International Journal of Electrical and Computer Engineering (IJECE)
Vol. 4, No. 5, October 2014, pp. 643~647
ISSN: 2088-8708

Voice-Based Door Access Control System Using the Mel Frequency Cepstrum Coefficients and Gaussian Mixture Model

Kayode Francis Akingbade, Okoko Mkpouto Umanna, Isiaka Ajewale Alimi
Department of Electrical and Electronics Engineering, School of Engineering and Engineering Technology,
Federal University of Technology, Akure, Nigeria

Article Info

Article history:

Received Jun 12, 2014
Revised Aug 20, 2014
Accepted Aug 26, 2014

Keyword:

Access control
PINs
Security
Smartcard
Voice

ABSTRACT

Access to an area or environment can be controlled by conventional and electronic keys, identity cards, personal identification numbers (PINs) pads and smartcards. Due to certain limitations of existing door access schemes deployed for security in buildings, this paper presents speaker recognition for building security as a better means of admission into important places. This is proposed due mainly to the fact that speech cannot be stolen, copied, forgotten, lost or guessed with accuracy. This paper, therefore presents design of an affordable voice activated door control system for building security. The proposed system uses the Mel Frequency Cepstrum and the Gaussian Mixture Model for feature extraction and template pattern matching respectively. The analysis of the result which is based on the false acceptance and rejection rates indicate a system accuracy of more than 80%.

Corresponding Author:

Kayode Francis Akingbade,
Department of Electrical and Electronics Engineering,
School of Engineering and Engineering Technology,
Federal University of Technology, Akure, Nigeria
Email: kfakingbade@futa.edu.ng

1. INTRODUCTION

The implementation of access control prevents unauthorized individuals to access secure areas, buildings, documents and services. The control system consists of two main stages namely, the identification and verification stages. People that want to access a secure facilities introduce themselves to the system in the identification stage and the verification stage check the validity of the identities of the introduced users. If the identity of the user is valid, then the user may access secure area with the assigned permissions. The access control system is used for numerous applications such as for logging on ATM machines, e-banking accounts or for physical security of a room or building as a whole [1].

Access control for buildings is an essential device for protecting important places in the building that have valuable or highly sensitive materials. Server and strong room of banks are important areas where extremely effective control system is required. There ways of security implementation in a building and door access control is an integral part of them. The door access control is a means of securing building by giving limited access to specific people and by keeping records of such accesses [2]. Smartcard according to [2, 3] is the most common authentication method for the door access controls. It has been observed that a card-based access system can only control the access of authorized cards that are pieces of plastic, but not the ownership of the card. It can be used illegitimately by an unauthorized person when in possession of it. Furthermore, systems using PINs require individual to enter specific numbers to gain entry but the shortcoming is that those who really enters the codes cannot be determined system.

The limitations of conventional security systems call for better ones. There are varieties of biometric methods that could be employed in access control system for verification of authorized person into important or sensitive places. An automatic verification of identity in terms of behavioral and/or physiological characteristics of a person is carried out in the biometric methods [2, 3]. The biometric device identifies people by certain unique features such as the fingerprint, voice, face and eye (iris). Additionally, the device can eliminate the need for card-based access system. In the light of this, biometric devices can reduce the need for reissue of lost or damaged cards as the fingerprint, voice, face and eye are rarely stolen or lost.

The advantages of voice as a biometrics method are expatiated in [2] among which are simplicity for the user, speed of authentication and level of false-rejection rate. To resolve problems of the PINs pads and smartcards-based door access control, this paper presents voice-based door access control system using the Mel Frequency Cepstrum and Gaussian mixture model for building security.

The paper is organized as follows. Section 2 describes the proposed system. Section 3 focuses on system design and implementation. Results and performance evaluation are discussed in section 4. Conclusions are drawn in section 5.

In the following sections, we will quickly go through feature extraction and Gaussian Mixture Model. Next, we look at the operation and implementation of the voice based door control system and finally, we present performance evaluation and results.

2. PROPOSED SYSTEM

Research in speaker recognition and speech recognition is presently mature. Speaker recognition is essentially used in access control systems to give access to individuals whose identities are validated from their previously stored voice records or models. This involves both speaker identification and speaker verification [4]. It is, however, different from speech recognition which relies on the shared characteristic of what is said and what is stored in order to make a decision. Both are employed in speaker identification and verification systems [5].This paper uses a text independent speaker identification and verification process where the phrase or word to be said is not known to the system.

The design is implemented in two parts namely the software and the hardware parts. For the software, we use the Mel Frequency Cepstral Coefficients (MFCCs) for feature extraction and the Gaussian Mixture Model (GMM) for template matching. We use MFCCs because they are very robust and are the dominant features used for speech recognition [6]. Also, GMMs are usually preferred because they offer high classification accuracy while still being robust to corruptions in the speech signal. Also, they are very successful when it comes to noise handling. This has led to the extensive use of GMM based speaker recognition systems. The hardware part uses such components as d.c. motors, the L293B H-Bridge integrated circuit, a parallel port and the door structure.

