RESEARCH ARTICLE

Journal of Medical Imaging and Health Informatics
Vol. 4, 1–7, 2014

Investigation of New Unsupervised Processing Methods for P300-Based Brain-Computer Interface

Mohammed J. Alhaddad[1], Mahmoud I. Kamel[1], and Yasser M. Kadah[2,*]

[1]*Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia*
[2]*Biomedical Engineering Department, Cairo University, Giza 12613, Egypt*

In Brain-computer Interface (BCI), the detection of activations is based on the experience gained through calibration or training sessions prior to actual use to build the classification model. This gives rise to several problems that include inter-session variability and time fading of accuracy after calibration. In this work, we investigate a new approach for brain-computer interface data that requires no prior training. The basic principle of this new class of unsupervised techniques is that the trial with true activation signal within each block has to be different from the rest of the trials within that block. Hence, a measure that is sensitive to this dissimilarity can be used to make a decision based on a single block without any prior training. The new approach is applied to experimental data for P300-based BCI for both normal and disabled subjects and compared to the classification results of the same data using the conventional processing techniques requiring prior calibration. Performance in different experiments assessed using classification block accuracy suggests that this approach can reach accuracies not very far from those obtained with training while maintaining robust performance in practice.

Keywords: Brain-Computer Interface, Adaptive System, Subspace Decomposition, Linear Vector Spaces.

1. INTRODUCTION

Brain-computer interface (BCI) offers hope for a communication channel with disabled patients who are not capable of using the normal communication channels. In spite of the major grounds covered by research in this area over the past decade, the challenge to make a reliable BCI system that combines mobility and accuracy remains open. Moreover, for some BCI techniques such as those based on detecting P300 signals in speller or one-of-multiple image selection tasks, existing commercial systems[1] require assistance from caregivers or patient's family to operate the system by the patient. So, there is an immediate need for developing technologies that would lead to the availability of such devices to patients at home with more independence and in less restrictive settings.

In P300-based BCI, a signal is triggered by an auditory or visual stimulus when participants are asked to watch for a particular target stimulus presented within a stream of other stimuli in an oddball paradigm.[2] In all previous P300-based BCI interfaces, detection of the P300 is based on experience gained through calibration or training sessions prior to actual use to utilize supervised training sets to build the classification model.[3] Several problems arise with this model including temporal variability of the signal (or inter-session variability) due to several reasons that include non-stationary brain dynamics and possible movement of electrode locations. This means that in practice, to communicate efficiently using such systems, the acquisition of data must be preceded by calibration with time difference as small as possible. As a result, the temporal persistence of such experience can be assumed to follow a training model close to an interpolation around the point at which the calibration/training was done, with most likely consequence is that the initial accuracy is expected to fade as time goes by after the initial training session. This required training imposes limitations on the utility (and hence commercialization) of the technology by individuals outside of research labs. Hence, efforts must be directed to develop methods that use unconventional decision models to overcome such limitations and achieve sufficient robustness for practical utility.

Previous work attempted to decrease the amount of calibration required for BCI and move toward a zero-training goal.[4] This method relies on observing the variation in training sessions and fitting such variation to spatial filters that can be used to make calibration sessions shorter. Even though the goal of this method is zero training, the approach relies on the utilization of prior information to allow future calibrations to be shorter or ideally no required. In that sense, it can be considered as a supervised training method with a more efficient training strategy that makes it possible for the training model to be more generalized and thus last longer. As a result, developing a P300-based BCI technology that can work adaptively without any prior calibration is still an open goal that once achieved would further the use of this important technology in real-life.

*Author to whom correspondence should be addressed.

The aim of this work is to investigate new methodologies for P300-based brain-computer interface data that require no prior training. This targets the development toward "plug-and-play" P300-based BCI devices whereby the device is taken out of the box and used immediately by the patient. The new method will be applied to experimental data and compared to the classification results of the same data using the conventional processing techniques requiring prior lengthy training sessions.

