Biomedical Signal Processing and Control 38 (2017) 55–66

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Research Paper

Characterization of phase space trajectories for Brain-Computer Interface

Khaled Sayed[a], Mahmoud Kamel[b], Mohammed Alhaddad[b], Hussein M. Malibary[b], Yasser M. Kadah[c,*]

[a] Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, USA
[b] Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia
[c] Electrical and Computer Engineering Department, King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

Article history:
Received 27 August 2016
Received in revised form 4 May 2017
Accepted 15 May 2017

Keywords:
Brain-Computer Interface (BCI)
Electroencephalogram (EEG)
Distance series (DS)
Moment invariants
Phase space reconstruction (PSR)

ABSTRACT

A new processing framework that allows detailed characterization of the nonlinear dynamics of EEG signals at real-time rates is proposed. In this framework, the phase space trajectory is reconstructed and the underlying dynamics of the brain at different mental states are identified by analyzing the shape of this trajectory. Two sets of features based on affine-invariant moments and distance series transform allow robust estimation of the properties of the phase space trajectory while maintaining real-time performance. We describe the methodological details and practical implementation of the new framework and perform experimental verification using datasets from BCI competitions II and IV. The results showed excellent performance for using the new features as compared to competition winners and recent research on the same datasets providing best results in Graz2003 dataset and outperforming competition winner in 6 out of 9 subject in Graz2008 dataset. Furthermore, the computation times needed with the new methods were confirmed to permit real-time processing. The combination of more detailed description of the nonlinear dynamics of EEG and meeting online processing goals by the new methods offers great potential for several time-critical BCI applications such as prosthetic arm control or mental state monitoring for safety.

1. Introduction

Brain computer interface (BCI) is an alternative communication pathway between the human brain and the external environment through a computer [1]. BCI systems are used to assist disabled people to control neuroprostheses and wheelchairs by detecting their brain electrical activity during different mental tasks [2–4]. Different techniques can be used to measure brain activity such as the electroencephalography (EEG), functional magnetic resonance imaging (fMRI), electrocorticography (ECoG), magnetoencephalography (MEG), and near infrared spectroscopy (NIRS) [5–8]. Among those techniques, the EEG-based BCI systems are the most widely used due to their relatively low cost, high temporal resolution, and convenience for users [9].

The human brain shows EEG activity over the sensorimotor cortex at mu (8–13 Hz) and beta (14–25 Hz) frequency bands when awake subject does not experience sensory or motor activities and this phenomenon is termed event-related synchronization (ERS) [10]. In contrast, the mu and beta rhythmic activities are attenuated when a subject processes motor commands or sensory stimuli and this phenomenon is termed as event-related desynchronization (ERD) [11]. Fortunately, such ERS/ERD changes can be elicited during the imagination of movements and hence can be used for EEG-based BCI systems operated by motor imagery [12,13].

Different feature extraction and classification algorithms were developed to interpret the brain electrical activity into commands for external computers or devices. Krusienski et al. studied the relative BCI performance using Phase-Locking Value (PLV) features and in combination with spectral power and coherence features [14]. Their results indicated that using spectral power features produced similar classification performance as using PLV, coherence, or any combination of these features. Brodu et al. introduced two sets of features based on multifractal cumulants and predictive complexity of the EEG signals [15]. The winner of BCI competition 2003 for dataset III extracted features using Morlet wavelets and used the Bayesian classifier to differentiate between the imagination of left and right-hand movements [16]. Xu et al. extracted statistical features over set of wavelet coefficients which were fed into a

* Corresponding author.
E-mail address: ykadah@kau.edu.sa (Y.M. Kadah).

http://dx.doi.org/10.1016/j.bspc.2017.05.007
1746-8094/

fuzzy support vector machine (FSVM) classifier to characterize the time-frequency distribution of the EEG signals [17]. Their results outperformed the winner of BCI competition 2003 for dataset III. Zhou et al. proposed extracting bispectrum-based features to characterize the non-Gaussian nature of EEG measurements leading to even better results for the same dataset using different classifiers [18]. On the other hand, Ang et al. proposed filter-bank common spatial pattern algorithm to optimize the subject-specific frequency band for classification of different motor imagery tasks [19]. Their technique was the winner of BCI competition 2008 for datasets 2a and 2b. Delgado et al. proposed new approach for classification of motor imagery tasks based on the hidden conditional random fields (HCRFs) [20]. The extracted features which include autoregressive (AR) modeling of the EEG signals followed by the calculation of the power spectrum were used to model the HCRFs. Their results outperformed the results obtained by the winner of BCI competition 2008 for dataset 2a [20].

Despite the relative success of the above techniques, many of the proposed feature extraction methods for EEG-based BCI systems assume linearity of the recorded EEG signals and hence ignore the well-established nonlinearity of brain electrical activity. Therefore, several methods were proposed to better model the nature of EEG signals using features derived from nonlinear dynamical modeling [21–23]. Many such features were proposed for the analysis of EEG signals at different mental states. Carlino et al. calculated the correlation dimension to differentiate between the EEG signals of healthy and schizophrenic patients [24]. Sakkalis et al. investigated three measures to detect the absence seizures; namely, a linear variance analysis approach, approximate entropy, and order index [25]. Hosseinifard et al. used four nonlinear features including correlation dimension, Lyapunov exponent, Higuchi fractal, and detrended fluctuation analysis (DFA) with KNN classifier to differentiate between normal and depression patients [26]. Banitalebi et al. calculated some chaotic indices such as mutual information, correlation dimension, Lyapunov exponent, and the minimum embedding dimension with multi-layer perceptron classifier and k-means support vector machine (KM-SVM) classifier to discriminate different motor imagery tasks [27]. Fang et al. proposed extracting features from the reconstructed phase space (RPS), which is a transformation of the EEG time series into a geometrical object embedded in a higher-dimensional space [28]. Their features included amplitude frequency analysis (AFA) and autoregressive (AR) modeling of RPS. Unfortunately, the calculation of most (if not all) of such features is notoriously time-consuming, which makes it impractical to develop online classification schemes with high information transfer rates as required for BCI. So, the development of techniques that account for nonlinear dynamical nature of EEG signals while meeting online processing goals would potentially be of significant impact on this field.

In this work, we propose a new processing framework that allows more detailed characterization of the nonlinear features of the EEG signals at real-time rates. In this technique, the phase space trajectory is reconstructed and characterized using two sets of features based on affine-invariant moments and a new distance series transform. Such features allow robust determination of the characteristics of the phase space trajectory while being rather simple to compute to meet real-time performance requirements. We describe the implementation of the new framework and perform experimental verification using data sets from BCI competitions.

2. Methodology

The phase space reconstruction method by Packard et al. [32] was proposed to reconstruct a system's attractor using one or more of its measured time series. Takens [33] showed that the reconstructed phase space (RPS) had the same dynamical properties as the true attractor of the system which produced the measured signals. Consequently, it is possible to reconstruct attractors with different topological properties using EEG measurements at different mental states assuming the human brain as a nonlinear dynamical system. Fig. 1 shows a graphical representation of the time-delay embedding procedure used to estimate RPS in this work. The embedding dimension in this figure was taken as 3 for clear illustration. For the EEG signal at the top plot, three coordinate points are required to create one point in a 3D phase space (bottom left plot). Each point in the phase space is represented by three values, which are the amplitude values of the signal at 3 consecutive time points separated by time lag τ. By repeating these steps for each time point, we can reconstruct the phase space of the underlying dynamical system and obtain an equivalent attractor as shown in the bottom right plot. The details of the time-delay embedding process used in this work are provided in Appendix A.

2.1. Preprocessing

The EEG signals were filtered using either Butterworth band-pass filter of order 3 or Chebyshev Type II band-pass filter of order 6. These filters and their parameters were utilized following the winners of competition II and IV respectively [16,19]. An additional filtering using spectral subtraction denoising method was used here for further noise rejection [36].

2.2. Feature extraction

Two sets of features are introduced in this work to represent the complexity in the phase space trajectory; namely, the moment invariant features and the distance series (DS) transform features. The moment invariant features are used to quantitatively characterize the shape of the RPS. On the other hand, DS features are computed after a transformation of the multidimensional RPS into a one-dimensional space. A set of DS-domain features are derived based on the raw values of the transformed RPS, their autoregressive model coefficients, magnitude of their discrete Fourier transform, and their wavelet decomposition coefficients.

2.2.1. Moment invariant features

Moments are quantitative measures that describe the distribution of a random variable whereby a set of moments with order ranging from zero to order infinity uniquely determine its probability density function. In this study, we consider the reconstructed attractor as m-dimensional object in the phase space and characterize its shape by moments.

