Robust classification of motor imagery EEG signals using statistical time-domain features

Article *in* Physiological Measurement · October 2013

DOI: 10.1088/0967-3334/34/11/1563 · Source: PubMed

Aida Khorshidtalab

Momoh Salami
International Islamic University Malaysia

Mahyar Hamedi

Robust classification of motor imagery EEG signals using statistical time–domain features

2013 Physiol. Meas. 34 1563

(http://iopscience.iop.org/0967-3334/34/11/1563)

IOP Publishing — Physiological Measurement
Physiol. Meas. **34** (2013) 1563–1579 doi:10.1088/0967-3334/34/11/1563

Robust classification of motor imagery EEG signals using statistical time–domain features

A Khorshidtalab[1], M J E Salami[1] and M Hamedi[2]

[1] Department of Mechatronics Engineering, International Islamic University Malaysia, Gombak, Malaysia
[2] Faculty of Bioscience and Medical Engineering, Universiti Teknolog Malaysia, 81310 Skudai, Johor, Malaysia

E-mail: aida.khorshid@student.iium.edu.my, momoh@iium.edu.my and hamedi.mahyar@ieee.org

Received 7 July 2013, accepted for publication 2 October 2013
Published 24 October 2013
Online at stacks.iop.org/PM/34/1563

Abstract
The tradeoff between computational complexity and speed, in addition to growing demands for real-time BMI (brain–machine interface) systems, expose the necessity of applying methods with least possible complexity. Willison amplitude (WAMP) and slope sign change (SSC) are two promising time–domain features only if the right threshold value is defined for them. To overcome the drawback of going through trial and error for the determination of a suitable threshold value, modified WAMP and modified SSC are proposed in this paper. Besides, a comprehensive assessment of statistical time–domain features in which their effectiveness is evaluated with a support vector machine (SVM) is presented. To ensure the accuracy of the results obtained by the SVM, the performance of each feature is reassessed with supervised fuzzy C-means. The general assessment shows that every subject had at least one of his performances near or greater than 80%. The obtained results prove that for BMI applications, in which a few errors can be tolerated, these combinations of feature–classifier are suitable. Moreover, features that could perform satisfactorily were selected for feature combination. Combinations of the selected features are evaluated with the SVM, and they could significantly improve the results, in some cases, up to full accuracy.

Keywords: electroencephalogram, brain–machine interface, feature extraction, support vector machine, fuzzy C-means, mutual information

1. Introduction

Brain–machine interface (BMI) is a direct communication pathway between the brain and an external electronic device aiming at translating brain activities into control commands

(Gilja *et al* 2011). It is for more than four decades that the possibility of brain–computer communication based on electroencephalogram (EEG) signals has been examined (Vidal 1973). EEG is the foundation of most of the current brain-controlled applications. EEG signals are known by their poor signal-to-noise ratio and their high dimensionality. Besides, their non-stationary characteristics and rapid variation over time and over sessions of recording pose a real challenge (Wolpaw *et al* 2000).

To achieve a good classification performance, the set of input features and the choice of the applied classifier are crucial (Pfurtscheller and Neuper 2001, Herman *et al* 2008). If a feature provides large interclass difference for different classes, the applied classifier can exhibit a better performance (Herman *et al* 2008). Several different algorithms with a variety of complexity and efficiency in different domains have been suggested for motor imagery EEG signal analysis (Bashashati *et al* 2007, Lotte *et al* 2007, Dornhege *et al* 2002, McFarland *et al* 2006, Rechy-Ramirez and Hu 2011). Classifiers mostly identify and differentiate different patterns of brain activities. Hence, a BMI system can be considered as a pattern recognition system partially. Performance of pattern recognition systems directly depends on the effectiveness of feature extraction and the applied classifier (Herman *et al* 2008). Among the tested features for EEG signals, time–domain features have low computational complexity. Thus, they could be considered as an appropriate option for real-time BMI systems (Geethanjali *et al* 2012).Although there have been several attempts to evaluate and compare the effectiveness of time–domain features, such as adaptive autoregressive model (Schlögl *et al* 1997a, 1997b, 2005, Ramoser *et al* 2000, Lee *et al* 2005) and common spatial filtering (Ramoser *et al* 2000, Lee *et al* 2005, Ziehe *et al* 2004, Pham and Cardoso 2001), there is almost no study investigating the feasibilities of statistical time–domain features, while this evaluation is necessary to verify their potential impacts.

The organization of this paper is as follows. In section 2, EEG signal acquisition, required experimental setups and applied feature extraction methods are explained. Section 2 also includes the proposed modified WAMP and modified SSC in addition to the applied classifiers. Experimental results, evaluation matrices and discussions are presented in section 3. Eventually, the paper ends with the conclusion and future works in the last section.

2. Materials and methods

2.1. EEG signal acquisition

Almost all multiclass movement-related potentials records are imaginations of left-hand movements, right-hand movements, feet (for three classes) or/and (for four classes) tongue movements. One of such major widely applied records is the BCI competition 2005 (Blankertz 2005).

In this study, by considering the fact that using more than two classes requires a suitable number of well-discriminable brain states, different sets of movement-related potentials are recorded. In essence, multiclass decisions should be derived from a decision space natural to human subjects such as movements of different body parts which have a somatotopically ordered layout in the primary motor cortex resulting in spatially discriminable patterns of EEG signals.

Instead of the aforementioned commonly used movement-related potentials, in this study, subjects were asked to think about the movements of their right hand and left hand, and the movement of their tongue to the right side of their mouth and the left side of their mouth.

2.1.1. Subject preparation and electrode placement. EEG signals are recorded from multiple electrodes placed on the subject's scalp, resulting in multichannel time series data. Three electrodes, known as C3, Cz and C4, were located on the subject's scalp based on the international 10–20 electrode placement system (Jasper 1958), in a mono-polar montage. These three electrodes cover the sensory motor cortex and they been recommended for recording motor imagery movements (Müller-Gerking *et al* 1999, Homan *et al* 1987). The reference electrode was located on the left mastoid, behind the ear, and the ground electrode was placed at Fpz, near the forehead.

The experiment consists of three runs for each of the four movements, 1 min in length. Each run is treated as one trial in the processing phase. The EEG data were recorded from 15 subjects, of which six are female. The subjects were in the range of 20–36 years old. During the recording signal, the subjects were asked to replace the desired movement with the imagination of the related movements.

