A P300-based BCI Classification Algorithm using Median Filtering and Bayesian Feature Extraction

Xiao-ou Li [#1], Feng Wang [#2], Xun Chen [*3], Rabab K. Ward [*4]

[#] *Shanghai Medical Instrumentation College*
School of Medical Instrument and Food Engineering, University of Shanghai for Science and Technology
101 Yingkou, Shanghai, China, 200093
[1 2] lixo@smic.edu.cn

[*] *Department of Electrical and Computer Engineering, University of British Columbia*
2332 Main Mall, Vancouver, BC, Canada, V6T 1Z4
[3] xunchen@ece.ubc.ca, [4] rababw@ece.ubc.ca

Abstract— **A brain computer interface (BCI) system translates a person's brain activity into useful control or communication signals. In this paper, an effective P300-based BCI identification algorithm using median filtering and Bayesian classifier is proposed to improve the classification accuracy and computation efficiency of P300-based BCI. Median filtering is firstly applied to remove noises and Bayesian Linear Discriminant Analysis (BLDA) is then employed for classification. Testing on the P300 speller paradigm in dataset II of 2004 BCI Competition III, we show that a 90% average classification accuracy can be achieved and the highest accuracy is 100%. The proposed method is also computationally efficient and thus it represents a practical implementation for man-computer communication control, especially for on-line applications.**

I. INTRODUCTION

Brain Computer Interface (BCI) is a complex intelligence system that makes it possible for people to communicate with their environment through EEG signals. Several types of brain activity reflected in EEG signals can be applied in BCI, such as alpha waves, evoked potential, P300, mu and beta rhythms, slow cortical potential, etc. BCI can be implemented by detecting the brain responses for sensory stimuli, attention direction and motion intention [1]. Compared with other types of EEG-based BCIs, the advantage of P300-based BCI is that less training is required, since P300 is one of the brain's built-in functions. The P300 signal is a positive deflection in central brain locations, appearing approximately 300ms after the presentation of target stimuli, and it is related to the cognitive psychology. The most widely used P300-based BCI is the P300 speller paradigm, where the goal is to spell desired characters through the P300 signals. If different P300 signals evoked by corresponding characters can be distinguished in the experiment, the spelling task can be realized through the computer simulation [2, 3].

Since the recorded P300 signals also contain some other brain activities, EMG, ECG and other artefacts, resulting in a low SNR, detection of target stimuli from a single trial is difficult. To improve the signal quality, the epochs are generally averaged over many trials. A required large number of repetitions decreases the spelling efficiency. Conventional filtering methods are often employed to extract clear P300 signals, but they may also filter the desired signals [4]-[6]. Median filtering is a nonlinear filtering method that can remove the interference pulse and meanwhile keep the local characteristic of the original signal effectively, and thus it was suggested to as a suitable preprocessing method to deal with pulse noises in P300 signals [7].

The key components for a BCI system are feature extractor and classifier that can distinguish different brain function patterns. Compared with nonlinear classification algorithms, linear classifiers have the desired properties such as simplicity and outstanding self-learning ability. Linear classifiers can yield classification accuracy on the range approximately 80%-90% [2]. Bayesian Linear Discriminant Analysis (BLDA) is an extension of Fisher's Linear Discriminant Analysis (FLDA) and actually is Bayesian regression under the evidence framework. It can prevent overfitting and reduce computation complexity [8].

Though both median filtering and Bayesian classifier have been shown to be attractive for P300-based BCIs, to other best knowledge, there is no work yet investigating the performance gain by combining both of them. In this paper, a P300 identification algorithm based on median filtering and Bayesian classifier is proposed. We believe that combining a good preprocessing algorithm with an efficient classifier will enhance the overall classification performance. Compared with BLDA combined with the Butterworth bandpass filtering [8], the proposed method can yield a higher classification accuracy. The proposed method can achieve fast convergence, and thus it is promising for on-line BCI applications [9].

II. METHODS

The block diagram for the proposed P300-based BCI is shown in Fig. 1. We now describe the details as follows.

