A Voice Based Recognition System Using Vector Quantization And Mel Frequency Cepstral Coefficients

Nitesh Bathma[1], Akant Kumar Raghuwanshi[2]

[1]M. Tech. (DC) VIT Bhopal, [2]Assistant Professor EC deptt., VIT Bhopal

***Abstract:* Voice acknowledgment are turning out to be increasingly helpful these days. Different fields for research in voice handling has been finished. In this work, the Mel Frequency Cepstrum Coefficient (MFCC) and Vector Quantization (VQ) has been utilized for making a content autonomous speaker recognizable proof framework. A few elements are removed from voice flag of talked words utilizing MFCC. The VQ-based strategies are parametric methodologies which utilize VQ codebooks comprising of a little number of delegate highlight vectors,. Voice acknowledgment frameworks are the productive options for such gadgets where writing gets to be troublesome.**

***Keywords:* MATLAB, Mel Frequency Cepstral Coefficients (MFCC), Speaker Recognition, Vector Quantization(VQ).**

I. INTRODUCTION

The Voice is the most well-known and essential method of correspondence among individuals. Human voice passes on a great deal more data, for example, sexual orientation, feeling and personality of the speaker. Voice Recognition can be characterized as the way toward changing over voice flag to a grouping of words by means an Algorithm .The target of voice acknowledgment is to figure out which speaker is available in view of the individual's portrayal [1].The most prominent ghastly based parameter utilized as a part of acknowledgment approach is the Mel Frequency Cepstral Coefficients called MFCC.MFCCs are coefficients, which speak to sound, in light of impression of human sound-related frameworks. By utilizing hamming window, voice flag is partitioned into various pieces of brief span so Fourier change can be connected. In this work, the Mel recurrence Cepstrum Coefficient (MFCC) highlight has been utilized for outlining a content autonomous speaker distinguishing proof framework. The extricated voice components (MFCC's) of a speaker are quantized to various centroids utilizing vector quantization calculation. These centroids constitute the codebook of that speaker. MFCC's are figured in preparing stage and again in testing stage. Speakers articulated same words once in an instructional course and once in a testing session later. The Euclidean separation between the MFCC's of every speaker in preparing stage to the centroids of individual speaker in testing stage is measured and the speaker is distinguished by least Euclidean distance[11].The code is created in the MATLAB environment and plays out the recognizable

proof agreeably.

II. DATABASE

The database comprises of two sections of voice tests that are recorded in an earth controlled space to have all potentially less

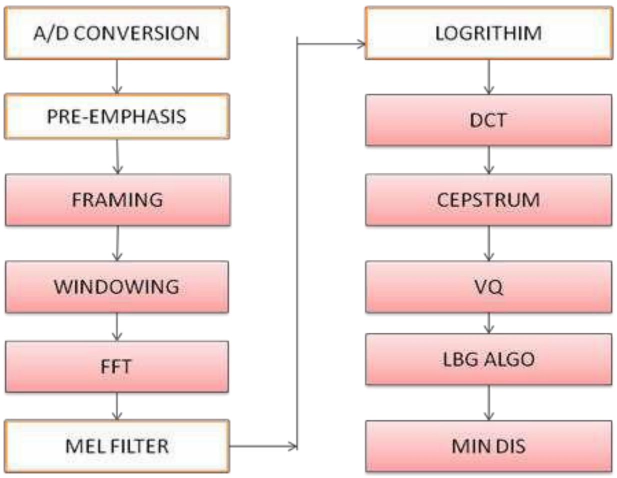

Fig 1.Complete Flow Chart.

acoustical meddles to the nature of voice test amid the recording time. The initial segment is instructional meeting contains add up to ten speakers, talking words. All voice signs are recorded under most comparable condition, for example, the same length of recording time and the level of sound adequacy

.In testing session when a Matlab program is executed it hypothesizes to the client to pick any voice test from the specimens that are prerecorded in the database. MFCC at the back end removes the elements of the picked voice test. At that point the euclidean separation between the prerecorded specimens and the example for test is measured. The finish square outline of the procedure is appeared in fig 1.

III. FEATURE EXTRACTION

The extraction of the acoustic components [2] from the voice flag is an essential assignment to create a superior acknowledgment execution. A few component extraction calculations can be utilized, for example, Linear Predictive Coefficients (LPC), Linear Predictive Cepstral Coefficients (LPCC), Mel Frequency Cepstral Coefficients (MFCC) and so forth [4].

1. MFCC are the most vital components, which are required among different sorts of voice applications.
2. It gives high precision comes about for clean voice.
3. MFCC can be viewed as the "standard" components in speaker and in addition voice acknowledgment.

A. Input

The speaker test voice flag is given with the assistance of inbuilt receiver in webcam. Webcam is utilized for giving the info utterence. It gives clear discussions without irritating foundation commotion because of the inherent mic with Right Sound innovation. The edge rate is 30 fps. The voice flag is appeared in fig 2.

