Ensemble SWLDA Classifiers for the P300 Speller

Garett D. Johnson and Dean J. Krusienski

University of North Florida
Jacksonville, FL 32224
dean.krusienski@unf.edu

Abstract. The P300 Speller has proven to be an effective paradigm for brain-computer interface (BCI) communication. Using this paradigm, studies have shown that a simple linear classifier can perform as well as more complex nonlinear classifiers. Several studies have examined methods such as Fisher's Linear Discriminant (FLD), Stepwise Linear Discriminant Analysis (SWLDA), and Support Vector Machines (SVM) for training a linear classifier in this context. Overall, the results indicate marginal performance differences between classifiers trained using these methods. It has been shown that, by using an ensemble of linear classifiers trained on independent data, performance can be further improved because this scheme can better compensate for response variability. The present study evaluates several offline implementations of ensemble SWLDA classifiers for the P300 speller and compares the results to a single SWLDA classifier for seven able-bodied subjects.

1 Introduction

A brain-computer interface (BCI) is a device that uses brain signals to provide a non-muscular communication channel [16], particularly for individuals with severe neuromuscular disabilities. The P300-event related potential is an evoked response to an external stimulus that is observed in scalp-recorded electroencephalography (EEG). Based on multiple studies in healthy volunteers [7], and initial studies in persons with physical disability [9][12][15], the P300 speller has potential to serve as an effective communication device for persons who have lost or are losing the ability to write and speak.

Many classification techniques have been investigated for the P300 Speller [1][2][6][8][13]. Several of these studies indicate that simple linear classifiers perform as well or better than more complex nonlinear classifiers for discriminating the P300 response. Recent work on ensemble linear classifiers shows the potential for further performance improvements [10]. The study uses an ensemble of linear classifiers trained on independent data using support vector machines (SVM). It has been shown that linear classifiers trained using stepwise linear discriminant analysis (SWLDA) can perform favorably compared to training using SVMs [8]. The objective of the present study is to investigate whether an ensemble of SWLDA classifiers offers any performance advantages over a single SWLDA classifier for the P300 speller by evaluating several offline implementations of ensemble SWLDA classifiers and comparing the results to a single SWLDA classifier for able-bodied subjects.

J.A. Jacko (Ed.): Human-Computer Interaction, Part II, HCII 2009, LNCS 5611, pp. 551–557, 2009.

The P300 Speller. The P300 Speller described by Farwell and Donchin [5] presents a 6 x 6 matrix of characters as shown in Figure 1. Each row and each column are intensified; the intensifications are presented in a random sequence. The user focuses attention on one of the 36 cells of the matrix. The sequence of 12 flashes, 6 rows and 6 columns, constitutes an Oddball Paradigm [4] with the row and the column containing the character to be communicated constituting the rare set, and the other 10 intensifications constituting the frequent set. Items that are presented infrequently (the rare set) in a sequential series of randomly presented stimuli will elicit a P300 response if the observer is attending to the stimulus series. Thus, the row and the column containing the target character will elicit a P300 when intensified, because this constitutes a rare event in the context of all other character flashes. With proper P300 feature selection and classification, the attended character of the matrix can be identified and communicated.

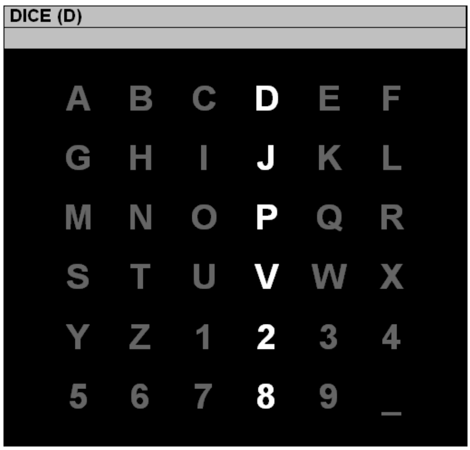

Fig. 1. The 6x6 matrix used in the current study. A row or column intensifies for 100 ms every 175 ms. The letter in parentheses at the top of the window is the current target letter "D." A P300 should be elicited when the fourth column or first row is intensified. After the intensification sequence for a character epoch, the result is classified and online feedback is provided directly below the character to be copied.

2 Data Collection

The data were collected by the Wadsworth Center BCI Laboratory in accordance with New York State Department of Health Institutional Review Board.

Participants. Seven able-bodied people (six men and one woman ages 24-50) were the participants in this study. The participants varied in their previous BCI experience, but all participants had either no or minimal experience using a P300-based BCI system.

