Institute of language, communication and the brain

ILCB

Summer school

INDEPENDENT COMPONENTS ANALYSIS

An Introduction to the Theory and Use in EEG Pre-processing and Analysis

ILCB SUMMER SCHOOL , SEPTEMBER 2-6, 2019

Aix-Marseille université

CNRS

A*Midex Initiative d'excellence Aix-Marseille

Inserm

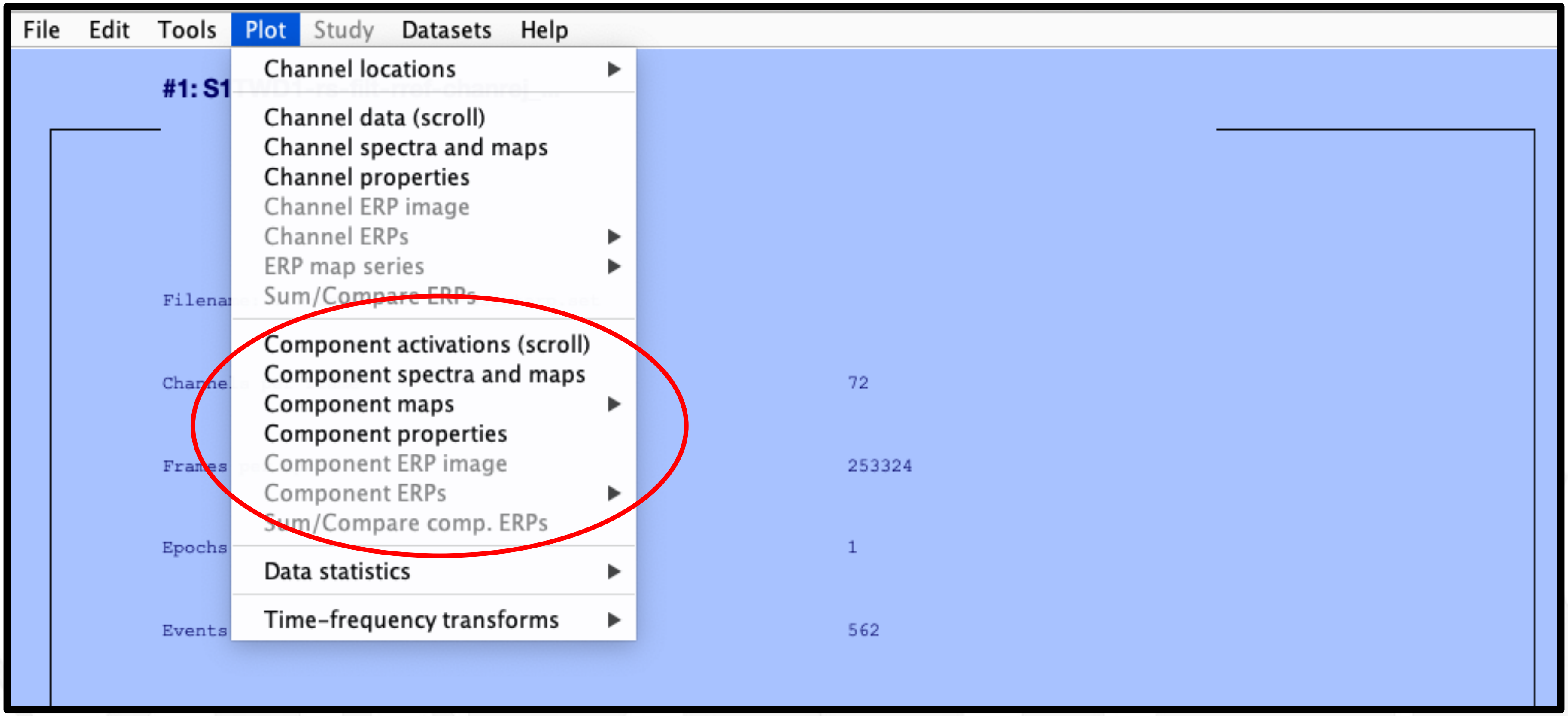
Investissements d'avenir

Université d'Avignon

Institute of language, communication and the brain

ILCB

INDEPENDENT COMPONENTS ANALYSIS



Used a tool for decomposing EEG signals into activity patterns that are presumed to correspond to activity in single active cortical areas and sources of non-cortical artifacts.

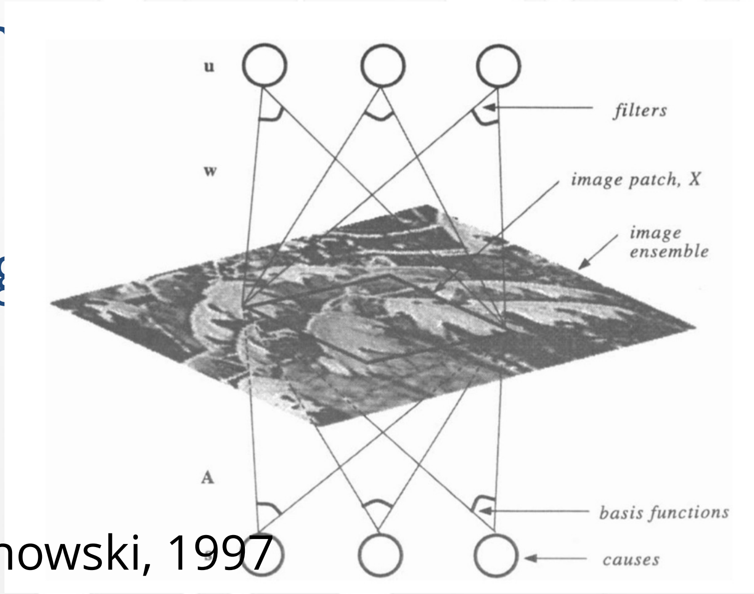
INDEPENDENT COMPONENTS ANALYSIS

A specific method for performing Blind Source Separation (BSS).

A data-driven method for decomposing a set of **observations** into a set of **statistically independent** components.

Independent Component Analysis (I**CA**) **demixes mutually independent sources** em
multivariate signals.

Used in biomedical signal processin
image and video analysis…



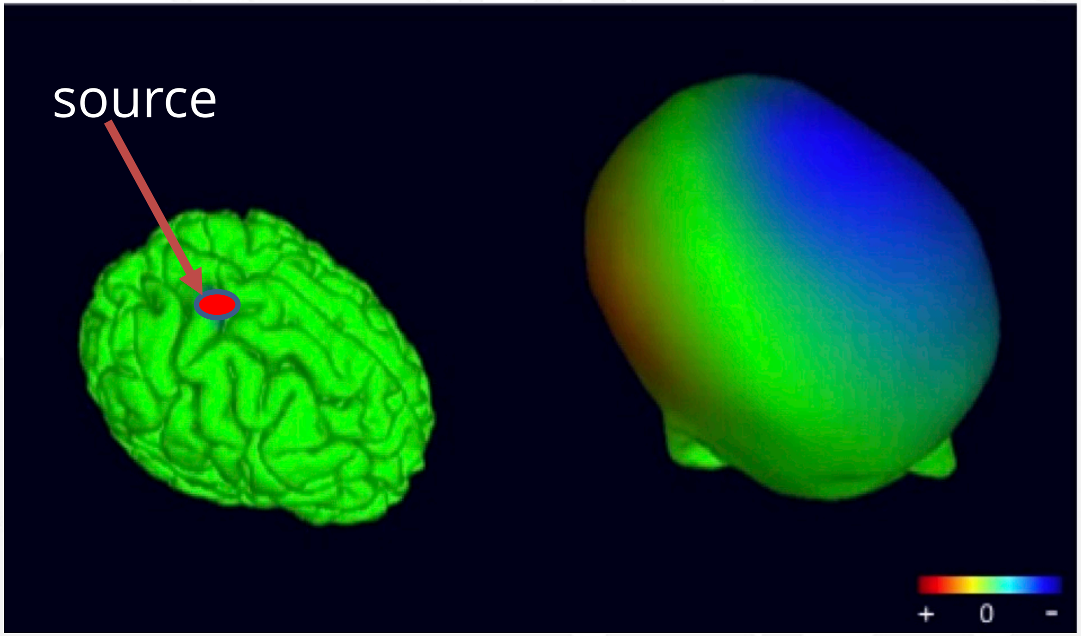
Bell & Sejnowski, 1997

ICA IN EEG ANALYSIS – MOTIVATION

In EEG, what factor/s underlie how the source signals are mixed to form the mixture signals?

Neural activity is conducted to the scalp and sensors by **volume conduction**.

The activity at each sensor is a weighted sum of activity from different sources as well as sources of non-brain artifacts.

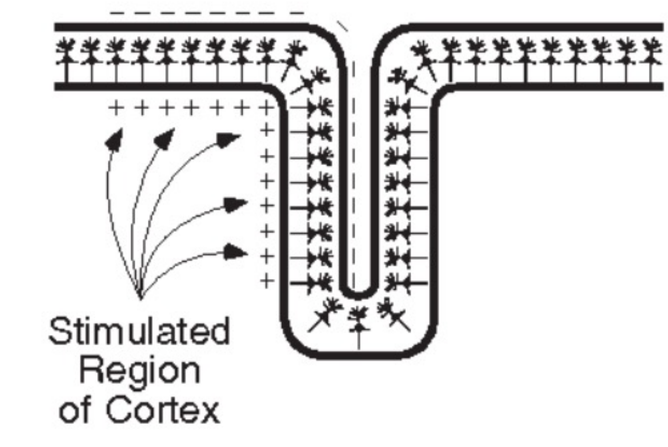


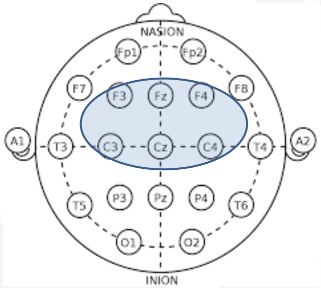
Akalin, Acar et al, 2011

ICA IN EEG ANALYSIS - MOTIVATION

Differences in cortical folding patterns between individuals can produce different net source orientations…

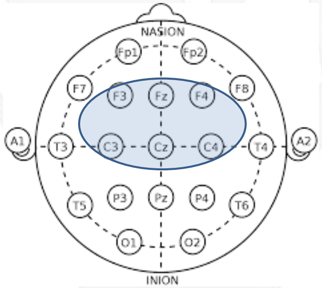
Inter-individual differences in how the activity is projected to the sensors at the scalp level





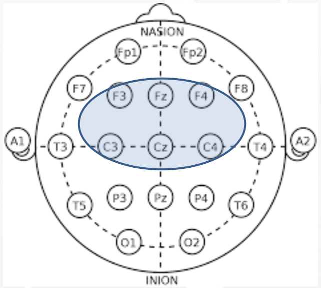
subject1

vs



subject2

…



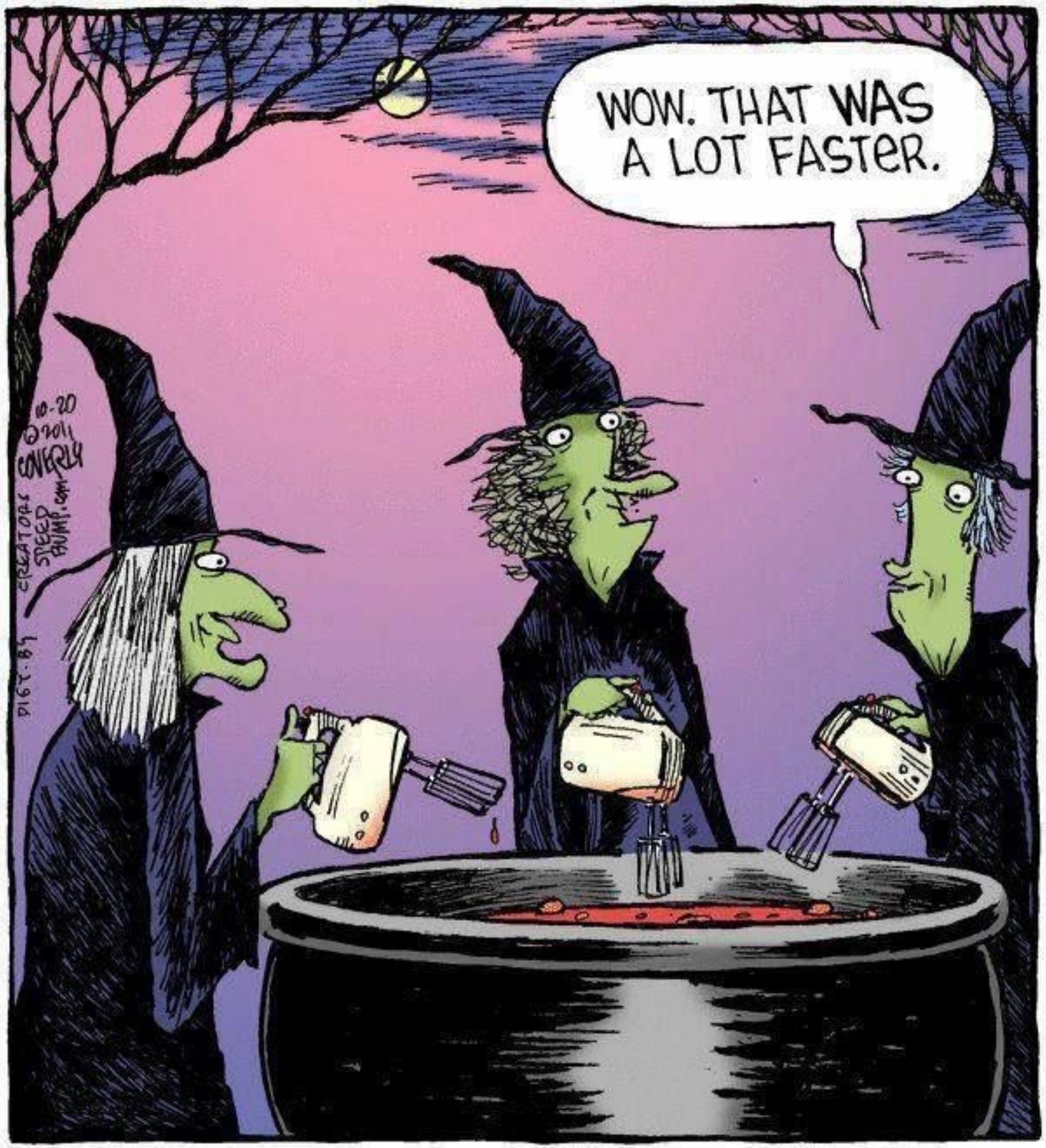
subject N

Accurate???

