Med Biol Eng Comput (2012) 50:319–328
DOI 10.1007/s11517-012-0871-2

ORIGINAL ARTICLE

Classification of multichannel EEG patterns using parallel hidden Markov models

Dror Lederman · Joseph Tabrikian

Received: 1 February 2011 / Accepted: 8 February 2012 / Published online: 10 March 2012

Abstract In this paper, a parallel hidden-Markov-model (PHMM)-based approach is proposed for the problem of multichannel electroencephalogram (EEG) patterns classification. The approach is based on multi-channel representation of the EEG signals using a parallel combination of HMMs, where each model represents a particular channel. The performance of the proposed algorithm is studied using an artificial EEG database, and two real EEG databases: a database of two classes of EEGs elicited during a task of imagery of hand upward and downward movements of a computer screen cursor (db Ia), and a database of two classes of sensorimotor EEGs elicited during a feedback-regulated left–right motor imagery task (db III). The results show that the proposed algorithm outperforms other commonly used methods with classification rate improvement of 2 and 10% for db Ia and db III, respectively. In addition, the proposed method outperforms a support vector machine classifier with a linear kernel, when both classifiers utilize the same feature set. The results also show that a model architecture which includes a left-to-right scheme with no skips, five states and three Gaussians, outperforms the other tested architectures due to the fact that it allows a better modeling of the temporal sequencing of the EEG components.

Keywords Brain–computer interface · Classification · Event-related potentials · Parallel hidden Markov model

D. Lederman (✉) · J. Tabrikian
Department of Electrical and Computer Engineering,
Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel
e-mail: drorle@ee.bgu.ac.il

J. Tabrikian
e-mail: joseph@ee.bgu.ac.il

1 Introduction

Recently, automatic classification of single-trial electroencephalogram (EEG) patterns has received increasing attention in a wide range of biomedical engineering applications. EEG provides a potential nonmuscular communication channel for severely disabled persons, such as those suffering from amyotrophic lateral sclerosis (ALS) or locked-in syndrome, since some mental tasks yield distinguishable EEG signals that can be used to control an assistant device, i.e. brain–computer interface (BCI) [33]. Consequently, the idea of developing a BCI, i.e., to use the brain electrical signals in order to control remote devices, has emerged [8, 26, 29, 31]. In the recent years, EEG signals have become an important tool for the study of cognitive processing in human subjects [12, 18].

The feasibility and reliability of using brain signals as a nonmuscular communication channel heavily depends on robust and accurate recognition of the EEGs corresponding to respective mental processes. In the last 20 years, a considerable amount of work has been invested in developing methods for EEG and event-related potentials (ERP) analysis. Research in this field has faced various engineering and physiological problems. These include, among others, low signal-to-noise ratios (SNRs), overlapping spectra between the ERP and the spontaneous EEG, high intrasubject and intersubject variability, signals characterized by nonstationarity and nonergodicity, unknown number and location of intracranial sources, and unknown intrabrain wave propagation channels.

Various types of signals have been used in BCI research. These signals include Mu rhythm [32], steady-state visual evoked potentials [9], single-trial ERPs based on combining gamma-band power with slow cortical potentials [17],

movement-related cortical potentials and event-related (de)synchronization [12, 21], and rhythmic macroscopic EEG [14]. Generally, the types of signals used in this field can be categorized into: oscillatory EEG components, slow cortical potentials and event-related potentials.

Classification of various mental tasks has been extensively investigated based on EEG and ERP analysis. Various methods have been proposed, such as common spatial subspace decomposition [30], autoregressive with exogenous input (ARX) parametric modeling [4], support vector machine (SVM) [11, 15, 28] and local temporal common spatial patterns [29]. A thorough review on classification algorithms for EEG-based BCI can be found elsewhere [15]. In several studies [1, 5, 6, 19, 20, 25, 27] hidden Markov models (HMMs) have been used to classify EEGs during various mental tasks. For instance, Tavakolian et al. [27] utilized HMMs for classification of different EEG patterns with a feature vector constructed of autoregressive (AR) and adaptive autoregressive (AAR) coefficients. Argunsah et al. [1] also used HMMs with dimensionality reduction performed using principle-component analysis (PCA).

The fact that none of these methods is optimal in any sense, led to several works in this field. Furthermore, most of the research works in the field of EEG analysis have been aimed at analyzing the EEGs recorded in one or two channels, independently of the other channels [20]. Analysis of multichannel EEGs shows that different channels reflect different aspects of the brain activity due to the presentation of an external stimulus. Therefore, utilizing the mutual information among the different channels may add crucial interchannel information and consequently improve the classification performance.

Recently, Gupta et al. [10] introduced parametric multichannel fusion models to utilize the complementary brain activity information recorded from multiple channels.

In this paper, a model-based approach for multichannel classification of different EEG patterns is introduced based on a parallel combination of HMMs. The proposed approach is applicable to various types of EEG patterns such as ERPs as well as to any multisensor signals. The main contribution of this paper is thus the introduction of a model-based approach for classification of multi-channel EEG signals and the application of this approach for the classification of response-related EEG signals which resulted in improved performances comparing to the previous reported results.

The paper proceeds as follows. Section 2 discusses the rationale of the parallel hidden Markov model (PHMM)-based approach and the features employed by the classification system. The results of the classification experiments are summarized in Sect. 3. A discussion appears in Sect. 4.

2 Methods

The main idea of using HMMs to represent different EEG patterns relies on the fact that HMMs can model non-stationary signals and provide automatic dynamic time warping [7, 24]. The model incorporates several states, which represent different statistical distributions. The transitions between the states are represented using probabilities that are calculated based on the available data. HMMs are capable of representing signals that their statistical properties change over time, and compare signals which are similar, but are timely-scaled with respect to each other, thus overcoming the problem of time and morphology variability of response-related EEG.

The following sections present the preprocessing phase and the features employed by the classification algorithm, the classification scheme, and the training and recognition algorithms.

