Relevant EEG Features for the Classification of Spontaneous Motor-Related Tasks

José del R. Millán[1], Marco Franzé[1], Josep Mouriño[1], Febo Cincotti[2], Fabio Babiloni[3]

[1]*ISIS, Joint Research Centre of the EC, 21020 Ispra (VA), Italy*

[2]*Fondazione Santa Lucia, Via Ardeatina 306, 00179 Rome, Italy*

[3]*Dept. Fisiologia Umana e Farmacologia, Università La Sapienza, P.le A. Moro 5, 00185 Rome, Italy*

Abstract. There is a growing interest in the use of physiological signals for communication and operation of devices for the severely motor disabled as well as for healthy people. A few groups around the world have developed *brain-computer interfaces (BCI)* that rely upon the recognition of motor-related tasks (i.e., imagination of movements) from on-line EEG signals. In this paper we seek to find and analyze the set of relevant EEG features that best differentiate spontaneous motor-related mental tasks from each other. This study empirically demonstrates the benefits of heuristic feature selection methods for EEG-based classification of mental tasks. In particular, it is shown that the classifier performance improves for all the considered subjects with only a small proportion of features. Thus, the use of just those relevant features increases the efficiency of the brain interfaces and, most importantly, enables a greater level of adaptation of the personal BCI to the individual user.

Correspondence to: J. del R. Millán

(Tel: +39-0332-785751; Fax: +39-0332-789185; e-mail: jose.millan@jrc.it).

1 Introduction

There is a growing interest in the use of physiological signals for communication and operation of devices for the severely motor disabled as well as for healthy people. In particular, a few groups around the world rely upon the recognition of motor-related tasks (i.e., imagination of movements) from on-line EEG signals (e.g., Wolpaw and McFarland, 1994; Kalcher et al., 1996; Millán et al., 2000). Thus people can for instance communicate using their brain activity by selecting letters from a virtual keyboard. This alternative communication channel is called a *brain-computer interface (BCI).*

BCIs are based on the analysis of EEG phenomena associated to spontaneous mental activity. Thus, Birbaumer et al. (1999) 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 (Kalcher et al., 1996; Müller-Gerking et al., 1999). Wolpaw and coworkers focus on the sensorimotor cortex too, but they measure continuous changes of the mu and beta rhythms amplitude (Wolpaw and McFarland, 1994; McFarland et al., 1997). We analyze also continuous variations of EEG rhythms, but not only on specific frequency bands. Our approach aims at discovering individual EEG features embedded in the continuous EEG signal.

In this paper we seek to find the set of relevant EEG features that best differentiate spontaneous motor-related mental tasks from each other. Any BCI that works with this reduced set will improve its efficiency, both in terms of recognition rate and computational resources. Moreover, the use of just relevant features enables a greater level of adaptation of the personal BCI to the individual user. Indeed, building individual interfaces greatly increases the likelihood of success, as demonstrated for all subjects we have worked with despite the short training time of most of them (Millán et al., 2000).

2 Experimental protocol

In a first set of experiments, five voluntary young subjects (three males and two females, all right-handed) participated in a single recording session where they performed several mental tasks, including imagination of internally self-paced repetitive extensions of the right middle finger (RI) and left middle finger (LI). In these experiments we aim to find and compare the relevant features of different subjects in well-controlled conditions for the recognition of the two motor-related tasks. In a second set of experiments, a sixth subject (male and left-handed), was trained for five consecutive days on two similar motor imagery tasks. The purpose here is to analyze the evolution of the relevant features as the subject masters the BCI for the discrimination of right versus left imagined movements.

2.1 Experiment 1

EEG signals were recorded from 26 scalp electrodes placed onto standard locations of the extended 10-20 international system and referred to a linked-ear reference (see Fig. 1). Sampling frequency was 400 Hz. We also recorded ocular and muscular activity to detect possible eye as well as hand movements, and removed the corresponding EEG samples from further analysis.

At the beginning of a recording session, subjects remained in a resting state[1] for 60 s. The average EEG activity of this period is used as a baseline for subsequent analysis of the mental tasks. Then, subjects performed a given mental task for 10 to 15 s immediately after the operator instructed them to do so. Every subject executed several times each mental task during the recording session.

2.2 Experiment 2

The subject used a portable EEG system with 8 scalp electrodes referred to a linked-ear reference (see Fig. 1). Sampling frequency was 128 Hz. In this case, the subject was not constrained to imagine repetitive extensions of the middle finger, but could choose any kind of movement of the right and left arms.

