Spatial Filtering

Spatial Filtering

- Spatial filtering techniques take as input brain signals recorded from several different locations (or “channels”) and transform them in one of several ways.
- Possible goals include
  - enhancing local activity
  - reducing noise that is common across channels,
  - decreasing the dimensionality of the data,
  - finding projections that maximize discrimination between different classes

Bipolar

- Extract bipolar signals

$$\widetilde{s_{i,j}} = s_i - s_j$$

- Highlight the **electrical potential differences** between the two electrodes of interest (i and j).

Laplacian

- *Laplacian filtering*, extracts local activity at electrode *i* by subtracting the average activity present in the four orthogonal nearest neighboring electrodes

$$\tilde{\boldsymbol{s}} = \boldsymbol{s_i} - \frac{\mathbf{1}}{\mathbf{4}} \sum\nolimits_{\boldsymbol{i \in \theta}} \boldsymbol{s_i}$$

Common Average Referencing

- *Common average referencing* (CAR), enhances the local activity at electrode *i* by subtracting the average over all electrodes

$$\tilde{\boldsymbol{s}}_{\boldsymbol{i}} = \boldsymbol{s}_{\boldsymbol{i}} - \frac{\mathbf{1}}{\boldsymbol{N}} \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{N}} \boldsymbol{s}_{\boldsymbol{i}}$$

