APPLYING A HYBRID GENETIC ALGORITHM IN THE DESIGN OF A SELF-PACED BRAIN INTERFACE WITH A LOW FALSE POSITIVE RATE

Mehrdad Fatourechi [1], *Gary E. Birch* [1,2,3], and *Rabab K. Ward* [1,3]

[1]Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada V6T 1Z4
[2] Neil Squire Society, Burnaby, BC, Canada V5M 3Z3
[3] Institute for Computing, Information and Cognitive Systems, Vancouver, BC, Canada V6T 1Z4
Email: {mehrdadf, rababw} @ece.ubc.ca, garyb@neilsquire.ca

ABSTRACT

A new hybrid genetic algorithm (HGA) for optimization of a self-paced brain interface (SBI) is proposed. To identify intentional control (IC) commands in the noisy background EEG signal, the proposed SBI uses features extracted from three neurological phenomena-- movement-related potentials as well as changes in the power of Mu and Beta rhythms. To identify the IC commands, for each neurological phenomenon, a multiple classifier system (MCS) is designed. Then a 2nd -stage MCS combines the outputs of the individual MCSs and generates the final decision. The HGA selects the optimal subset of features, the optimal parameter values of the classifiers, as well as the best configuration for combining the MCSs. Analysis of the data of four subjects shows an average TP= 56.18%,and an average FP=0.14%, a significant improvement over our previous SBI design.

Index Terms—Biomedical signal processing, Wavelet transforms, Genetic algorithms, Pattern classification, Electroencephalography

1. INTRODUCTION

A brain interface (BI) allows individuals to control devices using their brain signals only. Self-paced BIs (SBIs) form a sub-class of BI systems that allow such control at any time and at the user's own pace. An SBI detects an intentional control (IC) command in the EEG. At other times, the user is in a No Control (NC) state.

Unfortunately, the performance of EEG-based SBIs is still not suitable for most practical applications. Several methods can be employed to boost the performance of SBIs. They include the employment of more sophisticated signal processing algorithms, incorporating the spatial information as well as the use of more than one neurological phenomenon. To achieve low false positive (FP) rates, we proposed in [1] the simultaneous use of three neurological phenomena. The proposed SBI used features extracted from movement-related potentials (MRP), changes in the power of Mu rhythms (CPMR) and changes in the power of Beta rhythms (CPBR). The main rationale behind using these specific neurological phenomena is that they are time locked to the onset of movement. The evidence from the literature supports this hypothesis [2, 3].

In [1], a two-stage multiple-classifier system (MCS) that detects an IC command using features extracted from MRP, CPMR and CPBR is proposed. For each neurological phenomenon and EEG channel, one feature was extracted; thus three features were generated per channel. Each feature was extracted using matched filtering with a specific template that corresponded to the specific neurological phenomenon, and was classified using a K-nearest neighbor classifier. The process was repeated for all EEG channels. For each neurological phenomenon, an MCS combined the outputs of the individual classifiers. Next, a 2nd -stage MCS was used to combine the outputs of the three MCSs designed in the first stage. To reduce the dimensionality of the feature space, a genetic algorithm (GA) selected a subset of the features. For simplicity, the values of the parameters of all classifiers were assumed to be the same, and were selected through an exhaustive search.

Several factors could limit the application of this system. While, it achieved a low FP rate (1.25%), this was at the expense of having a low TP rate (the average TP rate on the data of four subjects was 29.05%). Moreover, because of the computational complexity involved, the MCS for each neurological phenomenon was designed separately. Since all three MCSs contributed to the overall performance, this process is sub-optimal.

To improve the performance of the SBI proposed in [1], in this paper, we introduce more sophisticated feature extraction and feature classification methods. Specifically, a hybrid genetic algorithm (HGA) is proposed .The HGA consists of a GA followed by a local search (LS). The GA selects the best subset of features and the optimal values of classification parameters, and the LS determines the best structure for the 2nd -stage MCS. Analysis of the data of four able-bodied subjects shows that the proposed method achieves significantly better performance than our previous SBI.

2. METHODS

The improved SBI is shown in Figure 1. Its components are described below.