2.1 Preprocessing and feature extraction

In the preprocessing stage, an 8th-order Chebyshev II low-pass filter with a cutoff frequency is used in order to reduce the effect of artifacts, after which feature extraction is performed. The cutoff frequency was empirically found to be 30 Hz for the first database (db Ia) and 60 Hz for the second database (db III).

The goal of the feature extraction process is to represent the EEG signal in a reduced feature space, on which the classification is performed. The EEG signal is often characterized by its time domain components. For example, response-related EEG and ERPs are often characterized by the existence of positive and negative peaks at some latencies, such as P300 component which refers to a positive peak that appears around 300 msec. Generally, the EEG temporal dynamics are one of the most important discriminative features between different types of EEG patterns. Therefore, classification of different EEG patterns has been commonly performed based on the detection of positive and negative peaks at specific time points. This implies that features based upon signal amplitude, energy, and their derivatives may possess excellent discriminative capability. In addition, we consider the following features which have been successfully used in many other classification problems: linear predictive coding (LPC) and LPC-based features such as LPC-derived cepstrum (LPCC), partial correlation coefficients (PARCOR) and energy filter bank coefficients [7, 24].

The feature extraction process is performed as follows. First, the signal is divided into frames, each 10 msec long. Each frame is then multiplied by a Hamming window. The frame length is chosen so as to assure quasi-stationarity of the signal in each frame, on the one hand, and allow

adequate number of samples in order to reduce features-estimation errors, on the other hand. The frame length of 10 msec was empirically chosen. It should be noted that no significant differences between frames with length of 5, 10 or 20 msec were observed. However, longer frames affect the features variance and hence yield worse performance. In addition, a 25% overlapping was used.

Next, the first $p + 1$ autocorrelation coefficients are estimated from each windowed frame of the signal, and a vector of LPC coefficients is calculated for each windowed frame using the Levinson–Durbin recursive algorithm [16]. The LPC coefficients are used to calculate other static features such as LPCC, PARCOR and log area ratio (LAR). Other static features, such as mean frame energy and log-energy, are estimated directly from the time domain signal. The dynamic features of the signal are estimated using a first-order orthogonal polynomial approximation over a finite length window with the following formula [7]:

$$\Delta F(l,m) = \frac{\sum_{k=-K}^{K} kF(l-k,m)}{\sum_{k=-K}^{K} k^2}, \quad 1 \le m \le Q, \qquad (1)$$

where $F(l, m)$ is the static feature, m is the feature index, Q is the size of the feature set and l is the frame number. K was chosen in this work to be 2, which is a reasonable choice that balances between the "weight" given to the current frame, and the need to represent the temporal changes as compared with adjacent frames [7]. $\Delta F(l,m)$ is referred to as the "Del" of the relevant feature. The "Del Del" of the static feature is also computed by applying (1) on the Del feature. Figure 1 presents a block diagram of the feature extraction procedure. Two examples of the features calculated for randomly-selected signals from the two EEG classes are presented in Fig. 2. For the purpose of demonstration, only four features are presented- the first energy filter bank coefficient, log energy and the corresponding dynamic features. The temporal changes in the signals are clearly represented using these features, although one cannot visually distinguish between the two classes of EEGs based on the 4-dimensional space presented here.

2.2 Classification scheme

Figure 3 shows a block diagram of the PHMM classification system. The figure presents the training phase, during which the models are estimated and stored in memory, and the testing phase, in which the system performance is evaluated.

2.2.1 *The training algorithm*

During the training phase, the supervised classification system is fed with observations from the channels at each class, and accordingly it estimates the parameters of the HMM which characterizes the channel at the class. The HMM parameters are initially estimated using the k-means algorithm [24], and then the Baum–Welch algorithm [7, 24]. A thorough description of these algorithms appear in the referenced literature. The calculations were made iteratively 100 times, but if the log-likelihood values became stable for five successive iterations, convergence was assumed and the iterations were stopped.

2.2.2 *The recognition algorithm*

Most of the EEG and ERP classifiers proposed in previous studies, have been concerned with classification of the data

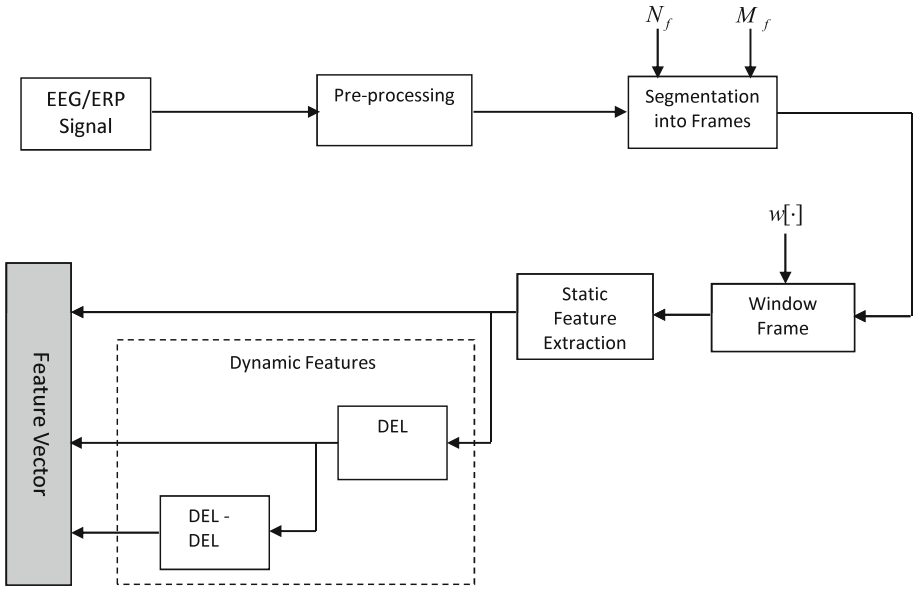

Fig. 1 Block diagram of the preprocessing and feature extraction procedure. N_f represents the number of samples in each frame, M_f represents the number of overlapped samples between consecutive frames and w[·] represents the time window used for short-time analysis

