Multiresolution analysis of an information based EEG graph representation for motor imagery brain computer interfaces

Keywords: Multiresolution analysis, EEG data graph representation, motor imagery, brain computer interfacing, wavelet lifting, mutual information

Abstract: Brain computer interfaces are control systems that allow the interaction with electronic devices by analysing the user's brain activity. The analysis of brain signals, more concretely, electroencephalographic data, represents a big challenge due to its noisy and low amplitude nature. Many researchers in the field have applied wavelet transform in order to leverage the signal analysis benefiting from its temporal and spectral capabilities. In this study we make use of the so-called second generation wavelets to extract features from temporal, spatial and spectral domains. The complete multiresolution analysis operates over an enhanced graph representation of motor imaginary trials, which uses per-subject knowledge to optimise the spatial links among the electrodes and to improve the filter design. As a result we obtain a novel method that significantly improves the performance of classifying different imaginary limb movements without compromising the low computational resources used by lifting transform over graphs.

1 INTRODUCTION

The analysis of brain signals applied to the operation of computer devices defines a human-machine interaction paradigm known as brain-computer interfacing (BCI) (Dornhege, 2007). This kind of interfaces not only benefit the historically targeted group of disabled users, who may not have at their disposal any other mechanisms of interaction with their surroundings, but also mainstream users (Allison et al., 2008).

In this study we will focus on the use of a tailored wavelet to extract Motor Imagery (MI) related information from electroencephalographic (EEG) data. The imagination of limb movements produces a series of short lasting amplifications and attenuations in the EEG data known as event related desynchronisation (ERD) and event related synchronisation (ERS) (Pfurtscheller and Lopes da Silva, 1999).

The study of ERS/ERD has proven to be a hard task. EEG data is noisy and of low amplitude, there is no inter-subject pattern consistency, and features that make the ERS/ERD patterns recognisable appear at different time intervals, different scalp locations and different frequency bands.

Wavelets have been profusely applied in the BCI domain as they allow a meaningful temporal-spectral analysis of the EEG data. Shifts and dilations of a mother wavelet function provide a series of orthogonal subspaces resulting in what is known as multiresolution analysis (MRA) (Daubechies, 2006). The first generation wavelets present a major disadvantage of difficult design. Commonly, researchers make use of well established wavelet families even though the wavelet function features may not completely fulfil the needs of the domain of study.

Wavelet lifting or second generation wavelets defines a framework that eases the task of developing new wavelet families (Sweldens, 1998) (Sweldens and Schrder, 2000). The lifting scheme consumes less computational resources than the first generation wavelets, and it allows MRA of domains that the first generation wavelets are incapable of.

In (Asensio-Cubero et al., 2013) a new MRA system for BCI data analysis was proposed using lifting scheme over graphs to fully explore the three domains involved in ERS/ERD patterns evolution. Graph EEG data representation is a natural way of describing the spatio/temporal relations among electrode readings. The purpose of this study is to extend the static graph representation by automatically building an enhanced graph in which the connections represent meaningful relationships among different electrodes. For this purpose we used mutual information as it provides a measurement of how much information one channel shares with another channel.

The paper is organised as follows. The data acquisition is detailed in Section 2.1, Section 2.2 explains the lifting scheme over graphs, Section 2.3 describes how the graphs are built, Section 2.4 focuses on the feature extraction technique, pattern description and classification methods, and the experimental method-

ology is described in Section 2.5. The obtained results along with discussions are presented in Section 3. Finally, the conclusions are drawn in Section 4.

2 METHODS

2.1 Data Acquisition and Preprocessing

The first dataset was recorded at the BCI Laboratory at the University of Essex. The protocol was set up as follows: The electrode placement followed the 10-20 international system and 32 channels were recorded with a sampling frequency of 256 Hz. During the recording session the subject was sitting on an armchair in front of a computer screen. A fixation cross was showed at the beginning of the trial at $t = 0s$. At $t = 2s$ a cue was shown indicating the imaginary movement class to perform. The end of the trial was marked when the fixation cross and cue disappeared at $t = 8s$. The subjects were asked to perform 120 trials of each of the three imaginary movements (right hand, left hand and feet). A total of 12 subjects participated in the recording sessions, half of them were naive on the use of BCI systems, 58% of the subjects were female, and the ages ranged from 24 to 50. During the result analysis these subjects were identified by the prefix *E-X*, with *X* being the subject number.

