A BISPECTRUM APPROACH TO FEATURE EXTRACTION FOR A MOTOR IMAGERY BASED BRAIN-COMPUTER INTERFACING SYSTEM

Shahjahan Shahid, Rakesh Kumar Sinha and Girijesh Prasad

Intelligent Systems Research Center, Magee Campus
University of Ulster, Londondery, Northern Ireland, UK
phone: +44 28 71375339, fax: + +44 28 71375417 email: s.shahid@ulster.ac.uk
web: http://isrc.ulster.ac.uk

ABSTRACT

Existing feature extraction techniques for BCI systems are developed based on traditional signal processing techniques assuming that the signal is Gaussian and has linear characteristics. But the motor imagery (MI) related EEG is highly non-Gaussian, non-stationary and non-linear. This paper proposes an advanced, robust but simple feature extraction procedure for MI based BCI system. This novel approach uses higher order statistics technique, the bispectrum, and extracts the non-linear features from EEG. Along with a linear classifier (LDA), the proposed technique has been applied to an MI based BCI system. The performance (classification accuracy, mutual information and Cohens kappa) of the system is evaluated and compared with the power spectrum based BCI. It is observed that the proposed technique extracts more pragmatic information resulting in better and consistent cross-session detection accuracy and Cohens kappa. It is concluded that the bispectrum based feature extraction is a promising technique for detecting different brain states.

1. INTRODUCTION

A brain-computer interfacing (BCI) system establishes a direct communication channel between brain and a control or communication device. This system is useful to the people with total dysfunction of neuromuscular system due to communicational breakdown between brain and spinal cord. The efficiency of a BCI system highly depends on three operations [12]: recording of the cerebral (brain) signal e.g., electroencephalogram (EEG), electrocorticogram (ECoG); extraction of information from the recorded signal; and translation of the extracted information to control a device.

The cerebral electrophysiological processes associated with motor imagery (MI) are reported as a spatiotemporal process [2]. A movement or MI of left or right hand results in a desynchronization of μ-band (8-13 Hz) oscillations in the contralateral EEG along with simultaneous synchronization of central β-bands (18-26 Hz) oscillations in the ipsilateral EEG [8]. Various digital filtering and signal processing techniques are therefore applied to extract features from the MI related EEG signals. Among variety of algorithms, common spatial patterns (CSP) have successfully been used for the pre-processing in BCI [9]. In power spectrum based techniques, the band power value [6] and autoregressive (AR) based feature extraction [10] attracted a lot of attention in BCI. The joint time-frequency and time-scale have also been found well suited for the EEG analysis; hence, wavelet based methods have been used in BCI [5].

Due to highly non-stationary and non-linear nature of MI related EEG signals, these traditional feature extraction methods often fail to provide good performance from one experiment session to another, i.e., these techniques do not provide time invariant separable features. To overcome the nonlinearity effect in BCI, [13] introduced higher order statistics (HOS) along with traditional techniques: the algorithm deals with 12 features where 8 features are extracted by traditional signal processing methods and 4 features are by HOS. Since the algorithm combines traditional features with HOS, it cannot overcome the errors due to non-Gaussianity and nonlinearity in EEG signals. In order to account these factors in more effective way, we propose(in Section 2) a new feature extraction technique comprising of the bispectrum of EEG signals. To evaluate the effectiveness of the proposed technique, we apply it to the Technical University of Graz dataset provided in the BCI-competition IV and classify the features by a linear discriminant analysis (LDA) method. The evaluation results and corresponding discussion are reported in Section 3 and Section 4, respectively.

2. METHOD

An MI related BCI system is normally developed in two phases: training phase and evaluation or application phase. In the training phase, the EEG signals - recorded from motor cortex area while a subject performs the imagination of motor movement, are processed to find the best (on the basis of highest accuracy) parameters for feature extraction and feature classification process. With known class labels during EEG recording, the selection of optimal parameters is made by tuning various parameters related to feature extraction and feature classification subsystem (i.e. a classifier); e.g., signal segmentation and frequency bands for filtering. In evaluation stage, the BCI system uses those optimal parameters and provides communication signal to control a device.

In this study we have developed a BCI system comprising of a new feature extraction technique (described in section 2.1) and a widely used linear discriminant analysis (LDA) classifier: Fishers LDA. The performance of our BCI system has been assessed by three standard statistical measures (see section 2.3).

