Biomedical Signal Processing and Control 18 (2015) 11–18

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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

An optimized feature selection and classification method for using electroencephalographic coherence in brain–computer interfaces

Rocio Salazar-Varas, David Gutiérrez*

Center for Research and Advanced Studies (Cinvestav), Monterrey's Unit, Apodaca, N.L. 66600, Mexico

ARTICLE INFO

Article history:
Received 28 August 2014
Received in revised form 18 October 2014
Accepted 6 November 2014
Available online 29 December 2014

Keywords:
Brain–computer interfaces
Electroencephalography
Coherence
Linear discriminant

ABSTRACT

We propose a method to use electroencephalographic (EEG) coherences as features in a brain–computer interface (BCI). The coherence provides a sense of the brain's connectivity, and it is relevant as different regions of the brain must communicate between each other for the integration of sensory information. In our case, the process of feature selection is optimized in the sense that only those statistically significant and potentially discriminative coherences at a specific frequency are used, which results in a feature vector of reduced-dimension. Next, those features are classified through an optimized linear discriminant, where the best discriminating hyperplanes are selected such that the area under the receiver operating characteristics (ROC) curve is maximized. Overall, the proposed EEG coherence selection and classification method can provide efficiency rates similar to those obtained with other methods in BCI, but with the advantage of blindly selecting and optimal combination of features out of all the possible pairwise coherences. We demonstrate the applicability of the proposed method through numerical examples using real data from motor and cognitive tasks.

1. Introduction

A brain–computer interface (BCI) is a communication system that allows a subject to act on his/her environment solely by means of his/her thoughts, i.e. without using the brain's normal output pathways of muscles or peripheral nerves [1]. Non-invasive BCIs rely on electroencephalographic (EEG) measurements of the brain's activity to read out the intentions of the subject and translate them into commands for a computerized system.

The translation from the brain activity to a command is usually achieved by means of a *feature generator* that extracts feature values from the EEG signals that correspond to the underlying neurological mechanism employed by the user for control. Next, a *feature translator* classifies the features into logical control signals, such as a two-state discrete output. Many methods have been proposed so far to carry out the extraction/classification processes in BCI, and a very comprehensive review about them can be found in [2]. In general, feature extraction methods are closely related to specific neuromechanisms, while feature classification algorithms are determined by the type of features that they discriminate.

Here, we examine the use of the EEG coherence as feature in a BCI. The coherence provides a sense of the brain's connectivity, and it is relevant to measure it as different regions widely distributed over the brain must communicate between each other in order to provide the basis for integration of sensory information, as well as for many functions that are critical for learning, memory, information processing, perception, and behavior. Transient periods of synchronization of oscillating neural discharges have been proposed to act as an integrative mechanism that may bring a widely distributed set of neurons together into a coherent ensemble that underlies a cognitive act [3], and many studies have used the EEG coherence to quantify such synchronization process (see [4] and references therein). In [5], the patterns in the coherence were studied during sequential and simultaneous tasks, while in [6], signals corresponding to spontaneous EEG, imaginery movement, and movement execution were classified based on the coherence using hidden Markov models and a multilayer perceptron. Nevertheless, the only attempt known to us of using the coherence in the context of BCI can be found in [7]. There, the use of the coherence as a feature was assessed for the case of measuring the mean coupling between signals recorded from an electrode and its neighbors, and a few individual electrode pairs reflecting connectivity between fronto-centro-parietal and temporal lobes. Given the limited number of subjects tested and the coherences that were assessed, their

* Corresponding author at: Center for Research and Advanced Studies, Monterrey's Unit, Vía del Conocimiento 201, Parque de Investigación e Innovación Tecnológica (PIIT), Autopista al Aeropuerto Km. 9.5, Lote 1, Manzana 29, Apodaca, N.L. 66600, Mexico. Tel.: +52 81 1156 1740x4513; fax: +52 81 1156 1741.
E-mail address: dgtz@ieee.org (D. Gutiérrez).

http://dx.doi.org/10.1016/j.bspc.2014.11.001
1746-8094/

results do not allow for a statistical conclusion regarding general performance of the proposed measures. Nevertheless, the results in [7] suggest that coherence-based features might not perform as well as other features, but still could be relevant for classifying mental tasks.

Therefore, in this paper we propose an optimized method for feature selection and classification which is customized for the EEG coherence. The process of feature selection is optimized in the sense that only those statistically significant and potentially discriminative coherences at a specific frequency are used, which results in a feature vector of reduced-dimension. Next, those features are classified through an optimized linear discriminant, where the best discriminating hyperplanes are selected such that the area under the receiver operating characteristics (ROC) curve is maximized. Based on these ideas, the paper is organized as follows: the coherence is briefly reviewed in Section 2, then the proposed coherence-based feature selection and classification process is introduced; in Section 3, we show the applicability of our method through a series of numerical examples using real EEG data; in Section 4, we discuss the results and future work.

2. Methods

In this section we briefly review the concept of coherence, then we explain our proposed coherence-based feature selection and pose a classification procedure customized for those features.