Institute of language, communication and the brain

ILCB

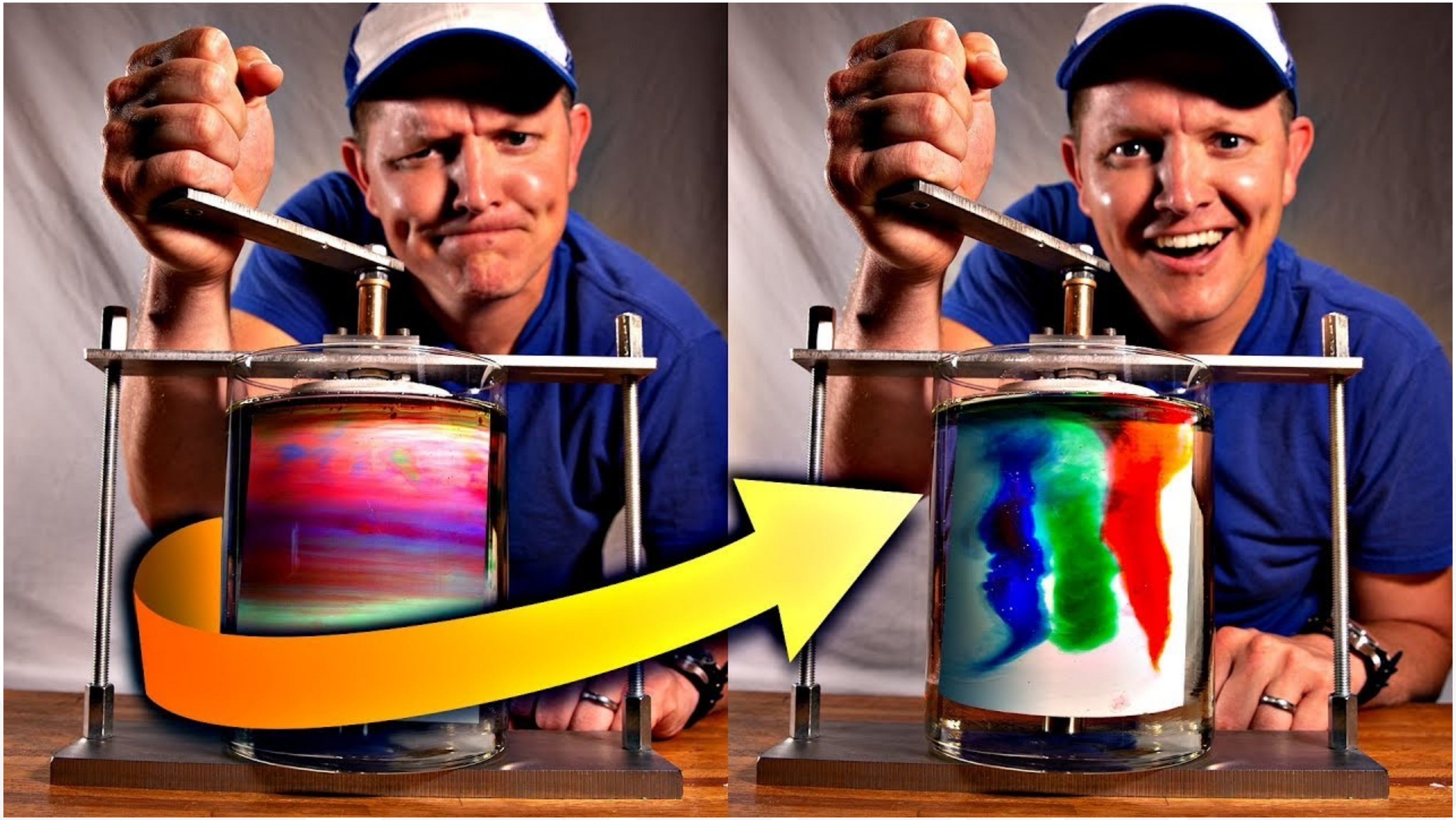
MIXING



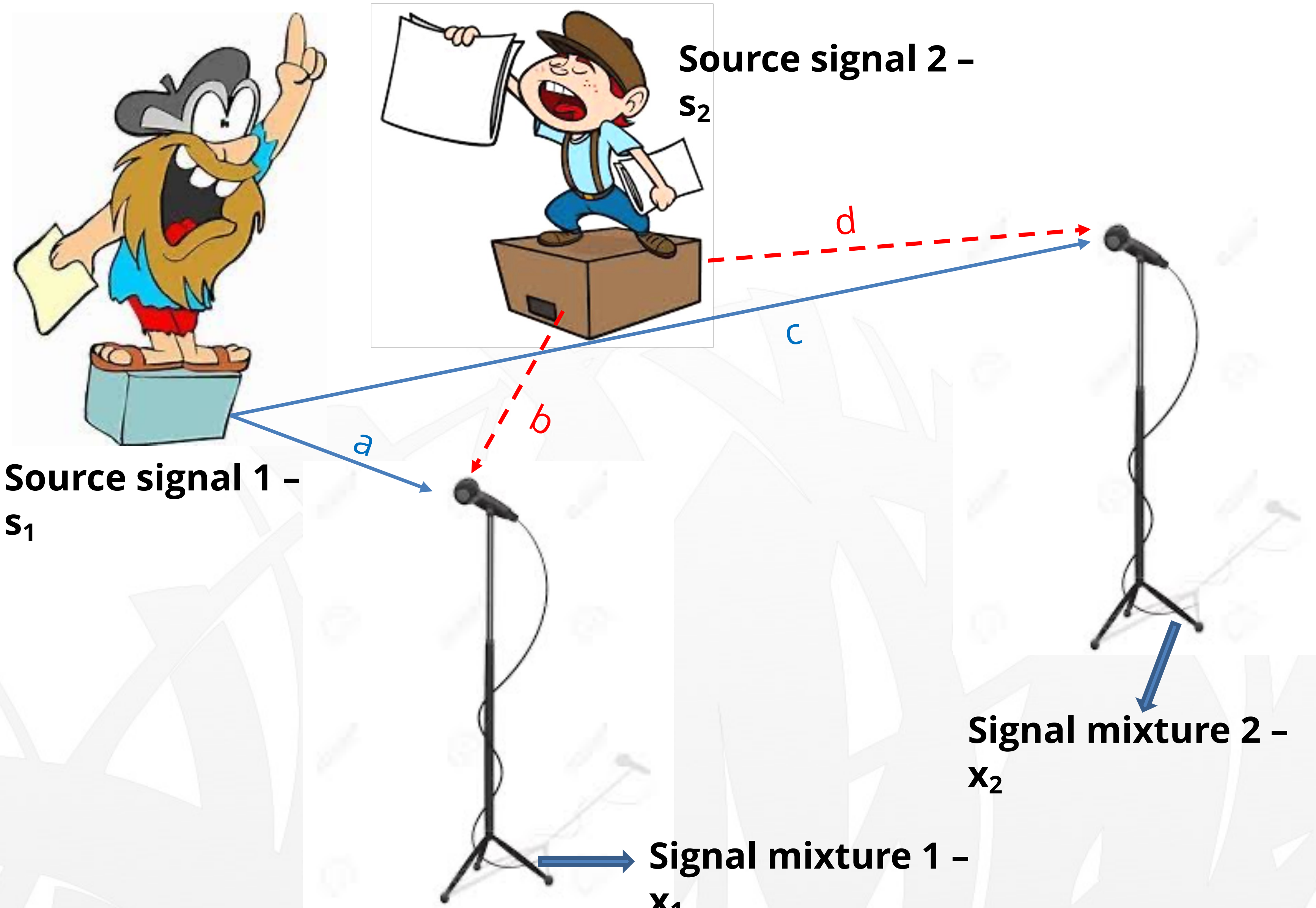
Institute of language, communication and the brain

ILCB

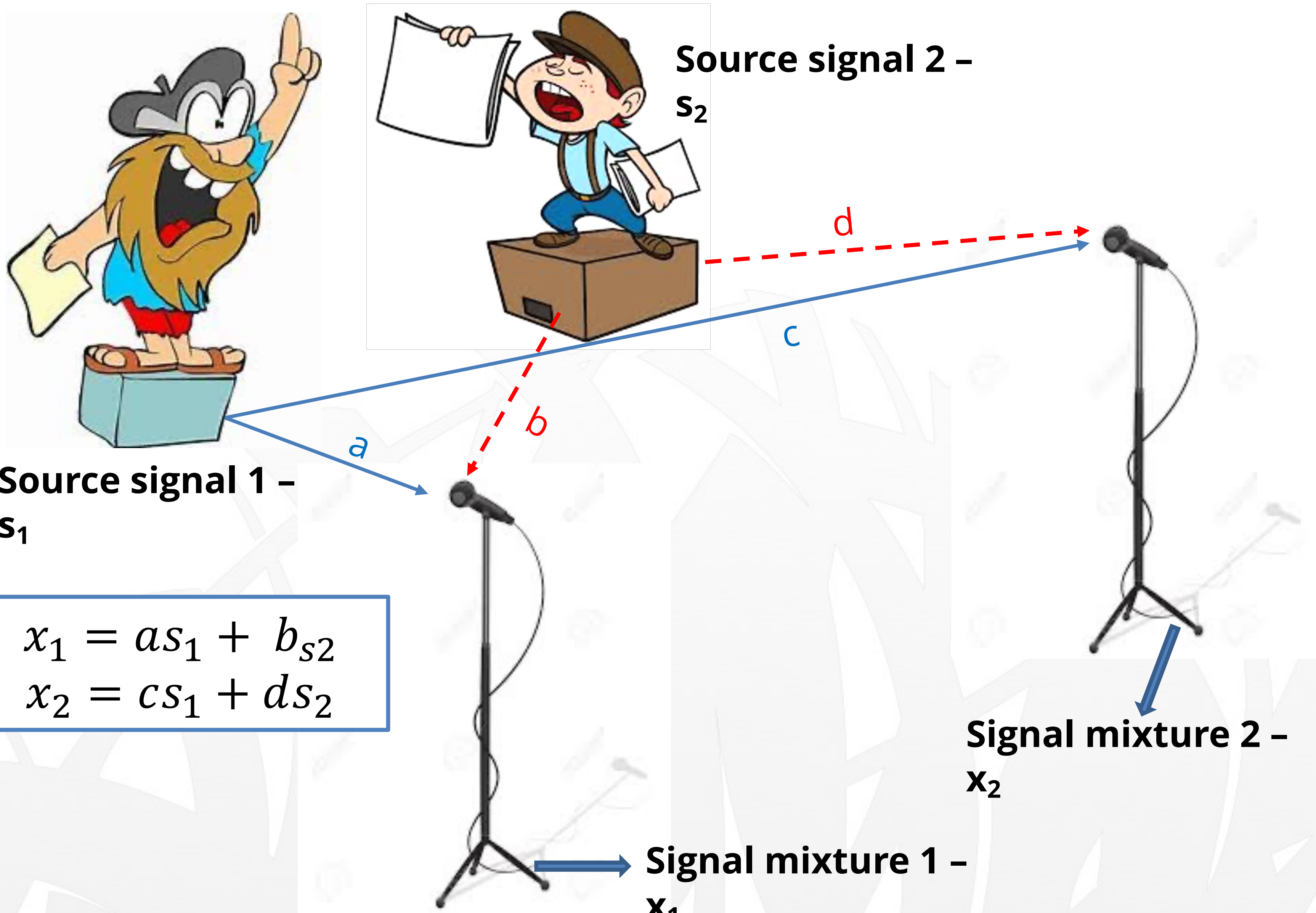
UNMIXING



THE COCKTAIL PARTY PROBLEM

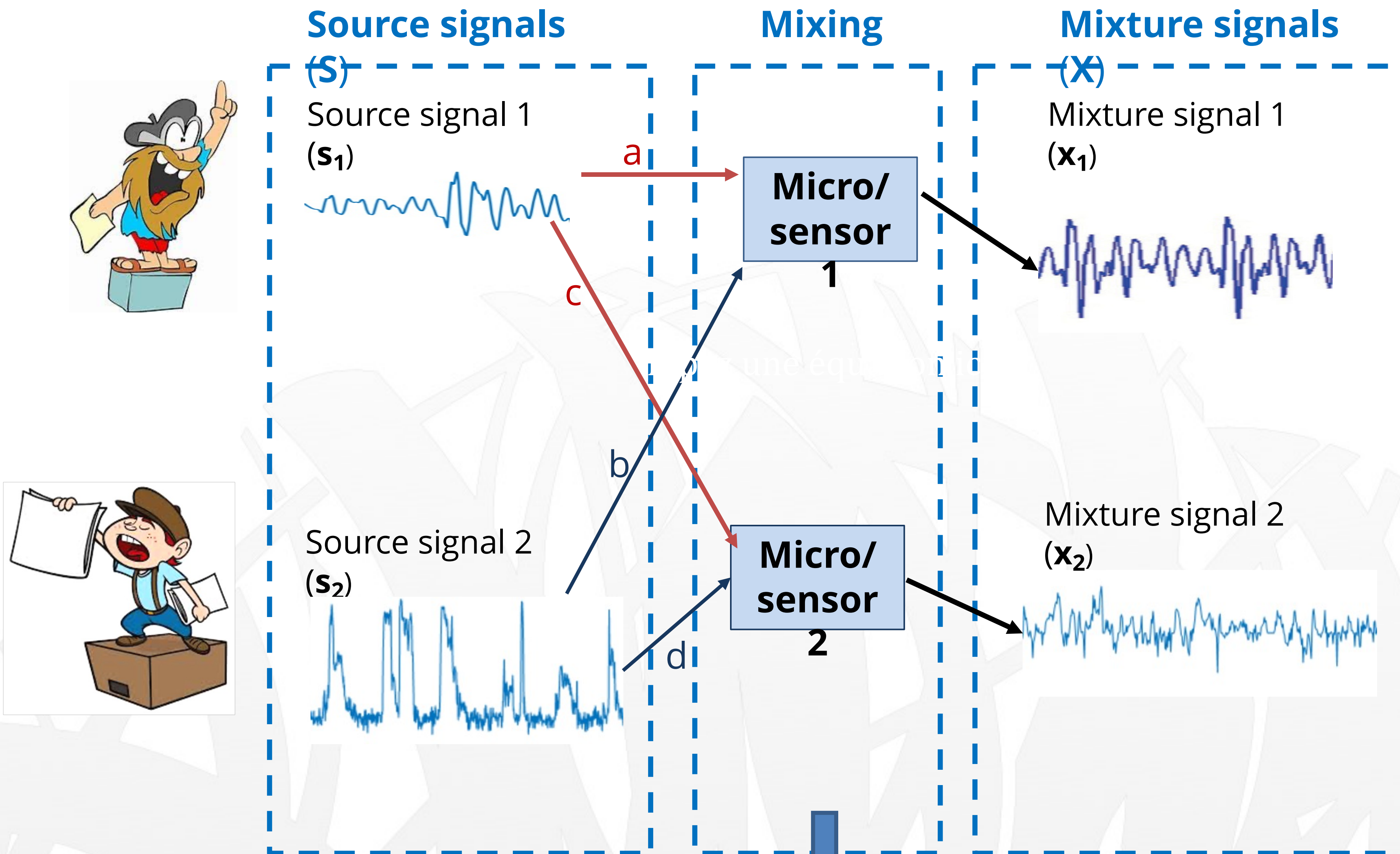


THE COCKTAIL PARTY PROBLEM



$$
\boxed{\begin{aligned} x_1 &= a s_1 + \ b_{s2} \\ x_2 &= c s_1 + d s_2 \end{aligned}}
$$

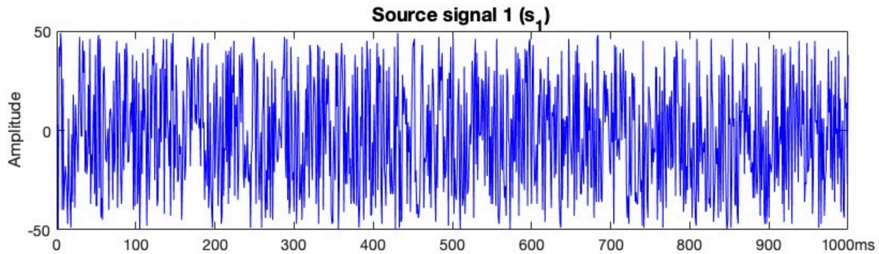
MIXING SIGNALS

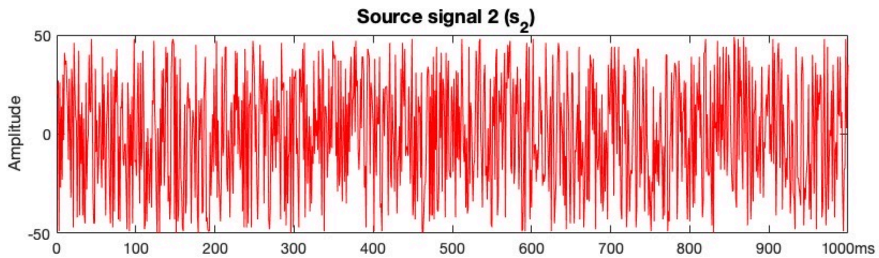


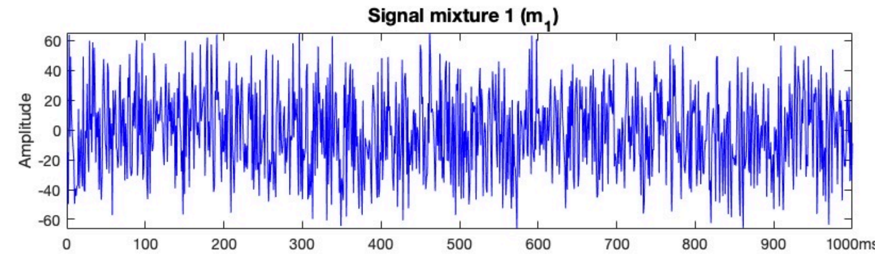
$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

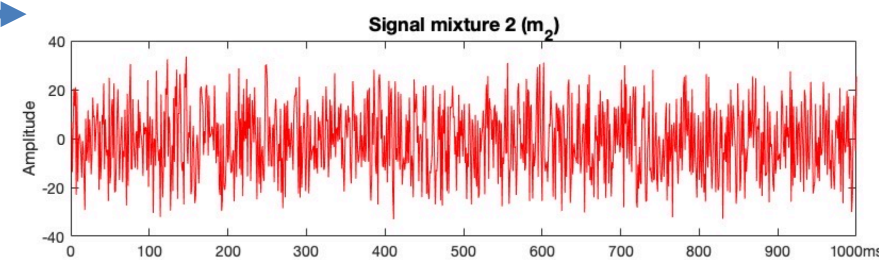
Institute of language, communication and the brain

MIXING SIGNALS









$$\boldsymbol{S} = (\boldsymbol{s}_1, \boldsymbol{s}_2)^{\mathsf{T}} = \begin{pmatrix} \boldsymbol{s}_1 \\ \boldsymbol{s}_2 \end{pmatrix} = \begin{pmatrix} s_1^1, s_1^2, \ldots & s_1^N \\ s_2^1, s_2^2, \ldots & s_2^N \end{pmatrix}$$

$$\mathbf{X} = \begin{pmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \boldsymbol{s}_1 \\ \boldsymbol{s}_2 \end{pmatrix} = \begin{pmatrix} a\boldsymbol{s}_1 + b\boldsymbol{s}_2 \\ c\boldsymbol{s}_1 + d\boldsymbol{s}_2 \end{pmatrix} = \mathbf{AS}$$

...where a = 1.54, b = 2.84, c = 1.42 and d = 0.3 are the mixing coefficients

The mixing coefficients transform, linearly, source signals (space **S**) to mixed signals (space **X**).

$$\mathbf{S} \rightarrow \boldsymbol{X} : \boldsymbol{X} = \boldsymbol{AS}$$

...where A ∈ $R^{p \times M}$ is the mixing coefficients matrix.

MIXING SIGNALS

What can we say about the cocktail party problem?

The problem involves separating a set of **source** signals from a set of mixture **signals.**

To extract the source signals we can only use what we know or assume about :

- The mixture signals
- The source signals.

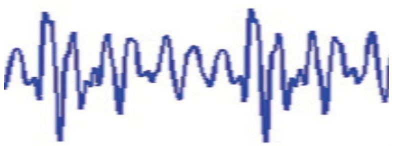
We cannot directly observe the source signals ➔ latent variables.

The mixing signals ➔ observed data

UNMIXING

Mixture signals (X)

Mixture signal 1 ($\mathbf{x_1}$)



Mixture signal 2 ($\mathbf{x_2}$)

Unmixing

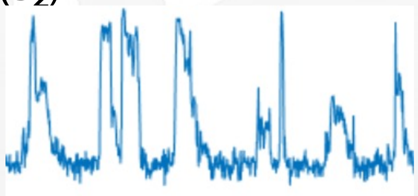
BSS (ICA)

$$\mathbf{W} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$$

Extracted signals (Y)

Source signal 1 ($\mathbf{s_1}$)

Source signal 2 ($\mathbf{s_2}$)



UNMIXING

Mixing signals

$$\begin{pmatrix} \boldsymbol{x}_1^1, \boldsymbol{x}_1^2, \dots & \boldsymbol{x}_1^N \\ \boldsymbol{x}_2^1, \boldsymbol{x}_2^2, \dots & \boldsymbol{x}_2^N \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \boldsymbol{s}_1^1, \boldsymbol{s}_1^2, \dots & \boldsymbol{s}_1^N \\ \boldsymbol{s}_2^1, \boldsymbol{s}_2^2, \dots & \boldsymbol{s}_2^N \end{pmatrix}$$

$$= (\boldsymbol{a}_1, \boldsymbol{a}_2)^T (\boldsymbol{s}_1, \boldsymbol{s}_2)$$

$$\mathbf{X} = \mathbf{AS}$$

Unmixing signals

$$\begin{pmatrix} \boldsymbol{s}_1^1, \boldsymbol{s}_1^2, \dots & \boldsymbol{s}_1^N \\ \boldsymbol{s}_2^1, \boldsymbol{s}_2^2, \dots & \boldsymbol{s}_2^N \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_1^1, \boldsymbol{x}_1^2, \dots & \boldsymbol{x}_1^N \\ \boldsymbol{x}_2^1, \boldsymbol{x}_2^2, \dots & \boldsymbol{x}_2^N \end{pmatrix}$$

$$= (\boldsymbol{w}_1, \boldsymbol{w}_2)^T (\boldsymbol{x}_1, \boldsymbol{x}_2)$$

$$\mathbf{Y} = \mathbf{WX} \approx \boldsymbol{S}$$

$\boldsymbol{W} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ is unmixing matrix

Note: Unmixing reverses mixing $\Longrightarrow$. $\boxed{\boldsymbol{W} = \boldsymbol{A}^{-1}}$

But I don't know **A** !!

Institute of language, communication and the brain

AIMS OF THIS SESSION

AIMS:

- Understand the assumptions that one makes when carrying out Independent Components Analysis.
- To promote judicious use of ICA.

OBJECTIVES:

- To introduce the basics of the ICA approach.
- To distinguish ICA from other related methods.
- To introduce the various measures that ICA uses to extract the independent components.
- To consider the use of ICA for EEG preprocessing and analysis.

INDEPENDENT COMPONENT ANALYSIS: ASSUMPTIONS 1

Can be used to seperate a set of signal mixtures into a set of source signals called **independent components**.

