IPASJ International Journal of Electronics & Communication (IIJEC)
Web Site: http://www.ipasj.org/IIJEC/IIJEC.htm
A Publisher for Research Motivation..... Email: editoriijec@ipasj.org
Volume 3, Issue 6, June 2015 ISSN 2321-5984

Person Identification using MFCC and Vector Quantization

[1]Anu L B , Dr Suresh D[2] , Sanjeev kubakaddi[3]

[1]Anu L B Mtech student Electronic, Channabasaveshwara institute of technology Tumkur, India.,

[2]Dr Suresh D S, is the Director of Channabasaveshwara Institute of Technology, Gubbi,Tumkur India.

[3]Mr. Sanjeev Kubakaddi M.Tech ITIE knowledge solution,Bangalore

ABSTRACT

Voice recognition is a process to identify the speaker on the basis of individual information within the speech wave. Recent development has made the voice recognition in the security system; this technique is mainly used in the speaker voice identification and control access like banking by telephone, voice dialling, and database access services. Voice recognition mainly involves two parts, one is the feature extraction and other one is the feature matching. The main approach is to isolate the speech recognition by Mel-Scale Frequency Cepstrum coefficient and Vector quantization. MFCC is used for feature extraction and vector quantization is used to minimize the amount of data to be handled. In feature matching find the vector quantized difference between input utterance of the new signal to test and codebooks stored in data base. From the VQ distortion result we can decide whether the new signal is accepted or rejected. Experimental results show the 100 percent accuracy.

Keywords: MATLAB, Mel Frequency Cepstral coefficient(MFCC), voice recognition, vector quantization.

1. INTRODUCTION

Digital processing speech signal and voice recognition algorithm is very important for fast and accurate automatic voice recognition system. The Speech is an important & first mode of Communication among the human being. The communication among human being and computer interaction is called human computer interface. Voice consist of plenty information, therefore its having too much of complexity. For voice recognition it is divided into two parts one is feature extraction and another one is feature matching. Speech has potential of being important mode of interaction with computer. This paper describes the speech recognition by using MFCC (Mel Frequency Cepstral Coefficient) and vector quantization technique. MFCC is used for feature extraction and vector quantization is used to find distortion between the input utterance of an unknown speaker and the codebook stored in the database is computed, based on vector quantization distortion a decision is made weather to accept or reject the test speaker. This technique is used for the speaker voice to verify their identity and control access to services such as telephone shopping, voice dialling, database access services, banking by telephone, information services, voice mail, remote access to computers, security control for confidential information areas.

Extracted features through MFCC are quantized to many centroids by the vector quantization algorithm. These centroids became codebook of the speaker. MFCC's are calculated in both training folder and in testing folder. Speakers produce same words in a training and testing session later. The Euclidean distance is calculated between the MFCC's of each speaker in training signal to the centroids of individual speaker in testing signal is calculated and the speaker is identified according to the minimum distance.

2 VOICE RECOGNITION

A voice analysis is done after taking an input through Goldwave software from a user. The design of the system involves manipulation of the input audio signal. At different stage, different operations are done on the input signal such as pre-emphasis, framing, windowing, Mel Cepstrum analysis and vector quantization (Matching) of the spoken word. The voice algorithms consist of two folders. The first one is training and the second one is referred testing phase. For speech/speaker recognition, the most commonly used acoustic features are mel-scale frequency Cepstral coefficient (MFCC). MFCC takes human perception sensitivity with respect to frequencies into consideration, and therefore are best for speech/speaker recognition. We shall explain the each step computation of MFCC.

