Independent Component Analysis: Theory and Application to EEG

Jason Palmer

Osaka University, Osaka, Japan
University of California San Diego, La Jolla CA

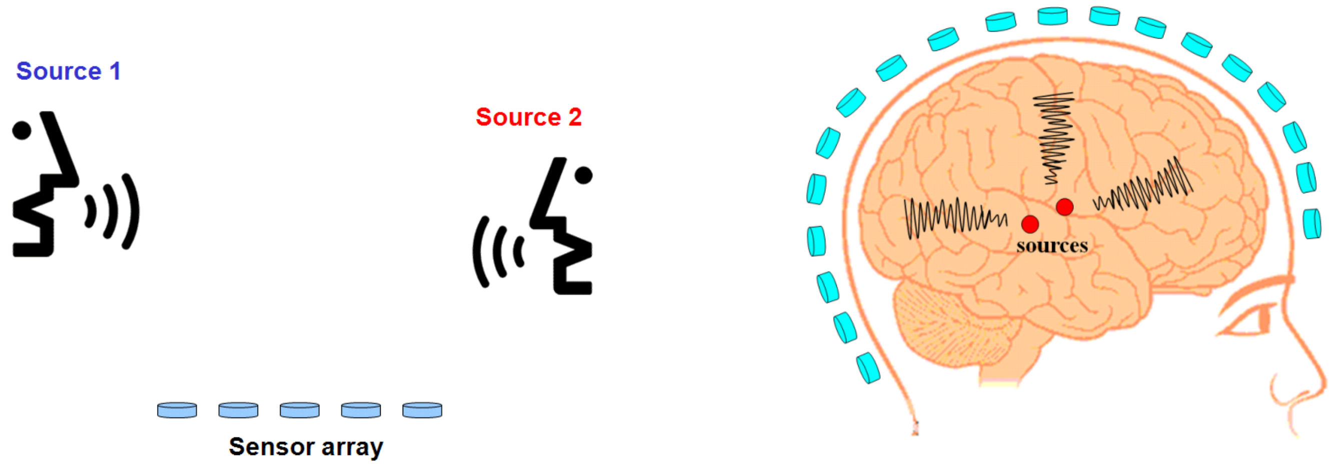
September 26, 2017

Outline

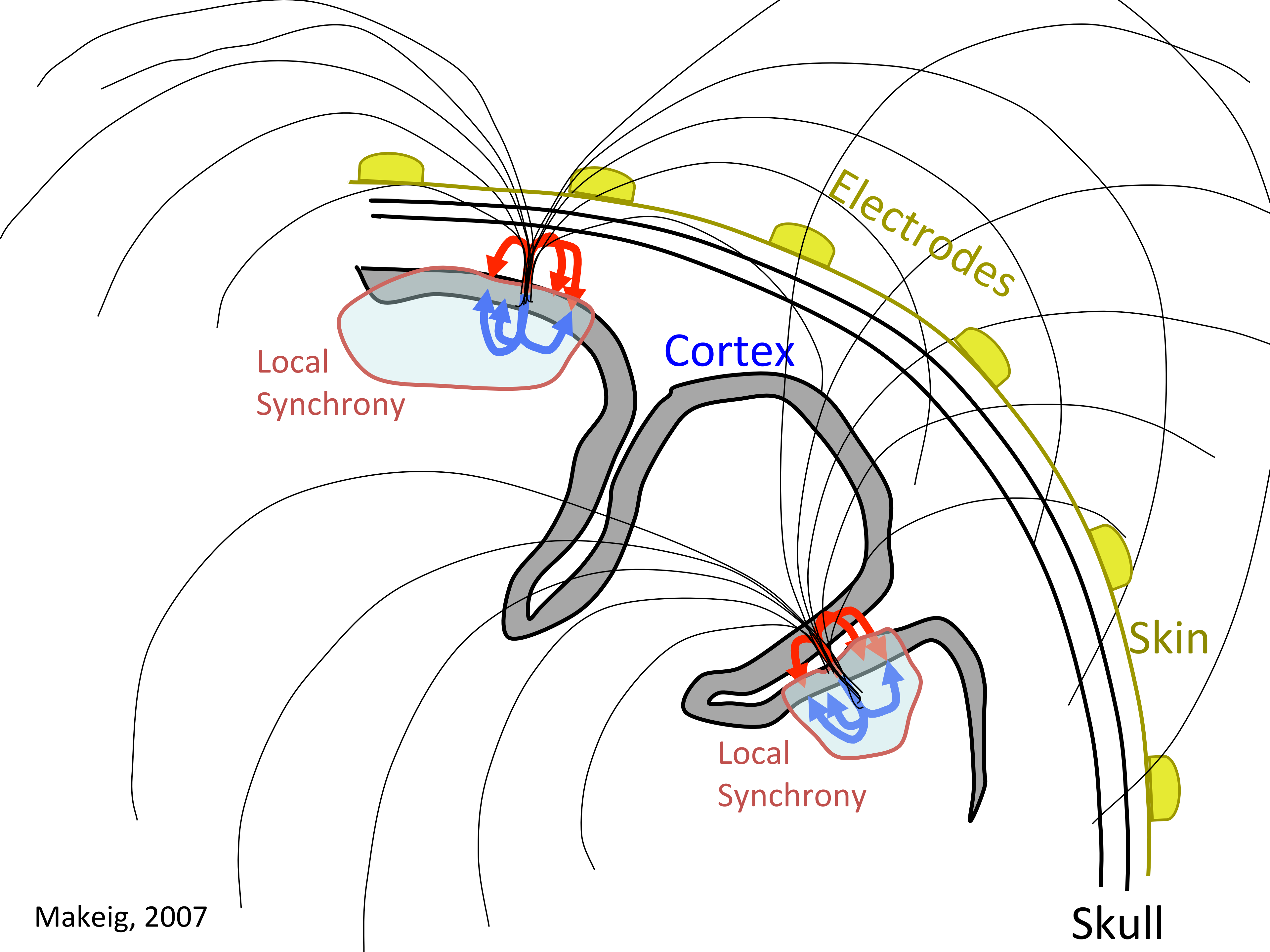
- EEG and the cocktail party problem
- The ICA model
- How does ICA work?
- Dependency and subspaces – does ICA still work?

Cocktail Party Problem

- EEG analysis as separation of ***multiple simultaneously active*** brain sources, similar to microphones recording and multiple simultaneous speakers, e.g. at a cocktail party

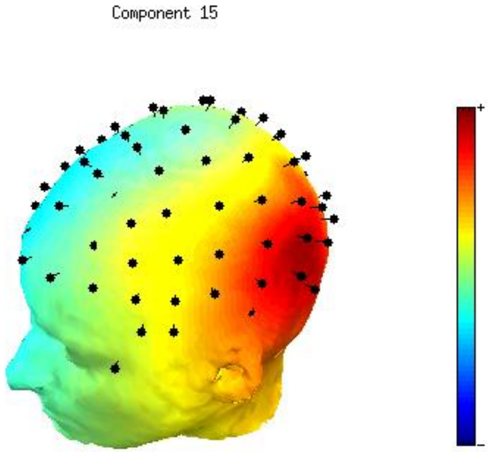


- ICA originally proposed for separation of multiple independent audio signals (early '90s)
- Scott Makeig proposed ICA for EEG source separation (1996), in collaboration with Tony Bell and Terry Sejnowski at Salk


Electrodes
Local
Synchrony
Cortex
Skin
Local
Synchrony
Makeig, 2007
Skull

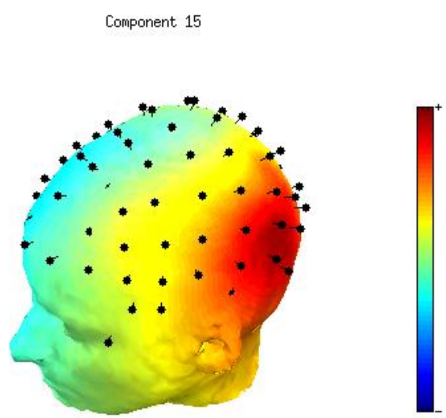
EEG Sources

- A source is essentially defined by the pattern of electrical potential that it projects onto the electrodes (by volume conduction)



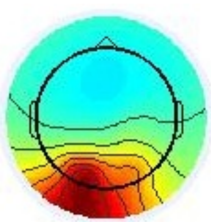
EEG Sources

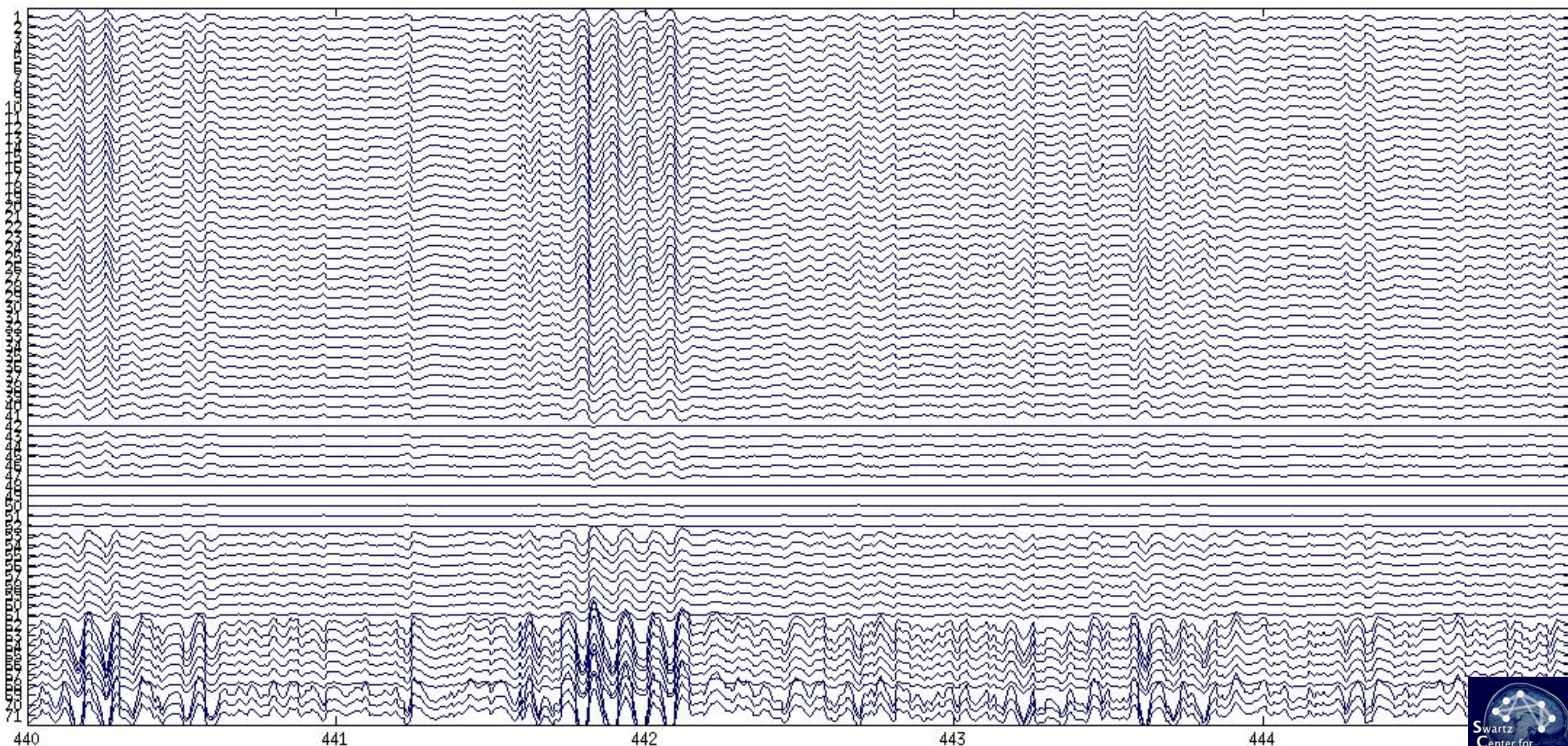
- Stationary source activity (local and stable) fluctuates, or oscillates, around zero, causing alternation of positive and negative potentials at the scalp



