FAST FOURIER TRANSFORM AND DISCRETE WAVELET TRANSFORM TO ENHANCE BIOMETRIC VOICE RECOGNITION

Najat Ahmed Elteraiki[1] and Mohammed M. Elsheh [2]

[1,2]Department of Information Technology, Libyan Academy, Misurata, Libya
sereen012015@gmail.com m.elsheh@it.misuratau.edu.my

ABSTRACT

This study investigates one of the technologies used in user recognition systems. Voice recognition is used as a tool to protect information systems from unlawful access. Other biometric techniques such as fingerprint, face print and iris print have shown some usage limitations such as big storage memory in the iris print technology. VRSs are based on the user's voice which is impossible to imitate and does not require costly facilities such as big storage memory. Using MFCC, DWT, and FTT, this study proposed an AVRS where the system trains and tests the voice of 10 speakers in two languages; English and Arabic. Previous studies have used only one language to train and test the speaker. The results of this study indicate that the rejection rate for Arabic words (12.9%) is higher than that for the English words (5.5%) which mean that the system is affected by the spoken words.

KEYWORDS

Voice Recognition, Fast Fourier Transform, Discrete Wavelet Transform

1. INTRODUCTION

Destroying and stealing information is one of the earliest risks of bad consequences to information systems or people. However, this unlawful practice has been largely exacerbated by the rapid advances of technology in the areas of communications and computers as well as by the large volume of information transmitted between the communication networks. Therefore, multiple information security techniques were introduced to protect information systems. Some of the first security techniques involve the use of user names and passwords as a means to identify the real user such as those used in computers' applications, ATMs and to gain access to buildings[1].

It has become a well-established fact that such continuous and rapid advances in the field of information technology has made information systems even more exposed to the realms of hackers (e.g. cyber attacks). As a result, IT scientists and researchers have come up with more sophisticated security techniques to protect and keep information systems secure, notably the use of biometrics as a means to identify the real user. Biometric is the use of biometric measurements of someone's body or behavioural characteristics, for example, their eyes, facial features, fingerprints, or voice to identify who that person is[2].

1.1 Biometrics

In information technology, biometric recognition, or simply biometrics, refers to the automatic recognition of individuals based on their physiological and/or behavioural characteristics, by using biometrics, the details required to confirm or establish the identity of users are based on the users' biometric characteristics such as their fingerprints or voices, unlike in the traditional security techniques where the access to systems is based on what the users possess (e.g., an ID card) or what they remember (e.g., a password)[3].

In essence, the use of biometrics in information technology is superior to traditional means of identification in that it does not rely on intangible details (e.g. passwords) but instead uses tangible details (e.g. human parts or behaviour-based characteristics) to identify the real user**.** Another advantage of the use of biometrics is the uniqueness of each individual's biometrics (i.e. scientifically each human has his/her own unique biometrics that are distinct from all other humans) which eliminates the risk of replicating the real user's biometrics hence, making it impossible for any potential illegitimate users to have fake access to information [4].

Biometrics systems for voice recognition often take three main separate processes: recording, extraction of templates, and comparing the templates. The purpose of the constraint is to collect and archive biometrics samples and create digital templates for future comparisons. By archiving raw samples, new alternative templates can be created in case a new or updated comparison-algorithm is introduced into the system. Practices that facilitate access to high-quality samples and improve the overall performance of the conformity process are particularly important for identity identification through the use of biometrics[1].

Biometrics is among of the most commonly used techniques for the identification and verification of the users of information systems. These include fingerprint, face-print, iris-print and voice print[5]. Since this paper is mainly focusing on using voice print, other techniques will not be discussed here and further information about them can be found in [6], [7], [8].

1.1.1 Voice-print

Voice-print technology represents another development in biometrics-based security systems. Unlike physiological biometrics that require the measurement of some bodily parts (e.g. fingerprint, iris-print, etc.) this technology looks into the voice of the system user, which makes it unique in nature compared to the other biometrics techniques. Scientifically, each person has their own unique voice features that distinguish them from all other people**.**This technology applies some mathematical models such as, the MFCC, DWT and FFT to analyze the user's voice. However, development of the processing techniques for digital signals in the 1960s led to an immediate automation in the user identification process[9].

This technology uses several ways to recognize the voice-print of a system user (i.e. the speaker) for example, by using the hearing, by using the sense of sight, or by using an automatic recognition process.

In this research, the author aims to demonstrate the way in which a system user is recognized automatically. This means some software programs have to be designed to compare and contrast voices to identify whether someone's voice is the same as their pre-recorded voice[10].

Why Voice-print?

