EEG Features and Spatial Filters towards “practical” BCI design

Fabien Lotte

Inria Bordeaux Sud-Ouest

France

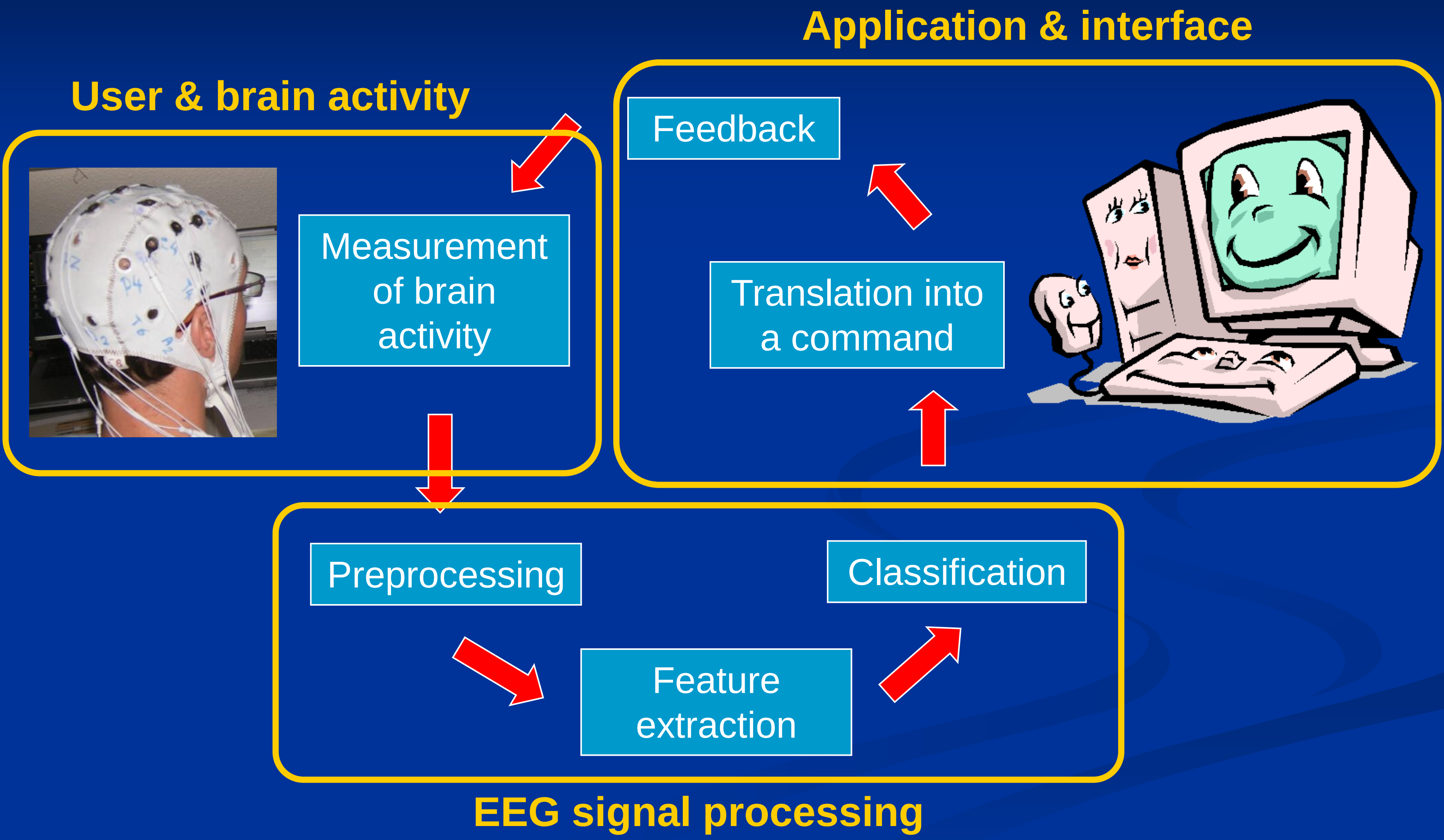
Designing a practical BCI?

- What is a « practical » BCI?
  - A BCI that can be used in real-life
  - A BCI that is (ideally) efficient anytime, anywhere
    - A least dependable! [Millan, BBCI Workshop, 2012]
- In other words
  - High classification accuracy
  - Robust to noise and non-stationarities
  - Short or no calibration time

Take home message

- A form of spatial filtering is essential
- Using a-priori knowledge helps

Architecture of a BCI



Application & interface

User & brain activity

Feedback

Measurement of brain activity

Translation into a command

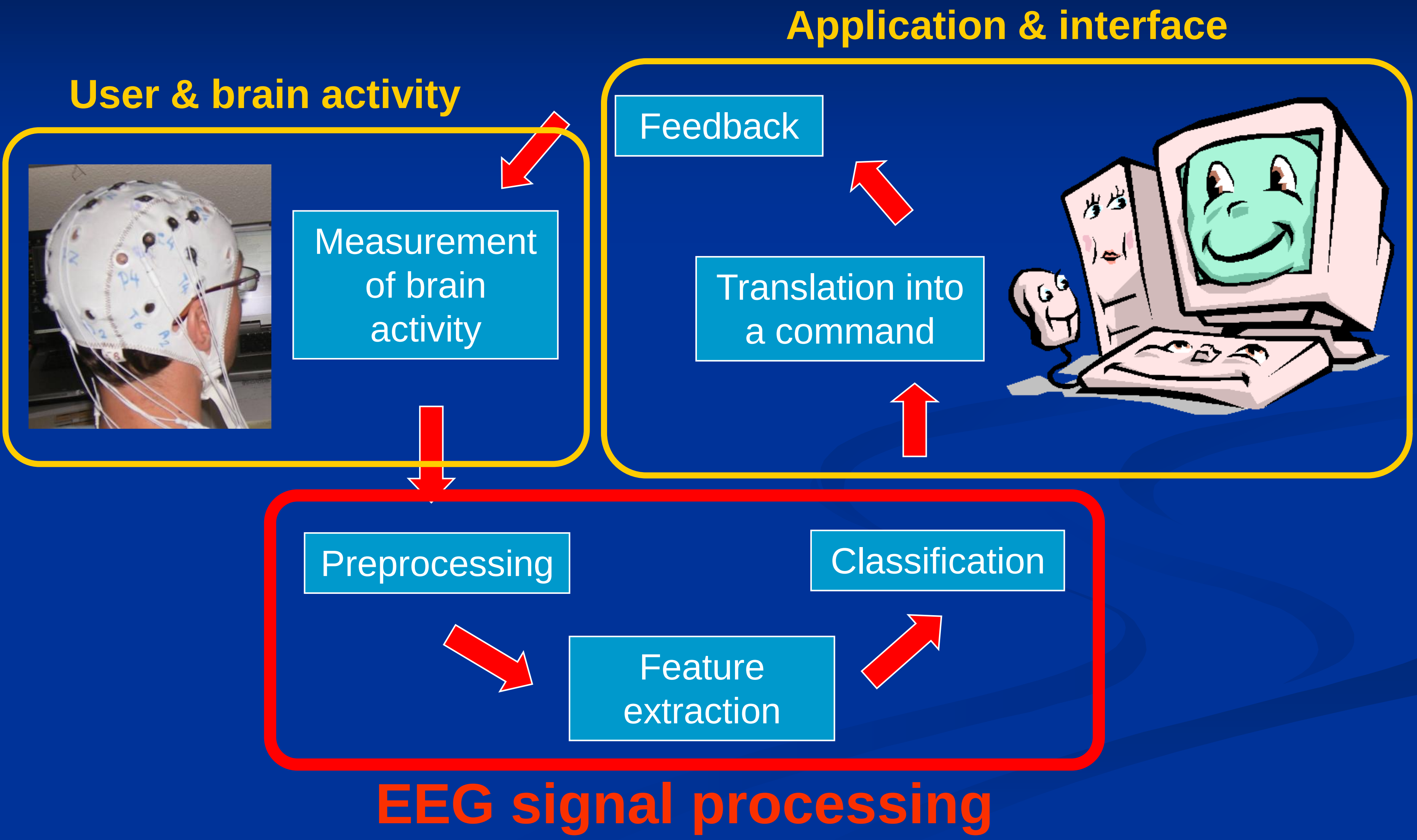
Preprocessing

Classification

Feature extraction

EEG signal processing

Today’s talk



Application & interface

User & brain activity

Measurement of brain activity

Feedback

Translation into a command

Preprocessing

Classification

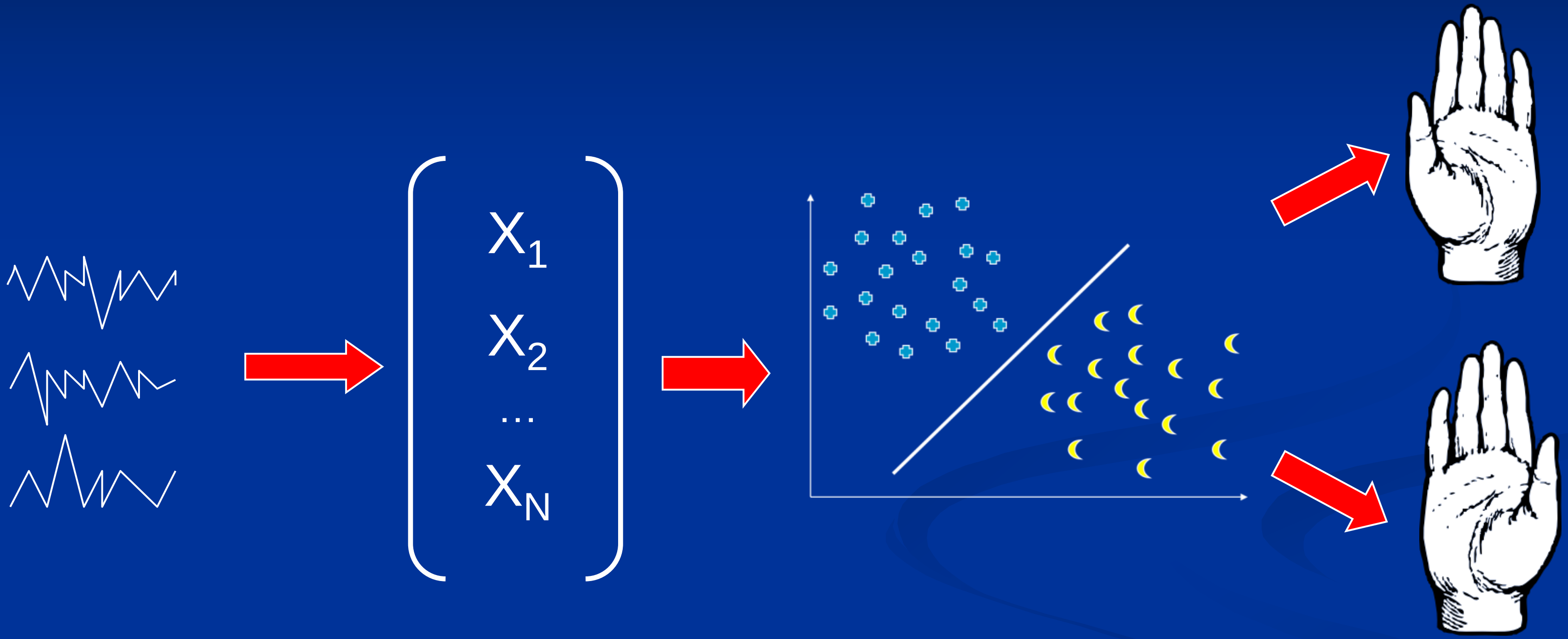
Feature extraction

EEG signal processing

Outline

1. Basic features for oscillatory activity
2. Inverse solution-based BCI
3. Common Spatial Patterns and extensions
4. Spatial filters for Evoked Potentials
5. Alternative features

