

**MULTI-MODAL ASSOCIATION LEARNING USING  
SPIKE-TIMING DEPENDENT PLASTICITY (STDP)**



**MASTER OF SCIENCE (INFORMATION TECHNOLOGY)  
UNIVERSITI UTARA MALAYSIA**

**2014**

## Permission to Use

In presenting this thesis in fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the Universiti Library may make it freely available for inspection. I further agree that permission for the copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to:

Dean of Awang Had Salleh Graduate School of Arts and Sciences

UUM College of Arts and Sciences

Universiti Utara Malaysia

06010 UUM Sintok

## Abstract

We propose an associative learning model that can integrate facial images with speech signals to target a subject in a reinforcement learning (RL) paradigm. Through this approach, the rules of learning will involve associating paired stimuli (stimulus–stimulus, i.e., face–speech), which is also known as *predictor-choice* pairs. Prior to a learning simulation, we extract the features of the biometrics used in the study. For facial features, we experiment by using two approaches: principal component analysis (PCA)-based Eigenfaces and singular value decomposition (SVD). For speech features, we use wavelet packet decomposition (WPD). The experiments show that the PCA-based Eigenfaces feature extraction approach produces better results than SVD. We implement the proposed learning model by using the Spike- Timing-Dependent Plasticity (STDP) algorithm, which depends on the time and rate of pre-post synaptic spikes. The key contribution of our study is the implementation of learning rules via STDP and firing rate in spatiotemporal neural networks based on the Izhikevich spiking model. In our learning, we implement learning for response group association by following the reward-modulated STDP in terms of RL, wherein the firing rate of the response groups determines the reward that will be given. We perform a number of experiments that use existing face samples from the Olivetti Research Laboratory (ORL) dataset, and speech samples from TIDigits. After several experiments and simulations are performed to recognize a subject, the results show that the proposed learning model can associate the predictor (face) with the choice (speech) at optimum performance rates of 77.26% and 82.66% for training and testing, respectively. We also perform learning by using real data, that is, an experiment is conducted on a sample of face–speech data, which have been collected in a manner similar to that of the initial data. The performance results are 79.11% and 77.33% for training and testing, respectively. Based on these results, the proposed learning model can produce high learning performance in terms of combining heterogeneous data (face–speech). This finding opens possibilities to expand RL in the field of biometric authentication.

**Keywords:** *spiking neural network, feature extraction, spike-timing-dependent plasticity, association learning, reinforcement learning.*

## Acknowledgement

‘Alhamdulillah’, praise be to Allah, The Most Beneficent, The Most Merciful.

I would not have been able to complete this journey without the aid and support of many people, to whom I am sincerely indebted and thankful.

First and foremost, my deepest thank you goes to my supervisor Dr Nooraini Yusoff, for without her support, guidance, and help this research would not have been successfully materialized. I cannot fully express my gratitude to the exceptional advice every time I seek enlightenment, the sharing of her knowledge from both theoretical and practical aspects, and her logical way of thinking have provided a good basis for this work. She guided me how to best synchronize and polish my ideas, and translate them into workable solutions.

I owe my most heartiest gratitude to my wonderful wife Enas Jassim and my gorgeous daughter Banan for their patience, understanding, unlimited support, and prayers for smoothness and blessed journey of my Master, without them, I could not pursue and accomplish this task so, thank you very much.

To my mother for her compassion and prayers, my father may Allah bless him, my sisters and brothers especially Dr Abbas Fadhil for his continuous support and encourage during the research. My parents in law Jassim Hadi and Aysha Majeed for their support and prayers which kept me survive along with my way.

My sincere appreciation to the most important people who have strongly contributed, in various ways, by providing the help and support needed: Prof Dr. Rahmat Budiarto, Prof. Dr. Jamal Al-Dabbagh, Ali Fadhil, Kareem Fadhil, Abdulaziz Fadhil, Maher Jassim, Samir Jassim, Ghaida Nazem, Dr. Nasser Ali, Dr. Ahmed Talib, Ammar Kadhum, Yassir Ahmed, Ishtar Khaldoun, Haitham Raed, Ali Adel, Mahdi Ahmed, Hayder Hashim, Mustafa Moosa, Marwah Saadi, Nawar Abbood, Hamza Sabry. Loai C. A. Alamro, and Enas Fadhil.

*Mohammed Fadhil Ibrahim*

# Table of Contents

Permission to Use .....	i
Abstract .....	ii
Acknowledgement .....	iii
Table of Contents .....	iv
List of Tables .....	vii
List of Figures .....	viii
List of Appendices .....	xi
<b>CHAPTER ONE INTRODUCTION .....</b>	<b>1</b>
1.1 Introduction .....	1
1.2 Problem Statement .....	3
1.3 Research Objectives .....	5
1.4 Scope of the Study .....	5
1.5 Significance of the Study .....	6
1.6 Organization of the Research.....	7
<b>CHAPTER TWO LITERATURE REVIEW.....</b>	<b>9</b>
2.1 Introduction .....	9
2.2 Biometric Technology Overview.....	9
2.2.1 Face Recognition.....	11
2.2.2 Speech Recognition.....	12
2.2.3 Fingerprint Recognition.....	13
2.2.4 Iris Recognition.....	13
2.2.5 Hand Geometry Recognition .....	14
2.2.6 Retina Recognition.....	14
2.2.7 Signature Recognition .....	15
2.2.8 Other Biometrics .....	16
2.3 Biometric Authentication and Classification.....	19
2.4 Multimodal Biometrics Authentication.....	23
2.4.1 Multimodal Biometric Systems Based on Fusion Technique.....	25
2.4.2 Biometric Authentication Based on Neural Networks Approach.....	29

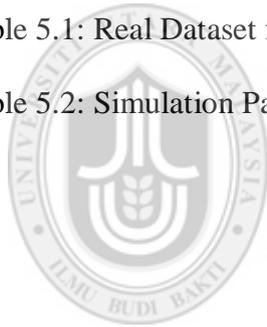
2.5 Face and Speech Recognition and Features Extraction .....	31
2.5.1 Face Features Extraction.....	32
2.5.2 Speech Feature Extraction .....	36
2.5.3 Face and Speech Recognition (Computational Intelligence Approach) ...	39
2.6 Face and Speech Authentication.....	41
2.7 Artificial Neural Networks (ANNs).....	43
2.8 Spiking Neuron Models.....	47
2.8.1 The Hodgkin-Huxley (HH) Model.....	49
2.8.2 Leaky Integrate-and-Fire (IF) Models.....	49
2.8.3 Izhikevich Spiking Neuron Model (IM).....	50
2.9 Reinforcement Learning (RL) .....	52
2.10 Summary.....	58
<b>CHAPTER THREE RESEARCH METHODOLOGY .....</b>	<b>61</b>
3.1 Research Methodology.....	61
3.2 The Learning and Classification Process .....	64
3.3 Phase I: Review of the Literatures .....	65
3.4 Phase II: Feature Extraction .....	65
3.4.1 Face Features Extraction.....	65
3.4.2 Speech Features Extraction.....	78
3.5 Phase III: Spike Encoding .....	83
3.5.1 The Rules of Synaptic Plasticity .....	85
3.5.2 Learning Strategy .....	87
3.6 Phase IV: Evaluation.....	89
<b>CHAPTER FOUR IMPLEMENTATION OF STDP AND FINDINGS.....</b>	<b>90</b>
4.1 Introduction .....	90
4.2 Preliminary Experiment .....	90
4.2.1 The Simulation Results.....	91
4.2.2 Inter-stimulus Interval .....	92
4.3 Multimodal Face-Speech Associative Learning .....	94
4.3.1 Feature Extraction .....	95
4.3.2 Face-Speech Training.....	95

4.3.3 Speech Encoding .....	97
4.3.4 Implementation of Learning (Face-Speech Features = 100, Number of Response Neurons = 100) .....	102
4.3.5 The learning Implantation (Number of Response Neurons = 200).....	105
4.4 Discussion.....	109
<b>CHAPTER FIVE LEARNING IN REAL WORLD.....</b>	<b>111</b>
5.1 Introduction .....	111
5.2 Data Collection .....	111
5.3 Face-Speech Feature extraction .....	112
5.4 Learning Implementation .....	112
<b>CHAPTER SIX DISCUSSION AND CONCLUSION .....</b>	<b>116</b>
6.1 Introduction .....	116
6.2 Conclusion (Objectives Achieved) .....	116
6.3 Future work.....	118



## List of Tables

Table 2.1: Biometric Comparison .....	18
Table 2.2: Biometrics Advantages and Disadvantages .....	18
Table 3.1: TIDigits Speakers' Numbers and Ages .....	79
Table 3.2: Description of Dialects and Distribution of Speakers .....	80
Table 4.1: Face-Speech Learning Samples .....	95
Table 4.2: The Initial Settings of the Association Learning .....	98
Table 4.3: The Face-Speech Learning Performance.....	98
Table 4.4: New Settings for the Learning Experiment Parameters .....	103
Table 4.5: Summary of the Learning Parameters and Performance .....	109
Table 5.1: Real Dataset for Training and Testing .....	112
Table 5.2: Simulation Parameters for Real Data Learning .....	113



**UUM**  
Universiti Utara Malaysia

## List of Figures

Figure 2.1: Types of Biometrics .....	16
Figure 2.2: Multimodal System Mechanism .....	24
Figure 2.3: Biometric System Based on Face and Iris .....	26
Figure 2.4: Score Level Fusion Approach .....	28
Figure 2.5: Multimodal Biometric System Using ANNs.....	31
Figure 2.6: Face Recognition Steps .....	33
Figure 2.7: Speech Recognition System Process.....	38
Figure 2.8: Face Recognition Based on Three ANN Classifiers .....	40
Figure 2.9: Thermal Face Recognition .....	41
Figure 2.10: Buffering Approach Biometric System.....	42
Figure 2.11: Neurons Network Example .....	48
Figure 2.12: Izhikevich Spiking Neuron Model .....	50
Figure 2.13: Associative Learning Experiment.....	53
Figure 2.14: Markov Decision Processes Representation.....	54
Figure 2.15: Description of Reinforcement Learning Process.....	58
Figure 3.1: Design Research Methodology.....	63
Figure 3.2: Process Flow of the Reinforcement Learning Using STDP for Multimodal Face-Speech Association Learning.....	64
Figure 3.3: The ORL Face Images Dataset .....	67
Figure 3.4: Example of Converting 2D Image into 1D Vector. ....	68
Figure 3.5: Face Image Normalization .....	69
Figure 3.6: Sample of Eigenfaces for ORL Data Set.....	72

Figure 3.7: The PCA Features Extraction Steps.....	73
Figure 3.8: Blocks Generation Process .....	75
Figure 3.9: SVD Dimensionality Reduction .....	77
Figure 3.10: Features Extraction Using SVD.....	77
Figure 3.11: Wavelet Packet Decomposition Tree at Three Levels .....	78
Figure 3.12: A Part of the Numeric Data for One Speech Sample.....	82
Figure 3.13: Wavelet Extracted Wave vs. Normal Speech Wave .....	83
Figure 3.14: Spiking Neural Network ( $N_E = 80\%$ , $N_I = 20\%$ ).....	84
Figure 3.15: Subpopulations of Neuron Stimulus .....	85
Figure 3.16: Face-Speech Association Learning Pseudo Code.....	89
Figure 4.1: : (A) Average Performance when the ISI Is within 10 m to 50 m. (B) The Firing Rate in the Target Response.....	92
Figure 4.2: The Performance of the Three Recall Correct Rates .....	94
Figure 4.3: The Learning Performance for the Face-Speech Model .....	99
Figure 4.4: Spike Raster Plot for the Network Activity after a Number of Trials in One Simulation .....	101
Figure 4.5: Winner State Details for Four Pairs Based on One Simulation.....	102
Figure 4.6: Spike Raster Plot for the Learning Simulation with 100 Features and the Number of Response Neurons=100.....	104
Figure 4.7: Winner State Details Based on the Number of Features= 100, Number of Response Neurons=100.....	105
Figure 4.8: Spike Raster Plot for the Network Activity with Number of Features=100, Number of Response Neurons=200 .....	106

Figure 4.9: Winner State Sample According to (Number of Features=100, Number of Response Neurons=200). ..... 107

Figure 4.10: Spike Raster Plot for the Network Activity (Number of Response Neurons=250) ..... 108

Figure 4.11: Winner State According for a Simulation with Response Group Neurons = 250 ..... 108

Figure 4.12: The Training and Testing Enhancement within Four Experiments ..... 110

Figure 5.1: Spike Raster Plot for the Network Activity during the Simulation of Real Data Experiment. .... 114

Figure 5.2: Averaged Percentage of Correct Recall with Real Data Experiment .... 115



## List of Appendices

Appendix A: Face and Speech Data File .....	144
Appendix B: Simulation and Learning .....	145



# CHAPTER ONE

## INTRODUCTION

### 1.1 Introduction

In general, human being depends on five senses to interact with the surrounding environment which are; sight, hearing, touching, smell, and taste. These senses enable the person to capture a huge amount of information to the brain. Then the brain analyzes, classifies, and recognizes this information in a way that is incredibly fast and accurate [1]. It is amazing for the brain to have such great capabilities to comprehend substantial physiological and behavioral biometric traits as well as to process the coming information in terms of human recognition. In computer systems, there are two methods that normally used to perform the authentication which are the traditional systems and the biometric systems.

Traditional person authentication approach can be knowledge-based like the password or PIN code, it also can be token based like an ATM card, credit card, and ID cards. This approach is less reliable and insufficient in terms of security performance [2, 3] because, it is difficult to differentiate between the genuine person and an imposter one. Furthermore, authentication elements like passwords or cards can be borrowed, stolen, and forgotten. That makes this approach suffers from a number of limitations which make it undesirable in terms person authentication [4].

Biometric identification approach is constraining on how to identify the individuals based on their physiological or behavioral characteristics. It based on what the person is, and what the person do. Biometric traits include fingerprint, iris,

gait, speech, palm geometry, face, facial thermo grams, retinal pattern, and signature. In contrast with the traditional approach, biometrics are difficult to be stolen, borrowed, or forgotten, that makes this approach is more sufficient and desirable in security performance [5]. The physiological and behavioral biometric traits can be, face, speech, gait, fingerprint, signature, hand geometry, ear shape, iris, retina, DNA, and so on.

Recently, the surveillance systems have been employed almost everywhere for security purposes such as, airports, banks, libraries, shopping, and many other aspects. Biometrics considered the most significant approach that used to verify and identify the individuals. Biometrics can identify the human being based on his or her physiological and behavioral traits[6]. This technology has taken the attention in the context of authentication systems.

Generally, there are two models that used in biometric authentication which are: Unimodal and Multimodal biometric system [2, 7]. Unimodal relies on single biometric trait to perform the authentication. However, multimodal authentication system relies on two or more biometrics to perform the authentication. Generally, multimodal system performs better than unimodal in terms of effectiveness and efficiency [2, 3, 8]. Since the unimodal depends on single biometric, thus, such models have low immunity against spoofing attacks, where spoofing one biometric is quite easier than spoofing two or more biometrics [9-12].

Learning the association features between two biometric traits is the approach that we adopt in our research. Among several approaches of learning the association, the reinforcement learning is one of the most realistic approach since it follows the

human behavior to learn about the environment [13]. Spike Neural Network (SNN) has proved its efficiency in the context of learning, due to its ability to simulate the human brain activities in terms of learning and training.

## 1.2 Problem Statement

Despite unimodal systems receiving significant enhancements in terms of accuracy and reliability, numerous challenges still face this approach and affect results negatively, such as noisy data, changeability, image illumination, intra-class variation, intra-class similarities, non-universality, and spoof attacks [2, 3, 8]. These limitations can be overcome by applying a multimodal authentication approach, which involves more than one biometric trait in the authentication process [9-12]. Hence, multimodal authentication systems are more reliable and desirable than unimodal systems. This approach helps overcome the limitations of unimodal systems and provides evidence presented by multiple human traits to improve recognition performance significantly, hinder spoof attacks, increase the degree of freedom, and minimize the identification failure rate. A variety of multimodal biometrics approaches have been recently adopted by researchers. These approaches include facial features and fingerprints [15], face and iris [16-18], face and ear [19, 20].

The commonly used recognition and classification techniques for biometrics can be statistical approach or computational intelligent approach (represented by Artificial Neural Network (ANN)). A statistical approach is typically less accurate and requires more resources than ANNs. [21, 22]. Accordingly, ANNs are generally

considered to be more accurate and effective than statistical approaches [23-26]. Brain-inspired SNNs represent third-generation of ANNs and are considered as a promising paradigm to generate new computational models [27]. SNNs can model complex information processes because of their ability to integrate and represent different information dimensions [28, 29], such as time, space, and frequency, as well as to deal with huge amounts of data in an adaptive and self-organized manner.

There are a number of algorithms that have been adopted in terms of performing the association learning such as the spike driven synaptic plasticity (SDSP) [30, 31], however, this algorithm considered as less efficient for fast on-line learning of complex spatio-temporal patterns [29]. In addition, SDSP in nature is a semi-supervised learning based technique where there should be a set of specific learning rules in order to perform the recognition, and this issue contrast with the reinforcement learning concept where the learning is performed as target-based behavior [32]. By using spike-timing-dependent plasticity (STDP) algorithm; SNNs can improve performance because of their behavior, which adopts the brain-like mechanisms in terms of associative learning based on RL. In spite of that the STDP has been employed in many learning aspects [25, 33, 34], however, there is a lack in terms of the application of this approach to multimodal contexts particularly when dealing with heterogeneous data samples.

In this study, we are proposing "Multi-modal Association Learning using Spike-Time Dependent Plasticity (STDP)". For an exploration of our study, there are key questions that we would like to pursue which are "What is the most suitable feature extraction method?", "How to develop SNN algorithm that can associate

face-speech biometrics and target the person?", and "How to evaluate the accuracy of proposed model?".

### **1.3 Research Objectives**

The main objective of this study is to apply Spike-time Dependent Plasticity (STDP) in multimodal association learning. Given the main objective, the specific study objectives are:

- a) To identify the most suitable face feature extraction method for optimal representation of the biometric input (i.e. face image).
- b) To utilize one of the existing speech feature extraction methods for optimal representation of the biometric input (i.e. speech signal).
- c) To develop SNN algorithm that can associate face-speech biometrics and target the person.
- d) To evaluate the accuracy of the face-speech association learning based on the percentage of current recall.

### **1.4 Scope of the Study**

This research focuses on the multimodal biometric systems based on face and speech traits. It is mainly concentrated on how to perform an associative learning using ANNs outlines. Among all ANNs algorithms, in this research we adopt STDP algorithm due to the efficiency and effectiveness of this algorithm.

For face biometric, this research adopts two face feature extraction techniques which are the Principal Component Analysis (PCA) and the Singular Value Decomposition (SVD) in order to test and analyze the performance for each technique, as well as determine the most significant technique in terms of extracting the facial features. All of the images that considered in our study are frontal images with different facial expressions. In addition, all face images that used in training and testing are gray scale level images.

For speech biometric, this research adopts the TIDigits speech data set [35] which was collected in 1984 for designing and evaluating the speech recognition systems. Thus this research focuses on the prerecorded speech samples. In the speech feature extraction phase, we adopt Wavelet Packet Decomposition (WPD) method to extract the speech features.

The classification process is the core of the proposed method, thus the accuracy of the sensors that used to capture the image face and the speech traits (i.e. The camera and the microphone) plays a key role of the overall system performance, hence the inaccurate results caused by the deficiency of such sensors is considered as out of this research scope.

### **1.5 Significance of the Study**

SNNs are considered as among the most appropriate techniques for biometric authentication because of their accuracy and speed [27]. SNNs can also simulate certain brain activities and learn biometrics. In addition, these algorithms can be used to process data with multiple dimensions, such as speed and time, as well as large

amounts of data[28, 29]; that makes this algorithm performs better in terms of learning. Adopting the reinforcement learning process in biometric authentication is a key point of this research, since this approach present high level of performance due to the behavior of this approach which adopts the human behavior in terms of learning. Hence, by implementing the integration of feature extracted (face-speech) and STDP based learning in spiking neural network; we come out with multimodal association learning using reinforcement approach that can be used in terms of person authentication, especially when dealing with face and speech biometrics which are easy to capture and implement. In addition, associative learning can be implemented with different types of biometrics in order to enhance the authentication performance.

The main contributions of this study are as follows:

1. Identification of suitable face feature extraction method for STDP-based learning in SNNs.
2. Application of multimodal associative learning that uses RL based on face–speech biometrics.

## 1.6 Organization of the Research

The following parts of our research are composed of five chapters which are represented by (*chapter 2 to chapter 6*) and can be described as follows:

**Chapter Two:** In this chapter, biometric types and identity authentication systems are discussed. In addition, the architecture of learning systems and methods used in

biometric learning is investigated. Feature extraction methods are also reviewed to select the method that will be applied in the present study. Learning methods are evaluated and implementing STDP to produce an optimal RL model is discussed in this chapter.

**Chapter Three:** In this chapter, the research design and the phases of the study are presented. Feature extraction, the learning network architecture, and the setting of the proposed learning are also described.

**Chapter Four:** In this chapter, the implementation of the proposed learning model in a series of learning simulations is discussed. The ability of the model to learn associative face–speech under different pair settings is also described.

**Chapter Five:** In this chapter, actual data experiment implementation and the obtained results are analyzed to evaluate the performance of the proposed model.

**Chapter Six:** This chapter summarizes the research conclusions. Potential future implementations of the proposed model are also suggested.

# **CHAPTER TWO**

## **LITERATURE REVIEW**

### **2.1 Introduction**

Biometric authentication is presented in several applications. Many research works have been presented in terms of biometric authentication systems [4]. In this chapter, we are going to view the most significant researches that deal biometric technology. We also will discuss the concept of biometric authentication systems and what are the pros and cons of such system. Then the multimodal systems will be viewed to explore the main difference between the performance of multimodal and unimodal biometrics. A comparison of the different approaches for the feature extraction and recognition process will be displayed. Furthermore, this chapter is describing the approaches and the technologies that were adopted in order to implement the learning of such systems and how the reinforcement learning can be employed in order to perform biometric learning. A review of the learning techniques and models will also be stated.

### **2.2 Biometric Technology Overview**

The word "Biometric" comes from the ancient Greek, "bio" means living of creatures and "metric" means the ability to measure [4, 7, 36]. Biometric systems basically are pattern recognition systems that can be operated by collecting biometric information for human being. Such systems beginning with capturing the biometric characteristics using a specific device (sensor) and then the features of the biometric

are extracted, then these features will be compared with a specific template and finally perform the recognition.

So why the biometrics?, Generally, the biometrics are used to describe the characteristics or particular process [36]. The characteristics are the measurable parts of the biometric which can be behavioral or physiological. The process on the other hand, is the method that used to recognize a person depending on his or her biometric characteristics. Traditional authentication systems normally based on:

- Something we know (knowledge-based), like passwords and Personal Identification Number (PIN).
- Something we have (token-based), such as, ID card, credit card.

Such authentication techniques normally come with a number of disadvantages, for example, they can be stolen, forgotten, or borrowed [4, 6, 37]. On the other hand, biometric traits are based on what we are, that means these biometric cannot be forgotten or stolen or borrowed. In addition, biometrics can be integrated with traditional systems to enforce the security [3, 4, 7].

When dealing with biometric systems, we must differentiate between the behavioral and physiological biometric traits. The physiological biometric is the stable characteristic which is not changeable with the time, such as, fingerprint, hand geometry, retina, iris, and face. The behavioral biometric on the other hand are those characteristics which may change over the time or specific conditions such as, speech, gait, keystroke, and signature.

The computer systems are very fast developing technology, consequently, the threats for such systems have also been increased, and hence, it is very important to develop a high security system to perform the user authentication. Biometrics play a critical key role in terms of high security performance. There are many types of biometrics that can be employed to develop such authentication systems. Each biometric has its own characteristics which may or may not be suitable for specific systems. In the coming sections, we will give a description of the most common biometrics as well each their specification.

### **2.2.1 Face Recognition**

This technology relies on the identification of particular person depending on the image of the face [38]. The characteristics which commonly considered in this kind of biometric are:

- The shape of the eyes, mouth, and nose.
- The distance between facial biometrics.
- The expression of the face.

In order to capture the face image, any camera can be used. There are many approaches that have been adopted to analyze the facial image and perform the recognition such as, Principal Component Analysis (PCA) [39], Local Feature Analysis (LFA) [40, 41], Elastic Graph Theory (EGT) [42], Artificial Neural Networks (ANNs) [43].

Face authentication is considered to be one of the most acceptable biometric authentication technique [44]. The users normally do not have big concerns of such authentication because it is not required for direct interaction with the sensor (camera). In addition, the camera is a low cost device that makes it easy to deploy and implement due to its availability. Furthermore, Face image can be captured remotely (i.e. without the user consent), which makes the face recognition on the top of biometric authentication techniques which employed in surveillance systems.

### **2.2.2 Speech Recognition**

It is the biometric that uses the person voice features in order to perform the recognition. It is a technology that enables the machine to identify the speaker's words. The recognition of the speech involves the analysis process for the pitch, cadence, tone, and the frequency of the speaker voice [36]. A number of approaches used to perform speech recognition such as, Pattern Matching Algorithm [45], Hidden Markov Models (HMM) [46], Artificial Neural Networks (ANNs) [47].

Similar to face recognition, speech recognition does not need for direct interaction with the users. Also, it can be captured remotely and without pre knowledge from the user side. Speech characteristics can be captured using a lower cost device which is the microphone, this device normally embedded with computers and mobile devices. The thing that makes speech recognition very suitable to use in such authentication systems [1]. Speech recognition is not based on the spoken word, but it is based on voice print. The voice print represents the characteristics of speech features.

### **2.2.3 Fingerprint Recognition**

Fingerprints represent the pattern of fingerprint biometric. The minutiae features of the fingerprint are: whorls, loops, arches, and ridges. All these features can be extracted from the image of the fingerprint. Many approaches have been adopted to perform the fingerprint recognition. The main benefit of using finger print biometric is low-error rate for such biometric. However, there are a number of disadvantages combined with fingerprint authentication systems such as, some people have no distinctive fingerprint, the high sensitivity for dry and wet fingers, and the latent oily image which stays from previous image for another user may cause problems [48]. Moreover, there are legal issues that hinder the deployment of fingerprint development which is some people do not like to for their fingerprint to be documented.

### **2.2.4 Iris Recognition**

Iris is the colored area that surrounded the pupil of the eye. It is considered to be one of the most important biometric which can uniquely identify the human being. The iris feature analysis involves the ring, freckles, and furrows in the colored ring of the eye [1, 36]. In spite of iris recognition produces a significant outcome which can be at a high level of accuracy, however, it requires the user to keep looking at the sensor for a while to extract the feature as well as it is need to be very close to the sensor, which make people have a big concern about applying the light directly to their eyes, thus, iris biometric has a low level of acceptance among users and it became insufficient in terms of surveillance systems.

### **2.2.5 Hand Geometry Recognition**

This kind of biometric measures the hand features such as, joints, shape, and palm size. It is one of the simplest biometrics. The advantages of such biometric are that it can be implemented in a wide range of applications. On the other hand, hand geometry normally not of the distinctive kind, hence, it is not efficient when used in a large number of the population [49]. In addition, the palm or hand geometry is prone to change through aging especially for children.

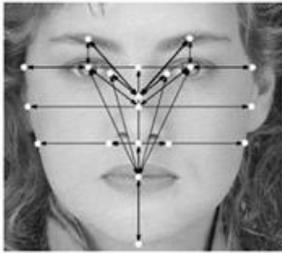
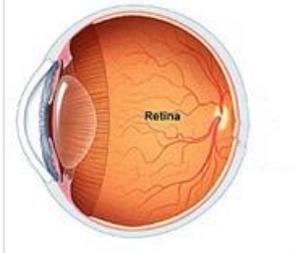
### **2.2.6 Retina Recognition**

Retina biometric is represented by the blood vessels in the white area which is located on the back of the eye. These blood vessels absorb the light easily compared with the surrounding tissues [36]. Scanning the retina can be performed by applying low-energy infrared ray on the person's eye as he or she looks to the scanner. The information that involved in the blood vessels is difficult to spoof due to the difficulties of faking the retina pattern. Retina biometric has a high level of accuracy, however, the person who wants to be recognized must position his or her eye very close to the sensor within half inch distance, also, the person is required to keep constraining on the sensor without any movement [1]. Furthermore, this technique does not work properly if the individual wearing eyeglasses. All these issues can make the use of such biometric limited to few authentication purposes, such as nuclear or high security government organization [49].

### 2.2.7 Signature Recognition

In many cases, signature can indicate the uniqueness of the person. The features that used to perform signature recognition are the pressure, speed, and the overall signature shape [48]. Other features can be taken into account regarding to signature biometric such as directions and length of the strokes. Signature is one of the behavioral biometrics, which means it is prone to be changed, also, it can be affected by the emotional and physical conditions [1]. Signature features can be captured using special pens or tablet device, which can capture 2D of features like shape and pressure [4, 8, 50]. However, these devices have two main drawbacks:

- The signature that resulted from these devices looks different from the original one.
- During signing process, the person cannot see what he or she is writing because he or she must look at the device's monitor to watch the signature, the thing that effects on the result of this biometric type and may give inaccurate performance. Figure 2.1 shows samples of biometric.

		
<i>Face Recognition</i>	<i>Fingerprint Recognition</i>	<i>Iris Recognition</i>
		
<i>Hand Geometry Recognition</i>	<i>Retina Recognition</i>	<i>Signature Recognition</i>

*Figure 2.1: Types of Biometrics*

### 2.2.8 Other Biometrics

There are some other biometric technologies that use a variety of behavioral and physiological traits. Some of these biometrics are available in commercial aspects, and some are still in the emerging period. The following are the most well-known biometrics:

- DNA matching.
- Thermal imaging.
- Ear shape.
- Human gait.
- Body odor.

- Blood pulses.
- Vein scan.

Commonly, there are many biometrics that can be selected to perform user recognition. However, none of them can be the best biometric. Each biometric has its own advantages and disadvantages. The selection of a particular biometric depends on the authentication system type as well as the user requirements. Some of the systems need a high level of security like nuclear fields and some of them need only for convenient level of accuracy like attendance systems.

A general comparison can be made for different types of biometrics according to seven categories [49, 51], which are:

- **Universality:** This means how the biometric characteristics appear for each person.
- **Uniqueness:** means how the biometrics can be adequately good to recognize an individual from the other one.
- **Permanence:** it measures the ability for the biometric in terms of aging resistance.
- **Collectability:** it measures how much easier to capture the biometric to perform the authentication process.
- **Performance:** explains the robustness, the speed, and the accuracy of the biometric system.
- **Acceptability:** shows the level of the reliability for the biometric technology.
- **Circumvention:** indicates how much easier to fake a specific biometric.

Table 2.1 shows a comparison for the biometrics according to the seven categories mentioned above.

*Table 2.1: Biometric Comparison (Reproduced from [49])*

Biometrics	Universality	Uniqueness	Permanence	Collectability	Performance	Acceptability	Circumvention
Face	H <sup>(*)</sup>	L	M	H	L	H	L
Speech	M	L	L	M	L	H	L
Hand	M	M	M	H	M	M	M
Iris	H	H	H	M	H	L	H
Signature	L	L	L	H	L	H	L
Fingerprint	M	H	H	M	H	M	H
Retina	H	H	M	L	H	L	H

To describe the main advantages and disadvantages of biometrics, Table 2.2 shows in details the biometric comparison.

