Published in final edited form as:

Disabil Rehabil Assist Technol. 2015 May ; 10(3): 216–220. doi:10.3109/17483107.2014.884174.

Electrode subset selection methods for an EEG-based P300 brain-computer interface

Michael T. McCann^{a,*},¹, David E. Thompson^a, Zeeshan H. Syed^b, and Jane E. Huggins^{a,c}

^aUniversity of Michigan Biomedical Engineering Department, 2200 Bonisteel, Blvd., Ann Arbor, MI 48109-2099, USA

^bUniversity of Michigan Electrical Engineering and Computer Science Department, 1301 Beal Ave., Ann Arbor, MI 48109-2122, USA

^cDepartment of Physical Medicine and Rehabilitation, University of Michigan Hospital, Room C640 MIB Ann Arbor, MI 48109-5841, USA

Abstract

Purpose—An electroencephalography (EEG)-based P300 speller is a type of brain-computer interface (BCI) that uses EEG to allow a user to select characters without physical movement. In general, using fewer electrodes for such a system makes it easier to set up and less expensive. This study addresses the question of electrode selection for EEG-based P300 systems.

Method—Data from 13 subjects collected with a 16-electrode cap was analyzed. The optimal subsets of electrodes of sizes 1–15 were calculated for each subject and for the group as a whole. The methods of exhaustive search, forward selection, and backward elimination were then compared to each other and to these optimal subsets.

Results—The results show that, while none of the methods consistently picked the best-performing electrode subsets, all methods were able to find small electrode subsets that provided acceptable accuracy both for individuals and for the whole group. The computationally intensive exhaustive search method provided no statistically significant increase in performance over the much quicker forward and backward selection methods.

Conclusions—The forward and backward selection methods are preferred for electrode selection.

Keywords

Brain-Computer Interface; Event-Related Potentials; P300 Speller; Channel Selection

*Corresponding author. Tel.: +1 734 476 8640; fax: +1 412 268 9580.

¹Present address: Center for Bioimage Informatics, C119-122 Hamerschlag Hall, Carnegie Mellon University, 5000 Forbes Ave. Pittsburgh, PA 15213, USA

Declaration of interest

The authors report no conflicts of interest.

II. Introduction

The P300-based brain-computer interface (BCI) paradigm is designed to allow a user to select letters without physical movement. In a typical P300-BCI setup, e.g. [1], a user looks at a grid of randomly flashing letters and counts the flashes of a desired letter. Each flash of the desired letter generates a P300 in the user's EEG that the BCI recognizes and that enables identification of the desired letter. EEG-based P300-BCIs have used as few as 1 electrode [1], and as many as 64 electrodes [2, 3]. The number of EEG electrodes used directly affects the cost and setup time of these systems; larger numbers of electrodes require more expensive amplifiers and more time to achieve the appropriate electrical impedance for each electrode during setup. Minimizing the number of electrodes, while still ensuring adequate performance, is therefore advantageous.

Feature selection has been studied in the context of BCIs [4], but that work focused on reducing the number of features (dimensionality reduction) to improve classification rather than reduction of the number of electrodes. While reducing the number of features can impact BCI performance, it has little direct impact on the cost of BCI equipment or the time required to setup the BCI. Electrodes represent the physical observation points on the scalp while features refer to the particular characteristics within the EEG that are thought to be of importance. Thus, multiple features may be drawn from a single electrode (e.g. representing EEG at different time points). These features are then assigned importance values (weights) representing their importance within the final BCI classifier. While some BCI classifier algorithms (e.g. SWLDA) may eliminate features that are found to be unimportant, these algorithms are not typically designed to eliminate entire electrodes. Further, even if a P300-BCI classifier does not use any features from an electrode, the experimenters are typically unaware that the particular electrode is unnecessary to the BCI setup and therefore do not gain a benefit in cost or setup time.

