

Selection of Relevant Features for Classification of Movements from Single Movement-Related Potentials Using a Genetic Algorithm

E. Yom-Tov, G. F. Inbar

Faculty of Electrical Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel
elad@ieee.org, inbar@ee.technion.ac.il

I. INTRODUCTION

Brain-computer interfaces (BCI's) are devices intended to help disabled people to communicate with a computer using the brains' electrical activity as the sole medium. Currently, such devices are realized using feedback [1], visual evoked potentials (EP's) [2] and a combination of feedback and imagined movements [3].

Our work is aimed at constructing a BCI based on Movement-Related Potentials (MRP's). These potentials can be recorded from the scalp when a person performs a voluntary movement, or imagines such a movement. The main problem hindering the construction of such a BCI is that MRP's are recorded from the scalp at an unfavorable signal to noise ratio (SNR) in the order of $-15[\text{dB}]$ [4]. Nevertheless, such a BCI offers a major advantage over most existing BCI's in that it operates asynchronously, i.e. the user can initiate communication without an external queue, in contrast with many current BCI's that require the user to respond to computer-generated queues.

An MRP-based BCI consists of three main blocks: A detector, a classifier, and a decoder. The detector locates time instances where the electroencephalographic (EEG) signal contains an imbedded MRP. Designs for such detectors have been suggested in [5] and [6]. Next, a classifier resolves which limb corresponds to the imbedded MRP. It is this block that the current article attempts to solve. Finally, the detector transforms a sequence of imagined movements into letters or actions, as in [7].

One of the major difficulty one encounters when designing a classifier is choosing relevant features from the vast number of possible features. Feature selection is necessary because irrelevant features are known to cause the classifier to have poor generalization, increase the computational complexity, and require many training samples to reach a given accuracy [8].

Past attempts at feature selection have usually focused on selecting features from a relatively small number of features drawn from one family [9], or on selecting a family of

features from several possible feature families by testing the performance of each feature family against several subjects [10]. These methods do not produce the best possible performance because for each subject a different feature family may be best. Indeed, it may be the case that the best features for classification are found in several feature families, and thus restricting the search to one family of features results in sub-optimal performance.

The goal of our study is to classify contralateral finger movements using MRP's buried in on-going EEG noise. This is achieved by generating a large bank of features using several feature extraction techniques, and selecting a small number of them using a genetic algorithm. These few features are then used as input to a support vector machine classifier.

II. METHODOLOGY

Four subjects (male, 27-30 years old) participated in the study. The subjects do not suffer from neurological or muscular disorders. Informed consent was obtained from the subjects.

Each subject was seated on an armchair, with his palms on a table and his feet on a footstool. Micro-switches were placed under both index fingers. The subject was told to press the micro-switches in random order, self paced, as briefly as possible. The subject was requested to pause for approximately 3 seconds between each movement, and the experimenter inform him when he was too quick. The experiment lasted for 22 minutes and 20 seconds, during which the subject made between 80 and 150 movements of each finger.

Cortical potentials were recorded using electrodes placed over F_{P1} , F_{P2} , F_3 , F_4 , C_3 , C_4 , T_3 and T_4 , all referenced to an electrode over C_z (according to the 10-20 system, using an Electro-Cap). The electrodes were Ag-AgCl surface electrodes, circular, with a 6-mm diameter. Resistance between electrodes was approximately $5K\Omega$. The state of the micro-switches was recorded in order to synchronize events in the EEG with external events.

The EEG channels were amplified using a custom made optically isolated amplifier with a gain of 10,000 and a 0.01-40Hz pass band. The amplified signals were digitized and sampled, together with the micro-switch states, at 250Hz. The samples were saved to disk, and processing was performed offline using Matlab.

