Class 10

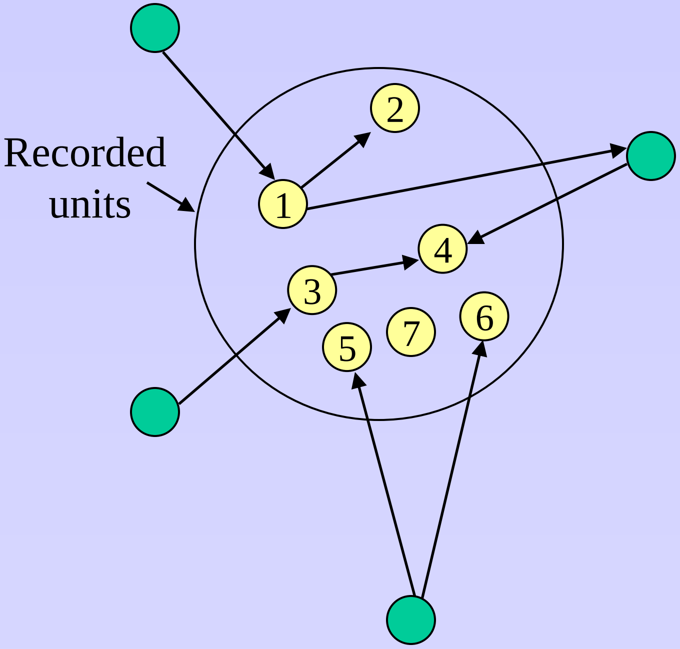
Dimension Reduction

PCA-ICA

Population code

- Population code: how are units 'functionally' grouped?
- Units working together are 'correlated' ←→ during a set of experiments (e.g. multiple stimulations) units are not functionally independent.

<u>Why are two neurons correlated?</u>



- Because of *interconnections* within the population recorded (1,2)
- Because of *interconnections* through some un-recorded neuron (1,4)
- Because of *common inputs* (5,6)

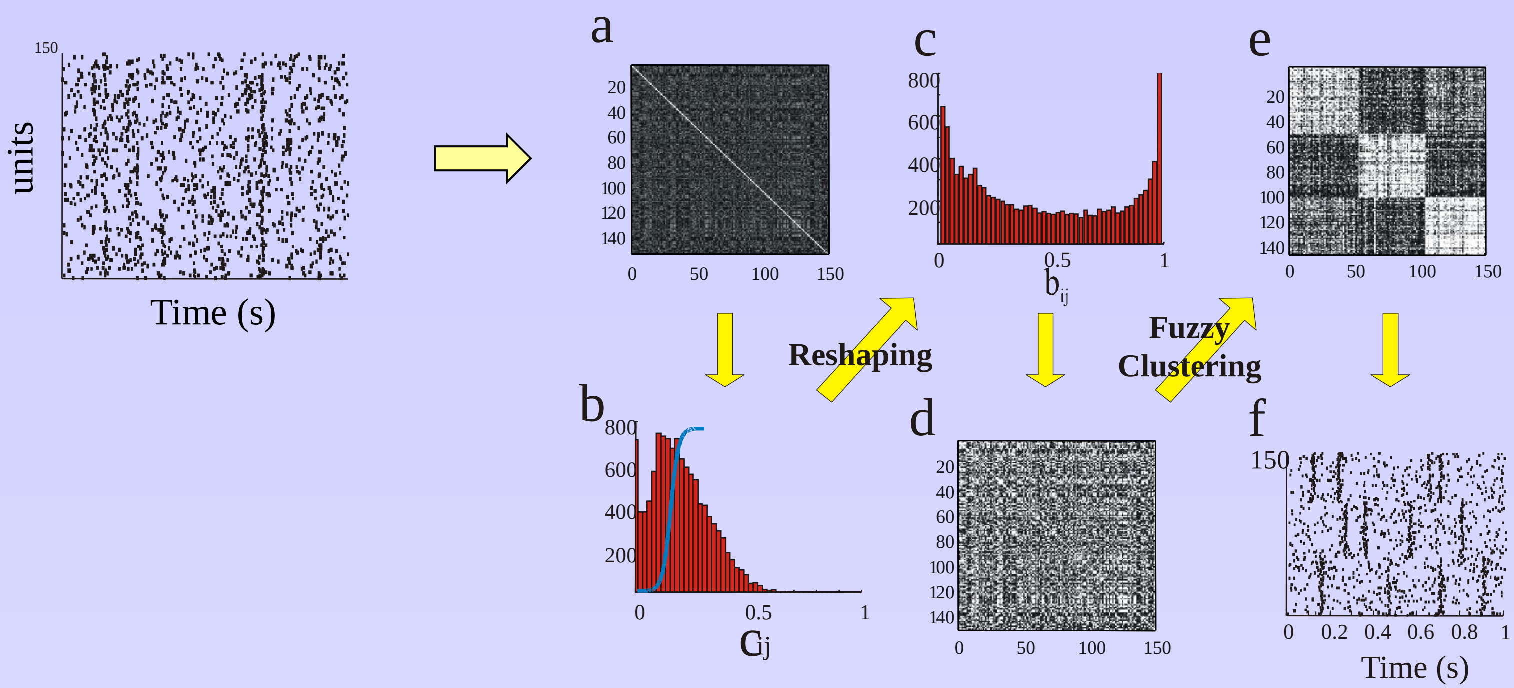
- **Goal**: Find groups of neurons such that the correlation within a group is high, and such that the groups are as uncorrelated as possible.

(1,2,4) (3,4) (5,6)

Population

- **Goal**: Find groups of neurons such that the correlation within a group is high, and such that the groups are as uncorrelated as possible.

● **Method 1**: Use Clustering



➔ Precise temporal information, short time scales clusters, pair-wise correlations

Principal Component Analysis

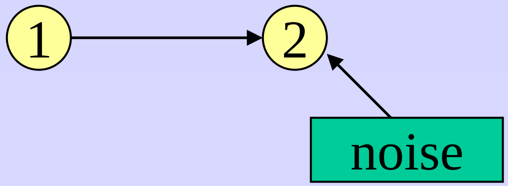
- **Goal**: Find groups of neurons such that the correlation within a group is high, and such that the groups are *as uncorrelated as possible*.

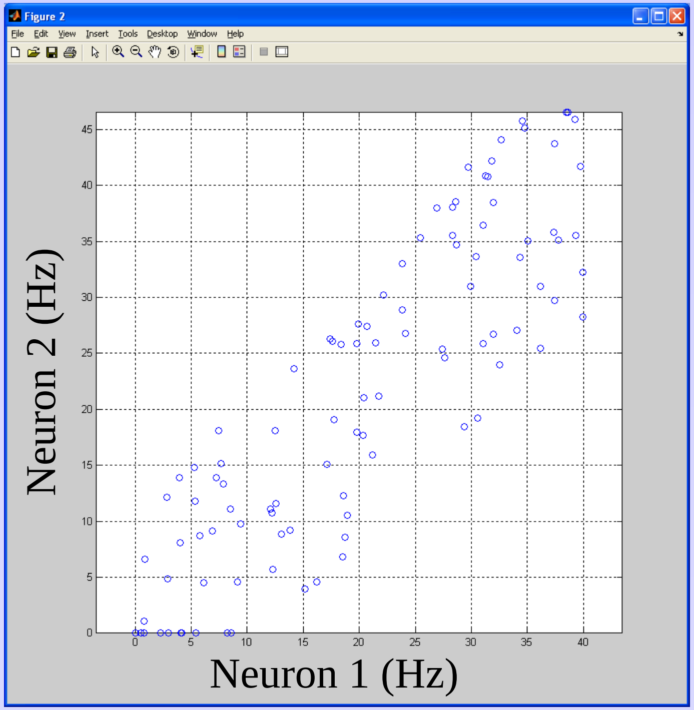
$\Leftrightarrow$

Find linear combinations of neurons outputs that give a *maximal variance*.

- **Method 2**: Use Principal Component Analysis

- For each experimental condition, each neuron is characterized by one number: Firing rate.





Principal Components Analysis

Collect data: N neurons, E experiments

- Step 1: Form a data structure

$F_{\text{experiment, neuron}}$

	Neuron 1	Neuron 2	Neuron 3	Neuron N
Experiment 1	F11	F12	F13	F1N
Experiment 2	F21	F22	F23	F2N
.	F31	F32	F33	F3N
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
Experiment E	FE1	FE2	FE3	FEN