2.1.2. Experimental setup and EEG acquisition. EEG signals were recorded via the g.tec® device, which is known to be one of the most accurate devices with high resolution available for recording bio-signals. Signals were recorded at the rate of 512 Hz. Subjects were free of medication and central nervous system abnormalities, and had no prior experience with EEG-based systems. In the recording environment, only intense sound disturbances were avoided. There has also been no use of bio-feedback to help the subjects perform these thinking tasks better. All runs for a subject were conducted on the same day with several breaks in between.

2.2. Feature extraction techniques

Measuring brain activity through EEG signals leads to the acquisition of a large amount of data. Feature extraction highlights important data and eliminates redundant or non-informative data, which is a transformation of the raw signal to a feature vector. This transformation causes dimensionality reduction, which naturally speeds up the classification process (Herman *et al* 2008). Time–domain features are computed based on the signals' amplitudes, and they require no transformation or complex calculation.

Among the most common statistical time–domain features, 15 were selected and examined for EEG data in the current work. As it is known that more gain can be expected from the combination of single features, if these features provide complementary information (Rechy-Ramirez and Hu 2011), therefore, several features are selected for feature combination. The applied features (Rechy-Ramirez and Hu 2011) are available in the appendix.

For the experimental evaluation, each trial is divided into hundred segments. Each segment's length is 255 ms, and segments do not overlap. This study only considers an offline analysis of the BCI experiment since the offline processing is the most stable form for any evaluation or assessment (Blankertz *et al* 2004).

2.2.1. Slope sign change (SSC). SSC is a feature that represents a frequency aspect of the EEG signal with the number of times that the slope of a waveform changes its sign. In the SSC, a threshold is included to reduce the noise. Given three consecutive samples, $x_{i-1},$ x_i and x_{i+1}, the SSC is incremented if

$$\begin{gathered}\{x_i > x_{i-1} \text{ and } x_i > x_{i+1}\} \quad \text{or} \quad \{x_i < x_{i-1} \text{ and } x_i < x_{i+1}\} \\ \text{and} \\ |x_i - x_{i+1}| \geqslant T \quad \text{or} \quad |x_i - x_{i-1}| \geqslant T. \end{gathered} \tag{1}$$

2.2.2. The Willison amplitude (WAMP). The WAMP, considering equation (2), counts the number of times that the absolute value of difference between the EEG signal amplitude of two consecutive samples exceeds a predetermined threshold value:

$$\mathrm{WAMP}_K = \sum_{i=1}^{N-1} f(x)$$
$$f(x) = \begin{cases} 1 & |x_i - x_{i+1}| > T \\ 0 & \text{otherwise} \end{cases}. \tag{2}$$

In equations (1) and (2), N is the length of the segment, k is the current segment, x_i is the current point of signal, i is the index of the current point and T is the threshold value.

2.3. Proposed modified WAMP and modified SSC

Both WAMP and SSC are time–domain features which have been suggested for EMG signals among bio-signals (Rechy-Ramirez and Hu 2011). In Khorshidtalab *et al* (2012a), (2012b) and Vigneshwari *et al* (2013), they have been applied for EEG signals and turned to be remarkably discriminant to EEG data. The noise level of the signal has direct effect on the performance of these two methods. The right threshold value for a signal could be very different from the right threshold value for the same signal after filtering and signal conditioning (Beim Graben and Kurths 2003). Although these methods have been suggested and are known to be discriminant, there is no suggestion or method for defining an exact, applicable threshold value in the literature for them. In most papers and works, heuristic methods are applied by researchers for finding a suitable value for T (Rechy-Ramirez and Hu 2011, Khorshidtalab *et al* 2012a, 2012b, Vigneshwari *et al* 2013). Not only is going through trial and error for an acceptable value time consuming, but it also does not guarantee stable processing, without further supervision, since finding the value of T through trial and error might face failure by not finding any appropriate threshold values at all.

In order to have feature values with distribution close to the normal distribution, one of the possibilities is to apply the log-transform values instead (Pfurtscheller and Neuper 2001). In other words, when the feature set is extracted from the EEG signal and is ready to be passed to the classifier, in this stage, the actual data can be replaced with their equal logarithmic value; this transformation has been done for all the feature values in the current work. This simple manipulation of data can make a difference in many aspects.

2.3.1. Modified WAMP. If a piece of EEG signal which is divided to several segments is available, by using the following method an exact T can be found.

The T value should be small enough so that this method delivers at least one count per each segment. To find the T value that meets the condition, it is necessary to know all the data points in all segments, which are not approximately equal with their next points, within a predefined quantization interval. Considering these points as a set, this set is called a set of '*acceptable points*'. Then we calculate the distance of each accepted point from its next point. Therefore, for each segment, we have a set of '*acceptable points*' and a set of '*distance*'.

If we choose the maximum value of each *distance* set for each segment, there would be a new set which is for the whole signal and includes a value from each set. The minimum value of the last set is to be chosen as the T.

If there are several classes to be compared, to have a valid and suitable T for all classes, this process should be repeated for the signal of each class. Among the Ts belonging to each class, the minimum T should be chosen for the final T. We demonstrate that this T is better compared to other possible values for T. Since this value is the minimum possible value in the last set, it causes one count for the segment it is chosen from and one or more counts for the other segments. Any greater value than the selected T would deliver zero as the feature value for some segments. In this case, it is evident that two different signals which belong to two different classes seem identical for the classifier that faces them with zero values.

Any value less than the selected T will possibly deliver more number of counts for each segment. Comparing more counts when we have the minimum counts, where none of them is zero, is also discriminant enough to be no advantage.

The procedure of finding the right T is as follows.

