ELSEVIER

Volume 178, issue 6, 15 March 2008

ISSN 0020-0255

INFORMATION SCIENCES

Informatics and Computer Science
Intelligent Systems
Applications

AN INTERNATIONAL JOURNAL

Available online at
ScienceDirect
www.sciencedirect.com

Available online at www.sciencedirect.com

ScienceDirect

Information Sciences 178 (2008) 1629–1640

INFORMATION SCIENCES
AN INTERNATIONAL JOURNAL

www.elsevier.com/locate/ins

Classifying mental tasks based on features of higher-order statistics from EEG signals in brain–computer interface

Shang-Ming Zhou [a,*], John Q. Gan [b], Francisco Sepulveda [b]

[a] *Centre for Computational Intelligence, School of Computing, De Montfort University, Leicester LE1 9BH, UK*
[b] *Department of Computing and Electronic Systems, University of Essex, Colchester CO4 3SQ, UK*

Received 29 June 2006; received in revised form 20 November 2007; accepted 21 November 2007

Abstract

In order to characterize the non-Gaussian information contained within the EEG signals, a new feature extraction method based on bispectrum is proposed and applied to the classification of right and left motor imagery for developing EEG-based brain–computer interface systems. The experimental results on the Graz BCI data set have shown that based on the proposed features, a LDA classifier, SVM classifier and NN classifier outperform the winner of the BCI 2003 competition on the same data set in terms of either the mutual information, the competition criterion, or misclassification rate.

Keywords: Brain–computer interfaces; Classification; Electroencephalogram (EEG); Feature extraction; Higher-order statistics; Bispectrum

1. Introduction

The recent decade has witnessed a rapid development of brain–computer interface (BCI) technology. An independent BCI is a communication system for controlling a device, e.g. computer, wheelchair or a neuroprosthesis, by human intensions, which does not depend on the brain's normal output pathways of peripheral nerves and muscles but relies on the detectable signals representing responsive or intentional brain activities [33]. Current techniques for monitoring brain activities include electroencephalogram (EEG), Electrocorticogram, Positron Emission Tomography (PET), functional Magnetic Resonance Imaging (fMRI), and Magnetoencephalography (MEG), among which EEG has been popularly used for BCI implementation due to its low cost, non-invasive nature, and its comparatively easily recording brain signals [11,12,33]. What is more, EEG data indicates that neural patterns of meanings in each brain occur in trajectories of discrete steps, whist the amplitude modulation in EEG wave is the mode of expressing meanings [7]. Although these EEG wave packets do not represent external objects, they embody and implement the meanings of objects for each

* Corresponding author. Tel.: +44 (0) 116 2551551x6709; fax: +44 (0) 116 2078159.
E-mail address: smzhou@ieee.org (S.-M. Zhou).

0020-0255/$ - see front matter
doi:10.1016/j.ins.2007.11.012

individual, in terms of what they portend for the future of that individual, and what that individual should do with and about them [7,8]. The information obtained in EEG can be extracted for social communication.

However, a successful EEG-based BCI system very much depends on whether the following two requirements can be satisfied: (1) The extracted EEG features are able to differentiate the task-oriented brain states; and (2) The methods for classifying such features in real-time are efficient. Specifically, it is essential to be able to extract complex spatial and temporal patterns from noisy multi-channel data obtained from EEG measurements [9], as well-studied classification methods are available in the field of machine learning [14,29,35]. The pros and cons of linear and nonlinear classification methods for BCI research can be found in [17].

Currently, most existing schemes for extracting EEG features are based on autoregressive (AR) models or adaptive AR models (AAR) [1,3,8,10,22,30], and power spectral density (PSD) [2,34] (see [33] for a review). In practice, physiologically meaningful EEG features can be extracted from various frequency bands of recorded EEG signals. McFarland et al. reported that the imagined movement signals could be reflected in the β rhythm (13–22 Hz) [15]. Pfurtscheller showed that μ (8–13 Hz) and/or β rhythm amplitudes could serve as effective inputs for a BCI to distinguish a movement or motor imagery [21]. Moon et al. [16] employed a smoothing algorithm for the power of $\alpha(\mu)$-band (8–13 Hz) and of θ-band (5–7 Hz) frequencies of EEG curve and the variation of pulse width obtained from ECG curve, to generate their corresponding trend curves, then the three trend features are applied to a fuzzy system to estimate the mental workload. In [20], a fuzzy ARTMAP neural network based BCI was proposed, in which two different spectral analyzes methods were used to obtain the PSD of the EEG signals from 0 to 50 Hz.

However, conventional methods for feature extraction based on AR models and PSD assume linearity, Gaussianality and minimum-phase within EEG signals, i.e., the amplitudes of EEG signals are normally distributed, their statistical properties do not vary over time, and their frequency components are uncorrelated. Under these assumptions, the EEG signal is considered as a linear superimposition of statistically independent sinusoidal or other wave components, and only frequency and power estimates are considered while phase information is generally ignored. In reality, however, EEG signals are generated by a typical nonlinear system consisting of, for example, post-synaptic neurons whose firing action potentials are based on whether their membrane potential is greater than a threshold. Thus EEG signals would have many sinusoidal components of distinct frequencies, interacting nonlinearly to produce one or more sinusoidal components at sum and difference frequencies [18], which cannot be completely characterized by autocorrelation functions, as done by AR models or PSD estimation methods.

