# Extracting common spatial patterns based on wavelet lifting for brain computer interface design

Javier Asensio-Cubero
School of Computer Science
and Electronic Engineering
University of Essex
United Kingdom
Email: jasens@essex.ac.uk

John Q. Gan
School of Computer Science
and Electronic Engineering
University of Essex
United Kingdom
Email: jqgan@essex.ac.uk

Ramaswamy Palaniappan
Dept of Engineering
School of Technology
University of Wolverhampton
United Kingdom
Email: palani@wlv.ac.uk

***Abstract*—Brain computer interfacing (BCI) offers the possibility to interact with machines uniquely relying on the user's thoughts. Although wavelet analysis has been used in the BCI field there is evidence that standard wavelet families, such as Daubechies, may not be the optimal approach. In this study, we developed a novel wavelet lifting scheme, specifically for BCI design. The lifting transform in this new approach is based on common spatial patterns (CSP), which allows to exploit the signal characteristics in temporal, spectral and spatial domains simultaneously. Experimental results show that in BCI applications the new wavelet outperforms several first generation wavelet families in terms of classification accuracy and resource consumption.**

## I. Introduction

Brain computer interfaces (BCI) allow the user's thoughts to be utilised to operate a computer or a machine directly [1]. These kind of interfaces present great advantages to disabled users who are not able to interact with a computer relying on common input devices. Regular uses also benefit from this human computer interaction paradigm, specially under circumstances where the user attention is compromised such as during surgery or piloting a plane [2].

In this study, we focused on limb movement imagery based BCI, usually referred to as motor imagery (MI). Event related desynchronisation (ERD) and event related synchronisation (ERS) [3] are short lasting amplitude amplifications/attenuations present in the electroencephalographic (EEG) data during the MI development. Our aim is to analyse these events extracting the maximum amount of useful information from temporal, spectral and spatial domains.

The time span given to the subject for imagining limb moment is called a trial and its duration may vary from four to eight seconds depending on the experiment. Common spatial patterns (CSP) [4] is widely used in MI analysis as it is able to locate the active sources while maximising the variance among two or more classes. Usually CSP is applied for trial-by-trial classification.

Wavelets, or the first generation wavelets (FGWs), offer a multiresolution analysis framework [5]. The basic idea behind wavelet analysis is the projection of the original signal onto the children orthogonal subspace at different coarse levels. Intuitively the signal is represented using small blocks (shiftings and dilations of the wavelet function $\psi$ at different resolutions), obtaining a more compact representation in both time and frequency domains. FGWs need a specific $\psi$ function that is suitable to the domain of application and, therefore, different wavelet families are designed for different applications. Constructing an ad-hoc wavelet family for specific needs is far from easy, this is the reason why researchers commonly use well studied families such as Daubechies or Coifflets.

The wavelet lifting scheme, known as the second generation wavelets, offers a new approach to multiresolution analysis where the steps to build new wavelets are more straightforward. Only two functions or filters have to be designed: the predict $P(x[n])$ and update $U(d[n])$ functions. By following a few simple steps, we can obtain a lossless multiresolution analysis tool. Other inherent advantages of the lifting scheme include the low resource consumption, as it can be calculated in-place; they can even be applied where the Fourier transform does not exist (FGWs rely on the Fourier transform for their design) such as unevenly sampled data or functions defined over surfaces and spheres. FGWs can also be implemented within the lifting scheme framework [6].

Though the lifting scheme offers an attractive alternative, there has not been real efforts to adapt them for use in the BCI domain. We can only find some studies that use the lifting scheme as a mere implementation approach, using the filters of FGWs like Daubechies [7] [8], which is much faster than the standard discrete wavelet transform algorithm. This low resource consumption is highly valuable in the BCI field, specially during online evaluation where a real time response is required.