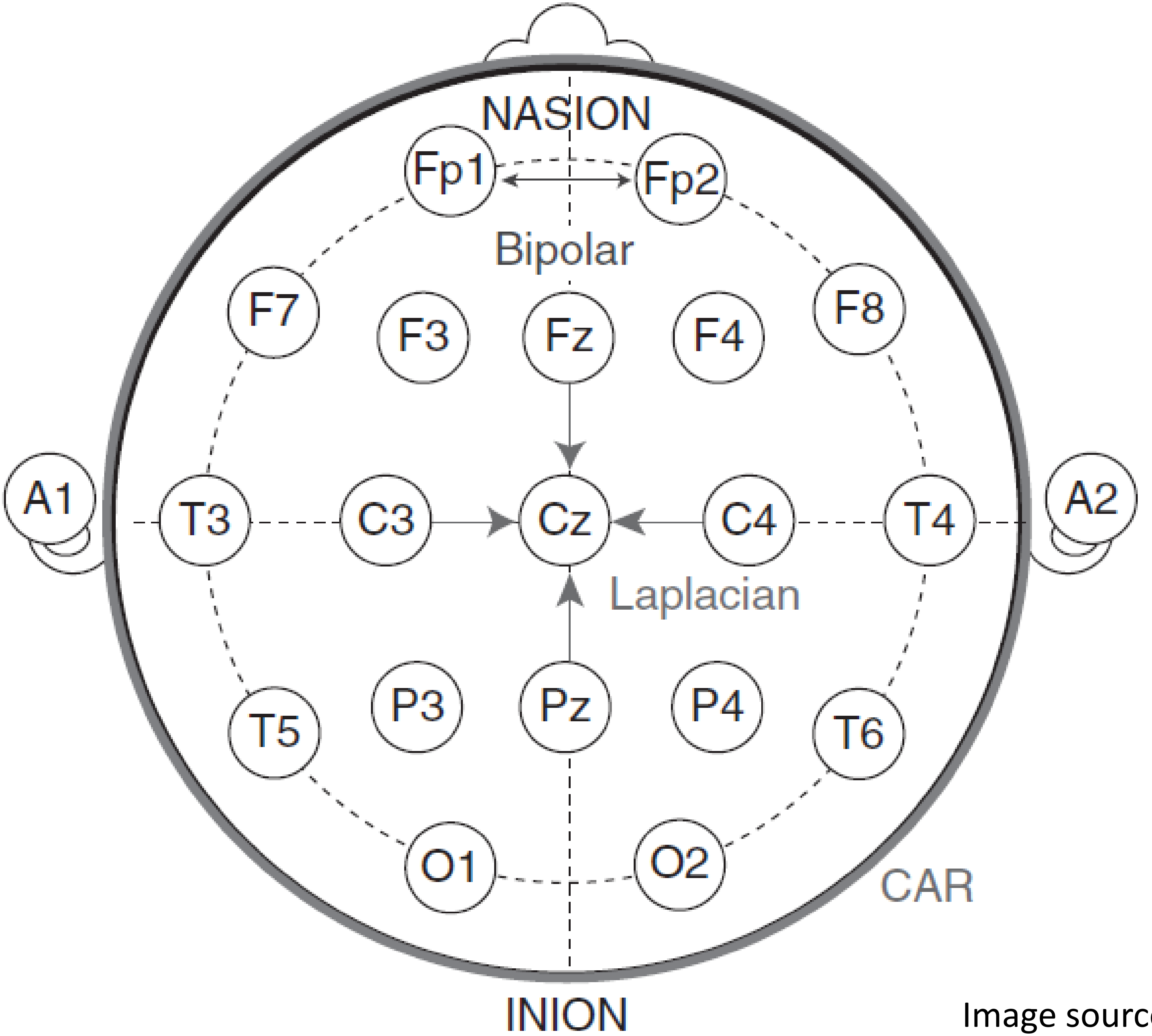


Image source: Rajesh P.N, Rao- Brain Computer Interfacing: An Introduction

Principal Component Analysis

- The goal in *principal component analysis* (PCA) (also called the *Hotelling transform*) is to discover the underlying statistical variability in the data and reduce the data's dimensionality from D to a much smaller number of dimensions L ($L << D$).

- PCA achieves this goal by
  - Finding the directions of maximum variance in the D-dimensional data
  - Rotating the original coordinate system to align with these directions of maximum variance

Principal Component Analysis

- Most natural signals, including brain signals are redundant
- In the case of EEG measurements from N electrodes
  - Measurements from nearby electrodes may be correlated
  - Underlying rhythms across multiple electrodes.
- PCA attempts to find the dominant directions of variability in the data.
- New data points can be projected along the “principal” directions. Each projection is called a “principal component”
- The resulting *L*-dimensional vector can be used as a feature vector for classification or other purposes in BCI applications

PCA - Steps

- Suppose we are given $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_M$ (N x 1) vectors

N: # of features

M: # data

Step 1: compute sample mean

$$\bar{\mathbf{x}} = \frac{1}{M}\sum_{i=1}^{M}\mathbf{x}_i$$

Step 2: subtract sample mean (i.e., center data at zero)

$$\Phi_i = \mathbf{x}_i - \bar{\mathbf{x}}$$

Step 3: compute the sample covariance matrix Σ_x

$$\Sigma_{\mathbf{x}} = \frac{1}{M}\sum_{i=1}^{M}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T = \frac{1}{M}\sum_{i=1}^{M}\Phi_i\Phi_i^T = \frac{1}{M}AA^T$$

where $A=[\Phi_1\ \Phi_2\ ...\ \Phi_M]$
i.e., the columns of A are the Φ_i

(N x M matrix)

PCA - Steps

Step 4: compute the eigenvalues/eigenvectors of Σ_x

$$\Sigma_x = \lambda_i u_i$$

where we assume $\lambda_1 > \lambda_2 > \ldots > \lambda_N$

Note : most software packages return the eigenvalues (and corresponding eigenvectors) is decreasing order – if not, you can explicitly put them in this order)

Since Σ_x is symmetric, $<u_1,u_2,\ldots,u_N>$ form an orthogonal basis in R^N and we can represent any $\mathbf{x} \in R^N$ as:

$$\mathbf{x}-\bar{\mathbf{x}} = \sum_{i=1}^{N} y_i u_i = y_1 u_1 + y_2 u_2 + \ldots + y_N u_N$$

$$y_i = \frac{(\mathbf{x}-\bar{\mathbf{x}})^T u_i}{u_i^T u_i} = (\mathbf{x}-\bar{\mathbf{x}})^T u_i \qquad if \ \ \|u_i\| = 1$$

$$\mathbf{x}-\bar{\mathbf{x}}: \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ x_N \end{bmatrix} \rightarrow \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ y_N \end{bmatrix}$$

i.e., this is just a "change" of basis!