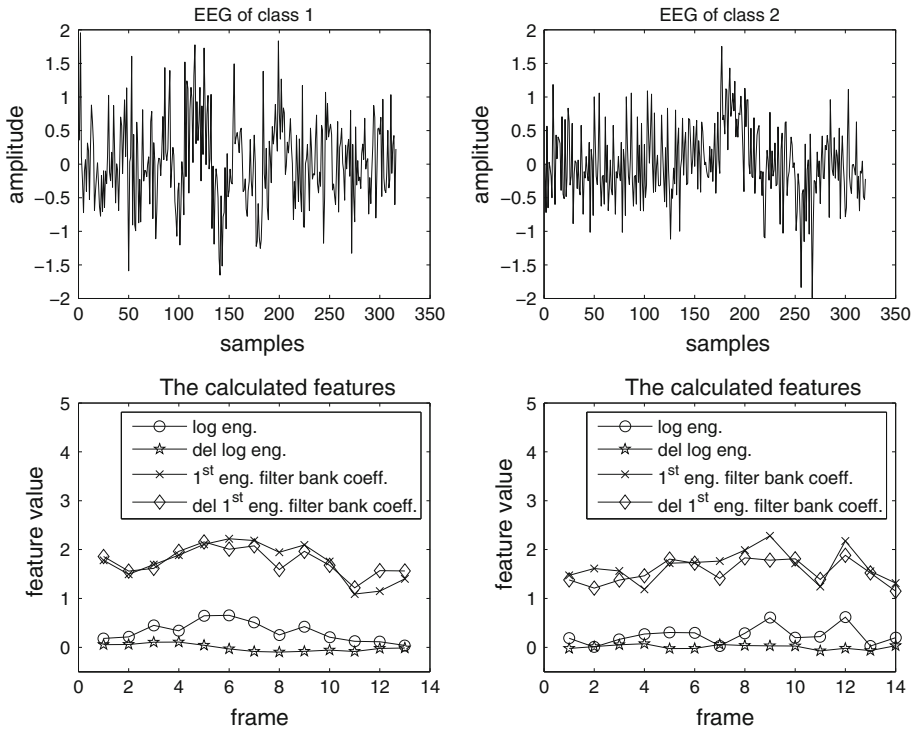

Fig. 2 Two examples of the features calculated for randomly-selected signals from the two EEG classes. In this figure, only four of the features are presented, namely the first energy filter bank coefficient, log energy and the corresponding dynamic features

extracted from one or two channels. When using data from multiple channels, recorded from different scalp electrodes, it is essential to allow a different statistical model for each channel. According to the PHMM approach, this is achieved by assigning a different HMM for each one of the class channels, forming a generalized model for each class. The PHMM concept is presented in Fig. 4, which refers to the common case of two classes of EEGs/ERPs.

During the recognition phase, the likelihood of each one of the HMMs based on the observation vector, O, is computed. The selected model is the one whose HMM provides the highest likelihood. The maximum likelihood (ML) decision rule is therefore given by:

$$m = \arg_l \left\{ \max_{l=1,\ldots,L} \left(\max_{s=1,\ldots,S} (P(O|\lambda_{ls})) \right) \right\}, \quad (2)$$

where λ_{ls} represents the sth HMM of the lth class. $P(O|\lambda_{ls})$ denotes the conditional probability density function (pdf) of the observation vector given the model λ_{ls}, and it is estimated using the Viterbi algorithm [24].

The PHMM achieves better representation for each class using several HMMs, one for each channel. The main drawback of the PHMM-based approach is that it requires a large training set in order to allow reliable parameter estimation for each channel model. This problem is solved by employing an adaptation algorithm [13]. According to this approach, instead of training each channel model using the dataset of the particular channel, which is relatively small, the whole training dataset of the class is used to initially train each model. Then, the parameters of each one of the channel models are adapted using the channel dataset. In this manner, less data are required to estimate each model, because these data are used for adaptation and not for the initial estimation of the model parameters.

3 Results

This section presents the experiments carried out in order to evaluate the performance of the PHMM approach. Section 3.1 presents the results obtained in the simulation experiments. The results of the experiments with real EEG data are reported in Sect. 3.2.

3.1 Simulation experiments

In order to evaluate the performance of the proposed algorithm under different SNRs and with different parameters, an artificial EEG database was created. While most of the studies, which are based on an artificial database for evaluation of estimation and classifications methods, use random signals [23], in this work the artificial database was generated out of real averaged EEG signals. In this way, no underlying assumption regarding the signals or their statistical properties is imposed. For this purpose, ensemble averages of the trials of db Ia (described in details in the next section) were calculated for each class. A total of 134 trials were used to calculate the averages. These averaged signals were used as basic patterns, based on which different trials were generated for each class. Each trial was

