# Acquisition and Analysis of Neuronal Data 2009
# BCI – Lecture #01


Benjamin Blankertz

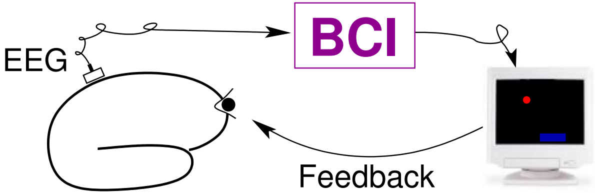
Machine Learning Laboratory, Berlin Institute of Technology

Fraunhofer FIRST (IDA)

blanker@cs.tu-berlin.de


12-Jun-2009

# Brain-Computer Interaction




**BCI:** Translation of human intentions as derived from brain signals into a technical control signal

**Motivation:**

- tools for paralyzed patients or handicaped people: mental typewriter, BCI-controlled prosthesis/wheelchair
- Novel neuroscientific tool to investigate the behaving brain
- Additional communication channel in human-computer interaction

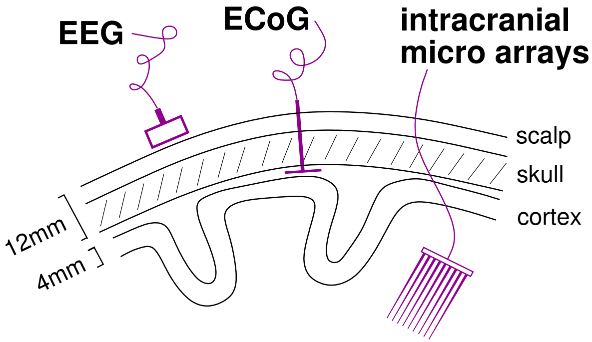
# Videos from Feedback Sessions

Using the Machine Learning approach most users can control BBCI applications in their very first session.

- **Cursor Control (1d)**
- **Brain-Pong (one player)** **Replay (two players)**
- **Pinball (Adam's Family)**
- **Mental Typewriter** **Replay**
- **Mental State Monitoring (here cognitive workload)**

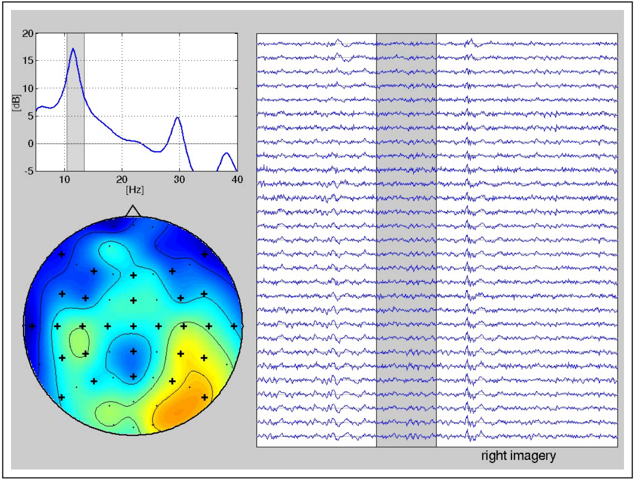
# Ways to Measure Brain Activity

Electrical brain activity in BCIs is acquired in different ways:

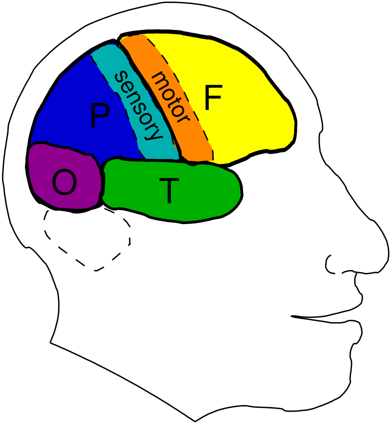


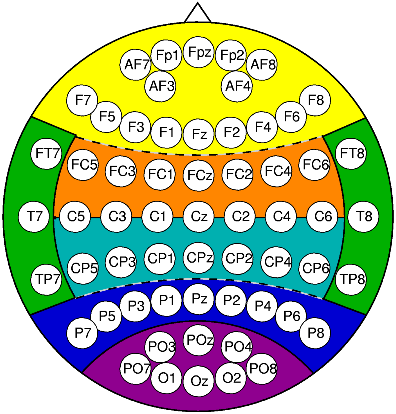

- EEG: electroencephalogram
- ECoG: electrocorticogram

# A Look at EEG Signals



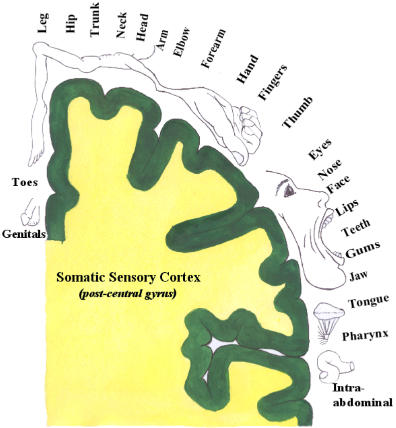
# Areas of the Brain

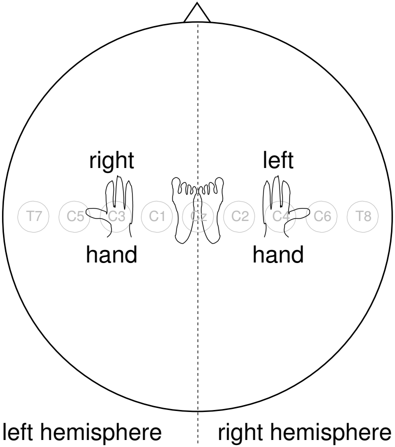


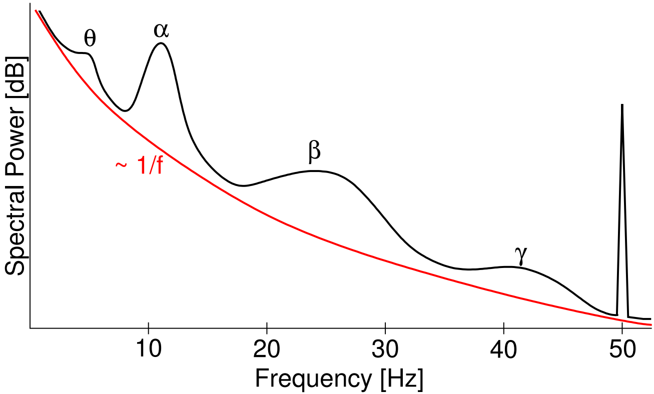

Brain lobes: **F**rontal, **P**arietal, **T**emporal, **O**ccipital.

# Topographic Mapping in Somatic Sensory Area



# Spectrum of Macroscopic Brain Activity

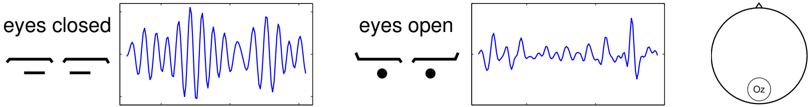



- The figure shows an idealized spectrum.
- Some brain rhythms are named according to their origin, e.g., $\mu$, $\sigma$, $\tau$.

# Modulation of Brain Rhythms

Most rhythms are idle rhythms, i.e., they are **attenuated** during activation.

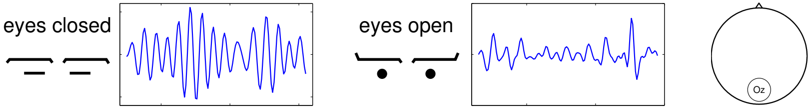
- $\alpha$-rhythm (around 10 Hz) in visual cortex:



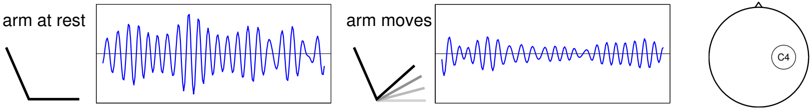
# Modulation of Brain Rhythms

Most rhythms are idle rhythms, i.e., they are **attenuated** during activation.

- $\alpha$-rhythm (around 10 Hz) in visual cortex:

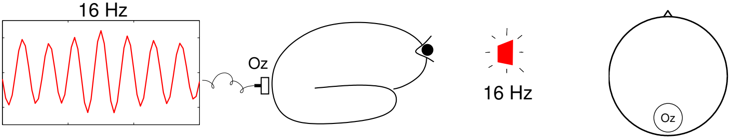



- $\mu$-rhythm (around 10 Hz) in motor and sensory cortex:



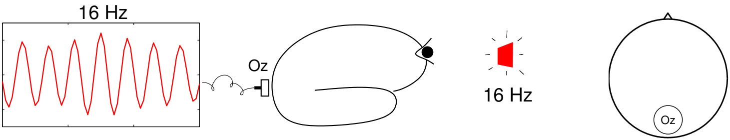
# Evoked Brain Activity

Visual stimuli of constant blinking frequency elicits Steady-State Visual Evoked Potentials (SSVEP) in visual cortex *if focused*:

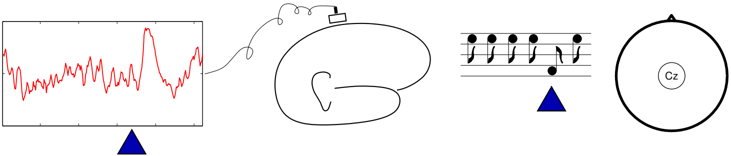


# Evoked Brain Activity

Visual stimuli of constant blinking frequency elicits Steady-State Visual Evoked Potentials (SSVEP) in visual cortex *if focused*:

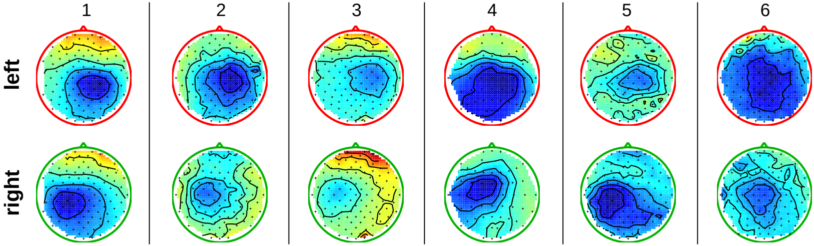



An infrequent stimulus in a series of standard stimuli evokes a P300 component at central scalp position *if attended*:



# Subject-to-Subject Variability

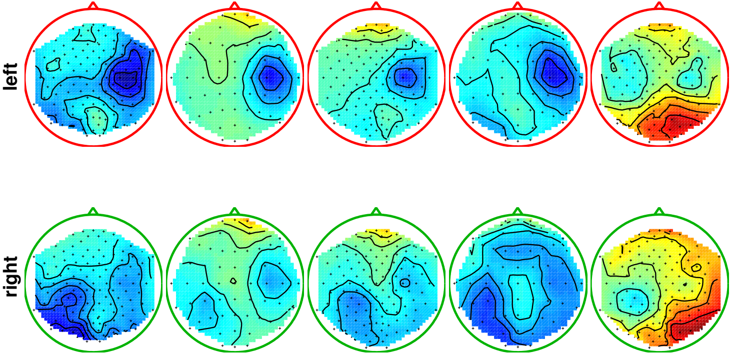
- Experiment: 6 subjects performed left vs. right hand finger tapping.
- Even though the task involves a highly **overlearned motor competence**, the **averaged** brain patterns exhibit a great diversity between subjects:




➤ An optimal system needs adaption for each user.

