Classification of EEG Signals Using Different Feature Extraction Techniques for Mental-Task BCI

Sarah M. Hosni, Mahmoud E. Gadallah, Sayed F. Bahgat, Mohamed S. AbdelWahab
Ain shams university, faculty of computer and information sciences
Abbasiya, Cairo, Egypt

Abstract-**The use of Electroencephalogram (EEG) or "brain waves" for human-computer interaction is a new and challenging field that has gained momentum in the past few years. If several mental states can be reliably distinguished by recognizing patterns in EEG, then a paralyzed person could communicate to a device like a wheelchair by composing sequences of these mental states. In this research, EEG from one subject who performed three mental tasks have been classified using Radial Basis Function (RBF) Support Vector Machines (SVM) to control overfitting. A method for EEG preprocessing based on Independent Component Analysis (ICA) was proposed and three different feature extraction techniques were compared: Parametric Autoregressive (AR) modeling, AR spectral analysis and power differences between four frequency bands. The best classification accuracy was approximately 70% using the parametric AR model representation with almost 5 % improvement of accuracy over unprocessed data.**

I. Introduction

The past two decades have seen an explosion of scientific interest in a completely different and novel approach of interacting with computers: The EEG-based brain computer interface (BCI) technology [17,11]. This challenging new line in human-computer interaction added a new dimension to electroencephalogram (EEG) research. In addition to being an important diagnostic tool that enabled the study of the relation between characteristics of brain activity signals and brain abnormalities, EEG signals are being investigated as a new mode of human-computer interaction motivated by numerous studies that have demonstrated correlations between EEG signals and several mental tasks [11,18]; If a small number of mental states can be reliably classified, then a person could compose sequences of such states to indicate commands to a computer, just as letters are composed to form words.

The growing field of BCI research is still in its infancy; the current BCI systems are designed mainly to enable severely motor disabled patients to communicate through thoughts alone and no BCI system has become commercially available. However, this technology might become one day a revolutionary communication channel enabling users to control computer applications and devices.

So far the accuracy of classification has been one of the main pitfalls of the current mental task BCI systems. Enhancing the accuracy may be achieved through improvements in the three main stages of the mental task BCI, namely, the EEG preprocessing, the feature extraction and the techniques used for feature classification.

In this paper, we emphasize on the problem of EEG preprocessing and representation. Although many feature extraction techniques were proposed based on either frequency or non frequency domain information, distinguishing features are still buried in the data and the quest of the best techniques to extract information from EEG signals with which we can discriminate mental states is still going on. As a consequence, new techniques must be proposed and existing techniques must be fairly compared in order to reach a conclusion on the most appropriate feature extraction techniques.

Work reported in this paper is related to a project initiated in Purdue University by Keirn and Aunon [18], and extended later by Charles Anderson [4,5,6]. In their study, Keirn and Aunon investigated the use of five different mental tasks as a new mode for communication between man and computer, the tasks were: baseline task, mental multiplication, geometric figure rotation, mental letter composing, and visual counting. These tasks were chosen to invoke hemispheric brainwave asymmetry [18] as such, Keirn and Aunon proposed that these tasks are suitable for brain-computer interfacing.

The data were recorded using six recording channels (electrodes) from four subjects performing the five mental tasks; the recordings of mental tasks were conducted for several 10 seconds trials, trials were made in two sessions, each session was conducted in a different day. A simple Bayesian classifier was applied to data collected from pairs of tasks from both single and combined sessions and features were extracted from a single quarter-second (and two-second) window per trial, chosen as close to the "middle" of the task as possible, assuming that during that period the subject was most likely concentrating on the requested mental task. Only eyeblink free data segments were used to avoid spikes and noise.