*Table 2.2: Biometrics Advantages and Disadvantages*

Biometric	Advantages	Disadvantages
Face	<ul style="list-style-type: none"> <li>• High Flexibility.</li> <li>• Can be captured remotely.</li> <li>• Short time required.</li> <li>• Low cost sensor.</li> </ul>	<ul style="list-style-type: none"> <li>• Influenced by aging.</li> <li>• High sensitivity to image conditions like illumination.</li> </ul>
Speech	<ul style="list-style-type: none"> <li>• Not required to the user consenting.</li> <li>• No direct interaction with the sensor.</li> <li>• Low cost</li> </ul>	<ul style="list-style-type: none"> <li>• Easy to fake using pre-recorded voice.</li> <li>• Influenced by noise and illness.</li> </ul>
Hand	<ul style="list-style-type: none"> <li>• No sensitivity to noise.</li> <li>• Easy to use</li> </ul>	<ul style="list-style-type: none"> <li>• Less accuracy.</li> <li>• Big size scanner.</li> <li>• Causing hand injury.</li> </ul>
Iris	<ul style="list-style-type: none"> <li>• High accuracy</li> </ul>	<ul style="list-style-type: none"> <li>• Time consuming.</li> <li>• Not suitable for children.</li> </ul>
Signature	<ul style="list-style-type: none"> <li>• Difficult to mimic the signature style.</li> <li>• Easy to use.</li> </ul>	<ul style="list-style-type: none"> <li>• Less accuracy.</li> <li>• Changeable.</li> </ul>
Retina	<ul style="list-style-type: none"> <li>• Difficult to fake.</li> <li>• High accuracy.</li> <li>• Aging resistance.</li> </ul>	<ul style="list-style-type: none"> <li>• Time consuming.</li> <li>• Difficult to implement.</li> </ul>

\*: H: High, M: Medium, L: Low

### 2.3 Biometric Authentication and Classification

The fast growing of computer, networking, communications, and mobile technology requires a reliable technique to identify and verify the persons. Biometrics represents a method to uniquely recognize the individuals. Person biometric authentication refers to the technology which can measure and analyze the physiological and behavioral human traits [36], such as face, speech, iris, retina, palm geometry, fingerprint, gait, and so on. Many studies have been conducted in biometrics authentication approach.

Among all the proposed methods, ANNs have proved their efficiency in terms of performance, time consuming, and accuracy [52-55]. Person authentication using the ANN approach can give better results than statistical approaches by employing a good technique, also it can rise up the accuracy of the recognition up to 99.25% [23]. ANNs are used to build a better classification and consequently the authentication system will be more effective in terms of classification [26].

Recently, many systems have adopted the term brain-like to describe the new generation of ANNs. This approach tries to manipulate the data in a way that is similar to human brain works. Thus, there is a big interest about building intelligent systems especially in the biometric field [25]. The brain-like model tries to adopt the brain network structure to manage the connectivity. The main idea for such systems is that the neurons can implement pattern recognition by adopting spike timing processing. The concept is that the neurons can process and exchange the data at spike level [56]. Spiking Neural Networks (SNNs) are the method that the neuroscientists used to study the activity of single or group of neurons.

SNNs are considered to be the third generation of neural network. This new trend has the ability of modeling complex information process because of its ability to integrate and represent a variety of information diminutions such as space, time, and frequency, as well as dealing with large amount of data in self-organizing and adapting manner. Deep machine learning is a new trend which is currently emerging, can be implemented using SNNs which is adopting the brain learning behavior [27].

Spiking Time Dependent Plasticity (STDP) is one of SNNs algorithms, the process of the STDP is that the neurons listen to the incoming spike trains, once a particular neuron fires, it strongly prevents the other neurons form take the same activity, this means prevent the other neurons form learn the same pattern [24]. Accordingly, the neurons will be self-organized by covering different patterns in a powerful distributed scheme. This represents how the brain can encode and decode the information easily. STDP plays a key role through detecting the repetitive patterns and make a response to them. It is currently a significantly well-established physiological mechanism of activity driven synaptic regulation. It has also been shown to do a better job than more conventional reorganization. Hence, adopting STDP mechanism can make the authentication systems more robust and effective.

In [38, 49], the authors have investigated the performance and security concerns of biometric technique. They stated the challenges of these systems as well as the rapid development for such systems. The study has also presented a comparison among biometric technologies according to seven categories, Table 2.1. According to this study, there are several attacks that can threat biometric systems

and influence their performance. The result of this study is that the biometrics still needs to be enhanced to reach the user acceptance level.

In [4], the researchers tried to present the importance and the growth of the biometric systems, they gave a detailed overview about the development as well as the technologies for these systems. In addition, they listed the biometric technologies with specific details of the analysis techniques that used with each biometric. The results of the study were illustrated that the biometric based authentication systems are still far from the perfect solution but at the same time it represent a high level of performance and efficiency in comparison with other authenticated systems.

In [7], the study provided an overview of biometric authentication systems, advantages and disadvantages, strengths and limitations. Classification methods have also stated in the study, the integration scenarios, and fusion techniques. The study has presented considerations of selecting a particular biometric characteristic according to the biometric specification and user requirements. Biometric challenges and limitations also addressed as well as the ways to face such attacks that threats the privacy issues.

A generic framework has been presented in [57], to analyze the security template and the privacy in biometric authentication systems. The study analyzes the weaknesses of authentication protocols in terms of user and data privacy. In biometric systems there are a number of problems that need to be solved. Employing a stronger security protocols and applying an encryption techniques to the features level can help to enhance the biometric systems security.

Biometric systems technical issues and challenges has been discussed in [58]. The performance of the biometric systems can be measured according to the accuracy of system recognition which involves two metrics False Accept Rate (FAR) and False Reject Rate (FRR). The study stated that the ideal biometric authentication system must meet the user requirement as well as reaches the level of universality. The acceptance of any biometric system can be determined according to the security level and privacy performance [6, 59, 60].

Many threats that faces the development of biometric systems [61]. In order to design a secure biometric system, it is important to figure out and evaluate the well-known threats. Spoof attack is one of the most famous threats that faces he biometric systems. Many scenarios have proved that the spoof attacks can crack the biometric systems. In this study, the author has proposed a method to evaluate the biometric system against spoofing attacks. Two models of match score have been presented, these models employ the information of genuine and imposter samples that have been collected for training of the biometric system.

Some biometric issues, technologies and challenges have been addressed in [62], the author has stated aspects where the biometric systems can be used such as, healthcare, banking, finance, energy access control, military, passports, airports, and so on. A comparison among the types of biometrics has been made. The study outlines the most common biometric and the consideration of selecting a particular biometric according to the efficiency and the performance. FAR and FRR are the main metrics that used to evaluate the performance of such systems. The author has come up with a conclusion that the face recognition is the leading biometric in large

population surveillance systems due to the high level of performance and the flexibility of this kind of biometric. In addition, the challenges that face the biometric systems can be overcome with the rapidly growing and development of pattern recognition.

#### **2.4 Multimodal Biometrics Authentication**

Multimodal systems are taking too much interest and acceptance among authentication systems developers and stakeholders. These systems normally rely on two or more biometrics to achieve the recognition process, therefore, such systems have proved their ability to overcome unimodal limitations for example; it is quite difficult to spoof two biometrics at the same time. Many studies have proved that multimodal biometric systems present better performance than unimodal in terms of accuracy and performance [5].

Multimodal biometric systems represent a method that can be used to promote the security performance and overcome the limitations of unimodal systems by having the evidence from multiple biometric sources such as face and speech [7]. Since the multimodal systems are difficult to be spoofed, then the possibility of getting a high accuracy decision system will be high as well. In other words, accepting an imposter as genuine rate (FAR) and rejecting an authorized person rate (FRR) will be decreased. Many studies have been presented to explore the multimodal systems as well as the issues and the consideration related to such systems such as [1, 6, 16, 19, 60, 63].

A performance analysis of multimodal systems has been presented in [64]. The authors have developed a multimodal system by combining face, iris, and fingerprint biometric traits. They also presented a diagram which represents a mechanism of multimodal system, Figure 2.2. The study proved that the performance of multimodal biometric systems is achieving better in terms of performance.

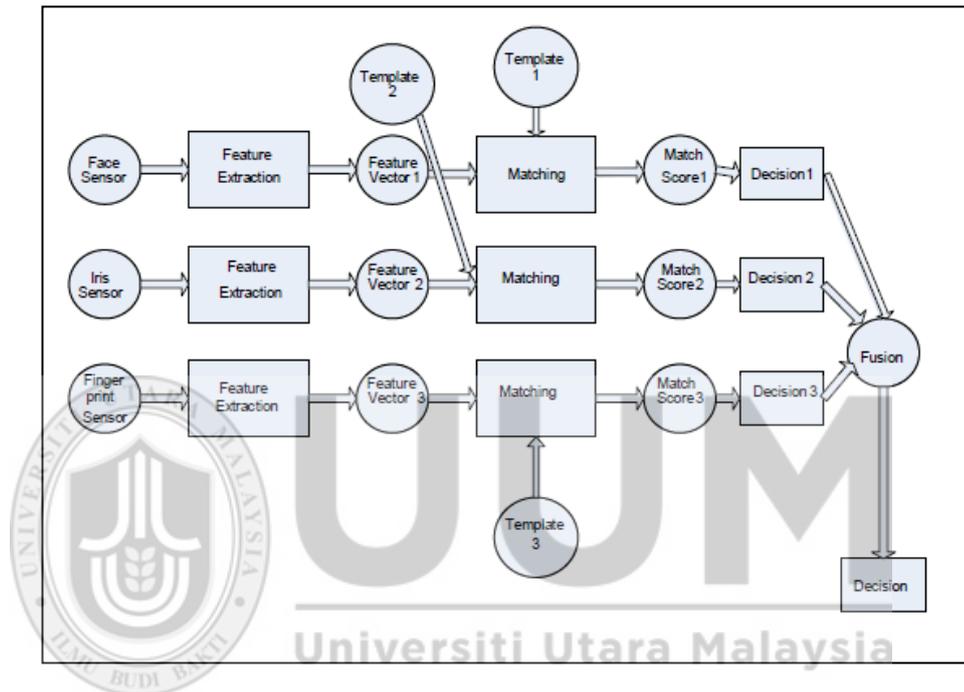


Figure 2.2: Multimodal System Mechanism (Reproduced from [64])

Multimodal biometric systems have many advantages in comparison with unimodal. Combining multi biometric traits captured from different sensors and by employing an efficient fusion scheme can enhance the overall accuracy significantly [3]. These systems provide a high resistance against the spoofing attacks. Recently biometric authentication has got a considerable enhancement in the accuracy and reliability, however, multimodal biometric can enhance the performance accuracy, deterring spoof attacks, rise up the degree of freedom, and decrease the failure rate. The key to the high performance of multimodal systems is the effective fusion

scheme which is responsible to combine the presented biometric traits [37]. Multimodal biometric systems are getting the huge acceptance among the designer because of the high performance and the accuracy which surpasses the performance of the unimodal systems. In addition, employing a multi-modality system such as (face and fingerprint) can overcome the limitation of a single modality (unimodal) system [2].

#### **2.4.1 Multimodal Biometric Systems Based on Fusion Technique**

A person authentication systems have been presented in [16], based on face and iris features. Feature level of fusion is used in this study and the result showed that this method of fusion can outperform the other fusion techniques. To improve the performance of face biometric recognition systems, face and iris recognition system has been presented in [17], the result of the study stated that by removing the redundant features for face and iris biometrics can give the optimal recognition performance. In [19], a multimodal face and ear based system have been presented. The system used Principal Component Analysis (PCA) approach to extract the biometric features, the results have shown that the performance of this system is present better outcomes that using a face or ear individually. The accuracy of the proposed system was 92.24% with FAR of 10% and FRR of 6.1%.

A study of a face and palm based authentication system was proposed in [65], in this study, the authors have presented a different fusion scheme to address the problem of high dimensional feature space. They adopted two fusion levels which are; the match score level and feature score level. The study compares a data set of

250 virtual people. The results of the study showed a significant level of improvement in a ratio of 6% in comparison with the performance of feature fusion using Log-Gabor method. In addition, the study stated that the hybrid (multimodal) systems give better performance than the unimodal.

In [66], a development of score-level fusion algorithms based on face and fingerprint recognition. The results of the study show that the proposed methods present better performance than that if the face or fingerprint recognition system are designed individually. To overcome the limitation of unimodal systems, a multimodal biometric system based on the face iris recognition by using Support Vector Machine (SVM) in has presented in [18]. The results show that system which are based on multiple biometric traits are performing better and can surpass the unimodal biometric systems as shown in Figure 2.3.

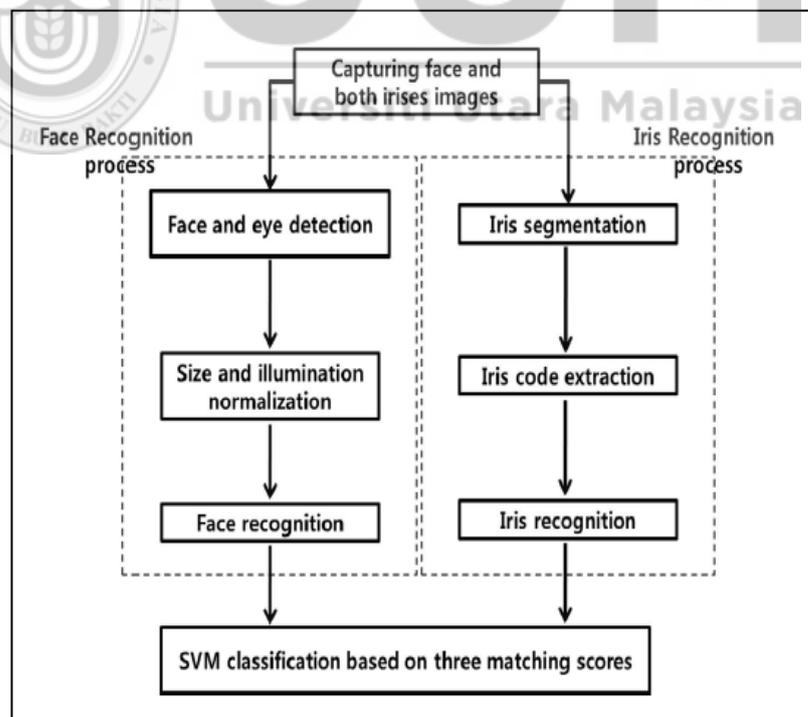


Figure 2.3: Biometric System Based on Face and Iris (Reproduced From [18])

A novel study of the multimodal feature extraction and user recognition has been presented in [67], the study relies on subclass discrimination analysis (SDA) approach. The authors proposed two ways to overcome the problem of singularity which caused by calculations. The study adopts two types of biometric which are face and palm print traits. The results of the study show that the multimodal authentication is better than unimodal in terms of recognition performance.

A multimodal biometric authentication system based on face and signed biometric has been presented in [68]. The system combines two types of biometrics to perform high significant level of recognition. The study adopts match-score level of fusion due to the easiness to combine and access the scores presented by different sources. The study has been conducted on a sample of 40 users as a data base to prove that the max-of-score fusion method gives better and more significant level of authentication performance if compared with the unimodal.

An evaluation of the score fusion rules under multimodal biometric systems and the efficiency against spoof attacks, a study was presented in [69], the study was based on face and fingerprint biometrics, it shows the differences between single biometric systems and the multimodal. It states the efficiency and the robustness of multimodal systems against spoof attacks. The study aims to evaluate several levels of fusion rules. Accordingly the study gives the designer a clear comparison and opens the way to select the most efficient technique. A ranking of eight score fusion rules has been reported in the study. This ranking has been made according to the values of FAR and FRR metrics. The results present evidence which is that the

ranking of multi fusion rules can strongly provide a way to select the most appropriate fusion to face the spoofing attacks.

The fusion can be done at various levels in terms of multimodal biometric systems. According to [44], the most common fusion approach is matching score level. In this study, face and signature traits have been selected to conduct the experiment. A database of 17 users is selected, and the study proved that the fusion of multiple traits can perform better in terms of biometric recognition than that using unimodal. Figure 2.4 shows the experiment's setup. The results also show that the accuracy of multimodal system outperforms the unimodal in about 10%. In addition, this rate can be improved by adopting more sufficient pattern recognition technique.

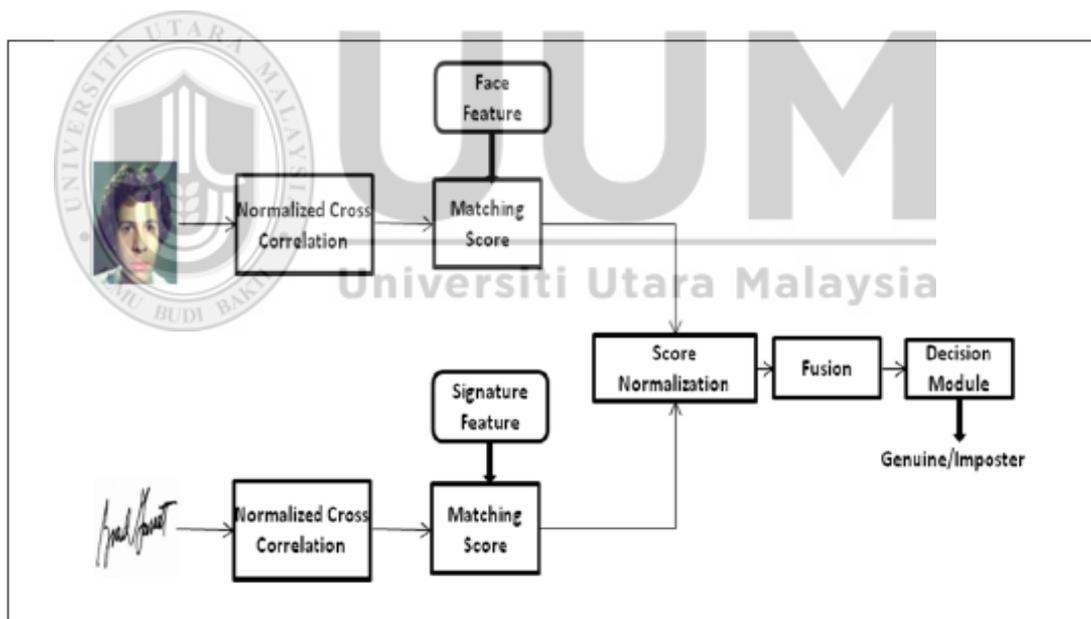


Figure 2.4: Score Level Fusion Approach (Reproduced from [44])

To sum up, from all previous literatures we can see that the multimodal biometric systems are present more efficient and more effective performance in comparing with unimodal. Multimodal systems are more robust in terms of facing

many threats that the biometric systems may face, which means it is quite difficult to fake more than one biometric at the same time.

Despite the fusion approaches have produced an easy implementation, there are some difficulties combined with such approach, such as it is difficult to come up with good fusion when the biometrics are quite different in terms of data like speech and face [21, 70]. That makes this approach is limited to specific multimodal systems.

#### **2.4.2 Biometric Authentication Based on Neural Networks Approach**

The effectiveness of personal authentication systems can be determined according to the fast and accurate recognition. The rapid expansion and the advancement of Artificial Neural Network (ANN) make the recognition process become faster and more accurate as well as evolving the learning capabilities.

A neural network technique has been adopted in [23], to design a personal authentication system based on iris traits. To locate inner and outer boundaries of the iris, a fast algorithm has proposed. The classification of the iris pattern is performed by employing neural networks. The effectiveness and the robustness of adopting neural networks in authentication systems can be noticed clearly according to the learning and training strategies on ANN. The results of the study showed that the recognition accuracy was at 99.25%.

A comparison among the classification techniques which are employed in biometric systems have been done in [26]. Biometric recognition plays a key role in

our daily life in terms of security and privacy. Several techniques have been adopted to develop biometric authentication systems as well as to perform a pattern recognition process. Among these techniques, neural networks have got a big interest according to the highest level of accuracy and the effectiveness as well. In the context of user authentication systems, there are two important issues should be considered, which are; accepting the genuine user and rejecting the imposter one. The decision about these two matters can be done in the classification step of the system process. Neural networks have proven their efficiency in terms of the classification process.

The precision and the performance have got the extreme attention in the world of biometric authentication [22]. This study produces a new approach to the field of biometric pattern recognition by adopting Chaotic Neural Networks (CNN). This method facilitates the learning of data pattern as well as train the classifier to perform the biometric authentication. The problem of high dimensionality can be overcome by employing a classifying method for the extracted features. Figure 2.5 shows the proposed system. By combining multiple biometric traits as well as the neural network techniques in authentication systems, the accuracy and performance will be at a higher level than using other techniques.

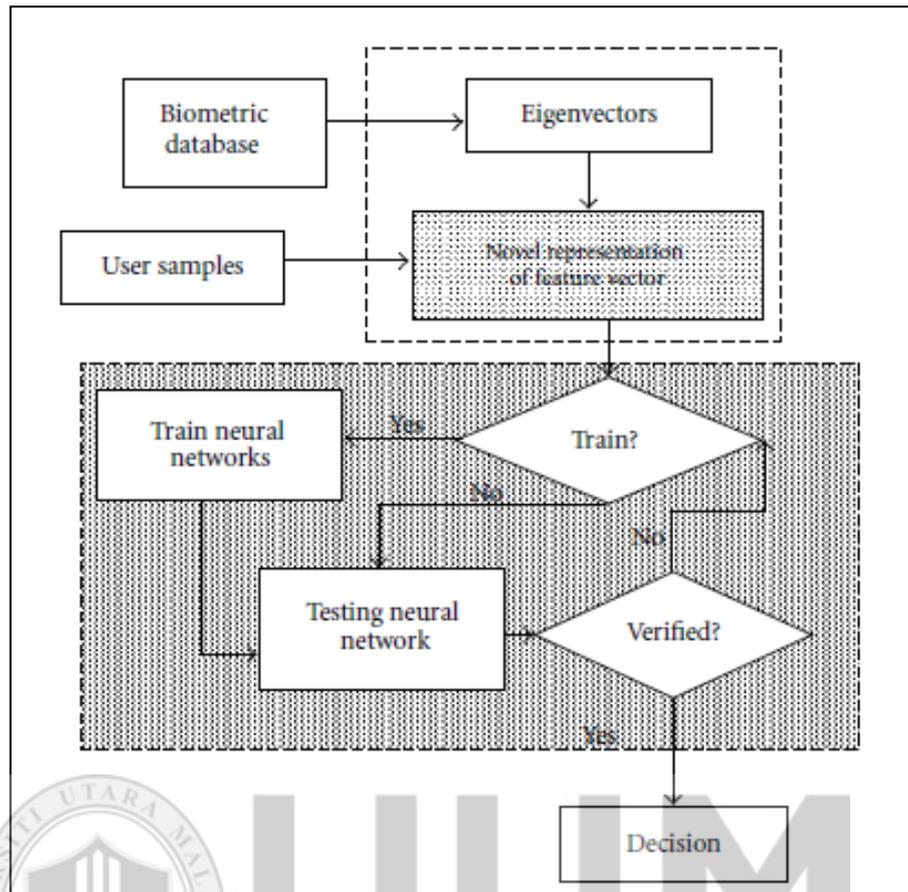


Figure 2.5: Multimodal Biometric System Using ANNs (Reproduced from [43])

Face and fingerprint authentication system based on neural network has been presented in [15]. The main objective of this study is to reduce the error rate and enhance the performance. The proposed system can recognize the individual faster, the recognition accuracy increased, and the system produced a better performance.

## 2.5 Face and Speech Recognition and Features Extraction

Since the multimodal biometric system depends on extracting the most dominant features from multiple biometrics; the critical step of such process is the way that used to combine the feature in order to make the decision. Feature extraction is the process of extracting the main characteristics to implement the

recognition. The main aim of feature extraction is to reduce the amount of features in an effective way to perform the discrimination. The performance of any system relies on the discrimination efficiency as well as the robustness toward features degradation. Hence, selecting the appropriate feature extraction method plays a key role in authentication performance [7].

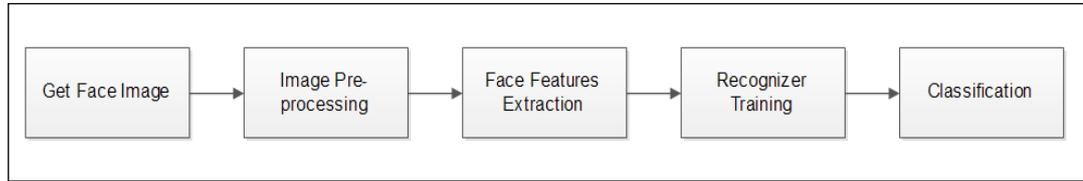
For face biometric, there are several techniques which used to extract the features such as, Principal Component Analysis (PCA) [71], Singular Value Decomposition (SVD) [67, 72], Gabor Wavelet, and Artificial Neural Networks (ANNs) [73]. For speech biometric, there also are several approaches such as, Mel-Frequency Cepstral Coefficient (MFCC) [74], Wavelet Packet Decomposition (WPD) [75, 76], Hidden Markov Models (HMMs) [46],

### **2.5.1 Face Features Extraction**

There is no doubt that human beings can recognize faces in the age of five years or earlier [39]. It looks like an automated process in our brain, also we can recognize people we know even though they are wearing glasses or hats, or even they have beard or long hair. In addition, our brains can recognize people when they get older. For human brain and its incredible capability of processing; the task of recognizing human faces seems to be trivial, but for computers, it is really challenging to perform such recognition task.

Since the efficiency and effectiveness of any classification process strongly depend on the size of the trained data set as well as the quality of the extracted features [67], the features extraction and dimensionality reduction become more

important and more challenging process. Figure 2.6 shows the steps of face recognition systems.



*Figure 2.6: Face Recognition Steps*

Despite the similarities of human faces, human beings are able to recognize each other according to the uniqueness in the facial characteristics; these characteristics can be the eyes (distance between eyes and the eyes shape), mouth, nose, jaw edge, and cheekbones. The recognition of the face is based these characteristics. The task of face recognition can be hindered by a number of challenges such as, illumination, face expression, and the accuracy of the sensor [4]. On the other hand and in comparison with other biometrics like iris, fingerprint, retina, and so on, the face recognition is the most acceptable biometric in the field of human surveillance [1]. Moreover, there are a number of factors that can make the face recognition technique on the top authentication technologies such as:

- Low cost hardware: face recognition can be performed by getting the face image using a camera which is quite suitable if compared with other sensor devices like an iris.
- The capture of the face can be done remotely, that is mean there is no direct interaction between the user and the sensor (camera) [20].

- Since a camera can capture the human face remotely and sometimes without user consent, face recognition is more suitable for surveillance systems.

PCA is one of the most dominant data analysis technique which aims to reduce the dimensionality and overcome the curse of dimensionality [71, 77, 78]. PCA plays a key role in the context of feature extraction through the ability of removing the noise as well as specifying the redundant information by extracting the most dominant features from the original dataset [79-81].

A comparison study of face recognition methods has been made in [82], in addition, face feature extraction has also been addressed such as, Gabor wavelet transform and Principal Component Analysis (PCA) extraction techniques. The study has also discussed the pros and cons of feature extraction techniques, the performance and limitation. The evaluation of face recognition algorithms can help significantly to extract the best facial features and consequently can help to increase the overall system performance. PCA approach has performed well when the image organized and tested first. Eigenfaces has produced a better solution which is suitable for face recognition. Gabor wavelet performs the face recognition in high level of stability. All mentioned comparisons have discussed in details in this study to provide good information in the way of increasing the feature extraction performance.

A face recognition novel approach has been presented in [55]. The authors have presented an efficient face recognition approach by applying different vector and Kernel Principal Component Analysis (KPCA). The feature extraction plays a

vital role in face recognition process. One of the common methods of feature extraction is PCA which proved its efficiency and effectiveness in terms of facial features extraction [83-85].

In [86], the authors proposed a method to overcome the limitation conditions which is a Selective Illumination Enhancement Technique (SIET) for face feature extraction. They also proposed Threshold Discrete Wavelet Transform (TDWT) to improve the performance of face feature extraction. The feature extraction stage individually examined to enhance each stage. To select the feature vector, a Binary Practical Swarm Optimization (BPSO) algorithm employed. The performance of feature extraction normally influenced by a variation of the image conditions like illumination. The result of the study has presented that the proposed method presents an efficient method for face feature extraction and better performance produced throughout the reduction of features extracted.

A color based feature extraction technique has been presented in [87-89]. After presenting the color pixels of skin region of the face, the obtained image statistics will prone to binarization, it will be transformed to gray scale image. The purpose of this step is to eliminate the hue and saturation and to focus only on illumination. The latter then will be transformed into a binary image because; the face image features are darker than the background colors that used for feature extraction based on PCA. Since PCA has a number of limitations, Template Based Technique (TBT) contains several algorithms to extract facial features such as, template based mouth and eye detection. The method does not need a complex mathematical calculations as well as pre-knowledge about the features. This

technique is considered as an easy method to perform and implement face feature extraction process.

SVD is one of the common feature extraction methods which is used with pattern recognition systems. It is a mathematical method used to define and order the matrix's dimensions. In addition, SVD is considered as one of the most efficient tools for data analysis and signal processing [90, 91]. In terms of pattern recognition, SVD produced a robust method in terms of image dimensionality reduction [92, 93]. The singular values that are related to any given matrix generally consist of information that describes the level of noise and the energy. SVD represents a method for extracting the most significant features of an image. SVD based image feature extraction produces a set of image characteristics that can help to enhance the recognition rate due to the ability to increase the image contrast [72, 94, 95].

### **2.5.2 Speech Feature Extraction**

Speech recognition technology commonly employed in several of daily activities such as giving the direction coordinates to GPS device while driving, sending text messages, access to specific information, security issues such as person authentication systems for mobiles and PCs, and many other applications that can be driven using the speech recognition technology. Recently, human speech recognition has got a wide interest among the researchers in the computer science field. The early systems were performing the recognition task by capturing the speakers' voice and then trying to match the wave sounds [96]. However, the speech kind of biometric is normally influenced by the environment and has a high sensitivity to noise,

furthermore, the speech also sensitive to special circumstances that may the speaker faces such as illness, aging, and even though the speaker may speak differently according to different occasions [36]. Hence, the idea of matching the speech wave becomes irrelevant in terms of recognition.

After that, there was a major enhancement in the way that used to perform the speech recognition which is using the statistical modeling such as HMM [46] to capture a large amount of speech data from many speakers in order to produce more robust statistical model for speech recognition. Due to this change, speech recognition systems have been developed dramatically, however, there still be noticeable error rate in these systems [97].

The current trend of speech recognition is based on the simulation of brain's behavior to perform the recognition by using ANNs techniques [47]. ANNs are used to overcome HMM's limitations where the data samples increase, then the recognition accuracy will increase accordingly. Since STDP is considered the third generation of ANNs and it is mainly adopts the brain behavior, then the classification step requires more robust feature extraction method to come out with high recognition rate and less error rate.

