**Indian Journal of Science and Technology,** *Vol 7(2), 104–112, February 2014*

*ISSN (Print): 0974-6846*
*ISSN (Online) : 0974-5645*

# Biometric Voice Recognition in Security System

**Hairol Nizam Mohd. Shah*, Mohd. Zamzuri Ab Rashid, Mohd. Fairus Abdollah, Muhammad Nizam Kamarudin, Chow Kok Lin and Zalina Kamis**

Universiti Teknikal Malaysia Melaka (UTeM), Faculty of Electrical Engineering, Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia; hnizam@utem.edu.my, zamzuri@utem.edu.my, mfairus@utem.edu.my, nizamkamarudin@utem.edu.my, Devil_Danny89@yahoo.com, zalina_fz@yahoo.com.sg

## Abstract

A voice recognition system is designed to identify an administrator voice. By using MATLAB software for coding the voice recognition, the administrator voice can be authenticated. The key is to convert the speech waveform to a type of parametric representation for further analysis and processing. A wide range of possibilities exist for parametrically representing the speech signal for the voice recognition system such as Mel-Frequency Cepstrum Coefficients (MFCC). The input voice signal is recorded and computer will compare the signal with the signal that is stored in the database by using MFCC method. The voice based biometric system is based on single word recognition. An administrator utters the password once in the training session so as to train and stored. In testing session the users can utter the password again in order to achieve recognition if there is a match. By using MATLAB simulation, the output can obtain either the user is being recognized or rejected. From the result of testing the system, it successfully recognizes the specific user's voice and rejected other users' voice. In conclusion, the accuracy of the whole system is successfully recognizing the user's voice. It is a medium range of the security level system.

**Keywords:** Biometric, Mel-Frequency Cepstrum Coefficients (MFCC), Voice Recognition

## 1. Introduction

Nowadays, a lot of residential area and the companies are using all kinds of security system to make sure their property is secured such as using password and User ID/Pin for protection. Unfortunately, all these security system is not secured at all because the pin code can be hacked, the ID card can be stolen and duplicated. Based on the reasons, a whole new technology of security system must bring out to increase back the confidential of the civilian about the security system[1].

A biometric technology is the one which use the user features parameter as the password. The feature parameters of everyone is unique, even the users are twins. Therefore, the voice recognition system is safe for the administrator user. Voice is the most natural way to communicate for humans. In this thesis, the issue of voice recognition is studied and a voice recognition system is developed for certain word being spoken[1].

Voice biometric technology for authentication user is more convenient and accurate. This is because the biometrics characteristic if an individual are unique and belongs to the personal until the user dead. It is convenient for the user because nothing to be carried or remembered and would not scare the ID card being stolen or password being hacked.

From a technological perspective it is possible to distinguish between two broad types of ASR: Direct Voice Input (DVI) and Large Vocabulary Continuous Speech Recognition (LVCSR). These systems will analyze users' specific voice and use it to fine tune the recognition of that user's speech, resulting in more accurate transcription.

**Author for correspondence*

The voice recognition system contains two main modules which are feature extraction and feature matching. Feature extraction is the process that extracts a small amount of data from the voice signal that can later be used to represent each user while feature matching involves the step to identify the unknown user by comparing extracted features from admin voice input with the ones from a set of known user.

The speech signal and its characteristics can be represented in two different domains which are time and frequency domain. An utterance is the vocalization of a word or words that represent a single meaning to computer.

## 2. Background

Human voice is an amazing tool. Each human has a unique tone, rhythm, frequency and pitch to express including where they stop in phrases and how quickly they speak depending on where they are in a phrase[2]. Obviously the average male has a lower voice than the average female but the average range of each person's voice is unique. Humans have the interesting characteristic of different accents when they speak. Even across one certain word there are several variances on the way words and in turn sound is produced. The highest value of the frequency that a human can produce is about 10 kHz while the lowest value is about 70 Hz.

Voice recognition is the process by which a computer identifies spoken words. It can split into two types which are text dependent and text independent. Text dependent is about the keywords or the phrases for the voice recognition while text independent is not specific on the text being said and is more flexible[3].

Hidden Markov Model (HMM) is one of the text dependent methods. At first, the voice of the user is speak through the microphone in .wav file is recorded. The voice signal is then using the A2D converter convert the analog signal to the digital signal[4]. Each utterance is converted to a Cepstrum domain in training phase. The features parameter of the user is then extracted out and takes to compare with the reference voice sample to produce a likelihood ration. Likelihood ration which is the comparison between the fit of two models to expresses how many times more likely the data are under one model than the other. After the likelihood ration, a decision is make either the user voice is being accepted or the impostor voice is being rejected. Figure 1 shows speaker verification process.

The speaker recognition system used to train automatically and computationally feasible to use. The HMM isolate the unwanted noise signal and models the spectral representation to compose by a mixture of Gaussian function. The spectral envelope is fitted to a number of Gaussian decided in the system set up. It would consist of Ceptstral Coefficients.