2.1 Bispectrum based Feature Extraction

A bispectrum of a signal is the expectation of three frequency components: two direct frequency components and the complex conjugate component of the sum of those two frequencies of a random signal [7]. Knowing the Fourier frequency components of a random signal $x(t)[t = 1,2,..]$, its bispectrum, $B_x(k,l)$, can be estimated as [7]

$$B_x(k,l) = E\{X(k)X(l)X^*(k+l)\} \quad (1)$$

where $E\{\bullet\}$ denotes the statistical expectation, $X(\bullet) = \text{FFT}[x(t)]$; k,l are the discrete frequency indices and * denotes the complex conjugate term of $X(\bullet)$. The bispectrum is a complex measurement and, therefore, it has magnitude and phase components. A bispectrum of a random signal provides supplementary information to the power spectrum of that signal. Also a bispectrum of an additive statistically independent multi-source signal is the sum of their individual bispectrums. Like other HOS measurements, the bispectrum of a Gaussian or independent, identically distributed (i.i.d.) signal is theoretically zero. One of popular bispectrum estimation procedures follows the formula given below [7]

$$B_x(k,l) = \frac{1}{N}\sum_{i=1}^{N} X_i(k)X_i(l)X_i^*(k+l) \tag{2}$$

where i is the epoch index and N is the total number of epoch within considered signal duration. Since any system noise is generally assumed as Gaussian and/or i.i.d. signal, the Bx in Eq. (2) is free from its effect. For the BCI system development, the bispectrum was computed using Eq.(2) where the preprocessed (here 8 -14Hz or 14-27Hz band passed) EEG signal was considered as $x(t)$. In order to characterize temporal bispectral information, we compute the sum of absolute log-bispectrum over all bifrequencies in the non-redundant region, θ (i.e., $0 \geq k \geq (f_s/2)$, $l \geq k$, $2k+l \geq f_s$, where f_s is the sampling frequency). Mathematically, the proposed feature vector for a BCI can be written as,

$$B(m) = \sum_{k,l\in\theta} |\log[B_x(k,l)]| \tag{3}$$

where m is a time index related with the time period for which the bispectrum is estimated. The block diagram of the proposed technique is illustrated in Fig. 1

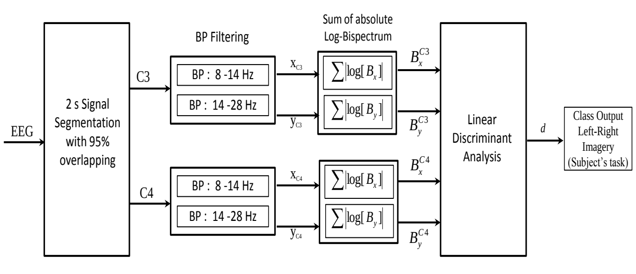

Figure 1: Block diagram of proposed features extraction techniques with linear discriminant analysis classifier.

2.2 Classification Technique: Fisher's LDA Classifier

The extracted features (e.g. $B(m)$) are often not straightforward to classify. In order to find the best combination of features that separates different events/classes, we used one of the most powerful and robust methods, Fishers LDA classifier [3]. Fishers LDA estimates a hyper-plane in the feature space to separate the features into the two different classes. It basically finds the separation between two distributions by the ratio of two group-variances: between-class variance to the within-class variance. The separation between two distributions (separability factor) can be written as [3]

$$S = \frac{\sigma^2_{between}}{\sigma^2_{within}} = \frac{(w\mu_{y=1} - w\mu_{y=0})^2}{w^T\nu_{y=1}w + w^T\nu_{y=0}w} \tag{4}$$

where μ_y is the mean and ν_y the variance of the feature distribution (B) within two classes (i.e., $y=0$ and $y=1$); and w is known as the weight vector. From Eq.(4), it can be shown that the maximum separation occurs when

$$w = (\nu_{y=0} + \nu_{y=1})^{-1}(\mu_{y=0} - \mu_{y=1}) \tag{5}$$

The weight vector w is the normal to the discriminant hyperplan. Fisher's LDA decision plane uses the following representation to classify the feature vector $B(m)$ as [3]

$$d(m) = B(m)w^T + b \tag{6}$$

where b is the bias (threshold). The features are assigned to one class or the other depending on the sign of $d(m)$.