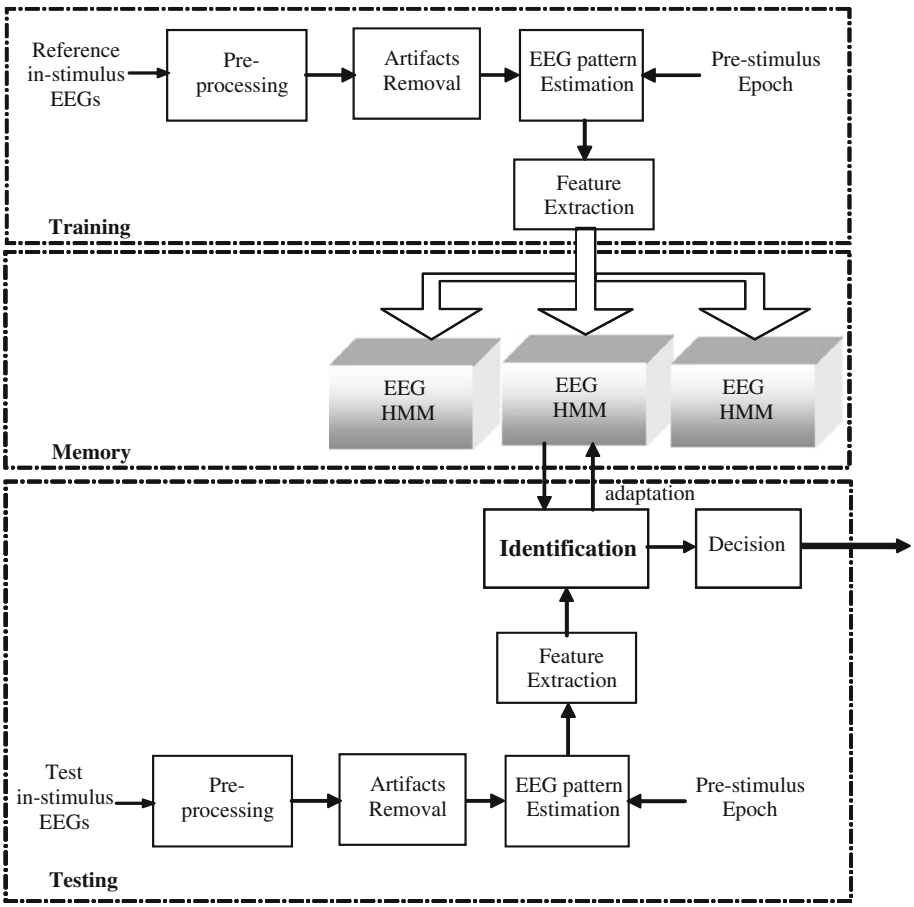

Fig. 3 A block diagram of the proposed classification system. *Upper part* training algorithm in which features are extracted from each signal to be saved in memory (*middle part*) for later processing. *Lower part* testing algorithm in which the features of the candidate signal are estimated and classified to one of the EEG/ERP HMMs based on the ML criterion

created by performing nonlinear time warping on a real averaged EEG and adding a zero-mean white Gaussian noise with variance σ^2, such that each trial is a noisy time-scaled version of an averaged EEG. This technique relies on the assumption that trials of the same class differ in time latency and time scaling. A total of 1,000 trials were generated. Figure 5 shows an example of the warping function used to create the time-scaled EEG.

A left-to-right continuous density HMM with five states and three Gaussian mixtures per state is employed (an example is presented in Fig. 7). This model architecture was initially chosen since it corresponds best to the temporal dynamics of the EEG signals. By using different states, where each state represents different signal distribution, this model is capable of representing the changes in the signal over time, hence reflecting the temporal dynamics. However, as reported in the next section, other architectures were considered as well.

The feature vector consists of ten uniformly-distributed energy filter bank coefficients, in the range of 0.5–60 Hz, and of log-energy and del log-energy. The classification performance was evaluated using ten fold cross-validation, according to which the dataset is divided into ten folds. In each iteration, nine folds are used to train the models, and the last fold is used for testing, i.e., calculating the classification error. This process is repeated 10 times and the overall classification error rate is calculated accordingly.

Figure 6 presents the equal error rate (EER) as a function of SNR. It can be seen that for SNRs smaller than −15 dB, the EER is more than 40%. For SNRs around −10 dB, the EER is approximately 25%, and for SNRs greater than 0 dB the EER decreases from about 10% for an SNR of 0 dB to about 0.7% for an SNR of 10 dB.

3.2 Real data experiments

Two real EEG databases were used in this work. These databases were published during the 2003 BCI competition [3]. In this competition, the participant research groups were given a labeled training dataset and an unlabeled testing dataset, and were asked to provide classification decisions on every signal in the testing dataset. It should be noted that at the time of performing the experiments reported here, the labels of the testing dataset were already available. However, in order to ensure that the experiments are performed in the same conditions as in the BCI competition, only the labeled dataset was used to adjust the classifier's parameters (using ten fold cross-validation),

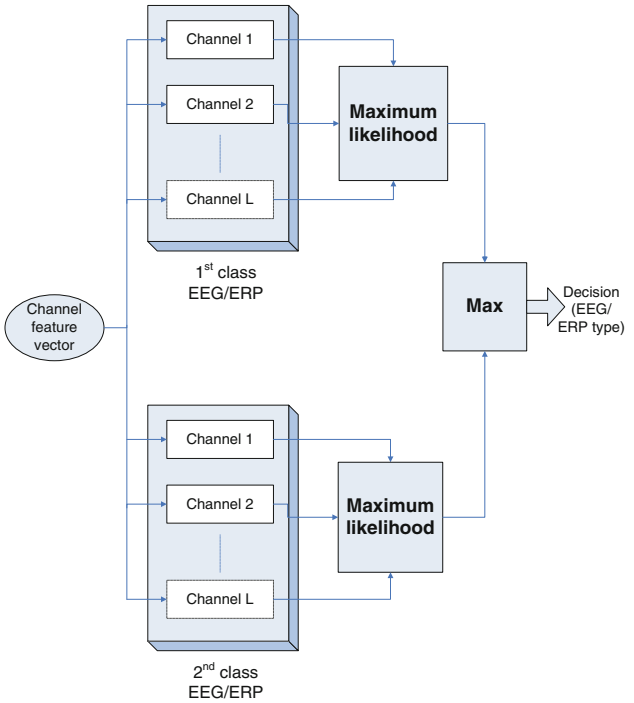

Fig. 4 The PHMM concept. Each of the two EEG/ERP HMMs is constructed of L HMMs connected in parallel. The HMM which provides the highest likelihood, is the selected model

