‘Futurizm 14’, (14th & 15th March 2014), ‘Pune District Education Association’s College of Engg.Manjari (BK)’, ‘Pune 412307’,Chief Editor:‘Dr.K.R.Harane’,Editor: ‘Prof D.O.Patil’, ‘Prof R.M.Sahu’.

International Journal of Innovative Technology & Adaptive Management (IJITAM)
ISSN: 2347-3622, Volume-1, Issue-6 , March, 2014

Improved Performance Based Method for Text Independent SpeakerIdentification

Mrs.Vishakha V. Jadhav[1], Prof. Vijay M. Sardar[2]
[1]PG student, Department of Electronics and Telecomm Engg., JSCOE, Hadapsar, Pune University, Pune, India.
[2]Head,Department of Electronics andTelecomm Engg., JSCOE, Hadapsar, Pune University, Pune, India.
[1]sanvishakha@gmail.com, [2]vijay.sardar11@gmail.com

***Abstract*—Speaker identification is the computing task of recognizing speaker's identity based on their voices. The classification of speech depends on the extraction of several key features like Mel Frequency Cepstral Coefficients (MFCC) from the speech signals of speaker.A unique identity for each person who has enrolled for speaker identification can be built using a statistical model like Gaussian Mixture Model (GMM) and features extracted from the speech signals. Using Vector Quantization (VQ) technique, a decision function is proposed to decrease the training model for GMM in order to reduce the processing time. In the proposed modeling, the superiority of VQ is takento differentiate the male and female speaker .Then, GMM is applied into the subgroup of speaker to get the accuracy rates.**

***Keywords*—GMM, MFCC, VQ.**

I. INTRODUCTION

Automatic speaker recognition is the process of recognizing person from a spoken phrase. This system operates in two modes: to identify a particular person or to verify a person’s claimed identity [1]. Speaker identification applications aim to determine which registered speaker provides a given utterance from a set of known speakers [1].

The various feature matching approaches in speaker recognition are, Dynamic Time Warping (DTW), Vector Quantization (VQ), Hidden Markov Models (HMM), Gaussian mixture model (GMM), Support Vector Machine (SVM). Among these pattern classification approach, the use of GMM are most common due to it can be performed in a completely text independent situation [7].

The process of speaker identification is divided into two main phases. During the first phase, speaker enrollment, speech samples are collected from the speakers, and they are used to train their models. In the second phase, identification phase, a test sample from an unknown speaker is compared against the speaker database.

Mel Frequency Cepstral Coefficients (MFCC)

MFCC are the coefficients obtained in the MFC representation. The Mel Frequency Cepstrum (MFC) can be defined as the short time power spectrum of speech signal which is calculated as the linear cosine transform of the log power spectrum on a non –linear mel scale of frequency. The block diagram of MFCC is as shown in Fig. 1.

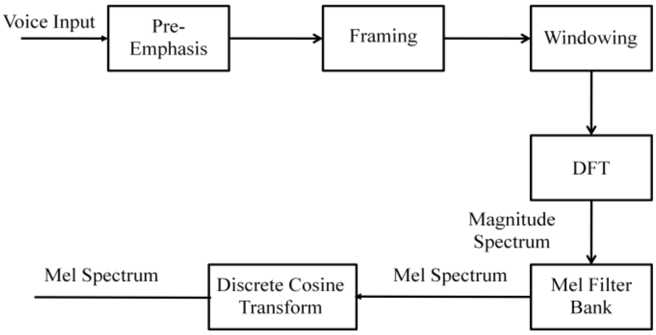

Fig.1. Block diagram of MFCC

Classification Methods

There are two major types of models for classification: stochastic models and template models.

In stochastic models, the pattern matching is probabilistic and results in a measure of the likelihood, or conditional probability, of the observation given the model. Here, a certain type of distribution is fitted to the training data by searching the parameters of the distribution that maximize some criterion.

It includes: Gaussian Mixture Model (GMM), Hidden Markov Model (HMM) and Artificial Neural Network (ANN), also linear classifier.

