# Analysis the effect of PCA for feature reduction in non-stationary EEG based motor imagery of BCI system

**Article** *in* Optik - International Journal for Light and Electron Optics · February 2014

DOI: 10.1016/j.ijleo.2013.09.013



**3 authors:**

Xinyang Yu
Chung-Ang University


Pharino Chum
Institute of Technology of Cambodia


Kwee-Bo Sim
Chung-Ang University

Optik 125 (2014) 1498–1502

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

# Analysis the effect of PCA for feature reduction in non-stationary EEG based motor imagery of BCI system

Xinyang Yu, Pharino Chum, Kwee-Bo Sim*

*School of Electrical and Electronics Engineering, Chung-Ang University, Seoul, Republic of Korea*

ARTICLE INFO

*Article history:*
Received 1 January 2013
Accepted 15 September 2013

*Keywords:*
EEG
BCI
Motor imagery
Feature extraction
Principal component analysis

ABSTRACT

Many brain computer–interface (BCI) systems depend on imagined movement to control external devices. But how to extract the imagination feature and classify them to control systems is an important problem. To simplify the complexity of the classification, the power band and a small number of electrodes have been used, but there is still a loss in classification accuracy in the state-of-art approaches. The critical problem is the machine learning art that when the signal into source has property of non-stationary causing the estimation of the population parameter to change over time. In this paper, we analyzed the performance of feature extraction method using several spatial filter such as common average reference (CAR), Laplacian (LAP), common spatial pattern analysis (CSP) and no-spatial filter techniques and feature reduction method using principle component analysis (PCA) based 90% rule variance and leave-one-out correct classification accuracy selection method; where support vector machine is the classifier. The simulation with non-stationary data set from BCI competition III-Iva shows that CAR best performance CSP method in non-stationary data and PCA with leave-one-out CCA could maintain CCA performance and reduced the trading off between training and testing 13.96% compared to not using PCA and 0.46% compared PCA with 90% variance.



## 1. Introduction

A brain–computer interface, or BCI, is a communication and control system that creates a non-muscular output channel for the brain [1]. The language of this communication is in part imposed on the brain that by the use of distinct brain signal features that the BCI system extracts and uses for device control and in part negotiated by the continuous and mutual adaptations of both the user and the system. The University of all over the world is committed to the development of the BCI system. Application of new sensors and EEG signals is to be used for exploiting motor imagery in order to produce more and more signal to control complicated BCI system. BCI systems have measured specific features of brain activity and translate them into command for the device control. For example, movement of arbitrary limbs changes the brain activity in the related cortex. In fact, already the preparation of movement or the imagination of movement also changed the so-called sensory rhythms. $\alpha$ rhythm activity is recorded from the sensorimotor areas, also called $\mu$ rhythm activity.

The EEG has numerous regular rhythm. The most famous are the occipital $\alpha$ rhythm, the central mu and beta rhythm. People can desynchronize the $\alpha$ rhythm and more increase activity by moror imangiry. This desynchronization reflects a decrease of oscillatory activity related to an internally or externally-paced event and is known as Event-Related Deynchronization (ERD). The opposite, namely the increase of rhythmic activity, was termed Event-Related Synchronization (ERS). ERD and ERS are characterized by fairly localized topographic and frequency specificity. The ERD/ERS patterns can be volitionally produced by motor imagery, which is the imagination of movement of limbs without actual movement [2]. In general, the EEGs are recorded over primary sensorimotor cortical areas often displays 8–10 Hz ($\mu$ rhythm) and 18–26 Hz ($\beta$ rhythm) activity. Some published paper had shown that people can learn to control the amplitude of $\mu/\beta$ rhythm in the absence of actual movement or sensation. Because $\mu/\beta$ rhythms are included in the ERD and ERS of the brain signals. It can be as part of spatial features for motor imagery

In order to improve the classification accuracy, in this paper, spatial filter and principal component analysis method have been proposed for the feature extraction and reduction, and SVM is introduced for the classifier.

* Corresponding author at: School of Electrical and Electronics Engineering, Chung-Ang University, Seoul, Republic of Korea. Tel.: +82 02 820 5319; fax: +82 02 817 0553.
*E-mail addresses:* yuxinyang@cau.ac.kr (X. Yu), Pharino@cau.ac.kr (P. Chum), kbsim@cau.ac.kr (K.-B. Sim).