It assumes that the source signals extracted are **statistically independent.**

It assumes that the observed data is a **linear** mixture of the underlying sources**.**

MUTUAL INDEPENDENCE

Examples…

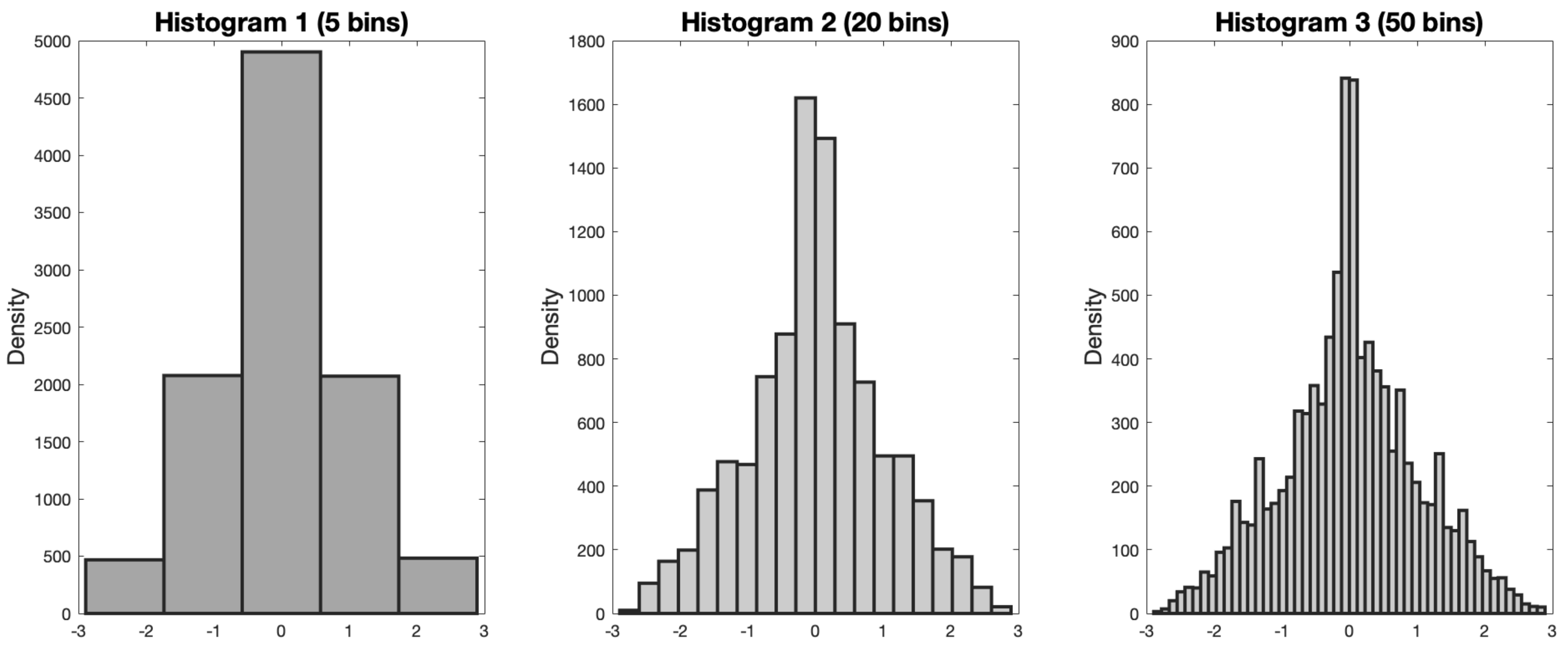
MUTUAL INDEPENDENCE: Discrete events

Probability of getting heads 4 times in a row …

As every coin toss has an outcome that is independent of all other coin tosses…

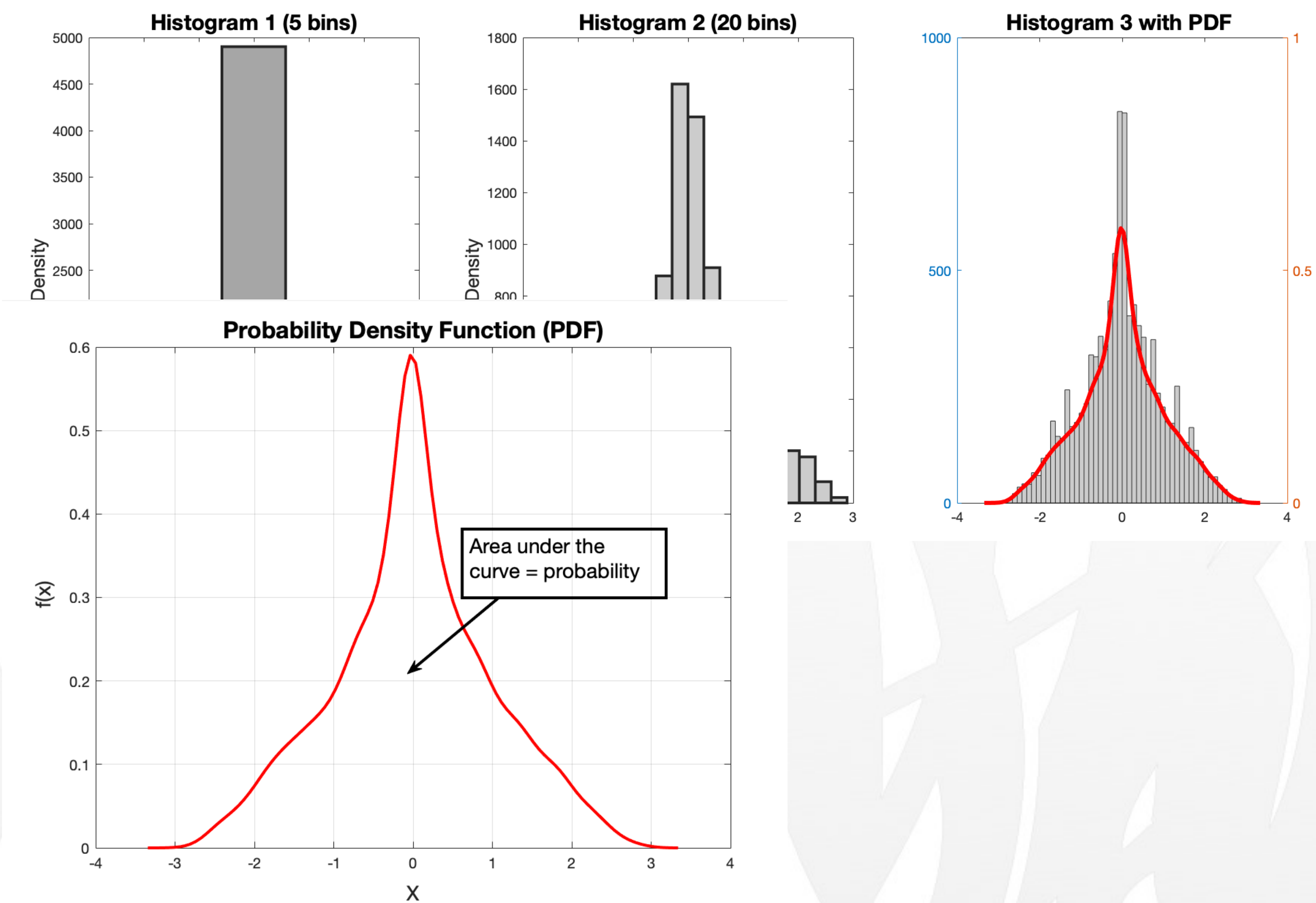
$$P \;\; p_h \times p_h = \prod_{i=1}^{N} p_h = p_h^N \qquad = P(heads)^4 = .5^4 = .0625$$

MUTUAL INDEPENDENCE: Continuous variables



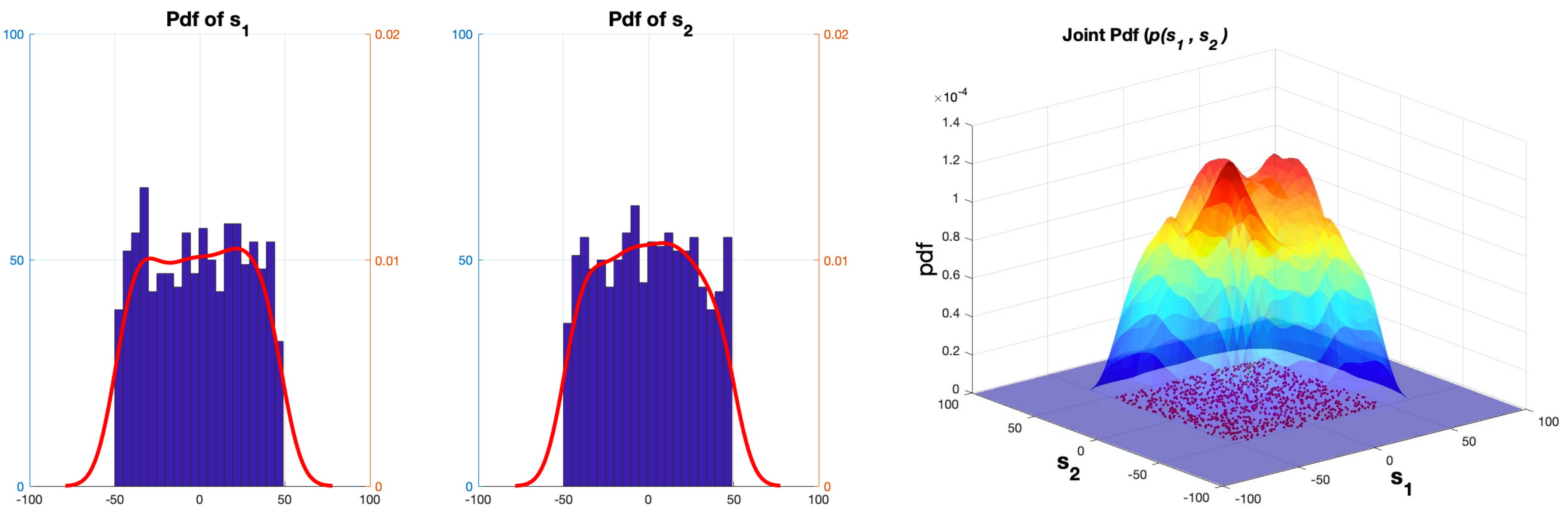
Institute of
language, communication
and the brain

MUTUAL INDEPENDENCE: Continuous variables



MUTUAL INDEPENDENCE

For Continuous Variables:

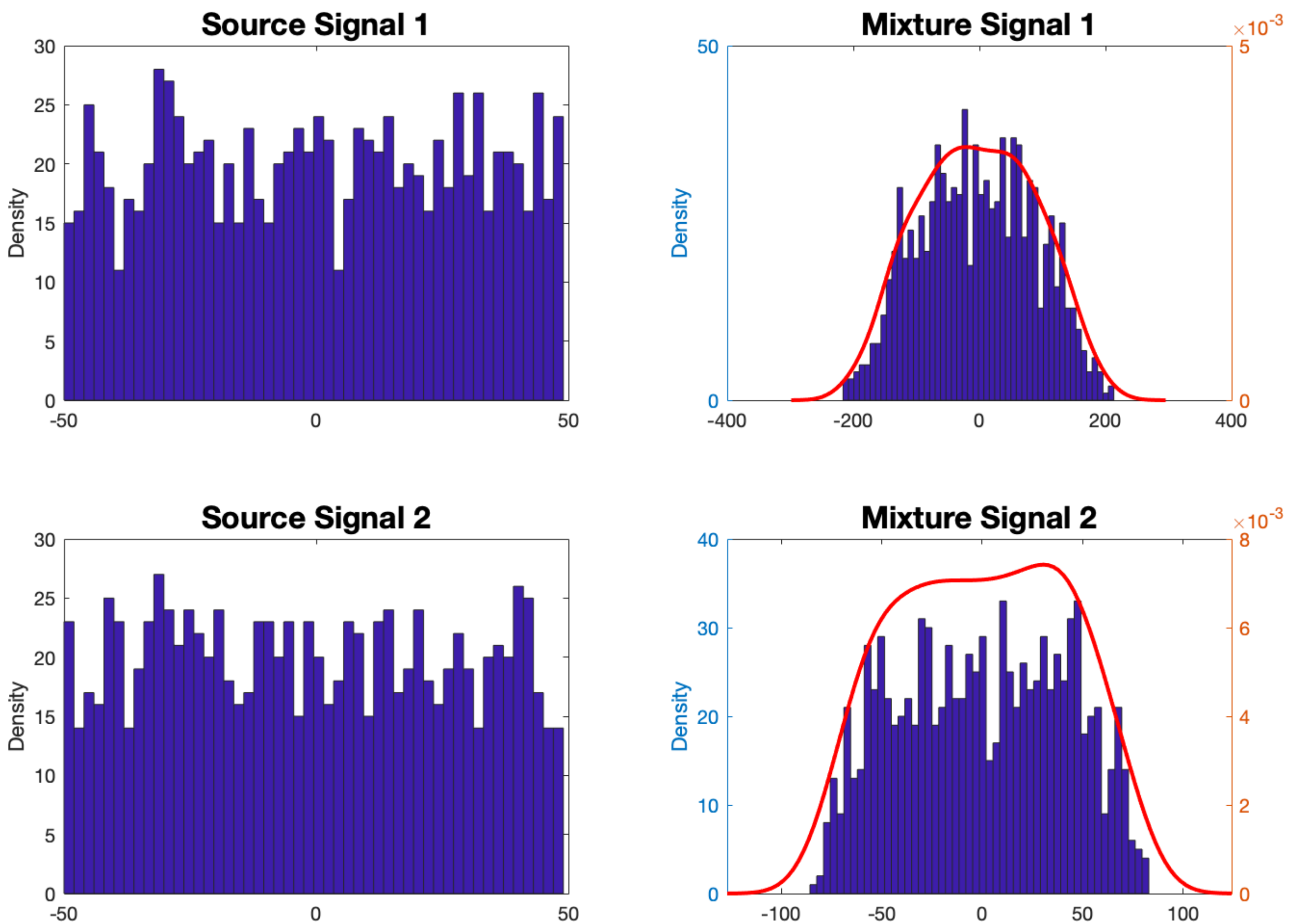


Statistically independent if:

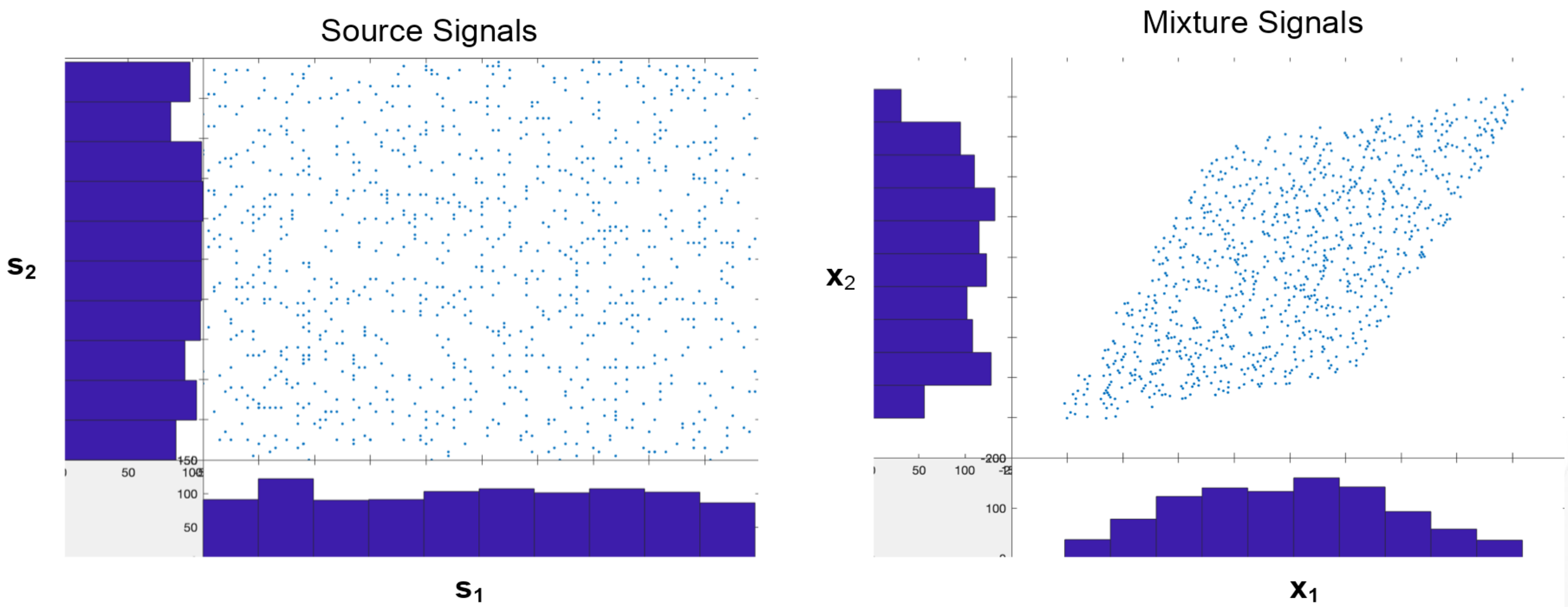
$$\underbrace{P_{s_1,s_2}(s_1,s_2)}_{\text{Joint pdf}} = \underbrace{p_{s_1}(s_1) \times p_{s_2}(s_2)}_{\text{marginal pdfs}}$$

Institute of language, communication and the brain

INDEPENDENT COMPONENT ANALYSIS: ASSUMPTIONS 2



INDEPENDENT COMPONENT ANALYSIS: ASSUMPTIONS 2



The pdf (probability density function) of signal mixture is **MORE <u>GAUSSIAN</u>** than that of the source signals.

PROPERTIES OF SOURCE SIGNALS

The Central Limit Theorem:

– The distribution of the sum of independent random variables tends toward a Gaussian distribution.

– The sum of two independent random variables usually has a distribution that is closer to gaussian than any of the two original random variables.

We can rotate each row vector in **W** until we find a weight vector will extract a source signal that is as **non-Gaussian** as possible.

In ICA maximising non-Gaussianity ➔ maximizing…a measure of non-Gaussianity

INDEPENDENT COMPONENT ANALYSIS

Can be used to seperate a set of signal mixtures into a set of source signals called **independent components**.

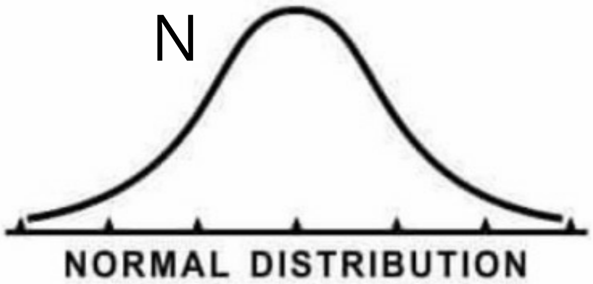
It assumes that the source signals extracted are **statistically independent.**

It assumes that the observed data is a **linear** mixture of the underlying sources.

