12-31-2010

Advanced Signal Processing Techniques for Single Trial Electroencephalography Signal Classification for Brain Computer Interface Applications

Kun Li
University of South Florida

Advanced Signal Processing Techniques for Single Trial Electroencephalography Signal

Classification for Brain Computer Interface Applications

by

Kun Li

Keywords: P300, BSS, ICA, SWLDA, Blind Tracking, Variance Analysis

Dedicated to my parents

ACKNOWLEDGMENTS

I would like to take this opportunity to express my gratitude to my Major Professor, Dr. Ravi Sankar, for his invaluable guidance and support during the course of this work. He has been a great source of inspiration during my work here. Without him, this work would not have been possible.

I also thank Dr. Donchin, Dr. Arbel, Dr. Wang and Dr. Jaroszeski for their guidance in the fulfillment of this dissertation and graduate program and for their valuable advice and comments that helped me to improve the quality of my dissertation.

My special thanks go to Ismail Butun, Murad Khalid and all the iCONS group members, who gave me constant support, immense help and encouragement during this work.

TABLE OF CONTENTS

LIST OF TABLES

iii

LIST OF FIGURES

ABSTRACT

Brain Computer Interface (BCI) is a direct communication channel between brain and computer. It allows the users to control the environment without the need to control muscle activity [1-2]. P300-Speller is a well known and widely used BCI system that was developed by Farwell and Donchin in 1988 [3]. The accuracy level of the P300-BCI Speller as measured by the percent of communicated characters correctly identified by the system depends on the ability to detect the P300 event related potential (ERP) component among the ongoing electroencephalography (EEG) signal. Different techniques have been tested to reduce the number of trials needed to be averaged together to allow the reliable detection of the P300 response. Some of them have achieved high accuracies in multiple-trial P300 response detection. However the accuracy of single trial P300 response detection still needs to be improved. In this research, two single trial P300 response classification methods were designed. One is based on independent component analysis (ICA) with blind tracking and the other is based on variance analysis. The purpose of both methods is to detect a chosen character in real-time in the P300-BCI speller. The experimental results demonstrate that the proposed methods dramatically reduce the signal processing time, improve the data communication rate, and achieve overall accuracy of 79.1% for ICA based method and 84.8% for variance analysis based method in single trial P300 response classification task. Both methods showed better

performance than that of the single trial stepwise linear discriminant analysis (SWLDA), which has been considered as the most accurate and practical technique working with P300-BCI Speller.

CHAPTER 1

INTRODUCTION

1.1 Overview of Brain Computer Interfaces

Brain Computer Interface (BCI) is a channel established between the human brain and computer or computer controlled electronic devices for communication purpose. It can translate people's intent into meaningful action in the real world solely by processing their brain waves. Research on BCIs started in the 1970s at the University of California Los Angeles (UCLA) under a grant from the National Science Foundation [4-5]. After this research, J. Vidal expressed the brain computer interface in his papers [5] which was considered as the first appearance of *brain computer interface* in scientific literature. The BCI systems can be classified into invasive systems [6-36] and non-invasive systems devices [37-48] while the invasive systems include fully-invasive and partially-invasive systems. Fully-invasive BCIs utilize electrodes that are implanted directly into the grey matter of the brain during neurosurgery. They provide the best quality signals for measuring the P300 component. However the Fully-invasive BCIs come with high risk of brain surgery and the signal may become weaker or even lost due to the scar-tissue build-up caused by body rejection of foreign materials. Partially-invasive BCI devices are implanted inside the skull but rest outside the grey matter. Partially-invasive BCI reduces the risk of forming scar-tissue compared to invasive BCIs and produce relatively high

1

quality signals (better than non-invasive systems but inferior to than fully-invasive systems). Non-invasive BCI are designed to work on the surface of the scalp without any implanted electrodes. They are easy to apply and remove. As a trade-off, non-invasive BCIs produce the weakest signals with poor resolution because the electromagnetic waves generated by the neurons are dispersed and blurred by the skull. This makes determination of the signal generating area of the brain and the actions of individual neurons a great challenge. Although the quality of the signals provided by non-invasive BCIs are not as good as the signals provided by invasive and partially-invasive BCIs, non-invasive BCIs still play a very important role in BCI research because they are safer, simpler, more practical, and can be designed as portable with low cost. Therefore, if the signal quality problem can be solved by signal processing techniques, the non-invasive BCIs will become quite a promising choice for BCI system designing.

EEG signal is the recording of electrical activity along the scalp produced by the activity of neurons within the brain [49]. It can be reliably and easily collected and processed in real time. These attributes make the EEG based method *best suited* for designing a practical non-invasive BCI system. Guger *et al*. [50] conducted an experiment to verify that EEG is a useful and reliable signal that can be easily controlled by most people to perform a task by using a BCI system constructed on EEG signal analysis. In their experiment, ninety nine people were asked to spend 20 to 30 minutes on a two-session BCI investigation. The first session consisted of 40 trials conducted without feedback. The data collected will be used to set up a subject-specific classifier that can provide the subject with feedback. Then the second session—40 trials in which the subject had to

2

control a horizontal bar on a computer screen—was conducted. The result of this experiment shows that 93% of the subjects were able to achieve classification accuracy above 60% after two sessions of training which provides evidence that EEG signal is a useful candidate for BCI construction. Researchers have developed several types of EEG based BCI systems. These systems rely on the finding that the brain reacts differently to different stimuli, based on the level of attention given to the stimulus and the specific processing triggered by the stimulus. Some EEG based BCI systems [51-53] require the user learn to produce self-regulated, stable EEG signal, such as alpha or mu rhythm. This learning process may take several weeks, and since there are only two states (on and off) available, it is less effective when performing multiple choices tasks.

The P300-Speller [3] developed by Farwell and Donchin in 1988 is another type of EEG based BCI system that relies on a brain response known as the P300, whose attributes have been studied for over four decades. P300-Speller is a non-invasive BCI without requirement of subject's training. In this research, all the experiments were performed on this well known and widely used BCI system. The P300-Speller and its base scenario — Oddball paradigm — are briefly discussed in the following section.

1.2 Event-Related Potential, Oddball Paradigm and P300-Speller

Event-related potentials (ERP) are voltage fluctuations that are associated in time with some physical and mental occurrences. These potentials can be extracted from the EEG signals that were recorded from the human scalp by means of signal processing techniques [54]. Task-based ERP component is defined as those aspects of the waveform associated with particular antecedent conditions and experimental manipulations. P300

3

response is such an ERP component that is elicited by rare events presented within the so-called Oddball Paradigm [55], in which each event in a sequence of events can be categorized into one of two categories and correct categorization is necessary to the subject's assigned task. When one of the categories occurs relatively infrequently, members of this rare category can elicit a P300. Early observations of the P300 response were reported in mid-1960s. In 1964, Chapman and Bragdon [56] found that event related potential (ERP) response to visual stimuli differed depending on whether or not the stimuli were meaningful. In their experiment, a large ERP peak appeared around 300 ms following the meaningful stimulus which is termed as P300 response and is shown in Figure 1. P300 response is an uncontrollable ERP signal generated by the brain. In mid-1980s it was used in the lie detection, which was known as "guilty knowledge test" [57]. This practice has recently enjoyed increased legal permissibility. In 1988, Farwell and Donchin designed a BCI based on P300 responses—P300-Speller as shown in Figures 2-3. In this BCI system, the user is presented with an oddball paradigm. The rows and columns in a matrix of letters and numbers are intensified in a random sequence. The user focuses attention on one letter in the matrix. Intensifications of the row and column of the attended letter compose a rare event. Intensifications of the other rows and columns compose the frequent events. Thus, the intensifications of rows and columns containing the attended letter elicit a P300 response, while rows and columns not containing this letter do not elicit a P300 response. Therefore, by examining in real-time which row and which column elicited a P300 response, the system is capable of detecting the character communicated by the user with high accuracy. The successful use of the P300-BCI system does not require any training of the user. However, for optimal use, the algorithm

4

detecting the P300 response needs to be "calibrated" based on the pattern of electrical brain activity of a specific user.
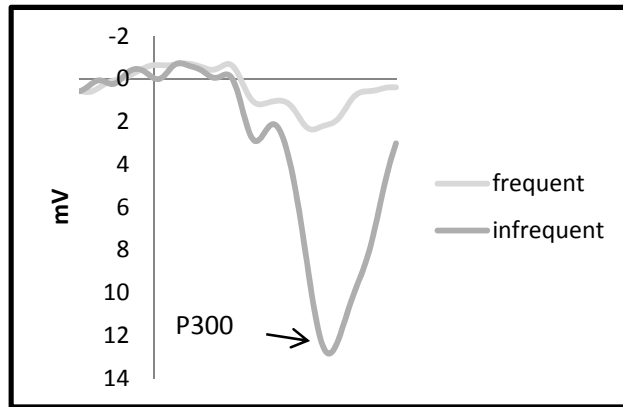
Figure 1: P300 response. P300 response is an event related potential which is triggered by the infrequent events in an oddball paradigm.
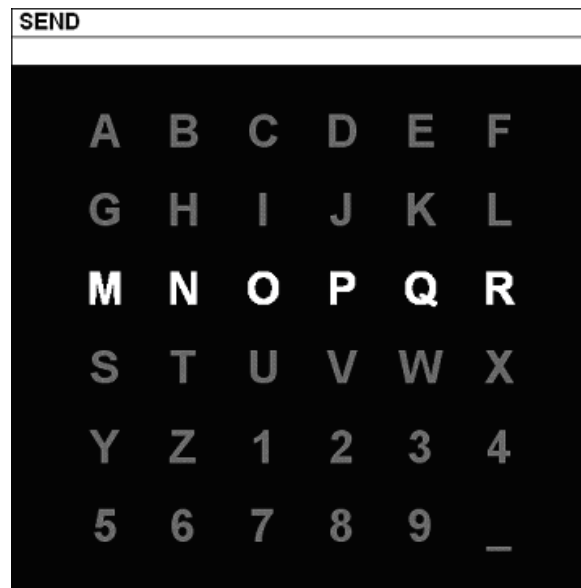
Figure 2: The P300-Speller proposed by Farwell and Donchin. On the screen, the rows and columns in the matrix were flashed alternately and the flashing row/column that contains the target character will elicit P300 responses [58].
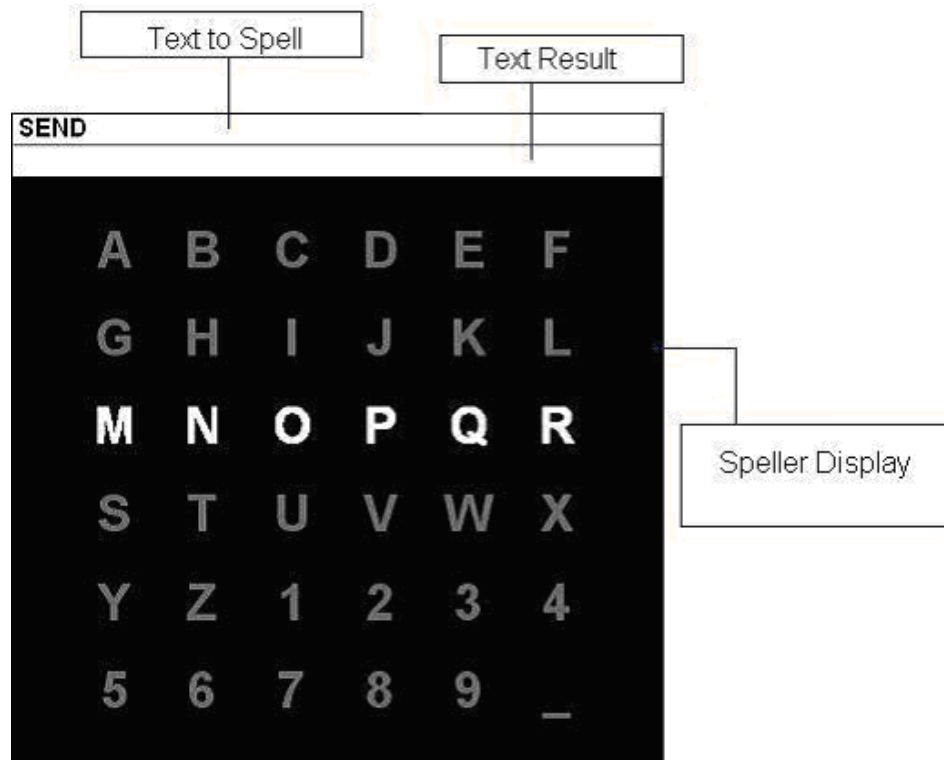
Figure 3: Elements of the user's screen of P300-Speller. *Text to Spell* indicates the pre-defined text. The speller will analyze evoked responses, and will append the selected test to *Text Result* [58].

1.3 Current State-of-the-Art of P300-Speller Based BCI System

In 1990s, there was a significant increase in research area of signal processing and BCI system design. A lot of techniques about the EEG signal feature extraction and classification for BCI systems have been introduced and thoroughly investigated by numerous researchers. Farwell and Donchin have used stepwise linear discriminant analysis (SWLDA), peak pick, area and covariance to do the feature extraction and classification [3]. Anderson *et al*. [59] designed a multivariate autoregressive model for EEG classification. Devulapalli [60] suggested using principal component analysis (PCA) with autoassociative networks to fulfill the task. Samar *et al*. [61] and many other

6

researchers [62-66] applied wavelet transform for EEG signal classification. Kaper *et al.* [67], Qin *et al.* [68] and other researchers [69-71] used support vector machine (SVM) as the feature extractor and classifier. Jung *et al.* [72] and Barros *et al.* [73] suggested using independent component analysis (ICA) to remove the artifacts from EEG signal while many other researchers [74-79] also used ICA to process and classify the multichannel signals. Some other authors suggest using matched filter [80], genetic algorithm (GA) [81-82] and other techniques for feature extraction. Generally, these techniques can be divided into two categories, linear and non-linear techniques. Krusienski *et al.* [91] compared most of these techniques and concluded that SWLDA is the most accurate and practical technique working with P300 Speller. In Chapter 2, several widely used and effective methods including SWLDA, SVM, matched filter, Wavelet Transform and ICA are discussed. The processing results of single trial SWLDA by using BCI2000[1] system are used as our benchmark in Chapters 3-5.

1.4 Motivation and Objective

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disease that affects nerve cells in the brain and the spinal cord. The progressive degeneration of motor neurons will destroy the patient's ability of control any voluntary muscles. In later stage,

---

[1]BCI2000 is a general-purpose system for brain-computer interface (BCI) research. It can also be used for data acquisition, stimulus presentation, and brain monitoring applications. BCI2000 development has been sponsored by a NIH (NIBIB/NINDS) Bioengineering Research Partnership grant to Jonathan Wolpaw. Current development is sponsored by a NIH (NIBIB) R01 grant to Gerwin Schalk.

patients who suffer from this disease may become totally paralyzed which means they may completely lose the traditional communication ability such as talking, writing or gesturing. There are approximately 5,600 people in the U.S. who are diagnosed with ALS each year. The incidence of ALS is two per 100,000 people which means the estimated Americans who may have this disease at any given time is about 30,000. ALS can strike anyone, cases of this disease have been found throughout the world without racial, ethnic or socioeconomic boundaries [83]. It is especially for these people's benefit that researchers are eager to design a BCI system to restore their communication ability. As mentioned before, a lot of signal processing techniques have been investigated, developed and designed for the construction of a reliable BCI system with high accuracy and processing speed. It has been shown that some of the techniques can be effectively in practical BCI systems, such as P300-Speller, and have been successfully applied to the ALS patients. These facts validated the feasibility of using BCI systems to restore the communication ability of the patients. In more than 20 years, researchers have made tremendous effort in developing new techniques to improve the communication speed and accuracy of P300-Speller. However, the communication speed is still at a low level (around 5 to 8 characters/min). To improve the performance of P300-Speller based BCI system is still a challenging problem that calls for effective solutions. The aim of this study is to develop fast and accurate single trial P300 response detection algorithms to further optimize the processing speed and accuracy of P300-Speller, thereby providing the ALS patients a reliable and high speed communication BCI system.

8

1.5 Contributions of this Dissertation

In this dissertation, the available techniques working with P300-Speller are investigated. A blind tracking based ICA algorithm is proposed for single trial EEG signal classification. This algorithm brought a new idea for applying ICA algorithm on EEG signals. By using the blindly generated random starting vector, a piece of random information that lies on the starting vector was actually removed from the rest of the components and therefore the entire independent component (IC) system will be rotated. A small portion of the non-Gaussianity of the rotated IC system may be sacrificed because the first vector is not computed according to the maximum non-Gaussianity criteria. However, this sacrifice may be beneficial to the construction of a more useful IC system since the non-Gaussian assumption for the P300 response and the background noises is not 100% valid (P300 repsonse could be the combination of several ICs and not completely independent to the background noise). In fact, ICA algorithms decompose a mixed signal into independent components without providing any information of mapping the ICs to P300 response. The proposed blind tracking based ICA algorithm created an opportunity to "modify" the IC system and make the relation of P300 response and ICs more clear and useful in the modified IC system. If we say that the traditional ICA algorithms provide us a fixed fair solution for the IC set computation, then the proposed algorithm provides us a dynamic optimized solution. In this work, the proposed algorithm has been applied to five different subjects and achieved an overall classification accuracy of 79.1%, which is 34.1% more accurate than the single trial SWLDA algorithm. The experimental result validates the effectiveness of the proposed blind tracking based ICA algorithm[1-5] and suggests a new direction of ICA research.

In addition to the blind tracking based ICA algorithm, another algorithm called variance analysis based single trial P300 response classification was proposed in this dissertation as well. In this algorithm, we mathematically derived a statistical parameter that can be used for target and non-target signals classification by analyzing their variances. The distribution of the statistical parameter was investigated and the classification rule was set up. This algorithm[6] is mostly established on statistical analysis and hypothesis testing. The processing results showed that this technique achieved an overall classification accuracy of 84.8% for five different subjects, which is about 40% more accurate than the single trial SWLDA.