0030-4026/$ – see front matter © 2013 Elsevier GmbH. All rights reserved.
http://dx.doi.org/10.1016/j.ijleo.2013.09.013

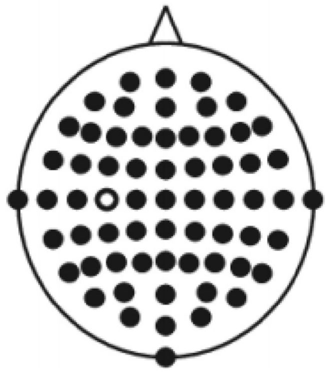

**Fig. 1.** Common average reference (CAR) filter sketch map.

## 2. Related works

### 2.1. Spatial filter

The typical motor imagery BCI system is separated by three major parts: preprocessing, feature extraction and classification of motor imagery patterns of EEGs. In this paper, our proposed method mainly for feature extraction steps. Before feature extraction step, we need a spatial filtering acted on EEG signals. The purpose of the spatial filter is to reduce the effect of spatial blurring from the raw signal. The spatial blurring occurs as effect of the distance between the sensor and the signal sources in the brain, because of the inhomogeneity of the tissues between the brain areas. I presented several kinds of spatial filter as following. In this paper, we used them to deal with the EEG signals.

*2.1.1. Common average reference (CAR)*

CAR a spatial filer could be considered as the subtraction of the common activity of EEG, which left only the idle activity of each individual EEG in specific electrode (Fig. 1).

The potential of each electrode after the filter could be computed as in Eq. (1), where $x_i^{CAR}(t)$ is the filtered output of electrode $i$th, $x_j(t)$ is the potential between $j$th electrode and the reference, $C$ is total number of all electrode on the scalp.

$$x_i^{CAR}(t) = x_i(t) - \frac{1}{C}\sum_{j=1}^{C} x_j(t) \tag{1}$$

*2.1.2. Laplacian spatial filter*

The Laplacian is a 2D isotropic measure of the 2nd spatial derivative of an image [3]. The Laplacian is often applied to an image that has first been smoothed with something approximating a Gaussian smoothing filter in order to reduce its sensitivity to noise, also has a good effect on processing the signal. Gaussian distribution on the scalp surface, and attempts to invert the process that blurred the brain activities detected on the scalp.

The approximations are further simplified as in Eq. (2), where $x_i^{LAP}(t)$ is filtered signal of electrode $i$, $x_i(t)$ is the potential of the electrode $i$ compare to the reference electrode. $\omega_{ij}$ is the constant weight compute using Eq. (3), where $d_{ij}$ is the Euclidian distance from electrode $i$ to electrode $j$. $S_i$ is the set of neighborhood electrodes of center electrode $i$. The montage of neighborhood electrode could be the small size and the large size as the depicted in Figs. 2 and 3. Size of neighborhood electrode usually is 4 choosing in vertical and horizontal, where $x_i$ is the center electrode.

$$x_i^{LAP}(t) = x_i(t) - \sum_{j \in S_i} \omega_{ij} x_j(t) \tag{2}$$

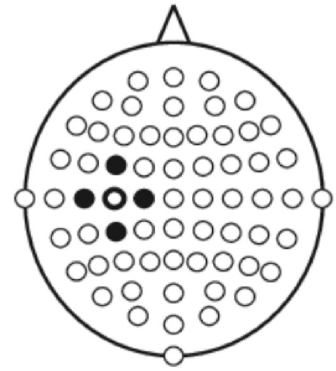

**Fig. 2.** Small Laplacian.

$$\omega_{ij} = \frac{1/d_{ij}}{\sum_{j \in S_i} 1/d_{ij}} \tag{3}$$

*2.1.3. Common spatial pattern*

The technique CSP algorithm is following by projecting of EEG segment from several electrodes to the orthogonal projection by mean of discriminant between two classes (events) [4]. Let $\Sigma^{(+)} \in R^{C\times C}$ and $\Sigma^{(-)} \in R^{C\times C}$ be the estimates of the covariance matrices of the band-pass filtered EEG signal in two conditions (e.g. right hand imagination and foot imagination movement).

