A study on temporal segmentation strategies for extracting common spatial patterns for brain computer interfacing

Javier Asensio-Cubero, John Q. Gan, Ramaswamy Palaniappan

Abstract— **Brain computer interfaces (BCI) create a new approach to human computer communication, allowing the user to control a system simply by performing mental tasks such as motor imagery. This paper proposes and analyses different strategies for time segmentation in extracting common spatial patterns of the brain signals associated to these tasks leading to an improvement of BCI performance.**

I. Introduction

Brain Computer Interfaces (BCI) are communication systems that use human thoughts as a control signal [1]. These systems are particularly valuable for paralysed users who may not be able to interact with computers in any other manner. From a non-disabled user's point of view, BCIs can enrich the human computer interaction where aspects from their mental state, such as emotions and error related activity, can be taken into account. Regular users suffering from an induced disability (situations where the user concentration or attention may be compromised, such as surgeons or pilots) may also benefit from this kind of human-computer interaction [2].

BCIs are classified into several paradigms depending on which mental state or signal type is utilised [3]. In this study, we focused on motor imagery (MI) using electroencephalography (EEG) as a method of recording the signals. When a subject performs a limb movement, several areas on the brain cortex are activated due to different neuron populations' firing signals. Some of these populations show activity even if the subject does not perform movement at all, just imagining the limb movement is sufficient to produce changes of state in the motor cortex [4].

Limb movement imagery is characterised by short lasting amplitude attenuations/amplifications in the EEG signals known as even related desynchronisation (ERD) and event related synchronisation (ERS) [5][6]. Many BCI designs rely on ERD/ERS to discriminate MI movements (such as hands, feet, fingers, tongue, etc) [7],[8]. ERD/ERS components can be found in temporal, spatial and spectral domains. Different researches use different techniques to find the most discriminant features in each domain. For example, many studies focus on spatial components such as common spatial patterns (CSP) [9]. Some researches try to extract relevant information from the ERD/ERS time course using techniques like local discriminant bases (LDB) [10]. Many studies combine elements from two or three different domains, such as PARAFAC based methods [11][12], common sparse spectral spatial pattern (CSSSP) [13], filter bank common spatial pattern (FBCSP) [14] and wavelet common spacial pattern (WCSP) [15].

The time duration given to the subject for imagining the limb moment is called a trial, and it is where the ERD/ERS occurs. Depending on the experiment protocol, the trial duration may vary from four to eight seconds. The classification of the data obtained from the feature extraction can be performed sample by sample, giving a classification result for every sample in the input data, or trial by trial, where only a single prediction is given for the trial sample set.

CSP has been popularly used in the literature for feature extraction for BCI due to its ability of locating the active sources while maximising the variance among two or more classes. Usually CSP is applied for trial by trial classification. This paper applies CSP using various segments of the trial aiming to capture both spatial and temporal features from EEG signals for BCI applications.

This paper is organised as follows: Section II explains the methodology: data acquisition (Section II-A), feature extraction methods (Section II-B), classification techniques (Section II-C) and time segmentation strategies (Section II-D). Section III describes the obtained results and conclusions are drawn in Section IV.

II. Methods

A. Data Acquisition

The data used for this study is obtained from the BCI competition IV (data set 2a [16]) which is publicly available, allowing us to place our outcomes with the best ranked methods. The data contains four different classes: imaginary movement of right hand, left hand, feet and tongue, from nine different subjects. The subjects sat in an arm-chair facing a computer screen with 22 electrodes placed on the scalp following the international 10-20 location system (as shown in Figure 1). Initially, at $t = 0$, a fixation cross was printed on the screen, after two seconds $t = 2$ an arrow was displayed indicating which imaginary class to perform and this cue was shown until $t = 3.25$. The fixation cross disappeared at $t = 6$ and denoted the end of the trial. The EEG data was recorded at 250Hz and band pass filtered between 0.5 and 100 Hz. During preprocessing, an elliptic band pass filter was applied to filter the data in pass band range of 8 to 30 Hz.