IPASJ International Journal of Electronics & Communication (IIJEC)
Web Site: http://www.ipasj.org/IIJEC/IIJEC.htm
A Publisher for Research Motivation..... Email: editoriijec@ipasj.org
Volume 3, Issue 6, June 2015 ISSN 2321-5984

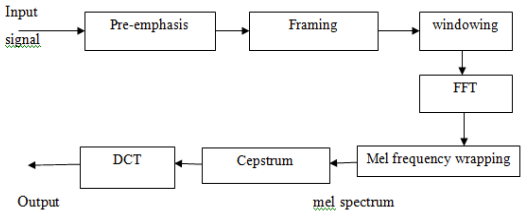

Fig 1:- Block diagram of MFCC

(a) Pre-emphasis

The speech signal x(n) is sent to a high-pass filter

$$y(n) = x(n) - a*x(n-1) \quad (1)$$

Where $s_2(n)$ is the output signal and the value of a is typically between 0.9 and 0.99. The goal of pre-emphasis is to compensate the high-frequency part that was suppressed during the sound production. It can also amplify the high-frequency formants.

(b) Frame blocking

The input speech signal is segmented into frames of 20~30 ms. Usually the frame size (in terms of sample points) is equal to power of two in order to facilitate the use of FFT. Otherwise we need to do zero padding to the nearest length of power of two. If the rate of sample that is sample rate is 16 kHz and the frame size is 320 sample, then the frame duration is 320/16000 = 0.02 i,e sec = 20 ms. if the overlap is 160 points, then the frame rate is 16000/(320-160) = 100 frames per second.

(c) Hamming windowing

Each frame has to be multiplied with a hamming window in order to keep the continuity of the first and the last points in the frame. If the signal in a frame is denoted by x(n), n = 0,…N-1, then the signal after Hamming windowing is x(n)*w(n), where w(n) is the Hamming window defined by:

$$w(n, a)=(1- a)- a \cos(2pn/(N-1)), \quad 0 \leqq n \leqq N-1 \quad (2)$$

(d) Fast Fourier Transform or FFT

Spectral analysis of different pitches in speech signals corresponds to different energy distribution on frequency scale. Therefore FFT is used to obtain the magnitude frequency response of each frame.

When we perform FFT on a frame, we assume that the signal within a frame is periodic and continuous when wrapping around. If not, still perform FFT but the in-continuity at the frame's first and last points is likely to introduce undesirable effects in the frequency response. we have two strategies:

Multiply every frame by a hamming window to increase its continuity at the first and last points.

Choose a frame of a variable size such that it always contains an integer multiple number of the fundamental periods of the speech signal.

The second strategy encounters difficulty in practice since the identification of the fundamental period is not a big problem. Moreover, unvoiced sounds do not have a fundamental period at all. Consequently, usually adopt the first strategy to multiply the frame by a hamming window before performing FFT.

(e) Mel-frequency wrapping

Human perception of frequency contents of sounds for speech signal does not follow a linear scale. Therefore for each tone with an actual frequency is measured in Hz, a particular pitch is measured on a scale called the 'mel' scale. The mel frequency scale is a linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000Hz .As a reference point ,the pitch of a 1 KHz tone ,40dB above the perceptual threshold is defined as 1000 mels. So we can use the following approximate formula to compute the mels for a given frequency *f* in Hz.

$$\text{Mel}(f) = 2595*\log10(1 + f/700) \quad (3)$$

The mel scale filter bank is a series of triangular band pass filters that have been designed to simulate the band pass filtering believed to occur in the audible system. This corresponds to number of band pass filters with constant bandwidth and spacing on a mel frequency scale.

(f) Discrete cosine transform or DCT

In this step apply DCT on the 20 log energy E_k obtained from the triangular band pass filters to have L mel-scale Cepstral coefficients.

$$C_m = S_{k=1}^{N} \cos [m*(k-0.5)*p/N]* E_k, \quad m=1,2, ..., L \quad (4)$$

Where N is the number of triangular band pass filters, L is the number of Mel Scale Cepstral Coefficients. Set N=20 and L=12 performed FFT. DCT converts the frequency domain into a time domain called quefrency domain. The obtained features are same as cepstrum, thus it is referred to as the mel-scale Cepstral coefficients (MFCC).