Report Documentation Page

Report Date 25 Oct 2001	Report Type N/A	Dates Covered (from... to) -
Title and Subtitle Selection of Relevant Features for Classification of Movements From Single Movement-Related Potentials Using a Genetic Algorithm		Contract Number
		Grant Number
		Program Element Number
Author(s)		Project Number
		Task Number
		Work Unit Number
Performing Organization Name(s) and Address(es) Faculty of Electrical Engineering Technion-Israel Institute of Technology Haifa 32000, Israel		Performing Organization Report Number
Sponsoring/Monitoring Agency Name(s) and Address(es) US Army Research, Development & Standardization Group (UK) PSC 802 Box 15 FPO AE 09499-1500		Sponsor/Monitor's Acronym(s)
		Sponsor/Monitor's Report Number(s)
Distribution/Availability Statement Approved for public release, distribution unlimited		
Supplementary Notes Papers from 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, October 25-28, 2001, held in Istanbul, Turkey. See also ADM001351 for entire conference on cd-rom.		
Abstract		
Subject Terms		
Report Classification unclassified	Classification of this page unclassified	
Classification of Abstract unclassified	Limitation of Abstract UU	
Number of Pages 3		

Five types of feature extraction methods were used in the study. These are:

- a. AR: The autoregressive coefficients of orders one through eight, obtained using the Yule-Walker method [11].
- b. PSD: The estimated power spectral density calculated using Welch's averaged, modified periodogram method [12], in the 0 to 32[Hz] frequency range. Bin size was 2[Hz].
- c. Barlow: The mean frequency and the mean amplitude in two spectral bands: 5-15[Hz] and 10-13[Hz].
- d. Mean: The mean amplitude difference between every pair of recorded electrodes.
- e. STD: The standard deviation of the amplitude difference between every pair of recorded electrodes.

The features were extracted for each micro-switch press in three time intervals:

- a. From 1.1[sec] before the micro-switch press until 1[sec] after it. These times contain the whole movement-related potential.
- b. From 0.4[sec] before the micro-switch press until the instant that it was pressed. This time interval corresponds to parts of the preparatory potential.
- c. From 0.3[sec] before the micro-switch press until 0.3[sec] after it. This time period corresponds to the Pre-motion potential, the Motor potential and the feedback potentials of the MRP [13].

This feature extraction resulted in 1092 features for each micro-switch press. Attempting to classify the samples using all 1092 features resulted in errors not significantly smaller than those obtained by chance. Therefore, we attempted to select a small number of features, which would (hopefully) give better classification results.

A Genitor type [14] genetic algorithm was used in order to select a small number of significant features from the feature set. The use of a genetic algorithm, rather than algorithms based on probability density estimation such as [15], was warranted by the relatively small number of training examples.

Five-fold cross validation and the SVMlight software package [16] was used to build and test classifiers. The score of a feature group was calculated as the average percentage of correctly classified tests examples.

The genetic algorithm proceeds as follows:

1. Randomly partition the a set of N features into $\lfloor N / N_F \rfloor$ groups, where N_F is the number of features to be used for classification.

2. Classify the examples using the feature groups and order them according to their score (defined above).
3. Discard one-third of the groups that have the lowest score, and build the same number of new groups by randomly selecting half the features from the remaining two-thirds of the groups, and combining them.
4. Repeat steps 2-3 for a predetermined number of iterations.

It was experimentally determined that repeating steps 2-3 for more than 30 iterations did not, usually, result in significant improvements of the classifier, and thus the algorithm was run for that number of iterations.

The algorithm was tested on the data recorded from the four subjects. For each subject we attempted to pick between 1 and 180 relevant features. In order to obtain a good estimation of the algorithms' performance, it was run 5 times for each of the number of features.

III. RESULTS

Fig. 1 shows the average success rate of the classifiers as a function of the number of features used for classification. From this figure it is evident that the best classification rates were obtained using 20 features, and that the degradation caused by using only 10 features was small. Considering that additional features generate a higher computational load, we chose to use 10 features for classification.

Based on these results, we examined which channels, time intervals, and feature extractors were selected most frequently as relevant to the classification problem. The results of this test are shown in Fig. 2.