Feature extraction in context: the pattern recognition approach



EEG signals

Ex: signal recorded during left or right hand motor imagery

Feature extraction

Ex: band power in the μ and β rhythms for electrodes located over the motor cortex

Classification

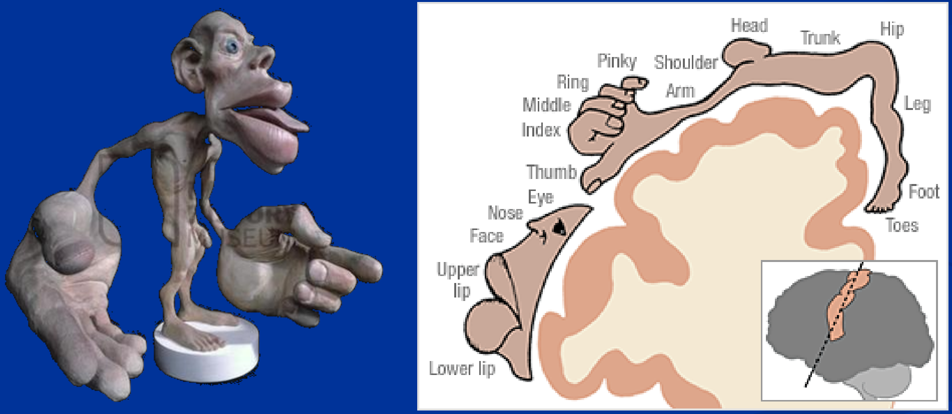
Ex: Linear Discriminant Analysis (LDA)

Estimated class

Ex: Left or Right (imagined hand movement)

Oscillatory activity based BCI

- Example: Motor Imagery (MI)
  - Imagination of limb movements
  - Contralateral ERD in μ (~8-12 Hz) or β (~16-24 Hz) during MI + β ERS (rebound) after MI



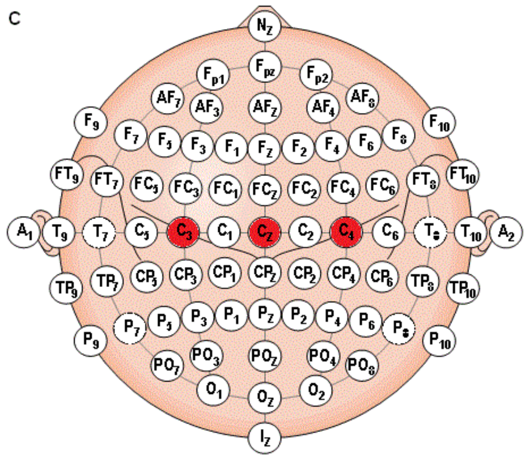
Penfield homonculus [Penfield54]

ERD/ERS = Event Related (De)Synchronization

Pfurtscheller & Neuper, “Motor imagery and direct brain-computer communication”, Proceedings of the IEEE, 2001

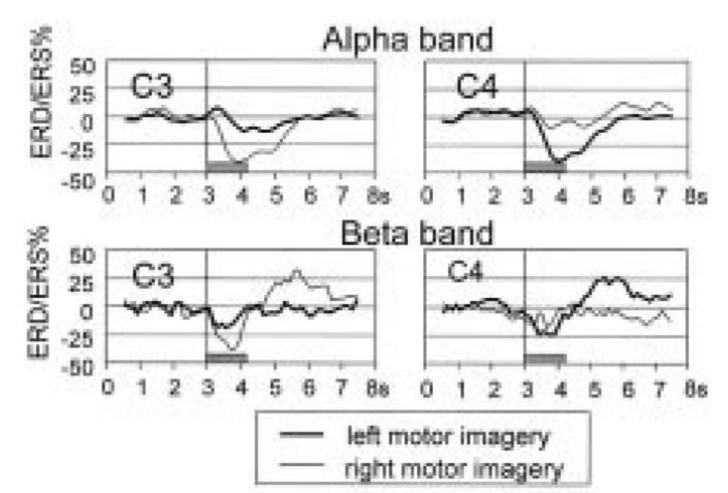
Example:
Features for MI-based BCI

- **Spatial** information
  - focusing on channels
    - C3: right hand MI
    - Cz: foot MI
    - C4: left hand MI



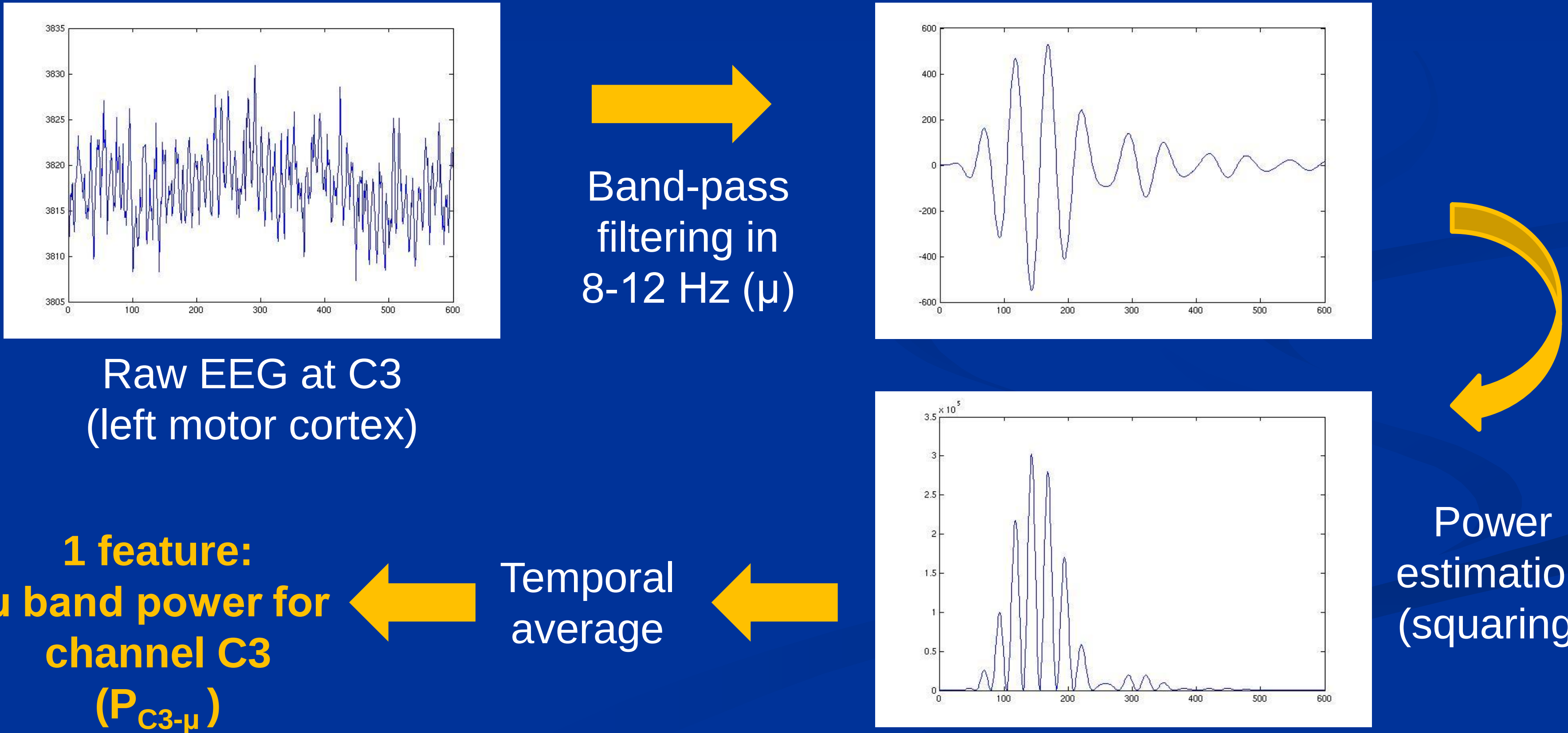
Channels C3, Cz, C4

- **Spectral** information
  - Focusing on frequency bands
    - μ (mu: ~8-12 Hz)
    - β (beta: ~16-24 Hz)



Band power features

- Signal power in a given frequency band



Raw EEG at C3
(left motor cortex)

Band-pass
filtering in
8-12 Hz (μ)

Power
estimation
(squaring)

Temporal
average

1 feature:
μ band power for
channel C3
($P_{C3-\mu}$)

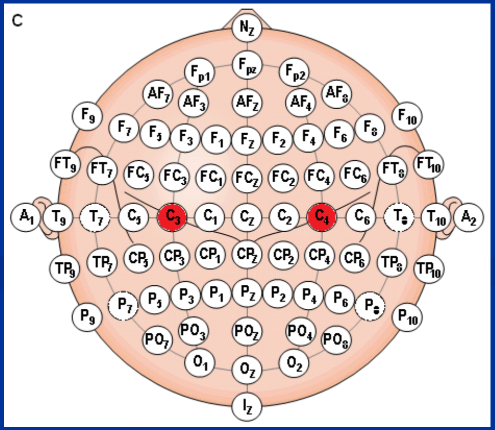
Band power features (2)

- Other ways to compute them
  - Periodogram (Fourier Decomposition)
  - Power spectral density from AutoRegressive (AR) coefficients
  - Wavelet scalogram (time-scale representation)
  - Spectrogram (time-frequency decomposition then spectrums are averaged over time)
  - Etc.

Herman et al, IEEE TNSRE, 2008
Brodu et al, IEEE SSCI, 2011

Basic features for left and right hand motor imagery

- Computing band power features P
  - In frequency bands
    - μ (8-12 Hz) & β (16-24 Hz)
  - For channels
    - C3 & C4
- Gathering them into a feature vector
  $v = [P_{C3\text{-}\mu}, P_{C4\text{-}\mu}, P_{C3\text{-}\beta}, P_{C4\text{-}\beta}]$
- v is used as input to the classifier



C3 and C4
electrode location

That's all? It's that simple?