Task, Procedure, and Design. The participant sat upright in front of a video monitor and viewed the matrix display. The task was to focus attention on a specified letter of the matrix and silently count the number of times the target character intensified, until a new character was specified for selection. All data was collected in the copy speller mode: words were presented on the top left of the video monitor and the character currently specified for selection was listed in parentheses at the end of the letter string (see Figure 1). Each session consisted of 8-12 experimental runs; each run was composed of a word or series of characters chosen by the investigator. The rows and columns were intensified for 100 ms with 75 ms between intensifications. One character epoch (i.e., one trial) consisted of 15 intensifications of each row and column. Specifically, the classification was performed after every row and column has been intensified 15 times. Each session consisted of 36 character epochs, equivalent to 6480 stimuli (row/column intensifications). A single session, lasting approximately one hour, per participant was collected per day over several weeks.

Data Acquisition. The EEG was recorded using a cap (Electro-Cap International, Inc.) embedded with 64 electrodes distributed over the scalp, based on the International 10 – 20 system [14]. All channels were referenced to the right earlobe, and grounded to the right mastoid. The EEG was bandpass filtered 0.1 – 60 Hz and amplified with a SA Electronics amplifier, digitized at a rate of 240 Hz, and stored. All aspects of data collection and experimental procedure were controlled by the BCI2000 system [11].

3 Response Classification

Responses were collected from the 8 ear-referenced channels shown in Figure 2. The channel selection and data preprocessing are based on results found in [7]. For each of the 8 channels, 800-ms segments of data following each intensification were extracted. The segments were then moving average filtered and decimated to 20Hz. The resulting data segments were concatenated by channel for each intensification, creating a single feature vector for training the classifiers.

Several ensemble classification schemes were investigated, where each linear classifier in the ensemble was trained using Stepwise Linear Discriminant Analysis (SWLDA) [3]. SWLDA is a technique for selecting suitable predictor variables to be included in a multiple regression model that has proven successful for discriminating P300 Speller responses. A combination of forward and backward stepwise regression was implemented. Starting with no initial model terms, the most statistically significant predictor variable having a p-value < 0.1, was added to the model. After each new entry to the model, a backward stepwise regression was performed to remove the least significant variables, having p-values > 0.15. This process was repeated until the model includes a predetermined number of terms, or until no additional terms satisfy the entry/removal criteria. In this case, the final discriminant function was restricted to contain a maximum of 60 features [7].

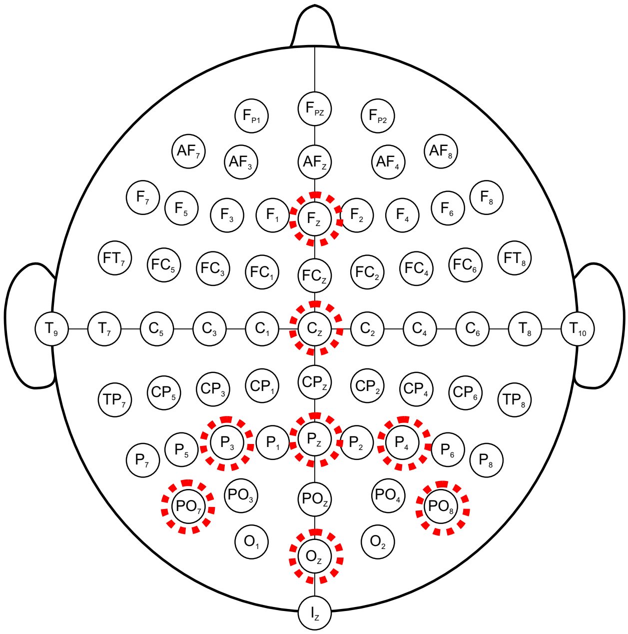

Fig. 2. The electrode montage used in the current study [8]. The 8 electrodes selected for analysis are indicated by the dotted circles.

The ensemble classifiers used for this study consist of *N* linear classifiers, each independently trained using SWLDA. For each subject, the first session of data was used to train the ensemble classifier. This session's data was divided into *N* equal, non-overlapping segments, where each segment was used to train an individual classifier in the ensemble. The ensemble classifiers were evaluated using the 4 subsequent sessions for each subject. For a given set of test data, each classifier in the ensemble independently evaluates the inner product of its respective SWLDA weights and the feature vector corresponding to each response, forming *N* scores, one for each classifier. The resultant score for each intensification response was generated as a weighted sum of the *N* individual classifier scores. The predicted matrix symbol was determined as the maximum of the sum of resultant scores for the respective row and column intensifications, respectively:

$$predicted\ row = \max_{rows}\left[\sum_{i_{row}}\sum_{n=1}^{N} w_n \cdot x_{i_{row}}\right] \tag{1}$$

$$predicted\ column = \max_{columns}\left[\sum_{i_{column}}\sum_{n=1}^{N} w_n \cdot x_{i_{column}}\right] \tag{2}$$

The following values of *N* were evaluated: *N* = [1, 2, 4, 6, 9, 12]. In addition, three schemes for weighting the individual classifiers in the ensemble to produce the single classification score were investigated: a simple average of the *N* classifier scores (AVG), weighting each classifier score by its respective accuracy on the training data (ACC), and using only the maximum output of all classifiers as the resultant ensemble score: i.e., all other classifiers are weighted by zero (MAX). The coefficients for each of these weighting schemes were also determined using the same training set used for training the SWLDA classifiers.