As demonstrated in Fig. 2, the most frequently selected channels are those located over the motor cortex (C3, C4 and T4). This is to be expected because the main physiologic difference between the movements of the two fingers is the area of the motor cortex that activates them. AR was the most

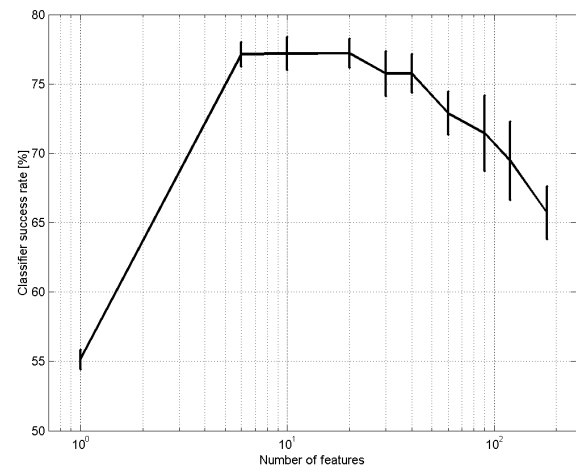


Fig. 1: Classification success rate as a function of the number of features. The best performance is obtained using 20 features. Error bars show the standard deviation of the success rate.

useful feature extractor of the five tested, a finding that is in agreement with [10]. Finally, the most influential times for classification were those before and just after the movement. This can be explained by the fact that the area corresponding to the different limbs on the motor and somatosensory cortices are activated during those times, and are thus useful in distinguishing between the two types of movement.

IV. DISCUSSION

MRP-based BCI's consist of three elements: A detector, a classifier, and a decoder. In this paper we have addressed a possible design for a classifier to distinguish between movements of contralateral fingers using MRP's imbedded in EEG.

Our results show that it is possible to select as few as 10 subject-specific features and attain an average of 77% accuracy in classification. Although this is a modest success rate, one should keep in mind that it was obtained without user feedback, that is: in an offline system. Experience has shown that allowing subject interaction can dramatically

increase performance. For example, the real-time system in [3] started with a 30% success rate, which improved to over 70% success rate after five days of training. We therefore hypothesize that implementing the above algorithm in a real-time system will, after relatively few training sessions, produce much higher classification accuracy than that attained by the off-line system.

