# Performance Enhancement in Binary EEG Signal Classification using Sparse Representation

**Younghak Shin, Seungchan Lee, Soogil Woo, and Heung-No Lee***

**Presenter : Heung-No Lee, Ph.D.**

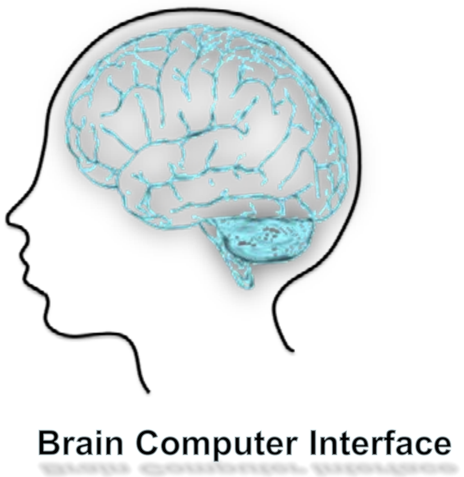
GIST, Dept. of Information and Communication, INFONET Lab.

http://infonet.gist.ac.kr

Gwangju Institute of Science and Technology

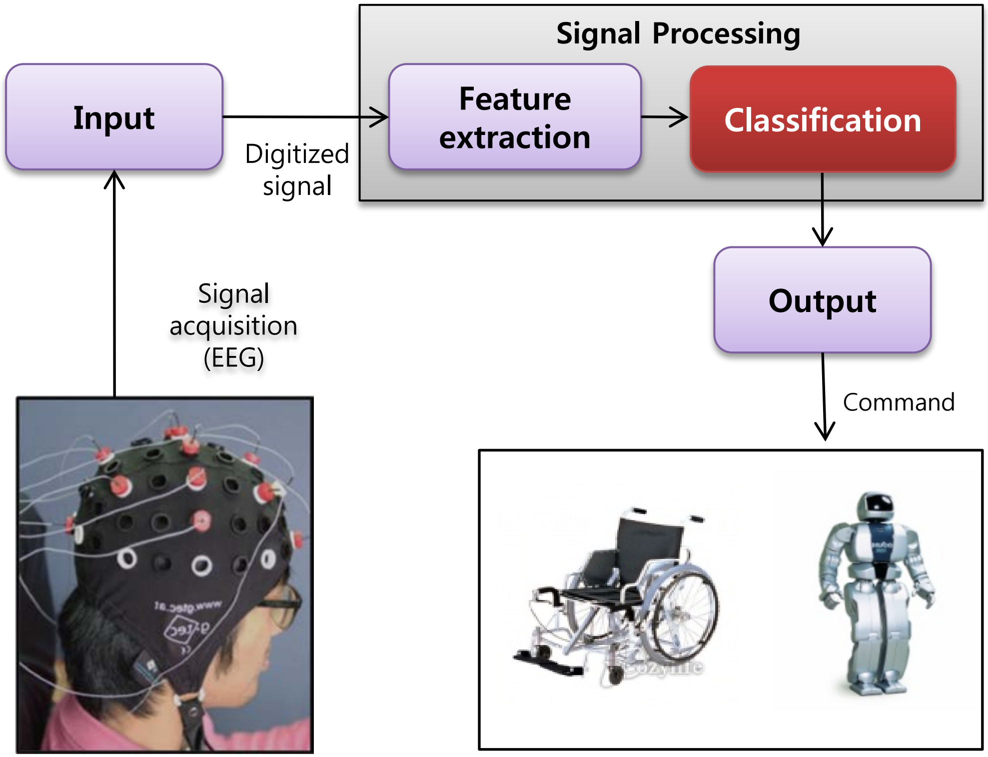
# Outline of Presentation

- Introduction
- Methods
- Results and Discussions
- Conclusions
- Reference
- Appendix



# EEG based BCIs

- Brain Computer Interface system (BCIs)

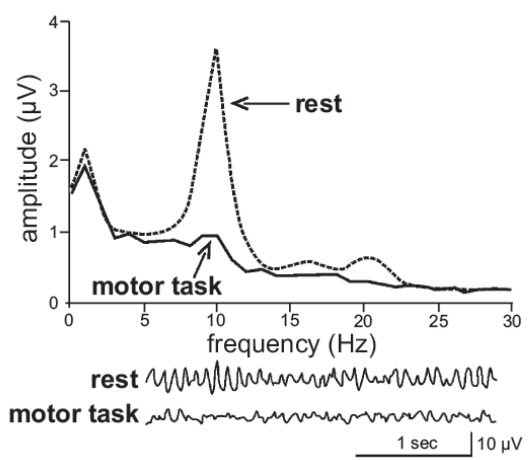



– In the BCIs, classification is needed to transform the extracted feature of a user's intention into a computer command to control the external device.

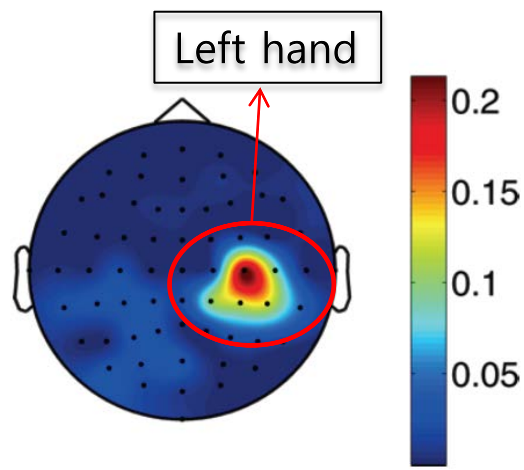
– EEG signals are very noisy and non-stationary. Therefore, powerful signal processing methods are needed.