Previous work [3, 5] has compared pre-selected electrode subsets (sizes 3, 3, 6, and 19; and 4, 8, 16, and 32, respectively) in the P300-BCI paradigm. However, to our knowledge, no large-scale, systematic comparison of electrode selection methods for a P300-BCI has been carried out. While specific electrode subsets such as these may be generally effective in people without physical impairments, the inability to identify a P300-BCI configuration that works for all subjects shown in [6] indicates that even within physiologically appropriate electrode locations, subject-to-subject variations may lead to variations in preferred electrode subsets. Further, the use of BCIs by people with conditions such as cerebral palsy, stroke, and multiple sclerosis, in which impairment is a direct result of damaged brain tissue, may be incompatible with standard electrode locations and require user-specific electrode subsets for which selection methods are needed.

Evaluating all possible electrode subsets is an obvious and attractive approach because it is guaranteed to find the best possible performance on the training data. However, the computing power necessary to carry out such an analysis, termed an exhaustive search algorithm, quickly becomes impractical as the number of electrodes increases. Exhaustive search requires testing 2^n possible electrode subsets. For the 16 electrode caps used by many groups, there are $2^{16} = 65,536$ possible subsets to evaluate for each subject. Exhaustive

search of all possible subsets of 32 electrodes is impractical because it would require evaluation of over 4.2 billion subsets per subject. A recent study, [7], indicated the effectiveness of backward elimination for electrode subset selection.

The current study addresses the question of how reducing the number of electrodes used affects P300-BCI accuracy and compares methods for selecting reduced electrode subsets for individuals as well as groups of users. The optimal electrode subsets out of 16 possible electrodes are calculated for each of 13 subjects, as well as for the group as a whole. Because the focus is solely on the impact of the number of available electrodes, subject-specific classifiers are calculated for each of these subsets. Three electrode subset selection methods are then compared in terms of their ability to find these electrode subsets based only on a restricted amount of training data.

III. Materials and methods

A. Subjects and data collection

This study presents an offline analysis of data collected according to the protocol described in [8]. Briefly, data were collected from 13 subjects, (n=6 male, 7 female), with mean age 30.5 ± 13.5 years. Participants signed consent forms approved by the appropriate institutional review boards (IRBs). Subjects wore a 16-electrode EEG cap (with electrodes at positions F3, Fz, F4, T7, C3, Cz, C4, T8, CP3, CP4, P3, Pz, P4, PO7, PO8, and Oz in the modified 10–20 system) with right mastoid reference and left mastoid ground. These electrode locations are commonly used in P300 BCI research. They include, for example, the single electrode used in [1] and 16 of the 19 electrodes used in [3]. During setup, electrode impedances below 10 kOhms were established. The signals were amplified using a g.USBamp (g.tec Medical Engineering, Austria) and digitized at 256 Hz. The BCI2000 system ([9]) controlled data collection and stimulus presentation.

Subjects selected characters from a 6 by 6 P300-BCI matrix in three sessions. The first session began with a 19-character long training run with the number of sequences set to 15 (that is, each row and column was intensified 15 times before a character was selected) and no feedback. The least squares method (LS, as implemented by the P300GUI utility included with BCI2000) was used to generate a set of linear classifier weights from all 16 electrodes for feedback during subsequent runs and to select the associated number of stimulus sequences as described in [8]. The number of sequences selected for individual subjects ranged from 4 to 10.

In each of the three sessions, subjects selected characters to form 3 sentences, each 23 characters in length. Subjects used a BACKSPACE selection to correct any mistakes they made. Counting corrections, these sessions contain an average of 97.4 ± 26.4 selections.