REFERENCES

- [1] N. Birbaumer, A. Kubler, N. Ghanayim, T. Hinterberger, J. Perelmouter, J. Kaiser, I. Iversen, B. Kotchouby, N. Neumann, and H. Flor, "The thought translation device (TTD) for completely paralyzed patients," *IEEE Transactions on Rehabilitation Engineering*, vol. 8(2), pp. 190-193, 2000.
- [2] E. Donchin, K.M. Spencer, and R. Wijesinghe, "The mental prosthesis: assessing the speed of a P300-based brain-computer interface," *IEEE Transactions on Rehabilitation Engineering*, vol. 8(2), pp. 174-179, 2000.
- [3] J. del R. Millán, J. Mouriño, F. Cincotti, F. Babiloni, M. Varsta, and J. Heikkinen, "Local neural classifier for EEG-based recognition of mental tasks," *IEEE Transactions on Rehabilitation Engineering*, vol. 3, pp. 632-636, 2000.
- [4] D.H. Lange and G.F. Inbar, "Modern techniques in ERP research" in: *Modern Techniques in ERP Research*, U. Windhorst, H. Johansson, Eds. Berlin: Springer, 1999, pp. 997-1024.
- [5] G.E. Birch and S.G. Mason, "Brain-computer interface research at the Neil Squire Foundation," *IEEE Transactions on Rehabilitation Engineering*, vol. 8(2), pp. 193-195, 2000.
- [6] E. Yom-Tov and G.F. Inbar, "On the detection of Movement-Related Potentials from on-going EEG for use in a Brain-Computer Interface," EE Publication No. 1239, Department of Electrical Engineering, Technion.
- [7] J. Perelmouter and N. Birbaumer, "A binary spelling interface with random errors," *IEEE Transactions on Rehabilitation Engineering*, vol. 8(2), pp. 227-231, 2000.
- [8] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial Intelligence*, vol. 97, pp. 245-271, December 1997.
- [9] D. Flotzinger, M. Pregenzer, and G. Pfurtscheller, "Feature selection with distinction sensitive learning vector quantization and genetic algorithms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 6, pp. 3448-3451, 1994.
- [10] A. Schlögl, B. Schack, G. Florian, K. Lugger, M. Pregenzer and G. Pfurtscheller, "Classification of Single trial EEG: A comparison of different parameters," in *Proceedings of the 1996 IEEE International Conference on Systems, Man, and Cybernetics*, H. Witte, U. Zwiener, B. Schack, A. Doering, Eds. Jena: Druckhaus Mayer, 1996, pp. 266-268.
- [11] C.W. Therrien, *Pattern Recognition: Principles and Practice*, Englewood Cliffs: Prentice-Hall, 1992, pp. 411-415.
- [12] C.W. Therrien, *Pattern Recognition: Principles and Practice*, Englewood Cliffs: Prentice-Hall, 1992, pp. 595-596.
- [13] L. Deecke, W. Becker, B. Grözing, P. Scheid, and H. Kornhuber, "Human brain potentials preceding voluntary limb movements," *Brain*, vol. 33(supp), pp. 87-94, 1973.
- [14] D. Whitley, "A genetic algorithm tutorial," *Artificial Intelligence Magazine*, vol. 4(2), pp. 65-85, June 1994.
- [15] D. Koller and M. Sahami, "Toward optimal feature selection," *Machine Learning - Proceedings of the Thirteenth International Conference*, pp. 284-292, 1996.
- [16] T. Joachims, "Making large-Scale SVM learning practical," in: *Proceedings of the 1999 Conference on Neural Information Processing Systems*, B. Schölkopf, C. Burges, and A. Smola, Eds. Cambridge: MIT-Press, 1999.

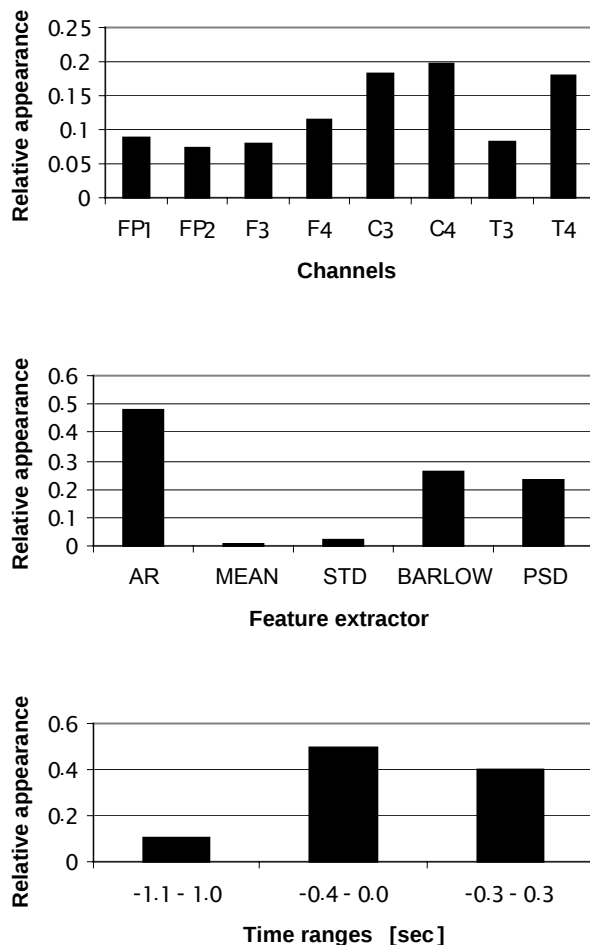


Fig. 2: Relative usage of features as the 10 best features for classification as a function of the recording channel, feature extractor and time interval.