Automatic Identity Recognition Using Speech Biometric

Mohammad A. M. Abushariah, PhD
Assal A. M. Alqudah, MA
King Abdullah II School for Information Technology,
The University of Jordan, Amman, Jordan

doi: 10.19044/esj.2016.v12n12p43 URL:http://dx.doi.org/10.19044/esj.2016.v12n12p43

Abstract

Biometric technology refers to the automatic identification of a person using physical or behavioral traits associated with him/her. This technology can be an excellent candidate for developing intelligent systems such as speaker identification, facial recognition, signature verification...etc. Biometric technology can be used to design and develop automatic identity recognition systems, which are highly demanded and can be used in banking systems, employee identification, immigration, e-commerce…etc. The first phase of this research emphasizes on the development of automatic identity recognizer using speech biometric technology based on Artificial Intelligence (AI) techniques provided in MATLAB. For our phase one, speech data is collected from 20 (10 male and 10 female) participants in order to develop the recognizer. The speech data include utterances recorded for the English language digits (0 to 9), where each participant recorded each digit 3 times, which resulted in a total of 600 utterances for all participants. For our phase two, speech data is collected from 100 (50 male and 50 female) participants in order to develop the recognizer. The speech data is divided into text-dependent and text-independent data, whereby each participant selected his/her full name and recorded it 30 times, which makes up the text-independent data. On the other hand, the text-dependent data is represented by a short Arabic language story that contains 16 sentences, whereby every sentence was recorded by every participant 5 times. As a result, this new corpus contains 3000 (30 utterances * 100 speakers) sound files that represent the text-independent data using their full names and 8000 (16 sentences * 5 utterances * 100 speakers) sound files that represent the text-dependent data using the short story. For the purpose of our phase one of developing the automatic identity recognizer using speech, the 600 utterances have undergone the feature extraction and feature classification phases. The speech-based automatic identity recognition system is based on

the most dominating feature extraction technique, which is known as the Mel-Frequency Cepstral Coefficient (MFCC). For feature classification phase, the system is based on the Vector Quantization (VQ) algorithm. Based on our experimental results, the highest accuracy achieved is 76%. The experimental results have shown acceptable performance, but can be improved further in our phase two using larger speech data size and better performance classification techniques such as the Hidden Markov Model (HMM).

Keywords: Identity Recognition, Speech, Biometrics, Mel-Frequency Cepstral Coefficient, Vector Quantization

Introduction

Rapidly changed computer technology and fast growth of communication ways, makes everyday work easy and managed. Technology takes place everywhere, in business, education, market, security…etc. However, communication between human and these technologies become the main concern of many research areas, especially for developing automatic identity recognition systems. However, biometric technologies are among the most important technologies used in this area.

Biometric technology refers to the automatic identity recognition using physical or behavioral traits associated with him/her. Using biometrics, it is possible to establish physiological-based systems that depend on physiological characteristics such as fingerprint, face recognition, DNA… etc, or behavioral-based systems that depend on behavioral characteristics such as gait, speech…etc, or even combining both of them in one system. Therefore, biometrics technologies can be excellent candidates for developing intelligent systems such as speaker identification, facial recognition, signature verification...etc. In addition, biometric technologies are flexible enough to be combined with other tools to produce more secure and easier to use verification systems.

As the society becomes more electronically connected due to the evolution of information technology, individuals are expected to have more electronic transactions on daily basis, which make their authentication and identification very essential. As a result, traditional person identification approaches such as using a Personal Identification Number (PIN), or an ID card become insufficient and cannot satisfy the security requirements of online transactions (Kartik et al., 2008).

Individuals are created with unique physiological or behavioral characteristics. Therefore, such characteristics can be used as indicators to easily identify them. However, different biometrics may require different requirements, which makes adopting it very difficult. As stated by Jain et al.

(2004), performance, acceptability, and circumvention are among the major issues to be considered in order to develop a practical biometric system.

Another issue to consider is selecting the appropriate biometric that can best serve the population, and meet the specified recognition accuracy, speed, and resource requirements. The intended system must be accepted by the users, and must be robust enough to any sort of attacks (Jain et al., 2004).

The main motivation behind the selection of speech as the biometric for developing the automatic identity recognition system is that speech is one of the main features individuals use for person authentication. Ease of collecting the biometric data is another motivation behind selecting speech too.

This research emphasizes on the development of behaviometrics system for automatic identity recognition using speech biometric technology. English digits are selected for developing the speech based system.

Background and Related Work

Human from ancient time have used body characteristics such as speech and face in their daily life; from simple basic actions such as recognizing someone to allow him/her enter your house using his/her speech or face, to critical situations such as defining criminal identity from his/her fingerprint, face or any other characteristic.

Automatic identity recognition technology transformed the traditional manual recognition of human characteristics, to automated systems based on one characteristic or even combined more than one. The goal of automatic identity recognition technology, in board sense, is to create machines that can receive human's information such as fingerprint, speech, face, signature… etc, and process this information to recognize his/her identity.

Automatic identity recognition has been defined differently in respect to the various, yet different, applications and domains for which they are used. Researchers and scientists have defined automatic identity recognition systems according to the way they use them in their research work. However, all automatic identity recognition systems aim to automatically recognize human identity from input human characteristics.

On the other hand, human characteristics need standard methods in order to be recognized. Biometrics or biometric authentications consist of different methods to recognize human upon one or more biometric characteristics.

Since dealing with computers, network technologies, and automated systems become essential in our daily lives, biometric technology makes accessing computer systems, workplaces, networks…etc, easier, more friendly, secured, and controlled (Wang et al., 2010).

Biometric systems provide a number of advantages over the traditional manual systems, not just because they reduce processing costs and fraud rates, but they also reduce error rates, improving convenience, improving scalability, increasing physical safety, and improving accuracy (Pato et al., 2010; Philips, 2002).

Biometric system is a pattern recognition system that works by taking biometric data from a person, extracting a feature set from that data, and comparing feature set with a template one in the database (Jain et al., 2004; Phillips, 2002; Jayasekara et al., 2006; Choras et al., 2006).

