Brain-Computer Interfacing

WS 2018/2019 – Vorlesung #10

Benjamin Blankertz

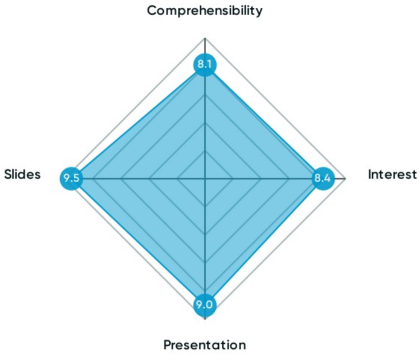
Lehrstuhl für Neurotechnologie, TU Berlin

benjamin.blankertz@tu-berlin.de

09 · Jan · 2019

Feedback from the Last Lecture

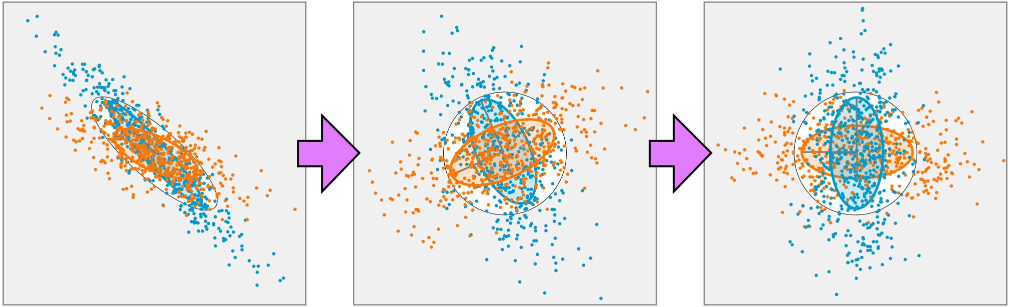
Feedback BCI Lecture #09



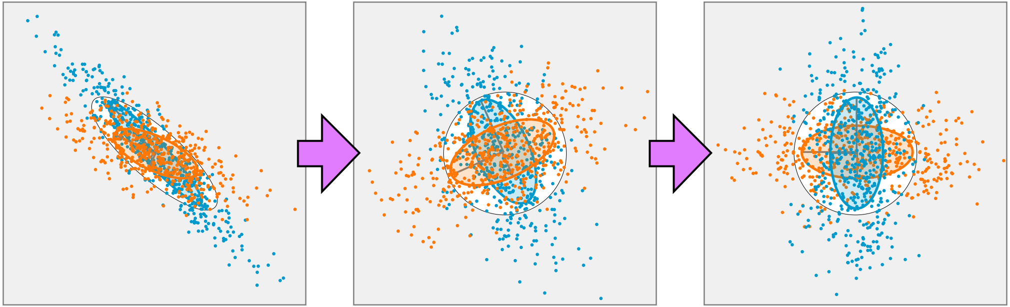
12

https://www.mentimeter.com/s/af0f8da488f3cf62df07c857bf4a1314/a504886b4879

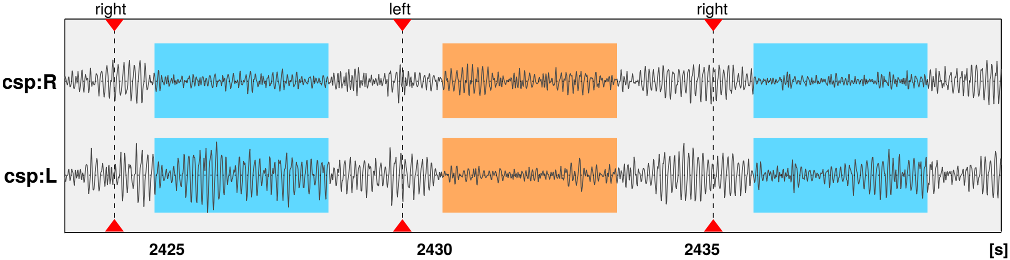
Recap: Illustration of CSP in Two Steps



Recap: Illustration of CSP in Two Steps



In the transformed space, the signals look like this:



Today's Topics

- An alternative approach to CSP with the Rayleigh coefficient
- CSP in the framework of the linear model
- CSP's susceptibility to outliers
- Linear spatial filtering prior to CSP
- Overview of some BCI applications based on the modulation of brain rhythms

An Alternative Approach to CSP

Next, we discuss an alternative approach to derive CSP analysis which provides useful background information.

The previously introduced approach to obtain CSP filters via generalized Eigenvalue decomposition emerged like deus ex machina. The approach presented here will show, **how** one derives the solution from the idea of the optimization.

An Alternative Approach to CSP

Next, we discuss an alternative approach to derive CSP analysis which provides useful background information.

