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

Assessment of Preprocessing on Classifiers Used in the P300 Speller Paradigm

H. Mirghasemi, M. B. Shamsollahi, *Member, IEEE*, and R. Fazel-Rezai, *Senior Member, IEEE*

Abstract— **Artifact removal is an essential part in electroencephalogram (EEG) recording and the raw EEG signals require preprocessing before feature extraction. In this work, we implemented three filtering methods and demonstrated their effects on the performance of different classifiers. Bandpass digital filtering, median filtering and facet method are three preprocessing approaches investigated in this paper. We used data set IIb from the BCI competition 2003 for training and testing phase. Our accuracy varied between 80% and 96%. In our work, we demonstrated that the problems of choosing the classifier and preprocessing methods are not independent of each other. Two of our approaches could achieve the 96% accuracy i.e. 31 of 32 characters were predicted correctly. These two approaches have different classifier and different preprocessing method. It means that the performance of each classifier can be enhanced with a specific preprocessing method. In our approach, we used only three electrodes of 64 applied electrodes. Therefore it can noticeably reduce the time and cost of EEG measurement.**

I. INTRODUCTION

A Brain-Computer Interface (BCI) is a system that uses EEG signals as input to enable user to communicate with his environment. P300 speller paradigm is a kind of BCI which use P300 potential to spell the intended character of user. P300 potential is a positive peak which occurred about 300 ms after observing an infrequent event among the other frequent events [1]. In P300 speller user faces a 6 by 6 matrix of characters. Each row or column is flashes and the user makes a selection by counting how many times the row or column containing the desired character flashes. The flashing of each row or column for each character is repeated 15 times. In response to flashing of each row or column corresponding to desired character, the P300 peak is occurred. So, if we are able to detect P300 potentials, we can determine the desired character and spelling will be possible [2].

Manuscript received April 24, 2006.

R. Fazel-Rezai is with the Brain Image and Signal Processing (BISP) Laboratory, Department of Electrical and Computer Engineering, University of Manitoba, Winnipeg, MB R3T 5V6 Canada (corresponding author: phone: 1-204-474-9490; fax: 1-204-261-4639; email: fazel@ee.umanitoba.ca).

H. Mirghasemi is with the BDP laboratory, Department of Electrical Engineering , Sharif University of Technology, Tehran, Iran (hmirghasemi@ee.sharif.edu)

M. B. Shamsollahi is with the BDP laboratory Department of Electrical Engineering , Sharif University of Technology, Tehran, Iran (mbshams@sharif.edu)

II. METHODS

A. Data

In this work, we used the dataset from BCI 2003 competition. These dataset contains three sessions. In first two sessions, the signals were labeled i.e. the signal which belongs to desired characters are labeled by 1 and the other signals by 0. Therefore, these two sessions can be used for training data and similar to competition conditions, the last session is used as testing dataset.

Data were recorded from 64 electrodes. However we did not use all of them. We limit our methods to only three electrodes. The channels Cz, Pz and Fz were selected. The location of channels is defined based on 10-20 standard [3]. We used only three electors because our goal was to assess the effect of preprocessing methods on the performance of classifiers. It is clear that if we increase the number of channels, we can achieve the higher accuracy.

B. Preprocessing

To reduce the effect of noise, we should implement preprocessing methods on EEG signals, Since EEG signals are low in amplitude, noise reduction is vital. For this paper, we used three preprocessing methods. The first method, bandpass filtering, is a conventional method in BCI [3][4]. Other two methods, median filtering and facet approach, have not been used for P300 preprocessing in BCI research. They are usually used in image processing. After preprocessing, all the data were normalized to an interval of [-1 1].

C. Feature reduction

Since P300 component has a special temporal pattern, therefore we do not need to use other features such as band power, power spectrum and auto-regressive (AR) model parameters. Therefore, the samples of EEG signals are used as feature. To reduce the dimension of extracted feature, i.e. samples of filtered signal, principal component analysis (PCA) was used. We decreased the dimension of feature input from 144 (number of samples of each epoch during 600 ms after stimuli onset) to 21 using PCA.

