國立臺灣大學工學院機械工程學研究所

碩士論文

Department of Mechanical Engineering
College of Engineering
National Taiwan University

Master Thesis

基於獨立成分分析自動去除腦波內眨眼干擾波及其應用

Automatic Removal of Eye-blink Artifacts in EEG Based on ICA and Its Applications

王璟萍

Ching-Ping Wang

指導教授：黃漢邦 博士

Advisor: Han-Pang Huang, Ph.D.

中華民國 101 年 7 月

July, 2012

摘要

腦波訊號在量測時往往伴隨許多干擾波混雜其中，例如眼電、肌電、心電訊號等。這將嚴重影響腦機介面的性能及臨床診斷的準確率，而這些干擾波中，又以眼電訊號對腦波的影響最為嚴重，而獨立成分分析已經被證明能夠有效的將腦波及干擾波分離。

本篇論文提出一個自動去除眼電訊號的方法，透過獨立成分分析將腦波及眼動分離，接著以樣本熵、碎型維度、峰度作為特徵抽取。另一方面，本研究以實際的眼動訊號做單類別分類器的訓練，並利用此訓練後的單類別分類器作為眼動成分的自動選擇器。在移除被選出的眼動成分後，將剩下的腦波成分以獨立成分分析所得到的去混合矩陣，重建去眼動的腦波訊號。根據實驗的結果，證明此方法能夠在較少的資訊損失下獲得去除眼動的腦波訊號。

接著以兩個實際應用驗證其性能。第一個應用為 P300 拼字機，透過本研究提出的方法將可提升拼字成功率 3~9%，證實此方法能有效去除眼動訊號而保留 P300 訊號。第二為睡眠腦波上的應用，使用台中澄清醫院睡眠中心資料庫做睡眠腦波的分析，以獨立成分分析前處理完畢後的腦波訊號進行睡眠階段的分類，可提升睡眠階段的 R1 分類率平均 1.56%，提升整體睡眠階段分類率 0.44%。

關鍵字：腦電圖(Electroencephalograph)、獨立成分分析(Independent Component Analysis)、眨眼干擾波(eye-blink artifact)、單類別分類器(One-Class Classification)

Abstract

Electroencephalogram (EEG) recordings are often contaminated by, for example, ocular artifacts, muscle artifacts, heart signals, and line noise. The influence of such contaminated EEG signals on the performance of a brain–computer interface and on clinical examination is serious. Among the types of noise, the influence of eyeblink artifacts is the most adverse. However, Independent Component Analysis (ICA) has been proven as an effective tool to separate artifacts from EEG signals.

In the thesis, an approach is presented to remove eyeblink artifacts automatically. Artifacts from brain waves signals are separated by ICA, based on features of with sample entropy, fractal dimension, and kurtosis. This study sets actual eyeblinks as the training data for one-class classification, and the trained one-class classifier is used as an automatic selector for eye blink artifacts. After removing the eye-blink artifacts, the remaining brain waves use the demixing matrixes from ICA to reconstruct eyeblink-removed brain waves. According to the results of experiments, the proposed method confirms that eyeblink artifacts can be removed with less loss of information.

Two applications were implemented to verify the performance of the proposed approach. One application is the P300 speller. Through the approach of this research,

the classification rate is improved by 3~9%, thus confirming that this method can remove eyeblink artifacts and retain P300 signals. The second application used to verify the results is one done on sleeping brain waves: the analysis was done with the database of the sleep center of Taichung Cheng Ching Hospital. Stages of sleep were classified by pre-procedure brain waves. Through the method, the R1 classification rate of stages of sleep is increased by 1.56%, and the classification rate of the full stage is increased by 0.44%.

Keywords: Electroencephalograph, Independent Component Analysis, eye-blink artifact, One-Class Classification

Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Introduction to EEG and the Brain–Computer Interface

1.1.1 Introduction to Electroencephalography

In recent years, many research teams have been working on brain wave applications. Brain waves can be recorded by electroencephalography (EEG), magnetoencephalography (MEG), functional Magnetic Resonance Imaging (fMRI), and Positron Emission Tomography (PET). The first human EEG was recorded in 1924 by German physiologist Hans Berger[1], and now EEG research has become one of the most popular methods in brain research activities because the EEG signal devices are cost effective.

EEG, a non-invasive method to record the brain's potential activity in the human brain, is measured through adhesive electrodes on the scalp. After brain waves cross over the skull, EEG signals are measured in micro volts.

EEG electrode placement is usually based on the International10-20 system[2], as shown in Figure 1-1. Each site on the scalp has a letter and a number or another letter. The letter is used to identify the lobe: F refers to the frontal lobe, T to the temporal lobe,

C to the central lobe, P to the parietal lobe, and O to the occipital lobe. The number is used to identify whether the electrode is placed on the right or left hemisphere. Even numbers refer to the right hemisphere, and odd numbers to the left. The letter Z refers to the middle line.

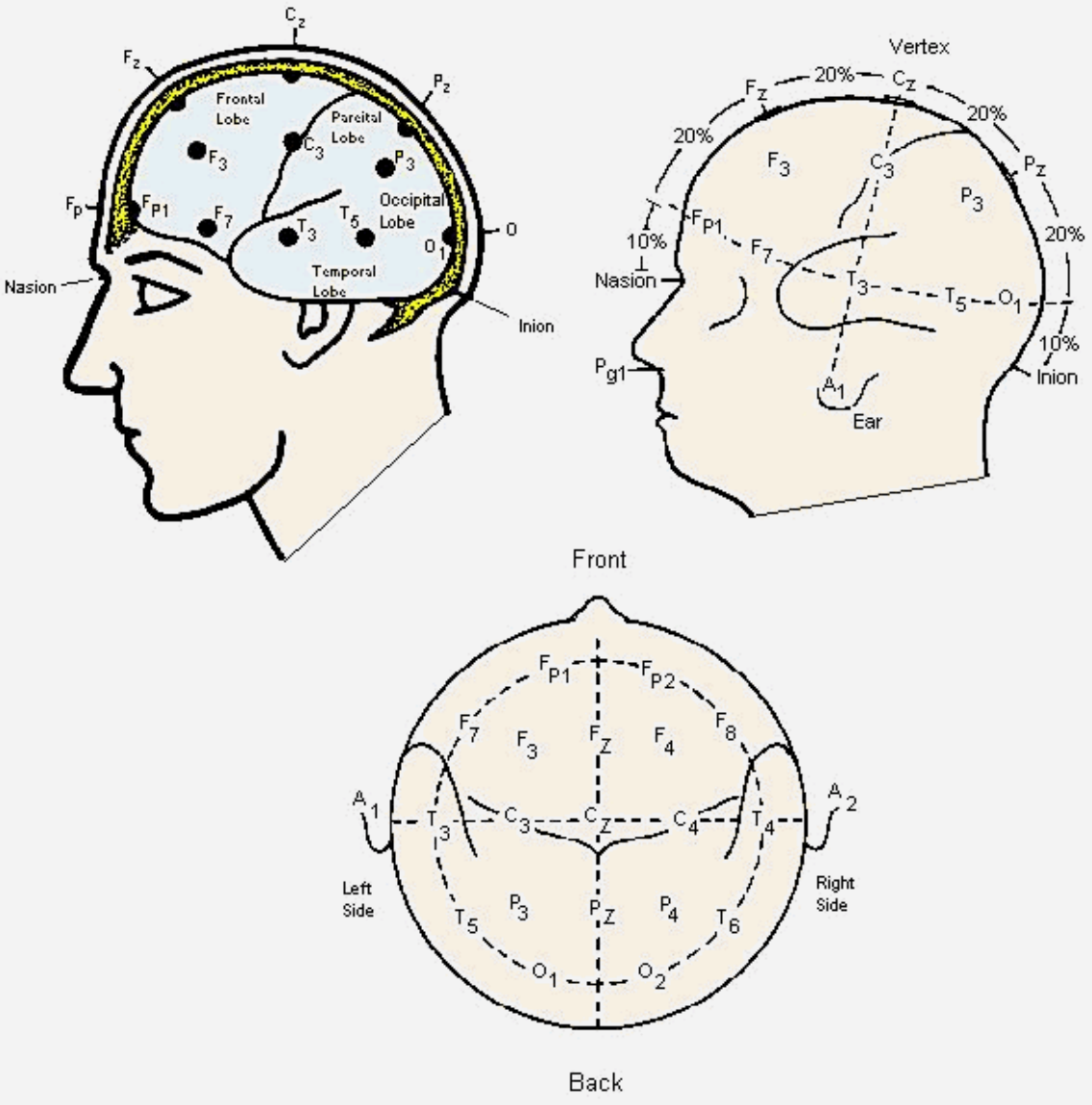


Figure 1-1 Electrode positions in the 10-20 system

Brain waves can be divided into four categories based on their frequency range(also known as background activity): δ、θ、α、β. These categories are shown in Table 1-1.

Table 1-1 Frequency band of background activity

Rhythm	Frequency band
δ rhyth m	< 4 Hz
θ rhyth m	4 ~ 8 Hz
α rhyth m	8 ~ 13 Hz
β rhyth m	>13 Hz

1.1.2 Introduction to the Brain–Computer Interface

One of the main applications of EEG is the brain-computer interface (BCI), a novel and useful way for users to communicate with their environment [3, 4]. Such systems can provide a method of communication for neurologically impaired patients, for example, those with amyotrophic lateral sclerosis or brain cord injury. Figure 1-2 shows the basic framework of the BCI system: signals from the brain are recorded by electrodes on the scalp and processed to extract specific signal features that reflect the user's intent. These features are translated into commands that operate a device.

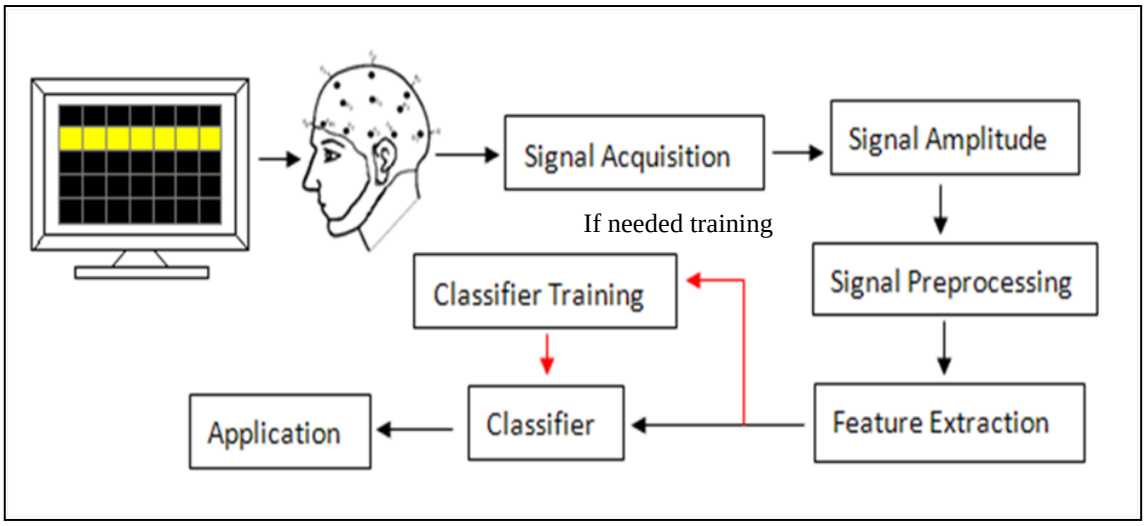


Figure 1-2 Framework of the brain–computer interface

1.2 Artifacts in EEG Recordings

During the period of the EEG recording, EEG signals are contaminated by, ocular artifacts, muscle artifacts, heart signals, and line noise. EEG signals containing artifacts often significantly affect the performance of the brain–computer interface.

1.2.1 Ocular Artifacts

Ocular artifacts are divided into two types: eye-blink and eye-movement. Eye-blink artifacts result from eyes opening and closing. The frequency range of ocular artifacts is about 0.03 to 20 Hz.

Negative potential activity occurs when the eyes close, and positive potential activity occurs when the eyes open, which is shown in Figure 1-3.

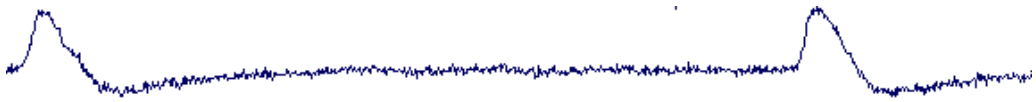


Figure 1-3 Eye-blink artifacts

In general, the closer the forehead, the more eye-blink artifacts mixed with brainwaves. The frequency band of eye-blink artifacts overlaps with the P300 potential and α activity. Therefore, having a procedure for removing eye-blink artifacts without losing brain waves is important.

Eye movement also causes ocular artifacts, which occur at a relatively low frequency, as shown in Figure 1-4.

Figure 1-4 Eye movement artifacts

1.2.2 Muscle Artifacts

Muscle artifacts result from activity in the jaw or head muscles and may occur at any location on the head. If the subject moves his or her body or strains during an EEG recording, muscle artifacts will be mixed within the EEG (Figure 1-5). The frequency range of muscle artifacts is about 50 to 5000 Hz.

Figure 1-5 Muscle artifacts

1.2.3 Heart Artifacts

Heart artifacts result from heart beats or mastoid reference channels, because the mastoids are usually used for reference electrodes. However, since there are arteries near the mastoid, heart artifacts may be mixed with the EEG. The frequency range of heart artifacts is about 0.03 to 10 Hz.

1.2.4 Other Artifacts

In addition to the physiological signals, white noise, shown in Figure 1-6, may be produced through a loosening electrode or the conductive adhesive. Poor grounding also may cause 60Hzof power line noise.

Figure 1-6 White noise

1.3 Objectives and Motivation

EEG has the advantages of being non-invasive and inexpensive. However, acquiring a clean EEG signal is almost impossible. Therefore, the removal of artifacts is crucial.

The most common method is measuring electrooculography (EOG) and EEG simultaneously and then removing the EOG from the EEG by linear regression. This method requires pasting four electrodes around the subject's eyes, which can make the subject uncomfortable. Another problem is that if the measured EOG contains brain waves, the linear regression could lose the EEG data.

Independent component analysis (ICA) has proven to be an effective tool to separate artifacts from EEG signals, whether the artifacts are eye movements, ECG, EMG, or power line noise[5-8]. After ICA separates artifacts from the EEG, artifact components can be removed to obtain clean EEG signals. However, choosing the component after ICA decomposition is still difficult. In order to reconstruct a corrected EEG signal, the independent artifact component that needs to be rejected must first be identified. At present, most studies still choose it manually, a process which is time consuming and may obstruct real-time measurement. Therefore, an automatic selection method of the independent component (IC) is necessary.

1.4 Related Works

Some studies on automatic artifact removal have been proposed over the last few years. Woestenburg et al. proposed a regression method that was applied with an EOG reference [9]. Boudet et al. suggested a global artifact removal method based on ICA and frequency bands [10]. Schlogl et al. presented an automatic method for reducing EOG artifacts based on regression analysis [11]. Patidarand Zouridakis proposed a hybrid approach for removing semi-simulated artifacts based on iterative ICA and fuzzy clustering [12]. Edmonds et al. presented a semi-automatic tool designed for clustering

independent components based on the correlation of ICA inverse weights and found independent components that are similar to a user-defined template[13]. Ong et al. suggested an automatic artifact removal method based on a weighted support vector machine[14]. Cantero et al. focused on the performance of four independent component analysis (ICA)algorithms (AMUSE, SOBI, Infomax, and JADE) to separate muscle activity from EEG during sleep[15]. Dan-hua et al. introduced an automatic procedure based on ICA and sample entropy in order to identify the components related to eyeblink artifacts [16]. Lu et al. proposed an automatic method to choose the eyeblink artifact components based on the pattern of scalp topographies[17]. Castellanos and Makarov presented an approach to identify eyeblink and ECG artifacts based on wavelet-enhanced ICA that applies wavelet thresholding [18]. Mammone introduced an automatic artifact approach based on Wavelet Transform and ICA [19]. Guerrero-Mosquera proposed a method based on adaptive filtering and ICA[20]. However, no standard automatic IC selection procedure has come out so far.

