Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Recent Studies in the Design of a Self-paced Brain Interface with Low False Positive Rate

Mehrdad Fatourechi, *Student Member*, *IEEE*, Rabab K. Ward, *Fellow, IEEE,* and Gary E. Birch, *Member*, *IEEE*

***Abstract*—The findings of two recent studies that aim at developing a self-paced brain interface (BI) system with low false positive rates are discussed. The first study examines the use of information extracted from different neurological phenomena and the second study examines the wavelet coefficients extracted from a single neurological phenomenon. The analysis of the data of two subjects shows that both are successful at yielding low false positive rates. These studies also show that for each subject, a unique set of features and EEG channels lead to superior performance.**

I. Introduction

A self-paced Brain Interface (BI) system aims at assisting individuals with severe motor disabilities. It allows them to control their environment by thinking of it only and at their own pace [1]. Such a system is different from a synchronized BI, where a user can initiate a command only during certain periods specified by the system.

In a self-paced BI system, the state in which the user is intentionally attempting to control a BI system is called an Intentional Control (IC) state. At other times, the user is said to be in a no-control (NC) state, where he/she may be performing some action other than trying to control the BI system. To operate in this paradigm, self-paced BI systems are designed to respond only when the user is in an IC state and to remain inactive when the user is in an NC state.

The performance of a self-paced BI system is usually evaluated through two metrics: A true positive (TP) rate and a false positive (FP) rate. A FP rate is the percentage of misclassifying a NC trial as an IC trial, and a TP rate is the percentage of correctly classifying an IC trial. The FP rates of current self-paced BI systems are still very high for practical applications, due to the very noisy nature of the brain's electrical signals. This makes the correct detection of patterns associated with control commands difficult. Nevertheless, it is crucial to keep the FP rate as low as possible in order to prevent user frustration. We expect that this goal can be achieved by exploring various types of information (neurological, spatial, temporal, and frequency-related) embedded in the features.

In this paper, we report on two recent studies carried out with the aim of developing a self-paced BI system with low FP rates and discuss their findings. In the first study, the information from different neurological phenomena is used in the design a self-paced BI system. The second study solely focuses on the MRPs.

Both systems were evaluated offline using the data collected from two able-bodied subjects. Both studies led to the design of a self-paced BI system with low FP rates. In the first study, this was achieved by using the neurological/spatiotemporal information of (movement related potentials (MRPs) and changes in the power of brain rhythms such as Mu or Beta) . In the second study; this was achieved by using the spatiotemporal/frequency information of wavelet features.

Manuscript received April 22, 2006. This work was supported in part by the NSERC under Grant 90278-06 and CIHR under Grant MOP-72711. This publication was carried using the WestGrid computing resources.

M. Fatourechi is with the Image and Signal Processing Lab, University of British Columbia, Vancouver, BC Canada (phone: 604-822-4988; fax: 604-822-9013; e-mail: mehrdadf@ece.ubc.ca).

R. K. Ward is the director of the Institute for Computing, Information and Cognitive Systems (ICICS) , Vancouver, BC Canada and a Professor in the University of British Columbia, Vancouver, BC Canada (e-mail: rababw@ece.ubc.ca).

G. E. Birch is the executive director of the Neil squire Society. He is also an adjunct Professor in the University of British Columbia , Vancouver, BC Canada (e-mail: garyb@neilsquire.ca).

II. Data Collection

The data of two right-handed, able-bodied male subjects (coded as *AB1* and *AB4*) were used in this study. Both had signed consent forms prior to participation in the experiment.

The EEG signals were recorded from 13 monopolar electrodes positioned over the subjects' supplementary motor area and primary motor cortex (according to the International 10-20 System at F_1, F_z, F_2, FC_3, FC_1, FC_z, FC_2, FC_4,C_3 , C_1, C_z,C_2 and C_4 locations). The data were collected from the subjects as they performed the flexion of the right index finger (see [2] for more details).All signals were sampled at 128 Hz and referenced to the ear electrodes. For each subject, the data collected on a period of 5 days (an average of 80 trials per day) was used.

