การพิสูจน์ตัวตนโดยใช้ช่องสัญญาณคลื่นสมองที่น้อยที่สุด

นายปรีชา ตั้งเกรียงกิจ

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์ ภาควิชาคณิตศาสตร์และวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2554

PERSONAL IDENTIFICATION USING MINIMUM NUMBER OF EEG SIGNALS

Mr.Preecha Tangkraingkij

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Computer Science
Department of Mathematics and Computer Science
Faculty of Science
Chulalongkorn University
Academic Year 2011

Thesis Title	PERSONAL IDENTIFICATION USING MINIMUM NUMBER OF EEG SIGNALS
By	Mr.Preecha Tangkraingkij
Field of Study	Computer Science
Thesis Advisor	Professor Chidchanok Lursinsap, Ph.D.
Thesis Co-advisor	Siripun Sanguansintukul, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctoral Degree

……………………………………………….. Dean of the Faculty of Science

(Professor Supot Hannongbua, Dr.rer.nat.)

THESIS COMMITTEE

……………………………………………….. Chairman

(Suphakant Phimoltares, Ph.D.)

……………………………………………….. Thesis Advisor

(Professor Chidchanok Lursinsap, Ph.D.)

……………………………………………….. Thesis Co-advisor

(Siripun Sanguansintukul, Ph.D.)

……………………………………………….. Examiner

(Assistant Professor Tayard Desudchit, M.D.)

……………………………………………….. External Examiner

(Piyawat Srichaikul, Ph.D.)

……………………………………………….. External Examiner

(Assistant Professor Sukanya Phongsuphap, Ph.D.)

ปรีชา ตั้งเกรียงกิจ : การพิสูจน์ตัวตนโดยใช้ช่องสัญญาณคลื่นสมองที่น้อยที่สุด. (PERSONAL IDENTIFICAION USING MINIMIM NUMBER OF EEG SIGNALS) อ. ที่ปรึกษาวิทยานิพนธ์หลัก : ศ.ดร. ชิดชนก เหลือสินทรัพย์, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม : อ. ดร. สิริพันธุ์ สงวนสินธุกุล, 79 หน้า.

เทคโนโลยีชีวภาพเช่น ลายพิมพ์นิ้วมือ การสแกนม่านตาหรือเยื่อภายในลูกตา การจดจำลักษณะใบหน้าถูกใช้ในการพิสูจน์ตัวตน เทคโนโลยีชีวภาพเกี่ยวกับการรับรู้โดยใช้คลื่นสมองได้กลายเป็นเครื่องมือพิสูจน์ตัวตนที่น่าสนใจ เนื่องจากสมองมีโครงสร้างที่ซับซ้อนที่สุดที่รู้จักกันมาและสัญญาณคลื่นสมองยากที่จะเลียนแบบหรือขโมย ในวิทยานิพนธ์ฉบับนี้สัญญาณคลื่นสมอง (EEG) ถูกใช้เพื่อพิสูจน์ตัวตน เนื่องจากบุคคลแต่ละบุคคลจะมีรูปแบบของคลื่นสมองที่แตกต่างกัน สัญญาณคลื่นสมองสามารถถูกจัดเก็บได้ในหลายตำแหน่งที่แตกต่างกันไป แต่สัญญาณที่มากเกินไปอาจทำให้ลดความเร็วและความถูกต้องในการจดจำลดลง วิธีในทางปฏิบัติเป็นการรวมการวิเคราะห์องค์ประกอบอิสระ (independent component analysis) สำหรับแยกแยะสัญญาณที่เกิดจากการผสมสัญญาณหลายสัญญาณให้ได้สัญญาณเดิมและใช้โครงข่ายประสาทแบบมีการสอน (supervised neural network) เพื่อการแยกแยะบุคคลจึงได้ถูกนำเสนอ จากตำแหน่งที่ถูกจัดเก็บสัญญาณคลื่นสมอง 16 ตำแหน่ง ตำแหน่งที่ถูกคัดเลือกซึ่งมีความสัมพันธ์กัน 4 ตำแหน่งของข้อมูล 1,000 จุดคือ (F_4, C_4, P_4, O_2), 1,500 จุด (F_8, F_3, C_3, P_4) และ 3,000 จุด (F_{p1}, F_4, P_4, O_2) โดยขั้นตอนวิธีของ SOBIRO การเลือกตำแหน่งดังกล่าวถูกใช้พิสูจน์ตัวตนของบุคคลในกลุ่ม 20 คน ได้ด้วยความถูกต้องสูงและยังสามารถแยกแยะบุคคลที่ไม่ได้อยู่ในกลุ่มอีกด้วย ตำแหน่งที่มีความสำคัญสำหรับการพิสูจน์ตัวตนคือตำแหน่ง P_4 ซึ่งตำแหน่งนี้เป็นส่วนของ parietal lobe ของสมอง

ภาควิชา คณิตศาสตร์และวิทยาการคอมพิวเตอร์	ลายมือชื่อนิสิต..............................
สาขาวิชา วิทยาการคอมพิวเตอร์	ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก.......
ปีการศึกษา 2554	ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม.......

5073842523 : MAJOR COMPUTER SCIENCE

KEYWORDS : IDENTIFICATION / INDEPENDENT COMPONENT ANALYSIS / PATTERN RECOGNITION / NEURAL NETWORK / ELECTROENCEPHALOGRAM

PREECHA TANGKRAINGKIJ : PERSONAL IDENTIFICATION USING MINIMUM NUMBER OF EEG SIGNALS. ADVISOR : PROF. CHIDCHANOK LURSINSAP, Ph.D., CO-ADVISOR : SIRIPUN SANGUANSINTUKUL,Ph.D., 79 pp.

Biometrics such as fingerprints, retinal or iris scanning and face recognition are actively used for identifications. Cognitive biometrics using brain signals have become interesting identification tools because the brain is the most complex biological structure known and its wave signals are very difficult to mimic or steal. In this dissertation, EEG signals are used to identify a person as different persons have different EEG patterns. EEG signals can be measured from different locations. However, many signals can degrade recognition speed and accuracy. A practical technique combining independent component analysis (ICA) for signal cleaning and a supervised neural network for person identification is proposed. From 16 different EEG signal locations, four truly relevant locations of 1,000 data points (F_4, C_4, P_4, O_2), 1,500 data points (F_8, F_3, C_3, P_4), and 3,000 data points (F_{p1}, F_4, P_4, O_2) by SOBIRO algorithm were selected. This selection was used to identify a group of 20 persons with high accuracy and can separate the persons who are not in the group. The significant location for identification is position P_4 which is the parietal lobe of the brain.

Department : Mathematics and Computer Science Student's Signature ....

Field of Study : Computer Science Advisor's Signature ....

Academic Year : 2011 Co-advisor's Signature ....

Acknowledgements

I am sincerely and heartily grateful to my supervisor, Professor Dr. Chidchanok Lursinsap, whose encouragement, supervision and support from the preliminary to the concluding level enabled me to develop an understanding of the subject. His overt enthusiasm and integral view on research have made a deep impression on me. He was the most enthusiastic and generous research supervisor, without his contributions and guidance this would not have been possible.

I am grateful to my thesis co-advisor, Dr. Siripun Sanguansintukul for providing additional comments that help improve this dissertation and introducing me the art of research. I am deeply grateful her generosity. I would like to thank Assistant Professor Tayard Desudchit for providing all data. His valuable suggestions and comments made this work feasible. I also express my thankfulness to my dissertation committee: Dr. Suphakant phimaltares, Dr. Piyawut Srichaikul and Assistant Professor Sukanya Phongsuphap.

Associate Professor Suchada Siripant is an admired person. I would like to thank for her suggestions in the Advanced Virtual and Intelligent Computing Center (AVIC) through all my Ph.D. student life. Also, it is my pleasure to acknowledge the financial from the Centre of Excellence in Mathematics, the Commission on Higher Education, Thailand and the 90th Anniversary of Chulalongkorn University Fund (Ratchadaphiseksomphot Endowment Fund).

Finally, I am deeply grateful to my beloved parents and my family for their everlasting supports and encouragements throughout my study.

Contents

	Page
Abstract (Thai)……	iv
Abstract (English)……	v
Acknowledgements……	vi
Contents……	vii
List of Tables……	ix
List of Figures……	xii
List of Abbreviations……	xiii
List of Symbols……	xv
Chapter	
1 INTRODUCTION……	1
1.1 Problem and Motivation……	1
1.2 Objective……	2
1.3 Scope of Work……	2
1.4 Contribution……	3
1.5 Research Methodology……	4
1.6 Dissertation Organization……	4
2 BACKGROUND AND LITERATURE REVIEWS……	5
2.1 Background……	5
2.1.1 Biometrics……	5
2.1.2 Electroencephalography (EEG) ……	10
2.2 Literature Reviews……	11
2.2.1 Independent Component Analysis (ICA)……	11
2.2.2 Neural Classification Concept……	14
3 METHODOLOGY OF EEG IDENTIFICATION PROCESS……	16
3.1 Collecting EEG Signals……	17
3.2 Automatically Allocating the Significant Range of EEG Signals……	20
3.3 Selecting Approriate ICA Algorithms……	22

Chapter | Page

3.3.1 Experimental Results.. 23

3.4 Selecting Minimum Relevant Channels.. 25

3.4.1 Experimental Results.. 25

3.5 Confirming of the 4-Channel Combinations... 28

3.5.1 Determining Outsider Subject.. 32

3.5.2 Experimental Results.. 33

3.5.3 Compare 4 Channel with 5, 6, 7 Channel Combinations............... 45

3.5.4 Experimental Results.. 45

3.6 Exploring Conditions Perform the Best for Identification............................ 51

3.6.1 Experimental Results.. 52

3.7 Explaining Biological Significance of the Selected Channels..................... 53

4 CONCLUSION AND FUTURE WORK... 55

References.. 56

Appendices.. 60

Appendix A List of Publications... 61

Biography.. 62

List of Tables

Table Page

3.1 10-20 system EEG locations and their corresponding channel names.. 18

3.2 The accuracy percentage of each ICA algorithm when applied to 5,10 and 20 subjects…………………………………………………………………. 24

3.3 The percentage of accuracy for one channel when applied to 5, 10 and 20 subjects with ERICA algorithm…………………………………….. 26

3.4 The percentage of accuracy for 2 channels when applied to 5, 10 and 20 subjects with ERICA algorithm………………………………………….. 27

3.5 The percentage of accuracy for 3 channels when applied to 5, 10 and 20 subjects with ERICA algorithm………………………………………….. 27

3.6 The percentage of accuracy for 4 channels when applied to 5, 10 and 20 subjects with ERICA algorithm………………………………………….. 28

3.7 Percentage of accuracy for 4-channel combinations, 500 data points with different ICA algorithms……………………………………………... 29

3.8 The percentage of accuracy for 1-20 subjects (insider) when tested with the best 4-channel combinations, ERICA algorithm, and 500 data points in experiment 1…………………………………………………….. 34

3.9 The percentage of accuracy for 21-40 subjects (outsider) when tested with the best 4-channel combinations, ERICA algorithm, and 500 data points in experiment 1…………………………………………………….. 35

3.10 Selected 4-channel combination having percentage of accuracy higher than 95% with SOBIRO algorithm, and 1,000 data points in experiment 2……………………………………………………………….. 37

3.11 The percentage of accuracy for 40 subjects tested with the best 4-channel combination (ch13, ch14, ch15, ch16), SOBIRO algorithm, and 1,000 data points in experiment 2……………………………………. 38

Table Page

3.12 Selected 4-channel combination having percentage of accuracy higher than 95% with SOBIRO algorithm, and 1,500 data points in experiment 3……………………………………………………………….. 40

3.13 The percentage of accuracy for 40 subjects tested with the best 4-channel combination (ch6, ch9, ch10, ch15), SOBIRO algorithm, and 1,500 data points in experiment 3…………………………………………. 41

3.14 Selected 4-channel combination having percentage of accuracy higher than 95% with SOBIRO algorithm, and 3,000 data points in experiment 4……………………………………………………………….. 43

3.15 The percentage of accuracy for 40 subjects tested with the best 4-channel combination (ch1, ch13, ch15, ch16), SOBIRO algorithm, and 3,000 data points in experiment 4…………………………………………. 44

3.16 The percentage of accuracy for 40 subjects tested with the best 5-channel combination (ch5, ch6, ch9, ch10, ch15), SOBIRO algorithm, and 3,000 data points……………………………………………………….. 47

3.17 The percentage of accuracy for 40 subjects tested with the best 6-channel combination (ch1, ch5, ch6, ch9, ch10, ch15), SOBIRO algorithm, and 3,000 data points…………………………………………... 48

3.18 The percentage of accuracy for 40 subjects tested with the best 7-channel combination (ch5, ch6, ch7, ch9, ch13, ch14, ch15), SOBIRO algorithm, and 3,000 data points…………………………………………... 49

3.19 The difference between the accuracy percentage of inside and outsider subjects for 4, 5, 6, 7 channels experimented by SOBIRO algorithm and 3,000 data points…………………………………………. 50

Table Page

3.20 The percentage of accuracy of flash light, calculation of math, and impressive image compare with relax condition………………………….. 52

3.21 The number of occurrences of each channel…………………………….. 54

List of Figures

Figure		Page
2.1	A typical EEG biometric system for personal identification consisting of four main modules…………………………………………………………	9
2.2	Simplified model of brain sources and EEG signals……………………...	12
2.3	The general model of ICA and extraction of sources…………………….	13
2.4	One layer network with m inputs and k neurons………………………….	14
3.1	The locations of electrode placements on the scalp using 10-20 system and the corresponding channel numbers (a) Locations of electrodes (b) Channel numbers of the electrodes………………………	17
3.2	The data points of observed EEG signals and source EEG signals after ICA process (a) Some observed EEG signals from channels ch2, ch10, ch11, and ch12 of a subject (b) The EEG signals after being processed by ICA algorithm (ERICA)………………………………………	19
3.3	Combine 15 windows to the group of 3,000 data points…………………	21
3.4	Format of data partitioning to form the training, validating, and testing groups……………………………………………………………………….	23
3.5	The EEG signals 3,000 data points are divided into six groups for testing………………………………………………………………………..	30
3.6	The EEG signals 3,000 data points are divided into three groups for testing………………………………………………………………………..	31
3.7	The EEG signals 3,000 data points are divided into two groups for testing………………………………………………………………………..	31
3.8	The EEG signals 3,000 data points, all for testing………………………..	32
3.9	Compare the accuracy percentage of insider and the different between the accuracy percentage with the outsider wrongly identified to insider for 4, 5, 6 and 7 channel combination………………………….	50

