國立交通大學

生醫工程研究所

碩 士 論 文

運用腦電波之身分辨識系統
及其長時調變機制

EEG-based Person Identification System and Its Longitudinal Adaptation

研 究 生：鄭佳怡

指導教授：陳永昇　博士

中 華 民 國　一○二 年 五 月

運用腦電波之身分辨識系統及其長時調變機制

EEG-based Person Identification System and Its Longitudinal Adaptation

研 究 生：鄭佳怡　　Student：Chia-Yi Cheng

指導教授：陳永昇　　Advisor：Yong-Sheng Chen

國 立 交 通 大 學

生 醫 工 程 研 究 所

碩 士 論 文

A Thesis

Submitted to Institute of Biomedical Engineering

College of Computer Science

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Computer Science

May 2013

Hsinchu, Taiwan, Republic of China

中華民國一〇二年五月

EEG-based Person Identification System and Its Longitudinal Adaptation

A thesis presented

by

Chia-Yi Cheng

to

Institute of Biomedical Engineering

College of Computer Science

in partial fulfillment of the requirements

for the degree of

Master

in the subject of

Computer Science

National Chiao Tung University

Hsinchu, Taiwan

2013

EEG-based Person Identification System and Its Longitudinal Adaptation

Copyright © 2013
by
Chia-Yi Cheng

摘　　要

生物特徵辨識因其廣泛性、可攜性與不易被複製等特質，而逐漸成為身分辨識的重要方法。運用腦電波進行生物特徵辨識是近年興起的研究主題，過去相關研究指出腦電波帶有個體獨特性的資訊且難以被竊取，故在高安全需求的系統上有相當高的應用潛力。然而腦電波辨識在準確度與穩定性上仍有改善空間。

本研究針對腦電波身分辨識系統提出了兩階段式身分辨識與長時調變機制，來分別改善準確率與穩定度。前者可在維持正確接受率的情況下，進一步降低錯誤接受率；後者則可因應腦電波特徵隨時間之變化來調變系統。兩階段式系統中包含了分類部分以及驗證部分。腦電波訊號經過自身回歸模型（autoregressive model, AR）與頻帶特徵的運算後，再透過多分類組別之支持向量機（support vector machine, SVM）來進行分類。而驗證部分是由候補者篩選(candidate selection)、資料重新表述(re-representation)和身分驗證三步驟所構成。我們首先運用線性區分分析（linear discriminant analysis, LDA）進行資料的重新表述，再由兩分類群組之支持向量機與最近相鄰分類器驗證分類結果。

另一方面，長時調變機制則分為漸進式學習（incremental learning）與在分類前先行調整資料分佈的適應性系統(adaptive system)。在適應性系統裡，我們探討了主成分分析與領域調整（domain adapatation）兩種方法。該系統是根據新資料來調整訓練資料的分佈，再利用調整過的內容重建分類器並進行身分辨識。

本研究共招募二十三位受試者，並在其自發性手指抬動的情況下收取 90 筆無眼動干擾之腦電波訊號。根據系統評估的實驗結果，兩階段式身分辨識能在兼顧正確接受率的情況下，有效地將錯誤接受率由 9.5%降低至 5.4%（正確接受率的改變幅度為 90.5%略降為 87.7%）。在穩定性的研究上，頻譜特徵是所有我們使用的特徵值中最不易隨時間改變的。若採用該特徵，主成分調整式身分辨識系統可增加 5.2%的辨識正確率；漸進式學習系統更可將正確辨識率自 12.2%提升至 58.9%。

Abstract

Biometrics has been viewed as an alternative to conventional person identification methods due to its universality, portability, and resistance in duplication. Previous study indicated that electroencephalography (EEG) carries discriminative information for distinguishing individuals and has great potential to meet the requirements of high security level. Therefore, it is essential to improve the reliability and stability of the EEG-based biometrics system for the promotion of its applicability.

In this study, we first developed a two-stage EEG-based person identification system to improve the reliability by lowering the false acceptance rate while maintaining the true acceptance rate. The second purpose was to improve the stability of the EEG-based identification system using longitudinal adaptation. In addition to the classification stage, the proposed two-stage identification system re-examines the classification results in another verification stage, which consisted of candidate selection, re-representation, and verification steps. The classifier applied in the classification step was a multi-class support vector machine (SVM), while a binary SVM or kNN classifier was utilized in the verification step. Additionally, the linear discriminant analysis (LDA) was utilized to re-represent the training data before the verification step.

On the other hand, the longitudinal adaptation was accomplished by either incremental learning or an adaptive system. In the adaptive system, we applied the principal component analysis (PCA) or domain adaptation (DA) to transform the data distribution toward the newly acquired data, and then we construct a new classifier with the adapted data.

For each of the 23 participants, 90 EOG-free trials of the EEG recordings were acquired during the lifting of left index finger. The experimental results showed that our two-stage identification process reduced the false acceptance rate from 9.5% to 5.4% while maintaining the true acceptance rate (form 90.5% to 87.7%). The evaluation results showed that the power spectral density (PSD) was the most stable feature, and the PCA-based adaptive system can improve the true acceptance rate by 5.2% when using the PSD features. Furthermore, the identification system using incremental learning can even improve the true acceptance rate from 12.2% to 58.9%.

誌　　謝

學生時期的終末，非常感激指導教授陳永昇老師與陳麗芬老師的悉心指導，使我無論在研究初期的文獻汲取、研究期間的難題討論、乃至最後階段的論文撰寫，各方面上都獲益良多。此外，也同樣感謝擔任口試委員的陳朝欽老師與王才沛老師在論文研究上的意見指導。有老師們的提攜，才使得這份研究成果臻於完備。同時，更對不厭其煩來參與實驗每一位受測者致上十二萬分的感激！

而在兩年的研究生活中，謝謝 BSP 實驗室的成員們陪我度過趕作業的抓狂時間和出遊解壓的瘋狂時間。謝謝晴懌、志維和伊婷在被論文進度壓得喘不過氣時陪我聊天放鬆；謝謝家仁學長和泰成學長幫忙實驗打膠的繁瑣過程；謝謝小雞學姐和慧伶學姐在我遇到瓶頸與撰寫論文時給予的諮詢建議；謝謝怡地和奭彥學長在口試當天幫忙打點飲食及問題記錄，並不斷給予鼓勵；特別感謝小白學長從專題生時期開始的帶領，舉凡修課作業的教導、研究過程中的細節討論、更甚是投影片的修改，都相當感激！

最後，謝謝六年來的優良室友和系桌大家庭。謝謝你們體諒我在口試前的語無倫次與焦亂情緒，更謝謝你們一直以來的陪伴與包容！有你們才讓我的研究生活鮮活而彌足珍貴。

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter, we will first introduce the biometrics and its categories. Secondly, we describe a measurement of brainwaves– electroencephalography– we used in our person identification system. In the next section, we describe EEG-based biometrics and then list some related works in this field. Finally, the motivation and purpose of this study are described in the last section.

1.1 Biometrics

As an alternative to conventional ID cards, keys, and passwords, biometrics has been gaining importance in the ubiquitous access control systems because of its advantages of high commonality, high uniqueness, easy acquisition, persistence, portability, and resistance in duplicating. Biometrics refers to the recognition of individuals by their biometric characteristics which are hard to be stolen or lost. Amongst varying biometric characteristics, Jain et al. [9] listed several requirements of an ideal biometric trait: universality, distinctiveness, permanence, collectability, performance, acceptability, and circumvention. The last three requirements are set for building a practical biometric system. The "universality" promises that everyone can use the characteristics to identify themselves. To be used as a recognition pattern, the distinctiveness– which means that every individual can be sufficiently distinct with the characteristic– is a basic and important requirement. The permanence and collectability confirm that the characteristic is stable and available to be used in the recognition system.

Biometric characteristics can be categorized into two groups: the physical and behavioral traits. Examples of physical traits are fingerprints, irises, faces and DNA, and these traits are usually anatomical parts of a human body. Most of the physical traits are identified on the outer part of human bodies and are easily obtained and used in the biometrics. On the other hand, behavioral biometrics recognizes individuals by their behavior, such as typing rhythm, signature, and gait. These kinds of traits have no exact shapes but can be acquired by a machine in recordings. With these recordings, the behavioral traits can then be widely used in the biometrics.

However, physical traits suffer from the risk of violent snatch whereas explicit behaviors can be observed and imitated. In addition, the discriminatory information of the fingerprint (probably the most widely-used biometric trait) has not yet been fully testified [18]. The admissibility of the fingerprint has been challenged due to possible false associations. Recently, electroencephalography (EEG) has been proposed as a new biometric trait [8, 13, 14, 17, 21]. Although the persistence of individual characteristics in EEG is yet to be investigated, its inherent advantage of implicit features results in the extreme difficulty in reproduction.

1.2 Electroencephalography

Magnetoencephalography (MEG), functional magnetic resonance imaging (fMRI), and EEG are non-invasive measurements that can be used to record brain activity. These measurements have been extensively used for brain research. The measurements of fMRI are the BOLD (blood-oxygen-level-dependent) effects relatively slow with high spatial resolution. On the contrary, MEG and EEG have high temporal resolution, but low spatial resolution due to the limited amount of sensors. Compared to fMRI and MEG, EEG has advantages of portability and lower cost.

EEG measures the change of the electric potential on the scalp induced by a group of active neurons on the cerebral cortex. Event-related potentials (ERPs) are brain responses measured by EEG when responding to specific stimuli or autonomous actions. This time-locked and phase-locked characteristic of brain activity allows trials of EEG recordings to be aligned and averaged to increase the signal-noise rate (SNR) by cancelling out the random noises.

The standard of EEG montage, called the international 10-20 system as shown in Figure 1.1, is established for EEG studies. In this system, the distances of adjacent electrodes are either 10% or 20% of the geodesic length from nasion to inion. Based on this standard, typical numbers of electrodes include 19 to 32, 64, 128, and 256.

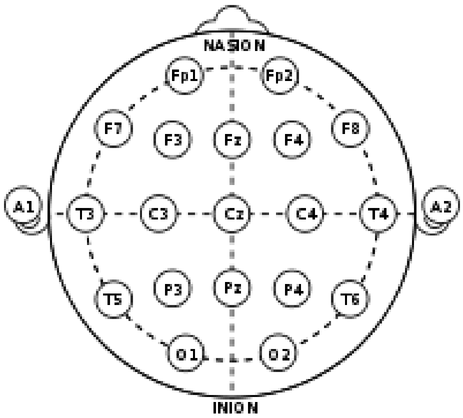

Figure 1.1: The international 10-20 system.

1.3 EEG-based biometrics

Although EEG is a new trait in the person recognition research field, EEG classification made its debut in 1970s as a key role in the construction of a natural interface called the brain-computer interface (BCI) [24, 25]. The original intention of the BCI is to provide an option to those who are suffering motor disability. A BCI system establishes a direct communication between machine and human bypassing the motor pathway [26]. In BCI systems, discriminative features are extracted from EEG signals and translated to the corresponding control commands for machines.

Since a BCI system will be operated across users, the extracted features from EEG signals should be discriminative in different tasks while keeping the consistency among populations to accommodate the BCI systems to the inter-subject variation. On the contrary, EEG-based identification systems aim to differentiate amongst people while performing the same requested task. In this case, discrepancy in the extracted features facilitates the recognition of person identities.

There are three major steps in EEG-based biometrics including EEG data acquisition, feature extraction, and classification. The person recognition system can be categorized into person authentication and identification systems [9, 12]. The person authentication, or verification, system is to validate the identity claimed by the clients. That is, the clients of the authentication part of the system will answer either “Accept. The identity you claimed belongs to you.” or “Reject. You are not the person you claimed.” The authentication system constructs a customized personal classifier for each client in the database. According to the claims, the authentication system can verify the claimed identity using the corresponded classifier by one-to-one matching as shown in Figure 1.2 [10].

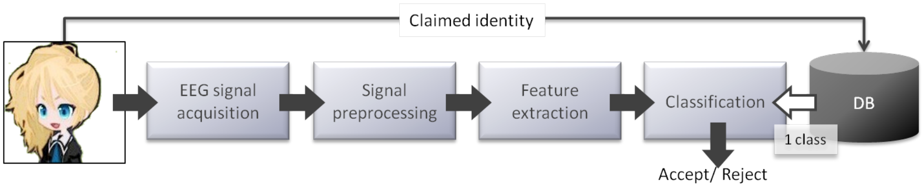

Figure 1.2: Architecture of the person authentication system.