2.1. Feature Extraction

The intention here is to have a model of the speech waveform that is sufficiently an accurate representation to the speech. It has been observed that the speech signal is a slowly time varying signal (quasi-stationary). This means that when observed over a sufficiently short period of time (between 5 and 100 ms), its characteristics are fairly stationary but change over long periods (0.2s or more) in order to reflect the different sounds being spoken. Therefore, to characterize the speech signal, the Mel Frequency Cepstral Coefficients (MFCCs) which is a tool for short time spectral analysis is employed. We refer to [6] for a complete description of the procedures for obtaining the MFCCs features. In this work, the programming platform used for voice processing and software development is MATLAB.

2.2. Gaussian Mixture Model

In [7-10], a Gaussian Mixture Model is described as a weighted sum of M component Gaussian densities given by the equation,

$$p(x|\lambda) = \sum_{i=1}^{M} \omega_i g(x|\mu_i, \Sigma_i)$$

Where x is a D-dimensional continuous-valued data vector (measurement or features),$\omega_i, i = 1, \dots, M,$ are the mixture weights, and $g(x|\mu_i, \Sigma_i), i = 1, \dots, M,$ are the component Gaussian densities with mean vectors μ_i and covariance matricesΣ_i.Each component density is D-variate Gaussian function of the form,

$$g(x|\mu_i, \Sigma_i) = \frac{exp\left\{-\frac{1}{2}(x-\mu_i)' \Sigma_i^{-1}(x-\mu_i)\right\}}{(2\pi)^{D/2}|\Sigma_i|^{1/2}}$$

The mixture weights satisfy the constraint that $\sum_{i=1}^{M} \omega_i = 1$. The parameters of the complete Gaussian model are collectively represented by the notation,

$$\lambda = \{\omega_i, \mu_i, \Sigma_i\} \qquad i = 1, \ldots, M.$$

In training the GMM, these parameters are estimated such that they best match the distribution of the training vectors [Fuzzy mixture Model for Speaker Recognition].

3. SYSTEM DESIGN AND IMPLEMENTATION

The process begins with the recording and training of voice samples otherwise called enrolment, which could be done either in real time or using a pre-recorded sample. A database for each of these samples exists such that any newly recorded speech would be saved there and not be lost either before or after the recognition process. For optimal results as in this case, it is very important that the recorded speech be obtained through the same means and if possible, processes every time. This is because the intrinsic properties of different microphones vary and could greatly affect the quality of the signal and the recognition system in general.

It is in this process that the analogue speech signal is converted to a digital signal by sampling. The analogue signal is conditioned with anti-aliasing filtering (and additional filtering if required to compensate for any channel impairments). The anti-aliasing filter limits the bandwidth of the signal to approximately the Nyquist rate (half the sampling rate) before sampling. This digitized speech is then further analyzed to extract the features that would be used for the recognition algorithm. Figure 1 shows the series of processes that the voice samples would undergo for a typical case where a verified I.D is enrolled and its model is subsequently compared with the features of a claimed I.D.

The hardware of this project is designed and built using a simple door prototype made with wood (plywood) having two DC motors. The DC motors are lightweight and consume less power, which implies that the batteries would last much longer. These motors provide the needed rotational displacement for the door to open and they are controlled by an H-Bridge IC (L293B). This IC is in turn driven directly by the parallel port of the system connected via a parallel port cable and controlled through MATLAB.