Both of the proposed algorithms introduced new ideas to the research of single trial P300 response classification by using P300-Speller. They dramatically reduced the processing time and increased the classification accuracy. The communication speed has been improved from 12.8 characters/ min (SWLDA) to 30.6 characters/min and 20.5 characters/min for blind.

[1] [84] K. Li, R. Sankar, Y. Arbel and E. Donchin, "P300 Based Single Trial Independent Component Analysis on EEG Signal", *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience, Lecture Notes in Computer Science*, Volume 5638/2009, 404-410, 2009.

[2] [85] K. Li, R. Sankar, Y. Arbel and E. Donchin, "Single trial independent component analysis for P300 BCI system", *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4035-4038. September 2009.

[3] [86] K. Li, R. Sankar, Y. Arbel and E. Donchin, "Single trial independent component analysis for the P300 BCI system", *Fourth International Meeting,* Asilomar, California May 31 - June 4, 2010.

[4] [87] K. Li, R. Sankar, Y. Arbel and E. Donchin, "Blind tracking based single trial independent component analysis for P300 BCI system", accepted with minor changes by *IEEE Transaction on Neural Systems and Rehabilitation Engineering*, 2010.

[5] [88] K. Li, R. Sankar, Y. Arbel and E. Donchin, "Single trial independent component analysis for the P300 BCI system", submitted to *The Journal of Neural Engineering*, 2010.

[6] [89] K. Li, R. Sankar, Y. Arbel, and E. Donchin, "A new single trial P300 classification method", submitted to *IEEE Transaction on Neural Systems and Rehabilitation Engineering*, 2010.

tracking based ICA algorithm and variance analysis based classification algorithm respectively.

1.6 Dissertation Outline

In Chapter 2, several widely used EEG signal processing techniques including stepwise linear discriminant analysis (SWLDA), support vector machine (SVM), matched filter, wavelet transform, and independent component analysis (ICA) are discussed. The mathematic derivation of the algorithms and their underline assumptions are presented. Generally, this chapter will give the readers a basic idea of the mentioned techniques and their advantages/disadvantages. With this background information, the readers can understand our proposed algorithms more easily.

In Chapter 3, the first proposed single trial P300 response classification method—Blind Tracking Based ICA is discussed. First, the BCI system and the data acquisition process are introduced. Then the experiment data structure is displayed. After that, the concept of "blind tracking" is explained, and the underlining assumptions and the steps of the proposed algorithm are presented in detail. Finally, the processing results are presented and discussed.

In Chapter 4, the second proposed single trial P300 response classification method—Variance Analysis Based P300 Response Classification is discussed. First, the BCI system and the data acquisition process are introduced. Then the experiment data structure is displayed. After that, the underlining assumptions and the steps of the

11

proposed variance analysis algorithm are presented in detail. Finally, the processing results of the single trial P300 response classification are presented and discussed.

In Chapter 5, the processing results of the proposed algorithms are summarized and discussed. The feasibility and the performance of the algorithms, and the impact to the research area are concluded. The problems and challenging issues that still need to be investigated are discussed and possible solutions are proposed.

12

CHAPTER 2

DESCRIPTION OF AVAILABLE SIGNAL PROCESSING TECHNIQUES FOR P300-SPELLER

2.1 Introduction

P300 response has been observed since mid-1960s [56][90]. It is an event related potential (ERP) elicited by infrequent, task-relevant stimuli. P300 response is considered to have stable presence, amplitude and timing and be related to a person's reaction to the stimulus but not to the physical attributes of the stimulus. Based on these properties and the fact that P300 response is usually elicited in oddball paradigm, Farwell and Donchin designed a well known BCI — P300-Speller, which has been proven to be a practical and reliable BCI system for several decades. A lot of techniques have been developed to work with P300-Speller. In this chapter, several techniques including SWLDA, SVM, matched filter, Wavelet, and ICA, will be discussed.

2.2 Stepwise Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a well known method for dimensionality reduction and classification that project high-dimensional data onto a low dimensional space where the data achieves maximum class separability. The derived features in LDA

13

are linear combinations of the original features, where the coefficients are from the transformation matrix. The optimal projection or transformation in classical LDA is obtained by maximizing the ratio of the between-class variance to the within-class variance as described in equation (1).

$$J(w) = \frac{w^t S_b w}{w^t S_w w} \tag{1}$$

Where w is the transformation matrix, S_b and S_w are the between-class variance and within-class variance, respectively and t represents the transpose operation.

Stepwise Linear Discriminant Analysis (SWLDA) combines the LDA with both forward and backward regression for feature selection to construct a multiple regression model as the classifier with significant features. The combined forward and backward stepwise regression starts with no initial model term. The most statistically significant features that are described as predictor variables in the classifier are added in the classifier if their p-value < 0.1. After adding each new entry to the classifier, a backward regression is performed to remove the least significant predictor variables that have p-value>0.15. This process is repeated till the classifier includes a predetermined number of terms, or till no additional terms satisfy the entry/removal criteria.

SWLDA has the advantage of automatic feature selection because during the process, insignificant features have been completely removed from the classifier. Therefore using less training data is less likely to corrupt the classification results. A weak point of SWLDA is that the convergence of the feature selection is not guaranteed if the model (classifier) is inadequate or if there is no discriminable information contained in the

14

features. This problem can be solved by properly configuring the system before the classifier development.

Krusienski *et al.* [91] have applied SWLDA with P300-Speller for P300 response classification. In their work, the EEG signal was recorded at 240 Hz sampling frequency by using a cap embedded with 64 electrodes whose locations are distributed over the entire scalp. To identify a character presented in the P300-Speller, the six rows and six columns will be intensified in random order for 15 times before the target character classification. The 800 ms long EEG segments (192 samples) that follow each intensification are extracted and concatenated as the feature vector. Both backward and forward regression are applied to the feature vector to remove the insignificant features and then the transformation matrix is derived for the significant features used to detect the presence or absence of the P300 response in an EEG signal.

By using SWLDA as the classification method, Krusienski *et al.* achieved at least 60% accuracy for all participants. Three of the five participants performed above 90% accuracy with fewer than 15 sequences. This indicates that the classification can be performed on a minimal number of sequences without compromising accuracy and can increase the communication rate. (According to Donchin *et al.*'s work [92], the average classification accuracy obtained by using SWLDA as the classifier can reach above 90% with 8 sequences.) In Krusiensiki *et al.*'s work, the online analysis and the offline analysis are separated by quite a few months, which means after the offline data were collected, the five participant were asked to do the online part several months later.

15

Therefore, the classification result also proved the stable and robust nature of EEG signal to the P300-Speller.

2.3 Support Vector Machine

Support Vector Machine (SVM) is a supervised learning method used for classification and regression. Here "supervised" means that for a set of given training data, SVM will build up a model that can predict whether a new sample belongs to one category or the other. Actually, what SVM does is to construct a hyperplane or set of hyperplanes in a high or infinite dimensional space to separate different categories. Usually a good separation can be achieved by finding the separation hyperplane that has the greatest distance to the nearest training data points of any class. The idea of SVM, which was discussed in detail by Burges [93], is briefly outlined here.

For the two classes shown in Figure 4, an easy way to perform binary classification is to construct a hyperplane described by the weight vector w and the bias term b. Given a training data set of l samples, each sample is denoted as x_i and the corresponding class labels y_i

$$(x_1, y_1), \dots (x_l, y_l) \in R^N \times \{-1,1\} \tag{2}$$

A SVM algorithm needs to find such a hyperplane that separates the two categories and maximizes the distances from the hyperplane to any of the nearest data points. Meanwhile, the selection of the optimal hyperplane subjects to some constraints that will be discussed later. The category label of an incoming data x can be predicted by using equation (3)

16

$$f(x) = w \cdot x + b \tag{3}$$

In equation (3) the incoming data vector x was projected on the weight vector w. Since w is perpendicular to the separating hyperplane, the sign of this projection would reveal the predicted class label. The separating hyperplane can be described by the vector w and bias term b, and w can be calculated by using the vectors on the margin (bordered circles in Figure 4) only. These necessary vectors are called support vectors.

We want to choose w and b to maximize the distance between the parallel hyperplanes that separate the data. These hyperplanes can be defined by the equations

$$w \cdot x - b = 1 \tag{4}$$

and

$$w \cdot x - b = -1 \tag{5}$$

The distance between these two hyperplanes is $\frac{2}{\|w\|}$. Therefore to maximize this distance, we need to minimize $\|w\|$. At the same time, we have to prevent data points from falling into the margin. Hence the minimization has to subject to the following constraint.

$$w \cdot x_i - b \geq 1 \; for \; x_i \; from \; the \; first \; class \tag{6}$$

or

$$w \cdot x_i - b \leq -1 \; for \; x_i \; from \; the \; second \; class \tag{7}$$

Equations (6) and (7) can be rewritten as:

17

$$c_i(w \cdot x_i - b) \geq 1 \; for \; all \; 1 \leq i \leq n \qquad (8)$$

where c_i is class label for x_i.

Now the optimization problem becomes:

minimize $\|w\|$ subject to (for any $i = 1, \dots, n$) $c_i(w \cdot x_i - b) \geq 1$.

This optimization problem is not easy to solve because it depends on $\|w\|$, which includes computation of a square root. Fortunately, substitute $\|w\|$ by $\frac{1}{2}\|w\|^2$ will not change the solution of this problem.

After substitution the problem is now to minimize $\frac{1}{2}\|w\|^2$ subject to $c_i(w \cdot x_i - b) \geq 1$. It can be easily solved by using quadratic programming technique [94-95].

Employing Lagrange multipliers α_i to clearly define this constrained problem gives:

$$min_{w,b} \, max_\alpha \{\frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i \, [c_i(w \cdot x_i - b) - 1]\} \qquad (9)$$

The solution to this problem is:

$$w = \sum_{i=1}^{n} \alpha_i c_i x_i \qquad (10)$$

In 1995, Cortes and Vapnik [96] proposed a "soft margin" method that allows for mislabeled examples. This method can find a hyperplane that separates the data as cleanly as possible when there is no such a hyperplane that can correctly split all the data. They introduced slack variables, ξ_i, to measure the degree of misclassification. With the slack variables, ξ_i, the constraint becomes:

18

$$c_i(w \cdot x_i - b) \geq 1 - \xi_i \tag{11}$$

and the optimization problem is now becoming:

$$\min_{w,\xi}\{\tfrac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i\} \tag{12}$$

subject to (for any $i = 1,\dots n$) $c_i(w \cdot x_i - b) \geq 1 - \xi_i$.

The solution to this optimization problem is:

$$w = \sum_i^{N_s} c_i\alpha_i x_i \tag{13}$$

where N_s denotes the number of resulting support vectors. Substituting (13) into (3) yields

$$f(x) = \sum_i^{N_s} c_i\alpha_i\,(x \cdot x_i) + b \tag{14}$$

It has been shown [93] that the replacement of the dot product $x \cdot x_i$ by a positive definite symmetric kernel function $K(x, x_i)$ transforms the given data space into a (usually higher dimension) feature space. This leads to the nonlinear discriminant function

$$S(x) = \sum_i^{N_s} c_i\alpha_i\,K(x, x_i) + b \tag{15}$$

This technique provides more flexible decision boundary in the data space, which may increase classification accuracy. In Kaper *et al.*'s work [67], they chose the Gaussian kernel

$$K(x, x_i) = e^{-\frac{\|x-x_i\|^2}{2\sigma^2}}. \tag{16}$$

19

The performance of SVM classifier depends on the regularization of parameter C and the bandwidth σ of the Gaussian kernel. To achieve good classification results, both parameters need to be carefully adjusted.

Kaper *et al.* used the P300-Speller proposed by Farwell and Donchin in 1988. The EEG signals were collected by using a cap embedded with electrodes. 600 ms segments of the signals following the intensification of each rows or columns were extracted as the experimental data, which are bandpassed (0.5-30 Hz) and normalized to an interval of [-1, 1] prior to the application of SVM classifier. The SVM classifier was trained by using two positive samples and two negative samples. The efficiency of the two parameters, C and σ, were assessed by *cross-validation* [97]. Now the SVM classifier is ready to be applied to incoming signals. Usually, the incoming signal is too noisy to be correctly classified by using only one sequence. Therefore, the classification results of several sequences have to be combined to give the final decision. They used the value of equation (15) as a score and combined sequences by summing the scores from corresponding rows/columns from different sequences as in equation (17).

$$s_i^{row/column} = \sum_{k=1}^{n} S(x_{ik}^{row/column}) \qquad (17)$$

This idea was coming from maximum contrast classifiers (MCC) [98], which suggests that the score can be interpreted as density difference. The row/column with the highest total score after n trials is chosen as the target row/column with P300 presence. By using this method, an accuracy of 84.5% was achieved for P300 classification after five trials with the two parameters set as $C = 20.007$ and $\sigma = 27.359$ [67].

20
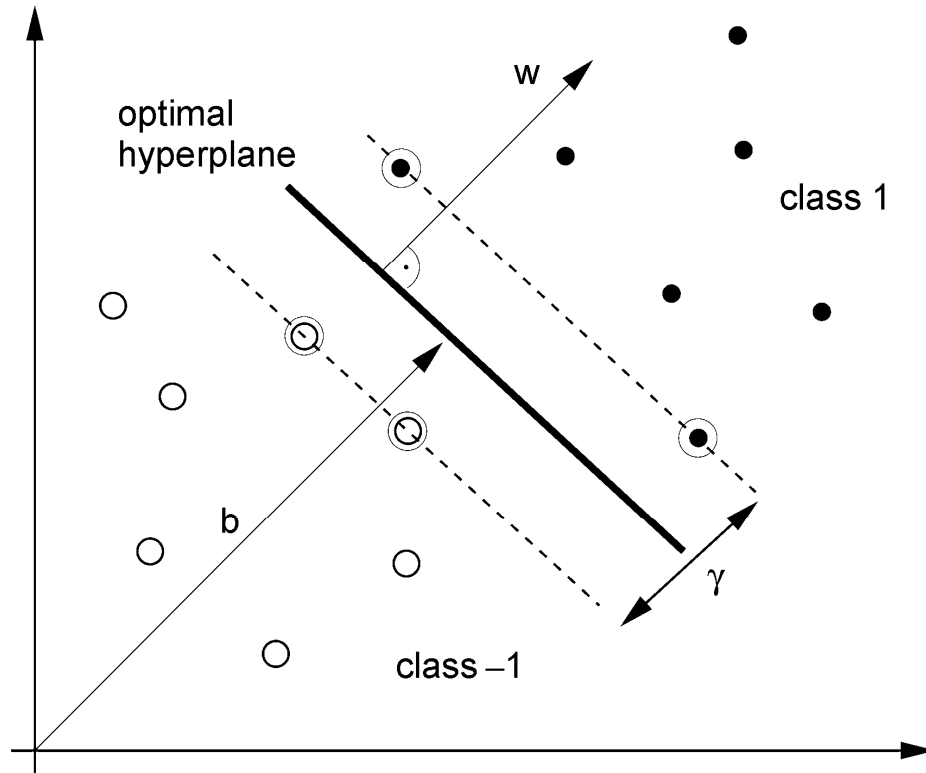
Figure 4: SVMs find the optimal hyperplane (solid line) to separate two classes by maximizing the margin γ. It can be described by the vector w and the bias term b. Only support vectors (bordered circle) are necessary to calculate w and b. (This Figure was quoted from [67])

2.4 Matched Filter

In signal processing, a matched filter is obtained by correlating a known signal, or template, with an unknown signal to detect the presence of the template in the unknown signal. It is actually the convolution of the unknown signal and the conjugated time-reverse version of the template. Matched filter is used to maximize the signal to noise ratio (SNR) when there are background stochastic noises. This technique is commonly used in radar, in which a known signal is sent out, and the reflected signal is examined

for common elements of the outgoing signal. It is also used in image processing, for example, to improve the SNR for X-ray images.

The matched filter is the linear filter, h (in equation (18)), that maximizes the output SNR.

$$y(n) = \sum_{k=-\infty}^{\infty} h(n-k)x(k) \tag{18}$$

Let us write the observed signal x as:

$$x = s + v \tag{19}$$

where s is the desirable signal and v is the background noise.

Thus the covariance matrix of the noise is given by:

$$R_v = E\{vv^H\} \tag{20}$$

By maximizing the SNR, the matched filter can be derived as:

$$h = \alpha R_v^{-1}s \tag{21}$$

where α is an arbitrary number. Normally, the expectation power of the filter output due to the noise (y_v) is normalized to unity, which means

$$E\{|y_v|^2\} = \alpha^2 s^H R_v^{-1}s = 1 \tag{22}$$

Then α can be solved as:

$$\alpha = \frac{1}{\sqrt{s^H R_v^{-1}s}} \tag{23}$$

Consequently, the matched filter *h* is:

$$h = \frac{1}{\sqrt{s^H R_v^{-1} s}} R_v^{-1} s \tag{24}$$

In frequency domain, matched filter can be considered as applying the greatest weighting to spectral components that have the greatest signal to noise ratio. This technique is often used in signal detection [99].

Matched filter is commonly used with other classification techniques to produce the final classification result. Serby *et al.* [80] have combined matched filter with independent component analysis to classify the P300 component in the EEG signals collected by using P300-Speller. In their work, the EEG data were collected by using the International 10-20 system Electro-Cap at 250 Hz sampling frequency. Afterwards, the EEG signals were low pass filtered with the cut-off frequency set as 6 Hz. Then only 3 signals from electrodes C_Z, P_Z and F_Z (locations of these electrodes are shown in Figure 5) were fed to the ICA algorithm. The ICA algorithm returned three independent components. One of them was considered associated with P300 response and the other two were omitted. Then every 500 ms (from 100 ms to 600 ms) long segment posterior to the beginning of any intensification in the P300 source was passed to the matched filters that were constructed by using the P300 templates of every row/column. The row/column that had the highest correlation with the incoming signal was considered as a target signal candidate. The final decision was made after several repetitions of the process and a predefined threshold was satisfied. With this method, a communication rate of 5.45 symbols/min with an accuracy of 92.1% was achieved.