Applications of biometric system may operate in two modes as follows:

1) Identification Mode: the system identifies a person by searching the templates of all users in the database, and tries to find a match. Therefore, to identify human identity, the system has to conduct one-to-many comparison to get the identification result (Kinnunen et al., 2006; Jin et al., 2009; Jain et al., 2004). It is important to highlight that identification in negative recognition applications (established only using biometrics) is a critical component, since a single person cannot use more than one identity (Jain et al., 2004).

2) Verification Mode: the system verifies user's identity by comparing captured biometric data with a template or templates stored in the system's database. In such system, the user claims an identity via personal identification number (PIN), smart card, or user name, then the system has to conduct one-to-one comparison to take a decision whether this identity is true or not (Guru et al., 2009; Jain et al., 2004; Phillips, 2002).

In any biometric system the first step is the enrollment, where new user information or characteristic such as speech, fingerprint, or signature is captured and checked, then features are extracted and stored in the system's database for future use as illustrated in Figure 1 (Jain et al., 2004; Pato et al., 2010).

Based on the system's type, verification or identification mode will be processed whenever the system is used. Figure 1 shows block diagrams of the required steps for verification and identification biometric system modes. User enrollment block diagram is common to both modes (Jain et al., 2004).

Biometric characteristics can be divided into two classes: physiological and behavioral characteristics, which can be obtained from speech, signature, hand-and-finger geometry, face shape, DNA, gait pattern, ear shape, fingerprint, iris scan, retinal scan…etc (Jain et al., 2008; Jin et al., 2009; Jayasekara et al., 2006; Jain et al., 2004).

Physiological characteristics are biometric characteristics, which reflect human characteristics that describe characteristics related to the shape

of the body such as fingerprint, iris, face, hand and finger geometry, DNA, and ear (Jain et al., 2008).

Behavioral biometric or also referred to as behaviometrics is related to the behavior of a person, in which the identification and/or verification of human depends on the way they provide information to the system. Required information could be passed to the system through speech, signature, lip movement… etc (Wikipedia, 2015; Revett, 2008).

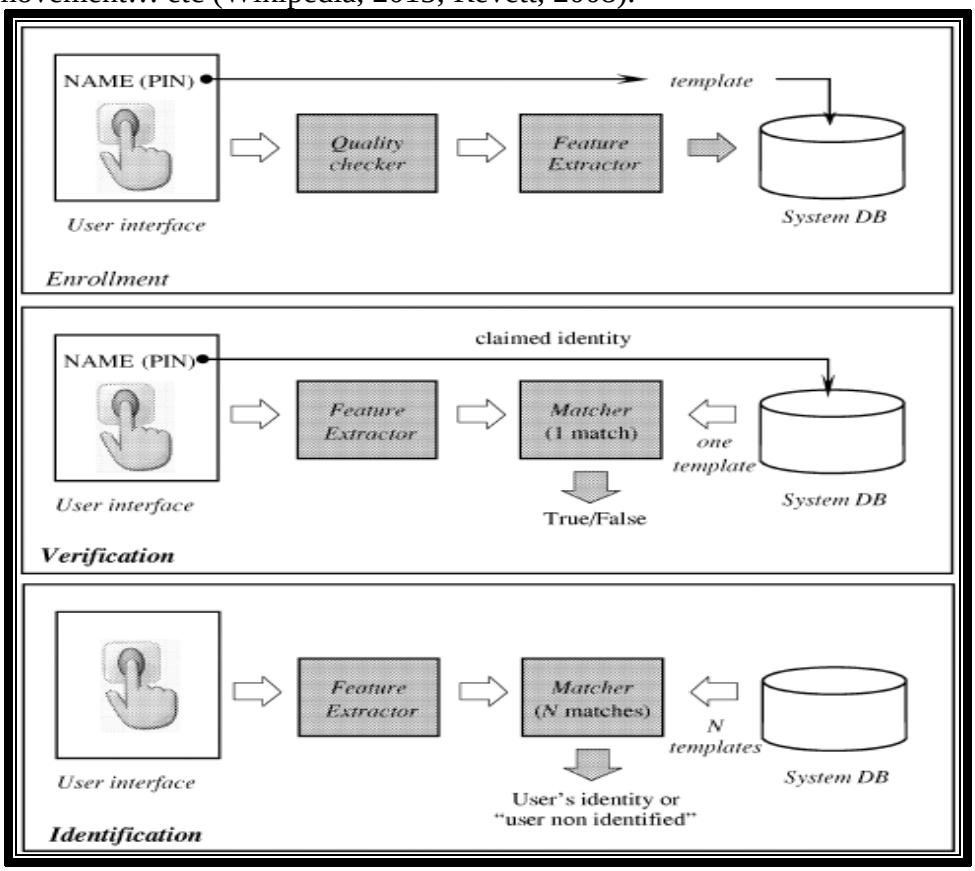

Figure 1: Block Diagrams of Enrollment, Verification, and Identification Tasks of a Biometric System (Jain et al., 2004)

Due to the reality that each person has distinct characteristics, biometrics nowadays become very effective personal identifiers. They are dependent and integral to an individual, therefore, they are more reliable, cannot be forgotten, and less likely to be lost or stolen compared to other identification methods such as identification cards, passwords, and PINs.

Consequently, the research community has recently witnessed a dramatic increment in the request for biometrics systems especially in identification and verification purposes, resulting in an increasingly widespread use of this technology for almost all aspects of our daily lives.

Among the many behaviometrics, speech is selected for this research and also for investigating the literature and related work in the next sections.

Feature Extraction Techniques for Speech Behaviometric Systems

Speech behaviometric is an important characteristic used for person identification systems. Such systems consist mainly of two phases; 1) features extraction phase, which works as a front-end followed by 2) features classification phase for generalized representation of extracted features (Singh et al., 2011a; Eriksson et al., 2005).

Selection of features for such systems is not an easy task; therefore, features for automatic identity recognition systems using speech should posses the following attributes (Zulfiqar et al., 2009):

1) Features should be easy to measure and extract.
2) Features occur naturally in speech.
3) Should not be affected by human's physical state and ambient noise.
4) Does not change over time.
5) Utterance variations; fast and slow talking rates.