Based on the authors' knowledge, for the first time, in this paper we introduce a novel lifting transform based on CSP, which is particularly suitable for handling BCI data. The lifting scheme decomposes the signal in both temporal and frequency domains, and CSP is used to enhance the spatial and temporal information from those electrodes and time sequences that contribute to better classification performance. These two steps are computed simultaneously obtaining a supervised multiresolution analysis decomposition. Once the EEG data is decomposed using either the FGW or the lifting scheme, CSP is applied again as feature extraction method. CSP has proven to give satisfactory performance in BCI studies. It has

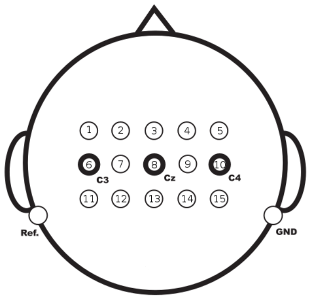


Fig. 1. Using 15 electrodes placed according to the 10-20 international standard

been discussed thoroughly in the literature and it is consistent with the proposed lifting scheme here.

The paper is organised as follows. The data structure and preprocessing is detailed in Section II-A, Section II-B describes the wavelet lifting scheme design and Section II-C focuses on the feature extraction technique and pattern description. The classification and postprocessing methods are covered in Section II-D. The results obtained along with discussions and conclusions are presented in Section III.

## II. Methods

### A. Data Acquisition and Preprocessing

The data used for this study was obtained from the BCI Competition IV (dataset 2a [9]), which was made publicly available. The data contains four different classes: imaginary movement of right hand, left hand, feet and tongue, from nine different subjects. The subjects sat facing a computer screen with 22 electrodes placed on the scalp following the international 10-20 electrode location system. A fixation cross was shown on the screen at $t = 0s$. After two seconds $(t = 2)$ a cue was displayed indicating which imaginary movement to perform. The cue was removed from display at $t = 3.25s$. The fixation cross disappeared at $t = 6s$ indicating 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 and 288 trials (72 for each class) were acquired per session. The first session dataset was used as training data while the second session was used for evaluation in our experiments.

For our experiments, we analysed two different classes: right hand and tongue movement, and a subset of the available electrodes was chosen, which are FC1, FC2, FC3, FC4, FCz, C1, C2, C3, C4, Cz, CP1, CP2, CP3, CP4 and CPz, totalling 15 electrodes (as shown in Figure 1).

For each trial, we only considered those samples from $t = 2s$ to $t = 7s$ (1250 samples per trial). A sliding window covering 375 samples was applied over EEG signals with an overlap of 50 samples; this allowed us to obtain better grained temporal/spatial information from every trial. A Hamming window was applied to reduce the effects of the boundary points. The resulting dataset consists of 1296 labeled segments per class from the total of 144 available trials.

### B. Common Spatial Pattern Based Wavelet Lifting Scheme Design

Orthogonal approaches are very useful for signal representation, but they also have the drawback that requires a strong theoretical background to develop the best base for the given problem. The lifting scheme simplifies this issue, offering a flexible yet potent framework to build new orthogonal basis with multiresolution capabilities [6].

The lifting scheme consists of the iteration of three basic operators [10]:

- **Split:** Divide the original data into two subsets. These usually are the odd and even indexed elements, $x_o[n]$ and $x_e[n]$.
- **Predict:** Obtain the wavelet coefficients as the error of predicting $x_o[n]$ in base of $x_e[n]$ using the *predict operator* $\mathcal{P}$.

$$d[n] = x_o[n] - \mathcal{P}(x_e[n]) \tag{1}$$

- **Update:** Calculate the coarse approximation of the original signal $c[n]$ by combining $x_e[n]$ and $d[n]$ using the *update operator* $\mathcal{U}$.

$$c[n] = x_e[n] + \mathcal{U}(d[n]) \tag{2}$$

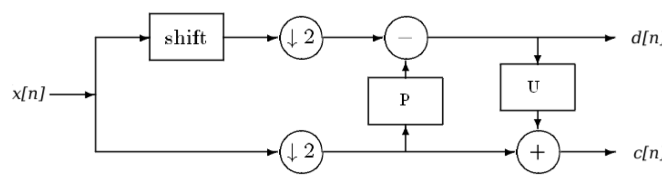


Fig. 2. The wavelet lifting scheme as a block diagram

The strength of the lifting scheme stems from the element selection (choosing even or odd elements is just an example), and specially the prediction and update function design [6]. The lifting scheme also allows obtaining the first generation wavelets thereby reducing the processing time as compared to the steps using discrete wavelet transform.