For template models, the pattern matching is deterministic. This approach makes minimal assumptions about the distribution of the features.Template models are considered to be the simplest ones. It includes: Dynamic Time Warping (DTW) and Vector Quantization (VQ) models.

Feature Matching

The speech produced by the speaker whose identity is to be recognized, will be compared with all speaker’s models in the database. Then, the speaker identity will be determined.

Zero Crossing Rate

ZCR is one of the methods used for taking a voiced or unvoiced(V/UV) decision.

Zero Crossing Rate gives information about the number of zero-crossings present in a given signal. Intuitively, if the number of zero crossing are more in a given signal, then the signal is changing rapidly and accordingly the signal may contain the high frequency information on the similar lines, if the number of zero crossing are less, hence the signal is changing slowly and accordingly the signal may contain low frequency information. Thus ZCR gives indirect information about the frequency content of the signal. [9]

The ZCR in case of stationary signal is defined as,

$$z = \sum_{n=-\infty}^{\infty} |sgn(s\ (n)) - sgn(s\ (n\text{-}1))|$$

$$\text{Where } sgn(s\ (n)) = 1 \text{ if } s\ (n) \geq 0$$

$$= -1 \text{ if } s\ (n) < 0$$

The zero crossing count is an indicator of the frequency at which the energy is concentrated in the signal spectrum.

II. Literature Survey

Douglas Reynolds, et. al. [2] has proposed the use of Gaussian Mixture Models (GMM) for robust text independent speaker identification. The individual Gaussian Components of a GMM are shown to represent some general speaker-dependent spectral shapes that are effective for modelling speaker identity. The Gaussian mixture speaker model was specifically evaluated for identification tasks using short duration utterances from unconstrained conversational speech, possibly transmitted over noisy telephone channels. The component Gaussians were first shown to represent characteristic spectral shapes (vocal tract configurations) from the phonetic sounds which comprise a person’s voice. By modeling the underlying acoustic classes, the speaker model is better able to model the short-term variations of a person’s voice, allowing high identification performance for short utterances. The Gaussian mixture speaker model was also interpreted as a nonparametric, multivariate pdfmodel, capable of modeling arbitrary feature distributions.

R. Saeidi et al. [4] introduced a Gaussian mixture model (GMM) classifier, which is called as, GMM identifier, as an efficient post processing method to enhance the performance of a GMM based speaker verification system; such as Gaussian Mixture Model Universal Background Model (GMM-UBM) and Structural Gaussian Mixture Models with Structural Background Model (SGMM-SBM) speaker verification schemes. The proposed classifier shows good performance while its computational load is almost negligible compared to the main GMM system. Experimental results show the superior performance of this post-processing method in comparison with a neural-network post-processor for such applications.

PoonamBansal, et. al. [3] proposed an automatic speaker identification scheme to identify or verify a person, by identifying his/her voice. All speaker identification systems contain two main phases, training phase and testing phase. In the training phase the features of the words spoken by various speakers are extracted and the feature matching takes place during testing phase.The raw speech signal is transformed into a compact but effective representation that is more stable and discriminative than the original signal by Feature extractor. The feature thus extracted is stored in the database. During the recognition phase the extracted features are compared with the template in the database. In the proposed Speaker Identifier (SI) the features extracted are LPCC, Mel-Frequency Cepstrum coefficients (MFCC), delta MFCC (DMFCC) and Delta-Delta MFCC (DDMFCC).Vector Quantization (VQ) is used for speaker modeling process. The final recognition decision is made based on the matching score: Speaker model with the smallest matching score is selected as a speaker of the test speech sample. Speaker identification rate was observed to be 96.59% in text independent case and increases by 3.5% in reference to text dependent, as the feature vector size is increased to 36 by including 12 DMFCC and 12 DDMFCC recognition rate gets increased by 0.4%. Better performances could be seen when applying this approach itself or mixed with Hidden

Markov Model (HMM) in isolated-word speech recognition.

