Motor Imagery Task Classification Using a New Signal-dependent Orthogonal Transform based Feature Extraction

Mostefa Mesbah[1, 2], Aida Khorshidtalab[3], Hamza Baali[4] and Ahmed Al-Ani[5]

[1] Department of Electrical and Computer Engineering College of Engineering,
Sultan Qaboos University, P O Box: 33, Muscat 123, Sultanate of Oman
[2] School of Computer Science and Software Engineering, The University of Western Australia,
35 Stirling Highway, Perth, WA 6009, Australia
[3] Intelligent Mechatronics System Research Unit, Department of Mechatronics Engineering,
International Islamic University Malaysia (IIUM), Kuala Lumpur, Malaysia
[4] Malaysia Industry Transformation, Technology Park Malaysia, 57000 Kuala Lumpur
[5] Faculty of Eng and IT, University of Technology Sydney, Ultimo, NSW 2007 Australia
`mmesbah@ ieee.org,  ida.khorshidtalab@gmail.com,`
`baaliha@yahoo.fr, Ahmed.Al-Ani@uts.edu.au`

Abstract. In this paper, we present the results of classifying electroencephalographic (EEG) signals into four motor imagery tasks using a new method for feature extraction. This method is based on a signal-dependent orthogonal transform, referred to as LP-SVD, defined as the left singular vectors of the LPC filter impulse response matrix. Using a logistic tree based model classifier, the extracted features are mapped into one of four motor imagery movements, namely left hand, right hand, foot, and tongue. The proposed technique-based classification performance was benchmarked against those based on two widely used linear transform for feature extraction methods, namely discrete cosine transform (DCT) and adaptive autoregressive (AAR). By achieving an accuracy of 67.35%, the LP-SVD based method outperformed the other two by large margins (+25% compared to DCT and +6 % compared to AAR-based methods).

Keywords: Brain-computer interface, channel selection, feature extraction, linear prediction, orthogonal transform

1   INTRODUCTION

The aim of Brain-computer interface (BCI) is to set a direct communication link between the brain and external electronic devices whereby brain signals are translated into useful commands. Such communication link would assist people suffering from severe muscular (motor) disabilities with an alternative means of communication and control that bypass the normal output pathways [1-3]. In this paper, we focus on an important sub-component of BCI systems, namely feature extraction. This sub-

component's aim is to identify a set of features that are effective in discriminating between different classes of interest.

Transform based approaches form an important class of feature extraction techniques. Their aim is to find a more compact lower-dimensional representation in which most of the signal's information is packed in a few number of uncorrelated coefficients. By eliminating irrelevant features (transform coefficients), these methods allow extracting effective features that preserve the generalization capability while lessening the computational complexity associated with the classification stage [4]. These transform-based approaches can be subdivided into linear and nonlinear, supervised and unsupervised, and signal dependent and signal independent methods. The most widely used linear techniques are PCA and LDA. The first one is unsupervised and aims at maximizing the variance of the projected data, using the eigenvectors of the sample covariance matrix, onto a low-dimensional subspace called principal subspace. In contrast, the latter is supervised and attempts to find a linear mapping that maximizes linear class separability of the data in a low-dimensional space [5].

Recently, the authors introduced a signal-dependent linear orthogonal transform, referred to as LP-SVD transform [6]. The transform has the advantage of forming the transformation matrix using only the AR model parameters, instead of the data samples as in the case of PCA. This transform is used in this paper to map EEG data into a new domain where only a few spectral coefficients contain most of the signal's energy. A subset of these transform coefficients, in conjunction with the LP coefficients and the error variance, were used as features in the classification of EEG into four class motor imagery tasks. The feature extraction method was validated using BCI IIIa competition dataset and its classification capability was assessed against two state-of-the-art methods based on DCT and AAR transforms.

The rest of the paper is organized as follows. Section II (*a*) describes the EEG data, its acquisition, and its pre-processing. Section II introduces the LP-SVD transform and how it is used in feature extraction. Section III compares the classification performance of the proposed LP-SVD based technique against two methods based on two of the most widely used linear transform for feature extraction. Section IV concludes the paper.