2.3 Evaluation Procedure

2.3.1 *Accuracy*

Two types of accuracy were computed for each time point of MI paradigm (a computer controlled thinking procedure see [1] for more detail). The classifiers output (the sign of $d(m)$) was compared with actual event(left or right)of MI. With a confusion matrix (CM) for each time point of MI the left and right accuracies were computed from the below formulas:

$$\begin{aligned}\text{Left Accuracy} &= \frac{\text{True Negative in CM} \times 100}{\text{Number of Actual Left}}\\ \text{Right Accuracy} &= \frac{\text{True Positive in CM} \times 100}{\text{Number of Actual Right}}\end{aligned} \tag{7}$$

Considering all trials and its actual labels (left and right), average left and right accuracies were computed for each time point of paradigm. The mean of left and right accuracy was called here as overall accuracy.

2.3.2 *Cohen's Kappa Coefficient*

Cohens kappa coefficient (also called Cohens kappa) is a statistical measure which provides an index of inter-rater reliability. It is considered to be an improvement over using the percent of accuracy, as the procedure of computing accuracy (see Eq. (7)) does not involve the false positive or false negative effects. From the definition of Cohens kappa it can be written as [11]

$$\kappa = \frac{P_o - P_c}{1 - P_c} \tag{8}$$

where P_o is the relative observed agreement between raters, and P_c is the hypothetical probability of chance agreement. The maximum possible value of Cohen's kappa is limited to 1 and then the raters are in complete agreement If there is no agreement among the raters (other than what would be expected by chance), $\kappa = 0$. To understand the entire computational procedure of Cohen's kappa, readers can refer [11].

2.3.3 *Mutual Information*

The mutual information of two random variables is a quantity that measures the mutual dependence of the two variables, i.e., how much one variable tells us about another. For two discrete variables X and Y whose joint probability distribution is $P_{XY}(x,y)$, the mutual information between them, denoted $I(X:Y)$, is given by

$$I(X:Y) = \sum_{x\in X}\sum_{y\in Y} P_{XY}(x,y)\log\left(\frac{P_{XY}(x,y)}{P_X(x)P_Y(y)}\right) \tag{9}$$

where $P_X(x)$ and $P_Y(y)$ are the marginal probability distribution function of X and Y respectively: $P_X(x) = \sum_y P_{XY}(x,y)$ and $P_Y(y) = \sum_x P_{XY}(x,y)$.

If two random variables X and Y are independent, their joint probability distribution $P_{XY}(x,y) = P_X(x)P_Y(y)$ and the mutual information $I(X:Y) = 0$. In addition, a high value of mutual information indicates a large reduction in uncertainty.

3. EXPERIMENTAL SETUP AND RESULTS

The proposed feature extraction technique has been devised to a BCI system with the data provided by the Technical University of Graz (TUG) BCI lab as part of the BCI competition IV data-2b. The descriptions of signal and its recording paradigm are reported in [1]. Our BCI system works with only 2 channels (C3 and C4) of EEG signal. For each subject, the system was trained with the EEG from 3rd session (i.e., signal from B0ϕ03T where ϕ is the subject number), and evaluated (or tested) on the EEG from 4th and 5th sessions (B0ϕ04E and B0ϕ05E). Since the MI task occurred after 3rd second, we trained and evaluated the system with the signals from 3s to 8s time course of a trial of the MI paradigm.

In the training phase, we tuned the system for highest accuracy. The tuning parameters were the size of EEG segment for bispectrum computation, number of bandpass filters, the cut-off frequencies of band-pass filter(s)and the parameters for bispectrum estimation. With different combination of tuning parameters a 5-fold cross-validation was carried out to find the possible classification accuracy (see Eq. (7)) by the LDA classifier. In our experiment, we obtained best classification accuracy for all 9 subject if the BCI system contained two band-pass filters (with 8-14Hz and 14-27Hz) and the segmentation of EEG signal is 2s with 0.1s step size. In order to get an optimal LDA classifier (i.e. LDA coefficients) for each subject, we fixed these parameters and re-trained the system with the whole training dataset. Note that the test (evaluate) datasets were kept hidden from the training, cross-validation and optimization steps.