Two sets of features were used; the first set was a frequency based representation that included asymmetry ratios and power values for each channel from four standard frequency bands: delta (1-3 Hz), theta (4-7 Hz), alpha (8-13 Hz), and beta (14-29 Hz). Asymmetry ratios given by (R - L)/(R+L), were calculated for all combinations of right to left electrodes for each frequency band, where R and L are power values from right and left hemisphere electrodes.

The second set of features was generated from the 6th order Autoregressive (AR) model coefficients calculated by the Burg method and concatenated together from all channels.

Keirn and Aunon concluded that it is possible to distinguish between the various mental tasks with a reasonable degree of accuracy, even those involving the same hemisphere, laying the groundwork for what was later called: the mental task BCI. They also found that the accuracies fell for the combined sessions compared to the single sessions due to the changing nature of EEG signals. Their classification results showed that the AR coefficients performed slightly better than the frequency based features resulting in 82.3% accuracy obtained by averaging results over all subjects and all possible task pairs.

In [4], Anderson et al extended Keirn and Aunon's work by replacing the Bayesian classifier with neural networks to classify EEG from only one subject performing two mental tasks: a baseline task and mental multiplication. 73% accuracy was obtained using Keirn and Aunon's frequency-based representation compared to other signal representations consisting of unprocessed data or a Karhunen-Loève transform of the data which resulted in classification performance that was not significantly better than chance; and only half second eyeblink free EEG windows were used covering the whole 10 seconds duration of a trial instead of only one EEG segment.

In [5], neural networks of various sizes were used to classify EEG from one subject performing three mental tasks: a baseline task, mental multiplication, and mental letter composing. Features were extracted from half-second EEG windows as 6th order AR model coefficients concluding that a single hidden layer of twenty input units performed best resulting in 76% accuracy.

In [6], attempts to differentiate between five tasks performed by four subjects resulted in 31-54% classification accuracy using 6th order AR model coefficients.

Palaniappan et al [12] proposed a frequency based feature extraction technique using spectral power and power difference in four bands from the six channels to classify pairs of mental tasks for each subject. Each pair included the baseline task combined with every other mental task from the remaining four. They extended previous work first by using gamma band (30 - 50 Hz) EEG in addition to the commonly used delta, theta, alpha and beta bands (below 30 Hz) and then by the use of spectral power differences from six channels instead of the spectral power asymmetry ratios. They suggested that it is possible to construct a simple BCI with subjects either thinking of a single mental task or relaxing; they also proposed "individual" BCI design which is suitable for use by a particular individual by selecting the best mental task combination with the baseline task for each subject.

As these previous results are reported for different number of mental tasks, different number of trials and sometimes different classifiers, the objective of our work is the evaluation and comparison of the three EEG representations recommended by previous work [4,5,12] in order to determine which of these three representations results in the best classification accuracy; and by allowing a fair comparison between them; we reach a recommendation on the most suitable feature extraction technique for a practical mental task BCI design. A major difference between our work and previous studies is the use of Independent Component Analysis (ICA) [14,16] for EEG preprocessing and the study of its effect on the classification accuracy. Removing artifacts (noise) from EEG signals rather than simply rejecting contaminated segments is recommended especially when the latter may result in a considerable loss in the amount of data presented to the classifier and in addition, it may be more suitable for a practical BCI. We also chose to classify three mental tasks not only two; we argue that a practical BCI must provide more than two commands to the user.

The remaining sections are organized as follows. In Section 2, we describe the methods and algorithms used to acquire, process, and classify the data. Results of the classification experiments are discussed in Section 3, and conclusions are presented in Section 4.