D. Classification

We used four classifies, Fisher Linear Discriminant (FLD), Kennel Fisher Discriminant (KLD), and two types of neural networks (NNs) with 2 layers of perceptron (2LPNN) and 3 layers of perceptron (3LPNN). These classifiers are explained in Section IV. To use them, first they should be trained. Therefore, we should have the data labeled. Typically, training data are labeled by 1 and 0 to show the presence or absence of P300 potentials, respectively. To make these data suitable for our programs, we changed the label to 1 and -1 in which -1 indicates the absence of P300 potentials. After training the classifiers, we implement our trained classifiers to testing dataset (i.e. the last session). Applying classifiers to each signal, we obtain a number between 1 and -1. If the output of classifier is near to 1, it shows that this signal has more probability to contain P300 potential. Using this fact, we apply classifier to signals of each row or column for each character. The row and column which have more positive output are selected as desired row or column. Some of our classifiers need parameters which should be determined using optimization. These classifiers include: KFD and NNs. To determine these parameters, we applied the trained classifier to part of our training data for each step. In each step, we changed the parameters and then compared the result of total steps. The parameters that are related to step showing the best performance were chosen as desired parameters.

III. PREPROCESSING METHODS

The following preprocessing methods were applied to P300 data set to reduce noise.

A. Bandpass Digital Filtering

A conventional method of preprocessing method in P300 speller paradigm is bandpass filtering [3][4]. The raw signals are filtered with a digital bandpass filtering [0.5 30] Hz. To do this, we first transferred the raw signal in frequency domain. Then we removed the components of signals which do not belong to this interval. Finally, signals were transferred back to time domain.

B. Median Filtering

Let $\Gamma = \{x_1, ..., x_{2p+1}\}$ be a segment of input signal. In median filtering, the central sample in filtered signal is replaced by the median value of $\{x_1, ... x_{2p+1}\}$. This filtering can reduce the random variations in EEG signals. For our paper, we chose p equal to 5. Unlike to bandpass filtering, this approach is related to time domain properties of signal.

C. Facet Method

In image processing methods, it is assumed that the pixel grid is a discrete approximation of an underlying intensity surface. The facet model assumes that the true surface is a continuous piecewise and that image is a noisy sampled version of it. The facet model is based on the minimization of the error between the image thought of as a piecewise continuous level intensity surface and observed data [5].

The facet model idea can be used in EEG signal denoising. We assume that the true EEG pattern in each segment with length of $2p+1$ is a line and we then modify the p^{th} value of the segment to reduce the difference between signal and its linear model [5]. Similar to median filtering, we choose p equal to 5.

The neighbors of the central sample in segment are used to evaluate the facet model. Modification can be done by comparing the difference between observed signal at central point of segment and its corresponding value in linear model. If this difference is higher than a defined threshold, we modified the observed signal at the central point of segment with the average of observed signals in the segment. To define a threshold, we can use statistical properties that are evaluated via facet model. For example, the average error between the neighborhood samples and their corresponding values in linear model can be used to define a threshold.

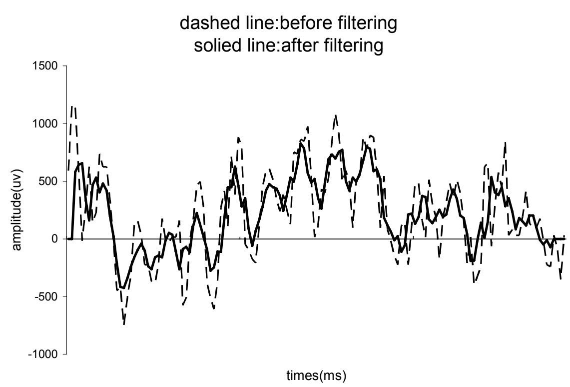

Fig1. The signal before and after facet approach filtering

As we can see, in this approach the variance of filtered signal in P300 duration is fewer than that of original signal.