1.5 Thesis Organization

This thesis is organized into six chapters. Chapter 1 introduces brain waves, identifies related studies, and presents the objectives and motivation of the study. Chapter

2 then introduces independent component analysis and one-class classifiers. The approaches of feature extraction are also introduced here. Next, Chapter 3 presents the approach framework and the experiment design, in addition to presenting the artifact removal methods. Chapter 4 presents the results of the application of the P300 Speller, while Chapter 5 presents the results of the application of sleep stage classification. Finally, Chapter 6 offers conclusions and proposes future works.

Chapter 2 Background Materials

2.1 Independent Component Analysis

Independent component analysis (ICA) is a higher order statistics method for finding the independent components with maximum non-Gaussian distributions. It is useful for the blind source separation (BSS) problem. It assumes the observed signals, $x(t) = \{x_i(t), i = 1,..., M\}$ to be a linear combination of statistically independent sources $s(t) = \{s_i(t), i = 1,..., M\}$. ICA further supposes that the sources are unknown and the number of sources is equal to the number of observers.

The basic ICA model is as follows:

$$\mathbf{x} = \mathbf{As} \tag{2-1}$$

Define

$$\mathbf{W} = \mathbf{A}^{-1} \tag{2-2}$$

ICA tries to estimate a demixing matrix **W**. Through the demixing matrix, we can find the independent sources as

$$\mathbf{s} = \mathbf{Wx} \tag{2-3}$$

ICA is used to identify the signals of the hidden components. These independent components (IC) can be separated by ICA [21-24]. The scalp topographies can also be obtained by ICA. It is a good feature for classification of the artifacts. There are many algorithms available for the ICA. In this thesis, the extended infomax ICA algorithm[25] implemented by EEGLAB is used .

2.1.1 The extended Infomax ICA algorithm

The aim of the infomax ICA algorithm is to make components have a maximum likelihood for each other. The extension of the infomax algorithm tries to separate the mixed sub-Gaussian and super-Gaussian source distributions.

The algorithm is described as below:

Step 1: Center the data to make its mean zero.

$$\mathbf{x}' = \mathbf{x} - \mathrm{E}\{\mathbf{x}\} \tag{2-4}$$

Step 2: Whiten the data to give **z**.

Linearly transform the observed data vector $\mathbf{x}'$ by linearly multiplying it with some matrix **V**.

$$\mathbf{z} = \mathbf{V}\mathbf{x}' \tag{2-5}$$

Using the eigenvalue decomposition (EVD) of the covariance matrix

$$\mathrm{E}\{\mathbf{z}\mathbf{z}^{\mathrm{T}}\} = \mathbf{E}\mathbf{D}\mathbf{E}^{\mathrm{T}} \tag{2-6}$$

D and E is the eigenvalue and eigenvector matrix of covariance matrix of $\mathbf{x}'$.

$$\mathbf{V} = \mathbf{E}\mathbf{D}^{-1/2}\mathbf{E}^{\mathrm{T}} \tag{2-7}$$

Step 3: Choose an initial (e.g., random) separating matrix **W**

Step 4: Calculate **u** = **Wz**.

Step 5: Update the separating matrix by

$$\Delta\mathbf{W} \propto [\mathrm{I} - \mathrm{k_i}\tanh(\mathbf{u})\,\mathbf{u}^{\mathrm{T}} - \mathbf{u}\mathbf{u}^{\mathrm{T}}] \times \mathbf{W} \tag{2-8}$$

$$\mathrm{k_i} = \mathrm{sign}(\mathrm{E}\{\mathbf{u}^4\} - 3(\mathrm{E}\{\mathbf{u}^2\})^2)\begin{cases}\mathrm{k_i} = 1 & \text{supergaussian}\\ \mathrm{k_i} = -1 & \text{subgaussian}\end{cases} \tag{2-9}$$

Step 6: If it not converged, go back to step 4.

2.1.2 Scalp Topography

Applying ICA to EEG signals, the resulting independent components (IC) will be separated. Every row of the demixing matrix gives the weights of the components in respect to each scalp electrode[26], as shown in Figure 2-1.The rows of the matrix

describe the weight of the independent components in respect to different channels, so every row gives a scalp topography that corresponds to a different component.

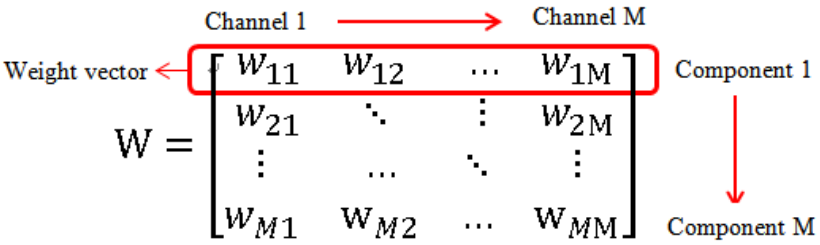


Figure 2-1 ICA demixing matrix

The scalp topographies represent the scale distribution associated with different independent components. From the scalp topographies, we can find which component is an eye-blink component or not. If it is an eye-blink component, the energy should be focused on the front. On the contrary, the energy of the EEG components should be distributed throughout the head. Figure 2-2 shows different components corresponding to scalp topographies.

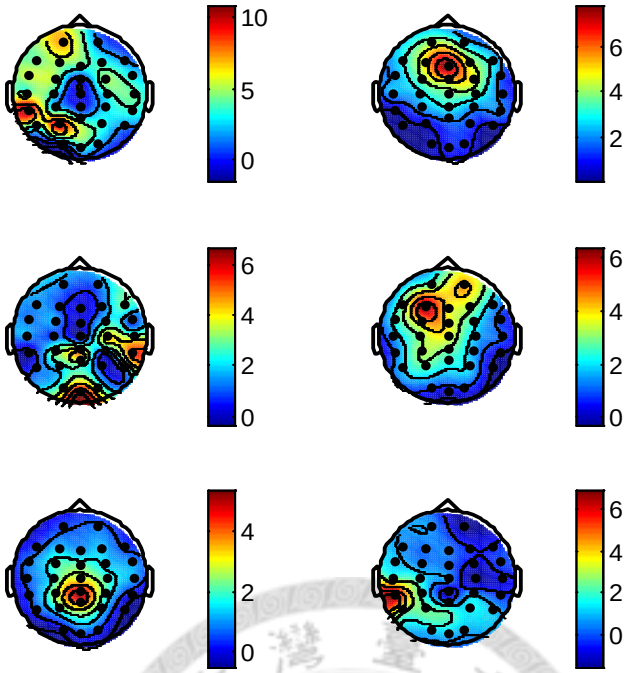


Figure 2-2 Scalp topographies of the independent components

Therefore, the vectors of scalp topographies are available features for IC selection.

In this study, a labeled procedure to identify the eye blink artifact component base on scalp topographies, time domain, and frequency domain, as shown in Figure 2-3, is needed manually and these labels are stored for calculating classification results.

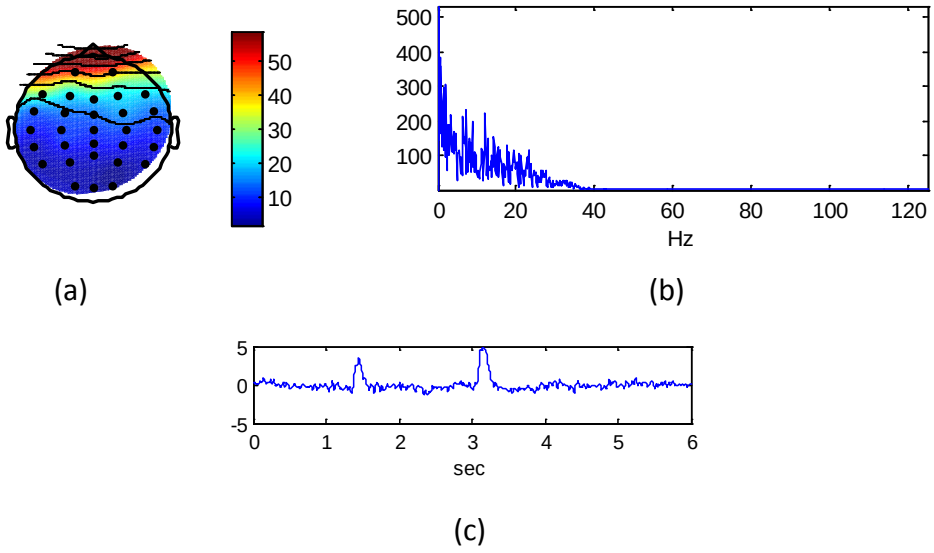


Figure 2-3 The features of an eye blink artifact. (a) Scalp topography; (b) In frequency domain; (c) In time domain

2.2 One-Class Classifiers

2.2.1 Support Vector Data Description (SVDD)

SVDD is a one-class classification tool proposed by Tax et al.[27, 28]. It is derived from the Support Vector Machines (SVM), which comprise a multi-class classifier[29]. The aim of SVDD is to find a hypersphere which has a minimum volume to contain all (or most) targets. The boundary of the hypersphere can be used to reject outliers. Since SVDD only needs a category of data, it is useful when another category of data is difficult to obtain.

Support vector data description (SVDD) is a one-class classifier. Given a data set $\{x_i, i = 1,..., N\}$, SVDD tries to find a hypersphere with the minimal volume to contain most of the data. The sphere is defined by center a and radius R. If a training set contains some outliers, the volume of the sphere will be very large. Therefore, the slack variables ξ_i are introduced and cost function may change to a minimum[30]:

Primal Problem:

$$F(R, \mathbf{a}, \xi_i) = R^2 + C\sum_i \xi_i , \tag{2-10}$$

subject to

$$(\mathbf{x}_i - a)^T (\mathbf{x}_i - a) \le R^2 + \xi_i \ \forall i, \xi_i \ge 0 . \tag{2-11}$$

Dual Problem:

Construct the Lagrangian with Lagrange multipliers α_i to maximize with respect to α_i

$$L = \sum_i \alpha_i (\mathbf{x}_i \cdot \mathbf{x}_i) - \sum_{ij} \alpha_i \alpha_j (\mathbf{x}_i \cdot \mathbf{x}_j) , \tag{2-12}$$

with new constraints

$$\sum_i \alpha_i = 1, a = \frac{\sum_i \alpha_i \mathbf{x}_i}{\sum_i \alpha_i} = \sum_i \alpha_i \mathbf{x}_i, \tag{2-13}$$

$$0 \le \alpha_i \le C, \quad i = 1,...N. \tag{2-14}$$

The radius of the sphere *R* can be obtained by the center a and the support vector on the boundary:

$$R^2 = (\mathbf{x}_k \cdot \mathbf{x}_k) - 2\sum_i \alpha_i (\mathbf{x}_i \cdot \mathbf{x}_k) + \sum_{ij} \alpha_i \alpha_j (\mathbf{x}_i \cdot \mathbf{x}_j) \tag{2-15}$$

To determine whether a test data $\mathbf{z}$ is accepted, we calculate the distance to the center of the sphere. The test object $\mathbf{z}$ is accepted when

$$f(\mathbf{z}) = \|\mathbf{z} - \mathbf{a}\|^2 \le R^2, \tag{2-16}$$

$$\|\mathbf{z} - \mathbf{a}\|^2 = (\mathbf{z} \cdot \mathbf{z}) - 2\sum_i \alpha_i (\mathbf{z} \cdot \mathbf{x}_i) + \sum_{ij} \alpha_i \alpha_j (\mathbf{x}_i \cdot \mathbf{x}_j). \tag{2-17}$$

In this work, the kernel function is

$$K(\mathbf{x}_i \cdot \mathbf{x}_j) = \exp(-\|(\mathbf{x}_i \cdot \mathbf{x}_j)^2 / s^2\|). \tag{2-18}$$

2.2.2 Gaussian Mixing Model (GMM)

Gaussian mixing model is a linear combination of gaussian distributions. It is to estimate the density of the target data. Therefore, it used as a one-class classifier by setting a threshold on this density.

Gaussian density for *d*-dimesional space as below:

$$\emptyset(x;\theta) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} e^{\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} \quad (2\text{-}19)$$

$$\mu = E\{x\} \quad (2\text{-}20)$$

$$\Sigma = E\{(x-\mu)^T(x-\mu)\} \quad (2\text{-}21)$$

So, *k*-component GMM can be written as:

$$f_k(x) = \sum_{j=1}^{k} \pi_j \phi(x;\theta_j) \quad (2\text{-}22)$$

$$\sum_{j=1}^{k} \pi_j = 1 \text{ and } \pi_j \geq 0 : j \in 1, \ldots, k \quad (2\text{-}23)$$

where π_j = mixing weight, $\phi(x;\theta_j)$ = mixture component

In general, parameters can be estimated by EM algorithm and steps are as follows:

Step 1: Numbers of mixtures decided beforehand.

Step 2: Updates the parameters of given *k*-component mixture with respect to data set $X_n = x_1, \dots, x_n$ such that likelihood of X_n is never smaller under newmixtures.

Step 3: Estimates by iterating following equations for all components $j \in 1, \dots, k$:

$$P(j|x_i) = \pi_j \emptyset(x_i; \theta_j)/f_k(x_i) \quad (2\text{-}24)$$

$$\pi_j = \sum_{i=1}^{n} P(j|x_i)/n \quad (2\text{-}25)$$

$$\mu_j = \sum_{i=1}^{n} P(j|x_i)x_i/(n\pi_j) \quad (2\text{-}26)$$

$$\Sigma_j = \sum_{i=1}^{n} P(j|x_i)(x_i - \mu_i)(x_i - \mu_i)^T/(n\pi_j) \quad (2\text{-}27)$$

2.2.3 Kernel Principal Component Analysis (KPCA)

Kernel Principal Component Analysis is a reconstruction method in one-class classification. It uses prior knowledge about target data to model the data. Reconstruction method is assumed that outlier data do not satisfy the model constructed from the target data. The reconstruction error of test data is used as a distance to the target data. And the empirical threshold has to be obtained by using the training data for one-class classification.

Kernel principal component analysis is a deformation of principal component analysis (PCA). The goal of PCA is to find a linear subspace containing the maximum

variance in the data. And KPCA is to find a non-linear subspace containing the maximum variance in the data.

2.3 Features Extraction

2.3.1 Kurtosis

Kurtosis is commonly defined as the fourth cumulant divided by the square of the second cumulant, which is equivalent to the fourth moment around the mean divided by the square of the variance of the probability distribution minus 3[31].

Consider a scalar random variable x with zero-mean, and then the kurtosis is defined as follows:

$$\mathrm{kurt}(\mathrm{x}) = \mathrm{E}\{\mathrm{x}^4\} - 3(\mathrm{E}\{\mathrm{x}^2\})^2 \tag{2-28}$$

2.3.2 Sample Entropy

Assume that the data vector is $x_1, x_2, \dots, x_N$, and then the algorithm of sample entropy can be described as follows:

Step 1: Define the following vector:

$$\boldsymbol{X_i} = [x_i,\ x_{i+1}, \dots, x_{i+m-1}] \quad i = 1 \sim N - m + 1, \tag{2-29}$$

where *m* is the embedded dimension;

Step 2: Define that the distance of $\boldsymbol{X_i}$ and $\boldsymbol{X_j}$ is the largest D-value of the corresponding elements and that is to say:

$$D[\boldsymbol{X_i},\ \boldsymbol{X_j}] = max_{k=0\sim m-1}[|x_{i+k} - x_{j+k}|] \quad (2\text{-}30)$$

Step 3: Given the threshold *r* , calculate the number of $d[\boldsymbol{X_i}, \boldsymbol{X_j}]$ which is less than *r* . Mark the mean of all number as

$$B_i^m(r) = \frac{1}{N-m-1}\{Number\ of\ D[\boldsymbol{X_i},\ \boldsymbol{X_j}] < r\},\ \ i = 1\sim N-m,\ i \neq j. \quad (2\text{-}31)$$

Step 4: Define $B^m(r)$ as the mean of $B_i^m(r)$,

$$B^m(r) = \frac{1}{N-m}\sum_{i=1}^{N-m} B_i^m(r) \quad (2\text{-}32)$$

Step 5: The value of *m* is added to 1. And then calculate $B_{m+1}(r)$.

Sample entropy can be estimated by

$$SampEn(m,r) = -ln(\frac{B^{m+1}(r)}{B^m(r)}), \quad (2\text{-}33)$$

where *m* =2, *r* =(0.1~0.25) * SD, SD is the standard deviation of the data vectors.

2.3.3 Fractal Dimension

An expression, which is in order to calculate the fractal dimension of a waveform, is obtained. The expression starts from the definition of Hausdorff dimension (D_h).