Another method used in voice recognition is Fusion Classifiers System used a minimum amount of input data to produce correct decision[1]. Each of the different classifiers shows the information about the voice patterns and combines with others classifiers; the system can achieve a better grade of security level. Input voice of the authenticated user as voice data, x. Perceptual Linear Prediction (PLP) coefficients are used as features. The model, S is set as the authenticated user. The voice is recorded and the feature parameter is extracted by using three different algorithms namely GMM, MFN and SVM. These three different algorithms used to calculate the match score between each authenticated user. Each different classifier will show the different feature parameter of the user and combined all the weight match scores of the classifiers fusion to give the decision whether the user is accepted or rejected. The system is determined through the False Acceptance Rate (FAR) and False Rejection Rate (FRR). Figure 2 shows the structure of the fusion classifier system.

The voice recognition system is implemented by using the MATLAB (SIMULINK). A 'voice reference template' is taken to compare against the voice authenticated user. A user must speak his/her name and saved in the form of .wav file. To recognize the user voice, several variables like pitch, dynamics and waveform are included and executed by using the function block that available in SIMULINK[5]. There is few process involved such as measurement of energy level of silence compare to energy level of short duration of the signal. Then, remove noise from the input

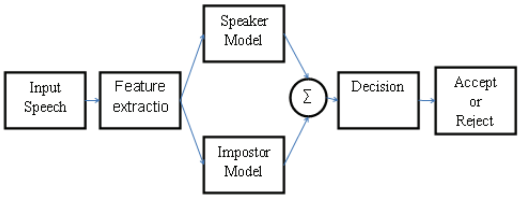


**Figure 1.** Speaker Verification.

voice signal by get through Digital Filter Design block. Next, extract features like determined pitch contour by computing autocorrelation, determined the format frequencies of three different vowels and also determined average energy spectral density using autocorrelation and FFT. The input voice of user will take to compare with the voice reference template. The output result is adjustable by setting the security level. If inside the range of the security level, logic "1" is produced otherwise logic "0" is produced.

Another method such as Dynamic Time Warping (DTW) algorithm is to measure similarity between two time series which may vary in time or speed[6]. The warping is used to find corresponding regions and determined the similarity between the two time series. The function of DTW is to compare two dynamic patterns and measure its similarity by calculating a minimum distance between the users[7]. DTW can analyze the data that can be turned into a linear representation and applied to the audio.

Two time series Q and C, of length n and m respectively represent in Equation 1 and 2.

$$Q = q_1, q_2, \ldots, q_i, \ldots, q_n \tag{1}$$

$$C = c_1, c_2, \ldots, c_j, \ldots, c_m \tag{2}$$

An n-by-m matrix where element of matrix contains distance d between the two points $q_i$ and $c_j$ is constructed. The distance is calculated by using the Euclidena distance computation in Equation 3.

$$d(q_i, c_j) = (q_i - c_j)^2 \tag{3}$$

While the accumulated distance is calculated by using Equation 4.

$$D(i,j) = \min\left[D(i-1, j-1), D(i-1, j), D(i, j-1)\right] + d(i,j) \tag{4}$$

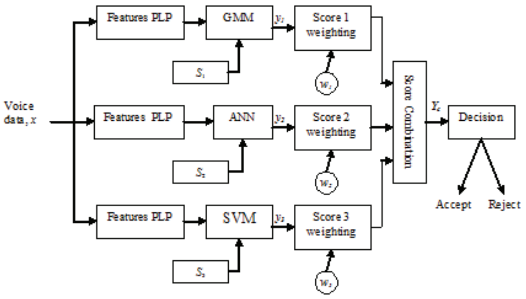


**Figure 2.** Fusion classifiers system.

The result of minimum distance between input and template signal will show at the path. Figure 3 shows about the Dynamic Time Warping.

The method that approach in voice recognition system is Gaussian Mixture Model (GMM)[8]. GMM is recognize the keyword and it divided into two steps which are isolate speech from utterance record and modeled statistical distribution of speaker characteristics in statistical way. Model of a user is calculated in training phase and stored in database[9]. As an example, the input voice signal is 8 kHz and filter by the low pass filter with a 3.8 kHz cutoff frequency in order to flat the spectrum. The signal is then window by the rectangular window of 25 ms overlapping to generate feature vector of 22 components by 10 Mel-cepstrum, 10 different Mel-cepstrum, energy and differential energy. There are two ways to represent speaker identity. First is the vocal tract configuration and the other one is linear combination of Gaussian basic function to represent a large class of distribution. Each model can has one covariance matrix per component. Gaussian mixture density is weight sum of M component densities by the Equation 5.

$$p(\vec{x} \mid \lambda) = \sum_{i=1}^{M} p_i b_i(\vec{x}) \tag{5}$$

Each $b_i$(x→) is Gaussian function of the form using Equation 6.

$$b_i(\vec{x}) = \frac{1}{(2\pi)^{\frac{D}{2}} \sum_i 1^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\vec{x} - \vec{\mu}_i)\right\} \tag{6}$$

And each of the speaker is represent by the GMM are referred by λ. Each speaker that identified is compared with the model in the database.