List of Abbreviations

EEG	Electroencephalography
MEG	Magnetoencephalography
fMRI	Function Magnetic Resonance Imaging
PET	Positron Emission Tomography
BOLD	Blood Oxygen Level Dependency
AR	Autoregressive
FFT	Fast Fourier Transform features
LVQ	Learning Vector Quantizer
CG	Computational Geometry
VEP	Visual Evoked Potential
ICA	Independent Component Analysis
CCEP	The Chulalongkorn Comprehensive Epilepsy Program
DNA	Deoxyribonucleic Acid
EMG	Electromyography
BCI	Brain Computer Interface
BBS	Blind Source Separation
AMUSE	Algorithm for Multiple Unknown Source Extraction is Based on the EVD
ERICA	Equivariant Robust Independent Component Analysis Algorithm
EVD2	SOS BSS Algorithm Based on Symmetric EVD
EWASOBI	Eficient Weights Adjusted Second Order Blind Identification
FAJDC4	Fast Approximate Diagonalization of Cummulant Matrices
FJADE	Flexible Joint Approximate Diagonalization of Quadricovariance Matrices
FOBI-E	Fourth Order Blind Identification with Transformation matrix E
JADEop	Robust Joint Approximate Diagonalization of Eigen matrices (with optimized numerical procedures)

JADETD	HOS Joint Approximate Diagonalization of Eigen matrices with Time Delays
MULCOMBI	Multi-Combination of WASOBI and EFICA
POWERICA	Power iteration for ICA
QJADE	Quadratic Join Approximate Diagonalization of Cummulant Matrices
SAD	Sequential Approximate Diagonalization
SIMBEC	Simultaneous Blind Signal Extraction Using Cumulants
SOBI	Second Order Blind Identification
SOBI-BPF	Robust SOBI with Bank of Band-Pass Filters
SOBIRO	Robust Second Order Blind Identification with Robust Orthogonalization
SONS	Second Order Nonstationary Source Separation
SYMMETRIC	Symmetric prewhitening
THINICA	Thin algorithm for Independent Component Analysis
UNICA	Unbiased Quasi Newton Algorithm for ICA
WASOBI	Weights Adjusted Second Order Blind Identification
MLP	Multilayer Perceptron
EDF	European Data Format

List of Symbols

s	noiseless input vector of EEG signals
o	observed vector of EEG signals
M	constant mixing matrix
v	unknown noise vector
y	estimated signal vector
D	de-mixing matrix
X_m	input signals of node *m*
$W_{k,m}$	weight of the neuron *k*
n_k	linear combination output of node *k* due to input signals
b_k	bias of node *k*
f( )	activation function
a_k	output signal of node *k*
f^{μ}	the class of vector x^{μ}
x^{μ}	feature vector of pattern μ
$\varnothing_i(.)$	the kernel function
w_i	weight of link *i*
$g(.)$	activation function of the output neuron
F_{P1}	frontal polar 1 location
F_7	frontal 7 location
T_3	temporal 3 location
T_5	temporal 5 location
F_{P2}	frontal polar 2 location
F_8	frontal 8 location
T_4	temporal 4 location

T_6	temporal 6 location
F_3	frontal 3 location
C_3	central 3 location
P_3	parietal 3 location
O_1	occipital 1 location
F_4	frontal 4 location
C_4	central 4 location
P_4	parietal 4 location
O_2	occipital 2 location
$G_1^{(500)}$	the first group of 500 data points
$G_2^{(500)}$	the second group of 500 data points
$G_3^{(500)}$	the third group of 500 data points
$G_4^{(500)}$	the fourth group of 500 data points
$G_5^{(500)}$	the fifth group of 500 data points
$G_6^{(500)}$	the sixth group of 500 data points
$G_1^{(1000)}$	the first group of 1000 data points
$G_2^{(1000)}$	the second group of 1000 data points
$G_3^{(1000)}$	the third group of 1000 data points
$G_1^{(1500)}$	the first group of 1500 data points
$G_2^{(1500)}$	the second group of 1500 data points
$G_1^{(3000)}$	the first group of 3000 data points
$G_i^{(m)}$	the set of *m* data points of signal group *i*
$a(j, G_i^{(m)})$	the percentage of accuracy of subject *j* tested by the *m* data points in signal group *i*
$\min_{\forall i} a(j, G_i^{(m)})$	the minimum percentage of accuracy of subject *j*
$\max_{\forall i} a(k, G_i^{(m)})$	the maximum percentage of accuracy of subject *k*

k	the subject is determined as an outsider
$\hat{k}$	the subject which the neural network determines

CHAPTER I

INTRODUCTION

1.1 Problem and Motivation

Biometrics information such as fingerprints, retinal or iris scanning, face recognition are actively used for identifications [1]. Cognitive biometrics using brain signals have become interesting as identification tools. The brain is the most complex biological structure known to man and its wave signals are very difficult to mimic or steal. To understand how the brain functions, many techniques such as electroencephalography (EEG), magnetoencephalography (MEG), function magnetic resonance imaging (fMRI), and positron emission tomography (PET) have been utilized. Each technique has its own strengths and weaknesses.

Researchers have used EEG to analyze patterns and images of the human brain because EEG has a desirable property for excellent time resolution and low cost of instrumentation. MEG has the same temporal resolution as EEG but has a much better spatial resolution. Nevertheless, MEG requires sophisticated devices that can be operated only in special facilities. fMRI measures changes of blood flow in the brain. It is an indirect method for measuring neural activity based on BOLD (Blood Oxygen Level Dependency). The changes of blood flow that occur in capillary beds in the specific regions of the brain are thought to represent various neuronal activities. PET is able to monitor glucose and oxygen metabolism as well as neurotransmitter activity in different areas within the brain. This can correlate with the level of activities in the particular region in the brain. Both fMRI and PET are limited to temporal resolutions and require sophisticated devices.

It has been shown in previous studies that EEG is unique and can be used for biometric identification and authentication. Paranjape [2] used an autoregressive (AR) model and discriminant function analysis to identify the EEG signals. Poulos [3,4] applied Fast Fourier Transform (FFT) and AR model for features extraction. Their classification was based on the Learning Vector Quantizer (LVQ) and Computational Geometry (CG) approaches. Palaniappan [5-8] examined the Visual Evoked Potential

(VEP) as an improved method for employing EEG biometric features. Marcel [9] proposed use of statistical framework based on a Gaussian mixture and maximum a-posteriori models for personal authentication.

In this dissertation, EEG signals were used to identify a person as different persons have different EEG patterns. EEG signals can be measured from different locations. Too many signals can degrade the recognition speed and accuracy. A practical technique combining Independent Component Analysis (ICA) for signal cleaning and a supervised neural network for classifying signals was proposed.

1.2 Objective

The objectives of this work strive to acquire locations (channels) on the scalp that are the most promising for personal identification. Furthermore, the minimum number of channels necessary for the identifications will also be explored. The all EEG signals are cleansed by Independence Component Analysis technique. The clean EEG signals of each person are identified by a neural network.

1.3 Scope of work

In this dissertation, the scope of work is constrained as follows:

1. The EEG data were collected from the normal patients at Chulalongkorn Comprehensive Epilepsy Program (CCEP) under the Patronage of Professor Dr.Her Royal Highness Princess Chulabhorn.
2. The subjects were resting with their eyes open and 16 electrodes were placed on their scalp.
3. Independence component analysis technique was used for signal cleaning.

4. A supervised neural network technique was performed on pattern recognition task for identifications.

1.4 Contribution

This dissertation proposed a new method for biometric identification by brain wave signals (EEG). The contributions of this research are as follows:

1. The personal identification by EEG has more advantage than other biometric personal identifications because EEG signals are difficult to mimic or steal.

2. The technique combines outstanding features of both Independent Component Analysis (ICA) and supervised neural network for personal identification.

3. The methodology can correctly determine the subjects who are in the group with high accuracy. In addition, it can separate subjects who are outside the group.

1.5 Research Methodology

1. Reviewing and study the research papers that were related to the EEG use for identification.

2. Developing a new method for identification by technique combining Independent Component Analysis (ICA) for signal cleaning and a supervised neural network for identification EEG signals.

3. Collecting EEG signals.

4. Automatically allocating the significant range of EEG signals.

5. Selecting appropriate ICA algorithms.

6. Selecting minimum relevant channels.

7. Confirming of the 4-Channel combinations.

8. Exploring conditions perform the best for identification.

9. Explaining biological significance of the selected channels.

1.6 Dissertation Organization

The remainder of the dissertation is organized as follows. Chapter 2 introduces background and literatures reviews. The methodologies of EEG identification process and their experiment results are showed in chapter 3. Chapter 4 concludes the dissertation.

CHAPTER II

BACKGROUND AND LITERRATURE REVIEWS

In this chapter, the background on the biometrics for personal identification and Electroencephalography are described. Literatures related to prototype methods, Independent Component Analysis (ICA) and Neural Classification Concept are also reviewed.

2.1 Background

In this section, the background of biometrics and electroencephalography are reviewed.

2.1.1 Biometrics

The biometrics systems have been used by humans for thousands of years to recognize each other. The first evidence that showed humans have used biometric as early as pre-historical age is the evidences found in the caves estimated 31,000 years old. The caves were adorned with pre-historical pictures signed with fingerprint stamps of the authors. The other evidence was found in China in the 14th century. The explorer Joao de Barros reported the Chinese merchants were stamping children's palm prints and footprints on paper with ink to distinguish the young children from one another. This is one of the earliest known cases of biometrics in the use and is still being used today.

The term "biometrics" is derived from the Greek words “bio” means life and “metric” means to measure. Biometrics became an interesting topic now in regarding to computer and network security. The first real biometric system was created by French anthropologist Alphonse Bertillion in 1870. He developed an identification system called Bertillonage, a method of bodily measurement which got named after him. The bertillonage is based on detailed records of body measurement, physical description and photographs. Despite their imprecise measures and difficulties in apply methodology, the Bertillonage was an important advance on criminal and people identification. His system was used by police authorities throughout the world, until it began to fail when it was discovered that many people shared the same anthropological measures.

Biometrics characteristics can be divided into three main classes. The first physiological biometrics relate to the shape of the different parts of body [1], [30] such as:

- DNA: Deoxyribonucleic Acid (DNA) is the one-dimensional ultimate unique code for each person. The exception is identical twins. This biometric has some drawbacks: 1) contamination and sensitivity, 2) no real-time application is possible, and 3) privacy issues.
- Ear shape: The shape of the ear and the structure of the cartilaginous tissue of the pinna are distinctive. Matching the distance of salient points on the pinna from a landmark location of the ear is the suggested method of recognition in this case.
- Face recognition: Facial images are the most common biometric characteristic used by humans to make a personal recognition. The most popular approaches to face recognition are based on the location and shape of facial attributes such as eyes, eyebrows, nose, lips and chin, and their spatial relationships, or the overall analysis of the face image that represents a face as a weighted combination of a number of canonical faces. Facial recognition system should be able to automatically detect a face in an image, extract its features and then recognize it from a general viewpoint (i.e., from any pose) which is a rather difficult task. Another problem of face is a changeable social organ displaying a variety of expressions.
- Fingerprints: A fingerprint is the pattern of ridges and valleys located on the tip of each finger. Fingerprints have been used for personal identification for many centuries. Patterns have been extracted by creating an inked impression of the fingertip on paper. The matching process involves comparing the two-dimensional minutiae patterns extracted from the user's print with those in the template. The problem with the fingerprint recognition

systems are that they require a large amount of computational resources, including the cut and bruises on the fingerprints.

- Hand and finger geometry: The hand geometry recognition is based on a number of measurements the lengths and width of the fingers and the location of joints, shape and size of palm. Hand geometry information may not be invariant during the growth period of children. There is even verification systems available that are based on measurements of only a few fingers instead of the entire hand because of these devices are smaller than those used for hand geometry.
- Infrared thermogram: (facial, hand or hand vein). It is possible to capture the pattern of heat radiated by the human body with an infrared camera. That pattern is considered to be unique for each person but image acquisition is rather difficult where there are other heat emanating surfaces near the body. Infrared sensors are expensive which is a factor inhibiting wide spread use of the thermograms.
- Iris recognition: Its complex pattern can contain many distinctive features such as arching ligaments, furrows, ridges, crypts, rings, corona, freckles and a zigzag collarette. Iris scanning is less intrusive than retinal because the iris is easily visible from several meters away. Responses of the iris to changes in light can provide an important secondary verification that the iris presented belongs to a live subject. Irises of identical twins are different.
- Odor: Each object spreads around an odor that is characteristic of its chemical composition and this could be used for distinguishing various objects. This would be done with an array of chemical sensors, each sensitive to a certain group of compounds. A component of the odor emitted by a human body is distinctive to a particular individual but deodorants and perfumes could lower the distinctiveness.
- Palm print: Like fingerprints, palms of the human hands contain unique pattern of ridges and valleys. Although palm is larger than a finger, palm

print is expected to be even more reliable than fingerprint. Palm print scanners need to capture larger area with similar quality as fingerprint scanners, so they are more expensive.

- Retina: Retina scan is rich in structure that supposed to be a characteristic of each eye and each individual. Since it is protected in an eye itself, and it is not easy to change or replicate the retinal vasculature. It is claimed to be one of the most secure biometric.

