ISSN: 2349-6037

International Journal of Multidisciplinary and Scientific Emerging ResearcH (IJMSERH)

Volume 2, Issue 3, July 2014

Comparative Study of MFCC And LPC Algorithms for Gujrati Isolated Word Recognition

H. B. Chauhan

Dept. of Computer Scholar, BVM Engineering College, Vallabh Vidhyanagar, India

ABSTRACT: The study performs feature extraction for isolated word recognition using Mel-Frequency Cepstral Coefficient (MFCC) for Gujarati language. It explains feature extraction methods MFCC and Linear Predictive Coding (LPC) in brief. The paper compares the performances of MFCC and LPC features under Vector Quantization (VQ) method. The dataset comprising of males and females voices were trained and tested where each word has been repeated 5 times by the speakers. The results show that MFCC is performed better feature extractor for speech signals.

KEYWORDS: feature extraction, LPC, MFCC, VQ, Gujarati database

I. INTRODUCTION

The Speech recognition is the analytic subject of speech processing in machines. Human speech recognition is thousands of year's old and known better as automatic speech recognition (ASR). Speech recognition systems have been developed for the languages like Hindi [1] [2], Malayalam [3] [4], Tamil [5], Marathi [6], Telugu [7], Panjabi [8], Urdu [9], etc…, in India. Isolated speech recognition using MATLAB® is done for Gujarati words. In relative work [10] Dr. C. K. Kumbharana has completed Gujarati word detection for "ક, ચ, ર, ત and મ", using MFCC function.
The study is performed with two different feature extraction algorithms and used training and testing data from distinct words like આઠ (Eight), ત્રણ (Three) and ગુજરતી (Gujaraati), etc. Each speaker had spoken the 10 words with 0 to 10 numbers and some word with 5 utterances of each. So, for four speakers, total 200 utterances of the words were recorded. The isolated words in Gujarati were recorded using built in microphone of laptop using the RecordPad Software [11] and stored in .wav format. The data had been recorded in closed rooms where background noise was present. This kind of recording of the speech data in such noisy environment will be useful in robust automatic speech recognition system.

The paper is divided into five sections. Section I gives introduction. The feature extraction using MFCC and LPC is describes in Section II and Section III, respectively. The results are analysed in Section IV followed by conclusion and future work in Section V.

II. FEATURE EXTRACTION USING MFCC

Mel Frequency Cepstral Coefficient (MFCC) was introduced by Davis and Mermelstein in the 1980 AD. It is very common and one of the best method for feature extraction method especially for automatic speech and speaker recognition system. An application of hand gesture recognition, MFCC was used as feature extractor by converting input image into 1D signal with SVM classifier [12]. The MFCC coefficients can be used as audio classification features to improve the classification accuracy, is used for the music features, and then BPNN algorithm recognizes the music classes [13].

Before introduction of MFCCs, Linear Prediction Coefficients (LPCs) and Linear Prediction Cepstral Coefficients (LPCCs) and were the main feature type for ASR [14]. MFCC is used in speaker verification with speaker information like, contents and channels [15]. MFCCs are a feature widely used in automatic speech and speaker recognition. There is computation for extracting the cepstral features parameters from the Mel scaling frequency domain. The steps of MFCC are given bellows,

ISSN: 2349-6037
International Journal of Multidisciplinary and Scientific Emerging ResearcH (IJMSERH)
Volume 2, Issue 3, July 2014