A. Median Filtering

The length of the filtering window is defined as $n = 2k + 1$ (odd) or $n = 2k$ (even), the signal length is N, and the observed data are $s_1, s_2, \cdots, s_N$, and $N > n$. When the window moves in P300 signal sequences, the output $med(s_i)$ of the median filter is

$$med(s_i) = \begin{cases} s_{(k+1)} & n = 2k+1 \\ [s_{(k)} + s_{(k+1)}]/2 & n = 2k \end{cases} . \quad (1)$$

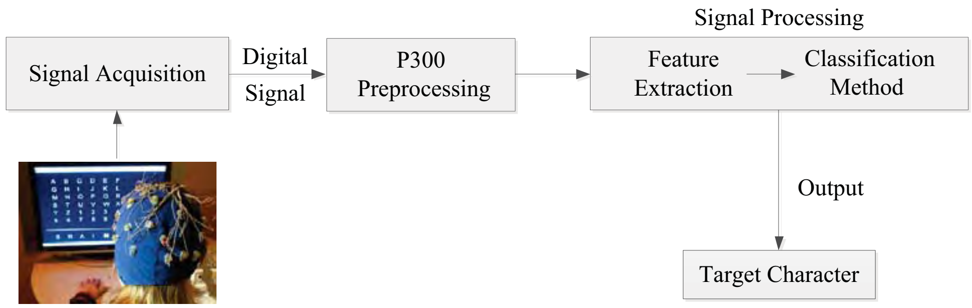

Fig.1. The block diagram for the proposed P300-based BCI

In Equation (1), $s_{(k)}$ is the k-th maximum in the $2k+1$ (or $2k$) observed data. It tends to treat the median value in the $2k+1$ (or $2k$) observed data as the output. For example, the length of the filtering window is 5, the observed data are {2, 1, 3, 1.5, 2.5}, then the median filter output is 2.

If the width of a pulse signal is greater than or equal to $(k+1)$, with the window length $n = 2k+1$, the signal will be preserved in the filtered P300 signal sequences, otherwise it will be removed. This is the main property of the median filtering for removing pulse noises and preserving detail signals. In the meanwhile the filter smoothes discrete pulse signals whose duration is less than half of the window length, e.g. it can be used to process spike and sharp waves in P300. After preprocessing the original signals using the above median filter, we now obtain the filtered P300 signals, represented by the vector $\boldsymbol{y}$. $\boldsymbol{y}$ will be used in BLDA for classification in Section II.B.

B. Bayesian Linear Discriminant Analysis

BLDA is an extension of Fisher's linear discriminant analysis [10]. It performs the regression computation in a Bayesian fashion, and the regularization is used to prevent overfitting in high dimensional and noisy signals. In Bayesian regression, an assumption is that the target y is linearly related to the vector $\boldsymbol{x}$ with additive white Gaussian noise n, i.e. $y = \boldsymbol{w}^T\boldsymbol{x} + n$.

The likelihood function for the weight $\boldsymbol{w}$ is

$$p(\boldsymbol{D}|\beta,\boldsymbol{w}) = \left(\frac{\beta}{2\pi}\right)^{\frac{N}{2}}\exp\left(-\frac{\beta}{2}\|\boldsymbol{X}^T\boldsymbol{w}-\boldsymbol{y}\|\right), \quad (2)$$

where $\boldsymbol{D}$ denotes the pair $\{\boldsymbol{X}, \boldsymbol{y}\}$, $\boldsymbol{X}$ is the matrix obtained from the training vectors, $\boldsymbol{y}$ is the filtered P300 signals, β is the inverse variance of noise, and N means the sample size.

In [10], the prior distribution of $\boldsymbol{w}$ is

$$p(\boldsymbol{w}|\alpha) = \left(\frac{\alpha}{2\pi}\right)^{\frac{d}{2}}\left(\frac{\varepsilon}{2\pi}\right)^{\frac{1}{2}}\exp\left(-\frac{1}{2}\boldsymbol{w}^T\boldsymbol{I}'(\alpha)\boldsymbol{w}\right), \quad (3)$$

where the regularization square

$$\boldsymbol{I}'(\alpha) = \begin{bmatrix} \alpha & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \varepsilon \end{bmatrix}_{(d+1)\times(d+1)},$$

where α is the hyperparameter that yields the best prediction performance, d is the vector number, the weight $\boldsymbol{w}$ follows an isotropic, zero-mean Gaussian distribution, and ε is a very small value.

