

-
-
-
-

Independent Component Analysis and Its Applications



Tzyy-Ping Jung (鍾子平)

Center for Advanced Neurological Engineering
Swartz Center for Computational Neuroscience
Institute for Neural Computation
University of California, San Diego, USA

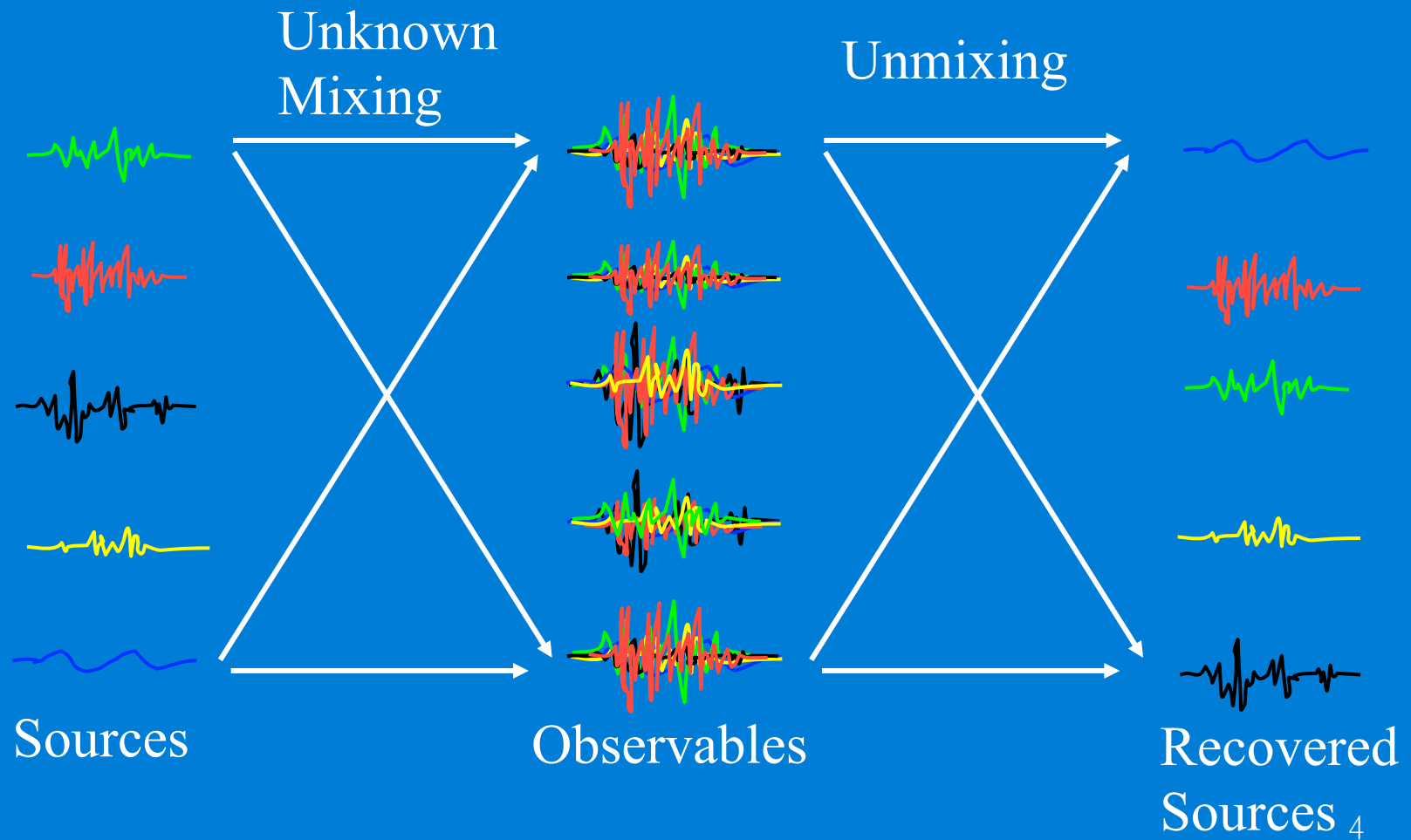
&

Department of Computer Science
National Chiao-Tung University, Hsinchu, Taiwan

Outline

- Blind Source Separation:
 - Solving the “cocktail party problem”
- Applications
 - Speech separation and clarity
 - EEG/ERP
 - fMRI (if time permits)

Blind Source Separation



Brief History of ICA

- Herault & Jutten ("Space or time adaptive signal processing by neural network models", *Neural Nets for Computing Meeting*, Snowbird, Utah, 1986): **Seminal paper, neural network**
- Comon (1994): **Approximation of MI by 4th order statistics**
- Bell & Sejnowski (1995): **Information Maximization**
- Amari et al. (1996): **Natural Gradient Learning**
- Cardoso (1996): **JADE**
- Hyvärinen & Oja: **Nonlinear PCA, FastICA**

- Applications of ICA to biomedical signals
 - EEG/ERP analysis (Makeig, Bell, Jung & Sejnowski, 1996; Jung et al., 1997; Makeig et al., 1997; Jung et al., 2001)
 - fMRI analysis (McKeown, Jung et al. 1998, Jung et al., 2001)
 - ECG analysis (Cardoso 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)

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

Independent Component Analysis

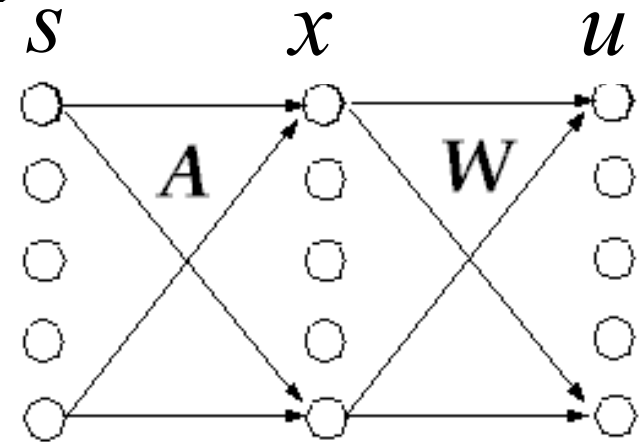
ICA is a method to recover a version, \hat{s} , of the original sources by multiplying the data by a unmixing matrix,

$$\mathbf{u} = \mathbf{W}\mathbf{x},$$

where \mathbf{x} is our observed signals, a linear mixtures of sources,

$$\mathbf{x} = \mathbf{A}\mathbf{s}.$$

While PCA simply decorrelates the outputs (using an orthogonal matrix \mathbf{W}), ICA attempts to make the outputs **statistically independent**, while placing no constraints on the matrix \mathbf{W}



WA after learning:

-4.09	0.13	0.09	-0.07	-0.01
0.07	-2.92	0.00	0.02	-0.06
0.02	-0.02	-0.06	-0.08	-2.20
0.02	0.03	0.00	1.97	0.02
-0.07	0.14	-3.50	-0.01	0.04

Statistical Independence

Statistical Independence:

$$f_{\mathbf{s}}(\mathbf{s}) = \prod_{i=1}^N f_{s_i}(s_i)$$

Or the mutual information:

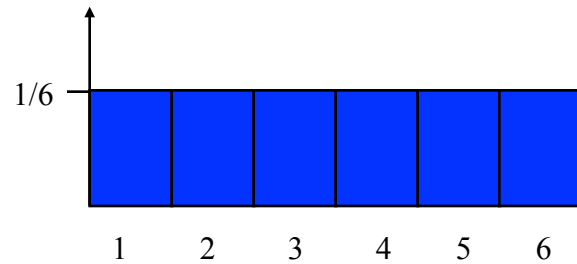
$$I(s_i, s_j) = E \left[\ln \frac{f_{\mathbf{s}}(\mathbf{s})}{\prod_{i=1}^N f_{s_i}(s_i)} \right] = 0, \text{ for } \forall i \neq j$$

The problem of blind separation is to find \mathbf{W} such that the linear transformation $\mathbf{u} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s}$ reestablishes the condition of statistical independence.

Entropy

$$H(X) = - \sum_{x \in X} p(x) \log(p(x))$$

Dice: 1/6



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

ICA learning rule

How to make the outputs statistically independent?


Minimize their redundancy or mutual information.

Entropy:
$$H(X) = - \sum_{x \in X} p(x) \log(p(x))$$

Joint entropy
$$H(X, Y) = - \sum_{(x, y) \in X \times Y} p(x, y) \log(p(x, y))$$

Mutual Information
$$I(y_1, y_2) = H(y_1) + H(y_2) - H(y_1, y_2)$$

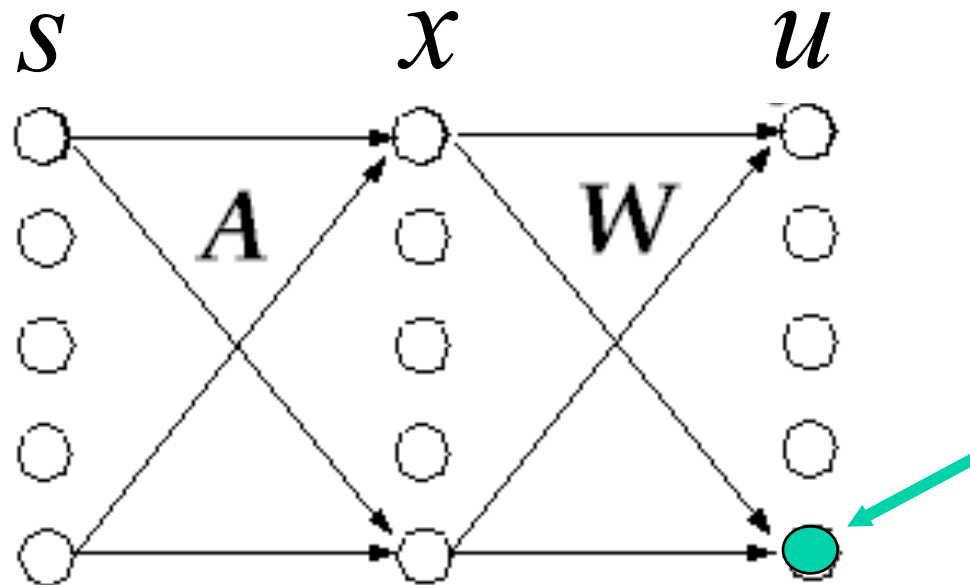
Minimizing $I(y_1, y_2) \rightarrow$ **Maximizing $H(y_1, y_2)$**


=0 if the two variables
are independent


ICA learning rule

$$\Delta W = \frac{\partial H(y_1, y_2, \dots)}{\partial W} \underbrace{W^T W}$$

Independent Component Analysis

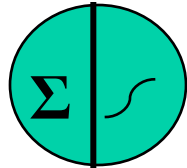


ICA is a method to recover a version, of the original sources by multiplying the data by a unmixing matrix,

$$\mathbf{u} = \mathbf{W}\mathbf{x}$$

InfoMax (Bell & Sejnowski, 1995)

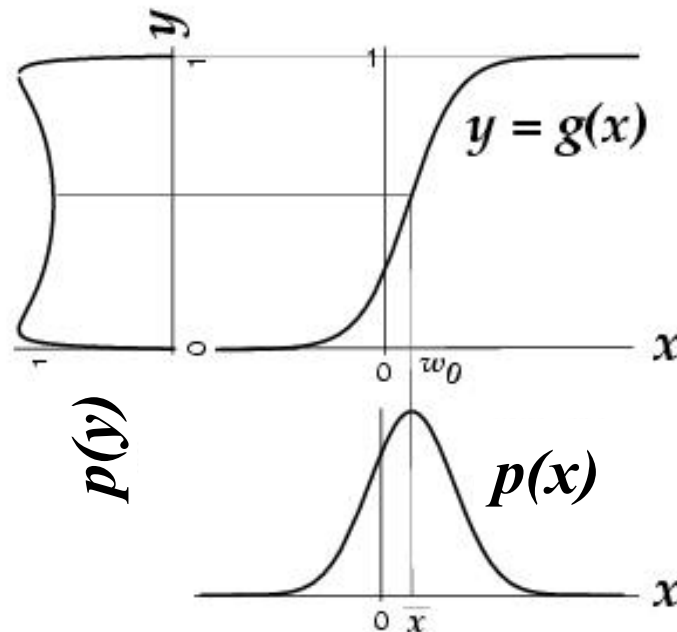
To make the u_i independent, we need to operate on non-linear transformed output variables, $y = g(u)$, such as



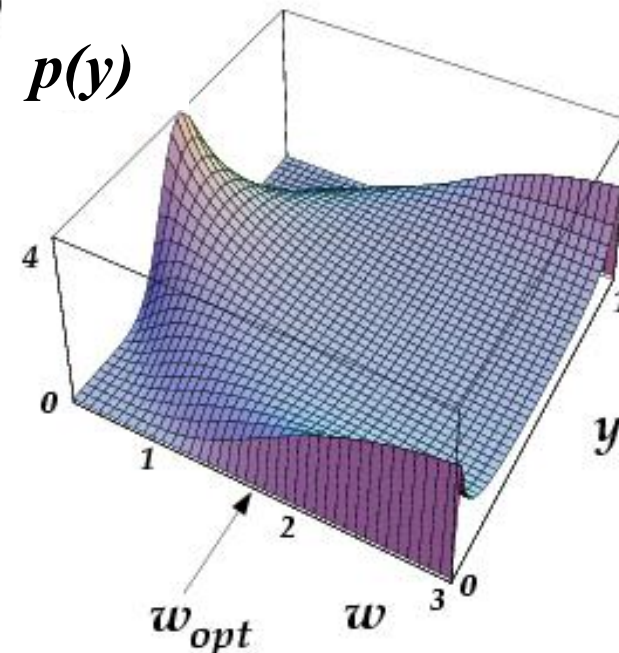
$$y = \frac{1}{1 + e^{-u}}, \quad u = \mathbf{W}x + w_0$$

The non-linear function provides all the higher-order statistics necessary to establish independence.

(a)



(b)



15

From Bell & Sejnowski *Neural Compu.* 1995.

ICA learning rule

The learning rule:

$$\Delta \mathbf{w} \propto \frac{\partial H(\mathbf{y})}{\partial \mathbf{w}} \mathbf{w}^T \mathbf{w} = [\mathbf{I} + \phi \mathbf{u}^T] \mathbf{w},$$

where $\phi_i = (\partial / \partial u_i) \ln(\partial y_i / \partial u_i)$.

