33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

# Performance of common spatial pattern under a smaller set of EEG electrodes in brain-computer interface on chronic stroke patients: a multi-session dataset study

Wing-Kin Tam, Zheng Ke, Kai-Yu Tong *Senior Member IEEE*

***Abstract*—Brain-computer interface (BCI) uses non-muscular channel of the nervous system for communication. Common Spatial Pattern (CSP) is a popular spatial filtering method used to reduce the effect of volume conduction on EEG signals. It is thought that CSP requires a large number of electrodes to be effective. Using a 20-session dataset of motor imagery BCI usage by 5 stroke patients, we demonstrated that after channel selection, CSP can still maintain a high accuracy with low number of electrodes using a newly proposed channel selection method called CSP-rank (higher than 90% with 8 electrodes). The results showed that using only the first session for channel selection, a high accuracy can be maintained in subsequent sessions. CSP-rank has been compared to the popular support vector machine recursive feature elimination (SVM-RFE). The results showed that the CSP-rank required less electrodes to maintain accuracy higher than 90% (a minimum of 8 compared to 12 of SVM-RFE) and it attained a higher maximum accuracy (91.7% compared with 90.7% of SVM-RFE). This could support clinicians to apply more BCI in routine rehabilitation.**

## I. Introduction

Brain-computer interface uses non-muscular channel of the nervous system to communicate with the outside world [1]. It allows total lock-in patient to communicate with the others. Recently it has been applied in rehabilitation [2]. Common spatial pattern (CSP) is a popular spatial filtering method in brain-computer interface (BCI). Due to the volume conduction property of EEG signal, signals recorded at the scalp are often "smeared" and have a low spatial resolution [3]. There are quite a few spatial filters available (e.g. common average reference, Laplacian filter [1], linear inverse method [4]) to reduce this "smearing" effect. CSP receives special appeal because one of its features is to increase the separation between filtered signals of different classes, and so it can enhance the classification performance. CSP tries to find a spatial filter that maximizes the difference in variance between two classes of data. This fits well with the operation of a motor imagery BCI, which is based on the modulation of the variance of EEG signal (i.e. the band power).

Manuscript received 13 April 2011. The project was funded by the Research Grants council of the Hong Kong Special Administrative Region, China (grant no. PolyU 5292/08E)

W.K.Tam is with the Department of Health Technology and Informatics, the Hong Kong Polytechnic University, Hong Kong, China, SAR

Z. Ke is with the Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong, China, SAR

K.Y.Tong is with the Department of Health Technology and Informatics, the Hong Kong Polytechnic University, Hong Kong, China, SAR (Corresponding author, tel: 852-2766 7662; E-mail: k.y.tong@polyu.edu.hk)

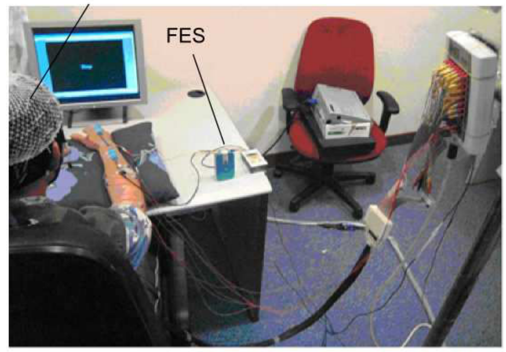


Fig. 1. A photograph showing the set up of the BCI training with a chronic stroke patient

There is a common understanding that CSP can only be effective if there are large amount of electrodes available [5], [6], but using a large number of electrodes would imply longer time spent in channel preparation. It may also increase the cost of a BCI system as more amplifiers are needed. This will undermine the practical efficiency of using a BCI system. Several studies have used CSP as a channel selection method [6], [7], [8], but the performance of CSP under low number of electrode is seldom studied. The previous studies are also mainly focused on single session dataset. EEG signals tend to be non-stationary across different measuring sessions [9]. Whether the results obtained from one session of recording (as in [6]) can be applied to multiple sessions of data still needs to be further studied.