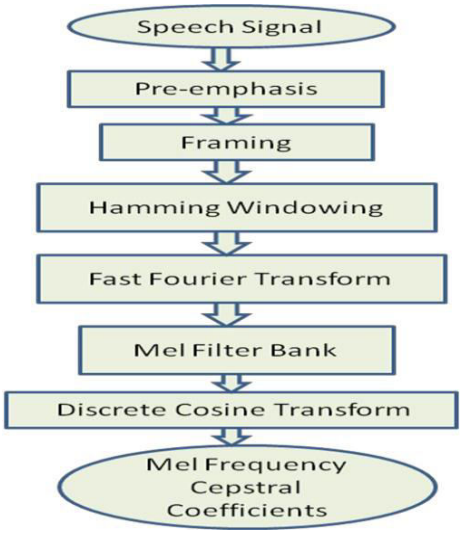


Figure-1 Flow of MFCC

As shown in figure-1, the signal is passed through very first stage of emphasizes which will increase the energy of the signal at higher frequency to compensate the high-frequency part that was suppressed during the sound production mechanism of humans. Now, the boosted signal is segmented into frames of 20~30 ms of the frame size with overlap of 1/3~1/2. Here, the sample rate is 8 kHz and the frame size is 256 sample points are used, then the frame duration is 256/8000 = 0.032 sec = 32ms. Each frame will be multiplied with a hamming window in order to keep the continuity of the first and the last points in the frame. MATLAB® provides the command for generating the curve of a Hamming window, also. There is FFT performed to obtain the magnitude frequency response of each frame which is assumed of periodic within frame. The triangular bandpass filters are used to extract an envelope like features. The multiple the magnitude frequency response by a set of triangular bandpass filters to get the log energy of each triangular bandpass filter which will give nonlinear perception for different tones or pitch of voice signal. The Mel frequency M(F) related to the common linear frequency f is by the following equation[16]:

$$M(F) = 1125 * \ln (1 + f / 700) \quad \ldots \ldots \ldots (1)$$

Then Discrete Cosine Transform (DCT) is applied on the log energy to have different mel-scale cepstral coefficients. The DCT converts the signal from frequency domain into a time domain. Because, the features are similar to cepstrum, it is referred to as the mel-scale cepstral coefficients. MFCC can be used as the feature for speech recognition. For better performance can generated by adding the log energy and perform delta operation. As new features in MFCC, Delta cepstrum can be generated which has advantage in the time derivatives of energy of signal. It can use for finding the velocity and acceleration of energy with MFCC. MFCC base speaker recognition system in MATLAB® can significantly increase the accuracy rate of training and recognition, and reduce the data required by calculation at higher recognition rate [17].

III. FEATURE EXTRACTION USING LPC

Linear predictive coding (LPC) method was developed in the 1960s by [18] and being used for speech vocal tracing because it represents vocal tract parameters and the data size are very suitable for speech compression. [19]. In this paper, a modified LPC coefficients approach is used for speech processing for representing the spectral envelope of speech in compressed form. This method gives encoding good quality speech at low bit rate and provides accurate estimates of speech parameters by describing the intensity, the residue signal. The information can be stored or transmitted somewhere else. A dialect-independent wavelet transform (WT) is based Arabic digits classier is proposed wavelet transformed with the LPC and the classier by probabilistic neural network (PNN) [20]. There was speaker's classification between male and female using nearest neighbour method, calculating Euclidean distance from the Mean value of Males and Females of the generated mean. There are 13 MFCCs and 13 LPCs coefficients are computed for Audio portion is extracted from Indian video songs [21].

There are basic four steps of LPC processor, Pre-emphasis where the digitized speech signal is flatten to make less susceptible to finite precision effects signal processing. In second step of frame blocking, the output signal is blocked

ISSN: 2349-6037
International Journal of Multidisciplinary and Scientific Emerging ResearcH (IJMSERH)
Volume 2, Issue 3, July 2014

into frames of *N* samples, with adjacent frames separated by no. of M samples. In Windowing, there is to window each individual frame so as to minimize the signal discontinuities at the starting and ending of each frame, same as in MFCC. Autocorrelation Analysis will auto correlate each frame of windowed signal in order to give highest autocorrelation value. In final step of LPC Analysis converts each frame of $p + 1$ autocorrelations into LPC parameter set by using Durbin's method. In equation, each sample of the signal $\tilde{x}(n)$ is expressed as a linear combination of the previous samples $x(n - i)$ it is called linear predictive coding [22]. Here, a_i are the predictor coefficients.

$$\tilde{x}(n) = \sum_{i=1}^{p} a_i \; x(n - i) \quad \ldots \ldots \ldots (2)$$

The error estimation for true signal value x(n) in one dimensional linear prediction is given by [22],

$$e(n) = x(n) - \tilde{x}(n) \quad \ldots \ldots \ldots (3)$$

For signals in multidimensional the error metric is given by,

$$e(n) = \| x(n) - \tilde{x}(n) \| \quad \ldots \ldots \ldots (4)$$

LPC and MFCCs coefficients combined can use for dynamic or runtime feature extraction. These both combine can use as feature vector for Emotions of speaker identified like Angry, Boredom, neutral, happy and sad [23]. The Hindi alphabet is done for emotion identification using syllables occur in the pattern Consonant Vowel Consonant (CO3VCO3) [24].