The second is behavioural biometrics. This type of biometrics relates to the behaviour of a human. Some examples of this type of biometrics are:

- Gait: Gait is the peculiar way one walks and it is a complex spatio-temporal biometrics. It is not supposed to be very distinctive but can be used in some low-security applications. Gait is a behavioral biometric and may not remain the same over a long period of time, due to change in body weight or serious brain damage. Since video-sequence is used to measure several different movements this method is computationally expensive.
- Signature: The way a person signs his or her name is known to be characteristic of that individual. Signatures are a behavioral biometric that change over a period of time and are influenced by physical and emotional conditions of a subject. In addition to the general shape of the signed name, a signature recognition system can also measure pressure and velocity of the point of the stylus across the sensor pad.
- Typing rhythm (Keystroke): Keystroke is the action of tracking (or logging) the keys struck on a keyboard. It is believed that each person types on a keyboard in a characteristic way. This is also not very distinctive but it offers sufficient discriminatory information to permit identity verification.

The third biometrics is a combination of physiological and behavioural biometrics.

- Voice: The features of an individual's voice are based on physical characteristics such as vocal tracts, mouth, nasal cavities and lips that are used in creating a sound. These characteristics of human speech are invariant for an individual, but the behavioral part changes over time due to age, medical conditions and emotional state. The pattern matching algorithms used in voice recognition are similar to those used in face recognition. A disadvantage of voice-based recognition is that speech features are sensitive to a number of factors such as background noise.

Biometric systems provide two different functions.

1. Authentication (or verification). This system validates a person's identity by comparing the captured biometric data with his/her own biometric stored in the system database.
2. Identification. This system recognizes an individual by searching data of all users in the system database for a match.

A biometric system is a pattern recognition system with two stages of operation, the first stage is **enrollment** and the second stage is **recognition** as shown in Figure 2.1. Enrollment is the stage in which some biometric reference information about the person is stored in the system database. The second stage is recognition. During this stage, the system scans the user's biometric trait, extracts features and performs a matching process against the reference biometric information stored in the system database.

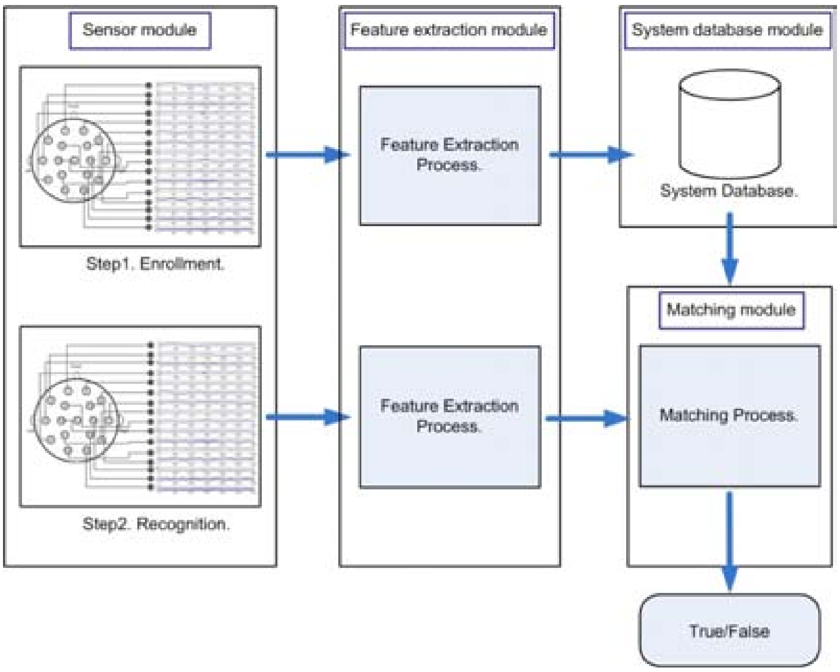


Figure 2.1: A typical EEG biometric system for personal identification consisting of four main modules.

In this work, we investigate the EEG as biometrics for a personal identification system. Figure 2.1 depicts a diagrammatic block visualization of EEG personal identification system used in the course of the works. This depicted system consists of the following four main modules.

1. Sensor module: This module records the EEG signals of an individual at the enrollment and the recognition stages.
2. Feature extraction module: The EEG signals are cleaned and their features are extracted. A neural network is deployed for recognition these signals. The recognition accuracy of each individual subject is registered for further authenticating subjects who are in the system database (insider) and subjects who are not in the system database (outsider).
3. Matching module: This module determines whether the measured EEG signal belongs to the signal already collected in the database. If it matches any signal then the individual subject is identified as an insider. Otherwise, that one is identified as an outsider.
4. System database module: This module stores the EEG signals of all enrolled individuals into the biometric system database.

2.1.2 Electroencephalography (EEG)

EEG is the measurement of electrical activity produced by the brain as recorded from electrodes placed on the scalp. The strength of each signal is considered rather low and the signal measured from any location of scalp can be interfered by signals from other locations due to the activities of the brain such as eye tracking and EMG (electromyography). In addition, the noise in EEG may be created by the surrounding large electrical potentials from the environment. Brain waves are categorized into five basic groups according to their frequencies as follows: 1) Delta (1-4 Hz), 2) Theta (4-8 Hz), 3) Alpha (8-12 Hz), 4) Beta (12-30 Hz), and 5) Gamma (30-50 Hz). Although none of these waves are ever emitted alone, the state of consciousness of the individual may result in one frequency being more pronounced than others.

Brain waves were first recorded in 1874 by Richard Caton, who connected equipment directly to the cerebral cortex of a rabbit. In 1929, Hans Berger published the first information on scalp-recorded brain waves in humans. The differential input was first amplified by B. H. C. Matthews in 1934 and revolutionized the high-gain amplification of biologic electrical signals, including brain waves. In 1935 Frederic Gibbs, Hallowell Davis, and William Lennox, published the first EEG paper in English dealing with epilepsy in humans. EEG is now extensively used in diagnosing epilepsy and in the study of how the brain functions in both animals and humans.

EEG is the most valuable diagnosis of epilepsy. It is also used to help predict a person's chance of recovery after a change in consciousness. The most advanced form of EEG usages is applied in the basic Brain Computer Interface (BCI), neuroscience research and statistical signal processing.

2.2 Literature Reviews

This section reviews the independent component analysis and the neural classification concept.

2.2.1 Independent Component Analysis (ICA)

The EEG signal is intrinsically a mixture of the other signals, such effects as delays, reverberations, non-linear distortions [27]. It is assumed that the EEG signals from the electrodes on the scalp picks up brain sources and non-brain sources related to movements of eyes and muscles. The objective of ICA is to clean and separate the individual signals from different areas of the brain.

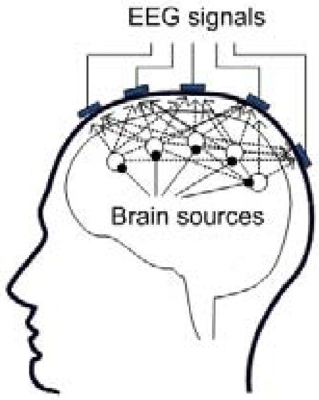


Figure 2.2: Simplified model of brain sources and EEG signals.

In this respect, each signal can be viewed as a vector. Typically, when concerning a vector space, it is assumed that the bases of the space are orthonormal. The accuracy of this assumption is not confirmed since the actual structure of the space is unknown. To cope with these unknown factors, all bases are assumed to be independent and non-orthogonal since any orthogonal space is just a special case of non-orthogonal space. We considered this non-orthogonal space as the natural space of the vector distribution. The separation of these mixed signals is based on this natural space. A signal separation technique called independent component analysis (ICA) was deployed to decompose multivariate data into a linear sum of non-orthogonal vectors with basic coefficients being statistically independent [10]. Let $\mathbf{s} = [s_1 s_2 \ldots s_n]^T$ be a noiseless input vector of n EEG signals at current time t. These n signals are mixed together by using a constant mixing matrix **M**. Let $\mathbf{o} = [o_1 o_2 \ldots o_n]^T$ be the observed vector of n EEG signals obtained from mixing matrix **M** and **s** by the following computation.

$$\mathrm{o} = \mathrm{Ms} + \mathrm{v} \tag{2.1}$$

v is an unknown noise vector. In other words, **o** is the set of EEG signals recorded from the subject's scalp. **s** are the actual signals produced within the brain. Obviously, the values of each $\mathbf{s}_i$ and each element in mixing matrix **M** are unknown. The only known value is the value of each $\mathbf{o}_i$. To derive the values of **s** and mixing matrix **M**, signal vector **o** is multiplied by a de-mixing **D** as follows.

$$y = Do \tag{2.2}$$

The result is an estimated signal vector $\mathbf{y} = [y_1 y_2 ... y_n]^T$ of signal vector **s**. Generally, the accurate value of each element in matrix D is measured by Kullback-Liebler divergence. However, the correctness of this divergence is based on how the probability of each s_i is assumed and how each element in de-mixing matrix D is adjusted.

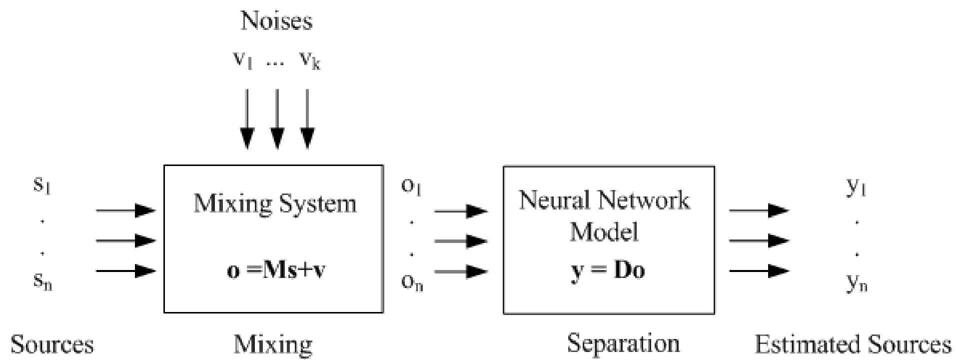


Figure 2.3: The general model of ICA and extraction of sources.

There are many proposed ICA algorithms to estimate the elements in de-mixing matrix **D** [10] and brain signals **s**. In this paper, the following ICA algorithms are considered: AMUSE [10], ERICA [11-12], EVD2 [10], EWASOBI [13-14], FAJDC4 [10], FJADE [10], FOBI-E [10], JADEop [15-16], JADETD [10], MULCOMBI [10], POWERICA [10], QJADE [10], SAD [10], SIMBEC [17-19], SOBI [20-21], SOBI-BPF [22-23], SOBIRO [24], SONS [10], SYMMETRIC [10], THINICA [10], UNICA [10], and WASOBI [13-14].

2.2.2 Neural Classification Concept

Neural network is a process paradigm that mimics the structures and functions of the human nervous system. Pattern recognition is an important application which can be implemented using a feed-forward neural network that has been trained accordingly. During the training, the network learns to associate output with input patterns. When the network is used, it identifies the input pattern and tries to output the associated output pattern similar to the way the human brain works.

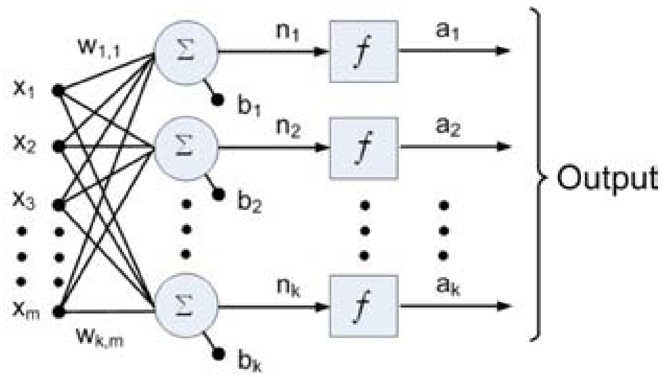


Figure 2.4: One layer network with m inputs and k neurons.

Figure 2.4 shows a neural model for one-layer network with m inputs and k neurons in mathematical terms, we can describe a neuron n_k using the following equations.

$$n_k = \sum_{j=1}^{m} w_{km} x_m \tag{2.3}$$

$$a_k = f(n_k + b_k) \tag{2.4}$$

where

$x_1, x_2, \ldots, x_m$ are input signals.

$w_{1,1}, w_{2,1}, \ldots, w_{k,m}$ are weights of neuron k.

$n_1, n_2, \ldots, n_k$	are activation values.
$b_1, b_2, \ldots, b_k$	are bias.
$f()$	is an activation function.
$a_1, a_2, \ldots, a_k$	are output signals of the neuron.

The separated signals from the ICA process cannot be directly used to identify a person. The relevant features must be extracted from these signals and the problem of identification a person is transformed into a classification problem. There are several classification techniques. However, the features extracted in this study form a feature vector and concern the real vector space with unknown vector distribution. Thus, a classifier based on neural computing is adopted for identification a person since it is rather a powerful technique for managing the unknown vector distribution of the space.

CHAPTER III

METHODOLOGY OF EEG IDENTIFICATION PROCESS

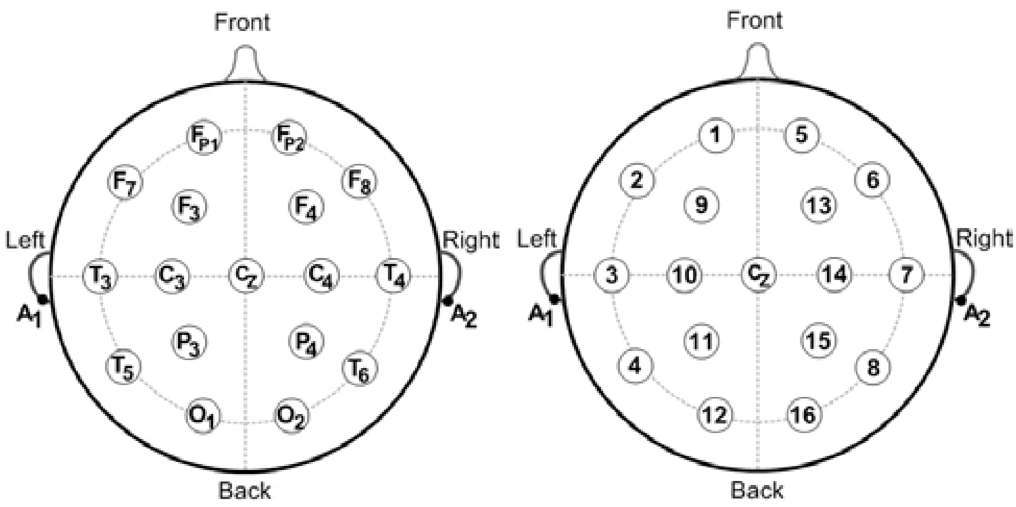
A major factor surrounding the viability of this dissertation is the detail of the identification process. Accuracy of the results is based on the group of subjects in the experiments, however there is no standard benchmark data set for comparison. The identification process in this dissertation consists of the following 7 main procedures:

1. Collecting EEG signals
2. Automatically allocating the significant range of EEG signals
3. Selecting appropriate ICA algorithms
4. Selecting minimum relevant channels
5. Confirming of the 4-Channel combinations
6. Exploring conditions perform the best for identification
7. Explaining biological significance of the selected channels

The details of all procedures are given in the next sections.