To overcome this limitation, this paper proposes a new set of features for EEG-based BCI systems, which includes higher-order statistics based on the bispectrum of EEG signals. To evaluate the effectiveness of this feature set, the LDA (linear discriminant analysis) classifier, support vector machine (SVM) classifier, and neural network (NN) classifier are adopted to classify the Graz BCI data set which was used in the BCI-competition 2003 [4]. On the other hand, it is known that in pattern recognition, error rate is the most commonly used criterion in measuring the performances of different methods. However, error rate only considers the signs of the classifier outputs, but not the degrees of memberships of patterns belonging to each class, so error rate measure just provides classification accuracy of the used classifier for us, does not give us the information how much confidence about the classifying result is. In order to combine classification accuracy and confidence, in BCI-competition 2003 on the Graz BCI data set, entropy based mutual information (MI) [26,27] obtained from classifying results was used as the criterion to compare the performances of different methods. Greater MI of classifying results by a classifier indicates this classifier produces the results with higher confidence. In this paper, the classifying results obtained by the LDA, SVM, and NN classifiers based on the proposed feature set are extensively compared with the ones achieved by the BCI-competition 2003 winning methods [5,25] on the same data set in terms of the criteria of MI and misclassification rate.

2. Bispectrum based feature extraction

2.1. The definition and properties of the third-order cumulant and bispectrum

For a non-Gaussian third-order stationary random process $\{x(t)\}$, its third-order cumulant in a discrete form is defined as

$$C_{3x}(m,n) = E[x(k)x(k+m)x(k+n)], \tag{1}$$

where E is the expectation over the process multiplied by 2 lagged versions of itself [23,19]. The corresponding bispectrum is defined as the 2D Fourier transform of the third-order cumulant:

$$B_x(\omega_1,\omega_2) = \sum_{m=-\infty}^{+\infty}\sum_{n=-\infty}^{+\infty} C_{3x}(m,n)\cdot\exp[-\mathrm{j}2\pi(m\omega_1+n\omega_2)]. \tag{2}$$

If a random process $x(t)$ is Gaussian, then $C_{3x}(m,n)=0$, thus non-Gaussian process can be detected by this property. If $z(t)=x(t)+w(t)$, where $w(t)$ is Gaussian and independent of $x(t)$, then $C_{3z}(m,n)=C_{3x}(m,n)$. Therefore, colored or white noise processes are suppressed and the bispectrum of a non-Gaussian signal can be recovered. By using higher-order statistics, the standard minimum-phase assumption, which is necessary when the process is characterized by linear model based on Gaussian or only second-order statistics are used, may be removed. Furthermore, higher-order cumulants can give evidence of nonlinearity, while the autocorrelation sequence cannot. These properties would be very useful as features for EEG signal classification. For instance, Fig. 1 shows an EEG signal corresponding to a left-hand motor imagery, its bispectrum and the diagonal slice, while Fig. 2 illustrates an EEG signal corresponding to a right-hand motor imagery, its bispectrum and the diagonal slice. At least in the data samples shown, it is obvious that the bispectrum provides distinctive features for the two types of EEG signals. This suggests the bispectrum may be useful for signal classification.

