International Journal of Trend in Scientific Research and Development (IJTSRD)
International Open Access Journal | www.ijtsrd.com
ISSN No: 2456 - 6470 | Volume - 2 | Issue – 6 | Sep – Oct 2018

Automatic Attendance System Using Speaker Recognition

Dr. Zaw Win Aung
Technological University (Loikaw), Loikaw, Myanmar

ABSTRACT

The main aim of this paper is to develop automatic attendance system using speaker recognition technique. The proposed system is software architecture which allows the user to access the system by making an utterance from microphone and the attendance of corresponding user is marked in the Microsoft Office Excel. The proposed system automates the whole process of taking attendance. The system uses text dependent open-set speaker identification with MFCC features and vector quantization based speaker modeling for authenticating the user. A simple Euclidean distance scoring is used as the classifier. For decision making, the new approach, mean value threshold is proposed to optimize the system performance. The database consists of 60 speech samples which were collected from 20 speakers consisting of 10 male speakers and 10 female speakers. The experimental study shows that the proposed system with mean value threshold achieves better results in recognition accuracy and verification performance than the system with other threshold setting.

Keyword: *automatic attendance system; speaker recognition; voice biometric; mean value threshold; maximum value threshold*

I. INTRODUCTION

In many institutions and organizations the attendance is a very important factor for various purposes and its one of the important criteria is to follow for students and organization employees. The previous approach in which manually taking and maintains the attendance records was very inconvenient task. As an alternative solution, biometrics technologies can be introduced to construct a more powerful version of attendance system. Biometric is an authentication technique that recognizes unique features in each human being. Some of the commonly used biometric features include voice, face, signature, finger print, handwriting, iris, DNA, Gait, etc. Biometrics techniques are widely used in various areas like building security, forensic science, ATM, criminal identification and passport control [1]. In the proposed automatic attendance system voice biometric is used for obtaining the attendance because each person has unique characteristic in voice that can be captured and analyzed to make the proposed automatic attendance system more efficient and effective. Voice recognition can be divided into two, which are speech recognition and speaker recognition. Both are using voice biometric differently. Speech recognition is the ability to recognize what have been said while speaker recognition is the ability to recognize who is speaking.

Speaker recognition can be classified into identification and verification. Speaker identification is the process of determining which registered speaker provides a given utterance. Speaker verification, on the other hand, is the process of accepting or rejecting the identity claim of a speaker. Speaker identification task is further classified into open- and closed-set tasks. If the target speaker is assumed to be one of the registered speakers, the recognition task is a closed-set problem. If there is a possibility that the target speaker is none of the registered speakers, the task is called an open set problem. In general, the open-set problem is much more challenging. In the closed-set task, the system makes a forced decision simply by choosing the best matching speaker from the speaker database no matter how poor this speaker matches. However, in the case of open-set identification, the system must have a predefined tolerance level so that the similarity degree between the unknown speaker and the best matching speaker is within this tolerance. In this way, the verification task can be seen as a special case of the open-set identification task, with only one speaker in the database [2]. Speaker recognition methods can also be divided into text-

dependent and text-independent methods. In a text-dependent system, the system knows the text spoken by the person. In a text-independent system, on the other hand, the system must be able to recognize the speaker from any text. For the proposed system, text dependent open-set speaker identification in speaker recognition is used.

Matsui et al. proposed a text-prompted speaker recognition method, in which key sentences are completely changed every time the system is used [3]. The system accepts the input utterance only when it determines that the registered speaker uttered the prompted sentence. Because the vocabulary is unlimited, prospective impostors cannot know in advance the sentence they will be prompted to say. This method not only accurately recognizes speakers, but can also reject an utterance whose text differs from the prompted text, even if it is uttered by a registered speaker. Thus, a recorded and played back voice can be correctly rejected.

Research on increasing robustness became a central theme in the 1990s. Matsui et al. [4] compared the VQ-based method with the discrete/continuous ergodic HMM-based method, particularly from the viewpoint of robustness against utterance variations. They found that the continuous ergodic HMM method is far superior to the discrete ergodic HMM method and that the continuous ergodic HMM method is as robust as the VQ based method when enough training data is available. They investigated speaker identification rates using the continuous HMM as a function of the number of states and mixtures. It was shown that speaker recognition rates were strongly correlated with the total number of mixtures, irrespective of the number of states. This means that using information about transitions between different states is ineffective for text-independent speaker recognition and, therefore, GMM achieves almost the same performance as the multiple-state ergodic HMM.

II. SYSTEM ARCHITECTURE

Most speaker recognition systems contain two main modules: feature extraction and feature matching while the proposed system contains three main modules: feature extraction, feature matching and decision making. Mel-frequency Cepstrum Coefficients (MFCC) is applied for feature extraction to extract a small amount of data from the voice signal that can later be used to represent each speaker. For feature matching, Vector Quantization (VQ) approach using Linde, Buzo and Gray (LBG) clustering algorithm is proposed because it can reduce the amount of data and complexity. For decision making, the new approach, mean value threshold is proposed to calculate the speaker specific threshold which improves the recognition accuracy and optimizes the verification performance by reducing False Acceptance Rate (FAR) and False Rejection Rate (FRR). Fig. 1 and Fig. 2 show the training and testing phases of the proposed automatic attendance system.