- Yes… but this basic design is far from being optimal
  - Only 2 channels

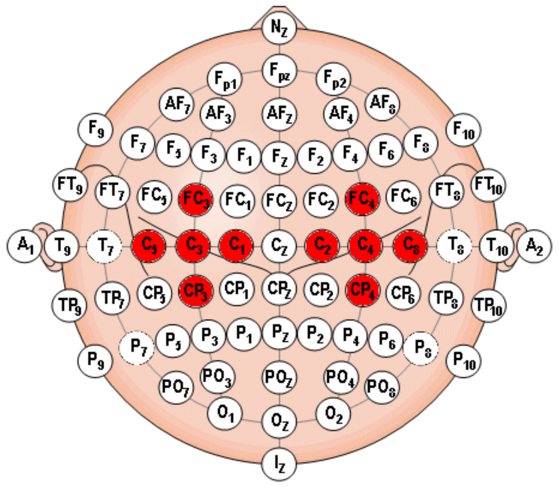
    ⇒Information might be missing
  - Fixed channels (C3 & C4)

    ⇒The optimal channels are subject-dependant
  - Fixed frequency bands (8-12 Hz, 16-24 Hz)

    ⇒The optimal frequency bands are subject-dependant

Using more channels

- Extracting features from neighboring channels as well
- Problem
  - More channels
  - ⇒ More features
  - ⇒ Need for more training data
  - Redundancy and correlation between channels
- Solution
  - Spatial filtering!



Spatial Filtering

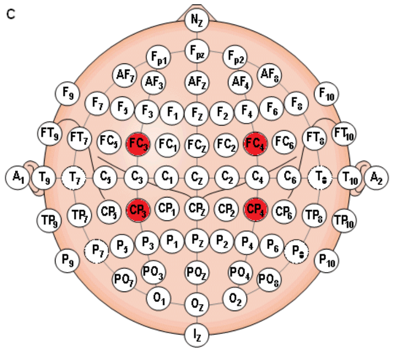
- Definition
  - Using a small number of new channels defined as a linear combination of the original ones

$$x' = \sum_i w_i x_i = wX$$

- Due to the smearing effect of the skull and brain, the underlying brain signal is spread over several channels

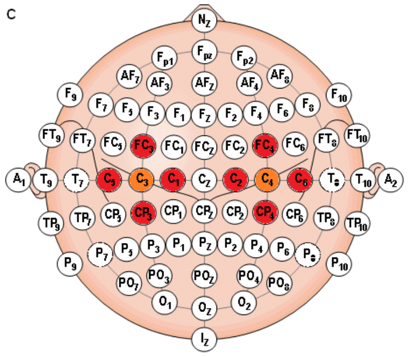
Some basic spatial filters

- Bipolar filters
  - C3’ = FC3 - CP3



Bipolar

- Laplacian filters
  - C3’= 4*C3 - FC3 - C5 - C1 - CP3



Laplacian

[McFarland et al, EEG and Clinical Neurophysiology, 1997]

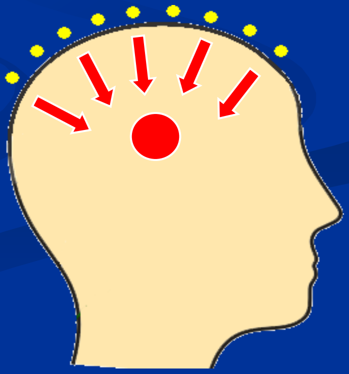
Inverse Solutions

- Context
  - EEG are scalp measurements ***m*** resulting from the mixing ***A*** of several unknown sources **s**

$$m=As$$ ← **forward model**

  - The brain is generally modeled as a 3D grid of voxels

- Inverse solutions [Baillet01]
  - Estimate the sources **s** from the scalp measurements ***m***



$$s=Tm$$ ← **Inverse/backward model**

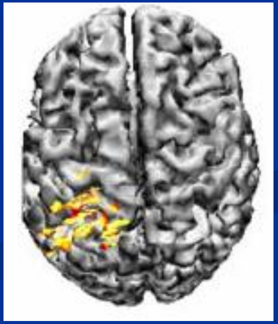
Some remarks

- Need for a head model **A**
  - Forward model: ***m*=*A*s**
  - model the conduction effect of the skull, skin, etc.
  - can be realistic (from MRI) or generic (spheres)
- The *inverse problem* is ill-posed
  - Inverse model: **s=*Tm***
  - An infinity of possible solutions
  - Constraints are needed
- Inverse solutions CANNOT recover the activity of individual neurons from scalp EEG
  - the information is not in EEG signals

2 main types of inverse solutions

- Equivalent dipole solutions (a.k.a parametric)
  - estimate the position, amplitude and orientation of few sources
  - each source is modelled by an equivalent dipole

- Distributed solutions (a.k.a non-parametric)
  - estimate the amplitudes and orientations of many voxels distributed in all the cortex/brain



From [Qin et al, 2005]

Inverse solutions for BCI

- Mostly explored for motor imagery BCI
- Equivalent dipole approaches
  - The location of the source dipole is used to identify the mental task
  - May require strong a-priori knowledge

    [Kamousi et al 2005] [Qin et al 2004]
- Distributed approaches
  - Relevant brain voxels need to be identified
  - These voxels current density is used as features
  - Features from the source space have been shown to outperform features from the sensor space

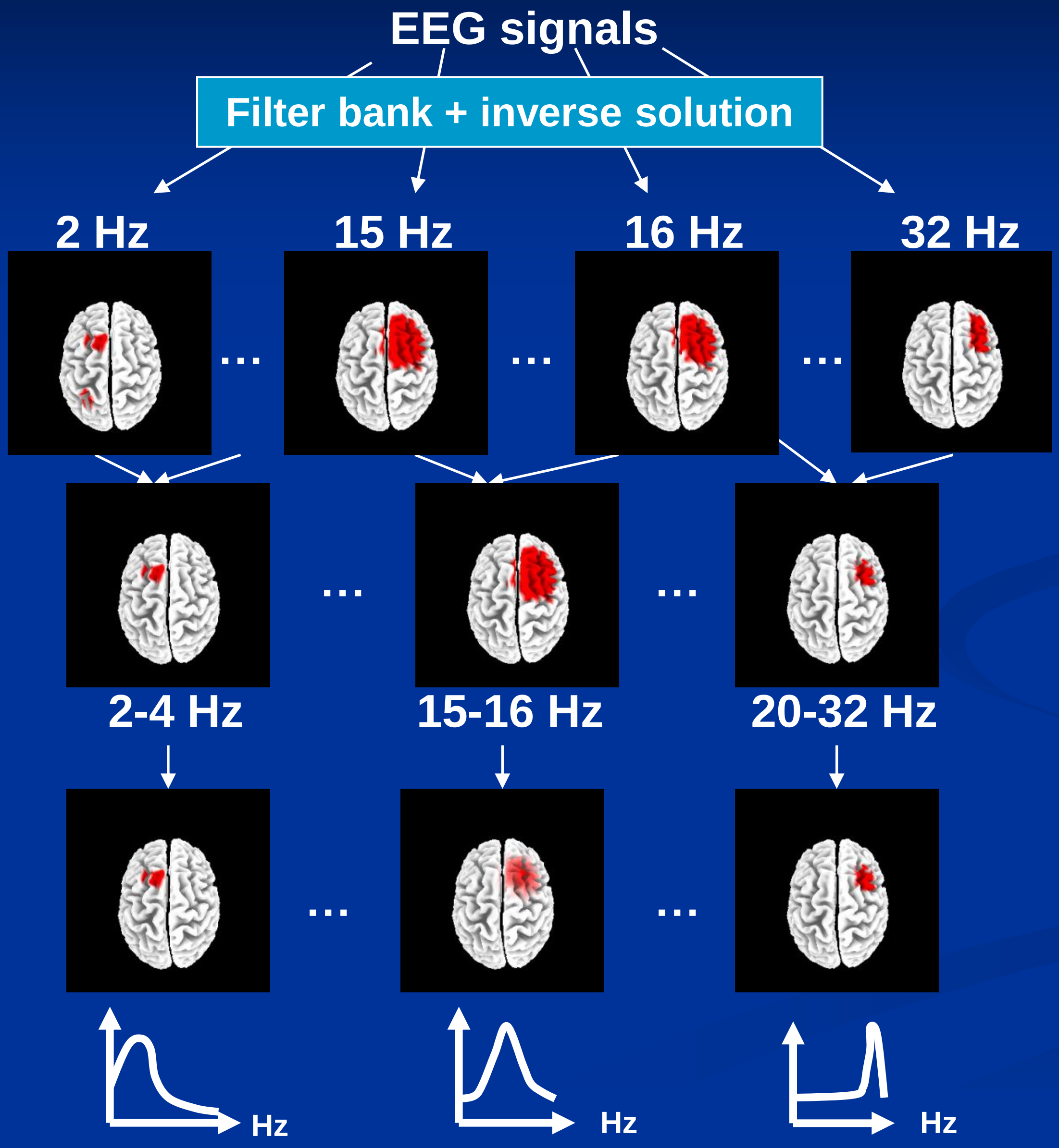
[Congedo, Lotte & Lecuyer 2006] [Lehembre et al 2008] [Menendez et al 2005][Besserve et al, 2011]

Example

- The inverse solution-based feature extraction algorithm FuRIA
  - **Fu**zzy **R**egion of **I**nterest **A**ctivity
  - Based on the sLORETA inverse solution [Pascual-Marqui01]
  - Learning
    - Automatic identification of Regions Of Interest (ROI) and frequency bands whose activity should discriminate mental states
  - Feature extraction
    - Computation of the activity in these ROI and frequency bands

Lotte, Lécuyer and Arnaldi, *IEEE Trans. on Sig Proc.*, 2009

FuRIA



EEG signals

Filter bank + inverse solution

2 Hz … 15 Hz … 16 Hz … 32 Hz

… …

2-4 Hz

15-16 Hz

20-32 Hz

… …

Training set

1. Identification of discriminant voxels and frequencies

(Statistical analysis)

2. Creation of ROI and frequency bands

(gathering through clustering)

3. Fuzzification of ROI and frequency bands

(fuzzy set theory [Zadeh96])

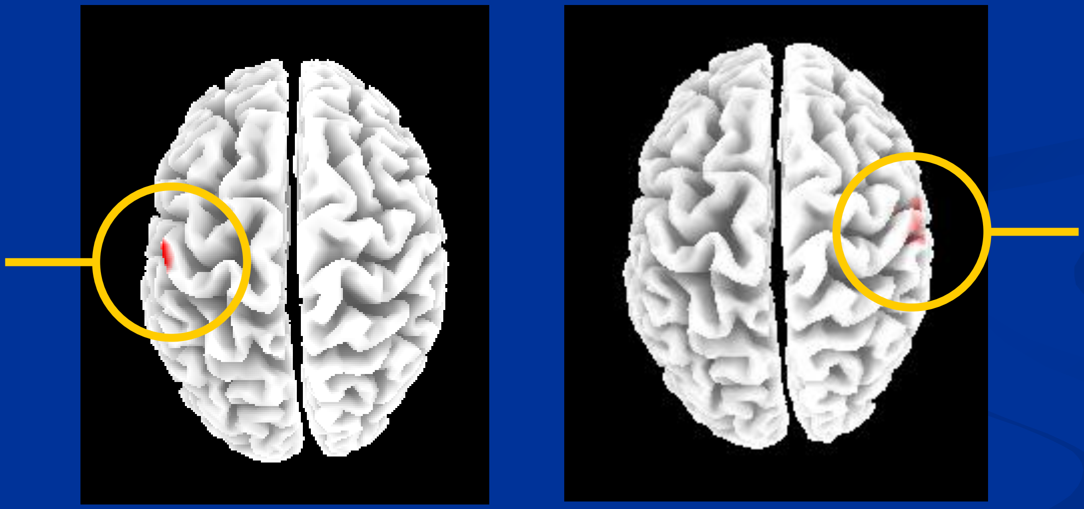
Evaluation

- EEG data set used
  - Data set IV, « BCI competition 2003 »
    - Finger movement intention, 2 classes, 1 subject
  - Data set IIIa, « BCI competition 2005 »
    - Motor imagery, 4 classes, 3 subjects

- Classifier used
  - Support Vector Machine (SVM)

Feature interpretability

- Finger movement intention data (BCI competition 2003, data set IV)
  - Only 2 features learnt



Right motor cortex

Left motor cortex

14-31 Hz

14-28 Hz

- ROI and frequencies found consistent with the literature [Pfurtscheller99, Wang04]

Results

- Comparison with BCI competition results
  [Blankertz04, Blankertz06]

BCI competition 2003 (%)	
FuRIA	**84**
Winner	**84**
2nd	81
3rd	77
4th	75

BCI competition 2005 (%)				
	S1	S2	S3	Average
FuRIA	90.56	69.17	**88.33**	82.68
Winner	86.67	**81.67**	85.00	**84.44**
2nd	92.78	57.50	78.33	76.20
3rd	**96.11**	55.83	64.17	72.04

=> Inverse solutions can lead to efficient BCI design!