II. METHODOLOGY

A. EEG Data Acquisition

All the data used in this paper were obtained previously by Keirn and Aunon [18]. The data are available online at http://www.cs.colostate.edu/~anderson. Data were acquired using the following procedure: The subjects were seated in an Industrial Acoustics Company sound controlled booth with dim lighting and noiseless fans for ventilation. An Electro-Cap elastic electrode cap was used to record from positions C3, C4, P3, P4, O1, and O2, shown in Fig. 1 and defined by the 10-20 system of electrode (sensor) placement [7]. The electrodes were connected through a bank of Grass 7P511 amplifiers and bandpass filtered from 0.1–100 Hz. Subjects' EEG were recorded while they performed five mental tasks. Data were recorded at a sampling rate of 250 Hz with a Lab Master 12 bit A/D converter mounted in an IBM-AT computer. Eyeblinks were detected by means of a separate channel of data (EOG channel), and subjects were asked not to make any physical movements to avoid any unwanted muscle artifacts.

In our work, the data from one subject performing three mental tasks were analyzed, the three tasks were:

(1) Baseline task. The subjects were asked to relax as much as possible.

(2) Letter task. The subjects were instructed to mentally compose a letter to a friend or relative without vocalizing.

(3) Math task. The subjects were given nontrivial multiplication problems, such as 49 times 78.

Data were recorded for 10 seconds during each task and each task was repeated five times per session. Subjects attended two sessions recorded on separate weeks, resulting in a total of ten trials for each task. For our analysis, we chose only three trials per session as experiments involving five trials per session in previous work were made possible by parallel hardware [5]. With a 250 Hz sampling rate, each 10 second trial produces 2,500 samples per channel; these are divided into half-second segments for preprocessing and classification, producing 20 segments per trial with each segment consisting of 125 data points. All channels were concatenated including the EOG channel as rows of a matrix, where each half-second EEG segment produces a 7x125 matrix of voltage measurements.

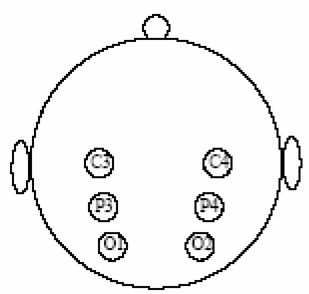

Fig. 1, Placement of the electrodes according to the 10-20 system [4].

B. EEG Preprocessing

Recognition and elimination of the artifacts added to the EEG signal during the recording sessions is an essential task to facilitate accurate classification. The most corruptive of the artifacts is due to eye blinks because it produces a high amplitude signal called electrooculogram (EOG) that can be many times greater than the EEG signals of interest (see Fig. 2).

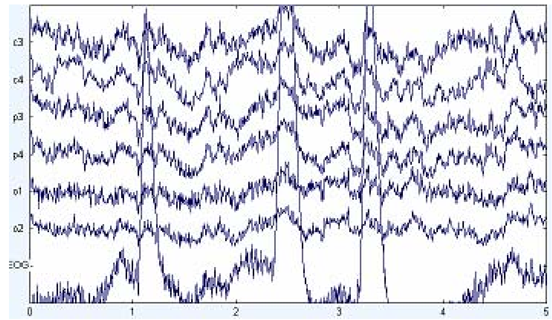

Fig. 2: 5 seconds of EEG contaminated by eyeblink artifacts

One possible solution to this problem is the rejection of contaminated EEG segments, based on either automatic or visual detection; however, this may lead to discarding valuable periods of EEG recordings especially when limited data are available or when blinks occur too frequently.

In this paper our objective is to remove eyeblinks while preserving the essential EEG signal, we propose a method for automatic detection and correction of eyeblink artifacts based on spatial manipulation of ICA components. ICA is a relatively recent method for Blind Source Separation (BSS) [2] that attained a wide attention and growing interest after A.J. Bell and T.J. Sejnowski published their approach based on the infomax principle [1] in the mid 90's. Recently, Lee et al. [16] extended the ability of the infomax algorithm to perform blind source separation on linear mixtures of sources having either sub or super-Gaussian distributions. In our work we implemented ICA using the efficient extended Infomax algorithm. For specific details of the algorithm, the reader is referred to [16].

