Cross Models for Twin Recognition

Da Tong Gu

A thesis submitted to Auckland University of Technology in
partial fulfilment of requirements for the degree of Master
of Computer and Information Science (MCIS)

2016
School of Engineering, Computer and Mathematical Sciences

Abstract

Nowadays, biometrics has become a popular tool in personal identification as it utilizes physiological or behavioral characteristics to identify individuals. Recent advancement in computer science has increased the accuracy of biometrics to a higher level. However, there are still a number of existing problems, such as complex environment, aging and unique problems. Twin identification is notably one of the most challenging issues, because they resemble each other in terms of biometrics. The similarity affects the use of biometrics in general cases and raised the potential risk of biometrics in access control.

This thesis presents and compares four methods for twin recognition, namely, ear recognition, speaker recognition and lip movement recognition. Our results show that speaker recognition has the best performance with 100% accuracy. This is much higher than that of face recognition and ear recognition (with 58% and 53% respectively) while movement recognition that yields 76% accuracy.

The objectives of this thesis are to investigate whether it is possible to identify twins between each other by using biometrics and find out which recognition approach is the best one. Comparing to the-state-of-the-art technologies, our work has taken a further step in twin recognition by using biometrics.

In future, we will take behavioral analysis of twins into consideration, since their growing-up environment may have impact on their behaviors. We assume that other clues for twin recognition could be discovered from the behavioral analysis.

Keywords: cross models, twin recognition, speaker recognition, face recognition, ear recognition, lip movement recognition, precision, recall, F-measure.

Table of Content

ABSTRACT II

LIST OF FIGURES V

LIST OF TABLES VI

DECLARATION VII

ACKNOWLEDGEMENTS VIII

CHAPTER 1 INTRODUCTION 1

1.1 BACKGROUND AND MOTIVATION 1
1.2 THE AIMS OF THIS THESIS 5
1.3 RESEARCH QUESTIONS 5
1.4 THE STRUCTURE OF THIS THESIS 6

CHAPTER 2 LITERATURE REVIEW 8

2.1 FACE DETECTION 9
2.2 FACE RECOGNITION 10
2.2.1 Principal Component Analysis (PCA) 10
2.2.2 Independent Component Analysis (ICA) 11
2.2.3 Linear Discriminant Analysis (LDA) 11
2.2.4 Local Binary Patterns (LBP) 12
2.2.5 Active Appearance Models (AAM) 12
2.2.6 Support Vector Machines (SVM) 12
2.3 EAR RECOGNITION 13
2.4 SPEAKER RECOGNITION 15
2.4.1 Pitch 15
2.4.2 Cepstrum 15
2.4.3 Mel-frequency cepstral coefficient (MFCC) 15
2.5 LIP MOVEMENT RECOGNITION 16

CHAPTER 3 METHODOLOGY 17

3.1 RESEARCH ENVIRONMENT 18
3.2 DATA SET 19
3.3 METHODS AND EVALUATIONS 20
3.3.1 PCA algorithm 20
3.3.2 MFCC 21
3.3.3 VQLBG 22
3.3.4 Evaluation 23
3.4 ALGORITHMS 24
3.4.1 Four algorithms 24
3.4.2 Four flowcharts 28

CHAPTER 4 FINDINGS 34

4.1 RESULTS 35

4.1.1 Speaker recognition *35*

4.1.2 Face recognition *38*

4.1.3 Lip movement recognition *39*

4.1.4 Human ear recognition *40*

4.2 EVALUATIONS 40

CHAPTER 5 DISCUSSIONS AND ANALYSIS **44**

5.1 ANALYSIS 45

5.2 LIMITATIONS 47

CHAPTER 6 CONCLUSION AND FUTURE WORK **48**

6.1 CONCLUSION 48

6.2 LIMITATIONS AND FUTURE WORK 49

REFERENCES **50**

APPENDIX I **60**

List of Figures

Figure 1.1 Twins of months old (Left A Right B) 4

Figure 1.2 Twins of 2, 6 and 10 years old 5

Figure 1.3 Twins of 15, 18 and 23 years old (ageing) 5

Figure 2.1 Integral Images 9

Figure 3.1 Face samples of twins. The left one is twin A and the right one is twin B 20

Figure 3.2 Samples of twin ears. (a) twin A (b) twin B 20

Figure 3.3 Work flow of the face recognition system 30

Figure 3.4 Work flow of the lip movement recognition system 31

Figure 3.5 Work flow of the speaker recognition system 32

Figure 3.6 Work flow of the ear recognition system 33

Figure 4.1 Pitches of the test samples. 35

Figure 4.2 Spectrogram of the twin A. 36

Figure 4.3 Spectrogram of the twin B. 37

Figure 4.4 Result of face recognition (a) twin A (b) twin B 39

Figure 4.5 Results of lip movement recognition. 40

Figure 4.6 The accuracy of four methods in twin recognition. 41

Figure 4.7 Precision of four methods in twin recognition. 41

Figure 4.8 Recall of four methods in twin recognition. 42

Figure 4.9 F-measure of four methods in twin recognition 43

List of Tables

Table 3.1 Sample list of speaker recognition. 19
Table 4.1Classification matrix of speaker recognition..38
Table 4.2 Classification matrix of face recognition. 39
Table 4.3 Classification matrix of lip movement. 40
Table 4.4 Classification matrix of ear recognition. 40

Declaration

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (expect where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Signature: Dutong Gu

Date: 28 May 2016

Acknowledgements

I would like to thank the Auckland University of Technology (AUT) to provide the research environment for this research project. Also I appreciate my twin brother Xiao Yi Gu for collecting relevant data. I am grateful for my parents for their support.

Chapter 1 Introduction

1.1 Background and motivation

Biometrics has been broadly employed in access control, digital security, and forensics. It is the technology automatically recognizing a person by using the physiological or behavioral characteristics (Jain et al., 2004). These characteristics, such as face, ear, fingerprint, iris, voice, and Deoxyribonucleic Acid (DNA), are collected by equipment like digital camera or voice recorder. The collected data are then abelled, trained, and applied to identify individuals. There is a variety of biometrics that have been applied to our daily life, for example, face recognition, DNA analysis, fingerprint identification signature verification (Qi & Hunt 1994) and speaker recognition, etc.

To achieve biometrics, different methods need different hardware. For example, fingerprint requires fingerprint scanner, DNA necessitates analysis machine and voice has to use microphone. The features are extracted and analyzed by these facilities. Biometrics can be used in many fields like verification of criminal investigation (Chellappa, Wilson & Sirohey, 1995). For example, banks use speaker recognition to verify the customer who is the correct person of that account.

Biometric has the advantages that make it popular in recent years. Unlike passwords, biometrics is much difficult to be shared and guessed (Jain, 2007). For example, passwords combined with numbers or alphabetical letters normally contain personal information including the date of birt. Complicated ones are hard to be remembered and easy to be forgeted. On the other hand, it is tough to replicate the biometric information of a person, meanwhile biometrics does not need to be remembered.

Nowadays, face recognition is one of the most commonly used biometrics (Turk &

Pentland, 1991) which has been studied for decades (Bruce & Young, 1986). Compared with other biometrics, face recognition has its own merits. Firstly, it is easy to get the recognition hardware. Face recognition only requires a camera to get the face information and a computer to analyze the data. Secondly, it can be used in the scenario that the targeted persons are unaware, which means it has the capacity of concealment. Numerous research reports have indicated that face recognition is useful in judicial investigation (Pentland, Moghaddam & Starner, 1994; Phillips, Moon & Rizvi 1997).

Face recognition generally contains two categories: online and offline recognition (Lawrence, Giles, Tsoi & Back, 1997). Online face recognition refers to real-time face recognition which employs a live camera to record videos and analyze the input simultaneously. Offline face recognition means the face recognition utilizes those previously recorded data.

A vast number of models have been proposed in face recognition such as eigenfaces (Kshirsagar, Baviskar & Gaikwad, 2011; Turk & Pentland, 1991), Local Binary Patterns (LBP) (Ojala, Pietikainen & Maenpaa, 2002), Elastic Bunch Graph Matching (EBGM) (Wiskott, Fellous, Kruger & vonderMalsburg, 1997; Hanmandlu, Gupta & Vasikarla, 2013), Trace Transform (Kadyrov & Petrou, 2001; Manjunath, Chellappa & Malsburg, 1992; Srisuk, Petrou, Kurutach & Kadyrov, 2003), Kernel Model (Yang, 2002), Neural Network (Zhao Huang & Sun, 2004; Rowley, Baluja & Kanade, 1998), Hidden Markov Model (HMM) (Nefian, & Hayes, 1999), Hausdorff Distance (Guo, Lam, Lin & Siu 2003), Nearest Feature Line (NFL) (Li & Lu, 1999), Support Vector Machines (SVM) (Guo, Li & Chan, 2000; Jonsson, Matas, Kittler & Li, 2000; Li, Ming, Yang & Pan, 2006) and 3D model (Moghaddam Lee, Pfister & Machiraju, 2003; Blanz & Vetter, 2003).

Face recognition has problems in twin recognition. One of the problems in face is the complex surroundings. Complex surroundings related problems include illumination problem, background problem and moving object problem (Cui, 2014). A human face recognition system that is able to be adaptive to a complex environment was developed by Sung and Poggio (1998), which addressed the problem.

The change of illumination is one of the most representative issues and various methods have been correspondingly discussed to tackle them. A skin-color extraction algorithm has been developed (Huynh-Thu, Meguro & Kaneko, 2002) which is able to capture face in the different illumination. In addition, a proper color space can solve the illumination problem in face recognition (Kovac, Peer & Solina, 2003).

Another issue in twin recognition is the ageing problem because wrinkles appear in a face with aging. One of the proposed models to solve this problem is called Age Classification, which detects the age of a person in photo based on his face (Kwon & da Vitoria Lobo, 1994; Nakano, Yasukata & Fukumi, 2004; Lanitis, Draganova & Christodoulou, 2004; Horng, Lee & Chen 2001). Another approach is to devise a model extracting the features concerning the age from the sample of face images containing already known age, and then reconstructing face images of each person for each age group. After that these reconstructed images are used as training samples for face recognition (Geng, Zhou & Smith-Miles, 2007; Hussein, 2002; Suo, Min, Zhu, Shan & Chen, 2007; Leta, Conci, Pamplona & Itanguy, 1996).

The third problem in twin recognition is the uniqueness of human faces (Gauthier & Logothetis, 2000). Twin recognition using face recognition is one of the most difficult tasks (Mahalingam & Ricanek, 2013), which has greatly challenged the conventional face recognition. Recently, a handful of researchers have paid their attention to this problem and have obtained notable achievements.

Human visual system can work much better than computational machines in twin recognition(Biswas, Bowyer & Flynn, 2011). The research outcomes have proved that human beings are good at utilizing facial marks such as moles and scars to effectively distinguish twins. In addition, if we can spend more time to carefully examine the difference, the performance will become much better. Klare, Paulino and Jain (2011) identify twins with human facial features such as eigenfaces as the level 1, Local Binary Patterns (LBP) as the level 2 and facial marks as level 3. The mouth shapes are relatively helpful to separate twins in typical data sets; especially the focus is on the age changing of twins (Le, Luu, Seshadri & Savvides, 2012). Therefore, they have employed facial aging features in twin identification. In another study, Vijayan et al. (2001) utilize a 3D model to recognize twins. Albeit a spate of good recognition has been achieved, it is still far from entirely solving the problem, e.g. closing the gap to human visual system.

Figure 1.1, 1.2 and 1.3 are the photos of a pair of twins. Figure 1.1 was taken when the twins were two months old. It is really hard to distinguish them. Figure 1.2 was the time when they were between two to ten years old, among these three images (a), (b) and (c), the left child is twin B and the right one is twin A. The three photos shown in Figure 1.3 were taken between 15 and 23 years old, among these three photos Twin A is at the left side and Twin B is at the right side. The twins look very similar in all the photos below.

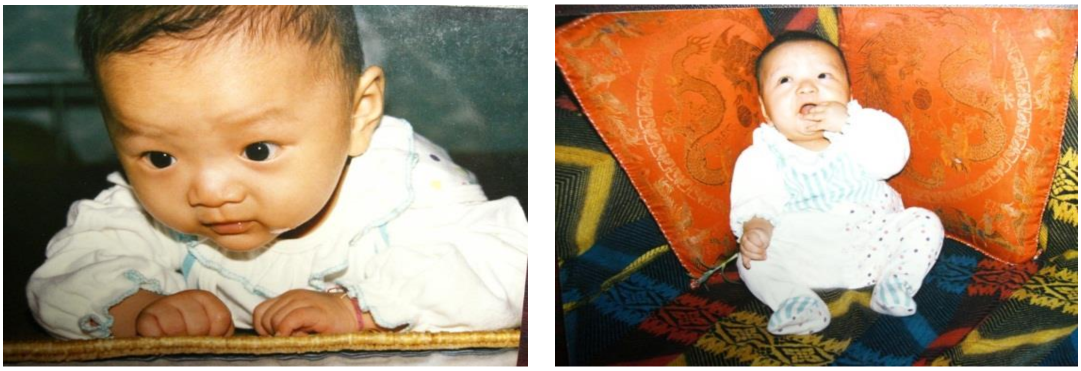
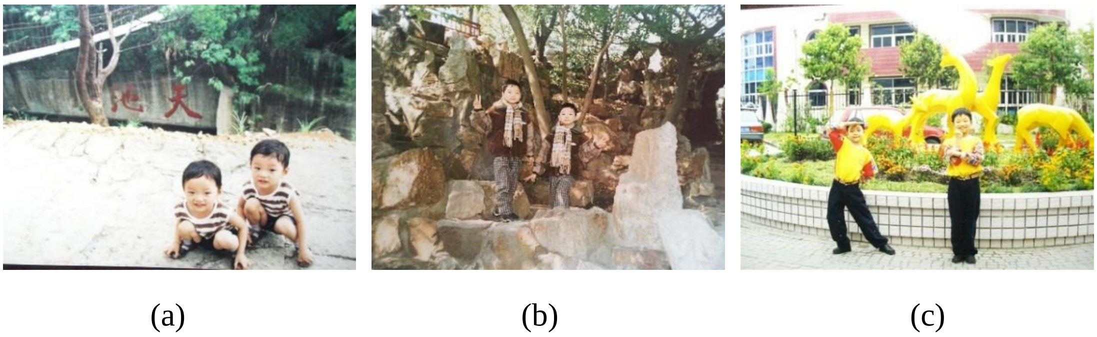
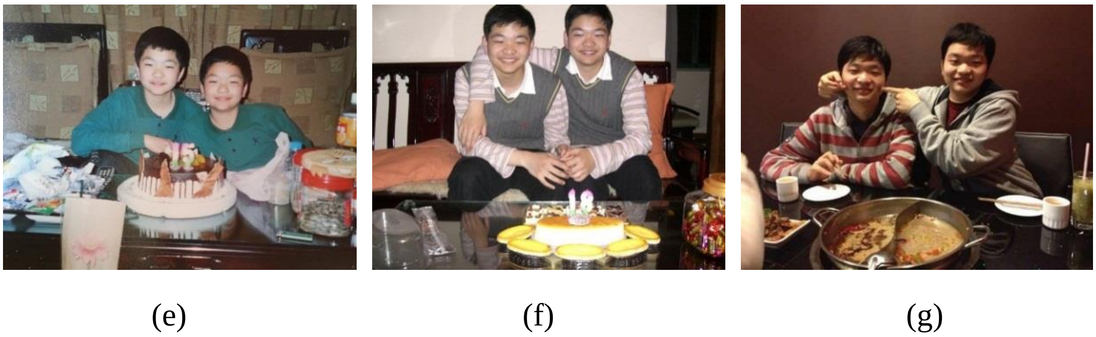


Figure 1.1 Twins of months old (Left A Right B)



(a) (b) (c)

Figure 1.2 Twins of 2, 6 and 10 years old



(e) (f) (g)

Figure 1.3 Twins of 15, 18 and 23 years old (ageing)

1.2 The Aims of This Thesis

The Chapter 1.1 introduces biometrics. Twin identification by DNA analysis has been introduced in literature, provided that expensive facilities are accessible (Ollikainen, Smith, Joo, Ng, Andronikos, Novakovic & Craig, 2010; Stewart, Evans, Bexon, van der Meer, & Williams, 2015). Even those, it still is significant to twin identification. Face recognition as one of the most important biometrics is also introduced before. The performance of face recognition is likely to become worse for recognizing twins. To address this issue, this thesis will analyze and compare the other three recognition methods based on biometrics including ear recognition, speaker recognition and lip movement recognition. The aim of this thesis is to find out the best method particularly for twin recognition.

1.3 Research questions

According to the related work, face recognition does not have excellent performance in twin recognition. Therefore, the research questions of this thesis are:

Question 1: Is it possible to recognize twins by using biometrics?

The following three points show the possibility for twin recognition by using biometrics.

(1) For physiological characteristics, twins are not exactly the same. For example, facial marks of twins usually are different (Srinivas, Aggarwal, Flynn, & Vorder Bruegge, 2012).

(2) For behavioral characteristics, habits of a person are formed in the long-term growing-up environment, even if the habits of twins are different.

(3) Multimodal biometric is possible to increase the successful recognition rate of twins (Sun, Paulino, Feng, Chai, Tan & Jain, 2010).

From the results of the experiments in this thesis, the performance of the four methods for face recognition, speaker recognition, ear recognition and lip movement recognition will be discussed.

Question 2: Which method has the best performance in twin recognition?

In this thesis, four biometrics approaches including face recognition, ear recognition, speaker recognition and lip movement recognition, will be analyzed and evaluated in terms of their performance in twin recognition. Finally, the approach with the best performance will be further discussed.

1.4 The Structure of This Thesis

This thesis contains five chapters. The first chapter details the introduction. In Chapter 1, the background and the aim of this thesis will be introduced.

Chapter 2 is literature review. Numerous algorithms for face recognition, ear recognition, speaker recognition and lip movement recognition will be reviewed. The chapter is to provide readers insight of how the biometrics works in twin recognition.

Chapter 3 is related to Methodology. In Chapter 3 we will mainly detail design of this thesis project. The algorithms and flow charts of four methods will be explained in this chapter.

Chapter 4 is related to findings and discussions. The test environment and resultst data will be shown in this part. After that，the performance of each method will be analyzed and the research questions will be answered one by one in this chapter. Samples of experimental data will also be described in this section.

The last Chapter is the conclusion and future work. In this part, we will conclude the thesis and vision future work.

Chapter 2 Literature Review

In this chapter many recognition methods based on biometrics will be introduced. The first one is face recognition which contains two-fold: face detection and face recognition. Then other three recognition methods include ear recognition, speaker recognition and lip movement recognition are detailed. Ear recognition is the biometrics based on the shape of human ears. This method can be divided into two parts namely using whole ears and using ear parts. Speaker recognition utilizes voice to identify twins. This chapter presents three features in speaker recognition: pitch, cepstrum and Mel-Frequency Cepstral Coefficient (MFCC). Lip movement recognition uses speech habits to recognize a person which needs to track the shape of the lip when the speaker is speaking. Usually, it is used to combine with speaker recognition method in order to improve the accuracy.

2.1 Face detection

Face detection is the first step of the twin identification. The definition of face detection is to detect faces from the given images. A vast majority of research work has been studied in this area. From the research results of Tanaka and Farah (1993), using the whole face is more effective than using the partial of it in face detection.