In evaluation phase, we first setup the BCI system with the computational parameters obtained in training phase; and started to extract time-varying signed distance (TSD) signal from evaluation dataset. We assessed this output signal with the actual labels provided in `http://www.bbci.de/competition/iv/results/#labels`. The assessment program we used was developed by TUG BCI team provided in Biosig toolbox (an open source software) and the program also follows the same formulas as in the section 2.3.

The results of this experiment were obtained in two ways: (a) evaluation of bispectrum (*BSP*) based feature extraction in the BCI system; and (b) its performance over the results by a power spectral density (*PSD*) based technique. In designing a *PSD* in a BCI system, the EEG power was computed by AR-Yule method. The PSD based feature extraction technique with LDA classifier was adopted from [4]. We followed the same procedure for training and evaluation phase as we did for *BSP* based BCI system.

Table 1 shows the training and evaluation results obtained by *BSP* and *PSD* based feature extraction methods. The results of training phase reflects the performance of feature extraction method and the used optimal classifier. Note, the feature extraction parameters were fixed for all subjects; but each subject used different classifier which were fixed at their training and evaluation phases.

In order to observe the performance of task classification accuracy (CA), kappa and mutual information over each time point of the MI paradigm, we have plotted these measures in Fig. 2 and Fig. 3 from two subjects B04 and B06 respectively. Each plot illustrates a comparative performance obtained by *BSP* and *PSD* based feature extraction techniques. In addition, each of these plots also displays the measures for training and evaluation phases which uses same classifier.

4. DISCUSSION AND CONCLUSION

As we see in Table 1, the training phase accuracies by the *BSP* technique for the left and right MI are often close to each other, i.e., the LDA finds suitable discriminant hyperplane in the feature space and hence, classifies in a balanced way. But the same measurements are mostly imbalanced with *PSD* technique: e.g., subjects B03, B05, B07 and B09. The average overall-accuracy in *BSP* based training phase is 82% which is 8% higher than that of the *PSD* based training phase. Again, the average mutual information and the kappa values are found higher in *BSP* based training phase. It is also observed that the training phase kappas by *BSP* technique for subject B01, B05, B06, B07 and B09 are remarkably higher than corresponding kappas by the *PSD* based technique. These observations conclude that the technique *BSP* provides consistent and distinct features to the classifier.

Further, considering the training phase kappas by *BSP* it can be concluded that the quality of features from subjects B01, B05 and B06 are moderate as the kappas for these subjects are around 0.60; whereas the features from subject B04 are very much compatible to our BCI system as the kappa is close to 1; and the signals from subjects B02 and B03 are not good for our BCI system as the kappa is less than 0.4 (its mutual information is also close to 0). In fact, the classifier does not find distinct features from the EEG of subjects B02 and B03.

The results of evaluation phase show that both *BSP* and *PSD* based techniques are found highly session sensitive, since they fail to provide balanced left and right accuracies ($< 10\%$) in the evaluation session (see subjects B01, B02, B05, B06, B07 and B09 for *PSD* based technique; and subjects B01, B05, B06, B07 and B09 for *BSP* based technique). But with moderate or good quality signal the kappas by *BSP* technique in evaluation session are found more than 60% and similar to its corresponding training phase kappas. The evaluation kappas by *PSD* based technique are mostly below 40%. It is evident that the *PSD* based approach has much higher extent of uncertainty than the *BPS* based technique.

A similar comparative observation can be made from Fig. 2 and Fig. 3: with *BSP*, the whole distribution of kappa and mutual information along the time course of paradigm is always higher than that of *PSD* based technique. For good signal (e.g., B04), the kappa and mutual information are quite similar in both techniques - the maximum performance is obtained after 5s of the paradigm course and it stays high till the end of the time course. But with moderate signal (from B06) the distribution of accuracies, kappa and mutual information by the *BSP* based technique remain reasonably higher after its maximum value - this behavior was not observed in the distributions by *PSD* based technique. Therefor, the signal quality independent consistent feature distribution empowered the robustness of *BSP* technique.

To see how much distinct the features are by *BSP* and

Table 1: Training and evaluation results (maximum measurements of accuracy, mutual information and kappa) of 9 subjects analyzed by *BSP* and *PSD* methods respectively. This result is achieved with the optimum classifiers' parameters: frequency bands 8-14Hz (μ-band) and 14-27Hz (β-band); optimal LDA coefficient obtained in the training phase for each subject.