Independent Component Analysis (ICA)

Independent component analysis (ICA) is a method for solving the blind source separation problem. The use of ICA for blind source separation of EEG data is based on a plausible assumption that EEG data recorded at multiple scalp sensors are linear sums of temporally independent components arising from spatially fixed, distinct or overlapping brain or extra-brain networks [9]. The goal of ICA is to recover statistically independent sources given only sensor observations that are unknown linear mixtures of the unobserved independent source signals. In contrast to correlation-based transformations, ICA reduces the statistical dependencies of the signals, attempting to make the signals as independent as possible which makes ICA capable of isolating artifactual components from EEG data since they are usually independent of each other.

Let $x(t)$ represent n-dimensional vectors which correspond to the n continuous time series from the n EEG channels. Then $x_i(t)$ corresponds to the continuous electrode readings from the i^{th} EEG channel. Because various underlying sources are summed via volume conduction to give rise to the scalp EEG, each of the $x_i(t)$ is assumed to be an instantaneous linear mixture of n unknown components or sources $s_i(t)$, via the unknown mixing matrix A.

$$x(t) = A\ s(t) \tag{1}$$

ICA uses the EEG measurement x(t) and nothing else to generate an unmixing matrix W that approximates A^{-1}, to recover a version, $\hat{s}(t)$, of the original sources s(t),

$$\hat{s}(t) = W\ x(t) \tag{2}$$

Sensor space projections, which indicate the effect of a given component, in isolation, on all sensors are given by the estimated mixing matrix,

$$\hat{A} = W^{-1} \tag{3}$$

Where each column of W^{-1} represents a spatial map describing the relative projection weights of the corresponding recovered source at each of the EEG channels: the first column of W^{-1} is the spatial map of the first independent component and so on. The columns of W^{-1} matrix are called the scalp topography (the scalp map) of the components which can be visualized in Fig. 3. These scalp maps provide evidence of the components' physiological origin (e.g. eye activity should project mainly to far frontal sites).

The proposed ICA-based method for eyeblink artifacts correction

The use of ICA for removing artifacts from EEG signals is not new, However, in most previous work, it was used as an offline preprocessing method mainly for two reasons; the first one is related to the manual selection of artifactual components by visual inspection which is impractical, time consuming and the second is that ICA sufficient amount of data to be able to separate the independent sources. Therefore our goal was not only avoiding artifact rejection and minimizing data loss but also we wanted to correct EEG data automatically, as a very important step for an application such as BCI that require online and real time operation.

Our ICA-based preprocessing technique consists of three steps:

(1) ICA decomposition. This is an offline training phase to obtain the unmixing matrix W, The idea is to train ICA on a whole 10-second trial and use the unmixing matrix W as a spatial filter to separate seven independent components from EEG data (the number of independent components is equal to the number of channels [16]).

We used Infomax (ICA), implemented with Matlab 6p5 in EEGLAB interface [3], the 10-second trial was chosen arbitrarily, we chose the fifth trial of our subject's letter task shown in Fig. 2 (we note that this trial is not one of our six trials chosen for classification).

(2)Eyeblink components identification. The preprocessing phase actually begins by applying the spatial filter W obtained from the training phase to the EEG data segments to separate the independent components or brain sources. Independent sources related to eyeblinks must then be identified. The identification process is based on the components' scalp topographies. The scalp topography of each component provide evidence of its physiological origin. An eyeblink component's scalp map has a strong far-frontal projection [8] (see Fig. 3). A simple rule for eyeblink components identification can be developed based on this fact and then these components must be removed from the components matrix $\hat{s}_j(t)$ in order to reconstruct a clean EEG segment.