In this study, we used a 20-session dataset of stroke patients using a motor imagery based BCI to show that the CSP-rank method, which is based on the sorting of CSP filter coefficients, was effective in reducing the number of electrodes. We provided evidence that CSP can still perform well using a low number of electrodes after simple channel selection. The results also showed that using only 1 calibration session for channel selection, its results can be applied on the subsequent 19 sessions with high accuracy. A large number of electrodes can be reduced with only small impact in classification accuracy.

## II. Methodology

### A. Data description

Five chronic stroke patients were recruited for the study (3 females, 2 males. Age: 50.4±17.3. Affected side: 3L, 2R). They were instructed to sit in a comfortable chair in a sound proof room, facing a 17-inch screen. 64-channel EEG following the extended international 10-20 system was

978-1-4244-4122-8/11/$26.00 ©2011 IEEE 

recorded from each patient using a Neuroscan Synamps2 (Neuroscan Inc, Hernon, USA) and SCAN 4.3 software at 250Hz. Electrode impedance was kept under 5kΩ. The experimental setup was shown in Fig. 1.

The experiment followed the famous "basket-paradigm". A diagram showing the paradigm was shown in Fig. 2. First a "+" sign was shown on the screen for 3s. After that, a blue basket was displayed on the screen, on the same side as the affected side of the patient. The patients were asked to perform motor imagery on their affected side (the MI phase). A feedback ball moved from the bottom to the top of the screen at uniform speed in 5.6s. If the EEG signal was classified as motor imagery, then the feedback ball would move horizontally towards the affected side of the subject,

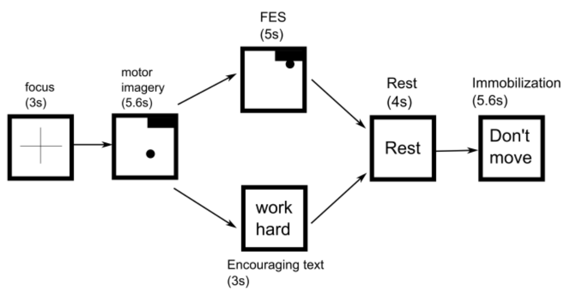


Fig. 2. Experimental paradigm showing 1 trial of the BCI experiment

otherwise it moved to the opposite side. Further details of the online classification were given at [10]. If the feedback ball hit the basket, functional electrical stimulation (FES) would be applied on extensor carpi radialis of the subject for 5s to extend the finger for hand opening exercise. Otherwise, an encouragement text will be shown for 3s. The subject would then be allowed to rest for 4s. After that, he was asked to remain stationary for 5.6s without any motor imagery (the Immobilization phase). The classification task was to distinguish the EEG signal during motor imagery (MI) and that during the Immobilization phase. One BCI session consisted of 80 trials. Each trial contained data from both classes (i.e. MI and Immobilization phase).

*B. Common spatial pattern*

Algorithm for common spatial pattern is given below [3], [11] :

For two signal matrices, $\boldsymbol{X}_1$ of class 1 and $\boldsymbol{X}_2$ of class 2, where column is channel, the covariance matrices are given by:

$$\boldsymbol{R_1} = \frac{\boldsymbol{X}_1\boldsymbol{X}_1^T}{trace(\boldsymbol{X}_1\boldsymbol{X}_1^T)} \qquad \boldsymbol{R_2} = \frac{\boldsymbol{X}_2\boldsymbol{X}_2^T}{trace(\boldsymbol{X}_2\boldsymbol{X}_2^T)}$$

The average covariance $\overline{\boldsymbol{R_1}}$ *and* $\overline{\boldsymbol{R_2}}$ are obtained by averaging the covariance matrices $\boldsymbol{R}_1$*and* $\boldsymbol{R}_2$ over all the trials of the respective class.

