Device Activation based on Voice Recognition using Mel Frequency Cepstral Coefficients (MFCC's) Algorithm

Hassan Mohammed Obaid Al Marzuqi [1], Shaik Mazhar Hussain[2], Dr Anilloy Frank[3]

[1,2,3]*Middle East College, Muscat, Oman*
[2,3]*Faculty, Dept. of Electronics and Communication Engineering, Middle East College, Muscat, Oman*

Abstract - The proposed paper presents the design and Implementation of a Security based system using Voice to activate any Device/circuit. This system is implemented using MATLAB and Arduino Uno as a testing component. The proposed system consists of two phases one is recording the user's voice signals. The extracted voice features are stored in the data base. In the second phase, the stored voice samples are compared with the samples in the database and determines whether it is matched or not. Mel Frequency Cepstral Coefficients (MFCC) Algorithm is used to extract the features of speaker's voice as it is found to be one of the most accurate techniques that could simulate the human audibility behavior. The testing device used in the proposed system is Ardunio Uno. Finally, The detailed simulation results were presented in this paper

Key Words: Mel Frequency Cepstral Coefficients; Device/circuit activation; Arduino Uno

1. INTRODUCTION

Security and Privacy are the major challenges that needs to ensure of handling any kind of sensitive information in any sector that includes managing Medical records, financial data, personnel information etc. So far, various technologies have been developed for security applications such as face scanning, voice printing and hand printing as discussed in [1]. One of the popular methods of securing the data is Password and User ID protection as discussed in [1]. With the rise of artificial intelligence, Voice recognition system has gained a lot of prominence. This technology gives friendly access to the users simply by speaking and enabling the devices. Neural and Fuzzy logic based automatic speaker recognition system is designed in MATLAB as discussed in [2]. A voice recognition system is a system where the voice characteristics of a person is identified. Voice recognition system is of two types, one is text dependent and the other one is text independent. In literature, Researchers have defined Speech/voice recognition in different ways. As in [3], Speech recognition is defined as the use of a machine to identify a person's voice based on his spoken phrases. Several algorithms have been developed for the design and implementation of voice recognition system. As in [4], a voice recognition system is developed for developing navigation autopilot system. To extract the voice features Linear Predictive Coding (LPC) algorithm is used and implemented using MATLAB. In [5], A Voice recognition system is developed for activated ground control station using MATLAB. An open source tool called Hidden Markov Model Tool Kit (HMMT) is used for voice recognition system. Similar to speech recognition system for security

EXISTING WORKS

Several literature surveys were studied and analyzed. A detailed comparative analysis is done. As mentioned in [10] , Home appliances are controlled using Zig bee through Voice commands. The system is designed using smart phone applications. However, the use of zig bee can lead to rigid problems. One of the problem could be loss of connectivity. In one of the article [11], voice recognition system is discussed based on word matching. However, this approach is not effective since it does not guarantee for Privacy. In a couple of papers [12] [13] , Speaker recognition methods were discussed which were not rely on specific audio patterns. In some of the papers [14] [15] , Mel Frequency Cepstral Coefficient technique (MFCC) is used for music modelling. In another article [16] , features of Speech and music have been extracted using MFCC to discriminate them. Very similar to this, several authors in [17] [18] have used Mel frequency Cepstral Coefficients (MFCC) for feature extraction which is found to be more effective method.

PROPOSED METHOD

The proposed block diagram is shown below in figure 1. The proposed work is carried out in two phases one is extraction phase and the second is Matching phase.

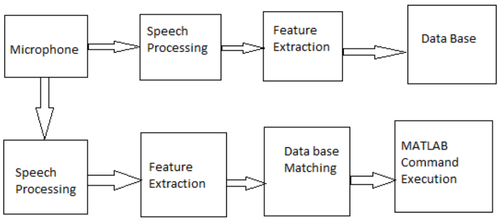

Figure 1 Proposed block diagram

Feature Extraction Phase

In this phase, The voice characteristics of a person is extracted through microphone and stored in data base. The speech processing techniques are applied after the first stage. Speech processing is carried out by sampling the signals in frame and reducing the spectral distortion. The signal is then converted to frequency domain. Filtering is done to get triangular shape and finally the signal can be retrieved through Discrete Cosine Transform (DCT) technique.