As in the previous case, the baseline is the average resting pattern computed over an initial period of 60 s. Now, however, the subject chose which of the two tasks to undertake and when to stop doing it. For the training and testing of the classifiers, the subject informed an operator of the task he/she was ready to perform next. A recording session lasted 5 minutes or more. In a day, the subject performed 4 recording sessions with a break of 5 to 10 minutes in between. In subsequent analysis, we removed the transitions between mental tasks to clean off possible artifacts[2] and to reduce the risk of mislabeling.

[1] In the resting state, subjects have eyes opened and do not undertake any particular mental task.

[2] The subject informs the operator of the task he/she intends to do next by pronouncing aloud the name (i.e., “left” or “right”).

3 Signal processing

Raw signals are transformed by means of 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 over the six centro-parietal electrodes C3, Cz, C4, P3, Pz, and P4. With the twenty-six-channel EEG system the SL is computed globally by means of a spherical spline of order 2 (Perrin et al., 1989, 1990). 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 taking into account their distance to the electrode of interest (McFarland et al., 1997; Zhou, 1993). In particular, each centro-parietal electrode uses the following three neighbors: F3,Cz,P3 for C3; C3,Cz,Pz for P3; C3,C4,Pz for Cz; P3,Cz,P4 for Pz; F4,Cz,P4 for C4; C4,Cz,Pz for P4. The superiority of SL-transformed over raw potentials for the operation of BCI has been demonstrated in different studies (e.g., McFarland et al., 1997; Babiloni et al., 2000).

In the first experiment with a conventional EEG system, for each SL channel, the analyzed features are the power spectral density components (estimated with a *Welch periodogram*) of 2-second long epochs, each starting 1 s after the previous one. Epochs are divided into segments of 1 s, with a Hann window of the same length applied to each segment, and 50 % overlapping between the segments. This gives a frequency resolution of 1 Hz. Then, the power components are referred to the corresponding values of the baseline and transformed in dB. Finally, 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 138 features (6 channels times 23 components each) and there are around 40 samples for each task of every subject. In the second experiment with the portable EEG system, the length of the sequences and segments are 1 and 1/2 seconds, respectively. In this case, the frequency resolution is 2 Hz and EEG samples have 72 features (6 channels times 12 components each). There is a minimum of 200 samples for each task for every day.

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 (Varsta et al., 2000).

4 Feature selection

In the case of EEG-based classification of mental tasks, we know that irrelevant features may act as noise that hinder the recognition problem. *Feature selection* aims at finding those *relevant* components for which the performance of the learned classifier is the best. Thus, feature selection and induction—i.e., the process of learning the appropriate classifier—are closely related. Depending on how these two algorithms are related, we have three different approaches:

1. *embedded* methods: the induction and the feature selection algorithms are *indivisible*;
2. *filter* methods: the feature selection *precedes* the induction algorithm; and
3. *wrapper* methods: the feature selection algorithm *uses* the induction algorithm.

One example of the first category is *C4.5* (Quinlan, 1993). This algorithm builds a well-known classifier called *decision tree*. In such a tree, each node corresponds to an attribute and each outgoing arc to a node at the next level is associated to a possible value (or to a range of values) of that attribute. A leaf of the tree specifies the expected value of the class. In a good decision tree, each node should be associated to the most *informative* attribute among the attributes not yet considered in the path from the root. This can be achieved using the notion of *entropy* to compute the *information gain* due to the attribute we are considering. C4.5 can be considered an embedded approach to feature subset selection because it builds a tree in which only the most relevant features appear.

Filter methods like *Relief* (Kira and Rendell, 1992), selects the most relevant features according to some criterion and then the induction algorithm builds a classifier using only this set of features. The idea behind this algorithm is that a feature is relevant if its values are very different for two samples of different classes, and very similar for two samples of the same class. To rank a feature, Relief finds two nearest neighbors for every sample of the training set, one from the same class (*nearHit*) and one from the opposite class (*nearMiss*). At the end, the weight of every feature is a real value in the range [-1,1]. Relief updates the weights of the features (which are initialized to zero in the beginning) according to the intuitive idea that a feature is more relevant if it has different values for the sample and its nearMiss, whereas it is less relevant if it has quite different values for the sample and its nearHit. In the experiments reported below, we actually use a version of Relief that makes it more appropriate to cope with noisy samples (Kononenko, 1994). In this case, the algorithm uses *n* nearest Hit/Miss. In the experiments reported below, *n*=5.