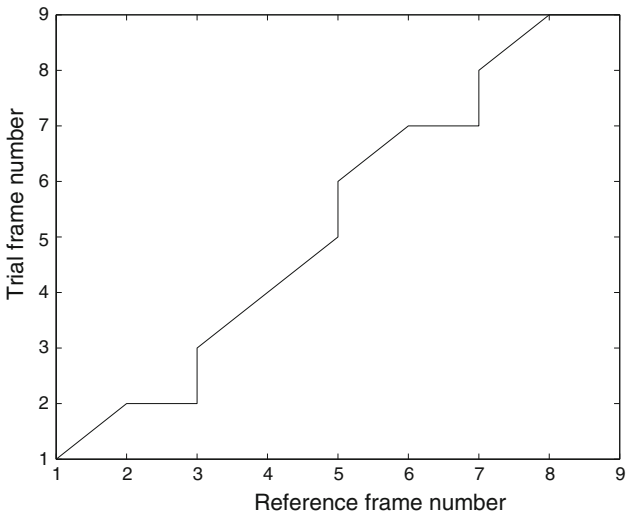

Fig. 5 An example of a random warping function. y axis- trial frame number, x axis- reference frame number. The figure presents scaling of the trial frames axis comparing to the reference frames axis

while the classifier performance was evaluated based on the 'unlabeled' dataset. It should be also noted that the classification errors were calculated just as in the competition such that comparison between the results is straightforward.

The first database (marked as db Ia) consists of two classes of EEGs elicited during a task of imagery of hand

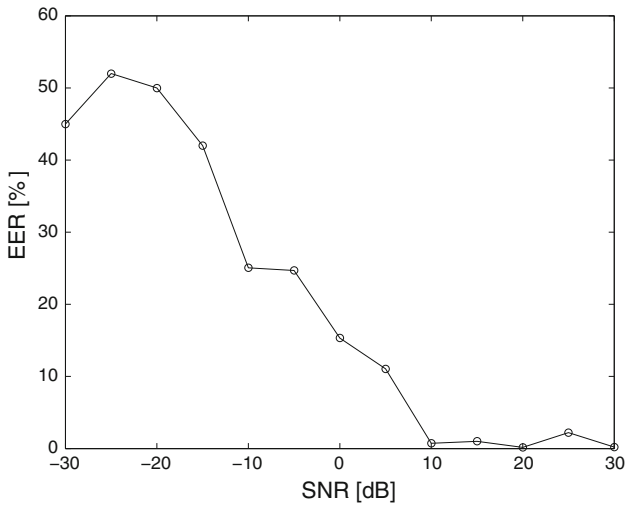

Fig. 6 EER as a function of SNR for the artificial EEG database. The feature vector consisted of 10 energy filter bank coefficients, log-energy and del log-energy

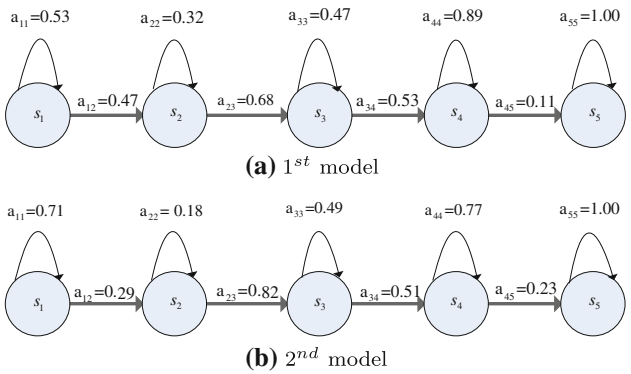

Fig. 7 Five states HMMs for the two EEG classes and their transition probabilities

upward and downward movements of a computer screen cursor [2]. The data were acquired from a single healthy subject at the University of Tuebingen, Germany, as described elsewhere [22]. The subject was asked to move a cursor up or down on a computer screen, while his/her cortical potentials were taken. During the recording, the subject received visual feedback of his slow cortical potentials (Cz-Mastoids). Cortical negativity led to an upward movement of the cursor, while cortical positivity led to a downward movement of the cursor on the screen. Six EEG electrodes were all referenced to the vertex electrode Cz as follows: Channel 1, channels A1–Cz [international 10–20 system (A1-left mastoid)]; channel 2, A2–Cz; channel 3, 2 cm frontal of C3; channel 4, 2 cm parietal of C3; channel 5, 2 cm frontal of C4; channel 6, 2 cm parietal of C4. The signals were acquired using a 16-bit A/D converter (Computer Boards PCIM-DAS1602)

at a sampling frequency of 256 Hz. The trials structure was as follows: 6 s inter-trial intervals, task presentation from 0.5 to 6.0 s and feedback period from 2.0 to 5.5 s. Only the signals from the feedback phase were provided such that each trial contained 896 samples for each channel. The training labeled data set consists of 268 trials and the validation set consists of 293 trials.

The second database (marked as db III) consists of two classes of sensorimotor EEGs elicited during a feedback-regulated left–right motor imagery task [22]. The data were provided by the Department of Medical Informatics, Institute for Biomedical Engineering, University of Technology, Graz, Austria. The EEG signals were acquired from a single healthy subject during a feedback session. Each subject sat in a relaxing chair with armrests. The task was to control a feedback bar by means of imagery left or right hand movements. The order of left and right cues was random. The experiment included 7 runs with 40 trials each. All runs were conducted on the same day with few breaks of several minutes each. Three bipolar EEG channels (anterior '+', posterior '−') were acquired using a G.tec amplifier and Ag/AgCl electrodes over C3, Cz and C4. The signals were sampled at 128 Hz and filtered between 0.5 and 30 Hz.

The data set consists of 140 labeled and 140 unlabeled trials, with an equal number of left and right hand trials. Each trial has a 9 s duration; after a 3 s presentation period a visual cue (arrow) is presented pointing either to the left or to the right. This is followed by another 6 s period for performing the imaginary task.

In the first classification experiment, two left-to-right PHMMs with five states per HMM and three Gaussians for each state, were trained. The feature vectors were constructed from ten energy filter bank coefficients.