Note : most software packages normalize u_i to unit length to simplify calculations; if not, you can explicitly normalize them)

PCA - Steps

Step 5: dimensionality reduction step – approximate $\mathbf{x}$ using only the first K eigenvectors (K<<N) (i.e., corresponding to the K largest eigenvalues where K is a parameter):

$$\mathbf{x}-\bar{\mathbf{x}}=\sum_{i=1}^{N} y_i u_i = y_1u_1 + y_2u_2 + \ldots + y_N u_N$$

approximate $\mathbf{x}$ *by* $\hat{\mathbf{x}}$
using first K eigenvectors only

$$\hat{\mathbf{x}}-\bar{\mathbf{x}}=\sum_{i=1}^{K} y_i u_i = y_1u_1 + y_2u_2 + \ldots + y_K u_K$$

(reconstruction)

$$\mathbf{x}-\bar{\mathbf{x}}: \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ x_N \end{bmatrix} \rightarrow \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ y_N \end{bmatrix} \rightarrow \hat{\mathbf{x}}-\bar{\mathbf{x}}: \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ y_K \end{bmatrix}$$

note that if K=N, then $\hat{\mathbf{x}}=\mathbf{x}$
(i.e., zero reconstruction error)

What is the Linear Transformation implied by PCA?

- The linear transformation y = Tx which performs the dimensionality reduction in PCA is:

$$\hat{\mathbf{x}} - \bar{\mathbf{x}} = \sum_{i=1}^{K} y_i u_i = y_1 u_1 + y_2 u_2 + \ldots + y_K u_K$$

$$(\hat{\mathbf{x}} - \bar{\mathbf{x}}) = U \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_K \end{bmatrix}$$

$where\ \ U = [u_1\, u_2 \ldots u_K]\ \ N\, x\, K$ matrix

i.e., the columns of U are the the first K eigenvectors of Σ_x

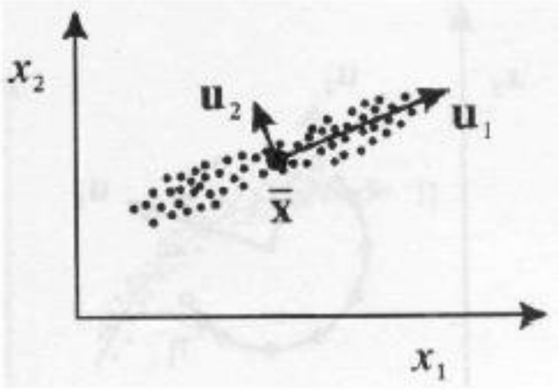
$$\Rightarrow \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_K \end{bmatrix} = U^T (\hat{\mathbf{x}} - \bar{\mathbf{x}})$$

$T = U^T$ K x N matrix

i.e., the rows of T are the first K eigenvectors of Σ_x

Interpretation of PCA

- PCA chooses the eigenvectors of the covariance matrix corresponding to the largest eigenvalues.
- The eigenvalues correspond to the variance of the data along the eigenvector directions.
- Therefore, PCA projects the data along the directions where the data varies most.
- PCA preserves as much information in the data by preserving as much variance in the data.



u_1: direction of max variance
u_2: orthogonal to u_1

How do we choose K ?

- *K* is typically chosen based on how much information (variance) we want to preserve:

Choose the smallest K that satisfies the following inequality:

$$\frac{\sum_{i=1}^{K} \lambda_i}{\sum_{i=1}^{N} \lambda_i} > T \quad where\ T\ is\ a\ threshold\ (e.g., 0.9)$$

- If T=0.9, for example, we "preserve" 90% of the information (variance) in the data.

- If K=N, then we "preserve" 100% of the information in the data (i.e., just a "change" of basis and $\hat{\mathbf{x}} = \mathbf{x}$)

Data Normalization

- The principal components are dependent on the *units* used to measure the original variables as well as on the *range* of values they assume.