In contrast, the person identification system reports the identity of a client from a database trained beforehand. For this purpose, an identification system finds the best candidate using the 1-N matching [21]. Figure 1.3 shows the architecture of an identification system. In EEG-based identification system, a client only provides his/her EEG recordings and the system will determine who he/she is.

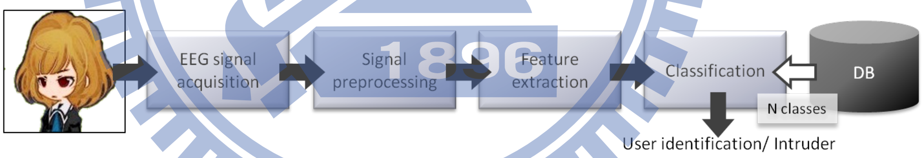

Figure 1.3: Architecture of the person identification system.

To increase the reliability of person recognition system, it is essential to detect the imposters and intruders in different stations. The imposter refers to a client belonging to the database but impersonating someone else belonging to the database, while the intruder is a client even not belonging to the database but claiming the identity whose owner belongs to the database. In the authentication system, each customized classifier should be able to reject the invalid claims made by the imposter or intruder. Inherently, the identification system does not face problem of imposters and use an additional class for intruders to prevent them from getting an arbitrary identity.

1.4 Related works

1.4.1 Task design

In order to distinguish individuals, the EEG-based biometrics should correct recognize people by their EEG signal while they are doing the same thing. The one of the simplest tasks which subjects are asked to do is resting [28]. It is also an earliest task that had been used in EEG-based biometrics. Poulos et al. [21] had proposed a EEG-based biometrics using eye-closed resting state signal. The correct verification rate was 72%-84%. This results confirmed that the EEG carried the genetically-specify information. Another famous task–visual evoked potential (VEP)– had been applied in several study [11,16].When acquired EEG signal, the subjects were seated in front of a monitor which would display visual stimuli. These visual stimuli were usually light flash. In [12, 23], the EEG signals were acquired when the subjects do the motor imagery tasks. The subjects were asked to imagine moving their left or right hand. Both of these two researches had the same conclusion that imagination of left hand movement is more efficient to recognize individuals than imagination of right hand movement.

1.4.2 Feature extraction

There are several kinds of features utilized in EEG-based biometrics, such as temporal waveform [11],or frequency spectrum [4, 20]. In [28], they built an EEG-based identification with eye-closed resting data and utilized linear and non-linear type of features. The linear features were autoregressive model (AR) and frequency spectrum, while the non-linear features were the complexity and chaotic characteristic of brainwaves. The correct classification rate of the linear features was 97.29% which is much higher than that of non-linear features (44.14%). Therefore, they came a conclusion that the linear features are more suitable to build a EEG-based biometrics.

1.4.3 Classification

A personal classifier plays a significant role in a personal authentication to execute one-to-one matching. The personal classifier can be attained by setting a personal threshold [13] or constructing a binary classifier [22]. The simple ideal of a threshold approach is to make an acceptance or a rejection by the difference between the given data and data stored in the database. The previous researches had defined the differences as Manhattan distance [15], or likelihood probability [12]. In Figure 1.4, the person authentication system would compute a pair of thresholds $\{Th_1, Th_2\}$ for each class and operated as the procedure shown on Figure. The Dt is the Manhattan distance computed from the given data. The opinion of the classifier with thresholds was computed only by the data belongs to the class and not considering the data of other classes in the database. In contrast with Palaniappan who computed the opinion only with the corresponding data, [12] obtained the threshold for a class both considering the data from the class itself and others. The opinion ($\Lambda(X)$) here was defined as the difference between the log likelihood probability of being the client and the log of the probability of not the client, which had been shown in Equation 1.1.

$$\Lambda(X) = logP(X \mid \lambda_C) - logP(X \mid \neg\lambda_C) \tag{1.1}$$

When $\Lambda(X) \geq Threshold$, the authentication would accept that the given input belongs to the identity he or she claimed. Otherwise, the system would reject the given person as a imposter.

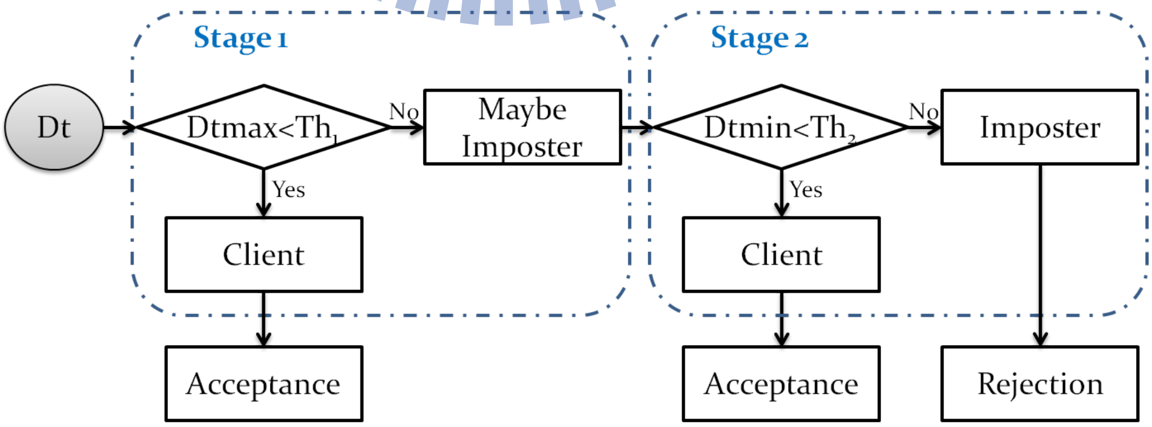

Figure 1.4: The procedure of the two-stage authentication proposed by Palaniappan.

On the other hands, a binary classifier used in [1] is the support vector machine (SVM). For each class, the authentication system would label the data in the database with only two types: 'subject' and 'not subject'. The more about the SVM classifier will be described in Section 2.3.3.

The classifier used in person identification is usually a multi-class classifier. The k-nearest neighbors (kNN) classifier which makes the decision by the distance has been used in pervious study [16,27]. In [7], the performance of the identification in a database with 20 subjects acquired in resting-state can achieve nearly 70% accuracy rate. Another classifier–LDA– had also been used in a person identification [19]. The performance with a database containing 40 subjects acquired in resting state is 85% when using an autoregressive model to capture the features.

1.5 Thesis scope

The resistance of being duplicating and violently snatched, the potential of EEG of being used as a new trait on biometrics is bright expectation. By appropriative task designing, feature extraction and classification, the discriminatory information carried by the brainwave has been excavated and used to distinguish individuals. However, to utilize EEG-based biometrics in systems with high security requirement, the limitations of EEG with respect to the reliability and stability become essential topics. The reliability of EEG-based biometrics contains not only the true acceptance rate but also the false acceptance rate. The true acceptance rate indicates that the clients can be correctly identified by the EEG-based identification. While the false acceptance rate refers to that a client might enter the system with a wrong identification decided by the EEG-based systems. Furthermore, an intruder would invade the system because an identification system accepts all given person and does not consider that the given person is totally not belonging to the database. On the other hand, the limitation on stability is that the performance of EEG-based biometrics was decreased caused by the time gap between the training and testing data [7, 12]. Neither authentication nor identification system can avoid the decreasing tendency of performance.

The first purpose of this study aims to design an identification system which focuses on lowering the false acceptance rate to defend against the imposter as well as intruder while maintaining the true acceptance rate. The lifting of the left index finger is the task of our system because that the left hand motor task has higher discriminant than the right hand (we introduced in Section 1.4.1). In addition, we attempt to find a more stable feature and a longitudinal adaptation for the EEG-based identification system to conserve the decreasing on performance over time span.

In the next chapter, we will first describe our proposed identification system composed of a two-stage identification and an adaptive system, and then introduce each part of these two systems in detail as well as the evaluated criteria. The evaluated results of systems will be recorded in chapter 3. The discussion and conclusions of our study can be given in the last two chapters.

Chapter 2

Materials and Methods

2.1 Materials

2.1.1 Subjects

Twenty-three subjects joined the experiment, and 14 of them had three EEG recordings acquired on three different days with time interval more than one month. Additionally, these fourteens include one subject who had 19 EEG recordings obtained in two years. The interval between the first day and second day when the subjects came on was 9 months, while that between the first day and third day was a year. The whole subjects consisted of 12 males and 11 females. They all were the students in Chiao-Tung University and had been diagnosis any brain disease or drug addicted. The mean age when the subjects first came is 22 (21-25 years old). Most of our subjects are right-handed, excluding one male and one female are left-handed.

Table 2.1: The demography data of the subjects.

	Came at least 1 time	Came at least 3 times	Came 19 times
Male	12	10	0
Female	11	4	1
Left-handed	2	1	0
Right-handed	21	13	1
Total	23	14	1

2.1.2 Paradigm

The self-paced left index finger lifting is the paradigm using in this study. The EEG acquisition experiments were progressed in a quiet and light room with no other signal interference excepting the experimental devices. subjects would set comfortably and put their left hand on a board which can detect the finger lifting as well as send a trigger to the computer. The subjects were instructed to lift their finger while counting from one to eight in silent and also do not blink during lifting. Each experiment included three sessions of

50 times of the counting and lifting behavior.

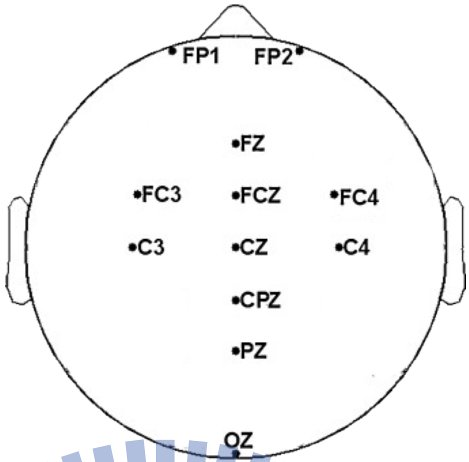

Figure 2.1: The 12 channels utilized in this study.

2.1.3 EEG recording

The EEG signal was recorded with NeuroScan 4.3 software and caps with 32 electrodes located on the standard of the International 10-20 system. The sampling rate of the EEG recordings was 500 Hz with 16 bit A/D conversions. Twelve channels (FP1, FP2, FZ, FC3, FCZ, FC4, C3, CZ, C4, CPZ, PZ, and OZ) which were related to the motor task were applied in this study. The mean value of the earlobe electrodes A1 and A2 was used as the reference. The impedance of the selected channels were kept below 5 $k\Omega$. The digital filter within 5-30Hz was also applied here.

2.2 System overview

In order to investigate two topics, we proposed two systems: a two-stage identification and an adaptive system. Figure 2.2 showed these two proposed systems. In the two-stage identification, the first three steps were the basic architecture of identification systems. Our design was to lower the false classification rate by verifying the output class from the SVM classifier. Therefore, the forth step: candidate selection was proposed. In the candidate selection step, we would list some most probable identities of the given person based on

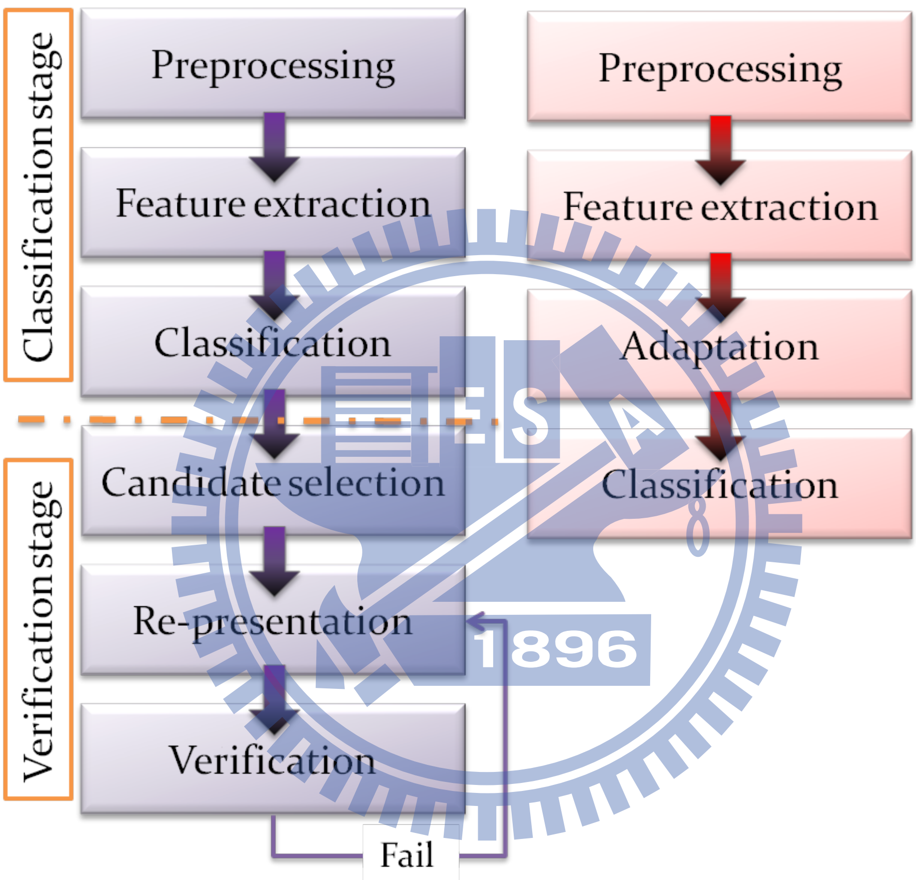

Figure 2.2: Architecture of the proposed systems. Architecture on the left-hand side is the two-stage identification wsystem, and the other is the adaptive system.

the classification results and then we verified the candidates by the order of the likelihood probability. Before the verification step, the features belonging to the candidates were re-represented into a wide separated distribution. Finally, the identification would stop with identifying the given person as one of the candidates, or an intruder if no candidate was verified success.