2. METHODS

The basic principle of the new class of unsupervised techniques is that the trial with P300 signal within each block has to be different from the rest of the trials within that block. In other words, if we have an N-trial block, one signal has a different form from all the other $(N-1)$ signals. This means that if we could find a measure that can be sensitive to this dissimilarity (or alternatively, the similarity of signals without activation), then we can indeed make a decision based on a single block without any prior training. In the following, we present a number of such measures and discuss how they will be used to separate the activated trial from the rest of trials in the same block. The assumption in all these methods remains that there is only one activated trial within each block. The input to each of these methods is a set of trials $\{x_1, x_2, \ldots, x_N\}$ given as a collection of $M \times 1$ vectors. The general block diagram for all methods is presented in Figure 1.

2.1. Outlier Detection Method

In this method, each trial is formulated as a vector in M-dimensional space where M is the number of points within each trial. The assumption underlying this method is that the vectors of all trials that exhibit no P300 activation will be similar and that they are all different from the one that has a P300 activation. Hence, a distance measure is used to compute the distance between each trial and all other trials in a pairwise manner. Then, for each trial, the sum of all distances with other trials is used to differentiate the one trial with the largest distance from all other trials. In mathematical form, the trial with P300 signal is computed as the solution to the optimization problem given by,

$$\max_i \sum_{\text{all } j \neq i} \|x_i - x_j\| \tag{1}$$

This means that this method detects the trial that is the furthest from all other trials. The norm in this equation was used as the 2-norm.[5] Possibility of using other norm definitions is possible but this one was selected to make its concept more visible by appealing to the common Euclidean distance as the measure used in this problem.

2.2. Correlation Method

The detection of the P300 signal relies on its characteristic shape and onset that are unique and help distinguish this type of activation from any other type. It is also very common for the literature working with P300 signals to show the P300 signal form their data by simple averaging of a number of trials with known activation presence. Here, we borrow an activation detection method from functional magnetic resonance imaging given the similarity between this technique and that of P300 based BCI. In particular, if the activation signal shape is somewhat known, it is possible to detect its presence by simple correlation of a "template" activation and each trial signal. If we have N trials within a block, then the trial with the strongest correlation with the template activation should most likely be the one with P300 signal. Since the onset of the true P300 signal varies between 300 and 500 ms, the template activation is used with different amounts of time shift to detect such correlation to make sure that such variability is taken into consideration. In a mathematical form, the activated trial is found by solving the following optimization over all trials in the block:

$$\max_{\text{all } i,\, \Delta t} \{x_i^T s_{\Delta t}\} \tag{2}$$

Here, $s_{\Delta t}$ is the template activation signal shifted in time by Δt. It is possible to constrain the range of time shifts to include only those with onset within the known range of P300 signal. However, this was not done in this work and the range of shifts was extended to be the full range of $[-M, M]$ for the M-sample signals.

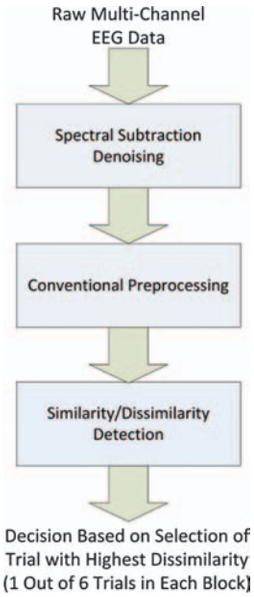

Fig. 1. Block diagram of the new approach.

2.3. Dot Product Method

This method is very similar to the Outlier Detection method above with the only exception in that the measure is here the dot product of the two trial vectors rather than the norm of their difference. This dot product relies on the fact that the dot product designates the component of one vector onto the other or basically the cosine of their angle if they both have similar magnitudes.[6] That is, similar vectors have higher dot products and vice versa. So, the optimization is here to find a trial that has the smallest dot product with all remaining vectors. Consequently, Eq. (1) is modified to be as follows:

$$\min_i \sum_{\text{all } j \neq i} x_i \cdot x_j \tag{3}$$

2.4. Cross-Correlation Method

This method bears similarities to both the dot product method and the correlation method. In particular, rather than computing the

dot product between two trials, it computes the cross-correlation between them and obtains the peaks of this cross-correlation function as the measure of similarity for this method. So, it is a generalization of the concept in the dot product method and also a variant of the correlation method whereby the template signal is just a different trial in the same block. In a mathematical form, we select the trial that satisfies the following optimization problem:

$$\max_i \left\{ \sum_{j \neq i} \max_{\Delta t} x_i^T x_{j,\Delta t} \right\} \tag{4}$$