# Session-to-Session Variability

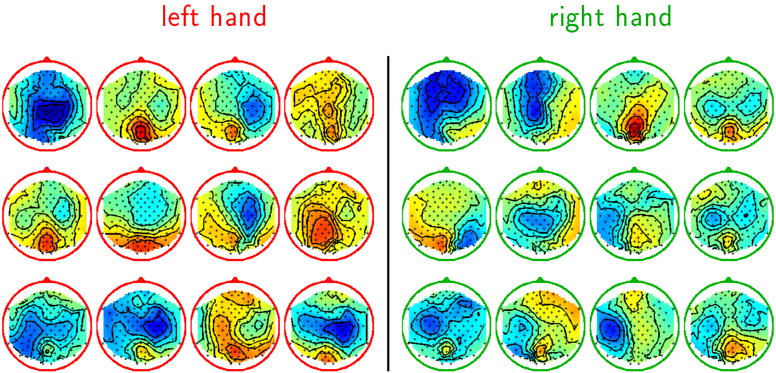
- Experiment: **One subject** imagined left vs. right hand movements on different days.
- Even though each ERD map represents an **average** across 140 trials, they exhibit an apparent diversity.



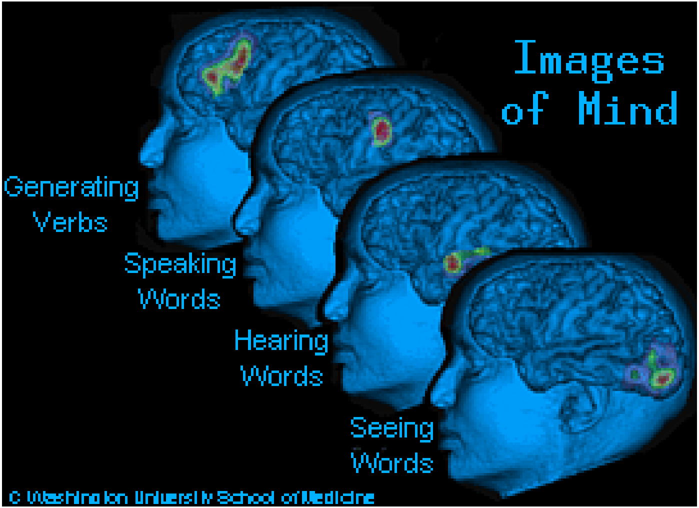

➤ An optimal system needs adaption for (or within?) each session.

## Trial-to-Trial Variability

- Experiment: One subject imagined left vs. right hand movements.
- Topographies show power in the **alpha band** during trials of 3.5 s.
- They exhibit an extreme diversity, although recorded from **one subject** on **one day**.

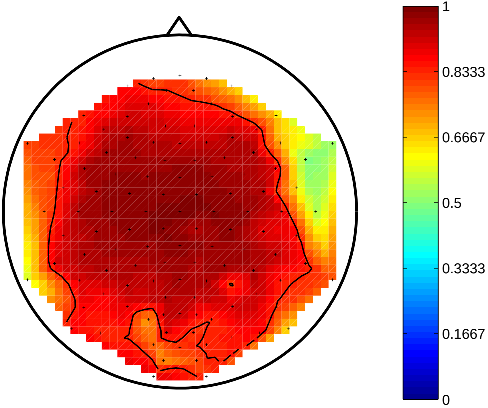


# Large Number of Simultaneously Active Areas



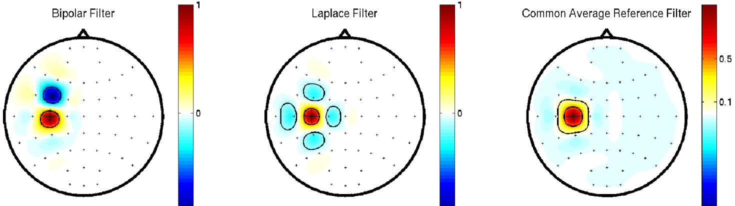
# Spatial Smearing

- Raw EEG scalp potentials are known to be associated with a large spatial scale owing to volumne conduction.
- In a simulation of Nunez et al [1] only half the contribution to one scalp electrode comes from sources within a 3 cm radius.