For super-Gaussian,

$$\phi_i = 1 - 2y_i \text{ (for logistic nonlinearity).}$$

For sub- and/or super-Gaussian,

$$\phi_i = \begin{cases} + \tanh(u_i) - u_i & \text{kurtosis} < 0 \\ - \tanh(u_i) - u_i & \text{kurtosis} > 0 \end{cases}$$

- Remove the mean

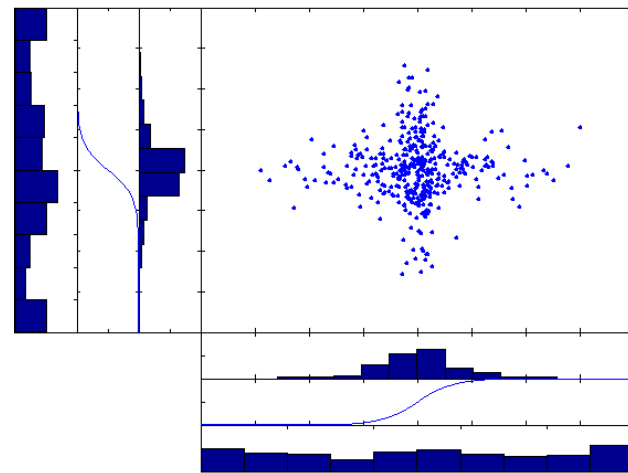
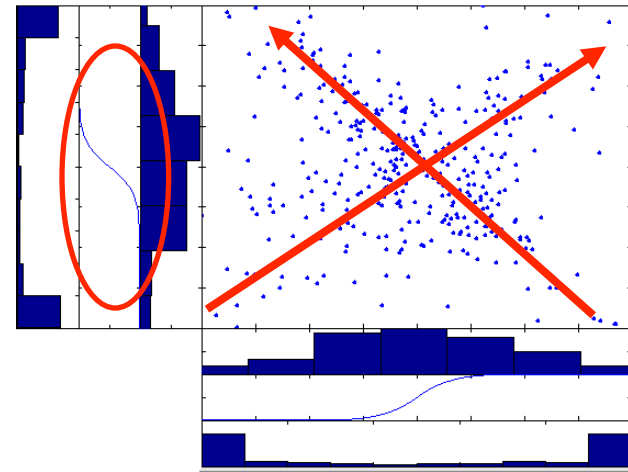
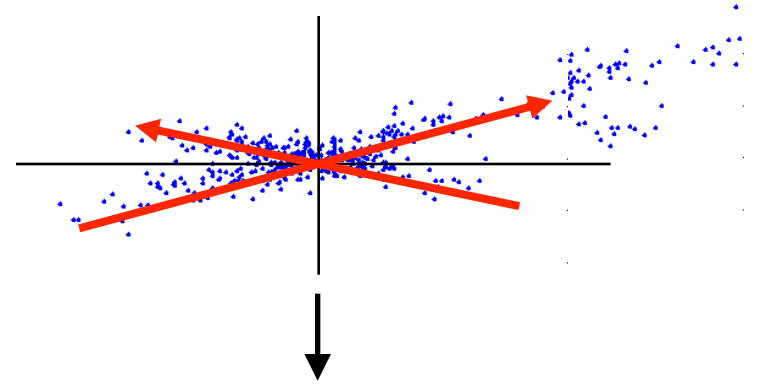
$$\mathbf{x} = \mathbf{x} - \langle \mathbf{x} \rangle.$$

- 'Sphere' the data by diagonalizing its covariance matrix,

$$\mathbf{x} = 2 \langle \mathbf{x} \mathbf{x}^T \rangle^{-1/2} (\mathbf{x} - \langle \mathbf{x} \rangle).$$

- Update \mathbf{W} according to

$$\Delta \mathbf{W} \propto \frac{\partial H(\mathbf{y})}{\partial \mathbf{W}} \mathbf{W}^T \mathbf{W} = [\mathbf{I} + \phi \mathbf{u}^T] \mathbf{W}.$$

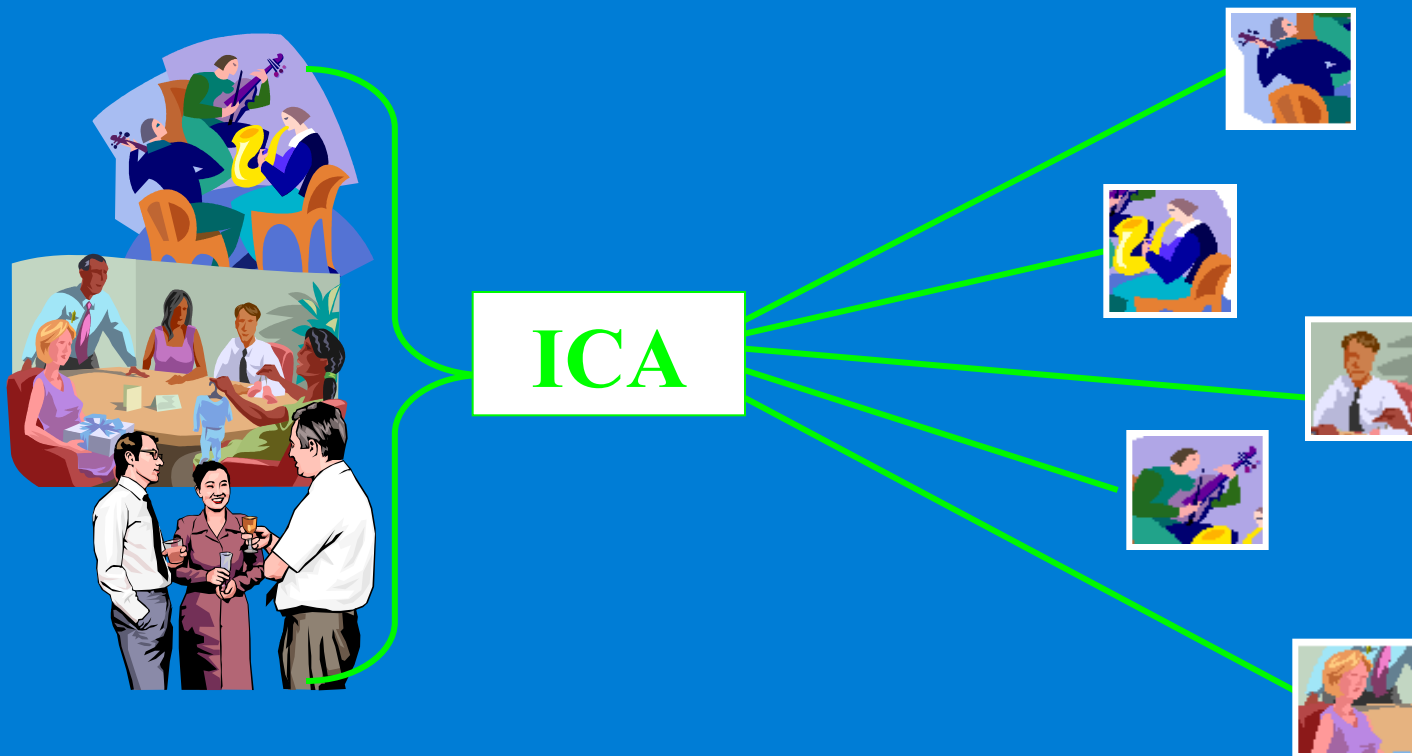


•
•
•

ICA Applications

- Speech enhancement (noisy speech recognition)
- Biomedical signal processing (EEG, ERP, fMRI, MEG)
- Image processing

Example: Speech Separation



Sound Separation Demonstration



Car Kit Demonstration

March 8, 2005

 **VOICE**

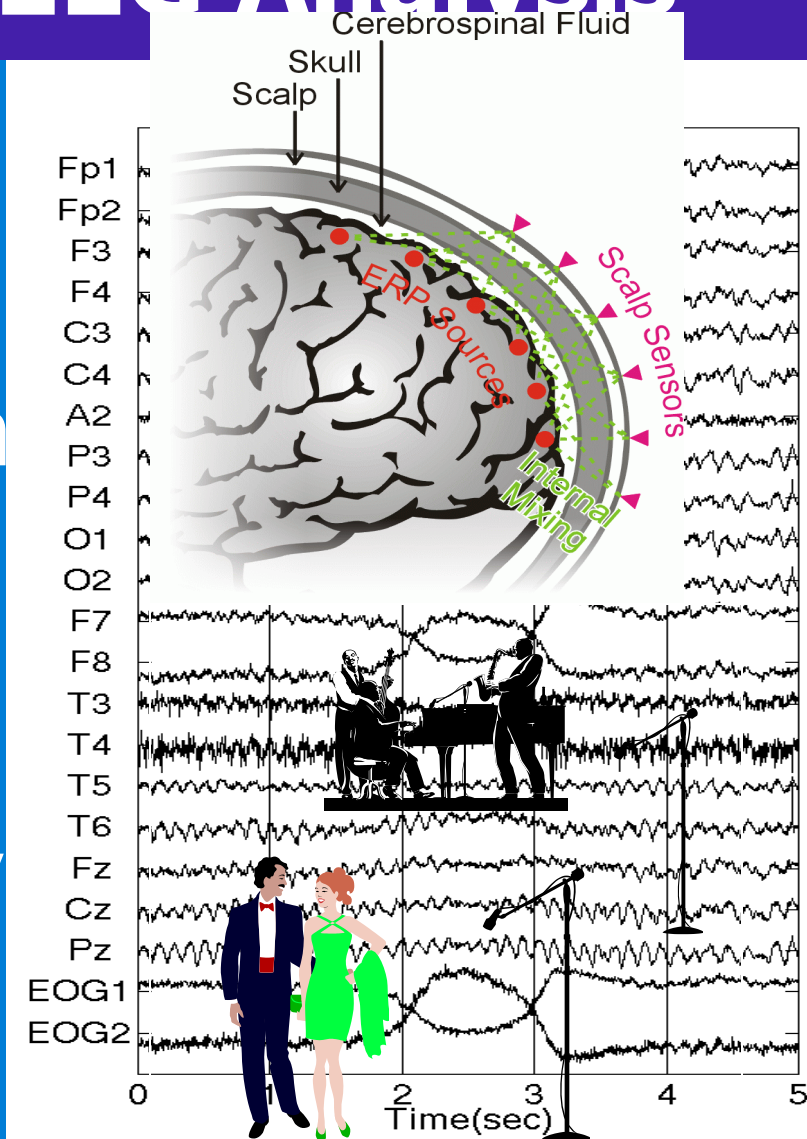
•
•
•

ICA Applications

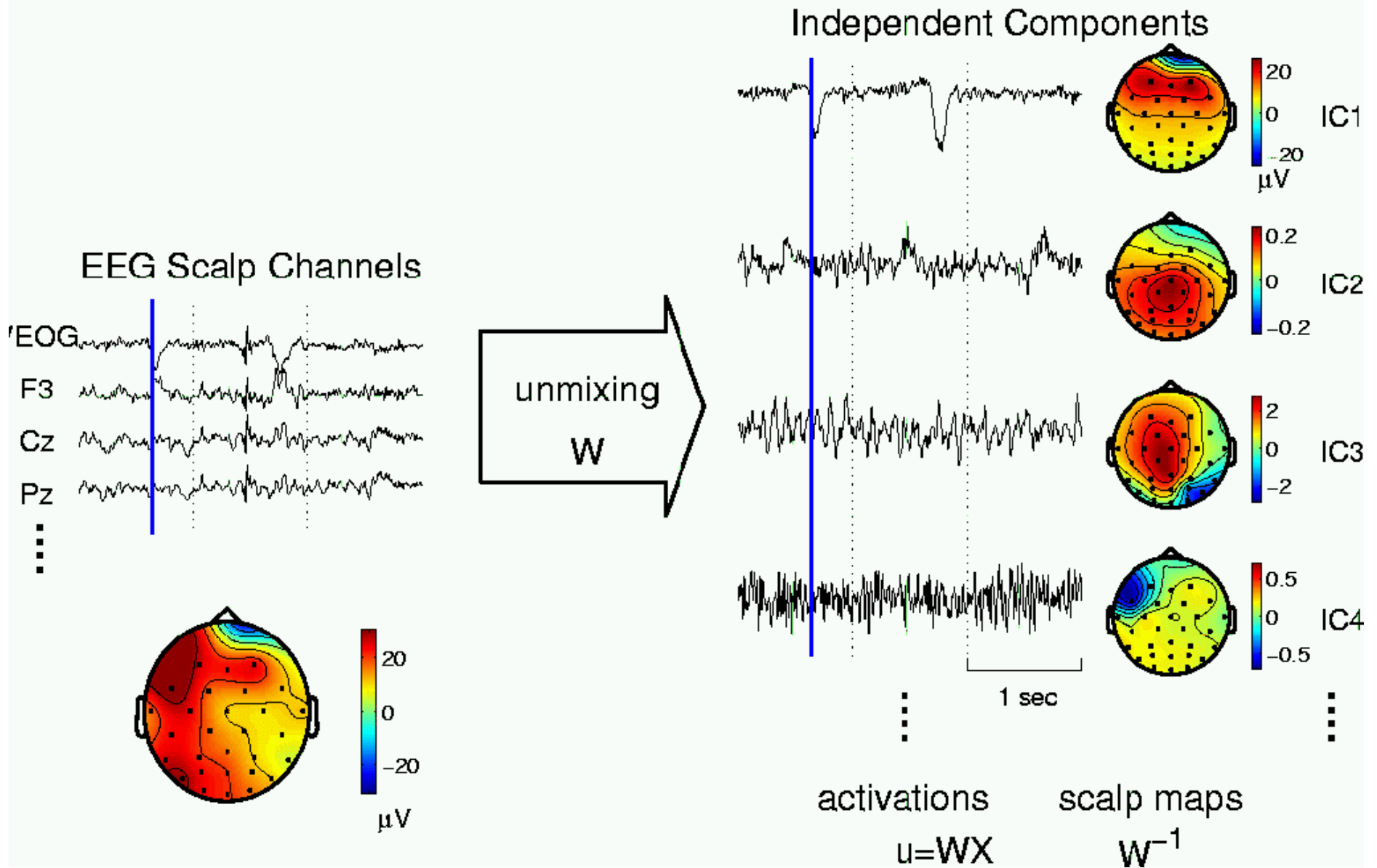
- Speech enhancement (noisy speech recognition)
- Biomedical signal processing (**EEG**, **ERP**, **fMRI**, MEG)
- Image processing

Challenges of EEG Analysis

- Pervasive artifacts
- EEG recordings are mixtures of all brain activities arising from different networks
- Response variability
- Inverse problem
- etc



ICA decomposition

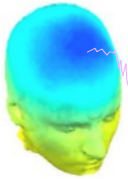


From Jung et al., *Clinical Neurophysiology*, 2000.

ICA/EEG Assumptions

- Mixing is linear at electrodes
- Propagation delays are negligible
- Component time courses are independent
- Number of components \leq number of channels.

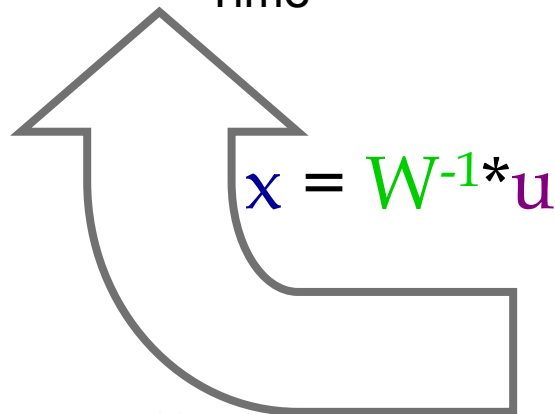
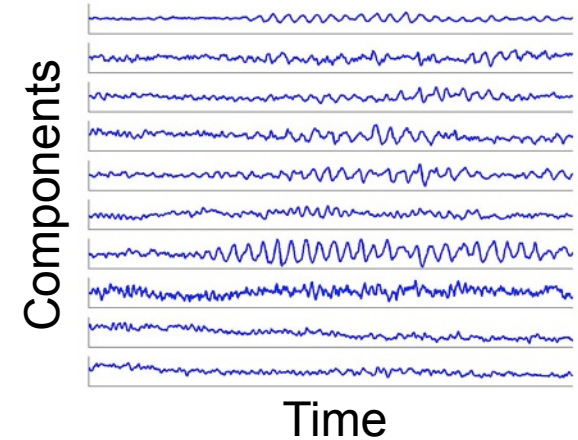
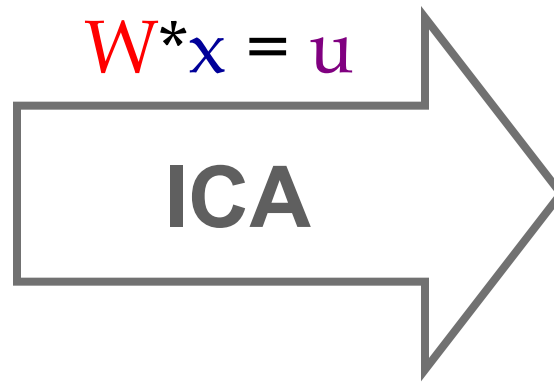
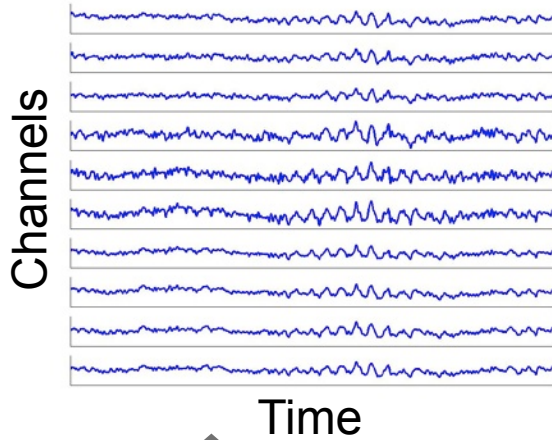
Independent Component Analysis