EEG of one source

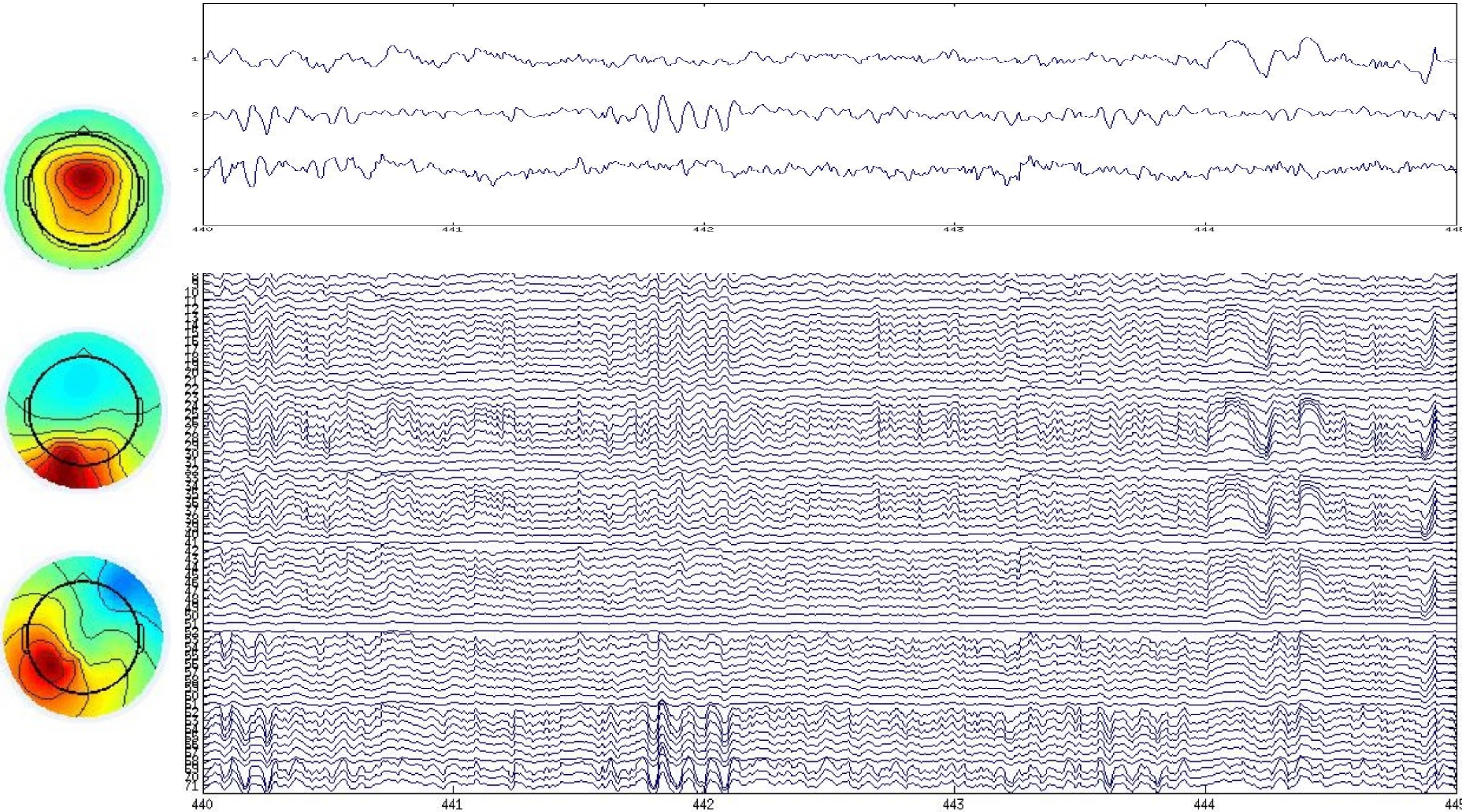
- EEG electrodes record the source activity weighted by different values depending on electrode location relative to the source





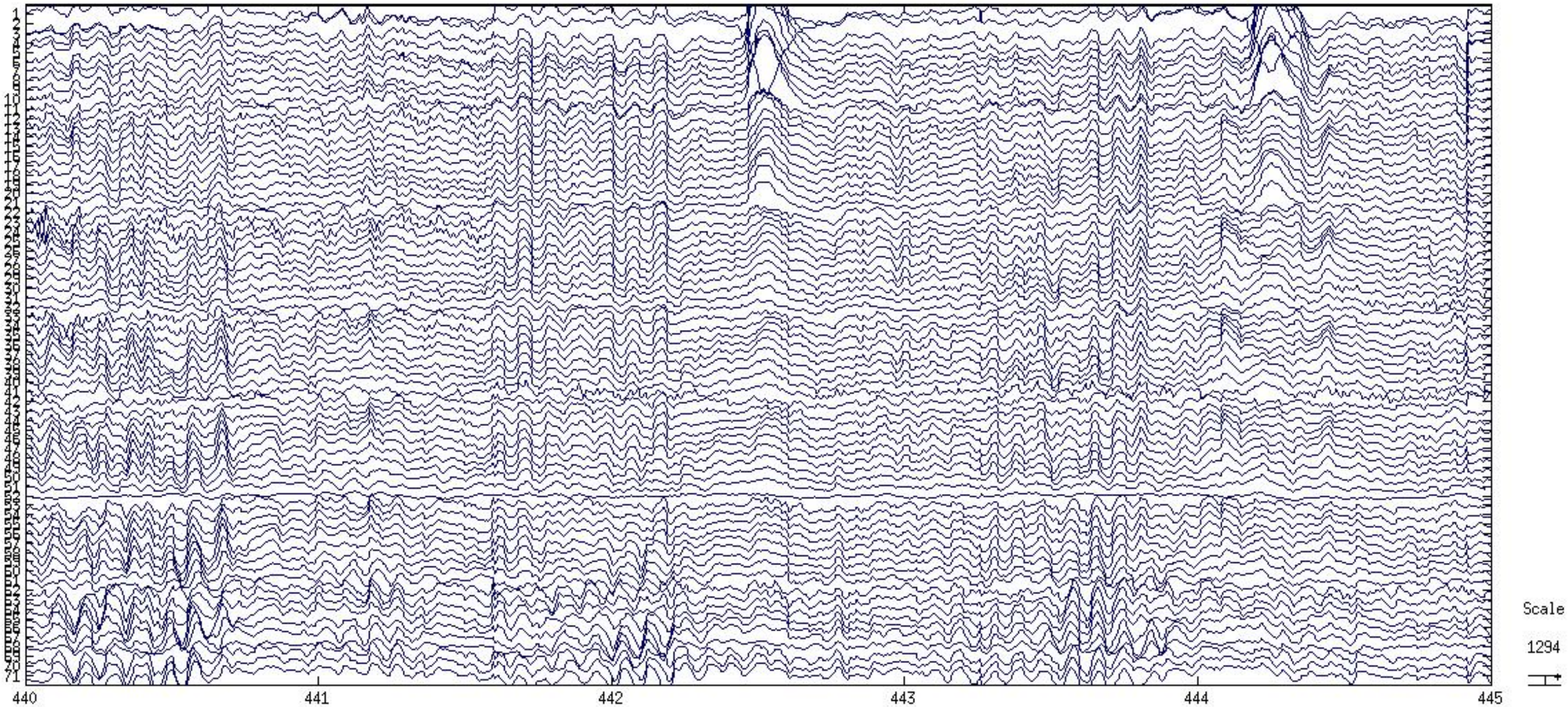
EEG of three sources

- EEG records multiple sources that are simultaneously active



EEG Data

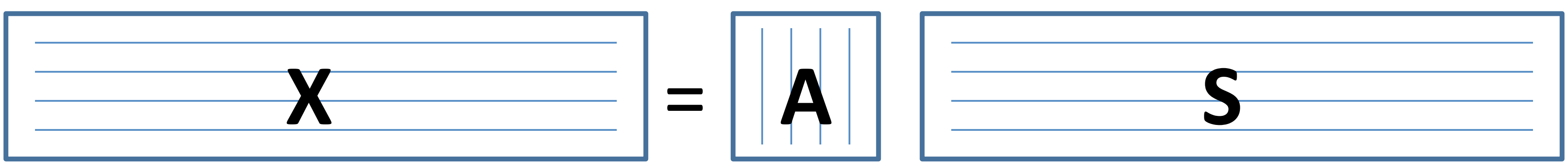
- Raw EEG records large number of simultaneously active sources
- From physics, we know that EEG at one instant is simply the sum of all source activity at that instant



Linear Superposition

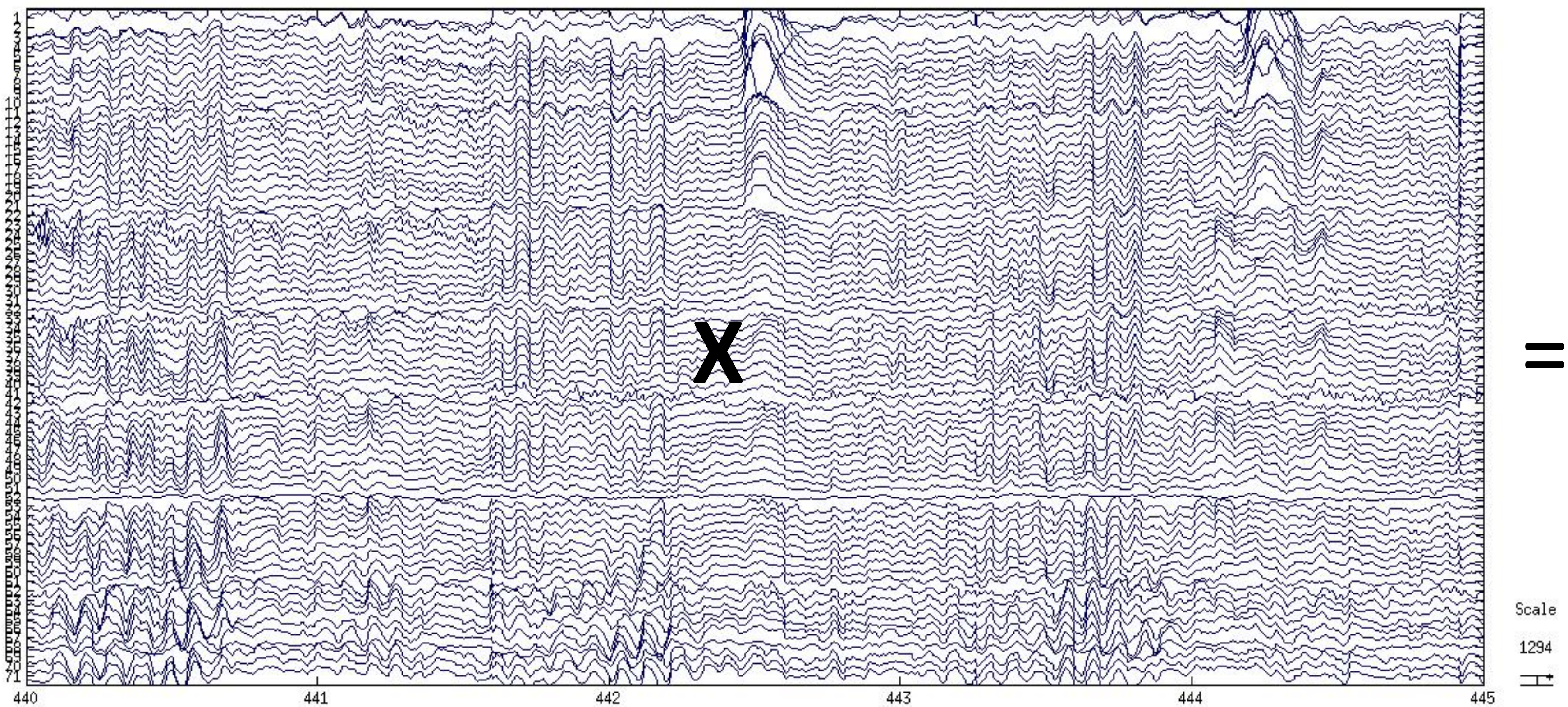
- Let the EEG data be represented by the vector of time varying electrode potentials $\mathbf{x}(t)$, and let the source activities be $s_i(t)$, $i = 1, \ldots, n$
- Let the scalp maps (patterns of potential) be represented by vectors $\mathbf{a}_i$, $i = 1, \ldots, n$
- The EEG data is the sum:

$$\mathbf{x}(t) = s_1(t)\,\mathbf{a}_1 + s_2(t)\,\mathbf{a}_2 + \ldots + s_n(t)\,\mathbf{a}_n$$



Decomposition of EEG

- Given the EEG data, **X**, we would like to decompose it into source scalp maps multiplied by source activity, **X** = **AS**, with **A** and **S** unknown