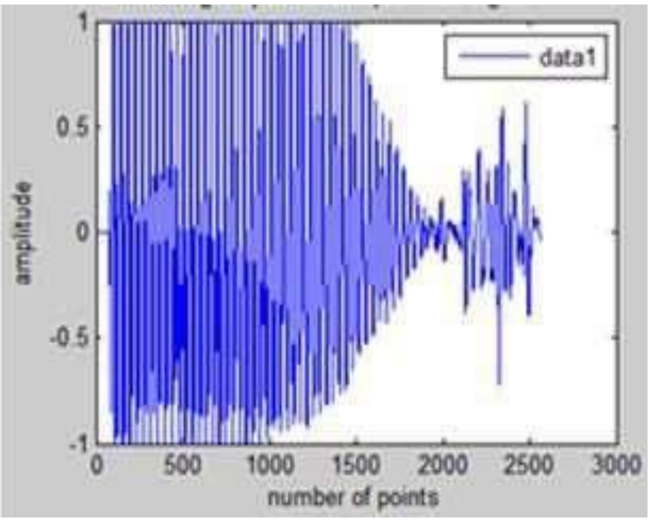

Fig 2.Voice sample voice signal.

B. Frame Blocking

The process of segmenting the voice samples obtained from analog to digital conversion (ADC) into a small frame with the length within the range of 20 to 40 msec. The voice signal is divided into frames of N samples. Adjacent frames are being separated by M (M<N).Typical values used are M

= 100 and N= 256.The framing is shown in fig 3.

C. Windowing

Hamming window is used as window shape by considering the next block in feature extraction processing chain and integrates all the closest frequency lines[6]. In voice recognition, the most commonly used window shape is the hamming window [3][7].The result obtained after windowing is given in fig 4.

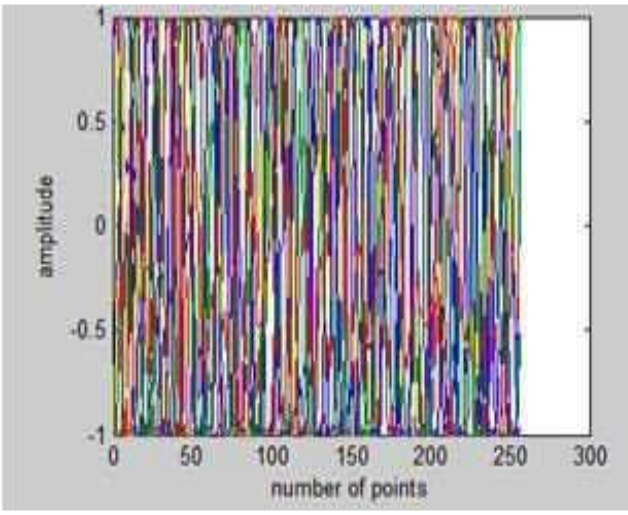

Fig 3.Framing of voice signal.

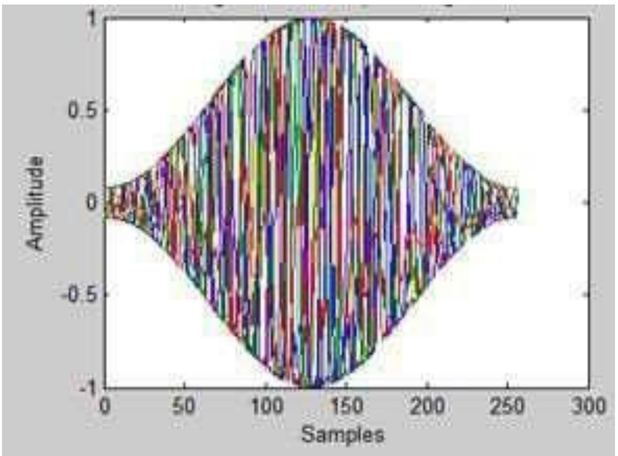

Fig 4.Windowed Voice Signal.

D. Fast Fourier Transform

To convert each frame of N samples from time domain into frequency domain FFT is being used. The Fourier Transform is used to convert the convolution of the glottal pulse and the vocal tract impulse response in the time domain into multiplication in the frequency domain [5].

