Proceedings of the 1st International IEEE EMBS
Conference on Neural Engineering
Capri Island, Italy • March 20-22, 2003

Comparison of Different Feature Classifiers for Brain Computer Interfaces

F. Cincotti[1], A. Scipione[4], A. Timperi[4], D. Mattia[1], M. G. Marciani[1,5], J. Millán[3], S. Salinari[4], L. Bianchi[5], and F. Babiloni[2]

[1] Fondazione Santa Lucia - IRCCS, Rome, Italy, [2] Dip. Fisiologia umana e Farmacologia, Università "La Sapienza", Rome, Italy, [3] Swiss Federal Institute of Technology, CH-1015 Lausanne, Switzerland, [4] Dipartimento Informatica e Sistemistica, Università "La Sapienza", Rome, Italy, [5] Dip. Neuroscienze, Università "Tor Vergata", Rome, Italy

Abstract - **Changes in EEG power spectra related to the imagination of movements may be used to build up a direct communication channel between brain and computer (Brain Computer Interface; BCI). However, for the practical implementation of a BCI device, the feature classifier plays a crucial role. In this paper, we compared the performance of three different feature classifiers for the detection of the imagined movements in a group of 6 normal subjects by means the EEG. The feature classifiers compared were those based on the Hidden Markov Models (HMM), the Artificial Neural Network (ANN) and on the Mahalanobis distance (MD). Results show a better performance of the MD and ANN classifiers with respect to the HMM classifier.**

Keywords - **Brain Computer Interface, Hidden Markov Model, Artificial Neural Network, Mahalanobis distance.**

I. INTRODUCTION

In the recent years it has been demonstrated that the opening of a communication channel between brain and computer (Brain Computer Interface; BCI) is possible by using EEG changes in power spectra related to the imagination of movements (reviewed in [1,2]). These EEG variations are specifically located in centro-parietal scalp areas and can be recognized by several features classifiers. Several classifiers have been proposed in the literature for the BCI devices [1]. It is worth of note that most devices operate in a context of synchronous mode, i.e by using experimental paradigms in which the user can change the mental state only in particular time instants. Ideally, a BCI device should give the opportunity to the user to change his mental states in an asynchronous way, without any restriction of time as it happens in synchronous BCI. However, it is not clear if the classifiers commonly used in a synchronous BCI could successfully operate in asynchronous devices. Hence, although the performances of several feature classifiers have been characterized previously by different Authors using synchronous BCI, there are no studies about comparisons of such classifiers for asynchronous BCI devices.

In this study we tested the performance of three feature classifiers in a context of asynchronous BCI device. The classifiers used are those based on the Mahalanobis distance [3], on Artificial Neural Networks [4] and on Hidden Markov Models [5]. Another key issue in the developing of an efficient BCI device is related to the use of a minimal number of electrodes. Hence, the comparison of of the different classifiers was performed by using EEG signals from a reduced set of electrode.

II. METHODOLOGY

A. EEG recordings and pre-processing

The EEG potentials were recorded from a group of 6 normal subjects while they are imaging right or left hand movements. Electrodes were placed in scalp centro parietal zones corresponding to the C3, P3, C4 and P4 positions of the international 10-20 system. The EEG sampling rate was 128 Hz. Depending on the particular study described below, different subsets of the scalp electrodes were used to select the features to be classified.

The EEG features employed in this study for the comparison of the classification performance between all the classifiers analyzed were the spectral parameters. More specifically, we used the Welch algorithm to estimate the power spectrum of each signal and considered values in the frequency band 8-30 Hz. Windows are 0.5 seconds long, what gives a frequency resolution of 2 Hz. Thus, an EEG sample is represented by $n \cdot 12$ features, where 12 are the spectral component for each one of the n channels used. The periodogram, and hence an EEG features vector sample, is computed every 1/2 second. In a few subjects for whom a realistic head model based on sequential MR images was available, we estimated the cortical activity during the mental imagery task using a linear estimation algorithm. Such reconstruction was performed in order to check if the electrodes employed in this study were appropriately located to gather the relevant electrical patterns elicited by the subject's mental activity from the cortical surface.

B. The Hidden Markov Classifier

We consider a HMM for which the observed sequence is produced by a set of continuous density process (CD-HMM), in particular we use a Gaussian density distribution, with covariance matrix bound to be diagonal.