Lotte, Lécuyer and Arnaldi, *IEEE Trans. on Sig Proc.*, 2009

BUT!

- Inverse solutions are “just” spatial filters
  - May be just as good as other filters (e.g., CSP)
- Inverse solutions make the problem harder
  - 10-100 EEG signals before – 1000-10000 sources after
  - Computational complexity – dimensionality reduction?
- Inverse solutions are no magic
  - There is no more information in the signals obtained with an inverse solution than in raw EEG signals.
  - A lot of information is simply missing in raw EEG signals
- Requires many EEG channels
  - May not be practical/comfortable

Why using Inverse solutions for BCI?

- Why using them?
  - Because you have a strong spatial a-priori
  - To obtain an interpretable model
  - Because you want to use connectivity
  - You don’t want to use other spatial filters

- When not using them?
  - You have few channels
  - You have few computational resources
  - You have faster and better supervised Spatial Filters

Ex: Common Spatial Patterns (CSP)!

Common Spatial Patterns (CSP) informally…

- Find spatial filters *w* such that the variance of the filtered signal is maximal for one class and minimal for the other class

  ➔ Variance of band-pass filtered signal

  (we typically use 8-30 Hz by default)

  = band-power of this frequency band

  ➔ CSP learns spatial filters that lead to optimally discriminant band-power features

Ramoser et al, "Optimal spatial filtering of single trial EEG during imagined hand movements", IEEE Trans. On Rehab. Eng., 2000

CSP formally

- It consists in extremizing

$$J(w) = \frac{wX_1X_1^Tw^T}{\underbrace{\underbrace{wX_2}_{\text{Spatially filtered signal}}X_2^Tw^T}_{\text{Variance of the spatially filtered signal}}} = \frac{wC_1w^T}{wC_2w^T}$$

C_i: EEG spatial covariance matrix for class i

w: spatial filter to optimize

X_i: multichannel EEG signals from class i

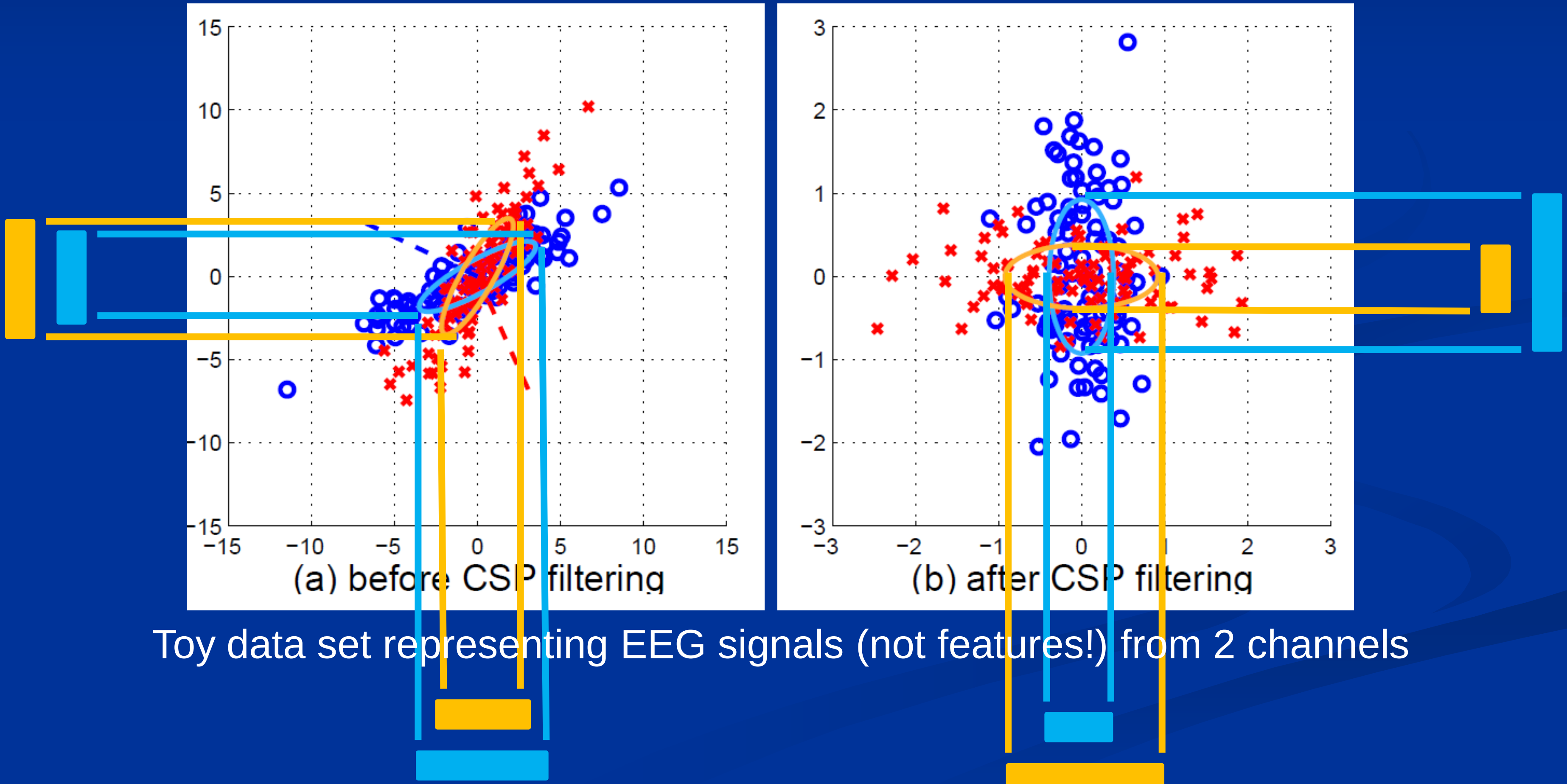
- Solved by Generalized EVD (GEVD) of C_1 and C_2
  - We typically use 6 CSP filters hence obtained
- Once the filter *w* obtained, the feature *f* used is

$$f = \log(\mathrm{var}(wX)) = \log(wCw^T)$$

CSP in action

🟨 **Average feature value for class 1**

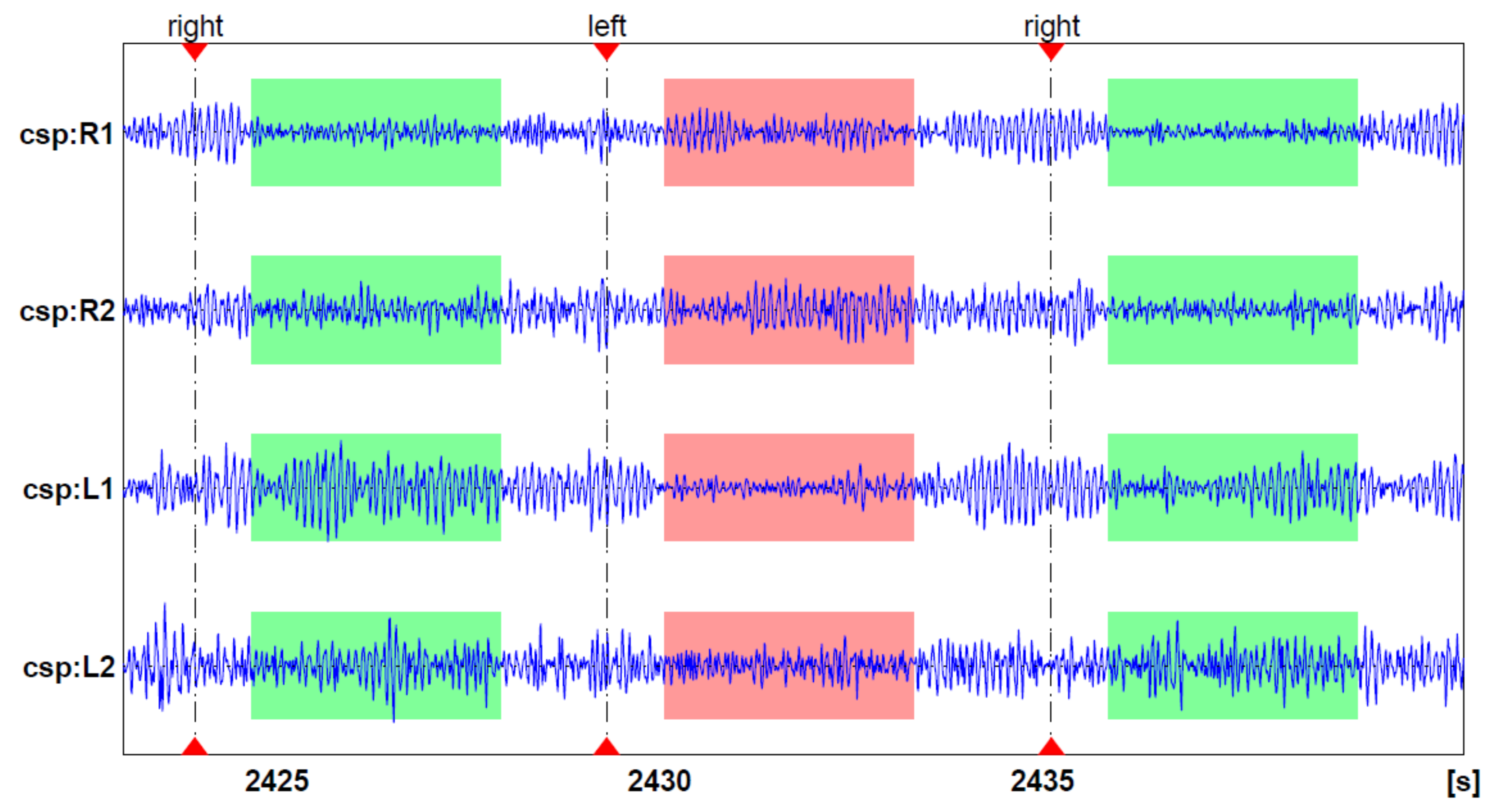
🟦 **Average feature value for class 2**



Toy data set representing EEG signals (not features!) from 2 channels

Blankertz et al, “Optimizing spatial filters for robust EEG single-trial analysis”, IEEE Signal Processing Magazine, 2008

CSP in action (2)



Examples of 4 CSP filtered signals, during left and right motor imagery

(from Blankertz et al, IEEE Sig. Proc. Mag., 2008)

CSP pros and cons

- Pros
  - Lead to high classification performances
  - Interpretable
  - Versatile (works for any ERD/ERS BCI)
  - Computationally efficient & simple to implement
  - → One of the most popular & efficient approach
- Cons
  - Non robust to noise and non-stationarities [Grosse-Wentrup08]
  - Prone to overfitting [Reuderink09]
  - Need to estimate the relevant frequency band

Towards a more robust CSP?

