

A Local Neural Classifier for the Recognition of EEG Patterns Associated to Mental Tasks

José del R. Millán, Josep Mouriño, *Student Member, IEEE*, Marco Franzé, Febo Cincotti, Markus Varsta, Jukka Heikkonen, and Fabio Babiloni

Abstract—This paper proposes a novel and simple local neural classifier for the recognition of mental tasks from on-line spontaneous EEG signals. The proposed neural classifier recognizes three mental tasks from on-line spontaneous EEG signals. Correct recognition is around 70%. This modest rate is largely compensated by two properties, namely low percentage of wrong decisions (below 5%) and rapid responses (every 1/2 s). Interestingly, the neural classifier achieves this performance with a few units, normally just one per mental task. Also, since the subject and his/her personal interface learn simultaneously from each other, subjects master it rapidly (in a few days of moderate training). Finally, analysis of learned EEG patterns confirms that for a subject to operate satisfactorily a brain interface, the latter must fit the individual features of the former.

Index Terms—Brain–computer interface, local neural classifier, spontaneous EEG activity.

I. INTRODUCTION

OVER the last years evidence has accumulated to show the possibility to recognize a few mental tasks from on-line EEG signals and have them associated to simple commands such as “move cursor up” (e.g., [1]–[6]). This alternative communication channel is called a brain–computer interface (BCI). Some groups—especially Wolpaw’s [1] and Birbaumer’s [5]—have demonstrated that some subjects can learn to control their brain activity through appropriate training in order to generate fixed EEG patterns that the BCI transforms into external actions. Anderson’s approach [3] lies at the other extreme in that only the BCI is trained. We, as other teams—in particular Pfurtscheller’s [7], but see also Penny and Roberts’ work [6]—adopt a broader approach based on a mutual learning process whereby the user and the BCI are coupled and adapt to each other.

We build *individual* BCI. We cannot expect a classifier built with EEG data from a few persons to generalize universally across subjects since no two people are the same, especially

from the physiological standpoint.¹ For instance, it has been found that string musicians have larger somatosensory cortical areas associated to the fingers than average [8]. Also, it is well known that there exists considerable variability in alpha frequency between subjects (e.g., [9]). These few examples illustrate that our exact cortical organization and dynamic are individual and reflect our personal life experience. We bring the individuality principle still further in that every single subject may choose the strategies to undertake those mental tasks (e.g., thinking on moving a finger, the hand, or the whole arm). In this way, we believe firmly, users can regularly generate those EEG patterns that are better distinguished by their personal BCI. In Birbaumer’s and Wolpaw’s approaches, subjects are also let at their own in what respect the selection of strategies to achieve the desired EEG patterns.

BCIs are based on the analysis of EEG phenomena associated to spontaneous mental activity. Thus, Birbaumer *et al.* [5] measure shifts of slow cortical potentials over the vertex. Other groups look at local variations of EEG rhythms. Pfurtscheller’s team works with event-related desynchronization over sensorimotor cortex at specific time intervals after the subject is commanded to undertake a mental task [2], [7]. Wolpaw and coworkers focus on the sensorimotor cortex too, but they measure continuous changes of the mu and beta rhythms amplitude [1], [10]. We analyze also continuous variations of EEG rhythms, but not only over the sensorimotor cortex and on specific frequency bands. The reason is that a number of neurocognitive studies has demonstrated that different mental tasks activate local cortical areas at different extents (e.g., imagination of movements [11], [12]; subtractions² [13], [14]; and cube rotation [15]). Our approach seeks to discover individual EEG patterns for three mental tasks embedded in the continuous EEG signal.

There is a large consensus that an efficient and practical BCI should exhibit the following properties: 1) high scores of correct recognition; 2) low percentage of wrong decisions to avoid users’ frustration; and 3) rapid responses, on the order of a second, to increase the bit rate of the communication channel. In this paper, we investigate different classifiers for the recognition of mental tasks from on-line spontaneous EEG signals. It turns out that, of these different classifiers we have explored, only

Manuscript received October 6, 2000; revised December 14, 2001. This work was supported by the European ESPRIT Programme (LTR Project 28193—ABI). This article is an extended version of a paper selected among the most outstanding contributions to IJCNN2000.

J. del R. Millán, J. Mouriño, and M. Franzé are with ISIS, Joint Research Centre of the EC, Ispra (VA), Italy.

F. Cincotti is with Fondazione Santa Lucia, Rome, Italy

M. Varsta and J. Heikkonen are with the Laboratory of Computational Engineering, Helsinki University of Technology, Espoo, Finland.

F. Babiloni is with the Human Physiology Institute, Università La Sapienza, Rome, Italy. He is also with the Laboratory of Computational Engineering, Helsinki University of Technology, Espoo, Finland.

Publisher Item Identifier S 1045-9227(02)04442-9.

¹Quite significantly, in Birbaumer’s and Wolpaw’s approaches the exact thresholds for the fixed EEG patterns to be recognized are manually tuned for each individual subject.

²An alternative arithmetic operation could be multiplication, as utilized by Anderson [3]. In this latter case the dominant active area seems to be the frontal cortex—working memory—that being involved in practically any mental task to various extents may hinder recognition by a reduced number of electrodes.

a simple kind of local neural network fulfills the requirements above. This local neural classifier is embedded in a portable BCI, called adaptive brain interface (ABI), which allows subjects to operate brain-actuated devices such as a virtual keyboard, a computer game and a wheelchair (interested readers may visit the site <http://sta.jrc.it/abi> for some details and illustrations of these applications).

II. EXPERIMENTAL PROTOCOL

An obstacle to the deployment of BCI systems is the acquisition of high-quality EEG signals outside shielded laboratory settings by means of robust, easy-to-use equipment and a few electrodes. To this end we have built a portable EEG system.

EEG signals have been acquired using this portable system as well as a clinical equipment. In both, scalp electrodes are referred to a linked-ear reference. The portable EEG system has eight scalp electrodes, whereas the clinical system supports 26 scalp electrodes and six non-EEG channels. In particular, EMG and EOG signals are recorded to confirm the absence of electromyographic activity during the movement imagination tasks and to detect ocular artifacts, respectively. We have removed EEG samples where muscular activity of the arms was detected.