Two sessions of EEG data were recorded from each subject, 288 trials (72 for each class) were acquired per

Javier Asensio-Cubero University of Essex; email: jasens@essex.ac.uk
John Q. Gan University of Essex; email jqgan@essex.ac.uk
Ramaswamy Palaniappan University of Essex; email: rpalan@essex.ac.uk

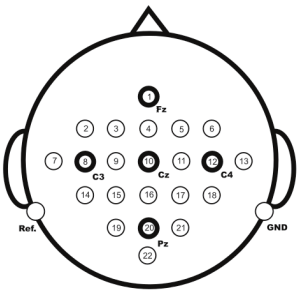

Fig. 1. Using 22 electrodes placed according to the 10-20 international standard [16]

session. The first session dataset is used as training data while the second session is used for evaluation in our experiments.

B. Common Spatial Patterns

Methods like Principal Component Analysis (PCA) [17] and Independent Component Analysis (ICA) [18] rely on statistical relationships to extract the most relevant features from a data set and have been extensively applied in domains such as video compression or image processing. CSP is based on PCA decomposition and can be regarded as a supervised blind source separation technique [19] which maximises the variance between two different classes. As this is the method that we are using for feature extraction stage, an introduction to its basics is given next.

Let us consider a matrix X_i of EEG data captured during an interval of length T, namely a trial. The dimension of X_i will be $N \times T$ as the signal is captured from N different electrodes. Consider X_i centred and scaled $X_i = \frac{1}{\sqrt{T}} X_i^{orig}(I_t - 1_t 1_t^T)$ where I_t is the $T \times T$ identity matrix and 1_t is a T dimensional vector with ones in it [9]. Now we estimate the covariance matrix for all the X_i samples as:

$$\Sigma = \bar{X}\bar{X}^T$$

where $\bar{X}$ is the mean of the X_i samples.

As Σ is symmetric, the eigenvalue decomposition results in:

$$\Sigma = W\Lambda W^T$$

with W being a matrix containing the eigenvectors of Σ and Λ a diagonal matrix with its eigenvalues. Now we choose the eigenvectors associated with the highest eigenvalues to build $\tilde{W}$, which contains the set of principal components.

From an intuitive point of view, this process is simply the projection of a sample against a subspace built upon the covariance of the whole sample set of X. The new orthogonal base built on the eigenvectors assures that only those components with more variance will survive allowing us to discard redundant components which contain less information.

CSP is an extension to PCA where two different classes of data are taken into account (e.g. left hand motor imagery vs right hand motor imagery). Therefore, the set of samples X is divided into $X^{(+)}$ and $X^{(-)}$, their simultaneous estimated covariance matrix decomposition is given by [9][20]:

$$\Sigma^{(+)} = W\Lambda^{(+)}W^T$$
$$\Sigma^{(-)} = W\Lambda^{(-)}W^T$$

where W is determined in such a way that $\Lambda^{(+)} + \Lambda^{(-)} = I$. Large values of $\lambda_j^{(+)}$ mean that the corresponding w_j obtains high variance in the positive class and low variance in the negative one (and vice-versa). Now it remains to choose those eigenvectors w_j that maximises the variance for both classes. This discrimination can be performed based on the discriminative activity S_d and the common activity S_c:

$$S_d = \Sigma^{(+)} + \Sigma^{(-)}$$
$$S_c = \Sigma^{(+)} - \Sigma^{(-)}$$

The eigenvectors are selected by solving the following maximisation problem:

$$\max_{w \in \mathbb{R}} = \frac{w^T S_d w}{w^T S_c w}$$

From the previous steps, we obtain a set of spatial patterns W that can be used to extract the most important features from EEG signals for BCI applications.

C. Classification

In this study, Linear Discriminant Analysis (LDA) is used as the classifier. In spite of its simplicity, this model has proved to achieve comparable results to other approaches such as support vector machines and artificial neural networks [21]. The main benefit comes from its low computational resource consumption, being much faster than the other mentioned methods. The linear discrimination is based on the discrimination function:

$$g(X) = W^T X + w_0$$

where X is the sample to discriminate, W is the weight matrix and w_0 is the bias or threshold whose values are determined by the training data using the Fisher's criterion [22] . The classification is performed simply by deciding that $X \in C_1$ if $g(X) > 0$ or $X \in C_2$ otherwise [23].

