Independent Component Analysis of High-density Scalp EEG Recordings

Jason Palmer[1], Scott Makeig[1], Arnaud Delorme[1], Julie Onton[1], Zeynep Akalin Acar[1], Ken Kreutz-Delgado[2], Bhaskar D. Rao[2]

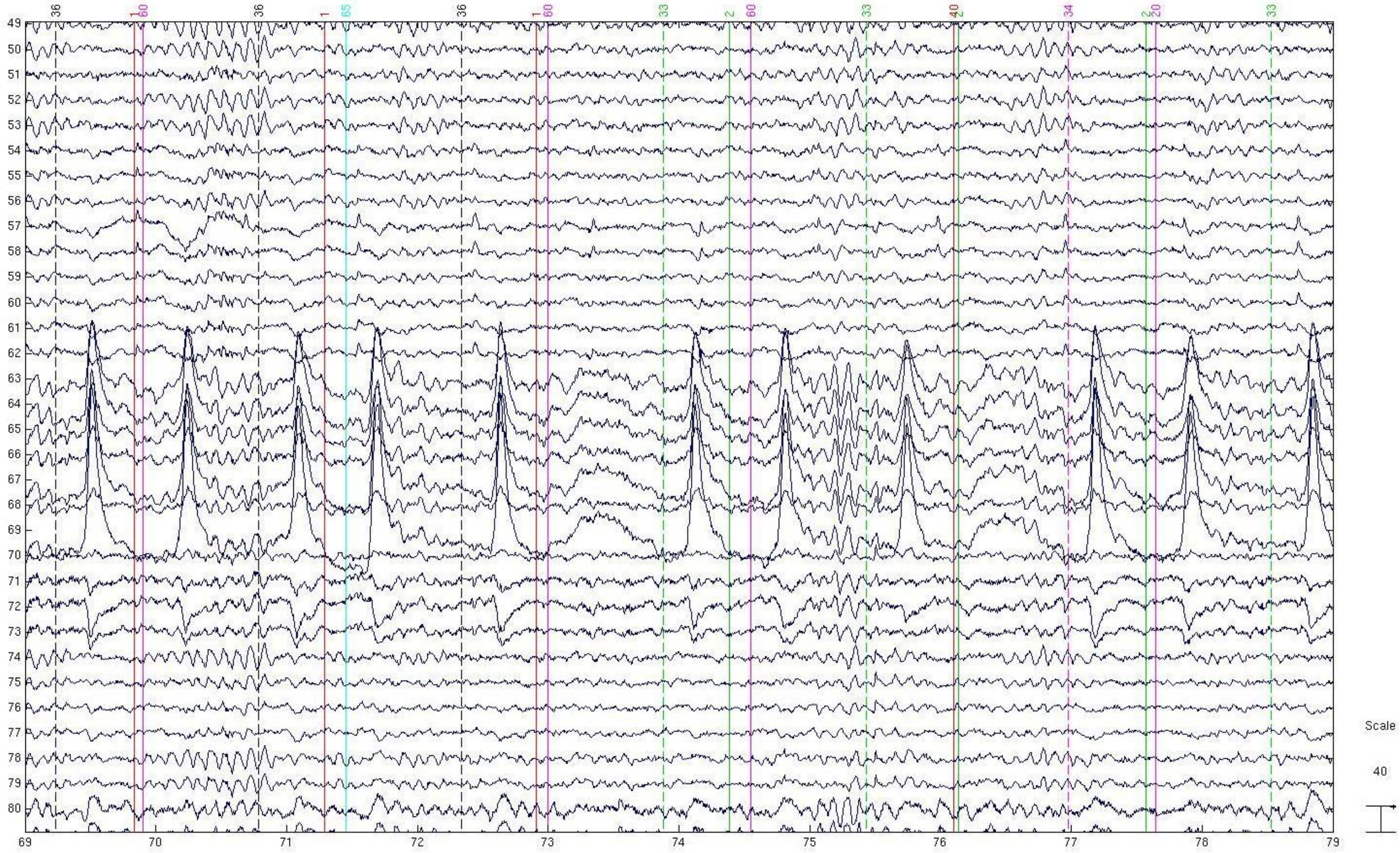
[1]Swartz Center for Computational Neuroscience,
Institute for Neural Computation
[2]Department of Electrical and Computer Engineering
University of California San Diego, La Jolla CA

Outline

- EEG data
- PCA and Sphering
- ICA
  - Strategies for performing ICA
  - How do different algorithms compare?
  - Maximum Likelihood and Mutual Information
- What about dependence among EEG sources?