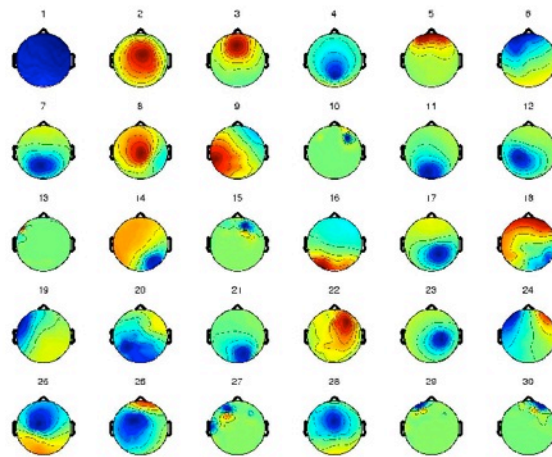
$x = \text{scalp EEG}$

$W = \text{unmixing matrix}$

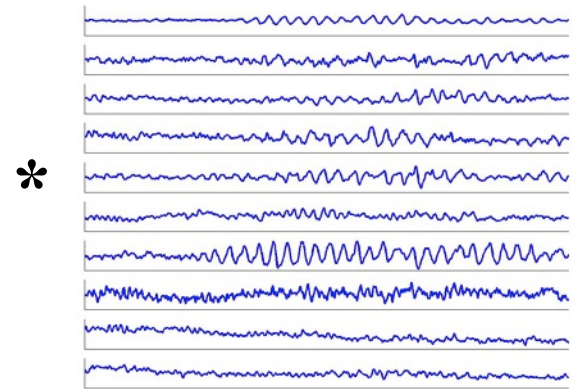
$u = \text{sources}$



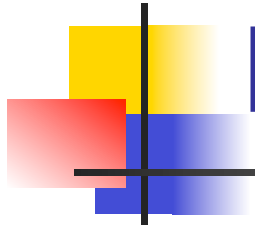
W^{-1} (scalp projections)



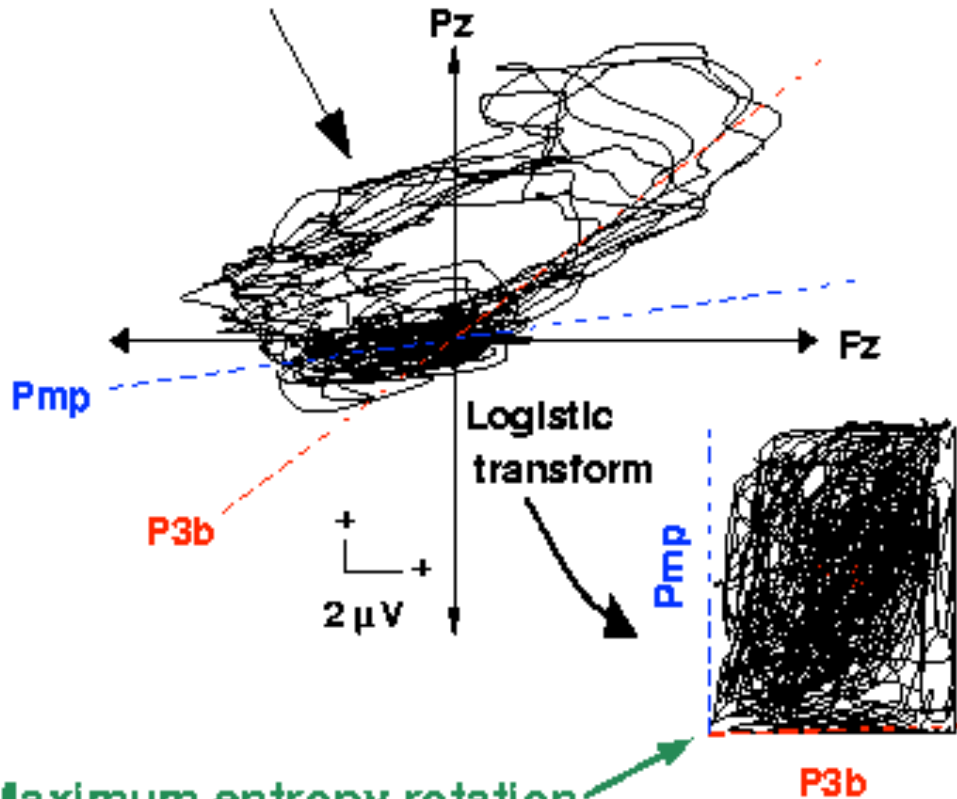
$u = \text{sources}$



Independent components of EEG/ERP



10 Target responses



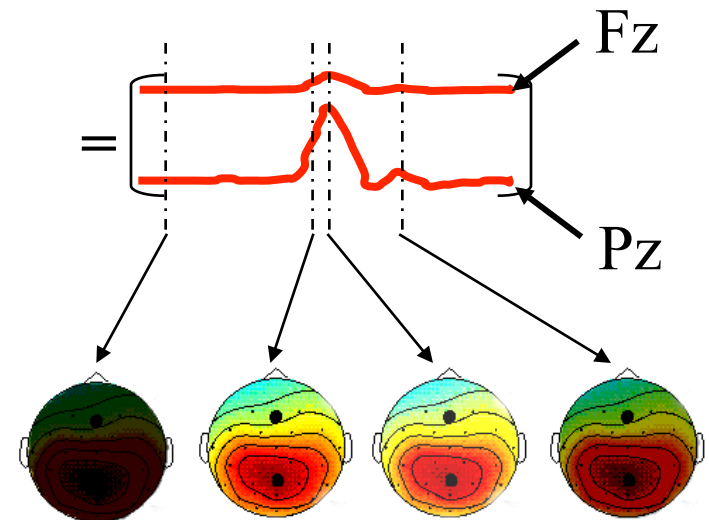
Maximum entropy rotation

From Makeig et al., *JNS*, 1999.

How is ICA scalp map plotted?

$$\mathbf{X} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} \text{Red waveform} \\ \text{Blue waveform} \end{bmatrix}$$

$$\mathbf{X}_{P3b} = \begin{bmatrix} 0.4 & A_{12} \\ 8.2 & A_{22} \end{bmatrix} \begin{bmatrix} \text{Red waveform} \\ \text{Blue waveform} \end{bmatrix}$$



From Jung & Makeig, 2009.

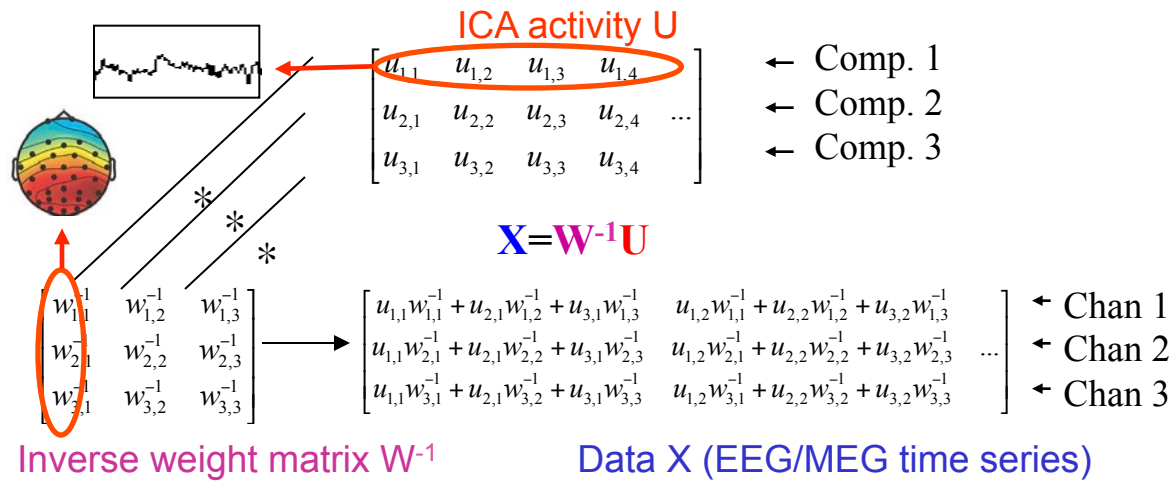
Frequently Asked Questions

- **What is temporal and spatial ICA?**

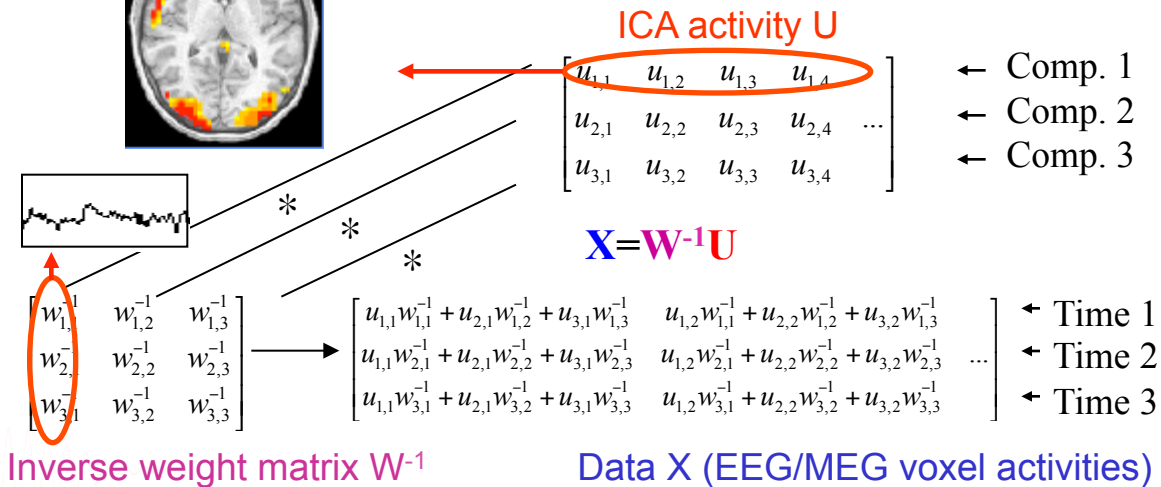
For EEG, we are looking at temporally independent brain activities arising from different brain networks.

For fMRI, the independence is considered over voxels because of brain modularity. i.e.,
Simplistically, "Different places do different things."

Temporal ICA



Spatial ICA



Frequently Asked Questions (cont.)

- **How much data is enough data?**

There is no fixed limit to the number of points needed for a "good" ICA solution - and in fact no fixed way to judge whether an ICA solution is "good" or not.

Frequently Asked Questions (cont.)

- **Pre-ICA procedures**
 - Check the rank of the data (if not full rank, use PCA)
 - 'Messy' channels or epochs should be removed
 - Ultra-low frequency activity should be removed, including the DC offset (a.k.a. remove baseline')
- **Check ICA solution prior to further analysis**
 - Review component scalp maps and check their 'dipolarity'
 - If component maps are 'messy', remove 'messy' epochs/channels and try again...

Frequently Asked Questions (cont.)

- How should the activations be scaled?

$$U=WX, X=W^{-1}*U$$

The strength of source activity is distributed between the columns of W^{-1} and the rows of U .

- Ordering of ICs

Not well-defined and intuitive.

- Can ICA separate 'correlated' source activities?

Practical Issues with ICA of EEG/ERP

1. Apply ICA to averaged ERPs

- How many time points are needed for training?
Suggestion: At least several times number of variables in the unmixing matrix.
- Which EEG processes may express their independence in the ERP training data?
Suggestion: Decompose the concatenated collection of ERP averages in response to the experimental stimulus and task conditions.
- ICA decomposition of averaged ERPs must be interpreted with caution.

Practical Issues with ICA of EEG/ERP

2. Apply ICA to continuous EEG data

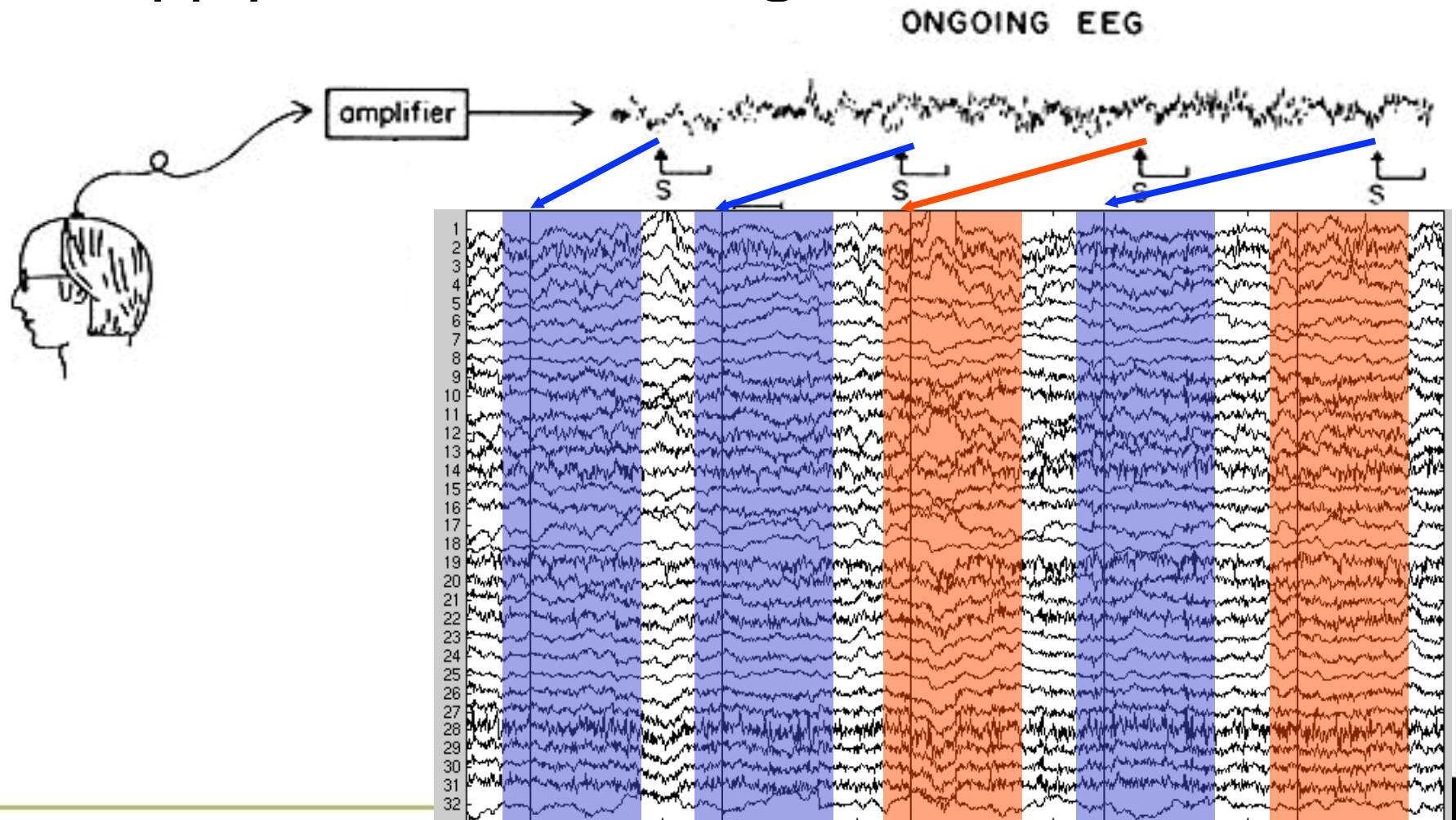
- Are components spatially stationary through time?