Generally speaking, the moment of order p for an m-dimensional object with density function $\rho(S_1, S_2, \ldots, S_m) = \rho(S)$ where S_i is a column in the reconstructed phase space $Y_i(m)$ is given by the Riemann sum as,

$$M_{p_1 \ldots p_m} = \sum_{j=1}^{K} \ldots \sum_{j=1}^{K} S_{1j}^{p_1} \ldots S_{mj}^{p_m} \rho(S)\, dS_1 \ldots dS_m, \tag{1}$$

where $p_1 + p_2 + \ldots + p_m = p$, $0 \le p < \infty$, and K is the length of the embedded dimensions S_i, $i = 1, 2, \ldots, m$. Similarly, the central moments are given by,

$$\mu_{p_1 \ldots p_m} = \sum_{j=0}^{K} \ldots \sum_{j=0}^{K} \left(S_{1j} - \overline{S_1}\right)^{p_1} \ldots \left(S_{mj} - \overline{S_m}\right)^{p_m} \rho(S)\, dS_1 \ldots dS_m, \tag{2}$$

where $\overline{S_1} = \frac{M_{1\ldots0}}{M_{0\ldots0}}, \ldots, \overline{S_m} = \frac{M_{0\ldots1}}{M_{0\ldots0}}$. From the above expressions, such moments will vary if an affine transformation is applied to RPS, which is a clear disadvantage when trying to characterize such

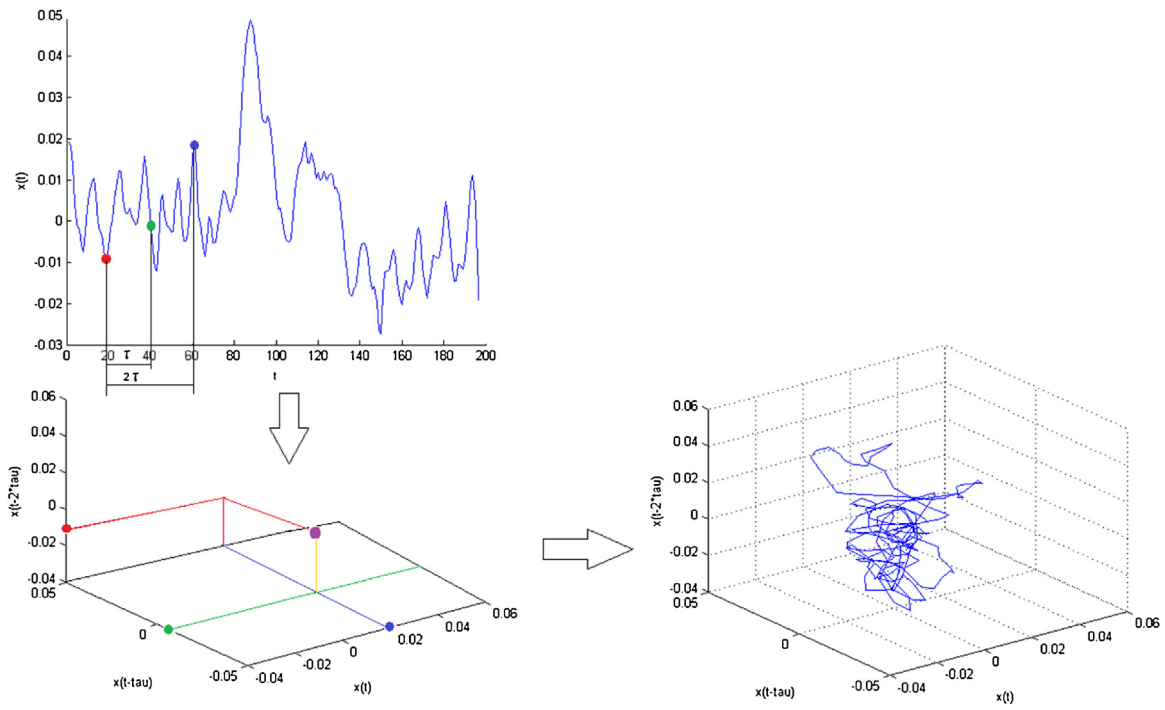

Fig. 1. Time-delay embedding procedure whereby the 3D coordinates of each point in the phase space are obtained as the values of the signal amplitudes at times t, t+τ, and t+2τ. So, the coordinates of the pink point in the phase space in the bottom left 3D plot is obtained as original signal values of the red, green, and blue points in the top 2D plot. Repeating this for all time points t of the original signal results in the complex phase space trajectory shown to the right. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

shapes in a robust manner. The central moments have the advantage of being invariant under translation but still vary under coordinate rotation.

Mamistvalov [37] presented the general fundamental theorem of moment invariants (GFTMI) to help in calculating invariant moments for m-dimensional objects. He proved that the sums of the major minors of the discriminants of matrix D containing the central moments of order two calculated for the m-dimensional RPS are absolute invariants of orthogonal transformations. This matrix is given as,

$$D = \begin{bmatrix} \mu_{2\ldots0} & \ldots & \mu_{1\ldots1} \\ \ldots & \cdots & \ldots \\ \ldots & \cdots & \ldots \\ \ldots & \cdots & \ldots \\ \mu_{1\ldots1} & \cdots & \mu_{0\ldots2} \end{bmatrix}. \tag{3}$$

Since the central moments in the elements of this matrix are invariant to coordinate translation, the calculated sums of major minors of the discriminants of matrix D will be invariant for translation, rotation and reflection transforms. This property allows them to be more robust in characterizing RPS.

As an example of how such moment invariants are calculated, consider the moment invariant features extracted for an EEG segment x of length $N = 100$ samples and optimal delay-time $\tau = 5$ s while embedding dimension $m = 3$. Hence, the RPS is given as,

$$Y_l(3) = \begin{bmatrix} Y_1 \\ Y_2 \\ \ldots \\ Y_K \end{bmatrix} = \begin{bmatrix} x_1 & x_6 & x_{11} \\ x_2 & x_7 & x_{12} \\ \ldots & \ldots & \ldots \\ x_K & x_{K+5} & x_{K+10} \end{bmatrix}, \tag{4}$$

where $K = N - (m-1) \times \tau = 100 - 2 \times 5 = 90$. The location of every point in the RPS is determined by the corresponding Y_l value i.e. the first point Y_1 is located at (x_1, x_6, x_{11}) and the last point Y_{90} is located x_{90}, x_{95}, x_{100}). To calculate the moments M and the central μ, $\rho(S)$ is set to be a vector of ones of length K to indicate that there is a point at the corresponding location Y_l. The D matrix is constructed as,

$$D = \begin{bmatrix} \mu_{200} & \mu_{110} & \mu_{101} \\ \mu_{110} & \mu_{020} & \mu_{011} \\ \mu_{101} & \mu_{011} & \mu_{002} \end{bmatrix}, \tag{5}$$

The number of major minors of the matrix D equals to the number of the embedding dimension m. For example, for $m = 3$, the sums of major minors of the discriminants of matrix D termed O_i can be calculated as,

$$O_1 = \mu_{200} + \mu_{020} + \mu_{002}. \tag{6}$$

$$O_2 = \begin{vmatrix} \mu_{200} & \mu_{110} \\ \mu_{110} & \mu_{020} \end{vmatrix} + \begin{vmatrix} \mu_{020} & \mu_{011} \\ \mu_{011} & \mu_{002} \end{vmatrix} + \begin{vmatrix} \mu_{200} & \mu_{101} \\ \mu_{101} & \mu_{002} \end{vmatrix}. \tag{7}$$

$$O_3 = \begin{vmatrix} \mu_{200} & \mu_{110} & \mu_{101} \\ \mu_{110} & \mu_{020} & \mu_{011} \\ \mu_{101} & \mu_{011} & \mu_{002} \end{vmatrix}. \tag{8}$$

In this work, we use this procedure to calculate m features, O_i with $i = 1,2,\ldots,m$, with optimal embedding dimension m estimated for the m-dimensional RPS of the dataset used.

2.2.2. Distance series features

The distance series (DS) transform domain defined by Sayed et al. [47] is used to transform the multi-dimensional RPS into a one-dimensional series computed as the Euclidian distances between every point in the phase space trajectory and the origin. The raw values of the DS transform domain, which represent how the phase space trajectory evolves with time and reflects the change of the system states are used in combination with the autoregressive

model coefficients, the magnitude of the Fourier transform, and the wavelet packet decomposition coefficients of the DS as features for the developed BCI system.

