Speaker Identification Based on Voice Samples using MFCC and GMM

Avinash Kumar[1] Ankit Raj[2] Avantika Shee[3] Kundan Kumar[4] Nagarathna R[5]
[1,2,3,4]Student [5]Assistant Professor
[1,2,3,4,5]Department of Telecommunication Engineering
[1,2,3,4,5]Dayananda Sagar College of Engineering, Bangalore

Abstract— Speaker Identification is the task of claiming a speaker's identity based on the characteristics contained in the voice signal. The identification is done by dividing the task in two phases named Feature Extraction and Feature Identification. Various methods are available for extraction of a voice signal such as Mel Frequency Cepstral Coefficients (MFCCs), Linear Frequency Cepstral Coefficients (LFCCs), Linear Predictive Cepstral Coefficients (LPCC), Perceptual Linear Prediction Cepstral Coefficients (PLPCCs) etc. Similarly, modern methods available for identification include Vector Quantization (VQ), Hidden Markov Model (HMM), Gaussian Mixture Models (GMM) etc.In this paper, based on our survey on the grounds of noise reduction, computational time, accuracy and efficiency, we propose that an MFCC-GMM model is best suited for identifying a speaker.

Key words: Speaker Identification, Feature Extraction, Feature Identification, Mel-Frequency Cepstral Coefficients (MFCC), Gaussian Mixture Model (GMM).

I. INTRODUCTION

Speech is the primary communication medium between people. This communication process has a complex structure consisting not only of the transmission of voice but also include gestures, the language, the subject and the capability of the listener that contributes to this process.

Therefore from the last five decades people have come forward to investigate various aspects of speech such as mechanical realization of speech signal, human machine interaction and speech recognition. The staggering rate at which the use of personal electronic devices such as cell phones and PDA's is increasing has motivated the scientific community to evolve human like interfaces. This necessitates the demand of an accurate speech recognition system which can be possibly language and speaker independent, for command and query, used for separation of speakers and their authentication and for text to speech conversion of human quality.

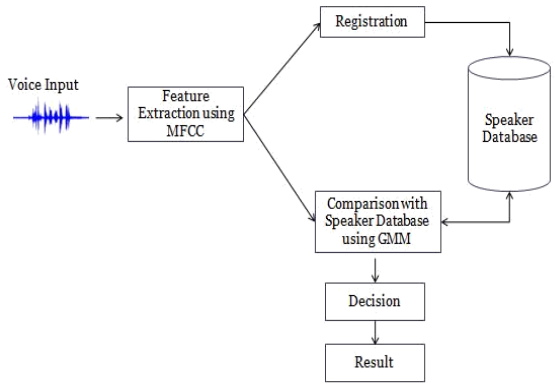

Fig. 1: Block Diagram For a Simple Speaker Identification System

One of the main objectives of speech recognition lies in the idea of communicating with a machine using voice signals. Though there already exists various means of communicating with machines, like with keyboard, images etc. but they are all slow, clumsy or complex. Hence speech plays a very vital role in making this communication easier. The ultimate goal of the Automatic Speech Recognition (ASR) is to build systems which can comfortably respond and act upon the human spoken language/instruction.

The idea of a Speaker Identification system is to match an input voice signal with the preloaded database of known voice signals. This is achieved in two stages:

(1) Speaker Enrollment
(2) Speaker Verification

Speaker Enrollment: In this stage, the voice samples from various speakers are taken and a database of these smaples is maintained. Feature extraction of each voice signal is done using MFCC.

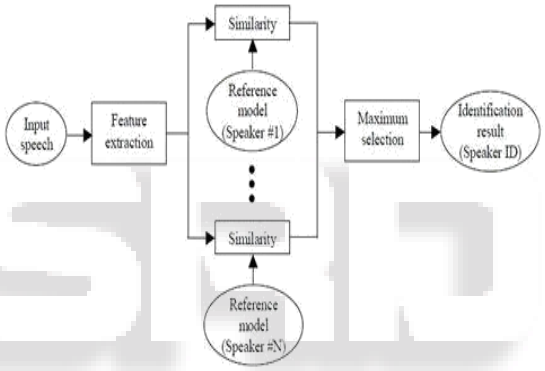

Fig. 2: Speaker Enrollment

Speaker Verification: In this stage, a test voice is taken and compared with all the samples of the database. The technique used is GMM which helps to return the best matched output.

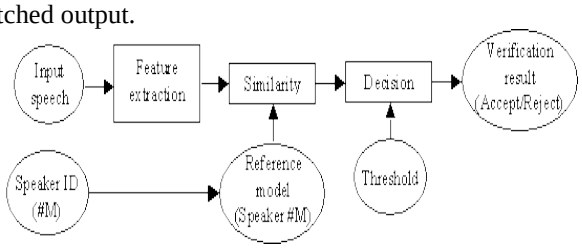

Fig. 3: Speaker Verification

The Speaker Recognition System is divided in two categories:

(1) Text-Dependent

In this category, the sample in the enrollment and the verification phase must be same. The results of a Text-Dependent system are very accurate and can be applied for authentication purposes, voice passwords etc.

(2) Text-Independent

In this category, the sample in the enrollment and the verification phase is different. The results are not that accurate as compared to the previous system but still satisfactory results can be achieved. This system can be applied in forensic experiments.

All rights reserved by www.ijsrd.com

II. SIGNIFICANCE

With the increasing technology speaker recognition finds its application in many fields. It can be used in access control like a Home Automation System for handling the devices using voice commands. It can be applied for transaction authentication where actions such as telephone credit card purchases, bank wire transfers and fraud detection can be made simpler. Even forensics can have a Speaker Identification system for voice sample matching.