We have gathered data from eight voluntary young subjects (five males and three females) according to the following experimental protocol. All subjects, except one, are right-handed.

We want our experimental protocol to fit the real conditions in which users would work. Thus, people operating a cursor decide spontaneously where to move toward without waiting for and responding to external cues, keep that action until their target is achieved, and want fast responses and feedback. This means that the experimental protocol cannot depend on external events—i.e., decisions are spontaneous and self-paced.

Another critical aspect of the experimental protocol is the set of *mental tasks* to recognize (and differentiate from each other). We utilize both cognitive tasks (e.g., arithmetic) and motor-related ones (e.g., imagination of left-hand movement). Tasks are chosen so that they activate cortical areas at different extents, as mentioned in the previous section. The five mental tasks considered in this study are “relax,” imagination of “left” and “right” hand movements, “cube rotation,” and “subtraction.” The tasks consist on getting relax, imagining repetitive self-paced movements of the hand, visualizing a spinning cube, and performing successive elementary subtractions by a fixed number (e.g., $64 - 3 = 61$, $61 - 3 = 58$, $58 - 3 = 55$, . . .), respectively. Relax is done with eyes closed, whereas the other tasks are performed with eyes opened. The different experiments reported later aim at recognizing several combinations of two or three tasks.

As a baseline for these five tasks, we use the *average resting pattern* computed over an initial period of 60 s. That is, when the BCI is turned on a signal indicates to the subject to remain with eyes opened but not undertaking any particular task. Then, after 1 min, another signal tells the subject to start operating the BCI.

During a recording session, the subject is seated and spontaneously concentrates on a mental task. The subject performs a task during 10 to 15 s, and he/she chooses when to stop doing it

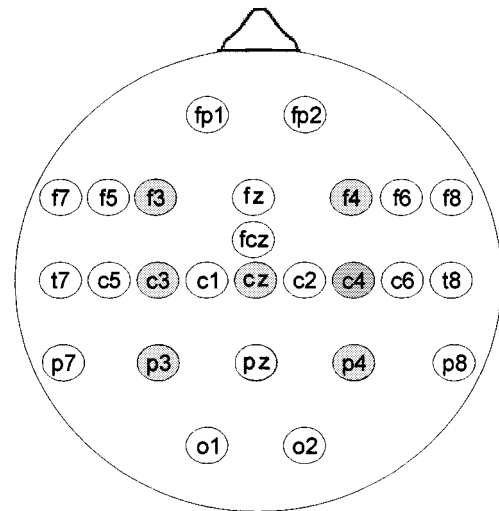


Fig. 1. Electrode montage. All the signals are recorded with respect to a linked-ear reference. The clinical system acquires EEG signals from all 26 electrodes shown in the figure, while the portable system records only from those eight indicated in gray.

and which task to undertake next. Each recording session lasts about 5 min. In a day, subjects perform four recording sessions with a break of 5 to 10 min in between. For the training and testing of the classifiers, the subject informs an operator of the task he/she is ready to perform next. Currently, this is done by pronouncing aloud the name of the task (e.g., “cube” or “left”). The operator enters manually the corresponding label. Then we remove from the recording 2 s before and 2 s after every transition to clean off the artifacts introduced by this “communication” and to reduce the risk of mislabeling.

III. SIGNAL PROCESSING

The sampling rates of the portable and clinical EEG systems are 128 and 400 Hz, respectively. EEG potentials are recorded from the electrodes shown in Fig. 1.

The main operation in the temporal domain is a *spatial filtering* whereby new potentials should represent better the cortical activity due only to local sources below the electrodes. In particular, we compute a *surface Laplacian (SL)* derivation [16] over the six centro-parietal electrodes C3, Cz, C4, P3, Pz, and P4. After that, we apply a second-order 4–45 Hz bandpass Butterworth filter. With the clinical EEG system the SL is computed globally by means of a spherical spline of order 2 [17], [18] using all 26 channels to minimize variance in the estimation. On the contrary, with the portable EEG system the SL is estimated locally using a finite difference method that, for each position of interest, subtracts the mean activity at neighboring electrodes (for details, see [10], [19]). The superiority of SL-transformed over raw potentials for the operation of BCI has been demonstrated in different studies (e.g., [10], [20]).

The analyzed features are the power spectrum density components (estimated with a *Welch periodogram*) of each SL channel obtained and transformed in the following way. With the portable EEG system, sequences are 1-s long and segments of 1/2 s are averaged. This gives a frequency resolution of 2 Hz. A *Hann window* is applied to each segment, and the overlapping

between the segments is 50%. Then, the power components are transformed in dB and the values in the frequency band 8–30 Hz are normalized according to the total energy in that band. Thus an EEG sample is represented by 72 features (6 channels times 12 components each). The periodogram, and hence an EEG sample, is computed every 1/2 s. With the conventional EEG system, the length of the sequences and segments are 2 and 1 s, respectively. In this case, the frequency resolution is 1 Hz and EEG samples have 138 features (six channels times 23 components each).

Finally, it is worth noting that, for our experimental protocol, periodogram features lead to better or similar performances than more elaborated features such as parameters of autoregressive (AR) models and wavelets [21].

IV. NEURAL CLASSIFIERS

A. Linear Classifiers

Initially, we have explored the use of two kinds of linear classifiers for the recognition of mental tasks. The first classifier is the well-known Fisher linear discriminant [22], whereas the second is based on the signal space projection (SSP) algorithm [23]. In a previous study, it was found that SSP classifiers achieve significantly better recognition rates than Fisher classifiers [24]. Thus, in the sequel we skip any discussion on Fisher linear discriminant analysis.

