FIRST ATTEMPTS TOWARDS A BRAIN COMPUTER INTERFACE (BCI) IN EGYPT

M. Zaky Rasmy[1], Osama M. Yousry[2], Tamer M. Adel[1], Walaa O. Gamal Al Din[1], and Yasser M. Kadah[1]

[1]Department of Biomedical Engineering, Cairo University, Giza, Egypt

[2]Institute of Biomedical Engineering, Lübeck University, Schleswig-Holsteins, Germany

e-mail: m.zaky@k-space.org

Abstract-**Brain Computer Interface (BCI) offers people with severe neuromuscular disorders a new communication channel with the outside world using only their thoughts. This paper presents the graduation project (Biomedical Engineering Department, Cairo University, 2004/05 grade) work as the first attempt of BCI research in Egypt. We developed a complete BCI system. We applied Spectral Subtraction Denoising for artifact removal, hypothesis t-test & PCA for feature extraction, & finally, we applied 4 different classifiers (Bayes minimum error, minimum distance, K-NN, feed-forward Neural Network classifiers) on data set 1, BCI competition III.**

Keywords **- BCI, ECoG, SSD, t-test, PCA, Bayes, minimum distance, K-NN, Neural Networks classifiers.**

I. INTRODUCTION

Brain-Computer Interface (BCI) is a communication system, which enables the user to control special computer applications by using only his or her thoughts. It was defined in the first international meeting devoted to BCI research held in June 1999 at the Rensselaerville Institute near Albany, New York: "A brain computer interface is a communication system that does not depend on the brains normal output pathways of peripheral nerves and muscles" [1]. Every movement, perception and thought we perform is associated with distinct neural activation patterns. BCI records the signals produced by the brain, picks out specific patterns from these signals and classifies these patterns into different categories; these classified categories can be associated with simple computer commands.

Electrocorticographic activity (ECoG) is an invasive technique, recorded from the cortical surface. ECoG has higher spatial resolution than Electroencephalography (EEG) (i.e., tenths of millimeters versus centimeters), broader bandwidth (i.e., 0–200 Hz (full) versus 0–80 Hz), higher amplitude (i.e., 50–200 μV maximum versus 10–50 μV), and far less vulnerability to artifacts such as EMG. At the same time, because ECoG is recorded by subdural electrode arrays and thus does not require electrodes that penetrate into cortex, it is likely to have greater long-term stability and might also be safer than single neuron recording [2].

We worked on data set 1 <motor imaginary in ECoG recordings, session-to-session transfer>; BCI Competition III [3]. The electrical brain activity was recorded using 8x8 ECoG platinum electrode grid which was placed on the contralateral (right) motor cortex. All recordings were performed with a sampling rate of 1000Hz. Every trial consisted of either an imagined tongue or an imagined finger movement and was recorded for 3 seconds duration. The data set contains a labeled train data set & un labeled test data set. The goal is to correctly classify the test data set.

II. METHODOLOGY

A. Artifact Removal

To correctly analyze brain recorded data in any BCI system, the first step is to filter the brain signal from unwanted noise. We can classify the recorded brain activity to the following: True activation (wanted brain signal), Physiological fluctuations (signals from inside human body; eye movement, eye blink, muscle activity, heart pulse …), & Random noise (signals from technical artifacts out side human body; electrical line noise …). The latter two components are considered as nuisance and must be removed for correct results.

ECoG recorded data contains negligible physiological artifacts compared to EEG recording; thus, we'll apply minimum artifact removal processing using a new adaptive signal-preserving technique for noise suppression in brain recorded data (EEG, ECoG …) based on spectral subtraction. The technique was originally proposed for event-related functional magnetic resonance imaging (fMRI) data [4]; we are going to apply the same concept on our data. We'll apply minimum artifact removal processing using Spectral Subtraction Denoising (SSD) to remove technical artifacts.

We will consider a model that is composed of the sum of one deterministic component d(t) incorporating both the true signal and the physiological noise and an uncorrelated stochastic component η(t)

$$s(t) = d(t) + \eta(t) \tag{1}$$

Since these two components are assumed independent, the corresponding power spectra are related by

$$P_{ss}(\omega) = P_{dd}(\omega) + P_{nn}(\omega) \tag{2}$$

Hence, an estimate of the power spectrum of the deterministic component takes the form

$$P_{dd}(\omega) = P_{ss}(\omega) - P_{nn}(\omega) \tag{3}$$

That is, the signal power spectrum ($P_{dd}(\omega)$) is obtained by spectral subtraction of the noisy signal ($P_{ss}(\omega)$) and noise ($P_{nn}(\omega)$) power spectra. In order to compute the deterministic signal component from its power spectrum, the magnitude of the Fourier transform can be obtained as the square root of the power spectrum. The problem now becomes that of reconstructing the signal using magnitude only information about its Fourier transform. Here we rely on an estimate