Haar-like feature developed by Viola and Jones (2001) was used in real-time face recognition. The benefit of Haar-like features is the fast calculation. To achieve this, the Integral Image is adopted as shown in equation (2.1).

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \qquad (2.1)$$

In equation (2.1) i is the original image, $i(x', y')$ is the value of the point (x', y') in the image and $ii(x, y)$ is the value of Integral Images.

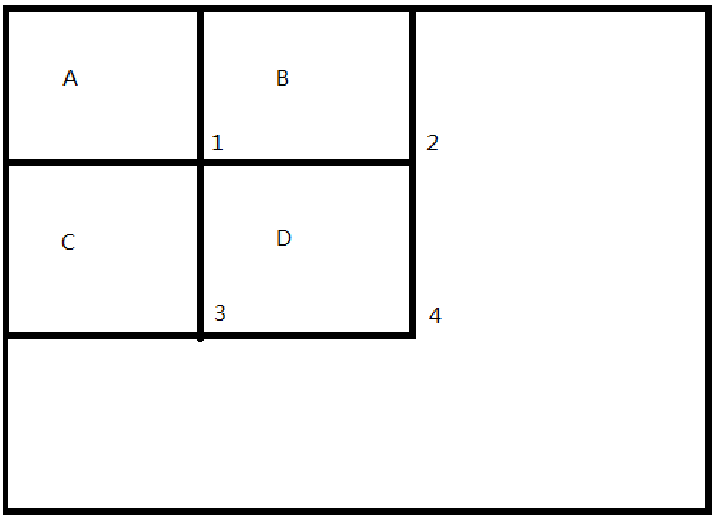


Figure 2.1 Integral Images

In Figure 2.1, the value of point 1 is the sum of pixel value of A in grayscale. The value of the point 2 is sum of all the pixel value of rectangular region A + B. The value of point 3 is the pixel value in rectangular range A and C, and the value of point 4 is the same thing in range A, B, C and D. The Integral Images make it possible to collect the value in a rectangular area rapidly. For example, if we want to get the sum of all pixel value in rectangle D, equation (2.1) can be used to calculate it,

$$D = p_4 - (p_3 + p_2) + p_1 \qquad (2.1)$$

where, p_1, p_3, p_3 and p_4 are the value of point 1, 2 ,3 and 4 respectively. D is the value of rectangle D.

2.2 Face recognition

There are a few methods in face recognition for twins. Dimensionality-reduction approach is the primary mission in these algorithms. The methods usually take the entire face as a whole for consideration (Delac, Grgic & Grgic, 2005).

2.2.1 Principal Component Analysis (PCA)

Principal Component Analysis (PCA) which adopts the feature eigenface is one of the effective algorithms in face recognition (Gottumukkal & Asari, 2004; Moghaddam & Pentland, 1997). PCA greatly reduces the computational complexity by diminishing dimension of features. In PCA method, each image is read as a vector. A matrix *M* is created to save all the training image as shown in equation (2.2),

$$\mathrm{M} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ . \\ . \\ v_n \end{pmatrix} \tag{2.2}$$

A method called 2D-PCA was developed for face recognition (Yang, Zhang, Frangi &Yang, 2004). Different from the traditional PCA algorithm, 2D-PCA adopts the 2D image matrices instead of image vectors. The testing result shows the method is more effective than traditional one.

The simplicity of PCA makes it easy to combine with other methods. The PCA and LDA are combined for face recognition (Zhao, Chellappa & Krishnaswamy, 1998). They used PCA to detect human faces and LDA is used to recognize those faces. PCA and a 3-dimensional (3-D) Facial Shape Model (FSM) were combined to make an age changing model for face recognition (Choi, 1999). PCA on wavelet sub-band is used to get more effective results (Feng, Yuen & Dai, 2000).

2.2.2 Independent Component Analysis (ICA)

ICA is a generalization of PCA. ICA and PCA are all belong to linear transformation. PCA is based on second-order statistics while ICA minimizes the high-order statistics (Delac, Grgic & Grgic 2005). According to the research work (Bartlett, Movellan & Sejnowski, 2002) ICA has better performance than PCA when the expression and testing days changed. PCA has worse performance than ICA when the recognizing face is moving (Draper, Baek, Bartlett & Beveridge, 2003). ICA is more effective than PCA in the properly compressed and whitened space (Liu & Wechsler, 1999). Another advantage of ICA is that can solve the problem of blind source separation which means there is no or little information for computer to separate data (James & Hesse, 2004).

2.2.3 Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) called Fisherface is another algorithm which is wildly used in face recognition area. It is one of the most effective dimensionality reduction techniques to solve the classification problems (Yu & Yang, 2001). The aim of LDA is to find the optimal discriminant feature. The two functions are shown in (2.3) and (2.4). As same as the Eigenfaces, Fisherfaces also are affected by illumination.

Between-class scatter matrix (S_B):

$$\mathrm{S}_B = \sum_{i=1}^{c} N_i (u_i - u)(u_i - u)^T \tag{2.3}$$

Within-class scatter matrix (S_w):

$$S_w = \sum_{i=1}^{c} \sum_{j=1}^{N_i} (x_j^i - u_i)(x_j^i - u_i)^T \tag{2.4}$$

where *u* is the mean of all samples, u_i is the mean of all samples in class *i*, N_i is the number of the samples in class *i* and x_j^i is the sample in class *i*.

Compared with these dimensionality-reduction approaches, PCA has better performance when the sample size is small (Martinez & Kak, 2001). Another advantage of PCA is that multiple images can be input while LDA only supports single impute image. In addition, PCA, ICA and LDA were compared in four different distance metrics: L1, L2, cosine, and Mahalanobis (Delac, Grgic & Grgic, 2005).

Apart from dimensionality-reduction approach there are also model-based approach and learning-based approach.

2.2.4 Local Binary Patterns (LBP)

Local Binary Patterns (LBP) is a kind of feature used in face recognition. Firstly, it divides image into small cells. Then, it compares each pixel in cell to its neighbors. If the grayscale values of the eight neighbor pixels are greater than the center pixel. The final step is to attach all the normalized histogram. This will be the feature vector of that image. The research work shows that facial images can be well described by LBP (Ahonen, Hadid & Pietikainen, 2006).

2.2.5 Active Appearance Models (AAM)

Active Appearance Models (AAM) is developed to trace feature points in a face (Cootes, Wheele, Walker & Taylor, 2002). After captured the feature points of the face, it will compare the distance of these points to analyze whether they are matching. In addition, this method is effective on catching the expression.

2.2.6 Support Vector Machines (SVM)

Support Vector Machines (SVM) which is developed by V. Vapnik is a kind of machine learning method (Osuna, Freund & Girosi, 1997). It uses associated learning algorithms to analyze sample data and recognize patterns. In twin recognition, using the facial components rather than the whole face significantly simplifies the Support Vector Machines (SVM) (Heisele, Ho & Poggio, 2001).

2.3 Ear recognition

Recently, twin recognition using human ears becomes a new class of relatively stable biometrics that has drawn attention of researchers. Human ear recognition as a biometric does not have significantly changes and this makes it can be used as a biometric (Kurniawan, Shafry, & Rahim, 2012). Ear recognition has its advantages and disadvantages. Compared with the face recognition, ears are not affected by human emotions (Kurniawan, Shafry, & Rahim, 2012).

Methods of ear recognition are divided into two-fold 2D and 3D. For 2D ear recognition, it still contains two classes. For the first class, the ear is considered as a whole, the second class of recognizes ear using the region of interest.

Methods using whole ears

This kind of methods regards the whole ear as an entirety. Generally, it contains statistical-based method and geometric-based method.

Statistical-based methods

Similar to the face recognition, statistically based methods analyze an ear image by statistical tools. The ear image is treated as a matrix. Then the method like PCA is used to get the features and reduce the redundancy of the data.

PCA as a statistical-based method is developed to recognize human ears (Chang, Bowyer, Sarkar & Victor, 2003). Their work analyzes the performance with the changing of the ageing, illumination and pose. In addition, they also compared face and ear with PCA. The result shows that there is no significant difference between them. The performances are 71.6 % for the ear recognition and 70.5 % for face recognition.

Fisher Linear Discriminant (FLD) are utilized to recognize human ears (Li, Yuan, Sang & Li, 2013). USTB ear database is used to recognize human ears and 92.5% recognition rate is obtained.

Geometric-based Methods

Geometric-based method uses shapes of human ear to identify twins.

Two methods: Concentric Circle Method (CCM) and Contour Tracing Method (CTM) are used to recognize human ears (Choras, 2004). The center point of a human ear is its center of mass and concentric circles are created by this point and predefined radius. Then the intersections of these circles and ear edges are feature points of the ear. CTM is the method to trace the ear edges so as to find the contour of bifurcation, intersection and ending. In this work, CCM is found better than that of CTM.

A method called max-line is a kind of geometrically based method to recognize the ear (Shailaja & Gupta, 2006). It uses the longest line from one side of the edges of human ear to the other side as the feature. The successful recognition rate of this method is around 80%.

Methods using ear parts

In these methods ear parts are employed for twin recognition. We decomposed a human ear into many parts. The benefit of these methods is that human ear recognition will be less affected by occlusion, because it only needs partial ear.

Improved non-negative matrix factorization with sparseness constraints (INMFSC) is this kind of ear recognition method, it recognized ear by separating it into three parts (Yuan, Mu, Zhang & Liu, 2006). Each part uses non-negative matrix factorization to get the feature vector. The occlusion problem is able to be solved by recognizing human ears using the sub-regions.

2.4 Speaker recognition

Voice recognition as another meaningful biometrics has been considered in this thesis. Three features namely pitch, cepstrum and MFCC which are popular in voice recognition will be introduced.

2.4.1 Pitch

The voice consists of voiceless and voiced sound. The vocal cords vibrate when the sound is voiced while for the voiceless it does not. The voiced sound can be decomposed into periodic waveforms. Pitch is the lowest frequency of these waveforms, it does not effect by which the person says, so it can be used as a feature of speaker recognition.

This feature now is combined with other methods together. For example, it can be combined used with CMM-based speaker recognition system, the work shows the performance of this system is 20% higher than the conventional method (Jian-wei, Shui-fa, Xiao-li & Bang-jun, 2009).

2.4.2 Cepstrum

Cepstrum is a widely used method in speaker recognition which utilizes the Inverse Fourier Transform (IFT) to the logarithm of the estimated spectrum of the original signal. It can be seen as rate of changes in the different spectrum bands.

2.4.3 Mel-frequency cepstral coefficient (MFCC)

MFCC is a special cepstrum which is more suitable for the human ear. MFCC is one of the best feature in describing the human voice feature. The performance of using Dynamic Time Warping (DTW) with MFCC is better than only using MFCC (Muda, Begam & Elamvazuthi, 2010).

2.5 Lip movement recognition

Lip movement recognition is a kind of biometrics that is proposed in recent years which is always used with speaker recognition to improve accuracy. These kinds of methods identify the person by using the information of dynamics in mouth region (Aleksic, 2009).

A system combines speech recognition and lip movement recognition has been developed, a fast lip localization algorithm is colored to track the lip movement (Chan, Zhang & Huang, 1998).

Chapter 3 Methodology

Twin recognition is a problem in biometric. The high similarity between twins makes it tough to be distinguished. Section 3.1 shows our research environment. The methods used in this thesis are presented. Although twin recognition has been discussed before, there is little work related to comparisons of the performance of those recognition methods. In this chapter four methods used in this thesis are introduced namely face recognition, ear recognition, speaker recognition and lip movement recognition. Accuracy, precision, recall and F-measure are used to evaluate these four methods.

3.1 Research environment

Our research project is undertaken in indoor environment so as to ensure that the data collected for speaker recognition contain little noises, face recognition has sufficient and constant illumination and have simplified background that without moving objects.

Twin recognition is one of the main problems in biometrics and pattern recognition. Biometrics refers to recognize human beings by the characteristics (Jain et al., 2004). Normally, it is classified into physiological and behavioral biometrics. Physiological biometrics refers to utilizing human's physiological characteristics, such as face, finger print, hand geometry to identify this person, while behavioral biometrics exploits human's action characteristics for identification (Jain et al., 2004). For example, gait is one of behavioral characteristics.

In computer science, pattern recognition is a branch of artificial intelligence comprising supervised and unsupervised learning (Bishop, 2006). Supervised learning refers to using data set that already has correctly labeled output to train the algorithm. This data set is also called training data. Unsupervised learning algorithm not only needs to differentiate the training data, but also classifies new data. Generally, pattern recognition is able to solve classification problem.

In pattern recognition, the problems of twin recognition are mainly discussed in face recognition. The convenience of data collection makes it widely used and easily to be implemented. However, the high similarities of biometrics between twins impede its development.

The section 3.2 will introduce our questions and methods in biometrics. Through completely study these questions, the method with prominent performance is expected to be identified for twin recognition, which is also the main objective of this thesis.

In this thesis, Matlab and Visual Studio 2012 are utilized as our development tools. The voice recognition and ear recognition are implemented in Matlab, while face recognition and lip movement recognition are conducted in Visual Studio with VC++.

3.2 Data set

The samples in this experiment are collected by ourselves. All of the data come from the twin brother. The data contain images and audios as well as labeling information.

For speaker recognition, 50 simple words spoken by the twins are recorded as waveform files. 25 of them are used for training and the rest of them are adopted for testing. Table 3.1 is the whole list of these words.

Table 3.1 Sample list of speaker recognition.

Word List				
1. hello	2. yes	3. O.K.	4. No	5. thanks
6. good	7. hi	8. no problem	9. what	10.why
11. I	12. you	13. come	14. in	15. out
16. more	17. go	18. do	19. he	20. she
21. had	22. mother	23. father	24. on	25. wait
26. easy	27. normal	28. hard	29. example	30. like
31. hate	32. run	33. eat	34. white	35. black
36. red	37. yellow	38. blue	39. fast	40. slow
41. ball	42. name	43. final	44. people	45. sea
46. hill	47. ice	48. juice	49. clod	50. hot

For face recognition, the samples are the facial images of the twins. Figure 4.1 shows samples of twin A and twin B for the purpose of comparisons.

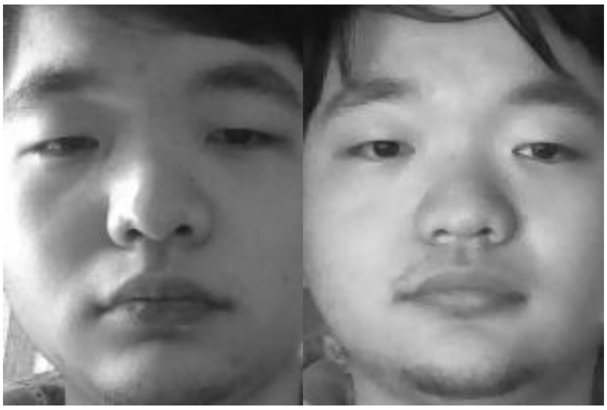
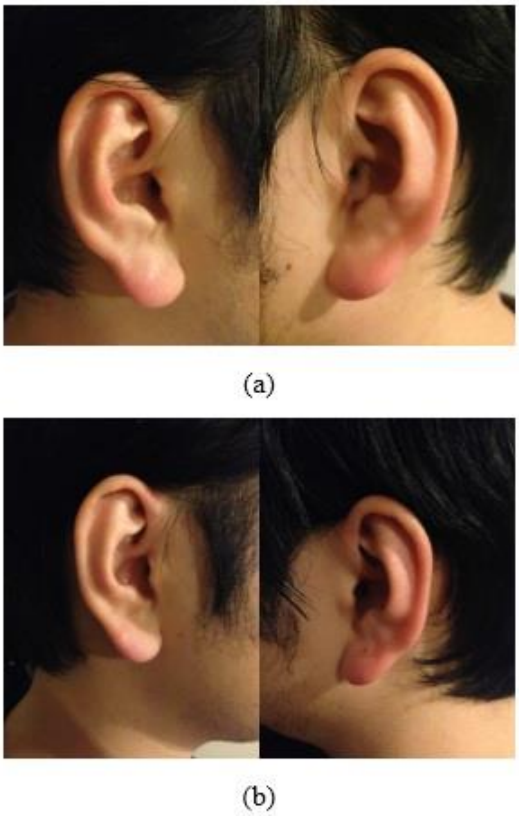


Figure 3.1 Face samples of twins. The left one is twin A and the right one is twin B

Figure 3.2 shows some samples of ear recognition. The top two photos are from twin A and the others are from twin B.



(a)

(b)

Figure 3.2 Samples of twin ears. (a) twin A (b) twin B

3.3 Methods and evaluations

3.3.1 PCA algorithm

Principal Component Analysis (PCA) is a dimensionality reduction method which can reduce the size of the data and makes it twin recognition easy. In this thesis PCA is employed to catch the features of face and ear images.

Firstly equation (3.1) calculates the average value of all feature vectors of training images, f_n is the feature vector and M is the number of the training image,

$$\Psi = \frac{1}{M}\sum_{n=1}^{M} f_n \tag{3.1}$$

The second step is to find difference of each feature vector like equation (3.2),

$$\Phi_i = f_i - \Psi \tag{3.2}$$

The next step is to get the eigenvalues and eigenvectors. Equation (3.3) is the covariance matrix which is utilized to calculate the similarity of two matrix.

$$\mathrm{C} = \mathrm{AA}^T \tag{3.3}$$

In this function the matrix A is$[\Phi_1, \Phi_2, \Phi_3, \dots \dots, \Phi_m]$.

$$\mathrm{AA}^T v_i = u_i v_i \tag{3.4}$$

According to the definition of eigenvectors, the equation (3.4) can be used to calculate the eigenvalues and eigenvectors. In this equation u_i is eigenvalues and v_i is eigenvectors. After we get the eigenvector, the equation (3.5) can be used to calculate the eigenface e_i,

$$\mathrm{e}_i = \sum_{k=1}^{M} v_{ik}\Phi_k \qquad\qquad i = 1,2, \dots \dots, M \tag{3.5}$$

Then next step is to recognize the face. In this part, firstly equation (3.6) is used to calculate the weight of the test face image. The equation is the set of these weights.

$$\omega_k = u_k^T(\nu - \Psi) \tag{3.6}$$

$$\Omega^T = [\omega_1, \omega_2, \omega_3, \dots \dots, \omega_m] \tag{3.7}$$

After we get the weighty sets, Euclidean Distance between Ω and Ω_k needs to be calculated. In equation (3.8), Ω_k is the average of all face images that belongs to class k, and Ω is the testing image. The test will belong to class k, if ε_k is smaller enough.

$$\varepsilon_k = ||\Omega - \Omega_k|| \tag{3.8}$$

3.3.2 MFCC

Mel-Frequency Cepstral Coefficient (MFCC) is a kind of features that is popular used in voice recognition. To get the MFCC feature, the first step is frame blocking. In this step the signal is segmented 20 to 30 msec. Each block contains N samples and the next block moves *M* samples. In this thesis *N* is set on 256 and *M* 80.

A hamming window is used to minimize the disruptions of signal. In equation $\mathrm{s(n)}$ is the signal, $\mathrm{w(n)}$ is the hamming window and $\mathrm{S'(n)}$ is the output signal. Function (3.10) is taken advantage to build the hamming window. In this thesis β equals 0.46.

$$\mathrm{S'(n) = s(n) \times w(n)} \tag{3.9}$$

$$\mathrm{W(n,a)} = (1-\beta) - \beta \times \cos\left(\frac{2\pi n}{N-1}\right),\ 0 \le \mathrm{n} \le \mathrm{N} - 1 \tag{3.10}$$

After that, the signal is transformed from time domain to frequency domain by using Fast Fourier Transform (FFT).