The Hausdorff dimension of a set in a metric space can be expressed as follows:

$$D_h = lim_{\epsilon \to 0} \frac{-ln[N(\epsilon)]}{ln(\epsilon)}, \tag{2-34}$$

where $N(\epsilon)$ is the number of self-similar structures of linear size ϵ needed to cover the whole structure.

A line of length L would be divided into $N(\epsilon) = {}^{L}/_{(2 \cdot \epsilon)}$ segments of length $2 \cdot \epsilon$.

$$D_h = lim_{\epsilon \to 0}[\frac{-ln(L)+ln(2 \cdot \epsilon)}{ln(\epsilon)}] = lim_{\epsilon \to 0}[1 - \frac{ln(L)+ln(2)}{ln(\epsilon)}] = lim_{\epsilon \to 0}[1 - \frac{ln(L)}{ln(\epsilon)}] \tag{2-35}$$

Because the metric space does not change under the linear transformation, it is convenient linearly to transform a waveform into another in a normalized space.

Normalize every point in the abscissa as:

$$x_i^* = \frac{x_i}{x_{max}}, \qquad y_i^* = \frac{y_i - y_{min}}{y_{max} - y_{min}} \tag{2-36}$$

Because of the value of x_{min}, the initial point, must be zero.

Take $\epsilon = {}^{1}/_{2n}$ and then D_h becomes as:

$$D_h \approx D = 1 + \frac{ln(L)}{ln(2 \cdot n)}, \tag{2-37}$$

where N is sample size, n equals to N-1 and L is normalized length.

Chapter 3 Proposed Approaches

3.1 Flowchart

In the testing step, the original EEG is separated into independent components by ICA. The same features extraction in the training step is used for each independent component. The artifact components are then removed by the IC selector. After finishing the IC selection, we can reconstruct a corrected EEG by inverse ICA.

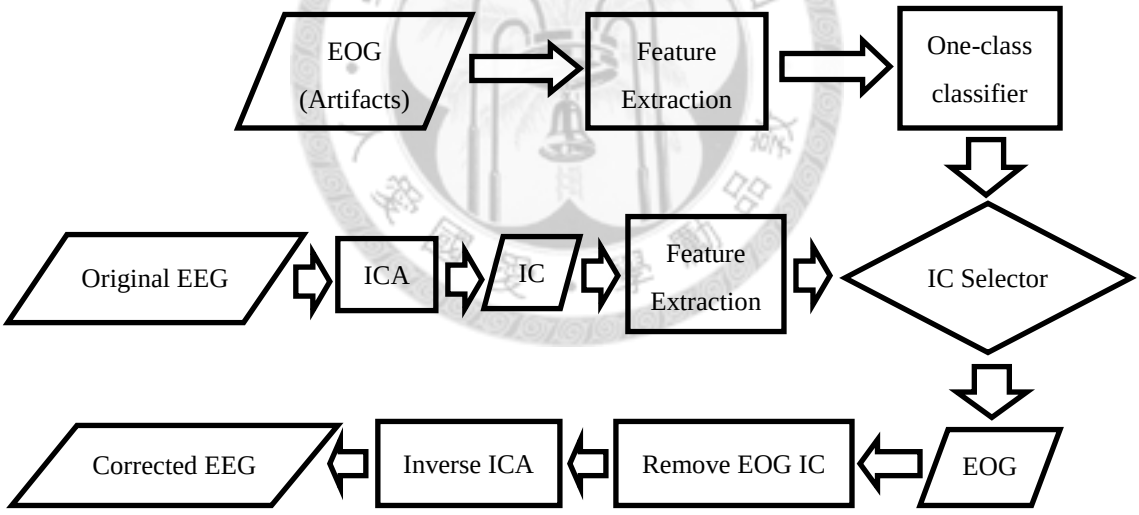


Figure 3-1 Flowchart

On the other hand, a labeled procedure is needed for verifying results in advance. This procedure is to identify which component is an eye-blink artifact. We perform this

step manually by time-domain signals, frequency domain signals, and scalp topographies.

3.2 Data Acquisition

The brain waves were recorded from 15 electrodes mounted on a 64-channel cap. The position of electrode sites was according to the international 10-20 system. The raw signals were amplified and filtered by a NeuroScan system coupled with a SynAmps2. During the experiment, impedances were kept at no more than 5000 ohms. The sampling rate was 250 Hz and was filtered with a cutoff frequency of 0.5–40 Hz.

Two subjects were involved in this experiment. In the training step, the EOG was recorded by additional electrodes for the training classifier. During the EOG recording, the subjects were asked to blink naturally. Because the frequency of an eye-blink is approximately once every 3–5seconds, each epoch length was set at 8 seconds ($8 \times 250 = 2000$ points).Each subject made 100 epochs each in the training step and the test step. All epochs were used for analysis and the basis of each epoch was shifted to zero. In the test step, the EEG was recorded for testing the classifier. No task was being executed by the subject when the EEG signals were recorded.

The training data was used to find the IC selector for the automatic removal of eye-blink artifacts. The test data was used to measure the efficiency of the classifier.

3.3 Independent Component Analysis

Before the feature extraction, the EEG was separated from the independent component (IC) by ICA. Figure 3-2 and Figure 3-3 show the raw data and their ICs. It can be seen that eyeblink artifacts are extracted to component 1 in Figure 3-3.

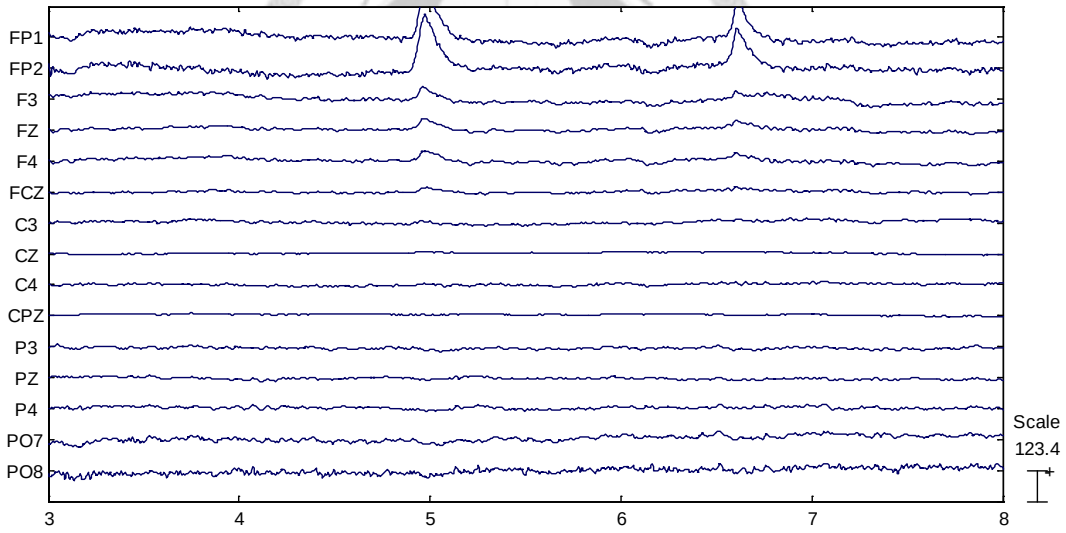


Figure 3-2 Raw EEG signals

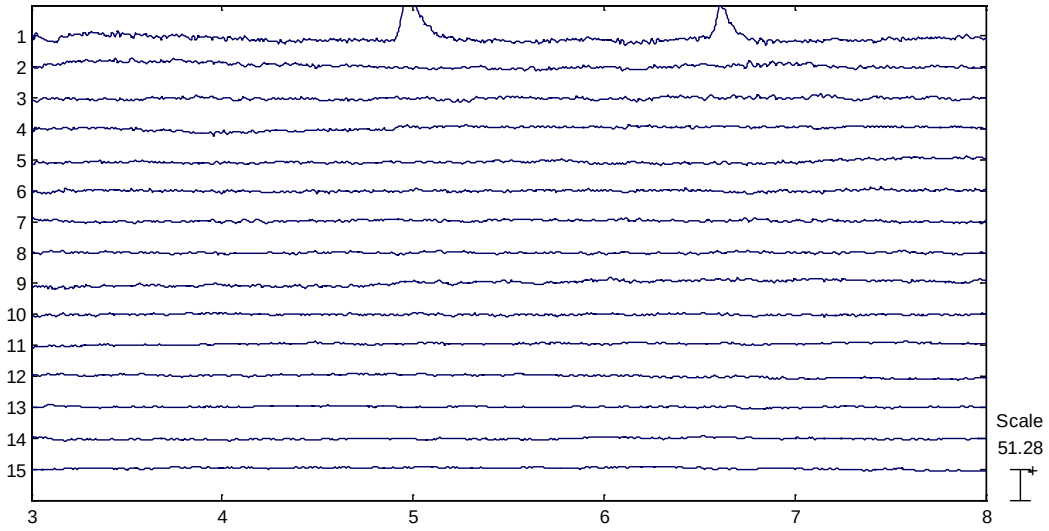


Figure 3-3 Independent component

3.4 Features Extraction

In this study, the vector of features given from fractal dimension, sample entropy, and kurtosis were used as classified data. The feature spaces of EEG and EOG are shown in Figure 3-4.

Fractal dimension is small in the EOG signals and large in the EEG signals. Kurtosis is small in the EEG signals and large in the EOG signals. The sample entropy is small in the EOG signals and large in the EEG signals. Then EOG signals and EEG signal can be separated by these features.

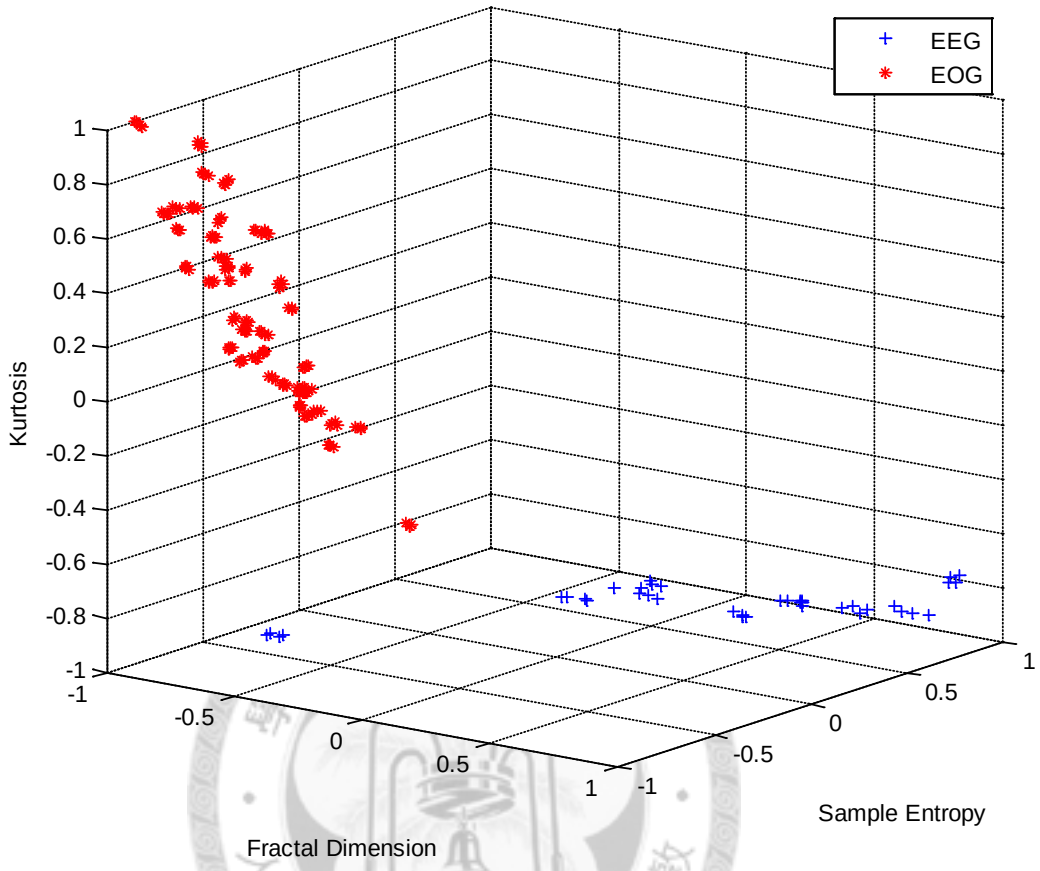


Figure 3-4 The distribution of sample entropy, kurtosis and fractal dimension

3.5 IC Selector

This study used an IC selector to select blink artifacts from independent components in the EEG. EOG signals were used as training data for the one-class classifier, and an IC selector was obtained.

Then, we calculate the distance from the center to the test data in one-class classifier. Compared with the radius of decided boundary, the data is rejected if the test data is outside the boundary.

3.6 Blink Artifacts Removal

An independent component matrix and an inverse demixing matrix can be obtained in ICA. Every row of the independent component matrix describes a independent component statistically. And the number of independent components is the same as the number of electrodes.

The IC selector identified artifact components from independent components. Then the vectors of the independent components corresponding to artifacts are replaced by zero vectors. New independent component matrix multiplied by inverse demixing matrix in ICA. After this procedure, the corrected EEG signals were obtained.

3.7 Classification Performance Indicators

In a multi-class problem, the number of data per classification is different. Errors occur because the traditional calculation of the classification rate is easily influenced by

the imbalance of data per classification. Thus, the study uses the weighted accuracy and kappa index of balance loss to evaluate the performance of the classification.

3.7.1 Weighted Accuracy

Before the classification performance indicators are introduced, they must be classified as True Positive, False Positive, True Negative, and False Negative. True Positive indicators are positive data that are judged positive. False Positive indicators are negative data that are judged positive. True Negative indicators are negative data that are judged negative. False Negative indicators are positive data that is judged negative. These classifications are shown in Table 3-1.

Table 3-1 Confusion Matrix

Predicted / **Actual**	Predicted Positive Class	Predicted Negative Class
Actual Positive class	True Positive (*TP*)	False Negative (*FN*)
Actual Negative class	False Positive (*FP*)	True Negative (*TN*)

Balance loss is defined as follows:

$$Balance\ loss = \frac{1}{2}\left(\frac{FN}{\#Positive} + \frac{Fp}{\#Negative}\right) \quad (3\text{-}1)$$

$$\text{Weighted Accuracy}(\%) = (1 - \text{balance loss}) \times 100\%, \quad (3\text{-}2)$$

where #Positive is the numbers of positive data, and #Negative is the numbers of negative data.

3.7.2 Kappa Index

The kappa index is a widely applied identical index that is an identical ratio without considering random identical states. The kappa index of agreement for categorical data was developed by Cohen[32, 33]. The general classification rate is influenced by the discrepancy of two classifications of data. The value of kappa is influenced by the data structure and represents the actual state more correctly. The range of kappa is from -1 to 1 and usually locates from 0 to 1.

Kappa is defined as follows:

$$Kappa = \frac{P_o - P_e}{1 - P_e}, \tag{3-3}$$

where P_o is the probability that the prediction agrees with the true label and P_e is the hypothetical probability of agreement by chance.

If the kappa index= 1, the raters are in complete agreement. If the kappa index= 0, there is no agreement between the raters.

The variations are calculated as follows:

Table 3-2 Inter-observer variation

		Observer 1 - Results		
		Yes	No	Total
Observer 2 - Results	Yes	a	b	m_1
	No	c	d	m_0
	Total	n_1	n_0	n

$$p_o = \left(\frac{a+d}{n}\right) \tag{3-4}$$

$$p_e = \left(\frac{n_1}{n} \times \frac{m_1}{n}\right) + \left(\frac{n_0}{n} \times \frac{m_0}{n}\right) \tag{3-5}$$

where *a* and *d* represent the numbers of the two observers that agree; *b* and *c* represent the numbers of the two observers that disagree.

Table 3-3 Interpretation of kappa

Kappa	Strength of Agreement
0.81~1.00	Almost Perfect
0.61~0.80	Substantial
0.41~0.60	Moderate
0.21~0.40	Fair
0.00~0.20	Slight
<0.00	No Agreement, less thanchance

3.8 Simulation and Results

3.8.1 Simulation

Simulated EEG signals and simulated EOG signals of ten channels have been generated to verify the results, such as Figure 3-5 and Figure 3-6. Simulated EEG signals contain one trial of 8s of noise with sampling frequency 250Hz. Simulated EOG signals include two peaks with frequency 5Hz at 3s and 6s. The peaks in different channels were multiplied by different weights.