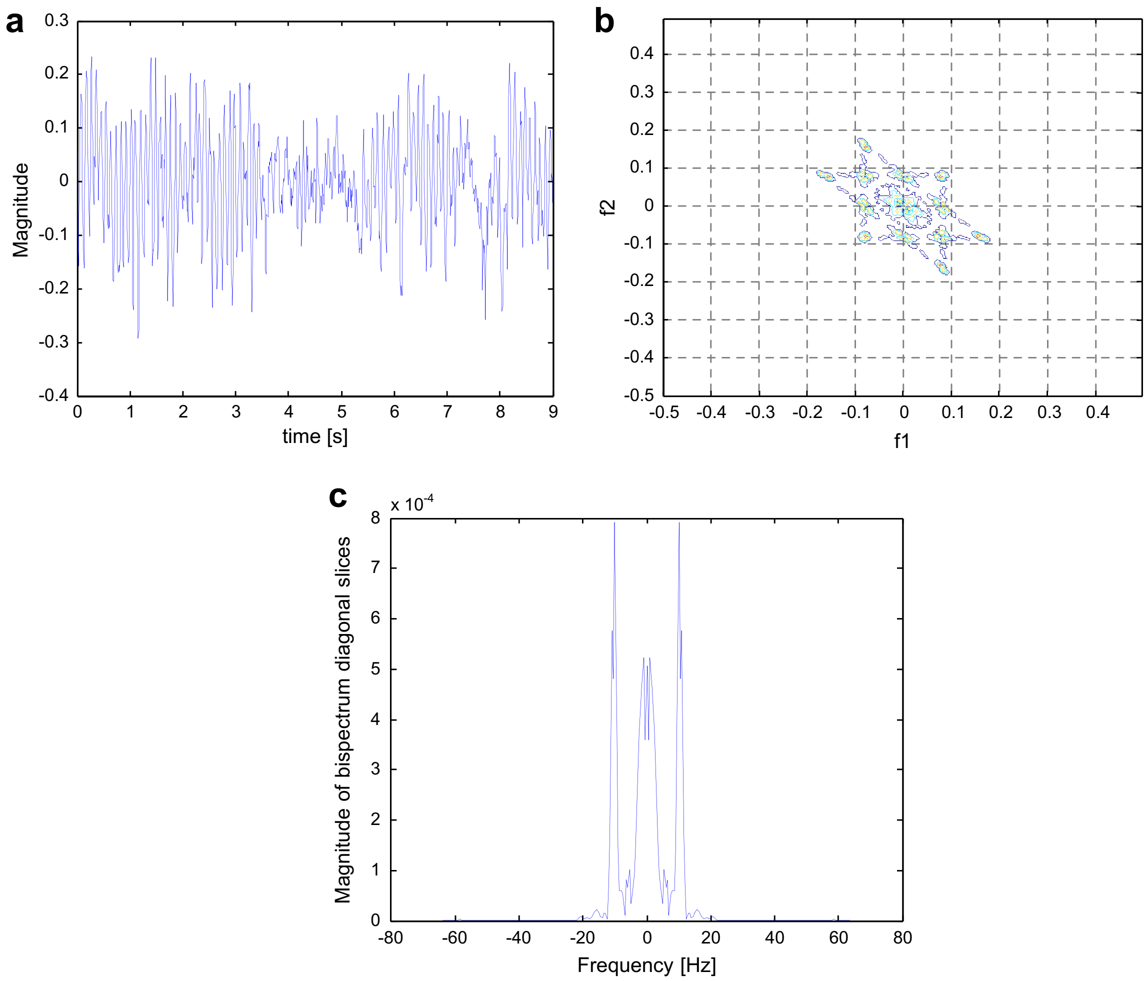

Fig. 1. (a) An EEG signal corresponding to a left-hand motor imagery; (b) a contour plot of the magnitude of the estimated bispectrum on the bi-frequencies (f_1,f_2) plane; and (c) the diagonal slice of the bispectrum.

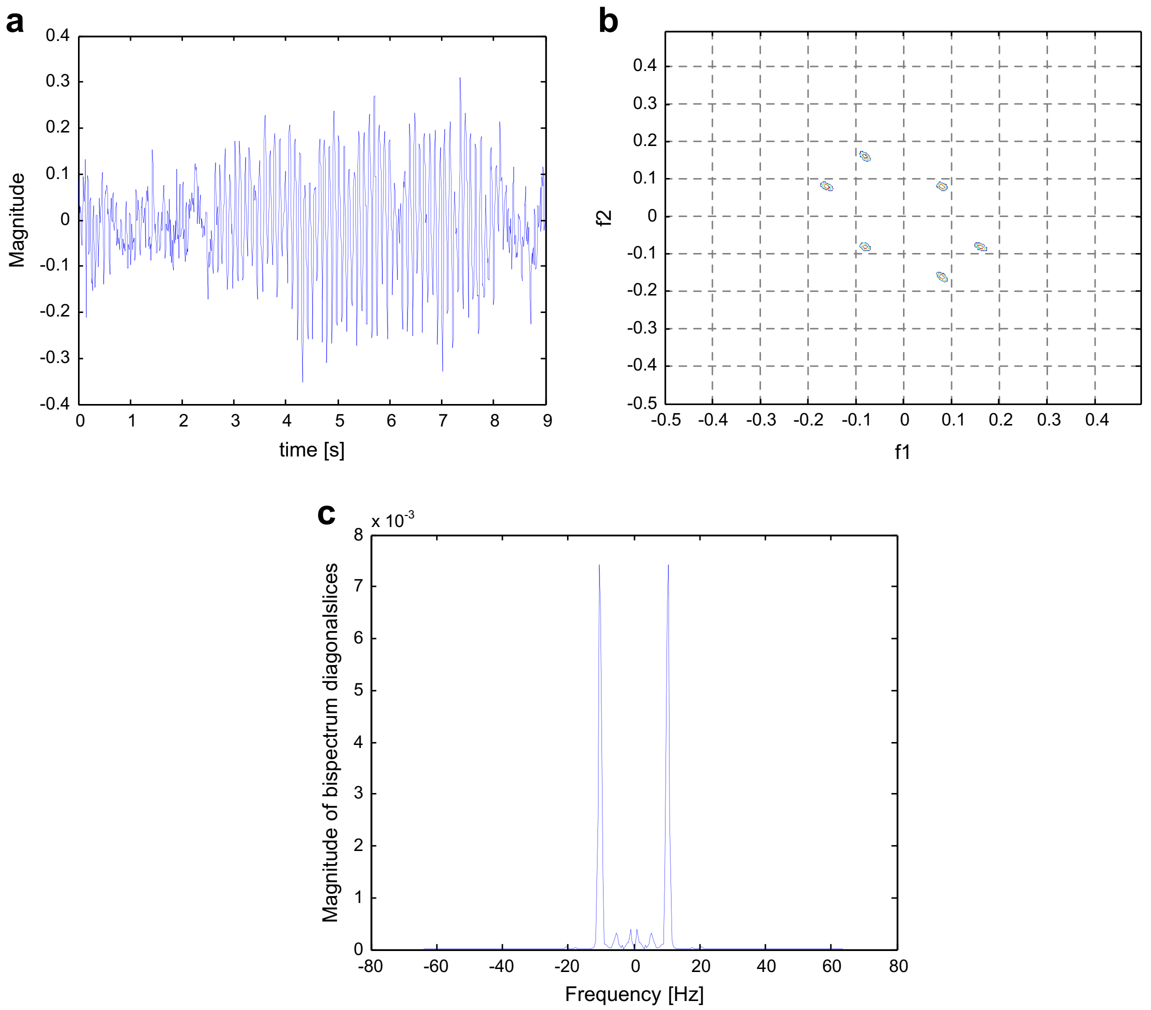

Fig. 2. (a) An EEG signal corresponding to a right-hand motor imagery; (b) a contour plot of the magnitude of the estimated bispectrum on the bi-frequencies (f_1,f_2) plane; and (c) the diagonal slice of the bispectrum.

2.2. Feature extraction

In order to characterize the temporal and frequency information within EEG data, this paper proposes to use the following hybrid features for an EEG-based BCI system:

(1) Four coefficients of the AR model obtained by the Burg method [28].
(2) Four features related to PSD:
  (i) peak frequency of the PSD;
  (ii) peak value of the PSD;
  (iii) the first-order spectral moment of the PSD:

$$m_1(\mathrm{PSD}) = \sum_{k=1}^{N} k \cdot \mathrm{PSD}_k \quad (3)$$

  and
  (iv) the second-order spectral moment of the PSD:

$$m_2(\mathrm{PSD}) = \sum_{k=1}^{N} (k - m_1)^2 \cdot \mathrm{PSD}_k. \quad (4)$$

(3) Four features related to the third-order statistics:

(i) the sum of logarithmic amplitudes of the bispectrum

$$H_1 = \sum_{\omega_1,\omega_2 \in F} \log(|B_x(\omega_1,\omega_2)|), \tag{5}$$

where F is the frequency range to be considered.