2.5. Singular Value Decomposition (SVD) Method

The issue of representation of a set of vectors is a well-known problem in mathematics and also has wide utility in many applications. Some of the well-known solutions are based on the principal component analysis (PCA) that allows the computation of the so-called "principal component" that best represents a set of vectors by inspecting the eigenvalues of the different eigenvectors in the eigen-decomposition of the problem and finding the eigenvector with much larger eigenvalue from the rest. As the set of vectors become more and more independent, it becomes more difficult to find a single vector that can best represent them all ending with the ideal case of orthonormal basis that result in all unity eigenvalues. In our context, the assumption is that $(N-1)$ trials are somewhat similar (at least not independent). Therefore, if we perform such analysis for each of the possible $(N-1)$ trials and using a sparsity measure for the resultant eigenvalues of each decomposition that can detect how close each of these sets of vectors to the idea case of a single outstanding eigenvalue, it can be possible to detect the trial with activation as the remaining vector.[7] That is, the most sparse set of eigenvalues of all decomposition denote that these trials are not activated and it turn the remaining vector has the P300 signal. The 1-norm measure was selected as the sparsity measure for the singular values in our implementation. In a mathematical form:[6]

$$\text{For Trial } i,\ A_i = \{x_j\}_{1 \leq j \leq N,\, j \neq i} = U \Sigma V^T \tag{5}$$

where U and V are orthogonal matrices of size $M \times M$ and $(N-1) \times (N-1)$ respectively, and Σ is a $M \times (N-1)$ matrix with its upper $(N-1) \times (N-1)$ matrix taking a diagonal form with singular values s_j^i on the diagonal and the lower $(M-N+1) \times (N-1)$ a zero matrix. The activated trial is hence taken as the solution to the following optimization problem,

$$\max_{i,\, \text{all } A_i} \|\{s_1, s_2, \ldots, s_{N-1}\}\|_1 \tag{6}$$

That is, we select the trial that has all the remaining trials forming a matrix with the largest 1-norm for its singular values (similar to the strategy used with compressive sensing[8]).

3. EXPERIMENTAL VERIFICATION

The experimental P300-based BCI data of Hoffmann et al.[3] were used to test the developed no-training unsupervised methods and compare it to their results that were obtained with 3 sessions of training of a Bayesian Linear Discriminant Analysis (BLDA) classifier. To make that comparison directly applicable, we followed the exact same sequence of preprocessing and classification in this paper. The description of the data set is found in detail in Hoffmann et al.[3] but a summary will be provided here. The duration of one run was approximately one minute and the duration of one session including setup of electrodes and short breaks between runs was approximately 30 min. One session comprised on average 810 trials, and the whole data for one subject consisted on average of 3240 trials. The experimental paradigm consists of flashing one of six images in a random order after asking the subject to count how many times a particular image appears. So, the six stimulus images appear in 6 consecutive trials, usually termed a block. The P300 signal is triggered by the appearance by the image of interest only (i.e., the one currently being counted and not the other five images) and can be detected from EEG signals to indicate the subject selection. In the supervised BLDA method, four-fold cross-validation was used to estimate average classification accuracy for each subject. So, each result from this classifier needs 3 sessions for training to compute. On the other hand, the proposed techniques work directly on the data without any prior training. This is a major difference between the previous methods and this work.