2.1. Coherence

Let us define $x_m(n)$ as the m-th EEG measurement, for $m = 1, 2, \ldots, M$, obtained from a set of available sensors $S = \{s_1, s_2, \ldots, s_M\}$ and acquired at time samples $n = 1, 2, \ldots, N$. Then, the *auto-spectral densities* of signals $x_j(n)$ and $x_k(n)$, with $j, k \in \{1, 2, \ldots, M\}$ and $j \neq k$, are given by

$$P_{\{*\}}(f) = \sum_{\tau=-\infty}^{\infty} \mathrm{E}\{x_{\{*\}}(n)x_{\{*\}}(n-\tau)\}e^{-j2\pi\tau f}, \tag{1}$$

where $\{*\}$ indicates either j or k, $\mathrm{E}\{\cdot\}$ indicates the expected value, and f is the frequency. Note that (1) corresponds to the Fourier transform of the auto-correlation of $x_{\{*\}}(n)$. Similarly, the *cross-spectral density* is given by

$$P_{jk}(f) = \sum_{\tau=-\infty}^{\infty} \mathrm{E}\{x_j(n)x_k(n-\tau)\}e^{-j2\pi\tau f}. \tag{2}$$

Therefore, based in (1) and (2), the coherence between $x_j(n)$ and $x_k(n)$ is defined as [8]

$$\gamma^2_{j,k}(f) = \frac{|P_{jk}(f)|^2}{P_j(f)P_k(f)}. \tag{3}$$

The coherence is a measure of the degree of correlation of the spectral power in an specified bandwidth between two signals acquired from two electrodes. High coherence implies a large degree of communication between different brain regions whereas low coherence suggests relative independence [9].

In our case, we are interested in analyzing the connectivity between a selection of L sensors, i.e.,

$$S' = \left\{s'_1, s'_2, \ldots, s'_L\right\} \subset S, \tag{4}$$

with $L \ll M$ for S' to be a proper subset of sensors chosen out of S. Let us refer to the measurements on two of those sensors as $x_{l_1}(n)$ and $x_{l_2}(n)$, such that $s'_{l_1}, s'_{l_2} \in S'$ and $l_1 \neq l_2$. Then, the $D = \binom{L}{2}$ pairwise coherences $\gamma^2_{l_1,l_2}(f_s)$ for each of the selected L sensors can be computed through (3). Note that those coherences can be obtained for different frequencies of interest. In our case, we select a frequency, denoted by f_s, in which the largest differences between tasks are expected based on physiological information. Nevertheless, a method like the one proposed in [10] can be used to perform a subject-specific estimation of the principal time-varying frequencies by means of non-stationary time series models, then the most significant frequency components of the EEG could be used as f_s in the scheme here proposed.

Next, the coherences obtained at f_s for a set of sensors are arranged into D-dimensional feature vectors $\boldsymbol{y}$, where each of its elements corresponds to a pairwise coherence $\gamma^2_{l_1,l_2}(f_s)$. In our case, the optimality criterion to choose the best feature vector out of the $\binom{M}{L}$ possible ones is based in the *statistical significance* of the pairwise coherences.

2.2. Statistical significance

In addition to computing the coherence between EEG measurements, it is necessary to assess its significance in order to assure that such coherence is indeed a reflection of the connectivity between different brain areas. Such assessment is usually performed in terms of a $100(1-\alpha)\%$ confidence interval (where the significance level is denoted by α), and different methods have been previously proposed for such assessment. In this paper we use the method proposed in [11], where an ensemble of K pairs of surrogate time series (which share the features of the original EEG measurements but are completely uncoupled) are generated as realizations of two linearly independent stochastic processes. The coherence between each pair of surrogate series is calculated, then an empirical sampling distribution of the coherence is estimated from all the surrogate data. The threshold below which measurements $x_{l_1}(n)$ and $x_{l_2}(n)$ are regarded as non-coherent (denoted by η_{l_1,l_2}) is set at the $100(1-\alpha)$ percentile of the estimated sampling distribution. Hence, we are interested only in those coherence values surpassing the threshold, i.e.

$$\gamma^2_{l_1,l_2}(f_s) > \eta_{l_1,l_2}(f_s). \tag{5}$$

2.3. Optimal feature vector

Once the significance of the coherence is determined, next we are interested in those cases where the EEG signals from different events can be discriminated. Let us consider the case of $i = 1, 2, \ldots, I$ different classes, each of them comprised by those $t = 1, 2, \ldots, T_i$ trials (independent EEG measurements) meeting the condition in (5). Furthermore, if we denote the coherence between signals for a given trial and class by $\{\gamma^2_{l_1,l_2}(f_s)\}_{i,t}$, then the mean class-coherence is given by

$$\mu_i = \frac{1}{T_i}\sum_{t=1}^{T_i}\{\gamma^2_{l_1,l_2}(f_s)\}_{i,t}, \tag{6}$$

where μ_i is used instead of $\mu_i(f_s)$ for notational convenience. Based on (6), we can set up the following hypothesis test in order to determine if the events can be discriminated through their EEG coherences:

$$\begin{aligned} &H_o : \mu_1 = \mu_2 = \cdots = \mu_I \\ &H_a : \text{any negation of } H_o. \end{aligned} \tag{7}$$

Independently of the method used to accept or reject the null-hypothesis (equal/different variances, paired/unpaired samples), we rely in the p-values of the corresponding statistical test in order to select the optimal coherences of L sensors. Hence, if we denote

such p-values as p_{l_1,l_2}, the most likely coherences will be those for which

$$p_{l_1,l_2} \ll \alpha. \tag{8}$$

Therefore, the optimal feature vector $\boldsymbol{y_{OPT}}$ will be the one in which the coherences attain the condition in (5), and the p-values of the between-classes test are such that the null-hypothesis in (7) is rejected, but also meet (8). In this way, we assure that the coherences are indicative of brain connectivity, and they are most likely to be discriminated based on differences between tasks.