For longitudinal adaptation, we applied two methods. One was the adaptive system, and the other was the incremental learning. The assumption of adaptation system was that the original data formed a distribution space and the given data formed another distribution space which caused the SVM classifier cannot correct classify the given data. This transformation of EEG signal with respect to the distribution might induce by time or physical state. Based on this assumption, we adapted the data distribution of the training data base with the given data before training the classification model. The unsupervised adaptations were applied here because the given data belonged to anonymous class.

2.3 Two-stage identification system

2.3.1 Signal preprocessing

The following preprocessing procedures were performed with EEGLAB (a MATLAB toolbox) [3]. After setting the information of our EEG recordings, we first segmented the signal into two-second epochs starting at one second before the trigger. The baseline correction was operated with a relative smooth period to remove the drift. Next, we rejected the epochs which might be interfered with the eye blinking. The epochs would be rejected if either the voltage of FP1 or that of FP2 burst extreme high values. There were 90 EOG free of epochs for each subject. In the last step, we pruned the epochs into one second which was the active region (starting at 100ms before the triggers). Figure 2.3 shown the grand average of FC4 with respect to the preprocessing signal.

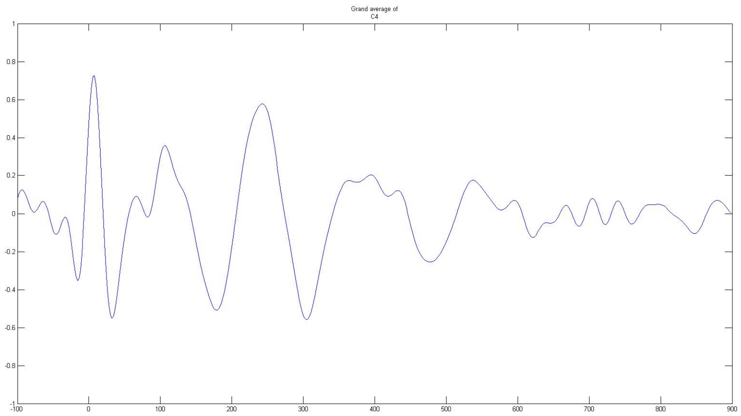

Figure 2.3: The grand average of FC4 with respect to the preprocessing signal.

2.3.2 Feature extraction

Due to the low SNR of the EEG signal, the raw data (the data after the pre-processing) was still in chaos. It would take amount of time, even hardly, for the system to discover the discriminatory information. Because the raw data carried huge useless information, an essential feature extraction can well present the discriminatory information of the EEG and also decrease the classification time. Five feature extraction methods were used in our study, and they could classify into three types:

- Time domain feature: autoregressive model,
- Frequency domain feature: power spectral density, power spectral density, and inter-hemispheric power difference,
- Time-frequency domain feature: discrete wavelet transform.

The features were computed trial wise and channel separation. That is, for each trial, a feature would be computed once for each channel and concatenated in a feature vector.

Autoregressive model (AR)

Autoregressive (AR) model is a predict model of time series that estimates the current observation with the previous observations. Many time-varying series exhibit serial correlation on themselves. The goal of AR model is to find this linear association within signal. The notation AR(p) indicates an AR model of order p. That means the model predicting the observation with p previous points and defined as:

$$X_t = \sum_{i=1}^{P} a_i X_{t-i} + e_t, \tag{2.1}$$

where a_i is a parameter of X_{t-i} and e_t refers to the error term. The AR parameters were estimated by Burgs method, and the order of the AR was eight chosen with the best classification results (Figure 2.4).

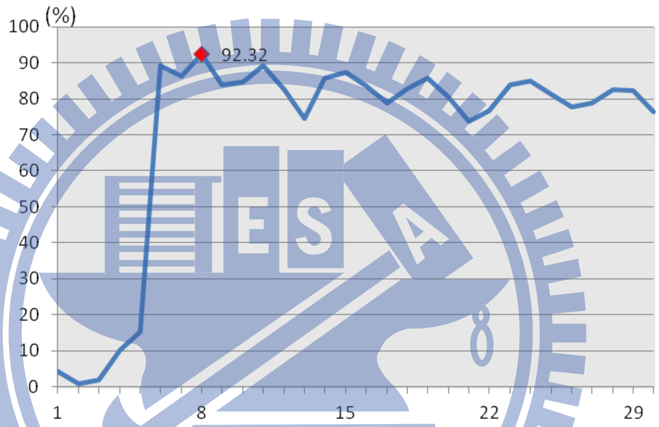

Figure 2.4: Accuracy of the SVM with different AR orders.

Power spectral density (PSD)

Power spectral density (PSD), also be known as power spectrum or spectral density, describes the distribution of energy on frequency scale. It is a common method to analyze a time-series in frequency domain. That is, a time-series can be decomposed with several frequency bases. In our study, PSD was transformed from a time-series by the fast Fourier transform (FFT). The formula of FFT is:

$$X(k) = \sum_{n=0}^{N-1} x_n e^{-i2\pi k \frac{n}{N}}, \tag{2.2}$$

where $k = 0,\ldots,N-1$. We used the $|X(k)|$ as our PSD features.

Band power(BP)

The power of frequency bands is one descriptor of the brain activity. The composition of the power on frequency bands is an evidence to explicate the behavior or constitution of an individual. In this study, band power was integrated with PSD by gathering the total power into following three bands: θ(5-7Hz), α(8-12Hz), and β (13-30Hz). Therefore, the dimensional size of BP for each trial was 36.

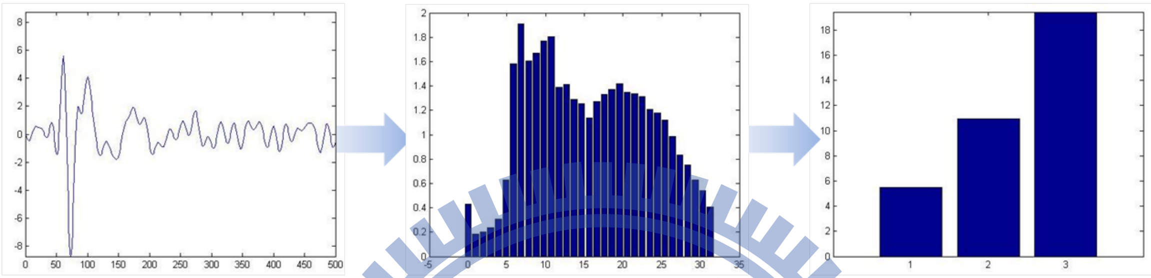

Figure 2.5: The computation process of BP.

Interhemispheric power difference (IHPD)

Interhemispheric power difference (IHPD) focuses on the relationship between the left and right hemispheres. It can be obtained by computing the difference of band power between each pair of electrode in left and right hemispheric. The IHPD was computed as [15]:

$$Power_{difference} = (P_1 - P_2)/(P_1 + P_2), \tag{2.3}$$

where P_1 is the band power on one band of a channel and P_2 is the band power on same band of another channel in the opposite hemisphere. In our study, the system consists of three electrodes on the left hemispheric (FP1, FC3, C3), and other three (FP2, FC4, C4) on the right side. Overall, there are nine differences on each band and the total size of IHPD is 27.

Discrete wavelet transform (DWT)

Wavelet transform is a decomposition method that decomposes the signal with time-scale wavelet functions; that is the basis function only vary in time extension. The Discrete wavelet transform (DWT) is one of them but samples the wavelet discretely. The DWT can both capture the time and frequency information of the original signal. When computing the one level DWT, the input signal will pass through a low-pass and a high-pass filter respectively, and then down sample by two. The approximation coefficient of the input signal can be calculated with the low-pass filter; on the other hand, the detail part is obtained from the output of the high-pass filter. The second level of DWT uses the approximation part computed at the first level as the input signal and repeats the filtering step as below.

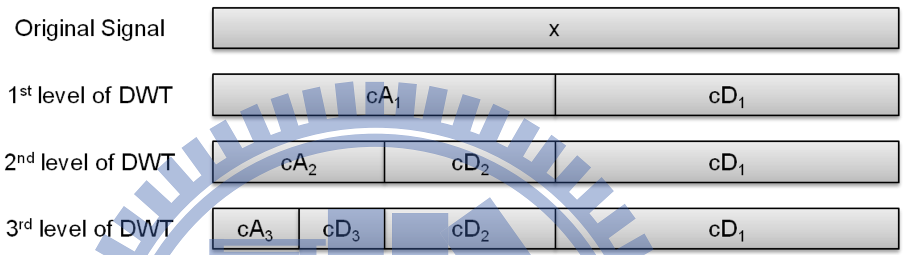

Figure 2.6: The coefficients of DWT with respect to different levels.

In this study, we use the detail part of the third level DWT with a set of Haar wavelets. These parameters are chosen by optimizing the performance of the classification (shown Figure 2.7).

Corr.rate	Haar	size	DB2	size	DB3	size	DB4	size
D1	42.2	250	40.9	251	31.3	252	1.3	253
A1	42.2	250	42.6	251	42.6	252	42.2	253
D1+A1	42.2	500	42.6	502	42.6	504	42.2	506
D2	42.2	125	43.0	127	39.6	128	33.5	130
A2	42.6	125	41.7	127	41.3	128	42.2	130
D2+A2	42.2	250	41.7	254	41.3	256	42.2	260
D3	43.9	63	41.7	65	41.3	66	37.0	68
A3	42.6	63	41.3	65	41.7	66	41.7	68
D3+A3	42.6	126	41.7	130	42.6	132	42.2	136
D4	29.6	32	40.0	34	37.0	35	33.9	37
A4	35.2	32	32.2	34	31.3	35	33.0	37
D4+A4	41.3	64	39.6	68	40.4	70	38.3	74

Figure 2.7: Accuracy of the SVM with different DWT features.

2.3.3 Classification

The classifier used in this study was SVM. SV learning can lead to a high performance in practical classifiers based on its simple ideas. SVM was a supervised classification method. It had to construct a classification model with a set of labeled training data before being tested with the unlabeled given data. The goal of SVM was to find a hyper-plane which could well separate the training data based on their class label and had the maximal margin between different classes. The hyper-plane was defined as:

$$w^T x - b = 0. \tag{2.4}$$

Figure 2.8 shows the example of two-dimensional data. Assume there were a set of training data $\{x_i, y_i\}, i = 1, , n$ and $x_i \in \Re^d, y_i \in \{+1, -1\}$. We wanted to find an optimal separating hyper-plane which allows $y = -1$ on the same side and $y = +1$ on the opposite side and had the maximal margin. The solid line was the optimal separating hyper-plane we wanted, and the dash lines were the support hyper-plane which were parallel with the optimal separating hyper-plane and closest to the data. Therefore, the problem of finding the optimal separating hyper-plane had become the problem to find a pair of support hyper-plane whose margin was the maximum. This problem could be solved as follows:

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2} \parallel w \parallel^2 \\ \text{subject to} \quad & y_i(w^T x_i - b) - 1 \geq 0 \quad \forall i \end{aligned} \tag{2.5}$$

The multi-class and binary SVM classifiers we applied in the classification and verification step were materialize with LIBSVM [2]. The kernel type was linear and the SVM type was C-SVM.