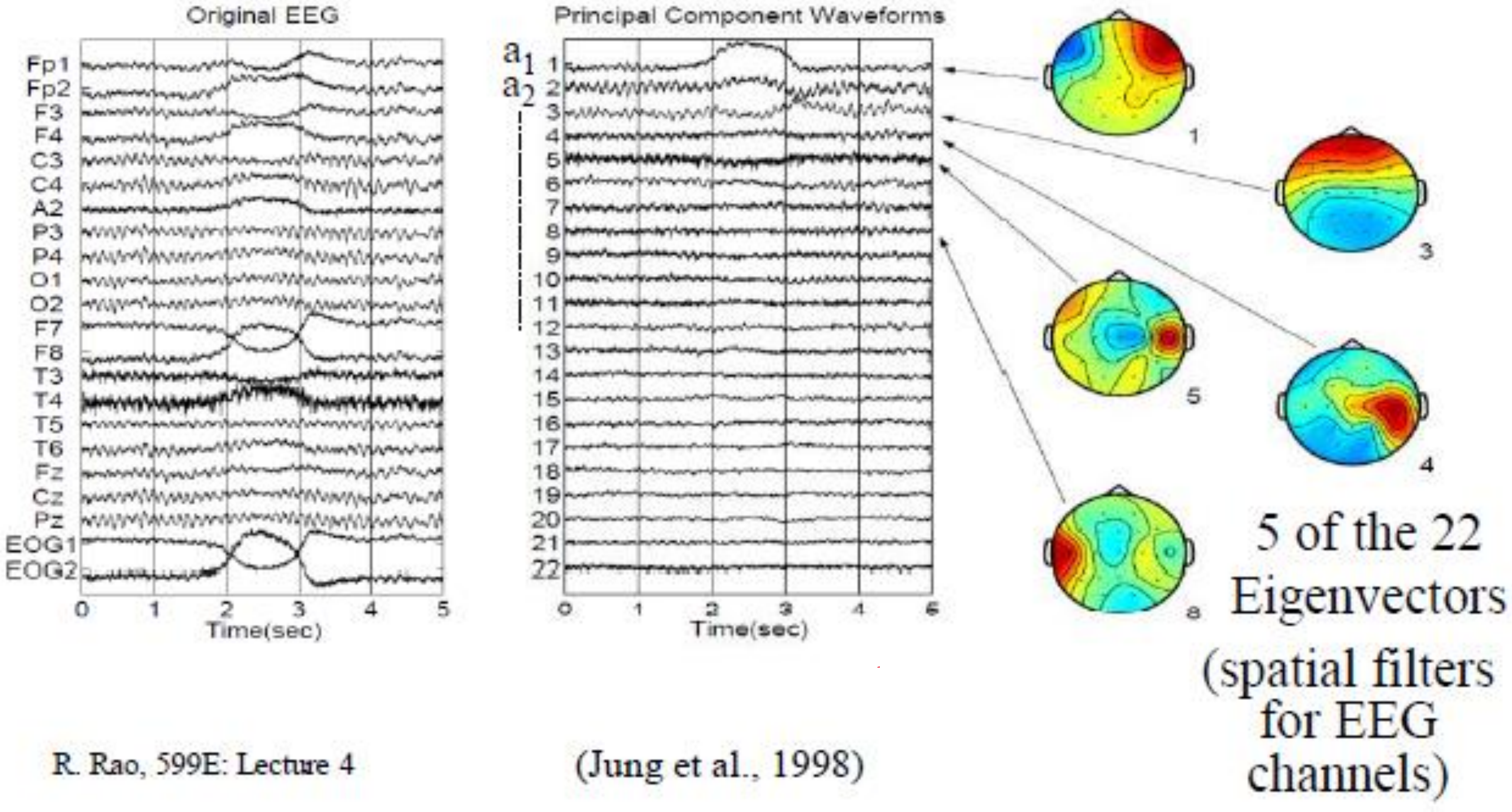
- Data should always be normalized prior to using PCA.

- A common normalization method is to transform all the data to have zero mean and unit standard deviation:

$$\frac{x_i - \mu}{\sigma}$$

where μ and σ are the mean and standard deviation of the i-th feature x_i

PCA applied to EEG



R. Rao, 599E: Lecture 4

(Jung et al., 1998)