The signal will be put through Mel Filter Bank in this step. Equation (3.11) shows how to make the given frequency f to Mel scale. It is one of the most widely used MFCC transform function (Ganchev, Fakotakis & Kokkinakis, 2005).

$$\mathrm{F(Mel)} = 2595 \times \log 10\,(1 + \frac{f}{700}) \tag{3.11}$$

3.3.3 VQLBG

Linde–Buzo–Gray (VQLBG) is Vector Quantization (VQ) using the Linde-Buzo-Gray algorithm. Linde-Buzo-Gray algorithm is to derive a good codebook for VQ (Linde, Buzo and Gray, 1980). In the equation (3.12), c_i is the codeword and X_{ij} is the feature of MFCC,

$$\mathrm{c}_i = \frac{1}{M}\sum_{j=1}^{M} X_{ij} \tag{3.12}$$

The equation (3.13) and (3.14) show the splitting law of the codebook. In these two equations *m* is the number of codevectors.

$$C_m^+ = C_m(1+\varepsilon) \tag{3.13}$$

$$C_m^- = C_m(1-\varepsilon) \tag{3.14}$$

After we get the initial codebook, iterations are considered to improve it. Equation (3.15) is applied to calculate the distortion. In this equation *n* is the times of iteration and *C* is the codebook.

$$\mathrm{D}^{(n)} = \sum_{k=1}^{K} \min d(X_k, C) \tag{3.15}$$

Equation (3.15) is the average distortion. The iteration will stop if the average distortion is less than a threshold value. Otherwise, the average value of each cluster of the current codebook will be calculated as the new codeword and the iteration will be continued.

$$D_{ave}^{(n)} = \left|\frac{\mathrm{D}^{(n-1)} - \mathrm{D}^{(n)}}{\mathrm{D}^{(n)}}\right| \tag{3.16}$$

3.3.4 Evaluation

In this thesis, four methods are employed to conduct the twin recognition: face recognition, ear recognition, speaker recognition and lip movement recognition. To compare the performance, this thesis take account of these four measurements accuracy, precision, recall and F-measure. These four measurements are for binary classification because the twin recognition is a binary classification problem. Therefore this method is suitable for this thesis.

Accuracy means the correct percentage of a classification. However, sometimes it is not enough to show the performance of a classifier machine. For example, there are two classes, A covers 99% and B covers 1%. The correct rate will be 99% if the negtive classification machine actually does not conduct the classified work just put all the samples into A. In this case, accuracy cannot evaluate the classifier accurately. Therefore, the precision, recall and F-measure are combined to evaluate the performance of a classification machine. Precision is also called the positive predictive value. It refers to the percentage of the real correct objects among the objects predicted as the correct ones. For example, precision of twin A is the correct rate of all the result predicted as twin A by the system. Higher value of precision means the higher accuracy rate. Recall is also called true positive rate. Recall is for evaluating the catching extent of a classification machine which means the machine catches how many samples for one class. Precision and recall are the essential components of F-measure. F-measure can be regarded as the score for the performance of the classification machine, which is related to precision and recall. The best score is 1 and the worst score is 0.

$$\text{Accuracy} = \frac{\text{AA+BB}}{\text{TC}} \times 100\% \tag{3.17}$$

$$\text{Precision}_A = \frac{\text{AA}}{\text{AA+BA}} \times 100\% \tag{3.18}$$

$$\text{Recall}_A = \frac{\text{AA}}{\text{AA+AB}} \times 100\% \tag{3.19}$$

$$\text{Precision}_B = \frac{\text{BB}}{\text{BB+AB}} \times 100\% \tag{3.20}$$

$$\text{Recall}_B = \frac{\text{BB}}{\text{BB+BA}} \times 100\% \tag{3.21}$$

$$\text{F} - \text{measure}_A = \frac{2\times\text{Precision}_A\times\text{Recall}_A}{\text{Precision}_A+\text{Recall}_A} \times 100\% \tag{3.22}$$

$$\text{F} - \text{measure}_B = \frac{2\times\text{Precision}_B\times\text{Recall}_B}{\text{Precision}_B+\text{Recall}_B} \times 100\% \tag{3.23}$$

where *AA* means twin A is recognized as twin A, *BB* means twin B is recognized as twin B, *AB* means twin A is recognized as twin B and *BA* means twin B is recognized as twin A.

3.4 Algorithms

3.4.1 Four algorithms

Algorithm 3.1: Face recognition

Input: Training face images and testing face images

Output: the name of the face

```
Read the training face images
Apply cvCalcEigenObjects to calculate eigenvector of the training face image
Use cvEigenDecomposite to reflect training face image into PCA space and create the eigenface of each person $X_i$
Detective human face with Haar-like feature
If face detected
Normalized image with cvNormalize
```

```
Use cvEigenDecomposite to reflect testing face image into PCA space and saved as $X'$
If the distance between training face image class $X_i$ and testing face image $X'$ is less than threshold value
Display the name of the face
```

In algorithm 3.1 the inputs are training face images, Haar-like feature is used to detect human face, the function cvCalcEigenObjects is applyed to get the eigenvector and eigenvalue of the training face image and *cvEigenDecomposite* is employed to calculate the PCA space for both testing and training image. The testing image needs to be normalized with the method *cvNormalize* firstly. Then it will project into PCA space and calculate the distance with all training image classes. The nearest one is regarded as the recognized person. The output is the name of this person.

Algorithm 3.2: Lip movement recognition

Input: Training videos and testing Video

Output: the name of the face

```
Read the training videos
Detective human face with Haar-like feature
Use the flandmark_detect to detective the mouth-corner
Save the coordinate of two points $x_1, y_1, x_2, y_2$
Calculate the average value of the mouth-corner
        $X = (x_1 + x_2)/2,\ Y = (y_1 + y_2)/2$
Calculate the average value of each frame $\bar{X}, \bar{Y}$
Read the testing video
Detective human face
If face detected
Use the flandmark_detect to detective the mouth-corner
Save the coordinate of two points $x_1', y_1', x_2', y_2'$
```

```
Calculate the average value of the mouth-corner
  $X' = (x'_1 + x'_2)/2,\ Y' = (y'_1 + y'_2)/2$
Calculate the distance between point $(X',\ Y')$ and point $(\bar{X},\ \bar{Y})$
If the distance between training value and testing value is less than threshold value
Display the name of the person
```

Algorithm 3.2 is to explain lip movement recognition. As same as face recognition, Haar-like feature is used to detect human face. After that a method called *flandmark_detect* which is based on Deformable Part Models (DPM) is considered to get the coordinate of the mouth-corner (Uřičář, Franc & Hlaváč, 2012). The average value of two mouth-corners is the feature of this method. The distance between feature vectors of testing video and training video is applied to classify them. The output is the name of the person that the method recognized.

```
Algorithm 3.3: Speaker recognition
Input: Training sound files and testing sound file
Output: the number of recognized person
Read training sound files
$i = 1$
If $i$ < size of training set, then
    Get the MFCC feature of the training sound [i]
    Use VQLBG to create the VQ code book of the training sound [i]
    Save the VQ code book as code book [i]
Read testing sound file
Get the MFCC feature of the testing sound file
$j = 1$
If $j$ < size of training set, then
Calculate the distance between code book [i] and testing sound files
    If distance < nearest distance, then
```

```
        nearest distance ← distance
        the number of recognized person ← i
Return the number of recognized person
```

Algorithm 3.3 is for speaker recognition. The input is voice files. The MFCC feature vectors are calculated from both the training set and the testing set. The distance between these feature vectors is used to identify twins. The output is the ID number of the voice file which is the most similar one.

Algorithm 3.4: ear recognition

Input: Training ear images and testing ear image

Output: the number of recognized person

```
Read training ear images
Transform all images into vector
Combine them together and save as a matrix allsamples. Each row is a sample and the column the matrix is the pixel of each image.
Calculate the average value of each column of the matrix and save as a vector samplemean
Matrix xmean is each row of allsamples minus samplemean
Squared the matrix xmean
Use eig function in Matlab to get the eigenvalues and eigenvectors of the matrix xmean
Calculate PCA space with the eigenvalues and eigenvectors.
Read the testing image
Transform the testing image into PCA space.
Calculate the distance between testing image and training images
Save the number of the nearest one as the number of recognized person
Return the number of recognized person
```

The Algorithm 3.4 is related to ear recognition. The input is humane ear images. It adopts *eig*(·) function in Matlab to calculate the eigenvalues and eigenvectors. Then all the images are projected to the PCA space. The distance between training and testing images is calculated to identify them. The output of this algorithm is the ID number of recognized person.

3.4.2 Four flowcharts

Figure 3.3, 3.4, 3.5 and 3.6 show the algorithms in twin recognition. The details of them will be discussed in this Section.

Figure 3.3 is the workflow of the face recognition. In this method testing data is recorded by a camera, the cascade classifier is taken to detect the face. PCA is conducted to collect the features from the face image. Euclidean Distance is employed to classify the face image. The output of this method is the name of the twins.

The lip movement recognition is based on behavioral biometrics. When a speaker is talking, we assume that the mouth-corners are moving. It is said that different speakers have different habits of lip movement that generate various experimental data. Therefore it belongs to behavioral characteristics to identify a person.

Figure 3.4 is the workflow of the lip movement recognition. In this method it uses cascade classifier to detect a face, and use Deformable Part Models to detect mouth-corners. After that, the average of these two points is calculated as one of the features of this method. The output of this method is the labels of the twins.

For speaker recognition, MFCC, as one of the effective features, is chosen in the experiment. Compared with cepstrum, MFCC is based on human auditory system (Ganchev, Fakotakis & Kokkinakis, 2005). The perception of human auditory system does not follow a linear scale for voice signal (Tiwari, 2010). Thus the nonlinear "Mel Scale" is more suitable for speaker recognition. In this method, the output is the ID number of the training audio file of the recognized person.

Figure 3.5 is the workflow of speaker recognition which contains two parts: training part and testing part. In the training part, it extarcts MFCC feature. The VQLBG is also implemented to create VQ codebook for MFCC features. In the testing part, Euclidean Distance between MFCC features of the testing data and VQ codebook is calculated to classify voice files.

For ear recognition, it is also based on physiological characteristics. The PCA is chosen as the feature selection algorithm. The ear images are taken account for the experiments.

Figure 3.6 is the workflow for ear recognition. In this method, firstly we input the training ear images, and then calculate the PCA features of them. The next step is to extract features of testing ear images. After that, we compare all the PCA features using SVM and find the nearest one. The ID number of the found picture is the output of the algorithm.

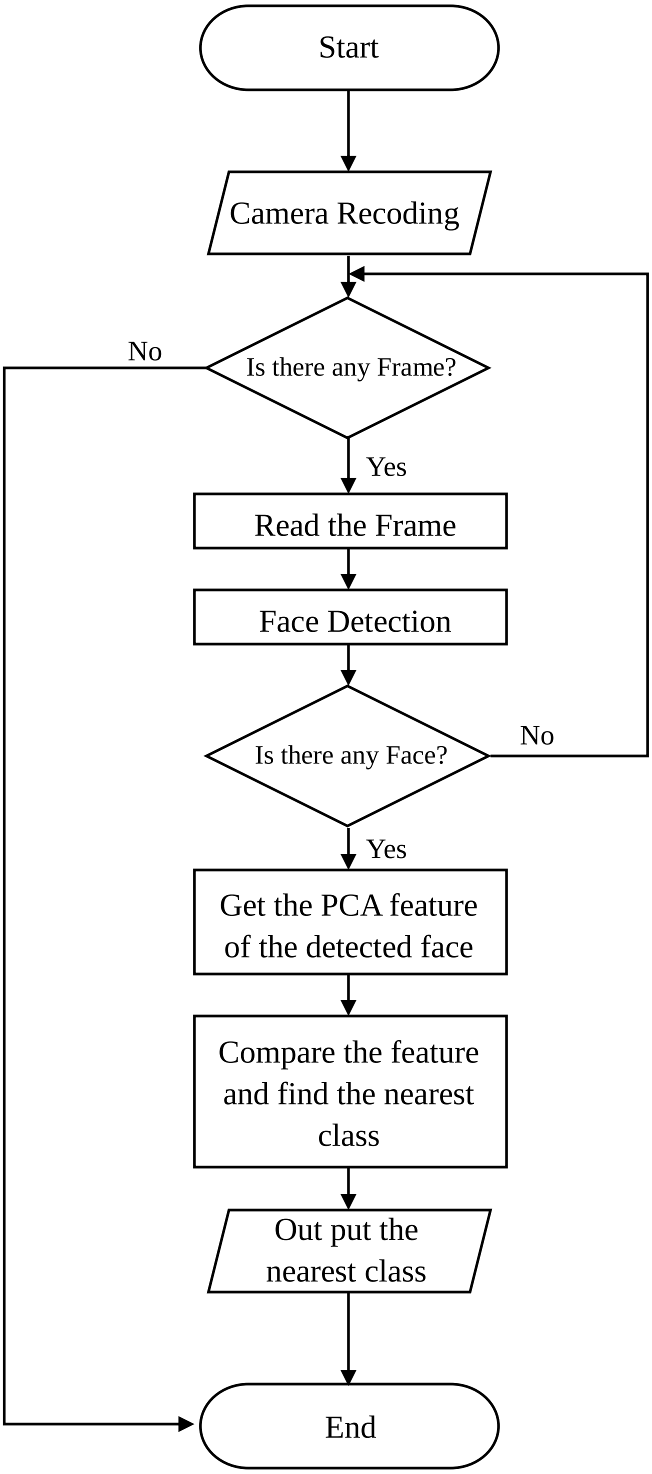

Start
Camera Recoding
Is there any Frame?
No
Yes
Read the Frame
Face Detection
Is there any Face?
No
Yes
Get the PCA feature of the detected face
Compare the feature and find the nearest class
Out put the nearest class
End

Figure 3.3 Work flow of the face recognition system

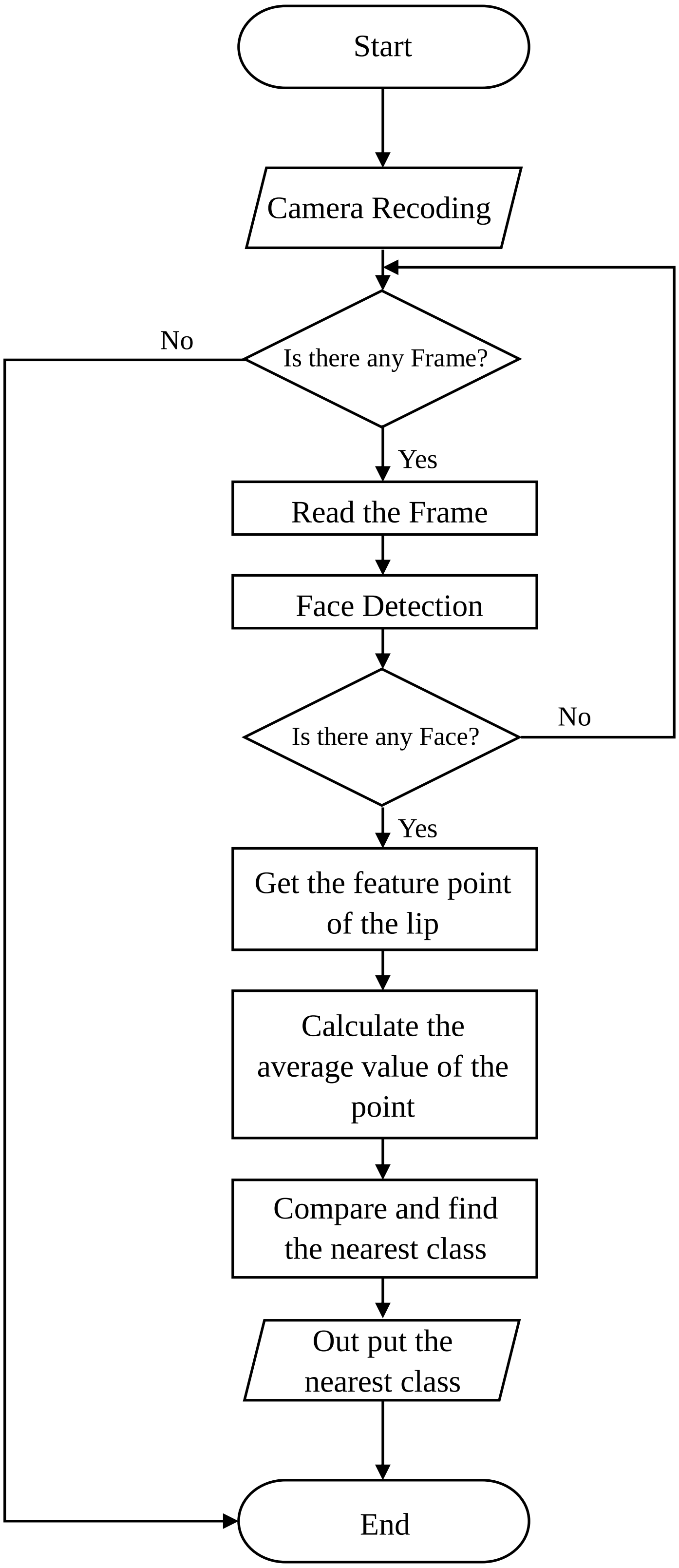

Start
Camera Recoding
Is there any Frame?
No
Yes
Read the Frame
Face Detection
Is there any Face?
No
Yes
Get the feature point of the lip
Calculate the average value of the point
Compare and find the nearest class
Out put the nearest class
End

Figure 3.4 Work flow of the lip movement recognition system

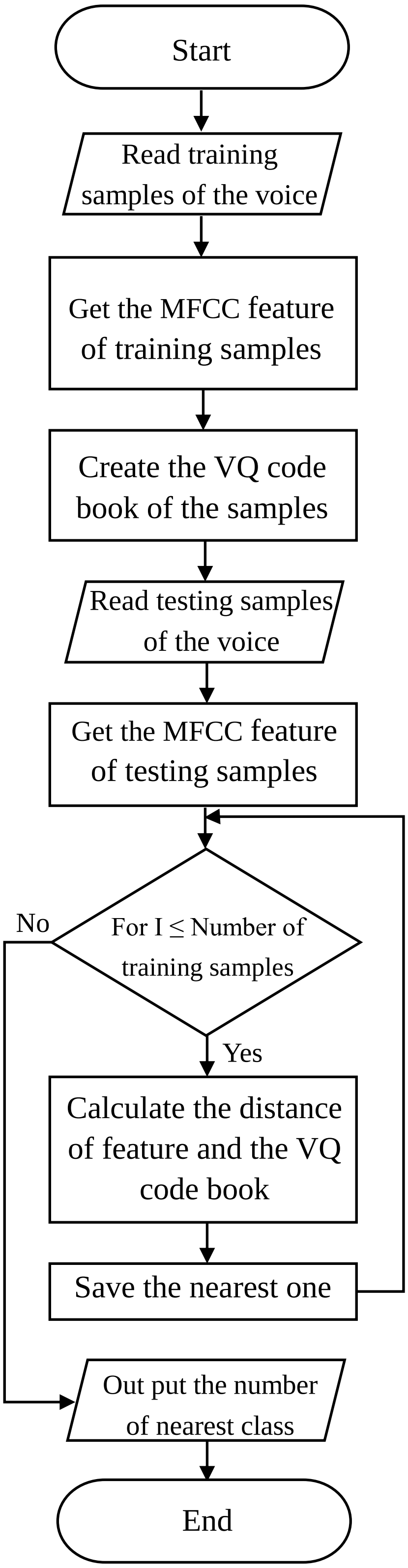

Start
Read training samples of the voice
Get the MFCC feature of training samples
Create the VQ code book of the samples
Read testing samples of the voice
Get the MFCC feature of testing samples
For I ≤ Number of training samples
No
Yes
Calculate the distance of feature and the VQ code book
Save the nearest one
Out put the number of nearest class
End