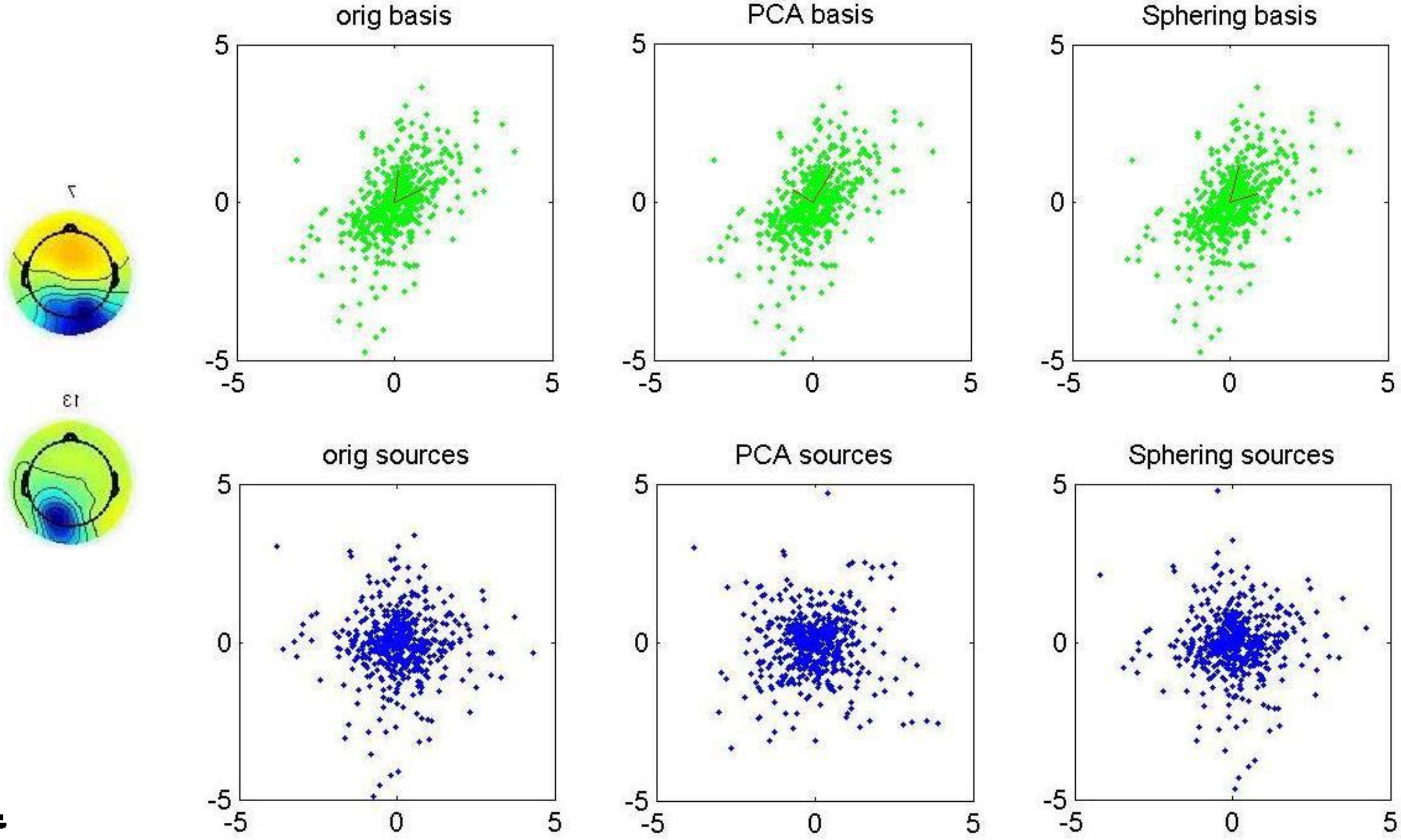
Example of raw EEG

- EEG measures superposition of source activity—neighboring channels correlated
- Volume conduction makes analysis of coherence in different regions problematic



Decorrelation

- Infinite ways to decorrelate—PCA, “Sphering”
- Sphering changes the data least of all such



Linear superposition model

- Basic linear model:

$$\text{n x N} \longrightarrow X = \overset{\text{n x n}}{A}Y \longleftarrow \text{n x N}$$

- Eigen-decomposition of covariance:

$$XX^T /N = UDU^T$$

- PCA decomposition:

$$X = (UD^{1/2})(D^{-1/2}U^TX) = A_{PCA}Y_{PCA}$$

- Sphering decomposition:

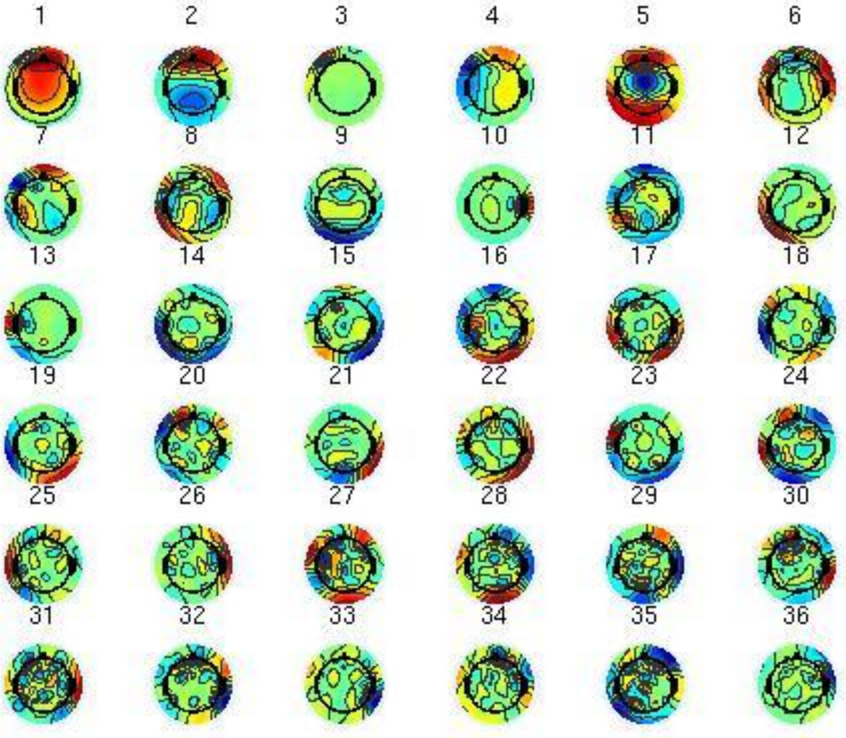
$$X = (UD^{1/2}U^T)(UD^{-1/2}U^TX) = A_{SPH}Y_{SPH}$$

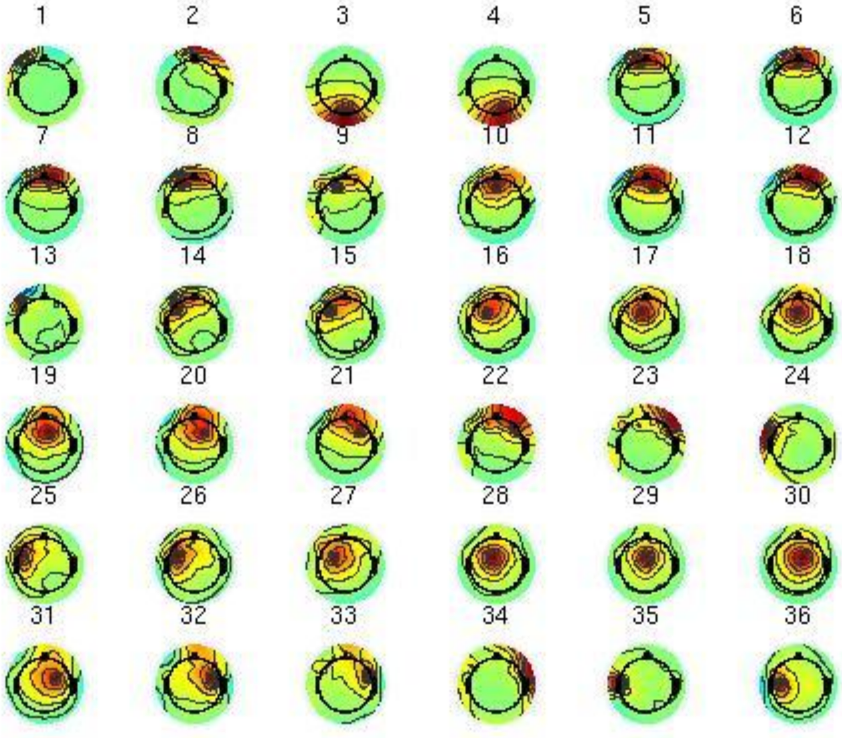
- ICA then decomposes $Y_{SPH} = A_{ICA}Y_{ICA}$ so that:

$$X = \underset{\text{icawinv}}{(A_{SPH}A_{ICA})}\,Y_{ICA}\,, \quad Y_{ICA} = (\underset{\text{icaweights}}{W_{ICA}}\,\underset{\text{icasphere}}{W_{SPH}})\,X$$

PCA and Sphering components

- PCA components (left) are eigenvectors – orthonormal, not realistic
- Sphering components (right) – all radial, localized





ICA Algorithms – strategies

- Look for sources with *independent* activity
- Mutual information and likelihood
  - Approx. MI via cumulant expansion of source density
  - Maximum likelihood
    - Fixed source densities – Infomax, FastICA
    - Adaptive / parametric source densities – Pearson, Amica, Extended Infomax
- Multiple lag decorrelation – SOBI, AMUSE, etc.
- Tensor diagonalization – JADE, SHIBBS, FOBI
- Multiple lag tensor diagonalization – JADE-TD

