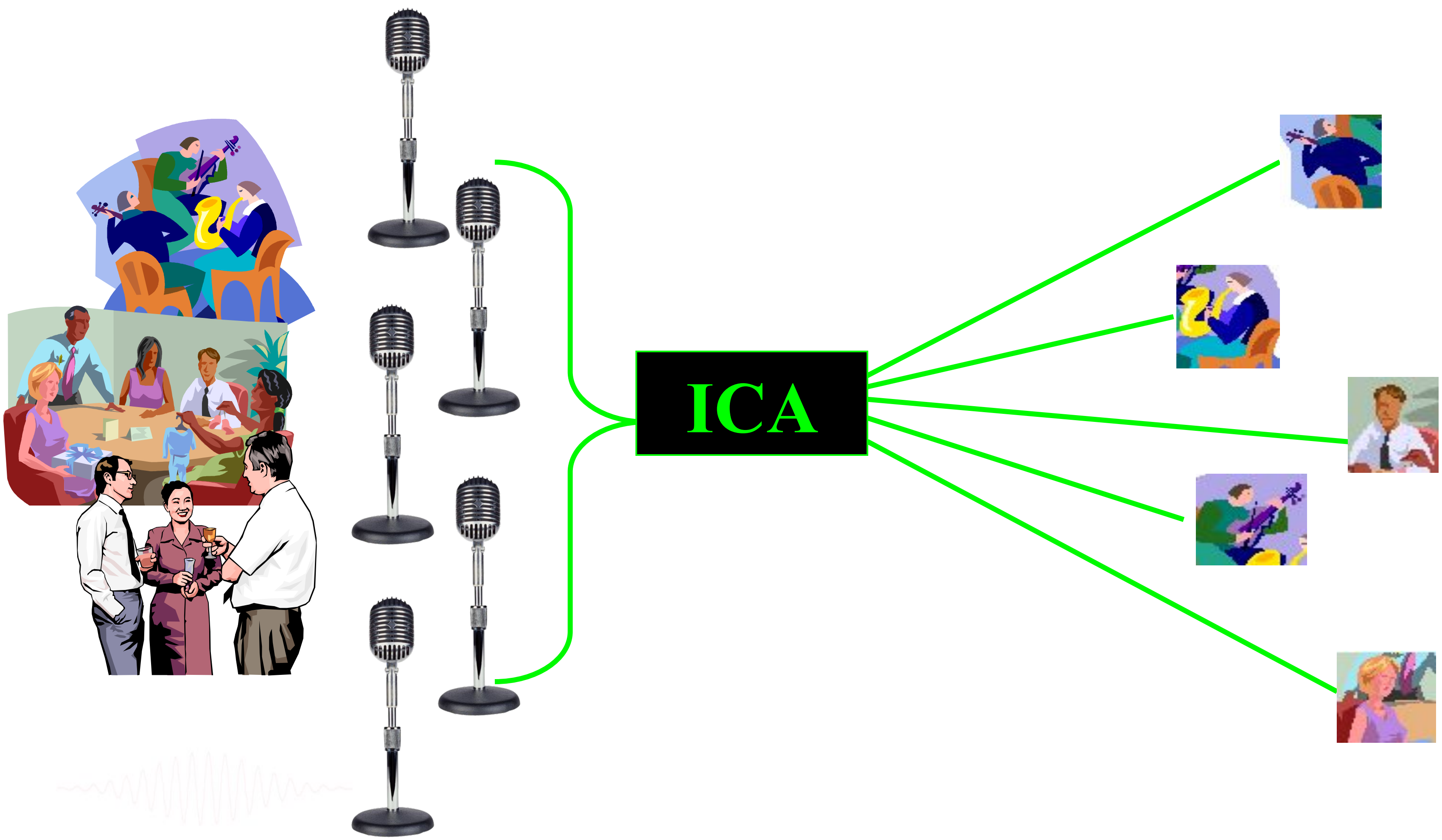
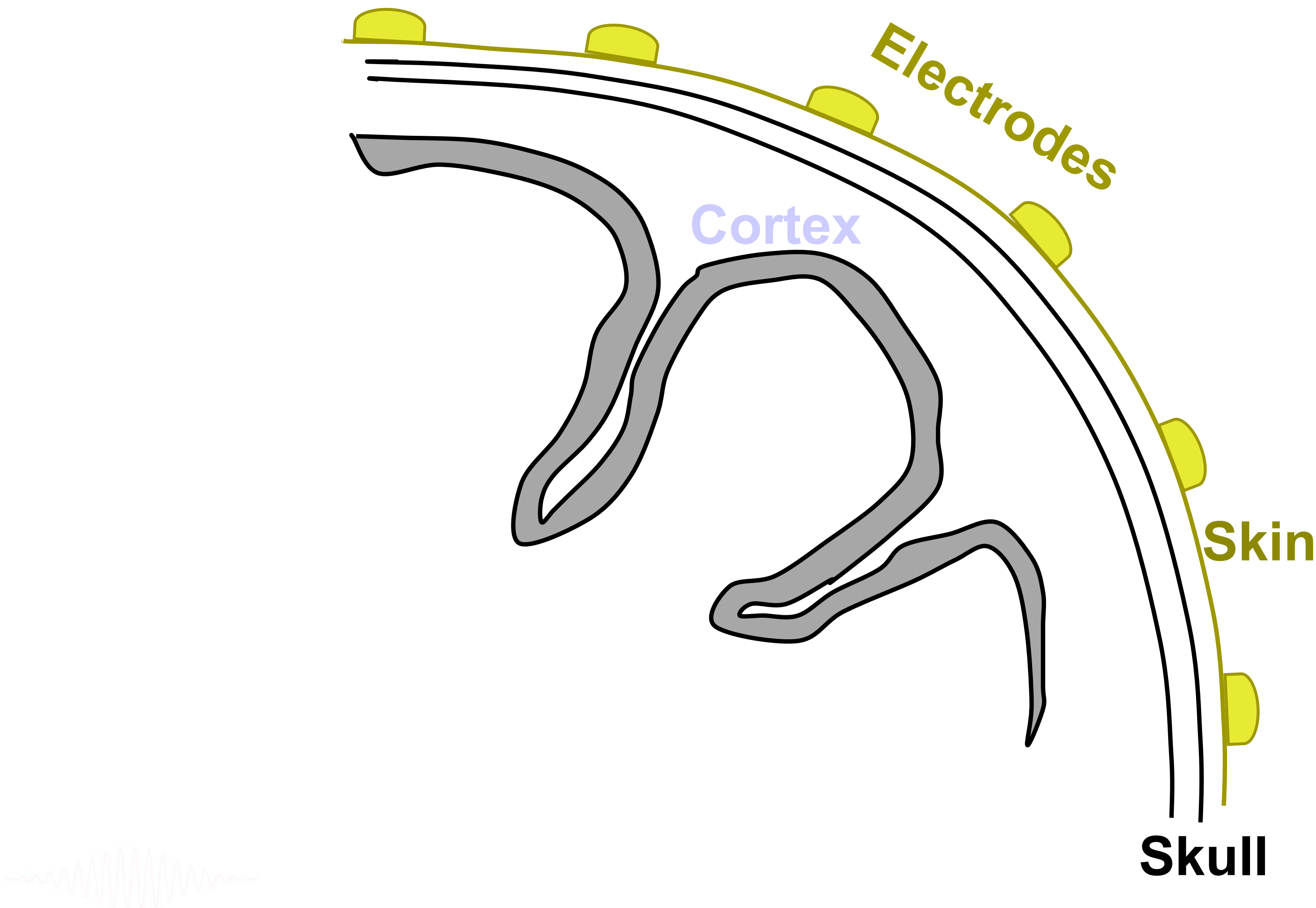
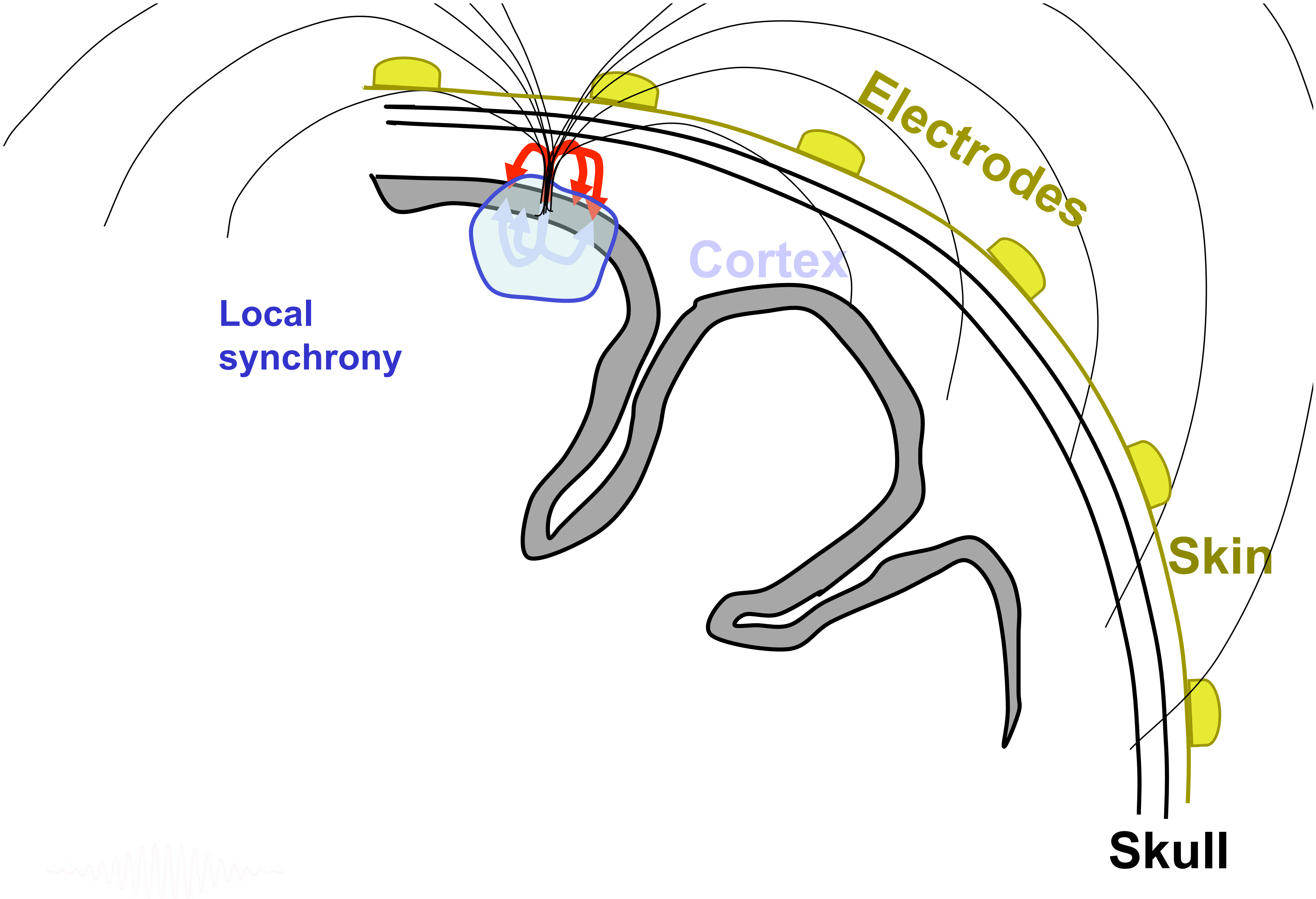
Independent component analysis applied to biophysical time series and EEG

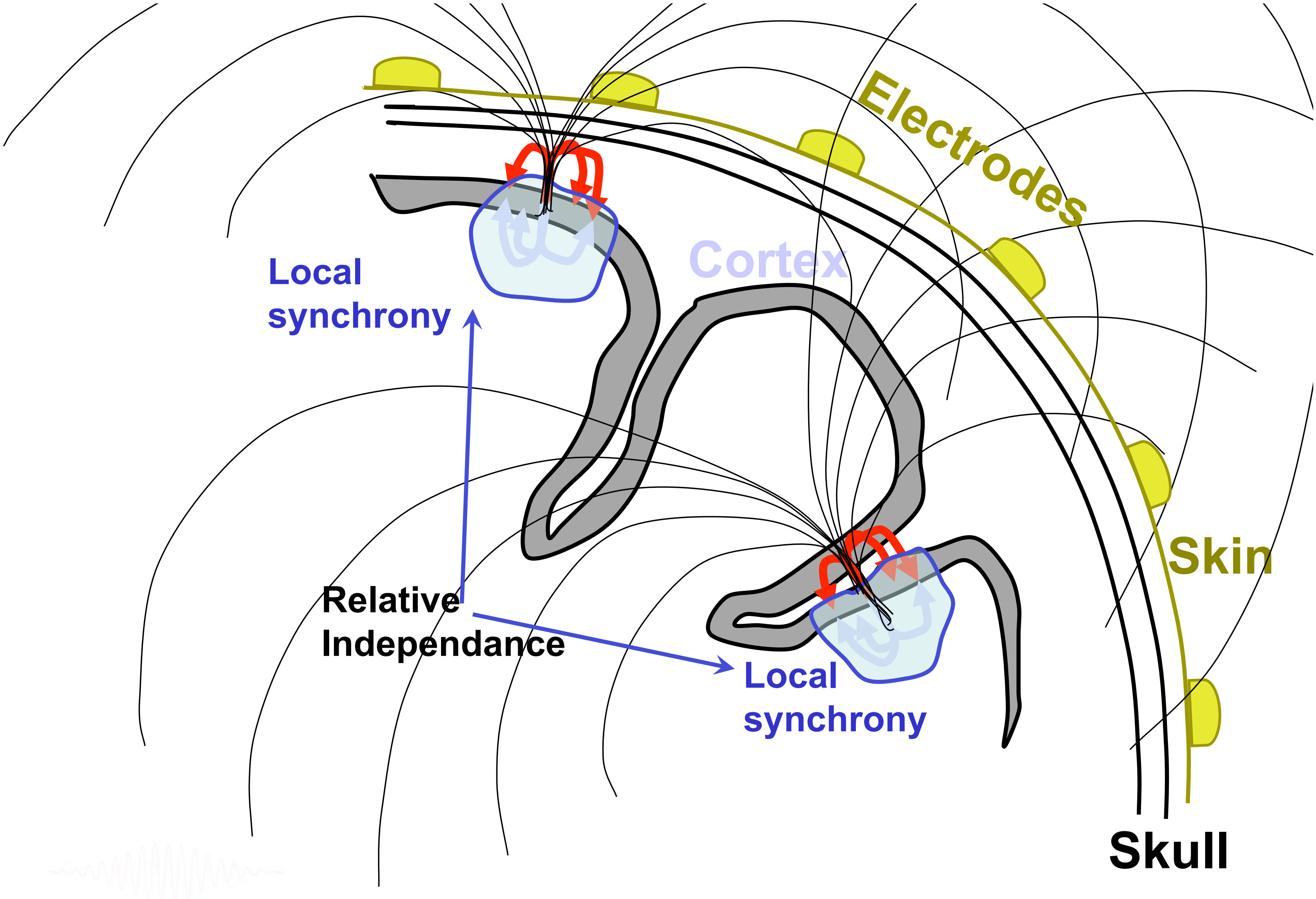
EEGLAB Workshop, Rio De Janeiro, 2014, Arnaud Delorme