Perform Eigen vector decomposition (EVD) on the sum of average covariance:

$$\boldsymbol{R} = \boldsymbol{R_1} + \boldsymbol{R_2} = \boldsymbol{U}_0\boldsymbol{\lambda}\boldsymbol{U}_0^T$$

where $U_0$ is the matrix of eigen vectors and $\lambda$ the diagonal matrix of eigenvalues. Find the whitening transform matrix:

$$\boldsymbol{P} = \boldsymbol{\lambda}^{-\frac{1}{2}}\boldsymbol{U}_0^T$$

Then transform the covariance matrices:

$$\boldsymbol{S}_1 = \boldsymbol{P}\overline{\boldsymbol{R}_1}\boldsymbol{P}^T \qquad \boldsymbol{S}_2 = \boldsymbol{P}\overline{\boldsymbol{R}_2}\boldsymbol{P}^T$$

Do EVD again:

$$\boldsymbol{S}_1 = \boldsymbol{U}\boldsymbol{\lambda}_1\boldsymbol{U}^T \qquad \boldsymbol{S}_2 = \boldsymbol{U}\boldsymbol{\lambda}_2\boldsymbol{U}^T$$

$S_1$ and $S_2$ share common eigenvectors, so $\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2 = 1$

Since $\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2 = 1$ , the eigenvector corresponding to the largest eigenvalue in $\boldsymbol{\lambda}_1$ will have the smallest eigenvalue in $\boldsymbol{\lambda}_2$, and vice versa. The 2 eigenvectors corresponding to the largest eigenvalue in $\boldsymbol{\lambda_1}$ *and* $\boldsymbol{\lambda_2}$ are extracted. They are called $\boldsymbol{U_1}$ *and* $\boldsymbol{U_2}$ respectively

The spatial filter can be found by:

$$\boldsymbol{SF}_1 = \boldsymbol{U}_1\boldsymbol{P} \qquad \boldsymbol{SF}_2 = \boldsymbol{U}_2\boldsymbol{P}$$

Then the signal matrix can be projected by

$$\boldsymbol{Y}_1 = \boldsymbol{SF}_1\boldsymbol{X} \qquad \boldsymbol{Y}_2 = \boldsymbol{SF}_2\boldsymbol{X}$$

where $\boldsymbol{Y}_1$ *and* $\boldsymbol{Y}_2$ are filtered feature vector for class 1 and 2 respectively.

*C. Channel selection methods*

*1) CSP-rank method*

A method called CSP-rank based on the sorting of CSP filter was proposed. The CSP algorithm produced two spatial filters $\boldsymbol{SF_1}$ and $\boldsymbol{SF_2}$ (the eigenvectors that correspond to the largest and smallest eigenvalue respectively) for class 1 and 2 respectively. They are the spatial filter coefficients to generate two new filtered signals from the original EEG signal.

Another way to look that these filter coefficients is that they assign different weights to different electrodes based on their importance. If the coefficient of a particular electrode is large, then that means the electrode will contribute more to the resulting filtered signal and hence it is more important. Conversely, removing the electrode with a small coefficient has less effect on the resulting signal due to its smaller contribution. The CSP-rank method first sorted the absolute value of the filter coefficients in each filter respectively, then take the electrode with the next largest coefficient in turn from the two spatial filters (e.g. take the first electrode from the sorted $\boldsymbol{SF_1}$, second from $\boldsymbol{SF_2}$, third from $\boldsymbol{SF_1}$ again etc.). If an electrode is already taken, then simply move on to the next coefficient in the same spatial filter until a new electrode is reached.

*2) Support-vector machine recursive feature elimination (SVM-REF)*

Support vector machine (SVM) tries to search for a hyperplane that maximize the margin between the support vectors of 2 classes. For a soft-margin SVM that accommodate non-linearly separable cases, it tries to solve the following optimization problem [12], [13]:

$$\min_{w,b} \ \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i \tag{8}$$

Subject to $y_i(\mathbf{w}^t\mathbf{x_i} + b) \geq 1 - \xi_i$

For training sample $\boldsymbol{x_i} \in \Re^d, i = 1, 2, ..., n$, where n is the total number of training samples, each with a class label $y_i \in \{1, -1\}$ where $\boldsymbol{y} \in \Re^n$. d is the number of features in each sample. $\xi_i$ is the slack variable. It allows some samples to be in the margin or misclassified just in case they are not linear separable. C is the soft margin constant that determines how tolerant the classifier is to misclassified or margin errors.