The second dataset is from the BCI Competition IV (dataset 2a) and follows a similar acquisition protocol. The full experiment description can be found in (Brunner et al., 2008). The data covers four different types of MI movement data: right-hand, left-hand, feet and tongue recorded at 250Hz. There are a total of 288 trials recorded, for each of the nine subjects. The subjects belonging to this dataset are identified by the prefix *C-X*.

For this study we utilised a subset of 15 electrodes, covering the major area of the motor cortex (Figure 1). The original data was filtered from 8 to 30 Hz in order to attenuate external noise and artifacts. Each each trial X_i of T samples was scaled by applying $X_i = \frac{1}{\sqrt{T}} X_i^{orig}(I_t - 1_t 1_t^{'})$, where I_t is the $T \times T$ identity matrix and 1_t is a T dimensional vector with ones in it.

The competition data was already divided into an training and evaluation subsets. The data from the University of Essex was split using the first two acquisition runs (180 trials) as training data and the last two runs (180 trials) as evaluation set.

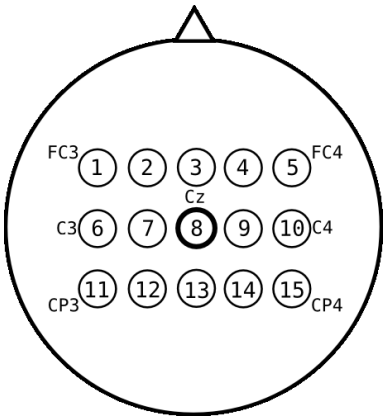

Figure 1: Numbering of the 15 electrodes used during the experimentation, which were allocated from FC3 to FC4, C3 to C4, and CP3 to CP4.

2.2 Wavelet Lifting over EEG Graphs

The first generation wavelets represents signals in terms of shifts and dilations of the basis functions known as mother wavelet. The design of this function obeys a set of restrictions assuring an accurate orthogonal decomposition of the original data. The main benefit of wavelet analysis over other orthogonal systems, such as the Fourier transform or the cosine (or sine) transform, is its multiscale capability. Wavelets allow to analyse the data not only in the frequency domain, but also in the temporal domain at different levels. (Daubechies, 2006) (Mallat, 1989).

The use of the first generation wavelets is pervasive in many domains where signal processing is involved. The BCI field is no exception and we can find their applications in different paradigms such as P-300 (Daubechies (Perseh and Sharafat, 2012)) and MI (Daubuchies, Coifflets and Symlets (Carrera-Leon et al., 2012)).

One major issue to cope with when working with the wavelet transform is that wavelet function design is an extremely complex task, and therefore, researchers apply common families in their studies despite the mother wavelet may not be suitable for the domain of study. The introduction of the second generation wavelets, also known as wavelet lifting (Sweldens, 1996), alleviates this problem making the design of complete multiresolution systems more straight forward. The wavelet lifting is capable of handling data where Fourier analysis is not suitable (and therefore first generation wavelets either) such as unevenly sampled data, surfaces, spheres (Schrder and Sweldens, 1995), trees (Shen and Ortega, 2008) and graphs (Narang and Ortega, 2009) (Martinez-Enriquez and Ortega, 2011).

A lifting scheme consists of iterations of three ba-

sic operations (Claypoole Jr et al., 1998):

- **Split:** Separate the original signal x into two subsets, referred as odd (x_o) and even (x_e) elements.
- **Predict:** The error of predicting x_o in base of x_e using a *predictor operator* $\mathcal{P}$ conforms the wavelet coefficients d.
- **Update:** The coarser approximation of the original signal is calculated by combining x_e and d using an *update operator* $\mathcal{U}$.

A lifting transform over graphs can be defined as follows (Narang and Ortega, 2009). Let us consider a graph $G = (V,E)$ where V is the node set of size $N = N_o + N_e$ and E the edges linking those nodes. V is divided into the even and odd sets and E is represented using the adjacency matrix Adj. We rearrange V and Adj so that the odd set of nodes (a vector V_o of size $N_o \times 1$) is gathered before the even set (a vector V_e of size $N_e \times 1$), obtaining the following structure:

$$\begin{aligned} \tilde{V} &= \begin{pmatrix} V_o \\ V_e \end{pmatrix} \\ \tilde{Adj} &= \begin{pmatrix} F^{N_o \times N_o} & J^{N_o \times N_e} \\ K^{N_e \times N_o} & L^{N_e \times N_e} \end{pmatrix} \end{aligned} \quad (1)$$

The submatrices F and L in $\tilde{Adj}$ in Equation (1) are discarded as they link elements within the same node sets. The block matrices J and K contain only edges linking nodes from different node sets.