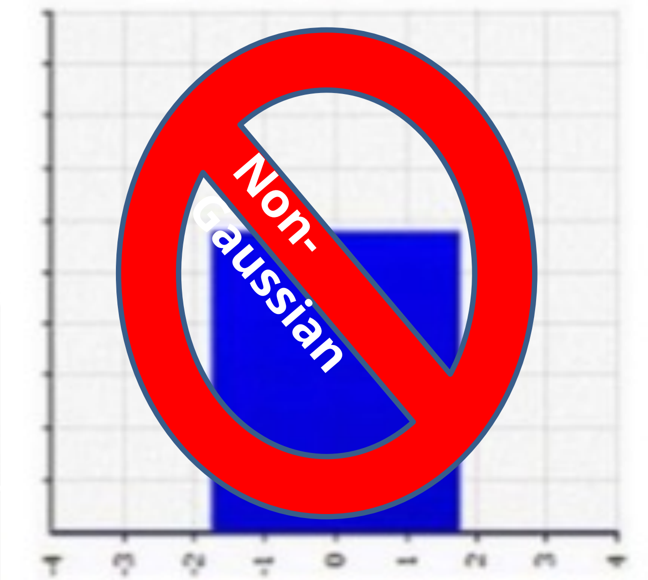
It assumes that the source signals have a **non-normal** or **non-Gaussian distribution.**

DISTRIBUTIONS

GAUSSIAN



SUB-GAUSSIAN

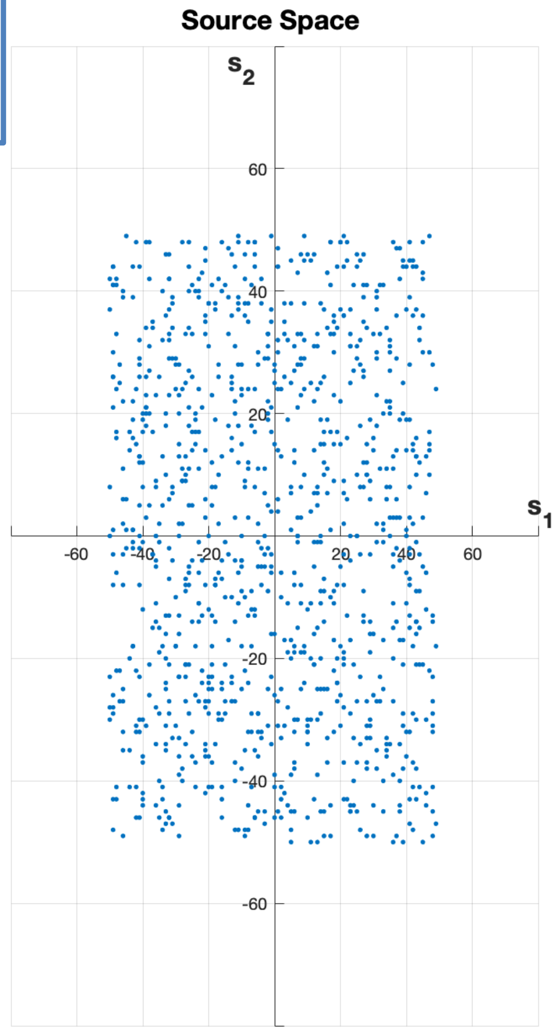


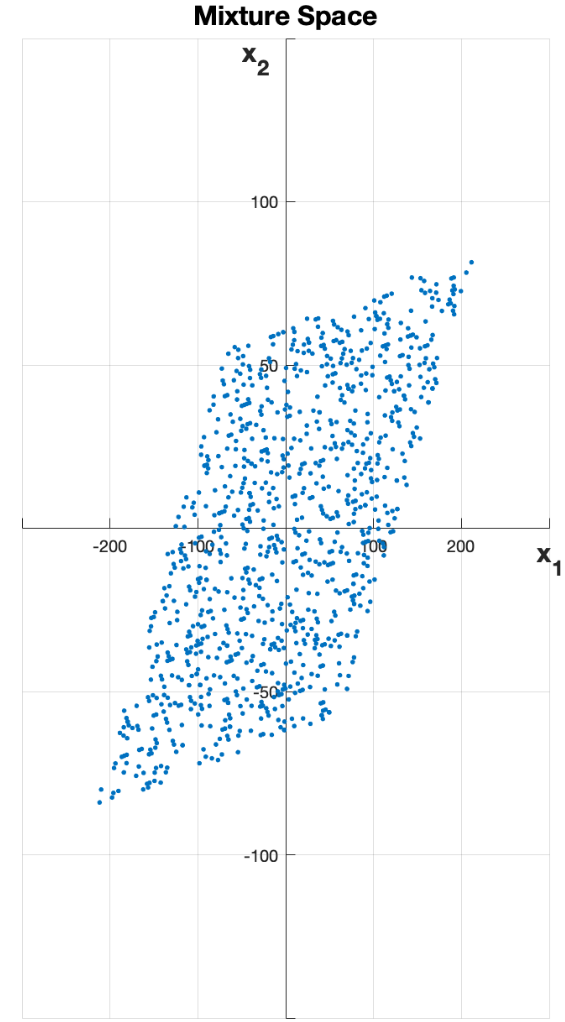
SUPER-GAUSSIAN

ICA: Simple Illustration

Joint Distributions

Uniform density





The two source signals, $\mathbf{s_1}$ (x-axis) and $\mathbf{s_2}$ (y-axis) form space **S**

The two mixture signals, $\mathbf{x_1}$ (x-axis) and $\mathbf{x_2}$ (y-axis) form space **X**

Institute of language, communication and the brain

ILCB

ICA: Simple Illustration

$$A = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

Mixing coefficients of $\mathbf{s_1}$

Mixing coefficients of $\mathbf{s_2}$

$$A = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

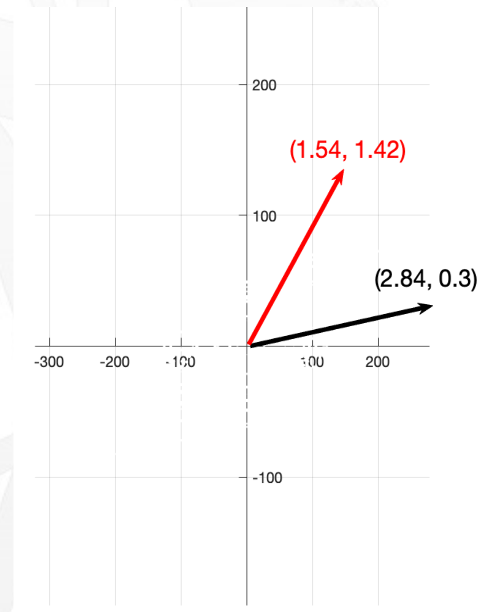
2 mixture signals

2 source signals

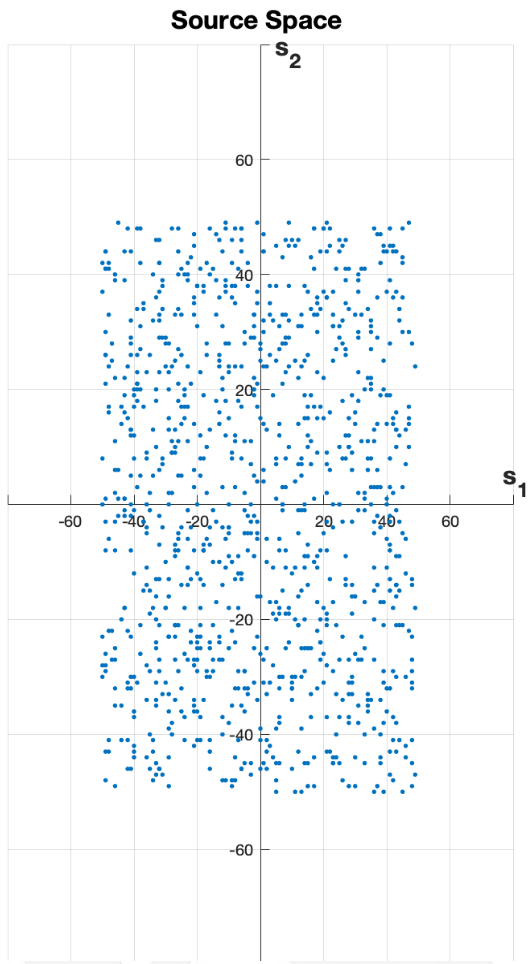
A is square.

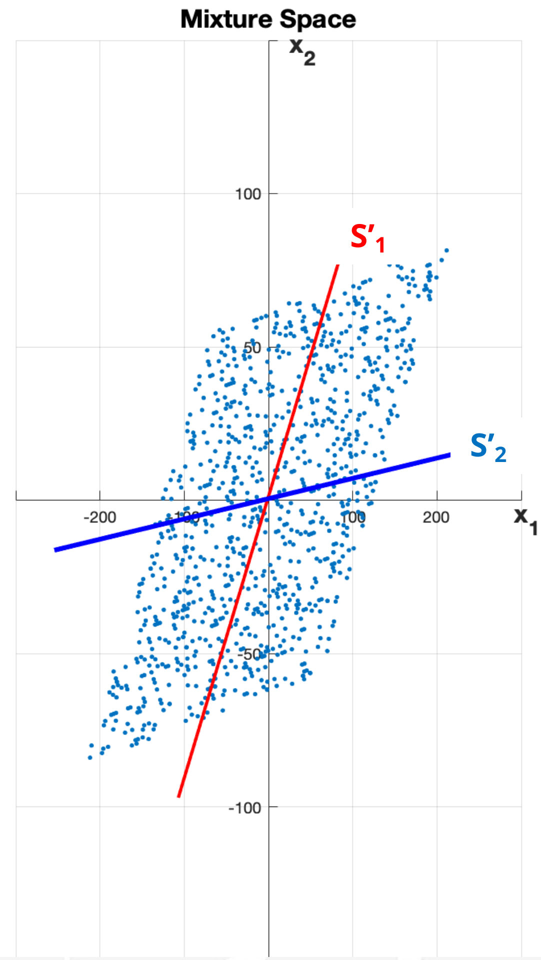
But A also should be **full rank…**

$$\mathbf{A} = \begin{pmatrix} 1.54 & 2.84 \\ 1.42 & 0.3 \end{pmatrix}$$



ICA: Simple Illustration



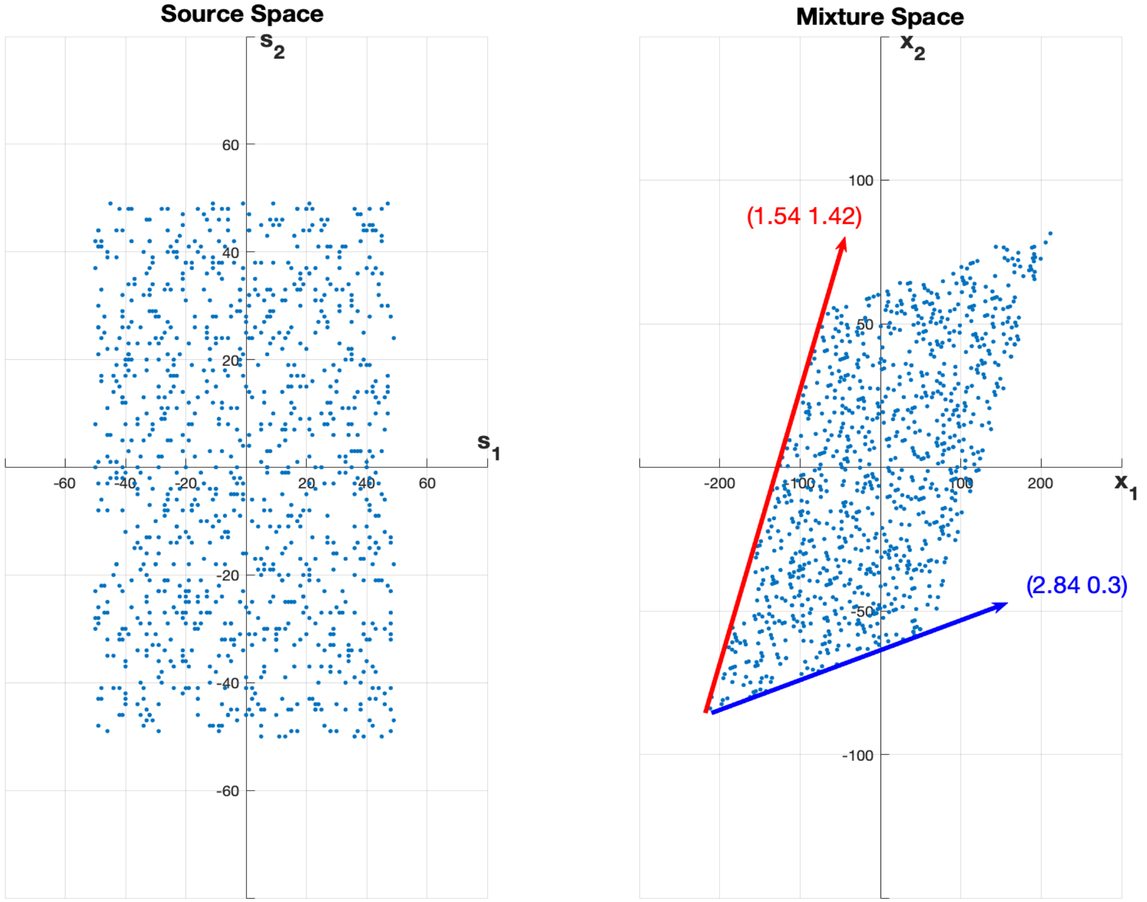


Mixing Matrix: A = $\begin{pmatrix} 1.54 & 2.84 \\ 1.42 & 0.3 \end{pmatrix}$

$$s'_1 = As_1 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ c \end{pmatrix} = \begin{pmatrix} 1.54 \\ 1.42 \end{pmatrix}$$

$$s'_2 = As_2 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} b \\ d \end{pmatrix} = \begin{pmatrix} 2.84 \\ 0.3 \end{pmatrix}$$

ICA: Simple Illustration



The edges of the parallelogram have an orientation that is in the same direction as the columns of the mixing matrix, A.

Does this mean that we could, in principle, estimate the ICA model by estimating the joint density of the mixtures and locating the edges of the parallelogram?

ICA: Simple Illustration

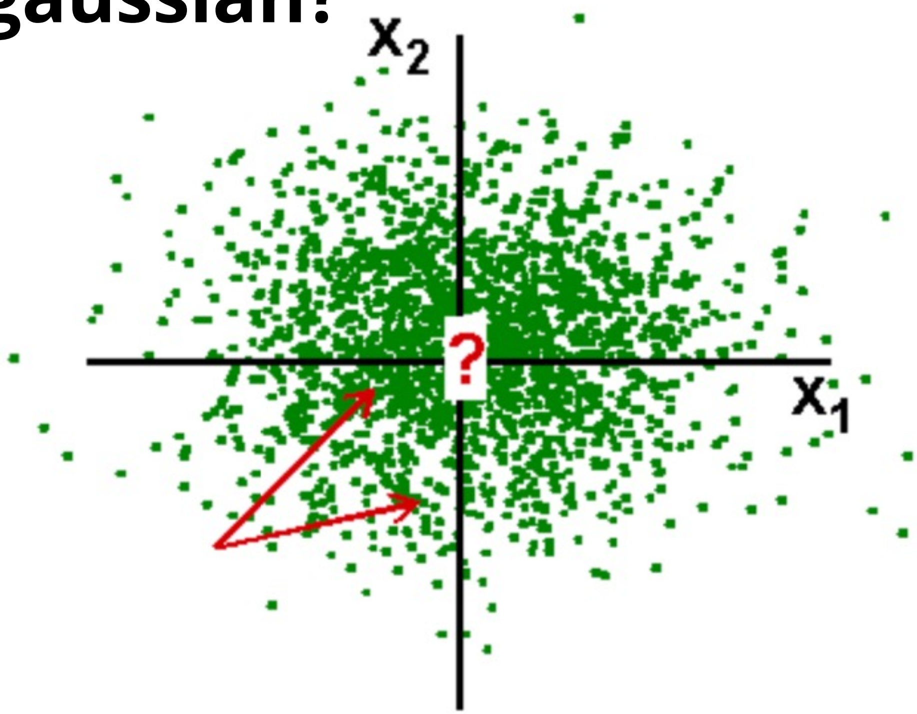
Unfortunately no. This only works for sources with uniform density.

We need a method that will work for sources with different densities…

But, it does demonstrate that each latent variable has its own orientation…

DISTRIBUTIONS

Why won't ICA work if our source signals are gaussian?



The density of the mixture is completely symmetric…

No information on the possible orientations of the columns of the mixing matrix, **A**.

THE UNMIXING PROBLEM

The goal of ICA is to find an unmixing matrix, W, that is a linear mapping between source signals, S, and our mixture signals, X:

$$y_i = Wx_i = WAs_i$$
$$y_i \approx s_i$$

Where y_i are maximally **statistically independent** and **non-Gaussian**.

Institute of language, communication and the brain ILCB

PROPERTIES OF SOURCE SIGNALS:
Measuring non-Gaussianity

Kurtosis:

$$kurt(y) = \boldsymbol{E(y)}^4 - 3(E(y^2)^2)$$

where y has zero mean and is normalised

A measure of the amount of probability in the tails of a probability density function.

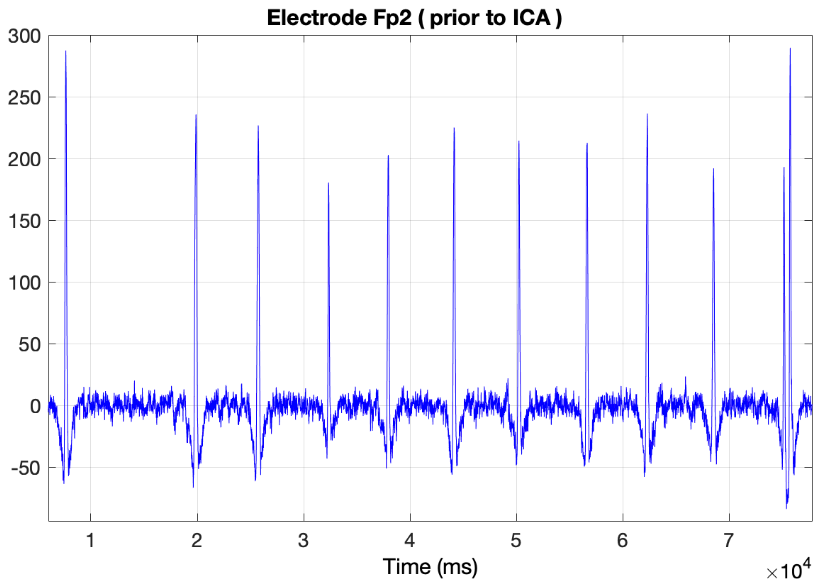
Kurtosis of a Gaussian distribution = 3

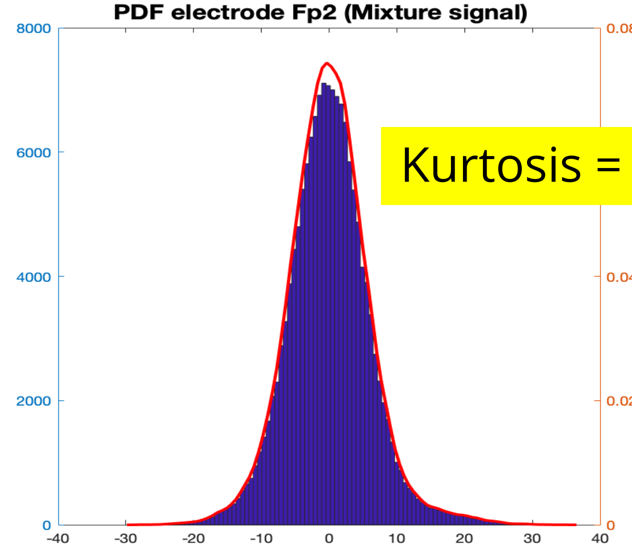
Note: Excess kurtosis of a Gaussian distribution = 0 (kurtosis -3)

If kurtosis < 3 (or negative) ➔ lighter tails in PDF.

