Brain Computer Interfacing

berlin brain computer interface

CHARITÉ CAMPUS BENJAMIN FRANKLIN

Klaus-Robert Müller, Carmen Vidaurre, Matthias Treder, Siamac Fazli, Jan Mehnert, Stefan Haufe, Frank Meinecke, Felix Biessmann, Michael Tangermann, Gabriel Curio, Benjamin Blankertz et al.

BCI MLSSP 2012 Topics

Part I

- Physiology, Signals and Challenges
- Event-Related Desynchronization and BCI

Part II

- Nonstationarity SSA et al.
- Multimodal data

Part III

- Event Related Potentials and BCI
- Applications

Some BCI Groups (not an exhaustive list!) from !2003!

Schwarz, Pittsburg: Invasive

Chapin, Rochester: Invasive

Nicolelis, Duke: Invasive

Kennedy, Atlanta: Invasive

Levine, Michigan: Invasive

Wolpaw, Albany: BCI 2000, 2D, Patients

Donchin, Beckman: P300: Spelling

Anderson, UC, CSU: NN for BCI, invasive

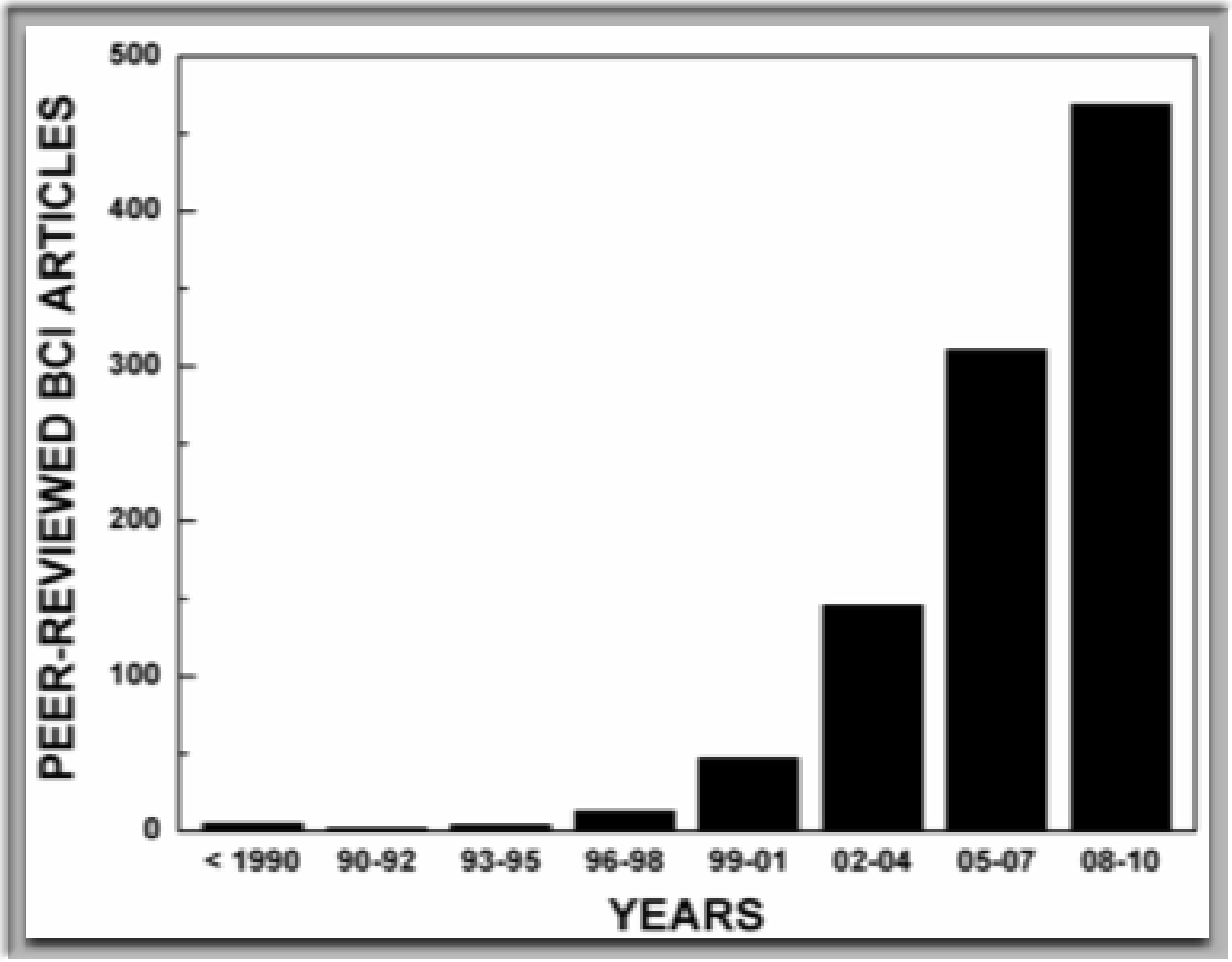
Sadja and Parra, NY: SP, Rapid Visual Stimulation

Birbaumer, Kübler TÜ: SCPs, TTD, Patients

- Pfurtscheller, Graz: ERD, Patients
- Bayliss, Rochester: P300 & VR
- Penny, Roberts, Sykacek, Oxford: Bayes & BCI
- Birch and Mason, UBC BCI
- Moore, Georgia BCI
- Allison, UCSD
- Millan, EPFL: brain states control robot
- Donoghue, Brown U, invasive patient study
- Cuntai Guan, Singapore: P300
- Gao, Beijing: P300
- **BBCI**: Let the machines learn

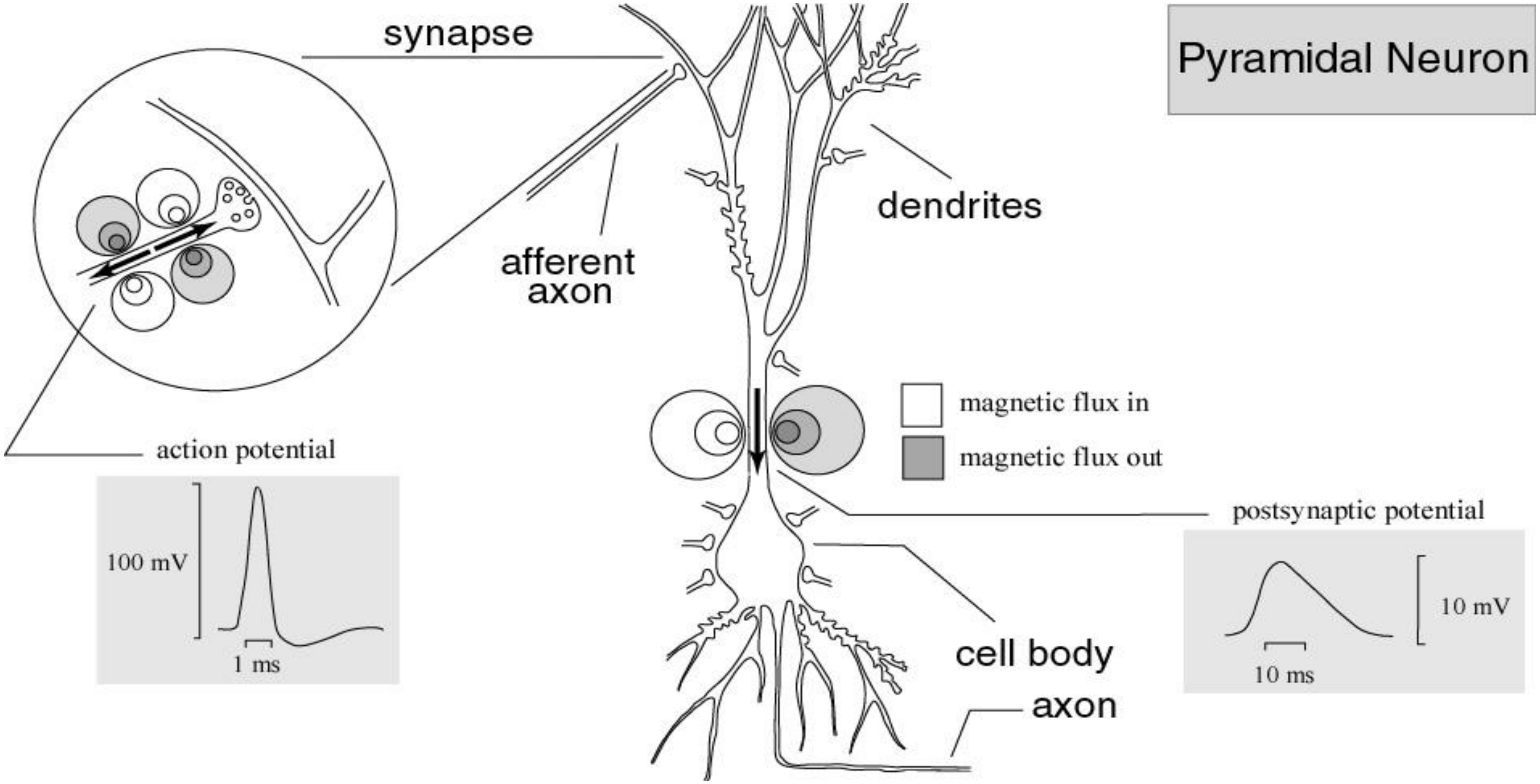
!Note that this is historical slide and MOST groups are missing!

Increasing Interest by Scientists



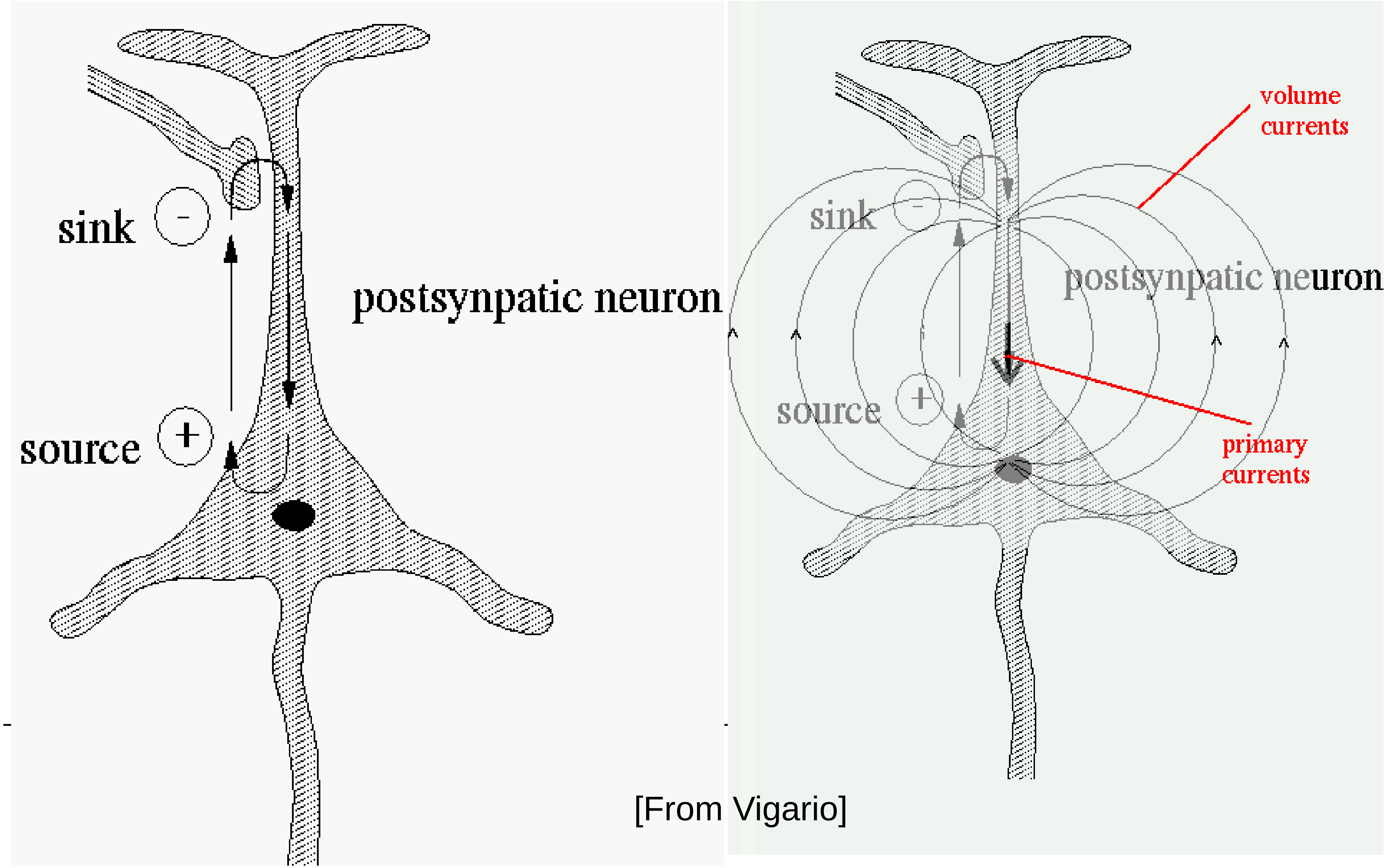
Courtesy of Dr. Jon Wolpaw, Wadsworth Center

The origins of EEG and MEG (short recap.)



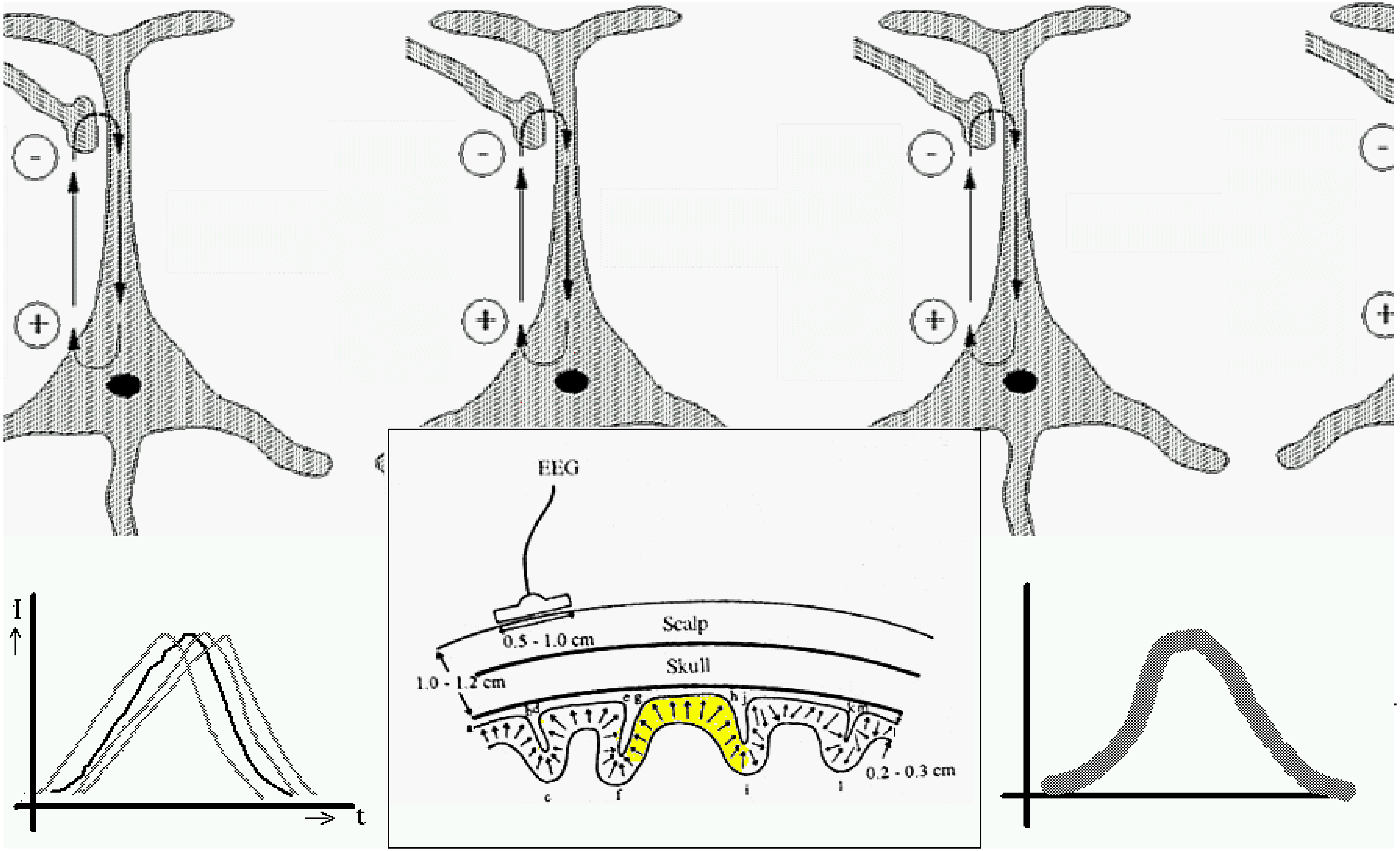
[From Vigario]

From single units to patch of dipoles



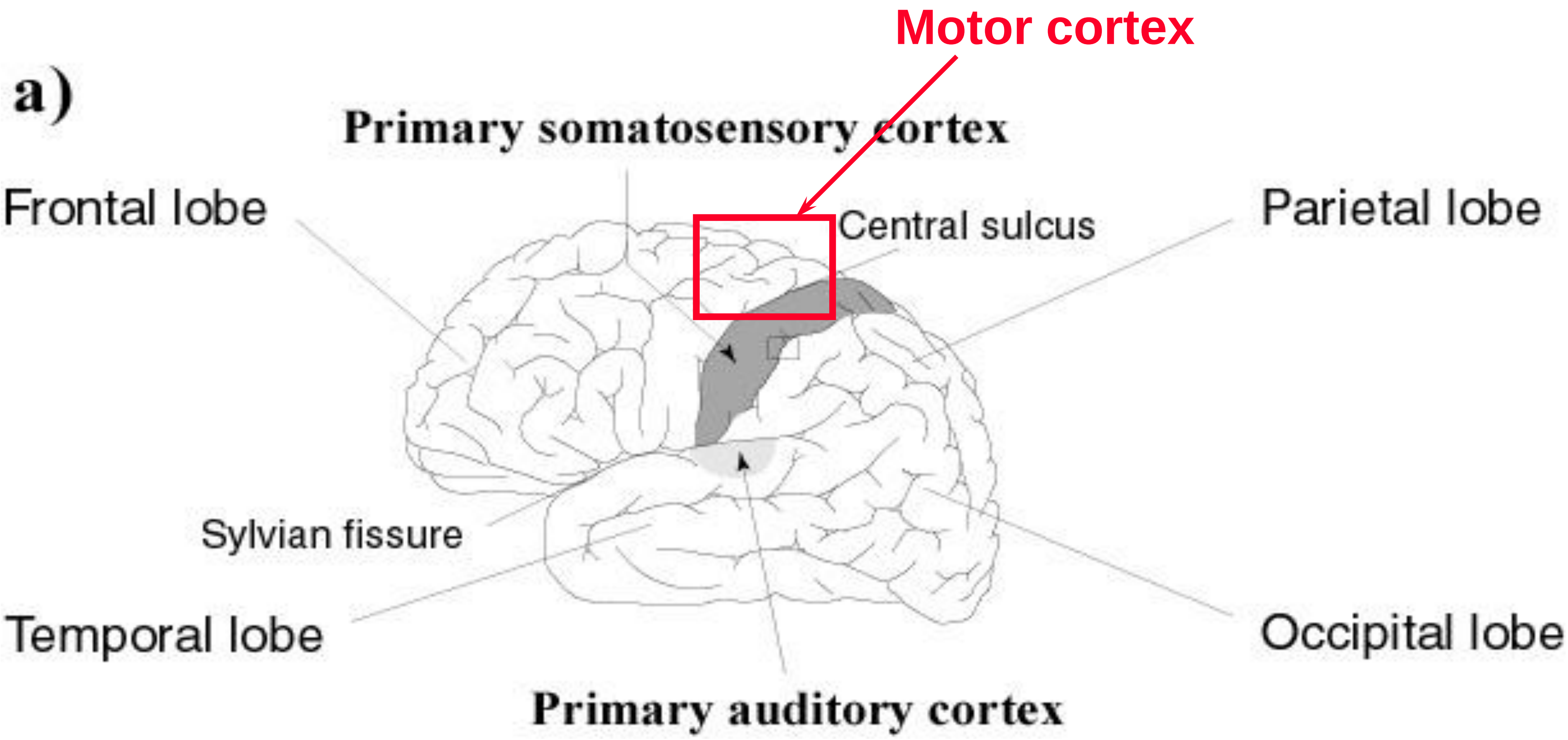
[From Vigario]

From single units to patch of dipoles (cont.)



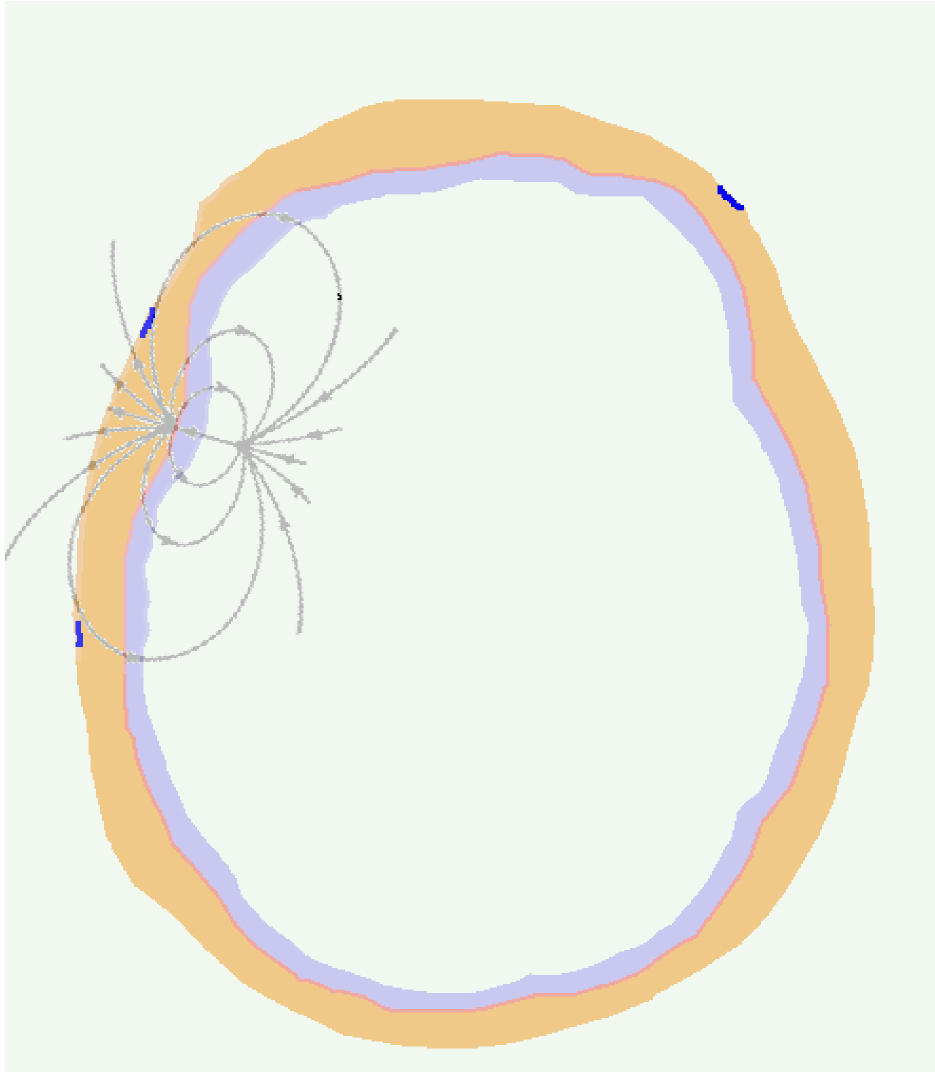
[From Vigario]

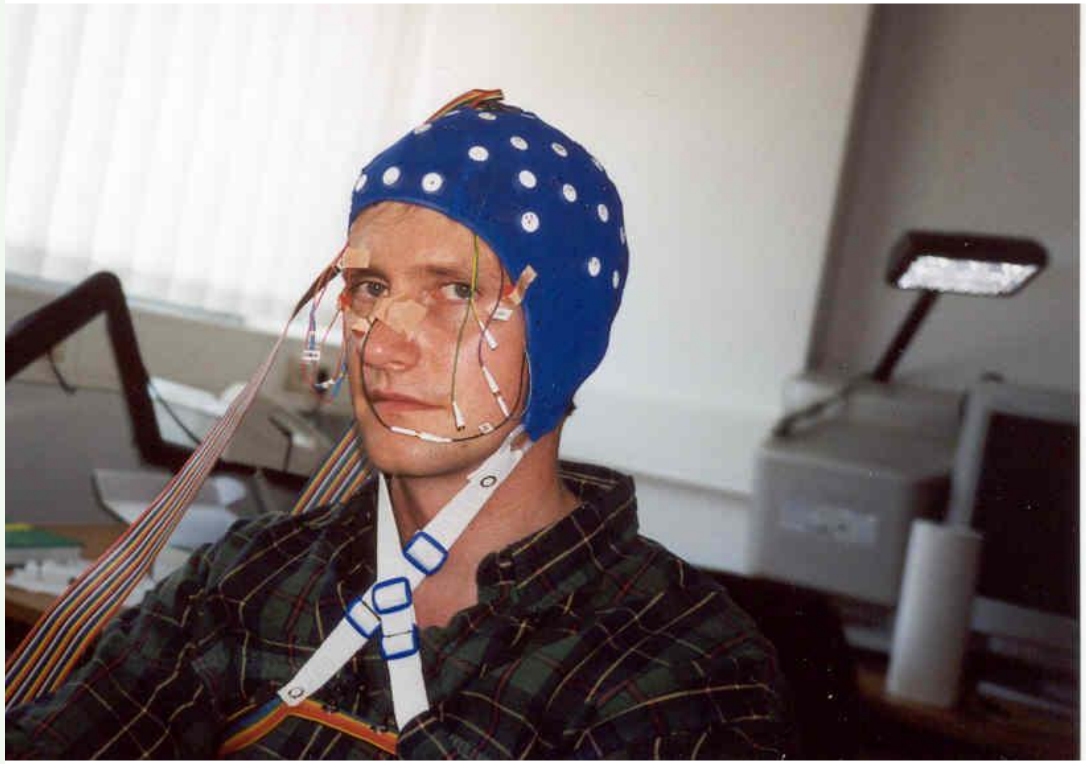
A glance at the cerebrum



[From Vigario]

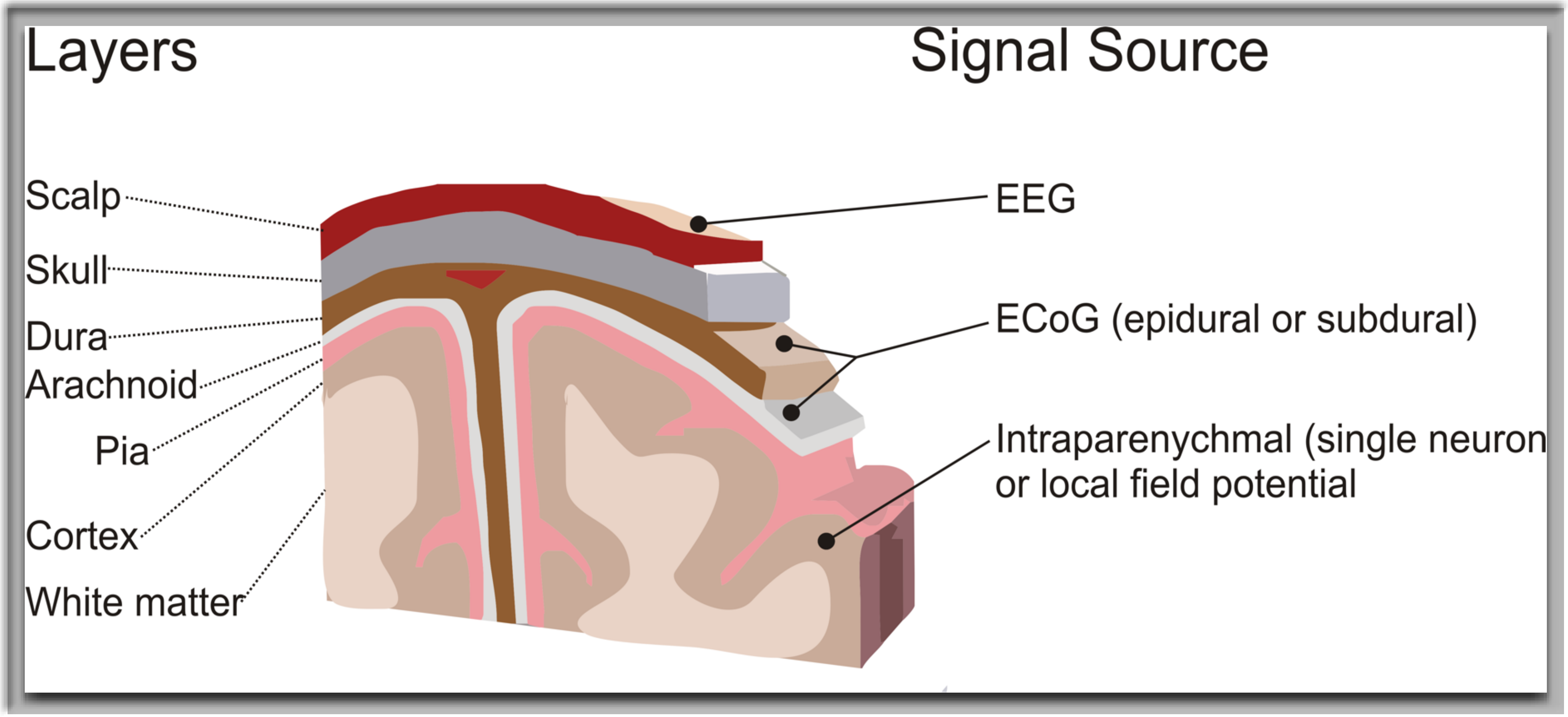
From dipole patches to EEG





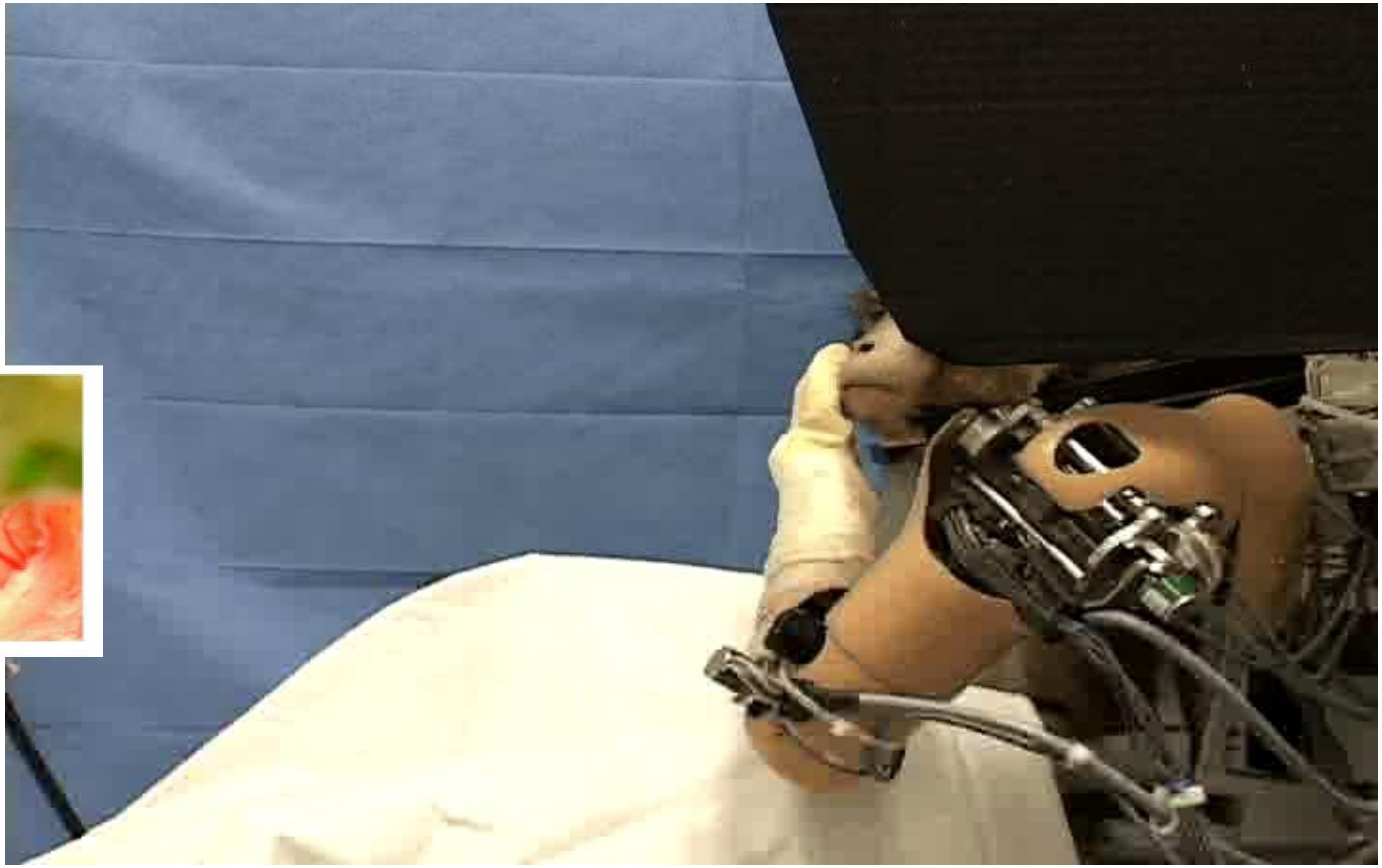
[From Vigario]

Invasive vs noninvasive Brain Computer Interfacing



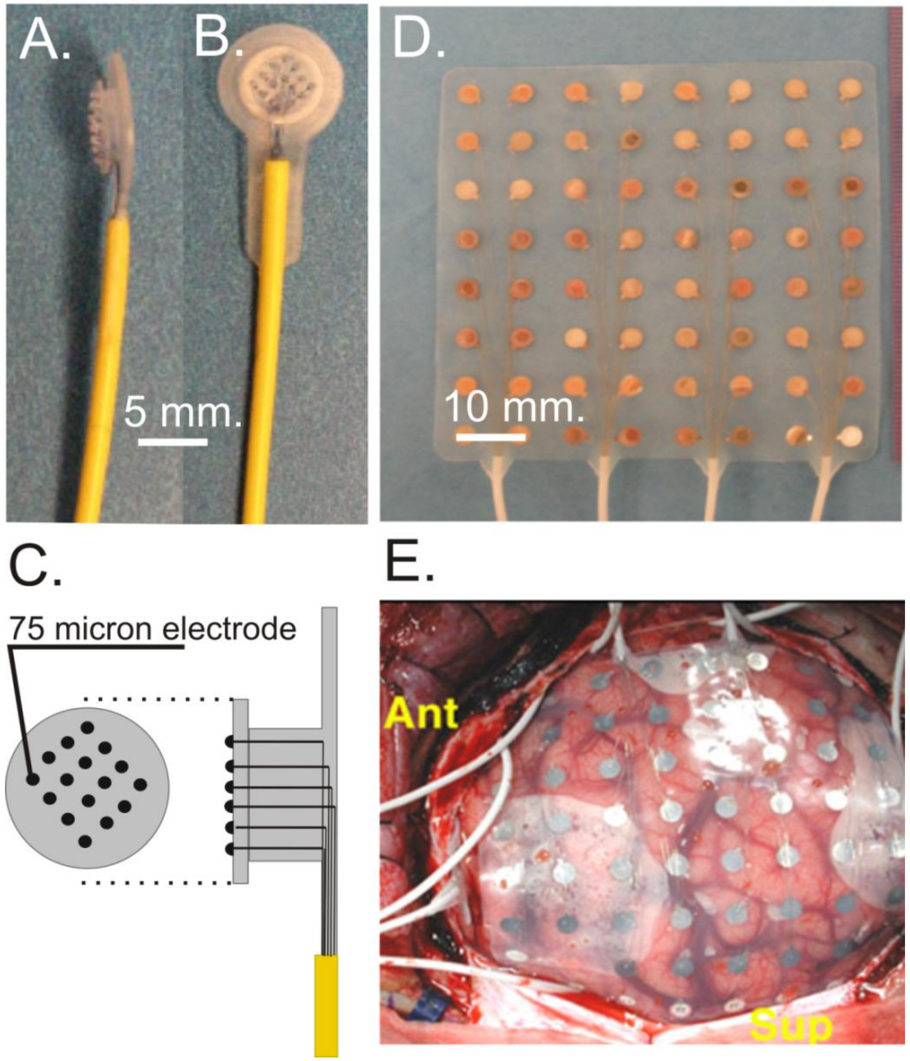
[From Schalk]

Invasive BCI at it's best



[From Schwartz]

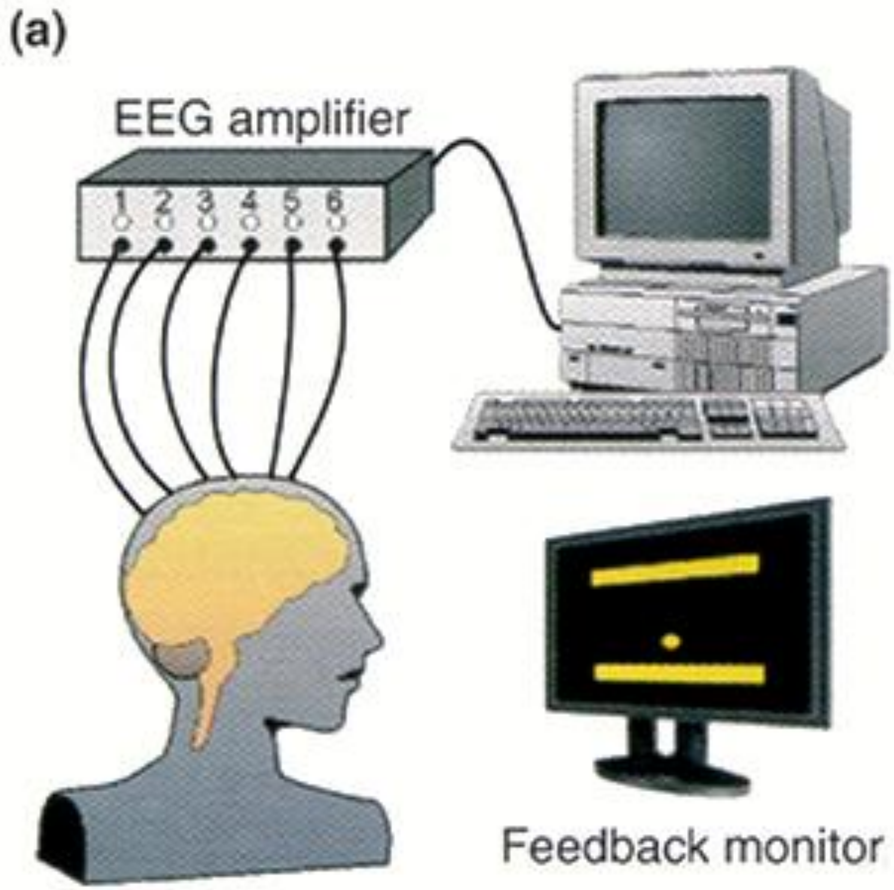
ECOG

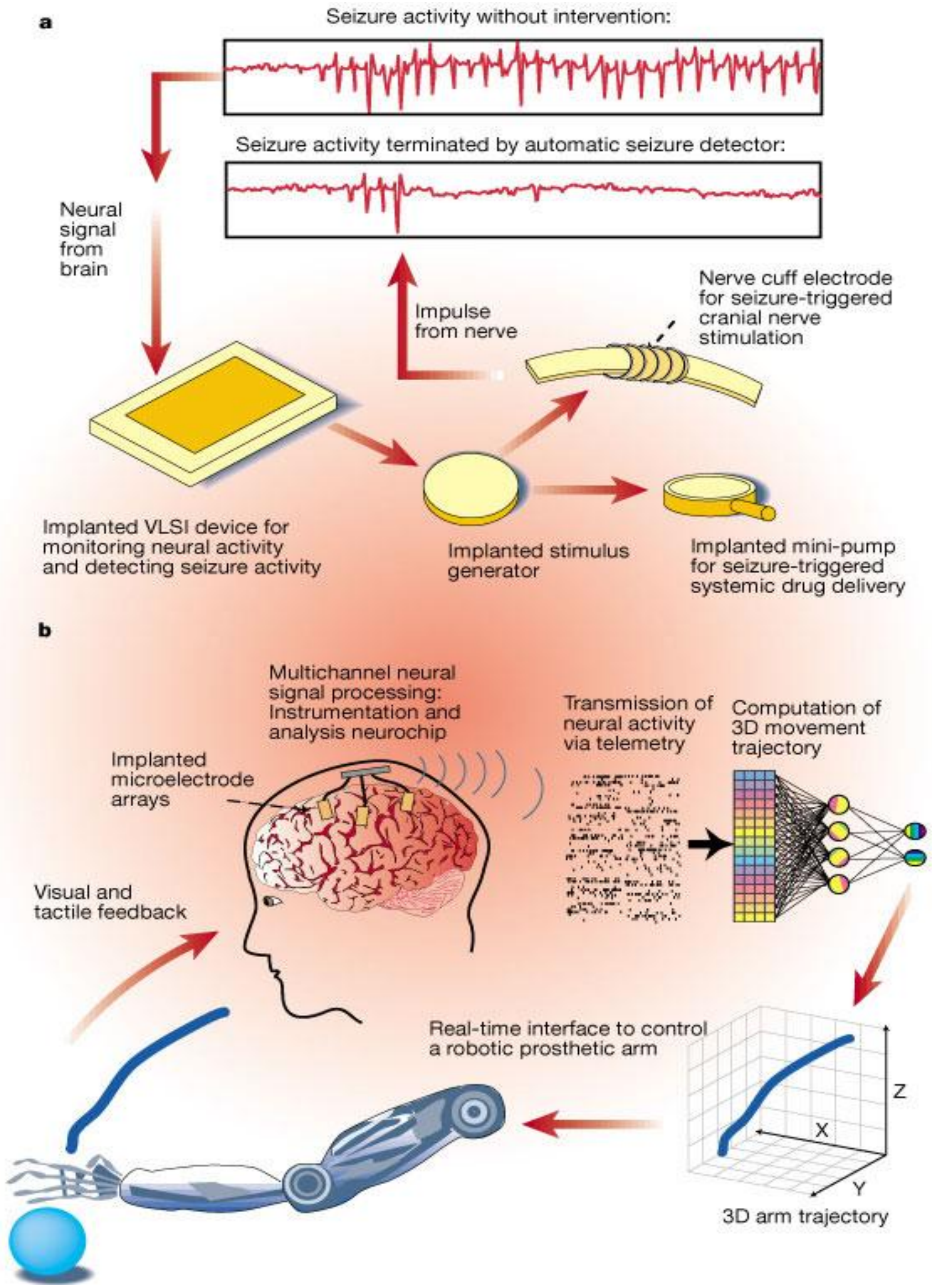


- presurgical localization of area causing epilepsy
- excellent possibilty to learn about brain for human subject

[From Schalk]

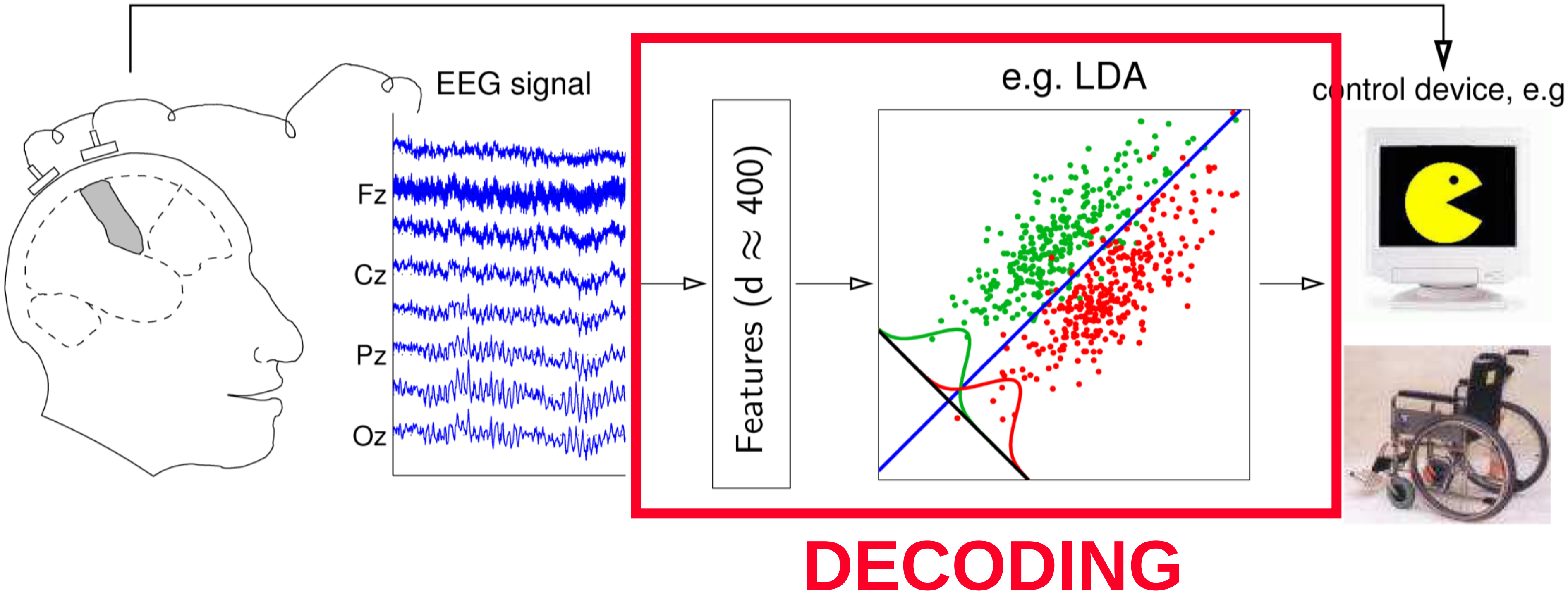
Invasive vs noninvasive Brain Computer Interfacing





[From Birbaumer et al., Nicolelis et al]

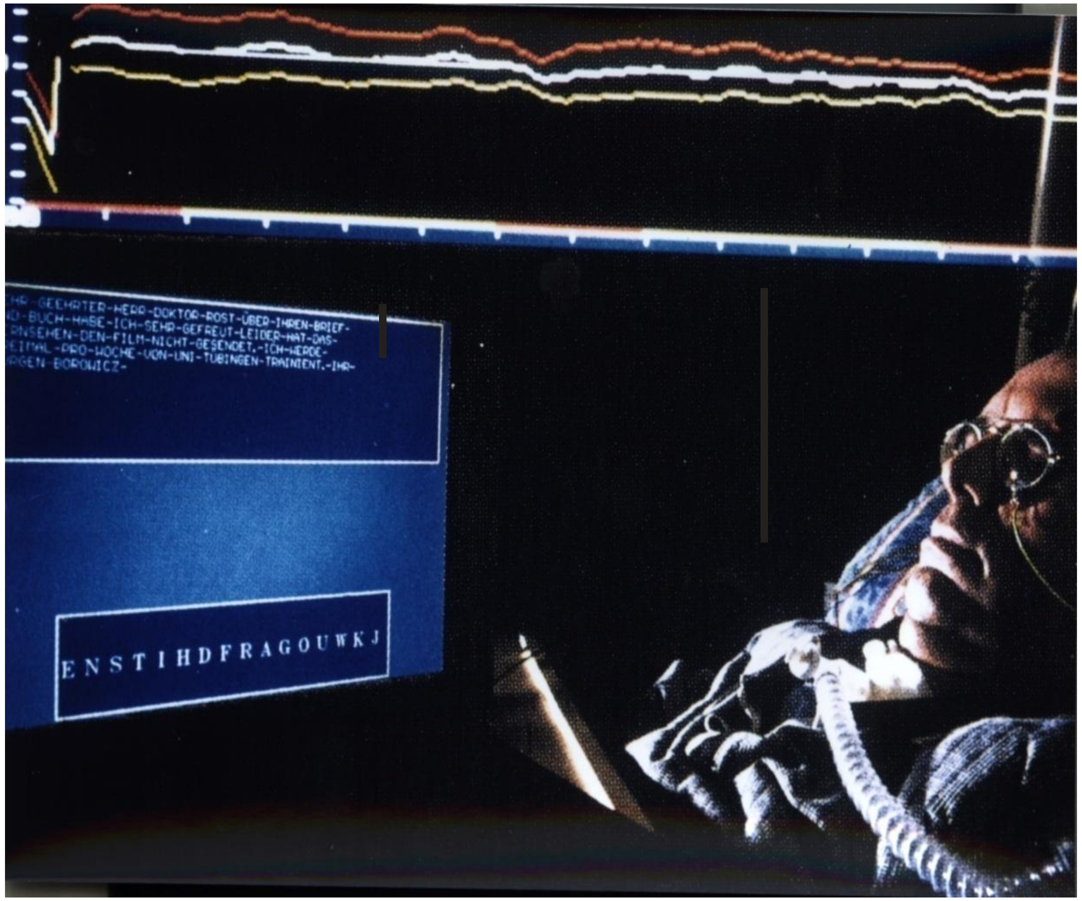
Noninvasive Brain-Computer Interface



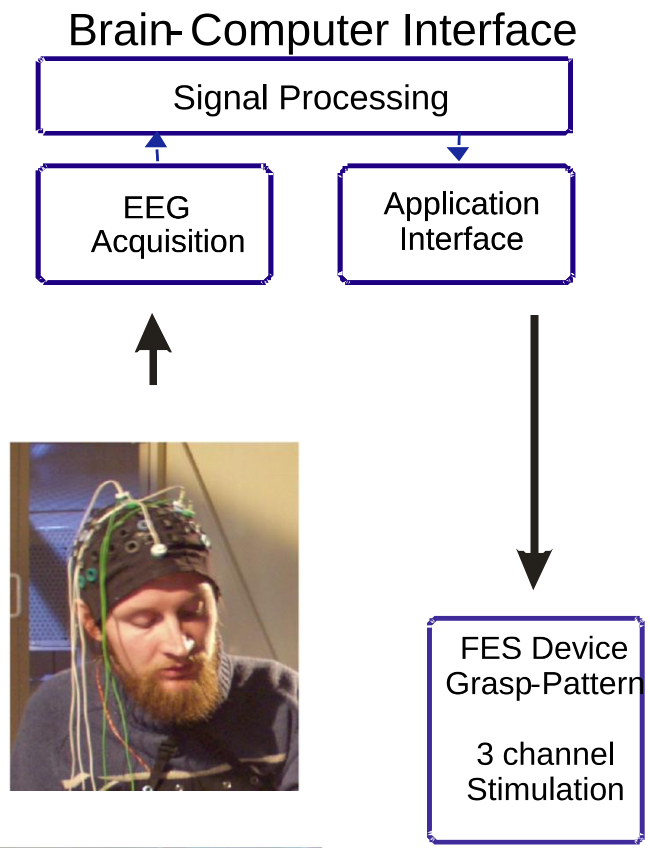
BCI: Translation of human intentions into a technical control signal **without using activity of muscles or peripheral nerves**

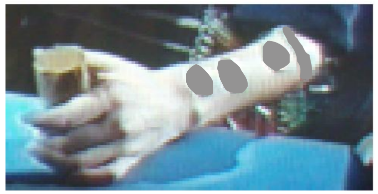
‚Brain Pong‘ with BBCI

Noninvasive BCI: clinical applications



[From Birbaumer et al.]

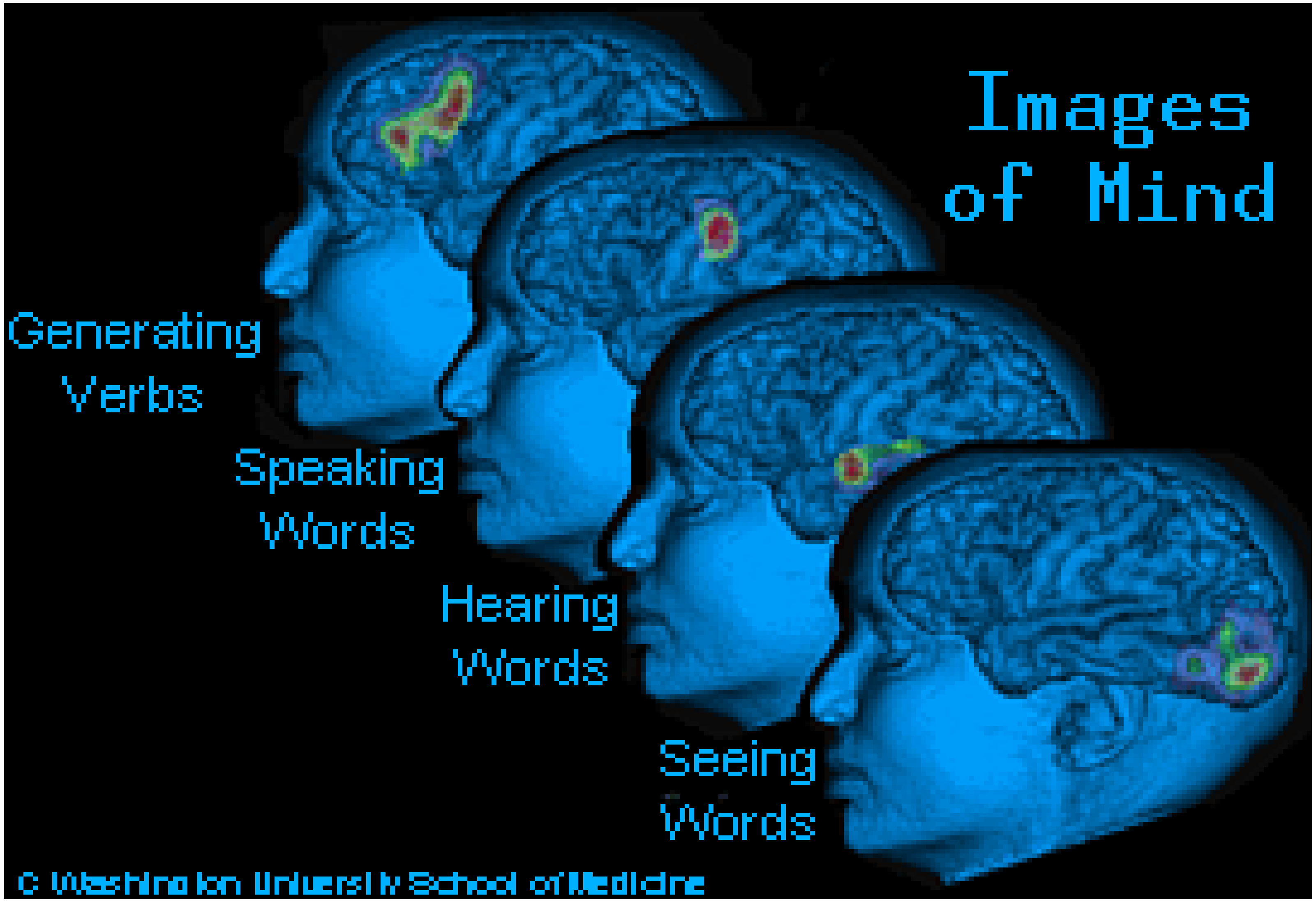


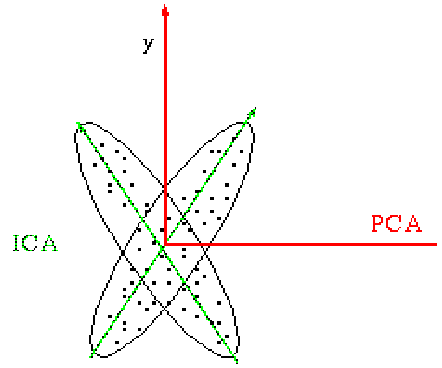


[From Pfurtscheller et al.]

BBCI: Leitmotiv: *›let the machines learn‹*

The cerebral cocktail party problem





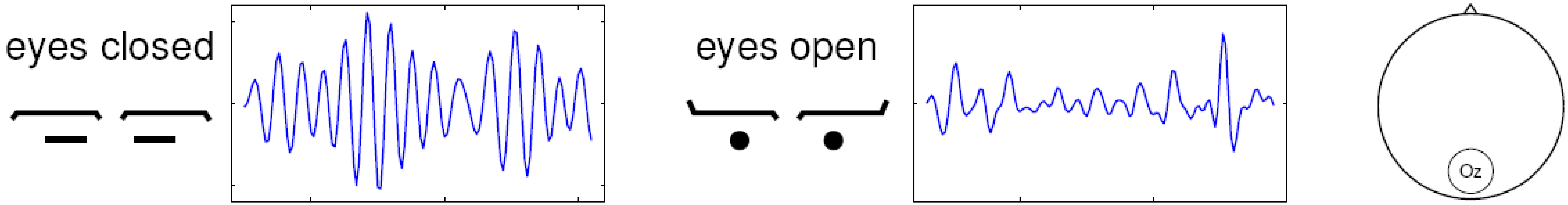
- use ICA/NGCA projections for artifact and noise removal
- feature extraction and selection

[cf. Ziehe et al. 2000, Blanchard et al. 2006]

Towards imaginations: Modulation of Brain Rhythms

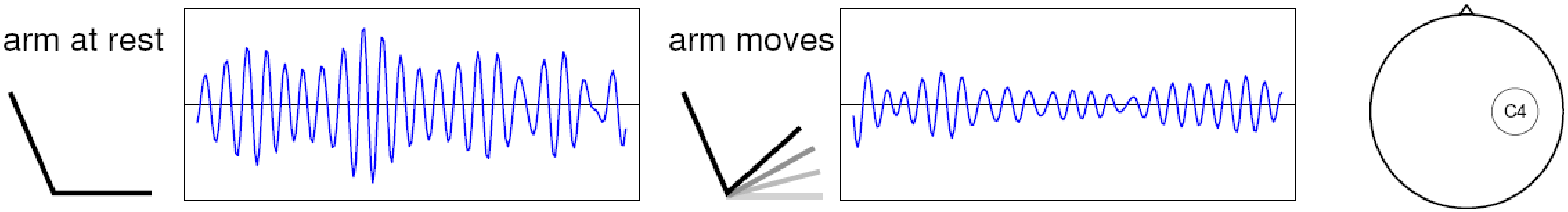
Most rhythms are idle rhythms, i.e., they are **attenuated** during activation.

