SCIREA Journal of Electrical Engineering

http://www.scirea.org/journal/DEE

August 12, 2020

Volume 5, Issue 4, August 2020

MFCC based performance analysis of VQ and GMM Speaker identification system

Chandar Kumar, Dr.Engr.Zahid Ali, Suresh Kumar, Syed Zain ul Abedin Abid, Chaman Lal

Faculty of Engineering, Science and Technology, Indus University, Karachi, Sindh, Pakistan

chandar.kumar@indus.edu.pk(Chandar Kumar), arain.zahid@indus.edu.pk(Dr.Engr.Zahid Ali), sureshkumar@indus.edu.pk(Suresh Kumar), zain.abidi@indus.edu.pk(Syed Zain ul Abedin Abid), chaman.lal@indus.edu.pk (Chaman Lal)

Abstract

Speaker identification is the key area of digital signal processing where the synthesis and noise reduction of speech are the core research areas. Speaker identification system is influenced by the background noise which directly affects the efficiency of system and is still reflected as a challenging question in speaker identification system. Several useful techniques for feature extraction have been proposed and refined. In this paper, the performance of GMM and VQ has been investigated on the basis of their effects in text dependent speaker identification and proposed the optimum techniques for MFCC based speaker identification system.

Keywords: vector –quantization, speaker identification, Gaussian mixture

I. Introduction

Speech processing plays a vital role in digital signal processing; in this area speaker identification is an inspiring and perplexing question. Speaker's identification is a unique identification procedure that is based on individual information in the speech signal. The voice of speaker utilized this technique to prove their individuality and gets the access of services where the security issues are extremely concern such as voice mail, data base access service, information service. It can easily detect the speakers by the several discriminating characteristics of human speech. The energy contained by the speech in band of zero frequency to five kilo hertz. The characteristics of speech signal vary significantly over the time function. Speaker identification system is used in two different categories are text-dependent and text-independent, in text-dependent, The enrolment and recognition must deal with the same text which enhance the working condition with the supportive speakers and in the text independent is more suitable in case of non-supportive speaker because it is unknown of expression of speaker [1]. Two distinguished phases is severed by each of speaker identification system. Training or enrolment phase comes in first phase whereas the testing or operational phase in second phase. In training section, samples of every speaker speech have been used to train the system by creating a reference model of speaker. In first phase (i-e, training phase) the system can shape a reference model for speaker through the samples of speech provided by every single enrolled speaker after this, the recorded samples move to testing phase, where the decision is made by the correlating with prerecorded reference model.

In [2, 3] the MFCC is measured in the front-end of the utmost consistent and resourceful front-end application for the use of speaker identification because it has the capability to indicate the audio, is set on observation. The most recent stage in the development of speaker recognition exploration mainly examines speaker explicit corresponding facts comparative to MFCC [4]. It can be noticed that the speaker recognitions have significantly enhanced the performance especially when corresponding facts is Ex-Ored with MFCC in characteristics level via both mild concatenation and merging visuals results. The pitch of speech signal is the vital element [5], residual-phase, prosody, dialectical characteristics [6-8]. In [9], it is discovered that, through the view of the filter-bank, the same specifics can effortlessly be reserved by the speech energy spectrum's high frequency. Specific speaker's features are employed to illustrate at portion of high-frequency of the speech energy spectrum. Upturned Mel Frequency Cepstral Coefficient has been chosen to extract the features through vector quantization method [10]. In recent times [11-12], this analysis can be realized speedily in improvement, which will lead the way to the advancement of numerous configuration based techniques for instance VQ, GMM, HMM and BTM with features of development, [13-14] showed the Pattern recognition and Linde, Buzo and Gray (LBG) set of rules founded VQ (i-e, vector quantization) method is exploited for feature's perception. This paper showed the execution of VQ and GMM has been examined and proposed the optimum technique for MFCC based speaker identification.

Section II discusses the key feature extraction techniques involved in this research. Section III deliberates the process of speaker model creation; Section IV provides the detailed results with supporting considerations and lastly the paper will be concluded.

II FEATURE EXTRACTION

From the angle of the automated job of speaker identification, it must be convenient to study about the speech-signal's features in term of both speaker and speech. In this identification technique, the most important pace is to extract the essential specifics for accurate observation which is appropriate for operational-molding in term of size and form. Through the feature extrication the original speech signal is transported into a compact, at this stage the raw signal is not as discriminative and stable as the operative interpretation. Before the extrication of features of the speech-signal is preliminary processing through framing and windowing techniques. The analog speech signal is divided into equally spaced to form the frames, where $\mathcal{M}$ indicates the samples per frame. The overlying of successive frames thru $\mathcal{M}$ with one another is reflected in [15]. The windowing method is preferred to use in framing the signal in so as to protect the signal at its end-points from swift variation. Consequently, the windowed frames are formed in term of segments by the multiplication of both window-function and frames respectively. The most suitable and commonly used for frames is Hamming-window.