For channel selection, the ranking score of a channel is given by [14], [15]:

$$R_j = (w_j)^2 \tag{9}$$

where $w_j$ is the weight for channel j in the projection vector **w** returned by the SVM algorithm. In each iteration, the channel with the lowest ranking score is eliminated from the training samples. Then SVM is applied to the sample again to generate a new projection vector **w.** The process continues until there is only one channel left. In this study, LIBSVM [16], a library for support vector machines developed at the National Taiwan University, was used for the SVM training. Model selection of parameter C in equation (2) was done by a 10-fold cross-validation with the training data. The value of C that gave the highest validation accuracy was chosen.

### 3) *Random selection*

The electrodes are just ranked randomly in Random selection to act as a control for comparison.

### *D. Signal processing and cross-validation*

There were totally 20 sessions of BCI data. Each session was carried out on a different day. For each subject, all trials in the first day of training were used for channel selection. The trials were first band-pass filtered between 8-12Hz and then fed into the different channel selection algorithms. For SVM-RFE, the log-variance of each channel was used as the feature. After channel selection, a list of channel sorted by their importance in channel selection was obtained.

Assuming the list was sorted in descending order of importance, using the first N (N ranged from 2 to 50, at a step of 2) channel only, classification and 10-fold cross-validation was done on all subsequent BCI session (i.e. day 2 to day 20). For each subsequent BCI session, its 80 trials were divided into 10-blocks. Each block respectively acted as the testing set and the rest acted as the training set. For the training set, the data was first filtered between 8-12Hz, in the mu-rhythm band [17]. A CSP filter was trained and then applied on the data, and then a Fisher's Linear Discriminant (FLD) was trained on the log-variance of the spatially filtered signal. The trained spatial filter and classifier are then used to classify the testing set. Accuracies from all folds of the validation were averaged. Finally, a grand average of accuracy is obtained by averaging the accuracy across the sessions (session 2 to session 20) and all the subjects.

## III. Results

The paired t-test was used to compare the accuracy between different methods at different number of electrodes. The number of electrodes that showed statistical difference ($p<0.05$) between CSP-rank or SVM-RFE against Random selection was marked in Fig. 3. Statistical difference can be observed between 14 to 26 electrodes. There was no statistical difference between CSP-rank and SVM-RFE.

The relation between the average classification accuracy and number of electrodes was shown in Fig. 3. CSP-rank and SVM-RFE had better performance than Random at nearly all number of electrodes. Surprisingly, Random selection was able to obtain accuracy as high as 89.6%, although using more electrodes. Even at 10 electrodes, its accuracy was still above 85%. However, when the number of electrode dropped below 10, the accuracy rapidly deteriorated.

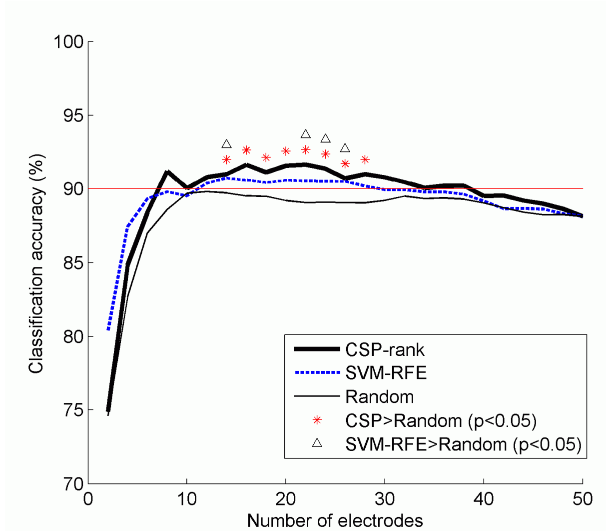


Fig. 3. Relation between the average classification accuracy and the number of electrodes for different channel selection algorithms. The number of electrodes showing statistical difference between CSP-rank and Random, SVM-RFE and Random was marked with asterisk and triangle respectively. The point of 90% accuracy was marked with a horizontal line.