LDA only allows to discriminate between two different classes, this problem can be solved using different discriminant functions, one per class. The discriminant function $g_i(X)$ will classify the unseen input X as C_i if $g_i(X) < 0$ or as the meta-class MC_i if $g_i(X) > 0$, $MC_i = \{C_j\}_{j=0, j\neq i}^{N}$ having N different classes in the training set. Therefore, the final label for X is given by :

$$LDA\_label = \arg\min_{i \in N} g_i(X)$$

In oder to measure the classifier performance the kappa [24] value is used along with the classification accuracy. The kappa value is defined as $\kappa = \frac{p_o - p_c}{1 - p_c}$, where p_o is the proportion of units on which the judgement agrees (output from the classifier and the actual label), and p_c is the proportion of units for which the agreement is expected by chance (0.25 for four classes).

D. Time Segmentation and Classification Strategies

As already mentioned, we are going to discuss different strategies of time segmentation within the trial from the instant $t = 2$ to the instant $t = 7$ while extracting common spatial patterns. Three different strategies of time segmentation are applied as depicted in Figure 2. The aim is to investigate the performances of the common approach of applying one single CSP transformation to the whole trial as compared to applying CSP to each segmented trial:

(a) No segmentation is performed by applying CSP directly to the whole trial;
(b) Uniform segmentation (all segments with the same size) without overlapping;
(c) Segmentation with overlapping or sliding window; this approach requires two parameters, the segment length and the overlapping size. In this case the segmentation is performed sliding the first segment a given number of samples as shown in Figure 2;

In every case we will obtain a set of independent features using CSP for each segment.

We will explore two different variants in terms of how the patterns are built for the LDA classification:

(a) One pattern per segment using the features extracted with CSP;
(b) Feature fusion (FF) by joining the different features after applying CSP to every segment. In order to avoid overfitting the classifier, this strategy is only tested with a small number of segments and features;

When it comes to classification, we are going to test two different approaches:

(a) One LDA is applied for all the segments;
(b) One LDA is applied for each segment in the trial;

From these experiments, we will be able to understand whether it is better to have one model for specific parts of the trial or one model that classifies all the segments in the trial.

The last strategy is to apply a voting window to every segment, such that the classification of a segment will depend on the $K-1$ previous outputs from the classifier. Thus, the label for the instant t_i will be $label_{t_i} = mode(\{LDA\_label_{t_{i-k}}\}_{k=0}^{K-1})$ .We expect to assess whether the output for given point can be improved using previous neighbouring data.

For all the experiments, the training set is divided for a ten-fold cross-validation classification, from which the best number of features to select from CSP is obtained. Based on this, a classifier is trained using the whole training set and tested on the evaluation set.

III. Results

In this section, we present the results obtained from applying the strategies proposed in Section II-D. Obviously we cannot explore all the possible combinations as some of them are incompatible or may not make much sense, e.g. applying majority voting on a segment of length three. In

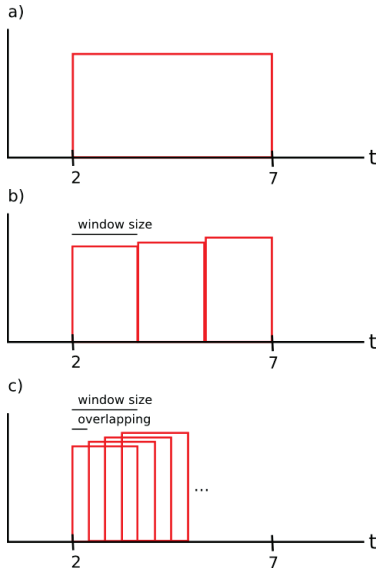

Fig. 2. Different time segmentation strategies a) No segmentation b) Segmentation without overlapping and c) Segmentation with overlapping

