14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

PARALLEL SPACE-TIME-FREQUENCY DECOMPOSITION OF EEG SIGNALS FOR BRAIN COMPUTER INTERFACING

Kianoush Nazarpour, Saeid Sanei, Leor Shoker, and Jonathon A. Chambers

Centre of Digital Signal Processing, Cardiff School of Engineering, Cardiff University,
Cardiff CF24 3AA, United Kingdom
Emails: {NazarpourK, SaneiS, ShokerL, ChambersJ}@cf.ac.uk

ABSTRACT

The presented paper proposes a hybrid parallel factor analysis-support vector machines (PARAFAC-SVM) method for left and right index imagery movements classification. The spatial-temporal-spectral characteristics of the single trial electroencephalogram (EEG) signal are jointly considered. Within this novel scheme, we develop a parallel EEG space-time-frequency (STF) decomposition in μ band (8-13 Hz) at the preprocessing stage of the BCI system. Using PARAFAC, we elaborate two distinct factors in μ band for each EEG trial. SVM classifier is utilised to classify the spatial distribution of the movement related factor. This factor is distinguished by its spectral, temporal, and spatial distribution.

1. INTRODUCTION

Brain computer interfacing (BCI) based on EEG activities is to enable the people suffering severe neurological disabilities but cognitively intact to operate computers by intention rather than by physical contact. In BCI the disabled body imagines mental tasks and the computer identifies the pattern of input EEG signals. It has been well established that the planning and the execution of voluntary (imagery or real) movements cause pre-movement attenuation and post-movement increase in amplitude of μ rhythm over the sensory motor cortex, mainly over the contralateral hemisphere [1].

For BCI, several EEG processing approaches have been addressed to enhance the correct recognition rate of the mental tasks. Most of these studies rely on the temporal and/or spectral features of the preprocessed EEG signals [2–4]. In [5] the spatial as well as temporal and spectral information have been considered by means of multivariate autoregressive (MVAR) modelling of the multi channel EEG. Approaches based on analysis of joint time, frequency, and space correlations are introduced in [6] where the EEG signals are classified with respect to the correlative time-frequency representations (CTFRs) of different channels.

Generally, existence of non-relevant potentials over the scalp in parallel with the motion related signals restrains the performance of BCI systems. Background activity of the brain, motion and ocular artifacts are of such interferences. In [7] we have contributed to ocular artifacts removal while this work expertly meets the exclusion of the brain background activities in BCI by means of PARAFAC.

In the early paper on the PARAFAC [8] it was used in order to decompose EEG signals. In [9] PARAFAC was reinstated, termed as topographic component analysis and employed to study the event related potentials (ERPs). Topographic time-frequency decomposition of the EEG was also adopted in [10] wherein the atoms were simultaneously characterised by their temporal-spectral and spatial signatures. The authors in [10] extracted physiologically significant activities in the EEG imposing mathematical constraints that in [11] have been figured out unnecessary. It has long been known that unique multi-linear decomposition of multi-way arrays of data is possible using PARAFAC [12]. We demonstrate that PARAFAC is capable of successfully space-time-frequency decomposition of the EEG for BCI and annihilation of the brain background potentials. The inherent uniqueness of the PARAFAC solution leads to single trial EEG decomposition with a minimum a priori assumptions [12] where previous applications of PARAFAC to EEG data have considered only averaged EEG signals.

Studies of medical imaging (PET & fMRI) have established that cortical sensorimotor systems are activated during imagery as well as real motions. It has also been well established that planning and execution of movement leads to a short-lasting and circumscribed amplitude attenuation following by an amplification in the μ rhythm (8-13 Hz) known as event-related (de-)synchronisation (ERD/ERS) [1]. Due to the fact that these brain activities are spatially smeared when volume conducted through the scalp, their exact localization is rather difficult and entails complex computations. Also the clearest ERD/ERS, to be utilised in BCI, may occur at different frequency bands and different time points. Index finger (imagery) movement produces a short-lasting amplitude attenuation (ERD) following by an amplification of the (ERS) μ rhythm, mainly in the contralateral central area [1].