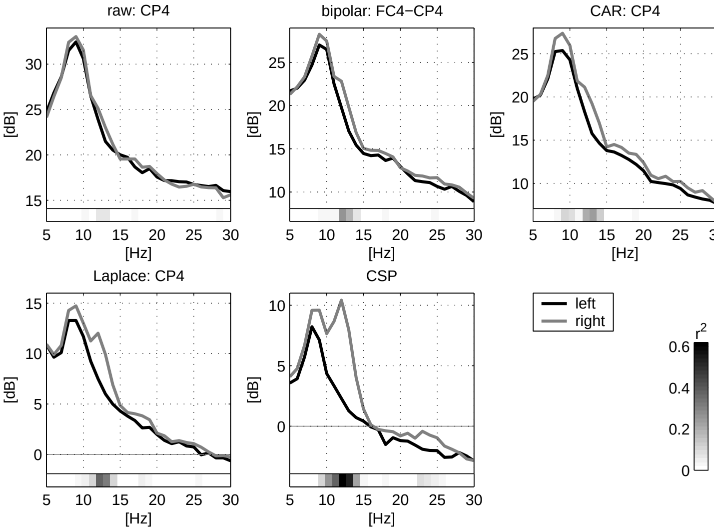


# Some Spatial Filters

- **Bipolar:** Subtract values from two electrode positions, e.g.: $\text{Bip}_{\text{C3,FC3}} = \text{C3} - \text{FC3}$
- **Common Average Reference (CAR):** Subtract the average of all EEG electrodes ($\mathcal{C} = \{\text{F3,Fz,F4,C3,Cz,C4}, \dots\}$) from the given electrode: $\text{C3}_{\text{CAR}} = \text{C3} - \frac{1}{|\mathcal{C}|} \sum_{C \in \mathcal{C}} C$
- **Laplace (Lap):** Subtract from each channel the average of its immediate neighbours: $\text{C3}_{\text{Lap}} = \text{C3} - {}^1\!/_4(FC3 + C1 + CP3 + C5)$
- **Common Spatial Patterns (CSP):** A data-driven method that can be used to find optimized filters that reflect amplitude modulations of brain rythms. See also **PCA**, **ICA**.



# The Need for Spatial Filtering



# Technical Note

## Conventions in Notation

- $x, N, \gamma$ (italics): scalar
- $\mathbf{x}, \boldsymbol{\mu}$ (bold face, lower case): column vector
- $\mathbf{X}, \boldsymbol{\Sigma}$ (bold face, upper case): matrix
- For a finite set $\mathcal{C}$, we denote its cardinality (number of elements) by $|\mathcal{C}|$.

## Some math

If $x_1, \ldots, x_k$ are iid Gaussian distributed according to $\mathcal{N}(\mu_k, \sigma_k^2)$, the following holds:

- $\alpha x_1 \sim \mathcal{N}(\alpha\mu_1, \alpha^2\sigma_1^2)$
- $\sum_{k=1}^{K} x_k \sim \mathcal{N}(\sum_{k=1}^{K} \mu_k, \sum_{k=1}^{K} \sigma_k^2)$

# Today's Topic

Investigation of **Event-Related Potentials (ERPs)**.
ERPs are brain responses that are time-locked to some *event*. The event may be an external sensory stimulus or internal, associated with the execution of a motor, cognitive, or psychophysiologic task.

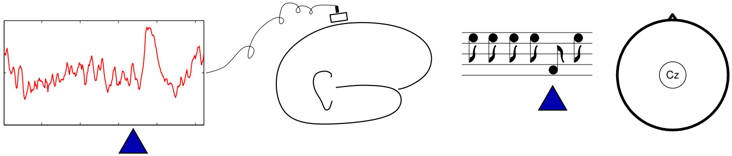
A subclass are the **evoked potentials (EPs)** which reflect the processing of the physical stimulus, rather than 'higher' processes, that might involve memory, expectation, or attention.

# Oddball Experiments

In the standard oddball paradigm [2], a series of stimuli is presented to the subjects. There are two types of stimuli

- *standard:* frequent stimuli
- *deviant:* infrequent stimuli (typically 10–20%)

During the experiment, the subject is required to *attend* the deviant stimuli, e.g., by counting them, or by giving a motor response. The most prominent ERP in an oddball paradigm is the P300 component, a positive reflection at central/parietal position roughly 300 ms after stimulus.



# BCI using the P300 Component


In the Donchin setup [3] the subject concentrates on a letter of a 6×6 symbol matrix. Rows and columns are highlighted several times in random order.

P300 (and other) components are most strongly elicited when the row resp. column is flashed which contains the selected letter.

**Video P300 Speller**

# Averaging across Trials

In raw EEG the signal-to-noise ratio is bad. The common technique to investigate a signal of interest is to record a long series of trials and then to average across trials.

One can consider each single trial $x_k(t)$ as composed of a (phase-locked) event-related source $s(t)$ (ERP) and 'noise' $n_k(t)$:

$$x_k(t) = s(t) + n_k(t) \qquad \text{for } k = 1, \ldots, K$$

Here $n_k(t)$ can be assumed to be iid $\mathcal{N}(0, \sigma_n^2)$ distributed (for a fixed $t$).

# Averaging across Trials

In raw EEG the signal-to-noise ratio is bad. The common technique to investigate a signal of interest is to record a long series of trials and then to average across trials.

One can consider each single trial $x_k(t)$ as composed of a (phase-locked) event-related source $s(t)$ (ERP) and 'noise' $n_k(t)$:

$$x_k(t) = s(t) + n_k(t) \qquad \text{for } k = 1, \dots, K$$

Here $n_k(t)$ can be assumed to be iid $\mathcal{N}(0, \sigma_n^2)$ distributed (for a fixed $t$).