Suggestion: Perform separate decompositions of subsets of the recorded data, each consisting of periods during which the sources may be stationary.

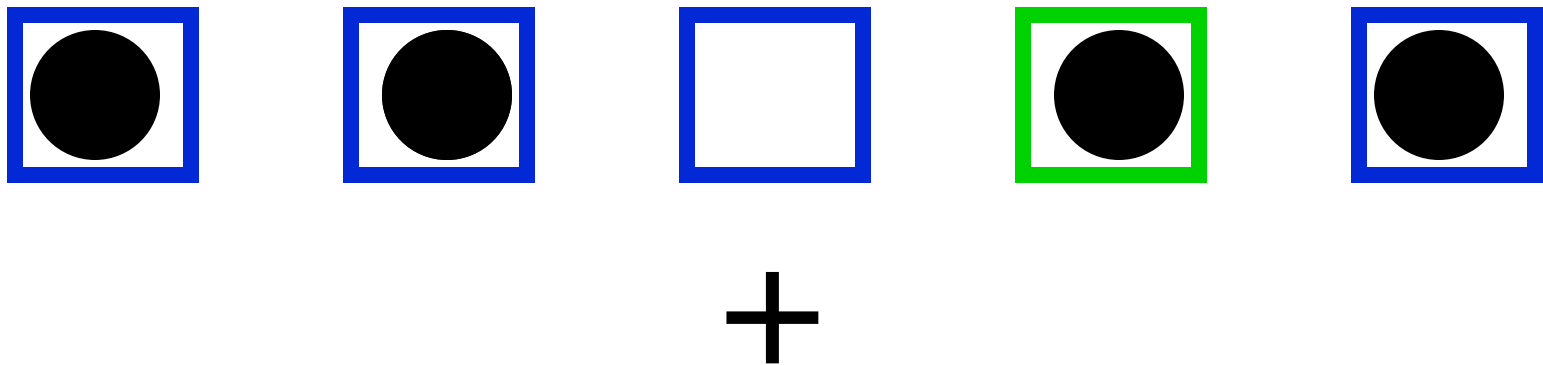
Or, you can use a mixture of ICA model.

Practical Issues with ICA of EEG/ERP

3. Apply ICA to unaveraged event-related EEG



Experiment

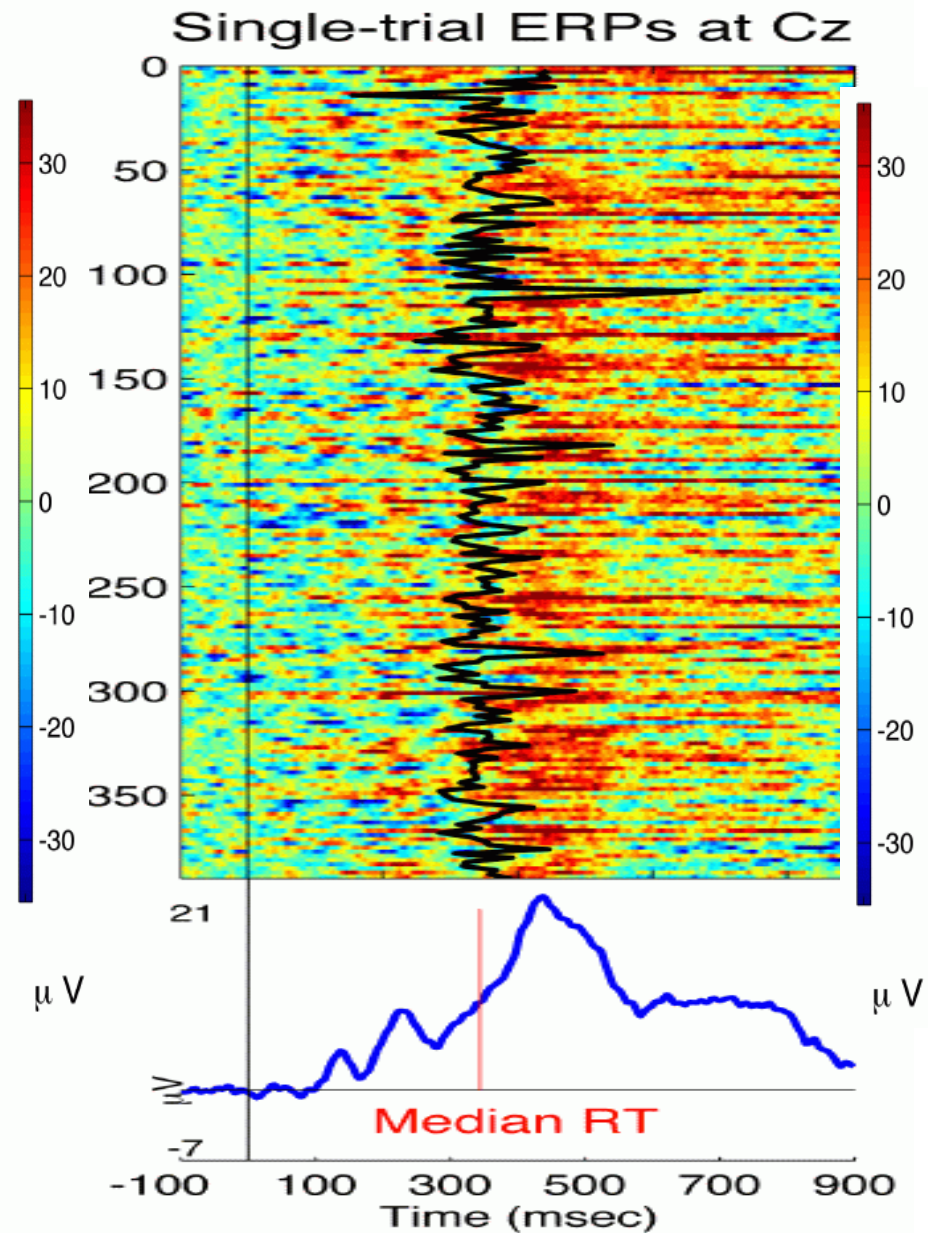
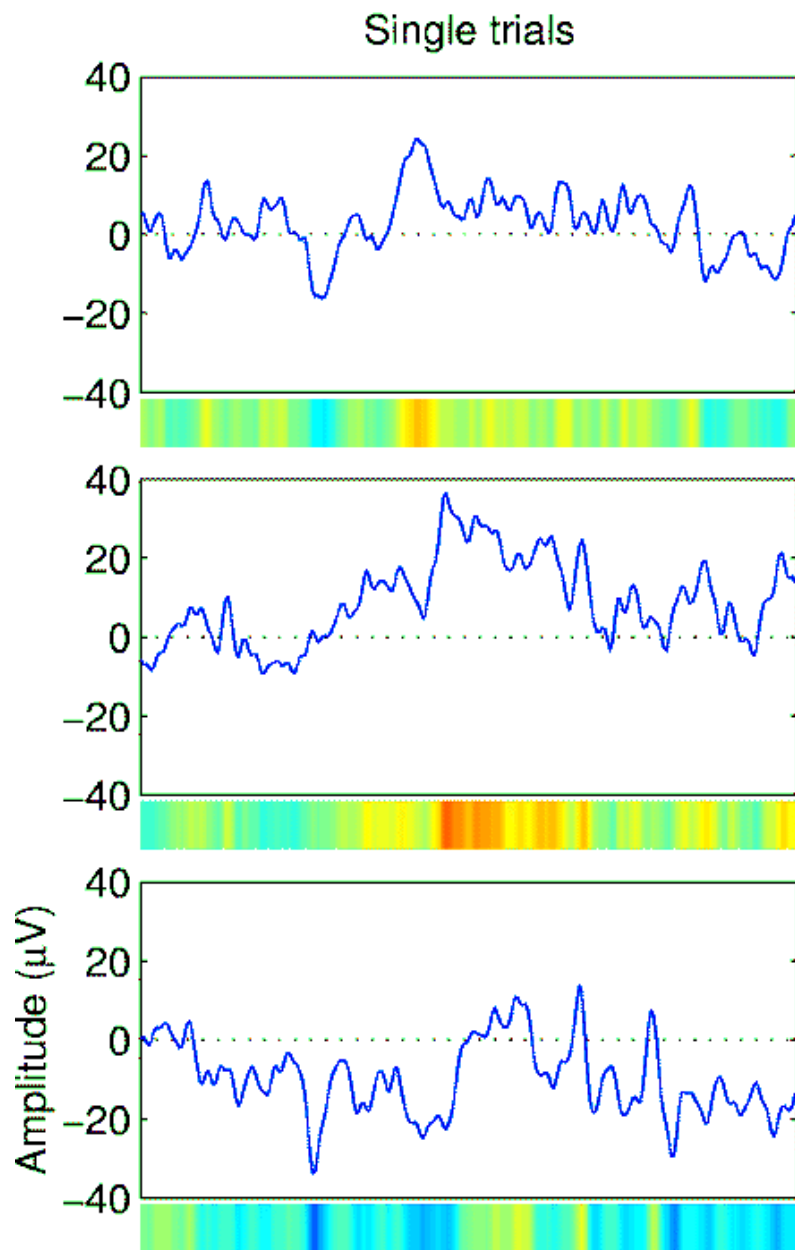


Task: Fixate **cross** while covertly attending to **green box**. Press button when **circle** is flashed in green box.

Subject: 28 normal control, 14 autistic and 8 cerebellar lesion subjects.

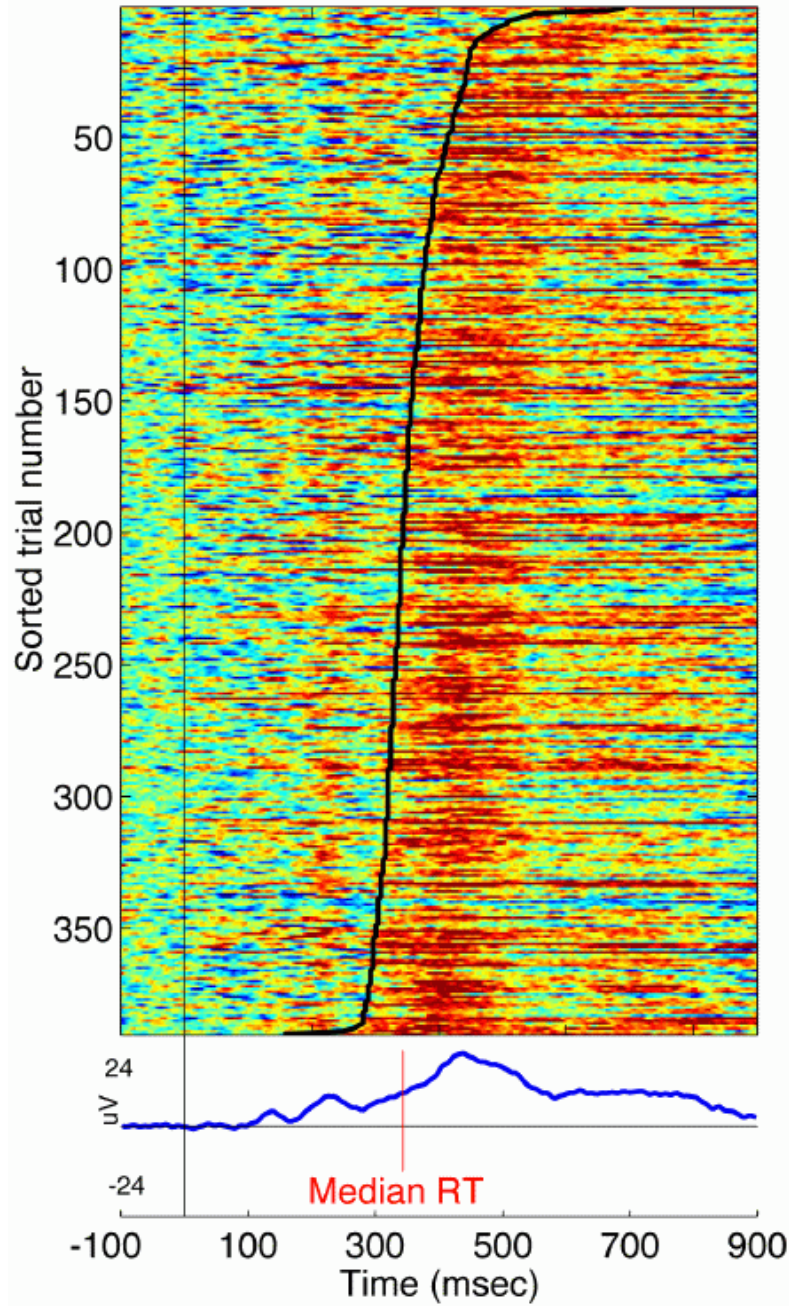
Session: 30 72-s task blocks, including 120 **targets** and 480 nontargets in each of the 5 locations.

ERP Image

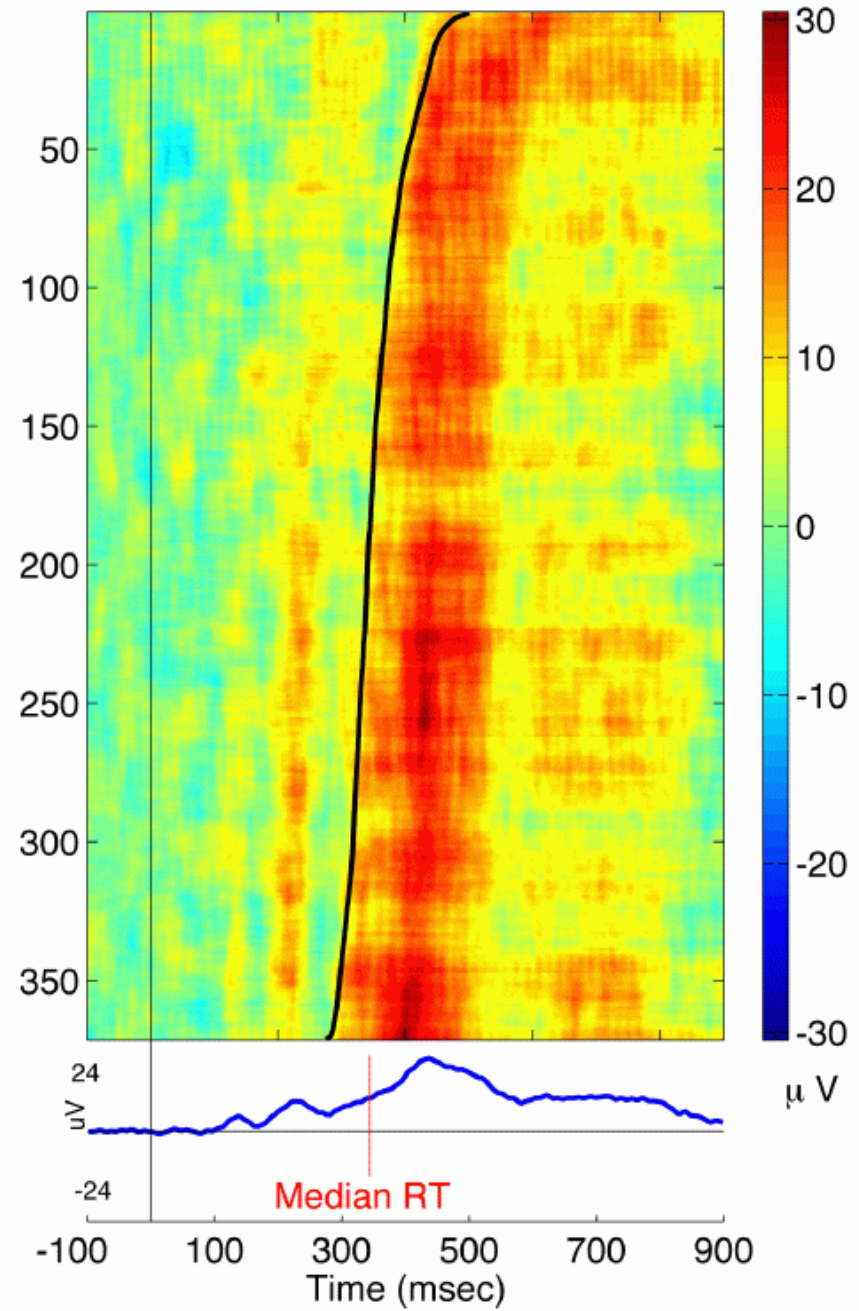


From Jung et al., *NIPS*, 1999.