The standard preprocessing operations were applied to the data including referencing, bandpass filtering with cut-off frequencies set to 1.0 Hz and 12.0 Hz, downsampling by a factor of 64, single trials extraction, windsorizing and finally amplitude normalization. Additionally, for the new approach, signal denoising based on spectral subtraction was employed to the raw data before the above preprocessing.[9] Other methods were used based on wavelet denoising and other types of filters can also be used for similar results.[10, 11] The denoising block is placed before the standard preprocessing steps above as shown in Figure 1. The number of electrodes was selected as 4, 8, 16 or 32 depending on the experiment with the same electrode configurations in the data set used.[3] Then, the samples from the selected electrodes were concatenated into feature vectors to be used for classification using either supervised BLDA[3] or based on the new approach in this work. The dimensionality of the feature vectors was $N_e \times N_t$, where N_e denotes the number of electrodes (selected as 4, 8, 16, or 32) and N_t denotes the number of temporal samples in one trial (32 samples in our experiments). The results of the different methods proposed are compared to each other and to supervised BLDA classification. The performance is measured using the block accuracy measure which is most relevant comparison criterion in this application. The block accuracy considers the data as blocks of 6 trials where only one of them should be selected with P300 signal showing while the others are not. If the classification results indicate anything other than only one activation at the correct image, it considers the whole block as incorrect. The results using different numbers of blocks were achieved by summing the signals from the selected number of blocks and using the sum as the new signal for classification using the proposed techniques. For the new approach, the block accuracy estimation experiments were repeated 24 times for independent sets of blocks containing trials from the same session and from different sessions for a given subject to avoid any bias and obtain accurate final results. The results are computed as block accuracy results for each subject, and average block accuracy results for all subjects. Also, relative block accuracy results were obtained by dividing the block accuracy results of the proposed methods by the block accuracy of the reference supervised BLDA method to allow better assessment of the performance.

4. RESULTS AND DISCUSSION

The block accuracy results of using the new approach on 4 sample subjects are shown in Figure 2. The figure presents the results using

(a) outlier detection method,
(b) correlation method,
(c) dot product method,
(d) cross correlation method,

