

國立交通大學

資訊工程學系

碩士論文

運動想像腦電波之連續式判別



Continuous Discrimination of EEG Recordings
During Motor Imagery

研究生：陳怡岑

指導教授：陳永昇 博士

中華民國九十四年八月



Continuous Discrimination of EEG Recordings During Motor Imagery

A dissertation presented

by

Yi-Tsen Chen

to

Department of Computer Science
and Information Engineering

in partial fulfillment of the requirements

for the degree of

Master of Science

in the subject of

Computer Science and Information Engineering

National Chiao Tung University

Hsinchu, Taiwan

2005

Continuous Discrimination of EEG Recordings During Motor Imagery

Copyright © 2005

by

Yi-Tsen Chen



國立交通大學

論文口試委員會審定書

本校 資訊工程學系 碩士班 陳怡岑 君

所提論文:

運動想像腦電波之連續式辨別

Continuous Discrimination of EEG Recordings

During Motor Imagery

合於碩士資格水準、業經本委員會評審認可。

口試委員：

林倚婷 _____

陳麗芬 _____

陳永昇 _____

指導教授：

陳永昇 _____

系主任：

張明峰

中華民國九十四年七月十四日




學生：陳怡岑

指導教授：陳永昇

國立交通大學資訊工程研究所碩士班

摘 要



腦機介面是一種利用腦電波訊息作為人類與機器溝通的一種新媒介，隨著腦機介面近年來技術的發展，使得所謂的以意念控制機器的想法逐漸有了實現的空間。此技術若能開發成功，除了對於那些患有神經肌肉損傷的病患將有莫大的幫助外，相信也可以廣泛應用在工業控制上。我們可以藉由腦電波儀取得一個受試者在執行某些心智動作時所引發出的腦電波，經由精密分析，進一步地將腦電波轉換為機器控制指令。這樣的溝通系統可以分成同步與非同步兩大類，分類的依據在於受試者執行心智動作的時間點是否決定於自我意願，一個非同步的腦機介面意味著使用者可以在任何時間點下達命令；相反地，一個同步的腦機介面則意味著使用者必須在特定的時間點才可下達命令。目前以非同步的腦機介面較具有實用價值。

在此論文中，受試者將藉由左右手動作的想像來控制系統，我們的問題在於如何較精確且穩定地分辨受試者在某個時間點所下達的命令（或心智狀態）。為了建立一個較穩定且具有較精確度的非同步腦機介面，我們需要藉由訊號處理與樣型辨識的技術，包含了訊號前處理、特徵擷取、特徵選取、與辨認。細部來說，收到的腦電波必須先經由濾波器濾除雜訊和去除生理雜訊，例如眼動。之後，我們會套用時頻分析(time-frequency analysis)來分析腦電波，由於這些時頻特徵元素數量龐大，因此，我們還必須應用特徵選取的技術，包含有 t 測試(t-test)和向前特徵選取法(forward feature selection)來降低特徵點的維度和選取出具有分辨力的特徵點。最後，我們採用單群分辨法(one-class classification)來分辨這些特徵點。

我們分析的資料是來自腦機介面國際型比賽 (BCI competition III) 所提供的資料。在這三筆資料中，我們取出C3、Cz、C4三個位置的訊號加以分析，經由向前特徵選取法處理後，僅取兩個特徵點來分辨受試者的命令。我們的分析結果得到 (1) 左 (右) 邊動作想像和其休息狀態FN和FP平均為20%和30%。(2) 左 (右) 邊動作想像和非左 (右) 邊動作想像FN和FP平均為30%和25%。



Abstract

Brain-computer interface (BCI) provides a communication channel for patients with severe neuromuscular disorders to signal their intentions directly with their brain activities, instead of the normal output pathways between the brain and muscles. When a subject is performing specific tasks, the electroencephalographic (EEG) signals induced by her/his neuronal activities are recorded, analyzed, and translated to the corresponding commands for computers or other devices. A BCI system can be synchronous or asynchronous depending on whether the task is cue-triggered or self-paced. Asynchronous BCI systems are more practical yet more complicated due to the requirement of continuous analysis of ongoing EEG without any timing information about the mental status of the subject.

Toward building an asynchronous BCI system, we develop signal preprocessing and classification techniques, including signal preprocessing, feature extraction, feature selection, and classification, that can be used to continuously discriminate between EEG recordings when the subject is resting or performing left-hand/right-hand imagery tasks. The acquired EEG signals are first filtered for artifact removal. Then we use Morlet wavelet to extract time-frequency components. During the training stage, these abundant components are examined through t statistic and forward feature selection and the components with large discernment capability can be determined. During the classification stage, the EEG signals go through the same preprocessing procedure and the discriminative wavelet components are calculated. Then, two one-class classifiers are applied to discriminate the resting state from the left-hand motor imagery and from the right-hand motor imagery, respectively. The one-class classifier focuses only on the feature distribution of one of the motor imagery task on which the subject concentrates. In this way, we do not need to model the widespreading distribution of the resting state, which may comprise slight but fickle mental task.

We obtained three datasets from the website of the BCI competition III. From these EEG recordings, three channels (C3, C4, and CZ) are employed and two most discriminative wavelet components are selected by using the proposed techniques. According to our experiments, the false negative (FN) and false positive (FP) recognition rates for both left- and right- hand motor imagery tasks are about $\approx 20\%$ and $\approx 35\%$, respectively. We

also applied the proposed techniques to discriminate the left-hand (right-hand) motor imagery class from the non-left-hand (non-right-hand) motor imagery class which comprises of resting and right-hand (left-hand) motor imagery EEG data. In this case, the FN and FP recognition rates are about $\approx 30\%$ and $\approx 30\%$, respectively. These experiments clearly demonstrate the stability and accuracy of the proposed techniques.



誌謝

本篇論文得以完成，首先要感謝多年來苦心栽培我的雙親，有他們的期許與支持，我才得以完成此碩士學位。再則，感謝我的指導教授陳永昇老師以及陳麗芬老師，兩年來在研究上的引領與指導，使我慢慢地具備解決問題的能力；除了在研究方面提供不少寶貴意見外，在待人接物上更給了我不少良性的典範與教誨，充實了兩年來每一天的生活。最後，也感謝實驗室的同儕們，對於我駑鈍之資不加嫌棄；盡心盡力地解說我所發問的問題；在我需要幫忙時，不吝的支援；在我低落沮喪時，綿綿的鼓勵；感謝這些人們兩年來的一切，他們是這份論文可以順利完成最主要的推力，同時也是豐富我兩年時光的主因，壯碩我這棵小樹的主要泉源。





Contents

List of Figures	ix
List of Tables	xi
1 Introduction	1
1.1 Motivation	2
1.2 Background	3
1.2.1 The human brain	3
1.2.2 Electroencephalography	4
1.2.3 Brain signal	6
1.3 Brain-computer interface	14
1.3.1 Components of brain-computer interface systems	16
1.3.2 Categorization of brain-computer interface systems	19
1.4 Thesis scope	22
2 Survey of brain-computer interface systems	23
2.1 Brain-computer interface systems today	24
2.1.1 Farwell P300-based brain-computer interface	26
2.1.2 Graz ERD/ERS-based brain-computer interface	28
2.1.3 Wolpaw and MacFrand μ -rhythm-based brain-computer interface	31
2.1.4 Birbaumer SCP-based brain-computer interface	32
2.1.5 Middendorf SSVEP-based brain-computer interface	33
2.2 Comparisons of brain-computer interface systems	36
2.3 Limitations of brain-computer interface systems	36
2.3.1 Habituation	37
2.3.2 Interference and distraction	38
2.3.3 Impaired visual system	38
2.3.4 Instability of EEG signals	38
2.3.5 Variability of a specific brain pattern	39

3	Material	41
3.1	Experiment paradigm	42
3.2	Datasets	42
4	Single-trial EEG classification	45
4.1	Signal preprocessing	46
4.1.1	Artifact removal	48
4.1.2	Bandpass filter	48
4.1.3	Reference-free spatial filter	49
4.2	Feature extraction	51
4.2.1	Event-related desynchronization and Event-related synchronization	53
4.2.2	Common spatial filter	54
4.2.3	Haar wavelet transformation	56
4.2.4	Morlet wavelet transformation	57
4.3	Feature selection	58
4.3.1	t statistic	59
4.3.2	Forward feature selection	62
4.4	Classification	63
4.4.1	Support vector classifier	64
4.5	Experiment results	67
4.5.1	Summary	68
5	Continuous EEG classification	71
5.1	One-class classification	74
5.1.1	Why one-class classification	76
5.2	Classifier	77
5.2.1	Support vector classifier data description	77
5.2.2	k-nearest neighbor data description	78
5.2.3	Self-organizing map data description	80
5.3	Classifier design	82
5.4	Implementation issues	83
5.4.1	Simulation	83
5.4.2	Sliding window	86
5.4.3	Unbalance features	87
5.4.4	Constraints	87
5.4.5	Five kinds of classification	89
5.4.6	Summary	90
5.5	Experiment results	90
5.5.1	Summary	93
6	Conclusions	105





List of Figures

1.1	BCI	3
1.2	Human brain structure	4
1.3	The mapped functional areas of the human brain.	5
1.4	The 10-20 system of electrodeplacement.	6
1.5	Common brain rhythms	8
1.6	A contralateral localized mu ERD and a occipital localized alpha rhythm ERS.	9
1.7	An example of SCP.	10
1.8	An example of P300.	11
1.9	The flowchart of ERP and ERD/ERS	13
1.10	ERD/ERS during imagination of hand movement.	14
1.11	ERD/ERS in mu and beta bands for a motor imagery experiments.	15
1.12	Flowchart of BCI.	17
2.1	An overview of BCI today.	27
2.2	Show the changes of ERD/ERS in a right hand movement.	29
2.3	The changes of beta ERD/ERS in a right hand movement imagery experiment.	30
2.4	Wolpaw mu rhythm based BCI.	32
2.5	A SSVEP-based BCI in flight simulator.	35
3.1	Experiment paradigm.	42
4.1	Flowchart of synchronous analysis.	47
4.2	EyeBlink	49
4.3	EyeMovement	50
4.4	EKG	51
4.5	Montage	52
4.6	Analyzed electrodes.	53
4.7	Show ERD/ERS of data G01 for hand motor imagery tasks.	54
4.8	The relationship of time and scale resolution in Haar wavelet.	57
4.9	Haar wavelet	58
4.10	The averaging Morlet wavelet coefficients of data G01	59

4.11	An example of 2 groups' distribution.	60
4.12	Show the relationship between discriminability, mean, and variance of two groups.	60
4.13	Show the t-statistic of Morlet wavelet coefficients of data G01	61
4.14	An example of forward feature selection.	63
4.15	A linear support vector classifier	65
5.1	Flowchart of analyzing asynchronous BCI.	73
5.2	Threshold of one-class classification.	75
5.3	An example of one-class	76
5.4	A hypersphere around the target data.	78
5.5	One example of kNN with $k = 5$	79
5.6	Examples of lattice grids.	80
5.7	Examples of updating BMU.	81
5.8	One method of classifying asynchronous BCI mental tasks in the work, <i>Method_A</i>	84
5.9	The other method of classifying asynchronous BCI mental tasks in the work, <i>Method_B</i>	85
5.10	The sliding windows of a signal.	86
5.11	The inter-trial variation of cue-onset latency time.	88
5.12	Experiment flowchart.	91
5.13	Time-frequency map of G02.	92
5.14	Time-frequency map of G03.	93
5.15	Time-frequency map of G04.	94
5.16	The ROC curves of using different numbers of feature of Analysis1 for G02.	95
5.17	The ROC curves of using different numbers of feature of Analysis1 for G03.	96
5.18	The ROC curves of using different numbers of feature of Analysis1 for G04.	97
5.19	The ROC curves of the three classifiers when analyzing G02-L and G02-R of analysis 2.	97
5.20	The ROC curves of the three classifiers when analyzing G03-L and G03-R of analysis 2.	98
5.21	The ROC curves of the three classifiers when analyzing G04-L and G04-R of analysis 2.	98
5.22	The ROC curves of the three classifiers when analyzing G02-L and G02-R of Analysis3.	99
5.23	The ROC curves of the three classifiers when analyzing G03-L and G03-R of Analysis3.	99
5.24	The ROC curves of the three classifiers when analyzing G04-L and G04-R of Analysis3.	100
5.25	A plot of illustrating F_{LR} and F_{RL}	102

List of Tables

1.1	The comparisons of MEG and EEG.	5
1.2	Common brain rhythms in different frequency bands.	7
2.1	The comparisons of BCIs.	37
4.1	The four situations of classifying an object in two-class classification. . . .	64
4.2	The accuracy of synchronouse BCI classification by suing ERD/ERS. . . .	68
4.3	The accuracy of synchronouse BCI classification by suing CSP.	68
4.4	The accuracy of synchronouse BCI classification by suing Haar wavelet transformation.	69
4.5	The accuracy of synchronouse BCI classification by suing Morlet.	70
5.1	The relationship of the(<i>ResultA</i> , <i>ResultB</i>) and final assignment <i>ResultC</i> in <i>MethodB</i> when classifying an asynchronous BCI.	83
5.2	Analysis4. In this figure, LN_L represents <i>number of results assigned to L</i> , N_R the number of results assigned to R, LN_r the number of results assigned to resting	100
5.3	Results of Analysis2. The upper is the results of our method and the bottom the method proposed by Graz group in 2004.	101
5.4	Analy3 Result	102
5.5	Mean of Analysis4. In this table, LN_L represents <i>number of results assigned to L</i> , N_R the number of results assigned to R, LN_r the number of results assigned to resting	102
5.6	Analysis5. In this figure, LN_L represents <i>number of results assigned to L</i> , N_R the number of results assigned to R, LN_r the number of results assigned to resting	103



Chapter 1

Introduction



1.1 Motivation

After the industry evolution, machines are invented and utilized in many fields. It seems no doubt that the computer is the most significant machine in this century. In the recent decades, computers become essential for human beings. Many researchers devote themselves to develop the best interface between human and computers to assist the manipulation. We have witnessed the magical evolution of the human-computer interface designed in touch, vision, voice, and a combination of these in the last two decades.

Today, the typical interfaces between human and computer are *graphic user interface (GUI)*, *voice command and controller*, and *touch control*; however, they have some limitations. For the GUI and the touched-control, the user must be free of motor-disable disease. For the voice-control, the user must have the ability to speak. Among the three interfaces, the voice-control interface is the most attractive one to communicate with computers or other machines, yet it still can not supply to *Amyotrophic Lateral Sclerosis (ALS)* patients. Obviously all the above interfaces are not suitable to the ALS. For the ALS, thoughts might be the last one for them to communicate the world.

Over the past decades people began interested in achieving a direct communication between human and computers. They wish to design an interface to provide a new communication channel for everyone even for the patients suffering from severe neuromuscular disorders who are forced to accept a reduced quality of life dependent on other individuals. Although many aids have been created to liberate these individuals, there still exists some limitations for those with sever disabilities to utilize these aids. Therefore, the interface between human and human brain are emerged. The interface is called *Brain-Computer Interface (BCI)*, consisting knowledge in many fields such as neuroscience, psychology, signal processing, machine learning.

BCI was first proposed in the 1970's by the United States Department of Defense in order to assist pilots in their aircraft. Unfortunately the speed of computers those day was too slow and the memory was too small to deal with the huge brain signals, so the idea was failed. Today these advanced hardware equipment makes it possible to achieve this idea by developing BCI systems. Such an interface was defined as " A brain-computer interface is a communication system that does not depend on the brain normal output pathways of



Figure 1.1: BCI

peripheral nerves and muscles” in the first international meeting devoted to BCI research held in June 1999 [43]. It takes our brain signals as its inputs, recognize what is the user’s intention and finally output the transition command to the system to trigger the application, such as Figure 1.1.

In short, in recent decades, BCI systems attract greater interests for the following reasons. First, it is the most potential and the most attractive interface. Second, many researchers are mainly fostered by the will to give the patients with different severe motor impairments or ALS an new communication channel.

In section 1.2, we first illustrate a brief overview of human brain structures and introduce the technique of monitoring brain activation, EEG. Then we describe some specific brain patterns. In section 1.3, we give the details of the components and the categorization of a BCI system. In the next chapter, we will show some examples of existing BCI systems today.

1.2 Background

1.2.1 The human brain

The human brain has four structures: *cerebral cortex*, *cerebellum*, *brain stem*, *hypothalamus* and *thalamus*, but what we concern is the cerebral cortex which is a 2-4 mm thick sheet of the gray tissue, to where the activation which we interest occur close. The cerebral cortex can be divided into two *hemispheres* by the *longitudinal fissure*. Each of the hemispheres can then be divided into four lobes called *frontal*, *parietal*, *occipital* and *temporal* lobes by two deep grooves, *Rolandic fissure* and *Sylvian fissure*.

At the beginning of the twentieth century, the famous German neuroanatomist Korbinian Bordmann divided the cortex into different zones depending on the functions of the

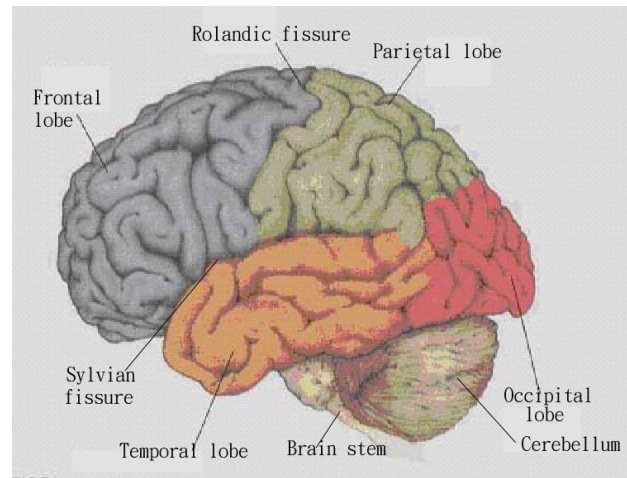


Figure 1.2: Human brain seen from the left side. Some of the important structure are indicated

zones. In general, our cortical areas fall into four main categories: primary sensory areas, motor areas, secondary sensory areas, and association areas. The *primary sensory areas* are those that are first receive stimuli from the external environments. For example, *primary somatosensory cortex*, also called S1, which is located posterior to the Rolandic fissure can receives tactile stimuli from the skin. Motor area is concerned with controlling voluntary movements, by receiving signals from thalamic nuclei and sending outputs to motor control neurons in the brain stem and spinal cord. For instance, the *primary motor cortex M1* which is anterior to the Rolandic fissure can handle the muscular activity and each part of the M1 is involved in the movement of a specific part of the body.

1.2.2 Electroencephalography

There exists various non-invasive techniques to monitor the brain activity such as functional Magnetic Resonance Imaging (fMRI) [14], magnetoencephalography (MEG) [12], Positron Emission Tomography (PET), and electroencephalography (EEG) [34]. Among these methods, MEG and EEG are suitable for BCI systems because they can give the instantaneous continuous recording of brain activity. The comparisons between MEG and EEG are illustrated in table 1.1. Because EEG is more portable, cheaper and doesn't need

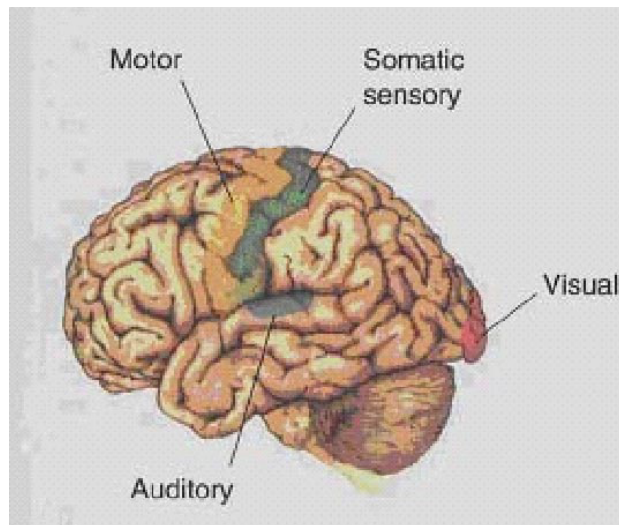


Figure 1.3: Human brain seen from the left side. The mapped functional areas of the brain are indicated

to be measured in a shielded room, it is more attractive to be used in BCI than MEG in spite of EEG's low spatial resolution. Today almost all of BCIs are EEG-based.

The spatial resolution of EEG depends on the number of electrodes, which are placed based on the *international 10-20 system* [19]. "The 10-20 system is the relationship between the location of an electrode and the underlying area of cerebral cortex". The "10" and "20" means the 10% and 20% interelectrode distance. How the electrodes are placed in 10-20system is shown in Figure 1.4.

When recording EEG signals, we will explore a position with zero potential to be the

Table 1.1: The comparisons of MEG and EEG.

	MEG	EEG
measurement	magnetic field	electrical potential
sensitive source orientation	tangential	radial
portability	no	yes
spatial resolution	high	low
volume conductor effect	less	more

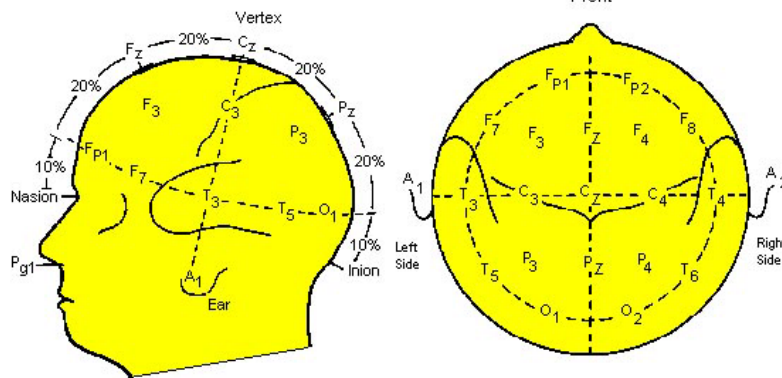


Figure 1.4: The 10-20 System of Electrode Placement. The "10" and "20" means the 10% and 20% interelectrode distance. The 'F', 'C', 'P', 'O', 'T' represent with respect to frontal, central, parietal, occipital, temporal lobe. The odd number is placed in the left side and the even number the right side [19]

reference. Unfortunately, on the scalp positions, there is no electrode with zero potential. Therefore, we usually take the right ear as the reference known as monopolar recording. This means that the measured EEG signals are reference-dependent. To convert reference-dependent signals to reference-free signals, we can apply reference-independent derivation to the reference-dependent signals such as *Common average reference (CAR)*, *Bipolar reference*, *Laplacian reference*. We will give more details about the three approaches in Chapter 4.

1.2.3 Brain signal

Neurons can send action potentials, postsynaptic potentials to other neurons by through the dendrites and axons. An action potential is initiated when the voltage is larger than the threshold about -40 mV. In fact, what EEG measures are the potentials integrating from the thousands or billions synchronously activated neurons but not the potential of one neuron for it is too small to be detected. The recorded EEGs include spontaneous electrical activity of the cerebral cortex and the cortical response evoked by external or internal events, namely event-related potential. In general, it is believed that what EEGs

Table 1.2: Common brain rhythms in different frequency bands.

Band.	Frequency [Hz]
Delta	< 3.5
Theta	4-7.5
Alpha	8-13
Beta	>13

measures is the summation of excitatory and inhibitory postsynaptic potentials of cerebral cortex with some contribution of granular and glia cell activity [27].

There exists three typical processing techniques of the measured EEGs, which are commonly used in BCI systems:

- Brain rhythm
- Event-related potential(ERP)
- Event-related desynchronization(ERD) and event-related synchronization(ERS)

Brain rhythm and ERD/ERS are based on analysis of spontaneous brain activity while ERP on cortical responses to events.

Brain Rhythm

Human brain has various different rhythmic activities according to the different frequency range. Typical rhythmic activities include delta, theta, alpha, beta and gamma rhythms. We indicate the relationship between the rhythms and frequency ranges in table 1.2. In Figure 1.5, an example for these brain rhythms [25] is illustrated.

In this thesis, we only concern the alpha (μ) and beta rhythms because they will react to the imagination of hand movement which is used as the predefined mental tasks in this work. The alpha, mu and beta rhythms are described as follows:

- 1. Alpha rhythm** Alpha rhythm is between 8-13 Hz with amplitude mostly below 50uV for adults. Alpha rhythm is considered as the primary rhythm of normal adult brain.

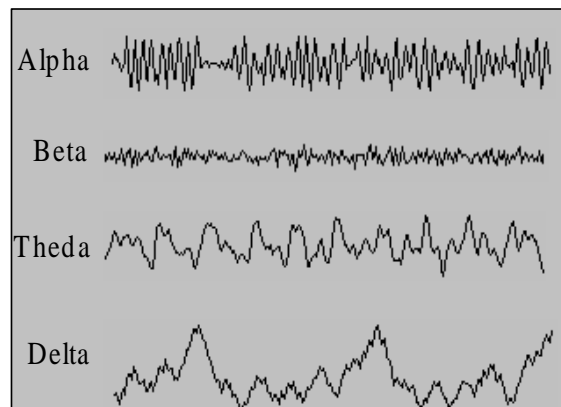


Figure 1.5: An example of four brain rhythms [21].

It is frequently seen in occipital areas when people is awake , close their eyes and under conditions of physical relaxation [45].

- 2. Mu rhythm** Mu rhythm is in the alpha frequency band. The main difference between mu and alpha rhythm is that the mu rhythm is usually blocked or attenuate with contralateral movements or the thought of a movement, but not react to eye open or closing. For example, when executing a movement, the central mu rhythm is desynchronized; however, the occipital alpha rhythm is synchronized. Figure 1.6 shows a contralateral localized mu event-related desynchronization (ERD) and a occipital localized alpha rhythm event-related synchronization (ERS) in a right hand movement task. In the next section we will give an overview of event-related desynchronization/event-related synchronization ERD/ERS.

During a movement, the elicited contralateral mu rhythm is interpreted as an unspecific presetting, priming of neurons in motor areas. Besides motor-relative tasks, the mu rhythm can also be induced by a flicker stimulation [41], a reading task [37].

- 3. Beta rhythm** The beta rhythm is defined as a frequency of above 14Hz and below 30Hz, with amplitude usually being seldom larger than 30uV and irregular. Beta rhythm is also the background feature of normal adult brain. It can be found over the frontal and cerebral regions. The central beta rhythm can be elicited when people perform a

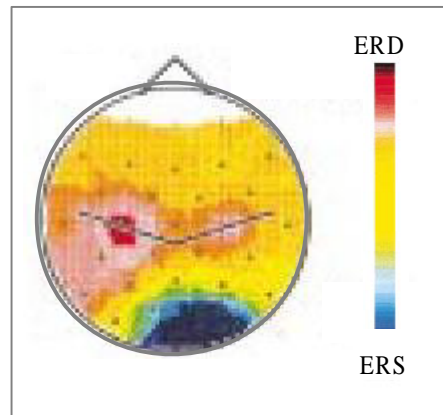


Figure 1.6: The contralateral localized mu ERD and a occipital localized alpha rhythm ERS in a right hand movement task is illustrated. The topography is at about the movement execution moment [36].

voluntary movements.

Event-related potentials(ERPs)

Several kinds of internally or externally paced events will result in time-locked and phase-locked brain signals. Almost all of these kinds evoked activities have a more or less fixed time-delay to the stimulus. These time-locked and phase-locked brain signals are called *event-related potentials (ERP)* or *evoked potentials (EP)*. ERP can be viewed as potential changes of the neurons when our brain deal with mental tasks.

Usually the brain signals of a mental task is smaller than the ongoing brain signals, thus concealed in the irregular and noisy ongoing brain signals. In order to extract the ERPs, *synchronous averaging* are performed, implying we have to execute the same mental tasks more than once. Due to the non-time-locked and non-phase-locked of the noise, after applying synchronous averaging, most of the noise will be eliminated, therefore enhancing the signal-to-noise ratio and obtaining the time-locked and phase-locked signals, ERP.