23
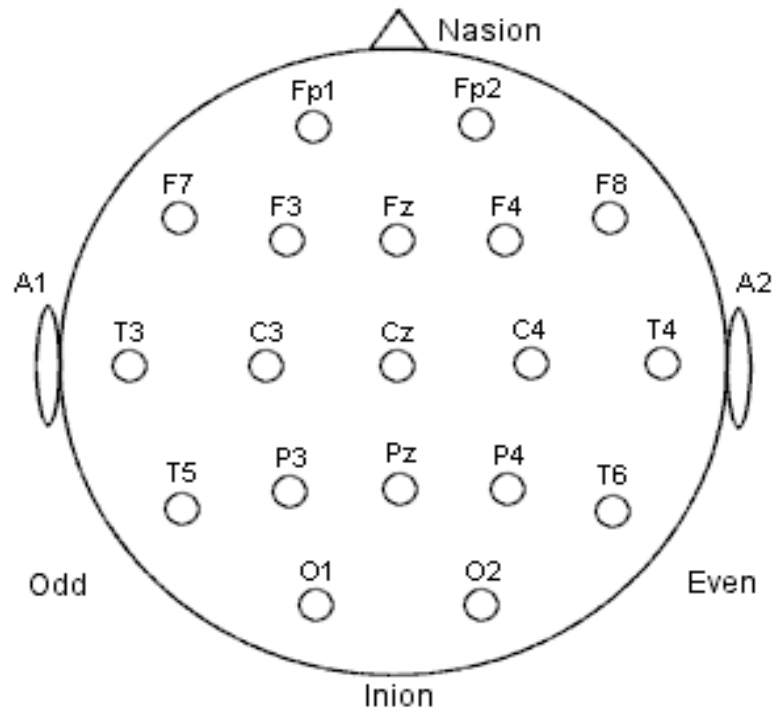
Figure 5: The international 10-20 system: An internationally recognized method to describe and apply the location of scalp electrodes in the context of an EEG test or experiment.

2.5 Wavelet Transformation

Wavelet is a mathematical tool that can be used to extract information from different data. Its development goes back to early 20th century, starting with Haar's work [100]. From 1975 to early 1990s, there was a burst of wavelet development. A lot of notable contributions had been made by many outstanding researchers, such as George Zweig [101], Jean Morlet [102-103], Alex Grossmann [103], Yves Meyer [104] and Stephane Mallat [105].

Technically, wavelet is a mathematical function used to divide a given function or continuous-time signal into different scale components that have been assigned a frequency range. A wavelet transform uses the wavelets to represent a function. Wavelet transform has advantages over traditional Fourier transform in representing functions that have discontinuities and sharp peaks, and for accurately deconstructing and reconstructing finite, non-periodic and non-stationary signals. Wavelet transforms are classified into continuous wavelet transforms (CWTs) and discrete wavelet transforms (DWTs). Both of them are continuous-time transforms and used to represent continuous-time signals. The difference is that CWTs operate over every possible scale and translation whereas DWTs use a predefined subset of scale and translation values.

In continuous wavelet transforms, a given signal is projected on a continuous family of frequency bands. For example the signal can be represented on every frequency band of [*f, 2f*] for all positive frequencies $f > 0$. The original signal can be reconstructed by a suitable integration over all the frequency components. The frequency bands or subspaces are scaled versions of a subspace at scale 1. Usually the subspaces are generated by shifting a generating function $\psi \in L^2(\mathbb{R})$, the mother wavelet. Some of the famous mother wavelets are shown in Figures 6-8.
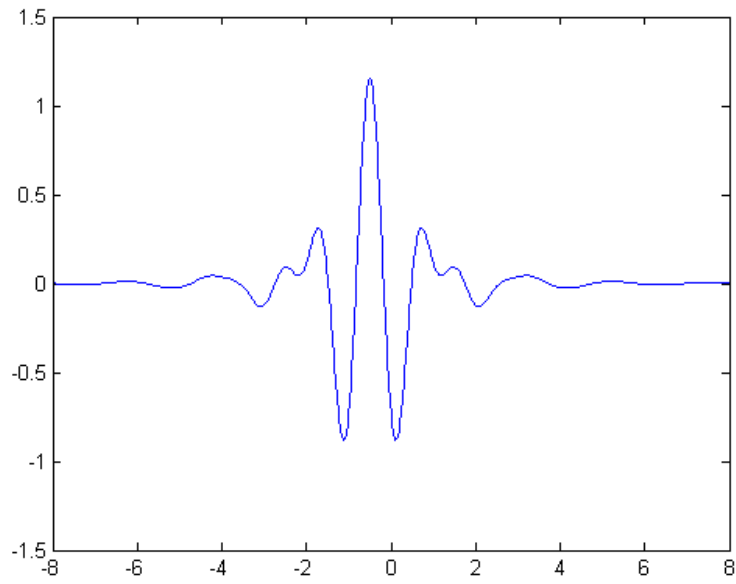
Figure 6: Meyer mother wavelet [104]
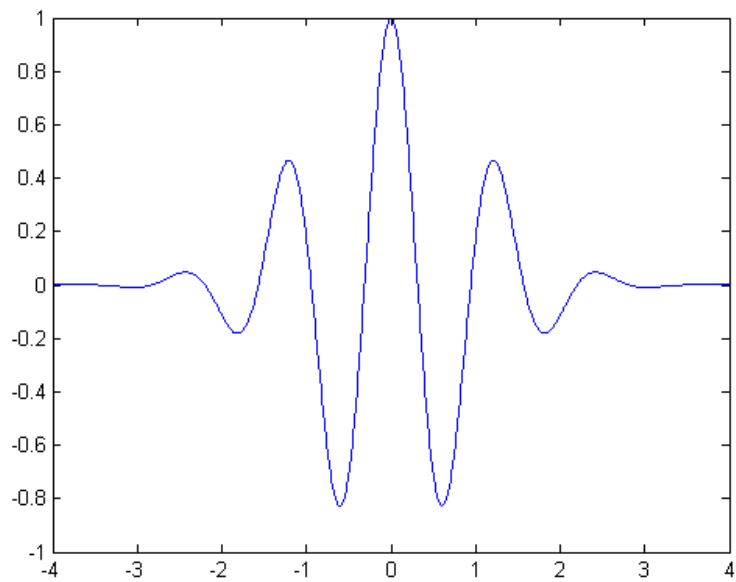
Figure 7: Morlet mother wavelet [103]

26
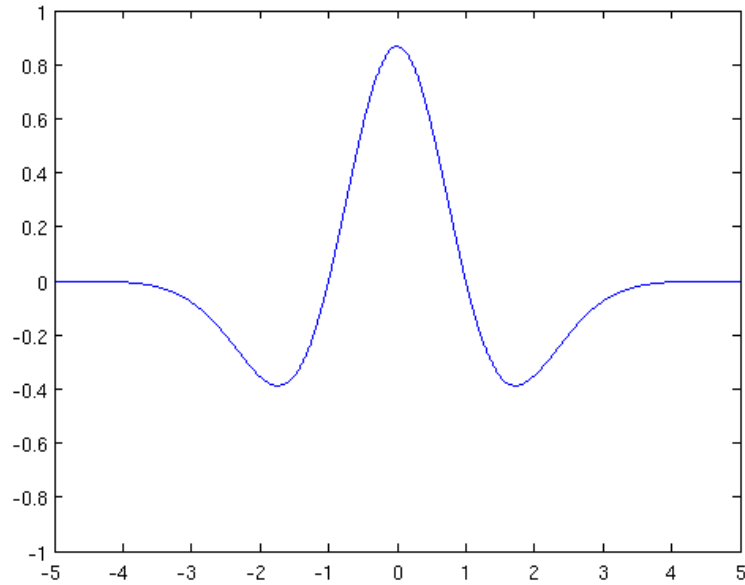
Figure 8: Mexican Hat mother wavelet [106]

The child wavelets of scale or frequency band $[\frac{1}{a}, \frac{2}{a}]$ are generated by the function

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-b}{a}) \tag{25}$$

where a is a positive number that defines the scale and b is a real number that defines the shift.

Then the projection of a function x onto the subspace of scale a has the form

$$x_a(t) = \int WT_\psi\{x\}(a,b) \cdot \psi_{a,b}(t)db \tag{26}$$

with wavelet coefficients

$$WT_\psi\{x\}(a,b) = \langle x, \psi_{a,b}\rangle = \int x(t)\overline{\psi_{a,b}(t)}\,dt \tag{27}$$

27

For discrete wavelet transform, it follows the same idea. However it is computationally impossible to analyze a signal x using all the wavelet coefficients. Therefore only a discrete subset of the wavelet coefficients is chosen to reconstruct the signal.

$$x(t) = \sum_{m \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} \langle x, \psi_{m,n} \rangle \cdot \psi_{m,n}(t) \qquad (28)$$

where $\psi_{m,n}$ is the child wavelet function that comes from a tight frame of $L^2(\mathbb{R})$.

In fact, CWT is a kind of template matching [107], i.e., a computation of the cross covariance between the signal and a predefined waveform, which is shifted in time and varied in scale [67]. The local extrema of the wavelet coefficients are the points of the best match between the signal and the template in the time-frequency domain. By combing the CWT building up on a modified Mexican Hat function and two-sample t-test, Vladimir Bostanov designed a feature extraction algorithm that works with P300-Speller. With this feature extraction algorithm, he had achieved 82.6% and 54.4% accuracies for two different data sets provided by BCI Competition 2003, respectively.

2.6 Independent Component Analysis

Blind source separation (BSS), also known as blind signal separation, is the separation of a set of target signals from a set of mixed signals, without the aid of information about the source signals or the mixing process. BSS relies on the assumption that the source signals do not correlate with each other, i.e. the signals are mutually statistically independent or decorrelated. BSS plays a very important role in signal processing and has been explored by many researchers. A famous example of the BSS is the "cocktail party" problem. If you have been to a cocktail party as shown in Figure 9, you most certainly

28

know how hard it can be to extract an interesting conversation from the noisy background signal of the crowd. So the cocktail party problem is the task of hearing a sound of interest in this sort of complex auditory setting [108]. The human hearing system can segregate the mixing sound and concentrate on the component of interest very well. In digital signal processing, an equivalent method that can accomplish this task needs to be developed. Independent Component Analysis (ICA) is one of the BSS methods which can decompose a mixed signal into statistically independent components (ICs) by maximizing their non-Gaussianity [109]. Typically, ICA applications include separating mixing signal, removing artifacts from brain signal recording, finding hidden factors in financial time series, and reducing noise in natural images. Figures 10-12 gives an illustration of ICA application for signal separation. (Figure 10 shows the original signals, Figure 11 shows the mixed signals of the two original signals and Figure 12 shows the reconstructed signals after the application of ICA algorithm [110].)

ICA algorithms are established on the assumption that the original signals are mutually statistically independent with maximized non-Gaussianity. Without non-Gaussianity the estimation of original signal is not possible. Therefore non-Gaussianity is used as a leading principle in ICA estimation. Widely used ICA algorithms include infomax, FastICA, and JADE. In the following paragraphs, the development of FastICA algorithm will be discussed.

29

Figure 9: A typical Manhattan cocktail party. The listener must focus on the conversation of interest and neglect all other background noises. (Image from Breakfast at Tiffany's: Paramount Pictures.) [108]
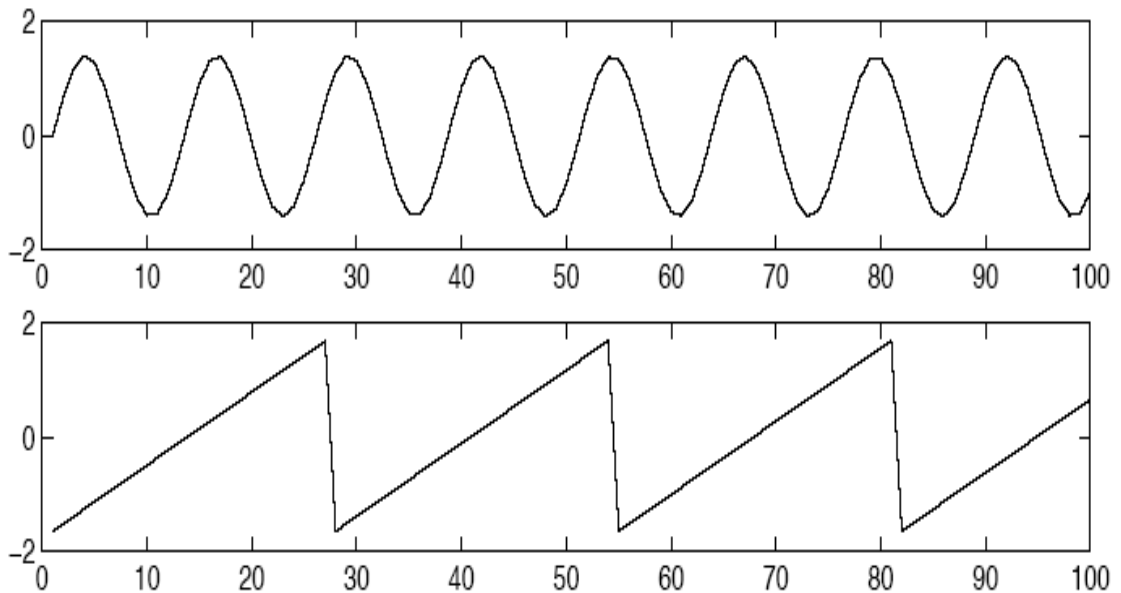
Figure 10: The source signals used as illustration of the ICA separation
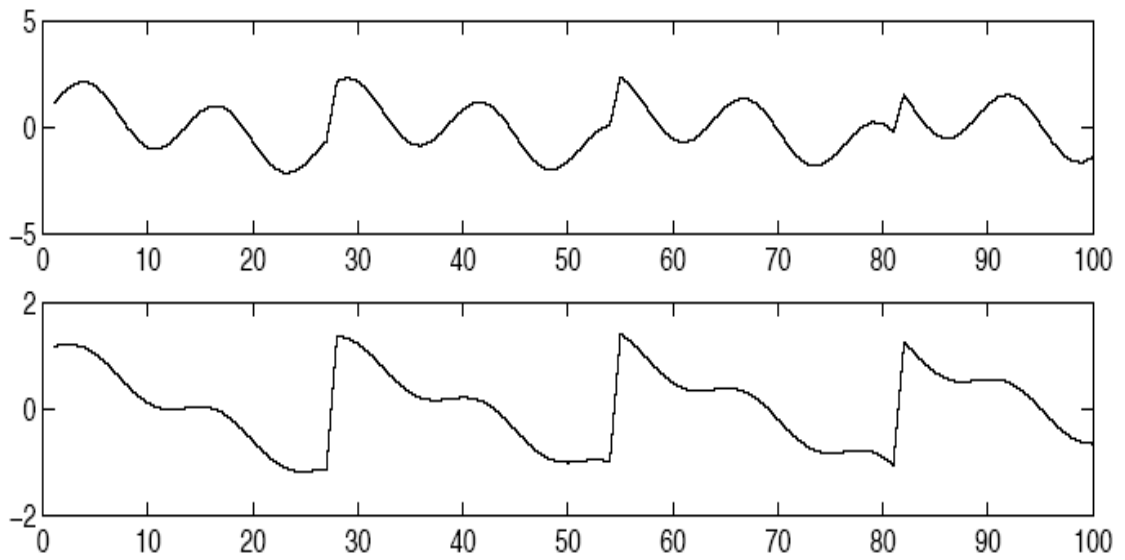
Figure 11: The observed mixtures of the source signals in Figure 10.
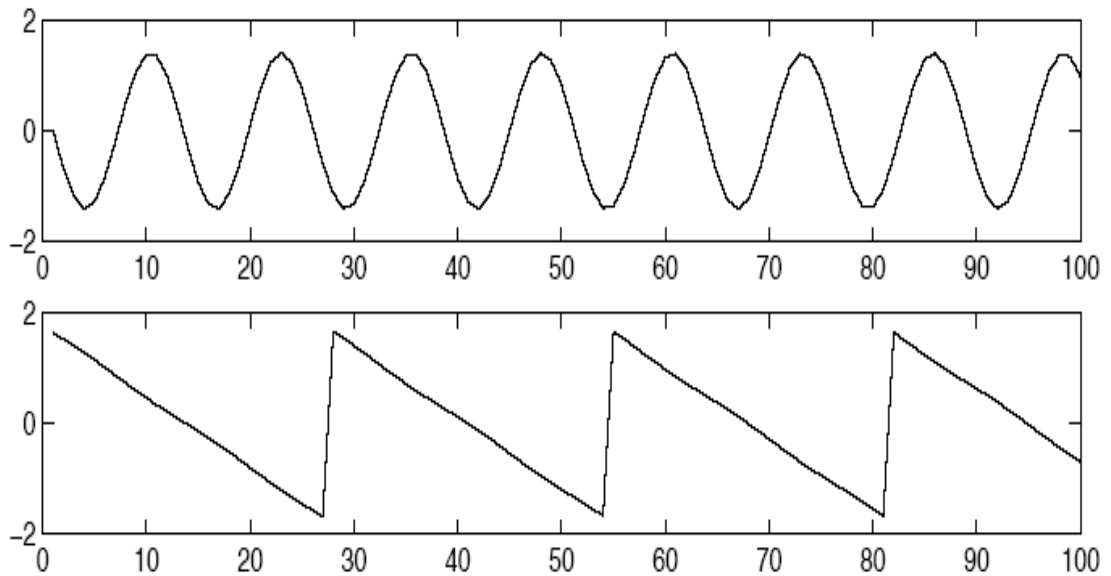
Figure 12: The estimates of the original source signal by using the observed mixing signals in Figure 9 only. The original signals were accurately estimated, up to multiplicative signs. (Figure 10-12 are quoted from [110])

31

Before we start the discussion of development of the FastICA algorithm, a whitening process is introduced. For any signal *x,* the whitening process (that is a linear transformation of the observed signal) is applied to reduce the parameters to be estimated and relive the computation load. The components of the transformed signal $\tilde{x}$ are uncorrelated with their variance equal unity (see equation (29)).

$$\varepsilon\{\tilde{x}\tilde{x}^T\} = I \tag{29}$$

The whitening transformation is always possible. A popular method is to use the eigenvalue decomposition of the covariance matrix, $\varepsilon\{xx^T\} = EDE^T$, where E is the orthogonal matrix of eigenvectors of $\varepsilon\{xx^T\}$ and D is the diagonal matrix of its eigenvalues. The whitening transformation is operated by

$$\tilde{x} = ED^{\frac{1}{2}}E^T x \tag{30}$$

If the observed signal *x* is distributed by an ICA data model as:

$$x = As \tag{31}$$

where *s* is the matrix of independent components and *A* is the activation matrix (*s* and *A* will be discussed later). Substituting equation (31) into equation (30) gives

$$\tilde{x} = ED^{\frac{1}{2}}E^T As = \tilde{A}s \tag{32}$$

where $\tilde{A}$ is an orthogonal matrix since

$$\varepsilon\{xx^T\} = \tilde{A}\varepsilon\{ss^T\}\tilde{A}^T = \tilde{A}\tilde{A}^T = I \tag{33}$$

32

Therefore the number of parameters to be estimated is reduced from n^2 (in A) to $\frac{n(n-1)}{2}$ (in $\tilde{A}$) because $\tilde{A}$ has only $\frac{n(n-1)}{2}$ degrees of freedom.