On the other hand, the identification time in automatic identity systems depends on the following factors (Kinnunen et al., 2006):

1) The number of feature vectors and their dimensionality.
2) The complexity of features classification techniques.
3) The number of speakers.

Feature extractor is the first component in an automatic identity recognition system using speech and works as a front-end for the system (Singh et al., 2011a; Singh et al., 2011b; Kinnunen et al., 2006).

Feature extraction transforms the raw speech signal into a compressed one but effective version that is more stable and discriminative than the original signal. The output of the front-end part or feature extraction is very important since the quality of other system components such as features classification depends on it. The intention of using feature extraction is to reduce the dimension of the extracted vectors, which reduces the complexity of the system in return (Singh et al., 2011b; Eriksson et al., 2005).

Based on the literature investigation, there are various feature extraction techniques for behaviometric systems for automatic identity recognition using speech. This research has selected the most commonly used feature extraction techniques, which include Mel-Frequency Cepstral Coefficient (MFCC), Perceptual Linear Prediction (PLP) and Linear Predictive Coding (LPC).

MFCC features show better performance slightly more than PLP and LPC. MFCC analysis provides better performance than PLP resultant cesptral in an unconstrained monophone test, and MFCC is widely used in

speech based applications for speech and speaker recognition systems (Hönig et al., 2005; Milner, 2002; Singh et al., 2011a).

Table 1 shows a summary of features extraction techniques. This summary compares between MFCC, PLP, and LPC in terms of filters applied, the relevant variables, inputs and outputs. Output quality of these techniques varies from one technique to another; Table 2 shows critical analysis of MFCC, PLP, and LPC in order to determine the quality of the output of each technique (Abushariah, 2006).

Table 1: Summary of Features Extraction Techniques (Abushariah, 2006)

Process	Technique	Filters Applied	Relevant variables/Data Structures	Input	Output
Features Extraction	MFCC	Mel Filter Bank	Statistical Features MFCC Coefficients	Digital Sound Samples	MFCC Coefficients
	PLP	Bark Filter Bank	Statistical Features PLP Coefficients	Digital Sound Samples	PLP Coefficients
	LPC	All-pole Filter	Statistical Features LPC Coefficients	Digital Sound Samples	LPC Coefficients

Table 2: Critical Analysis of Features Extraction Techniques (Abushariah, 2006)

Criteria	MFCC	PLP	LPC
Main Task	Extracts features based on frequency domain using the Mel scale that represents the human ear scale	Approximates the psychophysical attributes of the human hearing process and estimates the auditory properties of human ear	Predicts the current speech sample based on analyzing past speech samples
Detect Voice and Unvoiced Sound	Able	Able	Unable
Speaker Dependence	Moderate Speaker Dependence	Low Speaker Dependence	High Speaker Dependence
Ability to be Used in Noisy Conditions	Good	Good	Poor
Motivation Representation	Perceptually Motivated Representation	Perceptually Motivated Representation	Speech Production Motivated Representation
Filter Bank	Triangular Mel Filters	Critical-Band Filters	All-Pole Filters
Amplitude Compression	Logarithmic	Cubic-Root	Auto-Regressive Modeling
Spectral Smoothing	Cepstral	LPC-Based Smoothing	Cepstral
Suitable Applications	Speaker and Speech Recognition, Emotion Detection	Speech Recognition	Speaker and Speech Recognition, Emotion Detection

Features Classification Techniques for Speech Behaviometric Systems

The second major stage in automatic identity recognition systems based on speech is features classification. The main purpose of this stage is to generalize representation of the extracted features, and the quality of the output of this stage is strongly determined by the quality of the output of the features extraction stage (Singh et al., 2011a; Eriksson et al., 2005).

Features classification for speech applications such as identity recognition is known as Pattern Recognition (PR). The main goal of pattern recognition is to split one class from others, whereby each individual class is known as a pattern, and to extract patterns based on specific conditions (Zheng et al., 2005; Huang et al., 2001).

After determining the best features for pattern recognition, the output is used to design the classifier using different approaches. In addition, optimization is applied in all PR stages from preprocessing whereby optimization ensures that the quality of the input pattern is the best; in feature selection and extraction whereby some optimization techniques are used to obtain optimal features subsets; and in classification whereby the error rate is minimized in order to complete the PR processes (Zheng et al., 2005).

Based on literature investigation, it is found that features classification techniques including Vector Quantization (VQ), Artificial Neural Networks (ANN) and Hidden Markov Models (HMM) are among the most frequently used techniques for speech based automatic identity recognition systems, which are discussed in the following section. From speech recognition perspective, a summary of the features classification techniques is shown in Table 3.

Table 3: Summary of Features Classification Techniques (Abushariah, 2006)

Process	Technique	Relevant variables/Data Structures	Input	Output
Features Classification	HMM	Markov Chain	Sub-word Features (e.g. phonemes)	Comparison Score
	ANN	Number of Layers, Neurons and Initial Weights	Statistical Features (LPC, PLP, MFCC)	Final Weights
	VQ	Initial VQ Codebooks and Signal Distortion Value	Statistical Features (LPC, PLP, MFCC)	Final Codebook

A critical comparison between the above mentioned feature classification techniques is presented in Table 4.