Example: Speech Separation

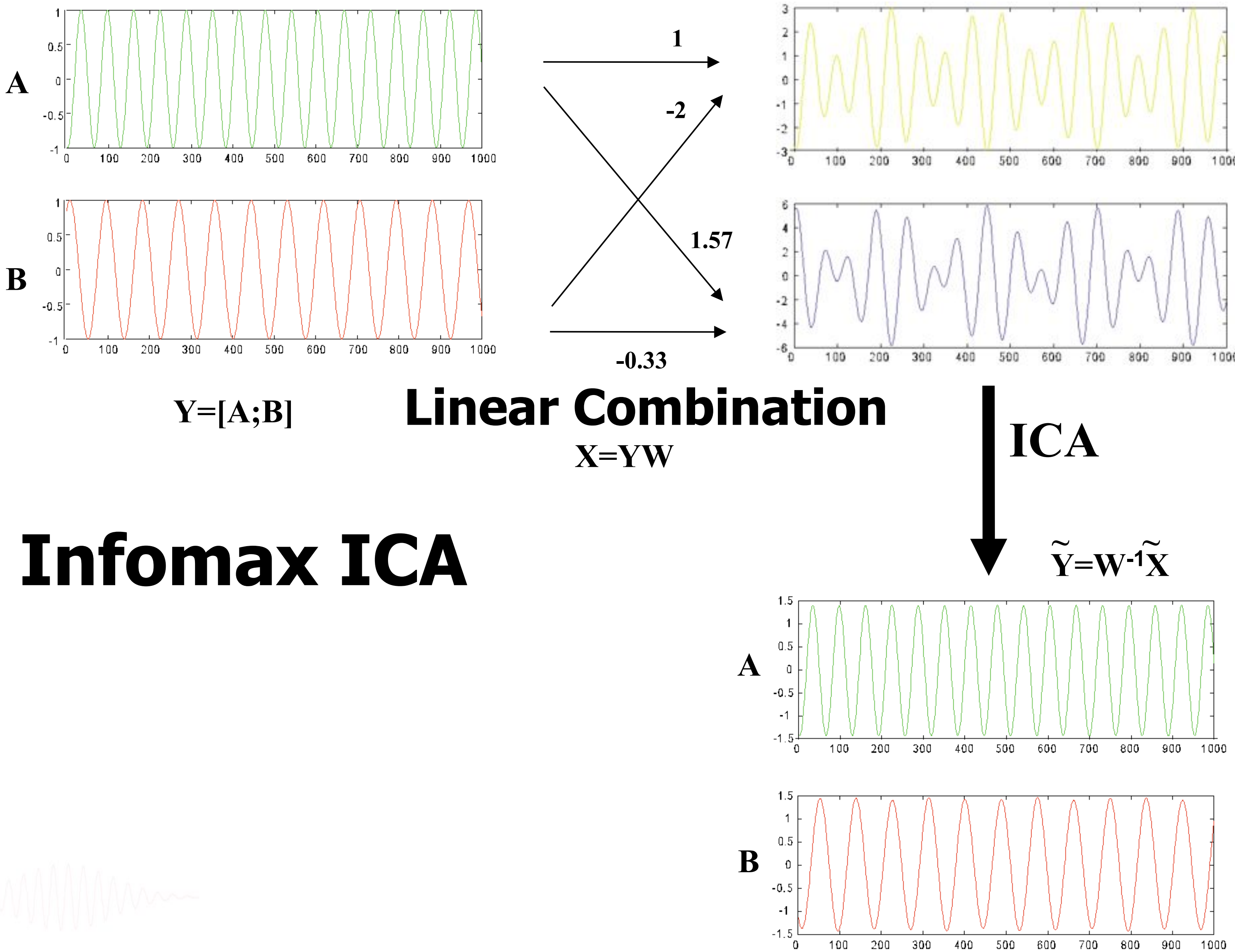



Electrodes
Cortex
Skin
Skull


Electrodes
Cortex
Local
synchrony
Skin
Skull

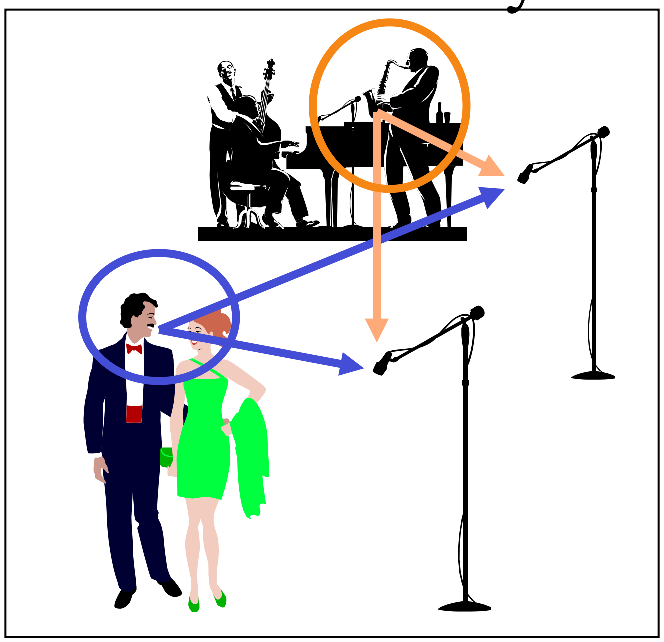

Electrodes
Cortex
Skin
Skull
Local synchrony
Relative Independance
Local synchrony

Infomax ICA

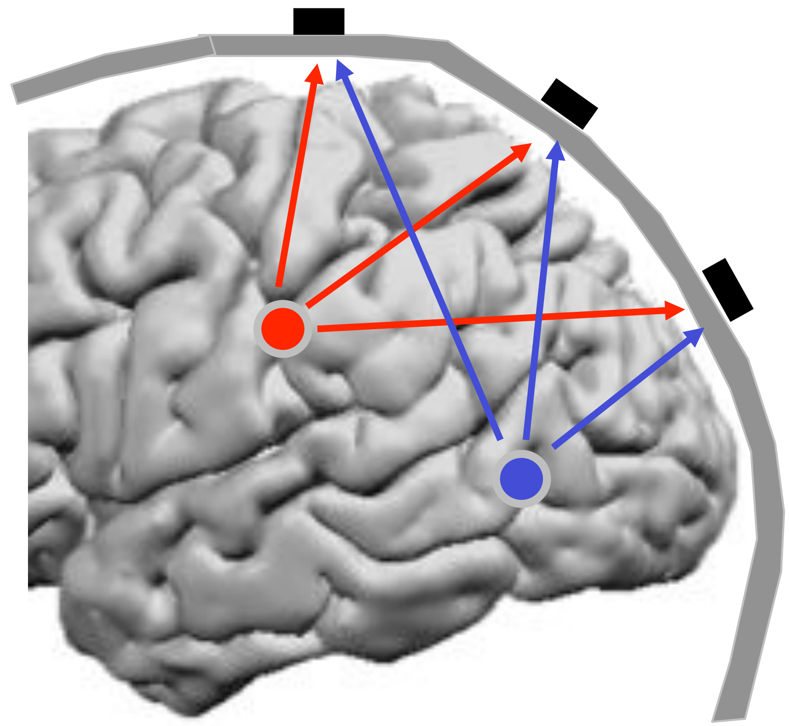

A
B
Y=[A;B]
1
-2
1.57
-0.33
Linear Combination
X=YW
ICA
$\tilde{Y}=W^{-1}\tilde{X}$
A
B

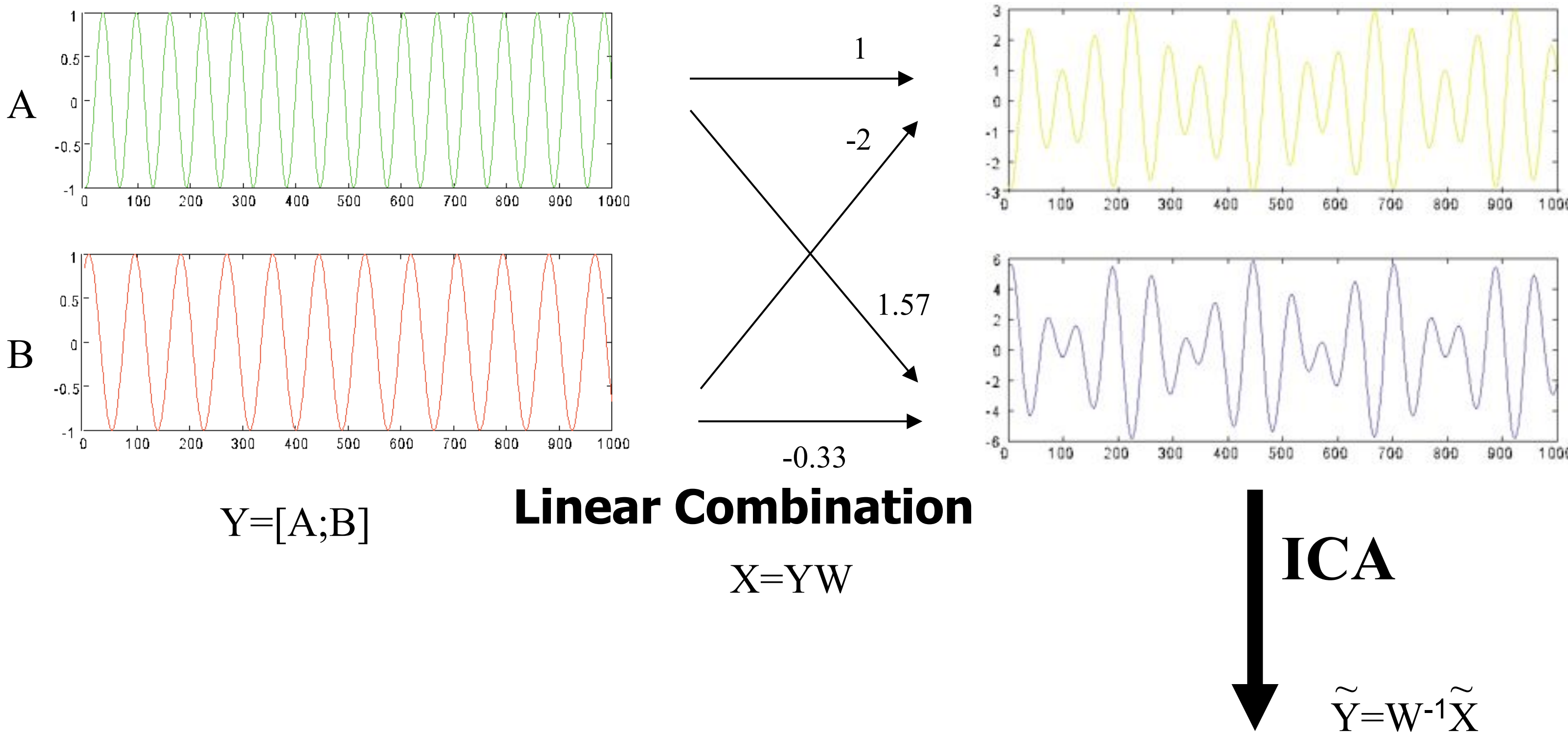
Independent component analysis

Cocktail Party

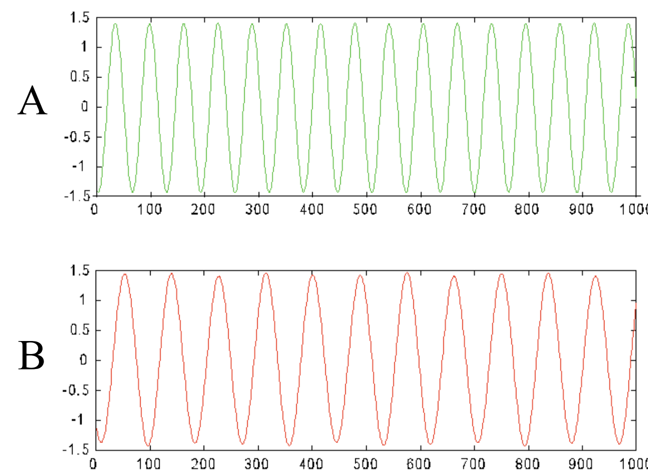


Mixture of Brain source activity





ICA is a method to recover a version, of the original sources by multiplying the data by a unmixing matrix



$$U = WX$$

ICA activity

Data

Data X

$$\begin{array}{cl} \begin{bmatrix} 3 & 2 & 5 & 4 & 3 & 2 & \\ 0 & -2 & -5 & -1 & 1 & -1 & \cdots \\ -1 & 2 & 0 & 1 & 0 & -3 & \end{bmatrix} & \begin{array}{l} \leftarrow \text{Channel 1} \\ \leftarrow \text{Channel 2} \\ \leftarrow \text{Channel 3} \end{array} \end{array}$$

$$\begin{bmatrix} 5 & 3 & -2 \\ 1 & 2 & 4 \\ 0 & -1 & 3 \end{bmatrix} \longrightarrow \begin{array}{cl} \begin{bmatrix} 3*5+0*3-1*(-2) & 2*5+(-2)*3+2*(-2) & \\ 3*1+0*2-1*4 & 2*1+(-2)*2+2*4 & \cdots \\ 5*1-5*2+0*4 & 5*1-5*2+0*4 & \end{bmatrix} & \begin{array}{l} \leftarrow \text{Comp. 1} \\ \leftarrow \text{Comp. 2} \\ \leftarrow \text{Comp. 3} \end{array} \end{array}$$

Weight matrix W

ICA activity U

$$X = W^{-1}U$$

Data

ICA activity U

$$\begin{array}{cc} \begin{bmatrix} 3 & 2 & 5 & 4 & 3 & 2 & \\ 0 & -2 & -5 & -1 & 1 & -1 & \ldots \\ -1 & 2 & 0 & 1 & 0 & -3 & \end{bmatrix} & \begin{array}{l} \leftarrow \text{Comp. 1} \\ \leftarrow \text{Comp. 2} \\ \leftarrow \text{Comp. 3} \end{array} \end{array}$$

* * *

$$\begin{bmatrix} 5 & 3 & -2 \\ 1 & 2 & 4 \\ 0 & -1 & 3 \end{bmatrix} \longrightarrow \begin{array}{cc} \begin{bmatrix} 3*5+0*3-1*(-2) & 2*5+(-2)*3+2*(-2) & \\ 3*1+0*2-1*4 & 2*1+(-2)*2+2*4 & \ldots \\ 5*1-5*2+0*4 & 5*1-5*2+0*4 & \end{bmatrix} & \begin{array}{l} \leftarrow \text{Chan 1} \\ \leftarrow \text{Chan 2} \\ \leftarrow \text{Chan 3} \end{array} \end{array}$$

Inverse weight matrix W^{-1}

Data X

Historical Remarks

- Herault & Jutten ("Space or time adaptive signal processing by neural network models“, *Neural Nets for Computing Meeting*, Snowbird, Utah, 1986): Seminal paper, neural network
- Bell & Sejnowski (1995): Information Maximization
- Amari et al. (1996): Natural Gradient Learning
- Cardoso (1996): JADE

- **Applications of ICA to biomedical signals**
  – EEG/ERP analysis (Makeig, Bell, Jung & Sejnowski, 1996).
  – fMRI analysis (McKeown et al. 1998)

ICA Theory – Cost Functions

Family of BSS algorithms

- Information theory (Infomax)
- Bayesian probability theory (Maximum likelihood estimation)
- Negentropy maximization
- Nonlinear PCA
- Statistical signal processing (cumulant maximization, JADE)

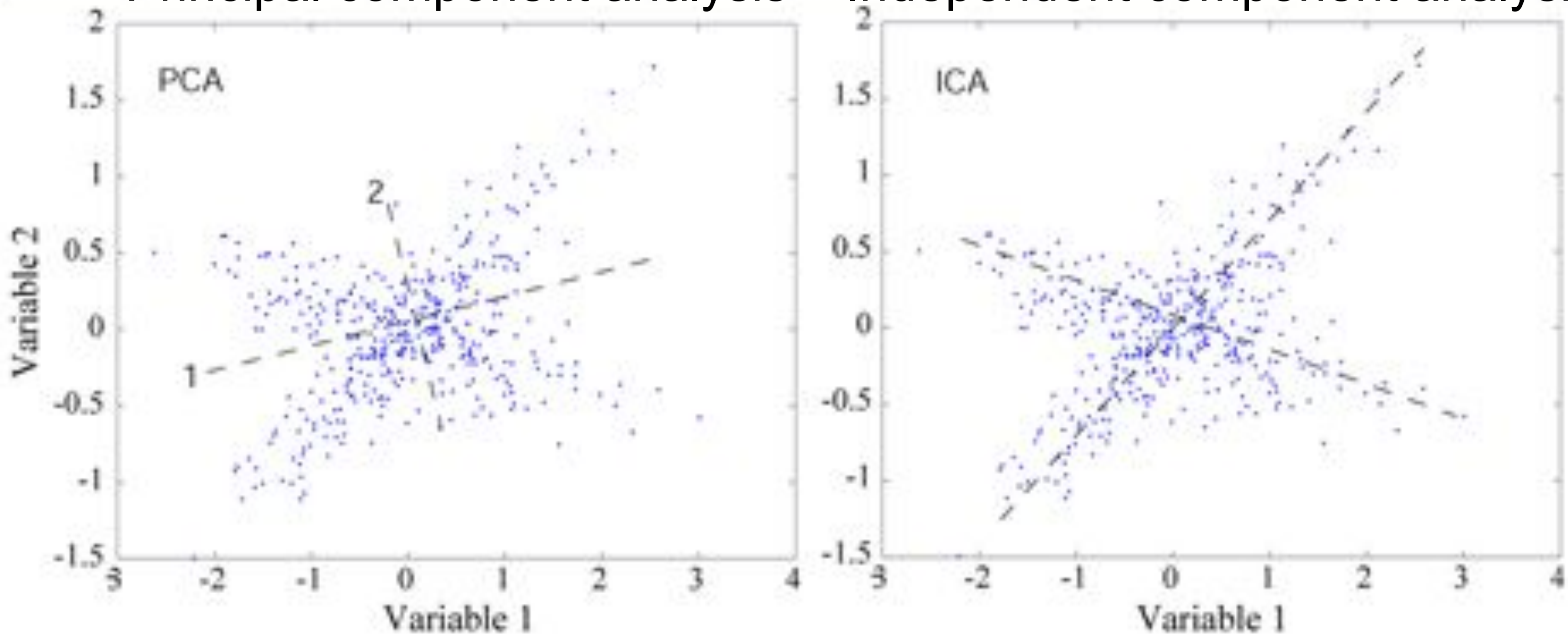
A unifying Information-theoretic framework for ICA

- Pearlmutter & Parra showed that InfoMax, ML estimation are equivalent.
- Lee et al. (1999) showed negentropy has the equivalent property to InfoMax.
- Girolami & Fyfe showed nonlinear PCA can be viewed from information-theoretic principle.

