Asensio-Cubero, Javier, Gan, John Q. and Palaniappan, Ramaswamy (2013) ***Extracting optimal tempo-spatial features using local discriminant bases and common spatial patterns for brain computer interfacing.*** **Biomedical Signal Processing and Control, 8 (6). pp. 772-778. ISSN 1746-8094.**

Extracting optimal tempo-spatial features using local discriminant bases and common spatial patterns for brain computer interfacing

Javier Asensio-Cubero[a], John Q. Gan[a], Ramaswamy Palaniappan[b]

[a]*University of Essex,Wivenhoe Park, Colchester, Essex CO4 3SQ,United Kingdom*
[b]*University of Wolverhampton, Shifnal Road, Telford, TF2 9NT, United Kingdom*

Abstract

Brain computer interfaces (BCI) provide a new approach to human computer communication, where the control is realised via performing mental tasks such as motor imagery (MI). In this study, we investigate a novel method to automatically segment electroencephalographic (EEG) data within a trial and extract features accordingly in order to improve the performance of classifying MI data. A new local discriminant bases (LDB) algorithm using common spatial patterns (CSP) projection as transform function is proposed for automatic trial segmentation. CSP is also used for feature extraction following trial segmentation. This new technique also allows to obtain a more accurate picture of the most relevant temporal-spatial points in the EEG during the MI. The results are compared with other standard temporal segmentation techniques such as sliding window and LDB based on the local cosine transform (LCT).

Keywords: brain computer interface, motor imagery, local discriminant bases, common spatial patterns.

1. Introduction

Brain computer interfaces (BCI) are communication systems that use human thoughts as control signals to operate machines [1]. These systems are particularly valuable for paralysed users who may not be able to interact with computers in any other manner. From the non-disabled user's point of view, BCIs can enrich human computer interaction where the subject's mental states, such as emotions and error related activity, can be taken into account. Regular users suffering from an induced disability (situations where the user's 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 analysed [3]. Whenever a limb movement is performed, several areas on

*Corresponding author: Javier Asensio-Cubero; Phone Number +447587223814
Email addresses: `jasens@essex.ac.uk` (Javier Asensio-Cubero), `jqgan@essex.ac.uk` (John Q. Gan), `palani@wlv.ac.uk` (Ramaswamy Palaniappan)

the brain cortex are activated due to firing signals from different neuronal populations. Some of these populations show activity even if no real movement is performed at all; just imagining limb movement is sufficient to produce changes of state in the motor cortex [4]. The BCI studied in this paper is based on motor imagery (MI), using EEG as a method for signal acquisition.

Limb movement imagery is characterised by short lasting power amplifications /attenuations along the μ band in EEG, which are known as event 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 researchers 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 researchers try to extract relevant information from the ERD/ERS time course using techniques like local discriminant bases (LDB) [10]. Other 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 spatial pattern (WCSP) [15].

The time span given to the subject for imagining limb movement is denominated a trial, within which 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 whole trial. CSP has been popularly used in the literature for feature extraction due to its ability in locating the active sources while maximising the variance among two or more classes. Usually CSP is applied for trial by trial classification sub-dividing the trial in different frequency bands.

In this paper, we aim to study the benefits of using CSP in different temporal stages of the trial development instead of focusing on different frequency bands. Our motivation is originated from two different grounds. Firstly, we hope to decrease the computational effort derived from the feature extraction when an unseen trial is to be processed. Secondly, unifying the spatial analysis provided by CSP with a temporal study, we will be able to construct a temporal-spatial profile that may be useful for further studies.

In contrast to exhaustive methods such as using sliding windows, LDB finds a subset of signal segments in the temporal domain to build a discriminant base. Traditional approaches [16] use the local cosine transform (LCT) as transform function, which has certain major disadvantages. Firstly, each channel is transformed independently losing the spatial context during the transform. Secondly, the resulting coefficients from a discrete cosine transform need an extra step for feature selection prior to the classification phase.