= F

Principal Components Analysis

- Step 2: Compute the correlation matrix C of F. *Corrcoef()*

$$\mu_i = \langle F_{k,i} \rangle$$

(average across experiments)

$$\mathrm{cov}(n_i, n_j) = \langle (F_{ki} - \mu_i)(F_{kj} - \mu_j) \rangle$$

$$c_{i,j} = cor(n_i, n_j) = \frac{\mathrm{cov}(n_i, n_j)}{\sigma_i \sigma_j}$$

$$C = \begin{pmatrix} 1 & 0.99 \\ 0.99 & 1 \end{pmatrix}$$

- Step 3: Diagonalize C. *eig()*

$$C.V = V.D$$

$$C = V.D.V^{-1}$$

eigenvectors

$$V = \begin{pmatrix} -0.707 & 0.707 \\ 0.707 & 0.707 \end{pmatrix}$$

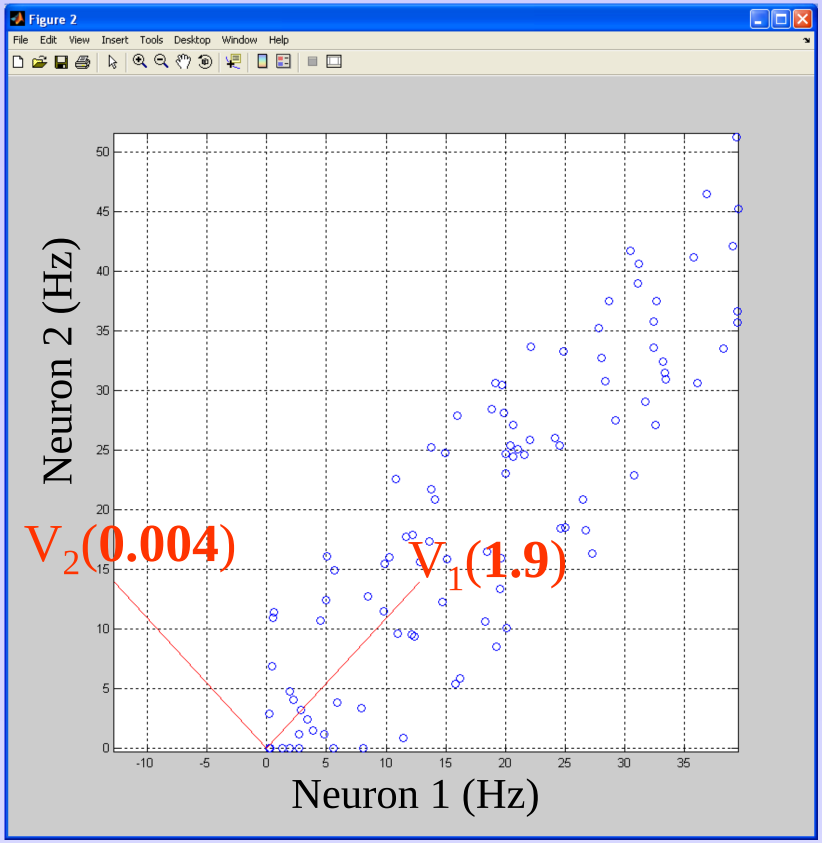
V_1 V_2

eigenvalues

$$D = \begin{pmatrix} 0.004 & 0 \\ 0 & 1.99 \end{pmatrix}$$

Principal Components Analysis

- Eigenvalues and Eigenvectors = Principal components



Explained variance (%)

$$EV_i = \frac{Eigenvalue_i}{\sum_j Eigenvalue_j}$$

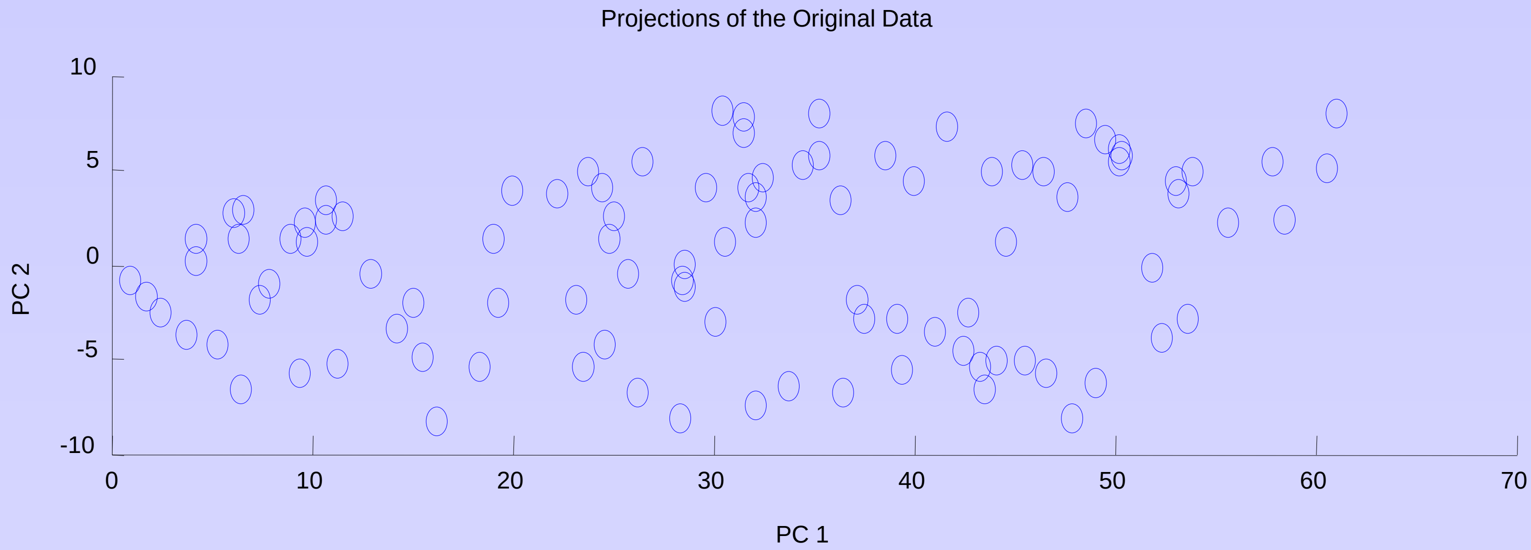
EV_1=99.9%

- **Kaiser Criterion**: Any eigenvector associated with an eigenvalue less then 1 contains less variance than the original dimensions.

→ Only keep eigenvectors with eigenvalues > 1

Principal Components Analysis

- Step 4: Sort the PC by their eigenvalue. Project to PC space (keep only the most significant components)

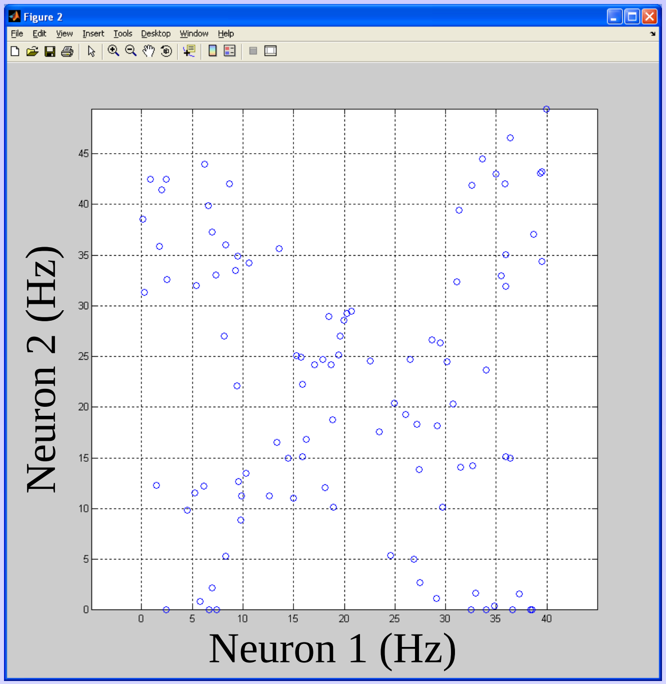


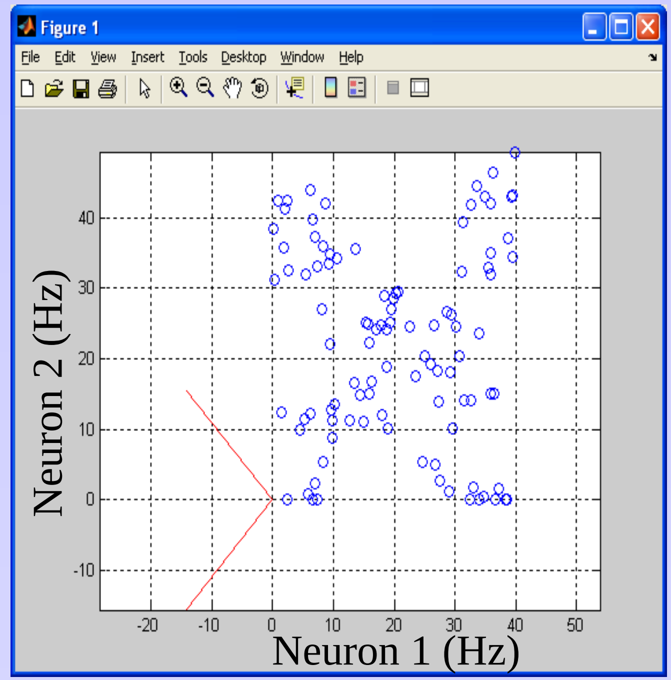
- Step 5: ‘Analyze’ the coordinate of the significant PCs:

  → do they ‘mean’ anything?