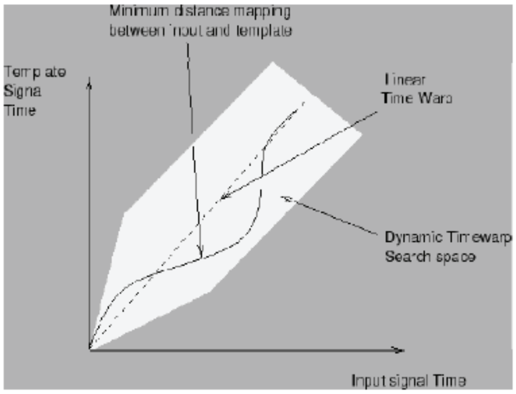

Minimum distance mapping between input and template
Template Signal Time
Linear Time Warp
Dynamic Timewarp Search space
Input signal Time

**Figure 3.** Dynamic Time warping (DTW)[6].

Furthermore, the method that approach to the voice recognition system is the pitch tracking that is a simple method based on the pitch detection via autocorrelation[10]. This method is to estimate pitch based on detect the highest value of autocorrelation function. The perception of pitch is related to the periodicity in time waveform.

$$R[m] = E\{x*[n]x[n+m]\} = \frac{1}{2}\cos(\omega_o n + \phi) \qquad (7)$$

From the Equation 7, it can find the pitch period by computing the highest volume of autocorrelation. For a low pitch male voice is as low as 40 Hz while high pitch female or child voice is as high as 600 Hz. With using the pitch track method, the input voice signal pitch can be easily detected and set the pitch of the speaker as the authenticated user. Figure 4 shows the pitch track with autocorrelation method.

The last method that would be approached to this project is Vector Quantization (VQ)[11]. VQ is a process of mapping vectors from a large vector space to a finite number of regions in that space. Each region is called a cluster and can be represented by its center called a codeword and collection of it named codebook. In training phase, each speaker will have a specific VQ codebook is generated for each known speaker by clustering user training acoustic vectors. The result of the distance from a vector to the closest codeword is name VQ-distortion. Speaker corresponding to the VQ codebook with smallest total distortions can be identified. Figure 5 shows the example of the vector quantization codebook formation.

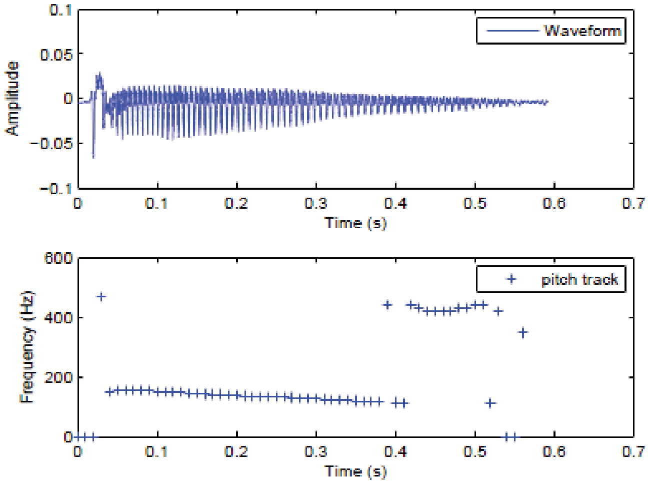


**Figure 4.** Pitch track with the autocorrelation method[11]

## 3. Methodology

In this project, MATLAB and ARDUINO will be used. MATLAB software is used for the voice recognition part while the ARDUINO software focus on the communication system part such as control the LED indicator switch, LCD screen display and the on/off of the magnet door. During the training phase, the input voice from the microphone will be extracted the actual uttered speech by the silence detection than apply hamming to smooth the voice signal. By using the MFCC, the energy feature of the user is extracted and saves as the reference template. The voice input signal from the testing phase will be check it is match with the reference template or not then calculate out its result. If the result is in the range with the reference template, then the voice is accepted and otherwise. Figure 6 shows the block diagram of the voice recognition system.

The voice recognition is divided into two phase which are training phase and testing phase. In training phase, the voice is recorded as long as one second. After that, silence detection will detect the actual uttered speech. The signal is then windowed. At first transformation, is the Fast Fourier Transform that convert voice signal from time domain into frequency domain. The MFCC convert to Mel Frequency Cepstrum. (MEL: change the scale of frequencies; CEPSTRUM: log followed by inverse Fourier Transform). A spectrum shows information of the frequency components while cepstrum show information about how the frequencies change to determine the energy within each window.