2.4. Classification

Once the data are mapped to the optimal vectors $\boldsymbol{y_{OPT}}$, we use a linear discriminant to separate the features into groups corresponding to the classes. In the case of $I = 2$ classes, the discriminating hyperplane is computed through the following normal vector [12]:

$$\boldsymbol{v} = (\beta\{\hat{R}_{yy}\}_1 + (1-\beta)\{\hat{R}_{yy}\}_2)^{-1}(\{\overline{\boldsymbol{y}}\}_2 - \{\overline{\boldsymbol{y}}\}_1), \tag{9}$$

where $\{\hat{R}_{yy}\}_i$ and $\{\overline{\boldsymbol{y}}\}_i$ are, respectively, the covariance matrix and the mean of the feature vectors in the i-th class, and β is the adjustment factor, which is in the range [0, 1] and is chosen based on an optimization criterion. Note that $\beta = 1$ gives the special case of the Fisher's linear discriminant (FLD). In our case, the criterion to determine β is the maximization of the area under the receiver operating characteristics (ROC) curve, in which a perfect classification is achieved when the area is equal to 1, while in the case of an area close to 0.5 the classifier is regarded to work by chance. The ROC-based evaluation is widely used in many areas [13] and it has been previously used to evaluate other BCI systems (see, e.g., [14]). Under these conditions and given a feature vector $\boldsymbol{y_{OPT}}$, the discriminant function can be written as

$$d = \boldsymbol{v}^T\boldsymbol{y_{OPT}} + v_o, \tag{10}$$

with

$$v_o = -\frac{\beta\sigma_1^2\boldsymbol{v}^T\{\overline{\boldsymbol{y}}\}_2 + (1-\beta)\sigma_2^2\boldsymbol{v}^T\{\overline{\boldsymbol{y}}\}_1}{\beta\sigma_1^2 + (1-\beta)\sigma_2^2}, \tag{11}$$

and

$$\sigma_i^2 = \boldsymbol{v}^T\{\hat{R}_{yy}\}_i\boldsymbol{v}. \tag{12}$$

Finally, the discrimination rule is the following:

$$\begin{aligned} &\text{if} \quad d > 0 \quad \boldsymbol{y}_{\text{OPT}} \in \text{Class1}, \\ &\text{otherwise} \quad \boldsymbol{y}_{\text{OPT}} \in \text{Class2}. \end{aligned} \tag{13}$$

The discriminant defined by (9)–(12) has been proven to offer better performance in comparison to the traditional FLD, specially in cases when the covariances of the two classes are markedly different. More details on this optimized version of the linear discriminant can be found in [12].

3. Numerical examples

We performed a series of numerical experiments to show the applicability of the proposed method using two different data sets of real EEG measurements. The first one corresponds to motor imaginery, and the second to cognitive tasks. For both data sets, we evaluated the search for optimal feature vectors from combinations of $L = 3$ sensors as previous studies have revealed that at least three modular structures of effective brain networks are involved in cognitive control [15]. Therefore, we analyze the case of three-sensor networks as proof of the concept.

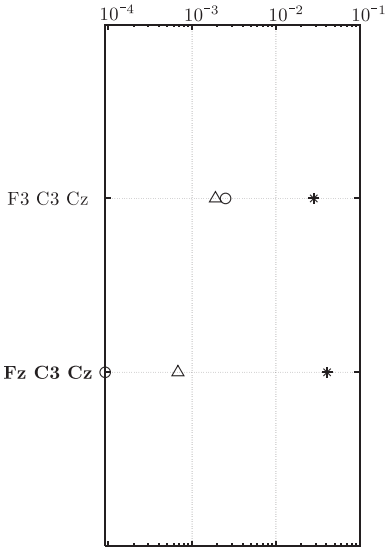
3.1. Motor imaginery tasks

We use real EEG data which corresponds to the IVa data set from the BCI competition III [16]. The data set is composed by the recordings of five healthy subjects (labeled *aa, al, av, aw* and *ay*) who were sitting in a comfortable chair with their arms relaxed on the armrests while performing a motor imaginery task. Visual cues indicated for 3.5 s which of the following two motor imageries the subject should perform: right hand (referred to as *Class 1*) or right foot (*Class 2*). The presentation of target cues was paused by periods of random length between 1.75 and 2.25 s, in which the subject could relax. The brain activity was recorded using an EEG array of 118 sensors located on the scalp according to the international 10–20 placing system, but in the following analysis we used only $M = 16$ sensors, namely $S = \{$F7, F8, Fp1, Fp2, F3, Fz, F4, C3, Cz, C4, P3, Pz, P4, O1, Oz, O2$\}$. This set of sensors are a close match to the available sensing positions of a mobile EEG system (e.g., the sixteen-sensors system of Cognionics, Incorporation). The data for each subject comprised 280 trials (140 from each class) and they were originally acquired at a sampling frequency of 1000 Hz. The data were band-pass filtered between 1 and 100 Hz, and a notch filter with 60 Hz cut-off frequency was applied to them as well. Finally, we downsampled the data to 250 Hz.

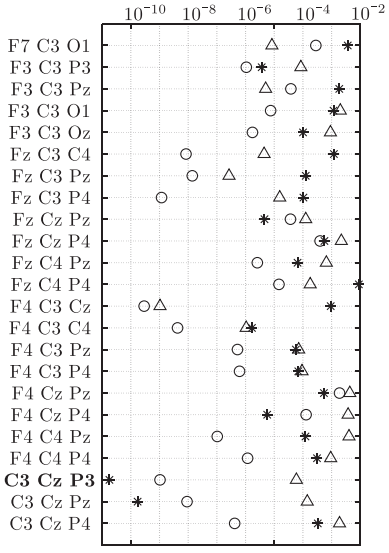
Under these conditions, we computed all the pairwise coherences of $L = 3$ selected sensors. For each of the possible $\binom{16}{3} = 560$ combinations of sensors, we evaluated the statistical significance of the $\binom{3}{2} = 3$ coherences (i.e., $\gamma^2_{1,2}$, $\gamma^2_{1,3}$, and $\gamma^2_{2,3}$) that fully characterized the connectivity between the sensors, but the evaluation was performed only for 35% of the total available trials in the frequency range of $f_s = 8, 9, 10, \ldots, 18$ Hz. The choice of frequencies was based on previous work with this data set, in which the largest differences in power between tasks were found in the range 8–18 Hz (see, e.g., [10] Fig. 2 for the case of subject *al*). In an iterative fashion, we randomly chose a trial and evaluated the significance of the coherence, as explained in Section 2.2, at the $\alpha = 0.01$ level and for $K = 400$ pairs of surrogate time series. Each time a coherence resulted statistically significant, we stored the combination of sensors and the coherence values for the corresponding frequency.

