SINGLE TRIAL BCI OPERATION VIA WACKERMANN PARAMETERS

Ian Daly, Nitin Williams, Slawomir J. Nasuto, Kevin Warwick, Douglas Saddy

University of Reading, Reading, UK

ABSTRACT

Accurate single trial P300 classification lends itself to fast and accurate control of Brain Computer Interfaces (BCIs). Highly accurate classification of single trial P300 ERPs is achieved by characterizing the EEG via corresponding stationary and time-varying Wackermann parameters. Subsets of maximally discriminating parameters are then selected using the Network Clustering feature selection algorithm and classified with Naive-Bayes and Linear Discriminant Analysis classifiers.

Hence the method is assessed on two different data-sets from BCI competitions and is shown to produce accuracies of between approximately 70% and 85%. This is promising for the use of Wackermann parameters as features in the classification of single-trial ERP responses.

Index Terms— Brain Computer Interfaces, Wackermann parameters, single-trial classification, event-related potentials, Network Clustering

1. INTRODUCTION

Brain Computer Interfaces (BCIs) provide a novel means for control of a computer system via the recording of a user's neural signals and the transformation into control commands for software or devices. As such they provide a way for a user to control a computer directly from the signals extracted from their brain without the need for muscle movement. They therefore allow users who would not otherwise be able to control a computer - such as those suffering from paralysis, Amyotrophic Lateral Sclerosis (ALS) etc. - to do so; retaining some control over their environment and ability for communication [1].

Electroencephalography (EEG) is a commonly used technique for obtaining the signals used for BCI control. Electrodes are placed over the surface of the user's scalp and referenced to one or more reference channels. The electrophysiological activity generated by the simultaneous firings of large numbers of cortical neurons may then be recorded by sampling between frequencies of 250 Hz to 2 kHz. The recorded signal is of the order of microvolts and after amplification and processing, can be used for BCI control.

The Event Related Potential (ERP) is defined as a change in the EEG in response to the presentation of particular stimuli or events and emerges in the average of multiple recordings of EEG after repetitions of that same event. The P300 ERP is a particular type of ERP, commonly used in BCI control, which occurs in response to the presentation of a rare and predefined stimuli such as an "oddball" event in a sequence. When such an oddball event is presented to the subject, a P300 - characterized by an increase in amplitude in the averaged waveform compared to pre-stimulus activity lasting for approximately 200ms - is expected to occur approximately 300ms later and be most prominent in the signals recorded on electrodes placed over the parietal cortex.

BCI control may be achieved via the P300 by presenting the user with a series of options. The user is instructed to select one option and concentrate on it. The options are then displayed to the user in quick succession in a randomized order - often via highlighting each choice individually on a screen displaying all choices - and the user concentrates on their chosen option. When the user's chosen option is highlighted one can expect a P300 event to occur in the recording of the users EEG over the parietal cortex approximately 300ms after stimulus presentation.

There are a number of factors that make detection of the ERP difficult. In particular there is a great deal of background noise present in the EEG which is uncorrelated to the ERP but generally of equal or greater amplitude and which may mask the ERP making single trial detection very difficult. Therefore the traditional method for detecting the presence of an ERP such as the P300 is to record EEG after several presentations of the trigger stimulus to the subject, interspersed with distractor stimuli, take the average of the recordings and look for a shift in amplitude that differs relative to baseline. For BCI purposes this means that every possible option has to be highlighted to the user several times in order for a sufficient number of recordings to be made for the ERP to be detectable in the average.

This need to present each choice to the subject multiple times makes the operation of the BCI much slower and can lead to a frustrating experience for the user [2]. The averaging technique is also not guaranteed to be 100% accurate due to inter-trial variability and this may result in misclassifications of the user instructions resulting in poorer quality of communication and control. Additionally there is a growing desire to develop BCIs for able bodied subjects

for uses such as computer game control. Considered together, there is a strong need for accurate BCI control on a single-trial basis for both disabled and able-bodied subjects.