- How to make CSP robust and stable?
  - With limited training data
  - With noisy and non-stationary training data

- Idea: add a-priori information into the learning process
  - Use a regularization framework to penalize unlikely and/or undesired solutions (i.e., spatial filters)
  - Add a priori information to stabilize statistical estimates

Using a Regularized CSP

Regularized CSP (RCSP)

CSP	RCSP
Goal: extremizing $\frac{wC_1w^T}{wC_2w^T}$	Goal: maximizing $\frac{w\tilde{C}_1w^T}{w\tilde{C}_2w^T \boxed{+\alpha P(w)}}$ and $\frac{w\tilde{C}_2w^T}{w\tilde{C}_1w^T \boxed{+\alpha P(w)}}$ Penalty term with $\tilde{C}_i = (1-\beta)C_i \boxed{+\beta G_i}$ Stabilization term

Lotte & Guan, *IEEE Trans. BioMed. Eng.*, 2011

What prior knowledge to use? Spatial knowledge to deal with noise

- Neighboring neurons are responsible for similar brain functions + EEG is smeared due to volume conduction

  => (close) neighboring electrodes should measure similar brain signals and thus have similar contributions

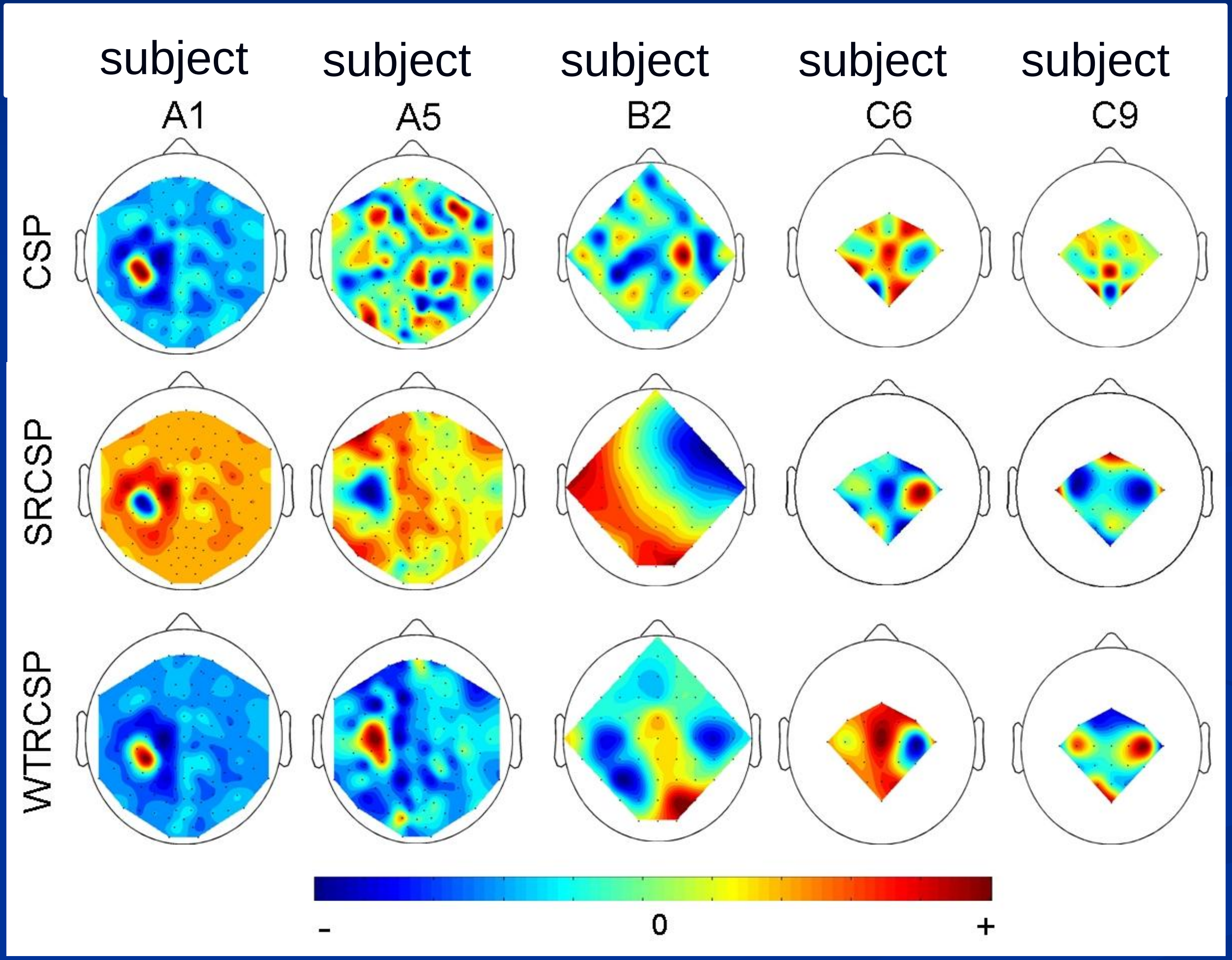
$$P(w) = \sum_{i,j} \underbrace{G(i,j)}_{\textbf{proximity of two electrodes}} \underbrace{(w_i - w_j)^2}_{\textbf{weight difference between electrodes}}$$

- For a given task, not all brain regions are involved

$$P(w) = w^T D w \quad \text{with} \quad D(i,j) = \begin{cases} \text{channel} \quad \text{i} \ \text{"uselessnes s"} & \textit{if } i = j \\ 0 & \textit{if } i \neq j \end{cases}$$

  - Can be used for subject-to-subject transfer!

Spatial filters obtained



Average classification accuracy (%)
(n = 12 subjects,
2 classes)

73.1 %

78.7 %

77.6 %

Lotte & Guan, *IEEE Trans. on Biomedical Engineering*, 2011

Regularization terms to deal with non-stationarities

- Invariant CSP [Blankertz, NIPS 09]
  - To find filters invariant to a given non-stationary noise sources (e.g., occipital alpha activity)

$$P(w) = w^T C_{noise} w$$

C_{noise}: covariance matrix of the noise source

⇒ Minimize known noise influence

- Stationary CSP [Samek, JNE 12]
  - To minimize non-stationarity

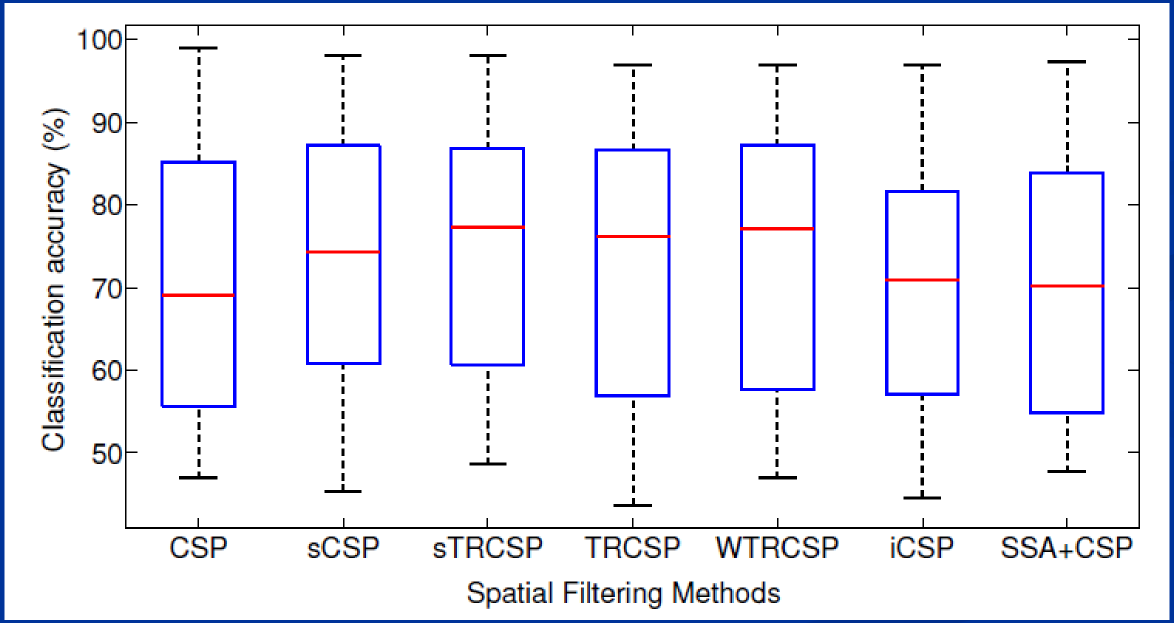
$$P(w) = w^T (\overline{\Delta}_1 + \overline{\Delta}_2) w$$

with

$$\overline{\Delta}_i = \frac{1}{N} \sum_{k=1}^{N} \mathrm{P}(C_i^k - C_i)$$

Combining multiple regularization terms

$$P(w) = \sum_i \lambda_i P_i(w)$$



Combining a Stationary + a Tikhonov regularization term (sTRCSP) [Samek et al, JNE, 2012]

Regularization terms to deal with limited data

- Covariance matrix shrinkage

$$\tilde{C}_i = (1-\beta)C_i + \beta I$$

- Using data from other subjects (previously recorded) as the stabilisation term

$$\tilde{C}_i = (1-\beta)C_i + \beta G_i$$

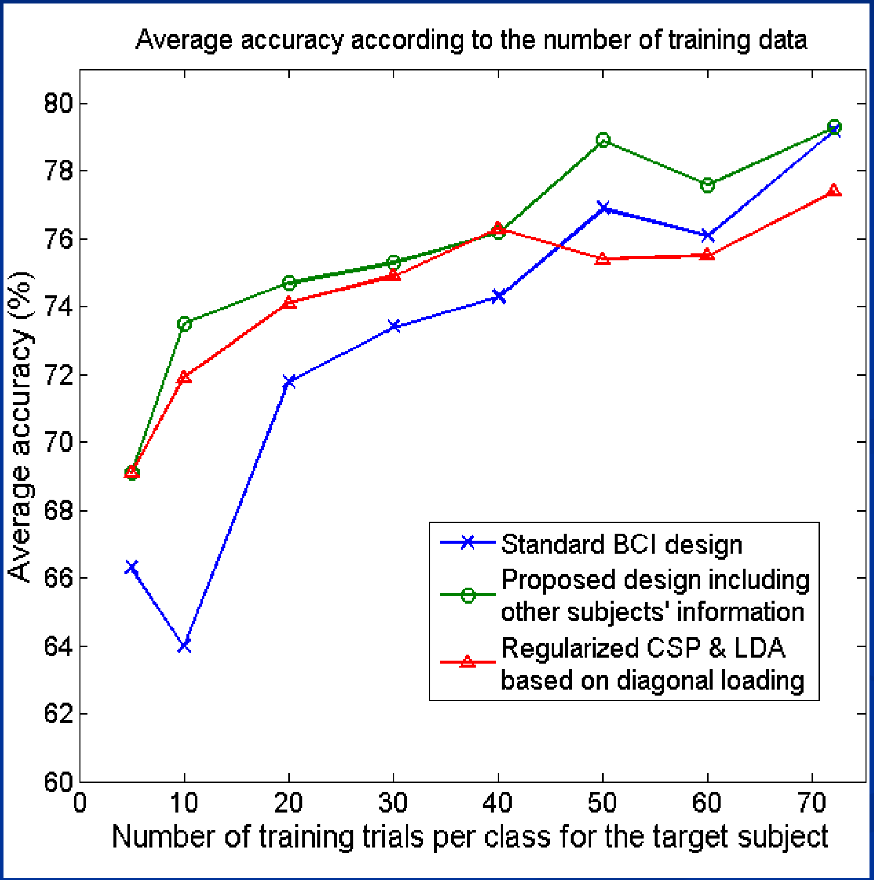
with $$G_i = \frac{1}{\left|S_t(\Omega)\right|} \sum_{k \in S_t(\Omega)} C^k$$

C^k: covariance matrix for the kth previous subject $S_t(\Omega)$: selected subset of previous subjects Ω

- Advantage: enable learning with less data

  => calibration time reduction!