An IC trial consisted of data collected from $-t_a$ second before to t_b seconds after the movement onset (measured as the finger switch activation). The trials for which the EOG activity exceeded a pre-defined threshold (±25 μV) were automatically rejected. The NC trials were selected as follows: a window of width (t_a+t_b) seconds was considered. The window was shifted over the EEG signals (collected during the No Control sessions) by a step of 16 samples (0.1250 sec). For each window where artifacts were not detected, features were extracted.

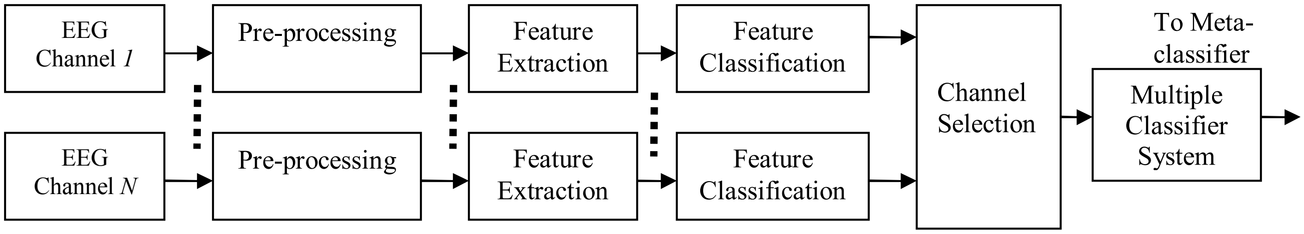


Fig.1. the proposed scheme for classifying each of the neurological phenomena in Study I.

III. STUDY I- A SELF-PACED BI SYSTEM BASED ON MULTIPLE NEUROLOGICAL PHENOMENA

The first study examines the use of multiple neurological phenomena in the development of a self-paced BI system.

It is known that an internally paced movement results in the generation of three different responses in the EEG signal: a movement-related potential (MRP) [3], an event related desynchronization (ERD) and an event-related synchronization (ERS) [4].

Averaging the EEG data with respect to the movement onset results in slow potentials called movement related potentials (MRPs). MRPs start about 1.5–1 s before the movement and have bilateral distribution [3].

A well-known neurological observation is that when a subject is not performing a movement task, large populations of neurons in the respective cortex fire in rhythmical synchrony. These rhythms, such as the Mu and Beta rhythms, are called the idling rhythms. Voluntary movement results in a circumscribed desynchronization in the Mu and Beta bands, localized close to the sensorimotor areas [4]. Since this attenuation is due to loss of synchrony in neural populations, it is termed ERD. The enhanced rhythmic activity after the movement is called ERS.

A number of studies provide some evidence that MRPs, and the changes in the power of the Mu rhythms (CPMR) and the Beta rhythms (CPBR) (usually characterized in the literature as ERD/ERS) provide complementary information in exploring the cognitive functions of the brain [5, 6]. There is also some evidence on the differences between the Mu and Beta rhythms. For example, it is shown that the reactivities of the Mu and Beta rhythms related to the movement onset are different [7].

Although most BI researchers use a single neurological phenomenon as the source of control, the use of multiple neurological phenomena has been reported [8, 9]. Some studies considered the simultaneous application of MRPs and the changes in the power of brain rhythms in the design of synchronized BI systems [8, 10, 11]. However, such a study has not been carried out for a self-paced BI system. We have carried out such study with the aim of designing a self-paced system with low FP rates.

Different methods have been proposed for handling a high-dimensional feature space (carrying temporal, spatial, frequency-related and perhaps neurological information). These methods include extracting only a few features from 1 or 2 EEG channels resulting in a low-dimension feature space [12, 13], feature weighting [9] and filtering irrelevant or less useful EEG channels [14]. Since we intended to keep as much spatiotemporal information as possible and at the same time, avoid the computational problems associated with dealing with high-dimension data, we divide the feature space to feature spaces with smaller sizes. This is done by using the theory of multiple classifier systems (MCSs). First, a separate classifier was designed for each of the neurological phenomenon (see Fig.1) . Then a Meta classifier was used to combine the outputs of these classifiers and form the final classifier (See Fig. 2).

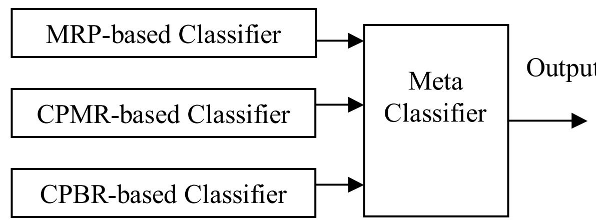


Fig.2. A MCS (Meta classifier) combines the outputs of the individually designed MCS classifiers in Study I.