2.3.4 Candidate selection

The candidates were selected with the likelihood probabilities which were the output of SVM classifier. SVM classified the input to the class which had the maximum likelihood probability. The maximum probability could be viewed as how confident the SVM confirmed the input belong to. If the maximum probability was lower than a half, the system would choose three candidates with the largest three likelihood probabilities and moved on

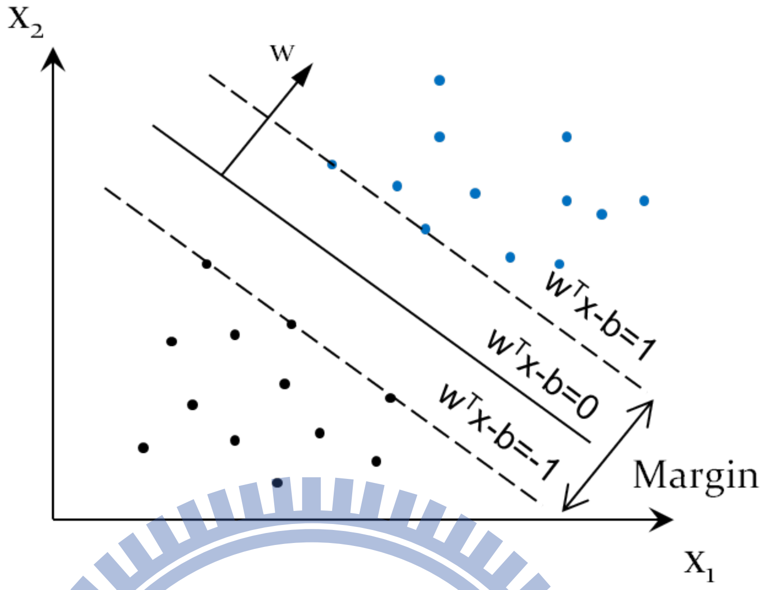

Figure 2.8: Example of 2D data for SVM classification.

to the next step. On contrast, the system identified the input based on classification result of the SVM classifier.

2.3.5 Re-representation

Linear discriminant analysis (LDA) is used in this step to re-represent the feature distribution. LDA is a supervised feature selection method that acquires a linear combination of features which can well separate the classes. In the Fishers linear discriminat, the separability of classes scattered in l-dimension is defined as the ratio of the variance between the classes to the variance within the classes. In order to use the Fishers criteria, the following matrices are defined:

- Within-class scatter matrix

$$S_w = \sum_{i=1}^{M} P_i \Sigma_i \tag{2.6}$$

- Between-class scatter matrix

$$S_b = \sum_{i=1}^{M} P_i \left(\mu_i - \mu_0\right) \left(\mu_i - \mu_0\right)^T \tag{2.7}$$

where Σ_i and P_i are the covariance and the probability of class i, μ_i is the mean vectors of class i and μ_0 is the global mean.The goal of LDA is to maximize the

$$J(W) = \frac{W^T S_b W}{W^T S_w W}, W* = argmax \tag{2.8}$$

by solving the W.

2.3.6 Verification

There were two kinds of classifiers used here to verify the identification of the input. One was the SVM classifier, and the other was a k-nearest neighbors (kNN) classifier. The former one was the same as the classifier used in classification step, while the latter was a classifier making the decision with the closest training data. A kNN classifier classified the input data by the voting of its k nearest neighbors whose distances to this data were the shortest. The distance of the kNN classifier we used here was the Euclidean distance and defined as follows:

$$D(x, y) = \sqrt{\sum_{i=1}^{n} \left(x_i - y_i\right)^2} \tag{2.9}$$

In verification step, the number of classes in the data pool was down from 23 to three. The system trained a one-against-rest classifier for each candidate and verified the candidates one by one. If the first candidate was verified succeed, the system would decide that the input belong to this candidate. Otherwise, the system went through to the second candidate with second largest likelihood probability, and so on. The system would identify the input as an intruder if it could not verify any candidate successfully.

2.4 Longitudinal adaptation

2.4.1 Adaptive system

The adaptation step was process with the procedure shown in Figure 2.10. We used both of training data and testing data to find the adaptive subspace, whose dimensionality was

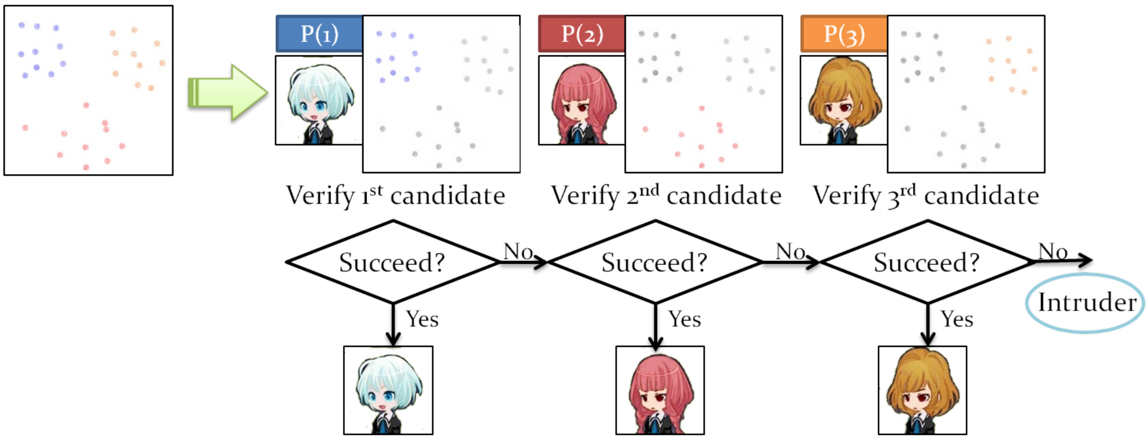

Figure 2.9: The procedure for one-against-rest verification.

smaller than the original, and project the training data and testing data onto this subspace.

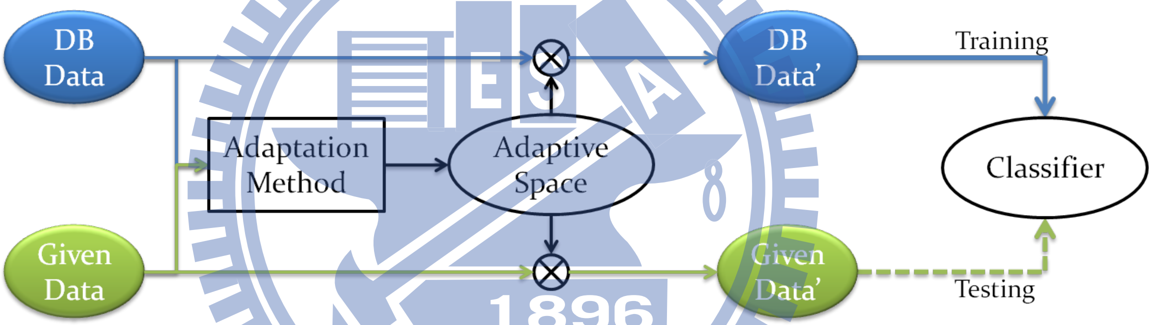

Figure 2.10: The procedure for adaptation step.

There were two unsupervised methods utilized in the adaptation step: principal component analysis (PCA) and domain adaption (DA). The PCA found the adaptive subspace from a mixed data consisting of the training and testing data. While the DA first computed the adaptive subspaces, which were known as domains, of the training and testing data separately and then find the intermediate subspace from the training and testing domain.

Principal component analysis

PCA [6] was mathematical procedure that usually used to convert a set of data possibly consisting of interrelated variables into a set of principle components (PCs), which were linearly uncorrelated to each other, while retaining as much information measured by variance as the original variables. The number of PCs was usually smaller than that of original variables and PCs were in order of the variance it preserved so that PCA transformation had been widely used to reduce the data dimensionality.

Overall, the process to compute PCA can be simplified by following steps. First, compute the covariance matrix of all data. Secondly, obtain the eigenvectors of the covariance matrix in the order of decreasing of the corresponding eigenvalues. The last step, project the data onto the subspace formed by the first few eigenvectors, or named as PCs, which can well present the variance of the original data consisting of the training data and the given input.

Domain adaptation

The DA [5] was proposed to adapt and match the images taken from different cameras which were called source domain and target domain.

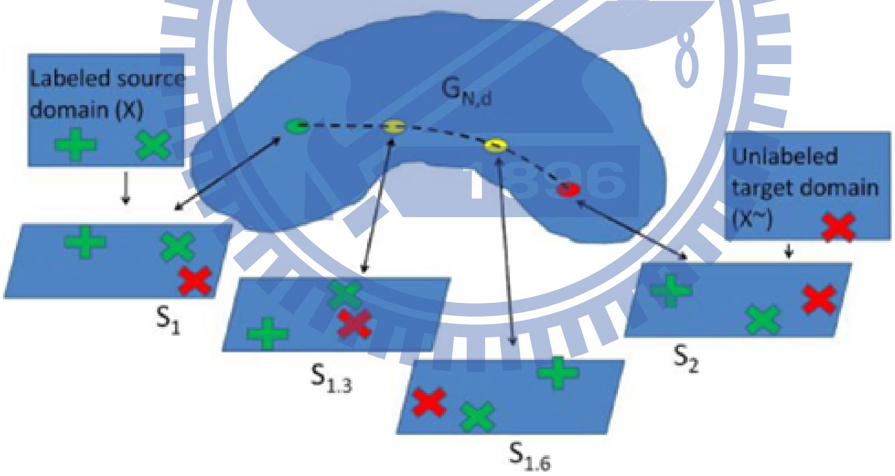

Figure 2.11: The concept of domain adaptation (source: Gopalan et al [5].)

The process of DA was described as follows: Perform subspaces for the source and target domain (here the source domain was defined as the training set, and the target domain was formed by the set of given data) by PCA separately. The dimension of these two subspaces measured by PCA must be the same. Next, compute the geodesic between these subspaces. Third, obtain the intermediate subspace by sampling points along the geodesic.

Finally, project the data both of the source and target domain onto the intermediate subspace. The intermediate subspace was chosen by the best classification results.

2.4.2 Incremental learning

Incremental learning was a supervised longitudinal adaptation method and had been utilized in EEG-based authentication system [12]. The concept of incremental learning was that the classifier would keep updating with the new recordings which were labeled. The classifier could construct a more general model by many EEG recordings acquired from a subject at different time, mental situation and physical status. In this study, the classifier would re-train the decision model with the database and a part of new data with label and be tested with the remained.

2.5 Evaluation methods

Because there were two different systems for different purposes, the evaluation step could also be stepwise divided into three phases. The first phase focused on optimizing the performance of the identification using EEG data acquired on the same day for each subject. In this phase, we attempted to find which or what combination of features was more suitable than others. Since the training and testing data were obtained on the same day, we used 10-fold cross-validation classification (CVC) to increase the reliability of the evaluated performance. The evaluation criteria used here were the true acceptance rate (TAR) and false acceptance rate (FAR). Intuitively, TAR means the probability that the system classifies the input to the correct class. In contrast, FAR refers to the probability of the false classification. Our purpose of this protocol was to lower down the FAR but maintain, or even improve, the TAR. Additionally, the evaluation of the two-stage identification system would consist of the legal test and intruder test. The former evaluated the performance with the given person belonging to the database, while the latter tested the identification system with an intruder (the intruder was not counted in our 23 subjects.).

The goal of the second phase was to find a more stable feature which can resist the decreasing of the performance caused by the time interval between the training and testing data. The stability of the feature was evaluated by the correct classification of the SVM and the coefficient of variation (CV). The CV is the ratio of the standard variation to the mean. We used the ratio of the CV of the single day to the CV of the triple days as a stability index. When the index value was near to one, it meant that the feature is stable. The number of the classes was 14 (only the subjects who had three recordings acquired on three different days were included).