Matching Feature Phase

In this phase, a data base is generated by extracting voice features. The voice features are compared with the data base using an algorithm called Mel Frequency Celpstral Coefficients (MFCC) algorithm. If the features are matched then Device is activated or else remains in deactivation mode. Figure 2 below shows MFCC block diagram that shows all the stages involved in processing the voice.

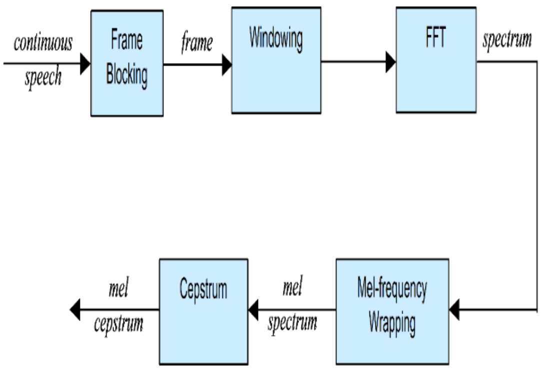

Figure 2 MFCC Block Diagram

Figure 2 above shows MFCC block diagram consists of various stages which is discussed in detail.

When the signal is received, Specific details about voice is extracted using MFCC which were discussed in detail below.

Frame Blocking: The received signal is separated in Frames and then sampled. Each frame contains 'N' Samples and continues till N-M samples and continues till speech signal is completed. The 'N' Samples has a value of 256 and 'M' has a value of 1000.

Windowing: This stage is used to reduce the spectral distortion

Fast Fourier Transform (FFT): This technique is used to convert Time domain to Frequency domain

Mel Frequency Wrapping: This technique is applied to convert wave shape to triangular wave.

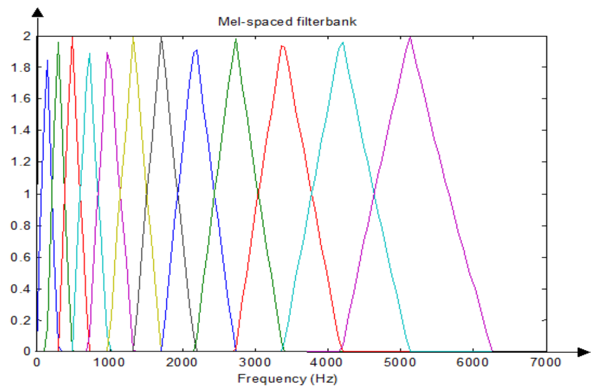

Figure 3 Triangular Wave

Cepstrum: In this stage, Discrete Cosine Transform (DCT) method is used to convert time domain to frequency domain. Using this process, Speakers Voice features can be used.

All the steps mentioned above are repeated for each frame for every 30mSec and compute a set of MFCC. This set is called as an Acoustic Vector.

Matching Phase

In this phase, Vector Quantization Technique is used as shown in Figure 4. Speaker voice recognition is identified with black circles and black triangles as a code word. The distance between the code words is called as VQ distortion and speaker with low VQ distortion is selected.

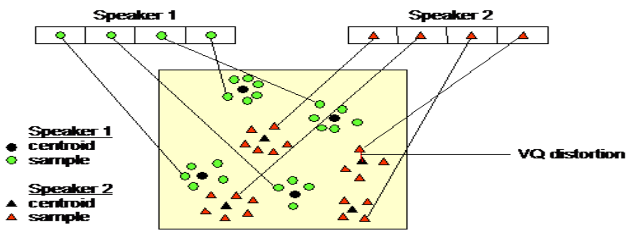

Figure 4 Vector Quantization (VQ)

The acoustic vectors are extracted from the input stages and assigns a set of Training Vector. A VQ code book is generated from the Training Vector. LBG Algorithm is used and 'L' Training Vector and 'M' Code book vector is clustered.

Figure 6 below shows the proposed flow chart. The flow chart shown below depicts the flow of proposed work. To testify the working of proposed system, An LED glow alarms the working of the system.