(ii) the sum of logarithmic amplitudes of diagonal elements in the bispectrum

$$H_2 = \sum_{\omega \in F} \log(|B_x(\omega,\omega)|), \tag{6}$$

(iii) the first-order spectral moment of the amplitudes of diagonal elements in the bispectrum

$$H_3 = \sum_{k=1}^{N} k \cdot \log(|B_x(\omega_k,\omega_k)|) \tag{7}$$

and

(iv) the second-order spectral moment of the amplitudes of diagonal elements in the bispectrum

$$H_4 = \sum_{k=1}^{N} (k - H_3)^2 \cdot \log(|B_x(\omega_k,\omega_k)|). \tag{8}$$

The above 12 features are extracted for each channel at every sampling point by using a sliding window, in which the first-order spectral moment and second-order spectral moment of PSD as useful statistical descriptors are used to convey information about the uncertainty of the PSD distribution, and the four features related to the third-order statistics are employed to characterize the nonlinear information within EEG signals.

3. Classification

In order to demonstrate the effectiveness of the proposed features in BCI applications, the LDA classification method, SVM method, and NN classification method are used in this paper and compared with others. For the sake of self-containment, this section briefly introduces the LDA, SVM, and NN classifiers.

3.1. LDA classifier

LDA firstly maps the data (feature vector) $\mathbf{x}$ to be classified by the following linear transformation:

$$y = \mathbf{w}^{\mathrm{T}}\mathbf{x} + w_0, \tag{9}$$

where $\mathbf{w}$ and w_0 are determined by maximising the ratio of between-class variance to within-class variance to guarantee maximal separability [6]. The within-class variance matrix is defined by

$$S_w = \sum_{i=1}^{K} \sum_{l=1}^{L_i} (x^l - \mu_i)(x^l - \mu_i)^{\mathrm{T}}, \tag{10}$$

where K is the number of classes, and μ_i the mean vector of the class i, L_i the number of samples within-class i, and the between-class variance matrix is defined by

$$S_b = \sum_{i=1}^{K} (\mu_i - \mu)(\mu_i - \mu)^{\mathrm{T}}, \tag{11}$$

where μ is the mean of the entire training sample set.

The classification is conducted as follows (for simplicity two class problem is used as an example):

$$\mathbf{x} \in \begin{cases} \text{class 1}, & \text{if } y > 0, \\ \text{class 2}, & \text{if } y < 0. \end{cases} \tag{12}$$

3.2. SVM classifier

The invention of SVM was driven by underlying statistical learning theory, i.e., following the principle of structural risk minimization that is rooted in VC dimension theory, which makes its derivation even more profound [31]. The SVMs have been a topic of extensive research with wide applications in machine learning and engineering. The output of a binary SVM classifier can be computed by the following expression:

$$y = \mathrm{sgn}\left(\sum_{i=1}^{N} \alpha_i y_i k(x_i, x) + b\right), \tag{13}$$

where $\{x_i, y_i\}_{i=1}^{N}$ are training samples with input vectors $x_i \in R^d$, and class labels $y_i \in \{-1, 1\}$, $\alpha_i \geqslant 0$ are Lagrangian multipliers obtained by solving a quadratic optimization problem, b is the bias, and $k(x_i, x_j)$ is called kernel function in SVM. The most commonly used kernel function is the Gaussian RBF function

$$k(x_i, x_j) = \exp\left(\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right). \tag{14}$$

The protruding characteristics of SVM lies in its elegant mechanism of handling nonlinear function classes [31], i.e., nonlinear information processing is carried out by means of linear techniques in an implicit high-dimensional feature space mapped by a nonlinear transformation $\phi(x_i)$ from original input space. Although, the analytical expressions of $\phi(x_i)$ is unknown, but because only the inner product operations $\phi(x_i)^{\mathrm{T}}\phi(x_j)$ are involved, the kernel functions can be used to substitute the inner product operations according to the Mercer theorem. Vapnik's theory [31] shows that the SVM solution is found by minimizing both the error on the training set (empirical risk) and the complexity of the hypothesis space, expressed in terms of VC dimension. In this sense, the decision function found by SVM is a tradeoff between learning error and model complexity.

3.3. NN classifier

NN characterized by parallel computing is a powerful machine learning scheme, which has achieved many successful applications. Promisingly, in some application fields, NN models have achieved human-like performance over more traditional artificial intelligence techniques. Now, NN has become a broad term which includes many diverse models and approaches. In this paper, we only focus on the most widely used network: multi-layer feedforward NN trained by backpropagation of error.

A multi-layer feedforward network has two or more layers of units, with the output from one layer serving as input to the next. There are no connections within a layer. The input layer has N neurons which are merely "fan-out" units, where N equals number of classification inputs, no processing takes place in these units. The layers with no external output connections are referred to as hidden layers, whilst there are M neurons in the output layer, where M equals number of classification outputs. In most cases, a feedforward NN with one hidden layer of units is used with a sigmoid activation function for the units.