2 Methodology

2.1 Data Acquisition and Pre-processing

The dataset IIIa from the BCI competition III (2005) [7] was used to evaluate the effectiveness of the proposed feature extraction method. It is a widely used benchmark dataset of multiclass motor imagery tasks recorded from three subjects; referred to as K3b, K6b and L1b. The multichannel EEG signals were recorded using a 64-channel Neuroscan EEG amplifier (Compumedics, Charlotte, North Carolina, USA). Only 60 EEG channels were actually recorded from the scalp of each subject using the 10-20 system and referential montage. The left and right mastoids served as reference and ground respectively. The recorded signal was sampled at 250 Hz and filtered using a bandpass filter with 1 and 50 Hz cut-off frequencies. A notch filter was then applied to suppress the interference originated from power lines. During the experi-

ments, each subject was instructed to perform imagery movements associated with visual cues. Each trial started with an empty black screen at $t = 0$ seconds. At time point t = 2 seconds, a short beep tone was presented and a cross '+' appeared on the screen to raise the subject's attention. At t = 3 seconds, an arrow pointed to one of the four main directions (left, right, upwards or downwards) was presented. Each of the four directions, indicated by this arrow, instructed the subject to imagine one of the following four movements: left hand, right hand, tongue or foot, respectively. The imagination process was performed until the cross disappeared at t = 7 seconds. Each of the four cues was randomly displayed ten times in each run. No feedback was provided to the subject. The recorded dataset from subject K3b consists of 9 runs, while the ones from K6b and L1b consist of 6 runs each, which resulted in 360 trials for subject K3 and 240 trials for each of the other two subjects.

2.2   The LP-SVD Transform

The LP-SVD transform is constructed using a two-step process, namely the estimation of LPC filter coefficients and the computation of the left singular vectors of LPC filter impulse response matrix using singular value decomposition (SVD).

Linear prediction (LP) consists of computing the current signal observation, $y(n)$, using a linear combination of its P past samples, namely, $y(n - i)$ for $i = 1, ..., P$. This can be expressed mathematically by [8]

$$y(n) = -\sum_{i=1}^{P} a_i y(n - i) + e(n), \tag{1}$$

where, a_i are the linear prediction coefficients (LPCs), P is the prediction order and $e(n)$ is the prediction error. Equation (1) can be written in a more compact form using the following matrix notations:

$$\boldsymbol{y} = \boldsymbol{He}, \tag{2}$$

where $\boldsymbol{y} = [y(1), ..., y(N)]^T$ and $\boldsymbol{e} = [e(1), ..., e(N)]^T$ are respectively the $N \times 1$ columns vectors of the data samples and the prediction residual, while $\boldsymbol{H}$ is the $N \times N$ impulse response matrix of the synthesis filter (also called LPC filter) whose entries are completely determined by the linear prediction coefficients a_i. The matrix $\boldsymbol{H}$ is lower triangular and Toeplitz. Applying the SVD to $\boldsymbol{H}$ gives:

$$\boldsymbol{y} = \boldsymbol{UDV}^T\boldsymbol{e} \tag{3}$$

$\boldsymbol{U}$ and $\boldsymbol{V}$ are the $N \times N$ orthogonal matrices containing the left and right eigenvectors of $\boldsymbol{H}$ and $\boldsymbol{D}$ is the $N \times N$ diagonal matrix of singular values [9].

We define the transformation that maps the measurement vector $(\boldsymbol{y})$ to a feature vector $(\boldsymbol{\theta})$ by [6]:

$$\boldsymbol{\theta} = \boldsymbol{U}^T\boldsymbol{y} \tag{4}$$

It is important to note that the transform operation $(\boldsymbol{U}^T\boldsymbol{y})$ by itself does not achieve

any dimensionality reduction. It only decorrelates and packs a large fraction of the signal energy into a relatively few transform coefficients as shown in Fig.1.
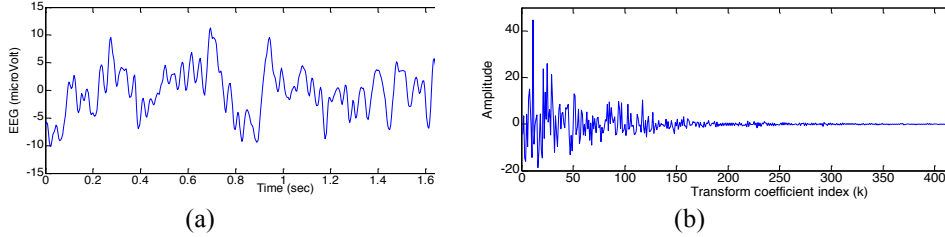

Fig. 1. Signal transformation using LP-SVD: (a) Original EEG signal trace from subject L1b, (b) Transform coefficients with AR(1) as a signal model.