The third phase would first show the decreasing tendency on TAR of the SVM and then evaluate how much the longitudinal adaptation can do to increase the performance. When observing the decreasing tendency, we used the EEG data of all 23 subject obtained on the first day to train the SVM classifier and test the classifier with the subject who had 19 EEG recordings. It could provide long-term variation within the performance of the EEG-based biometrics.

Chapter 3

Experimental Results

3.1 Performance of two-stage identification system

3.1.1 Feature selection and comparison

According to the previous study, we knew that features played significant roles in improving the performance of the EEG-based biometrics. In order to optimize the performance of the proposed person identification, we should choose a set, or a combination, of features which provided the generically-specific information of the EEG signal. Therefore, the feature selection and comparison evaluated began with single type of features. We first used raw data (the signal which was only processed with the pre-processing step) and five features as the input of the SVM separately.

The evaluation results are shown as Figure 3.1. The results indicated that the SVM using BP as the feature had the best correct classification rate amongst all features. The accuracy was 85.5% at single-trial level, and it could achieve 94.7% when averaging two trials as a unit. Figure 3.1 also shows the tendency of the classification results of the SVM with respect to the different number of trials. The accuracy of the SVM using BP as the feature became saturated when the number of averaging trials was accumulated to five.

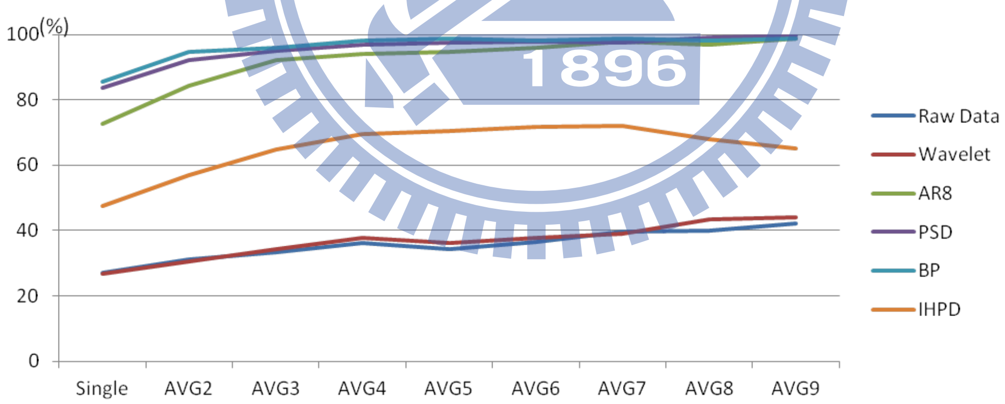

Figure 3.1: Classification results of each feature averaging with different number of trials.

This phenomenon was not quite clear in the performance of the SVM with other features. When we averaged six trials as a unit, the correct classification rate of the SVM using PSD as the feature was nearly saturated, but it still increased in a small amount. When the averaging number of trials came to eight trials, the performance of the classifier using PSD as the feature could achieve 99.1% which was higher than that using BP on the same situation (98.3%). In Table 3.1, we marked the saturation numbers of averaging trials of each feature in bold face. These saturation numbers were defined by the local maximum estimated from the single-trial level. It was obvious that the performance of the features was saturated with different number of averaging trials. In order to comparing the combination of features, we set the saturation number of averaging trials with single feature at six-trial level which was the mean saturation number amount of all features.

Table 3.1: Classification results using SVM with different features across number of trials. The size of each feature are 6000, 756, 96, 396, 36 and 27 from the top of table to the bottom sequentially.

Feature	Single	AVG2	AVG3	AVG4	AVG5	AVG6	AVG7	AVG8	AVG9
Raw Data	27.1%	31.2%	33.3%	**36.3%**	34.3%	36.5%	39.6%	40.0%	42.2%
DWT	26.9%	30.5%	34.3%	**37.8%**	36.1%	37.8%	39.1%	43.5%	43.9%
AR8	72.8%	84.2%	92.3%	94.1%	94.8%	96.1%	**97.8%**	97.0%	98.7%
PSD	83.8%	92.1%	94.9%	97.0%	97.4%	**97.8%**	97.4%	99.1%	99.6%
BP	85.5%	94.7%	96.1%	98.0%	**98.7%**	98.3%	98.7%	98.3%	98.7%
IHPD	47.4%	57.0%	64.8%	69.6%	70.4%	71.7%	**72.2%**	67.8%	65.2%

To observe the performance of the SVM classification using combination features, we attempted to use each pair of features as the input of the SVM. Each pair of features was not normalized before combination and just concatenated into one feature vector. The classification results at single-trial level showed that combining AR8 and BP could achieve the highest correct classification rate. The performance was 90.8% which was also higher than the correct classification rate using either AR8 or BP (AR8:72.8%, BP:85.5%).

We also increased the number of averaging trials of the SVM using combination features from single-trial to six-trial level which we decided as the saturation number of averaging trials of the SVM using single type of feature. According to the results shown in Table 3.2, we could observe that the combination of AR8 and BP obtained the highest correct classification rate when averaging fewer trials. If the number of averaging trials was increasing to four, the best combination feature would be AR8 and PSD (98.8%) instead of

AR8 and BP(98.5%). However, the larger number of averaging trials we utilized, the more time it took for the identification system to identify an individual. Because a subjects had to lift his or her left index finger more times and the interval between lifting was about two seconds. That is, a subject would enter the system after two or ten seconds. Consequently, we still chose the combination of AR8 and BP as the feature to be utilized in the proposed two-stage identification system, and set the saturation number of averaging trials at three since AR8 and BP was the best combination when the number of averaging trials was lower than four.

Table 3.2: Classification results of combination different two of features. In this study, the best combination of the features is combining AR8 and BP (90.5%).

Single Trial

f1\f2	AR8	PSD	BP	IHPD
DWT	42.0%	37.1%	57.1%	26.9%
AR8	–	89.6%	**90.5%**	74.3%
PSD		–	89.6%	83.8%
BP			–	85.4%

AVG2

f1\f2	AR8	PSD	BP	IHPD
DWT	45.5%	41.0%	64.6%	30.5%
AR8	–	95.4%	96.6%	84.7%
PSD		–	96.0%	92.3%
BP			–	94.8%

AVG3

f1\f2	AR8	PSD	BP	IHPD
DWT	49.9%	44.2%	69.6%	34.3%
AR8	–	97.1%	98.1%	92.6%
PSD		–	97.8%	95.1%
BP			–	96.1%

AVG4

f1\f2	AR8	PSD	BP	IHPD
DWT	53.7%	47.6%	75.2%	37.8%
AR8	–	98.9%	98.5%	94.1%
PSD		–	97.8%	97.2%
BP			–	98.0%

AVG5

f1\f2	AR8	PSD	BP	IHPD
DWT	55.7%	51.3%	77.4%	36.1%
AR8	–	98.7%	98.3%	94.8%
PSD		–	98.7%	97.4%
BP			–	98.7%

AVG6

f1\f2	AR8	PSD	BP	IHPD
DWT	58.7%	52.2%	82.2%	37.8%
AR8	–	99.6%	99.6%	96.1%
PSD		–	98.7%	98.3%
BP			–	98.3%

3.1.2 Performance of two-stage identification

Legal test

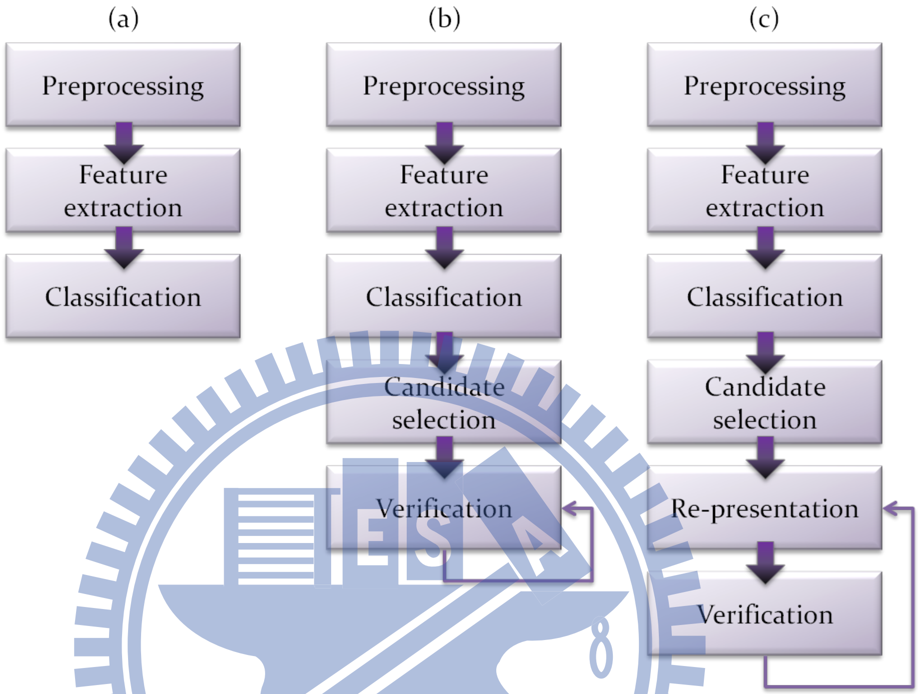

Figure 3.2: Evaluation procedures for the two-stage identification systems. (a) A basic identification used as a control group. (b) A two-stage identification without LDA represented the candidates. (c) The proposed two-stage identification system.

Since that the combination of AR8 and BP obtained the highest correct classification rate at single-trial level, the combination of AR8 and BP was applied as the feature in our two-stage identification system. The evaluated process was divided into three procedures shown in Figure 3.2. The procedure (a) was a basic identification system without representation and verification, while (c) was the two-stage identification we introduced in Section 2.3. The procedure (b) was a two-stage identification the same as (c) expected the re-representation step. Two classifiers introduced in Section 2.3.6, SVM and kNN, were applied to verify the results of the SVM in the two-stage identification systems. The parameter k of the kNN classifier was chosen by optimizing the performance, and it turned out that 3NN had the highest TAR and lowest FAR amongst the kNN classifiers with k

belonging to odd numbers which started from one to nine (shown in Table 3.3).

Table 3.3: Performance of kNN classifiers with different number of k

Verification method	1NN	3NN	5NN	7NN	9NN
TAR	88.5%	88.3%	87.8%	87.7%	87.5%
FAR	5.7%	5.3%	5.1%	5.1%	5.1%

Table 3.4: Performance comparison with respect to verification and re-representation.

Procedure	(a)	(b)		(c)	
Verification method	Non	SVM	3NN	SVM	3NN
TAR	90.5%	85.4%	88.4%	87.7%	88.3%
FAR	9.5%	5.3%	9.7%	5.4%	5.3%

Table 3.4 shows the TAR and FAR of these procedures at single-trial level. The identification with procedure (a) achieved the highest value on the true acceptance rate; however, it also nearly had the highest false acceptance rate. The FAR of the system built with procedure (b) using the SVM in verification was a half of that of the system built with procedure (a), but the TAR of the former was decreasing by 5%. While the system built with procedure and utilizing 3NN in verification could maintain the TAR but increase the FAR. According to the above findings, the two-stage identification without LDA in re-representation step could either lower the FAR but sacrifice much on TAR, or maintain the TAR but do nothing on decreasing the FAR. In this case, it seemed that applying the SVM in verification was closer to our purpose of lowering the FAR. On the other hand, the evaluation results of procedure (c) showed the proposed two-stage identification using either the SVM or 3NN classifier could reduce the FAR as well as retain the TAR. The loss on TAR was about 2% comparing to the simple identification, and the FAR was a half of the FAR of the simple one. It was quite clear that the proposed two-stage identification made a dent in lowering the FAR while maintaining the TAR.

Based on the results in Section 3.1.1, we also applied the proposed two-stage identification systems with different number of averaging trials from single trial to three trials.

Figure 3.3 shows the TARs and FARs of the basic identification and two-stage identification systems. The basic identification system, as procedure (a) in Figure 3.2, was labeled as 'Non' in Figure 3.3. The proposed two-stage identification systems were labeled with their classifier in the verification step. As the number of averaging trials increased, the differences between TARs of these three systems were closer. While the FAR of the basic identification at two-trial level was a half of that at single-trial level; it was even lower than the FAR of the two-stage identification systems at single-trail level. However, when averaging three trials as a unit, the FAR of the basic identification could not reach the FAR of the two-stage identification systems at two-trial level. It seemed that the best number of averaging trials of the two-stage identification systems was two in which the two-stage identification system could bring talent into full play.