Denoting the j^{th} column of W^{-1} by W^{-1}_j it represents the intensity distribution at each electrode (i.e. the scalp map) of the corresponding component $\hat{s}_j(t)$, where j = 1,..,7. w^{-1}_{ij} denote the ith element in W^{-1}_j where i = 1,..,7 as the number of sources is equal to the number of channels.
Then the eyeblink component identification rule is:

$$\begin{aligned} &\text{If } \max (W^{-1}_j) = w^{-1}_{7j} \\ &\text{Then } s'_j(t) = 0 \\ &\qquad W^{-1}_j = 0 \\ &\text{Else } s'_j(t) = \hat{s}_j(t) \end{aligned} \tag{4}$$

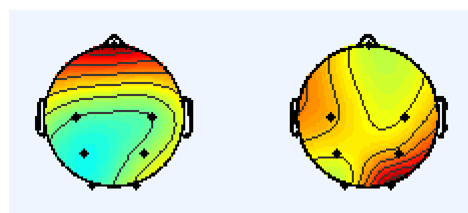

Fig. 3: scalp maps of 2 ICA components, the left corresponding to an eyeblink and the other is brain related.

We must note that by max $(W^{-1}_j) = w^{-1}_{7j}$ we mean that the ocular activity contributes most to this component activity regardless of the sign of that activity. An additional verification step is implemented in order to insure that the identified components are related to eyeblinks. The actual projection of the component onto the original data channels is computed by the outer product of the component activation with the corresponding scalp map (or inverse matrix column), the verification rule is:

An eyeblink component must project high activity at the EOG channel i^{th} amplitude exceeding 100μV

(3) EEG data correction. After identifying and removing artifactual components, EEG data are reconstructed using the new independent components matrix,

$$x'(t) = W^{-1}\ s'(t) \tag{5}$$

Where s'(t) is the matrix of the recovered sources ŝ(t) with rows representing artifactual sources set to zero and x'(t) is the corrected EEG segment (see Fig. 4).

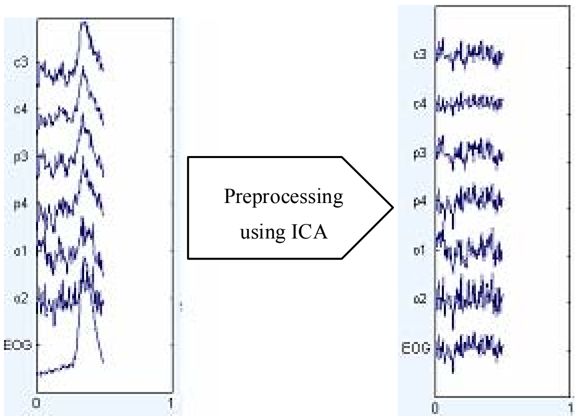

Fig. 4: half second EEG window of baseline task before and after ICA-based eyeblinks correction

C. *Feature Extraction*

Mental task BCI relies heavily on the assumption that different mental tasks can be discriminated by extracting significant features from EEG signals. We implemented three different feature extraction techniques on preprocessed EEG segments: AR modeling [4,5], AR spectral analysis [13] and power differences between four frequency bands [12]. The first is based on time domain features of the EEG segments while the second and the third are based on the frequency contents.

(1) Sixth-order Autoregressive (AR) model: The half second EEG segments were subjected to feature extraction using parametric AR modeling. Parametric modeling is a technique for time series analysis in which a mathematical model is fitted to a sampled signal. A real-valued, zero-mean, stationary, autoregressive process of order p is given by

$$x(n) = -\sum_{k=1}^{p} a_k x(n-k) + e(n) \qquad (6)$$

where p is the model order, x(n) is the signal at the sampled point n, ak are the real-valued AR coefficients and e(n) represents the error term independent of past samples. The term autoregressive implies that the process x(n) is seen to be regressed upon previous samples of itself, it assumes that the value of the current sample can be predicted as a linearly weighted sum of the p most recent sample values. The error term is assumed to be a zero-mean noise with finite variance. We used Burg's method [10] to estimate the AR coefficients. The method is more accurate as compared to other methods such as Levinson–Durbin as it uses the data points directly. Furthermore, the Burg algorithm uses more data points by minimizing both forward error and backward error [13].