The main disadvantage of filter approaches is that feature selection is completely independent from the induction algorithm, and the former cannot be guided by the classifier error rate. *Wrapper* methods (Kohavi and John, 1997) use this feedback from the classifier to make the selection. Feature selection can be viewed as a search in a state space, where each state represents a particular subset of features. Going from a state to its neighbors is done by using operators that delete or add features from/to the current set. Without other information, reasonable starting points to conduct the search are the empty set or the full set of features. In the former case, it is done a *forward selection*, whereas in the latter it is performed a *backward elimination.* In both cases it is necessary a search engine to find the neighboring node with the highest evaluation and a heuristic function to guide it. The search engine is the well-known *best first* (e.g., Russell and Norvig, 1995) that stops search when the last *j* expansions are unfruitful. In the experiments reported below, *j*=5.

For training and testing purposes, the induction algorithm is C4.5 and we have followed a *k*-fold cross-validation approach, with *k*=5. Hence, we have divided the available EEG samples for each subject into *k* subsets of equal size. Then, training is repeated *k* times, each leaving out one of the subsets that is used for testing the recognition rates of the system. Results reported below are the average of the *k* validations.

The main motivation for using decision trees is their simplicity as they are basically parameter-free. In other words, the result does not depend on the settings of parameters, as it is needed for other machine learning approaches. For instance, in a neural network it is required to properly tune the number of neurons, network architecture, learning rate, etc. This property allows making an objective comparison of relevant features for different subjects. In addition, decision trees (C4.5 in particular) have been successfully applied to a large variety of tasks, ranging from concept learning to pattern recognition based on numerical features. Indeed, in the case of EEG feature selection, a preliminary study found that C4.5 performed better or equal than other simple machine learning approaches such as Bayesian classifiers (Langley et al., 1992) and instance-based (Aha et al., 1991) that induce probabilistic classifiers.

Prezenger and Pfurtscheller (1999) have also explored the use of machine learning techniques for the selection of relevant features in a BCI. They determined the relevance of frequency components in two preselected electrodes, namely C3 and C4, by means of an algorithm called DSLVQ (Prezenger et al., 1994). This is a modified version of learning vector quantization, LVQ (Kohonen, 1997). The final selection of the most relevant features used by their classifiers was done manually.

5 Experimental results

5.1 Experiment 1

Table 1 reports the results obtained with the five subjects for the two-class recognition problem. It compares the generalization error and the number of features when the classifier is preceded by Relief or there is no feature selection. There is a significant improvement in performance for all five subjects, although, except for subject MJ, the number of selected features is less than 10% of the total (138 in this case). The average absolute improvement in the accuracy is 9.6%, which translates into a relative error reduction of 28.1%. Fig. 2 illustrates the combination of Relief and C4.5 for subject CL. Fig. 2 shows how the accuracy of the classifier varies according to the number of selected features, starting with the most relevant up to that with the lowest relevance weight. The classifier's error decreases and reaches its minimum with less than 10% of the total number of features. Then, the error increases as the classifier use additional features.

Table 2 gives results with the wrapper method for the same subjects and tasks shown in Table 1. It compares the generalization error, number of selected features and number of expanded nodes with forward and backward search directions. The number of expanded nodes indicates the computational cost of the method. Backward elimination is significantly better than forward selection for four out of five subjects. For subject TA is not significant. However, backward elimination is more expensive computationally (36% more expanded nodes). In both cases, forward and backward, the set of selected features is again less than 10% of the total number for all subjects. Compared to classifiers that work with all the features, the wrapper method with backward elimination induces classifiers whose average absolute improvement is 13.6%, which translates into an average relative error reduction of 39.8%. As expected, classifiers induced with wrapper methods for feature selection achieve better results than with filter methods (Table 1). Given that both kind of methods select similar number of features—except for subject MJ, for whom wrappers select 3 times less features—, it can concluded that wrapper algorithms find more relevant features. In Section 6 we will analyze these features.

As wrapper methods perform better than filter methods, whereas the latter are faster, we have combined them in a *mixed method* that combines the respective advantages of the original methods. The new method applies first Relief and then the wrapper algorithm starts backward elimination from the node with the features provided by Relief, namely those with a weight greater than 0.4. Table 3 shows the results achieved with this mixed approach, which confirm its validity. Indeed, its performance is similar to that of the wrapper/backward approach alone but expands significantly fewer nodes (37.6% less expanded nodes). It also selects similar features. In this case, the average absolute and relative improvements with respect to no feature selection are 13.2% and 38.6%, respectively.