# Motor imagery based BCIs

- In this study, we focus on the MI based BCI application
  - When subject imagine left or right hand movement, amplitude attenuation of mu rhythm appears at the contralateral area of cortex [Wolpaw1991] .
  - The mu(8-12Hz) and/or Beta(15~30Hz) rhythms originate above the sensory-motor cortex area.




[ Spectral focus ]

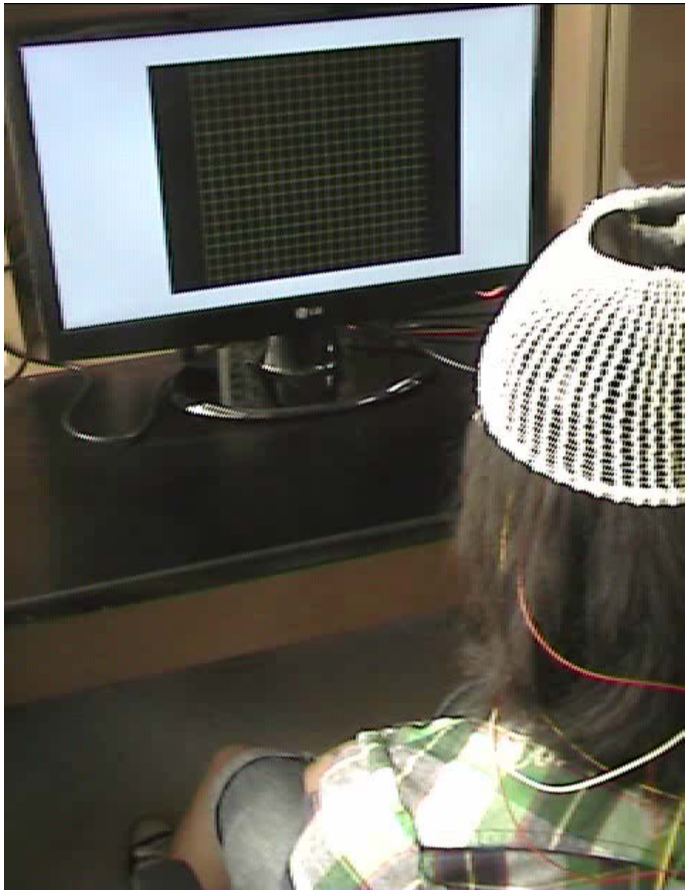



[ Position focus ]

# Motor imagery based BCIs

- INFONET BCI experiment
- 1-D cursor control experiment
- Performance evaluation:

| Run | Trial | Hit | Miss | Accuracy |
|---|---|---|---|---|
| 1 | 20 | 18 | 2 | 90% |
| 2 | 20 | 18 | 2 | 90% |
| 3 | 20 | 15 | 5 | 75% |
| 4 | 20 | 19 | 1 | 95% |
| 5 | 20 | 19 | 1 | 95% |



Experiment done on 2010/7/27

# Sparse Representation (SR)

- Sparse Representation has received a lot of attention in recent years.
- The problem of the SR is to find the most compact representation of a signal in terms of linear combination of atoms in an over-complete dictionary [Huang 2006].
- The problem of SR is to find the coefficient $\mathbf{x} \in \mathbb{R}^{N\times 1}$ :

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathbf{y} = \mathbf{A}\mathbf{x}$$

where, $\mathbf{A} \in \mathbb{R}^{M\times N}$ is known over-complete dictionary $(M \ll N)$

$\mathbf{y} \in \mathbb{R}^{M}$ is measured signal

- $\|\mathbf{x}\|_0$ denotes the L0 norm. Solving this minimization is NP hard.
- Recently developed Compressive Sensing theory [Donoho 2006] reveals that if solution is sparse enough, L1 norm solution is equivalent to the L0 norm solution.

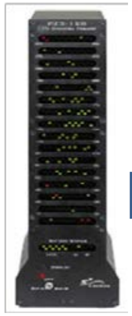
$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{y} = \mathbf{A}\mathbf{x}$$

# Motivation and Purpose

- For development of commercial BCIs, important issue is stable performance, *viz*. classification accuracy.
- Recently, Sparse Representation based Classification (SRC) method was studied in Face Recognition [Wright 09], and Speech Recognition area [Gemmeke 11].
- This SRC method has shown superior classification performance.
- In this study, we apply the SRC method to the motor imagery based Brain Computer Interface application.
- In addition, we compare the classification accuracy of SRC with that of conventional LDA and SVM classification methods.
- The LDA and SVM are most widely used classification methods.