In [10], the posterior distribution of $\boldsymbol{w}$ can be computed using Bayes rule by Equation (2) and (3) as

$$p(\boldsymbol{w}|\beta,\alpha,\boldsymbol{D}) = \frac{p(\boldsymbol{D}|\beta,\boldsymbol{w})p(\boldsymbol{w}|\alpha)}{\int p(\boldsymbol{D}|\beta,\boldsymbol{w})p(\boldsymbol{w}|\alpha)d\boldsymbol{w}}.$$

The posterior distribution is also Gaussian. The mean vector $\boldsymbol{m}$ and covariance matrix $\boldsymbol{C}$ of the posterior distribution satisfy the following forms respectively: $\boldsymbol{m} = \beta(\beta\boldsymbol{X}\boldsymbol{X}^T + \boldsymbol{I}'(\alpha))^{-1}\boldsymbol{X}\boldsymbol{y}$ and $\boldsymbol{C} = (\beta\boldsymbol{X}\boldsymbol{X}^T + \boldsymbol{I}'(\alpha))^{-1}$.

The probability distribution over regression targets for the input prediction vector $\hat{\boldsymbol{x}}$ is

$$p(\hat{y}|\beta,\alpha,\hat{\boldsymbol{x}},\boldsymbol{D}) = \int p(\hat{y}|\beta,\hat{\boldsymbol{x}},\boldsymbol{w})p(\boldsymbol{w}|\beta,\alpha,\boldsymbol{D})d\boldsymbol{w}.$$

The above predictive distribution is also Gaussian with its mean $\mu = \boldsymbol{m}^T\hat{\boldsymbol{x}}$ and its variance$\delta^2 = \frac{1}{\beta} + \hat{\boldsymbol{x}}^T\boldsymbol{C}\hat{\boldsymbol{x}}$. In the P300-based applications, only the mean value of the predictive distribution is used for taking decisions. In our case, there are many trails for each class and the summation of mean values over trials in each class is calculated. The class with maximum summation is then selected as the predictive result [10].

C. Procedure of the Proposed Identification Algorithm

Including the median filtering and BLDA described above, as illustrated in Fig. 1, we can describe the procedure of the proposed identification algorithm for the P300-based BCI as follows:

1) Signal Acquisition: 180 trials of EEG data (digitized at 240Hz) at Cz electrode are used for information analysis. These data include character intensification sequences, stimulus types and stimulus codes.

2) Preprocessing: P300 signals are extended to avoid the boundary effect. The median filtering described in section II.A is applied, where the filter window length is set to be 50 and the length of each trial is 240 (i.e., the analysis window length is 1000ms). The filtered signals are normalized into zero mean and unit variance. They are split into k subsets as the training and testing sets for the cross-validation purpose.

3) Feature Extraction: The parameters β, α, $\boldsymbol{m}$, $\boldsymbol{C}$ in the BLDA classifier are learned based on the training sets.

4) Classification: The testing sets are tested to evaluate the summation of mean value μ over trials. The class with maximum summation is then selected as the target character.

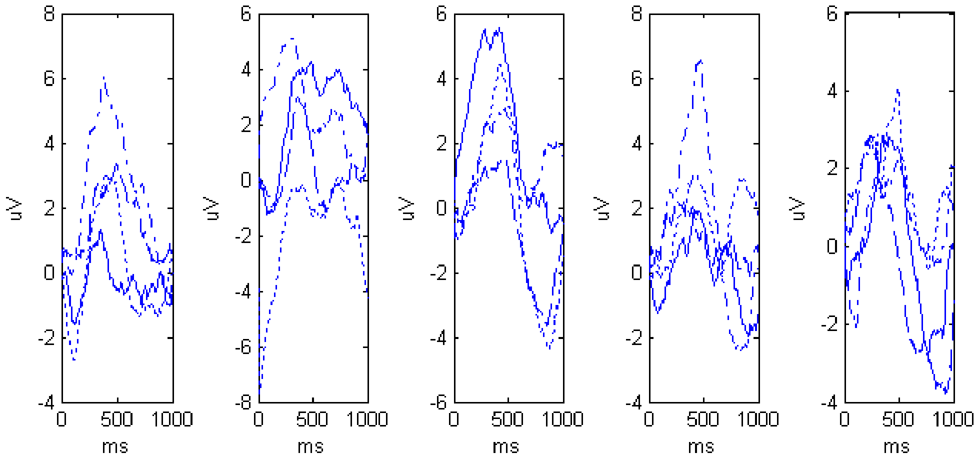

Fig.2. P300 waveforms corresponding to different target characters

TABLE I
THE RESULTS OF CORRECTED IDENTIFIED CHARACTERS WHEN USING DIFFERENT CROSS-VALIDATION SETUPS AND DIFFERENT NUMBERS OF ITERATION STEPS