The lifting transform functions are defined using a weighted version of the block matrices J and K:

$$\begin{aligned} D &= V_o - J^{\omega} \times V_e \\ A &= V_e + K^{\omega} \times D \end{aligned} \quad (2)$$

where the prediction and update functions are defined as a matrix product: $\mathcal{P} = J^{\omega} \times V_e$ and $\mathcal{U} = K^{\omega} \times D$, where J^{ω} and K^{ω} are the weighted adjacency block matrices and their actual values depend on the domain of application.

We repeat the process described in Equation(2) in each level $l+1$ assigning the approximation coefficients A in level l to V.

2.3 Automatic EEG Graph Building and Filter Design

In (Asensio-Cubero et al., 2013), a static EEG data graph representation was introduced. This representation had the benefit of keeping a channel oriented structure although no extra information was used to optimise the inter-channel links. In order to stablish which channels should be connected for each subject we made use of the mutual information of every pair of channels (Cover and Thomas, 2012)(Peng et al., 2005).

Mutual information measures the amount of information that one random variable Y contains about another random variable Z and is given by:

$$I(Y;Z) = \sum_{y \in Y} \sum_{z \in Z} p(y,z) log \frac{p(y,z)}{p(y)p(z)} \quad (3)$$

where $p(y,z)$ is the joint probability mass function and, $p(y)$ and $p(z)$ is the marginal probability mass function.

Consider a set of MI trials $X^{T \times C}$ of T samples and C channels. In order to stablish the relationships among the spatial locations we compute the mutual information $M(r,s) = I(c_r;c_s)$ for every pair of channels c_r c_s with $r \in \{1 \dots C\}$ and $s \in \{1 \dots C\}$. Note that the diagonal elements of M are set to zero (the mutual information of a channel with itself is ignored) and rest of non-zero elements normalised between zero and one.

The symmetric matrix M describes how all the channels are related to each other and this information can be used to build a specific graph representation for each subject.

Let us assume that the graph $G^x = (V^x, E^x)$ is embedding a trial X, where V^x defines the nodes and the edge set E^x is represented by a weighted adjacency matrix Adj^x:

$$Adj^x_{ij} = \begin{cases} a_{ij} & \text{If } v^x_i \text{ is connected to } v^x_j \\ 0 & \text{Otherwise} \end{cases} \quad (4)$$

For convenience, the odd set will correspond to the elements of X at odd values of t, and the even set at even values of t. Therefore, we obtain two different node vectors v^x_o and v^x_e.

The *predict* and *update* filters are computed in terms of the matrices M and Adj^x. The following steps are carried out in order to set the adjacency matrix values:

1. Apply a threshold th to the matrix M such that $M(r,s) = 0$ if $M(r,s) < th$, so only those channels with high mutual information values will be linked, and normalise the non-zero values between zero and one.
2. Set Adj^x such that for a given channel c and instant t it will be connected to the previous $t-1$ and following $t+1$ time instants with a weight $a_{ij} = 1$.
3. For all the other channels c_r and adjacent temporal values $t+1$ and $t-1$ set the weight $a_{ij} = M(c,c_r)$ in the corresponding entry of Adj^x, if $M(c,c_r) > 0$.

The resulting adjacency submatrices of F^x and L^x from Adj^x are empty. The *predict* and *update* matrices $J^{\omega x}$ and $K^{\omega x}$ (weighted versions the submatrices J^x and K^x) are computed row-wise as $J^{\omega x}_{ij} = \frac{J^x_{ij}}{\sum_{k=0}^{J} J^x_{ik}}$ and $K^{\omega x}_{ij} = \frac{J^x_{ij}}{2*\sum_{k=0}^{J} J^x_{ik}}$. It is noteworthy to mention that the obtained lifting filters are weighted Laplacian graph filters, and the design explained here assures that those channels that share high mutual information will contribute more to the detail coefficients than those that share low mutual information.