The Distance Series (DS) maps the complex variations of the multi-dimensional phase space trajectory Y_{ij} into a one-dimensional space D_i which can be calculated as:

$$D_i = \sqrt{\sum_{j=1}^{m} {Y_{ij}}^2}, i = 1, 2, \ldots, M. \quad (9)$$

Small changes in the DS values indicate that the RPS has slow trajectory movements while the large changes of such values indicate wide steps between consecutive points in RPS. So, this allows capturing more details about the underlying dynamics of the system than traditional complexity measures. Fig. 2 shows a simplified illustration of a hypothetical 3D phase space and its corresponding distance series transform. The Euclidian distances between each point in RPS and the origin in (a) are represented by the 1D series as shown in (b).

Dealing with the DS as a simple time series allows use of signal analysis methods to characterize its properties. Such methods include the autoregressive modeling of DS using Burg's algorithm. To calculate the model coefficients, the model order was chosen the same as Fang et al. [28]. Since the fluctuations in the DS values reflect how RPS trajectory points move in space, the magnitude of the Fourier transform of the DS time series is used to characterize RPS trajectory shape. For example, if the trajectory points have the same distance from the origin, the Fourier transform will be concentrated near DC with no significant high-frequency components. Otherwise, the Fourier transform will contain a spectrum that is characteristic to the global changes in the DS. Furthermore, the wavelet packed decomposition coefficients can also be used to provide a time-frequency analysis of the DS time series that allows the detection of local changes in the RPS. In this study, we report results using Daubechies 4 family with 3 levels for the wavelet packet decomposition of the DS to allow repeating this procedure. These parameters were selected as an example of a specific experiment we conducted to allow reproducing the reported results. Optimization of such selection was not attempted and can be considered for future work.

2.3. Feature selection and classification

Student's *t*-test and Fisher score [38] were used for feature selection in addition to principal component analysis (PCA) which was used for dimensionality reduction purposes. The feature vector for each class is formed by concatenating the features calculated from channels C3 and C4. The two feature vectors are fed to Student's *t*-test that assumes that the features calculated for the two classes have normal distributions with unknown but equal variance and tests the null hypothesis of whether they have equal means. Only features with scores below a significance level of 0.05 (p-value <0.05 indicating confidence in the rejection of the null hypothesis) are included in the classification stage. In some cases, the Fisher score was used after applying PCA to select the most relevant features from the new feature space produced from the projection of the extracted features on the principal components. In addition, different classifiers were tested including the K-nearest neighbor (KNN), support vector machines (SVMs) with radial basis function (RBF) kernel, linear discriminant analysis (LDA), and quadratic discriminant analysis (QDA). For each subject, the results obtained using the feature selection procedure as well as the classifier that produce the maximum classification accuracy using the moment invariant features or the DS based features are reported.

The classification procedure differs for each dataset. In Graz2003, we followed the procedure in [28] where the features are extracted from an EEG segment located inside a sliding window of 2 s within range t = 3–9 s. The classifier was trained and tested at the same time point. Alternatively, for Graz2008 dataset, we followed the classification procedure in [19] where a sliding window of 2 s is used within range t = 3-7.5 s for testing while a fixed window of 2 s length is selected to extract the training features. In other words, the classifier is trained with the features extracted from the fixed window and tested using the features extracted from the EEG segment inside the sliding window.

3. Experimental verification

Two datasets from BCI competitions II [29] and IV [30] were used in this work to verify the effectiveness of the proposed features. The first is dataset III from BCI competition II (termed Graz2003) while the second is dataset 2b from BCI competition IV (termed Graz2008). Both datasets were recorded during motor imagery tasks and were provided by the Laboratory of Brain-Computer Interface, Graz University of Technology [31].

Graz2003 dataset was collected over channels C3, Cz, and C4 from one normal female subject who was asked to control a feedback bar by the imagination of left and right hand movements. The dataset contains 280 trials to be divided equally into training and testing sets randomly. The recorded trials include equal numbers of trials for imagination of left and right hand movement. The measured EEG signals were sampled at a rate of 128 Sa/s and filtered using band-pass filter with corner frequencies 0.5 and 30 Hz. The length of each trial was 9 s. At t = 3 s, an arrow was displayed as cue to indicate the direction at which the subject should imagine the movement. The order of left and right cues during the experiment was random.

On the other hand, Graz2008 dataset was collected from 9 subjects with normal or corrected-to-normal vision. For each subject, five sessions were recorded where each session contains EEG measurements recorded using 3 bipolar channels (C3, Cz, and C4) as well as 3 monopolar EOG channels. The first two sessions were recorded without feedback and contained 120 trials each. The other three sessions were recorded during the presence of feedback and contained 160 trials each. The training data comprised the first three sessions while the last two sessions were provided for testing. The EEG measurements were sampled at a rate of 250 Sa/s and filtered using a band-pass filter with corner frequencies 0.5 and 100 Hz. The imagination period in the first two sessions was 4 s (t = 3–7 s), while the imagination period during the feedback sessions was 4.5 s (t = 3–7.5 s).

The first step in calculating the nonlinear dynamical modeling features for the EEG signals is the reconstruction of the phase space which requires the estimation of the embedding parameters that include the optimal delay-time and the optimal embedding dimension. The procedure in Appendix A was used to do that. The first minimum of the mutual information and Cao's algorithm were used to calculate these parameters for every EEG segment inside the sliding window in the training set. Fig. 3 shows the histograms of the optimal delay-time and the optimal embedding dimensions calculated for every EEG segment in the training set. We used the average of the calculated delay-time and embedding dimensions to represent the optimal values that should be used for the reconstruction of the phase space. The average delay-time over the EEG segments in the training dataset was 3 s while the average embedding dimension was 9.

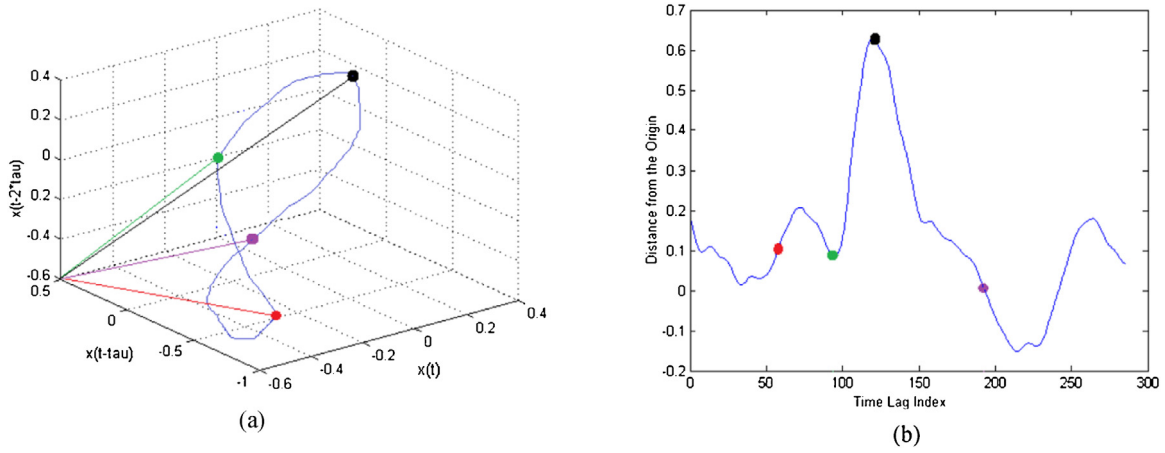

Fig. 2. Graphical representation of distance series (DS) signal generation. The distance between a fixed reference point (origin) to each point on the complex 3D phase space trajectory is computed as shown in (a). Then, the resultant DS signal is obtained as in (b) where each point in the 3D phase space trajectory is represented by one point in the 1D DS signal.