2.3 LP-SVD-based Feature Extraction

Our approach involves extracting features from each EEG segment. These features include the estimated LP coefficients (a_i), the prediction error variance (Vr), and a subset of the most significant transform coefficients $\boldsymbol{\theta}$. These features are described below.

According to the above LP analysis, the EEG vector is described in terms of all-poles filter coefficients and the prediction error. There are two classical approaches used to estimate the LP parameters, namely the autocorrelation and the covariance methods. In this study, we used the autocorrelation method as it guarantees the stability of the filter and allows the efficient Levinson-Durbin recursion to be used to estimate the model parameters [8]. Once the coefficients are estimated, the prediction error sequence can be computed using (1). The estimate of the prediction error $e(n)$ variance is given by:

$$Vr = \frac{1}{N-1} \sum_{n=1}^{N} (e(n) - \bar{e})^2,$$ (5)

where $\bar{e}$ is the arithmetic mean of the prediction error vector $\boldsymbol{e}$ and N is its length.

The data vector $\boldsymbol{y}$ is presented in the new coordinates $\{\boldsymbol{u}_i\}$ by the transform coefficients or scores θ_i. The transform coefficients corresponding to the K largest singular values are selected as features:

$$\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{U}}\boldsymbol{y} \text{ , The columns of } \widehat{\boldsymbol{U}} \text{ are } \{\boldsymbol{u}_1, \boldsymbol{u}_2, \dots, \boldsymbol{u}_K\}.$$ (6)

a. DCT –based feature extraction procedure

The DCT is a signal independent, real-valued, orthogonal transform that is asymptotically equivalent to the optimal principal component analysis (PCA) for highly correlated first-order stationary autoregressive signals [10]. The orthonormal basis vectors $\boldsymbol{w}_k$ of an N points discrete cosine transform (DCT-II) are giving by:

$$w_k = \begin{cases} \frac{1}{\sqrt{N}} (1,1,\dots,1)^T & \text{for } k = 1 \\ \frac{2}{\sqrt{N}} (\cos \frac{k\pi}{2N}, \cos \frac{3k\pi}{2N}, \dots, \cos \frac{(2N-1)k\pi}{2N})^T & \text{for } k = 2,\dots,N \end{cases} \quad (7)$$

The $N \times N$ orthogonal DCT matrix is then defined as $\boldsymbol{W} = (\boldsymbol{w}_1, \dots, \boldsymbol{w}_N)$. It follows immediately that the relation between a data vector $\boldsymbol{y}$ and its DCT transform $\boldsymbol{Y}$ is given by:

$$Y = W^T y \quad (8)$$

The resulting DCT coefficients represented by the vector $\boldsymbol{Y}$ are concentrated in the low-frequency subspace as shown in Fig.2. Dimensionality reduction using DCT is realized by using only these low frequency coefficients as features and discarding the remaining high frequency coefficients. This is illustrated by the following linear mapping.

$$\widehat{Y} = \widehat{W} y , \quad (9)$$

The columns of $\widehat{\boldsymbol{W}}$ are $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \dots, \boldsymbol{w}_K\}$

Fig. 2 shows an exemplary DCT coefficients vector of the EEG data of Fig.1. The energy of the transformed data is packed into the first few low frequency coefficients while all high frequency coefficients are relatively small.
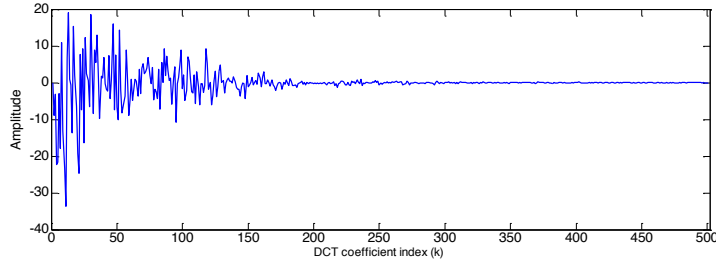