2.1. Feature extraction

The discrete wavelet transform (DWT) is a powerful tool for extracting time-frequency features of a signal. However, its use is limited because of its shift-variancy property. The use of a stationary wavelet transform (SWT) resolves the shift-variancy problem associated with the DWT by eliminating the downsampling operator from the multi-resolution analysis of the DWT [4]. However, the use of stationary wavelet coefficients as features, results in a dramatic increase in the dimension of the feature space. We thus propose a combination of SWT and

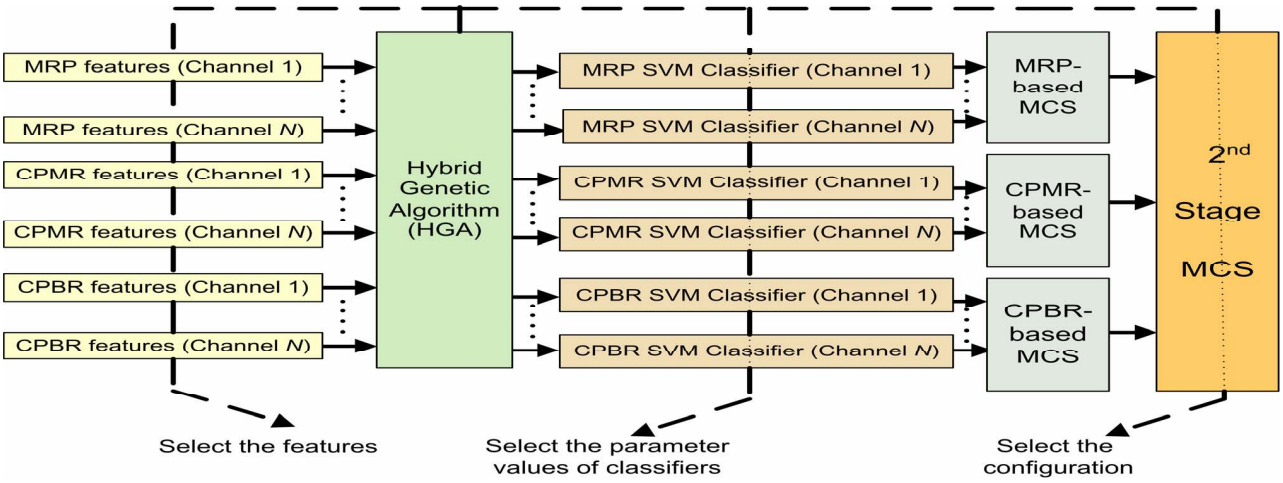

Figure 1. The overall structure of the proposed SBI (the dashed lines show the parameter values controlled by the HGA).

matched filtering to deal with the large-dimension feature space as described below.

For MRPs, after applying a 5-level SWT decomposition, the wavelet coefficients in the lowest approximation and detail levels (in the [0-2] and [2-4] Hz range) were selected. For each level, we averaged the coefficients extracted from the IC commands over the epochs in the training sets to generate the template of SWT features for that scale. The cross-covariance of the wavelet coefficients of the p'th epoch and the template at scale j was calculated, and the following feature (which represents the maximum of the cross-correlogram over a period of 0.125 seconds) was extracted:

$$F_{j,p} = \max(XCOR_{j,p}(n)), \quad (1)$$

where

$$n \in [t_{start} + t_{finish} - 0.0625, t_{start} + t_{finish} + 0.0625]$$

$XCOR(n)$ denotes the cross-covariance function. t_{start} and t_{finish} specify the length of the epochs. We chose a period of 0.125 seconds because the features of interest lie in the frequencies below 4 Hz and were generated by sliding a 0.125-second window over the EEG signals (see Section 3 for more details).

A second feature that provides information about the time instant when the maximum of the cross-correlogram has occurred was generated as follows:

$$T_{j,p} = t(XCOR_{j,p}(n) = F_{j,p}) \quad (2)$$

where in (2), t is the time operator. This process was repeated for all EEG channels, resulting in $4 \times N$ features for each neurological phenomenon, where N is the number of EEG channels.

As for the changes in the Mu and Beta rhythms (CPMR and CPBR), all epochs were band-pass filtered before feature extraction. For CPMR, the band pass was chosen from 8 to 12Hz. For CPBR, because of the relatively wide range of the Beta rhythms (typically in the range of 14-30Hz) , a user-customized band pass was chosen for each subject, as explained below. Both filters were linear-phase 32-point FIR filters. The amplitudes of the signals were squared to obtain the power values. The SWT was then applied, and the wavelet coefficients of the lowest approximation and detail levels were calculated. The process of feature extraction is then similar to that used above for the MRPs.

