ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 10, October 2016

Comparative Study on Feature Extraction Methods in Speech Recognition

Kaushal K. Mistry [1], Sunil S. Shah [2]

Sr. Lecturer, Department of Computer Engineering, B & B Institute of Technology, V. V. Nagar, Gujarat, India[1]

Sr. Lecturer, Department of Computer Engineering, B & B Institute of Technology, V. V. Nagar, Gujarat, India[2]

ABSTRACT: Speech recognition is an analysis side of the main topics machine speech processing. The synthesis side may be called speech production. These two taken together allow computers to cooperate with spoken language. My study specializes in isolated word speech recognition in Gujarati language. Special message recognition, in humans, is 1000s of years old. On our planet it would be traced backed millions of years to your dinosaurs. Our topic might far better be called automatic speech recognition (ASR). I give a limited survey of MFCC (Mel Frequency Cepstral Coefficient). A very simple computer experiment, using MatLab, into isolated word speech recognition is described in certain detail. I experimented with distinctive recognition algorithms and implemented training and testing data with two distinct vocabularies. My exercise and testing data was built-up and recorded with both men and female voices.

KEYWORDS: Artificial Intelligence, Speech Reorganization, MFCC, LPC.

I. INTRODUCTION

Over the last decade, speech recognition systems have developed into increasingly used in automated solutions with spoken language interfaces. Systems with spoken language interfaces generate a significant contribution to the detection of local language interfaces inside Indian scenario. There has been excellent progress in speech technology, with today's state-of-the-art systems to be able to transcribe unrestricted broadcast news special message data with good accuracy. Nevertheless, the number of speech popularity systems that support Indian languages is incredibly small and especially for Gujarati yet speech recognition system is not really available.

II. ISSUES IN AUTOMATIC TRANSCRIPTION OF CONTINUOUS SPEECH

There are several issues which help the automatic transcription of continuous speech a hard task. Translation of a continuous speech signal to a sequence of words is a hard task, as continuous speech doesn't need any natural pauses in concerning sub-word units. The conventional method of creating a speech recognizer for any language requires a substantial amount of segmented and labelled speech corpus with regard to training. However obtaining such a corpus can be a labour intensive and time consuming process To reduce the charge, both in terms of people effort and financial needs, whether it is required to adapt a recognition system for a new task or other words, the quality of automatic segmentation together with labelling is potentially of terrific significance for continuous speech popularity systems, as word-error rate greatly will depend on the accuracy of segmentation together with sub-word unit recognition.

The occurrence of words and grammatical improvements in spoken and written words are vastly different. Therefore, speech may not be viewed as an exact interpretation of written language into its spoken form. Channel variability brought on by the prevalent noise and entry to different microphones makes speech popularity a formidable task.

As speech recognition involves matching of speech signal with the already existing group of designs, the lexicon needs to end up kept small, in order to lower the search space. This factor another problem called out-of-vocabulary, so that the intended unit is not obtained in the lexicon. Automatic speech recognition (ASR) should also manage to handle out-

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 10, October 2016

of-vocabulary issues in a competent manner. Speaker variability, variation in speaking style, gender in the speaker and speaking rate, additionally affect speech recognition performance.

The main application of Artificial Intelligence will grow in technology. The notion of AI established fact due to its popularity with science fiction movies which outline humans interacting with machines as they simply would with other humans. Speech and gestures are definitely the natural means of communication made use of by humans to interact with the other. Speech Recognition makes it possible that you speak to a computer. Special message recognition software is the technological know-how that transforms spoken words inside alphanumeric text and navigational requires. Speech Recognition is used with legal and medical transcription, that generation of subtitles for stay sports and current affairs software programs on television. In naturally talked language, there are no stoppages between words, so it is difficult to get a computer to decide where phrase boundaries lie. Automatic speech recognition is a process by which a laptop computer maps an acoustic speech transmission to text.