Let A_i = sets of acceptable data points for each segment. Therefore,
$A_1 = \{a_{11}, \ldots, a_{1k}\}, A_2 = \{a_{21}, \ldots, a_{2k}\}, A_3 = \{a_{31}, \ldots, a_{3k}\}, \ldots, A_N = \{a_{N1}, \ldots, a_{Nk}\}$,
where a_{ij} is the jth point which belongs to the ith segment, which is not equal to its immediate next data point; $i = \{1, \ldots, N\}$ where $1 \leqslant j \leqslant M$.
Let D_i = sets of distances between each accepted point and its next point in the ith segment. Therefore,
$D_1 = \{d_{11}, \ldots, d_{1k-1}\}, D_2 = \{d_{21}, \ldots, d_{2k-1}\}, D_3 = \{d_{31}, \ldots, d_{3k-1}\}, \ldots, D_N = \{d_{N1}, \ldots, d_{Nk-1}\}$.
Let Z_i = maximum value of each set.
$Z_1 = \max(\{d_{11}, \ldots, d_{1k-1}\}), Z_2 = \max(\{d_{21}, \ldots, d_{2k-1}\}), \ldots, Z_N = \max(\{d_{N1}, \ldots, d_{Nk-1}\})$
and let $Z_{\text{class}1} = \{Z_1, Z_2, \ldots, Z_N\}$.
$T_{\text{class}1} = \min(Z_{\text{class}1}) = \min(\{Z_1, Z_2, \ldots, Z_N\})$.
For each class, the same procedure should be completed. Having P classes leads us to
$T_{\text{class}2} = \min(Z_{\text{class}2}) = \min(\{Z_1, Z_2, \ldots, Z_N\}), \ldots$ and $T_{\text{class}P} = \min(Z_{\text{class}P}) = \min(\{Z_1, Z_2, \ldots, Z_N\})$.
The final $\boldsymbol{T}$ is defined as follows:
$\boldsymbol{T} = \min(\{T_{\text{class}1}, T_{\text{class}2}, T_{\text{class}3}, \ldots, T_{\text{class}P}\})$,
and the condition should be changed to

$$f(x) = \begin{cases} 1 & |x_i - x_{i+1}| \geqslant T \\ 0 & \text{otherwise} \end{cases}.$$

2.3.2. Modified SSC. By considering a piece of EEG signal which is divided into several segments, trying this method could be a promising way to find a certain T value that causes at least one count for each segment.

To find the T value that meets the mentioned condition, we need to know all the points in every segment that satisfy the SSC's condition. If we consider these points as a set, we call this set a set of '*acceptable points*'. By calculating the distance of each accepted point from its previous point and to its next point, we can come up with a set named '*distances*'. In the *distance* set, each element itself is a set of two positive values. If any of these points should be counted, the T should be smaller than the smaller distance in each set of two distances each. Thus, between each two distance values, the smaller one should be chosen to proceed to the next set. Now, there is a new set for each segment which consists of the distance of each acceptable point from either its previous point or to its next point, the one that makes the smaller distance. If we choose the maximum value of each set for each segment, we will have a new set. In this new set which is for the whole length signal and includes only one value from each segment, the minimum value should be chosen for T.

Let S_i = sets of acceptable points for each segment.

$S_1 = \{s_{11}, \ldots, s_{1k}\}$, $S_2 = \{s_{21}, \ldots, s_{2k}\}$, S3 $= \{s_{31}, \ldots, s_{3k}\}, \ldots, S_N = \{s_{N1}, \ldots, s_{Nk}\}$,

s_{ij} = the jth point which belongs to the ith segment that could satisfy the SSC's condition, where M is the number of data points in each segment, $1 \leqslant j \leqslant M{-}2$ and $1 \leqslant I \leqslant N$.

Let D_i = set of distances of previous and next points to each acceptable point.

d_{ijp} = distance of the jth acceptable point in the ith segment with its immediate previous point.

d_{ijn} = distance of the jth acceptable point in the ith segment with its immediate next point:

$\min\,(d_{ijp}, d_{ijn}) = d_{ij}$,

where d_{ij} = the minimum distance value of each acceptable point, which it makes with its next or previous data point.

$MD_1 = \{d_{11}, \ldots, d_{1k}\}, MD_2 = \{d_{21}, \ldots, d_{2k}\}, \ldots, MD_N = \{d_{N1}, \ldots, d_{Nk}\}$.

Let us assume $\max(MD_i) = \max(\{d_{il}, \ldots, d_{ik}\}) = \{x_i\}$.

Then, $X = \{x_1\}\mathrm{U}\{x_2\}\mathrm{U}\ldots\mathrm{U}\{x_N\}$; therefore, $X = \{x_1, x_2, \ldots, x_N\}$

$T_{\text{class1}} = \min(\{X_{\text{class1}}\}) = \min\,\{x_1, x_2, \ldots, x_k\}$.

For each class, the same procedure should be completed. Having P classes leads us to

$T_{\text{class2}} = \min(\{X_{\text{class2}}\}) = \min\{x_1, x_2, \ldots, x_k\}, \ldots$ and $T_{\text{class}P} = \min(\{X_{\text{class}P}\}) = \min\{x_1, x_2, \ldots, x_k\}$.

The final T is defined as follows:

$T = \min(\{T_{\text{class1}}, T_{\text{class2}}, \ldots, T_{\text{class}P}\})$.

2.4. *Classification*

Support vector machine (SVM) and supervised fuzzy C-means (SFCM) are two different classifiers with different principles and approaches. To have a just and fair evaluation, both FCM and SVM are applied for evaluating the effectiveness of the mentioned time–domain features in this work. The comparison between their results provides a clearer picture of the capability of these features and classifiers for distinguishing those four mental tasks.

2.4.1. Support vector machine. The SVM is one of the best state-of-the-art classifiers with lower complexity compared to other classifiers such as neural network and fuzzy. The main idea behind the SVM is to find discriminant hyperplanes that separate the data which belong to different classes with the maximum possible margin. Maximizing the margins increases the generalization capabilities of the classifier. The SVM uses a regularization parameter that enables accommodation to outliers and allows error on the training set. With small increase of the classifier's complexity, a linear SVM can make nonlinear decision boundaries by using the 'kernel trick'. Generally, it is done by mapping the data to another space, mostly of much higher dimensionality, with the help of the kernel function.

The SVM has several advantages due to its margin maximization and regularization term. Insensitivity to overtraining and dimensionality, a few parameters that need to be tuned manually and good generalization properties are the other advantages of the SVM (Burges 1998). The applied SVM in this work is a multiclass SVM with a one-versus-one strategy. Polynomial kernel is used as the kernel function and the penalty factor, C, is set as 1000 through trial and error.

2.4.2. Supervised fuzzy C-means. FCM is a method of clustering that has recently been applied for the classification of various kinds of biomedical signals. Fuzzy clustering methods like FCM are more precise than crisp ones. Since a total commitment of a vector to a given class is not essential for each iteration, and it allows data to belong to two or more clusters at the same time, therefore, FCM is also known as an overlapping data clustering method. The advantage of a fuzzy model is that all the variables are continuous and differentiable. Therefore, the formulated problem is easier to solve. In the current work, the supervised version of FCM is applied to create a suitable direction during the training procedure. FCM is an unsupervised learning algorithm. A common problem with FCM is that the cluster structure does not necessarily correspond to the classes in the dataset. Consequently, the classification accuracy and thereafter, the efficiency is less. Class labels provide a convenient direction throughout the training procedure, as is being done in supervised learning methods. This method is called SFCM. Data grouping by categorized preliminary labeled samples or by modifying an available set of categories that reveal irregularities in the dataset is the major capability of supervised clustering. The principle of SFCM is to apply the labeled data samples to monitor the repetitive optimization procedure. Therefore, the obligation of association of a vector to a known class for each iteration is not essential (Krishnapuram and Keller 1996). The characteristics of the employed dataset such as the similarity or distance between respective data points are fundamental for clustering techniques. Distance, connectivity and intensity are the most common parameters where their measurement depends on the data and also the application. These parameters control the way of cluster formation. In this work, the Euclidean distance-based similarity measure is applied for class identification.