4 Results

The performance, averaged across subjects, of the various combinations of number of classifiers in the ensemble and classifier weighting methods is shown in Figure 3.

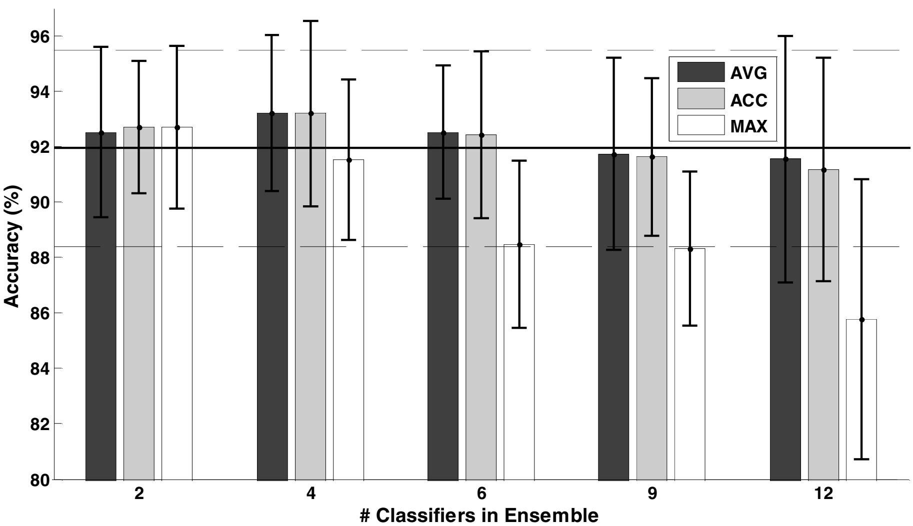

Fig. 3. The figure shows the classification accuracy, averaged across subjects, for the number of classifiers in the ensemble (2, 4, 6, 9, 12) and the three classifier weighting methods (AVG, ACC, MAX). The solid line indicates the average baseline performance using a single SWLDA classifier, with the dotted lines indicating the standard error. The error bars also indicate the standard error for each method.

A balanced one-way ANOVA on the accuracy across subjects after 15 intensifications did not reveal a statistically significant difference between any of the methods. The maximum classification accuracy achieved by an ensemble SWLDA classifier method for each of the seven subjects is compared to the accuracy achieved by a single SWLDA classifier in Table 1.

Table 1. For each subject, the table lists the baseline performance (± standard deviation across test sessions) using a single SWLDA classifier and the best performance achieved by an ensemble SWLDA classifer method. The numbers following the method abreviation indicate the number of classifiers in the ensemble. Note that multiple ensemble classifier methods resulted in the same performance results.

SUBJECT	BASELINE	BEST	METHOD
A	93.75 ± 7.3	94.44 ± 6.42	AVG6, ACC6
B	71.84 ± 6.27	81.28 ± 10.32	ACC4
C	93.75 ± 5.73	95.14 ± 4.74	AVG2,4,6; ACC2,4,6
D	98.61 ± 1.6	98.61 ± 1.6	AVG2, ACC2, MAX2
E	97.92 ± 2.66	99.3 ± 1.39	MAX2
F	89.83 ± 3.77	89.91 ± 5.75	MAX4
G	97.92 ± 2.66	97.92 ± 2.66	*All except* MAX4,6,9,12

5 Discussion

The results indicate that extending a single linear classifier to an ensemble linear classifier scheme can result in marginal to significant improvements in performance, depending on the subject as shown in Table 1. The improvement is only marginal for several subjects because the performance using a single linear classifier is already nearly maximized. However, it should be noted that in addition to the performance improvements, the variance across sessions is also decreased for several subjects.