N-fold cross-validation	Number of iteration steps	Number of sequences														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
4	28	2	1	2	2	2	2	2	2	2	2	2	3	3	2	3
4	23	2	1	3	3	4	4	4	3	3	5	5	5	5	4	5
4	26	1	1	1	2	2	2	2	0	1	1	2	3	3	3	4
4	20	1	1	1	1	2	2	1	2	4	2	3	3	4	3	3
3	29	1	1	3	3	4	4	4	4	4	5	5	5	5	4	5
3	31	1	1	1	2	2	2	2	1	1	1	2	3	3	3	4
3	20	1	1	1	1	2	1	1	2	3	2	3	3	4	3	3
2	18	1	2	2	2	3	2	2	4	4	5	5	5	5	5	5
2	24	1	0	1	2	1	3	2	1	1	1	1	2	2	3	4

III. EXPERIMENT RESULT ANALYSIS

A. Dataset

We test the P300 speller paradigm in dataset II of 2004 BCI Competition Ⅲ[11]. The subjects were asked to focus attention on a 6-by-6 matrix of characters, and the goal was to predict the correct character. Two out of 12 intensifications of rows and columns contained the desired character. The responses evoked by these infrequent stimuli were different from those evoked by the stimuli that did not contain the desired character.

B. Results

The averaged waveforms for target characters “D”, “E”, “O”, “R”, “U” after median filtering were plotted in Fig.2 (from left to right), and we can see that they were smooth and distinct. There were four waveforms for each target character in every figure. For each character, the waveforms share almost the time domain tendency and are time-locked around 300ms. The only difference was the amplitude. For different characters, the time domain waveforms were greatly different. The distinguishing patterns of different characters were useful for further classification.

The sets of 12 intensifications were repeated 15 times for each character epoch. The numbers of the target characters that were correctly identified under different training setups were listed in Table I. Here “0” means that no character was correctly identified and “5” means that all five characters were correctly identified. The data set was divided into k subsets, and k could be 2, 3, and 4, representing three different setups in the k-fold cross-validation. In each setup, one of the k subsets was used for testing and the remaining were used for training. A maximum 100% classification accuracy could be obtained under certain settings (as illustrated in Fig. 4). We noted that the required number of iterations was generally small (e.g. average 24 iterations), as shown in Table I, and thus the proposed method is appropriate for BCI real-time applications.

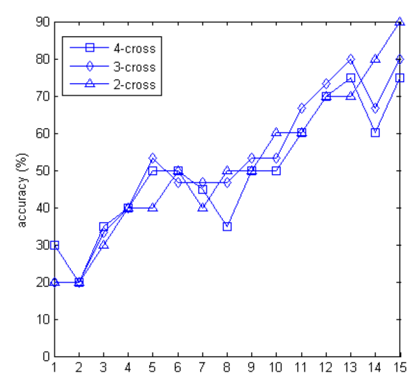

Fig.3. The average classification results when using different number of sequences