Many various ERPs have been proposed today such as *slow cortical potential (SCP)*, *P300*, *visual evoked potential (VEP)*, and *steady-state visual evoked potential (SSVEP)*. The way to label ERPS is often the latency and the electrical polarity or the type of given stimuli. An example of the former is P300, representing a *positive peak with latency 300ms*

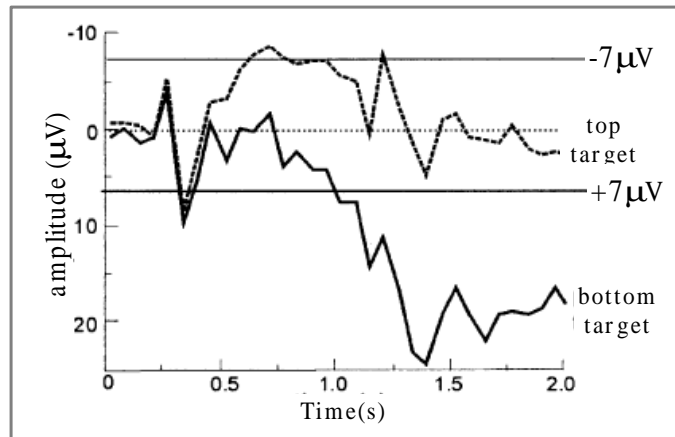


Figure 1.7: An example of SCP when the user performs a task that move a cursor toward a target at the bottom or the top [57].

posterior to the stimulus onset. For the latter, the representative ERP is visual event-related potential (VEP), elicited by a visual stimulus. Here, the details of these four ERPs, related to some existing BCI systems, are discussed.

- SCP:

SCPs are the slow voltage changes of the brain cortex, with a 0.5-10.0s potential shifts. They are settled in the frequency range below 1-2Hz. SCPs can be divided into two types, negative and positive. Negative SCPs represent the mobilization or readiness while positive SCPs represent ongoing cognitive and inhibition of neuronal activity [18]. In 1982, Lutzenberger gave an example that the subject could solve arithmetic problems faster after producing a negative SCP [18]. Figure 1.7 shows an example of the SCP when the user performs a task moving the cursor toward a target at the bottom or the top. As we see, bottom target often induce positive SCPs while top target negative SCPs. In this figure, we can also find that the SCP persist many seconds.

- P300:

Infrequent, particular or oddball stimuli in auditory, visual or somatosensory will evoke the P300. P300 is a positive peak reaching the maximum of about 300ms after

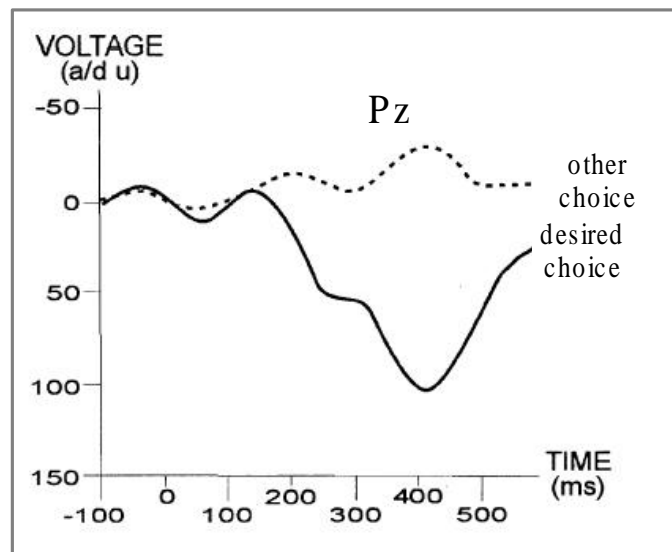


Figure 1.8: An example of P300 in a oddball visual stimuli experiment. [57]

the stimulus over the parietal areas. P300 is commonly utilized in a spelling BCI system [15, 35]. Because the P300 is a naive response to an infrequent stimuli, the user requires no user training to produce a P300 pattern.

- VEP:

Visual evoked potential (VEP) is induced when the user's eyes are stimulated by looking at a test pattern which often is a flashing pattern. To measure VEPs, the recording electrodes are placed over the visual cortex.

- SSVEP:

The SSVEP is a brain potential changes elicited by a brief visual stimulus modulated at a specific frequency. The SSVEP is characterized as an increase in EEG activity at the stimulus frequency.

Event-related desynchronization(ERD) and event-related synchronization(ERS)

Since Berger(1930), we have known some of our brain signals could block or desynchronize the ongoing brain activity. Because these types of signals are time-locked but not

phase-locked, we cannot use the simple linear method such as averaging to extract them, but use frequency analysis.

In 1977, Pfurscheller proposed the concepts of *event-related desynchronization(ERD)* and *event-related synchronization(ERS)*. He thought these types of brain signals represent specific changes of the ongoing brain activity and might consist, in general, of decrease or of increase of the power in a specific frequency bands. He defined the former as ERD and the latter as ERS. Different from the ERP, which can be thought as a series of changes of the post-synaptic response, the ERD/ERS can be considered as the controlling of brain oscillations.

The ERD/ERS is a percentage value to the power of a predefined interval signals, *reference* or baseline period. Usually the reference period is an interval before the target is performed. The typical algorithm of calculating ERD/ERS, (*ERD inter-trial variance*), is as follows [36]:

- 1 Specify the frequency bands we interest and afterward apply bandpass filtering.
- 2 Calculate the mean of the filtered data over all trials.
- 3 Subtract the mean from the filtered data in step1.
- 4 Square the amplitude samples from the previous step over all trials.
- 5 Average over time samples from the step4.
- 6 Obtain ERD/ERS by calculating the percentage relative to the power of the reference interval.

These steps can be found in Figure 1.9, which indicates the algorithms of calculating ERP, ERD/ERS. In this figure, there are two approaches we need to take a notice, band-power ERD and ERD inter-trial variance. Both are used to calculate ERD/ERS. The main difference between ERD and ERD inter-trial variance is that the latter subtract the mean of phase-locked signals from the original signals. The motive of performing subtraction is " a phase-locked power increase due to the ERP can mask the non-phase-locked power decrease(ERD) when the classical band-power method is used." In Figure 2.2, we present an example of calculating ERD/ERS.

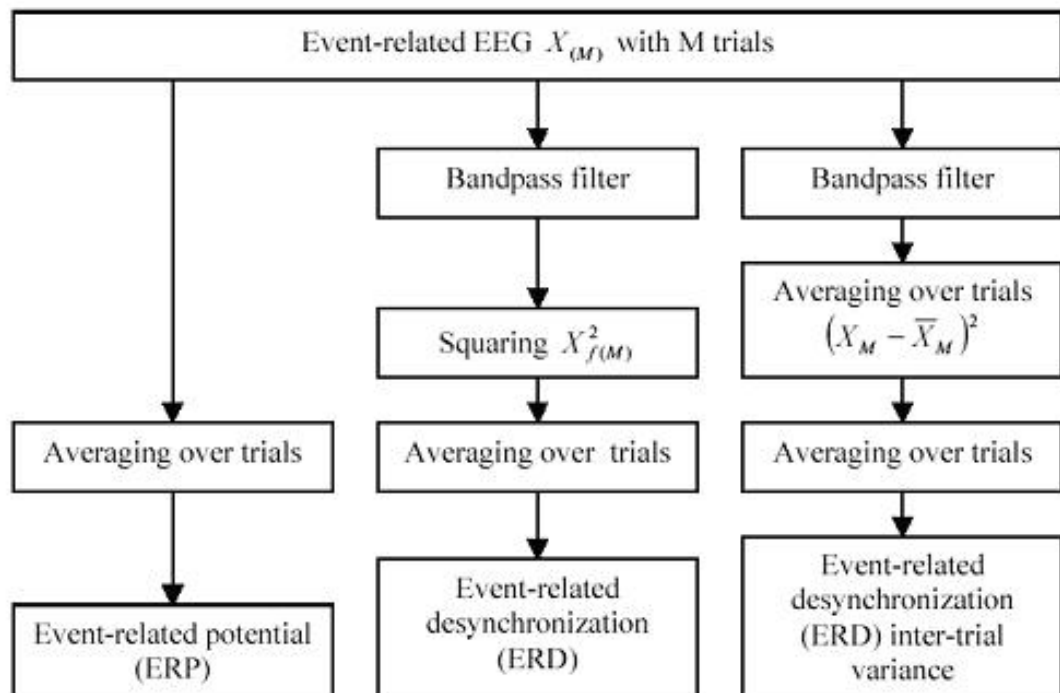


Figure 1.9: The flowchart ERP and ERD/ERS

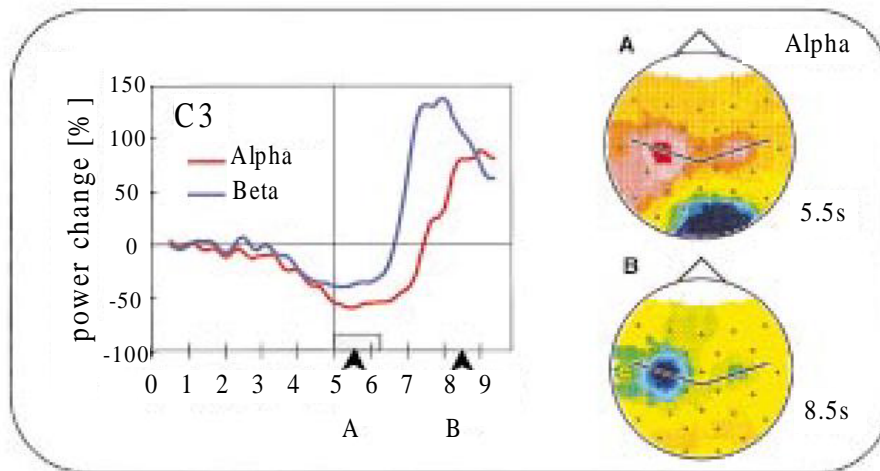


Figure 1.10: An example of ERD/ERS. [36]. The mental task is imagination of hand movement. The onset time is at 5s. In the left side two ERDs in alpha and beta band are illustrated. In the right side, two topographies are plotted with respect to at 5.5s and at 8.5s.

ERD/ERS can emerge when performing a movement or a thought of movements, perceptual, judgement task. For example, Figure 1.11 shows the ERD/ERS in mu bands. The data was measured over the position C3, Cz, C4 for performing hand motor imagery. The contralateral ERD can be found easily in this figure.

1.3 Brain-computer interface

As we mentioned, BCI was defined as a communication system not depending on the brain normal output pathways of peripheral nerves and muscles in the first international meeting in 1999. Similar to any communication system, a BCI has inputs, outputs, transformation elements, and a protocol that determines its operation. A user controls a device, such as a computer, by performing some predefined mental tasks. The association between mental tasks and machine action is determined by the univocal identification of the system after a series of processing.

Today, the BCI researchers design a BCI system based on two approaches, *pattern recognition approach* and *operant conditioning approach*.

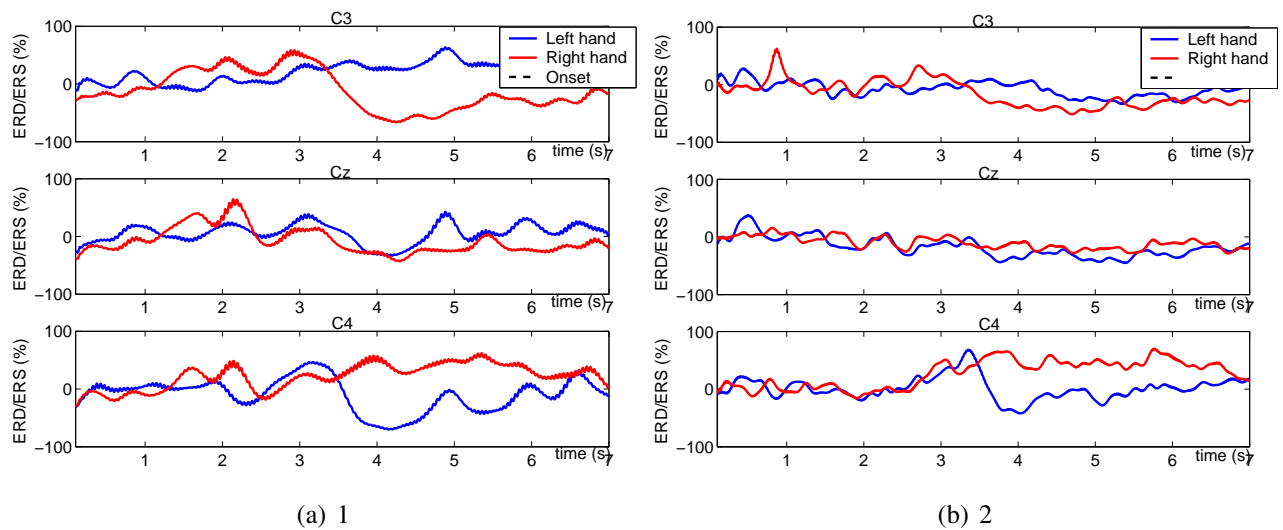


Figure 1.11: Show the ERD/ERS in mu and beta bands for a motor imagery experiments. The left plot and right plot are mu ERD/ERS and beta ERD/ERS respectively. Onset time is at 3.25s. After onset, the mu and beta contralateral ERD are elicited.

Pattern recognition approach:

This kind BCI system is based on classification with different cognitive mental tasks which are often predefined. Some brain hemispheric specialization studies have been used when choosing mental tasks, which suggest that the left hemisphere usually involves in verbal and other analytical functions and the right hemisphere in spatial and holistic processing. Commonly used mental tasks in current BCI systems include motor imagery, arithmetic, visual, spatial operations. In order to produce different and discriminable EEG patterns, the mental tasks should be carefully chosen to activate different parts of the brain. For example, the imagination of the left hand would activate the right motor cortex while imagination of the right hand the left motor cortex.

After determining mental tasks, a user will perform more than once these mental tasks to train a classifier used to recognize his wish. This classifier is built based on the technology of pattern recognition.

Operant conditioning approach:

The principle of operant conditioning approach based BCI systems is based on self-

regulation. In other words, we train a user to control the BCI systems by regulate his brain signals. Today, this kind of BCI systems is designed to self-regulate the brain rhythms or SCPs described in section 1.2.3. In order to achieve self-regulate these specific brain patterns, there are three important issues should be concerned [27].

- 1 Real-time feedback of the specific EEG activity.
- 2 Positive reinforcement of correct behavior.
- 3 Individual shaping schedule in which progressively more demanding tasks are rewarded.

It will takes a long time to train a user to achieve self-regulation, that is why the operant conditioning approach based BCI is not the main stream today. However, the recognition of a user's intentions in the system is more easier than that in a pattern-recognition BCI. Unlike the complex algorithms of training a classifier of the pattern-recognition-based BCIs, by searching for some specific changes such as amplitude, power in a narrow frequency band, or phase, the different mental tasks are distinguished in operant-conditioning-based systems.

The section focuses on giving an overview of a BCI system. First we will illustrate the BCI's components based on pattern-recognition approach. Afterwards, we will address the principles of categorization of BCI systems. In the next chapter some examples of the different BCIs to date are given.

1.3.1 Components of brain-computer interface systems

As we see in Figure 1.12, a BCI consists of several processes or procedures. These processes include six stages as follows:

- Measurement of EEG:

A typical EEG device consists of several components. These include electrode cap, EEG amplifier, computers and screens. In additions to the electrodes on the cap, we still need to place the electrodes around the eyes to measure EOG. In some situations,

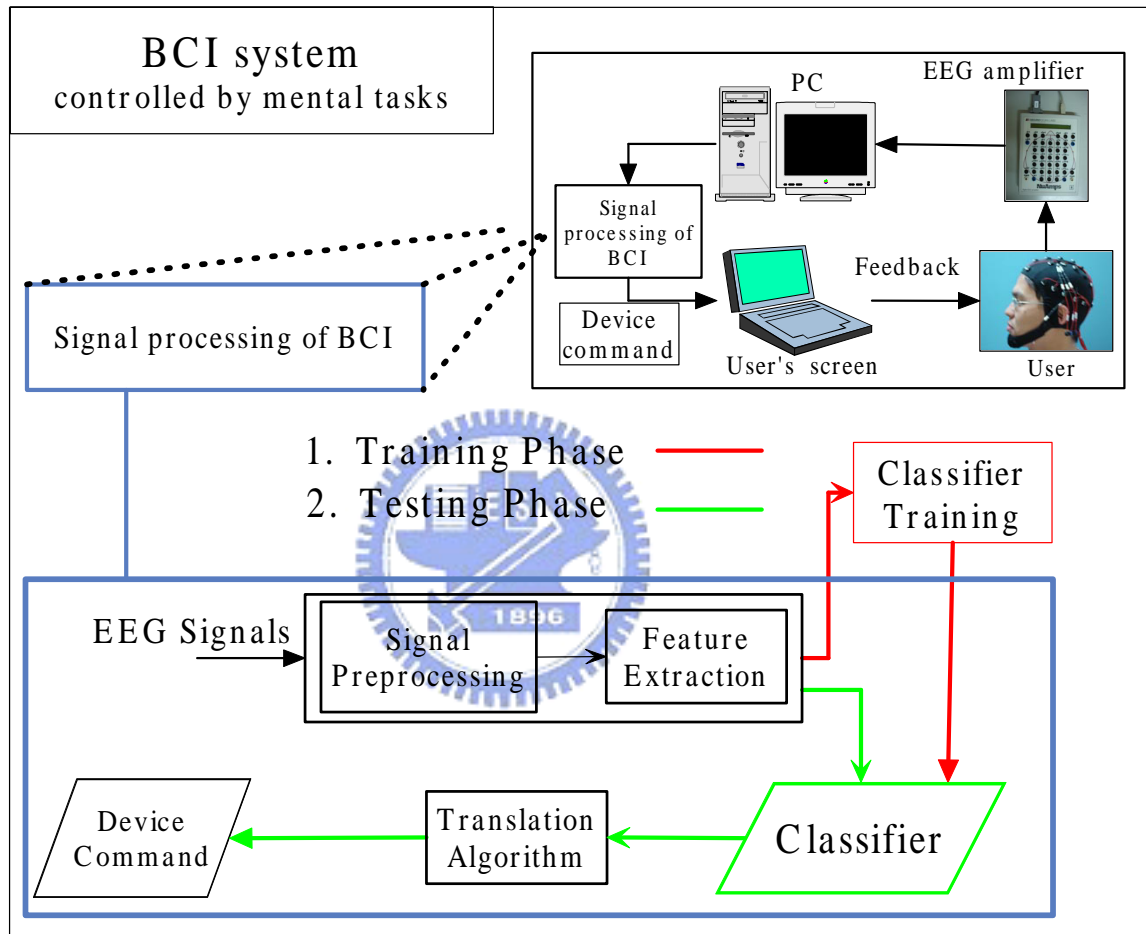


Figure 1.12: Flowchart of BCI.

we may also want to measure movement artifacts by means of some EMG channels. In preprocessing step, these additional electrodes will be used to perform artifact removal.

When measuring the brain signals, since the skull and the scalp cause smearing of the cortical signals, that is also called head volume conductor effect, the electrodes will be contaminated, thus containing relevant information. In order to reduce or correct the influence, in some BCI systems, the spatial filters as we will mention in Chapter 4 will be applied. These spatial filter can not only correct volume conductor effect but also make signals reference-free.

- Signal preprocessing:

The goal of the stage is to enhance the signal-to-noise ratio. Typical procedures include amplification, filtering, possible artifact removal. For the filtering, the bandpass filtering is usually applied. In addition, a notch filter is also used to suppress the 60 Hz power line interference. As for the artifact removal, almost all BCIs rule out the signals if the EOG or EMG is detected to be used or over a predefined threshold.

- Feature extraction:

In this stage, certain features are extracted from the preprocessed signals. ERP, ERD/ERS and brain rhythms are typically used features in a BCI system. Besides the above features, various feature extraction methods have been studied to extract more discriminative features, such as *common spatial filter*, *continuous wavelet transform*, *autoregress model (AR)* or *adaptive autoregress (AAR) model*, *power spectrum*. All the above methods can be found in BCI competition 2003 papers.

- Pattern classification:

The features extracted from feature extraction are fed to train a classifier. Many classification methods have been proposed in *pattern recognition* field. The classifier in a BCI can be anything from a simple linear model to a complex nonlinear or a machine learning models. In general, the BCI has two phases, *training phase* and *testing phase* which can be found in Figure 1.12. The training phase consists of a

repetitive process of cue-based mental tasks to train a classifier. In the testing phase, we use the classifier built in the training phase to recognize different mental tasks.

- Translation algorithm and device control

The goal of this step is to translate the classification output in previous step to an operator command. The command can be, e.g., a letter in a spelling system or a movement of a course on the user's screen or nothing to be performed when the classification is "resting" or "idle". The design of translation algorithm and device control depends on what applications the BCI want to provide with.

- Feedback and biofeedback

A feedback which make the user more easily adaptive to the system is a very important component for a BCI system. A feedback can indicate how well the asked mental activity was recognized by the system. When the system gives the feedback to a user, he will create a biofeedback which is the process that the user receives information about his biological state. By Biofeedback, the user can monitor his physiological states, shape his brain electrical behavior, and voluntary modification of his EEG response. Today, nearly all BCI systems provide a feedback to users. Among all feedback interfaces the cursor movement [13, 42, 56] is mostly utilized.

There is another issue we have to take a note. When we design a BCI system, we must to decide the experiment paradigm before all the steps we mentioned above. In the thesis, we are interested in the motor imagery experiment because the brain signals of motor related mental tasks or real motor movement have more detectable and more discriminative brain patterns. Besides, in the thesis, we only focus on the first five steps because these parts have direct influence on the performance of a BCI system.

1.3.2 Categorization of brain-computer interface systems

Invasive and non-invasive systems

We have mentioned the non-invasive brain monitor techniques in section 1.2.2. The BCI systems based on these techniques are non-invasive BCI systems and among

these techniques, EEG is the most suitable for a BCI system. The "non-invasive" means we don't have to directly record the brain activity by putting the electrodes into the brain, in which the user will be at medical risks. Therefore, the non-invasive BCI is less debatable. One of the disadvantages of the non-invasive BCI is the influence of the volume conductor effect, thus the quality of electrodes containing other noise overlapping the brain activity. Another disadvantage is that the signals we are measuring in an electrode are the activity of myriads of neurons from different areas which may not be what we are interested in [10].

On the contrary, for the invasive BCI systems, we have to put the electrodes into a user's skull to monitor the brain directly. This is the main drawback of invasive BCI systems. The other disadvantage of the invasive ones is that the quality of the signals decays over time. In order to correct the influence, usually new surgical interventions and implants are required. However, the signals in an invasive BCI are of a higher signal-to-noise ratio for the signals eliminate contamination by volume conductor effect, muscle artifact. Chaoin and Gaal [6, 7], Kennedy's group [39, 40] are of this type.

Due to the medical risks, present-day almost all of the BCI systems are non-invasive.

Pattern-recognition-based or operant-conditioning-based BCI systems

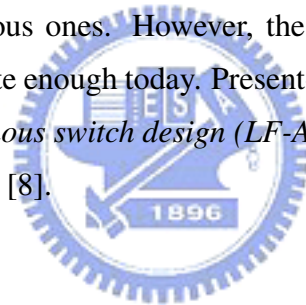
We have addressed concepts of pattern recognition approach and operant conditioning approach in the previous section. The pattern-recognition-based BCI is based on cognitive mental tasks while the operant-conditioning-based BCI is based on the self-regulation of the EEG response. BCI based on ERPs besides SCP, ERD/ERS are of the former type. Examples are Farwell and Donchin P300 based BCI, Graz ERD/ERS based BCI. For the latter, the representative work is Wolpaw μ -rhythm BCI [42] and Birbaumer [2].

Synchronous and asynchronous systems

In a synchronous BCI, the user is notified to perform a mental activity when a specific external cue is shown. That means this kind of system operates in a cue-paced mode

and has the information about the onset of the mental activity in advance. The analyses and classification of the brain signals in the system is limited to the predefined fixed time period. Besides, the system is active only during the predefined period as well. BCI systems based on evoked potentials and ERPs belong to this category, such as P300 [3, 11], SCP [4]. Besides EPRs, the BCI developed in Albany [55] and Graz [17] that analyzed the spontaneous EEG are also synchronous BCIs.

The BCI that a user can intend a mental activity whenever he wishes to perform such mental activity is an asynchronous BCI. In the asynchronous BCI, the brain signals are analyzed and classified continuously. We have to not only classify from the redefined mental tasks but also discriminate events from noise and nonevents such as resting or idling states. Such a BCI system is more flexible and attractive to be utilized in practice. Besides the above advantages, it also offer a rapider response time than synchronous ones. However, the classification in an asynchronous BCI system is not accurate enough today. Present-day this kind BCI systems include *low-frequency asynchronous switch design (LF-ASD)* [28], and Graz 2004 [51], *adaptive brain interface(ABI)* [8].



Online and offline BCI systems

The "online" means that the all processing of the BCI should be done in real-time or at least close to it. It is imperative when training a user to adapt to the BCI system because only in online mode, the system can provide feedbacks. The feedbacks make it possible for a user to monitor his physiological states and correct his behavior. For the offline BCI system, the aim is to evaluate a model to classify different mental tasks well. When training a classifier, the cross-validation test is often performed. Usually the model evaluated in the phase is then taken as the blueprint of the same kind online BCI.

1.4 Thesis scope

There are some problems making it hard to analyze EEGs. We list them as follows:

- 1 EEG signals are of very small amplitude and very sensitive to external noise.
 - 2 The sample dimension of EEGs is very high.
 - 3 User-generated artifacts may contaminate and smear the brain signals and even result in misleading conclusions regarding the real controlling skills of a user.
- The brain pattern may seriously vary over trials.

All of them hinder the analyses and make it difficult to obtain high recognition accuracy in a BCI. Indeed, in a pattern-recognition based BCI, the four factors will obstruct the extraction of discriminative features to distinguish these mental tasks. It will be more obvious when analyzing in an asynchronous BCI. To overwhelm these obstruction, first we enhance the SNR of EEGs by removing the external artifacts and user-generated artifacts. Then some techniques have to be applied to reduce the sample dimension and extract features to distinguish these mental tasks.

In the thesis, we consider the asynchronous BCI system with two mental tasks, imagination of left and right hands movement. In other words, we will classify three classes including two predefined active events and one for the idling state. The objective of the work is the development of an accurate recognition in an asynchronous BCI system. We focus our attention on how to overcome the above obstruction and how to build a better recognition model to distinguish the three classes well.

Chapter 2

Survey of brain-computer interface systems



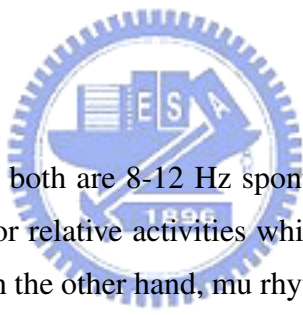
In chapter 1, we have mentioned that a BCI is a communication system which allows the user to control a device depending on only his thought, without through brain normal output pathway and muscles. We also have gave an overview of BCI system in section 1.3. In this chapter, we will first describe various EEG-based BCI systems today in section 2.1. Then we will do some comparisons between these existing BCI systems in section 2.2. Finally, in section 2.3, we discuss the limitations of BCI systems.

2.1 Brain-computer interface systems today

Present-day BCIs falls into various groups based on what kind of EEG signals they use. As we discussed in the previous chapter, the brain signals used in BCIs can be divided into the three classes:

- Brain rhythm:

- mu or alpha rhythms:



Mu and alpha rhythms both are 8-12 Hz spontaneous signals, but mu rhythm is associated with motor relative activities while alpha rhythm with sleep, eye closing and relaxing. In the other hand, mu rhythm is often emerged in the sensorimotor cortex; however, alpha rhythm is appeared over the occipital cortex. The BCI developed by Wolpaw and his colleagues training a user to control a system by self-regulating his mu rhythm amplitude [56] is representative work of this type BCI.

- ERP:

- SCP:

SCP is the slow voltage changes of the brain cortex, having a 0.5-10.0s potential shifts and settled below 1-2Hz. Usually negative SCPs represent the mobilization or readiness and positive SCPs represent ongoing cognitive and neural performance or inhibition of neuronal activity [18]. It have been shown that individual can control the movement of a cursor on the computer screen after

trained to regulate his SCP. BCI developed by Birbaumer [2] belongs to this type.