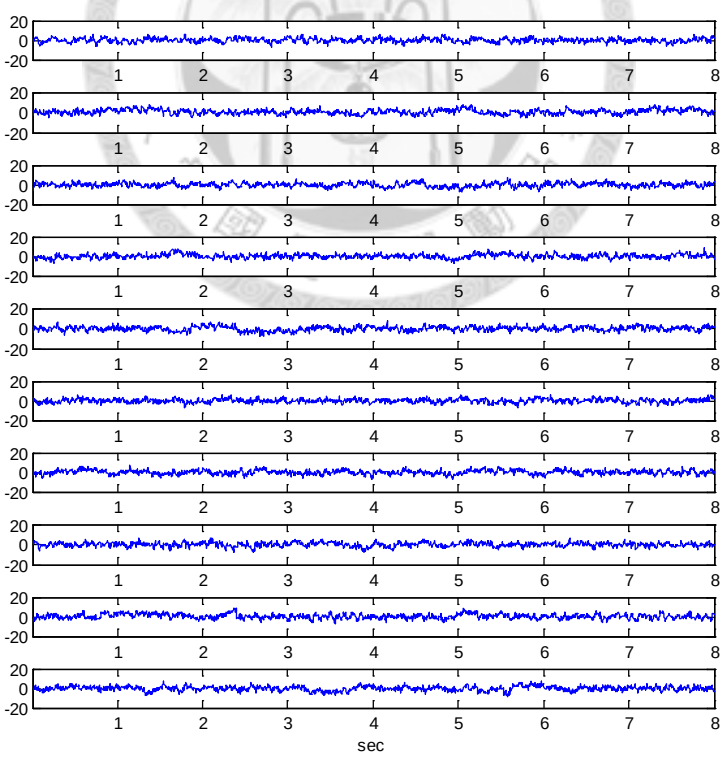


Figure 3-5 Simulated EEG signals

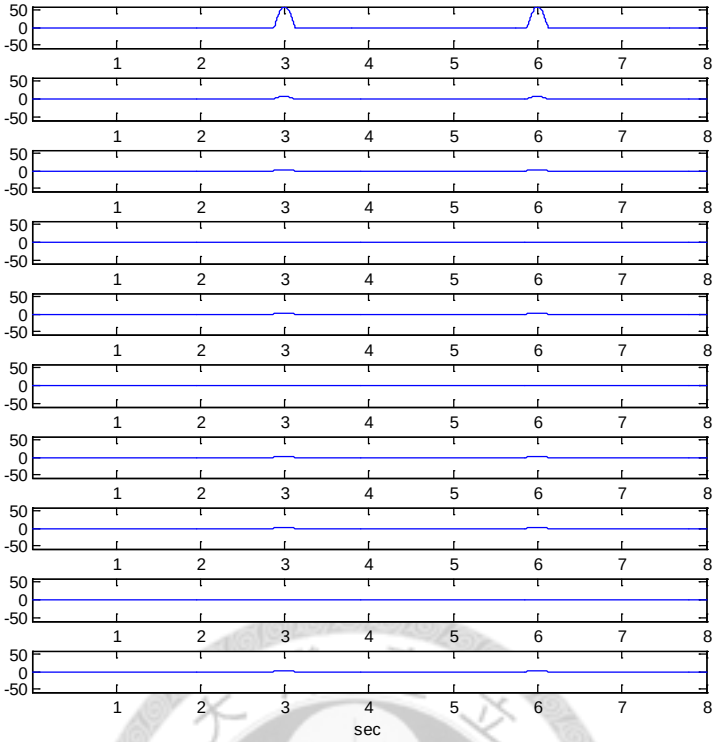


Figure 3-6 Simulated EOG signals

In order to meet the real measured signals, simulated EEG signals and simulated EOG signals are mixed, shows as Figure 3-7. Then, the simulated measured signals was separated from the independent component by ICA. Figure 3-8 shows the independent components. It can be seen that simulated EOG signals are extracted to component 1.

After ICA, feature extraction is executed for classification. In this simulation, SVDD is used to remove eyeblink artifacts. The corrected EEG signals are such as Figure 3-9.

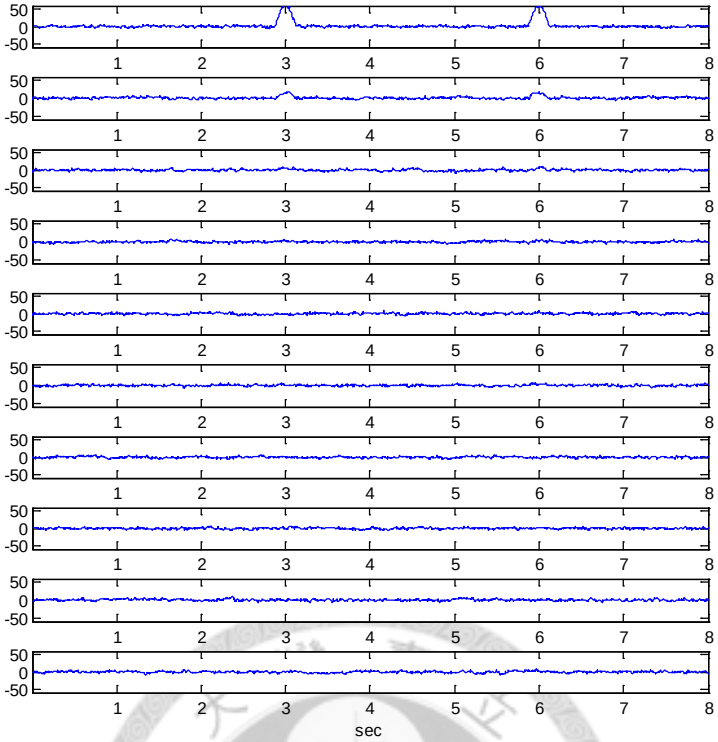


Figure 3-7 Simulated measurement signals

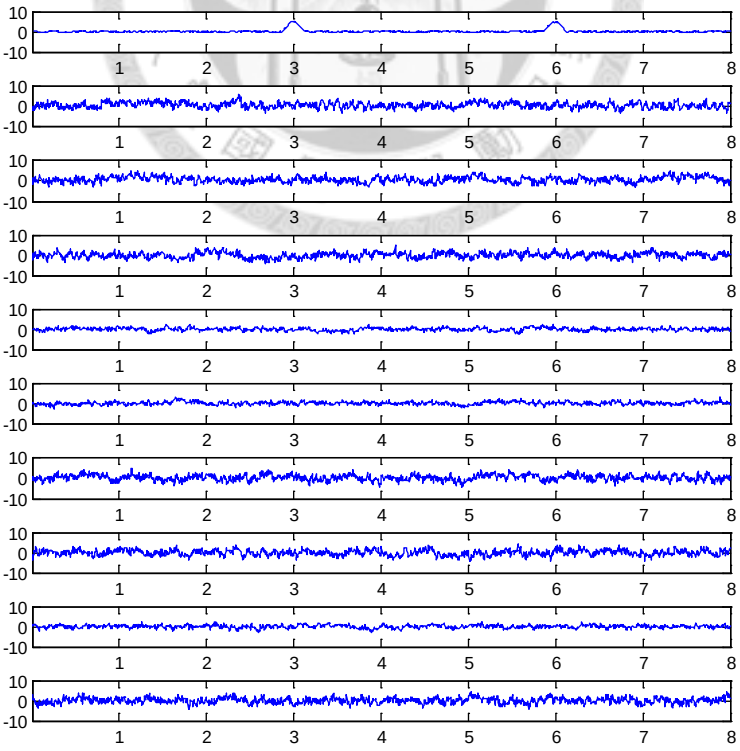


Figure 3-8 Independent components of simulated measurement signals

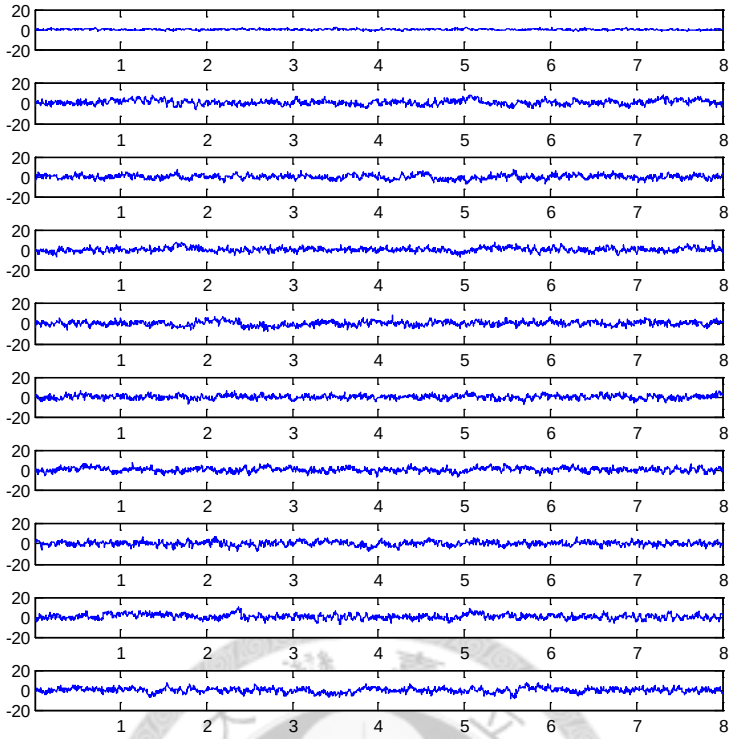


Figure 3-9 Corrected simulated measurement signals

Root mean square error (RMSE) is calculated to compare simulated EEG signals and corrected measurement signals, show as Table 3-4. In addition to channel 1, other channels are to achieve very low error. And mean RMSE is about 1.6590.

Table 3-4 Root mean square error between simulated EEG signals and corrected measurement signals

channel	1	2	3	4	5	6	7	8	9	10
RMSE	10.8329	2.5589	1.1420	0.6662	0.1814	0.5153	0.0374	0.2353	0.1168	0.3037

3.8.2 Classification Results

The center and radius of a one-class classifier were obtained by the training data. This study used three different one-class classifiers.

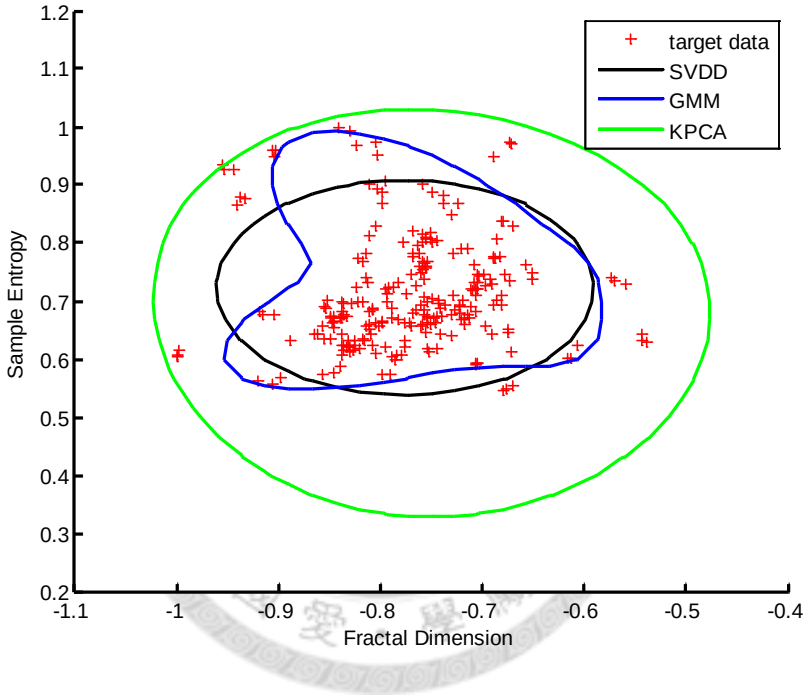


Figure 3-10 The boundary found by different one-class classifiers of the target data

We then calculated the distance from the center to the test data in each one-class classifier. When the radius of the decided hypersphere was compared, the data was rejected if the test data was outside the hypersphere. A false positive was the error of rejecting a target. A false negative was the error of failing to reject an outlier.

We calculated the ratio between the total number of correctly detected ICs and the total number of analyzed ICs, defined as the independent component detection rate (ICDR), which was employed by Mantini [34]. Therefore, in ten runs for subject 1, we achieved an average ICDR of 90.67%. We similarly achieved an average ICDR of 91.33% for subject 2. The results show that the performance of the system reached a satisfactory level.

Table 3-5 Classification rate of different one-class classifier

Subject / Method	Subject 1	Subject 2	Average
SVDD	92 %	95 %	93.5 %
GMM	90 %	91 %	90.5 %
KPCA	90 %	88 %	89.0 %
AVERAGE	90.67 %	91.33 %	91.0 %

The results verified that a non-eye blink IC could be effectively identified by using this approach.

3.8.3 Reconstruct the corrected EEG

The objective of the artifact removal was to obtain a corrected EEG. Figure 3-11 shows the flowchart of the corrected EEG reconstruction.

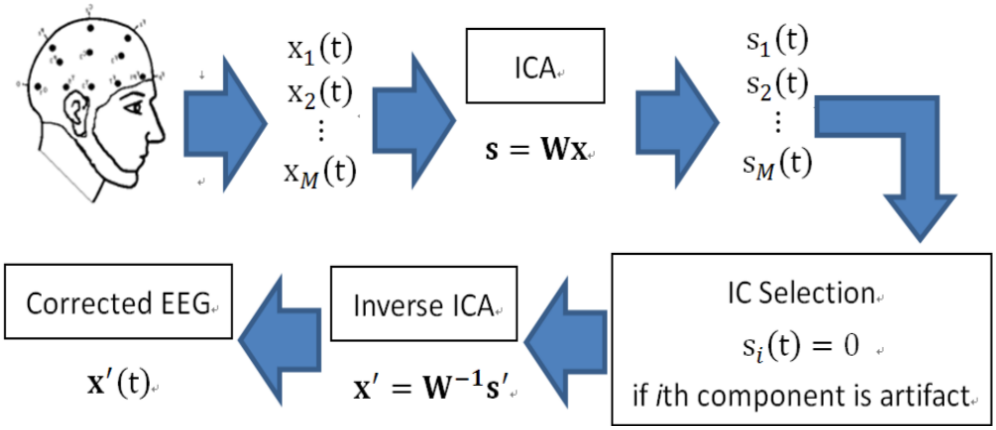


Figure 3-11 Flow chart of the corrected EEG reconstruction

After finishing the removal of eye-blink artifacts, the corrected EEG signal was reconstructed by inverse ICA. In order to obtain corrected EEG signals, replace the value of the independent components corresponding to artifacts by zero vectors and multiply the weights by the inverse demixing matrix in ICA. The correlated EEG signals are then obtained, as shown in Figure 3-12. The figure shows that the waveform of correlated EEG is similar to that of the original signal. They are different only in the time intervals between eyeblinks.

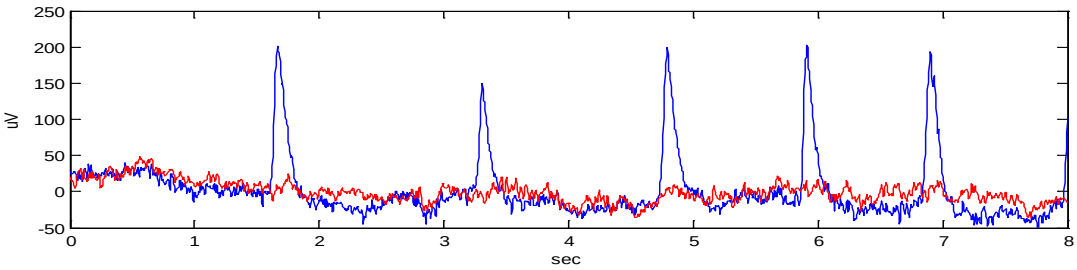


Figure 3-12 Comparison of original EEG with the corrected EEG at FP1

Compare the original EEG with the corrected EEG at FP1, which is displayed in Figure 3-12. The figure shows that the reconstructed EEG is similar to the original EEG. It also demonstrates that the method can remove eye-blink artifacts without obvious distortions.

