(\sum_j (W^L_{i_0 j} + \Delta W^L_{i_0 j}) f^{L-1}_j\right)$$

$$\delta^l_{i,i_0} = \partial_{f^l_i} f^L_{i_0} g^l_i$$

$$\delta^l_{i,i_0} = \sum_k g^l_i (W^{l+1}_{ki} + \Delta W^{l+1}_{ki}) \delta^{l+1}_{k,i_0}$$

$$\delta^L_{i,i_0} = \delta_{i,i_0} g^L_i$$

$$g^l_i = \partial_h f(h^l_i)$$

Linear Deep Network

$$\partial_{\Delta W_{ij}^{L}} E = \left(\partial_{\theta_{tr}} h_i^L\right) \partial_{\theta_{tr}} \left(h_j^{L-1}\right) - \lambda \Delta W_{ij}^L = 0$$

$$h_i^L = \sum_j (W_{ij}^L + \Delta W_{ij}^L) h_j^{L-1}$$

$$\sum_{j'} (W_{ij'}^L + \Delta W_{ij'}^L) \partial_{\theta_{tr}} \left(h_{j'}^{L-1}\right) \partial_{\theta_{tr}} (h_j^{L-1}) = \lambda \Delta W_{ij}^L$$

Linear Deep Network

$$\Delta W^L = \tilde{W}^L H^{L-1} H^{(L-1)T}$$

$$\tilde{W}^L = (\lambda - ||H^{L-1}||^2)^{-1} W^L$$

$$H_j^{L-1} = \partial_{\theta_{tr}} h_j^{L-1}$$

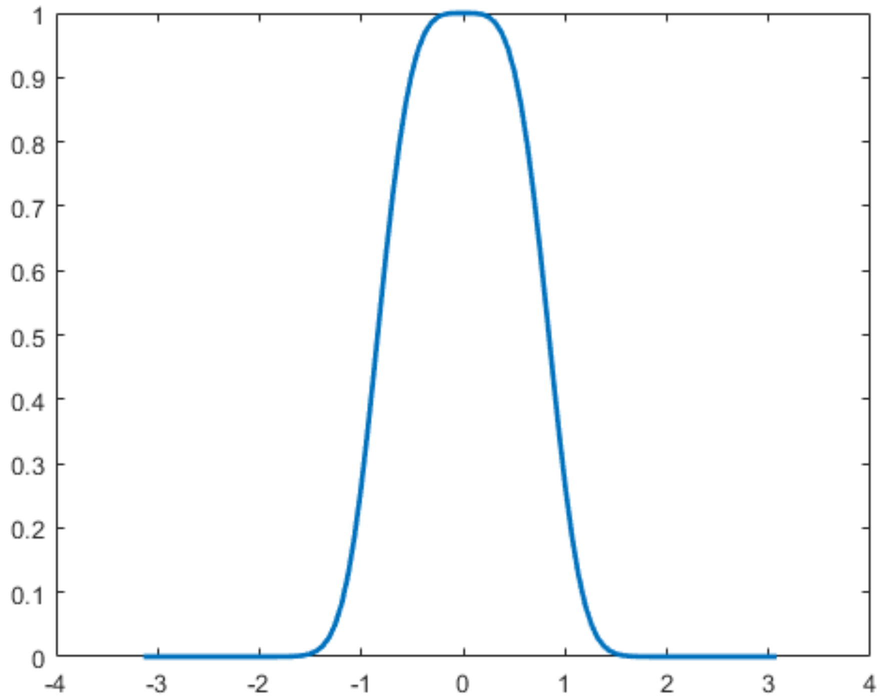
$$\lambda \Delta W_{ij}^l = S_i^l \partial_{\theta_r} h_j^{l-1}, l < L$$