PCA: More examples

- Example: 2 or more significant PCs



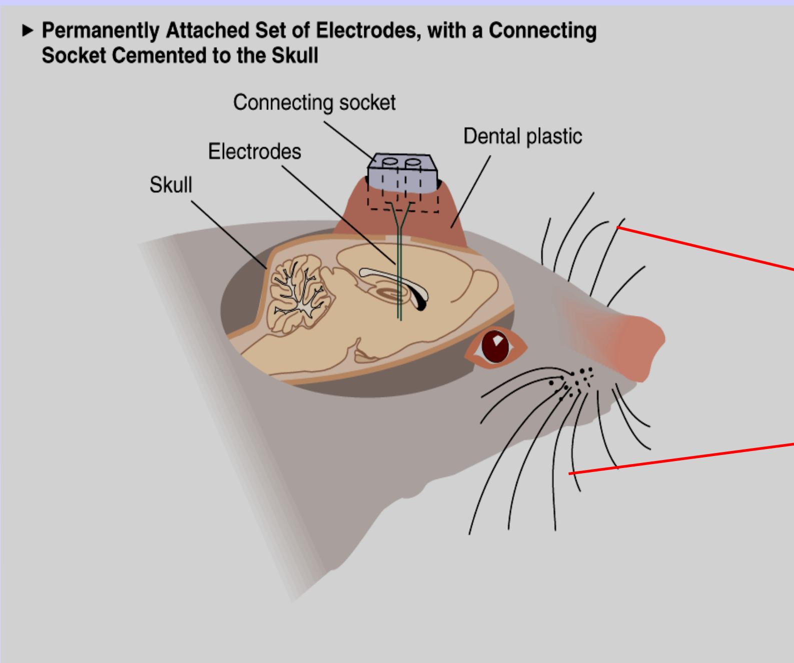


$$D = \begin{pmatrix} 0.97 & 0 \\ 0 & 1.02 \end{pmatrix}$$

$$V = \begin{pmatrix} -0.707 & -0.707 \\ -0.707 & 0.707 \end{pmatrix}$$

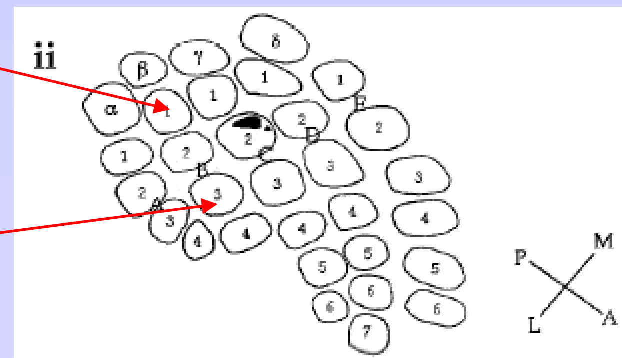
Principal Components Analysis

- In vivo behaving- recording from barrel cortex. 23 VPM neurons, multiple experimental conditions.



?

VentralPosteriorMedial
Nucleus Thalamus

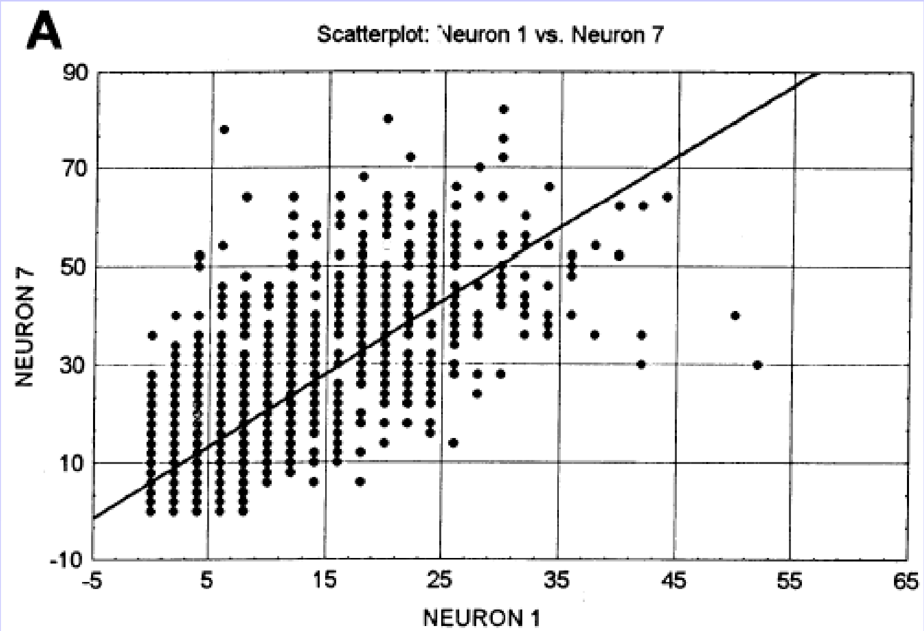


(Chapin, Nicolelis 1999)

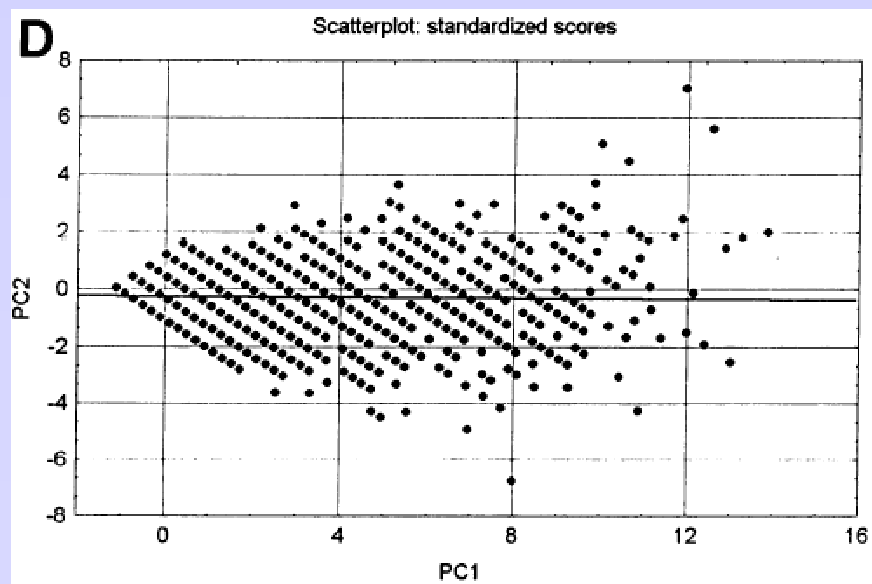
Barrel cortex

Questions: Are the VPM neural population response recorded co-related? If so, to what?

Principal Components Analysis

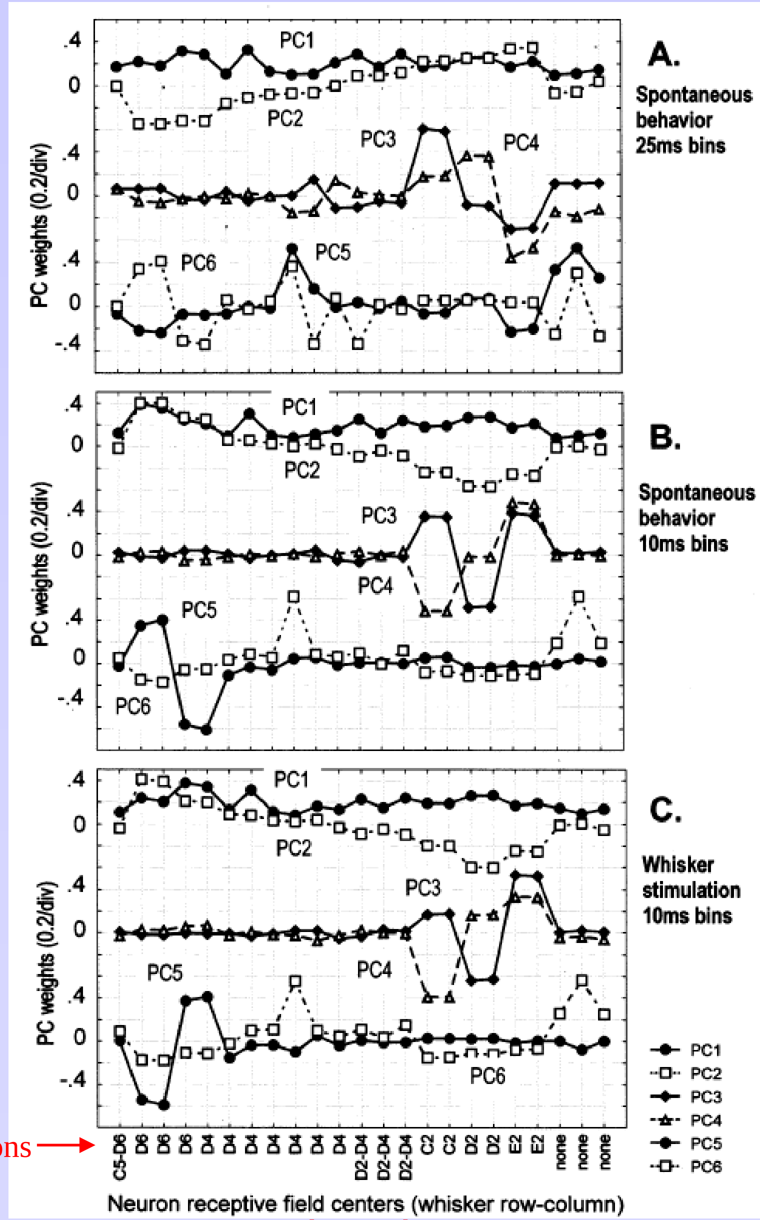


Simple case: 2 neurons



(Chapin, Nicolelis 1999)