The previously introduced approach to obtain CSP filters via generalized Eigenvalue decomposition emerged like deus ex machina. The approach presented here will show, **how** one derives the solution from the idea of the optimization.

The approach of formalizing an optimization objective as Rayleigh coefficient is widely used and very useful.

It is a stroke of luck, if a problem can be cast into the maximization (or minimization) of a Rayleigh coefficient, because it provides an easy and effective solution.

Optimization with the Rayleigh Coefficient

We define the Rayleigh coefficient wrt the sym. matrices $\mathbf{A}$ and $\mathbf{B}$ as

$$R_{\mathbf{A},\mathbf{B}}(\mathbf{w}) = \frac{\mathbf{w}^\top \mathbf{A}\mathbf{w}}{\mathbf{w}^\top \mathbf{B}\mathbf{w}}.$$

The minimum and maximum of R can be obtained by the following theorem:

*The **Min-Max Theorem** states: $d_1 \leq R_{\mathbf{A},\mathbf{B}}(\mathbf{w}) \leq d_C$, if $d_1 \leq \cdots \leq d_C$ are the **generalized Eigenvalues** of $\mathbf{A}$ and $\mathbf{B}$.*

Optimization with the Rayleigh Coefficient

We define the Rayleigh coefficient wrt the sym. matrices $\mathbf{A}$ and $\mathbf{B}$ as

$$R_{\mathbf{A},\mathbf{B}}(\mathbf{w}) = \frac{\mathbf{w}^\top \mathbf{A}\mathbf{w}}{\mathbf{w}^\top \mathbf{B}\mathbf{w}}.$$

The minimum and maximum of R can be obtained by the following theorem:

The ***Min-Max Theorem*** *states:* $d_1 \leq R_{\mathbf{A},\mathbf{B}}(\mathbf{w}) \leq d_C$, *if* $d_1 \leq \cdots \leq d_C$ *are the* ***generalized Eigenvalues*** *of* $\mathbf{A}$ *and* $\mathbf{B}$*.*

Let $\mathbf{w}_i$ be the corresponding Eigenvectors (i.e., $\mathbf{A}\mathbf{w}_i = \mathbf{B}\mathbf{w}_i d_i$). Then

$$R_{\mathbf{A},\mathbf{B}}(\mathbf{w}_i) = \frac{\mathbf{w}_i^\top \mathbf{A}\mathbf{w}_i}{\mathbf{w}_i^\top \mathbf{B}\mathbf{w}_i} =$$

Optimization with the Rayleigh Coefficient

We define the Rayleigh coefficient wrt the sym. matrices $\mathbf{A}$ and $\mathbf{B}$ as

$$R_{\mathbf{A},\mathbf{B}}(\mathbf{w}) = \frac{\mathbf{w}^\top \mathbf{A}\mathbf{w}}{\mathbf{w}^\top \mathbf{B}\mathbf{w}}.$$

The minimum and maximum of R can be obtained by the following theorem:

*The **Min-Max Theorem** states: $d_1 \leq R_{\mathbf{A},\mathbf{B}}(\mathbf{w}) \leq d_C$, if $d_1 \leq \cdots \leq d_C$ are the **generalized Eigenvalues** of $\mathbf{A}$ and $\mathbf{B}$.*

Let $\mathbf{w}_i$ be the corresponding Eigenvectors (i.e., $\mathbf{A}\mathbf{w}_i = \mathbf{B}\mathbf{w}_i d_i$). Then

$$R_{\mathbf{A},\mathbf{B}}(\mathbf{w}_i) = \frac{\mathbf{w}_i^\top \mathbf{A}\mathbf{w}_i}{\mathbf{w}_i^\top \mathbf{B}\mathbf{w}_i} = \frac{\mathbf{w}_i^\top \mathbf{B}\mathbf{w}_i d_i}{\mathbf{w}_i^\top \mathbf{B}\mathbf{w}_i} = d_i \tag{1}$$

Accordingly, the min (max) of R is attained for $\mathbf{w}_1$ (for $\mathbf{w}_C$).

CSP Optimization Problem

Let $\mathbf{X}_i \in \mathbb{R}^{C \times T_i}$ be the concatenation of all band-pass filtered trials of class i along the time dimension (T_i is the total number of time points of all trials of class i, and C being the number of channels).