The choice of the proper wavelet function. The proper wavelet function is usually chosen based on the similarity between the underlying neurological phenomenon and the shape of the wavelet function. Unfortunately, the choice of the wavelet function may become subjective. Moreover, it has been shown that the shape of the underlying neurological phenomenon may vary from one subject to another [5]. We thus propose an automatic method to select the wavelet function. For each subject, we define the Fisher ratio:

$$C(p,q,r,s) = \frac{(\mu_{IC}(p,q,r,s) - \mu_{NC}(p,q,r,s))^2}{\sigma^2_{IC}(p,q,r,s) + \sigma^2_{NC}(p,q,r,s)}$$
$$(p = 1,2,3 \quad ; \quad q = 1,2,...,N;$$
$$r = 1,2,...,N_{Features} \quad ; \quad s = 1,2,...,N_{WaveletFunctions}) \quad (3)$$

where $\mu_{IC(p,q,r,s)}$ and $\mu_{NC(p,q,r,s)}$ are the means and $\sigma^2_{IC}(p,q,r,s)$ and $\sigma^2_{NC}(p,q,r,s)$ are the variances of IC and NC classes for feature r of channel q extracted using the wavelet function s for neurological phenomenon p. N is the number of EEG channels , $N_{Features}$ is the total number of features, and $N_{WaveletFunctions}$ is the total number of wavelet functions. For each channel q and for each neurological phenomena p , the wavelet function that achieves the following objective is then chosen:

$$Max_{r,s}[(C(p,q,r,s))], \quad (p = 1,2,3; q = 1,2,...,N) \quad (4)$$

The wavelet functions were selected from a pool of Daubechies, Biorthogonal, Symlet and Coiflet wavelet functions (46 wavelet functions in total). Features were normalized prior to the calculation of the Fisher ratio.

The choice of proper CPBR frequency band. In order to select more discriminant CPBR features, the Fisher ratio in Eq. 3 was calculated for seven CPBR frequency bands: [14-18], [18-22], [22-26], [26-30], [18-26], [22-30] and [14-30]Hz. The averages of the maximum of Fisher ratios (as calculated in Eq.4) over all N EEG channels were compared, and the frequency band that resulted in the highest average was selected.

2.2. Feature classification

For each pair of neurological phenomenon and EEG channel, the features were classified using a support vector machine (SVM) classifier. The outputs of the classifiers were then combined using a two-stage MCS (see Figure 1).

2.2.1. Support vector machines (SVMs)

A total of $3\times N$ SVM classifiers with Gaussian kernels were designed. The performance of each classifier was determined by its regularization parameter C and the bandwidth σ of the kernel. Since there are $3\times N$ classifiers, $3\times N$ values had to be estimated for each (σ and C) parameter.

2.2.2. Multiple classifier systems (MCSs)

For each neurological phenomenon, an MCS that classifies the outputs of the individual SVM classifiers using a majority voting rule is used. Since there are N EEG channels for each neurological phenomenon, the outputs of N SVMs are combined. A 2nd-stage MCS combines the outputs of the three MCSs into five combinations, as described below. The output of this MCS is determined by an HGA. It forms the final classification label for each epoch as an IC or an NC state.

The following configurations were investigated for combining the outputs the MCSs in the first stage of the two-stage MCS: (1) configuration 1- using the AND operation to combine the binary outputs of MRP-based and CPBR-based MCSs; (2) configuration 2- using AND operation to combine the outputs of MRP-based and CPMR-based MCSs; (3) configuration 3- using the AND operation to combine the outputs of CPBR-based and CPMR-based MCSs; (4) configuration 4- combining the outputs of all MCSs according to the majority voting rule; (5) configuration 5- using the AND operation to combine the outputs of all MCSs .

2.3. Hybrid genetic algorithm (HGA)

The HGA has the following tasks: feature selection, determining the parameter values of the SVM classifiers, and selecting the best of the above 5 configurations.