ICA and PCA

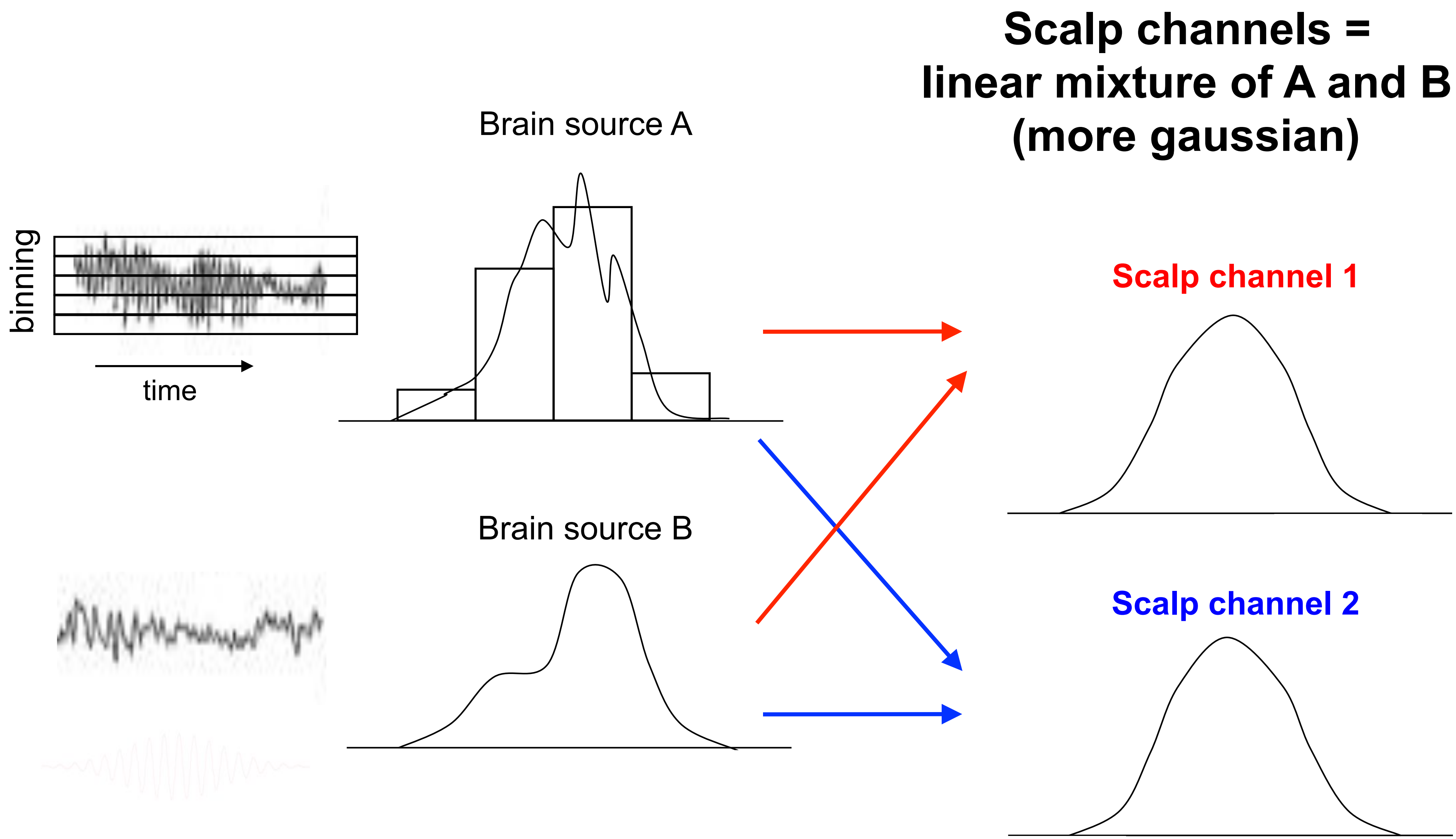
ICA is a method to recover a version, of the original sources by multiplying the data by a unmixing matrix,

Principal component analysis Independent component analysis



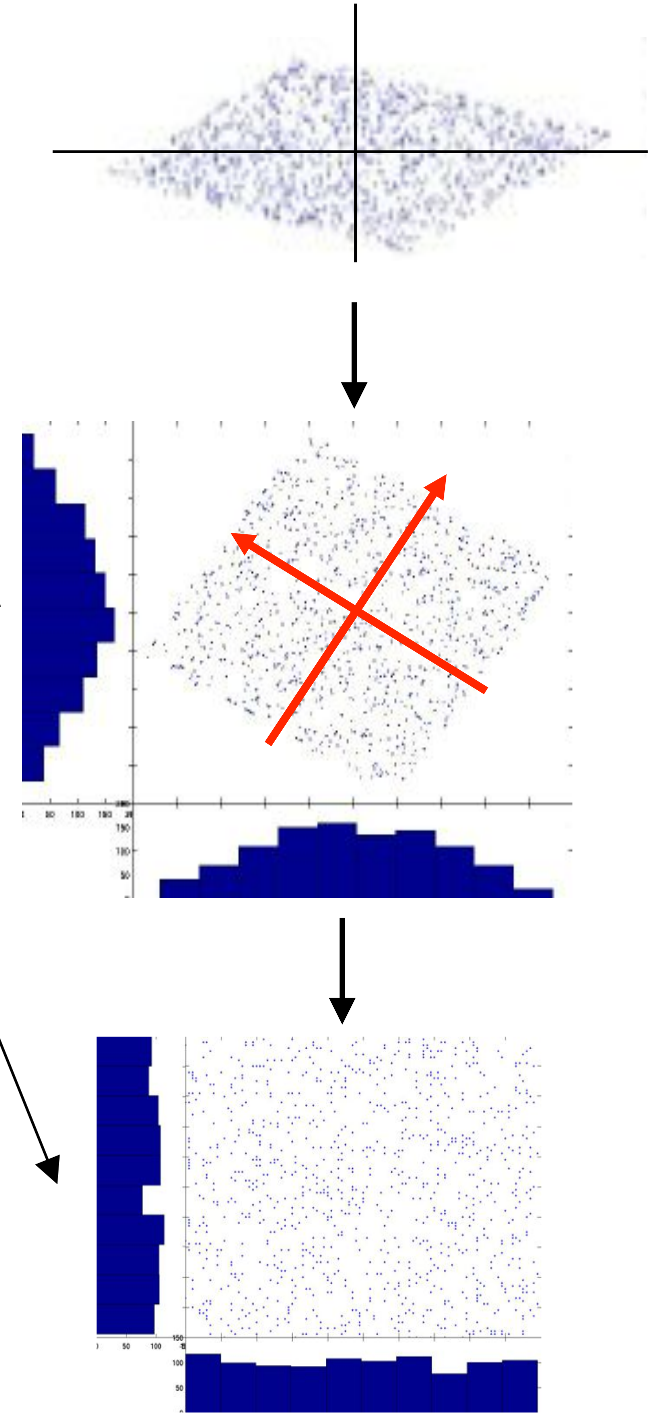
While PCA simply decorrelates the outputs (using an orthogonal mixing matrix), ICA attempts to make the outputs **statistically independent**, while placing no constraints on the minxing matrix.

Central limit theorem



ICA Training Process

Central limit theorem



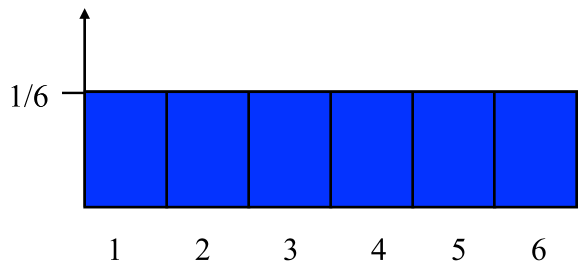
- Remove the mean
  $x = x - \langle x \rangle$
- 'Sphere' the data by diagonalizing its covariance matrix,
  $x = \langle xx^T \rangle^{-1/2}(x - \langle x \rangle)$.
- Update W according to

$$\Delta \mathbf{W} \propto \frac{\partial H(\mathbf{y})}{\partial \mathbf{W}} \mathbf{W}^T \mathbf{W}$$

Entropy

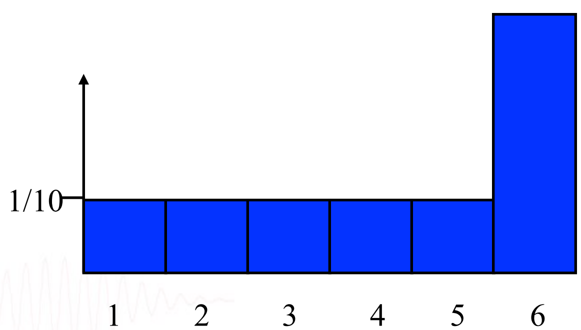
$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_b p(x).$$

Dice: 1/6



$$H = 6\left(-\frac{1}{6}\log_2\left(\frac{1}{6}\right)\right) = 2.58$$

Fake dice (make a 6 half of the time): entropy 2.16 (base 2)



$$H = 5\left(-\frac{1}{10}\log_2\left(\frac{1}{10}\right)\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right) = 2.16$$

Entropy

$$H(X) = -\sum_{x\in\mathcal{X}} p(x)\log_b p(x).$$

Joint entropy

$$H(X,Y) = -\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} p(x,y)\log_b p(x,y);$$

Mutual Information

$$H(y_1,y_2) = H(y_1) + H(y_2) - I(y_1,y_2).$$

Shannon in his landmark 1948 paper ``A Mathematical Theory of Communication.''

From http://planetmath.org/encyclopedia/ShannonsTheoremEntropy.html

Contingency table for stress and emotionality

	STRE						
	1	**2**	**3**	**4**	**5**	**6**	**Total**
EMOT= **1**	19	4					23
2	11	63	64	3	1		142
3	2	16	18	20	2	2	60
4	1	4	1	9	6	2	23
5			1	2	4	3	10
6				1	1	1	3
Total	33	87	84	35	13	8	

From http://tecfa.unige.ch/~lemay/thesis/THX-Doctorat/node149.html

Contingency frequencies for stress and emotionality

	STRE					
	1	**2**	**3**	**4**	**5**	**6**
EMOT= **1**	0.07	0.02				
2	0.04	0.24	0.25	0.01		
3	0.01	0.06	0.07	0.08	0.01	0.01
4		0.02		0.03	0.02	0.01
5				0.01	0.02	0.01
6						

Joint entropy 3.46; exercise: compute mutual information

$$H(X,Y) = -\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} p(x,y)\log_b p(x,y)$$

ICA learning rule

How to make the outputs statistical independent?

Minimize their redundancy or mutual information.

Consider the joint entropy of two components,

$$H(y_1, y_2) = H(y_1) + H(y_2) - I(y_1, y_2).$$

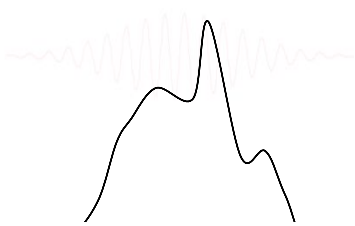
Maximizing $H(y_1, y_2) \Longrightarrow$ minimizing $I(y_1, y_2)$.

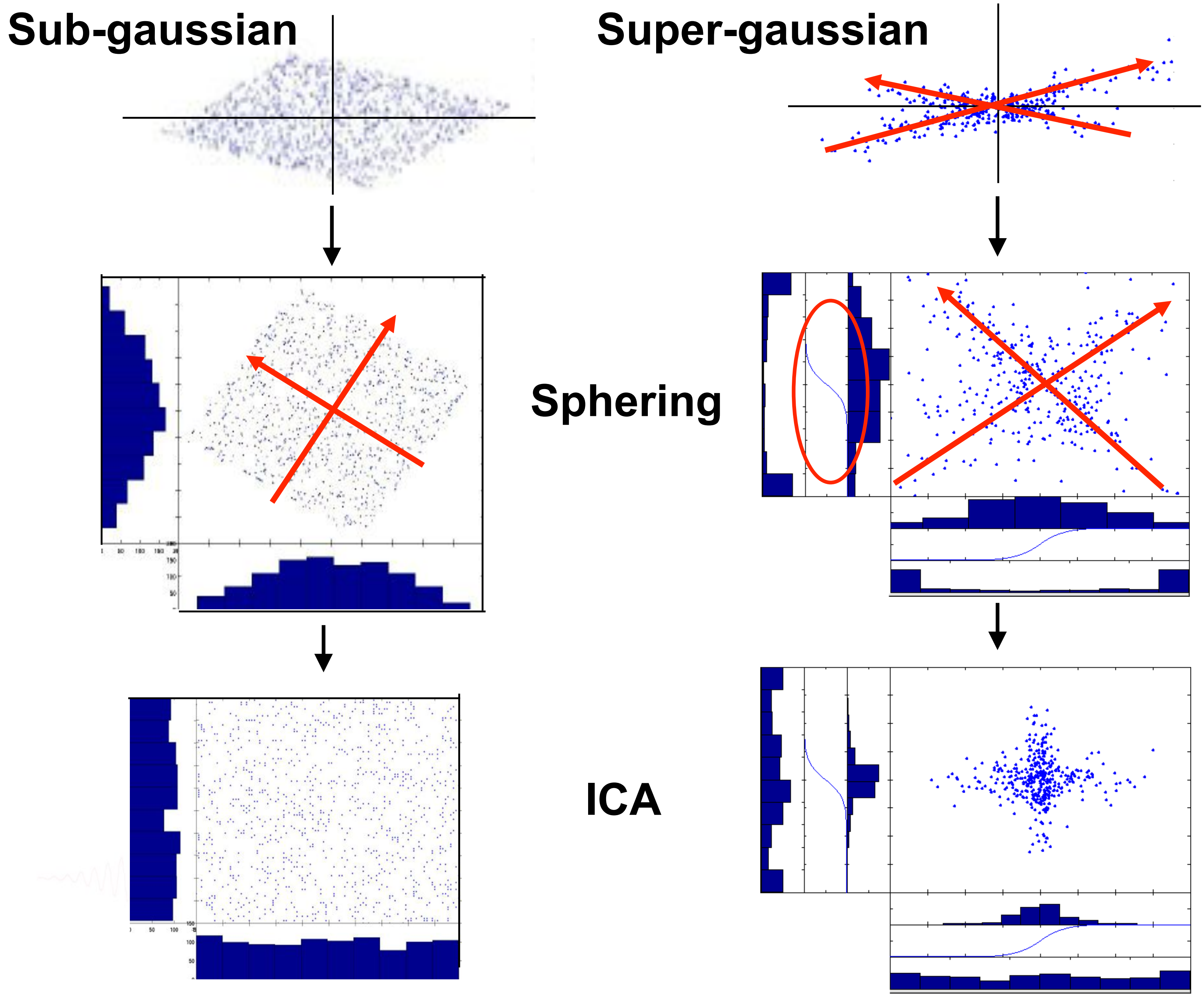
=0 if the two variables are independent

The learning rule:

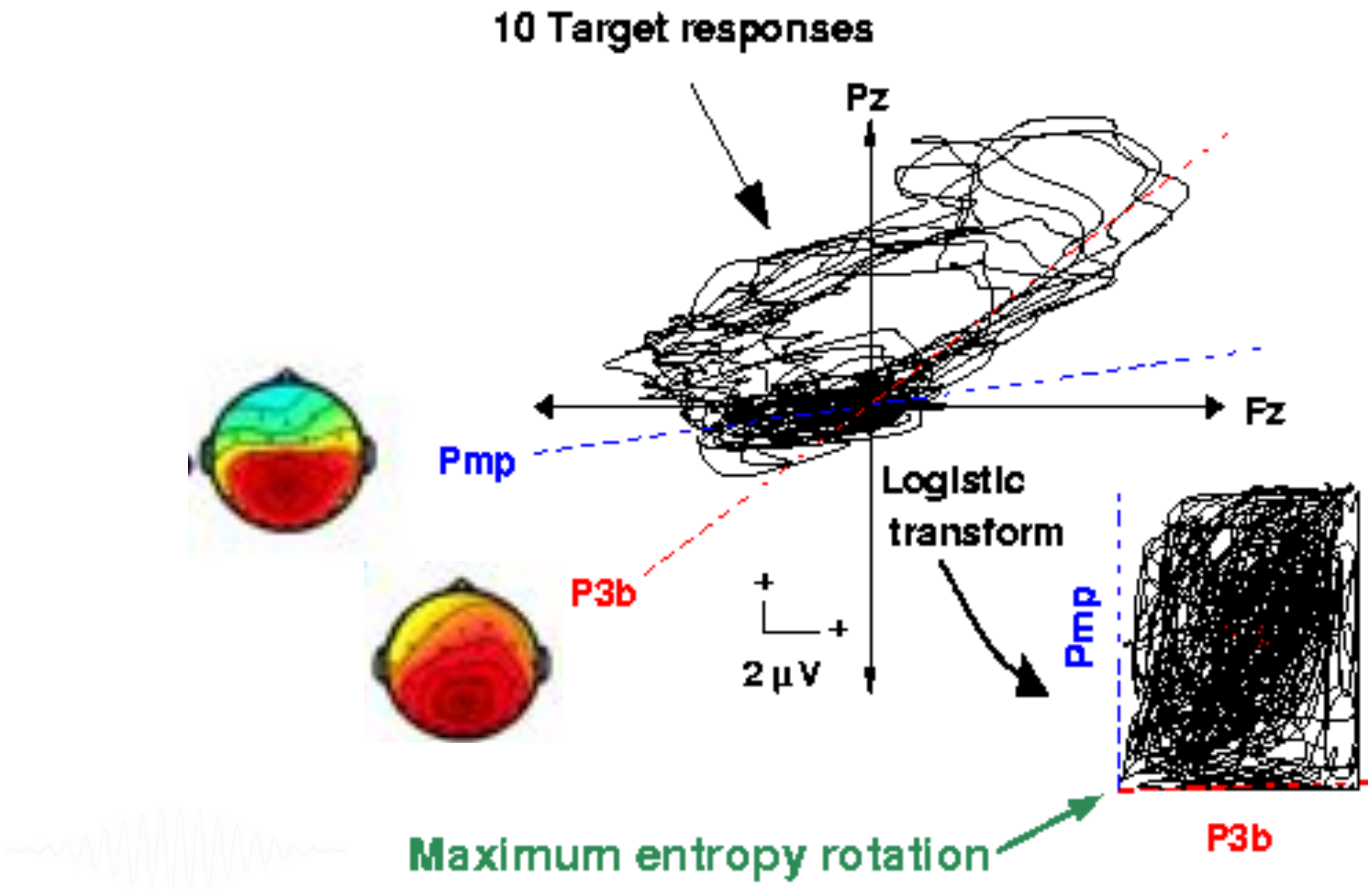
$$\Delta \mathbf{W} \propto \frac{\partial H(\mathbf{y})}{\partial \mathbf{W}} \underbrace{\mathbf{W}^T \mathbf{W}}_{\text{Natural gradient (Amari)}}$$