Fig. 2. DCT-II transform of the EEG signal shown in Fig.1(a)

3   EXPERIMENTAL RESULTS AND DISCUSSION

This section is divided into two parts. The first part is devoted to the AR model order selection. The second part evaluates the performance of the LP-SVD-based feature extraction method against two well-known related feature extraction methods. The classifier used to measure the performance is a logistic model tree implemented as part of the Weka software package with its default parameters [11]. This classifier, that uses SimpleLogistic, has a merit over other classifiers due to its use of LogitBoost. To evaluate the classification results, we used 10 fold cross-validation where the data is randomly split into 10 folds of equal size.

3.1 AR model selection

To investigate the appropriate AR model order and the number of transform coefficients to be retained as features, we performed a series of simulations. In this part, only the parameters characterizing the LP-SVD transform are used as features, namely, a subset of transform coefficients ($\widehat{\boldsymbol{\theta}}$), the LP coefficients ($a_i$) and the prediction error variance (Vr). The features were extracted from the electrode sites over the primary motor area C3, CZ, and C4. These are widely considered to be the most informative channels associated with motor imagery tasks [12].

We varied the AR model order from one to seven using the EEG segments from t = 3.5 sec to t = 5.5 sec (501 samples) from each trial. The best model order was selected based on the resulting classification accuracy. This criterion is more suitable, in the present context, than the commonly used one in signal representation (modeling), namely the tradeoff between the model order and the prediction error variance. Table I shows the classification results as function of the order of the AR model.

Table 1. AR Model Order Selection

AR model order	Subject		
	L1b	K3b	K6b
	Accuracy (%)		
1	**42.08**	**66.11**	**47.5**
2	32.5	63.61	28.75
3	29.58	57.5	28.33
4	26.25	58.61	36.25
5	30	60.83	28.33
6	22.5	56.94	24.16
7	20.83	57.5	32.91

For all subjects, the highest classification accuracy, on average, was obtained with first order AR model and using a subset of four transform coefficients with results ranging from 42.08% for subject l1b to 66.11% for subject K3b. Therefore, this model order and number of transform coefficients were used in subsequent analysis.

3.2 Feature extraction evaluation

This part compares the performance of the feature extraction method to those using similar approaches, which are based on signal modeling and orthogonal transform. These techniques are based on adaptive autoregressive (AAR) model [12] and discrete cosine transform (DCT). In particular, Schlögl et al [12] applied a third order adaptive autoregressive (AAR) model for EEG signal analysis. The extracted AAR coefficients, which provide dynamic information about the signal spectrum, served as features. The authors used three different classifiers namely, neural network based on k-nearest neighbour (kNN), support vector machines (SVM), and linear discriminant analysis (LDA) to classify the EEG signal into one of the four classes described earlier. The results showed that the SVM-based classifier achieved the best accuracies followed by LDA and then kNN. The authors also reported that the best results were obtained when using the features extracted from all 60 monopolar channels. In this

evaluation, we used these same channels to provide a fair comparison between the methods.

To find the adequate number of DCT coefficients that achieve the highest classification performance for the different subjects, we varied the number of retained DCT coefficients from 5 to 50 with a step size of 5. Table II summarizes the obtained classification results as a function of the number of retained DCT coefficients. The number of coefficients required, for subjects K6b, L1b and K3b, to achieve the highest classification accuracies were 15, 40, and 20, respectively.

The performances of the three feature extraction approaches mentioned above are summarized in Table III. It can be seen that when only the transform coefficients were used as features, the proposed approach outperformed the DCT-based one by up to 23 % in terms of accuracy (for subject L1b) with 10 times fewer number of features. Meanwhile, when the LP coefficient and the residual error variance were added to the LP-SVD transform coefficients, our technique performed better than the two methods for subjects L1b and K6b and achieved comparable results to the AAR-based method for subject K3b. On average, the improvement, in terms of accuracy was about +25% compared to DCT and +6 % compared to AAR-based methods. It is pertinent to point out that, unlike DCT which results only in the transform coefficients as features, our method results in other features, LPC coefficients and residual signal variance, that led to a better characterization of the signal. In addition, the DCT is signal independent while our proposed transform is signal dependent. These two facts explain the difference in performance between the two methods.