ManjotKaur Gill, et. al. [5] proposed the feature extraction by using MFCC (Mel Frequency Cepstral Coefficients). The speaker was modeled using Vector Quantization (VQ). A VQ codebook is generated by clustering the training feature vectors of each speaker and then stored in the speaker database. In this method, the K-means algorithm was used for clustering purpose. In the recognition stage, a distortion measure which based on the minimizing the Euclidean distance was used when matching an unknown speaker with the speaker database. VQ based clustering approach is best as it provides with the faster speaker identification process.

J. Pelecanos,et.al [6] proposed the use of a Vector Quantization Gaussian (VQG) as a more efficient alternative to the standard Gaussian Mixture Model for relatively well-clustered data. The VQG was more robust to mismatched speaker recognition conditions for the multi-background speaker system. This was possibly attributed to the method of estimation of the VQG mixture means, weights and variances. The VQG method provides a rapid means of training and testing to form a reliable and efficient speaker verification system.

III. 3. PROPOSED WORK

In baseline form, the VQ –based solution is less accurate than GMM, but it offers simplicity in computation. In baseline form, the GMM-based solution is more accurate, but results long time processing. To identify the speaker in large database, the accuracy and reduction in processing time is very important. So combining the advantages of VQ and GMM, we propose new hybrid method to identify a speaker by combining VQ as a decision and GMM as a Model to improve the system performance in terms of accuracy and time.

The overall structure of Speaker Identification System using VQ decision and GMM model is as shown in Fig.2. After MFCC feature extraction process, the speech signal will transform to a feature vector form. For the first stage of the classification, VQ classifier clusters the speaker model into two subgroups which is subgroup I (Female) and subgroup II (Male).

In next stage GMM is used within individual subgroup to find the desired speaker. GMM process will be applied in the particular subgroup to identify the speaker. The GMM classification engine will calculate log likelihood score for subgroup training speaker data and save it into a speaker model. While in testing phase, a comparison between training speaker and testing speaker will be done.

Vector Quantization

Vector Quantization (VQ) is a quantization technique used to compress the information and manipulate the data such in a way to maintain the most prominent characteristics. It works by dividing a large set of points into groups having approximately the same number of points closest to them. Each group is represented by its centroid.

LBG design algorithm

The LBG VQ design algorithm is an iterative algorithm. This algorithm requires an initial codebook. The initial codebook is obtained by the splitting method. In this method, an initial code vector is set as the average of the entire training sequence. This code vector is then split into two. The iterative algorithm is run with these two vectors as the initial codebook. The final two code vectors are split into four and the process is repeated until the desired number of code vectors is obtained. The algorithm is summarized in the flowchart of Fig.3.model using some statistical model like GMM statistical model.

Expectation Maximization (EM)

The Expectation-maximization algorithm can be used to compute the parameters of a parametric mixture model distribution. It is an iterative algorithm having two steps: an expectation step and a maximization step [2].