$$\mathbb{W}(\mathbb{m}) = 0.54 - 0.46 \cos \frac{2\pi n}{\mathcal{M} - 1} \qquad (1)$$

1. Mel Frequency Cepstral Coffiecient (MFCC)

David and Mermelstain have designed a technique for features extraction from speech signal is called Mel Frequency Cepstral Coffiecient. This technique has presented to be less Inclined for the change of the voice of speaker and surrounding environment. The base of this technique is an ear of human being is called as discrepancy. Bandwidths are precarious per frequency; filters are sequentially set apart and logarithmically at both frequencies correspondingly.

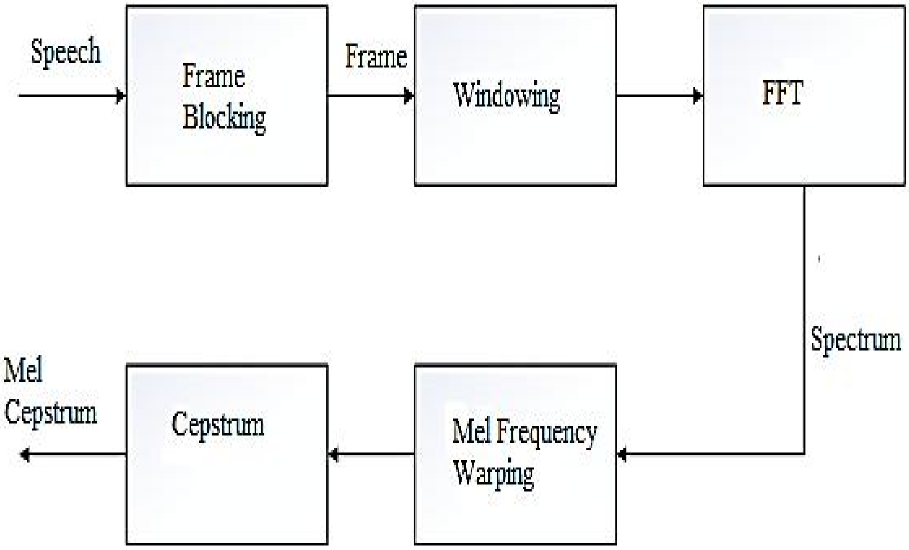

Fig. 1 Schematic drawing of MFCC.

The phonetically significant speech characteristics have been taken by filters. The sound of frequency is dignified in the Mel. It is bearing a resemblance to non-linearly to the average frequency, which is practically linear to 1 KHz and above Logarithmic. By using filter bank, the mal-spectrum is generated; [14], a filter is required for each preferred component of mal-frequency with triangular band-pass frequency response is calculated by following formula.

$$Mel(f) = 2595 \times \log(1 + f/700) \qquad (2)$$

Next to this step, through the discrete cosine transform the log Mel spectrum has converted again into time by the discrete cosine transform methods and created Mel Frequency Cepstral Coffiecient. It offers the superiority of the characteristics of local ethereal signal for the investigation of identified frame.

III SPEAKER MODEL CREATION

A. Gaussian Mixture Model (GMM):

The Gaussian mixture model is very convenient techniques to design the possible model of speech signal of respectively speaker. It is relatively sovereign between the various possibility models. Through the GMM techniques the feature extraction of speech-signal of respective speaker is demonstrated. The consequence of Gaussian mixture model is the summation of linear-components (Gaussian-distribution), represented by M is called mixture.

The factor of total Gaussian mixture density is mixture weights, covariance-matrices and means-vectors from all densities of element [15]. All speaker, in identification of speaker is categorized through a Gaussian mixture model and is stated to via model dint (λ), which is signified thru following system.

$$\lambda = \{\rho_i, \mu_i, \textstyle\sum_i\} i = 1, \dots \dots \dots \dots, \mathbb{M} \qquad (3)$$

The amount of training-vectors, factors of firmed possibility model is estimated through the technique of the repetitive EM (i-e, expectation-maximization) [12]. The expectation-maximization technique repetitively improves the specifications of Gaussian mixture model to frequently enhance the possibility of estimated model for the perceived specific trajectories, namely. For repetitions $\mathbb{k}$ and $\mathbb{k} + 1$, $\rho\left((X/\bar{\lambda})(\mathbb{k}+1)\right) \geq \rho\left({}^{X}\!/_{\lambda}(\mathbb{k})\right)$. This is the essential concept of the EM techniques, new model is assessed by means of primary model (λ) is

$$\rho(X/\bar{\lambda}) \geq \rho\left({}^{X}\!/_{\lambda}\right) \qquad (4)$$

This new model is the primary model for the next repetition and this process will be repeated till to reach at convergence edge.