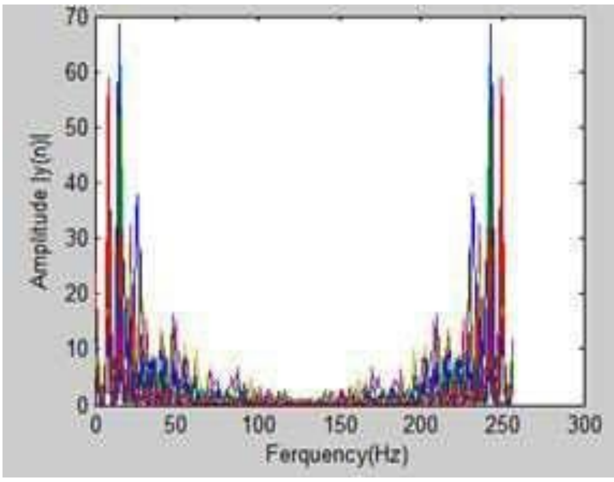

Fig 5.FFT of Windowed Voice Signal.

E. Mel Frequency Wraping

The human ear perception of frequency contents of sounds for voice signal does not follow a linear scale. Therefore, for the frequency measured in Hz, a subjective

pitch is measured on a scale which is called the „mel"scale. The mel frequency scale is a linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000Hz. To compute the mels for a given frequency f in Hz, the formula used is.

$$Mel(f) = 2595*\log10(1 + f/700)$$

F. Clustering Vectors: LBG Algorith

The outstanding calculation called the LBG calculation i.e of L training vectors into a set of M codebook vectors. These are the various stages used for designing an M-vector codebook in the LBG algorithm. It initially starts with the designing of a 1-vector codebook, then uses a division technique on the code words so as to get a 2-vector codebook and then this splitting process continues until the desired M-vector codebook is obtained.

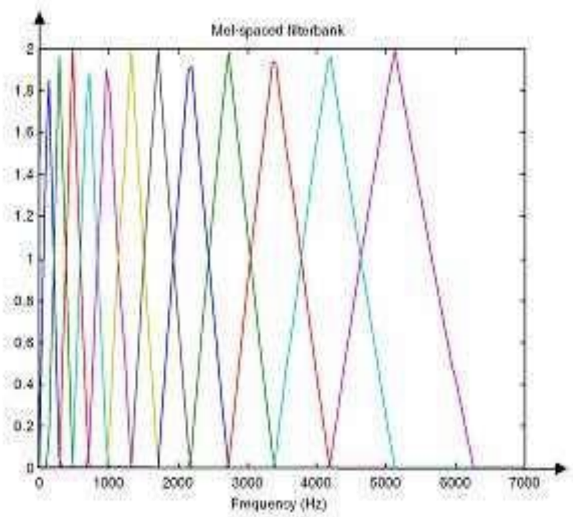

Fig 6 .Mel Scale Filter Bank.

G. Cepstrum

The cepstrum converts the mel spectrum into mel cepstrum i.e we convert the log mel spectrum back to time. The result is called the Mel Frequency Cepstrum Coefficients (MFCC). The cepstral representation of the voice spectrum provides a good representation of the signal for the given frame analysis. Because the mel spectrum coefficients are real numbers, we can convert them to the time domain using the discrete cosine transform (DCT).

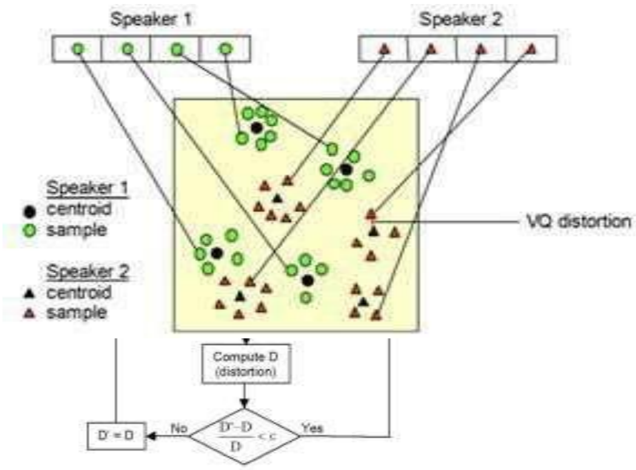

Fig 7.Two speaker codebook formation.

IV. FEATURE MATCHING

There are numerous sorts of coordinating systems utilized as a part of the speaker acknowledgment, for example, Dynamic Time Warping (DTW),Hidden Markov Modeling (HMM), and Vector Quantization(VQ). Vector Quantization system is utilized for highlight coordinating.

A. Vector Quantization(VQ)

Vector Quantization (VQ) is a procedure of speaking to the vectors from an expansive vector space into a predetermined number of locales present in that space. The every area so got is known as a group and can be controlled by its inside called the codeword. The accumulation of all code expressions of the vectors are known as a codebook. Separate from a vector to the nearest codeword of a codebook is called VQ-twisting. Input expression of an obscure voice is "vector-

V. COMPARSION OF THE PRESENTWORK & PREVIOUS WORK

PRESENT	PREVIOUS
1. Feature extraction- MFCC Feature matching- VQ. 2. This is text independent method. 3. Tolerance is the threshold value which is equal to 3 for measuring the distortion.	1. Feature extraction- MFCC Feature matching- DTW[8][9]. 2.This is text dependent method[10]. 3. Tolerance is 10% for identifying the speaker [10].