Single-trial ERPs at Cz



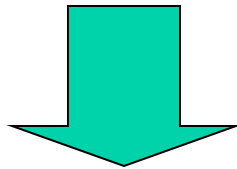
Single-trial ERPs at Cz



From Jung et al., *NIPS*, 1999.

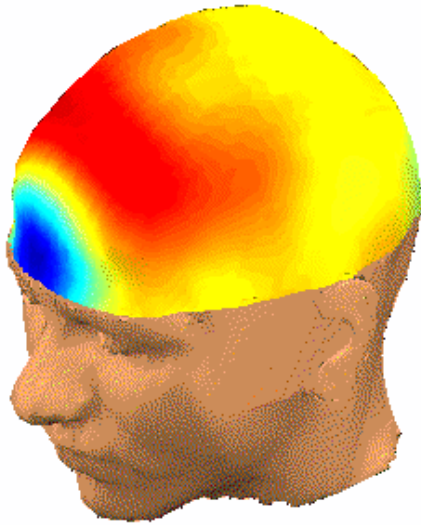
Analysis of Single-trial ERPs

ICA applied to ~600 (single-subject, 31-channel, 1-s) concatenated single-trial response epochs timelocked to detected **target stimuli**

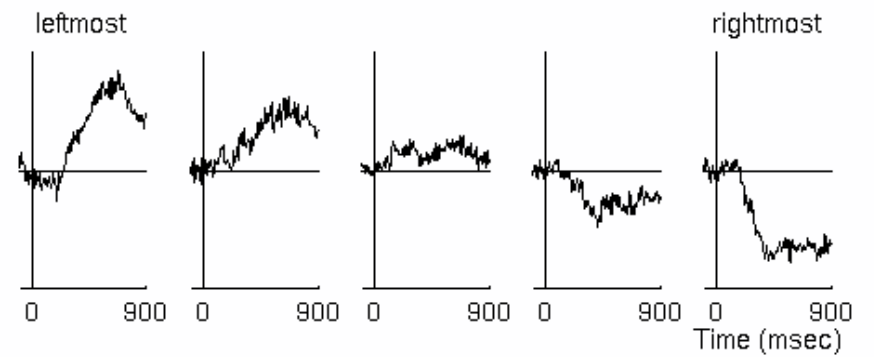
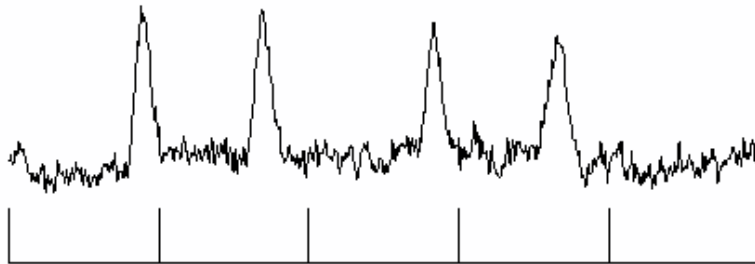
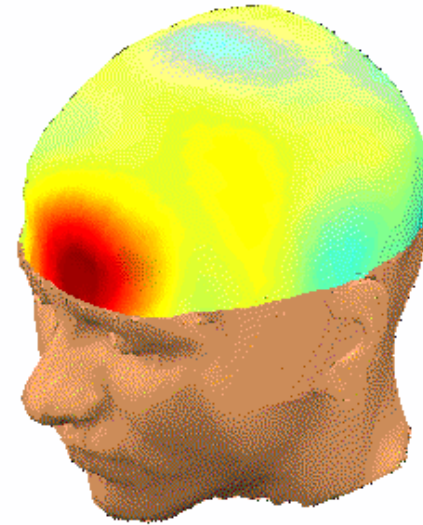


- 31** independent components having:
- **fixed spatial projections** to the scalp
 - **temporally independent time courses** of activation

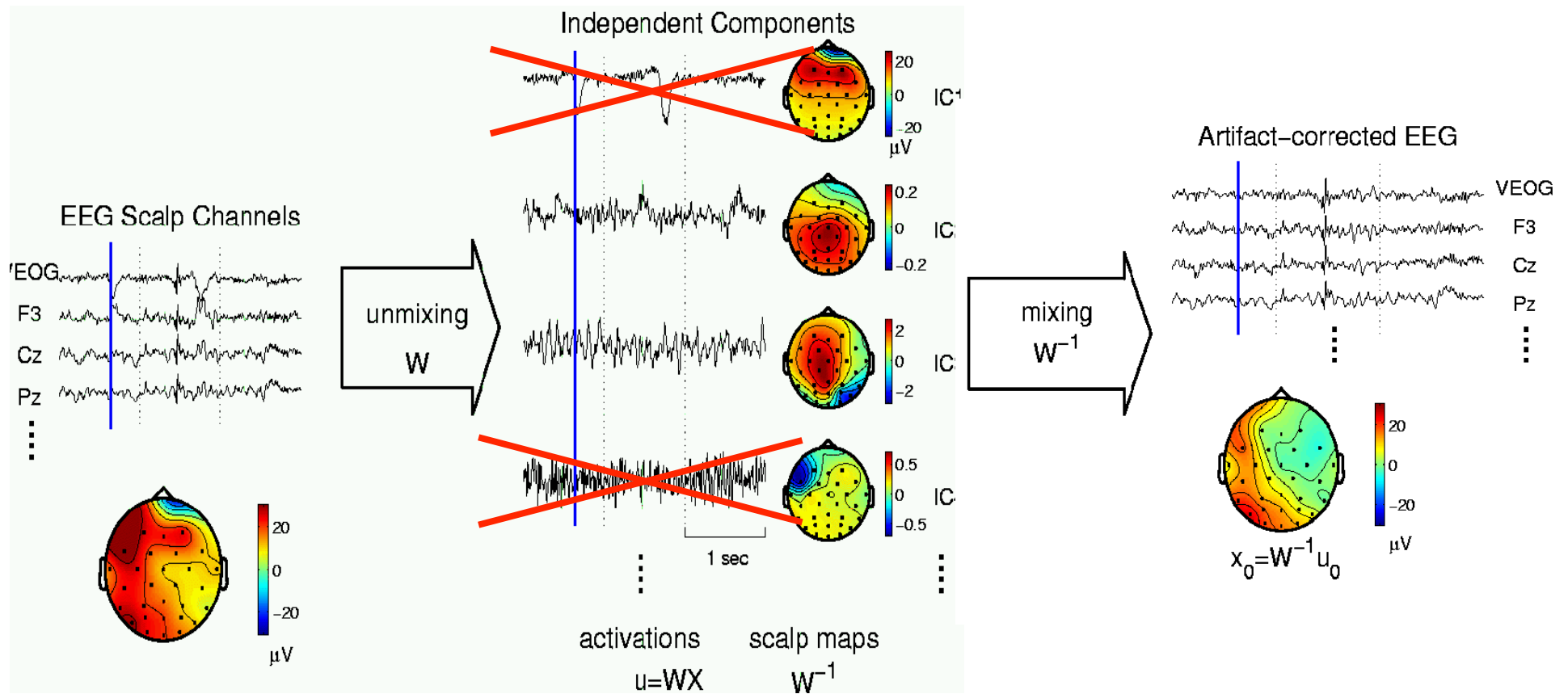
Component 1



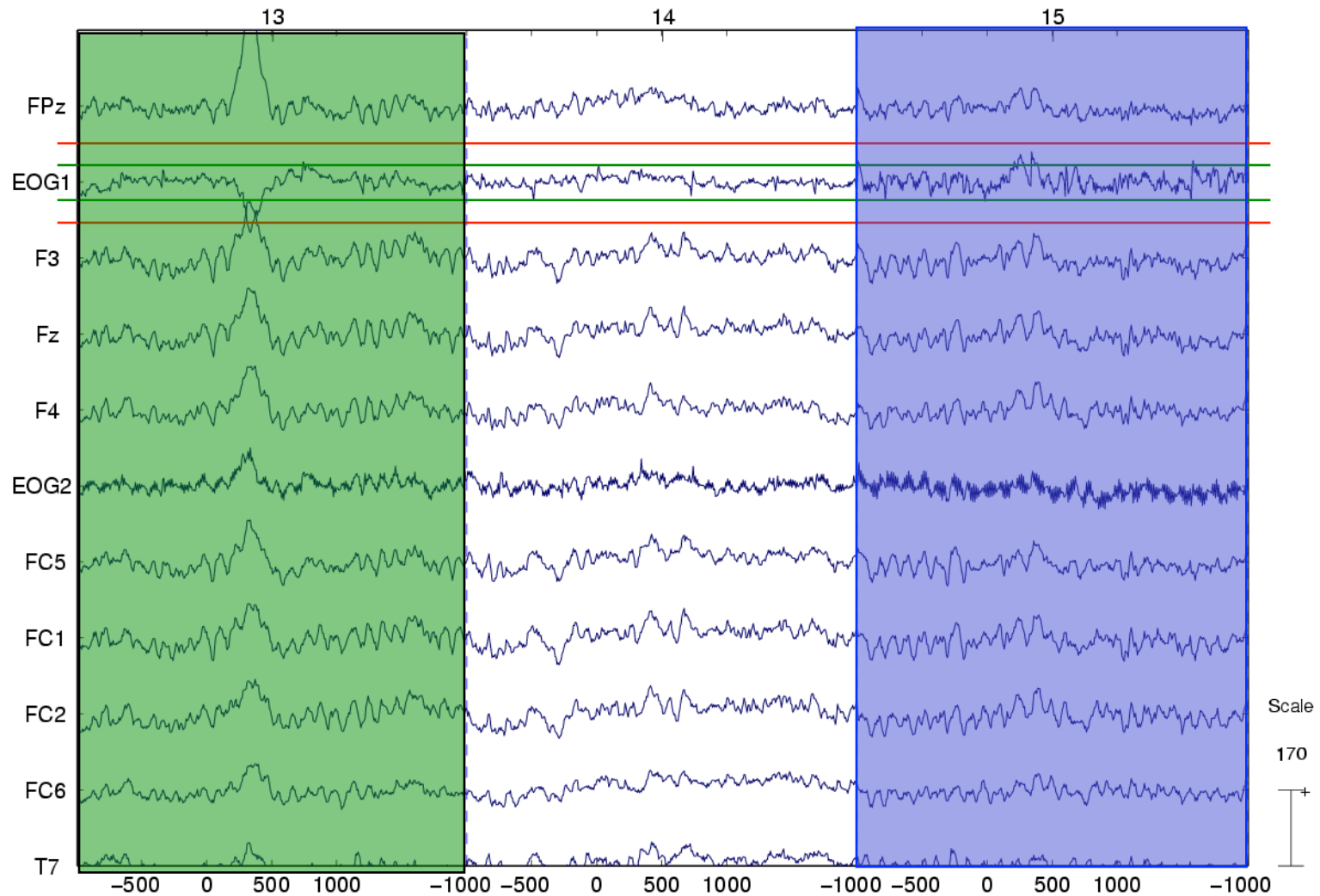
Component 2



ICA-based Artifact Correction



Split Single Trials based on EOG

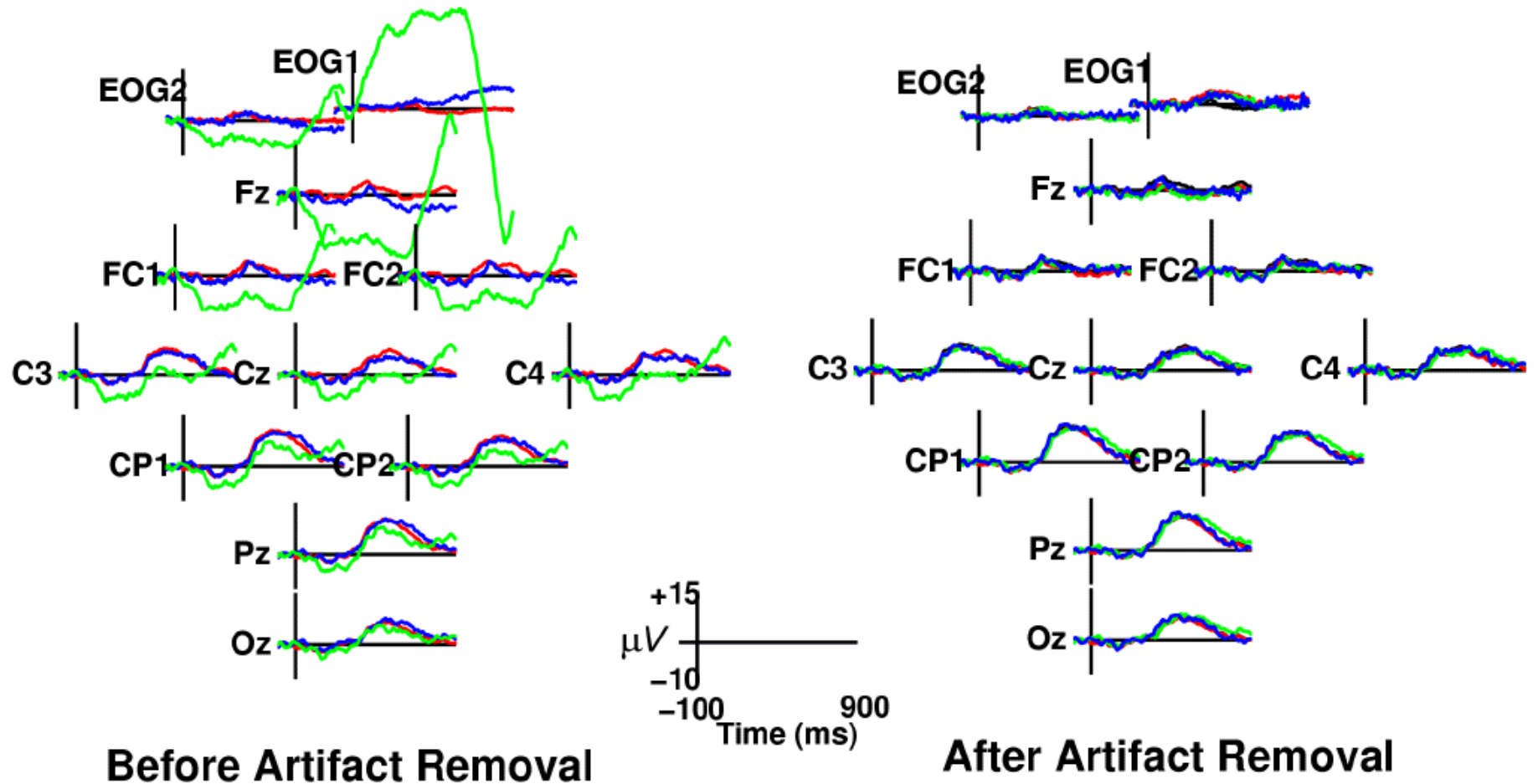


**Heavily
contaminated**

Clean trials

Contaminated

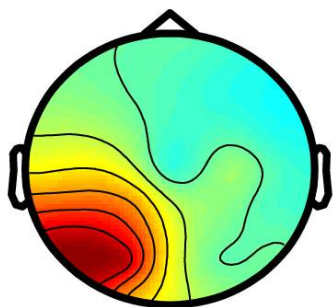
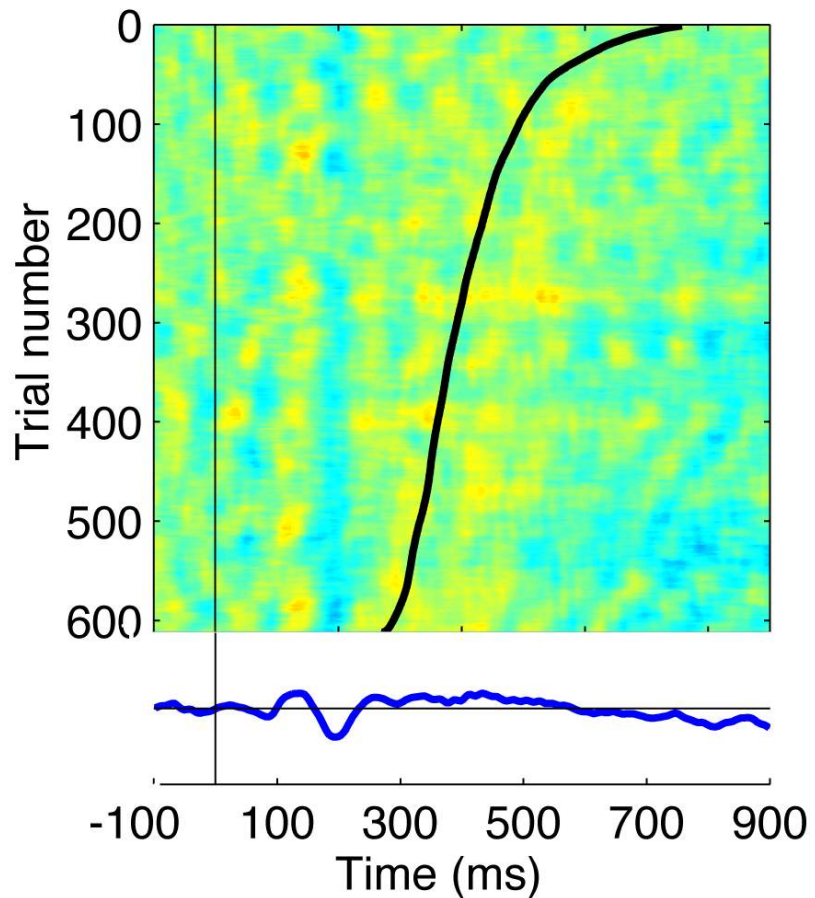
Averages of **Least**, **Moderately** and **Heavily** Contaminated Trials