2.4 Feature Extraction and Classification

One of the main drawbacks in the use of multiresolution analysis for signal classification is the large number of coefficients generated during the transform. In order to overcome this problem we use common spatial patterns (CSP) as a method for feature extraction and dimensionality reduction.

The different detail D^l and approximation A^l sets at different levels l were projected onto their own CSP subspaces $Y^{D_l} = W^T_{D^l} \times D^l$ and $Y^{A_l} = W^T_{A^l} \times A^l$. For clarity, we will refer to Y^{D_l} and Y^{A_l} using $\bar{Y}$.

For every $\bar{Y}$, we extracted the rows which maximised and minimised the variance between the two different classes (namely, the first m rows and last m rows) and calculated every feature as $f_k = var(\bar{\mathbf{y}}_k)$ with $k = \{1,2,\ldots,m,C-m,C-(m-1),\ldots,C\}$, obtaining a total of $F = 2*m$ features. In order to scale down the difference among the feature values, the logarithm $f^{log}_k = log(\frac{f_k}{\sum_{j=1}^{F} f_j})$ was computed (Ramoser et al., 2000).

For this study, $m \in 2,3,4$ was chosen using cross validation as explained in Section 2.5.

The features obtained from the CSP were classified using LDA as it provides a fair compromise between resource consumption and classification performance (Blankertz et al., 2006).

In order to measure the classification performance Kappa value (Cohen, 1960) was used instead of the classification ratio. Kappa value gives an accurate description of the classifier's performance, taking into account the per class error distribution. The Kappa value was computed as $\kappa = \frac{p_o - p_c}{1-p_c}$, where p_o is the proportion of units on which the judgement agrees (based on the output from the classifier and the actual label), and p_c is the proportion of units on which the agreement is expected by chance.

2.5 Experimental Methodology

After the data preprocessing, a temporal sliding window of one second with a fifth of second overlap was applied over each trial. The segmented data was then transformed using a lifting scheme over graphs (See Section 2.2 and Section 2.3) to the sixth level. The transformation resulted in twelve different coefficient sets, which were further processed to obtain the feature sets by selecting different number of CSP features (See Section 2.4).

The MRA approaches used for comparison were:

- Graph lifting scheme with static graphs (Asensio-Cubero et al., 2013). The static graph is built by linking the elements from the surrounding channels as shown in Figure 2. The filters are calculated analogously as explained in Section 2.3 but by setting the weights of the Laplacian filters to one.

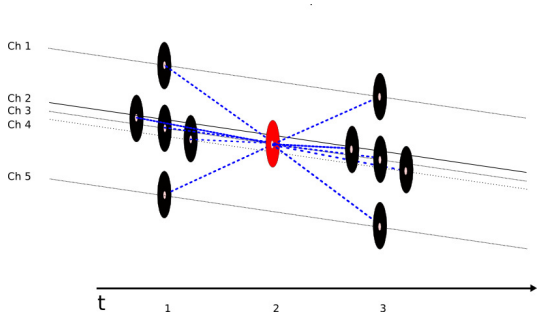

Figure 2: Detail of the graph after the even/odd split for the static approach. The even element (in red) is linked to the surrounding odd elements (in black) adding spatial information to the decomposition.

- Graph lifting scheme and mutual information driven graph building.

Each detail and approximation coefficient sets from the different temporal segments were classified with a separate LDA model after applying CSP. This led to a total of $n_s * l * 2$ LDA outputs, with n_s being the number of segments and l the number of levels. A majority voting approach was carried out in order to obtain the final classification output for each trial.

A cross-validation step using five folds was performed over the training data in order to select the two free parameters involved: the threshold applied to the mutual information matrix in Section 2.3 and the number of CSP features as explained in Section 2.4.

3 RESULTS AND DISCUSSION

From the analysis of the mutual information matrix M for the different subjects we learn that, in general, the standard deviations of the paired calculation do not differ much when compared among classes (two orders of magnitude smaller than the mean). Therefore, instead of computing a matrix M and a different graph to process one class against the others, we just use the whole set of data to generate the mutual information based adjacency matrix. This simplifies the model decreasing the execution time. It is also noteworthy that the performance of the transform calculation does not vary although the values of the filters applied changed.

