A comparison of classification methods for recognizing single-trial P300 in brain-computer interfaces *

Xiaolin Xiao[1], *Member, IEEE*, Minpeng Xu[1], *Member, IEEE*, Yijun Wang[3], *Member, IEEE*, Tzyy-Ping Jung[1,4], *Fellow, IEEE*, Dong Ming[1,2], *Senior Member, IEEE*

***Abstract*—P300s are one of the most popular and robust control signals for brain-computer interfaces (BCIs). Fast classifying P300s is vital for the good performance of P300-based BCIs. However, due to noisy background electroencephalography (EEG) environments, current P300-based BCI systems need to collect multiple trials for a reliable output, which is inefficient. This study compared a recently developed algorithm, i.e. discriminative canonical pattern matching (DCPM), with five traditional classification methods, i.e. linear discriminant analysis (LDA), stepwise LDA, Bayesian LDA, shrinkage LDA and spatial-temporal discriminant analysis (STDA), for the detection of single-trial P300s. Eight subjects participated in the classical P300-speller experiments. Study results showed that the DCPM significantly outperformed the other traditional methods in single-trial P300 classification even with small training samples, suggesting the DCPM is a promising classification algorithm for the P300-based BCI.**

***Index Terms*—Brain-computer interface (BCI), single-trial, P300, discriminative canonical pattern matching (DCPM).**

I. Introduction

As a directly link between the brain and the external world, brain-computer interfaces (BCIs) provide a new approach for users of communicating with their environments just by thinking [1]. Among all EEG features, Event-related potentials (ERPs), which are reliable brain responses to particular stimuli, have been widely used for controlling visual BCIs, such as P300 [2], N200 [3], motion visual evoked potential (mVEP) [4] and miniature asymmetric visual evoked potential (aVEP) [5]. The P300 is a specific ERP component that shows a peak waveform at approximately 300±500 ms after a small probability event is encountered [6]. BCIs based on P300 have many practical applications in both clinical and non-clinical fields.

A key component of a P300-based BCI is to discriminate the targets from non-targets. As background EEG signals are non-linear, non-stationary and often hundreds of times larger than P300s, it poses a great challenge to correctly recognize the P300s in single trials. Current P300 BCIs have to collect and average multiple trials to obtain a reliable output, which is time wasting and inefficient. Therefore, it's significant to develop a single-trial classification method for practical applications of P300 BCIs, such as a P300-speller.

Previously, researchers have studied extensively on the classification of single-trial P300s. Specifically, Krusienski *et al.* used a stepwise linear discriminant analysis (SWLDA) to recognize single-trial P300s and the online average accuracy was about 35% [7]. A Bayesian version of regularized LDA, i.e. bayesian LDA (BLDA), was introduced by Hoffmann *et al.* and it achieved an average accuracy about 60% for eight subjects including four disabled patients [8]. Blankertz *et al.* proposed shrinkage LDA (SKLDA), which used a shrinkage technique to regularize LDA and achieved about 70% in the single-trial P300 classifications [9]. Zhang *et al.* introduced the spatial-temporal discriminant analysis (STDA) algorithm, which can be regarded as a multidimensional extension of the LDA and the single-trial accuracy of P300s achieved about 61% [10]. Kaper *et al.* developed an SVM algorithm to classify the single-trial P300s and achieved an average accuracy of 64.56%. Convolutional neural networks (CNN) was first applied by Cecotti and Gräser, which achieved the highest accuracy of 78.19% in single-trial classifications with a large number of training samples [11].

Recently, we have developed a new algorithm, i.e. discriminative canonical pattern matching (DCPM)[5], to recognize the miniature ERPs. It uses the characteristic of EEG symmetry in space to suppress the background common-mode noise and then recognize the canonical patterns of ERPs. DCPM performs well in classifying the miniature asymmetric visual evoked potentials (aVEPs) with even as small as 0.5 microvolts in amplitude. However, it's still unclear what the advantage of DCPM is over other traditional ERP classification methods, and whether DCPM works for other ERPs, such as P300. To this end, this study compared the single-trial classification results obtained by the DCPM and other traditional methods including LDA, SWLDA, BLDA, STDA, and SKLDA for a P300-spellers.

