Brain Computer Interfacing

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**!

<table>
<thead>
<tr>
<th>Group</th>
<th>Projects/Techniques</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schwarz, Pittsburg</td>
<td>Invasive</td>
</tr>
<tr>
<td>Chapin, Rochester</td>
<td>Invasive</td>
</tr>
<tr>
<td>Nicolelis, Duke</td>
<td>Invasive</td>
</tr>
<tr>
<td>Kennedy, Atlanta</td>
<td>Invasive</td>
</tr>
<tr>
<td>Levine, Michigan</td>
<td>Invasive</td>
</tr>
<tr>
<td>Wolpaw, Albany</td>
<td>BCI 2000, 2D, Patients</td>
</tr>
<tr>
<td>Donchin, Beckman</td>
<td>P300: Spelling</td>
</tr>
<tr>
<td>Anderson, UC, CSU</td>
<td>NN for BCI, invasive</td>
</tr>
<tr>
<td>Sadja and Parra, NY</td>
<td>SP, Rapid Visual Stimulation</td>
</tr>
<tr>
<td>Birbaumer, Kühler TÜ</td>
<td>SCPs, TTD, Patients</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- 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

Motor cortex
From dipole patches to EEG

[From Vigario]
Invasive vs noninvasive Brain Computer Interfacing

Layers
- Scalp
- Skull
- Dura
- Arachnoid
- Pia
- Cortex
- White matter

Signal Source
- EEG
- ECoG (epidural or subdural)
- Intraparenchymal (single neuron or local field potential)

[From Schalk]
Invasive BCI at its best

[From Schwartz]
ECOG

- presurgical localization of area causing epilepsy
- excellent possibility 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.]

Brain-Computer Interface

- Signal Processing
- EEG Acquisition
- Application Interface

FES Device
Grasp-Pattern
3 channel Stimulation

BBCI: Leitmotiv: »let the machines learn«

[From Pfurtscheller et al.]
The cerebral cocktail party problem

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

Towards imaginations: Modulation of Brain Rhythms

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

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

  - eyes closed
  - eyes open

  ![Graphs showing $\alpha$-rhythm differences between eyes closed and open](image)

  **Single channel**

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

  - arm at rest
  - arm moves

  ![Graphs showing $\mu$-rhythm differences between arm at rest and in motion](image)

  **IMAGINATION** of left arm
Variance I: Single-trial vs. Averaging

Time Courses at Electrode C4

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

Diagram:
- Calibration session
  - Supervised measurement
  - Labeled trials
  - Feature extraction
  - Machine learning
  - Classifier
BBCI paradigms

Leitmotiv: ›let the machines learn‹

- healthy subjects *untrained* for BCI

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

→ infer the respective brain activities (ML & SP)

B: online feedback session
BBCl paradigms

Leitmotiv: ›let the machines learn‹

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

- instruction: imagine
  - squeezing a ball,
  - kicking a ball,
  - feel touch
Playing with BCI: training session (20 min)
Machine learning approach to BCI: infer prototypical pattern

Imagine left hand movements

Imagine right hand movements

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

**Remark:** Simultaneous selection of frequency band and interval is more appropriate.
**Goal:** Find spatial filters that optimally capture modulations of brain rhythms

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

Diagram:
- **Unknown Sources**:
  - Min variance for right
  - No class-specific influence on variance
  - Min variance for left

- **Observed Signals**:
  - $V^{-1}$ (projection)
  - EEG

- **Discriminative Signals**:
  - $V$ (filter)
  - csp:R1,2,...
  - csp:L1,2,...
Common Spatial Patterns for 2 classes

Original data: Each class has a specific spatial extension. Let $\Sigma_1$ and $\Sigma_2$ be the covariance matrices of the two classes. The blue cross visualizes the covariance 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: $\hat{\Sigma}_i = P\Sigma_iP^\top$.

Calculate orthogonal matrix $R$ and diagonal matrix $D$ 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.