Next, we applied the between-classes hypothesis test in (7) to the stored data in order to assess if there was a significant difference between the two different classes, but only in the cases when the coherences between the sensors being evaluated resulted significant over 70% of the time they were evaluated. In all cases, a two-sample t-test was used to evaluate (7): in the case where the two data samples had equal variances (which was determined by a preliminary F-test), the test statistic under the null hypothesis had a Student's-t distribution with $T_1 + T_2 - 2$ degrees of freedom; when the two data samples were from populations with unequal variances, the test statistic had an approximate Student's-t distribution with a number of degrees of freedom given by Satterthwaite's approximation.

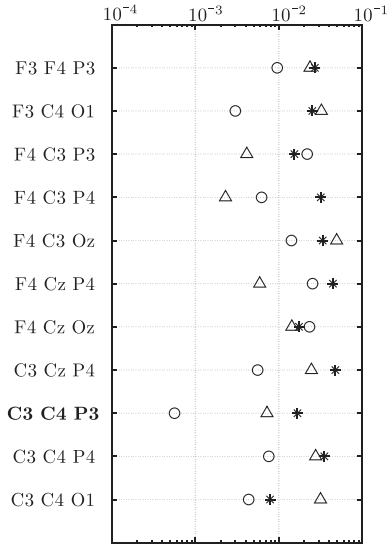
The results of the proposed feature selection scheme are shown in Fig. 1. Note that only a few of the possible 560 combinations of channels have the required p-values to reject the null-hypothesis in (7) and, from those, the optimal selection is the one for which $(p_{1,2} \ll \alpha) \wedge (p_{1,3} \ll \alpha) \wedge (p_{2,3} \ll \alpha)$. Also note that the frequency for which the best classification was attained differed among subjects. Nevertheless, all frequencies were in the range of μ-rhythm or its harmonic components, which is consistent with the imaginery task performed. It is interesting to notice that, in all cases, the optimal combination involved a significant coherence between different brain regions, either central–parietal or central–frontal (Subject *aw* presented both), which is also consistent with previously reported



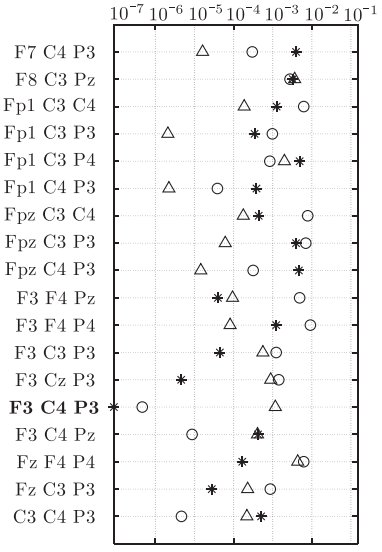
(a) Subject *aa*: $f_s = 17$ Hz, $\alpha = 0.05$.



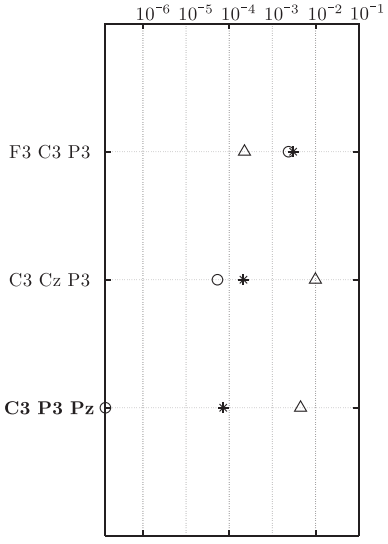
(b) Subject *al*: $f_s = 15$ Hz, $\alpha = 0.01$.



(c) Subject *av*: $f_s = 15$ Hz, $\alpha = 0.05$.



(d) Subject *aw*: $f_s = 17$ Hz, $\alpha = 0.01$.



(e) Subject *ay*: $f_s = 15$ Hz, $\alpha = 0.01$.

Fig. 1. *p*-Values (in logarithmic scale) for the subsets of EEG sensors that rejected the null-hypothesis in (7). The symbols ○, △, and * correspond to $p_{1,2}$, $p_{1,3}$, and $p_{2,3}$, respectively. In all cases, the channels whose coherences correspond to $\boldsymbol{y}_{\text{OPT}}$ are shown in **bold**.

evidence of functional coupling as result of somatosensory and motor events [17]. Finally, it is important to mention that Subjects *aa* and *av* did not fulfill the requirement of rejecting the null-hypothesis in (7) at the $\alpha = 0.01$ significance level, then we had to repeat the test with $\alpha = 0.05$ which is still acceptable according to other coherence-based studies with various physiological data (see e.g., [3,11]).