To address these problems this paper introduces a method to achieve accurate ERP based BCI control via the detection of P300 events on a single trial basis. Wackermann parameters are used to describe multivariate EEG time series as a trajectory in state space.

The spatial distribution of potentials at a given moment in time may be described by either traditional analysis in the time domain or by analysis in the state space domain. Wackermann parameters aim to characterize the potential field across the scalp at particular times using various state space measures and - unlike time domain descriptors of the data - takes into account correlations between different electrodes.

The state-space trajectories of epochs and sub-epochs which best differentiate the presence of a P300 event are identified via a filter based feature selection method. The selected features are then used to accurately classify P300 events on a single trial basis.

Section 2 first introduces Wackermann parameters, the data they will be applied to and the feature selection method that will be used with them before section 3 describes which parameters are selected as features of interest and the accuracies achieved with them. Section 4 then discusses the results and their application to single trial based BCI control before section 5 summarizes the approach taken and results achieved.

2. METHODS

2.1. Wackermann parameters

Wackermann parameters aim to characterize the evolution of multichannel EEG in state-space via the use of three parameters Σ, Φ and Ω [3].

For a period of multichannel EEG data *K* denotes the number of electrodes recorded, *N* the number of samples recorded on each electrode, Δ*t* the length of time step in the recording and f_{samp}, the sampling rate. The complete period of EEG recorded over *T* seconds is composed of *N* 'voltage vectors' where, $(N = \lfloor T f_{samp} \rfloor)$, u_n^k for all $n \in [1, \dots, N]$ and for all $k \in [1, \dots, K]$.

The EEG data must first be centered at zero so that the mean potential on all channels is zero ($\frac{1}{N}\sum_{n=1}^{N} u_n^i = 0$ for all $i \in [1, \dots, K]$). It must also be transformed to use average referencing such that the mean across channels is also zero ($\frac{1}{K}\sum_{n=1}^{K} u_n^i = 0$ for all $n \in [1, \dots, N]$). Two integral moments are then estimated

$$m_0 = \frac{1}{N}\sum_{n=1}^{N} \|u_n\|^2 \qquad (1)$$

$$m_1 = \frac{1}{N-1}\sum_{n=1}^{N-1} \|\Delta u_n\|^2 \qquad (2)$$

where $\Delta u_n \equiv u_n - u_{n-1}$

Two descriptors may then be defined

$$\Sigma = \sqrt{\frac{m_0}{K}} \qquad (3)$$

$$\Phi = \frac{1}{2\pi\Delta t}\sqrt{\frac{m_1}{m_0}} \qquad (4)$$

Σ^2 denotes the mean square of the global field power integrated over the length of the EEG epoch. Φ is referred to as a 'general frequency measure' and corresponds to the circular frequency of periodical field changes. The final parameter Ω provides a global measure of spatial synchronization and is defined from the eigen-decomposition of the *K* x *K* covariance matrix

$$C = \frac{1}{N}\sum_n u_n u_n^T \qquad (5)$$

Eigen decomposition of this covariance matrix produces eigen values $\lambda_1, \dots, \lambda_k$ with trace $T \equiv trC. \sum_i \lambda_i$ gives the total variance. After normalization

$$\lambda_i = \frac{\lambda_i}{T} \qquad (6)$$

the descriptor Ω may then be defined as

$$\log \Omega = -\sum_i \lambda_i \log \lambda_i \quad (7)$$

The parameters Σ, Φ and Ω may also be used to analyze the temporal dynamics of the EEG epoch in state space. The EEG is segmented into sub-epochs and for each of these the Wackermann parameters are calculated. Thus the temporal evolution of the Wackermann parameters may also be used as potential features to describe the data.

Wackermann parameters have been used to classify EEG into sleep and wakefulness classes and to group EEG by pharmacological effects and by correlates of sensory and motor processes [4]. However to the best of our knowledge they have not yet been applied to assist with the classification of BCI control / ERP data.