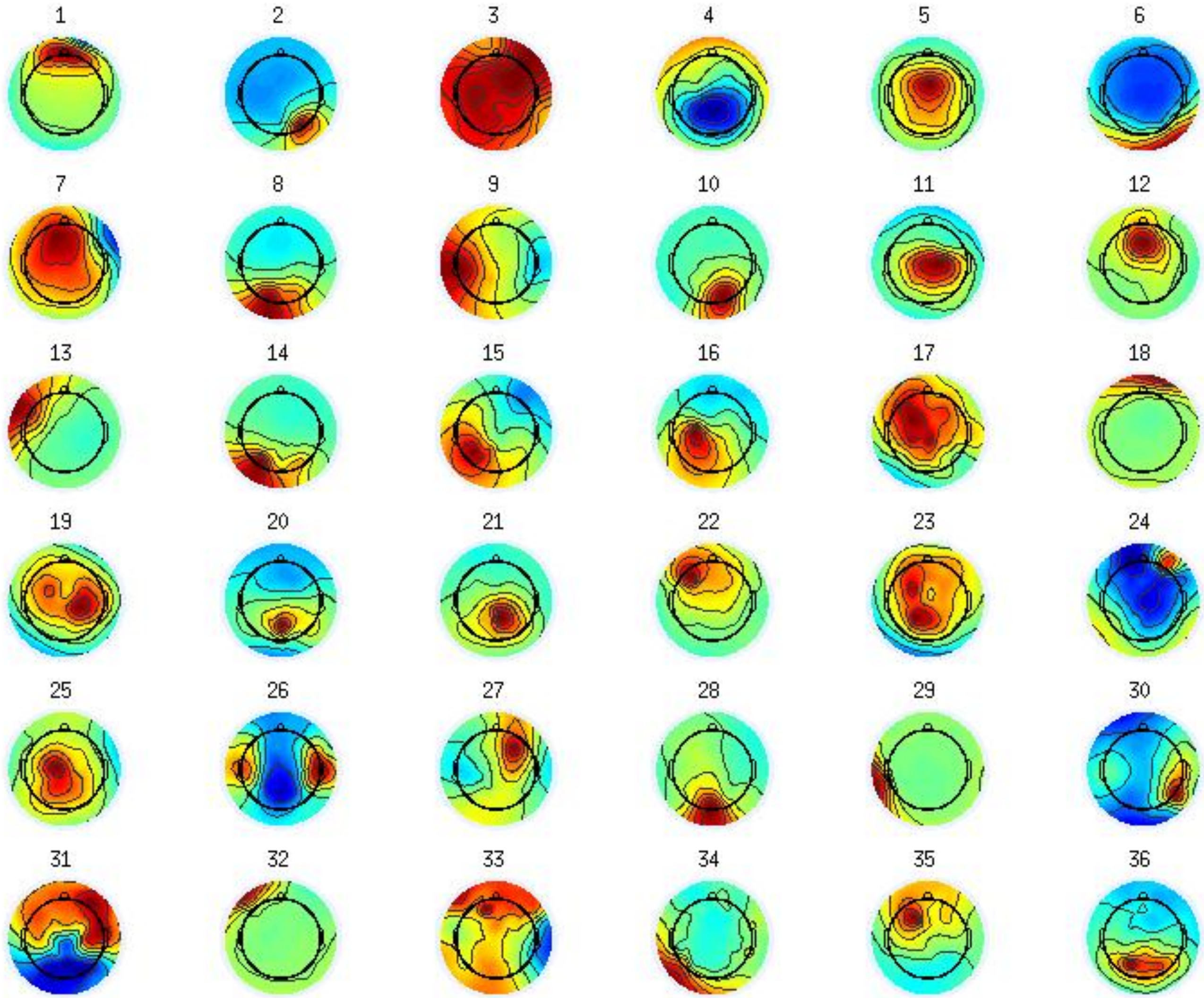


= AS

UCSD

Swartz Center for Computational Neuroscience

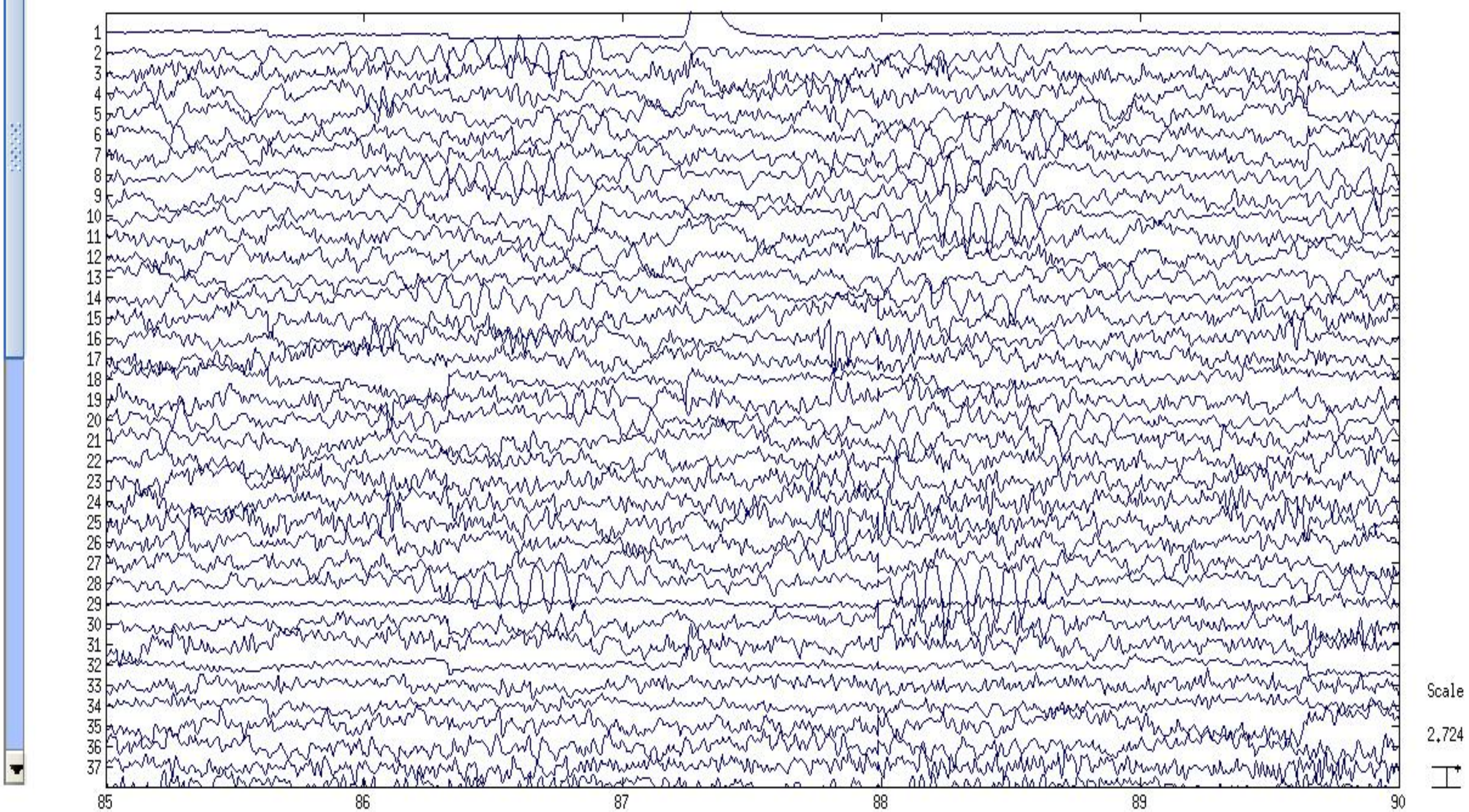
Typical ICA scalp maps



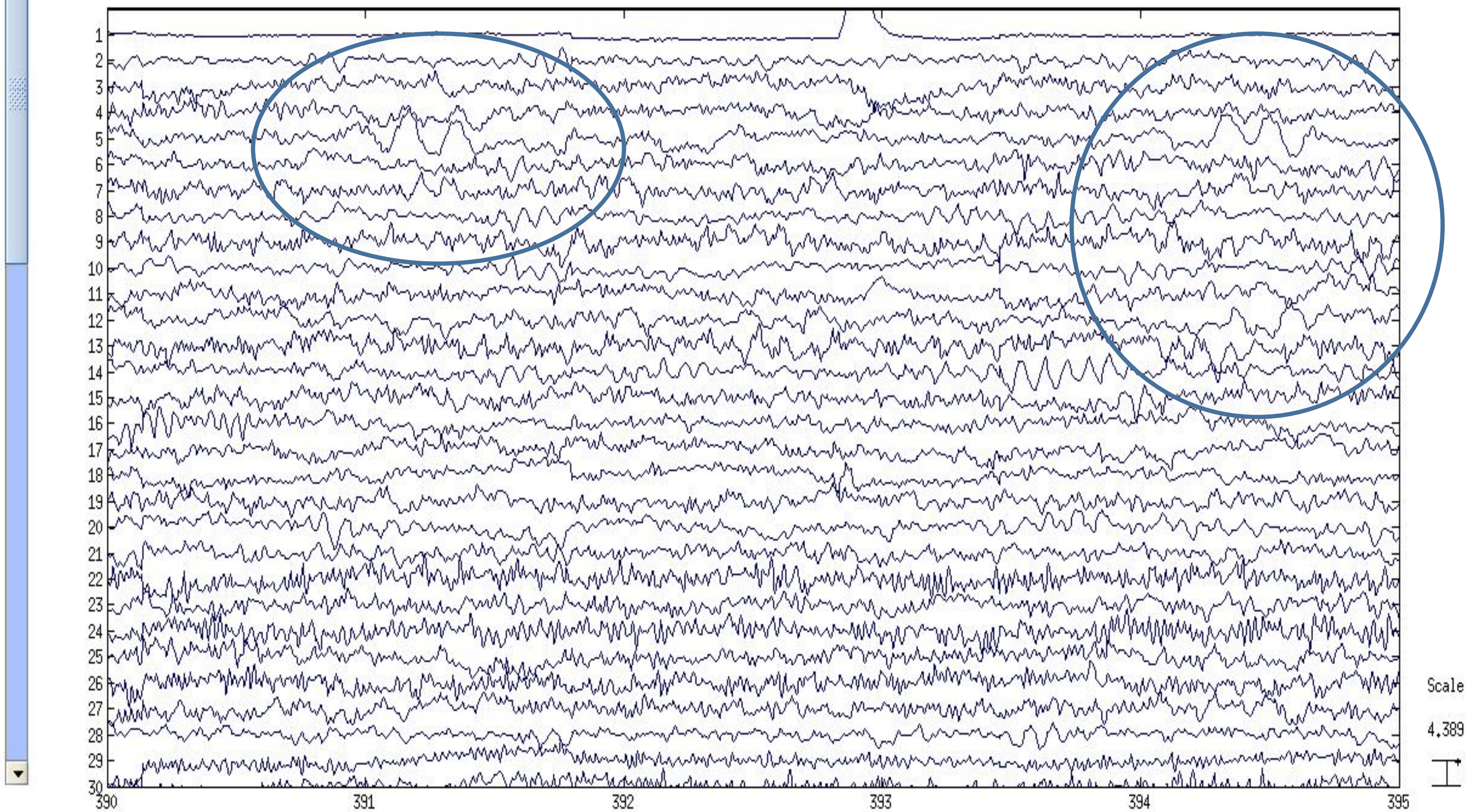
UCSD

Swartz Center for Computational Neuroscience

Typical ICA sources – Alpha



Typical ICA sources – Theta



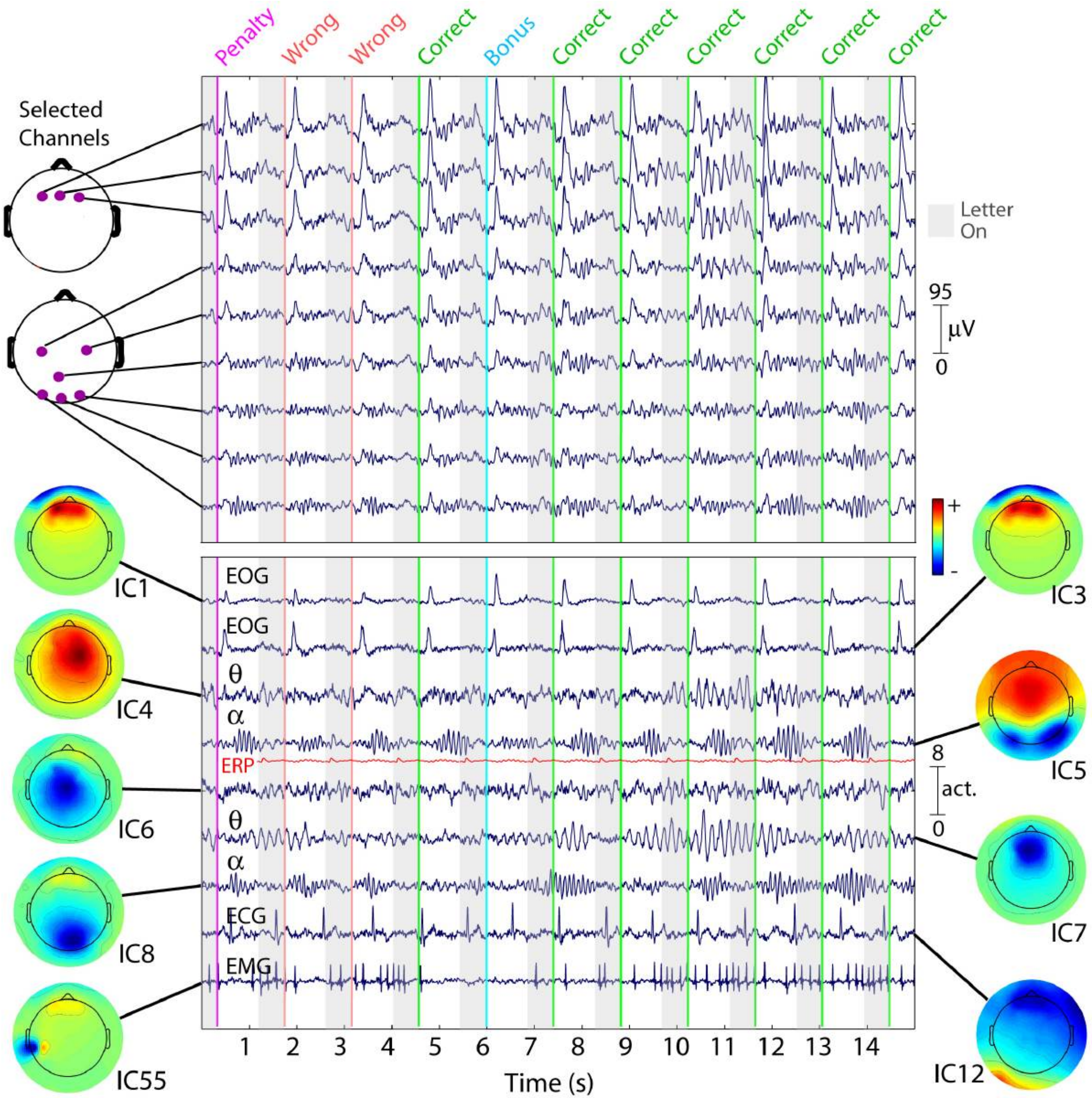
UCSD

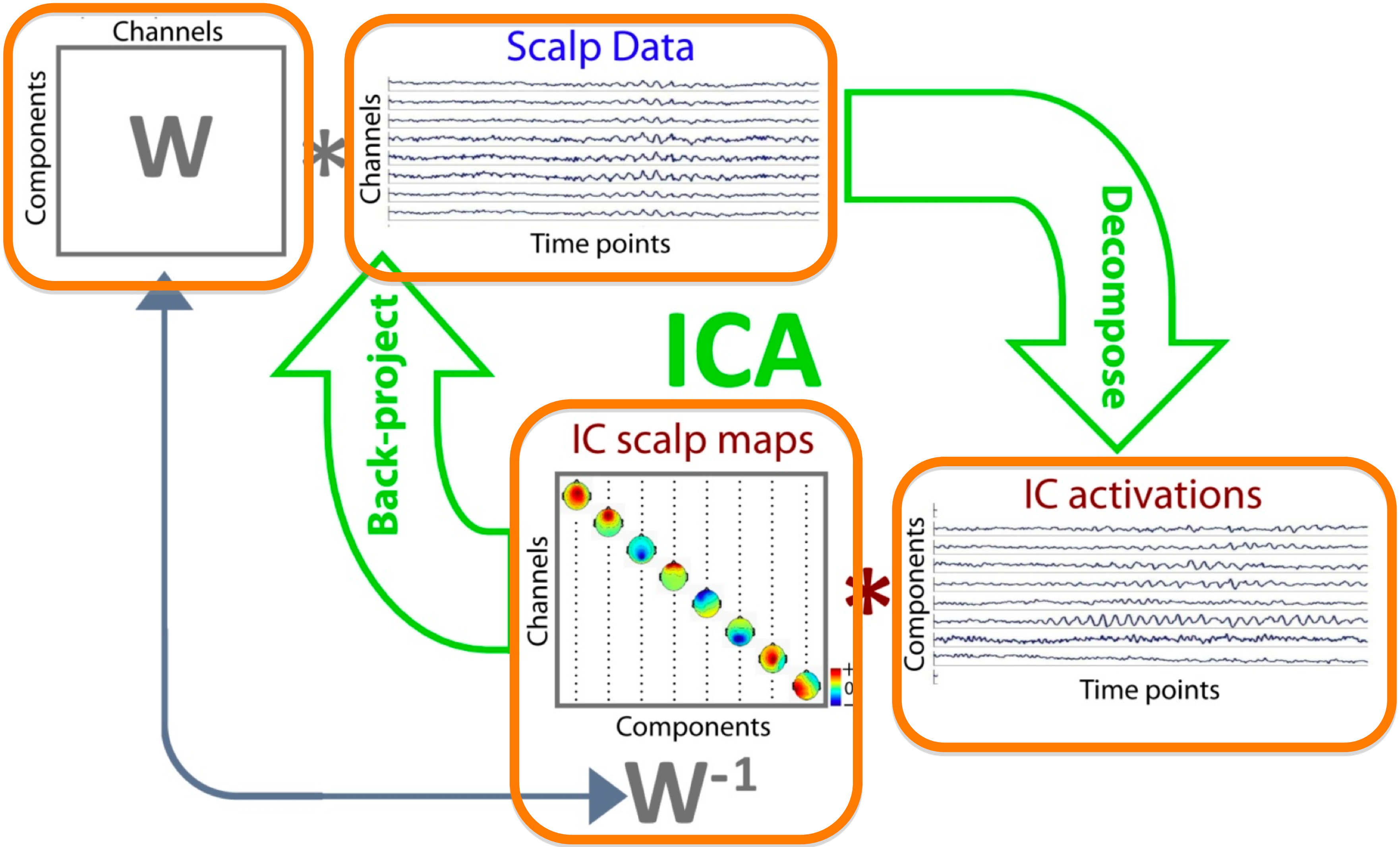
Swartz Center for Computational Neuroscience

Independent Component Analysis (ICA)

- ICA is used to separate raw EEG data (top) into independent sources (bottom)
- Often used for artifact removal (e.g. eye-blinks)