$$\Sigma^{(c)} = \frac{1}{\left|I_c\right|}\sum_{i \in I_c} x_i x_i^T, (c \in [+, -]) \tag{4}$$

where $I_c(c \in \{+, -\})$ the set of indices corresponding to trials is belongs to condition c and $\left|I\right|$ denotes the size of a set $I$ and $\boldsymbol{x} \in R^{C\times T}$ is a short segment of EEG signal corresponding to a trial of imagination movement; $C$ is the number of channels and $T$ is the number of sample time points in a trial. $X$ is already center and scale. The CSP analysis is given by the simultaneous diagonalization of the two covariance matrices.

$$\boldsymbol{w}^T \Sigma^{(c)} \boldsymbol{w} = \Lambda^{(c)}, \quad (\Lambda^{(c)\text{diagonal}}) \tag{5}$$

A discriminant equation can be constructed as in Eq. (6); the goal is to maximize the energy of another class while minimize energy of another class.

$$J(\boldsymbol{w}) = \frac{\boldsymbol{w}^T \Sigma^{(+)} \boldsymbol{w}}{\boldsymbol{w}^T \Sigma^{(-)} \boldsymbol{w}} \tag{6}$$

This optimization problem can be solved by the first observing that function $J(\boldsymbol{w})$ remains unchanged whether he filter $w$ is

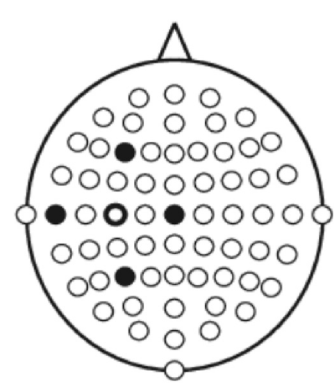

**Fig. 3.** Large Laplacian.

rescale. Indeed $J(k\boldsymbol{w})$=$J(\boldsymbol{w})$, with k a real constant, which means the rescaling by $k$ is arbitrary. As such, discriminant $J$(W) is equivalent to discriminant $\boldsymbol{w}^T\Sigma^{(+)}\boldsymbol{w}$ subjecting to the constraint $\boldsymbol{w}^T\Sigma^{(-)}\boldsymbol{w}=1$ as it is always possible to find a rescaling of $\boldsymbol{w}$ such that $\boldsymbol{w}^T\Sigma^{(-)}\boldsymbol{w}=1$. Using the Lagrange multiplier method, this constrained optimization problem accounts to maximizing Eq. (7), where $\lambda$ is Lagrange multiplier.

$$L(\lambda, \boldsymbol{w}) = \boldsymbol{w}^T\Sigma^{(+)}\boldsymbol{w} - \lambda(w^T\Sigma^{(-)}\boldsymbol{w} - 1) \quad (7)$$

Matrix $\boldsymbol{w}$ maximizes $L$ is such that the derivative of $L$ with respect to $\boldsymbol{w}$ equal to 0:

$$\frac{\partial L}{\partial \boldsymbol{w}} = 2\boldsymbol{w}^T\Sigma^{(+)} - 2\lambda\boldsymbol{w}^T\Sigma^{(-)} = 0 \quad (8)$$

$$\Sigma^{(+)}\boldsymbol{w} = \lambda\Sigma^{(-)}\boldsymbol{w} \quad (9)$$

The output signal from CSP could be found using Eq. (10).

$$y = \boldsymbol{w}^T x \quad (10)$$

### 2.2. Principle component analysis

Due to the instability of signal collection methods, we collected signal on the scalp have mixed with many interfering signal and have high dimensions [5,6]. Principal component analysis (PCA) is a mathematical procedure that uses an orthogonal transformationto convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. Suppose we have a random vector population $\boldsymbol{x}$, where $\boldsymbol{x}=(x_1,x_2, \ldots, x_n)^T$ and the mean of population is denoted by $\bar{\boldsymbol{x}} = E(\boldsymbol{x})$. The covariance matrix of the same data set is defined by $\boldsymbol{\Sigma} = E[(\boldsymbol{x} - \bar{\boldsymbol{x}})(\mathbf{x} - \bar{\boldsymbol{x}})^T]$. The components of $\Sigma$, denoted by $\Sigma_{ij}$, represent the covariance between the random variable components $x_i$, and $x_j$. The component $\sigma_{ii}$ is the variance of the component $x_i$. The covariance matrix is always symmetric. The variance of a component indicates the spread of the component values around its mean value. If two components $x_i$ and $x_j$ of the data are uncorrelated, their covariance is zero $\Sigma_{ij}=\Sigma_{ji}=0$. From a sample of vectors $\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots\boldsymbol{x}_N$; we can calculate the sample mean and the sample covariance matrix as the estimates of the mean and the covariance matrix. From a symmetric matrix such as the covariance matrix, we can calculate an orthogonal basis by finding its eigenvalues and eigenvectors. The eigenvectors $\boldsymbol{e}_i(i=1, \ldots, k \leq n)$ and the corresponding eigenvalues $\lambda_i$ are the solutions of Eq. (11).