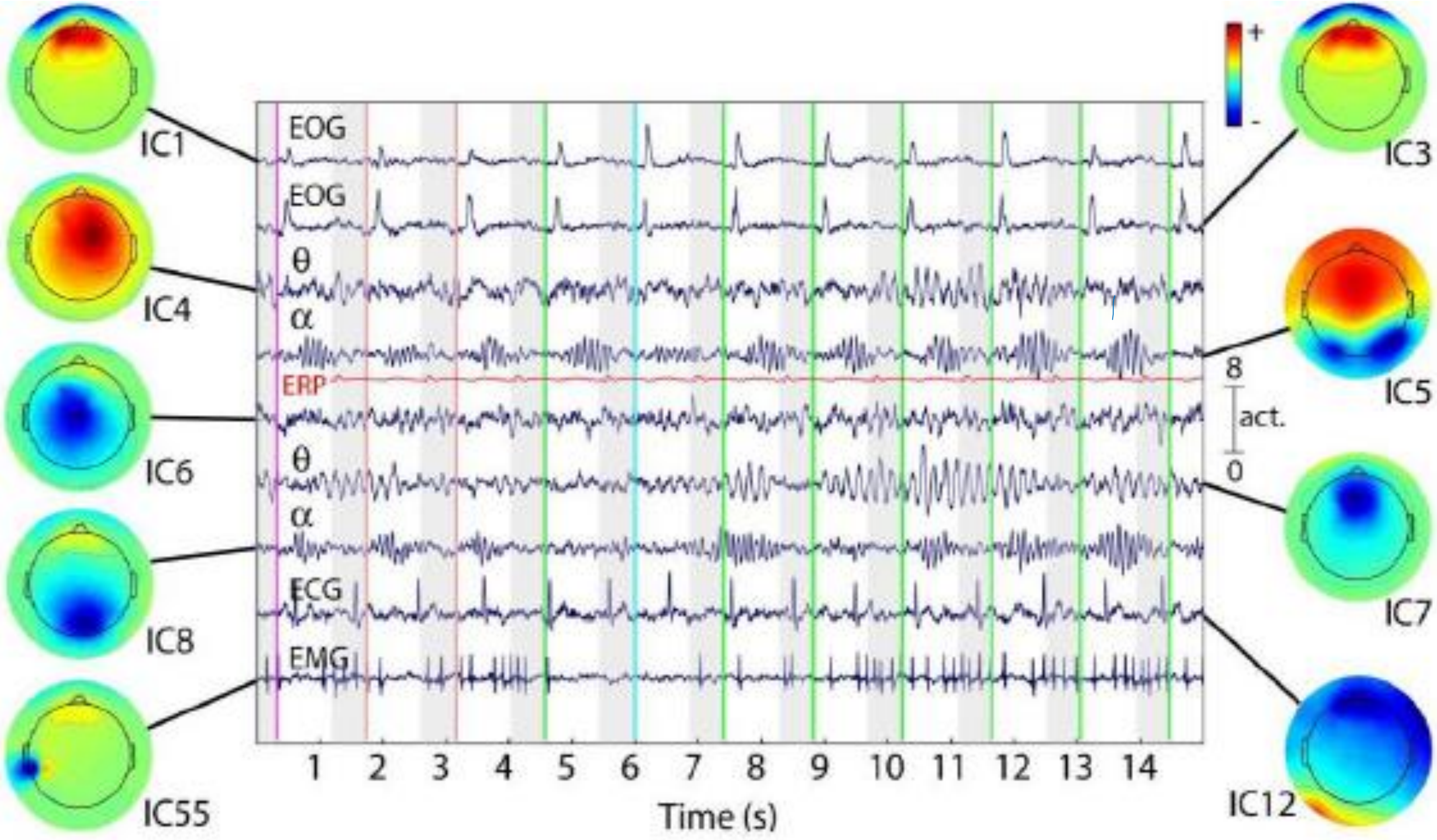
Independent Component Analysis

- PCA finds a matrix **V** that decorrelates the inputs but the resulting feature vector ***a*** may still retain higher order statistical dependencies
- There may be a possibility that the variables are independent.
- ICA tries to find a matrix W of filters (columns of W) such that the output **a** is **statistically independent**:

$$\boldsymbol{a} = \boldsymbol{W}^{\boldsymbol{T}}\boldsymbol{x} \text{ such that } \boldsymbol{P}(\boldsymbol{a}) \approx \prod_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{D}} \boldsymbol{P}(\boldsymbol{a}_{\boldsymbol{i}})$$