B. Vector Quantization (VQ)

Speakers Identification system is able to evaluate prospect distribution of the considered feature vectors. Also, it is not feasible to continue each single generated vector by training mode. Over a high dimensional space such distributions are distinct for the moment. Smallest part of template vector is quantized by each single feature vector. The method of mapping vectors to different regions from large space is known as vector-quantization. A cluster is made up of every single region and it symbolized by its center termed as code word. Code book of the estimated training material is the sum of the individual code words.

VQ is generates through cluster of individual training acoustic vectors for each single well known speaker, algorithm is utilized for this is K-Mean).

For execution the k-Mean algorithm following recursion points are necessary:

To design the k clusters by dividing the k points into equally spaced

To calculate the centroids of respectively cluster through calculating the mean of feature-vectors

To calculate the minimum distance to place the feature-vectors close to respective centroid and assign the groups.

Recurrence of point 2 and 3 until no centroids are transferrable

Find the distance between the centroid of test signal to nearby the object in the preparation databank. Find the shortest distance to identify the speaker.

IV. RESULT AND DISCUSSION

A. Effect of Variation in Filter Coefficients:

It can be seen from the data base of system that stored the 20 speakers containing half of both male and female speaker and the coefficients of filters varies between 12 and 25. The results show that from 12 to 20 the number of filtering elements of the filter has been increased; consequently, the identification-rate for both systems has also improved, however as the filters number is enlarged over and above 20, identifying unusual information from the identification signal has reduced the recognition rate. It is also seen that the performance of system using Mel Frequency Cepstral Coefficient and Gaussian mixture model reflects superior identification-rate as compare to the rest of the system

Table 1: I-R (identification rates based on filter coefficients)

No. of Filter Coefficients	MFCC and GMM IR (%)	MFCC and VQ IR (%)
12	83	81.5
16	85.5	84
20	87	86
25	84	83

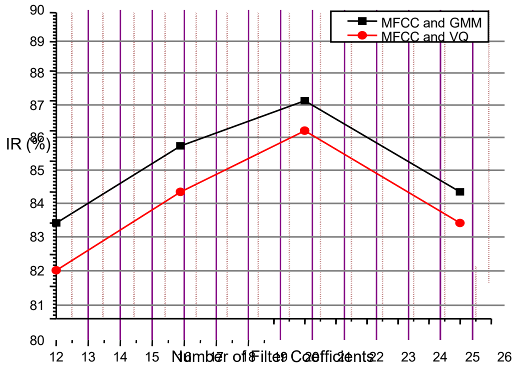

Fig. 3 Number of coefficients based Automatic Speaker Recognition analysis

B. Effect of Gender

10 of each male and female speaker have been considered to examine the effects of Gender on identification system rate. It was witnessed that the identification-system has a enhanced identification-rate for women's presenters as a result of high pitch speech signals. In the following table, it clearly shown that the individuality system with MFCC and GMM is a superior identification rate for women

Table 2: Gender based speaker identification rates

GENDER	MFCC and GMM IR (%)	MFCC and VQ IR (%)
Male	92	89.5
Female	93.5	91

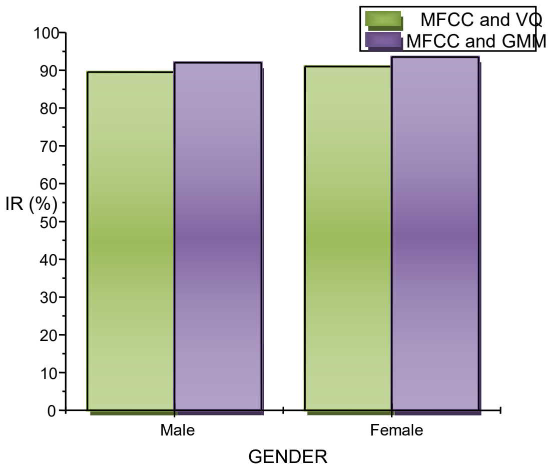

Fig. 4 Gender based Automatic Speaker Recognition analysis

C. Number of Speakers

This simulation also shows that the number of speakers also influences system performance. When the number of speakers is 5, both systems have a better identification rate of up to 95%, however as the number of speakers increased by 20, the system performance for speaker identification has reduced by 10 percent.