B. Data analysis

Every possible electrode subset (65,534 subsets) for each of the subjects was evaluated offline. The computation proceeded as follows: for each subject, and every possible subset, a unique set of weights was generated from the initial training run. Features consisted of the 800-ms of EEG following each flash decimated by a factor of 20. Weight generation was

carried out twice, once using stepwise linear discriminant analysis (SWLDA) ($p_{\text{enter}} = 0.1$, $p_{\text{remove}} = 0.15$, max features = 60), and once using LS (as used during data collection). Both methods were implemented using the P300GUI utility provided along with BCI2000. All further analysis was, therefore, also performed for both methods of weight generation. These methods were chosen because they are commonly used for P300 speller classification. Note that while the SWLDA itself carries out feature selection which might eliminate all features from a given electrode, we specified the maximum allowed subset size.

Training accuracy was recorded for every subset of the 16 electrodes. Training accuracy was defined as the number of correctly selected characters divided by the total number of selections in the training run.¹ Each of these 65,536 sets of weights was then used for offline analysis on the subject's copy-spelling data and the number of correctly selected characters was recorded. Copy-spelling accuracy was defined as the number of correct selections divided by the total number of selections in all copy-spelling data.

Calculations for each weight generation method took about 40 hours in total and were carried out using a high-performance computing cluster provided by the University of Michigan Center for Advanced Computing. This computation produced, for each subject, training and copy-spelling accuracies for every possible electrode subset at each size level (1, 2, 3, 4, ..., 15 electrodes) for both the SWLDA and LS methods. The optimal subset for each subject was the electrode subset with the highest accuracy on the subject's evaluation data, and the optimal subset for the subject group was the electrode subset with the highest average accuracy across the evaluation data for the group of subjects.

C. Data divisions for subset selection and evaluation

For purposes of evaluating the subset selection methods, data was divided into selection data and evaluation data. The subset selections methods used selection data to pick subsets and the accuracy of those subsets was tested on the evaluation data. Two divisions of the available data were used to represent two different real-world scenarios of electrode subset selection.

In the within-subject scenario, a new subject enters the laboratory, completes a training session, and then a custom electrode cap is designed for that subject, who returns when the custom cap is available. To simulate this scenario, for each subject, the selection data was the subject's training run and the evaluation data was the subject's copy-spelling runs.

In the cross-subject scenario, data from a large group of subjects is used to select a common electrode cap containing a particular electrode subset. That cap is then used for future subjects, with subject-specific classifiers created using the available electrodes. To simulate this scenario, for each subject, the electrode selection data was chosen to be all other subjects' copy-spelling runs and the subject's own copy-spelling data and associated classifier for the chosen electrode subset was used as the evaluation data. To determine what subsets would be picked for a hypothetical future subject, based on all the copy spelling data

¹Ties in training accuracy were broken by comparing accuracy at a reduced number of sequences (flashes). In the cases where a tie remained, the tie was broken randomly.

we have collected so far, we also applied the cross-subject method with all subjects' copy spelling data as electrode selection data.

D. Subset selection methods

Three subset selection methods were compared: the exhaustive search method, the forward selection method, and the backward selection method.

1) Exhaustive search—At each size level, the exhaustive search method picks the electrode subset with the highest accuracy on the selection data. This approach relies on finding training accuracy for each of the possible subsets and requires testing on the order of 2^n subsets, where n is the number of electrodes to pick from. Note that this method will find the best-performing subset on the selection data, but does not necessarily generalize to the evaluation data; thus it is not guaranteed to find optimal subsets.

2) Forward selection—The forward selection method (details can be found in [10] or other texts) belongs to the class of algorithms called greedy algorithms, so named because these algorithms assume that a good overall solution may be found by making a series of (easier) local optimizations. The forward selection method therefore adds new electrodes to its subset one-by-one, such that the subset of size $N+1$ always contains all the electrodes from the subset of size N . At each step, the next electrode to be added is the one that improves the accuracy on the selection data the most. This approach does not require the full computation outlined in the data analysis section and only requires testing on the order of n^2 subsets.

3) Backward elimination—The backward elimination method ([10]) is also from the class of greedy algorithms, and proceeds like the forward selection method except that the size 15 subset is selected first and electrodes are removed one-by-one. At each step, the electrode to be removed is the one that results in the smallest loss of accuracy on the selection data when removed. This approach also requires testing on the order of n^2 subsets.