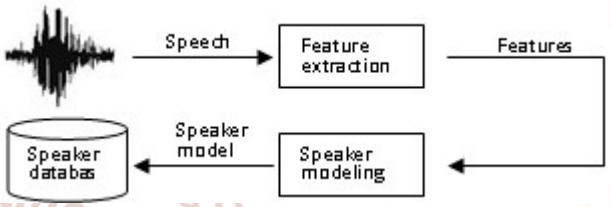

Fig.1. Training phase of automatic attendance system

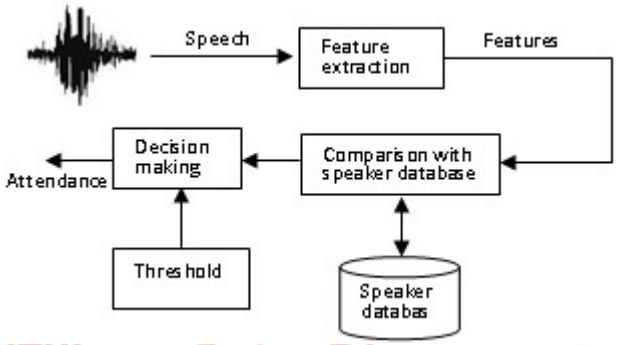

Fig.2. Testing phase of automatic attendance system

A. Decision Making

A central issue in speaker verification is how to make a decision. Essentially, a speaker verification system could make two types of mistakes during decision-making; one is the false acceptance (FA), and the other is the false rejection (FR). An impostor score that falls below the predefined threshold results in a false acceptance, while a genuine score that exceeds the predefined threshold results in a false rejection. A false acceptance is said to occur when an impostor is accepted, while a false rejection occurs when the system rejects a true client. The False Accept Rate (FAR) of a biometric system is the fraction of impostor scores falling below the threshold. Similarly, the False Reject Rate (FRR) of a system is defined as the fraction of genuine scores exceeding the threshold. A substantial task in decision-making is somehow to minimize both FA and FR errors during decision-making.

There have been a few approaches developed for decision-making. These approaches could be roughly classified into two categories: a priori threshold setting [5-8] and a posteriori threshold setting [9-10]. The basic idea of a priori threshold setting is to find a proper threshold somehow based on a training set, and then the resultant threshold will be applied to make a decision during test for any claimed voice token. The a priori threshold setting gives a feasible way to create real speaker verification systems. On the other hand, the a posteriori threshold is often set by finding the threshold of equal error rate (EER) that makes the FA rate equal to the FR rate for a given speaker system. In contrast to the a priori threshold setting, the a posteriori threshold setting provides a way to evaluate the discrimination capabilities of a particular speaker model in terms of a certain data set. Although such a method allows us to compare objectively the modeling performance, it is ultimately unrealistic for a real application. In this paper, the new approaches, mean value threshold and maximum value threshold, are proposed for decision making and the recognition accuracy and the verification performance of the two systems are compared.

B. Computing Mean Value Threshold and Maximum Value Threshold

This section describes the new approaches, mean value threshold and maximum value threshold, to calculate the speaker specific threshold from the training samples. In the training phase of speaker recognition system, each registered speaker has to provide samples of their speech so that the system can build or train reference models for that speaker. To compute mean value threshold and maximum value threshold, it is needed to collect more than one speech sample from each speaker in the training phase to improve recognition rate and to reduce False Acceptance Rate (FAR) and False Rejection Rate (FRR) of the system. Firstly, Euclidean distances between the test speech sample and the reference models of each speaker are calculated. Then, mean value threshold is computed by finding average value of all the distances between the test speech sample and the reference models of each speaker. Mean value threshold is computed as follow:

$$T_{mean} = \frac{1}{n} \times \sum_{i=1}^{n} d_i \qquad (1)$$

Where, T_{mean} is mean value threshold, d_i is distance between test speech sample and reference model of speaker and *n is* number of training speech samples for each speaker.

Maximum value threshold is computed by finding maximum value among the distances between the test speech sample and the reference models of each speaker. Maximum value threshold is computed as follow:

$$T_{max} = max\ (d_1, d_2, d_3,\ldots,d_n) \qquad (2)$$

where, T_{max} is maximum value threshold, $d_1, d_2, d_3,\ldots, d_n$ is distances between test speech sample and reference models of each speaker and *n is* number of training speech samples for each speaker.

III. IMPLEMENTATION

The proposed automatic attendance system is simulated in MATLAB with speech signal as input and produces the identity of speaker as output to mark the attendance of corresponding user in the Microsoft Office Excel. The database consists of 60 speech samples which were collected from 20 speakers consisting of 10 male speakers and 10 female speakers. Each speech sample is about 2 second long. The speaker is asked to utter his/her name three times in a training session and one time in a testing session later on. The same microphone is used for all recordings. Speech signals are sampled at 8000 Hz.