- Can be used to extract sources of interest for further analysis (e.g. theta or alpha, possibly gamma)

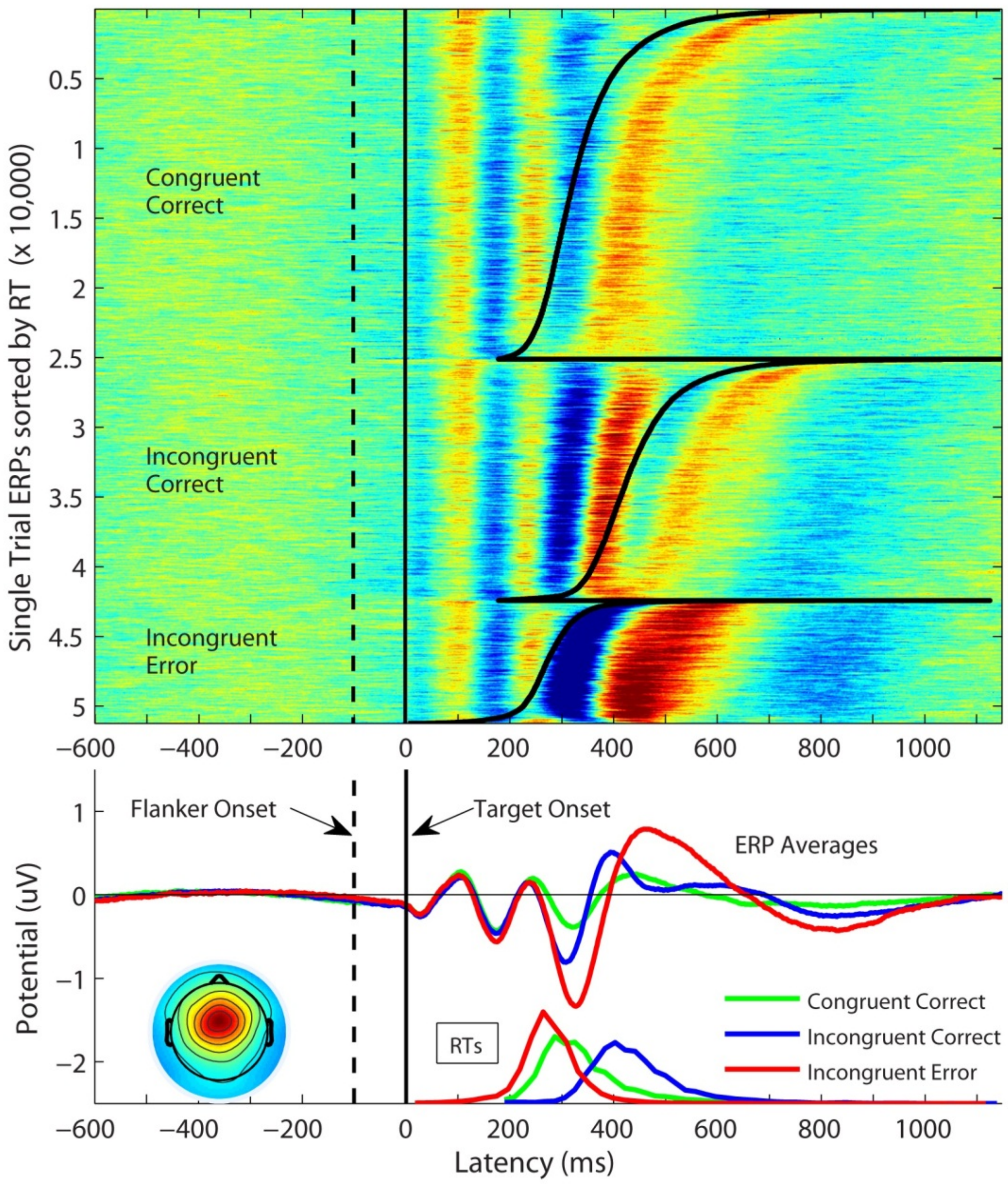
Makeig, 2007




Channels
Components
W
Scalp Data
Channels
Time points
Decompose
ICA
Back-project
IC scalp maps
Channels
Components
W-1
IC activations
Components
Time points

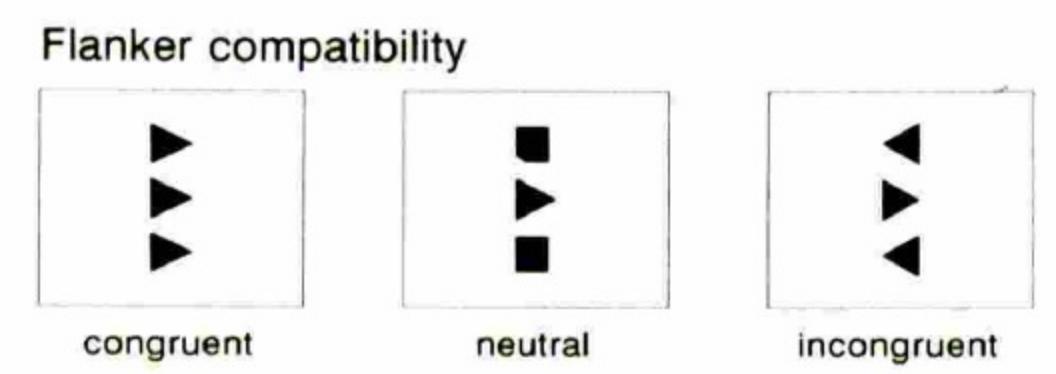
Makeig, 2007

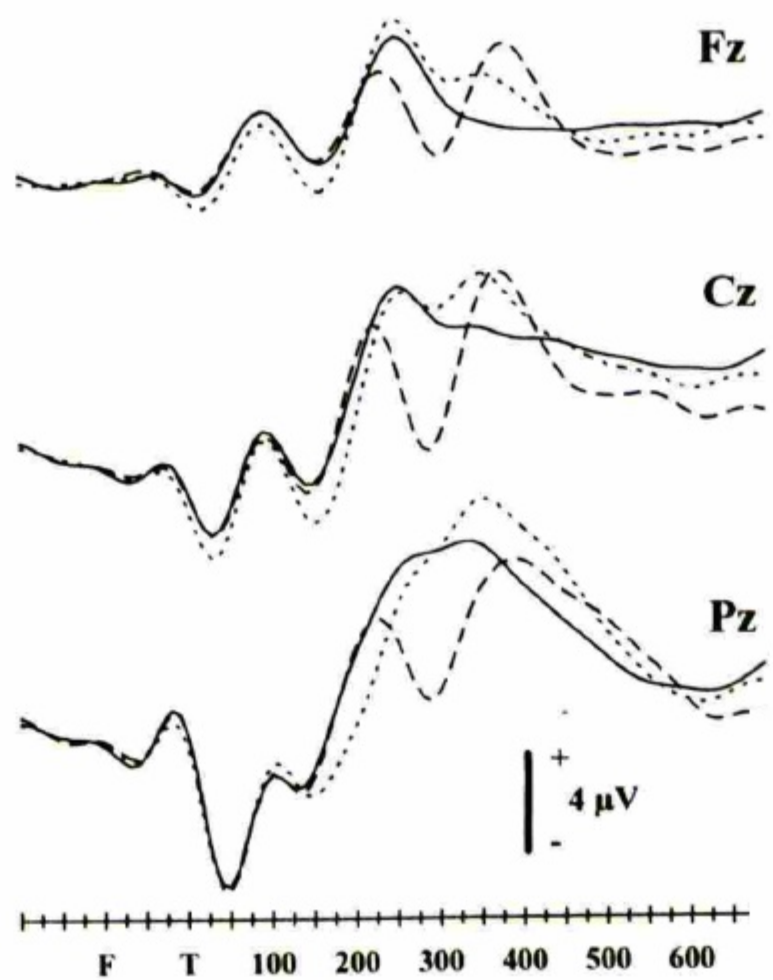
Central Midline Theta Component



McLoughlin et al., 2014

Kopp (1994,1996) and the Arrow Flanker Paradigm

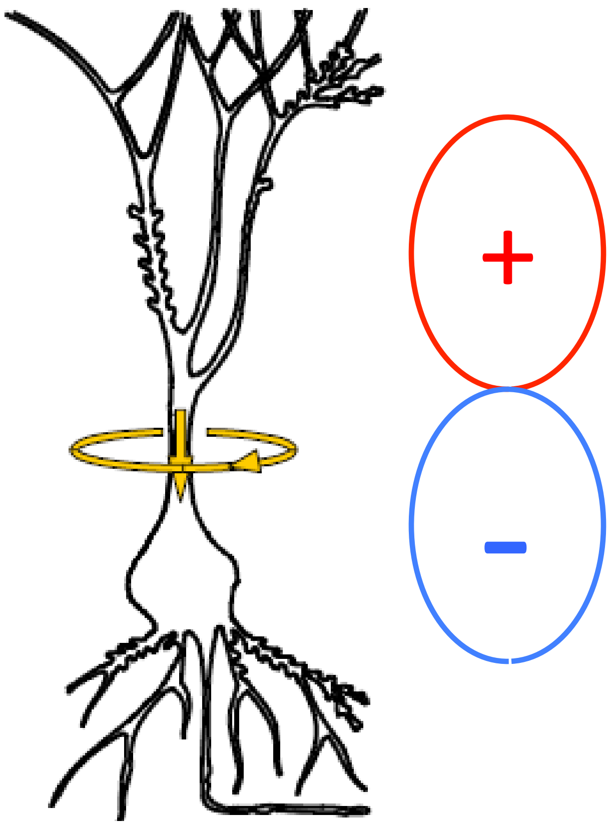




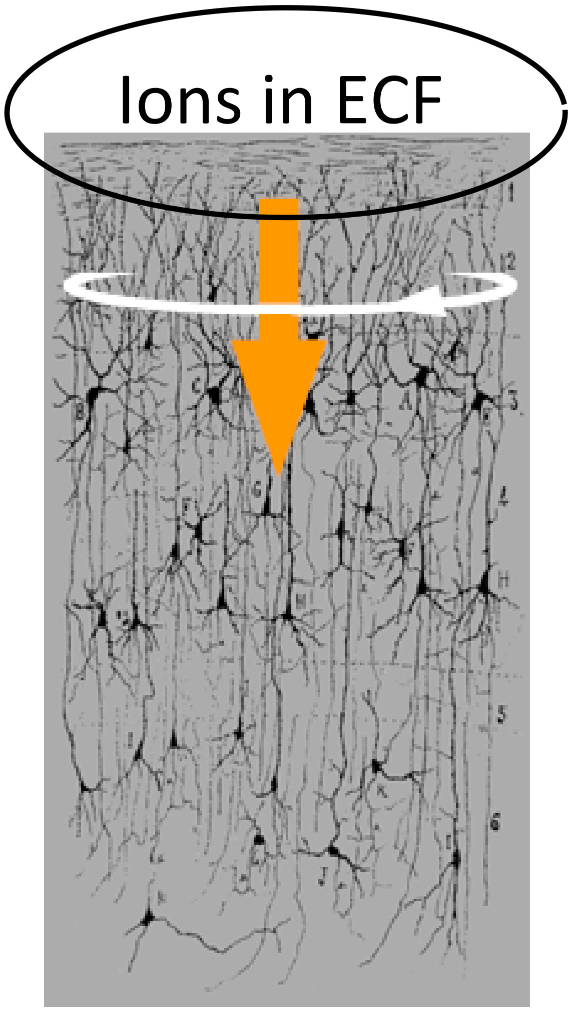
Note that the theta at Cz is conflated with the P300 waveform

Where do the dipoles come from?