Evaluations:
BCI competition IV, data set 2a



Lotte & Guan, ICASSP 2010

Sparse CSP for channel selection

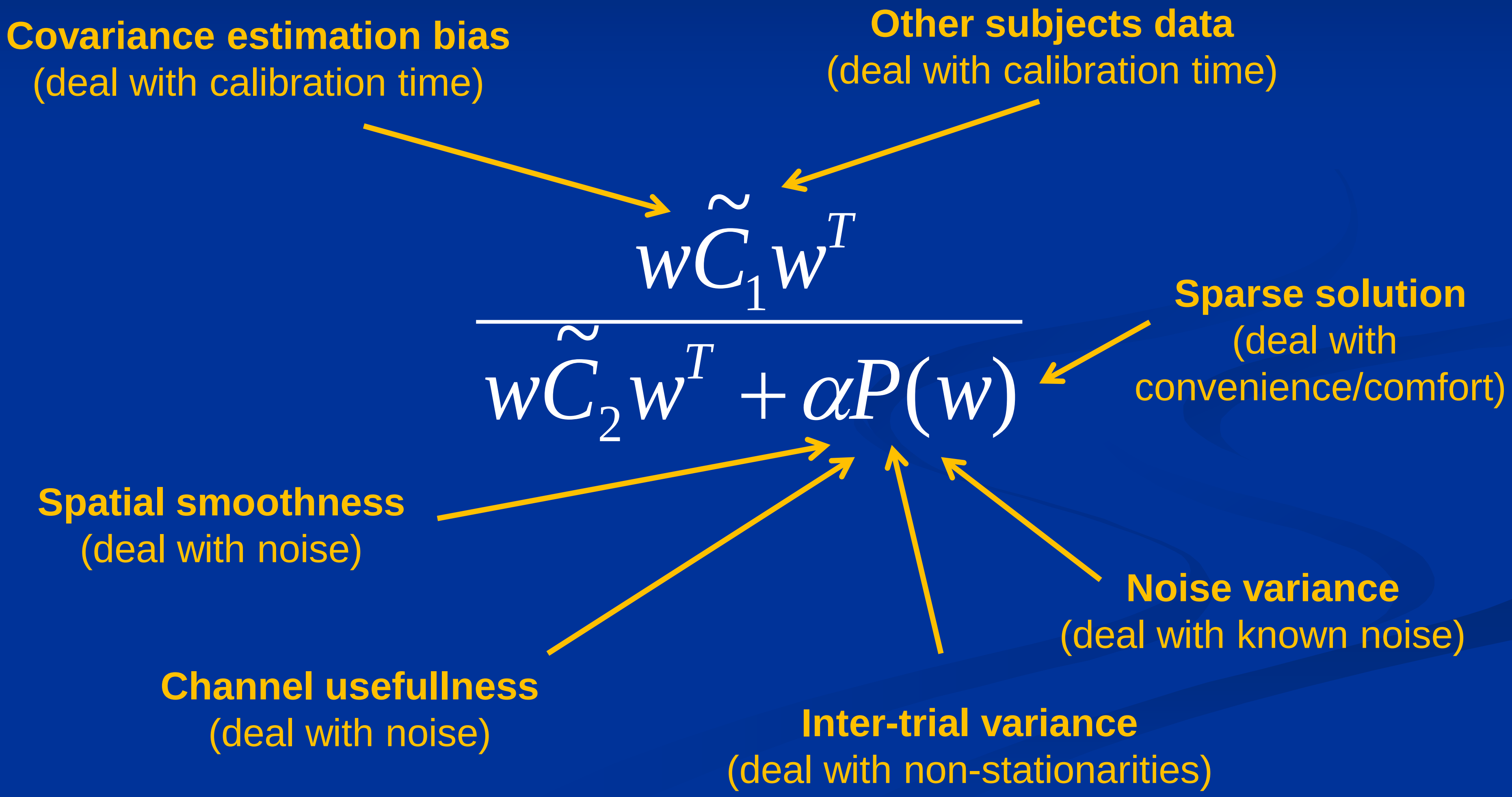
- Using less EEG channels is
  - More comfortable and less bulky
  - Faster to set-up

- Sparse CSP
  - Enforce the use of few channels
  - Also deal with noise

$$P(w) = \sum |w_i| = \|w\|_1$$

[Arvaneh et al, *IEEE TBME*, 2011][Farquhar et al, BCI workshop, 2006]

Using a-priori knowledge for CSP



$$\frac{w\tilde{C}_1 w^T}{w\tilde{C}_2 w^T + \alpha P(w)}$$

Covariance estimation bias
(deal with calibration time)

Other subjects data
(deal with calibration time)

Sparse solution
(deal with
convenience/comfort)

Spatial smoothness
(deal with noise)

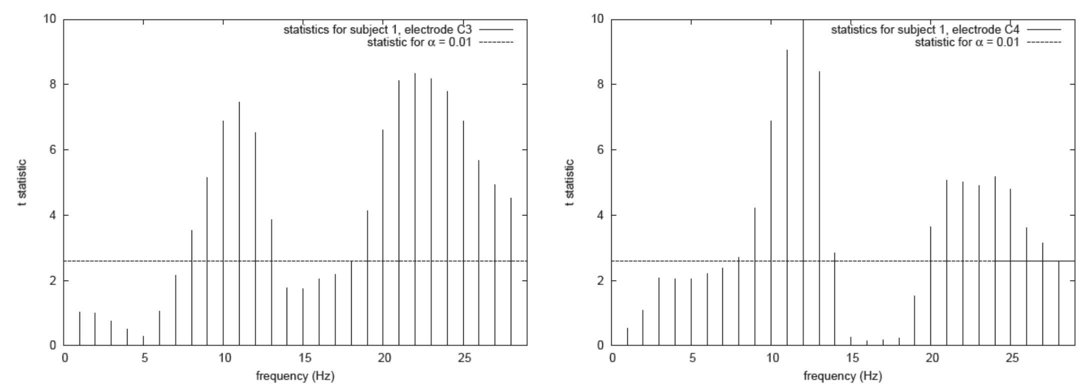
Noise variance
(deal with known noise)

Channel usefullness
(deal with noise)

Inter-trial variance
(deal with non-stationarities)

Using Subject-specific frequency bands

- How to find the optimal bands?
  - Manually (trial-and-errors)…
  - Looking at the average spectrum in each class
  - Computing statistics on the spectrum

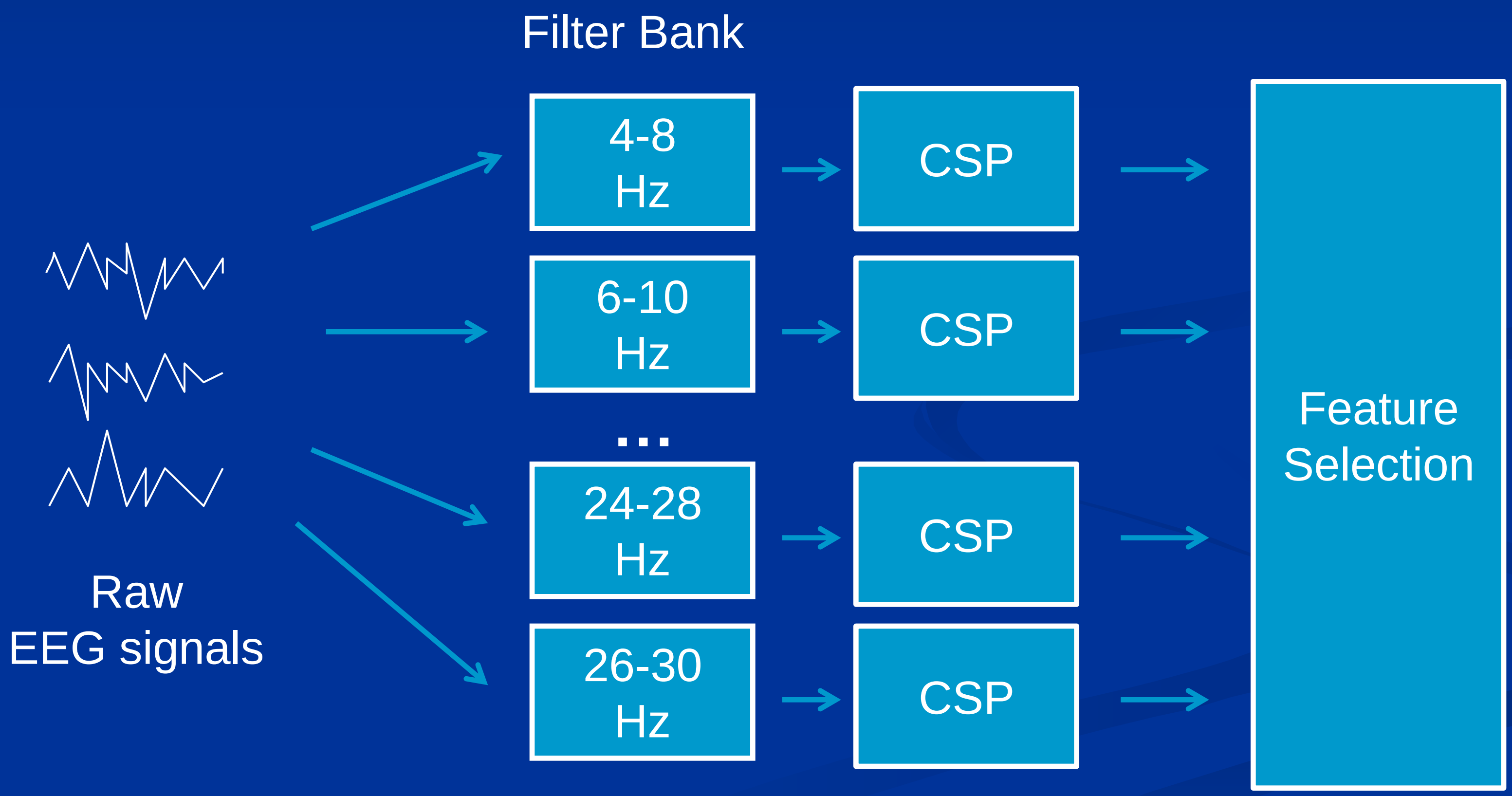


From Zhong et al, Patt. Rec. Let., 2008

  - Learning the optimal band-pass filter [Devlaminck11]

How to find the right frequency band with CSP?