UCSD

Swartz Center for Computational Neuroscience

Maximum Likelihood Framework

- Probabilistic model of EEG data is a classical linear model:

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

  where the sources $\mathbf{s}(t)$ are ***independent*** (density is product of marginal densities):

  $$p_{\mathbf{s}}(\mathbf{s}(t)) = p_1(s_1(t)) \cdot p_2(s_2(t)) \cdots p_n(s_n(t))$$

- We estimate the unmixing matrix $\mathbf{W}=\mathbf{A}^{-1}$ and estimate sources $\mathbf{y}$:

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

- Then the likelihood (prob. dens.) of one time point is:

  $$p_{\mathbf{x}}(\mathbf{x}(t)) = |\det \mathbf{W}|\ p_{\mathbf{s}}(\mathbf{y}(t))$$

- The log likelihood of the data **X** assuming temporal independence is:

  $$p(\mathbf{X}) = \prod_t p_{\mathbf{x}}(\mathbf{x}(t)), \qquad \log p(\mathbf{X}) = \sum_t \log|\det \mathbf{W}| + \log p_{\mathbf{s}}(\mathbf{W}\mathbf{x}(t))$$

- We maximize this function (optimize) with respect to **W**

Mutual Information Reduction (MIR)

- Entropy of linear transformation, $\mathbf{y} = W\mathbf{x}$

$$h(\mathbf{y}) \;=\; \log|\det W| + h(\mathbf{x})$$

- Mutual information (instantaneous) for linear transformation:

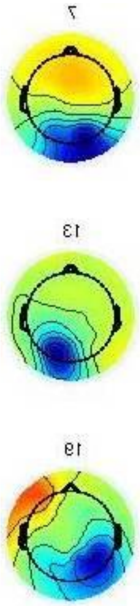
$$I(\mathbf{y}) \;=\; h(y_1) + \ldots + h(y_n) \;-\; \log|\det W| \;-\; h(\mathbf{x})$$

- Total mutual information reduction (MIR) due to linear transformation

$$\begin{aligned} \text{MIR} \;&=\; I(\mathbf{x}) - I(\mathbf{y}) \;=\; [h(x_1) + \ldots + h(x_n)] \;-\; h(\mathbf{x}) \\ &\qquad\qquad - [h(y_1) + \ldots + h(y_n)] \;+ \log|\det W| + h(\mathbf{x}) \\ &=\; \log|\det W| \;+\; [h(x_1) + \ldots + h(x_n)] \;-\; [h(y_1) + \ldots + h(y_n)] \end{aligned}$$