Now the non-Gaussianity, which is the most important role in the ICA algorithm development, will be discussed. First of all, let us investigate the reason why the non-Gaussianity is so important for ICA algorithm design. From equation (31), it is clear that

$$s = A^{-1}x \tag{34}$$

Suppose there is a linear combination that

$$a = w^T x = \sum_i w_i x_i \tag{35}$$

where w is a vector to be determined. If w were one of the rows of the inverse of A, this linear combination would equal one of the independent components.

Now let us make a change of variable. By setting

$$q = A^T w \tag{36}$$

we have

$$a = w^T x = w^T As = q^T s \tag{37}$$

Thus a is a linear combination of s_i with weights given by q_i. According to the central limit theory which states that a sum of two independent random variables is more Gaussian than any of the original variable, $q^T s$ is more Gaussian than any of the s_i and the Gaussianity will be minimized when it in fact equals to one of the s_i. (here s_i are

33

assumed having identical distributions). In this case, it is obvious that only one of the elements q_i of q is nonzero.

Therefore a vector w that maximizes the non-Gaussianity of $w^T x$ could be taken corresponding to a q which has one nonzero element. This means that

$$w^T x = q^T s \qquad (38)$$

equals one of the independent components. Now it is clear that maximizing the non-Gaussianity of $w^T x$ can help us find the independent components.

There are various methods to measure the non-Gaussianity. A classic measure is the kurtosis which is actually the fourth-order cumulant of a random variable. The kurtosis of a random variable x is denoted by kurt(x) and defined as:

$$kurt(x) = \varepsilon\{x^4\} - 3(\varepsilon\{x^2\})^2 \qquad (39)$$

The kurtosis is zero for Gaussian random variables and non-zero for most of the non-Gaussian random variables. By maximize equation (39) under the constraint that $\|w\| = 1$, the independent components can be computed. A weak point of using kurtosis to measure the non-Gaussianity is that the kurtosis is highly sensitive to the outliers. Therefore, a single value can make kurtosis large. A more robust measure of non-Gaussianity is the *negentropy*. In information theory, the entropy H of a discrete random variable X is defined as:

$$H(X) = -\sum_i P(X = \xi_i) \, log P(X = \xi_i) \qquad (40)$$

34

and for continuous random variables and vectors, it is defined as:

$$H(x) = - \int p(x) log p(x) dx \qquad (41)$$

To obtain a measure of non-Gaussianity that is zero for Gaussian variables and always non-negative, the *negentropy J* is constructed and defined as:

$$J(x) = H\left(x_{gauss}\right) - H(x) \qquad (42)$$

Negentropy is a statistically well justified measure of non-Gaussianity, however it is computationally very difficult. Therefore simpler approximations of the *negentropy* have been developed.

Classic approximations typically have the form similar to

$$J(x) \approx \frac{1}{12} \varepsilon \{x^3\}^2 + \frac{1}{48} kurt(x)^2 \qquad (43)$$

When the random variable *x* has zero mean and unit variance (as we have here), maximizing this approximation is simply equivalent to maximizing the absolute value of the kurtosis, which means this approximation suffers from the same non-robustness encountered by kurtosis. Therefore a more sophisticated approximation is developed.

By using the expectations of general non-quadratic functions to generalize the higher-order cumulant approximation, the *negentropy* can be expressed as:

$$J(x) \approx k_1(\varepsilon\{G_1(x)\})^2 + k_2(\varepsilon\{G_2(x)\} - \varepsilon\{G_2(v)\})^2 \qquad (44)$$

35

where k_1 and k_2 are positive constants, v is a Gaussian variable of zero mean and unit variance. G_1 and G_1 are non-quadratic functions, G_1 is odd and G_2 is even. This approximation of *negentropy* is better than the one given in equation (43) when G_1 and G_2 are wisely chosen. Two useful choices of G have been proved to be

$$G_1(x) = \frac{1}{a_1} logcosh a_1 x \tag{45}$$

and

$$G_2(x) = -e^{(\frac{-x^2}{2})} \tag{46}$$

where $1 \leq a_1 \leq 2$ is some suitable constant, often taken equal to 1.

In the case only one non-quadratic function G is used to approximate the *negentropy*, the approximation becomes:

$$J(x) \propto [\varepsilon\{G(x)\} - \varepsilon\{G(v)\}]^2 \tag{47}$$

Now let us discuss the fixed-point algorithm. An intuitive way of finding w as a fixed point is to make it equal to the gradient of the measure of non-Gaussianity and do the iteration with normalization of its norm to unity.

$$w \leftarrow \varepsilon\{xg(w^T x)\} \tag{48}$$

where $g(\cdot)$ is the derivative of a non-quadratic function G.

36

This iteration process does not, however, converge very well. A modified iteration process must be found. Since adding a multiplication of *w* on both side of equation (48) will not change the fixed point, we have

$$w = \varepsilon\{xg(w^T x)\} \Leftrightarrow (1 + \alpha)w = \varepsilon\{xg(w^T x) + \alpha w\} \qquad (49)$$

where α is a constant. And by mathematical derivation, the iteration process finally becomes

$$w \leftarrow \varepsilon\{xg(w^T x) - \varepsilon\{g'(w^T x)\}w\} \qquad (50)$$

This process will continue running till the *w* converges to the fixed point.

The ICA algorithm discussed above is called FastICA which was proposed by Aapo Hyvärinen [111] in 1999. This algorithm converges faster than other available ICA algorithms with high reliability and accuracy. Therefore it was employed to perform the independent component analysis in our work.

ICA was first applied to event-related potential (ERP) analysis by Makeig *et al.* [75] in 1997. After that, many researchers have employed ICA in their ERP research [73-79]. In 2003, Neng Xu *et al.* [112] designed an ICA-based subspace projection method that works with P300-Speller. In their work, after the data collection, they used infomax ICA algorithm to perform independent component analysis on the collected signal. Then the P300 component related ICs are chosen temporally and spatially. Those chosen ICs were then projected back on the scalp to obtain the scalp distribution of P300 potential. Finally they used the peak and area of the brain wave in the P300 window (from 275 ms to 370

37

ms after the beginning of the intensification) to determine target characters. By using this ICA-based subspace projection method, they achieved 100% accuracy of character prediction with 8 repeated trials.

Those techniques mentioned above are all capable of detecting the P300 response and predict the target character by analyzing the EEG data provided by P300-Speller. Krusienski *et al.* [113] investigated and compared these techniques. They concluded that SWLDA is the most practical and reliable P300 classification method working with P300-Speller. However this conclusion was made based on the multiple-trial P300 classification. SWLDA and other techniques share the same drawback. Although it is not a requirement for any of the mentioned algorithms, it is a critical step to achieve high classification accuracy by averaging several trials to remove the background noises and enhance the magnitude of P300 response before applying the P300 classifier on EEG signal. This "averaging" step slows down the communication process of the P300 BCI. For single trial P300 classification, without the "averaging" step, SWLDA might not be the best candidate.

It would clearly be advantageous to design a method that would allow reliable detection of P300 response in a single trial. To design such a method, not only the "averaging" step is a curdle to be passed on the road, the limitations of each of the techniques also need to be considered. SWLDA may propagate the error incurred in the feature extraction process. SVM needs to select the kernel function and adjust the parameters very carefully to obtain a good result. This process is very tedious and time consuming. Wavelet brings heavy computation load and needs to select the mother wavelet function wisely for

38

accurate classification. ICA cannot guarantee that the computed IC set will be correctly mapped to different features including P300 response.

After observing the shortcomings of each of the techniques and careful consideration of the advantages and disadvantages of these algorithms, we have proposed a blind tracking ICA algorithm to perform the single trial P300 classification task. This algorithm provides a solution for the feature mapping problem faced by ICA and dramatically improved the single trial P300 classification accuracy. In addition, we also proposed a simpler and practical P300 classification method—variance analysis based single trial P300 classification. Both of the proposed methods will be discussed in details in the following chapters.

39

SINGLE TRIAL P300 CLASSIFICATION METHOD: BLIND TRACKING BASED INDEPENDENT COMPONENT ANALYSIS

As mentioned in the previous chapter, ICA is a blind source separation technique. It can decompose a signal into statistically independent components (ICs). Since the ICs are related to different features of the signal, the problem is then to map them to the related features and determine which ones are related to P300 response. In other words, ICA has the ability to reveal the hidden features even if they are buried in the background noise. This ability makes it possible to detect P300 via a single trial. However, ICA is not guaranteed to find a standard IC set that can be clearly mapped to P300 response and other features due to the variations in decomposition. In this chapter, we propose a "blind tracking" method to acquire the standard IC set. This "blind tracking" method, working with single trial P300 classification of ICA, increased the processing speed and achieved high accuracy.

3.1 P300 Brain Computer Interface System

The P300-BCI system block diagram with the signal flow is shown in Figure 13. The shaded blocks indicate the components required for generating the reference templates

40

during the offline analysis and the solid blocks indicate the components required for detection and classification during the online analysis.

In the offline analysis, data is collected in the same way as online analysis (acquisition method is described in the next section). However there is no classification yet during this stage. After the data acquisition, the collected data (40 trials) is preprocessed to filter out high frequency noises and reduce the number of parameters to be estimated. The blind tracking is then performed to search for the standard IC set and standard activation matrix (the standard IC set and standard activation matrix are defined in section 3.5). The results are stored in the system.

In the online analysis, the activation matrix of an incoming signal is computed by using the stored standard IC set and compared to the stored standard activation matrix. The classification is done by using correlation based or discriminant analysis (DA) based majority voting scenario. In this work, 40 trials (480 flashes) were classified in the online analysis. Details of the sub-blocks in Figure 13 are discussed in the following sections.

3.2 Experiment Data Acquisition

Five young adults from the University of South Florida (USF) participated in the experiment. Each participant visited the BCI laboratory in the Department of Psychology at USF once for 120 minutes. Each participant sat upright in front of a computer monitor, which presented a 6x6 visual matrix of letters and numbers. The participant was asked to focus attention on a specified letter in the matrix and silently count the number of times the target character intensified, until a new character was specified for selection. The

41

EEG signal was recorded using a cap (Electro-Cap) embedded with 16 electrode locations distributed over the entire scalp as shown in Figure 14. The EEG signal was band pass filtered 0.1–60 Hz and amplified with an amplifier (20,000×), digitized at a rate of 160 Hz. A sample of the collected EEG signal is shown in Figure 15.
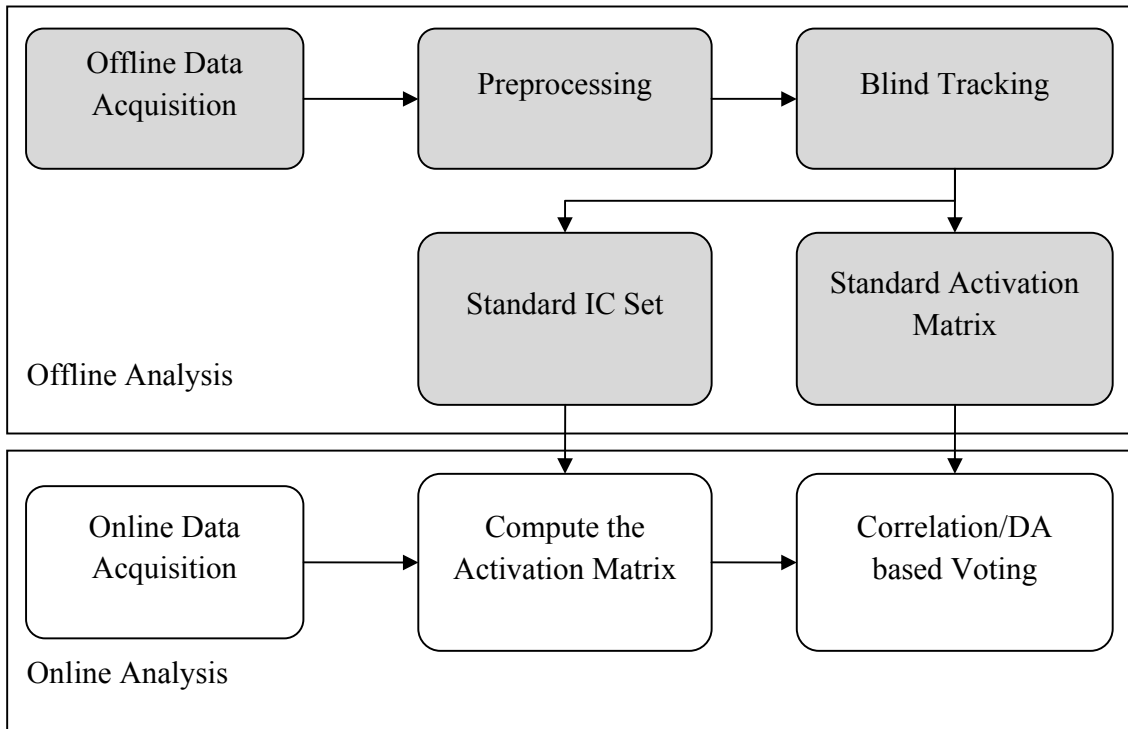
Figure 13: Diagram of the P300 BCI System

Figure 14: Illustration of the data acquisition equipments.
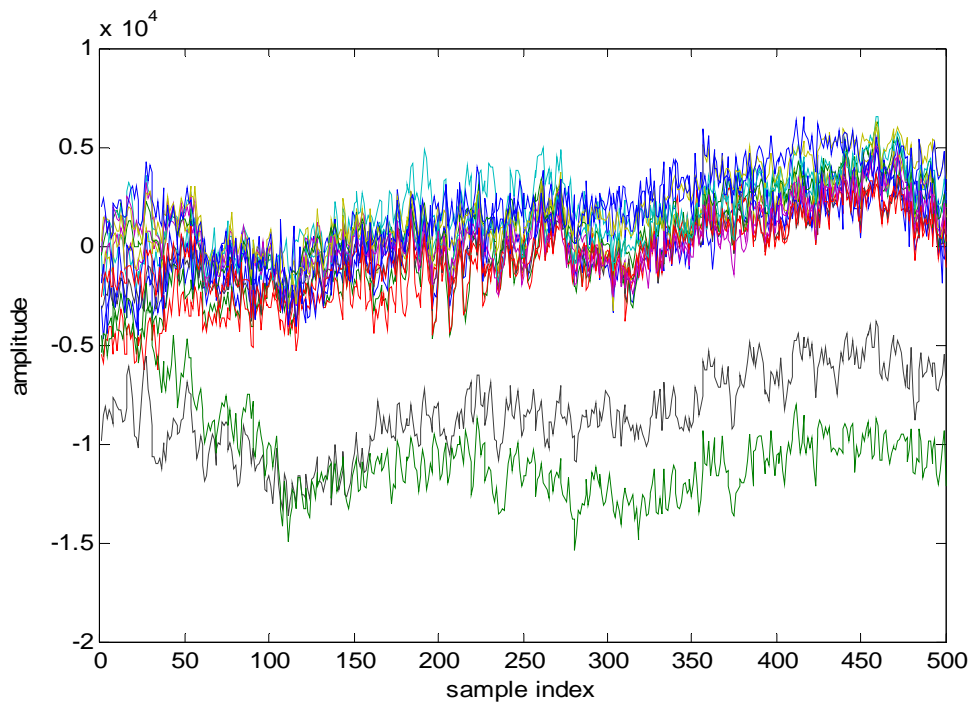
Figure 15: A sample of the collected EEG signal.

43

3.3 Data Structure

The rows and columns of the P300-BCI Speller were intensified for 75 ms with 100 ms interval between intensifications. Because of the delay of P300 occurrence, the EEG signal segments from 175 ms to 350 ms following each intensification were used as our experiment segments. 480 segments from each channel including 80 from target flash (the intensification of row or column that contains the desired character) and 400 from non-target flash (the intensification of row or column that does not contain the desired character) were extracted for offline analysis.

In order to simplify the performance assessment, the participants were asked to spell out some given words by using P300-BCI Speller in both offline and online data acquisition process in this work. However during the real time "spelling", the participant can arbitrarily choose any character he/she wants other than the specified one.