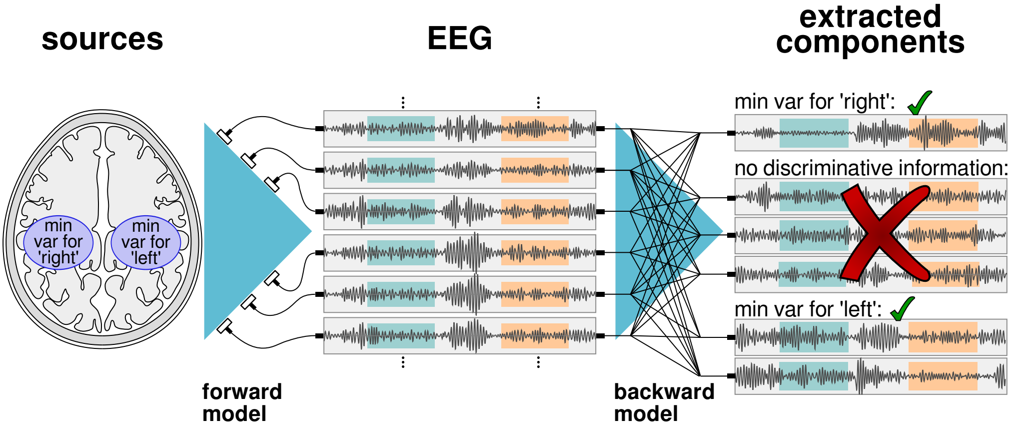
$\mathbf{\Sigma}_i = \frac{1}{T_i - 1} \mathbf{X}_i \mathbf{X}_i^\top \in \mathbb{R}^{C \times C}$ are the corresponding covariance matrices (mean does not need to be subtracted – it is zero anyway, due to band-pass filtering).

Then the **CSP** filter $\mathbf{w}_1$ that maximizes variance for class 1 is determined by the following optimization:

$$\underset{\mathbf{w} \in \mathbb{R}^C}{\operatorname{argmax}} \frac{\operatorname{var}(\mathbf{w}^\top \mathbf{X}_1)}{\operatorname{var}(\mathbf{w}^\top \mathbf{X}_1) + \operatorname{var}(\mathbf{w}^\top \mathbf{X}_2)} = \underset{\mathbf{w} \in \mathbb{R}^C}{\operatorname{argmax}} \frac{\mathbf{w}^\top \mathbf{\Sigma}_1 \mathbf{w}}{\mathbf{w}^\top (\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2) \mathbf{w}} \tag{2}$$

According to the Min-Max Theorem this optimization can be solved by generalized Eigenvalue decomposition. So we arrive at the same solution as before.

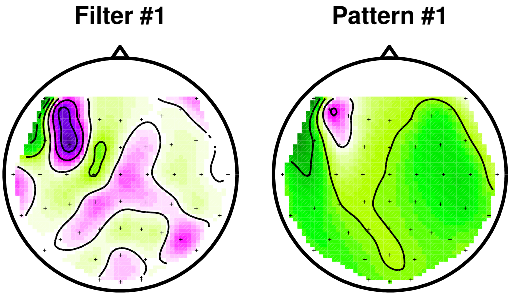
CSP analysis can be interpreted in terms of our linear model of the EEG:

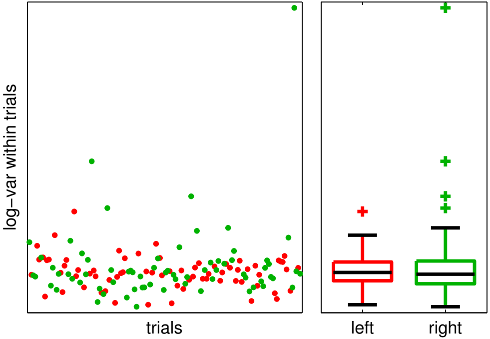


Note: For CSP analysis it does not matter, whether the extracted components correspond to single sources. The aim is just to extract most discriminative components.

CSP is prone to outliers

CSP filter/pattern corresponding to the ‘best’ eigenvalue (0.95) in a data set containing some artifacts:





This CSP solution is highly influenced by one single-trial in which channel FC3 has a high variance.

Note that the class-specific box-plots on the right show no difference in median of the variances (black line).

Remedy for CSP's Susceptibility

The first approach would be to make the estimation of CSP filters more robust (e.g., by a robust estimation of covariance matrices). But so far no convincing solution has been found.

Remedy for CSP's Susceptibility

Another approach is to have a different selection criterium for CSP filters. The extreme Eigenvalues may occur due to outliers. A simple and effective way is to calculate the variance of the filtered signal within each trial and then calculate the corresponding ratio of medians:

$$\text{score}(\mathbf{w}_j) = \frac{\mathrm{med}_j^{(1)}}{\mathrm{med}_j^{(1)} + \mathrm{med}_j^{(2)}}$$

where $\mathrm{med}_j^{(c)} = \text{median}_{i \in \mathcal{I}_c} \left(\mathbf{w}_j^\top \mathbf{X}_i \mathbf{X}_i^\top \mathbf{w}_j \right)$, $\mathbf{X}_i$ is the matrix of trial i (band-pass filtered signals), and $\mathcal{I}_c$ are the indices of trials of class c for $c \in \{1, 2\}$.
As with eigenvalues, a 'ratio-of-medians' score near 1 or near 0 indicates good discriminability of the corresponding spatial filter. The artifact pattern show above has a median score of 0.49.