TABLE I
MEAN KAPPA AND ACCURACY FROM CSP OVER THE WHOLE TRIAL

	Mean	1	2	3	4	5	6	7	8	9
CV										
Kappa	0.52	0.62	0.50	0.83	0.34	0.21	0.32	0.55	0.77	0.52
Acc	0.64	0.72	0.62	0.87	0.50	0.41	0.50	0.67	0.83	0.64
Eval.										
Kappa	0.42	0.54	0.38	0.67	0.32	0.13	0.22	0.65	0.55	0.38
Acc	0.57	0.65	0.54	0.75	0.49	0.35	0.41	0.73	0.66	0.53

Table I, the result of applying CSP over the whole trial is shown and will be used as benchmark where *CV* stands for cross-validation and *Acc* for accuracy. Regarding the kappa values and accuracy measurements, we have followed the competition procedure where only the segment with the best kappa is taken as the trial output.

Once we have the whole trial data, we can compare it against simple segmentation approaches without overlapping. The numbers of segments in which the trial has been divided are three, five, seven and nine. In this case we have tried also the approach of having one pattern per segment and applying feature fusion. The results shows that segmenting the trial leads to a better performance than using the whole trial (Table II).

Table III shows the performance of an analogous experiment using simple overlapping. The number of segments is again three, five, seven and nine, but the overlapping was calculated in such a way that it maintains the even distributions along the trial.

Our last experiment is configured to assess a more thorough assessment of overlapping effects. For this purpose, we set the segment length to be one second, the sliding window overlap to be 0.8 seconds and a voting window of 3 seconds to update the output for each segment. The results

TABLE II
MEAN KAPPA AND ACCURACY FROM CSP USING DIFFERENT NON-OVERLAPPED SEGMENTATIONS

N. Segments	Kappa CV	Acc CV	Kappa eval.	Acc eval.
No FF				
3	0.54±0.19	0.65±0.14	0.46±0.19	0.59±0.14
5	0.52±0.19	0.64±0.14	0.45±0.18	0.59±0.13
7	0.50±0.21	0.63±0.16	0.44±0.20	0.44±0.15
9	0.48±0.21	0.61±0.16	0.42±0.19	0.56±0.14
FF				
3	0.57±0.19	0.68±0.15	0.50±0.17	0.62±0.13
5	0.57±0.19	0.68±0.15	0.50±0.17	0.62±0.13
7	0.56±0.21	0.67±0.16	0.48±0.19	0.61±0.15
9	0.55±0.20	0.66±0.15	0.47±0.21	0.60±0.16

TABLE III
MEAN KAPPA AND ACCURACY FROM CSP USING DIFFERENT OVERLAPPED SEGMENTATIONS

N. Segments	Kappa CV	Acc CV	Kappa eval.	Acc eval.
No FF				
3	0.57±0.21	0.68±0.16	0.50±0.19	0.62±0.14
5	0.57±0.22	0.68±0.16	0.50±0.20	0.62±0.15
7	0.56±0.21	0.67±0.16	0.48±0.20	0.61±0.15
9	0.55±0.21	0.66±0.16	0.47±0.19	0.60±0.14
FF				
3	0.57±0.21	0.68±0.15	0.48±0.23	0.61±0.17
5	0.58±0.20	0.68±0.14	0.51±0.21	0.64±0.15
7	0.57±0.21	0.68±0.16	0.50±0.21	0.59±0.16
9	0.59±0.19	0.70±0.14	0.47±0.20	0.60±0.15

from this experiment are compared with the winner's of the competition (Table IV). This experiment also includes using a LDA for each segment and one LDA for all the segments within the trial respectively.