Entropy extremum



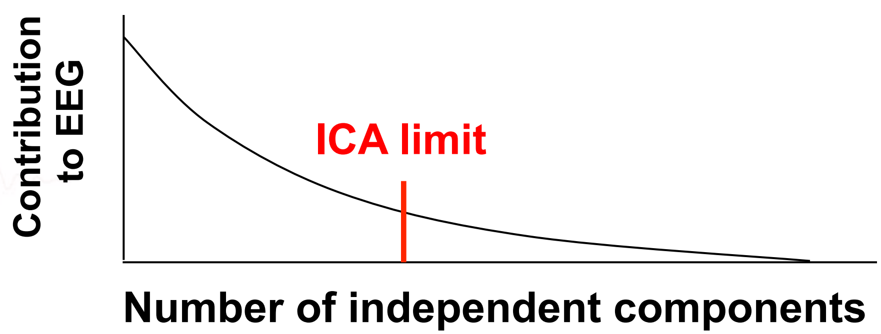

Sub-gaussian
Super-gaussian
Sphering
ICA

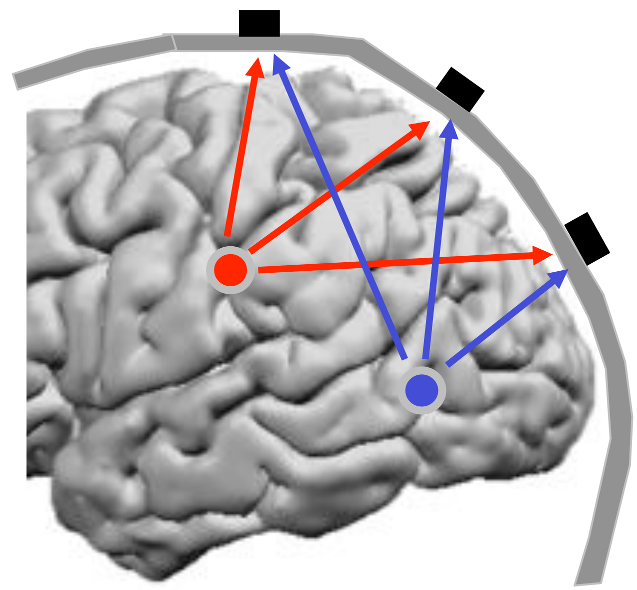
Independent components of EEG/ERP



ICA/EEG Assumptions

- Mixing is linear at electrodes **OK**
- Propagation delays are negligible **OK**
- Component time courses are independent ~
- Number of components less than the number of channels.





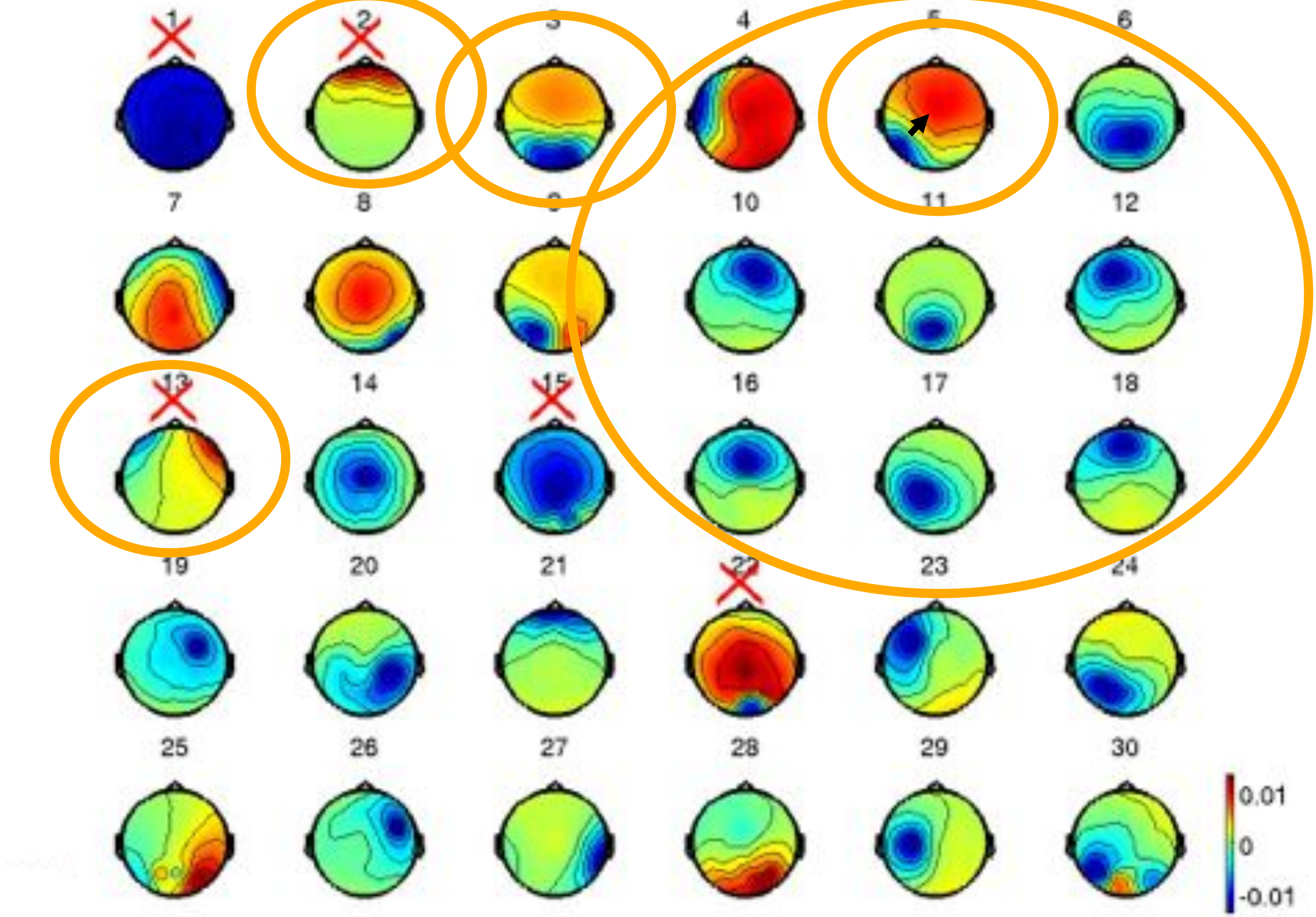
Independent Component Categories

- Artifacts
- Stimulus-locked activity
- Response-locked activity
- Non-phase locked activity
- Event-modulated oscillatory activity

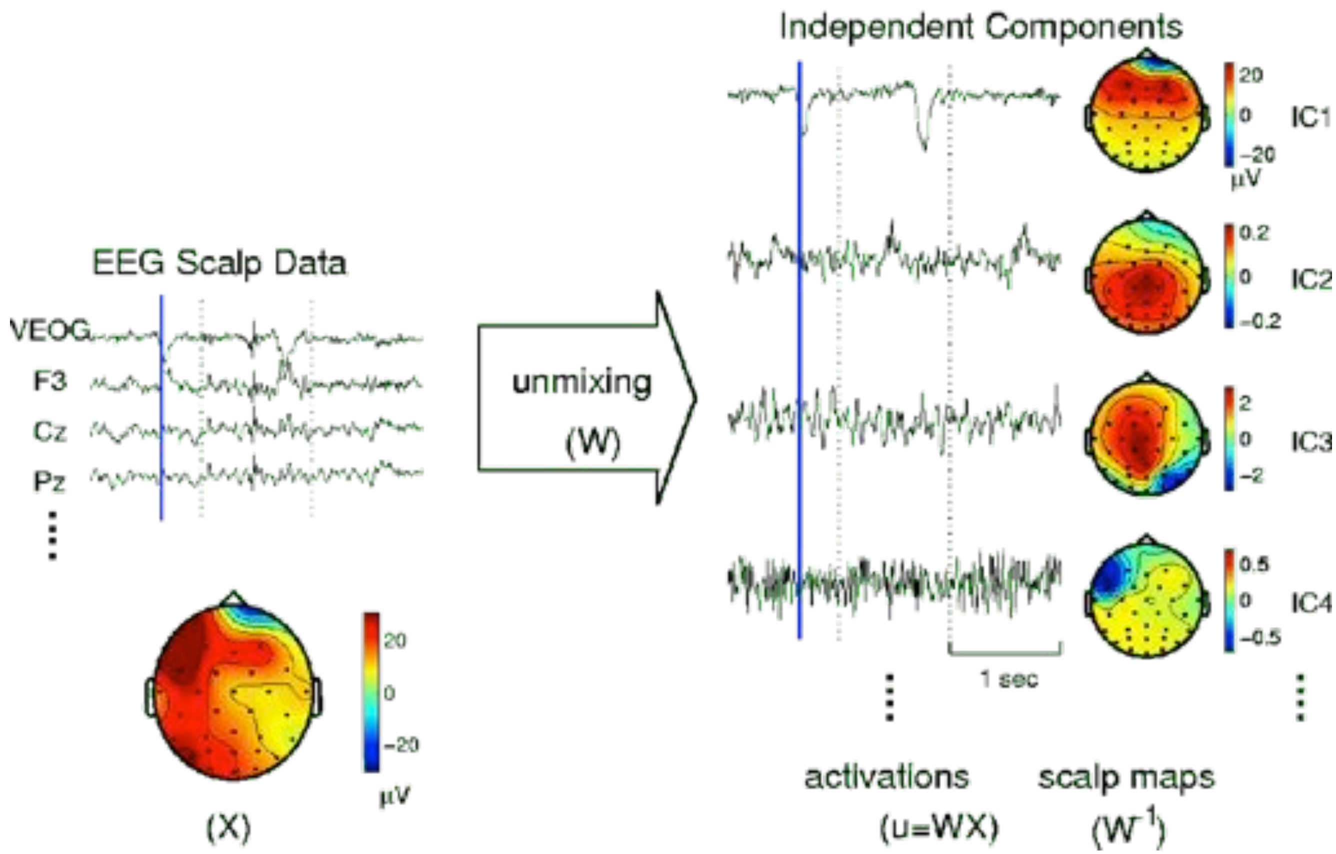
Characteristics of Independent Component of the EEG

- Concurrent Activity
- Maximally Temporally Independent
- Overlapping Maps and Spectra
- Dipolar Scalp Maps
- Functionally Independent
- Between-Subject Regularity

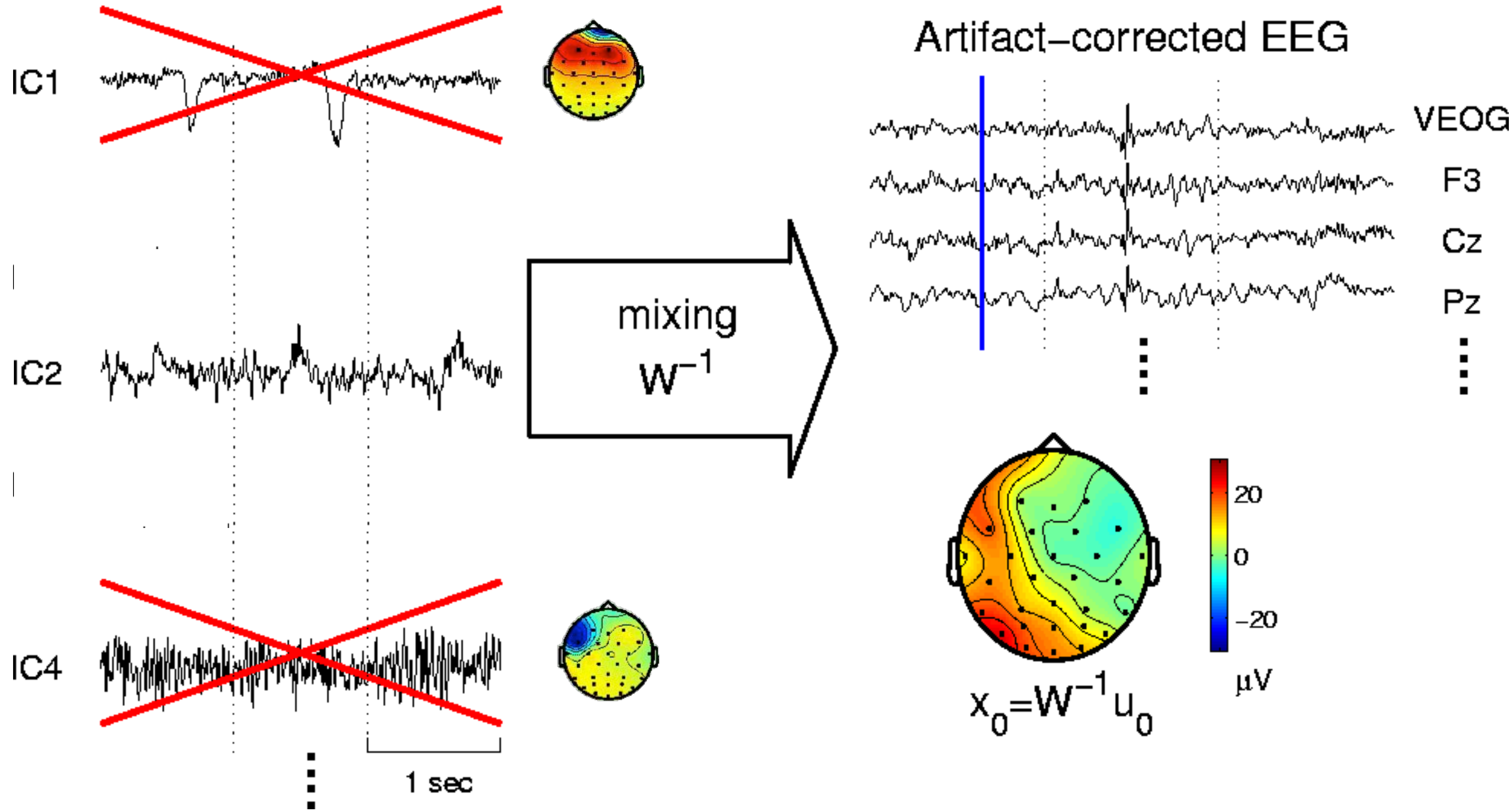
Largest 30 Independent Components
(single subject)



ICA Decomposition into Independent Components



Selective Projection onto Scalp Channels



$$\mathbf{X}=\mathbf{W}^{-1}\mathbf{U}$$

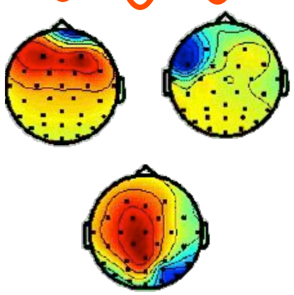
Data

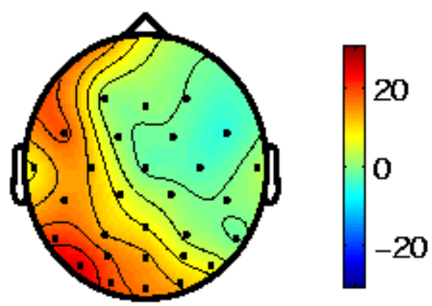
ICA activity U

$$\begin{array}{cl}\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ 0 & -2 & -5 & -1 & 1 & -1 & \dots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots \\ \dots & & & & & & \end{bmatrix} & \begin{array}{l}\leftarrow \text{Comp. 1} \\ \leftarrow \text{Comp. 2} \\ \leftarrow \text{Comp. 3} \\ \end{array}\end{array}$$

$$\begin{bmatrix} 5 & 3 & -2 & \dots \\ 1 & 2 & 4 & \dots \\ 0 & -1 & 3 & \dots \\ \dots & & & \end{bmatrix} \rightarrow \begin{array}{cl}\begin{bmatrix} 3*5+0*3-1*(-2) & 2*5+(-2)*3+2*(-2) & \dots \\ 3*1+0*2-1*4 & 2*1+(-2)*2+2*4 & \dots \\ 5*1-5*2+0*4 & 5*1-5*2+0*4 & \dots \\ \dots & \dots & \end{bmatrix} & \begin{array}{l}\leftarrow \text{Chan 1} \\ \leftarrow \text{Chan 2} \\ \leftarrow \text{Chan 3} \\ \end{array}\end{array}$$