2.2 Network Clustering

For a dataset of M trials recorded under two or more different conditions each trial may be represented by J features. Features may describe the data in a number of different ways such as its representation in the time and/or frequency domains; in this study they contain the Wackermann parameters - the vectors Σ, Φ and Ω. A feature vector contains the values a feature j takes across all trials [1,...,M]. It is important to note that there will be J+1 feature vectors where the (J+1)'th feature vector contains the class labels. This feature vector will hence be referred to as the class label vector (Cl).

The range of values contained in each feature vector is coarse grained into a number of partitions equal to the number of classes in the data set. This can be done via, for example, a k-means clustering algorithm which aims to minimize intra-cluster variance.

Each cluster of values is assigned a label and for each feature vector each value is replaced by its cluster label. This gives rise to a labeled feature vector L_j

It is now possible to define a weighted network of feature vectors where each feature vector is a node in the network and weights assigned to the edges are the inverse of the distances between these nodes measured by the mutual information.

Mutual Information (MI) is a measure of dependencies between random variables, X = [$X1$,$X2$,...,XN]. It acts as a measure of the amount of information one random variable gives about another and is defined as

$$MI(X) = \sum_{i}^{N} H(X_i) - H(X_1, X_2, \dots, X_N) \quad (8)$$

where $H(X)$ is entropy, $H(X_1, X_2, \dots, X_N)$ is the joint entropy of $[X_1, \dots, X_N]$. The entropy of a discrete random variable is defined as

$$H(X) = -\sum_{i=1}^{M_x} p(x_i) \log\left(p(x_i)\right) \quad (9)$$

where $p(x_i)$ is the probability of x_i and X contains values from the set $x_1, \dots, x_{M_x}$. MI is ≥0; MI equals zero if the variables are independent.

It is now possible to identify informative features by employing spectral clustering to identify feature vectors of interest [5]. One of the groupings obtained as a result will contain the class label vector. The elements of this grouping therefore define the features which are most closely associated with the class labels. Thus these features are taken to be the best features to allow classification of the different trials.

Eigen decomposition is applied to the Graph Laplacian matrix to find a set of eigenvalues and eigenvectors, defined as

$$Av_k = \lambda v_k \quad (10)$$

where $A = [a_{ij}]_{((J+1)x(J+1))}$ is the Graph Laplacian matrix, λ_k the eigenvalues for $k \in [1, \dots, (J+1)]$ and v_k the eigenvectors, each of dimension $(J+1)$.

Eigenvalues are then sorted in descending order and the number of eigenvalues in the top 10th percentile gives Kc the number of clusters. The top 10th percentile was selected based on tests on simulated data during development of the method. The eigenvectors that correspond to these Kc eigenvalues form a new matrix X and each row of X is re-normalized to have unit length. Each row in X is then treated as a point in $R^{Kc}$and rows are clustered into Kc clusters via the k-means algorithm. The cluster which contains the class label also contains the indices of the features which may be best at discriminating between the conditions under which trials were recorded.

The process can be summarized as follows.

1. Coarse grain the values each feature takes across all trials into C clusters where C is the number of different conditions under which the trials were recorded.

2. For each feature; form labeled feature vectors by replacing the values the features take across trials by their corresponding cluster labels.

3. Formulate the Graph Laplacian matrix

4. Perform Eigen-decomposition on the Graph Laplacian matrix

5. Apply spectral clustering; select Kc, the number of clusters, from the number of eigenvalues in the top 10th percentile and select features from the indices of the eigenvectors which cluster with the class labels.

2.3. Data

Our chosen method is to calculate the Wackermann parameters (Σ, Φ and Ω) on a set of EEG trials some of which are expected to contain a P300 event and some of which are not. Network Clustering is then applied to identify the subset of parameters which best differentiate between P300 present vs P300 absent conditions. To evaluate the performance of this method two data-sets are used; data from the Machine Learning in Signal Processing (MLSP) 'Mind reading' competition and data from the BCI data-set supplied by Ulrich Hoffmann [6] are both used to evaluate the performance of the method. Both EEG data-sets are first segmented into trials of length 750ms starting from 250ms before stimulus presentation and continuing until 500ms after presentation. The trials are then band-pass filtered between 1-30Hz and baseline corrected. The MLSP data is