The Median filtering and facet approach filtering has not been used in preprocessing of the P300 data sets in BCI research. These two filtering approaches are usually used in image processing.

IV. CLASSIFIERS

The following classifiers were used in this paper:

A. Fisher Linear Discriminant (FLD)

Let the input is a set $\Gamma = \{(x_1, y_1), ...(x_l, y_l)\}$ of binary-labeled $y_i \in \{-1, 1\}$ training vector. The x_i is the EEG sample and y_i is the correspond label of this sample. The y_i is 1 when the EEG signal has P300 component. The class separability is defined as [6]:

$$F(w) = \frac{\langle w.S_B w\rangle}{\langle w.S_w w\rangle} \tag{1}$$

where S_B is the between-class scatter matrix:

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T \tag{2}$$

and S_W is the within class scatter matrix defined as:

$$S_W = S_1 + S_2 \qquad S_y = \sum_{i\in Y_y}(x_i - \mu_y)(x_i - \mu_y)^T, y \in \{1,2\} \tag{3}$$

In the FLD, the parameter vector w of the linear discriminant function $f(x) = \langle w.x\rangle + b$ is determined to maximize the class separability. A classical approach is to set the w equals to $S_w^{-1}(m_1 - m_2)$. To find linear discriminant function $f(x) = \langle w.x\rangle + b$, we should determine b from the equation $f(m_1) = -f(m_2)$.

B. Kernel Fisher discriminant (KFD)

To solve nonlinear separable data, we can use the kernel methods to map the feature space to higher dimension feature space. By this method, we replace the linear discriminant function with a non linear discriminant function:

$$f(x) = \sum_{i=1}^{N} \alpha_i K(x, x_i) + b \tag{4}$$

In this work, we used the Gaussian kernel:

$$k(x, x_i) = \exp\left(\frac{-\|x - x_i\|^2}{2\delta^2}\right) \tag{5}$$

C. Neural Network (NN)

Neural network is a common classifier in pattern recognition problems. In our work, we use the one of the simplest type of neural networks, multi-layer perceptron. Multi-layer perceptron is extended version of perceptron. This extended version is more suitable for non separable data. In our work, we use both two-layer and three-layer perceptron to illustrate the effect of increasing the number of layers [7].

In the 2LPNN, tan-sigmoid transfer and linear transfer function were used for first hidden layer and second layer. To train the network, we used gradient descent learning approach. Also, to improve the performance of learning, we allowed change in the learning rate during learning.

In the 3LPNN, we used tan-sigmoid transfer function for first two layers and linear transfer function was used for last layer. The same learning approach was chosen for 3LPNN.

V. Results

First, we separated the training dataset into two groups. One time, the first session was used as training data and the second session was used as testing dataset. In second time, we replaced the roles of first session and second session. Each time, the classifiers were trained with one of these sessions and trained classifier was applied to the other session. The accuracy of these two steps was averaged as shown in table I and II.

As shown in tables I and II, the best accuracy of the FLD classifier was obtained when feature reduction and bandpass filtering was used. In the KFD, the higher accuracy is achieved if we use bandpass filtering as preprocessing method. Also, the accuracy of both of 2LPNN and 3LPNN are maximized when feature reduction is not used and facet approach is used for denoising.

Table I. Accuracy of sub testing data (PCA is not used)

	Bandpass	**Median**	**Facet**
FLD	**70.59**	**64.70**	67.65
KFD	**82.35**	**70.59**	82.35
2LPNN	**79.41**	**82.35**	82.35
3LPNN	64.70	58.82	79.41

Table II. Accuracy of sub testing data (PCA is used)

	Bandpass	**Median**	**Facet**
FLD	**82.35**	**79.41**	73.53
KFD	**82.35**	**82.35**	76.47
2LPNN	**58.82**	**67.64**	73.53
3LPNN	61.76	73.53	61.76

We applied trained classifiers (trained with first two sessions) on the last session. The accuracies of different classifiers and preprocessing methods when they applied to samples of signal are shown in table III. Table IV illustrates the accuracy of these classifiers and preprocessing methods when principal components of signal were used.