II. Materials and Methods

A. Experiment

The classical BCI paradigm followed Farwell and Donchin in 1988 [12]. Participants sat in front of a monitor in the experiment, and a 6 × 6 letter matrix of characters (see Fig. 1)

* Research supported by the National Key Research and Development Program of China (No. 2017YFB1300302), the National Natural Science Foundation of China (No. 81601565), and Tianjin Key Technology R&D Program (No. 16ZXHLSY00270). *(Xiaolin Xiao and Minpeng Xu contributed equally to this work.)* Correspondence e-mail: richardming@tju.edu.cn.

[1] College of Precision Instruments and Optoelectronics Engineering, Tianjin University, Tianjin, 300072 China.

[2] Academy of Medical Engineering and Translational Medicine, Tianjin University, Tianjin, 300072 China.

[3] State Key Laboratory on Integrated Optoelectronics, Institute of Semiconductors, Chinese Academy of Sciences, Beijing, 100083 China.

[4] Swartz Center for Computational Neuroscience, University of California, San Diego, CA 92093 USA.

978-1-5386-1311-5/19/$31.00 ©2019 IEEE

Figure 1. 6×6 character matrix in P300-speller

was displayed to them. For each character block, after a blank period without flashing stimulus for 2.5s, all rows and columns in this matrix were continuously and randomly lithtened for 100 ms and then extinguished for 75 ms. Everyone in the experiment should pay attention to a specified letter which appeared at the beginning of the block, and then silently counted the number of times of this character lightened. After this character block which was constituted by five rounds, a new specified letter would appear for the next block. A round consisted of six row and six column intensifications. The experiment in our study contained a total of forty character blocks, and 2,400 trials (400 targets and 2,000 non-targets) were collected from each subject.

Eight right-handed healthy subjects (2 females) participated in the study. EEG data were recorded using a Neuroscan Synamps2 system with 64 channels at 1000 Hz, and the positions of electrodes followed the 10/20 system. The reference electrode was placed at the parietal region near Cz, while the ground electrode was placed at the prefrontal region. Each channel was re-referenced to the average of the left and right mastoids firstly, and then down-sampled to 200 Hz. All data were filtered at 1-20 Hz using a band-pass Chebyshev filter. The time window from 0.05s to 0.75s for each trial with 16 channels (F3, Fz, F4, C3, Cz, C4, T7, T8, CP3, CP4, P3, Pz, P4, PO7, Oz and PO8) [13] were used for analysis and classification.

B. Linear Discriminant Analysis

As one of the most popular algorithms for BCI applications, LDA mainly uses hyperplanes to separate two classes of data. Suppose there are two classes (target and non-target) for P300 classification. Let $X_k \in R^{N_D \times N_k} (N_D = N_c \cdot N_t,\ k = 1,2)$ are the training samples of two patterns. N_c represents the number of channels, N_t represents the temporal points, and N_k represents the number of samples in class X_k. Each sample x_i is the concatenation of N_c and N_t . The means μ_k and empirical covariance matrices Σ_k are computed from

$$\mu_k = \frac{1}{N_k} \sum_{i \in X_k} x_i \,, k = 1,2 \tag{1}$$

$$\Sigma_k = \frac{1}{N_k - 1} \sum_{i \in X_k} (x_i - \mu_k)(x_i - \mu_k)^T \,, k = 1,2 \tag{2}$$

$$\Sigma = \frac{N_1}{N_1 + N_2} \Sigma_1 + \frac{N_2}{N_1 + N_2} \Sigma_2 \tag{3}$$

where Σ is the common covariance matrix, and the projection vector w is written as

$$\mathrm{w} = \Sigma^{-1}(\mu_1 - \mu_2) \tag{4}$$

Although LDA is applicable to online BCI system because of the very low computational requirement, it performs poorly in single-trial classification with few training samples. Therefore, some advanced versions of LDA were proposed to address the problem.