In the SSP method, given p vectors of n -dimensional “patterns” (S_1, S_2, \dots, S_p), the p components of the “activation” vector $\hat{A}(t) = \mathbf{S}^+M(t)$ weight the presence of each pattern in $M(t)$, the n -dimensional feature vector computed from the incoming EEG signals. \mathbf{S}^+ is the pseudoinverse of the projection matrix \mathbf{S} whose columns are the patterns (S_1, S_2, \dots, S_p).

SSP is similar to principal component analysis and related methods in that patterns S_i can be estimated directly from data. However, contrarily to those methods, SSPs patterns do not need to be orthogonal. In particular, each pattern S_i is estimated as the *average* of the available training data for the i th mental task. Alternatively, a SSP classifier can consist of several patterns computed by means of clustering techniques such as self-organizing maps (SOM) [25]. In this case, patterns of given mental task are obtained by training a SOM on samples of that task only.

A given EEG sample $M(t)$ is classified into the i th mental task when this is the component of the vector $\hat{A}(t)$ with the highest value. Even though this is probably the most elemental SSP-based classifier, the results achieved are quite promising.

B. Nonlinear Classifiers

SSP classifiers do not exploit class information during training. In this section, we analyze compact and local neural networks trained in a supervised manner.

Of the many multilayer perceptron (MLP) architectures we have tried, the strongest classifier is a ten-member committee trained with early stopping. All the member networks have a single hidden layer made of ten units and one output layer, where the output units (one per mental task) have a linear transfer function. The response of the committee is the average of the members’ output. The problem of overfitting is fought

with regularization methods, which favors simple models, and early stopping.

We have explored different types of local neural classifiers such as LVQ [25] and regularized RBF networks [26], [27]. Here, we will only describe a simple local neural classifier that achieves the best results and that is successfully used in the brain interfaces of all the users we have worked with.

In our local neural classifier, every unit represents a prototype of one of the mental tasks (or classes) to be recognized. The challenge is to find the appropriate position of the prototypes in the high-dimensional input space described above in order to differentiate the desired classes. The basic idea is that, during training, units are pulled toward the EEG samples of the mental task they represent and are pushed away from EEG samples of other tasks.

In a statistical framework, the discriminant function of class C_k for sample x is

$$y_k(x) = P(C_k | x) = \frac{p(x | C_k)P(C_k)}{p(x)}$$

where $P(C_k | x)$ is the posterior probability of class C_k , $P(C_k)$ denotes the prior probability of class C_k , and $p(x | C_k)$ is the class-conditional probability density function of x given that it belongs to class C_k . Assuming that each class-conditional density function is taken to be an independent normal distribution, equal prior class probabilities and dropping constant terms, then

$$y_k(x) = \frac{|\Sigma_k|^{-1/2} \exp(-1/2(x - \mu_k)^T \Sigma_k^{-1} (x - \mu_k))}{\sum_{j=1}^{N_c} |\Sigma_j|^{-1/2} \exp(-1/2(x - \mu_j)^T \Sigma_j^{-1} (x - \mu_j))} \quad (1)$$

where μ_k is the prototype (mean) of class C_k , Σ_k is the covariance matrix of class C_k , and N_c is the number of classes to be recognized. This is essentially a Bayesian classifier based on *quadratic discriminant analysis* [22].

In practice, a sample is assigned to the class C_k with the nearest prototype based on the *Mahalanobis* distance 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 experiments below, the probability threshold is set to 0.9.

To estimate the initial values of the units (i.e., the position μ_k and the receptive field Σ_k) we can use the maximum likelihood approach

$$\mu_k = \frac{1}{N_k} \sum_{n=1}^{N_k} x_k^n \quad (2)$$

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

where N_k denotes the number of training samples belonging to the class C_k and x_k^n is the n th sample of the class C_k in the training set.

We can go further and try to improve these initial estimations iteratively. First, we optimize the position of the prototypes of the different classes to minimize the mean square error through

gradient descent. Assuming diagonal covariance matrices,³ then 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 α is the learning rate, 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 (1), and y_j is the probability of the remaining classes. Second, after every iteration over the training set, we estimate again the new value of Σ_k using (3).

Finally, 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. These prototypes can be initially estimated with any clustering algorithm such as SOM [25].

V. EXPERIMENTAL RESULTS

Initial experiments were done without feedback and subjects' performance was analyzed off-line—i.e., learning and testing the classifiers on EEG data previously recorded. The purpose of the experiments was to investigate the limits of the neural classifier. After assessing off-line the validity of the local neural classifiers, a second group of experiments investigated their *on-line* performance. To this end, three subjects were trained in the presence of feedback for a few consecutive days (from three to five). One of the subjects, MJ, participated in all the experiments.

Feedback is provided by means of several colored buttons, one for each mental task to be recognized. A button lights up when the arriving EEG sample is classified as belonging to the corresponding mental task. Training is done as follows. The first day, the subject does not receive any feedback. With the EEG data recorded this day, we train off-line his/her first individual neural classifier. This neural classifier is then embedded in the BCI that is operated the following day. Again, the second individual classifier is tuned with the EEG data recorded this second day; this classifier is used the third day, and so on.

Initial experiments aimed at differentiating the two motor-related mental tasks from each other (i.e., imagination of left and right movements) with the clinical EEG system. To this end, five subjects (three males and two females) participated in a single recording session and data were evaluated off-line. This was the first time subjects used the system, and so they did not receive any feedback. After assessing the feasibility of the approach, we proceeded to the recognition of three mental tasks with the portable EEG system. In this case, four subjects (three males and one female) were trained either without or in the presence of feedback. The purpose of the experiments without feedback was to investigate the limits of the neural classifier.

Table I reports the generalization results for the five subjects obtained with the basic SSP classifier (i.e., one pattern per mental task estimated as the average of the training data) and the best SOM-based SSP classifier (i.e., the classifier made of $n \times m$ patterns estimated with the SOM algorithm that has the best performance, with n and m ranging from one to five). The basic

³We could have used the full covariance matrices. However, there exist two strong reasons not to do so. First, a very large number of samples is required to estimate the full matrices accurately. Second, their manipulation is very costly computationally and may prevent real-time operation of the BCI.