The rapid usage of information technology has led to a revolution in system hackings. Therefore, the organizations that use information systems such as in banking, communications and networking require the latest advances in security to keep their systems safe and secure.

Voice is a phenomenon that depends entirely on the speaker producing it. There are multiple measurable characteristics to the voice for example, its tone and density, which are unique and distinctive for each person. Undoubtedly, such unique characteristics make the voice a powerful phenomenon that can be used in security systems.

In addition, the advantage of measurability and comparability of the voice characteristics contribute to the robustness of the practicality of using VRSs. The voice signal is a well recognized tool in the measurability of voice characteristics and has been studied for many years[11].

In summary, VRSs can be used in almost all organizations to better protect and preserve their private data and information. Due to its robustness and effectiveness, this technology reduces the huge cost usually spent on information security[12, 13].

2. RELATED WORK

Many studies have been done in last few years related to voice recognition and its techniques, method and algorithms. Some of them are discussed in the reminder of this section.

In [14], the authors constructed DTW, GMM and SVM for speaker recognition and observed that one way of applying SVM is to present each utterance as a single N-dimensional vector with N being fixed. In this research, a speaker recognition system which includes pre-processing phase, feature extraction phase and pattern classification phase was performed.

The experiments were carried out using TIMIT speech database. It was observed that one way of applying SVM is to present each utterance as a single N-dimensional vector, where N is fixed. It was shown that an investigation on a better normalization function has to be done to ensure that the SVM classifier gets a better accuracy rate.

The study in [15] investigates the wavelet-transform-based method for feature extraction using NNT. The lead system in his study uses two techniques to extract voice features; first, it uses WPXX over five approximated DWT levels and second, it uses classification based on FFBNN. WPXX is used due to its higher capability of illustrating formants over the various band-pass of signal frequency. A thorough comparison among other methods is made in his study. For example, a text-dependent system is used, where passwords or PINs are added to the means of identification, to widen the application of this system for example, banks and hotels tend to use such systems.

The system has an excellent capability of tracking the features even with 0dB SNR. Finally, Daqrouq conducted more than four thousand trails to evaluate his system's performance. The results show a 94% classification rate.

In [16], the authors present a robust speaker identification system which is based on the wavelet transform method. A large feature vector is formed in this study by the concatenation of the MFCCs and polynomial coefficients extracted from the speech signals and their DWT. Their experimental results show that the proposed method is useful for feature extraction even with the presence of noise distortions and telephone degradations in the speech signals. The results also show that to reduce the noise levels, wavelet de-noising is required as a preprocessing step for the speech signals at low SNRs.

In[17], the aim is to use a technique that specifies the basic acoustic signal and noise automatically. Automatic Voice Signal Detection (AVSD) is sampled from the same voice of male and female based on the frequency content of the acoustic signal. The goal of his paper is to automatically investigate the fake voice signal in the security system. The frequency content is assigned to the domain frequency by calculating its DFT using the FT algorithm for more than 15 words. The audio signal detection algorithm is continually calculated to calculate the difference from the absolute rate of two adjacent audio windows, and compare them with a predefined threshold. The length of the voice signal is set to 1,024 seconds. AVSD will support both mono and stereo .wav file, but if a .wav file is recorded, a compressed .wav file format will be processed, and AVSD processing will be better to take less execution time. For this problem, Voice Signal Compression (VSC) is already developed and also published that compression 50% of the source .wav files in the same extension. AVSD is applied to generalize computing.

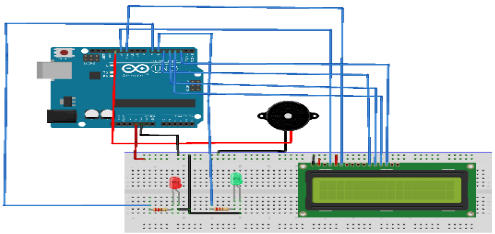
In [18], the researchers use DWT based on Mel-frequency cepstral coefficients (MFCC) and the traditional MFCC for a speaker identification system. They use pre-processing to remove the silent part of the speech signal, and a two-level DWT to extract the voice features. Their results show that MFCC-based DWT has an accuracy rate of 85%, whereas the traditional MFCC has an 80% rate of accuracy. The results also show that for feature extraction, the MFCC-based DWT and the LBG algorithm-based vector quantization techniques gave the maximum accuracy.