To represent each possible combination of features and parameter values of SVM classifiers, we defined a binary chromosome of length $L_{\text{Chromosome}}$ (see Figure 2). Bit i of the first $N_{features}$ bits of the binary chromosome specified whether or not feature i was selected by the HGA. A value of "1" indicated the presence of feature i and a value of "0" indicated its absence in a chromosome. We used the second part of the chromosome to select the parameter values of the $3\times N$ SVM classifiers. For each SVM classifier, two parameter values had to be determined: C and σ; for each parameter, 4 bits were used to represent the values of that parameter. For each chromosome, a local exhaustive search was then carried out to find the best of the 5 configurations in the 2nd-stage MCS. Suppose x denotes a model in Figure 2. The objective function for the HGA was defined as in Eq.5.

$$f_1 : Max_x[K\times\frac{\text{mean}(TP(x))}{\text{mean}(FA(x))\times(\text{Var}(TP(x))+\text{Var}(FP(x)))})]$$

$$K=\begin{cases}1, & \text{if} \quad \text{mean}\,(TP(x))>T_2\\ \dfrac{mean\,(TP(x))-T_2}{T_2-T_1}, & \text{if} \quad T_1<\text{mean}\,(TP(x))<T_2\\ 0, & \text{if} \quad \text{mean}\,(TP(x))<T_1\end{cases} \quad (5)$$

where *FA* (false activations) is the percentage of NC epochs that are affected by a false detection. The main difference between the FA and FP rate is that multiple FPs in an epoch are counted as one FA. The values of T_1 and T_2 in Eq. 5 were selected as 50% and 80%, respectively for all subjects except for subject AB3. For AB3, the value of T_1 seemed to be too high and resulted in the generation of chromosomes with high FA values. For this subject, the values of T_1 and T_2 were chosen as 33% and 50%, respectively. The "mean" operator is applied over the inner-validation sets (see Section 3).

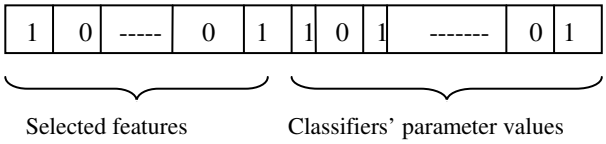

Figure 2. Structure of the chromosome

We implemented a lexicographic approach for sorting the chromosomes in the HGA population [6]. In this approach, the chromosomes were compared and ranked according to the values of $f_1(x)$ in Eq.5. Any ties were resolved by comparing the relevant chromosomes again with respect to another objective with a lower priority .If there was also a tie in the 2nd objective function, a 3rd objective function was used for comparison and so on. We defined the rest of the objective functions as follows (in the order of priorities):

$$f_2 : Min_x[\text{mean}(FA(x))\times\text{Var}(FA(x))] \quad (6)$$

$$f_3 : Max_x[\frac{\text{mean}(TP(x))}{\text{Var}(TP(x))}] \quad (7)$$

$$f_4 : Min_x[mean(FP(x))\times\text{Var}(FP(x))] \quad (8)$$

$$f_5 : Min_x[(N(x))] \quad (9)$$

$$f_6 : Min_x[(N_{Features}(x))] \quad (10)$$

Where in Eq. (9), N is the number of channels and in Eq.(10) $N_{Features}$ is the number of features. The "mean" operator is applied over the results obtained from inner-validation sets (see Section 3). The operators of the HGA were tournament-based selection (tournament size=3), uniform crossover (p=0.9) and uniform mutation (p=0.01). The sizes of the initial population and the populations in the next generations were chosen as 200 and 100, respectively. The population of HGA was randomized initially. Elitism was used to keep the best performing chromosome of each population in the subsequent populations, and the number of evaluations was set to 5000. If the improvement in the first objective of the best solution was found to be less than 1% for more than 10 consecutive generations, the HGA was terminated.

3. RESULTS

The data were collected from four right-handed (three males and one female) able-bodied subjects between 31 and 56 years old. They had all signed consent forms prior to participation in the experiment. The subjects performed a right index finger flexion. The EEG signals were recorded from 13 monopolar EEG channels (according to the International 10-20 System at F_1, F_z, F_2, FC_3, FC_1, FC_z, CF_2 , FC_4, C_3, C_1, C_z ,C_2, and C_4 locations) and were sampled at 128 Hz. The signals were then converted to bipolar EEG signals since it has been shown that bipolar electrodes are more likely to generate more discriminant features than monopolar electrodes [7]. The conversion was carried out by calculating the difference between adjacent EEG channels and resulted in the generation of 18 bipolar EEG channels (see [1] for details).