The proposed method in this paper consists of a CSP based lifting scheme that would filter out the signal features that do not contribute to the signal classification while enhancing those that lead to better classification. CSP is usually extracted using the generalised eigenvalue decomposition generating a set of spatial filters that separate those sequences on the temporal domain which maximise and minimise the variance among two different classes. The outcome of this decomposition is a set of spatial filters $W$ [4].

Let us assume that $X \in Z \times N$ is a segment of EEG data of $N$ samples and $Z$ channels to be transformed using the CSP based lifting scheme. Equations 1 and 2 are usually applied to one dimensional signals but we can reformulate them as matrix operations [11]:

$$D = X_o - P \tag{3}$$
$$C = X_e + U \tag{4}$$

where $X$ is divided into odd and even submatrices in the level of decomposition $l$, $X_e = X_{z,j}$ with $j = \{1 + m * 2^l\}_{m=0}^{M=N/2^l}$ and $X_o = X_{z,k}$ with $k = \{(1 + 2^{l-1}) + m * 2^l\}_{m=0}^{M=N/2^l}$. $P$ and $U$ are the predict and update transformation matrices. We propose to design the lifting operators as follows: $P = \tilde{W}' \times (\tilde{W} \times X_e)$ and $U = \hat{W}' \times (\hat{W} \times D)$, where $\tilde{W}$ is $W$ without the $J$ less discriminant $\mathbf{w}_j$ vectors and $\hat{W}$ is $W$ with only the $I$ most discriminant $\mathbf{w}_i$ vectors. $I$ and $J$ can be determined by trial and error, both are set to three in the current study.

By using the proposed method, during the multiresolution analysis process, those temporal parts of the signal which contribute less to the signal discrimination will be subtracted during the predict phase, whereas in the update part only those elements which actually lead to better discrimination are enhanced when $U$ is applied. This procedure differs from other lifting schemes where the even set is used to predict the odd set in terms of interpolation error and then this error is used in the update step. The newly proposed $P$ and $U$ will result in more appropriate coefficients for classification than using the residuals of an interpolation.

In this study, the performance of several families of FGWs (Haar; Daubechies 3 and 5; Coifflets 1, 3 and 5) have been compared with the CSP based lifting scheme. Every segment obtained from the process described in Section II-A is analysed to the 7th level of decomposition, obtaining 14 different sets of detail coefficients.

### C. Feature Extraction

The number of wavelet coefficients obtained from the multiresolution analysis described in Section II-B is excessive for the classification step without applying some sort of dimensionality reduction technique. In order to be consistent with the novel lifting scheme we apply CSP to the approximation and detail sets.

The projection of the detail coefficients $D_{i,l}$ from the transformed EEG segment $X_i$ into the spatial pattern subspace $W_l^D$ was computed following $Y_{i,l}^D = W_l^{D'} \times D_{i,l}$ for every level of decomposition. This procedure was applied analogously to the approximation coefficients $C_{i,l}$. For clarity, we will refer to $Y_{i,l}^D$ and $Y_{i,l}^C$ using $\bar{Y}$.

For every $\bar{Y}$, we extracted the first $m$ rows and last $m$ rows (those which maximised and minimised the variance between the two different classes) and computed every feature as $f_k = var(\bar{\mathbf{y}}_k)$ with $k = \{1, 2, \ldots, m, Z - m, Z - (m-1), \ldots, Z\}$, obtaining a total of $F = 2 * m$ features. In order to scale down the difference among the feature values, the logarithm $f_k^{log} = log(\frac{f_k}{\sum_{j=1}^{F} f_j})$ was computed.

