Performance Comparison of Speaker Identification using Vector Quantization by MFCC Algorithm

[1] Dr.Rajeev Mathur, [2] Mr Sanjay N.Sharma
[1] Principal, [2]Research Scholar
[1] Lachoo Memorial College of Science & Technology, [2]Department of Computer Science
[1] Jodhpur (Rajasthan), [2]Jodhpur National University

***Abstract* - In the proposed work, feature extraction algorithm is used for speaker identification system, the proposed feature extraction algorithm is Mel frequency Cepstrum Coefficient (MFCC). The extracted speech features (MFCC's) of a speaker using vector quantization algorithm are quantized to a number of centroids. The distance between centroids of individual speaker in testing phase and the MFCC's of each speaker in training phase is measured and the speaker is identified according to the minimum distance. The code performs the identification satisfactorily and is developed in the MATLAB environment.**

***Index Terms* - Speaker Identification, Vector Quantization (VQ), MFCC**

I. INTRODUCTION

The voice information available in the speech signal can be used to identify the speaker because anatomical structure of the vocal tract is unique for every person. MFCC is based on the human peripheral auditory system [1,2,3,4]. Voice comes under the category of biometric since differences in the anatomical structure are an intrinsic property of the speaker identity. Speaker recognition systems also involve two phases namely, training and testing. Training is registering the voice characteristics of the speaker. Testing is the actual recognition task. Feature vectors are used for building the reference model and also are representing the voice characteristics of the speaker are extracted from the training utterances.

The human perception does not follow a linear scale. Thus for subjective pitch is measured on a scale called the 'Mel Scale' and each tone frequency is measured in Hz, a .The mel frequency scale is a linear frequency spacing below 1000 Hz and logarithmic spacing above 1kHz.As a reference point, the pitch of a 1 kHz tone, 40 dB above the perceptual hearing threshold, is defined as 1000 Mels.

II. MEL FREQUENCY CEPSTRAL

Speech recognition performance is significantly affected by the extraction and selection of the best parametric representation of acoustic signals is an important task in the design of any speech. A compact representation would be provided by a set of mel-frequency cepstrum coefficients (MFCC), which are the results of a cosine transform of the real logarithm of the short-term energy spectrum expressed on a mel-frequency scale. The MFCCs are proved more efficient [5,6,7,8]. The MFCC algorithm includes the following steps.

Mel-frequency wrapping

Human perception does not follow a linear scale but has frequency contents of sounds for speech signal. Thus for subjective pitch is measured on a scale called the 'Mel Scale'and each tone frequency is measured in Hz. The melfrequency scale is logarithmic spacing above 1000Hz and a linear frequency spacing below 1000 Hz .As a reference point, the pitch of a 1 KHz tone ,40dB above the perceptual hearing threshold, is defined as 1000 mels. Therefore we can use the following approximate formula to compute the mels for a given frequency f in Hz.

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

We have used a filter bank, one filter for each desired mel-frequency component. The mel scale filter bank is a series of l triangular band pass filters that have been designed to simulate the band pass filtering believed to occur in the auditory system. Each filter bank has a triangular band pass frequency response and the spacing as well as the bandwidth is determined by a constant mel-frequency interval. This corresponds to series of band pass filters with constant bandwidth and spacing on a mel frequency scale.

Cepstrum

The result called the Mel Frequency Cepstrum Coefficients (MFCC) is formed by converting the log mel spectrum back to time .For a given frame analysis cepstral representation of the speech spectrum provides a good representation of the local spectral properties of the signal. We can convert them to the time domain using the discrete cosine transform (DCT) because the mel spectrum coefficients (and so their logarithm) are real numbers. In this final step log mel spectrum is converted back to time. The result is called the Mel Frequency Cepstrum Coefficients (MFCC).Figure 1.shows the MFCC algorithm.