E. Statistical analysis

Differences in accuracies between the methods and data divisions were tested using a 3-way repeated measures analysis of variance (ANOVA) with 15 levels of set size (1 through 15 electrodes), 3 levels of method (exhaustive search, forward selection, and backward elimination), and 2 levels of data division (within- and cross-subject). In addition, the results were compared to the accuracy achieved by the optimal subset for each subject and size level. This comparison was another 3-way repeated measures ANOVA with these optimal subsets included as an additional level of method.

IV. Results

For brevity, we present only data from the SWLDA weights. LS results are noted when they had differences in statistical significance.

A. Comparison of subset selection methods and data divisions

Table 1 shows accuracies on the evaluation data for each of the subset selection methods at several electrode subset sizes and illustrates the general equivalence of results produced by the different methods. The 3-way ANOVA showed that the main effect for subset size was significant ($F(14, 168) = 266.51, p < .0001$) with larger sets performing better. The main effects for method of electrode subset selection ($F(2, 24) = 2.96, p = .071$) and data division ($F(1, 12) = 1.48, p = .25$) were not significant.

The only significant interaction was between subset size and data division ($F(2, 24) = 5.47, p < .0001$). Post hoc testing revealed that the within-subject division resulted in significantly higher accuracies at size level 1 ($t(168) = 6.64, p < .0001$) while the cross-subject division resulted in significantly higher accuracies at size levels 3, 4, and 5 ($t(168) = 2.82, 3.49, 3.16, p = .0054, .0006, .0019$, respectively). The LS results showed a similar trend, with the within-subject division resulting in higher accuracies at size levels 1 and 2 ($t(168) = 10.09, 3.49, p < .0001, p = .0006$, respectively) and the cross-subject division resulting in higher accuracies at size levels 4, 5, 6, and 7 ($t(168) = 3.77, 4.05, 2.58, 2.71, p = .0002, p < .0001, p = .01, .007$, respectively).

When the optimal subsets were added into the comparison, the main effect for method became significant ($F(3, 36) = 107.28, p < .0001$) with the differences of means between the accuracy attained by the electrode section methods and optimal accuracy being 7.5% for exhaustive search, 8.3% for forward selection, and 7.0% for backward elimination. Thus, none of the selection methods finds subsets that perform optimally on the evaluation data, but the average difference between the methods is small (backward elimination has average percent accuracy 0.5 better than exhaustive search and 1.3 better than forward selection).

B. Cross-subject subset for a new subject

Figure 1 shows the results of electrode selection on the cross-subject data division for all three electrode selection methods. These results show the average accuracy of an electrode cap that is selected based on data from a population of subjects and then used to calculate a classifier for a never-before-seen subject. Figure 2 shows the electrode subsets that would be selected for a hypothetical new subject based on the copy-spelling data from all 13 subjects in the study (exhaustive method, other methods chose similar subsets). These subsets are optimal for our subjects and should generalize to a new subject with accuracies as in Figure 1.

V. Discussion

The comparison of subset selection methods show that none of the three methods evaluated provide significantly higher accuracy than the others. Although the exhaustive search method always finds the optimal subsets for the selection data, it generalizes to the evaluation data no better than the other methods. In fact, the best generalization was achieved by the backward elimination method, though the differences are both small and statistically insignificant.

In the absence of major differences in performance, the greedy algorithms are preferable because they have polynomial rather than exponential complexity, meaning that they are more feasible to apply to larger electrode selection problems (e.g. selecting subsets from a 64-electrode cap). Using one of the greedy algorithms to select a subset from a 64 channel cap would require evaluating $64 + 63 + 62 + \dots + 1 = 2080$ electrode subsets, while an exhaustive search of possible electrode subsets would require evaluating $2^{64} = 1.84 \times 10^{19}$ electrode subsets.