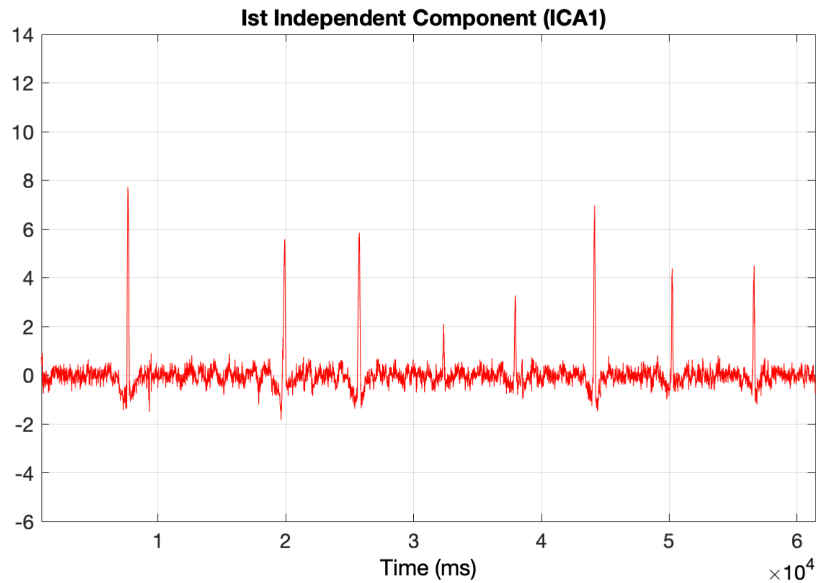
If kurtosis >3 (or positive) ➔ heavier tails in the PDF.

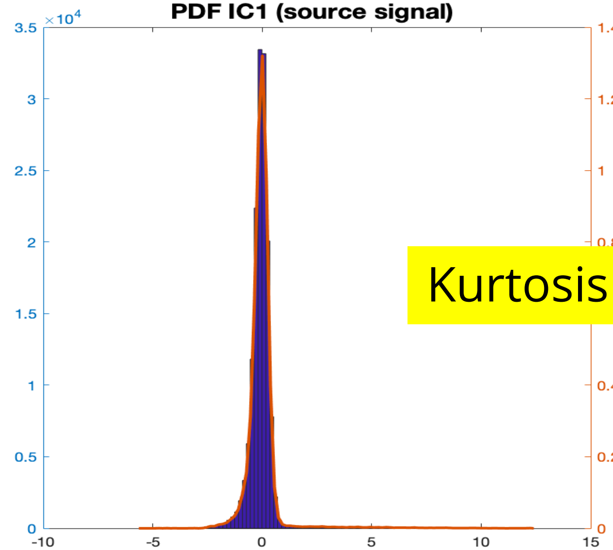
PROPERTIES OF SOURCE SIGNALS:
Measuring non-Gaussianity





Kurtosis = 4.52





Kurtosis = 58.03

PROPERTIES OF SOURCE SIGNALS:
Measuring non-Gaussianity with **Entropy**

Entropy is a measure of **uncertainty**.

The **entropy** of a random variable can be interpreted as the surprisal or the degree of information that the observation of the variable gives. The more unpredictable (random) the variable is, the larger the entropy value.

The entropy of variable is the log of the number, m, of equally probable outcome values:

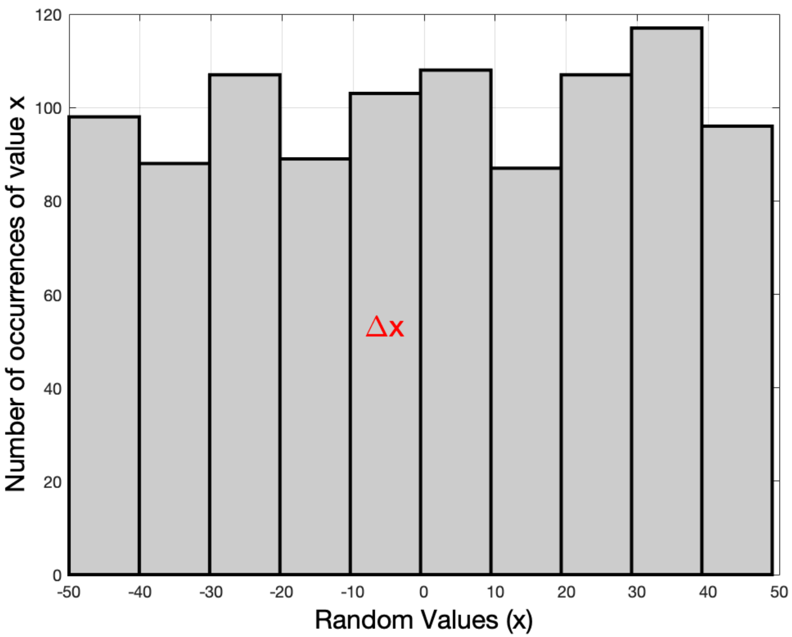
$$H(X) = log_2 m \text{ (bits)}$$

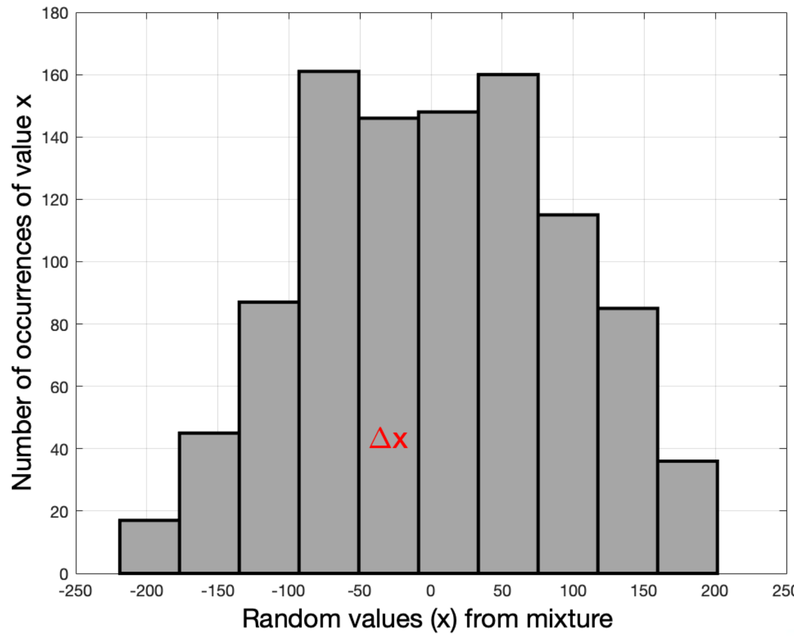
For a probability density, *p(x)*, the average suprise or entropy, H, is:

$$H(X) = \sum_{i=1}^{m} p(x_i) log \frac{1}{p(x_i)}$$

PROPERTIES OF SOURCE SIGNALS:
Measuring non-Gaussianity with **Entropy**

What type of distribution would give greater average entropy?





But this is for discrete random variables (coin tosses)…

For continuous variables – Gaussian distribution yields maximum entropy…

PROPERTIES OF SOURCE SIGNALS: **Negative Entropy**

<u>**Negative Entropy** or **Negentropy**</u>

For continuous random variables…

A measure of entropy relative to the entropy of a Gaussian.

$$J(y) = H(y_G) - H(y)$$

Where $H(y_G)$ is the entropy of a gaussian and y_G and y have the same variance.

$J(y) = 0$ for Gaussian distribution

An approximation of negentropy used in the FastICA algorithm (Hyvärinen, A. 1995a).

PROPERTIES OF SOURCE SIGNALS: **Mutual Information**

The mutual information, *I*, between two variables X and Y ($I(X,Y)$)…

The average reduction in uncertainty about the value of X provided by the value of Y, visa versa.

$$\mathrm{I}(X,Y) = \overbrace{H(X)}^{\text{entropy}} + H(Y) - \overbrace{H(X,Y)}^{\text{joint entropy}}$$

Note: If X and Y are independent: $H(X,Y) = H(X) + H(Y)$

➔Minimize $I(X,Y)$ to increase independence…

Minimization of mutual information used in algorithm (Bell & Sejnowski, 1995a, 1995b).

The Search for the Unmixing Matrix: Gradient Ascent

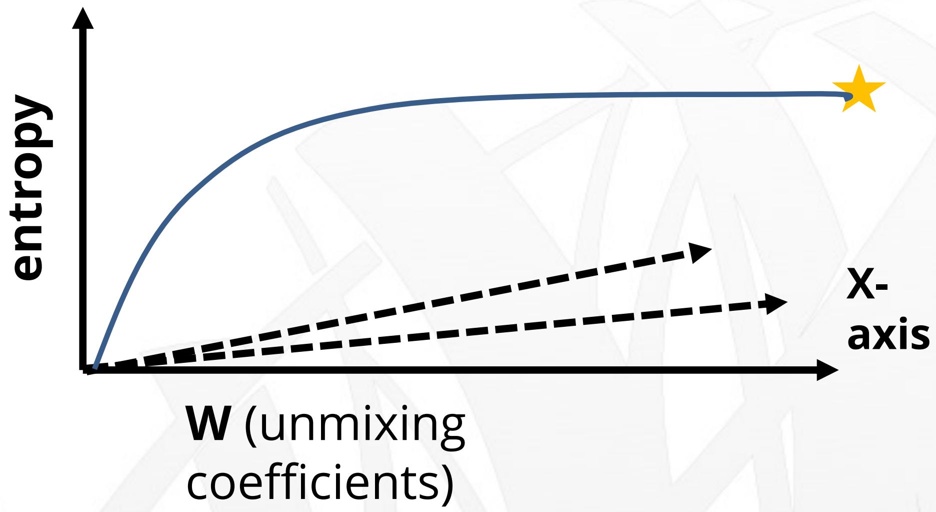
So…how do we find the optimal W that will maximise our measure of independence?

Example: maximise entropy …

The Search for the Unmixing Matrix: Gradient Ascent

So…how do we find the optimal W that will maximise our measure of independence?

Example: maximise entropy …



Change W in which direction to increase entropy???

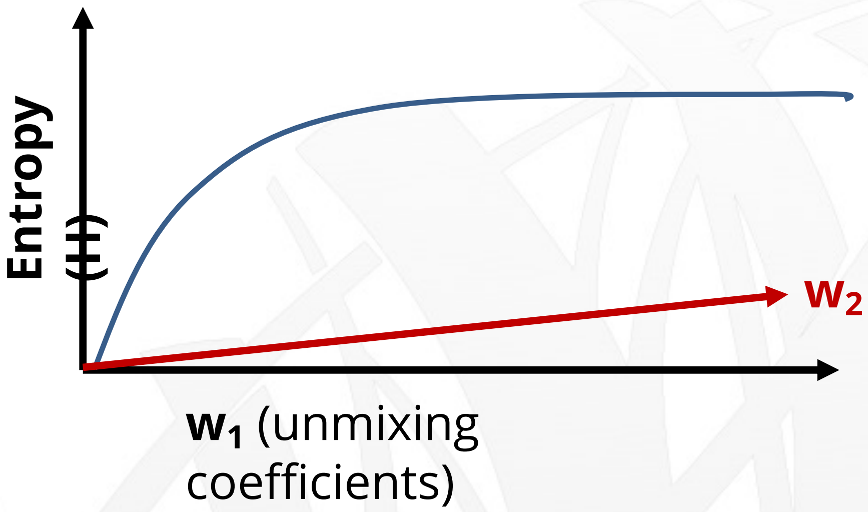
Choose a direction with steepest slope

But, we need to find a way of estimating the direction with the steepest ascent…

The Search for the Unmixing Matrix: Gradient Ascent

So…how do we find the optimal W that will maximise our measure of independence?

Example: maximise entropy …



Looking for the direction that will give the greatest increase in entropy…

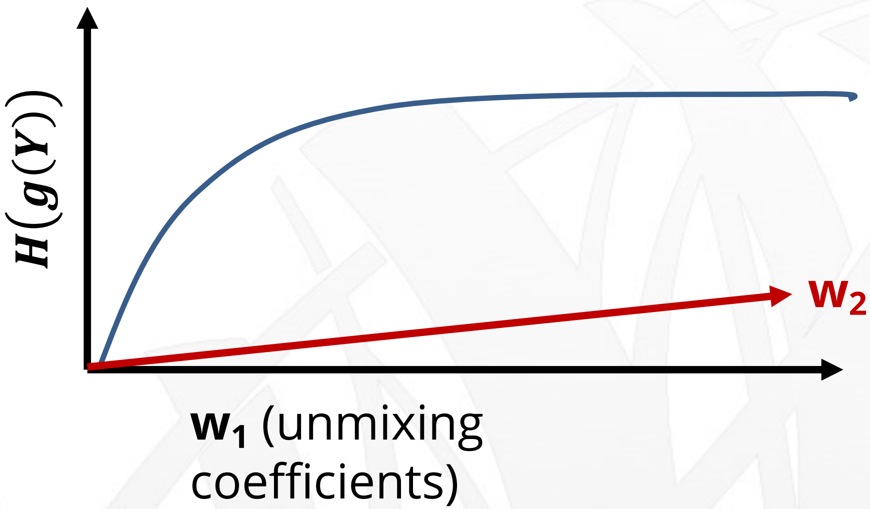
$$\frac{dH}{dw_i} = p(y_i) ln \frac{1}{p(y_i)}$$

Where $y_i = w_i x_i$.

The Search for the Unmixing Matrix: Gradient Ascent

So…how do we find the optimal W that will maximise our measure of independence?

Example: maximise entropy …



Looking for the direction that will give the greatest increase in entropy…

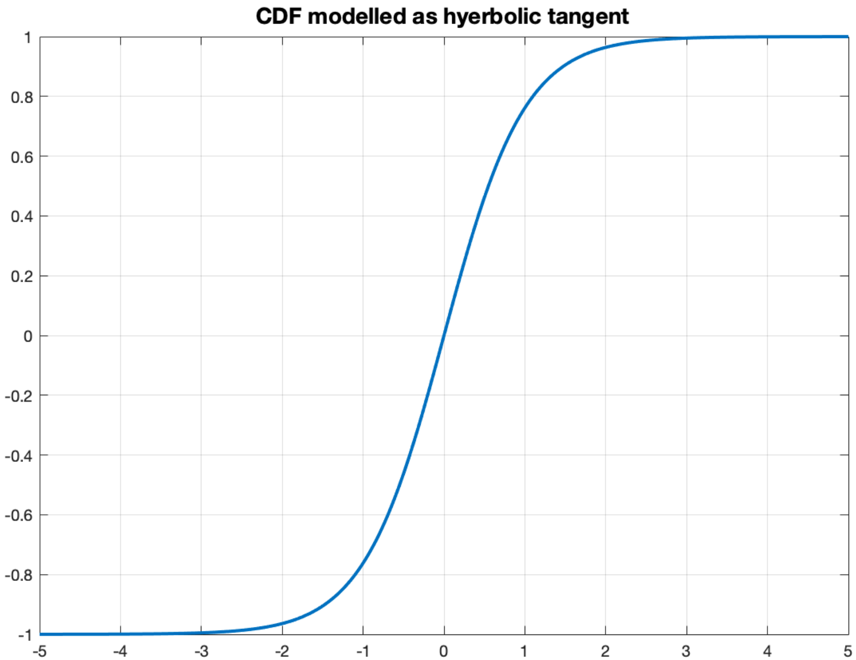
$$\frac{dH}{dw_i} = p(y_i) ln \frac{1}{p(y_i)}$$

Where $y_i = w_i x_i$.

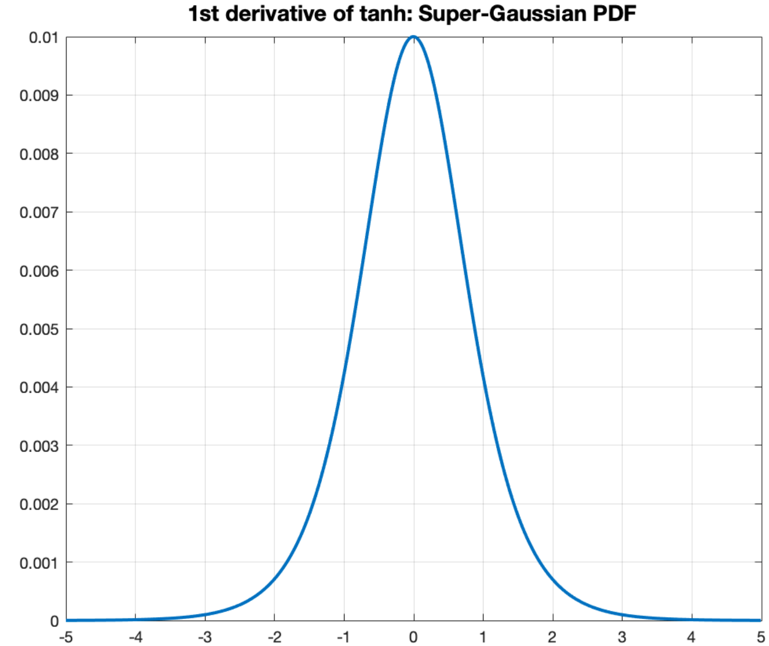
We can rename the y-axis: entropy as a function of the PDF of y_i or $\boldsymbol{H(p(Y))}$

But, generally the **cumulative density function** (CDF) is used: CDF of y_i or g(Y)

The Search for the Unmixing Matrix



CDF used a model for super-Gaussian signals. (Infomax)



Corresponding model PDF with super-Gaussian distribution.