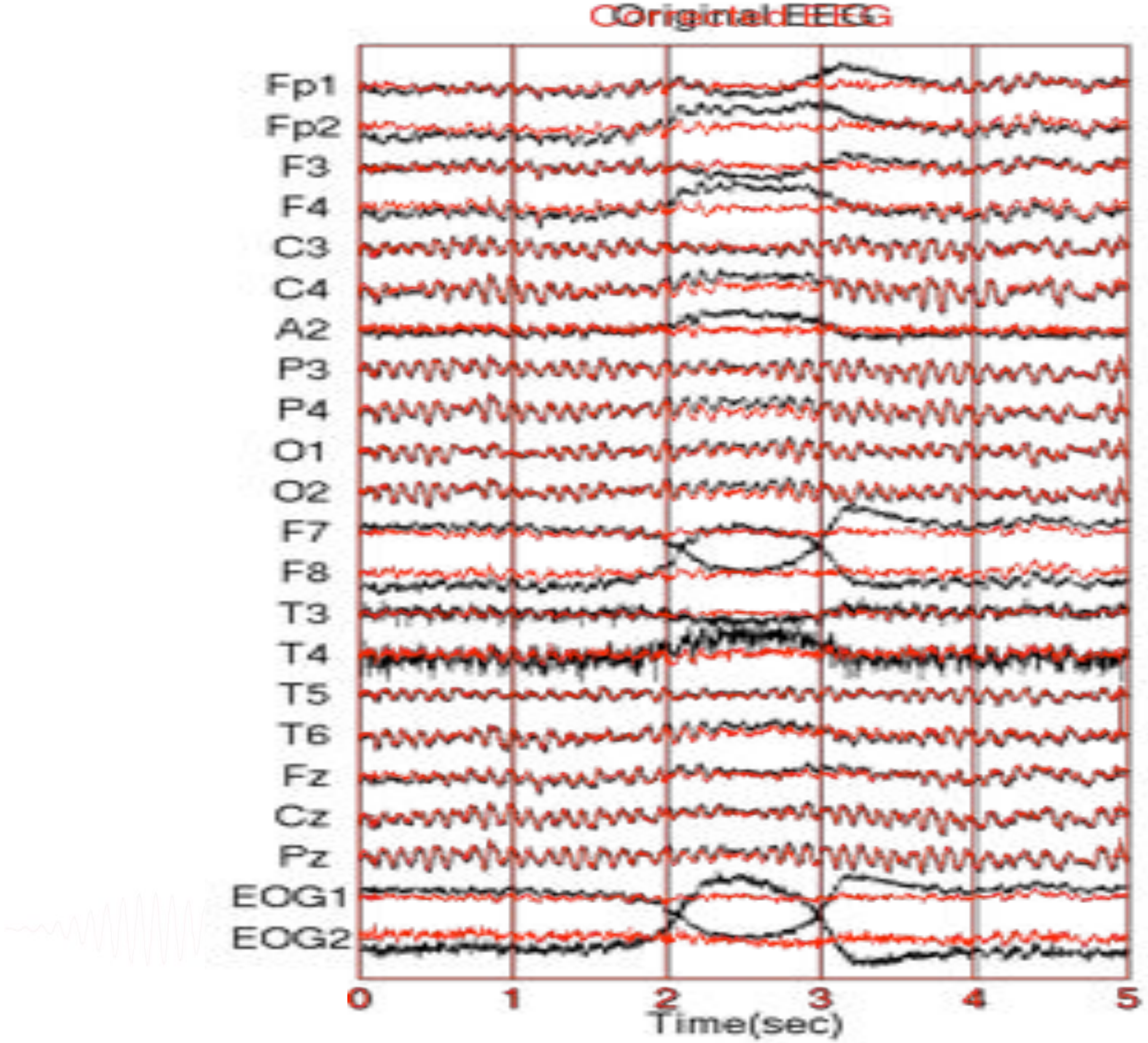
Data X



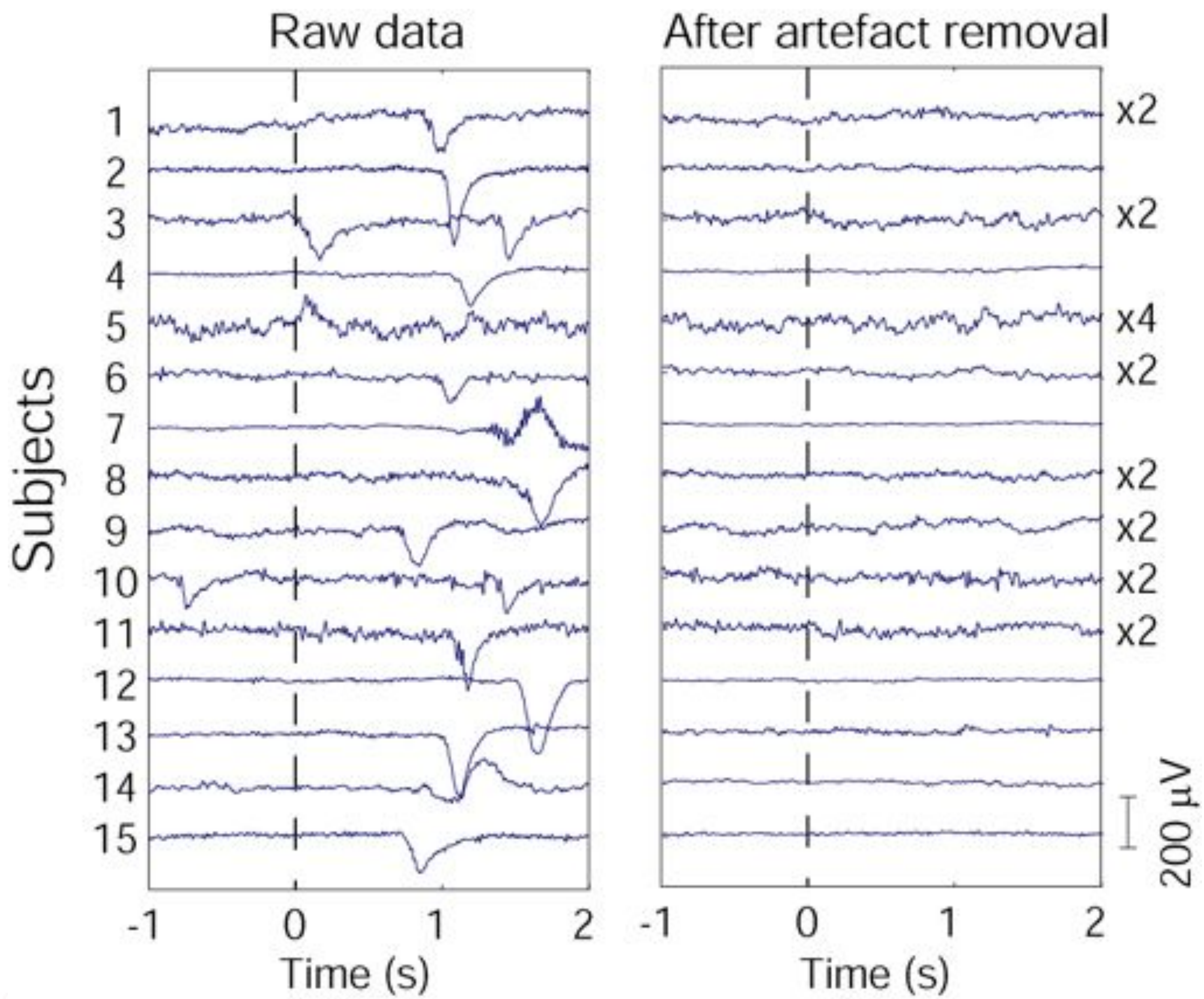


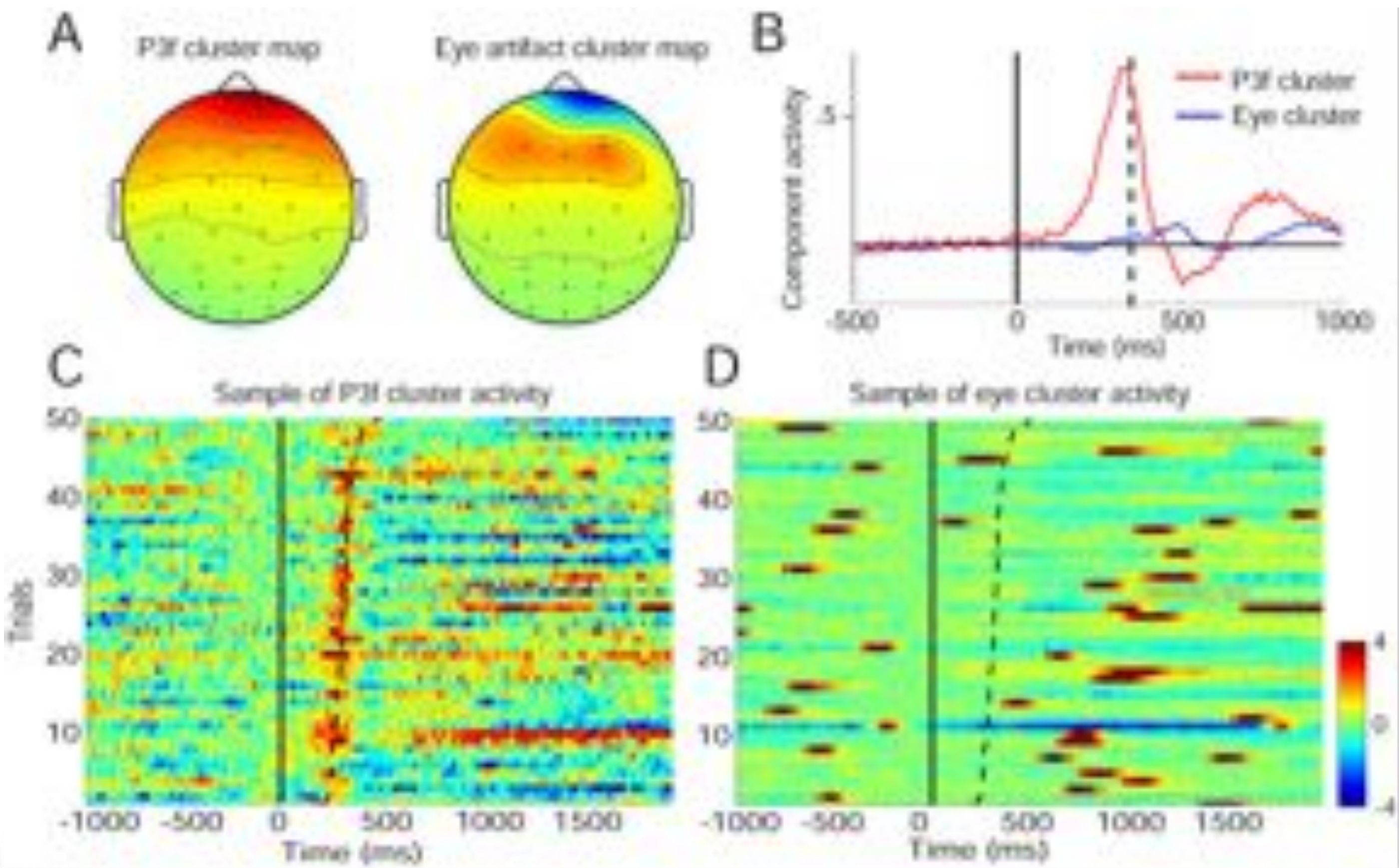
Inverse weight matrix W^{-1}

ICA-based Artifact Removal



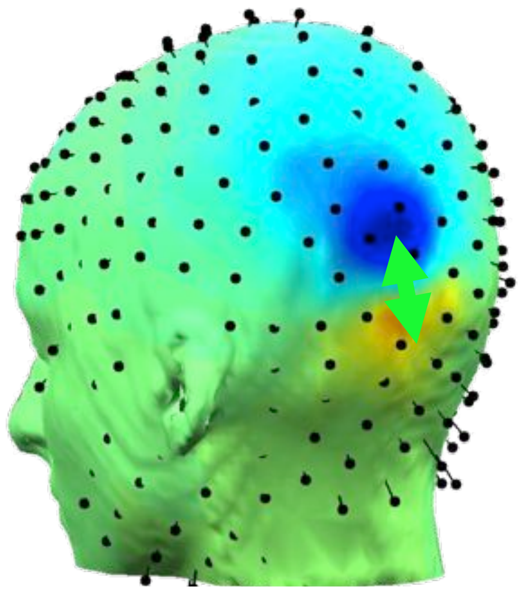
Artifact removal using ICA





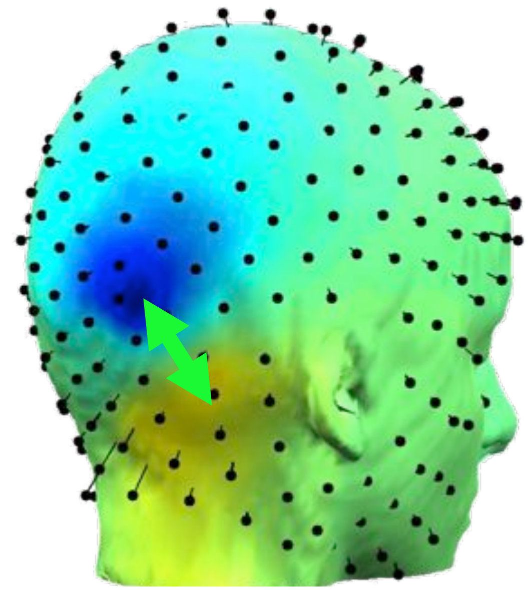
A
P3f cluster map
Eye artifact cluster map
B
P3f cluster
Eye cluster
Component activity
.5
-500
0
500
1000
Time (ms)
C
Sample of P3f cluster activity
Trials
50
40
30
20
10
-1000
-500
0
500
1000
1500
Time (ms)
D
Sample of eye cluster activity
50
40
30
20
10
-1000
-500
0
500
1000
1500
Time (ms)
4
0
-4