Table 3: Quantity of Speaker based identification rates

No. of Speakers	MFCC and GMM IR (%)	MFCC and VQ IR (%)
5	95	94
10	89	89
15	87	86
20	84.5	85

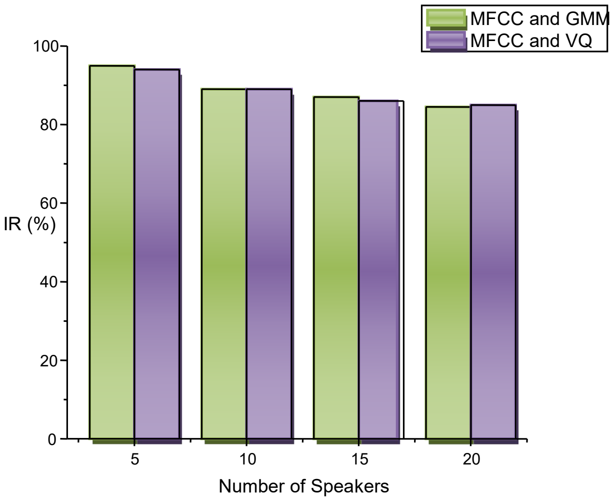

Fig. 5 Speakers based Automatic Speaker Recognition analysis

CONCLUSION

In this paper, the Mel Frequency Cepstral Coefficient and Vector Quantization speaker identification system have been examined according to the identification percentage in term of various factors for instance, filter-coefficients, gender of speaker and Speaker-count is selected to analyze system performance. The results have been shown that the system has a better identification rate for 20 speakers using MFCC and GMM It has also been observed that the increase in number of filters has been started to mitigate the performance of both systems. It is concluded form above analysis that the female speaker gives better identification rate in term of percentage as compared to male speaker in identification system.

REFERENCES

[1] Herbert Gish and Michael schimdt, ”Text Independent Speaker Identification” IEEE signal processing magazine, October 1994.

[2] Faundez-Zanuy M. and Monte-Moreno E. 2005 State-ofthe-art in speaker recognition , Aerospace and Electronic Systems Magazine, IEEE, 20(5), pp. 7-12

[3] K. Saeed and M. K. Nammous. 2005 Heuristic method of Arabic speech recognition, in Proc. IEEE 7th Int. Conf. DSPA, Moscow, Russia, pp. 528–530.

[4] Kumar, Chandar, Faizan ur Rehman, Shubash Kumar, Atif Mehmood, and Ghulam Shabir. "Analysis of MFCC and BFCC in a speaker identification system." In Computing, Mathematics and Engineering Technologies (iCoMET), 2018 International Conference on, pp. 1-5. IEEE, 2018.

[5] D. Olguin, P.A.Goor, and A. Pentland. 2009 Capturing individual and group behavior with wearable sensors, in Proceedings of AAAI Spring Symposium on Human Behavior Modeling

[6] S. B. Davis and P. Mermelstein. 1980 Comparison of Parametric Representation for Monosyllabic Word Recognition in Continuously Spoken Sentences, IEEE Trans. On ASSP, 28(4), pp. 357-365.

[7] R. Vergin, B, O Shaughnessy and A. Farhat. 1999 Generalized Mel frequency Cepstral coefficients for large-vocabulary speaker independent continuous-speech recognition, IEEE Trans. On ASSP,7(5), pp. 525-532.

[8] ur Rehman, Faizan, Chandar Kumar, Shubash Kumar, Atif Mehmood, and Umair Zafar. "VQ based comparative analysis of MFCC and BFCC speaker recognition system." In Information and Communication Technologies (ICICT), 2017 International Conference on, pp. 28-32. IEEE, 2017.

[9] S.Singh and Dr. E.G Rajan. 2007 A Vector Quantization approach Using MFCC for Speaker Recognition, International conference Systemic, Cybernatics and Informatics ICSCI under the Aegis of Pentagram Research Centre Hyderabad, pp. 786-790.

[10] K. Sri Rama Murty and B. Yegnanarayana. 2006 Combining evidence from residual phase and MFCC features for speaker recognition, IEEE Signal Processing Letters, 13(1), pp. 52-55.

[11] Prof. Ch.Srinivasa Kumar, Dr. P. Mallikarjuna Rao “Design Of An Automatic Speaker Recognition System Using MFCC, Vector Quantization And LBG Algorithm” international journal of computer science and Engineering (IJSCE); Aug 2011.

[12] Vibha Tiwari “MFCC and its applications in speaker recognition” International Journal on Emerging Technologies, 2010.

[13] Roma Bharti, Priyanka Bansal, “Real Time Speaker Recognition System using MFCC and Vector Quantization Technique”, International Journal of Computer Applications (0975 – 8887) Volume 117 – No. 1, May 2015.

[14] L. Rabiner and B. H. Jaung ,“Fundamentals of Speech recognition”, Prentice Hall Englewood Cliffs, New Jersey, 1993.

[15] Chakroborty, S., Roy, A. and Saha, G. 2007 Improved Closed set Text- Independent Speaker Identification by Combining MFCC with Evidence from Flipped Filter Banks , International Journal of Signal Processing, 4(2), pp. 114-122

[16] Reynolds, D.A.‘ Speaker identification and verification using Gaussian mixture speaker models’, Speech Communication, vol. 17, pp. 91-108,1995.