Subject	Bispectrum						Power spectrum density					
	Training Stage			Evaluation Stage			Training Stage			Evaluation Stage		
	Overall Accuracy (Left - Right) Max (in %)	Max. Mutual Info	Max. Kappa	Overall Accuracy (Left - Right) Max (in %)	Max. Mutual Info	Max. Kappa	Overall Accuracy (Left - Right) Max (in %)	Max. Mutual Info	Max. Kappa	Overall Accuracy (Left - Right) Max (in %)	Max. Mutual Info	Max. Kappa
B01	80 (81 - 79)	0.34	0.60	71 (96 - 46)	0.20	0.43	64 (61 - 68)	0.08	0.27	57 (42 - 72)	0.00	0.15
B02	66 (66 - 65)	0.07	0.31	68 (65 - 71)	0.05	0.36	65 (66 - 64)	0.02	0.19	59 (67 - 52)	0.01	0.18
B03	65 (65 - 61)	0.04	0.30	59 (59 - 60)	0.03	0.19	63 (56 - 70)	0.03	0.27	56 (53 - 59)	0.02	0.11
B04	99 (99 - 99)	1.22	0.98	97 (97 - 97)	1.04	0.95	96 (96 - 96)	0.87	0.93	93 (91 - 95)	0.77	0.86
B05	83 (82 - 84)	0.40	0.66	81 (86 - 76)	0.30	0.63	72 (60 - 84)	0.06	0.44	88 (94 - 81)	0.56	0.75
B06	81 (82 - 79)	0.36	0.61	83 (90 - 76)	0.47	0.66	71 (70 - 72)	0.15	0.43	69 (86 - 51)	0.10	0.38
B07	88 (93 - 84)	0.46	0.75	79 (72 - 86)	0.31	0.59	78 (82 - 72)	0.23	0.55	68 (81 - 54)	0.05	0.35
B08	90 (93 - 88)	0.62	0.80	95 (97 - 93)	0.92	0.90	85 (84 - 86)	0.50	0.71	90 (81 - 99)	0.67	0.80
B09	88 (86 - 90)	0.58	0.76	88 (82 - 94)	0.63	0.76	68 (72 - 63)	0.07	0.34	59 (49 - 70)	0.01	0.19
Average	82 (83 - 81)	0.454	0.641	80 (83 - 76)	0.439	0.607	74 (72 - 75)	0. 223	0.459	71 (71 - 70)	0.243	0.418