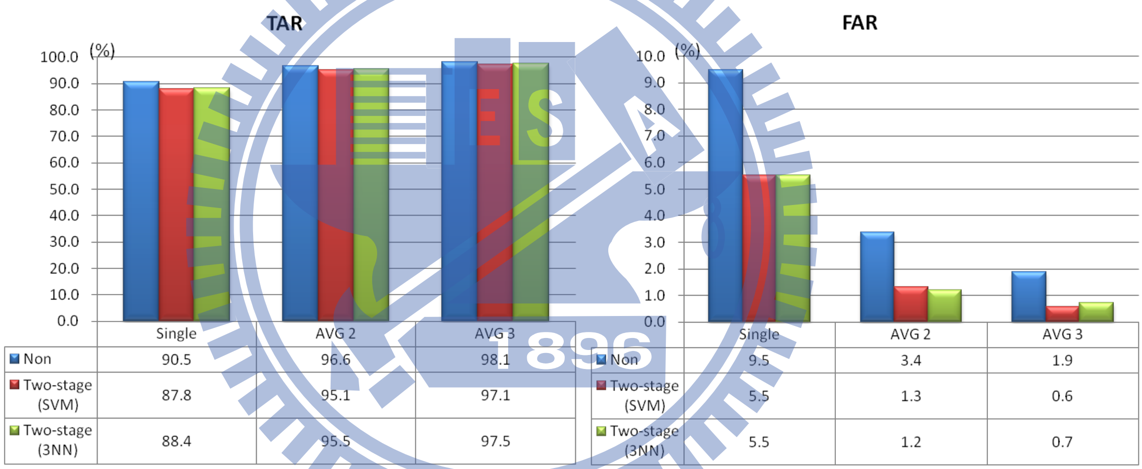

Figure 3.3: TAR and FAR of two-stage identification system for averaging different number of trials.

Intruder test

To confirm the proposed system can actually reduce the risk of being intruded, we also evaluated the performance with EEG recordings acquired from a person not belonging to the database. The results for the basic identification and two-stage identification systems with two different verification methods are shown in Table 3.5 from single-trial to three-trial level.

Table 3.5: Comparison of detection rates for intruders amongst different identification systems.

Trial level	Single Trial	AVG2	AVG3
Non	0.0%	0.0%	0.0%
Two-stage(SVM)	43.3%	66.7%	73.3%
Two-stage(3NN)	54.4%	64.4%	66.7%

It was rational that the basic identification without verification could not detect any intruders, while the successful detection rate of the two-stage identification system could reach 54.4% at single-trial level and increase much more by averaging two trials as a unit. When the number of averaging trials was two, the detection rate of the two-stage identification system using SVM in verification achieved 66.7% which was much more than that of the two-stage identification at single-trial level (43.3%). However, the benefit of increasing the number of averaging trials to three was not as much as the above. This phenomenon was quite obvious in the results of the system using 3NN in verification. The detection rate of the system using three trials as a unit only increased by 2.3% from that of the system at two-trial level. Therefore, we could conclude that two was the best number of averaging trials for the two-stage identification system which using motor movement as the task and the combination of AR8 and BP as the feature.

3.2 Stability of features

To observe the performance varying with respect to the time interval of the acquisition date between the training and testing data, we rebuilt and tested the SVM with 14 subjects who had participated the experiment at less three times. We first used 10-fold CVC to evaluate the performance the SVM of which the training and testing data were acquired on the same day. Then, we used whole EEG data acquired on the first day of these 14 subjects to train a SVM model and test with the EEG data acquired on the second and third day. The classification results are shown in Table 3.6. We could observe that the SVM with BP has the highest correct classification rate if the training and testing sets were acquired on the same day. The SVM using PSD as the feature, however, obtained the highest correct classification rate when the time interval between the training and testing data was more than a month. According to the correct classification rates, we could list the features ordered by the stability from high to low as following: PSD, BP, AR8, IHPD, DWT, and raw data.

Table 3.6: Comparison of classification results for the training and testing data acquired at the same or different days. The SVM using PSD as the feature had the highest accuracy rate amongst all features.

Training	Testing	Raw Data	DWT	AR8	PSD	BP	IHPD
Day1	Day1	31.6%	32.2%	76.5%	87.7%	86.8%	46.5%
Day2	Day2	31.0%	31.2%	69.9%	84.9%	87.2%	45.9%
Day3	Day3	40.9%	38.3%	70.2%	86.1%	87.8%	47.5%
Day1	Day2	8.8%	11.0%	28.1%	41.8%	36.3%	15.0%
Day1	Day3	10.0%	11.4%	16.0%	42.9%	34.9%	17.5%

Another index that we applied to evaluate the stability of features was the ratio of CV. For each feature, we first calculated CVs of different days for each 14 subject. Consequently, the subjects would have three CVs of each feature. These three three CVs were labeled with the acquisition day of the EEG signals, such as CV(Day1) meant that the CV was computed with the EEG recordings obtained on the first day. Additionally, we also computed a CV(Dall) of which the EEG data contained the recording acquired on the three days for subjects. Finally, we calculated the mean values of these CVs amount of 14 subjects and used the ratio of the CV of a single day to the CV of three days as the stability index. The results are shown in Table 3.7. According to evaluation results, we could also list the features in the order of stability from high to low by CVs: PSD, BP, DWT, IHPD, raw data and AR8.

Despite that there exited difference between two lists sorted by different stability indexes, both of the correct classification rate and ratio of CV indicated that PSD was the most stable feature amongst all features we used as shown in Figure 3.4 and Figure 3.5. However, the correct classification rate of the identification systems using PSD features

Table 3.7: Ratio of the CV of one day to the CV of three days.

	Raw Data	DWT	AR8	PSD	BP	IHPD
CV(Day1)/CV(Dall)	0.4	3.3	1.7	1.0	0.9	0.3
CV(Day2)/CV(Dall)	0.4	0.2	6.0	1.0	0.9	1.0
CV(Day3)/CV(Dall)	0.4	0.6	2.0	1.0	0.9	0.2
Mean	0.4	1.4	3.2	1.0	0.9	0.5

was decreased a half since the training and testing data were not acquired on the same days. The performance of the SVM of which the training and testing EEG recordings were acquired on the second day promised that it was not man-made muff induced the decreasing on performance. Therefore, we would investigate a longitudinal adaptation of the EEG-based identification system using PSD– the most stable feature.

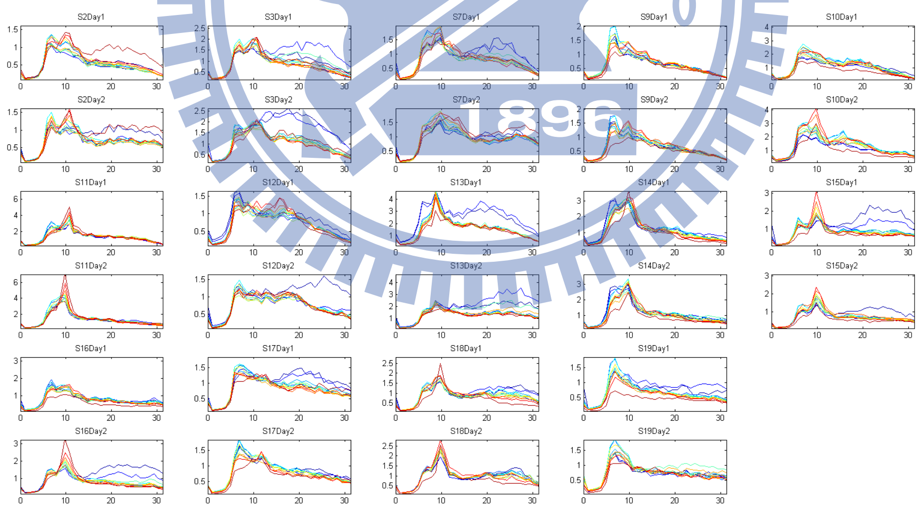

Figure 3.4: The PSD with respect to the EEG acquired on the first and second day for each subject who participated at least three times.

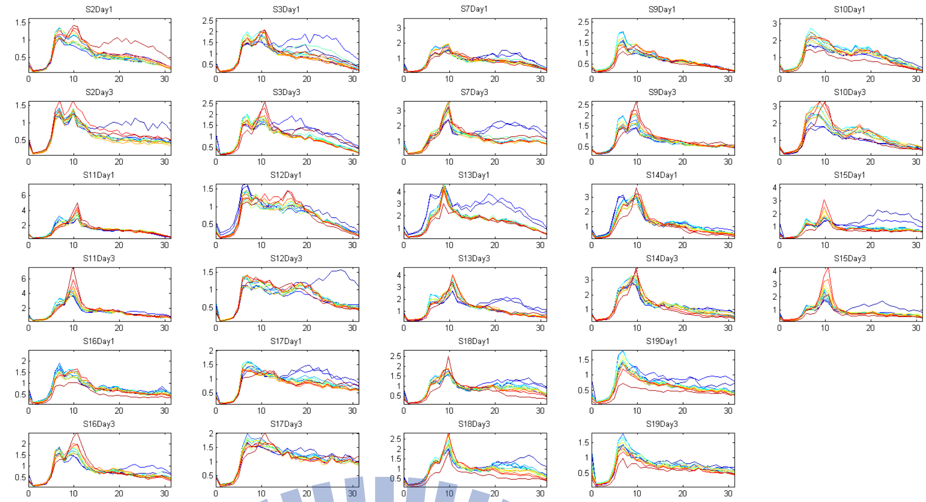

Figure 3.5: The PSD with respect to the EEG acquired on the first and third day for each subject who participated at least three times.

3.3 Performance of longitudinal adaptation

3.3.1 Adaptive system

Because the adaptive system was developed for improving the stability of the EEG-based identification system which the training and testing data were acquired on different days, we evaluated the adaptive system with three identification systems with different architecture:

1. A basic identification system without adaptation step (labeled as 'None'),

2. An identification system using PCA in the adaptation step (labeled as 'PCA'),

3. An identification system using DA in the adaptation step (labeled as 'DA').

The training pool of these systems was formed with the EEG recordings acquired on the first day amongst of 23 subjects, and the testing data was the EEG signals of the subject who participated 19 times. Amongst of recordings acquired on 19 different days, the data collected earliest was used to train the SVM model and others were used as the testing data separately. Table 3.8 shows the correct classification rates of three systems at single-trial level. Each row meant the correct classification rates of different systems, and each column meant that we test the identification system with EEG signal acquired on which day. In adaptive systems, the adaptation step were executed with the database contained only the EEG recordings acquired on the first day and the testing data. That is, when we tested the system with the recordings acquired on the third day, we did the adaptation based on the EEG signals obtained on the first and third day.

Table 3.8: Correct classification rate of the adaptive system at single-trial level.

Day#	**D2**	**D3**	**D4**	**D5**	**D6**	**D7**	**D8**	**D9**	**D10**
None	21.1%	56.7%	34.4%	3.3%	5.6%	4.4%	12.2%	1.1%	16.7%
PCA	35.6%	66.7%	53.3%	17.8%	10.0%	4.4%	15.6%	1.1%	24.4%
DA	31.1%	48.9%	25.6%	11.1%	6.7%	5.6%	15.6%	3.3%	13.3%
Day#	**D11**	**D12**	**D13**	**D14**	**D15**	**D16**	**D17**	**D18**	**D19**
None	0.0%	6.7%	1.1%	2.2%	2.2%	0.0%	11.1%	6.7%	33.3%
PCA	0.0%	11.1%	5.6%	3.3%	2.2%	1.1%	13.3%	10.0%	37.8%
DA	1.1%	7.8%	2.2%	5.6%	1.1%	2.2%	13.3%	4.4%	26.7%

According to the correct classification rate of system 'None', we could roughly observe that the correct classification rates were decreased as the time interval between the training and testing data increased. The correct classification rates of the adaptive system using PCA were all higher than the non-adaptive identification system. The accuracy of the above two identification systems were sometimes equal, when the accuracy of the non-adaptive system was lower than 5% (D7, D9, D11, and D15). On the other hand, the correct classification rates of system 'DA' were basically higher than that of system 'None'. Based on the comparison of the results of system 'PCA' and 'DA', it was obvious that adaptive system with PCA did much effort on improving the stability of EEG-based identification system.