TABLE I
GENERALIZATION RATES OF SSP CLASSIFIERS FOR RIGHT-HAND AND LEFT-HAND MOVEMENT IMAGINATION TASKS WITH THE CLINICAL EEG SYSTEM. IN THIS CASE, THE INPUT TO THE CLASSIFIER IS A VECTOR WITH 138 COMPONENTS. 75% OF THE AVAILABLE EEG SAMPLES ARE USED TO TRAIN THE CLASSIFIER AND THE REMAINING 25% TO TEST THE GENERALIZATION CAPABILITIES AND THUS THE SUBJECTS' PERFORMANCE

| Type | # Units | Subject | Right | Left | Total |
|-----------|---------|---------|-------|------|-------|
| average | 1 | CL | 89% | 53% | 71% |
| | 1 | MJ | 75% | 55% | 65% |
| | 1 | RA | 80% | 60% | 70% |
| | 1 | RB | 84% | 57% | 73% |
| | 1 | TA | 100% | 61% | 81% |
| SOM-based | 4 | CL | 78% | 67% | 72% |
| | 4 | MJ | 60% | 60% | 60% |
| | 2 | RA | 86% | 71% | 79% |
| | 2 | RB | 70% | 57% | 59% |
| | 2 | TA | 100% | 67% | 83% |

TABLE II
GENERALIZATION RATES OF THE LOCAL NEURAL CLASSIFIERS FOR LEFT-HAND AND RIGHT-HAND MOVEMENT IMAGINATION TASKS FOR SUBJECTS USING THE CLINICAL EEG SYSTEM. FIGURES IN BRACKETS INDICATE DIFFERENCE IN PERFORMANCE WITH RESPECT TO THE BEST SSP CLASSIFIER FOR THE CORRESPONDING SUBJECT (SEE TABLE I)

| # Units | Subject | Confusion Matrix (%) | | % Correct Responses | % Wrong Responses | |
|---------|---------|----------------------|-------|---------------------|-------------------|----------|
| | | Left | Right | | | |
| 1 | CL | Left | 90 | 10 | 85 (+13) | 15 (-13) |
| | | Right | 20 | 80 | | |
| 1 | MJ | Left | 80 | 10 | 80 (+15) | 10 (-25) |
| | | Right | 10 | 80 | | |
| 1 | RA | Left | 63 | 12 | 75 (-4) | 6 (-15) |
| | | Right | 0 | 88 | | |
| 1 | RB | Left | 71 | 14 | 76 (+3) | 12 (-15) |
| | | Right | 10 | 80 | | |
| 1 | TA | Left | 78 | 11 | 83 (+0) | 11 (-6) |
| | | Right | 11 | 89 | | |

SSP classifiers recognize quite well the right movement task for all five subjects (from 75% to 100%), but perform poorly for the left movement task (from 53% to 61%). On the other hand, the SOM-based SSP classifiers do not recognize the right movement task as well as the basic classifiers. But their performances on this task are still satisfactory for three of the subjects (from 78% to 100%) and, more importantly, the recognition rates of the left movement task increase considerably (they range now in between 57% and 71%). Thus the SOM-based SSP classifiers are to be preferred for they provide a more balanced recognition. In addition, it is worth noting that these classifiers reach their best results with a surprisingly small number of spatial patterns (either two or four for each mental task). Nevertheless, the SOM-based classifiers achieve more than 75% correct recognition for the combined task only for two of the subjects, for one other is close to that figure, and for the remaining two are quite far.

Table II reports the performance of the local neural classifier for the subjects using the clinical EEG system. Compared

TABLE III
GENERALIZATION RATES OF THE LOCAL NEURAL CLASSIFIERS FOR RELAX, LEFT-HAND AND RIGHT-HAND MOVEMENT IMAGINATION TASKS FOR TWO OF THE SUBJECTS USING THE PORTABLE EEG SYSTEM TRAINED WITHOUT FEEDBACK. IN THIS CASE, THE INPUT TO THE NEURAL CLASSIFIER IS A VECTOR WITH 72 COMPONENTS

| # Units | Subject | Confusion Matrix (%) | | | % Correct Responses | % Wrong Responses | |
|---------|---------|----------------------|------|-------|---------------------|-------------------|---|
| | | Relax | Left | Right | | | |
| 2 | MJ | Relax | 100 | 0 | 0 | 52 | 0 |
| | | Left | 0 | 42 | 0 | | |
| | | Right | 0 | 0 | 36 | | |
| 1 | MJR | Relax | 92 | 0 | 0 | 63 | 5 |
| | | Left | 0 | 51 | 5 | | |
| | | Right | 0 | 9 | 56 | | |

TABLE IV
ON-LINE CLASSIFICATION PERFORMANCE FOR THREE SUBJECTS AT THE END OF SEVERAL CONSECUTIVE DAYS OF TRAINING WITH FEEDBACK USING THE PORTABLE EEG SYSTEM

| # Units | Last Day of Training | Subject | Confusion Matrix (%) | | | |
|---------|----------------------|---------|----------------------|------|-------|----|
| | | | Relax | Left | Right | |
| 1 | 3 | MJ | Relax | 100 | 0 | 0 |
| | | | Left | 0 | 57 | 2 |
| | | | Right | 0 | 9 | 52 |
| 1 | 5 | CGS | Relax | 93 | 0 | 0 |
| | | | Left | 0 | 61 | 6 |
| | | | Right | 0 | 4 | 85 |
| 1 | 4 | MC | Relax | 76 | 0 | 1 |
| | | | Left | 0 | 24 | 6 |
| | | | Right | 0 | 9 | 21 |

with results in Table I, this local classifier clearly outperforms the SSP classifiers. Column "Correct Responses" (percentage of correct combined responses) shows that the local neural network performs much better for three out of five subjects, equally for one other and slightly worse for the remaining subject. For all five subjects the local neural classifiers achieve more than 75% correct recognition, and for three of them generalization is even over 80%. But the clearest evidence in improvement comes from the column "Wrong Responses." The local neural classifier makes much less errors than the SSP classifiers for all the subjects. For instance, for subject RA—who achieves 79% correct recognition with the SOM-based SSP classifier and 75% with the local network—wrong responses decrease from 21% to only 6%. This is extremely important from a practical point of view. However, the percentage of wrong responses is higher than desirable (in between 10% and 15% in most cases) although significantly less than the complementary to correct recognition figures (except for one of the subjects). It is worth noting that these generalization results are obtained with only one unit per mental task in all cases. Thus a neural classifier consists of just two units.