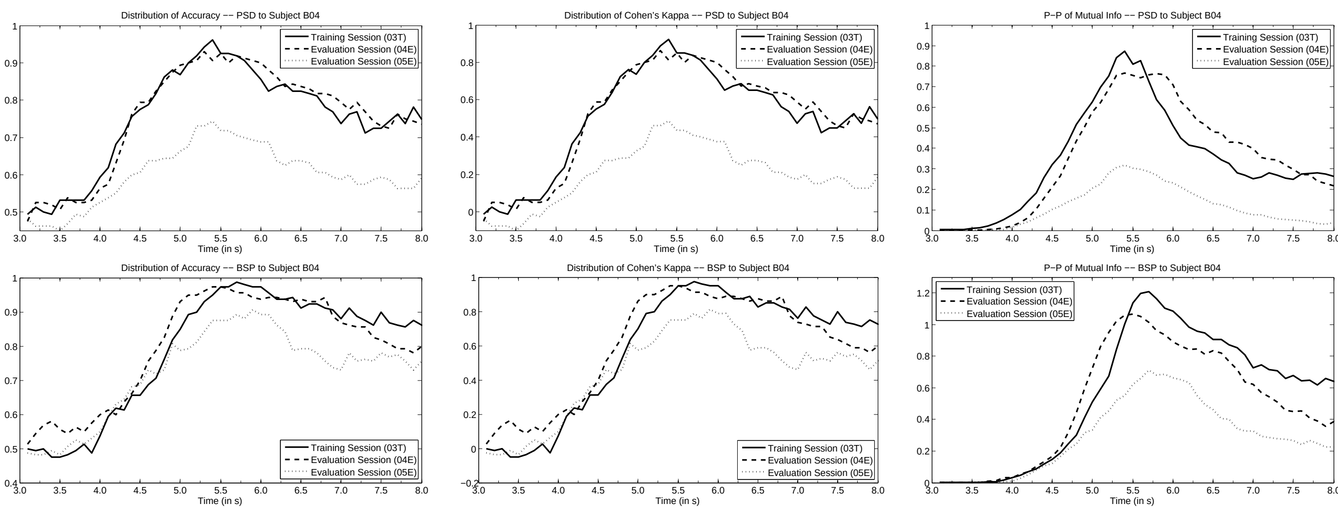

Figure 2: The values of accuracy (left), Cohens kappa (middle) and mutual information (right) along the time course of the paradigm for the subject B04. These were observed with *PSD* (top) and *BSP* (bottom) features extraction techniques.

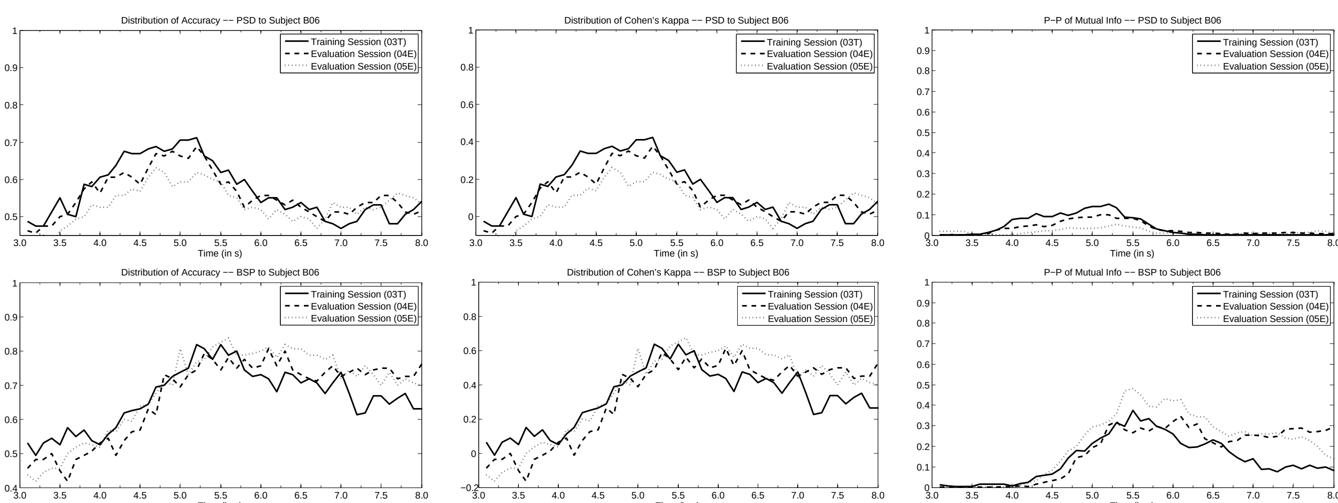

Figure 3: The values of accuracy (left), Cohens kappa (middle) and mutual information (right) along the time course of the paradigm for the subject B06. These were observed with *PSD* (top) and *BSP* (bottom) features extraction techniques.

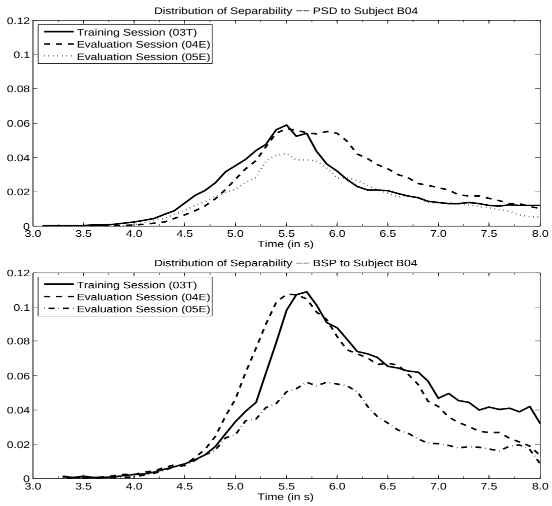

Figure 4: Point to point distribution of Separability in subject B04. The separability factor is computed from *PSD* and *BSP*.