Principal Components Analysis

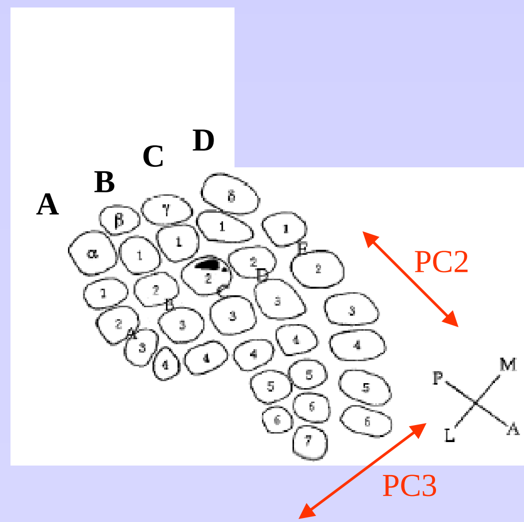


23 neurons →

Rostral ←------→ Caudal

- All 23 neurons. Interpreting the PCs…

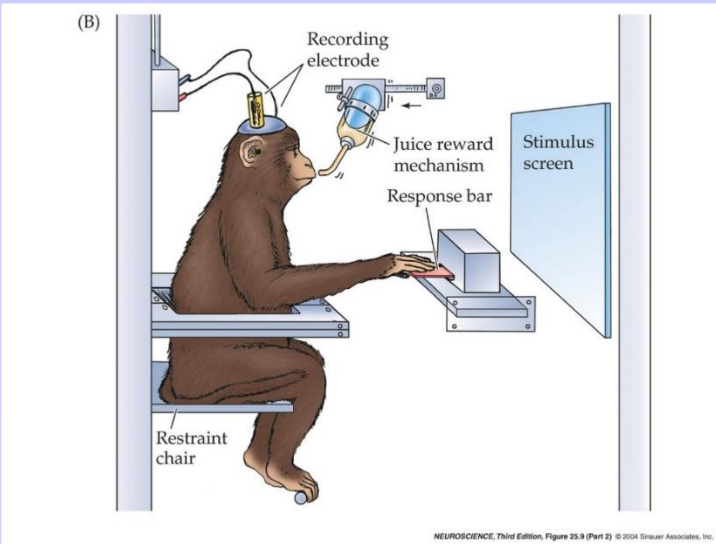
- PC1: non topographical. Global activity of all neurons.
- PC2: rostrocaudal gradient
- PC3: dorso-lateral gradient
- PC4: D2 Vs E2 neurons
- …



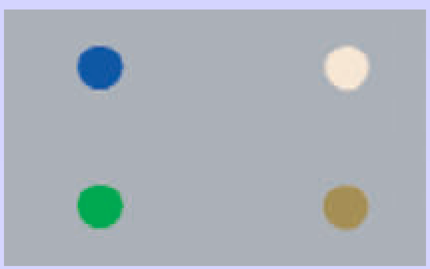
(Chapin, Nicolelis 1999)

Principal Components Analysis

- **I**nfero-**T**emporal cortex – face cells



(Matsumoto et al, 2005)

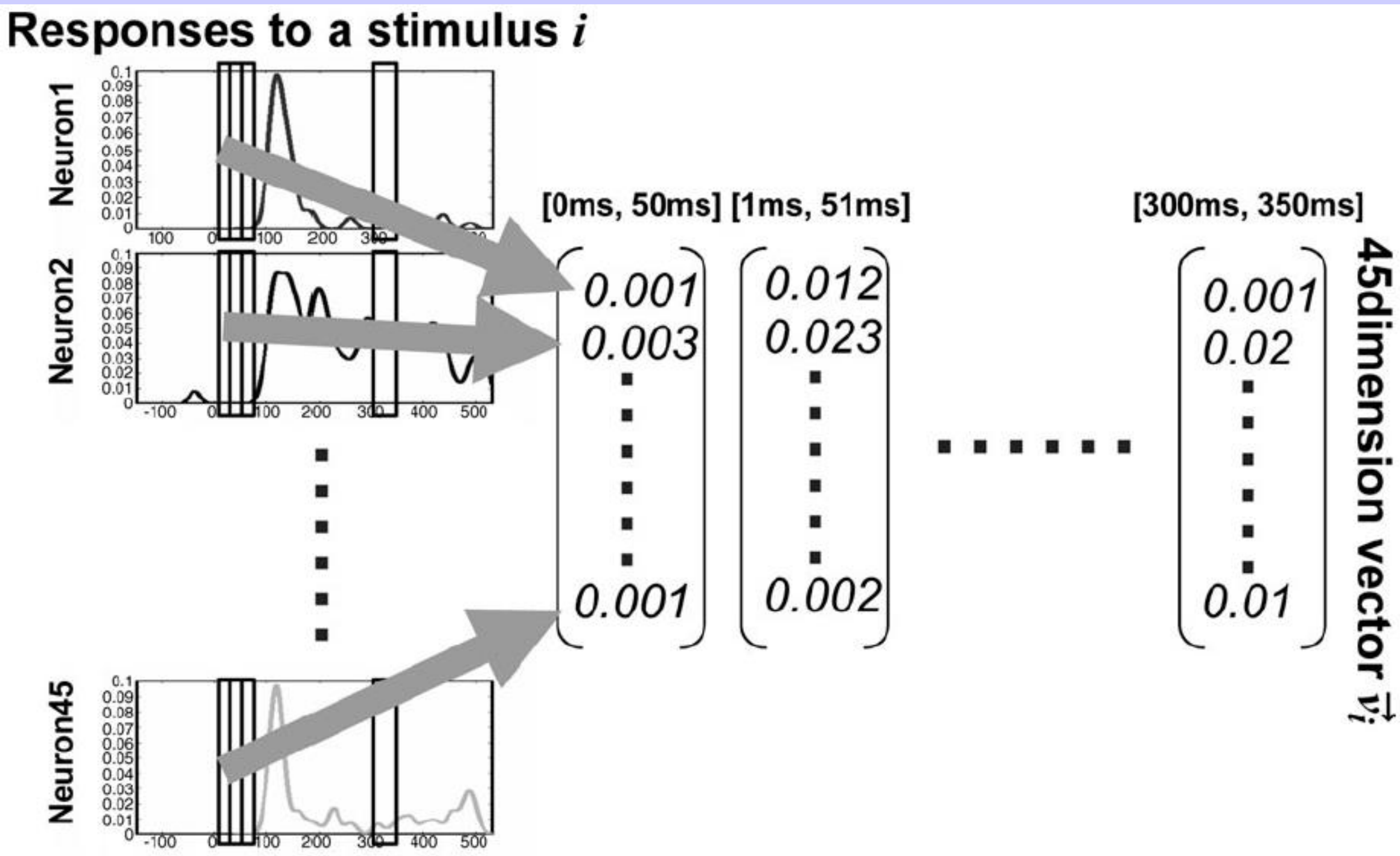


Large Stimuli variations: 38 stimuli. Identity, expression, human/monkey/objects

Questions: Are population responses indicative of specific aspects of the stimuli? Do they selectivity vary in time?

Principal Components Analysis

- For each stimulus: 45 neurons, 50 ms overlapping time windows



→ 45 x 300 Matrix

(Matsumoto et al, 2005)

Note: large overlap between dimensions → redundancy → dimension reduction

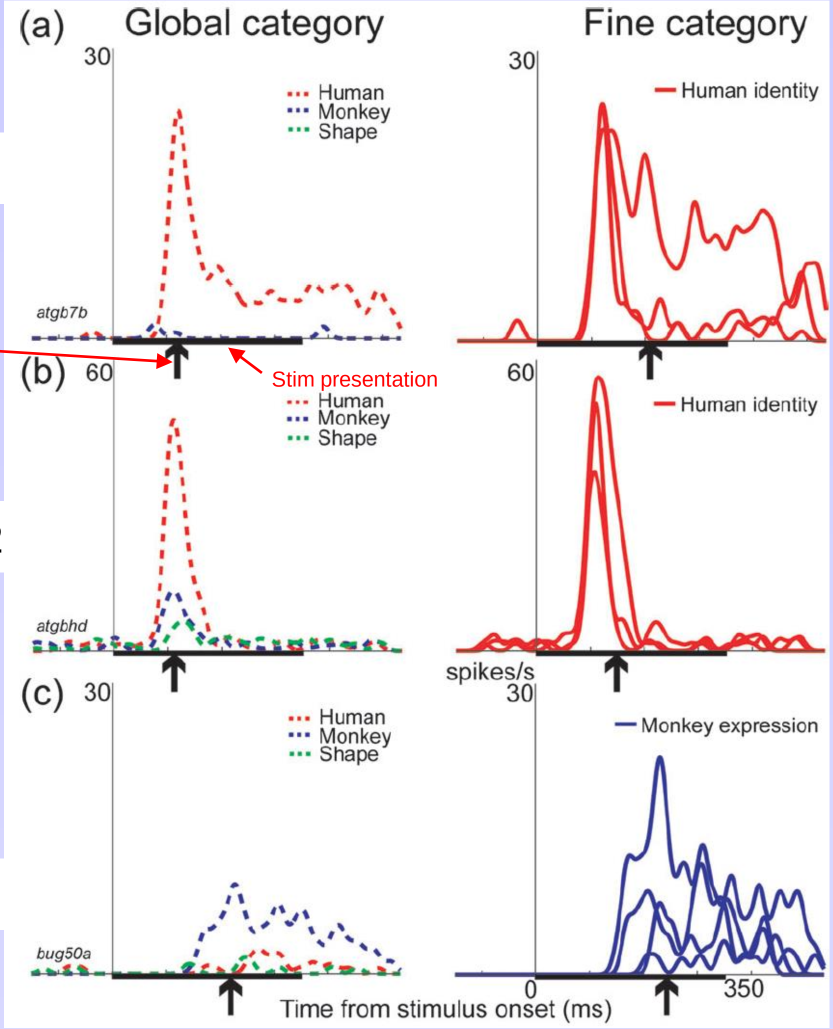
Principal Components Analysis

- IT neurons carry different types of information. 3 examples.