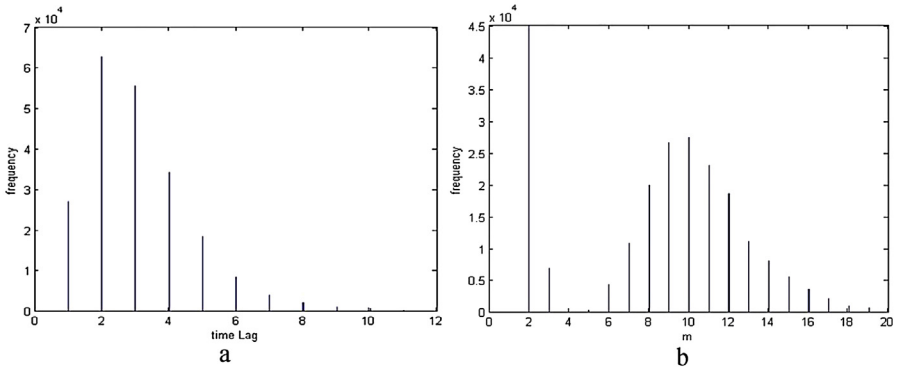

Fig. 3. Histogram of optimal delay-time (a) and the optimal embedding dimensions (b) calculated for every EEG segment in the training set using Cao's method. The average of such delay-time and embedding dimension values is used for the reconstruction of the phase space of testing data.

3.1. Graz2003 dataset results

The EEG signals were filtered using the Butterworth band-pass filter of order 3 with corner frequencies 0.5–30 Hz and the filtered signals were denoised using the spectral subtraction denoising with parameter α selected as 5 to produce the maximum classification accuracy as reported by Makary et al. [41]. Three sets of features were examined including moment invariant features, distance series (DS) features, and combination of both. The evaluation criterion used for Graz2003 dataset was the mutual information (MI), which represents the information transfer between the human brain and the computer [39,40].

3.1.1. Moment invariant features

The moment invariant features were calculated for channels C3 and C4 to construct the feature vector. Since the optimal embedding dimension was $m = 9$, the length of the feature vector will be 18 with 9 features from channel C3 and 9 features from channel C4. In addition, PCA was used for dimensionality reduction using the principal components that represent 99% of the variances in the feature space. Table 1 shows the best performance measures obtained over the classification period using the KNN classifier. As an illustration of the time variations of the error rate and mutual information, Fig. 4 shows the time courses of the MI and the error rate (ERR) for the best classification rate (KNN classifier with K = 9) over the classification period from 3 to 9 s.

Table 1
Performance measures using the moment invariant features with KNN classification. The best results shown in bold were obtained with K = 9.

Classifier	Max Accuracy (%)	Time of Max Accuracy (s)	Max MI (bit)
K = 1	85.71	5.86	0.4941
K = 3	87.86	5.80	0.6025
K = 5	87.86	5.82	0.6128
K = 7	88.57	5.80	0.6429
K = 9	**89.29**	**5.76**	**0.6825**
K = 11	88.57	5.80	0.6312
K = 13	87.14	5.80	0.5447

3.1.2. Distance series features

The DS values calculated for channels C3 and C4 were used to construct the feature vector. Fig. 5 shows the average DS time series over the training dataset calculated for a window that starts at $t = 4$ and ends at $t = 6$ s. It is evident that the curves of the DS calculated for channels C3 and C4 switch places between left and right imagery tasks. This indicates their potential to differentiate between the two classes.

Since the length of the EEG segment inside the 2 s sliding window at the sampling rate used is 256 samples, the number of RPS trajectory points of Eq. (4) (and hence the length of DS time series) calculated for one channel using $m = 9$ $\tau = 3$ will be 232 points. Consequently, the length of the DS feature vector will be 464 features. Table 2 shows the performance measures obtained using Student's t-test for feature selection and KNN for classification. An illustration

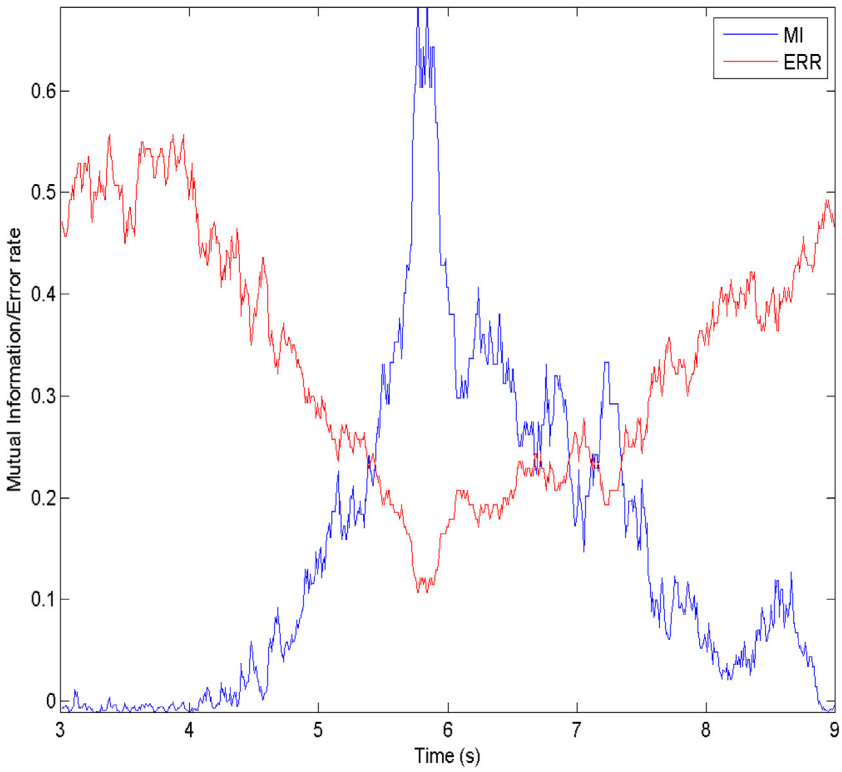

Fig. 4. Time course of the Mutual Information (MI) and Error Rate (ERR) variations for the case of best classification of moment invariant features using KNN classifier with K = 9 where the maximum accuracy (minimum ERR) was 89.29% at time of 5.76 s and MI of 0.6825.

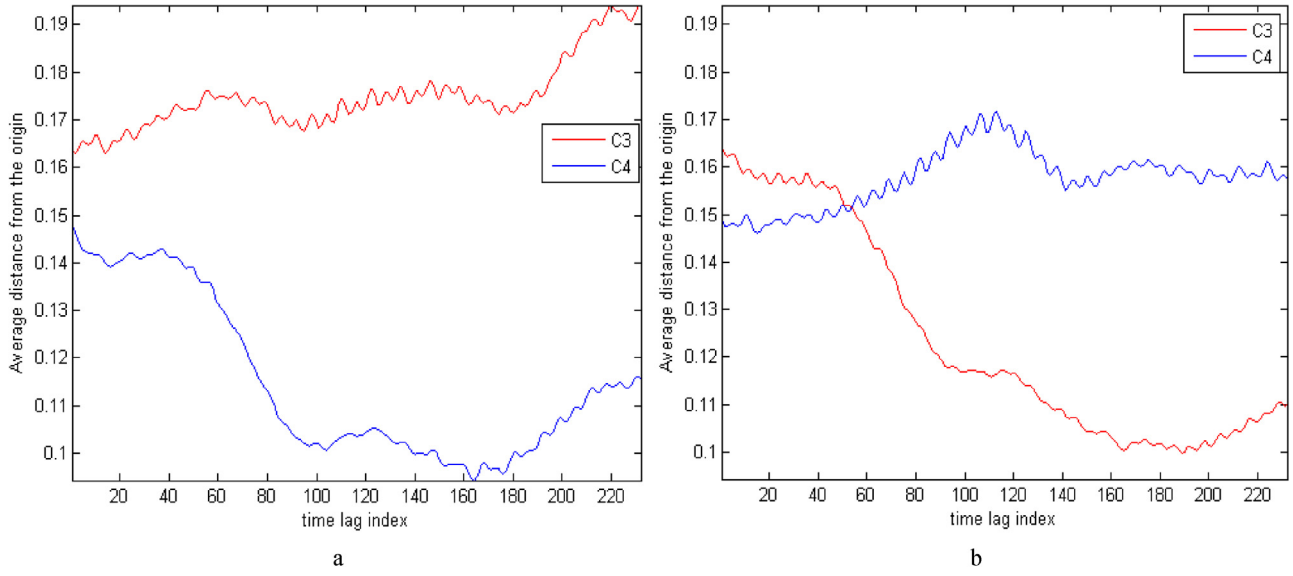

Fig. 5. Average distance series (DS) for Graz2003 dataset for (a) Class 1 (left), and (b) Class 2 (Right). As can be noticed, the curves of the average DS calculated for channels C3 and C4 switch positions between left and right imagery tasks. This illustrates the potential for DS signals to differentiate the two classes.

of the time courses of the MI and the ERR for the best classification results (KNN with K = 11) over the classification period is presented in Fig. 6.