All speech recognizers include an initial signal processing that converts a speech signal into its more convenient and compressed form called feature vectors. Feature extraction step is the most crucial step in the any recognition system. Similar to the face biometric, the accuracy and the efficiency of the recognizer have strongly depended on the extracted features quality, hence, more significant features produce

high recognition accuracy, Figure 2.7. For speech biometric, there are two common methods for feature extraction which are; Mel-Frequency Cepstral Coefficient (MFCC) [74, 98, 99], Wavelet Packet Decomposition (WPD) [100],

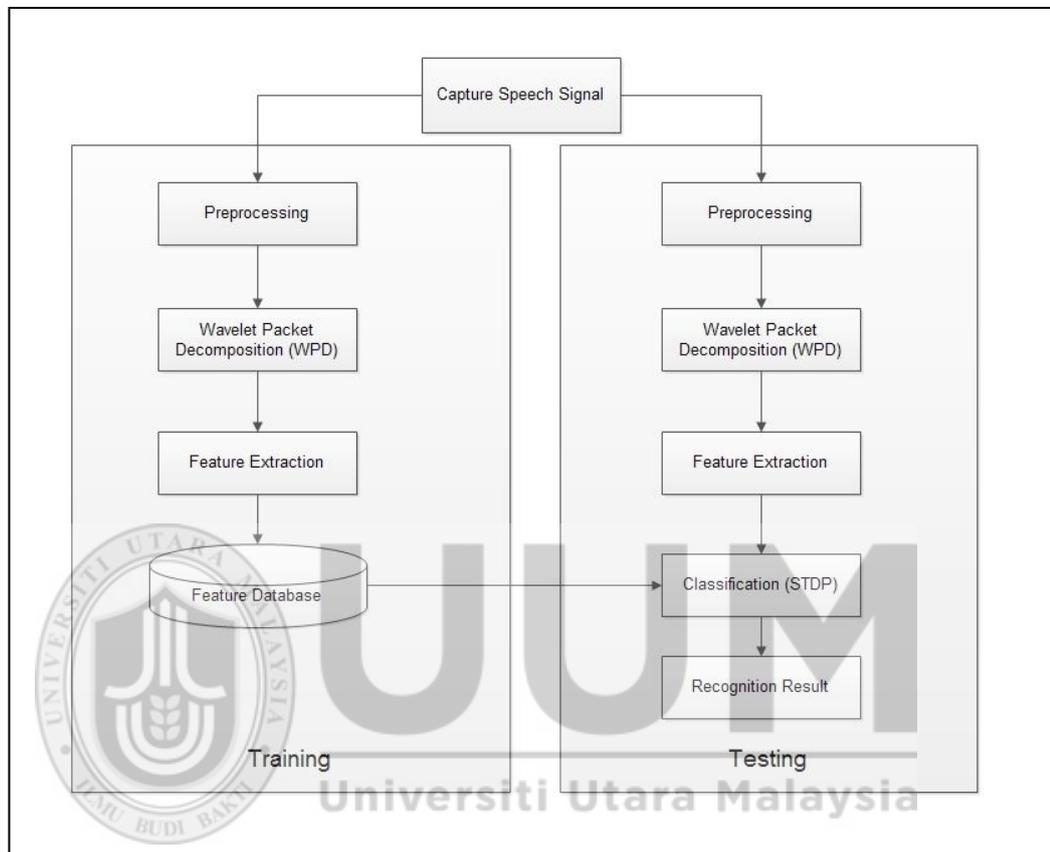


Figure 2.7: Speech Recognition System Process

Despite MFCC and LPC can give stable speech representation, they have some drawbacks when dealing with a non-stationary signals, in addition, they have a high level of sensitivity to the noise [100, 101]. WPD recently used as a robust speech feature extraction method. It proved its efficiency and effectiveness in terms of compressing and denoising the speech since the speech has a strong sensitivity toward the noise due to the microphone quality. Thus, noise reduction plays an important role in the speech feature extraction process. WPD as a speech feature

extraction method acts as a suitable choice in terms of reducing the features as well as the high efficiency when dealing with non-stationary signals [75, 76, 97, 102-108].

### **2.5.3 Face and Speech Recognition (Computational Intelligence Approach)**

Due to the desirable characteristics of spatial locality of Gabor filter, a filter extraction based on Gabor filter method has been presented in [73]. The extracted features are passed to the classifier which is based on Feed Forward Neural Network (FFNN) to reduce the feature in a way that simpler than PCA. A number of images have been tested in this study to demonstrate the efficiency and effectiveness of this approach. Several algorithms have been employed in terms of face recognition; however, two of them have high detective rate which is; Eigenfaces and Elastic graph matching. Eigenfaces perform well but still suffer from some drawbacks such as, the sensitivity to illumination and scaling. In contrast, more robustness can be earned with elastic graph matching method against the illumination. However, elastic graph en efficient in terms of time consuming and computational complexity, moreover, it is less desirable in commercial systems. Using Gabor wavelet seems to be good, thus, and in this study, a new approach has been presented by combining Gabor wavelet and FFNN. The results of the study have proven that this method can achieve better performance compared with elastic graph and Eigenfaces.

Generic-based feature extraction study is presented in [52], for optimizing the extracted features. The results show that this approach for biometric feature extraction can effectively reduce the number of features required for the recognition

performance. In [53], a novel cascade face recognition and feature extraction has been proposed. The system includes three Artificial Neural Networks (ANNs) classifiers to enhance the system efficiency and reliability. Figure 2.8 shows the three classifiers list. The result of the study showed that embedding the three classifiers of ANN can lead to high rate of accuracy for face recognition and more reliability as well. Figure 2.8.

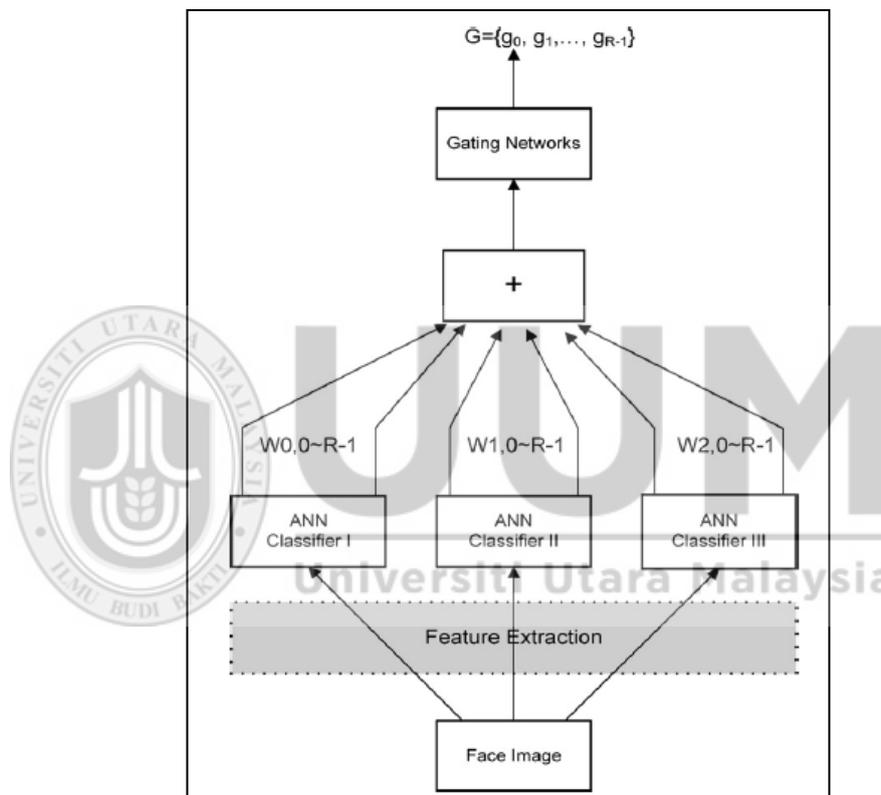


Figure 2.8: Face Recognition Based on Three ANN Classifiers (Reproduced from [15])

The performance of the face recognition process is normally has a sensitivity against the image variations and illumination conditions. Many algorithms have been presented to overcome these challenges such as, histogram and Eigenfaces [109]. In this study, a minutiae based on thermal face recognition method has been made as described in Figure 2.9. The system developed consists of three steps; face region

crop, feature extraction, and the classification and face recognition. The final recognition has been enhanced when using this method.

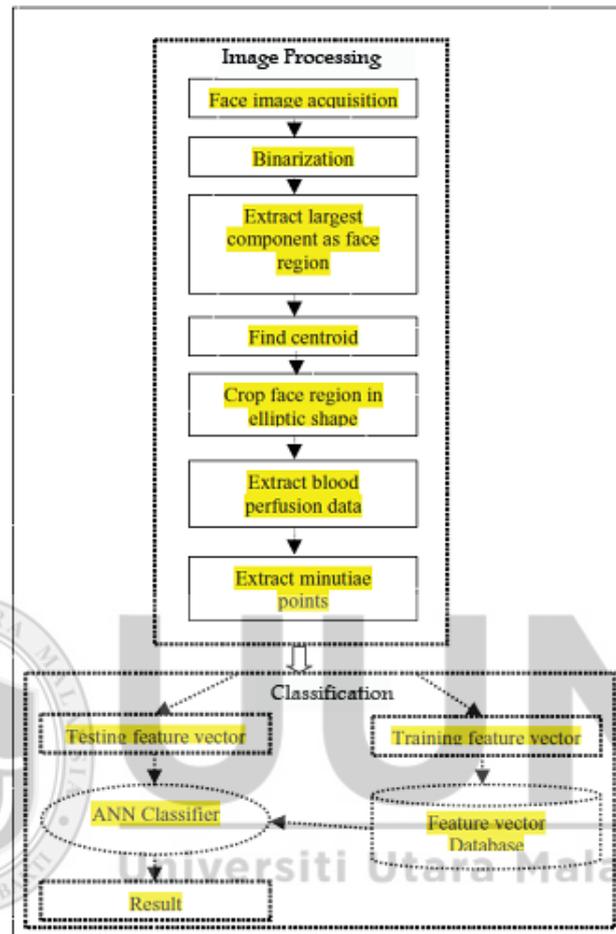


Figure 2.9: Thermal Face Recognition (Reproduced from [53] )

## 2.6 Face and Speech Authentication

Since the multimodal biometric authentication system performs better than the unimodal in terms of accuracy and effectiveness, many multimodal systems have been presented last few years. These systems normally combine two or more biometric to perform the identification and the verification for individuals. Among several human biometrics, this research is focused on face and voice biometrics. In

the upcoming paragraphs, a review of a number of studies will be done for the systems that employ face and speech biometrics.

A multimodal biometric system based on face and speech recognition has presented in [63]. The objective of face recognition is to identify the individuals based on their image face. This image will be compared with other stored images to perform the recognition. Face detection that used in this study is based on PCA method, while speech recognition is based on Gaussian Mixture Models (GMMs). The result of the study showed a higher stability authentication rate which can overcome the limitation of unimodal systems.

A general approach of multimodal systems has presented in [110]. The proposed system combines two types of biometric traits which are face and speech. The fusion level used in this study is on features level and decision level in order to make the system more robust. Buffering approach helps to enhance the system accuracy. Statistical methods have also been used; Figure 2.10 shows the model of the proposed system. The buffer is used to store the current information to enable the fusion of concurrent information and consequently enhance the overall accuracy.

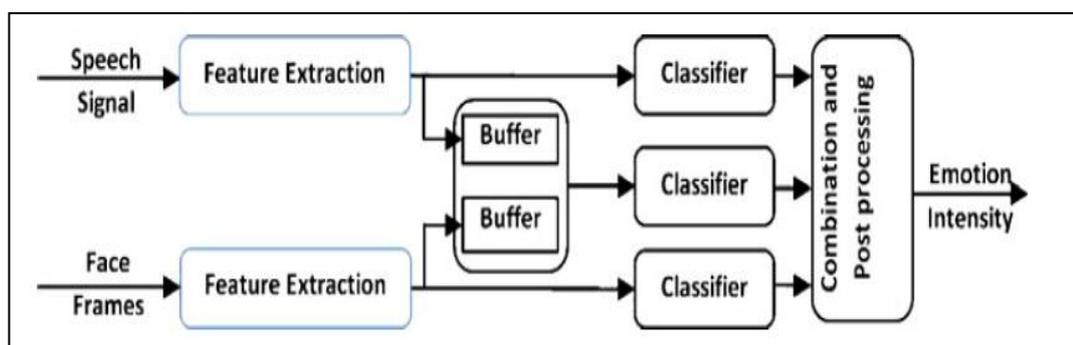


Figure 2.10: Buffering Approach Biometric System (Reproduced from [110] )

A multimodal person authentication system based on face and speech is presented [111]. The speech verification was based on Mel-Frequency Cepstral Coefficient (MFCC). The face verification system has been built using PCA and Linear Discrimination Analysis (LDA). The experimental results have clearly shown the efficiency of multimodal systems comparing with unimodal as well as the performance in at a high rate of accuracy.

In spite of the good performance of the systems that have been designed based on the face and speech, it is clearly can be seen that almost these systems have been built by adopting the statistical approaches. These approaches have a number of drawbacks that affect the performance of such systems. Firstly, the time consuming using statistical approaches can be noticed significantly. Secondly, statistical approaches normally require for high level of complexity which may hinder the performance of these systems. Lastly, these approaches have less ability for learning in comparing with artificial models.

## **2.7 Artificial Neural Networks (ANNs)**

The first generation of ANNs has been successfully employed used to in terms of problem solving such as the classification [112-114], however, the performance of these models suffers from several drawbacks [115]. Since the brain's neurons or spikes transmit the information across the brain using pulses, there should be a significant computational representation for such process especially when it depends strongly of the time. Therefore, the first generation of ANNs is insufficient in terms of brain activities representation.

In the second generation of ANNs, neurons use the continuous activation function instead of step or threshold function to compute their output signal. This neural computation is known as rate coding. According to this improvement of the output computation, it can model the intermediate frequency of pulsing that can approximate any analog function arbitrarily [116, 117]. There was another realistic improvement in the second generation of ANN models which is the learning efficiency. Learning is performed throughout the modification of weights by strengthening or weakening their effectiveness. Hence it can be modulate the incoming signals and accordingly can affect the strength level of output. The synaptic plasticity mechanism inherits some biological properties of a real neuron. This generation has proved its efficiency in terms of supervised information such as back propagation [118], and the self-organizing map (SOM) [119]. These two generations of ANN performed well in terms of problem solving in the context of classifications [114], clustering [120], and cognitive modeling [121]. However, the plausibility of these models regarding to biological neuron properties is minimal with several drawbacks.

The need to understand the remarkable capabilities of information processing for human brain has led to develop more complex processing models, namely brain-like models which are the Spiking Neural Networks (SNNs) [56], which represent the third generation of ANNs. These models adopt the training of spikes as the internal representation for the information instead of continuous variables. Nowadays, many researchers are trying to adopt SNNs to perform empirical studies [122-124].

The aim of any learning method is to produce an output for the neurons [125], each class labels relies on the classification problem. After producing a specific input dataset to the network, the corresponding spike train is broadcasted throughout the SNN which may also be produced in the firing stage of specific output neurons. There is also a possibility that may no output neuron is activated and the network remains silent. That means, the classification outcomes are undetermined.

The human brain has the incredible capability to learn the patterns with a variety of time scales, starting from milliseconds up to years and may up to millions of years (e.g. genetic information). Thus the brain is the extreme inspiration to develop a new machine learning techniques. Brain-inspired or brain-like Spiking Neural Networks (SNNs) [56] have the ability to learn using trains of spikes transmitted between spatially located synapses and neurons. In addition SNNs have proved their efficiency due to the ability to process and coordinate multi dimension information like time and space [27].

Spike-Timing Dependent Plasticity (STDP) algorithm rationally represents the physiological mechanism for activity-driven synaptic regulation. The neuron which provided with STDP can solve complex computational problems. The main idea of STDP is to enforce the connection of neurons each time the neuron fires. STDP plays a key role through detecting the repetitive patterns and make a response to them. It is currently a significantly well-established physiological mechanism of activity driven synaptic regulation [24, 33]. STDP uses correlations in the firing times of pre- and postsynaptic neuron for synaptic changes. In addition, through the third generation models, the memory capacity can also be maximized with

appropriate encoding strategies. For example, the parameter of transmission delay in a spiking neural network allows formation of polychronous groups that can store a vast amount of patterns [126, 127]. In encoding with polychronisation concept, a pattern is represented through a chain of neuron firings.

The Spike Driven Synaptic Plasticity (SDSP) is another algorithm that fall under SNNs category. It is a semi-supervised learning method [30, 31]. SDSP model of can learn to classify complex patterns in a semi-supervised fashion. The rule of SDSP learning introduces a long term dynamic of the synaptic weights depending on the value of the weight itself. If the weight is above a given threshold, then the weight is slowly driven to a fixed high value. In contrast, if the weight is driven by the learning mechanism to a low value, then the weight is slowly driven to a fixed low value. These two values represent the two stable states of SDSP learning method. In [31], the SDSP model is successfully employed to train and test a SNN for characters recognition. Each character was in the form of static image represented by feature vector, and each value of features vector is converted into spike rates, with. For each class, there were different training patterns used and there were specific number of neurons allocated, and trained for several thousand iterations. Rate coding of information was used rather than temporal coding, which is typical for unsupervised learning in SNN. In spite of SDSP has been successfully used for the recognition of mainly static patterns, the potential of the SDSP SNN model and its hardware realization have not been fully explored for spatio- and spectro-temporal data (SSTD), and not efficient for fast on-line learning of complex spatio-temporal patterns [29]. In addition, SDSP in nature is a semi-supervised learning based technique where there should be a set of specific learning rules in order to perform

the recognition, and this issue contrast with the reinforcement learning concept where the learning is performed as target-based behavior [32].

## 2.8 Spiking Neuron Models

Due to the development of neuroimaging technology, the brain activities have been delineated for better understanding, recording, and investigating in order to project the structural and functional behavior of the brain's activities. The human's brain consists of a huge number of neurons (Figure 2.11), each of them maps a connection with other neurons generating a network. The interaction between two neurons is described as receiving the impulse signals (input) and triggering an action (output). Generally, the biological neuron is composed of three major parts which are; the dendrites, the soma (or cell body), and the axon. The signals are received by dendrites from other neurons throughout a specific connection (synapses), the incoming signals are collected in soma (Figure 2.11), whenever sufficient signals received, then the cell is fired, and starts to transmit the signals the axon to other cells [56].

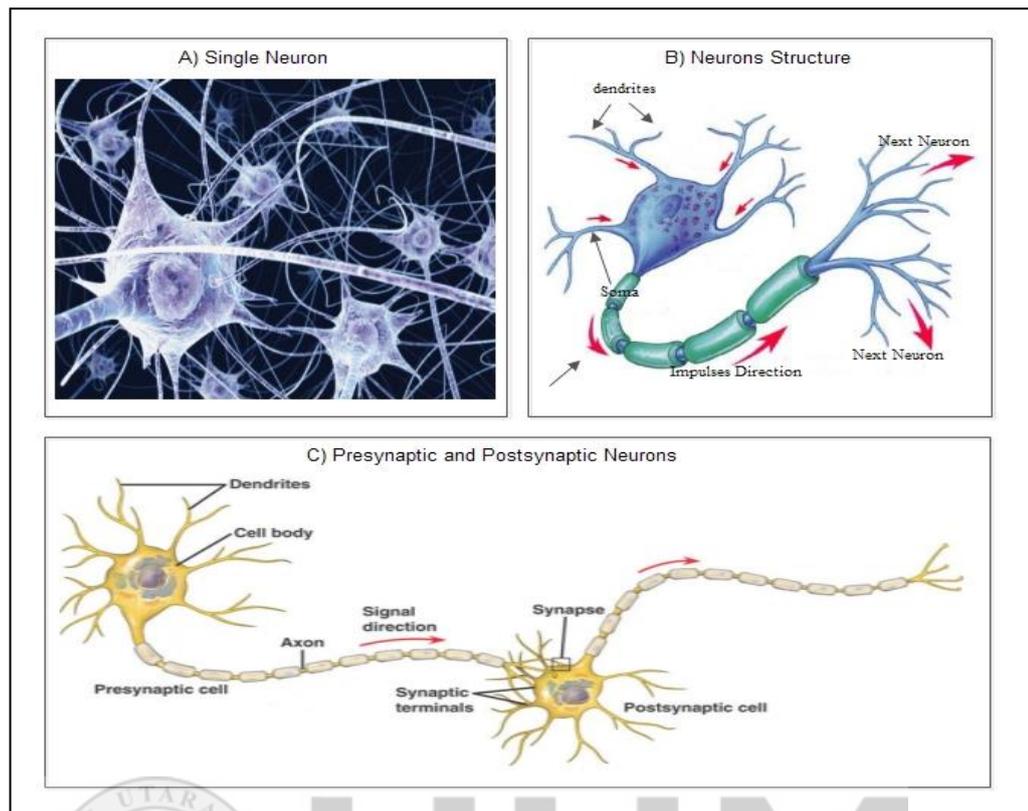


Figure 2.11: Neurons Network Example

The biological neurons communicate through the neurotransmitters which are specific chemical messengers [128]. A neurotransmitter modulates and boosts the signals between two neurons. The sender neuron (presynaptic) sends the signals (information) in the form of action which called spikes to the receiver neuron (postsynaptic). Postsynaptic neuron may or may not fire depending on the difference between the cell's interior and the surroundings. This difference is called membrane potential [56]. A certain threshold is determined and when the membrane potential reaches it, the neuron is triggered and generated a spike. Generally, a spiking neuron model depicts the accumulation of input signals that produce the spike which is specified by the increment and the decrement of the membrane potential value.

### 2.8.1 The Hodgkin-Huxley (HH) Model

This spiking neuron model is one of the most significant models in the context of computational neuroscience. It was involving an experiment conducted on a squid axon and it has been found that there are three main ion channels called sodium (Na), potassium ( $K$ ) and leakage currents carried by  $Cl^-$  ions. Hence, the HH model has been proposed in order to describe the conductivity of the currents that carried by a cell membrane. The following formula explains the HH model:

$$Cv' = I - g_K n^4 (v - E_K) - g_{Na} m^3 h (v - E_{Na}) - g_L (v - E_L) \quad (2.1)$$

Where:  $C$  is the membrane capacitance,  $v$  is the membrane potential, the variable  $I$  is the current's summation, the variables  $g_K$  and  $g_{Na}$  represent the amount of time dependent conductance functions, the variable  $g_L$  is the voltage-independent conductance that depicts the leakage channel. And lastly  $E_K$ ,  $E_{Na}$  and  $E_L$  are the reverse potentials for corresponding ions.

In spite of that the HH model is convenient to describe the squid neuron system; it has significant disabilities when dealing with more complex neuron systems such as human's brain [129]. Thus, this model considered to be inefficient to build brain like encoding system.

### 2.8.2 Leaky Integrate-and-Fire (IF) Models

IF is considered as the simplest integrate-and-fire (IF) model, thus this model in most widely used in SNN studies. It is represented by a simple routine, which is

the neuron integrates the input signals and then spike at a specific threshold. The leaky IF model can be formulated by the following formula:

$$v' = I + a - bv \quad (2.2)$$

*if  $v \geq v_{thresh}$ , then  $v \leftarrow c$*

Where the value of  $v$ , represents the membrane potential value,  $I$  is the input current, and  $a$ ,  $b$ ,  $c$ , and  $v_{thresh}$  represent the parameters of the model.

According to the single variable ( $v$ ), the model of leaky IF can only work as an integrator. The thing that makes this model is limited to only specific SNN models since the single IF model could not satisfy the performance of spiking process.

### 2.8.3 Izhikevich Spiking Neuron Model (IM)

This model was proposed in [13], it is based on two principals; computationally simple, and the capability of producing high level of firing patterns demonstrated by the real biological neurons. The model reproduces spiking process according to four basic parameters which are;  $a$ ,  $b$ ,  $c$ , and  $d$  [13, 129, 130].

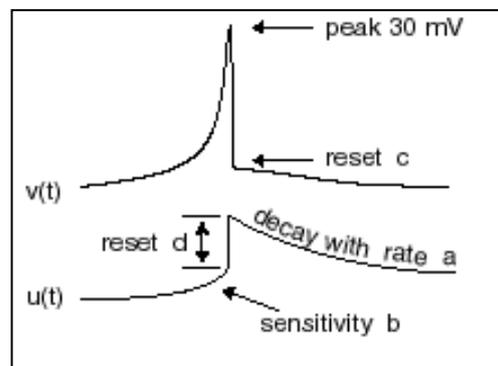


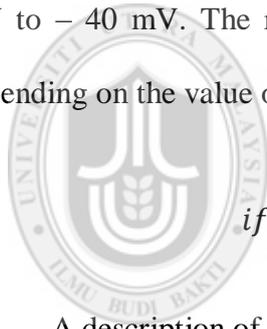
Figure 2.12: Izhikevich Spiking Neuron Model (Reproduced from [131])

The following formula describes the IM model:

$$v' = 0.04 v^2 + 5v + 140 - u + I \quad (2.3)$$

$$u' = a (bv - u) \quad (2.4)$$

Where  $v$  represents the membrane variable,  $u$  is the membrane recovery variable which calculates the activation of  $K^+$  and the inactivation of  $Na^+$  ionic currents. When the spike value reaches the peak ( $v_{peak} = +30 \text{ mV}$ ),  $v$  and  $u$  variables are at the reset state according to (4.5),  $v_{peak}$  is not a firing threshold, but the peak (cut off) of a spike. The dynamic firing threshold gives the model the similar behavior like real neurons, depending on the activity. Approximately the value is between  $-55 \text{ mV}$  to  $-40 \text{ mV}$ . The resting potential in the model is between  $-70$  and  $-60 \text{ mV}$  depending on the value of  $b$ .



$$\text{if } v \geq +30 \text{ mV, then } u \leftarrow u + d, v \leftarrow c \quad (2.5)$$

Universiti Utara Malaysia

A description of the variables ( $a$ ,  $b$ ,  $c$ , and  $d$ ) can be seen in the following list:

- Parameter  $a$ : the time scale of the recovery variable  $u$ , smaller values result in slower recovery (typical value,  $a = 0.02$ ).
- Parameter  $b$ : the sensitivity of the recovery variable  $u$  to the sub threshold fluctuations of the membrane potential  $v$  (typical value,  $b = 0.2$ ).
- Parameter  $c$ : the after-spike reset value of the membrane potential  $v$  caused by the fast high-threshold  $K^+$  conductance (typical value,  $c = -65 \text{ mV}$ ).
- Parameter  $d$ : after-spike reset of the recovery variable  $u$  caused by slow high threshold  $Na^+$  and  $K^+$  conductance (typical value,  $d = 2$ ).

From all the aforementioned, we can conclude that IM model has a significant level of simplicity due to it has only two equations with a variable that are easy to control their variables. In addition to its simplicity, IM is a plausible model since it can inherit the HH and leaky IF features [130]. This is the reason that we choose an IM model to encode the STDP approach.

## 2.9 Reinforcement Learning (RL)

In 1999, Erickson and Desimone presented a behavioral experiment on visual discrimination [132]. The experiment, known as “GO or NO-GO,” aimed to prove that an association process occurs in neurons to facilitate learning of visual association, and to test whether neurons that respond to the associated stimuli in the cortex are changed during the learning process. The responses of neurons from two monkeys were recorded during the experiment. The experiment involved showing a visual image, called a *predictor*, preceded by another visual image, called a *choice*, within a specific time delay. The neuronal activities of both monkeys were observed. The performances of the subjects, which were required to release or not release a bar followed by a reward, were recorded. A reward was given if the subjects realized that the *choice* accurately matched the *predictor* by releasing the bar, and vice versa (Figure 2.13). After a number of trials, the researchers found that the monkeys exhibited significant learning capabilities. In addition, this experiment proved that a correlation exists between the activities of neurons and delay time, thus indicating that neurons can learn the temporal sequences of stimuli. Moreover, this experiment proved that monkeys can learn faster by using associative learning.

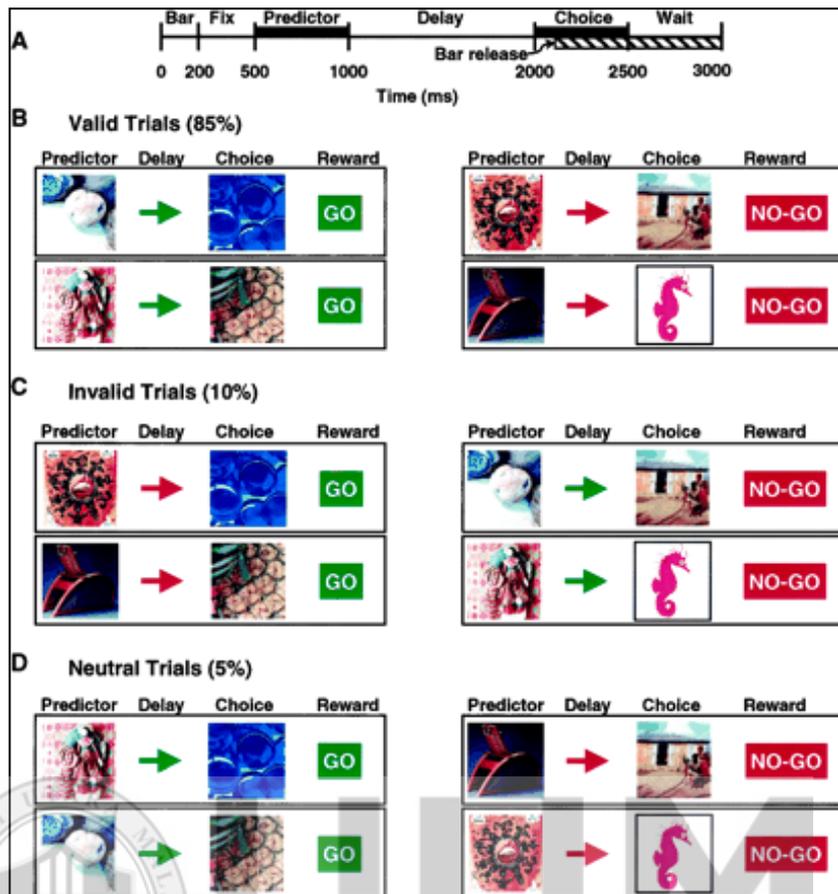


Figure 2.13: Associative Learning Experiment (Reproduced From [132])

Reinforcement learning algorithms enable the agent to learn an optimal behavior by interacting with some environment elements and learn from its obtained rewards [13, 133]. The agent uses a policy to control its behavior, where the policy is a mapping from obtained inputs to actions. RL is quite different from supervised learning where an input is mapped to a desired output by using a dataset of labeled training instances [134]. One of the main differences is that the RL agent is never know the optimal action; instead it receives an evaluation signal indicating the quality of the selected action.

Markov Decision Process (MDP) is one of the algorithms that used in performing RL classification [135-137]. It is a probabilistic temporal model of an agent interacting with its environment. It consists of the following [134, 138]:

- A set of states ( $S$ ).
- A set of actions ( $A$ ).
- A transition function  $T(s; a; s_0)$ .
- A reward function ( $R_s$ ).
- A discount factor ( $D$ ).

At each time,  $t$ , the agent is in some state  $S_t \in S$ , and takes an action  $A_t \in A$ . This action causes a transition to a new state  $S_{t+1} \in S$  at time  $t + 1$ . The transition function gives the probability distribution across the states at time  $t+1$ , such that  $T(S_t; A_t; S_{t+1}) = \Pr(S_{t+1} | S_t; A_t)$ . The reward function  $R_s$  specifies the reward being in state  $S$ . MDP can be represented as in Figure 2.14, where each node represents a single state.

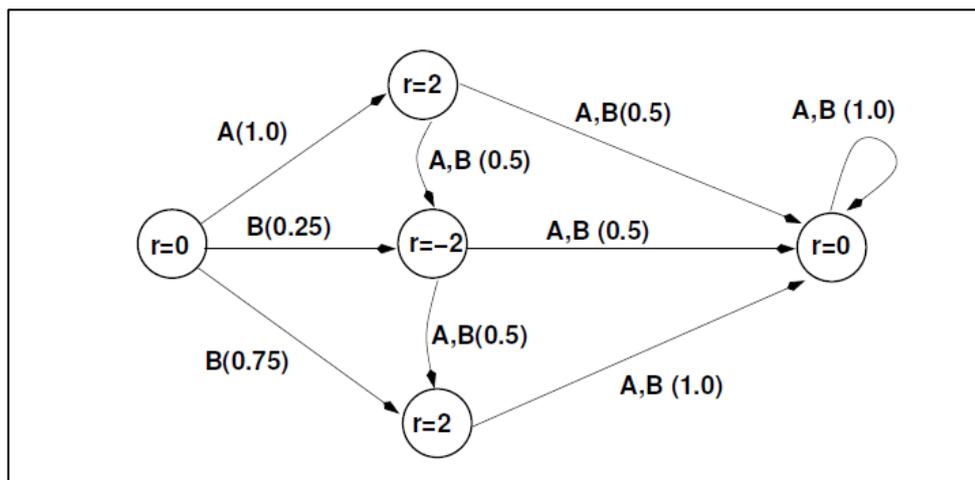


Figure 2.14: Markov Decision Processes Representation (Reproduced from [137])

Q-learning is another reinforcement learning algorithm [13], it is considered as an off-policy algorithm, which means that the agent learns about the optimal value-function while following another behavioral policy that includes exploration steps [139]. A disadvantage of Q-learning is that it can diverge when combined with another function, also, the off-policy algorithms do not modify the behavior of the agent to better deal with the action [140].

Another RL algorithm is the Actor-Critic (AC) algorithm, which is considered as on-policy method. In a way that different with Q-learning, AC method keep track of two functions; a Critic that evaluates states and an Actor that maps states to a preference value for each action. A number of Actor-Critic methods have been proposed [13, 141]. Despite the aforementioned algorithms have been successfully employed in terms RL; they share a number of limitations such as the deficiency when dealing with large and complex data, also, when the algorithms needs to select an action for test, it becomes much slower for classifying new data [13, 134].