IV. CONCLUSIONS

After the evaluation of different time segmentation approaches we can conclude that segmenting the trial leads to a better performance compared with the whole trial approach. Although as we can see in Figure 3 when the segment becomes too small the kappa value and accuracy decrease. This is due to the fact that when we reduce too much the segment size we are losing important temporal information used by CSP to compute the correlation among the different channels. Using feature fusion, which takes into account different segments within the trial, this effect gets attenuated,

TABLE IV
MEAN KAPPA AND ACCURACY FROM CSP USING OVERLAPPING SLIDING WINDOW AND COMPETITION WINNER'S RESULT

	Kappa CV	Acc CV	Kappa eval.	Acc eval
Winner	N/A	N/A	0.57±0.19	N/A
Multiple LDA	0.66±0.21	0.74±0.16	**0.59±0.22**	0.69±0.16
Single LDA	0.65±0.20	0.74±0.15	0.58±0.22	0.69±0.16

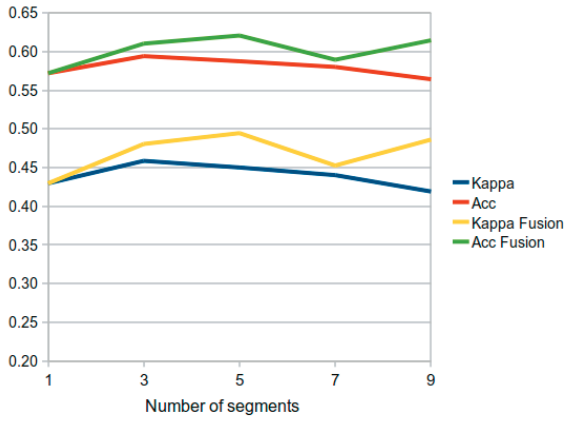

Fig. 3. Kappa and accuracy vs number of segments on the evaluation data

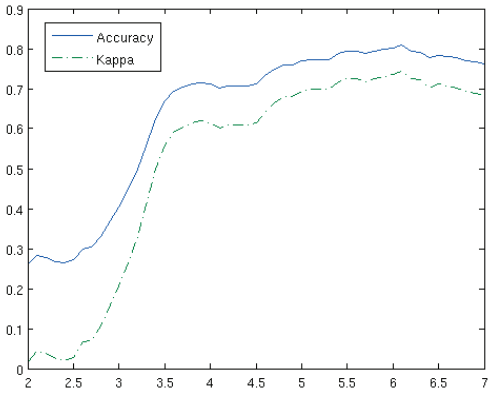

Fig. 4. Best subject's time course with voting window

apart from an increase in the performance probably due to the increase of temporal information within the patterns.

With the use of the overlapping and voting window, we have obtained better results than the winner of the BCI competition as shown in Table IV, even though the winner used FBCSP which is technically much more complex than our approach. In Figure 4 and Figure 5 we can observe the evolution of the kappa value and classification accuracy during the trial for the best subject, Figure 4 shows how the voting window helps to increase the accuracy. Notice from Table IV that the best result is obtained using one LDA per segment, although in practice this approach may be difficult to implement in on-line BCI and needs much more resources than using just a single LDA.

The effect of using the voting window is noteworthy as it boosts the classification accuracy; its effect may be similar to the feature fusion as it adds more temporal information to classifier and thereby makes it more accurate. This approach will be further investigated in the future in order to fully assess its usefulness, including the optimisation of the parameter settings.

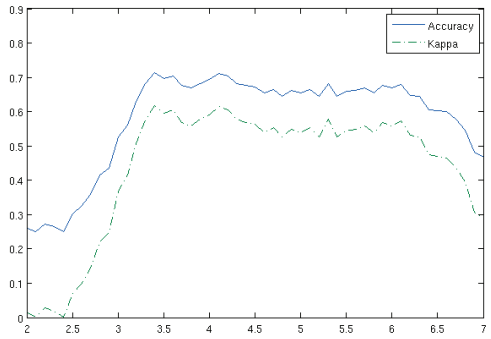

Fig. 5. Best subject's time course without voting window

ACKNOWLEDGMENTS

The authors would like to thank to the EPSRC for funding this research via an EPSRC DTA award.

REFERENCES

[1] A. Nijholt and D. Tan, "Brain-computer interfacing for intelligent systems," *IEEE Intelligent Systems*, vol. 23, no. 3, pp. 72–79, 2008.