Table 1: The EEG data structure in our experiment

Total number of EEG segments	480×16	Segment Length	175 ms
Sampling Frequency	160 Hz	Number of Samples in Each Segment	28
Intensification Duration	75 ms	Interval Time	100

The details of the sub-blocks are discussed in the following sections.

44

3.4 Preprocessing

All the extracted EEG signals from the 16 channels (electrodes) were low pass filtered to remove the background noise with cut-off frequency set as 10 Hz. Before the independent components (ICs) of the EEG signals being computed, the observed vector x of EEG signals were centered and whitened to obtain uncorrelated components and unit variances. Let us express an EEG signal x as in equation (31), which is $x = As,$ where s is the independent components set and A is the linear transformation from s to x. We call it the activation matrix. Then after the preprocessing, the number of parameters needs to be estimated in A reduced from n^2 to about $n(n-1)/2$ [109].

3.5 Independent Component Analysis

As described in Chapter 2, independent component analysis (ICA) is a statistical and computational technique for revealing hidden factors that underlie sets of random variables, measurements, or signals. It is a good solution to the famous "cocktail problem" which is a Blind Source Separation (BSS) problem. The following example describes the cocktail problem. Two speakers (S_1 and S_2) speak simultaneously in a room with two recorders (R_1 and R_2) recording their speech at different locations in the room. The recorded signals, $R_1(t)$ and $R_2(t)$, can be expressed as:

$$R_1(t) = a_{11}S_1(t) + a_{12}S_2(t)$$
$$R_2(t) = a_{21}S_1(t) + a_{22}S_2(t)$$

$$(51)$$

We can solve these equations for S_1 and S_2 with known values of a_{11}, a_{12}, a_{21} and a_{22}. Unfortunately these weights (a's) are unknowns and these equations can only be solved by ICA under the assumption that S_1 and S_2 are independent non-Gaussian signals. For

45

the EEG signal, a number of electrodes were put on different locations on the scalp to record the signal. So it can be considered as a "cocktail party" problem. Therefore, it is reasonable to apply ICA on EEG signal to identify those independent sources and map them to P300.

There are various ICA algorithms, such as Infomax [114], JADE [115] and FastICA [111] which can successfully compute the independent components by maximizing the non-Gaussianity or *negentropy* [109] (measurement of non-Gaussianity of the ICs). In this work, FastICA is chosen to perform ICA because it converges much faster than other algorithms with high reliability.

3.6 Blind Tracking of Optimal Independent Component Set

Although ICA can help us find the independent sources of some mixed signals, it is not guaranteed that the computed IC set can be clearly mapped to P300 response and other features because there are actually many ways to decompose the mixed signals into independent components. In our work, it is assumed that P300 related ICs are uncorrelated to the hyperplane defined by other ICs and pointed at some unknown directions in the multidimensional space. We need to identify those specific ICs pointing in an unknown direction from a "blind" beginning. Our scenario is to randomly choose a starting vector and compute all the ICs from that starting point. This process was repeated 50 times and therefore 50 different IC sets and activation matrices were computed as in equation (31). For each IC set, the activation status of each IC in each channel was described by the coefficients stored in the activation matrix A. The IC set whose activation matrices have the largest difference for target and non-target signals was

46

defined as the standard IC set. This optimal "standard" IC set may not always perfectly match the "real" standard IC set hidden in the signal. However, it is better than the other 49 IC sets.

In the IC sets computing process, the average of 400 preprocessed 175 ms EEG signals from non-target flash is set as the "standard non-target flash" signal, denoted as x_{nt}. Similarly, the average of 80 preprocessed EEG signals from target flash is set as the "standard target flash" signal, denoted as x_t (Figure 16 and17 show the samples of the averaged and lowpass filtered EEG signals).
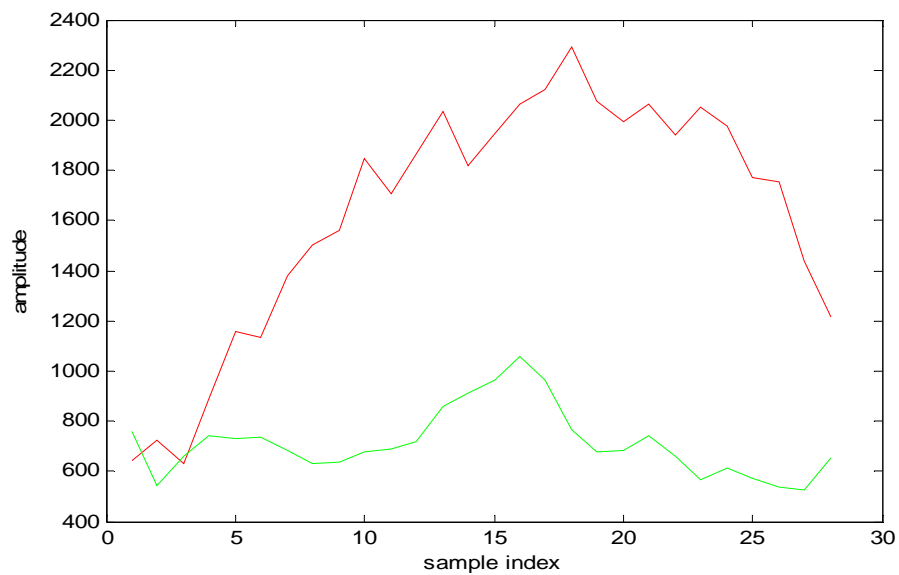
Figure 16: A sample of the averaged target and non-target EEG signal in channel 1.

47
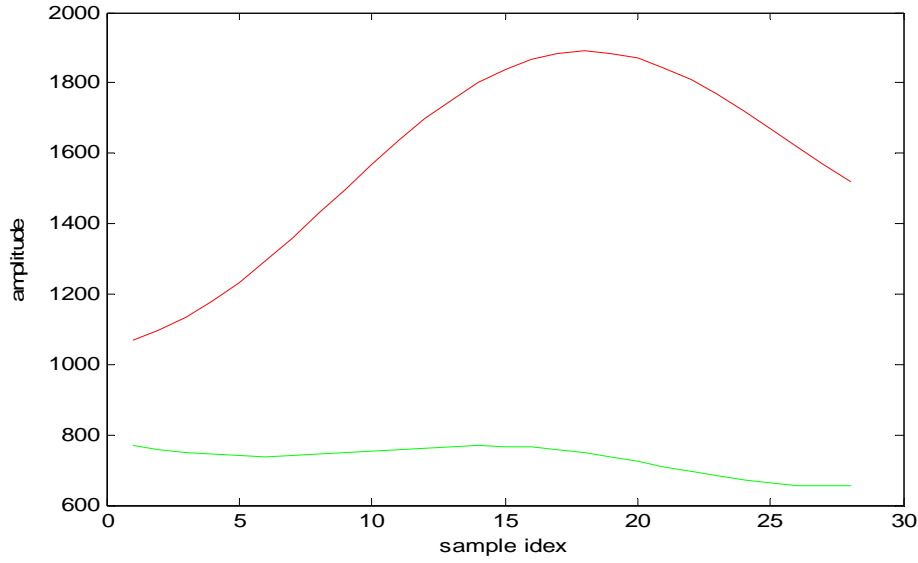
Figure 17: A sample of the averaged and lowpass filtered (cutoff frequency = 10 Hz) target and non-target signals.

By applying FastICA with a random starting vector, the independents components vector s and the mixing matrix A of x_t can be computed and expressed as:

$$x_t = A_t s_t \qquad (52)$$

Here, we assume that the EEG signal from target flash contains more components than those from non-target flash. This is reasonable since the EEG signal of target flash is constituted of "background noise" and P300 response while the EEG signal of non-target flash is constituted of "background noise" only. By substituting s_t and x_{nt} in equation (31), we can solve for A_{nt} that shows the activation status of the ICs underlying in x_{nt}. We investigated 50 A_{nt} and A_t matrices and chose the best pair that most clearly shows the differences between target and non-target signals as the reference activation matrices of target and non-target signals. Meanwhile we set the associated s_t as the standard IC set.

48

In the standard IC set, 3 ICs with the largest variation of their coefficients in A_{nt} and A_t are considered related to the P300 response. Their activation status in different channels will be used as the feature for P300 identification.

3.7 Correlation Method and Majority Vote Scenario

In this part, we use subject 1 as an example to describe the idea. All discussions are based on the processing results for subject 1. In our experiment, for subject 1, 16 ICs were computed and 3 of them, IC 2, 4 and 11, were considered having strong relation to P300. Their "activation pattern" (the activation status of the 3 ICs in the activation matrices) in all the 16 channels of the standard target/non-target flash are investigated and recorded as the reference pattern of target/non-target flashes. For an unknown incoming flash, its activation matrix is computed and the "activation pattern" of IC 2, 4, and 11 is extracted. If it is a target flash, the activation pattern of the P300 related ICs should be more similar to the target reference pattern, otherwise the activation pattern should be more like the non-target reference pattern. We use Pearson product-moment correlation coefficient ρ as the measurement of the similarity.

$$\rho = corr(i,r) \tag{53}$$

where i is a vector that represents the activation status of a chosen IC of an incoming signal, r is a vector that represents the activation status of the same IC in the target or non-target reference. According to the distribution of the correlation value ρ, we can appropriately choose the threshold value t that maximizes the correct target/non-target identification rate. We performed 1 and 3 ICs based classification in this work. The general classifying criteria can be expressed as:

49

If
$$\sum_j \rho_j \geq t, \tag{54}$$

the incoming signal is a target, otherwise it is a non-target, where $j = 1$ or 3, ρ_j is the correlation value according to the j-th P300 related IC. For 1 IC (IC4) based identification, t was set as 0.2 and for 3 ICs (IC2, IC4 and IC11) based identification, t was set as 0.5 (All the threshold values were chosen by maximizing the correct identification rate. Different subjects may have different threshold values according to their individual ρ distribution. Figures 18-20 show the distributions of the correlation values of the incoming signal and the target reference)
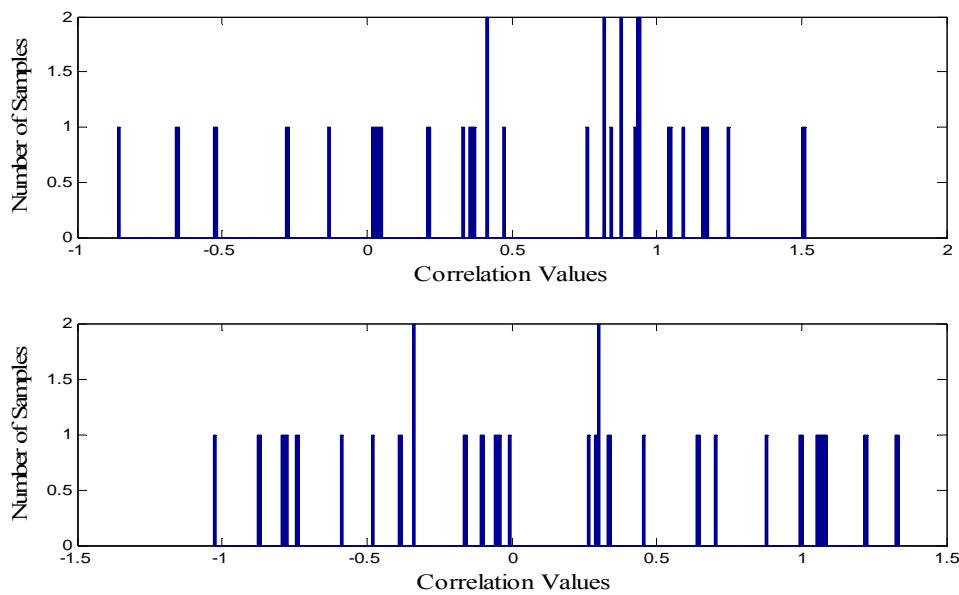
Figure 18: Top: The correlation value distribution of target signals and the standard target reference (Computed by using IC 4 only). Bottom: The correlation value distribution of non-target signals with the standard target reference (Computed by using IC 4 only).

50
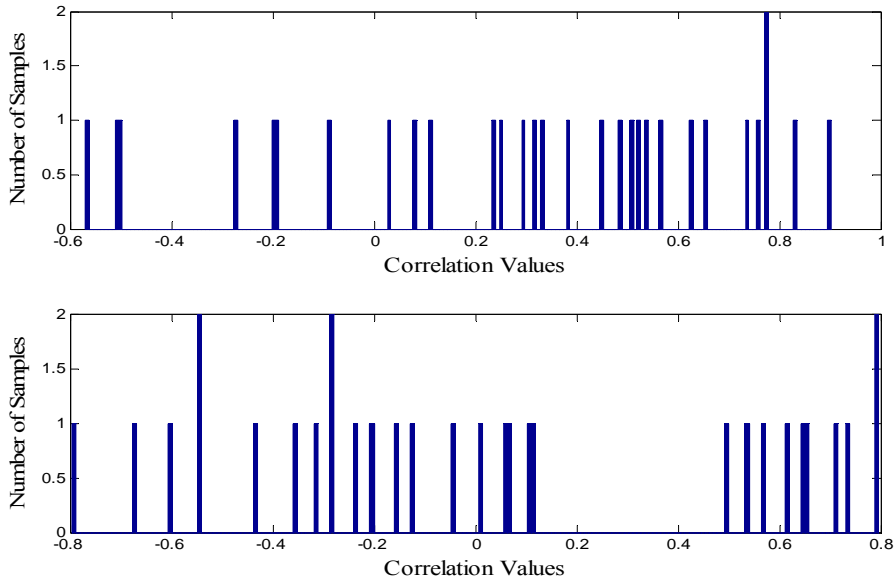
Figure 19: Top: The correlation value distribution of target signals and the standard target reference (Computed by using IC 2 and IC 4). Bottom: The correlation value distribution of non-target signals with the standard target reference (Computed by using IC 2 and IC 4).
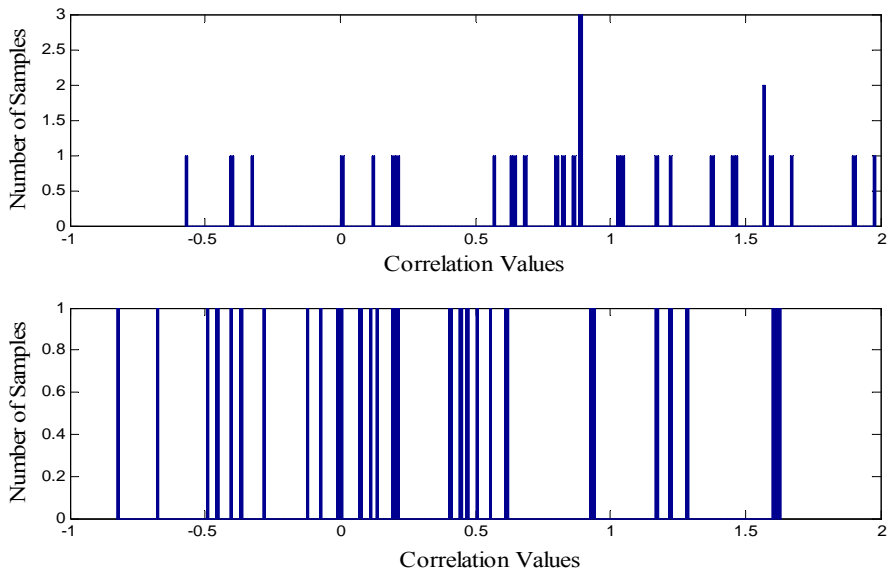
Figure 20: Top: The correlation value distribution of target signals and the standard target reference (Computed by using IC 2, IC 4 and IC 11). Bottom: The correlation value distribution of non-target signals with the standard target reference (Computed by using IC 2, IC 4 and IC 11).

51

In addition to directly summing the correlation values from corresponding ICs, we also use 3 P300 response related ICs to vote according to the following voting criteria:

If $\rho_1 \geq t_1$, then vote for target; Otherwise vote for non-target.

If $\rho_2 \geq t_2$, then vote for target; Otherwise vote for non-target.

If $\rho_3 \geq t_3$, then vote for target; Otherwise vote for non-target.

where ρ_1, ρ_2 and ρ_3 are correlation values that correspond to IC2, IC4 and IC11 respectively. t_1=0.3, t_2=0.2 and t_3=0.34. The majority vote of them determines the label of an incoming EEG signal.

The results of our experiment are shown in Tables 2-5.

Table 2: Classification results of correlation method by using 1 IC (IC 4)

Category	Correctly Classified	Incorrectly Classified	Total	Accuracy	Error Rate
Target	21	9	30	70%	30%
Non-target	19	11	30	63.3%	36.7%

Table 3: Classification results of correlation method by using 3 ICs (IC 2, 4 and 11)

Category	Correctly Classified	Incorrectly Classified	Total	Accuracy	Error Rate
Target	23	7	30	76.7%	23.3%
Non-target	21	9	30	70%	30%

52

Table 4: Classification results of discriminant analysis based majority voting (3 ICs)

Category	Correctly Classified	Incorrectly Classified	Total	Accuracy	Error Rate
Target	23	7	30	76.7%	23.3%
Non-target	22	8	30	73.3%	26.7%

Table 5: Classification results of correlation based majority voting of 5 subjects (3 ICs)

Subjects	Accuracy (%)		
	Target	Non-target	Overall
Subject 1	76.7	73.3	73.9
Subject 2	66.7	74.7	73.3
Subject 3	70	76.7	75.6
Subject 4	66.7	80.6	78.3
Subject 5	73.3	79.3	78.3
Average	*70.7*	*76.9*	*75.9*

3.8 Discriminant Analysis and Majority Vote Scenario

Although "Blind Tracking" can help us find an optimized standard IC set, the set is most likely not the best or optimal set yet. In fact, if the "Blind Tracking" is kept running for a great number of times, it is possible to find an IC set that matches the best standard IC set

fairly well. However, the tracking and testing process is rather time consuming. It may take us days or weeks to calibrate the standard IC set for each individual subject. To avoid this time consuming process, another classification method combining the Independent Component Analysis and Discriminant Analysis was designed and performed for this single trial classification problem. For each of the 16 ICs computed by using ICA algorithm, their activation status in the 16 channels were recorded and stored in a matrix. Half of the data stored in these matrices were used as training data for the discriminant analysis and the other half were used as testing data. The activation status of each IC in the 16 channels was classified into two categories, target and non-target. "Target" means the activation status of the specific IC indicates that the incoming signal is a target signal. "Non-target" means the activation status of the specific IC indicates that the incoming signal is a non-target signal. After the 16 ICs were classified, their "majority" vote (the majority number may vary for each individual subject, in this work, it was set as 8 for subject 1, 2, 4, 5, and 7 for subject 3) determined whether an incoming signal is a target signal. In our experiment, we also choose 3 of the 16 ICs that provided the highest classification accuracy to perform the majority voting. The results are summarized in the Tables 6-8.