For a multi-layer feedforward NN, the well known backpropagation algorithm is used to train this network from data. Although the backpropagation mechanism of training multi-layer networks was derived by Werbos [32], it was not popularized until Rumelhart et al. introduced the training algorithm-generalized delta rule in the late 80s of the 20th century [24]. Generally speaking, the backpropagation algorithm works as follows. At the output layer, the output vector is compared with the desired outputs. The error is calculated from the delta rule and is propagated back through the network to adjust the weights in the interests of minimizing the difference between the NN outputs and the desired outputs. Such networks can learn arbitrary associations by using differentiable activation functions. A theoretical foundation of backpropagation can be found in [24,32].

4. Experimental results

This section presents experimental results on a benchmark EEG data set which was used in the BCI-competition 2003 [4]. The results obtained using the proposed features and the classification methods of LDA, SVM and NN are compared to those of the competition winners.

4.1. Description of the Graz BCI data set [4,5]

In collecting the Graz BCI data set, the subject was asked to control a feedback bar by means of imagery left- or right-hand movements after a cue was indicated. The order of left and right cues was random. The experiment consists of 7 runs with 40 trials each. In the available data set there are 280 trials. As shown in Fig. 3, each trial lasts 9 s, in which the first 2 s was quiet. At $t = 2$ s an acoustic stimulus indicates the beginning of the trial, with a cross "+" displayed for 1 s. At $t = 3$ s, an arrow (left or right) was displayed as a cue, and at the same time the subject was asked to do motor imagery along the direction of the cue. Three bipolar EEG channels (anterior '+', posterior '−') were measured over C3, Cz and C4. The EEG was sampled with 128 Hz and was filtered between 0.5 and 30 Hz.

4.2. Experimental results

Because channel Cz shows its independence of the motor imagery, only channel C3 and C4 were used for feature extraction. Hence, 24 features were obtained for each trial of the EEG signal, which were sent to the used classifier. Because the cue (left or right) appeared at $t = 3$ s, in our experiments, only the data between $t = 3$ s and 9 s was used.

Based on the definitions in Section 2.2, features were extracted at every sampling point, with the sliding window size being 256 samples in order to capture the rich frequency information in the EEG signal. In our experiments on a PC with Dual Core CPU, it takes 3.14 ms to extract features from a pattern with 256 sampling points, so the proposed feature extraction method can be applied in real-time mental task classification. Table 1 illustrates examples of features extracted for four trials. In the 280 trials, 140 labeled trials were used to train the classifier. Because the true labels of the other 140 trials are now available in [4], they can be used to test the generalization performance of the trained classifier. The test data set was kept unseen in the feature extraction and the training of the classifiers. In this paper, the SVM with Gaussian kernel (14) is used.

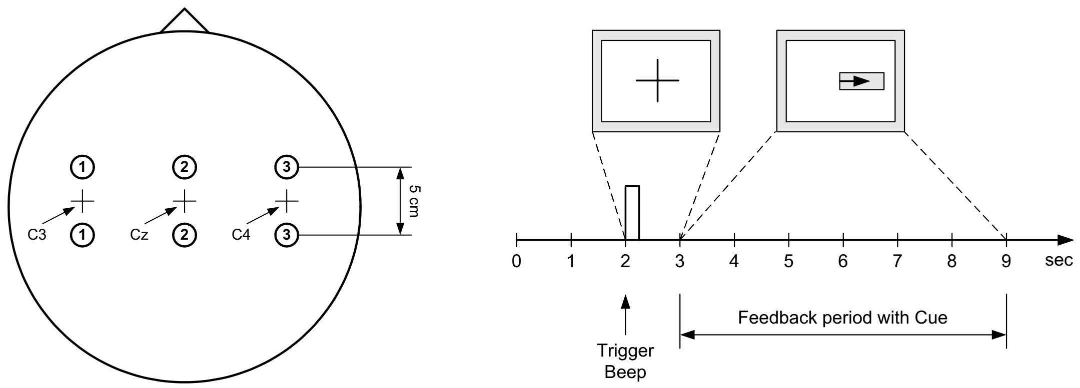

Fig. 3. Electrode positions (top) and timing scheme (bottom) for recording the Graz BCI data set [4].

Table 1
Examples of feature values of four trials

Trials	Feature values
1	−1.198 0.436 −0.066 0.049 −1.174 0.490 −0.083 −0.057 17.000 −3.524 0.572 39.251 2.000 −3.272 0.584 44.708 −361712.945 −1371.967 1.975 252.727 −352152.281 −1344.930 1.677 216.441
2	−1.389 0.747 −0.244 0.195 −1.455 0.933 −0.324 0.176 40.000 −3.084 1.009 55.782 44.000 −3.114 1.234 71.965 −356267.876 −1361.585 3.187 398.842 −353368.409 −1349.770 7.751 908.451
3	−1.386 0.638 −0.061 0.045 −1.265 0.514 −0.059 0.013 36.000 −3.287 0.905 51.329 2.000 −3.468 0.607 39.949 −360164.029 −1366.554 4.176 513.481 −356478.939 −1354.895 2.665 335.205
4	−1.237 0.638 −0.203 −0.029 −1.373 0.756 −0.224 0.089 2.000 −3.372 0.458 34.977 33.000 −3.393 0.912 57.449 −365534.236 −1398.412 1.231 159.589 −352820.815 −1357.196 5.573 674.378

And in the used NN with three layers, the input layer has 24 nodes for the features, the hidden layer has 15 nodes, and the output layer has two nodes for the classes of hand motor imagery, and backpropagation algorithm is used to train the NN. However, it should be noted that because the SVM classifier and NN classifier involve the choices of some hyper-parameters during construction, i.e., the SVM classifier needs to find optimal values for kernel parameter σ and the regularization parameter C, whilst the NN classifier needs to select appropriate learning rates, so in the off-line training process, the genetic algorithm (GA) is used to select the optimal values of these hyper-parameters for the SVM classifier and the NN classifier, respectively, in which the original 140 labeled training trials are separated into training data subset and validation data subset.