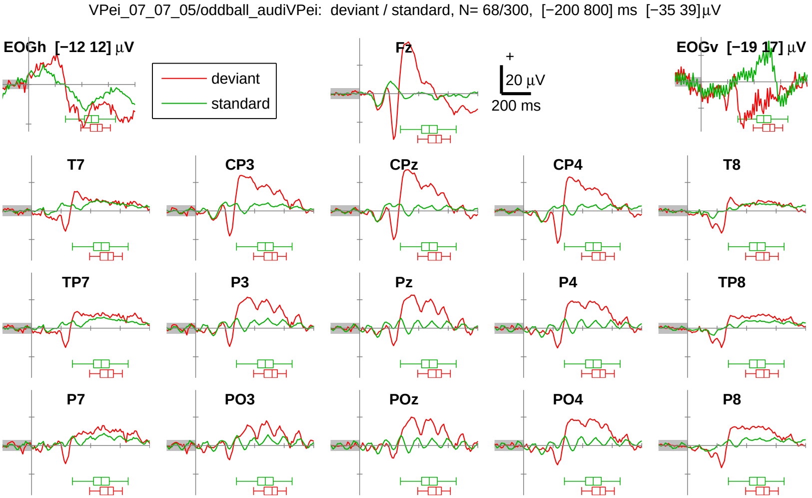
Then averaging across trials lead to

$$\frac{1}{K}\sum_{k=1}^{K} x_k(t) = s(t) + \frac{1}{K}\sum_{k=1}^{K} n_k(t)$$

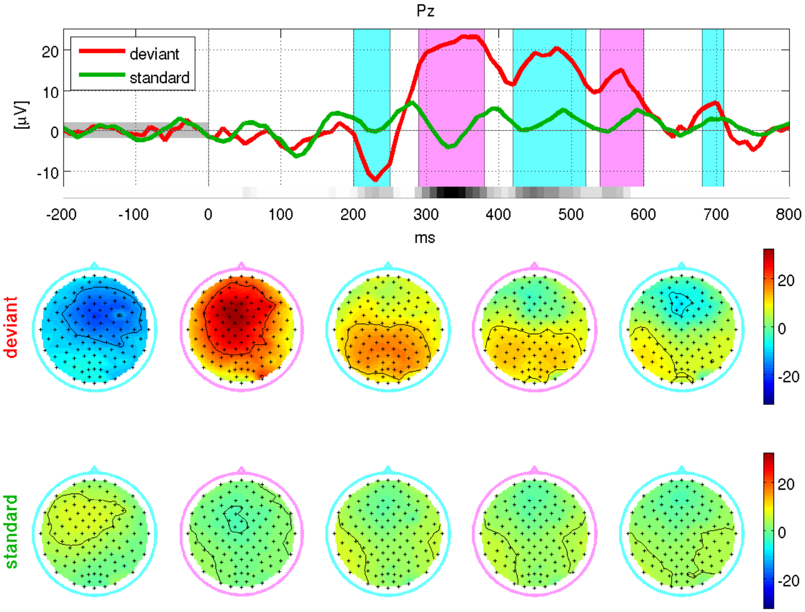
The noise in the average is $n(t) := \frac{1}{K}\sum_{k=1}^{K} n_k(t) \sim \mathcal{N}(0, \frac{1}{K}\sigma_n^2)$.

The amplitude of the noise goes down by a factor of $\sqrt{K}$ in an average across $K$ trials.

# Data of an Oddball Experiment



# Data of an Oddball Experiment



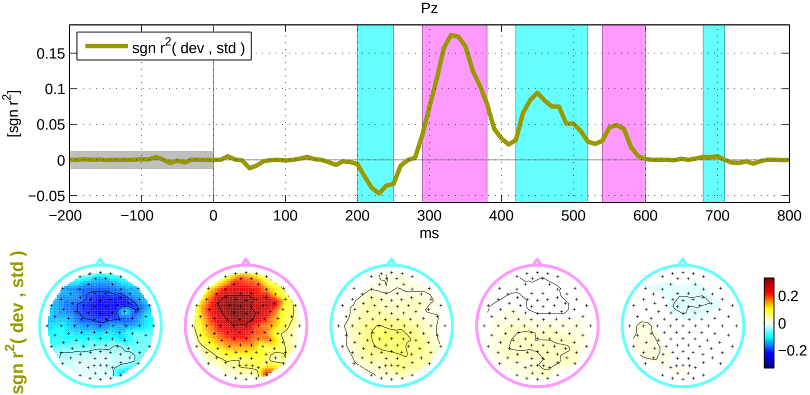
# Bi-serial Correlation Coefficient

One statistical way to measure how much information one feature $x_i$ carries about the labels $y_i \in \{1, 2\}$ is the bi-serial correlation coefficient $r$:

$$r(x) := \frac{\sqrt{N_1 \cdot N_2}}{N_1 + N_2} \, \frac{\mathrm{mean}\{x_i \,|\, y_i = 1\} - \mathrm{mean}\{x_i \,|\, y_i = 2\}}{\mathrm{std}\{x_i\}}$$

or the $r^2$-coefficient $r^2(x) := r(x)^2$, which reflects how much of the variance in the distribution of all samples is explained by the class affiliation.
The $r^2$-value is often used since it gives clearer topographical maps.
To retain the sign of the difference it can be multiplied by the sign of the $r$-value (sgn $r^2$).