Neuron 1

3 Different people

Max information across curves
(i.e. 'discriminability')



Neuron 2

3 Different people

Neuron 3

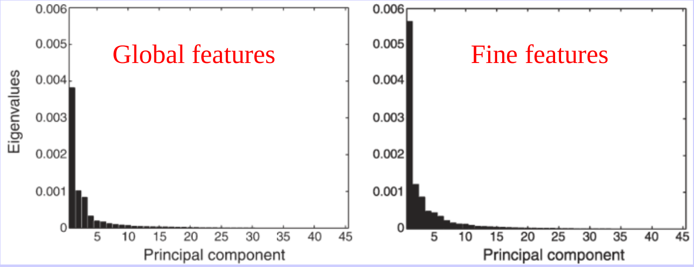
4 Different expressions

(Matsumoto et al, 2005)

Note: max discriminability occurs at different times for different features

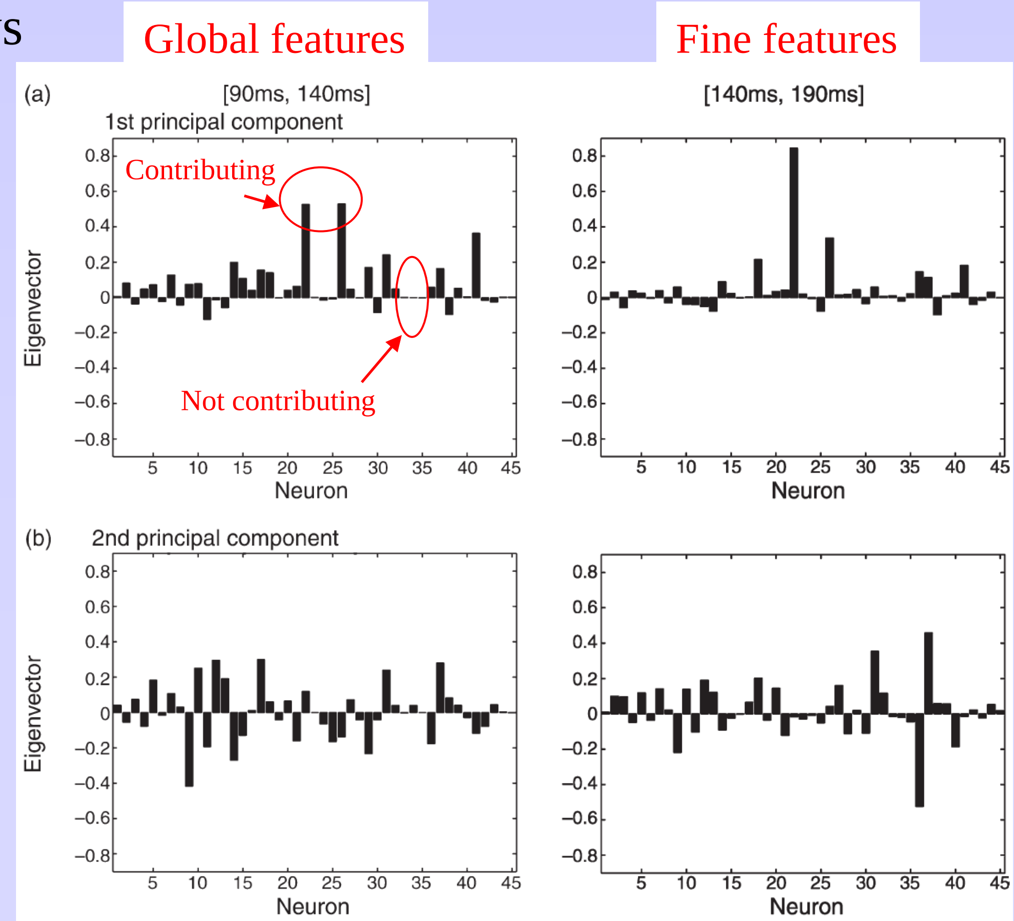
Principal Components Analysis

- Project on the first 2 PCs.
- Compute the mean vector for **global** features (human, monkey, shape)
- Compute the pair-wise sum of distances between centers (human-face, face-shape, human-shape) in 50ms windows
- Maximum occurs between 90 and 140ms



- For **fine** features (identity, expression, shape form), maximum occurred between 140 and 190 ms.

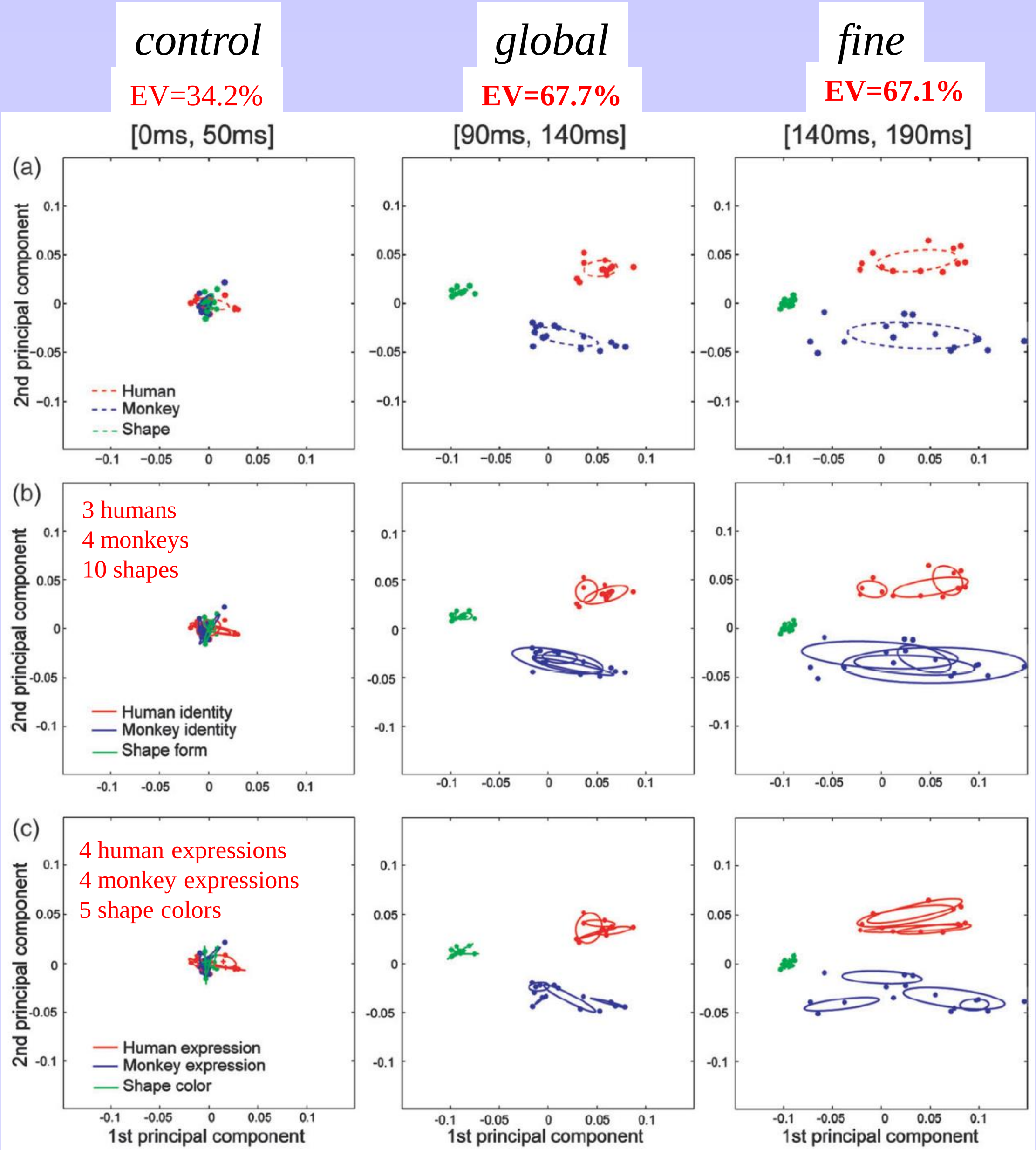
- Interpret the PCs. Which neurons are contributing?



(Matsumoto et al, 2005)

Principal Components Analysis

- High explained variance with only 2/45 PCs
- No discrimination in the 1st 50ms
- Shape is never discriminable
- Best **global** discriminability: 90-140 ms
- Best **fine** discriminability: 140-190 ms