In this study, we propose a new LDB approach, where, instead of LCT, CSP is used as transform function. This allows us to build up a coherent spatio-temporal analysis tool, which can simultaneously identify the significant temporal segments and spatial locations on the scalp to improve the EEG classification. The covariance decomposition performed by CSP adds insights from the spatial relationships into the LDB algorithm, whilst the features extracted from the transform are directly used for the pattern construction.

This paper is organised as follows: Section 2 explains the methodology: data acquisition (Section 2.1), time segmentation strategies (Section 2.2.1), feature extraction methods in spatial (Section 2.2.2) and temporal domains (Section 2.2.3), classification techniques (Section 2.3), and the experiment set up (Section 2.4). Section 3 describes and discusses the obtained results. Conclusions are drawn in Section 4.

2. Methods

2.1. Data Acquisition

During this study, we used three different data sets, one from the BCI Competition IV (dataset 2a [17]) and two collected in the BCI Laboratory at the University of Essex. They were acquired following a similar protocol.

The first data set 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, recorded from nine 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. Initially, at $t = 0$s, a fixation cross was printed on the screen, after two seconds (at $t = 2$s) an arrow was displayed indicating which imaginary class to perform and this cue was shown until $t = 3.25$s. Although the fixation cross disappeared at $t = 6$s there may be still MI related ERD/ERS patterns after this, i.e., there is an offset period of a couple of seconds. The EEG data was recorded at 250 Hz 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.

The second data set contains three classes from the imagery movement of right hand, left hand, and feet from three subjects. The EEG signals were recorded from five bipolar electrodes (FC3-PC3, FC1-PC1, PZ-FZ,FC2-PC2 and FC3-PC3) and sampled at 250 Hz. The trial structure is the same as the first data set.

The last set of data consists of 15 channels that were recorded at 256 Hz from five different subjects. The electrodes used were those longitudinally positioned from FC3 to FC4 (FC3, FC1, FCZ, FC2 and FC4), from C3 to C4, and from CP3 to CP4. The subjects were asked to perform three different imaginary movements: right hand, left hand and feet.

2.2. Feature Extraction

2.2.1. Time Segmentation

In this study, we explore two different time segmentation strategies: manually partitioning the trial, or relying upon an automatic method such as LDB to segment the signals. The EEG trial is processed from the instant $t = 2$s to $t = 7$s in order to capture information from the trial onset to offset. The EEG segments from the training trials within each partitioning interval are used to build a CSP matrix for extracting signal features that will be used in the later classification step.

1. **Manual segmentation**: Based on a sliding window this approach requires two parameters, the segment length and the overlapping size. Using this technique we evenly cover the trial during different development stages.

2. **Automatic segmentation**: The modified version of LDB described in Section 2.2.3 is used to perform an automatic trial segmentation. In this case, the parameters to set will be the transform technique, cost function, minimum window length and overlapping size. Both LCT and CSP as the transform technique will be investigated. The cost functions applied to evaluate the transformed signal segments are Fisher's criterion [18] and Davies-Bouldin Index (DBI) [19]. These cost functions were chosen because they measure the class separability based on the ratio of the inter-class distance to the intra-class deviation, similar to linear discriminant analysis (LDA), which is used as the classification method. The window length and overlapping size are set to be comparable to the manual segmentation experiments.

2.2.2. Common Spatial Patterns

Methods like principal component analysis (PCA) [20] and independent component analysis (ICA) [21] 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 [22] which maximises the variance between two different classes. As this is the method used in the 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 or a segment a trial. The dimension of X_i will be $N \times T$ as the signal is captured from N different channels.

Every evaluation trial X_i is projected onto the subspace composed of the spatial filters $W \in \mathbb{R}^{N \times N}$, which has been previously computed from the training data as the solution for the generalised eigenvalue decomposition:

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

where $\Sigma^{(+)}$ is the estimated covariance matrix for the trials belonging to class (+) and $\Sigma^{(-)}$ is the covariance matrix for the trials belonging to class (-). A large eigenvalue $\boldsymbol{\lambda}_j^{(+)}$ means that the corresponding vector $\mathbf{w}_j$ leads to high variance in the projected signal in the positive class and low variance in the negative one (and vice-versa).