To establish the usefulness of PARAFAC for BCI, we applied the decomposition of time-varying EEG spectrum for differentiating single trial left and right index finger imagination. PARAFAC, is not only able to identify the aforementioned ERD/ERS phenomena, extracts the brain background activity. We, omitting non-relevant factor, utiltised a SVM classifier to distinguish between left and right index movements. The feature set is the spatial distribution of the movement triggered μ rhythm.

This paper is organised as follows. In Section II, the procedure for recording the EEG signals and the preprocessing stages are presented. In Section III and IV, PARAFAC and SVM are briefly reviewed, respectively. The results are subsequently presented in Section V, followed by conclusions in Sections VI.

2. SIGNAL ACQUISITION AND PREPROCESSING

The EEG dataset used in this research has been made available by Dr. A. Osman of University of Pennsylvania for *NIPS2001 BCI Workshop*. EEG signals had been detected from 59 channels placed according to the international 10/20

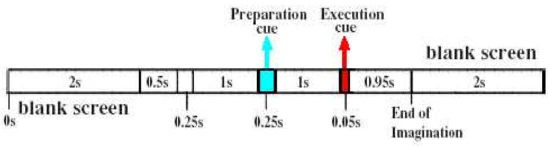

Figure 1: Time sequence of each EEG recording epoch.

system and sampled at the rate of 100 Hz. A fifth-order Butterworth filter was used for temporal bandpass filtering from 5 to 20 Hz, after baseline removal. The subjects were seated in front of a display screen and asked to imagine either left- or right-hand movement imagination for 180 trials - 90 left, 90 right. Each trial starts with one blank screen displayed for 2 s and lasts for 6 s as shown in Fig. 1. Two highly predictable timing cues as preparation and execution cues, were considered in the recording procedure. During the former, which starts at 3.75 s and lasts for 250 ms, a letter "L" or "R" appears on the screen indicating which hand movement should be imagined. The second cue begins at 5.0 s and displays an "X" for 50 ms to instruct the user to start imagination. Since the finger imagination activates the sensorimotor cortex area, the recorded signals from 15 channels, namely, $FC_3, FC_1, FC_z, FC_2, FC_4, C_3, C_1, C_z, C_2, C_4, CP_3, CP_1, CP_z, CP_2$, and CP_4, are considered in this paper.

2.1 Preprocessing

Preprocessing step consists of Laplacian spatial filtering and complex wavelet transform to set up the multi-way array;

2.1.1 Surface Laplacian Filtering

Scalp recorded EEG signals are manifestation of the noisy spatio-temporal superpositions (linear assumption) of electrical activities originating from different brain regions. In order to accentuate localized activity and reduce electrical diffusion in multichannel EEG, we used the spatial filtering technique. Assuming that the distances from a given electrode to its four directional neighboring electrodes are approximately equal, the surface Laplacian filtered EEG signal of channel i at time t, i.e. $EEG_i^{Lap}(t)$, is approximated as follows

$$EEG_i^{Lap}(t) = EEG_i(t) - \frac{1}{4}\sum_{j \in N_i} EEG_j \tag{1}$$

where EEG_i is the scalp EEG signal of the i^{th} channel, and N_i is an index set of the four neighboring channels.

2.1.2 Complex Wavelet Transform (CWT)

Each trial lasts for 6 s, but not all the time points during this period contain beneficial classification information of the different EEG patterns regarding left- and right index imagination. The important factor in BCI is the selection of time window and frequency band over optimal subset of electrodes. In this study instead of manually selection of fixed time-frequency intervals over small number of electrodes, say 2 when only C_3 and C_4 are considered, PARAFAC automatically extracts the time-space-frequency characteristics of EEG signal during MI. To setup a 3-way array, in the present study, Wavelet Transform (WT) is utilised to provide a time-varying representation of the energy of the signal in μ band over the 15 aforementioned electrodes. The complex *Morlet* wavelets $w(t, f_0)$,