# EEG data acquisition

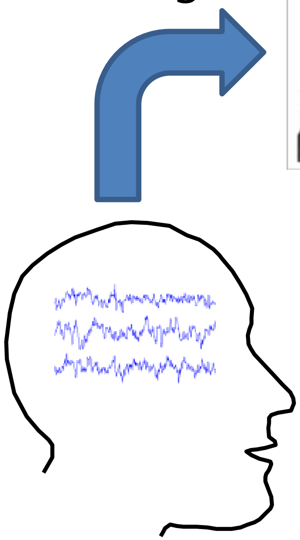
PZ3 Amplifier

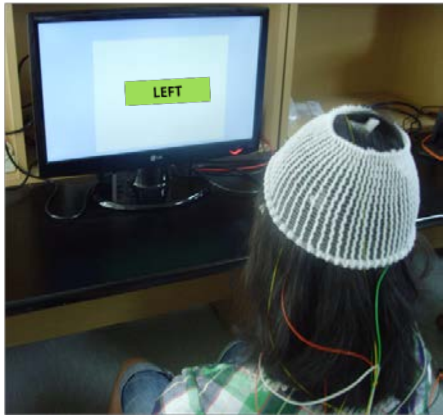


RZ5 Workstation

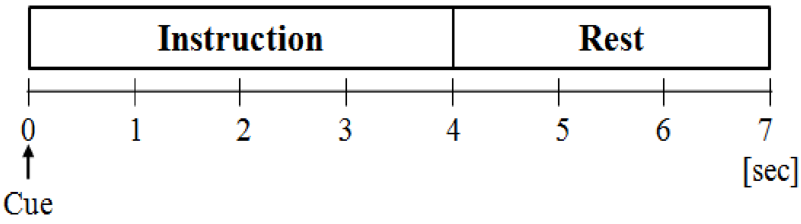
Desktop

EEG signal





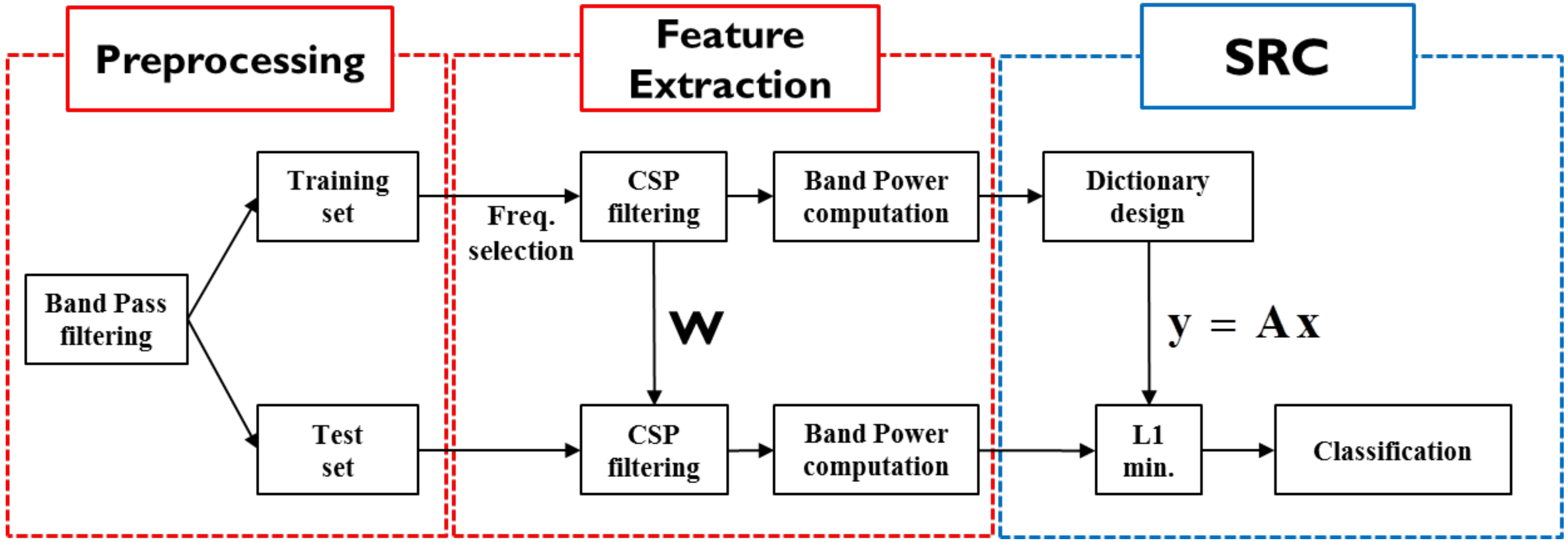
Experiment




- Collection of two class motor imagery data

: Left and Right hand imaginary movement

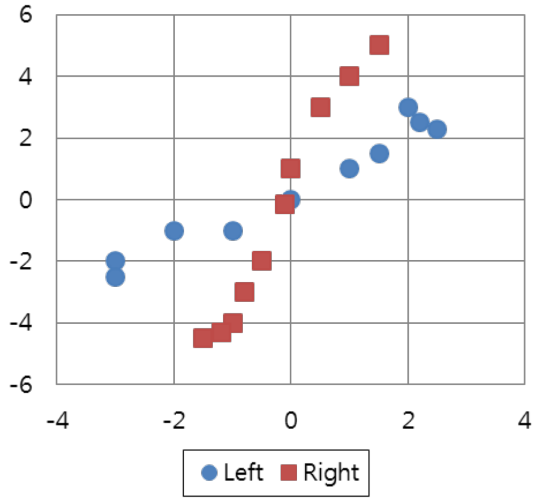
# Proposed SRC scheme



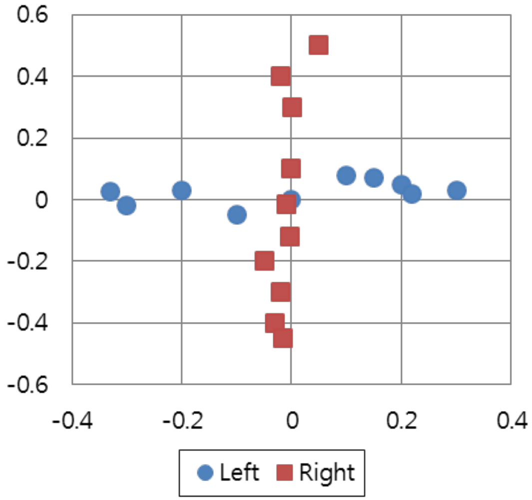

- For the SRC method, dictionary design is very important.
- We designed dictionary **A** using CSP filtering.
- To use a mu rhythm as a BCI feature, we compute the power of mu band.
- To find coefficient vector **x**, we use the L1 minimization tool for test signal **y**.