In the training phase, feature vectors are calculated from the input speech signal by MFCC feature extraction algorithm. Finally, the codebook or reference model for each speech signal is constructed from the MFCC feature vectors using LBG clustering algorithm and store it in the database. In the proposed system, in addition, mean value threshold is also computed from the training samples. In the testing phase, the input speech signal is compared with the stored reference models in the database and the distance between them is calculated using Euclidean distance. And then, the minimum distance is selected among the distances between the input speech signal and the stored reference models. If this minimum distance falls below the predefined mean value threshold, the system outputs the speaker ID which has minimum distance as identification result and the attendance of corresponding speaker is marked in the Microsoft Office Excel as verification result and the message "You have successfully marked your attendance." is displayed. Otherwise, the system determines it is none of the registered speakers and the message "Sorry, we are unable to mark your attendance." is displayed.

IV. EXPERIMENTAL RESULT

The purpose of this section is to illustrate the performance of the proposed system with mean value threshold comparing with the system with maximum value threshold. In order to show the effectiveness of the proposed system, the recognition accuracy or attendance marking accuracy and the verification performances, false rejection rate (FRR) and false acceptance rate (FAR), of the two systems were computed and compared. The database consists of 60 speech samples which were collected from 20 speakers consisting of 10 male speakers and 10 female speakers. Forty speech samples which were collected from 20 genuine speakers and 20 impostors were used as test speech samples. Table 1 shows experimental results of the two systems.

Table1. Experimental results of the two systems

No	Threshold setting	No: of Test Samples	Size of Database	Successful Recognition	FR	FA
1	Mean value threshold	40	60	40	0	0
2	Maximum value threshold	40	60	38	0	2

According to experiments, it is found that the attendance marking accuracy of the proposed system is 100 percent while the system with maximum value threshold achieves 95 percent of attendance marking accuracy. When verification performance is taken into account, the false rejection rate (FRR) is 0 percent for both systems and false acceptance rate (FAR) is 0 percent and 5 percent respectively for the proposed system and the system with maximum value threshold. The attendance marking accuracy and verification performance of the two systems are shown in Table 2.

Table2. Attendance marking accuracy and verification performance of the two systems

No.	Threshold setting	No: of Test Samples	Size of Database	Attendance marking accuracy	FRR	FAR
1	Mean value threshold	40	60	100%	0%	0%
2	Maximum value threshold	40	60	95%	0%	5%

V. CONCLUSION

The experimental result shows that using mean value threshold setting helps to optimize system performance. As a result, the recognition accuracy or the attendance marking accuracy of the proposed system is 100% and verification performance, False Acceptance Rate (FAR) is 0%, and False Rejection Rate (FRR) is 0%. Therefore, from this work it can be concluded that the proposed system is reliable to use in real world automatic attendance systems of institutions and organizations.

ACKNOWLEDGMENTS

The author would like to take the opportunity to thank all of his colleagues who have given him support and encouragement during the period of the research work. The author is also very thankful to his students who provided the speech samples for this research work. Finally the author would like to express his indebtedness and deep gratitude to his beloved parents, wife and son for their support and understanding during the research work.

REFERENCES

1. Anil K. Jain, Arun Ross and Salil Prabhakar, “An introduction to biometric recognition,” Circuits and Systems for Video Technology, IEEE Transactions on Volume 14, Issue 1, Jan. 2004 Page(s):4 – 20.
2. Kinnunen, T., "Spectral Features for Automatic Text-Independent Speaker Recognition", Licentiate's Thesis, University of Joensuu, Department of Computer Science, December 21, (2003).
3. T. Matsui and S. Furui, “Concatenated phoneme models for text-variable speaker recognition,” Proc. ICASSP, pp. II-391-394, 1993.
4. T. Matsui and S. Furui, “Comparison of text independent speaker recognition methods using

VQ-distortion and discrete/continuous HMMs", Proc. ICSLP, pp. II-157-160, 1992.

5. F. Bimbot, M. Blomberg, L. Boves, D. Genoud, H.P. Hutter, C. Jaboulet, J. Koolwaaij, J. Lindberg, J.B. Pierrot, "An overview of the CAVE project research activities in speaker verification", Speech Commun. 31 (2000) 1437–1462.
6. S. Furui, "Cepstral analysis technique for automatic verification", IEEE Trans. Acoustics Speech Signal Process. 29 (1981) 254–272.
7. J. B. Pierrot, J. Lindberg, J. Koolwaaij, H. P. Hutter, D. Genoud, M. Bolmberg, F. Bimbot, "A comparison of a priori threshold setting procedures for speaker verification in the CAVE project", Proceedings of the ICASSP, 1998, pp. 331–334.
8. C. Fredouille, J. Mariethoz, C. Jaboulet, J. Hennebert, J.F. Bonastre, C. Mokbel, F. Bimbot, "Behavior of a Bayesian adaptation method for incremental enrollment in speaker verification", Proceedings of the ICASSP, 2000, pp. 1197–1200.
9. A. E. Rosenberg, "Automatic speaker verification: review". Proc. IEEE 64 (1976) 475–487.
10. J. M. Naik, "Speaker verification—a tutorial", IEEE Commun. Mag. 28 (1990) 42–48.