In [19], the study discusses the use of speaker recognition for speech processing applications, and authentication. Their research demonstrates and analyzes the most common techniques used for feature extraction. Their results show that MFCC is one of the most widely used methods for feature extraction. They also discuss different feature classification techniques for speaker recognition. The Discrete Wavelet Transform with logarithmic Power Spectrum Density (PSD) for speaker formants extraction was used as classification features. The method of Feed Forward Back Propagation Neural Network was proposed. The system works with excellent capability of features tracking even with 0dB SNR. This proposed system was compared with K-means algorithm based clustering method. Text-dependent system was used so that the system can be applied in password or PINs identification in any security systems.

3.0 EXPERIMENTS AND RESULTS

The first step is to ask the user to record their voice in order for the system to perform the testing and analysis steps on the voice as well as extract and save the extracted features in the database. Next, the system asks the user to speak the same word that was recorded, and glean the characteristics of the voice and compare it with the voice recorded in the first step. Computer RAM2G, CPU 2GHz, and operating system 8.1 were used for testing.

After recording the input voice through the computer microphone, a micro–controller named "ATmeg228" is used with the aid of MATLAB program V2013 to make a decision on the speaker. The whole circuit board used in this in experiment is shown in Figure.1.



Figuer.1 the Circuit implementation

If the voice input is rejected, the micro–controller will trigger the buzzer to turn on making a red LED indicator start flashing and the LCD to display “SORRY” which means the system has determined the speaker as an impostor. If the voice input is accepted, the green LED indicator will turn on, LCD display will display “welcome” which means the system has determined the speaker as the real user.

From the experiments conducted on 10 speakers using 10 words per speaker (5 Arabic words, 5 English words), and ranging between 20 and 50 attempts per speaker. Figure 2 shows results for recognition user's voice by pronouncing the word (ok) that the maximum number of rejected attempts was 10 and the minimum number of rejected attempts was 1.

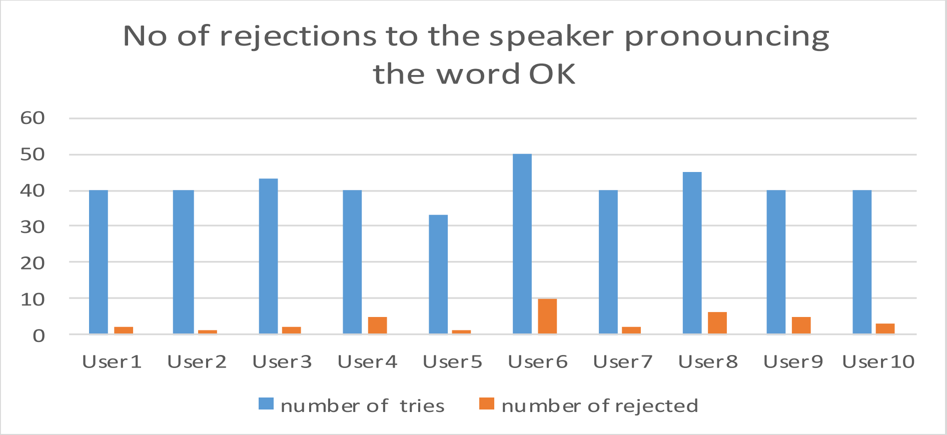


Figure 2 shows results for recognition user's voice by pronouncing the word (ok)

Figure 3 represents the system's capability to identify the user's voice by pronouncing the word (open) that the maximum number of rejected attempts was 11 and the minimum number of rejected attempts was 5.