54

Table 6: Classification results of discriminant analysis based majority voting of 5 subjects (16 ICs)

Subjects	Accuracy (%)		
	Target	Non-target	Overall
Subject 1	73.3	81.3	80
Subject 2	66.7	78.6	76.7
Subject 3	73.3	82.7	81.1
Subject 4	60	81.3	77.8
Subject 5	80	80	80
Average	*70.7*	*80.8*	*79.1*

Table 7: Classification results of discriminant analysis based majority voting of 5 subjects (3 ICs)

Subjects	Accuracy (%)		
	Target	Non-target	Overall
Subject 1	73.3	81.3	80
Subject 2	60	80	76.7
Subject 3	60	84	81.1
Subject 4	60	77.3	74.4
Subject 5	80	81.3	81.1
Average	*66.7*	*80.8*	*78.7*

Table 8: The results comparison of blind tracking based ICA and single trial SWLDA

Performance Metric	ICA (Correlation Method)	ICA (Discriminant Analysis)	SWLDA
Processing Time / Character (ms)	1225	1225	2100
Overall Accuracy (%)	75.9	79.1	45
Correctly Communicated Characters/min	28.2	30.6	12.8

3.9 Discussion of the Results

The 1 and 3 ICs based correlation method and 3 ICs based voting scenario were tested by 180 EEG signals including 150 from non-target flash and 30 from target flash. For 1 IC based correlation method, with t = 0.2, we achieved 70% and 63.3% accuracies for target and non-target identification, respectively. For 3 ICs based correlation method, with t = 0.5, these accuracies increased to 76.67% and 70%, respectively. The majority voting scenario provided the best identification accuracies of 76.67% and 73.3% for target and non-target, respectively. This experiment was repeated for the other four subjects. The overall accuracies achieved were 73.3%, 75.6%, 78.3%, and 78.3%, respectively. In our research we prefer to reduce the type II error because if we fail to identify a target flash, the identification process can be repeated till the target successfully identified. But if a signal is falsely identified as "target", this error will not be realized until the final character selection. Considering this, we may reduce the type II error by decreasing the t value. However, the tradeoff is that the processing time will increase due to repetition. In

56

this work, the discriminant analysis based voting scenario achieved 80%, 76.7%, 81.1%, 77.8% and 80% overall accuracies for 5 individual subjects when all 16 ICs were involved in the voting. It obtained higher accuracies than that those achieved by correlation method based voting. However, there is no significant improvement for target identification accuracy. When only 3 ICs were involved in the voting, the overall accuracies turned out to be 80%, 76.7%, 81.1, 74.4%, and 81.1%. It did not show significant improvement for the overall accuracies by reducing the number of voting ICs. Meanwhile the decrease of target classification accuracies for some subjects was observed. As mentioned before, in this work we prefer reducing the type II error. Therefore, 16 ICs involved voting is more suitable for this classification problem.

The proposed single trial ICA algorithm for P300 response classification significantly reduces the processing time by removing the time consuming step due to "averaging" used in other algorithms. Furthermore, this algorithm will stop and start the next "target searching" whenever it hits a "target". Thus the expecting character identifying time is given by $\varepsilon(t) = 7$ flashes $= 175 \times 7 = 1225$ ms. A comparison of the time efficiency and accuracy between this method and single trial SWLDA is provided in Table 8.

Our algorithm achieved an overall average accuracy of 79.1% in 1225 ms while SWLDA achieved only 45% overall average accuracy in 2100 ms. In other words, the P300 based single trial ICA algorithm is 171.4% more time efficient than the single trial SWLDA with 34.1% more overall average accuracy. Moreover, comparing the data communication speed in terms of correctly communicated characters per minute, our

57

method achieved 28.2 characters/min and 30.6 characters/min for correlation and discriminant analysis based majority voting while SWLDA achieved 12.8 characters/min.

CHAPTER 4

SINGLE TRIAL P300 CLASSIFICATION METHOD: VARIANCE ANALYSIS BASED P300 CLASSIFICATION

In the previous chapter, we discussed the development of a single trial independent component analysis (ICA) method with blind tracking of standard IC set to detect a chosen character in real-time in the P300-BCI speller that achieved an overall accuracy of 79.1% [88]. However, it requires the manual selection of the optimal standard IC set. This process is very time consuming and makes the algorithm less practical for implementation. This motivated us to develop a simpler and practical approach to solving the problem. In this work by taking a completely different approach to currently used signal processing methods, we propose a simple statistical analysis based method for P300 response detection and classification with high accuracy.

4.1 P300 Brain Computer Interface System

The block diagram with the signal flow of the proposed BCI system is shown in Figure 21. The shaded blocks indicate the components required for generating the reference templates during the offline analysis and the solid blocks indicate the components required for detection and classification during the online analysis.

59

In the offline analysis, after the data acquisition, the collected data is preprocessed to filter out high frequency noises and the signal is segmented as target and non-target signals. The average of 60 target signals is used as standard target signal and stored in the system.

In the online analysis, the incoming signal is transformed by using the standard target signal and the variance of the transformed signal is analyzed for P300 detection. Details of the sub-blocks shown in Figure 21 are discussed in the following sections.
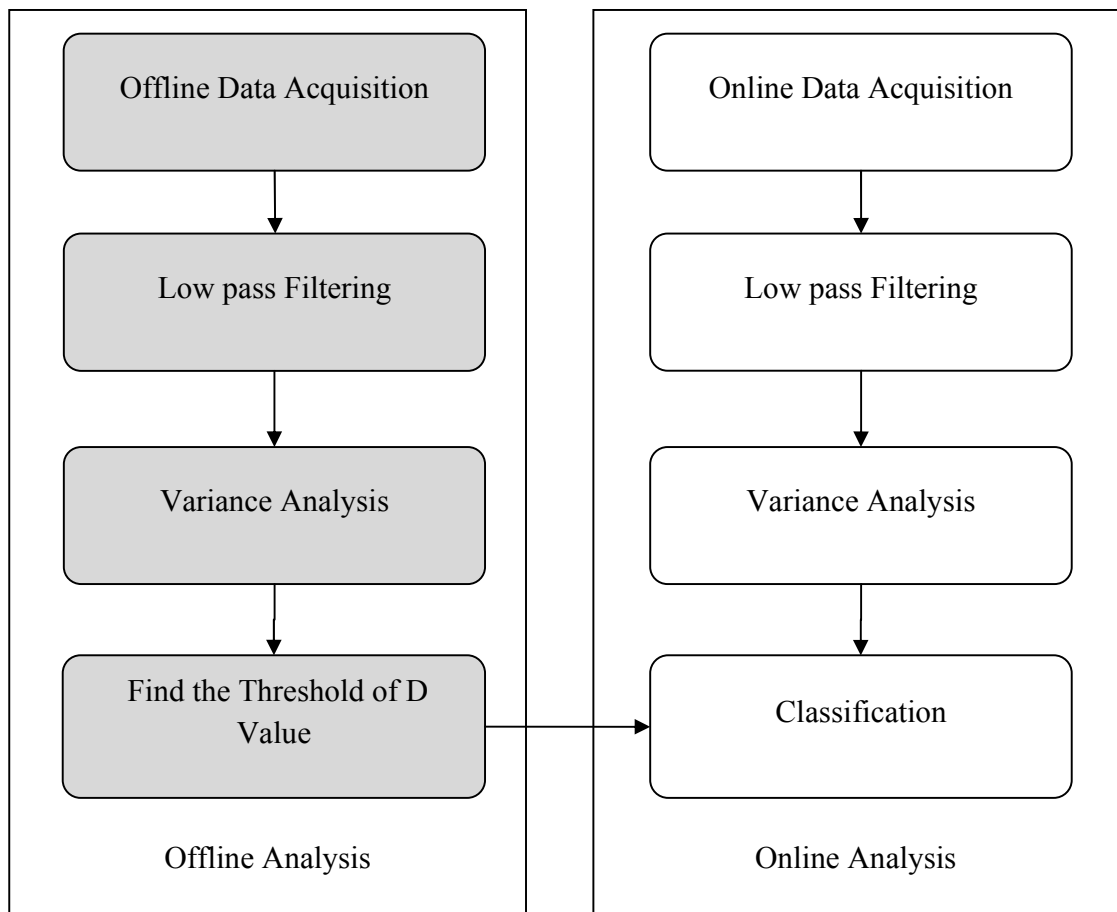
Figure 21: The proposed variance analysis based P300-BCI system

4.2 Data Acquisition and Structure

The data acquisition process and the data structure are the same as that described in section 3.2. After the data collection, 360 segments from each channel including 60 from target flash (the intensification of row or column that contains the desired character) and 300 from non-target flash (the intensification of row or column that does not contain the desired character) were extracted for offline analysis.

In order to simplify the performance assessment for this work, the participants were asked to spell out some given words by using P300-Speller in both offline and online data acquisition process. However during the real time "spelling", the participant can arbitrarily choose any character he/she wants other than the specified one.

4.3 Preprocessing

All the extracted EEG signals from the 16 channels (electrodes) were low pass filtered to remove the background noise with cut-off frequency set as 10 Hz. 60 target signals were segmented, and their average was stored in the system as standard target signal.

4.4 Variance Analysis

The standard target signal was used to transform the incoming signal. Then the variance of the transformed signal was analyzed to identify the incoming signal as target or non-target. An observed target signal can be expressed as:

$$x_o = x_t + n \tag{55}$$

where x_o denotes the observed target signal whose variance is σ_o^2, x_t denotes the real target signal whose variance is σ_t^2 and n denotes the background noise which is considered as random and stable with variance equal to σ_n^2. Here the independence of the real target signal and the background noise is assumed. Therefore,

$$\sigma_o^2 = \sigma_t^2 + \sigma_n^2. \tag{56}$$

If we average k number of the observed target signals and denote the result as x_{st}, then the variation of the background noise in x_{st} decreases to $\frac{\sigma^2}{k}$. The average target signal, x_{st} can be expressed as:

$$x_{st} = \frac{1}{k}\sum_{i=1}^{k} x_{oi} = \frac{1}{k}\sum_{i=1}^{k}(x_{ti} + n_i) \tag{57}$$

where, x_{oi} is the i-th observed target signal, x_{ti} is the i-th real target signals, n_i denotes the background noise with the *i*-th observed target signal. (Notice that all n_i have the same variance, σ_n^2.), and $k = 60$ (the number of target signals) Ideally the target signals are all identical. Therefore, the variance of x_{st}, denoted by σ_{st}^2, can be written as:

$$\sigma_{st}^2 = \sigma_t^2 + \frac{\sigma_n^2}{k} \tag{58}$$

Thereafter, an incoming signal was transformed for the purpose of variance analysis. Denoting the incoming signal by x_o, we define the transformation 1 as:

$$T_1 = \frac{1}{2}(x_{st} + x_o) \tag{59}$$

and transformation 2 as:

62

$$T_2 = \frac{1}{2}(x_{st} - x_o) \tag{60}$$

In the event that the incoming signal is a target signal, the variance of T_1 is given by:

$$Var(T_1) = Var\left(\frac{1}{2}(x_{st} + x_o)\right) = \sigma_t^2 + \frac{(k+1)\sigma_n^2}{4k} \tag{61}$$

and the variance of T_2 is given by:

$$Var(T_2) = Var\left(\frac{1}{2}(x_{st} - x_o)\right) = \frac{(k+1)\sigma_n^2}{4k} \tag{62}$$

The difference of variances of T_1 and T_2, denoted by D is:

$$D = Var(T_1) - Var(T_2) = \sigma_t^2 \tag{63}$$

In the event that the incoming signal is a non-target signal, the variance of T_1 and T_2 is given by:

$$Var(T_1) = Var(T_2) = \frac{\sigma_t^2}{4} + \frac{(k+1)\sigma_n^2}{4k} \tag{64}$$

Therefore, the difference of variances of T_1 and T_2 is:

$$D = Var(T_1) - Var(T_2) = 0 \tag{65}$$

As indicated by equations (63) and (65), D can be used as a classification measurement since it equals σ_t^2 for target signal and 0 for non-target signal. In practice, there is rather small possibility that D exactly equals σ_t^2 and 0 for target and non-target signal, respectively. However, the D value of non-target signal can be expected to be significantly smaller than that of target signal.

63

Due to this reason, the classification can be refined by setting a threshold α for D with the classification rule defined as:

$$x \in \{\text{Target signal}\}, \text{when } D > \alpha$$

and

$$x \in \{\text{Nontarget signal}\}, \text{when } D < \alpha$$

where x denotes the incoming signal.

4.5 Threshold Value Determination

The way to choose α wisely should follow the general idea that the choice of α should maximize the classification accuracy. Therefore the characteristics of D values were investigated. The means and standard deviations of D values for channel 9, 12 and 15 are given in Table 9. (These 3 channels were used in this work.)

Table 9: The mean and standard deviation of D values for channels 9, 12, and 15

Channel	Parameters	Target Signal	Non-target Signal
9	$\overline{D}$	0.941	-0.153
	S	1.470	1.368
12	$\overline{D}$	0.724	-0.068
	S	1.128	1.155
15	$\overline{D}$	1.315	-0.045
	S	1.030	1.238

64

Note: $\overline{D}$ indicates the sample mean of D values and S indicates the sample standard deviation. All numbers has been normalized by dividing 10^5.

First of all, the D value is assumed to be normally distributed. This assumption was verified by running the *Kolmogorov-Smirnov test* with normal distribution. The null hypothesis (the samples are drawn from normal distribution) cannot be rejected under 0.00001 significant level which indicates it is 99.99999% sure that the D values are coming from normal distribution. According to the results of *Kolmogorov-Smirnov test*, it is a robust assumption that D value is normally distributed. Hence the *two-sample T-test* was performed to test whether the D values of target signal and non-target signal are coming from distributions with equal means. The null hypothesis (the D values of target and non-target signals have equal mean) was rejected under 0.00001 significance level which indicates it is 99.99999% sure that the D value of target signal and non-target signal are coming from different classes with different means.

Now let us investigate the probability density functions (PDF) of the D values of target and non-target signals. Since the D value is normally distributed, its PDF can be written in terms of with the mean μ and the standard deviation σ of the random variable D:

$$f(d) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(d-\mu)^2}{2\sigma^2}} \tag{66}$$

The plot of D value's PDFs of target and non-target is shown in Figure 22.

65
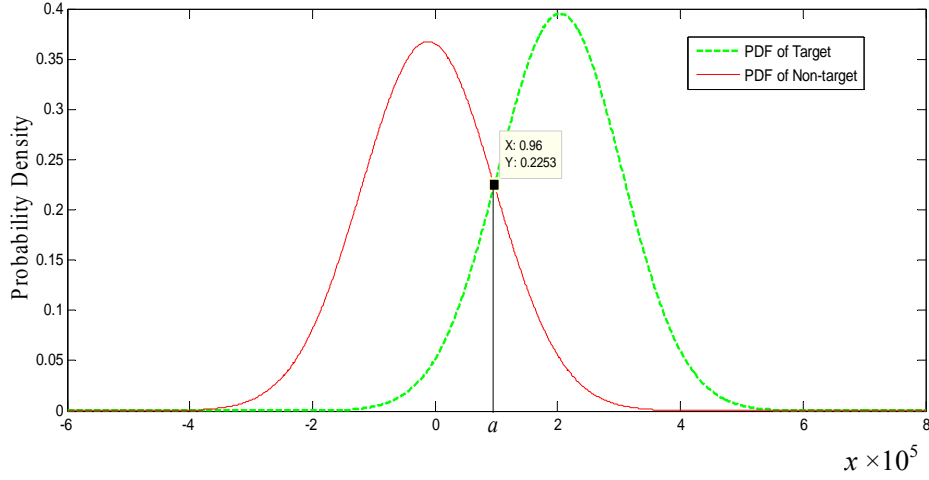
Figure 22: The PDF of the D values of target and non-target signals

The bell shape distribution curve to the right in Figure 22 indicates the PDF of target signal and the bell shape distribution curve to the left in Figure 22 indicates the PDF of non-target signal. α is the point at which the green and red curves intersect. It is easily seen for any interval beyond point α, the integral under the green curve is greater than that under red one which means if a signal has $D > \alpha$, it is more probable to be a target signal. For the same reason, if a signal has $D < \alpha$, it is more likely to be a non-target signal. Therefore, α is chosen as the optimal threshold value that minimizes the probability of incorrect classification and hence maximizes the classification accuracy.

α value can be determined by solving this equation:

$$\frac{1}{s_t\sqrt{2\pi}}e^{-\frac{(\alpha-\bar{D}_t)^2}{2s_t^2}} = \frac{1}{s_n\sqrt{2\pi}}e^{-\frac{(\alpha-\bar{D}_n)^2}{2s_n^2}} \tag{67}$$

where $\bar{D}_t$ and S_t are the sample mean and standard deviation of D values of target signal respectively and $\bar{D}_n$ and S_n are the sample mean of D values of non-target signal respect-

66

ively. Therefore the left hand side is the PDF of the *D* value of target signal and the right hand side is the PDF of the *D* value of non-target signal.