– P300:

Infrequent, particular, or oddball stimuli in auditory, visual or somatosensory will evoke the P300 over the 300ms after stimuli onset. The P300 is usually used in a spelling BCI system [15, 35]. Because the P300 is a naive response to a particular or oddball stimuli, the P300-based BCI requires no user initial training.

- VEP and SSVEP : A brief or flashing visual stimuli will evoke VEPs over the visual cortex. Sutter [48, 49] uses VEPs to determine the user's direction of looking or grazing in a matrix of flashing stimuli on the user's screen. The system of Sutter classifies the brain signals when a flicking stimuli is shown to the user. SSVEP is also used to build a BCI system. There are two approaches to design a SSVEP-based BCI. One is that the user is trained to control his SSVEP's amplitude to exert operator action. The other is to provide multiple SSVEPs to control the BCI system [31]. The former needs a user training but the latter not. Sutter's BCI system and Middendorf SSVEP-based BCI are the representative works based on the two brain patterns.

• – ERD/ERS:

The typical BCI based on ERD/ERS is the Graz BCI. The system discriminates the motor relative tasks, such as imagination of hands or feet movement, relying on ERD/ERS. When we perform motor relative tasks, the amplitude of the contralateral ERD over the sensorimotor is stronger than that in the ipsilateral side. Depending on the amplitude of ERD/ERS, the system can classify the user's intend.

An overview of present-day BCIs can be found in Figure 2.1. The figure are organized in seven subfigures, A~G. In this figure, 'A' indicates the demanded equipment in a BCI system, including user's screen, EEG amplifier and scalp, a computer recording EEGs. In 'B', the electrode placement are illustrated. Electrodes with color black, green, red and

blue are associated with different BCI systems with respect to P300-based BCI 'C' and ERD/ERS-based BCI 'D', mu-rhythm-based BCI 'E', SCP-based BCI 'F', and SSVEP-based BCI 'G'.

The details of these BCIs based on the above brain signals will be described in the following sections. We will deal with the following BCI systems:

- Farwell P300-based BCI, (C) in Figure 2.1
- Graz ERD/ERS-based BCI, (D) in Figure 2.1
- Wolpaw μ -rhythm BCI, (E) in Figure 2.1
- Birbaumer SCP-based BCI, (F) in Figure 2.1
- Middendorf SSVEP-based BCI, (G) in Figure 2.1

2.1.1 Farwell P300-based brain-computer interface

Farwell [15] proposed a spelling BCI system in 1988 based on P300 EEG signals. EEGs are usually recorded over the parietal areas in the system.

A user faces a screen that display 6X6 matrix, consisting of letters, numbers or other symbols, shch as the right top corner in Figure 2.1 (C). One raw or column chosen randomly flashes every a predefined fixed time and each raw and column will be flashed twice in a complete trial. The user is asked to select a symbol in advance and focus his attention on it, simultaneously counting the numbers of times that the row or column containing the symbol is flashed. This is a oddball paradigm which will elicit a P300 over 300ms after stimulus onset. When the raw or column contains the selected symbol is flashed, the user will produce a larger P300 response, compared to other raws and columns. This can be found in Figure 2.1 (C), showing an example of these P300s induced by target and non-target stimulus.

As we mention in the previous chapter, the P300 is a ERP, detected by averaging the EEG signals over many trials. After applying artifact removal and filtering, averaging is performed to each raw and column to measure the P300 amplitude. The amplitude of

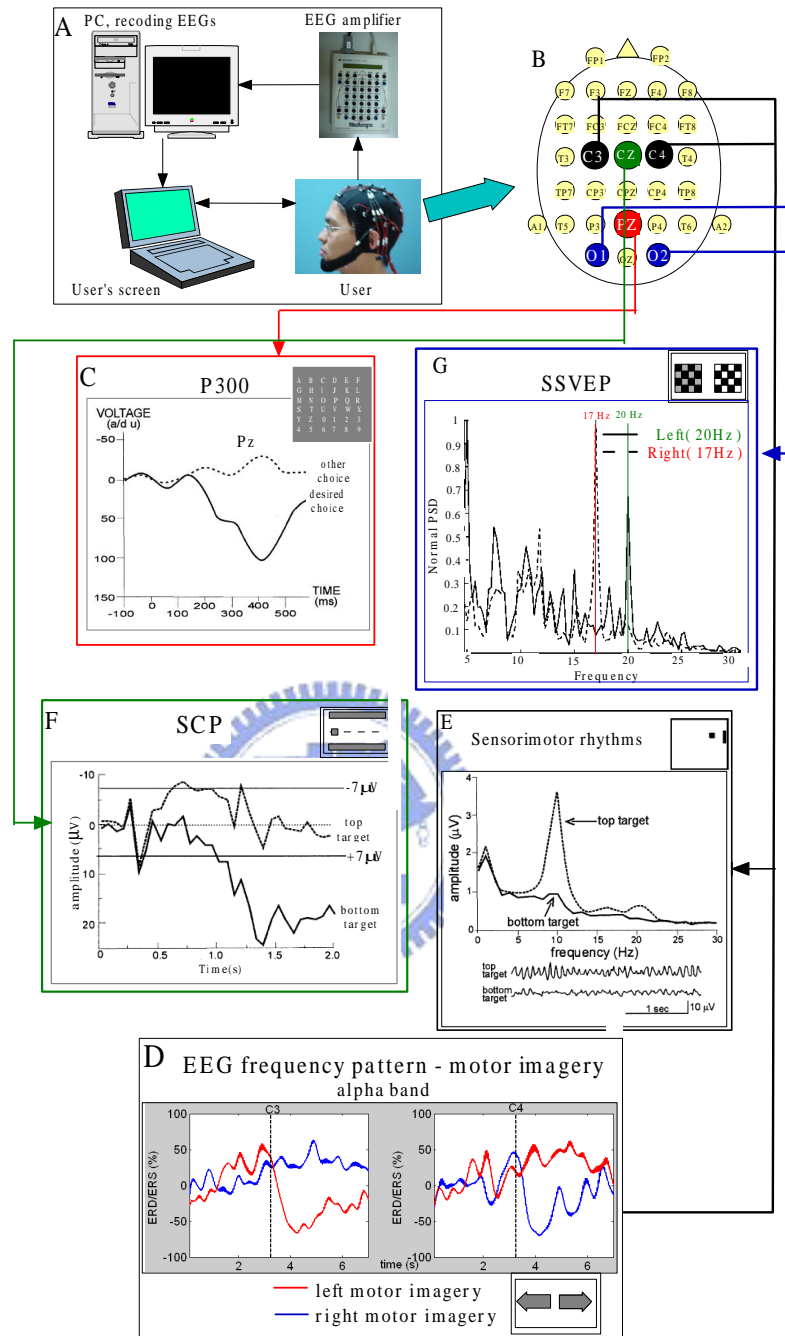


Figure 2.1: An overview of BCI today. 'A' indicates the demanded equipment in a BCI system, including user's screen, EEG amplifier and scalp, a computer recording EEGs. In 'B', the electrode placement are illustrated. Electrodes with color black, green, red and blue are associated with different BCI systems with respect to P300-based BCI (C) and ERD/ERS-based BCI (D), mu-rhythm-based BCI (E), SCP-based BCI (F), and SSVEP-based BCI (G). [27]

the row or column containing the desired letter is much larger than the others (see in Figure 2.1.1). The BCI then determines the user's intent by means of the P300 amplitude. For example, if the selected symbol is "B", "R", "A", "I", "N", then after each block the fixated letter is added to the screen so that user is conscious of slowly spelling out the word "BRAIN" through a succession of five blocks [46].

In Farwell [15] system, the user can communicate at a rate of 2.3 letters/min, after training a user for a few sessions, with accuracy rate about 95%. Since 1988, the P300-based BCI today can yield a communication rate of one word(about 5 letters) per minute. In related work, the P300 is recorded in a virtual environment [23], suggesting that a single-trial P300 amplitude may be used to control environment.

The drawbacks of the BCI is:

- The system is very slow because first the user must wait for visual stimuli presentation. Second, we need to perform average over many trials to detect the P300 response.
- A user may be fatigued for the flashes in the system.

The advantages of the BCI is:

- P300 is naive, thus requiring no initial user training.
- The accuracy is high.



2.1.2 Graz ERD/ERS-based brain-computer interface

In 1977, Pfurscheller proposed the concepts of *event-related desynchronization(ERD)* and *event-related synchronization(ERS)*. He addressed that the contralateral ERD at mu and beta band is induced during a movement or imagination of a movement. He utilized this phenomenon to develop a BCI system based on the mental tasks of imagination of a movement, which usually are imagination of hand movement. The changes of mu ERD/ERS and beta ERD/ERS in a movement related task are described as follows:

mu ERD/ERS : Pfurscheller argued that voluntary movement will elicit a circumscribed ERD at mu band, localized over the sensorimotor areas. For example, during a hand

movement, the contralateral mu ERD is started 2s prior to movement onset and becomes symmetric bilaterally immediately before the movement execution. After execution of the movement the central region induces a localized mu ERS [36]. Figure 2.2 shows these changes in topography mapping.

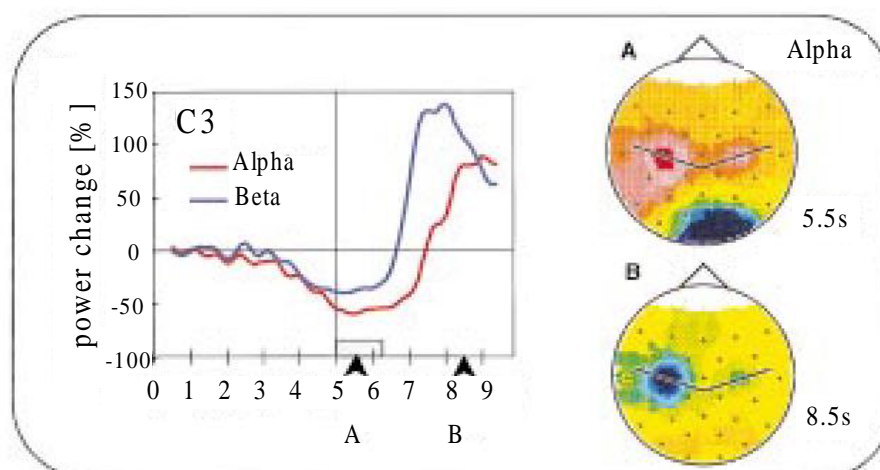


Figure 2.2: Show the changes of ERD/ERS in a right hand movement. In the left side, the alpha(mu) and beta ERD/ERS of the task in electrode C3 is given. In the right side, the two topographies shows the ERD/ERS changes for an interval at the movement execution(A) and at the post-movement recovery period. We can see the localized ERD over contralateral sensorimotor areas and a ERS in the occipital, both at 8-12 Hz. After the movement, the contralateral mu ERS is induced [36].

beta ERD/ERS : The beta ERS is found 1s posterior to the movement completion over the sensorimotor areas. In a right hand movement imagery, for example, during the movement execution, a simultaneous contralateral beta ERD and ipsilateral beta ERS is occurring. After the termination of movement, a contralateral beta ERS is appearing. See these changes in Figure 2.3.

The system use these above phenomena of ERD/ERS at mu and beta band to detect the user's wish. The type systems present-day are based on performing different motor imagery tasks, almost left and right hand motor imagery as the command to control a device. The feedback of the system often is a cursor movement controlled by the magnitude of the ERD/ERS.

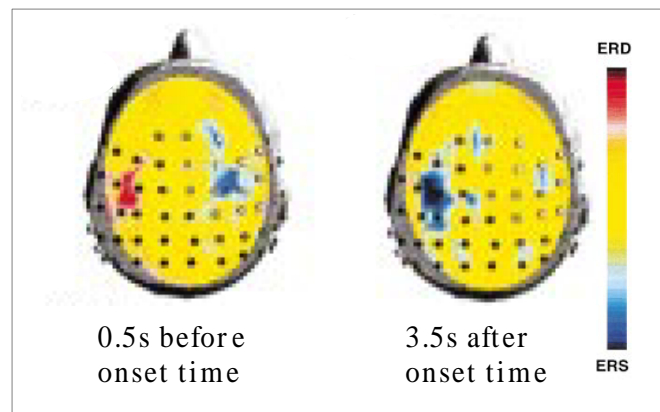


Figure 2.3: In a right hand movement imagery experiment, during the movement execution, a simultaneous contralateral beta ERD and ipsilateral beta ERS is occurring (left side). After the termination of movement, a contralateral beta ERS is appearing (right side). The signal was measured in C3 [36].

Research up to date focus on how to classify the different motor imagery tasks. Approaches include using other features, such as performing common spatial filtering (CSP), autoregressive model (AR model), wavelet transformation. There are studies aimed at discussion of the classifier such as *support vector machine (SVM)*, *linear discriminant analysis (LDA)*, *learning vector quantization (LVQ)*.

Future work in the system is to develop an asynchronous BCI system based on ERD/ERS. In 2004 they have proposed a simulation of asynchronous BCI system [52].

The drawbacks of the BCI is:

- The system may be slow because the interval between two successive trials lasts about some seconds for the event-related changes of brain activity need time to develop and recover.

The advantages of the BCI is:

- The time of training a user to controlling the system is short.
- The recognition accuracy is high.

2.1.3 Wolpaw and MacFraland μ -rhythm-based brain-computer interface

The system proposed by Wolpaw and MacFraland use the recorded brain rhythms to control the cursor movement. This system is based on a self-regulation of the mu or beta rhythm amplitude over sensorimotor areas. That means this is an operant conditioning based system.

In the one-dimensional mode, targets are located in the top or bottom edge of the screen and the user has to put the cursor to hit the targets vertically. Simultaneously the cursor move horizontally from left to right in a predefined fixed rate. In the two-dimensional mode, targets can appear anywhere of the screen and the user must move the cursor both vertically and horizontally. The cursor movement is usually determined by the amplitude of the mu or beta rhythm, or both. For example, in [29], the cursor movement is a linear function of mu rhythm amplitude. Usually the amplitude increase of rhythms will move the cursor to the target at the top, while decrease will cause the cursor to the target at the bottom. Figure 2.4 shows the interface of cursor movement in the system. In Figure 2.4, there are two parts, one for 1-dimensional and the other for 2-dimensional.

Usually power spectrum is applied to analyze signals. As see in Figure 2.1 E, top target will lead to a sharp peak at mu band while the bottom to a smooth curve. By means of this phenomenon, the system can recognize the user's intent between top target or bottom target.

Two important issues will influence the system performance. One of them is the targets number while the other is the rapidity of horizontal cursor movement. A greater number of targets and a faster movement will lead to a higher performance [29].

The users usually have to learn over a series of 40 min sessions to control the system. The procedures of training a user to achieve self-regulation are [57]:

1. In the initial sessions, the user is asked to perform motor imagery tasks to control the cursor movement. The imagination is done by thinking all the details of the action.
2. After step 1., the user now is required to move the cursor like they perform conventional motor acts yet without any thinking about the details of the action.

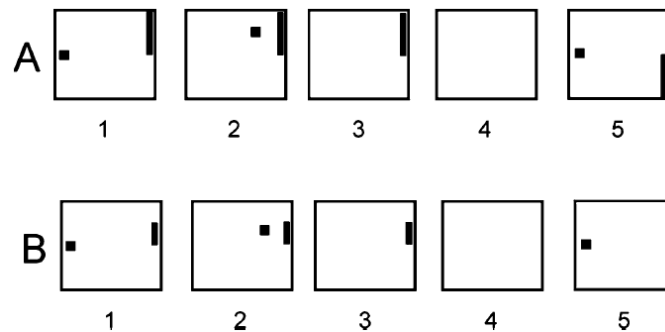


Figure 2.4: Mu rhythm based BCI developed by Wolpaw. The (A) is one-dimensional target task and (B) is the two-dimensional target task. (1) one of the targets and the cursor are present on the screen, (2) the cursor begins to move across the screen with its vertical movement controlled by the user, (3) the target flashes when it is hit by the cursor, (4) the screen is blank for a brief interval, (5) the next trial begins. [29].

The drawbacks of this system is that:

- 1. This is an operant conditioning based BCI, so it need a long time to train a user. In addition, in several papers examined, it appears that not all subjects can achieve self-regulation.

The advantages of this kind BCI is that:

- The system will have a good performance about 80-95% to control the cursor movement if a user is completely trained.

2.1.4 Birbaumer SCP-based brain-computer interface

This BCI is also called *thought translation device* proposed at the University of Tbingen in Germany. TTD is an another operant conditioning based BCI. The user is trained to control the BCI system for a long time by means of self-regulating his SCP amplitude. The SCPs can be extracted after performing filtering, artifact removals. Depending on the amplitude of the user SCP, his intent is determined. This system has been employed to some ALS or patients with paralysis in their home or nursing home [27]. classify the user's intend

In the system, the user faces a screen and choose to move the cursor to the top or the bottom on the screen by controlling his amplitude of SCP. The vertical movement of the cursor is decided by the amplitude of measured SCP. An example of a interface used in SCP-based BCI is given in the right corner of Figure 2.1 (F). In this BCI system, the user choose to move the cursor to the top or the bottom on the screen by regulating his amplitude of SCP. The vertical movement of the cursor is determined by the amplitude of his SCP. As illustrated in Figure 2.1 (F), usually a positive SCP will lead the cursor to the bottom target while a negative SCP to the top target. The cursor movement, feedback in the system, is necessary in order to let the user be able to reinforce his SCP whenever. The final selection made by TTD is highlighted in the top or bottom on the screen.

language support program(LSP) is a technique which can be incorporated into TTD. By means of LSP and TTD, the users can choose a letter or word by a series of binary choice. The performance of LSP and TTD cause a 65-90% accuracy and transfer rate 0.15-3.0 letters/min. The future work of the system is to increase the performance by using the predictive algorithm.

The drawback of TTD is that:

- 1. This is a operant conditioning based system, so it also takes a long time to train a user to fulfill self-regulation.

The advantages of TTD are that:

- The system will have a performance about $> 75\%$ to control the cursor movement if a user is completely trained.
- The SCP pattern is well understood based on the neurophysiological basis and is universally presented in cortical cells [27].
- The rule to train a user how to self-regulate SCPs have been known [33].

2.1.5 Middendorf SSVEP-based brain-computer interface

SSVEP is a response to a visual stimuli at a specific frequency. The SSVEP is characterized as an increasing in EEG activity at the stimulus frequency. The EEGs are recorded

over the visual cortex. One of the typical visual stimuli is the black and white checkboard pattern, as plotted in right top corner of Figure 2.1 (G), which is an example of a interface of the SSVEP-based BCI . The analyzed signals in Figure 2.1 (G) is the amplitude of the bipolar SSVEP over O1 and O2 electrodes. As we can see in this subfigure, if the left checkboard pattern is modulated at 20 Hz, a high peak at 20Hz will be elicited, while if the right checkboard pattern is modulated at 17 Hz, it will lead to a high peak at 17Hz. The system in this figure can determine the user's wish depending on the power at the specific frequency which the stimulus are modulated at. The other commonly used stimuli is a white fluorescent tubes that are luminance modulated at 13.25Hz and mounted behind a translucent diffusing panel as the visual stimuli [30,44].

The feedback in the system is a horizontal light bar to present the amplitude of SSVEP. There are two approaches to develop a SSVEP-based BCI. In one way, the user is trained to control the amplitude of his SSVEP. In the other, multiple SSVEPs are provided to control the system. The former is based on operant conditioning approach, while the latter is based on pattern recognition approach respectively.

- *operant conditioning approach SCP-based BCI*

There are some thresholds, adjustable for individuals, used to determine the control command. In a binary command system, two thresholds are employed to decide the command. Typically, if the VEP magnitude is above the upper threshold in a specific time, then one control is generated while if below the lower threshold, the other control is generated. When this system is combined with a simple flight simulator. For example, when the user's SSVEP is over the upper threshold, which is 75% of the samples, in a one-half second interval, the simulator will roll to right in one degree. On the contrary, when the user's SSVEP is below the lower threshold, the simulator will roll to left in one degree.

In the combined system, it needs about 30-min user training to achieve some level of control. For a trained operator, there are 80-95% accuracy to roll the simulator to the correct direction. The SSVEP-based BCI in flight simulator is shown in Figure 2.5 [5].

The drawbacks of this kind BCI is that:

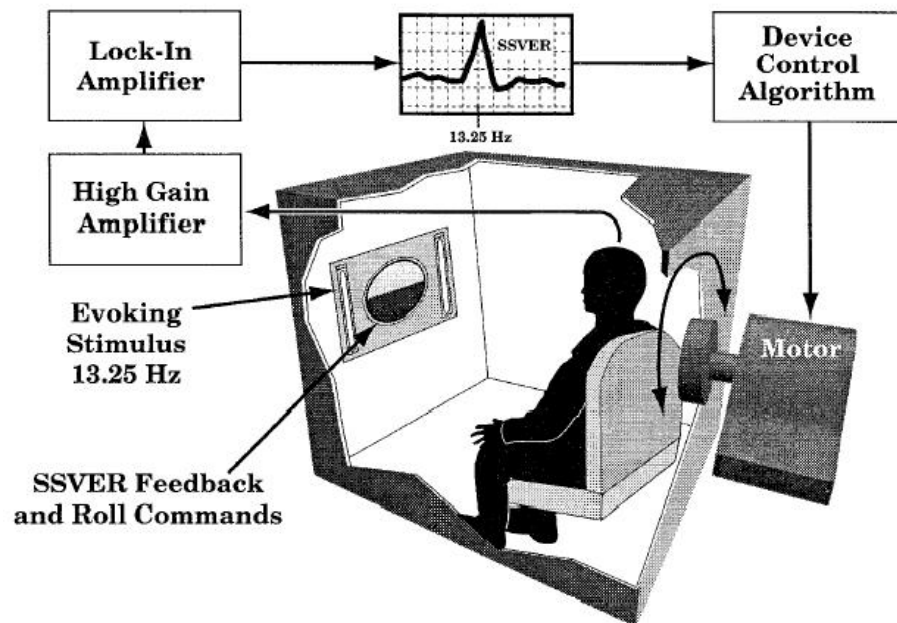


Figure 2.5: A SSVEP-based BCI and the flight simulator [5] are shown.

- 1. How to train a user to accomplish the self-regulating has no clear and definite principles.
- 2. The system assumes the user is always looking at the screen. But in fact the flashing visual stimuli will easily fatigue the user.

The advantages of this kind BCI is that:

- The recognition rate is better than the current systems and may be used by a user with sufficient eye control.
- *pattern recognition approach SCP-based BCI* This kind of BCI is much like the Sutter VEP-based BCI [49]. In the system, the task is to select various virtual buttons on the screen. The user choose the desired button by looking at it. There are two stimulus differentiating at the frequencies to generating two different SSVEPs. When the user gaze at the virtual button, the SSVEP is increasing at the button's modulation frequency. That is the SSVEPs is characterized by an amplitude increase at the stim-

uli's frequency. The system is also using the amplitude of the different SSVEPs to determine what button the user is looking.

The drawbacks of this kind BCI is that:

- 1.The system assumes the user is always looking at the screen. But in fact the flashing visual stimuli will fatigue the user easily.

The advantages of this kind BCI is that:

- The recognition rate is better than the current systems and may be used by a user with sufficient eye control.
- The user training is shorter.

2.2 Comparisons of brain-computer interface systems

Several features of existing BCI systems are compared in Table 2.1

2.3 Limitations of brain-computer interface systems

Here, we discuss some possible limitations in a BCI system [27]. These include:

- Habituation
- Interference and distraction
- Impaired visual system
- Instability of EEG signals
- Variability of a specific brain pattern
- Relation of users and brain patterns

Not all above limitations are present in all BCI systems.

Table 2.1: The comparisons of BCIs.

System.	Farwell	Graz	Wolpaw	Birbaumer	Sutter	Middendorf
Training Time	Minutes	2-2.5 hours	15-20 sessions	Months	10-60 min.	6 hours
Transition Rate	2-3 letters/min.	?	20-25 <i>bit/min</i>	?	10-12 words/min.	?
Accuracy	95%	90%	80%	> 75%	95%	80%
Interface	6X6 symbols	cursor movement	cursor movement	cursor movement	8X8 symbols	flight simulator
Control signals	P300	mu and beta ERD/ERS	mu rhythm	SCP	VEP	SSVEP
Electrodes	PZ	C3, C4	C3, C4	Cz	O1, O2	O1, O2
Operant or pattern-recog.	N	N	Y	Y	N	Y/N
Feedback	Y	Y	Y	Y	Y	Y

2.3.1 Habituation

Researches about the relation of habituation and P300 have been proposed [9, 24, 38, 47]. These studies argue that the P300 event-related potential is well-known to habituate after a large number of repeated presentation of oddball paradigm trial blocks because the updating of the stimulus environment neural model becomes automated. The habituation will lead to a declined amplitude of P300, with no changes in the latency, for both auditory and visual, olfactory oddball stimuli. Therefore, for the P300-based BCI system, it has the possibility to crash the system when the habituation results in such a seriously declined P300 that the system can not recognize P300 signals. The habituation may be happened even for other EEG response elicited by external stimuli, such as VEP and SSVEP.

However, for the self-regulation based BCI systems, the long training period is well-known. It reveals that for this kind of BCI system, the habituation will not influence the

corresponding brain responses. The EEG responses remain the same for weeks, months or years.

2.3.2 Interference and distraction

When the user have to communicate with the world by the system for a long time, interference and distraction may be the potential problems. For example, in the P300-based spelling system, if the user wishes to spell some sentences, he has to concentrate not only on the row or column containing the target symbol and simultaneously count the number of the occurrence of such row and column but also on the characters, words, sentences to be communicated. On the other hand, in ERD/ERS-based BCI or SCP-based BCI, the user has to perform the different mental tasks as well as to pay attention to the movement of the cursor or on the feedback. In all the above BCI systems, interference and distraction may have the possibility to lead to a low performance system or even a failed system.

2.3.3 Impaired visual system

This problem is occurred when the BCI system is based on visual event-related potentials. There are two issues leading to the problem. For one thing, as we have mentioned in section 2.1.5, the user have to fixate the visual stimulus, which is usually a flashing pattern or flicking pattern, for a long time. Often the users are easily fatigued by such kinds of stimulus. For another, this kind of BCI assumes the user always have to gaze at the target symbol during a trial. It implies whenever the user move his eyes or close his eyes, a communication error occurs.

2.3.4 Instability of EEG signals

Instability of EEG signals is a general problem for all BCI systems. When measuring EEGs, we have to inject conductive jelly into the electrodes, serving as media to ensure lowering of contact impedance at electrode-skin interface. Usually the conductive jelly will be dried after two or three hours, resulting in a noisy signals. This restricts the flexibility of EEG-based system because it can not be used continuously for a long time. On the other

hand, the amplitude of measured EEGs is so small such that it easily smeared, contaminated or hidden by other artifacts.

2.3.5 Variability of a specific brain pattern

It is well-known that a BCI system is a user-specific system. For different users, the BCI system has to be developed associated to its user. However, in addition to the problem of user-specific property, BCI systems still encounter the problem that even for the same user, his specific brain patterns, for example the P300s for the same stimuli, is neither always the same. When this used EEG pattern varies too much over trails, the performance of the BCI may be low and communication errors happen. For example, in the ERD/ERS-based BCI, we find that the possibility of that the mu-ERD of a user may occur at 8-10 Hz yesterday while at 10-12 Hz today is not low. The solution may be that update the BCI system every time when the user want to use it, yet it is very consuming and meaningless. It is still not clear how to solve the problem today.





Chapter 3

Material



3.1 Experiment paradigm

In the thesis, the dataset we analyzed is experimented as the paradigm showing in Figure 3.1. A fixation cross is shown on the screen up to 2s and the following is a warning tone lasting from 2s to 3s. Afterwards, a visual cue, an arrow, is given from 3.00s to 4.25s to indicate left or right corresponding to a left- or right-hand movement imagination. From 4.25s to 8s, the user has to perform the imagination of hand movement [51].