We used cross covariance (matched filtering) to extract the features. Matched filtering measures the similarities between two sequences and outputs a single feature per trial. Each single-trial sequence was matched with three template signals (see below).

The process of creating the MRP templates was as follows : First, for each channel, the EEG signal was filtered to the frequencies below 4Hz [1] . The IC trials were selected from $-t_a$ =-1 s. to t_b =2 s. with respect to the finger switch onset so as to account for MRPs and changes in the power of brain rhythms. The IC trials in the training session were then averaged to create the MRP template for that particular channel.

If the changes in the brain rhythms, characterized as the ERD and ERS (defined as in [4]) are used as the sources of control , the choice of a reference interval will be problematic. A reference interval should be defined some seconds before the occurrence of the movement. The ERD is then defined as the percentage of power decrease and ERS is defined as the percentage of power increase with respect to this reference interval [15]. However, in [16] it was shown that such approach may yield misleading results depending on the value of the reference signal's power . Thus, instead of using the ERD or ERS templates as defined in [4] , we focused on the time course of power as in [16] .

The Mu and Beta power templates were created as follows: First, for every EEG channel, all the trials were bandpass filtered. The band pass for the Mu rhythms was chosen from 8 to 12Hz and for the Beta rhythms, it was chosen between 18 to 26Hz. The amplitude signals were then squared in order to estimate the power values. The resulting power signals were then low-pass filtered below 4 Hz to ensure that smooth shapes were obtained for the power templates. The IC trials in the training sessions were then averaged to create the templates for the power of Mu and Beta rhythms.

After calculating the cross covariance between a trial X_n and a template Y_n ($C_{xy}(n)$), its maximum over a period of 0.125 s. was extracted as the feature:

$$F_i = \max(C_{xy}(n)), n \in [t_a + t_b - 0.0625, t_a + t_b + 0.0625] \quad (1)$$

We used K-Nearest neighbor (KNN) classifier with the LVQ3 learning algorithm for classification. Prior to applying the LVQ3 learning algorithm, K-means was used for forming the initial clusters. For each channel and for each neurological phenomenon, a classifier was trained based on the values of a single feature.

We then used the multiple classifier systems (MCS) framework to combine the outputs of classifiers. To create independent (diverse) classifiers, we used both the spatial information of channels and the presence of different neurological phenomena.

Each neurological phenomenon was first classified separately. For each neurological phenomenon, a K-NN classifier was designed for each EEG channel and then the individual classifiers' outputs were combined in a MCS framework to generate the output corresponding to that particular neurological phenomenon (Fig.1). For each neurological phenomena, a genetic algorithm (GA) was applied to select the best combination of channels. Each chromosome in the GA consists of 1's and 0's with a 1 (or 0) indicating the presence (or absence) of a particular feature. The operators of the GA were tournament-based selection (tournament size =3) , uniform crossover, uniform mutation , the size of the initial population=100, the size of the population in the next generations=50, random initialization for initializing the GA, Elitism of the best performing chromosome and the number of evaluations=1500.

The selected features then participated in a majority vote scheme. A lexicographic approach [17] was used for ranking the chromosomes in the population. The objectives were ranked according to their priorities before optimization. The objective function with the highest priority was defined as follows:

$$f(x) = \begin{cases} 0, & TP < 20\% \\ \dfrac{TP(x)}{FP(x)}, & otherwise \end{cases} \quad (2)$$

In equation 2 , x is a chromosome and f is the objective function . This objective function gave a higher priority to chromosomes which had a higher average of TP rate and lower average of FP rate on the validation sets . We also postulated that TP rates below 20% would be too low for the successful operation of a self-paced BI system (even though their FP rates might be very low at the same time), so they were considered "unfit". The other objectives were chosen as 1) the average of FP rate over the validation sets", and 2) the number of features, respectively.

Once the individual MCSs were designed for each neurological phenomenon, their outputs were combined using a meta-classifier. The majority vote scheme was then used for combining the outputs of the classifiers (see Fig.2).