Figure 3.5 Work flow of the speaker recognition system

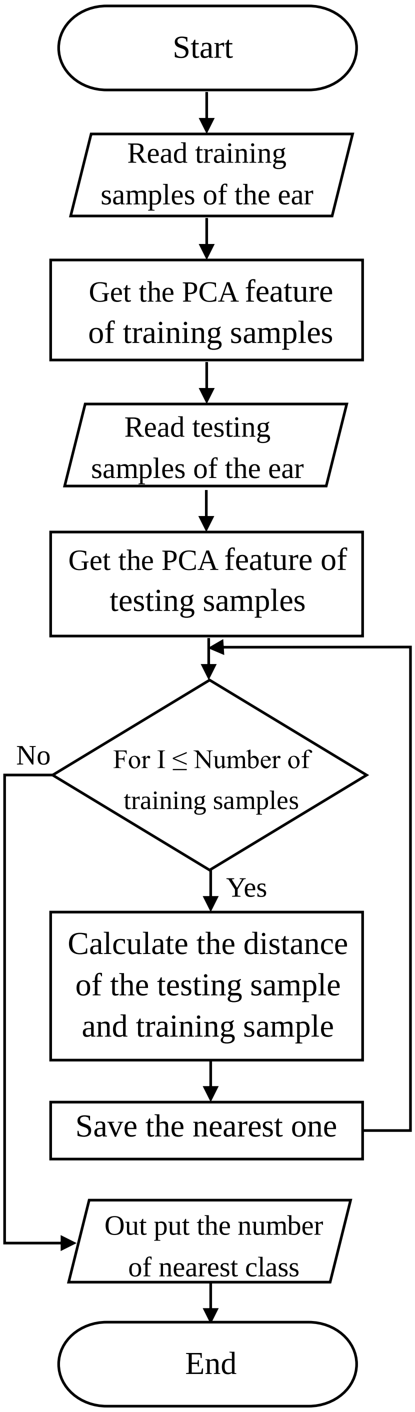

Start
Read training samples of the ear
Get the PCA feature of training samples
Read testing samples of the ear
Get the PCA feature of testing samples
For I ≤ Number of training samples
No
Yes
Calculate the distance of the testing sample and training sample
Save the nearest one
Out put the number of nearest class
End

Figure 3.6 Work flow of the ear recognition system

Chapter 4 Findings

In this thesis, four biometrics methods are implemented to test their performance in twin recognition. They are speaker recognition, ear recognition, face recognition and lip movement recognition. These methods include physiological and behavioral characteristics of the twins. The face recognition and ear recognition are related to the physiological characteristics, and the lip movement recognition tackles with the behavioral characteristics. Speaker recognition is concerned with both physiological and behavioral characteristics.

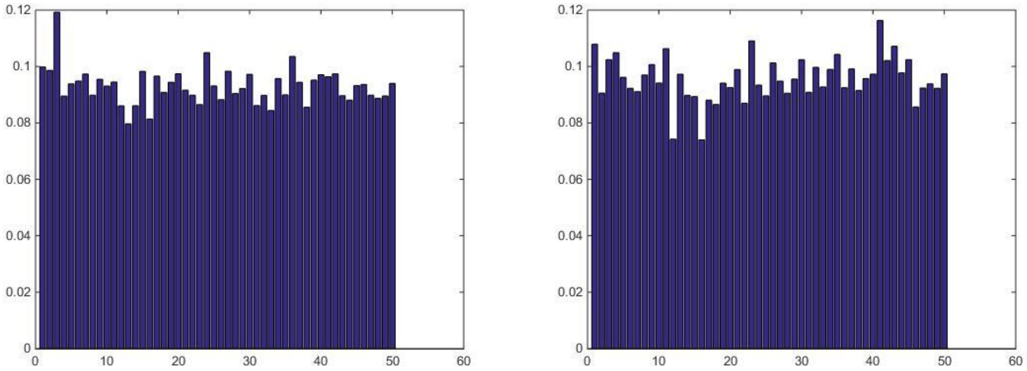
This chapter has two parts. The first part shows the result of twin recognition. For each method, there is a classification matrix containing four states: twin A is recognized as twin A, twin A is recognized as twin B, twin B is recognized as twin B and twin B is recognized as twin A. The second part of this chapter is the evaluation score of these four methods. The accuracy, precision, recall and F-measure are employed jointly to evaluate these methods. These marks can evaluate the system comprehensively.

4.1 Results

4.1.1 Speaker recognition

Speaker recognition is developed based on Matlab platform. The algorithm MFCC is employed to obtain the feature.

The capacity of the samples is 100, and all the sample words are listed in Table 3.1. The recording tool installed in the operating system is utilized for generating the voice files. In this test, 25 words spoken by each of the twins are assigned as training set and the other 50 voice files are regarded as the test examples.

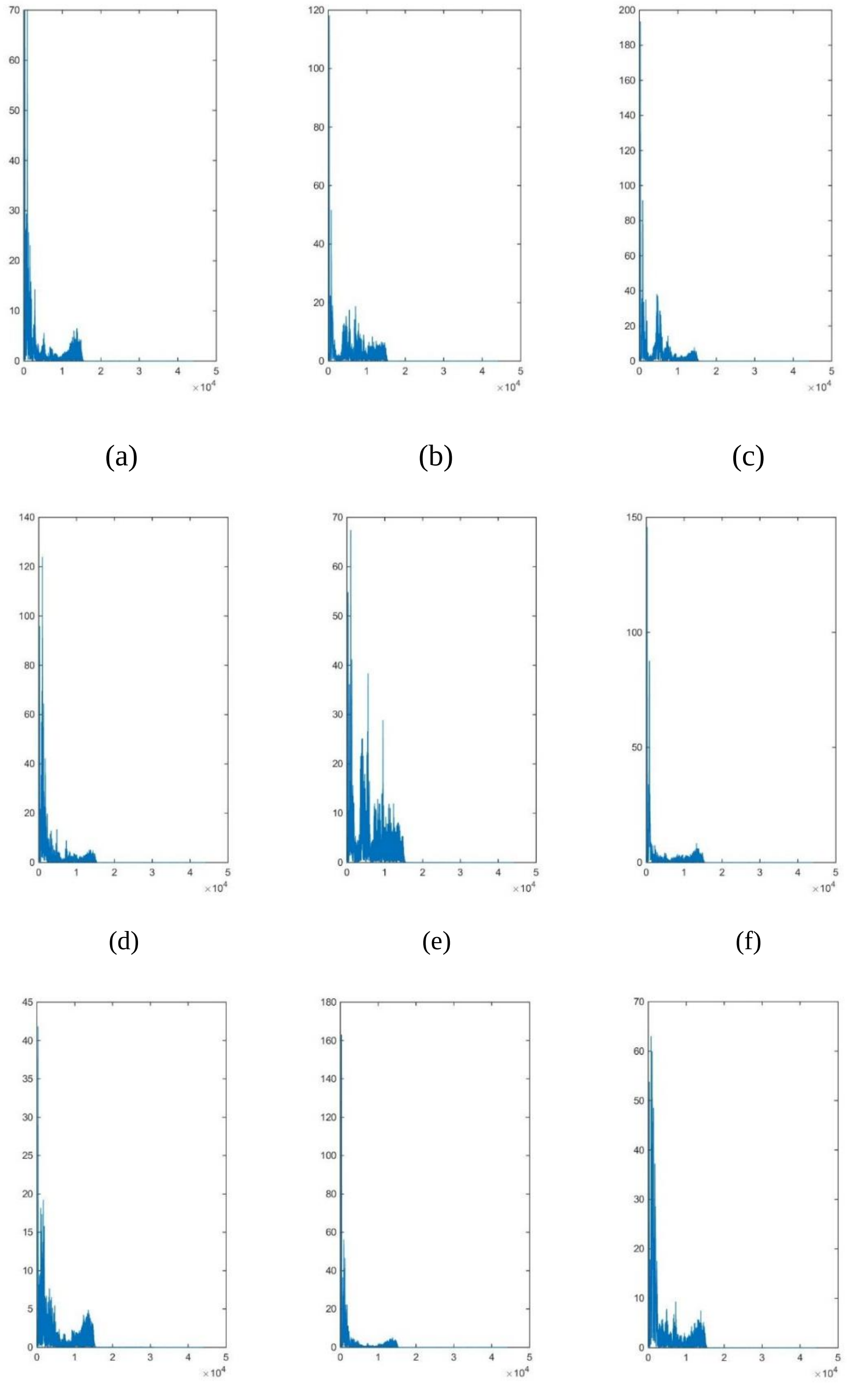


(a) (b)

Figure 4.1 Pitches of the test samples. (a) twin A (b) twin B

Figure 4.1 is the pitch of the 100 samples in speaker recognition. The left one is twin A and the right one is twin B. Most of them are around 0.08 to 0.1. There is no significant differences between them.

Figure 4.2 and 4.3 present the spectrogram of the twin A and twin B. The vertical axis is the amplitude. The figure 4.2 shows all of the highest amplitude in each picture is lower than 200, while in figure 4.3 the highest one reaches 1000.

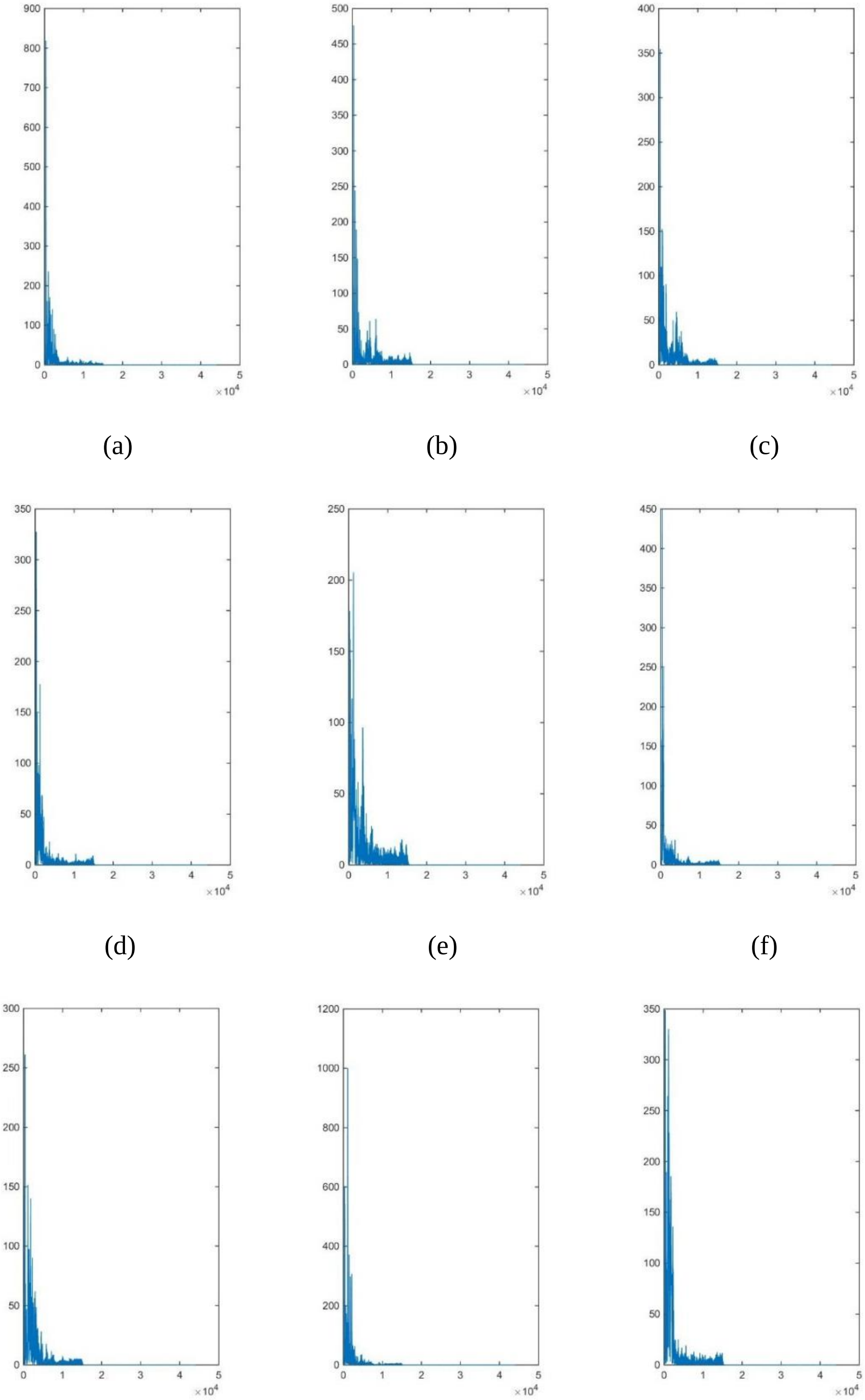


(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 4.2 Spectrogram of the twin A. Vertical axis is the amplitude of each voice and horizontal axis is frequency.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 4.3 Spectrogram of the twin B. Vertical axis is the amplitude of each voice and horizontal axis is frequency.

Equation (4.1) is to calculate the spectrogram of a voice signal,

$$S' = F^{-1}(ln|F(S(t))|) \tag{4.1}$$

where $S(t)$ is the signal of the voice. $F(\cdot)$means Fourier transform. $F^{-1}(\cdot)$ means inverse Fourier transform. S' is thought as the result signal.

For the speaker recognition, Table 4.1 shows the results of speaker recognition. The left row is the actual situation and the first line is the resultant list. For example, for the second line, the number 25 indicates that 25 sound samples of twin A recognized as twin A, while no sample of twin A is incorrectly recognized as twin B in speaker recognized system.

Table 4.1Classification matrix of speaker recognition

	Be recognized as twin A	**Be recognized as twin B**
Twin A	25	0
Twin B	0	25

4.1.2 Face recognition

For face recognition, Haar-like features are extracted to detect the human face from the background and PCA is used for face recognition. Face recognition is deployed in Visual Studio 2012 with VC++. The results are shown in figure 4.4, the left one is the twin A and the right one is for twin B.

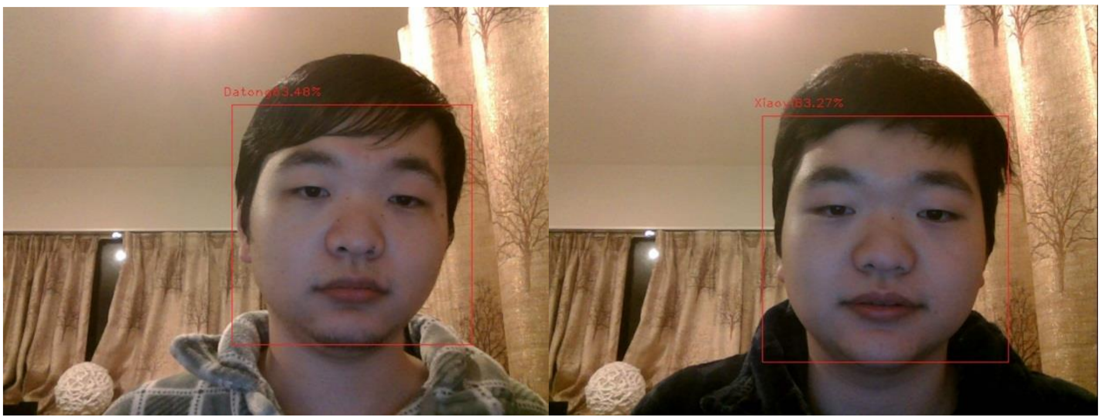


Figure 4.4 Result of face recognition (a) twin A (b) twin B

Table 4.2 shows the performance of face recognition in twin recognition. The stub of the table is the true situation and the box head of the table is the result of face recognition. For example, the number 128 refers to the number of frames that recognized twin A as twin A, and the number 80 means the number of frame that recognized twin A as twin B.

Table 4.2 Classification matrix of face recognition.

	Be recognized as twin A	**Be recognized as twin B**
Twin A	128	80
Twin B	107	116

4.1.3 Lip movement recognition

The lip movement recognition is also developed based on the Visual Studio 2012 with VC++. It used Deformable Part Models (DPM) to detect the corners of the mouth in a face. Figure 4.5 is the combination of lip movement recognition with the face detection. It displays the testing environment, the left one is twin A and the right one is twin B. The two green points highlight the corners of the mouth. The red point in the middle of the face highlights the nose.

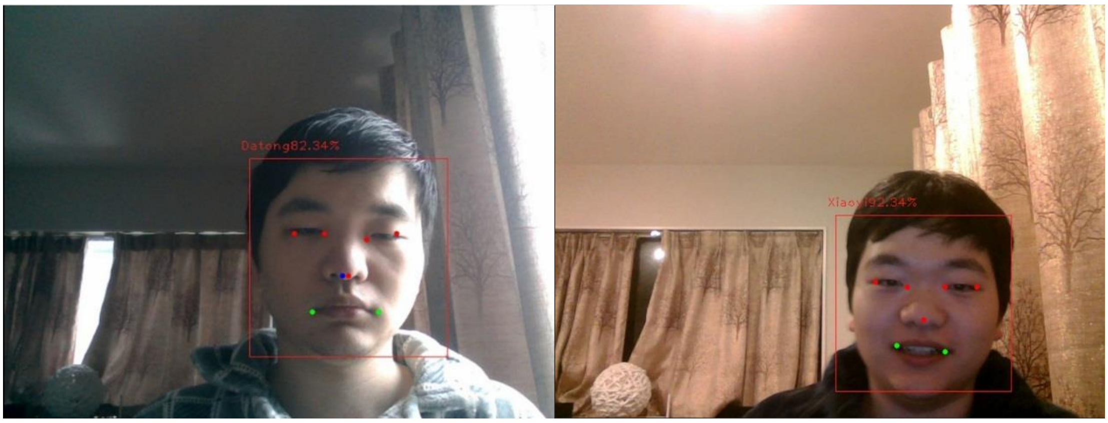


Figure 4.5 Results of lip movement recognition.

Table 4.3 shows the result of lip movement recognition. The stub of the table is the true situation and the box head of the table is the result of lip movement recognition.

Table 4.3 Classification matrix of lip movement.

	Be recognized as twin A	**Be recognized as twin B**
Twin A	158	65
Twin B	41	166

4.1.4 Human ear recognition

Human ear recognition is developed based on Matlab platform. It also exploits the PCA algorithm. The samples of the test include 30 human ear images for each of the twins. 10 of them are used to training and 20 of them are served as testing set for the recognition.

Table 4.4 shows the result of the ear recognition. For example, 12 means 12 ear images of twin A are recognized as twin A.

Table 4.4 Classification matrix of ear recognition.