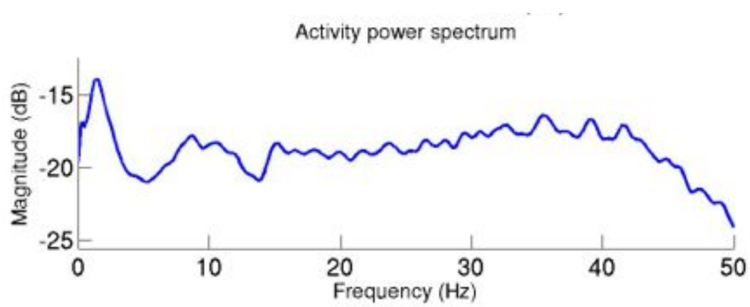
IC39

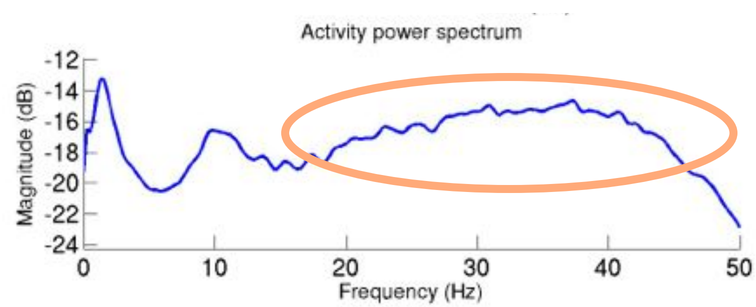


Two Neck Muscle Processes

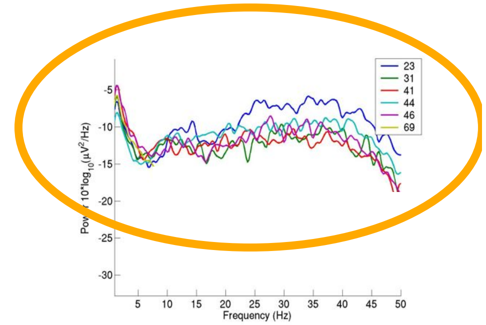
IC31

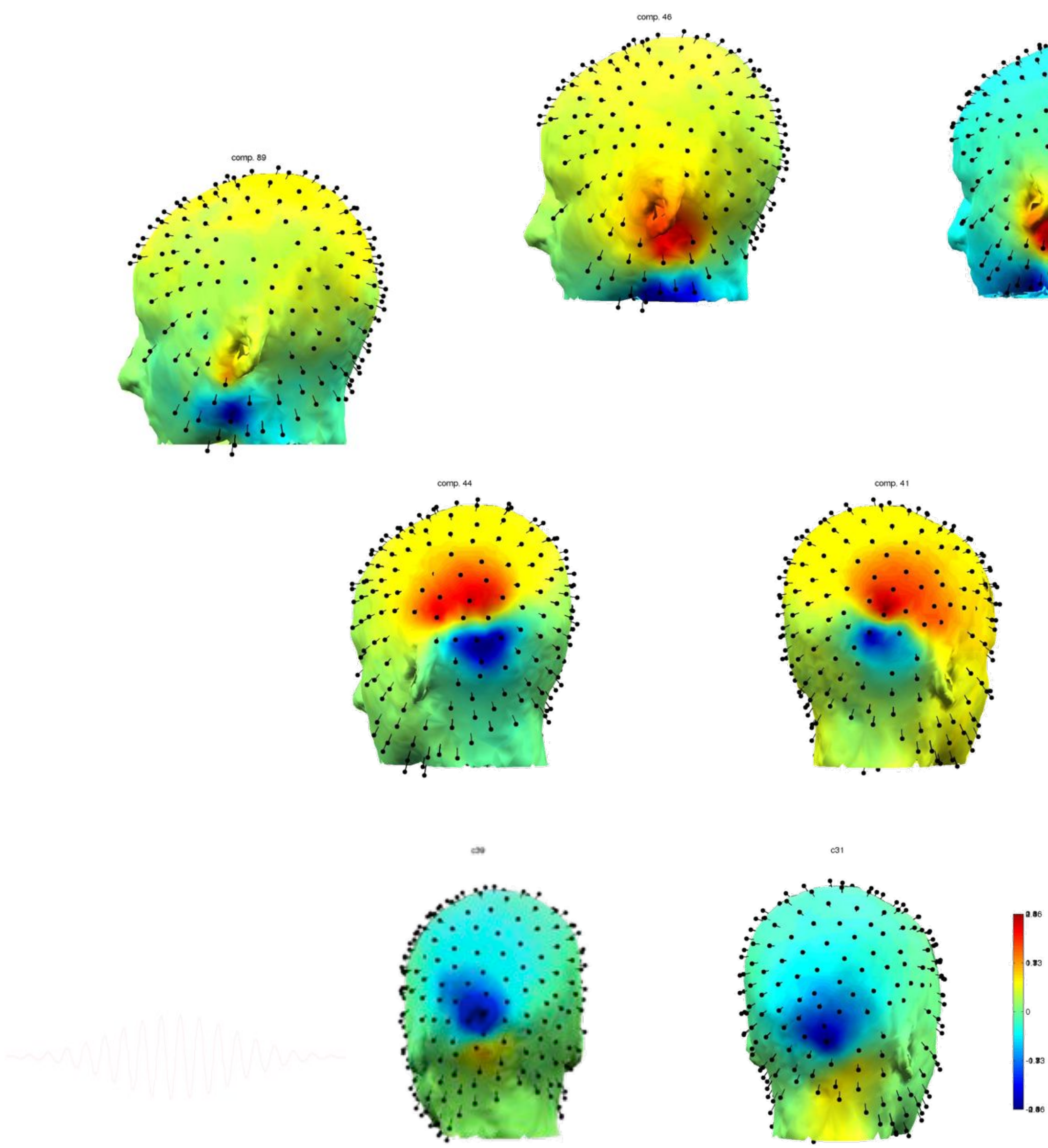


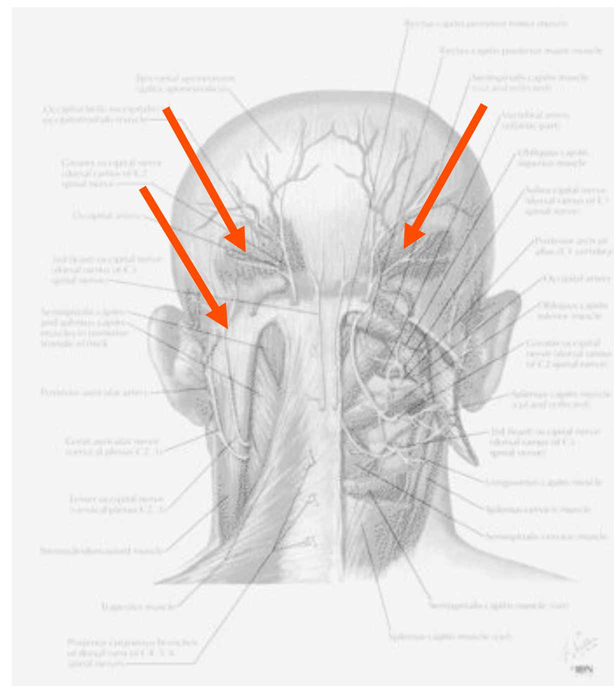




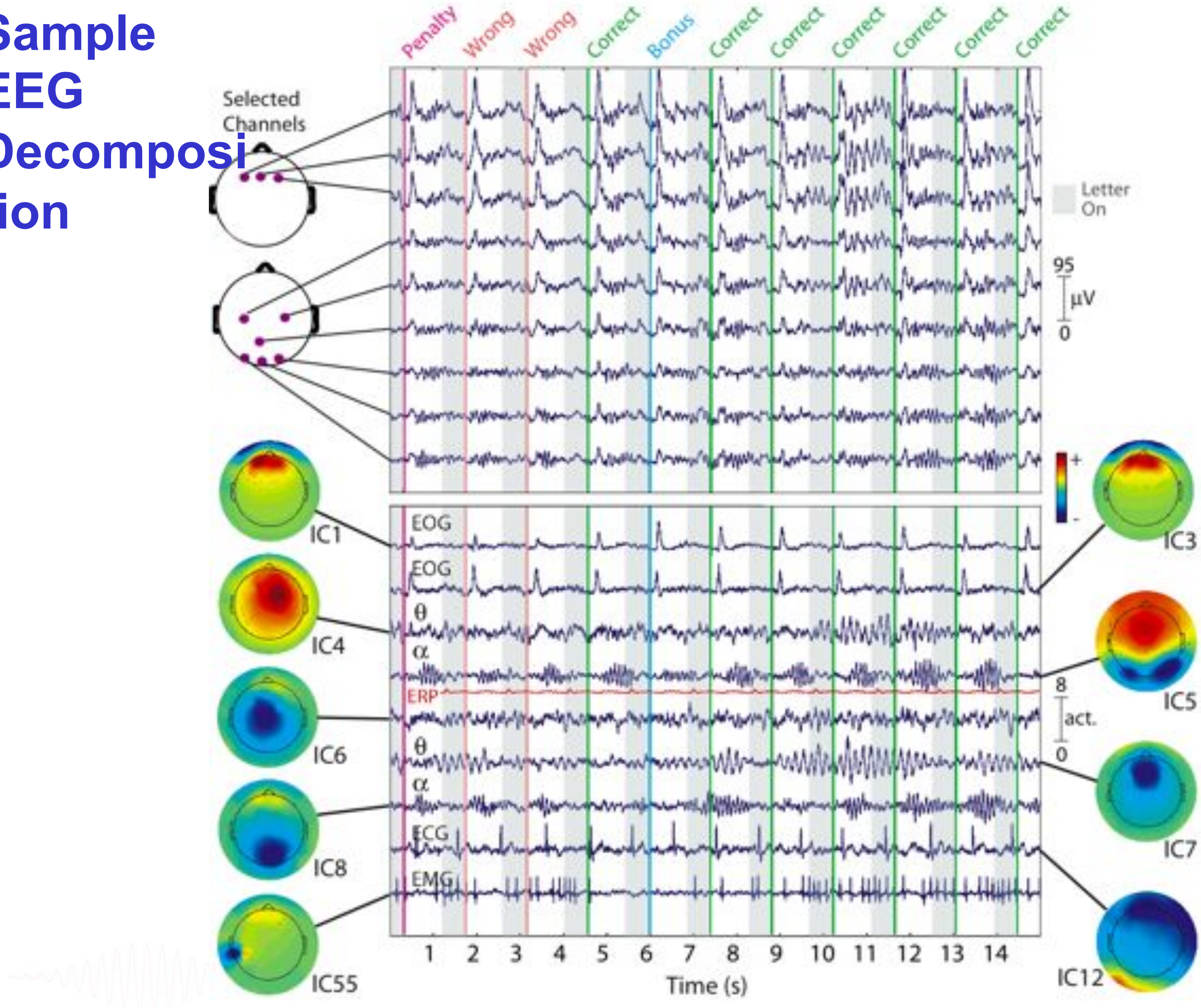
Some Independent EMG Components



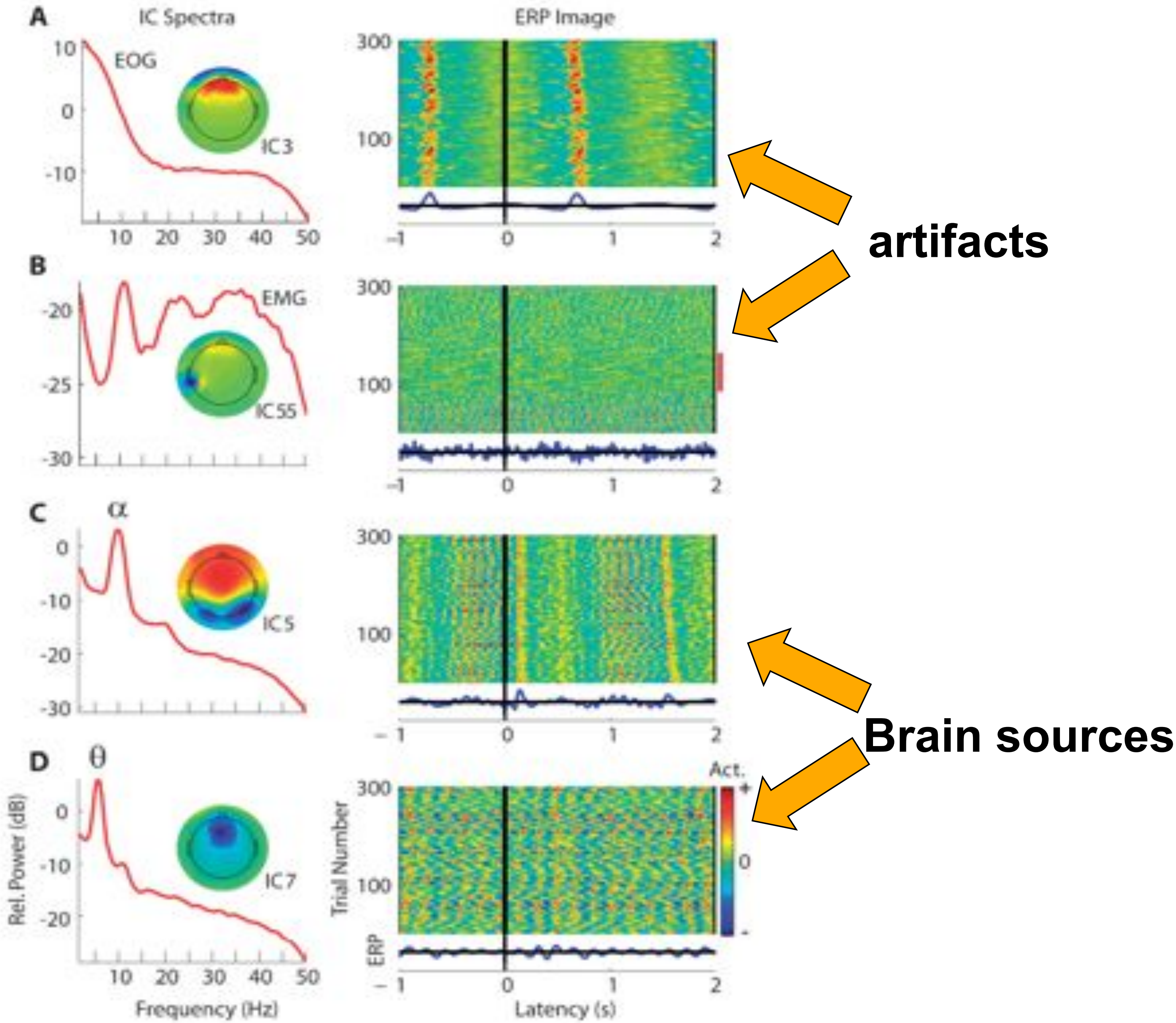




Sample EEG Decomposition

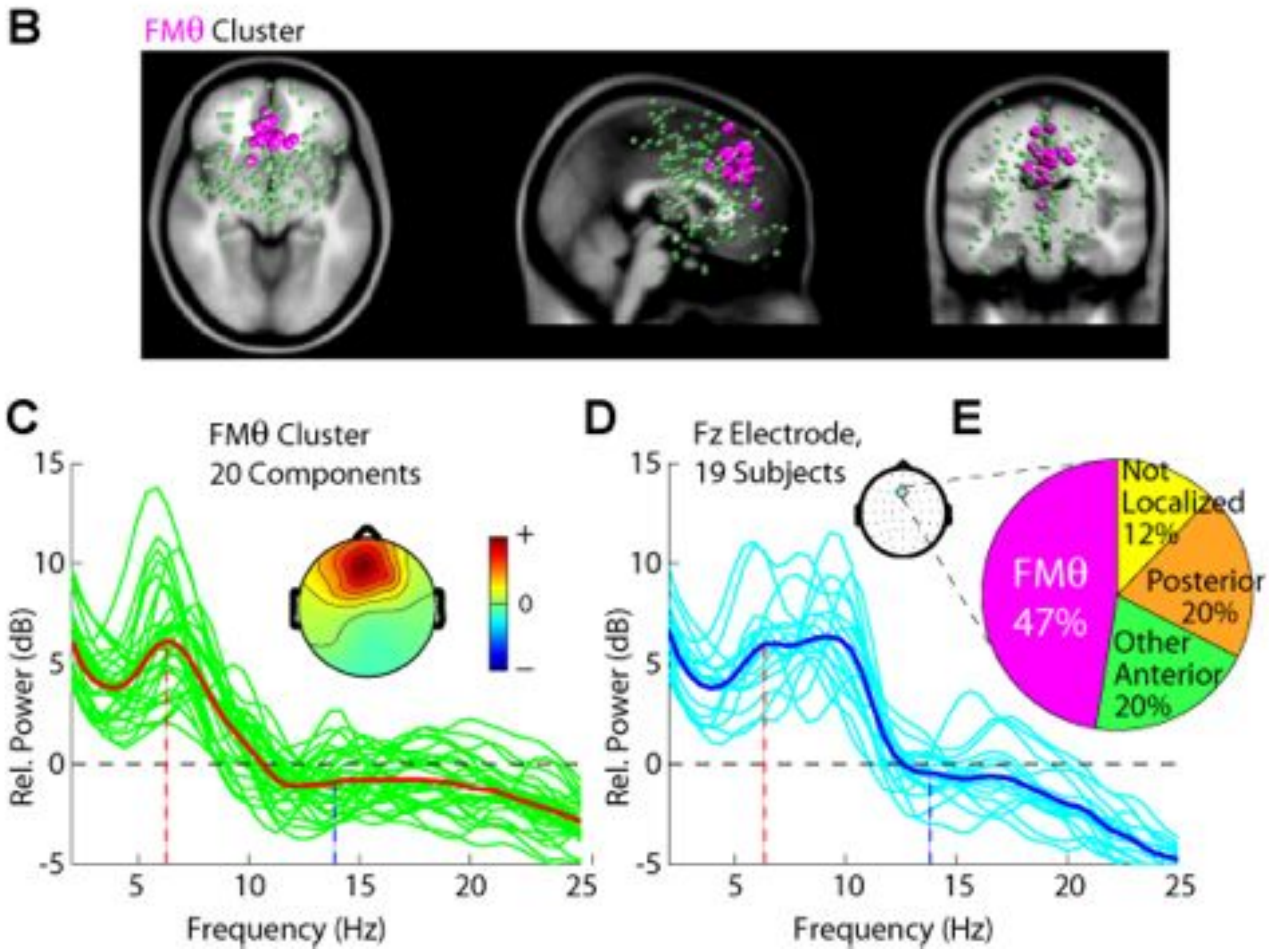


Credit: J. Onton



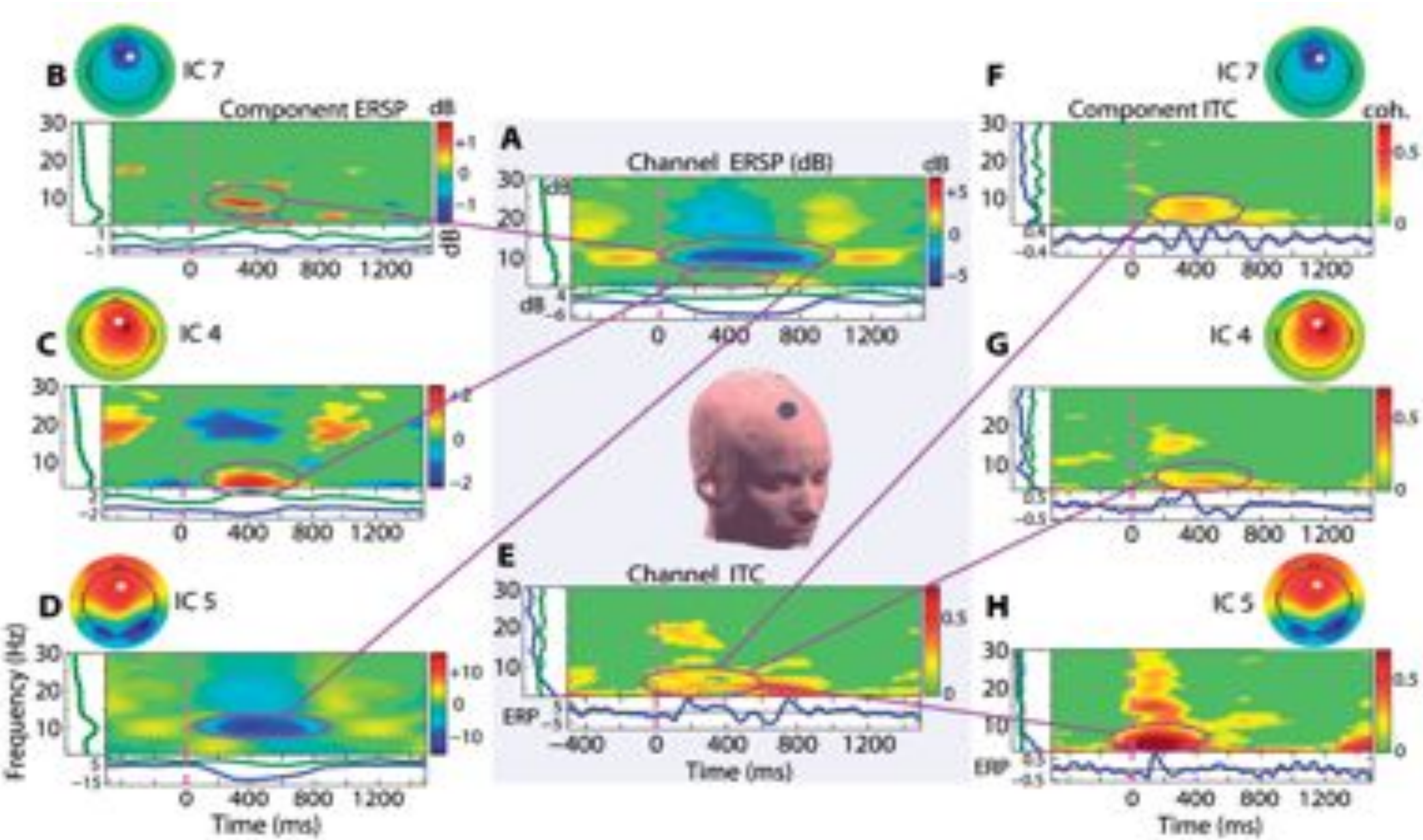
Credit: J. Onton