The steps training phase are exactly repeated in the testing phase. From the spectrum value get from energy within each window is used to determine value of mean square error and average pitch. The value is then compared with the training phase. If the result is rejected, the system will display "You are not the same user" while the

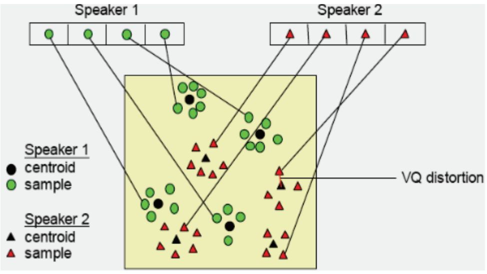


**Figure 5.** Vector quantization of codebook information.

result is accepted, they system will display “You are the same user”. Figure 7(a)–(c) shows block diagram voice recognition systems.

First, an input voice signal is recorded at a sampling rate above 10000 Hz from the microphone. This sampling frequency was chosen is to minimize the effects of aliasing in the analog-to-digital conversion. The voice signal is stored in 10000 sample vector. Figure 8 shows of the uttered speech of word “Hello”.

The actual uttered speech is extracted with the silence detection and the others will be ignored by filtering. Hamming window is applied to each window in order to decrease the spectral distortion created by the overlap window. Hamming window can improve the sharpness

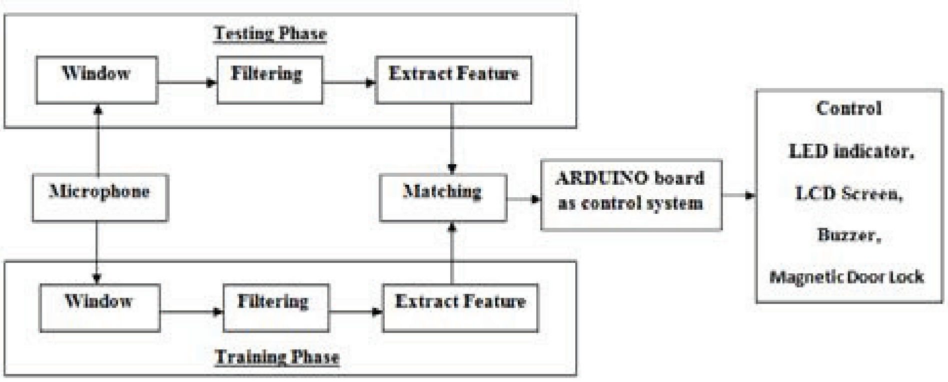


**Figure 6.** Block diagram of the voice recognition system.

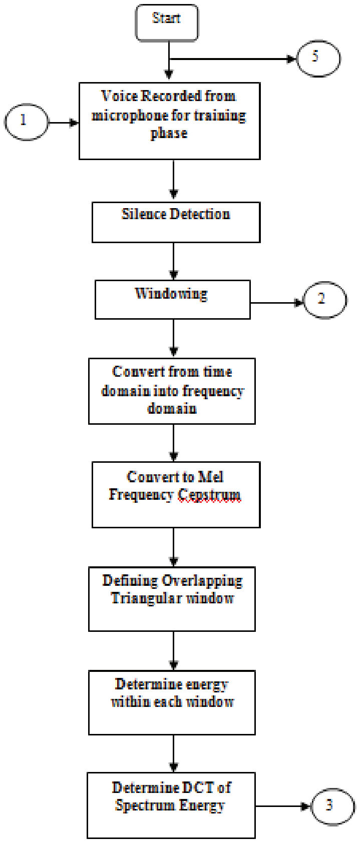

Start
5
1
Voice Recorded from microphone for training phase
Silence Detection
Windowing
2
Convert from time domain into frequency domain
Convert to Mel Frequency Cepstrum
Defining Overlapping Triangular window
Determine energy within each window
Determine DCT of Spectrum Energy
3

**Figure 7(a).** Voice recognition system.

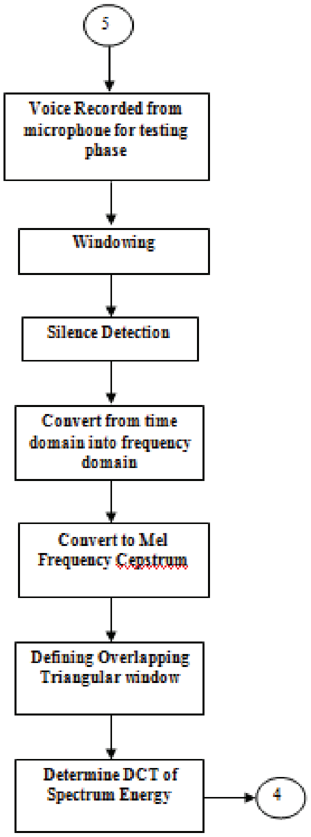

5
Voice Recorded from microphone for testing phase
Windowing
Silence Detection
Convert from time domain into frequency domain
Convert to Mel Frequency Cepstrum
Defining Overlapping Triangular window
Determine DCT of Spectrum Energy
4