- α-rhythm (around 10 Hz) in visual cortex:



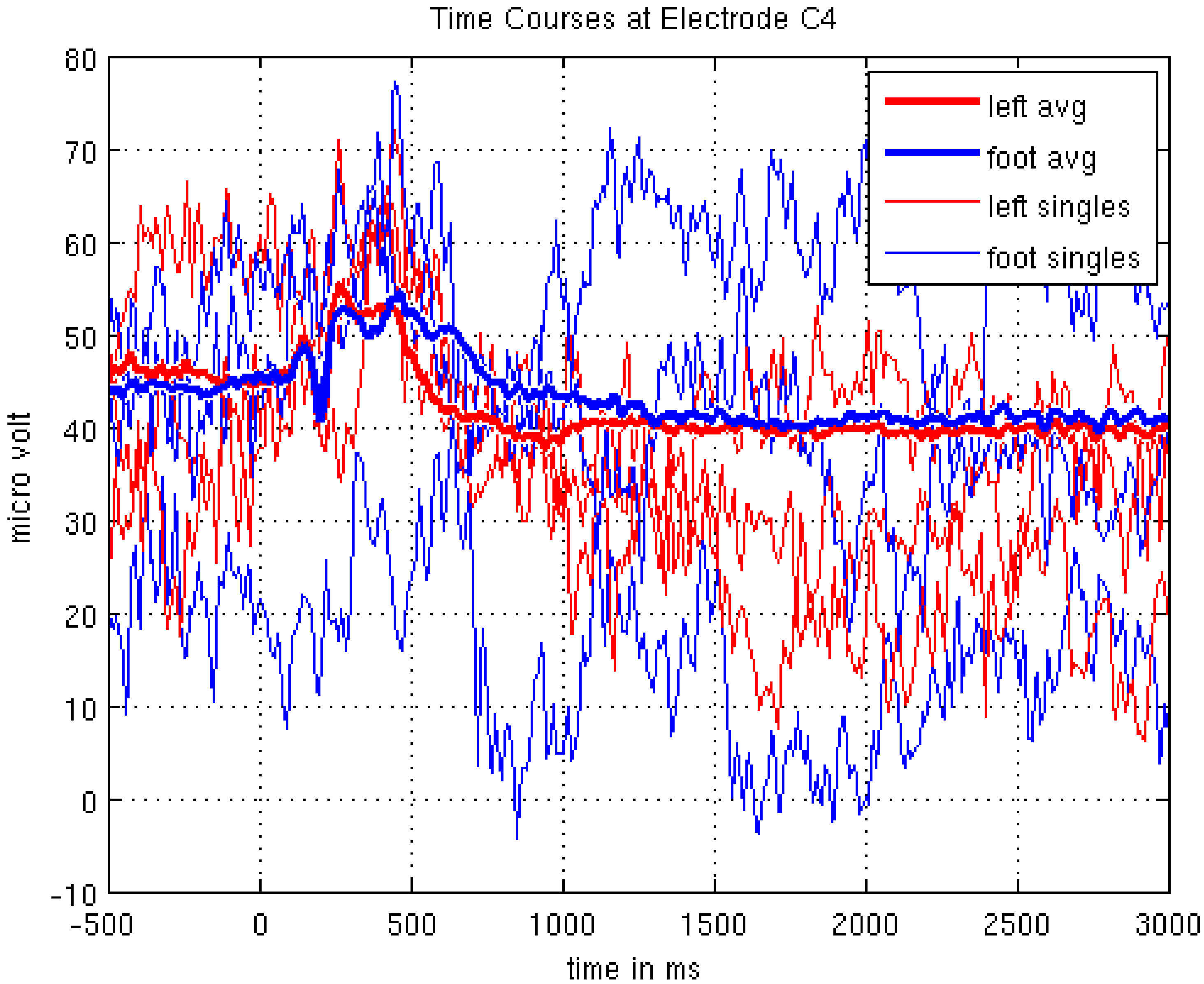
Single channel

- μ-rhythm (around 10 Hz) in motor and sensory cortex:



IMAGINATION of left arm

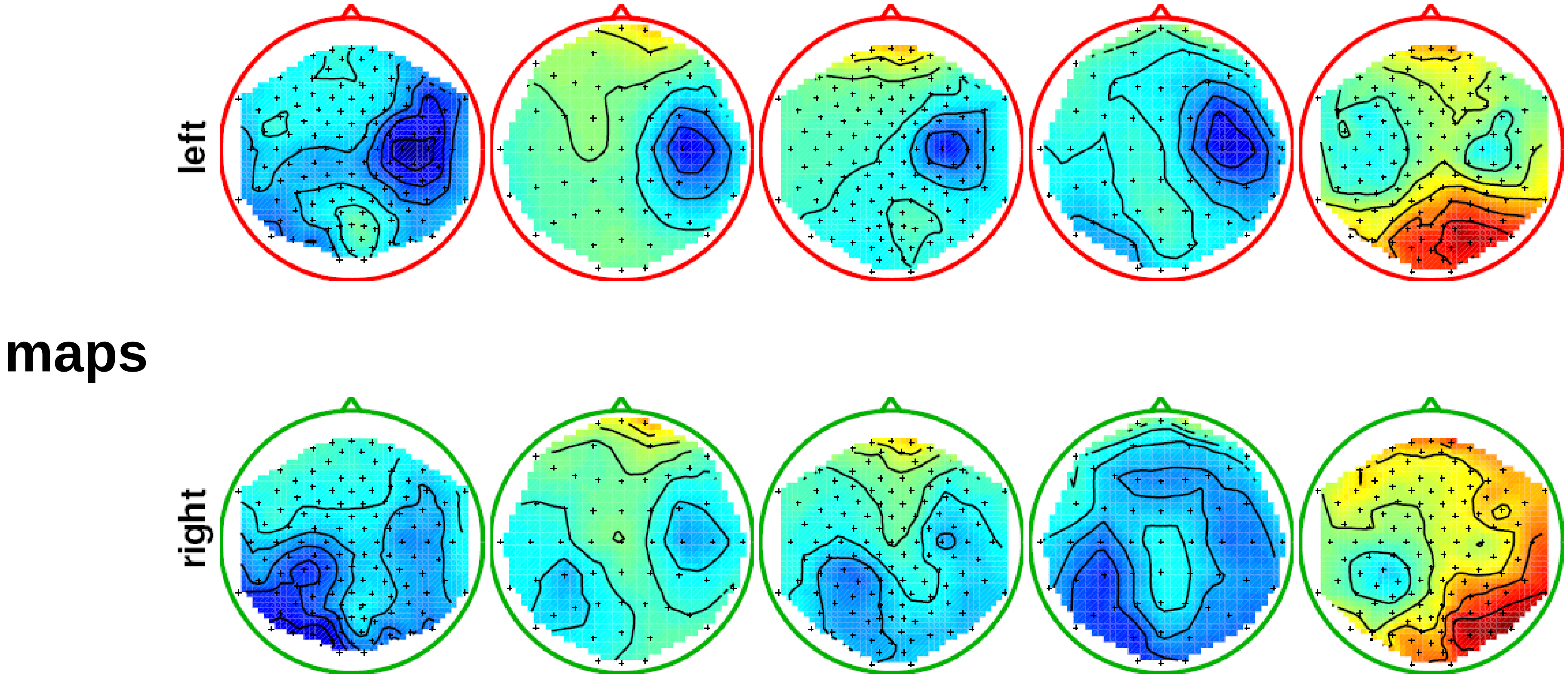
Variance I: Single-trial vs. Averaging



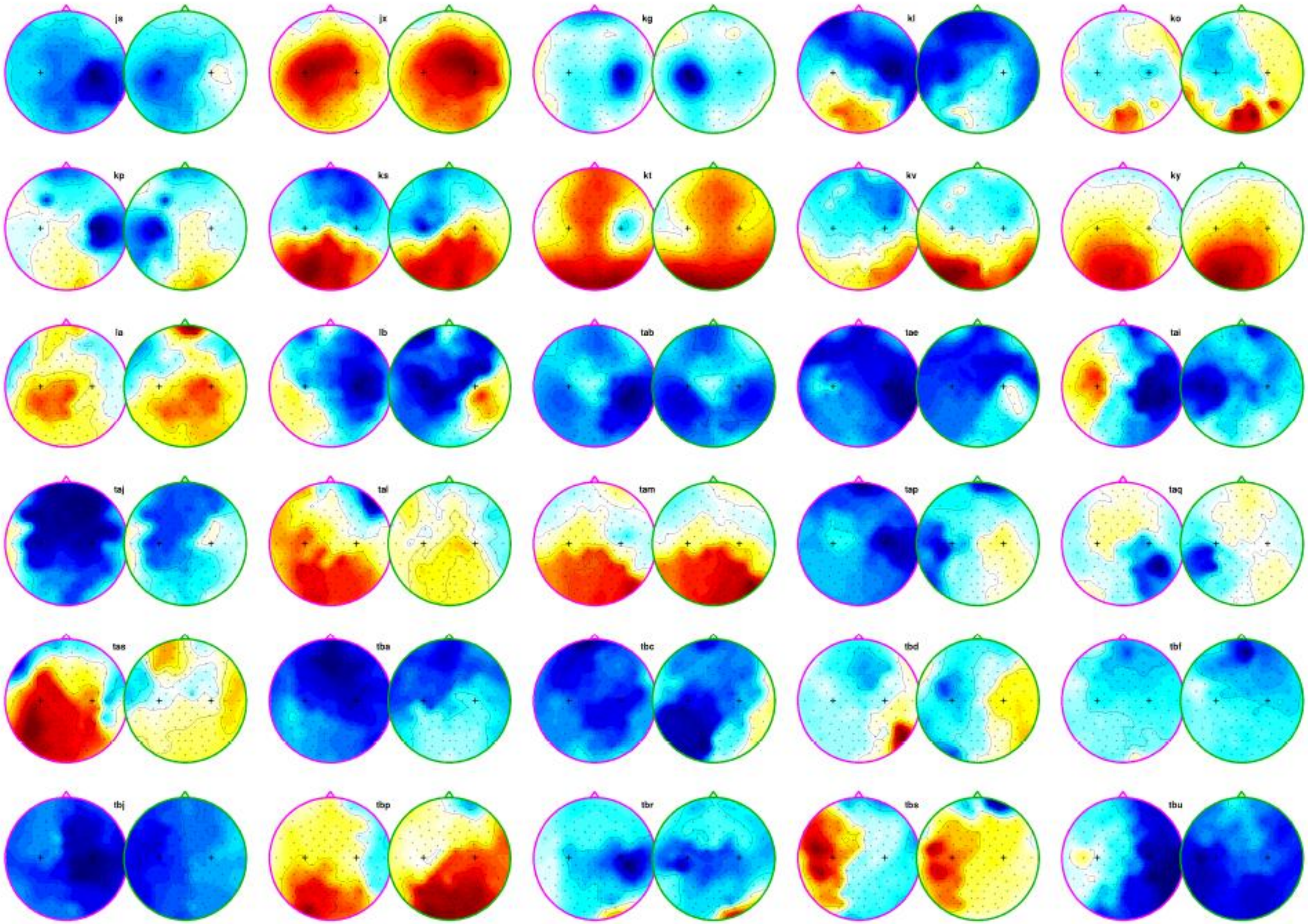
Single channel

Variance II: Session to Session Variability

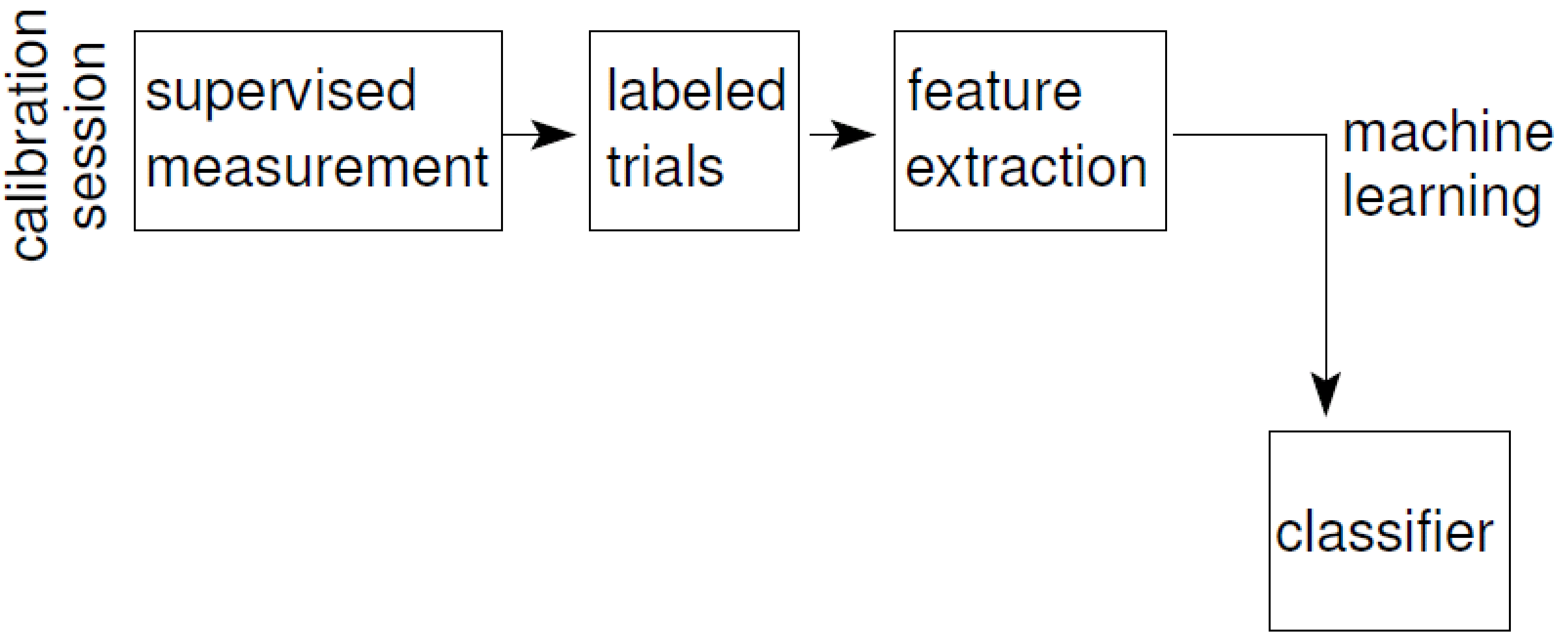
- Experiment: **One subject** imagined **left** vs. **right** hand movements on different days.
- Even though each ERD map represents an **average** across 140 trials, they exhibit an apparent diversity.



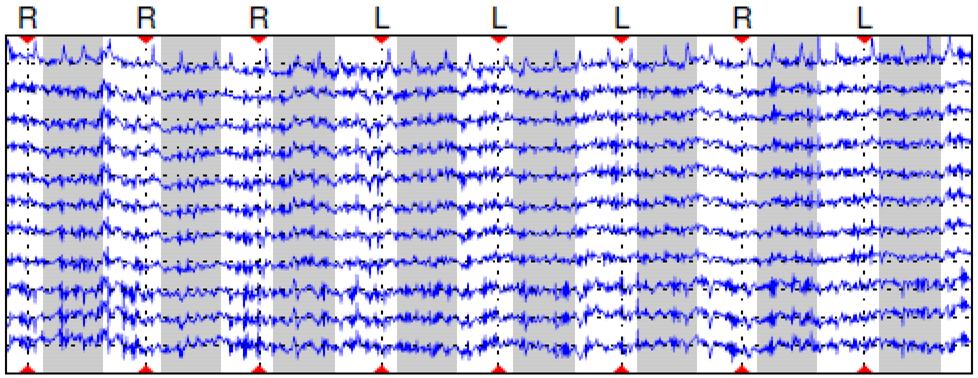
Variance III: inter subject variability [l vs r]



BCI with machine learning: training



offline: calibration (10–20 minutes)



collect training samples

BBCI paradigms

Leitmotiv: *›let the machines learn‹*

- healthy subjects *untrained* for BCI

A: training 20min: right/left hand **imagined** movements

→ infer the respective brain acivities (ML & SP)

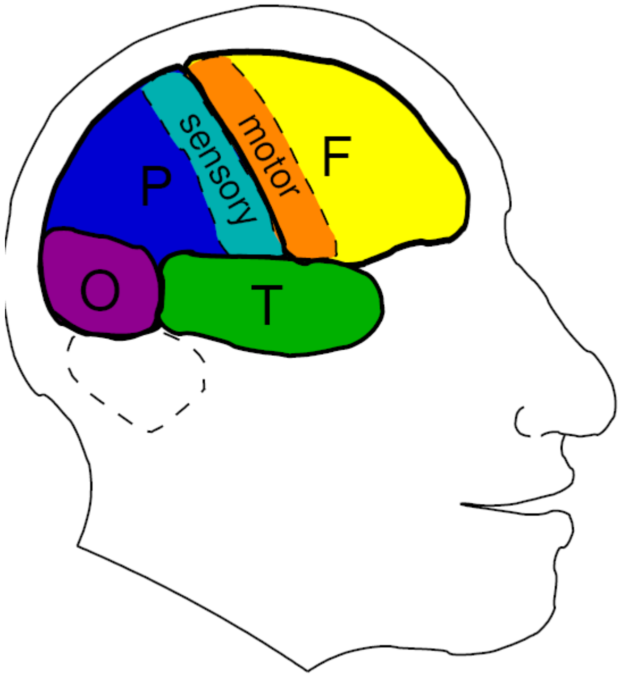
B: online feedback session

BBCI paradigms

Leitmotiv: *›let the machines learn‹*

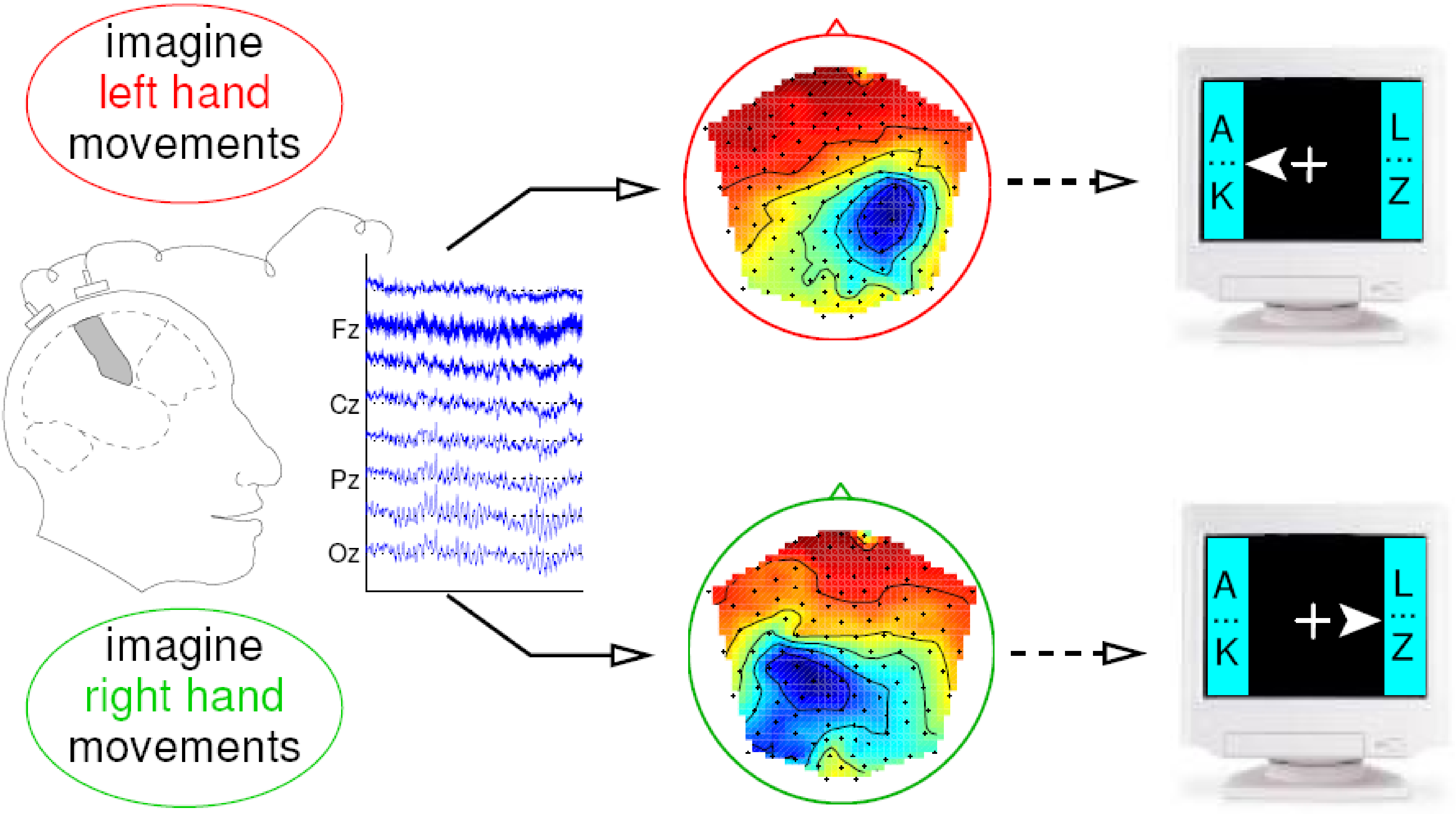
- healthy subjects (BCI *untrained*) perform "imaginary" movements (ERD/ERS)

- instruction: imagine
  - squezzing a ball,
  - kicking a ball,
  - feel touch



Playing with BCI: training session (20 min)

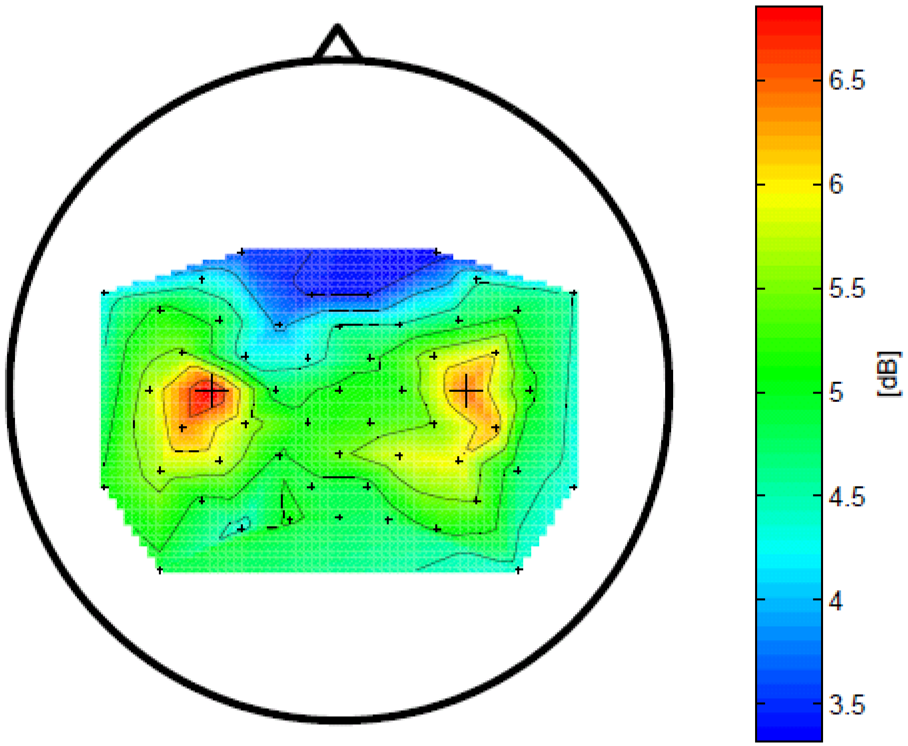
Machine learning approach to BCI: infer prototypical pattern



Inference by CSP Algorithm

Average topology of idle SMR

For each Laplace filtered channel in a relax recording, the strength of the local rhythm was estimated. The grand average over 80 participants is displayed as topographic mapping:

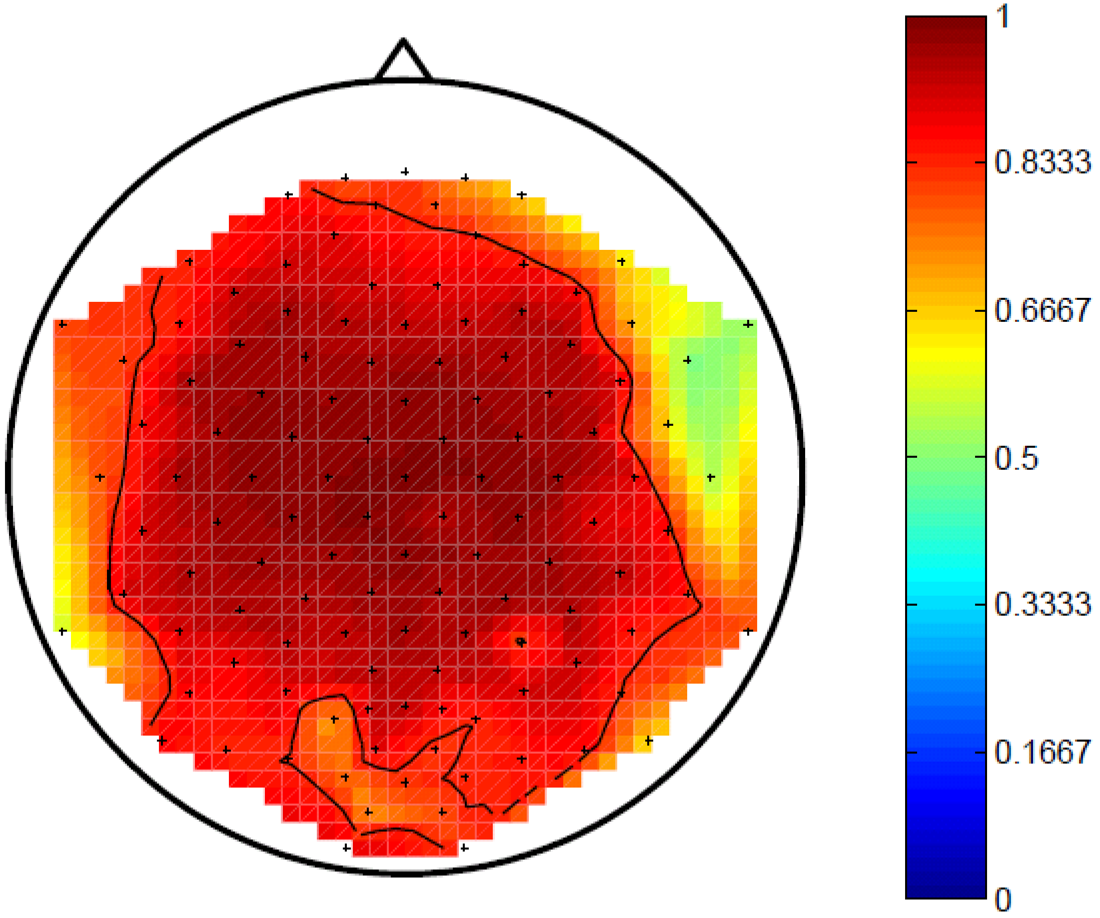


Conclusion

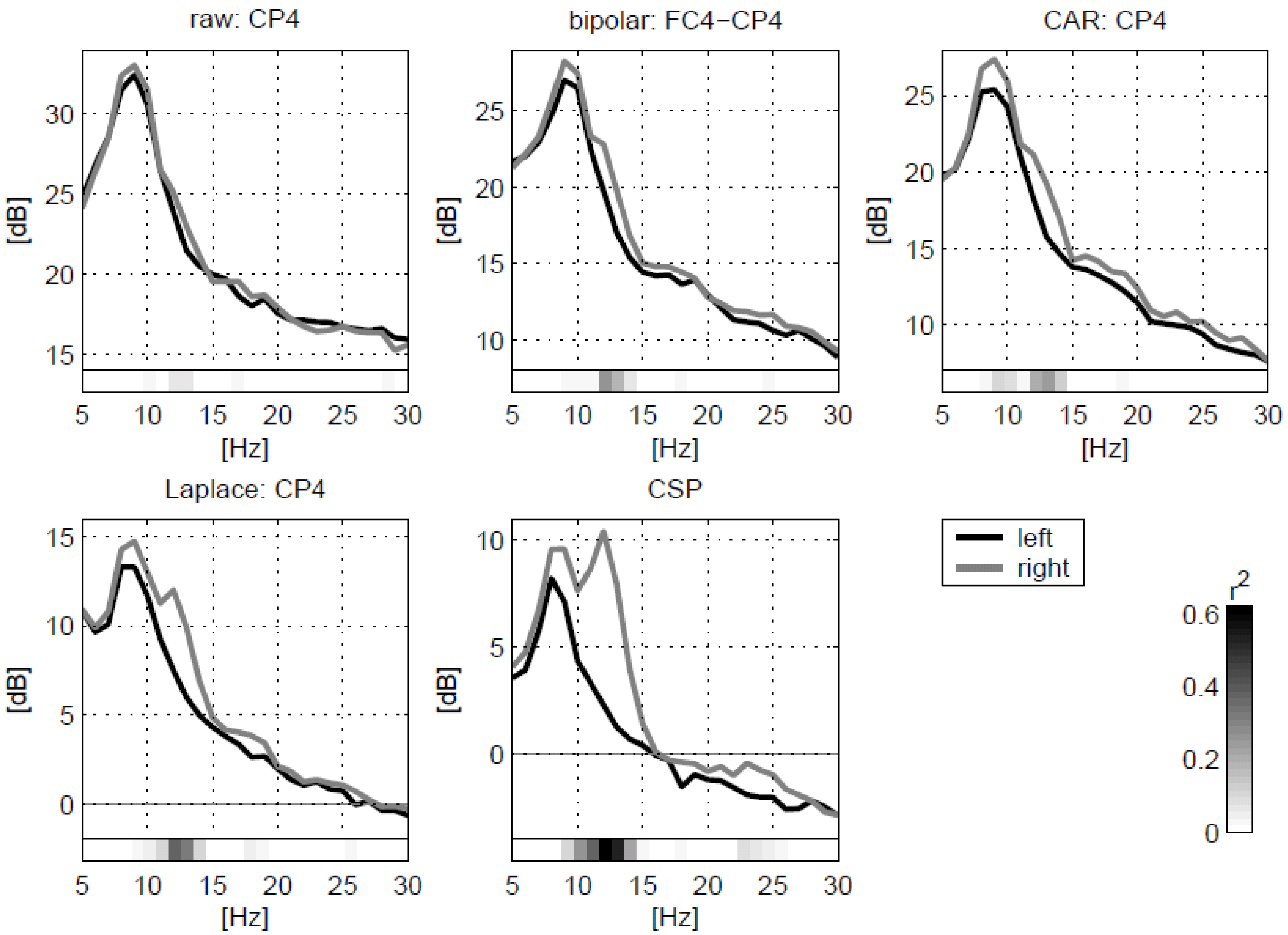
Locations C3 and C4 are good candidates to observe SMR modulations. These cover the sensorimotor areas of the right and the left hand.

Spatial Smearing

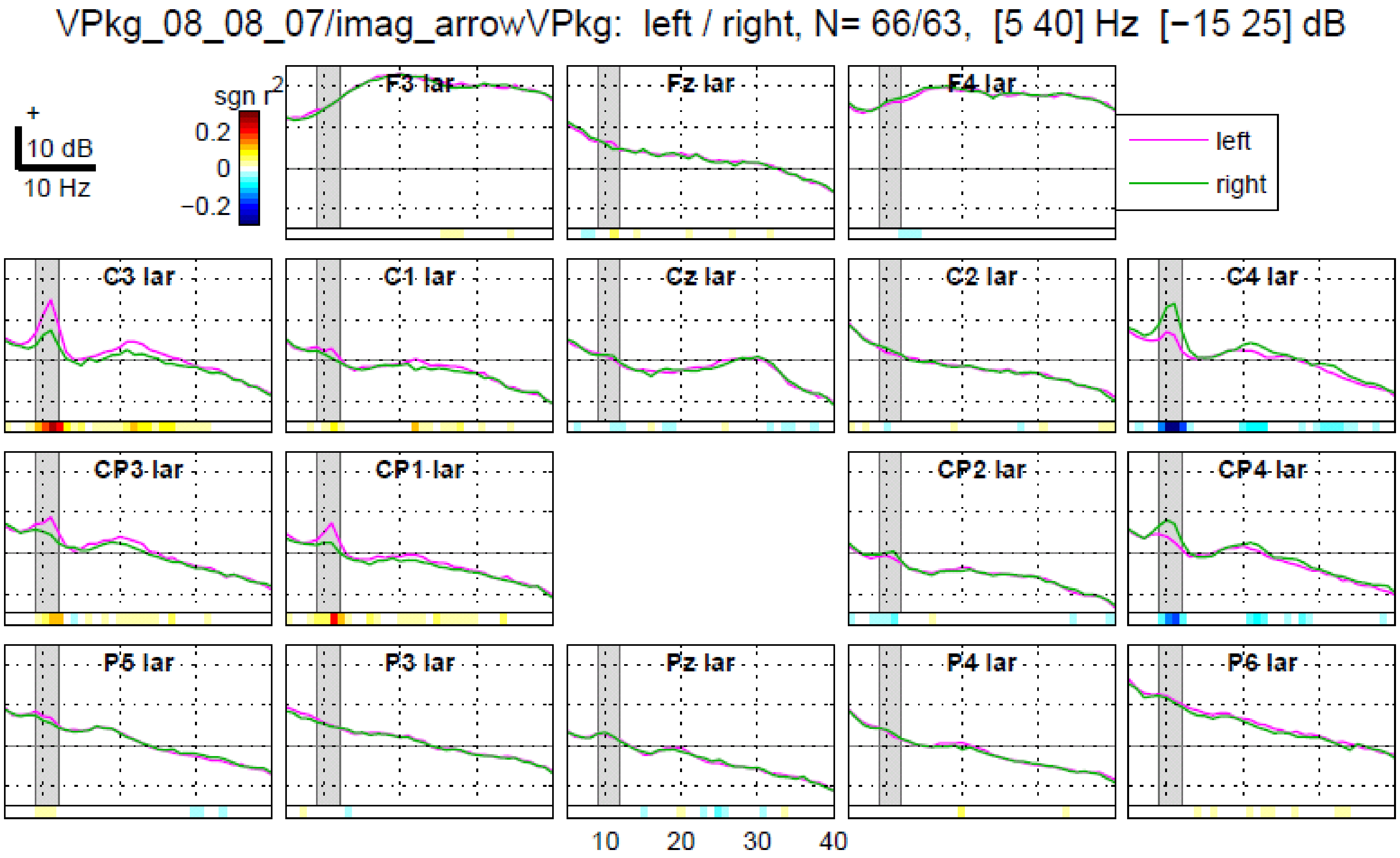
- Raw EEG scalp potentials are known to be associated with a large spatial scale owing to volumne conduction.
- In a simulation of Nunez et al [8] only half the contribution to one scalp electrode comes from sources within a 3 cm radius.



The need for spatial filtering

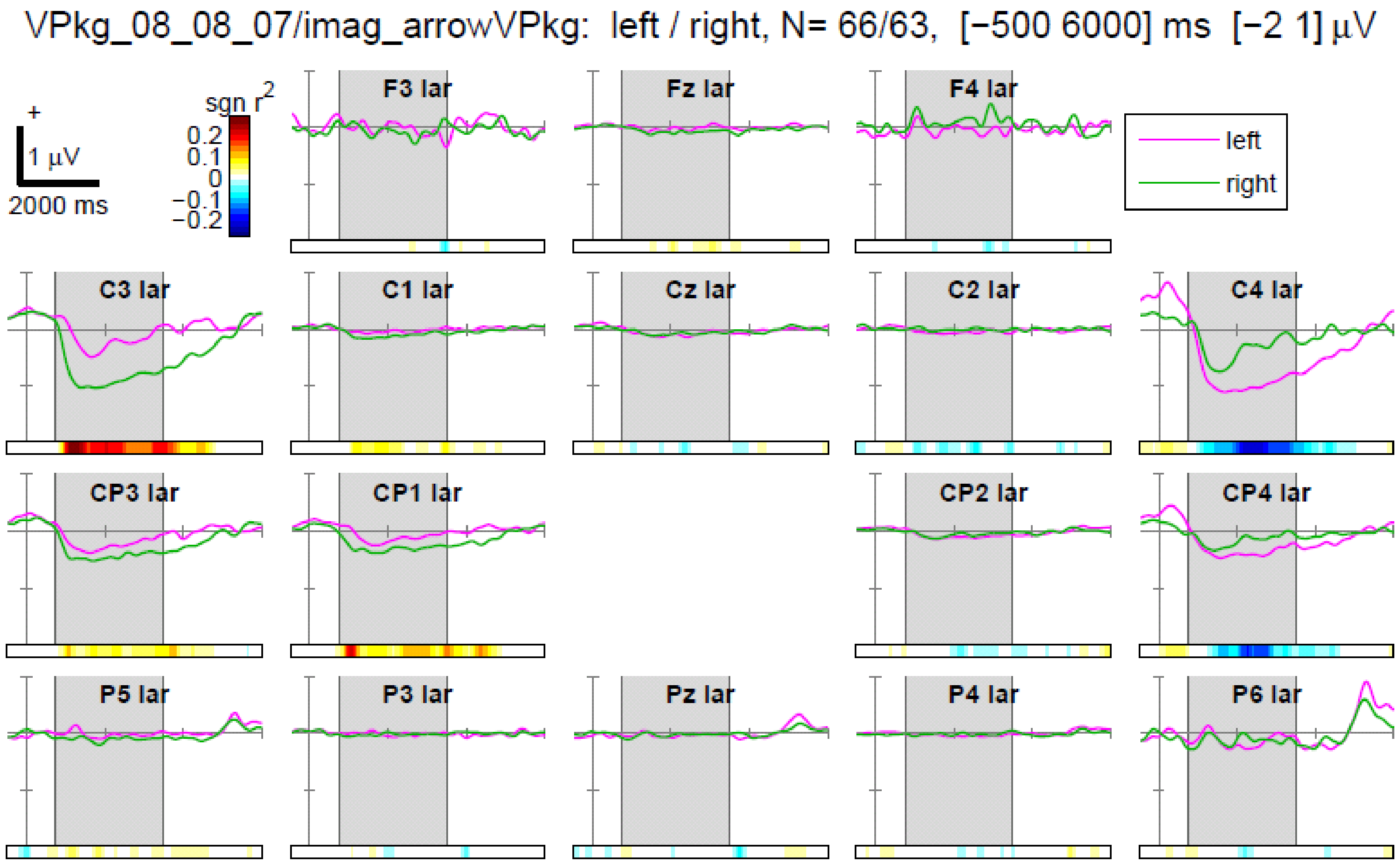


Analysis of motor imagery conditions: spectra



First step: determine a suitable frequency band that shows good discrimination between the conditions.

ERD curves of motor imagery



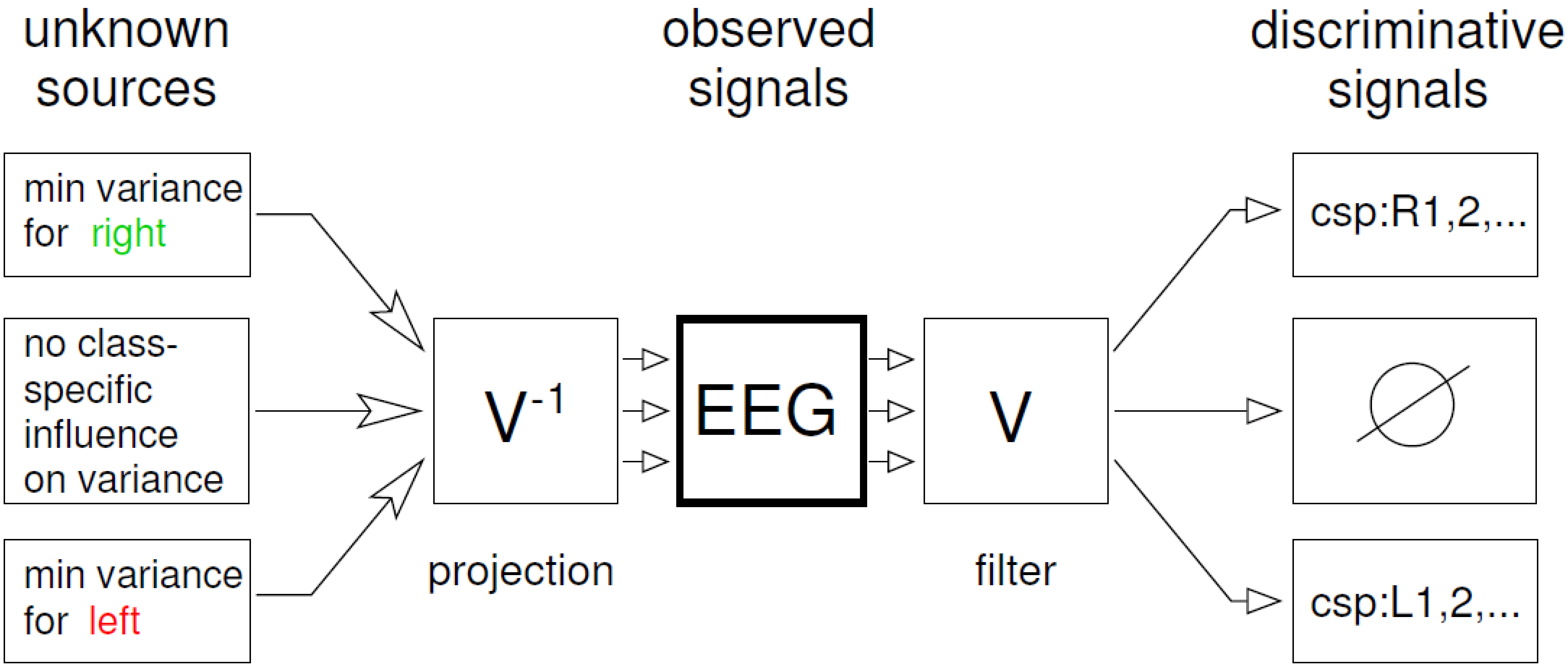
Second step: determine a suitable time interval during which discrimination is most prominent.

Remark: Simultaneous selection of frequency band and interval is more appropriate.

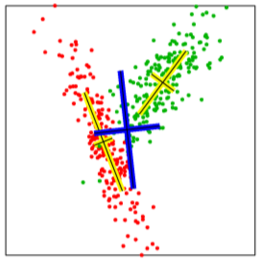
Common Spatial Pattern Analysis

Goal: Find spatial filters that optimally capture modulations of brain rhythms

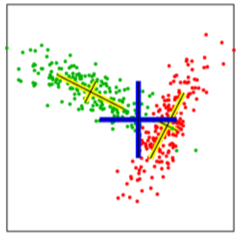
Observation: power of a brain rhythm $\sim$ variance of band-pass filtered signal.



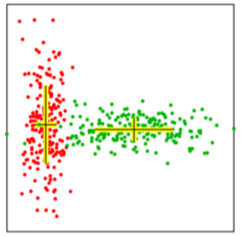
Common Spatial Patterns for 2 classes



Original data: Each class has a specific spatial extension.
Let Σ_1 and Σ_2 be the covariance matrices of the two classes.
The blue cross visualizes the covarianz matrix of $\Sigma_1 + \Sigma_2$.



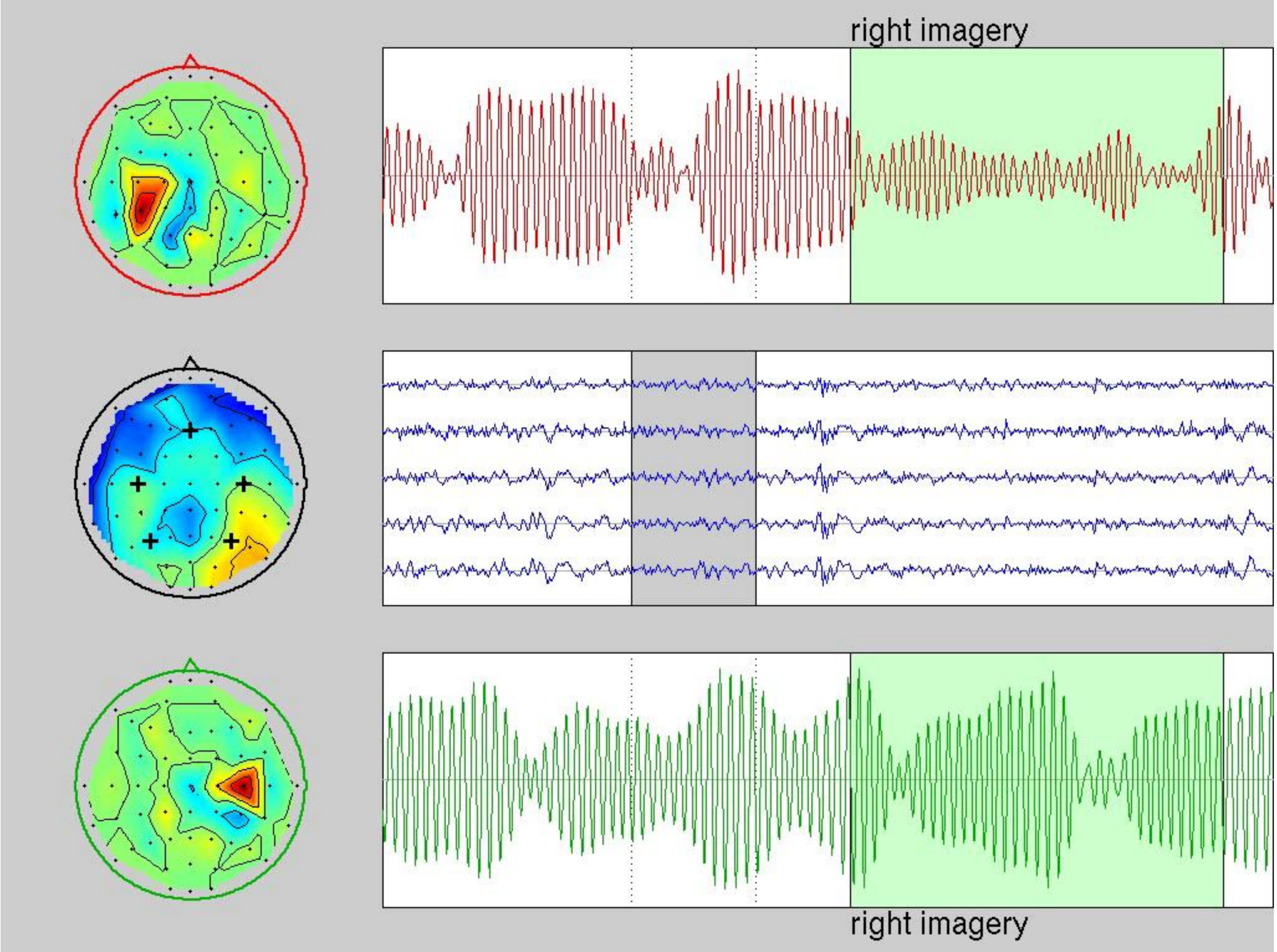
Make a whitening of $\Sigma_1 + \Sigma_2$, i.e., determine matrix P such that $P(\Sigma_1 + \Sigma_2)P^\top = I$ (possible due to positive definiteness of $\Sigma_1 + \Sigma_2$).
➤ Principal axis of the classes are perpendicular. Define: $\hat{\Sigma}_i = P\Sigma_i P^\top$.



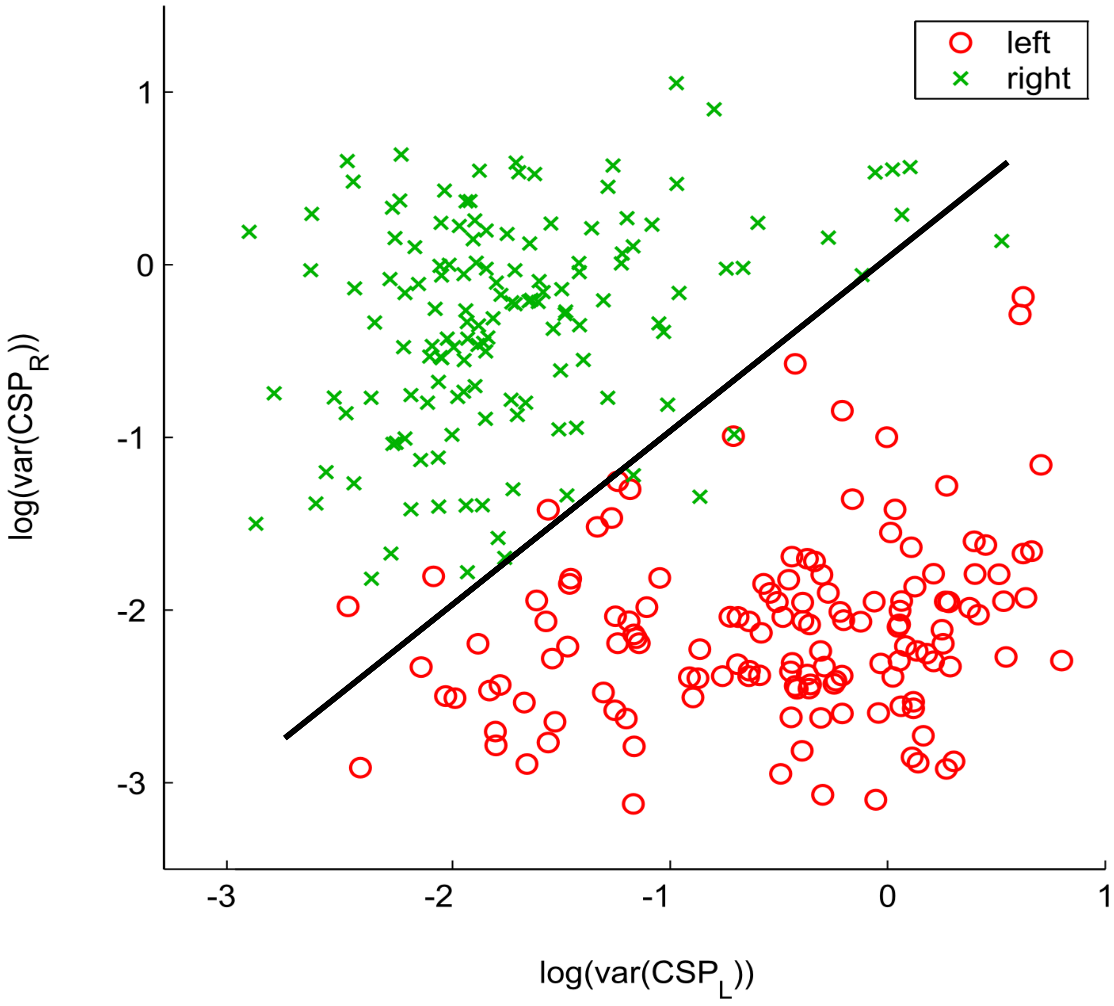
Calculate orthogonal matrix R and diagonal maxtrix D by spectral theory such that $\hat{\Sigma}_1^\top = RDR^\top$. Therefore $\hat{\Sigma}_2^\top = R(1-D)R^\top$ since $\hat{\Sigma}_1 + \hat{\Sigma}_2 = I$.
➤ Variance along the axis of input space is complementatory with respect to the two classes.

Essential idea for multi-class extension:
CSP is based on the **simultaneous diagonalization** of two covariance matrices with corresponding eigenvalues summing up to 1.

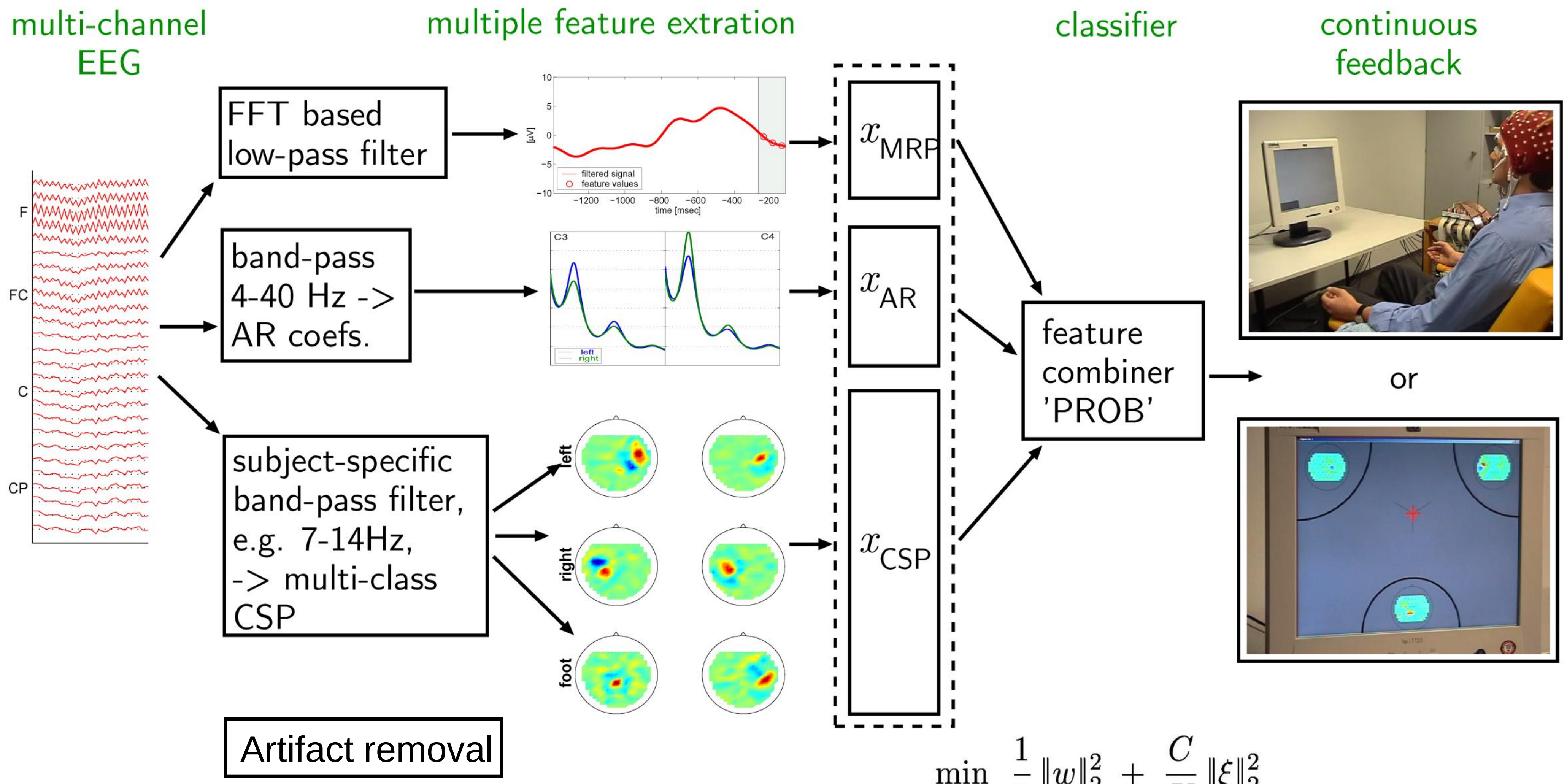
[cf. Blankertz et al. 2008, Lemm et al. 2005, Dornhege et al. 2006, Tomioka & Müller 2010]


right imagery
right imagery

Distribution of EEG features



BBCI Set-up

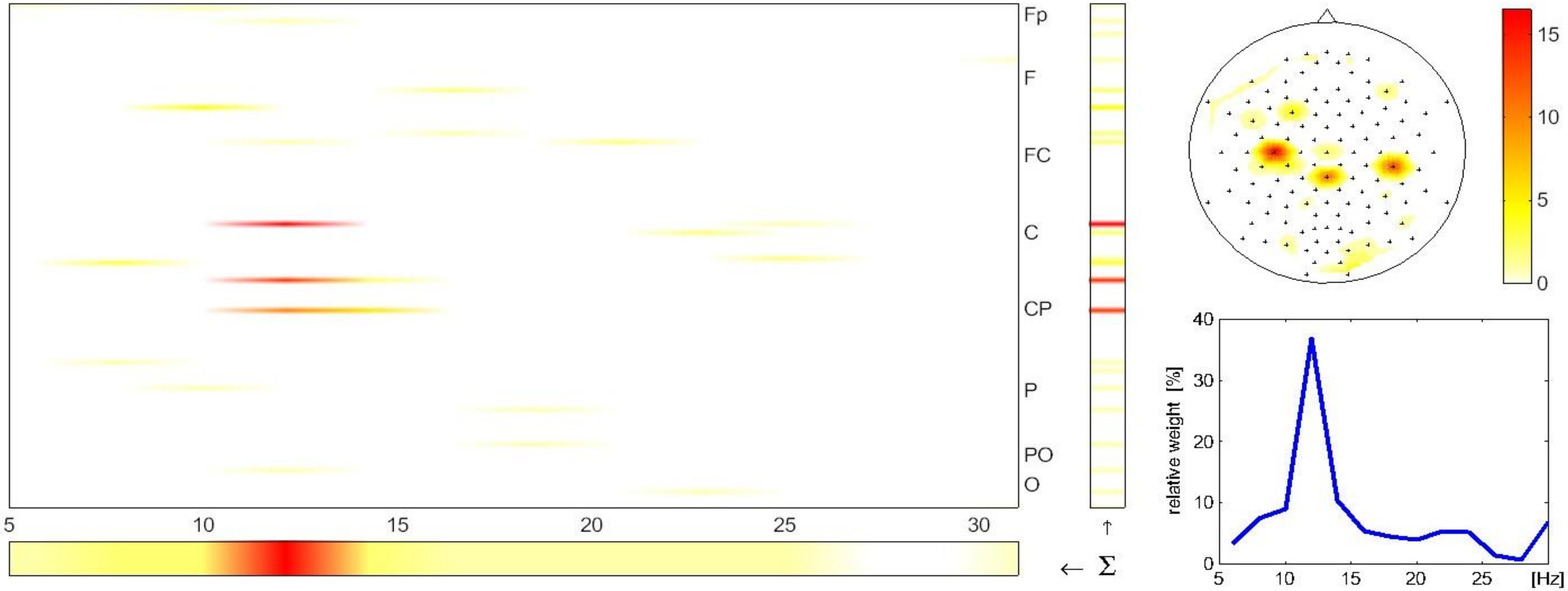


$$\min_{w,b,\xi} \ \frac{1}{2}\|w\|_2^2 + \frac{C}{K}\|\xi\|_2^2$$

$$\text{subject to} \qquad y_k(w^\top x_k + b) = 1 - \xi_k \quad \text{for } k = 1, \ldots, K$$

[cf. Müller et al. 2001, 2007, 2008, Dornhege et al. 2003, 2007, Blankertz et al. 2004, 2005, 2006, 2007, 2008]

What can Machine Learning tell us about physiology?

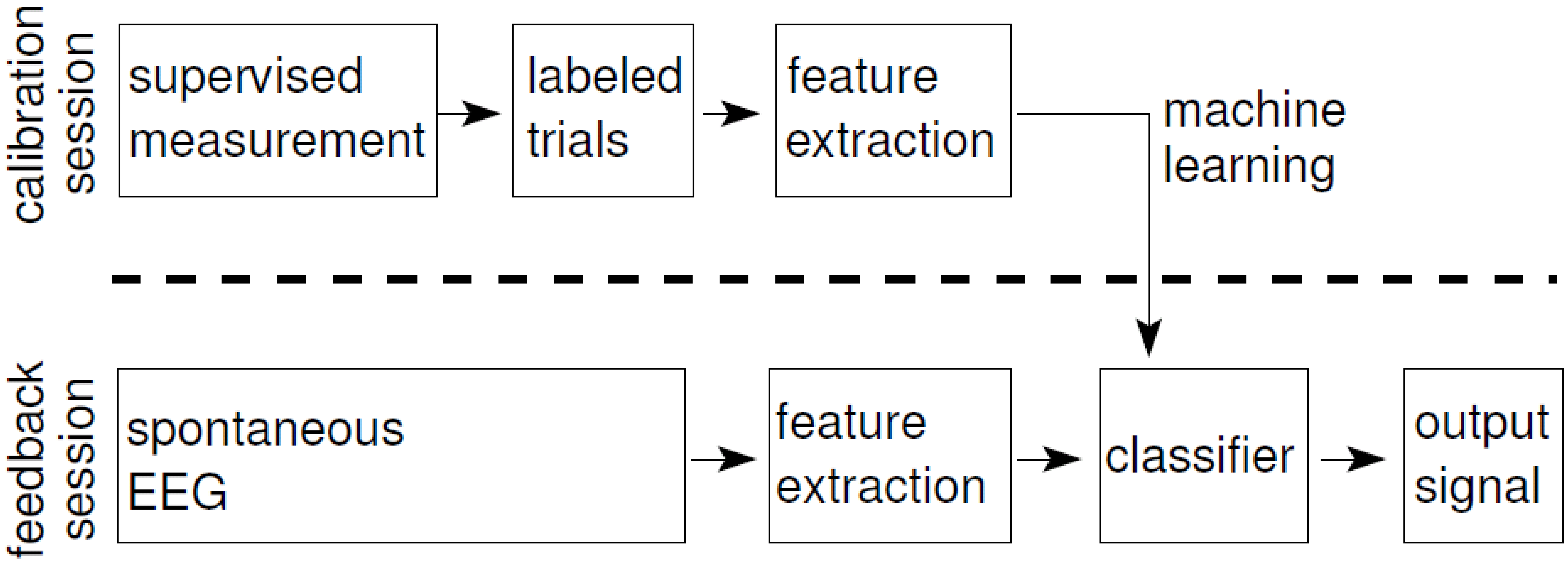


$$\min_{w,b,\xi} \; \frac{1}{2}\|w\|_1 \; + \; \frac{C}{K}\|\xi\|_1$$

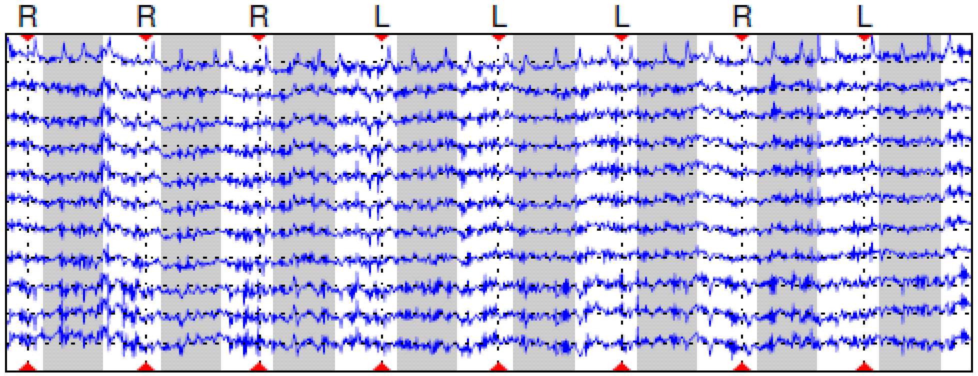
$$\text{subject to} \qquad y_k(w^\top x_k + b) = 1 - \xi_k \quad \text{for } k = 1, \ldots, K$$

[cf. Blankertz et al. 2001, 2006]

BCI with machine learning: feedback

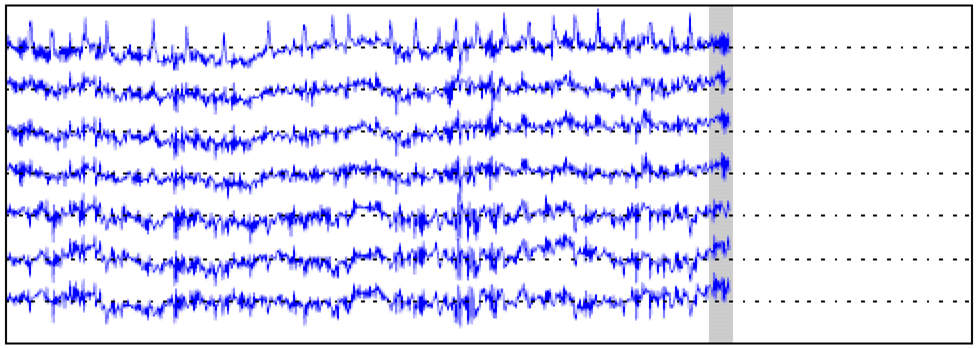


offline: calibration (10–20 minutes)



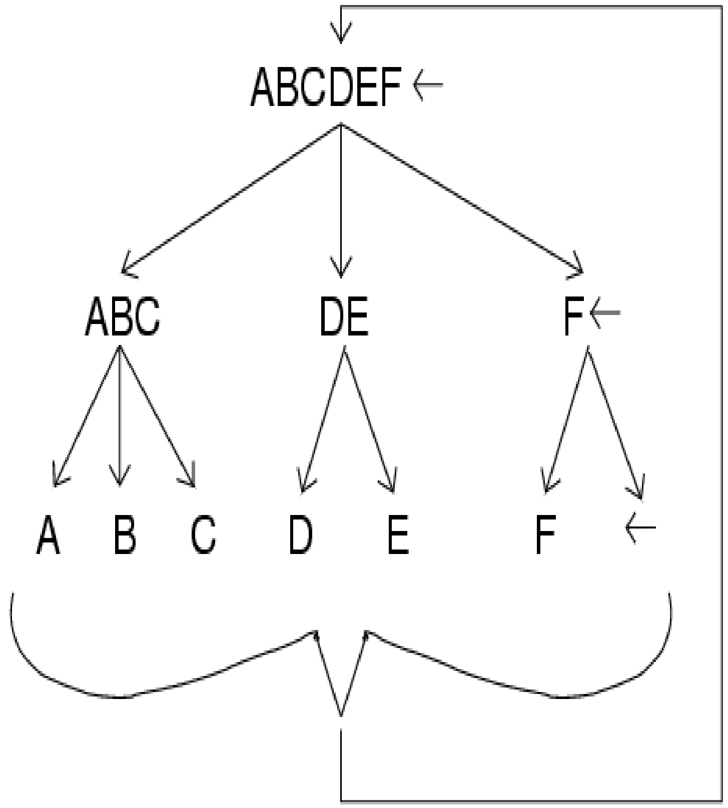
collect training samples

online: feedback (up to 6 hours)

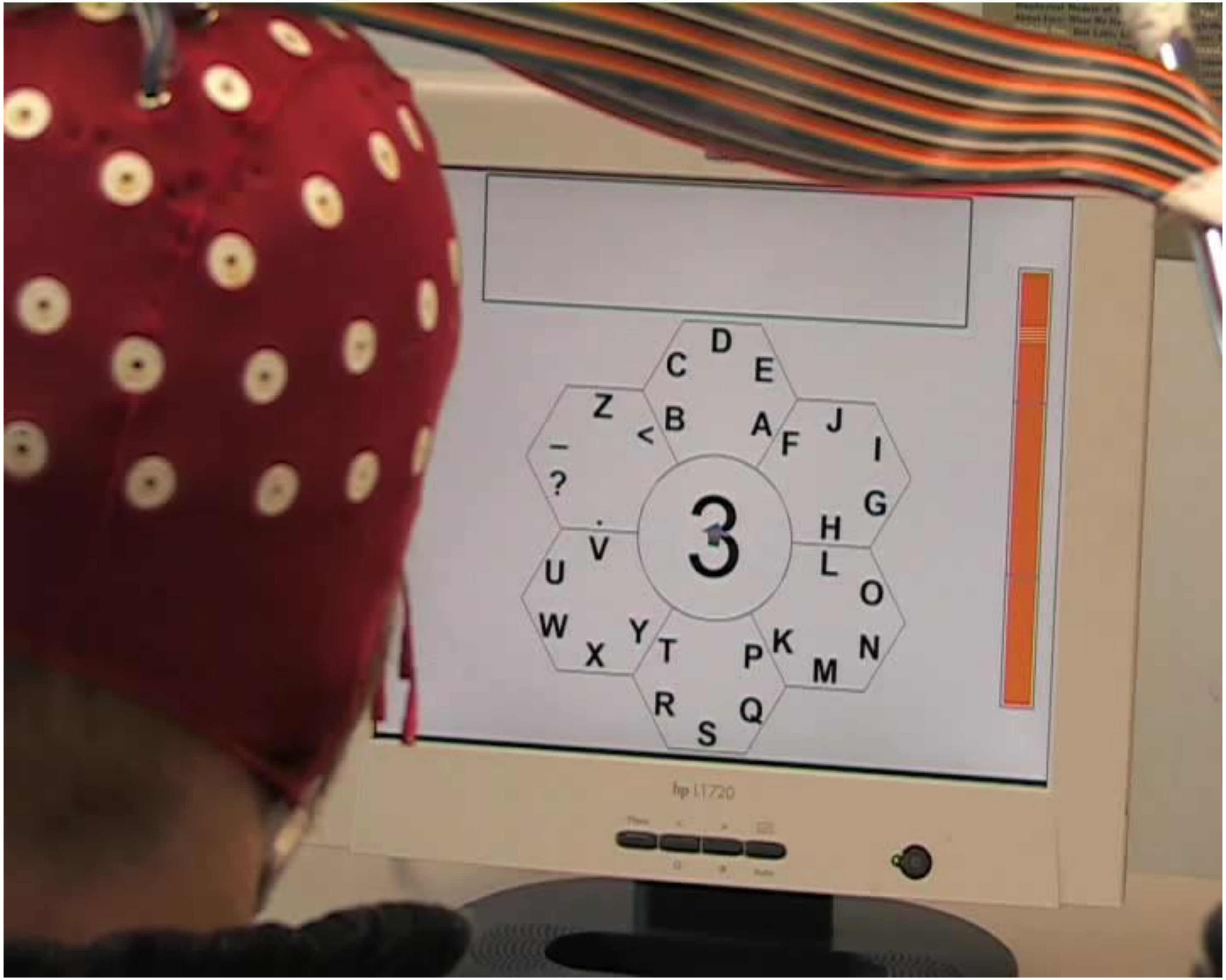


classification of sliding windows ($\leq$ 1s)

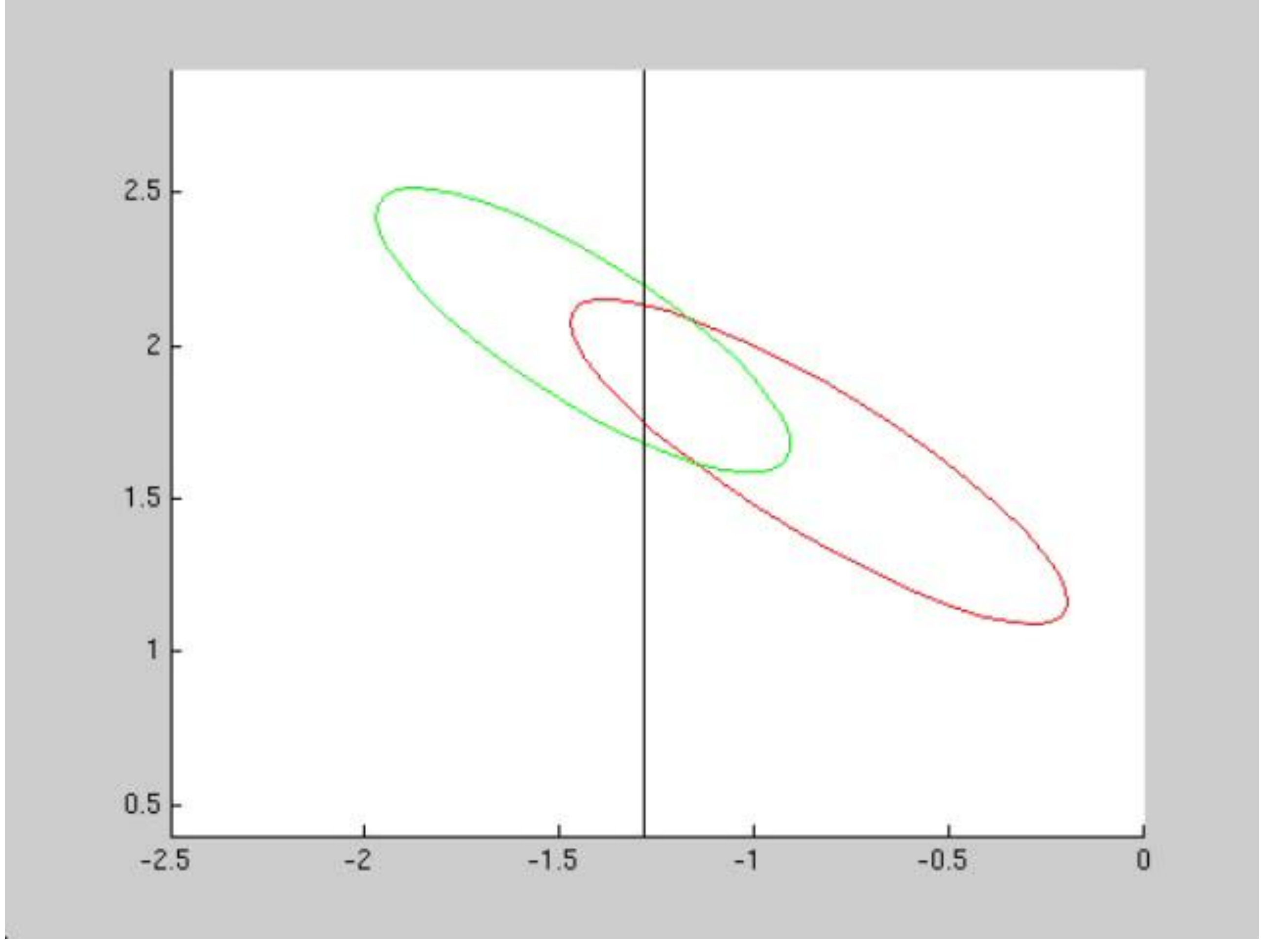
Spelling with BBCI: a communication for the disabled I



Spelling with BBCI: a communication for the disabled II



Variance IV: Shifting distributions within experiment



Interlude: Caveats in Validation

When machine learning techniques are used for classification of EEG single-trials, the expected performance of a method has to be evaluated carefully, and there are several possible pitfalls.