$$\boldsymbol{\Sigma}_i\boldsymbol{e}_i = \lambda_i\boldsymbol{e}_i \quad (11)$$

If the data vector has $n$ components, the characteristic equation becomes of order $n$. This is easy to solve only if $n$ is small. Solving eigenvalues and corresponding eigenvectors is a non-trivial task, and various methods exist. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to the preceding components [6].

Suppose one has a data set of which the sample means and the covariance matrix has been calculated. Let $\boldsymbol{A}$ be a transition matrix which consists of eigenvectors of the covariance matrix as the row vectors. By transforming an EEG data set into a vector, we can get the output $\boldsymbol{y}$ of the process of dimensionality reduction.

$$\boldsymbol{y} = \boldsymbol{A}(\boldsymbol{x} - \bar{\boldsymbol{x}}) \quad (12)$$

The components of $\boldsymbol{y}$ can be seen as the coordinates in the orthogonal base. We can reconstruct the original data vector $\boldsymbol{x}$ from $\boldsymbol{y}$ using the property of an orthogonal matrix $\boldsymbol{A}^{-1}=\boldsymbol{A}^T$.

$$\boldsymbol{x} = \boldsymbol{A}^T\boldsymbol{y} + \bar{\boldsymbol{x}} \quad (13)$$

The original vector $\boldsymbol{x}$ was projected on the coordinate axes defined by the orthogonal basis. The original vector was then reconstructed by a linear combination of the orthogonal basis vectors. Instead of using all the eigenvectors of the covariance matrix, we may represent the data in terms of only a few basis vectors of the orthogonal basis. If we denote the matrix having the $K$ first eigenvectors as rows by $\boldsymbol{A}_K^T$, we can create a similar transformation as seen Eq. (13).

$$\boldsymbol{y} = \boldsymbol{A}_K(\boldsymbol{x} - \bar{\boldsymbol{x}}) \quad (14)$$

$$\boldsymbol{x} = \boldsymbol{A}_K^T\boldsymbol{y} + \bar{\boldsymbol{x}} \quad (15)$$

This means that we project the original data vector on the coordinate axes having the dimension $K$ and transforming the vector by a linear combination of the basis vectors. This minimizes the mean-square error (MSE) between the data and this representation with the given number of eigenvectors. In this paper, the PCA is applied to the training set to find the transformation matrix $\boldsymbol{w}$ which are uses to calculate the final feature using Eq. (16), where the transposed matrix of $\boldsymbol{w}$ has dimension of $K \times N$.

$$\boldsymbol{y} = \boldsymbol{w}^T\boldsymbol{x} \quad (16)$$

## 3. Discriminant feature selection using PCA

We proposed method studying on the effect of feature reduction of the spatial filter using PCA algorithm with different approaches of spatial filter in time-varying EEG data. Method consists of filtering EEG signal with the temporal filter using finite impulse response (FIR) method with 1 s hamming window to create the pass band filter of 7–30 Hz frequency range for the best activity of motor imagery [2]. Then the less noisy EEG goes through the process spatial filter. In here, authors study the 4 cases: no spatial filtering, spatial filter using CAR method, spatial filter using small Laplacian method with 4 neighbor electrodes and the famous CSP algorithm to determine the effect of all spatial filters for EEG especially in case of non-stationary EEG data. Author selected for full spatial information (electrode number is reduce after spatial filtering process) and compute the logarithm of feature based on Eq. (17), where log(.) is logarithm operation and diag(.) is selecting the diagonal element of a matrix, where $\boldsymbol{x} \in R^{C\times T}$ is a short segment of EEG signal corresponding to a trial of imagination movement; $C$ is the number of channels and $T$ is the number of sample time points in a trial.

$$\boldsymbol{x}_{diag} = \log(diag(\boldsymbol{x}^T\boldsymbol{x})) \quad (17)$$