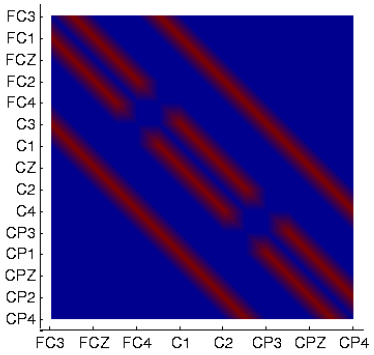

Figure 5: Values of the matrix M for the static graph approach.

Figure 3 and Figure 4 are graphical representations of the values of the matrix M for different users and with different thresholds. In Figure 3 we find two examples to show that there exists a clear correlation between the electrode spatial location and the mutual information, the parallel lines crossing the figure diagonally corresponds to high mutual information values between adjacent electrodes. This correlation is more evident if we compare it with with Figure 5, which corresponds to the matrix M of the static approach. Although in Figure 4 this effect is still noticeable, it is more obvious how, for these specific subjects, the inter-electrode information is more prominent in particular locations of the motor cortex, concretely, in the frontal and central lobes for subject E-7 and eminently central for subject E-8.

After applying the experimental methodology described in Section 2.5 we can analyse the impact of the automatic graph building on the classification results. Figure 6 and Figure 7 show how the median Kappa values change when different threshold values are applied. It is clear that the behaviour of the method is dependant on the subject of analysis. Some

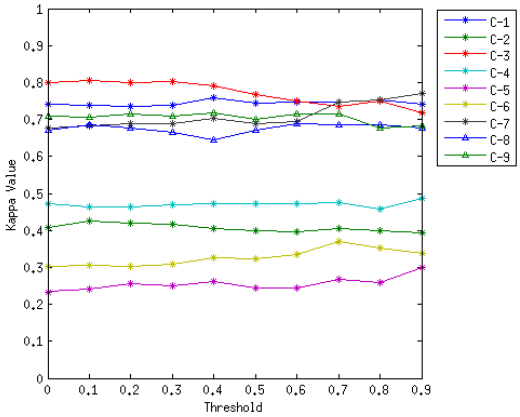

Figure 6: Median of the Kappa value in function of the threshold applied to the matrix M for the competition dataset

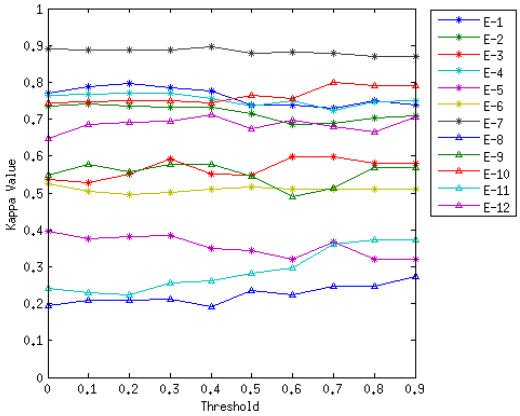

Figure 7: Median of the Kappa value in function of the threshold applied to the matrix M for the Essex dataset

subjects, such as E-8, C-4 and C-2, are not significantly affected by the change of the threshold value, although, on the other hand, we find subjects where the Kappa value fluctuates around 0.1 depending on the threshold value such as in subjects C-3, C-7, E-9 and E-11.

After selecting the two parameters, the threshold and the number of features used in CSP, from the cross-validation results we can compute the classification performance on the evaluation data. Table 1 shows the Kappa values and mean accuracy for the Essex dataset and Table 2 for the competition dataset.

For 80% of the 21 subjects the proposed method achieved a higher Kappa value. For some of the subjects this improvement rises the Kappa value by 0.1 when compared to the static approach. In order to obtain enough data to perform a significance test we

Table 1: Results on the Essex dataset in terms of Kappa values. The mean accuracy is included at the bottom.

Subject	GLS	GLS + Mutual Information
E-1	**0.757**	0.741
E-2	0.736	**0.744**
E-3	0.539	**0.644**
E-4	**0.730**	0.712
E-5	0.392	**0.393**
E-6	**0.529**	0.488
E-7	0.883	**0.891**
E-8	0.210	**0.263**
E-9	0.565	**0.581**
E-10	0.757	**0.774**
E-11	0.237	**0.321**
E-12	0.648	**0.707**
Mean Kappa	0.582 ± 0.21	0.605 ± 0.19
Mean Acc	0.723 ± 0.14	0.737 ± 0.13

Table 2: Results on the competition dataset in terms of Kappa values. The mean accuracy is included at the bottom.