Our IC epochs consisted of data collected over an interval containing the onset of movement (measured as the finger switch activation). The data obtained from t_{start} =1.0 second before to t_{finish} =1 second after the movement onset were analyzed.

We recorded our NC sessions as follows. During an NC session, the subject was asked to count the number of times that a white ball bounced off the screen. The NC sessions therefore contained attentive as well as non-attentive NC data. Each NC session lasted approximately two minutes; with up to 2 such sessions were recorded each day. We then selected the NC periods as follows. A window of width $(t_{start}+t_{finish})$ seconds was slid over each EEG signal collected during the NC sessions by a step of 16 samples (0.1250 sec). For each 1-second window where artifacts were not detected, features were extracted. The IC and NC epochs for which the electrooculography (EOG) activity exceeded a pre-defined threshold (±25 μV) were automatically rejected.

We used a five-fold nested cross-validation to evaluate the performance of the system. The inner cross-validation set was used for model selection and the outer cross-validation set was used to estimate the generalization error. For each outer cross- validation set, 20% of the data were used for testing and the rest were used for training and model validation. In order to select the models, the datasets were further divided into five folds. For each fold, 80% of the data were used for training the classifier and 20% were used for model validation.

The test results are shown in Table 1. The columns show the subject IDs, and the first row shows the selected configuration. For subjects AB1 and AB4, the combination of MRP and CPBR led to superior results (configuration 1), while for subjects AB2 and AB3, the combination of all three neurological phenomena using the AND operation was the best configuration (configuration 5).

The next three rows show the average of the TP, FA and FP rates on the test sets, consisting of the average of five runs over the outer cross-validation set. The numbers in parentheses show the standard deviations.

Table 1. Comparison of the performance.

	AB1	**AB2**	**AB3**	**AB4**
Results from current study				
Configuration	1	5	5	1
Average TP	58.64 (8.56)	64.16 (7.53)	46.89 (10.38)	55.06 (5.33)
Average FA	1.24 (1.13)	0.16 (0.35)	2.26 (1.39)	0.79 (0.82)
Average FP	0.15 (0.14)	0.02 (0.04)	0.28 (0.17)	0.12 (0.15)
Results from [1]				
Configuration	4	4	4	4
Average TP	26.00 (9.49)	26.96 (3.95)	31.31 (13.46)	31.88 (8.96)
Average FP	0.13 (0.09)	0.28 (0.15)	3.45 (3.99)	1.15 (0.31)

4. DISCUSSION AND CONCLUSIONS

A novel, improved self-paced brain interface is proposed. This SBI system uses features extracted from three neurological phenomena (MRP, CPMR and CPBR) to identify IC commands.

We compared the results with those reported in [1] (see the last three rows in Table 1), as both studies used similar experimental and evaluation paradigms. The average TP rate showed a significant increase to 56.19% compared to 29.04% from [1] , and the average FP rate also decreased significantly to 0.14% compared to 1.25% from [1] (see Table 1). This decrease in the FP rate also translated into the proposed design having an average of 1.11 FPs every 100 seconds. The improved design of another SBI called the LF-ASD, had an average FP of 1 every 6 seconds [8], while and the SBI proposed in [1] had a FP of 1 every 10.5 seconds. The improved SBI is thus able to recognize a longer NC period without having a FP, while achieving an acceptable TP rate depending on the target application. These results indicate that the new SBI is superior to that in [1]. This superior performance, however, entails an increase in the complexity of the system.

REFERENCES

[1] M. Fatourechi, G. E. Birch and R. K. Ward, "Recent studies in the design of a self-paced brain interface with low false positive rate," in *IEEE Int. Conf. of the Engineering in Medicine and Biology Society (EMBC'06),* 2006, pp. 2944-2949.

[2] 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.

[3] 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.

[4] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications," Wavelets *and Statistics,* vol. 103, pp. 281–299, 1995.

[5] A. Bashashati, M. Fatourechi, R. K. Ward and G. E. Birch, "User customization of the feature generator of an asynchronous brain interface," *Ann. Biomed. Eng.,* vol. 34, pp. 1051-1060, Jun. 2006.

[6] T. Back, D. B. Fogel and T. Michalewicz, *Evolutionary Computation.* Bristol and Philadelphia: Institute of Physics Publishing, 2000,

[7] 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.

[8] 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.