The number of classes is given as 4, and in the training, phase and labels of data are provided for the classifier. By means of the given information, the cluster center of each cluster and the initial distance-based similarity were computed.

3. Results and discussion

3.1. General assessment

Each feature as a three-dimensional single feature vector, with respect to three chosen electrodes (C3, Cz, C4), has been evaluated with the SVM. Random selection of samples, using shuffle function, for training sets is performed each time. After the selection, data are segmented. To assess the classification performance, the generalization error was estimated by fivefold cross-validation. The reported standard deviation is calculated from the accuracy of the fivefold cross-validations. The ratio of the training data to the test data is 70:30. Each time the classifier is tested with the data of the same subject that it was trained with. Thus, the following is an assessment of a subject-specific BCI. The average accuracy and standard deviation of each single feature are represented in table 1 for the SVM.

Regarding the standard deviation in the tables, small standard deviations indicate that, regardless of the feature ability in distinction, the feature performs similarly. On the contrary, a large standard deviation indicates that the feature is not robust and had difficulty dealing with the chaotic behavior of the EEG signal. Naturally, the best feature is the one with the highest mean value and the lowest standard deviation.

The best average result is achieved by WAMP and SVM with an average of 87.88 $\pm$ 6.53% for all subjects. The worst result is obtained by skewness and SVM. The obtained result is an average of 26.6% and a standard deviation of 2.9.

The overall result is acceptable as every subject has at least one of their performances greater than 80%, except subject 8 with a maximum of 79.82% and subject 13 with 78.83%

Table 1. Average classification accuracy standard deviation of features evaluated with the SVM.

Feature	Average accuracy (%)	Standard deviation	Minimum accuracy (%) (by subject)	Maximum accuracy (%) (by subject)
MAV	72.98	10.83	58.15 (S2)	92.64 (S9)
MAX	50.27	12.44	34.83 (S1)	75.16 (S4)
SSI	72.15	11.17	56.67 (S12)	90.67 (S9)
WAMP	87.88	6.53	77.70 (S14)	98.16 (S5)
WL	51.13	9.45	31.67 (S1)	68.52 (S6)
MMAV1	71.85	10.64	56.80 (S2)	90.84 (S9)
MMAV2	51.71	12.23	33.49 (S2)	70.14 (S3)
IEEG	73.68	11.02	58.63 (S2)	93.65 (S9)
VAR	46.54	6.16	36.65 (S13)	58.82 (S14)
SSC	82.03	11.27	65.15 (S13)	99.33 (S2)
STD	46.10	6.15	37.81 (S13)	58.17 (S14)
Skewness	26.60	2.90	20.83 (S14)	31.84 (S4)
Kurtosis	45.19	9.55	35.62 (S2)	69.49 (S9)
MV	31.22	4.46	24.00 (S12)	46.70 (S11)
RMS	74.47	10.05	59.98 (S2)	93.97 (S9)

Table 2. Average classification accuracy and standard deviation of features evaluated with FCM.

Feature	Average accuracy (%)	Standard deviation	Minimum accuracy (%) (by subject)	Maximum accuracy (%) (by subject)
MAV	62.86	16.66	35.17 (S13)	91.00 (S9)
MAX	54.99	12.23	34.00 (S15)	80.41 (S3)
SSI	59.58	13.91	39.50 (S2)	92.93 (S9)
WAMP	79.77	10.19	54.00 (S12)	96.00 (S5)
WL	42.44	9.10	31.63 (S8)	64.60 (S6)
MMAV1	59.47	15.39	38.29 (S2)	38.29 (S2)
MMAV2	51.81	16.89	16.67 (S13)	83.87 (S11)
IEEG	65.03	15.91	37.89 (S1)	88.89 (S3)
VAR	34.69	10.72	19.44 (S7)	57.14 (S11)
SSC	74.45	17.59	46.67 (S9)	99.00 (S6)
STD	39.84	10.95	22.39 (S13)	61.67 (S11)
Skewness	38.84	11.84	21.00 (S11)	60.00 (S3)
Kurtosis	39.74	10.95	22.13 (S1)	59.60 (S3)
MV	38.58	7.98	26.32 (S13)	55.50 (S11)
RMS	65.64	19.22	30.00 (S1)	98.00 (S9)

accuracy. Additionally, 9 out of 15 subjects had at least one of their performances with more than 90%. These results show that for BMI applications, which can tolerate a few errors, these feature–classifiers can be considered as potential options.

To be sure of the result obtained by the SVM, each feature is reassessed with FCM in exactly the same circumstances. The average accuracy and standard deviation of each feature are also calculated, and are represented in table 2 for FCM.

For 11 out of 15 features, the SVM performed better than FCM in terms of average accuracy. Only for four features, namely, MV, MAX, MMAV2 and skewness, FCM had higher average classification accuracy. It could be observed from a comparison between tables 1 and 2 that variation of the obtained results with FCM is substantial compared to SVM. This fact is quantified by more variation in the average value as well as higher standard deviation of each feature assessed by this classifier, which is represented in figure 1, for ease of comparison.

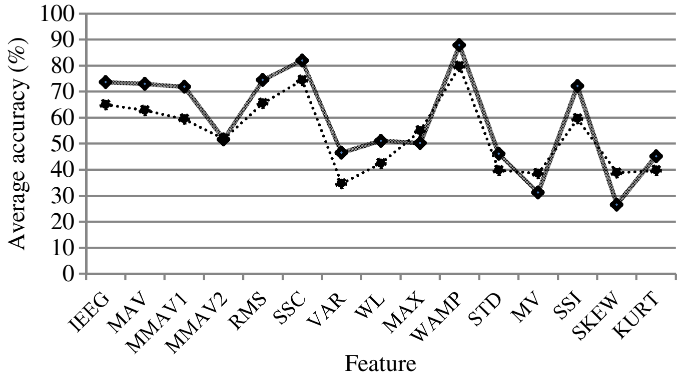

Figure 1. Average classification accuracy of features evaluated with the SVM (solid line) and FCM (dotted line).