3.1 Collecting EEG Signals

EEG signals were collected from 40 normal subjects (18 men and 22 women) from Chulalongkorn Comprehensive Epilepsy Program (CCEP) at Chulalongkorn Hospital in Bangkok. The age range of the subjects is between 10 and 40 years. The EEG signals were recorded while subjects were resting with their eyes open from 16 electrode channels attached to their scalp. According to 10-20 system, the following locations on the scalp are considered: F_{P1}, F_7, T_3, T_5, F_{P2}, F_8, T_4, T_6, F_3, C_3, P_3, O_1, F_4, C_4, P_4, O_2. Figure 3.1(a) shows these locations and Figure 3.1(b) denotes the corresponding channel numbers on the scalp. The pairs of signal location and its channel name are summarized as follows.



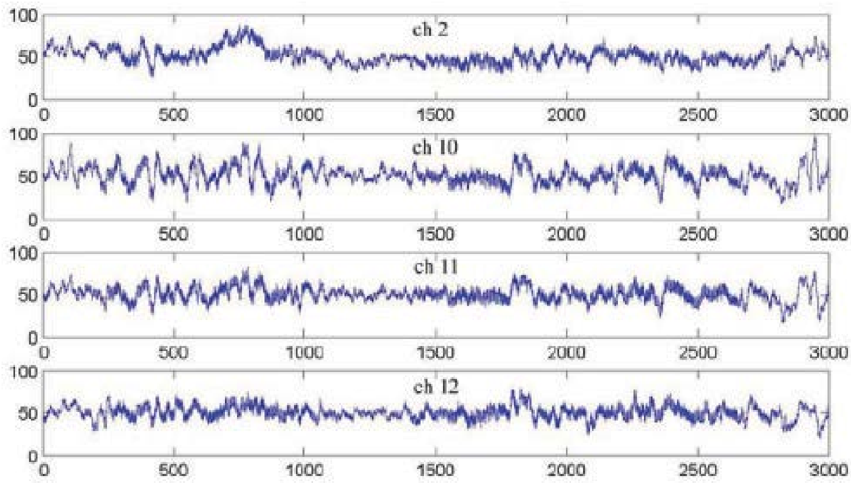
(a) (b)

Figure 3.1: The locations of electrode placements on the scalp using 10-20 system and the corresponding channel numbers. (a) Locations of electrodes. (b) Channel numbers of the electrodes.

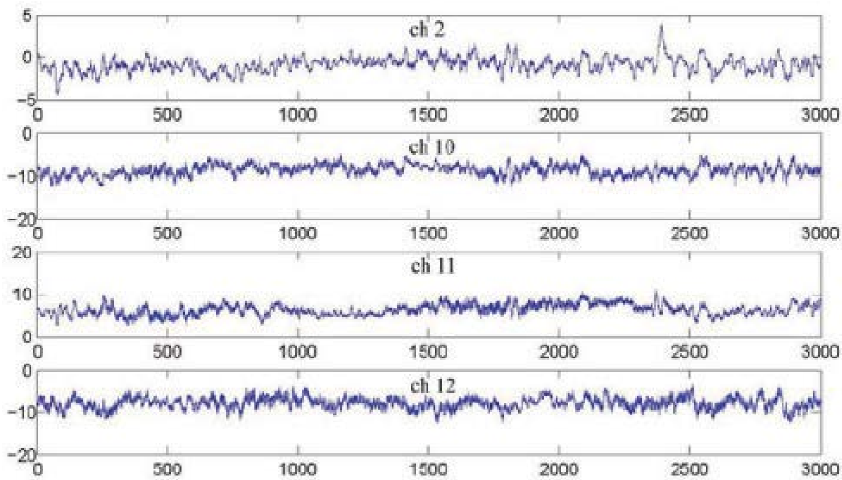
Table 3.1: 10-20 system EEG locations and their corresponding channel names.

Locations	Channel Names
F_{p1}	ch 1
F_7	ch 2
T_3	ch 3
T_5	ch 4
F_{P2}	ch 5
F_8	ch 6
T_4	ch 7
T_6	ch 8
F_3	ch 9
C_3	ch10
P_3	ch11
O_1	ch12
F_4	ch13
C_4	ch14
P_4	ch15
O_2	ch16

The recording sessions used mono-polar montage with reference at the mastoid area A_1 and A_2 as shown in Figure 3.1 The EEG amplifier was Grass model 8 plus. The sampling rate was 200 Hz. EEG data were notch filtered at 60 Hz. The data were digitized by BMSI board using Stellate Harmony EEG software and exported as EDF (European Data Format). For each subject, in each trial, 3,000 data points (15 second recording) were simultaneously collected from each of 16 channels. An example of the raw EEG signals obtained from channels 2, 10, 11, and 12 of a subject is shown in Figure 3.2(a).



(a)



(b)

Figure 3.2: The data points of observed EEG signals and source EEG signals after ICA process. (a) Some observed EEG signals from channels ch2, ch10, ch11, and ch12 of a subject. (b) The EEG signals after being processed by ICA algorithm (ERICA).

3.2 Automatically Allocating the Significant Range of EEG Signals

The Electromyography (EMG) signals included in the EEG signals were collected in the previous step and were removed from the EEG signals since they were irrelevant. The EEG signals are small when compared with EMG. This point is used to find the significant range of EEG signals.

Automatically allocating the significant range of EEG signals started by dividing raw EEG signals into groups of 3,000 data points. The next step found the average value in every group. The smallest average value group was selected as the significant range of EEG signals. The algorithm for automatically selecting the EEG signals is as follows:

1. Divide the raw EEG signals into 200 data points windows.
2. Slide and combine 15 windows of 200 data points windows from the previous step to obtain a group of 3,000 data points. Figure 3.3 illustrates how this process is done.
3. Change the data value in every window to the positive numbers.
4. Find the average of the data value in each group.
5. Compare the average value of each group to find the group with the smallest value.
6. The group with the smallest average value is then selected as the significant range of EEG signals.

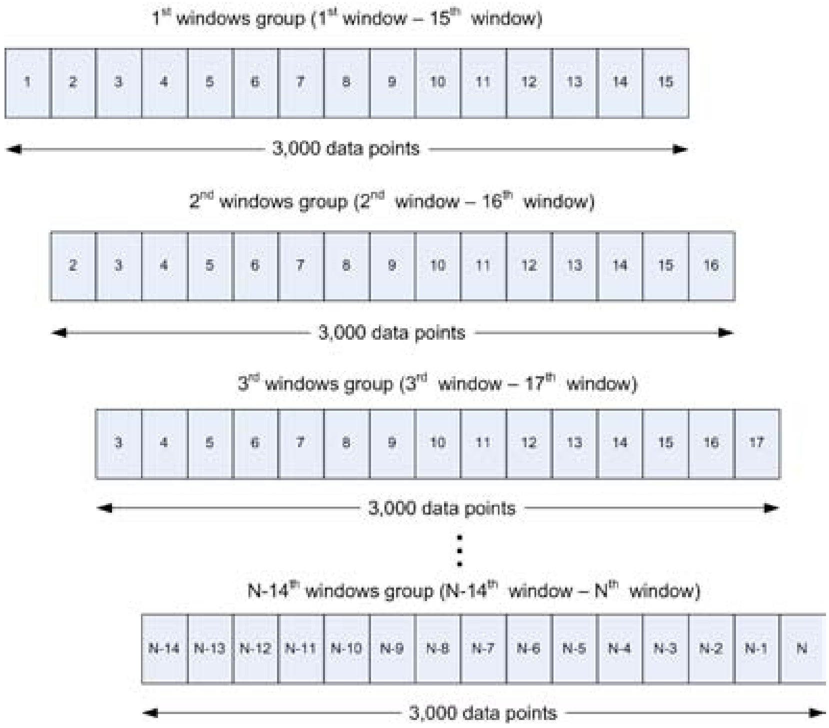


Figure 3.3: Combine 15 windows to be the group of 3,000 data points.

3.3 Selecting Appropriate ICA Algorithms

The collected signals from EEG electrodes were considered as the observed signals **o** in ICA since each observed signal o_i, for $1 \leq i \leq 16$, was a mixture of 16 EEG channels from different locations on the scalp. The purpose of this step is to select a set of appropriate ICA algorithms for this problem. For each subject, 3,000 observed EEG signals were sampled and separated to obtain each actual source signal s_i by deploying each ICA algorithm from 22 different ICA algorithms implemented in ICALAB [10]. An example of the EEG signals after being processed by ICA algorithm (ERICA) from channels 2, 10, 11, and 12 of a subject is shown in Figure 3.2(b).

To select the appropriate ICA algorithms, the obtained source signals were divided into training, validating, and testing patterns for neural classification. An appropriate ICA algorithm should give high classification accuracy. To prepare the source signals for neural classification, a sequence of 3,000 data points were partitioned into six sample sets of length 500 data points each. Three trials based on 5, 10, and 20 subjects were conducted. In each sample set, every sequence of 10 data points were grouped as follows: (a) the first six data points were grouped as training patterns, (b) the next two data points were grouped as validating patterns, and (c) the last two data points were grouped as testing patterns.

Figure 3.4 illustrates how training, validating, and testing patterns were grouped from the signals of channels 1 to 16. The numbers 1, 2, and 3 denote the training, validating, testing patterns, respectively. Based on this grouping scheme, it can be seen that 60% of the data points are for training, 20% are for validating, and the other 20% are for testing.

Since 16 channels were simultaneously considered in this process, each input pattern including training, validating, and testing, consisted of 16 elements. All training patterns were learned by a multi-layer perceptron with scaled conjugate gradient backpropagation learning rule. A 3-layer feed-forward neural network with 20 hidden neurons and ***n*** output neurons was deployed. The value of ***n*** is equal to the number of subjects. For example, there are five output neurons in case of five subjects. Hyperbolic tangent was used as the kernel function and activation function.

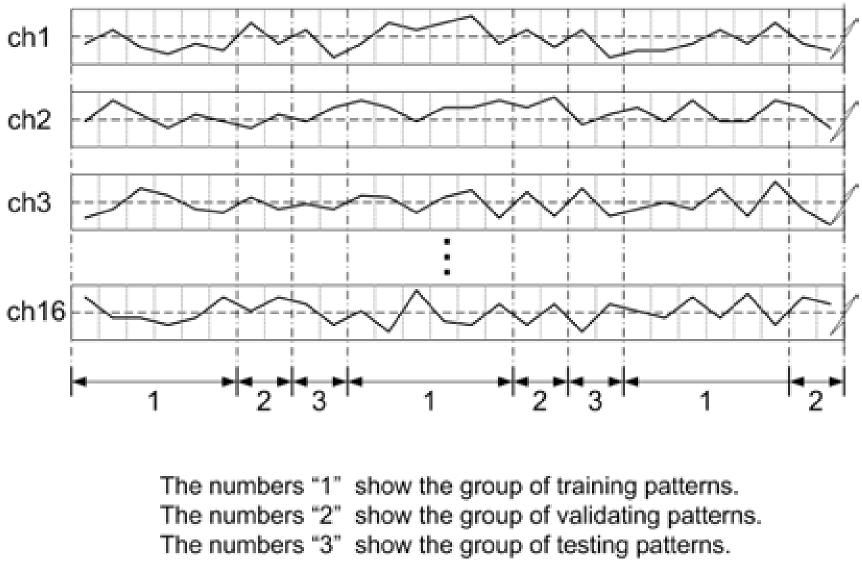


Figure 3.4: Format of data partitioning to form the training, validating, and testing groups.

3.3.1 Experimental Results

Table 3.2 summarizes the accuracy of neural classification with different ICA algorithms when applied to 5, 10, and 20 subjects. 8 ICA algorithms namely ERICA, EWASOBI, JADEop, SIMBEC, SOBI, SOBI-BPF, SOBIRO, WASOBI were the appropriate ICA candidates for further experiments because these algorithms give 100% accuracy for all subject numbers.

Table 3.2: The accuracy percentage of each ICA algorithm when applied to 5, 10, and 20 subjects.

Algorithms	5 Subjects	10 Subjects	20 Subjects
AMUSE	91.44	78.54	57.16
ERICA	**100.00**	**100.00**	**100.00**
EVD2	57.16	84.78	55.66
EWASOBI	**100.00**	**100.00**	**100.00**
FAJDC4	100.00	100.00	95.00
FJADE	100.00	100.00	95.00
FOBI-E	96.20	81.02	44.85
JADEop	**100.00**	**100.00**	**100.00**
JADETD	94.85	75.96	51.16
MULCOMBI	100.00	100.00	95.00
POWERICA	100.00	100.00	95.00
QJADE	100.00	100.00	90.00
SAD	90.88	70.34	45.67
SIMBEC	**100.00**	**100.00**	**100.00**
SOBI	**100.00**	**100.00**	**100.00**
SIOBI-BPF	**100.00**	**100.00**	**100.00**
SOBIRO	**100.00**	**100.00**	**100.00**
SONS	94.40	74.38	50.43
SYMMETRIC	100.00	100.00	89.62
THINICA	100.00	100.00	95.00
UNICA	100.00	100.00	95.00
WASOBI	**100.00**	**100.00**	**100.00**

3.4 Selecting Minimum Relevant Channels

In order to determine the minimum number of relevant channels for personal identification, all possible channel combinations, i.e. ch1 to ch16 were tested by starting from one channel, two channel combinations, three channel combinations and four channel combinations. Since only one channel could not a satisfied results. See more discussion in section 3.4.1. Here, Erica algorithm was employed for all channel combinations. The best channels are those combinations that produce the highest classification accuracy for all groups of subjects. Next, 2, 3, and 4 combinations were considered. In case of two channels, there are (${}^{16}C_2$) or 120 possible combinations. The signals from each combination set were processed and classified by the process discussed in the previous section. A 3-layer feed-forward neural network with 20 hidden neurons and *n* output neurons was deployed. Hyperbolic tangent was used as the kernel function and activation function.