- Similar to ML since entropy $h(y) = E\{-\log p(y)\}$

Dipolarity and biological plausibility

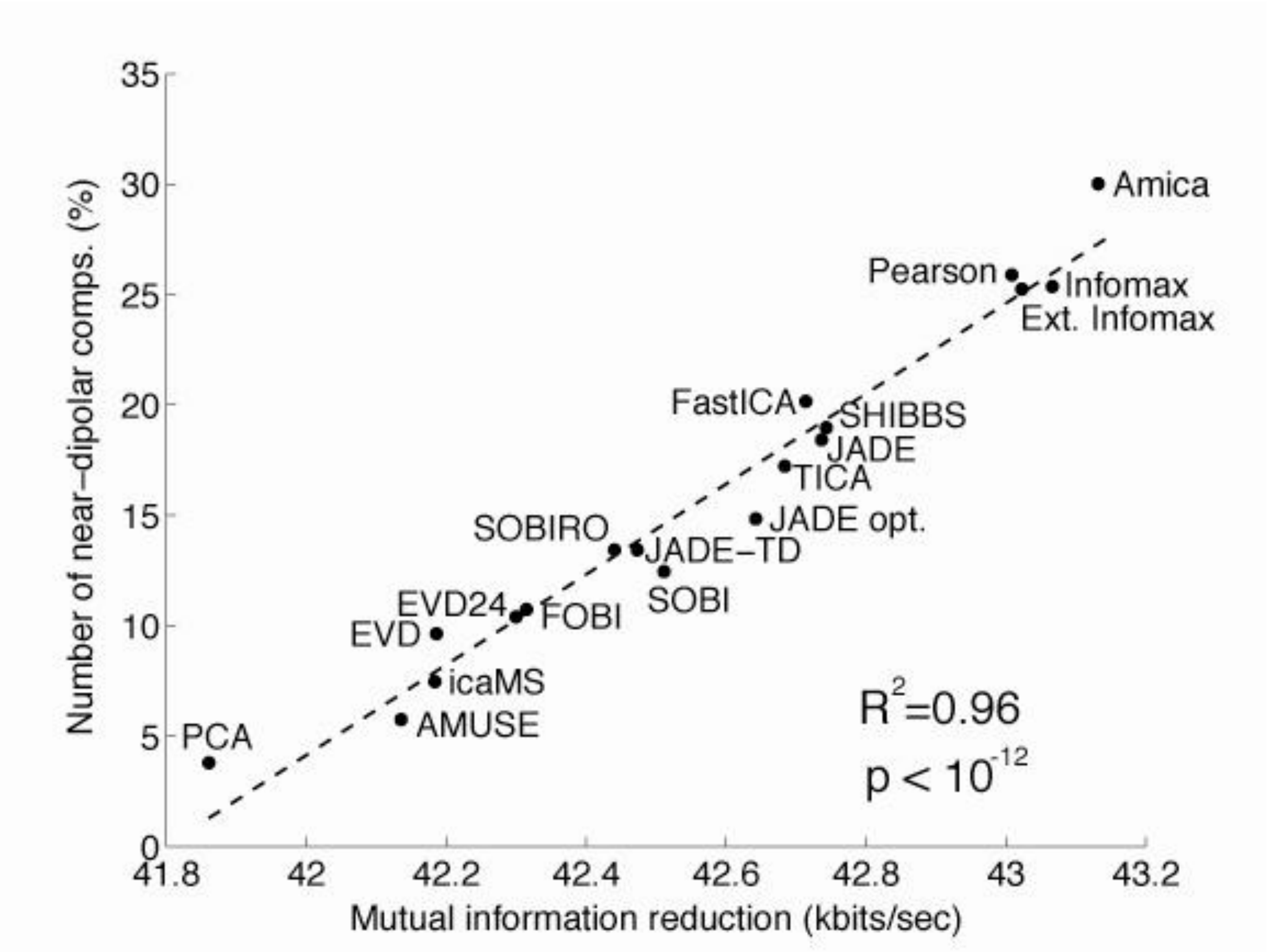


- Dipolarity is measured by fitting a single dipole (projection) to the measured component map and computing *residual variance*

- 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%)

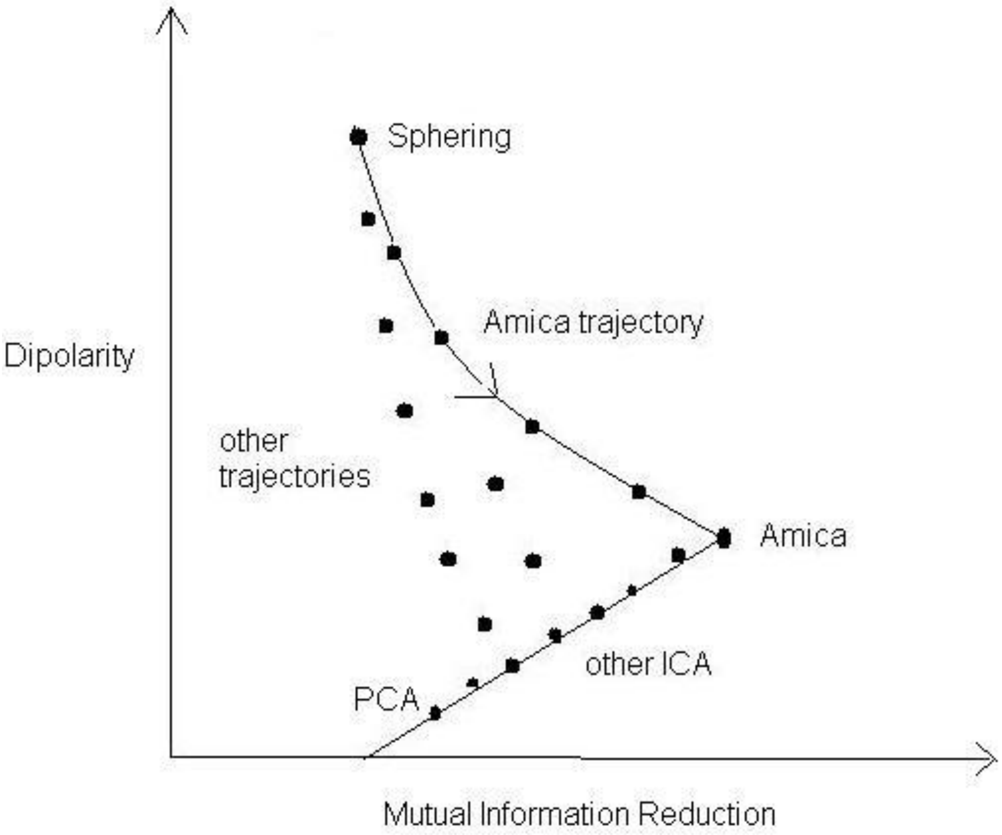
Comparison Dipolarity vs. MIR

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



Artificial dipolarity of sphering

- The Sphering decorrelating basis (not plotted in previous plot) scores high dipolarity because it consists mainly of radial dipoles (with high MI)



What does this tell us?