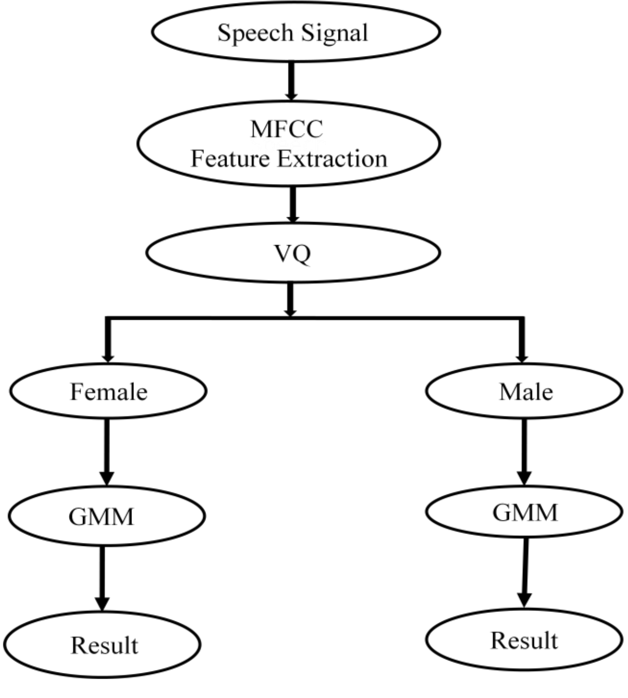

Fig. 2. Speaker Identification System

Gaussian Mixture Model

After extracting features we need to create a speaker

*a) The expectation step:*With initial guesses for the parameters of our mixture model, "partial membership" of each data point in each constituent distribution is computed by calculating expectation values for the membership variables of each data point. That is, for each data point x_j and distribution y_i, the membership value $y_{i,j}$ is:

$$y_{i,j} = \frac{a_i f_Y(x_j, \theta_i)}{f_X(x_j)}$$

*b) The maximization step:*With expectation values in hand for group membership, plug-in estimates are recomputed for the distribution parameters. The mixing coefficients a_i are the means of the membership values over the N data points.

$$a_i = \frac{1}{N} \sum_{j=1}^{N} y_{i,j}$$

Speech enhancement

The background noise is the most common factor degrading the quality of speech signal. The noise reduction is used to reduce the noise level without affecting the speech signal quality.

Speech enhancement aims to improve speech quality by using various algorithms. The objective of enhancement is improvement in intelligibility and/or overall perceptual quality of degraded speech signal using audio signal processing techniques.

Enhancing of speech degraded by noise, or noise reduction, is the most important field of speech enhancement, and used for many applications such as mobile phones, teleconferencing systems, speech recognition, and hearing aids.

The algorithms of speech enhancement for noise reduction can be categorized into three fundamental classes: filtering techniques, spectral restoration, and model-based methods.

- Filtering Techniques
- Spectral Subtraction Method
- Wiener Filtering
- Signal subspace approach (SSA)
- Spectral Restoration
- Minimum Mean-Square-Error Short-Time Spectral Amplitude Estimator (MMSE-STSA)
- Speech-Model-Based

Here we propose the Spectral Subtraction Method for noise reduction. The spectral subtraction method is a simple and effective method of noise reduction. In this method, an average signal spectrum and average noise spectrum are estimated in parts of the recording and subtracted from each other, so that average signal-to-noise ratio (SNR) is improved. The output of this method for a signal is as shown in figure.4.

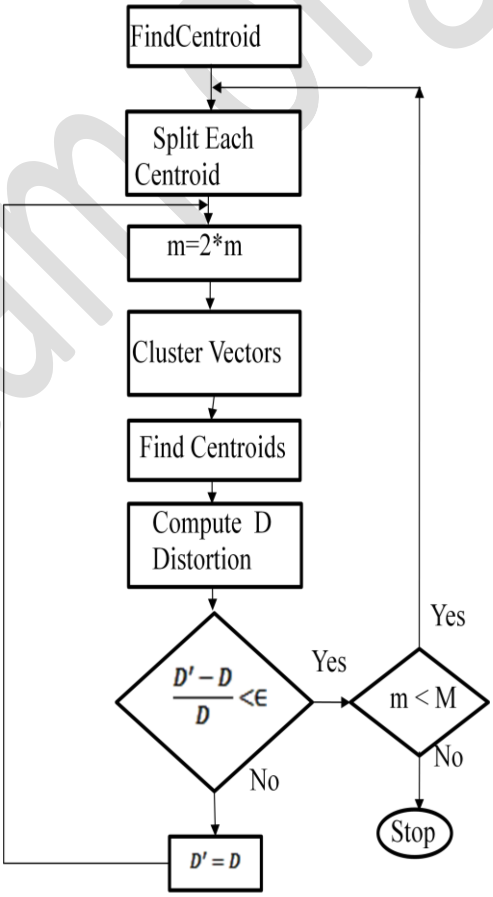

Fig. 3. Flowchart of VQ-LBG algorithm

IV. RESULTS

A database is prepared by recording voice signal. Following steps are applied onthe database.