C. Advanced Versions of LDA

1) Stepwise LDA

SWLDA has been always employed in the case of small sample size because of its effectiveness in feature dimension reduction [7]. A combination of forward and backward stepwise analysis is used to select suitable features in the discriminant model. The SWLDA uses ordinary least squares regression to weight input features and then predict labels of two classes. The discriminant process begins with no initial features, the most significant input feature which the p-value is less than 0.1 will be added to the model, and then the least significant input features which the p-value is more than 0.15 will be removed from the model. This discriminant process repeats until the discriminant model includes enough features or no other features meet the entry and removal condition [14].

2) Bayesian LDA

Based on the traditional LDA, BLDA constructs the covariance matrix of training samples using the prior knowledge from experiments [15]. BLDA is a probabilistic algorithm on the basis of Bayesian regression, and it performs better than LDA for the conditions with few training sets or strong noise contamination [8]. The Bayesian regression framework poses a more efficient solution aiming at how to choose the hyperparameters. The details of the algorithm can be found in [15].

3) Shrinkage LDA

As a common method for compensating the systematic bias of the estimated covariance matrices and shrinkage parameter for high-dimensional feature spaces [9], the SKLDA has shown its superior performance when using insufficient training samples[10]. It improves the traditional LDA through adjusting the extreme eigenvalues of the covariance matrix towards the average eigenvalue. The poorly estimated covariance matrix Σ_c is computed from

$$\tilde{\Sigma}_c = (1 - \lambda)\Sigma_c + \lambda_c v_c \mathbf{I},\ \lambda_c \in [0,1] \tag{5}$$

$$v_c = \left(tr(\Sigma_c)\right)/(D) \tag{6}$$

where λ_c is a shrinkage parameter, v_c represents the average eigenvalue of Σ_c, D is the dimensionality of the feature space, and $\mathbf{I}$ is an identity matrix. More details of calculating the optimal shrinkage parameter in different directions of shrinkage can be found in [16].

4) Spatial-Temporal Discriminant Analysis

Unlike concatenating temporal points and spatial channels that adopted in the LDA and the three aforementioned advanced methods, STDA constructs a new spatial-temporal feature sample instead of the vectorized feature sample. The STDA adopts discriminant analysis collaboratively both in the spatial and temporal dimensions by constructing two-way feature samples, and then it optimizes features in the spatial and temporal ways alternatingly. This method reduces feature dimensionalities, and hence improves veracity of classifier even with few training samples. Meanwhile, it assists to enhance generalization capability of the trained classifier. The details of the algorithm can be found in [10].

D. Discriminative Canonical Pattern Matching method

The DCPM method is originally proposed in our previous study [5], and it contains three major parts: (1) constructing of discriminative spatial patterns (DSPs); (2) constructing of CCA patterns; (3) pattern matching.

Let $X_k \in R^{N_c \times N_t \times N_s}, (k = 1,2)$ are training samples of two patterns, $Y \in R^{N_c \times N_t}$ is the testing sample, where N_c represents the number of channels, N_t represents the temporal points, and N_s represents the number of samples. The template of the certain pattern is $\hat{X}_k \in R^{N_c \times N_t}$, which is computed by averaging training samples corresponding to pattern k. The covariance matrix can be expressed as

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} \hat{X}_1\hat{X}_1^{\ T} & \hat{X}_1\hat{X}_2^{\ T} \\ \hat{X}_2\hat{X}_1^{\ T} & \hat{X}_2\hat{X}_2^{\ T} \end{bmatrix}. \quad (7)$$

The variances of X_k are

$$\sigma_1^{\ 2} = \frac{1}{N}\sum_{i=1}^{N_s}\left(X_{1,i} - \hat{X}_1\right)\left(X_{1,i} - \hat{X}_1\right)^T. \quad (8)$$

$$\sigma_2^{\ 2} = \frac{1}{N}\sum_{i=1}^{N_s}\left(X_{2,i} - \hat{X}_2\right)\left(X_{2,i} - \hat{X}_2\right)^T. \quad (9)$$

