International Journal of Advances in Applied Sciences (IJAAS)
Vol. 4, No. 4, December 2015, pp. 151~156
ISSN: 2252-8814

Speech Recognition Using MFCC and VQLBG

M. Suman, K. Harish, K. Manoj Kumar, S. Samrajyam
Electronics and Computers Dept., K L University, Vijayawada, India

Article Info

Article history:

Received Sep 18, 2015
Revised Nov 14, 2015
Accepted Nov 27, 2015

Keyword:

Euclidean distance
Feature extraction
Feature matching
MFCC
Signal
Vector quantization

ABSTRACT

Speaker Recognition is the computing task of confirmatory a user's claimed identity mistreatment characteristics extracted from their voices. This technique is one of the most helpful and in style biometric recognition techniques in the world particularly connected to areas in that security could be a major concern. It are often used for authentication, police work, rhetorical speaker recognition and variety of connected activities. The method of Speaker recognition consists of two modules particularly feature extraction and have matching. Feature extraction is that the method during which we have a tendency to extract a tiny low quantity of knowledge from the voice signal that will later be used to represent every speaker. Feature matching involves identification of the unknown speaker by scrutiny the extracted options from his/her voice input with those from a collection of identified speakers. Our projected work consists of truncating a recorded voice signal, framing it, passing it through a window perform, conniving the Short Term FFT, extracting its options and Matching it with a hold on guide. Cepstral constant Calculation and Mel frequency Cepstral Coefficients (MFCC) area unit applied for feature extraction purpose. VQLBG (Vector Quantization via Linde-Buzo-Gray) algorithmic rule is used for generating guide and feature matching purpose.

Corresponding Author:

M. Suman,
Electronics and Computers Dept.,
K L University,
Vijayawada, India.
Email: suman.maloji@kluniversity.in

1. INTRODUCTION

Speech process is one in every of most significant branches in digital signal process. Speech signals are often used for speech recognition, speaker recognition or voice command recognition systems. The task of recognition is to see the identity of a speaker. To acknowledge voice, the voices should be acquainted just in case of personalities still as machines. The second element of recognition is testing, particularly the task of scrutiny AN unidentified auditory communication to the coaching knowledge and creating the identification.

Depending upon the appliance the realm of speaker recognition is split into 2 components. One is identification and different is verification. In recognition there are a unit 2 sorts, one is text dependent and another is text freelance. Recognition is split into 2 components: feature extraction and have classification. In recognition the speaker are often known by his voice, wherever just in case of speaker verification the speaker is verified mistreatment info.

The main purpose to grasp concerning speech is that the sounds generated by a person's area unit filtered by the form of the vocal tract together with tongue, teeth etc. This form determines what sound comes out. If we will confirm the form accurately, this could provide North American nation a correct illustration of the sound being created. The form of the vocal tract manifests itself within the envelope of the short time power spectrum, and therefore the job of MFCCs is to accurately represent this envelope.

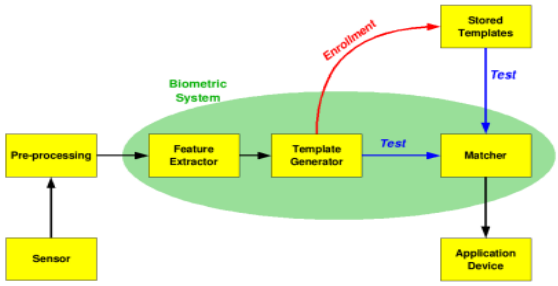
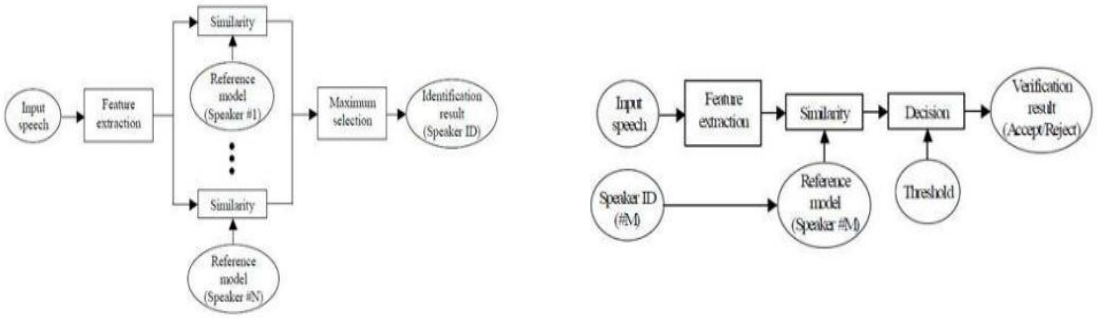


Figure 1. Basic Block Diagram of a Biometric System

2. IDENTIFICATION VS VERIFICATION

This class of classification is that the most significant among the heap. Automatic recognition and verification area unit usually thought-about to be the most natural and economical strategies for avoiding unauthorized access to physical locations or pc systems.