4.6 Single and Multi-Channel Classification

In this work, the single trial target/non-target classification was done by using both single and multiple channels. The employed channels have large differences in mean of the *D* values between target and non-target signals and meanwhile have relatively small standard deviation in D value. The first employed channel was channel 15. Then two channels, channels 15 and 12, were combined to conduct the classification. After that, channel 9 was combined with channels 12, and 15 to accomplish the classification task. The best accuracy was provided by the combination of channels 15 and 12.

The threshold of D values, α was set as 0.703, 0.953, and 1.394 for channel 15, the combination of channels 12 and 15 and the combination of channels 9, 12, and 15, respectively. The results of our experiment are shown in Tables 10-11. The accuracy/error rates are the percent of target/non-target flashes correctly/incorrectly classified.

Table 10: Classification results of single and multi-channel variance analysis based classification

Channel / α	Accuracy (%)			
	Theoretical Value		Practical Value	
	Target	Non-target	Target	Non-target
Ch 15 / α=0.703	72.38	75.60	75	78.33
Chs. 15 and 12 / α=0.953	85.88	83.71	88.33	81.67
Chs. 15, 12 and 9 / α=1.394	76.29	78.74	76.67	75

In Table 10, the theoretical accuracy is defined as the area underneath the PDF curve of D value from $-\infty$ to α or from α to $+\infty$ for target and non-target signal, respectively. The practical accuracy is given by the percentage of correct classification achieved by choosing a certain threshold, α, and applying it to the real data. The normality of D value distribution has been verified by *Kolmogorov-Smirnov test*. Therefore, the expectation of practical accuracy should converge to the theoretical accuracy in the long run. In our experiment, the theoretical accuracies associated with the single trial single channel classification employing channel 15 ($\alpha = 0.703$) were 72.38% and 75.60% for target and non-target signals, respectively. Practically, the classification accuracies are 75% and 78.33% for target and non-target signals, respectively. These accuracies are reasonably close to the theoretical values as expected. After single channel classification, the multi-channel classification was performed. For classification with channels 15 and 12

68

combined (α = 0.953), the practical accuracies increased to 88.33% and 81.67% for target and non-target, respectively while the theoretical accuracies increased to 85.88% and 83.71% for target and non-target signals, respectively. The explanation for the increase in accuracy is that the signal in channel 12 strengthens the signal in channel 15 and picks up information left out by channel 15.

However, it should be pointed out that by arbitrarily adding more channels for classification may not by itself provide better accuracy. The following experimental illustration confirms that conjecture. For classification with channels 15, 12 and 9 combined (α = 1.394), we achieved practical accuracies of 76.67% and 75% for target and non-target classification while the theoretical accuracies were 76.29% and 78.74% for target and non-target, respectively. An explanation for the decrease in the classification accuracies is that channel 9 brings in less information than "noise" which degrades the classification system. This fact implies that the classification accuracy depends on the signal to noise ratio from each contributing channel.

According to our experiment, classification with channels 15 and 12 combined provides the best accuracies for both target and non-target classification. The variance analysis based single trial P300 classification algorithm significantly reduces the processing time by removing the time consuming step due to "averaging" used in other multi-trial algorithms. A comparison of the performance between our method and single trial SWLDA is provided in Table 11. Our algorithm achieved an overall accuracy of 84.8% while SWLDA achieved only 45% overall accuracy. Moreover, comparing the data

communication speed in terms of the correctly communicated characters per minute, our

method achieved 20.5 characters/min while SWLDA achieved 12.8 characters/min.

Table 11: Performance comparison of variance analysis based classification and single
trial SWLDA

Performance Metric	Variance Analysis Based Classification	SWLDA
Overall Accuracy (%)	84.8	45
Correctly Communicated Characters/min	20.5	12.8

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

The aim of every brain computer interface is to translate simulated brain activity into a relevant computer command. The P300-Speller proposed by Farwell and Donchin in 1988 provided researchers a practical way to accomplish this aim. Thereafter, various techniques including the mentioned SWLDA, SVM, matched filter, Wavelet and ICA have been developed to work with P300-Spller for the P300 response classification. Although it is not a requirement for any of the mentioned algorithms, it is yet a critical step to achieve high classification accuracy by averaging several trials to remove the background noises and enhance the magnitude of P300 response prior to applying the P300 classifier. It would clearly be advantageous to employ a method that would allow reliable detection of P300 response in a single trial.

As early as in 1969, Donchin [116] had suggested the possibility of making meaningful comparisons between EEG records obtained with a single presentation of the stimulus and the average evoked potential. From 70s to 90s, researchers have done tremendous work in single trial ERP analysis [107]. They discussed the feasibility of single trial ERP analysis, investigated the factors that may affect the analysis, suggested possible

71

denoising and classification techniques and conducted various experiments on single trial ERP signal analysis. Their work directly or indirectly proved that single trial P300 classification on P300-Spller is possible, feasible and promising. Recently, some researchers [112][117] have applied independent component analysis (ICA) on P300 classification and achieved high classification accuracy in multi-trial experiments. ICA is a blind source separation technique that can decompose a mixing signal into statistically independent components. Therefore it is potentially capable of mapping the corresponding ICs to P300 response directly. With this feature mapping, the EEG signal can also be denoised by discarding the ICs associated with background noise.

The key problem of ICA based P300 classification is that the IC set computed by any available ICA algorithm is not guaranteed with a clear feature and IC mapping. The proposed blind tracking based ICA algorithm provided a solution to this problem. With this algorithm, an optimal IC set with relatively clear feature and IC mapping can be found. The classification results validated the effectiveness of our algorithm. For the same subject the overall accuracy was 66.7% with one IC based classification and 73.3% with three ICs based classification. The correlation based voting scenario provides an overall accuracy of 73.9% while the discriminant analysis based voting scenario provides an overall accuracy of 80%. It is clear that the two majority voting scenarios produced better classification than the one IC and three ICs based classification. This fact testified that the multi-variant analysis of signal provides better distinguishing features than the variant analysis of the EEG signal. The classification result of discriminant analysis based voting is better than that of the correlation based voting. This improvement may be due to the more effective feature combination in discriminant analysis.

In the variance analysis based method, a statistical parameter D has been mathematically derived for target signal classification. By using this parameter, we achieved an overall classification accuracy 84.8% which proved that it is a well defined and useful parameter. In this work, the D values based on 3 chosen channels are computed to provide best performance. Although we know that multi-channel analysis is better than a single channel analysis for EEG signal, the channels need to be carefully selected and appropriately combined to increase the signal to noise ratio. Indiscriminately adding more channels might decrease the classification accuracy.

Over the past forty years, P300-based BCI system has been successfully implemented using simple signal processing techniques such as signal averaging and LDA. Recently more advanced techniques have been used to process the P300 signal and made some achievement. However, for a single trial based target signal classification, the classification accuracy and processing speed still need to be highly improved for more effective communication. In this work, attempts have been made to develop single trial P300 classification methods to detect a chosen character in real-time in the P300-BCI speller. The results indicate that the proposed methods dramatically reduces the signal processing time, improves the data communication rate, and achieves an overall average accuracy of 79.1% for blind tracking based ICA algorithm and 84.8% for variance analysis base single trial P300 response identification.

The blind tracking based ICA algorithm provided 34.1% increase in accuracy and 139% more effective in communication speed over single trial SWLDA. The variance analysis based classification method provided 39.8% increase in accuracy and 60% more effective

in communication speed over single trial SWLDA. Therefore, the proposed methods can be considered to be promising and reasonable solutions for single trial EEG signal classification problem.

5.2 Recommendations for Future Research

There is still room for improving the processing speed and accuracy by optimizing the algorithm. For example, for blind tracking based ICA, we can weigh the voters or modify the voting rule to improve the performance of voting. We can also reduce the number of channels for the standard IC set tracking by using our blind approach, thereby, significantly reducing the computation load. For variance analysis based classification, we can construct a multi-dimensional space by using the D values of different channels as the axis and apply Linear Discriminant Analysis (LDA) to accomplish the classification task. In our experiment, we made an assumption that the P300 response occurs between 175 ms and 350 ms following a target flash, which is not true for some subjects because in some cases P300 shows up in the 350 ms to 500 ms range. This problem can be solved by using appropriate flashing and interval time. Although this work is built up on P300 response, there might be other signals involved when we do the variance analysis. We need to study more cases to test the robustness of these methods. Further optimization of our algorithms by involving statistical models to solve the non-stationary problem [118] will be researched in our future work. The ultimate goal is to further improve the accuracy of the single trial P300 analysis algorithms to make them more suitable for real-world applications and clinical use.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, "Brain-computer interfaces for communication and control", *Clin. Neurophsiol.*, vol. 113, pp. 767-791, Jun. 2002.

[2] U. Hoffmann, J. M. Vesin, and T. Ebrahimi, "Recent advances in brain-computer interfaces", *IEEE International Workshop on Multimedia Signal Processing (MMSP07),* Oct. 2007.

[3] L. A. Farwell and E. Donchin, "Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials", *Electroencephalography and Clinical Neurophysiology*, vol. 70, pp. 510-523, Dec.1988.

[4] J. Vidal, "Toward direct brain-computer communication", Annual review of biophysics and bioengineering vol. 2, pp. 157–180, Jun. 1973.

[5] J. Vidal, "Real-Time detection of brain events in EEG", Proceedings of IEEE, vol. 65, no. 5, pp. 633–641, May 1977.

[6] E. M. Schmidt, "Single neuron recording from motor cortex as a possible source of signals for control of external devices", *Ann. Biomed. Eng*. vol. 8, pp. 339–349, 1980.

[7] M. A. Nicolelis, L. A. Baccala, R. C. Lin, and J. K. Chapin, "Sensorimotor encoding by synchronous neural ensemble activity at multiple levels of the somatosensory system", *Science*, vol. 268, pp. 1353–1358, Jun. 1995.

[8] M. A. Nicolelis, D. Dimitrov, J. M. Carmena, R. Crist, G. Lehew, J. D. Kralik, and S.P. Wise*,* "Chronic, multisite, multielectrode recordings in macaque monkeys", *Proc. Natl. Acad. Sci. U. S. A.* vol. 100, pp. 11041–11046, Sep. 2003.

[9] M. A. Nicolelis, A. A. Ghazanfar, B. M. Faggin, S. Votaw, and L. M. O. Oliveira, "Reconstructing the engram: simultaneous, multisite, many single neuron recordings", *Neuron*, vol. 18, pp. 529–537, Apr. 1997.

[10] M. A. Nicolelis, and S. Ribeiro, "Multielectrode recordings: the next steps", *Curr. Opin. Neurobiol*. vol. 12, pp. 602–606, Oct. 2002.

[11] M. A. Nicolelis, "Actions from thoughts", *Nature* vol. 409, pp. 403–407, Jan. 2001.

[12] A. A. Ghazanfar C. R. Stambaugh, and M. A Nicolelis, "Encoding of tactile stimulus location by somatosensory thalamocortical ensembles", *J. Neurosci.* vol. 20, pp. 3761–3775, May 2000.

[13] D. J. Krupa, M. C. Wiest, M. G. Shuler, M. Laubach, M. A. Nicolelis, "Layer-specific somatosensory cortical activation during active tactile discrimination", *Science,* vol. 304, pp. 1989–1992, Jun. 2004.

[14] J. Wessberg, C. R. Stambaugh, J. D. Kralik, P. D. Beck, M. Laubach, J. K. Chapin, J. Kim, S. J. Biggs, M. A. Srinivasan, and M. A. Nicolelis, "Real-time prediction of hand trajectory by ensembles of cortical neurons in primates", *Nature,* vol. 408, pp. 361–365, Oct. 2000.

[15] M. D. Serruya, N. G. Hatsopoulos, L. Paninski, M. R. Fellows, and J. P. Donoghue, "Instant neural control of a movement signal", *Nature* vol. 416, pp. 141–142, Mar. 2002.

[16] D. M. Taylor, S. I. H. Tillery, A. B. Schwartz, "Direct cortical control of 3D neuroprosthetic devices", *Science,* vol. 296, pp. 1829–1832, Jun. 2002.

[17] J. M. Carmena, M. A. Lebedev, R. E. Crist, J. E. O'Doherty, D. M. Santucci, D. F. Dimitrov, P. G. Patil, C. S. Henriquez, and M. A. Nicolelis, "Learning to control a brain-machine interface for reaching and grasping by primates", *PLoS Biol.* 1, E42, Nov. 2003.

[18] M. A. Lebedev, J. M. Carmena, J. E. O'Doherty, M. Zacksenhouse, C. S. Henriquez, J. C. Principe, and M. A. Nicolelis, "Cortical ensemble adaptation to represent velocity of an artificial actuator controlled by a brain–machine interface", *J. Neurosci.* vol. 25, pp. 4681–4693, May 2005.

[19] C. Mehring, J. Rickert, E. Vaadia, S. Cardoso de Oliveira, A. Aertsen and S. Rotter, "Inference of hand movements from local field potentials in monkey motor cortex", *Nat. Neurosci.* vol. 6, pp. 1253–1254, Nov. 2003.

[20] J. Rickert, S. Cardoso de Oliveira, E. Vaadia, A. Aertsen, S. Rotter, and C. Mehring, "Encoding of movement direction in different frequency ranges of motor cortical local field potentials", *J. Neurosci.* vol. 25, pp. 8815–8824, Sep. 2005.

[21] B. Pesaran, J. S. Pezaris, M. Sahani, P. P. Mitra, and R. A. Andersen, "Temporal structure in neuronal activity during working memory in macaque parietal cortex", *Nat. Neurosci.* vol. 5, pp. 805–811, July 2002.

[22] H. Scherberger, M. R. Jarvis, and R. A. Andersen, "Cortical local field potential encodes movement intentions in the posterior parietal cortex", *Neuron* vol. 46, pp. 347–354, Apr. 2005.

[23] S. I. Tillery, and D. M. Taylor, "Signal acquisition and analysis for cortical control of neuroprosthetics", *Curr. Opin. Neurobiol*. vol. 14, pp. 758– 762, Dec. 2004.

[24] S. Musallam, B. D. Corneil, B. Greger, H. Scherberger, R. A. Andersen, "Cognitive control signals for neural prosthetics", *Science* vol. 305, pp. 258–262, July 2004.

[25] P. G. Patil, J. M. Carmena, M. A. Nicolelis, and D. A. Turner, "Ensemble recordings of human subcortical neurons as a source of motor control signals for a brain–machine interface", *Neurosurgery* vol. 55, pp. 27–35, July 2004.

[26] E. V. Evarts, "Pyramidal tract activity associated with a conditioned hand movement in the monkey". *J. Neurophysiol*. vol. 29, pp. 1011–1027, Mar. 1966.

[27] E. V. Evarts, "Relation of pyramidal tract activity to force exerted during voluntary movement", *J. Neurophysiol*. vol. 31, pp. 14–27, Jan. 1968.

[28] E. V. Evarts, "A technique for recording activity of subcortical neurons in moving animals", *Electroencephalogr. Clin. Neurophysiol*. vol. 24, pp. 83–86, Jan. 1968.

[29] J. M. Carmena, M. A. Lebedev, C. S. Henriquez, and M. A. Nicolelis, "Stable ensemble performance with singleneuron variability during reaching movements in primates", *J. Neurosci*. vol. 25, pp. 10712–10716, Nov. 2005.

[30] J. Wessberg, and M. A. Nicolelis, "Optimizing a linear algorithm for real-time robotic control using chronic cortical ensemble recordings in monkeys", *J. Cogn. Neurosci*. vol. 16, pp. 1022–1035, July 2004.

[31] D. Cohen, and M. A. Nicolelis, "Reduction of single-neuron firing uncertainty by cortical ensembles during motor skill learning", *J. Neurosci*. vol. 24, pp. 3574– 3582, Apr. 2004.

[32] M. A. Nicolelis, "Brain–machine interfaces to restore motor function and probe neural circuits", *Nat. Rev. Neurosci*. vol. 4, pp. 417–422, May. 2003.

[33] J. C. Sanchez, J. M. Carmena, M. A. Lebedev, M. A. Nicolelis, J. G. Harris, and J. C. Principe, "Ascertaining the importance of neurons to develop better brain–machine interfaces", *IEEE Trans. Biomed. Eng*. vol. 51, pp. 943–953, May 2004.

[34] R. F. Dodson, L. W. Chu, and N. Ishihara, "Cerebral tissue response to electrode implantation", *Can. J. Neurol. Sci*. vol. 5, pp. 443–446, Nov. 1978.

[35] R. L. Schultz and T. J. Willey, "The ultrastructure of the sheath around chronically implanted electrodes in brain", *J. Neurocytol*. vol. 5, pp. 621–642, Dec. 1976.

[36] V. S. Polikov, P. A. Tresco, and W. M. Reichert, "Response of brain tissue to chronically implanted neural electrodes", *J. Neurosci*. *Methods* vol. 148, pp. 1–18, Aug. 2005.

[37] J. K. Chapin, K. A. Moxon, R. S. Markowitz, and M. A. Nicolelis. "Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex", *Nat. Neurosci*. vol. 2, pp. 664–670, Jun. 1999.

[38] N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, B. Kotchoubey, A. Kubler, J. Perelmouter, E. Taub, and H. Flor, "A spelling device for the paralysed", *Nature* vol. 398, pp. 297–298, 1999.

[39] A. Kubler, B. Kotchoubey, J. Kaiser, J. R. Wolpaw, and N. Birbaumer, "Brain–computer communication: unlocking the locked in", *Psychol. Bull*. vol. 127, pp. 358–375, May 2001.