# sgn $r^2$-Difference Deviant minus Standard



# Classification of ERP Data: Spatially

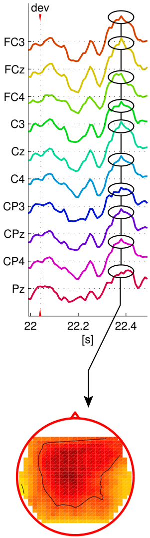
For the classification of ERP data, *feature vectors* (or features) need to be calculated from each single trial.

For the determination of *spatial features*, for each single trial the average within a specified interval is calculated.

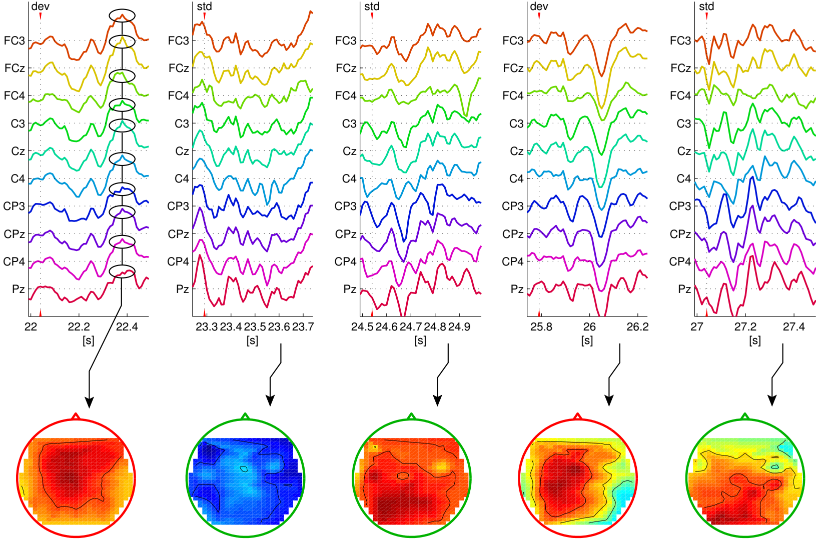
The selection of the interval can be performed by the experimenter using visualization of the data and statistical measures, e.g. $r^2$-values or by automatic procedures (not discussed here).

The performance of classification is typically measured by cross validation.

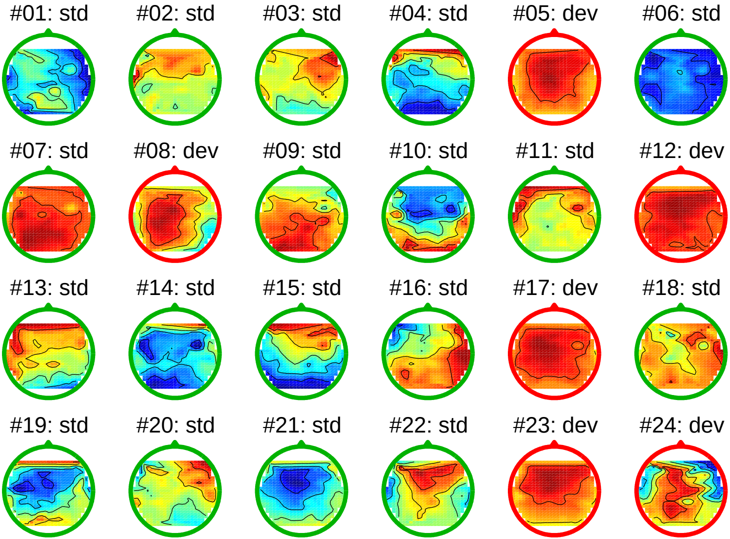
# Extraction of Spatial Features



# Extraction of Spatial Features

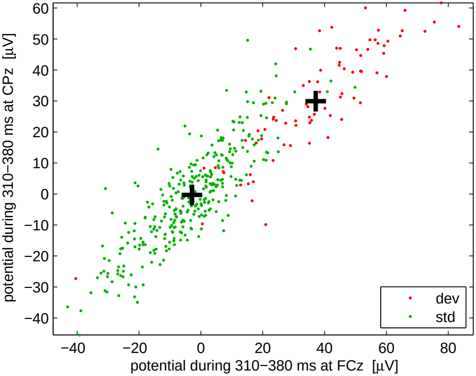


# Spatial Features



# Feature Distribution for Interval 310–380 ms

Distributions of features can only be easily visualized for 2 dimensions. Here we take two channels from the spatial features:



# Fisher's Discriminant Analysis

Let $\mathbf{x}_k$ be feature vectors of two conditions ($k$ in $\mathcal{C}_1$ resp. $\mathcal{C}_2$) and define

$$\mu_i = \frac{1}{|\mathcal{C}_i|} \sum_{k \in \mathcal{C}_i} \mathbf{x}_k,$$

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^\top, \quad S_i = \sum_{k \in \mathcal{C}_i} (\mathbf{x}_k - \mu_i)(\mathbf{x}_k - \mu_i)^\top$$

Fisher's Discriminant is that $\mathbf{w}$ which maximizes the Rayleigh quotient

$$J(\mathbf{w}) = \frac{\mathbf{w}^\top S_B \mathbf{w}}{\mathbf{w}^\top (S_1 + S_2) \mathbf{w}} \approx \frac{(\mathbf{w}^\top \mu_1 - \mathbf{w}^\top \mu_2)^2}{\mathrm{var}_{t \in \mathcal{C}_1}(\mathbf{w}^\top x(t)) + \mathrm{var}_{t \in \mathcal{C}_2}(\mathbf{w}^\top x(t))}$$