3.1.3. Combined features

Here, the feature vector was constructed using the moment invariant features in addition to the DS values calculated for channels C3 and C4. PCA was used for dimensionality reduction and the Fisher score was applied to the new feature space of the projected features onto the principal components that explain 99% of the variance. The number of features selected using the Fisher score was obtained using 10-fold cross-validation over the training data. Table 3 shows the performance measures obtained using PCA and Fisher score for dimensionality reduction and feature selection respectively and KNN classification. The time courses of the MI and the ERR for KNN with K = 5 are shown in Fig. 7.

Table 4 shows a comparison between the proposed methods and previous studies in the literature for the motor imagery tasks addressing the same Graz2003 dataset. The proposed methods outperformed the previous methods based on the competition

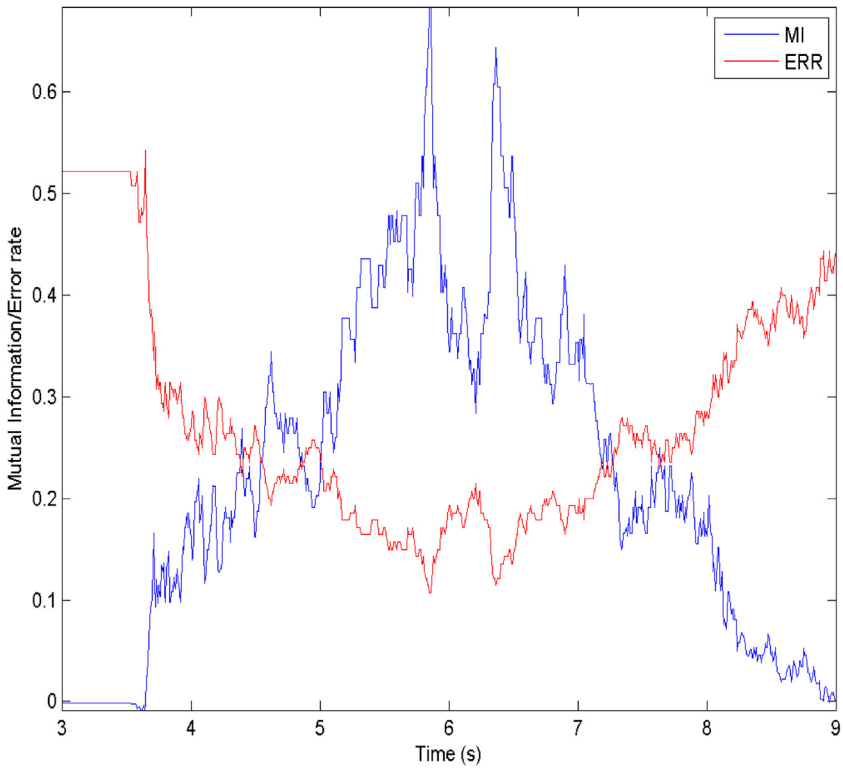

Fig. 6. Time course of the Mutual Information (MI) and Error Rate (ERR) variations for the case of best classification of distance series (DS) features using KNN classifier with K = 11 where the maximum accuracy (minimum ERR) was 89.29% at time of 5.84 s and MI of 0.6836.

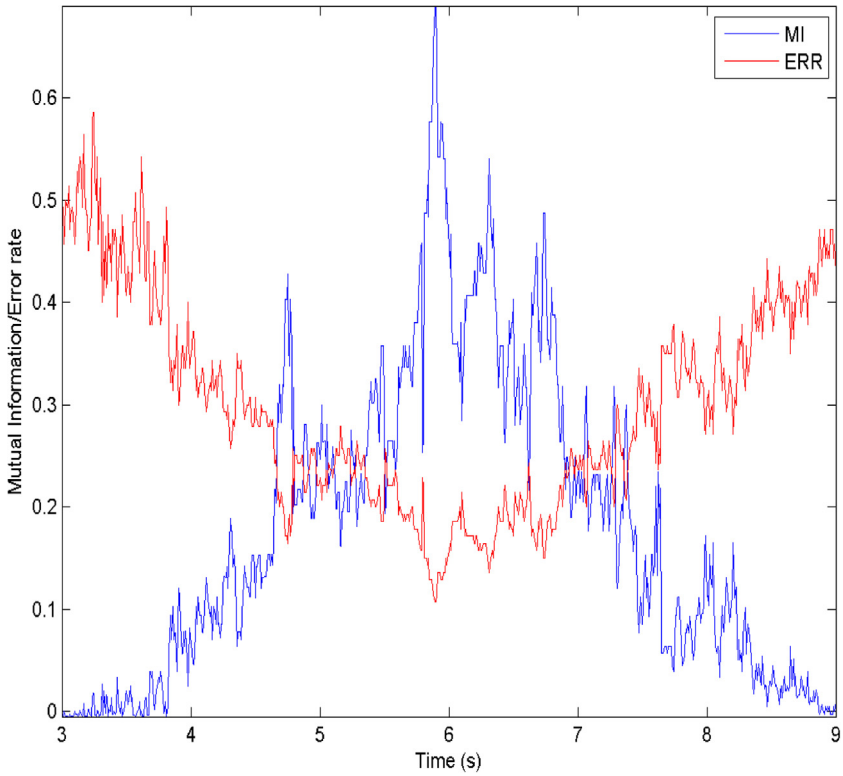

Fig. 7. Time course of the Mutual Information (MI) and Error Rate (ERR) variations for the case of best classification using the combination of moment invariant and distance series (DS) features using KNN classifier with K = 5 where the maximum accuracy (minimum ERR) was 89.29% at time of 5.88 s and MI of 0.6816.

criterion (i.e., MI). As shown in Table 4, the MI obtained by the competition winner [16] was 0.61 while the best result in [28] was 0.67. In this work, the MI values for the proposed methods using moment invariant features, distance series features, and combined features were 0.6836, 0.6825, and 0.6816 respectively.

Table 2
Performance measures using the DS based Features and KNN classification. The best results shown in bold were obtained with K values of 9 or 11.

Classifier	Max Accuracy (%)	Time of Max Accuracy (s)	Max MI (bit)
K = 1	82.14	6.11	0.3793
K = 3	85.71	5.99	0.5144
K = 5	84.29	5.90	0.4580
K = 7	85.71	5.82	0.5099
K = 9	87.85	**5.82**	0.6041
K = 11	**89.29**	5.84	**0.6836**
K = 13	86.43	5.82	0.5389

Table 3
Performance measures using Moment Invariant and DS-based Features and KNN classification. The best results shown in bold were obtained with K = 5.

Classifier	Max Accuracy (%)	Time of Max Accuracy (s)	Max MI (bit)
K = 1	82.86	5.85	0.3942
K = 3	85.00	5.84	0.4987
K = 5	**89.29**	5.88	**0.6816**
K = 7	86.43	5.87	0.5410
K = 9	87.14	5.94	0.5575
K = 11	87.14	5.84	0.5630
K = 13	87.14	6.30	0.5982

3.2. Graz2008 dataset results

We extracted the proposed features for Graz2008 dataset to evaluate their effectiveness in differentiating between the imagination of left and right hand movements. The phase space was reconstructed using $\tau = 3$ and $m = 9$ for channels C3 and C4 using the technique in Appendix A. We followed the procedure used by the winner [19] for training and testing phases. In this procedure, session 3 was selected by the cross-validation scheme to be used in the training phase for subjects 4, 5, 6, 7, 8, and 9 while sessions 1, 2, and 3 were used for subjects 2 and 3. On the other hand, only sessions 1 and 3 were used for subject 1. During the training phase, the features were extracted from a 2 s window which starts at $t = 3.5$ s and ends at $t = 5.5$ s for all subjects except subject 7 where the window starts at $t = 5.5$ s and ends at $t = 7.5$ s. Furthermore, sessions 4 and 5 were used for testing phase for all subjects. The evaluation criterion for Graz2008 dataset was Cohen's kappa coefficient [43].

Our goal in this work is to maximize the performance measures (i.e. Cohen's kappa coefficient and classification accuracy) calculated for each subject. Since we only use EEG data from one subject at a time for training and testing the system, we utilized different preprocessing, feature selection, and classification algorithms along with the proposed features to build a BCI system which produces the highest performance measures for each subject. The details of the different processing settings used for each subject are provided in Tables 5–7.