The rapid development of these technologies indicates that the next generation of technologies will move toward ubiquitous computing [142]. Recently, biometrics has become one of the richest fields in the context of research, particularly audiovisual recognition [143]. Facial recognition commonly requires a camera to track and capture human faces for identification purposes. This process normally involves segmentation, feature extraction, and classification [144]. To perform these tasks, an excessive amount of training data is required. Thus, system

performance is improved when the techniques used for different types of authentication systems consider the evolution of learning.

When the data set is small, the performance of a biometric system is poor. To overcome such problem, a study presented in [145] used a semi-supervised method based on face and gait biometrics. The most significant part of this previous study is employing co-training for both biometrics. Self-training has been used in biometrics via PCA methods, and is currently one of the hottest issues in semi-supervised learning methods [146]. The self-learning process can be described as follows.

- The classifier is trained by using a small amount of data.
- The classifier is then retrained. This process is repeated several times.

A learning discriminative local binary histogram (LBH) approach was used in [147] to perform effective recognition of human gender based on facial biometrics. This previous study adopted a new approach, which involves training a number of weak classifiers to enhance their performance, to perform the learning process. This study showed that the proposed learning approach yielded better classification performances than those obtained by using statistical approaches.

The joint discriminative dimensionality reduction and dictionary learning (JDDRDL) approach was proposed for facial recognition in [148]. This approach is different from PCA and linear discriminant analysis. The result of this previous study showed that the JDDRDL approach performs better than other approaches in terms of facial recognition and classification.

STDP is generally one of the most acceptable mechanisms used for learning with SNN [149]. However, STDP learning can only be performed with an unsupervised approach. In addition, STDP must be integrated with other suitable encoding strategies, which can delay the encoding process. Learning with STDP is also limited by its need to control synaptic changes as learning progresses. Thus, this process may generate unlimited growth (suppression) of weights even after learning has stabilized [126].

When the nature of learning is considered, our first idea is that we learn by interacting with our environment. When an infant waves his/her arms, plays, laughs, or looks at the things around him/her, he/she is not taught by a teacher; however, a direct connection and interaction exist between the infant and his/her environment [13]. By training this connection, a large amount of information on cause and effect becomes available. Furthermore, the aforementioned interaction configures the main source of knowledge on the environment and how to deal with such knowledge. When a person performs an activity, such as driving a car, this person is aware of how the environment responds to such activity. The idea of learning from our environment is the core concept behind all theories that deal with learning and intelligence. RL, which is a model of trial-and-error or the so-called “law of effect” in psychology, is a method of learning from interacting with the environment [150].

A few studies have reported on the application of SNN in RL. An abstract of algorithms, which is not based on explicit neural modeling, has been found [13]. Recently, however, studies on modeling RL in SNNs have increased. In RL, agents should upgrade their internal parameters to increase rewards based on the given

period [32, 151, 152]. This upgrade can be implemented by performing a sequence of trial-and-error action–rewards in response to environmental stimuli. In a manner different from those of supervised and unsupervised approaches, in which most of the learning cases are subjected to specific rules with a given initial state, agents explore and use their unknown identity states to establish a learning policy in RL. Thus, this form of learning has a high level of plausibility (Figure 2.15).

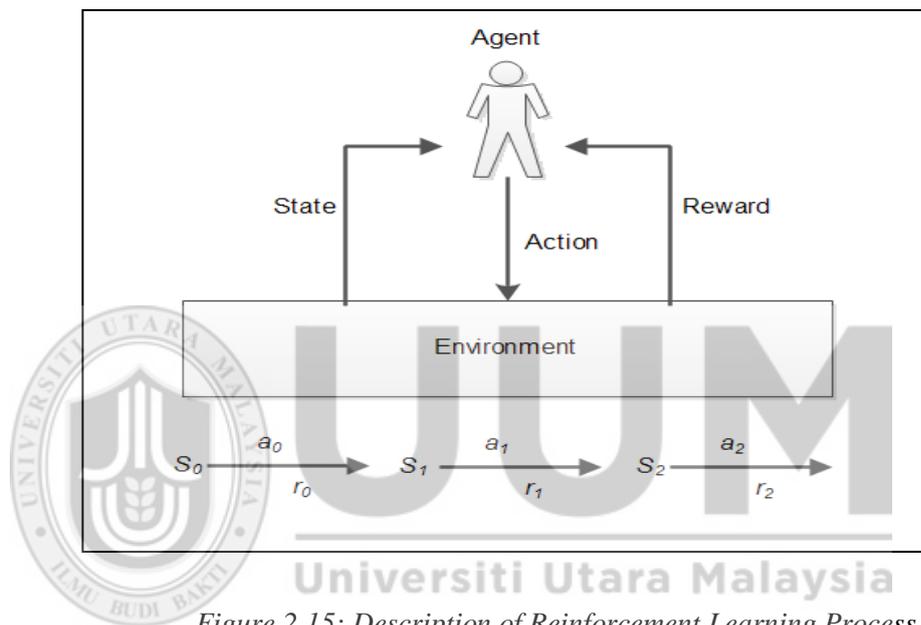


Figure 2.15: Description of Reinforcement Learning Process

## 2.10 Summary

From aforementioned sections, it can be clearly seen that the biometric authentication systems development is running very fast due to the development of computer system and information technology. Since, there are a variety of biometric traits; also there are variations of the techniques that used to build such authentication systems. The performance, accuracy, and time are playing the key role in the context of biometric recognition; hence, when designing an authentication system, there should be an appropriate selection of the techniques that will satisfy the

system and user requirements. According to the literature mentioned, ANNs can present the optimum solution so far. They are contributing significantly in consuming the time, increase the accuracy, as well as the ability of learning for neural networks makes them on the top of authentication techniques.

Feature extraction plays a key role in all authentication systems, since the accuracy of the classification results depends strongly on the quality of the features extracted. PCA and SVD are two common feature extraction methods that have been implemented widely in the field of facial feature extraction. These methods have proven their efficiency and effectiveness in terms of dimensionality reduction. For speech recognition; WPD is considered to be a good feature extraction that used to extract speech features. It was proved its significant performance due to the ability of denoising the speech as well as the high performance when dealing with non-stationary signals, the thing that makes it suitable for speech feature extraction and authentication systems, since the speech might not be pre-recorded.

One of the authentication systems' factors is applying an efficient learning technique. Reinforcement learning (RL) is a very effective learning approach since it adopts the human behavior in terms of learning. RL can be implemented by performing a sequence of trial-and-error action-rewards in response to environmental. In a way that is different from supervised and unsupervised approaches, where most of the learning cases subject to specific rules with given initial state, in the RL approach, agents explore and use their unknown identity states to establish a learning policy. That means this kind of learning has a high level of plausibility.

SNNs are presenting the most significant performance due to the ability of learning by adopting human-like learning techniques. The brain is an inspiration factor to adopt SNNs. Learning approach can contribute effectively to enhance the system accuracy and performance. In addition, SNNs have the ability to process multiple dimension information such as time and space. STDP is one of the most effective algorithms that belong to SSNs, which proved its ability to use the time more effectively.



# CHAPTER THREE

## RESEARCH METHODOLOGY

### 3.1 Research Methodology

In the previous chapter, we reviewed literature related to our research objectives to build a comprehensive understanding of aspects related to this study. This extensive literature review allows selection of suitable methods to conduct research activities and to come up with the research objectives. This chapter describes the application of the selected methods. The methodology consists of four main phases, which are described as follows.

- *Phase I:* this phase focuses on the multimodal biometric authentication and how such approaches can overcome unimodal authentication limitations. It also discusses the statistical approaches that used in multimodal authentication and states the differences with computational intelligence approaches. In addition, related research is also explored to express the main characteristics of authentication systems and the rapid development of multimodal systems. This phase presents a significant understanding of biometric technology and the development of such technologies in authentication systems. Furthermore, the development methods of and the associative learning method are evaluated for adoption and implementation in later phases.
- *Phase II:* This phase focuses on extracting facial and speech features, which is a critical step in all biometric authentication because the classification process primarily depends on the quality of the extracted features. For facial biometrics,

two feature extraction methods are adopted: the PCA-based Eigenfaces approach and SVD feature extraction. The facial data set used is from ORL. The speech feature extraction method used is WPD. The speech data set used is from TIDigits. Two facial feature extraction methods are used to test which one can perform better, and thus, satisfy the second objective. MATLAB 7.10 (R2010a) programming language is used to extract the biometrics. The outcome of this phase is a set of facial and speech features that will be used as input data for the next phase (i.e., classification).

- *Phase III:* In this phase, associative learning is implemented by using SNN based on STDP. RL is used as a machine learning approach following the trial-and-error concept. The feature selection mechanism and all processes used to encode spikes are demonstrated. A number of experiments with different network structures and parameters are performed. A test on actual data collected by the researcher is also conducted to test the performance of STDP, implement the result, and clarify the outcomes. MATLAB and C<sup>++</sup> are the programming tools used to encode the learning process of STDP, particularly for training and classification.
- *Phase IV:* Evaluation is performed to measure the accuracy and performance response of the multimodal association learning network. This phase involves analyzing the results according to the given output of the current recall accuracy. In this context, accuracy refers to the correctness of the learning model in associating face–speech biometrics with the target individual. The parameter we used to evaluate the model is performance rate, which represents the accuracy

ratio, and can be calculated as follows:  $\text{Performance} = (\text{number of correct calls} / \text{number of trials}) \times 100$ . Figure 3.1 presents an overall view of the phases of the research methodology.

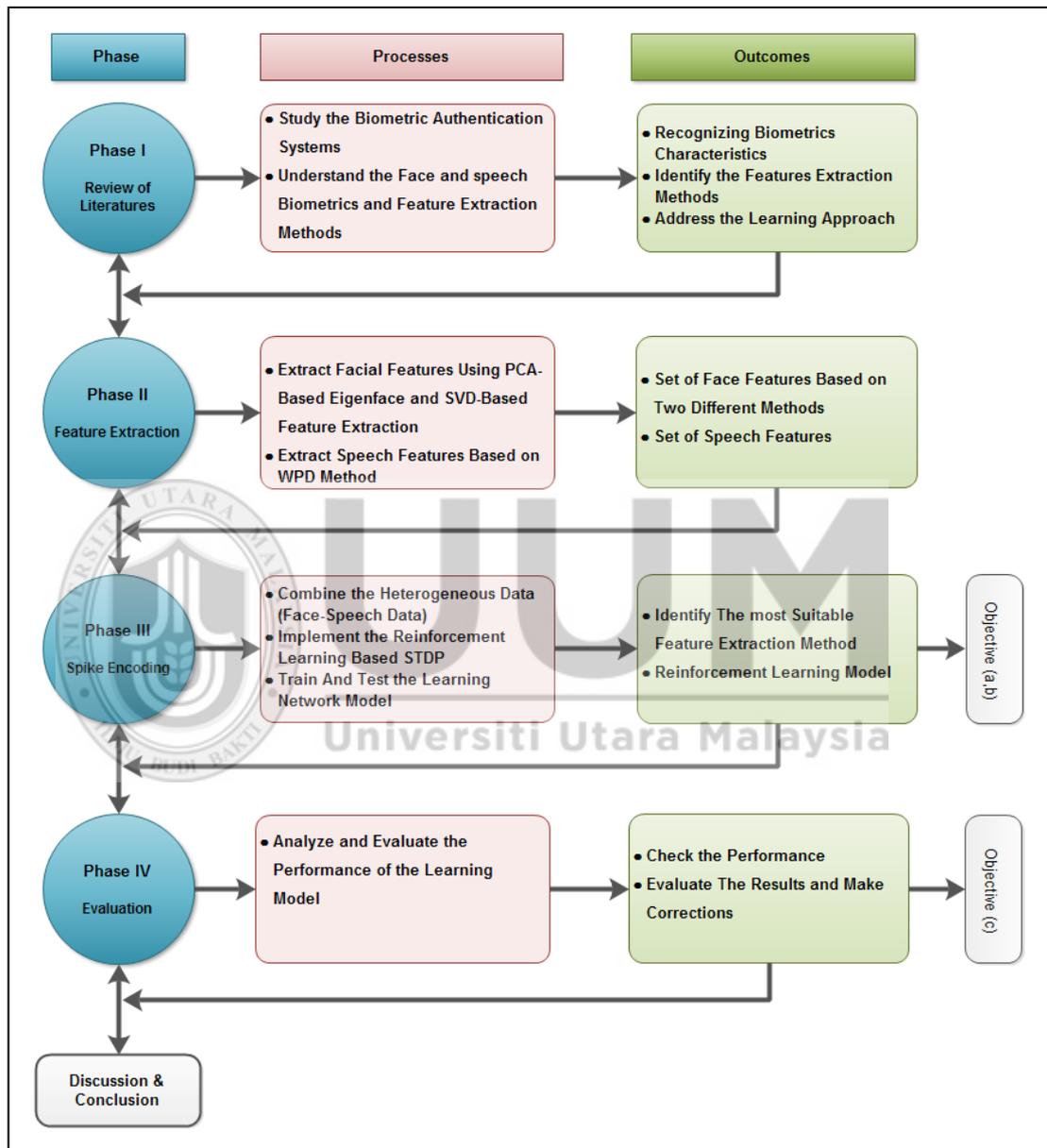


Figure 3.1: Design Research Methodology

### 3.2 The Learning and Classification Process

The authentication learning process starts by capturing the face image by using a camera and capturing voice by using a microphone. Feature extraction should then be performed to determine the most significant feature characteristics for accurate recognition. When the feature-extraction phase is completed, the classification process begins. Figure 3.2 shows the learning flowchart.

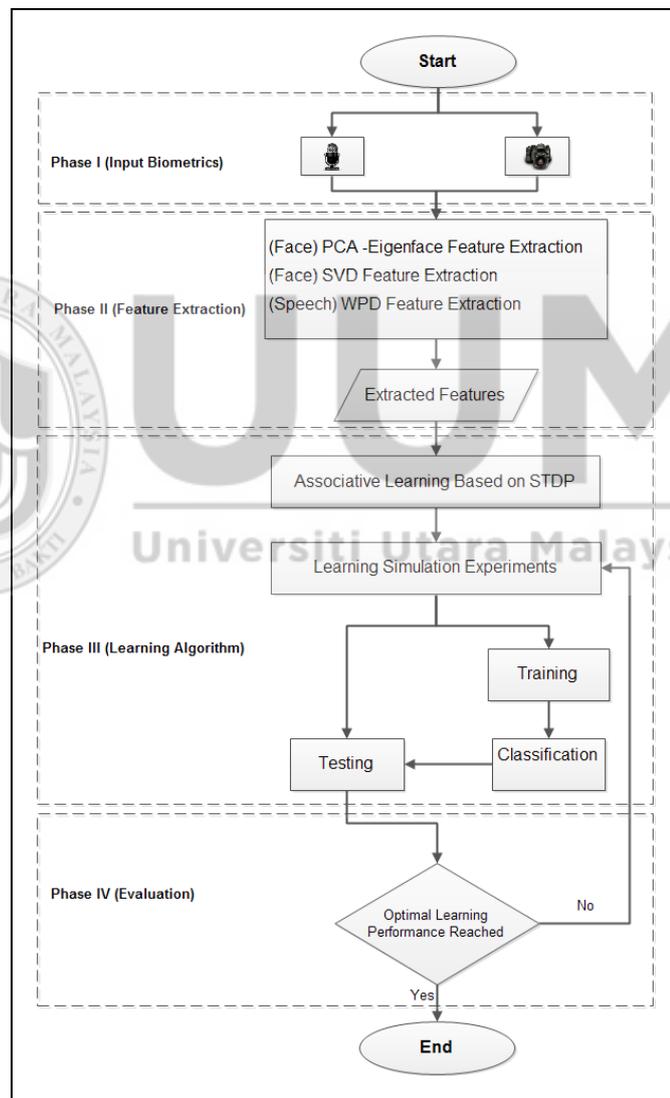


Figure 3.2: Process Flow of the Reinforcement Learning Using STDP for Multimodal Face-Speech Association Learning

### **3.3 Phase I: Review of the Literatures**

This section represented by Chapter1 and Chapter 2, which describe the multimodal authentication approach and the main related aspects. It also depicts the limitations of other biometric systems and how to overcome such limitation.

### **3.4 Phase II: Feature Extraction**

In this section, we describe facial feature extraction by using two common methods for facial images (PCA and SVD) because of the high level of performance and the efficiency of these methods in reducing dimensionality. For speech feature extraction, we use WPD to extract the most dominant speech features

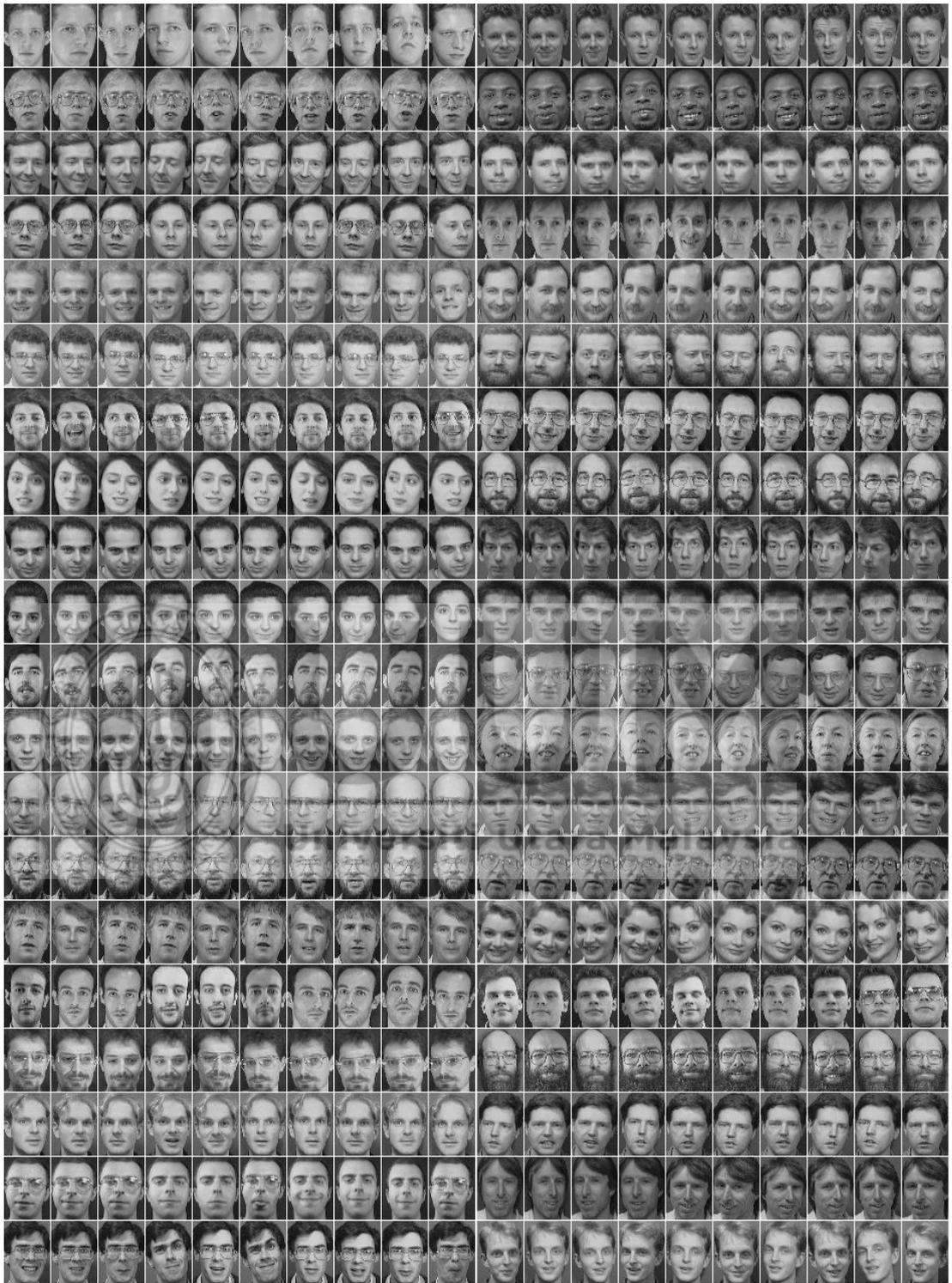
#### **3.4.1 Face Features Extraction**

Given that the efficiency and effectiveness of any classification process significantly depend on the size of the trained data set and the quality of the extracted features [67], feature extraction and dimensionality reduction are important and challenging processes. The present study adopts two algorithms to perform facial feature extraction and prepare data for classification: PCA and SVD.

In facial recognition, sample data must be prepared prior to training. In this study, we have selected the face data set from ORL (currently known as AT&T) [153]. The ORL data set contains facial images from 40 subjects with 10 facial expressions (smiling, sad, happy, with glasses, eyes closed, and so on) for each subject. The data set was collected between April 1992 and April 1994 at the Cambridge University Computer Laboratory [153, 154]. All images were captured

under different lighting conditions as well as in a homogeneous dark background. Figure 3.3 shows the ORL data set. The facial images in the data set are  $92 \times 112$  pixels in dimension, with 256 gray levels per pixel. The images are sorted into 40 folders, with each subject having one folder that contains his/her images with different facial expressions. According to the naming convention followed in this study. Given its variety in image illumination, poses, and expressions, the ORL data set has been used in several studies to test facial recognition algorithms [43, 155-159]. In the succeeding sections, we explain image preprocessing and feature extraction for PCA and SVD by using the ORL data set to produce an optimal set of features, and accordingly, obtain the best facial recognition results by using SDTP.





*Figure 3.3: The ORL Face Images Dataset [153]*

### 3.4.1.1 Face Features Extraction Using PCA

PCA is one of the most common data analysis techniques that aim to reduce dimensionality and overcome the curse of dimensionality [71, 77, 78]. PCA has a key role in feature extraction because of its ability to remove noise and specify redundant information by extracting the most dominant features from the original data set [79-81].

The PCA feature-extraction method starts by preparing the data set to train the recognizer. As mentioned earlier, the ORL facial image data set is used in this study. PCA does not process an image directly. An image must be converted first into vector form (i.e., a set of numerical values). The image dimensions in the ORL data set is  $112 \times 92$  pixels. To work with such images, they must be converted first into a column vector with 10304 values. This number is the product of the image dimensions, which is now in matrix form instead of in two dimensions. Figure 3.4 shows an example of converting image dimensions.

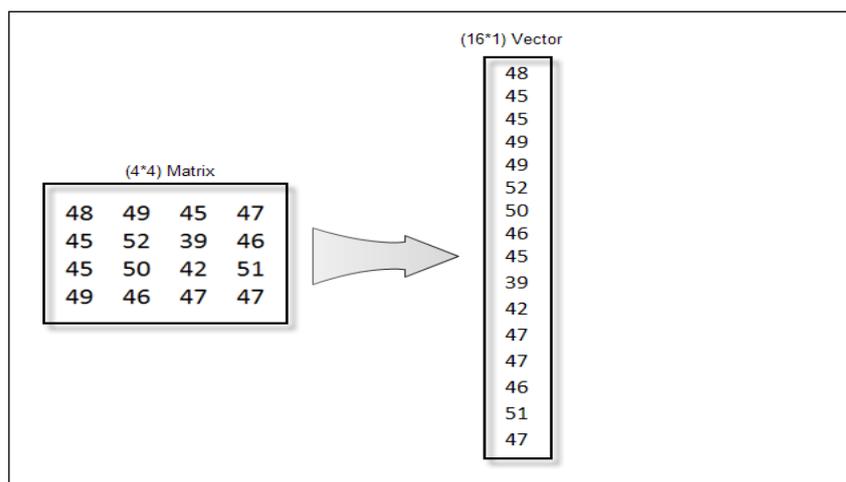
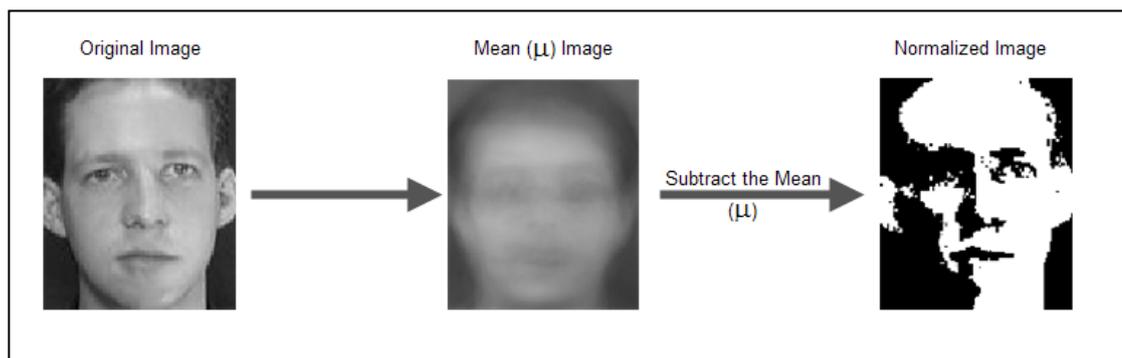


Figure 3.4: Example of Converting 2D Image into 1D Vector.

Now we have a single column vector for each image. By applying this process for all the images in the data set there will be a one matrix  $X$  ( $m*n$ ) that involves all images' features. The output matrix is constructed from a number of columns ( $n$ ) or observations and each column represents one single face image, in addition the number of rows ( $m$ ) reflects the number of features for each single image.

The next step is normalizing the feature vectors. This task is achieved by removing the features shared by the images in the data set. The common facial features in the data set can be determined by calculating the mean ( $\mu$ ) for all image vectors (i.e., the matrix  $X$ ). The mean face depicts the average features of the entire training set, that is, given that all facial images have eyes, eyebrows, nose, mouth, chin, and forehead; then, these common features do not considerably differentiate a particular image from others. Instead, other features, such as eye location, make an image unique. Hence, the mean ( $\mu$ ) has to be removed from each image to retain the most significant feature of each image. Figure 3.5 illustrates image normalization.



*Figure 3.5: Face Image Normalization*

The following formula is used to calculate the value of ( $\mu$ ) for the ORL dataset:

$$\mu = \frac{1}{n} \sum_{i=1}^n X_i \quad (3.1)$$

To obtain more significant features for the facial images, PCA includes a crucial step, which is, finding Eigenfaces [156, 160, 161]. Eigenfaces is produced to obtain facial images with the most dominant features. These faces can be calculated by finding the eigenvalues and eigenvectors for the covariance matrix of the training set [162-166]. The covariance matrix is represented by normalized image data, as previously stated. Calculating the eigenvalues and eigenvectors represents the core concept of dimensionality reduction. The following formula describes the means to calculate the covariance matrix ( $C$ ) of the facial image matrix ( $X$ ):

$$C = \frac{1}{n} \sum_{i=1}^n (X_i - \mu) (X_i - \mu)^T \quad (3.2)$$

Assuming that ( $A$ ) is the normalized matrix of the facial images, then:

$$A = (X_i - \mu) \quad (3.3)$$

The covariance matrix is then expressed as follows:

$$C = A \cdot A^T \quad (3.4)$$

The ORL data set includes 40 subjects, each with 10 facial expressions that are represented as 10 facial images with dimensions of  $112 \times 92$  pixels. Accordingly, a matrix of observations  $X$  ( $m \times n$ ) exists, where ( $m$ ) is the number of rows of the

matrix ( $X$ ), which is equal to 10304 ( $112 \times 92$ ), and ( $n$ ) is the number of columns of the same matrix, which is equal to 400 observations. Based on the covariance matrix formula [ $C = A.A^T$ ], and according to ( $X$ ) dimensions, the covariance matrix is matrix  $C$  ( $10304 \times 10304$ ), which is a huge matrix to process, and thus, will consume a large amount of memory. This problem will become increasingly complicated if we deal with a large data set of images. According to linear algebra, for any given observation ( $m \times n$ ) matrix, if the number of observations ( $n$ ) is bigger than the number of features ( $m$ ), then the covariance matrix ( $C$ ) can be calculated according to the following formula instead of the previous one:

$$C = A^T.A \quad (3.5)$$

By applying Formula 3.5, the results for a covariance matrix with dimensions  $C$  ( $400 \times 400$ ) will be obtained. Thus, this formula clearly produces the least number of features, and consequently, it requires less time and fewer computational processes than the other formulas.

As previously mentioned, eigenvalues and eigenvectors are the cores of dimensionality reduction. Hence, to obtain the most dominant features, we must calculate these factors for the covariance matrix. The formula used to calculate the eigenvalues and eigenvectors can be expressed as follows:

$$Cv_i = \lambda_i V_i \quad (3.6)$$

where  $i = 1, 2, 3, 4 \dots n$ ,

$(\lambda)$  represents the eigenvalue matrix, whereas  $(V)$  represents the eigenvector matrix.

Eigenfaces is considered as one of the most efficient feature-reduction techniques for facial recognition because of its speed and simplicity [167]. After obtaining eigenvalues and eigenvectors, each face in the data set is represented by a number of Eigenfaces, known as the vector of the feature. These Eigenfaces are used to train the recognizer to perform authentication. Figure 3.6 shows a sample of Eigenfaces for the ORL data set. The outputs of all previous steps are represented by a 2D matrix that includes the Eigenfaces extracted from the original facial images in the ORL data set. This matrix represents the input for our STDP algorithm, wherein classification is performed.



*Figure 3.6: Sample of Eigenfaces for ORL Data Set*

To sum up, Figure 3.7 illustrates the process of dimensionality reduction and facial feature extraction by using PCA and by obtaining a set of Eigenfaces for the data set.

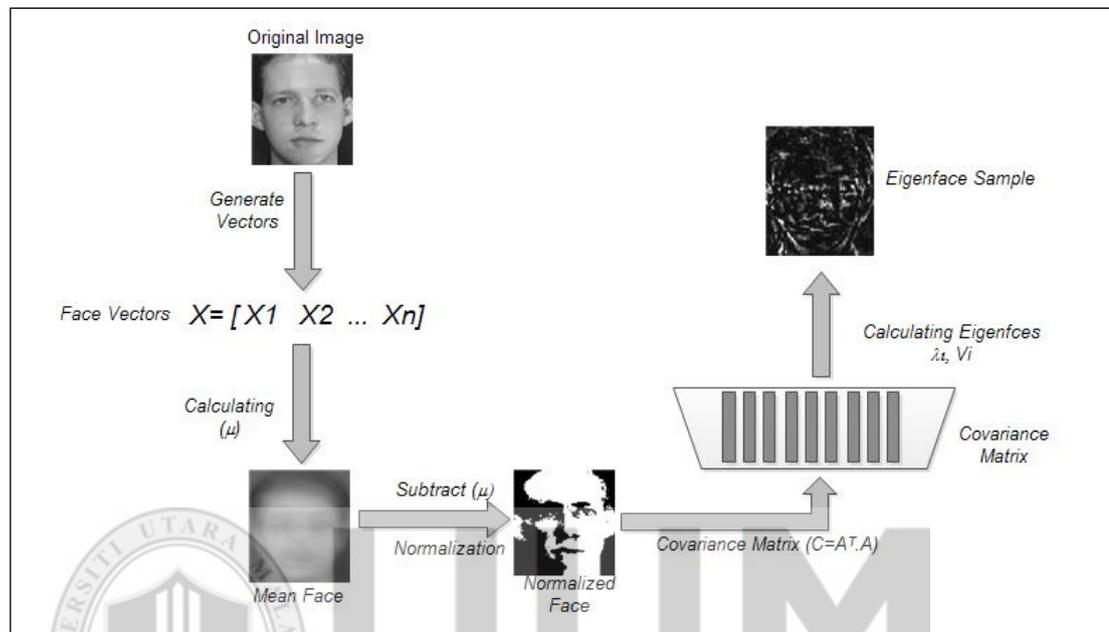


Figure 3.7: The PCA Features Extraction Steps

### 3.4.1.2 Face Features Extraction Using Singular Value Decomposition (SVD)

SVD is a well-known feature-extraction method used in pattern recognition systems, is a mathematical method used to define and order the dimensions of a matrix. SVD is also considered as one of the most efficient tools for data analysis and signal processing [90, 91]. Singular values related to any given matrix generally consist of information that describes noise and energy levels. SVD provides a method for extracting the most significant features of an image. The mathematical formula used to extract the singular values of any given matrix of observation ( $X$ ) with dimensions ( $m \times n$ ) can be formulated as follows:

$$X = U \Sigma V^T \quad (3.7)$$

where  $U$  ( $m \times m$ ) and  $V$  ( $m \times m$ ) are orthogonal matrices, and  $\Sigma$  is an ( $m \times m$ ) of diagonal singular values. The most significant property that makes SVD a robust technique for extracting features from facial images is the stability of the facial image.