Accuracy above 90% was maintained by CSP-rank for 8-38 electrodes, while SVM-RFE needed 12-28 electrodes. Both methods showed a tendency to produce higher accuracy when the number of electrodes was reduced from the maximum. They reached a peak in performance when the number of electrodes was between 10-20. When the number of electrodes was reduced below 8, the accuracy started to drop quickly. However, it should be noted that there was no similar trend in the Random selection method. Its accuracy remains relatively stable from 10 to 50 electrodes.

CSP-rank obtained the highest accuracy of 91.7% with 22 electrodes. SVM-RFE was able to obtain the highest of 90.7% with 14 electrodes. The random methods require a number as large as 32 to attain its maximum 89.6% accuracy.

## IV. Discussion

Our findings showed a smaller set of EEG electrodes with CSP could be applied for BCI. It was demonstrated with 20-session training on stroke patients. There is a common understanding among the BCI field that CSP requires a large number of electrodes to be effective. Several studies have used CSP as a channel selection method [6], [7], [8], [18] but usually in their assessment of classification accuracy, CSP was not used (except in [18], but [18] only used a fixed number of electrodes). The performance of CSP under lower channel count is not thoroughly examined. We used a 20-session BCI dataset from stroke patients to show that CSP could still maintain a high performance even with a low number of electrodes. Our study has used multiple BCI sessions to investigate the effect of channel selection, unlike most studies which only rely on cross-validation of a single session dataset. We showed that channel selection was effective despite the non-stationarity of EEG.

Our results showed that having more electrodes did not necessarily mean higher accuracy. This contrasts with the common standpoint that CSP requires large number of electrode to be effective. Both CSP-rank and SVM-RFE had the tendency to produce higher accuracy when the number of electrodes was reduced. This is consistent with previous studies as CSP was reported to have a high tendency to over-fit [19]. CSP was sensitive to artifact [9], [20]. Removing noisy channel can relieve the effect of over-fitting and artifact contamination. Statistical difference between specialized channel selection method and Random selection have been observed between 14 to 26 electrodes, suggesting that channel selection would be most effective among this range of electrodes.

CSP-rank had several advantages over SVM-RFE. At minimum, CSP-rank only required 8 electrodes to keep the classification above 90%, but SVM-RFE needed at least 12. The maximum accuracy obtained by CSP-rank was also higher. Computationally CSP-rank has major advantage over SVM-RFE as it is not a recursive method. Results can be obtained almost instantaneously. This feature allows CSP-rank to be embedded in the real-time signal processing pipeline and remove the channel not useful for channel selection. This can reduce the chance of over-fitting. The effectiveness of this approach in real-time processing will be addressed in future studies.

Our results suggested that although EEG signals are non-stationary, the location of the electrode important for classification was relatively stable across sessions. We performed the channel selection on the first BCI session and tested its performance on subsequent sessions. The average accuracy obtained in the testing sessions was very high (can be >90%), showing the spatial stability of the important channels. The high accuracy obtained by the Random selection method also showed that CSP was very capable to discriminate between two classes of motor imagery even given a random configuration of electrodes.

The authors were well aware that using CSP-rank as channel selection and again using CSP in classification may lend to bias. However, given the popularity of CSP in motor imagery BCI, this should be considered as one of the feature why CSP-rank is preferable. The tight integration between channel selection and spatial filter can lead to increased performance.

## V. Conclusion

Using a 20-session dataset, our results showed that CSP can still maintain a high accuracy under low number of electrodes after channel selection. The results of channel selection performed on the first calibration session for channel selection can be applied on subsequent sessions and maintain a high overall accuracy on chronic stroke patients. CSP-rank is a viable method for channel selection in motor imagery BCI using CSP. It required fewer electrodes and attained a higher accuracy compared with SVM-RFE. Generally, 8 electrodes were sufficient to maintain accuracy higher than 90%.

## VI. References


[1] J. Wolpaw, N. Birbaumer, D. Mcfarland, G. Pfurtscheller, and T. Vaughan, “Brain-computer interfaces for communication and control,” *Clinical Neurophysiology*, vol. 113, 2002, pp. 767-791.

[2] E. Buch, C. Weber, L.G. Cohen, C. Braun, M.A. Dimyan, T. Ard, J. Mellinger, A. Caria, S. Soekadar, A. Fourkas, and N. Birbaumer, “Think to move: a neuromagnetic brain-computer interface (BCI) system for chronic stroke.,” *Stroke*, vol. 39, Mar. 2008, pp. 910-7.