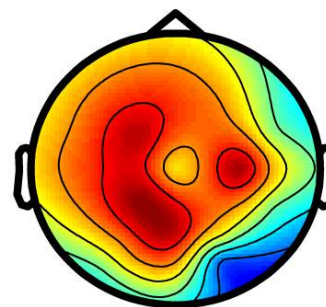
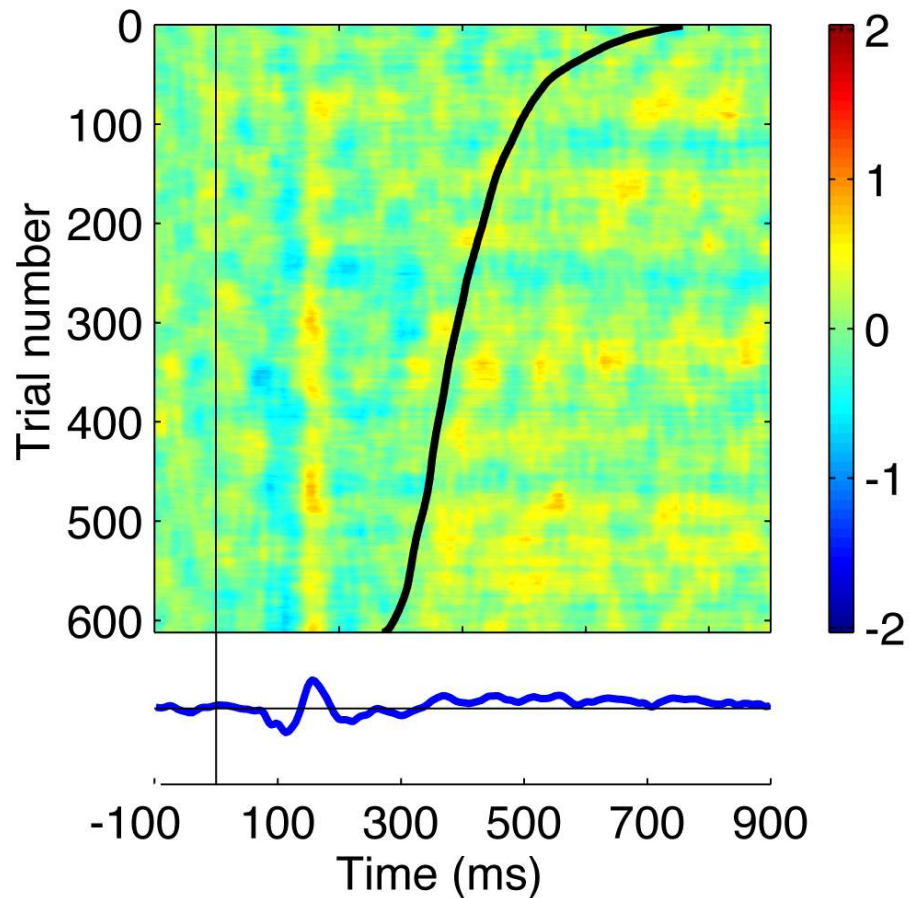
From Jung et al., *Clinical Neurophysiology*, 2000.

Stimulus-locked

IC7 activations

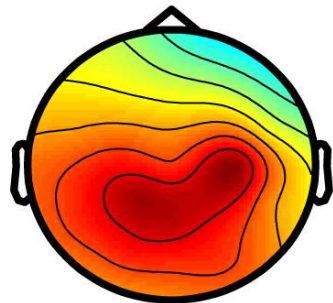
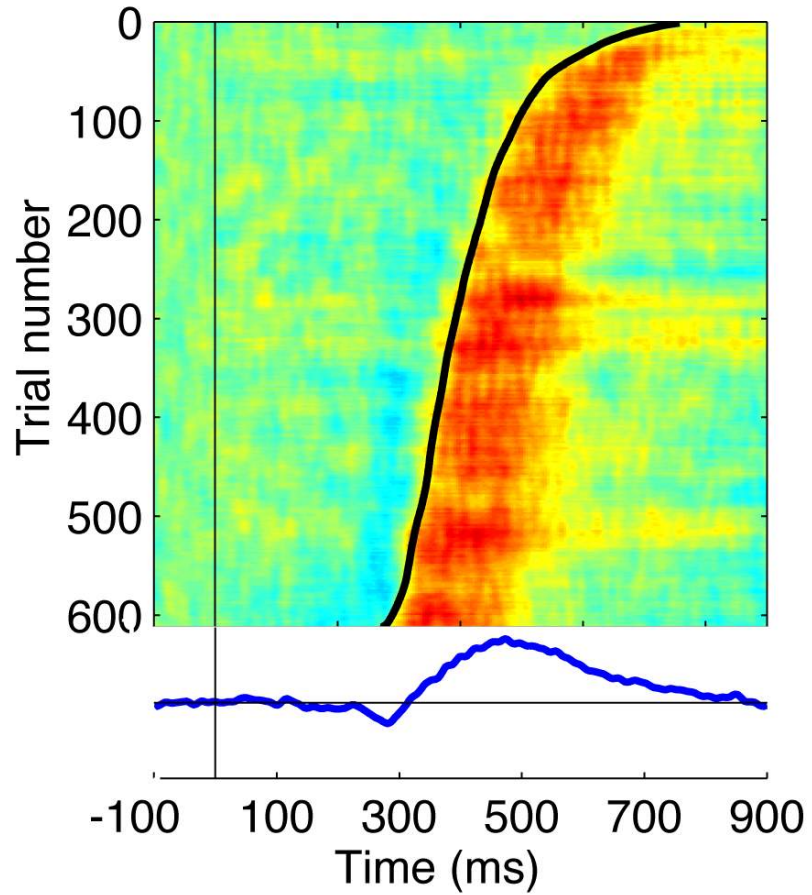


IC14 activations

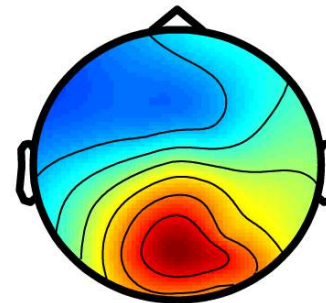
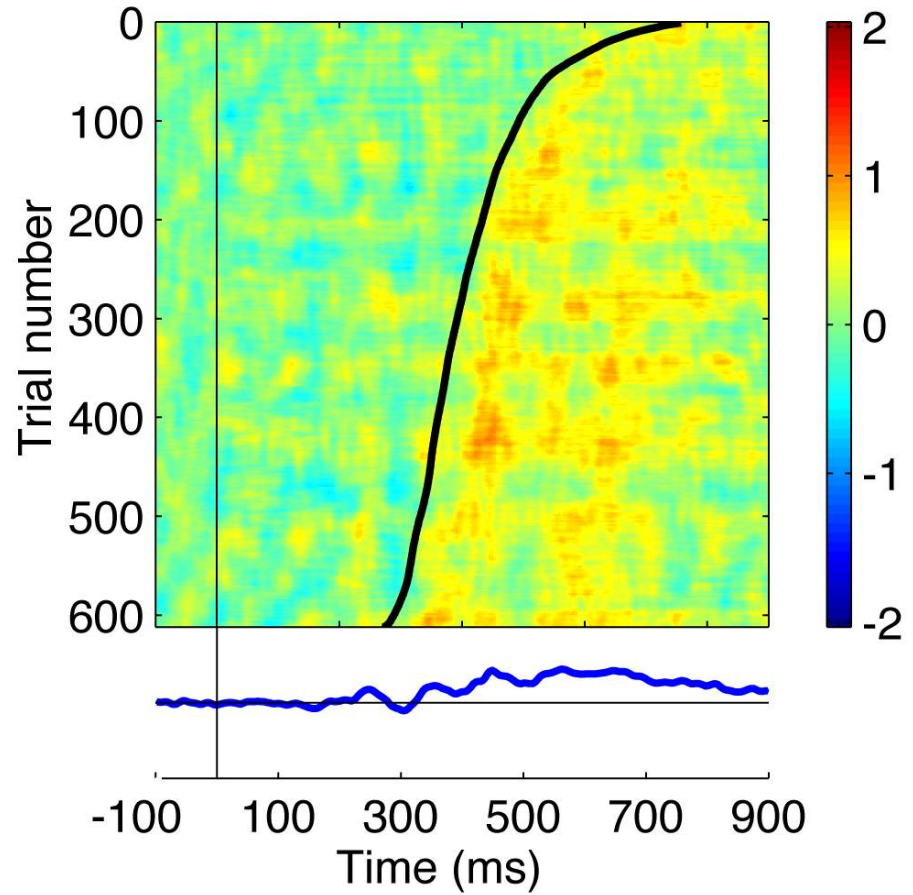


Response-locked

IC2 activations

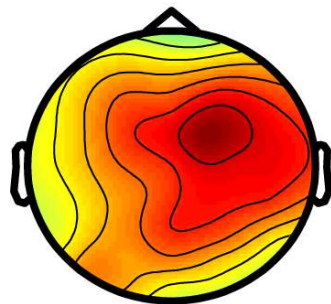
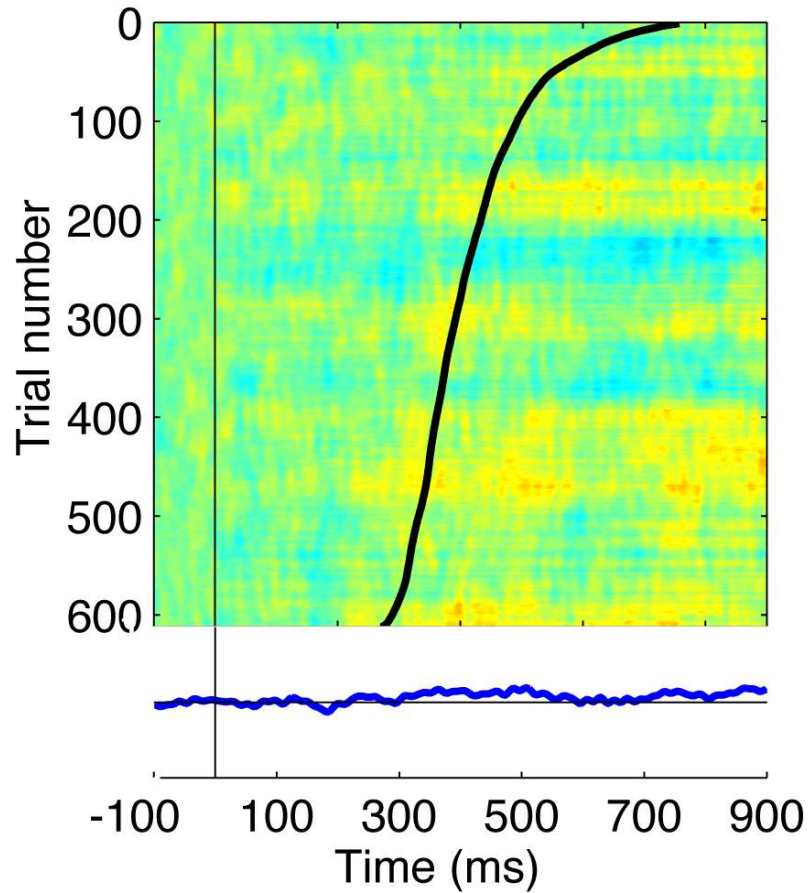


IC8 activations

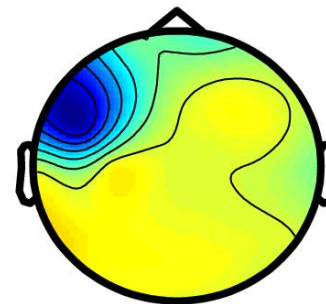
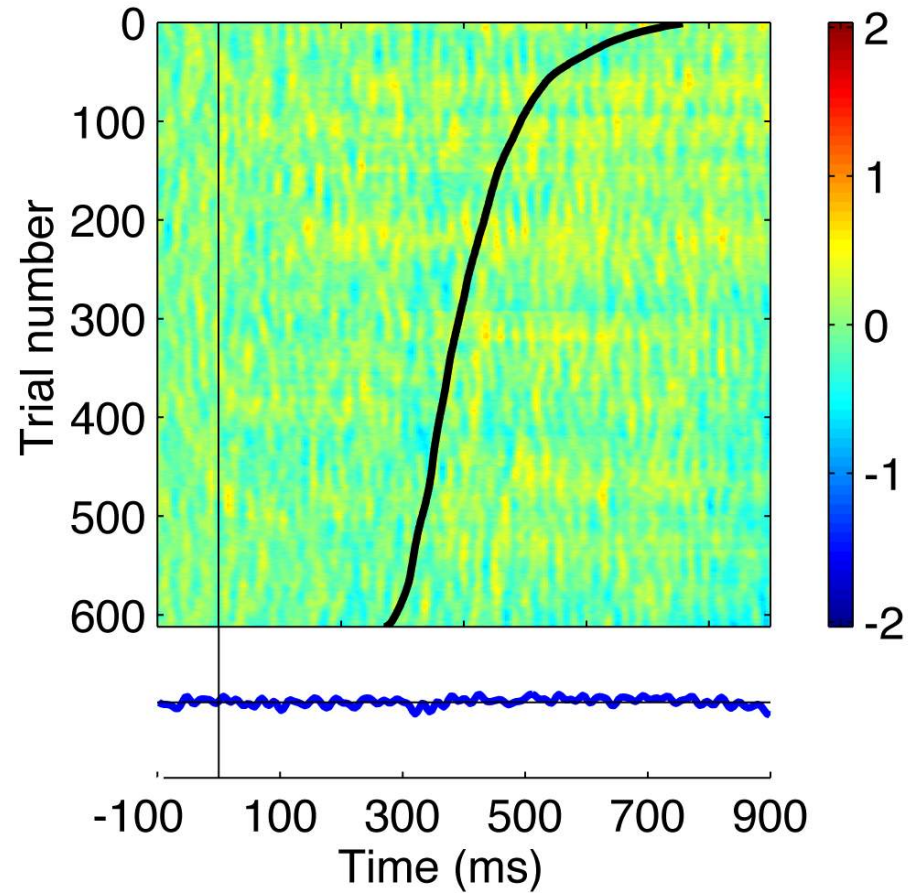