Subject	GLS	GLS + Mutual Information
C-1	0.754	**0.763**
C-2	0.410	**0.419**
C-3	0.800	**0.805**
C-4	0.484	0.475
C-5	0.243	**0.257**
C-6	0.317	**0.364**
C-7	0.629	**0.758**
C-8	0.661	**0.707**
C-9	0.698	**0.721**
Mean Kappa	0.555 ± 0.19	0.586 ± 0.20
Mean Acc	0.666 ± 0.15	0.689 ± 0.15

repeated the experiments by joining the validation and evaluation datasets for each subject and then performing a cross-validation with five folds. Ranksum test validates the hypothesis that the proposed mutual information based approach improves the performance significantly ($p = 0.027$) when compared with the static graph representation.

As a final remark we can mention that the proposed method obtains a Kappa value of 0.586 using the competition dataset, while the winner achieved 0.57. The small number of subjects in the competition data does not allow us to carry out a definitive significance test to compare both approaches.

4 CONCLUSIONS

In this study we have proposed a novel method to improve the EEG data representation based on static graphs by using the mutual information among the different channels. This new strategy for building the graph also has an impact on the filter design, allowing an automatic way to weight the contribution of the different spatial locations.

Comparing the mutual information matrices from different subjects we can observe how the initial static graph approach, where the surrounding electrodes were linked together, was appropriate as close electrodes tend to share similar information.

After applying the proposed methodology on the two data sets Kappa value was increased for an 80% percent of the subjects. This improvement has been proved to be statistically significant, obtaining for several subjects an improvement of 0.1 in their Kappa value.

The present study is also an example of the possibilities offered by the lifting transform, where MRA approaches can be easily implemented without the inherent complexity of the first generation wavelets.

The positive results presented hereby encourage us to explore new ways for optimising the graph representation of EEG data, along with different filter designs, that could eventually lead to more accurate motor imagery data classification.

ACKNOWLEDGEMENTS

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

REFERENCES

Allison, B., Graimann, B., and Grser, A. (2008). Why use a BCI if you are healthy. *Proceedings of BRAINPLAY, playing with your brain*.

Asensio-Cubero, J., Gan, J. Q., and Palaniappan, R. (2013). Multiresolution analysis over simple graphs for brain computer interfaces. *Journal of Neural Engineering*, 10(4):046014.

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

Brunner, C., Leeb, R., Muller-Putz, G. R., Schlogl, A., and Pfurtscheller, G. (2008). BCI competition 2008 - Graz data set A. `http://www.bbci.de/competition/iv/desc_2a.pdf`.

Carrera-Leon, O., Ramirez, J. M., Alarcon-Aquino, V., Baker, M., D'Croz-Baron, D., and Gomez-Gil, P. (2012). A motor imagery bci experiment using wavelet analysis and spatial patterns feature extraction. In *Engineering Applications (WEA), 2012 Workshop on*, pages 1–6. IEEE.

Claypoole Jr, R. L., Baraniuk, R. G., and Nowak, R. D. (1998). Adaptive wavelet transforms via lifting. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 3, pages 1513–1516.

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

Cover, T. M. and Thomas, J. A. (2012). *Elements of Information Theory*. John Wiley & Sons.

Daubechies, I. (2006). *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics.

Dornhege, G. (2007). *Toward Brain-Computer Interfacing*. The MIT Press.

Mallat, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence.*, 11(7):674–693.

Martinez-Enriquez, E. and Ortega, A. (2011). Lifting transforms on graphs for video coding. In *Data Compression Conference*, pages 73–82. IEEE.

Narang, S. K. and Ortega, A. (2009). Lifting based wavelet transforms on graphs. In *Conference of Asia-Pacific Signal and Information Processing Association*, pages 441–444.

Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(8):1226–1238.

Perseh, B. and Sharafat, A. R. (2012). An efficient p300-based bci using wavelet features and ibpso-based channel selection. *Journal of Medical Signals and Sensors*, 2(3):128.

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

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

Schrder, P. and Sweldens, W. (1995). Spherical wavelets: Efficiently representing functions on the sphere. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, pages 161–172. ACM.