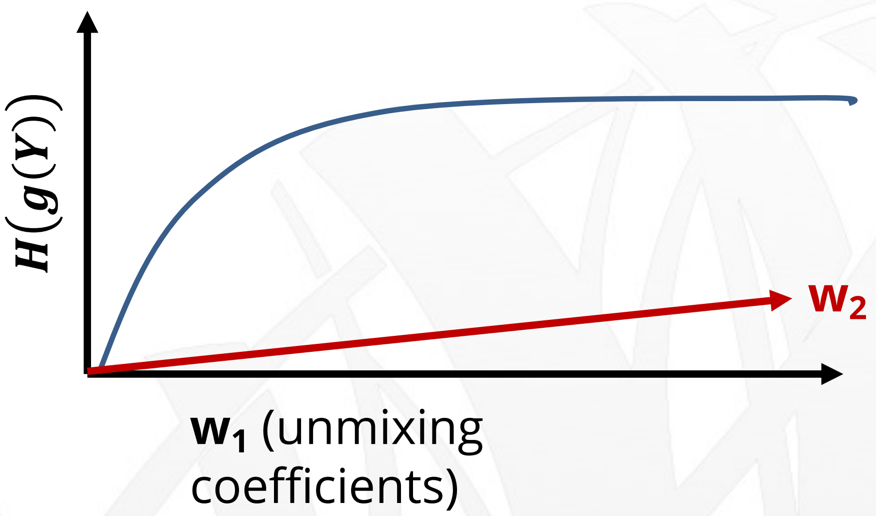
Assumption: An extracted signal, y_i , has maximum entropy if its cdf matches g .

➔ ICA as a method for extracting signals whose distributions match a specific PDF

The Search for the Unmixing Matrix: Gradient Ascent

So…how do we find the optimal W that will maximise our measure of independence?

Example: maximise entropy …



Looking for the direction that will give the greatest increase in entropy…

$$\frac{dH(g(y_i))}{dw_i} = p(y_i)ln\frac{1}{p(y_i)}$$

Where $y_i = w_i x_i$.

Update w_i in proportion to the gradient, $\frac{dH(g(y_i))}{dw_i}$.

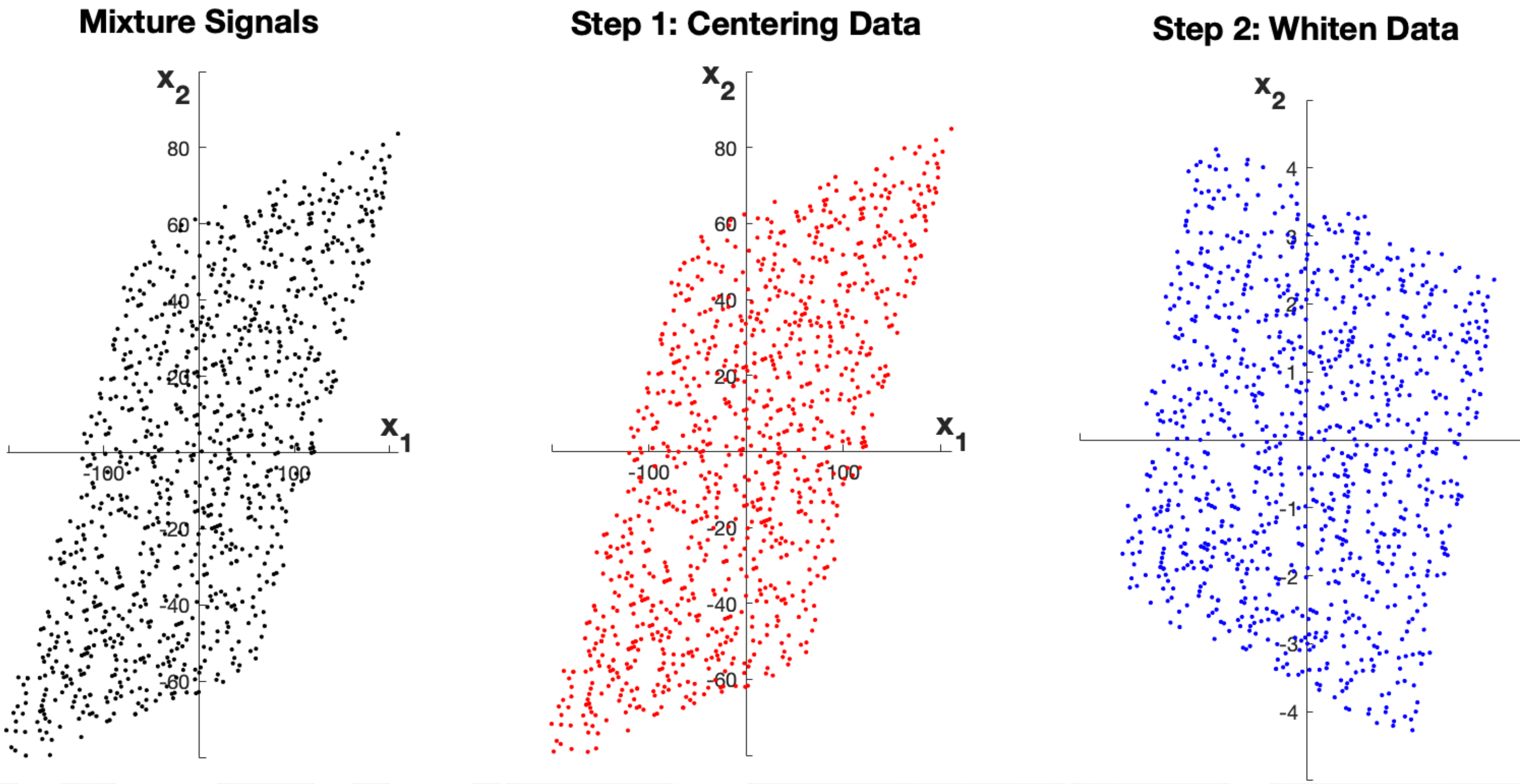
➔ Increment constant, η, needs to small (i.e. $^1/_{1000}$)

TO RECAP…

- The problem becomes one of searching for an unmixing matrix, **W**, that optimizes a measure of independence:
    - **Kurtosis**
    - Maximization of entropy or **negentropy**
    - Minimisation of **mutual independence**.
- The measures are based on our assumption about the sources and their probability densities:
    - Independent
    - Non-Gaussian
- Different ICA algorithms use different approaches to this optimization:
    - Gradient ascent…

ICA : MAIN STEPS

PRE-PROCESSING STEP

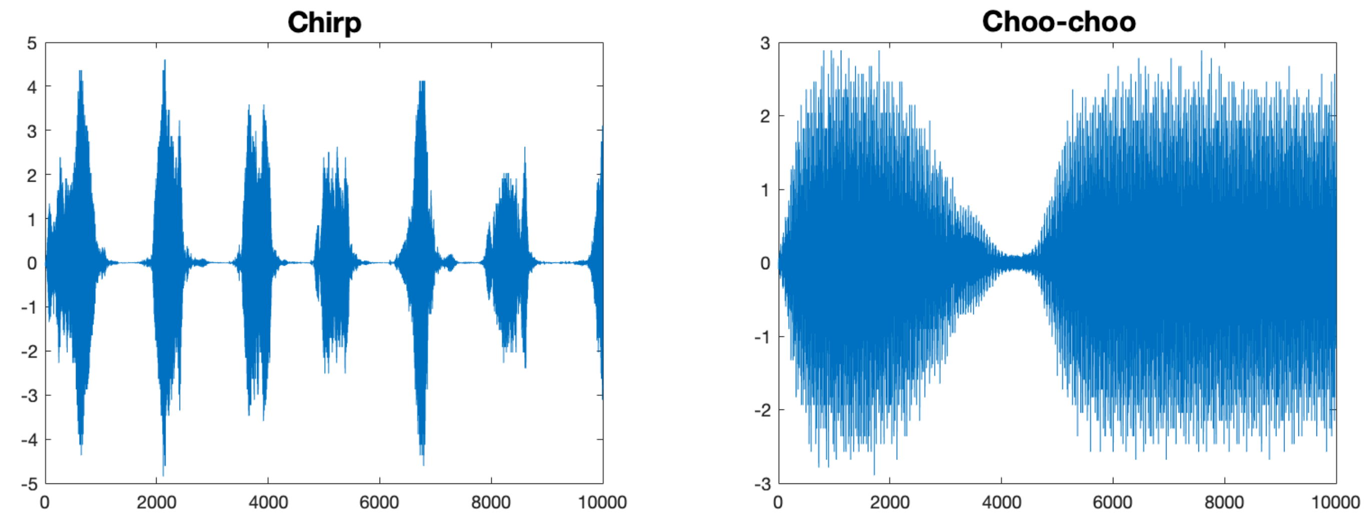


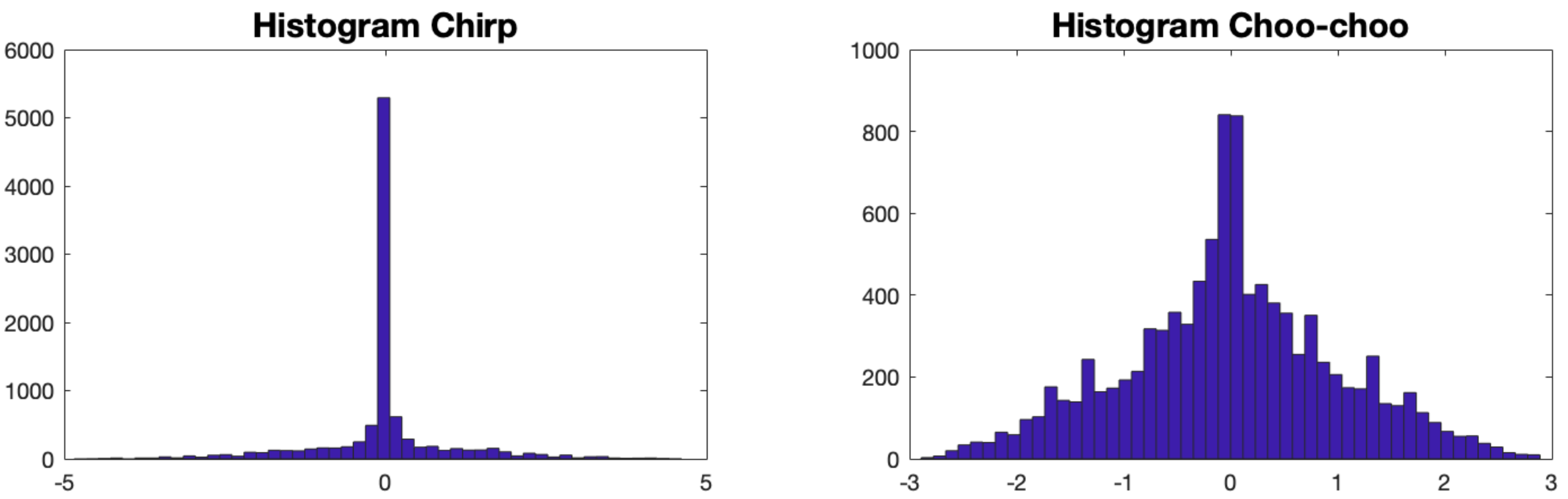
Subtract the mean

Decorrelate the signals and scale to unit variance

Institute of language, communication and the brain

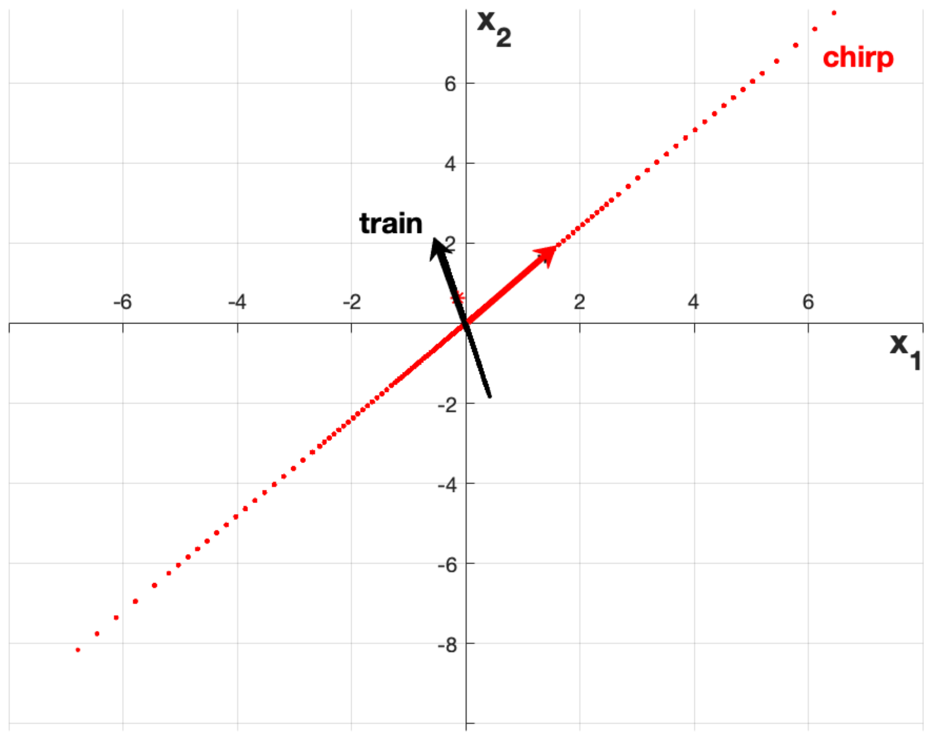
ICA EXAMPLE

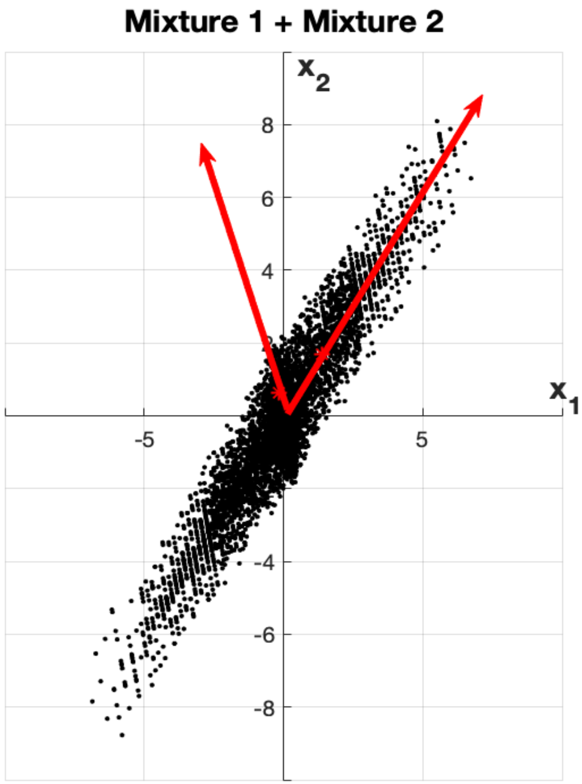




ICA EXAMPLE

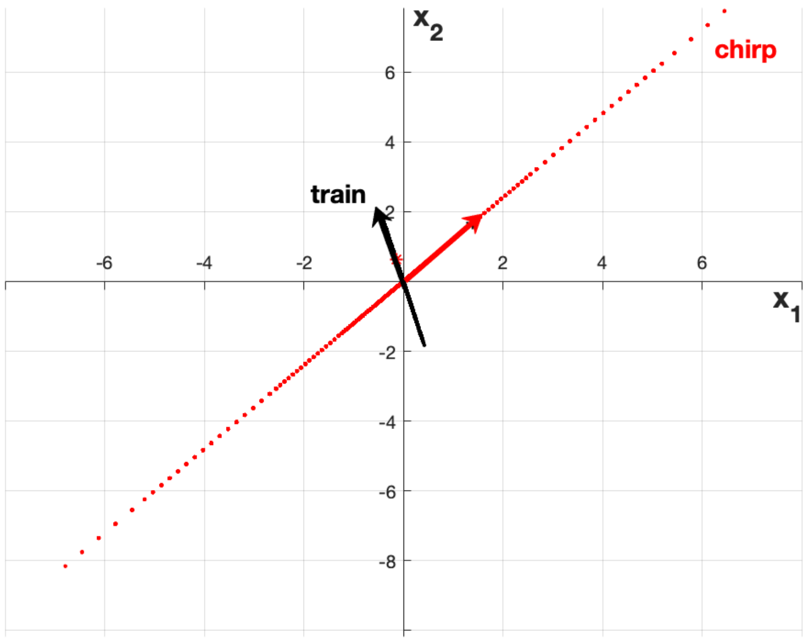
$S \rightarrow \mathbf{X}$: $X = AS$

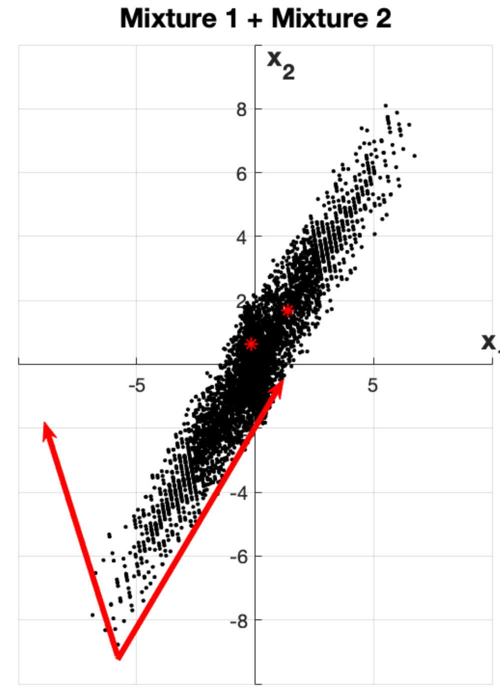




ICA EXAMPLE

$S \rightarrow \mathbf{X}$: $X = AS$





Mixture 1 ($\mathbf{x}_1$)

Mixture 2 ($\mathbf{x}_2$)

ICA: QUICK SUMMARY OF STEPS

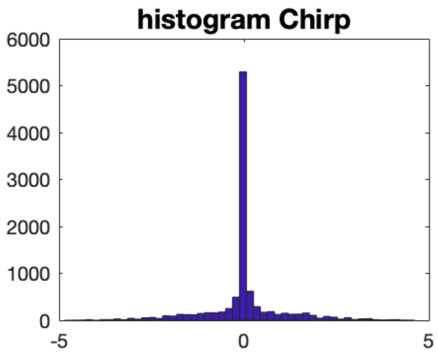
- Subtract off the mean of the mixture data in each dimension.

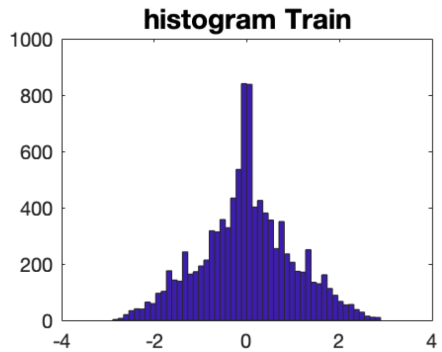
- Whiten the data by calculating the eigenvectors of the co-variance of the data .