Distribution of EEG features
BBCI Set-up

multi-channel EEG

FFT based low-pass filter

band-pass 4-40 Hz -> AR coefs.

subject-specific band-pass filter, e.g. 7-14Hz, -> multi-class CSP

Artifact removal

multiple feature extraction

$X_{MRP}$

$X_{AR}$

$X_{CSP}$

feature combiner 'PROB'

continuous feedback

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

subject to $y_k(w^T x_k + b) = 1 - \xi_k$ for $k = 1, \ldots, K$

What can Machine Learning tell us about physiology?

\[
\min_{\mathbf{w}, \mathbf{b}, \xi} \frac{1}{2} \| \mathbf{w} \|_1 + \frac{C}{K} \| \xi \|_1 \\
\text{subject to } y_k (\mathbf{w}^\top \mathbf{x}_k + b) = 1 - \xi_k \quad \text{for } k = 1, \ldots, K
\]

BCI with machine learning: feedback

### Offline: Calibration (10–20 minutes)
- Supervised measurement
- Labeled trials
- Feature extraction
- Machine learning

#### Feedback Session
- Spontaneous EEG
- Feature extraction
- Classifier
- Output signal

#### Collect training samples

### Online: Feedback (up to 6 hours)
- Classification of sliding windows (≤ 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)
- insufficient 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.
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:

$n\text{BlocksPerClass}=1$:

$n\text{BlocksPerClass}=2$:

$n\text{BlocksPerClass}=3$:

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

multi-channel EEG

FFT based low-pass filter

band-pass 4-40 Hz -> AR coeffs.

subject-specific band-pass filter, e.g. 7-14Hz, -> multi-class CSP

multiple feature extraction

classifier

continuous feedback

$x_{MRF}$

$x_{AR}$

feature combiner 'PROB'

$x_{CSP}$

minimum

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

subject to

\[ y_k(w^T x_k + b) = 1 - \xi_k \quad \text{for } k = 1, \ldots, K \]

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

\[ \{ (x_i, y_i) \mid y_i = f(x_i) + \epsilon_i \}_{i=1}^{n} \]

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

\[ \alpha^* = (\alpha_1^*, \alpha_2^*, \ldots, \alpha_p^*)^\top \] which minimizes

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

\[ \hat{f}(x) = \sum_{i=1}^{p} \alpha_i \varphi_i(x), \] choosing \[ w(x_i) = \frac{p_{fb}(x_i)}{p_{tr}(x_i)} \]

yields unbiased estimator even under covariate shift

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

Principal Component Analysis (PCA)

Uncorrelated sources
Orthogonal mixing

$X = A \begin{bmatrix} S(1) \\ \vdots \\ S(d) \end{bmatrix}$

Max. variance
Min. variance

Independent Component Analysis (ICA)

Independent sources
Arbitrary mixing

$X = A \begin{bmatrix} S(1) \\ \vdots \\ S(d) \end{bmatrix}$

Stationary Subspace Analysis (SSA)

Arbitrary mixing
Stationary sources
Non-stationary sources

$X_t = A \begin{bmatrix} S^s_t \\ S^n_t \end{bmatrix}$
Splitting into stationary and nonstationary subspace: SSA

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

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

[cf. Bünau, Meinecke, Kiraly, Müller PRL 09]
given: Epochs $X_i$ of Data points in $\mathbb{C}^n$

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

„stationary projection“
Inverting the SSA Mixing Model

**Model**

\[ x(t) = As(t) = [A^s \ A^n] \begin{bmatrix} s^s(t) \\ s^n(t) \end{bmatrix} \]

**Goal of SSA**

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

\[
\begin{bmatrix} \hat{s}^s(t) \\ \hat{s}^n(t) \end{bmatrix} = \hat{B}x(t) = \begin{bmatrix} \hat{B}^s \\ \hat{B}^n \end{bmatrix} x(t)
\]
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.

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

Algorithmic Approach

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

Divide the data into epochs (consecutive or sliding window)

Epoch 1 ...