recorded from a single subject as they view a series of satellite images some of which contain images of Surface to Air Missile (SAM) sites. As the images containing SAM sites are in the minority they are an oddball stimuli and subsequently a P300 event is expected in response to the subjects spotting them.EEG is recorded from 64 electrodes positioned according to the extended International 10/20 system at a sampling rate of 256Hz. Stimuli are presented in 75 blocks each containing 37 satellite images, thus a total of 2,775 satellite images are presented to the subject. Each image is 500 x 500 pixels in size and presented on an LCD screen approximately 43 cm in front of the subject. All images within a block are presented 100ms after each previous image with no inter-stimulus pause in presentation of images within the blocks. The subject was instructed to carefully watch the images and look out for SAM sites. Upon the identification of a SAM the subject was instructed to press the space bar. The subjects also pressed the space bar to begin a new block of presentations of 37 images. Thus two neural signatures are expected; one in response to SAM sites - the P300 - and the other in response to the subject pressing the space bar.

2.3.1. BCI data-set

The other data-set, taken from BCI related research in [6], is recorded at a sampling rate of 2048Hz from 32 electrodes as a subject views a series of images presented in random order. Six different colour photos of common household objects are shown to the subject sequentially in random order and the subject is instructed to lookout for a particular target image. Each image was displayed for 100ms with inter-stimulus intervals of 400ms. When the target image is shown a P300 event is expected. The target image is changed between runs of stimuli presentation with a total of 6 runs of different stimuli being presented to the subject.

2.4. Evaluation

The Naive Bayesian (NB) and Linear Discriminant Analysis (LDA) classifiers are used for classification of the selected feature sets.

Classification accuracy is evaluated via the application of a leave-one-out training and classification scheme. From data-sets of 50 target and 50 non-target trials training datasets are created from 49 target trials and 49 non-target trials. Classification algorithms are trained on these data-sets and then applied to classify the remaining left out target and non-target trials which were not included in the training. This procedure is repeated 50 times, with every target trial being tested once while the tested non-target is randomly chosen from the group of 50. The sum of correct classifications after the 50 passes is reported as the classification accuracy.

To be able to compare our analysis with that of previous studies using multiple repetitions, we used subject 6 from the BCI dataset. Data from first two sessions were training data while last two sessions were test data. In addition to single-trial classification accuracies, accuracies from averages of 2 to 10 trials was calculated. With the number of repetitions giving maximum classification accuracy, the same analysis was done on subject 7, subject 8 on the BCI data set to assess the ability of Wackermann parameters to be robust to individual differences.

Classifier	LDA	NB
Accuracy	0.693	0.779

Table 1. Accuracies achieved by the LDA and NB classifiers on feature sets identified from Wackermann parameters extracted from target and non-target trials by NC based feature selection on the MSLP data-set. All results are statistically significant ($p < 0.05$).

Classifier	LDA	NB
Accuracy	0.8211	0.8534

Table 2. Accuracies achieved by the LDA and NB classifiers on feature sets identified from Wackermann parameters extracted from target and non-target trials by NC based feature selection on the data-set from [6]. All results are statistically significant ($p < 0.05$).

3. RESULTS

The selected features for classifying P300's in both data-sets are from the Σ and Φ parameter vectors. The majority of the selected features are from the time dependent parameters around 300ms after stimulus presentation, although time-invariant Wackermann parameters are also selected.

Table 1 lists the accuracies achieved by the LDA and NB classifiers on the selected Wackermann parameters on the MLSP data-set. As can be seen NB classification performs better then LDA classification.

Table 2 lists the accuracies achieved via the same methods on the BCI data-set from [6]. Higher accuracies are achieved on this data-set then on the MSLP data-set with both the LDA and the NB classifiers.