All three methods perform better at higher subset size levels, which is the expected result because larger subsets should have more useful information. It is important to note, however, the asymptotic behavior of this trend, i.e. differences in accuracy become relatively minor above size level 4. In addition, the performance of small electrode subsets was usefully high for most subjects. For example, using the backward elimination method on within-subject data provided subsets of size 4 that performed adequately ($>70\%$) for 10 of 13 subjects.

The effect of training scenario was only significant for subsets under size 8. The results indicate that subject-to-subject variation was important for single- and double-electrode subsets, but that using information from multiple subjects yielded better accuracies at larger subset size levels. Analysis of data from larger groups of subjects may be able to identify subgroups (e.g. men or women) that have more consistent single- and double-electrode subsets.

The cross-subject selection results show the subsets that are most appropriate for the normal subjects studied. The prevalence of the parietal-occipital electrodes Pz, PO7, PO8, and Oz across all electrode subsets is consistent with [10, 11] and [13] in indicating that these electrodes are important for P300-BCI classification. The best of the 4 subsets compared in [3] was a 6-electrode subset that was nearly identical to the 6-electrode subsets selected in the current study. However, this electrode subset is not necessarily appropriate for people with impairments who would actually benefit from a BCI. Some conditions could possibly impact the brain in consistent ways so that common electrode subsets could be found. For example, subjects who cannot control eye movements may have different EEG patterns than those who can ([14]). However, the wide variety of disabling conditions accompanied by known brain abnormalities and the potential for neuroplastic reorganization of the brain tissue emphasizes the importance of identifying methods for selecting electrode subsets based on subject-specific data.

VI. Conclusions

This study presents a comparison of basic methods of subset selection for P300-BCI classification. The results show that exhaustive search, forward selection, and backward elimination all perform comparably. While none of the methods consistently selected optimal electrode subsets, they all selected subsets with useful accuracy even when as few as 4 electrodes were used. These small electrode subsets offer an attractive reduction in equipment cost and setup time for P300-BCIs.

In future work, we would like to apply a similar analysis to data from subjects with a specific motor impairment, such as amyotrophic lateral sclerosis (ALS). We expect electrode selection to be an important consideration for these potential BCI users because their condition is likely to reduce the effectiveness of the standard electrode sets.

Acknowledgment

The authors would like to thank the attendees of the Fourth International BCI Meeting for their insightful poster comments, which were of great help while preparing this manuscript.

The project described was supported by Grant Number R21HD054913 from the National Institute Of Child Health And Human Development (NICHD) in the National Institutes of Health (NIH). Any opinions, findings, conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of NICHD or NIH.