IV. RESULT ANALYSIS

Vector quantization (VQ) is used for comparing the trained data with new entered input data. It is a classical quantization technique that allows the modeling of probability density functions by the distribution of vectors. It divides a large set of points called vectors into groups having approximately the same number of points closest to them. The density matching property of VQ is powerful for identifying the density of large and high-dimensioned data [25]. All data points are represented by the index of their closest centroid which can be used for lossy data correction and density estimation. Vector quantization is the self organizing map model.

Using MFCC and LPC Features are extracted for the Gujarati words. The training data sets for the vector quantization are obtained by recording utterances of Gujarati words. The entered data are compared with already stored datasets. The comparison of both algorithms for three words is given in below charts.

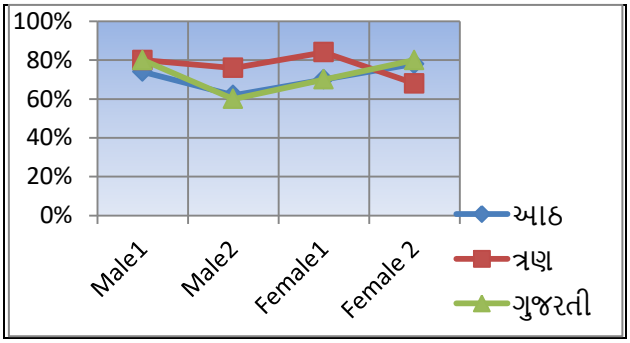


Chart -1 Result using LPC

The recognition accuracy is achieved by LPC is above 85%, shown in chart-1. The recognition accuracy is achieved by MFCC is more above 95%, shown in Chart-2.

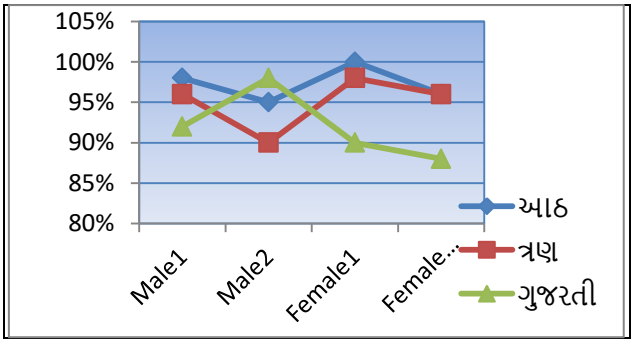
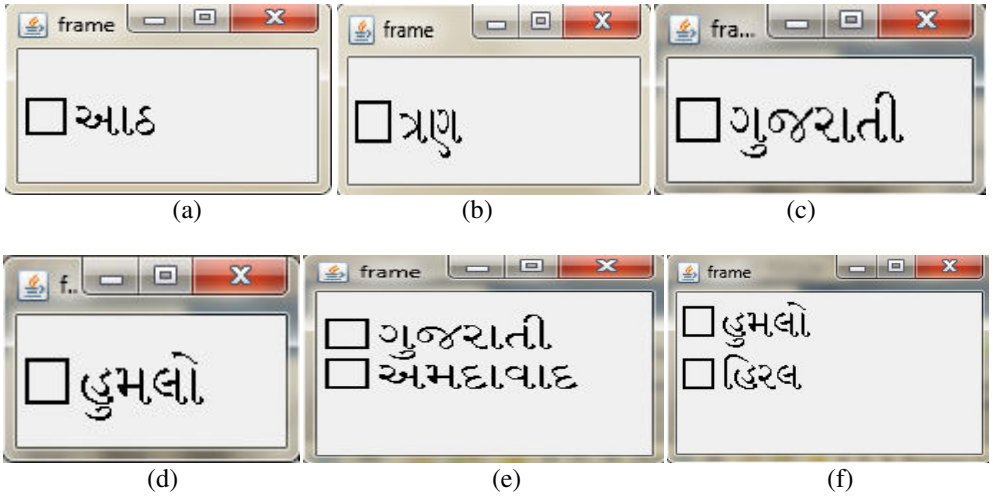