To extract facial features from the ORL facial data set, we first reduce each image in the data set to 50% of its original size. The images are resized to increase training and recognition speed, improve recognition accuracy, and overcome the curse of dimensionality [91]. Similar to that in PCA, we must convert the images in the data set into vectors for observations. To extract features, each image is divided into a number of blocks. Each block is treated as an individual matrix and processed to extract its features. A patch measuring ( $L \times W$ ) is virtually created, and it slides over all facial images starting from the top up to the bottom to scan image values and generate a number of blocks [168]. After resizing the images, the dimensions of each image are determined to be (Height = 56) and (Width = 46), as shown in Figure 3.8. Given that the patch width is the same as that of the image, we only need to determine the patch height to achieve block detentions. In the present experiment and according to [91], we consider the height of the patch as ( $L = 5$ ) pixels. Considering that the patch can only move one pixel each time toward the bottom of the image, a number of overlapping blocks are observed. The following formula provides the number of blocks for each image in the data set:

$$B = \frac{H - L}{L - P} + 1 \quad (3.8)$$

Where ( $P = L-1$ ) to ensure that the patch will only move one pixel each time. Hence, each image with dimensions ( $56 \times 46$ ) results in (52) blocks. We know that each block has (230) values, and thus, each image has (11960) values, which is a large number. Thus, we are still far from the optimum number of features, and additional dimensionality reduction processes should be conducted. The goal of SVD is to obtain only a single value for each block instead of (230) values.

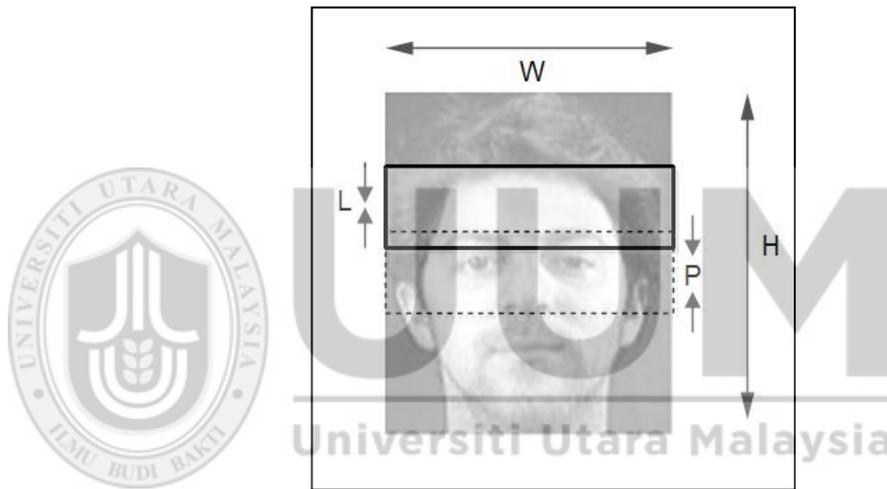


Figure 3.8: Blocks Generation Process

According to [91], SVD has three matrices ( $U, \Sigma, V$ ) for each block. Based on the experiment conducted in [91], three values present the best classification, namely,  $U(1,1)$ ,  $\Sigma(1,1)$ , and  $\Sigma(2,2)$ . Thus, only three single values instead of (230) exist for each block, and the entire image can be represented using (156) values. The number of features has decreased significantly. As previously stated, the goal of SVD is to obtain a single value for each block. Considering that we have three values and that SVD generates continuous values, then quantizing the values of each block

is necessary to obtain a single, unique value that can represent the entire block. Quantization has been adopted widely for generating features in pattern recognition systems. In addition, quantization is an efficient tool for image representation [169-171]. Quantization can be performed via a rounding or truncation process. According to the experiment in [91], the three aforementioned values can be quantized as follows:

- The value of  $U(1, 1)$  is quantized to level (18), i.e. in range of (0-17).
- The value of  $\sum(1, 1)$  is quantized to level (10), i.e. in range of (0-9).
- The value of  $\sum(2, 2)$  is quantized to level (7), i.e. in range of (0-6).

Assuming that  $X_1$ ,  $X_2$ , and  $X_3$  are the three aforementioned values; then, we need to obtain only a single value to represent these three values. To achieve this, we have to find the combination of these values. This combination can be calculated by using the following formula:

$$\text{Block Lable} = X_1 * 10 * 7 + X_2 * 7 + X_3 + 1 \quad (3.9)$$

where the numbers of each block range from (1 to 1260). Figure 3.9 shows an example of dimensionality reduction by using SVD.

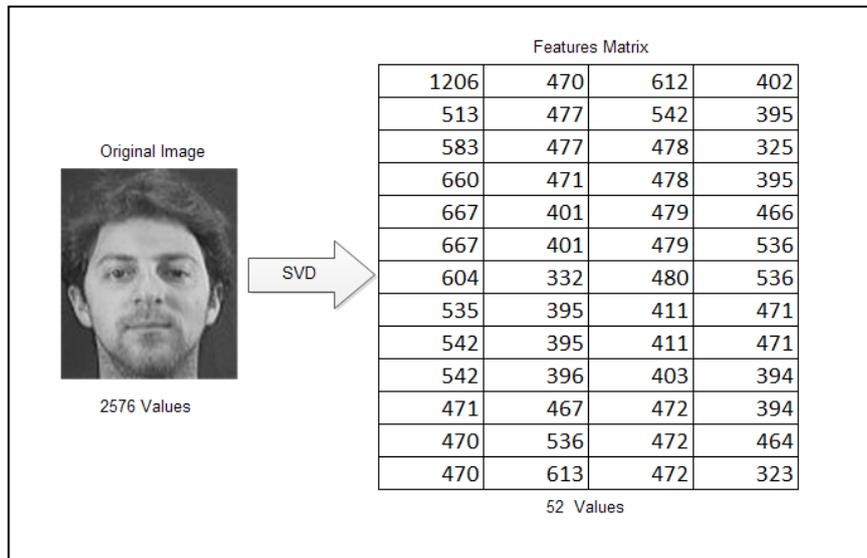


Figure 3.9: SVD Dimensionality Reduction

After performing SVD for the entire data set, we obtain the matrix of the extracted features, which is ready to be trained by using STDP. Figure 3.10 illustrates the overall SVD process for extracting facial features.

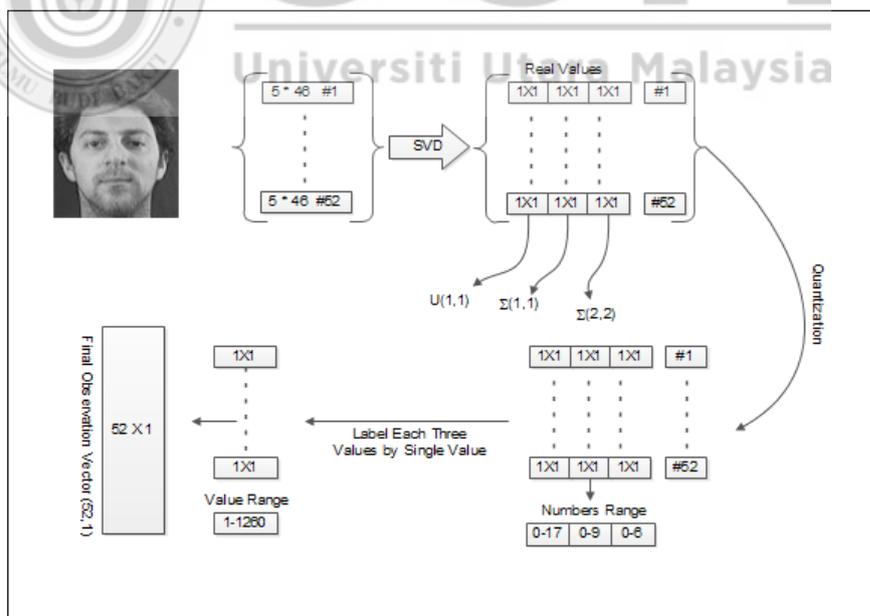


Figure 3.10: Features Extraction Using SVD

### 3.4.2 Speech Features Extraction

WPD is an implementation of the wavelet transform approach [172], which decomposes wave signals into a set of wavelets. WPD is considered to be a generalization of the wavelet decomposition approach that can present a high level of signal analysis [173]. It is indexed by using three commonly interpreted parameters, namely, position, scale, and frequency. For any given wavelet function, a library of bases called wavelet packet bases is normally generated. A particular method to code each signal of all bases is available. The wavelet packets can be used to expand a given signal numerous times. Then, we can select the most relevant decomposition of a given signal. In WPD, the signal is represented by using a tree that is composed of a specific numbers of levels, which are the wave packets, as shown in Figure 3.11. For example, the signal wave ( $S$ ) can be represented by the following ( $S = A_1 + AAD_3 + DAD_3 + DD_2$ ). This representation cannot normally be performed by other signal analysis approaches, thus making WPD an efficient extraction method for speech features.

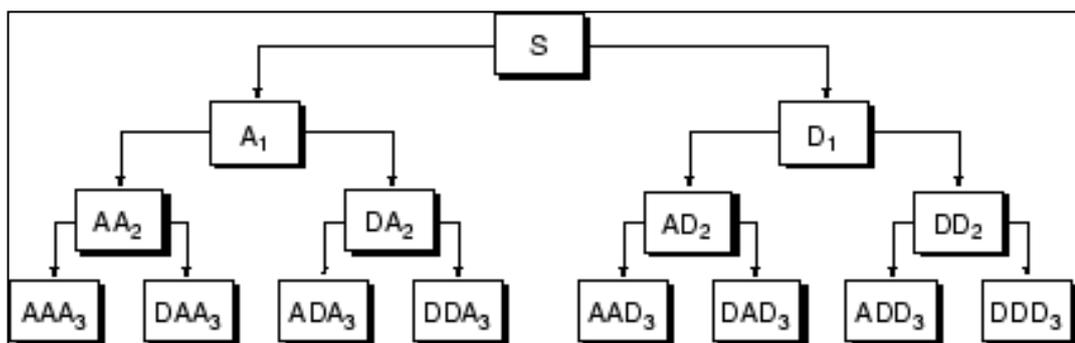
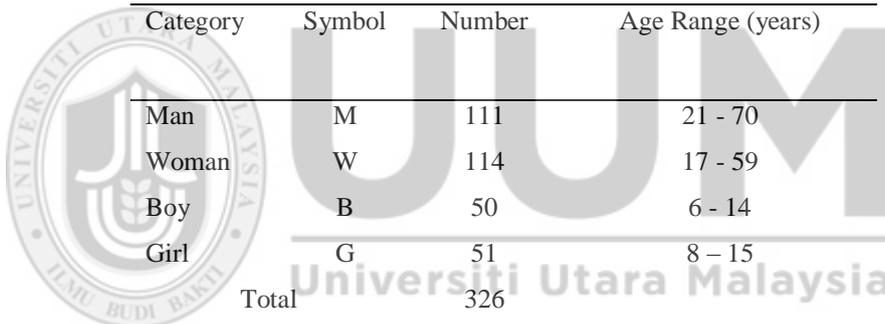


Figure 3.11: Wavelet Packet Decomposition Tree at Three Levels

### 3.4.2.1 Speech Dataset

Similar to facial feature extraction, a speech data set should be used for recognizer training. In this research, we adopt a common speech data set, that is, the TIDigits speech data set [35, 174]. TIDigits is one of the largest speech data sets adopted in numerous studies on designing and evaluating speech recognition systems, such as in [175, 176]. This data set contains approximately 25000 sequences of digits spoken by 326 speakers. The speech samples were obtained from different types of speakers (men, women, and children) in a quiet environment and then digitized at 20 kHz (Table 3.1).

Table 3.1: TIDigits Speakers' Numbers and Ages (Reproduced from [174])



Category	Symbol	Number	Age Range (years)
Man	M	111	21 - 70
Woman	W	114	17 - 59
Boy	B	50	6 - 14
Girl	G	51	8 - 15
Total		326	

The speech samples are represented by pronouncing numbers from “Zero” to “Ten,” as well as the word “Oh,” which is another way to pronounce “Zero.” The speakers are from different areas of the United States. Therefore, the multi-dialect issue has been considered. The data set is divided into 21 dialectal regions. Accordingly, the TIDigits speech data set has a high level of speech variety, thus making it robust. The details of this data set are shown in Table 3.2.

Table 3.2: Description of Dialects and Distribution of Speakers [174]

No	City	Dialect	M	W	B	G
01	Boston, MA	Eastern New England	5	5	0	1
02	Richmond, VA	Virginia Piedmont	5	5	2	4
03	Lubbock, TX	Southwest	5	5	0	1
04	Los Angeles, CA	Southern California	5	5	0	1
05	Knoxville, TN	South Midland	5	5	0	0
06	Rochester, NY	Central New York	6	6	0	0
07	Denver, CO	Rocky Mountains	5	5	0	0
08	Milwaukee, WI	North Central	5	5	2	0
09	Philadelphia, PA	Delaware Valley	5	6	0	1
10	Kansas City, KS	Midland	5	5	4	1
11	Chicago, IL	North Central	5	5	1	2
12	Charleston, SC	South Carolina	5	5	1	0
13	New Orleans, LA	Gulf South	5	5	2	0
14	Dayton, OH	South Midland	5	5	0	0
15	Atlanta, GA	Gulf South	5	5	0	1
16	Miami, FL	Spanish American	5	5	1	0
17	Dallas, TX	Southwest	5	5	34	36
18	New York, NY	New York City	5	5	2	2
19	Little Rock, AR	South Midland	5	6	0	0
20	Portland, OR	Pacific Northwest	5	5	0	0
21	Pittsburgh, PA	Upper Ohio Valley	5	5	0	0
22		Black	5	6	1	1
		Total	111	114	50	51

The TIDigits data set contains speech samples of (\*.wav) type. These samples must first be converted from the analog type, that is, from a wave sound to a set of numerical values. The speech samples are recorded with a sample frequency (SF) of 8000. Each speech sample has many numerical values. Figure 3.12 shows a part of the numerical values that represent the wave file ('1.wav'), which is the voice of the first speaker pronouncing the word "One." For this wave sound, which does

not exceed 3 s, (3600) values exist, which is obviously a huge amount of data, particularly when dealing with large speech data sets.

After applying the WPD method on all the speech samples in the training set, the number of features is reduced dramatically. Instead of having a large number of values, we obtain only (255) numerical values by decomposing level (7). All extracted features are stored in a single matrix where the columns correspond to the number of speech samples that must be trained. In our training set, we select (100) speech samples to train the system. Thus, the number of columns of the observation matrix is (100). The number of rows represents the number of extracted features, which in our case, is (255) values.



0.0006	-0.0035	0.0575	-0.0616	0.0830	0.1280	0.4727	-0.4173	-0.4305	-0.3902	0.5381	0.3769
0.0002	-0.0039	0.0690	-0.0064	0.1232	0.2625	0.6811	-0.4095	-0.5310	-0.4887	0.4326	0.3877
0.0002	-0.0042	0.0800	0.0566	0.1537	0.3322	0.7141	-0.4364	-0.5699	-0.5983	0.1448	0.2516
0.0002	-0.0049	0.0868	0.1230	0.1733	0.3381	0.6436	-0.4916	-0.5757	-0.6242	-0.0880	0.0899
0.0002	-0.0054	0.0903	0.1869	0.1702	0.2869	0.4247	-0.4100	-0.5390	-0.3966	-0.1897	0.0147
0.0004	-0.0058	0.0930	0.2432	0.1440	0.1808	0.1168	-0.1456	-0.4633	0.1419	0.0099	0.0307
0.0007	-0.0063	0.0922	0.2831	0.1019	0.0553	-0.1175	0.1526	-0.2749	0.5838	0.2764	0.0147
0.0008	-0.0066	0.0880	0.3002	0.0418	-0.0604	-0.2938	0.4632	0.0980	0.6055	0.3501	-0.0515
0.0008	-0.0064	0.0808	0.3003	-0.0211	-0.1552	-0.3351	0.7050	0.3926	0.3908	0.3490	-0.1275
0.0012	-0.0061	0.0695	0.2792	-0.0799	-0.2068	-0.1965	0.7085	0.4961	0.0171	0.2475	-0.1819
0.0010	-0.0060	0.0562	0.2382	-0.1306	-0.2188	-0.0020	0.5526	0.4969	-0.2901	0.1641	-0.1335
0.0012	-0.0050	0.0409	0.1918	-0.1633	-0.1986	0.2138	0.2684	0.2991	-0.2428	0.2699	-0.0339
0.0013	-0.0041	0.0230	0.1381	-0.1820	-0.1576	0.4162	-0.1063	0.0606	-0.0146	0.3722	-0.0009
0.0013	-0.0036	0.0053	0.0822	-0.1876	-0.1100	0.4848	-0.3057	-0.0886	0.3215	0.3372	-0.0087
0.0014	-0.0016	-0.0130	0.0357	-0.1842	-0.0609	0.4401	-0.3300	-0.1830	0.6085	0.1850	-0.0225
0.0017	-0.0001	-0.0317	-0.0038	-0.1780	-0.0178	0.3442	-0.2016	-0.0493	0.5626	-0.0375	-0.0271
0.0020	0.0012	-0.0481	-0.0344	-0.1678	0.0267	0.1689	0.1210	0.1870	0.3706	-0.2071	0.0095
0.0022	0.0033	-0.0630	-0.0527	-0.1607	0.0685	-0.0115	0.4145	0.3720	0.1706	-0.1956	0.0800
0.0026	0.0049	-0.0759	-0.0624	-0.1563	0.1136	-0.1129	0.5765	0.5406	0.0161	-0.0648	0.0950
0.0028	0.0062	-0.0855	-0.0677	-0.1529	0.1744	-0.1760	0.6168	0.5253	0.0953	0.0181	0.0693
0.0028	0.0079	-0.0917	-0.0685	-0.1536	0.2298	-0.1975	0.4580	0.3614	0.2430	0.0273	0.0695
0.0028	0.0098	-0.0954	-0.0661	-0.1548	0.2829	-0.1720	0.2003	0.2253	0.2439	-0.0476	0.0970
0.0027	0.0110	-0.0960	-0.0645	-0.1608	0.3181	-0.1541	-0.0016	0.0689	0.1460	-0.1728	0.1605
0.0023	0.0123	-0.0936	-0.0614	-0.1755	0.3015	-0.1542	-0.1381	-0.0391	-0.0529	-0.1658	0.2354
0.0018	0.0137	-0.0913	-0.0569	-0.1942	0.2557	-0.1568	-0.1916	0.0004	-0.3073	-0.0200	0.2465
0.0020	0.0143	-0.0877	-0.0509	-0.2173	0.1630	-0.1567	-0.1420	0.0150	-0.3569	0.0903	0.2067
0.0016	0.0150	-0.0844	-0.0408	-0.2355	0.0439	-0.1408	-0.0806	-0.0437	-0.1703	0.1024	0.1654
0.0015	0.0157	-0.0827	-0.0273	-0.2397	-0.0673	-0.0706	-0.0724	-0.0811	0.0353	0.0256	0.1225
0.0013	0.0153	-0.0831	-0.0096	-0.2238	-0.1674	0.0438	-0.1058	-0.1576	0.1620	-0.0949	0.0726
0.0012	0.0150	-0.0850	0.0115	-0.1645	-0.2195	0.1685	-0.1700	-0.2099	0.1407	-0.1151	0.0094
0.0010	0.0143	-0.0878	0.0316	-0.0767	-0.2194	0.2934	-0.2217	-0.1642	-0.0468	0.0359	-0.0629
0.0007	0.0130	-0.0917	0.0506	0.0328	-0.1830	0.3574	-0.1962	-0.0827	-0.2234	0.2047	-0.1259
0.0006	0.0117	-0.0950	0.0653	0.1657	-0.1078	0.3338	-0.0813	0.0274	-0.2441	0.2889	-0.1554

Figure 3.12: A Part of the Numeric Data for One Speech Sample

Figure 3.13 shows the effect of the wavelet in reducing dimensionality, and the difference between the normal and extracted speech waves.

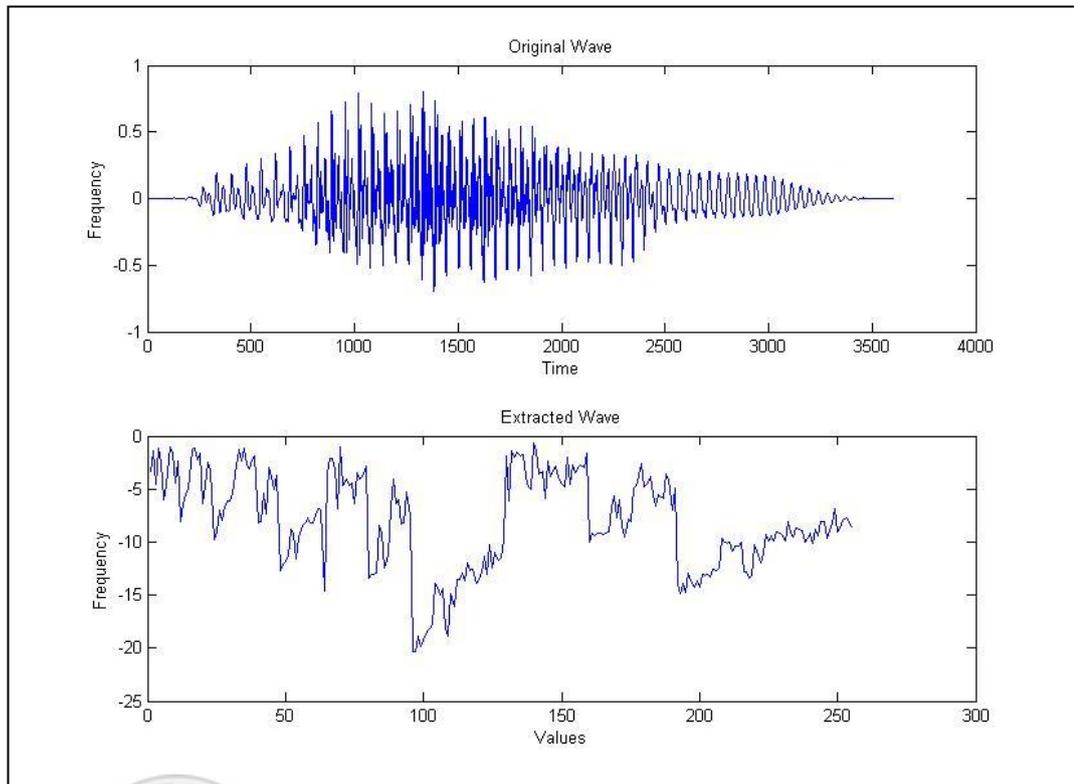


Figure 3.13: Wavelet Extracted Wave vs. Normal Speech Wave

After extracting the speech features, the data are now ready for classification by using STDP.

### 3.5 Phase III: Spike Encoding

In the context of learning experiments, we adopt a recurrent simulation for a neural network composed of 1000 neurons. Out of these neurons, 80% (800) are excitatory neurons ( $N_E$ ) and 20% (200) are inhibitory neurons ( $N_I$ ). This model is developed by Izhikevich [126]. Neuron connectivity is random and sparse within the value of probability ( $p = 0.1$ ), which indicates lack of self-feedback. Each neuron in the  $N_E$  group is randomly connected to 100 neurons, whereas each neuron in the  $N_I$  group is connected to 100 excitatory neurons (Figure 3.14).

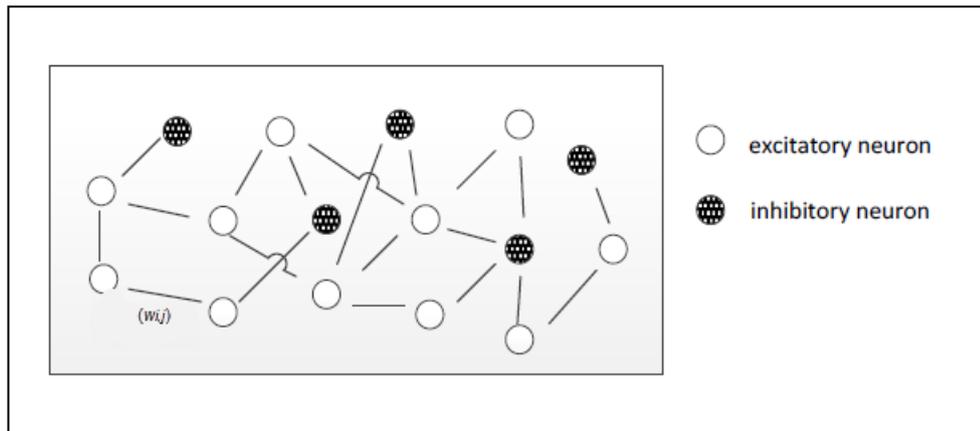


Figure 3.14: Spiking Neural Network ( $N_E = 80\%$ ,  $N_I = 20\%$ )

The delay in synaptic transmission has been set randomly within the range of 1 ms to 20 ms. The weights of synaptic neurons are set similar to those of excitatory neurons, that is, 1.0 mV, whereas those of inhibitory neurons are set to -1.0 mV. According to our network model, learning only affects the connections between  $N_E$  and  $N_E$  as well as between  $N_E$  and  $N_I$ , whereas the rest of the neurons are not updated (that is, not plastic). The range of adjustable weights (the excitatory synapses) is ( $0 \leq w \leq 4.0$  mV).

The population of the excitatory neurons is divided into subpopulations: the  $m$  stimulus groups ( $S$ ), the non-selective neurons ( $NS$ ), and the  $n$  of the response groups ( $R$ ). Each  $S$  consists of 50 neurons to represent a particular stimulus, whereas  $NS$  groups are supposed to be indiscriminate to other stimuli (Figure 3.15).  $R$  groups are composed of 100  $N_E$  neurons. All the dynamic properties of the neuron population are based on the Izhikevich model [131].

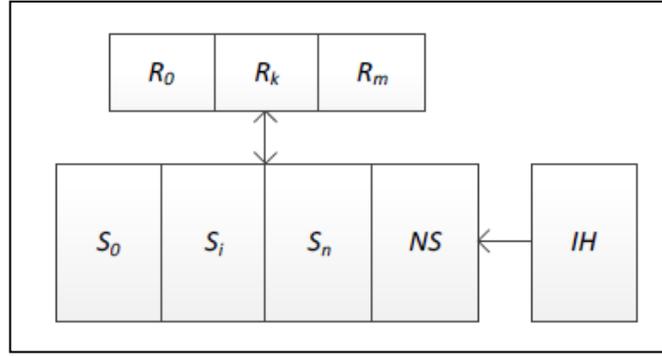


Figure 3.15: Subpopulations of Neuron Stimulus

### 3.5.1 The Rules of Synaptic Plasticity

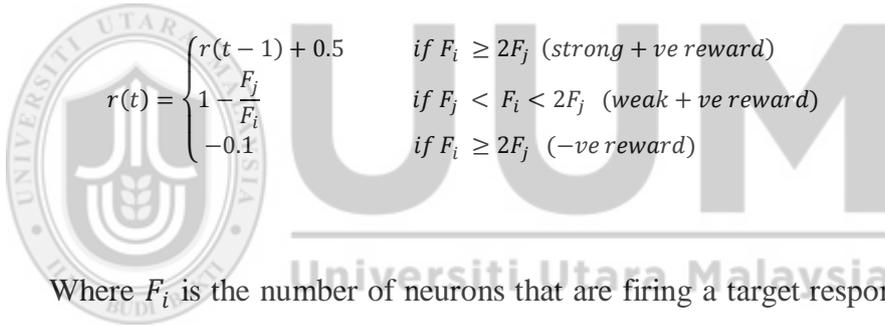
In our model implementation, we use learning according to the reward-based strategy to associate stimulus pairs with a target response. Our study presents a stimulus pair in the form of (*predictor-choice*), which is accordingly represented by ( $S_i, S_j$ ). The *predictor*  $S_i$  is an indicator of the response to the *choice* of another pair. The network is designed based on providing a positive reward in case of a correct response. However, a negative reward is given when the response is incorrect. The reward signal works as a reinforcement signal that strengthens synaptic changes. This signal is derived from the STDP function, as expressed in the following formula:

$$\Delta w_{stdp} = \begin{cases} A + e^{-\frac{\Delta t}{\tau}} & \text{if } \Delta t \geq 0 \\ A - e^{-\frac{\Delta t}{\tau}} & \text{if } \Delta t < 0 \end{cases} \quad (3.10)$$

Where  $\Delta w_{stdp}$  represents the spike-timing-dependent synaptic change that is produced by calculating the difference in firing times between the postsynaptic and

presynaptic neurons ( $\Delta t = t_{post} - t_{pre}$ ). The potentiation of the synapse is performed for each ( $\Delta t \geq 0$ ); otherwise, depreciation occurs. The value of changing the synapse is given by  $(A + e^{-\Delta t/\tau^+})$  for potentiation and by  $(A - e^{\Delta t/\tau^-})$  for depreciation. In such cases, ( $\Delta t = 0$ ) denotes the maximal change and  $\tau$  is the time constant measured in milliseconds (ms).

In terms of learning with STDP, the reinforcement reward signal is derived according to the  $r(t)$  function, which counts the amount of neurons that are firing ( $F$ ) in a particular response group within a time interval of 20 ms from the commencement of choice (3.11).



$$r(t) = \begin{cases} r(t-1) + 0.5 & \text{if } F_i \geq 2F_j \text{ (strong + ve reward)} \\ 1 - \frac{F_j}{F_i} & \text{if } F_j < F_i < 2F_j \text{ (weak + ve reward)} \\ -0.1 & \text{if } F_i \geq 2F_j \text{ (-ve reward)} \end{cases} \quad (3.11)$$

Where  $F_i$  is the number of neurons that are firing a target response. While  $F_j$  is the number of non-target firing group. From (3.12), we calculate the eligibility trace, which represents the summation of  $\Delta w_{stdp}$ . Hence, the change of synaptic weights can be calculated as follows:

$$w = w + \Delta w \quad (3.12)$$

### 3.5.2 Learning Strategy

The associative learning outlines applied in this study to perform visual recognition are adopted from the neuropsychological experiments used on monkeys [132, 177]. In these experiments, the stimuli were represented by a set of different images, and the learning process involved several trials. In each trial, a pair of stimuli was presented, followed by either a reward or a punishment, depending on the nature of the task. If the association between the stimulus pair was significant, then the monkeys received a reward, and vice versa. After several trials, the brain recording of the monkeys indicated that brain activities increased. The monkeys also expected to see the stimulus pair.

In our study, from a population of 1000 neurons, we choose  $n$  non-overlapping groups with 50  $N_E$  each, that is,  $S_i$  ( $0 \leq i < n$ ). Each particular group of neurons represents a stimulus. Other exclusive groups ( $m$ ) that are composed of 100  $N_E$  each are selected to function as response groups  $R_m$ , that is, when  $R_0 = A$  and  $R_l = B$ , where  $A$  and  $B$  represent the labels of the group. When response size is large, the connectivity among the stimulus groups typically increases.

For each simulation, the network model is given a set of pair-response  $(S_i, S_j) \rightarrow R_k$ , with various strategies for pairing according to the task. In the training phase, the stimulus pair (e.g., *predictor* =  $S_0$ , *choice* =  $S_l$ ) and the target response (e.g.,  $A$ ) are selected randomly for each learning trial. All 50 neurons within each stimulus group are stimulated by a super threshold current. The inter-stimuli interval (ISI) represents an experimental parameter that ranges from 10 ms to 50 ms.