(a) Speaker identification (b) Speaker Verification

Our paper is on recognition. Each the figures represent the ASI (automatic speaker recognition) systems. The on top of 2 area unit the block diagrams of each the processes whereas figure a pair of represent the sensible implementation of the systems.

2.1. Practical Examples of Indentification and Verification System

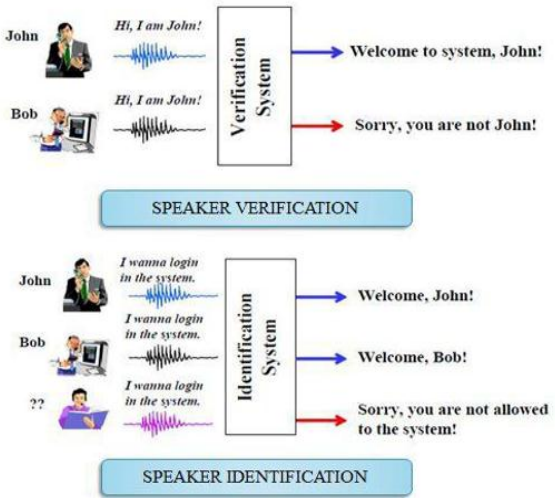


Figure 2. Practical examples of identification and verification systems

2.2. MODULES

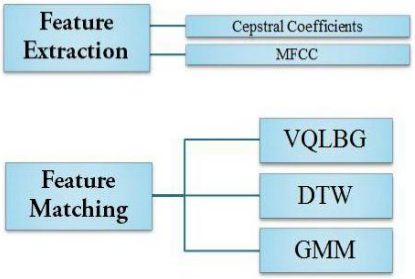


Figure 3. Module MFCC and VQLBG

We are using MFCC and VQLBG for feature extraction and feature matching purpose.

3. SPEAKER IDENTIFICATION

The main aim of this project is recognition that consists of scrutiny a speech signal from AN unknown speaker to an info of notable speaker. The system will acknowledge the speaker that has been trained with variety of speakers. Figure half dozen shows the elemental formation of recognition and verification systems. Wherever the recognition is that the method of crucial that registered speaker provides a given speech. On the opposite hand, speaker verification is that the method of rejecting or acceptive the identity claim of speaker. In many applications, voice is used because the key to substantiate the identities of a speaker area unit classified as speaker verification.

3.1. MFCC (Mel Frequency Cepstral Coefficients)

1. Frame the signal into short frames.
2. For each frame calculate the period gram estimate of the power spectrum.
3. Apply the Mel filter bank to the power spectrum and sum the energy in each filter.
4. Take the logarithm of all filter bank energies.
5. Take the DCT of the log filter bank energies.
6. Keep DCT coefficients 2-13, discard the rest.

But notice that only 12 of the 26 DCT coefficients are kept. This is because higher DCT coefficients represent fast changes in the filter bank energies and it turns out these fast changes actually degrade ASR performance. So we get a small improvement by degrading them.

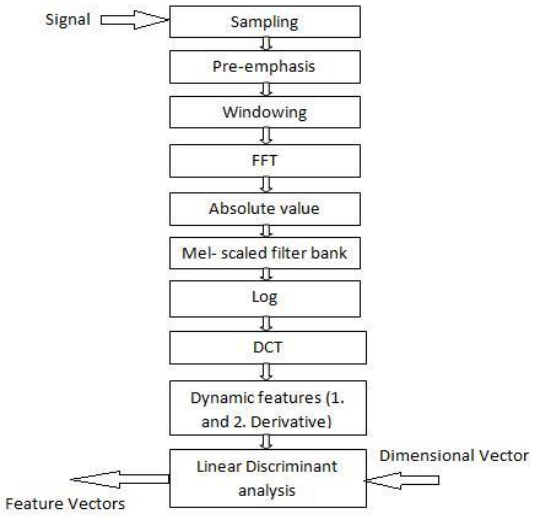


Figure 4. Pipeline of MFCC

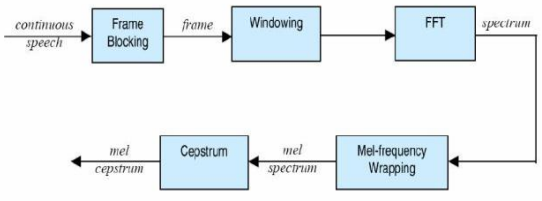


Figure 5. Block diagram of MFCC

The Mel-frequency Cepstrum constant (MFCC) technique is usually accustomed produce the fingerprint of the sound files. The MFCC square measure supported the well-known variation of the human ear's vital information measure frequencies with filters spaced linearly at low frequencies and logarithmically at high frequencies accustomed capture the necessary characteristics of speech. Studies have shown that human perception of the frequency contents of sounds for speech signals doesn't follow a linear scale. So for every tone with AN actual frequency, f, measured in cps, a subjective pitch is measured on a scale referred to as the Mel scale. The Mel-frequency scale is linear frequency spacing below a thousand cps and a power spacing higher than a thousand cps. As a point of reference, the pitch of a one kilohertz tone, forty decibel higher than the sensory activity hearing threshold, is outlined as a thousand Mel's.