$$S_i^l = \sum_k (W_{ki}^{l+1} + \Delta W_{ki}^{l+1}) S_k^{l+1}$$

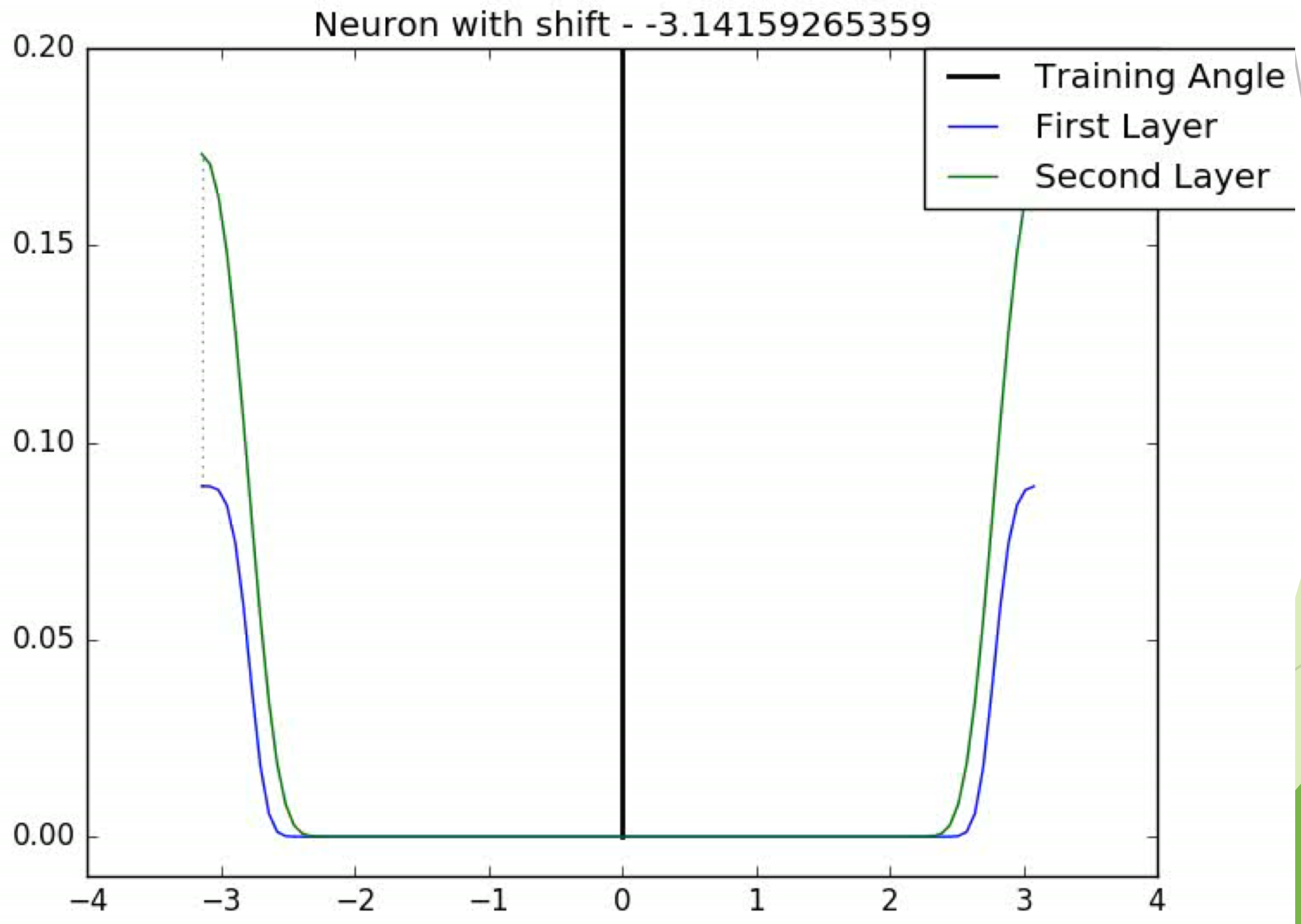
$$S_i^L = \partial_{\theta_r} h_i^L = H_i^L$$

Simulations

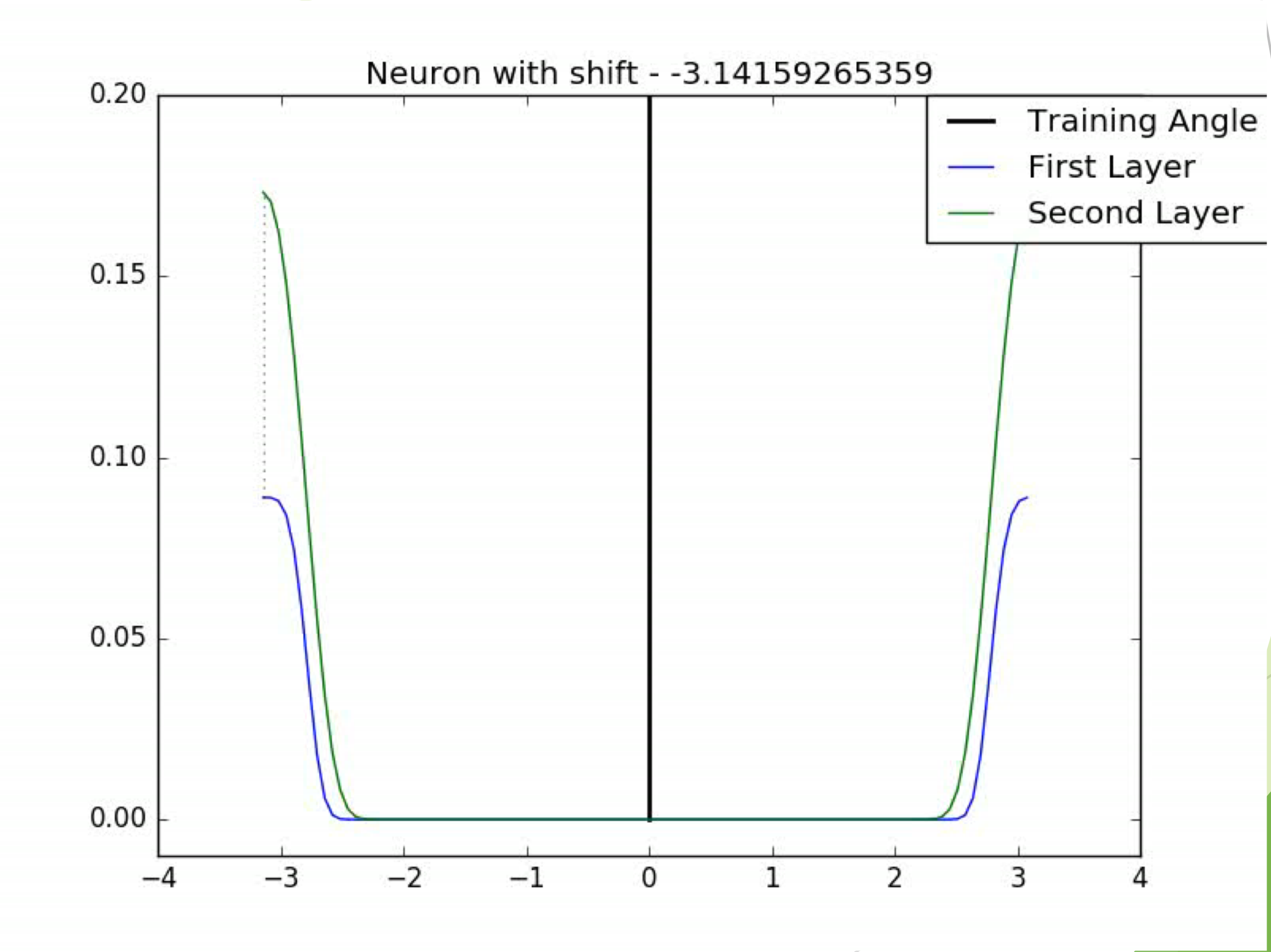
- The tuning curve - $\exp(-\frac{(1-\cos(x))^2}{a^2})$-