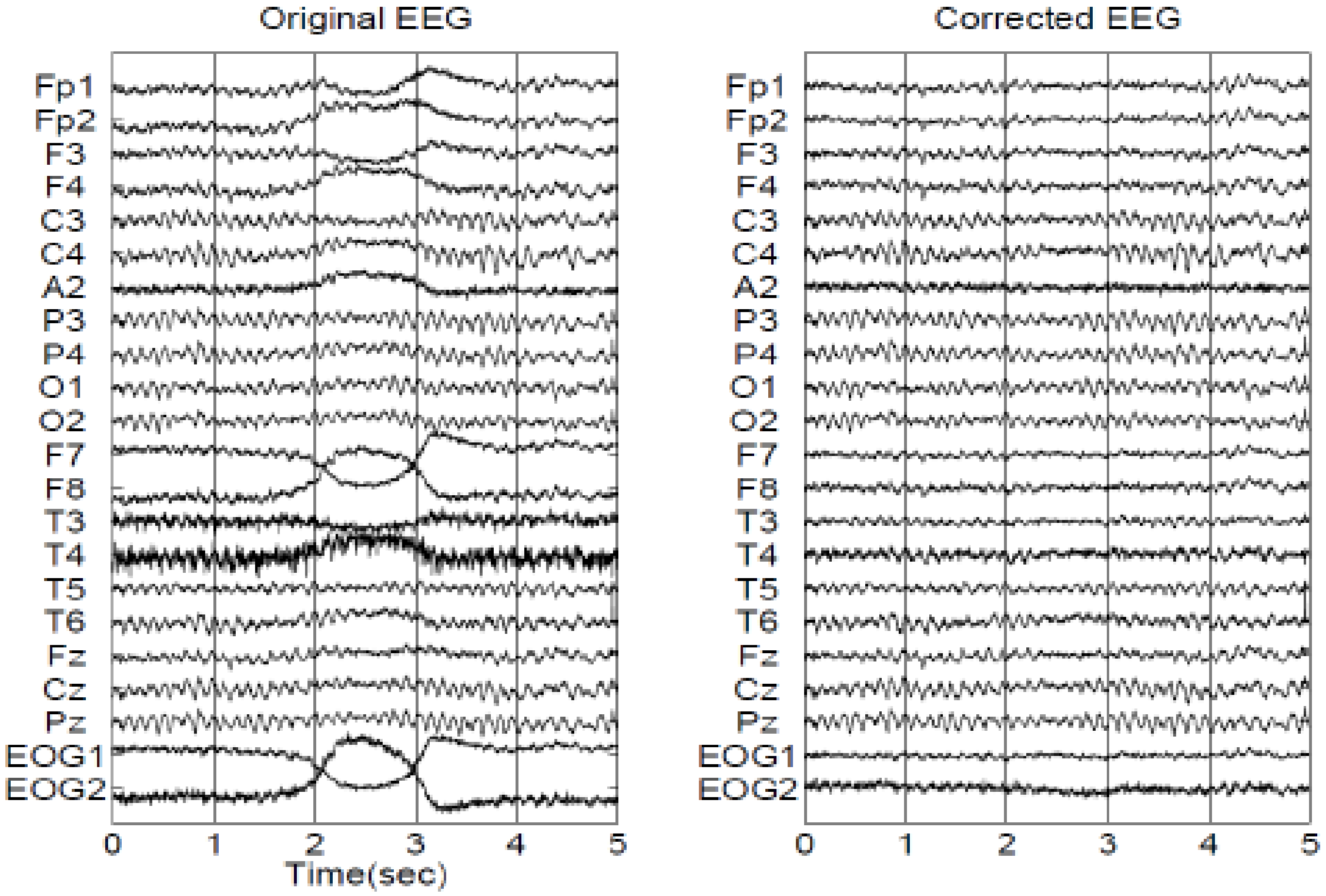
Independent Component Analysis

- ICA assumes sources are linearly mixed to produce x
- The feature vector dimension in ICA can be lesser than, equal to, or greater than the number of input dimensions.
- ICA has proved useful in a variety of settings in BCI applications, ranging from the use of the output vector **a** as a feature vector in classification.



Application of ICA to EEG data for isolating electrooculographic (EOG) (eye-related), electromyographic (EMG) (muscle-related) and electrocardiographic (ECG) (heart-related) artifacts, and unmixing putative source signals in the brain. Image (adapted from Onton and Makeig, 2006)

ICA for Artifact Removal in EEG



Common Spatial Pattern

- Supervised Technique
- Data is labeled with class to which each data vector belongs
  - E.g., EEG obtained for right versus left hand imagery
- CSP finds a matrix of spatial filters
  - the variance of the filtered data for one class is maximized
  - variance of the filtered data for the other class is minimized
- CSP filters can significantly enhance discrimination ability between the two classes