	Be recognized as twin A	**Be recognized as twin B**
Twin A	12	8
Twin B	11	9

4.2 Evaluations

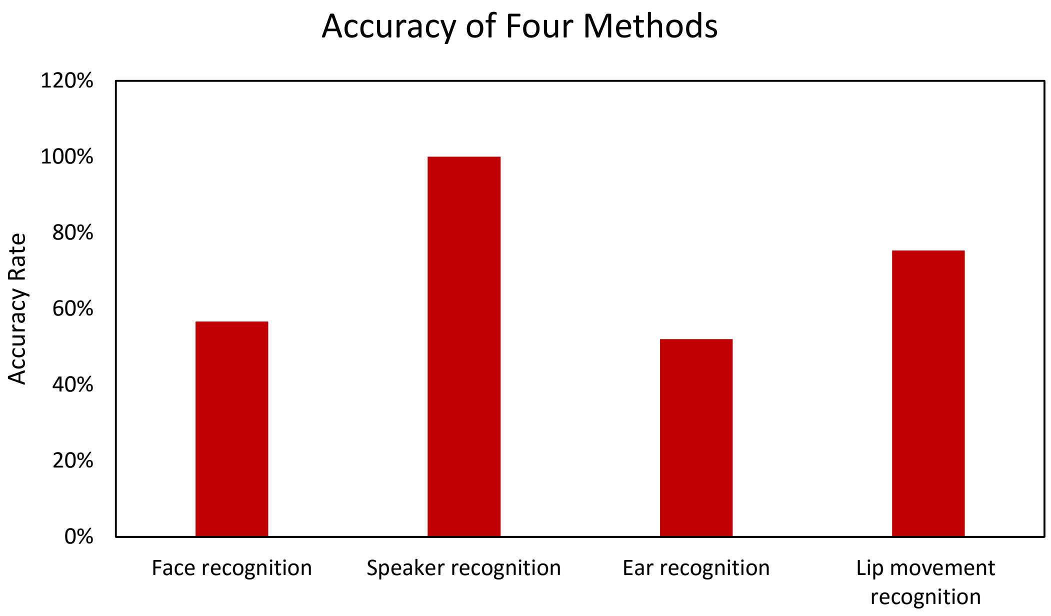


Figure 4.6 The accuracy of four methods in twin recognition.

Figure 4.6 shows the accuracy of four approaches in twin recognition. The speaker recognition has the highest rate of 100%. The lip movement recognition has the second highest correct rate that is 75%. The accuracy rate of face recognition is 57%. The ear recognition has the lowest result on accuracy with 53%.

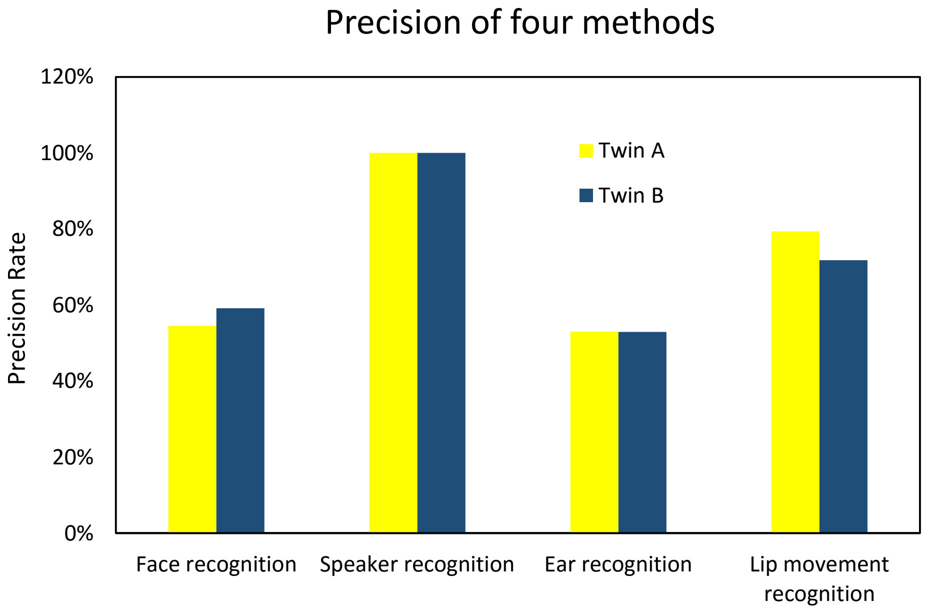


Figure 4.7 Precision of four methods in twin recognition.

Figure 4.7 is the precision of four methods in twin recognition. The blue bar in the chart is the twin A and the orange one is twin B. Considering twin A and twin B as a whole when analyzing this figure, for speaker recognition, the precision is 100%. Then the performance of lip movement system is significantly higher than other two methods. Ear recognition has the lowest precision rate in twin recognition.

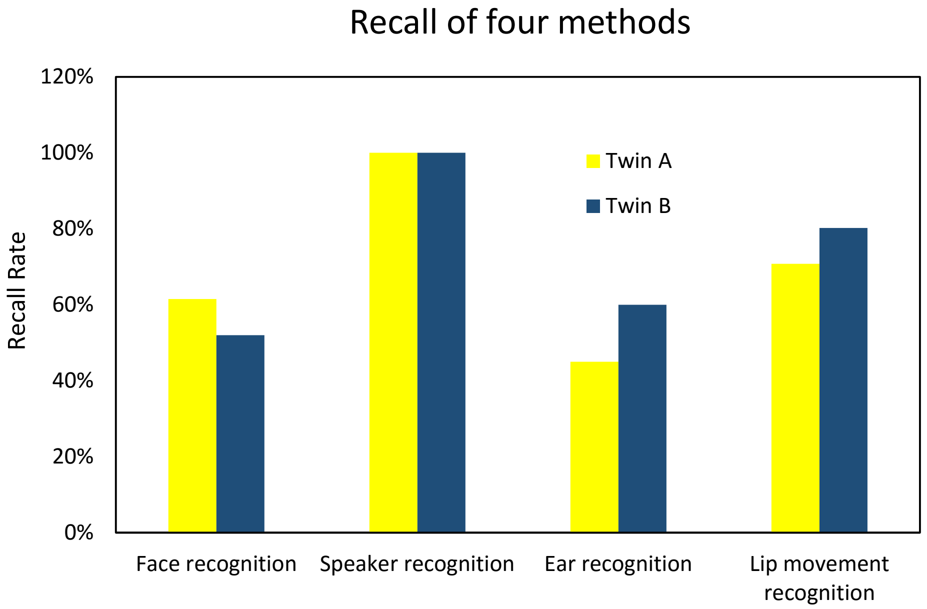


Figure 4.8 Recall of four methods in twin recognition.

Figure 4.8 is the recall of the four methods in twin recognition. Take into consideration of twin A and twin B as a whole when analyzing this figure, speaker recognition has the highest value. The recall of lip movement recognition shows high percent of 75%. The ear recognition and face recognition have the similar recall. The ear recognition has the lowest recall value.

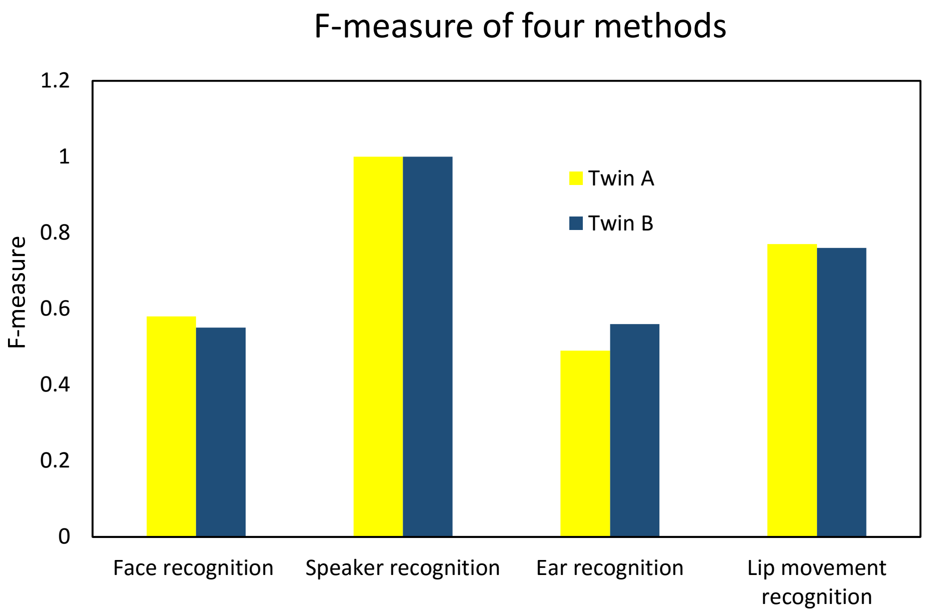


Figure 4.9 F-measure of four methods in twin recognition

Figure 4.9 shows the F-measure of the four methods, where '1' represents the highest mark and '0' represents the lowest mark. Considering twin A and twin B as a whole when analyzing this figure, the speaker recognition has the highest score 1. After that, lip movement has a good performance around 0.8. The F-measure of face recognition is nearly 0.6 while the ear recognition has the lowest mark of around 0.5.

Summary

Chapter 4 introduced the main findings of this research project which described the performance of the four methods in twin recognition. Finally, the most effective method is identified by evaluating these methods in terms of their accuracy, precision, recall and F-measure.

Chapter 5 Discussions and Analysis

In this chapter, these findings iare discussed. The coming sections analyzes the performance of these four methods: speaker recognition, face recognition, lip movement recognition and human ear recognition. In addition, the reasons of these results are further investigated in this chapter.

5.1 Analysis

The results show face recognition based on PCA does not have excellent performance in twin recognition. The accuracy rate is merely 57% and the F-measure is 0.58. Although there are minor differences between the twins in face recognition, but they are not enough for twin recognition.

Although some results claimed that it is possible to reach 90% confidence for twins recognition, but they are only ideal conditions such as: photos are taken in same day, studio lighting and the expression is neutral expression(Phillips, Flynn, Bowyer, Bruegge, Grother, Quinn & Pruitt, 2011).

The ear recognition algorithm shows it has the lowest performance among the four algorithms in twin recognition. The precision is only 53%. The reason why ear recognition does not have high performance is that the difference between the twins is too small to identify them.

Surprisingly, the results of twin speaker recognition show the precision is 100%. For the hypothesis, speaker recognition may have better correct rate than face recognition and lower than lip movement recognition. However, the results show it gets the highest mark.

There are two reasons of the high level in voice testing. The main characteristics of human voice is pitch, volume and timbre. In Figure 4.1, it shows that there is not significant difference between twin A and twin B in pitch. Therefore, pitch is not the main reason why the performance of speaker recognition is high.

The first reason why voice test is high is related to the volume which is the speaking habits of a person. We find that the amplitude of twin B is higher than twin A from Figure 4.4 and Figure 4.5. This means the volume of twin B is usually louder than twin A.

The second reason why the recognition precision is high is related to timbre, which is mainly affected by waveform. Although some waveforms as shown in Appendix I look similar such as the No. 3 and No. 7, most of them are different like No. 2, No 12, No. 27, No. 41 and etc. The timbre is affected by personal speech habits. The speech habits of twins are not exactly as same as each other. From the research of Nolan and Oh (1996), the learning also is the main reason that leads to the difference of voice between identical twins. This helps speaker recognition in twin identification.

The limitation of this test is that voice samples of the twins are recorded in the indoor environment which means there is nearly no noises. However, in the real world, noises always exist. This may affect the precession of the voice based twin recognition.

The lip movement recognition shows the second good performance among these four methods. Lip movement recognition is behavioral characteristics based recognition. The precision of twin recognition reaches 80%. Compared with the face and ear recognition, it is much better. In fact, the accuracy of twin recognition by average human is around 80% (Behravan & Faez, 2013).

The reason why lip movement recognition has good performance is that it adopts habit of human to conduct the recognition. Different people may have different personal habits, which are formed in the long-time growing-up environment. Even for the identical twins the habit are not identical.

Although, many article has investigated about twin recognition in biometrics. Most of them compare the performance between the same kind of biometric methods. It lakes comparisons between different kind of biometric methods. This thesis compares different kind of biometric methods in twin recognition. It can help people to choose proper biometric methods when they want to recognize twins.

5.2 Limitations

The limitations of this thesis mainly have two aspects: only one pair of twins and four methods. The data of the twins are just come from one pair of twins. Therefore, the method which has good performance on this pair of twins may probably not be effective on other twins.

The second limitation of this research is the kind of methods. Because of the time and facility, only face recognition system, speaker recognition system, lip movement system and ear recognition system are tested. However, we think other methods are still indispensable to be considered in future.

Chapter 6 Conclusion and Future Work

The aim of this thesis is to identify the differences between twins by using biometrics. The results show it is possible to completely recognize twins by using biometrics. Although face recognition and ear recognition do not show excellent performance in twin recognition, the results of other two methods are good enough. After test all the methods, the speaker recognition has shown the outstanding correct rate and the accuracy is perfect which is much higher than other methods. The reason is that speaker recognition is a kind of biometrics which contain both physiological and behavioral characteristics, this makes it more informative than other methods which only have physiological or behavioral characteristics.

6.1 Conclusion

To conclusion this thesis, Chapter 1 introduces the background and the motivation of this research work. Chapter 2 reviews many approaches of biometrics like face recognition, ear recognition and others. Chapter 3 shows the methodology of this research thesis. In this chapter, we introduce the algorithms which are used in this thesis such as the Principal Component Analysis (PCA), Mel-frequency cepstral coefficient (MFCC) and Vector Quantization Linde–Buzo–Gray (VQLBG). Also the samples of this research are introduced in this chapter including fifty voice samples of each of the twins. The ear and face photos of the twins are shown in this chapter.

In Chapter 4, the results are evaluated which show the speaker recognition has the best performance in twin recognition. On the other hand, the ear recognition is not so good performance (with only 53%). The correct rates of other two methods for twin recognition, namely, face recognition and lip movement recognition are 58% and 76% respectively.

In Chapter 5, we discussed the results of twin recognition by using biometrics. The accuracy, precision, recall and F-measure of speaker recognition and lip movement recognition are adopted to support our discussions. We would like to emphasize that speaker recognition is the best method for recognizing the twins among these methods again.

6.2 Limitations and Future work

One of the limitations of this face based twin recognition is that it does not take the environment changes like the illumination factor into considerations. The precision of face based twin recognition will be much better if we control the illumination properly. We will figure out this issue in future.

Regarding to future work, firstly data of more twins need to be collected. Specially, the twins in different age and gender groups need to be tested. Although there are some databases of the twins, most of them are still physiological characteristics like face, ear and finger print. More databases related to behavioral characteristics like the videos of walking posture need to be selected. Especially the database which contains multiple biometric characteristics of twins needs to be setup.

Secondly, other biometrics can be considered to recognize twins. Although the results of the speaker recognition are perfect in twin recognition, other methods still need to be test in this subject area. In this thesis lip movement recognition shows good performance in twin recognition, that means other behavioral characteristics such as handwriting and walking posture can be considered to be applied in twin recognition in future.

References

Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 28*(12), 2037-2041. doi:Doi 10.1109/Tpami.2006.244

Aleksic, P. S. (2009). Lip Movement Recognition. *Encyclopedia of Biometrics*, 904-908.

Bartlett, M. S., Movellan, J. R., & Sejnowski, T. J. (2002). Face recognition by independent component analysis. *Neural Networks, IEEE Transactions on*, *13*(6), 1450-1464.

Behravan, H., & Faez, K. (2013, May). Introducing a new multimodal database from twins' biometric traits. In *Electrical Engineering (ICEE), 2013 21st Iranian Conference on* (pp. 1-6).

Bishop, C. M. (2006). Pattern Recognition. *Machine Learning*.

Biswas, S., Bowyer, K. W., & Flynn, P. J. (2011, November). A study of face recognition of identical twins by humans. In *Information Forensics and Security (WIFS), 2011 IEEE International Workshop on* (pp. 1-6).

Blanz, V., & Vetter, T. (2003). Face recognition based on fitting a 3D morphable model. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *25*(9), 1063-1074.

Bruce, V., & Young, A. (1986). Understanding Face Recognition. *British Journal of Psychology, 77*, 305-327.

Burton, D. K. (1987). Text-dependent speaker verification using vector quantization source coding. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, *35*(2), 133-143.

Chan, M. T., Zhang, Y., & Huang, T. S. (1998, December). Real-time lip tracking and bimodal continuous speech recognition. In *Multimedia Signal Processing, 1998 IEEE Second Workshop on* (pp. 65-70). IEEE.

Chang, K., Bowyer, K. W., Sarkar, S., & Victor, B. (2003). Comparison and combination of ear and face images in appearance-based biometrics. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *25*(9), 1160-1165.

Chellappa, R., Wilson, C.L., & Sirohey, S. (1995). Human and Machine Recognition of Faces - A Survey. *83*(5), 705-740. doi:Doi 10.1109/5.381842

Chen, H., & Bhanu, B. (2007). Human ear recognition in 3D. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *29*(4), 718-737.

Choi, C. (1999, August). Age change for predicting future faces. In *Fuzzy Systems Conference Proceedings, 1999. FUZZ-IEEE'99. 1999 IEEE International* (Vol. 3, pp. 1603-1608).

Choraś, M. (2004). Ear biometrics based on geometrical method of feature extraction. In *Articulated Motion and Deformable Objects* (pp. 51-61).

Cootes, T. F., Wheeler, G. V., Walker, K. N., & Taylor, C. J. (2002). View-based active appearance models. *Image and vision computing*, *20*(9), 657-664.

Delac, K., Grgic, M., & Grgic, S. (2005). Generalization abilities of appearance-based subspace face recognition algorithms. *the 12th International Workshop on Systems, Signals & Image Processing*, 271-274.

Delac, K., Grgic, M., & Grgic, S. (2005). Independent comparative study of PCA, ICA, and LDA on the FERET data set. *International Journal of Imaging Systems and Technology*, *15*(5), 252-260.

Delac, K., Grgic, M., & Grgic, S. (2005). Statistics in face recognition: analyzing probability distributions of PCA, ICA and LDA performance results. In *Image and Signal Processing and Analysis, 2005. ISPA 2005. Proceedings of the 4th International Symposium on* (pp. 289-294).

Draper, B.A., Baek, K., Bartlett, M.S., & Beveridge, J. R. (2003). Recognizing Faces with PCA and ICA, *Computer Vision and Image Understanding (Special Issue on Face Recognition), Vols 91 (1-2),* 115-137.

Feng, G. C., Yuen, P. C., & Dai, D. Q. (2000). Human face recognition using PCA on wavelet subband. *Journal of Electronic Imaging*, *9*(2), 226-233.

Ganchev, T., Fakotakis, N., & Kokkinakis, G. (2005). Comparative evaluation of various MFCC implementations on the speaker verification task.

In *Proceedings of the SPECOM* (Vol. 1, pp. 191-194).

Gauthier, I., & Logothetis, N. K. (2000). Is face recognition not so unique after all?. *Cognitive Neuropsychology*, *17*(1-3), 125-142.

Geng, X., Zhou, Z. H., & Smith-Miles, K. (2007). Automatic age estimation based on facial aging patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *29*(12), 2234-2240.

Gottumukkal, R., & Asari, V. K. (2004). An improved face recognition technique based on modular PCA approach. *Pattern Recognition Letters*,*25*(4), 429-436.

Guo, G., Li, S. Z., & Chan, K. (2000). Face recognition by support vector machines. In *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on* (pp. 196-201).

Hanmandlu, M., Gupta, D., & Vasikarla, S. (2013). Face recognition using Elastic bunch graph matching. In *Applied Imagery Pattern Recognition Workshop (AIPR): Sensing for Control and Augmentation, 2013 IEEE* (pp. 1-7).

Heisele, B., Ho, P., & Poggio, T. (2001). Face recognition with support vector machines: Global versus component-based approach. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on* (Vol. 2, pp. 688-694).