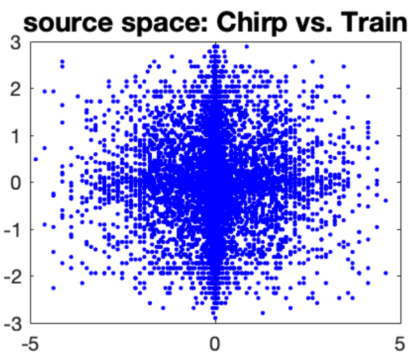
- Identify final rotation matrix that optimizes statistical inde-pendence .

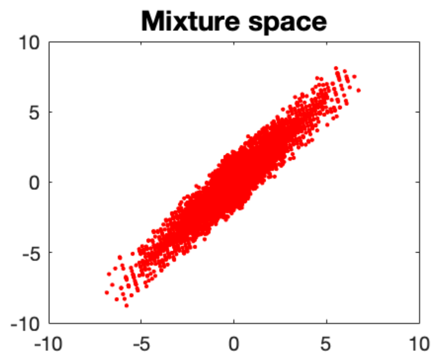
ILCB Institute of language, communication and the brain

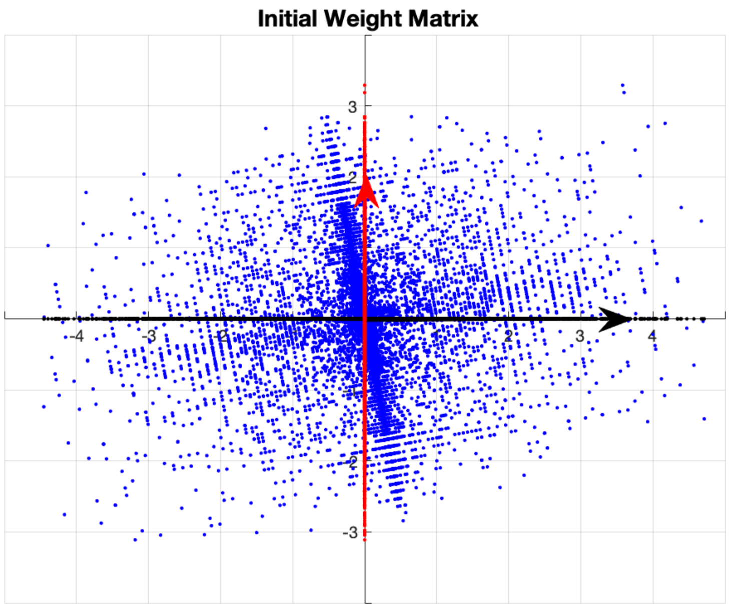
ICA : MAIN STEPS



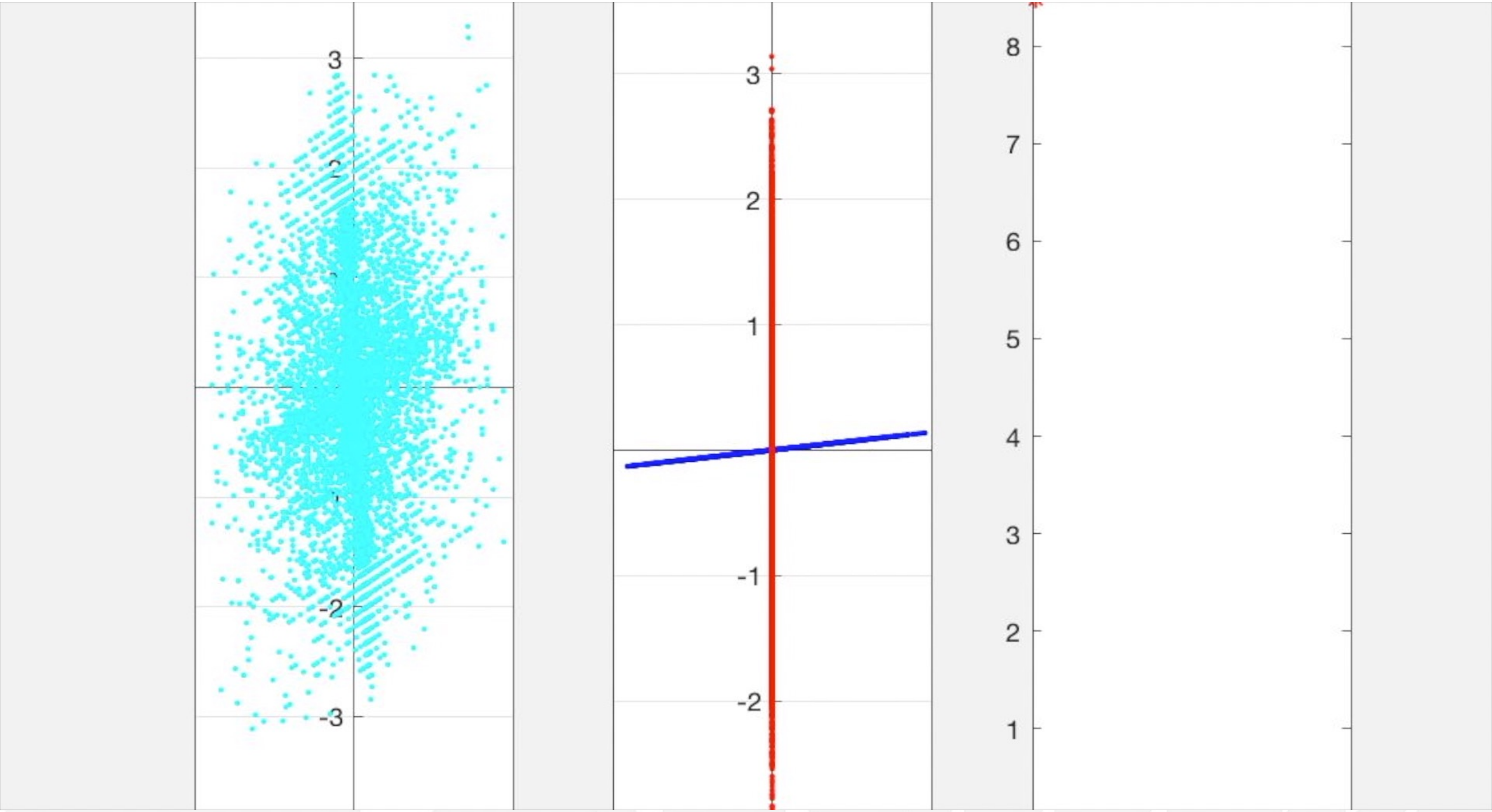




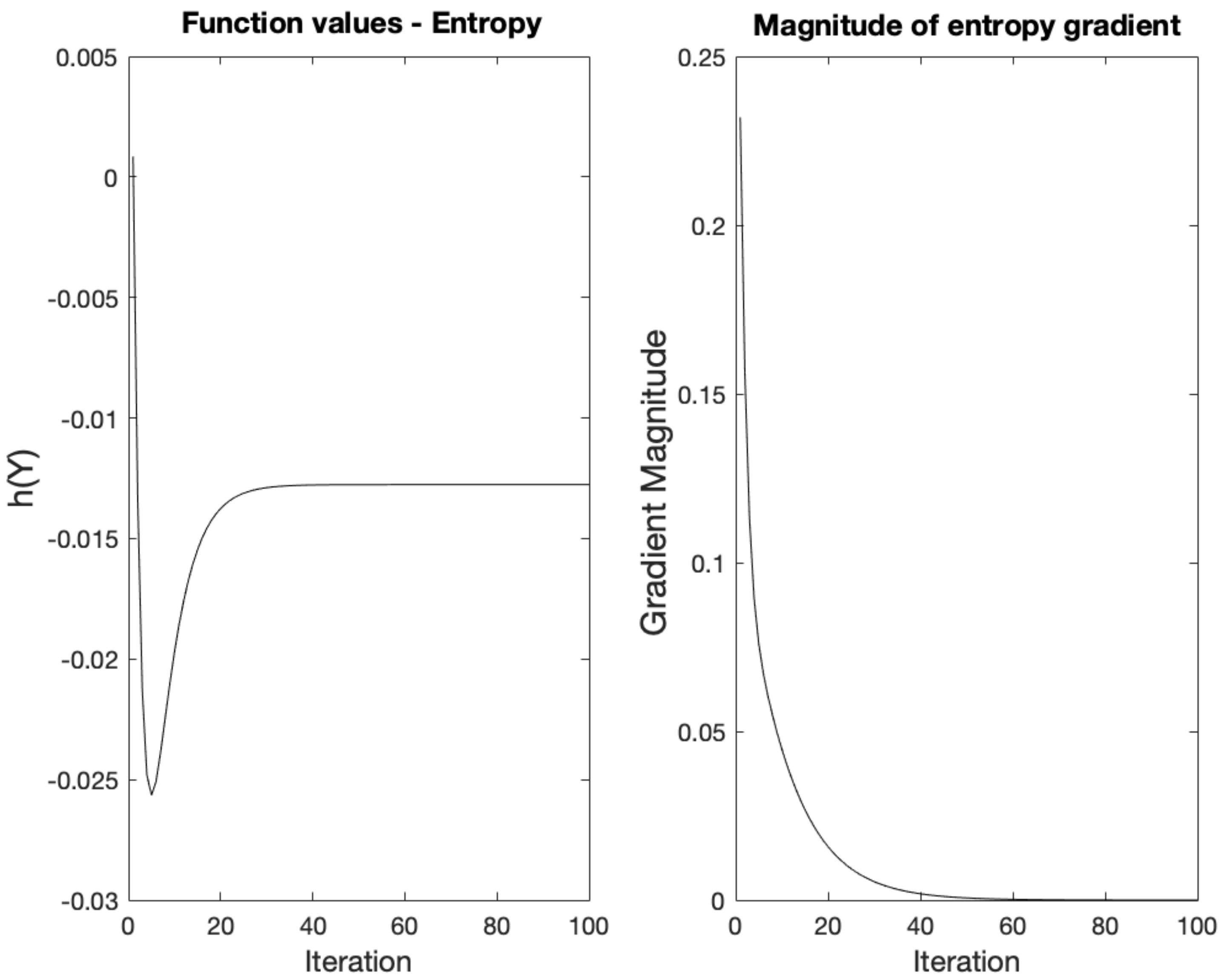




ICA : MAIN STEPS



ICA : MAIN STEPS



chirp

train

ICA: SOME AMBIGUITIES

- ICA extracts source signals but not in any specific order
- ICA extracts a scaled version of the source signals.
- The sign of the independence components is not important.

PRINCIPAL COMPONENTS ANALYSIS (PCA)

Another method for carrying out BSS.

It assumes that the observed data is a **linear** mixture of the underlying sources**.**

Can be used to seperate a set of signal mixtures into a set of **uncorrelated** signals call principal components (PC).

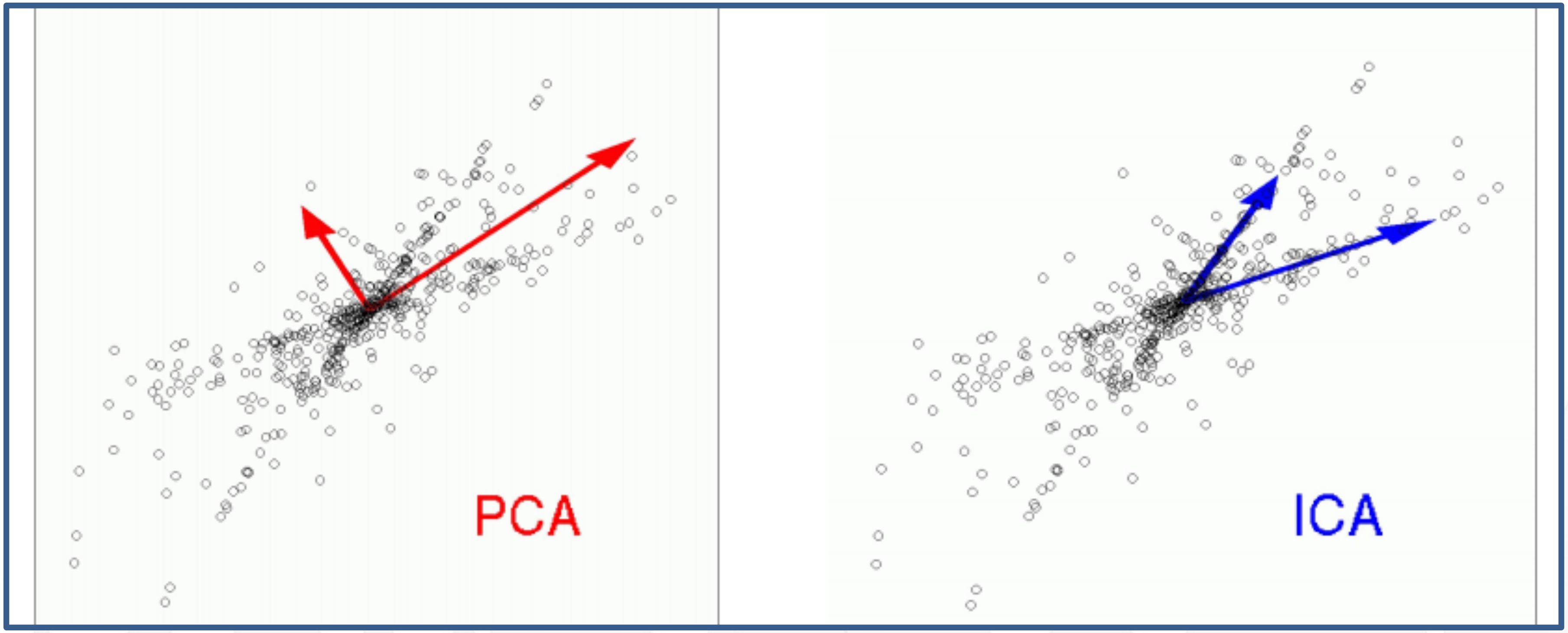
The signals that it extracts have a **Gaussian or normal** distribution.

It extracts signals that are **orthogonal**.

The extracted signals are ordered as a function of their variances. The 1st PC has maximum variance.

Institute of
language, communication
and the brain

PCA vs. ICA



Independence vs. Uncorrelated

Random variables X and Y are **uncorrelated** if: $corr(X,Y) = 0$

$corr(X,Y) = \dfrac{cov[X,Y]}{\sigma^X \sigma^Y}$, where $cov[X,Y]$ is covariance of X and Y, σ^X and σ^Y are standard deviation of X and Y, respectively.

Uncorrelated $\Longrightarrow covariance = 0$

If X and Y are linearly independent $\Longrightarrow$ X and Y are uncorrelated.

But if X and Y are uncorrelated $\not\Rightarrow$ X and Y are independent ...

TO RECAP…

What assumptions does ICA make about the source signals?

Can PCA extract source signals ?

What information can the mixture signals give us about the mixing process?

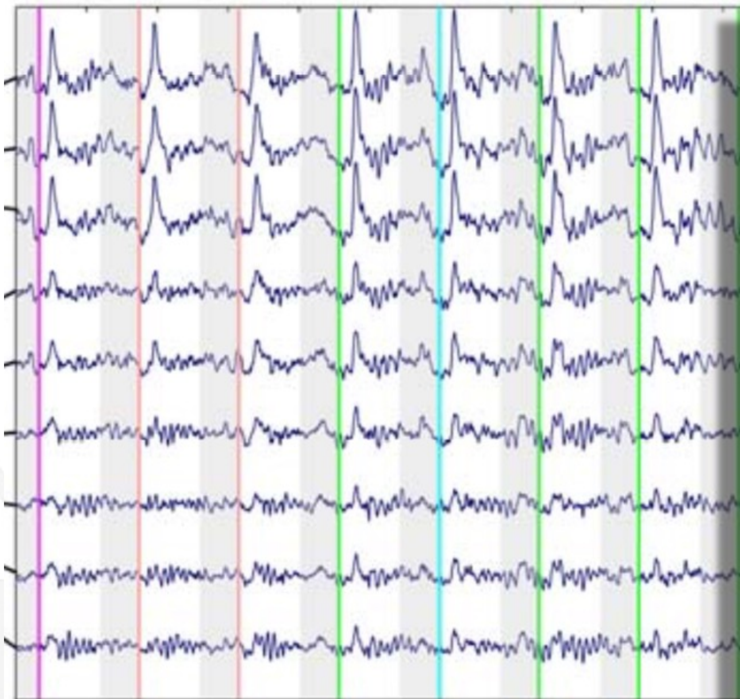
Independent Components Analysis in EEG Processing and Analysis

ICA in EEG Processing and Analysis

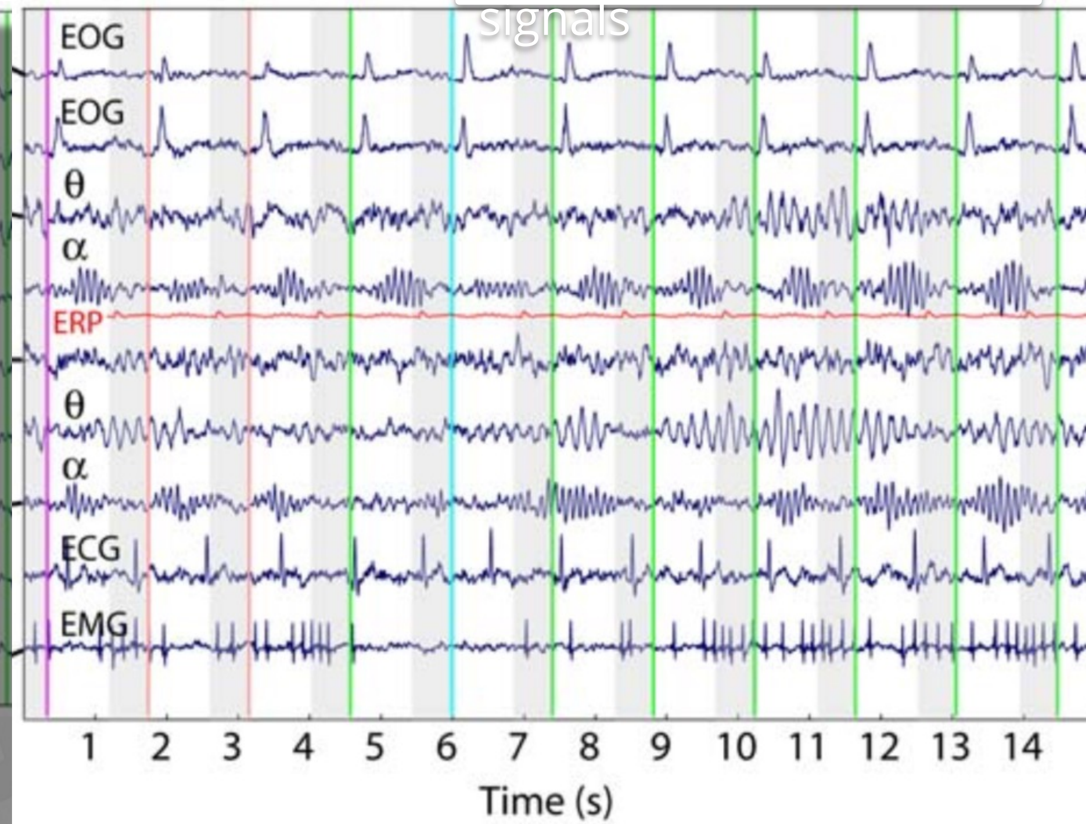
In EEG, what are our mixture signals?