Neuron

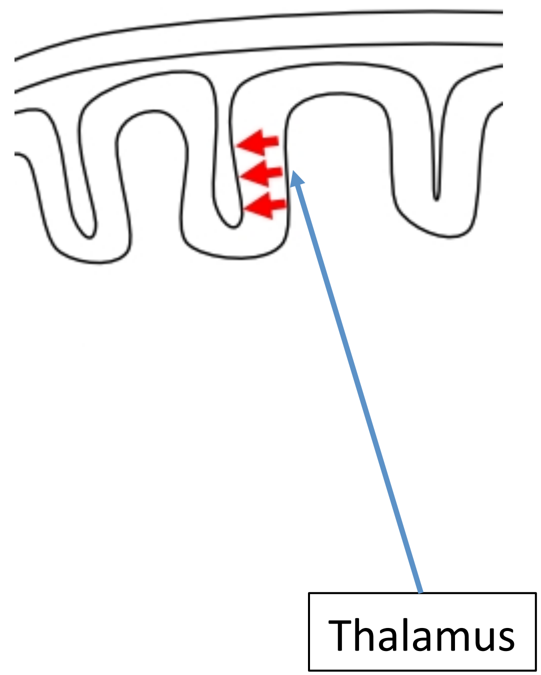


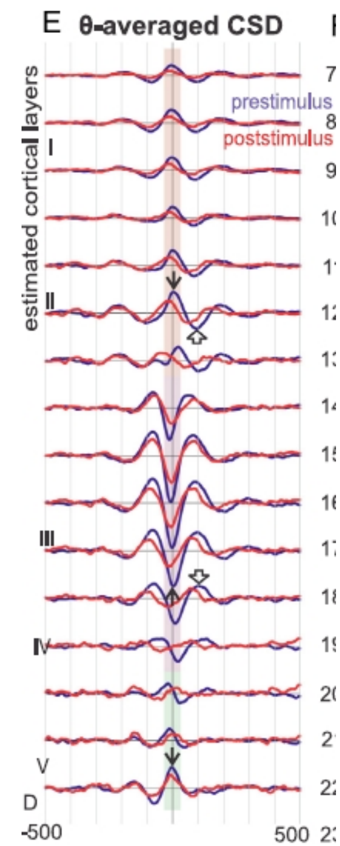
Cortex



Where do the dipoles come from?

- Cortical patches oriented perpendicular to cortex
- Recent paper: Halgren et al. (2015), proposed dipolar field arises from alternating firing of layers 2/3 and layers 5/6
- Layer 4 is input layer





How to separate? Decorrelation?

- Our first thought is decorrelation, i.e. find **A** and **S** such that the rows of **S** are orthogonal

- Unfortunately decorrelation is not unique, there are an infinite number of such **A**, **S** pairs

- One example is PCA, which projects onto the eigenvectors of covariance matrix:

$$\mathbf{X}\mathbf{X}^{T} / N = \mathbf{U}\mathbf{D}\mathbf{U}^{T}$$

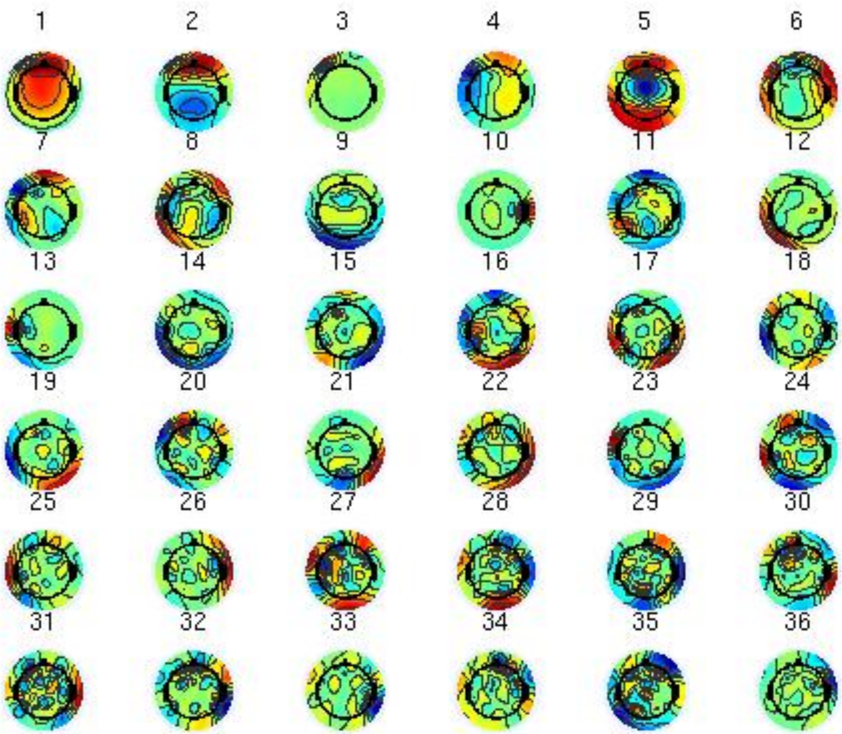
where the columns of **U** are the eigenvectors

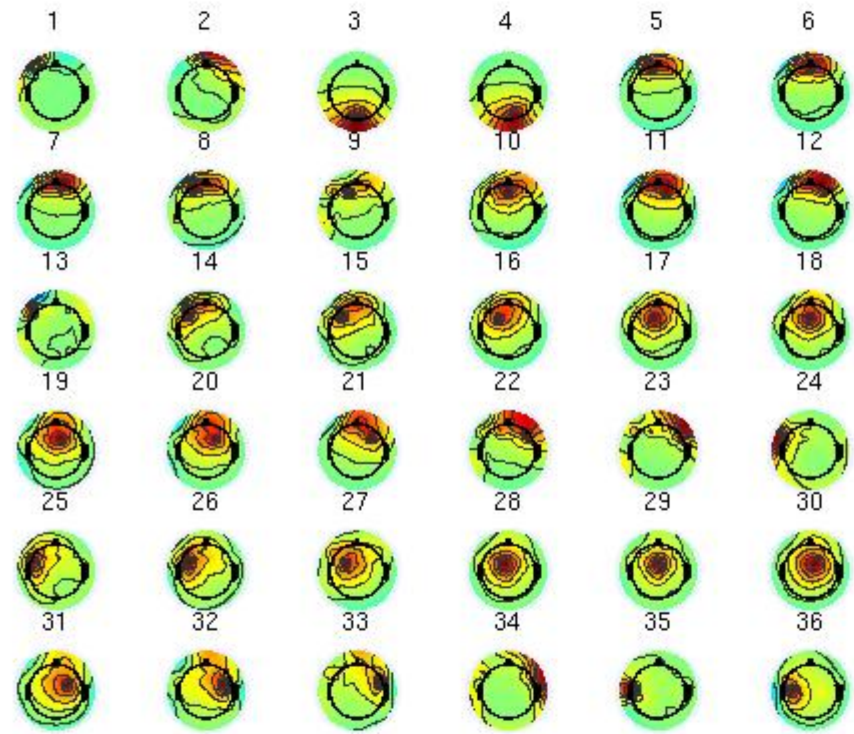
UCSD

Swartz Center for Computational Neuroscience

PCA and Sphering component maps

- PCA maps (left) are eigenvectors–orthogonal, unrealistic
- Sphering components (right) – all radial, localized





Independent Component Analysis

- Rather than try to reduce (or eliminate) correlation between sources, try to reduce statistical dependence

- Independence is defined mathematically by factorizability of the joint probability density:

$$p_s\big(s_1(t), s_2(t),\ldots, s_n(t)\big) = p_1\big(s_1(t)\big) \cdot p_2\big(s_2(t)\big) \cdots p_n\big(s_n(t)\big)$$

- Mutual information is a measure of how much the joint density differs from the product of the marginal densities, specifically it is the Kullback-Leibler divergence of joint from product of marginals

How do we find independent sources?

- A straightforward approach to ICA is based on the tendency of independent random variables to become “more Gaussian” when added together

- According to the Central Limit Theorem, the distribution of

$$(X_1 + X_2 + \ldots + X_N) / \sqrt{N}$$

tends to the Gaussian density as N goes to infinity

- And in fact, even the density of the sum of two independent random variables is in a sense more Gaussian than the density of either of the original variables

How do we find independent sources?

- According to our model, the independent sources are mixed linearly

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$$

- We generally seek an "unmixing" matrix transformation of the data to reproduce an estimate of the unknown sources

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) = \mathbf{W}\mathbf{A}\mathbf{s}(t)$$

- If we are successful, then **WA** = **I** (identity or a permutation matrix), and $\mathbf{y}(t)$ contains the original sources
- So each source estimate $y_i(t)$ is a linear combination of the observed EEG data $\mathbf{x}(t)$,

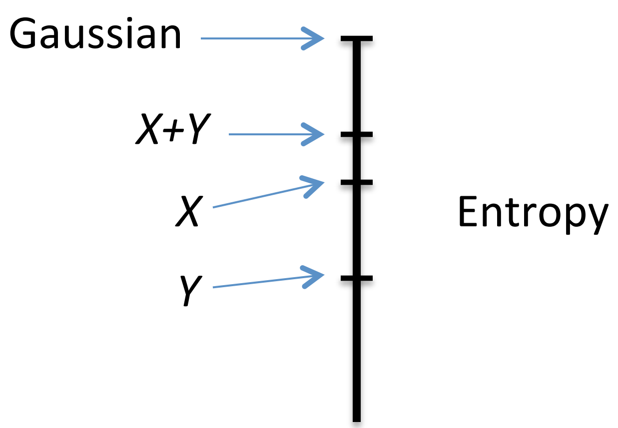
$$y_i(t) = \mathbf{w}^T\mathbf{x}(t)$$

- And we want $\mathbf{w}^T$ to be a row of the inverse of **A**

Measures of Non-Gaussianity: Entropy

- One commonly used measure is ***entropy***. If we limit consideration to variables with fixed variance, then the Gaussian distribution has maximum entropy
- This means that any ***non-Gaussian*** random variables with the same variance have ***lower entropy*** than Gaussian, and sums of random variables (normalized to unit variance) have higher entropy than the original variables

- To perform ICA using entropy, we attempt to minimize entropy



Measures of Non-Gaussianity: Kurtosis

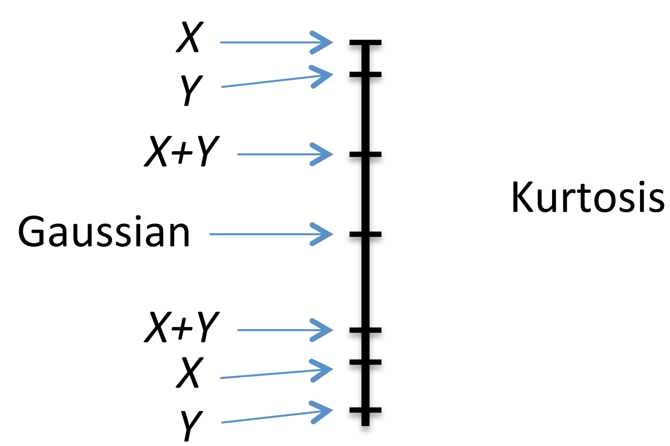
- Another commonly used measure is ***kurtosis***, which for a unit variance random variable is given by:

$$\text{Kurtosis}(X) = E\{X^4\} - 3$$

4^{th} moment of X

4^{th} moment of Gaussian

- So kurtosis is the difference between a moment of X and the same moment of Gaussian (both with unit variance)
- In this case, some random variables are on one side of Gaussian, some on the other
- Sums are closer to Gaussian
- ICA tries to push "away" from Gaussian



Measures of Non-Gaussianity: General

- More generally, we may consider the difference:

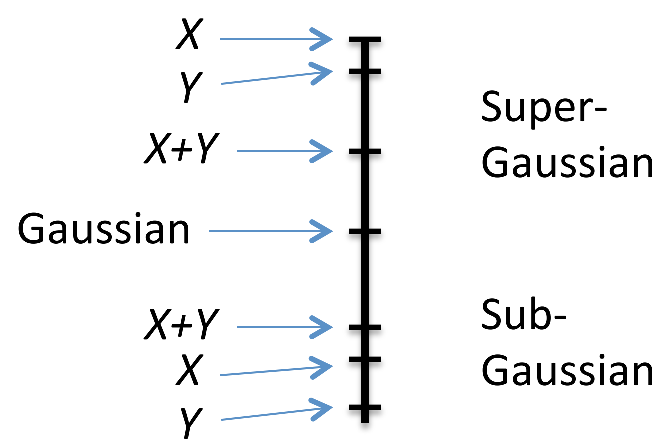
$$E\{G(X)\} - E\{G(Z)\}$$

moment of X — moment of Gaussian

- Variables on one side of Gaussian are called *super-Gaussian*, and variables on the other side are called *sub-Gaussian*

- Super-Gaussian variables are pushed "up" (maximize E{G(X)}), and sub-Gaussian are pushed "down" (minimize E{G(X)})

- Maximize the *magnitude* of the difference



Optimal Measures

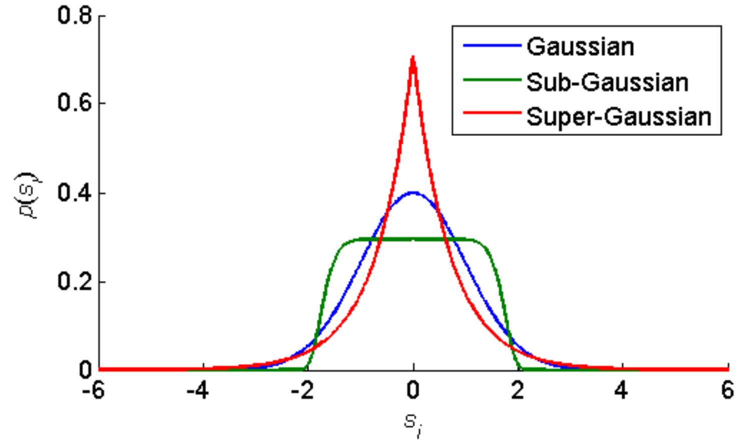
- The optimal measure to use in terms of estimation efficiency is based on the source density itself, and is related to entropy