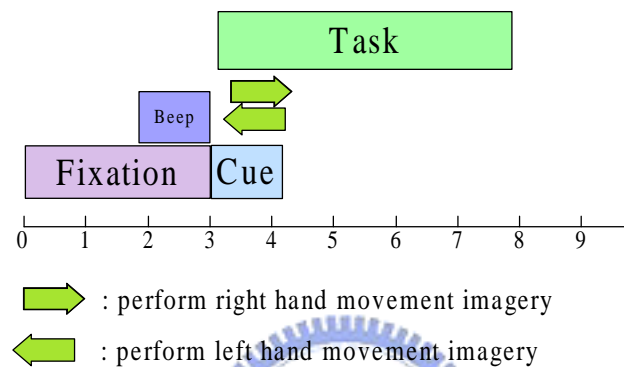


Figure 3.1:

3.2 Datasets

First we concern the datasets D01 ~ D07 which were made by ourselves following the above paradigm and a 32 Ag/Cl electrode cap from neuronscan. 32 channels were placed at the position of international 10/20-system. Signals are digitized at 1000Hz with 32 bit accuracy. These datasets will be analyzed in Chapter 4. Every dataset of D01 ~ D07 contains at least 200 trials EEGs. Each trial last 8s.

Besides the signals we made, we also analyze some dataset from BCI competition2003 and 2005 which list as follows.

1. G01

The 2003 BCI competition dataset of hand imagery from Graz group. There are 140 trails of left and right tasks (total 280). The signals were measured by bipolar

montage and digitized at 128Hz. In this dataset, only the three channels, C3, Cz, C4, are available. Each trial last 9s.

2. G02~G04

The 2003 BCI competition dataset of IIIa. Each dataset contains 4-class EEG data including motor imagery of left hand, right hand, foot and tongue. We use only the first two class in the work. The recording were measured with a 64-channel EEG amplifier from Neuroscan, using the left mastoid for reference, and is digitized at 250Hz. The signals were filtered between 1 and 50Hz. For G02, there are 45 trails of left and right tasks (total 90). For G03 and G04, there are 30 trails of left and right tasks (total 60).

We will use these competition datasets to assess our methods when analyzing an asynchronous BCI system.





Chapter 4

Single-trial EEG classification



In section 1.4, we have mentioned that there are four types of obstruction we have to overcome when we discriminate the user's intention, including:

- 1 The amplitude of EEG signals are small and very sensitive to external noise.
- 2 The dimension of EEG sample space is very high.
- 3 User-generated artifacts may contaminate and smear the brain signals and even result in misleading conclusions regarding the real controlling skills of a user.
- 4 The brain pattern may seriously vary over trials.

In this chapter, we present some methods to settle the above problems when analyzing a synchronous BCI system. We first discuss signal preprocessing including artifact removal, bandpass filtering, and some reference-independent filters in section 4.1. In section 4.2, several important approaches of features extraction covering ERD/ERS, common spatial filter(CSP), Haar wavelet, Morlet wavelet are considered. These four approaches aim to extract adequate features to characterize EEGs for different mental tasks, which simultaneously reducing the dimension of original sample space. However, the four analysis taking oscillation power change, spatial relation, time and frequency information into account respectively often result in a large number of features and thus a still high dimension data vector.

Therefore, in section 4.3, two methods of feature selection, t-statistic and forward feature selection, are used to further refine these features delivered by the feature extraction. Then we give an overview of support vector classifier (SVC) in section 4.4. Finally, in section 4.5, some experiments and analyses were done to discuss the influence of these approaches.

In Figure 4.1, we show a plot to indicate what we will discuss and present in this chapter and the processing order.

4.1 Signal preprocessing

First, typical types of artifact removal and bandpass filtering are applied to enhance SNR of EEGs. Afterwards, we present the principles of some spatial filters leading to the

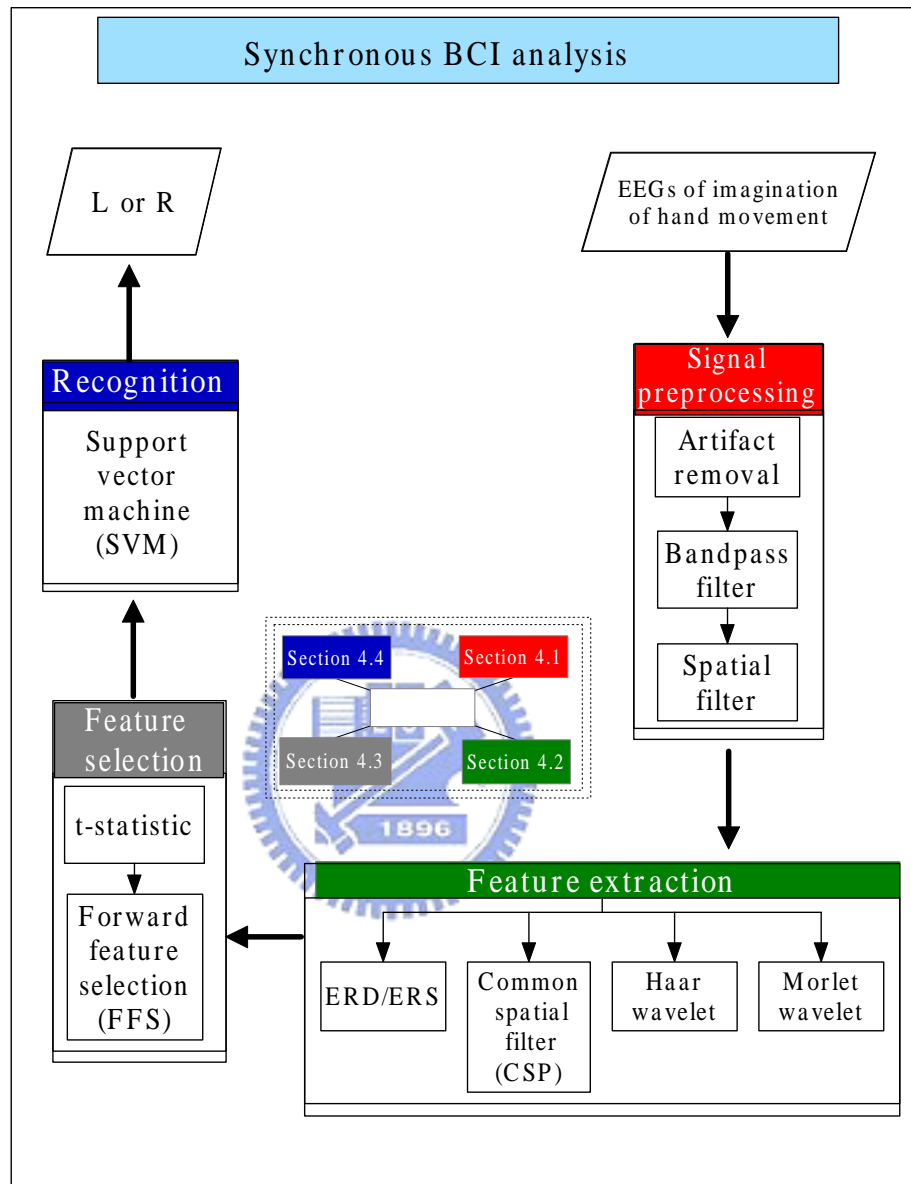


Figure 4.1: Flowchart. Indicate the processing order and what we will discuss and present in this chapter. After acquisition of EEGs for different mental tasks, we first apply signal preprocessing to enhance SNR, discussed in section 4.1. Then four methods of feature extraction are illustrated in section 4.2 and two methods of feature selection are given in section 4.3. In section 4.4, The classifier SVC is illustrated to recognize EEGs. Finally output the classification result, left or right in our case.

reference-free signals.

4.1.1 Artifact removal

When measuring EEGs, there exists many artifacts generated from the external or internal environment. The typical artifact of the former is the 60 Hz (or 50 Hz) powerline. For the latter, the representative artifacts are eye blinking, eye movement, body movement or muscle artifact and heartbeat. All the four artifacts are unavoidable when EEGs are recorded.

In this thesis, we deal with the artifacts: eye blinking or eye movement, which is measured by EOG channels, and body movement or muscle artifact, which is recorded by EMG channels. For the EOG artifact removal, usually a threshold is given to detect eye blinking and movement. For the movement artifact, a bipolar EMG is analyzed to detect body movement. When a trial signal is detected to be with using EOG or EMG, this trial will be ruled out. The heartbeat can be measured by reserving one channel for ECG. The ECG artifacts can be suppressed by applying a highpass filter which will be described in the next section.

The influence of eye blinking, eye movement, heartbeat are shown in Figure 4.2, Figure 4.3, Figure 4.4 respectively. These figures reflect that higher spikes will be elicited when these artifacts appear, leading to mistaken EEGs.

4.1.2 Bandpass filter

General processing of this step will include two steps:

- Because the heartbeat is at low frequency, we will apply a highpass filter to eliminate the influence of heartbeat.
- A notch filter is applied to suppress the powerline artifacts.

Both above steps can be achieved by applying a bandpass filter. Besides these processing, bandpass filter at different frequency band may be applied in different BCIs. In the work, we will applied a bandpass filter at cutoff 8 and 30 where the motor related mu and beta ERD/ERS are located.



Figure 4.2: Eye blinking artifact. This is an example of demonstrating eye-opening (green triangle) followed by a series of eye-blinking (blue triangle). The EOG in the last two channel measures these changes of eye blinking and eye opening. As the figure demonstrates, multiple channels, especially the frontal electrodes, are contaminated by eye opening and eye blinking. These artifact usually can be detected by given a threshold and signals over the threshold will be excluded [1].

4.1.3 Reference-free spatial filter

As We mentioned in Chapter 1, the EEG measurements are reference-dependent while we usually prefer reference-independent signals. We will give a brief description about three commonly used reference-free spatial filters.

1 Common average reference (CAR)

CAR is computed according to

$$V_{CAR(i)} = V_i - \sum_{n=1}^N V_n \quad (4.1)$$

where N is the number of scalp electrodes, V_i represent the potential of electrode i . Eq. 4.2 means that the average of all scalp electrodes are subtracted from all electrodes.

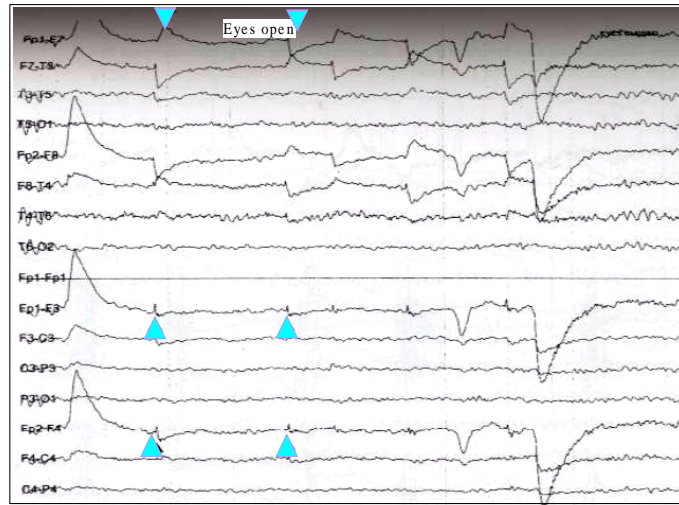


Figure 4.3: Eye movement artifact. This example demonstrates a series of eye movement (blue triangle). As the figure presents, many eye movement artifact are elicited in multiple channels, especially the frontal electrodes. These artifact usually can be detected by given a threshold and signals over the threshold will be ruled out [1].

2 Bipolar reference

Bipolar reference is calculated by subtracting the secondary electrode V_n from the active electrode i .

3 Laplacian reference

Laplacian reference is calculated as

$$V_{CAR(i)} = V_i - \sum_{n \in S(i)} V_n \quad (4.2)$$

where $S(i)$ is the neighbors of electrode i . If $S(i)$ is nearest-neighbor, then this a *small Laplacian reference*. If $S(i)$ is nest-nearest-neighbor, then this a *large Laplacian reference*.

An illustration of electrode locations with respect to the three spatial filters and ear reference is shown in Figure 4.5 [20]. Although we provide the three methods to enhance signals, we only discuss the inference of *Laplacian filter* and *bipolar reference* in section 4.5 because when analyzing the brain signals of motor related tasks, the two methods have better performance than others.

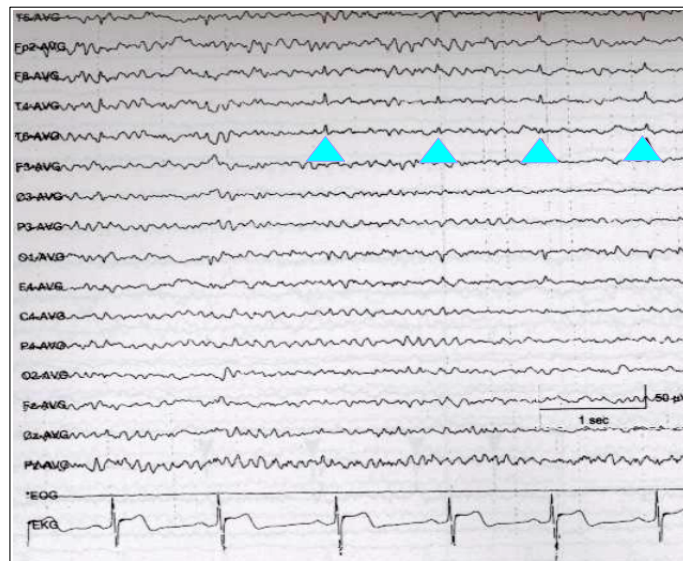


Figure 4.4: Electrocardiogram (EKG) artifact. This figure illustrates the EKG artifacts, mistaken for intermittent sharp spikes of EEG, in multiple channels, coincident with the EKG channel in the last channel. The EKG artifacts are indicated by cyan triangles [1].

4.2 Feature extraction

Because the signals of performing the predefined mental task are usually smaller than the ongoing EEGs, the signals we interested are concealed in the irregular ongoing signals and noise. To overcome the problem, we will perform some approaches of feature extraction to extract the signals we interested. Feature extraction is an important issue in a BCI system because it has a direct influence on the accuracy of the system. It is also a technique of decreasing the sample dimension by applying some transformations to each trial data. How to design the transformation and simultaneously retain the most discriminate features in a BCI system is a field, attracting many researchers' interest. Several approaches have been proposed, such as:

- ERD/ERS
- Common spatial filter
- wavelet transform : Haar

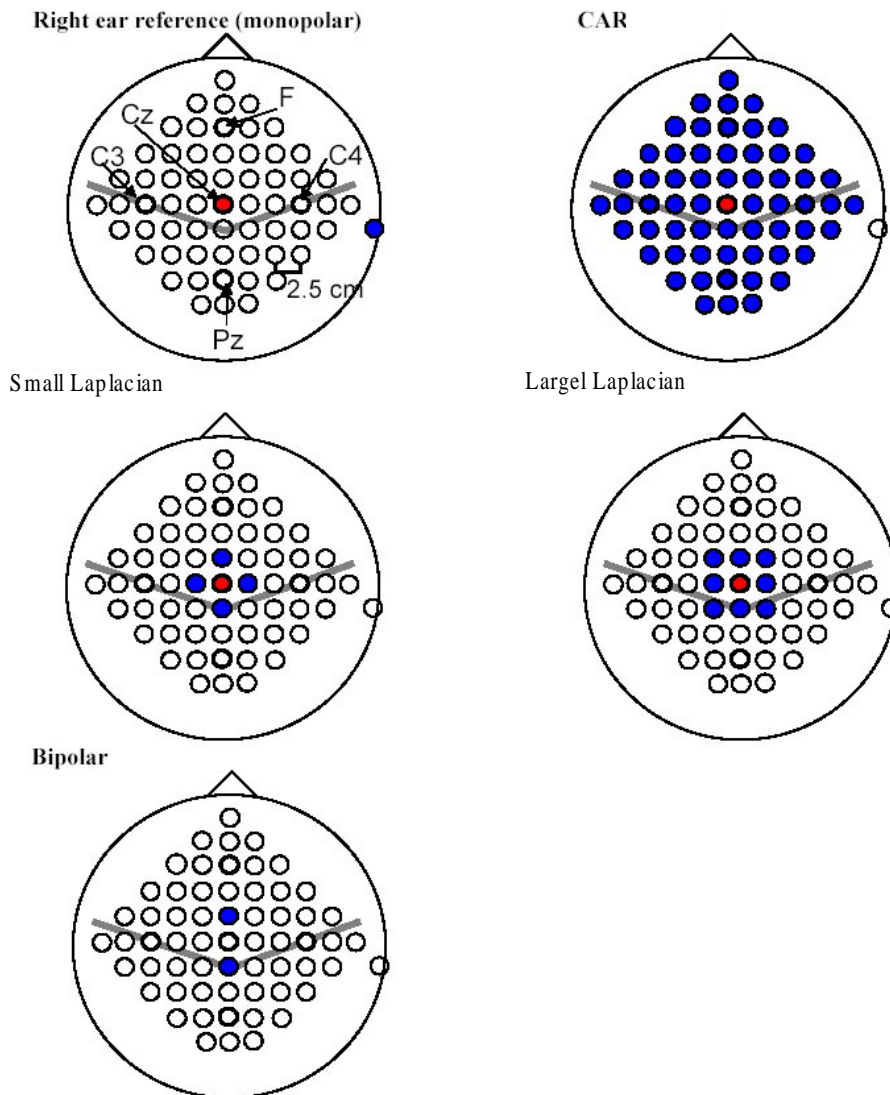


Figure 4.5: Electrodes placement of montages: right ear, common average reference(CAR), small Laplacian filter, large Laplacian filter and bipolar are shown [20].

- Time-frequency analysis : Morlet
- Power spectrum

In the section, we will focus on all but power spectrum because the power spectrum contains only the information on frequency domain, while the time-frequency analysis can contain both information on time and frequency domain.

When implementing the analysis of EEGs by means of the four approaches in this chapter, EEGs in the nine electrodes indicated in Figure 4.6, FC3,FCZ,FC4,C3,CZ,C4,CP3,CPZ and CP4, are used. The nine channels cover the sensorimotor and motor cortex activated during imagination of movement.

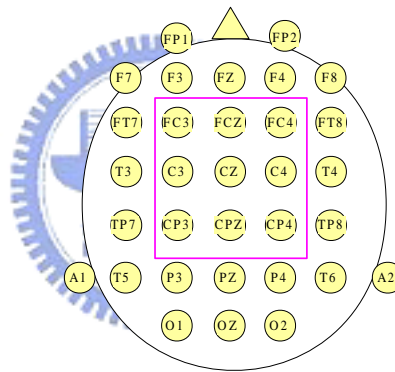


Figure 4.6: EEGs in the nine electrodes indicated in, FC3,FCZ,FC4,C3,CZ,C4,CP3,CPZ and CP4, will be analyzed by the approaches presented in this chapter.

4.2.1 Event-related desynchronization and Event-related synchronization

The concepts and algorithms of ERD/ERS have been illustrated in Chapter 1. For G01 dataset, the ERD/ERS of two mental tasks, left hand and right hand motor imagery, is plotted in Figure 4.7

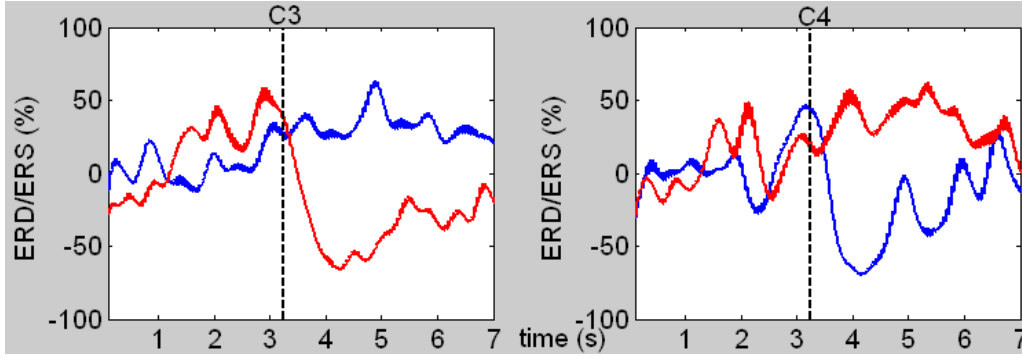


Figure 4.7: ERD/ERS of imagination of hand movement. The example demonstrates the ERD/ERS of imagination of left hand and right hand movement. The x-axis and y-axis represent 'time(1/128ms)' and the 'ERD/ERS value'. The 8-13 Hz bandpass filter is applied to extract the information of mu band. As we can see, the mu-ERD is contralaterally located over 3s 4s. In other words, when perform imagination of left hand movement, the mu-ERD is elicited over position C4 while imagination of right hand movement over position C3.

4.2.2 Common spatial filter

CSP is also called *common spatial subspace decomposition(CSSD)*. The principle of CSP is to simultaneously diagonalize the two covariance matrices associated with two classes [16]. Thus the limitation of CSP is that the most discriminative spatial filter is possible only to classify two classes. Approaches of multiclass extension of CSP by means of pairwise classification and vote are addressed in [26]. In the synchronous BCI, we classify only the left and right hand motor imagery, two classes, therefore without any problems of using CSP. However, when in asynchronous BCI, we encounter three-class problem and have to extent CSP to multiclass.

The algorithm of CSP [32] is as following:

- 1 Calculate the two covariance matrices, Σ_1 and Σ_2 , of the two classes.

$$\Sigma_1^i = \frac{\mathbf{V}^i \mathbf{V}^{iT}}{\text{trace}(\mathbf{V}^i \mathbf{V}^{iT})} \quad (4.3)$$

where \mathbf{V}^i is the raw data of trial i . The size of \mathbf{V}^i is $N \times T$, presented as N number of channels and T samples in time. The t means transpose.

$$\Sigma_1 = \sum_i \frac{\mathbf{V}^i \mathbf{V}^{iT}}{\text{trace}(\mathbf{V}^i \mathbf{V}^{iT})} \quad (4.4)$$

2 Calculate scatter matrix

$$\Sigma_S = \Sigma_1 + \Sigma_2 \quad (4.5)$$

, which can be diagonalize as

$$\Sigma_S = \mathbf{B}_S \lambda \mathbf{B}_S^t \quad (4.6)$$

Here \mathbf{B}_S is an $N \times N$ matrix of normalized eigenvectors, satisfying

$$\mathbf{B}_S \mathbf{B}_S^t = \mathbf{I}_{N \times N} \quad (4.7)$$

λ is the diagonal matrix of eigenvalues.

3 Perform the whitening transformation.

$$\mathbf{D}_1 = \mathbf{W} \Sigma_1 \mathbf{W}^t \quad (4.8)$$

$$\mathbf{D}_2 = \mathbf{W} \Sigma_2 \mathbf{W}^t \quad (4.9)$$

with

$$\mathbf{W} = \lambda^{(-1/2)} \mathbf{B}_S^t \quad (4.10)$$

\mathbf{D}_1 and \mathbf{D}_2 have the property of $\mathbf{D}_1 + \mathbf{D}_2 = \mathbf{I}_{N \times N}$ and of sharing the same eigenvectors.

4 Diagonalize \mathbf{D}_1 and \mathbf{D}_2 .

$$\mathbf{D}_1 = \mathbf{U} \Psi_1 \mathbf{U}^t; \quad (4.11)$$

$$\mathbf{D}_2 = \mathbf{U} \Psi_2 \mathbf{U}^t \quad (4.12)$$

, with the property $\Psi_1 + \Psi_2 = \mathbf{I}$

5 Calculate the projection matrix \mathbf{P}^t

$$\mathbf{P}^t = \mathbf{U}^t \mathbf{W} \quad (4.13)$$

Apply the projection matrix \mathbf{P}^t to each \mathbf{V}^i and then get projected data \mathbf{Z}^i as

$$\mathbf{Z}^i = \mathbf{P}^t \mathbf{V}^i \quad (4.14)$$

After performing the CSP, we get the projection matrix \mathbf{P}^t . Select the two most discriminative filter vectors and project them to each trial. This ensures the projected data having the property that the variance of the projected data contain the most relevant information for recognizing the two classes [32]. In fact, it makes the projected data of one class having the maximal variance while the projected data of the other class having the minimal variance.

Next projecting is that calculate the log-transformed normalized variance f_p^i of the each projected trial i .

$$f_p^i = \frac{\text{var}(\mathbf{Z}_i)_p}{\sum_{p=1}^2 \text{var}(\mathbf{Z}_i)_p} \quad (4.15)$$

We will take f_p as the features.

4.2.3 Haar wavelet transformation

Wavelet transform is a time-scale analysis and has the ability to perform local analysis, implying that analyze a localized area of a large signal. The property is shown in Figure 4.8. Many wavelet transformations have been proposed such as Haar, Daubechies, Morlet, Mexican Hat. Here we choose Haar wavelet transformation to process the brain signals because it is simple and easy to implement; the complexity is lower than other wavelets.

Assume the signals is $f(x)$, then the Haar wavelet coefficient of $f(x)$ at time t and scale s is

$$Wf(s, t) = \frac{1}{s} \int \Omega f(x) \Psi\left(\frac{x-t}{s}\right) \quad (4.16)$$

, where Ω is the length of signal and $\Psi(x)$ is the basis of Haar wavelet:

$$\Psi(x) = \begin{cases} 1, & \text{for } 0 \leq x < \frac{1}{2} \\ -1, & \text{for } \frac{1}{2} \leq x < 1 \\ 0, & \text{otherwise} \end{cases}$$

In Figure 4.9, we show the averaging result of Haar wavelet transformation over all trials of the data G01 with 8 scale levels.

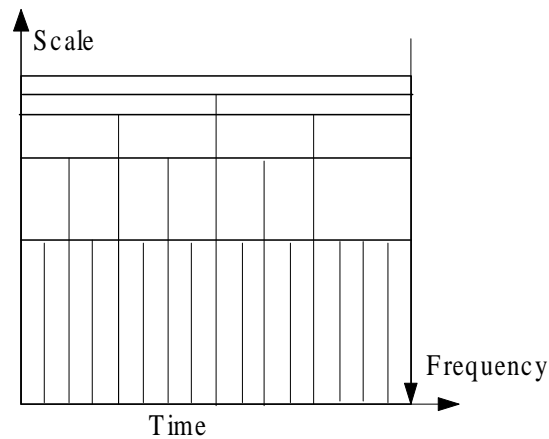


Figure 4.8: The relationship of time and scale resolution in Haar wavelet. Scaling a wavelet simply means stretching or compressing. Usually there is a correspondence between wavelet scales and frequency as:

1. Low scale \Rightarrow Compressed wavelet \Rightarrow High frequency
2. High scale \Rightarrow Stretched wavelet \Rightarrow Low frequency

4.2.4 Morlet wavelet transformation

Morlet wavelet is also a wavelet transformation but unlike Haar wavelet, it is a time-frequency analysis. The basis of Morlet wavelet is

$$Wf(f, t) = Ae^{\left(\frac{-t^2}{2\sigma t^2}\right)} e^{(2i\pi ft)} \quad (4.17)$$

The Morlet wavelet has more precise information of signals, but the complexity is a lot higher than Haar wavelet.

In Figure 4.10, we illustrate the averaging Morlet wavelet coefficients of dataset G01. We can also find the contralateral desynchronization at mu band about 3.75s 4.25s such as Figure 4.7. In effect, these are features that we want to extract in this step because it has the possibility to discriminate the tasks well.