For example: frontal midline theta cluster



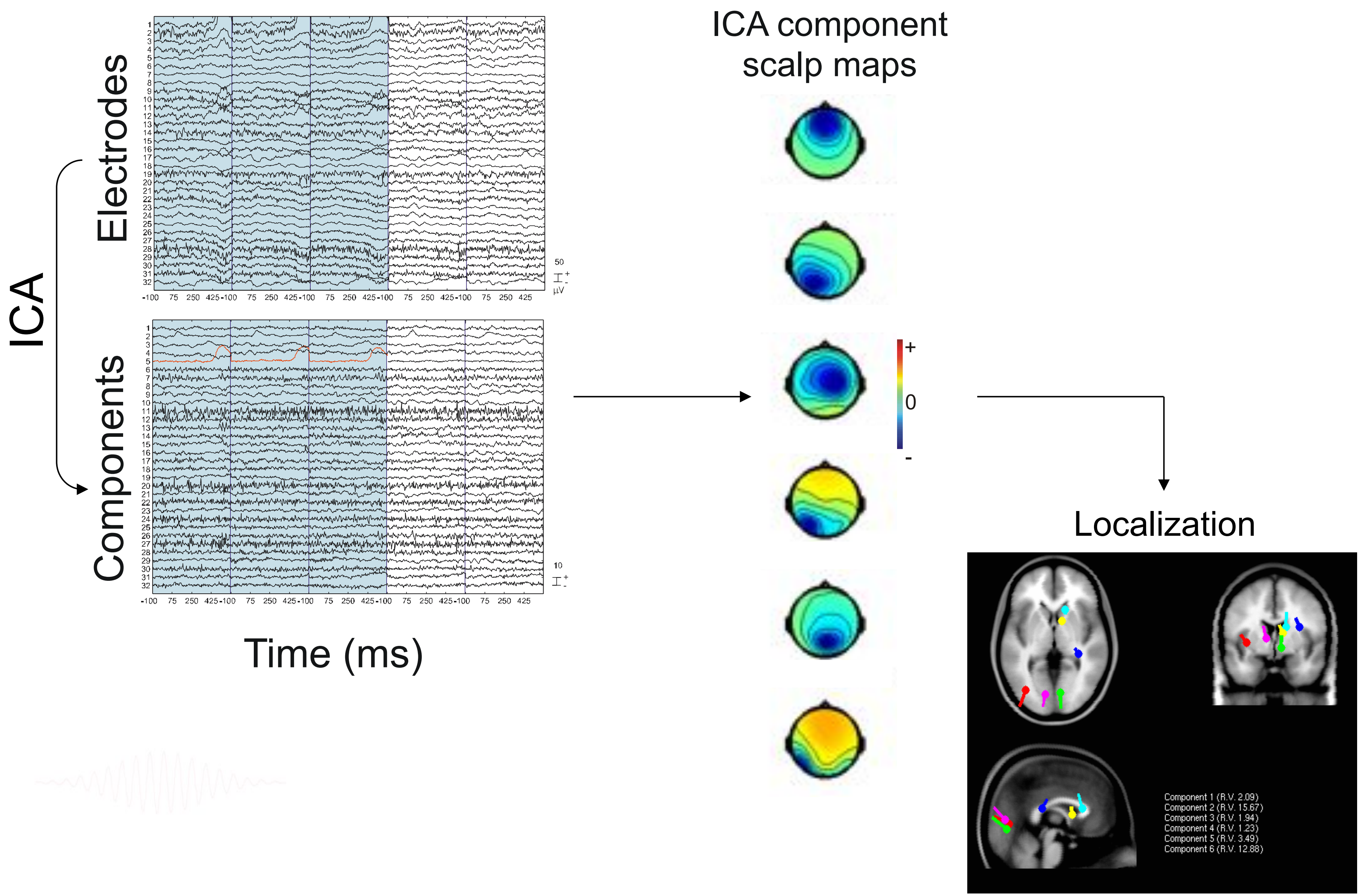
Onton, Delorme and Makeig, NeuroImage 27 (2005) 341 - 356

Why analyze source activity instead of channels?

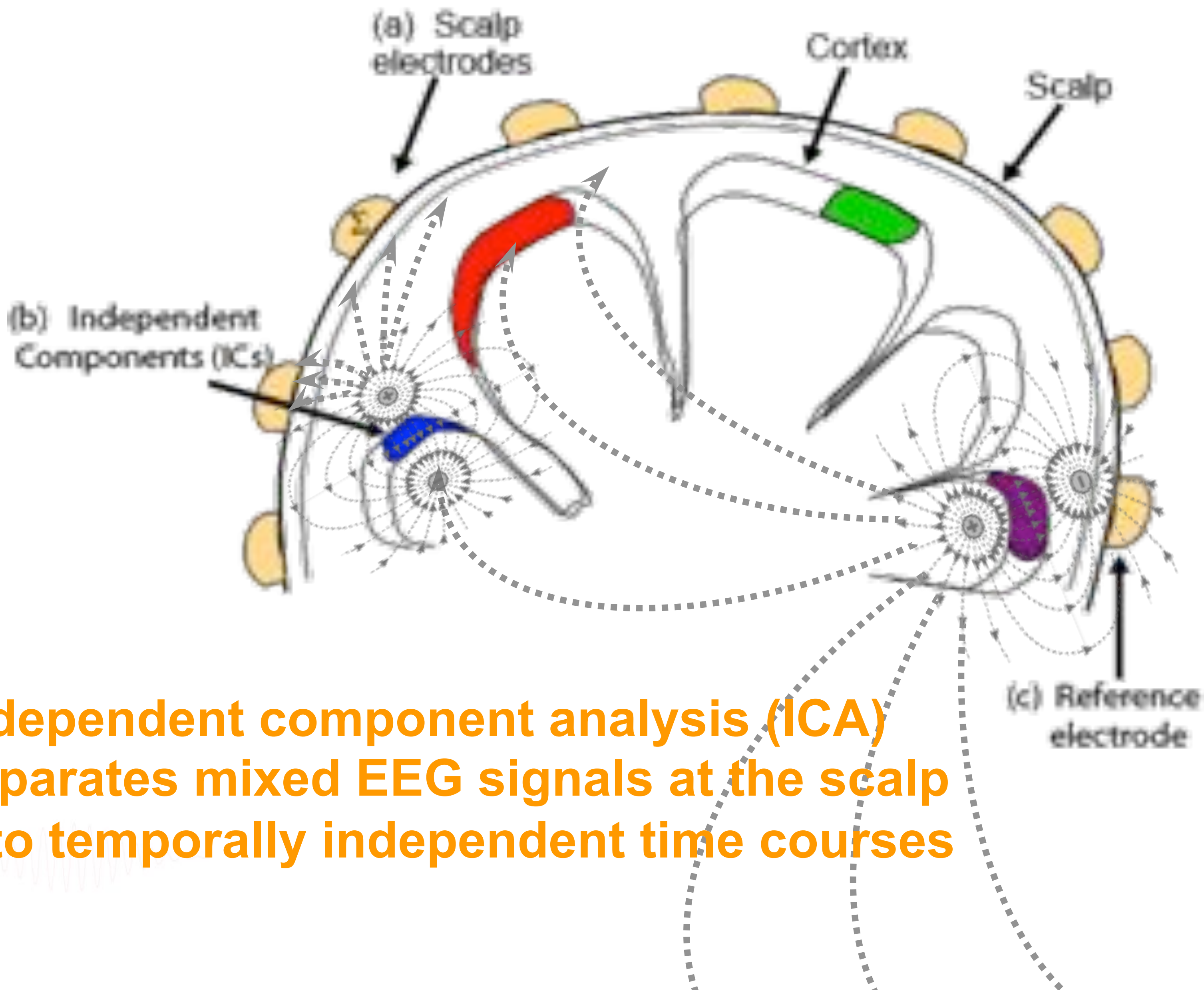


Credit: J. Onton

Localization

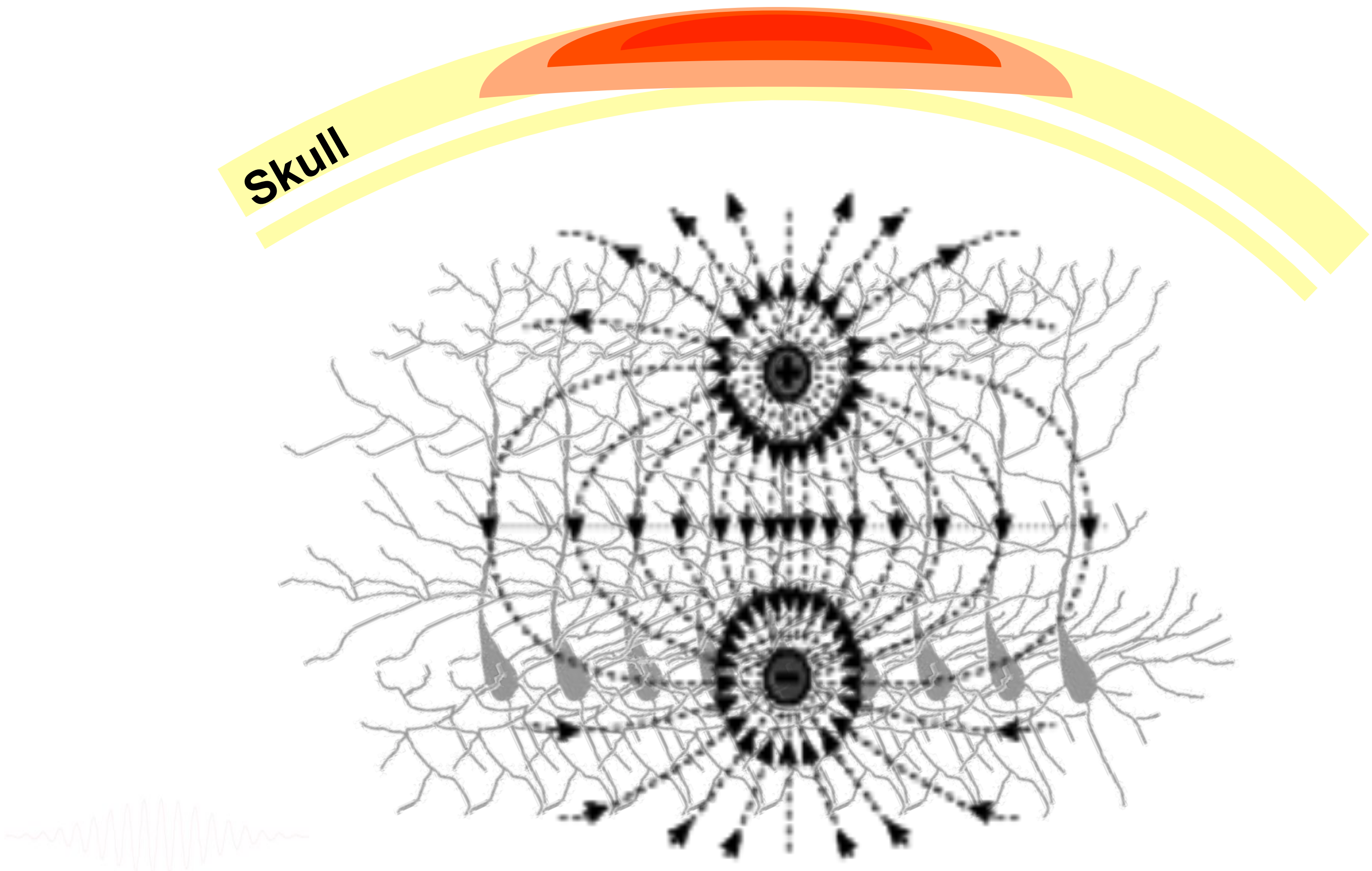


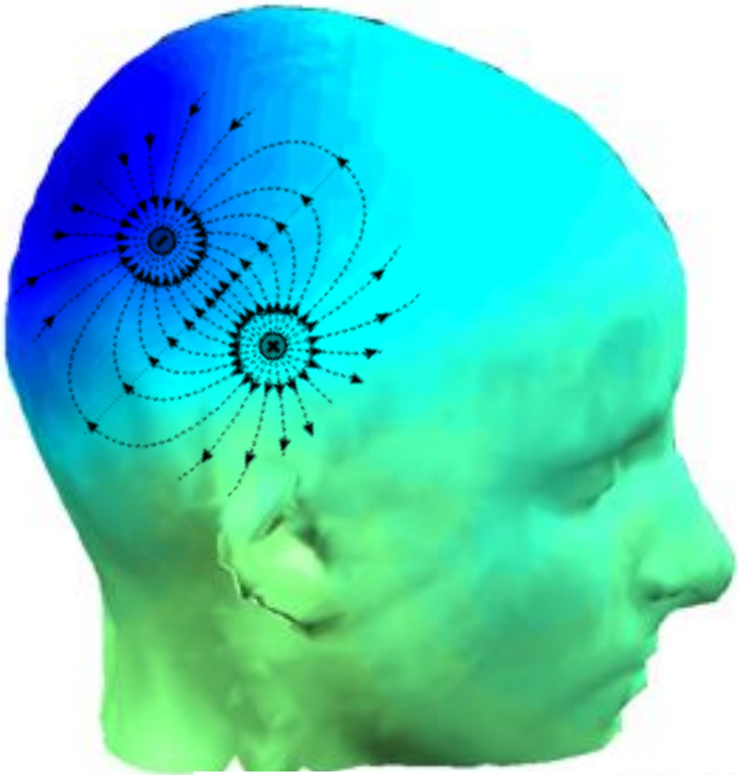
Separating EEG source activities



Independent component analysis (ICA) separates mixed EEG signals at the scalp into temporally independent time courses