# Fisher's Discriminant Analysis

Let $\mathbf{x}_k$ be feature vectors of two conditions ($k$ in $\mathcal{C}_1$ resp. $\mathcal{C}_2$) and define

$$\mu_i = \frac{1}{|\mathcal{C}_i|} \sum_{k \in \mathcal{C}_i} \mathbf{x}_k,$$

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^\top, \quad S_i = \sum_{k \in \mathcal{C}_i} (\mathbf{x}_k - \mu_i)(\mathbf{x}_k - \mu_i)^\top$$

Fisher's Discriminant is that $\mathbf{w}$ which maximizes the Rayleigh quotient

$$J(\mathbf{w}) = \frac{\mathbf{w}^\top S_B \mathbf{w}}{\mathbf{w}^\top (S_1 + S_2)\mathbf{w}} \approx \frac{(\mathbf{w}^\top \mu_1 - \mathbf{w}^\top \mu_2)^2}{\mathrm{var}_{t \in \mathcal{C}_1}(\mathbf{w}^\top x(t)) + \mathrm{var}_{t \in \mathcal{C}_2}(\mathbf{w}^\top x(t))}$$

The maximum is attained for the generalized eigenvector of the matrices $S_B$ and $(S_1 + S_2)$ with maximum eigenvalue. It can be calculated in a simpler way [4]:

$$\mathbf{w} = (S_1 + S_2)^{-1}(\mu_1 - \mu_2)$$

# FD for Classification

Let $\mathbf{x}_k \in \mathbb{R}^m$ be feature vectors of two classes ($k \in \mathcal{C}_1$ resp. $k \in \mathcal{C}_2$). Then the FD vector $\mathbf{w}$ as defined above separates $\mathbb{R}^m$ in two classes by virtue of the decision function:

$$f : \mathbb{R}^m \to \mathbb{R}; \quad \mathbf{z} \mapsto \begin{cases} -1 & \text{if } \mathbf{w}^\top \mathbf{z} + b < 0 \\ 1 & \text{else} \end{cases}$$

The bias can, e.g., be chosen as $b = -\mathbf{w}^\top(\mu_1 + \mu_2)/2$.

# FD for Classification

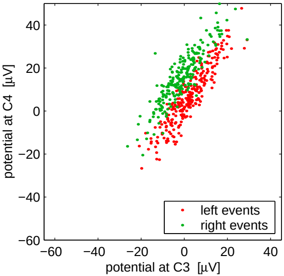
Let $\mathbf{x}_k \in \mathbb{R}^m$ be feature vectors of two classes ($k \in \mathcal{C}_1$ resp. $k \in \mathcal{C}_2$). Then the FD vector $\mathbf{w}$ as defined above separates $\mathbb{R}^m$ in two classes by virtue of the decision function:

$$f : \mathbb{R}^m \to \mathbb{R}; \quad \mathbf{z} \mapsto \begin{cases} -1 & \text{if } \mathbf{w}^\top \mathbf{z} + b < 0 \\ 1 & \text{else} \end{cases}$$

The bias can, e.g., be chosen as $b = -\mathbf{w}^\top(\mu_1 + \mu_2)/2$.

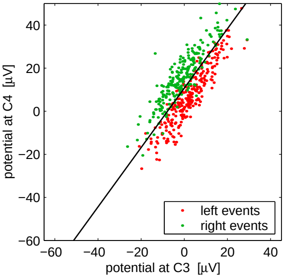
Under the assumption that both classes are distributed according to known Gaussian distributions with equal covariance matrices, the decision function $f$ minimizes the risk of misclassification.
**Note:** FDA is equivalent to Linear Discriminant Analysis.

# FDA Cartoon



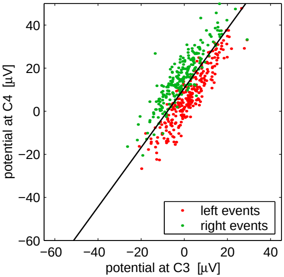
40
20
0
−20
−40
−60
potential at C4 [µV]
−60
−40
−20
0
20
40
potential at C3 [µV]
left events
right events

# FDA Cartoon

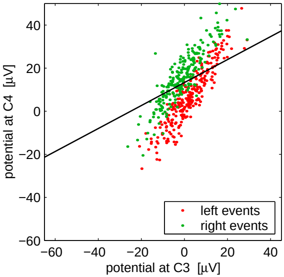


40
20
0
−20
−40
−60
potential at C4 [µV]
−60
−40
−20
0
20
40
potential at C3 [µV]
left events
right events

# FDA Cartoon

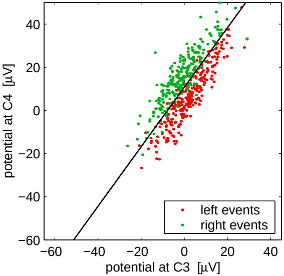


potential at C4 [µV]
potential at C3 [µV]
40
20
0
−20
−40
−60
−60
−40
−20
0
20
40
left events
right events

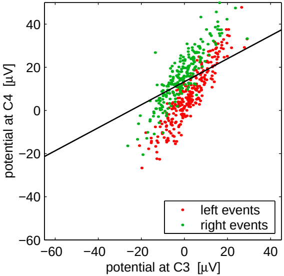



potential at C4 [µV]
potential at C3 [µV]
40
20
0
−20
−40
−60
−60
−40
−20
0
20
40
left events
right events