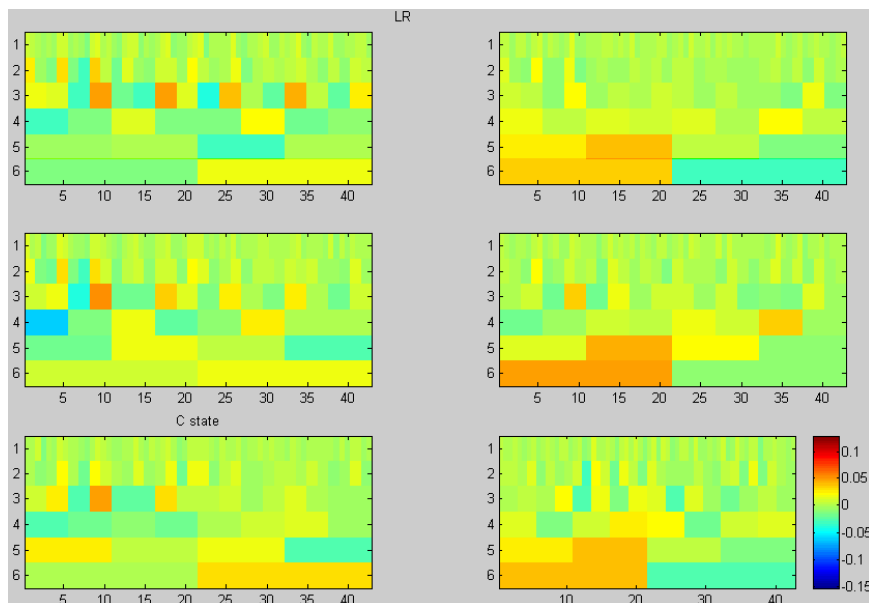


Figure 4.9: Haar wavelet

4.3 Feature selection

In some cases, the dimension of feature space may be too large to analyze it because the computational complexity is too high. In other cases, we have the probability to encounter the *small sample size* problem when the dimension of signal is far greater than the the small number of available training samples. In both cases, we must reduce the dimension of feature space. In this thesis, we apply *feature extraction* and *feature selection* to reduce the dimension of feature components after performing feature extraction.

Unlike feature extraction, feature selection is to discard some elements of a sample vector or a feature vector. It is to retrieve a subset from the measurement vector such that this subset is the most suitable for discrimination. How many features are left has no clear indices, yet it is a tradeoff. When the number is large, the complexity is high. On the contrary, when the number is small, we may lost information of the samples. We will find a reasonable subset at the training phase and meanwhile determine the number of left features.

We will introduce two approaches, *t-statistic* and *feature forward selection(FFS)* in the

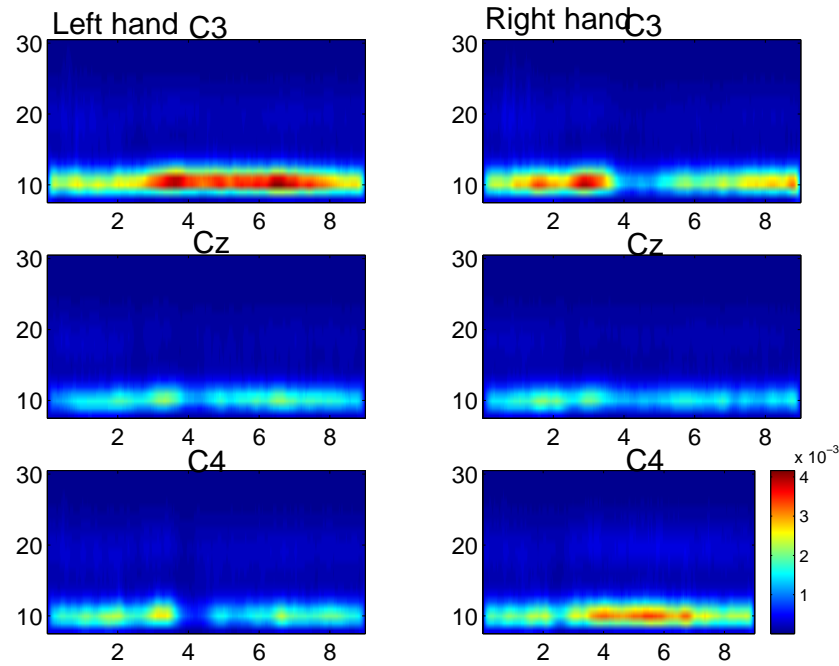


Figure 4.10: The averaging Morlet wavelet coefficients of dataset G01. The x-axis and y-axis represents respectively 'time' and 'frequency'. The color represent the coefficient values. We analyze the data from 4s to 8s and the frequency range is between 8 30 Hz. The three rows represent wavelet coefficients over C3, Cz and C4 positions and the left and right column represent the left hand imagery task and the right hand imagery task. As we can see in the figure, after performing Morlet wavelet transformation, we can easily find the contralateral desynchronization at mu band at about 3.75s-4.25s.

following. In this work, we first perform t-statistic to robustly extracted features and then apply FFS to the outputs of t-statistic to refine the feature subset.

When applying feature selection, we have to define a criterion to measure the discriminability of each features. The criterion associated with the two approaches is different and will be indicated in each respective section.

4.3.1 t statistic

The t-statistic assesses whether the means of two groups are statistically different from each other. In Figure 4.11 and Figure 4.12, we show an example of the discriminability between two groups. As Figure 4.12 displays, the mean and variance determine the dis-

criminability of the two groups. When the distance of the two means is large and the two variance are small, it has possibility to discriminate the two group very well.

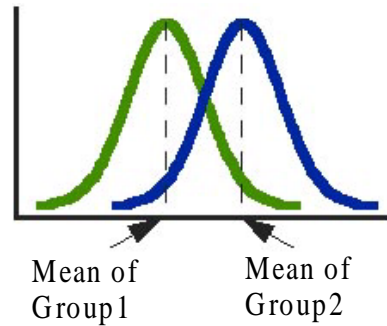


Figure 4.11: An example of 2 groups' distribution. [22]

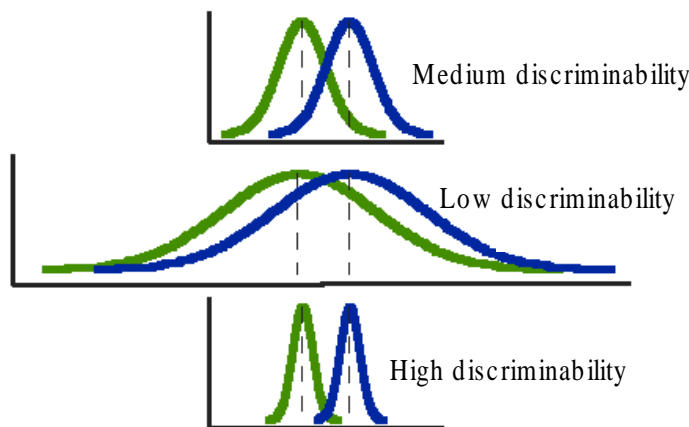


Figure 4.12: Show the relationship between discriminability, mean, and variance of two groups. [22]. As we can see, when the distance of the two means is large and the the distance of the two variance is small, it may be possible to discriminate the two group very well.

The formula of t-statistic is as follows :

$$t(\text{Group}_1^f, \text{Group}_1^f) = \frac{\bar{m}_1^f - \bar{m}_2^f}{\sqrt{(S_{12}^f)^2}} \quad (4.18)$$

where

$$(S_{12}^f)^2 = \left(\frac{1}{N_1} + \frac{1}{N_2}\right) \left(\frac{N_1\sigma_2^2 + N_2\sigma_1^2}{N_1 + N_2 - 2}\right) \quad (4.19)$$

The S_{12}^2 is the pooled variance between two groups in feature f , \bar{m}_1^f and \bar{m}_2^f , N_1 and N_2 represent the associated means and sample numbers of two groups in feature f . The criterion of estimating the discriminability in t-statistic is the value of

$$\|t(\text{Group}_1^f, \text{Group}_2^f)\|.$$

The higher the $\|t(\text{Group}_1^f, \text{Group}_2^f)\|$ is, the more discriminative the feature f is. We will leave the first $m\%$ of the feature and discard the others. The m is determined by the programmers.

Figure 4.13 shows the t-statistic of Morlet wavelet coefficients of data G01. We reserve the first 0.1% ($\alpha = 0.001$) discriminative features, which is the non-zeros value in the three maps, and discard the others, set to zeros. The higher the absolute value of the t-statistic is, the more discriminative the feature is.

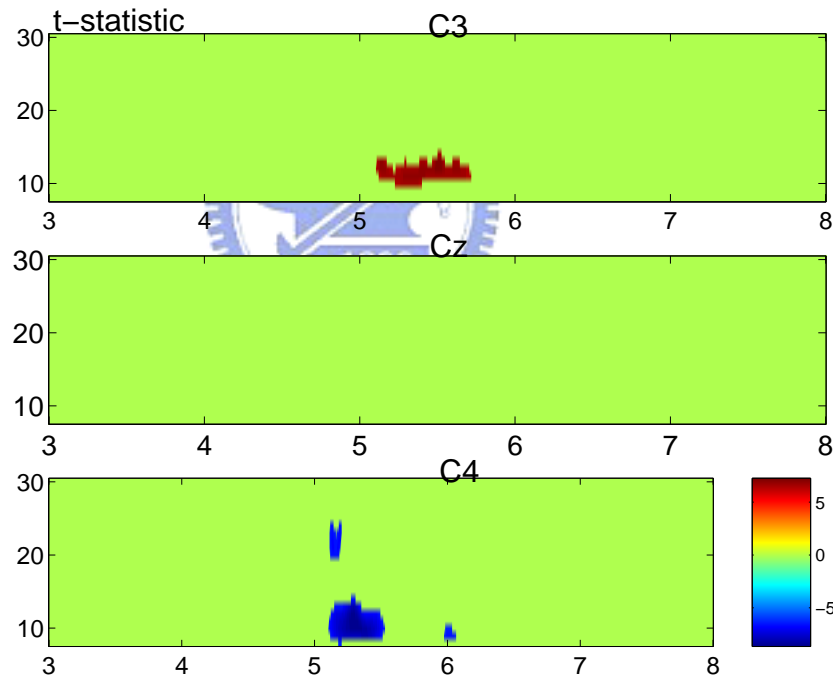


Figure 4.13: Show the t-statistic of Morlet wavelet coefficients of data G01. The Morlet wavelet coefficients have been illustrated in Figure 4.10. We reserve the first 0.1% ($\alpha = 0.001$) discriminative features, which is the non-zeros value in the three maps, and discard the others, set to zeros. Total 926 numbers of features are left and discard 78562 features. The higher the absolute value of the t-statistic is, the more discriminative the feature is.

4.3.2 Forward feature selection

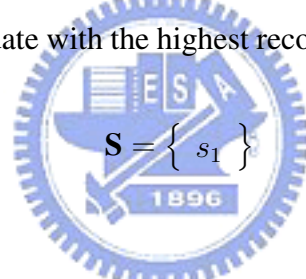
The principle of feature forward selection, also called *sequential forward selection*, is very simple. It is a *greedy algorithm*. At any iteration, feature forward selection always selects the feature such that when it emerges the selected feature subset in the previous iteration, the new feature subset can be with the highest discriminability than any other feature to be emerged into the subset. The algorithm is illustrated as follows:

Assume the feature set is

$$\mathbf{F} = \{ f_1, f_2, \dots, f_d \}$$

1 Using classification technique \mathbf{C} and *leave-one-out* to estimate the recognition error. Here we use *k-nearest neighbor* classifier with $k = 1$ to determine the classification accuracy.

2 The first feature s_1 we select must be the one with the highest recognition accuracy. That is if the f_i is the candidate with the highest recognition accuracy, we set $s_1 = f_i$ and



$$\mathbf{S} = \{ s_1 \}$$

in the first iteration.

3 We emerge picked feature f_i with the feature subset

$$\mathbf{S} \{ s_1, s_2, \dots, s_{i-1} \}$$

in the previous iteration. Then we get the new feature set

$$\mathbf{S} \{ s_1, s_2, \dots, s_i \}$$

4 Repeat the step3 until we have picked up all features.

Although feature forward selection guarantees that in every iteration, the new selected feature is the best candidate to emerge the former feature subset such that the error can be lower than any other one to be emerged into the former feature subset, it cannot ensure the feature subset we selected is the best features to discriminate these groups. It may be happened that the combination of the worst discriminative features, yet, is the best features

to discriminate these groups. However, if we want to find out the best features, we have to try all the combinations of all feature components. This is very time-consuming and impractical when the number of feature components is far great. Therefore, in spite of the possibility of selecting the secondary most discriminative feature subset, the feature forward selection is still a reliable approach to select features.

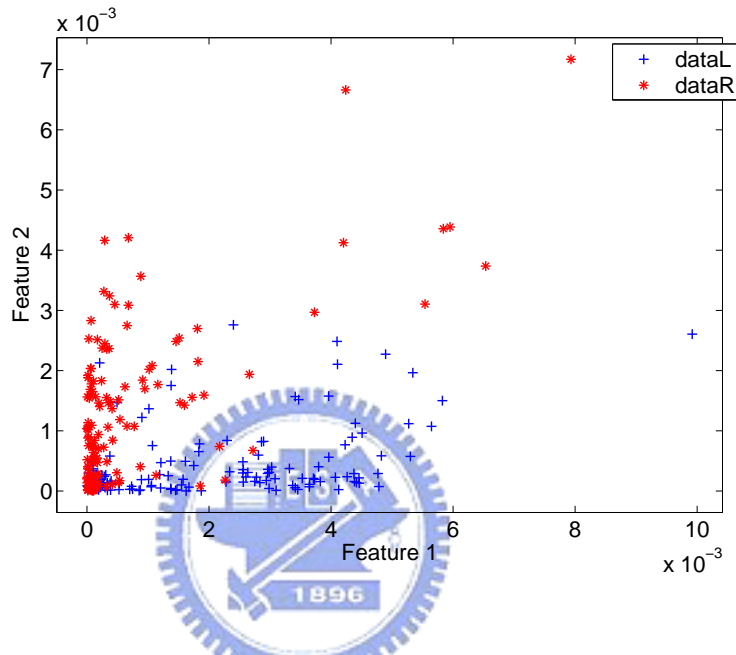


Figure 4.14: After applying Morlet wavelet transformation and t-statistic, we use forward feature selection to select two features, f_1 and f_2 . This figure shows the distribution of the dataset G01 in feature f_1 and f_2 . As we can see in the figure, the feature points of the two groups are discriminative.

4.4 Classification

What would you do if you were given some data of two groups and asked to classify them? In fact, this is a problem of *pattern recognition*. In a classification problem, we often use the collection of available data to train a specific classifier to recognize them. That is if we have dataset χ^{tr} :

$$\chi^{tr} = \{(\mathbf{x}_i, y_i) | i = 1, \dots, N\} \quad (4.20)$$

where \mathbf{x}_i represent the object i and y_i represent the label of \mathbf{x}_i . Then for the classification, a function $f(x)$ has to be inferred from the training data. The function of $f(x)$ in a two-class classification problem should be constructed to given an estimate label $y = f(x)$ when given a testing object x .

$$f : R^d \rightarrow \{-1, 1\} \quad (4.21)$$

where $\{-1, 1\}$ represents the labels of the two classes.

Table 4.1: The four situations of classifying an object in two-class classification.

	Object from target class	Object from outlier class
Assign class target to a object	true positive (TP)	false positive(FP)
Assign class outlier to a object	false negative(FN)	true negative(TN)

In Table 4.1, we describe the four situations of classifying an object in two-class classification. Usually, the classifier have to maximize true positive rate and simultaneously minimize the false positive rate. Today, there has existed various classification techniques. In the following we will describe the support vector classifier (SVC) applied to classify these mental tasks in this work.

4.4.1 Support vector classifier

SVC is a popular classifier. The basis principle of SVC is to find two decision hyperplanes such that the *margin* is largest. The margin is defined as the width of the largest "tube", which contains no samples as Figure 4.15. This figure shows an example of a linear support vector classifier for two classes. The two decision hyperplanes are $w^T z_n + b = 1$ and $w^T z_n + b = -1$. The linear classifier $w^T z_n + b = 0$ is sought such that for all n the following is satisfied

$$w^T z_n + b \geq 1 \quad \text{if } C_n = 1 \quad (4.22)$$

$$w^T z_n + b \leq -1 \quad \text{if } C_n = -1 \quad (4.23)$$

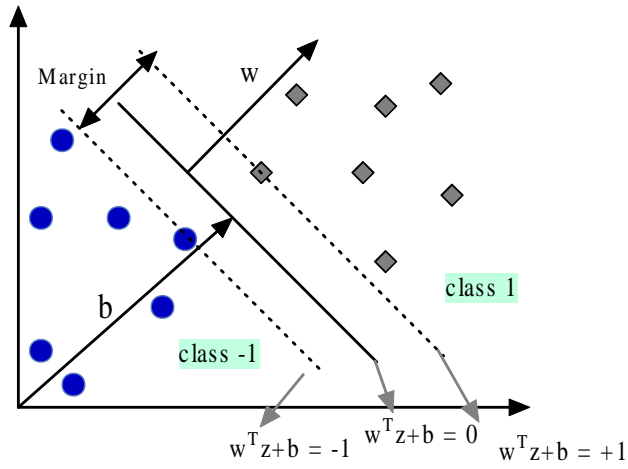


Figure 4.15: A linear support vector classifier for two classes. The support vector classifier find two decision hyperplan such that the margin is largest to classify the groups.

We can rewrite the two constrains Eq. 4.22, Eq. 4.23 into:

$$C_n(w^T z_n + b) \geq 1 \quad (4.24)$$

In order to maximize the margin, $\|\mathbf{w}\|^2$ should be minimized. Introducing *Lagrange multipliers*, α_n , we can obtain the solution of Eq. 4.24 by minimizing L :

$$L = \frac{1}{2} \|\mathbf{w}\|_2^2 + \sum_{n=1}^{N_s} \alpha_n (c_n (\mathbf{w}^T z_n + b) - 1) \quad \alpha_n \geq 0 \quad (4.25)$$

Set the partial derivates of L to 0 gives the constraints Eq. 4.26 and Eq. 4.27:

$$\mathbf{w} = \sum_{n=1}^{N_s} \alpha_n c_n z_n \quad (4.26)$$

$$\sum_{n=1}^{N_s} \alpha_n c_n = 0 \quad (4.27)$$

Substitute Eq. 4.26 and Eq. 4.27 into Eq. 4.24, we then can obtain the the so-called *dual form*:

$$L = \sum_{n=1}^{N_s} \alpha_n - \frac{1}{2} \sum_{n,m=1}^{N_s} c_n c_m \alpha_n \alpha_m (\mathbf{z}_n \cdot \mathbf{z}_m), \quad \alpha_n \geq 0 \quad (4.28)$$

if we apply a transformation Φ to all samples, mapping the original data space to a new feature space which improving the discriminativity before seeking the linear classifier $w^T z_n + b = 0$. This can be represented as:

$$\mathbf{x}^* = \Phi(\mathbf{x}) \quad (4.29)$$

Substitute Eq. 4.29 into Eq. 4.28, now we obtain:

$$L = \sum_{n=1}^{Ns} \alpha_n - \frac{1}{2} \sum_{n,m=1}^{Ns} c_n c_m \alpha_n \alpha_m (\Phi(\mathbf{z}_n) \cdot \Phi(\mathbf{z}_m)), \quad \alpha_n \geq 0 \quad (4.30)$$

We define a new function with two inputs called kernel function:

$$K(\mathbf{z}_n, \mathbf{z}_m) = \Phi(\mathbf{z}_n) \cdot \Phi(\mathbf{z}_m) \quad (4.31)$$

Resubstituting Eq.4.31 into Eq. 4.30 obtain:

$$L = \sum_{n=1}^{Ns} \alpha_n - \frac{1}{2} \sum_{n,m=1}^{Ns} c_n c_m \alpha_n \alpha_m K(\mathbf{z}_n, \mathbf{z}_m), \quad \alpha_n \geq 0 \quad (4.32)$$

Eq. 4.32 seems to replace inner product by a more general kernel function, leading to kernel-based SVM. Various kernel functions have been expended to the classification problem. Typical functions are polynomial kernels and Gaussian kernel. The Gaussian kernel with $\sigma^2 I$ as weighting function, also called *radial basis function kernel (Rbf)* is frequently used:

$$\mathbf{K}(z_n, z_m) = \exp\left(-\frac{\|\mathbf{z}_n - \mathbf{z}_m\|}{\sigma^2}\right) \quad (4.33)$$

When σ is small, the boundary of the classifier is very detailed. On the contrary, when σ is large, the boundary of the classifier is very smooth. Such kernel support vector classifier is often a non-linear classifier.

The advantages of SVC are :

- The performance is very competitive.
- It's parameters have a unique global optimum solution.
- When performing a nonlinear classifier, the extra computation is not too much.

The drawbacks are

- The complexity is depending on the number of samples, not on the dimension of sample space. When the number of samples is large, it may be not practical.
- SVM is a two-class classifier. When applying to more than two classes, we often use pairwise methods to handle multi-class classification problem.

4.5 Experiment results

In the section, the approaches introduced in the previous sections were applied to seven datasets. Here we shows the experiment results of using different reference-free spatial filters and feature extraction approaches. we will compare the performance of using different combination of reference spatial filter and feature extraction approaches. Except reference spatial filter and feature extraction approaches, all the other processing of the seven datasets, including artifact removal and bandpass filter for signal preprocessing, t-statistic for feature selection and finally SVM for classification, are the same. The results are illustrated in four tables, Table 4.2, Table 4.3, Table 4.4 and Table 4.5. We find that the recognition accuracy of different combination is:

$$(smallLaplacian, Haar) > (smallLaplacian, Morlet) > (smallLaplacian, CSP) > (smallLaplacian, ERD/ERS).$$

In these four tables, it reveals that ERD/ERS gives the worst accuracy; yet ERD/ERS should be the very good features for the two mental tasks. In effect, before doing these experiment analysis, we expected the accuracy of ERD/ERS should be satisfied, at least not so bad as the analysis result. We tried to look for the reasons and find out that this may be caused by the bad EEG raw datas. We find that the original raw data vary very seriously, which may result from the bad acquisition or the bad performance of mental tasks. Therefore, we can not get good performance if the ERD/ERS pattern of every trial is so different.

For time-scale and time-frequency analysis, the accuracy is better than ERD/ERS and CSP. This may be due to that the extracted features has more information than ERD/ERS

and CSP or that the two analysis approaches can extract better features to distinguish the two tasks.

Table 4.2: The accuracy of synchronous BCI classification by using ERD/ERS.

Subject.	$(A1 + A2)/2$	Laplacian	Bipolar
D1	56.43	65.38	51.92
D2	48.00	60.36	51.43
D3	55.00	56.00	51.50
D4	48.89	52.14	55.00
D5	48.00	51.67	49.44
D6	60.56	61.36	55.56
D7	56.67	49.17	56.67
Mean	53.36	56.58	53.07

Table 4.3: The accuracy of synchronous BCI classification by using CSP.

Subject.	$(A1 + A2)/2$	Laplacian	Bipolar
D1	67.14	64.29	51.07
D2	51.54	51.54	46.92
D3	55.50	55.00	51.50
D4	48.57	47.86	50.00
D5	71.11	79.44	45.56
D6	46.67	56.82	50.56
D7	67.08	80.83	54.58
Mean	58.23	62.25	50.03

4.5.1 Summary

In this chapter, we present several approaches to analyze EEGs in a synchronous BCI. In section 4.5, some experiments are done to compare these approaches. Among the various combinations of reference spatial filter and feature extraction methods, the Laplacian filter

Table 4.4: The accuracy of synchronouse BCI classification by suing Haar wavelet transformation.

Subject.	$(A1 + A2)/2$	Laplacian	Bipolar
D1	77.50	85.36	81.79
D2	78.33	88.33	83.33
D3	75.00	90.83	77.92
D4	81.87	94.37	83.13
D5	75.56	84.44	76.11
D6	82.73	88.64	85.00
D7	78.33	87.50	82.08
Mean	78.47	88.49	81.33

and Haar wavelet will give the best recognition accuracy. This is practical when real time analyzing EEGs for the implementation of Haar wavelet is easy and computation complexity is low.

The contribution of this chapter is to build a processing module of analyzing EEGs for the predefined mental tasks, left and right hand motor imagery. The processing module will be delivered to the procedures of analyzing asynchronous BCI. In next chapter, we give the details about how to analyzing EEGs in an asynchronous BCI system based on the results of this chapter.

Table 4.5: The accuracy of synchronous BCI classification by using Morlet.

Subject.	$(A1 + A2)/2$	Laplacian	Bipolar
D1	78.93	85.36	82.86
D2	67.31	71.15	73.08
D3	75.00	72.00	79.00
D4	62.14	69.29	66.43
D5	77.22	75.00	80.56
D6	67.22	77.73	77.78
D7	78.75	82.08	78.75
Mean	72.37	76.09	76.92

Chapter 5

Continuous EEG classification



In the previous chapter, we have illustrated the approaches of analyzing synchronous BCI systems and given several experiments to discuss and compare them. In this chapter, we will take the method with the best performance in the the previous chapter to analyze the asynchronous BCI classification.

We have already mentioned the obstruction of analyzing EEGs in a synchronous BCI system. All these hindrance should be overcome in an asynchronous BCI system, also. In addition to these mentioned problems, there are other difficulties encountered in an asynchronous BCI system. We list them as follows.

- 1 We has no information about when the tasks are performed, thus the analysis should be continuously performed to detect the onset time of a task. This is very different from that in a synchronous BCI system, where we only have to analyze the signals during a predefined period.
- 2 Unlike the analysis in a synchronous BCI, we have to not only recognize these predefined mental tasks, but also distinguish event states from nonevent states(/resting/ idle states). That means we have to deal with three classes, two active tasks and one for nonevents.
- 3 The high variance of resting states makes it hard to recognize between active and resting states because we can not obtain a clear rule to characterize the EEGs of resting states.

How to settle the three above problems is the aim of this chapter. In this chapter we will present our classification approaches for an asynchronous BCI system, which is the kernel of the thesis. This chapter is organized in five sections. First, one-class classification method is described in section 5.1, which is a technique of data description. We only illustrate a brief overview of this technique and if necessary, see the all details in [50]. Besides these illustration, the discussion of why we use one-class classification to handle our problem is given in section 5.1.1. Then, section 5.2 describe the perspectives of the three one-class classifier used in our work. In section 5.3, we propose a method based on one-class classifier to recognize the three classes, left, right and resting states. After these descriptions, we will discuss some important considerations when recognizing EEGs in an

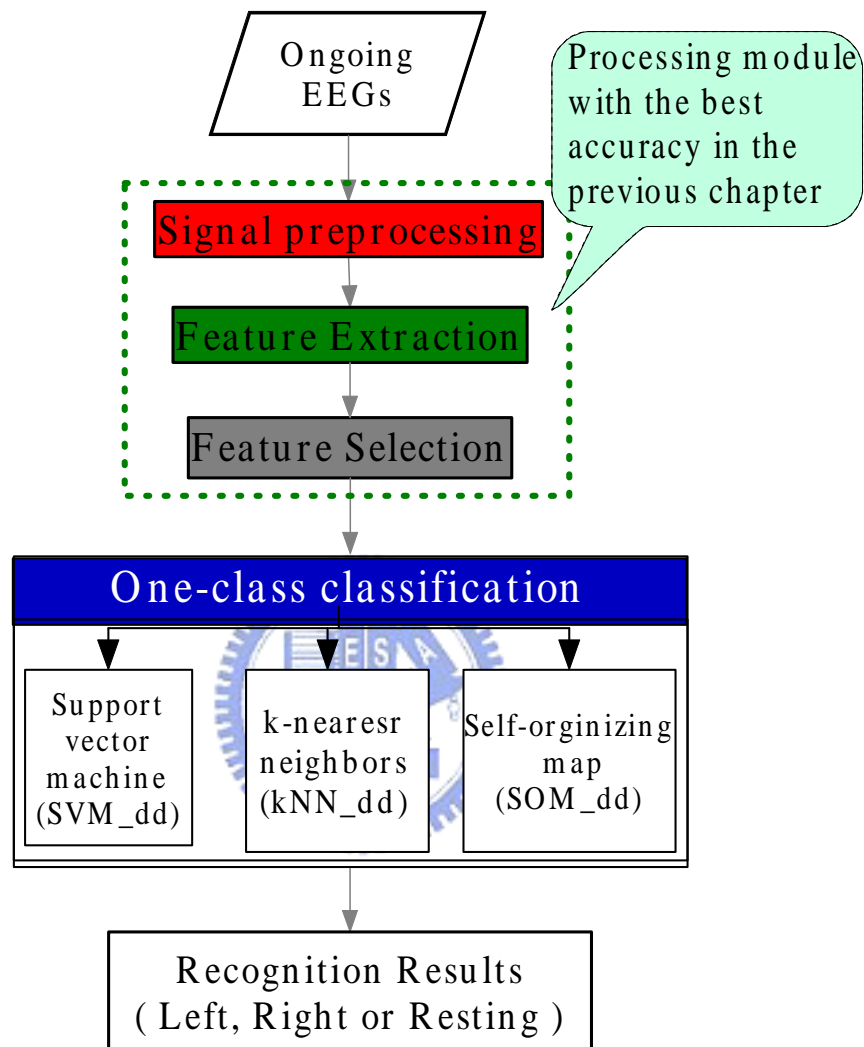


Figure 5.1: Flowchart of analyzing asynchronous BCI. We have two inputs, the raw EEG data and the processing module with best accuracy in the previous chapter. Because we have to continuously process signals, the method of sliding window is imported. These processed data(/feature vector) will be delivered to a classifier based on technique of one-class classification to recognize EEGs between these three mental tasks, left/ right hand motor imagery and resting states.

asynchronous BCI system in section 5.4. Finally, the experiments of the asynchronous BCI classification are given in section 5.5. In Figure 5.1, we give a flowchart to illustrate an overview of this chapter.