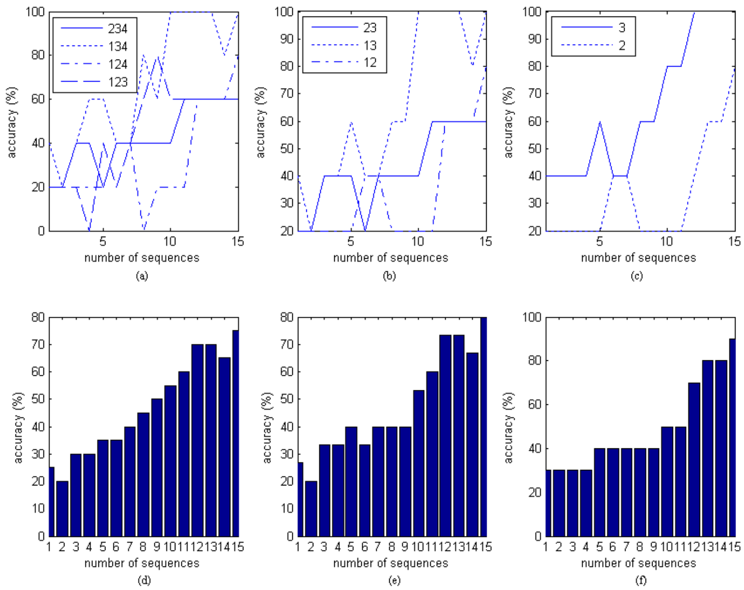

Fig.4. The classification results when combining BLDA with the wavelet analysis

The results of the average classification accuracy when using different number of sequences were shown in Fig. 3. Up to 90% average classification accuracy was achieved by BLDA combined with medican filtering. We noted that the performance of the proposed method was better than the previous study in [11], in which up to 89.5% average classification accuracy was achieved. In [11], BLDA was combined with Butterworth bandpass filtering [8].

We also examined the effect of combining BLDA with wavelet analysis (i.e., with biorthogonal B-spline wavelet and 5-level decomposition). The classification results were shown in Fig. 4. The cross-validation results were presented in Fig. 4(a), (b), (c) and the corresponding average accuracy results as the number of sequences increases were presented in Fig. 4(d), (e), (f). It can be seen that the wavelet-based method leaded to the similar classification accuracy. However, the wavelet-based method is more computationally expensive than the proposed approach.

IV. Conclusion

The proposed P300-based BCI identification algorithm employs both median filtering and Bayesian feature discriminant analysis. Better classification results were obtained. Median filtering can effectively remove pulse noises by using appropriate filtering windows. BLDA is a useful machine learning method, and it can prevent overfitting and reduce computation complexity.

Compared with BCI based on other EEG signals, no special training is required for subjects in the P300-based BCI which facilitates the data collection and real life applications. The proposed method can be a practical implementation for man-computer communication control, especially for on-line applications.

Acknowledgment

This work was supported by Shanghai Municipal Education Commission Project No.12YZ194, by Shanghai Science Technology Research Project No.11441902400 and by Qatar National Research Fund(QNRF) No.NPRP 09-310-1-058.

References

[1] X. Artusi, I. K. Niazi, M. F. Lucas, D. Farina, “Performance of a simulated adaptive BCI based on experimental classification of movement-related and error potentials,” *IEEE Journal on Emerging and Selected Topics in Circuirs and Eystems*, vol 1, pp. 480-488, 2011.

[2] A. E. Selim, M. A. Wahed, and Y. M. Kadah, “Machine learning methodologies in p300 speller brain-computer interface systems,” in *Proceedings of the 24th Nation Radio Science Conference*, 2009.

[3] L. A. Farwell, E. Donchin, “Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials,” *Electroencephalography & Clinical Neurophysiology*, vol 70, pp. 510-523, 1988.

[4] E. Donchin, K. M. Spencer, R Wijensinghe, “The mental prosthesis: assessing the speed of a P300-based brain-computer interface,” *IEEE Transactions on Rehabilitation Engineering*, vol 8, pp. 174-179, 2000.

[5] X. Lei, P. Yang, and D.Z. Yao, “An empirical Bayesian framework for brain computer interfaces,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol 17, pp.521-529, 2009.

[6] B. Rivet, A. Souloumiac, V. Attina, G. Gibert, “xDAWN algorithm to enhance evoked potentials:application to brain-computer interface,” *IEEE Transactions on Biomedical Engineering*, vol 56, pp.2035-2043, 2009.

[7] H. Mirghasemi, M.B. Shamsollahi, and R. R. Fazel, “Assessment of preprocessing on classifiers used in the P300 speller paradigm,” in *Proceedings of the 28th EMBS Annual International Conference*,2006.

[8] U. Hoffmann, A. Yazdani, J.M. Vesin, and T. Ebrahimi, ”Bayesian feature selection applied in p300 brain-computer interface,” in *Proceedings of the 16th European Signal Processing Conference*, 2008.

[9] M. Salvaris, F. Sepulveda, “Wavelets and ensemble of FLDs for P300 classification,” in *Proceedings of the 4th International IEEE EMBS Conference on Neural Engineering*, 2009.

[10] U. Hoffmann, J.M. Vesin, T. Ebrahimi, and D. Karin, “An efficient P300-based brain-computer interface for disabled subjects,” *Journal of Neuroscience Methods*, vol 167, pp.115-125, 2008.

[11] (2004) BCI Competition III Data sets. [Online]. Available: http://www.bbci.de/competition/iii/ index.html.