3.4.1 Experimental Results

Table 3.3 shows the result of one channel. The accuracy percentage for all subjects numbers could not reach 100 percent. As a matter of fact, the accuracy percentages are quite low. Table 3.4 shows the combination groups that give 100% accuracy in 5, 10 and 20 experimental subjects. The best result occurs when there are only 5 subjects. But when the number of subjects increases, the accuracy decreases. It is obvious that using only 2 channels is not effective enough to identify a large group of subjects. A similar procedure is adopted for the cases of 3-channel and 4-channel combinations. Table 3.5 illustrates the top 5 with the highest accuracy channel combinations on 3 channels. Note that, with 3 channels, there are (${}^{16}C_3$) or 560 possible combinations. Table 3.6 displays the rank in order of the channel combination using 4 channels. Here, there are (${}^{16}C_4$) or 1,820 possible combinations. Interestingly, in the case of 4-channel combinations, the groups of 5 and 10 experimental subjects can achieve 100% accuracy. Whereas, the group of 20 experimental subjects achieved almost 100% accuracy.

Table 3.3: The percentage of accuracy for one channel when applied to 5, 10 and 20 subjects with ERICA algorithm.

Channel	5 Subjects	10 Subjects	20 Subjects
ch 1	88.04	54.70	27.76
ch 2	80.96	59.34	36.64
ch 3	84.56	50.68	29.20
ch 4	65.88	64.86	35.15
ch 5	81.80	59.42	35.27
ch 6	69.96	52.90	31.10
ch 7	62.80	54.36	29.49
ch 8	65.40	43.64	30.25
ch 9	85.88	51.94	33.26
ch10	93.52	63.56	33.23
ch11	85.00	57.32	38.37
ch12	75.60	48.02	29.06
ch13	74.56	57.54	34.47
ch14	71.72	53.64	30.58
ch15	79.48	49.34	32.50
ch16	92.36	53.84	31.68

Table 3.4: The percentage of accuracy for 2 channels when applied to 5, 10 and 20 subjects with ERICA algorithm.

Channel	Combination	5 Subjects	10 Subjects	20 Subjects
ch 2	ch 7	100.00	91.96	61.67
ch 2	ch 9	100.00	91.12	59.04
ch 2	ch14	100.00	72.62	47.76
ch 2	ch15	100.00	74.28	48.44
ch 5	ch10	100.00	71.90	39.46
ch 5	ch15	100.00	83.76	20.34
ch 9	ch10	100.00	88.42	44.07
ch10	ch16	100.00	88.10	54.05
ch12	ch15	100.00	64.96	54.62

Table 3.5: The percentage of accuracy for 3 channels when applied to 5, 10 and 20 subjects with ERICA algorithm.

Channel	Combination		5 Subjects	10 Subjects	20 Subjects
ch 4	ch 5	ch10	100.00	100.00	90.52
ch 4	ch10	ch13	100.00	99.96	68.08
ch 4	ch13	ch15	100.00	100.00	75.16
ch 5	ch 9	ch10	100.00	99.98	55.34
ch 9	ch13	ch15	100.00	99.98	83.41

Table 3.6: The percentage of accuracy for 4 channels when applied to 5, 10 and 20 subjects with ERICA algorithm.

Channel Combination				5 Subjects	10 Subjects	20 Subjects
ch 2	ch10	ch11	ch12	100.00	100.00	98.71
ch 2	ch 4	ch 5	ch15	100.00	100.00	98.03
ch 4	ch 5	ch10	ch15	100.00	100.00	97.34
ch 2	ch 4	ch 7	ch11	100.00	100.00	96.96
ch 2	ch 4	ch13	ch15	100.00	100.00	96.43

3.5 Confirmation of the 4-Channel Combinations

Current assumptions stated that 4-channel combination is the best candidate for classification. To confirm this assumption, the neural network must be able to determine a person not belonging to one of 20 experimental subjects as an unknown person (outsider subjects). The determining algorithm called *Outsider* will be discussed in the next section. Three related factors must be concerned, namely

1. Eight selected ICA algorithms from Table 3.2. ERICA was used in last experiments section and the other 7 ICA algorithms, including EWASOBI, JADEop, SIMBEC, SOBI, SOBI-BPF, SOBIRO, WASOBI were used to confirm that 4-channel combination is the best candidate for classification. The results of applying 8 different ICA algorithms are summarized in Table 3.7.
2. Twenty additional experimental outsider subjects. These new subjects were experimented by the same process as discussed in Section 3.4.
3. More considered channels. To determine the minimum number of channels and the best combination of these channels, it is necessary to experiment with more channels and observe the situation when the identification accuracy is unchanged. Here, the numbers of channel from 5 to 7 were studied. The rationale of using these numbers will be clarified in Section 3.5.3.

Table 3.7: Percentage of accuracy for 4-channel combinations, 500 data points with different ICA algorithms.

ICA Algorithms	The best accuracy for 5 subjects	The best accuracy for 10 subjects	The best accuracy for 15 subjects
ERICA	1,502 combinations reach 100%	185 combinations reach 100%	ch 2 ch10 ch11 ch12 98.71%
SOBI-BPF	1,511 combinations reach 100%	218 combinations reach 100%	ch 3 ch 8 ch 9 ch12 98.57%
SOBIRO	170 combinations reach 100%	22 combinations reach 100%	ch 5 ch10 ch13 ch15 97.87%
SOBI	110 combinations reach 100%	4 combinations reach 100%	ch 6 ch 7 ch12 ch15 96.22%
WASOBI	1,814 combinations reach 100%	65 combinations reach 100%	ch 3 ch 6 ch 9 ch10 94.26%
JADEop	40 combinations reach 100%	4 combinations reach 100%	ch 9 ch11 ch14 ch15 93.48%
SIMBEC	110 combinations reach 100%	14 combinations reach 100%	ch 8 ch12 ch13 ch16 93.48%
EWASOBI	862 combinations reach 100%	34 combinations reach 100%	ch 4 ch 5 ch 8 ch13 91.17%

Before discussing the optimum number, the process of how to select the best combination of four relevant channels in detail was repeated as follows. Four experiments were conducted. In the first experiment, 3000 data signal points were partitioned into six groups. Each group contained 500 data points. In the second experiment, 3000 data signal points were partitioned into three groups. Each group had 1000 data points. In the third experiment, 3000 data signal points were partitioned into two separated groups. Each group consisted of 1500 data points. Lastly, in the fourth experiment, all 3000 data signal points were put into a single group.

Experiment 1: 3,000 data points signals from each subject were partitioned into six groups of 500 data points each. These 500 data points in the first group were divided into training, validating, and testing patterns for classification. But the data points from groups 2 to 6 were used for testing only. All six groups are named as $G_1^{(500)}$, $G_2^{(500)}$, $G_3^{(500)}$, $G_4^{(500)}$, $G_5^{(500)}$, and $G_6^{(500)}$.

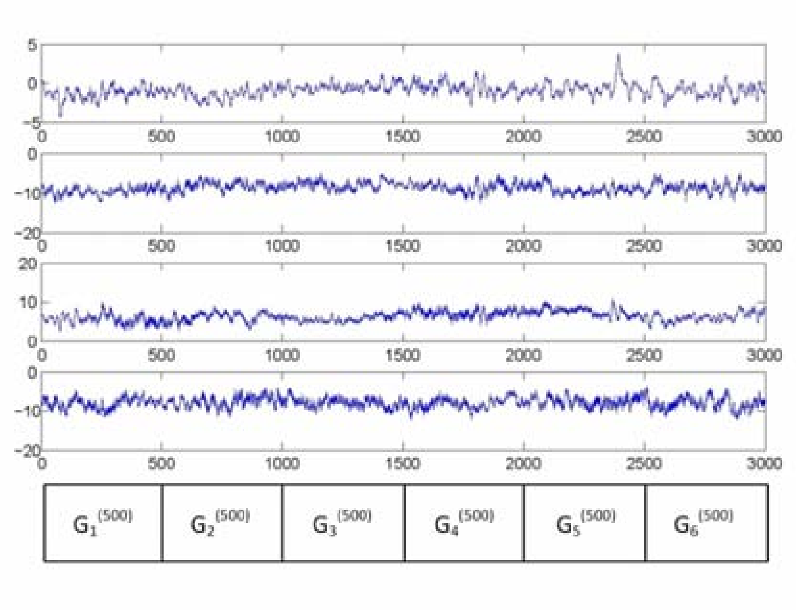


Figure 3.5: The EEG signal 3,000 data points are divided into six groups for testing.

Experiment 2: 3,000 data points signals from each subject were partitioned into three groups of 1000 data points each. These 1,000 data points in the first group were divided into training, validating and testing patterns for classification. But the data points from group 2 and 3 were used for testing only. All three groups are named as $G_1^{(1000)}$, $G_2^{(1000)}$, $G_3^{(1000)}$.

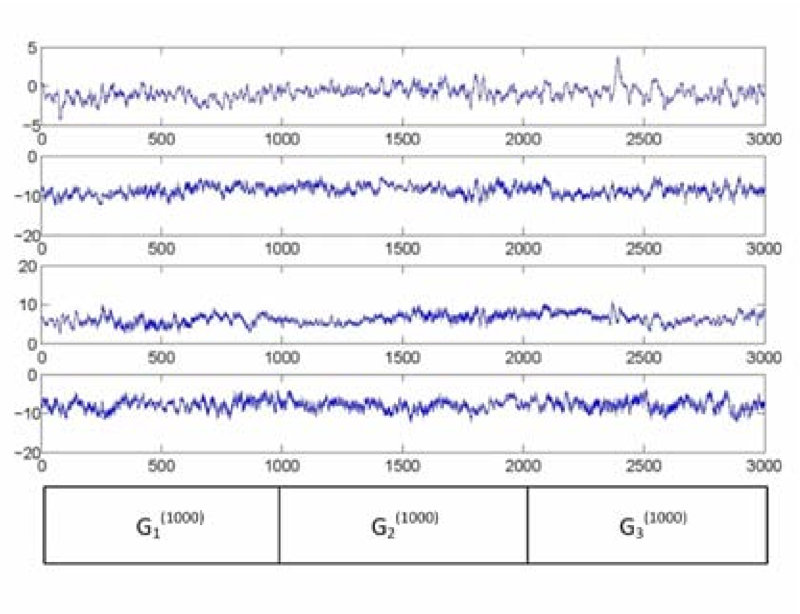


Figure 3.6: The EEG signal 3,000 data points are divided into three groups for testing.

Experiment 3: 3,000 data points signals from each subject were partitioned into two groups of 1500 data points each. The first group was divided into training, validating and testing patterns for classification. But the data points from another group were used for testing only. Both groups are named as $G_1^{(1500)}$, $G_2^{(1500)}$.

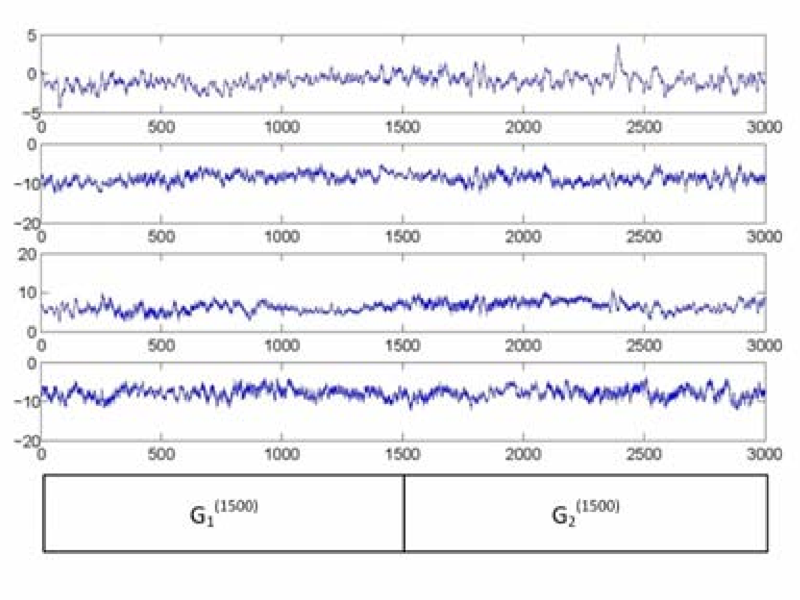


Figure 3.7: The EEG signal 3,000 data points are divided into two groups for testing.

Experiment 4: All 3,000 data point signals from each subject were used for this experiment. This group was divided into training, validating and testing patterns for classification which are named as $G_1^{(3000)}$. In addition to the above experiments, the neural network must be able to determine a person who does not belong to one of 20 experimental subjects as an unknown person. The determining algorithm called Outsider will be discussed in the next section.

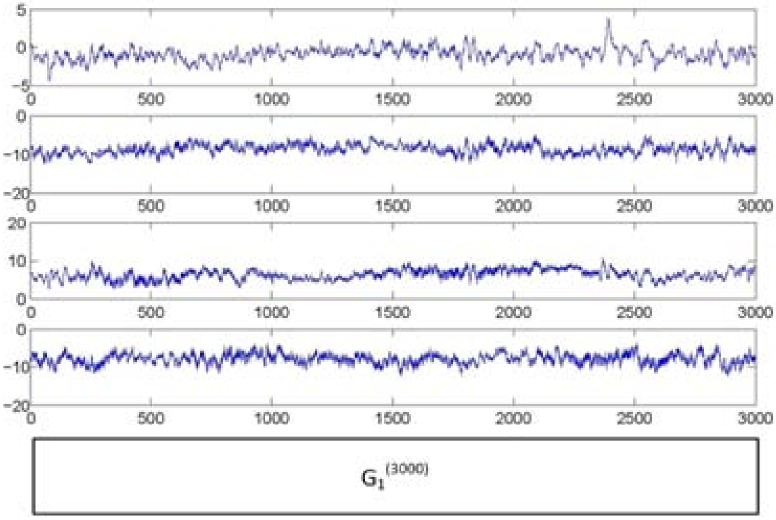


Figure 3.8: The EEG signal 3,000 data points, all for testing.