In terms of complexity, FCM is more complicated as it consumed considerably more time to deliver the final results compared to the SVM. The minimum and maximum results are also presented in both tables along with the subject who could obtain the result. Tracking the subjects' performances would be beneficial in the case of designing subject-specific systems.

3.2. Benchmarking

Although many parameters differ between the current experiment and the experiments in the BCI competitions, either BCI competition III or BCI competition IV, to evaluate the applied approaches, the obtained results in this work are compared with the obtained results of these two online databases. In the BCI competition III, dataset IIIa is cued motor imagery with four classes including left-hand, right-hand, foot and tongue movements, from three subjects. Regarding dataset IIIa, the approach of the third group could obtain classification accuracy as high as 94.81% for the first subject and as low as 41.11% for the second subject (Blankertz *et al* 2005). Variations of the results and the obtained accuracies are similar with the results obtained in this work. The distribution of a feature vector for each feature of the current work for one of the selected subject can be reviewed in Khorshidtalab *et al* (2012a), (2012b). In terms of the BCI competition IV, dataset IIa is cued motor imagery with four classes including left-hand, right-hand, feet and tongue movements from nine subjects. Compared to the obtained results for the BCI competition, dataset IIIa and our obtained results, the results are relatively low, which could be due to the distribution of data and the associated artifact. In dataset IIIa, the EEG signal of subject 5 and subject 6 were the problematic ones. Considering all five proposed approaches for classification, these two subjects could gain the two lowest accuracies among all. The three wining algorithms used the SVM as either the main classifier like the current work, or a part of classification algorithm, such as group 2 that used the ensemble method for classification which is bagging of SVM, *K*-nearest neighbors and the *linear discriminant analysis* (Tangermann *et al* 2012).

3.3. Assessment of the proposed features

After finding the right *T* through the proposed procedure for each subject, modified WAMP and modified SSC are evaluated with the SVM. The obtained accuracies for these two methods are shown in table 3.

Table 3. Classification accuracy for modified WAMP and modified SSC evaluated with the SVM.

Subject	Found T for WAMP (proposed method)	Accuracy (%)	Found T for SSC (proposed method)	Accuracy (%)
1	5.1	88.33	1.961	81.67
2	8.17	96.67	0.958	100
3	2.77	89.17	0.460	94.17
4	1.09	86.67	0.141	87.50
5	0.5	98.33	0.053	92.50
6	6.78	83.33	1.080	90.00
7	14.0	92.50	3.987	99.17
8	0.55	85.83	0.052	74.17
9	0.92	95.83	0.119	79.17
10	1.95	85.83	0.119	82.50
11	0.56	80.00	0.254	76.67
12	2.28	86.17	0.257	78.33
13	1.13	75.00	0.075	66.33
14	0.43	84.17	0.032	91.67
15	0.56	78.33	0.041	71.67

Table 4. Comparison of the classification accuracy for WAMP evaluated with the SVM.

Subject	Proposed T	$T = 10$	$T = 5$	$T = 1$	$T = 0.1$	$T = 0.01$
	Accuracy					
1	88.33	NA	87.5	55.83*	33.33*	25.83*
2	96.67	NA	95	90.83	50.83*	25*
3	89.17	NA	NA	82.50	73.33	35*
4	86.67	NA	NA	84.17	83.33	55*
5	98.33	NA	NA	NA	95	71.67
6	83.33	NA	80	42.59*	41.67*	25*
7	92.50	91.31	84.17	75	32.50*	25*
8	85.83	NA	NA	NA	84.17	54.17
9	95.83	NA	NA	NA	77.50	50.83*
10	85.83	NA	NA	85	59.17	39.17*
11	80.00	NA	NA	NA	67.50*	31.67*
12	86.17	NA	NA	84.17	65	31.67*
13	75.00	NA	NA	73.8	60.83	39.17*
14	84.17	NA	NA	NA	77.50	62.50*
15	78.33	NA	NA	NA	75.83	51.67*

Additionally, the related accuracy for each found value, through the proposed method, is compared with some other possible values for T. The experimental results are shown in tables 4 and 5, respectively. 'ACC' indicates accuracy and 'NA' expresses not available, which is zero accuracy in mathematical form. Values marked with star (*) are determined when misclassification occurs. The importance and sensitivity of finding the precise threshold value is clear from the EEG behavior of most of the subjects while facing different values for T. For instance, the proposed method for the WAMP found $T = 8.17$ for subject 2, and the SVM could classify it with 96.67%, while if $T = 10$, the classification is totally lost and misclassification occurs and for $T = 5$, the SVM could obtain 95%. The same trend can be observed from the EEG behavior of subject 6 while facing different threshold values for T.

Table 5. Comparison of the classification accuracy for SSC evaluated with the SVM.

Subject	Proposed T	$T = 1$	$T = 0.1$	$T = 0.01$
	Accuracy			
1	81.67	69.17	62.50	60.00
2	100	NA	100	100
3	94.17	NA	91.67	88.33
4	87.50	NA	85.00	75.00
5	92.50	NA	NA	85
6	90.00	88.33	79.17	86.67
7	99.17	92.50	91.67	90.83
8	74.17	NA	NA	65.00
9	79.17	NA	74.17	52.50*
10	82.50	NA	76.66	75.00
11	76.67	NA	74.17	68.33
12	78.33	NA	71.67	65.83
13	66.33	NA	NA	58.33
14	91.67	NA	NA	89.17
15	71.67	NA	NA	69.33

Table 6. Confusion matrix.

		Predicted class			
		Class 1	Class 2	Class 3	Class 4
Known class	Class 1	tp_1	e_{12}	e_{13}	e_{14}
	Class 2	e_{21}	tp_2	e_{23}	e_{24}
	Class 3	e_{31}	e_{32}	tp_3	e_{34}
	Class 4	e_{41}	e_{42}	e_{43}	tp_4

3.4. Evaluation metrics

Along with accuracy and standard deviation, three more metrics, namely sensitivity (*Se*), specificity (*Sp*) and normalized mutual information (*MI*) are applied to assess the performance of features capable of delivering average of more than 50% while their minimum performance is also more than 50%. These features are MAV, SSI, WAMP, MMAV1, IEEG, SSC and RMS.

According to the obtained results shown in figure 1, the SVM could perform better compared to FCM in the current study. Therefore, only the SVM is considered for further evaluation.