For each imagined movement V , considering the observation vectors belonging to the training set, we must estimate the parameters set of the HMM λ^v that optimize the likelihood of the model. The Baum-Welch algorithm is used to estimate the parameters of the HMMs. In Estimation-Maximization algorithms, the starting point for the estimation of the matrix of observation symbols probability distribution (**B**) is extremely important, especially if the process is continuous. Thus, for the estimation of **B** we used a segmental k-mean algorithm for a better evaluation of this matrix [6]. For each unknown imagined movement to be recognized, the Viterbi Decoder is carried out. The highest likelihood HMM is selected as follows:

$$v^* = \underset{1 \le v \le V}{\operatorname{argmax}}\left(P\left(O \mid \lambda^v\right)\right) \quad (1)$$

where the probability computation step is performed using the Viterbi algorithm [7].

C. The Local Artificial Neural Network

In our local neural classifiers, every unit represents a prototype of one of the mental tasks to be recognized. During training, units are pulled towards the EEG samples of the mental task they represent and are pushed away from EEG samples of other tasks. In a statistical framework, assuming that each class-conditional density function is taken to be an independent normal distribution and dropping constant terms, the discriminant function of class C_k for sample x is:

$$y_k(x) = -\tfrac{1}{2}(x-\mu_k)^T \Sigma_k^{-1}(x-\mu_k) - \tfrac{1}{2}\ln|\Sigma_k| + \ln P(C_k), \quad (2)$$

where $P(C_k)$ denotes the prior probability of class C_k, where μ_k is the prototype (mean) of class C_k, and S_k is the covariance matrix of class C_k. In practice, this means that a sample is assigned to the class C_k with the nearest prototype based on the Mahalanobis distance.

The initialization of μ_k and S_k is given according to the maximum likelihood approach:

$$\mu_k = \frac{1}{N_k}\sum_{n=1}^{N_k} x^n, \quad \Sigma_k = \frac{1}{N_k}\sum_{n=1}^{N_k}(x^n-\mu_k)(x^n-\mu_k)^T, \quad (3)$$

where N_k denotes the number of training samples belonging to the class C_k. Only the diagonal terms of S_k are computed, in order to keep the number of free parameter (and thus the estimation accuracy) to an acceptable level.

Using the expectation-maximization (EM) framework, we optimize the position of the prototypes of the different classes to minimize mean square error function through gradient descent. Assuming equal prior class probabilities and diagonal covariance matrices, for every sample x^n in the training set, μ_k is updated by:

$$\Delta\mu_k = \alpha[t_k(x^n) - y_k(x^n)]\Sigma_k^{-1}(x^n-\mu_k)\, y_k(x^n)\, y_j(x^n), \quad (4)$$

where t_k is the k^{th} component of the target vector in the form *1-of-c*, y_k is the posterior probability of class C_k given by Eq. 2, and y_j is the probability of the remaining classes. Finally, after every iteration over the training set, we estimate again the new value of S_k using Eq. 3.

The response of the network is the class C_k associated to the nearest unit from the arriving sample provided that y_k is greater than a given probability threshold; otherwise the response is "unknown" to avoid making risky decisions for uncertain samples.

In the case that the classes have several prototypes, then only the nearest prototype of a class is used for computing the probability of that class and for learning.

D. The Mahalanobis distance-based classifier

The classifier was a implementation of the concept of the Mahalanobis distance (MD, [3]). This classifier has only one prototype per mental task, computed as the mean vector of that class estimated using the corresponding EEG samples in the training set. The response of the classifier for the current EEG sample is just the class with the nearest prototype based on the Mahalanobis distance, which can be computed using the full covariance matrix for each mental task.

E. Statistical analysis

To test the recognition capabilities of the classifiers, we employed the k-fold cross-validation procedure for each subject, with k=3. Then, we did a statistical analysis of the average recognition scores. In particular we performed an analysis of variance (ANOVA), by using as dependent variable the classification rate obtained by the cross-validation procedure, while the classifiers used as first main factor (CLASSIFIER, with three levels; MD, ANN and HMM). Post-hoc tests were performed by using Scheffe's test at $p < 0.05$ level of significance.

III. RESULTS

Figure 1 shows the decrement of spectral power in alpha band in a subject by using a realistic cortical representation based on sequential MR images. The decrement of spectral activity occurs over the centroparietal areas, mainly sampled by the C3, C4, P3 and P4 electrodes.

The averaged correct classification scores for the recognition of mental imagery, obtained by using the three different classification methods adopted (MD, ANN and HMM) are 90%, 80%, and 65% respectively.