In computing the AR coefficients, order six was used based on the suggestions of Keirn, Aunon [18] and Anderson et al.[4,5,6]. Therefore, we had six AR coefficients for each channel, giving a total of 36 features for each EEG half segment of a mental task.

(1) Autoregressive (AR) Spectral Analysis: The goal of spectral estimation is to describe the distribution (over frequency) of the power contained in a signal. Methods for spectral estimation can be classified into parametric methods and FFT-based methods (nonparametric methods). Although the computationally efficient FFT with periodogram method is commonly applied for EEG spectral analysis, we chose to estimate the power spectral density using the parametric AR model because parametric methods can yield higher resolutions than nonparametric methods in cases when the signal length is short.

After the 6th order AR coefficients of the half second EEG segments were estimated using Burg algorithm, we obtained the power spectral density (PSD) values from 0 to 50 Hz for each channel giving a total of 750 features for each EEG half segment of a mental task.

(3) Band Powers and power differences: The existence of particular brain rhythms: delta (1-3 Hz), theta (4-7 Hz), alpha (8-13 Hz), beta (14-29 Hz) and gamma (30-50 Hz) underlying EEG activity naturally motivates the investigation of spectral characteristics of the EEG in the corresponding frequency bands [11].

For each half-second EEG window the spectral power for each band was computed after filtering, delta and theta bands were combined since their frequency range is small, this gave 4 power values for each channel and a total of 24 power values for each EEG window, next, power difference in each frequency band for all channel combinations was computed using

$$Power_{diff} = \left[\frac{P_1 - P_2}{P_1 + P_2}\right] \qquad (7)$$

Where P_1 is the power in one channel and P_2 is the power in another channel in the same spectral band this gave 120 spectral power differences. Concatenating these power differences to the 24 power values gave a total of 144 features for each half-second EEG window. Power differences were proposed by Palaniappan et al [12] to express the relative 'shift' in activity between 2 sites or from one frequency band to another instead of asymmetry ratios used in [18].

D. *Support Vector Machine Classifier*

The classifier implemented in our work is an RBF SVM. Support vector machine (SVM) has been widely used in pattern recognition and regression due to its computational efficiency and good generalization performance. The best way to apply SVMs to the multiclass case is an ongoing research problem, we trained our classifier using Platt's sequential minimal optimization (SMO) and DAGSVM algorithms [19]–[21]. More details on SVM can be found in [20].

The training data was selected from the full set of six trials as follows. Four of the six trials were used for the training set, one of the remaining trials was selected for testing and the last trial was used for validation which was used in order to determine the best values of the kernel parameters σ and C to use for the testing. A rigorous parameter selection search method was used were all possible combinations of parameters σ and C, from $2^{\wedge-16}$ to $2^{\wedge16}$ were examined with every possible selection for validation trial. Finally, the remaining five trials were compiled into one set of training data. The experiments were repeated for each of the 30 ways to partition the six trials in this manner and the results of the 30 experiments were averaged to produce the results shown in Table I. This choice of training paradigm is based on earlier results [4,5,6] to limit the amount of overfitting during training.

III. Results

In this section, we present the results of classification experiments for the best kernel parameters' combinations σ and C resulting from , using data from one subject performing the baseline, mental multiplication and Letter tasks. The parameters are denoted by a hyphenated pair of numbers representing the combination. We compare the classification performance using three representations of the preprocessed EEG signals using our ICA-based technique: 6th order AR model coefficients, PSD values obtained using autoregressive spectral analysis, and Power diff representation (spectral powers and power differences).

The classification results on test data given in Table 1 represent the percent of EEG test windows classified correctly, out of 60 test patterns (20 patterns for each task corresponding to the test trial), averaged over 30 runs, each run with different combinations of train, test, and validate sets for the best σ – C combinations. The total number of testing feature vectors for our subject was 23400.