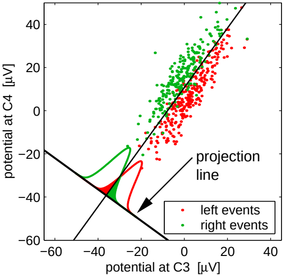
# FDA Cartoon



potential at C4 [µV]
potential at C3 [µV]
left events
right events

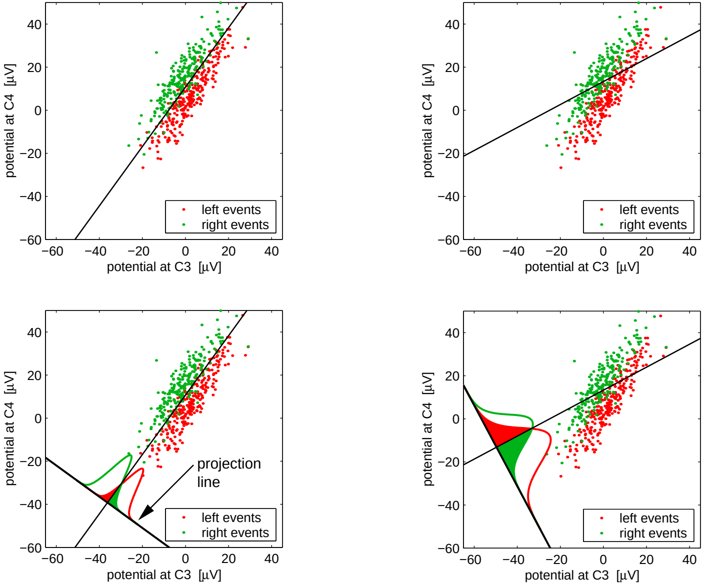



potential at C4 [µV]
potential at C3 [µV]
left events
right events

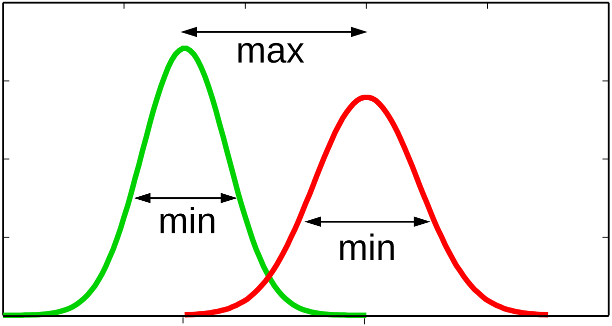



potential at C4 [µV]
potential at C3 [µV]
projection line
left events
right events

# FDA Cartoon


potential at C4 [µV]
potential at C3 [µV]
left events
right events
projection line
40
20
0
−20
−40
−60
−60
−40
−20
0
20
40

# FDA Optimization Criterium



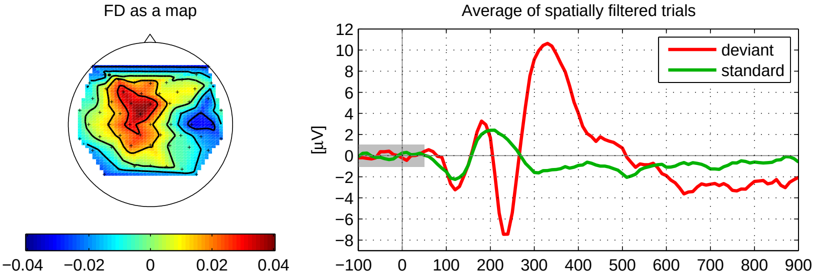
# FD as spatial filter

If the feature vectors $\mathbf{x}_k$ are purely spatial features, i.e. each component is exactly related to one channel, Fisher's Discriminant $\mathbf{w}$ can also be regarded as a spatial filter.
In the oddball data set, let $\mathbf{x}_k$ be the average value in the time interval 300 to 380 ms after simulus of the $k$-th trial, and $\mathcal{C}_1 =$'deviant' , $\mathcal{C}_2 =$'standard'.
Then $\mathbf{x}_k \in \mathbb{R}^{\#\text{chans}}$. Thus, the continuous EEG signals can be spatially filtered with $\mathbf{w}$ which results in one surrogate channel that shows pronounced behaviour for P300 components.
Let $\mathbf{X} \in \mathbb{R}^{\#\text{chans}\times\#\text{time points}}$ be continuous EEG signals. Then

$$\mathbf{X}_f := \mathbf{w}^\top \mathbf{X} \quad \in \mathbb{R}^{1\times\#\text{time points}}$$

is the result of spatial filtering: each channel of $\mathbf{X}$ is weighted with the corresponding component of $\mathbf{w}$ and summed up.
The obtained FD $\mathbf{w} \in \mathbb{R}^{\#\text{chans}}$ can be visualized as a scalp map.

# FD as Spatial Filter – Example



# References

P. L. Nunez, R. Srinivasan, A. F. Westdorp, R. S. Wijesinghe, D. M. Tucker, R. B. Silberstein, and P. J. Cadusch, "EEG coherency I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales", *Electroencephalogr Clin Neurophysiol*, 103(5): 499–515, 1997.

S. Sutton, M. Braren, J. Zubin, and E. John, "Evoked-potential correlates of stimulus uncertainty", *Science*, 150: 1187–1188, 1965.

L. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials", *Electroencephalogr Clin Neurophysiol*, 70: 510–523, 1988.

R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley & Sons, 2nd edn., 2001.