Table III gives the generalization results for two of the subjects using the portable EEG system trained without feedback. These results show the first two successful implementations of BCI's recognizing three mental tasks, namely relax, right-hand

imagined movement and left-hand imagined movement. The most appealing feature of these classifiers is the low percentage of wrong responses, 5% for MJR and even 0% for MJ. This excellent property does not come for free; the price to pay is a modest percentage of correct recognition for the tasks left-hand and right-hand movement imagination. Nevertheless, this does not preclude its practical operation, as the neural classifier is embedded in a BCI that makes decisions every 1/2 s. Thus, recognition of movement imagination tasks takes 1.5 s in average for subject MJ and 1 s for subject MJR. As for the previous subjects, the neural classifiers have a surprisingly small number of units per mental task.

Once we have assessed the validity of the local neural classifiers *off-line*—i.e., learning and testing the classifiers on EEG data previously recorded—we proceed to investigate their *on-line* performance. To this end, three subjects have been trained in the presence of feedback with the portable system for a few consecutive days (from three to five). One of the subjects is again MJ, whereas the other two had no experience with BCI before starting their training. In addition, subject CGS is left-handed. Table IV summarizes the performance of these subjects.

Figs. 2 and 3 show the on-line performance of subjects MJ and CGS over time, respectively. Subject MJ achieves recognition rates of 100%, 57%, and 52%—for relax, left-hand movement, and right-hand movement, respectively—whereas the wrong responses are just 0%, 2% and 9%. MJ's decrease in performance in the first day with respect to his previous level (Table III) can be due to two reasons. First, the subject did not practice the mental tasks between the two sessions. Second, he reported that, the first day, he was slightly disturbed by the feedback he was receiving. Nevertheless, MJ improves almost linearly his performance well over his previous level in only three days. Subject CGS achieves a more impressive control of his BCI. After just five days of moderate training he achieves recognition rates of 93%, 61% and 85%—for relax, left-hand movement and right-hand movement, respectively—whereas the wrong responses are only 0%, 6% and 4%. However, being the first time CGS works with a BCI, the evolution of his performance is not linear as for MJ.

As for the third subject trained with feedback, MC only achieved discrete on-line performance. On the fourth day of training, MC reached recognition rates of 76%, 24% and 21%—for relax, left-hand movement and right-hand movement, respectively—whereas wrong responses were 1%, 6% and 9%. However, based on the evolution of the other two subjects, we think that these figures should greatly improve with a few more training days. Unfortunately, MC could not come additional days in a row and since the experiments aimed to investigate how long it takes a subject to master the BCI, we decided not to continue weeks after.

To end this section, we explore the possibility of building a BCI based on the recognition of motor-related tasks versus cognitive tasks. Table V reports the generalization results for subject MJ when using relax, cognitive, and motor-related mental tasks. As motor-related task we have chosen left-hand movement imagination, the task for which MJ achieves better results. As cognitive task we have explored both subtraction and cube

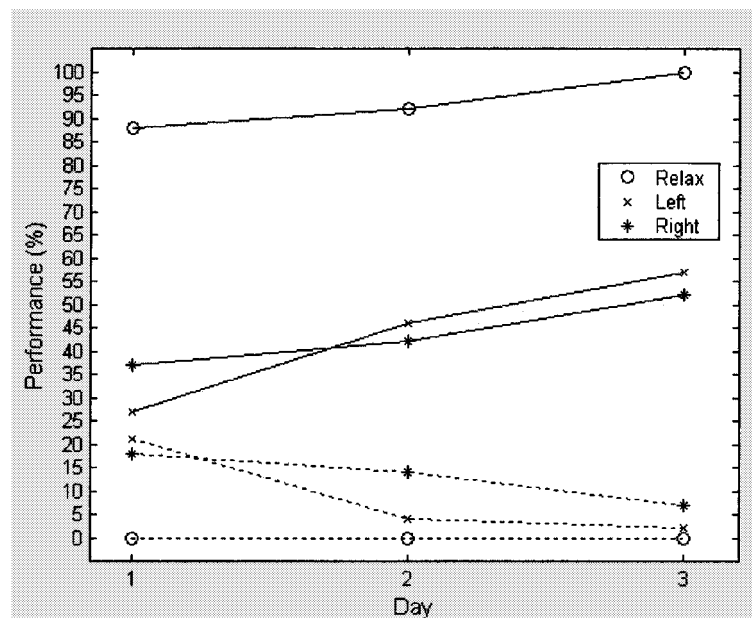


Fig. 2. Evolution of the on-line classification performance for subject MJ over three consecutive days. Solid lines represent the percentage of correct recognition for the corresponding task, whereas dashed lines represent the percentage of wrong recognition.