For a further improvement of stability, we increased the number of averaging trials. The mean values of correct classification rate amount of 18 days are shown in Table 3.9. When averaging two trials as a unit, the performance of system 'None' was extended by

3.7% from 12.2% to 15.9%, which was less than the increasing amount of correction rates affected by the adaptive system using PCA at single-trial level. That is, a EEG-based identification system could enhance more on the stability by using the adaptive system applied PCA in adaptation than using averaging trials. According to Table 3.8, however, when we tested the adaptive system with the data collected on the fifth day, the performance was not that ideal. Furthermore, when the testing data was acquired after the sixth day, the increasing amount of correct classification rates affected by the adaptive system was not obvious. Accordingly, we determined that the adaptive system could used to maintain the stability of EEG-based identification, but it still had its limitation.

Table 3.9: Performance of adaptive system with different number of trials.

	Single trial	**AVG2**	**AVG3**
None	12.2%	15.9%	20.0%
PCA	17.4%	18.6%	22.4%
DA	12.5%	16.0%	19.6%

3.3.2 Incremental learning

In order to evaluate the performance of incremental learning, we also tested the identification systems using incremental learning with the PSD features extracted from the EEG signals of the subject who participated 19 times. The database contained the EEG data acquired only on the first day of each subject, excepting the subject whose data was used to test the classifier. For this subject, acquiring date of her EEG data in the database was extended from the first day to the nineteenth day sequentially. For each extended, we would retrained and tested the SVM. These 19 classifiers were labeled by the training set of itself. For example, the classifier we trained with the EEG recordings of 23 subjects acquired on the first day was labeled as ’D1-D1’. While the classifier we trained with the EEG recordings of 22 subjects acquired on the first day as well as the EEG signal obtained on the first and second days of the subject who participated 19 times was labeled as ’D1-D2’. However, if all 90 recordings of that subject acquired on each day were all used to train the classifier, the classification model might be dominated by this subject. Because the subject had 19 times recordings more than others. Therefore, the incremental learning was processed by training with a part of older and newer recordings which total number was 90 the same as the number of EEG recordings for other subject. That is, every subjects all had 90 feature vectors in the database, we would use 45 recordings obtained on the first day and other 45 recordings acquired on the second day when constructing the 'D1-D2' classifier. The remained EEG recordings would be utilized to evaluate the performance of the classifier. The results are shown in Figure 3.6.

In Figure 3.6, values in the blue clocks were the correct classification rate of the SVM in which the database contained all the previous data collected before the testing data. That is, we used all older EEG signals to learn the classifier, and tested this classifier with the data collected on the next time. Additionally, the correct classification rates in the green blocks were evaluated by dividing the given data into a group utilized to learn the classifier and the other group used to test this classifier. These evaluated performance could be viewed as the upper bound of classification results.

For the blue blocks, most of the correct classification rates in these blocks were higher than the performance of the SVM in which the database only contained the EEG recordings collected on the first day (the 'D1-D1' classifier). When we used the recordings obtained on the fourth day to test the 'D1-D3' classifier which was incremental learning with the data acquired on the first three days, the correct classification rate reached 82.2% and was much higher than the rate of the classifier without incremental learning (34.4%). In addition to the mean value amount of correction rates, it was obvious that the incremental learning affected very much on enhancing the stability of EEG-based identification system. However, when we evaluated the performance with the data acquired on the seventeenth day, the accuracy of the 'D1-D16' classifier was lower than that of the 'D1-D1' classifier. This wired result could also be found in the performance comparison of the 'D1-D18' and 'D1-D1' classifiers with the testing data obtained on the nineteenth day. It was probably caused by the fixed number of trials in the database. In the case, the more number of acquisition dates was in the training set, the more information of recordings collected on each day might be diluted.

	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	D17	D18	D19	Mean
D1-D1	21.1	56.7	34.4	3.3	5.6	4.4	12.2	1.1	16.7	0.0	6.7	1.1	2.2	2.2	0.0	11.1	6.7	33.3	12.2
D1-D2	44.4	65.6	84.4	20.0	23.3	8.9	14.4	11.1	21.1	13.3	33.3	15.6	13.3	8.9	3.3	4.4	5.6	11.1	22.3
D1-D3	43.3	80.0	82.2	21.1	17.8	3.3	10.0	5.6	26.7	6.7	8.9	11.1	7.8	3.3	3.3	11.1	8.9	30.0	21.2
D1-D4	47.8	80.6	85.1	23.3	20.0	0.0	8.9	6.7	23.3	6.7	4.4	10.0	14.4	3.3	3.3	7.8	4.4	24.4	20.8
D1-D5	61.1	76.4	87.5	63.9	47.8	14.4	13.3	23.3	45.6	23.3	32.2	37.8	26.7	15.6	6.7	7.8	6.7	21.1	34.0
D1-D6	65.3	68.0	85.3	72.0	53.3	31.1	23.3	37.8	42.2	35.6	66.7	67.8	41.1	21.1	11.1	8.9	4.4	22.2	42.1
D1-D7	75.3	66.2	85.7	81.8	61.0	58.4	37.8	60.0	53.3	43.3	75.6	81.1	53.3	33.3	18.9	6.7	7.8	17.8	51.0
D1-D8	72.2	57.0	84.8	72.2	63.3	65.8	43.0	63.3	50.0	46.7	77.8	85.6	52.2	30.0	17.8	7.8	7.8	17.8	50.8
D1-D9	70.0	48.8	81.3	67.5	63.8	62.5	38.8	56.3	50.0	52.2	77.8	86.7	57.8	27.8	18.9	7.8	6.7	13.3	49.3
D1-D10	71.6	48.1	82.7	69.1	63.0	59.3	33.3	58.0	43.2	51.1	77.8	82.2	55.6	28.9	15.6	4.4	4.4	5.6	47.4
D1-D11	70.7	42.7	82.9	67.1	61.0	61.0	32.9	57.3	45.1	75.6	81.1	84.4	67.8	33.3	17.8	3.3	4.4	2.2	49.5
D1-D12	69.5	41.5	81.7	69.5	61.0	63.4	34.1	59.8	45.1	74.4	78.0	84.4	64.4	35.6	18.9	3.3	3.3	2.2	49.5
D1-D13	69.9	45.8	81.9	68.7	57.8	57.8	32.5	60.2	34.9	66.3	77.1	80.7	61.1	32.2	11.1	1.1	2.2	2.2	46.9
D1-D14	72.6	35.7	79.8	58.3	58.3	59.5	29.8	61.9	39.3	70.2	81.0	79.8	70.2	41.1	15.6	2.2	1.1	2.2	47.7
D1-D15	75.0	35.7	81.0	63.1	61.9	63.1	34.5	63.1	41.7	79.8	84.5	85.7	76.2	58.3	18.9	2.2	1.1	3.3	51.6
D1-D16	75.0	38.1	78.6	72.6	64.3	64.3	38.1	66.7	54.8	75.0	79.8	84.5	79.8	69.0	38.1	4.4	1.1	3.3	54.9
D1-D17	71.8	31.8	63.5	52.9	63.5	60.0	37.6	58.8	55.3	67.1	85.9	71.8	80.0	70.6	37.6	18.8	15.6	11.1	53.0
D1-D18	67.1	32.9	60.0	49.4	62.4	68.2	43.5	56.5	57.6	70.6	89.4	71.8	81.2	68.2	35.3	18.8	17.6	15.6	53.7
D1-D19	62.4	41.2	63.5	58.8	72.9	69.4	58.8	61.2	62.4	69.4	91.8	80.0	84.7	70.6	45.9	25.9	18.8	22.4	58.9

(%)

Figure 3.6: Classification results of the SVM using incremental learning as the longitudinal adaptation.

Chapter 4

Discussion

4.1 Length of epochs

When choosing the epoch size in the preprocessing step, we did not know which period of EEG recording carrying much discriminatory information while subjects lifting their left index finger. There were two periods we chose to use: "-100 ms - +400 ms" and "-100 ms +900ms". The former one concluded most responses of a movement, while the latter also concluded that as well as the post-movement. However, we finally decided using the latter one because the accuracy rate.

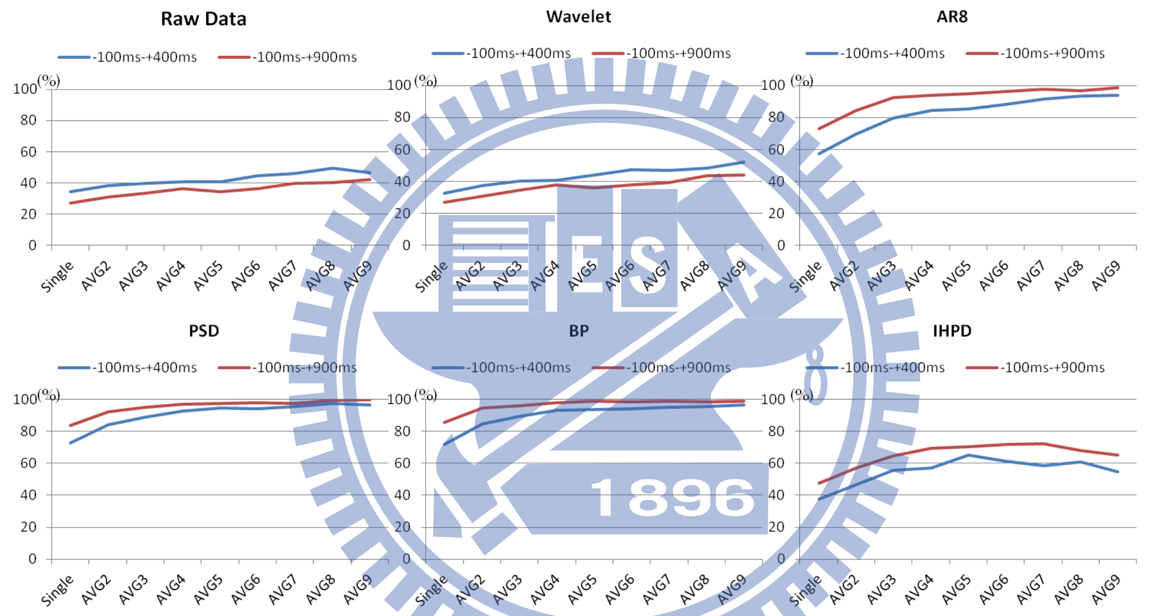

Figure 4.1: Performance of the identification with different epoch lengths using different features.

Figure 4.1 shows the correct classification rate of the identification with the epoch cut in these two periods. It was obvious that the SVM with shorter epochs performs better than the SVM with longer epochs when using raw data and DWT as the feature. In contrast, while using the other features, the comparing results were totally opposite. The classification results with respect to the SVM using other four features showed that longer epochs cover more discriminatory information.

Moreover, the identification with longer epochs using BP (we evaluated as the best feature in Section 3.1.1) as the feature was about 10% higher than that with shorter epochs at single-trial level. Although the gap between them was almost eliminated when averaging nine trials, we still chose "-0.1ms-+0.9ms" as the epoch length. Because the more trials we used to average into a unit, the more times it took for subjects to lift their index-finger.

4.2 Number of candidates

The candidate selection step played a significant role in our two-stage identification. The number of candidate became an important issue. The issue contained "when should the system constrain to choose candidates" and "how many candidates should be listed".

- *When should the system constrain to choose candidates?*

The first question could come out with a statistic solution. After the classification step, we got a list of classification results and the corresponded likelihood probabilities. That is, for each classification result, we had the probability which the SVM classified the input based on. We computed the mean of the likelihood probabilities for both correct and false classification results. The values are respectively 73% and 42%. Therefore, when the SVM classify the input with only 42% likelihood probability, this result might mostly be incorrect. On the opposite side, we can trust the classification result if the SVM classifies the input to a class with the probability more than 74%. According to the above opinions, here were two thresholds we can use: 50% and 80%.

- *How many candidates should be listed?*

In this study, the number of candidates was also set with a statistic result. Based on the likelihood probability of the SVM, we could find which position in the candidate list ordered by the likelihood probability the correct class of the testing data won. Accumulating the places, we got a result shown in Figure 4.2.

We could observe that more than 95% of the correct class would be in the first three places. Therefore, we chose the three candidates whose probability was the largest. While the fixed number of candidates is not the only choices, we also comparing the variable number of candidates.