3 VECTOR QUANTIZATION

The state-of-the-art in feature matching techniques used in speaker recognition includes Dynamic Time Warping (DTW), Hidden Markov Modeling (HMM), and Vector Quantization (VQ). In this paper, the VQ approach is used, due to fast, ease of implementation and having high accuracy. VQ is a process of identifying vectors from a large vector space to a set of number in the space. Each region is called a cluster and can be represented by its centre called a codeword. The group of all codewords is called a codebook.
Vector quantization (VQ) is a lossy data compression method based on principle of blocks coding. It is a fixed-to-fixed length algorithm.

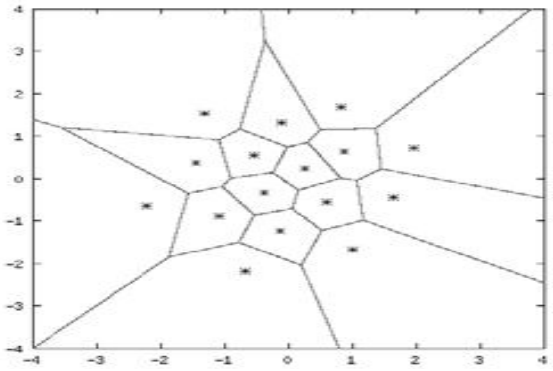

Fig2:- 2Dimensional VQ

Here, every pair of numbers falling in a particular region is approximated by a star associated with that region. In Fig2 shows that the stars are called codevectors and the regions defined by the borders are called encoding regions. The set of all codevectors is called the codebook and the set of all encoding regions is called the partition of the space. The below fig3 show the codebook formation of two speaker, green colour circle are the code vectors and the dark circle is the centroids of the speaker 1, the red colour triangles are the code vectors and dark triangles are the centroids of the speaker 2. The distance between the centroid and the code vector is called the vector quantization distortion.

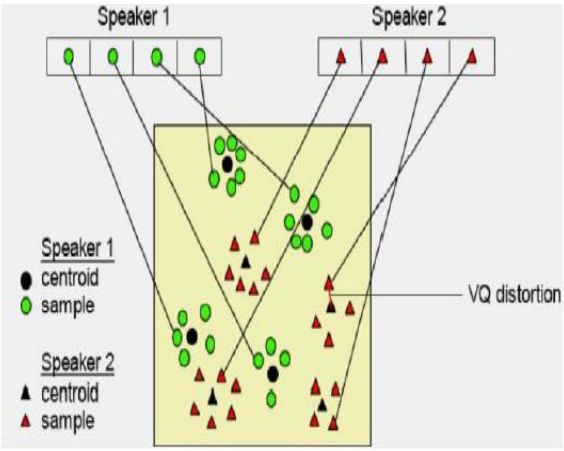

Fig3:- Codebook formation

(a) LBG Design Algorithm

The LBG VQ design algorithm is an iterative algorithm which alternatively solves the above two optimality criteria. The algorithm search for an initial codebook $C^{(0)}$. This initial codebook is obtained by the splitting method. In this method, an initial codevectors is set as the average of the entire training sequence. This codevectors is then divides into two. The iterative algorithm is proceed with these two vectors as the initial codebook. The final two codevectors are spitted into four and the process is repeated until the desired number of codevectors is obtained. The figure shows LBG

IPASJ International Journal of Electronics & Communication (IIJEC)
Web Site: http://www.ipasj.org/IIJEC/IIJEC.htm
A Publisher for Research Motivation..... Email: editoriijec@ipasj.org
Volume 3, Issue 6, June 2015 ISSN 2321-5984

algorithm flow chart. In the first block the centroid has to be found and that centroid is split to get initial centroid, therefore the main centroid is split. And one centroid and its cluster is taken to compute distortion between centroid and the corresponding codevectors, the cycle repeats until the smallest distance is obtained. The average distortion is to be less than e then the cycle repeats for new cluster until all the codevectors or checked for initial centroid