- Ex: The Filter Bank CSP (FBCSP)



Ang et al, « Filter Bank Common Spatial Patterns in Brain-Computer Interfaces », IJCNN, 2008

FBCSP Results

Method	Classification accuracy (%)	
	Data Set I (5 subjects)	Data Set II (11 subjects)
CSP	86.6	73.3
FBCSP	**90.3**	**81.1**

Efficiency of FBCSP (from Ang et al, IJCNN, 2008)

Winning algorithm of BCI competition 2008 on all EEG data sets
Ang et al, Pattern Recognition, 2011

Other Spatio-Spectral Filters

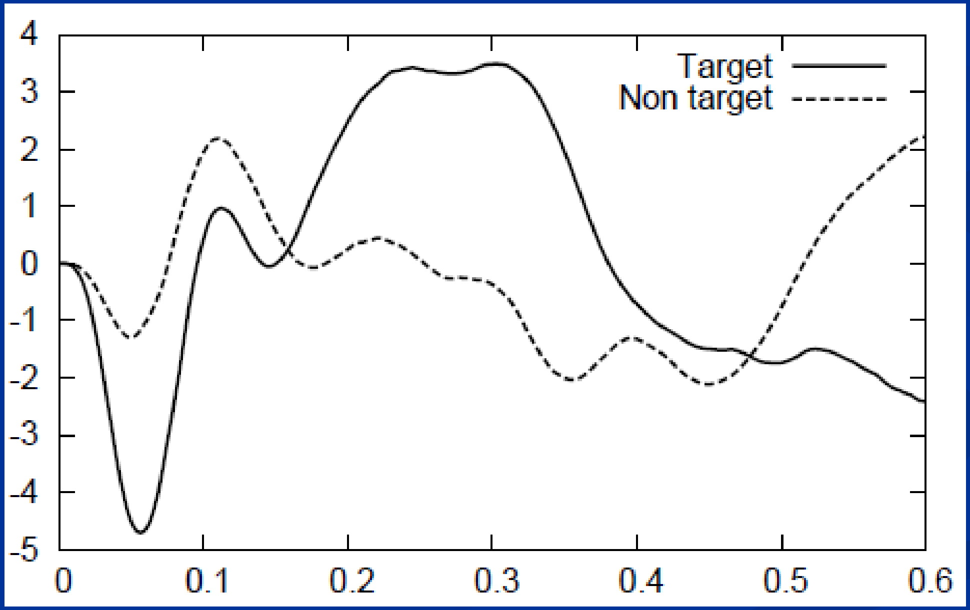
- Many CSP variants to do so
  - Common Spatio-Spectral Patterns (CSSP) [Lemm et al, TBME, 2005]
  - Common Sparse Spectral Spatial Pattern (CSSSP) [Dornhege et al, TBME 2006]
  - Spectrally Weighted CSP [Tomikia et al, 2006]
  - Discriminative CSP [Thomas et al TBME, 2009]
  - …
- All more efficient than basic CSP

Summary of features for ERD/ERS-based BCI

- Importance of Spectral information
  - Using band power features in relevant frequency bands is an efficient approach
- Importance of Spatial information
  - Using multiple channels is useful
  - Spatial filtering is a very efficient solution
    - Inverse Solutions
    - Common Spatial Patterns (CSP) is a must-try
    - Many CSP variants are available to deal with noise, non-stationarities, limited training data, etc.

Evoked potentials-based BCI

- Example: the P300
  - A brain signal occurring about 300ms after the user perceived a rare and relevant stimulus



EEG signal amplitude (μV)

Time (s)

Spatial Filters for Evoked Potentials: do we need them?

- classification of a signal window X

$$y = wF + b \qquad \text{with} \qquad F = WX$$

Classifier weights

Spatial filter weights

$$y = w(WX) + b = \hat{w}X + b$$

- Everything is linear!
  - The classifier can learn all the weights directly
  - But the learning problem is more difficult

Should we use spatial filters for Evoked Potentials?

- Example:

  Nc= 32 channels Nf = 3 filters Ns = 20 samples

$$w \in \Re^{1\times(Nf*Ns)} + W \in \Re^{Nf*Nc}$$

→ **156 parameters to learn**

$$\hat{w} \in \Re^{1\times(Nc*Ns)}$$

→ **640 parameters to learn**

- In practice
  - With enough data and/or a good classifier
    - No need for spatial filter with EP
    - It is “more” optimal to let the classifier do all the job
  - Otherwise it can be useful

Spatial Filtering for EP: Why not using CSP?

- A crucial information for classifying EP is the EEG time course

- CSP completely ignores this time course as it only considers the average power
  - ➔CSP may not be suitable for EP classification

- Other spatial filters have been specifically designed for EP classification

Fisher Spatial Filter (1)

- Fisher criterion for optimal class separability

$$\max J = \frac{tr(S_b)}{tr(S_w)}$$ with

$$S_b = \sum_k p_k (\overline{x_k} - \overline{x})(\overline{x_k} - \overline{x})^T$$ Between class-variance

$$S_w = \sum_k p_k \sum_{i \in C_k} (x_i - \overline{x_k})(x_i - \overline{x_k})^T$$ Within class-variance

x_i: feature vector, *p_k*: probability of class k

Fisher Spatial Filter (2)

- Fisher criterion to learn P300 spatial filters

$$\max J(w) = \frac{w\hat{S}_b w^T}{w\hat{S}_w w^T} \quad \text{with} \quad x_i = wX_i$$

- As CSP, solved by Generalized EVD
- Informal summary
  - Finds the spatial filter such that the spatially filtered EEG time course (i.e., the feature vector) is maximally different between classes

Hoffmann et al, ESANN, 2006

xDAWN spatial filter

- Maximize the signal-to-signal+noise ratio
  - ➔ Enhance the P300 response

$$\max J = \frac{w\overbrace{ADD^T A^T}^{\text{Variance of the time course of P300 responses (= signal)}}w^T}{w\underbrace{XX^T}_{\text{Variance of raw EEG (= signal+noise)}}w^T}$$

A: time course of a single
P300 response
(estimated from training data)

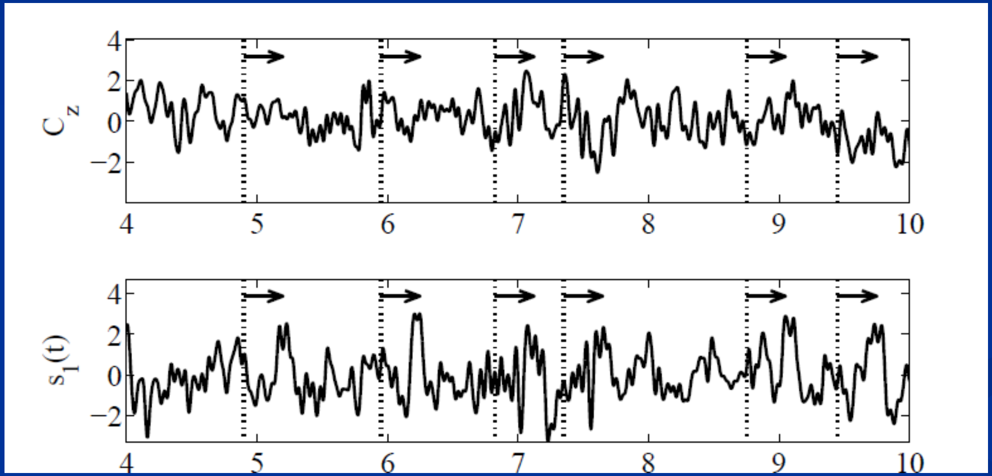
D: positions of target stimuli
(that should evoke P300)

- Again, solved by Generalized EVD

Rivet et al, IEEE TBME, 2009

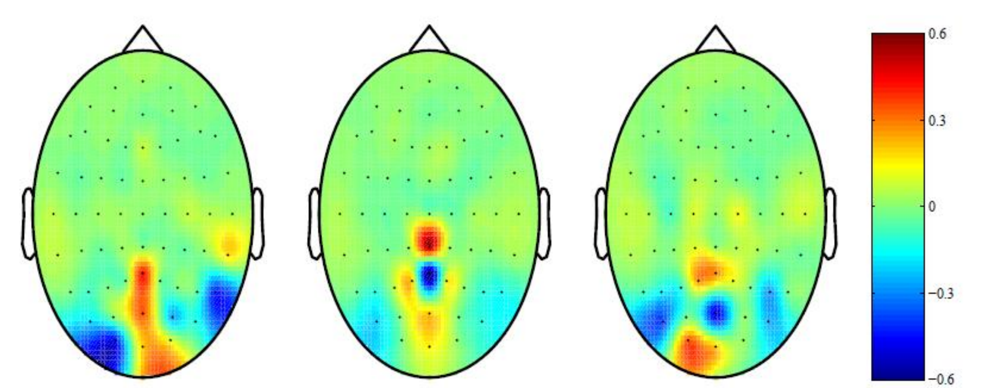
xDAWN spatial filter (2)

- Outcome of xDAWN spatial filtering



Raw EEG (Cz)