Criteria	HMM	NN	VQ
Main Task	Models the inherent spectral and temporal variations between multiple examples of the same phrase, word, or phonetic unit	Attempts to mechanize the recognition procedure according to the way a person applies intelligence in visualizing, analyzing, and characterizing speech based on a set of measured acoustic features	Encodes groups of data in order to exploit the relation among elements in the group to represent the group as a whole more efficiently than each element by itself
Computational Complexity	Complex	Complex	Simple
Learning Form	Supervised	Supervised	Unsupervised
Time Consumption	Very Time Consuming	Very Time Consuming	Fast
Pattern Recognition Approach	Statistical	Artificial Intelligence	Statistical
Suitable Applications	Large Vocabulary Speech	Isolated Words	Isolated Words and Large Vocabulary
Address Nonstationary Signals	Able	Unable	Able

Table 4: Critical Analysis of Features Classification Techniques (Abushariah, 2006)

Table 5 shows a comparison of results for speech based automatic identity recognition systems. Important legends for Table 5 are as follows:

EMD: Empirical Mode Decomposition
GRNN: Generalized Regression Neural Network
BPNN: Back-Propagation Neural Network
CHMM: Continuous Hidden Markov Models
GFM: Generalized Fuzzy Model
HMM3: Third-Order Hidden Markov Models.
PSD: Power Spectrum Density

No.	System	Reference Source	Applied Features Extraction Techniques	Applied Features Classification Techniques	Recognition Rate (%)
1	Bangali Text Dependent Speaker Identification Using Mel Frequency Cepstrum Coefficient and Vector Quantization	Bhotto et al., 2004	MFCC	VQ	70% to 85%
2	Speaker Identification Using Mel Frequency Cepstral Coefficients	Hasan et al, 2004	MFCC	VQ	57% to 100%
3	A Real Time Speaker Identification using Artificial Neural Network	Hossain et al., 2007	MFCC	ANN	96%
4	Text-Dependent Multilingual Speaker Identification for Indian Languages using Artificial Neural Network	Ranjan et al., 2010	LPC	ANN	83.29% - 92.78%
5	Speaker identification system using empirical mode decomposition and an artificial neural network	Wu et al., 2011	EMD	GRNN and BPNN	74.82% - 89.89
6	A high-performance text-independent speaker identification of Arabic speakers using a CHMM-based approach	Tolba., 2011	MFCC	CHMM	80%
7	GFM-Based Methods for Speaker Identification	Bhardwaj et al., 2013	MFCC	HMM- GFM	75% -90%
8	Novel third-order hidden Markov models for speaker identification in shouted talking environments	Shahin., 2014	MFCC	HMMBs	63% -64%
9	Speaker identification using vowels features through a combined method of formants, wavelets, and neural network classifiers	Daqrouq et al., 2015	PSD	ANN	90.09%
10	Efficient Online and Offline Template Update Mechanisms for Speaker Recognition	Anzar et al, 2016	MFCC	VQ	90%

Table 5: Comparison of Speech Based Automatic Identity Recognition Systems

Based on our literature review, we decided to select the combination of Mel-Frequency Cepstral Coefficients (MFCC) algorithm and the Vector Quantization (VQ) for phase one for the development of the speech based automatic identity recognition system, which is discussed further in the next sections.

Automatic Identity Recognition Using Speech Behaviometric

Based on literature investigation, the combination of MFCC and VQ is among the best choices for developing behaviometrics system for automatic identity recognition using speech. In addition, the Euclidean distance measure is found to be among the best choices too for features matching and evaluating the similarity or distortion. Therefore, our behaviometrics system for automatic identity recognition using speech as presented in this section uses MFCC algorithm for features extraction,

whereas it adopts the VQ for features classification, and finally it uses the Euclidean distance measure for features matching.

Architecture of the Speech-Based Automatic Identity Recognition System

The architecture of the automatic identity recognition system using speech includes certain components that are important for its implementation. The system's architecture is divided into two main phases. The first phase of the system's architecture is the training, whereas the second phase of the system's architecture is the testing/matching.

The system's architectures are designed using a pipe and filter architecture since each component of the architecture has a set of inputs and corresponding outputs and processes. Each filter in the system's architectures reads a stream of data on its input and produces a stream of data on its outputs. Figure 2 shows the system's training and testing/matching architectures.

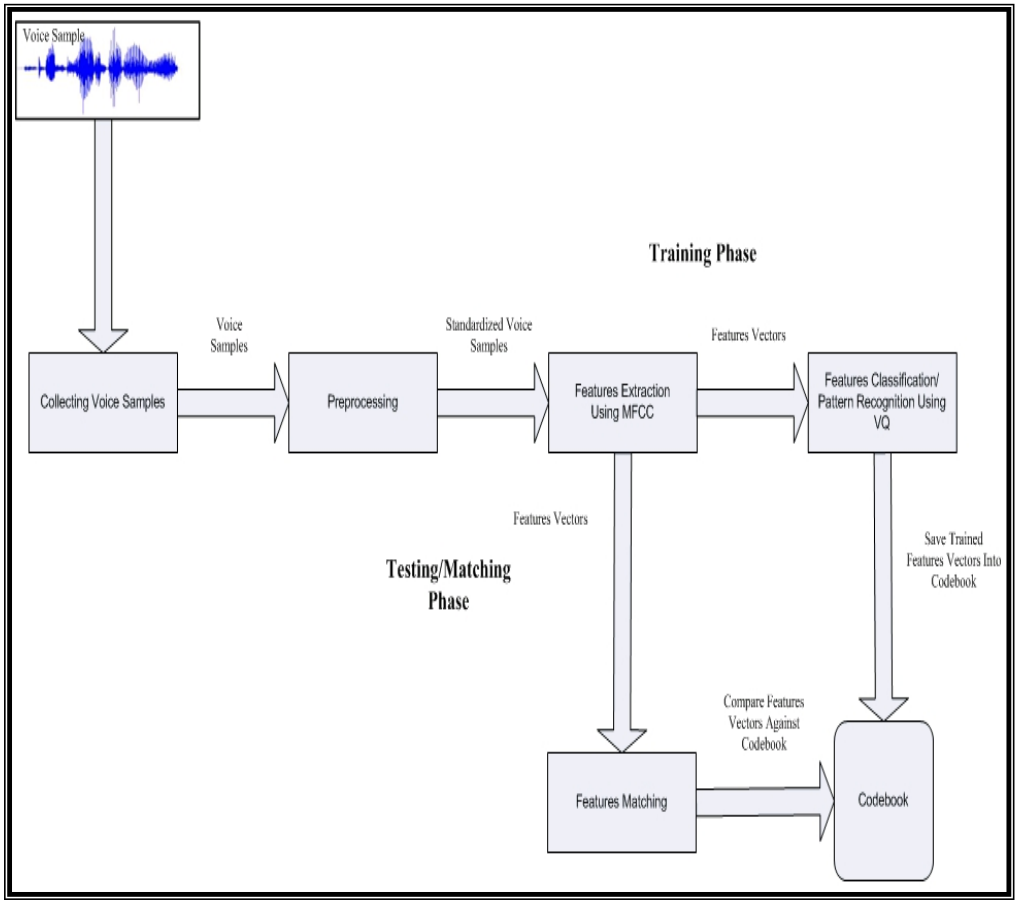

Figure 2: Architecture of the Automatic Identity Recognition System Using Speech

Implementation details of MFCC and VQ are explained further in the coming sections.