III. RELATED WORK

Shupeng Xu, Yan Liu and Xiping Liu [2] proposed in their paper a method for speaker recognition and speech emotion recognition based on GMM. The authors stated the reason of using this model was its ability to the dependency among extracted speech signal feature vectors and the multimodality in their distribution. Firstly, they extracted the Mel-Frequency Cepstral Coefficients (MFCC) from each frame of the speech signal as speech features, and then applied Gaussian mixture model as a statistical classifier. The authors achieved encouraging results with the proposed system.

Varsha Singh, Vinay Kumar Jain and Dr. Neeta Tripathi [5] in their paper presented a brief survey of feature extraction techniques used in Language Identification (LID) system. The objective of the Language Identification system was to automatically identify the specific language from a spoken utterance. The extraction of the features of acoustic signals is an important task because LID mainly depends on the language-specific characteristics. The performance and quality of the system is strongly affected by the feature extraction phase. Out of various methods available, the authors proposed MFCC technique is as one of the standard methods for feature extraction and accepted it as the baseline. MFCC are the results of the short-term energy spectrum expressed on a Mel-frequency scale. The MFCCs are proved more efficient better anti-noise ability than other vocal tract parameters.

Alfredo Maesa, Fabio Garzia, Michele Scarpiniti and Roberto Cusani (4) focused on showing time results of a text independent Automatic Speaker Recognition (ASR) system. They also intended to show how much accuracy can be achieved with such system. They developed a security control access gate and the system they designed was based on Mel-Frequency Cepstrum Coefficients (MFCC) and Gaussian Mixture Models (GMM). In the paper 450 speakers were randomly extracted from the *Voxforge.org* audio data-base, to improve their utterances a technique named spectral subtraction was used, then MFCC were extracted and these coefficients were statistically analyzed by GMM in order to build each profile. The proposed system when implemented in Matlab language took only 2 seconds for a single test run on a common PC and the approach recorded encouraging results with 96% accuracy.

IV. PROPOSED SYSTEM

We propose a Speaker Identification System which will automate the access of home appliances through voice commands. The voice commands of all the residents of a house to switch on or switch off various electrical devices will be stored in the system. When an instruction will be given in form of voice commands, the system will first check in the database whether the instruction is from an authorized user and based on the result in this stage it will take proper action. If the voice command matches any one of the already stored samples, the device will be turned on or off. Otherwise, no action will be taken.

We will implement MFCC for feature extraction and GMM will be used for feature identification. Using these algorithms we intend to develop a Home Automation System. This is a basic model with the scope of addition of other valuable features.

V. CONCLUSION

It's quite apparent that for a Speaker Identification System based on voice samples, MFCC-GMM is the best system available for feature extraction and feature identification. MFCC has least false acceptance rate and zero false rejection rate which makes it best suited for feature extraction. Other methods such as LPCC and BFCC shouldn't be used because they have high false rejection rate. In identifying a speech signal, out of the techniques such as VQ, GMM and HMM, GMM emerges out to be the most efficient and accurate method. Its results are very satisfactory even in the case of corrupted or unconstrained speech.

REFERENCE

[1] Vibha Tiwari, Dept. of Electronics Engineering, Gyan Ganga Institute of Technology and Management, Bhopal, (MP), INDIA, "MFCC and its applications in speaker recognition", International Journal on Emerging Technologies 1(1): 19-22(2010) ISSN : 0975-8364.

[2] Shupeng Xu, Yan Liu and Xiping Liu, Computer Science and Engineering Department, Changchun University of Technology, Changchun, China, "Speaker Recognition and Speech Emotion Recognition Based on GMM", 3rd International Conference on Electric and Electronics (EEIC 2013).

[3] Douglas A. Reynolds and Richard C. Rose, Member, IEEE, "Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models", IEEE Transactions on speech and audio processing, Vol. 3, No. 1, January 1995.

[4] Alfredo Maesa[1], Fabio Garzia[1,2], Michele Scarpiniti[1], Roberto Cusani[1], [1.]Dept. of Information, Electronics and Telecommunication Engineering, University of Rome, Rome, Italy, [2.] Wessex Institute of Technology, Southampton, UK, "Text-Independent Automatic Speaker Recognition System Using Mel-Frequency Cepstrum Coefficient and Gaussian Mixture Models", Journal of Information Security, October 2012.

[5] Varsha Singh[1], Vinay Kumar Jain, Dr. Neeta Tripathi[2], [1]Department of Electronics & Telecommunication, CSVTU University, [2]SSITM, CSVTU University, FET, SSGI, SSTC

All rights reserved by *www.ijsrd.com*

Jumwani, Bhilai, C. G, India, "A Comparative Study on Feature Extraction Techniques for Language Identification ", International Journal of Engineering Research and General Science Volume 2, Issue 3, April-May 2014.

[6] Tridibesh Dutta, Indian Statistical Institute, Kolkata, India, "Text Dependent Speaker Identification based on Spectrograms", Proceedings of Image and Vision Computing New Zealand 2007, pp. 238-243, Hamilton, New Zealand, December 2007.

[7] Ms. Arundhati S. Mehendale and Mrs. M.R.Dixit, Department of Electronics, K.I.T.'s College of Engineering, Kolhapur, "Speaker Identification", Signal & Image Processing: An International Journal (SIPIJ) Vol.2, No.2, June 2011.

[8] S. Selva Nidhyananthan, R. Shanntha Selva Kumari, Dept. of Electronics and Communication Engineering, Mepco Schlenk Engineering College, Virudhunagar District, Tamil Nadu, India, "Language and Text-Independent Speaker Identification System Using GMM", WSEAS Transactions on Signal Processing, Issue 4, Volume 9, October 2013.

All rights reserved by *www.ijsrd.com*