[2] B. Allison, B. Graimann, and A. Grser, "Why use a BCI if you are healthy?" in *BRAINPLAY 07 Brain-Computer Interfaces and Games Workshop at ACE (Advances in Computer Entertainment)*, 2007, p. 7.

[3] R. Palaniappan, C. S. Syan, and R. Paramesran, "Current practices in Electroencephalogram-Based Brain-Computer interfaces," *Encyclopedia of Information Science and Technology, 2nd ed.*, vol. II, pp. 888–901, 2009.

[4] G. Dornhege, *Toward Brain-Computer Interfacing*. The MIT Press, 2007.

[5] C. Neuper, M. Wortz, and G. Pfurtscheller, "ERD/ERS patterns reflecting sensorimotor activation and deactivation," *Progress in Brain Research*, vol. 159, pp. 211–222, 2006.

[6] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clinical Neurophysiology*, vol. 110, no. 11, pp. 1842–1857, 1999.

[7] J. Wang, L. Gao, H. Zhang, and J. Xu, "Adaboost with SVM-based classifier for the classification of brain motor imagery tasks," *Universal Access in Human-Computer Interaction. Users Diversity*, pp. 629–634, 2011.

[8] W. Y. Hsu, "EEG-based motor imagery classification using enhanced active segment selection and adaptive classifier," *Computers in Biology and Medicine*, vol. 41, no. 8, pp. 633–639, 2011.

[9] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K. R. Muller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Processing Magazine*, vol. 25, no. 1, pp. 41–56, 2008.

[10] N. F. nce, A. H. Tewfik, S. Arica, and S. Yagcioglu, "Analysis and visualization of movement related EEG activities using local discriminant bases," in *The Second International IEEE EMBS Conference on Neural Engineering, Washington, DC, USA*, 2005.

[11] Y. Wongsawat, S. Oraintara, and K. R. Rao, "Reduced complexity space-time-frequency model for multi-channel EEG and its applications," in *IEEE International Symposium on Circuits and Systems*. IEEE, 2007, pp. 1305–1308.

[12] K. Nazarpour, S. Sanei, L. Shoker, and J. A. Chambers, "Parallel space-time-frequency decomposition of EEG signals for brain computer interfacing," in *Proceedings of the 14th European Signal Processing Conference (EUSIPCO'06)*, 2006.

[13] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, and K. R. Muller, "Combined optimization of spatial and temporal filters for improving brain-computer interfacing," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 11, pp. 2274–2281, 2006.

[14] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (FBCSP) in brain-computer interface," in *IEEE International Joint Conference on Neural Networks*. IEEE, 2008, pp. 2390–2397.

[15] E. A. Mousavi, J. J. Maller, P. B. Fitzgerald, and B. J. Lithgow, "Wavelet common spatial pattern in asynchronous offline brain computer interfaces," *Biomedical Signal Processing and Control*, vol. 6, no. 2, pp. 121–128, 2010.

[16] C. Brunner, R. Leeb, G. R. Muller-Putz, A. Schlogl, and G. Pfurtscheller, "BCI competition 2008Graz data set a," 2009.

[17] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[18] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[19] L. Parra and P. Sajda, "Blind source separation via generalized eigenvalue decomposition," *The Journal of Machine Learning Research*, vol. 4, pp. 1261–1269, 2003.

[20] Z. J. Koles, "The quantitative extraction and topographic mapping of the abnormal components in the clinical EEG," *Electroencephalography and Clinical Neurophysiology*, vol. 79, no. 6, pp. 440–447, 1991.

[21] B. Blankertz, K. R. Muller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlogl, G. Pfurtscheller, J. R. Millan, M. Schroder, and N. Birbaumer, "The BCI competition III: validating alternative approaches to actual BCI problems," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 153–159, 2006.

[22] K. Fukunaga, *Introduction To Statistical Pattern Recognition*. Academic Press Professional, 1990.

[23] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. wiley New York:, 2001, vol. 2.

[24] J. Cohen, "A coefficient of agreement for nominal scales," *Educational and Psychological Measurement*, vol. 20, no. 1, pp. 37–46, 1960.