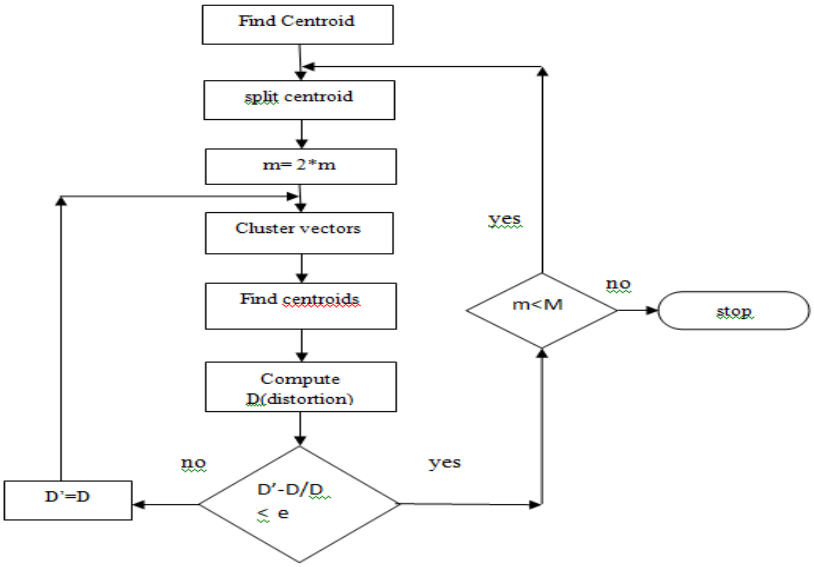

Fig4:- Flow chart of LBG algorithm

4 EXPERIMENTAL RESULTS AND ANALYSES

In this paper we have used eight different speaker signal and it is stored in train and test folder, first extraction features from MFCC and matching the feature using vector quantization, the eight signals are matched with both the train and test folder. And different plots are shown like fig5 power spectrum, mel scale filter bank, fig6 2D plot of acoustic vector and finally fig7recognition rate.

The experimental result shows that it is 100 % efficient in recognizing the eight person signals which is stored in train and test folder.

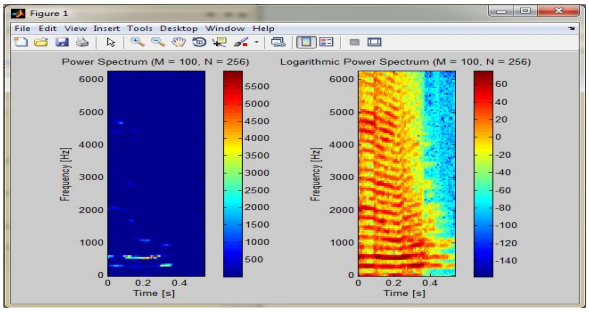

Fig5:- power spectrum

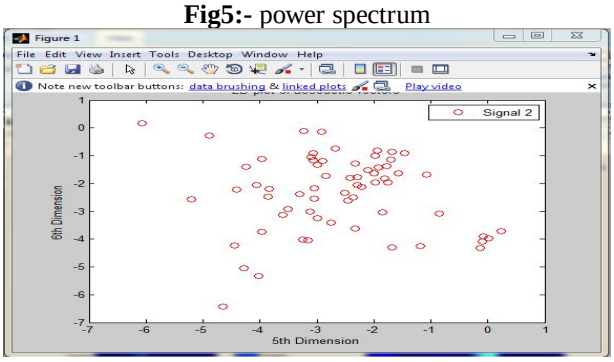

Fig6:- 2D Plot of acoustic vectors

Fig7:- Recognition rate in percentage

5 CONCLUSION AND FUTURE WORKS

The aim of this paper is to recognize the speaker using feature extraction and feature matching. Mel Frequency Cepstral Co-efficient is used for feature extraction and vector quantization is used for feature matching, both Mel frequency and hamming window gives the good performance. It also suggests that to obtain satisfactory result, the number of centroids has to be increased as the number of speakers in the database increases. In order to achieve the better performance the training sessions have to be repeated so as to update the speaker specific codebooks in the database as it is shown in psychophysical studies that there is a probability that human speech may vary over a period of 2-3 years. The experimental results are analyzed using MATLAB and gives the efficient result.