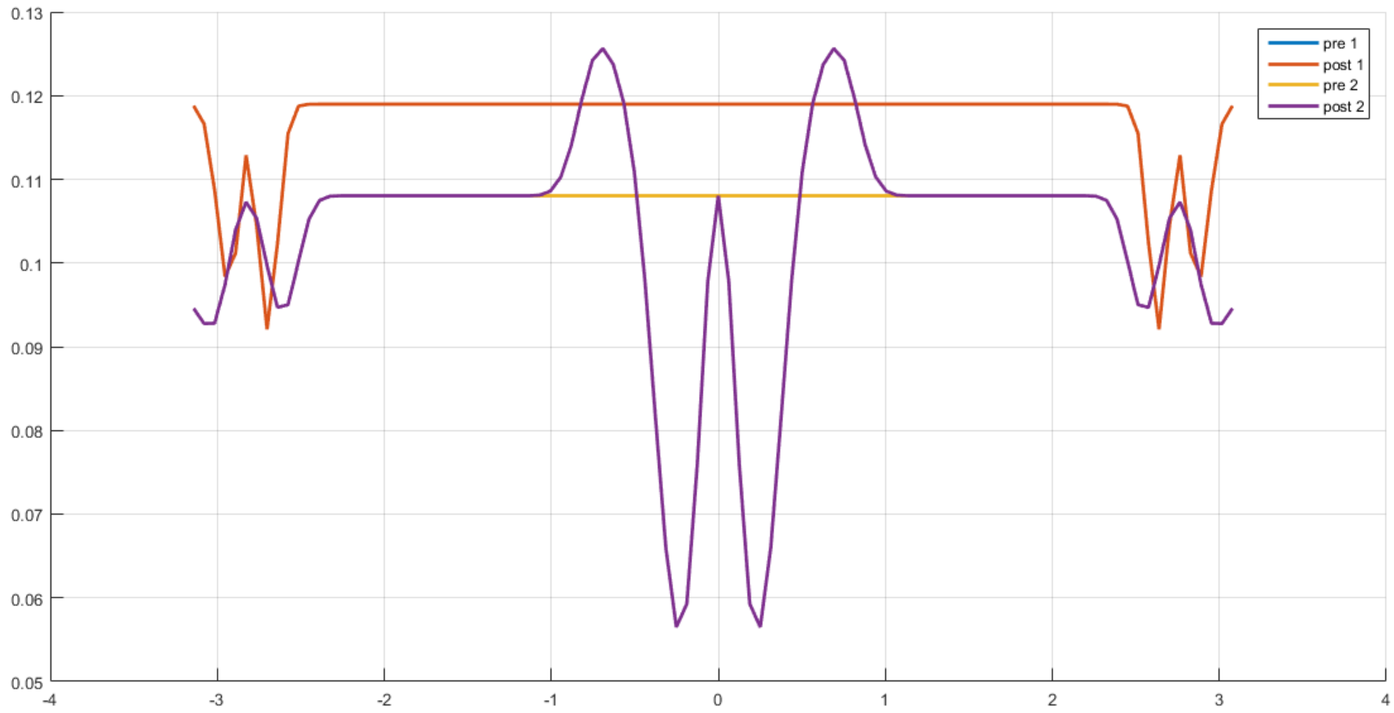
The changes due to the learning



The changes due to the learning



Fisher Information



Conclusions

Thanks

Prof. Adi Mizrahi

Ido Maor

Prof. Haim sompolinsky

Questions?

Thank you