The ANOVA performed on these data shows a significant decrease of data variance for the main factor CLASSIFIER ($p < 0.05$). Post-hoc tests confirmed the superiority of the MD and ANN classifiers with respect to

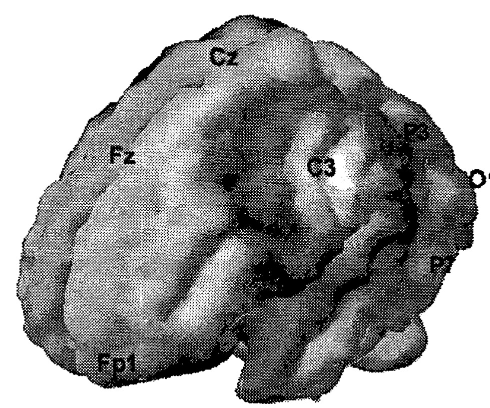

Figure 1. Estimated cortical current distribution of the power of EEG oscillations in the frequency range of 9-11 Hz for the subject FB during the mental imagery of right hand movements. The gray scale coded from white to light gray the desynchronization of the estimated current EEG rhythms with respect to the rest period, while from light gray to dark gray is coded the synchronization.

the HMM classifier ($p < 0.001$).

IV. DISCUSSION

In this paper we described some of the computational mechanisms useful to detect the imagination of motor acts in normal subjects by using EEG. Results indicate the possibility to recognize mental imagery with four electrodes by using a quadratic classifier based on the computation of Mahalanobis distance or an artificial neural network with local properties. These results are in line with recent advancements in the BCI area [1,2]. Besides, we observed that, the performance of the HMM classifiers is, on our group of subjects, below the other classifiers.

This latter observation does not appear to be attributable to an incorrect model design. In fact, the recognition scores obtained on the training set are high (about 90% for both of left and right mental imagery). Furthermore, the finding of low degree of HMM classifier performance is not in agreement with what already obtained by other Authors, who used a HMM based classifier for the classification of right and left mental imagery [5]. It should be noted, however, that some differences exist between these two implementations of HMM. For instance, in our implementation, the observed sequence is modeled as being produced by a set of continuous density process. A further remarkable difference resides in the type of experimental paradigm employed. In the aforementioned study, the experimental paradigm was of a synchronous type; namely, the subjects were required to perform the given task upon the presentation of a given stimulus. In our case, the subjects could change task at any moment of the acquisition, allowing a more "natural" utilization of the interface.

Even though the MD represents the most promising classifier in our experimental context, it should be mentioned that, under different validation conditions which are more similar to a realistic BCI utilization (i.e. testing data acquired in a session separate and subsequent to training data acquisition), the performance of the ANN classifier appears not to decline markedly.

These considerations converge in the global issue of the "benchmarking" of BCI implementation, where the choice stands between either to maximize the reproducibility of experiments, (e.g. acquisition of a common set of data once for ever) or to optimize acquisition for a given BCI, allowing the subject to take advantage of the feedback feature and thus to contribute with her/his learning process to the maximization of the brain interface performance.

REFERENCES

[1] Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtscheller G, Vaughan TM. Brain-computer interfaces for communication and control. *Clin Neurophysiol.* 2002 Jun;113(6):767-91

[2] Pfurtscheller, G. and Neuper, C. (2001). Motor imagery and direct brain-computer communication. *Proceedings of the IEEE,* **89**(7): 1123–1134

[3] Cincotti F., Mattia D., Babiloni C., Carducci F., Bianchi L., Millan J., Mourino J., Salinari S., Marciani M. and Babiloni F. Classification of EEG mental patterns by using two scalp electrodes and Mahalanobis distance-based classifiers. *Method of Inform. Med.*, 2002, 41:337-41

[4] Millán J., Mouriño J., Franzé M., Cincotti F., Varsta M., Heikkonen J., Babiloni F., A Local Neural Classifier for the Recognition of EEG Patterns Associated to Mental Tasks, *IEEE Trans Neural Network* 2002, (13):678-686

[5] Obermaier B., Guger C., Neuper C. and Pfurtscheller G. Hidden Markov models used for the offline classification of EEG data. *Biomed Tech.* 1999 Jun;44(6):158-62

[6] Lawrence R. Rabiner, Jay G. Wilpon and Biing-Hawang Juang. A segmental k-mean training procedure for connected word recognition, *AT&T Tech* 1986., 65(3):21-31.

[7] Lawrence R. Rabiner and Fellow. A tutorial on Hidden Markov Models and selected applications in speech recognition, *Proceedings of the IEEE*, Vol. 77, No. 2, February 1989.