Speech Samples Collection (Speech Recording)

Speech samples collection is a very important step towards producing automatic identity recognition systems with efficient performance. For the purpose of developing the behaviometrics system for automatic identity recognition using speech, a speech database is required. This database includes recordings of 20 (10 male and 10 female) speakers. Each speaker was asked to utter all English language digits (zero to nine) three times each. Therefore, each speaker has a total of 30 utterances. The total number of utterances in this speech database is computed as follows:

Total Recordings = [(10 digits * 3 utterances) * 20 speakers] = 600 utterances

The 600 utterances are divided into training and testing data sets. In order to test the importance of training data size, two versions of the system are developed depending on the training and testing data size. The first version of the system has a distribution of (200 utterances for training the system, and 400 utterances for testing the system), whereas the second version of the system has a distribution of (400 utterances for training the system, and 200 utterances for testing the system).

The recording sessions took place in a sound-attenuated studio. Speakers used the SHURE SM58 wired unidirectional dynamic microphone to utter the recordings. They also used the Beyerdynamic DT 231 Headphone in order to listen to instructions from the recording specialist. In addition, the YAMAHA 01V 96 Version 2 (Digital Audio Mixer) was used. Sony Sound Forge 8 was used on a normal Personal Computer (PC) located in the studio with Windows XP in order to record the utterances from the speakers.

At this stage, speech utterances have been collected in order to be used for training and testing the system. These collected utterances are used in features extraction, features training and features testing stages, which are explained further in the next sections. In addition, a comparison in terms of the recognition rates based on the gender of the speakers is shown in the experimental results and analysis section. Therefore, a set of ten English language digits (Zero, One, Two, Three, Four, Five, Six, Seven, Eight and Nine) were recorded by 20 (10 male and 10 female) speakers.

Features Extraction Using MFCC

Features extraction phase is important in order to extract unique characteristics of each digit. MFCC technique is among the dominating techniques for features extraction from speech signals. The computational process of the MFCC was shown earlier in Figure 3.

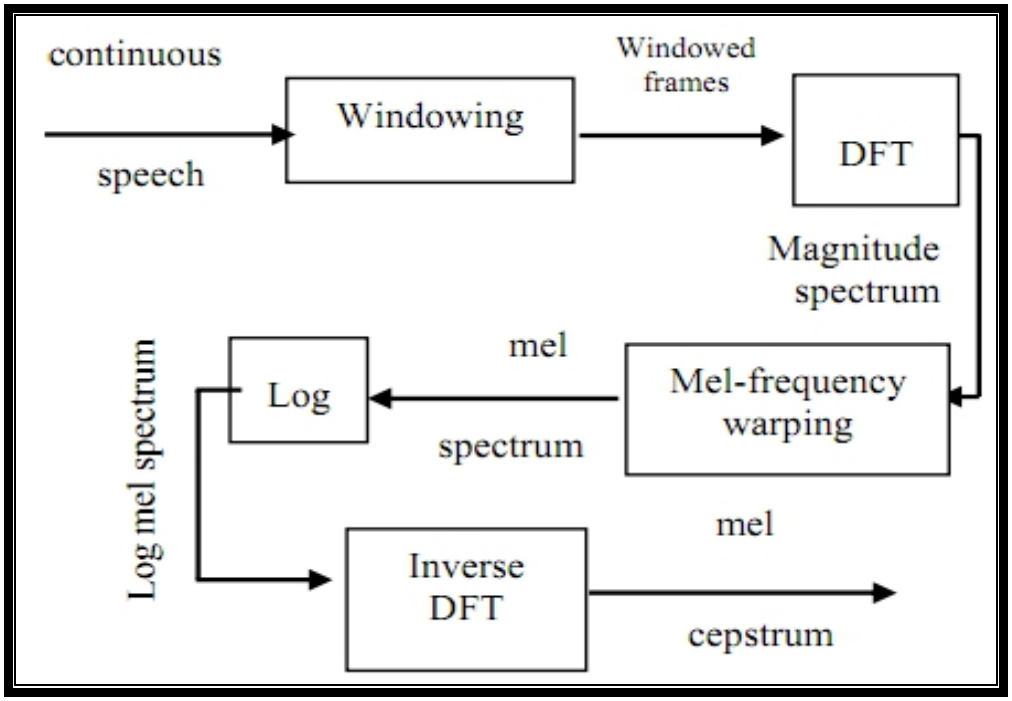

Figure 3: Block Diagram of MFCC Technique (Hossain et al. 2007)

Table 6 shows the most important parameters used in the MFCC features extraction technique.

Table 6: MFCC Main Parameters

Parameter	Defined Value
Sampling Rate (R_fs)	16000 Hertz per second
Frame Size (N)	256
Overlap Size (M)	156
Number of Filters (nof)	40

Based on the above parameters, the MFCC MATLAB code is executed. At this stage, the MFCCs are ready to form features vectors, which are then considered as inputs for the next section that trains them to form the VQ codebook. Each features vector has the vector size of [3237 * 1]. For the first version of the system, the features vectors have the size [3237 * 200] to be used for training the system using VQ, whereas the second version of the system has the features vectors size of [3237 * 400] to train the system using VQ.

Features Classification Using Vector Quantization (VQ)

This phase is divided into two parts, which are features training and features testing/matching. Features training enrolls an unknown speech utterance of a distinct English digit to the identification system's database.