Once the optimal features were selected, we computed the corresponding feature vectors $\boldsymbol{y}_{\text{OPT}} = [\gamma^2_{1,2}, \gamma^2_{1,3}, \gamma^2_{2,3}]^T$ for each trial and class. The resulting plots are displayed in Fig. 2. There, we observe that the two classes are arranged in clusters whose main difference can be credited to their covariances. For this reason, the classification was performed through the optimized linear discriminant described in Section 2.4, where the area under the ROC curves is used to evaluate the performance. The classifier was trained with 50% of the available data for each class (140 trials in total), and the testing data corresponded to the remaining 50%. In our case, the training process corresponded to calculate a consistent estimate of the covariance matrices $\{R_{yy}\}_i$ using the data selected for this purpose. Both the training and classification processes were repeated 200 times where the data of each trial was randomly placed in either the training or the testing data set, then the area under the ROC curve was computed each time. The corresponding mean and the standard deviation obtained from performing this iterative evaluation procedure for each subject are shown in Table 1. Our results show that the areas, in most cases, were greater than 0.7. Furthermore, these results are close to those obtained in [10], but with the advantage that here only three sensors were used.

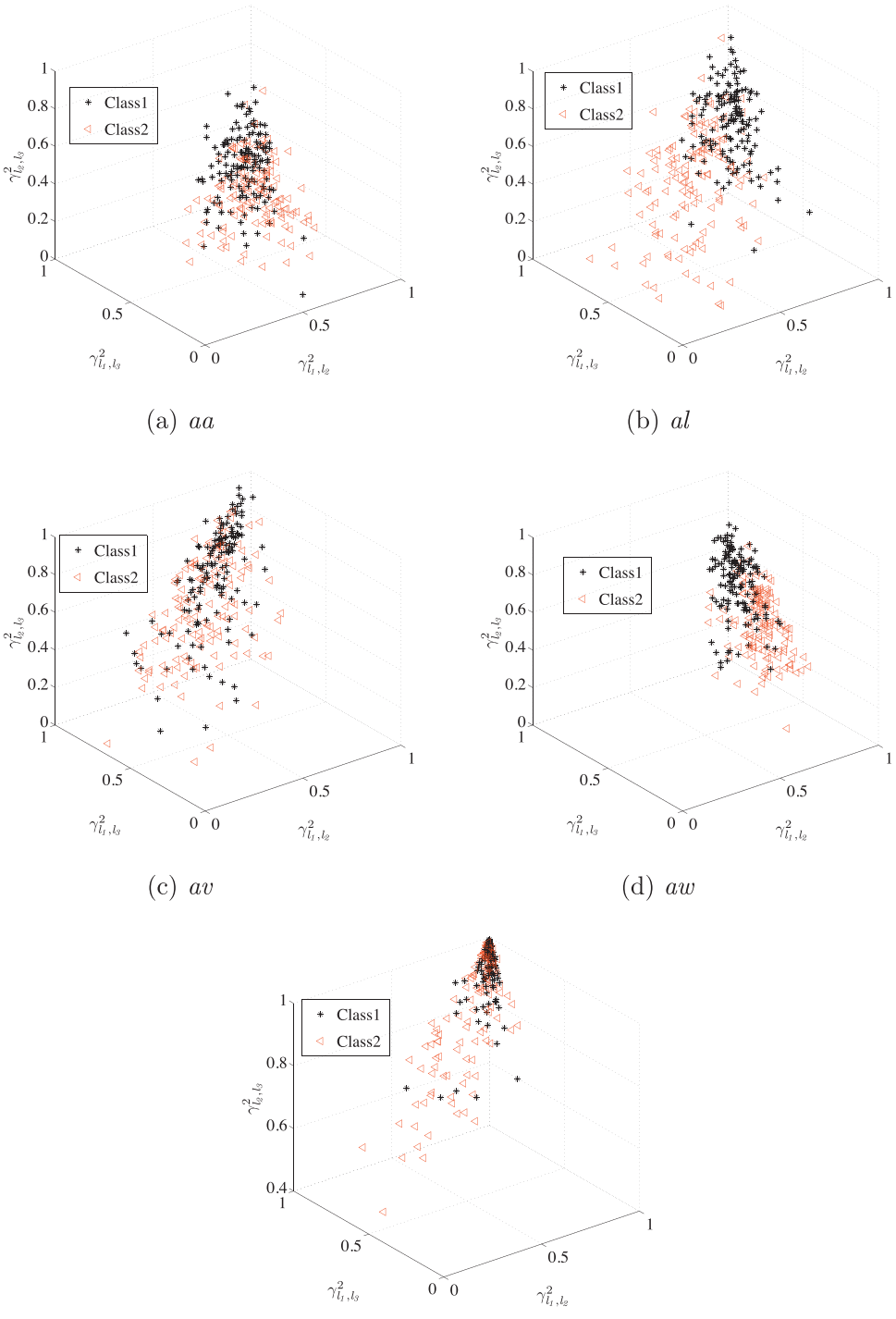


Fig. 2. Optimal coherence-based feature vectors for each subject.

Table 1
Area under the ROC curves with 95% confidence intervals in parenthesis.

Subject	Coherence-based method	Method in [10]
aa	0.725 (0.721, 0.729)	0.813 (0.806, 0.820)
al	0.905 (0.902, 0.908)	0.958 (0.954, 0.962)
av	0.659 (0.655, 0.663)	0.694 (0.690, 0.698)
aw	0.771 (0.767, 0.775)	0.841 (0.838, 0.844)
ay	0.730 (0.726, 0.734)	0.861 (0.856, 0.866)

Table 2
Classification accuracy under the BCI competition's conditions.