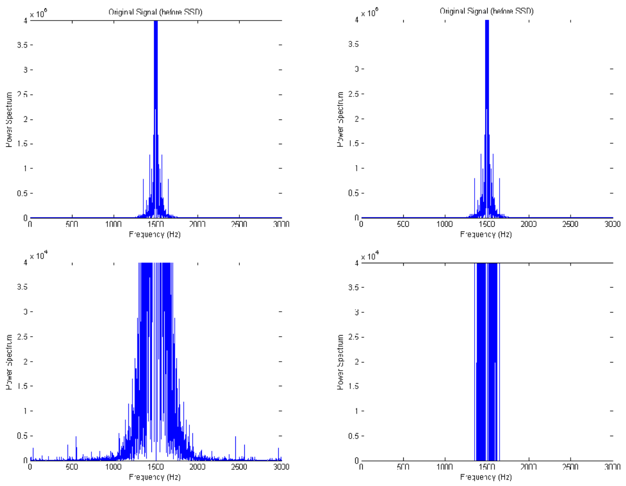

Fig. 1. Power Spectra of the original signal & the denoised signal using SSD. The bottom row shows the same power spectra with different vertical scale

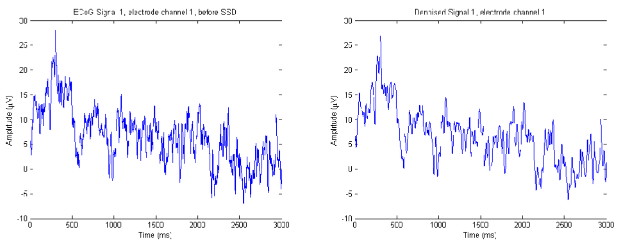

Fig. 2 Original signal before applying SSD and the denoised signal.

obtained from the phase of the Fourier transform of the original signal to overcome this problem. Hence, the Fourier transform of the processed signal can be expressed as

$$S_d(\omega) = \sqrt{P_{dd}(\omega)} \, . \, e^{j \text{ Phase } (S(\omega))} \qquad (4)$$

The enhanced deterministic signal $s_d(t)$ is then computed as the real part of the inverse Fourier transformation of this expression.

By examining the power spectra of the Train dataset signals, we estimated the random noise to be values smaller than 0.1×10^6. The power spectra of the signal 1, electrode channel 1 before and after processing with spectral subtraction is shown in Fig. 1 . Notice that the random noise (values less than 0.1×10^6 in the power spectrum) is removed in the Denoised signal.

The results of applying the Spectral Subtraction Denoising technique to process ECoG data are shown in Fig. 2. From the results shown, the application of SSD on ECoG signals was successful. As it can be observed, the noise in the original data was suppressed significantly in the output signal and the Denoised signal appears free of random noise signal components.

B. Feature Extraction

Feature extraction goal is to form a distinct set of features for each mental task to facilitate the representation & interpretation of the data. We applied hypothesis t-test & Principal Component Analysis (PCA) as simple feature extraction techniques to simplify classification process.

1) Hypothesis t-test

ECoG electrode grid placement in data set 1 covered the right motor cortex area along with surrounding cortex areas due to its size [3]. We are only interested in data collected form right motor cortex area, thus electrodes covering other areas are not significant in our study.

We applied hypothesis t-test [5] on data set 1 to reduce number of electrodes so that we use only the most representative ones. We separated train data set to 2 classes (+1, -1) & applied t-test; electrodes are rejected when the two classes are not separated using this electrode. We used MATLAB (MathWorks, Inc.) ttest2 function.

After applying the hypothesis t-test we reduced the number of electrodes from 64 to only 9 electrodes (electrodes number: 9,22,27,32,35,43,45,46,58). These 9 electrodes are the most representative electrodes for right motor cortex area.

Hypothesis testing generally is the first step to be done in any research; we applied hypothesis t-test as first step in our data analysis followed by SSD then PCA so that to work only on the 9 representative electrodes which significantly reduced processing time for upcoming signal processing (its put in this section in paper for logic sequence only)

2) PCA

PCA assumes ECoG observations are generated by the linear mixing of a number of source signals, S = XA, where S is matrix of source signals, X is the matrix of p, n dimensional observations, & A is the mixing matrix. PCA has Three assumptions: The number of sources is less than or equal to the number of observations, The mixing is linear & The mixing is instantaneous. PCA finds a linear transformation of a data set that maximizes the variance of the transformed variables subject to orthogonality constraints on the transformation and transformed variables.