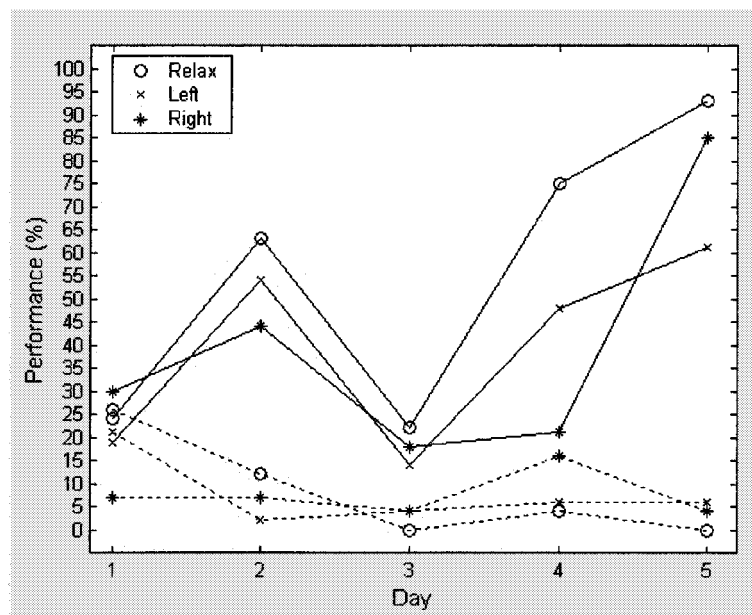


Fig. 3. Evolution of the on-line classification performance for subject CGS over five consecutive days. Solid lines represent the percentage of correct recognition for the corresponding task, whereas dashed lines represent the percentage of wrong recognition. This subject did not have any previous experience with BCI before this experiment.

rotation. Then, we have built two classifiers, the first based on the three tasks relax-subtraction-left and the second on the tasks relax-cube-left. As shown in the table, both classifiers perform exceptionally well. Indeed, they achieve over 90% of correct recognition on the combined task, while keeping the wrong responses below 2%.

VI. CONCLUSION

BCI systems evaluate subjects' performance in two different ways. On the one hand, performance is measured as a combina-

tion of many single responses, which are either added up until some threshold is reached [1], [10] or averaged and the result compared to some thresholds [5]. On the other hand, the BCI responds to single EEG samples (or single-trial EEG), which either are generated in response to external events and recorded only at specific times [2], [7] or are recorded continuously [3], [6] and even generated in a self-paced manner [4]. In this paper, we have followed the latter approach that, in principle, is more flexible—compared to event-related settings—and should allow faster responses—compared to making decisions only at the end of a period of time. On the other hand, it is true that achieving

TABLE V
GENERALIZATION RATES OF THE LOCAL NEURAL CLASSIFIERS FOR SUBJECT MJ WHEN USING THE PORTABLE EEG SYSTEM AND CONCENTRATING ON RELAX, A MOTOR-RELATED TASK (LEFT-HAND MOVEMENT) AND A COGNITIVE MENTAL TASK (EITHER SUBTRACTION OR CUBE ROTATION). IN BOTH CASES, THERE IS ONLY 1 UNIT PER MENTAL TASK

| Confusion Matrix (%) | | | | | | | |
|----------------------|-------|--------|------|-------|-------|------|------|
| | Relax | Subtr. | Left | | Relax | Cube | Left |
| Relax | 95 | 0 | 0 | Relax | 95 | 0 | 0 |
| Subtr. | 1 | 86 | 3 | Cube | 1 | 89 | 1 |
| Left | 0 | 1 | 90 | Left | 0 | 0 | 96 |

performances near to 100% correct responses is far more difficult with our approach, especially if more than two mental tasks are to be recognized. However, the fact that the task is more difficult does not imply that it is impossible to develop practical brain-actuated devices or to scale up the number of mental tasks. As discussed below, an operational BCI does not require perfect recognition, but only a considerably low level of errors and quick response times. Furthermore, to the best of our knowledge this is the only approach where more than two mental tasks have been reliably recognized.

Experimental results show that the proposed local neural classifier achieves recognition rates of 70% (or more) for three mental tasks from on-line spontaneous EEG signals. This figure is more than twice random classification, which for three tasks is 33.3%. This modest rate is largely compensated by two properties: wrong responses are below 5% and it makes decisions every 1/2 s. The first property implies that recognition is quite *robust* since the neural classifier hardly takes one class for another. This is extremely important for the user to accept a BCI as a reliable system. From the second property it follows that recognition takes 1 s in average, thus allowing the BCI to respond quickly. It is worth noting that these results are obtained with only one unit per mental task in all cases. This simplicity of the neural architecture makes our BCI very well suited for real-time operation.

For the sake of comparison, the performances of the MLP committee (ten networks with ten hidden units each), the LVQ network (16 codebooks per class for a total of 48 units) and the RBF network (48 units too) for subject MJ when trained with feedback are similar and slightly below 70% of correct classification. This recognition rate is similar to that of the new local classifier described in this paper (Fig. 2 and Table IV). However, the latter classifier just makes 4% of wrong decisions whereas the misclassification rates of the former rise to 30%. In addition, these MLP, LVQ, and RBF networks utilized substantially more computational resources than the new local neural classifier.

To the best of our knowledge, only Pfurtscheller's group has tried the recognition of three mental tasks. In [7] they achieved off-line classification rates in between 84% and 94% for three subjects who performed actual movements of their limbs. In [2] they attempted on-line recognition while four subjects imagined movements of their limbs, obtaining 45% of correct classification and 30% of wrong responses. These results were obtained with a BCI that responded to EEG samples recorded at specific