5.2 Experiment 2

Unlike the previous experiment, in this experiment the subject was trained in the presence of feedback. 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. The first day, the subject does not receive any feedback. Classifiers are learned and used as follows. With the EEG data recorded a day, we induce off-line the current classifier and evaluate its generalization performance. This classifier is then embedded in the brain interface that is operated the following day.

Table 4 shows the performance of the sixth subject over five consecutive days of training. It compares classifiers induced without feature selection and with the mixed method described above. For three out of five days—i.e., days 2, 4 and 5—the mixed method induces classifiers that perform significantly better than those obtained without feature selection. In day 1 both classifiers achieve similar performances while in day 3 the classifier that work with all the features perform significantly better. Interestingly, in this third day the performance of the classifiers degrade. This negative trend continues on the fourth day, to recover on the fifth. At the end of training, the mixed method yields absolute and relative improvements of 6.9% and 24.0%, respectively. The number of selected features is less than $1/3$ of the total.

Results of Table 4 raise a number of questions. Firstly, why performance does not improve linearly with training? As described in (Millán et al., 2000), the approach is based on a mutual learning process where the user and the BCI are coupled together and adapt to each other. This helps to achieve good performance rapidly (second day), but it may also make the subject lag behind the classifier. Thus, it is quite possible that in the third day the subject is using mental strategies that previously worked well and in the meantime the classifier has changed. Secondly, why is the number of relevant features higher in the second experiment than in the first? A possible explanation is the poorer estimation of the power spectrum; as explained in Section 3, the frequency resolution is 1 and 2 Hz in the first and second experiments, respectively. Indeed, comparing Fig. 3 and Fig. 4 one can see that in the second experiment it is common to have consecutive relevant features, while this is not the case in the first.

6 Discussion

Fig. 3 shows the relevant features selected for the different subjects in the first experiment using the wrapper method with backward elimination. This figure demonstrates empirically that the set of relevant features is individual for each subject. Indeed there exists no single feature shared by all the subjects. Moreover, we could have hardly selected these sets of features manually. Previous neurophysiological findings (e.g., Pfurtscheller and Neuper, 1997) seem to indicate that, for imagination of hand movements, the most relevant features are contralateral event-related desynchronizations (ERD) in the alpha band over the motor cortical areas (channels C3 and C4). It turns out that no feature in the alpha band (8–12 Hz) of channel C4 is selected for any of the subjects, and only a few features in the alpha band of channel C3 are chosen for three of the subjects. This does not mean that there is no ERD in the alpha band of channel C4. Averaged task-related alpha power maps shows that this ERD is present for both motor imagery tasks (although is higher for the left imagined movement) and so is not relevant for discrimination. On the other hand, Wolpaw and coworkers (e.g., McFarland et al., 1997) as well as Pfurtscheller's team (e.g., Pregenzer and

Pfurtscheller, 1999) have found that electrodes overlaying the hand representation (e.g., C3 and C4) are most appropriate for recognition of left/right motor imagery. This study indicates, and that is surprising, that features from midline electrodes Cz and Pz are also relevant. The involvement of parietal areas in motor tasks was reported by Gerloff et al.'s (1998) and Babiloni et al.'s (1999).

Why such a partial discrepancy with respect to previous neurophysiological findings? We put forward that the reason is the use of a different experimental framework. While the above-mentioned findings have been obtained in a classical event-related paradigm where the subject is synchronized with a cue stimulus that indicates when, what and the pace of the task to perform, ours is closer to real life as subjects are not passive but make decisions spontaneously and self-paced. This is especially true for experiment 2. On the other hand, although in experiment 1 an operator gives the instruction, the subject still imagines internally paced finger extensions.

Fig. 4 shows the relevant features selected as the sixth subject is trained over five consecutive days using the wrapper method with backward elimination. One can notice, at least, three patterns in Fig. 4. First, for every day, there are subsets of consecutive relevant features. As discussed before, this may be due to the low frequency resolution. Second, as in the previous experiments, only a small fraction of relevant features are located in the alpha band of the channels. Third and most importantly, for the days 2 and 5 in which performance is better, it can be appreciated a regular distribution of features mostly belonging to the central channels. In particular, no feature of the parietal channel P4 is selected any day. In addition, in these two days there is a significant subset of common features (see Table 5) that accounts for the 65% and 48% of the total in the second and fifth day, respectively. Moreover, quite a few of the remaining features are consecutive to the common ones—6% and 26%, respectively—what together represent more than 70% in both days.

To conclude, this study empirically demonstrates the benefits of heuristic feature selection methods for EEG-based classification of mental tasks. In particular, it is shown that the classifier performance improves for all the considered subjects with only a small proportion of features. Thus, the use of just those relevant features increases the efficiency of the brain interfaces and, most importantly, enables a greater level of adaptation of the personal BCI to the individual user.