**Figure 7(b).** Voice recognition system.

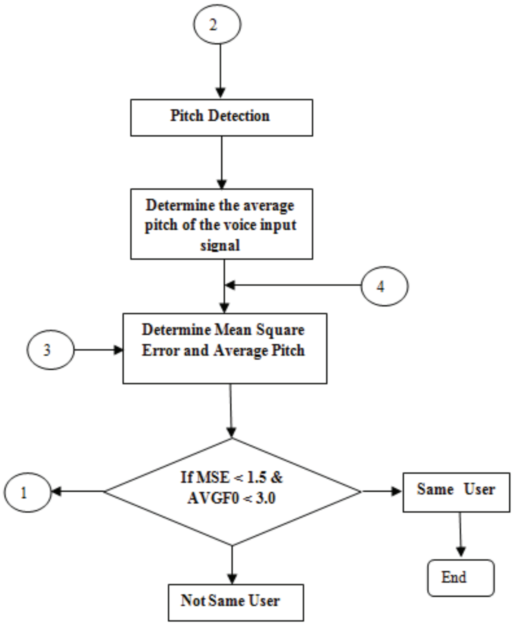

2
Pitch Detection
Determine the average pitch of the voice input signal
4
3
Determine Mean Square Error and Average Pitch
1
If MSE < 1.5 & AVGF0 < 3.0
Same User
End
Not Same User

**Figure 7(c).** Voice recognition system.

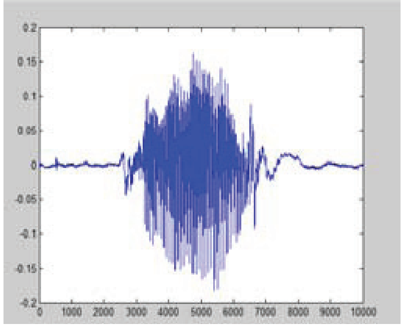


**Figure 8.** Uttered speech of "Hello".

of harmonics and removes discontinuities on the edges using Equation 8.

$$0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), n \ \varepsilon[0, N-1] \qquad (8)$$

The Fast Fourier Transform (FFT) is a powerful algorithm to calculate the discrete Fourier Transform and convert the signal from time domain to frequency domain. The FFT calculation time is 10 times lower than a classic DCT. The horizontal axis represents time while vertical axis represents frequency.

Feature extraction is a process that extracts a small amount of data from the input voice signal to represent the administrator speaker. This module converts a speech waveform to some type of parametric representation for further analysis and processing. MFCC method is used for coefficients calculation. Figure 9 shows block diagram MFCC approach mentioned variations.

Mel Filter Bank is use to determine the frequency content across each filter. The Mel filter bank is built from triangular filters. The filters are overlapped in such a way that the lower boundary of one filter is situated at the center frequency of the next filter. 1000 Hz was defined as 1000mels. An approximate formula to compute the Mels for a given frequency in Hz is using Equation 9.

$$Mel(f) = 2595 * \log_{10}(1 + f/700) \qquad (9)$$

The overlapping windows in the frequency domain can be directly used. The energy within each triangular window is obtained and followed by the DCT to achieve better compaction within a small number of coefficients and results known as MFCC. The data will be stored in the database and take to compare with the voice input at the testing phase with same steps of process.

## 4. Result

Two experiments are carried out to analyze the performance of the voice recognition system. One of the experiment is testing about the accuracy of own voice while another experiment is testing accuracy of other people's voice when my voice set as the reference template.

During a speaker uttered speech, his/her voice will produce a waveform and known as voice pattern. Every of the voice pattern is unique and different from other users. Therefore, the first and second experiments are used to analyzing the accuracy of the verification process.

Figure 10 shows the hardware setup for the voice recognition security system. With this system, the output data is able to read and transfer from the MATLAB by setting the baud rate at 9600 and all the I/O pins. A 5V is supply to the Arduiono Uno. Pin 2, 3, 4, 5, 6, and 13 all set as the output pin to connect with the LED indicator, buzzer, LCD display and magnet door lock. Arduino Uno is used for automatically reading the output data according to the condition of MATLAB software.

After the voice input is accepted and the systems determine as the admin user, then Arduino Uno there will activate the LCD display to display "WELCOME HOME, SIR", green LED indicator will turn on, and the magnet door lock will open. At the meanwhile, if the voice input is rejected and the system determine as the impostor user, Arduino Uno will activate the buzzer to turn on, red LED indicator will turn on, and LCD display will display "SORRY SIR! PLEASE RETRY AGAIN" and the magnet door lock remain lock.