3.5.1 Determining Outsider Subject

The group of first 20 subjects, i.e. subjects 1 to 20, were used for training or called the insider group. Any subject not belonging to the insider group is called outsider. Let $G_i^{(m)}$ be a set of ***m*** data points of signal group *i*, for $1 \le i \le M$. Suppose subject ***j*** belongs to an *insider* group. Let $a(j, G_i^{(m)})$ be the percentage of accuracy of subject ***j*** tested by the *m* data points in signal group ***i***. The minimum percentage of accuracy of subject ***j*** is defined as $\min_{\forall i} a(j, G_i^{(m)})$.

Suppose subject ***k*** is an unknown subject and it is determined as subject $\hat{k}$ by the neural network with percentage of accuracy of $a(k, G_i^{(m)})$, for $1 \le i \le M$. The maximum percentage of accuracy of subject ***k*** is defined as $\max_{\forall i} a(k, G_i^{(m)})$.

Subject *k* is determined as an outsider if

1. The neural network determines as different subject $\hat{k}$ in each group in each experiment. For example in Table 3.11, subject 32 is determined as subject 13 in sets $G_1^{(1000)}$ and $G_2^{(1000)}$ but it is determined as subject 20 in set $G_3^{(1000)}$.
2. The maximum percentage of accuracy of subject *k* smaller than the minimum percentage of accuracy of subject $\hat{k}$ is defined as follows:

$$\max_{\forall i} a(k, G_i^{(m)}) < \min_{\forall i} a(\hat{k}, G_i^{(m)}) \qquad (6.1)$$

3.5.2 Experimental Results

Experiment 1

The result of experiment 1 is summarized in Tables 3.8 and 3.9. Here, the observed signals from ch2, ch10, ch11, and ch12 by ERICA algorithm were used in the experiment since this combination gives the highest accuracy.

Three parameters used in ERICA algorithm were set as follows:

(a) Pre-whitening was set to *ye*s.

(b) The maximum number of iteration was to 1000.

(c) Ordering was set to *none*.

There are 40 subjects considered in the process. The observed signals from the first 20 subjects were used for training as well as testing. The other additional 20 subjects, i.e. subjects 21 to 40, were not involved in any training process but were used to determine subjects 21 to 40 as unknown persons. From the experiment, it can be seen that some outsider subjects were not correctly identified. The numbers in the parentheses in Table 3.9 denote the subjects wrongly identified by the neural network. For example, subject 21 in the first row of Table 3.9 was wrongly identified as subject 16 when being tested with the data points in groups $G_1^{(500)}$, $G_2^{(500)}$, $G_3^{(500)}$, $G_4^{(500)}$, $G_5^{(500)}$, and $G_6^{(500)}$, respectively. In this case $\max_{\forall i} a(k, G_i^{(m)})$ from Table 3.9, subject 21, $G_2^{(500)}$ was equal to 99.60 and $\min_{\forall i} a(\hat{k}, G_i^{(m)})$ from Table 3.8, subject 16, $G_5^{(500)}$ was equal to 89.00. Since $\max_{\forall i} a(k, G_i^{(m)}) \geq \min_{\forall i} a(\hat{k}, G_i^{(m)})$. It is obvious that using 500 data points can correctly identified only insider subjects but cannot correctly identified outsider subjects.

Table 3.8: The percentage of accuracy for 1-20 subjects (insider) when tested with the best 4-channel combinations, ERICA algorithm, and 500 data points in experiment 1.

Insider Subject	$G_1^{(500)}$	$G_2^{(500)}$	$G_3^{(500)}$	$G_4^{(500)}$	$G_5^{(500)}$	$G_6^{(500)}$
1	100.00	100.00	99.20	98.40	93.40	98.80
2	100.00	100.00	100.00	98.20	100.00	100.00
3	100.00	100.00	100.00	100.00	100.00	100.00
4	100.00	100.00	100.00	100.00	100.00	100.00
5	100.00	99.60	100.00	99.80	100.00	100.00
6	97.00	99.60	99.80	99.60	97.60	99.40
7	99.60	97.80	95.00	99.00	94.60	94.00
8	100.00	99.40	92.40	100.00	100.00	99.60
9	100.00	100.00	100.00	99.80	100.00	100.00
10	97.40	96.40	93.40	97.80	96.20	90.00
11	100.00	99.00	95.60	99.00	98.80	99.60
12	100.00	100.00	100.00	99.20	99.80	95.00
13	99.80	98.60	99.20	100.00	100.00	100.00
14	95.40	81.00	81.80	91.60	82.40	81.20
15	99.40	92.80	95.40	97.00	92.60	94.20
16	96.80	96.20	97.60	94.80	**89.00**	93.00
17	94.60	92.60	94.60	94.40	98.00	96.40
18	95.80	89.40	93.20	94.80	86.60	90.60
19	98.40	98.40	99.40	98.40	99.60	90.60
20	100.00	99.80	100.00	100.00	99.80	100.00

Table 3.9: The percentage of accuracy for subjects 21-40 (outsider) tested with the best 4-channel combinations, ERICA algorithm, and 500 data points in experiment 1.

Outsider Subject	$G_1^{(500)}$		$G_2^{(500)}$		$G_3^{(500)}$		$G_4^{(500)}$		$G_5^{(500)}$		$G_6^{(500)}$	
21	(16)	99.00	**(16)**	**99.60**	(16)	98.80	(16)	99.00	(16)	98.60	(16)	99.40
22	(11)	66.00	(11)	69.80	(11)	71.00	(11)	73.80	(11)	65.60	(11)	71.00
23	(7)	76.60	(7)	59.20	(7)	77.20	(7)	68.00	(7)	63.40	(7)	69.20
24	(6)	49.60	(6)	48.60	(6)	40.00	(6)	51.60	(6)	40.20	(6)	33.80
25	(7)	48.40	(7)	51.20	(7)	49.40	(7)	45.00	(7)	48.00	(7)	47.80
26	(7)	92.20	(7)	95.00	(7)	97.60	(7)	98.00	(7)	93.20	(7)	98.80
27	(6)	53.00	(19)	46.00	(19)	45.20	(19)	54.40	(6)	52.80	(6)	53.60
28	(7)	41.00	(7)	44.60	(7)	49.40	(7)	36.60	(7)	35.20	(7)	41.40
29	(1)	81.60	(1)	82.80	(1)	84.40	(1)	84.80	(1)	85.60	(1)	77.20
30	(7)	83.20	(7)	81.60	(7)	88.60	(7)	88.20	(7)	93.40	(7)	88.40
31	(13)	53.80	(13)	71.40	(13)	65.40	(13)	80.60	(13)	56.00	(13)	65.80
32	(6)	35.00	(19)	58.80	(19)	37.40	(19)	46.00	(19)	41.00	(19)	45.80
33	(20)	51.40	(20)	53.40	(20)	56.20	(20)	49.80	(20)	51.00	(20)	56.00
34	(13)	63.00	(13)	59.80	(13)	46.80	(13)	45.40	(13)	49.80	(13)	50.20
35	(12)	85.80	(12)	82.80	(12)	76.80	(12)	84.40	(12)	77.20	(12)	82.00
36	(7)	97.40	(7)	97.00	(7)	96.40	(7)	93.20	(7)	95.60	(7)	95.60
37	(7)	52.80	(7)	47.20	(7)	61.40	(7)	89.40	(7)	82.40	(7)	79.00
38	(14)	57.00	(14)	71.40	(14)	84.80	(14)	75.60	(14)	68.20	(14)	70.60
39	(15)	68.00	(15)	48.20	(15)	50.80	(15)	63.60	(15)	63.60	(15)	58.60
40	(14)	47.20	(14)	34.00	(14)	40.60	(14)	49.80	(14)	59.00	(14)	43.20

Experiment 2

Both ERICA and SOBI-BPF algorithms were used in this experiment. However, the results based on these algorithms are not satisfactory enough. Hence, SOBIRO algorithm was used instead. Two parameters used in SOBIRO algorithm were set as follows.

(a) Number of time-delayed covariance matrices was set to 100.

(b) Ordering was set to *none*.

Prior to the identification process, the most optimum 4-channel combination was required.

Table 3.10 summarizes the combinations that achieved an accuracy higher than 95%. The best combination is from ch13, ch14, ch15, and ch16. The signals from these channels were used to authenticate 20 subjects. The process of training and testing is as follows. The first 1000 data points in $G_1^{(1000)}$ were for training, validating and testing but the other data points in $G_2^{(1000)}$ and $G_3^{(1000)}$ were for testing.

Table 3.11 shows the identification accuracy of each subject. Subjects 21 to 40 were not involved in the training process and quantified as unknown persons. The numbers in the parentheses denote the subjects that were inaccurately identified by the neural network. Based on the outsider algorithm previously discussed, it can be seen that subjects 21 to 40 can be correctly determined as unknown persons.

Table 3.10: Selected 4-channel combination having percentage of accuracy higher than 95% with SOBIRO algorithm, and 1,000 data points in experiment 2.

Channel Combination				Average Accuracy percentage
ch13	ch14	ch15	ch16	98.51
ch 3	ch 6	ch 9	ch15	98.15
ch 3	ch 6	ch13	ch15	98.04
ch 6	ch13	ch14	ch16	97.74
ch 6	ch13	ch15	ch16	97.52
ch 5	ch 6	ch15	ch16	97.47
ch 3	ch12	ch13	ch16	96.66
ch 9	ch12	ch13	ch14	96.6
ch 2	ch 5	ch 6	ch16	96.54
ch 5	ch 6	ch 7	ch 8	96.4
ch 1	ch 9	ch15	ch16	96.18
ch 3	ch11	ch12	ch16	96.08
ch 4	ch12	ch13	ch16	95.94
ch 9	ch14	ch15	ch16	95.79
ch 3	ch 4	ch 5	ch 6	95.5
ch 3	ch 8	ch14	ch15	95.47
ch 2	ch 6	ch14	ch15	95.31
ch 6	ch11	ch12	ch13	95.28
ch 3	ch 4	ch15	ch16	95.21
ch 1	ch 8	ch 9	ch15	95.14

Table 3.11: The percentage of accuracy for 40 subjects tested with the best 4-channel combination (ch13, ch14, ch15, ch16), SOBIRO algorithm, and 1,000 data points in experiment 2.

Insider Subject	$G_1^{(1000)}$	$G_2^{(1000)}$	$G_3^{(1000)}$	Outsider Subject	$G_1^{(1000)}$	$G_2^{(1000)}$	$G_3^{(1000)}$
1	99.90	100.00	100.00	**21**	(17) 75.00	(17) 79.40	(17) 67.60
2	99.60	100.00	100.00	**22**	(7) 63.40	(7) 65.20	(7) 57.40
3	100.00	100.00	100.00	**23**	(18) 73.00	(18) 72.60	(18) 69.80
4	100.00	99.80	100.00	**24**	(18) 60.20	(18) 67.60	(18) 61.20
5	100.00	99.80	99.00	**25**	(11) 54.20	(11) 52.40	(11) 51.00
6	99.90	99.80	99.60	**26**	(20) 60.60	(20) 67.20	(20) 65.80
7	99.00	94.60	92.00	**27**	(7) 49.00	(7) 47.20	(7) 57.80
8	99.90	99.60	100.00	**28**	(7) 34.00	(7) 32.60	(7) 32.80
9	99.60	100.00	100.00	**29**	(6) 60.80	(6) 50.20	(6) 56.60
10	99.70	99.40	99.20	**30**	(18) 43.80	(18) 41.00	(18) 40.80
11	99.90	100.00	91.00	**31**	(11) 87.80	(11) 46.60	(11) 85.20
12	96.70	99.60	98.00	**32**	(13) 37.40	(13) 32.60	(20) 29.00
13	93.90	96.00	94.20	**33**	(13) 84.00	(13) 80.60	(13) 85.20
14	98.40	98.80	96.40	**34**	(20) 78.40	(20) 86.80	(20) 87.80
15	99.70	99.20	98.20	**35**	(20) 59.60	(20) 62.40	(20) 65.60
16	90.50	88.00	91.40	**36**	(18) 79.40	(18) 76.40	(18) 67.20
17	99.70	99.00	99.60	**37**	(15) 25.20	(15) 29.40	(15) 27.80
18	97.40	98.40	96.40	**38**	(19) 85.00	(19) 75.20	(19) 87.00
19	98.90	100.00	100.00	**39**	(7) 69.60	(7) 71.20	(7) 70.80
20	97.50	97.60	98.40	**40**	(12) 33.60	(12) 34.80	(12) 29.00

Experiment 3

The combination of channels ch6, ch9, ch10, and ch15 by SOBIRO algorithm was selected for the identification process. Table 3.12 summarizes the combinations having percentage of accuracy higher than 95%. The signals in group $G_1^{(1500)}$ of subjects 1 to 20 were for training, validating and testing but the signals in group $G_2^{(1500)}$ were for testing only. To test whether the neural network can correctly determine a person as an unknown person, subjects 21 to 40 were involved only in testing by the outsider algorithm. Table 3.13 shows the percentage of accuracy for 40 subjects. The numbers in the parenthesis denote the subject wrongly authenticated by the neural network. It can be seen that the network correctly determined subjects 21 to 40 as unknown persons.

Table 3.12: Selected 4-channel combination having percentage of accuracy higher than 95% with SOBIRO algorithm, and 1,500 data points in experiment 3.