This is performed by constructing a model of the digit based on its extracted features. In addition, features testing/matching computes a matching score, which refers to the similarity of the extracted features from the unknown digit and the stored digit models in the database. The unknown digit is identified by having the minimum matching score in the database.

Training the VQ codebook uses the LBG VQ algorithm, which clusters a set of *L* training vectors into a set of *M* codebook vectors. The LBG VQ requires the following steps as illustrated by Rabiner and Juang (1993):

1) Design a 1-vector codebook; which is the centroid of all training vectors. There is no iteration required at this stage.

2) Double the size of the codebook by splitting each codebook yn.

3) Nearest-Neighbor Search: find the codeword in the current codebook that is closest (in terms of similarity measurement) for each training vector, and assign the vector to the corresponding cell using the K-means iterative algorithm.

4) Centroid Update: update the centroid in each cell using the centroid of the training vectors assigned to that cell.

5) Iteration 1: repeat steps 3 and 4 until the average distance falls below a preset threshold.

6) Iteration 2: repeat steps 2, 3, and 4 until a codebook of size M is reached.

After executing the above mentioned steps for training the VQ codebook using the MATLAB code, the automatic identity recognition system now has trained sets of codebooks that are treated as the main databases of the system for testing/matching purposes.

In order to perform the testing/matching, the Euclidean distance measure is used in order to measure the similarity or the dissimilarity between two spoken digits. The matching of an unknown digit is performed by measuring the Euclidean distance between the features vector of the unknown digit to the codebook of the known digits in the database.

In the automatic identity recognition system, the Euclidean distance measure is applied on an unknown features vector compared against the trained codebook. The resulting outputs are the ID numbers assigned for each features vector in the trained codebook associated with the distances or the squared error values. This algorithm then picks up the ID number of the features vector that has the minimum distance to the unknown features vector.

Experimental Results and Analysis

The experimental work is evaluated based on the number of correct identification of the speech. This number is then divided by the total number

of the testing speech utterances, and then multiplied by 100 in order to get the percentage of the accuracy. Therefore, the accuracy (%) is calculated as follows:

Accuracy of the Speech System = [Number of Correctly Identified Utterances/Total Number of Tested Utterances] * 100

It is important to highlight that each utterance corresponds to one person. Therefore, the number of correctly identified utterances reflects on how many times the persons were correctly identified.

Testing and Evaluation of the Speech Based System

As highlighted earlier, the speech-based automatic identity recognition system has two versions that differ in the training data size. This is important in order to show the impact of the training data size on the overall accuracy of the system. Impact of training data size is examined for System 1 and System 2 as shown in Table 7, and Table 8, respectively. The accuracy for both systems is shown in form of confusion matrix in order to also identify the confusion of the result.

Based on the accuracy presented in Table 7 and Table 8, it is clearly seen that the training data has an impact on the overall accuracy of the system. It is found that the larger the training data, the higher the accuracy of the speech-based automatic identity recognition system.

Table 7: Confusion Matrix for System 1 (Training=10, Testing=20) for Each Speaker in the Speech-Based System

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Accuracy (%)
1	**12**	-	2	-	-	-	-	-	-	-	-	6	-	-	-	-	-	-	-	-	**60.00**
2	-	**14**	-	-	1	1	-	-	1	1	-	-	-	-	-	1	1	-	-	-	**70.00**
3	1	-	**17**	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	**85.00**
4	-	-	-	**20**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**100.00**
5	-	-	-	-	**9**	-	1	-	1	8	-	-	-	-	-	-	-	-	1	-	**45.00**
6	-	1	-	1	-	**13**	-	1	-	-	-	-	-	-	1	-	-	1	-	2	**65.00**
7	-	-	-	-	-	-	**19**	-	-	1	-	-	-	-	-	-	-	-	-	-	**95.00**
8	-	2	-	-	-	1	1	**10**	1	-	-	-	-	-	-	2	2	1	-	-	**50.00**
9	-	1	-	-	-	-	-	1	**14**	-	-	-	-	-	-	3	-	-	1	-	**70.00**
10	-	-	-	-	6	-	2	-	1	**9**	-	-	-	-	-	-	-	1	1	-	**45.00**
11	-	-	7	-	-	-	-	-	-	-	**13**	-	-	-	-	-	-	-	-	-	**65.00**
12	1	-	-	-	-	-	-	-	-	-	-	**15**	-	4	-	-	-	-	-	-	**75.00**
13	1	-	1	-	-	-	-	-	-	-	-	2	**14**	2	-	-	-	-	-	-	**70.00**
14	-	-	-	-	-	-	1	-	-	-	-	-	1	**17**	-	-	-	1	-	-	**85.00**
15	-	-	-	-	-	-	-	-	1	-	-	-	-	-	**17**	1	1	-	-	-	**85.00**
16	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	**16**	1	-	-	1	**80.00**
17	-	-	-	-	-	-	-	1	2	-	-	-	-	-	1	2	**12**	1	-	1	**60.00**
18	-	-	-	-	1	-	-	-	1	-	-	-	-	-	-	1	-	**17**	-	-	**85.00**
19	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	1	-	-	**17**	1	**85.00**
20	-	1	-	-	2	-	-	-	-	-	-	-	-	-	2	-	1	2	-	**12**	**60.00**
Average Results:																					**71.75**

Table 8: Confusion Matrix for System 2 (Training=20, Testing=10) for Each Speaker in the Speech-Based System