REFERENCES

[1] Md. Rashidul Hasan, Mustafa Jamil, Md. Golam Rabbani Md. Saifur Rahman" speaker identification using mel frequency cepstral coefficients" Electrical and Electronic Engineering, Bangladesh University of Engineering and Technology, Dhaka.

[2] Lindasalwa Muda, Mumtaj Begam and I. Elamvazuthi " voice recognition algorithms using mel frequency cepstral coefficient (mfcc) and dynamic time warping (dtw) techniques" journal of computing, volume 2, issue 3, march 2010.

[3] Aseem saxena, Amit kumar sinha, Shashank chakrawarti, Surabhi charu, Suresh Gyan Vihar University, Jaipur, Rajasthan, India "Speech recognition using Matlab" International Journal of Advances In Computer Science and Cloud Computing, Nov-2013

[4] Shivanker Dev Dhingra , Geeta Nijhawan , Poonam Pandit Student, Dept. of ECE, MRIU, Faridabad, Haryana, India. "Isolated speech recognition using mfcc and dtw" International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 2, Issue 8, August 2013.

[5] Santosh K.Gaikwad, Bharti W.Gawali, Pravin Yannawar Assistant Professor Department of CS& IT Dr.Babasaheb Ambedkar Marathwada University "A Review on Speech Recognition Technique" International Journal of Computer Applications Volume 10– No.3, November 2010

[6] Manan Vyas B.E Electronics, University of Mumbai "A Gaussian mixture model based speech recognition system using matlab" Signal & Image Processing : An International Journal (SIPIJ) Vol.4, No.4, August 2013.

[7] Robert Modic1, Børge Lindberg, Interactive Systems Laboratory University of Ljubljana, Slovenia, Bojan Petek #Center for PersonKommunikation Aalborg University, Denmark. "Comparative Wavelet and MFCC Speech Recognition Experiments on the Slovenian and English SpeechDat"

[8] Ferda Ernawan and Nur Azman Abu, Nanna Suryana Faculty of Information and Communication Technology Universitas Dian Nuswantoro Semarang, Indonesia. Faculty of Information and Communication Technology Universiti Teknikal Malaysia Melaka Melaka, Malaysia "Spectrum analysis of speech recognition via discrete tchebichef transform" International Conference on Graphic and Image Processing

[9] Noelia Alcaraz Meseguer "Speech Analysis for Automatic Speech Recognition" Norwegian University of Science and Technology Department of Electronics and Telecommunications, July 2009.

[10] Arun Rajsekhar. G. "Real time speaker recognition using MFCC and VQ" Department of electronics & communication engineering National Institute of Technology. 2008

[11] Dipmoy Gupta, Radha Mounima C. Navya Manjunath, Manoj PB Dept. of EC,AMCEC, Bangalore " Isolated Word Speech Recognition Using Vector Quantization (VQ)" International Journal of Advanced Research in Computer Science and Software Engineering Volume 2, Issue 5, May 2012

AUTHOR

Anu L B perceiving MTech Electronic Channabasaveshwara Institute of Technology, Gubbi,Tumkur India.

Dr. Suresh D S has completed B.E(IT),M.Tech, MBA & PHD. His area of research is in the field of embedded Systems. He authored a few Text books and has published papers in many national/international journals & Conferences. He has guided many projects sponsored by KSCST and has recently filed patent of his project on "Health Monitoring System". Prof. Suresh is one among few instrumental in giving a base to Toshiba in India for starting its business operations. He also professionally associated with many MNCs Like Future Techno Designs, Creative Labs etc. and many Government organizations like GTTC, Southern Railways, Unique Identification authority of India, District development forums etc. As resource person Prof Suresh D S has trained many engineers for WASE program at Wipro Bangalore.