The estimation of generalization performance requires a training and a test set. The estimation is only proper

- if the test set was not used in any way to determine parameters of the method, and
- if the samples in the test set are independent from the samples in the training set.

Although these principles are quite obvious, it happens that they are violated.
Unfortunately, even some published journal articles lack a proper validation of the proposed methods.

[cf. Blankertz et al 2011]

Hall of pitfalls in single-trial EEG analysis (and beyond)

- preprocessing methods that use statistics of the whole data set like ICA, or normalization of features (particularly severe for methods that use label information)
- features are selected on the whole data set, including trials that are later in the test set
- select parameters by cross validation on the whole data set and report the performance for the selected values
- artifacts/outliers are rejected from the whole data set (resulting in a simplified test set)
- unsufficient validation for paradigms with block design

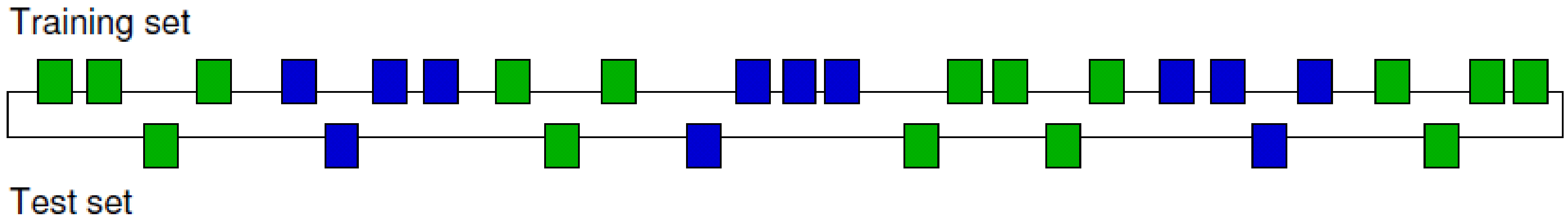
In this presentation we highlight the last issue.

Block design

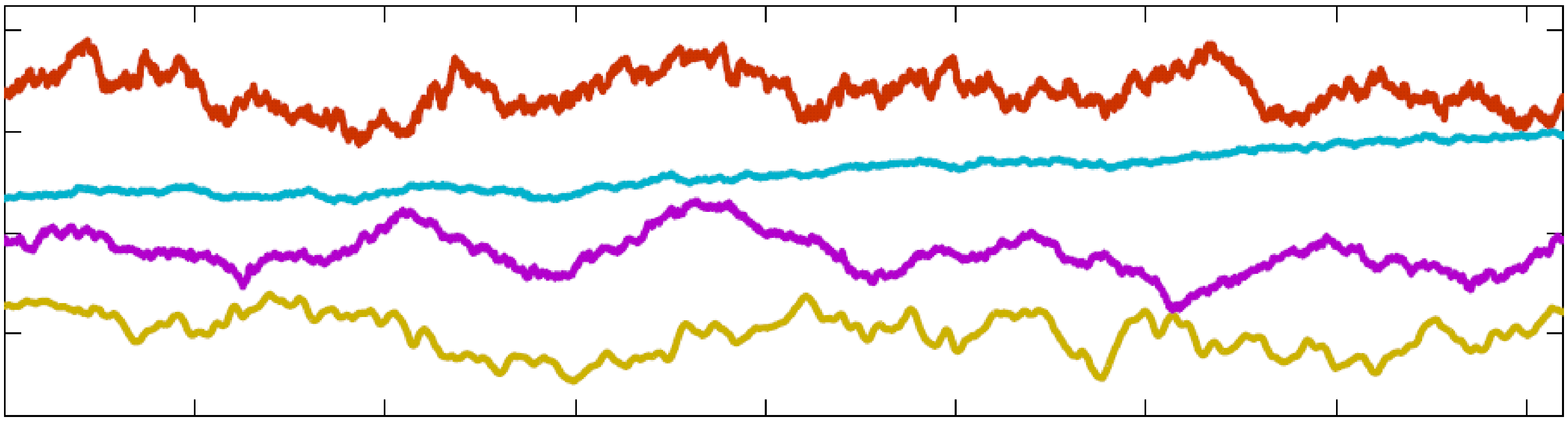
Assume the task is to discriminate between mental states in different conditions.
We say that an experiment has a block design, if the periods for which there is no alternation between conditions are longer than the intended change of states in online operation.

A problem arises, if the performance is estimated for such a data set by cross validation.

Slowly varying variables



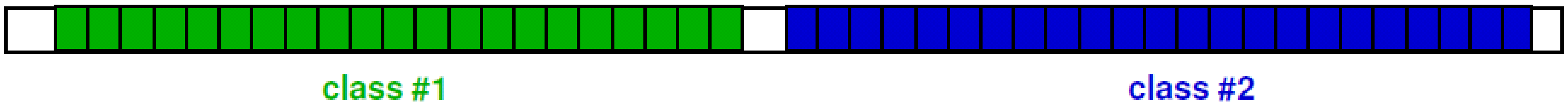
In EEG there are many slowly changing variables of background activity, therefore the single-trials are not independent. For an ordinary cross validation in a block design data set, the requirement of independence between training and test set is violated.



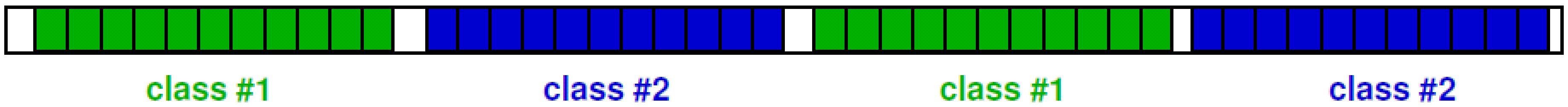
A validation test

To demonstrate impact of block design in cross validation, we perform cross validation in the following setting. Taking an arbitrary EEG data set, we assign **fake** labels (regardless of what happened during the recording) like this:

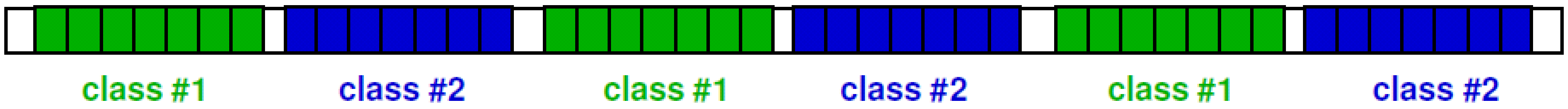
nBlocksPerClass=1:



nBlocksPerClass=2:



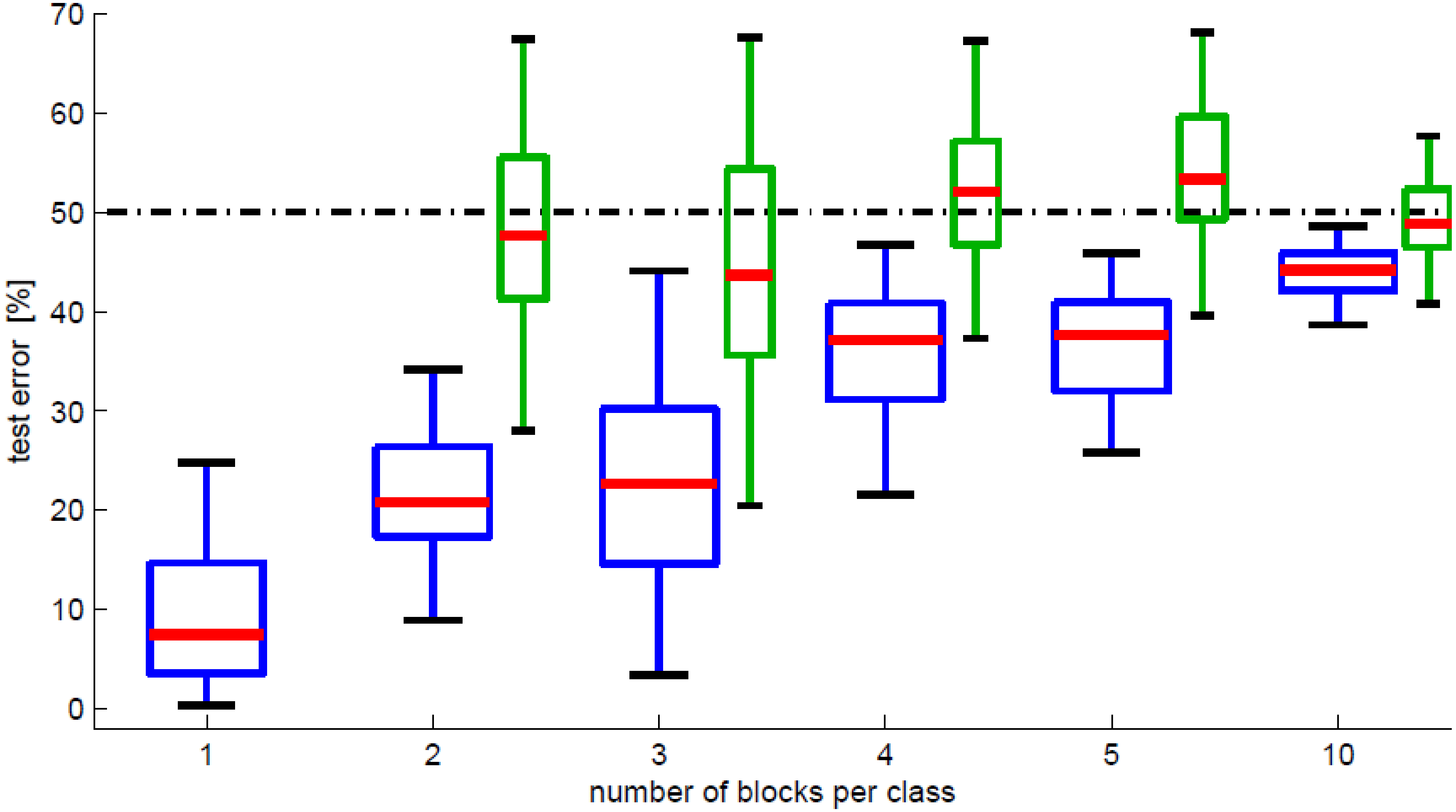
nBlocksPerClass=3:



and so on.

Results of validation test

From each block single-trials are extracted of length 1s. This procedure was performed for 80 EEG data sets. Blue boxplots show the results of cross-validation:



For comparison, results for **leave-one-block-out** validation are shown in green.

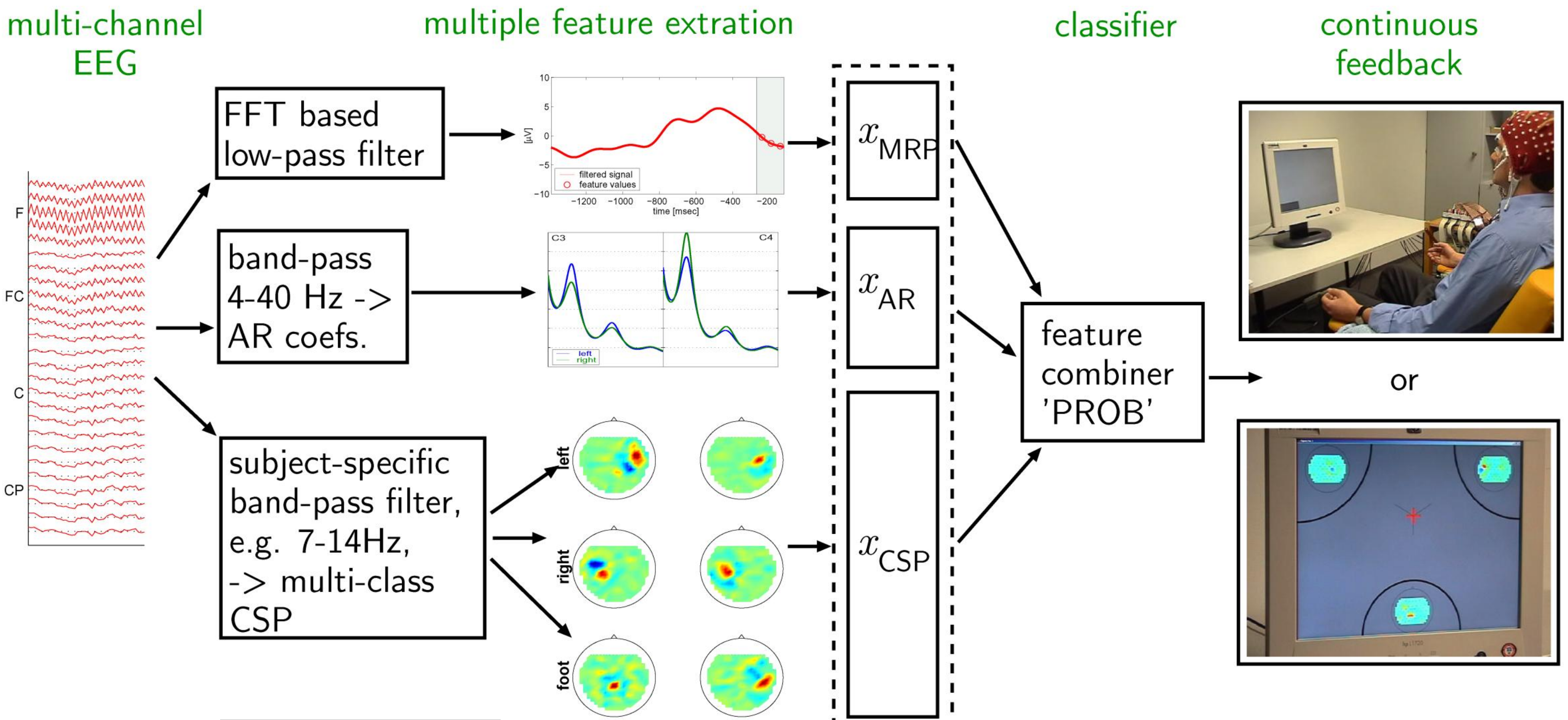
Further remarks & summary

- The severeness of the underestimation of the true error depends on the complexity of the features and the classifier.
- Cross validation in block design data might also give the correct result – but alternative evaluation is required.
- The situation gets worse if trials are extracted from overlapping segments.
- The most realistic validation is to train the methods on the first $N-1$ runs and to evaluate on the last run.
- Leave-one-block-out and leave-one-run-out have larger standard errors than cross validation.

Part II ML challenges

- Aleviating non-stationarity
- Multimodal sources

Recap: BBCI Set-up



Artifact removal

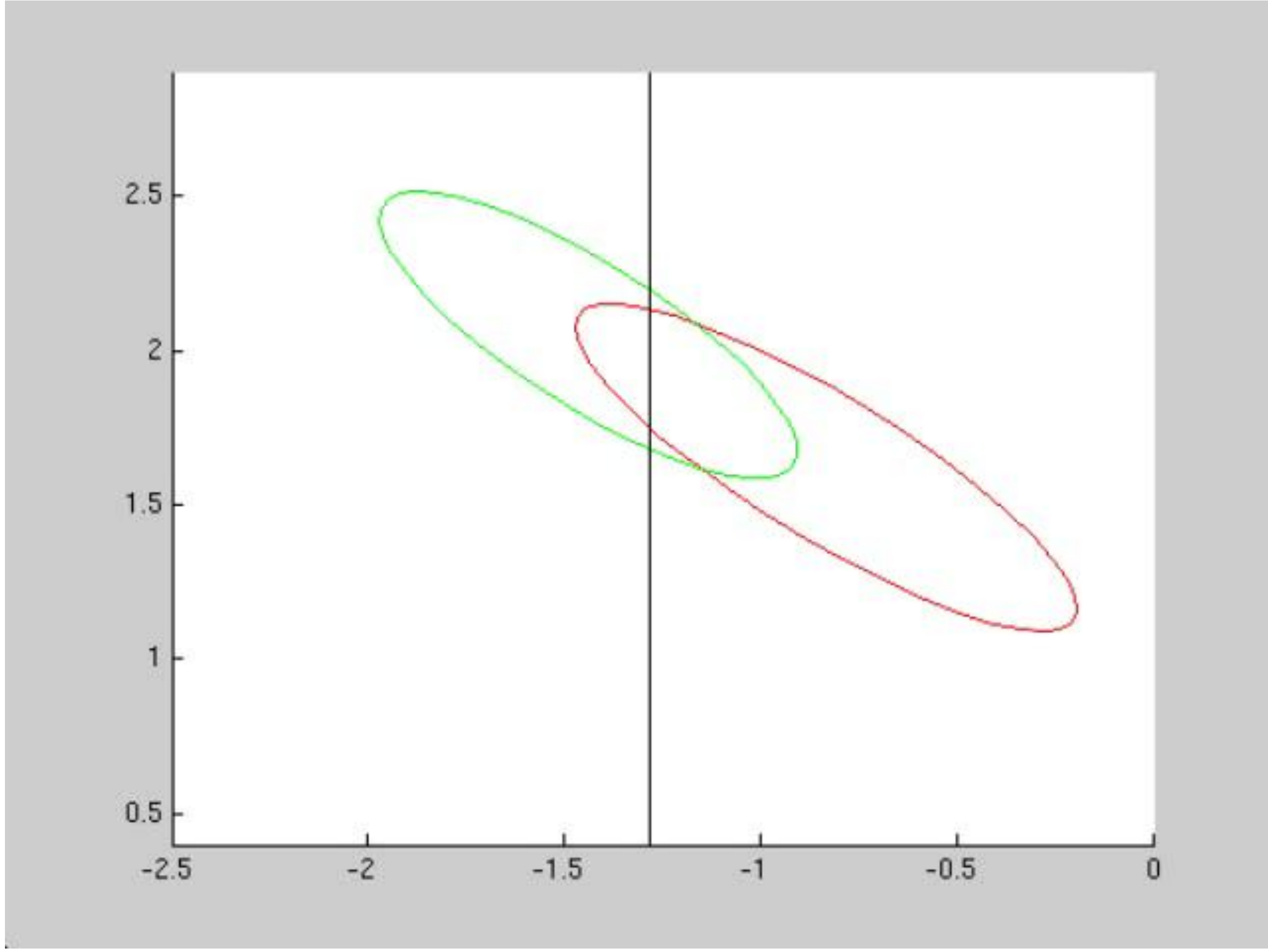
$$\min_{w,b,\xi} \ \frac{1}{2}\|w\|_2^2 + \frac{C}{K}\|\xi\|_2^2$$

$$\text{subject to} \quad y_k(w^\top x_k + b) = 1 - \xi_k \quad \text{for } k = 1, \ldots, K$$

[cf. Müller et al. 2001, 2007, 2008, Dornhege et al. 2003, 2007, Blankertz et al. 2004, 2005, 2006, 2007, 2008]

Nonstationarity in BCI

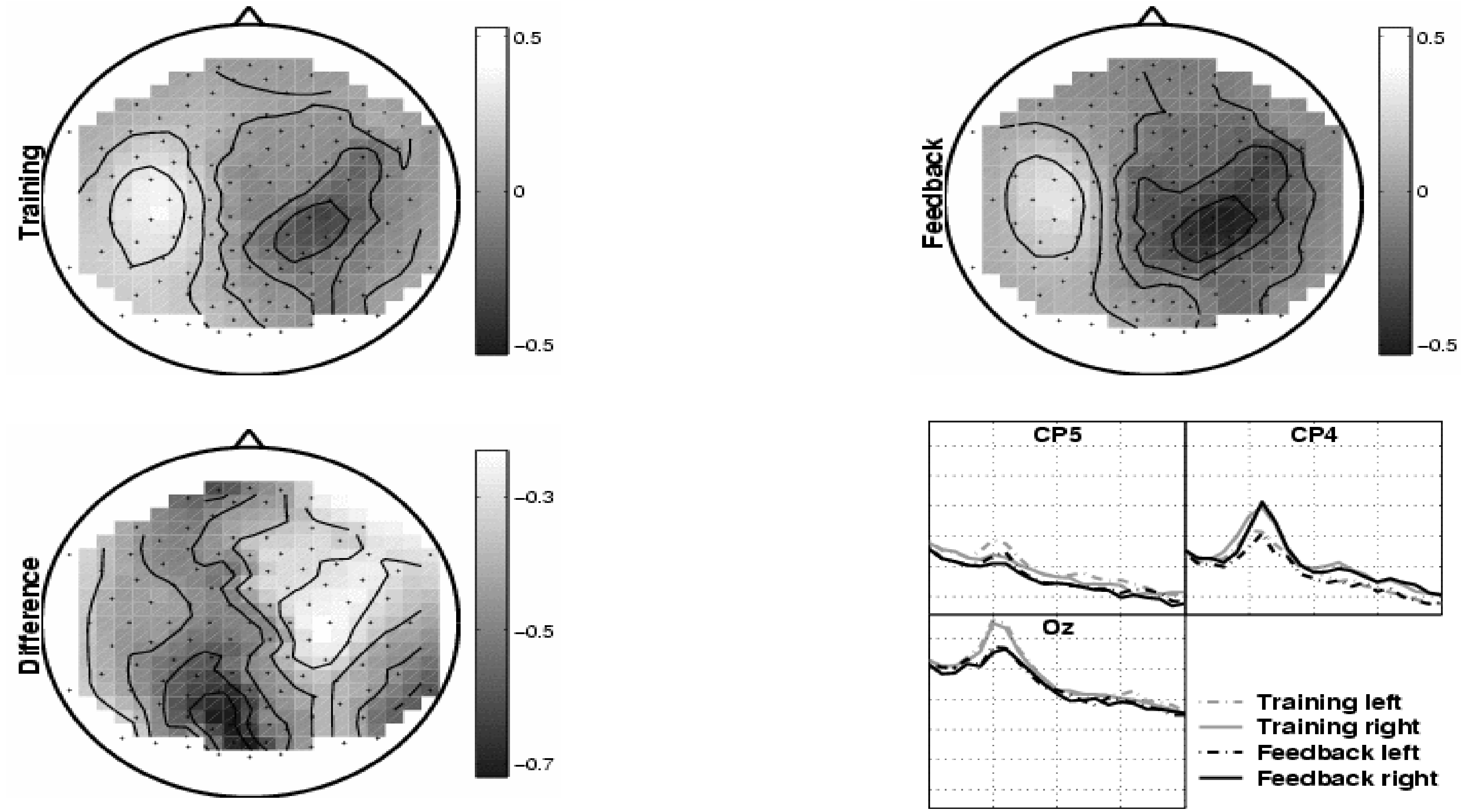
Variance IV: Shifting distributions within experiment



Mathematical flavors of non-stationarity

- Bias adaptation between training and test
- Covariate shift
- SSA: projecting to stationary subspaces
- Nonstationarity due to subject dependence: Mixed effects model
- Co-adaptation

Neurophysiological analysis



[cf. Krauledat et al. 07]

Weighted Linear Regression for covariate shift compensation

Given training samples

$$\{(\boldsymbol{x}_i, y_i) \mid y_i = f(\boldsymbol{x}_i) + \epsilon_i\}_{i=1}^{n}$$

for some function f and linearly independent basis functions $\Phi = \{\varphi_i(\boldsymbol{x})\}_{i=1}^{p}$, find

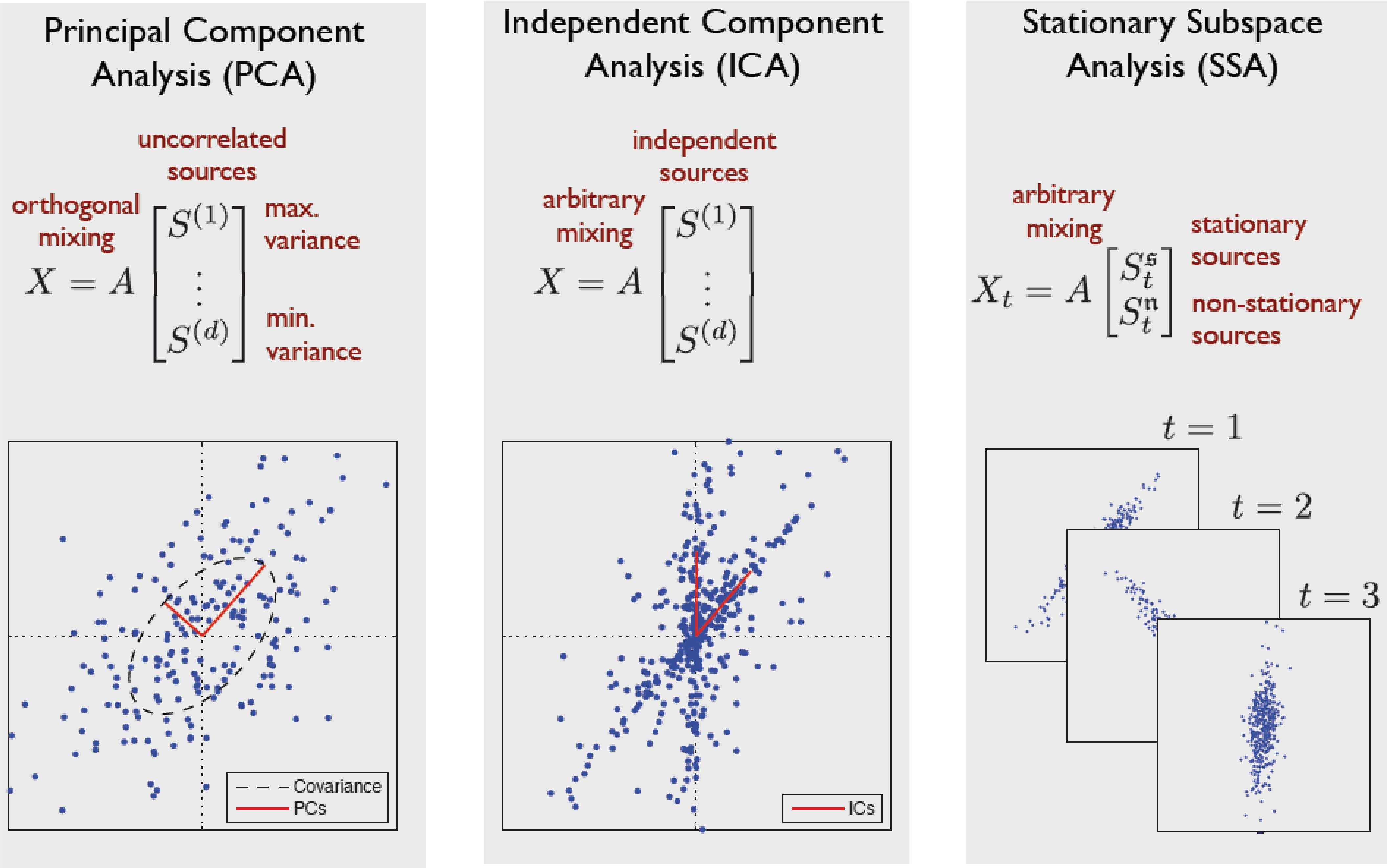
$\boldsymbol{\alpha}^* = (\alpha_1^*, \alpha_2^*, \ldots, \alpha_p^*)^\top$ which minimizes

$$\min_{\{\alpha_i\}_{i=1}^{p}} \left[\sum_{i=1}^{n} w(\boldsymbol{x}_i) \left(\hat{f}(\boldsymbol{x}_i) - y_i \right)^2 + \langle \boldsymbol{R}\boldsymbol{\alpha}, \boldsymbol{\alpha} \rangle \right].$$

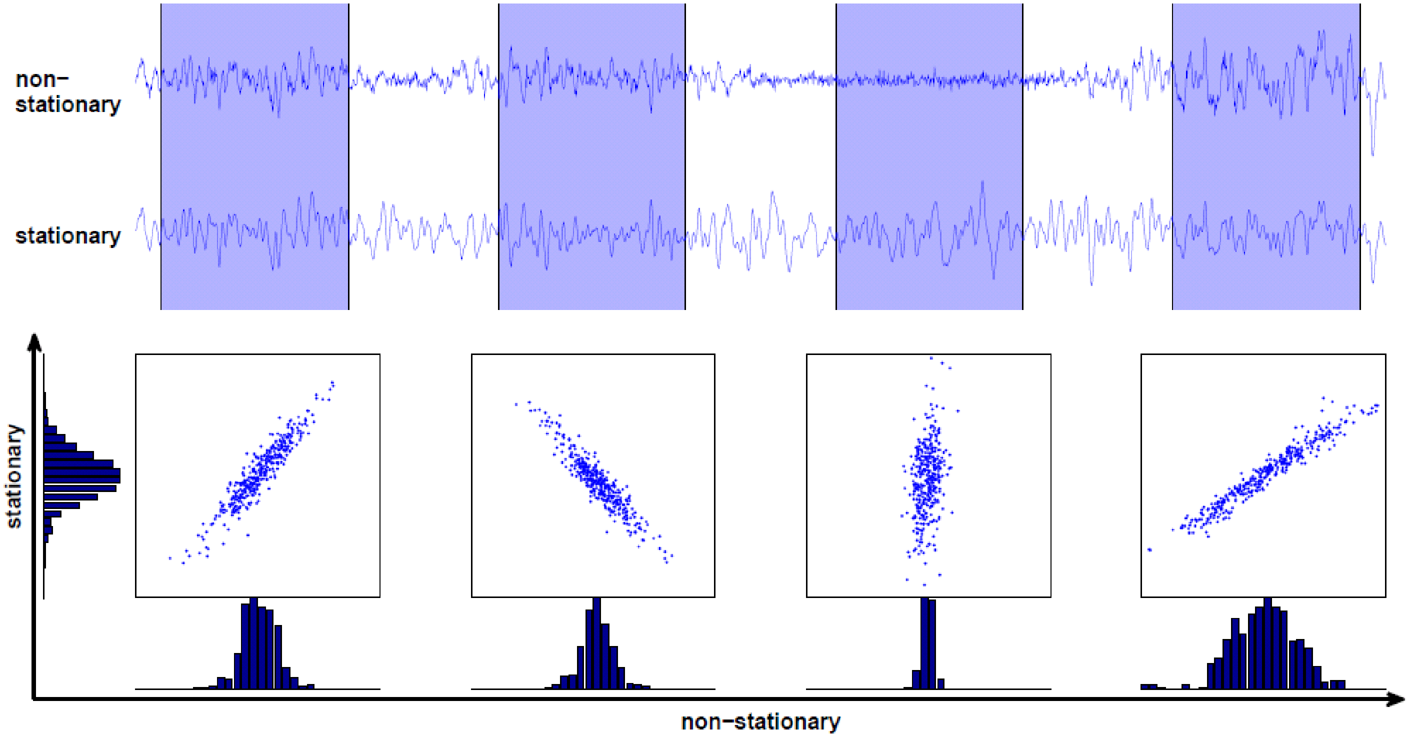
$\hat{f}(\boldsymbol{x}) = \sum_{i=1}^{p} \alpha_i \varphi_i(\boldsymbol{x})$, choosing $w(\boldsymbol{x}_i) = \frac{p_{fb}(\boldsymbol{x}_i)}{p_{tr}(\boldsymbol{x}_i)}$ yields **unbiased** estimator even under covariate shift

[cf. Sugiyama & Müller 2005, Sugiyama et al. JMLR 2007, see next week MLSS12]

Projection Methods: recap



Splitting into stationary and nonstationary subspace: SSA



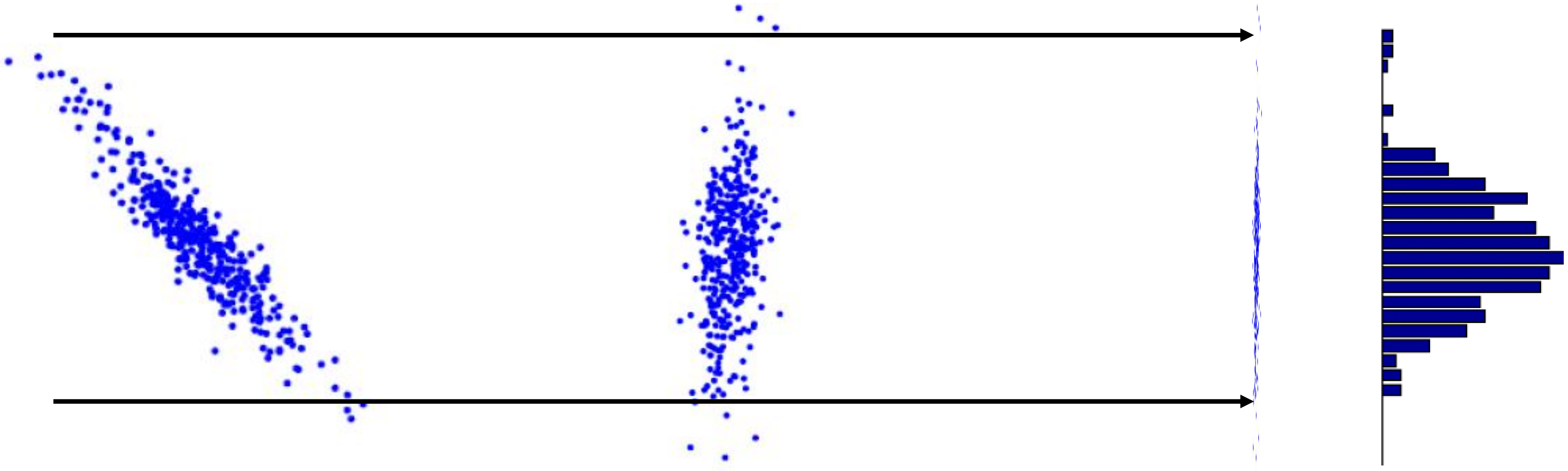
- d stationary source signals $s^{\mathfrak{s}}(t) \in \mathbb{R}^{d}$
- $D - d$ non-stationary source signals $s^{\mathfrak{n}}(t) \in \mathbb{R}^{(D-d)}$
- Observed signals: instantaneous linear superpositions of sources

$$x(t) = As(t) = \begin{bmatrix} A^{\mathfrak{s}} & A^{\mathfrak{n}} \end{bmatrix} \begin{bmatrix} s^{\mathfrak{s}}(t) \\ s^{\mathfrak{n}}(t) \end{bmatrix}$$

invert

[cf. Bünau, Meinecke, Kiraly, Müller PRL 09]

SSA



given: Epochs X_i of Data points in $\mathbb{C}^n$

wanted: Linear subspace S of $\mathbb{C}^n$ such that marginalized data sets $X_i\,|_S$ look the same

„stationary projection”

Inverting the SSA Mixing Model

Model

$$x(t) = As(t) = \begin{bmatrix} A^{\mathfrak{s}} & A^{\mathfrak{n}} \end{bmatrix} \begin{bmatrix} s^{\mathfrak{s}}(t) \\ s^{\mathfrak{n}}(t) \end{bmatrix}$$

Goal of SSA

Given only $x(t)$, find an estimate for the demixing matrix $\hat{B} = \hat{A}^{-1}$ that separates $\mathfrak{s}$-sources from $\mathfrak{n}$-sources.

$$\begin{bmatrix} \hat{s}^{\mathfrak{s}}(t) \\ \hat{s}^{\mathfrak{n}}(t) \end{bmatrix} = \hat{B}x(t) = \begin{bmatrix} \hat{B}^{\mathfrak{s}} \\ \hat{B}^{\mathfrak{n}} \end{bmatrix} x(t)$$

SSA: Algorithm idea

Stationarity in the context of SSA

A timeseries $x(t)$ is *weakly stationary*, if its mean and covariance is constant over time, i.e.

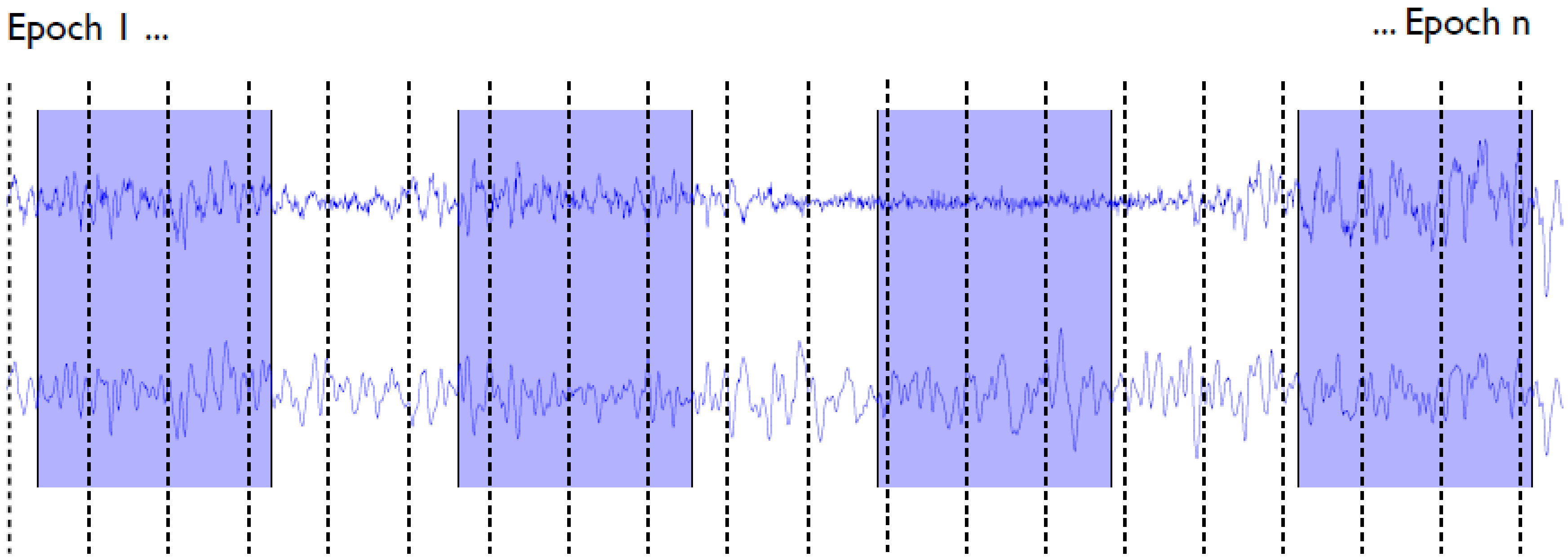
$$\mathbb{E}[x(t)] = \mathbb{E}[x(t+\tau)] \text{ and}$$
$$\mathbb{E}[x(t)^\top x(t)] = \mathbb{E}[x(t+\tau)^\top x(t+\tau)] \; \forall t, \tau.$$

Algorithmic Approach

Divide the timeseries into N epochs. Find the projection $\hat{B}^{\mathfrak{s}}$ to the stationary sources which minimizes the difference in mean and covariance between each epoch $(\hat{\mu}_i^{\mathfrak{s}}, \hat{\Sigma}_i^{\mathfrak{s}})$ and the whole dataset $(\bar{\hat{\mu}}^{\mathfrak{s}}, \bar{\hat{\Sigma}}^{\mathfrak{s}})$ for the estimated stationary sources.

Algorithm idea

Divide the data into epochs (consecutive or sliding window)



Estimate the epoch mean and covariance matrix.

μ_1, Σ_1 $\quad \dots \quad$ μ_n, Σ_n

Using Symmetries and Invariances

Symmetries

Without loss of generality, we can center and whiten the whole dataset and write the projection to the $\mathfrak{s}$-sources as

$$\hat{B}^{\mathfrak{s}} = RW$$

where $RR^{\top} = I$ is rotation matrix truncated to the first d rows and W is a whitening matrix. Thus we have set the mean and covariance of the estimated $\mathfrak{s}$-sources on the whole dataset to

$$\bar{\hat{\mu}}^{\mathfrak{s}} = 0 \text{ and } \bar{\hat{\Sigma}}^{\mathfrak{s}} = I.$$

SSA: Objective Function

Distance measure

To measure the distance between mean and covariance of two datasets we use the Kullback-Leibler divergence between Gaussians (Maximum Entropy principle).

The objective function

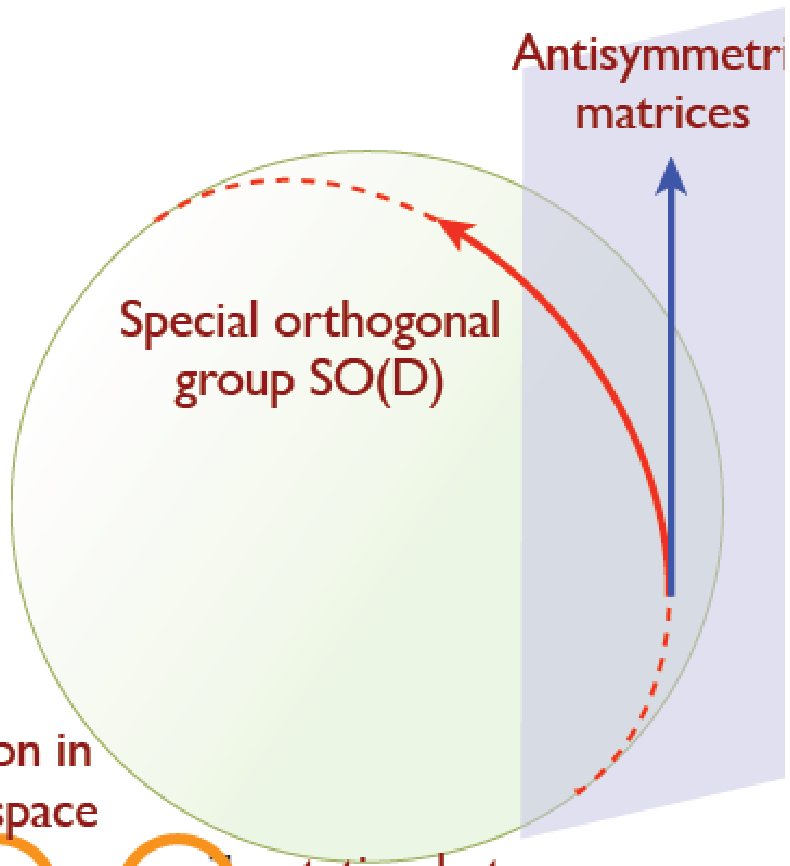
$$
\begin{aligned}
\hat{B}^{\mathfrak{s}} &= \underset{RR^\top = I}{\operatorname{argmin}} \sum_{i=1}^{N} \mathrm{KL}\left[\mathcal{N}(\hat{\mu}_i^{\mathfrak{s}}, \hat{\Sigma}_i^{\mathfrak{s}}) \;||\; \mathcal{N}(\bar{\hat{\mu}}^{\mathfrak{s}}, \bar{\hat{\Sigma}}^{\mathfrak{s}})\right] \\
&= \underset{RR^\top = I}{\operatorname{argmin}} \sum_{i=1}^{N} \mathrm{KL}\left[\mathcal{N}(\hat{\mu}_i^{\mathfrak{s}}, \hat{\Sigma}_i^{\mathfrak{s}}) \;||\; \mathcal{N}(0, I)\right] \\
&= \underset{RR^\top = I}{\operatorname{argmin}} \sum_{i=1}^{N} \left(-\log\det \hat{\Sigma}_i^{\mathfrak{s}} + {\hat{\mu}_i^{\mathfrak{s}}}^\top \hat{\mu}_i^{\mathfrak{s}}\right)
\end{aligned}
$$

Optimzing

Multiplicative update of the rotation part

$$B^{\text{new}} \leftarrow R B^{\text{old}}$$

update rotation



Parametrize the update R as the matrix exponential of an antisymmetric matrix M

$$R = \exp(M) \text{ with } M^\top = -M$$

Interpretation: M_{ij} rotation angle of axis i towards axis j

This leads to a gradient of the form:

$$\left.\frac{\partial L_{B^{\text{old}}}}{\partial M}\right|_{M=0} = \begin{bmatrix} 0 & Z \\ -Z^\top & 0 \end{bmatrix}$$

rotation in the s-space

rotation between the two spaces

rotation in the n-space

Spurious stationarity

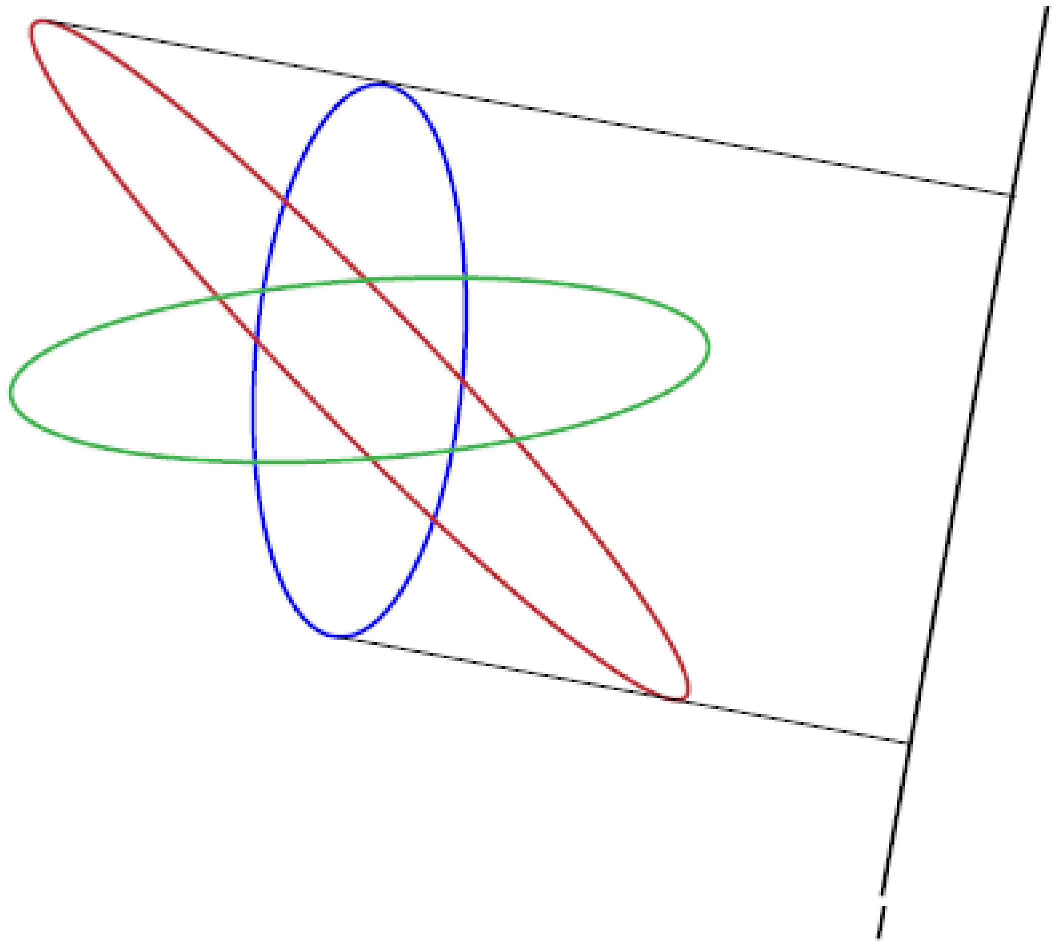
- Directions in the non-stationary space can appear stationary if we have not observed enough variation.
- The presence of *spurious stationary directions* renders the true solution unidentifiable.

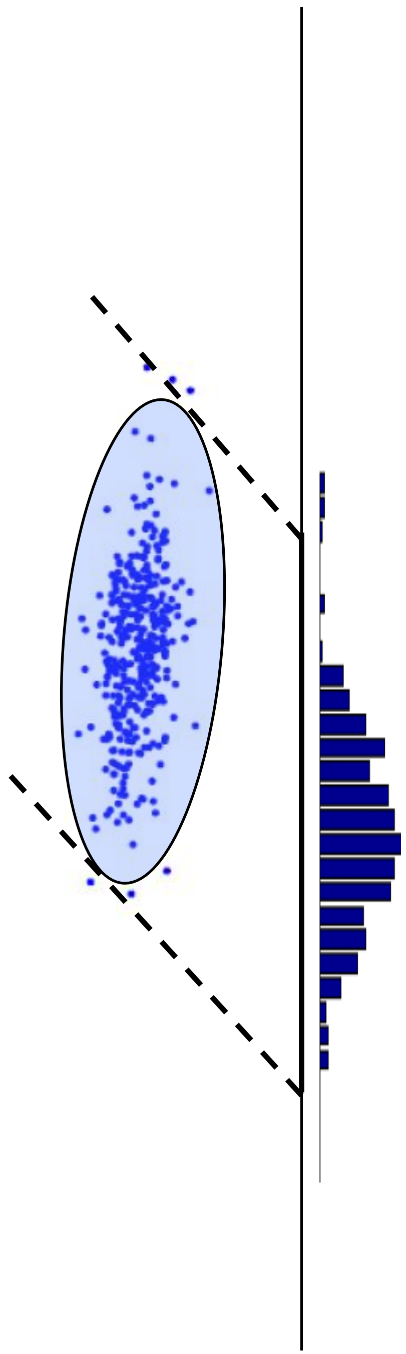
How many distinct epochs do we need to rule out spurious stationary directions?

SSA: how many epochs?