Then $\boldsymbol{x}_{diag}$ feature is used to compute PCA transformation matrix $\mathbf{w}_K$ and produce the $K \leq c$ dimension feature $\boldsymbol{y}_K$ using Eq. (16), $C$ is the total number of spatial information. Value of $K$ is to be determined using 2 different techniques for comparison their effectiveness. The first technique to selecting $K$ is based on the accumulation of the first largest eigenvalues exceed 90% of total sum of eigenvalues as describe in the PCA algorithm part, the first smallest $K$ that satisfy Eq. (18) is selected as the model for selecting number of principle component.

$$\frac{\sum_{i=1}^{K}\lambda_i}{\sum_{i=1}^{C}\lambda_i} \times 100\% \geq 90\% \quad (18)$$

The second technique for selecting $K$ for PCA algorithm is based on the correct classification accuracy rate of leave-one- out method. With $K$ dimension feature compute through transformation matrix

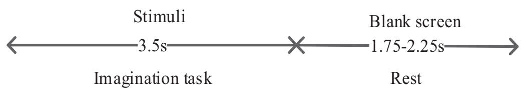


**Fig. 4.** Time scheme of imagination movement paradigm. The imagination movement started by a visual cue and proceeded of task for 3.5 s, then it followed by a blank screen for relaxation.

**Table 1**
Averaging results of correct classification accuracy (CCA) rate of all test data size (10%, 20%, . . ., 90%) in percentage of method with without feature reduction using PCA. S.F. is spatial filter type, where S.J. stands for subject name.

| S.J. | S.F. | | | |
|---|---|---|---|---|
| | No SF CCA (%) | CAR SF CCA (%) | LAP SF CCA (%) | CSP SF CCA (%) |
| aa | 60.36 | 72.32 | 59.47 | 57.58 |
| al | 87.39 | 87.94 | 91.53 | 77.39 |
| aw | 60.73 | 61.07 | 59.85 | 55.30 |
| av | 75.04 | 80.24 | 76.26 | 70.31 |
| ay | 78.26 | 80.13 | 68.30 | 63.21 |
| Mean | 72.35 | 76.34 | 71.08 | 64.76 |

$\boldsymbol{w}_K \in R^{K\times C}$, Leave-one-out with $N$ samples of training set use $N-1$ samples to train the classifier, which in this study we used support vector machine (SVM) with linear kernel, and compute the error of classifier based on 1 sample left. The performance of leave-one-out method is compute through the average error rate as in equation (19), where $f(y_i) = 1$ if SVM classified feature $y_i$ correctly, otherwise $f(y_i) = 0$. The value of $K$ is selected based on $\mathrm{argmax}_k CCA(K)$as the value for model $K$ for the testing data.

$$CCA(K) = 100\% \times \frac{1}{N}\sum_{i}^{N} f(y_i) \tag{19}$$

## 4. Experiments and results

### 4.1. EEG data

EEG signal was obtained from BCI competition III. Here we analysis for individual subject using dataset Iva of the competition data set. This data set was recorded from 5 healthy subjects using visual stimuli with total number of electrodes of 118 channels. Fig. 4 shows time schemes of experiment during recording EEG signal. Each trial conducted by visualization cue for 3.5 s while experimenting subject imagined the movement 1 of 3 motor imageries that subject should perform: left hand, right hand and foot. The target cues were intermitted by a period of random length, 1.75–2.25 s, in which subjects could relax [7,8]. BCI III data set Iva released for public use contains 2 classes (right hand and foot). Each subject has totally 280 trails with sampling frequency of 100 Hz. As the study goal is to test the algorithm in both effect of number of training data and multiple subjects analysis, data were divided as ratio of 10%, 20%, . . ., 90% for training sets [9]. The left over data was regarded as the test set. The results showed in this paper are the average of all different data set size.

### 4.2. Results and discussions

Tables 1–3 shows the averaging result of correct classification accuracy (CCA) rate of testing data of all size in percentage. The first column shows the subjects names in the BCI competition III data set Iva, The second, third,. . ., fifth column are the CCA of study corresponding to: No spatial filter, common average reference, Laplacian and common spatial pattern respectively. Each row shows the CCA of corresponding subject with specific method of spatial filter. The last row is the averaging CCA of all subjects. Table 1 shows the result of studied method without the feature reduction PCA. While Tables 2 and 3 show results of studied method of PCA with 90% rule of eigenvalues and the leave-one-out in CCA techniques.