Acknowledgements

This work has been supported by the European ESPRIT programme, LTR project number 28193—ABI.

References

Aha DW, Kibler D, Albert MK (1991). Instance-based learning algorithms. Machine Learning 6:37–66.

Babiloni C, Carducci F, Cincotti F, Rossini PM, Neuper C, Pfurtscheller G, Babiloni F (1999). Human movement-related potentials vs desynchronization of EEG alpha rhythm: A high-resolution EEG study. NeuroImage 10:658–665.

Babiloni F, Cincotti F, Lazzarini L, Millán J del R, Mouriño J, Varsta M, Heikkonen J, Bianchi L, Marciani MG (2000). Linear classification of low-resolution EEG patterns produced by imagined hand movements. IEEE Trans. on Reh. Eng. 8:186–188.

Birbaumer N, Ghanayim N, Hinterberger T, Iversen I, Kotchoubey B, Kübler A, Perelmouter J, Taub E, Flor H (1999). A spelling device for the paralysed. Nature 398:297–298.

Gerloff C, Richard J, Hadley J, Schulman AE, Honda M, Hallett M (1998). Functional coupling and regional activation of human cortical motor areas during simple, internally paced and externally paced finger movements. Brain 121:1513–1531.

Kalcher J, Flotzinger D, Neuper C, Gölly S, Pfurtscheller G (1996). Graz brain-computer interface II. Med. & Biol. Eng. & Comput. 34:382–388.

Kira K, Rendell LA (1992). The feature selection problem: Traditional methods and a new algorithm. Ninth National Conf. on Artificial Intelligence, 129–134.

Kohavi R, John G (1997). Wrappers for feature subset selection. Artificial Intelligence 97:273–324.

Kohonen T (1997). Self-organizing maps, 2nd ed. Springer-Verlag, Berlin.

Kononenko I (1994). Estimating attributes: Analysis and extensions of RELIEF. European Conf. on Machine Learning, 171–182.

Langley P, Iba W, Thompson K (1992). An analysis of Bayesian classifiers. Tenth Natl. Conf. on Artificial Intelligence, 223–228.

McFarland DJ, McCane LM, David SV, Wolpaw JR (1997). Spatial filter selection for EEG-based communication. Electroenceph. Clin. Neurophysiol. 103:386–394.

Millán J del R, Mouriño J, Babiloni F, Cincotti F, Varsta M, Heikkonen J (2000). Local neural classifier for

EEG-based recognition of mental tasks. Int. Joint Conf. on Neural Networks, vol. III, 632–636.

Müller-Gerking J, Pfurtscheller G, Flyvbjerg H (1999). Designing optimal spatial filters for single-trial EEG classification in a movement task. Clin. Neurophysiol. 110:787–798.

Perrin F, Pernier J, Bertrand O, Echallier JF (1989). Spherical spline for potential and current density mapping. Electroenceph. Clin. Neurophysiol. 72:184–187.

Perrin F, Pernier J, Bertrand O, Echallier JF (1990). Corrigendum EEG 02274. Electroenceph. Clin. Neurophysiol. 76:565.

Pfurtscheller G, Neuper C (1997). Motor imagery activates primary sensorimotor area in humans. Neurosci. Lett. 239:65–68.

Prezenger M, Pfurtscheller G (1999). Frequency component selection for an EEG-based brain to computer interface. IEEE Trans. on Reh. Eng. 7:413–419.

Prezenger M, Flotzinger D, Pfurtscheller G (1994). Distinctive sensitive learning vector quantization — A noise insensitive classification method. Int. Conf. on Neural Networks, 2890–2894.

Quinlan JR (1993). C4.5: Programs for machine learning. Morgan Kaufmann, San Francisco, CA.

Russell S, Norvig P (1995). Artificial intelligence: A modern approach. Prentice Hall, Upper Saddle River, NJ.

Varsta M, Heikkonen J, Millán J del R, Mouriño J (2000). Evaluating the performance of three feature sets for brain-computer interfaces with an early stopping MLP committee. 15th Int. Conf. on Pattern Recognition.

Wolpaw JR, McFarland DJ (1994). Multichannel EEG-based brain-computer communication. Electroenceph. Clin. Neurophysiol. 90:444–449.

Zhou PB (1993). Numerical analysis of electromagnetic fields. Springer-Verlag, Berlin.