Discriminative spatial pattern method constructs projection matrix W, which could be regarded as a spatial filter, and the separability of two samples could be larger after filtering by W,

$$S_w^{\ -1}S_B * W = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_{N_c} \end{bmatrix} * W. \quad (10)$$

$$S_B = \Sigma_{11} + \Sigma_{22} - \Sigma_{12} - \Sigma_{21}. \quad (11)$$

$$S_w = \sigma_1^{\ 2} + \sigma_2^{\ 2}. \quad (12)$$

where λ_i is the eigenvalue of the ith column of W. The correlation between $W^T\hat{X}_k$ and W^TY is calculated by CCA algorithm,

$$\mathrm{CCA}\left(W^T\hat{X}_k, W^TY\right) = \max_{U_k,V_k}\frac{\mathcal{E}\left[U_k^{\ T}W^T\hat{X}_kY^TWV_k\right]}{\sqrt{\mathcal{E}\left[U_k^{\ T}W^T\hat{X}_k\hat{X}_k^{\ T}WU_k\right]\cdot\mathcal{E}\left[V_k^{\ T}W^TYY^TWV_k\right]}}. \quad (13)$$

where $\mathcal{E}[\cdot]$ is the expectation. In pattern matching, the similarity between the training template and the testing signal is represented as a vector

$$\rho_k = \begin{bmatrix} \rho_{k1} \\ \rho_{k2} \\ \rho_{k3} \\ \rho_{k4} \\ \rho_{k5} \end{bmatrix} = \begin{bmatrix} \mathrm{corr}\left(W^T\hat{X}_k, W^TY\right) \\ -\mathrm{dist}\left(W^T\hat{X}_k, W^TY\right) \\ \mathrm{CCA}\left(W^T\hat{X}_k, W^TY\right) \\ \mathrm{corr}\left(U_k^{\ T}W^T\hat{X}_k, U_k^{\ T}W^TY\right) \\ \mathrm{corr}\left(V_k^{\ T}W^T\hat{X}_k, V_k^{\ T}W^TY\right) \end{bmatrix}, k = 1,2. \quad (14)$$

where $\mathrm{corr}(*)$ refers to the Pearson's correlation, $\mathrm{dist}(*)$ refers to the Euclidean distance. More similar it is between Y and $\hat{X}_k$, more larger the $\rho_{k,i}$ will be. After the summation of all coefficients from eq.(14), the predicted pattern of Y is

$$\hat{k} = \underset{k}{\mathrm{argmax}}\,\tilde{\rho}_k. \quad (15)$$

$$\tilde{\rho}_k = \sum_{i=1}^{5}\rho_{k,i}\,. \quad (16)$$

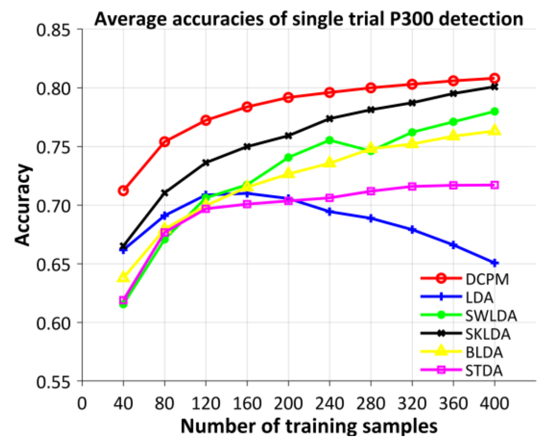

Figure 2. Single-trial P300 classification accuracies obtained by different methods.

III. Results and Analysis

We compared the single-trial P300 classification accuracy of the LDA, BLDA, SWLDA, SKLDA, STDA, and DCPM methods in different conditions of training samples. For each subject, 40 to 400 samples with a step of 40 samples (50% are target responses) were randomly selected from the data set to train the classifiers, while additional 400 samples (50% are target responses) were randomly selected from the remaining data to test the classifiers. To ensure fairness, all classification methods addressed the same training and testing samples. This procedure repeated for 10 times and the average accuracy and area under the receiver operating characteristic (ROC) curve were computed.