Estimate Epochs X_i by Gaussians $\mathcal{N}(\mu_i, \Sigma_i)$

Marginalized Gaussians are $\mathcal{N}(P_S^T \mu_i, P_S^T \Sigma_i P_S)$





How many epochs? Theoretical results

Theorem (Identifiability of SSA)

- *If the non-stationarity is expressed in* **both** *mean and covariances, the stationary subspace can be uniquely identified if*

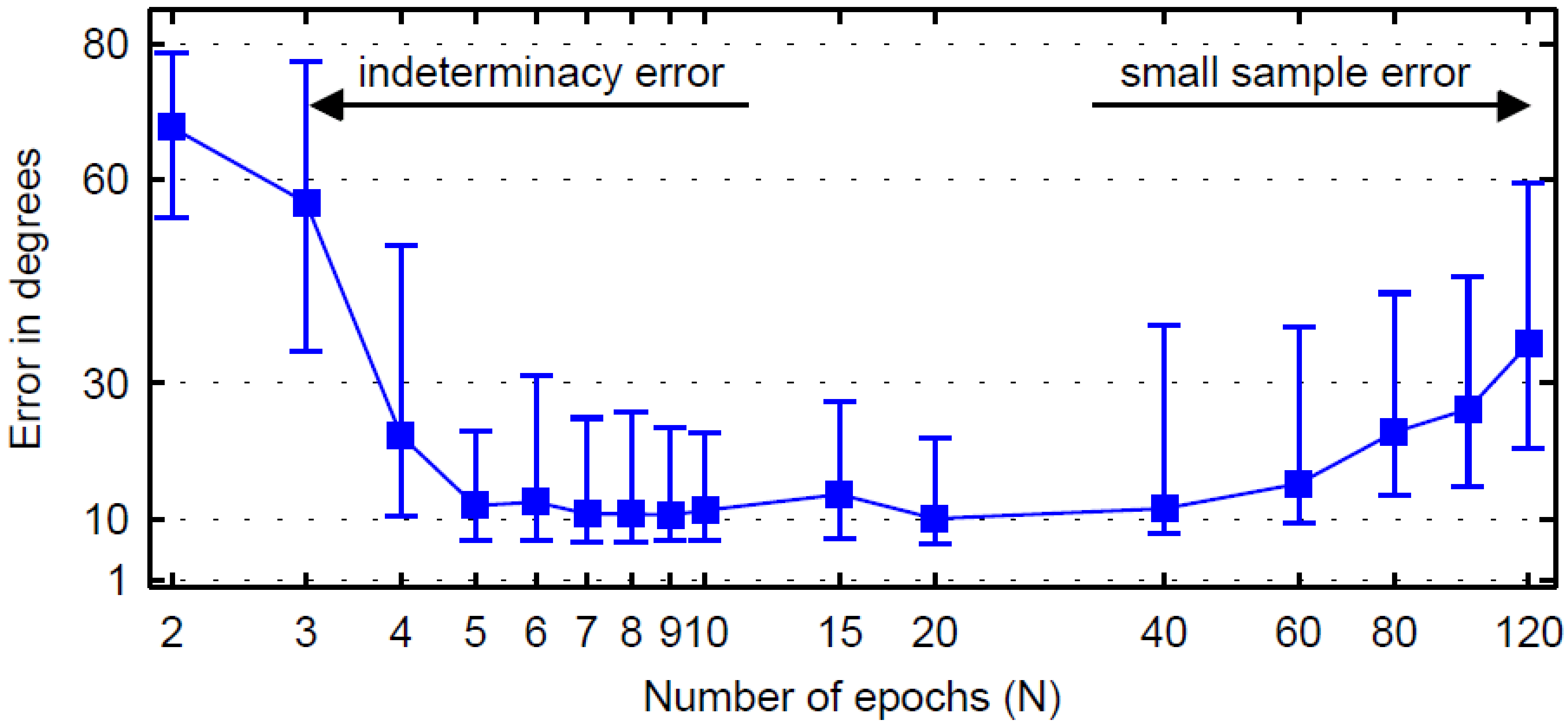
$$N > \frac{D - d}{2} + 2.$$

- *If the non-stationarity is* **only** *expressed in either mean* **or** *covariances, Identifiability is guaranteed for*

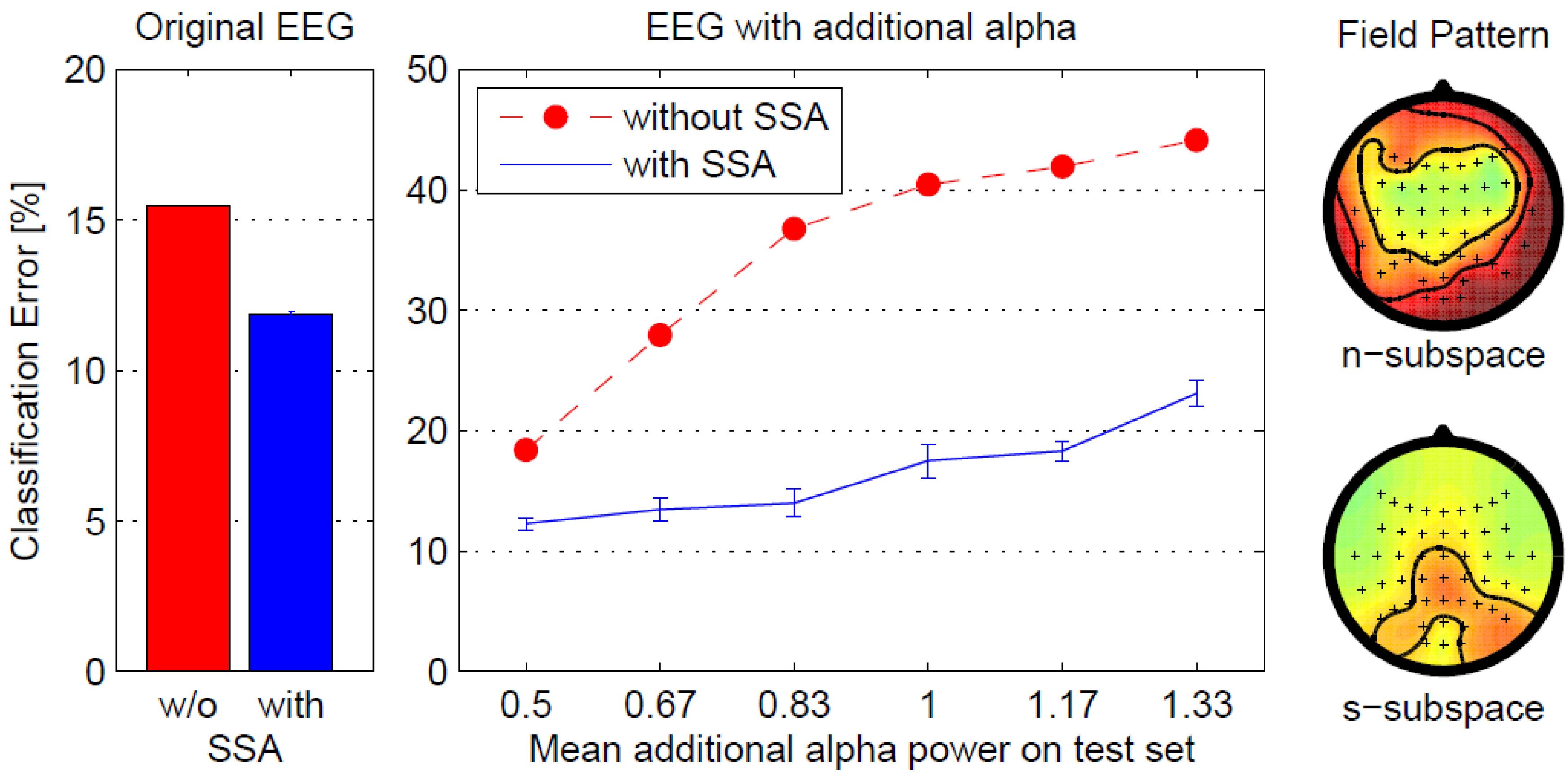
$$N > D - d + 1.$$

Simulations: toy data

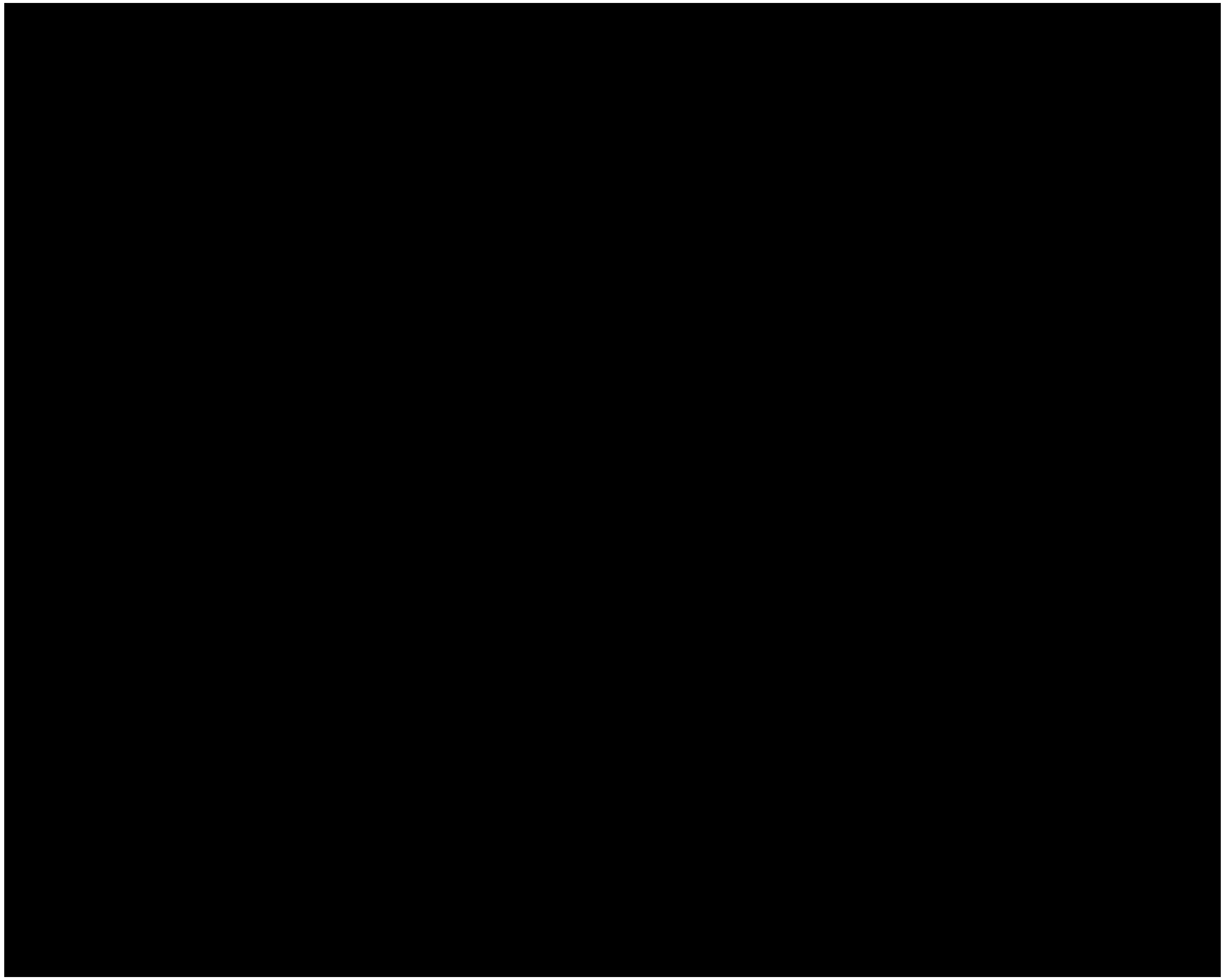
- 4 $\mathfrak{s}$-sources and 4 $\mathfrak{n}$-sources
- Fixed number of samples divided into epochs



Application to Brain-Computer-Interfacing

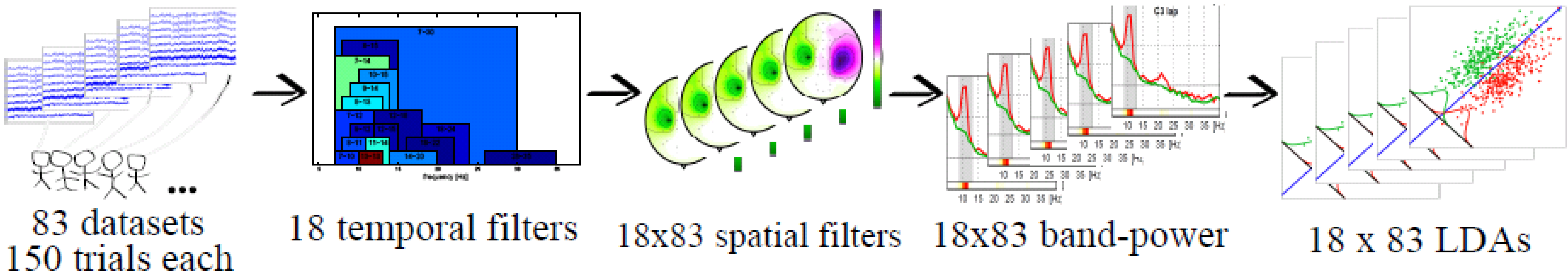


Real Man Machine Interaction



Towards a subject independent BCI decoder

- we end up with **1494 features** and $83 \cdot 150 =$ **12450 trials**
- to find a **subject-independent BCI**, we can perform ℓ_1-regularized regression (or others like LMM) using **leave-one-subject-out cross-validation**
- note that our trials have a **grouping** structure



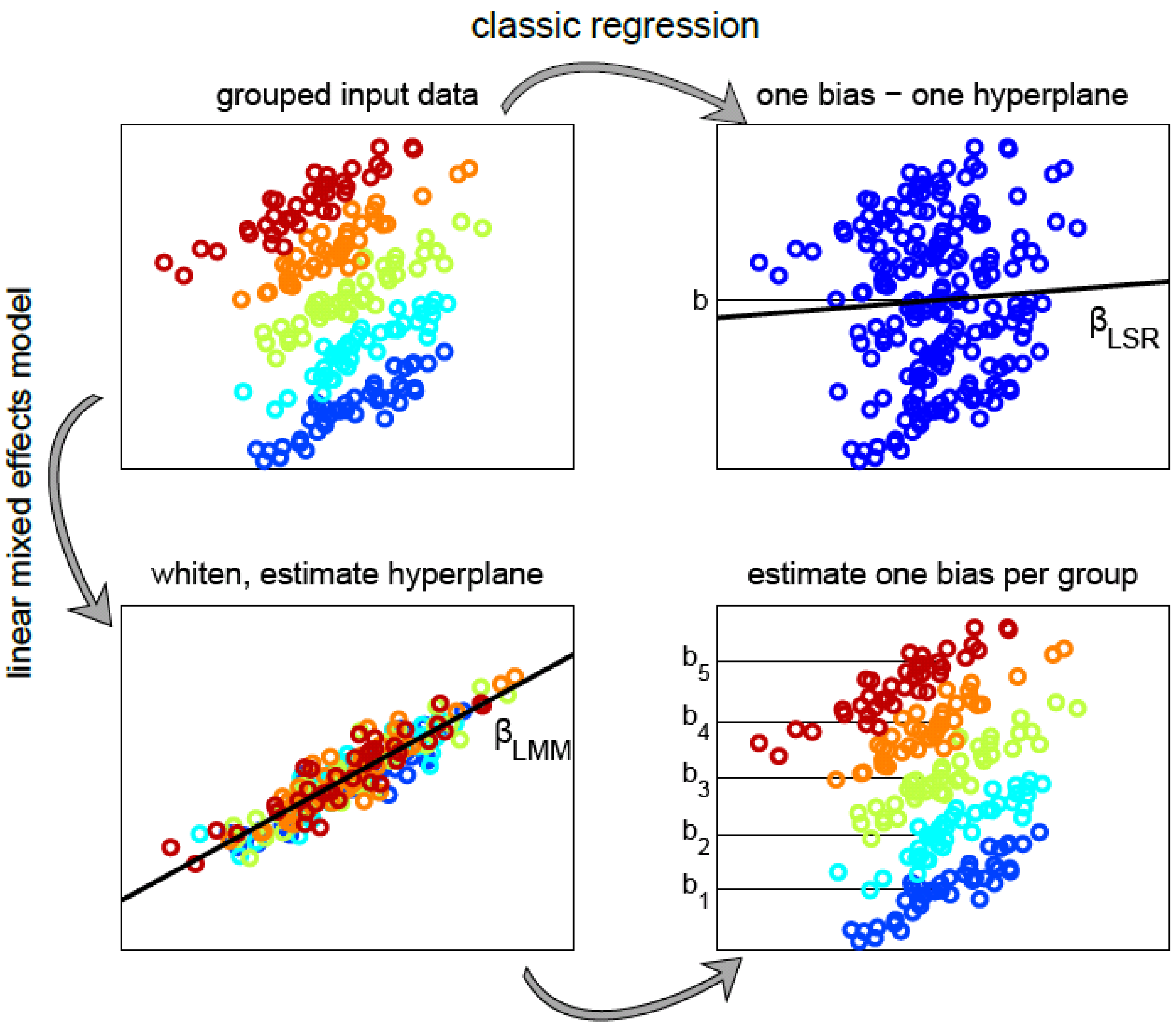
Model formulation

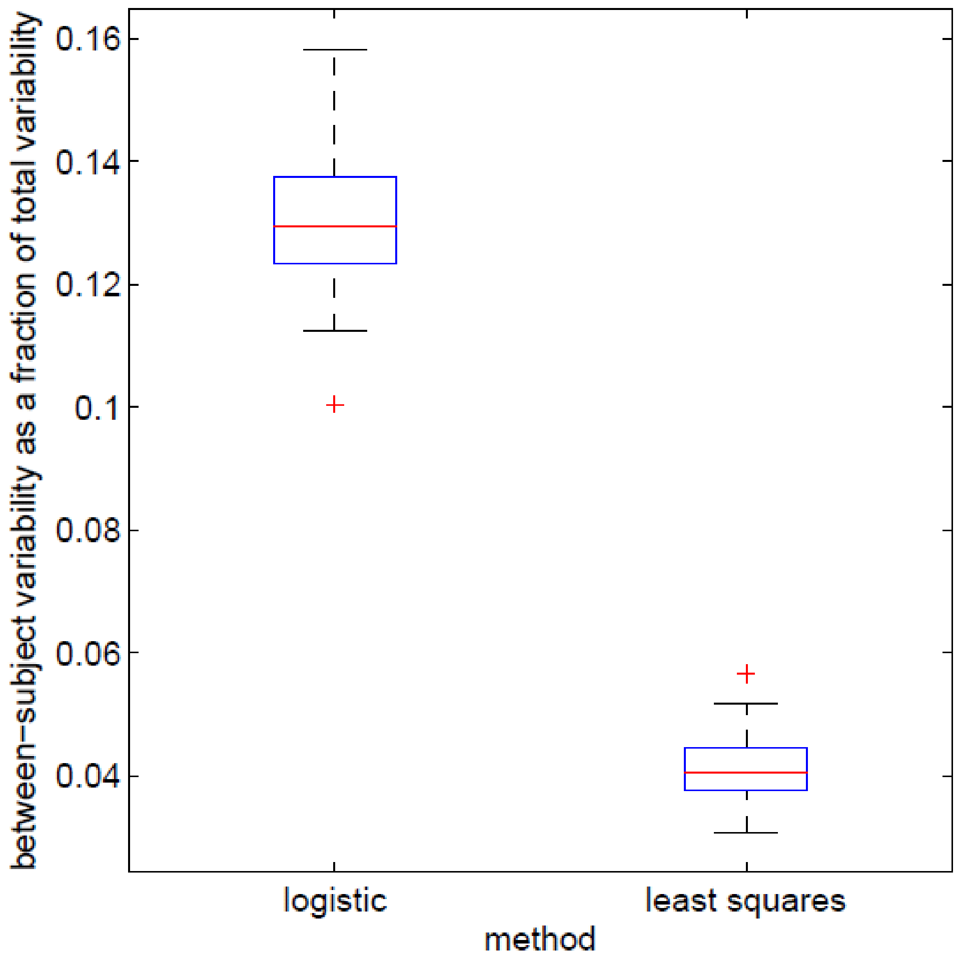
- Reminder – Linear regression:
  - $\mathbf{y} = \mathbf{X}\beta + \varepsilon$
- Mixed effects model with n groups:
  - $\mathbf{y}_i = \mathbf{X}\beta + \mathbf{Z}_i\mathbf{b}_i + \varepsilon_i \quad \forall i \in \{1 \dots n\}$
  - Consists of n simultaneous equations, one for each group
  - The equations are coupled by the common term $\mathbf{X}\beta$
  - Each equation has a group-dependent term $\mathbf{Z}_i\mathbf{b}_i$
  - In our case, each $\mathbf{Z}_i$ is simply a vector of ones, i.e. the corresponding $\mathbf{b}_i$ is scalar and represents the bias of group i
  - So-called **random intercepts model**

$$b_i \sim \mathcal{N}_q(0, \tau^2 I_q)$$
$$\varepsilon_i \sim \mathcal{N}_{n_i}(0, \sigma^2 I_{n_i})$$

- Since we expect our features to be redundant and are aiming for better interpretability, we enforce sparsity by adding an ℓ_1 penalty

Linear Mixed Effects Model: intuition





[Fazli, Müller et al. 2011]

Approach to „Cure“ BCI Illiteracy

Runs 1 - 3 | fixed Laplace | supervised

- Direct feedback -> Unspecific LDA classifier.
- Each trial, perform adaptation of the cls.
- Features: log band power (alpha and beta).
- Laplacian channels C3, C4 and Cz.

Runs 4 - 6 | CSP + sel.Lap. | supervised

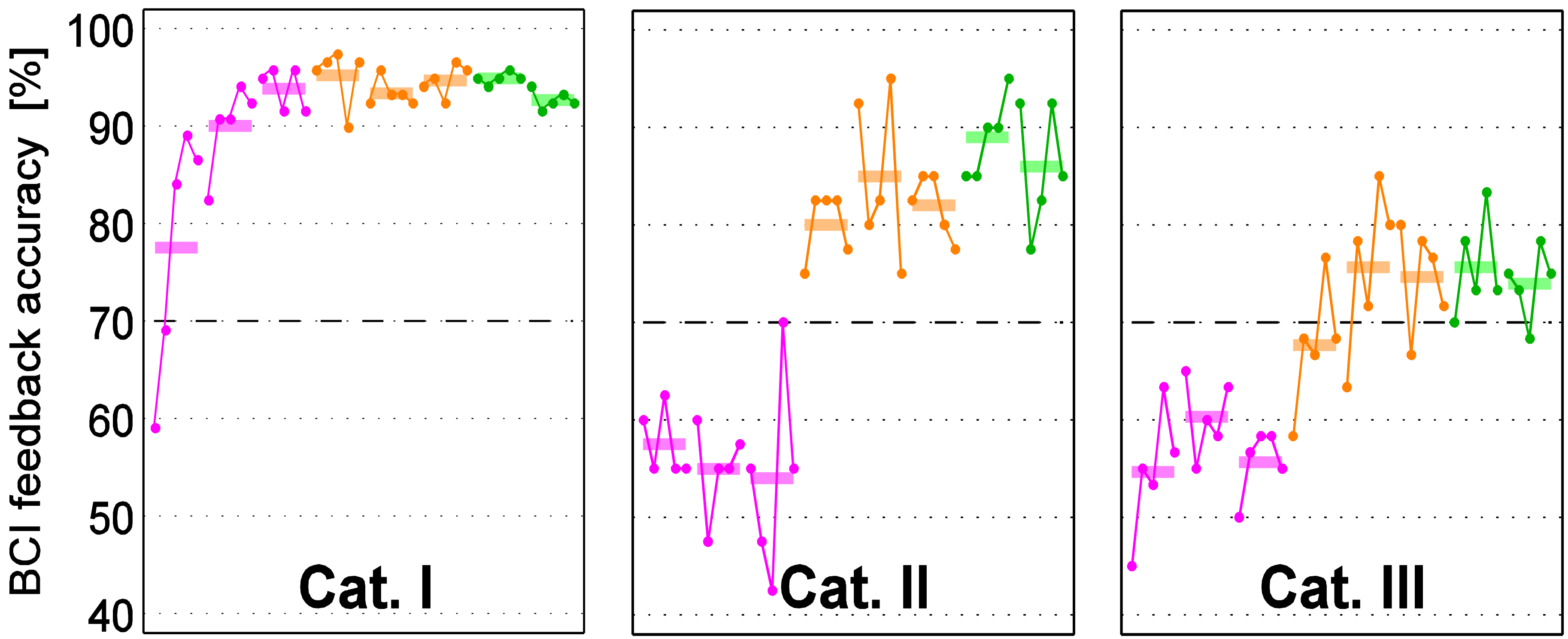
- Compute CSP and sel. Laps. from runs 1-3.
- Fixed CSP filters, automated laps. selection.
- Each trial retrain the classifier.

Runs 7 - 8 | CSP | unsupervised

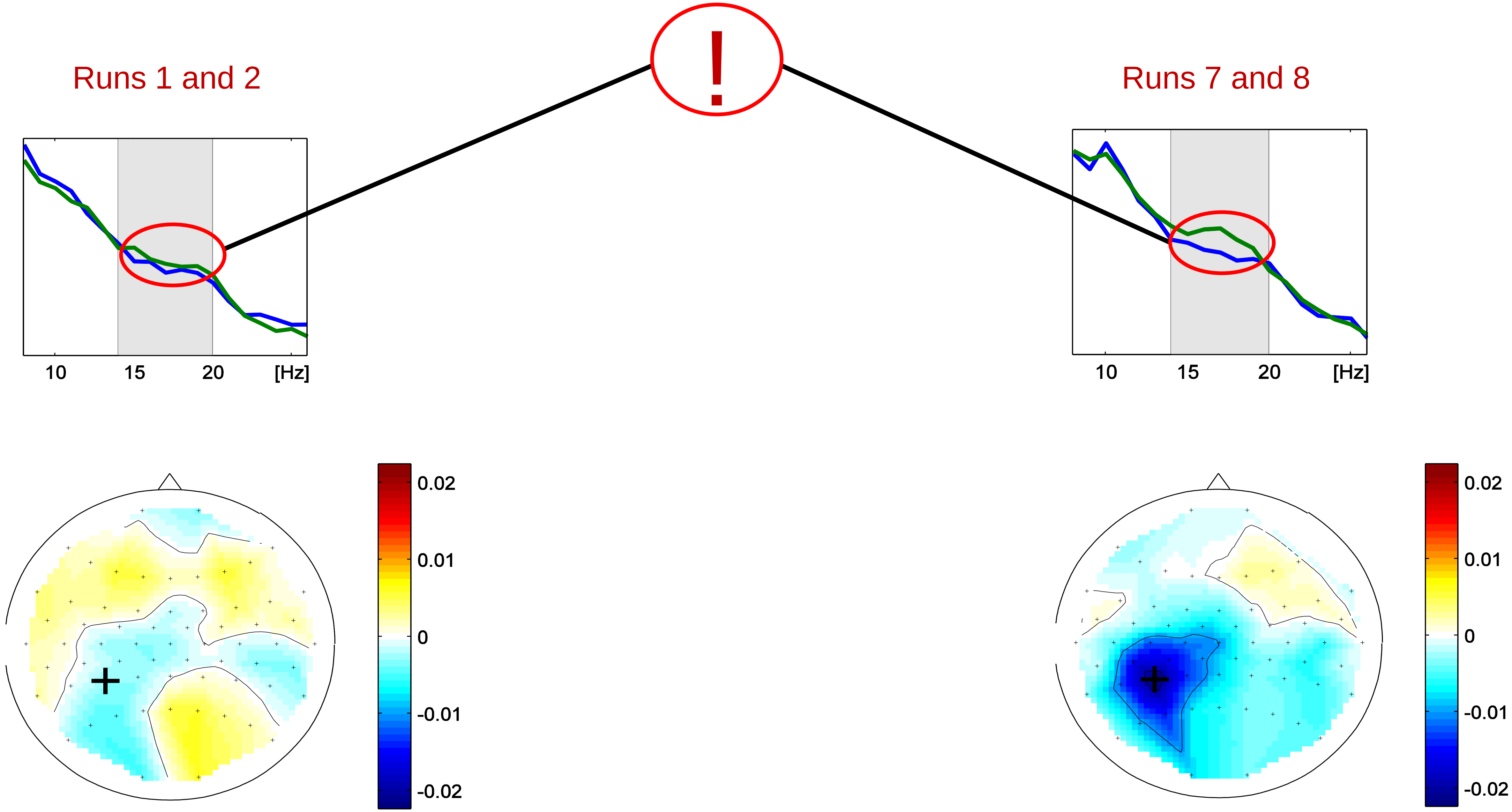
- Compute CSP from runs 4-6.
- Perform unsupervised adaptation of pooled mean.
- Update the bias of the classifier.

[cf. **Vidaurre**, Blankertz, Müller et al. Neural Comp. to appear]

Results (Grand Averages)



Example: one subject of Cat. III

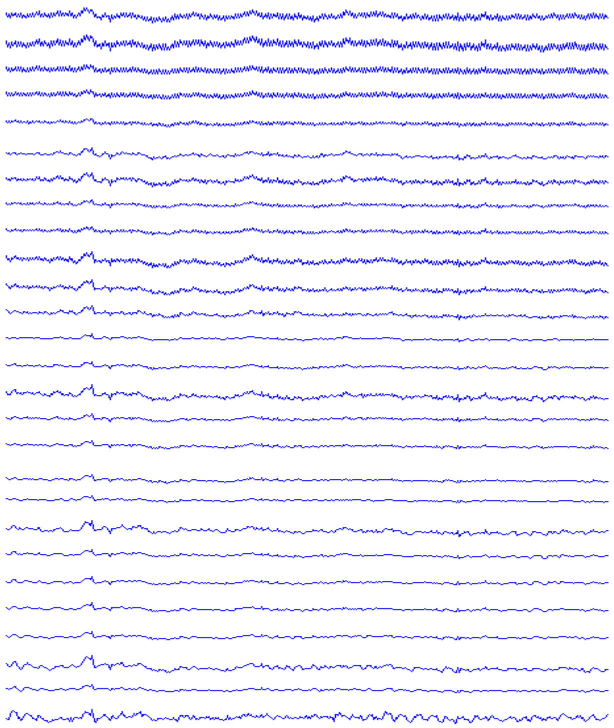


[cf. Vidaurre, Blankertz, Müller et al. 2009]

Multimodal

Different physiological Features

EEG signals



- Slow Features, e.g.
- Event Related Potential/Slow Cortical Potentials (ERP/SCP)

Independent???

Neurophysiology: YES

Maps

- Oscillatory Features, e.g.
- Event Related Desynchronization/ Synchronization (ERD/ERS)

…

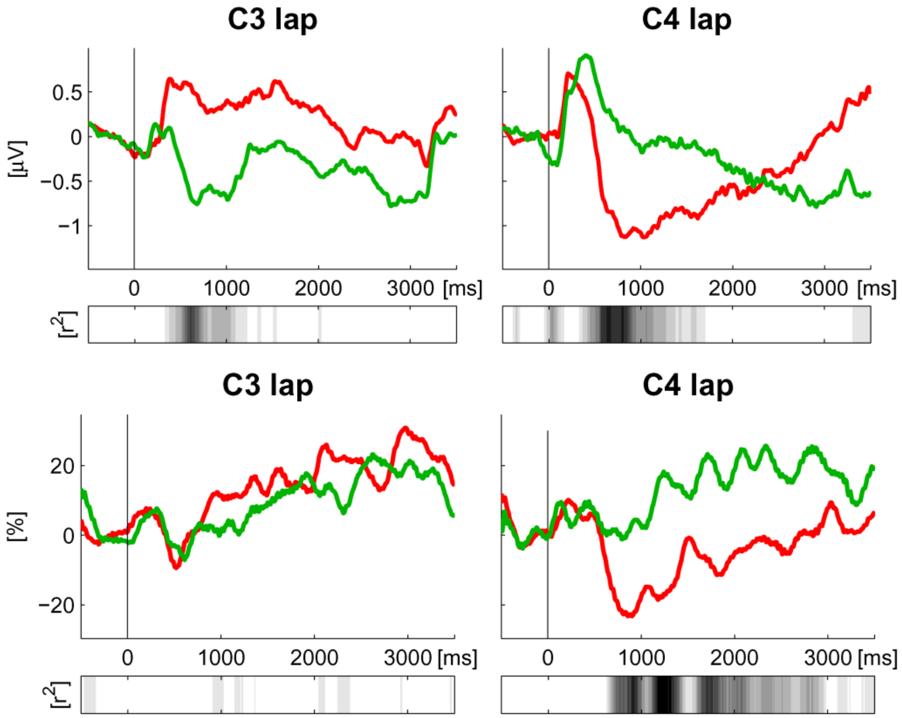
[Dornhege, et al. 2006]

Different physiological Features

Some mental activities or states are reflected by different neurophysiological features.
Motor related brain activity (actual movement, imagery, intentions) is reflected by

Lateralized Readiness Potential (LRP)

Event-Related Desynchronization (ERD)



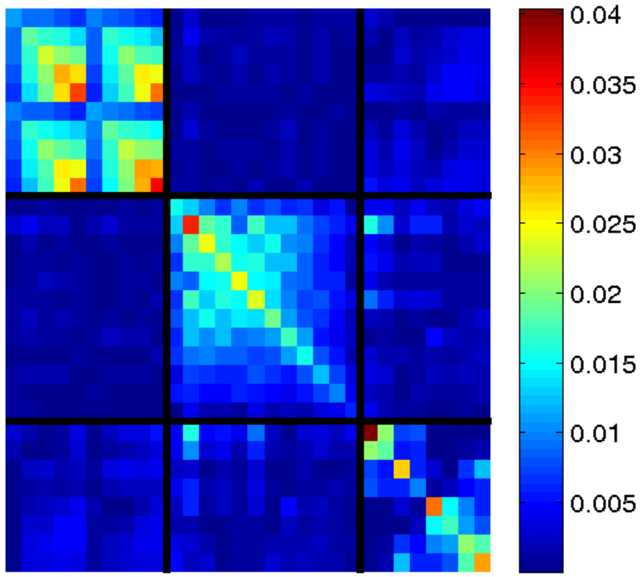
➤ **early** distinction between the signals of **left** and **right** trials.

➤ **long persisting** distinction between the signals of **left** and **right** trials.

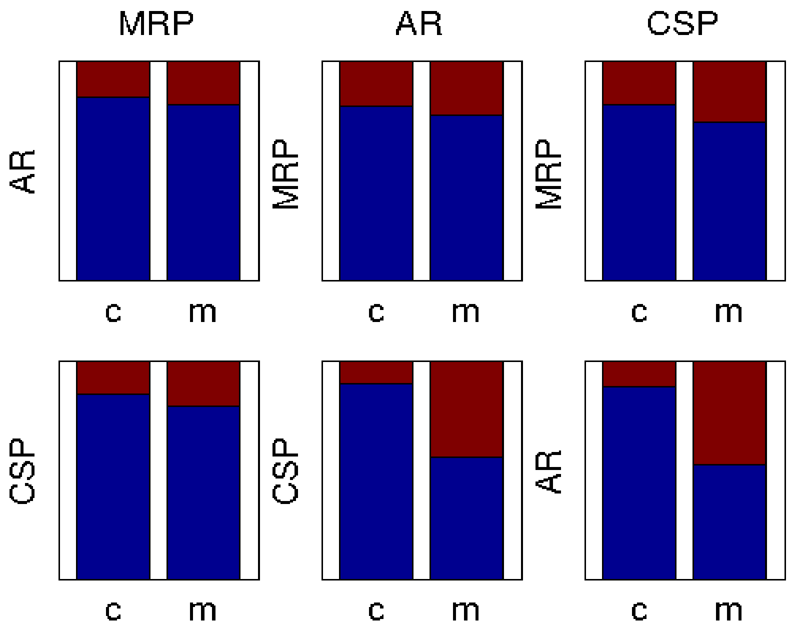
➤ As seen from the time courses, the LRP and the ERD seem to reflect *independent* cortical processes.

Independent Features

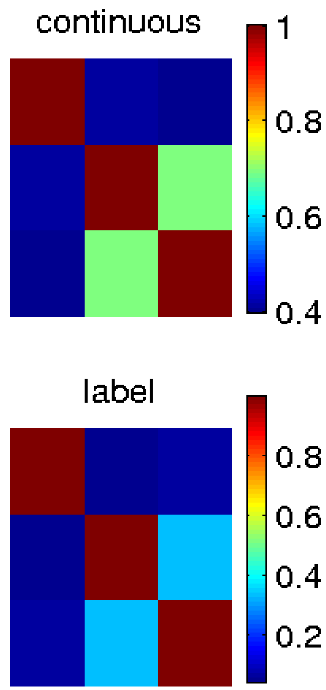
Covariance matrix between features



Distribution of misclassified and classified trials for different features (loo)



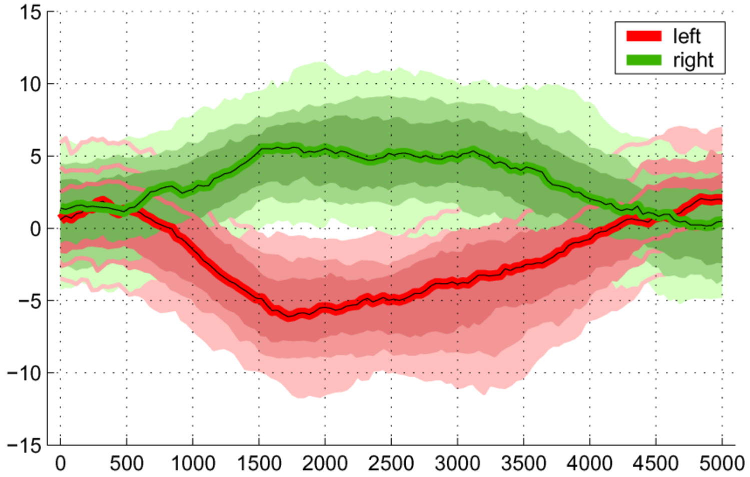
Correlation of classifier output (continuous/ label)



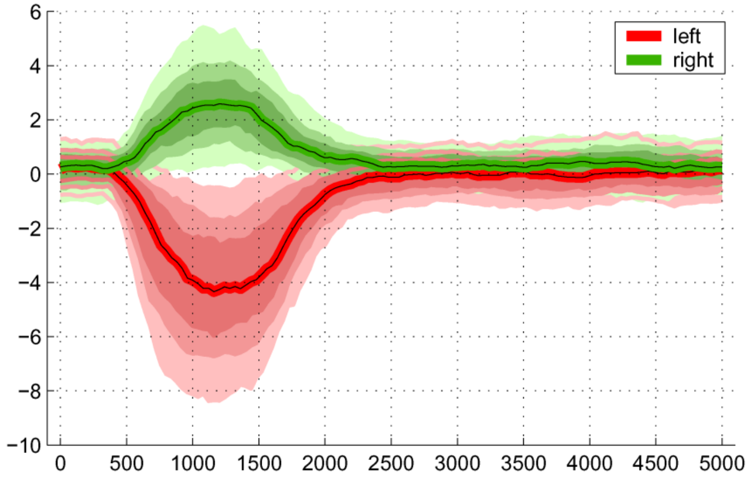
from left to right, top to bottom: MRP, AR,CSP

Combination Results

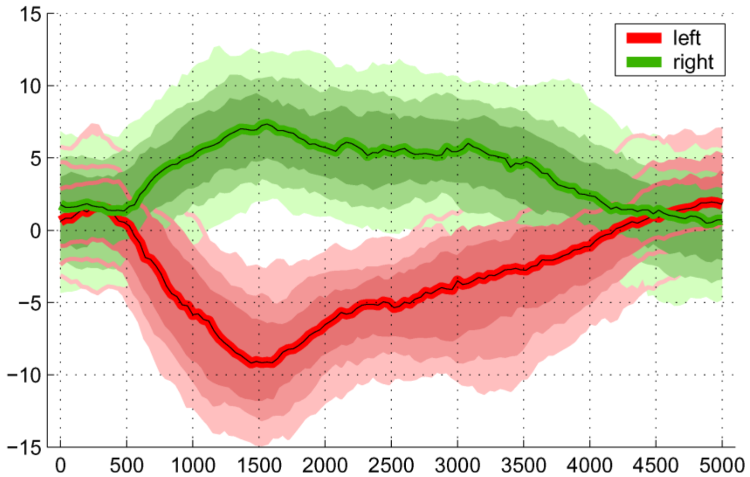
based on ERD features



based on LRP

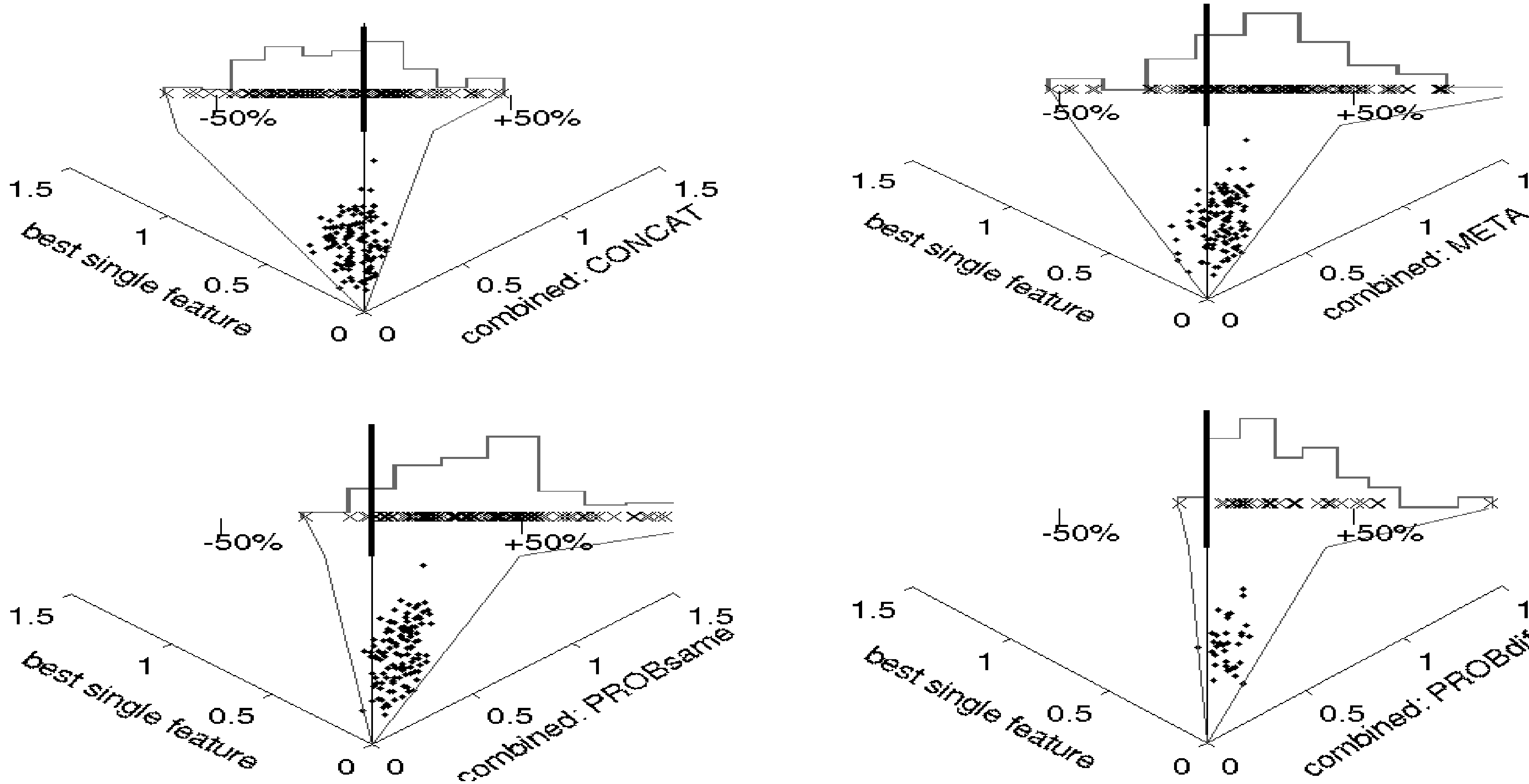


based on combination of the two



The combination of ERD and LRP features exploits the merits of the two: **rapid response** of LRP features and the **persistence** of ERD features.

Combination Results

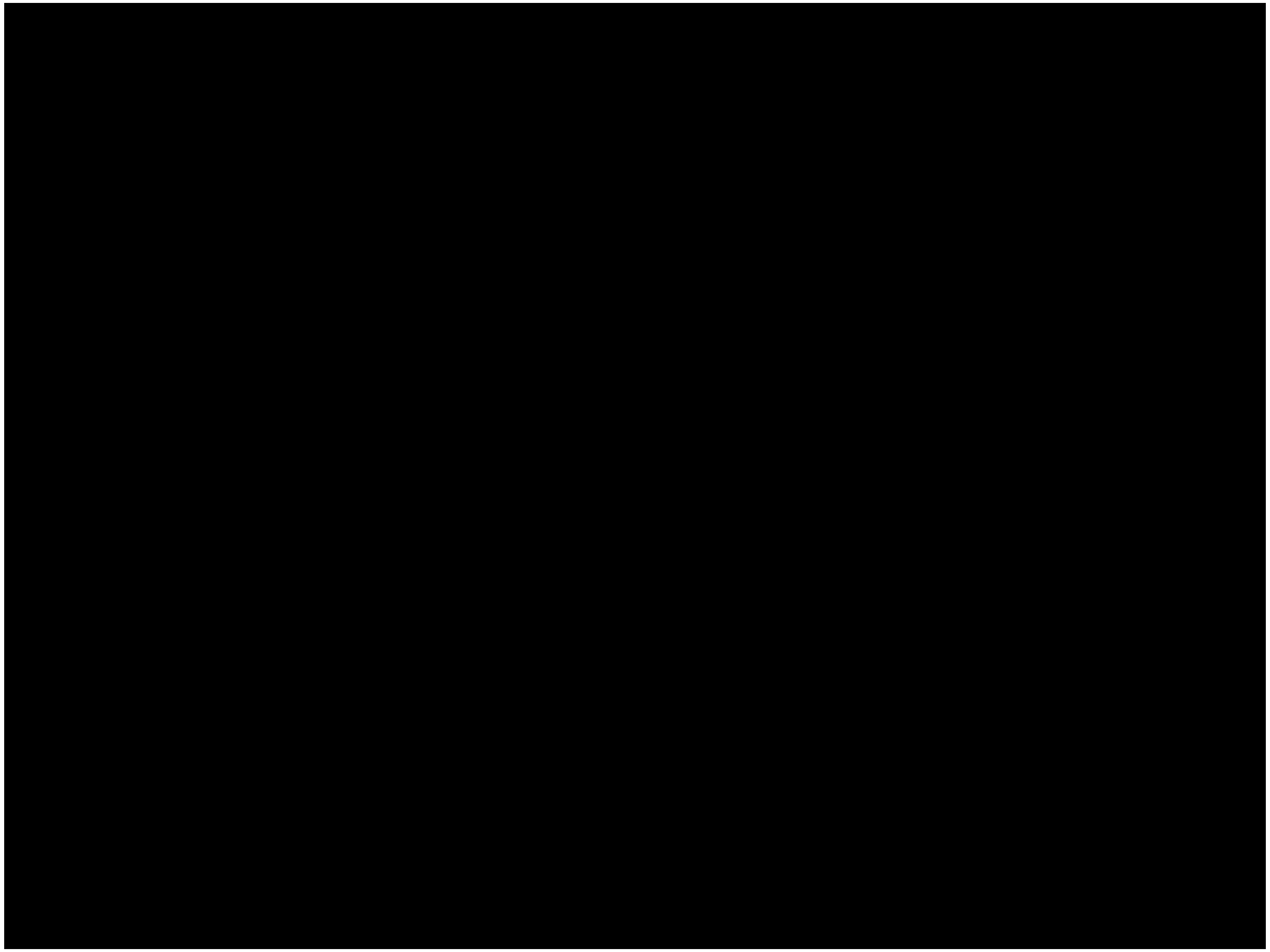


The figures show the Information Transfer Rate per decision for the best single feature compared to the suggested algorithms on all subset of classes out of the experiments we have done. Above each figure a histogram is plotted. For points right of the middle line the suggested algorithm outperforms the best single feature performance.

Example: NIRS-EEG Brain Computer Interfaces

[Fazli et al. Neuroimage 2012]

Photon Transport in the Human Brain Tissue



- Near-Infrared light can penetrate the brain
- ‚banana-shaped‘ measurement volume for non-invasive NIRS

Experimental Setup and Paradigm

EEG: 37 electrodes

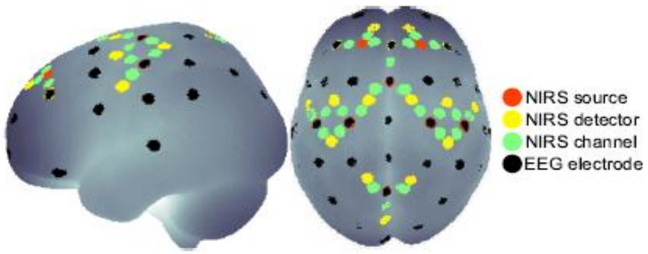
NIRS 26 channels (frontal, parietal, occipital)

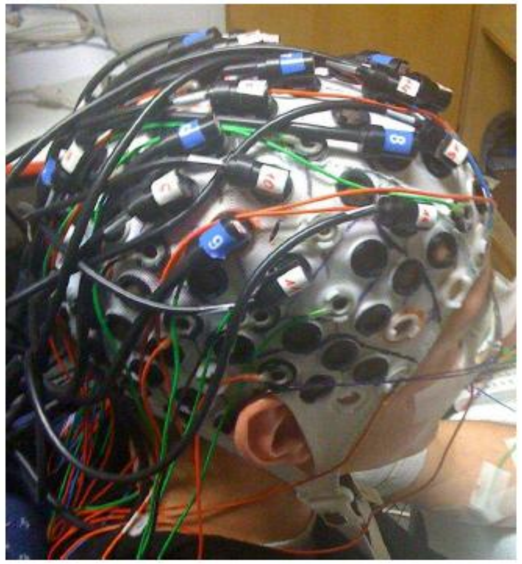
EEG-based cursor feedback (ISI = 15 s)

Executed movement vs imagery movements

Imagery movements: EEG-feedback for left and right motor imagery

Number of subjects: 14





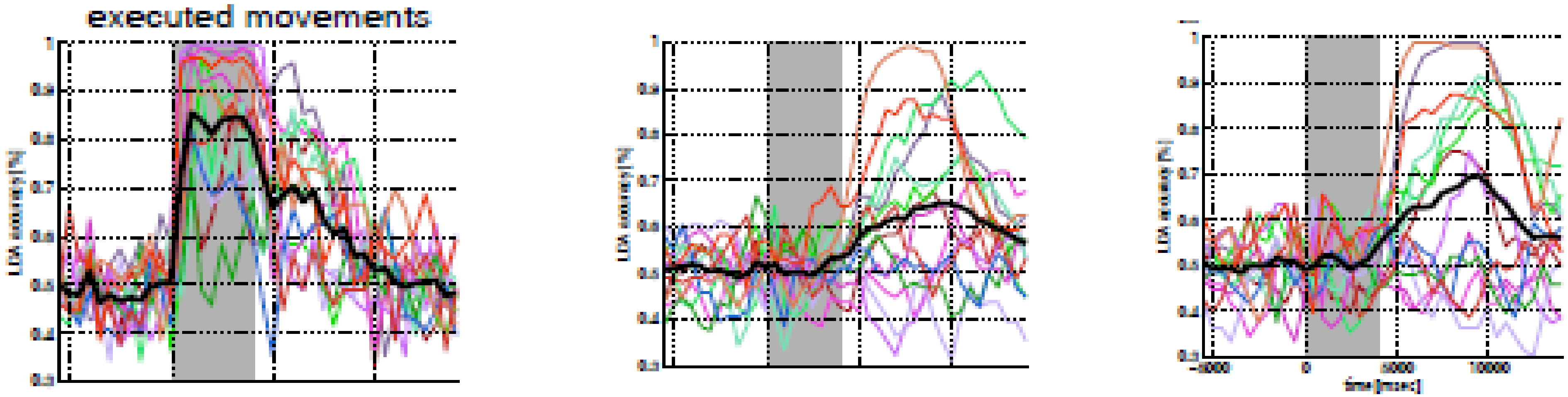
Can a simultaneous measurement of NIRS and EEG during Brain Computer Interfacing enhance the classification accuracy?

Are the results physiologically reliable?

Fazli et al. 2012

Temporal Dependency of Classification in Executed Movements

EEG HbO HbR



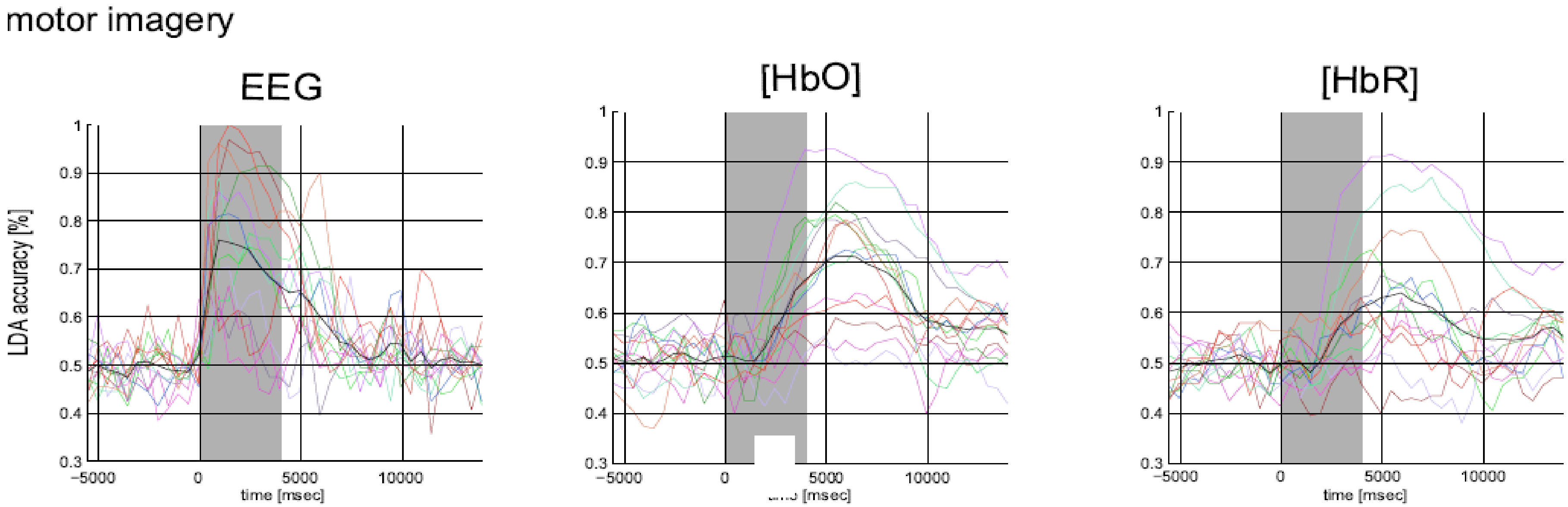
Fazli et al. 2012

EEG peaks earlier as compared to HbO and HbR

Physiological reliability: HRF shaped classification accuracies over time

Classification accuracy higher for EEG

Temporal Dependency of Classification in Motor Imagery



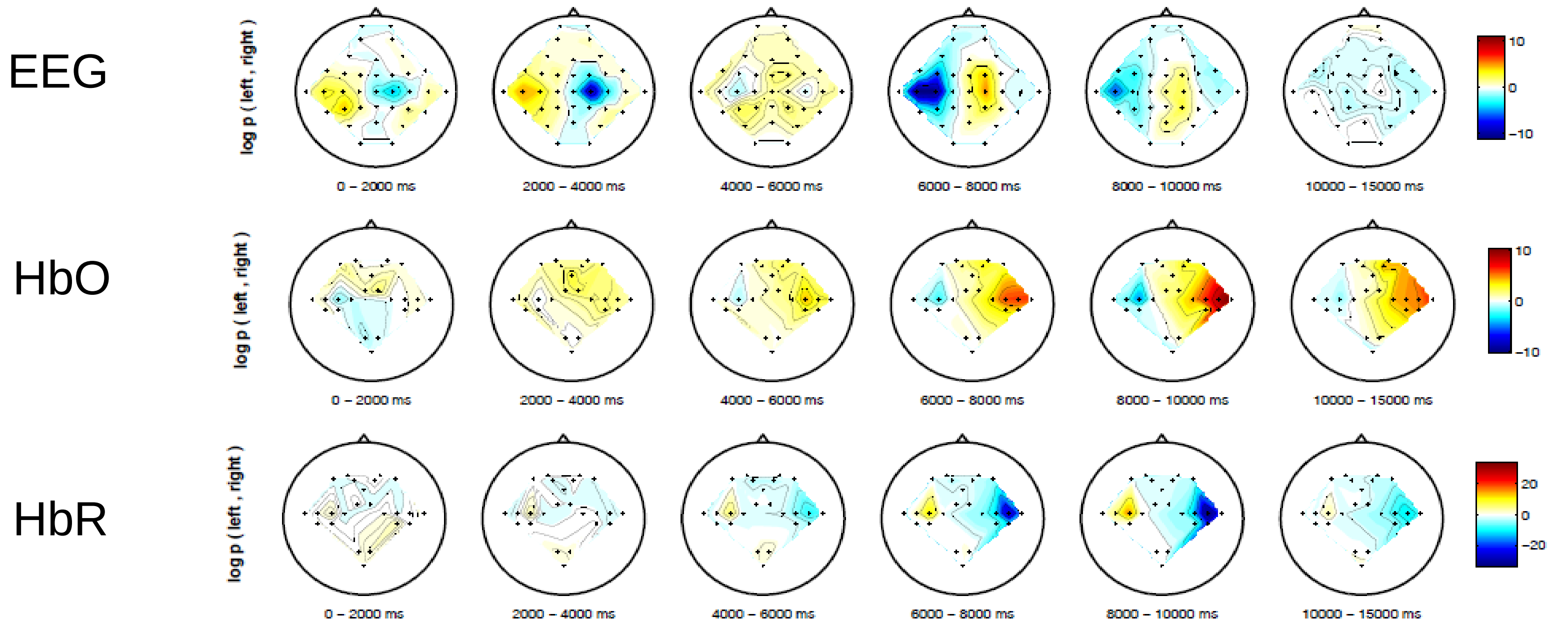
EEG peaks earlier as compared to HbO and HbR

Physiological reliability: HRF shaped classification accuracies over time

Classification accuracy higher for EEG

Classification accuracy lower than in executed movements

Topography for Executed Movements

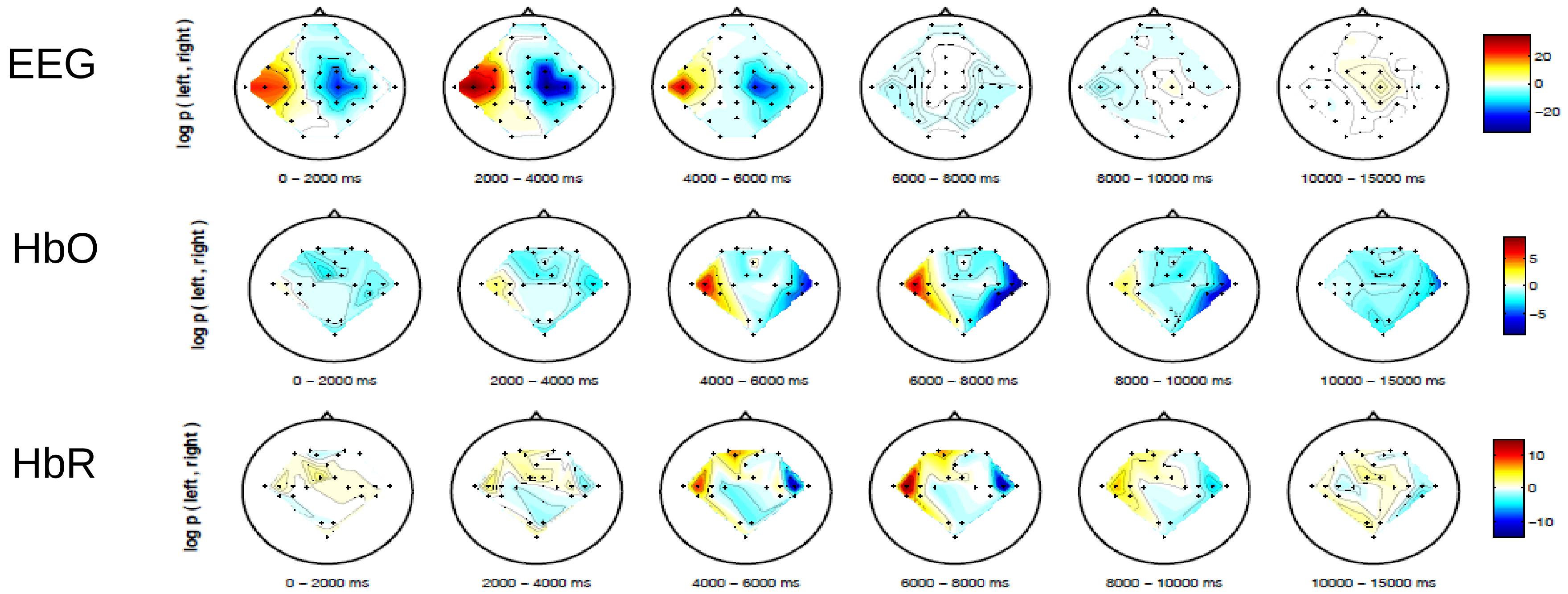


EEG earlier

NIRS has clear lateralization

HbO goes up, HbR down

Topography for Imagery Movements



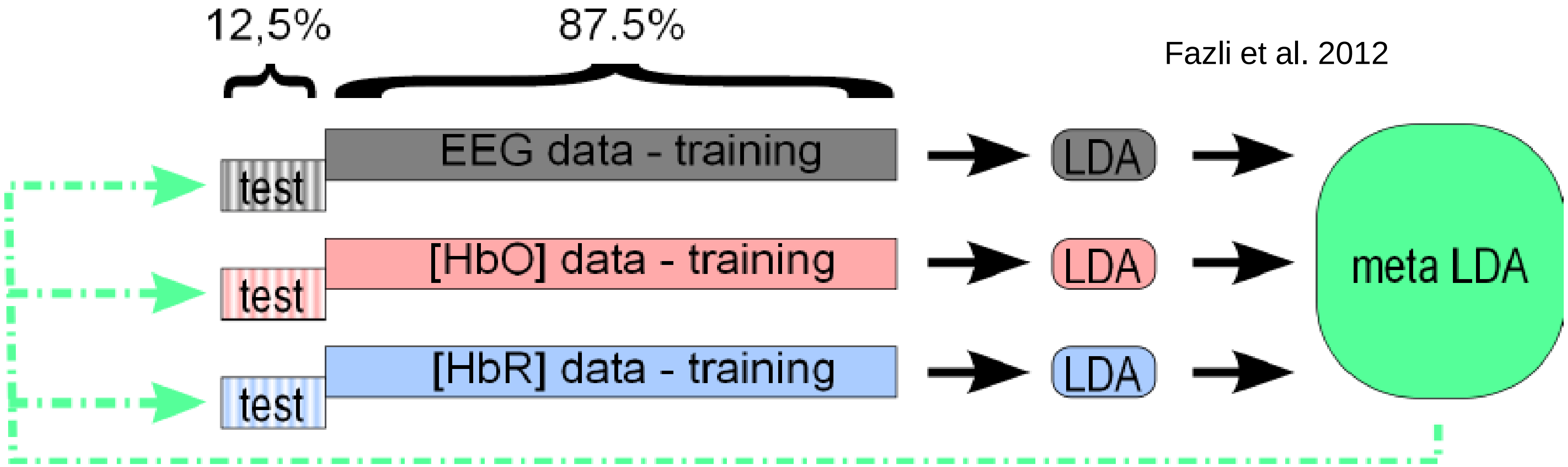
Similar results

EEG earlier

NIRS has clear lateralization

HbO goes up, HbR up (reason unsolved)

Combination of EEG and NIRS



Fazli et al. 2012

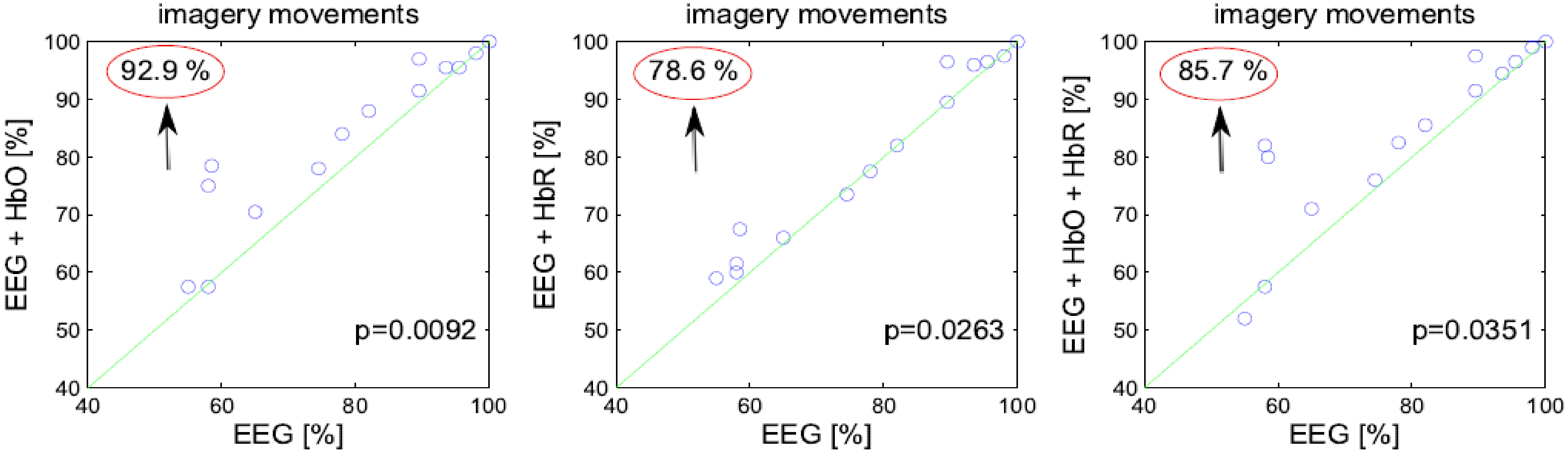
LDA classifier estimated for EEG, HbO and HbR (individually)

Meta-classifier estimated for combination in each subject

All within cross-validation (8 chronological splits)

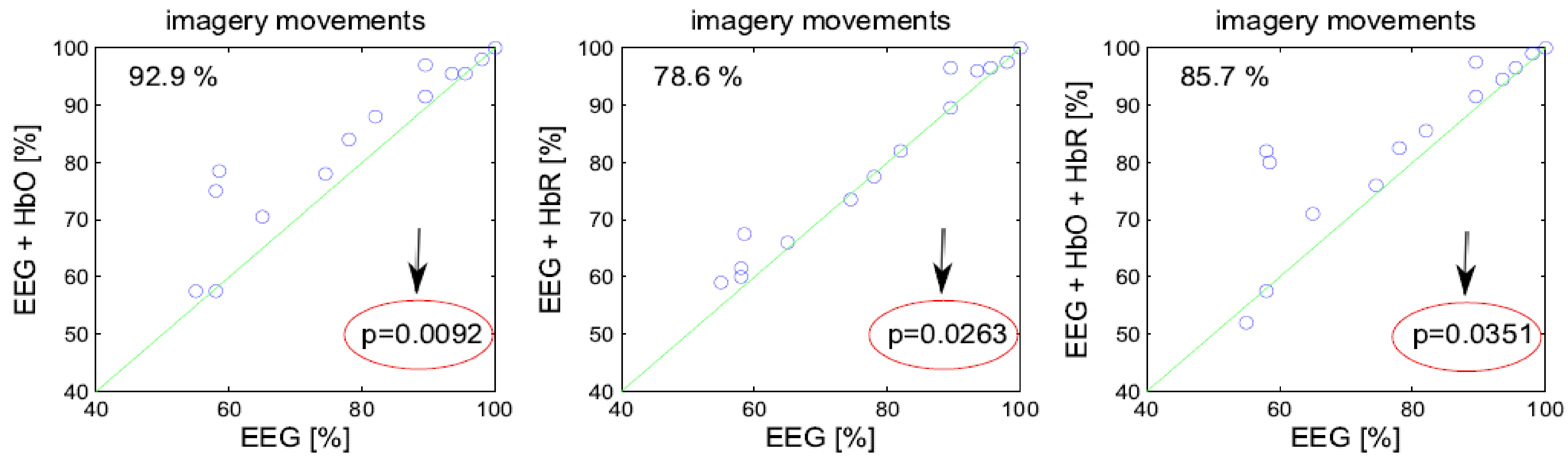
Feature Combination

Fazli et al. 2012



NIRS-EEG combinations have higher classification accuracies for vast majority of subjects

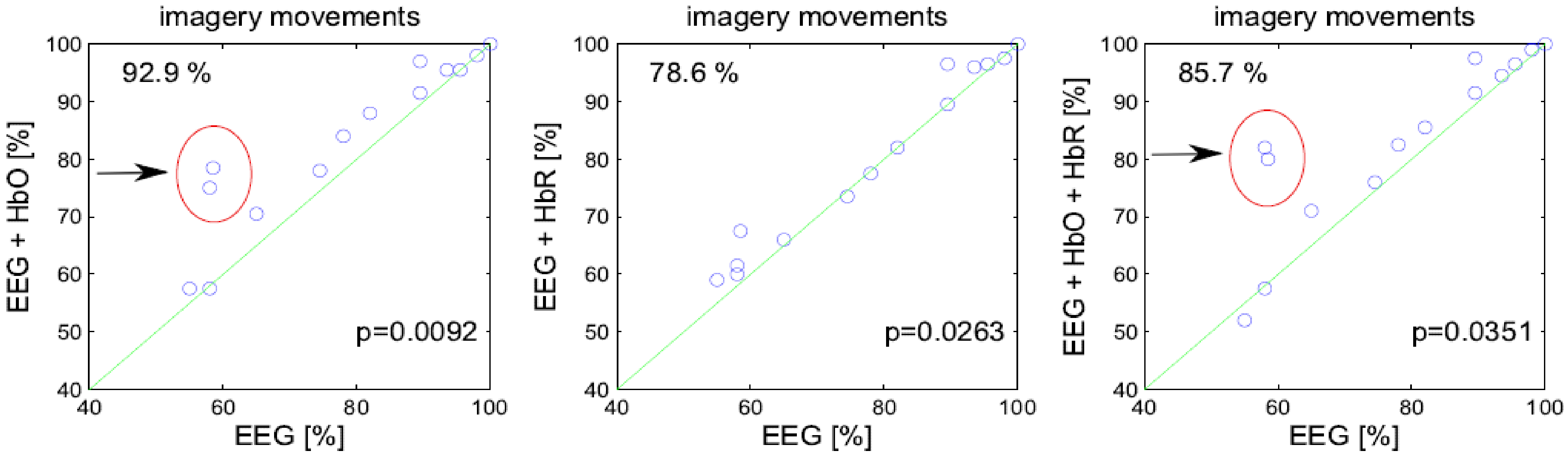
Feature Combination



Fazli et al. 2012

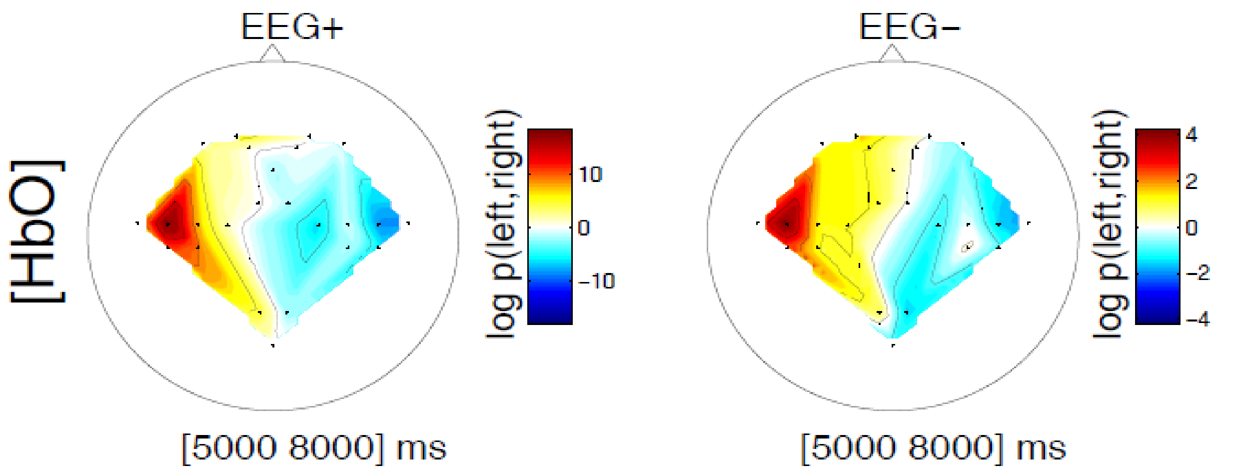
t-tests reveal a significant increase of classification accuracy for combination

Feature Combination



Some subjects, which were not classifiable with EEG become classifiable by a meta-classifier in combination with NIRS

Mutual Information



NIRS features for all correct EEG trials (EEG+) and incorrect EEG trials (EEG-)

Pattern is similar although the significance drops

NIRS can complement the EEG with physiological meaningful information

Discussion

Problems

- Different temporal properties of the measurement devices (e.g. EEG: 1000 Hz, NIRS: max. 10 Hz)
- Temporal lag between parameters
- Different signal qualities

Ideas to Overcome the Temporal Lag

- NIRS as a measure of subjects' attention to predict EEG-based performance
- NIRS as a localizer of the source of EEG signals
- NIRS as a 'stop', e.g. to discard a EEG-based classified trial when not confirmed by NIRS

Correlating apples and oranges

[Biessmann et al. Neuroimage 2012, Machine Learning 2010]

CCA: correlating apples and oranges

Given two (or more) multivariate variables

$$X \in \mathbb{R}^M, Y \in \mathbb{R}^N$$

CCA finds projections

$$w_x \in \mathbb{R}^M, w_y \in \mathbb{R}^N$$

that maximise the covariance between the variables

$$\underset{u}{\text{ar}} \begin{bmatrix} 0 & C_{xy} \\ C_{yx} & 0 \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix} = \alpha \begin{bmatrix} C_{xx} & 0 \\ 0 & C_{yy} \end{bmatrix} \begin{bmatrix} w_x \\ w_y \end{bmatrix}$$

kCCA: solving CCA on data kernels

Intuition behind the Kernel Trick:

The solution of CCA in kernel space is obtained by solving the generalised eigenvalue problem

$$\begin{bmatrix} 0 & K_x K_y \\ K_y K_x & 0 \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix} = \rho \begin{bmatrix} K_x^2 & 0 \\ 0 & K_y^2 \end{bmatrix} \begin{bmatrix} \alpha_x \\ \alpha_y \end{bmatrix}$$

The solutions in the input space can be recovered by

$$w_x = X\alpha_x$$

$$w_y = Y\alpha_y$$

No need to compute big covariance matrices!