Non-phase locked

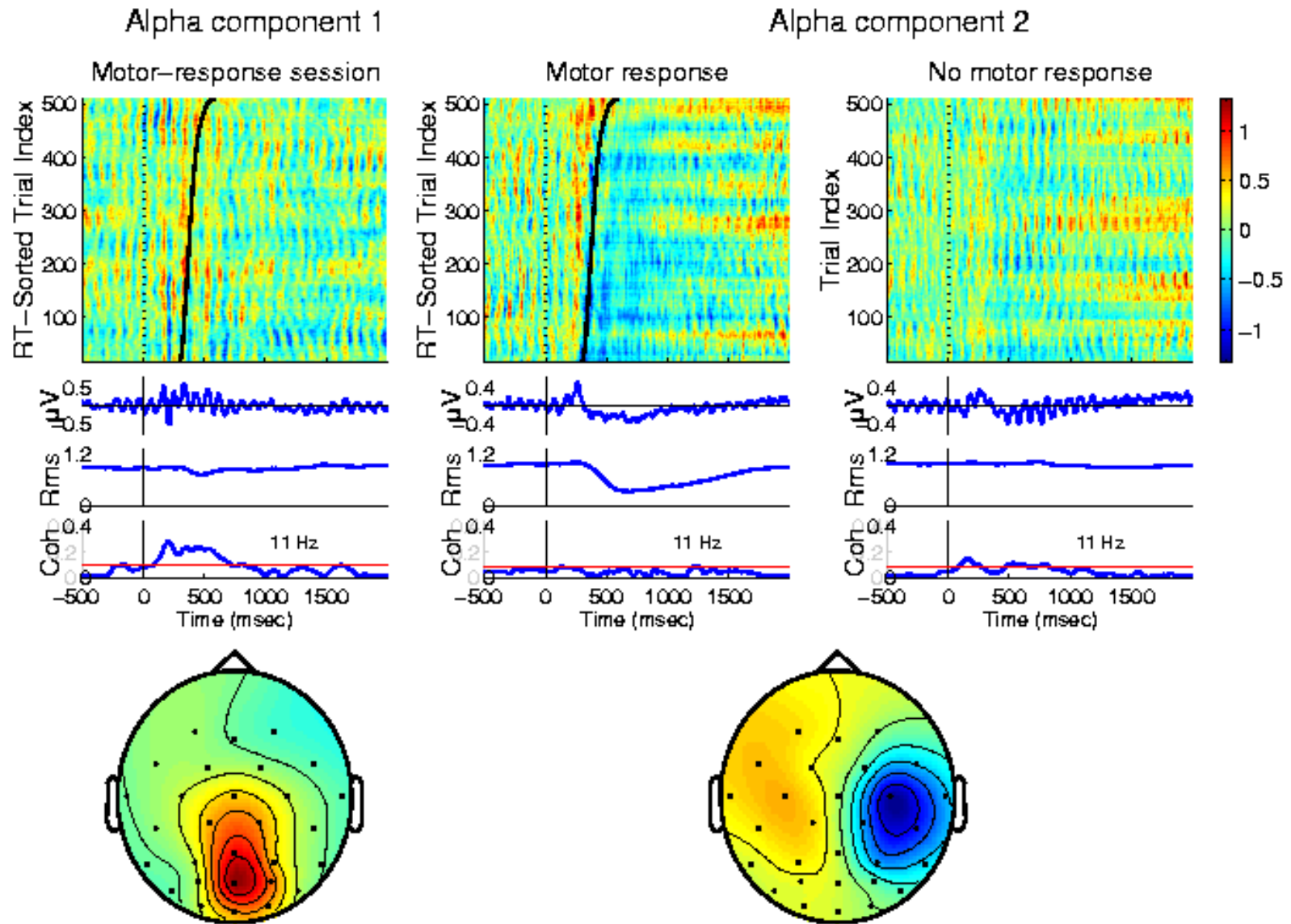
IC5 activations



IC23 activations



Event-modulated Oscillatory Activity



Characteristics of Independent Components

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

Conclusions

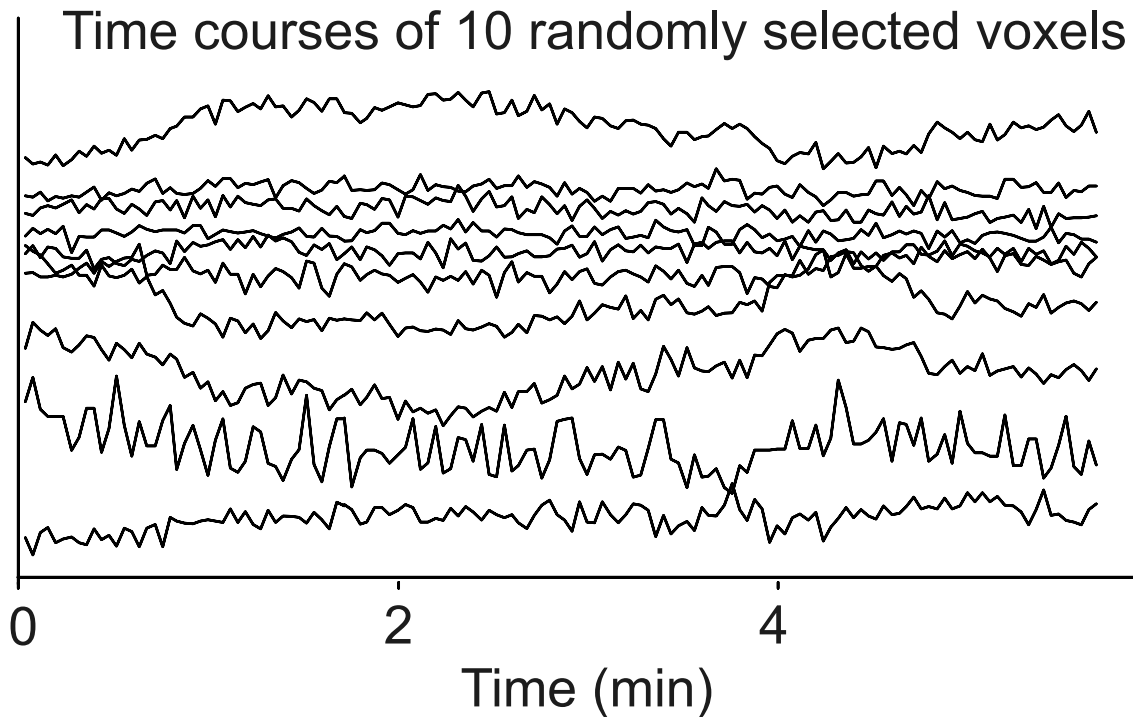
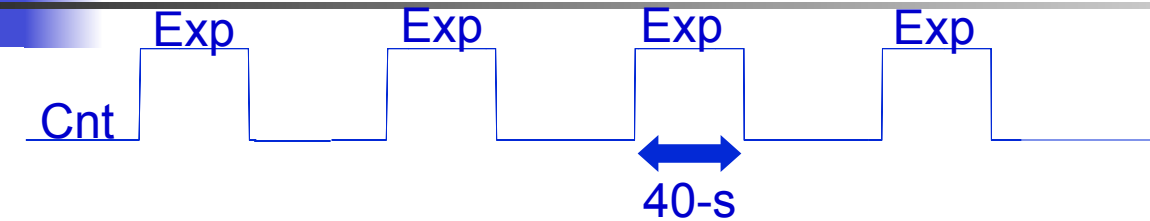
- ICA has proven successful in many data-analysis applications.
- Great care must be taken to examine the validity of the assumptions that are used by ICA to derive a decomposition of the observed signals and/or to evaluate the reliability and functional significance of the resulting components.

Magnetic Resonance Imaging (MRI and/or functional MRI)

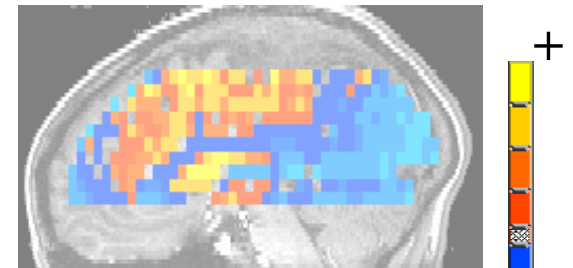


1. MRI is an imaging technique used to produce high quality images of the inside of the human body.
2. It is based on the magnetic susceptibilities of oxygenated hemoglobin (HbO_2) and deoxygenated hemoglobin (HbR) to track the blood-flow changes related to neuronal activity, which is referred to as blood-oxygen-level-dependent (BOLD) contrast.

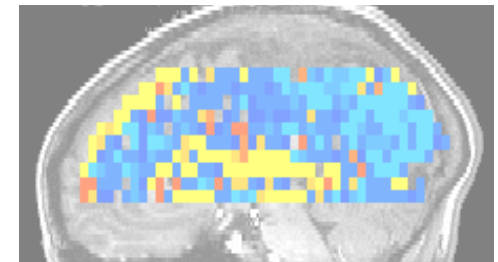
fMRI/BOLD Signal Complexity



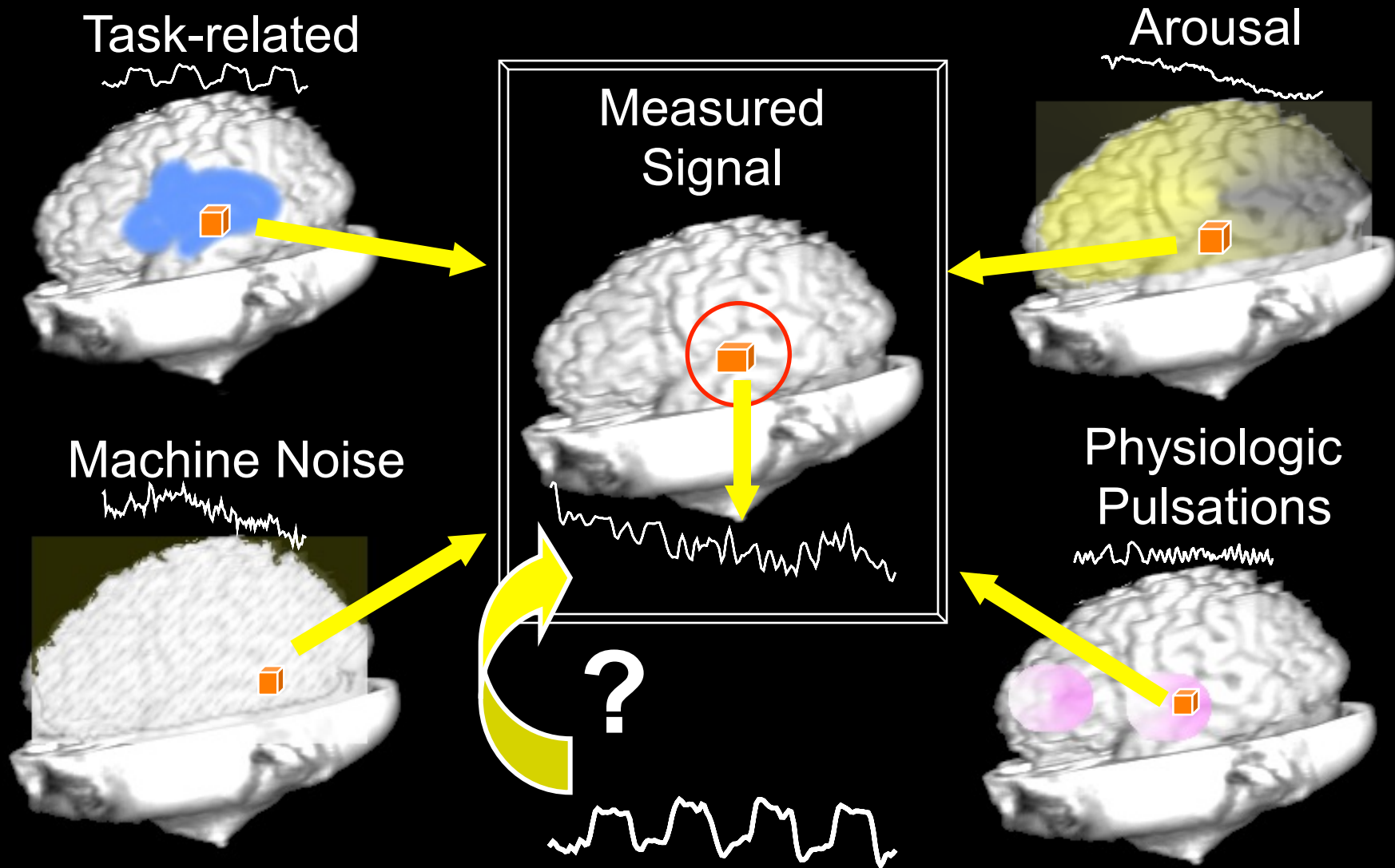
Mean Signal



Variance



ICA Applied to fMRI Data



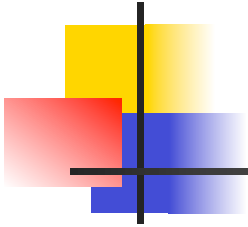
Analysis of Event-related fMRI Data

Model-based methods

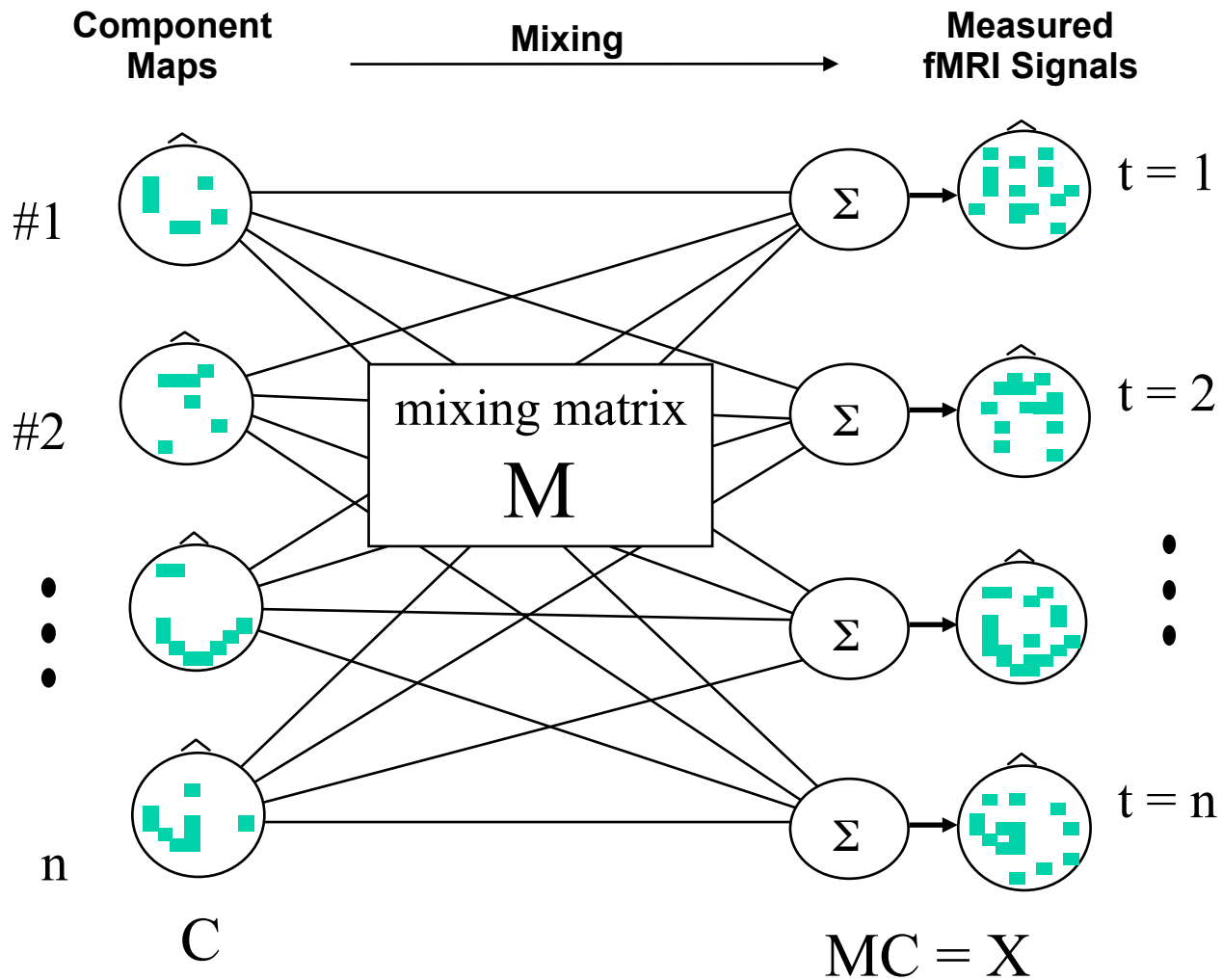
- **Require** *a priori* knowledge of the time course of the hemodynamic response
- **Assume** homogeneity across different brain regions
- **Allow** tests of statistical significance within an assumed data+noise model.