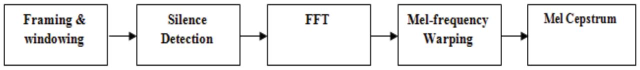


**Figure 9.** MFCC Approach.

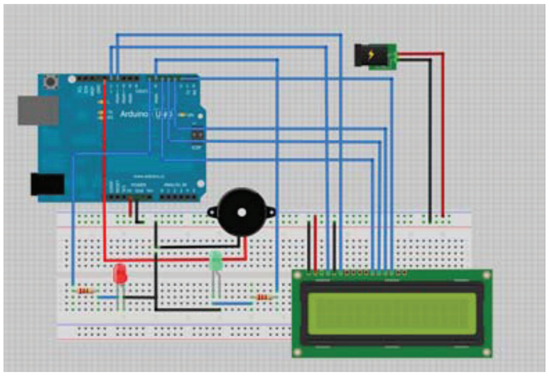


**Figure 10.** Hardware setup for the voice recognition security system.

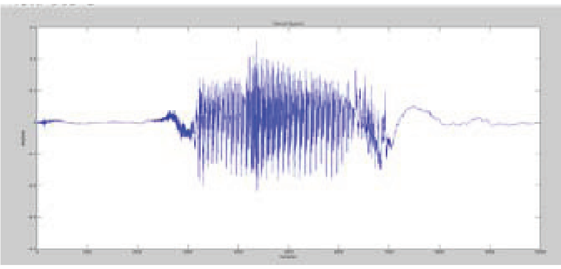


**Figure 11.** The word "HELLO" as voice input signal.

Microphone is used to record the user voice. The sampling frequency voice is set at the 10000 Hz and the duration period set as 1 second. Figure 11 shows the word "HELLO" is uttered by the user from the microphone.

The input voice signal is recorded as 1 second then the silence detection will extract only the uttered speech out and ignore the noise signal. Figure 12 shows the word "HELLO" after silence detection is move forward and the noise signal is being ignored.

After the actual speech signal is extracted by the silence detection, the Hamming window is used to smooth the input voice signal. Figure 13 is shows the word "HELLO" after smoother by Hamming Window.

After smoother the input voice signal, the signal is in time domain. The Fast Fourier Transform (FFT) is change the input voice signal from time domain to frequency domain. Figure 14 is shown the word "HELLO" at frequency domain.

Changing from time domain to frequency domain, the input voice signal is then calculated the energy by using a formula. Determine overlapping triangle window and the energy within each window. Figure 15 is shows the word "HELLO" after Mel-warping.

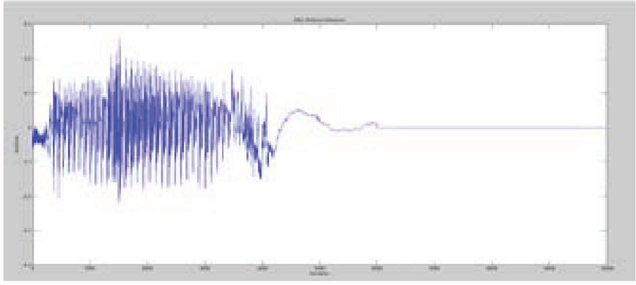


**Figure 12.** The word "HELLO" after silence detection.

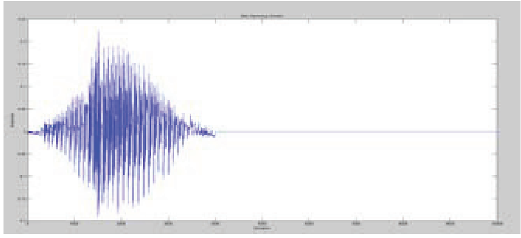


**Figure 13.** The word "HELLO" after using Hamming Window.

## 5. Discussion

By repeating the experiment to a recognized admin voice for 20 times, 5 times it failed to recognize administrator's voice. Thus our study able to muster 75% accuracy for this voice recognition system. The failure of the system to recognize the authenticated user's voice is due to energy depth variability of the speech uttered by the speaker. In the recognition algorithms it calculates the summation energy within each window and value of energy irrespective of whether the spectrum peaks at certain particular frequency. A user spoke loudly or soft, will affect to the energy of the voice signal. It will affect the output being accepted or rejected too. However, it would be better to improve the accuracy of the voice verification.

Table 1 shows the result for accuracy for voice recognition system.

Table 2 shows results for recognition admin user among 10 peoples which including admin user and the others user. Among the 10 peoples testing for the voice recognition system, there are admin user and Imposter E had been recognized by the voice recognition system, while the others are being rejected.

This experiment is doing the test with 10 different users, only one person is the authenticated user and the others are others people. Among the 10 people different gender or different age with the authenticated user, the voice recognition system is able to recognize admin's voice correctly. The different of gender and the different of age is to test whether can affect the accuracy of the voice recognition.

## 6. Conclusion

The voice recognition algorithm is developed by using MFCC method to extract the feature of the voice signal. The reference voice is being stored in training phase and compare with the voice in testing phase to match