Chart -2 Result using MFCC

ISSN: 2349-6037
International Journal of Multidisciplinary and Scientific Emerging ResearcH (IJMSERH)
Volume 2, Issue 3, July 2014

So, MFCC can bring better feature for Gujarati language tutor application in speech recognition. The input speech that matches trained database is converted into related text, and results are shown below figure 2 (a) describes digit 8,(b) describes digit 3, (c) describes word "Gujarati",(d) describes word "attack", (e) describes words "Gujarati" and "Ahmadabad", (f) describes words "attack" and "Hiral" (a name) in Gujarati language.



(a) (b) (c)

(d) (e) (f)

Figure-2 Gujarati Output Text

V. CONCLUSION AND FUTURE WORK

The approach is to implement for isolated speech recognition system for Gujarati language. The MFCC and LPC are used as speech feature extractor. The algorithms are followed by VQ method for testing, helps to conclude that MFCC is more accurate feature extractor for verity speech signals. The present work was limited to phonemes of Gujarati only. The further study can be done for continuous speech recognition using MFCC Features extraction algorithm and Hidden Markov Model (HMM) for testing and modeling purpose. The very large vocabulary speech recognition (VLSR) using MFCC with PLP Features extraction algorithm and HMM combined with Artificial Neural network (ANN) for better classification.

ACKNOWLEDGMENT

A Special thanks to Prof. Dr. Mayur M.Vegad of BVM engineering college, who insist a lot towards sincere and pre-eminent work.

REFERENCES

[1] Gaurav, Devanesamoni Shakina Deiv, Gopal Krishna Sharma, Mahua Bhattacharya, "Development of Application Specific Continuous Speech Recognition System in Hindi", *Scientific Research Journal of Signal and Information Processing, 394-401, Vol*-3, August 2012

[2] Ankit Kuamr, Mohit Dua, Tripti Choudhary, "Continuous Hindi Speech Recognition Using Gaussian Mixture HMM", *IEEE Students' Conference on Electrical, Electronics and Computer Science*, 2014

[3] Cini Kurian, Kannan Balakrishnan, "Malayalam Isolated Digit Recognition using HMM and PLP cepstral coefficient", *International Journal of Advanced Information Technology (IJAIT)*, Vol. 1, No.5, October 2011DOI: 10.5121/ijait.2011

[4] Cini Kurian, Kannan Balakriahnan, "CONTINUOUS SPEECH RECOGNITION SYSTEM FOR MALAYALAM LANGUAGE USING PLP CEPSTRAL COEFFICIENT", *International Journal of Computing and Business Research (IJCBR) ISSN* : 2229-6166 Volume 3 Issue 1 January 2012

[5] M. Chandrasekar, M. Ponnavaikko," Tamil speech recognition: a complete model", Electronic Journal, Technical Acoustics, http://www.ejta.org 2008

[6] Siddheshwar S. Gangonda, Dr. Prachi Mukherji, "Speech Processing for Marathi Numeral Recognition using MFCC and DTW Features " *International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 National Conference on Emerging Trends in Engineering & Technology ,VNCET-30 Mar'12*

[7] D. Nagaraju et al., "Emotional Speech Synthesis for Telugu", *Indian Journal of Computer Science and Engineering (IJCSE)*, ISSN: 0976-5166, Vol. 2 No. 4, Aug -Sep 2011