Moreover, in order to combine classification accuracy and confidence, in BCI-competition 2003 on the Graz BCI data set, the MI [26,27] was used as the criterion to compare the performances of different methods. Table 2 ranks the performances of the BCI-competition 2003 winning methods and the LDA, SVM and NN methods based on the proposed features in terms of MI criterion, whilst Table 3 illustrates the ranking order of the BCI-competition 2003 winning methods and the LDA, SVM and NN methods with the proposed features in terms of misclassification rate criterion. To show the time course of the mutual information, Fig. 4 depicts the time courses of the mutual information obtained by the BCI-competition 2003 winning methods [5,25]. As a comparison, Fig. 4 shows the time courses of the mutual information of the NN, LDA and SVM classifiers based on the proposed features. In Figs. 4 and 5, the increase of the mutual information indicates an increase in separability between left- and right-hand motor imagery. In our methods shown in Fig. 5 and most BCI-competition methods shown in Fig. 4, the MI values tend to be zero at time 9 s. This is reasonable, because according to the experimental settings described in Section 4.1, the hand motor imagery happens after the cue is displayed at time 3 s, and this process will not last long. At time 9 s, there is no hand motor imagery happening, so there is no much information leading to the separation of the left- and right-hand motor imagery. As a result, the MIs values at time 9 s should be zero, which has been validated by the proposed methods shown in Fig. 5 and most BCI-competition methods shown in Fig. 4. However, as shown in Fig. 4 some BCI competitors' MI values at time 9 s are not zero, the possible reason is as follows: the sliding windows used in these methods for online feedbacks are so large that they still cover the time sequences of hand motor imagery period at time 9 s.

Table 2
Ranking order of the proposed method and the BCI-competition 2003 winning methods in terms of the MI on the Graz BCI data set

Ranking	Methods	Maximal MI (bit)	Minimal misclassification rate (%)
1	NN with the proposed features	0.64	10.00
2	LDA with the proposed features	0.63	10.71
3	BCI_Comp2003_1st winner	0.61	10.71
4	SVM with the proposed features	0.58	10.00
5	BCI_Comp2003_2nd winner	0.46	15.71
6	BCI_Comp2003_3rd winner	0.45	17.14
7	BCI_Comp2003_4th winner	0.44	13.57

Table 3
Ranking order of the proposed method and the BCI-competition 2003 winning methods in terms of the error rates on the Graz BCI data set

Ranking	Methods	Minimal misclassification rate (%)	Maximal MI (bit)
1	NN with the proposed features	10.00	0.64
2	SVM with the proposed features	10.00	0.58
3	LDA with the proposed features	10.71	0.63
4	BCI_Comp2003_1st winner	10.71	0.61
5	BCI_Comp2003_4th winner	13.57	0.44
6	BCI_Comp2003_2nd winner	15.71	0.46
7	BCI_Comp2003_3rd winner	17.14	0.45

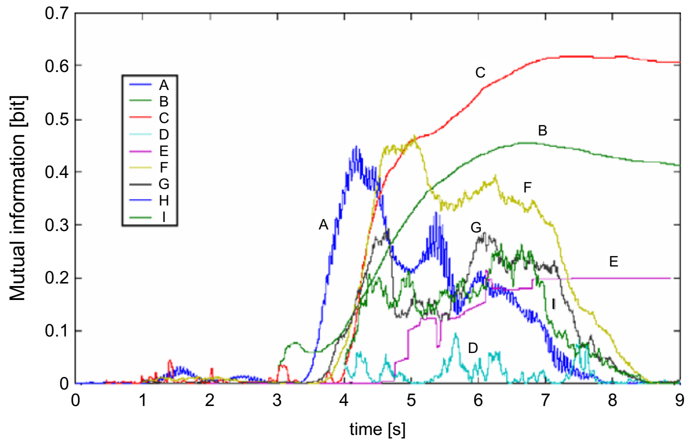

Fig. 4. The MI time courses of the BCI-competition winning methods (A–I: the serial numbers of competitors) [25]: at $t = 3$ s the cue (left or right in random order) was presented.

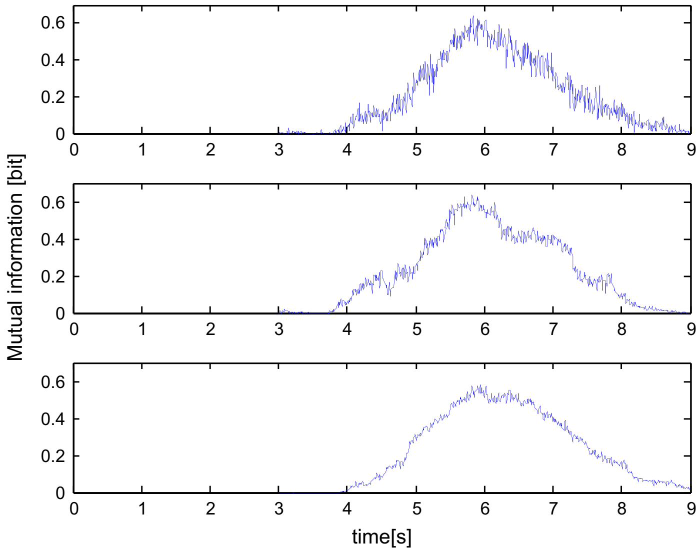

Fig. 5. The MI time courses of the NN (top), LDA (middle) and SVM (bottom) classifiers with proposed features: at $t = 3$ s the cue (left or right in random order) was presented.