TABLE 1: PERCENT OF TEST PATTERNS CLASSIFIED CORRECTLY BY RBF SVM FOR DIFFERENT COMBINATIONS OF σ AND C AND DIFFERENT SIGNAL REPRESENTATIONS.

Feature Extraction Algorithm					
σ - C	AR	σ - C	PSD	σ - C	Power diff
0.0313- 64	67.5	0.0039-16	54.4	0.0039-8	63
0.1250-4	68	0.0039-16	54.4	0.0039-1	61
0.0625-8	70	0.0039-8	56	0.0039-16	61
0.0078-2048	66	0.0039-4	55	0.0039-16	61
0.0313-16	69	0.0039-16	54.4	0.0039-2	61
0.0039-1024	67.5	0.0039-16	54.4	0.0039-8	63

Clearly the best classification accuracy is achieved with the AR method, giving an average accuracy of about 70% for a SVM with 0.0625-8 kernel parameters. Performance with the other representations is significantly lower especially the PSD representation.

To study the effect of our proposed ICA-based preprocessing technique on the classification accuracy, we trained the SVM with the highest classification accuracy for each feature extraction technique on raw EEG segments. Segments contaminated with eyeblinks were rejected as in [4,5,6,12]. The SVMs were trained using the same six-fold crossvalidation procedure. The results given in Table 2 represent the percentage of raw EEG test windows classified correctly versus the percentage of correct EEG segments when ICA is used for preprocessing. Results are averaged over the 30 possible combinations of training, testing and validation sets and the average percentage of rejected segments is 9%.

Table II: Effect of our ICA-based preprocessing technique on percent of test patterns correctly classified

Feature Extraction Algorithm	% Correct Raw EEG	% Correct using ICA-based preprocessing
AR	63.6	70
PSD	51	56
Power diff	58.6	63

Analyzing these results reveals how preprocessing EEG data with ICA enhanced the performance for each feature extraction technique by 6.4%, 5% and 4.4% respectively.

IV. Conclusions

EEG signals recorded from a subject performing three mental tasks were discriminated with almost 70% accuracy using an AR representation of half-second EEG windows and a RBF SVM classifier. Our results strongly suggests that the Parametric AR model-based representation produces significantly more accurate classification than do the frequency-based representations, commonly used in extracting features in EEG research.

A simple technique for removing eyeblinks based on ICA was proposed instead of traditional rejection method to avoid discarding useful data, to enhance classification accuracy and also to offer the user a much more flexible interface which is one of the most important issues in designing a practical mental BCI system. Analysis of EEG artifact removal is inherently difficult since no “clean” reference for the artifacts is available, EOG signals are also contaminated with EEG signals. An indirect way to evaluate the performance of the artifact- removal method is to study their effect on the classification performance.

By visual inspection, our proposed ICA-based preprocessing technique successfully removed eyeblink artifacts from half-second EEG windows which improved classification accuracy of 5% averaged over all the feature extraction methods, Comparison of the proposed technique with other artifact removal approaches is left for future work. After comparing the three EEG representations recommended by previous work [4,5,12], we recommend Parametric AR modeling for representing EEG windows preprocessed using ICA for best classification accuracy noting that this is also the smallest representation which adds another significant advantage .

In future, we plan to expand the work to include more subjects and to study the effect of different feature classification algorithms on the classification accuracy.

REFERENCES

[1] A. Bell and T. J. Sejnowski, “An information maximization approach to blind separation and blind deconvolution,” Neural Computation, vol. 7, pp. 1129–1159, 1995.

[2] A. Cichocki and S. Amari, Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications, JohnWiley and Sons, Ltd., Chichester, England, 2002.

[3] A. Delorme, S. Makeig, “EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis”, J. Neurosci. Methods., vol 134, pp. 9-21, 2004.