3.2.1. Moment invariant features

The moment invariant features were extracted for each subject at different preprocessing, feature selection, and classification techniques. Table 5 shows the different settings used for each subject as well as the classification accuracy and kappa value. The corner frequencies for Chebyshev Type II band-pass filter was 4 and 40 Hz while the corner frequencies for Butterworth band-pass filter was 0.1–30 Hz as used by the competition winners [16,19]. In addition, spectral subtraction denoising was used after bandpass filtration with parameter α where zero value of this parameter means no spectral subtraction processing. The number of features selected by Fisher score is 2 and the percentage of variance threshold used for PCA is 99%.

3.2.2. Distance series features

The DS features calculated for each subject in Graz2008 dataset showed the same behavior as the DS calculated for Graz2003 dataset which confirmed their effectiveness in differentiating between different motor imagery tasks. So, the feature vector was constructed using the raw DS values calculated for channels C3 and C4, in addition to the AR model coefficients, the magnitude of the Fourier transform, and the wavelet packet decomposition coefficients of these DS values. Table 6 shows the performance measures obtained using the DS based features and the corresponding settings used for each subject.

For subject 5, the Butterworth filter was used after applying the independent component analysis (ICA) [44] on the original dataset to remove the effect of the EOG signals on the measured EEG. We used FastICA [45] to calculate the independent components. For subjects 1, 2, 3, and 7, we applied PCA on the feature space and then used Fisher score to select the most discriminating features from the new feature space produced from the projection of the extracted features on the principal components that represent 99% of the variance. The number of features selected by the Fisher score was 2 for all subjects.

3.2.3. Combined features

Table 7 shows the performance measures obtained using the moment invariant features combined with the DS based features with the corresponding settings used to produce these results. ICA was used for subject 5 as above and the Fisher score feature selection was applied after using PCA for subjects 1, 2, 3, 6, and 8. Table 8 presents a comparison between the results of the proposed methods and other studies. The average kappa values obtained using the DS based features as well as using the combined feature vector can be noted to exceed the average kappa provided by the winner of the BCI competition IV dataset 2b [19]. The proposed methods also outperform most recent work for subjects 2, 3, 7, and 8 while the kappa value obtained for subject 4 is the same for all of the work in comparison. The best average kappa value was obtained by Delgado et al. which was slightly better than the proposed methods [20].

Table 4
Comparison Between the Proposed Methods and Other Studies for Graz2003 Dataset. All three proposed methods provided Maximum MI metrics that are better than those from previous work.

Method	Features	Classifier	Max MI (bits)	Min ERR	Time of Max MI (s)
Proposed 1	Moment Invariant	KNN, K = 11	**0.6836**	10.71%	5.84
Proposed 2	DS Values	KNN, K = 9	0.6825	10.71%	5.76
Proposed 3	Moment Inv. +DS Values	KNN, K = 5	0.6816	10.71%	5.88
Fang et al. [28]	AFAPS+ARPS	LDA	0.67	9.29%	5.76
Xu et al. [17]	Wavelet based features	FSVM	0.66	12.14%	5.92
Zhou et al. [18]	Higher order statistics	NN	0.64	10%	–
Zhou et al. [18]	Higher order statistics	LDA	0.63	10.71%	–
Competition winner [16]	Morlet Wavelets	Bayesian Classifier	0.61	10.71%	7.59

Table 5
Performance measures using the moment invariant features for Graz2008 dataset. The optimal processing methodology for each subject is shown with its corresponding performance metrics. The average performance metrics for all subjects are also presented.

Subject	Filter Type	α	Feature Selection	Classifier	Max Kappa	Max Accuracy (%)
S1	Chebyshev	0.5	PCA	QDA	0.40	69.7
S2	Chebyshev	1	PCA	QDA	0.32	66.1
S3	Butterworth	0	Fisher Score	KNN, K = 9	0.27	63.9
S4	Chebyshev	0	Fisher Score	QDA	0.95	97.7
S5	Butterworth	1	PCA	KNN, K = 5	0.59	79.5
S6	Chebyshev	0	PCA	LDA	0.54	76.8
S7	Chebyshev	0.5	PCA	KNN, K = 13	0.59	79.3
S8	Chebyshev	1	PCA	LDA	0.85	92.6
S9	Chebyshev	0	PCA	KNN, K = 11	0.71	85.3
Average					**0.58**	**79.0**

Table 6
Performance measures obtained using the DS based Features for Graz2008 dataset. The optimal processing methodology for each subject is shown with its corresponding performance metrics. The average performance metrics for all subjects are also presented.

Subject	Filter Type	α	Feature Selection	Classifier	Max Kappa	Max Accuracy (%)
S1	Chebyshev	0	PCA + Fisher Score	KNN, K = 7	0.43	70.1
S2	Chebyshev	0	PCA + Fisher Score	KNN, K = 11	0.29	64.5
S3	Chebyshev	0	PCA + Fisher Score	KNN, K = 11	0.26	62.6
S4	Chebyshev	6	Fisher Score	KNN, K = 1	0.95	97.7
S5	Butterworth	1.5	PCA	KNN, K = 11	0.85	92.7
S6	Chebyshev	0	PCA	LDA	0.54	77.2
S7	Chebyshev	0	PCA + Fisher Score	LDA	0.60	80.2
S8	Chebyshev	6	PCA	KNN, K = 13	0.90	95.2
S9	Chebyshev	1	PCA	KNN, K = 13	0.66	82.9
Average		**0.61**	**80.3**			

Table 7
Performance measures obtained using the moment invariant features and the DS based features for Graz2008 dataset. The optimal processing methodology for each subject is shown with its corresponding performance metrics. The average performance metrics for all subjects are also presented.

Subject	Filter Type	α	Feature Selection	Classifier	Max Kappa	Max Accuracy (%)
S1	Chebyshev	0	PCA + Fisher Score	QDA	0.39	68.9
S2	Chebyshev	0	PCA + Fisher Score	LDA	0.31	65.3
S3	Chebyshev	0	PCA + Fisher Score	KNN, K = 13	0.28	63.9
S4	Chebyshev	6	Fisher Score	KNN, K = 1	0.95	97.7
S5	Butterworth	1.5	PCA	KNN, K = 11	0.83	91.6
S6	Chebyshev	0	PCA + Fisher Score	LDA	0.48	74.0
S7	Chebyshev	0	Fisher Score	KNN, K = 13	0.62	81.0
S8	Chebyshev	0	PCA + Fisher Score	KNN, K = 13	0.90	94.8
S9	Chebyshev	0.25	Fisher Score	LDA	0.70	84.9
Average		**0.61**	**80.2**			

Table 8
Comparison Between the Proposed Methods and Other Studies for Graz2008 Dataset. The best results for each subject and for the average of all subjects are shown in bold. The proposed methods outperformed the competition winner [19] in 6 out of 9 subjects in addition to the average for all subjects.

Subject	Moment Invariants	Distance Series	Combined	Ang et al. [19]	Gan et al. [46]	Delgado et al. [20]
S1	0.40	0.43	0.39	0.40	0.42	**0.56**
S2	**0.32**	**0.29**	**0.31**	0.21	0.21	0.24
S3	**0.27**	**0.26**	**0.28**	0.22	0.14	0.18
S4	**0.95**	**0.95**	**0.95**	**0.95**	0.94	**0.95**
S5	0.59	0.85	0.83	0.86	0.71	**0.93**
S6	0.54	0.54	0.48	0.61	0.62	**0.72**
S7	0.59	0.60	**0.62**	0.56	0.61	0.51
S8	0.85	**0.90**	**0.90**	0.85	0.84	0.80
S9	0.71	0.66	0.70	0.74	**0.78**	0.71
Average	0.58	0.61	0.61	0.60	0.58	**0.62**

4. Discussion

Given that the ultimate goal of BCI is to allow real-time interaction between the subject and his/her environment, it is important to assess the processing requirements of the proposed methods to ensure that they would allow such performance. So, the processing time for the proposed methods were calculated on a modest processing platform using a PC with Intel® Core™ i7 2.66 GHz processor and 4 GB RAM running Windows 7 and Matlab 2011b. The results for Graz2003 dataset indicate that the average processing times were 0.0258, 0.0066, and 0.0432 s for moment invariant, distance series and combined features respectively. The corresponding average processing times for Graz2008 data set were 0.0283, 0.0871, and 0.11 s respectively. Since the sampling rate for Graz2003 dataset was 128 Sa/s while that for Graz2008 was 250 Sa/s, the processing time should be less than 0.008 s for Graz2003 or 0.004 s for Graz2008 datasets to provide strict real-time performance. However, since variations to be detected are usually

much slower than the sampling period, it is evident that all three proposed processing methodologies will allow practical real-time tracking of changes in the BCI signals. Furthermore, with the present availability of parallel processing hardware such as multi-core processors or graphics processing units (GPUs), it is possible to reduce the processing time to the desired time resolution required by the application. Nevertheless, this comes at the expense of more power consumption. Hence, we expect future work to focus on the trade-off between the two conflicting requirements in the realization of practical wearable BCI devices.