PSD, we computed separabilities for the subject B04 using Eq. (4) The separability plots (Fig. 4) demonstrate that the features by *BSP* based technique are more distinct than that of *PSD* based technique and the distinctness of features are observable till the end of paradigm. Due to its inherent properties with non-Gaussian and nonlinear signals, the *BSP* thus can extract more separable and distinct features for the BCI.

The same datasets were also processed and classified by the different renowned researchers (as competitors of the BCI Competition IV 2b-dataset). We have compared the performance of our proposed *BSP* based BCI with their achievements. The average result just crosses the winners achievements: our average kappa in the evaluation phase is 0.607, while it was 0.60 by the competition winner. Note, there is a major difference in feature extraction procedure between competitors and us: we have chosen common optimal parameters for feature extraction for all subjects but the competitors used subject specific parameters. In order to cope with the fixed frequency bands for all subjects, we had to sacrifice some good feature extraction results, which would increase the average performance of *BSP* based technique; e.g., with a pair of band-pass filters (21-25Hz and 25 -29Hz) to subject B05, it was possible to get higher accuracy (91%); kappa (0.82) and mutual information (0.66) in the evaluation phase. However, this optimal parameter is not good for all other subjects.

The novelty of this paper is the development of a new feature extraction method and its implementation to the BCI system. The feature extraction technique estimates the bispectral power from MI related EEG which deals with the measurements due to non-Gaussian and nonlinear behavior of EEG whereas the traditional technique assumes the EEG as the output of a linear system. Further, the proposed BCI system uses two band-pass filters by which LDA gets the bispectral information of ERD and ERS oscillation in the EEG. Finally, we advise to use the *BSP* technique with real-time MI based EEG signal: a PC with 3.33GHz Core2Duo CPU, it takes about 0.05s to extract features from 1s EEG signal.

Acknowledgment

Dr. Shahjahan Shahid is supported by InvestNI and the Northern Ireland Integrated Development Fund under the Centre of Excellence in Intelligent Systems project and Dr. Rakesh Kumar Sinha is supported by Leverhulme Visiting Fellowship scheme.

REFERENCES

[1] http://www.bbci.de/competition/iv/desc_2b.pdf.

[2] A. B. Barreto, A. M. Taberner, and L. M. Vicente. Classification of spatio-temporal EEG readiness potentials towards the development of a brain-computer interface. In *Proc. IEEE Southeastcon '96*, pages 99–102, 1996.

[3] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley-Interscience Publication, 2000.

[4] P. Herman, G. Prasad, T. M. McGinnity, and D. Coyle. Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification. *IEEE Trans Neural Syst Rehabil Eng*, 16:317–326, 2008.

[5] S. Lemm, C. Schfer, and G. Curio. BCI competition 2003–data set III: probabilistic modeling of sensorimotor mu rhythms for classification of imaginary hand movements. *IEEE Trans Biomed Eng*, 51:1077–1080, 2004.

[6] Y. Li, X. Gao, H. Liu, and S. Gao. Classification of single-trial electroencephalogram during finger movement. *IEEE Trans Biomed Eng*, 51:1019–1025, 2004.

[7] C. L. Nikias and M. R. Raghuveer. Bispectrum estimation: A digital signal processing framework. *Proceedings of the IEEE*, 75:869–891, 1987.

[8] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer. EEG-based discrimination between imagination of right and left hand movement. *Electroencephalogr Clin Neurophysiol*, 103:642–651, 1997.

[9] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans Rehabil Eng*, 8:441–446, 2000.

[10] A. Schloegl, K. Lugger, and G. Pfurtscheller. Using adaptive autoregressive parameters for a brain-computer-interface experiment. *in Proc IEEE Engineering in Medicine and Biology*, 4:1533–1535, 1997.

[11] J. Sim and C. Wright. The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy*, 85:257–268, 2005.

[12] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan. Brain-computer interfaces for communication and control. *Clin Neurophysiol*, 113:767–791, 2002.

[13] S.-M. Zhou, J. Gan, and F. Sepulveda. Classifying mental tasks based on features of higher-order statistics from EEG signals in brain-computer interface. *Inf Sci*, 178:1629–1640, 2008.