- The EEG sources really do have some delayed dependence. By trying to eliminate dependence at all lags, the time domain algorithms yield unrealistic (non-dipolar) components. Sophisticated algorithms that are instantaneous only, like JADE, do better.

- Algorithms that enforce decorrelation, like FastICA and JADE, seem to yeild less biologically plausible components. Sources actually have some dependence.

- Algorithms that don't enforce decorrelation, and that have adaptive source densities (like Ext. Infomax, Pearson, Amica) or have good density models to start with (Infomax) seem to do the best. There is a known higher variance in the component estimate when decorrelation is enforced, so this makes sense.

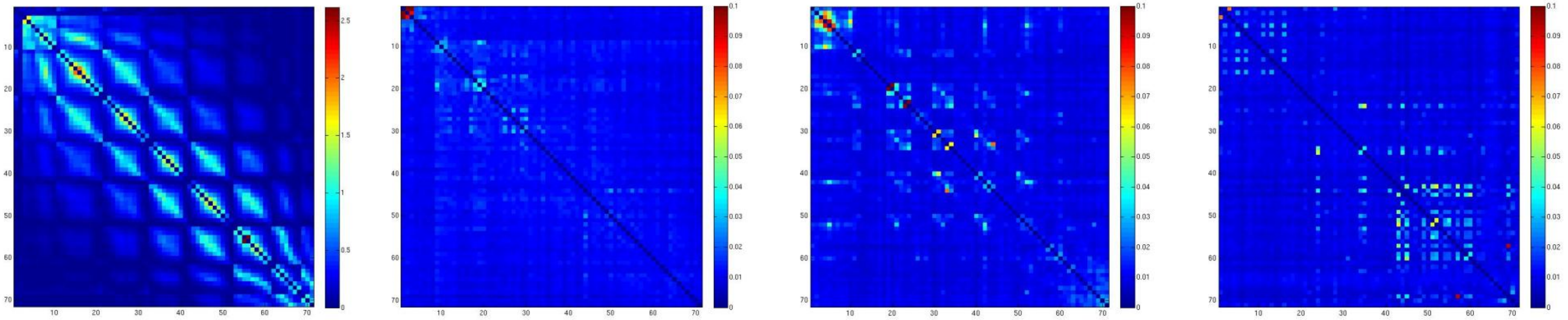
- Among the ML / min mutual info type algorithms, the better the source density is modeled, the better the algorithm does in both MIR and dipolarity. There is a known penalty in asymptotic minimum variance (CRLB) when source density model is misspecified.

Pairwise mutual information

- Pairwise mutual information (PMI):

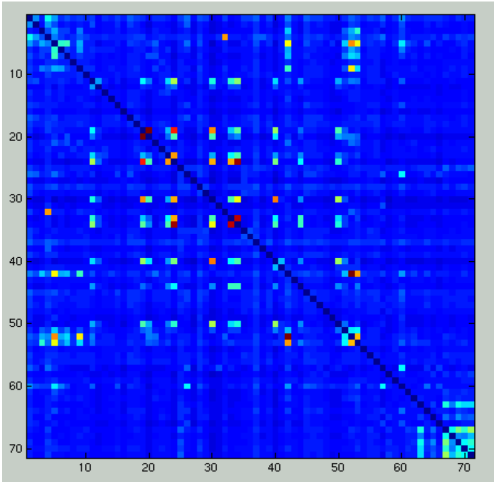
$$[\mathrm{M}]_{ij} \;=\; \mathrm{I}(x_i; x_j) \;=\; h(x_i) + h(x_j) - h(x_i, x_j)$$