Figure 7 shows the left-to-right HMMs and the transition probabilities, which were estimated using the Baum–Welch algorithm [7, 24]. The transition probabilities are denoted by a_{ij} for each couple of states $\{i, j\}$. It can be seen that transitions between successive states occur in non-zero probability, i.e., all states are utilized by the model and therefore there are no redundant states in this case.

Tables 1 and 2 present the classification error rates of the PHMM classifier comparing to the rates obtained using several other methods. Detailed description of these methods was provided only for the winning methods and can be found elsewhere [17, 14]. Note that only the methods with the four lowest error rates are described in these tables. The results show that the PHMM classifier provides the lowest error rates, with an improvement of 2 and 10%, for databases db Ia and db III, respectively, comparing to the best reported method.

Figure 8 presents the EER for different model parameters, for db Ia. The upper graph presents the EER as a function of the number of Gaussians, where the number of states is set to five. The lower graph presents the EER as a function of the number of states, where the number of Gaussians is set to three. The figure shows that the best performance is obtained when the number of Gaussians is either three or four and the number of states is between four and six. When the number of states and/or Gaussians is increased beyond a certain value, the performance degrade due to the "curse of dimensionality" phenomenon, since the available database is not large enough to estimate all the model parameters.

Each one of the studies reported in Tables 1 and 2 use different features, different classifiers, and different combinations of data channels. In order to understand whether the improvement achieved by the proposed method may be attributed to the PHMM-concept, the used feature set, or combination of both, we compared the performance of the PHMM classifier with a linear-kernel SVM classifier. For this purpose, we used an available C implementation,

Table 1 Comparison of the classification error rates for db Ia

Method	Features	Error rate (%)
Left-to-right PHMM with 5 states and 3 Gaussian/state (the current work)	10 energy filter bank coefficients, log-energy and Δlog-energy	9
Discriminant analysis (Mensh et al. [17])	DC-level and features from spectral analysis of high beta power band	11
Support vector machines (Ka-Min and Chung in [3])	Averaged EEGs amplitudes after downsampling to 25 Hz	12
Nonlinear support vector Machines (Tzu-Kuo and Huang in [3])	Time series features after linear support vector machines	15

Table 2 Comparison of the classification error rates for db III

Method	Features	Error rate (%)
Left-to-right PHMM with 5 states and 3 Gaussian/state (the current work)	10 energy filter bank coefficients, log-energy and Δlog-energy	1
Bayes classifier (Lemm et al. [14])	Morlet-wavelets at 10 and 22 Hz in channels C3 and C4	11
Linear discriminant analysis (Akash and Narayana in [3])	Ratio of AR spectral power in 4 frequency bands, channels C3 and C4	12
Several neural networks trained on different time regions (Saffari in [3])	Adaptive AR parameters (no details were given)	17

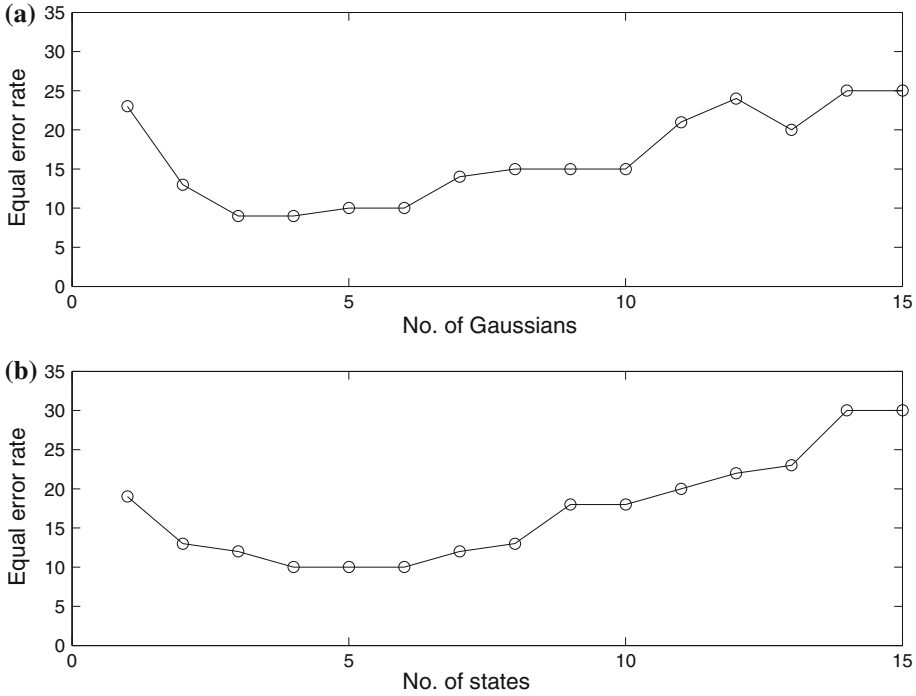

Fig. 8 EER for different HMM parameters: **a** as a function of the number of Gaussians where the number of states is set to 5; **b** as a function of the number of states where the number of Gaussians is set to 3

SVMLight (http://svmlight.joachims.org/), to classify the same feature matrices (i.e., ten energy filter bank coefficients, log-energy and Δlog-energy). The classification rates were 11 and 13% for db Ia and db III, respectively. Clearly, the PHMM classifier outperforms the linear-kernel SVM with the same feature set. It should be also noted that the SVM with energy filter bank coefficients performs slightly better (11% compared to 12%, db Ia) than an SVM with averaged EEG amplitudes (see Table 1). It can be concluded that the PHMM classifier proposed in this paper, yields the best classification rates among the above-mentioned methods.

It should be also mentioned that previous works which utilized an HMM-based classifier [25, 27], reported lower classification performance for db III, than the results obtained in this work. However, no details were given in these papers regarding the exact model architecture, signals and interchannels data processing, etc., and hence no further comparison between the two HMM-based classifiers can be performed.