Table 2. Performance (classification accuracy) of DCT-based feature extraction using 60 Monopolar Channels

Number of coefficients retained	Subject		
	K6b	L1b	K3b
5	37.91	38.75	33.61
10	43.75	41.25	36.94
15	**45.833**	40.41	37.77
20	40.00	42.08	**38.05**
25	42.50	42.50	33.88
30	41.25	41.66	35.83
35	40.41	42.91	36.38
40	40.00	**43.75**	33.611
45	39.16	42.50	35.00
50	38.75	39.16	34.44

Table 3. Comparative analysis of different features extraction approaches

Subject	DCT coefficients (best results)	AAR(3) Best SVM results [12]	LP-SVD coefficients (4 coefficients)	LP-SVD coefficients (4)+AR(1)+ Error variance
	ACC %			
L1b	43.75	53.90	52.50	58.75
K3b	38.05	77.20	55.00	76.66
K6b	45.83	52.40	67.08	66.66
Avg	42.54	61.16	58.19	**67.35**

4 Conclusion

In the present study, we presented a feature extraction approach based on the combination of autoregressive modeling and orthogonal transformation. Results of classification experiments, using a benchmark dataset from the BCI competition III, and comparison against closely related approaches, namely DCT and AAR, demonstrates that the proposed feature set is compact and offers a significant improvement in performance as judged by the classification accuracy. The number of transform coefficients was kept constant during all the experiments. It would be interesting to address the issue of parameter tuning in future studies. Future work will also include adding more features to improve the performance beyond the one obtained in this study.

References

1. Lotte, Fabien, Marco Congedo, Anatole Lécuyer, and Fabrice Lamarche. "A review of classification algorithms for EEG-based brain–computer interfaces."Journal of neural engineering 4 (2007)
2. Zander, Thorsten O., Christian Kothe, Sabine Jatzev, and Matti Gaertner. "Enhancing human-computer interaction with input from active and passive brain-computer interfaces." In Brain-computer interfaces, pp. 181-199. Springer London (2010)
3. Saa, Delgado, F. Jaime, and Mujdat Cetin. "Discriminative methods for classification of asynchronous imaginary motor tasks from EEG data." Neural Systems and Rehabilitation Engineering, IEEE Transactions on 21, no. 5: 716-724 (2013)
4. Ozertem, Umut, Deniz Erdogmus, and Robert Jenssen. "Spectral feature projections that maximize Shannon mutual information with class labels."Pattern Recognition 39, no. 7: 1241-1252 (2006).
5. Bishop, Christopher M. Pattern recognition and machine learning. Springer (2006)
6. Baali, Hamza, Rini Akmeliawati, Momoh Jimoh Eyiomika Salami, Aida Khorshidtalab, and Eul-Gyoon Lim. "ECG Parametric Modeling Based on Signal Dependent Orthogonal Transform." Signal Processing Letters, IEEE 21, no. 10: 1293-1297 (2014)
7. Blankertz, Benjamin, Klaus-Robert Müller, Dean J. Krusienski, Gerwin Schalk, Jonathan R. Wolpaw, Alois Schlögl, Gert Pfurtscheller, Jd R. Millan, Michael Schröder, and Niels Birbaumer. "The BCI competition III: Validating alternative approaches to actual BCI problems." Neural Systems and Rehabilitation Engineering, IEEE Transactions on 14, no. 2: 153-159 (2006)
8. Vaidyanathan, P. P. "The theory of linear prediction." Synthesis lectures on signal processing 2, no. 1: 1-184 (2007)
9. Strang, Gilbert. Computational science and engineering. Vol. 1. Wellesley: Wellesley-Cambridge Press, (2007)
10. Ahmed, Nasir, Paul J. Milne, and Stanley G. Harris. "Electrocardiographic data compression via orthogonal transforms." IEEE Transactions on Biomedical Engineering 6, no. BME-22: 484-487 (1975)
11. Hall, Mark, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. "The WEKA data mining software: an update."ACM SIGKDD explorations newsletter 11, no. 1: 10-18. (2009)
12. Schlögl, Alois, Felix Lee, Horst Bischof, and Gert Pfurtscheller. "Characterization of four-class motor imagery EEG data for the BCI-competition 2005." Journal of neural engineering 2, no. 4: L14 (2005)