TABLE III. ACCURACY (%) OF TESTING DATA
(PCA WAS NOT USED)

	Bandpass	Median	Facet	Raw
FLD	**90.32**	**83.87**	**80.64**	90.32
KFD	**90.32**	**87.09**	**83.87**	83.87
2LPNN	**93.54**	**93.54**	**96.77**	87.09
3LPNN	83.87	93.54	90.32	83.87

TABLE IV. ACCURACY (%) OF TESTING DATA
(PCA WAS USED)

	Bandpass	Media n	Facet	Raw
FLD	**96.77**	**90.32**	**87.09**	93.54
KFD	**93.54**	**93.55**	**93.54**	90.32
2LPNN	**93.54**	**96.77**	**96.77**	90.32
3LPNN	87.09	83.87	90.32	87.09

VI. CONCLUSIONS

Considering all four result tables, we can conclude that for different classification methods, specific preprocessing algorithm is appropriate. They are summarized in Table V. In this table, for each classifier, the best preprocessing algorithm is proposed.

TABLE V. PROPOSED PREPROCESSING ALGORITHM FOR DIFFERENT CLASSIFICATION METHODS

	Feature Reduction	Preprocessing
FLD	**PCA**	Bandpass
KFD	**PCA**	Bandpass
2LPNN	**None**	Facet Approach
3LPNN	None	Facet Approach

Although bandpass filtering is used widely for preprocessing, we can conclude that the performance of neural network is enhanced with facet approach and median filtering. In FLD and KFD, these two preprocessing methods show low performance. In these cases, the accuracy of using facet approach and median filtering is even lower when we do not use any preprocessing method.

Two algorithms showed the highest performance. These two algorithms are: 1) FLD with PCA feature reduction and bandpass filtering 2) 2LPNN facet approach for preprocessing. The first algorithm is more encouraging for several reasons: 1) FLD does not require optimization 2) PCA reduces the feature dimension, so the time of processing decreases. Also, this approach has high accuracy and it is not time consuming, because training and testing of FLD can be done quickly and data can be measured from only three electrodes. These features make this approach a promising tool for online P300-based BCI that requires accuracy, and high transfer rate.

REFERENCES

[1] J. R. Wolpaw, N.Birbaumer, D. J. MackFarland, G. Pfurstcheller and T. M. Vaugan, “Brain Computer Interface for communication and control”, *Clin. Neurophyiol.,* vol.11 3, pp.767-791, 2002.

[2] L. A. Farwell and E. Donchin, "Talking of top of your head: Toward a mental prostethis utilizing event-related brain potentials, *Electroencephalogr and Clin Neurophy,* vol. 70, pp.510-523, 1988.

[3] M. Kaper, P. Meinicke, U. Grosskathoefer, T. Lingner, H. Ritter, "Support Vector Machines for the P300 speller paradigm," *IEEE Trans. Biomed. Eng.* vol. 51, no. 6, pp. 1073-1076, 2004.

[4] Xu N, Gao X, Hong B, Miao X, Gao S and Yang F, "BCI Competition 2003_Data Set IIb: Enhancing P300 Wave Detection Using ICA-Based Subspace Projections for BCI Applications," *IEEE Transaction On Biomedical Eng,* vol. 51, pp. 1067-1072, 2004.

[5] T. Seemann, "Digital Iamge Processing using Local Segmentation" , Ph.D. thesis, School of Computer Science and Software Engineering, Faculty of Information Technology March University Australia , pp. 37-39, April 2002.

[6] S. Mika, G. Ratsch , and K.-R. Muller, “Fisher Discriminant Analysis with Kernels," *Neural Networks for Signal Processing,* vol. 4, pp. 41-48, 1999.

[7] S. Theodiris and K. Koutrombas, *Pattern Recognition.* Academic Press, pp. 77-88, 1999