Linear Spatial Filtering prior to CSP

Can CSP-based classification be enhanced by preprocessing the data with a linear spatial filter (like PCA, ICA or re-referencing like Laplace filtering)?
The question is difficult to answer in general, but two facts can be derived.

Let $\mathbf{B} \in \mathbb{R}^{C \times C_0}$ be the matrix representing arbitrary linear spatial filters (with dimension reduction if $C_0 < C$). We denote all variables corresponding to the $\mathbf{B}$-filtered signals by $\tilde{\cdot}$, the signals are $\tilde{\mathbf{X}} = \mathbf{B}^\top \mathbf{X}$. This implies $\tilde{\mathbf{\Sigma}}_1 = \mathbf{B}^\top \mathbf{\Sigma}_1 \mathbf{B}$ and $\tilde{\mathbf{\Sigma}}_2 = \mathbf{B}^\top \mathbf{\Sigma}_2 \mathbf{B}$.
The filter matrices calculated by CSP are denoted by $\mathbf{W}$ and $\tilde{\mathbf{W}}$.

Linear Spatial Filtering prior to CSP

Case 1: Matrix $\mathbf{B}$ is invertible. The CSP solution for $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ implies

$$\begin{aligned}(\mathbf{B}^{-1}\mathbf{W})^\top \tilde{\mathbf{\Sigma}}_1 \mathbf{B}^{-1}\mathbf{W} &= \mathbf{D} \\ (\mathbf{B}^{-1}\mathbf{W})^\top (\tilde{\mathbf{\Sigma}}_1 + \tilde{\mathbf{\Sigma}}_2) \mathbf{B}^{-1}\mathbf{W} &= \mathbf{I}\end{aligned}$$

which means that $\mathbf{B}^{-1}\mathbf{W}$ is a solution to the simultaneous diagonalization of $\tilde{\mathbf{\Sigma}}_1$ and $\tilde{\mathbf{\Sigma}}_2$. Since the solution is unique up to the sign of the columns, we obtain

$$\tilde{\mathbf{W}}\mathbf{\Lambda} = \mathbf{B}^{-1}\mathbf{W} \quad \text{for some diagonal } \mathbf{\Lambda} \text{ with } \quad \lambda_j \in \{-1, +1\}.$$

Accordingly, the filtered signals are identical up to the sign:
$\mathbf{W}^\top \mathbf{X} = \mathbf{\Lambda}\tilde{\mathbf{W}}^\top \mathbf{B}^\top \mathbf{X} = \mathbf{\Lambda}\tilde{\mathbf{W}}^\top \tilde{\mathbf{X}}$, so the features, the classifier and the classification performance **do not change**.

Linear Spatial Filtering prior to CSP

Case 2: Matrix $\mathbf{B}$ is not invertible. We use the formulation as Rayleigh coefficient.

$$\max_{\tilde{\mathbf{w}} \in \mathbb{R}^{C_0}} \frac{\tilde{\mathbf{w}}^\top \tilde{\mathbf{\Sigma}}_1 \tilde{\mathbf{w}}}{\tilde{\mathbf{w}}^\top (\tilde{\mathbf{\Sigma}}_1 + \tilde{\mathbf{\Sigma}}_2) \tilde{\mathbf{w}}} = \max_{\tilde{\mathbf{w}} \in \mathbb{R}^{C_0}} \frac{\tilde{\mathbf{w}}^\top \mathbf{B}^\top \mathbf{\Sigma}_1 \mathbf{B} \tilde{\mathbf{w}}}{\tilde{\mathbf{w}}^\top \mathbf{B}^\top (\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2) \mathbf{B} \tilde{\mathbf{w}}}$$
$$\leq \max_{\mathbf{w} \in \mathbb{R}^{C}} \frac{\mathbf{w}^\top \mathbf{\Sigma}_1 \mathbf{w}}{\mathbf{w}^\top (\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2) \mathbf{w}}$$

since every term on the left hand side of the inequality is covered on the right hand side for $\mathbf{w} = \mathbf{B}\tilde{\mathbf{w}}$. That means, the CSP-optimum for the unfiltered signals (right hand side) is greater than or equal to the CSP-optimum for the signals filtered by $\mathbf{B}$ (left hand side). Accordingly, the objective of CSP analysis (on the training data) can only get worse by using $\mathbf{B}$.