IV. STUDY II- THE DESIGN OF A SELF-PACED BI SYSTEM USING MRP WAVELET COEFFICIENTS

This study examines the application of the discrete wavelet transform (DWT) in the design of a self-paced BI system. The DWT can be used as a powerful feature extraction tool that can extract the time-frequency features that are similar to the shape of a particular wavelet function. As a result, DWT has an advantage over other feature extraction methods which solely operate in one domain such as the Fourier transform or AR modeling.

In the design of BI systems, DWT has been successfully applied to extract time and frequency information from the brain signals [9, 14, 18]. In this study, we explored the application of DWT to extract MRP features for driving a self-paced BI system.

Although the above studies provide promising evidence that the wavelet transform can be employed to extract features in BI systems, they used only one or two EEG channels. Thus, the high dimension feature space did not pose a serious problem. Using only one or two electrodes seems appealing, since its setup is fast and uses less hardware/software infrastructure. However, in most of the above-mentioned papers, relatively high amount of classification error was achieved. The above observations strongly motivate the use of more EEG electrodes. Using signals recorded from multiple channels allows us to explore the spatial information, which is expected to yield improvements in the classification accuracy.

The same recording protocol as the one described in Section II was also used in this study. Initial results of the analysis of the wavelet features showed high FP rates in the analysis of monopolar channels (FP>4%). In order to find a solution which yields low FP rates, the recorded signals were converted to bipolar EEG signals by calculating the difference between the adjacent EEG channels. The previous evidence has shown that bipolar signals "may" result in the generation of more discriminant features than those obtained from monopolar electrodes, [1]. This conversion resulted in the generation of the following 18 bipolar EEG channels: F_1-FC_1, F_1-F_z , F_2-F_z, F_2-FC_2 , FC_3-FC_1, FC_3-C_3, FC_1-FC_z, FC_1-C_1, FC_z-FC_2, C_1-C_z, C_2-C_4 , FC_2-FC_4 , FC_4-C_4 , FC_2-C_2 , FC_z-C_z , C_3-C_1 , C_z-C_2 and F_z-FC_z .

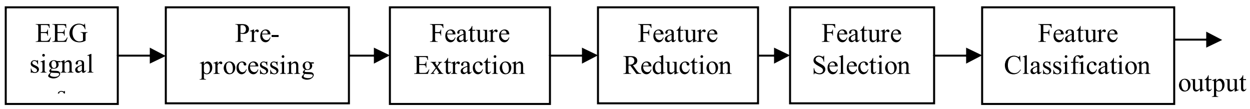


Fig.3. the overall scheme proposed for Study II.

The overall structure of the proposed scheme is shown in Fig.3. The IC trials were selected from $-t_a$ =-1.5 s. to t_b =1 s. with respect to the finger switch onset, as this range covers the start and finish of the MRP . The *rbio3.3* wavelet was chosen as the wavelet function because it has some similarities with the shape of the classic bipolar MRP pattern. Using a 5-level decomposition method resulted in wavelet coefficients corresponding to the following frequency bands (the sampling frequency is 128 Hz): [32-64], [16-32], [8-16], [4-8], [2-4], and [0-2] Hz.

Since an MRP is a low frequency neurological phenomenon [1], only the lowest frequency bands (i.e., 0-2Hz and 2-4Hz) were considered for further analysis. Even with this reduced feature space, the resultant feature space dimension, which was the product of the number of electrodes and the number of wavelet features per EEG signal, remained very high.

A hybrid feature selection algorithm was used to handle the dimensionality of the feature space : Mutual Information (MI) was employed first to filter out the less informative features and a GA was then used for selecting the optimal set of features.

For each subject, the wavelet coefficients (features) values corresponding to all the training set data were calculated. After calculating the values of MI for all $N_{features}$ features, they were ranked in descending order according to their MI value. The top P features were then selected. We arbitrarily chose $P=50$, to avoid having a feature space with a very high dimension. After the reduction of the dimension of the feature space, a GA was used for selecting a subset of m features from the top P features.

The fitness function and the GA parameters were identical to those described in Study I. Only the number of evaluations was set to 2000. A support vector machine then classified each chromosome in the GA population. We used the LIBSVM software [19].