The CSP projection results in $Y_i = W X_i$. In our experiments, we take the first m rows and last m rows in Y_i, which maximise the variance of one class while minimising the variance of the other class, and compute each feature as $f_l = var(\mathbf{y}_l)$ for $l = \{1, \ldots, m, N-(m-1), \ldots, N\}$, obtaining a total of $L = 2 * m$ features per class. In order to scale down the difference among the feature values we compute the logarithm $f_l^{log} = log(\frac{f_l}{\sum_{j=1}^{L} f_j})$ [23].

2.2.3. Local Discriminant Bases

An LDB algorithm selects a basis of local functions or temporal segments where the most substantial information about a signal can be found. The algorithm is simple and can be used in combination with other spatial approaches [24]. The basic idea is to represent the temporal-spatial information using a binary tree and then prune those branches and nodes which lead to poorer discrimination. Measuring this discrimination power is done through a cost function. Different domains have different requirements in terms of class separation. Typically we can use different distance measures to assess the interclass separability such as Euclidean distance or Fisher's criterion.

The common dyadic tree decomposition of the signal [25] presents an important caveat as the signal is not fully explored, i.e., only those portions enclosed within a node will be transformed and evaluated during the execution. In order to overcome this issue an alternative is proposed in [26] where a merge and divide strategy is adopted covering signal segments that are not visited in the binary tree configuration. This new approach adds adaptability to the LDB algorithm resulting in a more accurate basis.

The LDB algorithm used in this work is described in Algorithm 1. It starts with two children and one mother segments. Initially the segment sizes are defined by the configuration but will change during the execution. It should be noted that as we are handling several channels, the transform within a segment is calculated for each of them, and the cost function is applied taking all the information into account. Although in the BCI domain this approach has focused mainly on LCT [25], other techniques may be applied such as wavelet-packets [27]. We are not interested in the latter as it does not allow the exploration of the signal with variable segment lengths.

In this paper, we look into the possibility of directly using CSP as the transform. In this case we do not focus on the frequency decomposition but in the separability of the features extracted from the selected segment.

An issue that we face when using this version of LDB is that some selected segments have to be removed as they may not contribute to improving the final classification although they perform better than their siblings. However, these segments are easily identifiable as their cost function value is relatively smaller than the rest of the segments which conform to the basis.

2.3. Classification

In this study, LDA is used as the classification method. In spite of its simplicity, this model has proved to achieve results comparable to other approaches such as support vector machines and artificial neural networks [28]. Its main benefit comes from its low computational resource consumption.

A window based majority voting mechanism was used to obtain the class label for each segment, i.e., the predicted label of the current segment will depend on the predicted labels of the previous segments. Specifically, the label for the segment s_i will be $label_{s_i} = mode(\{LDA\_label_{s_{i-k}}\}_{k=0}^{K-1})$ as this has proven to improve the final classification accuracy [29].

Rather than the raw classification accuracy, the Kappa value [30] is used as performance measure as it gives a better picture of the ratio of the classifier accuracy taking into account the per class error distribution. The Kappa value is defined as $\kappa = \frac{p_o - p_c}{1 - p_c}$,

Algorithm 1: Local Discriminant Bases

Set the minimum window size to $winSize$
Set the overlapping size to $overlap$
$Child_1 = [0, winSize + overlap]$
$Child_2 = [winSize - overlap, 2 * winSize + overlap]$
$Mother = [0, 2 * winSize + overlap]$
$bestSegments = List()$
repeat
    Extract features by applying the transform (CSP or LCT) to the three segments, $Child_1$,$Child_2$ and $Mother$.
    Calculate the cost function (Fisher's criterion or DBI) using the features extracted in the previous step.
    if Cost function of any child element is less than the mother's **then**
        $bestSegments.add(Child_1)$
        $Mother = shift(Mother, winSize, overlap)$ {As the child gets the best cost function move the mother segment to the next interval}
        $Child_1 = Child_2$ {Redefine the children segments}
        $Child_2 = shift(Child_2, winSize, overlap)$ {Move the second child to the next interval}
    else
        $bestSegments.add(Mother)$
        $Child_2 = shift(Child_1, winSize, overlap)$ {Move the first child to the next interval}
        $Child_1 = Mother$ {The mother segment takes the role of first child}
        $Mother = expand(Mother, winSize, overlap)$ {Expand the mother segment to include the next interval}
    end if
until The end of the signal is reached

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 (e.g., 0.25 for four classes).