ID	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Accuracy (%)
1	**6**	-	1	-	-	-	-	-	-	-	1	2	-	-	-	-	1	-	-	-	**60.00**
2	-	**8**	-	-	-	-	-	-	-	1	-	-	-	-	-	1	-	-	-	-	**80.00**
3	1	-	**9**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**90.00**
4	-	-	-	**10**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**100.00**
5	-	-	-	-	**6**	-	1	-	-	3	-	-	-	-	-	-	-	-	-	-	**60.00**
6	-	-	-	-	-	**7**	1	-	-	-	-	-	-	-	1	-	-	-	-	1	**70.00**
7	-	-	-	-	-	-	**9**	-	-	-	-	-	-	-	-	-	-	1	-	-	**90.00**
8	-	-	-	-	-	-	1	**5**	-	-	-	-	-	-	-	3	-	1	-	-	**50.00**
9	-	-	-	-	-	-	-	-	**9**	-	-	-	-	-	1	-	-	-	-	-	**90.00**
10	-	-	-	-	5	-	1	-	-	**4**	-	-	-	-	-	-	-	-	-	-	**40.00**
11	-	-	5	-	-	-	-	-	-	-	**5**	-	-	-	-	-	-	-	-	-	**50.00**
12	-	-	-	-	-	-	-	-	-	-	-	**7**	-	3	-	-	-	-	-	-	**70.00**
13	-	-	-	-	-	-	-	-	-	-	-	-	**9**	1	-	-	-	-	-	-	**90.00**
14	1	-	-	-	-	-	-	-	-	-	-	-	1	**8**	-	-	-	-	-	-	**80.00**
15	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**9**	1	-	-	-	-	**90.00**
16	-	1	-	-	-	-	-	1	-	-	-	-	-	-	-	**8**	-	-	-	-	**80.00**
17	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	1	**8**	-	-	-	**80.00**
18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**10**	-	-	**100.00**
19	-	-	-	-	-	-	1	1	-	-	-	-	-	-	-	-	-	-	**8**	-	**80.00**
20	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	-	1	-	-	**7**	**70.00**
Average Results:																					**76.00**

This research work has a fair distribution of speakers' gender (10 males and 10 females). Therefore, the impact of gender is also examined in this work. It is important to highlight that the IDs (1 to 10) represent the male speakers, whereas the IDs (11 to 20) represent the female speakers in all experimental setup of this work. Table 9 shows the impact of the gender based on the results of System 1 and System 2 as shown earlier in Table 7 and Table 8.

Table 9: Impact of the Speakers' Gender on the Accuracy of System 1 and System 2 in the Speech-Based System

System	Average Accuracy (%) for Males	Average Accuracy (%) for Females
System 1	68.50	**75.00**
System 2	73.00	**79.00**

Based on the accuracy results presented in Table 9, it is found that the utterances collected from the female speakers performed better than those collected from the male speakers. This is due to the difference in the vocal characteristics of the male and female speakers.

Overall Experimental Results Analysis

Based on the overall accuracy of the speech based system, it is found that the size of the training data has a great impact on the accuracy. In all training and testing cases of our systems, it is found that both systems agree on this finding. Therefore, the final version of the speech-based system uses 20 utterances for each speaker to train the system and uses 10 utterances for each speaker to test the system. The larger the training data size, the higher the accuracy and the system will be able to recognize the speaker more accurately.

On the other hand, the impact of gender is also examined in our systems. The accuracy of females in the speech-based system is higher than the males. Speech characteristics are dynamic, and the reality of female voice is better than the male voice.

Overall, it is believed that the speech based system is able to identify male and female persons successfully. However, more improvements need to be taken into consideration such as increasing the number of persons and their data volumes. In addition, probably another biometric should be used that has static characteristics such as handwritten signatures, fingerprints, and others, which may result in better accuracy.

In order to improve the accuracy of our work, our next plan is to increase the speech data size. In fact, phase two of the data collection is already accomplished and we are in the process of development for the recognizer using more accurate and better techniques and classifiers such as the Hidden Markov Model (HMM), which will come out in a coming publication. Table 10 shows some details about our newly developed speech data corpus.

Criteria for Text-Dependent (Story)	**Total**	**Criteria for Text-Independent (Speakers' Full Names)**	**Total**
No. of speakers	100	**No. of speakers**	100
Story's number of sentences	16	**Speaker's Text Selection**	1
Story's number of repetitions for every sentence	5	**Speaker's number of repetitions for the selected text**	30
Total number of repetitions for all sentences	16 * 5 * 100 = 8000 Sound Files	**Total number of repetitions for all speakers**	1 * 30 * 100 = 3000 Sound Files

Table 10: Impact of the Speakers' Gender on the Accuracy of System 1 and System 2 in the Speech-Based System

This new speech corpus will be used mainly during our phase two of this research with a hope that better research contribution can be made.

Conclusion

Due to the fact that people in this technological era want to achieve their targets in easy and fast manner, the behaviometrics system for automatic identity recognition using speech is able to achieve the user's needs, whereby this system provides them with a very easy and fast way to recognize each other. This Behaviometrics system would make access available to any speech based systems and applications.

Based on our first research phase, it is found that the combination of MFCC and VQ techniques can work well in automatic identity recognition systems using speech behaviometric. This research phase also analyzed the impact of the size of training data and the gender of the participant. It is found that the larger the training data size, the higher the accuracy of the automatic identity recognition system. The gender impact is also examined in this project. It is found that female participants outperform the male participants. Finally, the system is able to achieve an accuracy of 76% using MFCC and VQ and such results are satisfactory at this level taking into account the small speech data size. However, this research work recommends larger speech data size and other combination of techniques such as MFCC and HMM should be investigated in our currently underway phase two of this research for better accuracy and performance optimization.

References:

Abushariah, M.A.M., (2006). A Vector Quantization Approach to Isolated-Word Automatic Speech Recognition. Master Thesis, Department of Software Engineering, Faculty of Computer Science and Information Technology, University of Malaya, Malaysia.

Anzar S.M., Amala K., Remya Rajendrana, Ashwin Mohana, Ajeesh P.S., Mohammed Sabeeh K., and Febin Aziz, (2016). Efficient online and offline template update mechanisms for speaker recognition. *Computers and Electrical Engineering*, Elsevier, Vol. 50, pp. 10 – 25.

Bhardwaj, S., Srivastava, S., Hanmandlu, M., and Gupta, J. R. P., (2013). GFM-based methods for speaker identification. *IEEE Transactions on Cybernetics*, Vol. 43, No. 3, pp. 1047-1058.

Bhotto, M.Z., and Amin, M.R., (2004). Bangali Text Dependent Speaker Identification Using Mel Frequency Cepstrum Coefficient and Vector Quantization, 3rd International Conference on Electrical and Computer Engineering, Dhaka, Bangladesh, pp.569-572.