Fang et al. [28] proposed feature extraction scheme based on the phase space reconstruction of the EEG signals. They assumed the value of the embedding dimension $m = 2$ based on some studies which indicated that many chaotic systems, such as Lorenz, Rossler and EEG are not sensitive to the selection of embedding dimension m, but rather sensitive to the selection of delay-time τ [42]. The delay time for RPS was selected as the delay time that produces the maximum performance measures using the cross-validation scheme over the training dataset assuming prior knowledge about the range of the delay time. No optimization procedures were used to select neither the optimal embedding dimension nor the delay-time. On the other hand, in our study, we used the first minimum of the mutual information and the Cao's algorithm to optimally select the delay-time and the embedding dimension respectively. In addition, we dealt with the RPS as one unit which represent m-dimensional object in the phase space to avoid extracting features from each dimension separately. The experimental results showed that the quantitative measures (i.e. the moment invariant features) that characterize the shape of the RPS can be used to distinguish between different motor imagery tasks. Furthermore, the proposed DS shows that the brain dynamics exhibit different behavior at different mental states and the DS can be used to describe the evolution of the underlying dynamical system. So, this indicates that the proposed framework is likely to be more robust given its optimized RPS and more descriptive RPS feature extraction.

The utilization of different preprocessing, feature selection, and classification algorithms along with the proposed features for the results of Graz2008 dataset indicate the wide inter-subject variability of the data that required different processing to be used to reach optimal results. Even though this seems as a drawback, in practice the optimal selection of such processing blocks can be done using a "fitting" session for each subject that identifies the optimal set of techniques to be used by the subject. This is an existing practice in other medical devices that require such custom fitting such as hearing aids.

The focus of this work was to propose a new set of features based on the nonlinear dynamical analysis of the EEG that can be used to obtain better performance measures for BCI systems. So, we used a simple classifier such as the KNN to emphasize that the enhanced performance is due to the proposed features and not the use of a sophisticated classifier. So, optimization of classification by use/comparison of more elaborate classifiers is open for further investigation.

Several aspects of the proposed framework need further research to investigate. Examples include using different distance metric for the distance series transformation (e.g., 1-norm), using different features to characterize the 1D distance series transform signals, and theoretical study of the implications of dimensionality reduction from the multidimensional phase space trajectory to the 1D distance series transform. Also, it is always important to consider the possible practical implementations for proposed methods to allow mobility. This includes deployment on suitable low-power embedded systems for example.

The fast processing time of the proposed method suggests its possible application to design real-time BCI systems. As an example, prosthetic arm control is an application where real-time performance of BCI system allows direct and intuitive control modality for disabled persons. Other time-critical applications include intelligent mental state monitoring to improve safety in operating heavy industrial machinery or driving. Further investigation of the use of the proposed method in such applications was left to future work.

5. Conclusion

In this study, we proposed a new framework for processing motor imagery datasets based on the characterization of the shape of phase space trajectories. Two sets of features were developed based on moment invariants and distance series transform features. The proposed methods were applied on two different motor imagery datasets from BCI competitions II and IV respectively. The experimental results indicate the potential of the proposed framework in improving the results obtained for this type of data sets and possible generalization for use with other data sets. Moreover, the required modest computation time suggest their practicality for most applications.

Acknowledgement

This project was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under grant no. (12-MED3081-03). The authors, therefore, acknowledge with thanks DSR technical and financial support.

Appendix A. Phase space reconstruction procedure

The time-delay method proposed by Packard et al. [32] is used here to reconstruct the phase space using the mutual information between time-shifted versions of the EEG signals and the Cao's algorithm for determining the optimal delay-time and the optimal embedding dimension respectively. Let $\{x_k : k = 1, 2, \ldots, N\}$ be the measured signal, the reconstructed phase space $Y(m)$ can be reconstructed as the following:

$$Y_i(m) = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdots \\ Y_M \end{bmatrix} = \begin{bmatrix} x_1 & x_{1+\tau} & \cdots & x_{1+(m-1)\tau} \\ x_2 & x_{2+\tau} & \cdots & x_{2+(m-1)\tau} \\ \cdots & \cdots & \cdots & \cdots \\ x_M & x_{N-m\tau} & \cdots & x_N \end{bmatrix}, \quad (I1)$$

where $M = N - (m-1)\tau$, N is the length of the measured signal, m is the embedding dimension, and τ is the delay-time.

A. Optimal Selection of Delay-Time

The time lag corresponding to the first local minimum of the mutual information between time-shifted versions of the EEG segments is selected to be the optimal delay-time τ. The mutual information between time series x_k and its delayed version $x_{k+\tau}$ can be calculated as,

$$I(\tau) = \sum_{i=1}^{N_s}\sum_{j=1}^{N_s} P[x_k(i), x_{k+\tau}(j)]\log_2 \frac{P[x_k(i), x_{k+\tau}(j)]}{P[x_k(i)]P[x_{k+\tau}(j)]}, \quad (I2)$$
$$k = 1, 2, \ldots, N - (m-1)\tau.$$

here, N_s is the total number of the probability states, $P[x_k(i)]$ is the probability of time series x_k belonging to state i, $P[x_{k+\tau}(j)]$ is the probability of the delayed version $x_{k+\tau}$ belonging to state j, $P[x_k(i), x_{k+\tau}(j)]$ is the joint probability of x_k belonging to state i, and $x_{k+\tau}$ belonging to state j concurrently. $P[x_k(i)]$ can be calculated using the formula proposed by [34] as follows:

$$P[x_k(i)] = \frac{n(i)}{N - (m-1)\tau}, \quad (I3)$$

where n(i) is the number of data points in time series x_k belonging to state i while $N-(m-1)\tau$ is the length of x_k. Probabilities $P[x_{k+\tau}(i)]$ and $P[x_k(i), x_{k+\tau}(j)]$ can be calculated using Eq. (I3) as well.

B. Optimal Selection of Embedding Dimension

The optimal embedding dimension m for every EEG segment in the training dataset was determined using Cao's method [35] that has been recently used to determine the minimum embedding dimension for scalar time series. Cao's method is based on the assumption that if m is the true embedding dimension for the given time series, then any two points that are close to each other in the m dimensional space will continue close in the higher order embedding dimensions (i.e. m + 1 dimensional space). In this method, the phase space is being reconstructed using m = 1 and the value of m is increased until the number of the false neighbors approaches zero. Cao's function is given by,

$$E(m) = \frac{1}{N-m\tau}\sum_{i=1}^{N-m\tau}\frac{\|Y_i(m+1)-Y_{n(i,m)}(m+1)\|}{\|Y_i(m)-Y_{n(i,m)}(m)\|}. \quad (I4)$$

here $Y_{n(i,m)}(m)$ is the nearest neighbor of $Y_i(m)$ in m dimensional space and $\|^*\|$ represents the maximum norm function. The minimum embedding dimension is determined as the value of m at which the function $E_1(m)$ approaches one with $E_1(m)$ defined as:

$$E_1(m) = \frac{E(m+1)}{E(m)}. \quad (I5)$$

The value of $E_1(m)$ may approach one even with random signals. Therefore, Cao defined a more robust function $E_2(m)$ to differentiate between random and deterministic time series such that:

$$E_2(m) = \frac{E^*(m+1)}{E^*(m)}, \quad (I6)$$

where,

$$E^*(m) = \frac{1}{N-m\tau}\sum_{i=1}^{N-m\tau}|x_{i+m\tau}-x_{n(i,m)+m\tau}|. \quad (I7)$$

here, $x_{n(i,m)+m\tau}$ are the nearest neighbors of $x_{i+m\tau}$. The time series is said to be deterministic if the value of $E_2(m)$ is not equal one for at least one value of m.

References

[1] J.R. Wolpaw, N. Birbaumer, W.J. Heetderks, D.J. McFarland, P.H. Peckham, G. Schalk, E. Donchin, L.A. Quatrano, C.J. Robinson, T.M. Vaughan, Brain-computer interface technology: a review of the first international meeting, IEEE Trans. Rehabil. Eng. 8 (2) (2000) 164–173.

[2] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, T.M. Vaughan, Brain–computer interfaces for communication and control, Clin. Neurophysiol. 113 (June (6)) (2002) 767–791.