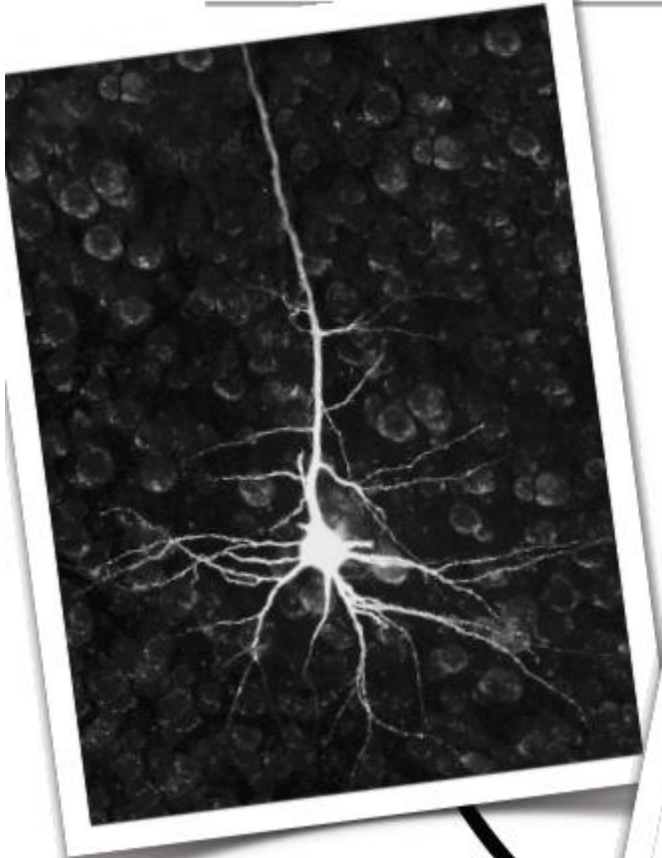
tkCCA: correlating apples and oranges over time

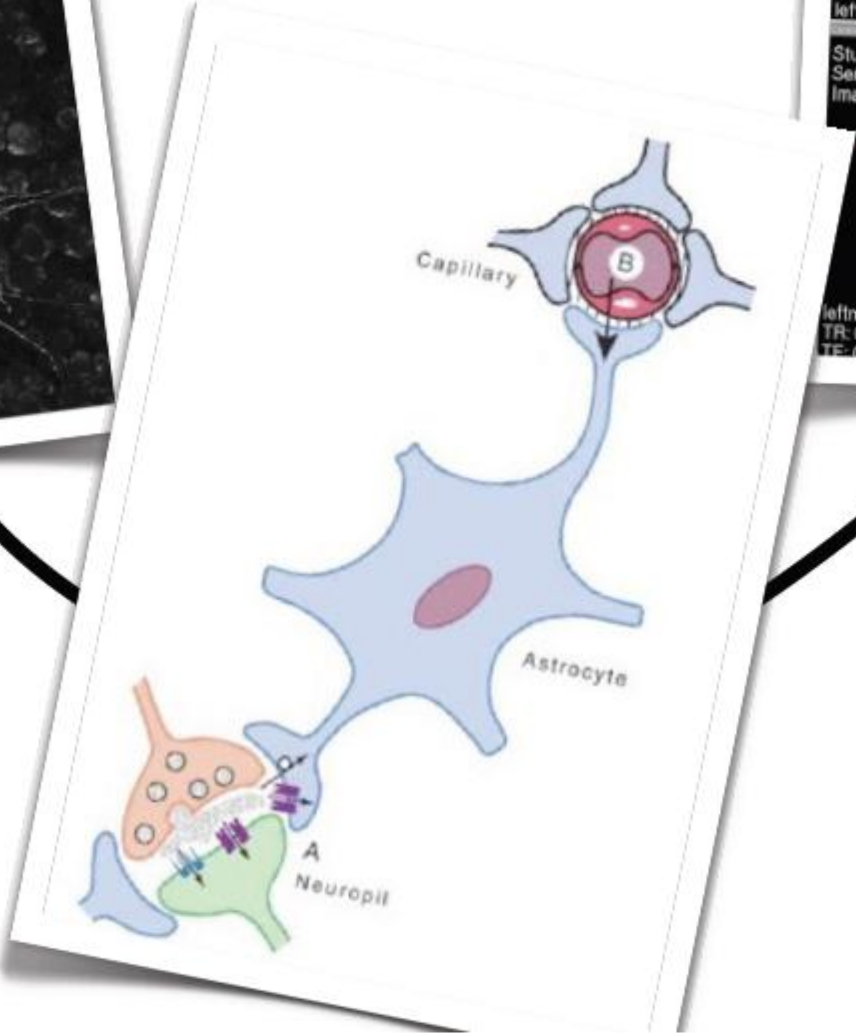
$$\underset{w_x(\tau), w_y}{\operatorname{argmax}} \operatorname{Corr}\left(\sum_{\tau} w_x(\tau)^\top x(t-\tau),\ w_y^\top y(t)\right)$$

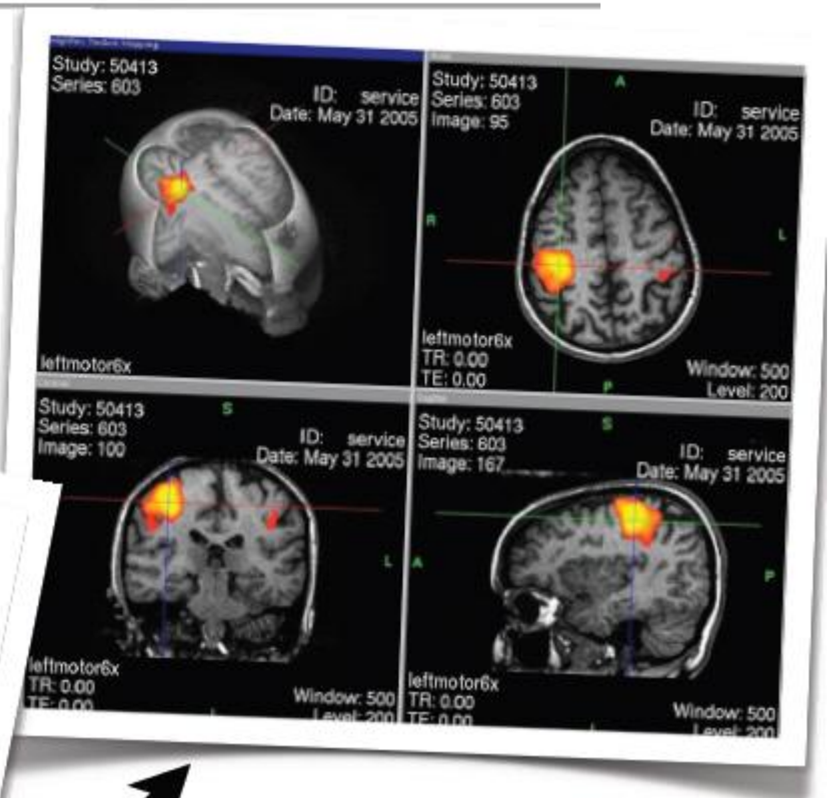
$$\tilde{X} = \begin{bmatrix} X_{\tau_1} \\ X_{\tau_2} \\ \vdots \\ X_{\tau_T} \end{bmatrix} \qquad \tilde{w}_x = \begin{bmatrix} w_x(\tau_1) \\ w_x(\tau_2) \\ \vdots \\ w_x(\tau_T) \end{bmatrix}$$

$$\underset{w_{\tilde{x}}, w_y}{\operatorname{argmax}} \operatorname{Corr}\left(\tilde{w}_x^\top \tilde{X}, w_y^\top Y\right)$$

Application: Neuro-Vascular Coupling

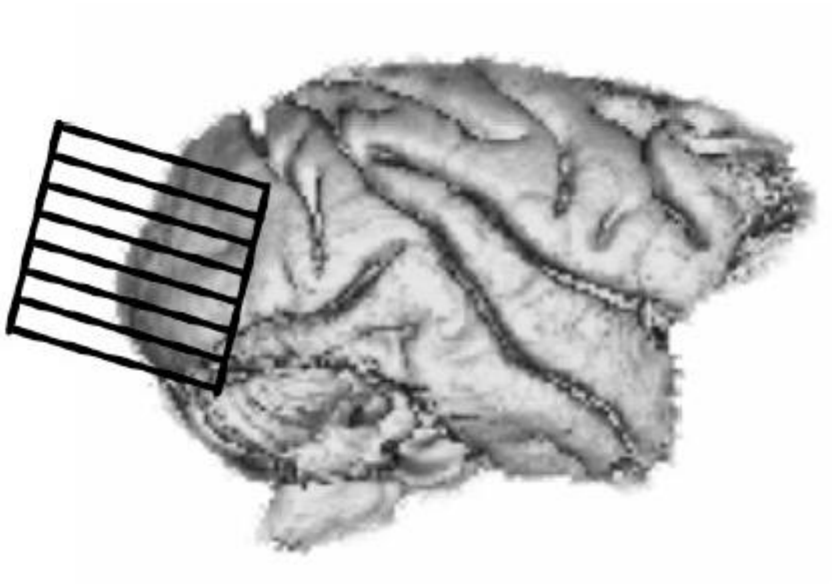


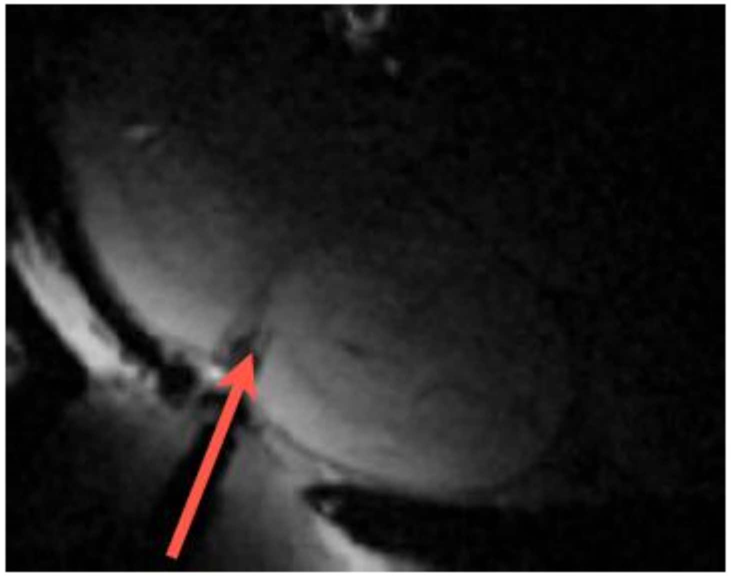




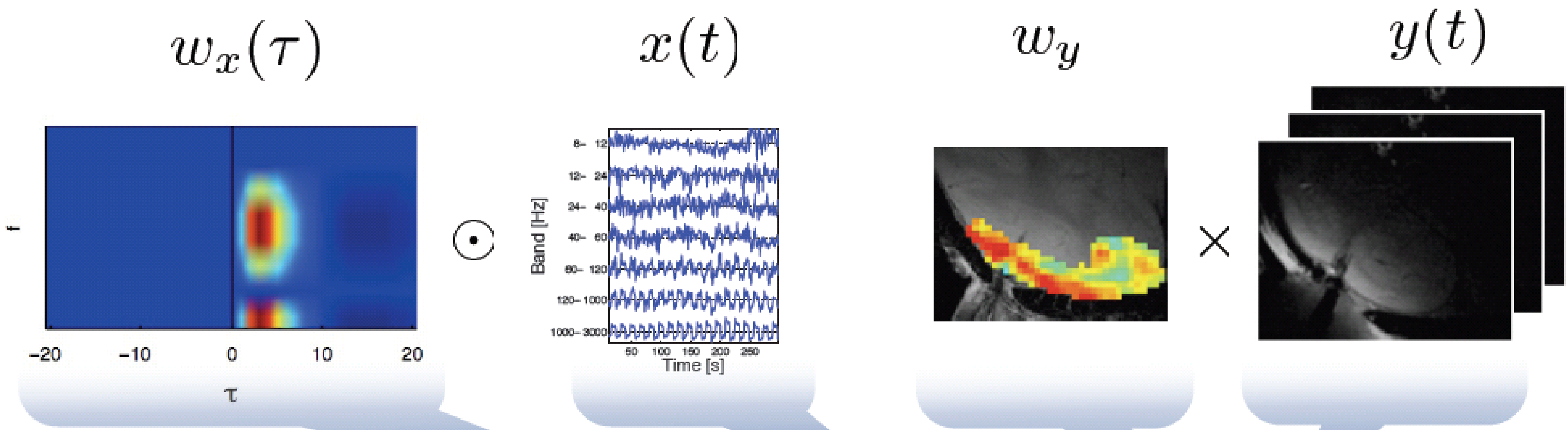
Experimental Setup

- **Simultaneous measurements of**
  - fMRI/ BOLD signal
  - Intracortical neural activity





Temporal Kernel CCA

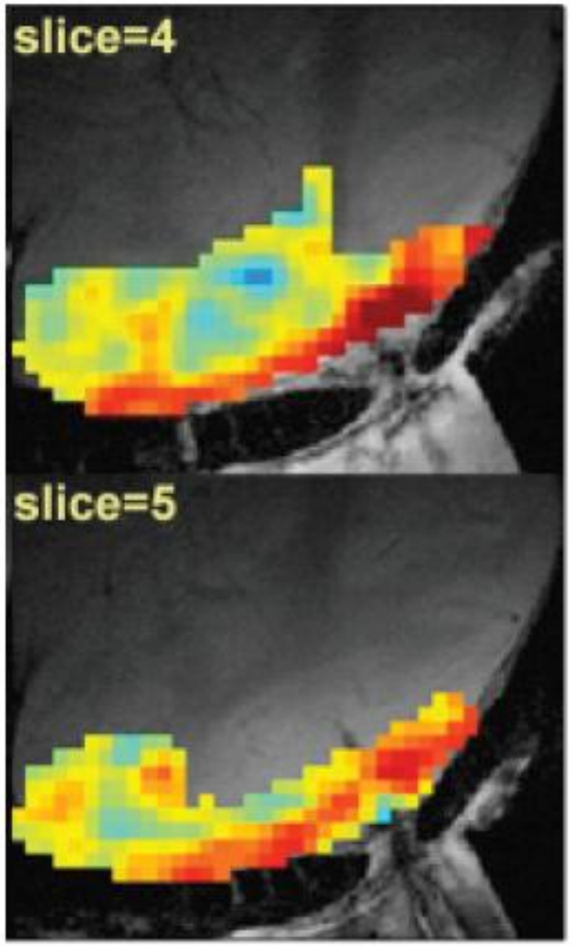


$$\underset{w_x(\tau),w_y}{\operatorname{argmax}} \operatorname{Corr}\left(\sum_{\tau} w_x(\tau)^\top x(t-\tau),\ w_y^\top y(t)\right)$$

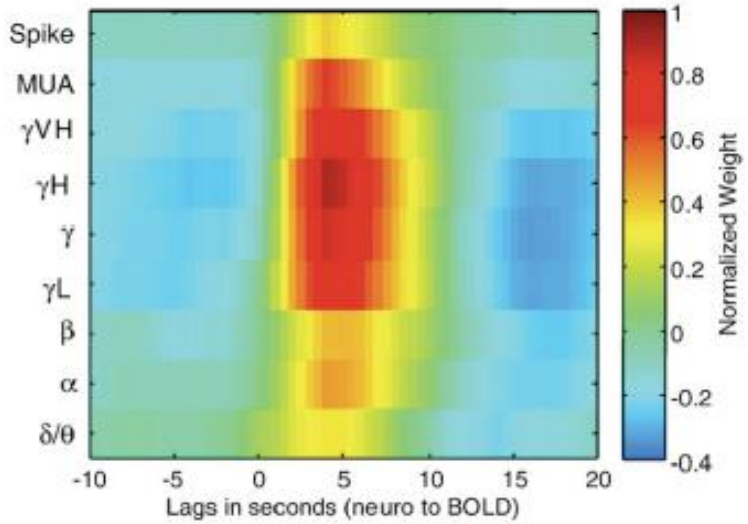
multivariate convolution of the neurophysiological signal with frequency dependent HRF

spatial weighting of voxels with activation pattern.

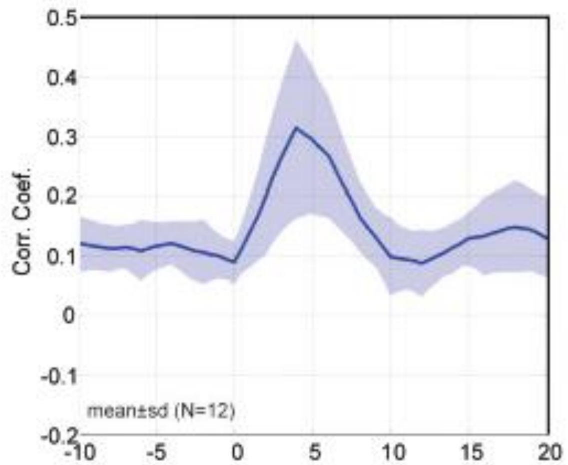
Results tkCCA: spatial dependencies and HRF



Spatial Dependencies



Haemodynamic Response Function



Canonical Correlogram

Murayama et al., "*Relationship between neural and haemodynamic signals during spontaneous activity studied with temporal kernel CCA*", Magnetic Resonance Imaging, 2010

Conclusion II

- **CCA**
  - finds projections for sets of variables that maximise correlation
- **kernel CCA**
  - extends CCA to non-linear dependencies
  - applicable to high dimensional data
- **Temporal kernel CCA**
  - extends kCCA to data with non-instantaneous correlations
  - computes multivariate convolution from one modality to another

FOR INFORMATION SEE:
www.bbci.de

Part III ERP analysis & applications beyond communication

Method:

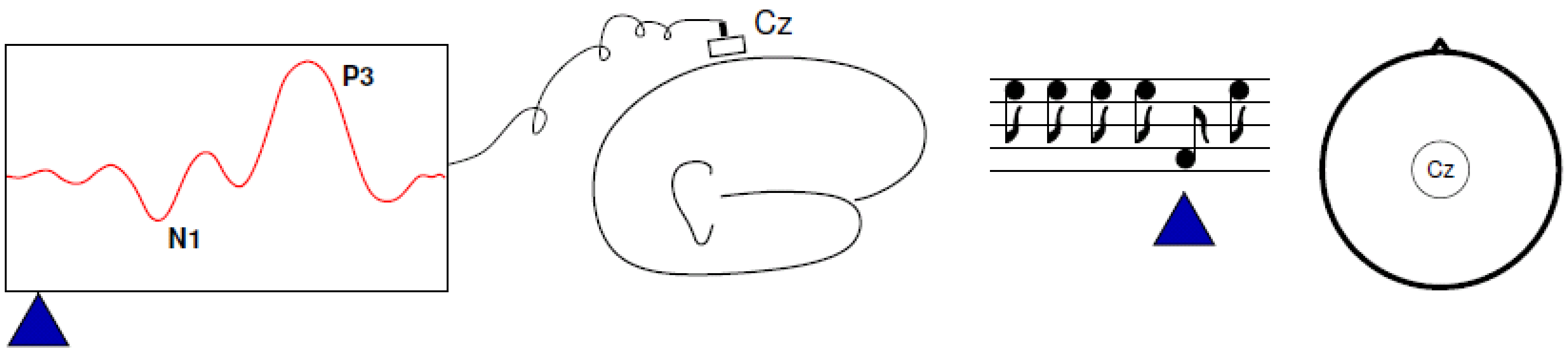
- classification of **spatio-temporal** features;
- *shrinkage* of the sample covariance matrix to counterbalance the estimation bias

Application:

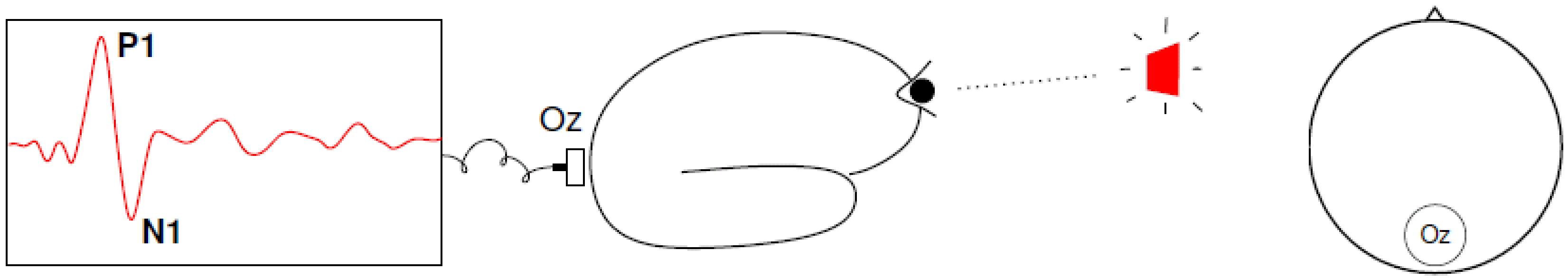
- classification of single-trial ERPs in an attention-based speller

Neurophysiological Background for ERPs

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



The presentation of a visual stimulus elicits a Visual Evoked Potential (VEP) in visual cortex *if focused*:

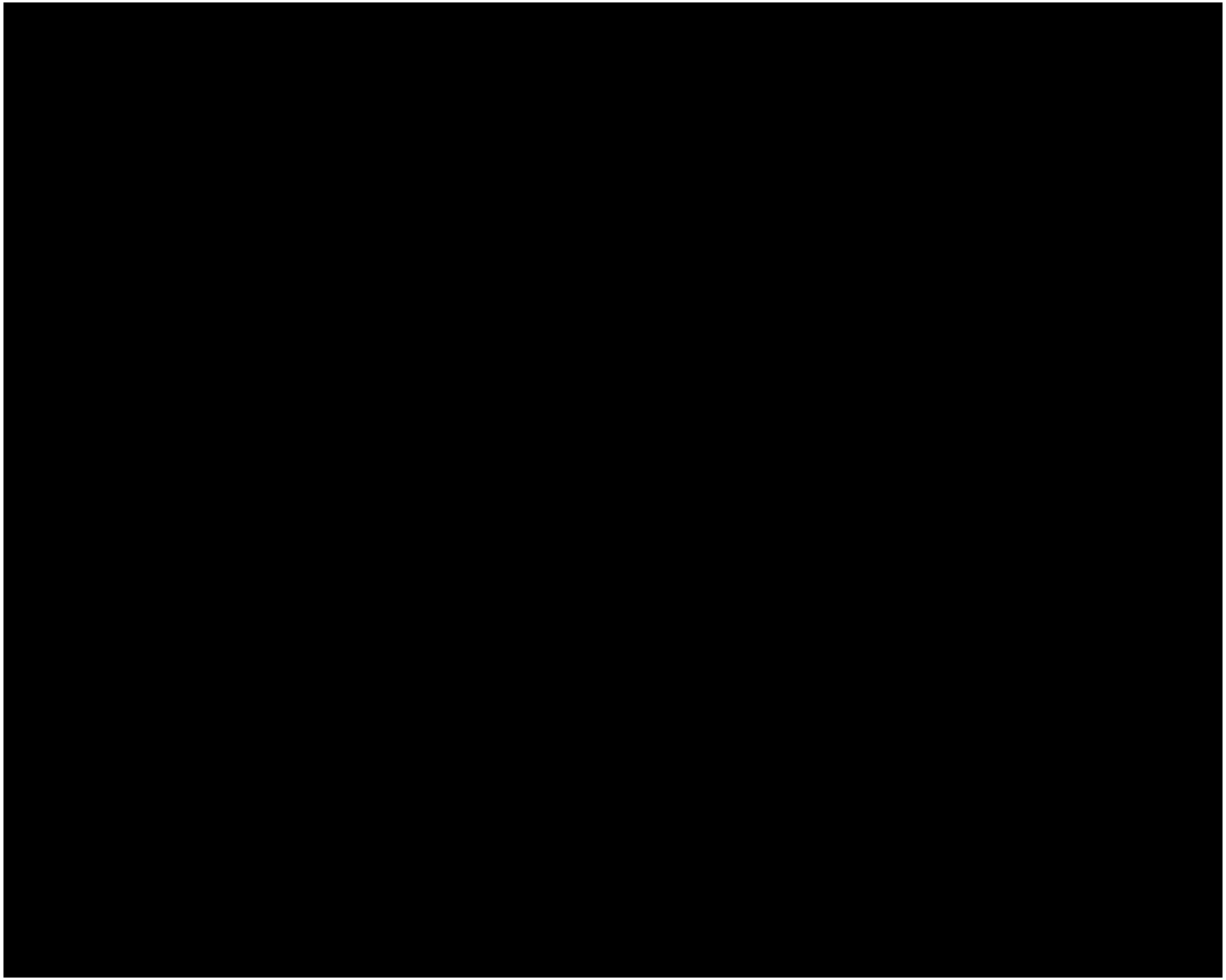


Experimental Design

Classic Matrix Speller

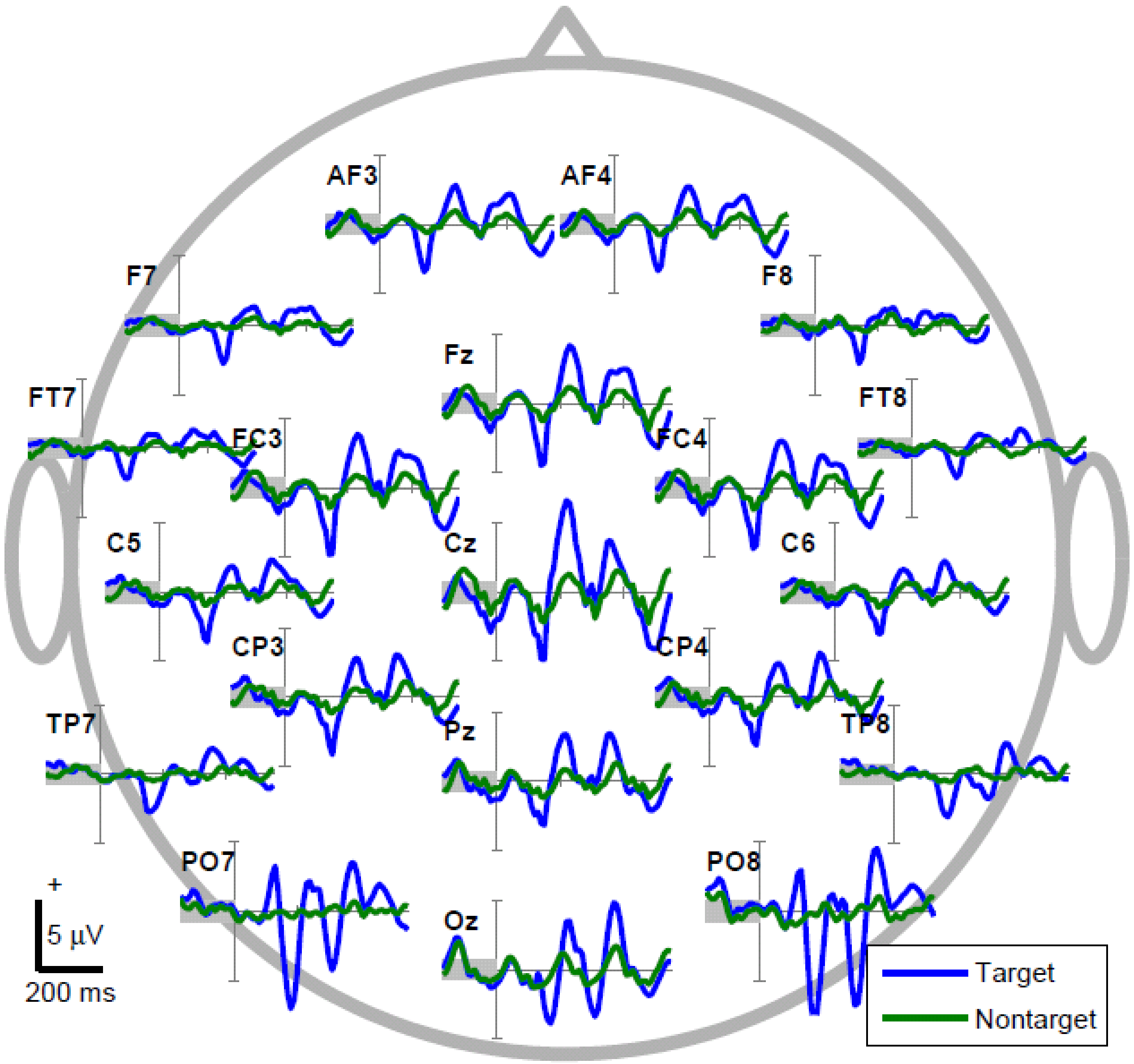
Attention-based Hex-o-Spell

P300 in action: Hex-o-spell



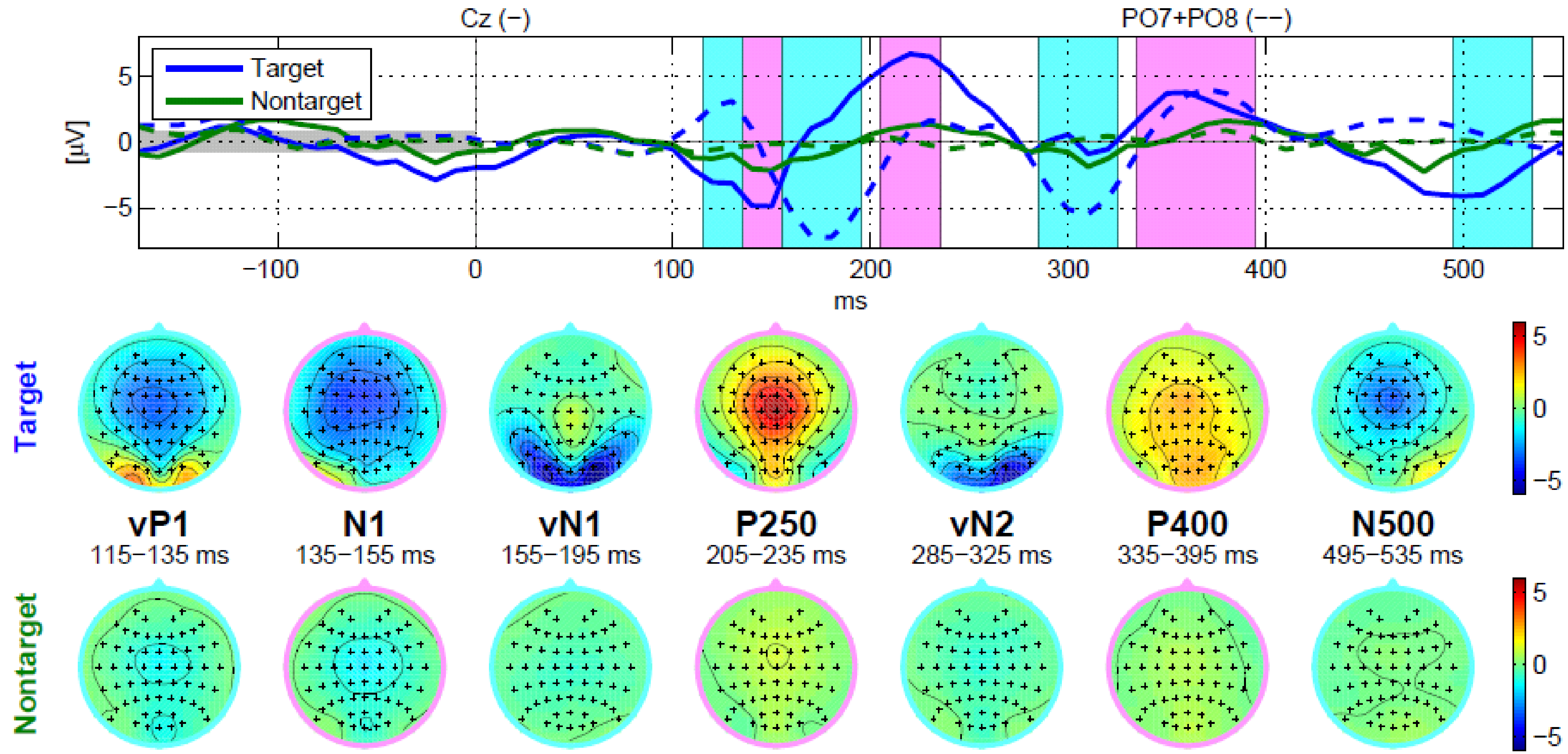
Single subject ERPs for Hex-o-spell

Data set for illustration of classification methods:



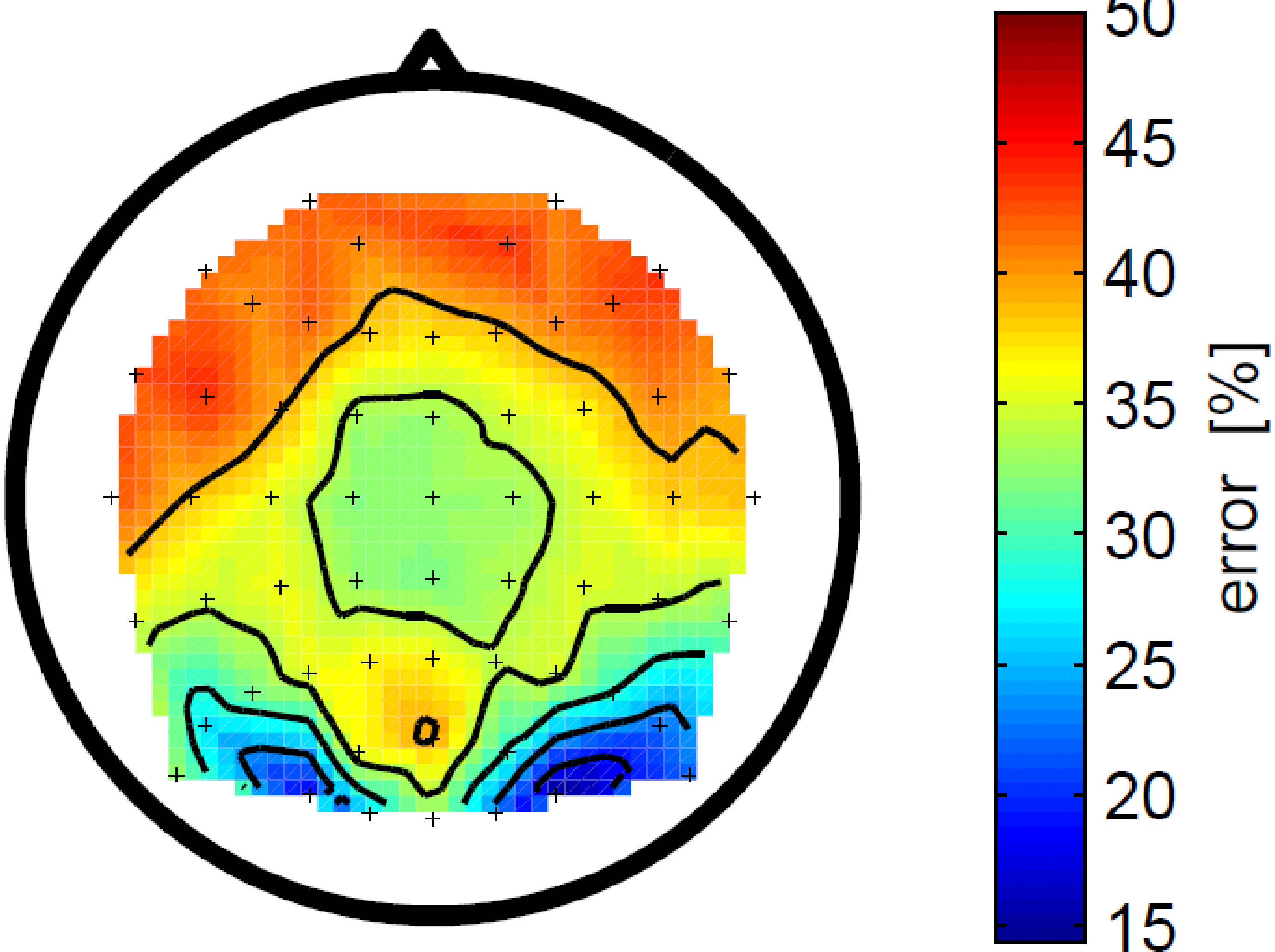
Topographies of ERP components

There are several ERP components that can be used to determine the attended symbol:

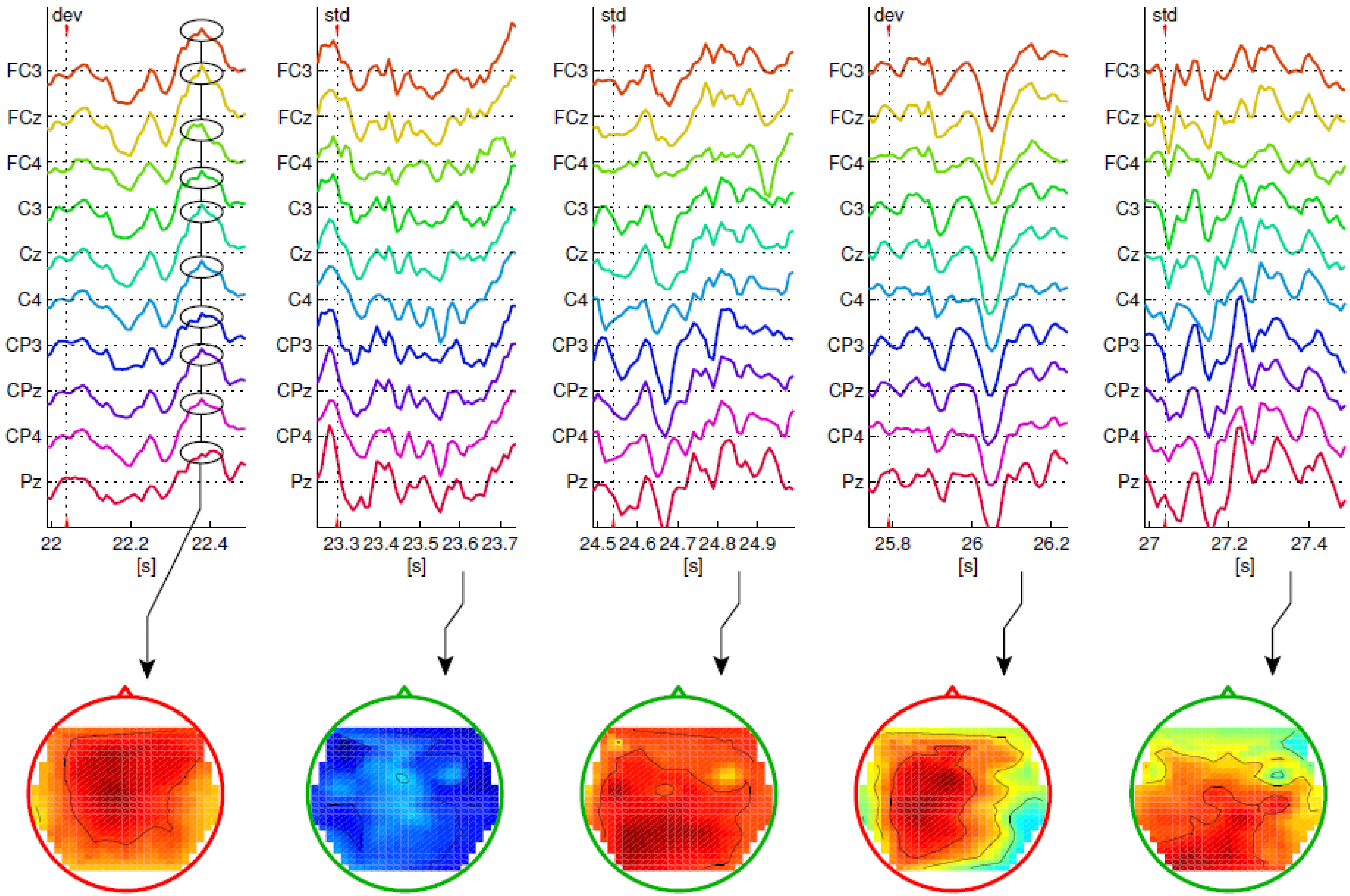


Classification of temporal features

As a first step: classification on raw time courses (115–535 ms) in single channels. The result is displayed as scalp map:

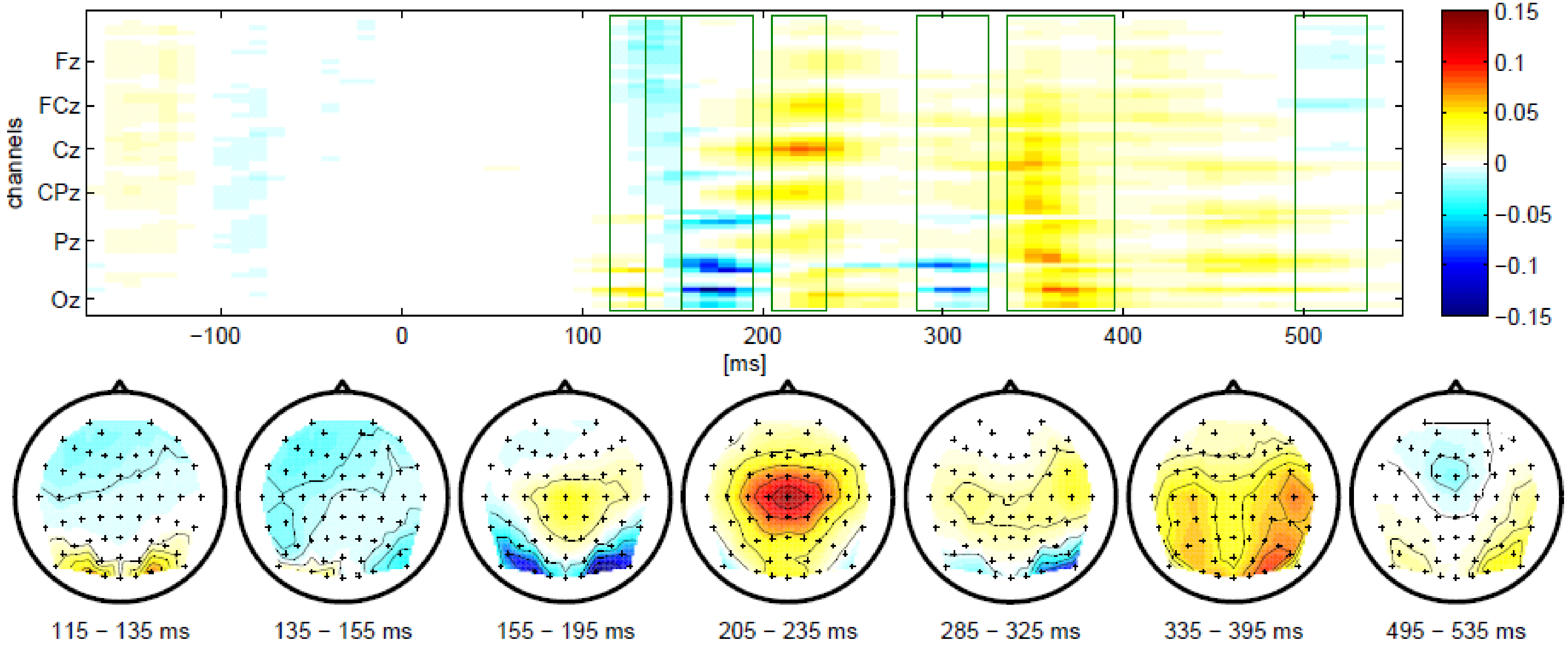


Extraction of spatial features



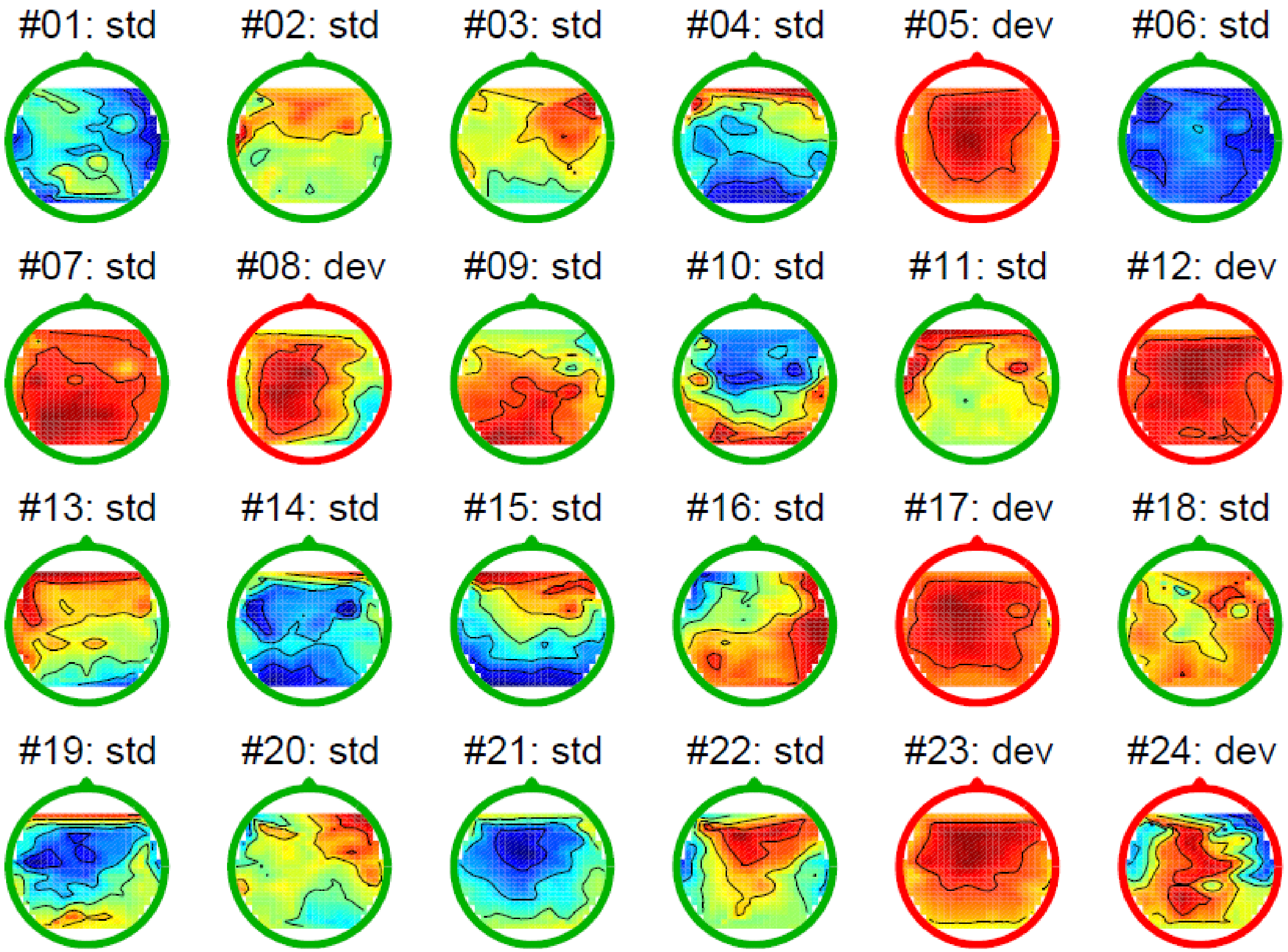
The r^2 matrix of differences

The temporal and spatial structure of the difference between ERPs of different conditions can be investigated by the signed r^2-matrix:



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

Spatial features



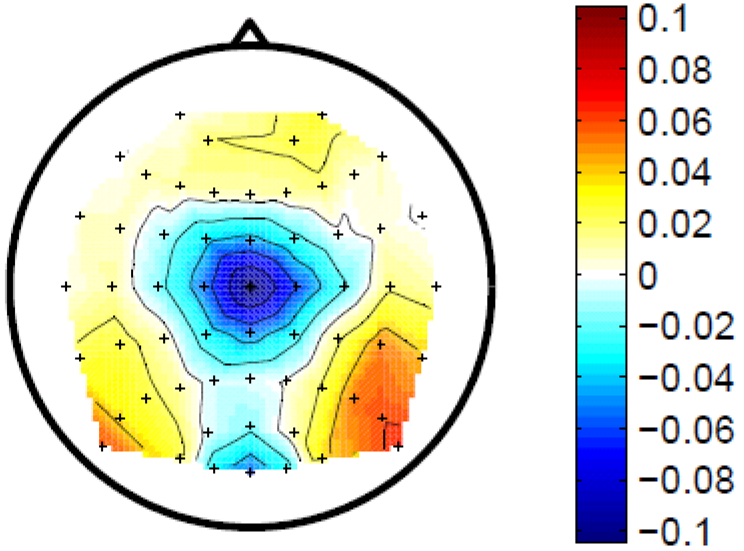
A linear classifier as a spatial filter

A linear classifier that was trained on *spatial features* can also be regarded as a **spatial filter**.
Let $\mathbf{w}$ be the LDA weight vector and $\mathbf{X} \in \mathbb{R}^{\#\text{chans} \times \#\text{time points}}$ be continuous EEG signals. Then

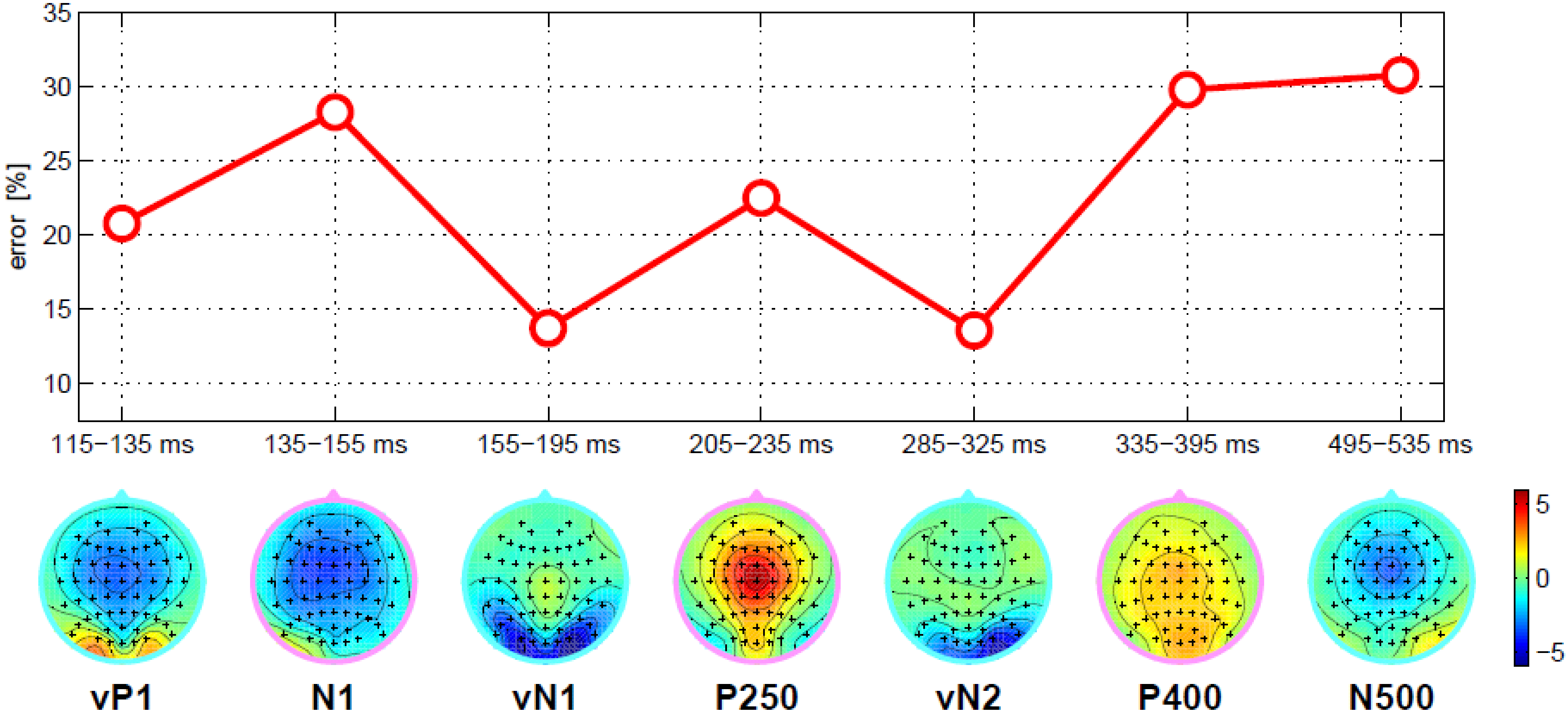
$$\mathbf{X}_f := \mathbf{w}^\top \mathbf{X} \quad \in \mathbb{R}^{1 \times \#\text{time points}}$$

is the result of spatial filtering: each channel of $\mathbf{X}$ is weighted with the corresponding component of $\mathbf{w}$ and summed up.