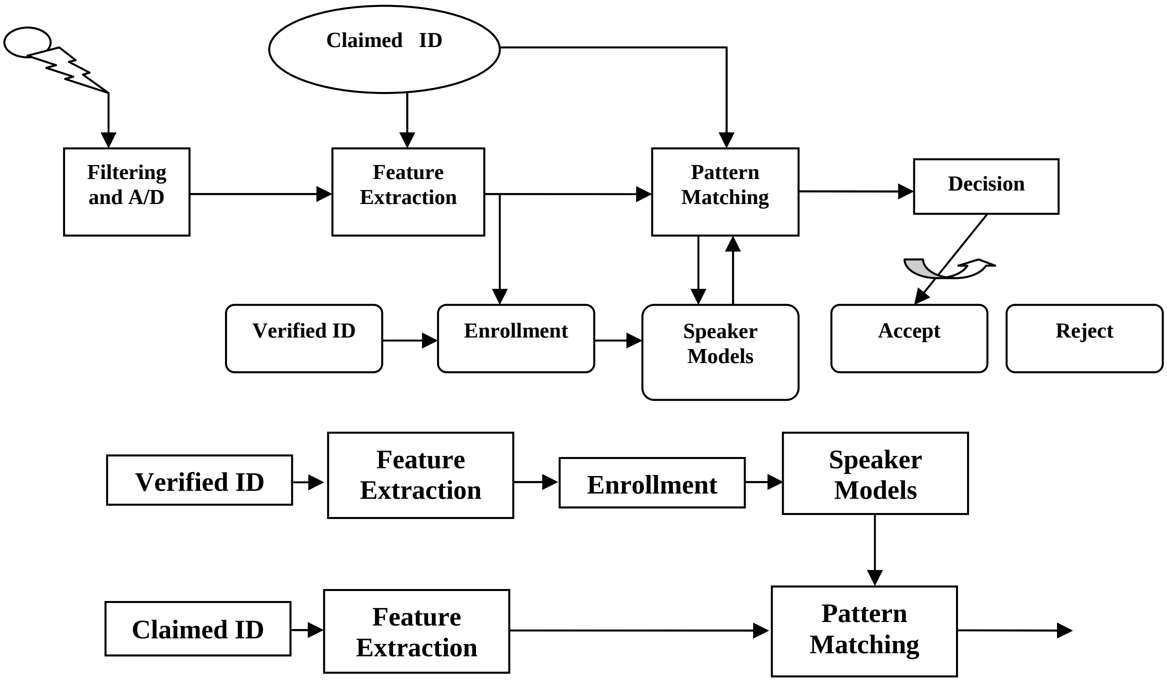

Figure 1. Operations on a typical analogue signal

4. RESULTS AND PERFORMANCE EVALUATION

A total of seven (7) voice samples from ten (10) different speakers that are recorded through the same process and at a sampling rate of **88.2KHz**is used for the performance evaluation. Since this system is not a text-dependent system, the voice samples are varied from names to numbers depending on the choice of the speaker. Furthermore, in assessing the system performance with respect to accuracy and reliability, we use the false accept rate and the false reject rate. Therefore, out of ten (10) verification trials each for every individual set,

$$\sum \frac{total\ false\ acceptance\ precentage}{total\ number\ of\ trials} \equiv \sum \frac{total\ number\ of\ false\ acceptance}{total\ number\ of\ trials}$$

$$FAR = 13.27\%$$

This invariantly means that the genuine acceptance probability of the system is;

$$100 - 13.27 = \mathbf{86.73}\%$$

The figures obtained for the FAR and Genuine Accept Rate (GAR) of this system clearly indicates that the system based on this test, has an efficiency of more than 80% so far. Similarly,

$$\sum \frac{total\ false\ rejection\ percentage}{total\ number\ of\ trials} \equiv \sum \frac{total\ number\ of\ false\ rejects}{total\ number\ of\ trials}$$

$$FRR = 18.5\%$$

The lower the False reject rate, the higher the efficiency of any biometric system. Additionally, this test also proves the efficacy of the given system. The performance of the Automated Speaker Recognition is summarized in Table 1.

Table 1. General Performance Automated Speaker Recognition

	FAR (%)	FRR (%)
Speaker 1	20	10
Speaker 2	0	20
Speaker 3	30	0
Speaker4	10	20
Speaker 5	50	40
Speaker 6	20	30
Speaker 7	10	10

5. CONCLUSION

This paper has described the design of a voice activated door control system. We have used the MFCCs for feature extraction while the GMM is used for pattern matching. We have also shown that the door control system could easily be assembled using cheap and easily available materials. Analysis of the results using standard performance metrics such as FAR and FRR produced accuracy (genuine acceptance probability) of more than 80%, which is high when compared with existing access control schemes.

REFERENCES

[1] E Dovgan, B Kaluža, T Tušar and M Gams. Agent-based Security System for User Verification. *International Joint Conference on Web Intelligence and Intelligent Agent Technology*. 2009: 331-334.

[2] WA Wahyudi and M Syazilawati. Intelligent Voice-Based Door Access Control System Using Adaptive-Network-based Fuzzy Inference Systems (ANFIS) for Building Security. *Journal of Computer Science*. 2007; 3(5): 274-280.

[3] SY Kung, MW Mak and SH Lin. *Biometric Authentication: Machine Learning Approach*. Prentice Hall. 2004.

[4] S Furui. Recent advances in speaker recognition. *Patter Recognition Letters*. 1997; 18: 859-872.

[5] JP Campbell. “*Speaker Recognition: a Tutorial*”. Proceedings of the IEEE. 1997; 85(9): 1437-1462.
[6] Sirko Molau, Michael Pitz, Ralf Schl¨uter, and Hermann Ney. Computing Mel Frequency Cepstral Coefficients on the Power Spectrum. *ICASSP*. 2001.
[7] D Tran and M Wagner. “*Fuzzy Normalization Methods for Speaker Verification*”. In Proc. ICSLP2000, Beijing, China. 2000; 1: 446-449.
[8] D Tran and M Wagner. “*A Proposed Likelihood Transformation for Speaker Verification*”. In Proc. ICASSP2000, Turkey. 2000; 2: 1069-1072.
[9] JC Bezdek. "*Pattern Recognition with Fuzzy Objective Function Algorithms*". Plenum Press, New York. 1987.
[10] JM Mendel. *Uncertain Rule-based Fuzzy Logic Systems: Introduction and New Directions*. Prentice-Hall, Upper Saddle River, NJ. 2001.