[3] G. Pfurtscheller, C. Neuper, Motor imagery and direct brain-computer communication, Proc. IEEE 89 (July (7)) (2001) 1123–1134.

[4] N. Birbaumer, Breaking the silence: brain-computer interfaces (BCI) for communication and motor control, Psychophysiology 43 (November (6)) (2006) 517–532.

[5] N. Weiskopf, K. Mathiak, S.W. Bock, F. Scharnowski, R. Veit, W. Grodd, R. Goebel, N. Birbaumer, Principles of a brain-computer interface (BCI) based on real-time functional magnetic resonance imaging (fMRI), IEEE Trans. Biomed. Eng. 51 (6) (2004) 966–970.

[6] G. Schalk, E.C. Leuthardt, Brain-computer interfaces using electrocorticographic signals, IEEE Rev. Biomed. Eng. 4 (2011) 140–154.

[7] J. Mellinger, G. Schalk, C. Braun, H. Preissl, W. Rosenstiel, N. Birbaumer, A. Kübler, An MEG-based brain-computer interface (BCI), Neuroimage 36 (3) (2007) 581–593.

[8] M. Strait, C. Canning, M. Scheutz, Limitations of NIRS-Based BCI for Realistic Applications in Human-Computer Interaction, 2013, pp. 2–3.

[9] L.F. Nicolas-Alonso, J. Gomez-Gil, Brain computer interfaces, a review, Sensors (Basel) 12 (January (2)) (2012) 1211–1279.

[10] G. Pfurtscheller, Functional brain imaging based on ERD/ERS, Vision Res. 41 (May (10–11)) (2001) 1257–1260.

[11] G. Pfurtscheller, A. Aranibar, Evaluation of event-related desynchronization (ERD) preceding and following voluntary self-paced movement, Electroencephalogr. Clin. Neurophysiol. 46 (February (2)) (1979) 138–146.

[12] G. Pfurtscheller, C. Brunner, A. Schlögl, F.H. Lopes da Silva, Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks, Neuroimage 31 (1) (2006) 153–159.

[13] G. Pfurtscheller, F.H. Lopes da Silva, Event-related EEG/MEG synchronization and desynchronization: basic principles, Clin. Neurophysiol. 110 (November (11)) (1999) 1842–1857.

[14] D.J. Krusienski, D.J. McFarland, J.R. Wolpaw, Value of amplitude, phase, and coherence features for a sensorimotor rhythm-based brain-computer interface, Brain Res. Bull. 87 (1) (2012) 130–134.

[15] N. Brodu, F. Lotte, A. Lécuyer, Exploring two novel features for EEG-based brain-computer interfaces: multifractal cumulants and predictive complexity, Neurocomputing 79 (2012) 87–94.

[16] S. Lemm, C. Schafer, G. Curio, BCI competition 2003—Data set III: probabilistic modeling of sensorimoto Mu rhythms for classification of imaginary hand movements, IEEE Trans. Biomed. Eng. 51 (6) (2004) 1077–1080.

[17] Q. Xu, H. Zhou, Y. Wang, J. Huang, Fuzzy support vector machine for classification of EEG signals using wavelet-based features, Med. Eng. Phys. 31 (7) (2009) 858–865.

[18] S.-M. Zhou, J.Q. Gan, F. Sepulveda, Classifying mental tasks based on features of higher-order statistics from EEG signals in brain–computer interface, Inf. Sci. (Ny) 178 (March (6)) (2008) 1629–1640.

[19] K.K. Ang, Z.Y. Chin, C. Wang, C. Guan, H. Zhang, Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b, Front. Neurosci. 6 (January) (2012) 39.

[20] J.F. Delgado Saa, M. Cetin, Hidden Conditional Random Fields for Classification of Imaginary Motor Tasks from EEG Data, 2011, pp. 171–175.

[21] S.A. Akar, S. Kara, F. Latifoğlu, V. Bilgiç, Estimation of nonlinear measures of schizophrenia patientsğ EEG in emotional states, IRBM 36 (July (4)) (2015) 250–258.

[22] O. Vyšata, M. Vališ, A. Procházka, R. Rusina, L. Pazdera, Linear and nonlinear EEG synchronization in Alzheimer's disease, Neurophysiology 47 (April (1)) (2015) 46–52.

[23] G.M. Bairy, S. Bhat, L.W.J. Eugene, U.C. Niranjan, S.D. Puthankattil, P.K. Joseph, Automated classification of depression electroencephalographic signals using discrete cosine transform and nonlinear dynamics, J. Med. Imaging Heal. Inform. 5 (June (3)) (2015) 635–640.

[24] E. Carlino, M. Sigaudo, A. Pollo, F. Benedetti, T. Mongini, F. Castagna, S. Vighetti, P. Rocca, Nonlinear analysis of electroencephalogram at rest and during cognitive tasks in patients with schizophrenia, J. Psychiatry Neurosci. 37 (July (4)) (2012) 259–266.

[25] V. Sakkalis, G. Giannakakis, C. Farmaki, A. Mousas, M. Pediaditis, P. Vorgia, M. Tsiknakis, Absence seizure epilepsy detection using linear and nonlinear EEG analysis methods, Proc. 35th IEEE Eng. Med. Biol. Soc. Con. 2013 (2013) 6333–6336.

[26] B. Hosseinifard, M.H. Moradi, R. Rostami, Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal, Comput. Methods Progr. Biomed. 109 (March (3)) (2013) 339–345.

[27] A. Banitalebi, S.K. Setarehdan, G.A. Hossein-Zadeh, A technique based on chaos for brain computer interfacing, 14th International CSI Computer Conference (2009) 464–469.

[28] Y. Fang, M. Chen, X. Zheng, Extracting features from phase space of EEG signals in brain–computer interfaces, Neurocomputing 151 (2015) 1477–1485.

[29] BCI Competition II [Online]. Available: http://bbci.de/competition/ii/ (Accessed 19 April 2015).

[30] BCI Competition IV [Online]. Available: http://www.bbci.de/competition/iv/ (Accessed 19 April 2015).

[31] Laboratory of Brain-Computer Interface [Online]. Available: https://bci.tugraz.at/ (Accessed 19 April 2015).

[32] N.H. Packard, J.P. Crutchfield, J.D. Farmer, S.R.S, Geometry from a time series, Phys. Rev. Lett. 45 (9) (1980) 712–715.

[33] F. Takens, Detecting strange attractors in turbulence, Dyn. Syst. Turbul. Warwick 1980 381 (1981).

[34] G. Williams, Chaos Theory Tamed, Josheph Henery Press, Washington, D.C, 1997.

[35] L. Cao, Practical method for determining the minimum embedding dimension of a scalar time series, Phys. D Nonlinear Phenom. 110 (1–2) (1997) 43–50.

[36] M.M. Makary, Processing Methodlogies for Brain Computer Interface, Cairo University, 2013.

[37] A.G. Mamistvalov, N-dimensional moment invariants and conceptual mathematical theory of recognition n-dimensional solids, IEEE Trans. Pattern Anal. Mach. Intell. 20 (8) (1998) 819–831.

[38] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, 2000, p. 680.

[39] a. Schlogl, C. Keinrath, R. Scherer, P. Furtscheller, Information transfer of an EEG-based brain computer interface, First Int. IEEE EMBS Conf. Neural Eng. 2003. Conf. Proceedings (2003) 641–644 (no. 10).

[40] G. Dornhege, T. Hinterberger, D.J. Mcfarland, Toward Brain Computer Interfacing, MIT Press, 2007.

[41] M.M. Makary, Y.M. Kadah, Improving P300 and SCP-based Brain computer interfacing by spectral subtraction denoising, 2nd Middle East Conference on Biomedical Engineering (2014) 228–231.

[42] R.-Y. You, X.-J. Huang, Phase space reconstruction of chaotic dynamical system based on wavelet decomposition, Chin. Phys. B 20 (February (2)) (2011) 020505.

[43] J. Cohen, A coefficient of agreement for nomial scales, Educ. Psychol. Meas. XX (1) (1960) 37–46.

[44] A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, John Wiley & Sons, 2004.
[45] FastICA. [Online]. Available: http://research.ics.aalto.fi/ica/fastica/ (Accessed: 17 June 2015).
[46] BCI Competition IV: Results. [Online]. Available: http://www.bbci.de/competition/iv/results/#dataset2b (Accessed 21 June 2015).
[47] K.S. Sayed, A.F. Khalaf, Y.M. Kadah, Arrhythmia classification based on novel distance series transform of phase space trajectories, in: Proc. 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, 2015, pp. 5195–5198.