CSP applied to EEG for Right/Left Hand Imagery

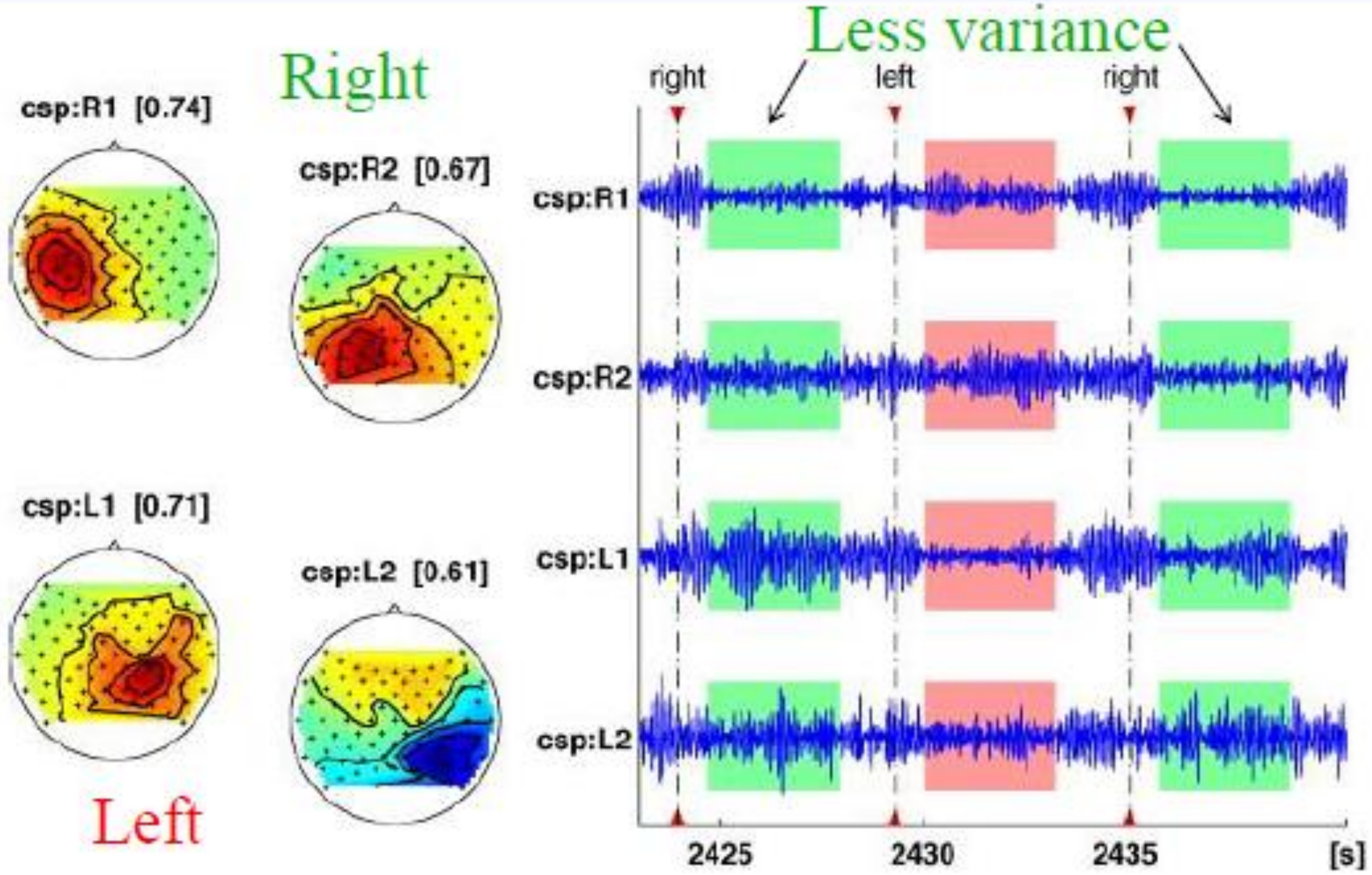
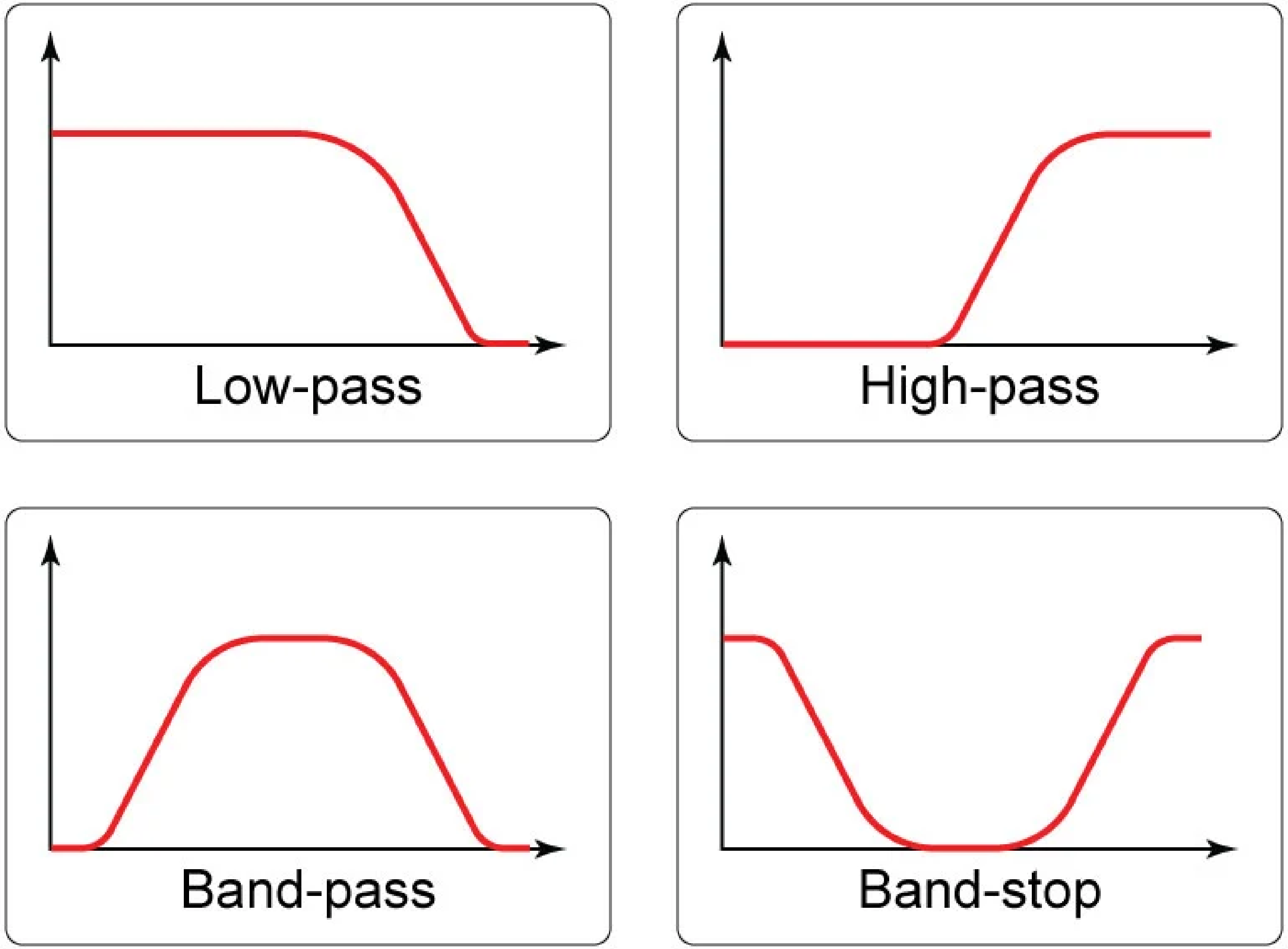