- However ICA can be performed in principle simply using kurtosis, or other more general fixed measures

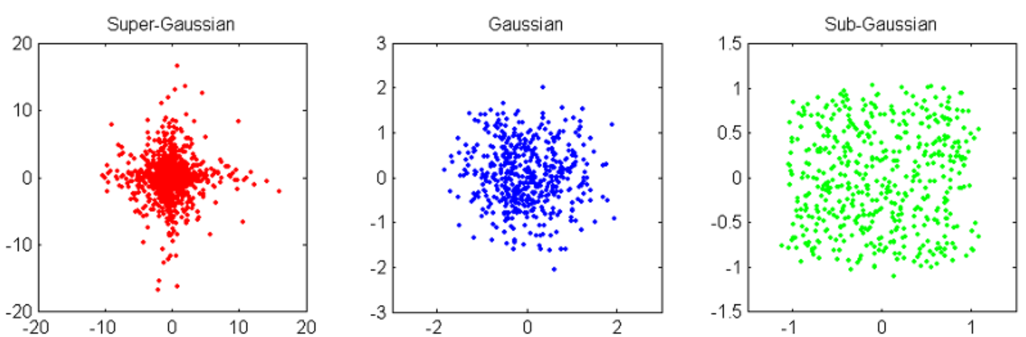
- Generally we only need to determine whether the source we are estimating is super-Gaussian or sub-Gaussian, to know whether to maximize or minimize kurtosis, or to know which of two particular measures to maximize

Sub- and Super-Gaussian Densities

- Gaussian: limiting distribution of sums of random variables
- Super-Gaussian: heavier tails, sharper peak, positive kurtosis
- Sub-Gaussian: light tails, like uniform density, negative kurtosis



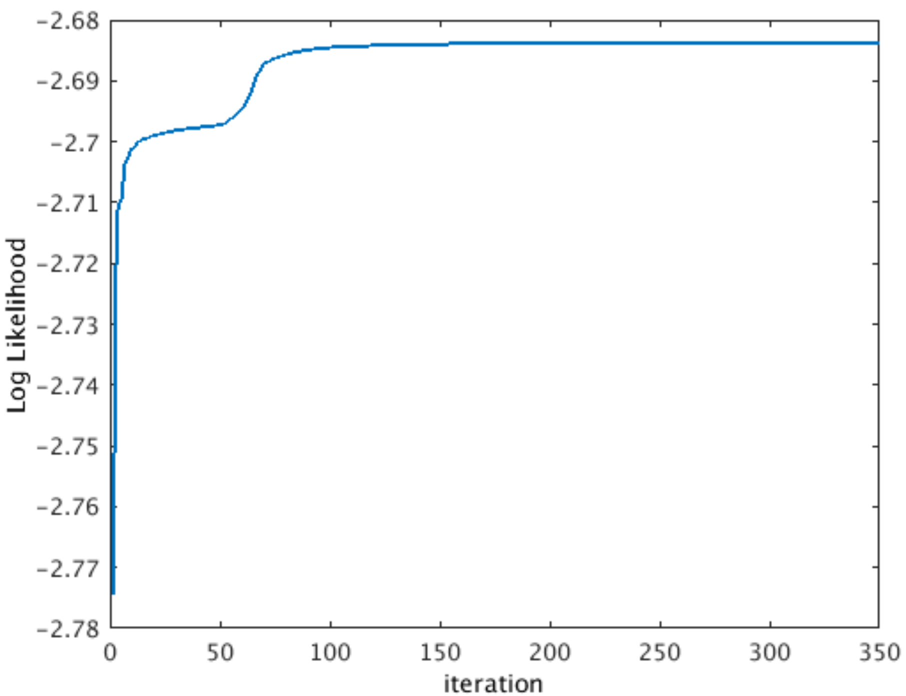
- Scatter plots of two independent random variables:

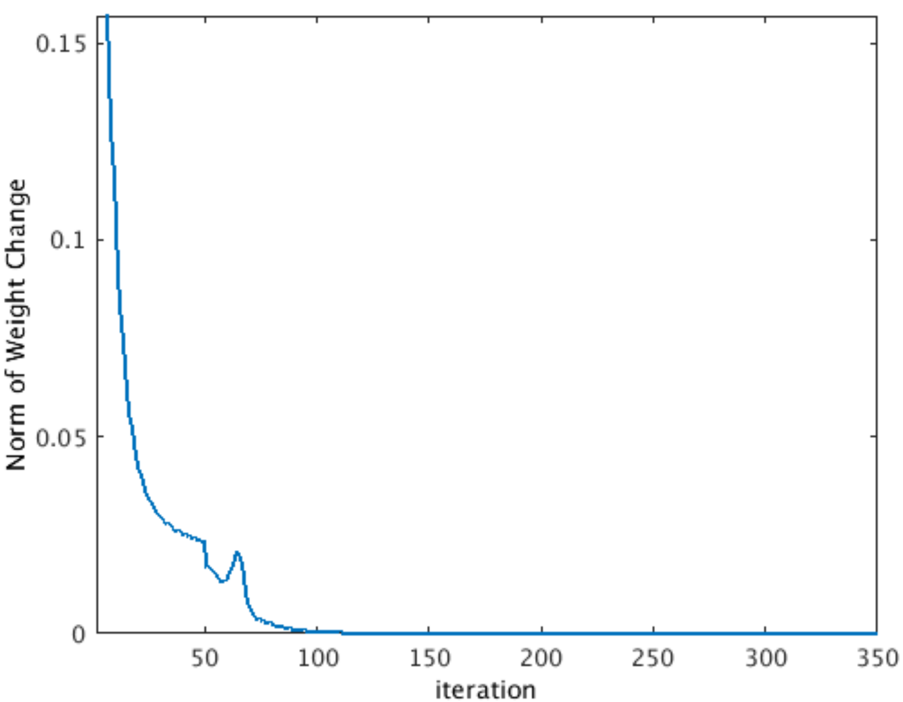


ICA Optimization

- Log Likelihood increases with iteration
- Change to Newton at iteration 50

- Norm of weight change decreases (parameters gradually stop changing)



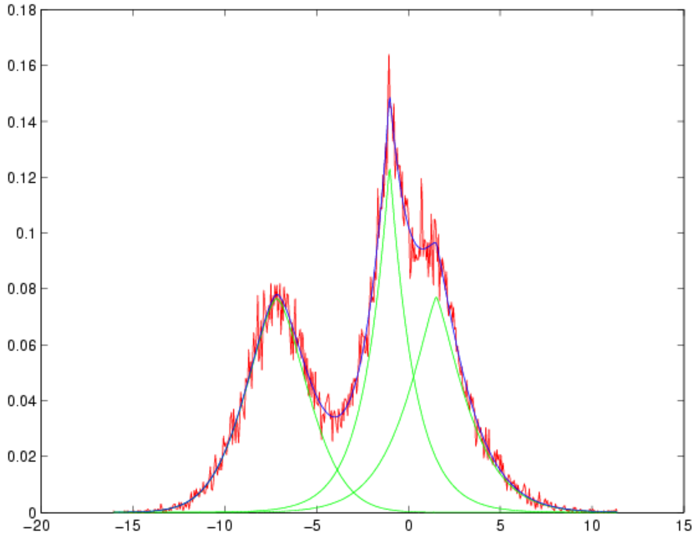


AMICA Source Density Mixture Model

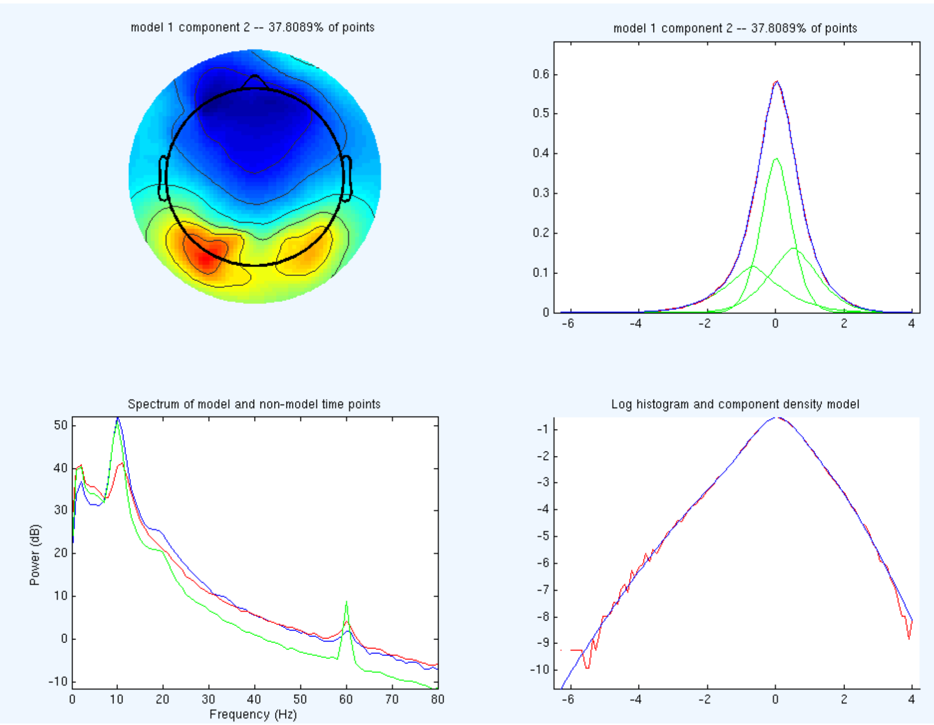
- Each source density mixture component has unknown location, scale, and shape:

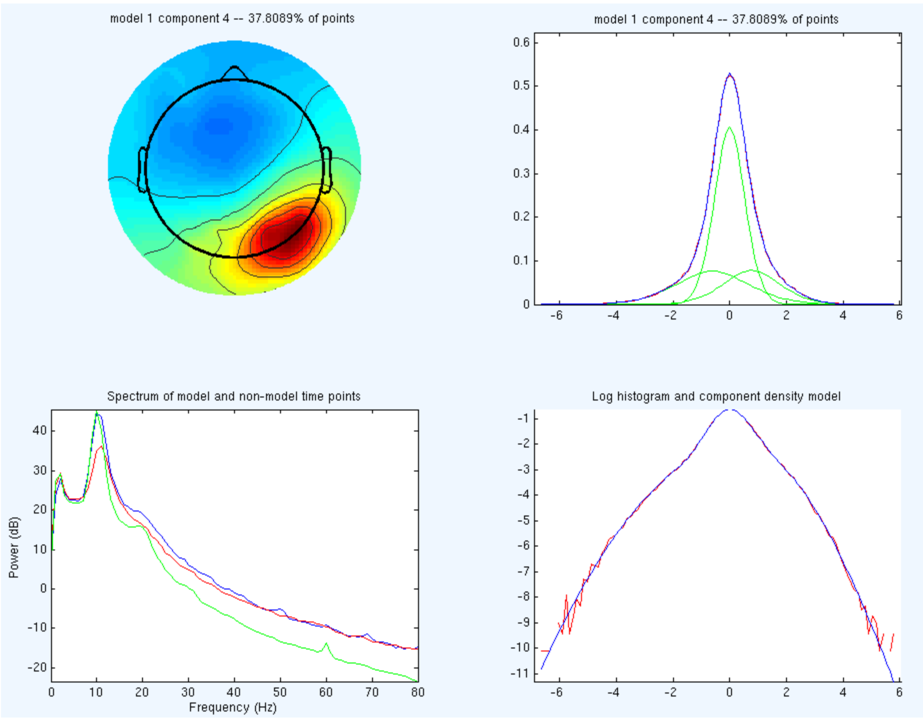
$$q_{hi}\big(s_i(t)\big) = \sum_{j=1}^{m} \alpha_{hij} \sqrt{\beta_{hij}}\, q_{hij}\Big(\sqrt{\beta_{hij}}\big(s_i(t) - \mu_{hij}\big)\,;\rho_{hij}\Big)$$