**Table 2**
Averaging results of correct classification accuracy (CCA) rate of all test data size (10%, 20%, . . ., 90%) in a percentage of method with feature reduction using PCA by eigenvalue 90% rule selection method. S.F. is spatial filter type, where S.J. stands for subject name.

| S.J. | S.F. | | | |
|---|---|---|---|---|
| | No SF CCA (%) | CAR SF CCA (%) | LAP SF CCA (%) | CSP SF CCA (%) |
| aa | 54.64 | 56.01 | 57.19 | 61.23 |
| al | 72.85 | 86.62 | 92.60 | 90.30 |
| aw | 53.27 | 63.93 | 54.22 | 56.33 |
| av | 62.12 | 72.10 | 72.54 | 79.43 |
| ay | 68.70 | 84.38 | 70.98 | 71.36 |
| Mean | 62.32 | 72.61 | 69.50 | 71.73 |

**Table 3**
Averaging results of correct classification accuracy (CCA) rate of all test data size (10%, 20%, . . ., 90%) in percentage of method with leave-one-correct classification accuracy selection method. S.F is spatial filter type, where S.J stands for subject name.

| S.J. | S.F. | | | |
|---|---|---|---|---|
| | No SF CCA (%) | CAR SF CCA (%) | LAP SF CCA (%) | CSP SF CCA (%) |
| aa | 60.70 | 63.37 | 55.96 | 60.61 |
| al | 83.27 | 86.17 | 92.59 | 88.59 |
| aw | 61.33 | 64.27 | 56.58 | 53.33 |
| av | 73.82 | 78.13 | 73.15 | 76.07 |
| ay | 79.20 | 82.20 | 69.72 | 68.00 |
| Mean | 71.66 | 74.83 | 69.60 | 69.32 |

The averaging CCA of all subjects is showed in Fig. 5. In the mean spatial filter method, result indicated that CAR method performance better than the other method at 74.59% while CCA of Laplacian spatial filter is at 70.06%, of no-spatial filter is at 68.78% and of CSP accuracy is at 68.60% in average. Even CSP is the famous for general BCI pre-processing and feature extraction but it performs poorly for variant EEG signal as BCI III data set Iva. While other spatial filter as CAR or Laplacian best CSP method, as the filters do not depend the estimation of covariance which is highly sensitive of time-varying signal. In term of feature reduction method, Fig. 5 shows that features of BCI III data set Iva are highly uncorrelated. As the result for reducing feature using PCA could diminish the information of the feature like in a case of PCA method with

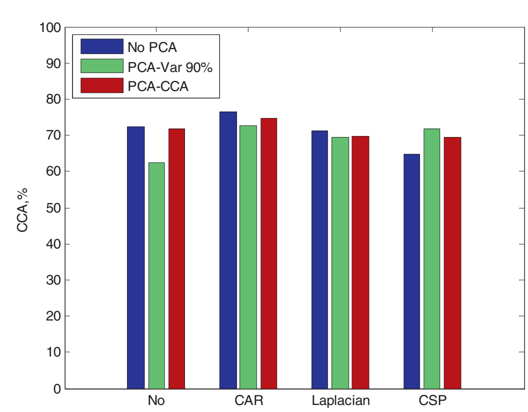


**Fig. 5.** Averaging CCA results of all subject. *X* axis show different types of spatial filters. *Y* axis shows CCA of each method. Blue bar indicates for no feature reduction method, green is for with PCA method using 90% rule of variance and red is for PCA method with leave-one-out CCA selection method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

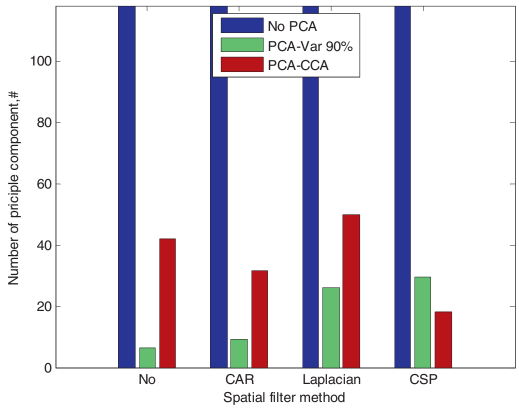


**Fig. 6.** Averaging number of features of all subject. *X* axis show different types of spatial filters. *Y* axis shows a number of features. Blue bar indicates for no feature reduction method, green is method with PCA using 90% rule of variance and red is for PCA method with leave-one-out CCA selection. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