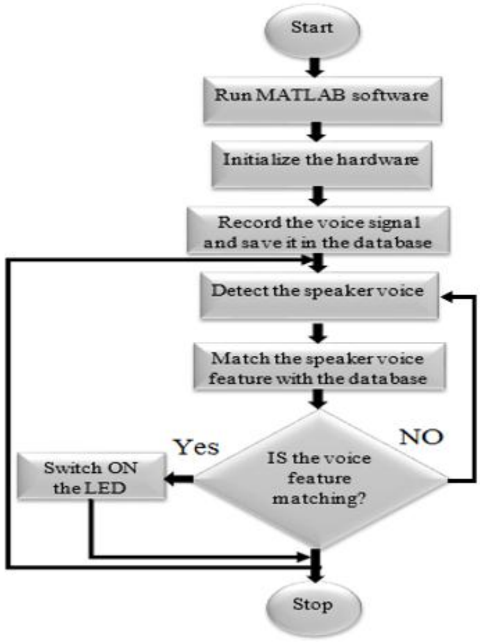

Figure 5 Flow chart of proposed system

Implementation Results

Figure 6 below shows the input voice signals that are used for testing.

Input Voice Signals

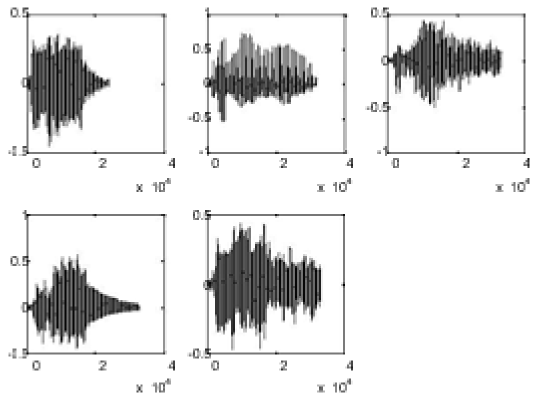

Figure 6 Input Voice Signals [7]

The below Figure 7 shows Prototype Voice signals and feature vectors.

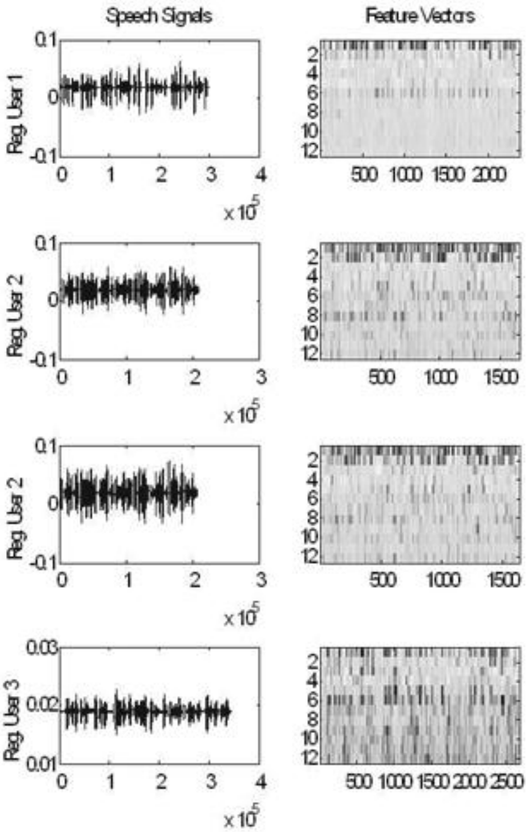

Figure 7 Voice signals and feature vectors [7]

The Simulation Result of the proposed system is depicted in the Figure 8 below.

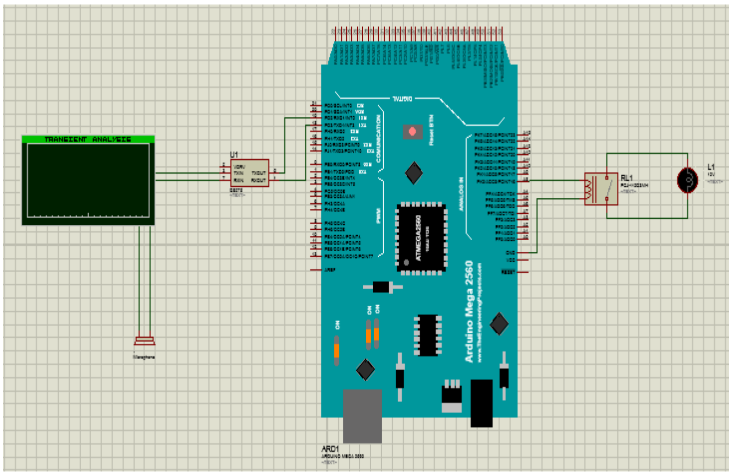

Figure 8 Simulation results of the proposed system

CONCLUSION

The research work is carried out in this paper is about activating any device using voice commands which could be used in societal applications to make the things secure . A detailed study is done on existing works. The work is executed in two phases. One is Extraction Phase and the second is Matching Phase. Mel Frequency Celpstral coefficients (MFCC) algorithm is used for extraction and matching the voice samples. As a part of testing, few voice samples are taken and their voice features are stored in data base and compared. Once matched, the device will be activated .The major contribution of our research work is based on security applications.