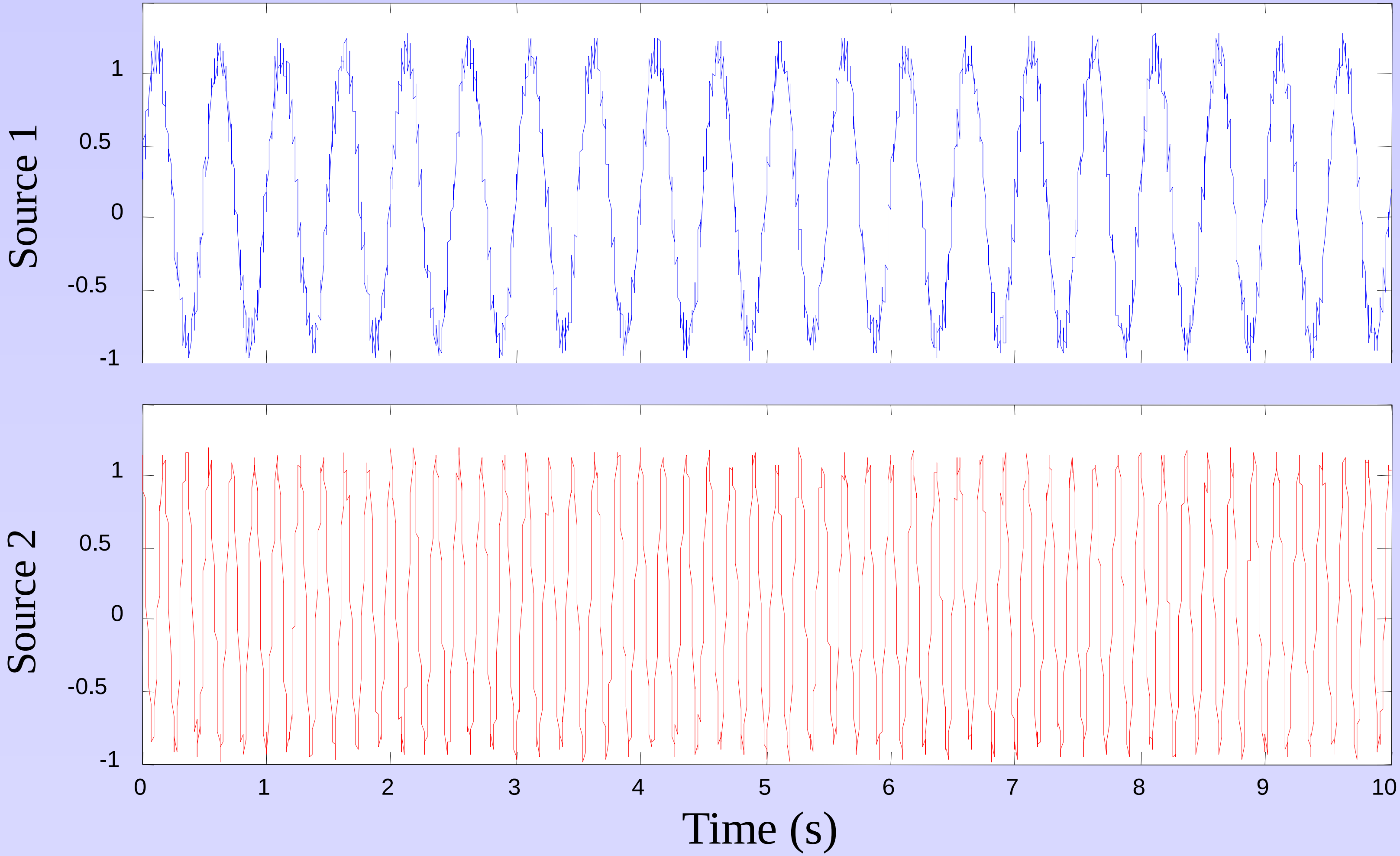


➔ Time of maximal discriminability in PC space, for various features of the stimulus set.

(Matsumoto et al, 2005)

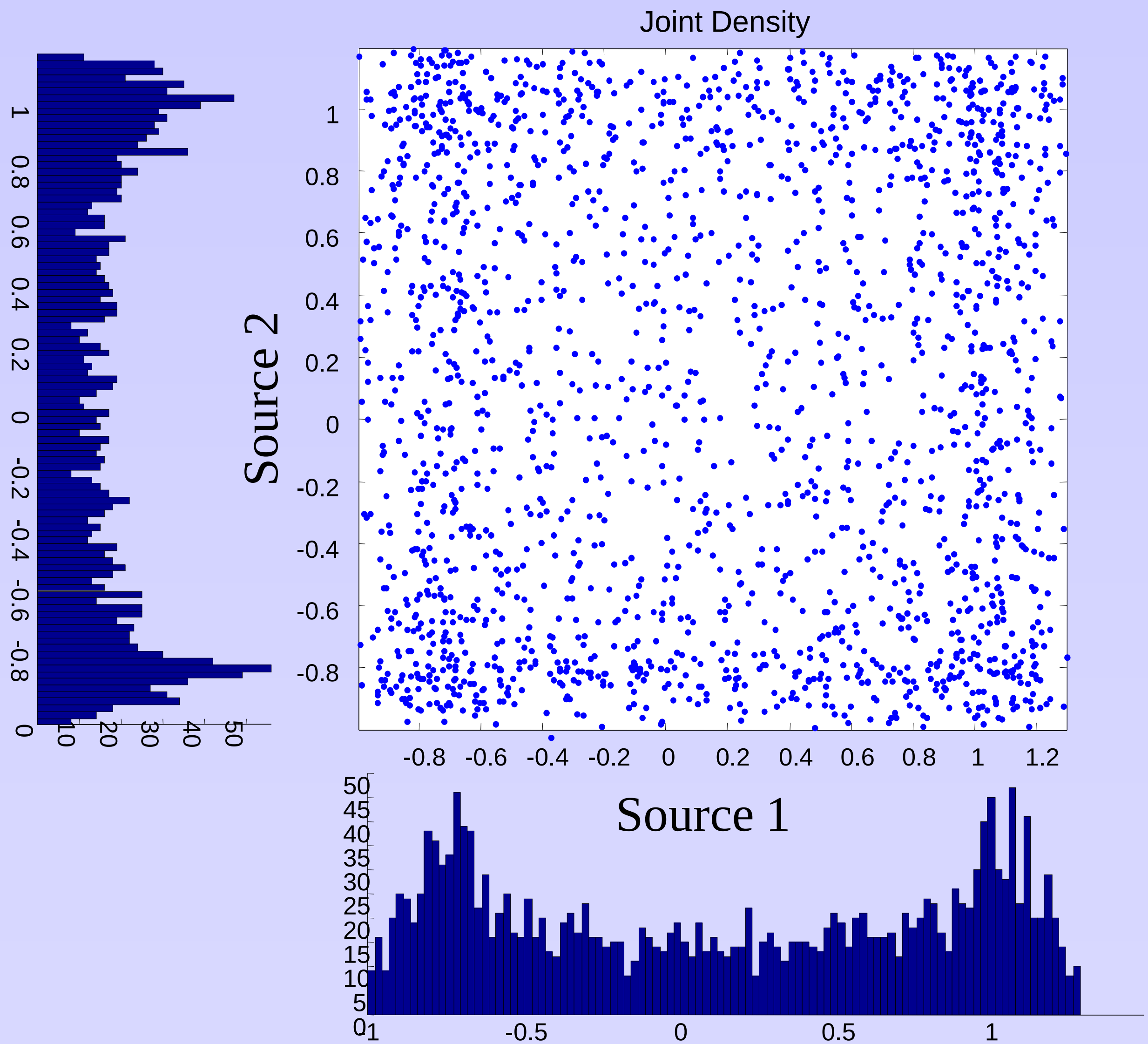
Independent Component Analysis: The basics

- The ground truth: 2 simultaneous noisy sources of information



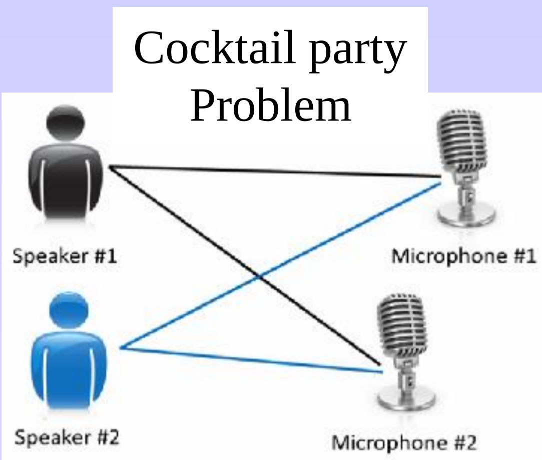
Independent Component Analysis: The basics

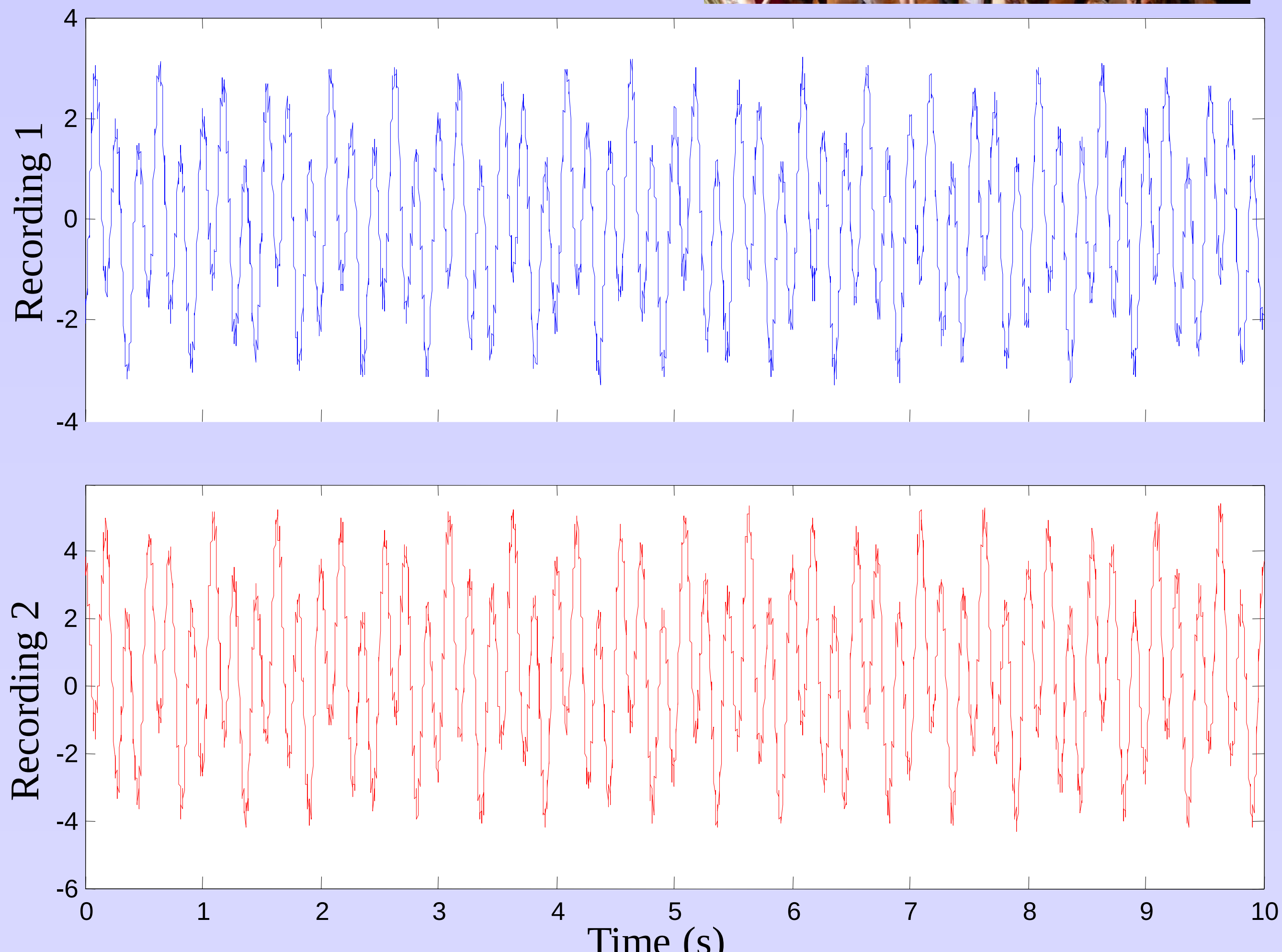
- Cannot be distinguished by amplitude alone