Estimate the epoch mean and covariance matrix.

\[ \mu_1, \Sigma_1 \]

... 

\[ \mu_n, \Sigma_n \]
Using Symmetries and Invariances

**Symmetries**

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

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

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 $s$-sources on the whole dataset to

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

\[
\hat{B}^g = \arg\min_{R R^T = I} \sum_{i=1}^{N} \text{KL} \left[ \mathcal{N}(\hat{\mu}_i^g, \hat{\Sigma}_i^g) \parallel \mathcal{N}(\tilde{\mu}_i^g, \tilde{\Sigma}_i^g) \right]
\]

\[
= \arg\min_{R R^T = I} \sum_{i=1}^{N} \text{KL} \left[ \mathcal{N}(\hat{\mu}_i^g, \hat{\Sigma}_i^g) \parallel \mathcal{N}(0, I) \right]
\]

\[
= \arg\min_{R R^T = I} \sum_{i=1}^{N} \left( -\log \det \hat{\Sigma}_i^g + \hat{\mu}_i^g \hat{\mu}_i^g^T \right)
\]
Optimizing

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: \( M_{i,j} \) rotation angle of axis i towards axis j

This leads to a gradient of the form:

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

rotation in the s-space

rotation between the two spaces

rotation in the n-space
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 $s$-sources and 4 $n$-sources
- Fixed number of samples divided into epochs
Application to Brain-Computer-Interfacing

Original EEG

Classification Error [%]

w/o with SSA

EEG with additional alpha

Field Pattern

n-subspace

s-subspace

Mean additional alpha power on test set
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 $\ell_1$-regularized regression (or others like LMM) using **leave-one-subject-out cross-validation**
- note that our trials have a **grouping** structure

![Diagram](image_url)
Model formulation

- Reminder – Linear regression:
  \[ y = X\beta + \varepsilon \]

- Mixed effects model with \( n \) groups:
  \[ y_i = X\beta + Z_i b_i + \varepsilon_i \quad \forall i \in \{1 \ldots n\} \]
  - Consists of \( n \) simultaneous equations, one for each group
  - The equations are coupled by the common term \( X\beta \)
  - Each equation has a group-dependent term \( Z_i b_i \)
  - In our case, each \( Z_i \) is simply a vector of ones, i.e. the corresponding \( b_i \) is scalar and represents the bias of group \( i \)
  - So-called random intercepts model

- Since we expect our features to be redundant and are aiming for better interpretability, we enforce sparsity by adding an \( \ell_1 \) penalty
Linear Mixed Effects Model: intuition

[Fazli, Müller et al. 2011]
Approach to „Cure“ BCI Illiteracy

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

- Runs 7-8: CSP
  - 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

Runs 1 and 2

 Runs 7 and 8

[cf. Vidaurre, Blankertz, Müller et al. 2009]
Multimodal
Different physiological Features

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

Maps

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

Independent???

Neurophysiology: YES

[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)**

- early distinction between the signals of left and right trials.

**Event-Related Desynchronization (ERD)**

- 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

The combination of ERD and LRP features exploits the merits of the two: rapid response of LRP features and the persistence of ERD features.
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 peaks earlier as compared to HbO and HbR

Physiological reliability: HRF shaped classification accuracies over time

Classification accuracy higher for EEG

Fazli et al. 2012
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

HbO

HbR

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

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

Meta-classifier estimated for combination in each subject

All within cross-validation (8 chronological splits)

Fazli et al. 2012
Feature Combination

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

Fazli et al. 2012
t-tests reveal a significant increase of classification accuracy for combination

Fazli et al. 2012
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

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

\[
\begin{bmatrix}
0 & C_{xy} \\
C_{yx} & 0
\end{bmatrix}
\begin{bmatrix}
w_x \\
w_y
\end{bmatrix}
= \alpha
\begin{bmatrix}
C_{xx} & 0 \\
0 & C_{yy}
\end{bmatrix}
\begin{bmatrix}
w_x \\
w_y
\end{bmatrix}
\]
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

\[
\begin{align*}
w_x &= X \alpha_x \\
w_y &= Y \alpha_y
\end{align*}
\]