2.4. Experimental Methodology

The experimental methodology used in this paper consists of four steps which will be applied to the available data (Figure 1):

1. Apply the segmentation strategy (sliding window or LDB);

2. Apply CSP to each segment to extract features. This process is repeated with different number of CSP features (i.e. two, four, six and eight);

3. Classify the patterns obtained using LDA (one LDA per segment);

4. The final output for the whole trial will rely on a majority voting among the LDA outputs from the different segments;

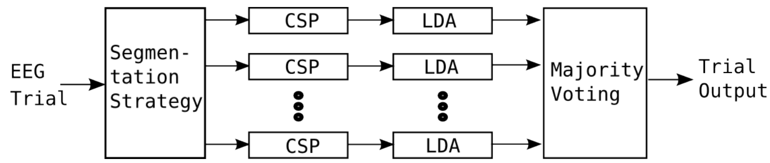

Figure 1: Diagram showing the different steps involved in the experiments. Every EEG trial goes through a segmentation process. These segments are processed using CSP to obtain the final patterns to be classified using an LDA. The output of every LDA is taken into account to give the final trial output.

The EEG data from each subject is divided into two sessions. The first session is used to test every configuration and identify the best parameter configuration using cross validation. The second session is used as evaluation data which leads to the results presented in Section 3.

In order to assure that the results are comparable we have used consistent segment lengths along the experiments which are 125 (0.5s), 150 (0.6s), 200 (0.8s) and 250 (1s) samples.

3. Results

3.1. Classification Performance

In this section, we present the results obtained from the execution of the different proposed strategies in terms of Kappa values. Table 1 displays the Kappa values of the best configurations (segment length and number of features extracted from CSP) obtained from the cross-validation analysis on the training set and applied to the evaluation set.

Ideally we would like to compare our results with the methods proposed in the competition (for subjects one to nine) but the lack of data and methodology description provided thwarts this goal. As a mere piece of information we mention that the two highest ranked proposals obtained 0.57 ± 0.18 and 0.52 ± 0.22 Kappa values respectively, meanwhile the sliding window approach scores 0.57 ± 0.20.

From the presented results, we can observe that LDB obtains better results than sliding window except for the case of the subjects 5 and 14. This indicates that selecting a best basis improves the classification ratio in most cases. It can be seen that LDB+CSP obtains the highest Kappa values on 65% of the subjects whilst LDB+LCT on the 24%; this indicates that replacing LCT with CSP would be an effective approach most of the time.

In order to test the significance of these results we adopted the Wilcoxon rank sum test [31] as it is designed for paired values and it shows a robust behaviour for small populations where the distribution is unknown [32]. From the significance analysis we learn that LDB+CSP+Fisher's criterion obtains a significant higher Kappa performance with $p < 0.05$ when compared with the sliding window approach, although this behaviour is not consistent for the rest of LDB based approaches. When it comes to comparing the performance between LDB+CSP and LDB+LCT, the common spatial pattern based approach performs significantly better ($p = 0.002$). Regarding the cost