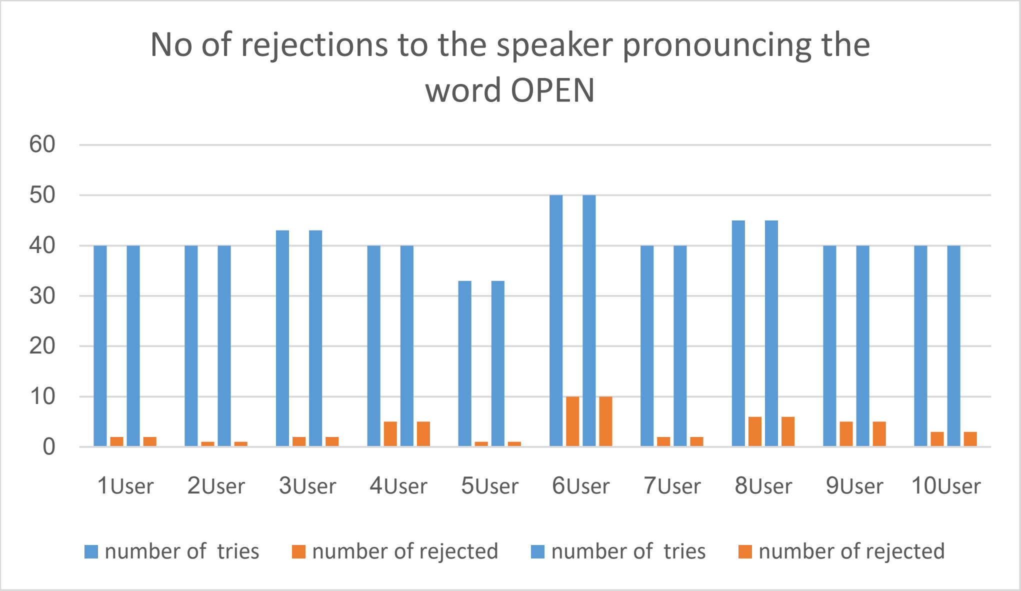


Figure 3 shows results for recognition user's voice by pronouncing the word (open)

It can be seen from Figure 4 which represents the system's capability to identify the user's voice by pronouncing the word (close) that the maximum number of rejected attempts was 4 and the minimum number of rejected attempts was 0.

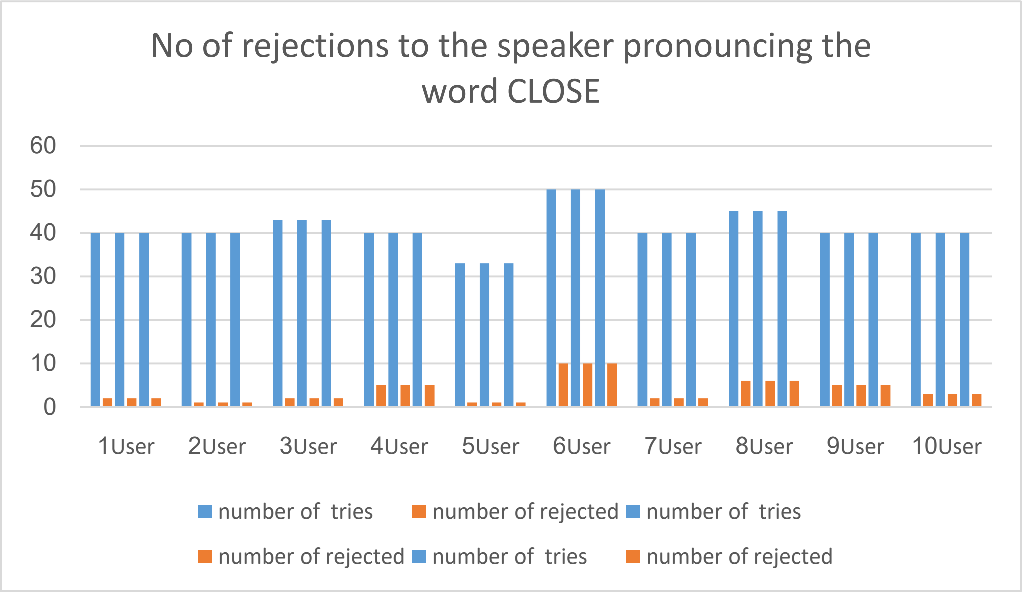


Figure 4 shows results for recognition user's voice by pronouncing the word (close)

It can be seen from Figure 5 which represents the system's capability to identify the user's voice by pronouncing the word (دخـول) that the maximum number of rejected attempts was 15 and the minimum number of rejected attempts was 7.