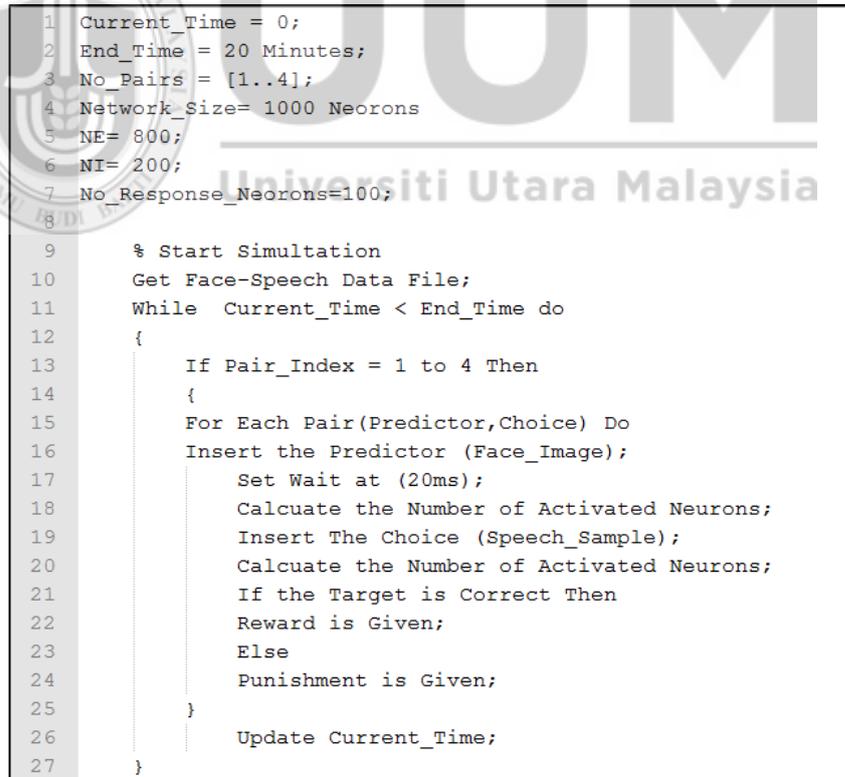
According to the value of a *choice*, we calculate the number of spikes triggered (fired) corresponding to the neurons in the response groups (*A* and *B*) within a response time interval of 20 ms. The next trial is performed upon reaching a delay of 100 ms. According to the initial experiment [131], an interval of 100 ms is the minimum value for stimulation. Therefore, the network is rewarded based on the number of spikes in *A* and *B* within a time interval of 20 ms. To initialize the learning process, we start the learning network with a pregenerated background activity for 100 ms. As stated previously, the background activity is performed by stimulating the neurons randomly by providing a 20 pA current for each ms; otherwise, the network is under an asynchronous state combined with the absence of motivated currents to target a group. When the time  $t = 0$ , the membrane potential is set to  $v = -60$  mV for each neuron, which is slightly higher than the initial membrane potential of the neurons (-56 mV). This process is used to provide several activities to the network before the learning trials and to facilitate the activation of neurons.

When the initialization of the network is completed, we start the learning trials by producing a *predictor* stimulus  $S_i$  within a specific time  $t = t_n$ , then all neurons are stimulated in  $S_i$  with a 1 ms pulse at a 20 pA current. After a specific ISI and at time  $t = t_n + \text{ISI}$ , we stimulate all neurons by applying the same amount of current to the *choice*  $S_j$  for 10 ms to 50 ms. We then select the best ISI according to the preliminary experiment. From the commencement of the choice stimulus, we notice that activation occurs within the first 20 ms. The winner is determined based on the highest number of activated neurons. As stated in (3.11) and according to the

value of activated neurons, we classify the performance of responses as strong positive, weak positive, or negative reinforcement.

### 3.6 Phase IV: Evaluation

The test for the trained network is performed with the same setting as that of the stimulus training. This test involves recalling learned pairs, unlearned pairs, and noisy stimuli. Noisy stimuli can be generated by stimulating a number of neurons randomly with a probability of less than 1.0. The result of the test reflects the average ratio of the performance within a specific number of trials, that is,  $\text{Performance} = (\text{the number of correct calls} / \text{number of trials}) * 100$ . To illustrate the process of our association learning model, Figure 3.16 shows the learning pseudo code.



```
1 Current_Time = 0;
2 End_Time = 20 Minutes;
3 No_Pairs = [1..4];
4 Network_Size= 1000 Neurons
5 NE= 800;
6 NI= 200;
7 No_Response_Neurons=100;
8
9 % Start Simulation
10 Get Face-Speech Data File;
11 While Current_Time < End_Time do
12 {
13     If Pair_Index = 1 to 4 Then
14     {
15         For Each Pair(Predictor,Choice) Do
16         Insert the Predictor (Face_Image);
17         Set Wait at (20ms);
18         Calculate the Number of Activated Neurons;
19         Insert The Choice (Speech_Sample);
20         Calculate the Number of Activated Neurons;
21         If the Target is Correct Then
22         Reward is Given;
23         Else
24         Punishment is Given;
25     }
26     Update Current_Time;
27 }
```

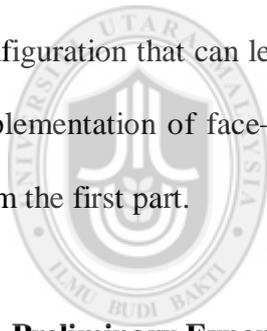
Figure 3.16: Face-Speech Association Learning Pseudo Code

## CHAPTER FOUR

### IMPLEMENTATION OF STDP AND FINDINGS

#### 4.1 Introduction

In this chapter, we propose an associative learning based on the STDP approach by combining facial and speech biometric data. For the simulation network, we adopt the Izhikevich learning model to perform the RL-based face–speech associative learning because this model is simple and does not require any spike template. This chapter has two main parts. The first part discusses the preliminary experiment, which has been performed to determine the optimal learning configuration that can learn associate face–speech data. The second part presents the implementation of face–speech associative learning based on the settings produced from the first part.



UUM  
Universiti Utara Malaysia

#### 4.2 Preliminary Experiment

To initiate the learning model, we have conducted an experiment to test and operate the learning model. We have started this experiment with a number of initial simulations by using a simple, pre-structured network to explore the best simulation parameters to be used later in learning. These parameters include the range of weight values and stimulation to background activity. We have coded the association between two neurons in the same group, as well as between *predictor* and *choice* groups. At this point, all neurons in the network model are not plastic because no learning has occurred yet. We have implemented the network according to a set of

weight values. We then investigate the behavior and properties of our learning model. Afterward, we use associative learning to determine a good firing rate via the STDP approach. All our simulations are conducted using MATLAB programming environment and C<sup>++</sup>. The rules of learning are implanted to the excitatory–inhibitory network by using the Izhikevich neuron model (IM).

#### 4.2.1 The Simulation Results

We train the network by using exclusive stimulus groups. We select 8 overlapping stimulus groups with 50  $N_E$  each. Moreover, 2 exclusive groups of 100  $N_E$  have been selected as response groups  $R_m$ , that is,  $R_0 = A$  and  $R_1 = B$ , where  $A$  and  $B$  are the group labels. Learning is performed as follows:

$$\{(S_0, S_1) \rightarrow A, (S_2, S_3) \rightarrow B, (S_4, S_5) \rightarrow A, (S_6, S_7) \rightarrow B\}.$$

We stimulate all 50 neurons in the predictor group  $S_i$  followed by the stimulation of the *choice*  $S_j$  group. The ISI between  $S_i$  and  $S_j$  is fixed to 10 ms given an average synaptic transmission delay that ranges from 1 ms to 20 ms. During implementation, learning target responses obtained within a correct mapping of pairs are 94.08% for training and 99.9% for testing. Given the high performance of learning by using this configuration, we have implemented this strategy in all upcoming learning simulations.

#### 4.2.2 Inter-stimulus Interval

We conduct this experiment to determine the most significant ISI interval. In the previous experiment, ISI has been set with a fixed value of 10 ms. At this point, we explain how the delay between the *predictor*  $S_i$  and the *choice*  $S_j$  in a pair can influence group responses. Similar to the previous process, the stimulus pair is selected randomly and submitted to the network. Neuron activation is observed at a period of 20 ms, starting from the *case*. The network has been trained based on a set of ISIs that is within the range of 10 ms to 50 ms.

The associative learning performance of the network of the stimulus pair is 82.2% and 91.07% for training and testing, respectively. These results are obtained when  $ISI \leq 20$  ms. However, when the time delay for the ISI exceeds 20 ms, the learning performance is below the acceptable range (Figure 4.1).

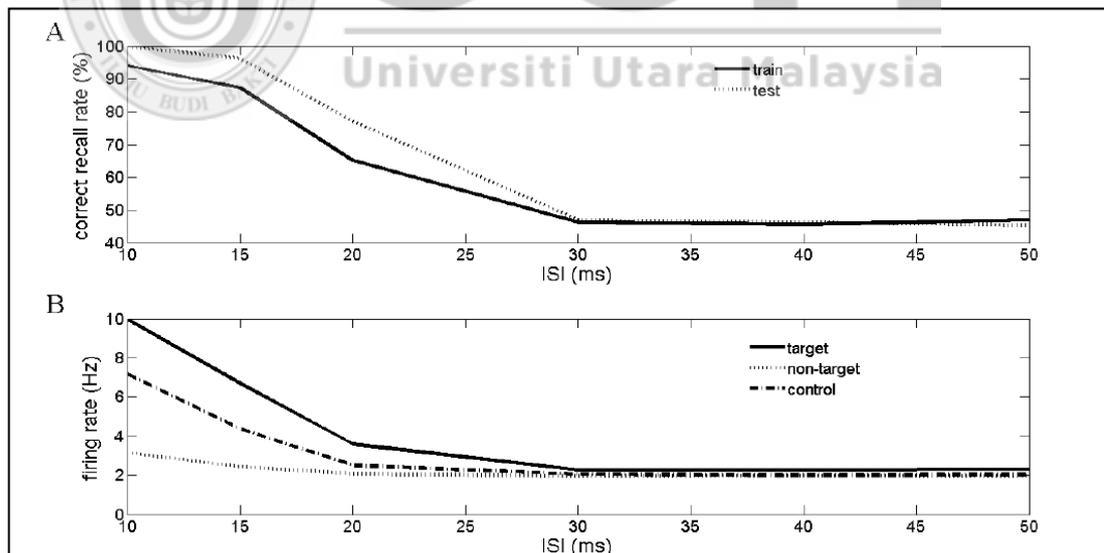


Figure 4.1: : (A) Average Performance when the ISI is within 10 ms to 50 ms. (B) The Firing Rate in the Target Response.

By further modifying the network, that is, by testing the network with unlearned pairs and changing pair position (putting the *predictor* as *choice*, and vice versa), we find that the best performance can be achieved when ISI = 15 ms. After training the network with the following pair responses, we then perform a number of trials. The recall performance average for more than 100 probes is 70% neurons (i.e., 35 out of 50 neurons) required to function as minimal activators with a minimum of 65.48% correct calls. This ratio results from a random selection of neurons.

We then perform a probe based on three conditions by using a selected group (not random). The first condition is *neutral*, in which trial is implemented on learned *choices*  $\{S_1, S_3, S_5, S_7\}$  only, that is, without their *predictors*. The second condition is *congruent*, which is based on the learned stimuli  $\{(S_0, S_1), (S_2, S_3), (S_4, S_5), (S_6, S_7)\}$ . The last condition is *incongruent*, which involves selecting the stimulus pair in the conflict response  $\{(S_0, S_3), (S_2, S_1), (S_4, S_7), (S_6, S_5)\}$ .

After more than 100 trials, the performance rate for the *neutral* condition is 53.93%, whereas the rate for the *congruent* condition is 95.85%. However, for the *incongruent* condition, the performance rate decreases to 42.28%. These results prove that *choice* has to be present with its correct predictor (Figure 4.2).

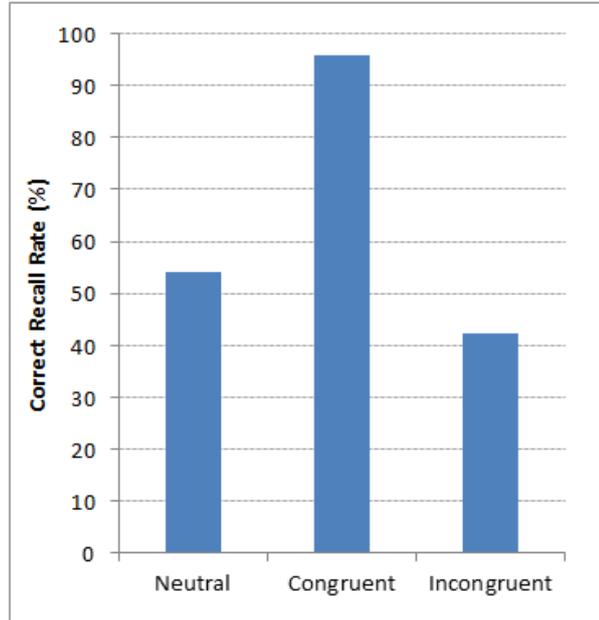


Figure 4.2: The Performance of the Three Recall Correct Rates

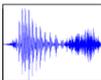
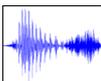
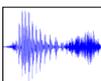
### 4.3 Multimodal Face-Speech Associative Learning

The previous sections were devoted to depict the network of the learning model as well as to describe the parameters and the structure of such model. In this section we are going to describe the implementation of STDP on the multimodal Face-Speech. Since we are dealing with heterogeneous data (face image pixels and speech wave sound), the representation of such non identical data was the main challenge. Furthermore, in this section, we will describe how to encode and combine these heterogeneous data and produce the stimuli groups (pairs  $(S_i, S_j)$ ) in order to train and test the network. In order to train the network with real Face-Speech features, the agent was rewarded if the response shows an accurate matching of the *predictor* (face image) with the *choice* (speech wave).

### 4.3.1 Feature Extraction

As stated in chapter three, the face and speech features have been extracted in order to train the learning network. The image faces were selected from ORL data set, and the speech samples were selected from TIDigits speech samples. For face feature extraction, we have implemented two different methods in order to perform this task which are; PCA-based Eigenface feature extraction, and SVD face feature extraction. On the other hand, for speech feature extraction, we have selected the WPD to extract the most dominant speech features. To train the network, we selected 4 images samples as well as four speech samples, since there is no standard data set for the both faces and speech; we assumed that the first speech sample belongs to the first subject (image face) and so on. (Table 4.1).

Table 4.1: Face-Speech Learning Samples

No	Predictor	Choice	Target
1			B
2			A
3			A
4			B

### 4.3.2 Face-Speech Training

As mentioned in chapter one, our second objective is to evaluate the learning performance of the associative learning based on two different face feature extraction methods. In our study, we have face features that has been extracted using PCA-

Based Eigenface and SVD face feature extraction. On the other hand, we extracted the speech features based on WPD, therefore, there were two different sets of data for the same sample which was illustrated in (Table 4.1) have to be learned.

As we know that the face and speech features were produced with different range of numbers; for PCA method the features range was in (0-255), the SVD features were in the range of (1 to 1260), and for speech features we have got various values since the WPD method depends on finding the logarithmic values for the wave sound encoding. Accordingly, we have performed more processing tasks on the data to be trained and to select only 50 features for each data type. The feature selection is performed as follows:

- A) *Face PCA-based Eigenface Data Selection:* according to PCA feature extraction (chapter 3); each face image was represented by 10304 feature values. Since our network model requires 50 neurons to represent each face sample, thus we selected 50 facial features randomly out of 10304. According to the range of the giving values was at (0-255), we have converted the values into (0, 1), where 1 represents the firing of the neuron and 0 represents the non-fired neurons.
- B) *Face based on SVD Data Selection:* for this feature extraction method, we have got 52 features to represent each single face image. We selected 50 features randomly in a way similar to the previous method. As we mentioned in chapter three, the range of face features was from 1 to 1260. Thus we selected the features that are greater than the average of these values to represent the activated neurons and the rest are the non-activated neurons.

C) *Speech Data Selection*: from the data that produced by WPD, we noticed that there were a large variety of the data given, in addition the majority of the data were negative values due to the logarithmic operations. Accordingly we have selected the values that represent the neuron activation according to the following ranges:

1.  $-10.0 < X < 0$
2.  $-15.0 < X < -5.0$
3.  $-20.0 < X < -10.0$

Where  $X$  represents a single value of speech features, the value of (1) -which will be selected according to the above ranges- represents the neuron activation, while the (0) means the neuron is off.

After the face-speech data have been prepared, we merged them in one data file to be presented in the simulation process (see appendix A).

#### **4.3.3 Speech Encoding**

Since we have one speech feature extraction; then the next step is to encode the speech values in order to perform the learning procedures. The aim of this experiment is to select the optimum range of speech values as well as to state the most efficient face feature extraction method. This task has been implemented according to the network settings in (Table 4.2).

Table 4.2: The Initial Settings of the Association Learning

No.	Parameter	Value
1	Network Size	1000 neurons
2	Number of Excitatory Neuron	800 neurons
3	Number of Inhibitory Neuron	200 neurons
4	Number of face and speech features	50 features
5	Number of response neurons	100 neurons
6	Number of pairs	4

We ran the simulation using C<sup>++</sup> and by implementing five simulations for each face-speech based on PCA, and Face-speech based on SVD (see Appendix B). After the simulation ends; we used Matlab to analyze the results of the learning model for the selected data; the rates of performance for both training and testing are listed in (Table 4.3). The winner is produced according to a target that can be seen as follows; [1 0 0 1] or [B A A B], which means that the winner for the first pair supposed to be B, and so on.

Table 4.3: The Face-Speech Learning Performance

Group No	Rang of Speech Values	Face PCA-based Eigenface Performance (%)		Face based SVD Performance (%)	
		Train	Test	Train	Test
1	$-10.0 < X < 0$	54.13	55.33	45.47	42.00
2	$-15.0 < X < -5.0$	50.54	47.33	41.19	44.66
3	$-20.0 < X < -10.0$	44.16	39.33	39.32	36.66

Based on the result given by the learning experiment which as stated in the table above, it is clearly can be seen that the performance of the data set that produced by PCA-based Eigenface was better than the SVD based approach and in all data ranges (Figure 4.3). In addition, we can see the speech values in the range of  $(-10.0 < X < 0)$  has produced the best performance in comparison with other ranges.

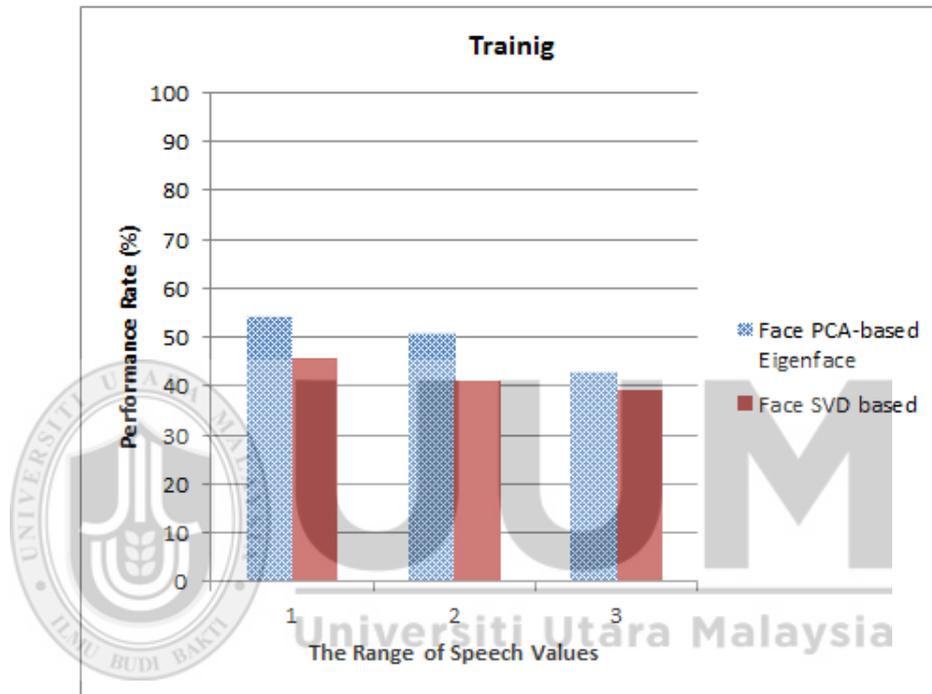
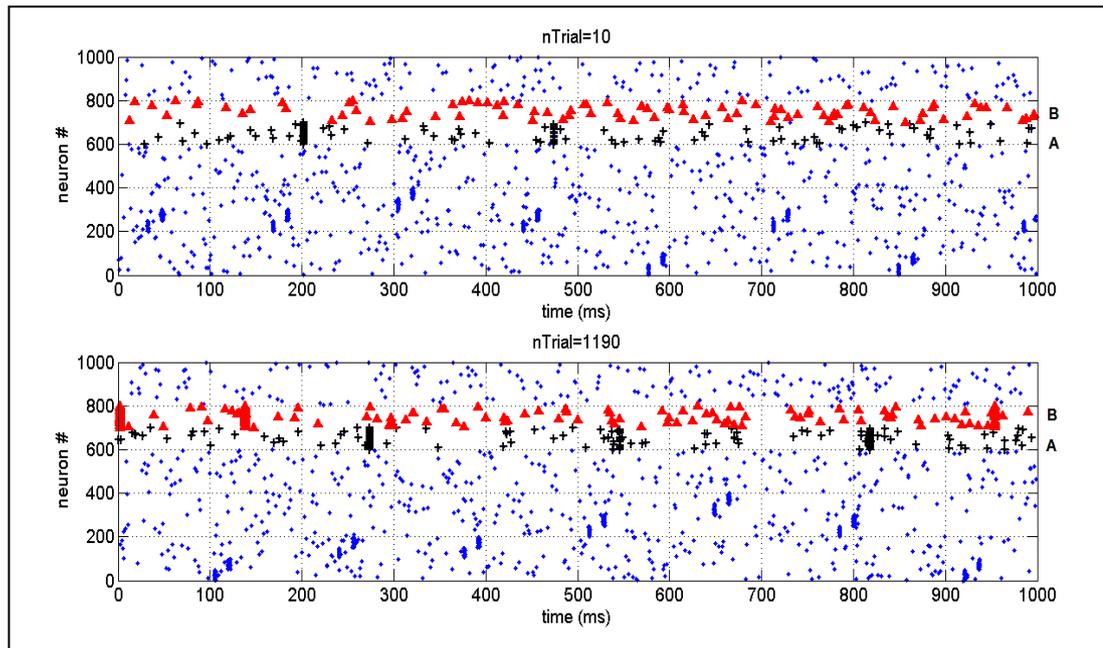


Figure 4.3: The Learning Performance for the Face-Speech Model

According to the result obtained, we can conclude that the PCA-based Eigenface feature extraction approach presents more significant facial features in comparison with SVD-based approach due to PCA's robustness and efficiency in terms of face feature extraction. Furthermore the quantization process that has been applied at the end of SVD face feature extraction (Chapter 3), was caused loosing of features, where the quantization in general is known as lossy data technique [178, 179], the thing that may cause missing of many features from the face which might

be important. Also the range of the data for the SVD method was (1 to 1260) which consists of a large variety in terms of values' range, in contrast to the PCA approach which consists of less range of values that are easier to represent since these values related to a real value of image's pixel. Therefore and according to the results; we can say that the PCA-based Eigenface feature extraction has more positive impact on our learning model; accordingly we have come out with the objective (*a*) of this this research.

In spite of PCA-based approach produces better performance, however, this performance rate is still low and not that much satisfied thus it can be classified as low positive performance. The train value (54.13%) means that the model has a low level of accuracy in terms of targeting the subject. In addition, from Figure 4.4, it is clearly can be seen that there is low activation progress if we compare between the trial 10 and the trial 1190 which is the last trail in each simulation. While there should be a noticeable enhancement within the trials progress. Hence, in order to get a better performance, there must be a number of improving steps that can lead to enhancing the accuracy.



*Figure 4.4: Spike Raster Plot for the Network Activity after a Number of Trials in One Simulation*

For more details, Figure 4.5 illustrates the four pairs' state according one simulation. As we can see that the winner state reflects the poor performance for such experiment due to the poor presentation of the winner according to a given target. Where there should be more significant targeting for the response.

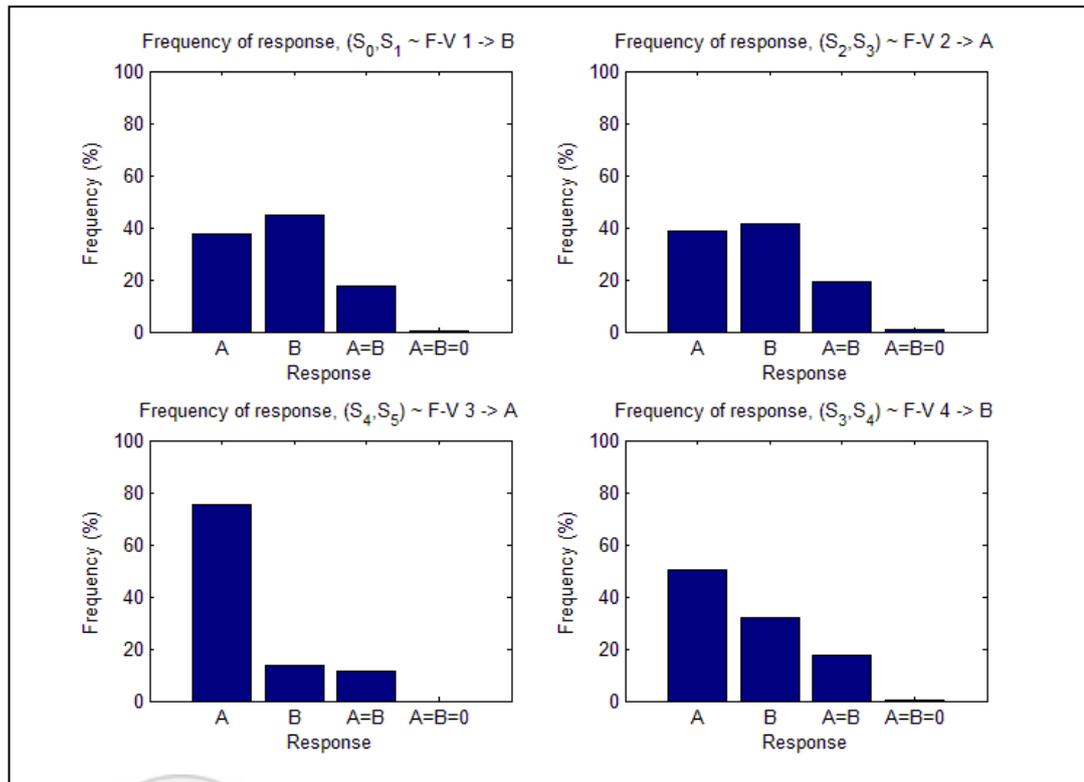


Figure 4.5: Winner State Details for Four Pairs Based on One Simulation

According to the results given in the previous experiment; we implemented a number of experiments with different network settings in order to explore the learning behavior.

#### 4.3.4 Implementation of Learning (Face-Speech Features = 100, Number of Response Neurons = 100)

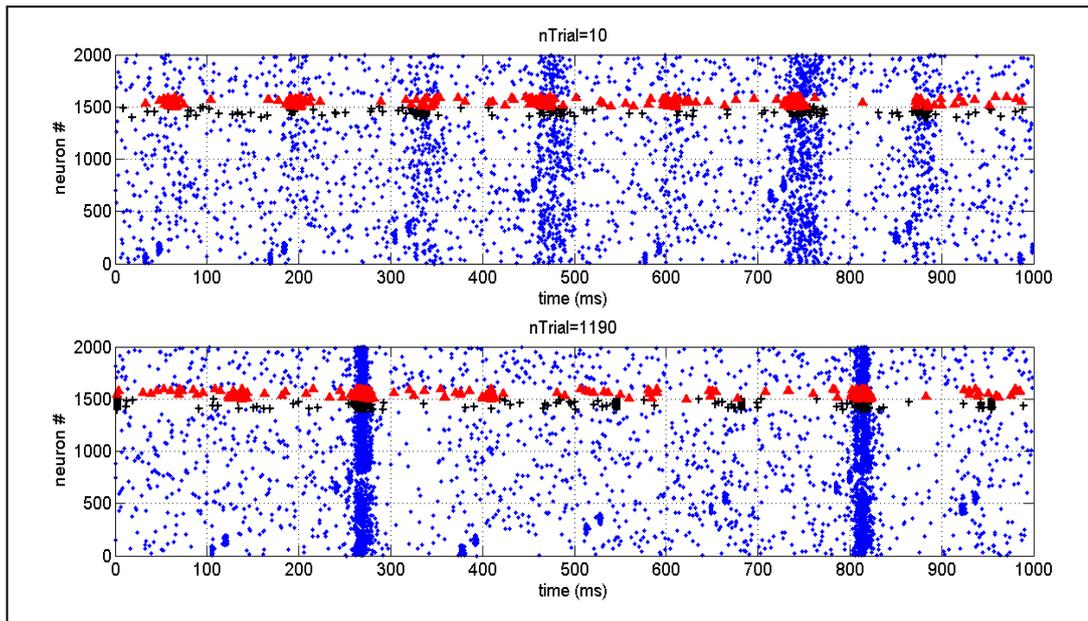
According to the results given in the first simulation, we continued with our simulation according to the model that gave us a better result. Since the PCA approach and the speech data with a range  $(-10.0 < X < 0)$  have presented the best performance; we have concentrated on this configuration in order to enhance the performance of the model. Accordingly, we increased the network size as well as

increase the number of features for both biometrics up to 100 features for each. The new setting of the model can be seen in (Table 4.4).

*Table 4.4: New Settings for the Learning Experiment Parameters*

No.	Parameter	Value
1	Network Size	2000 neurons
2	Number of Excitatory Neuron	1600 neurons
3	Number of Inhibitory Neuron	4000 neurons
4	Number of face and speech features	100 features
5	Number of response neurons	100 neurons
6	Number of pairs	4

As long as we are going to use 100 neurons to represent the face and speech data; we have selected 100 features for each of them instead of 50. This task can help to represent more face-speech features within the learning model. We performed the training task on these settings, and the results were 69.53% and 70.00% for training and testing, respectively. The results show a significant enhancement of the overall network performance. That means, enlarging the network size as well as increase the number of neurons within the response groups has a big impact on the network performance. That means the network starts to discriminate between the different predictor-choice pairs. Figure 4.6 shows the difference in the learning between the first trial and the last one of the training simulations. We can see that the number of activated neurons has been increased significantly which reflects the learning enhancement due to the new parameters.



*Figure 4.6: Spike Raster Plot for the Learning Simulation with 100 Features and the Number of Response Neurons=100*

To highlight the performance inside each simulation in the training phase Figure 4.7 shows the simulation results and the values for each response (A and B) in terms of the winner strategy and how the pairs being respond to the variety of *predictor-choice* parameters in terms of target the correct choice. As stated in this figure we can see the enhancement.