# CSP(Common Spatial Pattern) filtering

- CSP filtering is a powerful signal processing technique suitable for EEG-based BCIs [Blankertz 2008].
- CSP filters maximize the variance of the spatially filtered signal for one class while minimizing it for the other class.
- In our method, the CSP filtering was used to produce high *incoherence* between the two group of columns in the dictionary.
- Using the CSP filter, we form maximally uncorrelated feature vectors between the two classes

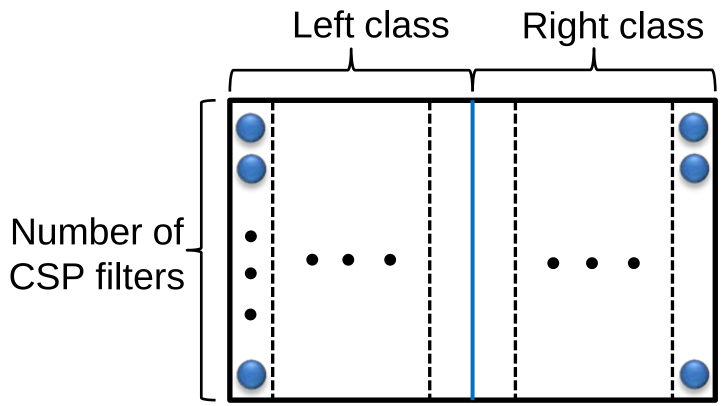



[ before CSP filtering ]




[ after CSP filtering ]

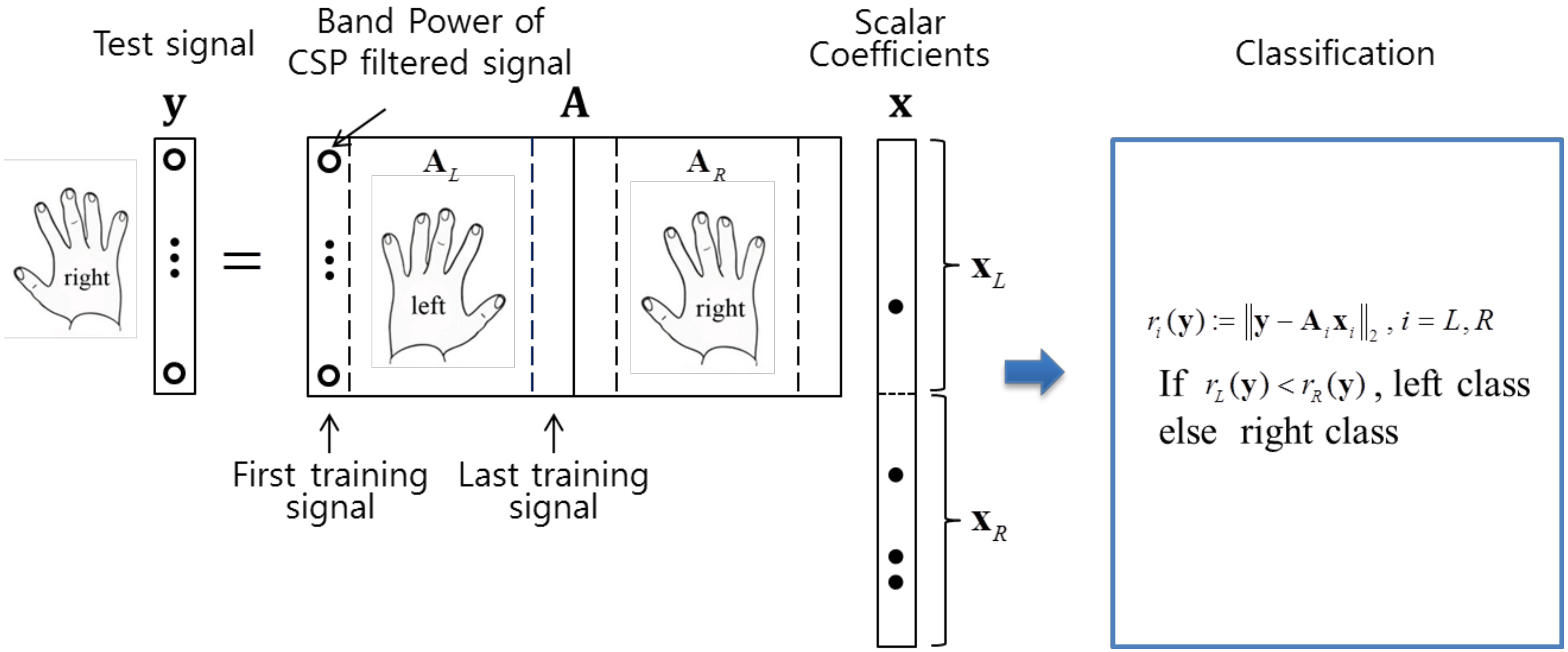
# Incoherent Dictionary




[Shin 2012]

- We use the CSP filtering to design a *incoherent* dictionary.
- When a dictionary is incoherent, a test signal from one particular class can be predominantly represented by the columns of the same class.
- Therefore, the incoherent dictionary promotes the sparse representation of the test signal under the L1 minimization.
- Sparsely represented a test signal helps in boosting the classification accuracy of the proposed method.

# Dictionary Design and Sparse Representation




- The sparse representation can be solved by L1 minimization [Candès 2006].
- For example, a test signal **y** of right class can be sparsely represented as the training signals of right class.
- However, EEG signals are very noisy, nonzero coefficients may appear in the indices corresponding to the left class.
- We use a minimum residual classification rule.