Our goal from PCA is to reduce the dimension of data. Then by projecting our "old" data into the subspace of the reduced dimension data, we get our "new" features dataset. We used a MATLAB (MathWorks, Inc.) program [6] based on the FASTICA algorithm [7] & applied the following algorithm:

1- Separate the dataset into two classes (one for +1 and the other for -1 direction)
2- Work only on one electrode channel each time
3- Reduce dimensions.
4- Project every signal on this reduced subspace by using the dot product.

We chosen the range of dimensions from 1 to 20 (Fig. 3), by this we maintained 85.8306 % of (non-zero) eigen values. We have a vector 1×20 for every signal from a single electrode channel (Fig. 4), so we applied this for all the signals and get a matrix its dimension equals the total number of signals×20×3000 which represent the "new" dataset of extracted features.

Now we finished 'pre-processing' on data set 1 (artifact removal & feature extraction) & it's now suitable for applying different classification techniques.

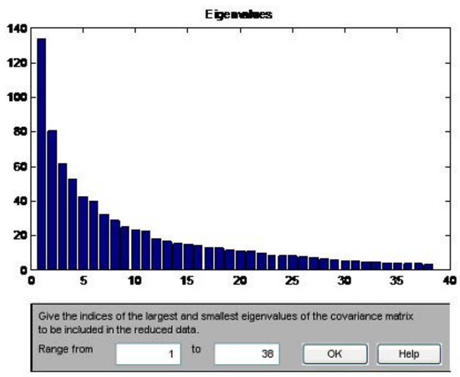

Fig. 3. Eigen values of covariance matrix

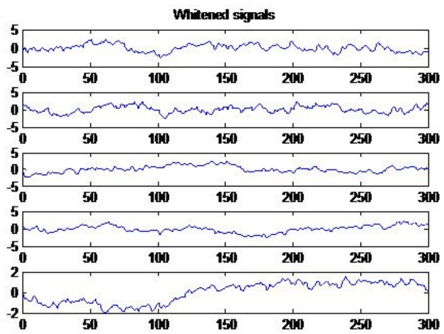

Fig. 4. PCA first 5 whitened signals as example

C. Classification

We used four classifiers, Bayes minimum-error, minimum distance, K-NN and Neural Network classifiers.

1) Bayes minimum-error classifier

The Bayes decision rule classifies an observation (test signal) to the class that has the highest a posteriori probability among classes [5], [8]. We assume dataset 1 to have a Gaussian joint probability density function (j-pdf) as in (5). In our application we have two classes, thus the a posteriori probability for the two classes is equal and equal 0.5 ($P(\omega_i=1/2)$).

$$f_x(x/\omega) = [1/(2\pi)^{N/2} |\sum_x|^{1/2}]. \exp [-1/2 (x-\mu_x)^T \sum_x^{-1} (x-\mu_x)] \quad (5)$$

The Bayes decision rule is: choose class $j \in \{1,2\}$ if

$$f_x(x/\omega_j)P(\omega_j) = \max \{ f_x(x/\omega_i)P(\omega_i)|_{i=1,2}\} \quad (6)$$

Bayes classifier doesn't hold when applied on data set 1; since the data has large dimensions (train data set 278 x 64 x 300). The covariance matrix is infinity; so we introduce a simplified approach based on Bayes decision rule. Since the covariance matrix is a diagonal matrix whose diagonal elements are the variances. Then we use the vector containing the variances instead of the whole covariance matrix in the decision rule. The class which gives the smallest result is the right class according to the modified Bayes decision rule [].

$$g_x(x/\omega) = [(x-\mu_x)^T (\sigma_x^2)^{-1} (x-\mu_x)] \quad (7)$$

The modified Bayes decision rule is: choose class $j \in \{1,2\}$ if

$$g_x(x/\omega_j) = \min \{g_x(x/\omega_i)|_{i=1,2}\} \quad (8)$$

2) Minimum distance classifier

The idea behind this method is that the mean should be a representative value for the class, defining usually the centre of all the sample vectors that were labeled as that class in input space [8].

3) Voting K-Nearest Neighbor (K-NN)

K-NN assigns a test sample to the class of the majority of its k-neighbors [8]. We used 1, 3, 5 neighbors (k = 1, 3, 5). We used Euclidean distance to calculate the metric distance between sample signal & its k neighbors. Fig. 5 shows distribution of 2 classes of train data set.

4) feed-forward Neural Network

Artificial neural networks (NNs) were originally developed with the goal of modeling information processing and learning in the brain! Neural networks are composed of simple elements, neurons operating in parallel [9].