As shown in Fig. 2, the proposed DCPM method achieved higher accuracies than the other methods for all numbers of training samples. The advantage of DCPM was more evident for fewer training samples. Specifically, when using only 40 training samples, DCPM achieved 71.23% that was 4.71% higher than the second-best method, i.e. SKLDA. However, the accuracy advantage was reduced to 0.71% for 400 training samples. A more detailed AUC analysis was shown in Fig. 3 and Table I. The average AUC of DCPM ranged from 0.78 (40 training samples) to 0.88 (400 training samples), which was statistically significantly higher than those of the other methods revealed by paired t-tests. It should be noted that DCPM had more advantages over the other methods with fewer training samples. Individual analyses showed DCPM achieved higher AUC than the others for most conditions. It could be that the DCPM could extract single-trial P300 signal by suppressing the background EEG signals, hence it could perform better than the other methods with limited number of training samples. The results demonstrate DCPM is an efficient classification algorithm not only for the miniature aVEP-speller but also for the classical P300-speller.

IV. Conclusion

In this study, we introduced the DCPM algorithm to the P300-speller. Compared to five other traditional classifiers, i.e. LDA, SWLDA, BLDA, SKLDA and STDA, DCPM had a

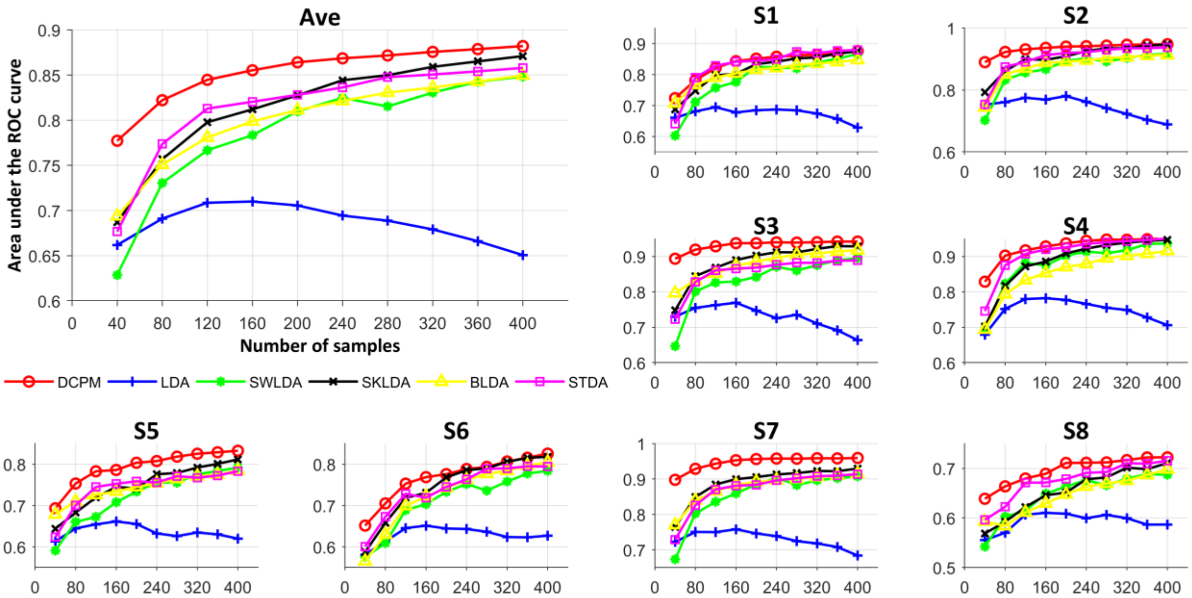

Figure 3. Average and individual AUCs of P300 identification obtained by different methods.

significantly better performance. The advantage of DCPM was more evident with insufficient training samples. The results of

TABLE I. STATISTICAL ANALYSIS RESULTS FOR AUC DIFFERENCE BETWEEN THE DCPM AND OTHER METHODS

Methods	Number of training samples									
	40	*80*	*120*	*160*	*200*	*240*	*280*	*320*	*360*	*400*
DCPM vs. LDA	††	††	††	††	††	††	††	††	††	††
DCPM vs. SWLDA	††	††	††	††	††	††	††	††	††	††
DCPM vs. SKLDA	††	††	††	††	†	†	†	**	*	*
DCPM vs. BLDA	††	††	††	††	††	††	††	††	††	†
DCPM vs. STDA	††	**	*	**	†	**	*	*	*	*

Note: ~nonsignificant, *p<0.05, **p<0.01, †p<0.005, ††p<0.001

this study demonstrate DCPM is a promising algorithm for P300-speller.