- Generalized Gaussian mixture model is convenient and flexible

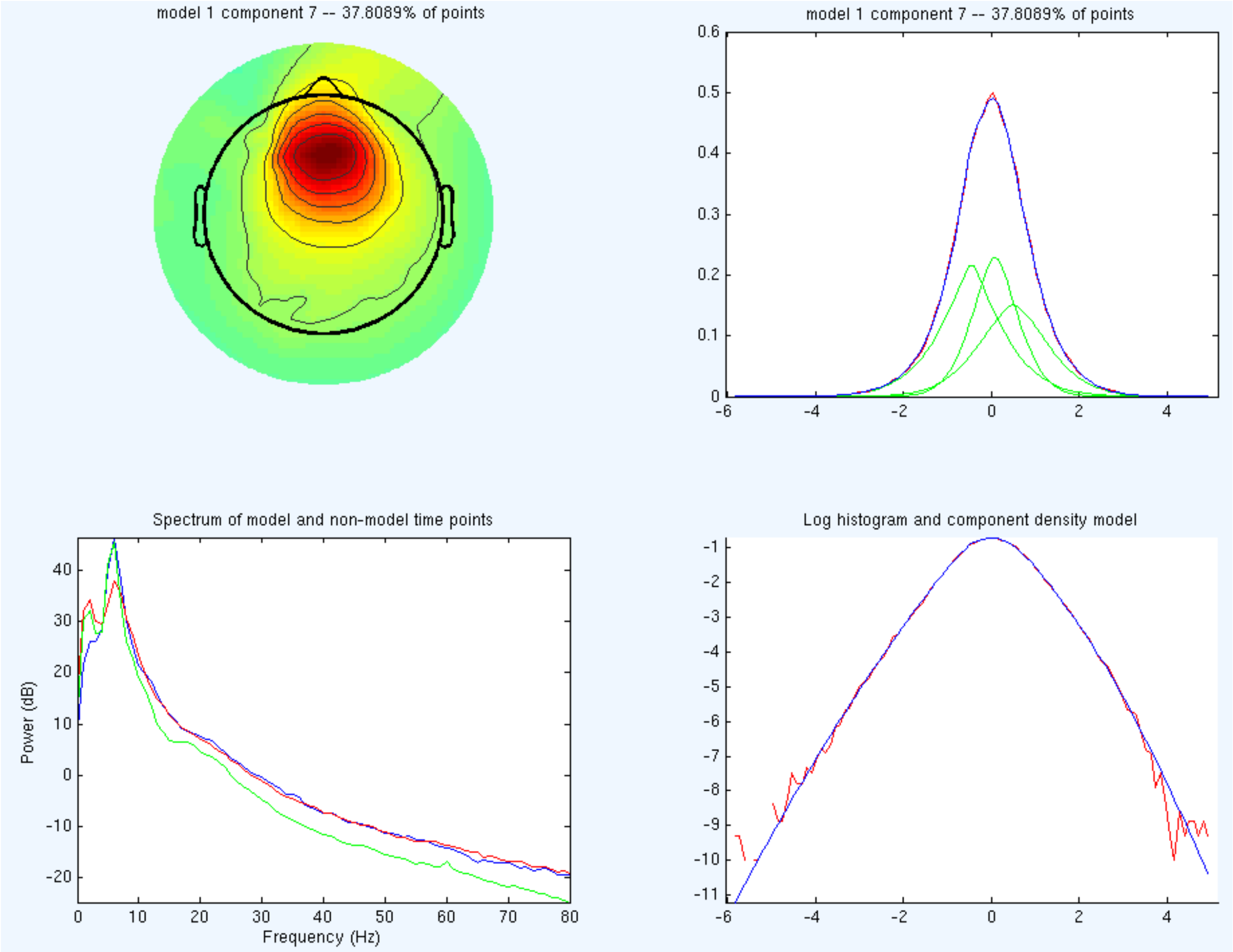


Alpha components



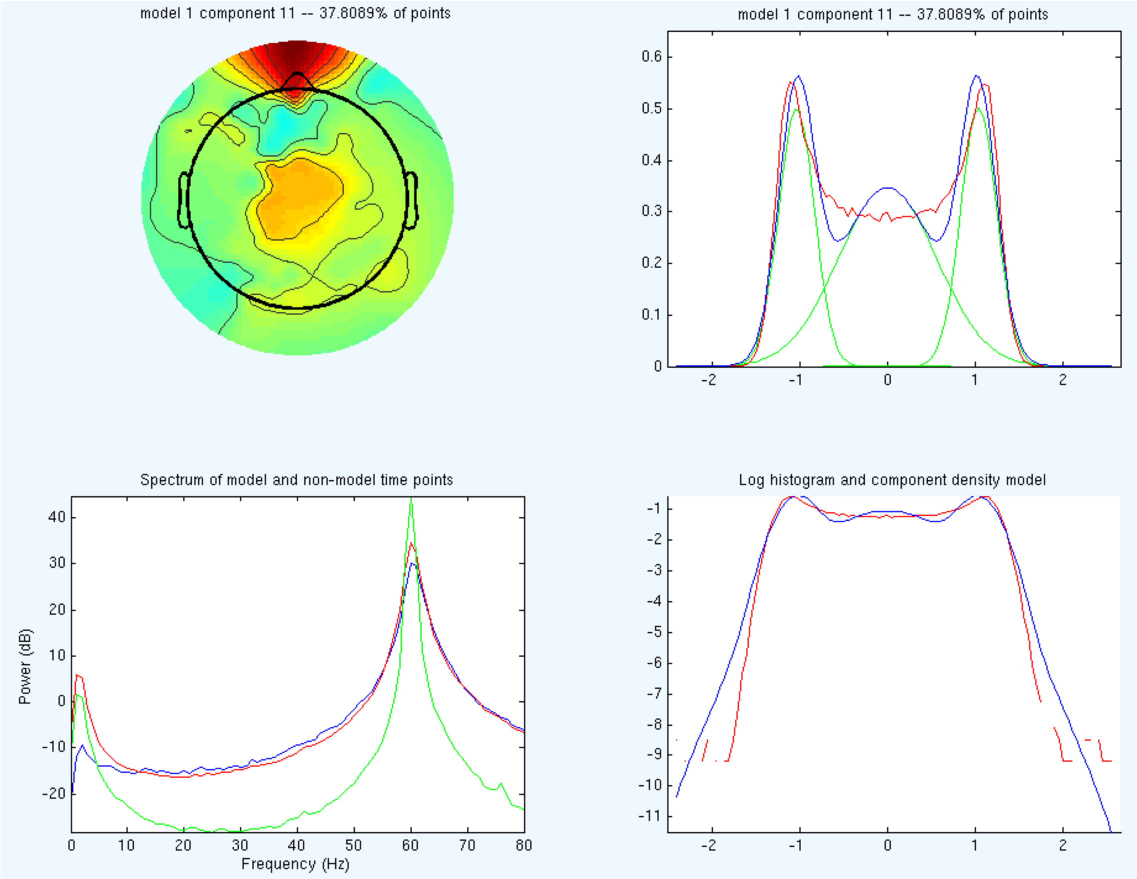


Frontal midline θ

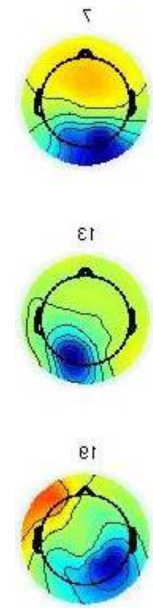


Power line component

- Sub-Gaussian component represented by mixture model of Generalized Gaussian densities



Dipolarity and biological plausibility

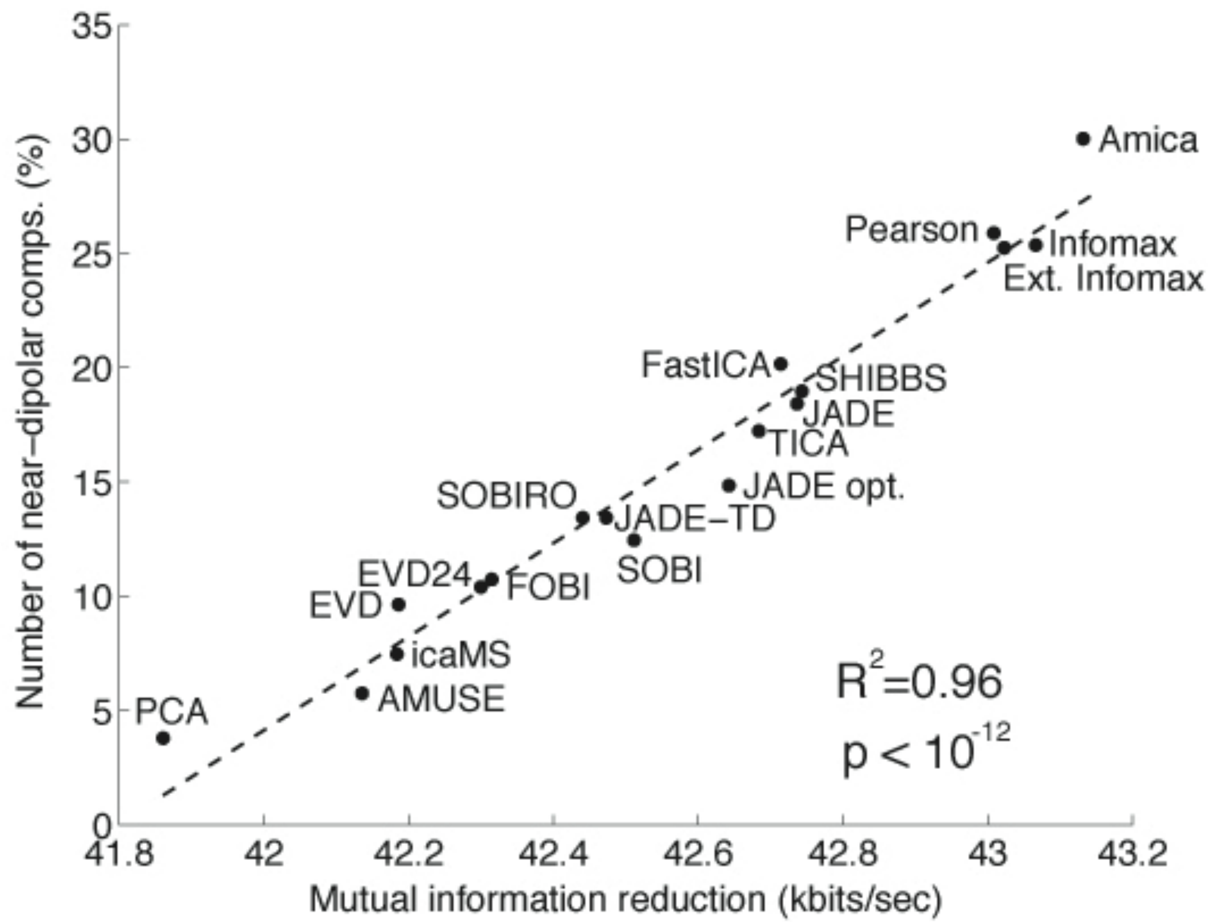


- Dipolarity is measured by fitting a single dipole (projection) to the measured component map and computing *residual variance*

- The dipolarity of a decomposition is the percentage of the estimated components with a residual variance (squared error in dipole fit) less than some threshold (typically 5%)

Comparison Dipolarity vs. MIR

Experiment with 14 datasets of 71 channel data, 22 ICA algorithms tested

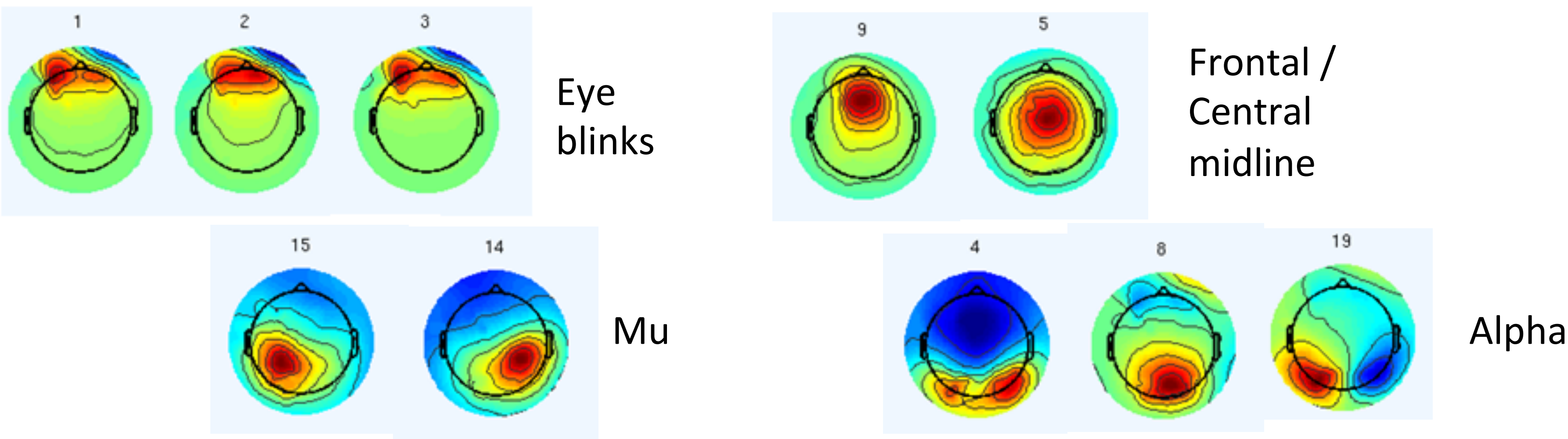


UCSD

Swartz Center for Computational Neuroscience

Dependent Source Subspaces

- The “sources” may not be independent, but may consist of dependent subspaces



- Real temporally extended source activity may take place in a “space” defined by a few component maps, rather than just one

UCSD

Swartz Center for Computational Neuroscience

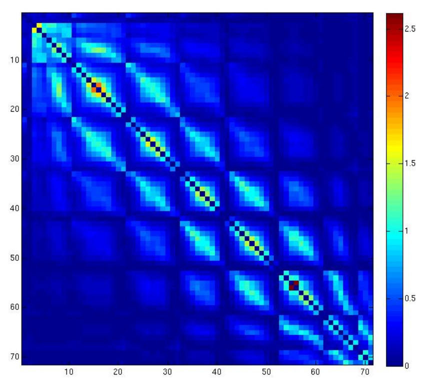
Measuring Independence: Pairwise mutual information

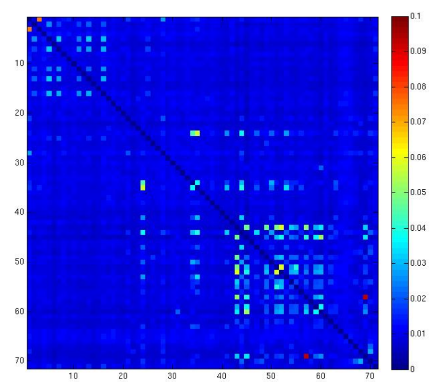
- Pairwise mutual information (PMI) between two random variable x_i and x_j:

$$[M]_{ij} = I(x_i; x_j) = h(x_i) + h(x_j) - h(x_i, x_j)$$

  PMI is a measure of dependence between sources, how non-factorial is the joint density