In terms of the criterion of BCI-competition 2003, the NN and LDA classifiers based on the proposed features achieve the maximum of the MI 0.64 and 0.63, respectively, both of which are greater than 0.61, the one achieved by the first winner of the BCI-competition 2003 on the Graz data set. On the other hand, in terms of the misclassification rates, the NN and SVM classifiers based on the proposed features achieve the error rates 10.0%, both of which are smaller than the ones achieved by all the winners of the BCI-competition 2003. Hence, the proposed EEG-based mental task classification systems outperform the winner of BCI-competition 2003 on the Graz BCI data set, which demonstrates the effectiveness of the propose feature set.

Furthermore, we use another mental task data set [13] to validate the proposed features for classifying the hand motor imageries, this data set was generated by different experimental settings and different subjects for hand motor imagery [13]. In our experiments, 121 samples are randomly selected for training and 121 samples for testing. Table 4 illustrates the experimental results using LDA, SVM and NN to classify this data set based on the proposed features, in which the SVM with Gaussian kernel (14) and the NN with three layers are used. In the NN, the input layer has 24 nodes for the features, the hidden layer has 15 nodes, and the output layer has two nodes, and backpropagation algorithm is used to train the NN. And GA is used to select the optimal

Table 4
Experimental results of the proposed features

Ranking	Methods	Maximal MI (bit)	Minimal misclassification rate (%)
1	NN with the proposed features	0.64	10.00
2	SVM with the proposed features	0.63	9.00
3	LDA with the proposed features	0.61	12.00

Table 5
Experimental results of the AR features

Ranking	Methods	Maximal MI (bit)	Minimal misclassification rate (%)
1	NN with AR features	0.38	21.00
2	SVM with AR features	0.25	18.00
3	LDA with AR features	0.27	22.00

values of the hyper-parameters in the SVM classifier and the NN classifier, respectively, in which the original 121 labeled training trials are separated into training data subset and validation data subset.

Moreover, as a comparison, the widely used AR features are extracted for LDA, NN and SVM classifiers to classify the hand motor imagery tasks [13], in which the AR features are obtained for each channel by using the Burg method [28] in AR model. The structures of SVM and NN remain the same, and GA is used to select the optimal values of the hyper-parameters in the SVM classifier and the NN classifier, respectively, in which the training data subset and validation data subset are same as the ones in the above experiments about the proposed features. Table 5 shows the experimental results using LDA, SVM and NN to classify this data set based on the AR features. It can be seen that the classifiers with the proposed features can achieve better performance than the ones with the AR features in terms of the criteria of MI and misclassification rate.

It is known that EEG signals are generated from human brain which is a system with highly nonlinear dynamics, but there is no evidence indicating that the activation of human brain is Gaussian. We believe that if the EEG-based BCI classification and coding quality is to be improved, then more of the information available in the EEG signals must be exploited, such as the information of nonlinearity and non-Gaussianality. The key advantages of the proposed feature extraction method lie in that:

(1) The proposed feature set contains high-order statistics information, while the widely used conventional features with only second-order measures (such as the power spectrum and autocorrelation functions) does not. As a consequence, non-minimum-phase signals, such as EEG signals, cannot be correctly characterized by the second-order measures, moreover, some types of phase coupling in EEG signals which is associated with nonlinearities cannot be correctly identified by the second-order measures.
(2) The proposed feature set is less affected by Gaussian background noise than the conventional features with only second-order measures due to the property of bispectrum: the bispectrum of Gaussian signal is zero.

The above experimental results have shown that the proposed feature set is very effective in identifying the different mental tasks from EEG signals.

5. Conclusion

In this paper a new feature extraction method is proposed for classifying EEG signals corresponding to left/right-hand motor imagery. The feature set includes higher-order statistics based on the bispectrum of EEG signals. Experimental results have shown that based on the proposed features, the LDA classifier, SVM classifier and NN classifier achieve better classification performance than the BCI-competition 2003 winner on the same BCI data set in terms of the criteria of either MI or misclassification rate.

Acknowledgement

The authors would like to thank the BCI group at the Graz University of Technology, Austria, for making their data available to the BCI research community.