Channel Combination				Average Accuracy percentage
ch 6	ch 9	ch10	ch15	98.8
ch 5	ch 6	ch 7	ch15	98.75
ch 6	ch 9	ch14	ch15	98.71
ch 6	ch 7	ch13	ch15	98.68
ch 2	ch 6	ch15	ch16	98.47
ch 3	ch 6	ch13	ch15	98.11
ch 1	ch 6	ch 9	ch15	98.02
ch 6	ch 8	ch 9	ch15	97.87
ch 1	ch 6	ch13	ch15	97.79
ch 9	ch10	ch13	ch15	97.61
ch 2	ch 6	ch10	ch15	97.39
ch 2	ch 6	ch 7	ch15	97.24
ch 9	ch13	ch15	ch16	96.91
ch 6	ch13	ch14	ch15	96.82
ch 3	ch 5	ch15	ch16	96.33
ch 3	ch 5	ch 7	ch15	96.19
ch 2	ch12	ch15	ch16	95.88
ch 3	ch 7	ch 9	ch15	95.67
ch 2	ch 6	ch 8	ch15	95.66
ch 3	ch 7	ch12	ch15	95.63

Table 3.13: The percentage of accuracy for 40 subjects tested with the best 4-channel combination (ch6, ch9, ch10, ch15), SOBIRO algorithm, and 1,500 data points in experiment 3.

Insider Subject	$G_1^{(1500)}$	$G_2^{(1500)}$	Outsider Subject	$G_1^{(1500)}$	$G_2^{(1500)}$
1	100.00	100.00	**21**	(10) 41.20	(10) 51.60
2	100.00	99.60	**22**	(5) 30.20	(5) 44.80
3	100.00	100.00	**23**	(11) 60.40	(11) 40.80
4	100.00	100.00	**24**	(17) 76.60	(17) 73.40
5	100.00	100.00	**25**	(10) 48.60	(10) 45.80
6	99.70	98.20	**26**	(7) 82.80	(7) 83.00
7	99.60	97.60	**27**	(7) 77.80	(7) 68.80
8	99.90	100.00	**28**	(14) 47.20	(14) 78.40
9	100.00	100.00	**29**	(8) 69.80	(8) 70.40
10	100.00	100.00	**30**	(19) 57.40	(19) 61.80
11	99.30	97.80	**31**	(17) 41.80	(17) 58.60
12	100.00	100.00	**32**	(12) 61.60	(12) 51.80
13	92.70	98.00	**33**	(10) 78.60	(10) 69.80
14	96.40	95.80	**34**	(15) 46.80	(15) 63.60
15	93.20	94.60	**35**	(13) 49.60	(13) 67.80
16	98.60	96.40	**36**	(15) 67.60	(15) 49.00
17	96.70	98.00	**37**	(15) 67.80	(15) 79.00
18	100.00	100.00	**38**	(2) 67.80	(2) 89.00
19	100.00	100.00	**39**	(15) 47.20	(15) 65.20
20	100.00	100.00	**40**	(17) 93.80	(17) 94.80

Experiment 4

SOBIRO algorithm was still used in this experiment. The combination of channels ch1, ch13, ch15, and ch16 was selected for the identification process since it gave the highest accuracy for 3,000 data points as shown in Table 3.14. The signals in group $G_1^{(3000)}$ of subjects 1 to 20 were for training, validating and testing and in group $G_1^{(3000)}$ of subjects 21 to 40 were for testing. To test whether the neural network can correctly determine a person as an unknown person, subjects 21 to 40 were involved only in testing by the outsider algorithm. Table 3.15 shows the percentage of accuracy for 40 subjects. It can be seen that the network correctly determined subjects 21 to 40 as unknown persons.

The results of experiments 1-4 suggested that only 500 data points were eligible to determine insider subjects but insufficient to use for identification of outsider subjects. From the experiments, we found that 1,000, 1,500, and 3,000 data points can correctly determine insider and outsider subjects with high accuracy.

From Tables 3.11, 3.13, and 3.15, the numbers in parentheses denote the subject numbers authenticated by the neural network. These subject numbers are used to determine outsider subjects. For any subject, if the subject number identified by the neural network is different in each experiment, then that subject is determined as the outsider subject. For example, subject 27 in Table 3.11 and Table 3.13 were determined as subject 7 but this subject in Table 3.15 was determined as subject 5. Hence, subject 27 was considered as the outsider subject.

Table 3.14: Selected 4-channel combination having percentage of accuracy higher than 95% with SOBIRO algorithm, and 3,000 data points in experiment 4.

Channel Combination				Average Accuracy percentage
ch 1	ch13	ch15	ch16	98.85
ch 5	ch 6	ch 7	ch15	98.73
ch10	ch13	ch15	ch16	98.70
ch 6	ch 9	ch10	ch15	98.67
ch 3	ch 5	ch13	ch15	98.37
ch 6	ch 7	ch 9	ch15	98.20
ch12	ch13	ch15	ch16	98.10
ch 6	ch13	ch15	ch16	98.05
ch 3	ch 6	ch13	ch15	97.99
ch 6	ch 9	ch13	ch15	97.71
ch 6	ch 9	ch12	ch15	97.38
ch 3	ch12	ch13	ch16	97.15
ch 9	ch14	ch15	ch16	96.82
ch 1	ch 3	ch13	ch15	96.64
ch 1	ch 5	ch 9	ch16	96.59
ch 3	ch 9	ch15	ch16	96.17
ch 9	ch12	ch15	ch16	96.05
ch 3	ch13	ch14	ch15	95.84
ch 3	ch 7	ch13	ch15	95.50
ch 3	ch 9	ch13	ch15	95.16

Table 3.15: The percentage of accuracy for 40 subjects tested with the best 4-channel combination (ch1, ch13, ch15, ch16), SOBIRO algorithm, and 3,000 data points in experiment 4.

Insider Subject	$G_1^{(3000)}$	Outsider Subject	$G_1^{(3000)}$
1	100.00	**21**	(12) 39.20
2	99.97	**22**	(20) 48.80
3	99.97	**23**	(20) 55.80
4	100.00	**24**	(20) 63.00
5	99.73	**25**	(1) 33.20
6	99.87	**26**	(19) 41.80
7	99.67	**27**	(5) 51.40
8	100.00	**28**	(20) 69.80
9	99.73	**29**	(3) 70.20
10	100.00	**30**	(14) 59.20
11	99.73	**31**	(11) 94.20
12	98.43	**32**	(20) 89.80
13	95.63	**33**	(5) 40.00
14	97.13	**34**	(12) 44.20
15	98.30	**35**	(15) 50.60
16	98.67	**36**	(18) 51.00
17	99.20	**37**	(20) 40.40
18	95.37	**38**	**(19) 95.80**
19	99.60	**39**	(20) 73.20
20	96.07	**40**	(4) 39.00

3.5.3 Compare 4 Channel with 5, 6, 7 Channel Combinations

In order to confirm the 4-channel combination is the best for classification. The combination of 5, 6 and 7 channels were explored and compared with 4-channel combination. The experiment based on 3,000 data points of SOBIRO algorithm and 40 subjects were conducted. In case of 5 channels, there are $({}^{16}C_5)$ or 4,368 possible combinations. In case of 6 and 7 channels, there are 8,008, and 11,440 combinations. The results of percentage accuracy of 5, 6, 7 channels were shown in Tables 3.16, 3.17, and 3.18, respectively.

3.5.4 Experimental Results

To compare the accuracy percentage of insider subjects for 4, 5, 6, and 7 channel combinations, Tables 3.15-3.18 using the first column and second column were considered. From Tables 3.15, the average accuracy percentage for authentication of 20 insider subjects in case of 4 channels from ch1, ch13, ch15, and ch16 is 98.85%. From Table 3.16, for 5 channel combination from ch5, ch6, ch9, ch10, and ch15, the average accuracy percentage is 99.87%. From Table 3.17, for 6 channel combination from ch1, ch5, ch6, ch9, ch10, and ch15, the average accuracy percentage is 99.97%. From Table 3.18, for 7 channel combination from ch5, ch6, ch7, ch9, ch13, ch14, and ch15, the average accuracy percentage is 100%.

To compare the accuracy percentage of outsider subjects for 4, 5, 6, and 7 channel combinations, Tables 3.15-3.18 using the third column and fourth column, were considered. From Table 3.15, the maximum percentage of accuracy for 4 channel combination is 95.80% from subjects 38. This indicates that subject 38 was incorrectly identified as subject 19. From Table 3.16, 5 channel combination, the maximum percentage of accuracy is 97.60% from subject 32. This indicates that subject 32 was inaccurately authenticated as subject 12. From Table 3.17, 6 channel combination, the maximum percentage accuracy is 99.40% from subject 39. This indicates that subject

39 was inaccurately authenticated as subject 15. From Table 3.18, 7 channel combination, the maximum percentage accuracy is 99.40% from subject 23. This indicates that subject 23 was inaccurately authenticated as subject 12.

To compare the difference between the accuracy percentage of outsiders, who were wrong predicted by neural network, and insiders for 4, 5, 6, 7 channel combinations, Table 3.15-3.18 were used for analysis. The accuracy percentage of outsiders is defined by $\max_{\forall i} a(k, G_i^{(m)})$. Whereas the accuracy percentage of insiders, who are wrong authenticated, is defined by $\min_{\forall i} a(\hat{k}, G_i^{(m)})$. From Table 3.15, the difference between $\max_{\forall i} a(k, G_i^{(m)})$ and $\min_{\forall i} a(\hat{k}, G_i^{(m)})$ for 4 channel combination is 3.8. In the same manner, from Tables 3.16-3.18, the differences of accuracy percentage are 2.4, 0.6, and 0.6, respectively. These results are summarized in Table 3.19. Notice that, for any subject, when the number of channels increases, the difference between the accuracy percentage in case of authenticating as an outside and as an insider is decreased as shown in Figure 3.9.

Table 3.16: The percentage of accuracy for 40 subjects tested with the best 5-channel combination (ch5, ch6, ch9, ch10, ch15), SOBIRO algorithm, and 3,000 data points.

Insider Subject	$G_1^{(3000)}$	Outsider Subject	$G_1^{(3000)}$
1	100.00	**21**	(8) 75.40
2	100.00	**22**	(4) 64.60
3	100.00	**23**	(15) 35.80
4	100.00	**24**	(15) 64.60
5	100.00	**25**	(12) 45.00
6	100.00	**26**	(7) 88.20
7	99.97	**27**	(7) 64.80
8	100.00	**28**	(14) 58.00
9	100.00	**29**	(15) 30.60
10	100.00	**30**	(15) 87.80
11	99.90	**31**	(7) 92.20
12	100.00	**32**	**(12) 97.60**
13	100.00	**33**	(14) 61.20
14	98.70	**34**	(15) 84.00
15	99.97	**35**	(15) 85.20
16	99.27	**36**	(15) 91.60
17	99.60	**37**	(15) 92.80
18	100.00	**38**	(2) 96.80
19	100.00	**39**	(15) 90.40
20	100.00	**40**	(15) 86.20
Average	99.87		

Table 3.17: The percentage of accuracy for 40 subjects tested with the best 6-channel combination (ch1, ch5, ch6, ch9, ch10, ch15), SOBIRO algorithm, and 3,000 data points.

Insider Subject	$G_1^{(3000)}$	Outsider Subject	$G_1^{(3000)}$
1	100.00	**21**	(8) 74.80
2	100.00	**22**	(14) 64.60
3	100.00	**23**	(15) 73.40
4	100.00	**24**	(15) 63.20
5	100.00	**25**	(12) 53.00
6	100.00	**26**	(2) 61.40
7	99.97	**27**	(15) 52.80
8	99.97	**28**	(14) 90.20
9	100.00	**29**	(16) 77.40
10	100.00	**30**	(19) 98.40
11	99.90	**31**	(17) 96.40
12	100.00	**32**	(12) 78.00
13	100.00	**33**	(5) 47.00
14	99.90	**34**	(15) 94.80
15	100.00	**35**	(15) 97.80
16	99.90	**36**	(15) 92.00
17	99.77	**37**	(15) 88.20
18	100.00	**38**	(2) 69.00
19	100.00	**39**	**(15) 99.40**
20	100.00	**40**	(15) 98.00
Average	99.97		

Table 3.18: The percentage of accuracy for 40 subjects tested with the best 7-channel combination (ch5, ch6, ch7, ch9, ch13, ch14, ch15), SOBIRO algorithm, and 3,000 data points.

Insider Subject	$G_1^{(3000)}$	Outsider Subject	$G_1^{(3000)}$
1	100.00	**21**	(12) 72.00
2	100.00	**22**	(5) 98.80
3	100.00	**23**	**(12) 99.40**
4	100.00	**24**	(12) 84.20
5	100.00	**25**	(12) 95.60
6	100.00	**26**	(7) 43.20
7	100.00	**27**	(7) 90.80
8	100.00	**28**	(14) 40.00
9	100.00	**29**	(6) 68.40
10	100.00	**30**	(12) 47.40
11	100.00	**31**	(6) 72.60
12	100.00	**32**	(7) 37.60
13	100.00	**33**	(15) 96.40
14	100.00	**34**	(11) 48.00
15	100.00	**35**	(12) 74.20
16	100.00	**36**	(12) 77.80
17	100.00	**37**	(12) 66.20
18	100.00	**38**	(19) 87.80
19	100.00	**39**	(1) 33.40
20	100.00	**40**	(12) 61.20
Average	100.00		

Table 3.19: The difference between the accuracy percentage of inside and outsider subjects for 4, 5, 6, 7 channels experimented by SOBIRO algorithm and 3,000 data points.