TABLES

Table 1. Generalization errors of classifiers for right-hand and left-hand movement imagination tasks induced by C4.5 without feature selection (FS) and with features selected by Relief.

Subject	**Generalization error**		**# Features used**	
	Without FS	Relief	Without FS	Relief
CL	0.322	0.215	138	6
MJ	0.329	0.234	138	30
RA	0.421	0.314	138	11
RB	0.338	0.232	138	2
TA	0.300	0.235	138	9
average	*0.342*	*0.246*	*138*	*11.6*

Table 2. Generalization errors of classifiers induced by C4.5 and using a wrapper method with best-first search and two different search directions, either forward (FW) or backward (BW).

Subject	**Generalization error**		**# Features used**		**# Nodes expanded**	
	FW	BW	FW	BW	FW	BW
CL	0.238	0.190	7	7	1630	3127
MJ	0.241	0.193	8	10	2168	3403
RA	0.267	0.229	5	7	1358	1903
RB	0.239	0.215	8	9	1770	2577
TA	0.208	0.202	4	6	1218	1770
average	*0.239*	*0.206*	*6.4*	*7.8*	*1629*	*2556*

Table 3. Generalization errors of classifiers induced by C4.5 without feature selection (FS) and with a mixed feature selection algorithm that combines Relief and wrapper methods.

Subject	**Generalization error**		**# Features used**		**# Nodes expanded**
	Without FS	Mixed	Without FS	Mixed	
CL	0.322	0.172	138	7	2063
MJ	0.329	0.223	138	7	2058
RA	0.421	0.256	138	12	1384
RB	0.338	0.232	138	6	1509
TA	0.300	0.168	138	4	957
average	*0.342*	*0.210*	*138*	*7.2*	*1594*

Table 4. Generalization errors of classifiers induced by C4.5 without feature selection (FS) and with a mixed feature selection as the same subject is trained over five consecutive days.

Day	**Generalization error**		**# Features used**	
	Without FS	Mixed	Without FS	Mixed
1	0.429	0.424	72	21
2	0.248	0.206	72	17
3	0.305	0.324	72	22
4	0.404	0.370	72	31
5	0.288	0.219	72	23

Table 5. Common relevant features for the days 2 and 5 in which the performance is better.

Channel	**Hz**
C3	10, 18
Cz	18, 26
Pz	12
C4	16, 18, 20, 24, 28, 30

FIGURES

Fig. 1.

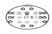


Fig. 2.

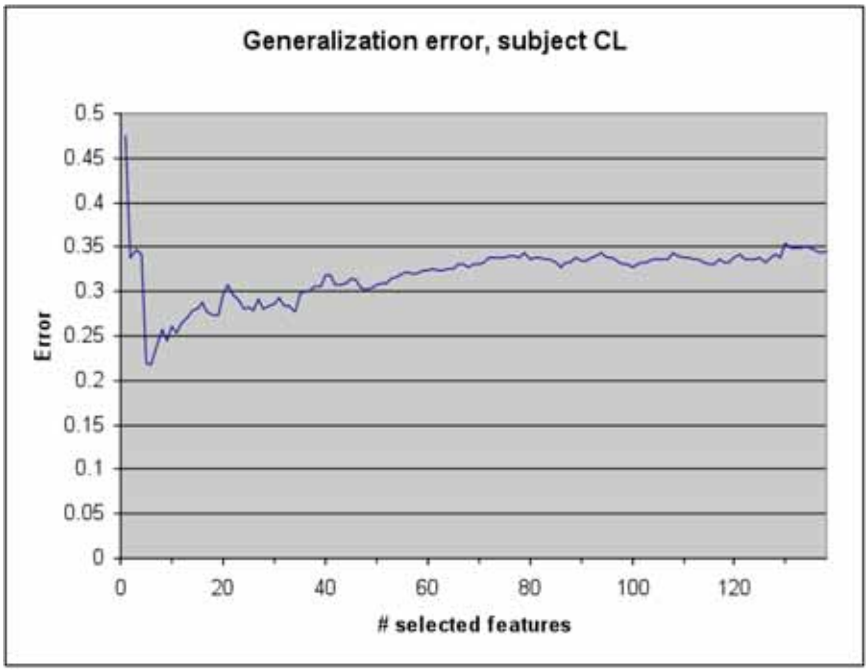

Generalization error, subject CL
Error
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
20
40
60
80
100
120
selected features

Fig. 3.