Figure 3 shows that the best classification results are achieved in this context by using between 2 and 6 classifiers in the ensemble, which all but MAX4 and MAX6 perform better on average than the single classifier baseline. The inferior performance of the MAX methods can be attributed to the increased variability of the individual classifiers when trained using smaller data sets. The AVG and ACC methods outperform MAX because, in contrast to the single classifier score used for MAX, they combine the scores from all classifiers, which acts to decrease the variability ensemble score. The degradation in performance with more than 6 classifiers can be attributed to the fact that the amount of training data was held constant, meaning that each individual classifier is trained on proportionally less data as the number of classifiers increases. Because P300 classification often remains stable over time [7], additional training data can be used to create a larger ensemble that can potentially better account for response variance.

Based on related P300 Speller classification studies [8][10], it is presumed that similar results would be attained using classifier training methods other than SWLDA. When employing a single classifier, it is straightforward to construct an ensemble in this fashion to potentially improve performance. By using the AVG or ACC method, an equivalent single classifier can be realized via the linear combination of the ensemble classifiers using the respective classifier weights. This is not possible with the MAX method because the computation of the maximum classifier output does not allow for the individual classifiers to be combined linearly.

References

1. Blankertz, B., Müller, K.R., Krusienski, D.J., Schalk, G., Wolpaw, J.R., Schlögl, A., Pfurtscheller, G., Millán, J.R., Schröder, M., Birbaumer, N.: The BCI Competition III: Validating Alternative Approaches to Actual BCI Problems. IEEE Transactions on Neural Systems and Rehabilitation Engineering 14(2) (2006)
2. Bostanov, V.: BCI competition 2003-data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram. IEEE Trans. Biomed. Eng. 51, 1057–1061 (2004)
3. Draper, N., Smith, H.: Applied Regression Analysis, 2nd edn., pp. 307–312. John Wiley and Sons, Chichester (1981)
4. Fabiani, M., Gratton, G., Karis, D., Donchin, E.: Definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. Advances in Psychophysiology 2, 1–78 (1987)
5. Farwell, L.A., Donchin, E.: Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. Electroenceph. clin. Neurophysiol. 70, 510–523 (1988)
6. Kaper, M., Meinicke, P., Grossekathoefer, U., Lingner, T., Ritter, H.: BCI competition 2003-data set IIb: Support vector machines for the P300 speller paradigm. IEEE Trans. Biomed. Eng. 51, 1073–1076 (2004)
7. Krusienski, D.J., Sellers, E.W., McFarland, D.J., Vaughan, T.V., Wolpaw, J.R.: Toward Enhanced P300 Speller Performance. Journal of Neuroscience Methods 167, 15–21 (2008)
8. Krusienski, D.J., Sellers, E.W., Cabestaing, F., Bayoudh, S., McFarland, D.J., Vaughan, T.M., Wolpaw, J.R.: A Comparison of Classification Techniques for the P300 Speller. Journal of Neural Engineering 3, 299–305 (2006)
9. Nijboer, F., Sellers, E.W., Mellinger, J., Jordan, M.A., Matuz, T., Furdea, A., Mochty, U., Krusienski, D.J., Vaughan, T.M., Wolpaw, J.R., Kubler, A.: A Brain-Computer Interface (BCI) for People with Amyotrophic Lateral Sclerosis (ALS). Clinical Neurophysiology 119, 1909–1916 (2008)
10. Rakotomamonjy, A., Guigue, V.: BCI Competition III: Dataset II – Ensemble of SVMs for BCI P300 Speller. IEEE Trans. Biomedical Engineering 55(3), 1147–1154 (2008)
11. Schalk, G., McFarland, D.J., Hinterberger, T., Birbaumer, N., Wolpaw, J.R.: BCI2000: A general-purpose brain-computer interface (BCI) system. IEEE. Trans. Biomed. Eng. 2004 51, 1034–1043 (2004)
12. Sellers, E.W., Donchin, E.: A P300-based brain-computer interface: Initial tests by ALS patients. Clin. Neurophysiol. 117, 538–548 (2006)
13. Serby, H., Yom-Tov, E., Inbar, G.F.: An improved P300-based brain-computer interface. IEEE Trans. Neural Syst. Rehabil. Eng. 13, 89–98 (2005)
14. Sharbrough, F., Chatrian, C.E., Lesser, R.P., Luders, H., Nuwer, M., Picton, T.W.: American Electroencephalographic Society guidelines for standard electrode position nomenclature. J. Clin. Neurophysiol. 8, 200–202 (1991)
15. Vaughan, T.M., McFarland, D.J., Schalk, G., Sarnacki, W.A., Krusienski, D.J., Sellers, E.W., Wolpaw, J.R.: The Wadsworth BCI Research and Development Program: At Home with BCI. IEEE Transactions on Neural Systems and Rehabilitation Engineering 14(2) (2006)
16. Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M.: Brain-computer interfaces for communication and control. Clin. Neurophysiol. 113, 767–791 (2002)