V. Conclusion

We have proposed hybrid VQ/GMM speaker identification system. We are intended to improve the computation andaccuracy of the speaker identification system by our method.

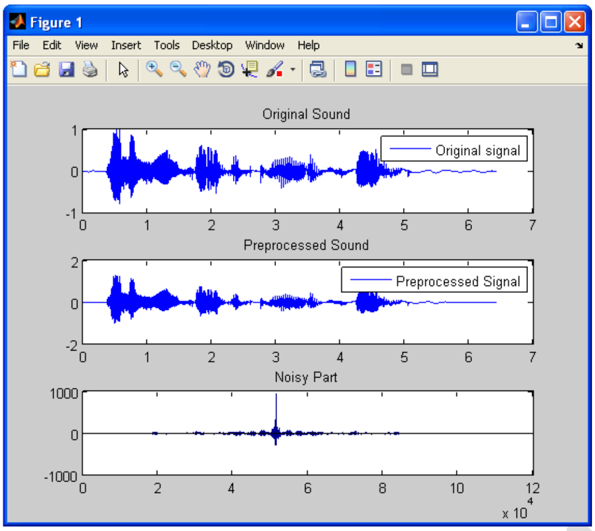

Fig. 4. Separation of Noisy Part

1. Following stages of MFCC are implemented:-
   - Pre–emphasis
   - Framing
   - Hamming windowing
   - Fast Fourier Transform
   - Mel Filter Bank Processing
   - Discrete Cosine Transform

   The output for above stages is as shown in figure 5.

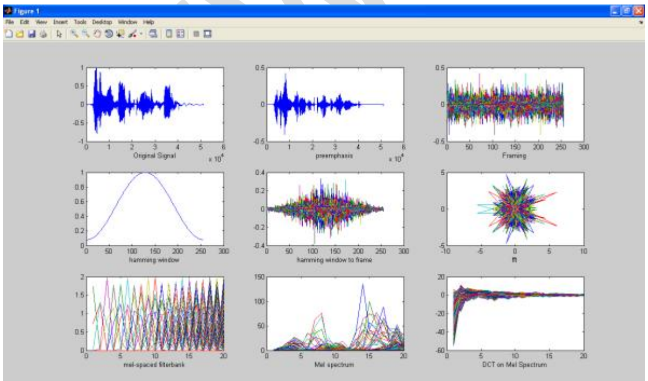

Fig.5. Output of MFCC

References

[1] Campbell, J.P., "Speaker Recognition: A Tutorial", Proc. of the IEEE, vol. 85, no. 9, 1997, pp. 1437-1462.

[2] Reynolds, D. A. and Rose, R. C. "Robust text-independent speaker identification using Gaussian mixture speaker models. IEEE Trans.Speech Audio Process. 3, 1995, pp 72–83.

[3] PoonamBansal, AmitaDev, ShailBala Jain, “ Automatic speaker identification using vector Quantization, “ Asian Journal of Information Technology pp938-942, 2007.

[4] R. Saeidi, H. R. SadeghMohammadi, M. KhalajAmirhosseini, "An efficient GMM classification post-processing method for structural Gaussian mixture model based speaker verification," to be Presented at ICASSP’06, Toulouse, France, May 2006

[5] ManjotKaur Gill, ReetkamalKaur, JagdevKaur, “Vector Quantization based Speaker Identification”, International Journal of Computer Applications (0975 – 8887) Volume 4 – No.2, July 2010.

[6] J. Pelecanos, S. Myers, S. Sridharan and V. Chandran “Vector Quantization based Gaussian Modeling for Speaker Verification”.

[7] Abdul Manan Ahmad and LohMun Yee, “Vector Quantization Decision Function for Gaussian Mixture Model Based Speaker Identification”, 2008 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS2008), Bangkok, Thailand.

[8] Dr.Shaila. D. Apte, “Speech and audio Processing”, Wiley India.