Subject	Sliding window	LDB+CSP +Fisher's	LDB+CSP +DBI	LDB+LCT +Fisher's	LDB+LCT+ +DBI
1	0.70	**0.75**	0.63	0.64	0.71
2	0.43	**0.50**	0.46	0.46	0.34
3	0.72	**0.74**	0.72	0.72	0.72
4	0.43	0.40	**0.45**	0.43	0.42
5	**0.24**	0.19	0.22	0.23	0.22
6	0.30	**0.41**	0.38	0.40	0.39
7	0.76	**0.78**	0.74	0.75	0.73
8	0.76	0.72	**0.78**	0.76	0.78
9	0.73	**0.78**	0.70	0.74	0.76
mean (data set 1)	0.57±0.20	0.59±0.19	0.57±0.18	0.58±0.21	0.57±0.18
10	0.72	**0.72**	0.72	0.62	0.65
11	0.56	0.62	0.59	**0.64**	0.55
12	0.86	0.88	0.90	**0.93**	0.88
mean (data set 2)	0.72 ± 0.14	0.75 ± 0.15	0.74 ± 0.15	0.73 ± 0.17	0.70 ± 0.17
13	0.59	0.61	0.61	0.59	**0.63**
14	**0.77**	0.73	0.65	0.70	0.69
15	0.57	0.57	**0.63**	0.48	0.56
16	0.84	0.80	0.84	0.73	**0.86**
17	0.87	0.86	0.84	0.84	**0.89**
mean (data set 3)	0.73 ± 0.13	0.71 ± 0.12	0.71 ± 0.11	0.67 ± 0.14	0.73 ± 0.14

Table 1: Kappa values from best configurations chosen by cross validation on the training set

function there is not significant difference between Fisher's criterion and DBI in terms of Kappa value.

In Table 2, we draw upon the results of different segmentation strategies, detailing the utilised segment lengths. For the LDB based methods, it is noteworthy that with different minimum segment lengths the outcomes are different. Due to the modifications made to the original LDB algorithm, when using a small initial segment we will explore bigger segments along the way, covering higher and lower frequencies more efficiently. From the experimental results, we can state that using the smallest segment size (125 samples), we obtain a statistically significantly better Kappa value than using the largest segments, with $p < 0.05$. If we focus on the overall performance, it is clear that the sliding window strategy leads to more consistent results, regardless of the segment length. This is likely because it uses almost the same segments regardless of the segment length as shown in Table 3.

In terms of the computational load during the evaluation phase sliding window is, in general, the highest resource consuming option, as for every segment a CSP transformation and an LDA classification has to be performed. As shown in Table 3, even using the smallest segment length, the number of segments required by LDB+CSP is about a half of that required by sliding window. Under certain configurations LDB is able to find a trade-off between computing effort and classification performance, obtaining no

Segment length	Sliding window	LDB+CSP +Fisher's	LDB+CSP +DBI	LDB+LCT +Fisher's	LDB+LCT+ +DBI
125	0.60 ± 0.21	0.62 ± 0.24	0.62 ± 0.22	0.58 ± 0.20	0.60 ± 0.21
150	0.60 ± 0.20	0.62 ± 0.24	0.63 ± 0.20	0.56 ± 0.20	0.60 ± 0.21
200	0.61 ± 0.22	0.61 ± 0.21	0.61 ± 0.21	0.55 ± 0.21	0.58 ± 0.21
250	0.61 ± 0.19	0.57 ± 0.23	0.58 ± 0.22	0.55 ± 0.21	0.56 ± 0.20

Table 2: Kappa values related to segment lengths

Segment length	Sliding window	LDB+CSP +Fisher's	LDB+CSP +DBI	LDB+LCT +Fisher's	LDB+LCT+ +DBI
125	23	10.83±4.27	18.81±7.71	16.83±6.43	28.30±3.47
150	23	5.51±2.22	9.24±4.06	10.22±3.12	13.87±2.27
200	22	2.86±1.33	4.41±2.14	5.37±1.36	6.66±1.23
250	21	2.10±0.92	3.13±1.45	4.16±1.10	4.56±1.06

Table 3: Number of segments resulted from different configurations

significantly worse results than sliding window but with much fewer CSP transformations. It is fair to mention that during the calibration process (cross-validation analysis) LDB is computationally heavier due to the nature of this algorithm that explores the different segments, but it is not an issue during the evaluation phase. Therefore, we would suggest choosing a segment length as small as possible so that the algorithm will explore a larger amount of frequencies and temporal locations along the trial, but keeping in mind that the computational load during the validation stage will increase if the segment length is too small.

3.2. Method Analysis and Spatio-Temporal Information

Besides the classification performance, the novel approach of combining CSP and LDB allows us to obtain a detailed picture about which areas contain more information during different periods of time. From the segments obtained via LDB we can directly draw upon the intervals that provide a better class dichotomisation. In Figure 2, we can observe that these segments are relatively associated to their cost function values. Different subjects develop different suitable segments during the LDB process, and different transforms tend to affect the outcomes of different intervals too.