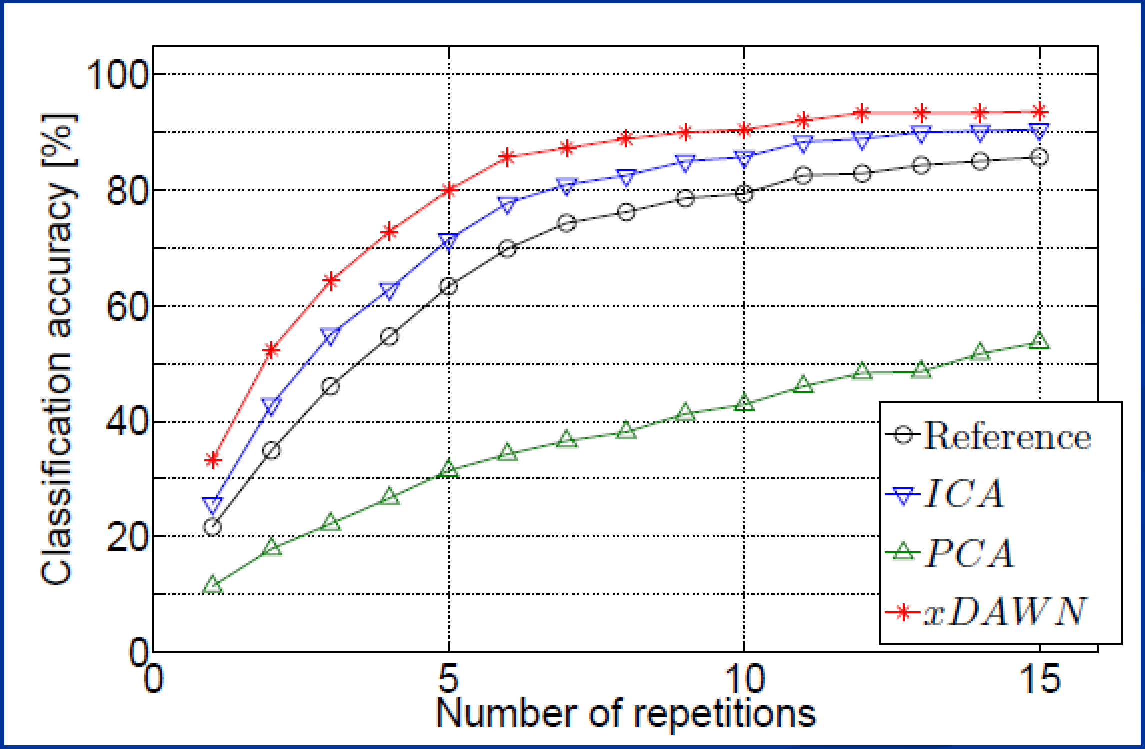
Spatially filtered EEG



Examples of 3 spatial filters
Obtained with xDAWN

From Rivet et al, GRETSI, 2007

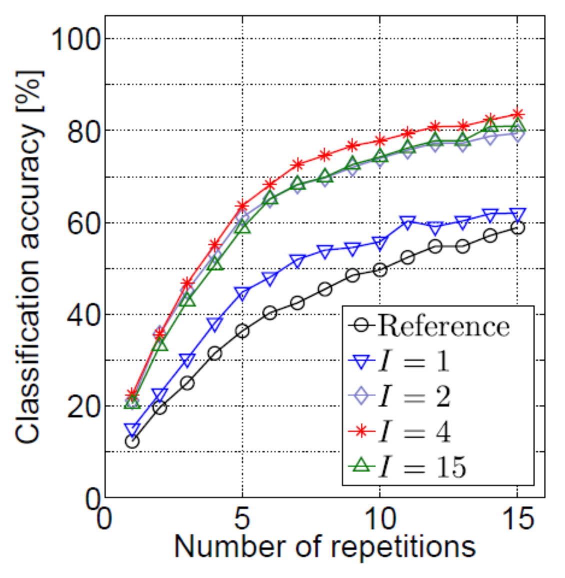
xDAWN vs other spatial filters



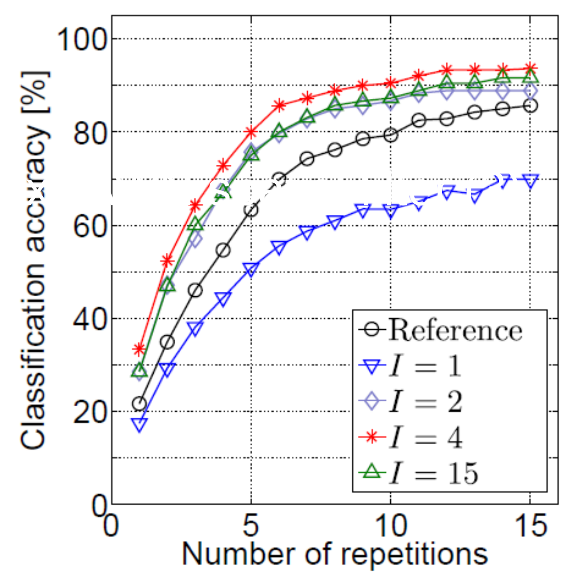
From [Rivet et al, IEEE TBME, 2009]

Impact of the training data size

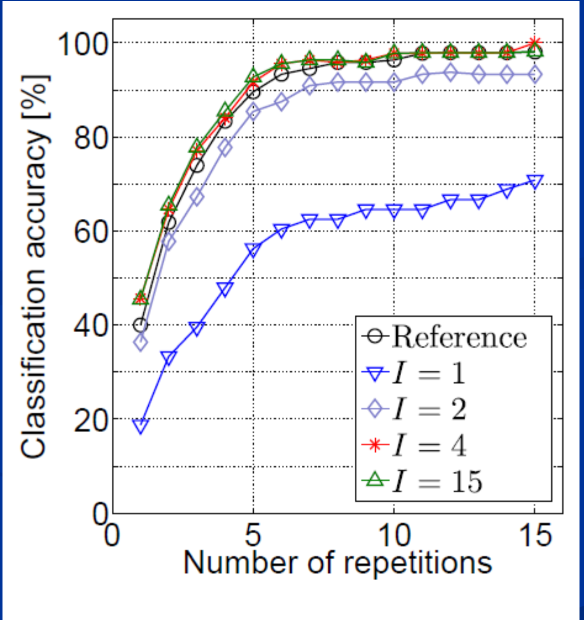
- xDWAN Performances with a BLDA



(a) $N_c = 2$



(b) $N_c = 5$



(c) $N_c = 20$

From [Rivet et al, IEEE TBME, 2009]

Alternative features

- ***Temporal*** representations [Vidaurre09]
  - Ex: Hjorth or Time Domain Parameters
- ***Connectivity*** measures [Grosse-Wentrup09]
  - Ex: Coherence, phase locking value, causality
- ***Chaos theory***-inspired measures [Brodu12]
  - Ex: Fractal dimension, multifractal cumulants
- ***Complexity*** measures [Balli11, Brodu12]
  - Ex: Entropy, predictive complexity
- …

Using multiple features

- Provides complementary information
  - Likely to increase performances
    [Dornhege et al, IEEE TBME, 2004]
  - Features that are weak alone may be efficient together
    [Brodu, Lotte, Lécuyer, Neurocomputing, 2012]
- May provide a kind of redundancy
  - Provide robustness to different kinds of noise
    [Fatourechi, JNE, 2008]

Spatial filters for alternative features?

- Band-power (MI) and Time-embedding (P300) are not the only valuable features
  - But they are basically the only ones with dedicated spatial filters
    - CSP for BP
    - xDAWN for Time-embedding
    - Using other features with such filters is suboptimal
- Could spatial filters for alternative features be useful?

Example: Time Domain Parameters (TDP)

- A measure of temporal variations

$$TDP^{(k)} = \left\|\Delta x^{(k)}\right\| = \sum_i \left\|x_{i+1}^{(k)} - x_i^{(k)}\right\|$$

$$\Delta x_{i+1}^{(k)} = \Delta x_{i+1}^{(k-1)} - \Delta x_i^{(k-1)} \qquad \Delta x_{i+1}^{(1)} = x_{i+1} - x_i$$

- Ex: with k=1 and the l_1-norm
  - TDP=waveform length
- TDP have been proved valuable alternative to band-power [Vidaurre09, Bruner11]

Spatial filters for TDP

- Objective function

$$J_{TDP^{(k)}} = \frac{\left\| w\Delta_{X_1}^{(k)} \right\|}{\left\| w\Delta_{X_2}^{(k)} \right\|} = \frac{w\Delta_{X_1}^{(k)^T} \Delta_{X_1}^{(k)} w^T}{w\Delta_{X_2}^{(k)^T} \Delta_{X_2}^{(k)} w^T}$$

- Like CSP, this is solved by GEVD

Evaluation

- Evaluation for k=1 (waveform length) on N=15 subjects

Features	Classification accuracy
Band Power	68 %
TDP	66.7 %
CSP	77 %
TDP-Spatial Filter	78.7 %
TDP-SF + CSP	80.1 %

Lotte, ICPR 2012

Alternative SF are valuable!

- Spatial Filters for « regression »
  - SPoC – Source Power Correlation Analysis [Dähne et al, BBCI workshop 2012]
  - Spatial Filter for Granger Causality [Winkler et al, BBCI workshop 2012]
- Many optimal spatial filters are yet to be designed
  - Complexity, connectivity, etc.

Do we really need spatial filters?

- For ERD/ERS

$$y = wF + b = w\log(WC_XW^T) + b \approx \hat{W}vec(C_X) + b$$

Classifier weights

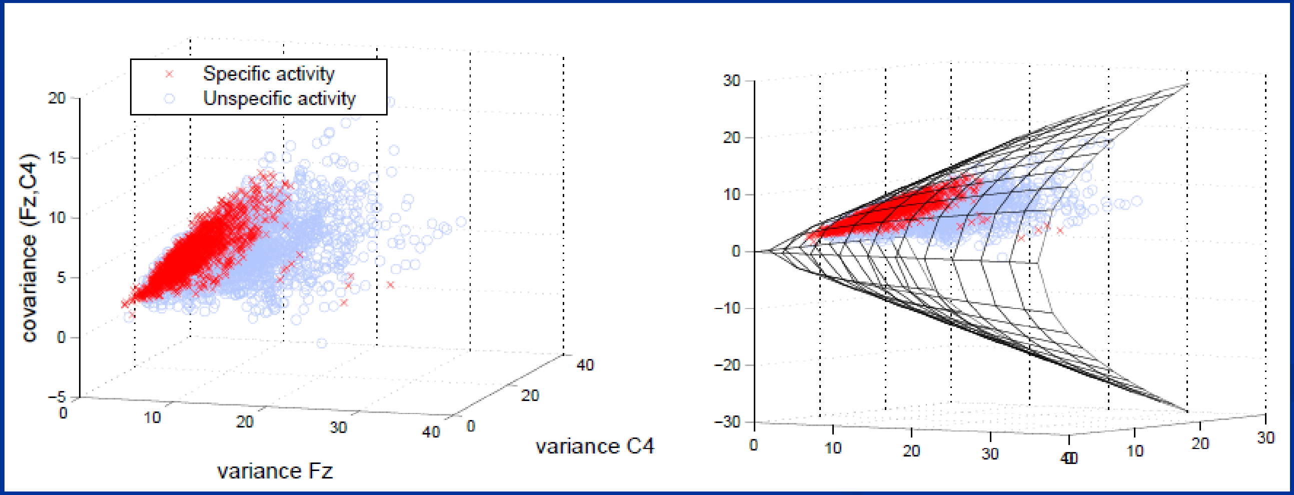
Spatial filter weights

- All the weights can be learned in the (vectorized) covariance matrix space
  - Enough data and a good regularized classifier is necessary

[Tomioka & Muller, NeuroImage, 2009] [Farquhar, IEEE TNN, 2009]

Classifying Covariance matrices

- The Riemannian distance



From [Barachant et al, BCI conference, 2010]

- Can be kernelized

$$y = \sum_i \alpha_i K(C, C_i) + b$$

[Barachant et al, IEEE TBME, 2012][Barachant et al, ESANN, 2012]

Open research challenges

- Finding features and filters that are
  - More Informative
    - ➔To reach better performances
  - Robust to noise & artifacts
    - ➔To use outside laboratories, with moving users
  - Invariant
    - ➔To deal with non-stationarity and session-to-session transfer
  - Universal
    - ➔To design subject-independent BCI

Conclusion

- **A form of spatial filtering is essential**
  - Regularized CSPs, xDAWN, covariance matrix space
  - the right filter must be used for the right problem
- **Using a-priori knowledge helps**
  - Neurophysiology, other subjects, noise, …

Thank you for your attention!

Any question?

Fabien Lotte
web: http://sites.google.com/site/fabienlotte/
http://openvibe.inria.fr
e-mail: fabien.lotte@inria.fr