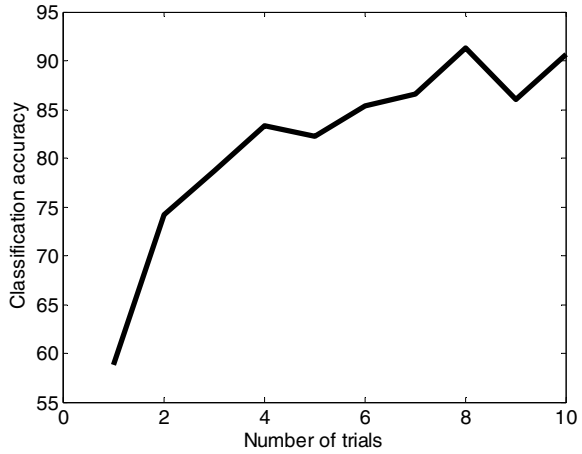

Figure 1. Plot of classification accuracy versus number of trials in average for subject 6 from BCI data set. Naive-Bayes classifier was used.

From Figure 1, it is also seen that classification accuracies generally increase with increasing number of repetitions, with a classification accuracy of 91% after 10 trials. For averages of 10 trials, high accuracies of 83.9% and 77% were achieved for subject 7 and subject 8 respectively, demonstrating the ability of Wackermann parameters to be robust to inter-subject variability.

4. DISCUSSION

Accurate single-trial classification of P300 ERPs is done via Wackermann parameters. The Network Clustering feature selection algorithm was used to select a combination of features which maximally differentiated trials with P300 and trials without a P300. These features were then used to classify single trials into the two groups, giving classification accuracies of 77.9% and 85.3% with a Naive Bayes classifier and 69.3% and 82.1% with LDA on the two different data-sets.

The selected features were Σ and Φ based features at different time points as well as stationary Σ and Φ parameters. Σ based features correspond to the effective voltage per channel while Φ based features correspond to the generalized frequency of the epoch.

Previous work has only attempted to classify averages of a few trials, owing to the extremely low SNR (signal-to-noise ratio) of single-trial responses. Most of these studies have involved off-line classification - for example, [7], [8] and [9], report results from classification of P300 data from the BCI competition 2003. All these classification schemes achieve close to 100% accuracy after 15 presentation of the target character, while [7] and [9] reported error free classification after only 4 presentations. Some other groups have attempted online classification within the framework of a P300 speller paradigm - [6] reported 100% classification on both able-bodied and disabled subjects, but only after the target image was presented 10 times. [10] and [11] both achieve near 100% classification on some subjects, but a minimum of 5 trials are required in [10] and at least 10 in [11]. After just a single presentation of the target stimulus, classification accuracies were in the range of only 30% for both [10] and [11]. However, the accuracy rates achieved with these studies are not directly comparable to our study since in these studies, approximately one in six stimuli were targets. Hence, the probability of correctly discriminating between target and non-target trials by chance was low. For our study, since number of targets and non-targets was equal, chance level was 50%. Table 3 lists accuracies achieved by the previous studies.

Figure 1 illustrates classification accuracy as a function of number of trials in the average. At a single-trial level, accuracy is lower than that obtained by the leave-one-out classification scheme which obtained a maximum accuracy of 85.3%. This might be because, in the former case, training and testing data were from different sessions while in the latter case, training and testing sets were from the same sessions. Hence, the lower accuracy might be a result of inter-session variability in the response. It is also seen that Wackermann-based classification on averages of multiple trials performs less well than previous studies on multiple trials. This might be because previous studies used combinations of different types of features – amplitude-based features, frequency-based features, auto-regressive features. Since our study looked at potential of Wackermann parameters, only Wackermann-based features was used. Since Wackermann features have been shown to be informative, they could be used in combination with existing features to potentially improve classification performance. High accuracies of subjects 6,7,8 from averages of 10 trials show that the accuracy achieved by Wackermann parameters is not subject-specific.

However, it is important to note that the usefulness of Wackermann parameters (specifically Σ and Φ based features selected by Network Clustering) has been demonstrated only on off-line classification of two P300 data-sets - a more thorough test of their usefulness should be done within the framework of an online BCI (such as a P300-based speller) on a number of subjects, before more general claims can be made.