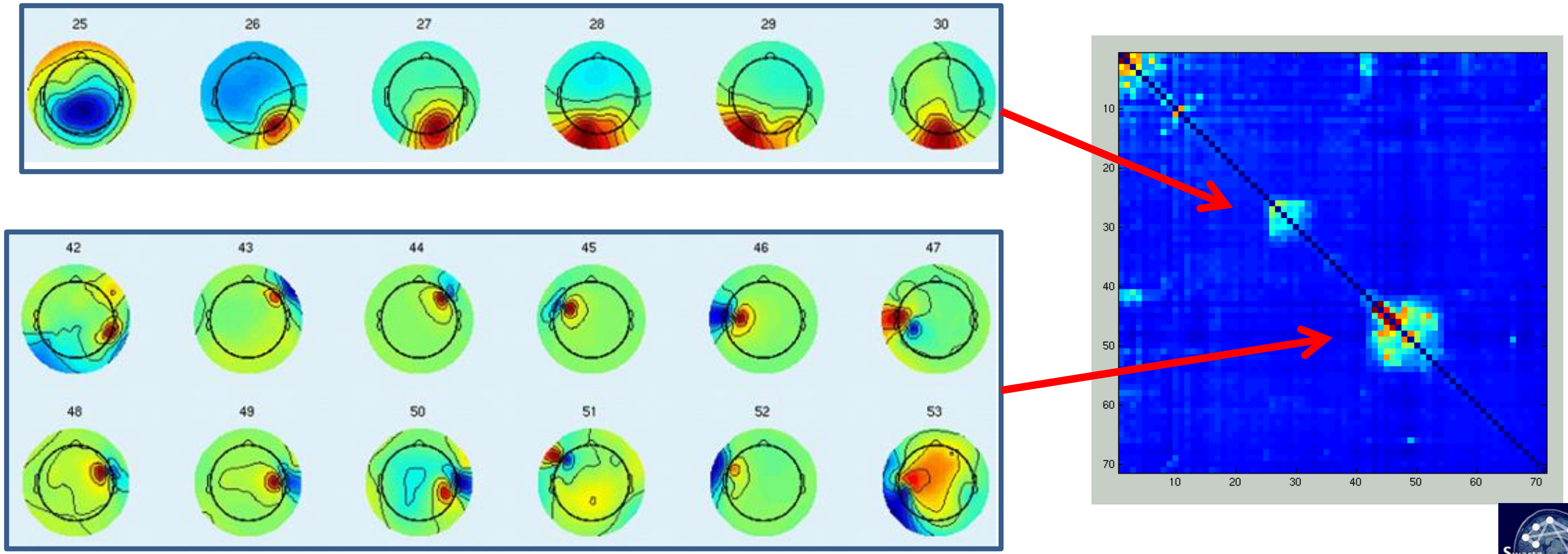
- Comparison of PMI for original data, PCA (data projected onto eigenvectors), Sphered data, ICA



Dependent subspaces

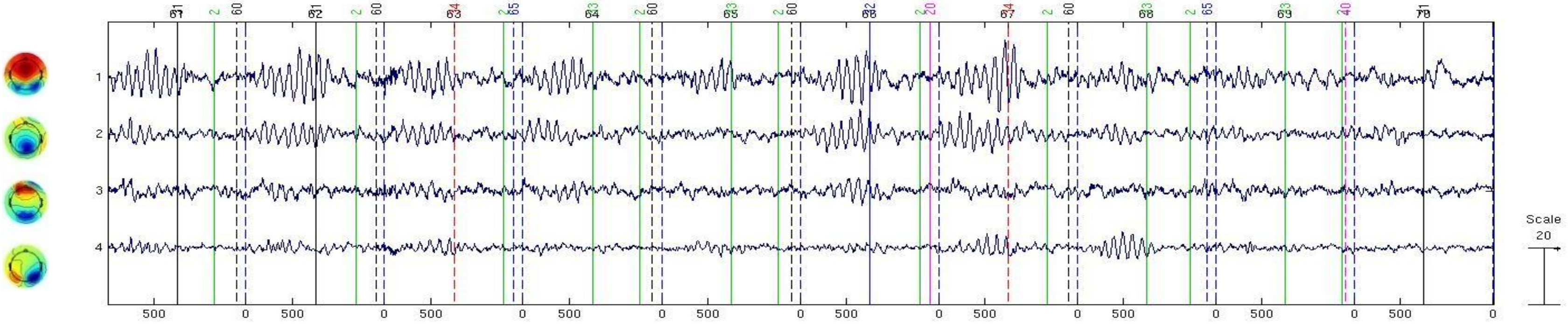
- Residual dependence structure can be seen using Pairwise Mutual Information (PMI) plot
- Block diagonalizing this matrix (heuristically), we see blocks corresponding to dependent subspaces of components





Alpha dependence

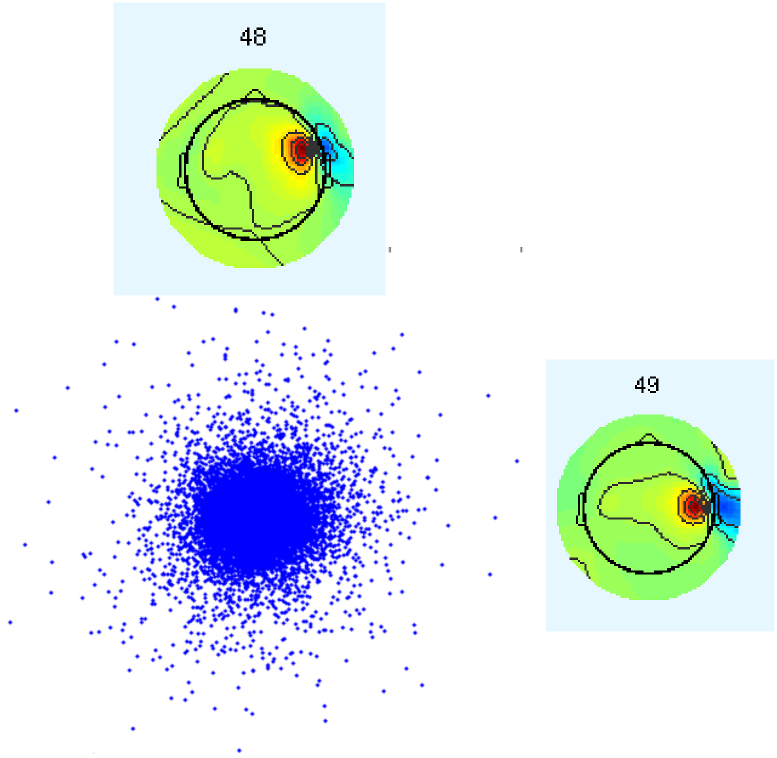
- Below several alpha components are shown