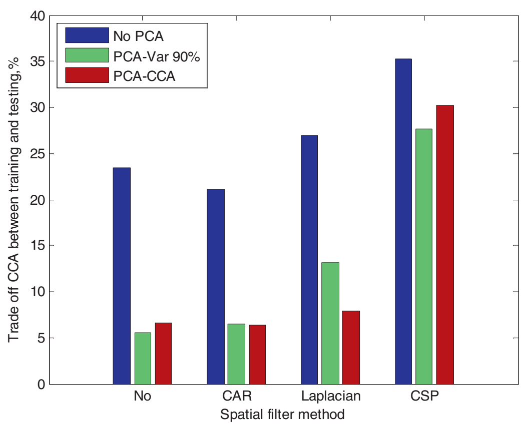


**Fig. 7.** Averaging trading off CCA (TCCA) from training to test data of all subjects. *X* axis show different types of spatial filters. *Y* axis shows TCCA of each method. Blue bar indicates for no feature reduction method, green is method with PCA using 90% rule of variance and red is for PCA method with leave-one-out CCA selection method. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

90% rule technique and gets a stable performance for PCA based CCA selection method compared to not using PCA at all.

But performance of CCA does not all determine the strong factor of feature extraction method. Number of features also plays an important role as the smaller number of feature could have a good generalization for application. Fig. 6 shows the comparison of the average number of feature creation by each method. Without PCA, the number of features is constant at 118, while for PCA with 90% variance selection is at 17.86 and PCA with CCA selection is at 35.35 average number of feature.

The last point we would like to discuss is the effectiveness of tracking CCA performance from the training data to testing data. This factor should be considered as an algorithm maybe performance well in training data but lack of generalization causing it to have low performance if test data or other subject data. Fig. 7 shows comparison of trading off between training and testing performance of both spatial filter method and feature reduction using PCA method. Comparing the spatial filter method, CAR has the lowest trading off performance than other spatial filter methods. Also the result shows that even PCA could not improve CCA performance for non-stationary EEG data, but PCA could improve the generalization of feature extraction from training to test real data and also could maintain CCA performance like PCA with CCA selection method.

## 5. Conclusion

In this paper, we analyzed the performance of spatial filter and PCA method for feature extraction and reduction for non-stationary EEG signal. Simulation result of noise and time-varying data shows that the common average reference filer has better performance than other universal spatial filter techniques. In time-varying data, PCA with leave-one-out CCA selection could not help improve accuracy but could maintain the classification performance while decreasing feature number effectively.

## Acknowledgment

This work was supported by the Min-career Researcher Program through an NRF grant funded by the MEST (No. 2012-0008726).

## References

[1] J.R. Wolpaw, C.B. Boulay, Brain signals for brain–computer interface, in: Brain–Computer Interface, The Frontiers Collection, Springer-Verlag Heidelberg, Berlin, 2010, pp. 29–46.
[2] G. Pfurtscheller, C. Neuper, Dynamics of sensorimotor oscillations in a motor task, in: Brain–Computer Interface, The Frontiers Collection, Springer-Verlag Heidelberg, Berlin, 2010, pp. 47–64.
[3] B. Hjorth, Principles for transformation of scalp EEG from potential field into source distribution, J. Clin. Neurophysiol. 8 (4) (1991) 391–396.
[4] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, et al., Optimizing spatial filters for robust EEG SingleTrial Analysis, IEEE Signal Proc. Mag. 25 (1) (2008) 41–56.
[5] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, Brain–computer interface for communication and control, J. Clin. Neurophysiol. 113 (6) (2002) 767–791.
[6] D.J. McFarland, G.W. Neat, J.R. Wolpaw, An EEG-based method for graded cursor control, Psychobiology 21 (1) (1993) 77–81.
[7] J. Le, A. Gevins, Method to reduce blur distortion from EEG's using a realistic head mode, IEEE Trans. Biomed. Eng. 40 (6) (1993) 517–528.
[8] G. Santhanam, S.I. Ryu, B.M. Yu, A. Afshar, K.V. Shenoy, A high-performance brain–computer interface, Nature 422 (2006) 195–198.
[9] G. Domhenge, B. Blankertz, G. Curio, et al., Boosting bit rates in non-invasive EEG single-trial classifications by feature combination and multi-class paradigms, IEEE Trans. Biomed. Eng. 51 (June (6)) (2004) 993–1002.