Horng, W. B., Lee, C. P., & Chen, C. W. (2001). Classification of age groups based on facial features. *Tamkang Journal of Science and Engineering*, *4*(3), 183-192.

Hussein, H. K. (2002). Towards realistic facial modeling and re-rendering of human skin aging animation. In *Shape Modeling International, 2002. Proceedings* (pp. 205-212).

Huynh-Thu, Q., Meguro, M., & Kaneko, M. (2002). Skin-color extraction in images with complex background and varying illumination. In *Applications of Computer Vision, 2002.(WACV 2002). Proceedings. Sixth IEEE Workshop on* (pp. 280-285).

Jain, A. K. (2007). Technology: Biometric recognition. *Nature, 449*(7158), 38-40. doi:http://dx.doi.org/10.1038/449038a

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

James, C. J., & Hesse, C. W. (2004). Independent component analysis for biomedical signals. *Physiological measurement*, *26*(1), R15.

Jian-wei, Z., Shui-fa, S., Xiao-li, L., & Bang-jun, L. (2009, August). Pitch in speaker recognition. In *Hybrid Intelligent Systems, 2009. HIS'09. Ninth International Conference on* (Vol. 1, pp. 33-36).

Jonsson, K., Matas, J., Kittler, J., & Li, Y. P. (2000). Learning support vectors for face verification and recognition. In *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*(pp. 208-213).

Kadyrov, A., & Petrou, M. (2001). The trace transform and its applications.*Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *23*(8), 811-828.

Karhunen, J., Oja, E., Wang, L., Vigario, R., & Joutsensalo, J. (1997). A class of neural networks for independent component analysis. *Neural Networks, IEEE Transactions on*, *8*(3), 486-504.

Klare, B., Paulino, A. A., & Jain, A. K. (2011, October). Analysis of facial features in identical twins. In *Biometrics (IJCB), 2011 International Joint Conference on* (pp. 1-8).

Kim, T. K., Kittler, J., & Cipolla, R. (2007). Discriminative learning and recognition of image set classes using canonical correlations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *29*(6), 1005-1018.

Kovac, J., Peer, P., & Solina, F. (2003). Human skin colour clustering for face detection. *IEEE Region 8 Eurocon 2003, Vol B,* 144-148.

Kshirsagar, V. P., Baviskar, M. R., & Gaikwad, M. E. (2011). Face recognition using Eigenfaces. In *Computer Research and Development (ICCRD), 2011 3rd International Conference on* (Vol. 2, pp. 302-306).

Kwon, Y. H., & da Vitoria Lobo, N. (1994). Age classification from facial images. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on* (pp. 762-767).

Kurniawan, F., Shafry, M., & Rahim, M. (2012, March). A review on 2D ear recognition. In *Signal Processing and its Applications (CSPA), 2012 IEEE 8th International Colloquium on* (pp. 204-209).

Lammi, H. K. (2004). Ear biometrics. *Lappeenranta University of Technology*.

Lanitis, A., Draganova, C., & Christodoulou, C. (2004). Comparing different classifiers for automatic age estimation. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, *34*(1), 621-628.

Lawrence, S., Giles, C. L., Tsoi, A. C., & Back, A. D. (1997). Face recognition: A convolutional neural-network approach. *Neural Networks, IEEE Transactions on*, *8*(1), 98-113.

Le, T. H. N., Luu, K., Seshadri, K., & Savvides, M. (2012, September). A facial aging approach to identification of identical twins. In *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*(pp. 91-98).

Leta, F. R., Conci, A., Pamplona, D., & Itanguy, I. (1996). Manipulating facial appearance through age parameters. In *Proc. Ninth Brazilian Symp. Computer Graphics and Image Processing* (pp. 167-172).

Li, S. Z., & Lu, J. (1999). Face recognition using the nearest feature line method. *Neural Networks, IEEE Transactions on*, *10*(2), 439-443.

Li, Y. Z., Ming, F., Yang, J. Y., & Pan, R. L. (2006, December). An Efficient Method of Nonlinear Feature Extraction Based on SVM. In *Control, Automation, Robotics and Vision, 2006. ICARCV'06. 9th International Conference on* (pp. 1-6).

Li, Y., Yuan, W., Sang, H., & Li, X. (2013, May). Combination recognition of face and ear based on two-dimensional fisher linear discriminant. In *Software Engineering and Service Science (ICSESS), 2013 4th IEEE International Conference on* (pp. 922-925).

Linde, Y., Buzo, A., & Gray, R. M. (1980). An algorithm for vector quantizer design. *Communications, IEEE Transactions on*, *28*(1), 84-95.

Liu, C., & Wechsler, H. (1999). Comparative assessment of independent component analysis (ICA) for face recognition. In *International conference on audio and* *video based biometric person authentication*.

Liu, C., & Wechsler, H. (2000). Evolutionary pursuit and its application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *22*(6), 570-582.

Mahalingam, G., & Ricanek, K. (2013, December). Investigating the effects of gender and age group based differences in identical twins. In *Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), 2013 Fourth National Conference on* (pp. 1-4).

Manjunath, B., Chellappa, R., & Malsburg, C.V. (1992). A feature based approach to face recognition, *Computer Vision and Pattern Recognition*. 373-377

Martínez, A. M., & Kak, A. C. (2001). Pca versus lda. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 23*(2), 228-233.

Moghaddam, B., Lee, J., Pfister, H., & Machiraju, R. (2003). Model-based 3D face capture with shape-from-silhouettes. In *Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on* (pp. 20-27).

Moghaddam, B., & Pentland, A. (1997). Probabilistic visual learning for object representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *19*(7), 696-710.

Mozaffari, S., & Behravan, H. (2011, May). Twins facial similarity impact on conventional face recognition systems. In *Electrical Engineering (ICEE), 2011 19th Iranian Conference on* (pp. 1-6).

Muda, L., Begam, M., & Elamvazuthi, I. (2010). Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. *arXiv preprint arXiv:1003.4083*.

Nakano, M., Yasukata, F., & Fukumi, M. (2004, January). Age classification from face images focusing on edge information. In *Knowledge-Based Intelligent Information and Engineering Systems* (pp. 898-904). Springer Berlin Heidelberg.

Nefian, A. V., & Hayes, M. H. (1999, March). An embedded HMM-based approach for face detection and recognition. In *Acoustics, Speech, and Signal*

Processing, 1999. Proceedings., 1999 IEEE International Conference on (Vol. 6, pp. 3553-3556).

Nolan, F., & Oh, T. (1996). Identical twins, different voices. *International Journal of Speech Language and the Law*, *3*(1), 39-49.

Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *24*(7), 971-987.

Ollikainen, M., Smith, K. R., Joo, E. J. H., Ng, H. K., Andronikos, R., Novakovic, B., ... & Craig, J. M. (2010). DNA methylation analysis of multiple tissues from newborn twins reveals both genetic and intrauterine components to variation in the human neonatal epigenome. *Human molecular genetics*, *19*(21), 4176-4188.

Osuna, E., Freund, R., & Girosi, F. (1997). Training support vector machines: an application to face detection. In *Computer vision and pattern recognition, 1997. Proceedings., 1997 IEEE computer society conference on*(pp. 130-136).

Pentland, A., Moghaddam, B., & Starner, T. (1994). View-based and modular eigenspaces for face recognition. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on* (pp. 84-91).

Phillips, P. J., Flynn, P. J., Bowyer, K. W., Bruegge, R. W. V., Grother, P. J., Quinn, G. W., & Pruitt, M. (2011). Distinguishing identical twins by face recognition. In *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on* (pp. 185-192).

Phillips, P. J., Moon, H., & Rizvi, S. (1997). The FERET evaluation methodology for face-recognition algorithms. *Computer Vision and Pattern Recognition 97,* 137-143

Qi, Y., & Hunt, B. R. (1994). Signature verification using global and grid features. *Pattern Recognition*, *27*(12), 1621-1629.

Rowley, H. A., Baluja, S., & Kanade, T. (1998). Neural network-based face detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions* *on,20*(1), 23-38.

Shailaja, D., & Gupta, P. (2006, December). A simple geometric approach for ear recognition. In *Information Technology, 2006. ICIT'06. 9th International Conference on* (pp. 164-167).

Srinivas, N., Aggarwal, G., Flynn, P. J., & Vorder Bruegge, R. W. (2012). Analysis of facial marks to distinguish between identical twins. *Information Forensics and Security, IEEE Transactions on*, *7*(5), 1536-1550.

Srisuk, S., Petrou, M., Kurutach, W., & Kadyrov, A. (2003). Face authentication using the trace transform. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on* (Vol. 1, pp. I-305).

Stewart, L., Evans, N., Bexon, K. J., van der Meer, D. J., & Williams, G. A. (2015). Differentiating between monozygotic twins through DNA methylation-specific high-resolution melt curve analysis. *Analytical biochemistry*, *476*, 36-39.

Suo, J., Min, F., Zhu, S., Shan, S., & Chen, X. (2007, June). A multi-resolution dynamic model for face aging simulation. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on* (pp. 1-8).

Sun, Z., Paulino, A. A., Feng, J., Chai, Z., Tan, T., & Jain, A. K. (2010, April). A study of multibiometric traits of identical twins. In *SPIE Defense, Security, and Sensing* (pp. 76670T-76670T). International Society for Optics and Photonics.

Sung, K. K., & Poggio, T. (1998). Example-based learning for view-based human face detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *20*(1), 39-51.

Tanaka, J.W., & Farah, M.J. (1993). Parts and Wholes in Face Recognition. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 46*(2), 225-245.

Tiwari, V. (2010). MFCC and its applications in speaker recognition.*International Journal on Emerging Technologies*, *1*(1), 19-22.

Turk, M. A., & Pentland, A. P. (1991). Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on* (pp. 586-591).

Uřičář, M., Franc, V., & Hlaváč, V. (2012). Detector of facial landmarks learned by the structured output SVM. *VISAPP*, *12*, 547-556.

Wiskott, L., Fellous, J. M., Kuiger, N., & Von Der Malsburg, C. (1997). Face recognition by elastic bunch graph matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 19*(7), 775-779.

Viola, P., & Jones, M. (2001). Robust real-time face detection. *Eighth IEEE International Conference on Computer Vision, Vol II*, 747-747.

Yang, M.H. (2002). Kernel Eigenfaces vs. Kernel Fisherfaces: Face recognition using Kernel methods. *Fifth IEEE International Conference on Automatic Face and Gesture Recognition,* 215-220.

Yang, J., Zhang, D., Frangi, A. F., & Yang, J. Y. (2004). Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *26*(1), 131-137.

Vijayan, V., Bowyer, K. W., Flynn, P. J., Huang, D., Chen, L., Hansen, M., ... & Kakadiaris, I. A. (2011, October). Twins 3D face recognition challenge. In*Biometrics (IJCB), 2011 International Joint Conference on* (pp. 1-7).

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (Vol. 1, pp. I-511).

Viola, P., & Jones, M. (2001). Robust real-time face detection. *Eighth IEEE International Conference on Computer Vision, Vol II*, 747-747.

Yu, H. & Yang, J. (2001). A direct LDA algorithm for high dimensional data with application to face recognition. *Pattern Recognition,* vol. 34. 2067-2070

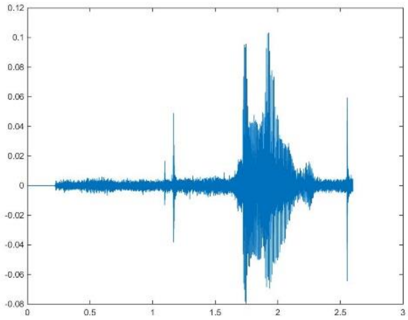
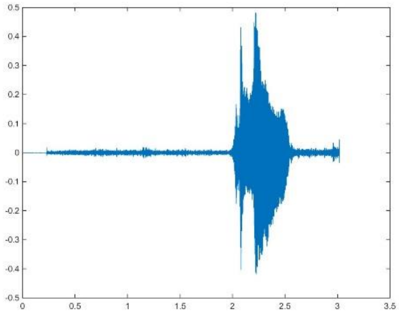
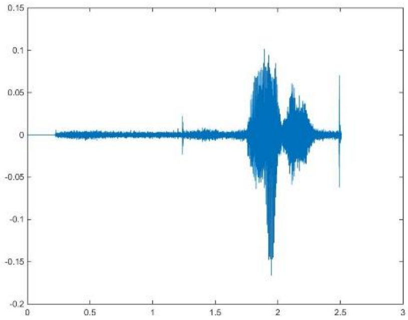
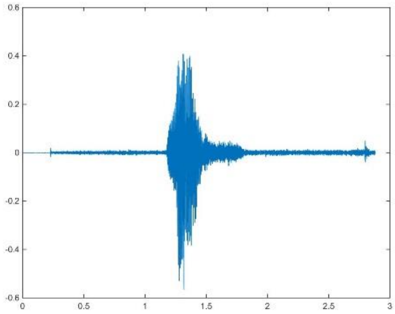
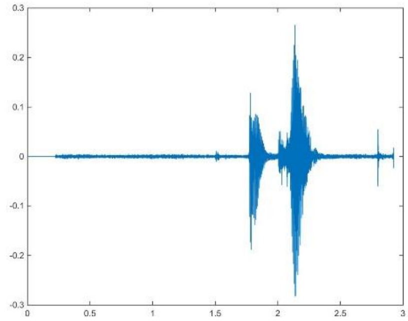
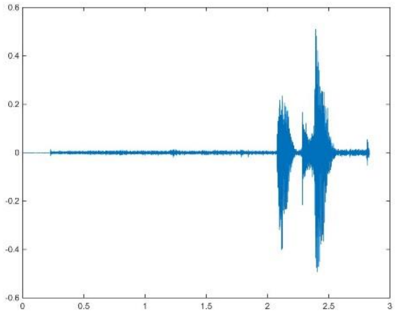
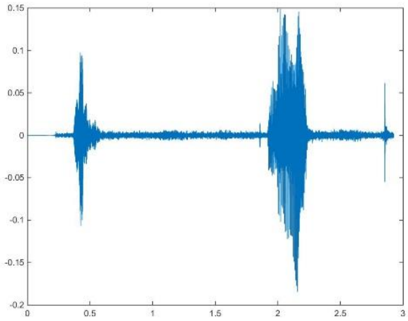
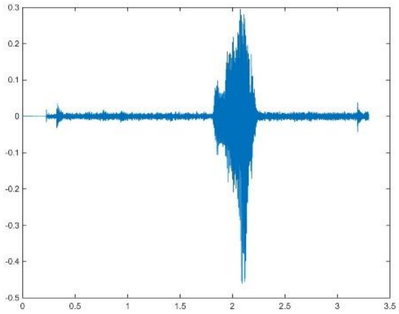
Yuan, L., Mu, Z. C., Zhang, Y., & Liu, K. (2006, August). Ear recognition using improved non-negative matrix factorization. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on* (Vol. 4, pp. 501-504).

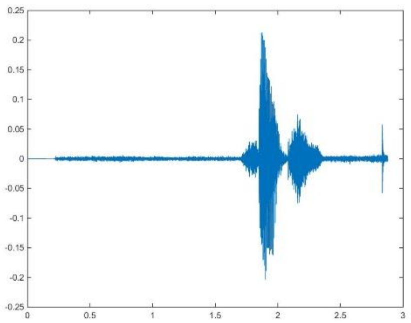
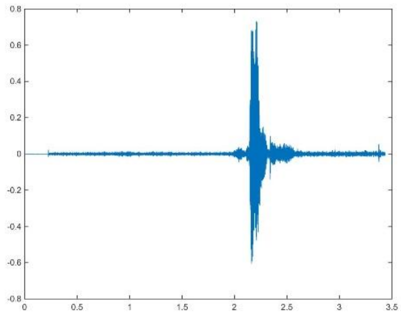
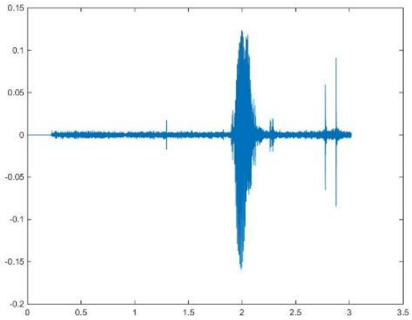
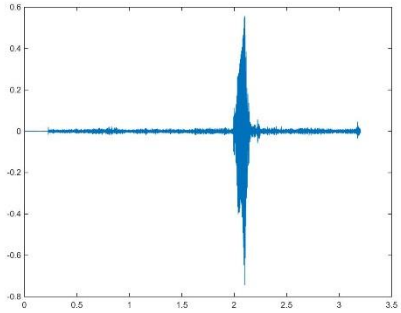
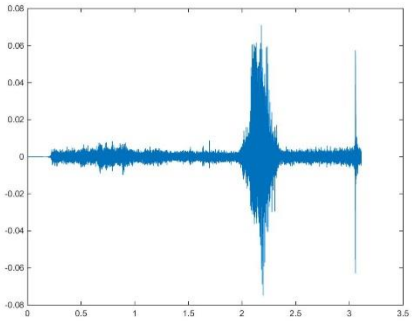
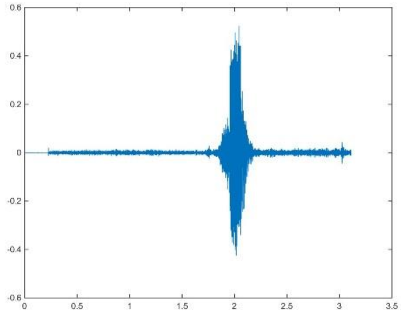
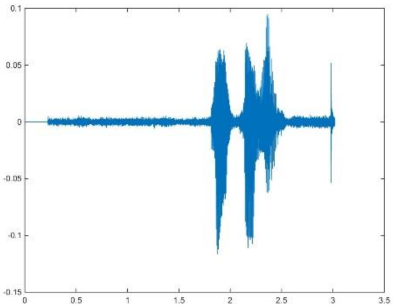
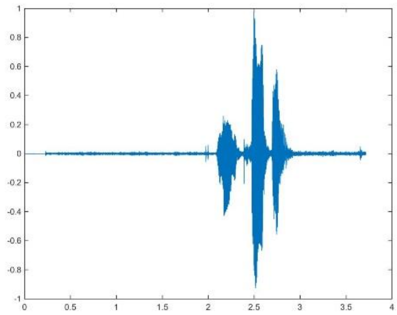
Zhao, W., Chellappa, R., & Krishnaswamy, A. (1998). Discriminant analysis of principal components for face recognition. *Automatic Face and Gesture Recognition - Third IEEE International Conference*, 336-341

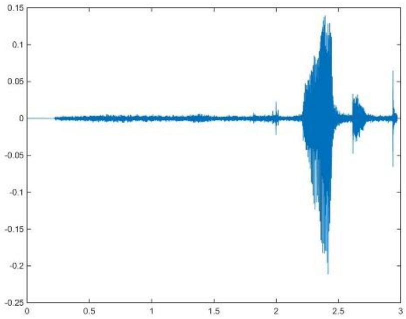
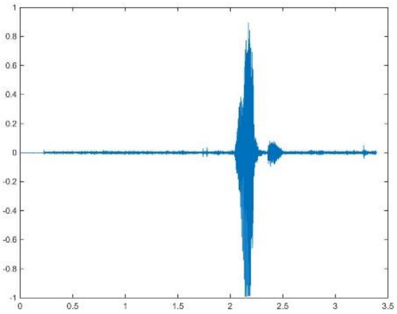
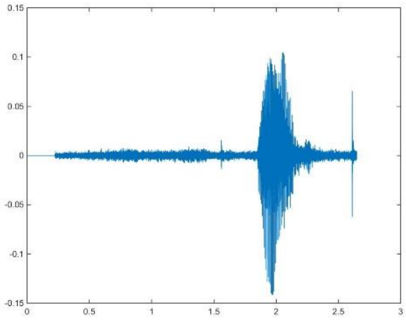
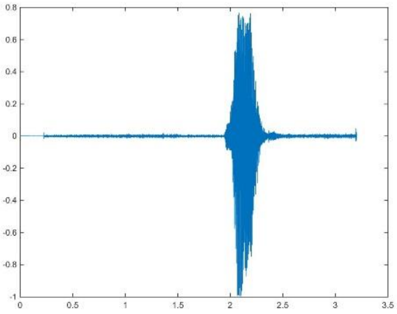
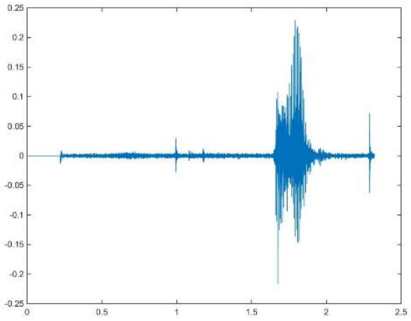
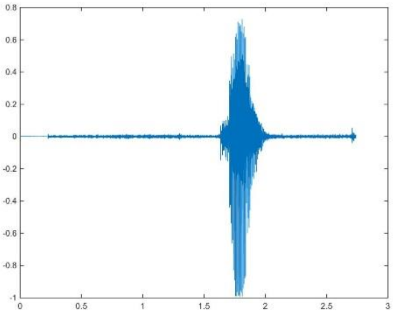
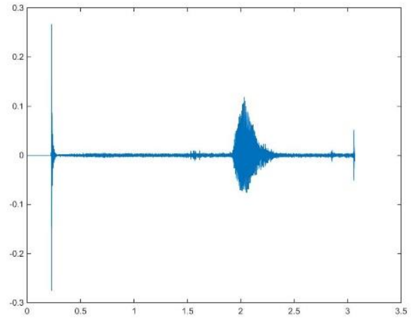
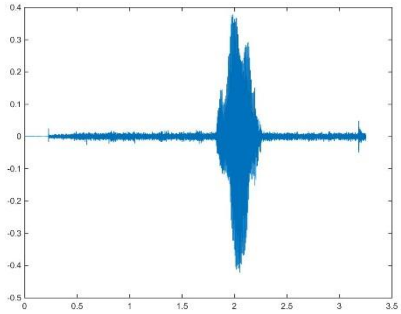
Zhao, Z. Q., Huang, D. S., & Sun, B. Y. (2004). Human face recognition based on multi-features using neural networks committee. *Pattern Recognition Letters*, *25*(12), 1351-1358.