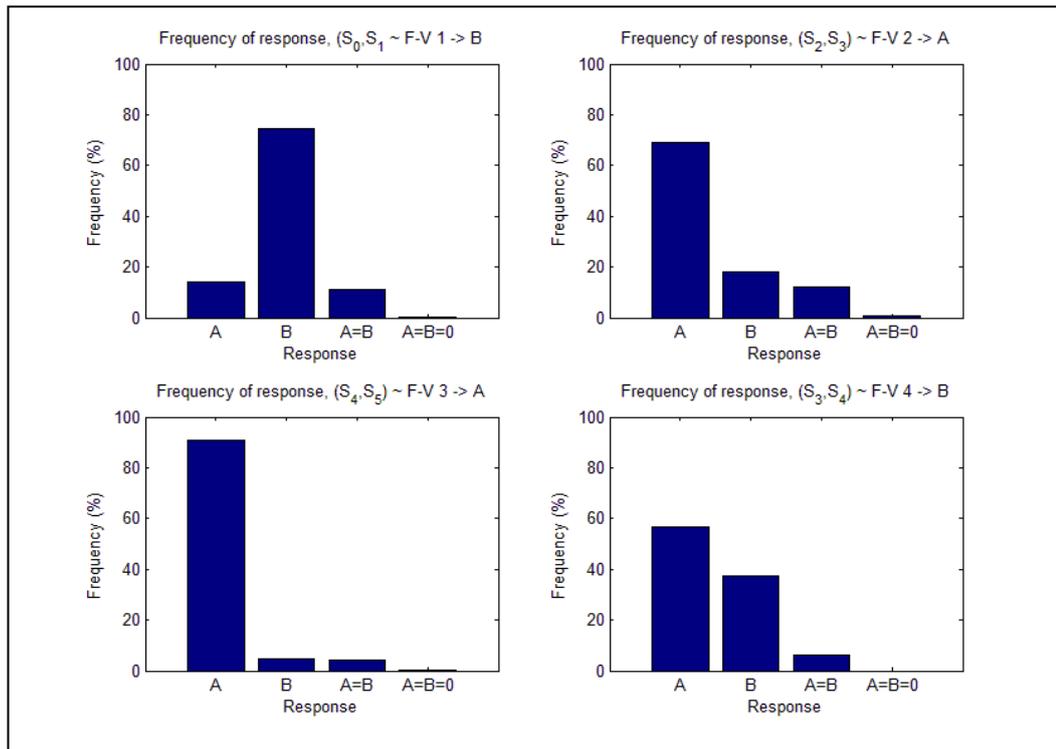


Figure 4.7: Winner State Details Based on the Number of Features= 100, Number of Response Neurons=100

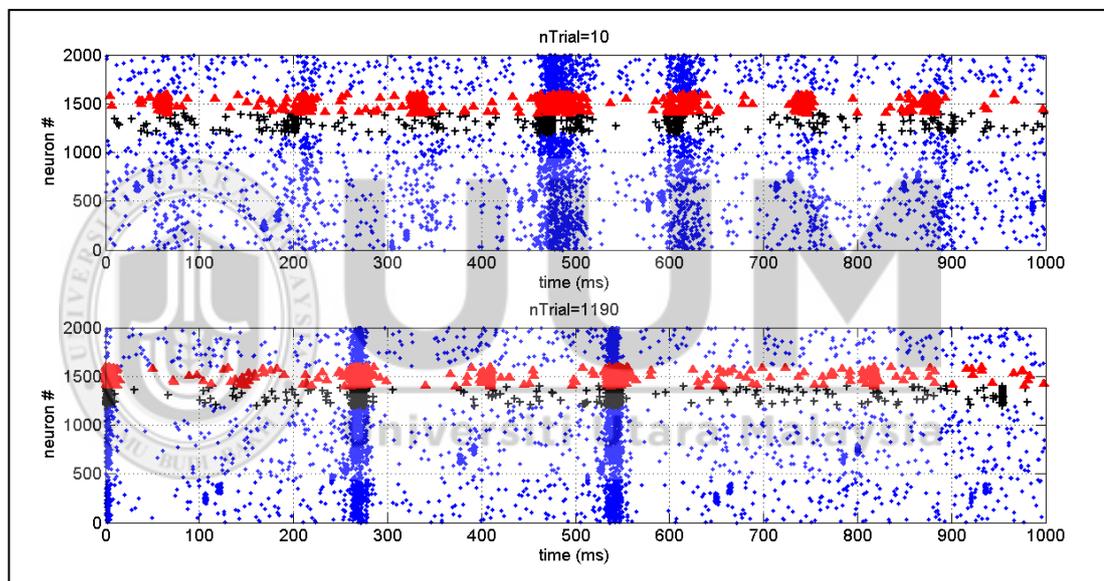
In order to test the model with more variety of parameters, we have implemented one more experiment. The purpose of this experiment is to clarify and describe the learning behavior of such model.

#### 4.3.5 The learning Implantation (Number of Response Neurons = 200)

As we mentioned, the major aim of this experiment to keep tracking the network performance. Thus we did a little adjustment to the model. We kept the number of neurons at 100 neurons and accordingly, we still keep using the amount of feature with 100 face-speech features to be represented by the network. The size of the network is 2000 with 1600 excitatory neurons and 400 inhibitory neurons. We

changed the number of response groups, instead of 100 neurons; we are going to use 200.

We performed the simulation with the previous settings and the results were 73.88% and 71.33% for training and testing, respectively. And here we can see the enhancement of the performance rate. That means the response group has positively influenced the performance and that can be seen clearly in Figure 4.8. Enlarging the number of response neurons contributes to reduce the level of confusion since there are many neurons that can be used to represent the pairs.



*Figure 4.8: Spike Raster Plot for the Network Activity with Number of Features=100,  
Number of Response Neurons=200*

Also the performance can be seen with more details in Figure 4.9, which shows high accuracy in terms of correct response.

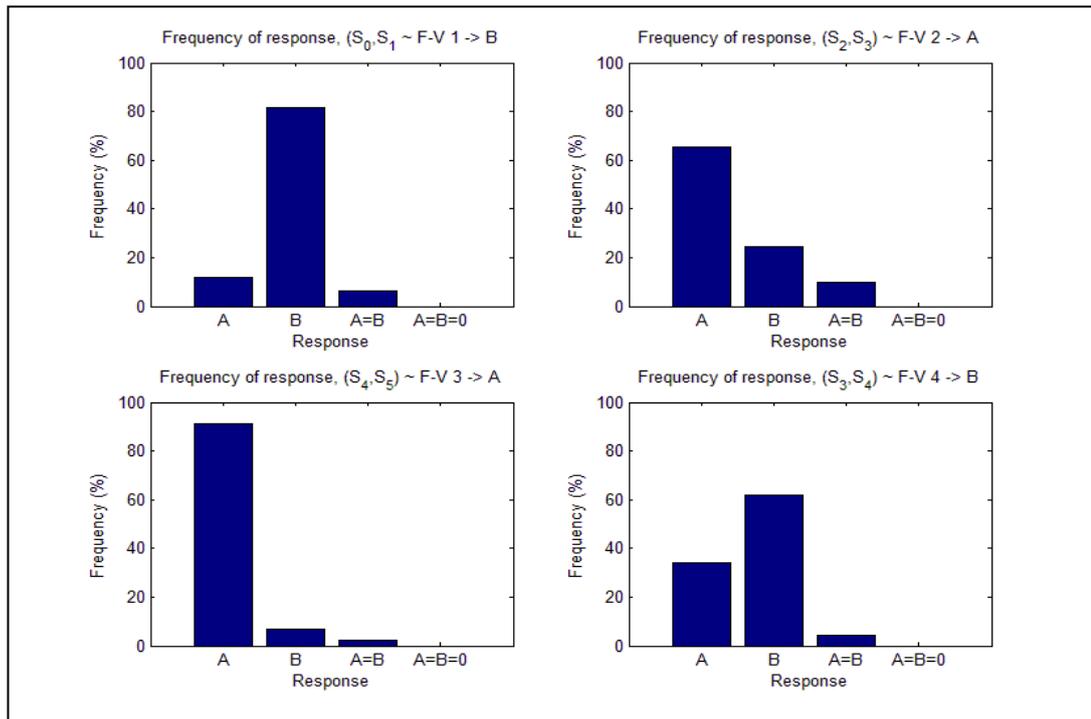


Figure 4.9: Winner State Sample According to (Number of Features=100, Number of Response Neurons=200).

We also performed a small adjustment and performed another simulation. We used the same network except increasing the amount of response groups up to 250 instead of 200. The results of the performance have shown better performance. The results of training and testing were 77.26% and 82.66% respectively. The performance increases again (Figure 4.10 , Figure 4.11)

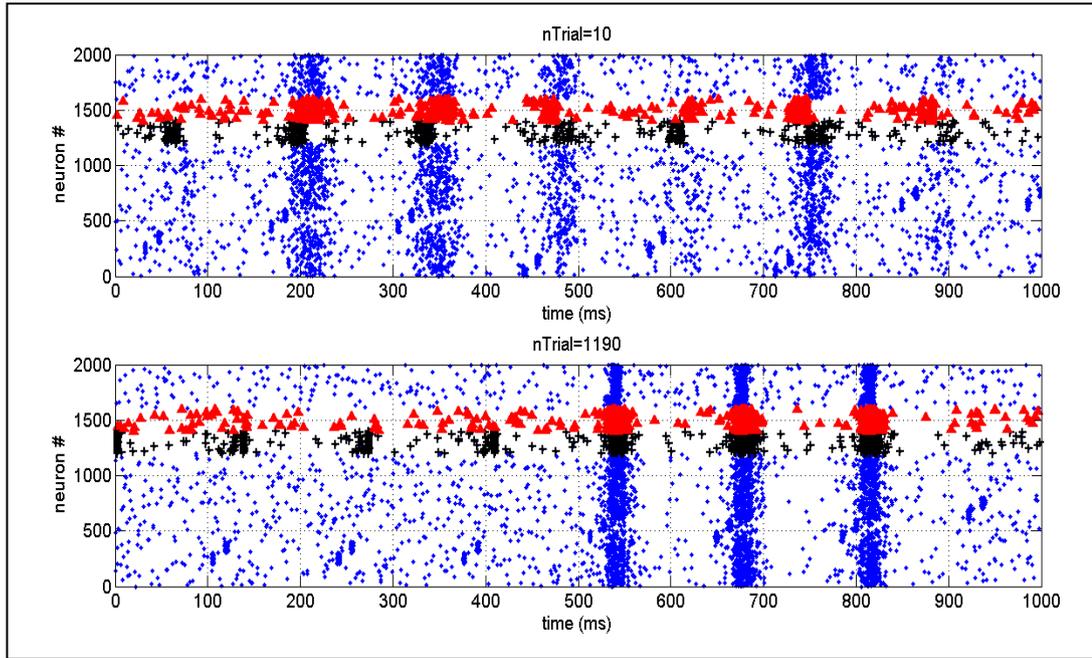


Figure 4.10: Spike Raster Plot for the Network Activity (Number of Response Neurons=250)

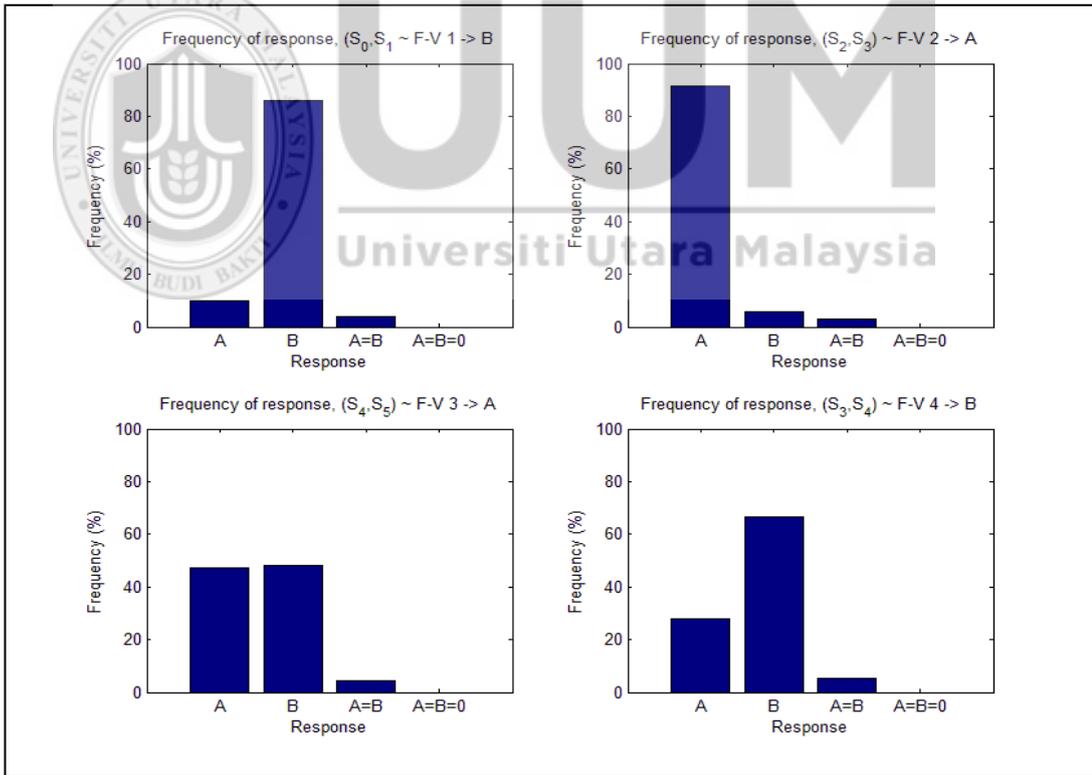


Figure 4.11: Winner State According for a Simulation with Response Group Neurons = 250

## 4.4 Discussion

From the previous sections, we have seen how the parameters could influence the learning performance. According to the feature extraction; we found out that the face features which have been extracted using PCA-based Eigenface have produced a better result than the SVD-based method. This is due to the SVD behavior which is based on a lossy technique of vector quantization. In terms of speech data that have been used to activate the neurons which are in the range of  $(-10.0 < X < 0)$  have presented the best performance in terms of learning and testing. According to that we have used this range in all our experiments (Table 4.5).

Table 4.5: Summary of the Learning Parameters and the Correspond Performance

Experiment	Network Size	Number of Excitatory Neuron	Number of Inhibitory Neuron	Number of Face-Speech features	Number of Response Neurons	Number of Pairs	Performance (%)	
							Training	Testing
1	1000 neurons	800 neurons	200 neurons	50	100 neurons	4	54.13	55.33
2	2000 neurons	1600 neurons	400 neurons	100	100 neurons	4	69.53	70.00
3	2000 neurons	1600 neurons	400 neurons	100	200 neurons	4	73.88	71.33
4 <sup>(*)</sup>	2000 neurons	1600 neurons	400 neurons	100	250 neurons	4	77.26	82.66

For the first network setting; the performance rate was low; however, adjusting these settings have improved reliable performance which is proven by a number of experiments (Figure 4.12). The big impact was obtained by adjusting the

\* These parameters presented the best performance according to our learning; hence we used these settings in our real data learning in Chapter 5.

amount of response group which can reduce the neurons confusion by assigning more neurons in order to target the correct choice, thus, we can see that the performance has been increased steadily from the first experiment until the last one. So for the first experiment we have got 54.13% of performance, while after performing four experiments with different network setting; the performance increased to 77.26%. And for testing, the performance has increased from 55.33% up to 82.66%, which indicate a good enhancement and better performance (Figure 4.12).

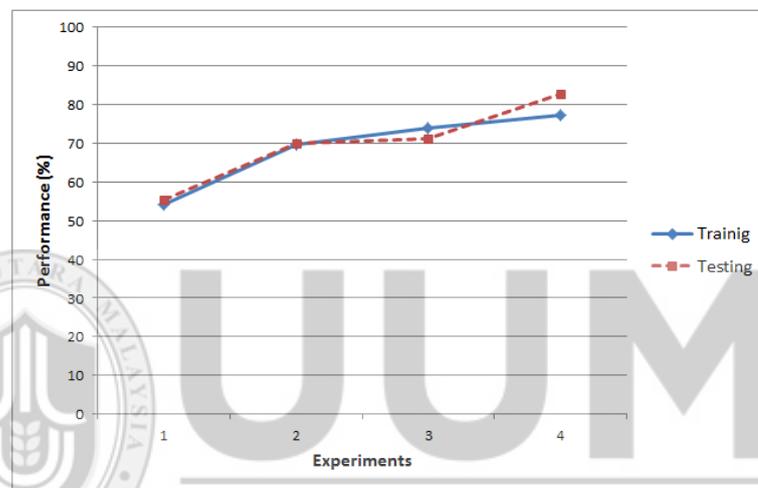


Figure 4.12: The Training and Testing Enhancement within Four Experiments

In spite of expanding the network structure has big influence on the performance rate; it requires more storage resources to be implemented, the issue that has to be taken into consideration where more neurons in the network require more resources. In other words, each face image and speech signal requires a number of neurons to be presented (according to our model 50 or 100). The set of neurons works like a sensor that corresponds to the face or speech. Hence increasing the number of responding groups is similar to adding more sensors which constrict the network structure expansion.

## **CHAPTER FIVE**

### **LEARNING IN REAL WORLD**

#### **5.1 Introduction**

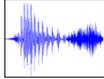
The previous chapter presented a simulation results to implement the reinforcement learning approach in order to achieve the face-speech authentication. We used a sample data which are available for researchers in order to design and test biometric-based systems. Since we are implementing face and speech biometrics; and there was face and speech data from different sources; therefore, we used a face image samples from ORL data set and speech samples from TIDigits dataset. In the implementation stages, we joined these samples by assuming that the first speech sample belongs to the first subject face and so on, this was the strategy that we followed to construct the response group. In this chapter, we present an implementation of our network learning model based on real data that we have collected by capturing the image face and recording the speech wave for real subjects.

#### **5.2 Data Collection**

Our face and speech data set have been collected from a group of people from international student who are studying a postgraduate in Universiti Utara Malaysia (UUM). We captured face images of these subjects as well as we have recorded a speech sample of those subjects, the speech sample involved the speaking of the number “one”. The devices that have been used in order to collect the data as

follows; to capture the face biometric, we used the camera with the specification; Sony *a77*, lens Sony 16-50 2.8mm, 24mp, and for speech collection we used a Dell laptop compatible microphone. Table 5.1 shows the collected sample data.

*Table 5.1: Real Dataset for Training and Testing*

Stimuli No	Face Image (Predictor)	Wave Sound (Choice)	Target
1			B
2			A
3			A
4			B

### 5.3 Face-Speech Feature extraction

In usual, there should be feature extraction process for the data before implementing the learning. For face biometric, we used the PCA-based Eigenface approach due to its efficiency in terms of extract the face features, and depending on the results that have been produced in chapter four. For speech biometric; we have used the WPD approach to extract the speech feature.

### 5.4 Learning Implementation

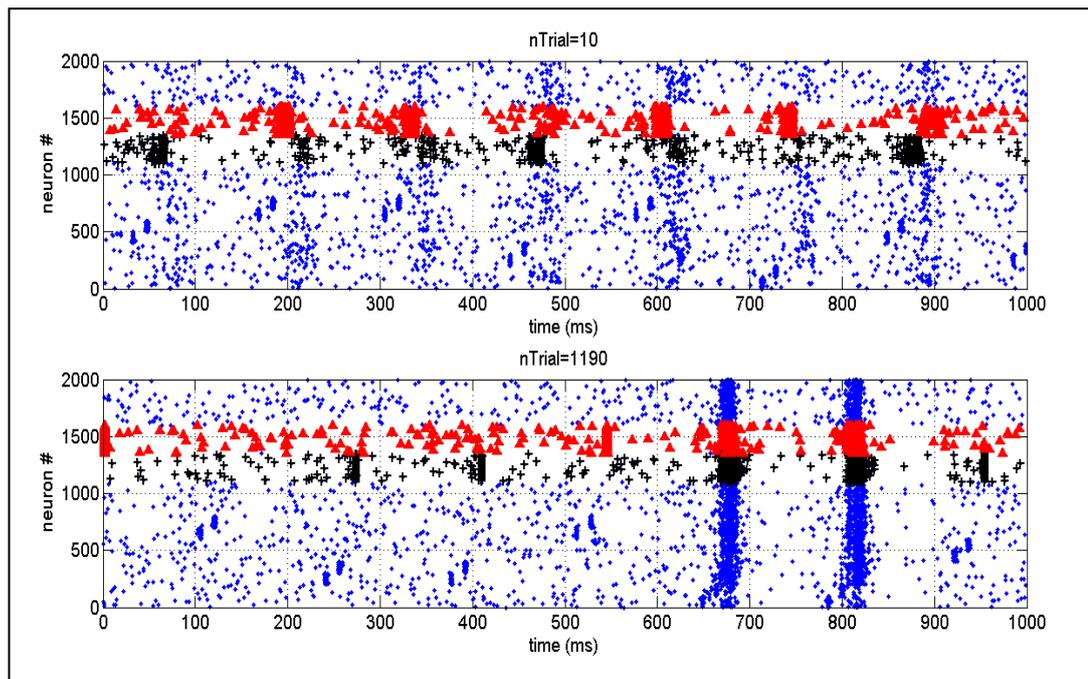
After the face and speech features have been extracted; we performed the learning task. According to (Table 4.5); we used the network settings which have presented the best performance according to our simulations. These settings are illustrated in ()

*Table 5.2: Simulation Parameters for Real Data Learning*

No.	Parameter	Value
1	Network Size	2000 neurons
2	Number of Excitatory Neuron	1600 neurons
3	Number of Inhibitory Neuron	4000 neurons
4	Number of face and speech features	100 features
5	Number of response neurons	250 neurons
6	Number of pairs	4

The reason of choosing this setting is because of the high performance that we have achieved from the previous implementation. For the learning task, the network model has been trained to associate a pair of face-speech with a target response, *A* or *B*. After we implemented the learning process; the results of the performance were 79.11% and 77.33% for training and testing, respectively. This illustrates that the network can learn the pair and produce correct response.

The results showed that our model can perform well with real data that have been collected, which proves that this model can be implemented and tested for any set of face-speech data, In addition, this step shows the possibility of implementing other types of biometric with other applications. To show the activities during the learning process; Figure 5.1 shows the stimulation activities that have been recorded during the simulation of the training process. It is obviously that the activation is high and the neurons are responding to the target the choice starting from the beginning of the training until the end of this process.



*Figure 5.1: Spike Raster Plot for the Network Activity during the Simulation of Real Data Experiment.*

To show the respond performance in terms of targets the choice; Figure 5.2 shows the winner state and the operation of responding to the pairs. As we can see that for each target *A* or *B*, our model responds to come out with the correct response. This indicates that this model is produced a good level of accuracy and efficiency.

From all aforementioned sections as well as chapters, we conclude that the reinforcement learning can be performed within the biometric technology. This can produce more sophisticated authentication systems due to the adoption of SNNs approaches. Encoding with STDP can be extended for more biometrics in terms of performing multimodal authentication.

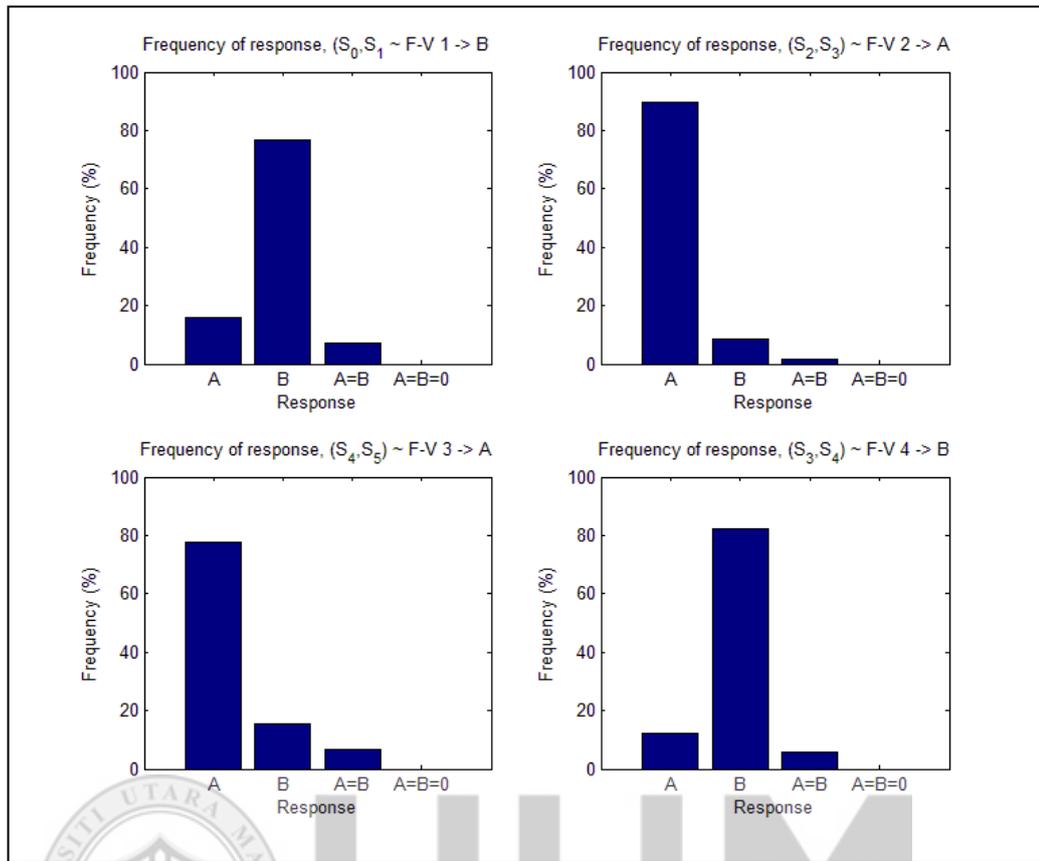
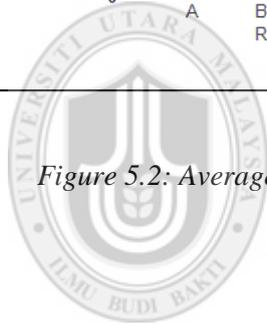


Figure 5.2: Averaged Percentage of Correct Recall with Real Data Experiment



UUM  
Universiti Utara Malaysia

## CHAPTER SIX

### DISCUSSION AND CONCLUSION

#### 6.1 Introduction

The previous chapters have presented the whole body of our research starting from identifying the problem, exploring the theoretical background, and implementing the spike encoding processes and associative learning for face and speech biometrics. This chapter concludes our research by presenting the research objectives that have been achieved. Basically this chapter highlights the research objectives, methods, as well as outcomes. The other part of this chapter describes the future trends of our research.

#### 6.2 Conclusion (Objectives Achieved)

In our study, we tested and compared two facial feature extraction methods which are: PCA-based Eigenface and SVD (Chapter 3, Section 3.4). According to the experiments; extracting the face feature using PCA-based Eigenface has presented better performance in comparison with SVD-based approach (Chapter 4, Section 4.3.3). This step led to come out with objective (*a*). For speech feature extraction, we used the WPD method to extract the most dominant speech features (Chapter 3, Section 3.5). By applying this process we come out with objective (*b*).

The proposed network model is an associative learning model based on RL and uses SNNs (Chapter 4). The learning rules were based on neuron firing rate and

spike timing according to the STDP approach. By developing the reinforcement learning for face and speech biometrics we come out with the objective (c). According to the results given by different experiments; we conclude that the number of response neurons has the biggest impact on the performance since in reduce the level of confusion (Objective (d)). We implemented the association learning on real data, and the results were similar to the previous performance (Chapter 5).

Although expanding the network structure significantly influenced the performance rate, this expansion requires additional storage resources. This issue must be considered because more neurons in the network require more resources. That is, each facial image and speech signal requires a number of neurons to be represented (according to our model, 50 or 100). The set of neurons functions like a sensor that corresponds to facial image or speech. Hence, increasing the number of responding groups has a similar effect as adding more sensors, which constricts network structure expansion. Thus, further investigations on encoding and feature extraction are required to improve recall rate.

The high performance of the proposed learning model opens possibilities for more applications in biometric authentication. The ability to combine heterogeneous data (face–speech) adds a new characteristic to STDP by incorporating authentication capability into the learning context. Our research findings demonstrate a new method with high response performance in encoding facial and speech biometrics.

### 6.3 Future work

Although the proposed model can be classified as a biometric learning approach, it can also be expanded for further implementation. Other stimuli need to be encoded to expand the amount of memory. In our model, we represent the stimulation structure of a fixed number of neurons (excitatory and inhibitory) with a limited degree of overlapping among the neuron groups. This setup imposes certain limitations to our model, particularly when dealing with large-scale applications. However, this problem can be overcome by adopting a network with polychromatic neuron groups. Such model allows each neuron to be related to multiple groups within a variety of synaptic transmission delays, and thus, increase the amount of memory because the model relies only on a few neurons to respond to a specific pair.

Even though we mentioned the study limitations, also we have highlighted the most significant advantages of our learning model, where it can learn a heterogeneous data which are encoded and combined within on the pair, and apply the learning procedure in a simple way using STDP that come out with a correlation. Despite these limitations, our learning model possesses several advantages, such as its ability to learn heterogeneous data that are encoded and combined within a pair, and to apply the learning procedure in a simple manner by using STDP to correlate spike timing with the rate of firing. Furthermore, our learning model can exhibit the difference between two facial feature extraction methods. Our model uses face–speech biometrics, that is, it captures such biometrics by using a camera and a microphone, which are embedded in smartphones and tablets, thereby allowing the application of such learning on mobile devices is possible.

Employing face and speech biometric in an associative learning model can be implemented in terms of authentication systems, since these biometrics can be captured remotely and without the user consent, which makes such authentication is suitable for surveillance systems. In addition, the proposed association learning can be extended to be implemented for further types of biometrics such as (iris, retina, fingerprint, signature and so on) in the context of multimodal biometric authentication.



## References

- [1] L. Huang, *Automated Biometrics of Audio-Visual Multiple Modals*: Florida Atlantic University, 2010.
- [2] T. B. Long, "Hybrid Multi-Biometric Person Authentication System," in *World Congress on Engineering and Computer Science*, San Francisco, USA, 2012.
- [3] S. Sudarvizhi and S. Sumathi, "A Review on Continuous Authentication Using Multimodal Biometrics," *International Journal of Emerging Technology and Advanced Engineering*, vol. 3, pp. 192-196, 2013.
- [4] D. Bhattacharyya, R. Ranjan, A. Farkhod Alisherov, and M. Choi, "Biometric authentication: A review," *International Journal of u-and e-Service, Science and Technology*, vol. 2, pp. 13-28, 2009.
- [5] A. K. Jain and K. Nandakumar, "Biometric Authentication: System Security and User Privacy," *IEEE Computer*, vol. 45, pp. 87-92, 2012.
- [6] A. K. Jain and A. Kumar, "Biometrics of next generation: An overview," *Second Generation Biometrics*, 2010.
- [7] S. Sahoo and T. Choubisa, "Multimodal Biometric Person Authentication: A Review," *IETE Technical Review*, vol. 29, p. 54, 2012.
- [8] J. L. Wayman, "Fundamentals of biometric authentication technologies," *International Journal of Image and Graphics*, vol. 1, pp. 93-113, 2001.
- [9] R. N. Rodrigues, L. L. Ling, and V. Govindaraju, "Robustness of multimodal biometric fusion methods against spoof attacks," *Journal of Visual Languages & Computing*, vol. 20, pp. 169-179, 2009.

- [10] R. N. Rodrigues, N. Kamat, and V. Govindaraju, "Evaluation of biometric spoofing in a multimodal system," in *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, 2010, pp. 1-5.
- [11] Z. Akhtar, G. Fumera, G. L. Marcialis, and F. Roli, "Robustness evaluation of biometric systems under spoof attacks," in *Image Analysis and Processing-ICIAP 2011*, ed: Springer, 2011, pp. 159-168.
- [12] Z. Akhtar, S. Kale, and N. Alfarid, "Spoof attacks on multimodal biometric systems," in *Int. Conf. Information and Network Technology*, 2011, pp. 46-51.
- [13] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction* vol. 1: Cambridge Univ Press, 1998.
- [14] R. Bellman, *Adaptive control processes: a guided tour* vol. 4: Princeton university press Princeton, 1961.
- [15] P. K. Nayak and D. Narayan, "Multimodal Biometric Face and Fingerprint Recognition Using Neural Network," *International Journal of Engineering*, vol. 1, 2012.
- [16] Z. Wang, E. Wang, S. Wang, and Q. Ding, "Multimodal biometric system using face-iris fusion feature," *Journal of Computers*, vol. 6, pp. 931-938, 2011.
- [17] H. F. Liau and D. Isa, "Feature selection for support vector machine-based face-iris multimodal biometric system," *Expert Systems with Applications*, vol. 38, pp. 11105-11111, 2011.