Se, *Sp* and *MI*, all three of them, demand for the confusion matrix which is shown in table 6 for four classes of data. In many research works, a binary confusion matrix for binary classification is reported while a multiclass confusion matrix is not commonly applied.

The confusion matrix shows how the predictions are made by the classifier. In the confusion matrix, the rows correspond to the known class of the data and the columns correspond to the predictions made by the model. The diagonal elements show the number of correct classifications made for each class, and the off-diagonal elements show the errors.

For multiclass classifiers, *Se* corresponds to the true positive rate and is defined by equation (3), and *Sp* corresponds to the true negative rate for each class (as in not being a member of a certain class) and is given by equation (4):

$$Se_1 = tp_1/\,(tp_1 + e_{21} + e_{31} + e_{41}) \tag{3}$$

Table 7. Average sensitivity (*Se*) and average specificity (*Sp*) for the selected features.

Feature	Average *Se*	Average *Sp*
WAMP	0.8583	0.9512
SSC	0.8389	0.9463
IEEG	0.7450	0.9150
RMS	0.7372	0.9124
SSI	0.7244	0.9081
MAV	0.7139	0.9046
MMAV1	0.7117	0.9039

$$Sp_1 = tn_1/(tn_1 + e_{21} + e_{31} + e_{41}), \tag{4}$$

where tn_i is $tn_1 = tp_2 + tp_3 + tp_4 + e_{23} + e_{32} + e_{24} + e_{42} + e_{34} + e_{43}$.

Table 7 shows the average *Se* and average *Sp* of the selected features vector for all subjects. The highest *Se* and *Sp* belongs to WAMP followed by SSC, which is the same as what has been obtained from the accuracy point of view. Interestingly, the rank of average *Se* and *Sp* for each of the selected features is the same as their rank of accuracy.

Different from the conventional evaluation criteria using performance measures, information-theory-based criteria present a unique beneficial aspect. Mutual information, which compares the amount of information carried by the classifier output and by the true class labels, is one of them (Schlögl *et al* 2002, 2003). Mutual information is an entropy type quantity, which provides a measure of the statistical relationship between variables and contains all the statistics of the related distributions. Thus, *MI* is a more general measurement than a simple cross-correlation:

$$MI_e\,(T, Y) = \sum_T \sum_Y P_e\,(T, Y) \quad \log_2 \frac{P_e\,(T, Y)}{P_e\,(T)\,P_e\,(Y)} \tag{5}$$

$$MI_e\,(T, Y) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{m} C_{ij} \log_2 \left(\frac{C_{ij}}{C_i \sum_{i=1}^{m} \left(\frac{c_{ij}}{n} \right)} \right)}{\sum_{i=1}^{m} C_i \log_2 \left(\frac{C_i}{n} \right)}. \tag{6}$$

In equation (5), *Y* is the output data set and *T* is the target data set, and P_e (*T; Y*) is the empirical probability density function of the joint distribution. When $MI = 1$, it indicates a full correlation between *Y* and *T*, while when $MI = 0$, complete independence between *Y* and *T* is expected.

In equation (6), C_{ij} represents the sample number of the *i*th class, which is classified as the *j*th class, and C_i is the total number for the *i*th class (Hu and Wang 2008).

MI for four classes of data is presented in table 8. For each feature–subject, the average mutual information of the four aforementioned classes is calculated and presented. The last row of the table represents the average mutual information per feature and the last column represents the average mutual information per subject. Values in light gray point out the result with maximum value and those in dark gray indicate the obtained results with minimum value per feature.

Interestingly enough, subjects who could perform better than the others and some features that could differentiate different classes' data points better than the other features have higher mutual information.

Although the information presented in table 8 is the analysis of the acquired EEG signal after processing and classification from the entropy point of view, it is also another proof of

Table 8. Average normalized mutual information for selected features.

Subject	MAV	SSI	WAMP	MMAV1	IEEG	SSC	RMS
1	0.0994	0.0520	0.6711	0.0368	0.0752	0.5024	0.0916
2	0.1337	0.1358	0.8880	0.1500	0.1536	0.9482	0.1625
3	0.6540	0.6549	0.7689	0.6110	0.5947	0.8215	0.7225
4	0.2762	0.3045	0.7312	0.4163	0.4149	0.6889	0.4711
5	0.4324	0.5234	0.9482	0.3731	0.6321	0.7962	0.5281
6	0.6278	0.6588	0.6544	0.7151	0.7057	0.7252	0.7040
7	0.3808	0.3037	0.7957	0.3876	0.3934	0.9694	0.3308
8	0.2592	0.3698	0.5582	0.2885	0.3929	0.4579	0.2978
9	0.8415	0.7319	0.8018	0.8364	0.8270	0.5147	0.8782
10	0.4359	0.4281	0.5150	0.5462	0.4979	0.6246	0.5577
11	0.5285	0.4904	0.6197	0.4506	0.4775	0.5454	0.5140
12	0.1853	0.1962	0.5953	0.1494	0.2794	0.4873	0.1701
13	0.2707	0.2733	0.5048	0.2954	0.3942	0.2689	0.3120
14	0.5114	0.5216	0.6181	0.4619	0.5132	0.7278	0.4420
15	0.2812	0.3218	0.5257	0.2431	0.2378	0.4465	0.2767
Average	**0.3945**	**0.3977**	**0.6797**	**0.3974**	**0.4393**	**0.6350**	**0.4306**
STD	**0.2097**	**0.1991**	**0.1378**	**0.2185**	**0.2036**	**0.1988**	**0.2259**

Table 9. Accuracy and mutual information for feature combination.

Subject	Accuracy (%)	Mutual information
1	90	0.6899
2	100	1
3	99.1667	0.9694
4	96.6667	0.9012
5	100	1
6	100	1
7	100	1
8	100	1
9	100	1
10	99.1667	0.9694
11	98.3333	0.9393
12	96.6667	0.8880
13	90	0.7466
14	93.3333	0.8147
15	88.3333	0.6862

the accuracy obtained by these feature–subjects as these analyses delivered closely relevant results.

3.5. Assessment of feature combination

Classification is applied to the concatenation of the selected single features. These features are MAV, SSI, WAMP, MMAV1, IEEG, SSC and RMS, which together make a 21-dimensional feature vector. These features were capable of delivering an average of more than 50%, while their minimum performance is also more than 50%. Regarding table 9, the performance is significantly enhanced. In addition to accuracy, mutual information reached its perfect state for several subjects. It should be noted that despite the considerable improvement obtained by applying combinations of features, single features are still preferred in some conditions due to some practical usage or certain constrains such as speed of processing.