[8] Vivek Sharma, Meenakshi Sharma," A Quantitative Study Of The Automatic Speech Recognition Technique", *International Journal of Advances in Science and Technology (IJAST) Vol I, Issue I*, December 2013
[9] Javed Ashraf, Naveed Iqbal, Naveed Sarfraz Khattak, Ather Mohsin Zaidi, "Speaker Independent Urdu Speech Recognition Using HMM", *Natural Language Processing and Information Systems Lecture Notes in Computer Science*, Volume 6177, 2010, pp 140-148
[10] C. K. Kumbharana , "Speech Pattern Recognition for Speech To Text Conversion", *etheses .saurashtrauniversity . edu /337/1/ kumbharana _ck_ thesis _cs .pdf* by CK Kumbharana - 2007
[11] RecordPad Sound Recording Software - NCH Software *www.nch.com.au/recordpad*
[12] Leena R Mehta, S.P.Mahajan, Amol S Dabhade, "COMPARATIVE STUDY OF MFCC AND LPC FOR MARATHI ISOLATED WORD RECOGNITION SYSTEM" *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 2, Issue 6,* June 2013
[13] Shikha Gupta, Jafreezal Jaafar Wan, Fatimah wan Ahmad, Arpit Bansal, "FEATURE EXTRACTION USING MFCC", *Signal & Image Processing : An International Journal (SIPIJ)* Vol.4, No.4, August 2013
[14] J.X. Jin, Debnath Bhattacharyya, "Research on Music Classification Based on MFCC and BP Neural Network", *Proceedings of the 2nd International Conference on Information, Electronics and Computer*, part of the series AISR, ISSN 1951-6851, volume 59, 2014
[15] Archit Kumar, Charu Chhabra, "Intrusion detection system using Expert system (AI) and Pattern recognition (MFCC and improved VQA)", *International Journal of Advance Research in Computer Science and Management Studies,* Volume 2, Issue 5, May 2014
[16] Jun Wang; Lantian Li; Dong Wang; Zheng, T.F., "Research on generalization property of time-varying Fbank-weighted MFCC for i-vector based speaker verification," *Chinese Spoken Language Processing (ISCSLP), 2014 9th International Symposium on* , vol., no., pp.423,423, 12-14 Sept. 2014
[17] Ganchev T, Fakotakis N, Kokkinakis G. "Comparative evaluation of various MFCC implementations on the peaker verification task", in: Proceedings of the Specom. 2005;1:191–194.
[18] Chenchen Huang Wei Gong Wenlong Fu, Dongyu Feng, "Research of speaker recognition based on the weighted fisher ratio of MFCC," *Mechatronic Sciences, Electric Engineering and Computer (MEC), Proceedings 2013 International Conference on* , vol., no., pp.904,907, 20-22 Dec. 2013
[19] K. Daqrouq, M. Alfaouri, A. Alkhateeb, E. Khalaf1 and A. Morfeq, "Wavelet LPC with Neural Network for Spoken Arabic Digits Recognition System", *British Journal of Applied Science & Technology,* ISSN: 2231-0843 ,Vol.: 4, Issue.: 8, March 2014
[20] K Rakesh, S Dutta, K Shama , "Gender Recognition using speech processing techniques in LABVIEW", *International Journal of Advances in Engineering & Technology*, May 2011
[21] Tushar Ratanpara, Narendra Patel "Singer Identification Using MFCC and LPC Coefficients from Indian Video Songs*", Emerging ICT for Bridging the Future - Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1, Advances in Intelligent Systems and Computing*, pp 275-282, Volume 337, 2015
[22] K. Ravi Kumar, V.Ambika, K.Suri Babu, "Emotion Identification From Continuous Speech Using Cepstral Analysis", *International Journal of Engineering Research and Applications (IJERA),* Vol. 2, Issue 5, pp.1797-1799, September- October 2012
[23] Bansal S.; Dev A., "Emotional hindi speech database," *Oriental COCOSDA held jointly with 2013 Conference on Asian Spoken Language Research and Evaluation (O-COCOSDA/CASLRE), 2013 International Conference , vol.,*4, Nov. 2013
[24] Soong, Frank K.; Rosenberg, Aaron E., Juang, Bling-Hwang, Rabiner, Lawrence R., "Report: A vector quantization approach to speaker recognition," *AT&T Technical Journal Murray Hill, New Jersey ,* vol.66, no.2, pp.14,26, March-April 1987
[25] Tarun Pruthi, Sameer Saksena, Pradip K Das, , "*Swaranjali: Isolated Word Recognition for Hindi Language using VQ and HMM*" Journal of Computing and Business Research, 1993