Shen, G. and Ortega, A. (2008). Comopact image representation using wavelet lifting along arbitrary trees. In *15th IEEE International Conference on Image Processing, 2008. ICIP 2008.*, pages 2808–2811. IEEE.

Sweldens, W. (1996). Wavelets and the lifting scheme: A 5 minute tour. *Zeitschrift fur Angewandte Mathematik und Mechanik*, 76(2):41–44.

Sweldens, W. (1998). The lifting scheme: A construction of second generation wavelets. *SIAM Journal on Mathematical Analysis*, 29(2):511.

Sweldens, W. and Schrder, P. (2000). Building your own wavelets at home. *Wavelets in the Geosciences*, pages 72–107.

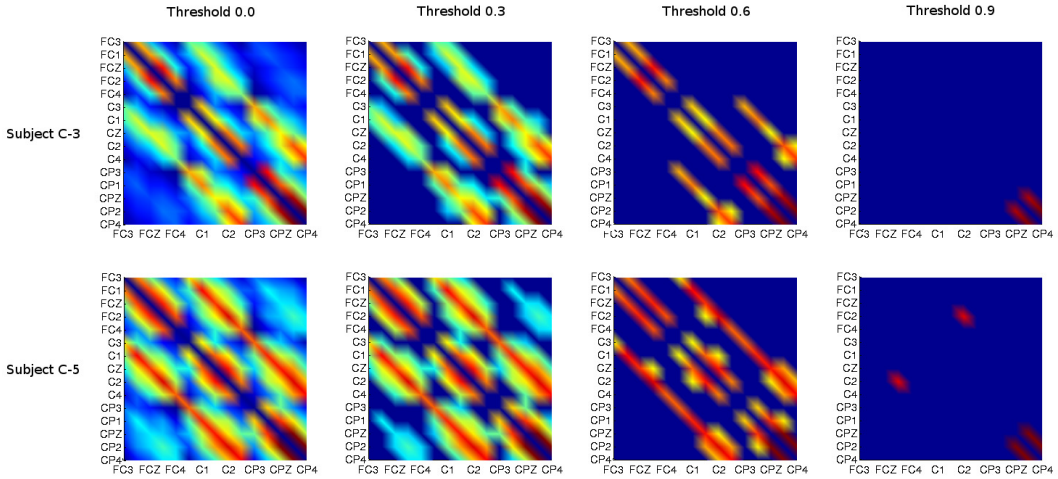

Figure 3: Representation of the values of M for subjects C-3 and C-5 applying different thresholds

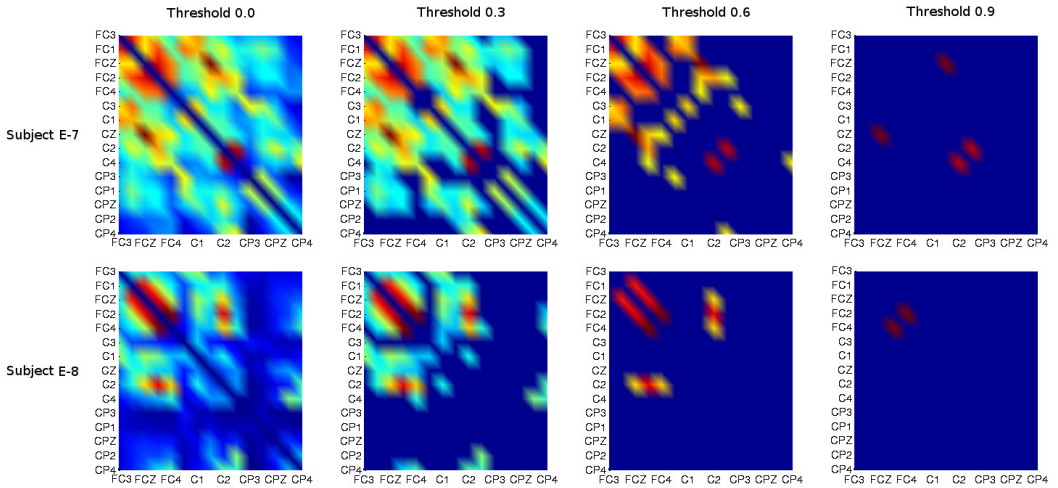

Figure 4: Representation of the values of matrix M for subjects E-7 and E-8 applying different thresholds