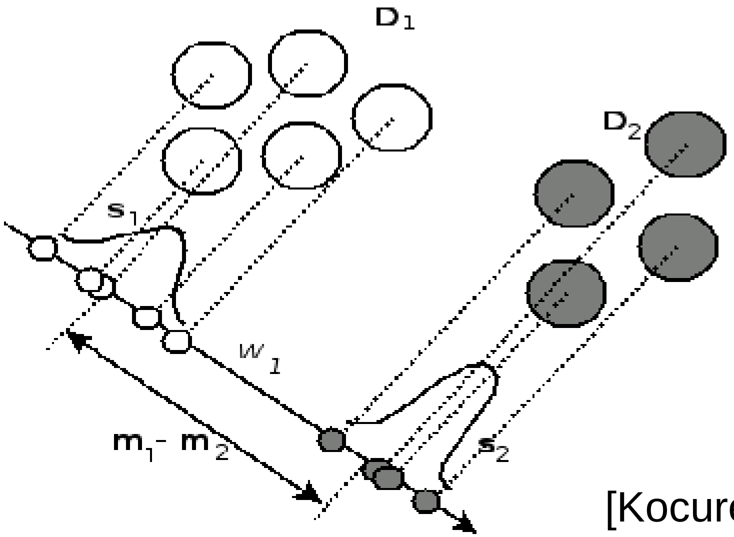
# Linear discriminant analysis (LDA)

- The LDA (also known as Fisher's LDA) approach aims to find the optimal direction, $\mathbf{w_1}$, to project data upon and maximize the Fisher ratio:

$$J(\mathbf{w}_1) = \frac{\mathbf{w}_1{}^T \mathbf{S_B} \mathbf{w}_1}{\mathbf{w}_1{}^T \mathbf{S_w} \mathbf{w}_1} \qquad \text{where, } \mathbf{S_B} = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^T$$

$$\mathbf{S_w} = \sum_i (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T$$

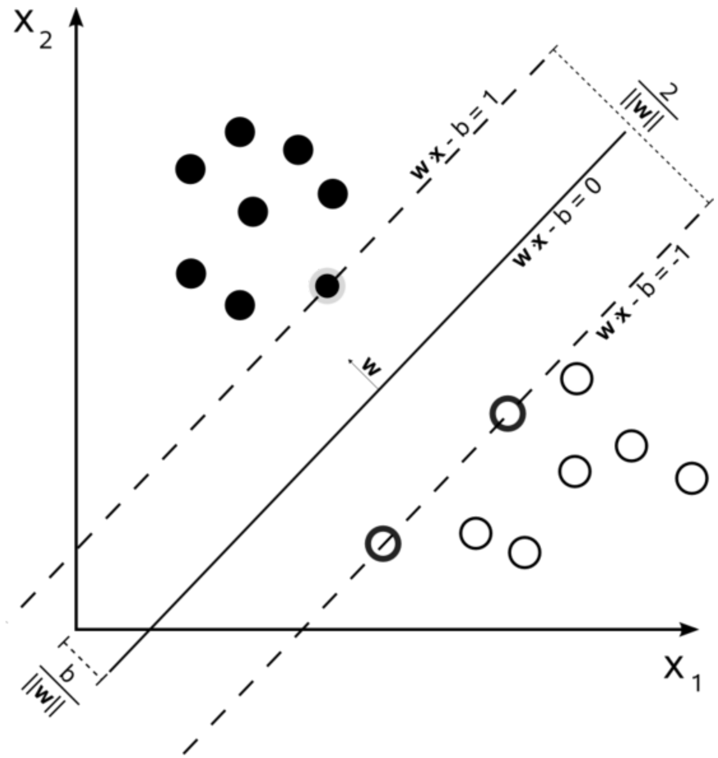
- The maximization of mean distances and minimization of class scatters.




[Kocurek 2007]

# Support vector machine (SVM)

- The idea of SVM is proposed by Vapnik aimed to find decision hyperplane with maximum margin which is the distance between the hyperplane and the nearest training feature vectors (support vectors).

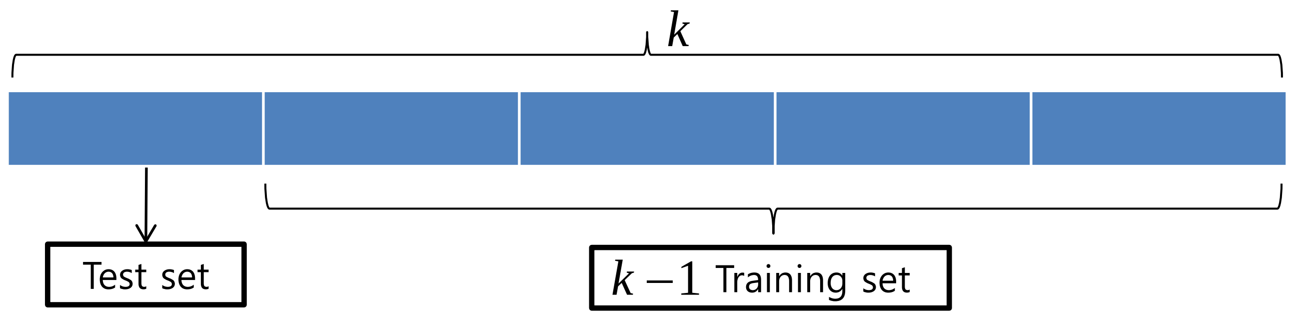



$$\text{minimize} \quad \frac{1}{2}\|\mathbf{w}\|^2,$$
$$\text{subject to} \quad y_n(\mathbf{w}^T\mathbf{x}_n + b) \geq 1$$
$$n = 1, \ldots, N$$

- In the BCI field, SVM has shown the robust classification performance in many experiments.

# Performance evaluation

- To evaluate the classification accuracy for each subject, we use the leave-one-out (LOO) cross-validation.
- LOO is useful for increasing the number of independent classification tests with a given limited data trials.
- Each time, one of the $k$ subsets is used as the test set and the other $k-1$ subsets are the training set.
- This method is repeated $k$ times with different subsets.