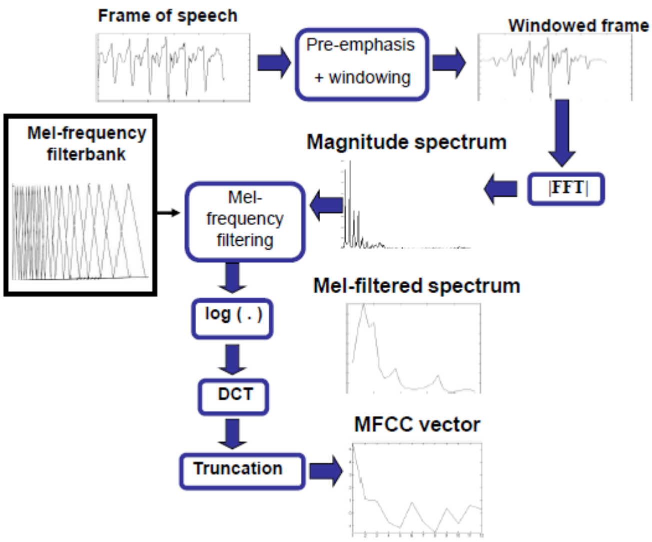


Figure 1. MFCC algorithm

III. VECTOR QUANTIZATION

Vector Quantization allows the modeling of probability density functions by the distribution of prototype vectors. Vector quantization is used for data compression.It works by dividing a large set of points or (vector) into groups having approximately the same number of points closest to them. Each group of vector is represented by its centroid point as in K-means and some other clustering algorithms. One unique property of VQ is there that is density matching property which is powerful, especially for identifying the density of large and high dimensioned data.It is used in speech recognition, CBIR, image segmentation, speech data compression [9,10,11,12]. The density matching property of vector quantization is powerful, especially for identifying the density of large and high-dimensioned data. Since data points are represented by the index of their closest centroid, commonly occurring data have low error, and rare data high error. Hence, Vector Quantization is also suitable for lossy data compression.

A vector quantizer maps k-dimensional vectors in the vector space Rk into a finite set of vectors Y = {yi : i = 1, 2, ..., N}. Each vector yi is called a code vector or a codeword and the set of all the code words is called a codebook. Associated with each codeword, yi, is a nearest neighbor region called Voronoi region, and it is defined by : Vi = {x ∈ Rk : || x – yi || < || x – yj ||, for all j ≠ 1}. Given an input vector, the codeword that is chosen to represent it is the one in the same Voronoi region [13,14,15,16].

.

IV. IMPLEMENTATION

A. PLATFORM EXPERIMENTATION

The implementation of model is done in MATLAB 2012a with basic system of Intel core 2 duo (2.93GHz) with 2GB RAM and minimum of 250GB hard disk for storage. The modules of model are run under MATLAB 2012a compiler. The operating system used is windows 7 for mat lab environment.

B AUDIO DATABASE

The speech samples used in this work are recorded using Audhocity software . The sampling frequency is 8000 Hz (8 bit, mono PCM samples). Table I shows the database description. The samples are collected from speakers of different age group ranging from 15 to 74 years. Ten iterations of four different sentences of varying lengths are recorded from each of the speakers. Ten samples per speaker are taken. These speech signals have an amplitude range of '-1' to '+1'.

TABLE I: Database Description

Parameter	Sample Characteristics
Language	English
No. of Speakers	40
Speech Type	Read Speech
Recording Condition	Normal (a silent room)
Sampling Frequency	8000 Hz
Recording Condition	8 bps

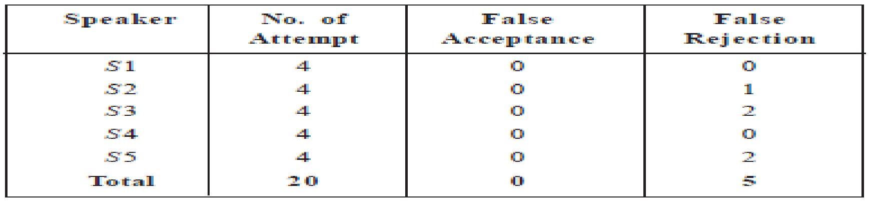
V. COMPARISON OF DIFFERENT RESULTS OF MFCC

The performance of the Mel-Frequency Cepstrum Coefficients (MFCC) may be affected by (1) the number of filters,(2) type of window.In this paper, several comparison experiments are done to find a best implementation [17].