Independent Component Analysis: The basics

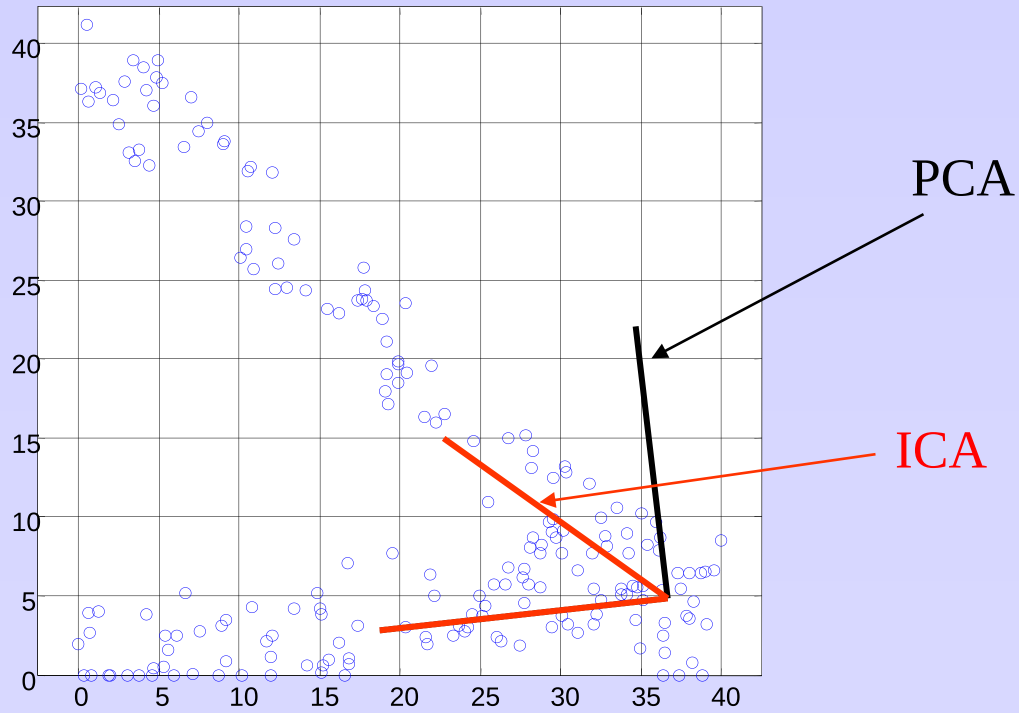
- The reality: Recording= Ft (Sources) = Mixture
- The sources are not distinguishable from the recordings alone.



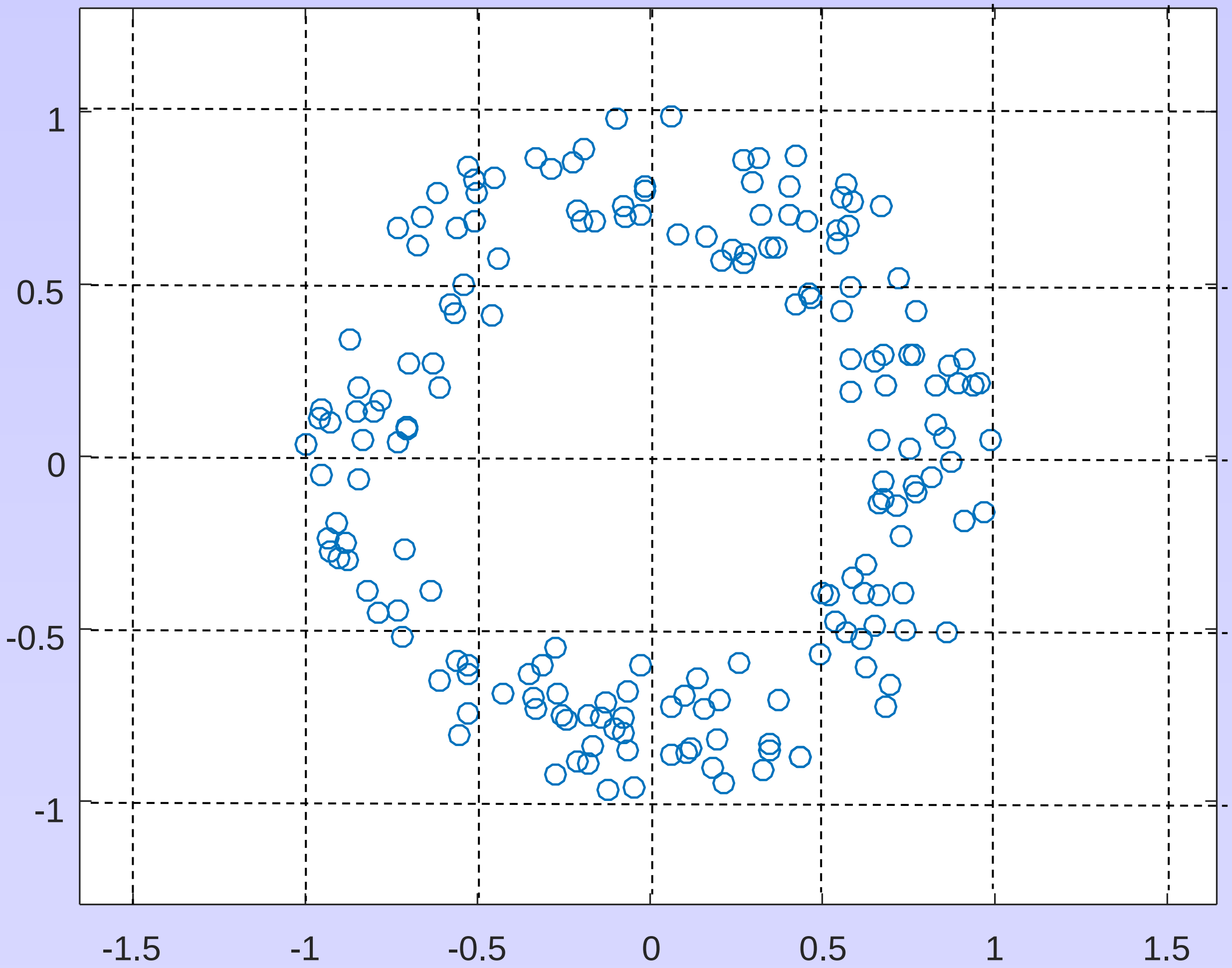


Independent Component Analysis: The basics

- ICA is a statistical method to find the underlying ‘sources’ hidden in a mixed set of signals.
- ICA works if sources are **non Gaussian**, and mutually **independent**
- ICA works if the sources are mixed **linearly**
- ICA is more general than PCA (and factor analysis): basis is not orthonormal
- PCA de-correlates (uses second order stats, e.g. variances), ICA de-mixes (e.g. uses kurtosis to assess Gaussian shapes)



- Is ICA always better than PCA?