Subject	Correct classification (%)
aa	73
al	89
av	63
aw	74
ay	69

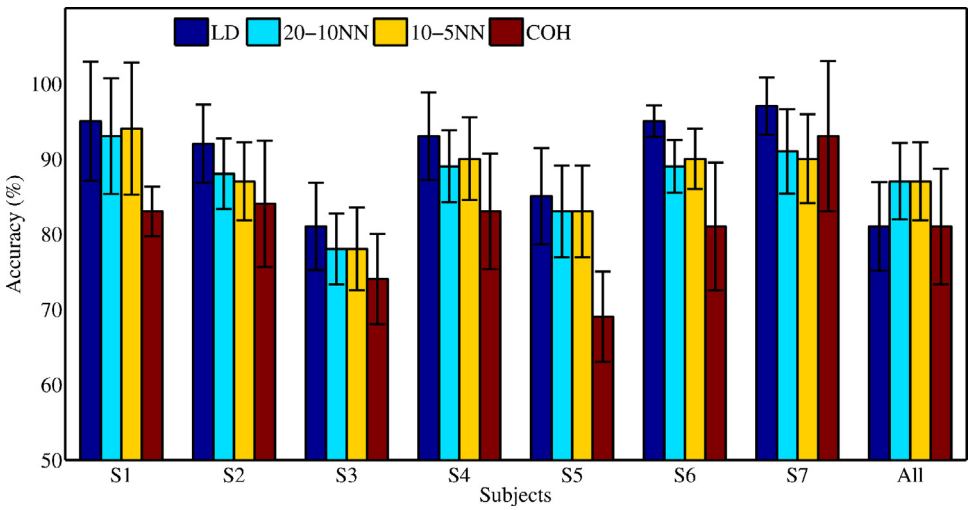


Fig. 3. Mean classification accuracy among all possible two-tasks combinations for each subject. Error bars show a ± standard deviation. The bars labeled as **COH** correspond to the results of the method here proposed, while **LD**, **20-10NN**, and **10-5NN** correspond to the results using the feature vectors and different classification strategies proposed in [19].

Another test performed on this data set corresponded to evaluate the classification accuracy according to the settings of the BCI Competition III. The results of this test for each patient are shown in Table 2, while the averaged accuracy corresponds to 73.6%. This result would have ranked the proposed method in the fourth position of the competition.

3.2. Cognitive tasks

In these numerical examples, the EEG data from [18] was used, which is freely available at http://www.cs.colostate.edu/eeg/main/data/1989_Keirn_and_Aunon. The data set is from seven subjects (labeled *Subj1, Subj2, Subj3, Subj4, Subj5, Subj6* and *Subj7*) performing five different events: resting (baseline measurement), computing a nontrivial multiplication problem (referred to as *Math*), mentally composing a letter to a friend (*Lett*), rotating a three-dimensional solid (*Rot*), and visualizing a sequence of numbers being written on a blackboard (*Count*). The subjects were placed in a dim, sound controlled room, and EEG measurements were acquired through electrodes placed at positions $S = \{C3, C4, P3, P4, O1, O2\}$, and referenced to two electrically linked mastoids at A1 and A2. The data were recorded at a sampling rate of 250 Hz with a Lab Master 12-bit A/D converter mounted in an IBM-AT computer. Eye blinks were detected by means of a separate channel of data recorded from two electrodes placed above and below the subject's left eye. Then, the subjects were asked to perform the five separate mental tasks previously described. Each task was recorded for 10 s and the experiment was repeated five times per session. The subjects participated in a different number of sessions: *Subj1*, *Subj3*, *Subj4*, and *Subj6* participated in two sessions; *Subj2* and *Subj7* in one; *Subj5* in three. The data were divided in 2-s segments that overlap by one quarter-second. Furthermore, the data were band-pass filtered between 1 and 100 Hz, and with a notch filter with 60 Hz cut-off frequency.

Under those conditions, we applied a similar procedure as the one applied to the motor imaginery data in Section 3.1. In this case, the total number of combinations that we assessed were $\binom{6}{3} = 20$. Given that the data corresponded to four different cognitive tasks, we evaluated all the possible two-class combinations. The resulting three-sensors combination corresponding to $\boldsymbol{y}_{\mathrm{OPT}}$, and the selected frequencies for all subjects are shown in Table 3. These results for the cognitive data show that, for the different task combinations, most of the sensors in the optimal features included the 01 and 02 sensors, or at least one of these occipital channels. This may be an indication of brain connectivity in relationship with the comprehension of visualized words, numbers, or an object, which involves a strong activity of the brain's occipital area. However, given the reduced number of available channels, such observation is only speculative.

In terms of evaluating the classification, the area under ROC curve corresponding to all the cases analyzed are shown in Table 4. As in the results of the data in Section 3.1, here we note that the classification achieves a good performance and with the advantage of using only three sensors blindly selected. Furthermore, our results can be also compared against those recently reported in [19] for the same data sets. There, the authors report an average accuracy in the classification of the same two-classes combinations of up to 98.7% for *Subj7* (best performing subject). However, obtaining such results involve the use of a feature vector obtained from decomposing each of the six EEG signals into five intrinsic mode functions [20], then computing six different metrics to each of them (root mean square, variance, Shannon entropy, Lempel–Ziv complexity, central frequency, and maximum frequency). Such process ends up generating (6 EEG channels) × (5 mode functions) × (6 metrics) = 180 elements in the feature vector. Hence, in [19] the authors propose a secondary feature selection method by which they are capable to reduce the dimension of their feature vector to 16 ± 7 (depending on the subject and on the mental task), but with a reduction in the average performance to 90.8% for *Subj7*. In our case, the overall performance for *Subj7* was 89.1%, but using a three-dimensional feature vector and with the capability of gaining insight of the brain process through the selected sensors and their connectivity.