[4] C.W. Anderson, S. V. Devulapalli, and E. A. Stolz. “Determining mental state from EEG signals using neural networks”. Scientific Programming, Special Issue on Applications Analysis, vol. 4, no.3, pp.171–183, Fall 1995.

[5] Charles W. Anderson and Zlatko Sijerĉiċ. “Classification of EEG Signals from Four Subjects During Five Mental Tasks”. In Solving Engineering Problems with Neural Networks: Proceedings of the Conference on Engineering Applications in Neural Networks, pp. 407–414, 1996.

[6] C. W. Anderson. “Effects of variations in neural network topology and output averaging on the discrimination of mental tasks from spontaneous electroencephalogram”. Journal of Intelligent Systems, vol.7, pp.165–190, 1997.

[7] Jasper, H. H. "Report of the committee on methods of clinical examination in electroencephalography: The ten-twenty electrodes system of the international Federation." Electroencephalography and Clinical Neurophysiology, vol. 10, no. 2, pp 371-375. (1958).

[8] T. Jung, C. Humphries, T. Lee, S. Makeig, M. McKeown, V. Iragui, T. Sejnowski, “Extended ICA removes artifacts from electroencephalographic recordings” in Advances in Neural Information Processing systems, vol. 10, pp. 894-900, 1998.

[9] T. Jung et al. “Removing electroencephalographic artifacts by blind source separation”. Psychophysiology, vol. 37, no. 2, pp.163-178, 2000.

[10] J. P. Burg, “A New Analysis Technique for Time Series Data”, Modern Spectrum Analysis, New York, IEEE Press, pp. 42-48, 1978.

[11] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, “Brain-computer interfaces for communication and control,” Clinical Neurophysiology, vol. 113, no. 6, pp. 767–791, June 2002.

[12] R. Palaniappan, “Brain Computer Interface Design Using Band Powers Extracted During Mental Tasks”, Proceedings of the 2 International IEEE EMBS Conference on Neural Engineering Arlington, Virginia · March 2005.

[13] R. Palaniappan and P. Raveendran, “Cognitive task prediction using parametric spectral analysis of eeg signals”, Malaysian Journal of Computer Science, vol. 14, no. 1, pp. 58-67, June 2001.

[14] S. Choi, A. Cichocki, H._M. Park and S.-Y. Lee: "Blind source separation and independent component analysis: A review", Neural Information Processing - Letters and Reviews, vol. 6, no.1, pp.1-57, January 2005.

[15] Simon Haykin. Neural Networks, a comprehensive foundation. Prentice Hall ,second edition, 1999.

[16] T.-W. Lee, M. Girolami, and T. J. Sejnowski. “Independent component analysis using an extended infomax algorithm for mixed sub-gaussian and super-gaussian sources”. Neural Computation, pp. 609_633, 1998.

[17] Wolpaw, J.R., Birbaumer, N., Hectderks, W.J., McFarland,D.J., Pecleham, P.H., Schalk, G., Donchin, E., Quatrano, L.A.,Robinson, C.J., Vaughan, T.M. “Brain-Computer InterfaceTechnology: A Review of the First International Meeting,” IEEE Transactions on Rehabilitation Engineering, vol. 8 no. 2, pp. 164-173, June 2000.

[18] Z. A. Keirn, and J. I. Aunon, “A New Mode of Communication Between Man and His Surroundings”. IEEE Transactions on Biomedical Engineering, vol. 37, no. 12, pp. 1209-1214, December 1990.

[19] J. C. Platt, “Fast training of support vector machines using sequential minimal optimization,” in Advances in Kernel Methods-Support Vector Learning, pp. 185–208, 1998.

[20] B. Sch¨olkopf and A. J. Smola, Learning with Kernels. MIT Press, 2002.

[21] J. C. Platt, N.Nello Cristianini, and J. Shawe-Taylor, “Large margin DAG’s for multiclass classification,” in Advances in Neural Information Processing Systems, vol. 12, pp. 547–553 , 2000.