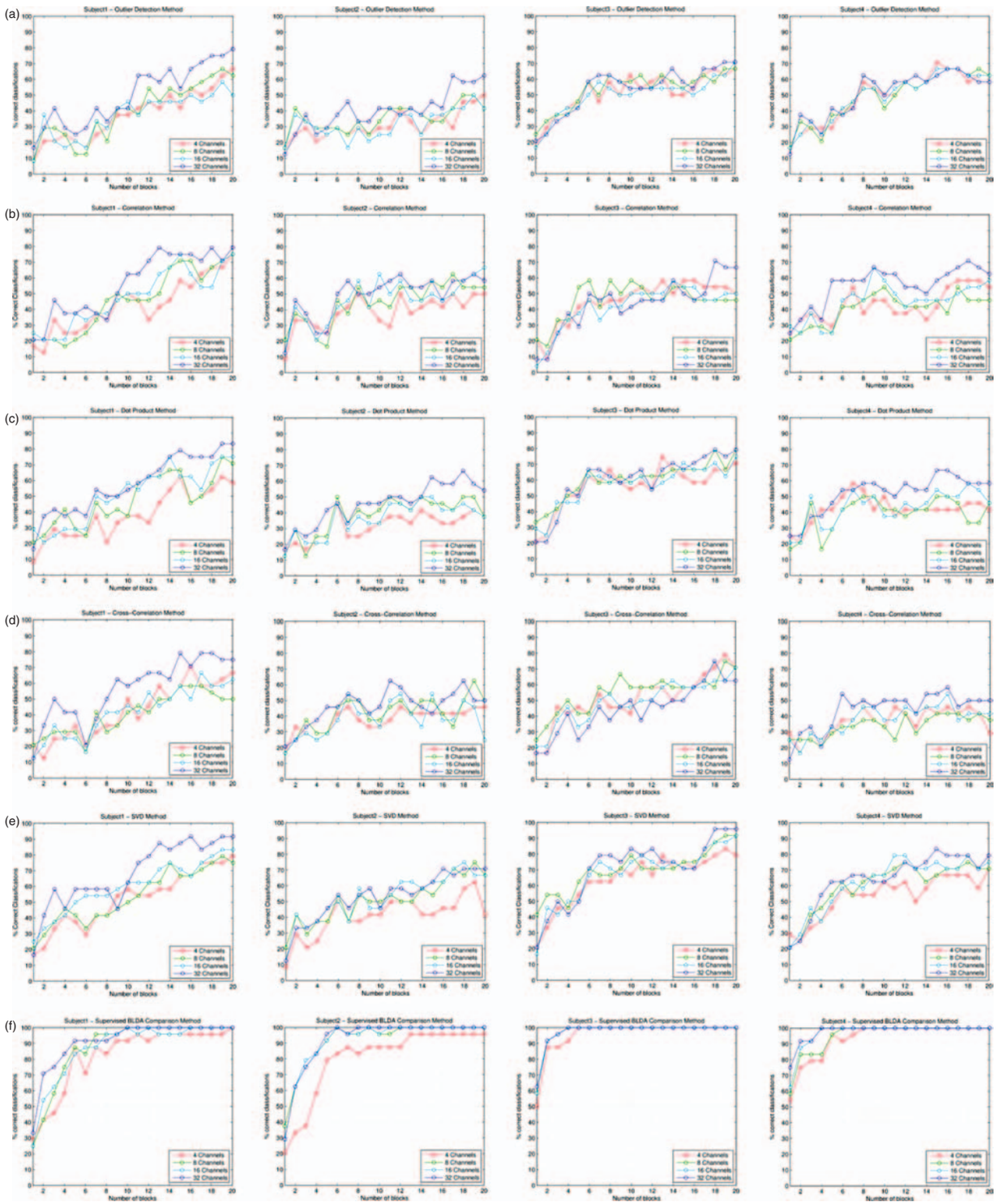

Fig. 2. Block accuracy results for 4 subjects using (a) outlier detection method, (b) correlation method, (c) dot product method, (d) cross correlation method, (e) singular value decomposition method, and (f) supervised classification using BLDA for comparison in different rows.

(e) singular value decomposition method, and
(f) supervised classification using BLDA for direct comparison, each on a separate row.

The results also show the cases of using 4, 8, 16 and 32 channel data on the same graph for each case/method. The relative block accuracy results are shown in Figure 3 with the same order of the methods for better interpretation. In Figure 4, the average block accuracies and relative block accuracies for all subjects are presented for each method to see an overall picture of the performance. In Table I, the low and high limits for the block accuracies for each method are given whereas those for relative block accuracies are given in Table II. In the following, the analysis of the results according to different parameters is presented.

- *Effect of Method*: The results from the sample individual cases show that the proposed method based on SVD provided the best performance reaching 95% block accuracy in some cases. The other 4 methods were comparable in block accuracy performance reaching accuracies above 80% in all cases. Examining the relative block accuracy curves, it is clear that all method range from 30% of the performance of the supervised BLDA method for low block averaging to above 90% in some cases with high block averaging and particularly for SVD. From the average performance curves and Tables I and II, the performance of the