We also evaluated the system performance with different model architectures. For an ergodic model, the classification errors were 12 and 8%, for db Ia and db III, respectively.

In order to evaluate the impact of different feature vectors, in the second experiment, various combinations of features were considered. Tables 3 and 4 summarize the results obtained for db Ia and db III, respectively. The results show that the best combination of features includes ten energy filter bank coefficients, log-energy and Δlog-energy. It is shown that this feature set is much better than a combination of LPC and ΔLPC, for instance, which is in common use in BCI applications.

Table 3 Classification error rates for different feature sets (db Ia)

Feature set	Error rate (%)
10 enrg. filter bank + log enrg. + Δlog enrg.	9
10 enrg. filter bank	13
10 enrg. filter bank + 10 Δenrg. filter bank	11
10 LPC	27
10 LPC + log enrg. + Δlog enrg.	22
10 LPC + 10 ΔLPC	25
10 LPCC	31
10 LPCC + 10 ΔLPCC	34
10 LAR	47
10 PARCOR	47
10 PARCOR + 10 ΔPARCOR	34
10 PARCOR + log enrg. + Δlog enrg.	28

4 Discussion

In this paper, a novel model-based approach for multichannel EEG classification is introduced. The EEG signal is first uniformly segmented into frames, out of which a feature set consisting of energy filter bank coefficients, log-energy and del log-energy, is extracted and feature matrices, which represent both temporal and spectral properties of the EEG signals, are constructed. Classification of the feature matrices is performed using an HMM-based

Table 4 Classification error rates for different feature sets (db III)

Feature set	Error rate (%)
10 enrg. filter bank + log enrg. + Δlog enrg.	1
10 enrg. filter bank	3
10 enrg. filter bank + 10 Δenrg. filter bank	14
10 LPC	49
10 LPC + log enrg. + Δlog enrg.	30
10 LPC + 10 ΔLPC	56
10 LPCC	26
10 LPCC + 10 ΔLPCC	48
10 LAR	51
10 PARCOR	53
10 PARCOR + 10 ΔPARCOR	44
10 PARCOR + log enrg. + Δlog enrg.	25

classifier. In order to model multi-channel EEGs, a parallel combination of HMMs is used, where each model represents a different channel of one class. An ML-decision rule is employed to classify the input EEG signal. The proposed algorithm is studied using an artificial EEG database and two real EEG databases. In the case of artificial EEG database, it was shown that the EER decreases rapidly from about 10% for an SNR of 0 dB to about 0.7% for an SNR of 10 dB. This result is expected, as the noise (in this case the background EEG) significantly impacts the classifier capability to discriminate between the two classes of EEG signals. This is demonstrated in Fig. 7 which shows that generally as the SNR increases, the EER decreases.

The results show that the best combination of features includes ten energy filter bank coefficients, log-energy and Δlog-energy. It is shown that this feature set yields much better performance than features such as LPC and ΔLPC, which are still in common use in BCI applications. These features efficiently model the spectral and temporal behaviour of signals in each class, allowing better discrimination between different types of EEGs.

It is shown that the PHMM classifier provides the lowest error rates, with an improvement of 2 and 10%, for databases db Ia and db III, respectively. The superiority of the PHMM classifier is due to its ability to model the nonstationarity of the signal using different states, which represent different statistical distributions. The PHMM classifier also provides automatic dynamic time warping, i.e., it is capable of comparing signals which are time-scaled with respect to each other, thus overcoming the difficult problem of time and morphology variability of response-related EEG.

The results show that the PHMM classifier with a left-to-right architecture without skips is capable of representing different types of EEGs and that it performs better than other commonly used classifiers. Moreover, a feature set which consists of energy filter bank coefficients with log-energy and Δlog-energy, is found to provide the best classification rates, outperforming other feature sets such as LPC and LPC-derived cepstrum. This is explained by the fact that the left-to-right architecture corresponds to the time-dynamics of the response-related EEG signals.

The results also show that the classification performance strongly depend on the SNR. It is therefore required to develop methods for single-trial EEG estimation in order to improve the classification performance.

It is worth mentioning that the approach proposed in this paper is applicable to many types of sequential multi-class discrimination tasks, by adapting the pre-processing and/or feature extraction.

Using a Dual Core2 2.0 GHz computer with 1 GB RAM, the PHMM training requires about 2–4 min/model to converge, depending on the specific configuration, and the validation process requires about 5 s per 1 s EEG signal. We believe that code optimization can significantly further reduce the processing time, thus allowing implementation of the system in real-time.

The processing time depends linearly on the number of channels incorporated in the PHMM architecture. Since the ML-calculation is performed using the Viterbi algorithm [24], for each frame the PHMM classifier requires $O(2L)$ operations per frame, assuming two classes of EEGs are considered, where L is the number of channels. Therefore, there is a trade-off between the required processing time, for practical BCI applications and the required accuracy.

A few concerns and improvements are left for future work. Both databases used in this work include EEG signals from a single subject. Despite the fact that there was a relatively large number of trials for each database, a much larger database which includes multiple subjects is required in order to thoroughly evaluate system performance and its usability.

In summary, a novel model-based approach for multi-channel EEG classification using energy filter bank coefficients and their derivatives was introduced and was shown to be efficient in classification of EEG recorded during two different mental tasks. The approach should be further evaluated using more diverse databases that include a large number of subjects and several mental and cognitive tasks.

Acknowledgments This work was initiated by Prof. Arnon Cohen, head of the signal processing laboratory at Ben-Gurion University of the Negev who passed away during the project. By completing this work we continue his legacy of honest and fair research.