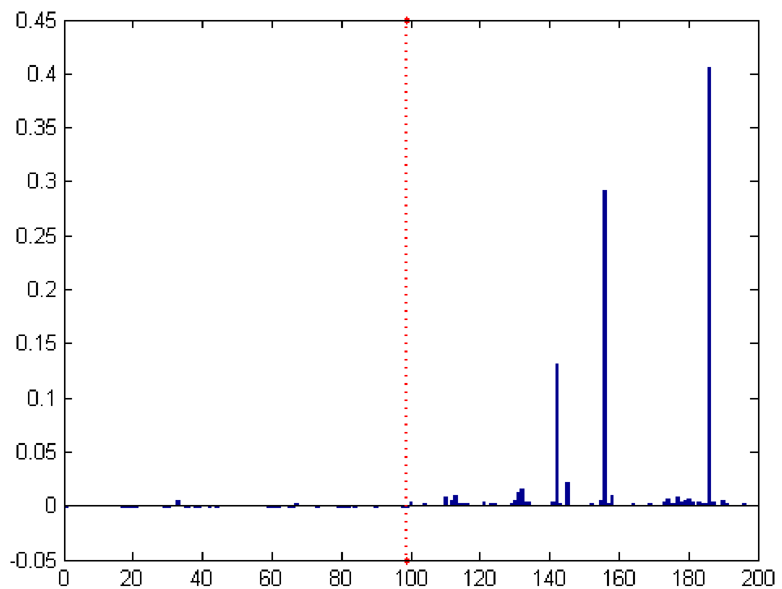



- The classification accuracy is calculated as :

$$\text{Accuracy}(\%) = \frac{\text{correct test trials}}{\text{total test trials}} \times 100$$

# Sparse representation results

- EEG Sparse representation
  - Sparse representation example of real EEG signals for one subject.
  - X-axis represents the number of total training trials (the number of columns of dictionary A).
  - Y-axis represents the recovered coefficients $\mathbf{x}$ in $\mathbf{y} = \mathbf{A}\mathbf{x}$.
  - The class of the test trial is the right hand imaginary
  - The test signal of right class sparsely represented with some training signals of right class



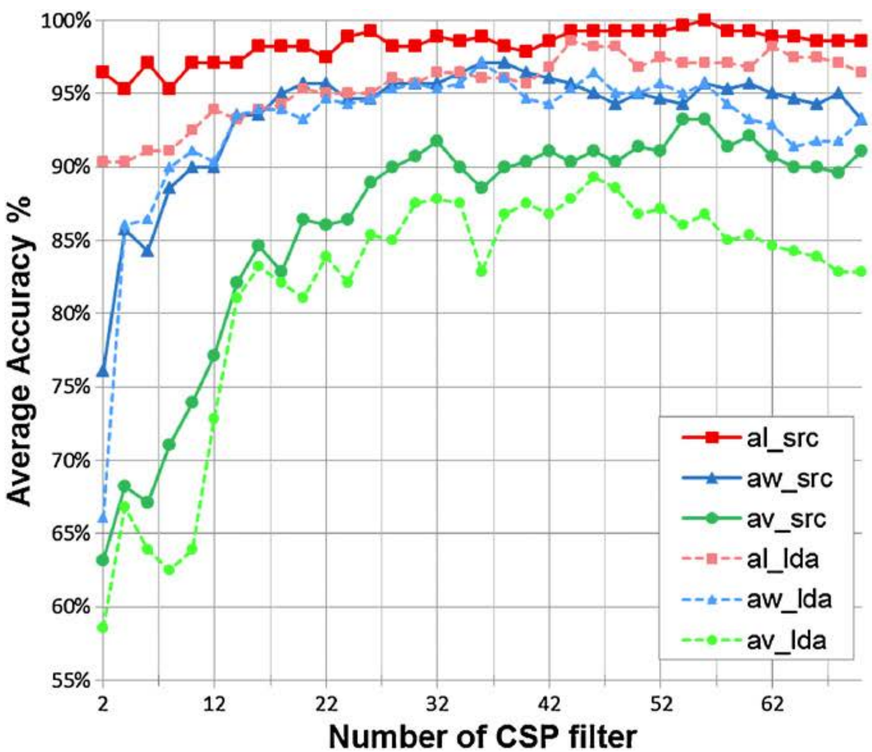
# Classification results

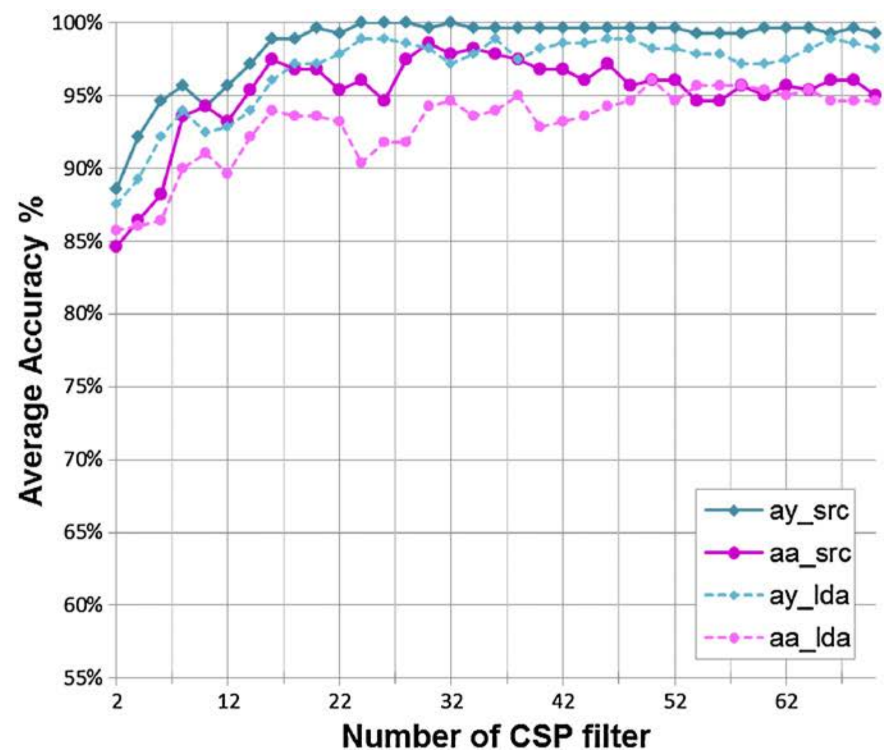
- BCI competition dataset (Data set IVa)
  - 5 subjects, 128 EEG channels
  - Right hand, and Right foot of motor imagery movements
  - 140 trial signals for each class
  - We use 16 CSP filters

| Subject | SRC Accuracy [%] | LDA Accuracy [%] |
|---|---|---|
| al | 98.93 | 96.43 |
| ay | 100 | 97.14 |
| aw | 95.71 | 95.36 |
| aa | 97.86 | 94.64 |
| av | 91.79 | 87.86 |
| Mean (SD) | 96.85 (3.25) | 94.29 (3.72) |