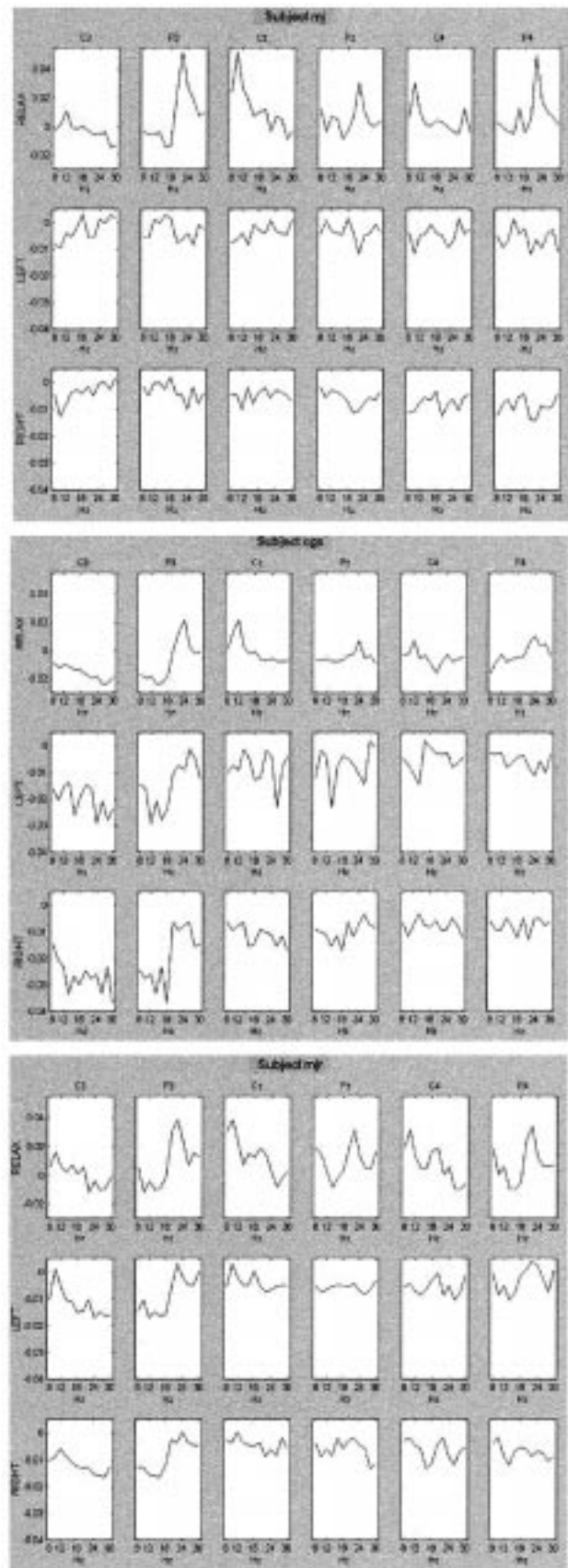


Fig. 4. Prototypes learned for subjects MJ, CGS and MJR, from top to bottom.

times, namely every 10 s or more. In this paper, we report satisfactory on-line, continuous recognition of three mental tasks,

namely 70% (or more) of correct classification and 5% (or less) of wrong responses.

Most groups have limited to use MLPs to build adaptive EEG-based classifiers (e.g., [3], [6]). We have also investigated their use [21], but our results clearly indicate that local neural classifiers—i.e., classifiers made of prototypes associated to the different mental tasks—are to be preferred. Pfurtscheller's group has also explored local neural classifiers [2] based on LVQ. We have explored different types of local neural classifiers such as LVQ and RBF networks. The novel and simple local neural classifier reported in this paper performs better than all the others for our experimental protocol and has made possible to develop individual BCI for all eight subjects we have worked with despite the short training time of most of them. Indeed, since the user and his/her personal BCI learn simultaneously from each other, subjects master it rapidly. This portable BCI is being used to operate some brain-actuated devices (see Introduction).

Analysis of learned EEG patterns confirms that for a subject to operate satisfactorily a BCI, the latter must fit the individual features of the former. This confirms that building individual interfaces greatly increases the likelihood of success. Fig. 4 clearly illustrates this claim. It shows the learned prototypes (or patterns) for three of the subjects who master the three mental tasks relax, left and right movement imagination. Briefly, there hardly exist features shared by these subjects. The same holds for the remaining subjects.

The work described here is being extended along two directions. The first one is the recognition of a larger set of EEG patterns. Initial results suggest that it is feasible to distinguish five mental tasks at the same levels of recognition reported in this paper for three mental tasks. The second area of current investigation is the incorporation of temporal information in order to improve the recognition rates. In particular, a previous study with off-line EEG signals confirms that an artificial neural network distinguishes EEG patterns better if it uses the temporal dynamics of brain activity [28]. This is not surprising since EEG signals carry temporal information. Ongoing research is exploring the use of time-delay approaches—i.e., the response of the neural classifier is based on the current as well as several previous feature vectors—and recurrent networks.

REFERENCES

- [1] J. R. Wolpaw and D. J. McFarland, "Multichannel EEG-based brain-computer communication," *Electroenceph. Clin. Neurophysiol.*, vol. 90, pp. 444–449, 1994.
- [2] J. Kalcher, D. Flotzinger, C. Neuper, S. Göllly, and G. Pfurtscheller, "Graz brain-computer interface II," *Med. Biol. Eng. Comput.*, vol. 34, pp. 382–388, 1996.
- [3] C. W. Anderson, "Effects of variations in neural network topology and output averaging on the discrimination of mental tasks from spontaneous electroencephalogram," *J. Intell. Syst.*, vol. 7, pp. 165–190, 1997.
- [4] J. del R. Millán, J. Mouriño, M. G. Marciani, F. Babiloni, F. Topani, I. Canale, J. Heikkonen, and K. Kaski, "Adaptive brain interfaces for physically-disabled people," in *Proc. 20th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 1998.
- [5] N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kübler, J. Perelmouter, E. Taub, and H. Flor, "A spelling device for the paralyzed," *Nature*, vol. 398, pp. 297–298, 1999.
- [6] W. D. Penny and S. J. Roberts, "EEG-based communication via dynamic neural-network models," in *Proc. Int. Joint Conf. on Neural Networks*, 1999.