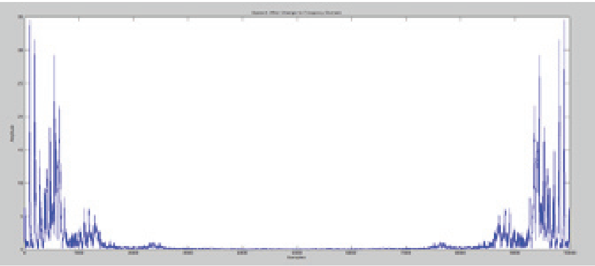


**Figure 14.** The word "HELLO" after FFT.

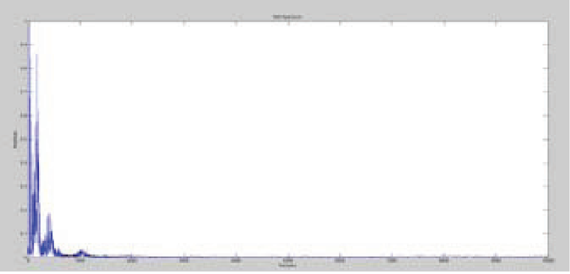
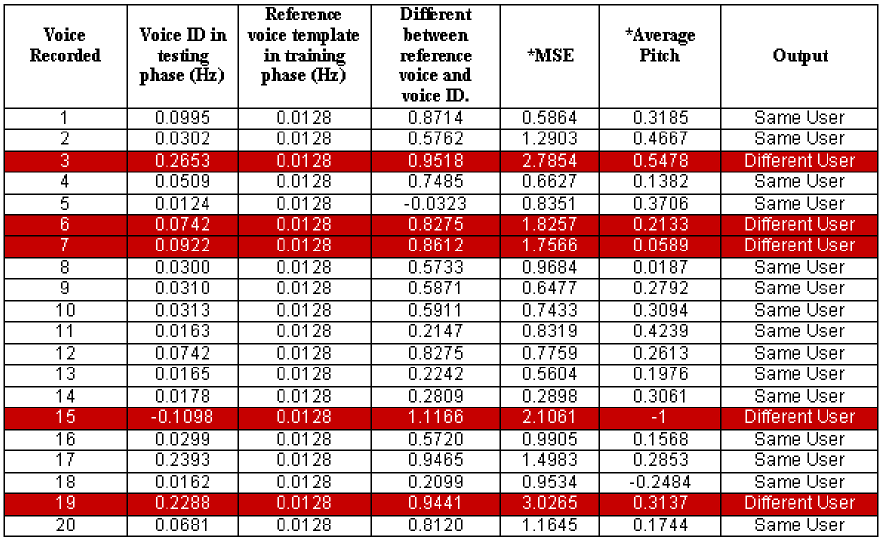


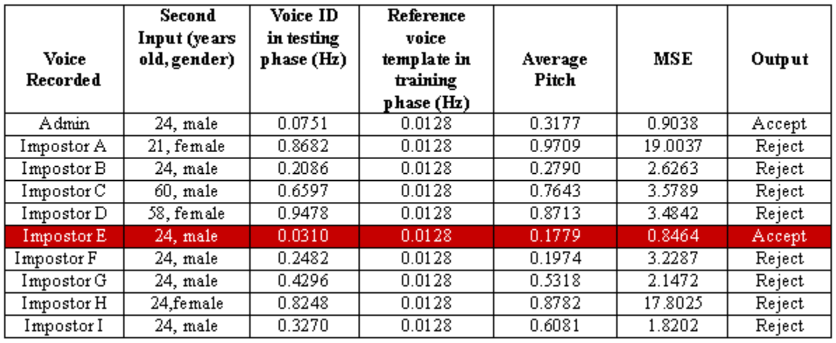
**Figure 15.** The word "HELLO" after Mel-warping.

**Table 1.** Result for accuracy for voice recognition system



| Voice Recorded | Voice ID in testing phase (Hz) | Reference voice template in training phase (Hz) | Different between reference voice and voice ID. | *MSE | *Average Pitch | Output |
|---|---|---|---|---|---|---|
| 1 | 0.0995 | 0.0128 | 0.8714 | 0.5864 | 0.3185 | Same User |
| 2 | 0.0302 | 0.0128 | 0.5762 | 1.2903 | 0.4667 | Same User |
| 3 | 0.2653 | 0.0128 | 0.9518 | 2.7854 | 0.5478 | Different User |
| 4 | 0.0509 | 0.0128 | 0.7485 | 0.6627 | 0.1382 | Same User |
| 5 | 0.0124 | 0.0128 | -0.0323 | 0.8351 | 0.3706 | Same User |
| 6 | 0.0742 | 0.0128 | 0.8275 | 1.8257 | 0.2133 | Different User |
| 7 | 0.0922 | 0.0128 | 0.8612 | 1.7566 | 0.0589 | Different User |
| 8 | 0.0300 | 0.0128 | 0.5733 | 0.9684 | 0.0187 | Same User |
| 9 | 0.0310 | 0.0128 | 0.5871 | 0.6477 | 0.2792 | Same User |
| 10 | 0.0313 | 0.0128 | 0.5911 | 0.7433 | 0.3094 | Same User |
| 11 | 0.0163 | 0.0128 | 0.2147 | 0.8319 | 0.4239 | Same User |
| 12 | 0.0742 | 0.0128 | 0.8275 | 0.7759 | 0.2613 | Same User |
| 13 | 0.0165 | 0.0128 | 0.2242 | 0.5604 | 0.1976 | Same User |
| 14 | 0.0178 | 0.0128 | 0.2809 | 0.2898 | 0.3061 | Same User |
| 15 | -0.1098 | 0.0128 | 1.1166 | 2.1061 | -1 | Different User |
| 16 | 0.0299 | 0.0128 | 0.5720 | 0.9905 | 0.1568 | Same User |
| 17 | 0.2393 | 0.0128 | 0.9465 | 1.4983 | 0.2853 | Same User |
| 18 | 0.0162 | 0.0128 | 0.2099 | 0.9534 | -0.2484 | Same User |
| 19 | 0.2288 | 0.0128 | 0.9441 | 3.0265 | 0.3137 | Different User |
| 20 | 0.0681 | 0.0128 | 0.8120 | 1.1645 | 0.1744 | Same User |