No need to compute big covariance matrices!
tkCCA: correlating apples and oranges over time

\[
\arg\max_{w_x(\tau), w_y} \text{Corr} \left( \sum_{\tau} w_x(\tau)^\top x(t - \tau), \; w_y^\top y(t) \right)
\]

\[
\tilde{X} = \begin{bmatrix} X_{\tau_1} \\ X_{\tau_2} \\ \vdots \\ X_{\tau_T} \end{bmatrix} \quad \Downarrow \quad \tilde{w}_x = \begin{bmatrix} w_x(\tau_1) \\ w_x(\tau_2) \\ \vdots \\ w_x(\tau_T) \end{bmatrix}
\]

\[
\arg\max_{\tilde{w}_x, w_y} \text{Corr} \left( \tilde{w}_x^\top \tilde{X}, w_y^\top Y \right)
\]
Application: Neuro-Vascular Coupling
Experimental Setup

» Simultaneous measurements of
  » fMRI/ BOLD signal
  » Intracortical neural activity
Temporal Kernel CCA

\[ w_x(\tau) \]

\[ x(t) \]

\[ w_y \]

\[ y(t) \]

\[ \text{argmax } \text{Corr} \left( \sum_{\tau} w_x(\tau)^\top x(t - \tau), w_y^\top y(t) \right) \]

- 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:

\[
r(x) := \frac{\sqrt{N_1 \cdot N_2}}{N_1 + N_2} \frac{\text{mean}\{x_i \mid y_i = 1\} - \text{mean}\{x_i \mid y_i = 2\}}{\text{std}\{x_i\}}
\]
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

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

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 displayed as a scalp map:
Classification results of spatial features

![Classification results of spatial features](image)
Extraction of spatio-temporal features
Spatio-temporal features are typically high-dimensional (here 59 EEG channels × 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 $x_1, \ldots, x_n \in \mathbb{R}^d$ be $n$ vectors drawn from a $d$-dimensional Gaussian distribution $\mathcal{N}(\mu, \Sigma)$.

For classification $\mu$ and $\Sigma$ have to be estimated from the data:

- $\hat{\mu} = \frac{1}{n} \sum_{k=1}^{n} x_k$
- $\hat{\Sigma} = \frac{1}{n-1} \sum_{k=1}^{n} (x_k - \hat{\mu})(x_k - \hat{\mu})^T$

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

This affects, e.g., classification with LDA:

Normal vector of LDA: $w = \hat{\Sigma}^{-1}(\mu_1 - \mu_2)$. 
Bias in estimating covariances II

The left graph shows the relationship between true covariance and empirical covariance. The right graph illustrates the behavior of eigenvalues with different sample sizes (N=50, 100, 200, 600) compared to the true eigenvalues.
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 $\hat{\Sigma}$ is replaced by

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

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

$$\tilde{\Sigma} = (1 - \gamma)VDV^T + \gamma\nu I = V((1 - \gamma)D + \gamma\nu I) V^T$$

we see that

- $\tilde{\Sigma}(\gamma)$ and $\hat{\Sigma}$ have the same Eigenvectors (columns of $V$)
- extreme Eigenvalues (large/small) are shrunk/extended towards the average $\nu$.
- $\gamma = 0$ yields LDA without shrinkage, $\gamma = 1$ assumes spherical covariance matrices.
LDA with shrinkage of the empirical covariance matrix has one free parameter ($\gamma$), 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 $\gamma$ ($x$-axis). Features vectors have 250 dimensions.
Investigating the impact of shrinkage

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

\[\begin{align*}
gamma &= 0 \\
w &\sim \hat{\Sigma}^{-1}(\mu_1 - \mu_2)
\end{align*}\]

\[\begin{align*}
gamma &= 1 \\
w &\sim \mu_1 - \mu_2
\end{align*}\]

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’ $n_k(t)$:

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

mean: phase-locked – ERP
cov: non-phase-locked – noise
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

(a) with little disturbance

(b)
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 $\gamma$. 
Optimal selection of shrinkage parameters

Let $x_1, \ldots, x_n \in \mathbb{R}^d$ be $n$ feature vectors and let $\hat{\mu} = \frac{1}{n} \sum_{k=1}^{n} x_k$ be the empirical mean.