For this study, $m$ is set to three, and therefore, we obtain a total of 84 features for every EEG segment after the multiresolution analysis and the feature extraction.

TABLE I
KAPPA VALUES FOR THE SLIDING WINDOW APPROACH

| Sub | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | mean |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0.93 | 0.31 | 0.91 | 0.52 | 0.29 | 0.20 | 0.51 | 0.63 | 0.94 | $0.58 \pm 0.28$ |

### D. Classification

In our experiments, linear discriminant analysis (LDA) was used as the classification method. Despite its simplicity, this model has proved to obtain similar classification accuracy to other approaches such as support vector machines and artificial neural networks [12]. The main benefit from LDA is its low computational requirements, being much faster than the other mentioned methods.

Once trained, this model was applied to classify every segment in the validation set. In order to get the final classification output for a single trial we applied a voting window, such that the classification of a segment will depend on $K$ previous segment outputs within the same trial. Thus, the label for the instant $t_i$ will be $label_{t_i} = mode(\{LDA\_label_{t_{i-k}}\}_{k=0}^{K})$, as this has proven to improve the final classification accuracy [13].

Instead of using the raw classification accuracy, we calculated the Kappa value [14]. This performance measure gives a better picture of the ratio of the classifier accuracy taking into account the per class error distribution. The Kappa value was calculated 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 (0.5 for two classes).

## III. Results and Conclusions

In Table I we show the results obtained from applying the sliding window method without performing the multiresolution analysis. The mean Kappa value over the nine subjects was relatively high enough as a benchmark for comparison (0.79 in terms of accuracy) as it is much higher than chance (random) classification.

Using wavelet analysis leads to an increase in the performance as shown in Table II. Obviously, the filtering process inherent in the multiresolution analysis helped to extract more significant features from the signal at different frequency bands. It is noteworthy that the wavelet family which performed better was the Haar wavelet, although this family is usually delegated to theoretical discussion in the wavelet literature as it is considered far too simple for real world applications. As we mentioned in the introduction, FGWs may not be suitable for the BCI field as they are not specifically designed to cope with this problem, the results presented here supported this notion.

Table II also gives us information on how the Kappa value behaves for different subjects. Subjects giving high performances (such as subject 1, subject 3 and subject 9) had consistent output irrespective of the wavelet family, whereas other subjects, such as subject 2, showed large differences (up to 0.39) in their Kappa value depending on the wavelet design.

TABLE II
KAPPA VALUES FOR HAAR, DAUBECHIES AND COIFFLETS FAMILIES

| Sub | haar | db3 | db5 | coif1 | coif3 | coif5 | mean |
|---|---|---|---|---|---|---|---|
| 1 | 0.93 | 0.94 | 0.93 | 0.90 | 0.90 | 0.94 | $0.92 \pm 0.01$ |
| 2 | 0.40 | 0.48 | 0.27 | 0.36 | 0.19 | 0.29 | $0.33 \pm 0.10$ |
| 3 | 0.91 | 0.93 | 0.93 | 0.93 | 0.94 | 0.93 | $0.93 \pm 0.01$ |
| 4 | 0.59 | 0.59 | 0.59 | 0.59 | 0.53 | 0.46 | $0.56 \pm 0.05$ |
| 5 | 0.43 | 0.29 | 0.40 | 0.31 | 0.45 | 0.44 | $0.39 \pm 0.06$ |
| 6 | 0.41 | 0.34 | 0.25 | 0.34 | 0.25 | 0.27 | $0.31 \pm 0.06$ |
| 7 | 0.61 | 0.59 | 0.66 | 0.59 | 0.55 | 0.55 | $0.59 \pm 0.04$ |
| 8 | 0.59 | 0.52 | 0.45 | 0.50 | 0.37 | 0.45 | $0.48 \pm 0.07$ |
| 9 | 0.94 | 0.97 | 0.97 | 0.97 | 0.98 | 0.97 | $0.96 \pm 0.01$ |
| mean | **0.64** $\pm$ **0.22** | 0.63 $\pm$ 0.25 | 0.60 $\pm$ 0.28 | 0.61 $\pm$ 0.26 | 0.57 $\pm$ 0.29 | 0.59 $\pm$ 0.28 | |