A Effect of number of filters

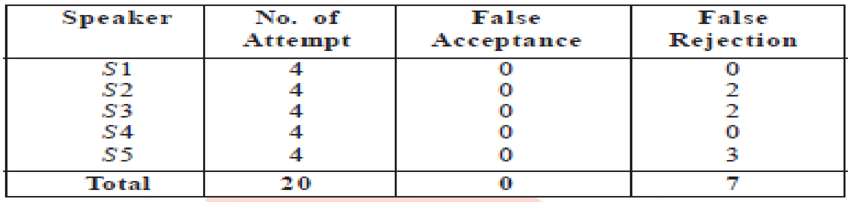
Results of the speaker recognition performance by varying the number of filters of MFCC to 12, 22, 32, and 42 are given. The recognizer reaches the maximal performance at the filter number K = 32. Too few or too many filters do not result in better accuracy. Hereafter, if not specifically stated, the number of filters is chosen to be K = 32.

Table II: MFCC with 12 filters: efficiency is 75 %



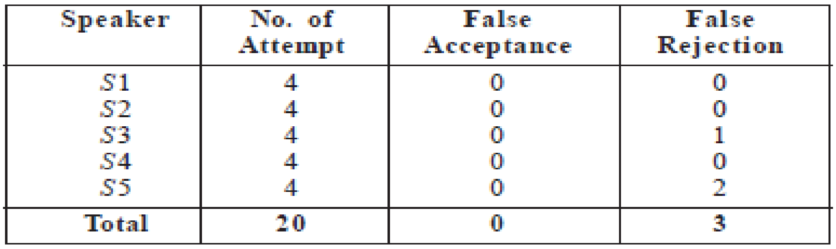
Speaker	No. of Attempt	False Acceptance	False Rejection
S1	4	0	0
S2	4	0	1
S3	4	0	2
S4	4	0	0
S5	4	0	2
Total	20	0	5

Table III:MFCC with 22 filters: efficiency is 65 %



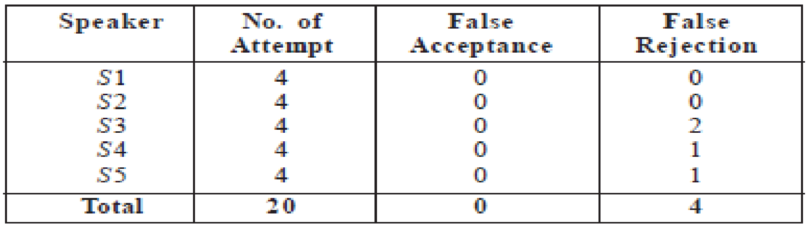
Speaker	No. of Attempt	False Acceptance	False Rejection
S1	4	0	0
S2	4	0	2
S3	4	0	2
S4	4	0	0
S5	4	0	3
Total	20	0	7

Table IV:MFCC with 32 filters: efficiency is 85 %



Speaker	No. of Attempt	False Acceptance	False Rejection
S1	4	0	0
S2	4	0	0
S3	4	0	1
S4	4	0	0
S5	4	0	2
Total	20	0	3

Table V:MFCC with 42 filters:efficiency is 80 %



Speaker	No. of Attempt	False Acceptance	False Rejection
S1	4	0	0
S2	4	0	0
S3	4	0	2
S4	4	0	1
S5	4	0	1
Total	20	0	4

VI. CONCLUSION

In this paper MFCC feature extraction technique for speaker recognition is discussed. Relative to the speaker discriminative vocal tract properties MFCC is well known techniques used in speaker recognition to describe the signal characteristics. The goal of this proposed work was to create a speaker recognition system, and apply it to a speech of an unknown speaker. By investigating the extracted features of the unknown speech and then compare them to the stored extracted features for each different speaker in order to identify the unknown speaker. In future we will try to improve this system to be a text independent speaker identification system

REFERENCES

[1] Lawrence Rabiner, Biing-Hwang Juang and B.Yegnanarayana,“Fundamental of Speech Recognition”,Prentice-Hall, Englewood Cliffs,2009.