A final test performed on this data set corresponded to evaluate the classification accuracy according to the settings in [19], whose result in terms of the mean and standard deviation among all possible tasks combinations for each subject is shown in Fig. 3. Again, note that our results are not significantly different from those obtained with more complex feature vectors and classification strategies.

Table 3
Subset of sensors (S') used for $\boldsymbol{y}_{\mathrm{OPT}}$ and selected frequencies for each subject.

Classes	*Subj1*		*Subj2*		*Subj3*		*Subj4*		*Subj5*		*Subj6*		*Subj7*	
	S'	f_s	S'	f_s	S'	f_s	S'	f_s	S'	f_s	S'	f_s	S'	f_s
Math vs. Lett	{C3, P4, O1}	18 Hz	{P4, O1, O2}	18 Hz	{C4, P3, P4}	16 Hz	{C3, O1, O2}	13 Hz	{C4, P3, O1}	9 Hz	{P4, O1, O2}	10 Hz	{C4, O1, O2}	18 Hz
Math vs. Rot	{P3, O1, O2}	18 Hz	{C4, O1, O2}	18 Hz	{P3, P4, O1}	18 Hz	{C4, P3, O1}	18 Hz	{P4, O1, O2}	14 Hz	{C3, C4, P4}	11 Hz	{C3, C4, P3}	16 Hz
Math vs. Count	{C4, O1, O2}	15 Hz	{C3, P3, O1}	18 Hz	{C3, C4, P4}	15 Hz	{P3, P4, O1}	9 Hz	{P3, O1, O2}	18 Hz	{P4, O1, O2}	18 Hz	{C4, O1, O2}	17 Hz
Lett vs. Rot	{P3, O1, O2}	18 Hz	{C4, O1, O2}	18 Hz	{P3, P4, O1}	16 Hz	{C4, P3, O1}	18 Hz	{P3, P4, O2}	9 Hz	{C3, P3, O1}	18 Hz	{C4, O1, O2}	18 Hz
Lett vs. Count	{C3, P4, O2}	9 Hz	{P4, O1, O2}	18 Hz	{C4, P3, O1}	18 Hz	{C3, P4, O2}	13 Hz	{C3, C4, P4}	13 Hz	{P3, P4, O2}	9 Hz	{C3, P3, O1}	10 Hz
Rot vs. Count	{C3, O1, O2}	18 Hz	{C3, O1, O2}	13 Hz	{C4, P3, P4}	16 Hz	{C3, C4, P4}	11 Hz	{P3, P4, O2}	9 Hz	{C3, P3, P4}	18 Hz	{C3, O1, O2}	18 Hz

Table 4
Area under the ROC curves with 95% confidence intervals in parenthesis.

Classes	*Subj1*	*Subj2*	*Subj3*	*Subj4*	*Subj5*	*Subj6*	*Subj7*
Math vs. Lett	0.843 (0.840, 0.846)	0.884 (0.880, 0.888)	0.716 (0.712, 0.720)	0.896 (0.892, 0.900)	0.689 (0.686, 0.692)	0.732 (0.729, 0.735)	0.981 (0.977, 0.985)
Math vs. Rot	0.894 (0.891, 0.897)	0.779 (0.773, 0.785)	0.707 (0.703, 0.711)	0.713 (0.709, 0.717)	0.719 (0.716, 0.722)	0.871 (0.868, 0.874)	0.699 (0.694, 0.704)
Math vs. Count	0.801 (0.798, 0.804)	0.694 (0.689, 0.699)	0.734 (0.730, 0.738)	0.772 (0.768, 0.776)	0.586 (0.583, 0.589)	0.852 (0.849, 0.855)	0.995 (0.992, 0.998)
Lett vs. Rot	0.837 (0.834, 0.840)	0.878 (0.874, 0.882)	0.699 (0.696, 0.702)	0.869 (0.865, 0.873)	0.742 (0.739, 0.745)	0.822 (0.819, 0.825)	0.920 (0.915, 0.925)
Lett vs. Count	0.786 (0.783, 0.789)	0.809 (0.804, 0.814)	0.807 (0.804, 0.810)	0.918 (0.914, 0.922)	0.638 (0.635, 0.641)	0.681 (0.678, 0.684)	0.778 (0.773, 0.783)
Rot vs. Count	0.781 (0.778, 0.784)	0.738 (0.734, 0.742)	0.675 (0.671, 0.679)	0.629 (0.625, 0.633)	0.681 (0.678, 0.684)	0.832 (0.829, 0.835)	0.973 (0.970, 0.976)

4. Conclusions

We presented a method based on EEG coherence to select an optimal feature vector with reduced-dimension for BCI applications. The selection was performed without prior knowledge about brain activity related to the tasks involved. However, by assessing the significance of the EEG coherence, we assure that the signals from selected channels indeed provide information related to the brain connectivity associated to the task. Our experiments with real EEG data show that the selected sensors are capable of discriminating different tasks (motor and cognitive) with good accuracy rates, while using a reduced number of sensors. Our results also show that, against previous belief, coherence-based features can provide a good performance, and it can be comparable to the one provided with other methods based on features with higher dimensionality.