In section 2.2.2, we explained how to obtain the spatial filters W from a set of labelled signals. From these filters, we can compute the corresponding spatial patterns $A = (W)^{-1}$ and project the most relevant components in A onto the scalp as shown in Figure 3 for subject two. This procedure allows us to observe the areas which maximise and minimise the variance for different tasks along the trial development.

Combining both CSP and LDB information, we obtain a detailed picture in domains of time and space, as shown in Figure 4. This representation allows us to easily locate which channels add more useful information during the trial development for a

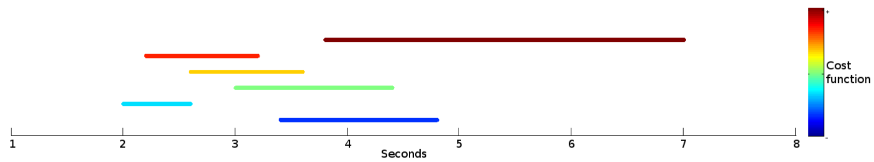

Figure 2: Most significant segments selected using LDB+CSP with Fisher's criterion as cost function for subject 2 (minimum segment size used here is 150 samples). The larger the value from the cost function is, the better the patterns are separable. Here it is possible to appreciate how LDB explores different segment sizes along the trial adapting to the user's data.

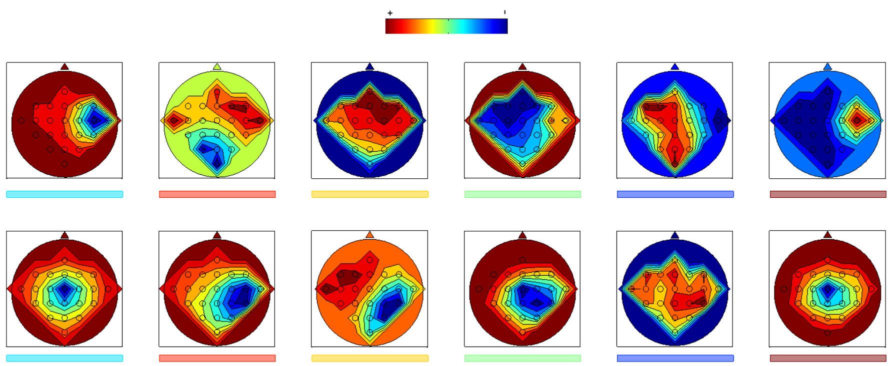

Figure 3: Spatial patterns from the segments shown in Figure 2 (subject 2). The upper row are the patterns which maximise the variance whereas the lower row are the patterns which minimise it. The small rectangle under each pattern is the segment's cost function value.

given class (in this case, right hand movement). The resulting structure could be easily applied in combination with other methods to improve the feature selection or to weight different spatio-temporal locations along the trial.

An example of this temporal-spatial decomposition can be found in Figure 4 *a)*, when the minimum segment size was set to 125 samples. It can be seen that the most meaningful interval is found around 4.1s and most channels contribute significant information. In the same figure, another highlighted segment appears from 3.9s to 4.0s, in which the class-relevant information appears around the channels FC4, C2, C4, C6 and CP4.

4. Conclusions

In terms of classification accuracy measured by the Kappa value, LDB+CSP performs significantly better than sliding window with CSP for feature extraction, and has an additional advantage of consuming less computational resources. However, its classification accuracy depends on the segmentation length. The proposed method achieved better classification performance than the BCI competition winner as well. However, due to the lack of information provided by the BCI competition winner, it is