However, this result holds only for the training data. If the prefiltering reduces artifacts, it is well possible that the generalization performance of CSP improves. On the other hand, the prefiltering could also discard discriminative information which would be detrimental for performance.

Part II

BCI as a Tool to Improve Devices, Interfaces, Infrastructure

- The approach relies on estimating brain-based measures of perception and cognitive states during interaction.

BCI as a Tool to Improve Devices, Interfaces, Infrastructure

- The approach relies on estimating brain-based measures of perception and cognitive states during interaction.
- E.g., a navigation system for a car may be optimized based on neural measures of workload or focused attention.

  [Kohlmorgen et al, 2007; Blankertz et al, 2010]

BCI as a Tool to Improve Devices, Interfaces, Infrastructure

- The approach relies on estimating brain-based measures of perception and cognitive states during interaction.
- E.g., a navigation system for a car may be optimized based on neural measures of workload or focused attention.
  [Kohlmorgen et al, 2007; Blankertz et al, 2010]
- Recent work employs neurotechnology to assess safety-critical aspects of infrastructure, such as harbors and bridges that require demanding maneuvers.
  [Miklody at al, 2016; Miklody et al, submitted]

BCI as a Tool to Improve Devices, Interfaces, Infrastructure

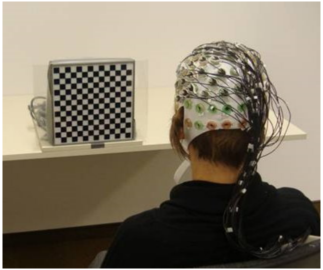
- The approach relies on estimating brain-based measures of perception and cognitive states during interaction.
- E.g., a navigation system for a car may be optimized based on neural measures of workload or focused attention.
  [Kohlmorgen et al, 2007; Blankertz et al, 2010]
- Recent work employs neurotechnology to assess safety-critical aspects of infrastructure, such as harbors and bridges that require demanding maneuvers.
  [Miklody at al, 2016; Miklody et al, submitted]
- Other papers demonstrate the potential for assessing the quality of audio and video codecs.
  [Wenzel et al, 2016; Acqualagna et al, 2015; Scholler et al, 2012, Porbadnigk et al, 2010,2011,2013]

Neuro-guided Product Development: Lighting Technology

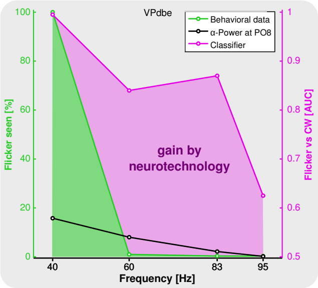
Aim:

Determine the threshold frequency that avoids *neural flicker*.

[Research in cooperation with Philips Research, Light and Brain]



- Stimuli: constant light, or flickering (40 to 100 Hz)
- Task: indicate whether stimulus was flickering



[Porbadnigk et al, 2011]

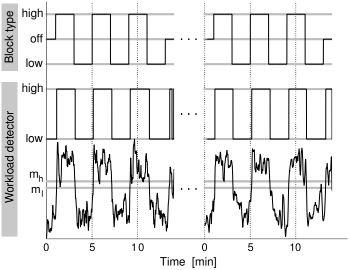
Similarly: **Enhance 3D-TV** [Wenzel et al, 2016]
[Joint studies with Sony AG]

Aim: Avoid that cognitive load during driving reaches dangerous levels:

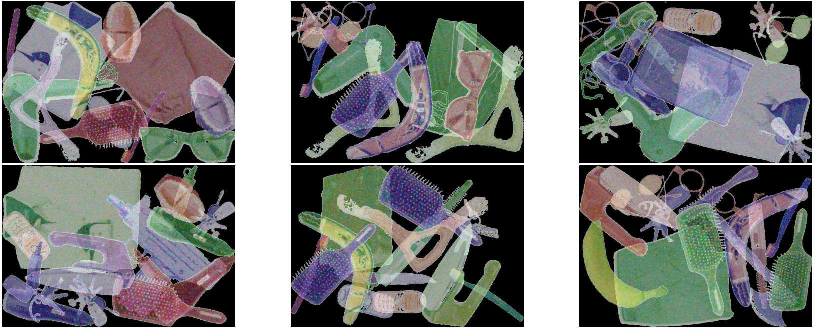
- *evaluate* new product features to avoid elevation of cognitive load
- develop and *validate* new features to reduce workload

[Research in cooperation with Daimler AG] [Kohlmorgen et al, 2007]