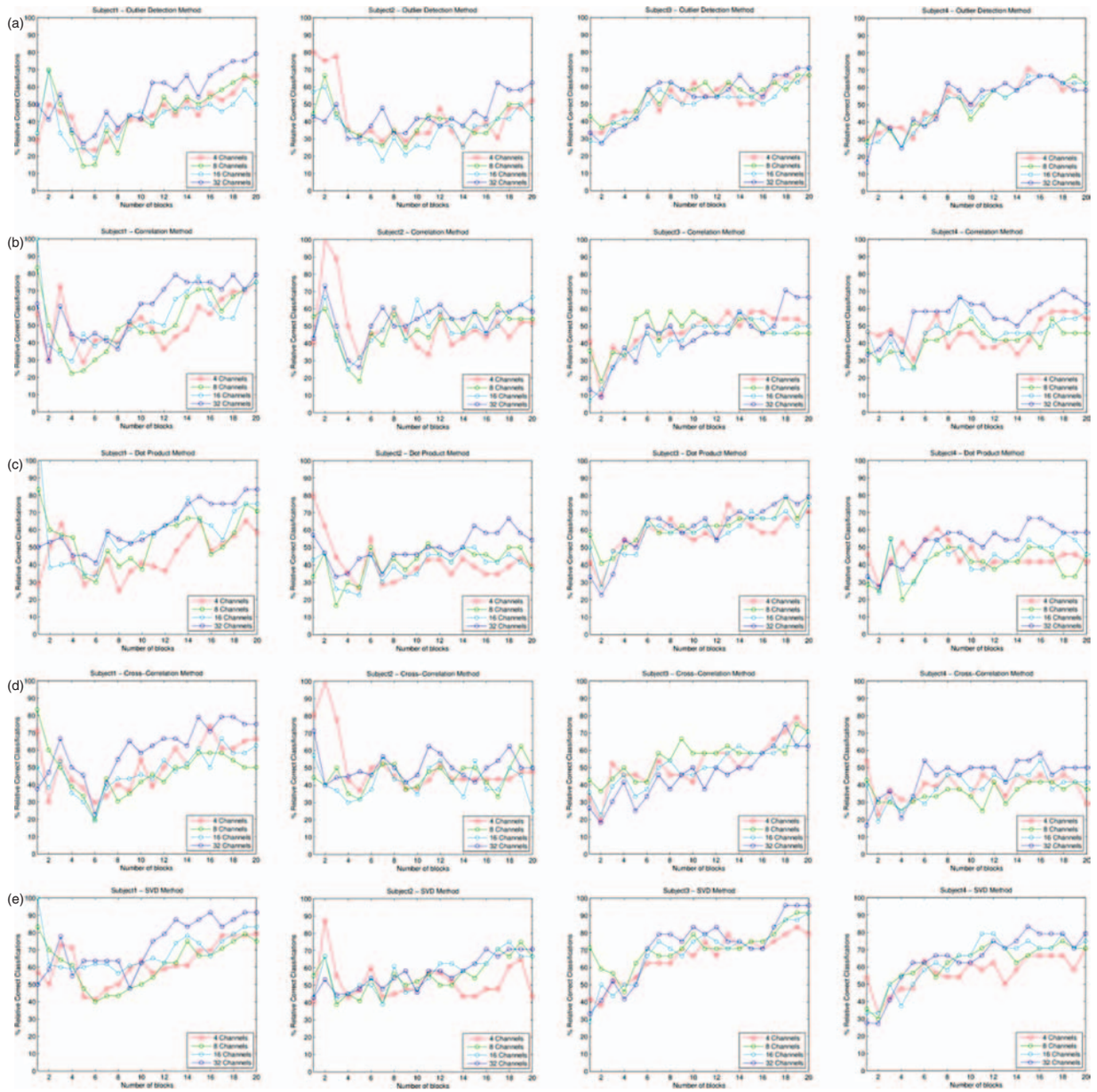

Fig. 3. Relative block accuracy results for 4 subjects using supervised BLDA results as a reference for comparison using (a) outlier detection method, (b) correlation method, (c) dot product method, (d) cross correlation method, and (e) singular value decomposition method in different rows.