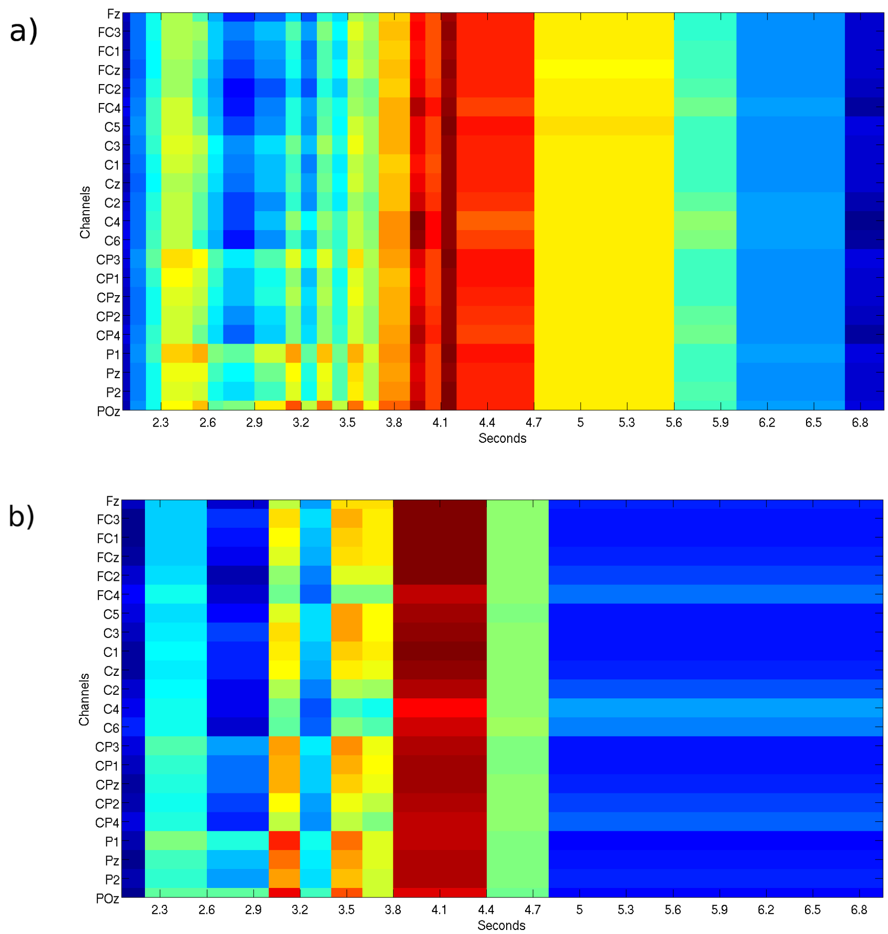

Figure 4: Accumulated spatial patterns along time during trial development for subject 2. Y-axis shows the different channels and X-axis represents the trial duration. *a)* minimum segment size was set to 125 samples; *b)* minimum segment size was set to 150 samples.

difficult to compare to the results obtained in the competition with a statistical significance test.

The combination of CSP and LDB gives detailed information on which instants and channels contain the most significant information for EEG classification during the trial development. We believe that this is an interesting contribution to LDB as the temporal-spatial decomposition can be exploited in order to improve existing methods or develop new ones.

When it comes to the use of LCT and CSP as transform function, the results indicate that CSP performs significantly better than LCT. This is likely due to the fact that during the discovery of the best basis, CSP already takes into account which parts of the signal will contribute better to the classification output.

When comparing automatic and manual partition techniques, LDB+CSP reduces computational resource consumption while obtaining classification performance comparable to the sliding window approach. The number of CSP projections (temporal segments) and LDA classifications involved is reduced by about a half during the evaluation stage, with comparable classification performance.

Acknowledgments

The first author would like to thank the EPSRC for funding his Ph.D. study via an EPSRC DTA award.

References

[1] A. Nijholt, D. Tan, Brain-computer interfacing for intelligent systems, IEEE Intelligent Systems 23 (3) (2008) 72–79.

[2] B. Allison, B. Graimann, 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, R. Paramesran, Current practices in electroencephalogram-based brain-computer interfaces, Encyclopedia of Information Science and Technology, 2nd ed. II (2009) 888–901.

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

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

[6] G. Pfurtscheller, F. H. Lopes da Silva, Event-related EEG/MEG synchronization and desynchronization: Basic principles, Clinical Neurophysiology 110 (11) (1999) 1842–1857.