# Classification results

- BCI competition dataset (Data set IVa)
  - We examine classification accuracies of SRC and LDA as a function of the number of CSP filters (feature dimensions) for each subject.



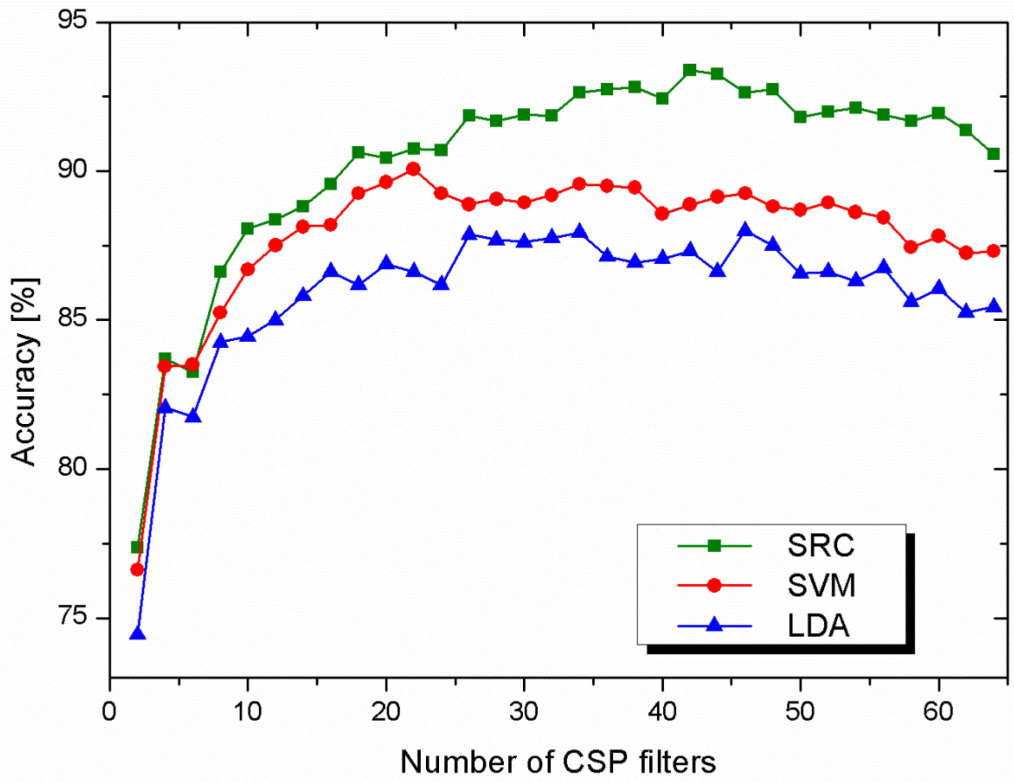
# Classification results

- Our own experimental dataset
  - 8 subjects, 64 EEG channels
  - Right hand, and Left hand of motor imagery movements
  - 100 trial signals for each class

| Subject | Average accuracy (%) | | | Max accuracy (%) | | |
|---|---|---|---|---|---|---|
| | LDA | SVM | SRC | LDA | SVM | SRC |
| **A** | 91.25 | 93.47 | 93.06 | 95.5 | 95.5 | 96 |
| **B** | 76.78 | 79.17 | 84.39 | 81 | 84.5 | 90 |
| **C** | 94.09 | 95.34 | 95.81 | 96.5 | 98 | 98 |
| **D** | 80.95 | 82.58 | 85.40 | 87.5 | 87 | 93.5 |
| **E** | 82.36 | 86.72 | 89.84 | 87.5 | 90.5 | 95.5 |
| **F** | 89.73 | 90.38 | 92.92 | 93 | 93.5 | 97.5 |
| **G** | 91.36 | 93.97 | 96.03 | 95 | 96.5 | 98 |
| **H** | 80.55 | 81.17 | 85.42 | 85 | 85.5 | 91 |
| **Mean** | 85.88 | 87.85 | **90.36** | 90.13 | 91.38 | **94.94** |
| **Std.** | 6.43 | 6.33 | 4.79 | 5.67 | 5.26 | 3.13 |
| **p-value** | 0.0007 | 0.0063 | | 0.0027 | 0.0053 | |

# Classification results

- Our own experimental dataset
  - We examine average classification accuracy for all subjects when the number of CSP filters is varied from 1 to 64.



# Conclusions

- We propose a sparse representation based classification (SRC) method for the motor imagery based BCI system.
- The SRC method needs a well-designed dictionary matrix made of a given set of training data.
- We use the CSP filtering to make dictionary uncorrelated for two different classes.
- We have compared with most widely used classification methods, LDA and SVM.
- The SRC method is shown to provide the best classification accuracy regardless of the number of CSP filters.

# Reference

[Wolpaw1991] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and C. A. Forneris, "An EEG-based brain-computer interface for cursor control," Electroencephalogr. Clin. Neurophysiol., vol. 78, pp. 252-259, 1991.

[Huang 2006] K. Huang and S. Aviyente, "Sparse Representation for Signal Classification," Neural Information Processing Systems, 2006.