References

1. Farwell LA, Donchin E. Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr Clin Neurophysiol*. 1988; 70:510–523. [PubMed: 2461285]
2. Krusienski DJ, Sellers EW, Cabestaing F, Bayouth S, McFarland DJ, Vaughan TM, Wolpaw JR. A comparison of classification techniques for the P300 speller. *J Neural Eng*. 2006; 3:299–305. [PubMed: 17124334]
3. Krusienski DJ, Sellers EW, McFarland DJ, Wolpaw JR. Toward enhanced P300 speller performance. *J Neurosci Methods*. 2008; 167:15–21. [PubMed: 17822777]
4. Dias N, Kamrunnahr M, Mendes P, Schiff S, Correia J. Feature selection on movement imagery discrimination and attention detection. *Comput Biol Med*. 2004; 48:331–341.
5. Hoffmann U, Vesina J-M, Ebrahimia T, Diserensb K. An efficient P300-based brain-computer interface for disabled subjects. *J Neurosci Methods*. 2007; 167:115–125. [PubMed: 17445904]
6. Sellers EW, Krusienski DJ, McFarland DJ, Vaughan TM, Wolpaw JR. A P300 event-related potential brain-computer interface (BCI): The effects of matrix size and inter stimulus interval on performance. *Biological Psychology*. 2006; 73:242–252. [PubMed: 16860920]
7. Cecotti H, Rivet B, Congedo M, Jutten C, Bertrand O, Maby E, Mattout J. A robust sensor-selection method for P300 brain-computer interfaces. *J Neural Eng*. 2011; 8
8. Thompson DE, Baker JJ, Sarnacki WA, Huggins JE. Plug-and-play braincomputer interface keyboard performance. *Proceedings of the 4th International IEEE/EMBS Conference on Neural Engineering*. 2009; 433–435
9. Schalk G, McFarland DJ, Hinterberger T, Birbaumer N, Wolpaw JR. BCI2000: A general-purpose brain-computer interface (BCI) system. *IEEE Trans Biomed Eng*. 2004; 51:1034–1043. [PubMed: 15188875]
10. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed.. New York: Springer; 2009. 2009.
11. Blankertz B, Muller KR, Curio G, Vaughan TM, Schalk G, Wolpaw JR, Schlogl A, Neuper C, Pfurtscheller G, Hinterberger T, Schroder M, Birbaumer N. The BCI competition 2003: progress and perspectives in detection and discrimination of eeg single trials. *IEEE Trans Biomed Eng*. 2004; 14:153–159.
12. Blankertz B, Muller KR, Krusienski DJ, Schalk G, Wolpaw JR, Schlogl A, Pfurtscheller G, Millan JR, Schroder M, Birbaumer N. The BCI competition III: validating alternative approaches to actual BCI problems. *IEEE Trans Neural Syst Rehabil Eng*. 2006; 51:1044–1051.
13. Kaper M, Meinicke P, Grossekhoefer U, Lingner T, Ritter H. BCI competition 2003-data set IIb: support vector machines for the P300 speller paradigm. *IEEE Trans Biomed Eng*. 2004; 51:1073–1076. [PubMed: 15188881]
14. Brunner P, Joshi S, Briskin S, Wolpaw JR, Bischof H, Schalk G. Does the 'P300' speller depend on eye gaze? *J Neural Eng*. 2010; 7

Implications for rehabilitation

- > A P300 speller is a type of brain-computer interface that allows a user to select characters without physical movement.
- > Using fewer electrodes reduces setup time and cost for an EEG-based P300 speller.
- > We show that acceptable P300 speller performance can be achieved with as few as 4 electrodes.
- > We compare methods of selecting electrode sets and identify fast and efficient methods for customizing electrode sets for individuals.

Figure 1.

Results from the cross-subject data division using the exhaustive method and SWLDA weights, with errorbars representing standard deviation. These results serve as a leave-one-out cross validation of the accuracy of the subsets presented in Figure 2.

Figure 2.

The subsets that would be selected for a new subject based on data from all 13 subjects (again, exhaustive method with SWLDA weights). Small black dots indicate positions in the 10–20 system, grey dots represent electrodes which were used in the study, and large black dots show the electrodes chosen for the subset.

Table 1

Summary of results at select size levels using SWLDA weights. The top of the chart shows the average accuracy (in percent) of the subsets selected by each method under both data divisions. The bottom row shows the accuracy of the optimal electrode subsets. For comparison, note that the accuracy attained when all 16 electrodes were used was $80 \pm 17\%$.

Data division	Method	Number of electrodes						
		1	2	3	4	5	8	15
Within-subject	Exhaustive search	37±19	55±22	65±16	71±17	69±21	77±16	80±18
	Forward selection	37±19	52±23	58±24	65±24	71±20	78±17	82±15
	Backward elimination	32±15	51±17	65±20	72±21	76±20	79±18	80±18
Cross-subject	Exhaustive search	22±10	52±23	69±16	76±20	78±18	79±19	80±17
	Forward selection	22±10	55±20	64±17	74±20	76±20	77±20	80±18
	Backward elimination	27±17	49±23	69±16	76±19	78±19	78±17	80±17
Optimal electrode subset		41±18	65±18	74±15	81±14	83±14	86±13	83±14