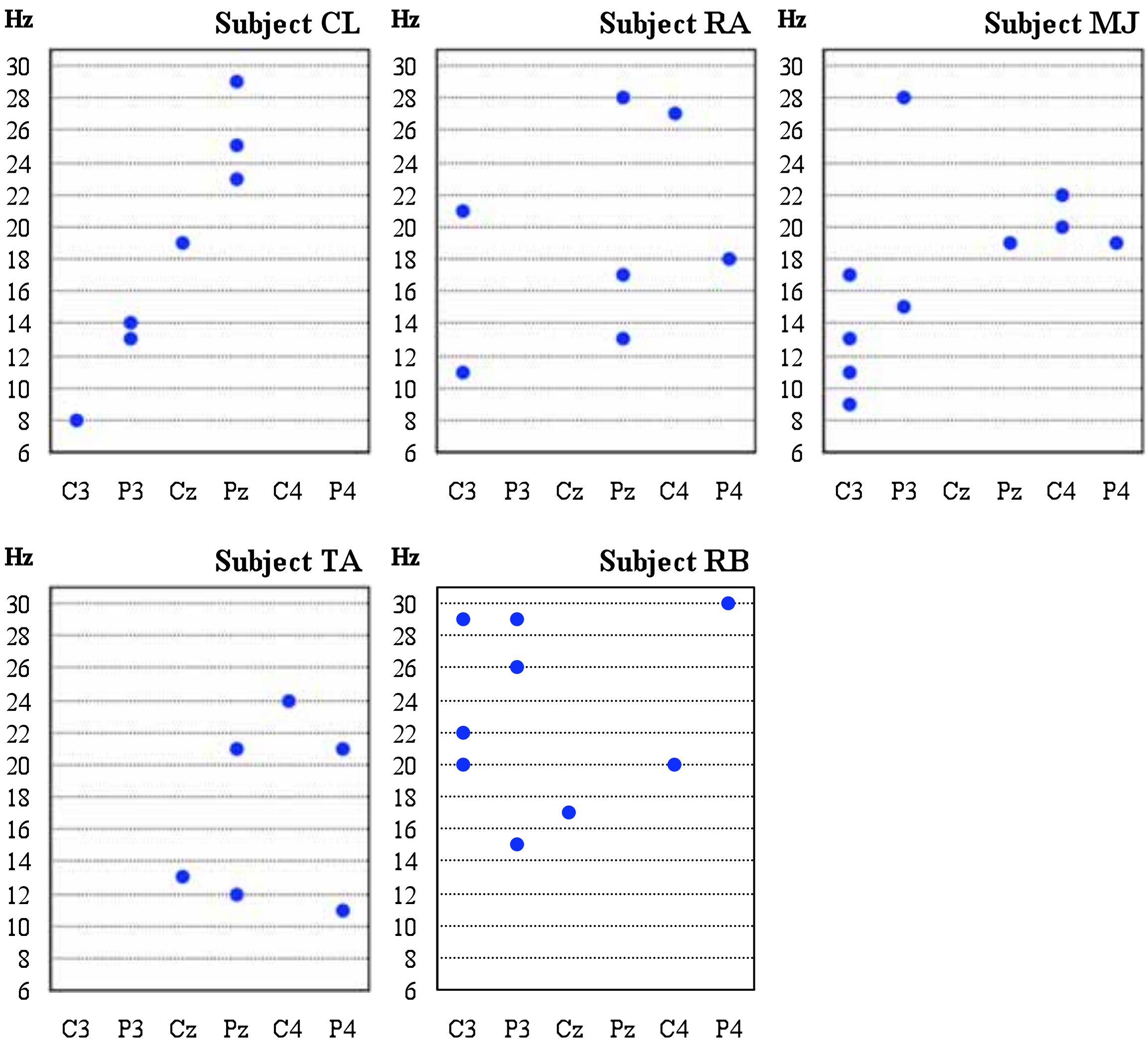
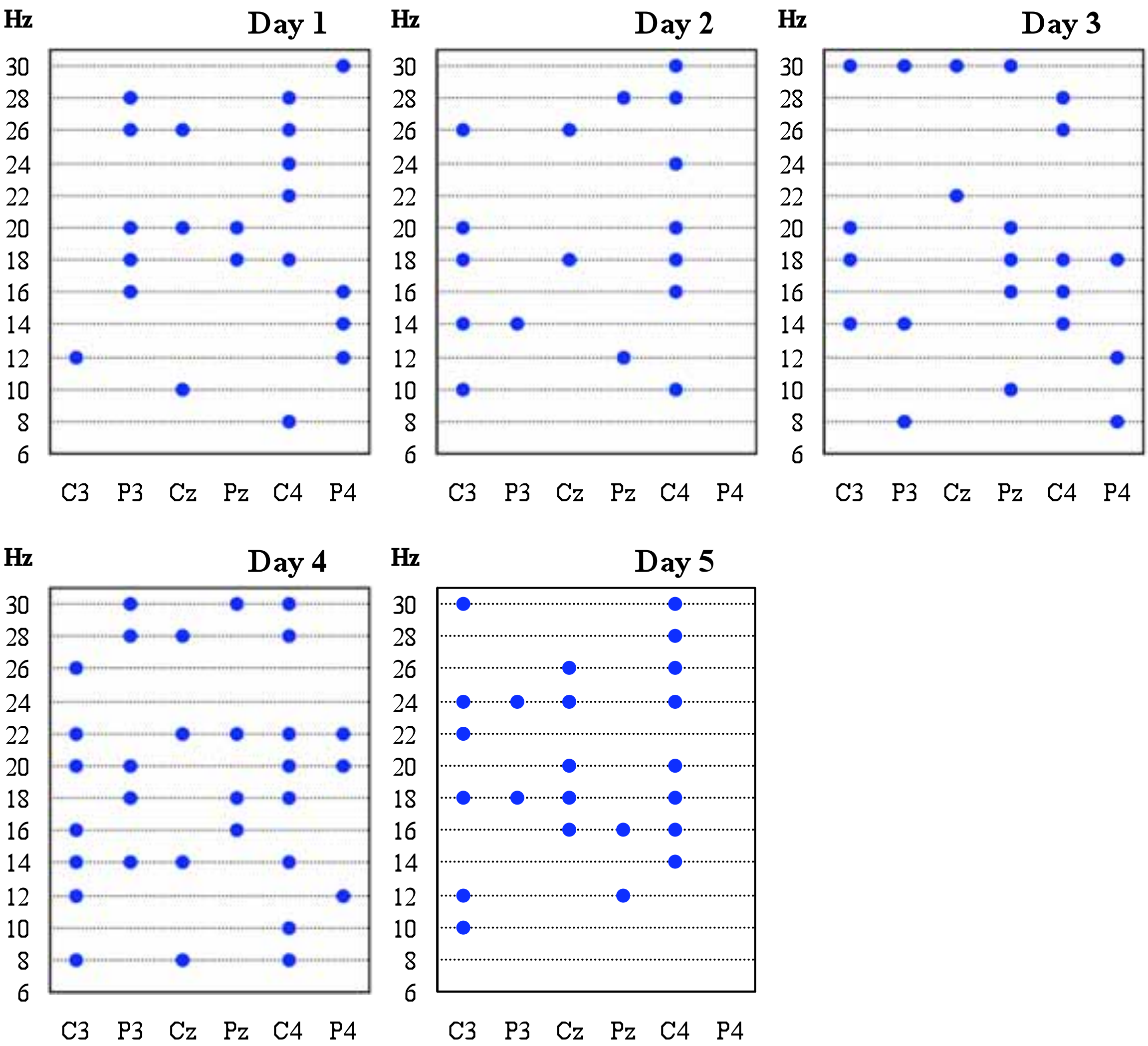