V. CONCLUSION AND FUTURE WORKS

The fundamental point of this venture was to perceive voice utilizing MFCC and VQ strategies. The component extraction was done utilizing Mel Frequency Cepstral

Coefficients {MFCC} and the element coordinating was finished with the assistance of Vector Quantization(VQ) system. A bending measure in view of minimizing the Euclidean separation was utilized while coordinating the obscure voice motion with the voice flag database. The exploratory results were investigated with the assistance of MATLAB and it is demonstrated that the outcomes are proficient. The framework is content autonomous. The procedure can be reached out for n number of speakers.

VI. REFERENCES

[1] G.I. Sapijaszko and Wasfy B. Mikhael " An Overview of Recent Window Based Feature Extraction Algorithms for Speaker Recognition," IEEE 55th International Midwest Symposium on Circuits and Systems , pp. 880-883, 2015.

[2] P. Premakanthan and W. B. Mikhael, "Speaker verification/recognition and the importance of selective feature extraction: review," in Circuits and Systems, 2014. MWSCAS 2014. Proceedings of the 44th IEEE 2014 Midwest Symposium on, 2014, pp. 57 -61 vol.1.

[3] H. Beigi, Fundamentals of speaker recognition [electronic resource] / Homayoon Beigi. New York: Springer Science+Business Media, LLC, 2013.

[4] P. Kumar, N. Jakhanwal, and M. Chandra, "Text Dependent Speaker Identification in Noisy Environment," in Devices and Communications

[5] (ICDeCom), 2011 International Conference on, 2012, pp. 1-4.

[6] J. Wu and J. Yu, "An improved arithmetic of MFCC in speech recognition system," in Electronics, Communications and Control (ICECC), 2011 International Conference on, 2011, pp. 719-722.

[7] M. R. Islam, M. F. Rahman, and M. A. G. Khan, "Improvement of speech enhancement techniques for robust speaker identification in noise," in Computers and Information Technology, 2009. ICCIT '09. 12th International Conference on, 2009, pp. 255-260.

[8] A. Revathi and Y. Venkataramani, "Text Independent Composite Speaker Identification/Verification Using Multiple Features," in Computer Science and Information Engineering, 2009 WRI World Congress on, 2009, pp. 257-261.

[9] R. J. Mammone, Z. Xiaoyu, and R. P. Ramachandran, "Robust speaker recognition: a feature-based approach," Signal Processing Magazine, IEEE, vol. 13, p. 58, 2008.

[10] M. G. Sumithra and A. K. Devika, "A study on feature extraction techniques for text independent speaker identification," in Computer Communication and Informatics (ICCCI), 2012 International Conference on, 2012, pp. 1-5.

[11] M. A. Hossan, S. Memon, and M. A. Gregory, "A novel approach for MFCC feature extraction," in Signal Processing and Communication Systems (ICSPCS), 2010 4th International Conference on, 2010, pp. 1-5.

[12] S. Biswas, S. Ahmad, and M. K. Islam Mollat, "Speaker Identification Using Cepstral Based Features and Discrete Hidden Markov Model," in Information and Communication Technology, 2007. ICICT '07. International Conference on, 2007, pp. 303-306.

[13] T. Schurer, "An experimental comparison of different feature extraction and classification methods for telephone speech," in Interactive Voice Technology for Telecommunications Applications, 1994., Second IEEE Workshop on, 1994, pp. 93-96.

[14] P. Delsarte and Y. Genin, "The split Levinson algorithm," Acoustics, Speech and Signal Processing, IEEE Transactions on, vol. 34, pp. 470- 478, 1986.

[15] R. A. Cole, M. Noel, "The CSLU speaker recognition corpus", ICSLP-1998,paper,0856

[16] Y. Yuan, P. Zhao, and Q. Zhou, "Research of speaker recognition based on combination of LPCC and MFCC," in Intelligent Computing and Intelligent Systems (ICIS), 2010 IEEE International Conference on, 2010, pp. 765-767

[17] A. Lawson, P. Vabishchevich, M. Huggins, P. Ardis, B. Battles, and A. Stauffer, "Survey and evaluation of acoustic features for speaker recognition," in Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, 2011, pp. 5444-5447

[18] R. Togneri and D. Pullella, "An Overview of Speaker Identification: Accuracy and Robustness Issues," Circuits and Systems Magazine, IEEE, vol. 11, pp. 23-61, 2011

[19] M.A.Anusuya , S.K.Katti "Voice Recognition by Machine: A Review" International journal of computer science and Information Security 2013.

[20] Sanjit K.Mitra, Digital Signal processing using MATLAB, MC-Graw Hill, 2000.