ACKNOWLEDGMENT

The authors sincerely thank all participants for their voluntary participation.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control," *Clin Neurophysiol,* vol. 113, no. 6, pp. 767-91, Jun 2002.

[2] G. Townsend and V. Platsko, "Pushing the P300-based brain-computer interface beyond 100 bpm: extending performance guided constraints into the temporal domain," *J Neural Eng,* vol. 13, no. 2, p. 026024, Apr 2016.

[3] B. Hong, F. Guo, T. Liu, X. Gao, and S. Gao, "N200-speller using motion-onset visual response," *Clin Neurophysiol,* vol. 120, no. 9, pp. 1658-66, Sep 2009.

[4] F. Guo, B. Hong, X. Gao, and S. Gao, "A brain-computer interface using motion-onset visual evoked potential," *J Neural Eng,* vol. 5, no. 4, pp. 477-85, Dec 2008.

[5] M. Xu, X. Xiao, Y. Wang, H. Qi, T. P. Jung, and D. Ming, "A Brain-Computer Interface Based on Miniature-Event-Related Potentials Induced by Very Small Lateral Visual Stimuli," *IEEE Trans Biomed Eng,* vol. 65, no. 5, pp. 1166-1175, May 2018.

[6] J. Polich, "Updating P300: an integrative theory of P3a and P3b," *Clin Neurophysiol,* vol. 118, no. 10, pp. 2128-48, Oct 2007.

[7] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "Toward enhanced P300 speller performance," *J Neurosci Methods,* vol. 167, no. 1, pp. 15-21, Jan 15 2008.

[8] U. Hoffmann, J. M. Vesin, T. Ebrahimi, and K. Diserens, "An efficient P300-based brain-computer interface for disabled subjects," *Journal of Neuroscience Methods,* vol. 167, no. 1, pp. 115-125, 2008.

[9] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K. R. Muller, "Single-trial analysis and classification of ERP components--a tutorial," *Neuroimage,* vol. 56, no. 2, pp. 814-25, May 15 2011.

[10] Y. Zhang, G. Zhou, Q. Zhao, J. Jin, X. Wang, and A. Cichocki, "Spatial-temporal discriminant analysis for ERP-based brain-computer interface," *IEEE Trans Neural Syst Rehabil Eng,* vol. 21, no. 2, pp. 233-43, Mar 2013.

[11] H. Cecotti and A. Graser, "Convolutional neural networks for P300 detection with application to brain-computer interfaces," *IEEE Trans Pattern Anal Mach Intell,* vol. 33, no. 3, pp. 433-45, Mar 2011.

[12] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalogr Clin Neurophysiol,* vol. 70, no. 6, pp. 510-23, Dec 1988.

[13] R. E. Alcaide-Aguirre and J. E. Huggins, "Novel hold-release functionality in a P300 brain-computer interface," *J Neural Eng,* vol. 11, no. 6, p. 066010, Dec 2014.

[14] D. J. Krusienski *et al.*, "A comparison of classification techniques for the P300 Speller," *J Neural Eng,* vol. 3, no. 4, pp. 299-305, Dec 2006.

[15] X. Lei, P. Yang, and D. Yao, "An empirical bayesian framework for brain-computer interfaces," *IEEE Trans Neural Syst Rehabil Eng,* vol. 17, no. 6, pp. 521-9, Dec 2009.

[16] J. Schafer and K. Strimmer, "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics," *Stat Appl Genet Mol Biol,* vol. 4, p. Article32, 2005.