$$w(t, f_0) = A\exp(\frac{-t^2}{2\sigma_t^2})\exp(2i\pi f_0 t) \tag{2}$$

with $\sigma_f = 1/2\pi\sigma_t$, and $A = (\sigma_t\sqrt{\pi})^{-1/2}$, is used here in which the trade-off ratio (f_0/σ_f) is 7 to create a wavelet family. The time-varying energy $E(t, f_0)$ of a signal at a specific frequency band is the squared norm of the convolution of a complex wavelet of the signal $EEG^{Lap}(t)$

$$\mathbf{X}^{I\times J\times K} = E(t, f_0) = |w(t, f_0) * EEG^{Lap}(t)|^2 \tag{3}$$

where $EEG^{Lap}(t)$ are the Laplacian-filtered multi channel EEG signals. $\mathbf{X}^{I\times J\times K}$ is a 3-way matrix indexed by I channels, and $J \times K$ of the estimated energy. The time window from 2.75 to 5.75s and the frequency band from 8 to 13Hz are chosen. PARAFAC extracts the underlying factors.

3. PARALLEL FACTOR ANALYSIS

Traditionally, the decomposition of EEG into its constituent components has been based on independent component analysis (ICA) methods. It is not rational to presume that all of the brain sources are mutually independent. Supposing the motion related potentials are synchronous, highly localized, and independent from the background neuronal activities, this assumption led us to exploit PARAFAC. The noteworthy distinction of PARAFAC is that the decomposition of multi-way data with PARAFAC is unique without further orthogonality or independence constraints [12].

Multi-channel EEG data are altered into time-frequency domain by CWT. The increase of dimensionality gives the 2-way array i.e. the matrix of space-time, an extra modality yielding a 3-way array of space-time-frequency. ICA can merely analyse such data by unfolding some modalities into others, reducing the multi-way array into matrices. The unfolding process makes the interpretation of the results doubtful since it removes some specific information endorsed by those modalities. Consequently, rather than unfolding these multi-way arrays into matrices, the data is analysed using the multi-way PARAFAC model. In matrix notation, the factor analysis method is expressed as

$$\mathbf{X}^{I\times J} = \mathbf{A}^{I\times F}\mathbf{S}^{J\times F^T} + \mathbf{E}^{I\times J} \tag{4}$$

where $\mathbf{A}$ are the factor loadings, $\mathbf{S}$ the factor scores, $\mathbf{E}$ the estimation errors, and F is the number of factors. Here, T denotes the transpose operation. Similarly, the PARAFAC of the 3-way array $\mathbf{X}^{I\times J\times K}$ is articulated by unfolding one modality to another as

$$\mathbf{X}^{I\times JK} = \mathbf{A}^{I\times F}(\mathbf{S}^{K\times F}|\odot|\mathbf{D}^{J\times F})^T + \mathbf{E}^{I\times JK} \tag{5}$$

where $\mathbf{D}$ are the factor scores corresponding to the second modality and $\mathbf{S}|\odot|\mathbf{D} = [\mathbf{s}_1 \otimes \mathbf{d}_1, \mathbf{s}_2 \otimes \mathbf{d}_2, \cdots, \mathbf{s}_F \otimes \mathbf{d}_F]$ is the Khatri Rao product. Equivalently, the j^{th} matrix corresponding to the j^{th} slice of the second modality of the 3-way array can be expressed as

$$\mathbf{X}^{I\times j\times K} = \mathbf{A}^{I\times F}\mathbf{D}_j^{F\times F}\mathbf{S}^{K\times F^T} + \mathbf{E}^{I\times j\times K} \tag{6}$$

14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

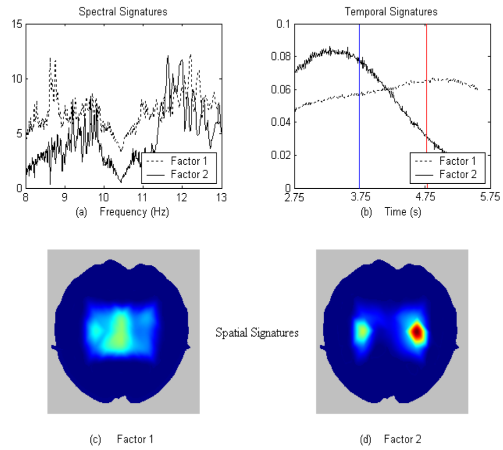

Figure 2: Sample STF decomposition of the 15 channel EEG signal recorded during LEFT index imagination. The factor demonstrated with solid line shows clear ERD in the contralateral hemisphere. (a) Spectral contents of the two identified factors, (b) Temporal profile of the identified factors, the onset of preparation and execution cues are in blue and red, respectively, (c) and (d) Topographic mapping for the two factors.