Reference	Method	Accuracy	Number of trials
[6]	Bayesian Linear Discriminant Analysis	100%	10
[7]	Continuous Wavelet Transform	100%	4
[8]	Ensemble of SVMs	97%	15
[9]	P300 subspace estimation	100%	5
[10]	Linear feature extraction	100%	5
[11]	Principal Component Analysis	100%	10

Table 3. Comparison of typical accuracies achieved with multi-trial averaging approaches to ERP classification. Maximum accuracies are listed against the associated numbers of trials averaged together and the methods that achieved them.

5. CONCLUSION

We investigated the potential of Wackermann parameters, selected by Network Clustering, for detecting the P300 event-related response. The Network Clustering feature selection algorithm selected Σ and Φ based features as particularly discriminative. On classification with these features, high classification accuracies were obtained (77.9% and 85.3% using Naive-Bayes classifiers on the MLSP and BCI datasets respectively). These results are remarkable considering classification was done on a single-trial basis and hence promise to increase transfer rate and speed of operation in P300 based BCI.

6. REFERENCES

[1] Jonathan R Wolpaw, "Brain-computer interfaces as new brain output pathways.," *The Journal of physiology*, vol. 579, no. Pt 3, pp. 613–619, March 2007.

[2] Nicola Neumann and Andrea Kubler, "Training Locked-in Patients : A Challenge for the Use of Brain Computer Interfaces," *Rehabilitation*, vol. 11, no. 2, pp. 169–172, 2003.

[3] J Wackermann, "Beyond mapping: estimating complexity of multichannel EEG recordings.," *Acta neurobiologiae experimentalis*, vol. 56, no. 1, pp. 197–208, January 1996.

[4] J Wackermann, "Towards a Quantitative Characterisation of Functional States of the Brain: From the Non-Linear Methodology to the Global Linear Description," *International Journal of Psychophysiology*, vol. 34, pp. 65–80, 1999.

[5] Andrew Y. Ng, Michael I. Jordan, and YairWeiss, "On Spectral Clustering: Analysis and an algorithm," *Advances in Neural Information Processing Systems*, vol. 14, pp. 849–856, 2001.

[6] Ulrich Hoffmann, Jean-Marc Vesin, Touradj Ebrahimi, and Karin Diserens, "An efficient P300-based brain-computer interface for disabled subjects.," *Journal of Neuroscience methods*, vol. 167, no. 1, pp. 115–125, January 2008.

[7] Vladimir Bostanov and Boris Kotchoubey, "Recognition of affective prosody: continuous wavelet measures of event-related brain potentials to emotional exclamations.," *Psychophysiology*, vol. 41, no. 2, pp. 259–68, 2004.

[8] A. Rakotomamonjy, V. Guigue, G. Mallet, and V. Alvarado, "Ensemble of SVMs for Improving Brain Computer Interface P300 Speller Performances," *Artificial Neural Networks: Biological Inspirations ICANN 2005*, vol. 3696, pp. 45–50, 2005.

[9] Bertrand Rivet, Antoine Souloumiac, Guillaume Gibert, and Virginie Attina, "P300 Speller brain-computer interface: enhancement of P300 evoked potential by spatial filters P300," in *16th Euro Sig Proc Conf EUSIPCO-2008*, Lausanne, Suisse, 2008.

[10] Nikolay Chumerin, Nikolay Manyakov, Adrien Combaz, Johan Suykens, Refet Yazicioglu, Tom Torfs, Patrick Merken, Herc Neves, Chris Van Hoof, and Marc Van Hulle, "P300 Detection Based on Feature Extraction in On-line Brain-Computer Interface," *KI 2009: Advances in Artificial Intelligence*, vol. 5803, pp. 339–346, 2009.

[11] Manoj Thulasidas, Cuntai Guan, and Jiankang Wu, "Robust classification of EEG signal for brain computer interface.," *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 14, no. 1, pp. 24–9, March 2006.

.