[40] A. Kubler, N. Diplpsych, J. Kaiser, B. Kotchoubey, T. Hinterberger, and N. Birbaumer, "Brain–computer communication: selfregulation of slow cortical potentials for verbal communication", *Arch. Phys. Med. Rehabil*. vol. 82, pp. 1533–1539, Nov. 2001.

[41] B. Obermaier, G. Muller, and G. Pfurtscheller, "Virtual keyboard controlled by spontaneous EEG activity", *IEEE Trans. Neural Syst. Rehabil.* Eng. vol. 11, pp. 422–426, Dec. 2003.

[42] B. Obermaier, C. Neuper, C. Guger, and G. Pfurtscheller, "Information transfer rate in a five-classes brain–computer interface", *IEEE Trans. Neural Syst. Rehabil*. Eng. vol. 9, pp. 283–288, Sep. 2001.

[43] H. Sheikh, D. J. McFarland, W. A. Sarnacki, and J. R. Wolpaw, "Electroencephalographic (EEG)-based communication: EEG control versus system performance in humans", *Neurosci. Lett.* vol. 345, pp. 89–92, July 2003.

[44] J. R. Wolpaw, "Brain-computer interfaces (BCIs) for communication and control: a mini-review", *Suppl. Clin. Neurophysiol*. vol. 57, pp. 607–613, Apr. 2004.

[45] N. Birbaumer, "Brain–computer-interface research: coming of age", *Clin. Neurophysiol*. vol. 117, pp. 479–483, Mar. 2006.

[46] D. P. Nowlis and J. Kamiya, "The control of electroencephalographic alpha rhythms through auditory feedback and the associated mental activity", *Psychophysiology* vol. 6, pp. 476–484, Jan. 1970.

[47] G. Pfurtscheller, D. Flotzinger, and J. Kalcher, "Brain-Computer Interface -- a new communication device for handicapped persons", *Journal of Microcomputer Applications,* vol. 16, pp. 293-299, July. 1993.

[48] B. Rebsamen, E. Burdet, C. T. Guan, H. H. Zhang, C. L. Teo, Q. Zeng, M. Ang, and C. Laugier, "A Brain-Controlled Wheelchair Based on P300 and Path Guidance", In *Proceedings of IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics.* Feb. 2006.

[49] E. Niedermeyer, and F. Lopes Da Silva, "Electroencephalography: Basic Principles, Clinical Applications, and Related Fields", *Lippincot Williams & Wilkins*. 2004.

[50] G. Guger, G. Edlinger, W. Harkam, I. Niedermayer, and G. Pfurtscheller, "How many people are able to operate an EEG-Based Brain-Computer Interface (BCI)?", *IEEE Transaction on Neural System and Rehabilitation Engineering*, vol. 11, no.2, pp. 145-147, July 2003.

[51] J. Kalcher, D. Flotzinger, Ch. Neuper, S. Gölly, and G. Pfurtscheller, "Graz brain-computer interface II: towards communication between humans and computers based on online classification of three different EEG patterns", *Med. & Biol. Eng. & Comput.*, vol. 34, pp. 382—388, 1996.

[52] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and C. A. Forneris, "An EEG-based brain-computer interface for cursor control", *Electroencephalography and Clinical Neurophysiology*, vol. 78, pp. 252-259, Mar. 1991.

[53] D. J. McFarland, G. W. Neat, R. F. Read, and J. R. Wolpaw, "An EEG-based method for graded cursor control", *Psychobiology* vol. 21, pp. 77-81, 1993.

[54] T. W. Picton, S. Bentin, P. Berg, E. Donchin, S. A. Hillyard, R. Johnson, JR., G.A. Miller, W. Ritter, D. S. Ruchkin, M. D. Rugg, and M. J. Taylor, "Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria", *Psychophysiology*, vol. 37, pp. 127-152, Cambridge University Press. Mar. 2000.

[55] N. K. Squires, K. C. Squires, and S. A. Hillyard, "Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man", *Electroencephalogr Clin Neurophysiol*. vol. 38, pp. 387-401. Apr. 1975.

[56] R. M. Chapman, and H. R. Bragdon, "Evoked responses to numerical and non-numerical visual stimuli while problem solving", *Nature* vol. 203, pp. 1155-1157. Sep. 1964.

[57] L. A. Farwell, and S. S. Smith, "Using brain MERMER testing to detect knowledge despite efforts to conceal", *J Forensic Sci.* vol. 46, pp. 135-143 Jan 2001.

[58] BCI2000 User Reference: P3SpellerTask, http://www.bci2000.org/wiki/index.php/User_Reference:P3SpellerTask

[59] C. W. Anderson, E. A. Stolz, and S. Shamsunder, "Multivariate autoregressive models for classification of spontaneous electroencephalogram during mental tasks", *IEEE Trans. Biomed. Eng.*, vol. 45, pp. 277-286, Aug. 1998.

[60] S. Devulapalli, "Nonlinear principal component analysis and classification of EEG during mental tasks", *M.S. Thesis*, Department of Computer Science, Colorado State Univ., Fort Collins, CO, 1996.

[61] V. J. Samar, A. Bopardikar, R. Rao, and K. Swartz, "Wavelet analysis of neuroelectric waveforms: A concept tutorial", *Brain. Lang.*, vol. 66, pp. 7-60, Jan. 1999.

[62] V. Bostanov, and B. Kotchoubey, "Recognition of affective prosody: Continuous wavelet measures of event-related brain potentials to emotional exclamations", *Psychophysiol.*, vol. 41, no. 2, p. 259, Mar. 2004.

[63] T. Demiralp, J. Yordanova, V. Kolev, A. Ademoglu, M. Devrim, and V. J. Samar, "Time-frequency analysis of single-sweep event-related potentials by means of fast wavelet transform", *Brain. Lang.*, vol. 66, pp. 129-145, Jan. 1999.

[64] J. Raz, L. Dickerson, and B. Turetsky, "A wavelet packet model of evoked potentials", *Brain. Lang.*, vol. 66, pp. 61-88, Jan. 1999.

[65] H. Heinrich, H. Dickhaus, A. Tothenberger, V. Heinrich, and G. H. Moll, "Single-sweep analysis of event-related potentials by wavelet networks-Methodological basis and clinical application", *IEEE Trans. Biomed. Eng.*, vol. 46, pp. 867-879, July. 1999.

[66] M. Ende, A. K. Louis, P. Maass, and G. Mayer-Kress, "EEG signal analysis by continuous wavelet transform techniques", in *Nonlinear Analysis of Physiological Data*, H. Kantz, J. Kurths, and G. Mayer-Kress, Eds, Berlin, Germany: Springer, pp. 213-219, 1998.

[67] M. Kaper, P. Meinicke, U. Grossekathoefer, T. Lingner, and H. Ritter, "BCI Competition 2003—Data Set IIb: Support Vector Machines for the P300 Speller Paradigm", *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, May. 2004.

[68] J. Qin, Y. Li, and W. Sun, "A semisupervised support vector machines algorithm for BCI systems", *Computational Intelligence and Neuroscience*, vol. 2007, article ID 94397, 9 pages, July. 2007.

[69] G. G. Molina, "BCI adaptation using incremental-SVM learning", *Proceedings of the 3rd international IEEE EMBS Conference on Neural Engineering*, Kohala Coast, Hawaii, USA, May 2-5, 2007.

[70] T. Lal, M. Schroeder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, and B. Schoelkopf, "Support vector channel selection in BCI," *IEEE Transactions on Biomedical Engineering*, vol. 51, no. 6, pp. 1003–1010, May 2004.

[71] A. Rakotomamonjy, V. Guigue, G. Mallet, and V. Alvarado, "Ensemble of SVMs for improving brain-computer interface p300 speller performances," in *15th International Conference on Artificial Neural Networks*, Nov. 2005.

[72] T. P. Jung, S. Makeig, C. Humphries, T. W. Lee, M. J. McKeown, V. Iragui, and T. J. Sejnowski, "Removing electroencephalographic artifacts by blind source separation", *Psychophysiol.*, vol.37, pp. 163-178, Mar. 2000.

[73] A. K. Barros, A. Mansour, and N. Ohnishi, "Removing artifacts from ECG signals using independent components analysis", *Neurocomputing*, vol. 22, pp. 173-186, Nov. 1998.

[74] J. F. Cardoso, "Multidimensional independent component analysis", in *Proc. IEEE int. Conf. Acoustics, Speech and Signal Processing*, vol. 4, pp. 1941-1944, May 1998.

[75] S. Makeig, T. P. Jung, A. J. Bell, and T. J. Sejnowski, "Blind separation of auditory event-related brain responses into independent components", *Proc. Nat. Acad. Sci.*, vol. 94, pp. 10979-10984, Apr. 1997.

[76] S. Makeig, M. Westerfield, T. P. Jung, J. Covington, J. Townsend, T. J. Sejnowski, and E. Courchesne, "Functionally independent components of the late positive event-related potential during visual spatial attention," *J. Neurosci.*, vol. 19, no. 7, pp. 2665–2680, Apr. 1999.

[77] R. Vigário, J. Särelä, V. Jousmäki, M. Hämäläinen, and E. Oja, "Independent component approach to the analysis of EEG and MEG recordings," *IEEE Trans. Biomed. Eng.*, vol. 47, pp. 589–593, May 2000.

[78] B. Hong, F. S. Yang, Y. F. Pan, Q. Y. Tang, K. Chen, and Y. M. Tie, "Single-trial estimation and analysis of PVEP based on independent component analysis," *Tsinghua Sci. Technol.*, vol. 6, no. 5, pp. 503–508, Dec. 2001.

[79] R. N. Vigario, "From principal to independent component analysis of brain signals," in *Proc. 23rd Annu. EMBS Int. Conf.*, Istanbul, Turkey, Oct. 25–28, pp. 1970–1973, Oct. 2001.

[80] H. Serby, E. Yom-Tov, and G. F. Inbar, "An Improved P300-Based Brain-Computer Interface", *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol.13, no.1, Mar. 2005.

[81] D. Whitley, R. Beveridge, C. Guerra, and C. Graves, "Messy genetic algorithms for subset feature selection", in *Proc. Int. Conf. on Genetic Algorithms*, T. Baeck, Ed.. Boston, MA, pp. 568-575, July 1997.

[82] J. Yang, and V. Honavar, "Feature subset selection using a genetic algorithm", in *Feature Extraction, Construction and Subset Selection: A Data Mining Perspective*, H. Liu and H. Motoda, Eds. Boston, MA: Kluwer Academic, pp.117-136, Mar. 1998.

[83] ALS Association: About ALS: Facts You Should Know, http://www.alsa.org/als/facts.cfm?CFID=6758224&CFTOKEN=5c763e966c7a843 7-C40528F0-188B-2E62-80AAAAB898699547

[84] K. Li, R. Sankar, Y. Arbel, and E. Donchin, "P300 Based Single Trial Independent Component Analysis on EEG Signal", *Foundations of Augmented Cognition. Neuroergonomics and Operational Neuroscience, Lecture Notes in Computer Science*, vol. 5638/2009, pp. 404-410, 2009.

[85] K. Li, R. Sankar, Y. Arbel, and E. Donchin, "Single trial independent component analysis for P300 BCI system", *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4035-4038. Sep. 2009.

[86] K. Li, R. Sankar, Y. Arbel, and E. Donchin, "Single trial independent component analysis for the P300 BCI system", *Fourth International Meeting*, Asilomar, California May 31 – June. 2010.

[87] K. Li, R. Sankar, Y. Arbel, and E. Donchin, "Blind tracking based single trial independent component analysis for P300 BCI system", accepted with minor changes by *IEEE Transaction on Neural Systems and Rehabilitation Engineering*, 2010.

[88] K. Li, R. Sankar, Y. Arbel, and E. Donchin, "Single trial independent component analysis for the P300 BCI system", submitted to *The Journal of Neural Engineering*, 2010.

[89] K. Li, R. Sankar, Y. Arbel, and E. Donchin, "A new single trial P300 classification method", submitted to *IEEE Transaction on Neural Systems and Rehabilitation Engineering*, 2010.

[90] S. Sutton, M. Braren, J. Zubin, and E. R. John, "Evoked-Potential Correlates of Stimulus Uncertainty", *Science*, vol. 150, pp. 1187-1188. Nov. 1965.

[91] D. J. Krusienski, E. W. Sellers, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "Toward enhanced P300 speller performance", *J Neurosci Methods*. vol. 167, pp. 15–21, Jan. 2008.

[92] E. W. Sellers, A. Kübler, and E. Donchin, "Brain-computer interface research at the University of South Florida Cognitive Psychophysiology Laboratory: the P300 Speller", *IEEE Trans Neural Syst Rehabil Eng*. vol. 14, pp. 221-224, July 2006.

[93] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition", *Data Mining and Knowledge Discovery*, vol. 2, pp. 121-167, Jun. 1998.

[94] J. Nocedal, and S. J. Wright, *Numerical Optimization (2nd ed.)*, Berlin, New York: Springer-Verlag. 2006.

[95] K. G. Murty, "Linear complementarity, linear and nonlinear programming", *Sigma Series in Applied Mathematics*. 3. Berlin: Heldermann Verlag. 1988.

[96] C. Cortes, and V. Vapnik, "Support-Vector Networks", *Machine Learning*, vol. 20, Sep. 1995.

[97] B. D. Ripley, *Pattern Recognition and Neural Networks*, Cambridge, U.K.: Cambridge University Press, 1996.

[98] P. Meinicke, T. Twellmanna, and H. Ritter, "Discriminative densities form maximum contrast estimation", in *Advances in Neural Information Processing Systems 15*. Cambridge, MA: MIT Press, 2003.

[99] P. M. Woodward, *Probability and Information Theory with Applications to Radar*, Norwood, MA: Artech House, 1980.

[100] A. Haar, "Zur Theorie der orthogonalen Funktionensysteme", *Mathematische Annalen*, vol. 69, pp. 331–371, 1910.

[101] http://scienceworld.wolfram.com/biography/Zweig.html, Zweig, George's Biography on Scienceworld.wolfram.com

[102] http://www.mssu.edu/seg-vm/bio_jean_p__morlet.html, Morlet, Jean's Biography on www.mssu.edu

[103] A. Grossmann, and J. Morlet, "Decomposition of Hardy functions into square integrable wavelets of constant shape", *SIAM J. Math. Anal.*, vol. 15, pp. 723--736, July 1984.

[104] Y. Meyer, *Wavelets and operators*, Cambridge University Press, Cambridge, 1992.

[105] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, iss. 7, pp: 674-693, July 1989.

[106] R. Brinks, "On the convergence of derivatives of B-splines to derivatives of the Gaussian function", *Comp. Appl. Math.*, vol. 27, no.1, Jun. 2008.

[107] F. T. Smulders, J. L. Kenemans, and A. Kok, "A comparison of different methods for estimating single-trial P300 latencies", *Electroencephalogr. Clin. Neurophysiol.*, vol. 92, pp. 107-114, Mar. 1994.

[108] J. H. McDermott, "The cocktail party problem", *Current Biology*, vol. 19, no.22, R1024-R1027, Sep. 2005.

[109] A. Hyvärinen, "Fast ICA by a fixed-point algorithm that maximizes non-Gaussianity", *Independent Component Analysis Principles and Practice*. Cambridge University Press, Mar. 2001.

[110] A. Hyvärinen, and E. Oja, "Independent Component Analysis: Algorithms and Applications", *Neural Networks*, vol. 13, pp. 411—430, May 2000.

[111] A. Hyvärinen, "Fast and Robust Fixed-Point Algorithms for Independent Component Analysis", *IEEE Transactions on Neural Networks*, vol. 10, pp. 626-634, May 1999.

[112] N. Xu, X. Gao, B. Hong, X. Miao, S. Gao, and F. Yang, "BCI Competition 2003-Data Set IIb: Enhancing P300 wave detection using ICA-Based subspace projections for BCI applications", *IEEE Transaction on Biomedical Engineering*, vol. 51, no. 6, May 2004.

[113] D. J. Krusienski, E. W. Sellers, F. Cabestaing, S. Bayoudh, D. J. McFarland, T. M. Vaughan, and J. R. Wolpaw, "A comparison of classification techniques for the P300 speller", *J. Neural Eng.* vol. **3,** pp. 299-305, Dec. 2006.

[114] A. J. Bell, and T. J. Sejnowski, "An information-maximisation approach to blind separation and blind deconvolution", *Neural Computation* vol. 7, pp. 1129—1159, Feb. 1995.

[115] J. F. Cardoso, and P. Comon, "Equivariant adaptive source separation", *IEEE Transactions on Signal Processing*, vol. 45, No. 2, pp. 434-444, Aug. 1996.

[116] E. Donchin, "Data analysis techniques in average evoked potential research", In E. Donchin & D. B. Lindsley (Eds.), *Average evoked potentials: Methods, results and evaluations* (NASA SP-191, pp. 199-236). Washington, DC: U.S. Government Printing Office. 1969.

[117] O. I. Khan, S. H. Kim, T. Rasheed, A. Khan, and T. S. Kim, "Extraction of P300 using constrained independent component analysis", *Conf Proc IEEE Eng Med Biol Soc*. 2009, pp. 4031-4034. Dec. 2009.

[118] W. D. Penny, S. J. Roberts, E. A. Curran, and M. J. Stokes, "EEG-based communication: A pattern recognition approach", *IEEE Trans. Rehab. Eng*., vol. 8, pp. 214—215, Jun. 2000.

ABOUT THE AUTHOR

Kun Li received a Bachelor Degree in Biological Science from Wuhan University, China in 1999. He worked as a bio-engineer in Shanghai Health-Digit Bio-tech Ltd. in the area of molecular biology and microarray development and analysis for years. After that, he came to U.S. and worked in H. Lee Moffitt Cancer Center as a research associate in the area of microarray analysis, pattern recognition and cancer prediction. Kun Li received a Master of Science in Electrical Engineering from University of South Florida in 2009. His research interests lie in the fields of signal processing, image processing, brain computer interface application, statistical analysis, and microarray analysis.