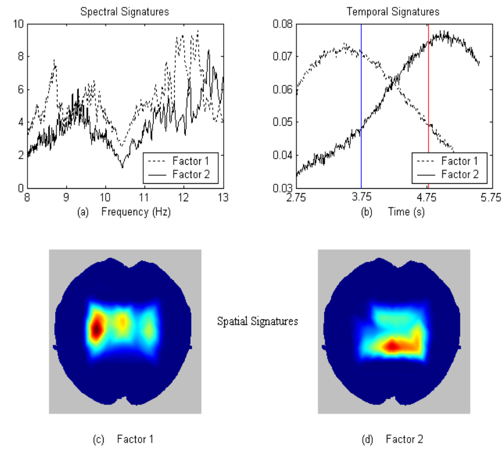

Figure 3: Sample STF decomposition of the 15 channel EEG signal recorded during RIGHT index imagination. The factor demonstrated with dotted line shows clear ERD in the contralateral hemisphere. (a) Spectral contents of the two identified factors, (b) Temporal profile of the identified factors, the onset of preparation and execution cues are in blue and red, respectively, (c) and (d) Topographic mapping for the two factors.

where $\mathbf{D}_j$ is a diagonal matrix having the j^{th} row of $\mathbf{D}$ along the diagonal. The main advantage of PARAFAC over ICA is that uniqueness is ensured under mild conditions, making it unnecessary to impose constraints such as statistical independence. Alternating Least Squares (ALS) is the most common way to estimate the PARAFAC model. In ALS, in order to decompose the tensor to parallel factors a cost function (normally the squared error) is minimised as

$$[\hat{\mathbf{A}}, \hat{\mathbf{S}}, \hat{\mathbf{D}}] = \arg \min_{\mathbf{A},\mathbf{S},\mathbf{D}} \|\mathbf{X}^{I\times JK} - \mathbf{A}(\mathbf{S}| \otimes |\mathbf{D})^T\|^2 \qquad (7)$$

which corresponds to optimising the maximum likelihood of a Gaussian noise model. This is done by alternating between re-estimating each parameter given the estimation of the other parameters. The algorithm can be initialised in several ways, i.e. by randomly defining all parameters and stopping when all parameters have converged [12].

4. CLASSIFICATION

We use the SVM [7] to classify the spatial signatures of the selected atoms. The goal of an SVM is to find an optimal separating hyperplane (OSH) for a given feature set. The OSH is found by solving the following constrained optimisation problem,

$$\begin{array}{ll} & \min_{\mathbf{z},b,\gamma_{i=1,\,2,\,\cdots,\,l}} \left(\frac{1}{2}\|\mathbf{z}\|^2 + C\sum_{i=1}^{l}\gamma_i\right) \\ \text{s.t.} & q_i(\mathbf{z}\cdot\mathbf{g}_i - b) + \gamma_i \geq 0 \qquad i = 1,\, 2,\, \cdots,\, l. \end{array} \qquad (8)$$

where l is the number of training vectors, and $q_i \in \{1, -1\}$ are the output targets, $\|\mathbf{z}\|^2 = \mathbf{z}^T\mathbf{z}$ is the squared Euclidean norm, and $(\cdot)$ is the dot product. The parameter $\mathbf{z}$ determines the orientation of the separating hyperplane, γ_i is the i^{th} positive slack parameter, and $\mathbf{g}_i$ is a vector containing the features $\mathbf{g}_i = [f_1(i) f_2(i) \cdots f_{15}(i)]^T$. The non-negative parameter C is the (misclassification) penalty term and can be considered as the regularisation parameter and is selected by the user. A larger C is equivalent to assigning a higher penalty to the training errors. The optimum value for C is found such that it minimises the cross validation error while yielding proper generalisation performance. Support vectors (SVs) are the points from the data set that fall closest to the separating hyperplane. Any vector that corresponds to a nonzero is an SV of the optimal hyperplane. It is desirable to have the number of SVs small to have a more compact and parsimonious classifier. The OSH (generally nonlinear) is then computed as a decision surface of the form