[Donoho2006] D.Donoho, "Compressed sensing," IEEE Trans. Information Theory, vol. 52, pp. 1289-1306, 2006

[Wright2009] John Wright, Allen Y. Yang, Arvind Ganesh, S. Shankar Sastry, Yi Ma, "Robust Face Recognition via Sparse Representation" IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pp. 210~227, February 2009

[Blankertz2008] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, Klaus-Robert Müller. "Optimizing Spatial Filters for Robust EEG Single-Trial Analysis," IEEE Signal Proc. Magazine, 25(1):41-56, January 2008.

[Candès 2006] E. Candès, J. Romberg, and T. Tao, "Stable Signal Recovery from Incomplete and Inaccurate Measurements," Comm. Pure and Applied Math., vol. 59, no. 8, pp. 1207-1223, 2006

[Shin 2012] Younghak Shin, Seungchan Lee, Junho Lee and Heung-No Lee*, Sparse representation-based classification scheme for motor imagery-based brain-computer interface systems, "Journal of Neural Engineering, no. 9 (2012) 056002″.

[Donoho 2001] Donoho D L and Huo X 2001 Uncertainty principles and ideal atomic decomposition IEEE Trans. Inf. Theory 47 2845–62

# Thank you for attention!

# CSP(Common Spatial Pattern)

Find vectors w satisfying the following optimization problems (Second order statistics)

$$\max_{w}\left(\frac{w_i^T C_R w_i}{w_i^T C_F w_i}\right), \quad C_R = X_R X_R^T, C_F = X_F X_F^T$$

$\Rightarrow \min_{w}\left(-w_i^T C_R w_i\right)$ subject to $w_i^T C_F w_i = 1$ From Lagrangian method,

$$\Rightarrow L(w_i, \lambda_i) = -w_i^T C_R w_i + \lambda_i (w_i^T C_F w_i - 1)$$

$$\frac{d}{dw_i} L(w_i, \lambda_i) = -C_R w_i + \lambda_i C_F w_i = 0$$

$$\Rightarrow C_R w_i = \lambda_i C_F w_i$$

$$\therefore \max_{w}\left(\frac{w_i^T C_R w_i}{w_i^T C_F w_i}\right) = \max_{w}\left(\frac{w_i^T \lambda_i C_F w_i}{w_i^T C_F w_i}\right) = \max_{w} \lambda_i$$

$$C_R w_i = \lambda_i C_F w_i$$

$$\Rightarrow \left|C_R - \lambda_i C_F\right|_{\det} w_i = 0$$

$$\left|C_R - \lambda_i C_F\right|_{\det} = 0$$

$$\left|X_R X_R^T - \lambda_i X_F X_F^T\right|_{\det} = 0 \quad \text{find max } \lambda_i$$

$\Rightarrow w$ is an eigen vector corresponding to the largest eigen value

# LDA(Linear Discriminant Analysis)

Find the optimal direction **w** to project data upon so that between-class variance is maximized and within-class variance is minimized.

Let's define linear projection : $\mathbf{y} = \mathbf{w}^T \mathbf{x}$, where $\mathbf{w}$ is optimal projection vector, $\mathbf{x}$ is sample($\mathbf{x}_{R,}\mathbf{x}_F$)

Define sample average: $\mathbf{m}_R = \frac{1}{N_i}\sum_i \mathbf{x}_R$, $\mathbf{m}_F = \frac{1}{N_i}\sum_i \mathbf{x}_F$

And after projection average: $\bar{m}_R = \frac{1}{N_i}\sum_i \mathbf{y}_R = \frac{1}{N_i}\sum_i \mathbf{w}^T\mathbf{x}_R = \mathbf{w}^T\mathbf{m}_R$

$$\bar{m}_F = \frac{1}{N_i}\sum_i \mathbf{y}_F = \mathbf{w}^T\mathbf{m}_F$$

Then we can define Between class scatter: $\mathrm{B_S} = \left|\bar{m}_R - \bar{m}_F\right| = \left|\mathbf{w}^T\mathbf{m}_R - \mathbf{w}^T\mathbf{m}_F\right|$

Also, We can define Within class scatter: $\mathrm{W_S} = \bar{s}_R{}^2 + \bar{s}_F{}^2$

$$\bar{s}_R{}^2 = \sum_{y\in R}(y - \bar{m}_R)^2,\ \bar{s}_F{}^2 = \sum_{y\in F}(y - \bar{m}_F)^2$$

Then we can define objective function:

$$\max_{\mathbf{w}} J(\mathbf{w}) = \frac{\text{Between class scatter}}{\text{Within class scatter}} = \frac{\mathrm{B_S}^2}{\mathrm{W_S}} = \frac{\left|\bar{m}_R - \bar{m}_F\right|^2}{\bar{s}_R^2 + \bar{s}_F{}^2} = \frac{\mathbf{w}^T\mathbf{S_B}\mathbf{w}}{\mathbf{w}^T\mathbf{S_w}\mathbf{w}}$$

# Recovery algorithm

$$\mathbf{y} = \mathbf{A}\mathbf{x}$$

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathbf{y} = \mathbf{A}\mathbf{x}$$

- The ell-0 norm is equivalent to the number of nonzero components in the vector **x**. This involves combinatorial search ; $\binom{N}{K}$

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{x}\|_2 \quad \text{subject to} \quad \mathbf{y} = \mathbf{A}\mathbf{x}$$

- The ell-2 norm solution is $\bar{\mathbf{x}} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{y}$.
- This obtained by Least-square method

$$\hat{\mathbf{x}} = \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{y} = \mathbf{A}\mathbf{x}$$

- If solution is sparse enough, ell-1 norm solution is equivalent to the ell-0 norm solution.
- This problem can be solved by standard linear programming in polynomial time.