Che, C.W., Lin, Q., and Yuk, D.S., (1996). AN HMM APPROACH TO TEXT-PROMPTED SPEAKER VERIFIGATION, *Proceeding of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Georgia, USA, pp. 673 – 676.

Choras, R. S., and Choras, M., (2006). Hand and Shape Geometry and Palmprint Feature for the Personal Identification, *Proceeding of IEEE International Conference on Intelligent Systems Design and Applications (ISDA),* Jinan, China, pp.1085-1090.

Daqrouq, K., and Tutunji, T.A., (2015). Speaker identification using vowels features through a combined method of formants, wavelets, and neural network classifiers. *Applied Soft Computing*, Vol. 27, pp. 231-239.

Eriksson, T., Kim, S., Kang, H.G., and Lee, C., (2005). An Information-Theoretic Perspective on Feature Selection in Speaker Recognition, *IEEE Signal Processing Letters*, Vol.12, No.7, pp.500-503.

Guru, D. S., and Prakash, H. N., (2009). On-line signature verification and recognition: An approach based on symbolic representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 6, pp.1059-1073.

Hasan, M.R., Jamil, M., and Saifur Rahman, M.G., (2004). Speaker Identification Using Mel Frequency Cepstral Coefficients. 3rd International Conference on Electrical & Computer Engineering, Dhaka, Bangladesh, pp.565-568.

Hönig, F., Stemmer, G., Hacker, C. and Brugnaa, F., (2005). Revising Perceptual Linear Prediction (PLP). *INTERSPEECH*, Lisbon, Portugal, pp. 2997-3000.

Hossain, M.M., Ahmed, B., and Asrafi, M., (2007). A Real Time Speaker Identification using Artificial Neural Network, *Proceeding of IEEE International Conference on Computer and Information Technology (ICCIT)*, Dhaka, Bangladesh, pp.1-5.

Huang, X., Acero, A., and Hon, H.W. , (2001). *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*. Prentice Hall, Upper Saddle River, NJ, USA.

Jain, A. K., Flynn, P., and Ross, A. A., (2008). *Handbook of Biometrics*, Springer, USA.

Jain, A.K., Ross, A., and Prabhakar, S., (2004). An Introduction to Biometric Recognition. *IEEE Transaction on Circuits and Systems for Video Technology*, Vol. 14, No.1, PP. 4-20.

Jayasekara, B., Jayasiri, A., and Udawatta, L., (2006). An Evolving Signature Recognition System. *First International Conference on Industrial and Information Systems (ICIIS*), Sri Lanka, pp. 529-534.

Jin, M., Kim, J., and Yoo, C.D., (2009). Humming-Based Human Verification and Identification. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Taiwan, pp.1453-1456.

Kartik, P., Prasanna, S.R.H., and Prasad, R.V.S.S.V., (2008). Multimodal Biometric Person Authentication System using Speech and Signature

Features, Technical Conference of IEEE Region 10 TENCON, Hyderabad, India, pp. 1-6.
Kinnunen, T., Karpov, E., and Fränti, P., (2006). Real-Time Speaker Identification and Varification. *IEEE Transactions on Audio. Speech, and Language Processing*, Vol. 14, No. 1, pp. 277-288.
Milner, B., (2002). A Comparison of Front-End Configurations for Robust Speech Recognition. *Proceeding of International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Orlando, Florida, USA pp.797–800
Pato, J.N., and Millett, L. I., (2010). *Biometric Recognition Challenges and Opportunities*, The National Academies Press, Washington, D. C., USA.
Phillips, J. P., (2002). Human Identification Technical Challenges. *Proceedings of the IEEE International Conference of Image Processing (ICIP)*, Rochester, New York, pp.49-52.
Rabiner, L. and Juang, B.H., (1993). *Fundamentals of Speech Recognition*, Prentice Hall, NJ, USA.
Revett, K., (2008). *Behavioral Biometrics: A Remote Access Approach*, John Wiley and Sons, Singapore.
Ranjan, R., Singh, S.K., Shukla, A., and Tiwari, R., (2010). Text-Dependent Multilingual Speaker Identification for Indian Languages using Artificial Neural Network, *Proceeding of IEEE International Conference on Emerging Trends in Engineering and Technology*, India, pp.632-635.
Shahin, I., (2014). Novel third-order hidden Markov models for speaker identification in shouted talking environments. *Engineering Applications of Artificial Intelligence*, Vol. 35, pp. 316-323.
Singh, S., and Rajan, E.G., (2011a). MFCC VQ based Speaker Recognition and its Accuracy Affecting Factors, *International Journal of Computer Applications*, Vol. 21, No. 6, pp.1-6.
Singh, S., and Rajan, E.G., (2011b). Vector Quantization Approach for Speaker Recognition using MFCC and Inverted MFCC, *International Journal of Computer and Applications*, Vol. 17, No.1, pp.1-7.
Tolba, H., (2011). A high-performance text-independent speaker identification of Arabic speakers using a CHMM-based approach. *Alexandria Engineering Journal*, Vol 50, No 1, pp. 43-47.
Wang, L., and Geng, X., (2010). *Behavioral Biometrics for Human Identification: Intelligent Applications*, Medical Information Science Reference, Hershey, New York.
Wikipedia, (2015). Biometrics, http://en.wikipedia.org/wiki/Biometrics , accessed on 10 August 2015.
Wu, J. D., and Tsai, Y. J. (2011). Speaker identification system using empirical mode decomposition and an artificial neural network. *Expert Systems with Applications*, Vol. 38, No. 5, pp. 6112-6117.

Zheng, L., and He, X., (2005). Classification Techniques in Pattern Recognition, *The International conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG)* , Czech Republic.
Zulfiqar, A., Muhammad, A., and Enriquez, A.M.M., (2009). A Speaker Identification System using MFCC Features with VQ Technique, *IEEE International Symposium on Intelligent Information Technology Application (IITA)*, China, pp. 115-118.