5.1 One-class classification

The literature has a lot of different terms to be used. The term of one-class classification is proposed by Moya (Moya et al.,1993), while the outlier detection (Ritter and Gallegos, 1997), novelty detection (Bishop, 1994b) are also used. These different terms originate from different applications but the kernel of all are one-class classification. In this section we just give the brief principles of one-class classification while the details can be found in [50].

In a one-class classification problem, there are two classes, *target class* and *outlier class*. Our task is to give a model to describe the target class and detect which object belongs to the target class. We first give an example to show the usefulness of one-class classification and then go about describing some perspectives.

We first give an example in the following to give some ideas about two-class and one-class classification. Given a collection of apples, pears, we are asked to detect apples from the pears. In this case, the typical solution is to build a classifier based on the conventional two-class problem to recognize the two types of fruits. This is not a complicated problem. However, if we are asked to distinguish apples from all the other types of fruits, such as bananas, plums, guavas ...etc., using a typical classifier to settle the problem seems not practical. In a conventional solution of the problem, we may built a model for every other fruits expects apples or build a model to characterize all the other fruits. The former seems consuming and meaningless. As for the latter, because the properties between all the other fruits besides apples is so different that we have high probability to obtain a bad model. In such condition, using the technique of one-class classification to settle the problem seems very attractive.

We proceed to deal with how the one-class classifier detect a target from the outliers. In a one-class classification problem, a new object is accepted only when the distance of

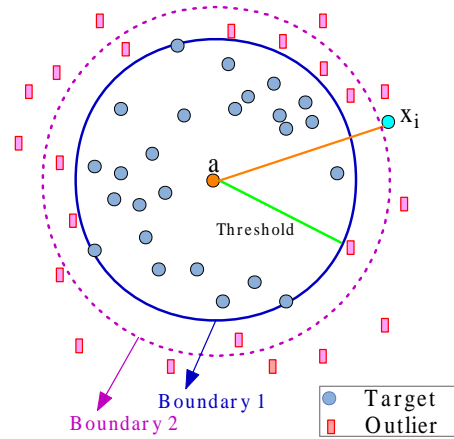


Figure 5.2: Threshold of one-class classification. If the boundary is 'Boundary 1', then x_i will be detected as an outlier, which is a FN. However, if the boundary is 'Boundary 2', then many outliers will be detected as a target, which is a FP. Therefore, different thresholds will result in different FN and FP.

the object to the target class is smaller than a threshold θ . The distance function differs with respect to different one-class classification methods. Much effort has been expended to develop methods to solve the one-class classification problem. There are three main approaches: the density estimation, the boundary methods and the reconstruction methods.

As for the threshold, it is the tradeoff between the fraction of the target class that is accepted, TP, and the fraction of the outlier that is rejected, FN, seeing Figure 5.2. The threshold θ is determined by a given fraction f_{TP} of the target class [50], as Eq. 5.1. Using Eq. 5.1 we can compute a threshold $\theta_{f_{TP}}$ on the training set for different f_{TP} , and measure the f_{FP} on a set of example outliers, thus obtaining a Receiver-Operating Characteristic curve (Metz, 1978).

$$\theta_{f_{TP}} : \frac{1}{N} \sum_i I(p(x_i) \geq \theta_{f_{TP}}) = f_{TP} \quad (5.1)$$

where $x_i \in \text{chi}^{\text{training}} \forall_i$

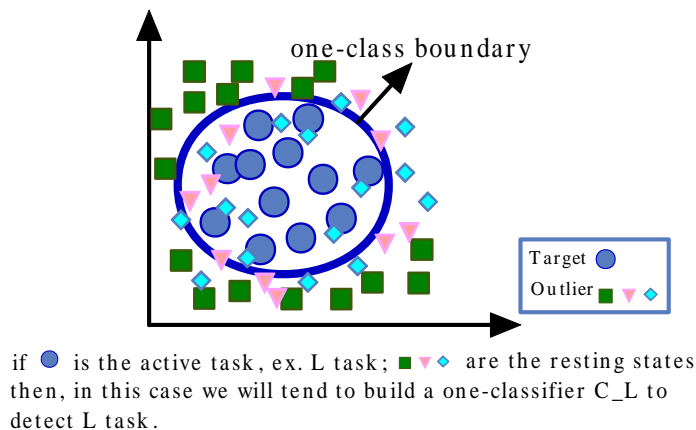


Figure 5.3: An example of one-class.

5.1.1 Why one-class classification

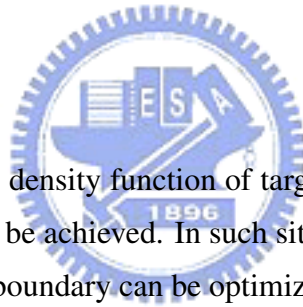
There are two main reasons we prefer one-class classifier. The first is that the wide-spreading distribution of the resting state due to the composition of many slight and fickle ongoing EEGs such as thinking, idling, and fixation. In the situation, it is hard to build a conventional model to detect resting states well. Second, resting states between training and testing phases are significantly different. This is because during testing phase, the user have to pay strong attention on the feedback and therefore the user is at rather active state despite of having no actual control intention. In this situation, we think that modeling a classifier to include resting data may have not much help to solve the continuous classification problem.

Based on the two reasons, we think that if we can build a classifier to positively recognize the two active tasks, implying that the classifier will return zero for nonactive states and one for active states, we have more possibility to deal with the problem well than the conventional methods. Therefore, using one-class classification to handle this problem seems attractive because the one-class classification always focuses on how to obtain a closed boundary around only one class (the target class or the active class), seeing Figure 5.3. This differs from the conventional two-class classification because for the latter, the constraint is to minimize the probability of classification error depending on the distribution of all classes.

Originally, the one-class classification is used to solve the classification problem when one class is severely undersampled or even completely absent. The main difference between one-class and two class is that the former assume only the target class is available. In the thesis, we expand the one-class classification to settle the recognition of our problem, although the outliers are available. We viewed the active tasks as target objects while resting states as outliers. In the section 5.3, we propose one classification methods based on one-class classifier to distinguish the three classes, left, right and resting states.

In the next section, we describe three one-class classifier, SVM_dd, kNN_dd and SOM_dd which fall into boundary-based, boundary-based and reconstruction-based method. The 'dd' here represent 'data description'. We don't use any density-based methods in this thesis because in such methods, it needs a large number of samples; however, the number of EEG trials for all mental tasks is relative small.

5.2 Classifier



To estimate a complete density function of target set needs a numerous samples; however, sometimes, it can not be achieved. In such situations, we turn to estimate a boundary around the target set. The boundary can be optimized by means of several methods associated with different boundary-based one-class classification, such as SVM_dd and kNN_dd.

Besides seek a boundary to describe the target set, we can also build a model to fit the data. This can be achieved by reconstruction methods, such as SOM_dd.

In the following, we will illustrate two boundary methods, SVM_dd and kNN_dd, and one reconstruction method, SOM_dd used in this work to discriminate these mental tasks and resting states.

5.2.1 Support vector classifier data description

Here we use a hypersphere as the boundary around the target set, shown in Figure 5.4, to illustrate the concepts of SVM_dd. This hypersphere is centered at \mathbf{a} with radius R . We wish the hypersphere can include all target objects. Assume target object is enumerated by

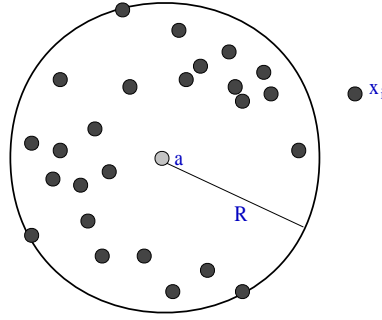


Figure 5.4: A hypersphere around the target data. This hypersphere is characterized by \mathbf{a} and R , where \mathbf{a} is the center and R is the radius. The solid circle represents the target object and one object x_i is outside the boundary.

indices i . Then for a target object \mathbf{x}_i , this constraint can be written as:

$$\|\mathbf{x}_i - \mathbf{a}\| \leq R^2 \quad \forall i \quad (5.2)$$

After introducing Lagrange multipliers and a series of evaluations, seeing details in [50], we can obtain the final decision rule that a new object \mathbf{z} is accepted when:

$$\|\mathbf{z} - \mathbf{a}\| \leq R^2 = ((\mathbf{z} \cdot \mathbf{z}) - 2 \sum_i \alpha_i (\mathbf{z} \cdot \mathbf{x}_i) + \sum_{i,j} \alpha_i \alpha_j (\mathbf{x}_i \cdot \mathbf{x}_j)) \leq R^2 \quad (5.3)$$

where α_i, α_j is Lagrange multipliers.

As in section 4.4.1, we also introduce the general kernel function and then obtain some kernel-based SVM_{dd} classifiers, such as Gaussian-kernel-based and polynomial-kernel-based. Using these kernel function can often improve the fit of the data and obtain more robust boundary. In the following analysis, we will use Gaussian kernel SVM_{dd} to recognize mental tasks. The Gaussian kernel has been described in Eq. 4.33. In Eq. 4.33, the parameter σ will influence the fit of the dataset. Given a specific target acceptance fraction f_{TP} , an optimal σ can be determined.

5.2.2 k-nearest neighbor data description

This section is organized in two parts, conventional kNN and the kNN_{dd} based on one-class classification. We will first introduce the conventional kNN classification technique

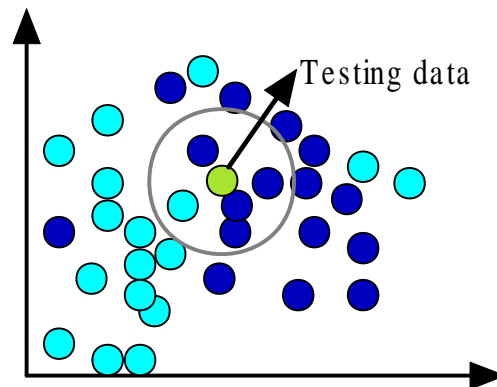


Figure 5.5: One example of kNN with $k = 5$. In the figure, the five nearest neighbors consist of two of class1(diamond) and three of class2(circle). Therefore, we assign class2 to the testing data.

to afterwards bring out the perspectives of kNN_dd.

k-nearest neighbor

The conventional kNN is of most practical interest because it doesn't need to estimate the probability. The principle of kNN is very simple. We sketch the algorithm of kNN as follows [53]:

Assume the training data set is D_{train} and use the Euclidean distance to calculate the distance between two samples. We want to classify a sample \mathbf{y} .

- 1 Find the *k-nearest neighbors* of \mathbf{y} . That means finding the k sample points with the first k shortest distance to \mathbf{y} in D_{train} .
- 2 Let N_i is the number of samples found with class C_i . Then the class C_j will be assigned to \mathbf{y} if N_j is greater than other N_i where $i = 1 \dots K$, and K is the number of these groups.

Figure 5.5 shows an example of kNN with $k = 5$.

How to choose the value of k is an important factor in a classification system. When k is large or the dimension of sample space is great, calculate the distance between samples is much expensive cost. Besides k , the computation complexity is also depending on the

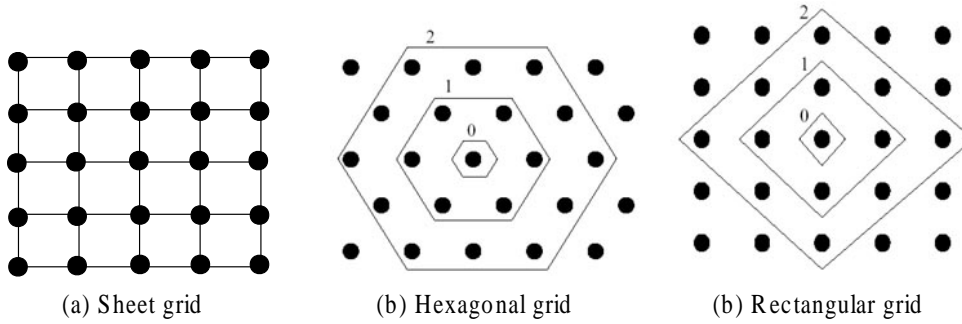


Figure 5.6: Three examples of lattice grids: (a) sheet grid (b) hexagonal grid (c) rectangular grid. For the (b) and (c), the innermost polygon represents the 0-neighborhood, the second represents the 1-neighborhood and the outmost is the 2-neighborhood [54].

number of samples and on the dimension of sample space. When the number of samples or the dimension of sample space is large, kNN may be not practical.

k-nearest neighbor data description

Almost all the properties of kNN_{dd} is the same with the kNN besides the following two properties.

1. A threshold θ is given.
2. In this type classifier, a testing object, z is accepted only when the distance from z to its first nearest neighbor, $NN^t_r(z)$, is smaller than the distance from its nearest neighbor $NN^t_r(z)$ to its nearest neighbor. That is z belongs to target class if

$$\|z - NN^t_r(z)\| \leq (\theta \times \|NN^t_r(z) - NN^t_r(NN^t_r(z))\|) \quad (5.4)$$

Beside using the distance $\|z - NN^t_r(z)\|$, the mean of distances to the k nearest neighbors or the distance to the mean of distances to the k nearest neighbors can be considered as well. These latter two approaches lead to a more robust classifier.

5.2.3 Self-organizing map data description

As section 5.2.2, we first present the perspectives of conventional SOM and then introduce the principles of SOM_{dd}. In effect, the main difference between the two methods is

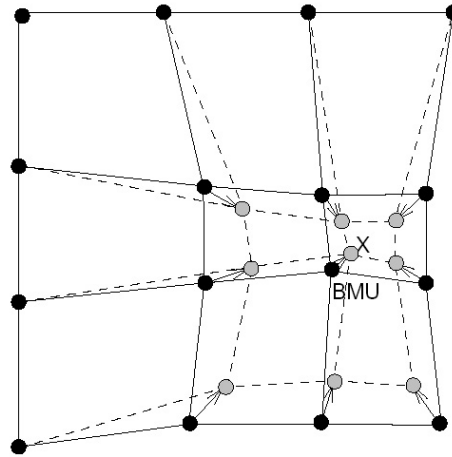


Figure 5.7: Update the weight vector and topological neighbors of the BMU such that the BMU and its topological neighbors move closer to x . The solid and dashed line represent the situation before and after updating, respectively [54].

the concept of the given threshold in SOM_{dd}.

Self-organizing map

SOM is an unsupervised neural network method. It consists of neurons represented by a d -dimensional weight vector where d is the dimension of sample vectors. These neurons are organized on a regular-dimensional grid. Examples of the grid can be sheet, rectangular and hexagonal lattice structures as shows in Figure 5.6. In the following, we will give a iterative algorithm of training SOMs to show how the SOMs organize themselves to cluster data on the map.

Sequential training algorithm This is an iterative algorithm. First, an object, \mathbf{x} , is chosen from the input data. We select the neuron, \mathbf{c} , with the shortest distance from \mathbf{x} to it. We call \mathbf{c} the Best-Matching Unit (BMU) [54]. After finding the BMU, we update the weight vector and topological neighbors of the BMU such that the BMU and its topological neighbors move closer to \mathbf{x} . This can see in Figure 5.7. The weight vector can be updated by:

$$\mathbf{W}_i^{(t+1)} = \mathbf{W}_i^t + \alpha(i)h(t)[\mathbf{x}(t) - \mathbf{W}_i(t)] \quad (5.5)$$

The $\mathbf{x}(t)$ is the input data, $\alpha(i)$ the learning rate and $h(t)$ the neighborhood kernel around the BMU c at time t .

The neighborhood kernel $h(t)$ weights how much a particular neuron in the grid is moved toward x . Usually $h(t)$ is a non-increasing function of time and of the distance of neuron i from the BMU c . An example of $h(t)$ is Gaussian function.

Self-organizing map data description

In SOM_dd, the target objects are represented by the nearest neuron (or prototype). Most often the distance is measured by the Euclidean distance. A object will be assigned target if the distance from it to the target neuron is smaller than the threshold.

5.3 Classifier design

After deciding using one-class classification to settle the problem in this work, we proceed to consider how to design the classifier based on one-class classification. There are two methods to settle the recognition of the three classes, left, right and resting states.

- A. At training phase, we first train a one-class classifier to distinguish between all active tasks and resting states. Then we follow to train a two-class classifier to discriminate left and right tasks. At testing phase, we first classify the testing data to be active or resting state. If active, we go on to discriminate the testing data to left or right motor imagery. See Figure 5.8.
- B. At training phase, we train two one-class classifiers. One is to classify between $\{left(target), right\ and\ resting\ states(outlier)\}$. The other is to classify between $\{right(target), left\ and\ resting\ states(outlier)\}$. At testing phase, we feed the testing data to the two one-classifiers to get $ResultA$ and $ResultB$. See Figure 5.9. The testing data will be assigned class $ResultC$ depending on Table 5.1, named as **Criterion1**.

We finally choose Method B to analyze EEGs because in Method.A, the performance of the one-class classifier between resting and active states may be bad. The reason is that

Table 5.1: The relationship of the($ResultA$, $ResultB$) and final assignment $ResultC$ in $MethodB$ when classifying an asynchronous BCI.

.	Left	Resting
Right	<i>Resting</i>	<i>Right</i>
Resting	<i>Left</i>	<i>Resting</i>

the target class (the active state) here consists of two types of data , left and right hand motor imagery. This may lead to a high possibility of a poor fit, arising from the uncluster of original samples. Thus, we prefer to using B than A to handle our problem.

5.4 Implementation issues

Some considerations have to be concerned when implementing the analysis of discrimination between the three classes. In this section, we will take the following six issues into account:

1. Simulation (using cue-based signals to simulate asynchronous BCI analysis)
2. Sliding window
3. Unbalance features
4. Constraints
5. Three analysis:
6. Using Morlet wavelet

5.4.1 Simulation

We first consider how to get the groundtruth of when the user perform a imagination task. It seems that today, we still has no idea about how to determine the onset time of this problem. Thus, almost all the present asynchronous BCI systems use *cue-based continuous*

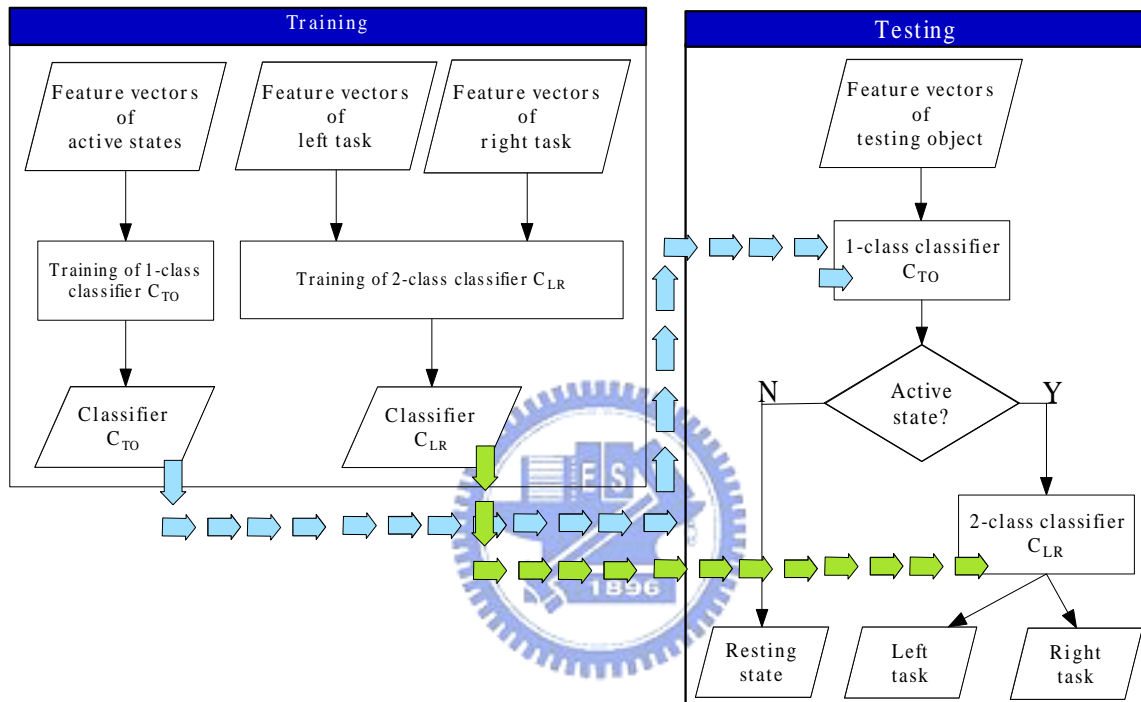


Figure 5.8: This is a method of classifying asynchronous BCI mental tasks in the work. At training phase, we first train a one-class classifier to classify between all active tasks including left and right motor imagery, and resting states. Then we also train a two-class classifier to discriminate left and right motor imagery. At testing phase, we first classify the testing data to active or resting state. If active, we go on to discriminate the testing data to left or right motor imagery.

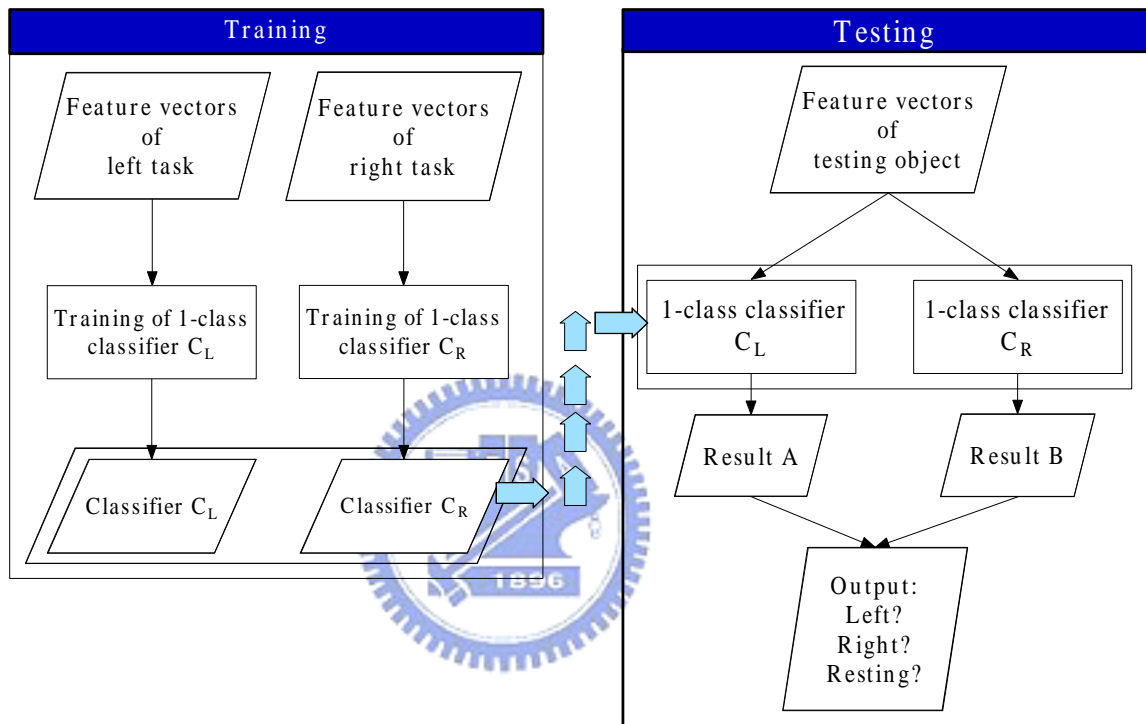


Figure 5.9: At training phase, we train two one-class classifiers. One is to classify between left(target) and, right and resting states(outlier). The other is to classify between right(target) and, left and resting states(outlier). At testing phase, we feed the testing data to the two one-classifiers to get $Result_A$ and $Result_B$. The output $Result_C$ is determined as Table 5.1.

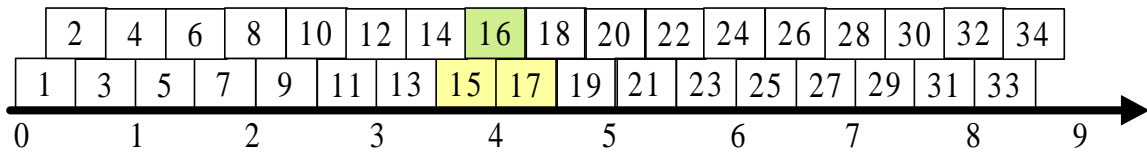


Figure 5.10: A signal consists of 34 sliding windows with sliding window size 0.5s and overlapping 50%. The active state is about at the 16th sliding window (green). However, because the onset of the active task may not be the same over all the trials, for example, a subject may perform a imagery task during 16th sliding window while perform another imagery task during 17th sliding window. Therefore, we think the 15th, 17th may be the secondary candidates periods of performing active tasks.

signals to simulate the asynchronous BCI analysis. This is unavoidable because we have to contain the information about the period during performing the active tasks when we assess the accuracy of the classifier. The experiment paradigm is as mentioned in Chapter 3. The cue-based continuous signals consist of two active states, left and right hand motor imagery, and the resting state. Every trial of the data is of 8s or 9s.

5.4.2 Sliding window

In an asynchronous BCI analysis, the goal is to determine what state of every time point is. However, if what to be analyzed is only the data of only one time point, then how to determine the state of every time point? In effect, this is impossible today because the information is far little. The intuitive approach to achieve the goal is to apply a sliding window at every time point and analyzing the signals in the period. However, this will cause the computation complexity too high. We then consider that when someone perform any mental tasks, his brain need a period of time to activate the corresponding cortex. Thus we think it reasonable to apply a sliding window with overlapping $m\%$ to recognize its state. The parameter m and the size of the sliding window, L , are important factors of the analysis. Indeed, in the following all the intermediate results and classification accuracy are based on the sliding window approach. In Figure 5.10, we show one case of sliding window in the work with $L = 0.5s$, $m = 50\%$. The highlight represent the possible period of performing a imagination mental task.

In Figure 5.10, we view the 15th, 17th as the secondary candidate periods of performing active tasks. How did we select the range of this period with the probability of performing an active task? Indeed, we have performed an experiment to verify the inter-trial variation of cue-onset latency time. We finally obtained the result that the inter-trial variation of cue-onset latency time was about $\pm 360ms$. The experiment we performed is as follows:

Experiment The paradigm is the same as that in Chapter 3 expect that after the user perform the imagination task, we asked him to perform a real movement of finger lifting to record this time. After performing about 200 trials, we calculated the variance of the recorded time to be the inter-trial variation of cue-onset latency time.

Although the estimated inter-trial variation of cue-onset latency time is not very precise, it is still a acceptable rule. Despite of the $\pm 360ms$ inter-trial variation, in our work, we used only $\pm 250ms$ as shown in Figure 5.11 to assess the accuracy performance.