4. Conclusion and future work

Two modified time–domain features along with a comprehensive assessment of time–domain features for a robust motor imagery EEG signal classification are proposed. The obtained results show that in average, the SVM and FCM have more or less similar capability in recognizing the selected mental tasks. The best collection of subsets is the one that minimizes the probability of misclassification, which is investigated and discussed in this paper from the entropy point of view by applying mutual information. Considering some of the characteristics of this investigation would help understand the ability of combination of these feature–classifiers better. No use of any filters for noise reduction, no elimination or rejection of signals partially, no use of artifact removal and last but not least, no use of any biofeedback show how robust some of these features could perform individually.

Feature selection in this work was based on their accuracy, and feature combination applied the simplest method of concatenating the selected features. For future work, the other methods of feature selection and more complicated techniques of feature combinations for a better result are suggested. Another possibility is to examine fusion classifiers with these time–domain features for the best possible result. Applying ensemble classifiers is another potential future work for improving the general results in terms of applying a single feature.

Appendix

(A) *Mean absolute value (MAV)*. Estimates the mean absolute value of each segment by adding the absolute value of all the values x_i –ith point, the current point of signal x, and dividing it by the length of the segment (Englehart and Hudgins 2003):

$$\mathrm{MAV}_K = \frac{1}{N}\sum_{i=1}^{N} |x_i| \,. \tag{A.1}$$

(B) *Maximum value (MV)*. Maximum peak value (Bronzino 1995) refers to the maximum absolute value of each considered segment, that is

$$x_k = \max |x_i| \,. \tag{A.2}$$

(C) *Simple square integral (SSI)*. SSI (Phinyomark *et al* 2009) calculates the energy of the EEG signal according to

$$\mathrm{SSI}_K = \sum_{i=0}^{N} (|x_i^2|). \tag{A.3}$$

(D) *Waveform length (WL)*. WL is the cumulative length of the waveform over the segment. It indicates a measure of waveform amplitude, frequency and duration, all within a single parameter (Englehart and Hudgins 2003):

$$\mathrm{WL}_k = \sum_{i=1}^{N-1} |x_{i+1} - x_i| \,. \tag{A.4}$$

(E) *Modified mean absolute value1 (MMAV1)*. MMAV1 is an extension MAV using the defined weighting function w_i (Phinyomark *et al* 2009):

$$\mathrm{MMAV1}_k = \frac{1}{N}\sum_{i=1}^{N} w_i\,|x_i|$$
$$w\,(i) = \begin{cases} 1, & 0.25N \leqslant i \leqslant 075N \\ 0.5, & \text{otherwise} \end{cases}. \tag{A.5}$$

(F) *Modified mean absolute value2 (MMAV2)*. MMAV2 has a continuous weighting function of w_i. This function is known as an improvement to the modified version of MAV (Phinyomark *et al* 2009):

$$\mathrm{MMAV2}_K = \frac{1}{N}\sum_{i=1}^{N} w_i\,|x_i|$$
$$w(i) = \begin{cases} 1 & 0.25N \leqslant i \leqslant 0.75N \\ \dfrac{4i}{N} & 0.25N > i \\ \dfrac{4(i-N)}{N} & 0.75N < i. \end{cases} \tag{A.6}$$

(G) *Integrated EEG (IEEG)*. IEEG computes the summation of the absolute values of EEG signals (Huang and Chen 1999):

$$\mathrm{IEEG}_k = \sum_{i=1}^{N} |x_i|\,. \tag{A.7}$$

(H) *Variance (VAR)*. It depicts the variation of each segment. Figure 3.23 is the distribution of VAR, in the log-transformed form, in the feature space (Oskoei and Hu 2006):

$$\mathrm{VAR}_k = \frac{1}{N}\sum_{i=1}^{N} (x_i - \bar{x})^2\,. \tag{A.8}$$

(I) *Standard deviation (STD)*. It is defined as below and represents the deviation of the mean for each segment (Rechy-Ramirez and Hu 2011):

$$\mathrm{STD}_k = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N-1}}. \tag{A.9}$$

(J) *Skewness*. This parameter measures the degree of deviation from the symmetry of a normal distribution. This measure has the value of zero when the distribution is completely symmetrical, or some nonzero values when the EEG distribution is asymmetrical, with respect to the baseline (Bronzino 1995):

$$\mathrm{Skewness}_k = \frac{\sum_{i=1}^{N} \frac{(x_i-\bar{x})^3}{N}}{\left[\sum_{i=1}^{N} \frac{(x_i-\bar{x})^2}{N}\right]^{3/2}}. \tag{A.10}$$

(K) *Kurtosis*. This parameter reveals the peakedness or the flatness of each segment's waveform distribution. Figure 3.27 depicts the distribution of kurtosis, in the log-transformed form, in the feature space (Bronzino 1995):

$$\mathrm{Kurtosis}_k = \frac{\sum_{i=1}^{N} \frac{(x_i-\bar{x})^4}{N}}{\sum_{i=1}^{N} \left[\frac{(x_i-\bar{x})^2}{N}\right]^2} - 3. \tag{A.11}$$

(L) *Mean value (MV)*. This parameter calculates the MV of each segment (Bronzino 1995):

$$\mathrm{MAV}_K = \frac{1}{N}\sum_{i=1}^{N} x_i. \tag{A.12}$$

(M) *Root mean square (RMS)*. RMS is a common, widely used feature for variety of biosignals (Bronzino 1995):

$$\mathrm{RMS}_k = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}. \tag{A.13}$$

In all the abovementioned equations, N is the length of the segment, k is the current segment, x_i is the current point of signal and i is the index of the current point.