In EEG, what are our source signals?

Sources: brain signals, non-brain artifact signals



Channel Data = mixtures



ICA in EEG Processing and Analysis

What assumptions does ICA make about the signals recorded at the scalp level?

That the signals are the linear mixtures of cortical and non-cortical artifact signals.

What assumptions does ICA make about the volume conduction?

That the volume conduction is **instantaneous** and **linear**.

That the cortical and non-cortical EEG sources are **temporally distinct** and **temporally independent**.

Physiologically plausible?

Institute of language, communication and the brain

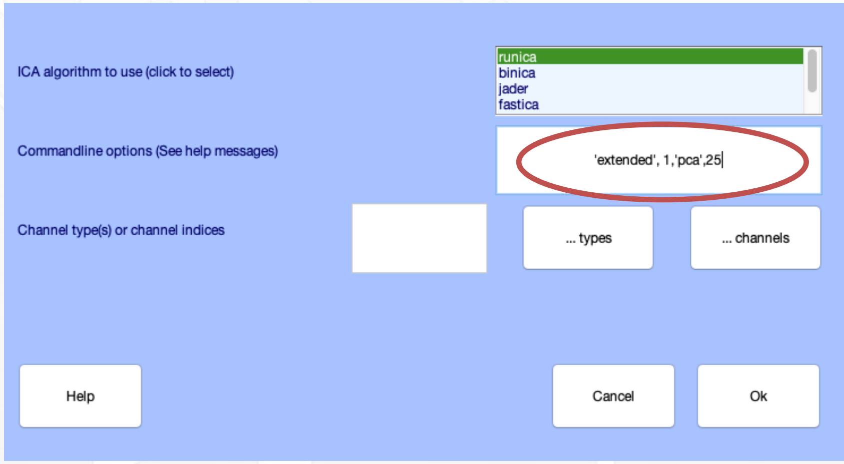
ICA in EEGLAB: runica()

- Applies the infomax ICA (Bell & Sejnowski, 1995).

- The *runica()* function alone extracts super-Gaussian signals (signals with transient structure such as eye-blinks, cardiac…).

- Option of « extended » ICA to extract **sub-Gaussian** signals (continuous signals such as line noise, alpha…)

- Possibility to define the number of Independent components « *ncomps* » in code but not advisable – use « PCA » instead.

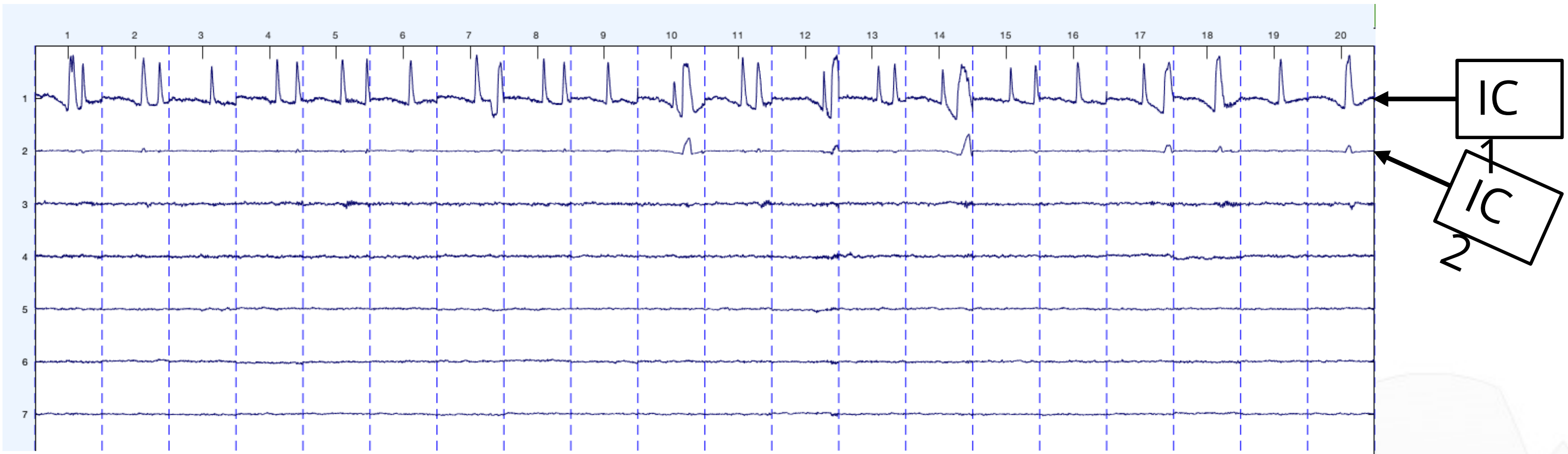


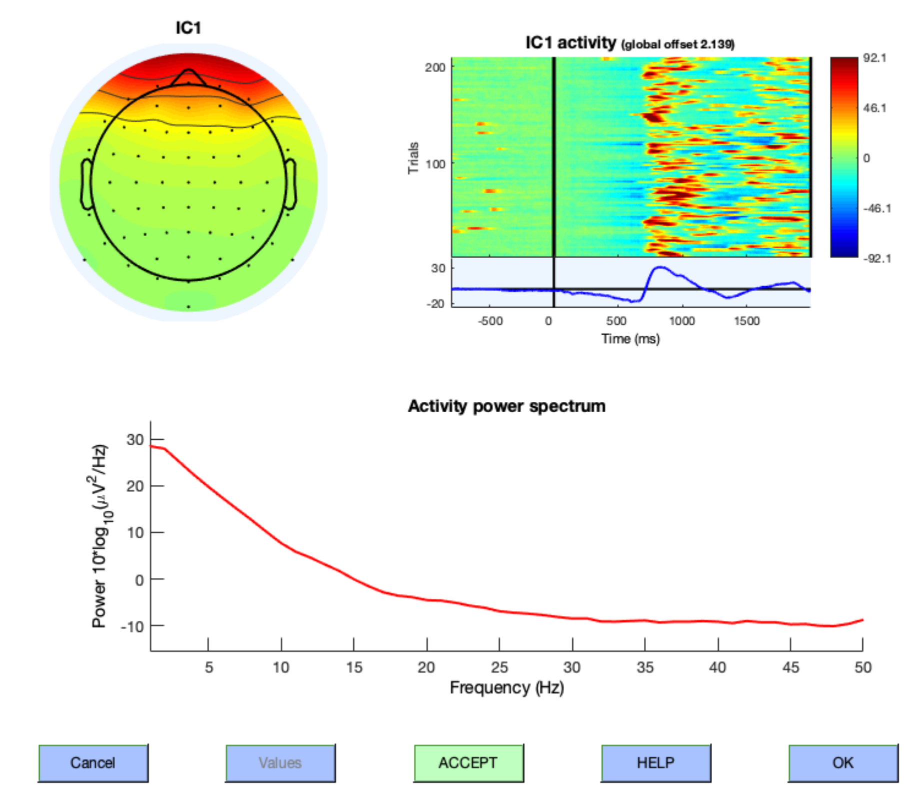
By default, *randomseed generator* is reset each time the function is run ➔ results will vary at each execution.

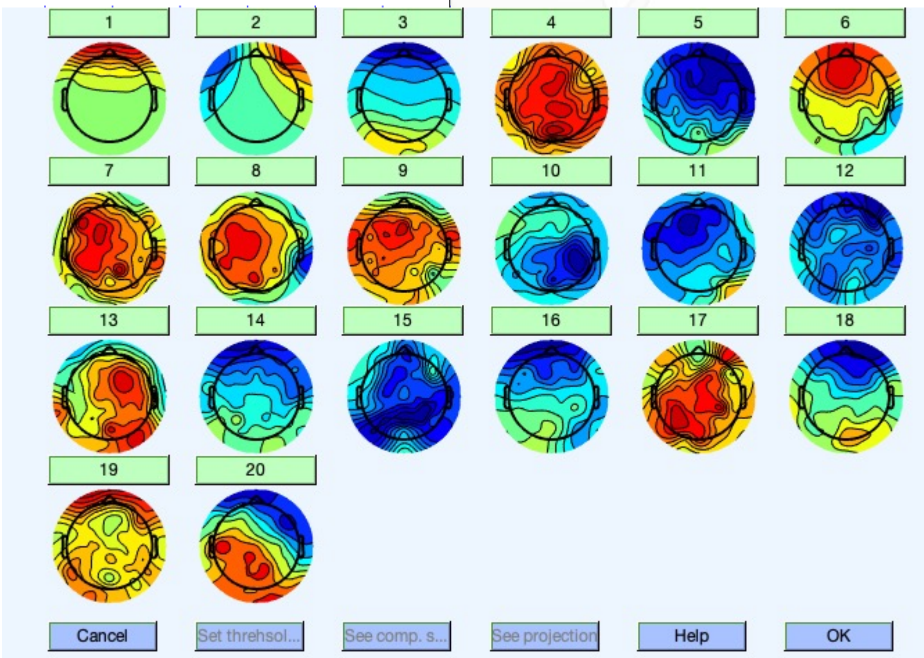
Possibility to change this using rand('seed',value)

```
if reset_randomseed
    rand('state',sum(100*clock));  % set the random number generator state to
end                                % a position dependent on the system clock
```

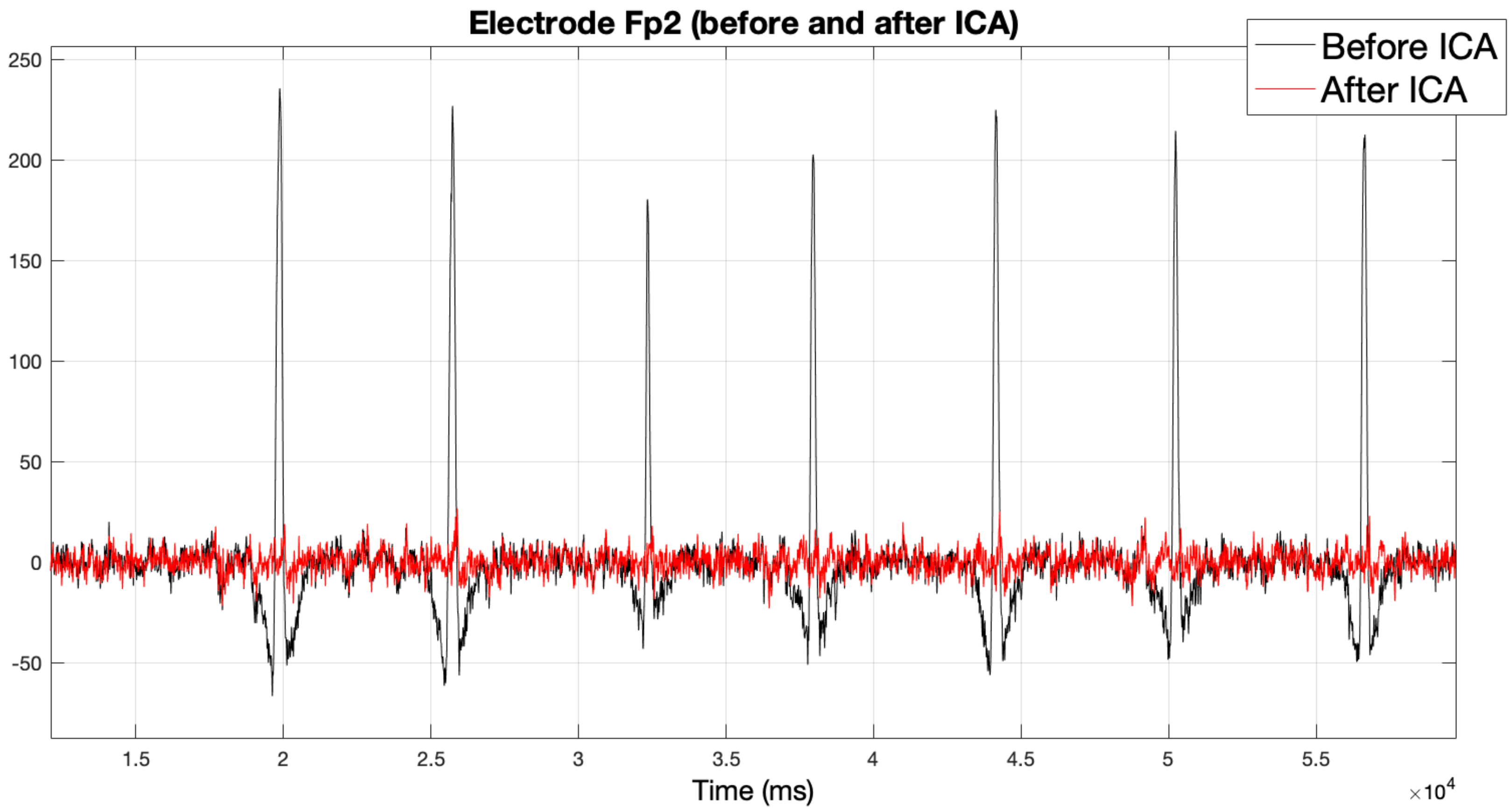
ICA in EEGLAB: So does it work??



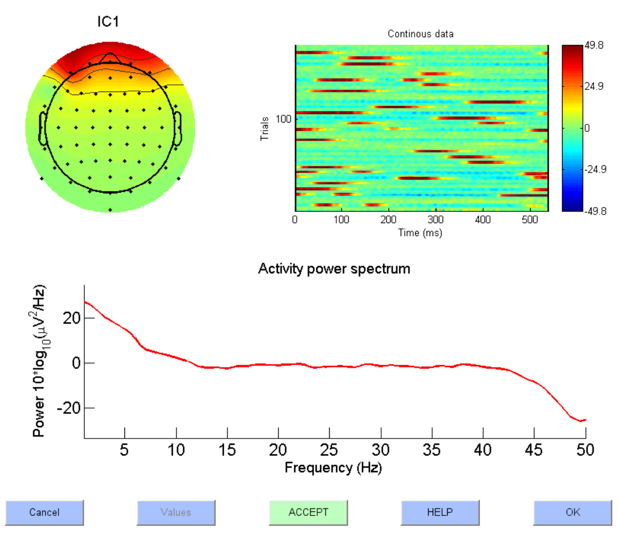


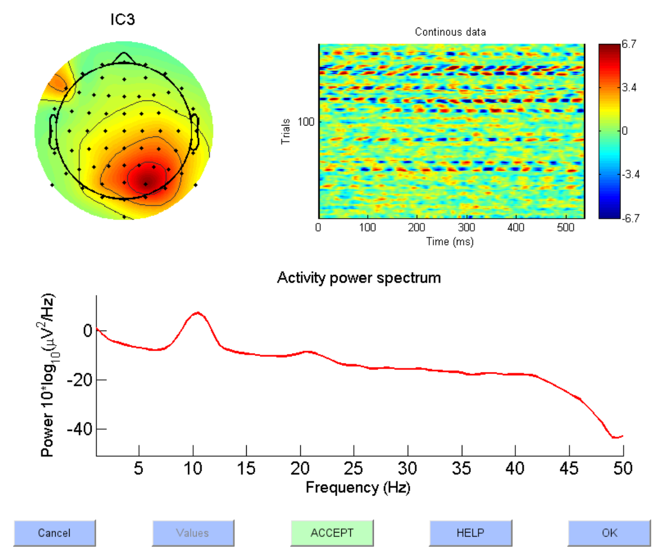


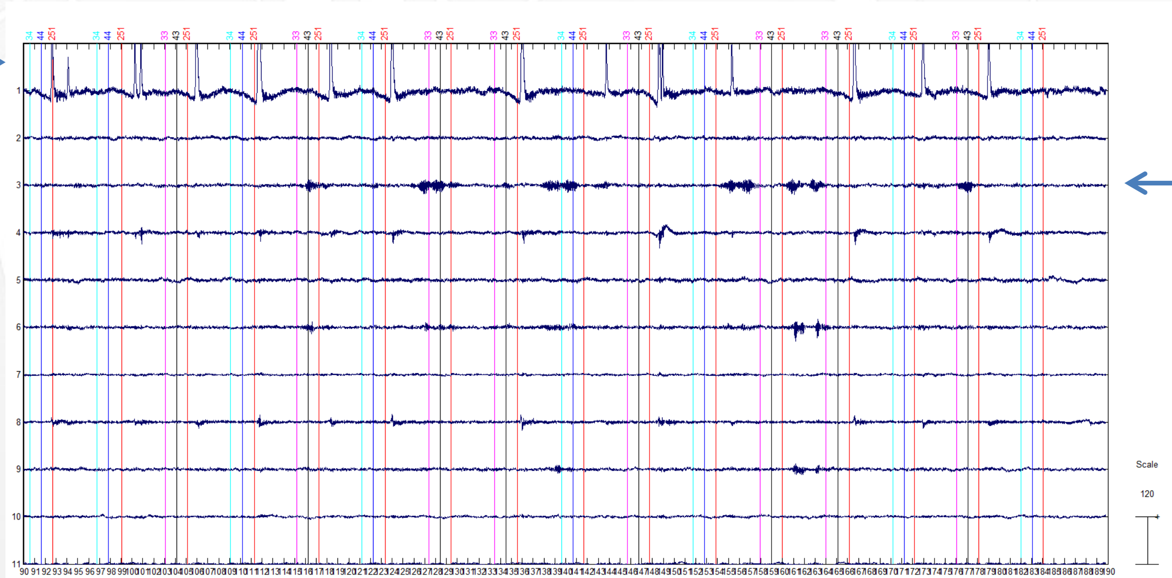
ICA in EEGLAB: So does it work??



ICA in EEGLAB: So does it work??







ICA in EEGLAB: Conditions

- Sufficient data…
- Do not perform on channels
  - that have been interpolated
  - corresponding to the reference channel
- No bridges formed between electrodes during EEG acquisition.

Reduces the **Rank** of the data

If in doubt…verify the rank using: `rank(EEG.data(channels,:))`

If carrying out on epoched data, data should be baseline normalised…

Is ICA the only way…

For the EEG and MEG other possible approaches:

Spatial Filtering methods such as

- **Signal Space Projection (SSP)**
- Signal Space Separation (SSS)
- Temporal Signal Space Separation (tSSS)

Signal Space Projection (SSP) also decomposes the signals into components.

But components do not need to statistically independent.

Thank you for your attention!

Institute of language, communication and the brain

INDEPENDENT COMPONENT ANALYSIS

Can be used to seperate a set of signal mixtures into a set of source signals called **independent components**.

It assumes that the source signals extracted are **statistically independent.**

It assumes that the observed data is a **linear** mixture of the underlying sources.

It assumes that the source signals have a **non-normal** or **non-Gaussian distribution.**

The unmixing matrix must square and **full rank…**