V. RESULTS

Both IC and NC datasets were randomized and divided in training, validation and test sets. The training set was used for training the classifier, the validation set was used to select the best model for the system. The configuration which yielded the best average performance on the validation set, was selected and the performance of the system once applied to the test set was reported. We used a 5-fold nested cross validation for evaluating the performance of the system. For each outer cross- validation set, 20% of the data was used for testing and the rest was used for training and model validation. In order to select the models, the datasets are further divided into 5-fold. For each fold, 80% of the data was used for training and 20% was used for model validation.

TABLE.1 TP AND FP RATES OF BOTH STUDIES

Subj	MeanTP(%)-Study I	MeanFP(%) -Study I	MeanTP(%) -Study II	MeanFP(%) -Study II
AB1	26.00 (9.49)	0.13 (0.09)	66.96 (4.79)	0.99 (0.39)
AB4	31.88 (8.96)	1.15 (0.31)	56.10 (4.90)	1.41 (0.75)

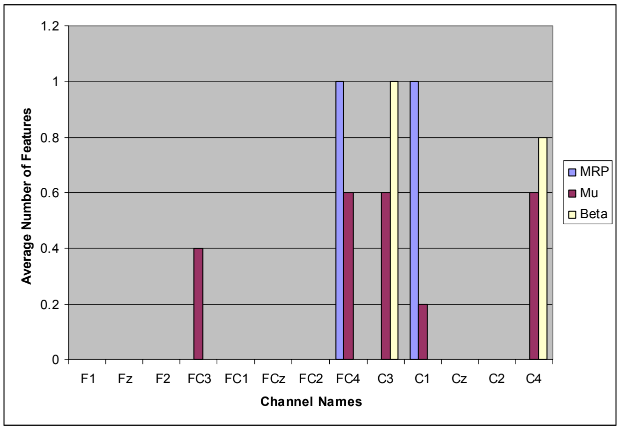
In Study I, the value of K in the *K-NN* classifiers, the number of codebooks and the parameters' values of the LVQ3 learning algorithm were all determined during model selection and through parallel runs of the code.

The results of applying the proposed method are shown in the 2nd and 3rd columns in Table 1 . Since a 5-fold nested cross-validation was used for the performance evaluation, the results were averaged over 5 runs of the outer validation sets. The numbers in parentheses are the standard deviations. As seen in this Table, low FP rates for both subjects were achieved for a reasonable TP rate. Fig.4 shows the average number of selected features per channel for each subject in Study I.

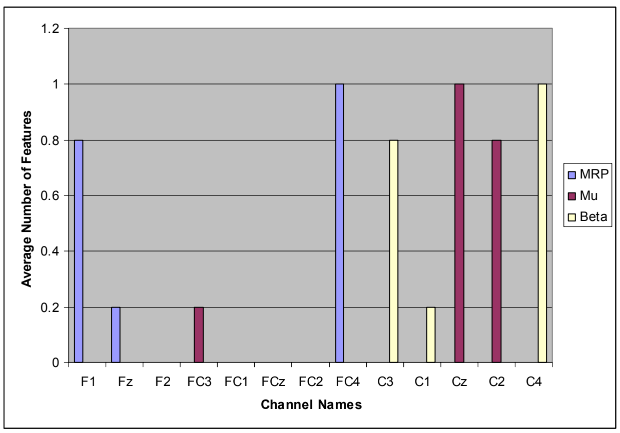
In Study II, we also performed a search on the classifier's parameters during the model validation. We found that a 5th degree polynomial Kernel function resulted in a superior performance compared to other Kernel functions studied in this paper (linear , polynomial and RBF kernel).The 4th and 5th columns in Table 1 show the results achieved on the test sets. Fig.5 shows the average number of selected features per channel for both subjects after applying the hybrid selection method in Study II.

VI. DISCUSSION AND CONCLUSIONS

In this paper, we presented the results of two separate studies on the design of self-paced BI systems with low FP rates. In the first study, we explored the application of using the spatiotemporal information of different neurological phenomena in order to design a new self-paced BI system. In the second study, the application of the wavelet transform in the design of a self-paced BI system was studied. The above results show that both systems had low FP rates. The results led to interesting observations.



(a)



(b)

Fig.4. The average number of selected features in Study I.
(a) Subject AB1, (b) Subject AB4.