We used MATLAB (MathWorks, Inc.) *newff* function to create feed forward back propagation network & *traingda* as network training function that updates weight and bias values according to gradient descent with adaptive learning rate. We created a 1 layer NN with 15 neurons & 800 epochs (The number of updates occurs to the network until the network error falls beneath the specified goal error). Applying the mentioned NN on train dataset, goal was met after 705 epochs (Fig. 6)

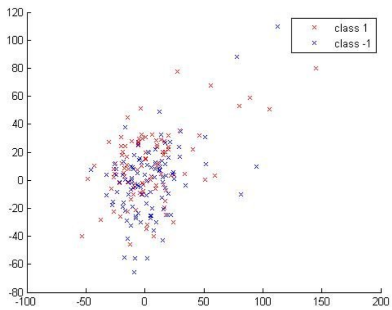

Fig.5. distribution of 2 classes of train data

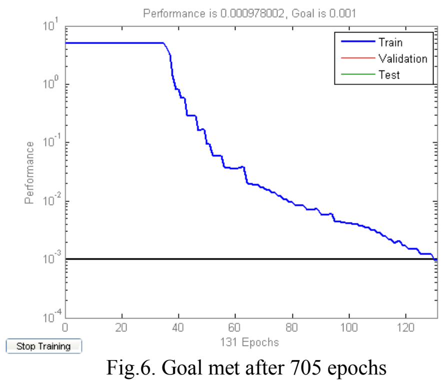

Fig.6. Goal met after 705 epochs

III. RESULTS

The results of classification of test data set using 4 different classifiers (Bayes minimum-error, minimum distance, K-NN and Neural Network classifiers) is shown in Table 1.

TABLE I
TEST DATA SET CLASSIFICATION RESULTS

Classifier	Result
Bayes Minimum Error	50 %
KNN (K=1)	51 %
KNN (K=1)	50 %
KNN (K=1)	54 %
Minimum Distance	60 %
Feed Forward Neural Network	62 %

The obtained results put as in the 19th rank out of 27 participants compared to data set I, BCI competition III announced results [10].

IV. DISCUSSION & CONCLUSION

The main aim of this paper is to present our approach to classification of data set I, BCI competition III. We located the most representative electrodes within ECoG implanted grid using hypothesis t-test - 9 electrodes from 64 electrodes - which decreased significantly the processing time without decreasing classification accuracy. We introduced SSD as a useful preprocessing technique to remove technical artifacts. PCA was applied for feature extraction. Finally we used Bayes minimum-error, minimum distance, K-NN and Neural Network classifiers. We suppose the obtained results are reasonable as primary trails.

This paper showed only our primary work as the first attempts towards BCI research in Egypt. The BCI project continues successfully in Biomedical Engineering Department, Cairo University. Further advanced work is already accomplished, that we hope we'll participate with it in BCI competition VI.

ACKNOWLEDGMENT

The authors would like to thank the rest of BCI graduation project (Biomedical Engineering Department, Cairo University, 2004/05) members; Marwa Mohamed, Nermeen Lotfy, Reem Abou Taleb & Tamer Taha for all their efforts; You are all part of this!

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, and T. M.Vaughan, "Brain computer interface technology:Areviewof the first international meeting," *IEEE Trans. Rehab. Eng.*, vol. 8, pp. 164–173, June 2000.

[2] Eric C. Leuthardt, Gerwin Schalk, J. R. Wolpaw, Jeffrey G. Ojemann and DanielWMoran; "A brain–computer interface using electrocorticographic signals in humans," *J. Neural Eng*. 1 (2004) 63–71

[3] Thomas Lal, Thilo Hinterberger, Guido Widman, Michael Schröder, Jeremy Hill, Wolfgang Rosenstiel, Christian Elger, Bernhard Schölkopf, Niels Birbaumer. "Methods Towards Invasive Human Brain Computer Interfaces," *Advances in Neural Information Processing Systems (NIPS)*, 2004

[4] Yasser M. Kadah, "Adaptive denoising of event-related functional magnetic resonance imaging data using spectrum subtraction," *IEEE Trans. on Biomedical Engineering*, vol. 51, no. 11, pp. 1944-1953, Nov. 2004.

[5] K. Sam Shanmugan, Arthur M. Breipohl; "Random Signals, Detection, Estimation and Data Analysis," John Wiley & Sons.

[6] Fast ICA program for MATLAB http://www.cis.hut.fi/projects/ica/fastica/code/dlcode.html

[7] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Networks*, vol. 10, pp. 626–634, May 1999.

[8] Andrew R. Webb; "Statistical Pattern Recognition," Second Edition; John Wiley & Sons.

[9] Yu Hen Hu, Jeng-Neng Hwang; "Hand Book of Neural Network Signal Processing," CRC press.

[10] Data set I, BCI competition III results http://ida.first.fhg.de/projects/bci/competition_iii/results/tuebingen/true_labels.txt