[2] S Furui, “50 years of progress in speech and speaker recognition research”, ECTI Transactions on Computer and Information

Technology, Vol. 1, No.2, November 2005.

[3] D. A. Reynolds, “An overview of automatic speaker recognition technology”, Proc. IEEE Int. Conf. Acoust., Speech,Signal Process. (ICASSP’02), 2002, pp. IV-4072–IV-4075.

[4] Joseph P. Campbell, Jr., Senior Member, IEEE, “Speaker Recognition: A Tutorial”, Proceedings of the IEEE, vol. 85, no. 9, pp. 1437-1462, September 1997.

[5] F. Bimbot, J.-F. Bonastre, C. Fredouille, G. Gravier, I Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-García, D.Petrovska-Delacrétaz, and D. A. Reynolds, “A tutorial on text-independent speaker verification,” EURASIP J. Appl. Signal Process., vol. 2004, no. 1, pp. 430–451, 2004

[6] D. A. Reynolds, “Experimental evaluation of features for robust speaker identification,” IEEE Trans. Speech Audio Process., vol. 2, no. 4, pp. 639–643, Oct. 1994.

[7] Tomi Kinnunen, Evgeny Karpov, and Pasi Fr¨anti, “Realtime Speaker Identification”, ICSLP2004

[8] Marco Grimaldi and Fred Cummins, “Speaker Identification using Instantaneous Frequencies”, IEEE Transactions on Audio, Speech, and Language Processing, vol., 16, no. 6, August 2008.

[9] Zhong-Xuan, Yuan & Bo-Ling, Xu & Chong-Zhi, Yu (1999). “Binary Quantization of Feature Vectors for Robust Text-Independent Speaker Identification” in IEEE Transaction on Speech and Audio Processing, Vol. 7, No. 1, January 1999. IEEE, New York, NY, U.S.A.

[10] R. M. Gray.: „Vector quantization‟, IEEE ASSP Marg., pp. 4-29, Apr. 1984.

[11] Y. Linde, A. Buzo, and R. M. Gray.: „An algorithm for vector quantizer design,” IEEE Trans. Commun.‟, vol. COM-28 no. 1, pp. 84-95, 1980.

[12] A. Gersho, R.M. Gray.: „Vector Quantization and Signal Compression’, Kluwer Academic Publishers, Boston, MA, 1991.

[13] F. K. Soong, et. al., “A vector quantization approach to speaker recognition”, At & T Technical Journal, 66, pp. 14-26, 1987.

[14] A. E. Rosenberg and F. K. Soong, “Evaluation of a vector quantization talker recognition system in text independent and text dependent models”, Computer Speech and Language 22, pp. 143-157, 1987.

[15] Jeng-Shyang Pan, Zhe-Ming Lu, and Sheng-He Sun.: „An Efficient Encoding Algorithm for Vector Quantization Based on Subvector Technique‟, IEEE Transactions on image processing, vol 12 No. 3 March 2003.

[16] F. Soong, E. Rosenberg, B. Juang, and L. Rabiner, "A Vector Quantization Approach to Speaker Recognition", AT&T Technical Journal, vol. 66, March/April 1987, pp. 1426.

[17] Md. Rashidul Hasan, Mustafa Jamil, Md. Golam Rabbani Md. Saifur Rahman , “Speaker Identification using Mel Frequency Cepstral Coefficients”, 3rd International Conference on Electrical & Computer Engineering ICECE held at Dhaka, Bangladesh , 28-30 December 2004.