3.9 Summary

Among the three one-class classifiers, SVDD showed the best performance. The average classification rate was 91%. The results demonstrate that the method is useful to remove eye-blink artifacts automatically. When one category of data is difficult to obtain, one-class classifier is a good choice.

The waveform of P300 potentials and eye-blink signals is similar. The results confirmed that the proposed an approach remove eye-blink without destroying P300 potentials.

Chapter 4 Application I: P300 Speller

4.1 Materials and Methods

4.1.1 Introduction to P300 Speller

Many EEG components can be used in control signals, including event-related desynchronization/synchronization (ERD/ERS) and event-related potentials (ERP). Among ERPs, P300 potentials have the advantage of requiring less training time. The task for invoked P300 potentials is the oddball paradigm [35],where an attended target stimulus is infrequently presented between non-target stimuli. The P300 potential of the human brain is a well-described positive deflection of the ongoing EEG signal, with a latency of 300 ms to a stimulus. The application of P300-based BCI is based mainly on Speller.

Farwell and Donchin developed a P300-based speller in 1988 [36]. It shows a 6×6 matrix of letters, numbers, and/or other symbols or commands. The subject focuses his/her attention on one of the 36 cells of the matrix. Every 125 ms, a single row or column flashes randomly; in a complete trial of 12 flashes, each row or column flashes twice. Thus, the row and the column containing the target character will elicit a P300. The

user makes a selection by counting how many times the row or column EEG over parietal cortex is digitized, the average response to each row and column is computed, and P300 amplitude for each possible choice is computed. As shown in Figure 4-1, P300 is prominent only in the responses elicited by the desired choice, and the BCI uses this effect to determine the user's intent. By detecting the P300, the BCI can spell the words in the human's mind.

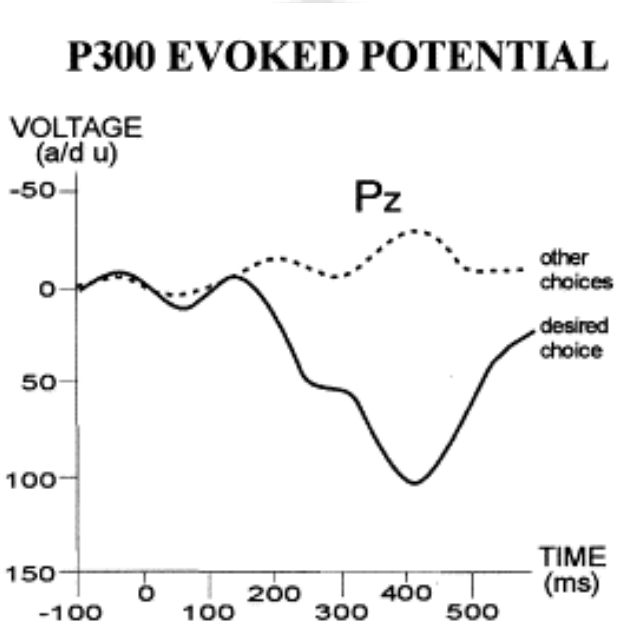


Figure 4-1 P300 evoked potential. A matrix of possible choices is presented on a screen, and scalp EEG is recorded over the centroparietal area while these choices flash in succession. Only the choice desired by the user evokes a large P300 potential (i.e., a positive potential about 300 ms after the flash) (Farwell and Donchin).

The goal of P300 Speller is to estimate to which letter in a 6×6 matrix with successively intensified row response and column response the subject was paying attention. Figure 4-2 shows the flowchart of P300 Speller.

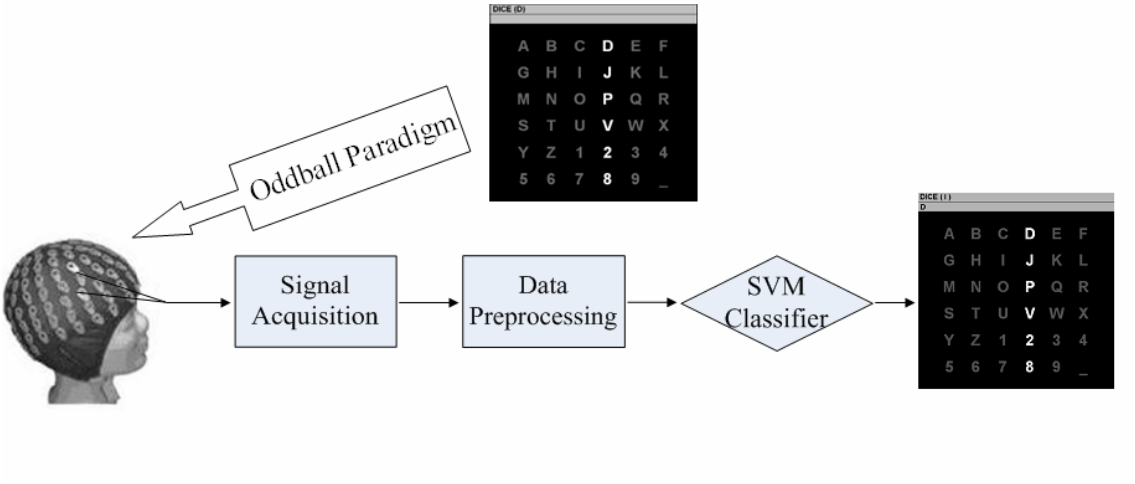
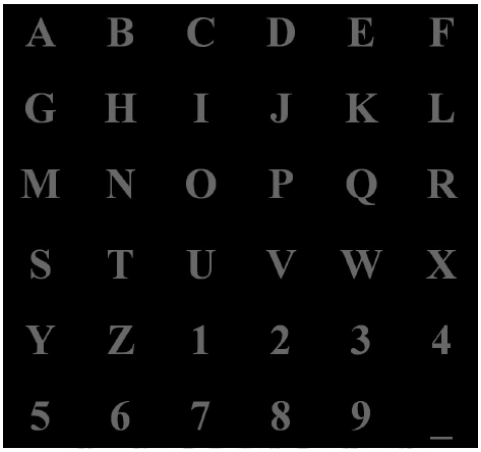


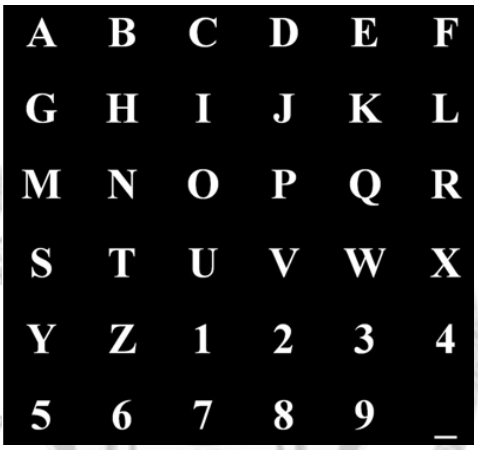
Figure 4-2 Flowchart of P300 Speller

4.1.2 P300 Panel

The panel is developed by Visual C#. The stimulus method is the oddball paradigm. By showing a visual keyboard on the monitor, a 6×6 matrix is designed in the P300 panel. Using the P300 potentials, 36 commands are designed in the P300 panel, as shown in Figure 4-3. The matrix is composed of 26 letters, numbers from 1–9, and "_". The user can spell words through the panel.



(a)



(b)

Figure 4-3 P300 panel (a) dark screen (b) light screen

There are four steps in the flash processes of the panel:

Step1: After the panel is started, the monitor shows the target that needs to be stared at for 4 seconds, as shown in Figure 4-4.

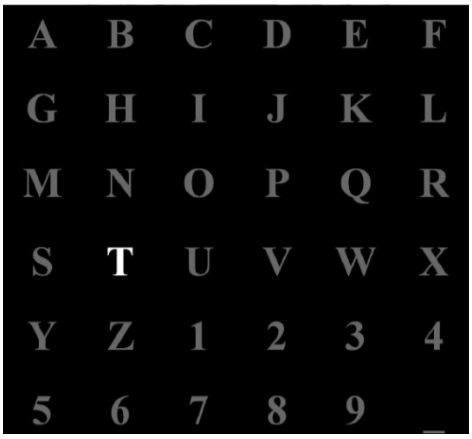


Figure 4-4 Prompt state panel

Step2: After the 4-second prompt, the panel returns to a dark screen and maintains that state for 2.5 seconds, as shown in Figure 4-3(a).

Step3: Steps 1 and 2 are the prompt stages. After the prompts, the panel starts to flash. The flash method is to show the column or row randomly; the sequences of columns and rows are independent. The flash interval of each column or row is 175 milliseconds, which includes 100 milliseconds of light and 75 milliseconds of dark. The interval of light and dark is called the inter-stimulus interval (ISI). The panel has six columns and six rows. One “round” means that each column and row has flashed one time, as shown in Figure 3-4. The interval of each round is 500 milliseconds. The panel does not stop flashing until ten rounds are completed, and the “trial” is defined by the process. The process interval, shown in Figure 3-5, is called inter-trial interval (ITI). Each trial has 180 flashes.

Step 4: After one trial is finished, the panel turns blank, which is the trigger for finishing a character. Before the next character test starts, the panel stops for 2.5 seconds to relax the experimental subject and help the subject keep his/her concentration.

During the process of flashes, the subject has to relax and focus on the target character, and count the times that the target flashes. The panel flashes the columns and rows randomly, so when the subject stares at the target, one column and one row are triggered against the target, which is the cross of one column and one row. In other words, one column and one row comprise the lower frequency of triggers in oddball block, and the other five columns and five rows (non-target) are the higher frequency. The probability of target is 2/12 (16.7%). The above method brings out the relative P300 electric potential.

4.1.3 Brain signal acquisition

Brain waves were recorded from 15 electrodes mounted on a 64-channel cap. The electrode sites were positioned according to the international 10-20 system. During the experiment, impedances were kept at no more than 5k Ohms. The raw signals were

amplified and filtered by a NeuroScan system coupled with a SynAmps2. The sampling rate was 250 Hz and was filtered with a cutoff frequency of 0.5–40 Hz.

4.1.4 Eye-Blinks Artifacts Removal

Before data preprocessing, eye-blink artifacts were removed from the EEG by the approach proposed in Chapter 3. Raw EEG signals and corrected EEG signals (after eye-blink artifact rejection) are shown in Figure 4-5 and Figure 4-6. Three corrected EEG datasets corresponding to different one-class classifiers are obtained here. SVM also was used to generate a set of corrected EEGs.

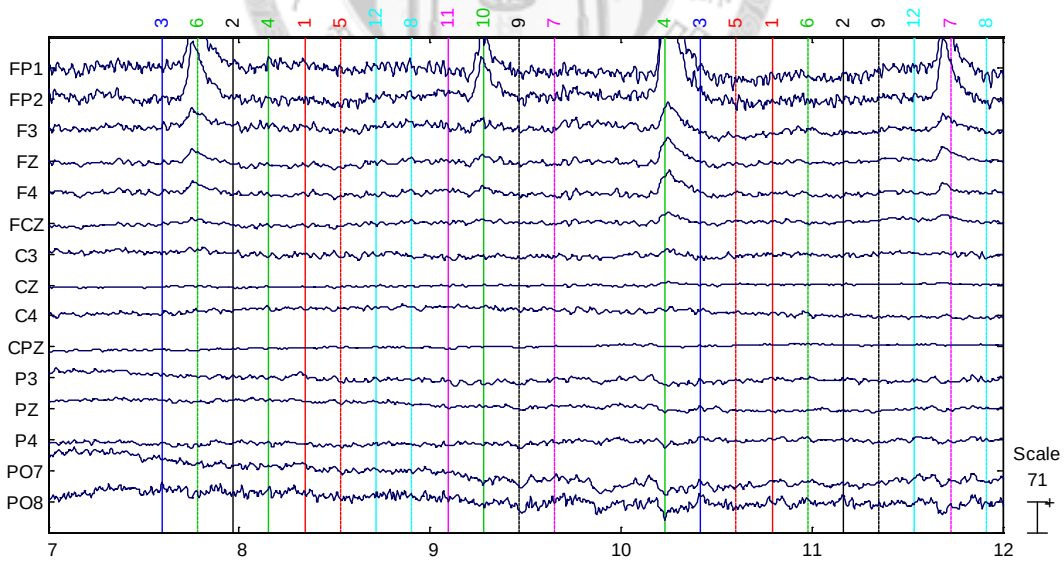


Figure 4-5 Raw EEG signals

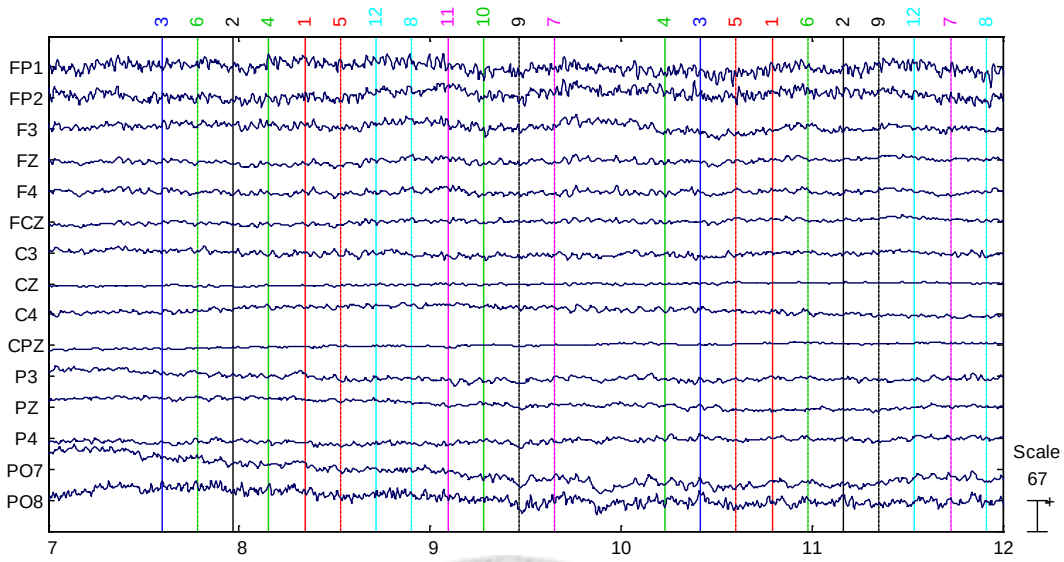


Figure 4-6 Corrected EEG signals (after eye-blink artifact rejection)

4.1.5 Data Preprocessing

In this study, a digital fourth-order Butterworth band pass filter (1~15Hz) was used to extract the EEG signals, and the data samples from 0 to 50ms after stimulation were used for a baseline correction to correct the drifted EEG signal.

The brain signals were collected from the electrode cap and the channels (Fz, FCz, Cz, CPz, C3, and C4) for the detection of P300 potentials were chosen; the reference voltage was the average voltage of PO7 and PO8. Because the evoked P300 potentials appeared about 300 ms after the stimulus, the data points of six channels were extracted from 0 to 500ms. Therefore, the total number of data points is 750 = 6 × 125. To

determine the accuracy of the P300 and non-P300 classifications, 82 epochs with P300 and 410 epochs without P300 were collected from each subject for training and testing.

4.1.6 Averaging Rounds

Because the P300 potentials are very small, they are easily concealed by background activity in an EEG. In order to make the P300 potentials become apparent, an averaging method is used. The P300 potentials are always positive, while the background wave is sometimes positive and sometimes negative; therefore, after superimposing on average, the P300 signal will be highlighted.

In P300 experiments, each trial contains ten rounds. In each round, each row and column in the panel is flashed randomly. Each row and column corresponds to a stimulus code. After the end of a trial, the data corresponding to the same stimulus code is averaged.

4.1.7 Brain signal classification

After data preprocessing, the features of the EEG signals are input into a classifier to predict the intentions of the patient. Here, a support vector machine (SVM) was chosen as the classifier. Below is a brief description of the SVM.

The training set$\{x_i, y_i\}_{i=1}^{N}$ contains input vectors $x_i \in X \subseteq \Re^D$ and their label $y_i \in \{-1, +1\}$.