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

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

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

$$z_{ij}(k) = ((x_k)_i - (\hat{\mu})_i)((x_k)_j - (\hat{\mu})_j)$$

Then the optimal shrinkage parameter $\gamma^*$ for which

$$\tilde{\Sigma}(\gamma^*) = \arg\min_S \|S - \Sigma\|_F^2$$

can be analytically calculated ([2]) as

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

BCI Technology
- machine learning
- adaptive signal processing
- experimental design
- neuroimaging hardware

BCI for Intentional Control
- Tools for patients
  - mental typewriter
  - brain actuated prosthesis
- Rehabilitation
  - NFT for stroke recovery
- HMI
  - Gaming, Art
  - Hands-free applications

Monitoring cognitive states
- Rehabilitation
  - NFT for ADHD
  - NFT for pain reduction
- HMI
  - Neurousability
  - Neuroergonomics
  - Car safety

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

Amplitude [µV]

Time [s]

From Birbaumer et al.]
[From Birbaumer et al.]
ERFAHRUNGEN-EINES-TTD-SCHREIBERS-BEIM-SCHREIBEN


[From Birbaumer et al.]
ECOG Decoding
ECOG

- presurgical localization of area causing epilepsy
- excellent possibility 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
(1) Collect data and divide into training and validation data sets

(2) Use the training data to estimate one or more encoding models for each voxel

(3) Apply the estimated encoding models to the validation data and evaluate prediction accuracy

(4) Use encoding models to derive decoding models. Apply them to the validation data to decode features.
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)

Reconstructing visual experiences from brain activity evoked by natural movies

Shinji Nishimoto, An T. Vu, Thomas Naselaris, Yuval Benjamini, Bin Yu, Jack L. Gallant

Supplemental movie S1
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

BCI Technology
- machine learning
- adaptive signal processing
- experimental design
- neuroimaging hardware

BCI for Intentional Control
- Tools for patients
  - mental typewriter
  - brain actuated prosthesis
- Rehabilitation
  - NFT for stroke recovery
- HMI
  - Gaming, Art
  - Hands-free applications

Monitoring cognitive states
- Rehabilitation
  - NFT for ADHD
  - NFT for pain reduction
- HMI
  - Neurousability
  - Neuroergonomics
  - Car safety

BCI for Assessing Signal Quality perception
Why Quality Assessment?

- Ensure user satisfaction
- Develop better compression algorithms

© www.eftrends.com
Approaches

Behavioral tests (standard)  EEG + BCI methods (novel)

- Continuous signal
- Objective measure
- Capture
  - subtle differences
  - non-conscious processing
## EEG Studies

<table>
<thead>
<tr>
<th>Domain</th>
<th>Stimuli</th>
<th>Cooperation Partner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Auditory</td>
<td>Phonemes</td>
<td>Telekom Laboratories - &quot; &quot;</td>
</tr>
<tr>
<td></td>
<td>Words</td>
<td></td>
</tr>
<tr>
<td>Visual</td>
<td>Flickering light</td>
<td>Philips Research</td>
</tr>
<tr>
<td></td>
<td>Video</td>
<td>Fraunhofer (HHI)</td>
</tr>
</tbody>
</table>
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)

<table>
<thead>
<tr>
<th>Phoneme Study</th>
<th>Word Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>stimulus</td>
<td>/a/</td>
</tr>
<tr>
<td>disturbed by</td>
<td>signal-correlated noise</td>
</tr>
<tr>
<td></td>
<td>/Haus/, /Schild/ by female/male speaker</td>
</tr>
<tr>
<td></td>
<td>bit rate limitation</td>
</tr>
</tbody>
</table>
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