Data-driven methods

- **Require** minimal space/time assumptions
- **Explore** time courses and spatial distribution of the data
- **Reveal** unforeseen activations (time-varying, site-dependent)
- **Provide no** noise model for statistical testing.

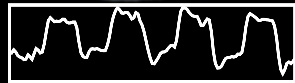
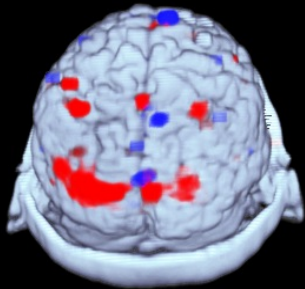


ICA Applied to fMRI Data

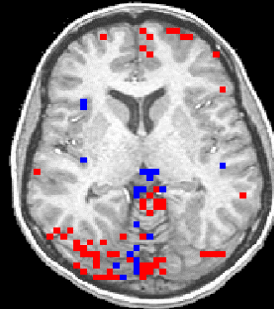


Independent fMRI Components

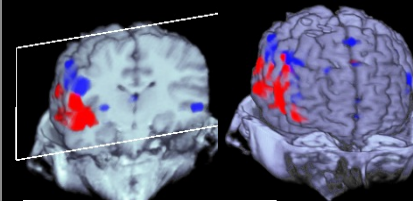
Consistently task-related



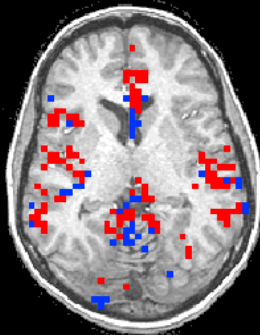
Transiently task-related



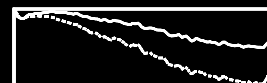
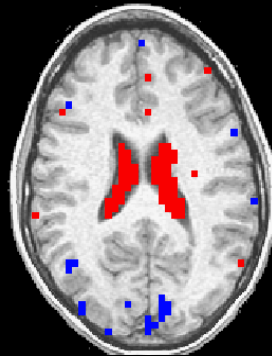
Abrupt head movement



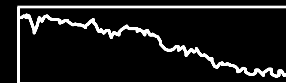
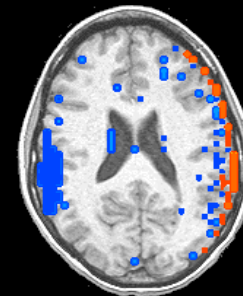
Quasi-periodic



Slowly-varying



Slow head movement



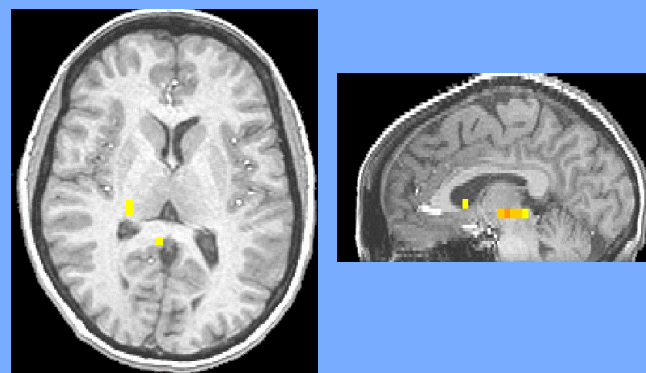
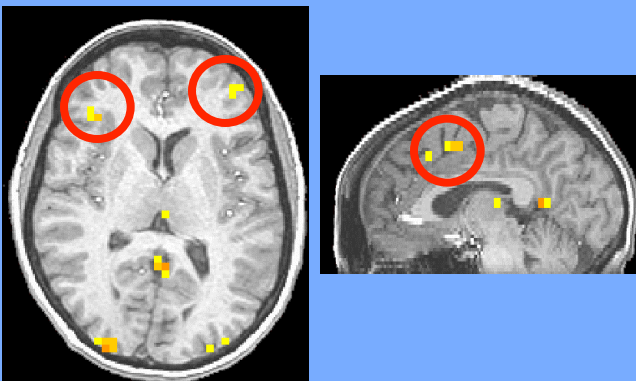
■ Activated
■ Suppressed

3 Subjects

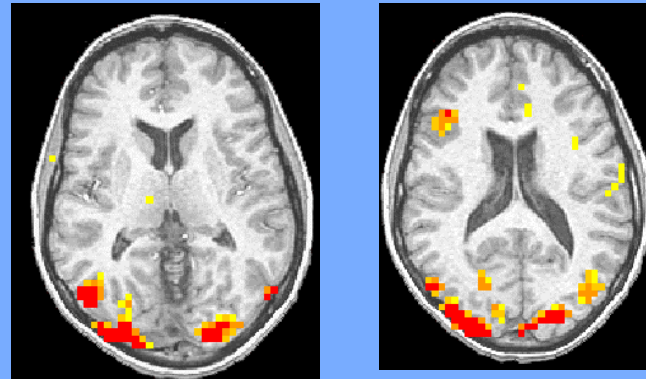
ICA

Correlation

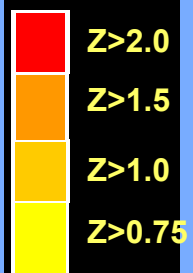
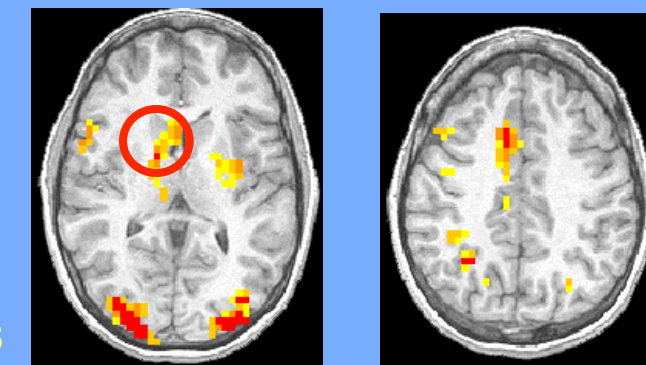
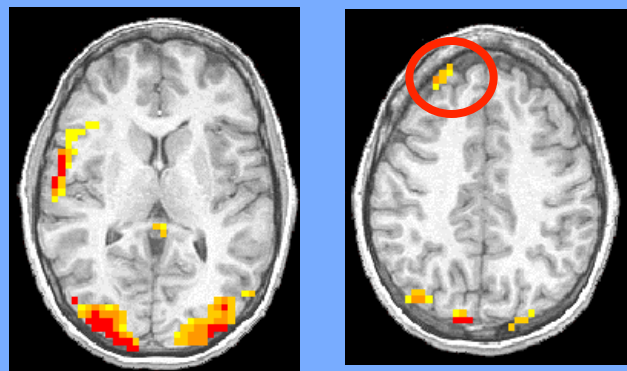
S1



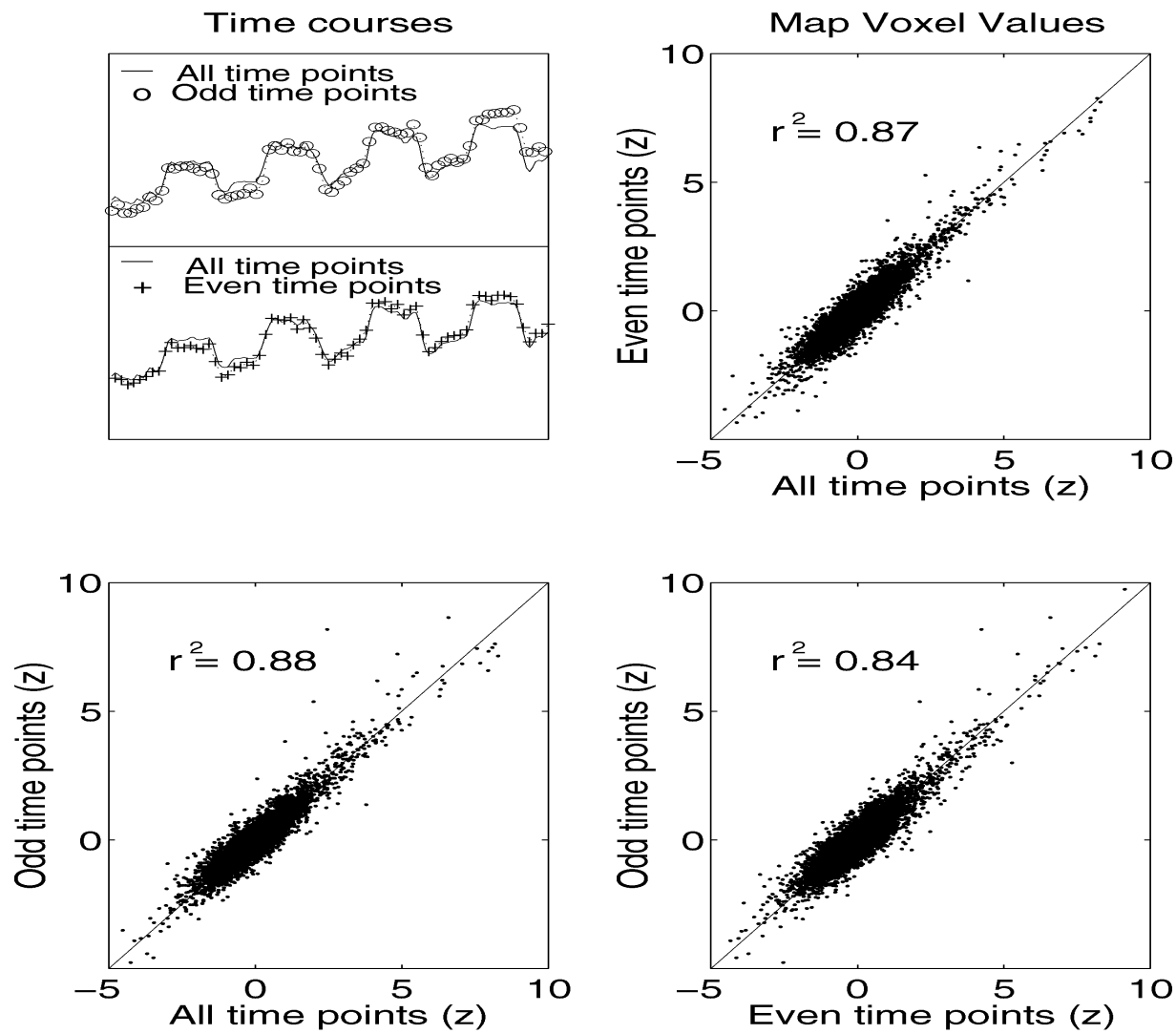
S2



S3

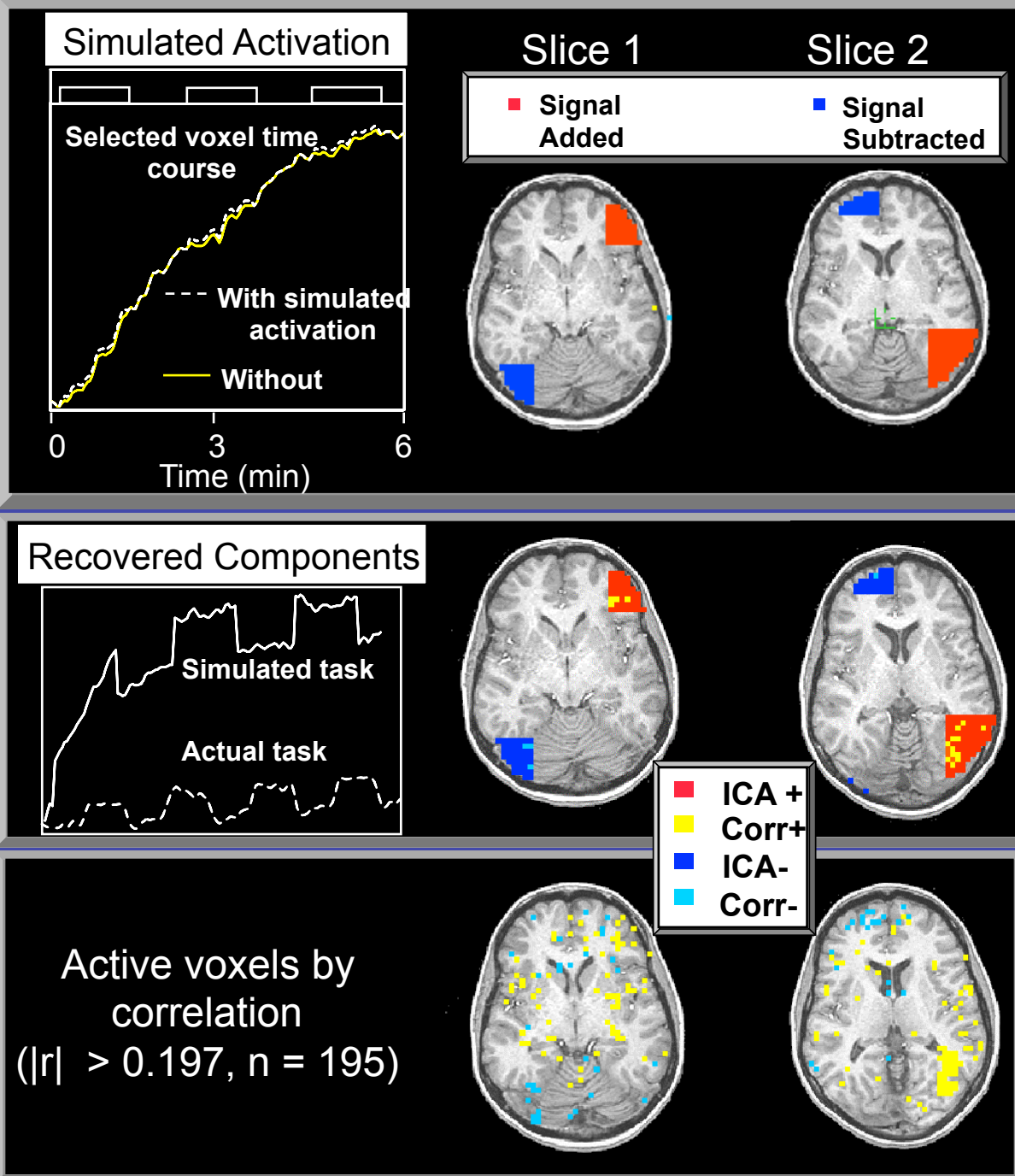


Stability of ICA Component Maps



Relative Task Sensitivity

ICA vs Correlation (Simulation)



FMRLAB, a Matlab-based open-source environment for fMRI analysis



J-R Duann