The basic problem of SVM[29] is as follows:

$$\begin{aligned} &min_{f,b,\xi}\ \frac{1}{2}\|w\|^2 + C\sum_i^N \xi_i\,, && \text{for } i = 1,2,\dots\dots,N, \\ &s.t.\ \ y_i(w^T\phi(x_i) + b) \geq 1 - \xi_i, && \forall i, \\ &\xi_i \geq 0, && \forall i, \end{aligned} \tag{4-1}$$

The dual problem of SVM[29] is as follows:

$$\begin{aligned} max_{f,b,\xi}\ -\frac{1}{2}\sum_{i,j}^N \alpha_i\alpha_j y_i y_j K(x_i, x_j) + \sum_i^N \alpha_i\,, \qquad & for\ i = 1,2,\dots\dots,N \\ s.t.\ \ \sum_i^N \alpha_i y_i, & \\ 0 \leq \xi_i \leq C, \qquad & \forall i \end{aligned} \tag{4-2}$$

Finally, we can obtain the decision function:

$$f(x) = \sum_i^N \alpha_i^* K(x, x_i) + b^* \tag{4-3}$$

The kernel type is chosen as Gaussian kernel:

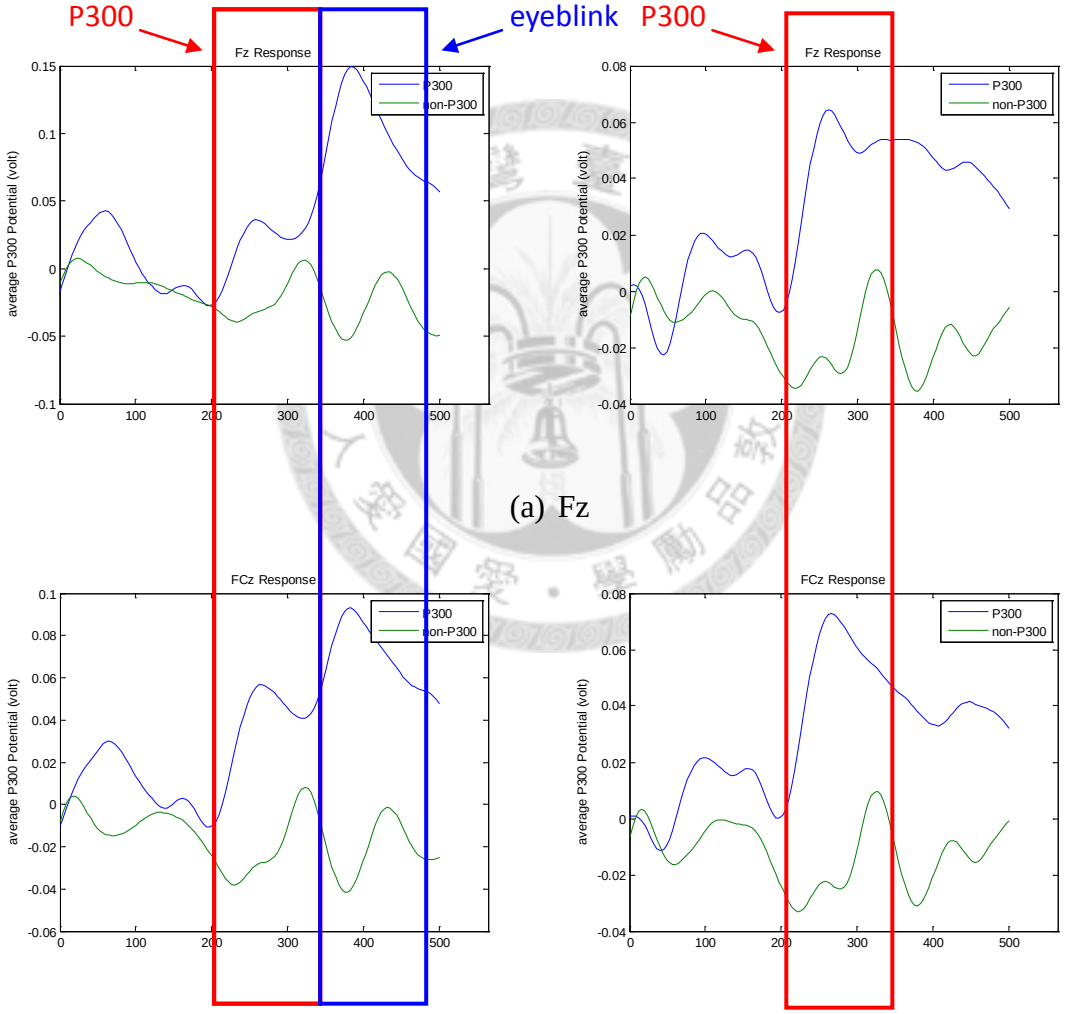
$$K_{Gaussian}(x_i, x_j) = \phi(x_i)\phi(x_j)^T = exp(-\frac{\|x_i - x_j\|_2^2}{2\sigma^2}) \tag{4-4}$$

The cross-validation is chosen as two-fold and ten runs.

4.2 Results

4.2.1 P300 Waveform

The P300 and non-P300 signals are plotted, averaged over ten rounds using different channels, as shown in Figure 4-7.



(a) Fz

(b) FCz

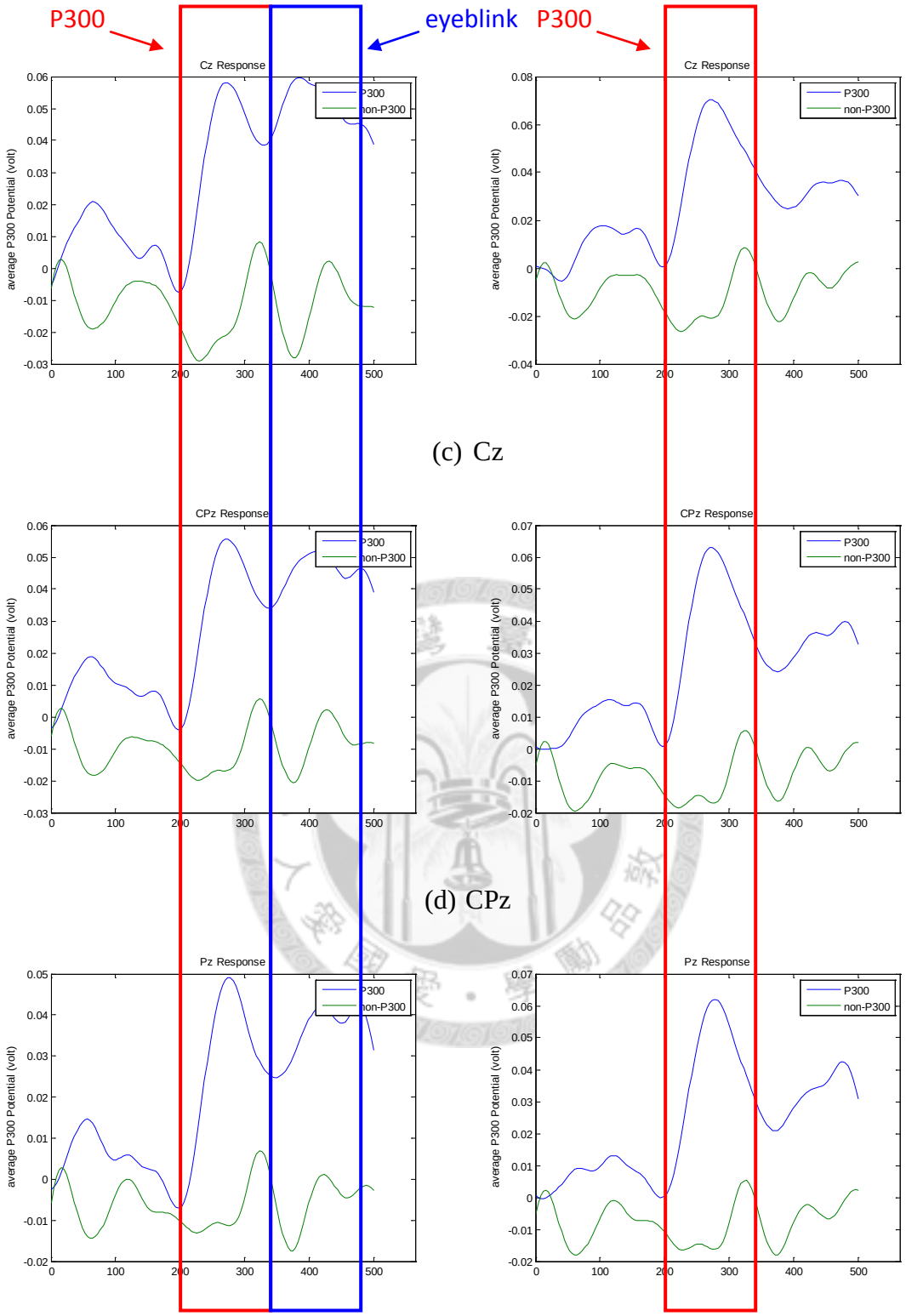


Figure 4-7 Original P300 waveform and P300 waveform after eye-blink removal at different channels. Left figures are original P300 potentials, and right figures are P300 potentials after eye-blink removal.

It can be seen that the P300 waveform becomes more obvious when the eye-blink signal is removed. The results confirmed that this method can remove eye-blink artifacts without P300 information loss.

4.2.2 Classification Results

The P300 or non-P300 signals are generated by averaging all signals from the six channels for each subject. The peak of the P300 signals is distributed from 150ms to 300ms, and the amplitude of the P300 signal is larger than that of the non-P300 signal. Therefore, we can use the difference between the P300 and non-P300 signals to estimate the intended letter.

The accuracy of the P300 and non-P300 classification is shown in Figure 4-8 and Table 4-1. The accuracy increases from one round to the average of ten rounds. The best accuracy achieved was 92.07%, which occurred when the eye-blink artifacts were removed by ICA and SVDD.

Table 4-1 Comparison of results from different one-class classifiers with different numbers of rounds

Number of Rounds	Original	ICA-SVDD	ICA-GMM	ICA-KPCA	ICA-SVM
1	58.17	50.00	57.56	55.37	53.54
2	63.78	73.05	63.66	67.68	62.07
3	72.07	78.41	72.07	74.39	61.59

4	82.93	84.15	82.93	83.41	70.24
5	82.56	87.56	82.56	84.27	72.68
6	85.61	89.15	85.61	84.88	80.24
7	86.59	89.27	85.98	86.71	81.71
8	89.88	90.73	89.88	89.63	82.56
9	90.49	92.68	90.49	92.07	83.78
10	90.00	92.07	90.00	89.27	88.66
Average	80.21	82.71	80.07	80.77	73.71

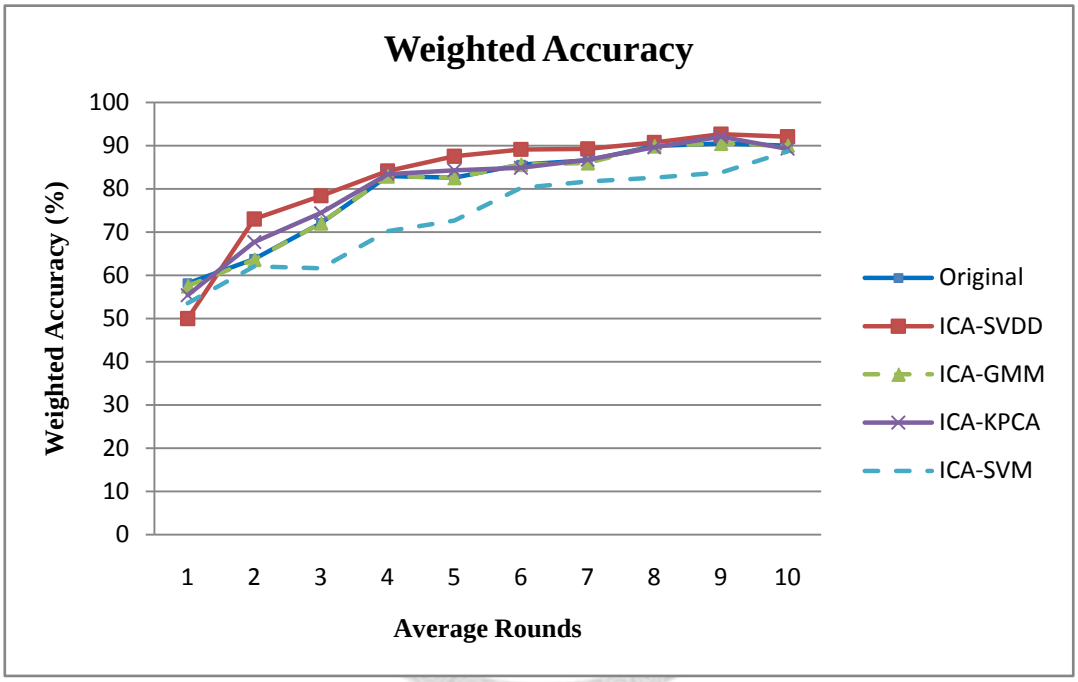


Figure 4-8 Weighted accuracy of eye-blink artifact removal with different one-class classifiers from 1 to 10 rounds

4.2.3 Correct of classification character vector

Each round included 12 stimulus codes. By using the rows and columns corresponding to the stimulus codes, the subject can spell out intended letters, as shown

in Figure 4-9. The correct classification character vector is shown in Table 4-2 and Figure 4-10.

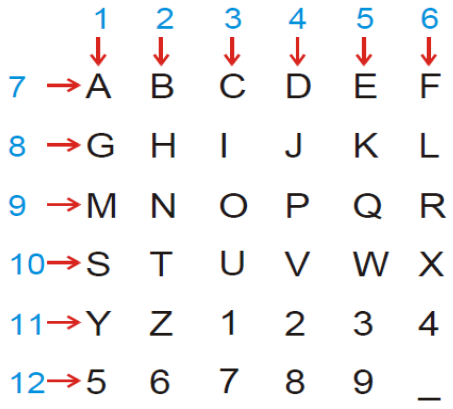


Figure 4-9 Each letter corresponds to a number

Table 4-2 Correction of classification character vector from different one-class classifier with different numbers of rounds

Number of Rounds	Original	ICA-SVDD	ICA-GMM	ICA-KPCA	ICA-SVM
1	21.95	21.95	21.95	24.39	7.32
2	36.59	36.59	36.59	31.71	31.71
3	39.02	39.02	39.02	36.59	31.71
4	68.29	68.29	68.29	70.73	31.71
5	70.73	73.17	73.17	78.05	43.90
6	70.73	70.73	70.73	75.61	48.78
7	70.73	73.17	73.17	73.17	58.54
8	78.05	78.05	78.05	80.49	70.73
9	82.93	80.49	80.49	78.05	70.73
10	82.93	82.93	82.93	82.93	78.05
Average	62.20	62.44	62.44	63.17	47.32

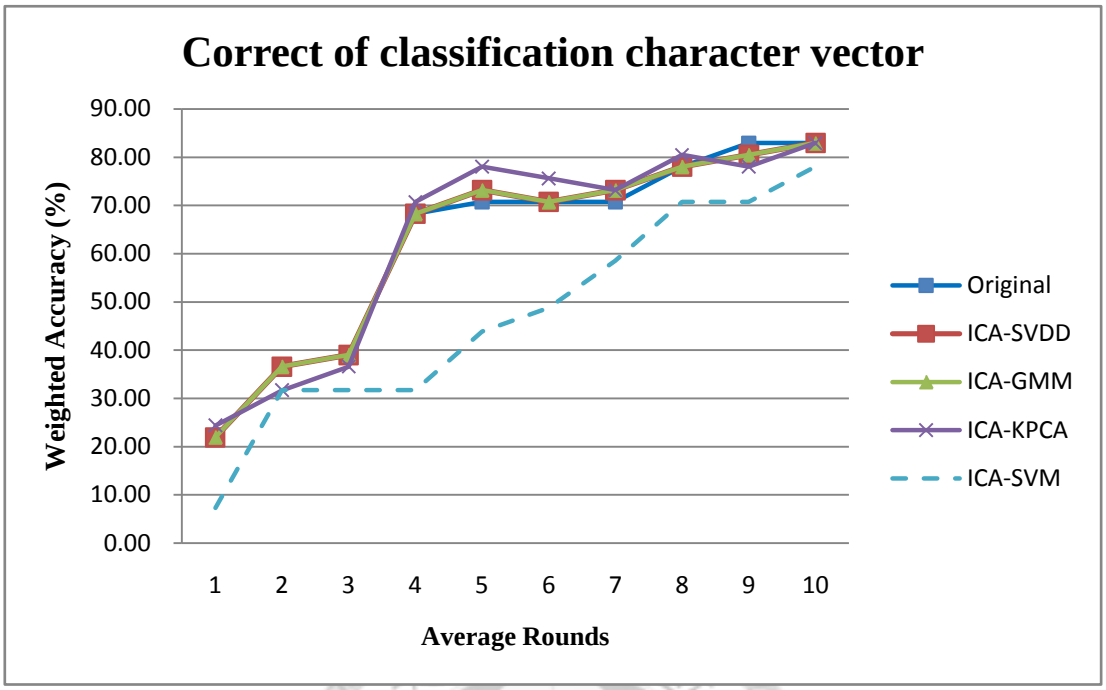


Figure 4-10 Correction of classification character vector of eye-blink artifact removal with different one-class classification from 1 to 10 rounds

4.3 Summary

In the classifications of P300 and non-P300, it is obvious that SVDD performs the best. It promotes 9.27% of the classification when the number of average rounds is two, and 5% of the classification when the number of average rounds is five.