[7] W. Y. Hsu, EEG-based motor imagery classification using enhanced active segment selection and adaptive classifier, Computers in Biology and Medicine 41 (8) (2011) 633–639.

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

[9] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, K. R. Muller, Optimizing spatial filters for robust EEG single-trial analysis, IEEE Signal Processing Magazine 25 (1) (2008) 41–56.

[10] N. F. Ince, A. H. Tewfik, S. Arica, 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] K. Nazarpour, S. Sanei, L. Shoker, 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.

[12] Y. Wongsawat, S. Oraintara, 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.

[13] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, K. R. Muller, Combined optimization of spatial and temporal filters for improving brain-computer interfacing, IEEE Transactions on Biomedical Engineering 53 (11) (2006) 2274–2281.

[14] K. K. Ang, Z. Y. Chin, H. Zhang, 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, B. J. Lithgow, Wavelet common spatial pattern in asynchronous offline brain computer interfaces, Biomedical Signal Processing and Control 6 (2) (2010) 121–128.

[16] N. F rat Ince, S. Arica, A. Tewfik, Classification of single trial motor imagery EEG recordings with subject adapted non-dyadic arbitrary timefrequency tilings, Journal of Neural Engineering 3 (2006) 235.

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

[18] K. Fukunaga, Introduction to Statistical Pattern Recognition, Academic Press Professional, 1990.

[19] J. C. Bezdek, N. R. Pal, Some new indexes of cluster validity, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 28 (3) (1998) 301–315.

[20] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemometrics and Intelligent Laboratory Systems 2 (1-3) (1987) 37–52.

[21] P. Comon, Independent component analysis, a new concept?, Signal Processing 36 (3) (1994) 287–314.

[22] L. Parra, P. Sajda, Blind source separation via generalized eigenvalue decomposition, The Journal of Machine Learning Research 4 (2003) 1261–1269.

[23] H. Ramoser, J. Muller-Gerking, G. Pfurtscheller, Optimal spatial filtering of single trial EEG during imagined hand movement, IEEE Transactions on Rehabilitation Engineering 8 (4) (2000) 441–446.

[24] N. F. Ince, A. H. Tewfik, S. Arica, Extraction subject-specific motor imagery time-frequency patterns for single trial EEG classification, Computers in Biology and Medicine 37 (4) (2007) 499–508.

[25] M. V. Wickerhauser, Adapted Wavelet Analysis from Theory to Software, AK Peters Ltd, 1994.

[26] R. S. Wu, Y. Wang, New flexible segmentation technique in seismic data compression using local cosine transform, in: Proceedings of SPIE- The International Society for Optical Engineering, Vol. 3813, Citeseer, 1999, pp. 784–794.

[27] B. Yang, G. Yan, R. Yan, T. Wu, Feature extraction for EEG-based braincomputer interfaces by wavelet packet best basis decomposition, Journal of Neural Engineering 3 (2006) 251.

[28] B. Blankertz, K. R. Muller, D. J. Krusienski, G. Schalk, J. R. Wolpaw, A. Schlogl, G. Pfurtscheller, J. R. Millan, M. Schroder, N. Birbaumer, The BCI competition III: Validating alternative approaches to actual BCI problems, IEEE Transactions on Neural Systems and Rehabilitation Engineering 14 (2) (2006) 153–159.

[29] J. Asensio-Cubero, R. Palaniappan, J. Gan, A study on temporal segmentation strategies for extracting common spatial patterns for brain computer interfacing, in: Proceedings of the 11th UK Workshop on Computational Intelligence, Vol. 1, Manchester, 2011, pp. 99–102.

[30] J. Cohen, A coefficient of agreement for nominal scales, Educational and Psychological Measurement 20 (1) (1960) 37–46.

[31] F. Wilcoxon, Individual comparisons by ranking methods, Biometrics Bulletin 1 (6) (1945) 80–83.

[32] P. D. Bridge, S. S. Sawilowsky, Increasing physicians awareness of the impact of statistics on research outcomes: comparative power of the t-test and wilcoxon rank-sum test in small samples applied research, Journal of Clinical Epidemiology 52 (3) (1999) 229–235.