Table 2. Result for recognize admin user among 10 peoples



| Voice Recorded | Second Input (years old, gender) | Voice ID in testing phase (Hz) | Reference voice template in training phase (Hz) | Average Pitch | MSE | Output |
|---|---|---|---|---|---|---|
| Admin | 24, male | 0.0751 | 0.0128 | 0.3177 | 0.9038 | Accept |
| Impostor A | 21, female | 0.8682 | 0.0128 | 0.9709 | 19.0037 | Reject |
| Impostor B | 24, male | 0.2086 | 0.0128 | 0.2790 | 2.6263 | Reject |
| Impostor C | 60, male | 0.6597 | 0.0128 | 0.7643 | 3.5789 | Reject |
| Impostor D | 58, female | 0.9478 | 0.0128 | 0.8713 | 3.4842 | Reject |
| Impostor E | 24, male | 0.0310 | 0.0128 | 0.1779 | 0.8464 | Accept |
| Impostor F | 24, male | 0.2482 | 0.0128 | 0.1974 | 3.2287 | Reject |
| Impostor G | 24, male | 0.4296 | 0.0128 | 0.5318 | 2.1472 | Reject |
| Impostor H | 24,female | 0.8248 | 0.0128 | 0.8782 | 17.8025 | Reject |
| Impostor I | 24, male | 0.3270 | 0.0128 | 0.6081 | 1.8202 | Reject |

the both results. The system is successfully recognize the authenticate user's voice and rejected all the others impostor's voice. The output result is divided into two categories which are accepted and rejected. If accepted, the Arduino will activate the magnet door to unlock. If the output is rejected, the Arduino will remain the magnet door as lock and the buzzer will alarm for 1 second.

## 7. Recommendation

A few recommendations are provided to improve the accuracy and the performance of the voice recognition system. Firstly, increase the accuracy of the voice recognition system. The background noise must be filtered completely to get a more accurate data. Increase the complexity of the voice recognition by limit the range of the amplitude or frequency of the voice signal. So that the system can recognize the admin's voice more accurate. More set of data experiment can be obtained to gain the accuracy of the system. Software can be improved to block any wrong data from been collected or display. Besides that, utilizing the system to be an embedded system based access control device.

## 8. Acknowledgement

The authors are grateful for the support granted by Universiti Teknikal Malaysia Melaka (UTeM) in conducting this research.

## 9. References

1. Mohamed S, Martono W. Design of fusion classifiers for voice-based access control system of building security. WRI World Congress on Computer Science and Information Engineering; 2009 31 Mar-2 Apr; Los Angeles, CA. p. 80–84.
2. de Krom G. Consistency and reliability of voice quality ratings for different types of speech fragments. J Speech Lang Hear Res. 1994 Oct 37:985–1000.
3. Campbell JP. Speaker recognition: a tutorial. Proceedings of the IEEE. 1997 Sep; 85(9):1437–62.
4. Shrawankar U, Thakare VM. Techniques for feature extraction in speech recognition system: a comparative study. International Journal of Computer Applications in Engineering, Technology and Sciences (IJCAETS). 2013; 412–18.
5. Rashid RA, Mahalin NH, Sarijari MA, Abdul Aziz AA. Security system using biometric technology: design and implementation of Voice Recognition System (VRS). International Conference on Computer and Communi ation Engineering. 2008 13–15 May; Kuala Lumpur.p. 898–902.
6. Muda L, Begam KM, Elamvazuthi I. Voice recognition algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) techniques. Journal of Computing. 2010; 2(3):138–143.
7. Lama P, Namburu M. Speech recognition with dynamic time warping using MATLAB; 2010 Spring; Report no: CS 525.
8. Stuttle M, Gales MJF. A mixture of gaussians front end for speech recognition. 7th European Conference on Speech Communication and Technology. 2001 Sep 3–7. Aalborg, Denmark; 2001. p. 675–78.
9. Janicki A, Biały S. Improving GMM-based speaker recognition using trained voice activity detection. International Conference on Signals and Electronic Systems (ICSES 2006).
10. Nearey TM. Speech perception as pattern recognition. J Acoust Soc Am. 1997; 101(6):3241–54.
11. Rao, PM, Kumar S. Design of an automatic speaker recognition system using MFCC, Vector quantization and LBG algorithm. International Journal on Computer Science & Engineering. 2011; 3(8):2942–54.