Hz
Subject CL
Hz
Subject RA
Hz
Subject MJ
30
28
26
24
22
20
18
16
14
12
10
8
6
C3 P3 Cz Pz C4 P4
Hz
Subject TA
Hz
Subject RB

Fig. 4.



Hz
Day 1
30
28
26
24
22
20
18
16
14
12
10
8
6
C3 P3 Cz Pz C4 P4
Hz
Day 2
30
28
26
24
22
20
18
16
14
12
10
8
6
C3 P3 Cz Pz C4 P4
Hz
Day 3
30
28
26
24
22
20
18
16
14
12
10
8
6
C3 P3 Cz Pz C4 P4
Hz
Day 4
30
28
26
24
22
20
18
16
14
12
10
8
6
C3 P3 Cz Pz C4 P4
Hz
Day 5
30
28
26
24
22
20
18
16
14
12
10
8
6
C3 P3 Cz Pz C4 P4

LEGENDS

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 8 indicated in gray.

Fig. 2. Generalization error as the number of selected features increases for subject CL. The features are sorted by a filter method (Relief). The induction algorithm used for the classification is a decision tree generator (C4.5).

Fig. 3. Relevant EEG features selected for subjects CL, MJ, RA, RB and TA using a wrapper method with backward elimination.

Fig. 4. Evolution of relevant EEG features over five consecutive days as a new subject masters the BCI. Features are selected using a mixed method that applies Relief and then a wrapper algorithm with backward elimination.