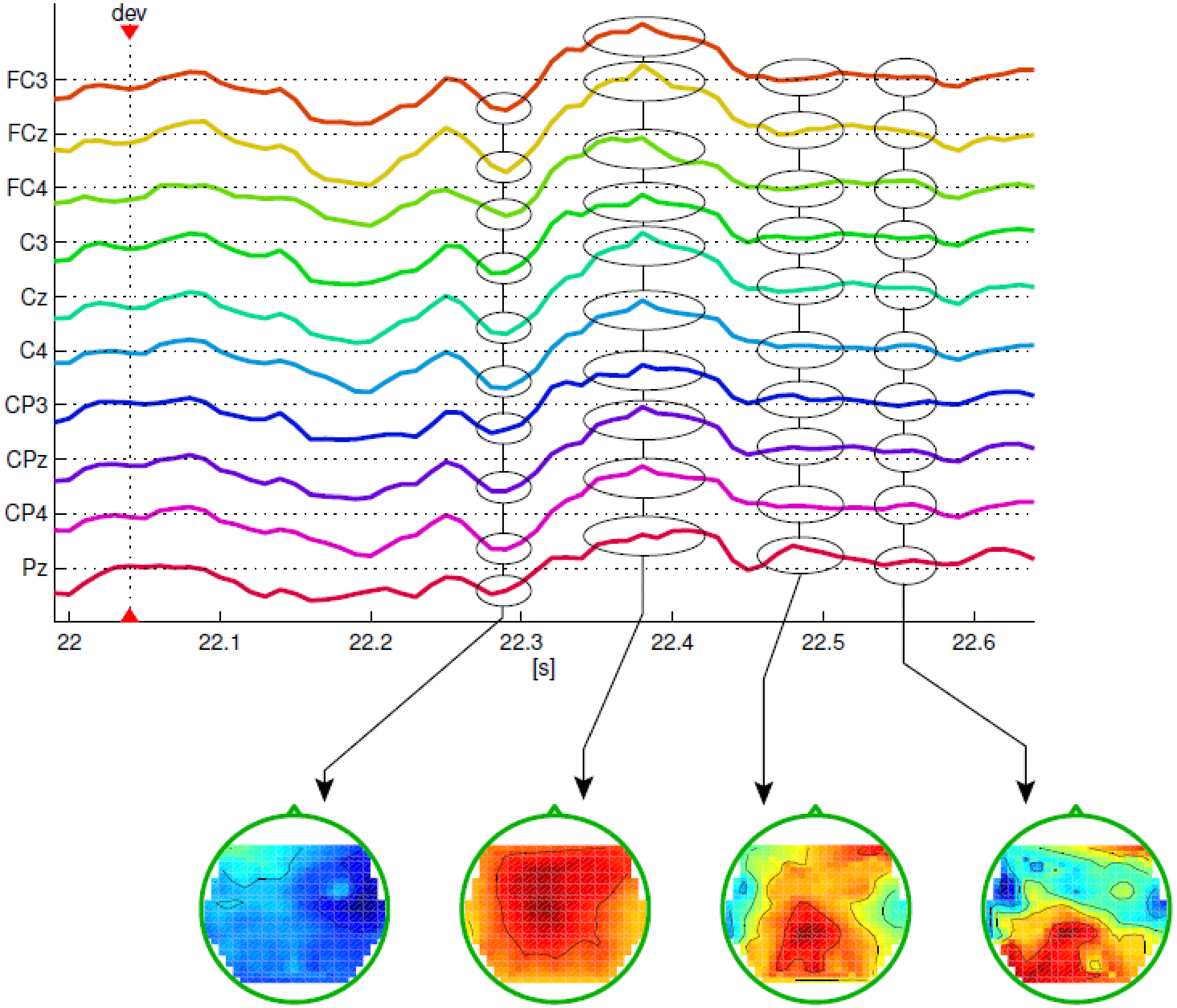
The weight vector of the classifier can be display as scalp map:



Classification results of spatial features

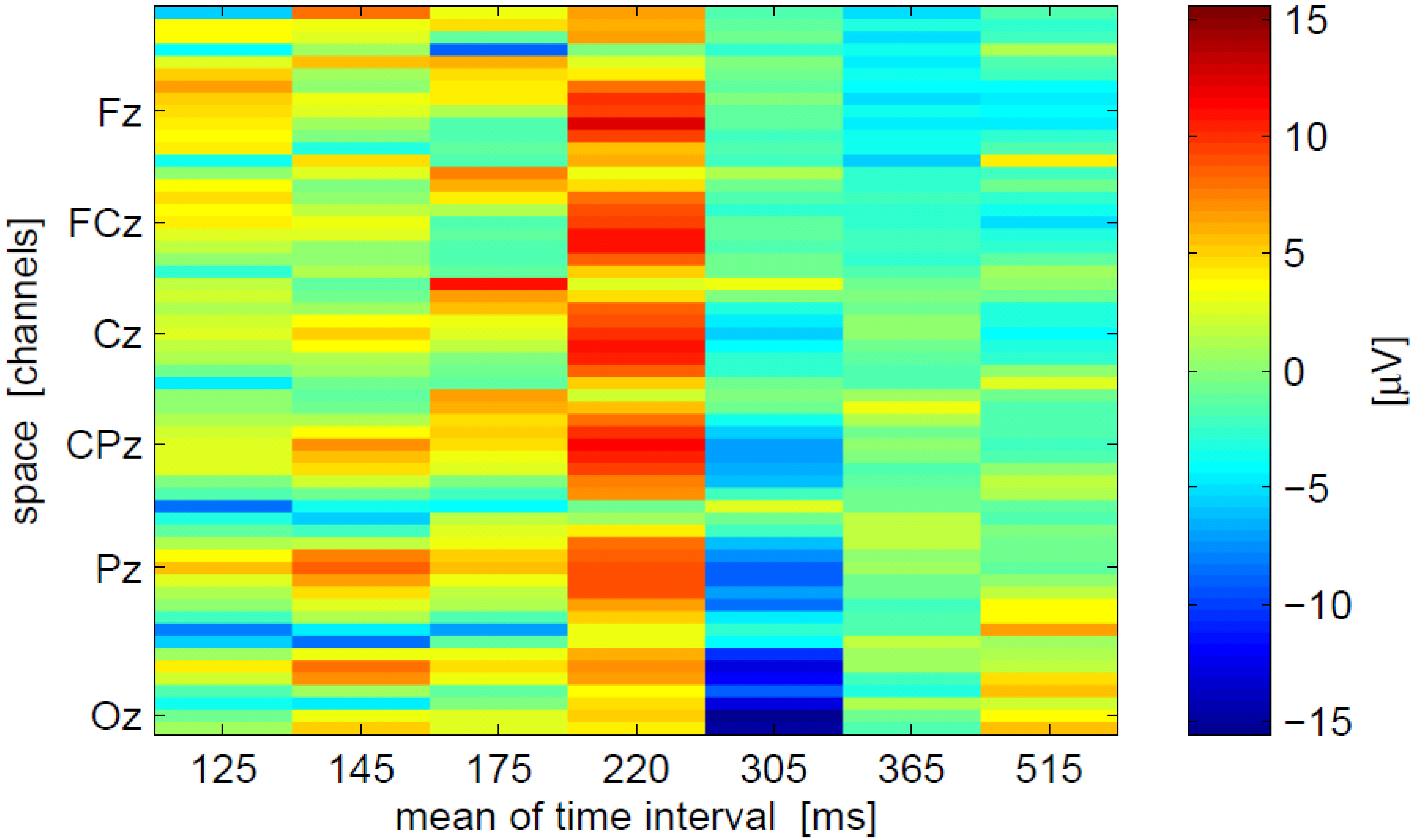


Extraction of spatio-temporal features

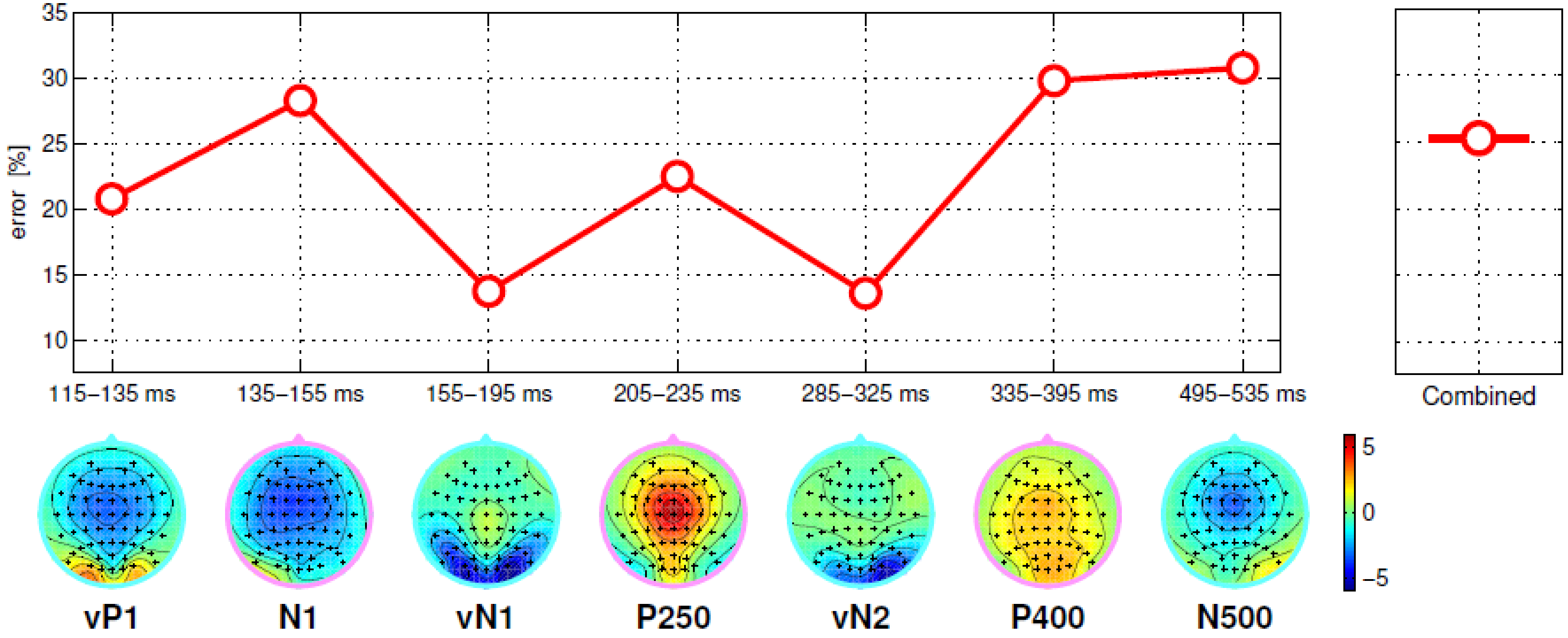


Spatio-temporal features

Spatio-temporal features are typically high-dimensional (here 59 EEG channels $\times$ 7 time intervals $=$ 413 dimensional features):



Classification results for spatio-temporal features



Although information was added, classification on the concatenated feature becomes worse: *overfitting*.

Bias in estimating covariances

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ be n vectors drawn from a d-dimensional Gaussian distribution $\mathcal{N}(\mu, \mathbf{\Sigma})$.
For classification μ and $\mathbf{\Sigma}$ have to be estimated from the data:

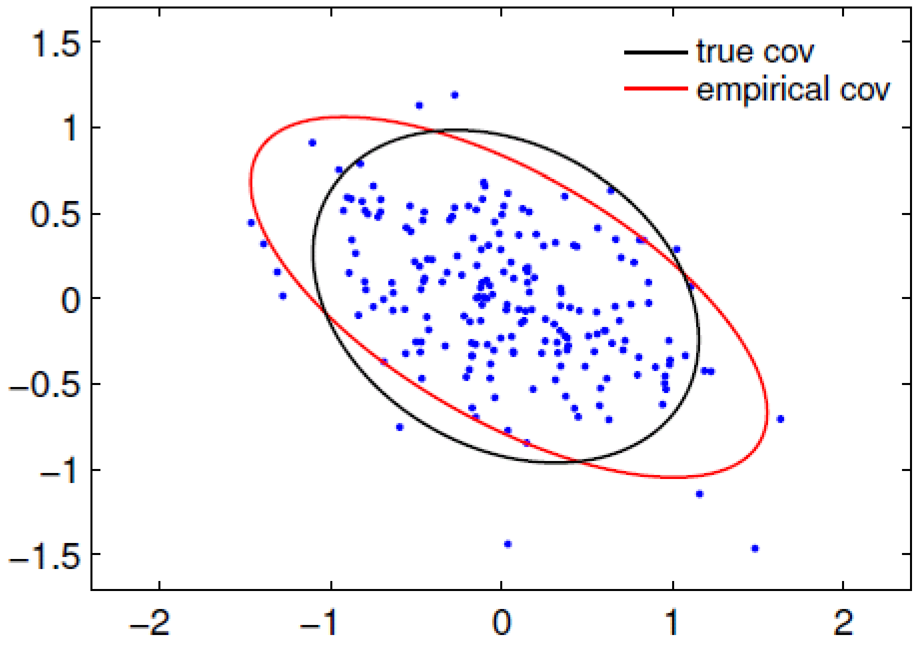
- $\hat{\mu} = \frac{1}{n} \sum_{k=1}^{n} \mathbf{x}_k$
- $\hat{\mathbf{\Sigma}} = \frac{1}{n-1} \sum_{k=1}^{n} (\mathbf{x}_k - \hat{\mu})(\mathbf{x}_k - \hat{\mu})^\top$

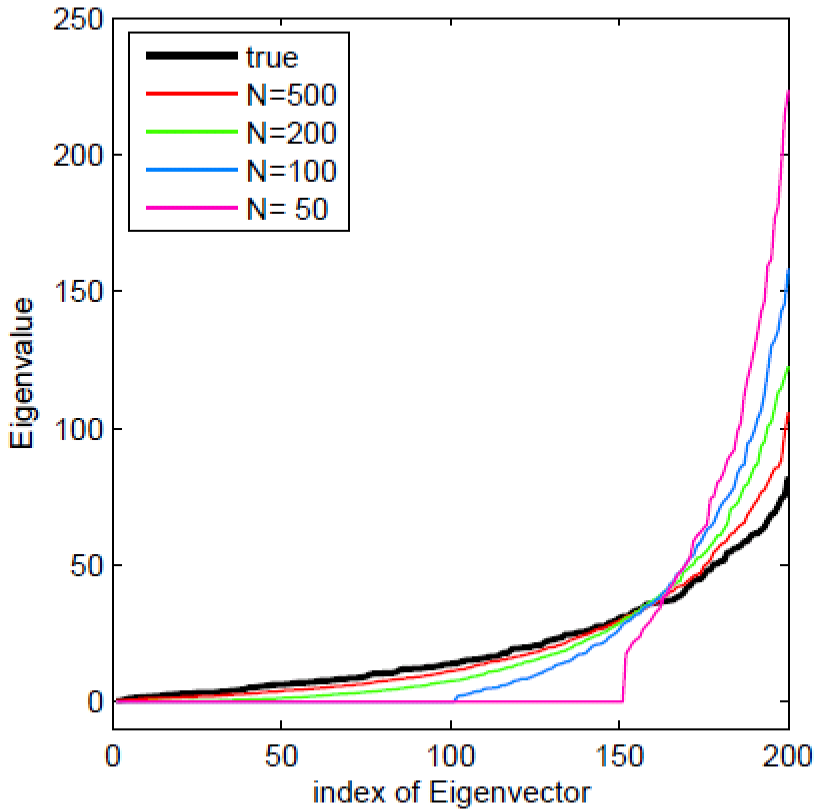
But, if the number of samples n is not large relative to the dimension d, the estimation is error-prone.
There is a systematical bias:

- Large Eigenvalues of $\hat{\mathbf{\Sigma}}$ are too large
- Small Eigenvalues of $\hat{\mathbf{\Sigma}}$ are too small

This affects, e.g., classification with LDA:
Normal vector of LDA: $w = \hat{\mathbf{\Sigma}}^{-1}(\mu_1 - \mu_2)$.

Bias in estimating covariances II





A remedy for classification

A simple way that can partly fix the bias is **shrinkage**: the empirical covariance matrix is modified to be more spherical.
In LDA the empirical covariance matrix $\hat{\mathbf{\Sigma}}$ is replaced by

$$\tilde{\mathbf{\Sigma}}(\gamma) = (1-\gamma)\hat{\mathbf{\Sigma}} + \gamma\nu\mathbf{I}$$

for a $\gamma \in [0, 1]$ and ν defined as average Eigenvalue $\mathrm{trace}(\mathbf{S}_i)/d$.
Since $\hat{\mathbf{\Sigma}}$ is positive semi-definite we can have an Eigenvalue decomposition $\hat{\mathbf{\Sigma}} = \mathbf{V}\mathbf{D}\mathbf{V}^\top$ with orthonormal $\mathbf{V}$ and diagonal $\mathbf{D}$.
From

$$\tilde{\mathbf{\Sigma}} = (1-\gamma)\mathbf{V}\mathbf{D}\mathbf{V}^\top + \gamma\nu\mathbf{I} = \mathbf{V}\left((1-\gamma)\mathbf{D} + \gamma\nu\mathbf{I}\right)\mathbf{V}^\top$$

we see that

- $\tilde{\mathbf{\Sigma}}(\gamma)$ and $\hat{\mathbf{\Sigma}}$ have the same Eigenvectors (columns of $\mathbf{V}$)
- extreme Eigenvalues (large/small) are shrunk/extended towards the average ν.
- $\gamma = 0$ yields LDA without shrinkage, $\gamma = 1$ assumes spherical covariance matrices.

Modelselection

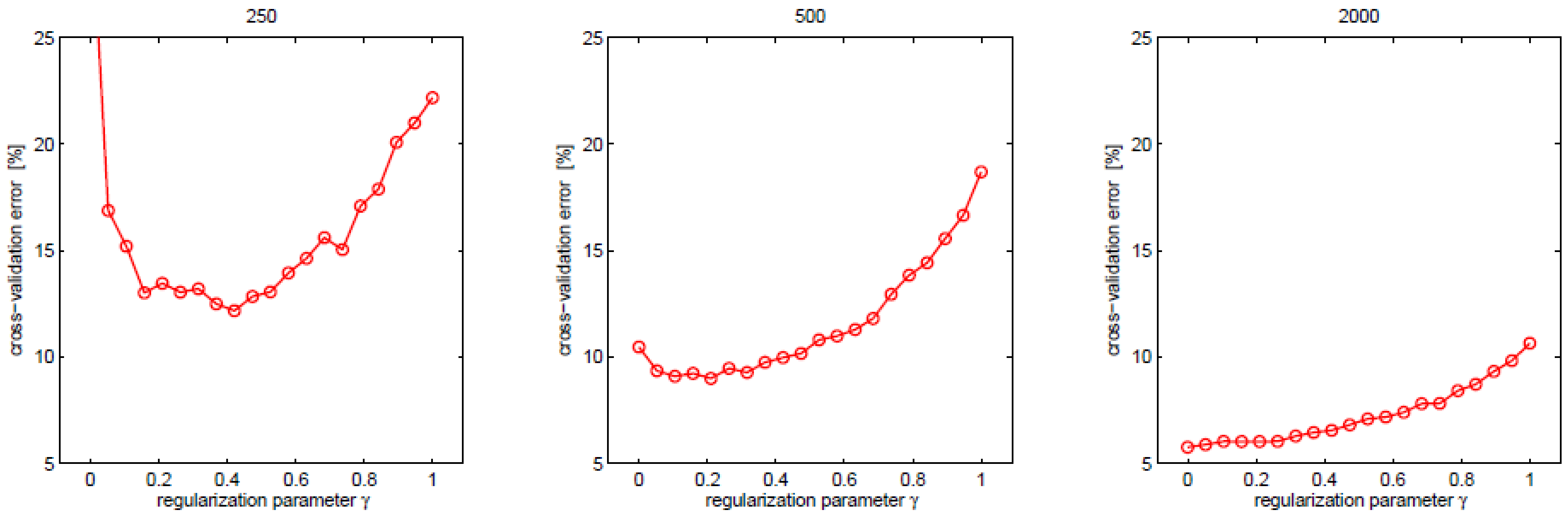
LDA with shrinkage of the empirical covariance matrix has one free parameter (γ), also called hyperparameter, that needs to be selected. There is no general way to do it.
Numerous strategies with different properties exist, e.g.

- empirical Bayes shrinkage estimator
- MDL: Minimum Description Length
- Model-selection based on cross-validation.
- ...

An easy (and also time-consuming) way is model-selection based on **cross-validation**.

Regularized LDA at work

Cross-validation results for different sizes of training data (250, 500, 2000) for different values of the regularization parameter γ (x-axis). Features vectors have 250 dimensions.

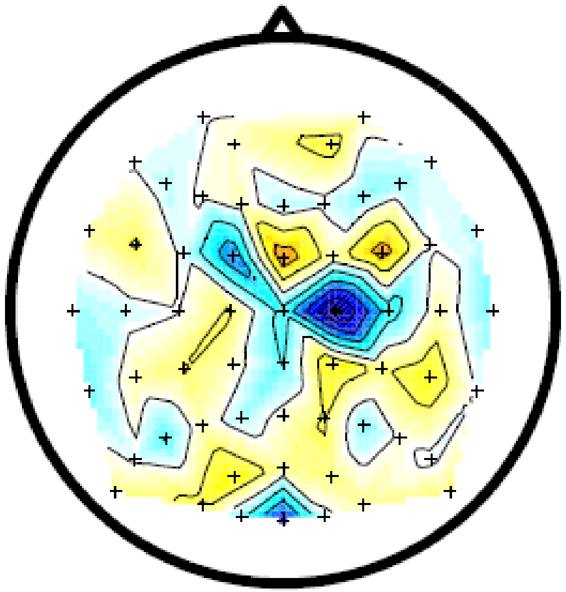


Investigating the impact of shrinkage

LDA: $w = \hat{\mathbf{\Sigma}}^{-1}(\mu_1 - \mu_2)$;   **shrinkage:** $\tilde{\mathbf{\Sigma}}(\gamma) = (1-\gamma)\hat{\mathbf{\Sigma}} + \gamma\nu\mathbf{I}$

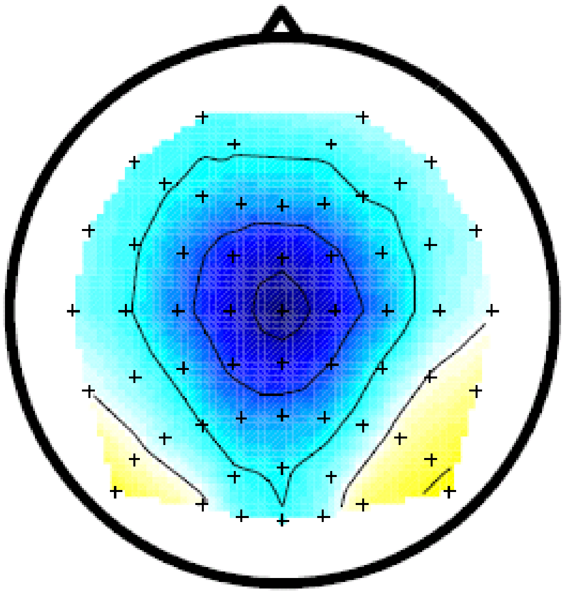
$\gamma = 0$

$w \sim \hat{\mathbf{\Sigma}}^{-1}(\mu_1 - \mu_2)$



$\gamma = 1$

$w \sim \mu_1 - \mu_2$

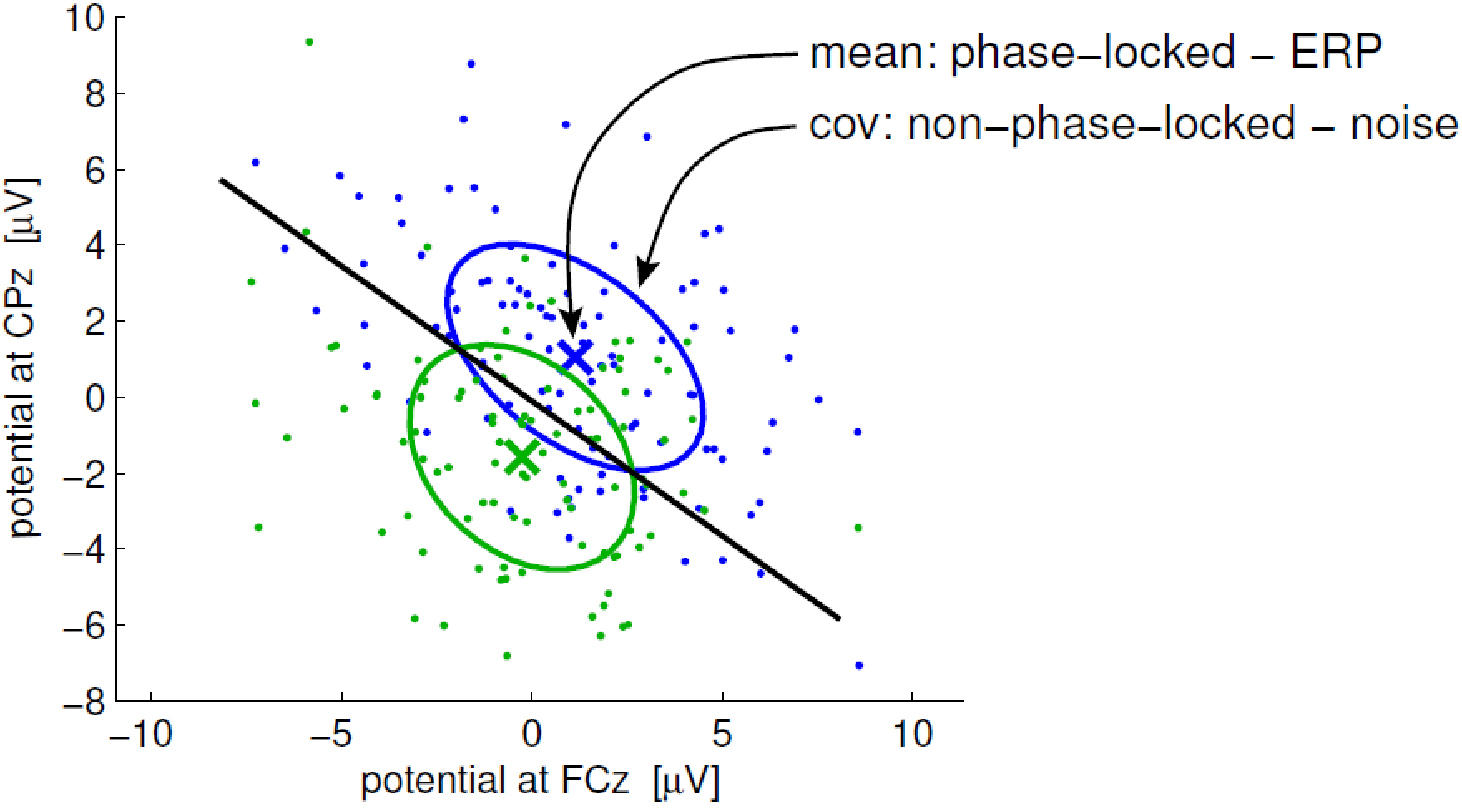


accounting for spatial structure of the noise

ERP and noise

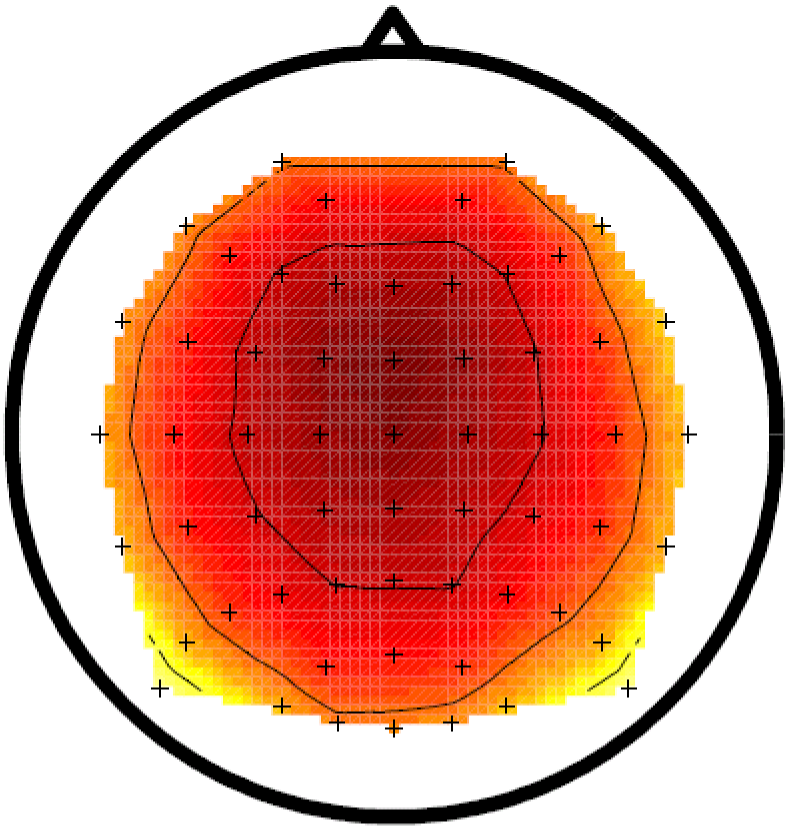
Simple assumption for ERPs: single trial $x_k(t)$ is composed of an ERP $s(t)$ and Gaussian 'noise' $\mathbf{n}_k(t)$:

$$\mathbf{x}_k(t) = \mathbf{s}(t) + \mathbf{n}_k(t) \qquad \text{for all trials } k = 1, \ldots, K$$

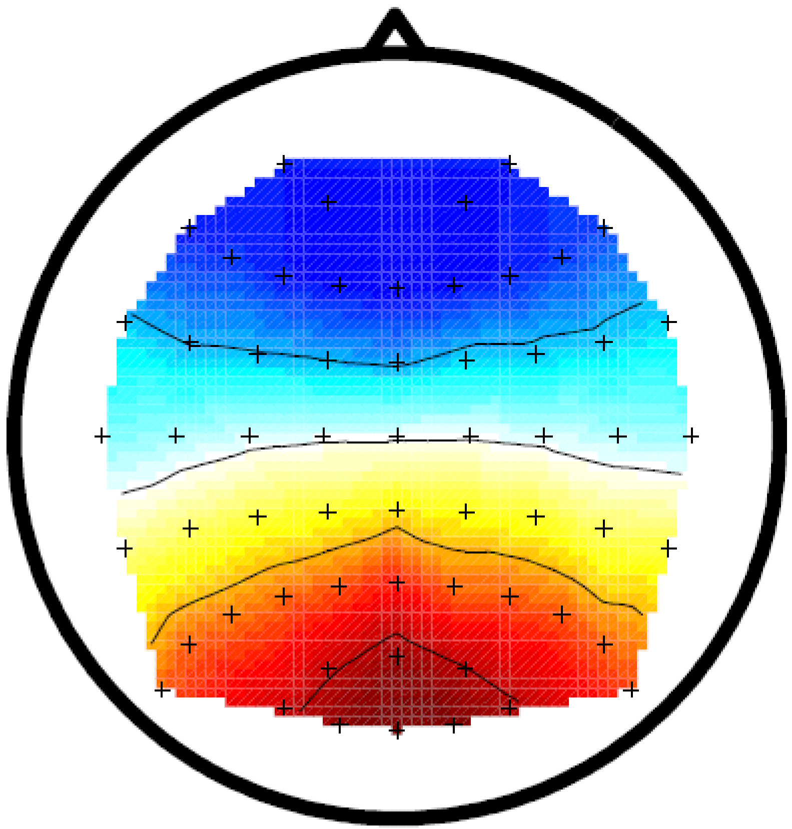


Spatial structure of noise

The two strongest principal components of the noise (covariance matrix) in this data set:

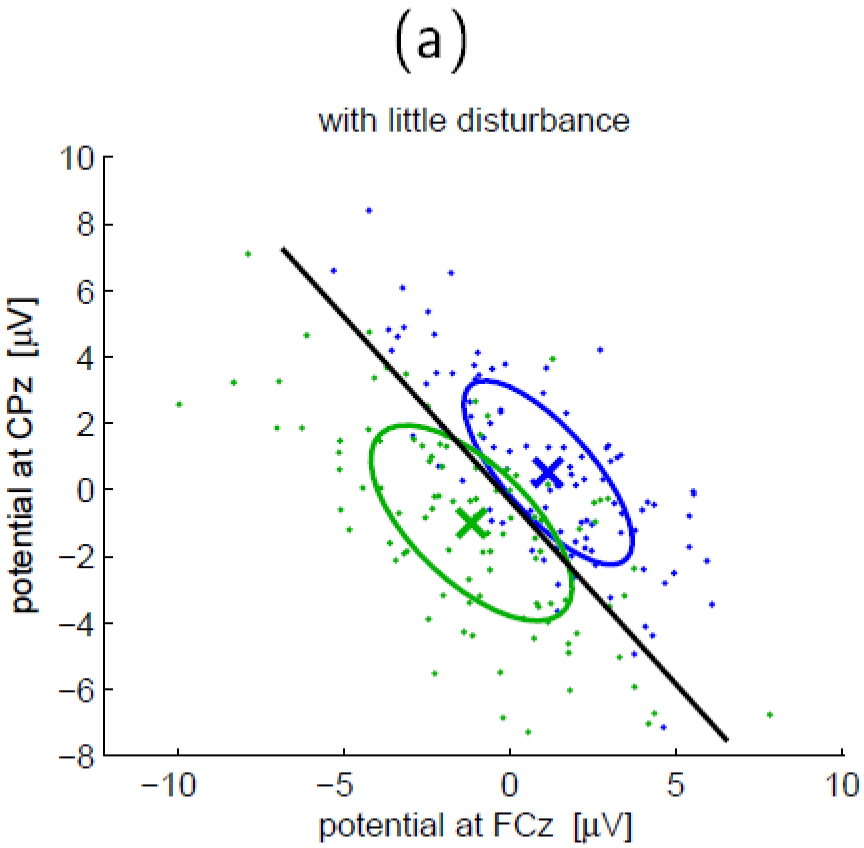


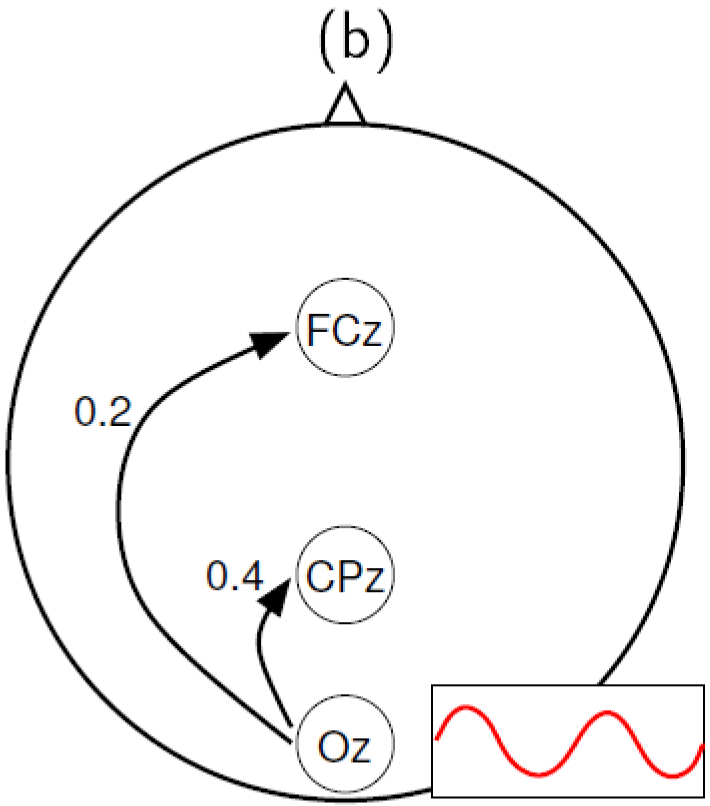
Trial-to-trial variation of P3



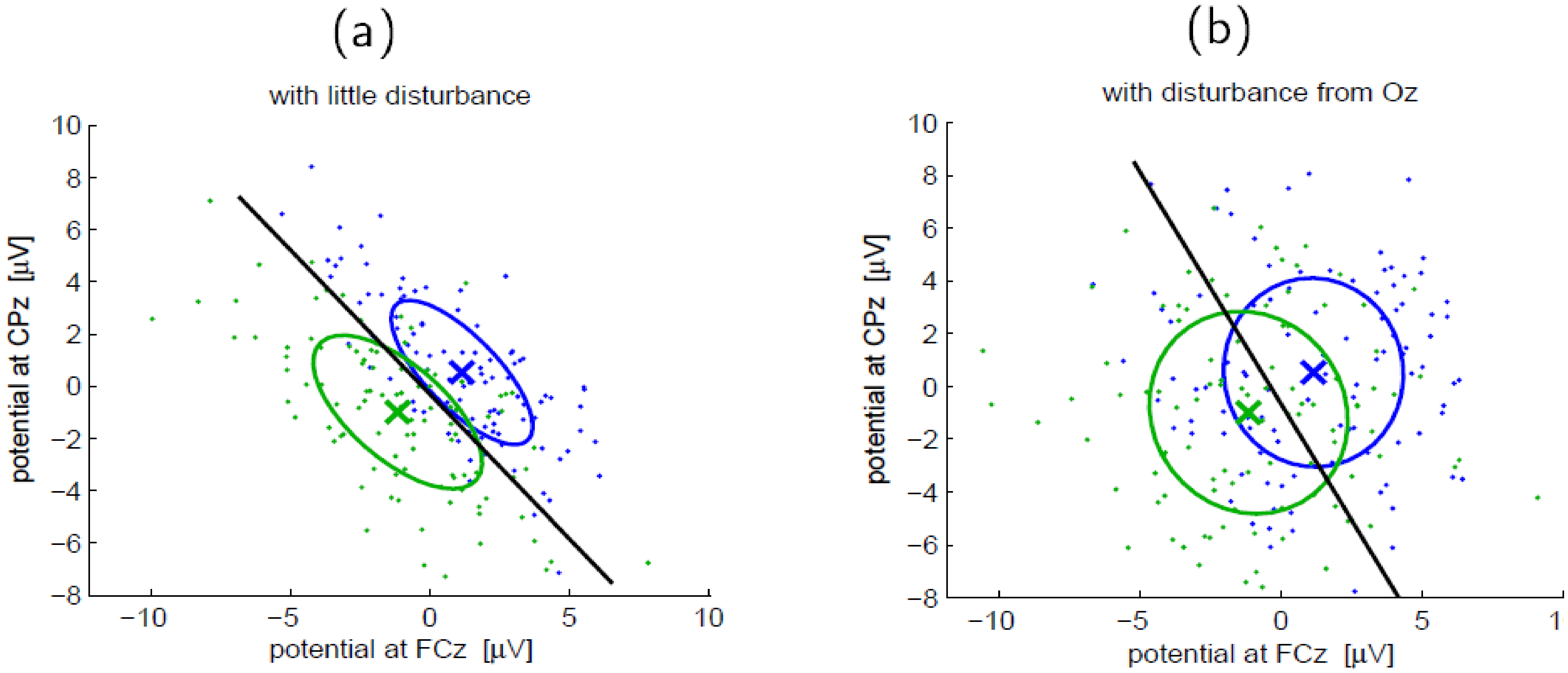
Visual alpha

Understanding spatial filters





Understanding spatial filters II

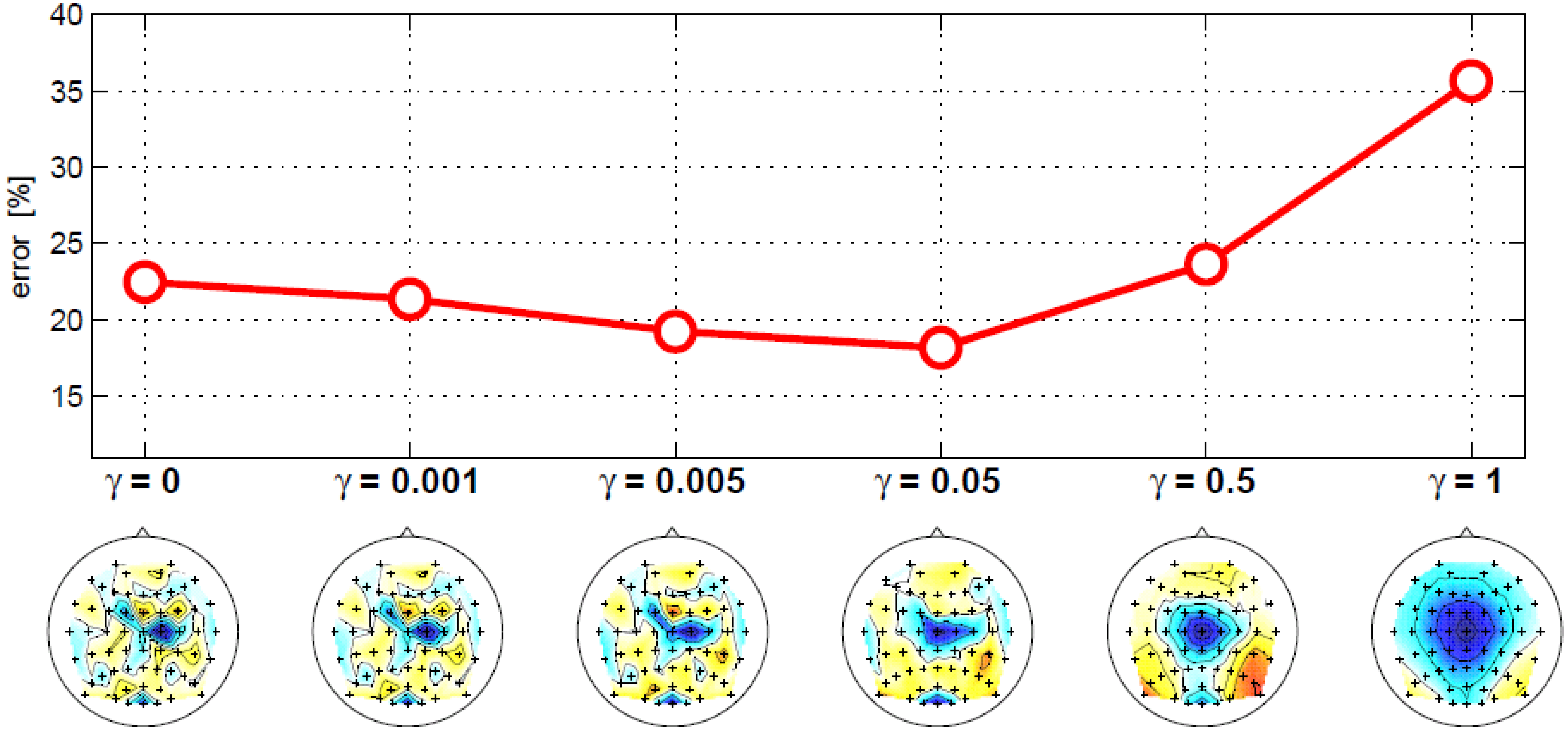


Two channel classification of (a): 15% error, (b): 37% error

When disturbing channel Oz is added to the data (3D): 16% error. Here, channel Oz is required for good classification although itself is not discriminative.

Impact of shrinkage on the spatial filters

With increasing shrinkage, the spatial filters (classifier) look smoother, but classification may degrade with too much shrinkage.



Maps of spatial filters for different values of γ.

Optimal selection of shrinkage parameters

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ be n feature vectors and let $\hat{\mu} = \frac{1}{n}\sum_{k=1}^{n} \mathbf{x}_k$ be the empirical mean.

Aim: get a better estimate of the true covariance matrix $\mathbf{\Sigma}$ (especially in case $n < d$) than the sample covariance matrix $\hat{\mathbf{\Sigma}} = \frac{1}{n-1}\sum_{k=1}^{n}(\mathbf{x}_k - \hat{\mu})(\mathbf{x}_k - \hat{\mu})^\top$ by selecting a γ in

$$\tilde{\mathbf{\Sigma}}(\gamma) := (1-\gamma)\hat{\mathbf{\Sigma}} + \gamma\nu\mathbf{I}.$$

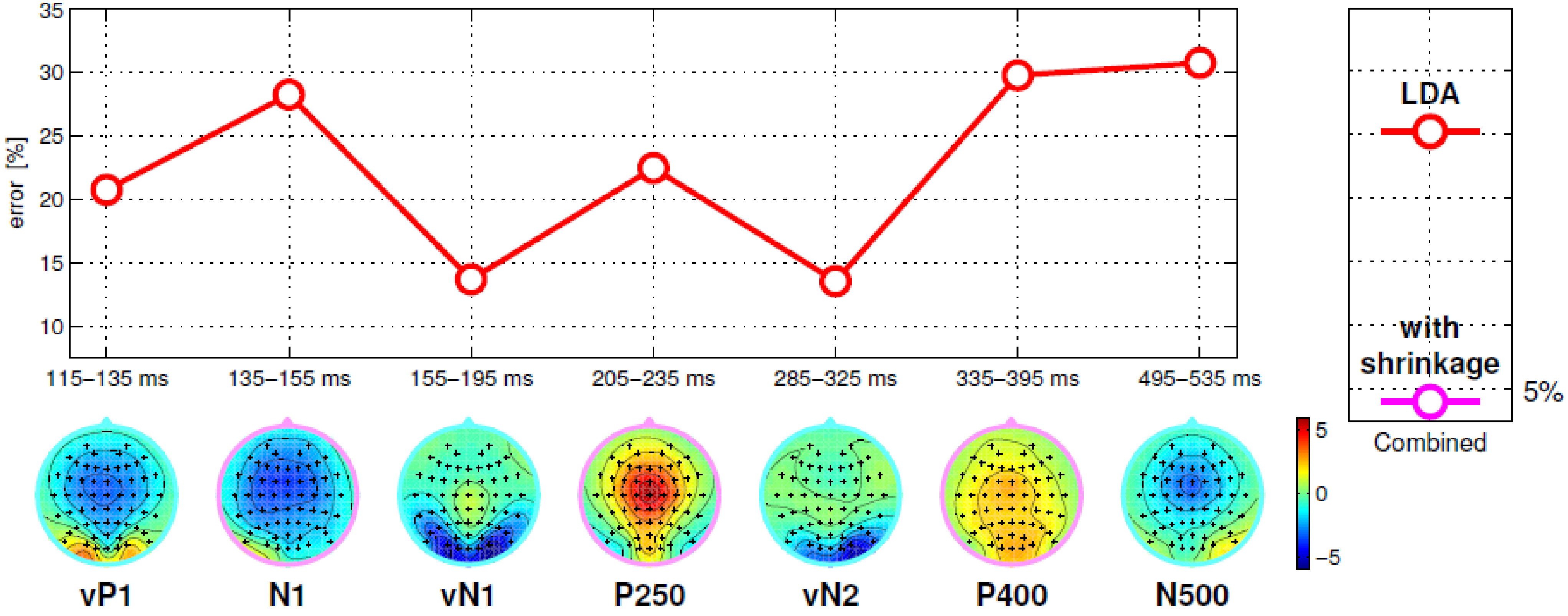
We denote by $(\mathbf{x}_k)_i$ resp. $(\hat{\mu})_i$ the i-th element of the vector $\mathbf{x}_k$ resp. $\hat{\mu}$. Furthermore we denote by s_{ij} the element in the i-th row and j-th column of $\hat{\mathbf{\Sigma}}$. We define

$$z_{ij}(k) = \left((\mathbf{x}_k)_i - (\hat{\mu})_i\right)\left((\mathbf{x}_k)_j - (\hat{\mu})_j\right)$$

Then the optimal shrinkage parameter $\gamma^\star$ for which $\tilde{\mathbf{\Sigma}}(\gamma^\star) = \operatorname{argmin}_{\mathbf{S}} \|\mathbf{S} - \mathbf{\Sigma}\|_F^2$ can be analytically calculated ([2]) as

$$\gamma^\star = \frac{n}{(n-1)^2} \frac{\sum_{i,j=1}^{d} \operatorname{var}_k(z_{ij}(k))}{\sum_{i\neq j} s_{ij}^2 + \sum_i (s_{ii} - \nu)^2}$$

Result of Classification with shrinkage

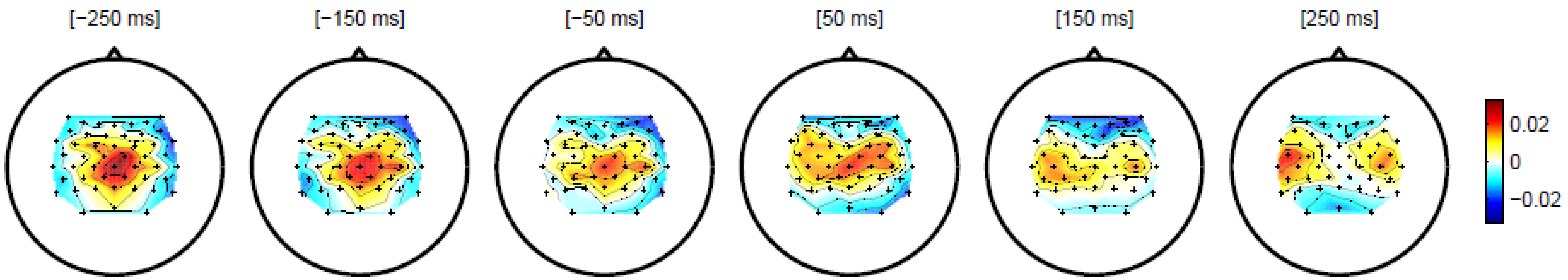


Using shrinkage the classification error could be drastically reduced to 4%.

Summary spatio-temporal classification

- Linear classification with shrinkage is a powerful method.
- Complete shrinkage ($\gamma = 1$) means neglecting the structure of the noise. In this case the classifier is the difference of the ERPs.
- The appropriateness of a linear separation depends on the way features are extracted and transformed.
- In contrast to non-linear classifiers, the weights of a linear classifier are informative.

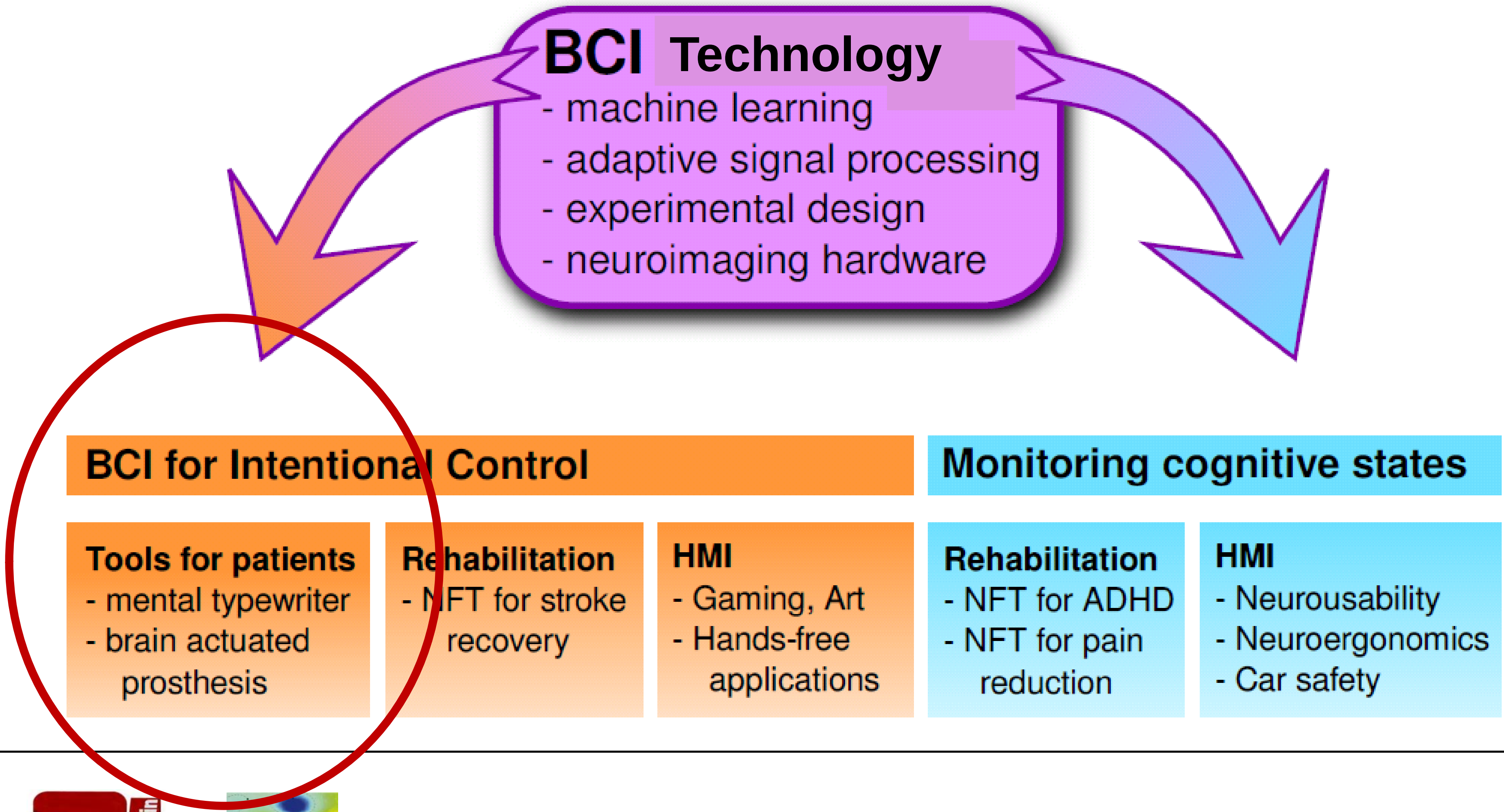
The weights of the trained classifier can be visualized as a sequence of scalp topographies:



Applications

Clinical Applications

Towards industrial applications of BCI Technology



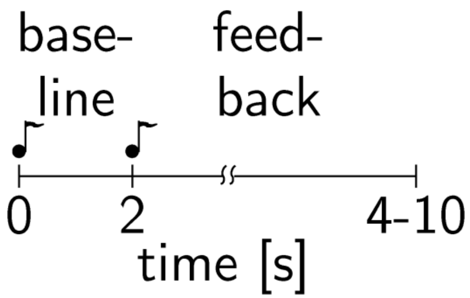
[Blankertz et al 2010 Front. Neurosci.]

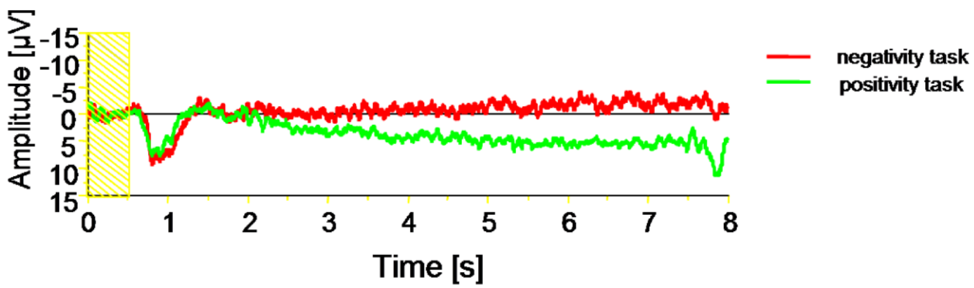
Operant conditioning: Tübingen Group

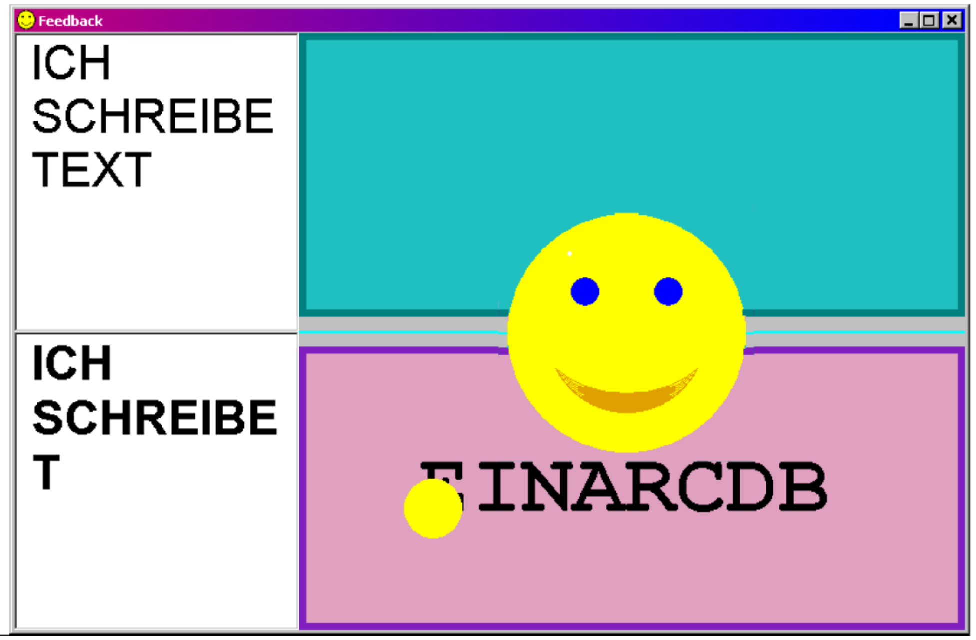
The **slow cortical potentials (SCPs)** at central scalp position can be voluntary controlled. But this learning process might require many training sessions.

The yellow ball travels at a constant speed from left to right, vertically controlled by SCPs. When the ball reaches the right border one of the targets gets selected.

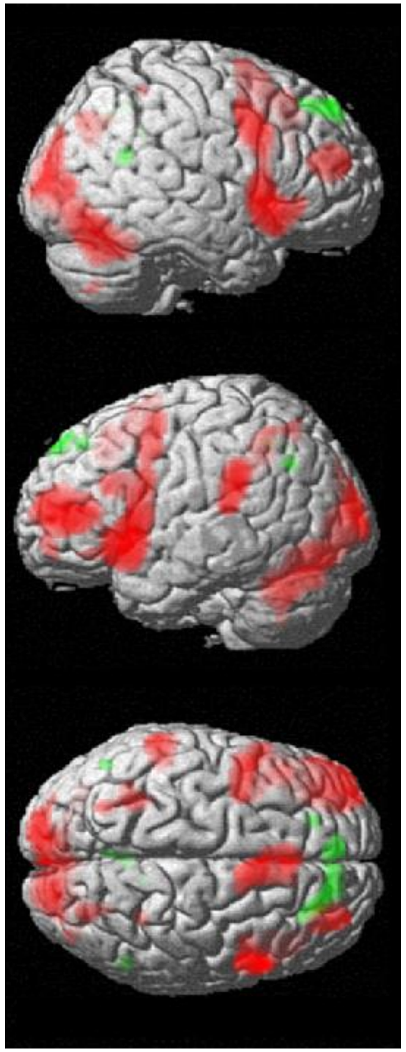
When an acceptable accuracy is reached after some training sessions, subjects are switched to a language support program.



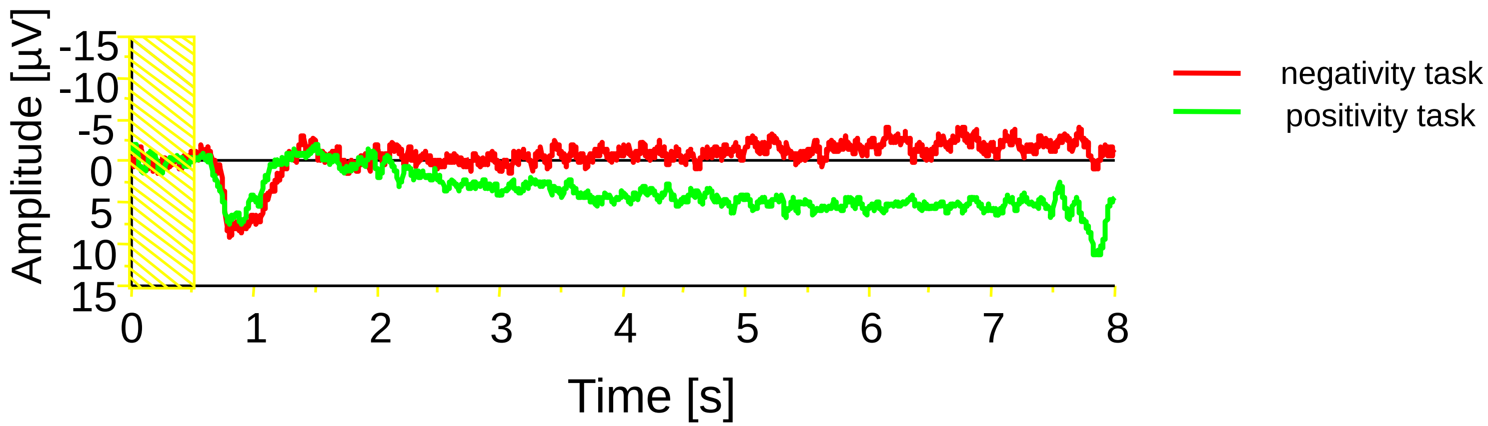




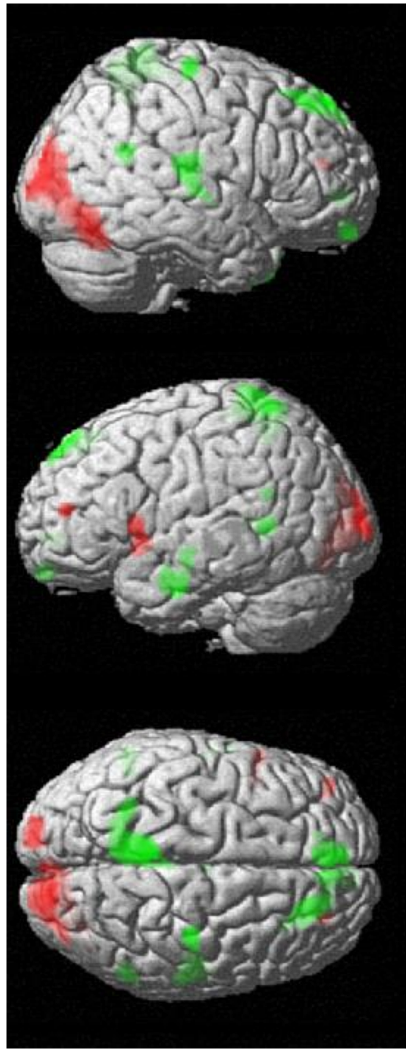
Non-Invasive: Tübingen. Birbaumer Lab: Slow Cortical potentials



Negativity task



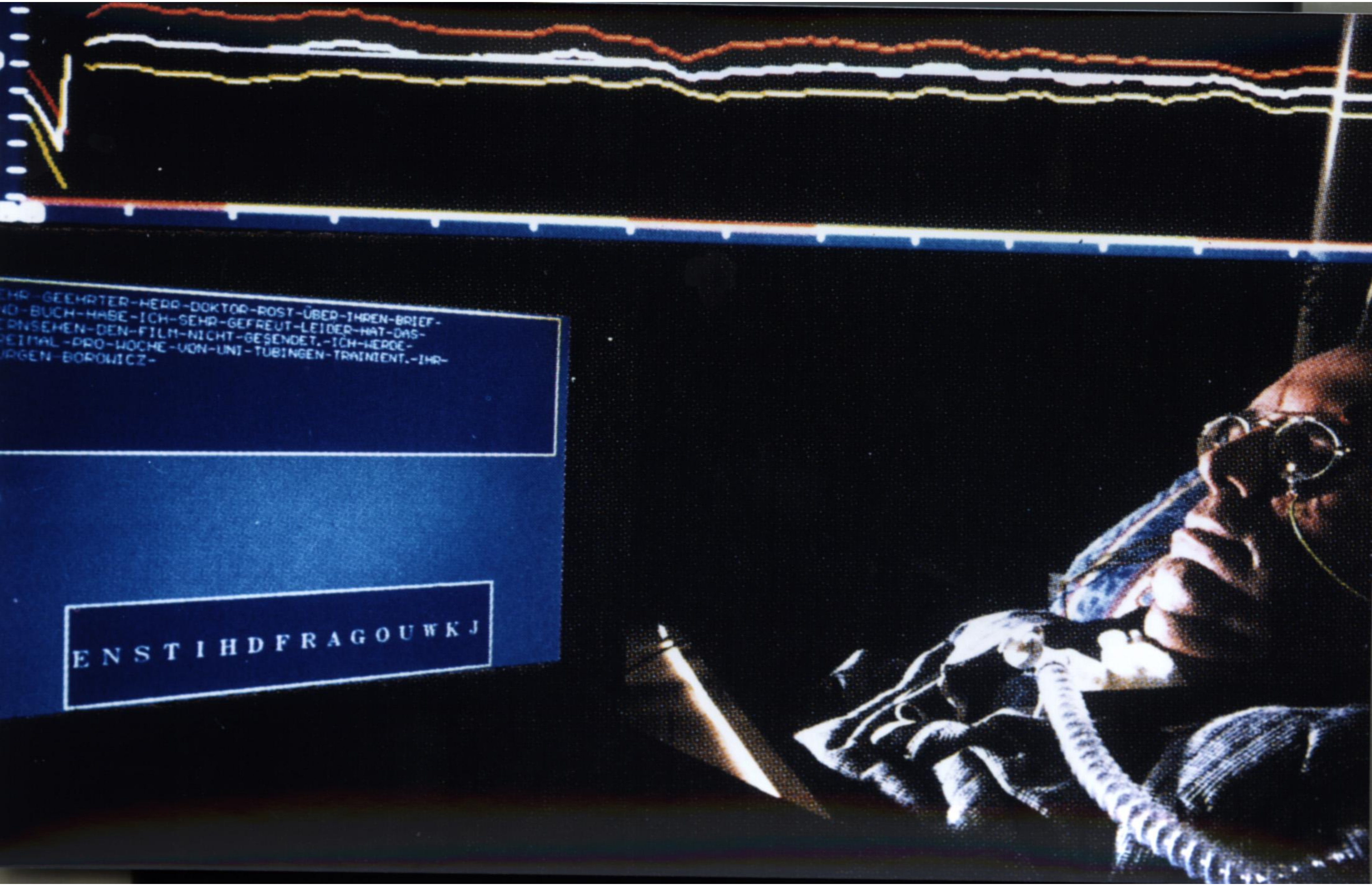
Positivity task



[From Birbaumer et al.]