$$f(\mathbf{g}) = \text{sgn}\left(\sum_{i=1}^{L_s} q_i \alpha_i K(\mathbf{g}_i^s, \mathbf{g}) + b\right) \qquad (9)$$

where $\text{sgn}(\cdot) \in \{\pm 1\}$, $\mathbf{g}_i^s$ are SVs, $K(\mathbf{g}_i^s, \mathbf{g})$ is the nonlinear kernel function (if $K(\mathbf{g}_i^s, \mathbf{g}) = \mathbf{g}_i^s \cdot \mathbf{g}$ the SVM is linear), and L_s is the number of support vectors. A Kernel for a nonlinear SVM projects the samples to a feature space of higher dimension via a nonlinear mapping function. Among nonlinear kernels, the radial- based function (RBF) defined as $K(\mathbf{g}_i^s, \mathbf{g}) = \exp(-|\mathbf{g} - \mathbf{g}_i|^2/2\rho)$, where ρ the adjustable parameter governs the variance of the function, is widely used due to quasi-Gaussian distribution for data sets with a large number of samples. In the following we describe the results of the new PARAFAC-SVM scheme, where the PARAFAC decomposes the EEGs and SVM classifies the extracted features.

5. RESULTS

Two sample results for right and left index imagination are shown in Fig. 2 and Fig. 3, respectively. Fig. 2-(a) shows the spectral content of the two factors identified by PARAFAC in the μ band. Fig. 2-(b) is of great interests where two temporal profiles are illustrated. Note that the blue and red vertical lines indicate onset of "L/R" and "X" cues for preparation and execution, respectively. Fig. 2-(b) shows that even before "X" which occurs at time point "3.75 s," the subject has started imagination. The dotted curves of Fig. 2-(a, b) correspond to Fig. 2-(c), where obviously occurs under C_3. Note that red and blue areas indicate the level of activity of the channels normalised between zero and one (left ear is left, and the nose is up). Eventually, this outcome comes along with previous researches where it is elaborated that an ERD can be recorded on the contralateral hemisphere in μ band [1]. The ERD must not necessarily be followed by an ERS, as the complete ERD/ERS is merely visible in an average over a large number of EEG trials [1]. The other factor, demonstrated by solid line, occurs simultaneously within brain but mostly in ipsilateral hemisphere. This factor shows that EEG signals of μ band have greater amplitude there.

Similarly, following the spectral signatures of the two factors of Fig. 3-(a), in Fig. 3-(b), the two profiles are shown. In Fig. 3, following the same logic as for Fig. 2, of the factors, demonstrated by solid line, indicates a clear ERD in the contralateral hemisphere under C_4 and the other factor, illustrated by dotted line, corresponds to the brain activities mostly in central and ipsilateral areas.

The hybrid PARAFAC-SVM approach after training with 90 trials; 45 for left and 45 for right index imagery movement, was tested using another 90 trials. In order to test the overall classification rate, classification were performed for a number of iterations. Two different kernels namely, Linear and Gaussian RBF, were examined for the SVM. For this dataset the value chosen for the parameter C was empirically found to be 79 and for the case of linear kernel the average number of SVs was found 63% of the training trials and the best achieved classification rate was 68.33%. For the RBF kernel the parameter ρ was set to 1 and C to 10. The optimal values for C and ρ have found using grid search. The average number of support vectors calculated when using the RBF kernel was 57% of the training trials and the best achieved classification rate was 76.22%. High dimensional feature space causes the considerable number of support vectors. Using transforms such as principle component analysis, sufficient number of features can be selected while retaining the discriminatory inormation. Using the above kernels any overfitting can be easily avoided. From Fig. 4, it can be verified that the multidimensional feature space is nonseparable due to the overlapping regions.