Image source: http://www.sciencedirect.com/science/article/pii/S0165027007004657

Artifact Reduction Techniques

- Thresholding
  - If the magnitude or some other characteristic of a recorded EOG or EMG signal exceeds a pre-determined threshold, the brain signals recorded during that epoch are deemed to be contaminated and rejected.
- Band-Stop and Notch Filtering
  - Band-stop filtering is a useful artifact reduction technique that attenuates the components of a signal in a specific frequency band and passes the rest of the components of the signal.
  - A notch filter set to the 59–61 Hz band (in the United States) for filtering out the 60 Hz power-line noise artifact.


Low-pass
High-pass
Band-pass
Band-stop

Artifact Reduction Techniques

- Linear Modeling
  - A simple way of modeling the effect of artifacts on a recorded brain signal is to assume that the effect is additive.
  - For example, if $EEG_i(t)$ is the EEG signal recorded from electrode i at time t, then a model of how the signal has been contaminated could be:

$$EEG_i(t) = EEG_i^{true}(t) + K \cdot EOG(t)$$

  - $EEG_i^{true}(t)$ is the uncontaminated ("true") EEG signal from electrode i at time t, $EOG(t)$ is the recorded EOG signal at time t and K is a constant.
  - Given an estimated value for K, one can obtain an estimate of the true EEG signal using:

$$EEG_i^{true}(t) = EEG_i(t) - K \cdot EOG(t)$$