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

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 $D = \begin{pmatrix} L \\ 2 \end{pmatrix}$ 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 $\begin{pmatrix} M \\ L \end{pmatrix}$ 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