6. CONCLUSION

We have presented a robust method for distinguishing between left and right index imagery movements from scalp EEGs using the resulting features of PARAFAC. The potential of PARAFAC to jointly Space-Time-Frequency decompose the time-varying spectrum of multichannel EEG, enables spatially localization of the ERD factor in contralateral hemisphere clearly in parallel with time and frequency. The classification is done by using the SVM. The experiments herein demonstrate the potential of proposed PARAFAC-SVM method to classify single trials EEG signals. Higher classification rates are achieved when the RBF kernel is used in SVM.

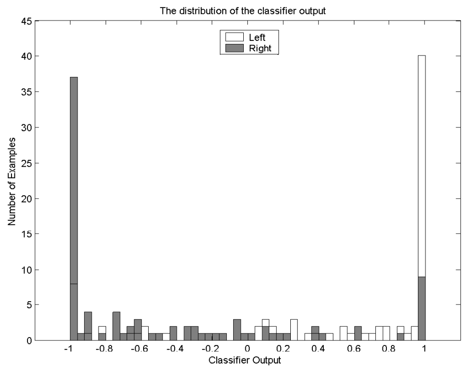

Figure 4: Histogram plot showing the output of the classifier pre $\mathrm{sgn}(\cdot)$, using the linear kernel.

REFERENCES

[1] G. Pfurtscheller and F. H. Lopes da Silva, editors. *Event-Related Desynchronisation*, Handbook of Electroencephalogr. Clin. Neurophysiol., Revised Series, vol. 6. Amsterdam: Elsevier, 1999.

[2] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and C. A. Forneris, "An EEG-based braincomputer interface for cursor control," *J. Electroencephalogr. Clin. Neurophysiol.*, vol. 78, pp. 252-259, 1991.

[3] D. P. Burke, S. P. Kelly, P. de Chazal, R. B. Reilly, and C. Finucane, "A parametric feature extraction and classification strategy for brain-computer interfacing," *IEEE Trans. Neur. Sys. Rehab. Eng.*, vol. 13, no. 1, pp. 12-17, Mar. 2005.

[4] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," in *Proc. IEEE,* vol. 89, no. 7, pp. 1123-1134, July 2001

[5] G. Ginter, K. J. Blinowska, M. Kaminski and P. J. Durka, "Phase and apmplitude analysis in time-frequency space-application to voluntry finger movement," *J. Neuro. Method.*, vol. 110, pp. 113-124, 2001.

[6] T. Ebrahimi, J.-M. Vesin, and G. Garcia, "Brain-computer interface in multimedia communication," *IEEE Sig. Process. Mag.*, vol. 20, no. 1, pp. 14-24, Jan. 2003.

[7] L. Shoker, S. Sanei, and J. Chambers, "Artifact removal from Electroencephalograms using a hybrid BSS-SVM algorithm," *IEEE Sig. Process. Lett.*, vol. 12 no. 10, pp 721-724, 2005.

[8] R. A. Harshman, "Foundation of the PARAFAC procedure: models and conditions for an 'explanatory' multi-modal factor analysis," UCLA Work. Pap. Phon. 16, 1-84, 1970.

[9] J. Moks, "Decomposing event-related potentials: a new topographic components model," *J. Biol. Psychol.* vol. 26, pp. 199-215, 1998.

[10] T. Koenig, F. Marti-Lopez, and P. A. Valdés-Sosa, "Topographic time-frequency decomposition of the EEG," *J. NeuroImage*, vol. 14, pp. 383-390, 2001.

[11] F. Miwakeichi, E. Martinez-Montes, P. A. Valdés-Sosa, N. Nishiyama, H. Mizuhara, and Y. Yamaguchia, "Decomposing EEG data into spacetimefrequency components using Parallel Factor Analysis," *J. NeuroImage*, vol. 22, pp. 1035-1045, 2004.

[12] R. Bro, *Multi-way Analysis in the Food Industry: Models, Algorithms and Applications*, PhD Thesis. University of Amsterdam (NL) and Royal Veterinary and Agricultural University, Matlab toolbox available [Online] at http://www.models.kvl.dk/users/rasmus/.