- [18] Y. G. Kim, K. Y. Shin, E. C. Lee, and K. R. Park, "Multimodal biometric system based on the recognition of face and both irises," *Int J Adv Robotic Sy*, vol. 9, 2012.
- [19] A. Darwish, R. A. Elghafar, and A. F. Ali, "Multimodal face and ear images," *Journal of Computer Science*, vol. 5, p. 374, 2009.
- [20] Z. Huang, Y. Liu, C. Li, M. Yang, and L. Chen, "A Robust Face and Ear based Multimodal Biometric System using Sparse Representation," *Pattern Recognition*, 2013.
- [21] M. Wöllmer, M. Al-Hames, F. Eyben, B. Schuller, and G. Rigoll, "A multidimensional dynamic time warping algorithm for efficient multimodal fusion of asynchronous data streams," *Neurocomputing*, vol. 73, pp. 366-380, 2009.
- [22] K. Ahmadian and M. Gavrilova, "Chaotic Neural Network for Biometric Pattern Recognition," *Advances in Artificial Intelligence*, vol. 2012, p. 1, 2012.
- [23] R. H. Abiyev and K. Altunkaya, "Neural network based biometric personal identification with fast iris segmentation," *International Journal of Control, Automation and Systems*, vol. 7, pp. 17-23, 2009.
- [24] T. Masquelier, R. Guyonneau, and S. J. Thorpe, "Competitive STDP-based spike pattern learning," *Neural computation*, vol. 21, pp. 1259-1276, 2009.
- [25] S. G. Wysoski, L. Benuskova, and N. Kasabov, "Evolving spiking neural networks for audiovisual information processing," *Neural Networks*, vol. 23, pp. 819-835, 2010.

- [26] P. Gomathi, "A Survey on Biometrics based Key Authentication using Neural Network," *Global Journal of Computer Science and Technology*, vol. 11, 2011.
- [27] N. Kasabov, "Evolving spiking neural networks and neurogenetic systems for spatio-and spectro-temporal data modelling and pattern recognition," in *Advances in Computational Intelligence*, ed: Springer, 2012, pp. 234-260.
- [28] K. Dhoble, N. Nuntalid, G. Indiveri, and N. Kasabov, "Online spatio-temporal pattern recognition with evolving spiking neural networks utilising address event representation, rank order, and temporal spike learning," in *Neural Networks (IJCNN), The 2012 International Joint Conference on*, 2012, pp. 1-7.
- [29] N. Kasabov, K. Dhoble, N. Nuntalid, and G. Indiveri, "Dynamic evolving spiking neural networks for on-line spatio-and spectro-temporal pattern recognition," *Neural Networks*, 2012.
- [30] S. Fusi, M. Annunziato, D. Badoni, A. Salamon, and D. J. Amit, "Spike-driven synaptic plasticity: theory, simulation, VLSI implementation," *Neural Computation*, vol. 12, pp. 2227-2258, 2000.
- [31] J. M. Brader, W. Senn, and S. Fusi, "Learning real-world stimuli in a neural network with spike-driven synaptic dynamics," *Neural computation*, vol. 19, pp. 2881-2912, 2007.
- [32] D. Baras and R. Meir, "Reinforcement learning, spike-time-dependent plasticity, and the BCM rule," *Neural Computation*, vol. 19, pp. 2245-2279, 2007.

- [33] B. Glackin, J. A. Wall, T. M. McGinnity, L. P. Maguire, and L. J. McDaid, "A spiking neural network model of the medial superior olive using spike timing dependent plasticity for sound localization," *Frontiers in computational neuroscience*, vol. 4, 2010.
- [34] T. Masquelier and S. J. Thorpe, "Learning to recognize objects using waves of spikes and Spike Timing-Dependent Plasticity," in *Neural Networks (IJCNN), The 2010 International Joint Conference on*, 2010, pp. 1-8.
- [35] R. G. Leonard and G. R. Doddington. (1984, 21-5-2013). *A Speaker Independent Connected Digit Database*. Available: <http://catalog ldc.upenn.edu/docs/LDC93S10/tidigits.readme.html>
- [36] R. Newman, *Security and Access Control Using Biometric Technologies*. United States: Course Technology Ptr, 2010.
- [37] T. Sabareeswari and S. L. Stewart, "Identification of a Person Using Multimodal Biometric System," *International Journal of Computer Applications IJCA*, vol. 3, pp. 12-16, 2010.
- [38] A. Khairwa, K. Abhishek, S. Prakash, and T. Pratap, "A comprehensive study of various biometric identification techniques," in *Computing Communication & Networking Technologies (ICCCNT), 2012 Third International Conference on*, 2012, pp. 1-6.
- [39] I. Marqu'es, "Face Recognition Algorithms," Del Pais Vasco, 2010.
- [40] F. Jiao, W. Gao, X. Chen, G. Cui, and S. Shan, "A face recognition method based on local feature analysis," in *Proc. of the 5th Asian Conference on Computer Vision*, 2002, pp. 188-192.

- [41] S. Arca, P. Campadelli, and R. Lanzarotti, "A face recognition system based on local feature analysis," in *Audio-and Video-Based Biometric Person Authentication*, 2003, pp. 182-189.
- [42] C. Kotropoulos, A. Tefas, and I. Pitas, "Frontal face authentication using morphological elastic graph matching," *Image Processing, IEEE Transactions on*, vol. 9, pp. 555-560, 2000.
- [43] M. Agarwal, H. Agrawal, N. Jain, and M. Kumar, "Face recognition using principle component analysis, eigenface and neural network," in *Signal Acquisition and Processing, 2010. ICSAP'10. International Conference on*, 2010, pp. 310-314.
- [44] M. Kazi, Y. Rode, S. Dabhade, N. Al-Dawla, A. Mane, R. Manza, *et al.*, "Multimodal Biometric System Using Face and Signature: A Score Level Fusion Approach," *Advances in Computational Research*, vol. 4, 2012.
- [45] X. Anguera, "Speaker independent discriminant feature extraction for acoustic pattern-matching," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, 2012, pp. 485-488.
- [46] R. Terashima, T. Yoshimura, T. Wakita, K. Tokuda, and T. Kitamura, "Prediction method of speech recognition performance based on HMM-based speech synthesis technique," *IEEJ Transactions on Electronics, Information and Systems*, vol. 130, pp. 557-564, 2010.
- [47] G. E. Dahl, D. Yu, L. Deng, and A. Acero, "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition," *Audio*,

- Speech, and Language Processing, IEEE Transactions on*, vol. 20, pp. 30-42, 2012.
- [48] M.-C. Cheung, M.-W. Mak, and S.-Y. Kung, "Intramodal and intermodal fusion for audio-visual biometric authentication," in *Intelligent Multimedia, Video and Speech Processing, 2004. Proceedings of 2004 International Symposium on*, 2004, pp. 25-28.
- [49] N. Dahiya and C. Kant, "Biometrics Security Concerns," in *Advanced Computing & Communication Technologies (ACCT), 2012 Second International Conference on*, 2012, pp. 297-302.
- [50] S. K. Bandyopadhyay, D. Bhattacharyya, P. Das, D. Ganguly, and S. Mukherjee, "Statistical Approach for Offline Handwritten Signature Verification," *Journal of Computer Science*, vol. 4, pp. 181-185, 2008.
- [51] A. Jain, L. Hong, and S. Pankanti, "Biometric identification," *Communications of the ACM*, vol. 43, pp. 90-98, 2000.
- [52] J. Adams, D. L. Woodard, G. Dozier, P. Miller, K. Bryant, and G. Glenn, "Genetic-based type II feature extraction for periocular biometric recognition: Less is more," in *Pattern Recognition (ICPR), 2010 20th International Conference on*, 2010, pp. 205-208.
- [53] P. Zhang and X. Guo, "A cascade face recognition system using hybrid feature extraction," *Digital Signal Processing*, 2012.
- [54] N. Khan, R. Ksantini, I. Ahmad, and B. Boufama, "A novel SVM+ NDA model for classification with an application to face recognition," *Pattern Recognition*, vol. 45, pp. 66-79, 2012.

- [55] Y. Wen, L. He, and P. Shi, "Face recognition using difference vector plus KPCA," *Digital Signal Processing*, vol. 22, pp. 140-146, 2012.
- [56] W. Gerstner and W. M. Kistler, *Spiking neuron models: Single neurons, populations, plasticity*: Cambridge university press, 2002.
- [57] K. Simoens, J. Bringer, H. Chabanne, and S. Seys, "A Framework for Analyzing Template Security and Privacy in Biometric Authentication Systems," *Information Forensics and Security, IEEE Transactions on*, vol. 7, pp. 833-841, 2012.
- [58] S. M. S. Ahmad, B. M. Ali, and W. A. W. Adnan, "Technical Issues and Challenges of Biometric Applications as Access Control Tools of Information Ssecurity," *International Journal of Innovative*, vol. 8, pp. 7983-7999, 2012.
- [59] B. Chen and V. Chandran, "Biometric template security using higher order spectra," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, 2010, pp. 1730-1733.
- [60] S. Chindaro, F. Deravi, Z. Zhou, M. Ng, M. C. Neves, X. Zhou, *et al.*, "A multibiometric face recognition fusion framework with template protection," in *SPIE Defense, Security, and Sensing*, 2010, pp. 76670U-76670U-6.
- [61] Z. Akhtar, "Security of Multimodal Biometric Systems against Spoof Attacks," PhD thesis, University of Cagliari, Italy, 2012.
- [62] L. Wang, "Some issues of biometrics: technology intelligence, progress and challenges," *International Journal of Information Technology and Management*, vol. 11, pp. 72-82, 2012.

- [63] M. Soltane, N. Doghmane, and N. Guersi, "Face and Speech Based Multimodal Biometric Authentication," *International Journal of Advanced Science and Technology*, vol. 21, pp. 41-56, 2010.
- [64] M. R. Manju, M. A. Rajendrana, and A. S. Nargunab, "Performance Analysis of Multimodal Biometric Based Authentication System," *International Journal of Engineering*, vol. 1, 2012.
- [65] R. Raghavendra, B. Dorizzi, A. Rao, and G. Hemantha Kumar, "Designing efficient fusion schemes for multimodal biometric systems using face and palmprint," *Pattern Recognition*, vol. 44, pp. 1076-1088, 2011.
- [66] Y. Tong, F. W. Wheeler, and X. Liu, "Improving biometric identification through quality-based face and fingerprint biometric fusion," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, 2010, pp. 53-60.
- [67] X. Jing, S. Li, Y. Yao, W. Li, F. Wu, and C. Lan, "Multi-Modal Biometric Feature Extraction and Recognition Based on Subclass Discriminant Analysis (SDA) and Generalized Singular Value Decomposition (GSVD)," in *Hand-Based Biometrics (ICHB), 2011 International Conference on*, 2011, pp. 1-6.
- [68] Y. Elmir, S. Al-Maadeed, A. Amira, and A. Hassaine, "A Multi-modal Face and Signature Biometric Authentication System Using A Max-of-scores Based Fusion," in *Neural Information Processing*, 2012, pp. 576-583.
- [69] Z. Akhtar, G. Fumera, G. L. Marcialis, and F. Roli, "Evaluation of Multimodal Biometric Score Fusion Rules Under Spoof Attacks," in

- Biometrics (ICB), 2012 5th IAPR International Conference on*, 2012, pp. 402-407.
- [70] M. Al-Hames and G. Rigoll, "Reduced complexity and scaling for asynchronous HMMs in a bimodal input fusion application," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, 2006, pp. V-V.
- [71] H. Ebied, "Feature extraction using PCA and Kernel-PCA for face recognition," in *Informatics and Systems (INFOS), 2012 8th International Conference on*, 2012, pp. MM-72-MM-77.
- [72] K. P. Chandar, M. M. Chandra, M. R. Kumar, and B. S. Latha, "Multi scale feature extraction and enhancement using SVD towards secure Face Recognition system," in *Signal Processing, Communication, Computing and Networking Technologies (ICSCCN), 2011 International Conference on*, 2011, pp. 64-69.
- [73] B. Gupta, S. Gupta, and A. K. Tiwari, "Face Detection Using Gabor Feature Extraction and Artificial Neural Network," *ABES Engineering College, Ghaziaba*, 2010.
- [74] M. A. Hossan, S. Memon, and M. A. Gregory, "A novel approach for MFCC feature extraction," in *Signal Processing and Communication Systems (ICSPCS), 2010 4th International Conference on*, 2010, pp. 1-5.
- [75] R. Sarikaya, B. L. Pellom, and J. H. Hansen, "Wavelet packet transform features with application to speaker identification," in *IEEE Nordic Signal Processing Symposium*, 1998, pp. 81-84.

- [76] S. Malik and F. A. Afsar, "Wavelet transform based automatic speaker recognition," in *Multitopic Conference, 2009. INMIC 2009. IEEE 13th International*, 2009, pp. 1-4.
- [77] P. Wallisch, M. Lusignan, M. Benayoun, T. I. Baker, A. S. Dickey, and N. Hatsopoulos, *MATLAB for Neuroscientists: an Introduction to Scientific Computing in MATLAB*: Elsevier, 2009.
- [78] P. Xanthopoulos, P. M. Pardalos, and T. B. Trafalis, "Principal Component Analysis," in *Robust Data Mining*, ed: Springer, 2013, pp. 21-26.
- [79] C. Juanjuan, Z. Zheng, S. Han, and Z. Gang, "Facial expression recognition based on PCA reconstruction," in *Computer Science and Education (ICCSE), 2010 5th International Conference on*, 2010, pp. 195-198.
- [80] B. Bozorgtabar, F. Noorian, and G. A. R. Rad, "Comparison of different PCA based Face Recognition algorithms using Genetic Programming," in *Telecommunications (IST), 2010 5th International Symposium on*, 2010, pp. 801-805.
- [81] Y. Luo, C.-m. Wu, and Y. Zhang, "Facial expression recognition based on fusion feature of PCA and LBP with SVM," *Optik-International Journal for Light and Electron Optics*, 2012.
- [82] J. Keche and M. Dhore, "Comparative Study of Feature Extraction Techniques for Face Recognition System," *International Journal*, 2012.
- [83] Z. Liang and P. Shi, "Kernel direct discriminant analysis and its theoretical foundation," *Pattern Recognition*, vol. 38, pp. 445-447, 2005.

- [84] Y. Wen and P. Shi, "An approach to numeral recognition based on improved LDA and Bhattacharyya distance," in *Communications, Control and Signal Processing, 2008. ISCCSP 2008. 3rd International Symposium on*, 2008, pp. 309-311.
- [85] Y. Wen, Y. Lu, J. Yan, Z. Zhou, K. M. von Deneen, and P. Shi, "An algorithm for license plate recognition applied to intelligent transportation system," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 12, pp. 830-845, 2011.
- [86] V. Vidya, N. Farheen, K. Manikantan, and S. Ramachandran, "Face Recognition using Threshold Based DWT Feature Extraction and Selective Illumination Enhancement Technique," *Procedia Technology*, vol. 6, pp. 334-343, 2012.
- [87] V. Vezhnevets, V. Sazonov, and A. Andreeva, "A survey on pixel-based skin color detection techniques," in *Proc. Graphicon*, 2003, pp. 85-92.
- [88] S. K. Singh, D. Chauhan, M. Vatsa, and R. Singh, "A robust skin color based face detection algorithm," *Tamkang Journal of Science and Engineering*, vol. 6, pp. 227-234, 2003.
- [89] B. G. Bhatt and Z. H. Shah, "Face feature extraction techniques: a survey," in *National conference on recent trends in engineering & technology*, 2011, pp. 13-14.
- [90] Z.-Q. Hong, "Algebraic feature extraction of image for recognition," *Pattern recognition*, vol. 24, pp. 211-219, 1991.

- [91] H. Miari-Naimi and P. Davari, "A New Fast and Efficient HMM-Based Face Recognition System Using a 7-State HMM Along With SVD Coefficients," *Iranian Journal of Electrical And Electronic Engineering*, vol. 4, pp. 46-57, 2008.
- [92] A. P. Gosavi and S. Khot, "Facial Expression Recognition using Principal Component Analysis with Singular Value Decomposition," *International Journal of Advance Research in Computer Science and Management Studies*, vol. 1, 2013.
- [93] R. W. Swiniarski and A. Skowron, "Rough set methods in feature selection and recognition," *Pattern recognition letters*, vol. 24, pp. 833-849, 2003.
- [94] H. Demirel, C. Ozcinar, and G. Anbarjafari, "Satellite image contrast enhancement using discrete wavelet transform and singular value decomposition," *Geoscience and Remote Sensing Letters, IEEE*, vol. 7, pp. 333-337, 2010.
- [95] H. Demirel, G. Anbarjafari, and M. N. S. Jahromi, "Image equalization based on singular value decomposition," in *Computer and Information Sciences, 2008. ISCIS'08. 23rd International Symposium on*, 2008, pp. 1-5.
- [96] J. Baker, L. Deng, J. Glass, S. Khudanpur, C.-H. Lee, N. Morgan, *et al.*, "Developments and directions in speech recognition and understanding," *Signal Processing Magazine, IEEE*, vol. 26, pp. 75-80, 2009.
- [97] P. K. Cherupalli and J. S. K. Gari, "A Wavelet based Feature Extraction for Voice-Lock systems," in *TENCON 2005 2005 IEEE Region 10*, 2005, pp. 1-4.

- [98] L. Muda, M. Begam, and I. Elamvazuthi, "Voice recognition algorithms using mel frequency cepstral coefficient (mfcc) and dynamic time warping (dtw) techniques," *arXiv preprint arXiv:1003.4083*, 2010.
- [99] M. J. Alam, T. Kinnunen, P. Kenny, P. Ouellet, and D. O'Shaughnessy, "Multi-taper MFCC features for speaker verification using I-vectors," in *Automatic Speech Recognition and Understanding (ASRU), 2011 IEEE Workshop on*, 2011, pp. 547-552.
- [100] S.-Y. Lung, "Feature extracted from wavelet decomposition using biorthogonal Riesz basis for text-independent speaker recognition," *Pattern Recognition*, vol. 41, pp. 3068-3070, 2008.
- [101] T. Kinnunen, "Spectral features for automatic text-independent speaker recognition," *Licentiate's Thesis*, 2003.
- [102] Z. Tufekci and J. Gowdy, "Feature extraction using discrete wavelet transform for speech recognition," in *Southeastcon 2000. Proceedings of the IEEE*, 2000, pp. 116-123.
- [103] R. N. Khushaba, A. Al-Jumaily, and A. Al-Ani, "Novel feature extraction method based on fuzzy entropy and wavelet packet transform for myoelectric Control," in *Communications and Information Technologies, 2007. ISCIT '07. International Symposium on*, 2007, pp. 352-357.
- [104] N. S. Nehe and R. S. Holambe, "New feature extraction methods using DWT and LPC for isolated word recognition," in *TENCON 2008 - 2008 IEEE Region 10 Conference*, 2008, pp. 1-6.

- [105] J.-D. Wu and B.-F. Lin, "Speaker identification using discrete wavelet packet transform technique with irregular decomposition," *Expert Systems with Applications*, vol. 36, pp. 3136-3143, 3// 2009.
- [106] M. Anusuya and S. Katti, "Comparison of different speech feature extraction techniques with and without wavelet transform to Kannada speech recognition," *International Journal of Computer Applications*, vol. 26, pp. 19-24, 2011.
- [107] B. S. Anami, V. B. Pagi, and S. M. Magi, "Wavelet-based acoustic analysis for determining health condition of motorized two-wheelers," *Applied Acoustics*, vol. 72, pp. 464-469, 6// 2011.
- [108] R. N. Khushaba, S. Kodagoda, S. Lal, and G. Dissanayake, "Driver Drowsiness Classification Using Fuzzy Wavelet-Packet-Based Feature-Extraction Algorithm," *Biomedical Engineering, IEEE Transactions on*, vol. 58, pp. 121-131, 2011.
- [109] A. Seal, S. Gangulyb, D. Bhattacharjee, M. Nasipuri, and D. K. Basu, "Minutiae Based Thermal Human Face Recognition using Label Connected Component Algorithm," *Procedia Technology*, vol. 4, pp. 604-611, 2012.
- [110] M. Mansoorizadeh and N. M. Charkari, "Multimodal Information Fusion Application to Human Emotion Recognition From Face and Speech," *Multimedia Tools and Applications*, vol. 49, pp. 277-297, 2010.
- [111] S. K. Sahoo and S. M. Prasanna, "Bimodal biometric person authentication using speech and face under degraded condition," in *Communications (NCC), 2011 National Conference on*, 2011, pp. 1-5.

- [112] Z. Ciota and M. Gawinowski, "Multilayer neural networks in handwritten recognition," in *Modern Problems of Radio Engineering, Telecommunications and Computer Science, 2004. Proceedings of the International Conference, 2004*, pp. 192-194.
- [113] F. Siraj, N. Yusoff, and L. C. Kee, "Emotion classification using neural network," in *Computing & Informatics, 2006. ICOCI'06. International Conference on, 2006*, pp. 1-7.
- [114] M. S. B. H. Salam, D. Mohamad, and S. H. S. Salleh, "Malay isolated speech recognition using neural network: a work in finding number of hidden nodes and learning parameters," *Int. Arab J. Inf. Technol.*, vol. 8, pp. 364-371, 2011.
- [115] N. Yusoff and A. Grüning, "Learning anticipation through priming in spatio-temporal neural networks," in *Neural Information Processing, 2012*, pp. 168-175.
- [116] B. DasGupta and G. Schnitger, "The power of approximating: a comparison of activation functions," *Mathematical Research*, vol. 79, pp. 641-641, 1994.
- [117] J. Vreeken, "Spiking neural networks, an introduction," *Technical Report UU-CS*, pp. 1-5, 2003.
- [118] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533-536, 1986.
- [119] T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological cybernetics*, vol. 43, pp. 59-69, 1982.

- [120] B. d. B. Pereira, C. R. Rao, R. L. Oliveira, and E. M. do Nascimento, "Combining unsupervised and supervised neural networks in cluster analysis of gamma-ray burst," *Journal of Data Science*, vol. 8, pp. 327-338, 2010.
- [121] G. B. Kaplan, N. Şengör, H. Gürvit, and C. Güzeliş, "Modelling the Stroop effect: A connectionist approach," *Neurocomputing*, vol. 70, pp. 1414-1423, 2007.
- [122] S. G. Wysoski, L. Benuskova, and N. Kasabov, "Adaptive learning procedure for a network of spiking neurons and visual pattern recognition," in *Advanced Concepts for Intelligent Vision Systems*, 2006, pp. 1133-1142.
- [123] N. K. Kasabov and N. Kasabov, *Evolving connectionist systems: the knowledge engineering approach*: Springer, 2007.
- [124] N. Kasabov, "To spike or not to spike: A probabilistic spiking neuron model," *Neural Networks*, vol. 23, pp. 16-19, 2010.
- [125] S. Schliebs and N. Kasabov, "Evolving spiking neural network—a survey," *Springer-Verlag*, 2013.
- [126] E. M. Izhikevich, "Polychronization: computation with spikes," *Neural computation*, vol. 18, pp. 245-282, 2006.
- [127] H. Paugam-Moisy, R. Martinez, and S. Bengio, "Delay learning and polychronization for reservoir computing," *Neurocomputing*, vol. 71, pp. 1143-1158, 2008.
- [128] T. Grabowski and G. J. Tortora, *Principles of anatomy and physiology*: John Wiley & Sons, 2003.

- [129] E. M. Izhikevich and J. Moehlis, "Dynamical Systems in Neuroscience: The geometry of excitability and bursting," *SIAM review*, vol. 50, p. 397, 2008.
- [130] E. M. Izhikevich, "Which model to use for cortical spiking neurons?," *Neural Networks, IEEE Transactions on*, vol. 15, pp. 1063-1070, 2004.
- [131] E. M. Izhikevich, "Simple model of spiking neurons," *Neural Networks, IEEE Transactions on*, vol. 14, pp. 1569-1572, 2003.
- [132] C. A. Erickson and R. Desimone, "Responses of macaque perirhinal neurons during and after visual stimulus association learning," *The Journal of neuroscience*, vol. 19, pp. 10404-10416, 1999.
- [133] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *arXiv preprint cs/9605103*, 1996.
- [134] M. A. Wiering, H. van Hasselt, A. D. Pietersma, and L. Schomaker, "Reinforcement learning algorithms for solving classification problems," in *Adaptive Dynamic Programming And Reinforcement Learning (ADPRL), 2011 IEEE Symposium on*, 2011, pp. 91-96.
- [135] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial intelligence*, vol. 101, pp. 99-134, 1998.
- [136] Y. Engel, "Algorithms and representations for reinforcement learning," Citeseer, 2005.
- [137] J. Boger, J. Hoey, P. Poupart, C. Boutilier, G. Fernie, and A. Mihailidis, "A planning system based on Markov decision processes to guide people with

- dementia through activities of daily living," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 10, pp. 323-333, 2006.
- [138] M. G. Lagoudakis and R. Parr, "Reinforcement learning as classification: Leveraging modern classifiers," in *ICML*, 2003, pp. 424-431.
- [139] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, pp. 279-292, 1992.
- [140] S. Singh, T. Jaakkola, M. L. Littman, and C. Szepesvári, "Convergence results for single-step on-policy reinforcement-learning algorithms," *Machine Learning*, vol. 38, pp. 287-308, 2000.
- [141] V. R. Konda and V. S. Borkar, "Actor-Critic--Type Learning Algorithms for Markov Decision Processes," *SIAM Journal on control and Optimization*, vol. 38, pp. 94-123, 1999.
- [142] M. Pantic, A. Pentland, A. Nijholt, and T. S. Huang, "Human computing and machine understanding of human behavior: A survey," in *Artificial Intelligence for Human Computing*, ed: Springer, 2007, pp. 47-71.
- [143] Z. Zeng, Y. Hu, Y. Fu, T. S. Huang, G. I. Roisman, and Z. Wen, "Audio-visual emotion recognition in adult attachment interview," in *Proceedings of the 8th international conference on Multimodal interfaces*, 2006, pp. 139-145.
- [144] L. Wang, W. Hu, and T. Tan, "Recent developments in human motion analysis," *Pattern Recognition*, vol. 36, pp. 585-601, 2003.

- [145] Q. Yu, Y. Yin, G. Yang, Y. Ning, and Y. Li, "Face and Gait Recognition Based on Semi-supervised Learning," in *Pattern Recognition*, ed: Springer, 2012, pp. 284-291.
- [146] F. Roli, L. Didaci, and G. L. Marcialis, "Adaptive biometric systems that can improve with use," in *Advances in Biometrics*, ed: Springer, 2008, pp. 447-471.
- [147] C. Shan, "Learning local binary patterns for gender classification on real-world face images," *Pattern Recognition Letters*, vol. 33, pp. 431-437, 2012.
- [148] Z. Feng, M. Yang, L. Zhang, Y. Liu, and D. Zhang, "Joint discriminative dimensionality reduction and dictionary learning for face recognition," *Pattern Recognition*, 2013.
- [149] N. Caporale and Y. Dan, "Spike timing-dependent plasticity: a Hebbian learning rule," *Annu. Rev. Neurosci.*, vol. 31, pp. 25-46, 2008.
- [150] J. Hu and M. P. Wellman, "Multiagent reinforcement learning: theoretical framework and an algorithm," in *ICML*, 1998, pp. 242-250.
- [151] P. Dayan, L. F. Abbott, and L. Abbott, "Theoretical neuroscience: Computational and mathematical modeling of neural systems," 2001.
- [152] R. V. Florian, "Reinforcement learning through modulation of spike-timing-dependent synaptic plasticity," *Neural Computation*, vol. 19, pp. 1468-1502, 2007.
- [153] C. U. C. Laboratory. (2002, 22 June 2013). *The Database of Faces*. Available: <http://www.cl.cam.ac.uk/research/dtg/attarchive/>

- [154] A. Harter, A. Hopper, P. Steggle, A. Ward, and P. Webster, "The anatomy of a context-aware application," *Wireless Networks*, vol. 8, pp. 187-197, 2002.
- [155] V. Maheshkar, S. Kamble, and S. Agarwal, "DCT-based unique faces for face recognition using Mahalanobis distance," in *Proceedings of the First International Conference on Intelligent Interactive Technologies and Multimedia*, 2010, pp. 208-212.
- [156] E. Gumus, N. Kilic, A. Sertbas, and O. N. Ucan, "Evaluation of face recognition techniques using PCA, wavelets and SVM," *Expert Systems with Applications*, vol. 37, pp. 6404-6408, 2010.
- [157] V. Maheshkar, S. Kamble, S. Agarwal, and V. Kumar, "Feature Image Generation Using Low, Mid and High Frequency Regions for Face Recognition," *The International Journal of Multimedia & Its Applications*, vol. 4, 2012.
- [158] M. S. El-Bashir, "Face Recognition Using Multi-Classifer," *Applied Mathematical Sciences*, vol. 6, pp. 2235-2244, 2012.
- [159] M. R. Faraji and X. Qi, "An effective neutrosophic set-based preprocessing method for face recognition," in *Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on*, 2013, pp. 1-4.
- [160] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, pp. 71-86, 1991.
- [161] P. Shemi and M. Ali, "A Principal Component Analysis Method for Recognition of Human Faces: Eigenfaces Approach," *International Journal*

of *Electronics Communication and Computer Technology (IJECCCT)*, vol. 2, 2012.

- [162] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, pp. 711-720, 1997.
- [163] D. Pissarenko. (2002, 4 September 2013). *Eigenface-based Facial Recognition*. Available: <http://openbio.sourceforge.net/resources/eigenfaces/eigenfaces.pdf>
- [164] J. Wang, K. N. Plataniotis, and A. N. Venetsanopoulos, "Selecting discriminant eigenfaces for face recognition," *Pattern Recognition Letters*, vol. 26, pp. 1470-1482, 2005.
- [165] J. Wang, K. N. Plataniotis, and A. Venetsanopolous, "Selecting kernel eigenfaces for face recognition with one training sample per subject," in *Multimedia and Expo, 2006 IEEE International Conference on*, 2006, pp. 1637-1640.
- [166] A. Yu, C. Elder, J. Yeh, and N. Pai. (2009, 17 September 2013). *Facial Recognition Using Eigenfaces: Obtainig Eigenfaces*. Available: <http://cnx.org/content/m33173/1.3/>
- [167] V. Akalin, "Face Recognition Using Eigenfaces and Neural Networks," Master, Electrical and Electronics Engineering, Middle East Tchnical University, 2003.

- [168] O. Sakhi. (2012, 13 August). *Face Recognition Guide Using Hidden Markov Models and Singular Value Decomposition (First ed.)*. Available: <http://www.facerecognitioncode.com/>
- [169] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 1794-1801.
- [170] Y.-L. Boureau, F. Bach, Y. LeCun, and J. Ponce, "Learning mid-level features for recognition," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 2010, pp. 2559-2566.
- [171] A. Coates and A. Ng, "The importance of encoding versus training with sparse coding and vector quantization," in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 2011, pp. 921-928.
- [172] A. Jayakumar and P. Babu Anto, "Comparison of wavelet packets and DWT in a gender based multichannel speaker recognition system," in *Information & Communication Technologies (ICT), 2013 IEEE Conference on*, 2013, pp. 847-850.
- [173] MathWorks. (2001, 5 October 2013). *Wavelet Packets*. Available: <http://www.mathworks.com/help/wavelet/ug/wavelet-packets.html?searchHighlight=Wavelet+Packets>

- [174] R. Leonard, "A database for speaker-independent digit recognition," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84.*, 1984, pp. 328-331.
- [175] J. Cardenas, E. Castillo, and J. Meng, "Noise-robust speech recognition system based on power spectral subtraction with a geometric approach," *Acoustics 2012 Nantes*, 2012.
- [176] H.-G. Hirsch, S. Ganapathy, and H. Hermansky, "Comparison of Different Approaches for Speech Recognition in Hands-free Mode," in *Speech Communication; 10. ITG Symposium; Proceedings of*, 2012, pp. 1-4.
- [177] Y. Naya, M. Yoshida, and Y. Miyashita, "Forward processing of long-term associative memory in monkey inferotemporal cortex," *The Journal of neuroscience*, vol. 23, pp. 2861-2871, 2003.
- [178] S. Dhawan, "A review of image compression and comparison of its algorithms," *International Journal of electronics & Communication technology*, vol. 2, 2011.
- [179] M.-H. Horng, "Vector quantization using the firefly algorithm for image compression," *Expert Systems with Applications*, vol. 39, pp. 1078-1091, 1// 2012.