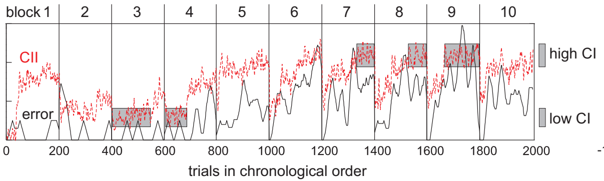
- Car driving on the highway
- Successful mitigation based on real-time measure of cognitive load

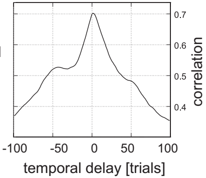


[Research in cooperation with Siemens AG]



Monotoneous, but demanding detection task (weapons in x-ray pictures)

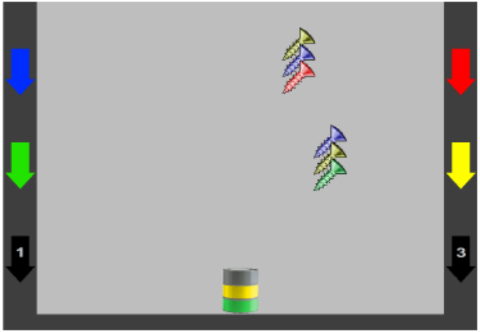


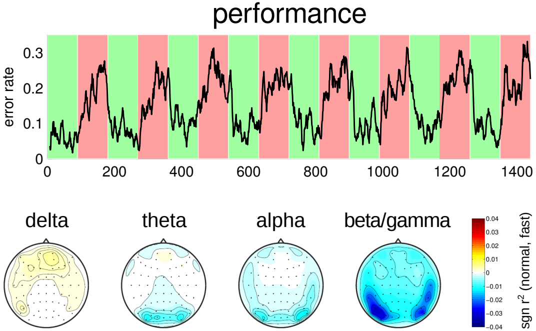


[Müller et al, 2008; Blankertz et al, 2010]

Aim

Mimicking task demands of industrial workplace under controlled laboratory conditions.



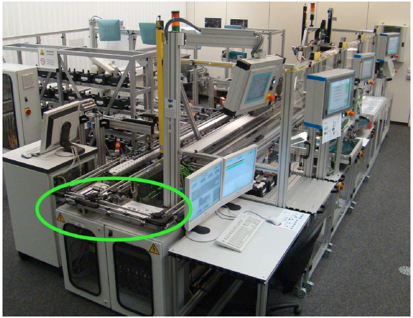


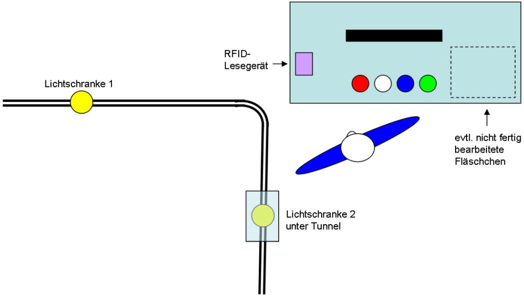
- Catching game controllable by touch display
- Variation of workload: Dropping of screws in two different speeds (90s intervals)
- N=6 participants [Schultze-Kraft et al, 2016]

Transfer to the Real Industrial Environment

Aim

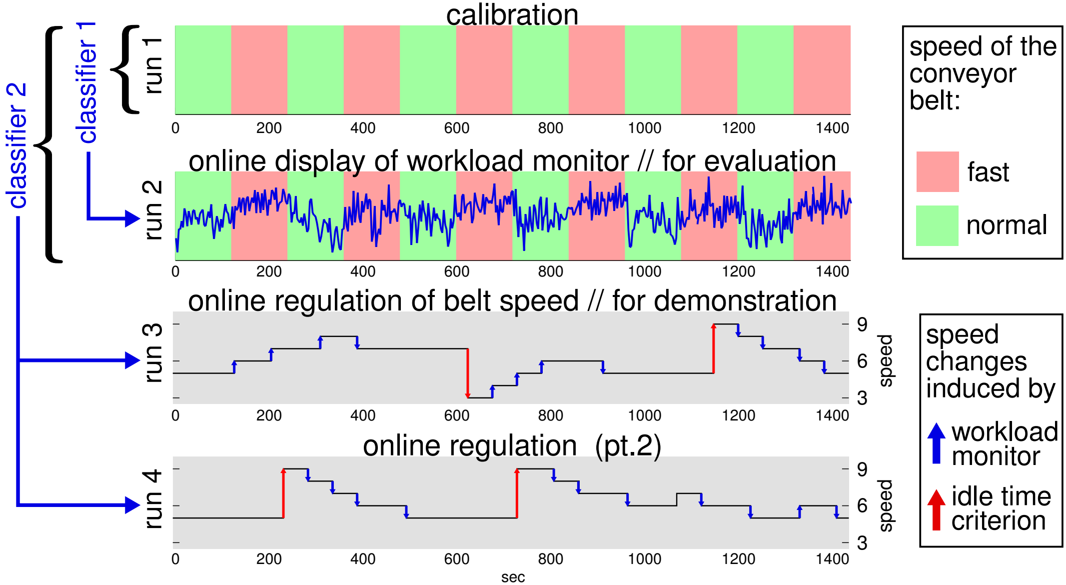
Adjust conveyor belt speed to momentary workload