References

[1] C.W. Anderson, E.A. Stolz, S. Shamsunder, Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks, IEEE Trans. Biomed. Eng. 45 (3) (1998) 277–286.
[2] F. Babiloni, F. Cincotti, L. Lazzarini, J. Millán, J. Mouriño, M. Varsta, J. Heikkonen, L. Bianchi, M.G. Marciani, Linear classification of low-resolution EEG patterns produced by imagined hand movements, IEEE Trans. Rehabil. Eng. 8 (2) (2000) 186–188.
[3] M. Basseville, A. Benveniste, Sequential segmentation of nonstationary digital signals using spectral analysis, Inform. Sci. 29 (1) (1983) 57–73.
[4] BCI Competition 2003. <http://ida.first.fraunhofer.de/projects/bci/competition/>.
[5] B. Blankertz, K.-R. Müller, G. Curio, T.M. Vaughan, G. Schalk, J.R. Wolpaw, A. Schlögl, C. Neuper, G. Pfurtscheller, T. Hinterberger, M. Schröder, N. Birbaumer, The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials, IEEE Trans. Biomed. Eng. 51 (6) (2004) 1044–1051.
[6] R.A. Fisher, The use of multiple measurements in taxonomic problems, Ann. Eugen. 7 (1936) 179–188.
[7] W.J. Freeman, A neurobiological interpretation of semiotics: meaning, representation, and information, Inform. Sci. 124 (2000) 93–102.
[8] W.J. Freeman, Comparison of brain models for active vs. passive perception, Inform. Sci. 116 (1999) 97–107.
[9] D. Garrett, D.A. Peterson, C.W. Anderson, M.H. Thaut, Comparison of linear, nonlinear, and feature selection methods for EEG signal classification, IEEE Trans. Neural Syst. Rehabil. Eng. 11 (2) (2003) 114–144.
[10] B. Graimann, J.E. Huggins, A. Schlögl, S.P. Levine, G. Pfurtscheller, Detection of movement-related desynchronization patterns in ongoing single-channel electrocorticogram, IEEE Trans. Neural Syst. Rehabil. Eng. 11 (3) (2003) 276–281.
[11] L. Kauhanen, T. Nykopp, J. Lehtonen, P. Jylänki, J. Heikkonen, P. Rantanen, H. Alaranta, M. Sams, EEG and MEG brain–computer interface for tetraplegic patients, IEEE Trans. Neural Syst. Rehabil. Eng. 14 (2) (2006) 190–193.
[12] J. Kronegg, G. Chanel, S. Voloshynovskiy, T. Pun, EEG-based synchronized brain–computer interfaces: a model for optimizing the number of mental tasks, IEEE Trans. Neural Syst. Rehabil. Eng. 15 (1) (2007) 50–58.
[13] R. Leeb, R. Scherer, F. Lee, H. Bischof, G. Pfurtscheller, Navigation in virtual environments through motor imagery, in: Proceedings of the Ninth Computer Vision Winter Workshop (CWWW'04), 4–6 February, 2004, Piran, Slovenia, pp. 99–108.
[14] P. Lingras, C. Butz, Rough set based 1-v-1 and 1-v-r approaches to support vector machine multi-classification, Inform. Sci. 177 (2007) 3782–3798.
[15] D.J. McFarland, L. McCane, L.A. Miner, T.M. Vaughan, J.R. Wolpaw, EEG mu and beta rhythm topographies with movement imagery and actual movement, Society for Neuroscience Abstracts, 1997, p. 1277.
[16] B.S. Moon, H.C. Lee, Y.H. Lee, J.C. Park, I.S. Oh, J.W. Lee, Fuzzy systems to process ECG and EEG signals for quantification of the mental workload, Inform. Sci. 142 (2002) 23–35.
[17] K.-R. Müller, C.W. Anderson, G.E. Birch, Linear and non-linear methods for brain–computer interfaces, IEEE Trans. Neural Syst. Rehabil. Eng. 11 (2) (2003) 165–169.
[18] J. Muthuswamyand, R.J. Roy, The use of fuzzy integrals and bispectral analysis of the electroencephalogram to predict movement under anesthesia, IEEE Trans. Biomed. Eng. 46 (3) (1999) 291–299.
[19] L. Nikias, A.P. Petropulu, Higher-Order Spectral Analysis – A Nonlinear Signal Processing Framework, Prentice-Hall, Englewood Cliffs, NJ, 1993.
[20] R. Palaniappan, R. Paramesran, S. Nishida, N. Saiwaki, A new brain–computer interface design using fuzzy ARTMAP, IEEE Trans. Neural Syst. Rehabil. Eng. 10 (3) (2002) 140–148.
[21] G. Pfurtscheller, EEG event-related desynchronization (ERD) and event-related synchronization (ERS), Electroencephalogr.: Basic Prin., Clin. Appl. (1999) 958–967.
[22] G. Pfurtscheller, C. Neuper, A. Schlogl, K. Lugger, Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters, IEEE Trans. Rehabil. Eng. 6 (3) (1998) 316–325.
[23] M.R. Raghuveer, C.L. Nikias, Bispectrum estimation: a parametric approach, IEEE Trans. Acoust. Speech Signal Process. 33 (1985) 1213–1230.
[24] D. Rumelhart, G. Hinton, R. Williams, Learning internal representations by error propagation, in: J. Anderson, E. Rosenfeld (Eds.), Neurocomputing, MIT Press, Cambridge, MA, 1988, pp. 675–695.
[25] A. Schlögl, Outcome of the BCI-competition 2003 on the Graz data set. <http://ida.first.fraunhofer.de/projects/bci/competition/results/TR_BCI2003_III.pdf>.
[26] A. Schlögl, C. Neuper, G. Pfurtscheller, Estimating the mutual information of an EEG-based brain–computer-interface, Biomed. Tech. 47 (1–2) (2002) 3–8.
[27] A. Schlögl, C. Keinrath, R. Scherer, G. Pfurtscheller, Information transfer of an EEG-based brain–computer interface, in: Proceedings of the First International IEEE EMBS Conference on Neural Engineering, Capri, Italy, 2003, pp. 641–644.
[28] R. Shiavi, Introduction to Applied Statistical Signal Analysis, Aksen Associates, MA, 1991.

[29] C.-Y. Tsai, On detecting nonlinear patterns in discriminant problems, Inform. Sci. 176 (2006) 772–798.
[30] A.C. Tsoi, D.S.C. So, A. Sergejew, Classification of electroencephalogram using artificial neural networks, in: J.D. Cowan, G. Tesauro, J. Alspector (Eds.), Advances in Neural Information Processing Systems, vol. 6, Morgan Kaufman, San Francisco, CA, 1994, pp. 1151–1158.
[31] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, 1995.
[32] P.J. Werbos, Beyond regression: new tools for prediction and analysis in the behavioral sciences, PhD Thesis, Harvard University, Cambridge, MA, USA, 1974.
[33] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, T.M. Vaughan, Brain–computer interfaces for communication and control, Clin. Neurophysiol. 113 (2002) 767–791.
[34] E. Yom-Tov, G.F. Inbar, Feature selection for the classification of movements from single movement-related potentials, IEEE Trans. Neural Syst. Rehabil. Eng. 10 (3) (2002) 170–177.
[35] S.-M. Zhou, J.Q. Gan, Constructing parsimonious fuzzy classifiers based on L2-SVM in high-dimensional space with automatic model selection and fuzzy rule ranking, IEEE Trans. Fuzzy Syst. 15 (3) (2007) 398–409.