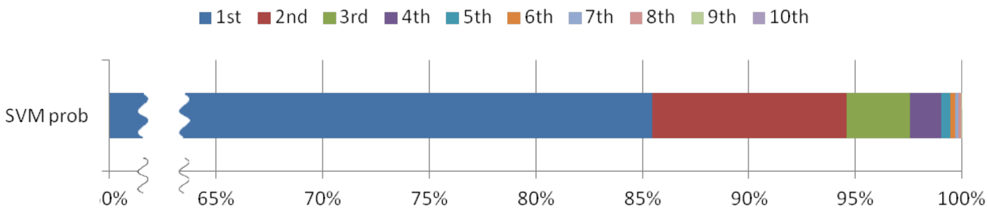

Figure 4.2: Accumulated percentage of the position of the test data in the candidate list ordered by the likelihood probability.

The variable number of candidate was deciding by the accumulated likelihood probabilities. We first sorted the probabilities in descend and then accumulated the probabilities beginning with the largest one until the sum of these probability is more than the threshold. The thresholds here were also set by the same way with that of the fixed number of candidates.

Table 4.1: Performance of two-stage identification systems with different criteria of candidate selection.

Verify type	SVM				3NN			
Num. of candidates	fixed		variable		fixed		variable	
Threshold	0.5	0.8	0.5	0.8	0.5	0.8	0.5	0.8
TAR	87.83%	83.29%	90.34%	81.64%	88.36%	84.15%	90.39%	82.08
FAR	5.51%	4.35%	7.44%	4.54%	5.51%	4.64%	7.58%	5.27

For the two-stage identification system with either fixed or variable number of candidates, the higher threshold (0.8) can lower the FAR but the TAR even more. The two-stage identification system set threshold at 0.5 can decrease the FAR and also maintain the TAR (comparing the simple identification system shown on Table 3.4). The two-stage identification system with threshold at 0.8 and floating number of candidate was the strictest criteria for the classifier in the classification step. The SVM should classify the input with more than 80% likelihood probability– otherwise the system would verify the classification results with the most number of candidates. However, comparing to the identification system with the same threshold and classifier in verification , the fixed number ones had higher TAR and lower FAR. It was obvious that considering more candidates does not help more to the one-against-all verification step.

Overall, the higher, or said strict, threshold can decrease more FAR but also sacrifice more TAR. If the correct classification rate of a classifier was quite confident, combining a lower threshold would apply the system more safety with a lower FAR. Additionally, the number of the candidates also depended on the correct classification result of the classifier. It was useless to choose too much candidates. In our two-stage identification system, the best number of candidates is three.

4.3 Relationship between EEG data and their acquisition dates

According to the Table 3.8 and Figure 3.6, we could observe that the decreasing tendency was not really related to the interval between the training and testing set. Because the performance evaluated with the testing data acquired on the third day (56.7%) was higher than the performance evaluated with the recordings collected on the second day (21.1%). Therefore, we used 10-fold CVC to classify these 19 groups of recordings labeled by their order of the acquired day. The classification results were shown in the form of the confusion matrix and we converted the matrix into an intensity image (Figure 4.3).

In Figure 4.3, the brighter meant that the more data were classified into that class. We could observe that the blocks on the diagonal line were obviously brighter than the remained blocks. There might be some differences between data collected on different day for the same subject, so that most data could be classified into the class labeled by the day when it was acquired. While the correct classification rate of D7 somehow lowered than others, and the signals of D7 were classified into the class around D7. This result provided a reason to explain why the 'D1-D7' classifier achieved the local maximum of the performance. Because D7 had the information similar to other days, such as D8, the information we learning from D1 to D7 would similar to that from D1 to D8 so that the 'D1-D7' classifier had the performance the same as the performance which the 'D1-D8' had. In Figure 3.6, when we used the recordings collected on 13th day as the testing data, the 'D1-D6' classifier had much improvement on the performance than that of the 'D1-D5'

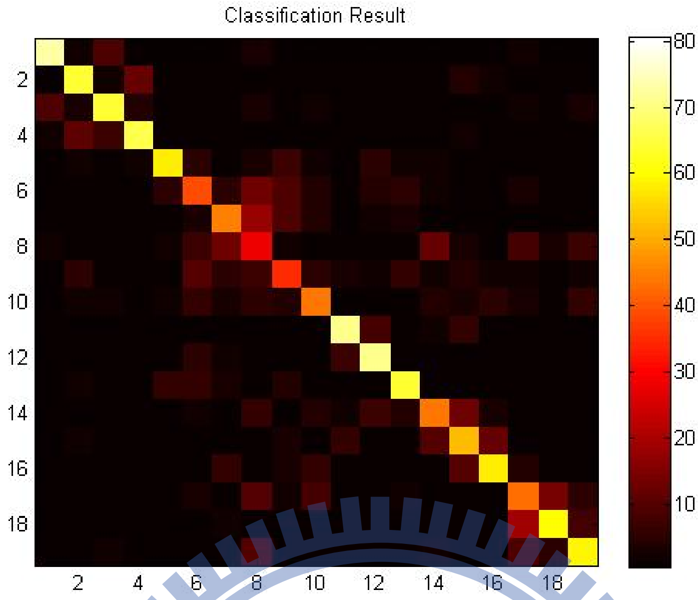

Figure 4.3: Image of the confusion matrix with respect to the intra-subject classification.

classifier, and the 'D1-D9' classifier had the highest performance amongst all classifiers with incremental leaning. We might assume that D6, D9, and D13 might similar to each other. In Figure 4.3, the brighter blocks on (D6,D13), (D6,D9) and (D9,D13) also provided an evidence that these days shared a part of information so that they would be classified into each other.

Based on above results, we knew that the information carried by the EEG would change during times, but these changes were not totally independent to the past. Sometimes, the changed information would similar to that of someday in the past. If the change of EEG was moved in the circle, maybe it might be possible to collect enough group of EEG recordings which can covered all the situation of this subject.

Chapter 5

Conclusions

In this thesis, we have proposed a two-stage system for EEG-based person identification. For classification stage in the proposed two-stage system, according to our experimental results of feature extraction, the combination of BP and AR8 can extract the most discriminatory information of human brainwaves. For the verification stage, the best threshold of the likelihood probability to choose candidates has been set at 0.5, and the most ideal number of candidates has been fixed upon three. In this way, our system can lower the FAR while maintaining the TAR. To re-represent the data distribution of candidates, the LDA has been applied and the re-representation step can further facilitate lowering the FAR while maintaining the TAR. Additionally, the two-stage identification system can correctly detect the intruders at 66.7% accuracy rates.

Amongst all features we applied, the most stable one is the PSD, which is shown by the classification result and the ratio of the CV of a single day to the CV of the three days. With the most stable feature, the longitudinal adaptation contains an unsupervised method (adaptive system) and a supervised method (incremental learning.) The adaptive system using PCA in the adaptation step can very much improve the accuracy of the EEG-based identification system by 5%. The performance of the adaptive system using DA in the adaptation step has not much improvement. However, when the time interval between the training and testing sets were longer than one year, the best longitudinal adaptation was the incremental learning. By training the decision model with EEG recordings acquired on several days, the identification system with incremental learning reached 60% correct classification rate. Overall, the EEG-based identification system using incremental learning as a longitudinal adaptation can improve much more on stability than that using the adaptive system. However, the unsupervised longitudinal adaptation (adaptive system) is more practical for the system users since they do not have to provide another group of EEG recordings to rebuild the training database before entering the system.

Bibliography

[1] C. Ashby, A. Bhatia, F. Tenore, and J. Vogelstein. Low-cost electroencephalogram (eeg) based authentication. *5th International IEEE/EMBS Conference on Neural Engineering (NER), 2011*, pages 442–445.

[2] C.C. Chang and C.J. Lin. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3):1–27, 2011.

[3] A. Delorme and S. Makeig. Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1):9–21, 2004.

[4] A. Ferreira, C. Almeida, P. Georgieva, A. Tome, and F. Silva. Advances in EEG-based biometry. In *7th International Conference on Image Analysis and Recognition*, pages 287–295, Povoa de Varzim, Portugal, June 2010.

[5] R. Gopalan, Li Ruonan, and R. Chellappa. Domain adaptation for object recognition: An unsupervised approach. In *IEEE International Conference on Computer Vision (ICCV), 2011*, pages 999–1006.

[6] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441, 1933.

[7] B. Hu, Q. Liu, Q. Zhao, Y. Qi, and H. Peng. A real-time electroencephalogram (eeg) based individual identification interface for mobile security in ubiquitous environment. In *IEEE Asia-Pacific Services Computing Conference (APSCC), 2011*, pages 436–441. IEEE.

[8] J.F. Hu. New biometric approach based on motor imagery EEG signals. In *2009 International Conference on Future Biomedical Information Engineering*, pages 94–97, Sanya, China, December 2009.

[9] A.K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, 14(1):4–20, 2004.

[10] T. Karthikeyan and B. Sabarigiri. Enhancement of multi-modal biometric authentication based on iris and brain neuro image coding. *International Journal of Biometrics and Bioinformatics (IJBB)*, 5(5):249, 2011.

[11] K. Malinka, P. Hanacek, and M. Trzos. Evaluation of biometric authentication based on visual evoked potentials. In *2011 International Carnahan Conference on Security Technology (ICCST)*, Barcelona, Spain, October 2011.

[12] S. Marcel and J.R. Millan. Person authentication using brainwaves (eeg) and maximum a posteriori model adaptation. *IEEE transactions on pattern analysis and machine intelligence*, pages 743–752, 2007.

[13] C. Miyamoto, S. Baba, and I. Nakanishi. Biometric person authentication using new spectral features of electroencephalogram (eeg). In *International Symposium on Intelligent Signal Processing and Communications Systems(ISPACS)*, pages 1–4, Bangkok, October 2008.

[14] I. Nakanishi, S. Baba, and C. Miyamoto. EEG based biometric authentication using new spectral features. In *International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pages 651–654, Kanazawa, Japan, January 2009.

[15] R. Palaniappan. Two-stage biometric authentication method using thought activity brain waves. *International journal of neural systems*, 18(1):59, 2008.

[16] R. Palaniappan and D.P. Mandic. Biometrics from brain electrical activity: A machine learning approach. *IEEE transactions on pattern analysis and machine intelligence*, pages 738–742, 2007.

[17] R. Palaniappan and K.V.R. Ravi. Improving visual evoked potential feature classification for person recognition using pca and normalization. *Pattern recognition letters*, 27(7):726–733, 2006.

[18] S. Pankanti, S. Prabhakar, and A.K. Jain. On the individuality of fingerprints. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(8):1010–1025, 2002.

[19] R.B. Paranjape, J. Mahovsky, L. Benedicenti, and Z. Koles. The electroencephalogram as a biometric. In *2001 Canadian Conference on Electrical and Computer Engineering*, volume 2, pages 1363–1366, May 2001.

[20] M. Poulos, M. Rangoussi, and N. Alexandris. Neural network based person identification using EEG features. In *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages 1117–1120, Phoenix, Arizona, March 1999.

[21] M. Poulos, M. Rangoussi, V. Chrissikopoulos, and A. Evangelou. Person identification based on parametric processing of the eeg. In *Proceedings of ICECS '99. The 6th IEEE International Conference on Electronics, Circuits and Systems, 1999.*, volume 1, pages 283–286 vol. 1. IEEE.

[22] A. Riera, A. Soria-Frisch, M. Caparrini, C. Grau, and G. Ruffini. Unobtrusive biometric system based on electroencephalogram analysis. *EURASIP Journal on advances in signal processing*, 2008(11), 2008.

[23] S.L. Sun. Multitask learning for eeg-based biometrics. In *19th International Conference on Pattern Recognition, 2008. ICPR 2008.*, pages 1–4.

[24] J. Vidal. Toward direct brain-computer communication. *Annual review of Biophysics and Bioengineering*, 2(1):157–180, 1973.

[25] J. Vidal. Real-time detection of brain events in eeg. *Proceedings of the IEEE*, 65(5):633–641, 1977.

[26] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T.M. Vaughan. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6):767–791, 2002.

[27] A. Yazdani, A. Roodaki, S.H. Rezatofighi, K. Misaghian, and S.K. Setarehdan. Fisher linear discriminant based person identification using visual evoked potentials. In *9th International Conference on Signal Processing, 2008. ICSP 2008.*, pages 1677–1680.

[28] Q. Zhao, H. Peng, B. Hu, L.L. Li, Y.B. Qi, Q.Y. Liu, and L. Liu. Towards an efficient and accurate EEG data analysis in eeg-based individual identification. In *7th International Conference on Ubiquitous Intelligence and Computing (UIC)*, pages 534–547, October 2010.