5.4.3 Unbalance features

In this section, what we want to illustrate is that when analyzing in asynchronous mode, the data number of resting states is usually far larger than the data number of active states. Using all resting-state data to select features or to train a classifier will lead to an unbalanced feature set. The other disadvantage is that the computation complexity will be too high. Therefore, we replace using all resting-state data by randomly selecting N data of resting states to train the classifier. The N is usually about the same with the number of active states.

5.4.4 Constraints

As mentioned in section 5.4.2, we view the 15th, 17th as the secondary candidate periods of performing active tasks. Therefore, we consider it reasonable if there exists one classification assigned to the active states during a predefined period such as the 15th 17th sliding windows (see Figure 5.10), then we decide that in this situation, a *hit* is given. That is: Assume we assign '1' to the target class and '0' to outlier class.

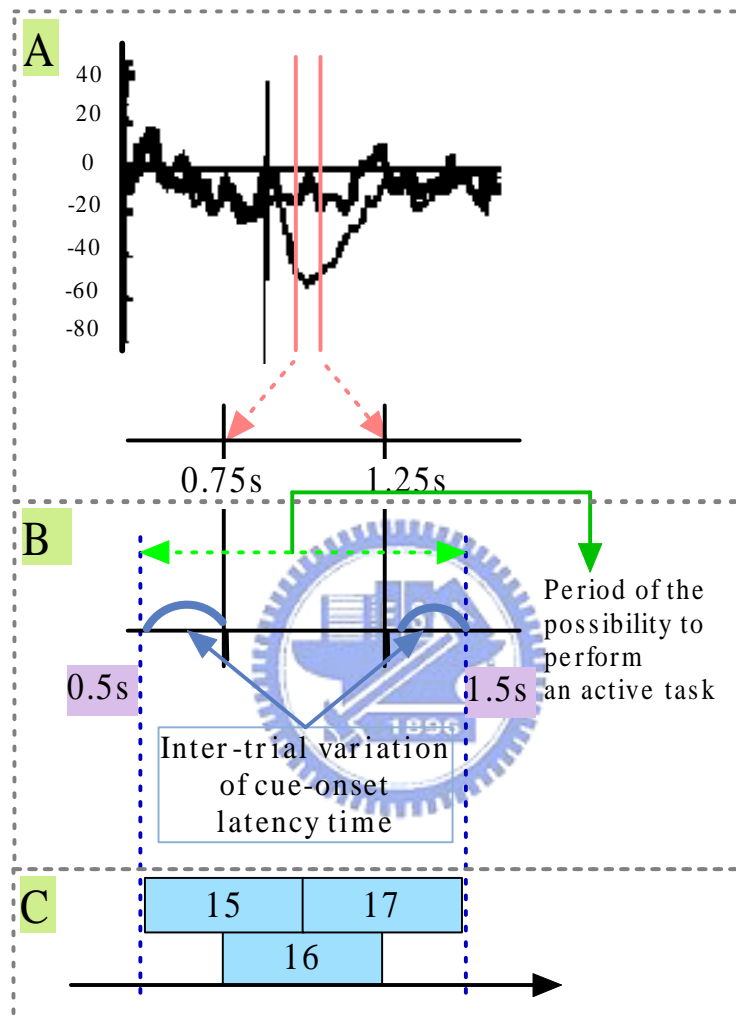


Figure 5.11: The inter-trial variation of cue-onset latency time. Part(A) is a ERD/ERS plot of performing hand motor imagery. We select the maximal ERD as the onset time of an active task. Part(B) illustrates the perspective of the inter-trial variation of cue-onset latency time. Finally, part(C) shows the corresponding sliding windows of the period of performing the task.

$$\text{If } (C_{15} | C_{16} | C_{17}) == 1, \text{ then this is a hit.} \quad (5.6)$$

We name this **Criterion2**.

5.4.5 Five kinds of classification

In our work, three analyses are performed to analyze asynchronous BCI signals. Two of them are the same as those in [52]. First, the left and right motor imagery is tested separately (analysis 1). Then the left motor imagery(target) is compared with right motor imagery and resting periods(outlier) (analysis 2_a) and the right motor imagery(target) is compared with left motor imagery and resting periods(outlier) (analysis 2_b). The final analysis is that the classification between three classes, left and right motor imagery and resting states by using method B in section 5.3 (analysis 3). To sum up, in the work, we will analyze the followings:

1. Analysis1.

(a) Discuss the influence of feature numbers when recognize the left motor imagery and resting states. We will perform the analysis of different feature number from 2 to 7.

2. Analysis2.

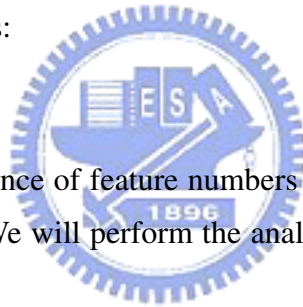
(a) Left motor imagery and resting states. (b) Right motor imagery and resting states.

3. Analysis3.

(a) Left motor imagery(target) is compared with the composition of right motor imagery and resting periods(outlier). (b) Right motor imagery(target) is compared with the composition of left motor imagery and resting periods(outlier).

4. Analysis4.

Classify between three classes: left, right motor imagery and resting states based on classification method B.



When performing analysis 4, assume we adopt all the above considerations; how to determine a *hit* should be concerned. The classification output belongs to the subset $\{r, L, R\}$ where r, L, R stand for resting states, left and right hand motor imagery task respectively. We define C_i the classification output at i th sliding window; there are 3^3 combinations outputs during the 15th 17th sliding windows. For what kinds of combinations will we assign this period to r, L or R ? In the following, we give a simple and intuitive approaches to decide the final output of this period. We name this approach **Criterion3**.

1. If $C_i \in \{R, r\}$ for $i = 15, 16, 17$, then the assignment is 'R'
2. If $C_i \in \{L, r\}$ for $i = 15, 16, 17$, then the assignment is 'L'
3. Otherwise, the assignment is 'r'

5.4.6 Summary

To sum up, section 5.4 propose three criteria when we analyzing asynchronous BCI systems.

Description 1. Table 5.1

Description 2. Eq. 5.6

Description 3. The itemize in the end of section 5.4.5.



5.5 Experiment results

We have mentioned that the aim of synchronous BCI analysis is to build a process order to the analysis of asynchronous BCI system. After Chapter 4 and this chapter, we summary the processes of the analysis by the flowchart, Figure 5.12. Before we go on the experiment analysis, there is an important point we first have to show clearly. In this figure, there are two difference from the Chapter 4 and Chapter 5. The one is that we do not apply Laplacian filter in signal preprocessing step. The other is that the classifier is different depend on what kind of analysis.

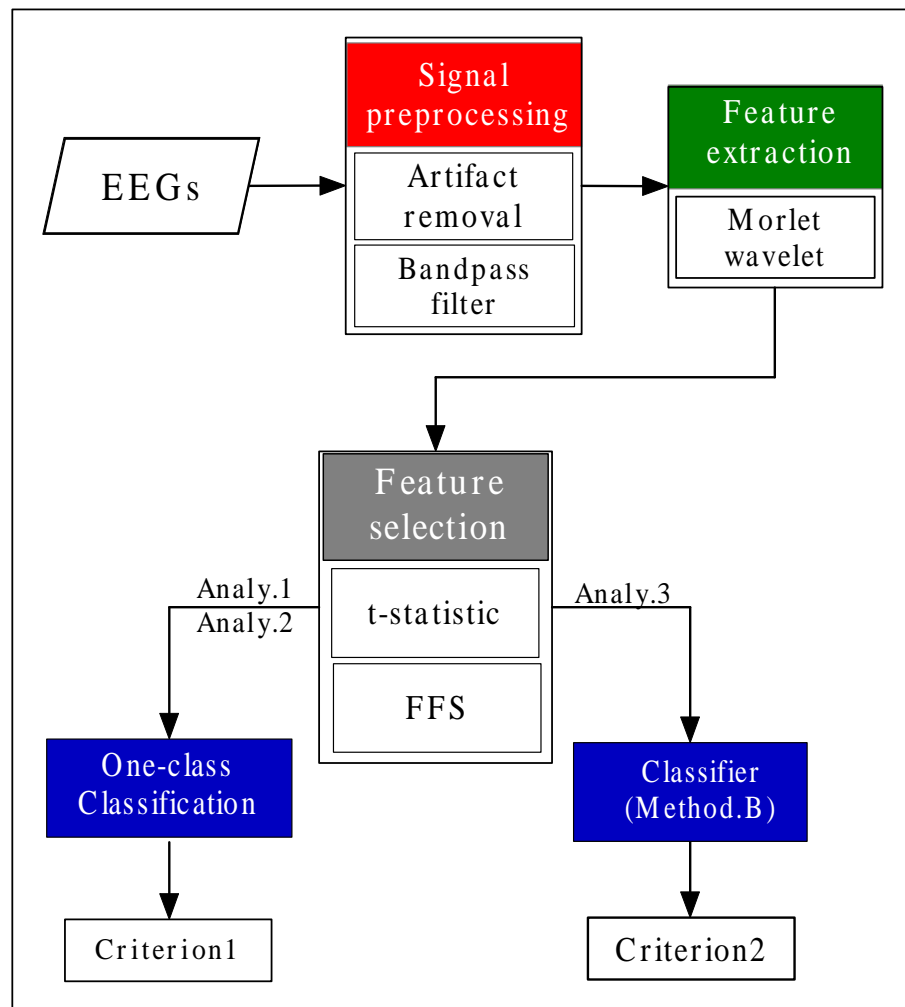


Figure 5.12: Experiment flowchart.

Laplacian filter has the functionality to enhance SNR and to make data reference-free. The reason we don't apply it is that we recognize the mental tasks using only the signals of the three electrodes C3, Cz and C4, implying the number of channels is too small.

In this Chapter, three datasets, G02~ G04, are analyzed in these analyses, Analysis 1~4, mentioned in section 5.4.5. In every aspect, for example analysis 1, the analysis for the three datasets is the same. We use the ROC curve to illustrate the performances of different classifiers with different parameters. Besides, in each ROC plot, we also indicate the classifier with the best classification accuracy.

Before we go about these analysis, we first demonstrate the mean-time-frequency map of every dataset in Figure 5.13, Figure 5.14 and Figure 5.15 to present the property of every dataset. After these time-frequency maps, the results of analysis 1 ~ 4 are shown in Figure 5.16, 5.17, 5.18, 5.19, 5.21, 5.21, 5.22, 5.23, 5.24 and Table 5.2.

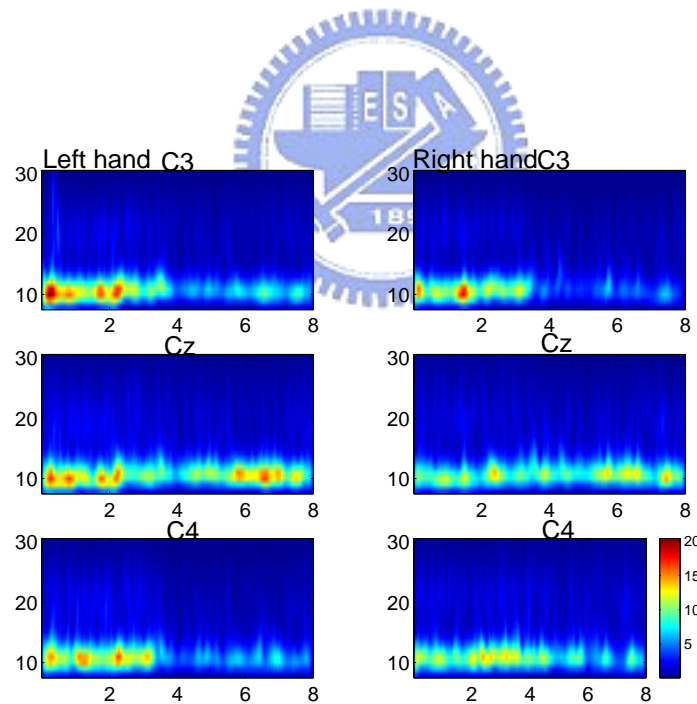


Figure 5.13: Time-frequency map of G02.

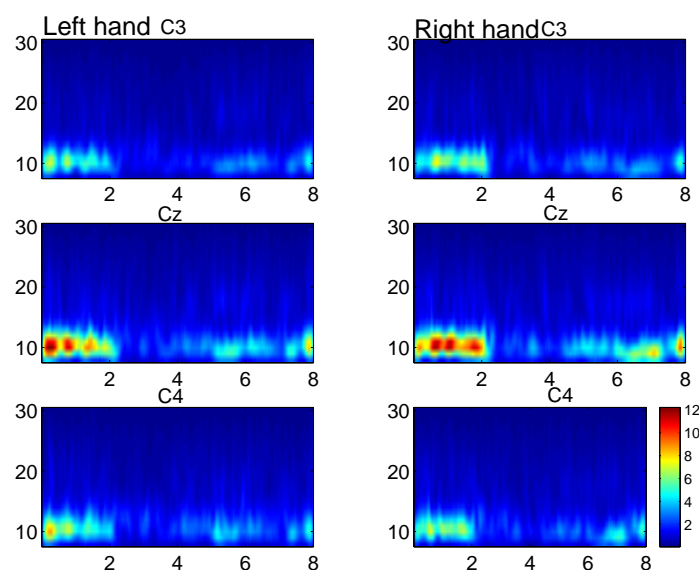


Figure 5.14: Time-frequency map of G03.

5.5.1 Summary

Results of Analysis1

In Figure 5.16, Figure 5.17 and Figure 5.18, we show the analysis results of using different number of features to analyze Analysis1. We discuss the number of features from 2 to 7. As these figures show, in some cases, using more features will give a higher recognition accuracy while in other case, only using 2 features can give the better accuracy than others of using 3~7 features. But in all these analyses, we can find that using two features to recognize Analysis1 can give an acceptable accuracy. Therefore, in the following analysis2 ~ analysis4 we will use only two features to perform the recognition.

Results of Analysis2

We then concern the three figures, Figure 5.19, Figure 5.20 and Figure 5.21 which illustrate the ROC curve of the analyses in Analysis2: (*Left motor imagery, resting states*) and (*Right motor imagery, resting states*) for the three datasets. In the three figures, we can find that not only the different thresholds but also the type of the classifier, 4NN_dd,

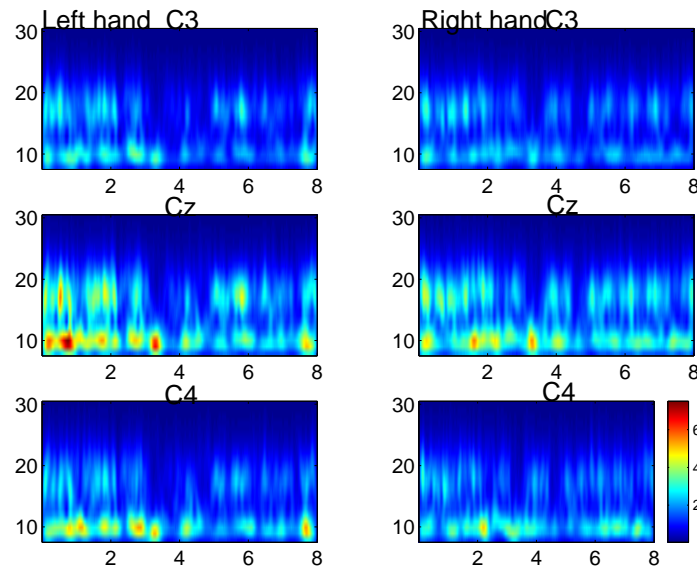


Figure 5.15: Time-frequency map of G04.

SOM_dd, SVC_dd, will influence the classification accuracy. The best classifier is indicated by a arrow with some commentaries including which type of classifier the value of threshold, and FN and FP.

Table 5.3 shows the analysis results of ours and Graz's [51] to give a comparison. For the analysis, we use only the EEGs of three channels, C3, CZ, C4, while Graz use 27 channles. In this table, it reveals that the three analysis results of ours have less variability than those of Graz's, implying that our method has a more stable performance. As for the classification results, the mean of the accuracy of ours and Graz is near the same, implying that despite of using few data information to train the classifier, our method still has a reasonable performance. This is very practice to a online BCI because the computation time is low due to the smaller processed data.

Results of Analysis3

As for Analysis3: (*Left, Right and Resting*) and (*Right, Left and Resting*), the three ROC curves, Figure 5.22, 5.23, 5.24, are displayed. Like the ROC curves in Analysis 1, we also indicate the best classifier in these figures, We finally summary the three figures

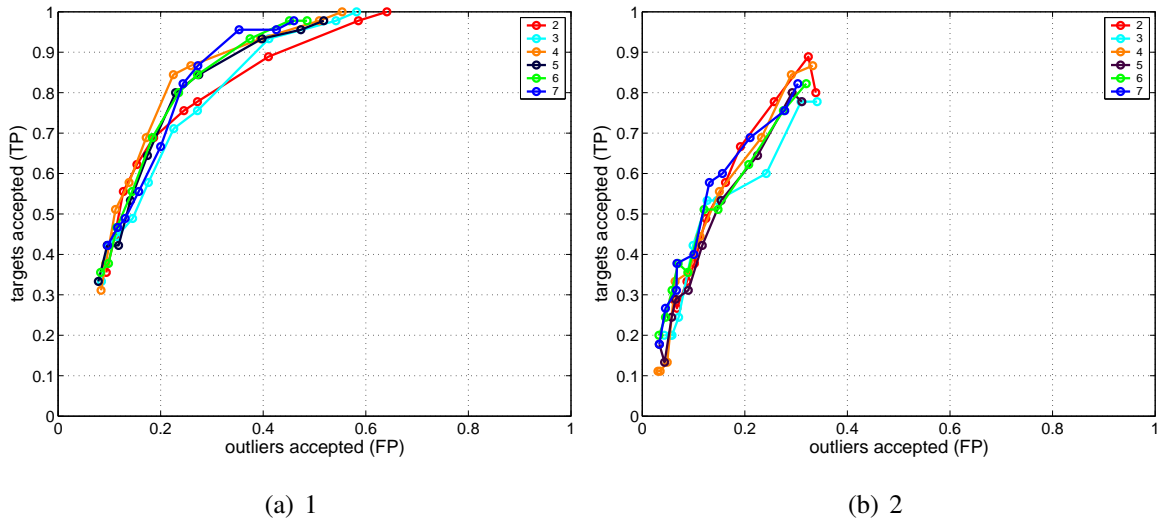


Figure 5.16: The ROC curves of using different numbers of feature of Analysis1 for G02. The left and right represent using the different classifiers, 4NN and SOM.

in Table 5.4. In the table, the F_{LR} represents the error of assigning a R-task data to L-task class using the trained L-classifier; as regards, the F_{RL} , the error of assigning a L-task data to R-task class using the trained R-classifier, where the L-classifier and R-classifier are trained by the one-class classification technique with treating the L-task and R-task as target data respectively. Figure 5.25 shows the perspective of F_{LR} and F_{RL} . The mean of the classification error is lower than 30%.

Results of Analysis4

Finally, we concern the Analysis4: classification between the three classes based on classification method B. The analysis results are shown in Table 5.2. In Table 5.5, we show the mean of these results, where we can find that the accuracy of classifying three class, L, R and resting, is poor. The most issue we have to point out is that almost all the error arises from assigning the active tasks to resting states. This is caused by our consideration of Criterion1 in section 5.3 and Criterion3 in section 5.4.5. We think that it may improve the accuracy by applying a two-class classifier to handle the situation when the output is (1,1) in Criterion1. Therefore, we finally perform another analysis, Analysis5, to discuss this idea.

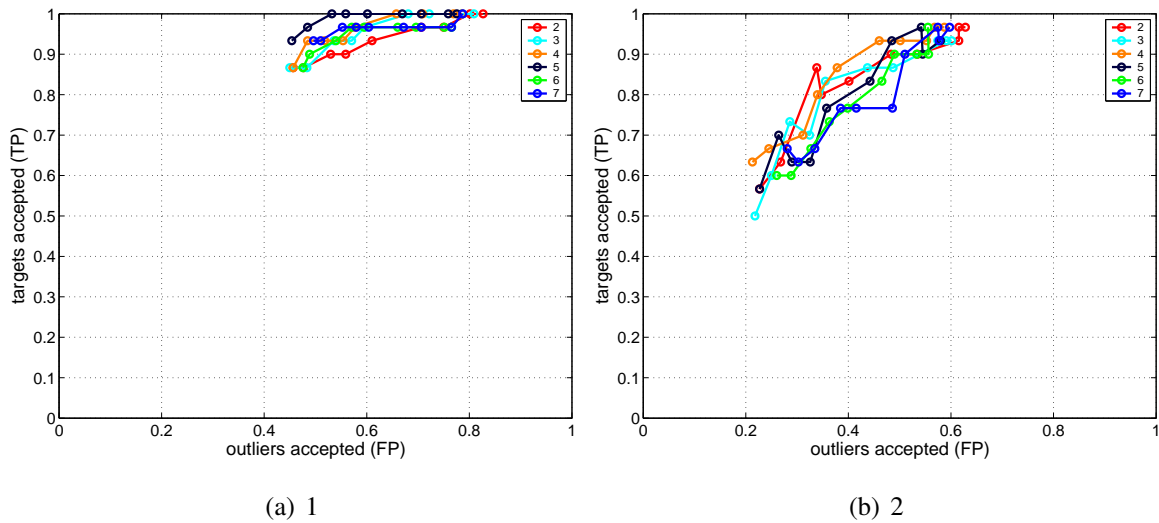


Figure 5.17: The ROC curves of using different numbers of feature of Analysis1 for G03. The left and right represent using the different classifiers, 4NN and SOM.

The analysis results of Analysis5 can be seen in Table /reftb:TbAnaly5. However, this table illustrates the improvement is just a little, implying adding a two-class LR-classifier can not settle the above problem well. We think there are two reasons.

First, we can use only the information of 0.5s-sliding-window to train the LR-classifier. Second, the 0.5s information may not really contains the active information in this period. The two problems we encounter in Analysis5 lead to a bad performance classification between left and right; thus it still can not settle the problem in Analysis4. If we can determine the real onset time of the active task to extract the real active information to train the classifier, we will have more possibility to obtain a good LR-classifier to handle the problem here. This may be the future work for the thesis.

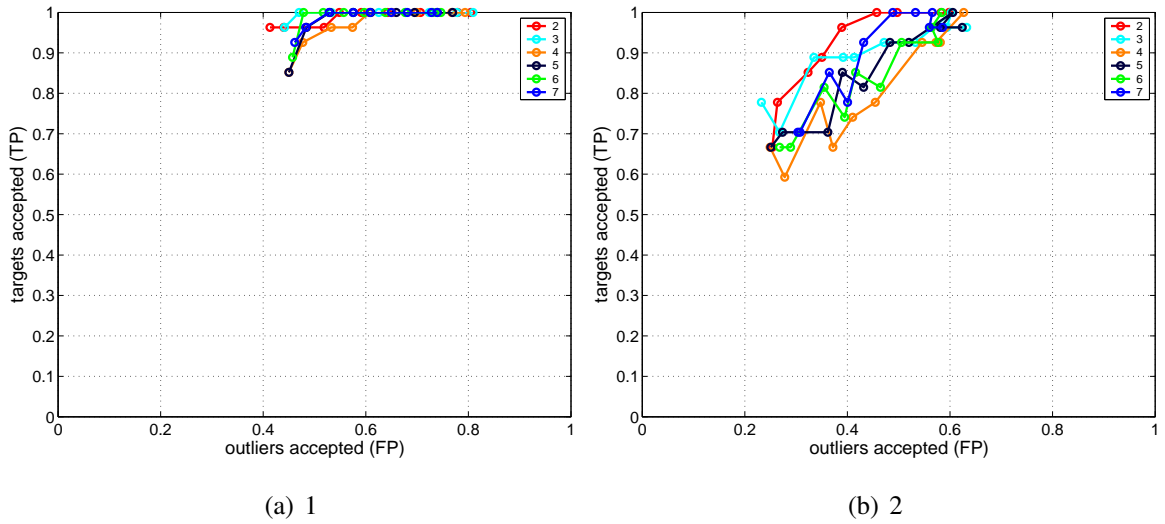


Figure 5.18: The ROC curves of using different numbers of feature of Analysis1 for G04. The left and right represent using the different classifiers, 4NN and SOM.

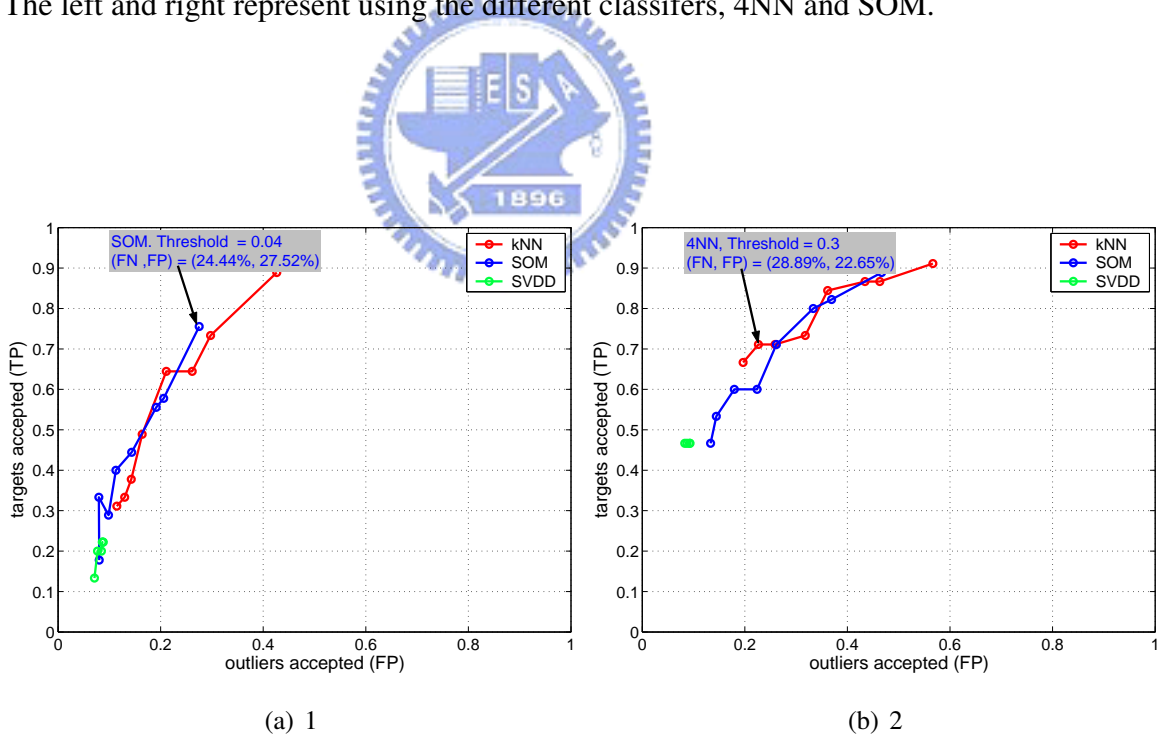


Figure 5.19: The ROC curves of the three classifiers when analyzing G02-L and G02-R of analysis 2.

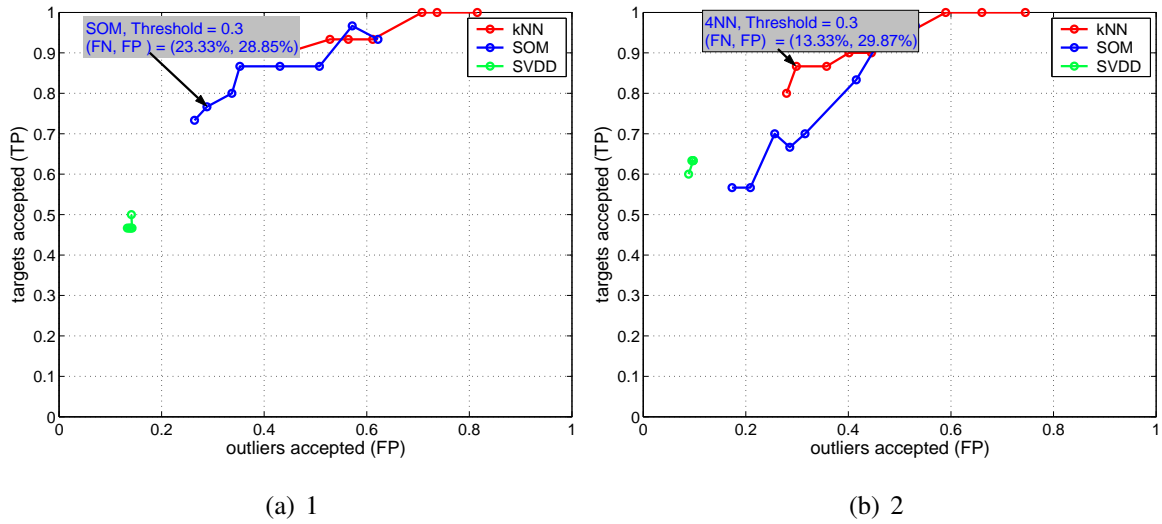


Figure 5.20: The ROC curves of the three classifiers when analyzing G03-L and G03-R of analysis 2.