[From Birbaumer et al.]

SCP



[From Birbaumer et al.]

ERFAHRUNGEN-EINES-TTD-SCHREIBERS-BEIM-SCHREIBEN

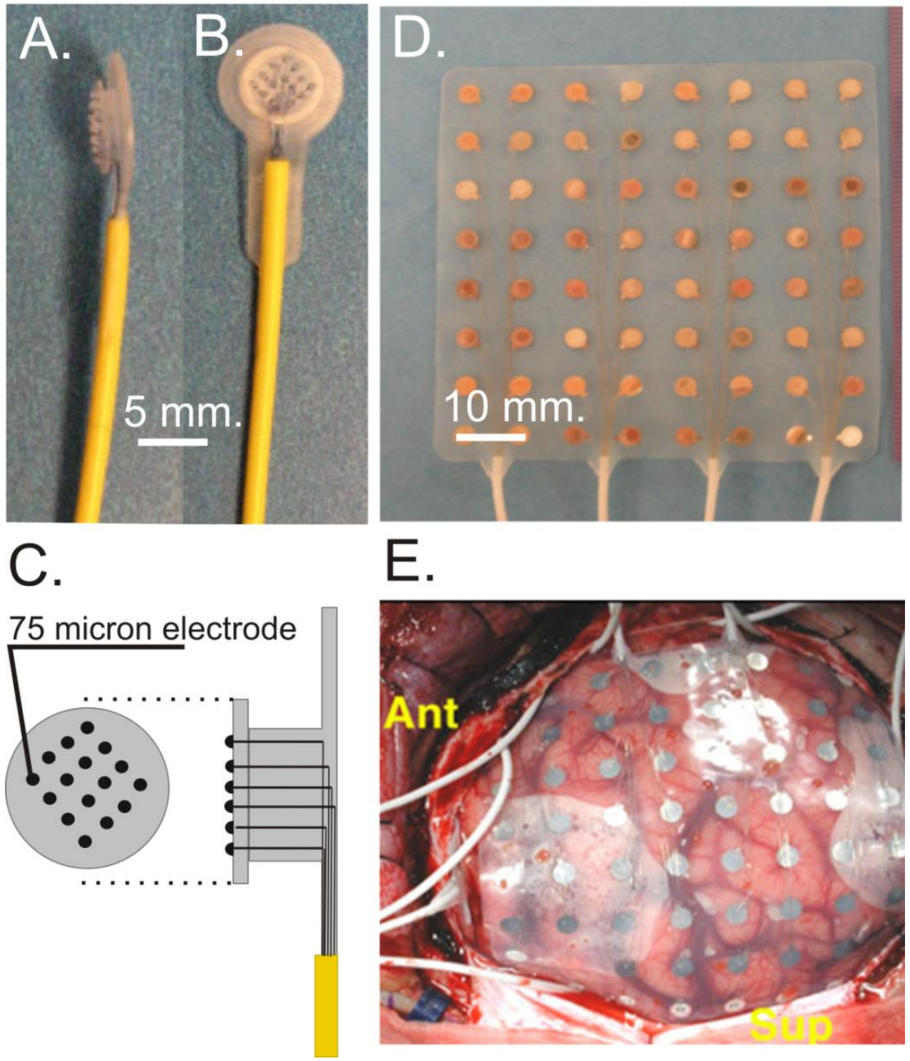
ZUERST-MÖCHTE-ICH-KURZ-NOCH-EINMAL-DIE-STRATEGIE-SCHILDERN,DIE-ICH-MIR-ZURECHT-GELEGT-HABE..BEIM-POSITIVIEREN-VERSUCHE-ICH-IN-DER-FEEDBACK-PHASE-DRUCK-IM-GEHIRN-ZU-ERZEUGEN,-WOHINGEGEN-BEIM-NEGATIVIEREN-ICH-ÜBER-ENTSPANNUNG--DES-GEHIRNS-SOWOHL-IN-DER-BASELINE-PHASE-ALS-AUCH-IN-DER-ANSCHLIESSENDEN-FEEDBACK-PHASE-EINE-ART-GEDANKLICHE-LEERE-HERZUSTELLEN-VERSUCHE.

UM-DRUCK-ZU-ERZEUGEN,BENUTZE-ICH-VERSCHIEDENE-DENKHILFEN.SO-STELLE-ICH-MIR-MIT-DEM-TACK-TON,-ALSO-MIT-BEGINN-DER-FEEDBACK-PHASE,VOR,-DASS-EINE-AMPEL-AUF-GRÜN-UMSPRINGT-ODER-EIN-LEICHTATHLET-MIT-DEM-STARTSCHUSS-LOSRENNT-ODER-EIN-PFEIL-VOM-GESPANNTEN-BOGEN-SCHNELLT-ODER-DER-CURSOR-INS-BUCHSTABENFELD-SPRINGT.-VORAUSSETZUNG-FÜR-DIE-WIRKSAMKEIT-DER-DENKHILFEN-IST-ALLERDINGS,-DASS-ICH-WÄHREND-DER-BASELINE-PHASE-GENÜGEND-SPANNUNG-BZW.-ERWARTUNG-AUFBAUE.IM-AMPELFALL-STELLE-ICH-MIR-DIE-GELBPHASE-VOR,BEIM-PFEILBILD-DAS-SPANNEN-DES-BOGENS,USW.DIESER-VON-GEDANKENBILDERN-HERVORGERUFENE-SPANNUNGSAUFBAU-LÄSST-SICH-IM-HIRN-LOKALISIEREN-,-UND-ZWAR-VON-DER-ZENTRALEN-ELEKTRODE-CZ-IN-RICHTUNG-ZU-ELEKTRODE-FZ.-DESHALB-NENNE-ICH-DAS-MAL-DIE-PHYSIOLOGISCHE-GRUNDLAGE-DES-OBEN-ERWÄHNTEN-DRUCKS.-DIE-ENTSTEHUNG-DER-BILDER-SELBST-LÄSST-SICH-DAGEGEN-NICHT-LOKALISIEREN-,SIE-KOMMEN-INSOFERN-AUS-DEM-NICHTS.-IHRE-VERWENDUNGSWEISE-ENTZIEHT-SICH-ABER-EINER-KONSTANTEN,UNBEGRENZTEN-UND-UNMITTELBAREN-WIEDERHOLBARKEIT.-DAS MAG AN MANGELNDER KONZENTRATION LIEGEN ODER AN DER FLÜCHTIGEN STRUKTUR VON GEDANKEN UND BILDERN,DIE LETZTLICH ZUR FOLGE HAT,DASS KEIN GEDANKE ODER BILD GLEICHEN INHALTS IDENTISCH REPRODUZIERT WERDEN KANN. ALS FEHLERQUELLE IST ZUNÄCHST DER UNPRÄZISE GEBRAUCH DER BILDER IN DER BASELINE FESTZUSTELLEN, WOBEI UNPRÄZIS VOR ALLEM ZU KURZ UND UNSCHARF BEDEUTET.MITUNTER VERGESSE ICH AUCH DIE ANWENDUNG DER BILDER IM EIFER DES KAMPFES MIT DEN BUCHSTABEN , WAS MEIST MIT EINER SUBJEKTIVEN ZEITVERKÜRZUNG EINHERGEHT,IN DER BEIDE PHASEN INEINANDER VERSCHWIMMEN.

[From Birbaumer et al.]

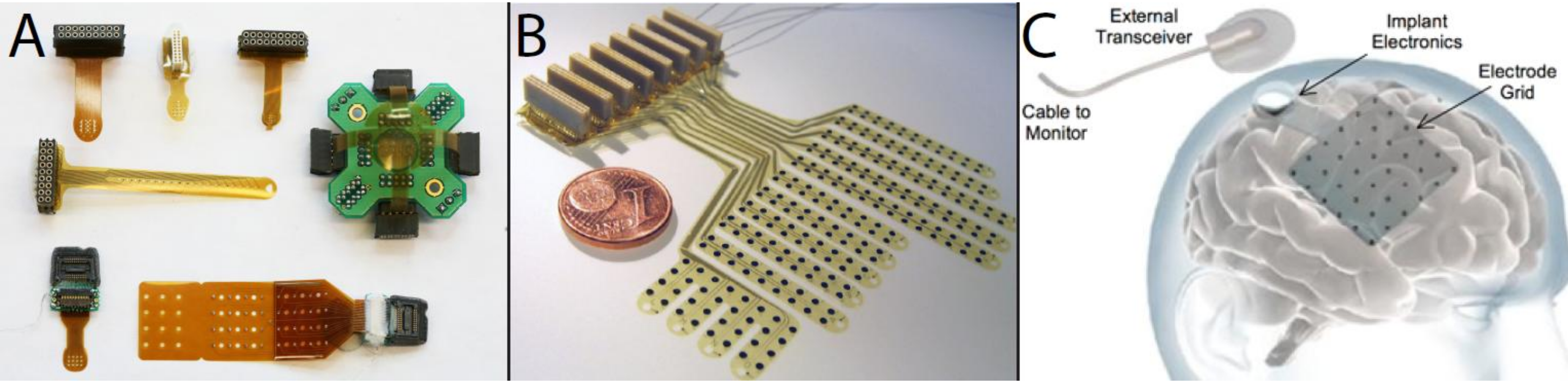
ECOG Decoding

ECOG

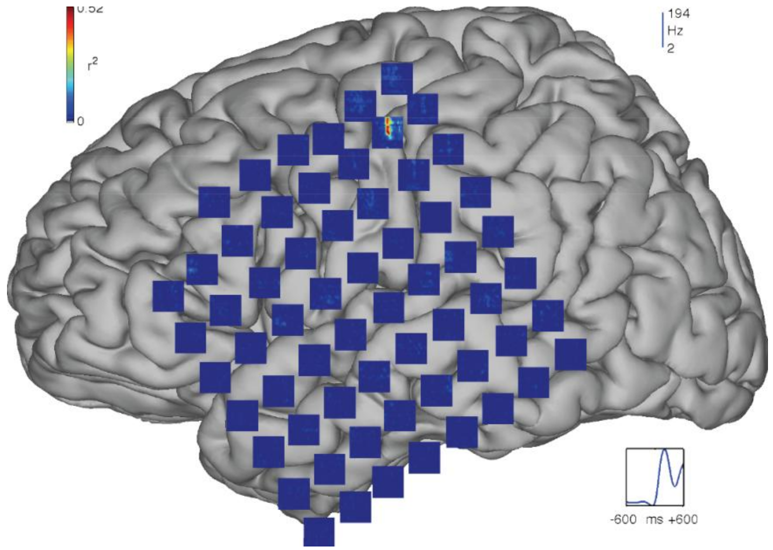


- presurgical localization of area causing epilepsy
- excellent possibilty to learn about brain for human subject

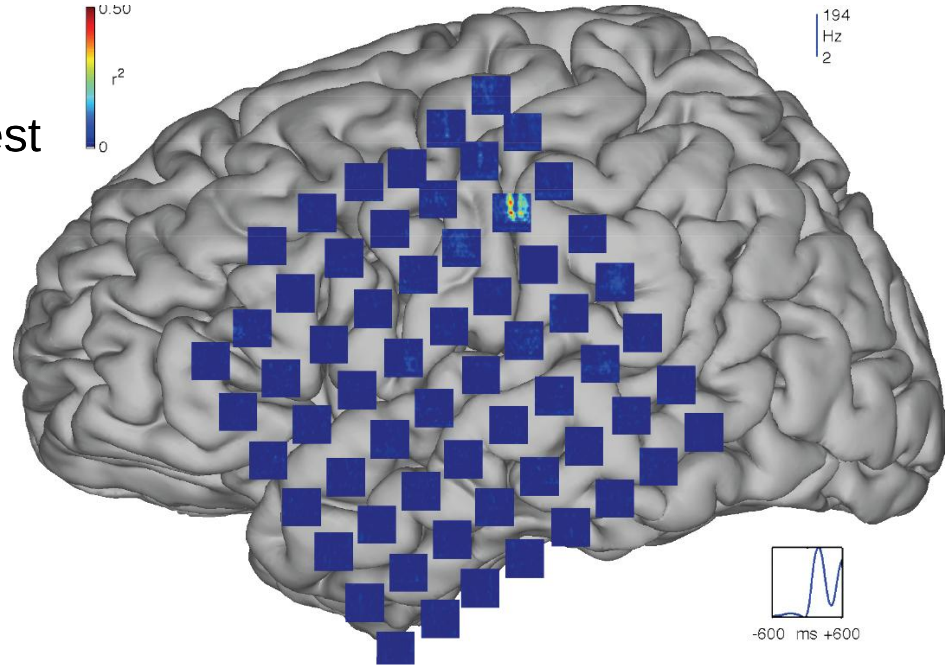
[From Schalk]



Index vs rest



Thumb vs rest



[From Schalk]

ECOG Analysis

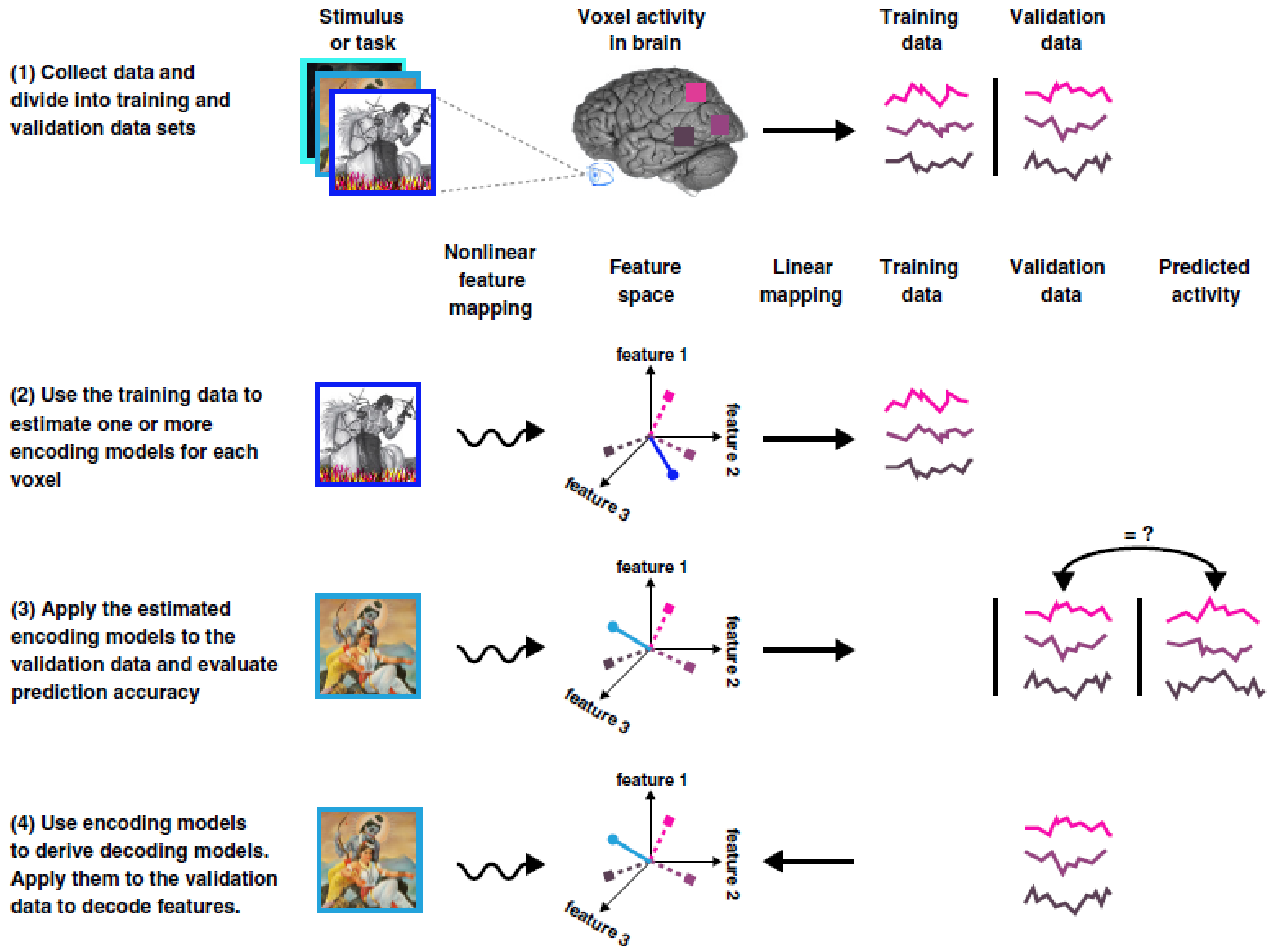
[From Schalk]

fMRI Decoding

Example: Which Video are you watching?

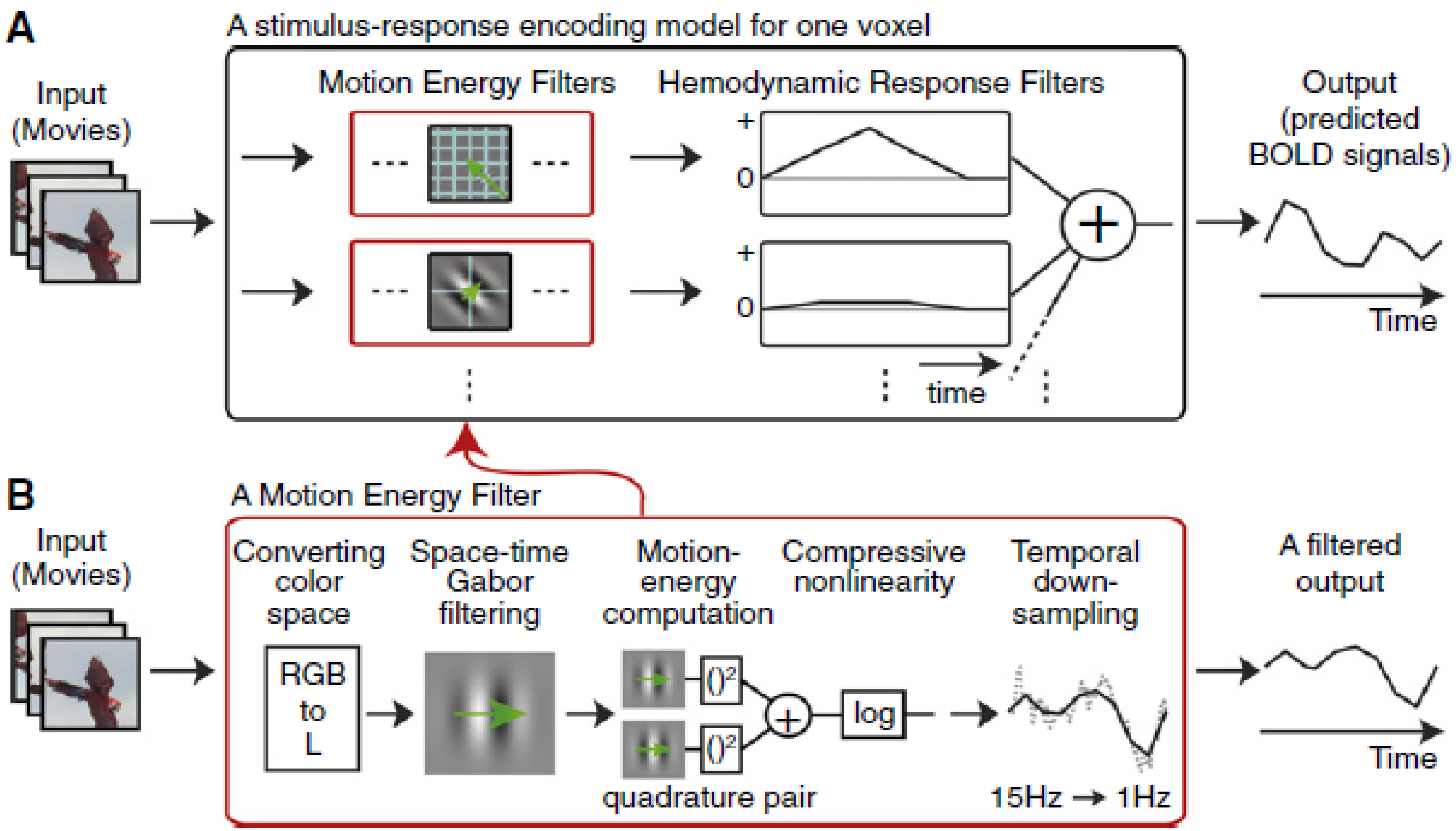
- Study: Reconstructing Visual Experience from Brain Activity Evoked by Natural Movies (Nishimoto 2011)
- Aim: validation of neurovascular coupling in the visual cortex

- Models of hemodynamics elicited by a movie for each voxel in early visual areas
- fMRI measurement of subjects watching movies
- Reconstruction of movies from the brains‘ activity



Stimulus or task
Voxel activity in brain
Training data
Validation data
(1) Collect data and divide into training and validation data sets
Nonlinear feature mapping
Feature space
Linear mapping
Training data
Validation data
Predicted activity
(2) Use the training data to estimate one or more encoding models for each voxel
feature 1
feature 2
feature 3
= ?
(3) Apply the estimated encoding models to the validation data and evaluate prediction accuracy
feature 1
feature 2
feature 3
(4) Use encoding models to derive decoding models. Apply them to the validation data to decode features.
feature 1
feature 2
feature 3

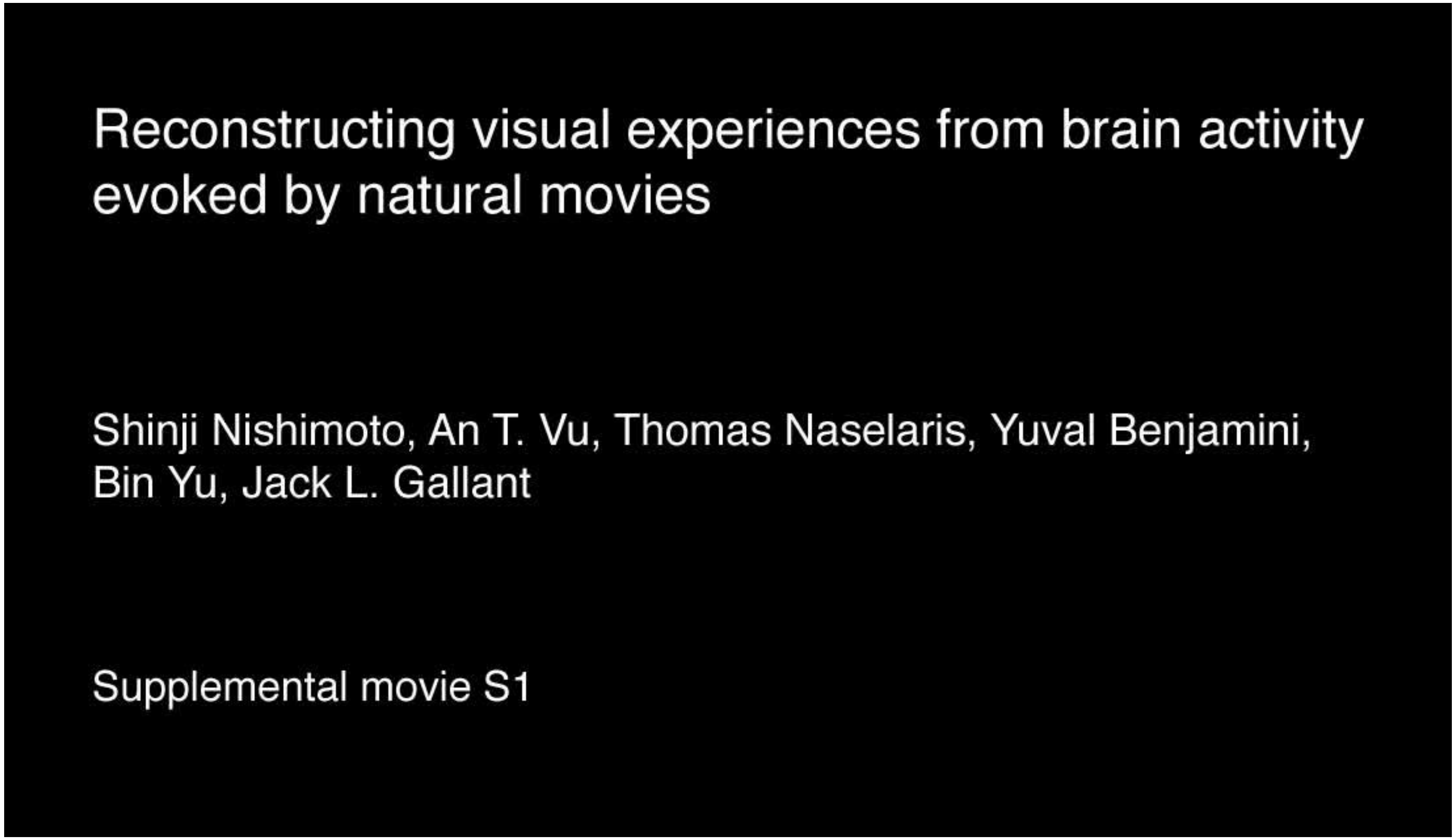
Example: Which Video are you watching?



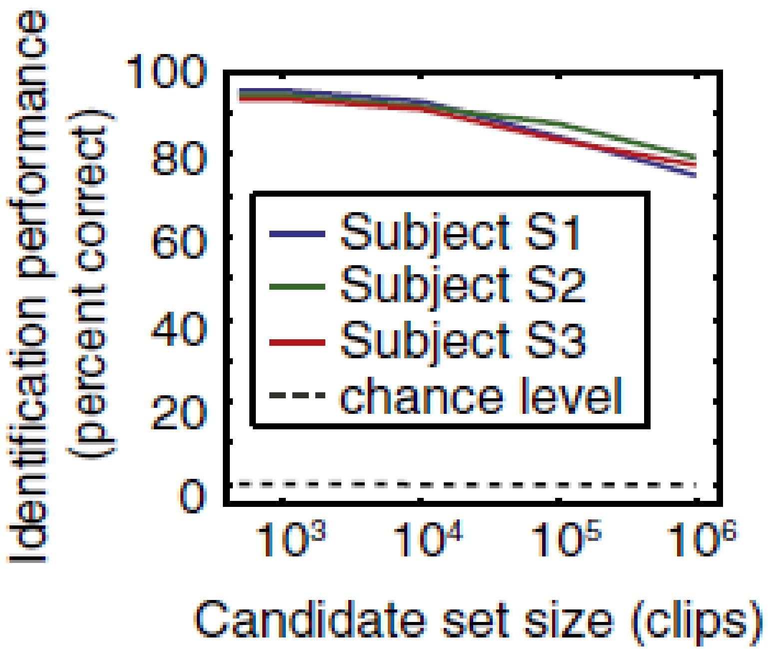
Motion energy of the pictures were calculated and fed to hemodynamic modeling

Example: Which Video are you watching?

- Bayesian fit to acquired data of 3 subjects watching 12 movie (each once)
- Test the approach on subject watching 9 other movies (each 10 times)

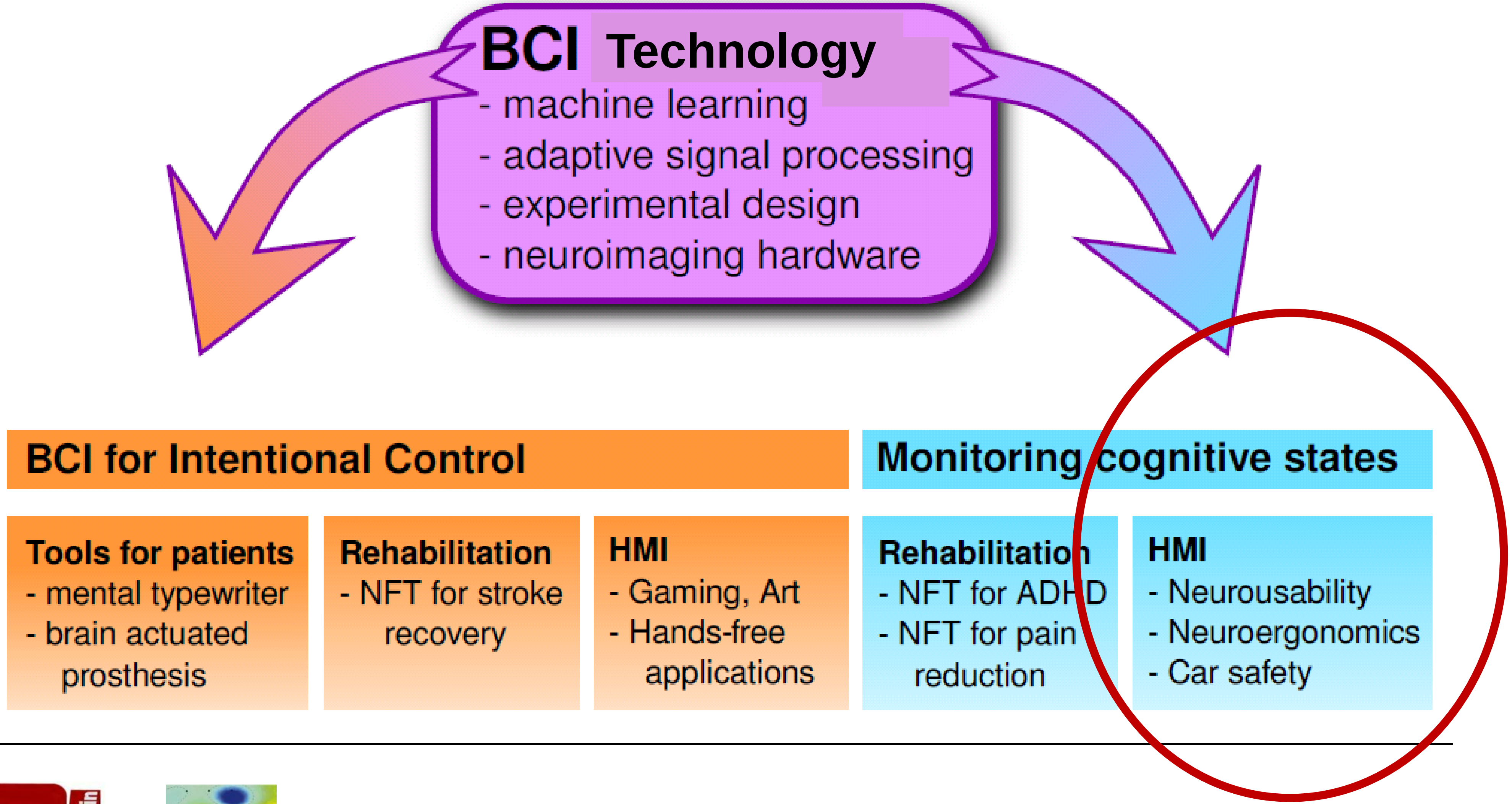


Example: Which Video are you watching?



The accuracy becomes worse when more films are included for decoding (not watched by the subjects) but remains high

Towards industrial applications of BCI Technology



[Blankertz et al 2010 Front. Neurosci.]

BCI for Assessing Signal Quality perception

Why Quality Assessment?

- Ensure user satisfaction
- Develop better compression algorithms

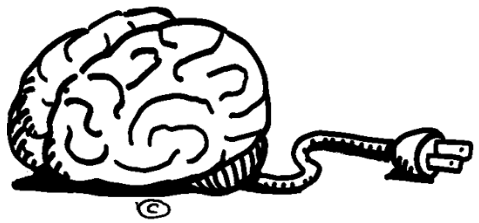
signal quality

profitability

Approaches

Behavioral tests
(standard)

EEG + BCI methods
(novel)



- Continuous signal
- Objective measure
- Capture
  > subtle differences
  > non-conscious processing

EEG Studies

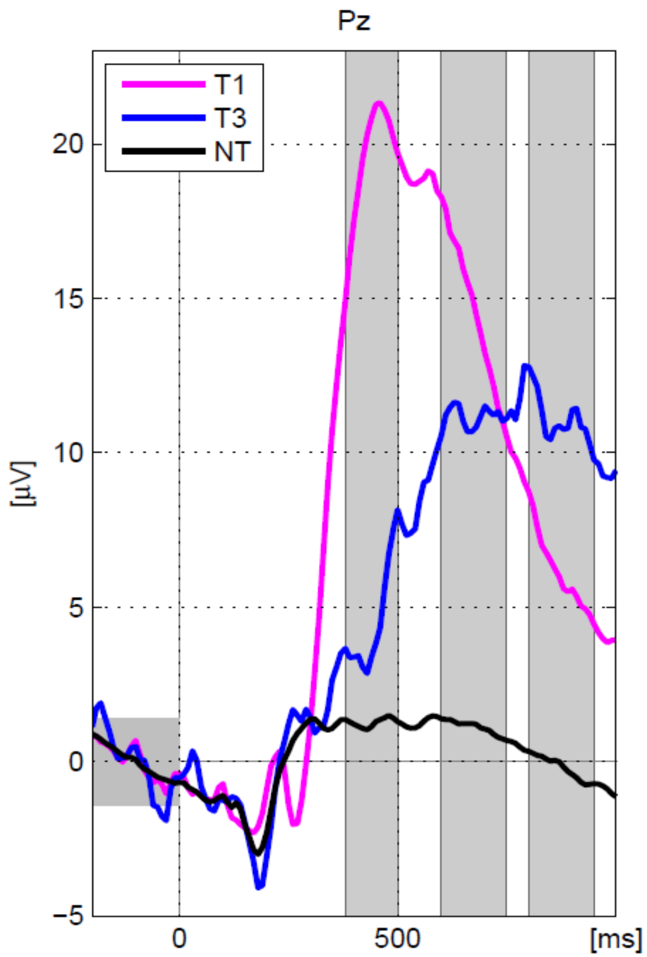
Domain	Stimuli	Cooperation Partner
Auditory	Phonemes	Telekom Laboratories
	Words	- " -
Visual	Flickering light	Philips Research
	Video	Fraunhofer (HHI)

Audio Quality

- Discrimination task: Is stimulus disturbed?
- Recording: button press, 64-channel EEG
- Stimuli:
  - 4 levels of degradation: strong (T1) – weak (T4),
  - undisturbed stimulus (NT)

	Phoneme Study	**Word Study**
stimulus	/a/	/Haus/, /Schild/ by female/male speaker
disturbed by	signal-correlated noise	bit rate limitation

Audio Quality



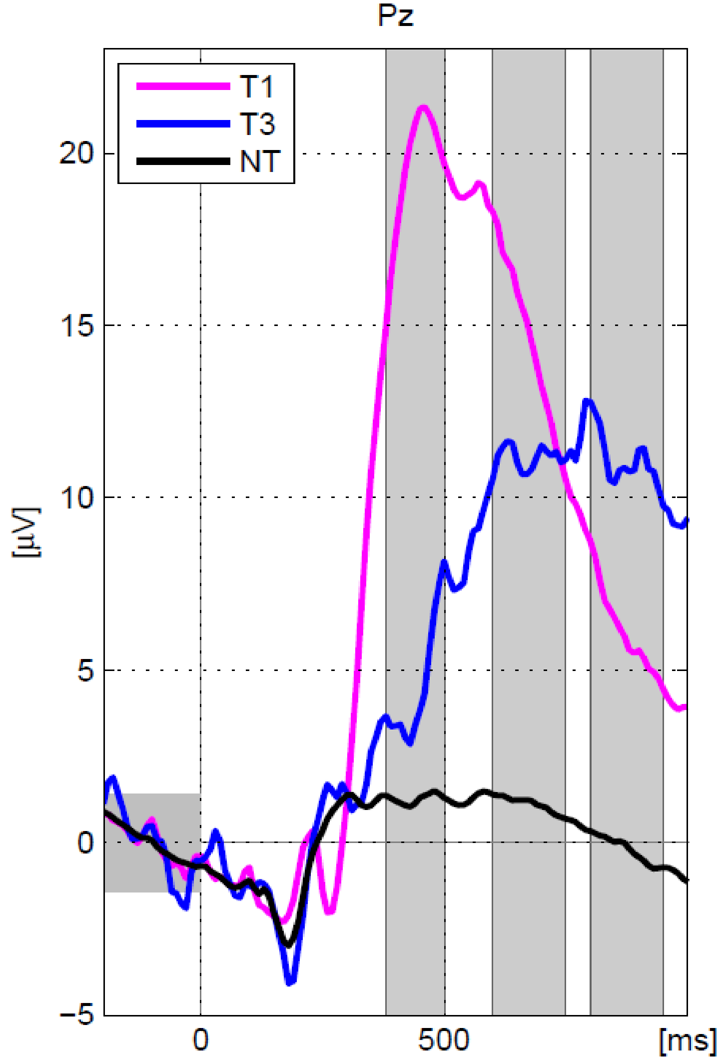
- Hits:
  The more subtle the noise, the lower the amplitude and the higher the latency of P3 component

  → 'Neural effort'
  → Quantification of hits

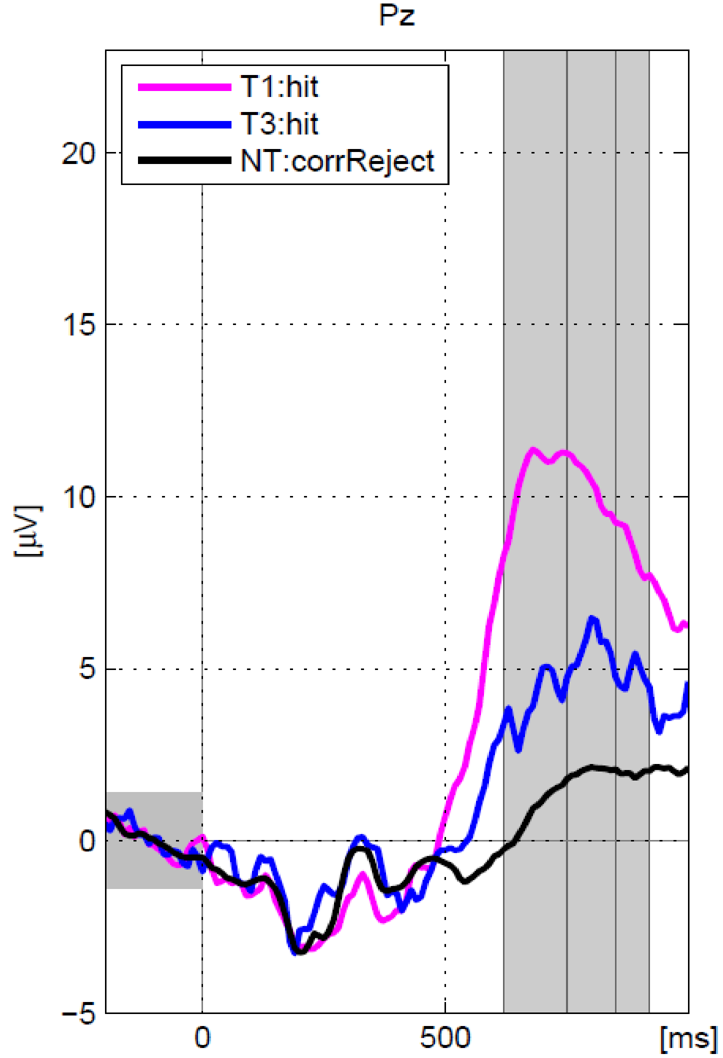
Grand average EEG signal (ERP):
stimulus T1 (strong degradation), T3 (weak degradation), NT (undisturbed).

Audio Quality

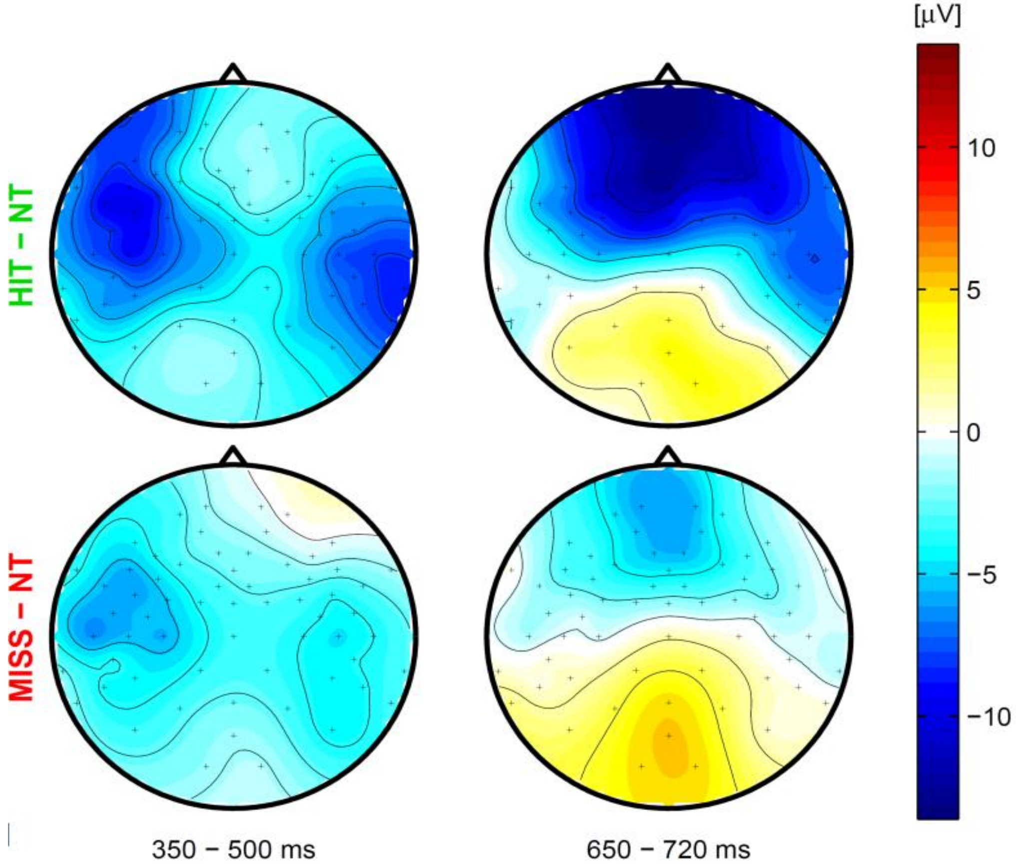
Phonemes



Words



Audio Quality



- Misses:
  Similarity to hits at the threshold of perception

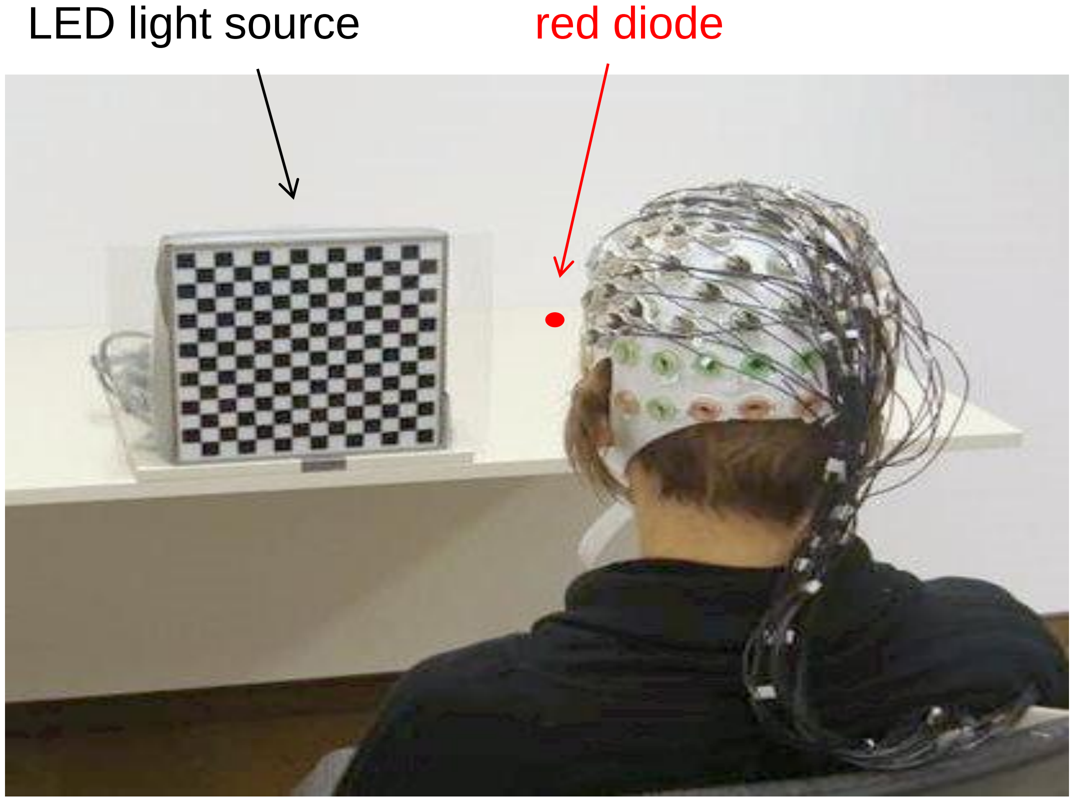
  → Non-conscious processing

  → Quantified by linear classification

Difference topographies at the threshold of perception:
hits / misses (low quality) – correct rejections (high quality)
(one participant, phonemes)

Visual Quality

- Discrimination task: Does the stimulus flicker?
- Recording: 64-channel EEG, button press
- Stimuli:
  - Constant wave light (CW)
  - 4 levels of flicker frequency:
    slow (S1) – fast (S4)



Visual Quality

- Added value of **EEG**

	Detected vs CW	Undetected vs CW		
Participant	S1	S2	S3	S4
VPdbe	40 Hz	60 Hz	83 Hz	95 Hz
VPik	50 Hz	70 Hz	85 Hz	100 Hz
VPdbf	50 Hz	70 Hz	85 Hz	100 Hz
VPdbd	40 Hz	50 Hz	60 Hz	70 Hz
VPow	50 Hz	70 Hz	85 Hz	100 Hz
VPfat	50 Hz	70 Hz	85 Hz	100 Hz

Stimulation frequencies [Hz] per participant; colored cells: significant neural response

- Orange: shown by EEG (t-test, univariate)

Visual Quality

- Added value of **EEG** and **ML**

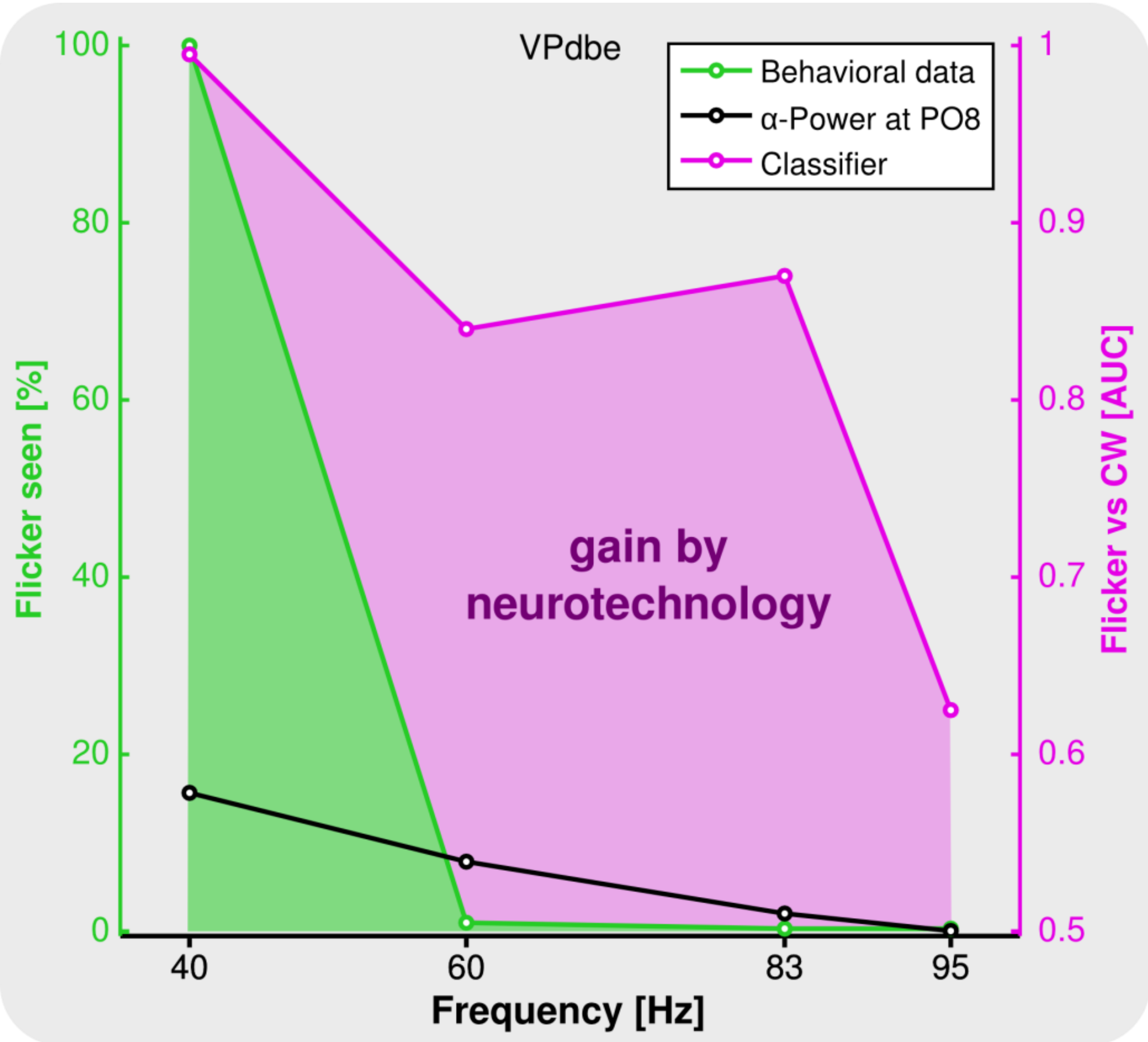
	Detected vs CW	Undetected vs CW		
Participant	S1	S2	S3	S4
VPdbe	40 Hz	60 Hz	83 Hz	95 Hz
VPik	50 Hz	70 Hz	85 Hz	100 Hz
VPdbf	50 Hz	70 Hz	85 Hz	100 Hz
VPdbd	40 Hz	50 Hz	60 Hz	70 Hz
VPow	50 Hz	70 Hz	85 Hz	100 Hz
VPfat	50 Hz	70 Hz	85 Hz	100 Hz

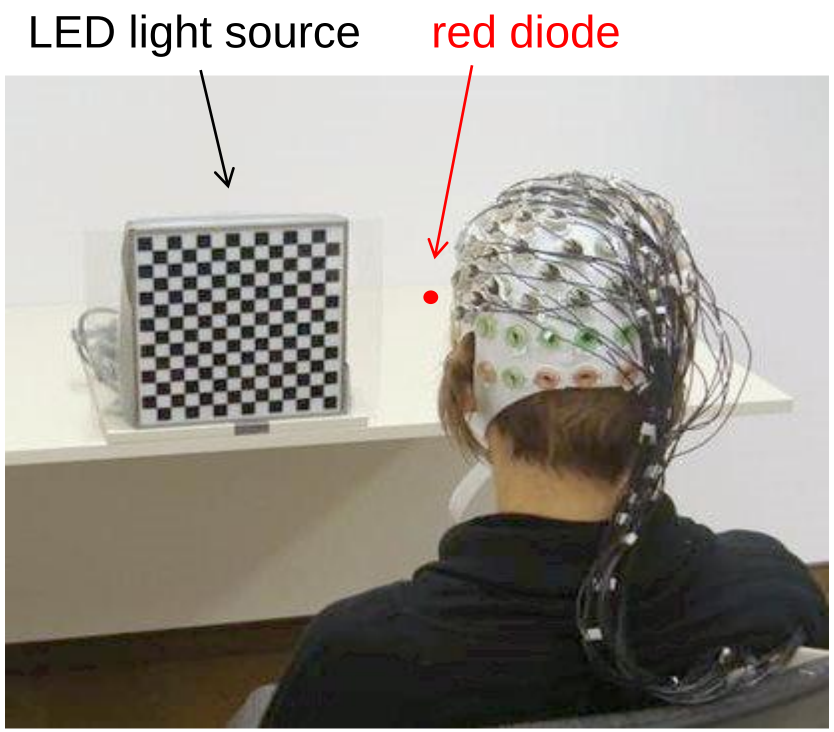
Stimulation frequencies [Hz] per participant; colored cells: significant neural response

- Yellow + orange: shown by ML (CSP+LDA, multivariate)

Visual Quality Gain by NT

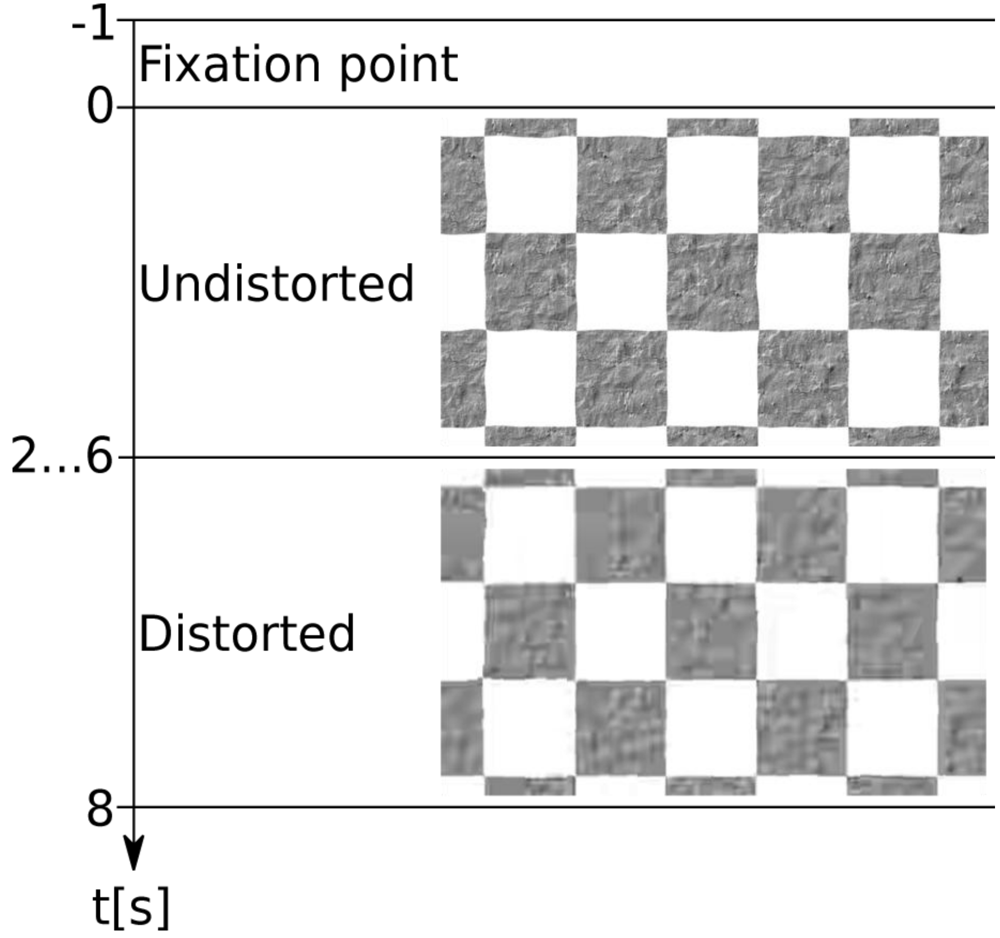
- Discrimination task: Does the stimulus flicker?
- Stimuli: slow (S1) – fast (S4) flfr & CW



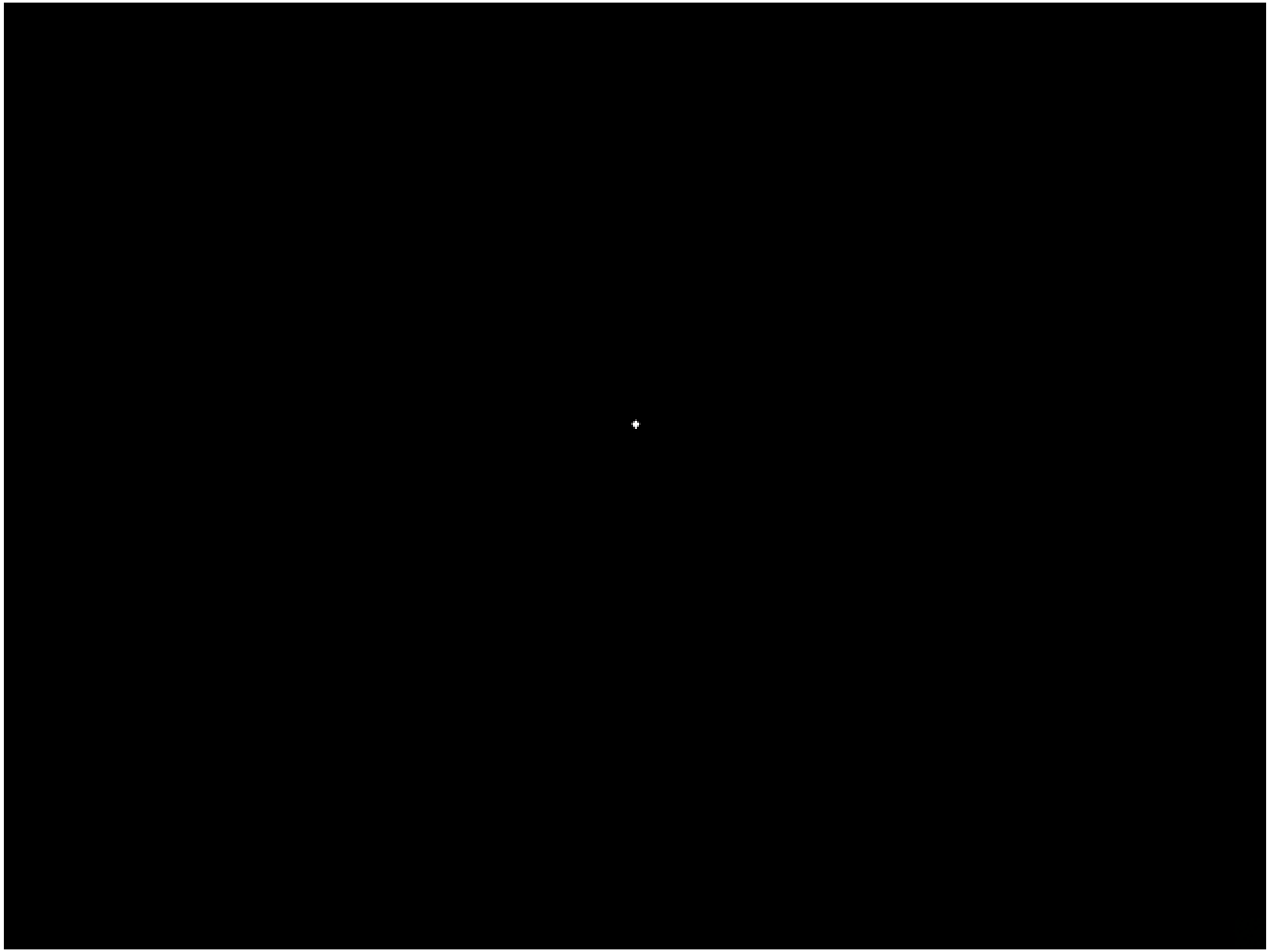


Video Quality

- Detection task: Does the quality change in the video?
- Stimuli:
  - artificially generated videos (8 sec) with a quality change
  - Undistorted baseline (BL), 8 levels of distortion (S1-8)
- Recording:
  64-channel EEG,
  button press



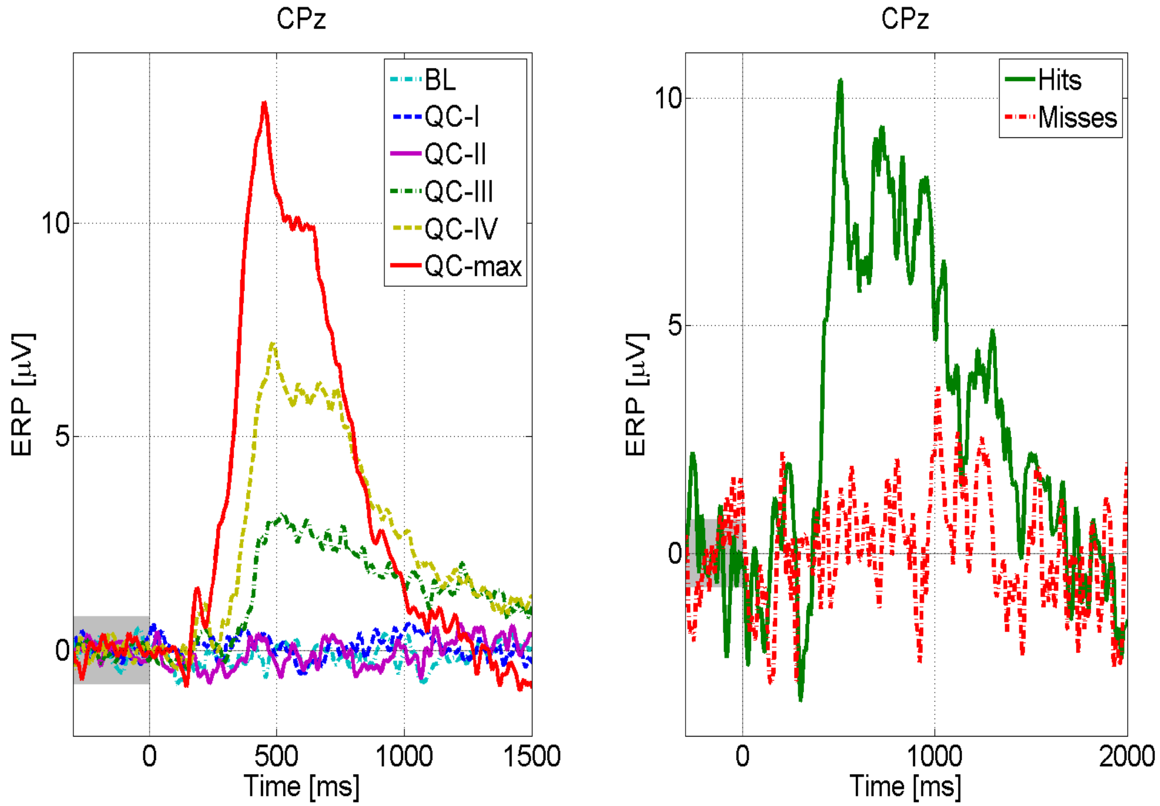
Video Quality



Video Quality

- P3 component is a graded neural index of quality perception (left)
- Effect depends on subjective perception (right)