The following formula is employed to calculate the Mel's for a specific frequency: Mel (f) = 2595*log10 (1+ f / 700). A diagram of the MFCC processes is shown in Figure 4.

The speech wave form is cropped to get rid of silence or acoustic interference which will be gift within the starting or finish of the sound file. The windowing block minimizes the discontinuities of the signal by tapering the start and finish of every frame to zero. The FFT block converts every frame from the time domain to the frequency domain. Within the Mel-frequency wrapping block, the signal is planned against the Mel spectrum to mimic human hearing. Within the final step, the Cepstrum, the Mel-spectrum scale is regenerate back to straightforward frequency scale. This spectrum provides a decent illustration of the spectral properties of the signal that is vital for representing and recognizing characteristics of the speaker.

After the fingerprint is formed, we are going to additionally stated as AN acoustic vector. This vector are keep as a reference within the information. once AN unknown sound file is foreign into Mat research lab, a fingerprint are created of it additionally and its resultant vector are compared against those within the information, once more mistreatment the geometrician distance technique, and an acceptable match are determined. This method is as stated as feature matching.

3.2. Vector Quantization

A speaker recognition system should able to estimate chance distributions of the computed feature vectors. Storing each single vector that generate from the coaching mode is not possible, since these distributions square measure outlined over a high-dimensional area. It's usually easier to begin by quantizing every feature vector to at least one of a comparatively tiny variety of model vectors, with a method referred to as vector quantization. VQ may be a method of taking an oversized set of feature vectors and manufacturing a smaller set of live vectors that represents the centroids of the distribution.

The technique of VQ consists of extracting little variety of representative feature vectors as AN economical means that of characterizing the speaker specific options. By means that of VQ, storing each single vector that we have a tendency to generate from the coaching is not possible.

By mistreatment these coaching knowledge options square measure clustered to create a codebook for every speaker. Within the recognition stage, the info from the tested speaker is compared to the codebook of every speaker and live the distinction. These variations square measure then use to form the popularity call.

3.3. K-Means Algorithm

The K-means formula may be a thanks to cluster the coaching vectors to urge feature vectors. During this formula clustered the vectors supported attributes into k partitions. It use the k means that of knowledge generated from mathematician distributions to cluster the vectors. The target of the k-means is to attenuate total intra-cluster variance, V.

The process of k-means formula used least-squares partitioning methodology to divide the input vectors into k initial sets. It then calculates the mean purpose, or center of mass, of every set. It constructs a brand new partition by associating every purpose with the highest center of mass. Then the centroids square measure recalculated for the new clusters, and formula recurrent till once the vectors now not switch clusters or instead centroids aren't any longer modified.

3.4. Euclidean Distance

In the speaker recognition section, AN unknown speaker's voice is diagrammatic by a sequence of feature vector then it's compared with the codebooks from the information. So as to spot the unknown speaker, this could be done by activity the distortion distance of 2 vector sets supported minimizing the Euclidean distance.

The Euclidean distance is that the "ordinary" distance between the 2 points that one would live with a ruler, which may be established by recurrent application of the philosophe.

The speaker with the lowest distortion distance is chosen to be identified as the unknown person.

4. EXPERIMENTAL RESULTS

To implement projected speaker recognition system, a system with some voice commands like 'Hello' is taken into account.