References

1. Argunsah AO, Cetin M (2010) A brain–computer interface algorithm based on hidden markov models and dimensionality

reduction. In: IEEE 18th Signal Processing and Communications Applications Conference (SIU), pp 93–96
2. Birbaumer N, Flor H, Ghanayim N, Hinterberger TI, Iverson E, Taub B, Kotchoubey A, Kübler A, Perelmouter J (1999) Brain-controlled spelling device for the completely paralyzed. Nature 398:297–298
3. Blankertz B (2003) BCI competition final results. http://ida.first.fraunhofer.de/projects/bci/competition/
4. Burke D, Kelly S, de Chazal P, Reilly R, Finucane C (2005) A parametric feature extraction and classification strategy for brain–computer interfacing. IEEE Trans Neural Syst Rehabil Eng 13(1):12–17
5. Chiappa S, Bengio S (2004) Hmm and iohmm modeling of EEG rhythms for asynchronous bci systems. In: European Symposium on Artificial Neural Networks
6. Cincotti F, Scipione A, Tiniperi A, Mattia D, Marciani MG, del R. Millan J, Salinari S, Bianchi L, Babiloni F (2003) Comparison of different feature classifiers for brain computer interfaces. In: Proceedings of the 1st International IEEE EMBS Conference on Neural Engineering
7. Deller JR, Proakis JG, Hansen JHL (2000) Discrete-time processing of speech signals. Macmillan Publishing Company, New York
8. Donchin E, Spencer KM, Wijesinghe R (2000) The mental prosthesis: assessing the speed of a P300-based brain–computer interface. IEEE Trans Rehabil Eng 8:174–179
9. Friman O, Volosyak I, Graser A (2007) Multiple channel detection of steady-state visual evoked potentials for brain–computer interfaces. IEEE Trans Biomed Eng 54(4):742–750
10. Gupta L, Chung B, Srinath MD, Molfese DL, Kook H (2005) Multichannel fusion models for the parametric classification of differential brain activity. IEEE Trans Biomed Eng 52(11):1869–1881
11. Kaper M, Meinicke P, Grossekathoefer U, Lingner T, Ritter H (2004) BCI competition 2003—data set IIb: support vector machines for the p300 speller paradigm. IEEE Trans Biomed Eng 51:1073–1076
12. Krause CM (1999) Event-related desynchronization (ERD) and synchronization (ERS) during auditory information processing. J New Music Res 28:257–265
13. Lee CH, Lin CH, Juang BH (1991) A study on speaker adaptation of the parameters of continuous density hidden markov models. IEEE Trans Speech Signal Proc 39(4):806–814
14. Lemm S, Shafer C, Curio G (2004) BCI competition 2003—data set III: probabilistic modeling of sensorimotor μ rhythms for classification of imaginary hand movements. IEEE Trans Biomed Eng 51(6):1077–1080
15. Lotte F, Congedo M, Lécuyer A, Lamarche F, Arnaldi B (2007) A review of classification algorithms for EEG-based brain–computer interfaces. J Neural Eng 4:1–13
16. Makhoul J (1975) Linear prediction: a tutorial review. In: Proc IEEE, vol 63
17. Mensh BD, Wefel J, Seung HS (2004) BCI competition 2003—data set Ia: combining gamma-band power with slow cortical potentials to improve single-trial classification of electroencephalographic signals. IEEE Trans Biomed Eng 5(6):1052–1056
18. Molfese DL (1990) Auditory evoked responses recorded from 16-month-old human infants to words they did and did not know. Brain Lang 38:345–363
19. Obermaier B, Guger C, Neuper C, Pfurtscheller G (2001) Hidden Markov models for online classification of single trial EEG. Pattern Recognit Lett 22:1299–1309
20. Obermaier B, Neuper C, Guger C, Pfurtscheller G (2001) Information transfer rate in a five-classes brain–computer interface. IEEE Trans Neural Syst Rehabil Eng 9(3):283–288
21. Pfurtscheller G, Kalcher J, Neuper C, Flotzinger D, Pregenzer M (1996) On-line EEG classification during externally-paced hand movements using a neural network-based classifier. Elec Clin Neuro 99:416–425
22. Pfurtscheller G, Neuper C, Flotzinger D, Pregenzer M (1997) EEG -based discrimination between imagination of right and left hand movement. Electroenceph Clin Neurophysiol 103:642–651
23. Quiroga RQ, Garcia H (2003) Single-trial event-related potentials with wavelet denoising. Clin Neurophysiol 114:376–390
24. Rabiner LR (1989) A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE 77(2):257–286. doi:10.1109/5.18626
25. Rezaei S, Tavakolian K, Nasrabadi AM, Setarehdan SK (2006) Different classification techniques considering brain computer interface applications. J Neural Eng 3(2):139–144
26. Schalk G, Wolpaw JR, McFarland DJ, Pfurtscheller G (2000) EEG-based communication: presence of an error potential. Clin Neurophysiol 111:2138–2144
27. Tavakolian K, Vasefi F, Naziripour K, Rezaei S (2006) Mental task classification for brain computer interface applications. Proc First Can Conf Biomed Comput 1:56–60
28. Thulasidas M, Guan C, Wu J (2006) Robust classification of EEG signal for brain–computer interface. IEEE Trans Neural Syst Rehabil Eng 14(1):24–29
29. Wang H, Zheng W (2008) Local temporal common spatial patterns for robust single-trial EEG classification. IEEE Trans Neural Syst Rehabil Eng 16(2):131–139
30. Wang Y, Zhang Z, Li Y, Gao X, Gao S, Yang F (2004) BCI competition 2003—data set IV: an algorithm based on CSSD and FDA for classifying single-trial EEG. IEEE Trans Biomed Eng 51:1081–1086
31. Wolpaw JR, McFarland DJ (1994) Multichannel EEG-based brain–computer communication. Electroencephalogr Clin Neurophysiol 90:444–449
32. Wolpaw JR, McFarland DJ, Neat GW, Forneris CA (1991) An EEG-based brain–computer interface for cursor control. Elec Clin Neuro 78:252–258
33. Wolpow J, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan T (2002) Brain–computer interfaces for communication and control. Clin Neurophysiol 113:767–791