Channel combination	Minimum accuracy percentage insider (%)	Maximum accuracy percentage outsider (%)	Different of accuracy percentage which outsider wrongly identified to insider
4-channel	subject 19 : 99.60	subject 38 : (19) 95.80	99.60-95.80 = 3.80
5-channel	subject 12 : 100.00	subject 32 : (12) 97.60	100.00-97.60 = 2.40
6-channel	subject 15 : 100.00	subject 39 : (15) 99.40	100.00-99.40 = 0.60
7-channel	subject 12 : 100.00	subject 23 : (12) 99.40	100.00-99.40 = 0.60

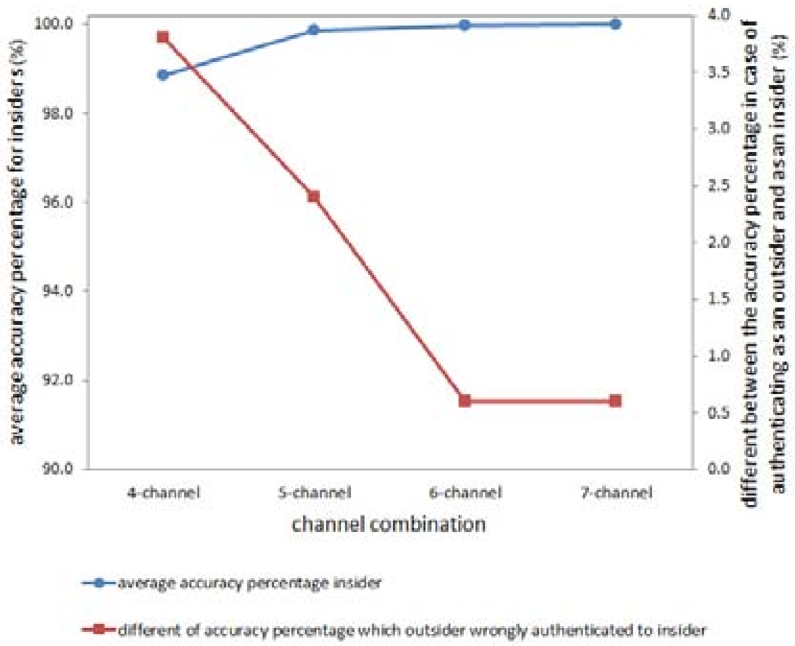


Figure 3.9: Compare the accuracy percentage of insider and the different between the accuracy percentage with the outsider wrongly identified to insider for 4, 5, 6 and 7 channel combinations.

3.6 Exploring Conditions Perform the Best for Identification

To enhance the performance efficiency for identification, four conditions were explored to find out which conditions perform the best for identification. EEG signals were collected from five subjects performing one stimulus condition, two mental thought conditions to compare with the relax condition. These conditions are:

1. Relax condition. The subjects were asked to relax and think of nothing in particular. This condition was used as a control and as a baseline measure of EEG signals.
2. Stimulus condition. The subject were flashed light in 60 seconds.
3. Mental thoughts condition for calculating math problems. The subjects were given math problems to solve in 60 seconds. These math problems are addition, subtraction, multiplication, and division. Subjects were asked to solve them without vocalizing or making any other physical movements.
4. Mental thoughts condition, thinking about the impressive image. The subjects were asked to think of the impressive image in 60 seconds.

The result of each condition when compared with relax condition is shown in Table 3.20

3.6.1 Experimental Results

From Table 3.20, the accuracy of identification of stimulus condition by flash light is decrease by 15.01% when compared with relaxing condition. Same as mental thoughts condition, the accuracy of calculation of math and thinking about the impressive image are decreased by 10.31% and 18.13%, respectively. This result shows the relax condition is the best for personal identification of this experiment.

Table 3.20: The percentage of accuracy of flash light, calculation of math, and impressive image compare with relax condition.

Flash light	Calculation of math	Impressive image
-15.01%	-10.31%	-18.13%

3.7 Explaining Biological Significance of the Selected Channels

In order to explore which channels are actually significant for the identifications, the number of occurrences of each channel from 4-channel combinations was counted from Tables 3.11 (1,000 data points), Tables 3.13 (1,500 data points) and Table 3.15 (3,000 data points). For example, with 1,000 data points, ch6 occurred 10 times, with 1,500 data points, ch6 occurred 13 times and 7 times with 3,000 data points.

Table 3.21 shows number of occurrences of each channel. The number of occurrences (frequency) is derived from Table 3.11, 3.13 and 3.15 with 1,000, 1,500, and 3,000 data point sample. For 1,000 data points, ch15 and ch16 occurred most frequently for 11 times, for 1,500 and 3,000 data points, ch15 occurred most frequently for 20 times and 18 times, respectively. Hence, it is obvious that ch15 is significant for identifications. Ch15 is at the position P_4 which is the parietal lobe of the brain.

The parietal lobe plays an essential role in integrating sensory information from various parts of the body, i.e. knowledge of numbers and their relations, and constructing the spatial coordinate system to represent the world around us. The principal function of position P_4 is perception (cognitive processing). The other functions are spatial relations, multi-modal interactions, praxis, and reasoning (non-verbal).

Table 3.21: The number of occurrences of each channel.

Channel	No. occurrences of 1,000 data points	No. occurrences of 1,500 data points	No. occurrences of 3,000 data points
ch 1	2	2	3
ch 2	2	5	-
ch 3	7	5	8
ch 4	3	-	-
ch 5	4	3	3
ch 6	10	13	7
ch 7	1	6	3
ch 8	3	2	-
ch 9	5	7	9
ch10	-	3	2
ch11	2	-	-
ch12	5	2	4
ch13	8	6	12
ch14	6	2	2
ch15	11	20	18
ch16	11	4	9

CHAPTER IV
CONCLUSION AND FUTURE WORK

In this study, a practical technique for identifying 16 standard EEG locations for EEG signals was proposed. The gathered signals were cleaned by applying Independent Component Analysis, then a supervised neural network was used to test the accuracy of the identification process. Our results concluded that a the group of four combinations is the minimum number of brain wave signals required for identification with high accuracy for insider and outsider subjects. The EEG signals 500 data points at locations F_7, C_3, P_3 and O_1 by ERICA algorithm can correctly identify only insider subjects but cannot correctly identify outsider subjects. The EEG signals 1,000 data points (at F_4, C_4, P_4, O_2), 1,500 data points (at F_8, F_3, C_3, P_4), and 3,000 data points (at F_{p1}, F_4, P_4, O_2) by SORIBO algorithm can correctly identify both insider and outsider subjects with high accuracy. The selected channels correspond to actual locations of the brain having the biological functions conforming to the identification.

In this study, data was collected from experimental subjects at the same period of time. One point of interest could be conceived within identification of this technique and whether it can be applied to the data collected from different time periods with more subjects. Such identification is proposed subject to further study within the realm of our proposed technique.

References

[1] A.K. Jain, A. Ross, and S. Prabhakar. An Introduction to Biometric Recognition. IEEE Trans. Circuits Syst. Video Technol 14, 1 (2004): 4-20.

[2] R.B. Paranjape, J. Mahovsky, L. Benedicenti, and Z. Koles, The Electroencephalogram as a Biometrics, Proc. Canadian Conf. Electrical and Computer Eng., vol. 2, pp. 1363-1366, 2001.

[3] M. Poulos, M. Rangoussi, N. Alexandris, and A. Evangelou. On the use of EEG features towards person identification via neural networks. Medical Informatics and the Internet in Medicine 26, 1 (2001): 35-48.

[4] M. Poulos, M. Rangoussi, N. Alexandris, and A. Evangelou, Person identification from the EEG using nonlinear signal classification. Methods of Information in Medicine 41, 1 (2002): 64-75.

[5] R. Palaniappan and K.V.R. Ravi, A New Method to Identify Individuals Using Signals from the Brain, Proc. Fourth Int'l Conf. Information Comm. and Signal Processing, pp. 15-18, 2003.

[6] R. Palaniappan, D.P. Mandic, Biometrics from Brain Electrical Activity: A Machine Learning Approach. Pattern Analysis and Machine Intelligence, IEEE Transactions 29, (2007): 738 - 742.

[7] R.Palaniappan, Method of identifying individuals using VEP signals and neural network. IEE Proc.-Sci. Meas. Technol 151, 1 (2004): 16-20.

[8] R. Palaniappan and D. P. Mandic. EEG Based Biometric Framework for Automatic Identity Verification. VLSI Signal Processing 2, 2 (2007): 243-250.

[9] S. Marcel and J. Millan, Person authentication using brain-waves (EEG) and maximum a posteriori model adaptation. IEEE Transactions on Pattern Analysis and Machine Intelligence 29, 4 (2007): 743-752.

[10] A. Cichocki, S. Amari, K. Siwek, T. Tanaka, et al. ICALAB toolboxes. Available online at http://www.bsp.brain.riken.jp/ICALAB

[11] S. Cruces, L. Castedo, A. Cichocki, Robust blind source separation algorithms using cumulants, Neurocomputing 49 (2002): 87-117.

[12] S. Cruces, L. Castedo, A. Cichocki, Novel Blind Source Separation Algorithms Using Cumulants, IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 5, pp. 3152-3155, 2000.

[13] Yeredor, A., Blind Separation of Gaussian Sources via Second-Order Statistics with Asymptotically Optimal Weighting, IEEE Signal Processing Letters 7, 7 (2000): 197-200.

[14] Doron, E. and Yeredor, A., Asymptotically Optimal Blind Separation Of Parametric Gaussian Sources - Lecture Notes in Computer Science (LNCS 3195): Independent Component analysis and Blind Source Separation, 5th International Conference on ICA, Granada, Spain, vol. 3195, pp. 390-397, 2004.

[15] J.-F. Cardoso and A. Souloumiac, Blind beam-forming for non Gaussian signals, IEE Proceedings-F, vol. 140, pp. 362-370, 1993.

[16] J.-F. Cardoso and A. Souloumiac, Jacobi angles for simultaneous diagonalization, SIAM Journal of Matrix Analysis and Applications 17, 1 (1996): 161-164.

[17] S. Cruces, A. Cichocki, S. Amari, Criteria for the Simultaneous Blind Extraction of Arbitrary Groups of Sources, The 3rd international conference on Independent Component Analysis and Blind Signal Separation. San Diego, California, USA, pp. 740-745, 2001.

[18] S. Cruces, A. Cichocki, S. Amari, The Minimum Entropy and Cumulant Based Contrast Functions for Blind Source Extraction, Lecture Notes in Computer Science, Springer-Verlag, vol. 2, pp. 786-793, 2001.

[19] S. Cruces, A. Cichocki, S. Amari, On a new blind signal extraction algorithm: different criteria and stability analysis. IEEE Signal Processing Letters 9 (2002): 233-236.

[20] A. Belouchrani, K. Abed-Meraim, J.F. Cardoso and E. Moulines, A blind source separation technique using second order statistics, IEEE Trans. on Signal Processing 45, 2 (1997): 434-444.

[21] A. Ziehe, K.-R. Mller, TDSEP - an efficient algorithm for blind separation using time structure, ICANN'98, pp. 675-680, 1998.

[22] A. Cichocki and A. Belouchrani, Sources separation of temporally correlated sources from noisy data using a bank of band-pass filters, Proc. of Third International Conference on Independent Component Analysis and Signal Separation (ICA-2001), pp. 173-178, 2001.

[23] R.R. Gharieb and A. Cichocki, Second order statistics based blind signal separation using a bank of subband filters, Journal of Digital Signal Processing 13 (2003): 252-274.

[24] A. Belouchrani, and A. Cichocki, Robust whitening procedure in blind source separation context, Electronics Letters 36, 24 (2000): 2050-2053.

[25] P. Tangkraingkij, C. Lursinsap, S. Sanguansintukul, T. Desudchit, Selecting Relevant EEG Signal Locations for Personal Identification Problem Using ICA and Neural Network, Eigth IEEE/ACIS International Conference on Computer and Information Science (icis 2009), pp.616-621, 2009.

[26] P. Tangkraingkij, C. Lursinsap, S. Sanguansintukul, T. Desudchit, Personal identification by EEG using ICA and neural network, Computational science and

its applications (ICCSA 2010), Lecture Notes in Computer Science vol. 6018, pp. 419-430, 2010.

[27] A. Cichocki, Blind Signal Processing Methods for Analyzing Multichannel Brain Signals, International Journal of Bioelectromagnetism 6, 1 (2004).

[28] A. Cichocki and S. Amari. Adaptive blind signal and image processing: learning algorithms and applications, New York, NY: Wiley, 2003.

[29] Ernst Niedermeyer, F. Lopes da Silva. Electroencephalography: Basic Principles, Clinical Applications. Page 139-141, Lippincott Williams, Wilkins, 2004.

[30] D.D. Daly , T.A. Pedley. Current Practice of Clinical Electroencephalography, 2nd ed. New York: Raven Press, 1990.

[31] W. van Drongelen, Signal Processing for Neuroscientists An introduction to the analysis of physiological signals. Academic press, 2007.

[32] S. Sanei and J.A. Chambers. EEG Signal Processing. Wiley, 2007.

[33] K. Delac, M. Grgic, A survey of biometric recognition methods, 46th International Symposium Electronics in Marine, ELMAR-2004, 16-18 June 2004, Zadar, Croatia.

[34] J. Malmivuo, R. Plonsey. Principles and Applications of Bioelectric and Biomagnetic Fields. Oxford university press, 1995.

[35] S.J. Luck. An Introduction to the Event-Ralated Potential Technique. The MIT Press, 2005.

[36] T.C. Handy. Event-Related Potentials A Methods Handbook. The MIT Press, 2005.

[37] V.K. Ingle and J.G. Proakis. Digital Signal Processing using MATLAB. Thomson, 2007.

[38] J. Alan Stuller. An Introduction to Signals and Systems. Thomson, 2008.

Appendices

Appendix A

List of Publications

Parts of this work are published in the following articles.

International Conference Proceedings

1. Preecha Tangkraingkij, Chidchanok Lursinsap, Siripun Sanguansintukul, Tayard Desudchit : Personal Identification by EEG Using ICA and Neural Network. ICCSA (3) 2010: 419-430

2. Preecha Tangkraingkij, Chidchanok Lursinsap, Siripun Sanguansintukul, Tayard Desudchit : Selecting Relevant EEG Signal Locations for Personal Identification Problem Using ICA and Neural Network. ACIS-ICIS 2009: 616-621

Biography

Mr. Preecha Tangkraingkij

PERSONAL DETAILS:

Date of Birth : December 7, 1965

Place of Birth: Bangkok, Thailand

EDUCATION:

- Ph.D. Program in Computer Science, Department of Mathematics and Computer Science, Chulalongkorn University, Thailand (Jun'07-May'12).
- M.S. Program in Computer Science, Nation Institute of Development Administration, Thailand (Mar'00-Dec'02).
- B.Eng. Program in Civil Engineering, Department of Civil Engineering, King Monkut's Institute of Technology Thonburi, Thailand (Jun'85-May'89).

SCHOLARSHIPS:

- Scholarship of the Centre of Excellence in Mathematics, the Commission on Higher Education, Thailand.
- Scholarship of the 90th Anniversary of Chulalongkorn University Fund (Ratchadaphiseksomphot Endowment Fund).