- [7] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clin. Neurophysiol.*, vol. 110, pp. 787–798, 1999.
- [8] T. Elbert, C. Pantev, C. Wienbruch, B. Rockstroh, and E. Taub, "Increased cortical representation of the fingers of the left hand in string players," *Science*, vol. 270, pp. 305–307, 1995.
- [9] M. Doppelmayr, W. Klimesch, T. Pachinger, and B. Ripper, "Individual differences in brain dynamics: Important implications for the calculation of event-related band power," *Biol. Cybern.*, vol. 79, pp. 49–57, 1998.
- [10] D. J. McFarland, L. M. McCane, S. V. David, and J. R. Wolpaw, "Spatial filter selection for EEG-based communication," *Electroenceph. Clin. Neurophysiol.*, vol. 103, pp. 386–394, 1997.
- [11] G. Pfurtscheller, "Event-related synchronization (ERS): An electrophysiological correlate of cortical areas at rest," *Electroenceph. Clin. Neurophysiol.*, vol. 83, pp. 62–69, 1992.
- [12] G. Pfurtscheller and C. Neuper, "Motor imagery activates primary sensorimotor area in humans," *Neurosci. Lett.*, vol. 239, pp. 65–68, 1997.
- [13] Y. Sakurai, T. Momose, M. Iwata, Y. Sasaki, and I. Kanazawa, "Activation of prefrontal and posterior superior temporal areas in visual calculation," *J. Neurol. Sci.*, vol. 139, pp. 89–94, 1996.
- [14] S. Dehaene, *The Number Sense: How the Mind Creates Mathematics*. Oxford, U.K.: Oxford Univ. Press, 1997.
- [15] P. Rappelsberger and H. Petsche, "Probability mapping: Power and coherence analysis of cognitive processes," *Brain Topography*, vol. 1, pp. 46–54, 1988.
- [16] B. Hjorth, "An on-line transformation of EEG scalp potentials onto orthogonal source derivations," *Electroenceph. Clin. Neurophysiol.*, vol. 39, pp. 526–530, 1975.
- [17] F. Perrin, J. Pernier, O. Bertrand, and J. F. Echallier, "Spherical spline for potential and current density mapping," *Electroenceph. Clin. Neurophysiol.*, vol. 72, pp. 184–187, 1989.
- [18] —, "Corrigendum EEG 02 274," *Electroenceph. Clin. Neurophysiol.*, vol. 76, p. 565, 1990.
- [19] P. B. Zhou, *Numerical Analysis of Electromagnetic Fields*. Berlin, Germany: Springer-Verlag, 1993.
- [20] F. Babiloni, F. Cincotti, L. Lazzarini, J. del R. Millán, J. Mouriño, M. Varsta, J. Heikkonen, L. Bianchi, and M. G. Marciani, "Linear classification of low-resolution EEG patterns produced by imagined hand movements," *IEEE Trans. Rehab. Eng.*, vol. 8, pp. 186–188, 2000.
- [21] M. Varsta, J. Heikkonen, J. del R. Millán, and J. Mouriño, "Evaluating the performance of three feature sets for brain-computer interfaces with an early stopping MLP committee," in *Proc. 15th Int. Conf. Pattern Recognition*, 2000.
- [22] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. San Diego, CA: Academic, 1990.
- [23] M. A. Uusitalo and R. J. Ilmoniemi, "Signal-space projection method for separating MEG or EEG into components," *Med. Biol. Eng. Comput.*, vol. 35, pp. 135–140, 1997.
- [24] F. Babiloni, F. Cincotti, L. Bianchi, G. Pirri, J. del R. Millán, J. Mouriño, S. Salinari, and M. G. Marciani, "Recognition of imagined hand movements with low resolution surface Laplacian and linear classifiers," *Med. Eng. Phys.*, vol. 23, pp. 323–328, 2001.
- [25] T. Kohonen, *Self-Organizing Maps*, 2nd ed. Berlin, Germany: Springer-Verlag, 1997.
- [26] C. M. Bishop, "Improving the generalization properties of radial basis function neural networks," *Neural Comput.*, vol. 3, pp. 579–588, 1991.
- [27] Z. Wang and T. Zhu, "An efficient learning algorithm for improving generalization performance of radial basis function neural networks," *Neural Networks*, vol. 13, pp. 545–553, 2000.
- [28] M. Varsta, J. del R. Millán, and J. Heikkonen, "A recurrent self-organizing map for temporal sequence processing," in *Proc. 7th Int. Conf. Artificial Neural Networks*, 1997.

José del R. Millán received the Ph.D. degree in computer science from the University of Politècnica de Catalunya, Barcelona, Spain, in 1992.

He was an Assistant Professor with the University of Politècnica de Catalunya for three years. He is a Research Scientist at the Joint Research Centre of the European Commission. Currently, he is a Visiting Professor at the Swiss Federal Institute of Technology, Lausanne. His work is mainly focused on the design of adaptive sensor-based systems and adaptive biosignal-based interaction.

Josep Mouriño (S'98) received the Bachelor's degree in industrial engineering and the M.Sc. degree in bioengineering from the Technical University of Catalonia, Barcelona, Spain, in 1995 and 1998, respectively.

Since 1997, he has been working at the Joint Research Centre of the European Commission in the project adaptive brain interfaces. He has been mainly involved in the design of the digital signal processing of the brain activity. His areas of interest include acquisition systems, digital signal processing, and pattern classification.

Marco Franzé received the Laurea degree in computer engineering in 1999 from the Politecnico di Milano, Milan, Italy.

He was with the Joint Research Centre of the European Commission in the project adaptive brain interfaces. He is now a software engineer developing business solutions. His interests are related to machine learning and feature selection.

Febio Cincotti received the Laurea degree cum laude in electronic engineering in 1998 from the University of Rome "La Sapienza," where he is pursuing the Ph.D. degree in biophysics.

Since 1999, he has been with IRCCS Fondazione Santa Lucia of Rome participating in the research on brain-computer interfaces. His research interests include EEG signal recording, analysis, and algorithm development for biological signals processing.

Markus Varsta, photograph and biography not available at the time of publication.

Jukka Heikkonen received the M.Sc. and Dr.Tech. degrees in information technology at the Lappeenranta University of Technology, Finland, in 1991 and 1994, respectively.

Since 1996, he has been working at the Laboratory of Computational Engineering of Helsinki University of Technology, Finland, where he is currently acting as an Academy Fellow of the Academy of Finland. His research interests include computer vision, pattern recognition, and neural networks.

Fabio Babiloni received the Master's degree in electronic engineering, summa cum laude, from the University of Rome "La Sapienza," in 1985 and the Ph.D. degree in neural and computational engineering at the Helsinki University of Technology in 2000.

Currently, he is a Researcher at the Dept. of Human Physiology of Univ. "La Sapienza." His research interests include high-resolution EEG, multimodal integration of EEG, MEG, and fMRI, and brain-computer interfaces.