• 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)

![Image of experiment setup](image-url)
## Visual Quality

- **Added value of EEG**

<table>
<thead>
<tr>
<th>Participant</th>
<th>Detected vs CW</th>
<th>Undetected vs CW</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S1</td>
<td>S2</td>
</tr>
<tr>
<td>VPdbe</td>
<td>40 Hz</td>
<td>60 Hz</td>
</tr>
<tr>
<td>VPik</td>
<td>50 Hz</td>
<td>70 Hz</td>
</tr>
<tr>
<td>VPdbf</td>
<td>50 Hz</td>
<td>70 Hz</td>
</tr>
<tr>
<td>VPdbd</td>
<td>40 Hz</td>
<td>50 Hz</td>
</tr>
<tr>
<td>VPow</td>
<td>50 Hz</td>
<td>70 Hz</td>
</tr>
<tr>
<td>VPFat</td>
<td>50 Hz</td>
<td>70 Hz</td>
</tr>
</tbody>
</table>

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**

<table>
<thead>
<tr>
<th>Participant</th>
<th>Detected vs CW</th>
<th>Undetected vs CW</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S1</td>
<td>S2</td>
</tr>
<tr>
<td>VPdbe</td>
<td>40 Hz</td>
<td>60 Hz</td>
</tr>
<tr>
<td>VPIk</td>
<td>50 Hz</td>
<td>70 Hz</td>
</tr>
<tr>
<td>VPdbf</td>
<td>50 Hz</td>
<td>70 Hz</td>
</tr>
<tr>
<td>VPdbd</td>
<td>40 Hz</td>
<td>50 Hz</td>
</tr>
<tr>
<td>VPow</td>
<td>50 Hz</td>
<td>70 Hz</td>
</tr>
<tr>
<td>VPFat</td>
<td>50 Hz</td>
<td>70 Hz</td>
</tr>
</tbody>
</table>

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

- 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 assistant system

Assess cognitive alertness

[Kohlmorgen, Müller et al 2007]
BCI for Assessing Upcoming decisions
Bereitschaftspotential over C3 (primary motor cortex of the right hand)

- Reactive
- Spontaneous

- 0.60 s
- 0.25 s

Movement right hand
EEG single-trial preprocessing

**raw EEG signal**

**windowing**

**Fourier coefficients (mag. shown)**

**filtering and downsampling**

- **baseline only**
- discarded bins
- selected bins

- δ
- δ
- α

- filtered signal
- feature values
Regularized Fisher Discriminant

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

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

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

\( \xi \): slack variables, \( C \): regularization strength (hyperparameter).

- use more robust loss functions, e.g. \( \ell_1 \)-norm or Huber-loss
Fisher’s Discriminant: Assumptions correct?

left events: $N(\mu=5.7, \sigma=19.6)$

right events: $N(\mu=3.9, \sigma=20.0)$
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

At -120 ms before keystroke, EEG error ≤ 10% after -230 ms before keystroke

End point 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.
The shape of thoughts to come
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 speed 100 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, INdW 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

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

<table>
<thead>
<tr>
<th>BCI competition I</th>
<th>BCI competition II</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 datasets</td>
<td>6 datasets</td>
</tr>
<tr>
<td>10 submissions</td>
<td>59 submissions</td>
</tr>
<tr>
<td>[Sajda et al., 2003]</td>
<td>[Blankertz et al., 2004]</td>
</tr>
</tbody>
</table>

## 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
FOR INFORMATION SEE: www.bbci.de

Machine Learning open source software initiative: MLOSS see
www.jmlr.org
Advances in Neurotechnology Workshop 2012

Bernstein Focus: Neurotechnology Berlin
http://bbc12.ml.tu-berlin.de

Ceci n’est pas une noix

September 17 - 19
Audimax der Humboldt-Universität zu Berlin
Unter den Linden 6 • 10099 Berlin
Biased selected references