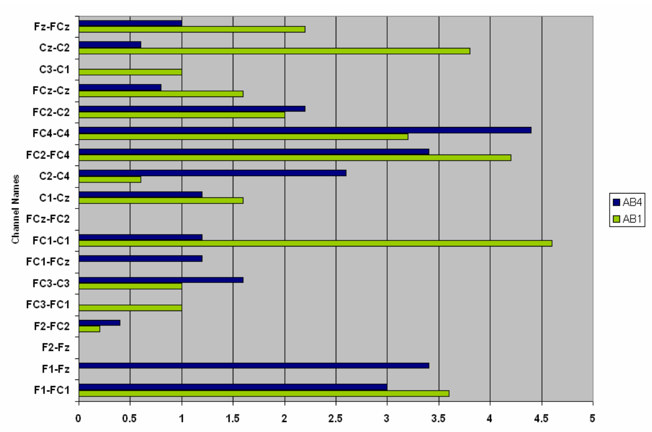


Fig.5. the average number of selected features in Study II.

1) Both studies show that it is not necessary that all channels are used for the successful operation of the system. Having fewer channels, in turn, leads to the speed-up of the setup of any self-paced BI system.

From Fig. 4 and Fig. 5, it is seen that in Study I, for Subject AB1, channels C1, C3, C4 and FC4 contributed more to the final classification scheme . Similarly, for Subject AB4, channels F1, FC4, C3, Cz, C2 and C4 contributed more to the final classification scheme. In Study II, for subject AB1, more features were selected from channels FC_1-C_1, F_1-FC_1, F_z-FC_z, FC_4-C_4, FC_2-FC_4 and C_z-C_2 , while for subject AB4, more features were selected from channels FC_4-C_4, FC_2-FC_4, F_1-F_z, C_2-C_4, F_1-FC_1,FC_2-C_2 as the best set. *These results support the hypothesis that proper channel selection for every subject is necessary in order to obtain superior performance.*

2) From Fig. 4 and Fig. 5, it is seen that the relevant features were unique for each subject. However, this finding should not come as a surprise. It has been shown that the selected features are not necessarily located in the standard frequency bands or on specific scalp locations and that the set of selected features are different from subject to subject [20]. *These studies support the notion that a customized BI system should be designed for each subject.*

3) Our initial findings also suggested that the design of a BI system with low FP rates using monopolar features extracted from a single neurological phenomenon is not straightforward. In Study I, when the MCS system was designed for each individual neurological phenomenon, the MRP-based system yielded a TP=54.92 and FP=3.03 for Subject AB1 and TP=59.87 and FP= 6.93 for Subject AB4 (the results obtained for CPMR and CPBR were inferior). Only when the outputs of these classifiers were combined, the FP rate decreased. In Study II, as mentioned in Section IV, the results of the analysis of both subjects showed FP>4%. *These findings suggest the use of bipolar features in future studies.*

It should be mentioned that it is difficult to directly compare the results of our study and other self-paced BI studies, as the number of subjects and their description (whether or not they are able-bodied), recording equipment, recording protocols, classification protocols, and neurological phenomena considered are different from one study to another.

Future studies will focus on combining the findings of these two studies in order to come up with a more sophisticated self-paced BI system with lower FP and higher TP rates, which has a more robust performance. Other prospects for future work include the extension of the obtained results to more subjects and also to subjects with motor disabilities, extension to continuous signals, online testing of the proposed method and automating the process of selecting the classifiers' parameters.

REFERENCES

[1] S. G. Mason and G. E. Birch, "A brain-controlled switch for asynchronous control applications," *IEEE Trans. Biomed. Eng.,* vol. 47, pp. 1297-1307, Oct. 2000.

[2] G. E. Birch, Z. Bozorgzadeh and S. G. Mason, "Initial on-line evaluations of the LF-ASD brain-computer interface with able-bodied and spinal-cord subjects using imagined voluntary motor potentials," *IEEE Trans. Neural Syst.*