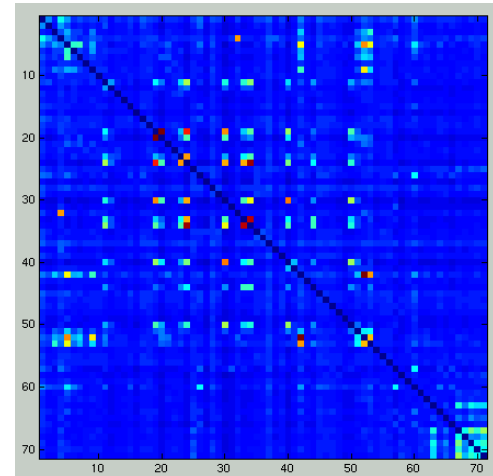
- Comparison of PMI for original data and ICA

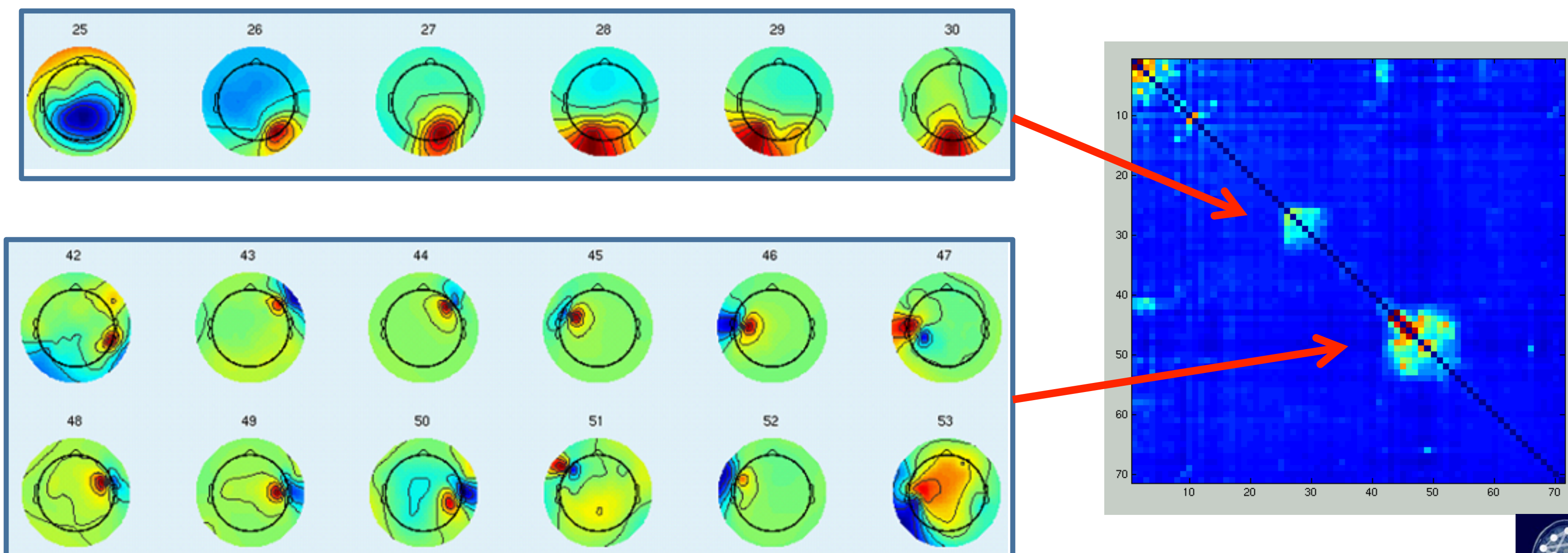




Dependent subspaces

- Residual dependence structure can be seen using Pairwise Mutual Information (PMI) plot
- Block diagonalizing this matrix (heuristically), we see blocks corresponding to dependent subspaces of components



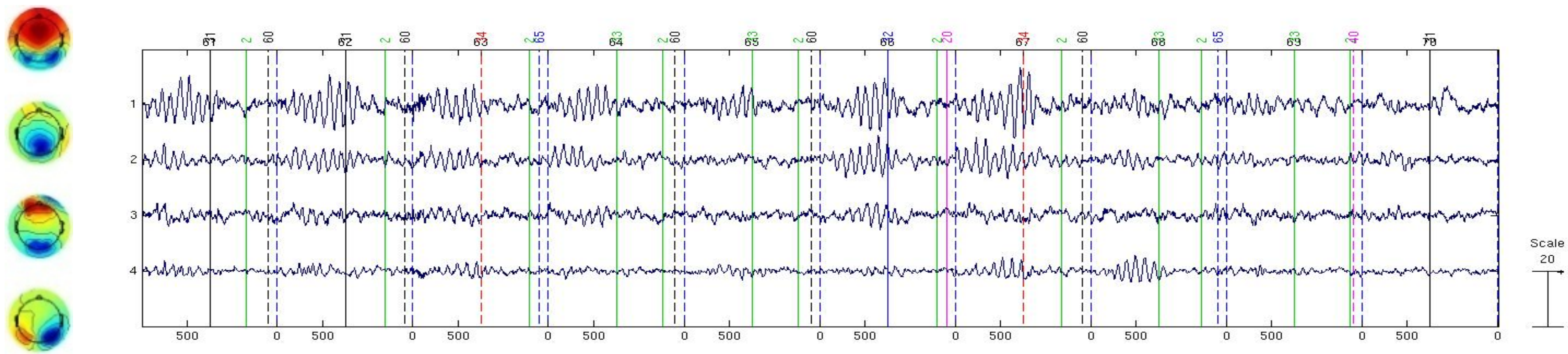


UCSD

Swartz Center for Computational Neuroscience

Alpha dependence

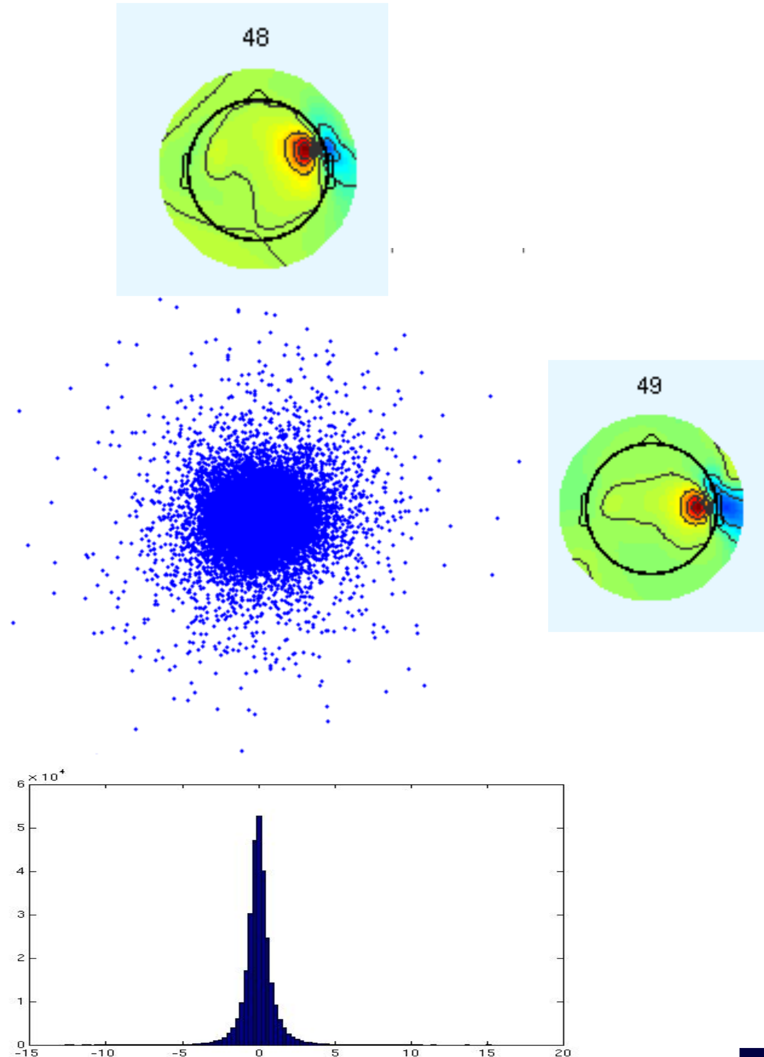
- Below four alpha components are shown



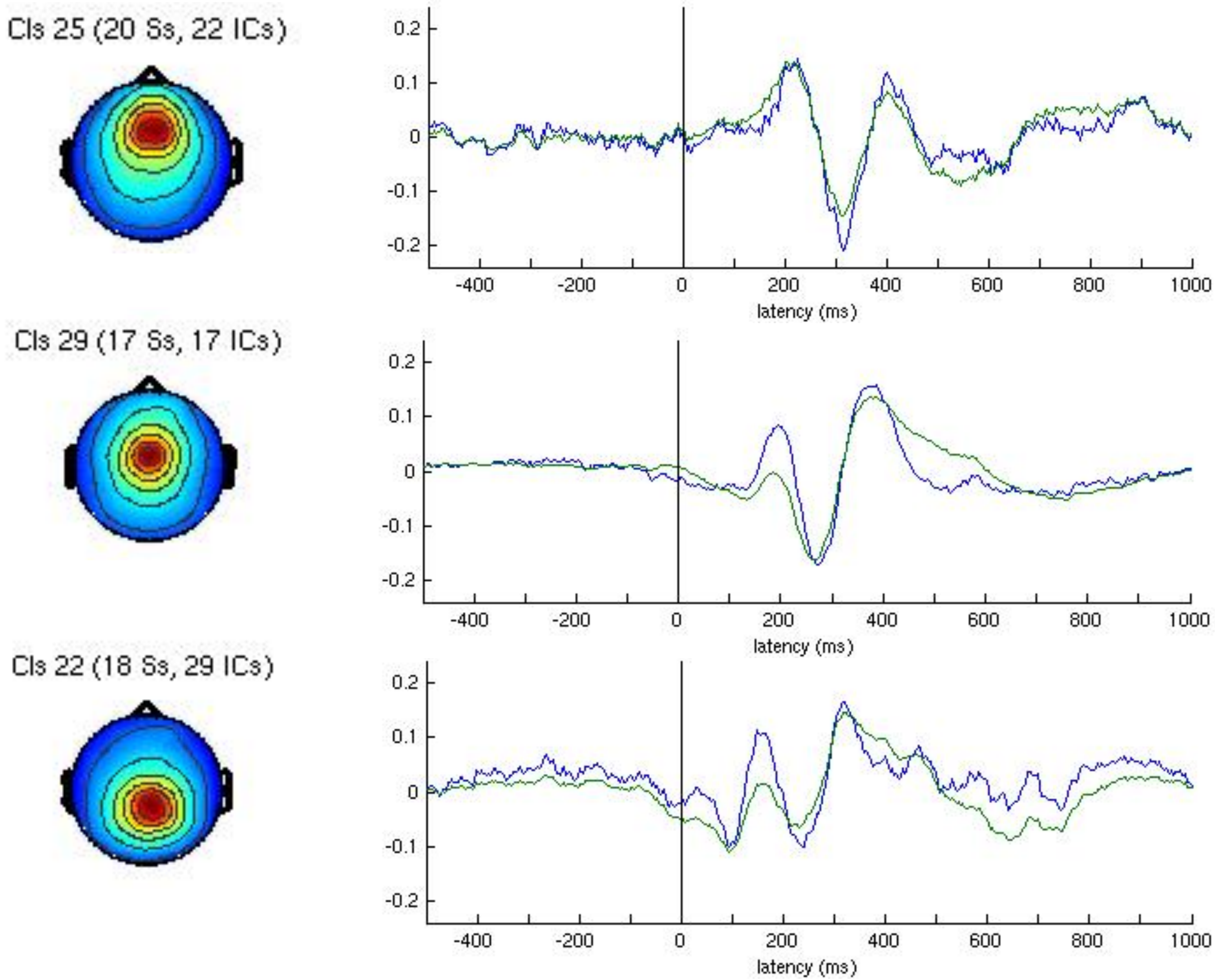
- This alpha activity exhibits dependence and coherence
- There is actually an alpha "subspace"
- Is alpha a "distributed dynamic" phenomenon?

Muscle dependence

- Muscle components tend to be active at the same time
- Activity is uncorrelated, but nevertheless dependent
- Activity is non-Gaussian, marginal histograms are "sparse"



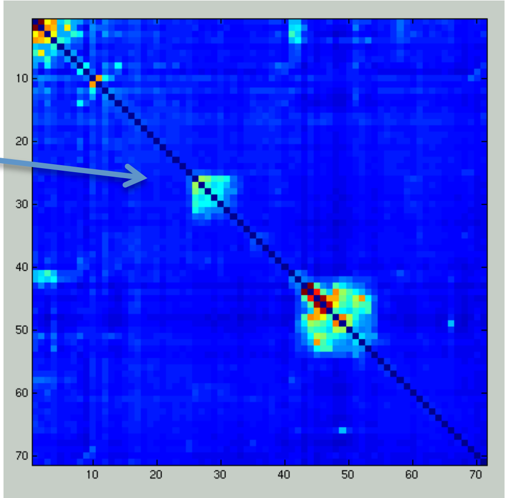
Multiple Theta Components



How does ICA perform Independent Subspace Analysis (ISA)?

- ICA attempts to minimize mutual information (dependence) in estimated sources
- ICA will generally separate (isolate) subspaces as well since the cost function (or "contrast function") can be reduced by by eliminating linear dependence (mixing) without increasing dependence within the dependent subspace

Dependence on this subspace is eliminated from other sources because any residual linear dependence increases the "cost function"



J. A. Palmer and S. Makeig, "Contrast Functions for Independent Subspace Analysis," Proceedings of the 10th International Conference on Latent Variable Analysis and Independent Component Analysis, Lecture Notes in Computer Science, Springer, 2012.

Conclusions

- Problem of separating EEG sources is similar to the "cocktail party problem" of separating simultaneous audio sources
- Adding random variables increases "Gaussianity". ICA works by reversing the process, "pushing" sources away from Gaussian
- Sources may exhibit residual dependency, but ICA generally separates dependent "subspaces" from other sources