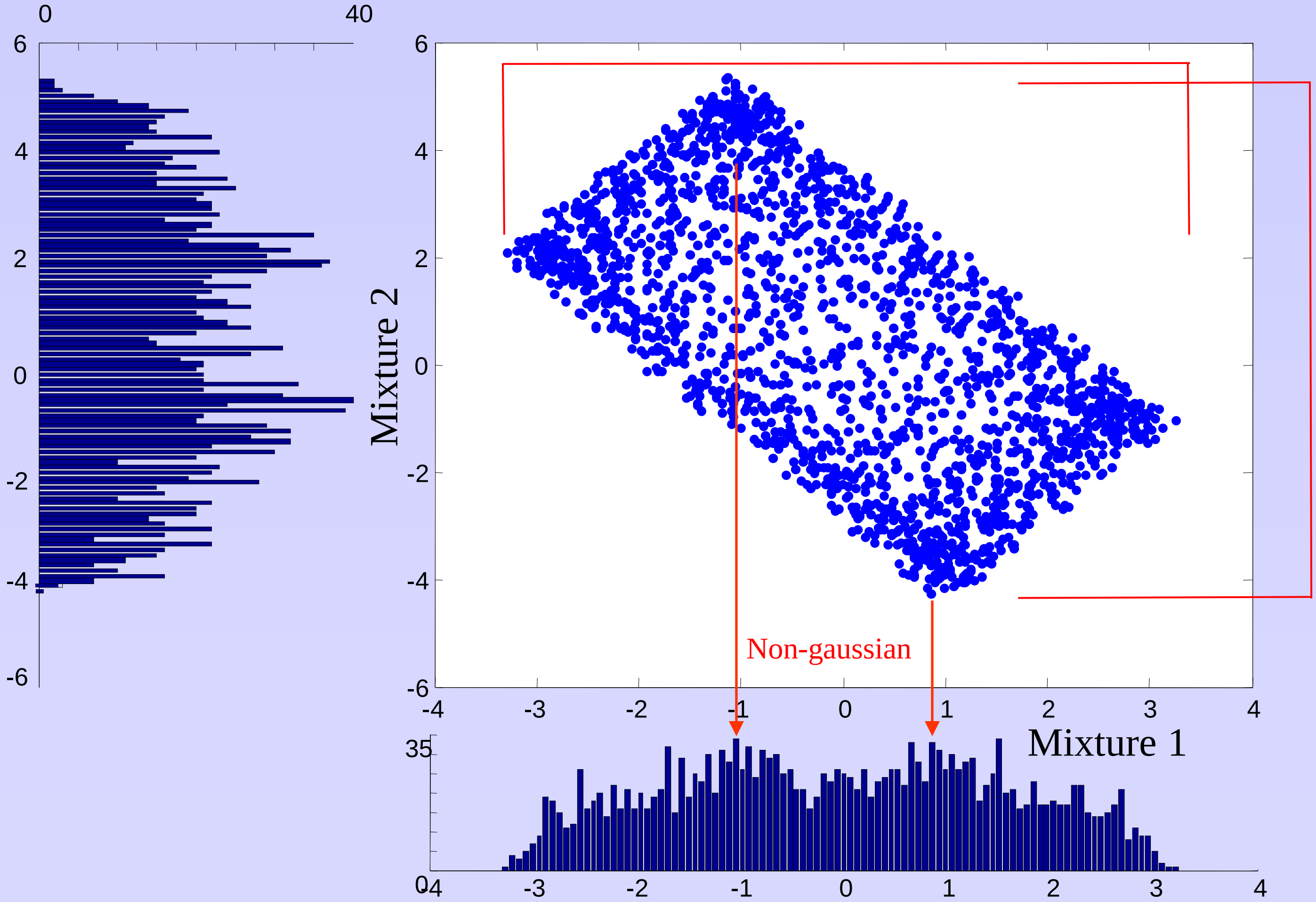


What are the PCs? ICs?

Independent Component Analysis: The basics

Step 1: Whitening

Rescaling to make distribution of values of equal variance and zero cross-covariance



Independent Component Analysis: The basics

Whitening = making the data 'spherical'

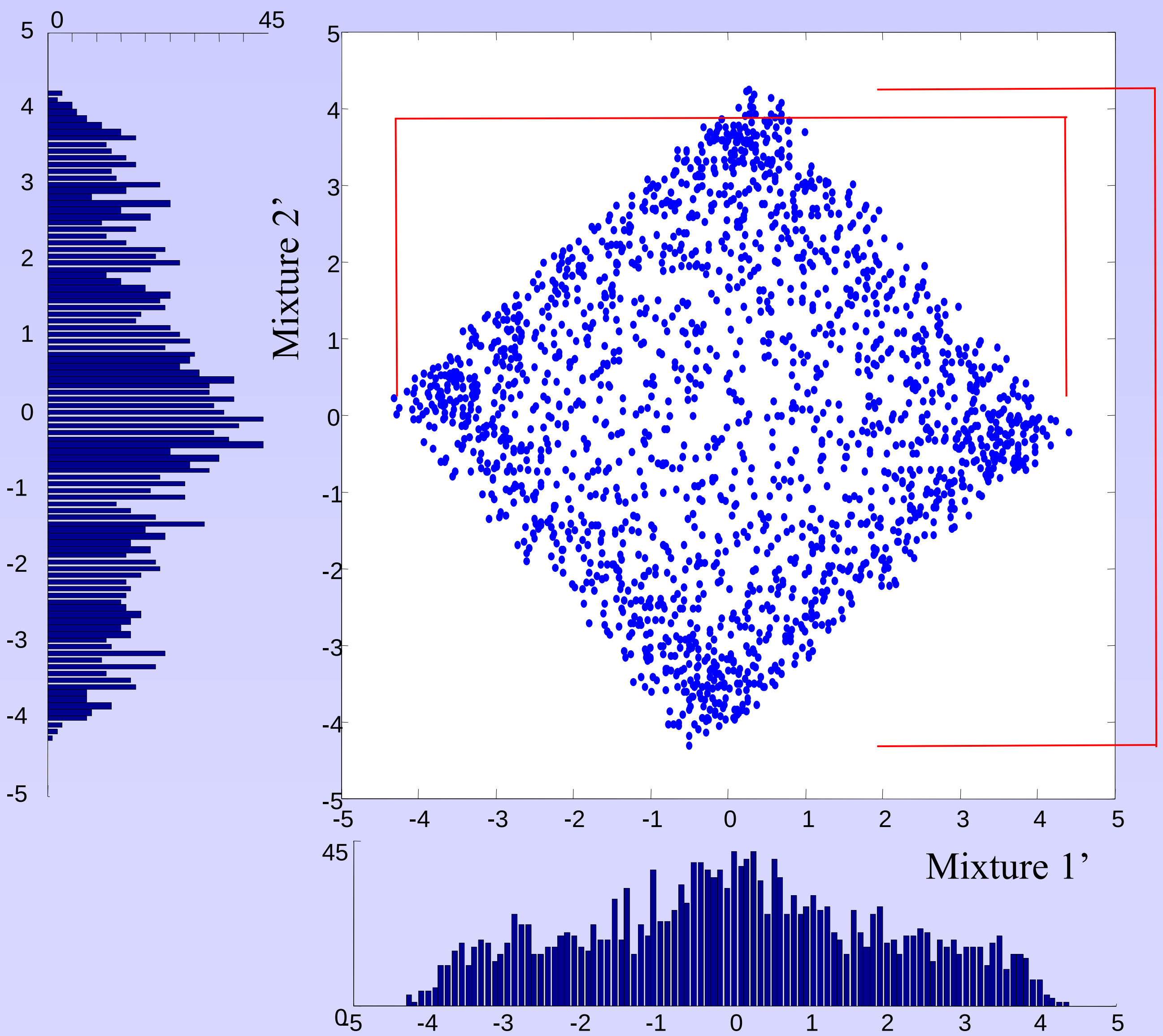
If X is the matrix containing the sources in columns:

$$Y = 2\left(\sqrt{\operatorname{cov}(X)}\right)^{-1} . \left(X - \langle X \rangle\right)$$

Note: in Matlab use: inv(), and sqrtm()

Independent Component Analysis: The basics

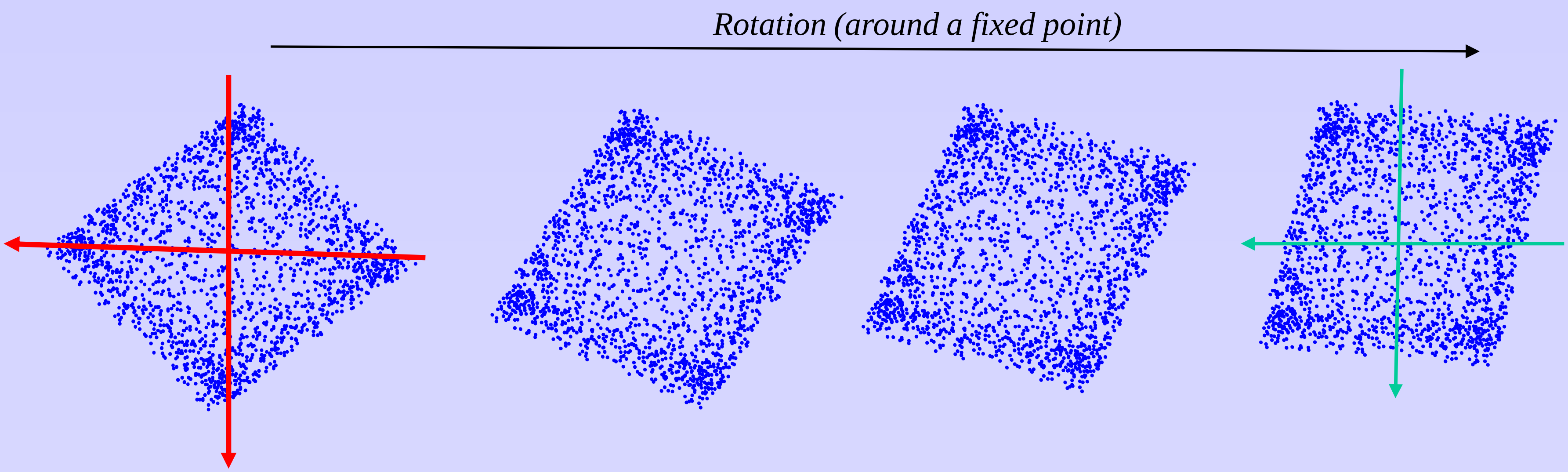
➔ equal variance, diagonal covariance matrix (no cross co-variance)



Independent Component Analysis: The basics

Step 2: Find a rotation of the joint-density that maximizes the non-normality of the distribution (i.e. makes them as ‘flat’ as possible, hence ‘independent’)

Central limit theorem: a mixture of independent variables is more gaussian than the original variables



→ Fixed point ICA (FastICA)

Independent Component Analysis: The basics

Step 3: fastICA (see code for details)

CD to the code2.5 folder
Execute fasticag.m
Load the data and leave 'name of variable' blank.