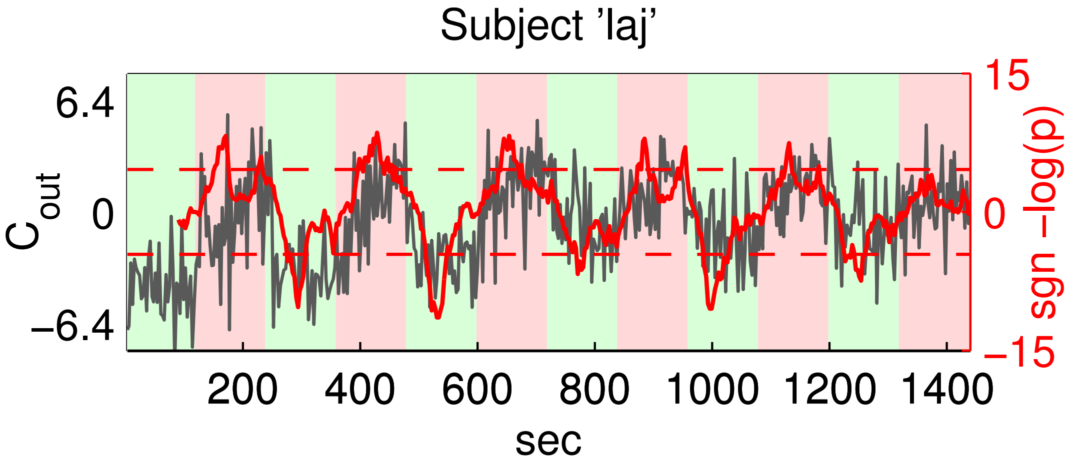


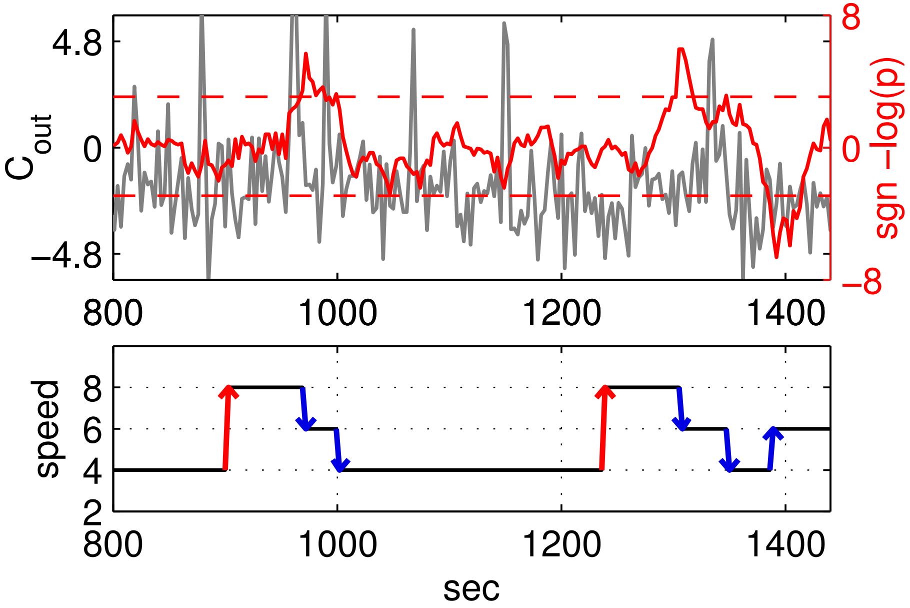


[Research in cooperation with Siemens AG]

Experimental Design




Subject 'laj'
6.4
0
–6.4
C_{out}
15
0
–15
sgn –log(p)
200
400
600
800
1000
1200
1400
sec


4.8
0
−4.8
C_out
8
0
−8
sgn −log(p)
800
1000
1200
1400
8
6
4
2
speed
800
1000
1200
1400
sec