REFERENCES

[1] N. H. M. M. A. S. A. A. A. A. Rozeha A. Rashid, "Security System Using Biometric Technology: Design and Implementation of Voice," in Proceedings of the International Conference on Computer and Communication Engineering 2008, Kuala Lumpur, Malaysia , 2008.

[2] S. Jain, P. Jha and R. Suresh, "Design and implementation of an Automatic Speaker recognition system using neural and fuzzy logic in Matlab," in 2013 INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING AND COMMUNICATION (ICSC), Noida, India, 2013.

[3] J. Campbell, "Speaker recognition: a tutorial," Proceedings of the IEEE, vol. 85, no. 9, pp. 1437 - 1462, 1987.

[4] H. K. Tezer and M. Yağımlı, "Navigation autopilot with real time voice command recognition system," in 2013 21st Signal Processing and Communications Applications Conference (SIU), Haspolat, Turkey, 2013.

[5] K. S, V. S and R. Kumaraswamy, "Development of automatic speech recognition system for voice activated Ground Control system," in 2015 International Conference on Trends in

Automation, Communications and Computing Technology (I-TACT-15), Bangalore, India, 2015.

[6] A. F. Haque, M. M. Rahman, A. Khatun, M. Younus and J. F. Chowdhury, "Voice and Irish based automatic moving camera," in 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), Noida, India, 2015.

[7] T. Barbu, "Comparing various voice recognition techniques," in 2009 Proceedings of the 5-th Conference on Speech Technology and Human-Computer Dialogue, Constant, Romania, 2009.

[8] T.Barbu, "Speech-dependent voice recognition system using a nonlinear metric," International Journal of Applied Mathematics, vol. 18, no. 4, pp. 501-514, 2005.

[9] R. D.A.Reynolds, "A supervised text-independent speaker," IEEE Transactions, vol. 3, no. 1, pp. 72-83, 1995.

[10] S. B. M. F. K. Faisal Baig, "Zigbee Based Home Appliances Controlling Through Spoken Commands Using Handheld Devices," International Journal of Smart Home, vol. 7, no. 1, pp. 19-26, 2013.

[11] N. Gupta, S. Rawat and D. Arya, "VOICE RECOGNITION SYSTEM".

[12] J. M. H. U. A. Z. a. A. Z. R. A. Cole, ""Survey of the state of the art in Human Language Technology"," Cambrdige university press, Cambridge, 1997.

[13] C. S. B. Squires, ""Automatic speaker recognition: An application of machine learning"," in In Proceeding of the 12th International Conference on Machine Learning, 1995.

[14] O. Mubarak, E. Ambikairajah and J. Epps, "Novel Features for Effective Speech and Music Discrimination," in 2006 IEEE International Conference on Engineering of Intelligent Systems, Islamabad , Pakistan, 2006.

[15] B. Logan, ""Mel Frequency Cepstral Coefficients for Music Modelling"," in In Proceedings of Int. Symposium on Music Information Retrieval, Plymouth, MA, 2000.

[16] O. Mubarak, E. Ambikairajah and J. Epps, "Novel Features for Effective Speech and Music Discrimination," in 2006 IEEE International Conference on Engineering of Intelligent Systems, Islamabad,pakistan, 2006.

[17] H. Zhou, A. Sadka and R. M. Jiang, "Feature extraction for speech and music discrimination," in 2008 International Workshop on Content-Based Multimedia Indexing, London , UK, 2008.

[18] L. Jiqing, D. Yuan, H. Jun, Z. Xianyu and W. Haila, "Sports audio classification based on MFCC and GMM," in 2009 2nd IEEE International Conference on Broadband Network & Multimedia Technology, Beijing,China, 2009.