- Non-conscious processing in 3 out of 11 participants



Summary

Audio Quality

Neuronal effort:
loss of quality is reflected in P3 latency/amplitude

- Non-Conscious Processing.
  use classification to single out trials where misses resemble hits

Visual Quality

Non-Conscious Processing:
high-frequency flicker can still elicit a neural response, even if it is not noticed behaviorally

- Machine Learning:
  classification reveals effect for additional participants and stimuli

BCI for Assessing Workload

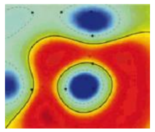
Nonclinical Application: tiredness monitoring

Application: Cognitive workload and drowsyness assessment

Assess workload with BCI and balance it by smart driver assistent system

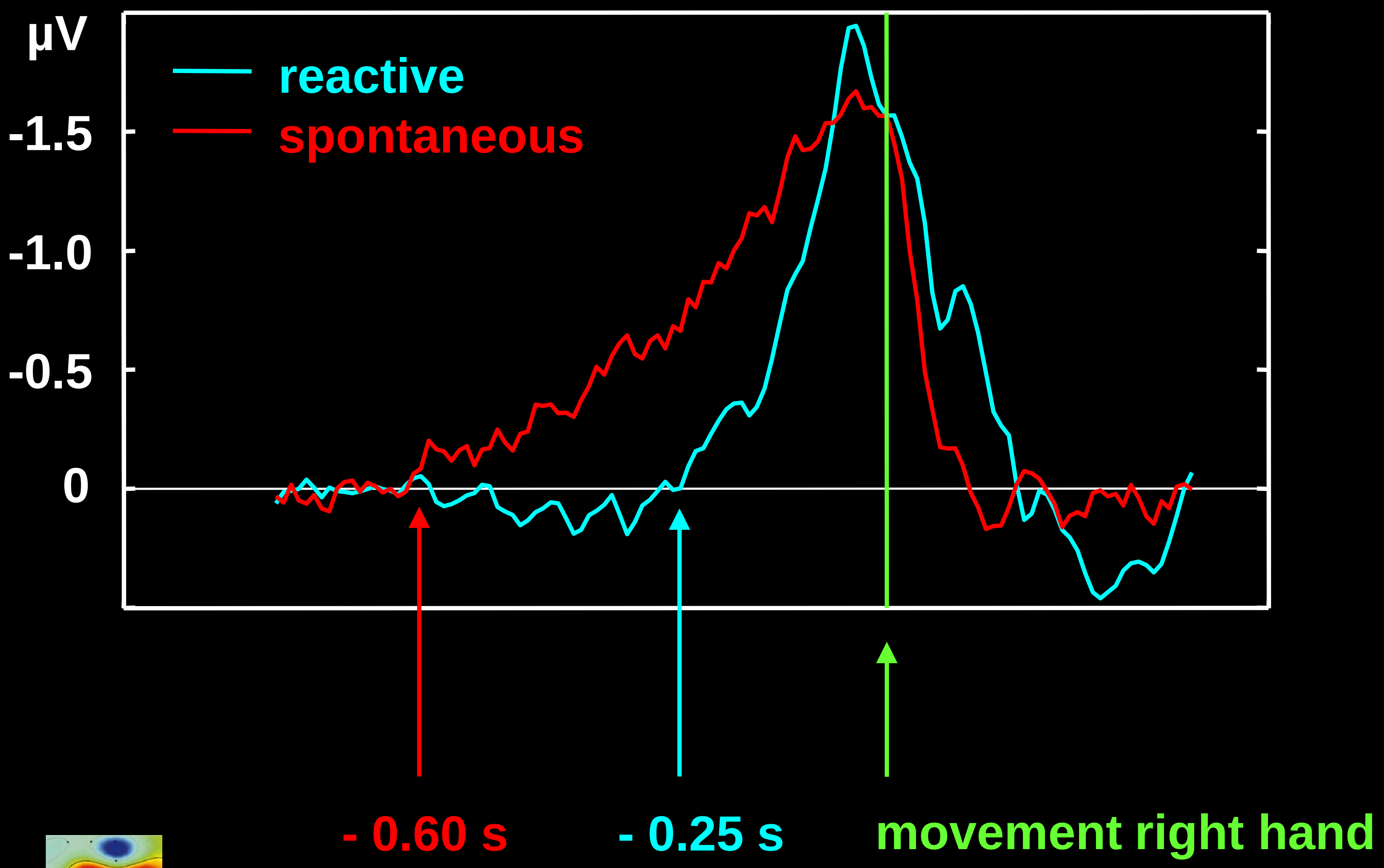
Assess cognitive alertness

[Kohlmorgen, Müller et al 2007]

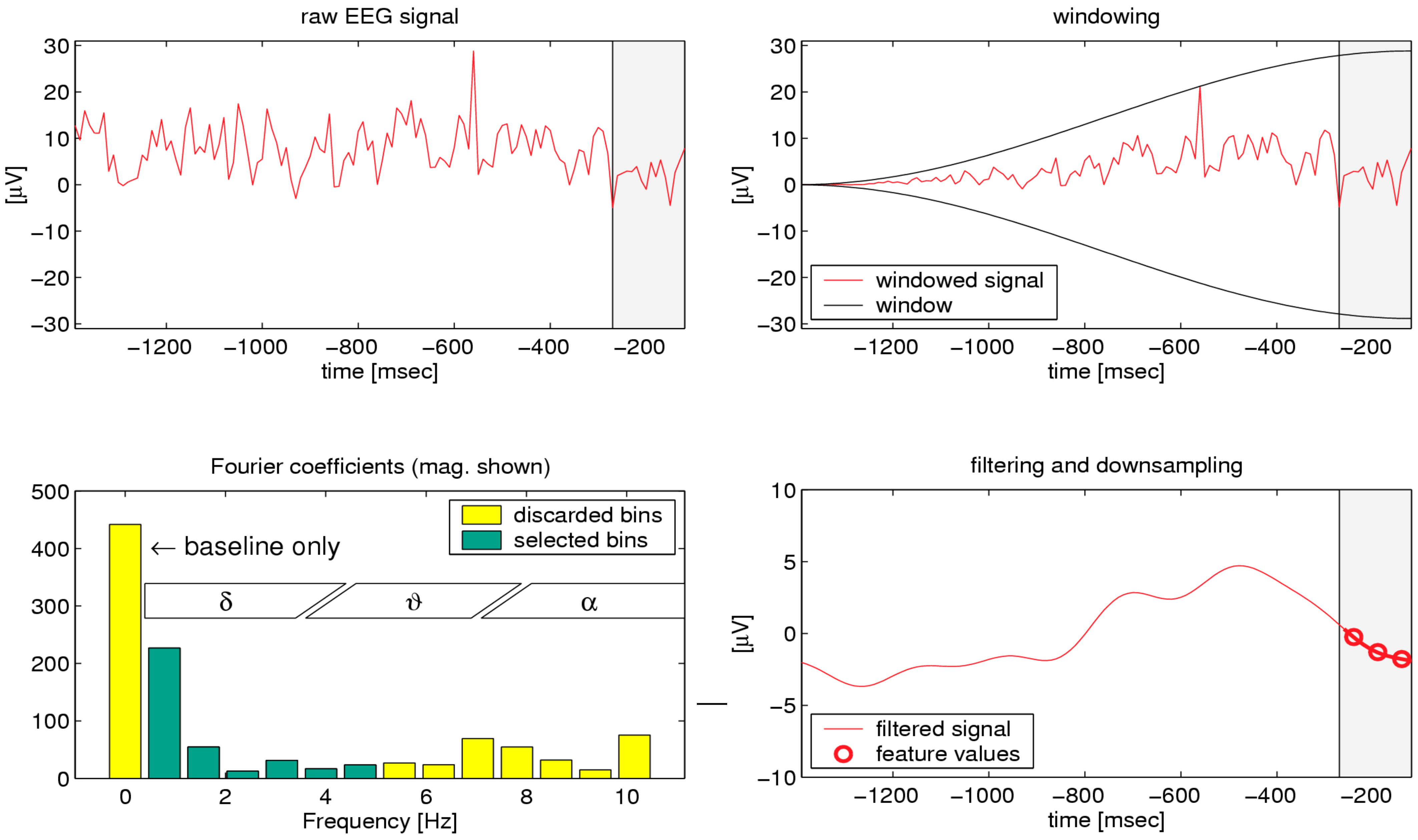


BCI for Assessing Upcoming decisions

Bereitschaftspotential over C3 (primary motor cortex of the right hand)



EEG single-trial preprocessing



Regularized Fisher Discriminant

- detect & remove outliers!
- limit the influence of single patterns, i.e. mistrust your data
- $y = sgn(\mathbf{w}^\top \mathbf{x} + b)$; **regularize!**, e.g. Regularized Fisher Discriminant (RFD, cf. Mika, Rätsch & Müller 2000): find $\mathbf{w}$ by solving the mathematical program

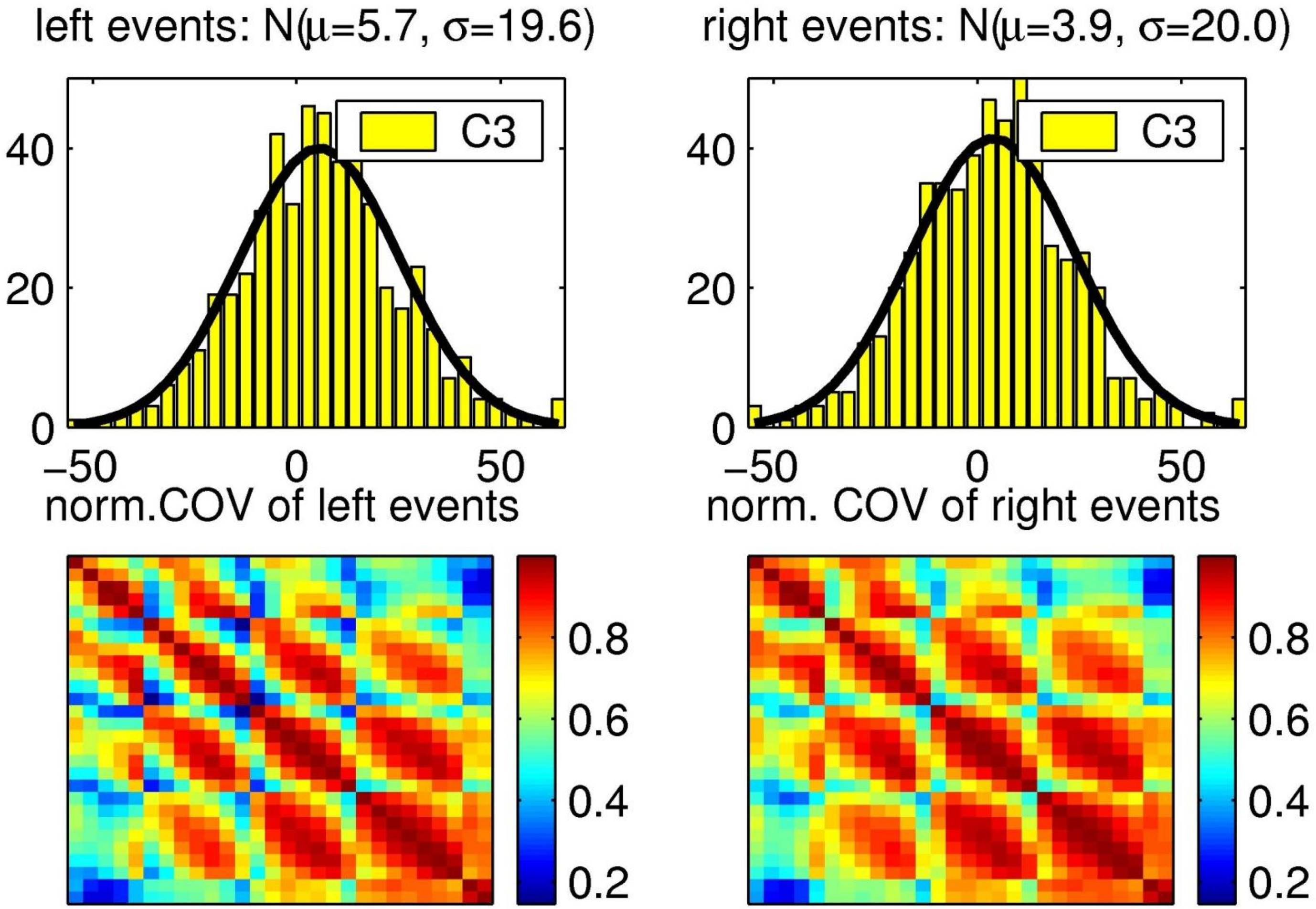
$$\min_{w,b,\xi} \ \frac{1}{2}\|w\|_2^2 + \frac{C}{K}\|\xi\|_2^2$$

$$\text{subject to} \quad y_k(w^\top x_k + b) = 1 - \xi_k \quad \text{for } k = 1, \ldots, K$$

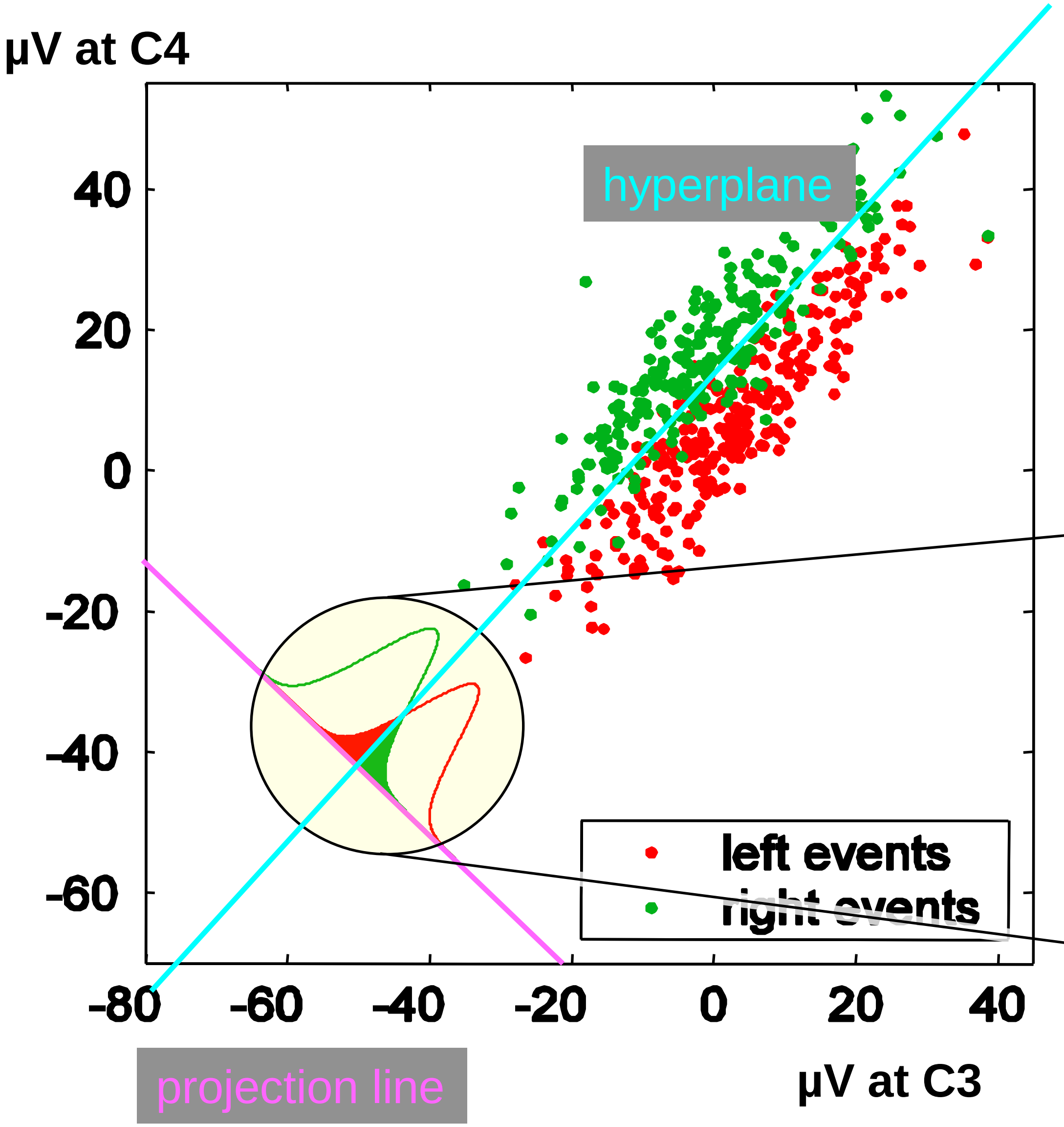
  ξ: slack variables, C: regularization strength (hyperparameter).

- use more robust loss functions, e.g. ℓ_1-norm or Huber-loss

Fisher's Discriminant: Assumptions correct?



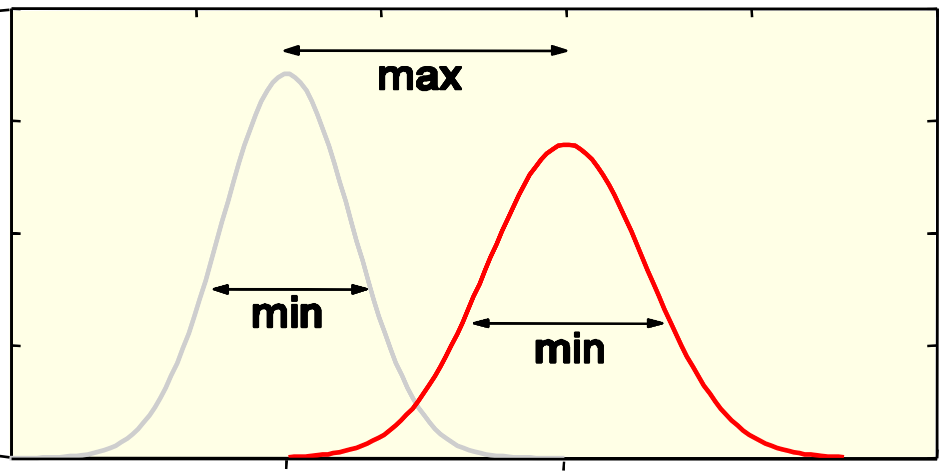
Linear Classification



1) Binary linear classification separates the feature space by a *hyperplane*.

2) *Projection line:* 'best' discriminating dimension

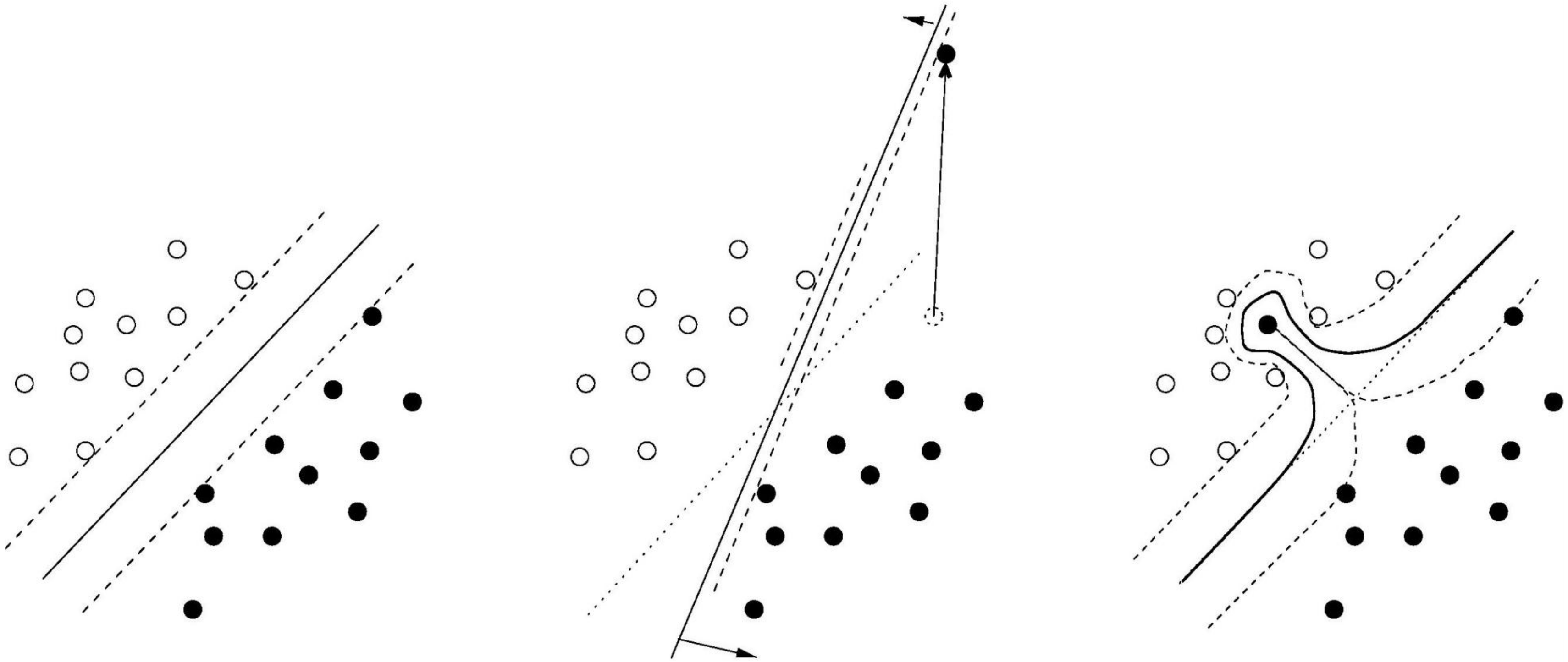
3) *Linear* classifications determine a projection line on a training set such that a *specific objective* is satisfied for the projected distributions.



E.g. Fisher Discriminant (FD) maximizes margin between means of the projected class distributions and minimizes intra-class variance.

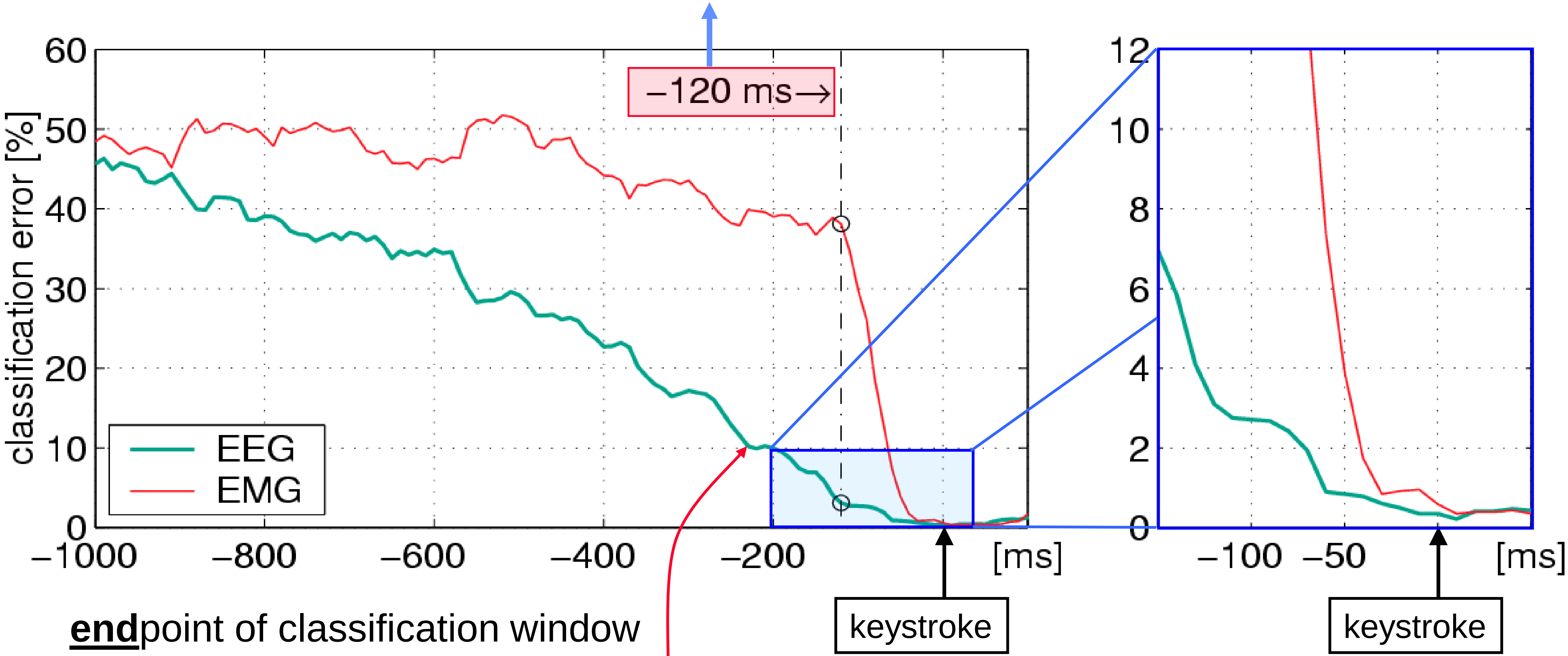
- Linear classifications yields good generalisation in case of limited training data.
- **BUT Regularize!**

Robustness against outliers is mandatory



Time development of classification error (FDA)

At an average keystroke interval of 2.1 sec ⇒ 22.9 bit/min



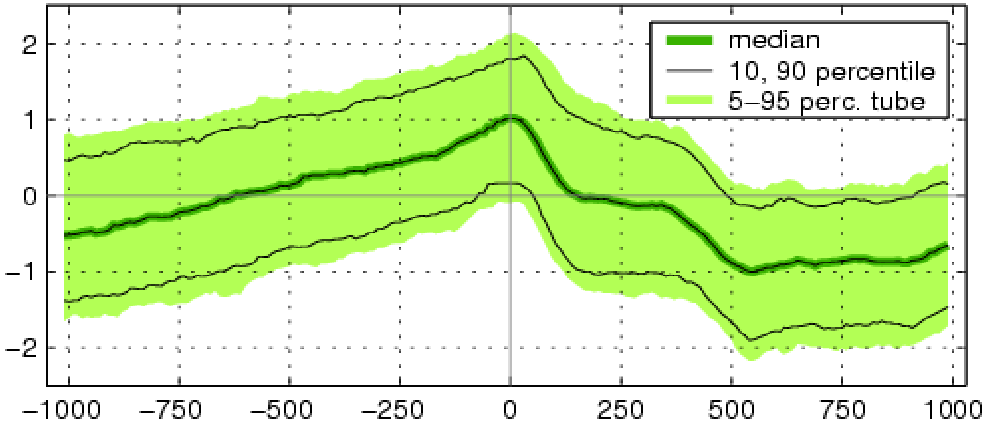
endpoint of classification window

EEG error ≤ 10% after -230 ms before keystroke

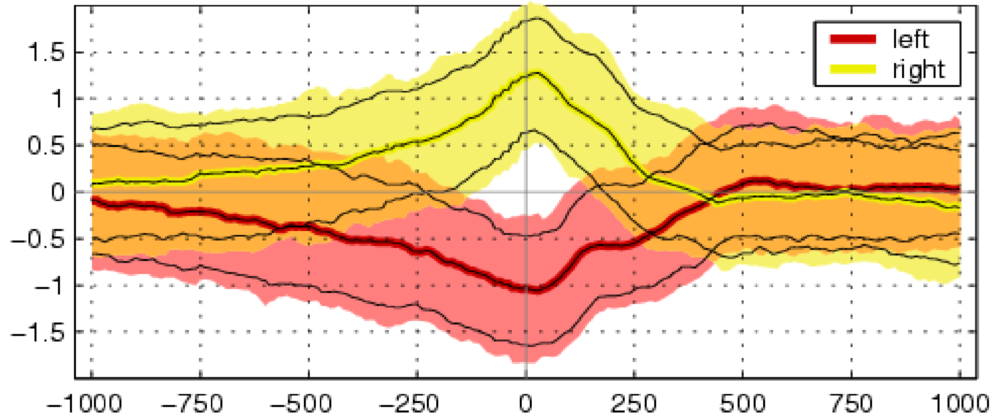
Steps towards online classification

- *no* usage of information about event *timing* (keystrokes)
- *ternary* decision: right – left – no movement
- *continuous* classification in sliding windows + *graded* output

detect upcoming movements

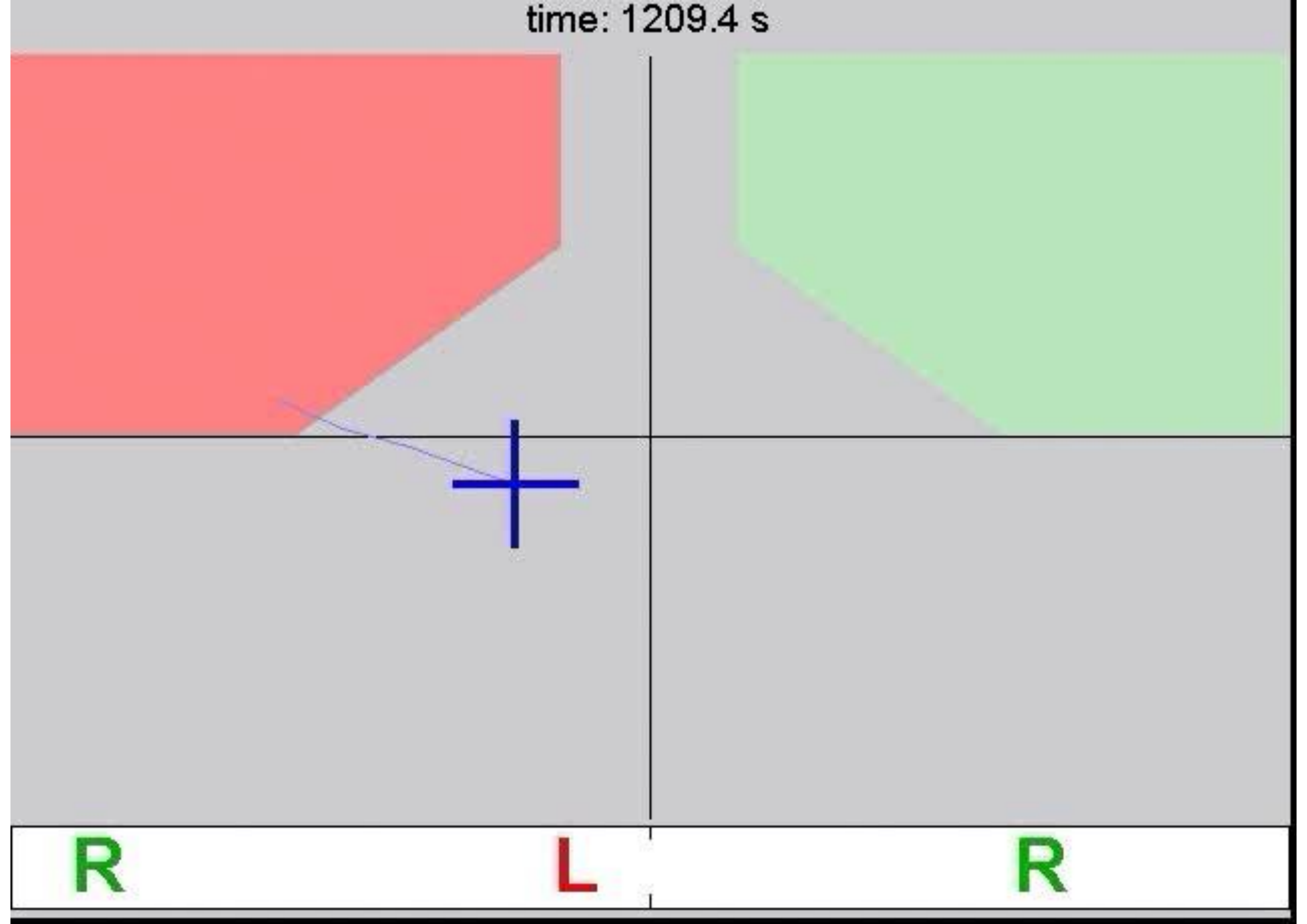


predict movement laterality

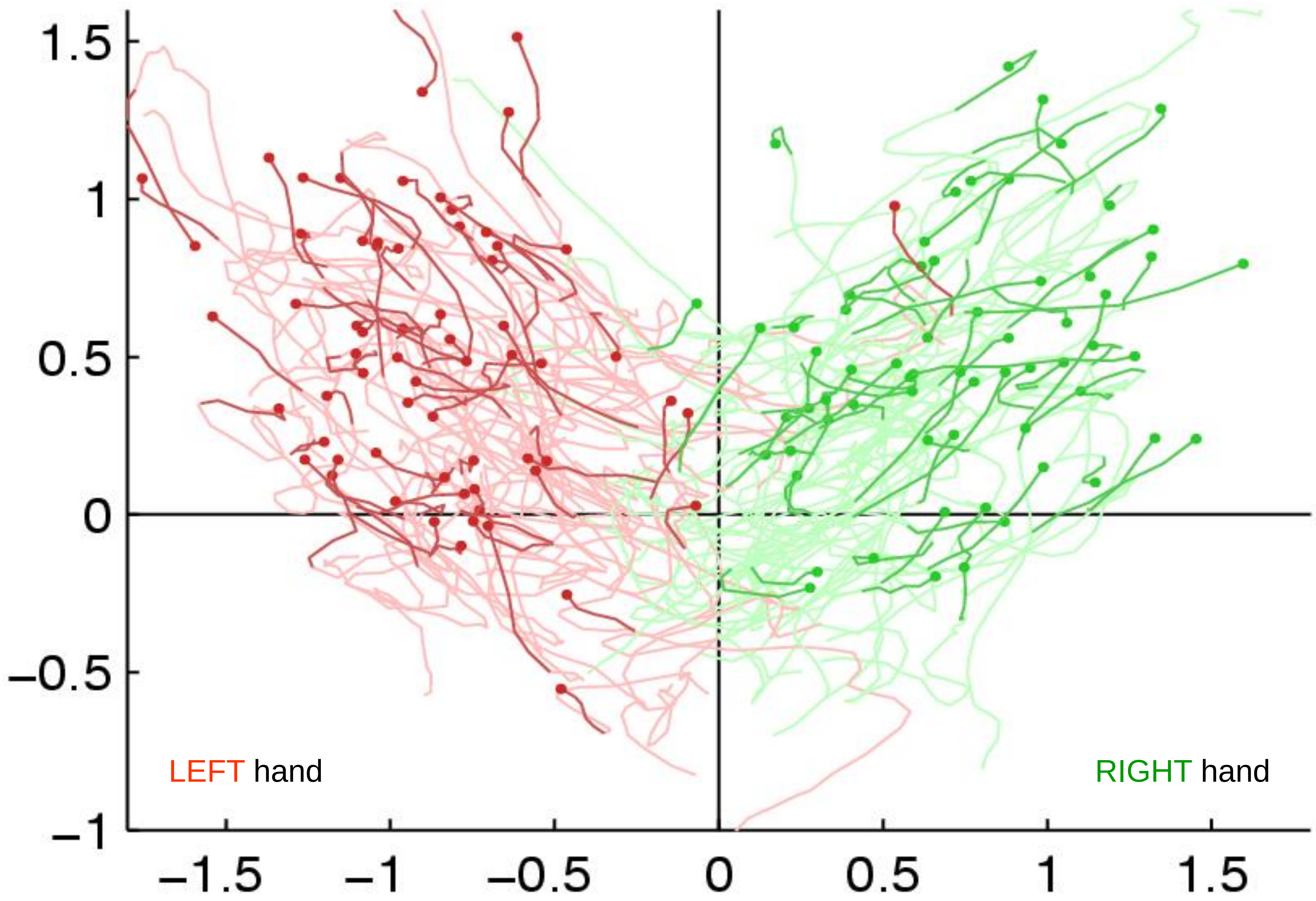


online 2-classifier combination: 10% error rate corresponding to 29 bits/min.

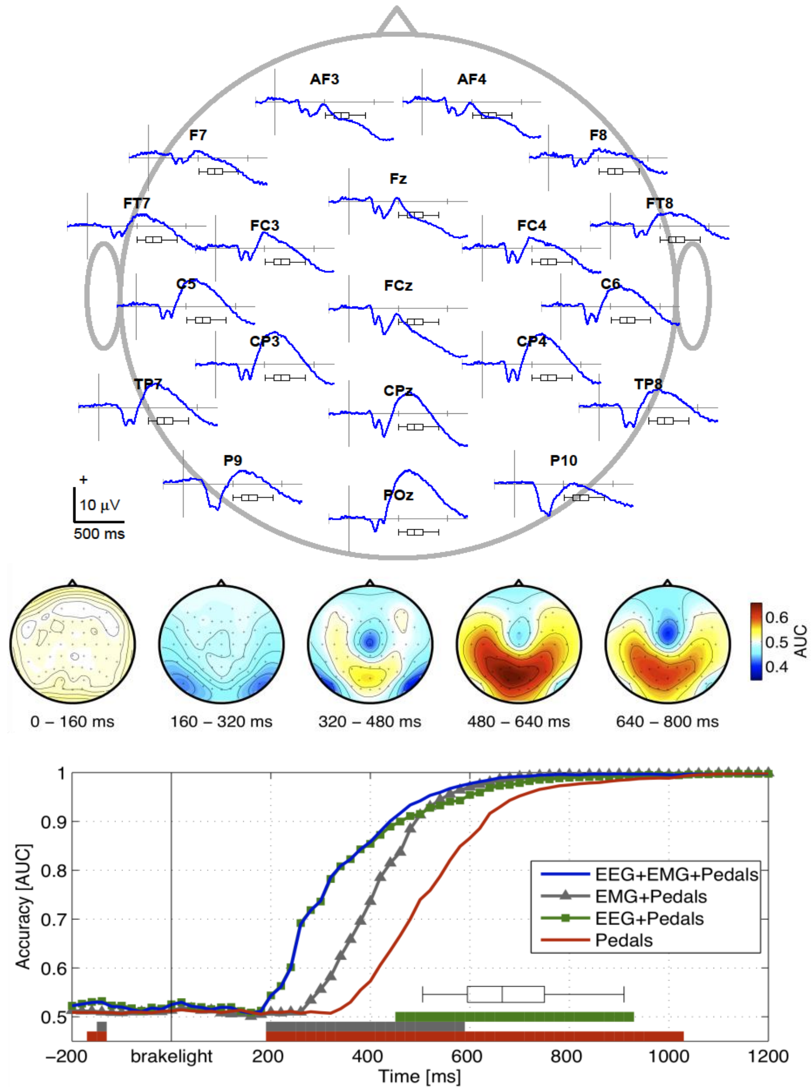
[redacted]


time: 1209.4 s
R
L
R

The shape of thoughts to come



Study: emergency breaking in driving simulator



- Highly specific sequence of EEG potentials 500 ms before breaking
    1) Perception of breaklight stimulus (‘visual evoked potentials’)
    2) Identification of emergency (‘P300’ component)
    3) Preparation of breaking movement (‘Bereitschaftspotential’)

- EEG (+EMG) features improve the pedal based breaking detector by 150 ms

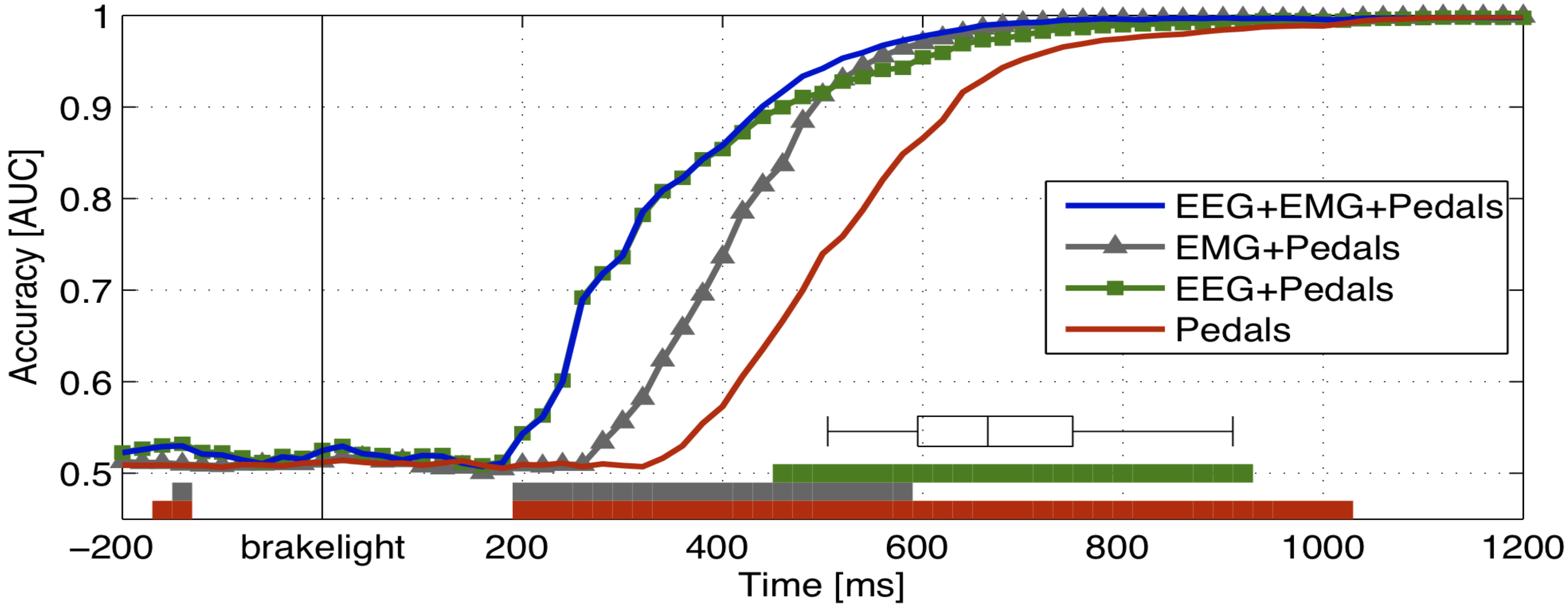
- 4 m less breaking space at speed100 km/h

[Haufe et al., EEG potentials predict upcoming emergency brakings during simulated driving. *J Neural Eng*. 2011]

Car Safety: Improving emergency braking

Driving simulator study with 20 participants

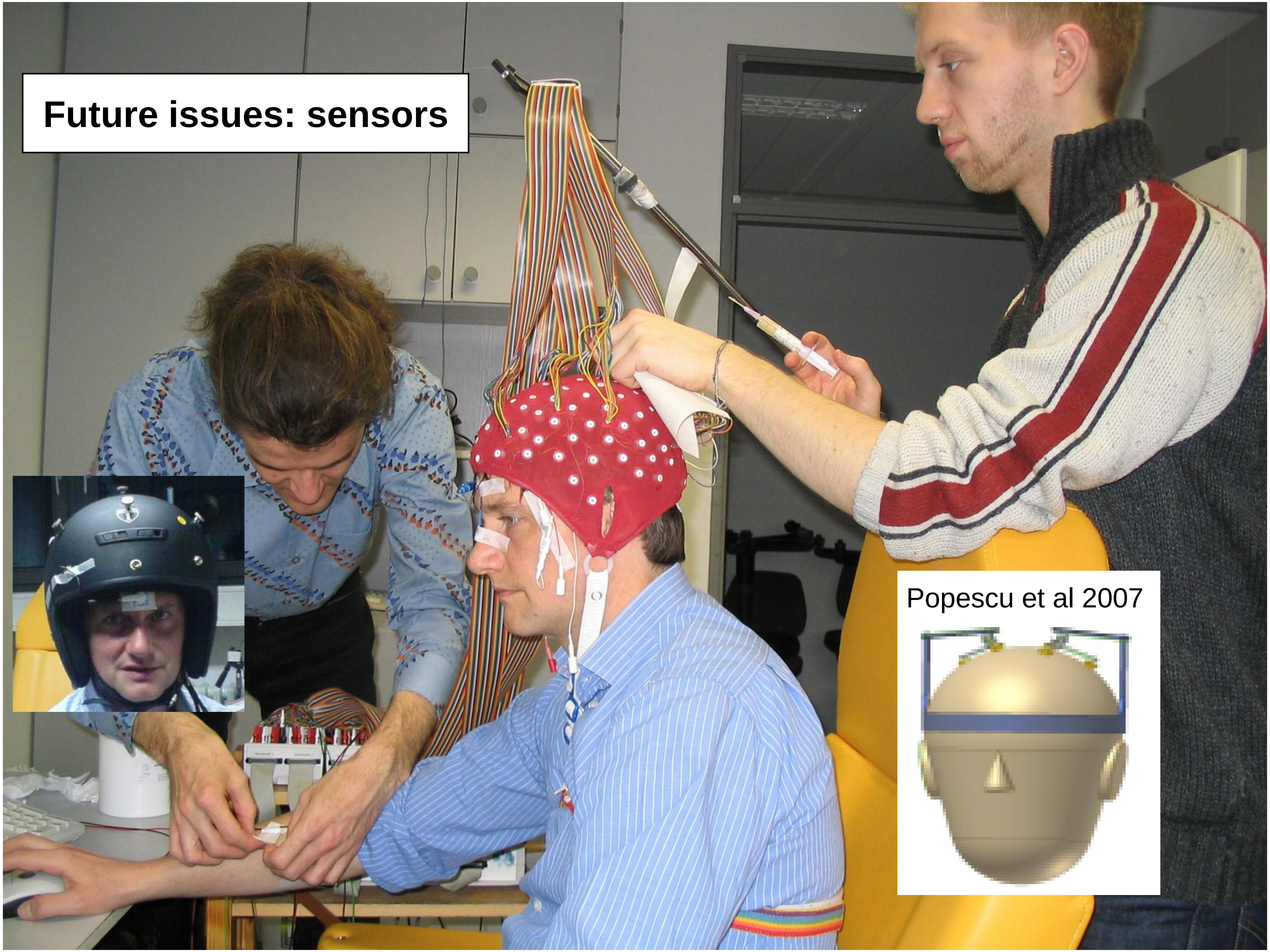
Task: tightly follow a computer-controlled car which performs unpredictable sudden brakes.



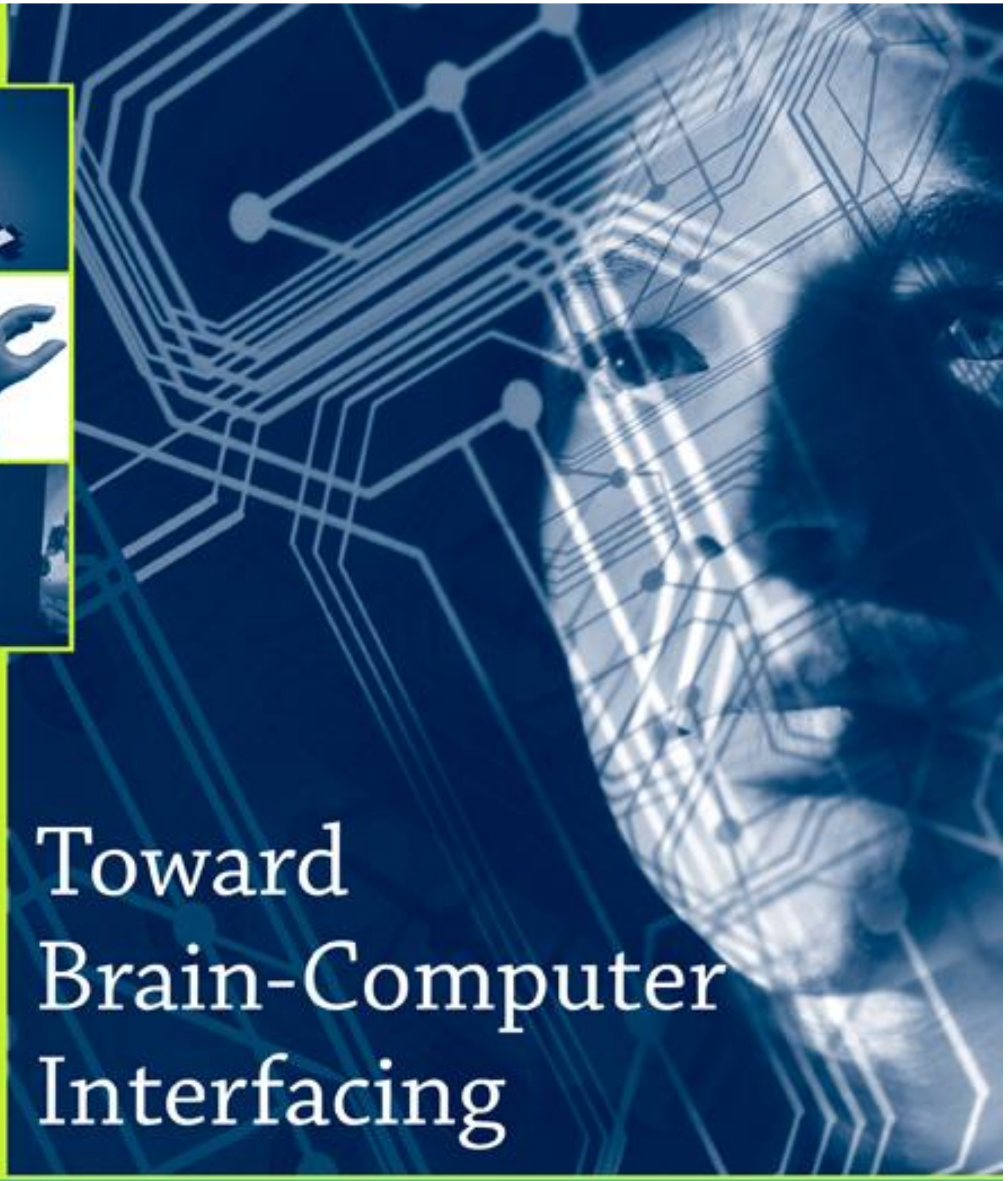
Conclusion

- BBCI: Untrained, Calibration < 10min, data analysis <<5min, BCI experiment
- 5-8 letters/min mental typewriter CeBit 06,10. Brain2Robot@Medica 07, lNdW 09
- Machine Learning and modern data analysis is of central importance for BCI **et al**
- Important issue of this talk: How to learn under nonstationarity?
- Solutions:
- SSA, i.e. project on stationary subspace and learn there, linear, sound & fast
- Modeling: covariate shift based CV: special
- mixed effects model
- co-adaptation, Multimodal
- tracking, invariant features etc

FOR INFORMATION SEE:
www.bbci.de


Future issues: sensors
Popescu et al 2007



Toward Brain-Computer Interfacing

edited by
Guido Dornhege, José del R. Millán, Thilo Hinterberger, Dennis J. McFarland, and Klaus-Robert Müller

foreword by Terrence J. Sejnowski

Thanks to BBCI core team:

Gabriel Curio
Florian Losch
Volker Kunzmann
Frederike Holefeld
Vadim Nikulin@Charite

Florin Popescu
Andreas Ziehe
Steven Lemm
Motoaki Kawanabe
Guido Nolte@FIRST

Yakob Badower@Pico Imaging
Marton Danozci

BBCI
berlin
brain computer
interface

Benjamin Blankertz
Michael Tangermann
Claudia Sannelli
Carmen Vidaurre
Siamac Fazli
Martijn Schreuter
Stefan Haufe
Laura Acqualagna
Thorsten Dickhaus
Frank Meinecke
Felix Biessmann@TUB

Matthias Krauledat
Guido Dornhege
Roman Krepki@industry

Collaboration with: U Tübingen, Bremen, Albany, TU Graz, EPFL, Daimler, Siemens, MES, MPIs, U Tokyo, TIT, RIKEN, Bernstein Center for Computational Neuroscience Berlin, Columbia, CUNY
Funding by: EU, BMBF and DFG

Overview of BCI Competitions

BCI competition I	BCI competition II
December 2001 – June 2002	December 2003 – June 2004
3 datasets	6 datasets
10 submissions	59 submissions
[Sajda et al., 2003]	[Blankertz et al., 2004]

BCI Competition III

- Dec 12th 2004 – May 31st 2005
- announcement of the results: between June 14th and 19th 2005
- 8 datasets from 5 different BCI groups with different tasks

For BCI IV Competition see www.bbci.de

berlin
brain computer
interface

FOR INFORMATION SEE: www.bbci.de

Machine Learning open source software initiative: MLOSS see

www.jmlr.org


Advances in Neurotechnology
Workshop 2012
BBCI
Bernstein Focus:
Neurotechnology Berlin
Ceci n'est pas une noix
september 17 - 19

F. Biessmann, F. C. Meinecke, A. Gretton, A. Rauch, G. Rainer, N. Logothetis, K. R. Müller, Temporal Kernel Canonical Correlation Analysis and its Application in Multimodal Neuronal Data Analysis, Machine Learning, 79(1-2):5-27, 2009

F. Biessmann, S. M. Plis, F. C. Meinecke, T. Eichele, K. R. Müller, Analysis of Multimodal Neuroimaging Data, IEEE Reviews in Biomedical Engineering , 4:26-58, 2011

Birbaumer, N, Ghanayim, N, Hinterberger, T, et al., A spelling device for the paralysed, Nature, 398(6725): 297-298, 1999

G. Blanchard, M. Sugiyama, M. Kawanabe, V. Spokoiny, K. R. Müller, In search of non-Gaussian components of a high-dimensional distribution, Journal of Machine Learning Research, 7:247-282, 2006

B. Blankertz, G. Curio, K. R. Müller, Classifying Single Trial EEG: Towards Brain Computer Interfacing, Advances in Neural Inf. Proc. Systems (NIPS 01), 14:157-164, 2002

B. Blankertz, G. Dornhege, C. Schäfer, R. Krepki, J. Kohlmorgen, K. R. Müller, V. Kunzmann, F. Losch, G. Curio, Boosting Bit Rates and Error Detection for the Classification of Fast-Paced Motor Commands Based on Single-Trial EEG Analysis, IEEE transactions on neural systems and rehabilitation engineering, 11(2):127-131, 2003

B. Blankertz, K. R. Müller, G. Curio, T. M. Vaughan, G. Schalk, J. R. Wolpaw, A. Schlögl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schröder, N. Birbaumer, The BCI Competition 2003: Progress and Perspectives in Detection and Discrimination of EEG Single Trials, IEEE transactions on bio-medical engineering, 51(6):1044-1051, 2004

B. Blankertz, K. R. Müller, D. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlögl, G. Pfurtscheller, J. d. R. Millán, M. Schröder, N. Birbaumer, The BCI Competition III: Validating Alternative Approachs to Actual BCI Problems, IEEE transactions on neural systems and rehabilitation engineering, 14(2):153-159, 2006

B. Blankertz, G. Dornhege, S. Lemm, M. Krauledat, G. Curio, K. R. Müller, The Berlin Brain-Computer Interface: Machine Learning Based Detection of User Specific Brain States, J Universal Computer Sci, 12(6):581-607, 2006

B. Blankertz, G. Dornhege, M. Krauledat, K. R. Müller, G. Curio, The non-invasive Berlin Brain-Computer Interface: Fast Acquisition of Effective Performance in Untrained Subjects, NeuroImage, 37(2):539-550, 2007

B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, K. R. Müller, Optimizing Spatial Filters for Robust EEG Single-Trial Analysis, IEEE Signal Processing Magazine, 25(1):41-56, 2008

B. Blankertz, C. Sannelli, S. Halder, E. M. Hammer, A. Kübler, K. R. Müller, G. Curio, T. Dickhaus, Neurophysiological Predictor of SMR-Based BCI Performance, NeuroImage, 51(4):1303-1309, 2010

B. Blankertz, M. Tangermann, C. Vidaurre, S. Fazli, C. Sannelli, S. Haufe, C. Maeder, L. E. Ramsey, I. Sturm, G. Curio, K. R. Müller, The Berlin Brain-Computer Interface: Non-Medical Uses of BCI Technology, Frontiers in neuroscience, 4:198, 2010

B. Blankertz, S. Lemm, M. S. Treder, S. Haufe, K. R. Müller, Single-trial analysis and classification of ERP components - a tutorial, NeuroImage, 56:814-825, 2011

P. v. Bünau, F. C. Meinecke, F. Király, K. R. Müller, Finding Stationary Subspaces in Multivariate Time Series, Physical Review Letters, 103:214101, 2009

G. Dornhege, B. Blankertz, G. Curio, K. R. Müller, Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms IEEE transactions on bio-medical Engineering, 51(6): 993-1002 , 2004

G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, K. R. Müller, Combined optimization of spatial and temporal filters for improving Brain-Computer Interfacing, IEEE transactions on bio-medical engineering, 53(11):2274-2281, 2006

G. Dornhege, J. d. R. Millán, T. Hinterberger, D. McFarland, K.R. Müller, Toward Brain-Computer Interfacing, MIT Press, 2007

S. Fazli, F. Popescu, M. Danóczy, B. Blankertz, K. R. Müller, C. Grozea, Subject-independent mental state classification in single trials, Neural networks, 22(9):1305-1312, 2009

S. Fazli, M. Danóczy, J. Schelldorfer, K. R. Müller, L1-penalized Linear Mixed-Effects Models for high dimensional data with application to BCI, NeuroImage, 56(4):2100 - 2108, 2011

S. Fazli, J. Mehnert, J. Steinbrink, G. Curio, A. Villringer, K. R. Müller, B. Blankertz, Enhanced performance by a Hybrid NIRS-EEG Brain Computer Interface, open access NeuroImage, 59(1):519-529, 2012

S. Haufe, M. S. Treder, M. F. Gugler, M. Sagebaum, G. Curio, B. Blankertz, EEG potentials predict upcoming emergency brakings during simulated driving, Journal of neural engineering, 8:056001, 2011

A. Kübler, K. R. Müller, An introducton to brain computer interfacing, Toward Brain-Computer Interfacing, MIT press, 2007

TN. Lal, M Schröder, T Hinterberger, J Weston, M Bogdan, N Birbaumer, B Schölkopf, Support Vector Channel Selection in BCI, IEEE Transactions on Biomedical Engineering, 51(6) 1003-1010, 2004

S. Lemm, B. Blankertz, G. Curio, K. R. Müller, Spatio-Spectral Filters for Improving Classification of Single Trial EEG, IEEE transactions on bio-medical engineering, 52(9):1541-1548, 2005

S. Lemm, B. Blankertz, T. Dickhaus, K. R. Müller, Introduction to machine learning for brain imaging NeuroImage, 56:387-399, 2011

MAL. Nicolelis, Actions from thoughts, Nature, 409(6818)403-7, 2001

K. R. Müller, C. W. Anderson, G. E. Birch, Linear and Non-Linear Methods for Brain-Computer Interfaces, IEEE transactions on neural systems and rehabilitation engineering, 11(2):165-169, 2003

K. R. Müller, S. Mika, G. Rätsch, K. Tsuda, B. Schölkopf, An Introduction to Kernel-based Learning Algorithms, IEEE Transactions on Neural Networks, 12(2):181-201, 2001

K. R. Müller, M. Tangermann, G. Dornhege, M. Krauledat, G. Curio, B. Blankertz, Machine learning for real-time single-trial EEG-analysis: From brain-computer interfacing to mental state monitoring, Journal of neuroscience methods, 167(1):82-90, 2008

G. Nolte, A. Ziehe, V. Nikulin, A. Schlögl, N. Krämer, T. Brismar, K. R. Müller, Robustly Estimating the Flow Direction of Information in Complex Physical Systems, Physical Review Letters, 100:234101, 2008

F. Popescu, S. Fazli, Y. Badower, B. Blankertz, K. R. Müller, Single Trial Classification of Motor Imagination Using 6 Dry EEG Electrodes, PloS one, 2(7), 2007

P. Sajda, A. Gerson, K. R. Müller, B. Blankertz, L. Parra, A Data Analysis Competition to Evaluate Machine Learning Algorithms for use in Brain-Computer Interfaces, IEEE transactions on neural systems and rehabilitation engineering, 11(2):184-185, 2003

W. Samek, C. Vidaurre, K. R. Müller, M. Kawanabe, Stationary Common Spatial Patterns for Brain-Computer Interfacing, J Neural Eng, 9(2):026013, 2012

S. Scholler, S. Bosse, M. S. Treder, B. Blankertz, G. Curio, K. R. Müller, T. Wiegand, Towards a Direct Measure of Video Quality Perception using EEG, IEEE transactions on image processing, 21(5): 2619-2629, 2012

P. Shenoy, M. Krauledat, B. Blankertz, R. P. Rao, K. R. Müller, Towards Adaptive Classification for BCI, Journal of neural engineering, 3(1):R13-R23, 2006

M. Sugiyama, M. Krauledat, K. R. Müller, Covariate Shift Adaptation by Importance Weighted Cross Validation, Journal of Machine Learning Research, 8:1027-1061, 2007

M. Tangermann, K. R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. Miller, G. Müller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre, S. Waldert, B. Blankertz, Review of the BCI Competition IV, Frontiers in neuroscience, 6(55), 2012

DM. Taylor, SI. Tillery, AB. Schwartz, Direct cortical control of 3D neuroprosthetic devices. Science. 296(5574):1829-32, 2002

R. Tomioka, K. R. Müller, A regularized discriminative framework for EEG analysis with application to brain-computer interface, NeuroImage, 49:415-432, 2010

M. S. Treder, N. M. Schmidt, B. Blankertz, Gaze-independent brain-computer interfaces based on covert attention and feature attention, Journal of neural engineering, 8(6):066003, 2011

B. Venthur, S. Scholler, J. Williamson, S. Dähne, M. S. Treder, M. T. Kramarek, K. R. Müller, B. Blankertz, Pyff - A Pythonic Framework for Feedback Applications and Stimulus Presentation in Neuroscience, Frontiers in neuroscience, 4:179, 2010

Biased selected references

C. Vidaurre, C. Sannelli, K. R. Müller, B. Blankertz, Machine-Learning Based Co-adaptive Calibration, Neural computation, 23(3):791-816, 2011

J. Williamson, R. Murray-Smith, B. Blankertz, M. Krauledat, K. R. Müller, Designing for uncertain, asymmetric control: Interaction design for brain-computer interfaces, Int J Hum Comput Stud, 67(10):827-841, 2009

Wolpaw, J. et al., Brain–computer interfaces for communication and control Source: Wolpaw, Jonathan. Clinical Neurophysiology, 113(6) 767-791, 2002

Ziehe, A., Müller, KR, Nolte, G, Mackert, BM, Curio, G, Artifact reduction in magnetoneurography based on time-delayed second-order correlations Source: IEEE Transactions On Biomedical Engineering, 47 (1) 75-87, 2000