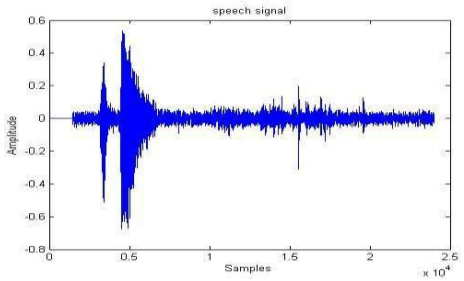


Figure 6. Original speech signal

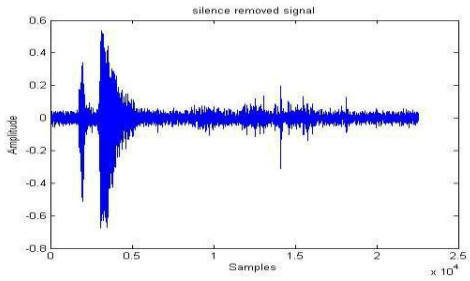


Figure 7. Silence removal signal

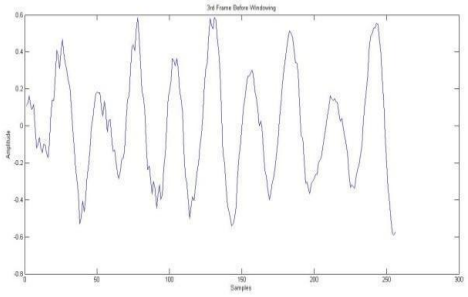


Figure 8. Framing

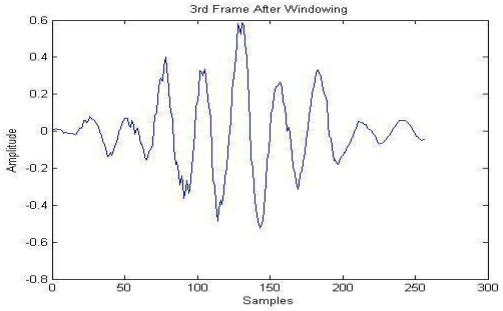


Figure 9. Windowing

Training part is completed in 2 forms. Initial system was trained with one repetition every for every} command and once in each testing sessions.

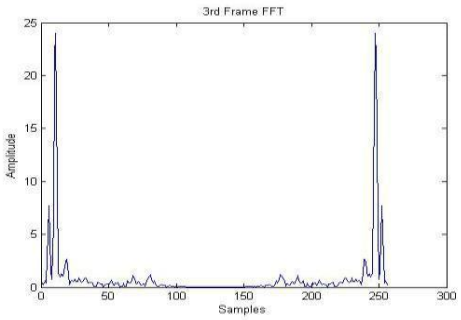


Figure 10. Fast Fourier Transform

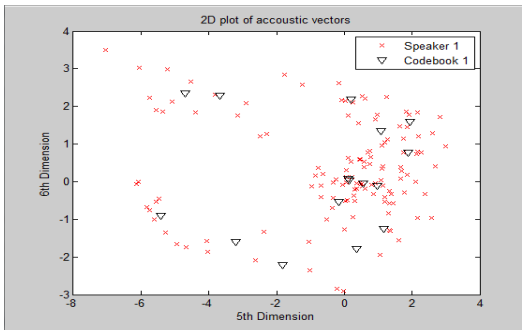


Figure 11. Code vectors

With this sort of coaching error rate is regarding thirteen. In second kind, speaker perennial the words four times in an exceedingly single coaching session, and so doubly in every testing session. By doing this negligible error rate in recognition of commands is achieved.

5. CONCLUSION

The goal of this project was to make a speaker recognition system, associate degreed apply it to a speech of an unknown speaker. By investigation the extracted options of the unknown speech and so compare them to the hold on extracted options for every totally different speaker so as to spot the unknown

speaker. The feature extraction is completed by victimization MFCC (Mel Frequency Coefficients). The operate 'melcepst' is employed to calculate the Mel cepstrum of a sign. The speaker was modelled victimization Vector quantization (VQ). A VQ codebook is generated by clump the coaching feature vectors of every speaker and so hold on within the speaker information. During this technique, the K means that formula is employed to try to to the clump. Within the Recognition stage, a distortion live that supported the minimizing the geometrician distance was used once matching associate degree unknown speaker with the speaker information.

REFERENCES

[1] Mahdi Shaneh and Azizollah Taheri, "Voice Command Recognition System Based on MFCC and VQ algorithms", *World Academy of Science, Engineering and Technology*, Vol. 33, 2009.
[2] Ms. Arundhati S. Mehendale and Mrs. M. R. Dixit, "Speaker Identification Signals and Image Processing", *International Journal (SIPIJ)*, Vol. 2, No. 2, June 2011.
[3] Jamel Price, Dr. Ali Eydgahi, "Design of an Automatic Speech Recognition System Using MATLAB", Chesapeake Information Based Aeronautics Consortium, August 2005.
[4] E. Darren. Ellis, "Design of a Speaker Recognition Code using MATLAB", Department of Computer and Electrical Engineering-University of Tennessee, Knoxville Tennesse.
[5] J. S Chitode, Anuradha S. Nigade, "Throat Microphone Signals for Isolated Word Recognition Using LPC", *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 2, No. 8, August 2012.
[6] B. Gold and N. Morgan, "Speech and Audio Signal Processing", John Wiley and Sons, New York, NY, 2000.
[7] Vibha Tiwari, "MFCC and its applications in speaker recognition", Dep't. of Electronics Engg., Gyan Ganga Institute of Technology and Management, Bhopal.
[8] E. Karpov, "Real Time Speaker Identification", Master`s thesis, Department of Computer Science, University of Joensuu, 2003.