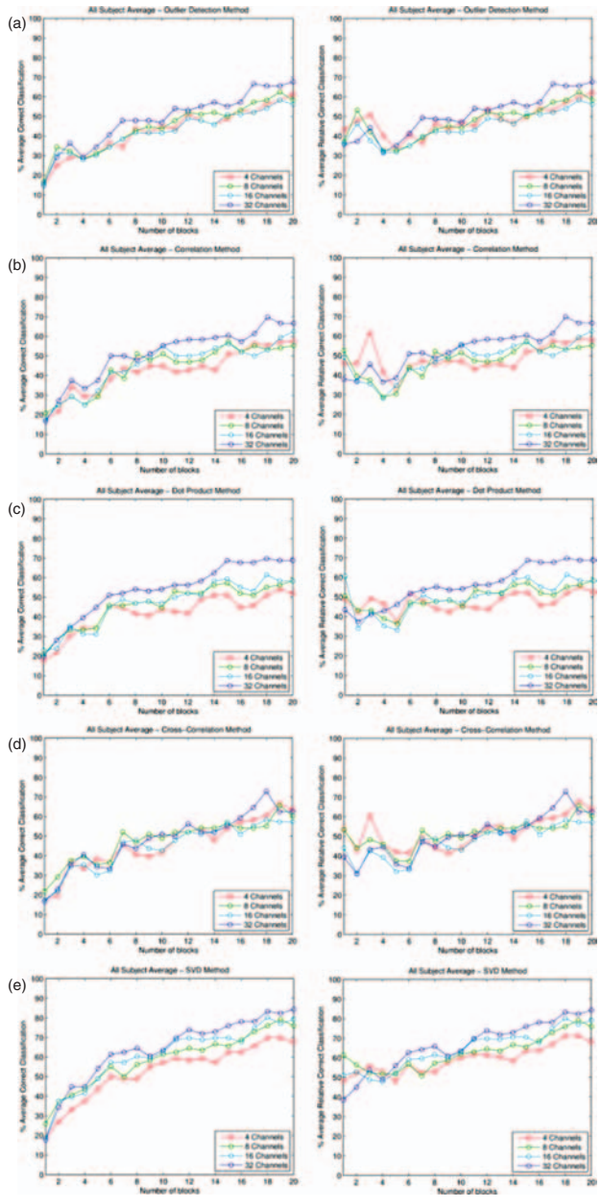

Fig. 4. Average block accuracy and Relative block accuracy results over all subjects for comparison using (a) outlier detection method, (b) correlation method, (c) dot product method, (d) cross correlation method, and (e) singular value decomposition method in different rows.

SVD method was clearly dominant on the high end of the range of block and relative block accuracies with average performance reaching 84.4%. The other proposed methods provide block and relative block accuracies around 70%. Their performances vary but they are within a close proximity of each other.

- *Effect of number of channels*: The difference is clear between the cases of 4 and 32 channels due to the implicit spatial averaging that occurs in using the higher number of channels. However, the difference in performance between the cases of 8 and 16 channels was not much and they both present accuracies in the middle between those of 4 and 32 channel data.
- *Effect of number of blocks*: There is a clear linear relationship between the block and relative block accuracies and the number of blocks used that is evident in all average curves. This is expected since the higher number of blocks allows more temporal averaging that improve the signal-to-noise ratio of the trials enhancing their separation procedures.
- *Variability among subjects*: Some variations among subjects were observed where the results from Subject 2 for example were significantly lower than those from other subjects.

From a global overview of results, one can observe that the new methods with no training requirement were able to achieve relative block accuracies of above 70% of the performance of the supervised BLDA method that require prior lengthy training with 3 full sessions. This is particularly important for such applications as P300-based BCI where the disabled person chooses one out of several images to indicate the need for a particular action. Such selection is usually done infrequently and with time separation that would require training to be repeated every time one selection has to be made, which would make this cumbersome for practical use. This demonstrates potential for the new approach that works adaptively without any prior knowledge or assistance from care givers or family members and without the training overhead required in supervised methods.

Given that the correct communication between the brain and the computer must include no ambiguity, a measure that considers the correct answer at the level of a whole block (i.e., one of six images) rather than an individual image on/off measure must be used. For example, if within a particular block 2 images out of 6 are classified as "selected" with only one of them a true selection, the usual accuracy would give a success rate of 5 out of 6, which is 83.3%. This is clearly incorrect because the message received was ambiguous. On the other hand, the block accuracy considers this whole block as incorrectly classified and would give a success rate of 0%, which is a realistic assessment of the utility of the received information. Other measures were used in other studies as well such as the bit rate. Here, given that we are

Table I. Performance of different methods in terms of their low and high block classification accuracies computed over all subjects as compared to the results from supervised comparison method that requires 3-session training at the bottom row.

	4-Channels		8-Channels		16-Channels		32-Channels	
Block accuracy limits	Low (%)	High (%)	Low (%)	High (%)	Low (%)	High (%)	Low (%)	High (%)
Outlier detection	14.6	61.5	16.7	62.5	14.6	58.3	15.6	67.7
Correlation	17.7	57.3	20.8	57.3	17.7	62.5	16.7	69.8
Dot product	17.7	54.2	20.8	58.3	**22.9**	61.5	**19.8**	69.8
Cross-correlation	17.7	66.7	21.9	65.6	17.7	58.3	16.7	72.9
SVD	**18.8**	**69.8**	**26.0**	**79.2**	18.8	**80.2**	17.7	**84.3**
Comparison method	38.5	98.9	44.8	100	43.8	100	50.0	100

Table II. Performance of different methods in terms of their relative low and high block classification accuracies computed over all subjects with reference to the results from supervised BLDA comparison method that requires 3-session training.