Our method is well suited for practical BCI applications where mobile EEG systems are used. In this type of systems, the number of sensors available might not be large, then it is important to find a subset of sensors that allows to reduce the computational load but, at the same time, optimizes the classification process by taking into account the connectivity information of the brain processes in a personalized manner. For the case when the search for an optimal feature is being performed in a highly dense EEG array (e.g., for the whole 118 sensors available in the data set in Section 3.1), our proposed method might seem exhaustive. However, for those cases we recommend an initial randomized search of possible combinations of L sensors in which candidates of the optimal solution are pre-selected, then those candidates can be used in a directed search through metaheuristic methods. In preliminary tests using this approach, we have been able to reach an optimal feature out of 10,000 random picks instead of exhaustively testing the whole $\binom{118}{3} = 266,916$ combinations. This opens the possibility of exploring cases with $L > 3$, as well as more dense measuring arrays, such as magnetoencephalographic (MEG) systems, for which coherence measures seem to reflect different aspects of neocortical dynamics [21].

Our future work will extend our approach to network theory, in which L channels could be seen as forming a *sensor network*, and their pairwise coherences could serve as connectivity weights. Having that in mind, our work will include a more rigorous assessment of the connectivity through other metrics, such as the partial directed coherence [22], so we could propose different network topologies and not only fully-connected networks as in this work.

Acknowledgment

This work was supported by the Mexican Council of Science and Technology (CONACyT) under Grant 220145.

References

[1] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, T.M. Vaughan, Brain–computer interfaces for communication and control, Clin. Neurophysiol. 113 (6) (2002) 767–791.
[2] A. Bashashati, M. Fatourechi, R.K. Ward, G.E. Birch, A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals, J. Neural Eng. 4 (2) (2007) R32.
[3] E. Rodriguez, N. George, J.-P. Lachaux, J. Martinerie, B. Renault, F.J. Varela, Perception's shadow: long-distance synchronization of human brain activity, Nature 397 (6718) (1999) 430–433.
[4] P.L. Nunez, R. Srinivasan, A.F. Westdorp, R.S. Wijesinghe, D.M. Tucker, R.B. Silberstein, P.J. Cadusch, EEG coherency I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales, Electroencephalogr. Clin. Neurophysiol. 103 (5) (1997) 499–515.
[5] S.T. Okuhata, S. Okazaki, H. Maekawa, EEG coherence pattern during simultaneous and successive processing tasks, Int. J. Psychophysiol. 72 (2) (2009) 89–96.
[6] A.P. Souza, L.B. Felix, C.A. Maia, C.J. Tierra-Criollo, Classification of imaginary movements using the magnitude-squared coherence feature extractor, in: Biosignals and Biorobotics Conference, IEEE, 2012, pp. 1–6.
[7] E. Gysels, P. Celka, Phase synchronization for the recognition of mental tasks in a brain–computer interface, IEEE Trans. Neural Syst. Rehabil. Eng. 12 (4) (2004) 406–415.
[8] S. Sanei, J.A. Chambers, EEG Signal Processing, John Wiley & Sons, 2008.
[9] S. Weiss, H.M. Mueller, The contribution of EEG coherence to the investigation of language, Brain Lang. 85 (2) (2003) 325–343.
[10] D. Gutiérrez, R. Salazar-Varas, Using eigenstructure decompositions of time-varying autoregressions in common spatial patterns-based EEG signal classification, Biomed. Signal Process. Control 7 (6) (2012) 622–631.
[11] L. Faes, G.D. Pinna, A. Porta, R. Maestri, G. Nollo, Surrogate data analysis for assessing the significance of the coherence function, IEEE Trans. Biomed. Eng. 51 (7) (2004) 1156–1166.
[12] T. Cooke, M. Peake, The optimal classification using a linear discriminant for two point classes having known mean and covariance, J. Multivar. Anal. 82 (2) (2002) 379–394.
[13] T. Fawcett, An introduction to ROC analysis, Pattern Recognit. Lett. 27 (8) (2006) 861–874.
[14] D. Gutiérrez, D.I. Escalona-Vargas, EEG data classification through signal spatial redistribution and optimized linear discriminants, Comput. Methods Programs Biomed. 97 (1) (2010) 39–47.
[15] J. Ying Liu, S. Moser, Aviyente, Network community structure detection for directional neural networks inferred from multichannel multisubject EEG data, IEEE Trans. Biomed. Eng. 61 (July (7)) (2014) 1919–1930.
[16] J.P. Donoghue, B. Blankertz, G. Curio, K. Muller, Boosting bit rates in non-invasive EEG single trial classification by feature combination and multi class paradigm, IEEE Trans. Biomed. Eng. 51 (6) (2004) 993–1002.
[17] C. Babiloni, A. Brancucci, F. Vecchio, L. Arendt-Nielsen, A.C.N. Chen, P.M. Rossini, Anticipation of somatosensory and motor events increases centro-parietal functional coupling: an EEG coherence study, Clin. Neurophysiol. 117 (5) (2006) 1000–1008.
[18] Z.A. Kerin, J.I. Aunon, A new mode of communication between man and its surrounding, IEEE Trans. Biomed. Eng. 37 (1990) 1209–1214.
[19] P.F. Diez, V.A. Mut, E. Laciar, A. Torres, E.M. Avila-Perona, Features extraction method for brain–machine communication based on the empirical mode decomposition, Biomed. Eng. Appl. Basis Commun. 25 (6) (2013), pp. 1350058/1-13.
[20] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung, Liu H.H., The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. Lond. Ser. A: Math. Phys. Eng. Sci. 454 (1971) (1998) 903–995.
[21] R. Srinivasan, W.R. Winter, J. Ding, P.L. Nunez, EEG and MEG coherence: measures of functional connectivity at distinct spatial scales of neocortical dynamics, J. Neurosci. Methods 166 (1) (2007) 41–52.
[22] B. Schelter, J. Timmer, M. Eichler, Assessing the strength of directed influences among neural signals using renormalized partial directed coherence, J. Neurosci. Methods 179 (1) (2009) 121–130.