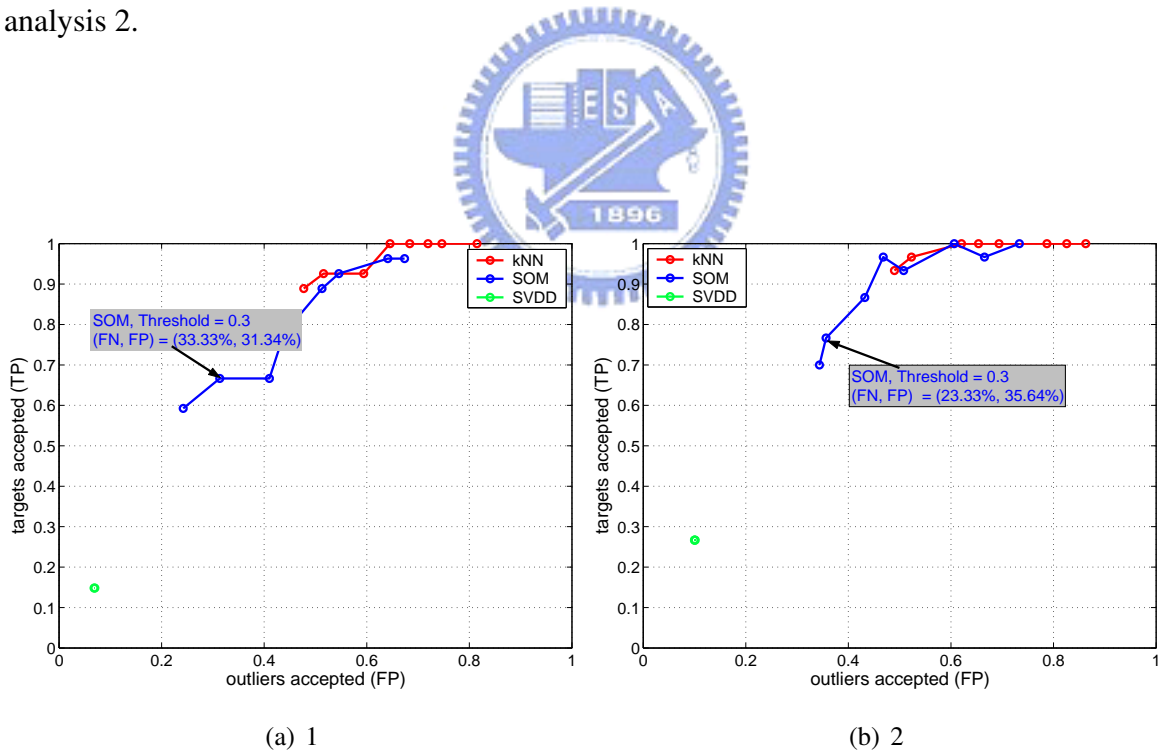


Figure 5.21: The ROC curves of the three classifiers when analyzing G04-L and G04-R of analysis 2.

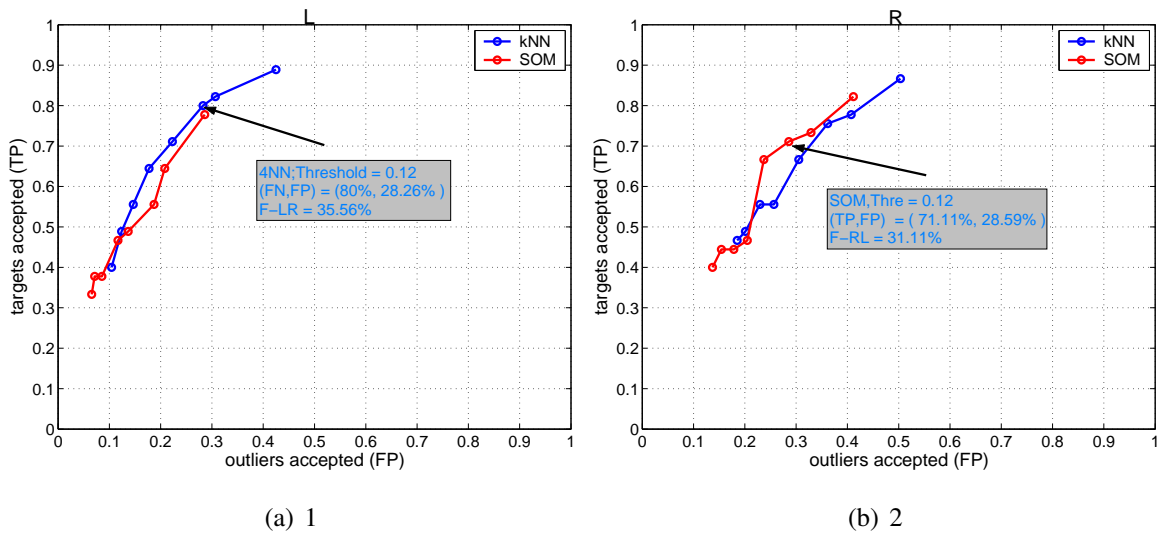


Figure 5.22: The ROC curves of the three classifiers when analyzing G02-L and G02-R of Analysis3.

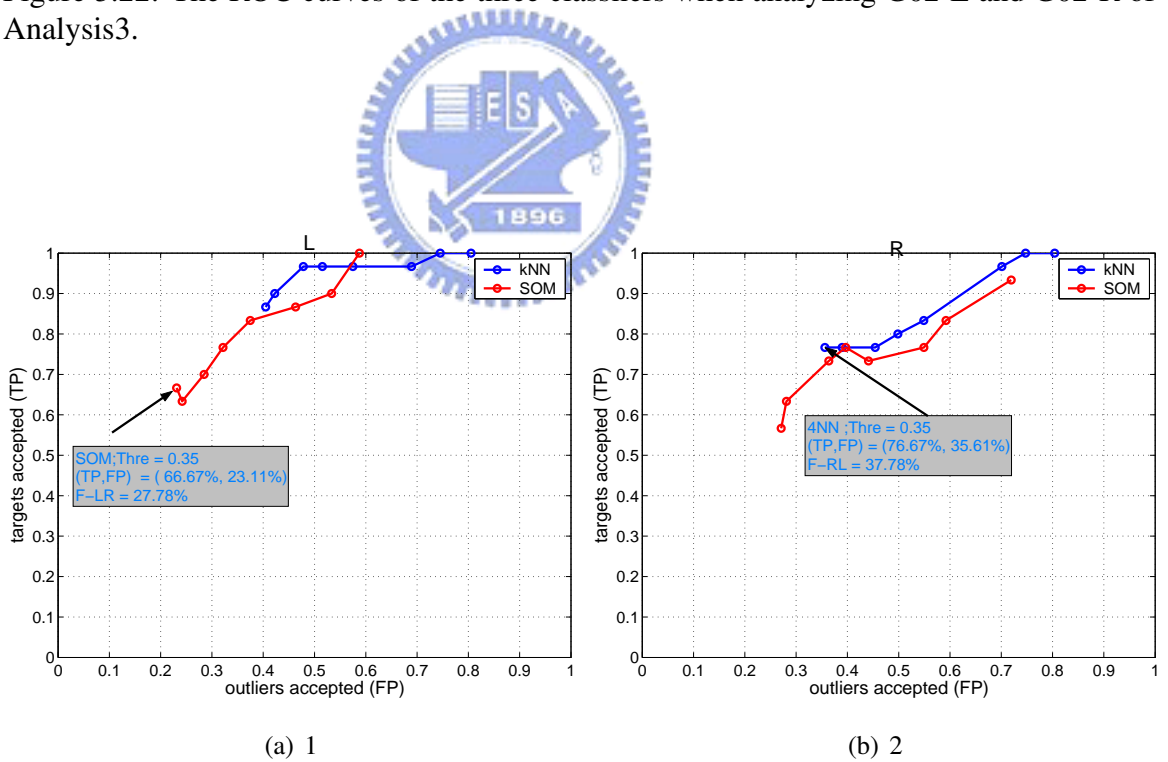


Figure 5.23: The ROC curves of the three classifiers when analyzing G03-L and G03-R of Analysis3.

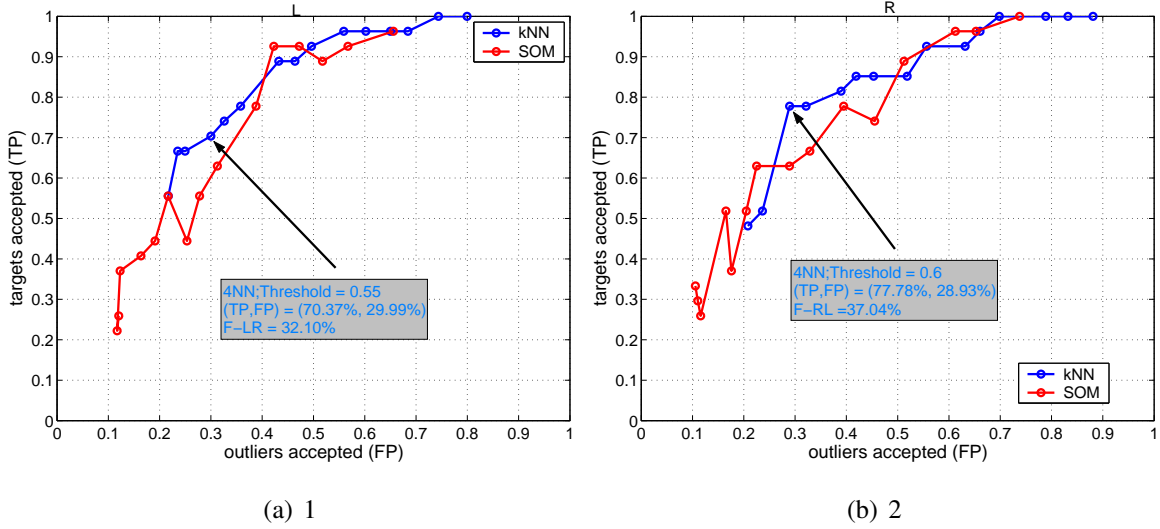


Figure 5.24: The ROC curves of the three classifiers when analyzing G04-L and G04-R of Analysis3.



Table 5.2: Analysis4. In this figure, LN_L represents number of results assigned to L, N_R the number of results assigned to R, LN_r the number of results assigned to resting

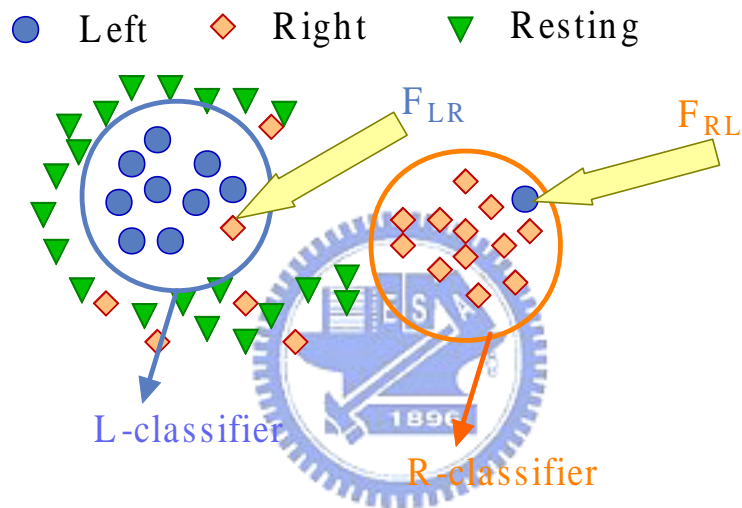
Dataset	Classifier	L or R	FN	FP	LN_r	LN_L	LN_R
G02	L:4NN,thre=0.12;	L	48.89	28.12	13	23	9
	R:SOM,thre=0.12;	R	60.00	33.59	13	14	18
G03	L:SOM,thre=0.35;	L	70.00	36.28	11	9	10
	R:4NN,thre=0.35;	R	46.67	35.00	9	5	16
G04	L:SOM,thre=0.30;	L	70.37	36.18	12	8	7
	R:SOM,thre=0.20;	R	70.37	35.19	12	7	8

Table 5.3: Results of Analysis2. The upper is the results of our method and the bottom the method proposed by Graz group in 2004.

	Dataset	Classifier		FN(%)	FP(%)
Our data	G02	L:SOM,thre=0.04;	L	24.44	27.52
		R:SOM,thre=0.16;	R	28.89	22.65
	G03	L:SOM,thre=0.35;	L	23.33	28.85
		R:SOM,thre=0.16;	R	13.33	29.87
	G04	L:SOM,thre=0.30;	L	33.33	31.34
		R:SOM,thre=0.20;	R	23.33	35.64
	Mean		L	27.03	29.24
			R	21.85	29.39
Graz 2004	g3	RP=non ;	L	24.00	16.00
			R	19.00	10.00
		RP=1.00;	L	14.00	16.00
			R	14.00	11.00
		RP=1.25;	L	14.00	15.00
			R	9.00	16.00
		RP=1.35;	L	11.00	28.00
			R	20.00	14.00
	g7	RP=non ;	L	21.00	37.00
			R	42.00	16.00
		RP=1.00;	L	16.00	19.00
			R	42.00	14.00
		RP=1.25;	L	14.00	24.00
			R	35.00	27.00
		RP=1.35;	L	25.00	21.00
			R	55.00	11.00
	i2	RP=non ;	L	6.00	75.00
			R	7.00	77.00
		RP=1.00;	L	9.00	59.00
			R	13.00	59.00
		RP=1.25;	L	12.00	59.00
			R	19.00	52.00
		RP=1.35;	L	23.00	49.00
			R	27.00	50.00
	Mean		L	0	0
			R	0	0

Table 5.4: Analy3 Result

	FN%	FP%	F_{LR} %	F_{RL} %
L	27.65	27.12	31.81	
R	24.81	31.04		35.31

Figure 5.25: A plot of illustrating F_{LR} and F_{RL} .Table 5.5: Mean of Analysis4. In this table, LN_L represents number of results assigned to L, N_R the number of results assigned to R, LN_r the number of results assigned to resting

Mean	L or R	FN	FP	LN_r	LN_L	LN_R
	L	59.26	38.48	12.3	14	7
	R	63.21	36.12	15	6	12

Table 5.6: Analysis5. In this figure, LN_L represents *number of results assigned to L*, N_R the *number of results assigned to R*, LN_r the *number of results assigned to resting*

Dataset	Classifier	L or R	FN	FP	LN_r	LN_L	LN_R
G02	L:4NN,thre=0.12;	L	44.44	46.27	14	25	6
	R:SOM,thre=0.12;	R	80.00	51.75	16	20	9
G03	L:SOM,thre=0.35;	L	56.67	56.43	12	13	5
	R:4NN,thre=0.35;	R	50.00	54.76	8	7	15
G04	L:SOM,thre=0.30;	L	66.67	48.15	11	9	7
	R:SOM,thre=0.20;	R	70.37	46.03	14	5	8



Chapter 6

Conclusions



In this chapter, we review the most important issues when analyzing EEGs in an asynchronous BCI system and then illustrate the contribution and advantages of our method.

The analysis of EEGs in an asynchronous BCI system in this thesis were:

Signal preprocessing

The aim of this step is to enhance the SNR of the EEGs. We applied artifact removal and bandpass filter to achieve the aim. When the number of the channels is large, some spatial filter may be applied to make EEGs reference-free. In our work, when analyzing asynchronous BCI, we do not apply spatial filter because we only use EEGs of three channels, C3, Cz, C4 to detect the user's intention.

Feature extraction

Because the signals of performing the predefined mental task are usually smaller than the ongoing EEGs, the signals we interested are concealed in the irregular ongoing signals and noise. To overcome the problem, we will perform some approaches to extract the signals we interested. In this work, we compare four feature extraction approaches including ERD/ERS, CSP, Haar wavelet and Morlet wavelet. We finally decide to use Morlet wavelet to analyze the EEGs in an asynchronous BCI system due to the analysis result in Chapter 4.

- Feature selection

Due to the dimension of features is usually high, we perform t-statistic and FFS to select more discriminative features to represent the original data. The two approaches has a satisfied and efficient results in the analysis of our work. We finally use only 2 features to present a data and have a reasonable classification accuracy in the analysis of Chapter 5. The property of few features will have a significant influence when analyzing a online BCI because the computation time is decreased a lot.

- Classification

We involve the technique of one-class classification to settle the problem of recognize between the active mental tasks and resting states. Because the resting state

is composed of a lot of different ongoing states, thus is a wide-spreading distribution. Besides, the resting states of training phase and testing phase is very different because the user having to pay attention to the feedback is very active at the testing phase. We think that if we can develop a good model of active states and output 0(zero) when the testing data is of resting states, then the problem in an asynchronous BCI may be solved as well. Thus, using the technique of one-class classification to handle the problem becomes very attractive for it can avoid to include a resting class and only focus on how to fit the data of active states. We also have discussed the two approaches of using two one-class classifiers and of using two one-class classifiers and a two-class classifier to recognize the left, right and resting states in Chapter 5.

In this thesis, the achievements can be summarized as follows.

1. We use few number of channels,3, to achieve the near same or better accuracy than Graz,2004 [51] of using 27 channels.
2. We develop a efficient feature selection process to select significantly discriminative features. For example, in our work, when analyzing EEGs of an asynchronous BCI, we use only 2 features to present the EEGs and delivered only the 2 information to the classifier yet still obtain a reasonable accuracy.
3. According to our experiments using 2005 BCI competition III datasets, the proposed methods can achieve higher stability and accuracy than Graz's method [51] proposed in 2004.



Bibliography

- [1] Eugene Tolunsky A. James Rowan. *Primer of EEG With a Mini-Atlas: With a Mini-Atlas*. Butterworth-Heinemann, 2003.
- [2] Kotchoubey B., Haisst S., Daum I., Schugens M., and Birbaumer N. Learning and self-regulation of slow cortical potentials in older adults. *Experimental Aging Research*, 26:15–35, 2000.
- [3] Jessica Bayliss and Dana Ballard. A virtual reality testbed for brain-computer interface research. *IEEE Transactions on Rehabilitation Engineering*, 8(2):118–190, 2000.
- [4] N. Birbaumer, N. Ghanayim, T. Hinterberger, I. Iversen, and B. Kotchoubey. A spelling devices for the paralyzed. *Nature*, 398:297–298, 1999.
- [5] G.L. Calhoun and G.R. McMillan. EEG-based control for human-computer interaction. *Hics*, 00(00):0–4, 1996.
- [6] J.K. Chapin and G. Gaal. Robotic control from realtime transformation of multi-neuronal population vectors. *Brain-Computer Interface Technology: Theory and Practice: First International Meeting Program and Papers*, 1999.
- [7] J.K. Chapin, K.A. Moxon, R.S. Markowitz, and M.A.L. Nicoleslis. Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nature Neurosci.*, 2(7):664–670, 1999.
- [8] J. d. R. Millan, J. Mourino, M. Franze, F. Cincotti, M. Varsta, J. Heikkonen, and

- F.Babiloni. A local neural classifier for the recognition of EEG patterns associated to mental tasks. *IEEE Transactions on Neural Networks*, 13:678–686, May 2002.
- [9] John Polich Daran Ravden. Habitation of p300 from visual stimuli. *International Journal of Psychophysiology*, 30:359–365, 1998.
- [10] R. Grave de Peralta, S. L. Gonzalez, J. del R. Milln, T. Pun, and C. M. Michel. Direct non-invasive brain computer interfaces. Proceedings of the 9th International Conference on Functional Mapping of the Human Brain, June 2003.
- [11] Emanuel Donchin et al. The mental prosthesis: Assessing the speed of a p300-based brain-computer interface. *IEEE Transactions on Rehabilitation Engineering*, 8(2):174–179, 2000.
- [12] Matti Hamalainen et al. Magnetoencephalography - theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2):413–497, 4 1993.
- [13] Niels Birbaumer et al. The thought translation device (tt) for completely paralyzed patients. *IEEE Transactions on Rehabilitation Engineering*, 8(2):190–193, 2000.
- [14] Peter Jezzard et al. *Functional MRI: An Introduction to the Methods*. Oxford University Press, 2001.
- [15] L.A. Farwell and E. Donchin. Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials. *Electroencephalogr. Clinical Neurophysiology*, 70:510–523, 1988.
- [16] Keinosule Fukunaka. *Introduction to statistical pattern recognition*. ACADEMIC PRESS, INC, 1990.
- [17] G.Pfurschteller and C.Neuper. Motor imagery and direct brain-computer communication. *Proceedings of the IEEE*, 89:1123–1134, July 2001.

- [18] T. Hinterberger, J. M. Houtkooper, and B. Kotchoubey. Effects of feedback control on slow cortical potentials and random events. *The Parapsychological Association Convention*, 00:0, 2004.
- [19] <http://faculty.washington.edu/chudler/1020.html>. 10-20 system.
- [20] http://www.cortechsolutions.com/g.BSanalyze_EEGtoolbox.htm. The manual of g.BSanalyze.
- [21] <http://www.neuro.uu.se/fysiologi/gu/nbb/lectures/EEGBas.html>. Eeg basis.
- [22] http://www.socialresearchmethods.net/kb/stat_t.htm. t-test.
- [23] Bayliss JD and Ballard DH. A virtual reality testbed for brainvcomputer interface research. *IEEE Transactions on Rehabilitation Engineering*, 8:188–190, 2000.
- [24] Kanehisa Morimoto Jingbo Pan, Tatsuya Takeshita. P300 habituation from auditory single-stimulus and oddball paradigms. *International Journal of Psychophysiology*, 37:149–153, 2000.
- [25] editor John G. Webster. *Medical Instrumentation, chapter4*. John Wiely & Sons Inc., 1998.
- [26] M. Joho and K. Rahbar. Joint diagonalization of correlation matrices by using newton methods with applicaiton to blind signal separation. *IEEE Sensor Array and Multi-channel Signal Processing Workshop SMA*, 2000.
- [27] Andrea1 et al. Kbler. Brainvcomputer communication: unlocking the locked in. *Psychological Bulletin*, 127(3):358–373, 2001.
- [28] S. G. Mason and G. E. Birth. A brain-controlled switch for asynchronous control applications. *IEEE Transations on Biomedical Engineering*, 47(10):1297–1307, OCTOBER 2000.
- [29] Dennis J. McFarland, William A. Sarnacki, and Jonathan R. Wolpaw. Brain-computer interface (BCI) operation: optimizing information transfer rates. *Biological Psychology*, 63:237–251, 2003.

- [30] Matthew Middendorf, Grant McMillan, Gloria Calhoun, and Keith S. Jones. Brain-computer interface based on the steady-state visual-evoked response. *IEEE Transactions on Rehabilitation Engineering*, 8(2):211–214, 2000.
- [31] Matthew Middendorf, Grant McMillan, Gloria Calhoun, and Keith S. Jones. Brain-computer interfaces based on the steady-state visual-evoked response. *IEEE Transactions on Rehabilitation Engineering*, 8(2):211–214, 2000.
- [32] Johannes Miller-Gerking, G.Pfurtscheller, and Henrik Flyvbjerg. Designing optimal spatial filters for single-trial EEG classification in a movement task. *Clinical Neurophysiology*, 110:787–798, 1999.
- [33] Birbaumer N., Elbert T., Canavan A. G. M., and Rockstro h B. Slow potentials of the cerebral cortex and behaviour. *Physiological Reviews*, 70(1):1–41, 1990.
- [34] Ernst NiederMeyer and editors Fernando Lopes da Silva. *Electroencephalography*. Lippincott Williams & Wilkins, 1999.
- [35] L. Otten and E. Donchin. The relationship between p300 amplitude and subsequent recall for distinctive events: Dependence on type of distinctiveness attribute. *Psychophysiology*, 37:644–661, 2000.
- [36] G. Pfurtschellera and F.H. Lopes da Silva. Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology*, 110:1842–1857, 1999.
- [37] G. Pfurtschellera and W. Klimesch. Functional topography during a visuoverbal judgement task studied with event-related desynchronization mapping. *Journal of Clinical Neurophysiology*, 9:120–131, 1992.
- [38] John Polich. Comparison of auditory p300 habituation from active and passive conditions. *International Journal of Psychophysiology*, 17:25–34, 1994.
- [39] P.R.Kennedy and R.A.E. Bakay. Restoration of neural output from a paralyzed patient by a direct brain connection. *NeuroReport*, 9:1707–1711, 1998.

- [40] P.R.Kennedy and R.A.E. Bakay. Direct control of a computer from the human central nervous system. *Brain-Computer Interface Technology: Theory and Practice: First International Meeting Program and Papers*, June 1999.
- [41] Brechet R and Lecasble R. Reactivity of mu-rhythm to flicker. *Electroencephelin Neurophysiol*, 18:721–722, 1965.
- [42] Jonathan R. and Wolpaw et al. Brain-computer interface research at the wadsworth center. *IEEE Transactions on Rehabilitation Engineering*, 8(2):222–226, 2000.
- [43] Jonathan R. and Wolpaw et al. Brain-computer interface technology: A review of the first international meeting. *IEEE Transations on Rehabilitation Engineering*, 8(2):164V173, 2000.
- [44] D. Regan. *Human Brain Electrophysiology*. New York: Elsevier, 1989.
- [45] A. James Rowan and Eugene Tolunsky. *Primer of EEG*. Butterworth Heinemann, 2003.
- [46] Richard HC Seabrook. The brain-computer interface: technique for controlling machines. 00.
- [47] Claire Murphy Spencer Wetter, John Polich. Olfactory, auditory, and visual erps from single trials: no evidence for habitation. *International Journal of Psychophysiology*, 54:263–272, 2004.
- [48] E.E. Sutter. The visual evoked response as a communication channel. Proc. IEEE/NSF Symp, June 1984.
- [49] E.E. Sutter. The brain response interface: Communication through visually-induced electrical brain responses. *Microcomput. Appl.*, 15:31–45, 1992.
- [50] David Martinus Johannes TAX. *One-class classificaiton- concept learning in the absence of counter-examples*. PhD thesis, 2001.

- [51] George Townsend, Bernhard Graimann, and Gert Pfurtscheller. Continuous EEG classification during motor imagery -simulation of an asynchronous BCI. *IEEE Transaction on Neural System and Rehabilitation Engineering*, 12(2):258–265, June 2004.
- [52] George Townsend, Bernhard Graimann, and G. Pfurtschellera. Continuous EEG classification during motor imageryxsimulation of an asynchronous BCI. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 12:258–265, 2004.
- [53] F. van der Heijden, R.P.W. Duin, D. de Ridder, and D.M.J Tax. *Classification, parameter estimation and state estimation*. John Wiley & Sons, Ltd., 2004.
- [54] Juha Vesanto, Johan Himberg, Esa Alhoniemi, and Juha Parhankangas. Som toolbox for matlab5. 2000.
- [55] J. R. Wolpaw and D. J. McFarland. Multichannel EEG-based brain-computer communication. *Electroencephalogr. Clinical Neurophysiology*, 90:444–449, 1994.
- [56] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and C. A. Forneris. An EEG-based brain-computer interface for corsur control. *Electroencephalogr. Clinical Neurophysiology*, 78:252–258, 1991.
- [57] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T.M Vaughan. Brainvcomputer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767–791, 2002.