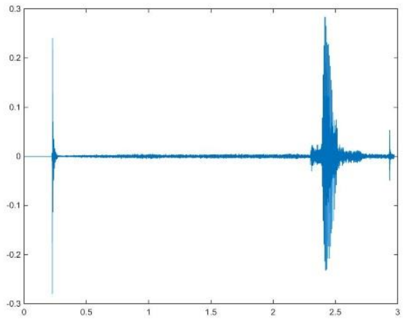
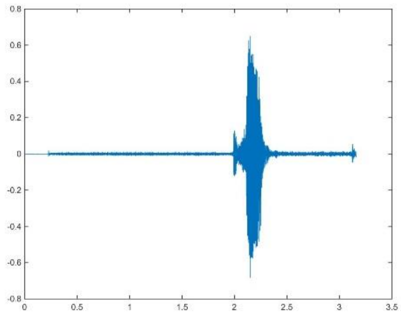
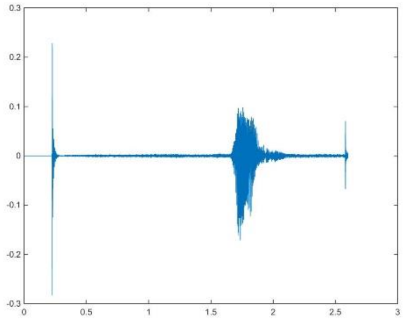
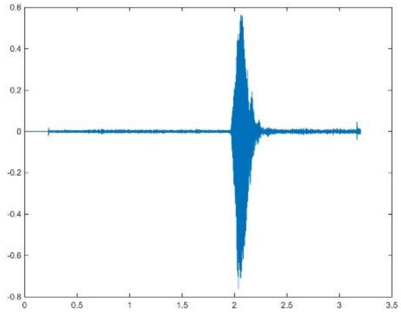
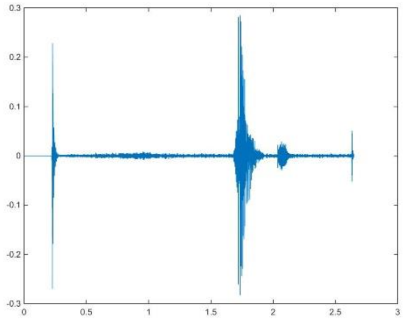
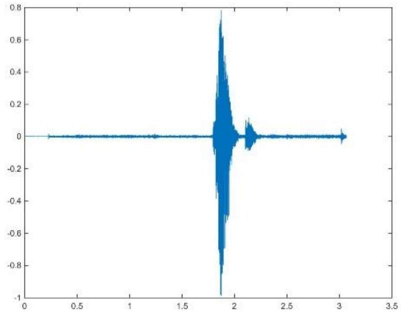
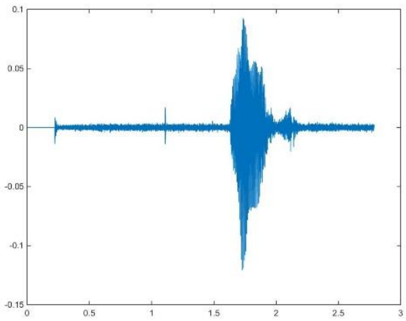
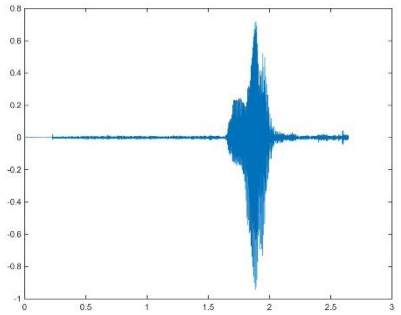
Appendix I

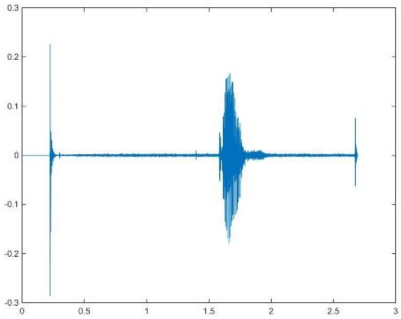
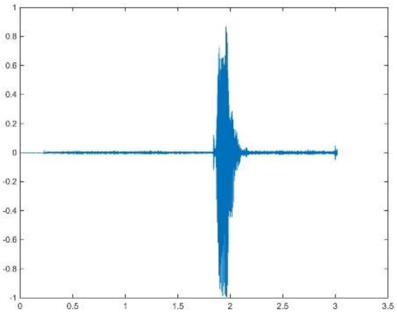
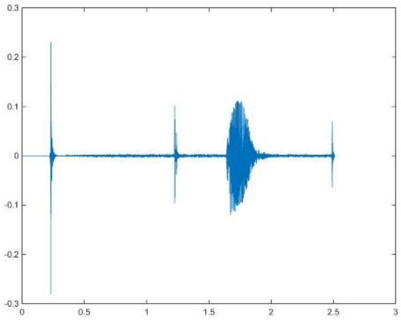
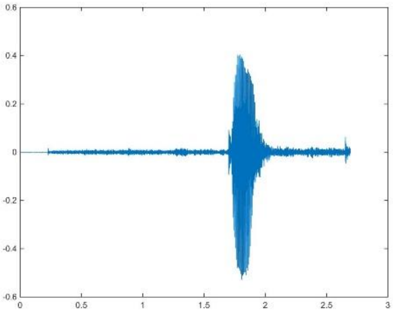
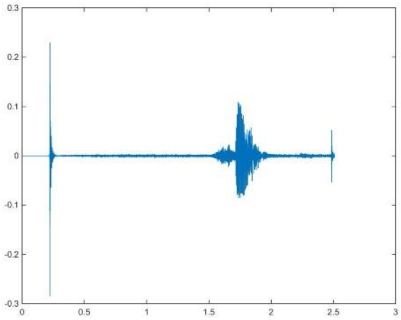
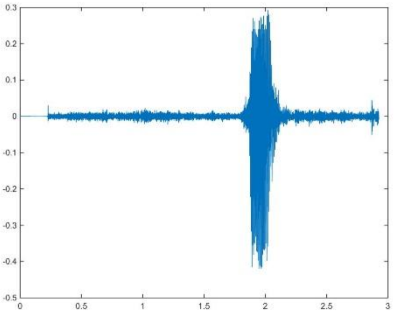
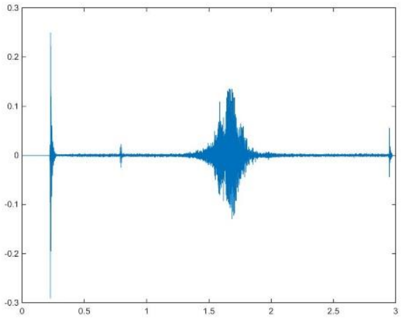
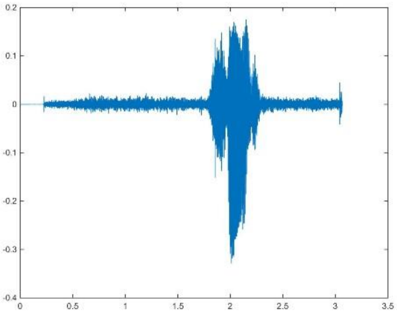
Waveforms of the twin pronunciation

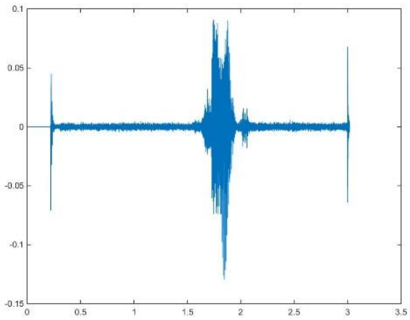
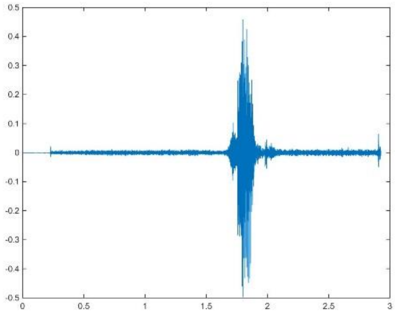
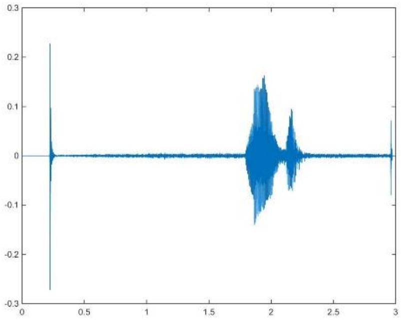
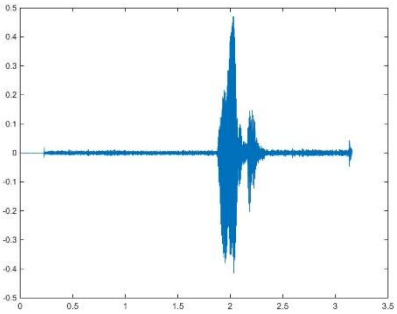
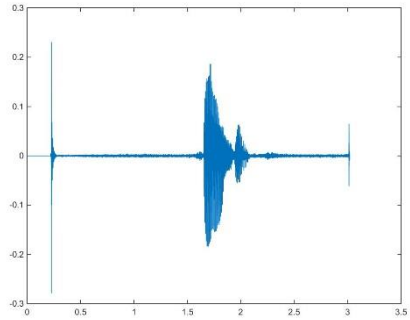
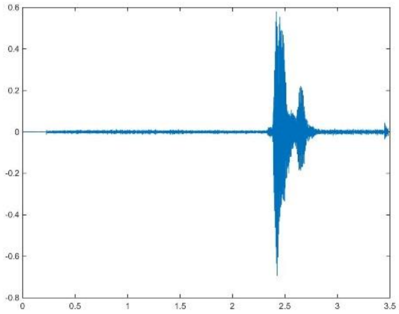
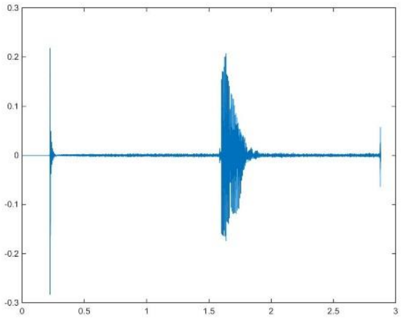
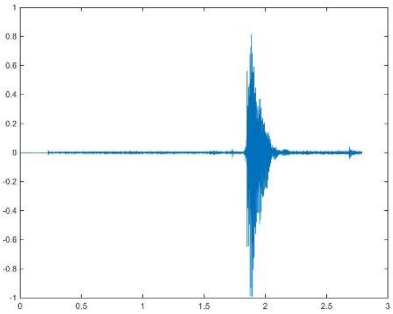
Words	Twin A's waveforms	Twin B's waveforms
1 (hello)		
2 (yes)		
3 (O.K)		
4 (no)		

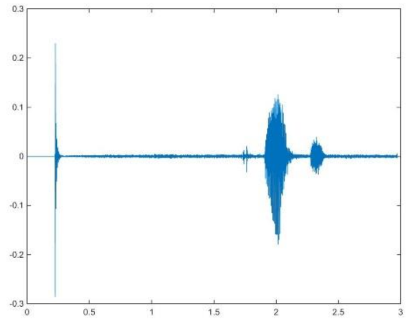
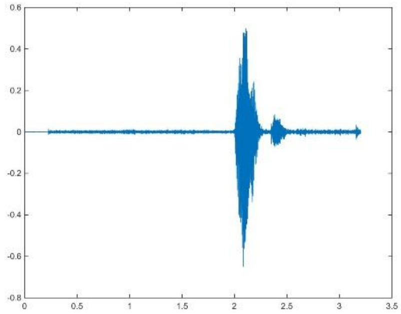
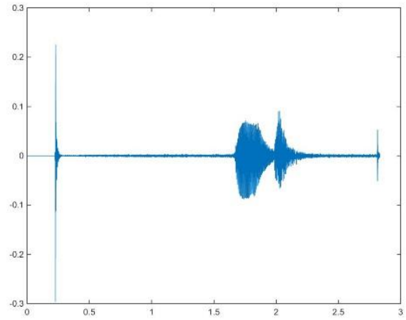
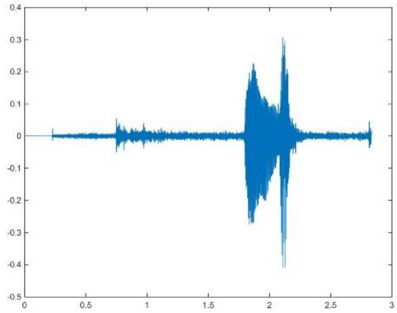
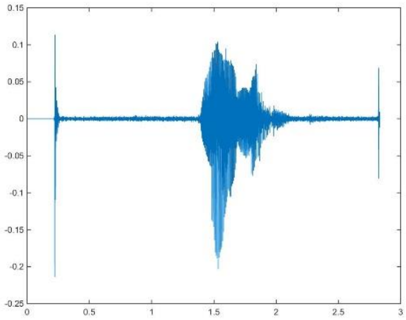
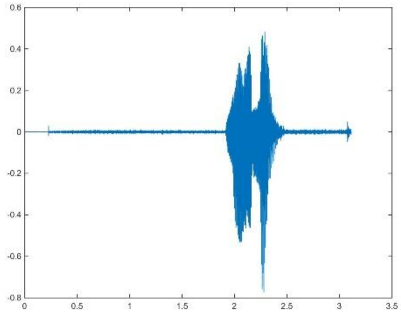
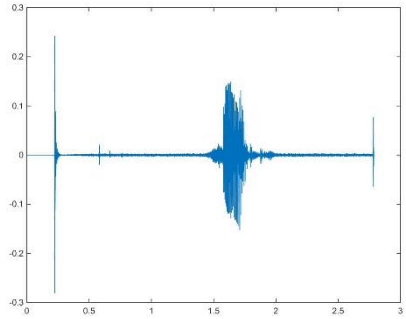
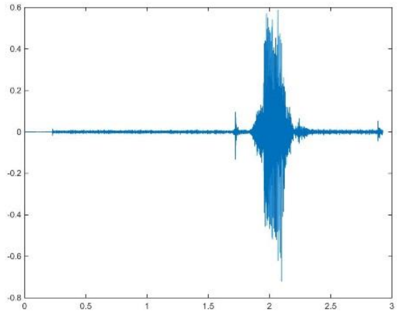
Words	Twin A’s waveforms	Twin B’s waveforms
5 (thanks)		
6 (good)		
7 (hi)		
8 (no problem)		

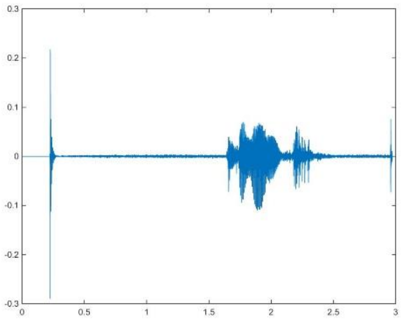
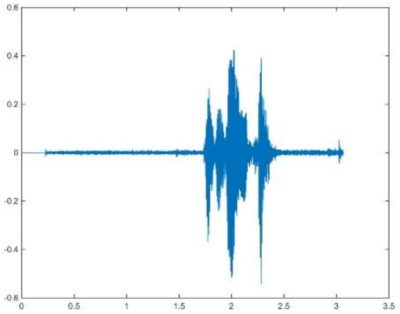
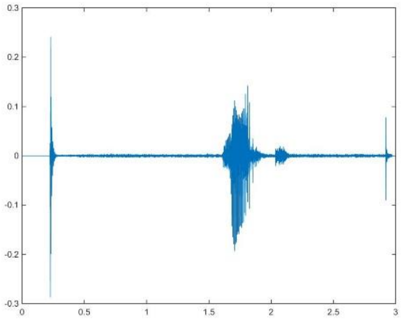
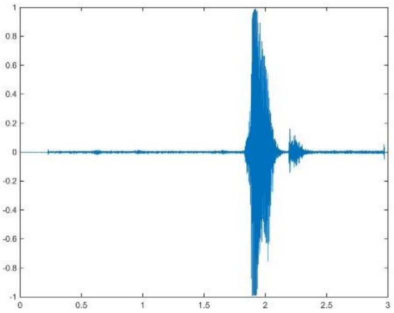
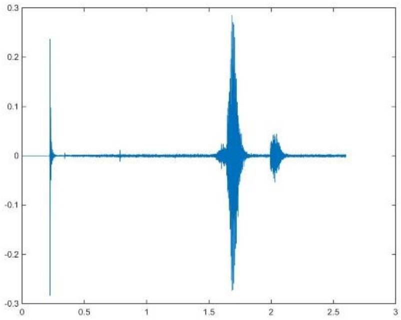
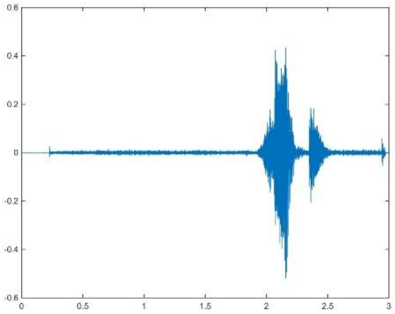
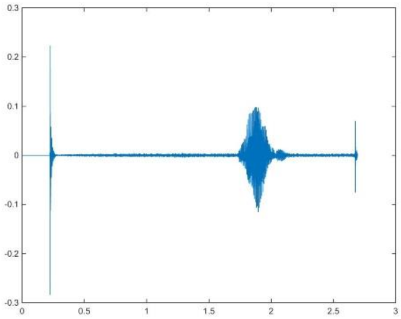
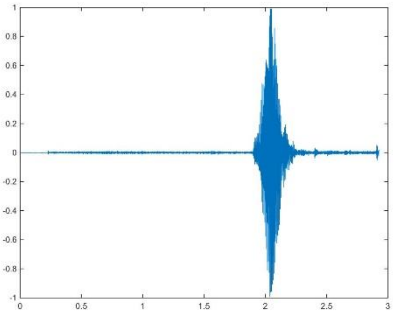
Words	Twin A’s waveforms	Twin B’s waveforms
9 (what)		
10 (why)		
11 (I)		
12 (you)		