[3] J. Muller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, “Designing optimal spatial fiters for single-trial EEG classification in a movement task,” *Clinical Neurophysiology*, vol. 110, 1999, pp. 787-798.

[4] F. Cincotti, D. Mattia, F. Aloise, S. Bufalari, L. Astolfi, F. De Vico Fallani, A. Tocci, L. Bianchi, M.G. Marciani, S. Gao, J. Millan, and F. Babiloni, “High-resolution EEG techniques for brain-computer interface applications.,” *Journal of neuroscience methods*, vol. 167, Jan. 2008, pp. 31-42.

[5] C. Guger, H. Ramoser, and G. Pfurtscheller, “Real-time EEG analysis with subject-specific spatial patterns for a brain-computer interface (BCI).,” *IEEE transactions on rehabilitation engineering*, vol. 8, Dec. 2000, pp. 447-56.

[6] J. Farquhar, N. Hill, T.N. Lal, and B. Schölkopf, “Regularised CSP for sensor selection in BCI,” *Proceedings of the 3rd International Brain-Computer Interface Workshop and Training Course*, Citeseer, 2006, p. 14–15.

[7] J. Lv and M. Liu, “Common Spatial Pattern and Particle Swarm Optimization for Channel Selection in BCI,” *Information and Control*, 2008, pp. 0-3.

[8] Y. Wang and S. Gao, “Common spatial pattern method for channel selelction in motor imagery based brain-computer interface,” *2005. IEEE-EMBS 2005. 27th Annual*, vol. 5, Jan. 2005, pp. 5392-5.

[9] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-robert Muller, “Optimizing Spatial filters for Robust EEG Single-Trial Analysis,” *IEEE Signal Processing Magazine*, vol. 25, 2008, pp. 41-56.

[10] F. Meng, K.-Y. Tong, S.-tak Chan, W.-wa Wong, K.-him Lui, K.-wing Tang, X. Gao, and S. Gao, “BCI-FES training system design and implementation for rehabilitation of stroke patients,” *2008 IEEE International Joint Conference on Neural Networks*, IEEE, 2008, pp. 4103-4106.

[11] Y. Wang, S. Gao, and X. Gao, “Common spatial pattern method for Channel Selection in Motor Imagery Based Brain-computer Interface,” *Engineering in Medicine and Biology Society 27th Conference*, Shanghai, China: .

[12] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, 1995, pp. 273-297.

[13] C.-wei Hsu, C.-chung Chang, and C.-jen Lin, “A Practical Guide to Support Vector Classification,” vol. 1, 2009.

[14] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene selection for cancer classification using support vector machines,” *Machine learning*, 2002, pp. 389-422.

[15] T.N. Lal, M. Schröder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, and B. Schölkopf, “Support vector channel selection in BCI,” *IEEE Transactions on Biomedical Engineering*, vol. 51, Jun. 2004, pp. 1003-10.

[16] C.-C. Chang; and C.-jen Lin, “LIBSVM : a library for support vector machines,” 2001.

[17] M. Cheng, W. Jia, X. Gao, S. Gao, and F. Yang, “Mu rhythm-based cursor control: an offine analysis,” *Clinical Neurophysiology*, vol. 115, 2004, pp. 745-751.

[18] J. Meng, G. Liu, G. Huang, and X. Zhu, “Automated selecting subset of channels based on CSP in motor imagery brain-computer interface system,” *2009 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Dec. 2009, pp. 2290-2294.

[19] N. Hill, T. Lal, M. Schröder, T. Hinterberger, G. Widman, C. Elger, B. Schölkopf, and N. Birbaumer, “Classifying event-related desynchronization in EEG, ECoG and MEG signals,” *Pattern Recognition*, 2006, p. 404–413.

[20] M. Kawanabe and C. Vidaurre, “Improving BCI Performance by Modified Common Spatial Patterns with Robustly Averaged Covariance Matrices,” *World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany*, Springer, 2009, p. 279–282.