After the number of average runs promoted, the eye-blink signals are averaged continuously, and the P300 signals become more obvious. Hence, the classification of signals with eye-blinks removed becomes less obvious as the number of average runs increases.

The SVM performs the worst, and it is concluded that the one-class classification is more suitable to remove eye-blink signals than the two-class classification.

Chapter 5 Application II: Sleep Stage Classification

5.1 Materials and Methods

5.1.1 Introduction to Sleep Stage

People's sleep can be classified into two main stages: rapid eye movement (REM) and non-rapid eye movement (NREM). NREM can be further classified into three stages: stages N1, N2, and N3. The following is a list of sleep stages and the criteria doctors use to determine into those stages of sleep:

1. Wake Stage:

(1)When the alpha wave of the zone in the back of the head is greater than 50% of one epoch, the person is in the Wake stage.

(2)If there are no visual alpha waves, one of these conditions must apply:

(a) A blinking frequency of 0.5-2 Hz

(b) The conditions of the reading types' rotations of eyeballs

(c) Irregular, rapid rotation of the eyeballs with normal or high tension in the lower jaw

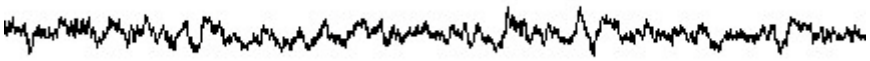


Figure 5-1 Brain signal during the Wake stage

2. Stage N1, or NREM 1: This stage is also known as the light sleep stage. It is defined as an obvious 4-7 cycle per second (cps) brain type comparative to low amplitudes and combined frequencies.

(1) If the wave one with the alpha wave reduces, is replaced by low amplitudes and combined frequencies and the time exceeds 50% of one page, the person is considered to be in stage N1.

(2)The activity that is without alpha wave:

(a) The 4-7 Hz theta wave with over 1 Hz reduction in the background frequency.

(b) The shape wave that is measured on the top of the head. (The shape of the wave is thin as a needle and the interval is within 0.5 seconds.)

(c) Slow rotations of the eyeballs.

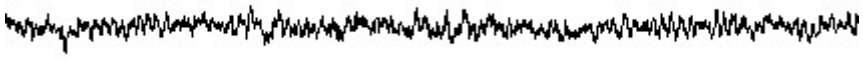


Figure 5-2 Brain signal during stage N1

3. Stage N2, or NREM 2: This stage is characterized by brain waves concretely displaying a spindle and K-complex. If one of them appears, judge as the state N2.

A spindle is a wave with series of definite 11-16 Hz short rhythms lasting at least 0.5 seconds. The most common frequency is 12-14 Hz. The spindle usually increases or decreases gradually and the amplitude which is measured on the center area of the head is maximum.

A K-complex is made up of a negative sharp wave followed by a slow positive wave. They must last at least 0.5 seconds. The maximum amplitude of a K-complex is generated on the forehead. (A K-complex is larger than 75 μV and lasts longer than 0.5 seconds.)

Figure 5-3 Brain signal during stage N2

4. Stage N3, or NREM 3: No matter the age of the sleeper, if a 30-second record of waves consists of more than 20% of slow waves 0.5-20 Hz, and the amplitudes are greater

than 75 μV, the person is considered to be in stage N3. The eyeballs usually do not rotate rapidly or slowly during stage N3.

Note that the patients with sleep apnea usually cannot reach NREMS which is in deep sleep.

Figure 5-4 Brain signal during stage N3

5. Stage R, or REM: The electromyogram (EMG) amplitude of REM sleep has the lowest EMG signal compared to other stages. REM sleep exhibits only low amplitude waves α waves, β waves, and so on).

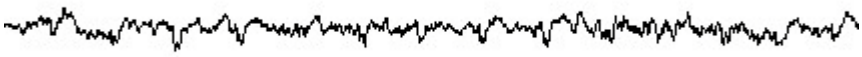


Figure 5-5Brain signal during stage R

Electroencephalogram (EEG) arousal: Repeated EEG arousals do not cause any changes to sleep stages or cause the body to move. However, EEG arousal does usually interrupt the persistence of sleep, so it could potentially result in that the patient wants to sleep during the day. The main characteristic of EEG arousal is a sudden change in brain wave frequencies sometimes followed by an increase in the amplitude of the brains that is

measured on the lower jaw. The main criterion of judgment parameter is the EEG of the center area of the head which is monopolar derivation (C4/A1 or C3/A2), or EEG of zone behind head (O1/A2 or O2/A1) and EMG of mandible of bipolar derivation.

The criteria of judgments: The frequencies of brain waves have sudden changes and last more than 3 seconds. They can include alpha waves, theta waves, or waves which are more than 16 Hz. Before the frequencies of waves change, they must demonstrate more than 10 seconds of stable sleep. In addition, the judgment of the EEG arousal of REM should follow increases of amplitudes of the lower jaw and last 1 second at least.

The judgment of sleep stage is determined by the doctor manually at present. It a time-consuming procedure and the doctor are required prior training. So this application try to classified sleep stage automatically, it hopes to help sleep stage classification in the future.

5.1.2 Dataset

The data presented here is from experiments conducted at the sleep center in Cheng Ching Hospital. The data includes EEG, EOG, EMG, and ECG readings. The sampling frequency is 200 Hz. The electrodes used to take measurements during the EEG are C3, F3, C4, O1, O2, and GND and are placed on the forehead. C3, F3, O1, and A2 are

referenced electrodes. C4, F4, O2, and A1 are referenced electrodes. The measure type of EMG is bipolar and measures the area of the jaw. The measure type of ECG is 2-lead. Each data set is made up of 30-second epochs. Every epoch has been given a sleep stage label by a doctor.

Table 5-1 Sleep EEG data –numbers of each stage

Subject	ID	Wake	N1	N2	N3	REM	Total epoch
1	A0002053	493 (50%)	98 (10%)	300 (31%)	0 (0%)	83 (9%)	974
2	A0000539	424 (46%)	90 (10%)	367 (40%)	0 (0%)	41 (4%)	922
3	A0002055	361 (38%)	138 (14%)	253 (26%)	137 (14%)	75 (8%)	964
4	A0002049	169 (18%)	201 (21%)	394 (41%)	0 (0%)	186 (20%)	950
5	A0002045	139 (16%)	41 (5%)	454 (53%)	73 (9%)	147 (17%)	854
6	A0002047	158 (16%)	70 (7%)	551 (54%)	0 (0%)	235 (23%)	1014
7	A0002056	129 (14%)	27 (3%)	612 (69%)	0 (0%)	128 (14%)	896
8	A0002054	154 (17%)	214 (23%)	419 (46%)	0 (0%)	123 (14%)	910
9	A0002051	128 (13%)	182 (18%)	472 (46%)	110 (11%)	116 (12%)	1008
10	A0002052	64 (7%)	40 (5%)	436 (49%)	134 (15%)	216 (24%)	890
11	A0000442	61 (7%)	54 (6%)	557 (62%)	78 (9%)	148 (16%)	898
12	A0002048	52 (5%)	35 (4%)	605 (62%)	57 (6%)	221 (23%)	970
13	A0001673	44 (5%)	47 (5%)	549 (63%)	43 (5%)	195 (22)	878
	Total	2376	1237	5969	632	1914	

5.1.3 Hardware and Software Interface

The software used to present the data is Alice 5. The interface of the software is as shown in Figure 5-6. The main functions of the software are measuring sleep stages and labeling the behavior of the disease of the subject while the subject is sleeping. For example, subjects with apnea, unusual twitching of legs, and so on were observed. The main functions of the manipulation are the data output and reading the sleep stage labels.

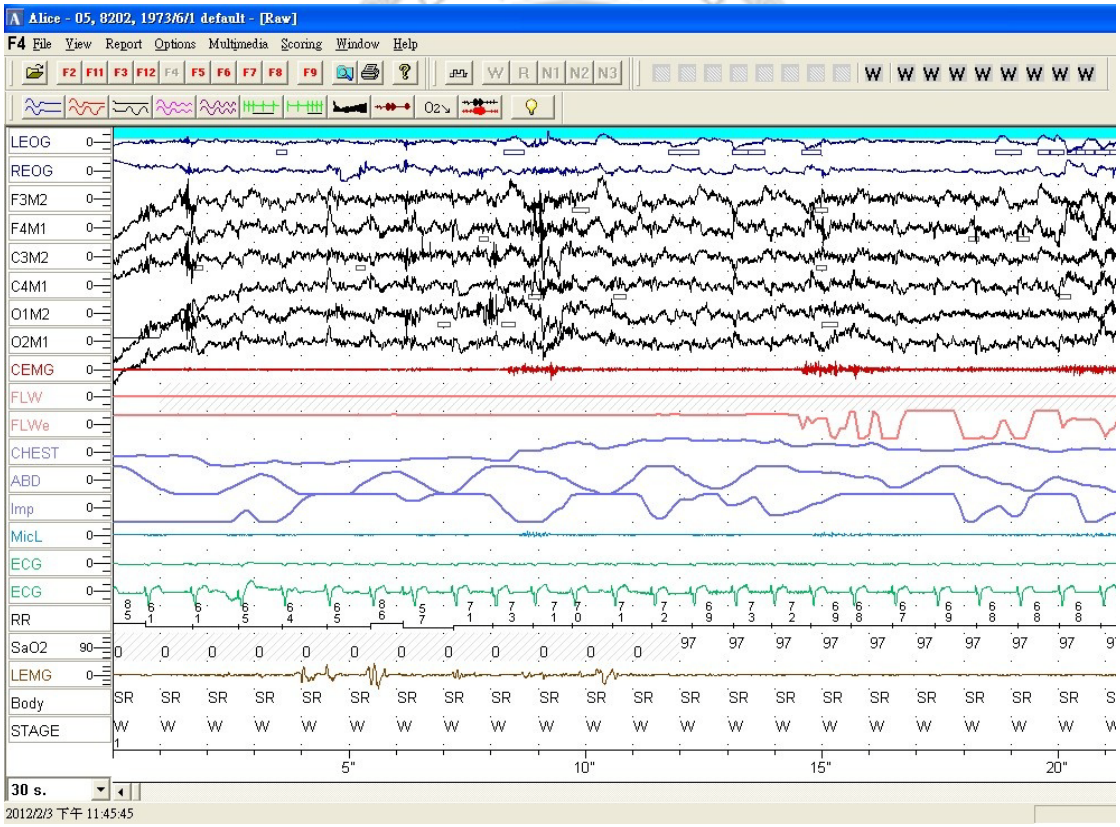


Figure 5-6 The interface of the Alice 5 software

The receiving of data needs the amplifier and the computer for transmission, as shown in Figure 5-7 and Figure 5-8.

Figure 5-7 The Alice 5 amplifier

Figure 5-8 The computer for Alice 5 transmission

5.1.4 Experiment setup

Two sensors were place on the chest and abdomen of the subject to measure the movement of the chest and abdomen. The 10-20 system was used to measure the proper locations of F3, F4, C3, C4, O1, O2, and GND on the head, as shown in Figure 5-9. Conductive adhesive was applied to the electrodes and they were pasted in the correct positions.

Figure 5-9 Confirming the positions of the electrodes according to the 10-20 system

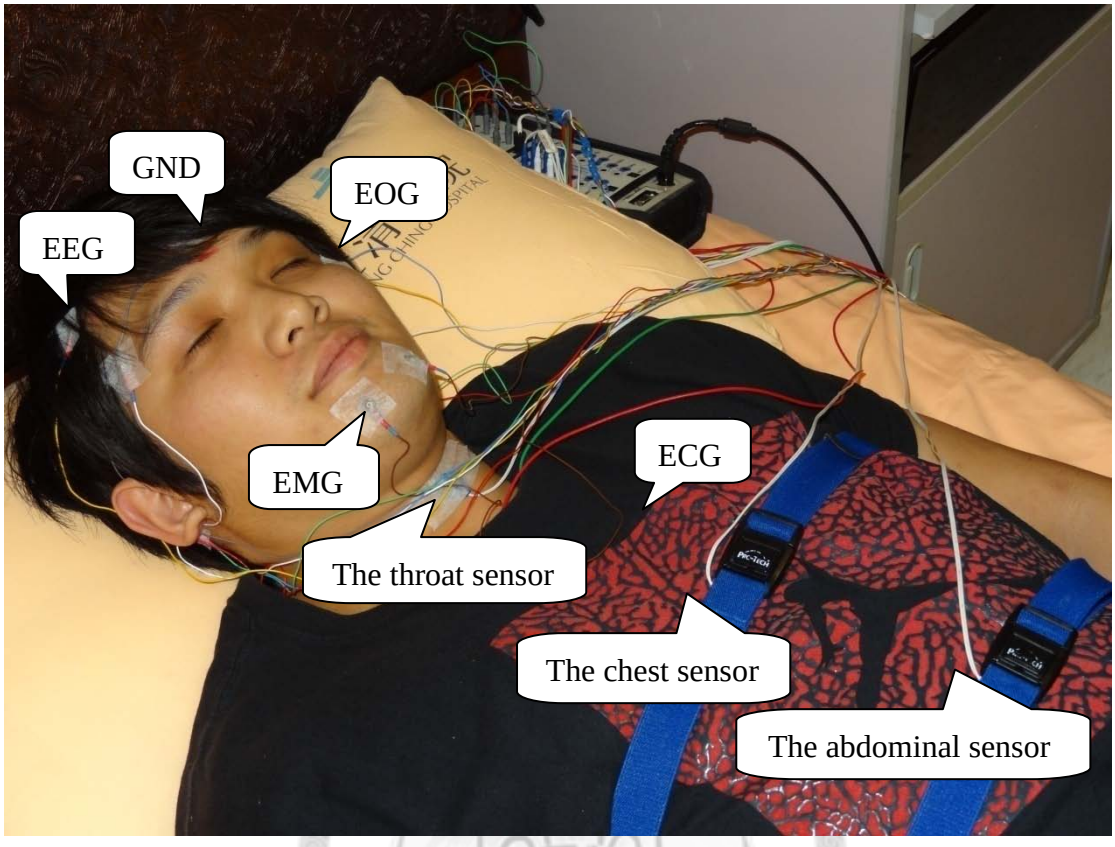


Figure 5-10 Positions of electrodes

After the electrodes were applied, the instruments were calibrated and tested. The subject opened his eyes for 30 seconds, closed his eyes for 30 seconds, moved his eyes up and down, turned his eyes left and right, took a deep breath, held his breath, blinked, ground his teeth, coughed, and moved his legs to calibrate the instruments and test whether they were working properly. After the calibration, the light was turned off and the observer set the computer software to “lights off” and began to record.

The electrodes were set in three pairs: C3 and C4, F3 and F4, and O1 and O2. Every pair of electrodes had different functions. The O pair of electrodes was used to

observe changes in α wave. The F pair of electrodes was used to observe changes in the big waves. The C pair of electrodes was used to evaluate sleep stages. The sleep stages are judged by a single electrode only. By recording different pairs of electrodes simultaneously, the experimenters could avoid problems (such as a lack of records of brain waves) resulting from single electrodes falling off due to the subject moving during sleep (if the C3 electrode fell off, the experimenter could use signals recorded by the C4 electrode during the same time interval to evaluate the sleep stage). This is shown in Table 5-2.

Table 5-2 Main purpose of each electrode

Electrodes	Main purpose
F3 or F4	Observe changes in big waves
C3 or C4	Main electrodes to evaluate sleep stages
O1 or O2	Change in á wave activity

5.1.5 Power of frequency band

Each epoch was split into six 5-second intervals. And then count the band energy ofδ, θ, α, σ, β waves in every data. The ranges of band are shown in Table 5-3. Then, use the mean value of the six intervals as classification features of every epoch. Hence, every epoch can get a feature vector of the length of 5.