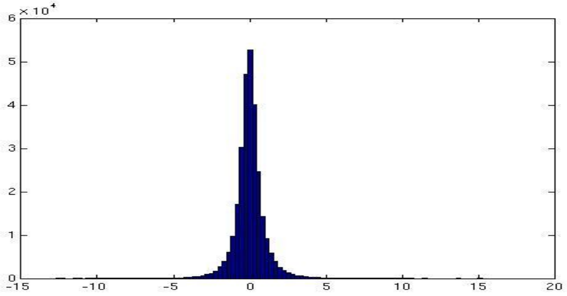


- This alpha activity exhibits dependence and coherence
- There is actually an alpha “subspace”
- Is alpha a “distributed dynamic” phenomenon?

Muscle dependence

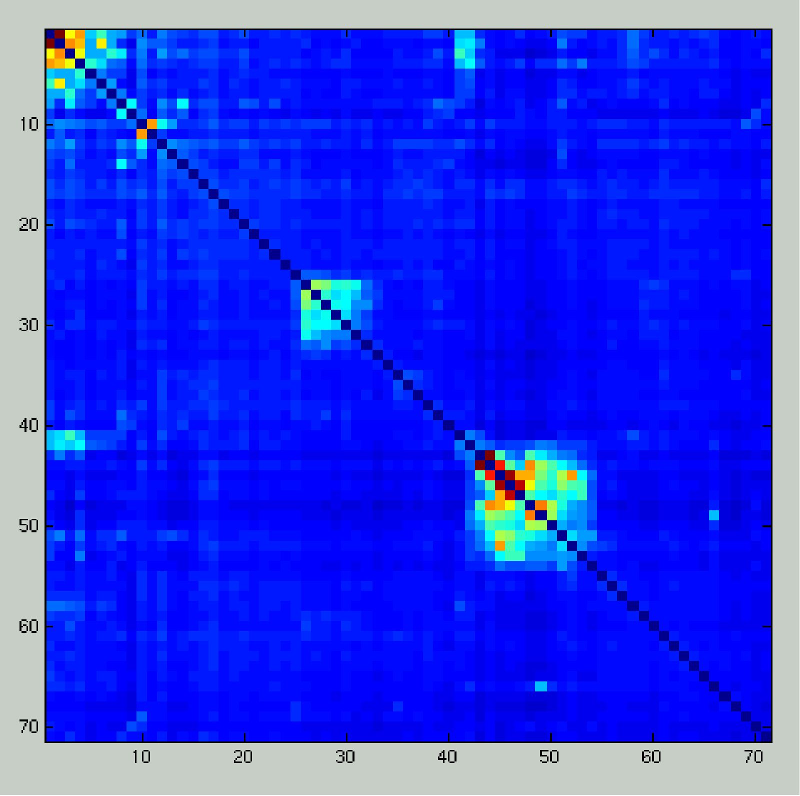
- Muscle components tend to be active at the same time
- Activity is uncorrelated, but nevertheless dependent
- Activity is non-Gaussian, marginal histograms are “sparse”





Variance Dependence and ICA

- We can show that minimizing the total mutual information will separate variance dependent sources



- PMI can be used to analyze dependence structure after ICA has been performed

Conclusion

- ICA is essentially an optimization problem
- Instantaneous ICA algorithms with adaptive source densities yield best EEG components
- Lagged decorrelation algorithms (SOBI, etc.) enforce decorrelation at all time shifts at the cost of biological plausibility
- Some EEG sources may be instantaneously dependent, e.g. alpha, and scalp muscle
- Strategy of minimizing mutual information nevertheless sound because dependent subspaces are separated from rest of sources

Thank you!

- Thanks to the Interbrain group for organizing the meeting and workshop
- Thanks to the Sloan-Swartz Foundation for supporting work at SCCN
- Thanks for your attention!