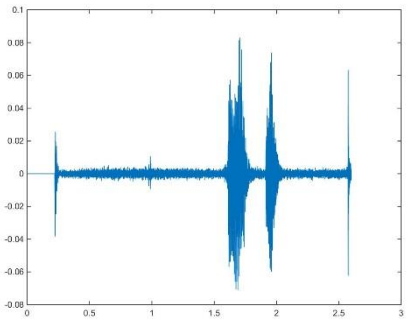
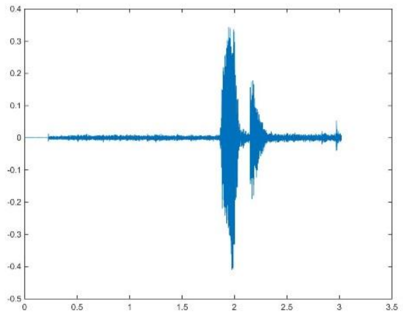
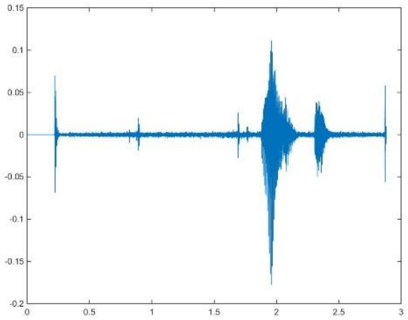
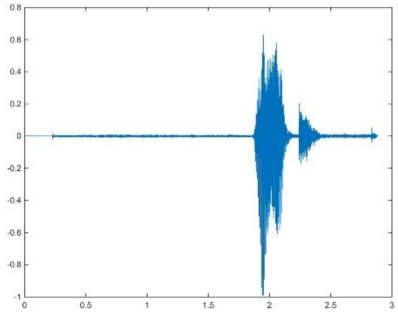
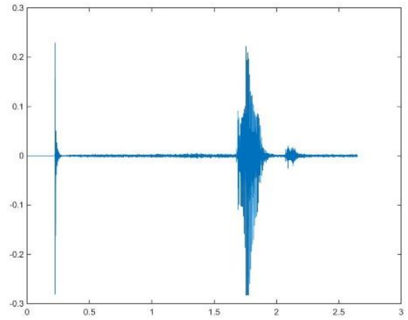
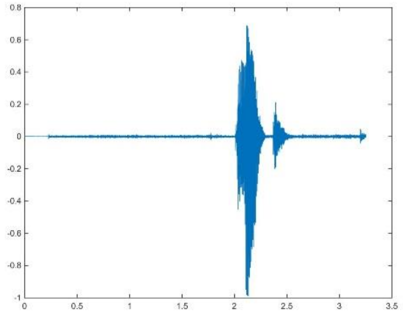
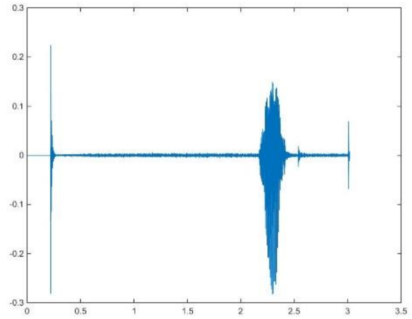
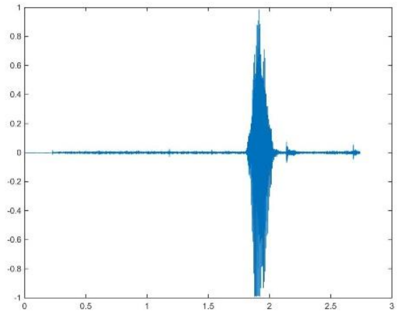
Words	Twin A’s waveforms	Twin B’s waveforms
13 (come)		
14 (in)		
15 (out)		
16 (more)		

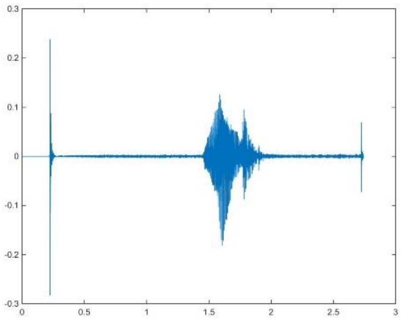
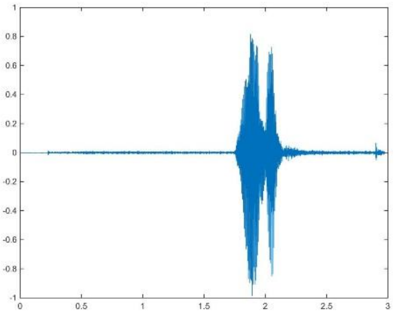
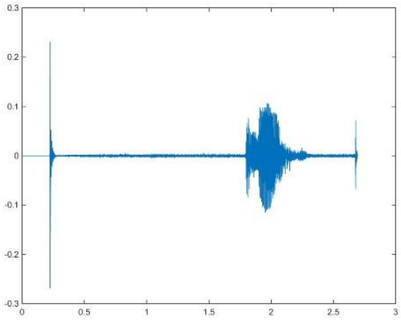
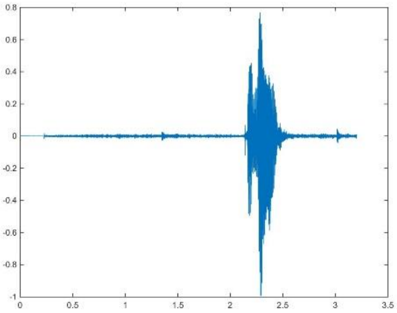
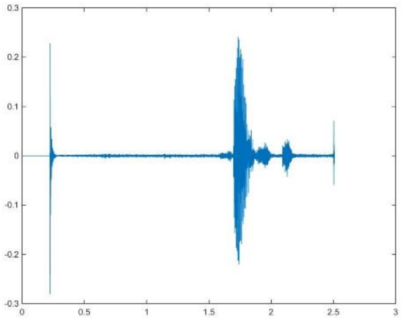
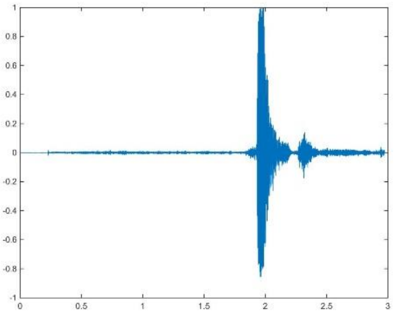
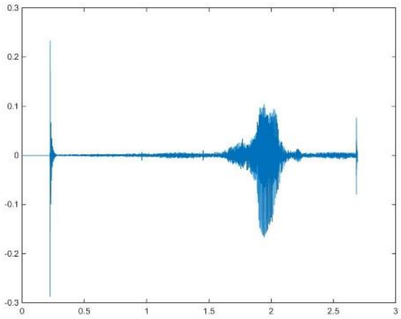
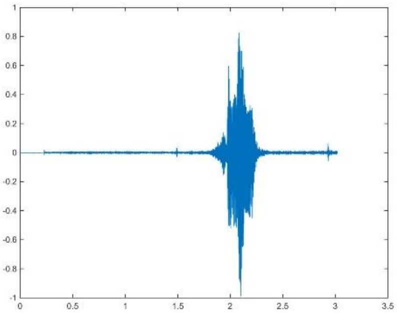
Words	Twin A’s waveforms	Twin B’s waveforms
17 (go)		
18 (do)		
19 (he)		
20 (She)		

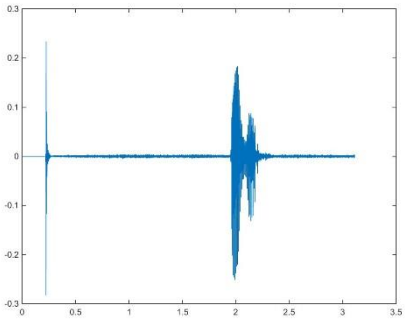
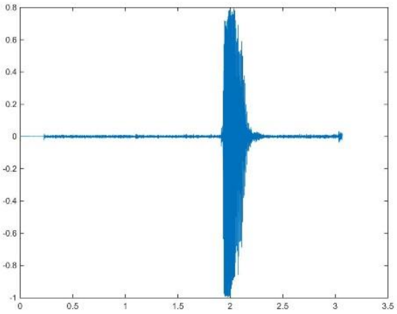
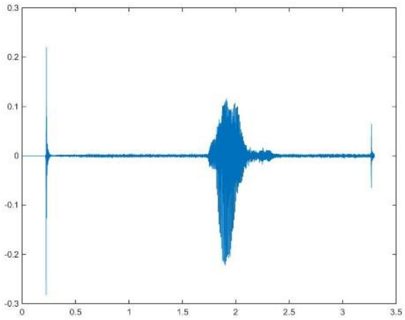
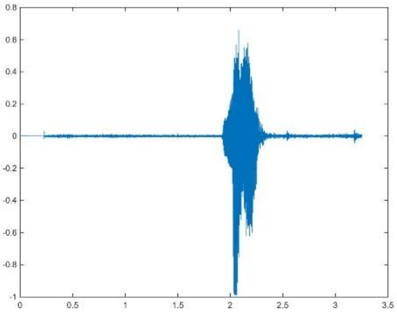
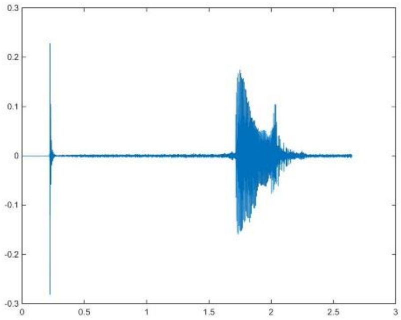
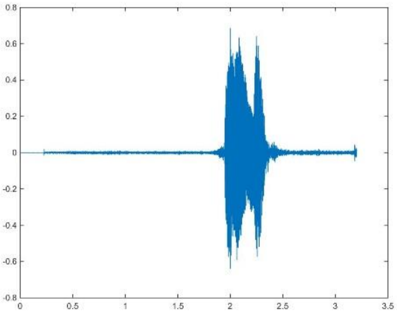
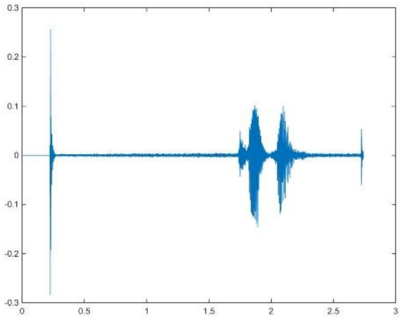
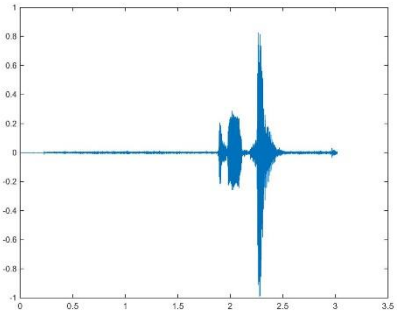
Words	Twin A’s waveforms	Twin B’s waveforms
21 (had)		
22 (mother)		
23 (father)		
24 (on)		

Words	Twin A’s waveforms	Twin B’s waveforms
25 (wait)		
26 (easy)		
27 (normal)		
28 (hard)		

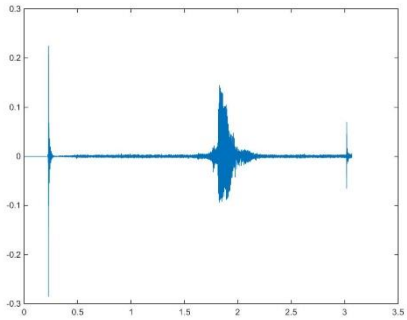
Words	Twin A’s waveforms	Twin B’s waveforms
29 (example)		
30 (like)		
31 (hate)		
32 (run)		

Words	Twin A’s waveforms	Twin B’s waveforms
33 (eat)		
34 (white)		
35 (black)		
36 (red)		

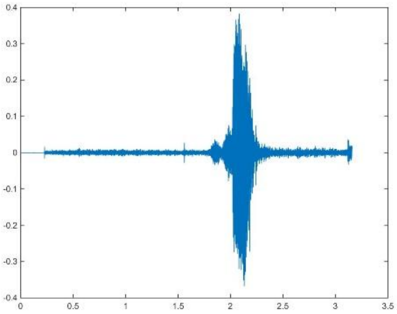
Words	Twin A’s waveforms	Twin B’s waveforms
37 (yellow)		
38 (blue)		
39 (fast)		
40 (slow)		

Words	Twin A's waveforms	Twin B's waveforms
41 (ball)		
42 (name)		
43 (final)		
44 (people)		

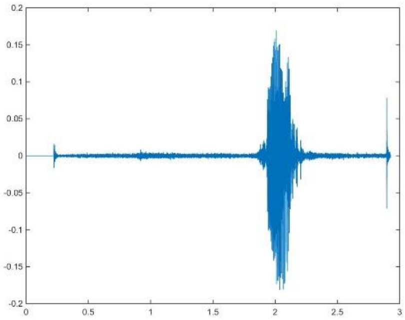
<table>
<tr><th>Words</th><th>Twin A’s waveforms</th><th>Twin B’s waveforms</th></tr>
<tr><td>45
(sea)</td><td>



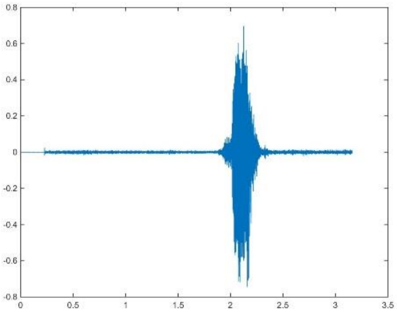
</td><td>



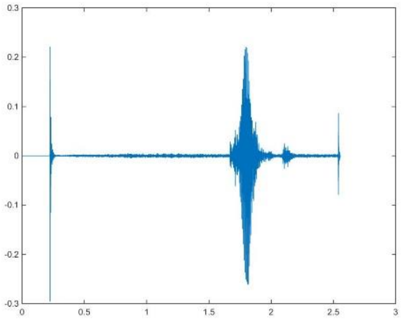
</td></tr>
<tr><td>46
(hill)</td><td>



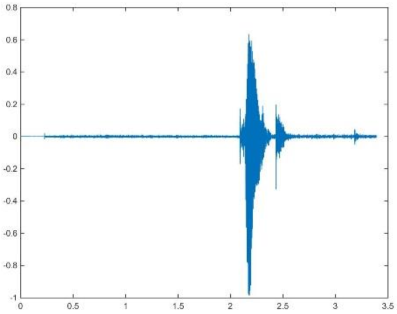
</td><td>



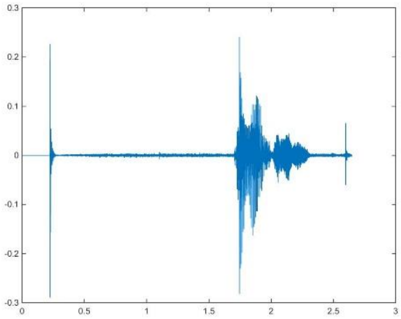
</td></tr>
<tr><td>47
(drink)</td><td>



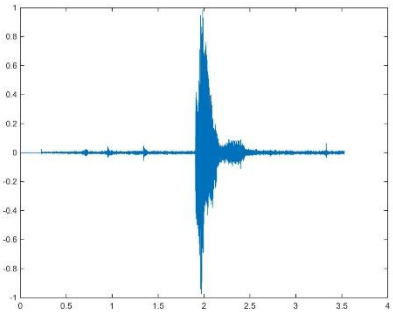
</td><td>



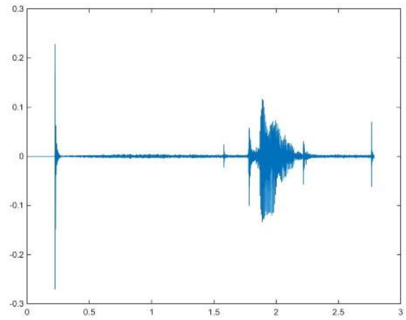
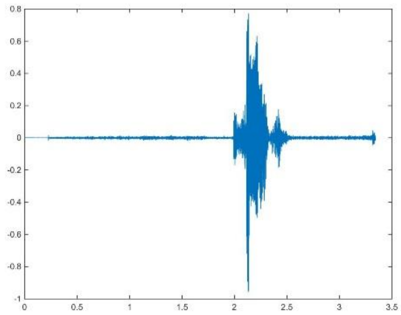
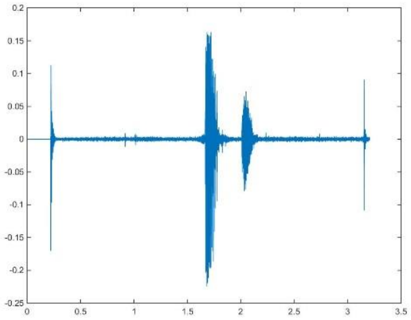
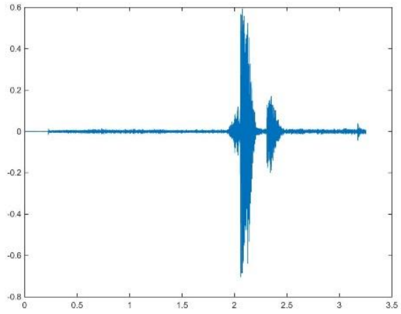
</td></tr>
<tr><td>48
(ice)</td><td>



</td><td>



</td></tr>
</table>

Words	Twin A’s waveforms	Twin B’s waveforms
49 (cold)		
50 (hot)	 0.2 0.15 0.1 0.05 0 -0.05 -0.1 -0.15 -0.2 -0.25 0 0.5 1 1.5 2 2.5 3 3.5	 0.6 0.4 0.2 0 -0.2 -0.4 -0.6 -0.8 0 0.5 1 1.5 2 2.5 3 3.5