Neurotechnolgy to Enhance Safety in the Maritime World

- *Adapting the training to learning progress*
- *Assess demand (wrt. safety) during passing under bridges and port entry*

[Miklody et al, 2016]

Lessons Learnt

After this lecture you should

- be familiar with the derivation via the Rayleigh coefficient as a general approach to optimization problems,
- be aware of the sensitivity of CSP to outliers and
  know a way how to avoid the selection of deteriorated filters,
- have an overview of BCI applications based on the modulation of brain rhythms

Feedback for Today's Lecture

Go to mentimeter questionnaire: at www.menti.com with **71 04 67**

https://www.mentimeter.com/s/6035114b72fe0d4c9b5553d6ad4ce7b0/800187d660ca

References I

- Acqualagna, L., Bosse, S., Porbadnigk, A. K., Curio, G., Müller, K.-R., Wiegand, T., and Blankertz, B. (2015).
  EEG-based classification of video quality perception using steady state visual evoked potentials (SSVEPs).
  J Neural Eng, 12(2):026012.
- Blankertz, B., Tangermann, M., Vidaurre, C., Dickhaus, T., Sannelli, C., Popescu, F., Fazli, S., Danóczy, M., Curio, G., and Müller, K.-R. (2010a).
  Detecting mental states by machine learning techniques: The Berlin Brain-Computer Interface.
  In Allison, B., Graimann, B., and Pfurtscheller, G., editors, *Brain-Computer Interfaces (Revolutionizing Human-Computer Interaction)*, The Frontiers Collection, pages 113–135. Springer, Berlin Heidelberg.
- Blankertz, B., Tangermann, M., Vidaurre, C., Fazli, S., Sannelli, C., Haufe, S., Maeder, C., Ramsey, L. E., Sturm, I., Curio, G., and Müller, K.-R. (2010b).
  The Berlin Brain-Computer Interface: Non-medical uses of BCI technology.
  Front Neuroscience, 4:198. Open Access.
- Blankertz, B., Tomioka, R., Lemm, S., Kawanabe, M., and Müller, K.-R. (2008).
  Optimizing spatial filters for robust EEG single-trial analysis.
  IEEE Signal Process Mag, 25(1):41–56.
- Kohlmorgen, J., Dornhege, G., Braun, M., Blankertz, B., Müller, K.-R., Curio, G., Hagemann, K., Bruns, A., Schrauf, M., and Kincses, W. (2007).
  Improving human performance in a real operating environment through real-time mental workload detection.
  In Dornhege, G., del R. Millán, J., Hinterberger, T., McFarland, D., and Müller, K.-R., editors, *Toward Brain-Computer Interfacing*, pages 409–422. MIT press, Cambridge, MA.
- Müller, K.-R., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., and Blankertz, B. (2008).
  Machine learning for real-time single-trial EEG-analysis: From brain-computer interfacing to mental state monitoring.
  J Neurosci Methods, 167(1):82–90.
- Porbadnigk, A. K., Antons, J.-N., Blankertz, B., Treder, M. S., Schleicher, R., Möller, S., and Curio, G. (2010).
  Using ERPs for assessing the (sub)conscious perception of noise.
  In *Conf Proc IEEE Eng Med Biol Soc*, volume 2010, pages 2690–2693.

References II

- Porbadnigk, A. K., Scholler, S., Blankertz, B., Ritz, A., Born, M., Scholl, R., Müller, K.-R., Curio, G., and Treder, M. S. (2011).
  Revealing the neural response to imperceptible peripheral flicker with machine learning.
  In *Conf Proc IEEE Eng Med Biol Soc*, volume 2011, pages 3692–3695.
- Porbadnigk, A. K., Treder, M. S., Blankertz, B., Antons, J.-N., Schleicher, R., Möller, S., Curio, G., and Müller, K.-R. (2013).
  Single-trial analysis of the neural correlates of speech quality perception.
  J Neural Eng, 10(5):056003.
- Scholler, S., Bosse, S., Treder, M. S., Blankertz, B., Curio, G., Müller, K.-R., and Wiegand, T. (2012).
  Towards a direct measure of video quality perception using EEG.
  IEEE Trans Image Process, 21(5):2619–2629.
- Schultze-Kraft, M., Dähne, S., Gugler, M., Curio, G., and Blankertz, B. (2016).
  Unsupervised classification of operator workload from brain signals.
  J Neural Eng, 13(3):036008. Open Access.
- Wenzel, M. A., Schultze-Kraft, R., Meinecke, F. C., Cardinaux, F., Kemp, T., Müller, K.-R., Curio, G., and Blankertz, B. (2016).
  EEG-based usability assessment of 3d shutter glasses.
  J Neural Eng, 13(1):016003.