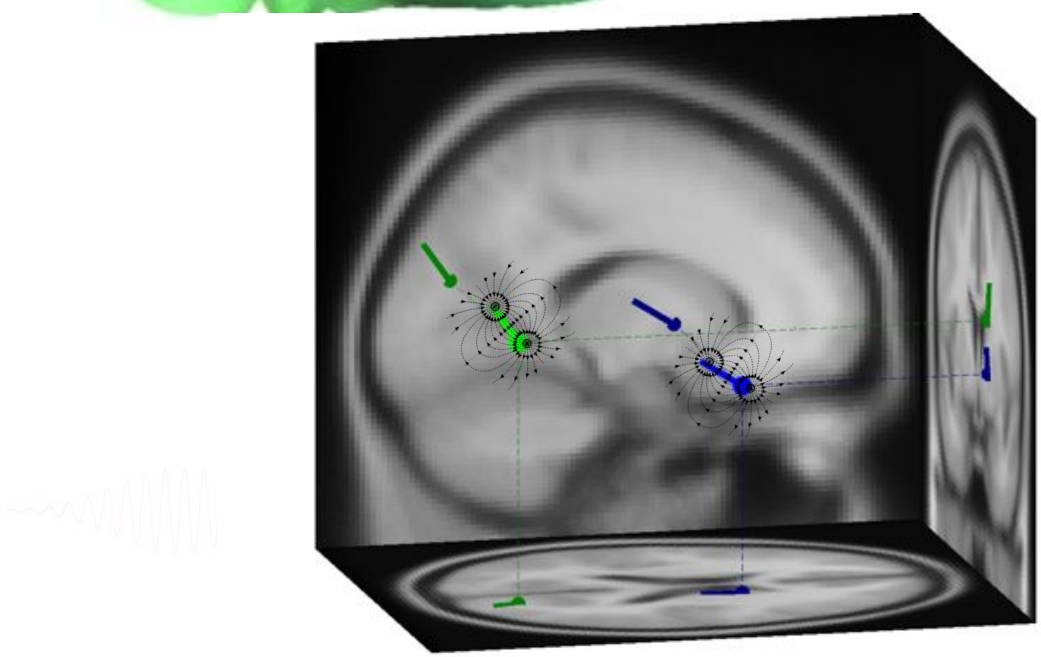
Patch of Cortex Acting as a Dipole

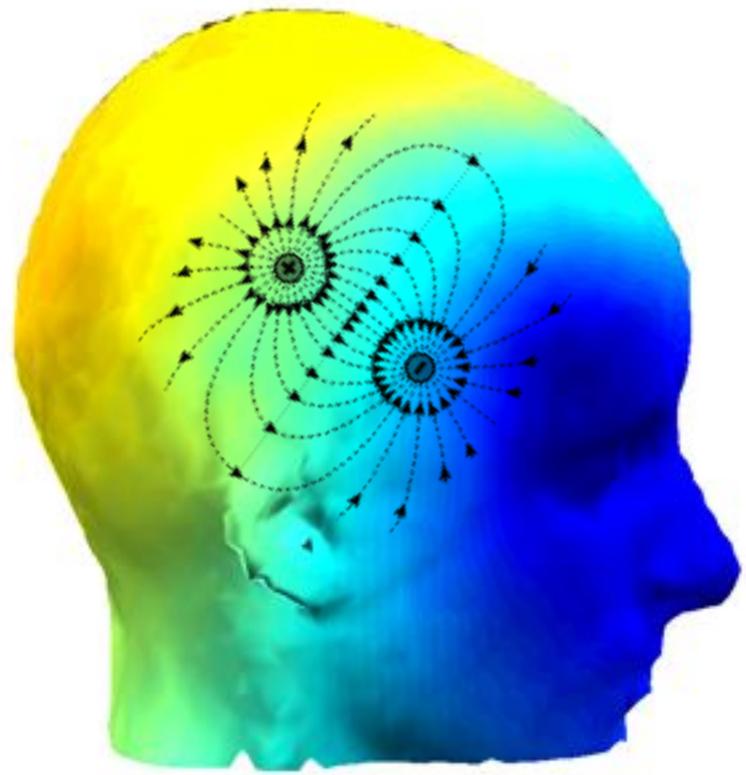


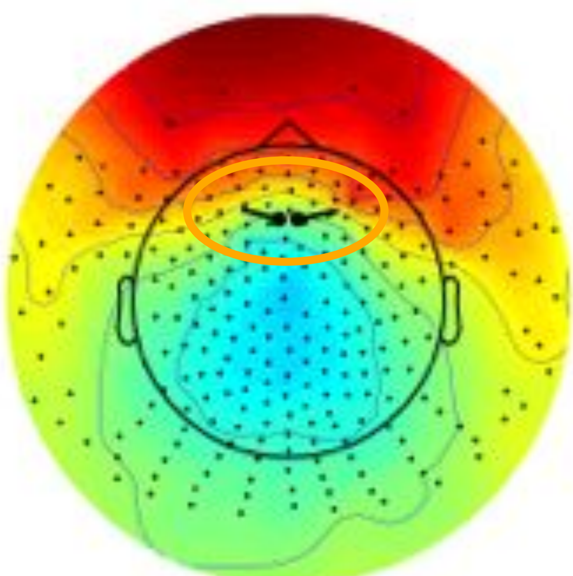


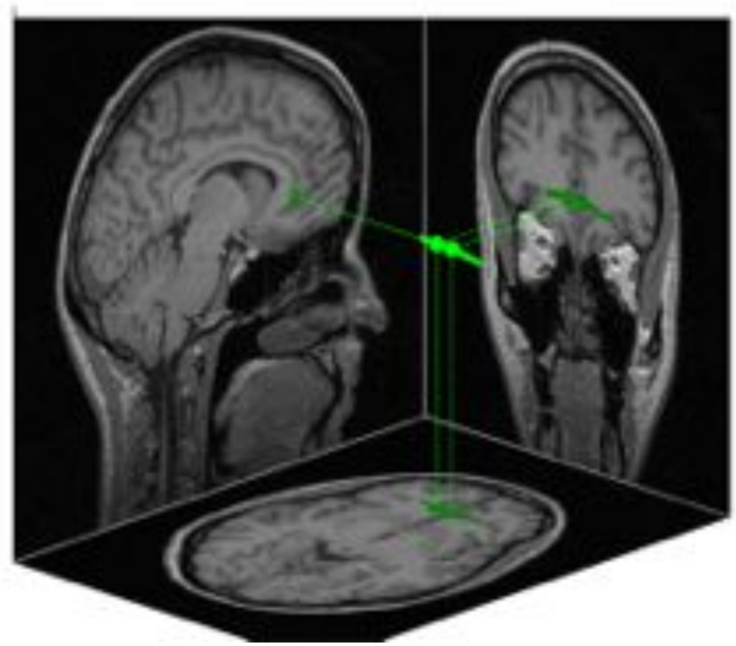
Dipolar Scalp Projections

ICA creates a spatial filter for each temporally independent source

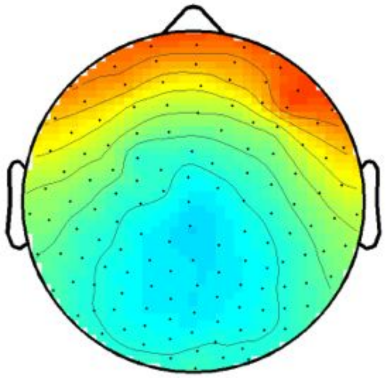




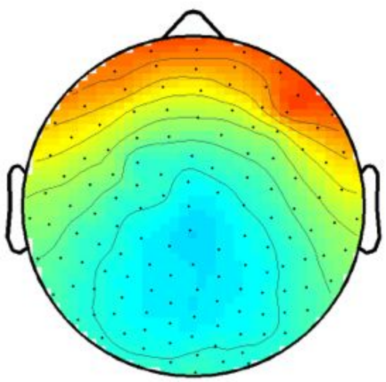


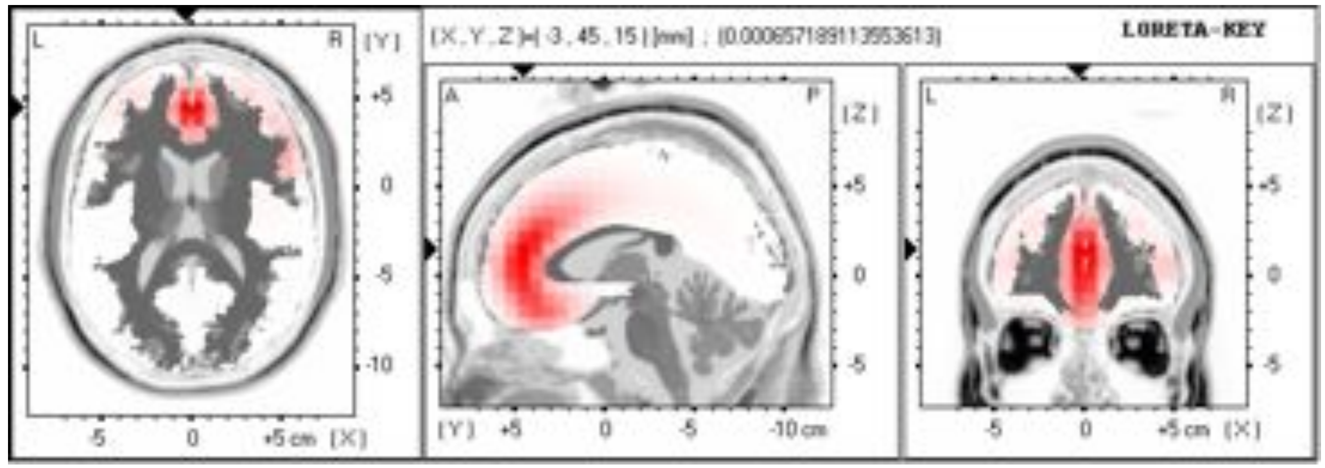


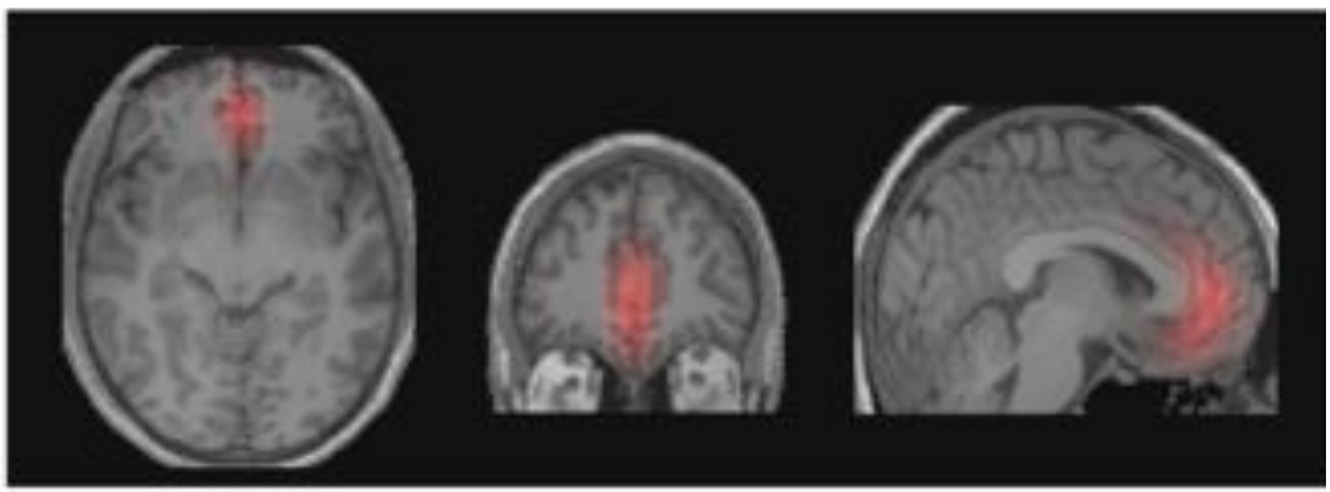
Localization of activity

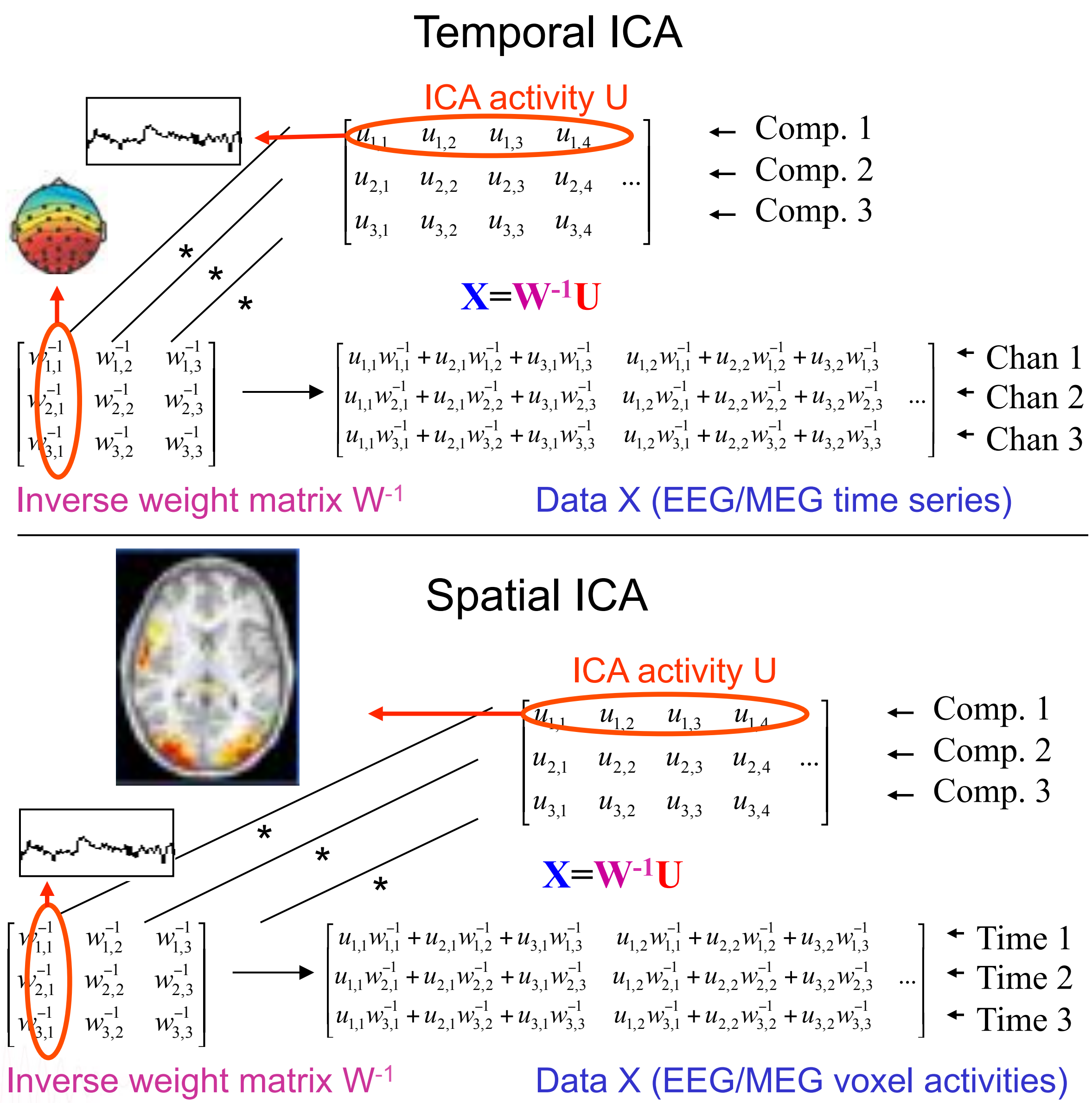


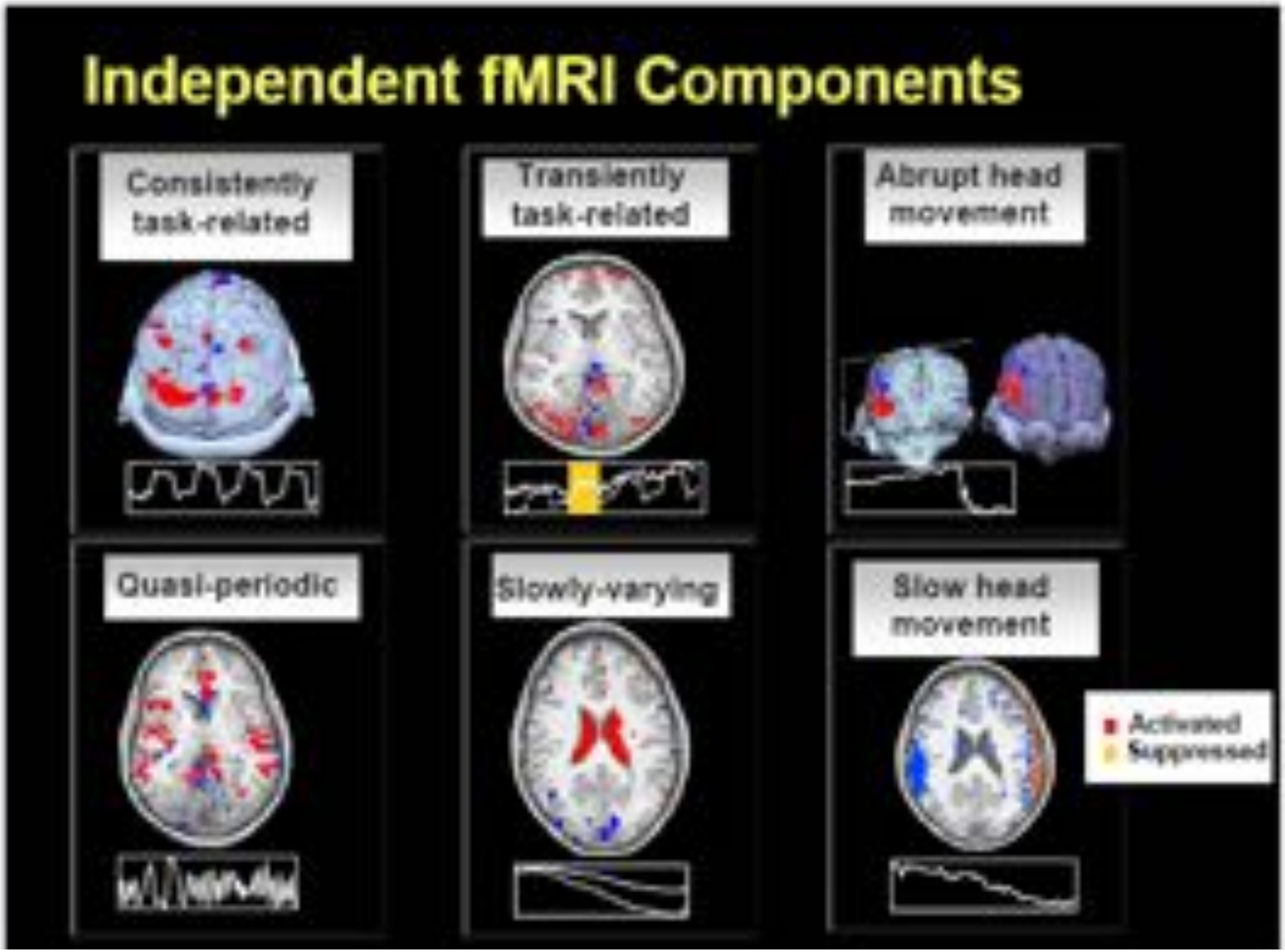
Localization of activity








Temporal ICA
ICA activity U
Comp. 1
Comp. 2
Comp. 3
X=W-1U
Chan 1
Chan 2
Chan 3
Inverse weight matrix W-1
Data X (EEG/MEG time series)
Spatial ICA
ICA activity U
Comp. 1
Comp. 2
Comp. 3
X=W-1U
Time 1
Time 2
Time 3
Inverse weight matrix W-1
Data X (EEG/MEG voxel activities)


Independent fMRI Components
Consistently task-related
Transiently task-related
Abrupt head movement
Quasi-periodic
Slowly-varying
Slow head movement
Activated
Suppressed

The end