Rehabil. Eng., vol. 10, pp. 219-224, Dec. 2002.
[3] L. Deecke, B. Grozinger and H. H. Kornhuber, "Voluntary finger movement in man: cerebral potentials and theory," *Biol. Cybern.,* vol. 23, pp. 99-119, Jul 14. 1976.
[4] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: basic principles," *Clin. Neurophysiol.,* vol. 110, pp. 1842-1857, Nov. 1999.
[5] C. Toro, G. Deuschl, R. Thatcher, S. Sato, C. Kufta and M. Hallett, "Event-related desynchronization and movement-related cortical potentials on the ECoG and EEG," *Electroencephalogr. Clin. Neurophysiol.,* vol. 93, pp. 380-389, Oct. 1994.
[6] C. Babiloni, F. Carducci, F. Cincotti, P. M. Rossini, C. Neuper, G. Pfurtscheller and F. Babiloni, "Human movement-related potentials vs desynchronization of EEG alpha rhythm: a high-resolution EEG study," *Neuroimage,* vol. 10, pp. 658-665, Dec. 1999.
[7] W. Szurhaj, P. Derambure, E. Labyt, F. Cassim, J. L. Bourriez, J. Isnard, J. D. Guieu and F. Mauguiere, "Basic mechanisms of central rhythms reactivity to preparation and execution of a voluntary movement: a stereoelectroencephalographic study," *Clin. Neurophysiol.,* vol. 114, pp. 107-119, Jan. 2003.
[8] G. Dornhege, B. Blankertz, G. Curio and K. R. Muller, "Boosting bit rates in noninvasive EEG single-trial classifications by feature combination and multiclass paradigms," *IEEE Trans. Biomed. Eng.,* vol. 51, pp. 993-1002, Jun. 2004.
[9] B. Graimann, J. E. Huggins, S. P. Levine and G. Pfurtscheller, "Toward a direct brain interface based on human subdural recordings and wavelet-packet analysis," *IEEE Transactions on Biomedical Engineering,* vol. 51, pp. 954-962, 2004.
[10] H. S. Liu, X. Gao, F. Yang and S. Gao, "Imagined hand movement identification based on spatio-temporal pattern recognition of EEG," in *1st International EMBS/BMES Conferences on Neural Engineering,* 2003, pp. 599-602.
[11] Y. Wang, Z. Zhang, Y. Li, X. Gao, S. Gao and F. Yang, "BCI Competition 2003--Data set IV: an algorithm based on CSSD and FDA for classifying single-trial EEG," *IEEE Transactions on Biomedical Engineering,* vol. 51, pp. 1081-1086, 2004.
[12] D. Coyle, G. Prasad and T. M. McGinnity, "Extracting features for a brain-computer interface by self-organising fuzzy neural network-based time series prediction," in 2004, pp. 4371-4374 Vol.6.
[13] J. F. Borisoff, S. G. Mason, A. Bashashati and G. E. Birch, "Brain-computer interface design for asynchronous control applications: improvements to the LF-ASD asynchronous brain switch," *IEEE Trans. Biomed. Eng.,* vol. 51, pp. 985-992, Jun. 2004.
[14] E. L. Glassman, "A wavelet-like filter based on neuron action potentials for analysis of human scalp electroencephalographs," *IEEE Trans. Biomed. Eng.,* vol. 52, pp. 1851-1862, Nov. 2005.
[15] G. Pfurtscheller, "Graphical display and statistical evaluation of event-related desynchronization (ERD)," *Electroencephalogr. Clin. Neurophysiol.,* vol. 43, pp. 757-760, Nov. 1977.
[16] S. Lemm, C. Schafer and G. Curio, "BCI Competition 2003--Data set III: probabilistic modeling of sensorimotor mu rhythms for classification of imaginary hand movements," *IEEE Transactions on Biomedical Engineering,* vol. 51, pp. 1077-1080, 2004.
[17] T. Back, D. B. Fogel and T. Michalewicz, *Evolutionary Computation.* Bristol and Philadelphia: Institute of Physics Publishing, 2000.
[18] T. Hinterberger, A. Kubler, J. Kaiser, N. Neumann and N. Birbaumer, "A brain-computer interface (BCI) for the locked-in: comparison of different EEG classifications for the thought translation device," *Electroencephalogr. Clin. Neurophysiol.,* vol. 114, pp. 416-425, 2003.
[19] C. Chang and C. Lin, *LIBSVM: A Library for Support Vector Machines.* 2001,
[20] J. Millan, M. Franze, J. Mourino, F. Cincotti and F. Babiloni, "Relevant EEG features for the classification of spontaneous motor-related tasks," *Biol. Cybern.,* vol. 86, pp. 89-95, Feb. 2002.