References

Bashashati A, Fatourechi M, Ward R K and Birch G E 2007 A survey of signal processing algorithms in brain–computer interfaces based on electrical brain signals *J. Neural Eng.* **4** R35–57

Beim Graben P and Kurths J 2003 Detecting subthreshold events in noisy data by symbolic dynamics *Phys. Rev. Lett.* **90** 100602

Blankertz B 2005 BCI Competition III www.bbci.de/competition/iii

Blankertz B *et al* 2005 Results of the BCI Competition III BCI Meeting (www.bbci.de/competition/iii/results/bci_competition_iii_results_list.pdf)

Blankertz B *et al* 2004 The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials *IEEE Trans. Biomed. Eng.* **51** 1044–51

Bronzino J D (ed) 1995 *The Biomedical Engineering Handbook* vol 1 (Boca Raton, FL: CRC Press)

Burges C J C 1998 A tutorial on support vector machines for pattern recognition *Data Min. Knowl. Discovery* **2** 121–67

Dornhege G, Blankertz B, Curio G and Müller K R 2002 Combining features for BCI *Advances in Neural Information Processing Systems (NIPS 02)* vol 15 ed S Becker, S Thrun and K Obermayer (Cambridge, MA: MIT Press) pp 1115–2

Englehart K and Hudgins B 2003 A robust, real-time control scheme for multifunction myoelectric control *IEEE Trans. Biomed. Eng.* **50** 848–54

Geethanjali P, Mohan Y K and Sen J 2012 Time domain feature extraction and classification of EEG data for brain computer interface *FSKD'12: 9th IEEE Int. Conf. on Fuzzy Systems and Knowledge Discovery* pp 1136–9

Gilja V, Nuyujukian P, Chestek C A, Cunningham J P, Byron M Y, Ryu S I and Shenoy K V 2011 Brain–machine interface *US Patent* application 12/932070

Herman P, Prasad G, McGinnity T M and Coyle D 2008 Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification *IEEE Trans. Neural Syst. Rehabil. Eng.* **16** 317–26

Homan R W, Herman J and Purdy P 1987 Cerebral location of international 10–20 system electrode placement *Electroencephalogr. Clin. Neurophysiol.* **66** 376–82

Hu B G and Wang Y 2008 Evaluation criteria based on mutual information for classifications including rejected class *Acta Autom. Sin.* **34** 1396–403

Huang H P and Chen C Y 1999 Development of a myoelectric discrimination system for a multi-degree prosthetic hand *Proc. IEEE Int. Conf. on Robotics and Automation* vol 3 pp 2392–7

Jasper H H 1958 The ten–twenty electrode system of the international federation *Electroencephalogr. Clin. Neurophysiol.* **10** 371–5

Khorshidtalab A, Salami M J E and Hamedi M 2012a Evaluating the effectiveness of time-domain features for motor imagery movements using SVM *ICCCE'12: IEEE Int. Conf. on Computer and Communication Engineering ICCCE'12* pp 909–13

Khorshidtalab A, Salami M J E and Hamedi M 2012b Evaluation of time-domain features for motor imagery movements using FCM and SVM *JCSSE'12: IEEE Int. Joint Conf. on Computer Science and Software Engineering* pp 17–22

Krishnapuram R and Keller J M 1996 The possibilistic C-means algorithm: insights and recommendations *IEEE Trans. Fuzzy Syst.* **4** 385–93

Lee F, Scherer R, Leeb R, Neuper C, Bischof H and Pfurtscheller G 2005 A comparative analysis of multi-class EEG classification for brain computer interface *Proc. 10th Computer Vision Winter Workshop* pp 195–204

Lotte F, Lecuyer A, Lamarche F and Arnaldi B 2007 Studying the use of fuzzy inference systems for motor imagery classification *IEEE Trans. Neural Syst. Rehabil. Eng.* **15** 322–4

McFarland D J, Anderson C W, Muller K R, Schlögl A and Krusienski D J 2006 BCI meeting 2005—workshop on BCI signal processing: feature extraction and translation *IEEE Trans. Neural Syst. Rehabil. Eng.* **14** 135–8

Müller-Gerking J, Pfurtscheller G and Flyvbjerg H 1999 Designing optimal spatial filters for single-trial EEG classification in a movement task *Clin. Neurophysiol.* **110** 787–98

Oskoei M A and Hu H 2006 GA-based feature subset selection for myoelectric classification *ROBIO'06: IEEE Int. Conf. on Robotics and Biomimetics* pp 1465–70

Pfurtscheller G and Neuper C 2001 Motor imagery and direct brain–computer communication *Proc. IEEE* **89** 1123–34

Pham D T and Cardoso J F 2001 Blind separation of instantaneous mixtures of nonstationary sources *IEEE Trans. Signal Process.* **49** 1837–48

Phinyomark A, Limsakul C and Phukpattaranont P 2009 A novel feature extraction for robust EMG pattern recognition arXiv:0912.3973

Ramoser H, Muller-Gerking J and Pfurtscheller G 2000 Optimal spatial filtering of single trial EEG during imagined hand movement *IEEE Trans. Rehabil. Eng.* **8** 441–6

Rechy-Ramirez E J and Hu H 2011 Stages for developing control systems using EMG and EEG signals: a survey *Technical Report* CES-513 (Essex, UK: School of Computer Science and Electronic Engineering, University of Essex)

Schlögl A, Keinrath C, Scherer R and Pfurtscheller G 2003 Information transfer of an EEG-based brain–computer interface *Proc. 1st Int. IEEE EMBS Conf. on Neural Engineering* pp 641–4

Schlögl A, Lee F, Bischof H and Pfurtscheller G 2005 Characterization of four-class motor imagery EEG data for the BCI-competition 2005 *J. Neural Eng.* **2** L14

Schlögl A, Lugger K and Pfurtscheller G 1997a Using adaptive autoregressive parameters for a brain–computer-interface experiment *EMBS'97: Proc. IEEE 19th Annu. Int. Conf. on Engineering in Medicine and Biology Society* vol 4 pp 1533–5

Schlögl A, Neuper C and Pfurtscheller G 1997b Subject specific EEG patterns during motor imaginary *EMBS'97: Proc. 19th Annu. Int. Conf. IEEE Engineering in Medicine and Biology Society* vol 4 pp 1530–2

Schlögl A, Neuper C and Pfurtscheller G 2002 Estimating the mutual information of an EEG-based brain–computer interface *Biomed. Tech.* **47** 3–8

Tangermann M *et al* 2012 Review of the BCI competition IV *Front. Neurosci.* **6** 55

Vidal J J 1973 Toward direct brain–computer communication *Annu. Rev. Biophys. Bioeng.* **2** 157–80

Vigneshwari C, Vimala V, Vignesh S V and Sumithra G 2013 Analysis of finger movements using EEG signal *Emerging Technol. Adv. Eng.* **3** 583

Wolpaw J R, Birbaumer N, Heetderks W J, McFarland D J, Peckham P H, Schalk G and Vaughan T M 2000 Brain–computer interface technology: a review of the first international meeting *IEEE Trans. Rehabil. Eng.* **8** 164–73

Ziehe A, Laskov P, Nolte G and Müller K R 2004 A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation *J. Mach. Learn. Res.* **5** 777–800