TABLE III
KAPPA VALUES FOR CSP LIFTING SCHEME

| Sub | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | mean |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0.90 | 0.31 | 0.88 | 0.59 | 0.45 | 0.41 | 0.69 | 0.63 | 0.95 | $0.65 \pm 0.22$ |

The results obtained by the CSP based wavelet lifting scheme are displayed in Table III. From a Wilcoxon rank-sum test we can conclude that the CSP lifting scheme is not significantly better than Haar or Db3 ($p < 0.05$) but, on the other hand, the improvement is significant when compared to db5 and the Coifflets family. Regarding the performance obtained by individual subjects, it is noted that the lifting approach obtained worse results from the best scoring subjects than FGWs. It is fair to mention that this method was able to raise the accuracy for the not so accurate subjects such as subject 7 and subject 8.

Regarding the computational performance, an average Haar wavelet decomposition would take $2311.2s$ (for the 77760 signals involved for a single user), whereas the lifting approach processes the same amount of data in $54.9s$, in spite of being a supervised method.

We can conclude that using a tailored lifting scheme leads to an improvement in terms of Kappa value when compared to Coifflets and High order Daubechies. It is interesting though that Haar family obtained comparable results to our method despite the fact that it is usually neglected in wavelet related studies in the BCI field.

One important outcome of this research is laying the foundations for a previously unexplored topic in signal processing related to BCI. The future work along the effort of this study will concentrate on deeper analysis of the coefficients obtained from the CSP based lifting scheme with new adaptive strategies for dividing the data in even and odd sets and investigating other paradigms of supervised and unsupervised lifting scheme designs.

## ACKNOWLEDGMENT

The first author would like to thank the EPSRC for funding his Ph.D. study 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," *Proceedings of BRAINPLAY, playing with your brain*, 2007.

[3] 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.

[4] 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.

[5] I. Daubechies, *Ten Lectures On Wavelets*. Society for industrial and applied mathematics, 2006.

[6] W. Sweldens, "The lifting scheme: A construction of second generation wavelets," *SIAM Journal on Mathematical Analysis*, vol. 29, no. 2, p. 511, 1998.

[7] M. Murugappan, M. Rizon, R. Nagarajan, S. Yaacob, I. Zunaidi, and D. Hazry, "Lifting scheme for human emotion recognition using EEG," in *Proceedings of IEEE International Symposium on Information Technology*, vol. 2, 2008, pp. 1–7.

[8] B. Graimann, J. E. Huggins, S. P. Levine, and G. Pfurtscheller, "Toward a direct brain interface based on human subdural recordings and wavelet-packet analysis," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 954–962, 2004.

[9] C. Brunner, R. Leeb, G. R. Muller-Putz, A. Schlogl, and G. Pfurtscheller, "BCI competition 2008 - Graz data set A," *http://www.bbci.de/competition/iv/desc_2a.pdf*, 2008.

[10] R. L. Claypoole Jr, R. G. Baraniuk, and R. D. Nowak, "Adaptive wavelet transforms via lifting," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 3, 1998, pp. 1513–1516.

[11] E. Martinez-Enriquez and A. Ortega, "Lifting transforms on graphs for video coding," in *Proceedings of Data Compression Conference (DCC)*, 2011, pp. 73–82.

[12] 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.

[13] J. Asensio-Cubero, R. Palaniappan, and J. Q. 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*, Machester, 2011, pp. 99–102.

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