	4-Channels		8-Channels		16-Channels		32-Channels	
Block accuracy limits	Low (%)	High (%)	Low (%)	High (%)	Low (%)	High (%)	Low (%)	High (%)
Outlier detection	32.9	62.0	32.7	62.5	31.3	58.3	31.9	67.7
Correlation	33.1	61.8	28.9	57.3	28.2	62.5	36.5	69.8
Dot product	37.1	55.3	36.3	58.3	33.2	61.5	37.3	69.8
Cross-correlation	41.5	67.8	37.3	65.6	31.2	58.3	30.9	72.9
SVD	**48.1**	**71.3**	**50.9**	**79.2**	**47.8**	**80.2**	**38.5**	**84.4**

comparing methods with no training to others with long training, the commonly used definition of bit rate is clearly flawed because it fails to account for the time needed for the required training and the fading of such training with time. Therefore, it was not possible to utilize this measure in this work.

The results for a particular number of blocks were achieved by summing the signals from the selected number of blocks and using the sum as the new signal for classification using the proposed techniques. Another approach that can be used is to calculate the proposed classification metrics from each block and then sum up all metrics from the desired number of blocks then make the classification decision based on this sum. The approach we used gave a better performance and hence was preferred over this alternative. The analysis of the problem shows that this is due to the nonlinearity of the computed measures that makes the average of the individual block measures completely different from the measure of average of blocks.

The applications of the new approach include developing plug-and-play P300 based BCI devices that require no training and work straight out of the box. Even though the block accuracy of such devices will be lower than the conventional methods with prior training, its adaptive nature and availability for immediate use without calibration boosts the robustness of their performance and practical utility.

5. CONCLUSIONS

The results of a new approach for processing P300-based brain-computer interface data that allows classification of trials within a block without prior training are presented. The new method was verified using experimental data and compared to the results obtained with conventional processing with lengthy training. Promising results were obtained suggesting potential for the new approach in making the P300 based BCI technology easier to implement as plug-and-play device with no prior calibration required and capable of adaptively follow any changes in the subject's condition.

Acknowledgments: The authors are thankful to the research group that conducted the research in Hoffmann et al.[3] for making their data and processing code available. We would like to thank the other members of our team for their efforts in the BCI project. This research was funded by the Deanship of Scientific Research (DSR), King Abdulaziz University, under grant, No. (16-15-1432 HiCi). The authors, therefore, acknowledge with thanks DSR technical and financial support.

References and Notes

1. http://www.intendix.com/ (accessed on July 24, 2013).
2. A. Kubler and K.-R. Muller, An Introduction to Brain-Computer Interfacing, in Toward Brain-Computer Interfacing, edited by G. Dornhege, J. R. Millan, T. Hinterberger, D. J. McFarland, and K.-R. Muller, The MIT Press, Cambridge **(2007)**.
3. U. Hoffmann, J.-M. Vesin, T. Ebrahimi, and K. Diserens, An efficient P300-based brain-computer interface for disabled subjects. *Journal of Neuroscience Methods* 167, 115 **(2008)**.
4. M. Krauledat, M. Tangermann, B. Blankertz, and K.-R. Müller, Towards zero training for brain-computer interfacing. *PLoS One* 3, e2967 **(2008)**.
5. G. Strang, Introduction to Linear Algebra, 4th edn., Wellesley Cambridge Press, New York **(2009)**.
6. G. Golub and C. F. Van Loan, Matrix Computations, Johns Hopkins University Press, Baltimore **(1996)**.
7. A. Hyvarinen, J. Karhunen, and E. Oja, Independent Component Analysis, Wiley, New York **(2001)**.
8. E. Candes and M. Wakin, An introduction to compressive sampling. *IEEE Signal Processing Magazine* 25, 21 **(2008)**.
9. Y. Kadah, Adaptive denoising of event-related functional magnetic resonance imaging data using spectrum subtraction. *IEEE Trans. Biomedical Engineering* 51, 1944 **(2004)**.
10. Z. A. Mustafa, B. A. Abrahim, I. A. Yassine, N. Zayed, and Y. M. Kadah, Wavelet domain bilateral filtering with subband mixing for magnetic resonance image enhancement. *J. Med. Imaging Health Inf.* 2, 230 **(2012)**.
11. B. A. Abrahim, Z. A. Mustafa, I. A. Yassine, N. Zayed, and Y. M. Kadah, Hybrid total variation and wavelet thresholding speckle reduction for medical ultrasound imaging. *J. Med. Imaging Health Inf.* 2, 114 **(2012)**.

Received: 26 October 2013. Accepted: 20 December 2013.