Table 5-3 Range of frequency for each band

Frequency band	Range
δ band	0.5 - 4.5 Hz
θ band	4.5 - 8 Hz
α band	8 - 12 Hz
σ band	12 - 16 Hz
β band	16 - 30 Hz

5.1.6 Sleep Stage Classification

SVM is used for the classification of sleep stages. Every stage is classified by different levels of classifications. The sequences are N3, Wake, and N2. After those stages, REM and N1 are classified. When N3, Wake, and N2 are classified, band energy acts as classified features. After that, use ICA to extract the ocular signals before the REM and N1 classification, as shown in Figure 5-11. In this application, ocular signals are made as the available features rather than artifact. Here ocular signals are extracted by ICA. And SVDD is made as IC selector. EOG signals are used to training SVDD.

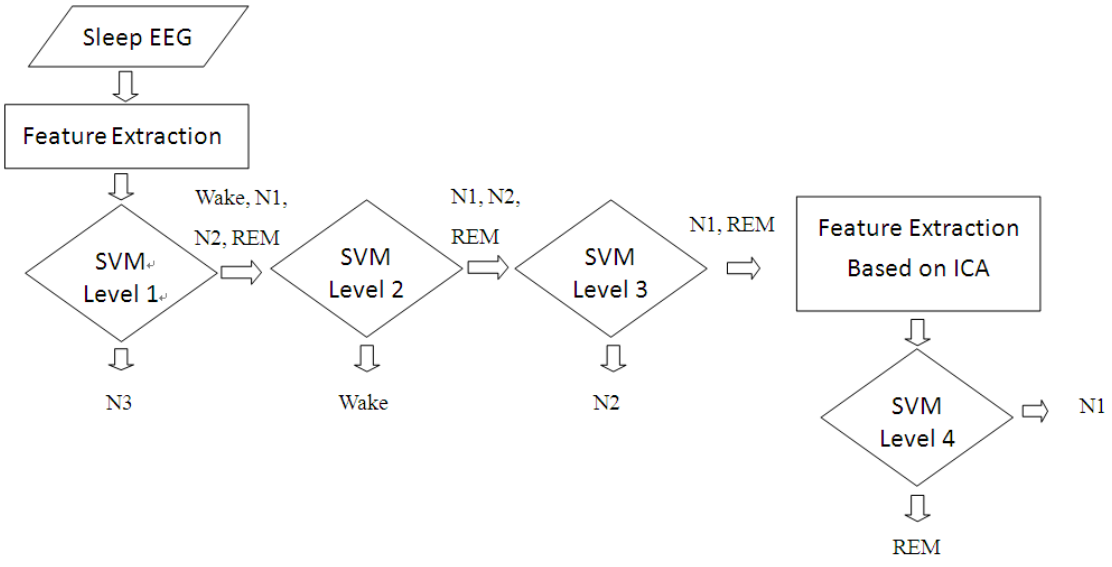


Figure 5-11 The flow chart of sleep stage classification

5.2 Results

If a data which is measured by a subject is the test data when the classification is counted, other subjects are the training data. For example, when the subject 4 is the test data, subject 1~3 and 5~13 are the training data.

5.2.1 N1 vs. REM Classification

The classifications of N1 and REM are the most difficult of all stages. Therefore, the main purpose of the test is to promote the classification of N1 and REM. Therefore, not only band energy of EEG but also fractal dimension and kurtosis of ocular signals act as classified features. The results are shown in Table 5-4 and

Table 5-5.

Table 5-4 The weighted accuracy of N1 vs. REM classification

	EEG	**EEG and ocular IC**
Subject 1	84.24%	85.21%
Subject 2	68.40%	68.96%
Subject 3	68.96%	68.59%
Subject 4	62.68%	62.83%
Subject 5	61.09%	58.45%
Subject 6	53.15%	54.07%

Subject 7	70.52%	75.88%
Subject 8	80.98%	85.06%
Subject 9	59.42%	60.40%
Subject 10	62.08%	63.15%
Subject 11	56.73%	61.36%
Subject 12	64.47%	66.35%
Subject 13	62.55%	65.23%
Average	65.79%	67.35%

Table 5-5 The kappa index of N1 vs. REM classification

	EEG	**EEG and ocular IC**
Subject 1	0.687	0.701
Subject 2	0.434	0.451
Subject 3	0.369	0.361
Subject 4	0.258	0.261
Subject 5	0.253	0.205
Subject 6	0.092	0.119
Subject 7	0.212	0.285
Subject 8	0.567	0.642
Subject 9	0.155	0.171
Subject 10	0.155	0.182
Subject 11	0.182	0.295
Subject 12	0.195	0.223
Subject 13	0.328	0.378
Average	0.299	0.329

5.2.2 All Sleep Stage Classification

The results of the classification of all sleep stage counts are shown in Table 5-6 and Table 5-7.

Table 5-6 The weighted accuracy of all stages classification

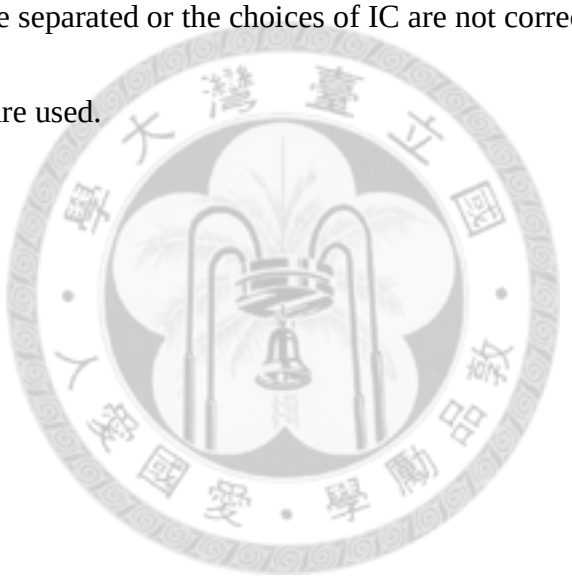
	EEG	**EEG and ocular IC**
Subject 1	78.09%	78.44%
Subject 2	77.23%	77.45%
Subject 3	61.03%	60.62%
Subject 4	62.91%	63.65%
Subject 5	59.38%	59.24%
Subject 6	71.93%	72.30%
Subject 7	67.48%	68.11%
Subject 8	77.64%	78.36%
Subject 9	61.09%	61.27%
Subject 10	66.24%	67.45%
Subject 11	50.93%	52.41%
Subject 12	45.96%	46.05%
Subject 13	56.92%	57.14%
Average	64.372%	64.807%

Table 5-7 The kappa index of all stages classification

	EEG	**EEG and ocular IC**
Subject 1	0.780	0.781
Subject 2	0.750	0.752
Subject 3	0.521	0.518
Subject 4	0.376	0.387
Subject 5	0.597	0.595
Subject 6	0.717	0.720
Subject 7	0.322	0.327
Subject 8	0.608	0.615
Subject 9	0.477	0.479
Subject 10	0.613	0.634
Subject 11	0.449	0.458
Subject 12	0.295	0.297
Subject 13	0.641	0.639
Average	0.550	0.554

5.3 Summary

The classifications of N1 and REM are the most difficult to classify. The goal is to promote the classification of N1 and REM. Using ICA as features extracting can increase the classifications of N1 and REM by 1.56%. Using ICA to draw out EOG signals obviously does promote the classifications of N1 and REM. One possible reason is that EOG signals cannot be separated or the choices of IC are not correct under the condition that a few electrodes are used.



Chapter 6 Conclusions and Future Works

6.1 Conclusions

This thesis proposed an approach to remove eyeblink artifacts in EEG automatically. Independent component analysis can successfully separate the artifacts from the EEG signals. In this study, fractal dimension, sample entropy, and kurtosis are used as the feature rather than the activity vector. One-class classifiers are used to act as the automatic independent components selector. Three different one-class classifiers are compared with each other. The results demonstrate that the approach is useful to remove eye-blink artifacts automatically.

Two applications are used to verity this approach in this thesis. In the first application, eye-blink artifacts are removed by this approach before P300 and non-P300 classification. It successfully improves the classification rate. Sleep stage classification is the second application. Classification rate did not significantly increase in this application. This may be with less number of the electrodes.

6.2 Future works

However, other artifacts such as eye movement signals, muscle artifacts, heart signals, and line noise are still unable to be effectively removed by this method. Other ways to reject all the artifacts are still underway.

A good method of feature extraction can help get a better classification. How to find more representative features is still an important problem in the future studies.

For real-time applications, automation is essential. Full automation in artifact rejection can reduce the amount of time wasted and human error. The findings of the present study will serve as directions for part of the future research.

References

[1] L. F. Haas, "Hans Berger (1873–1941), Richard Caton (1842–1926), and electroencephalography," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 74, pp. 9, 2003.

[2] H. H. Jasper, "The ten-twenty electrode system of the International Federation," *Electroencephalography and Clinical Neurophysiology*, pp. 371-375, 1958.

[3] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, and T. M. Vaughan, "Brain-computer interface technology: a review of the first international meeting," *IEEE Transactions on Rehabilitation Engineering*, vol. 8, pp. 164-173, 2000.

[4] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw, "BCI2000: a general-purpose brain-computer interface (BCI) system," *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 1034-1043, 2004.

[5] C. J. James and O. J. Gibson, "Temporally constrained ICA: an application to artifact rejection in electromagnetic brain signal analysis," *IEEE Transactions on Biomedical Engineering*, vol. 50, no.9, pp. 1108-1116, 2003.

[6] Z. Weidong, Z. Jin, Z. Hao, and J. Liu, "Removing Eye Movement and Power Line Artifacts from the EEG based on ICA," in *Engineering in Medicine and Biology Society, 27th Annual International Conference of the*, pp. 6017-6020, 2005.

[7] B. Azzerboni, M. Carpentieri, F. La Foresta, and F. C. Morabito, "Neural-ICA and wavelet transform for artifacts removal in surface EMG," *Proceedings of the IEEE International Joint Conference on Neural Networks*, pp. 3223-3228 vol.4, *2004*.

[8] M. T. Akhtar, C. J. James, and W. Mitsuhashi, "Modifying the Spatially-Constrained ICA for Efficient Removal of Artifacts from EEG Data," in *Bioinformatics and Biomedical Engineering (iCBBE) 4th International Conference on*, pp. 1-4, 2010.

[9] J. C. Woestenburg, M. N. Verbaten, and J. L. Slangen, "The removal of the eye-movement artifact from the EEG by regression analysis in the frequency domain," *Biological Psychology*, vol. 16, pp. 127-147, 1983.

[10] S. Boudet, L. Peyrodie, P. Gallois, and C. Vasseur, "A global approach for automatic artifact removal for standard EEG record," *Proceedings of the 28th IEEE EMBS Annual International Conference*, New York City, USA, pp. 5719-5722, 2006.

[11] A. Schlögl, C. Keinrath, D. Zimmermann, R. Scherer, R. Leeb, and G. Pfurtscheller, "A fully automated correction method of EOG artifacts in EEG recordings," *Clinical Neurophysiology,* vol. 118, pp. 98-104, 2007.

[12] U. Patidar and G. Zouridakis, "A hybrid algorithm for artifact rejection in EEG recordings based on iterative ICA and fuzzy clustering," *Proceedings of the 30th Annual International IEEE EMBS Conference*, Vancouver, British Columbia, Canada,, pp. 50-53, 2008.

[13] F. Campos Viola, J. Thorne, B. Edmonds, T. Schneider, T. Eichele, and S. Debener, "Semi-automatic identification of independent components representing EEG artifact," *Clinical Neurophysiology,* vol. 120, pp. 868-877, 2009.

[14] S. Shi-Yun, S. Kai-Quan, O. Chong Jin, E. Wilder-Smith, and L. Xiao-Ping, "Automatic EEG Artifact Removal: A Weighted Support Vector Machine Approach With Error Correction," *IEEE Transactions on Biomedical Engineering,* vol. 56, no.2, pp. 336-344, 2009.

[15] M. Crespo-Garcia, M. Atienza, and J. Cantero, "Muscle Artifact Removal from Human Sleep EEG by Using Independent Component Analysis," *Annals of Biomedical Engineering,* vol. 36, pp. 467-475, 2008.

[16] Z. Dan-hua, T. Ji-jun, and C. Yu-quan, "An ICA-based method for automatic eye blink artifact correction in multi-channel EEG," *Proceedings of the 5th International Conference on Information Technology and Application in Biomedicine*, Shenzhen, China, pp. 338-341, 2008.

[17] Y. Li, Z. Ma, and W. Lu, "Automatic removal of the eye blink artifact from EEG using an ICA-based template matching approach," *Physiol Meas,* vol. 27, pp. 425-436, 2006.

[18] N. P. Castellanos and V. A. Makarov, "Recovering EEG brain signals: Artifact suppression with wavelet enhanced independent component analysis," *Journal of Neuroscience Methods,* vol. 158, pp. 300-312, 2006.

[19] N. Mammone, F. La Foresta, and F. C. Morabito, "Automatic Artifact Rejection From Multichannel Scalp EEG by Wavelet ICA," *Sensors Journal, IEEE,* vol. 12, pp. 533-542, 2012.

[20] C. Guerrero-Mosquera, V. Navia, x, and A. zquez, "Automatic removal of ocular artefacts using adaptive filtering and independent component analysis for electroencephalogram data," *Signal Processing, IET,* vol. 6, pp. 99-106, 2012.

[21] A. Kachenoura, L. Albera, L. Senhadji, and P. Comon, "Ica: a potential tool for bci systems," *Signal Processing Magazine, IEEE,* vol. 25, pp. 57-68, 2008.

[22] J. K. Aapo Hyvärinen, Erkki Oja, "Independent Component Analysis," *John Wiley & Sons,* New York, 2001.

[23] S. Lisha, L. Ying, and P. J. Beadle, "Independent component analysis of EEG signals," *Proceedings of IEEE International Workshop on VLSI Design and Video Technology*, Suzhou, China, pp. 219-222, 2005.

[24] D. Djuwari, D. K. Kumar, and M. Palaniswami, "Limitations of ICA for Artefact Removal," *Proceedings of IEEE 27th Annual International Conference on Engineering in Medicine and Biology Society*, Shanghai, China, pp. 4685-4688, 2005.

[25] T.-W. Lee, M. Girolami, and T. J. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources," *Neural Comput.*, vol. 11, pp. 417-441, 1999.

[26] T.-P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. J. Sejnowski, "Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects," *Clinical Neurophysiology*, vol. 111, pp. 1745-1758, 2000.

[27] D. M. J. Tax and R. P. W. Duin, "Support vector domain description," *Pattern Recognition Letters*, vol. 20, pp. 1191-1199, 1999.

[28] R. P. W. Duin and D. M. J. Tax, "Outliers and Data Descriptions," *In Proceedings of the Seventh Annual Conference of the Advanced School for Computing and Imaging (ASCI)*, 2001.

[29] M. A. Hearst, S. T. Dumais, E. Osman, J. Platt, and B. Scholkopf, "Support vector machines," *Intelligent Systems and their Applications, IEEE*, vol. 13, pp. 18-28, 1998.

[30] D. M. J. Tax and R. P. W. Duin, "Support Vector Data Description," *Mach. Learn.*, vol. 54, pp. 45-66, 2004.

[31] K. V. Mardia, "Measures of multivariate skewness and kurtosis with applications," *Oxford University Press*, 1970.

[32] A. J. Viera and J. M. Garrett, "Understanding interobserver agreement: the kappa statistic," *Family medicine*, vol. 37, pp. 360-363, 2005.

[33] K. J. Berry and P. W. Mielke, "A Generalization of Cohen's Kappa Agreement Measure to Interval Measurement and Multiple Raters," *Educational and Psychological Measurement*, vol. 48, pp. 921-933, 1988.

[34] F. R. Mantini D, Romani GL, Pizzella V., "Improving MEG source localizations: An automated method for complete artifact removal based on independent component analysis," *NeuroImage*, vol. 40, pp. 160-173, 2007.

[35] M. Fabiani, Gratton, G., Karis, D., Donchin, E., " Definition, identification, and reliability of measurement of the P300 component of the event-related brain potential.," *Adv. Psychophysiol.*, vol. 2, pp. 1-78, 1987.

[36] 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, 1988.