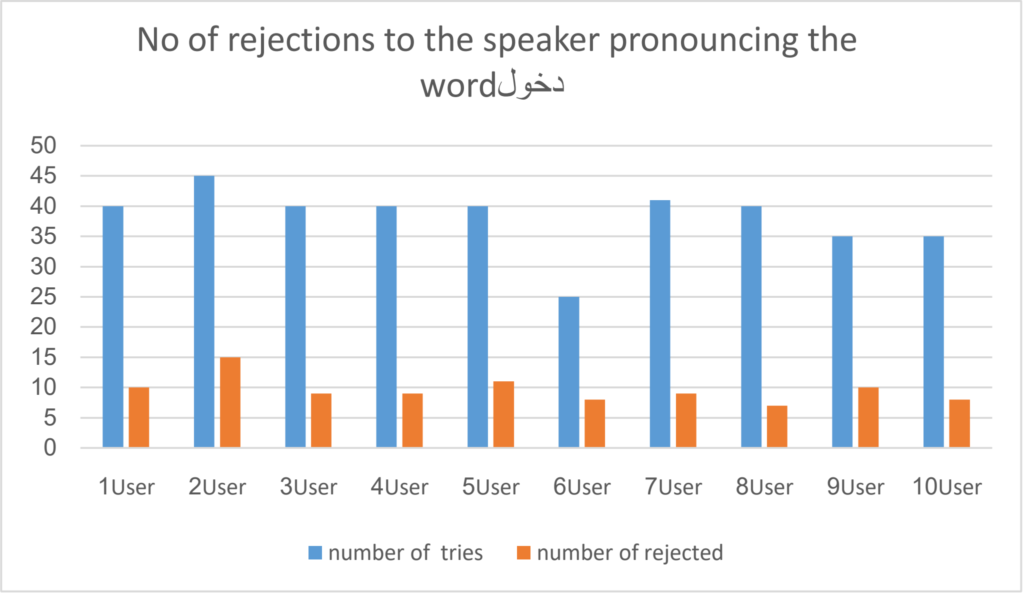


Figure 5 shows results for recognition user's voice by pronouncing the word (دخول)

4. CONCLUSION

An automatic voice recognition system was designed and tested on ten unknown speakers using 10 words; 5 Arabic words and 5 English words. The system extracts the specific features of the speaker's voice and compares them to the features already stored for each speaker in order to identify the identity of the speaker (i.e. real user or imposter). The researcher in this study applied MFCC, FFT, and DWT to extract the voice features, analyze them and produce a result based on the analysis. The feature is extracted using MFCC to see how well the system can recognize the speaker through their voice and whether or not the words spoken by the speaker have an effect on the rate of error in voice recognition. The results of the research found that the rate of error in the Arabic words is higher than that in the English words, and that the error rate between the words in one language is variant.

REFERENCES

1. Jain, A.K., A. Ross, and S. Prabhakar, *An introduction to biometric recognition.* IEEE Transactions on circuits and systems for video technology, 2004. **14**(1): p. 4-20.
2. Kaur, G., D. Singh, and R. Kaur, *A Review on Basic of Voice Recognition.* 2015.
3. Jain, A., L. Hong, and S. Pankanti, *Biometric identification.* Communications of the ACM, 2000. **43**(2): p. 90-98.
4. Jain, A.K., R. Bolle, and S. Pankanti, *Biometrics: personal identification in networked society*. Vol. 479. 2006: Springer Science & Business Media.
5. Anand, R., et al., *Biometrics Security Technology with Speaker Recognition.* International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 2012. **1**(10): p. pp: 232-236.
6. Gaur, S., V. Shah, and M. Thakker, *Biometric recognition techniques: a review.* International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 2012. **1**(4): p. 282-290.
7. Jadhav, I.S., V. Gaikwad, and G.U. Patil, *Human Identification using Face and Voice Recognition 1.* 2011.
8. Rani, M.U., J. Goutham, and M. Parthiban, *IMPROVED AUTHENTICATION USING ARDUINO BASED VOICE AND IRIS RECOGNITION TECHNOLOGY.*
9. Campbell, J.P., *Speaker recognition: A tutorial.* Proceedings of the IEEE, 1997. **85**(9): p. 1437-1462.
10. Furui, S., *An overview of speaker recognition technology*, in *Automatic speech and speaker recognition*. 1996, Springer. p. 31-56.
11. Rabiner, L.R. and M.R. Sambur, *An algorithm for determining the endpoints of isolated utterances.* Bell Labs Technical Journal, 1975. **54**(2): p. 297-315.
12. Shah, H.N.M., et al., *Biometric voice recognition in security system.* Indian Journal of Science and Technology, 2014. **7**(2): p. 104.
13. Yuschik, M. and R. Slezak, *Voiceprint identification system*. 2002, Google Patents.
14. Yee, L.M. and A.M. Ahmad, *Comparative study of speaker recognition methods: Dtw, gmm and svm.* 2007.

15. Daqrouq, K., et al. *Speaker identification system using Wavelet Transform and neural network*. in *Advances in Computational Tools for Engineering Applications, 2009. ACTEA'09. International Conference on*. 2009. IEEE.
16. Shafik, A., et al., *A wavelet based approach for speaker identification from degraded speech.* International Journal of Communication Networks and Information Security (IJCNIS), 2009. **1**(3).
17. Kumar, S., A. Shastri, and R. Singh, *An Approach for Automatic Voice Signal Detection (AVSD) Using Matlab.* International Journal of Computer Theory and Engineering, 2011. **3**(2): p. 240.
18. Yadav, S.S. and D. Bhalke, *Speaker identification system using wavelet transform and VQ modeling technique.* International Journal of Computer Applications, 2015. **112**(9).
19. Ramgire, J.B. and S.M. Jagdale, *A Survey on Speaker Recognition With Various Feature Extraction And Classification Techniques.* 2016.