III. FEATURE EXTRACTION USING MFCC AND LPC

Feature extraction of Mel Frequency Cepstral Coefficient (MFCC) has been introduced by Davis and Mermelstein inside 1980 AD. It is quite normal and one of the most practical methods for feature extraction method designed for automatic speech and speaker popularity system. An application of give gesture recognition, MFCC was implemented as feature extractor by remodelling input image into 1D transmission with SVM classifier [4]. The MFCC coefficients can be installed as audio classification features to boost the classification accuracy, is raised for the music features, and in that case BPNN algorithm recognizes the popular music classes.

Feature extraction of Linear predictive coding (LPC) method originated in the 1960s by [10] and available for speech vocal tracing since it represents vocal tract parameters along with the data size are very well suited for speech compression[11]. In the following paper, a modified LPC coefficients approach is utilized for speech processing for addressing the spectral envelope of special message in compressed form. This method gives encoding top quality speech at low bit rate and accurate estimates of speech variables by describing the intensity, that residue signal. The information may be stored or transmitted somewhere better. A dialect independent wavelet transforms (WT) situated Arabic digits classier is proposed wavelet transformed along with the LPC and the classier just by probabilistic neural network (PNN) [12]. There would be speaker's classification between male together with female using nearest neighbour process, calculating Euclidean distance from the Mean value of Men and women of the generated mean.

There are actually 13 MFCCs and 13 LPCs coefficients are generally computed for Audio portion is usually extracted from Indian video tunes.

Different feature extraction methods are as mentioned in below table-1

Sr.No	Method	Property	Procedure for Implementation
1	Principal Component analysis (PCA)	Non linear feature extraction method, Linear map, fast, eigenvector-based	Traditional, eigenvector base method also known as karhuneu-Loeve expansion; good for Gaussian data
2	Linear Predictive coding (LPC)	Static feature extraction method,10 to 16 lower order coefficient	It is used for feature Extraction at lower Order
3	Cepstral Analysis	Static feature extraction method, Power spectrum	Used to represent spectral envelope
4	Mel-frequency	Power spectrum is	This method is used for find our

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 10, October 2016

	cepstrum co-efficient (MFCCs)	computed by performing Fourier Analysis	Features
5	Wavelet	Better time resolution than Fourier Transform	It replaces the fixed bandwidth of Fourier transform with one proportional to frequency which allow better time resolution at high frequencies than Fourier Transform
6	Dynamic feature extractions i)LPC ii)MFCCs	Acceleration and delta coefficients i.e. II and III order derivatives of normal LPC and MFCCs coefficients	It is used by dynamic or runtime Feature
7	RASTA filtering	For Noisy speech	It is find out Feature in Noisy data
8	Integrated Phoneme subspace method (Compound Method)	A transformation based on PCA+LDA+ICA	Higher Accuracy than the existing Methods

Table 1: Feature Extraction Methods

IV. DETAILED DISCUSSION OF MFCC AND LPC

Feature extraction using MFCC in Details

As shown in Fig 1, in advance of introduction of MFCCs, Linear Conjecture Coefficients and Linear Conjecture Cepstral Coefficients and were the most crucial feature type for automatic special message recognition [6]. MFCC is popularly applied to speaker verification with speaker information like, contents and channels. [7] MFCCs are feature widely used in instant speech and speaker recognition. There does exist computation for small extracting the Cepstral features parameters in the Mel scaling frequency domain. The steps of MFCC get bellows shown in figure-1, the signal is passed through to start with stage of emphasizes which raises the energy of the transmission at higher frequency to compensate the high-frequency part that's suppressed during the sound output mechanism of humans. Now, the boosted signal is segmented inside frames of 20~30 ms in the frame size with overlap with 1/3~1/2. Here, the sample rate is 8 kHz along with the frame size is 256 sample points are utilized, then the frame duration is usually 256/8000 = 0. 032 Securities and Exchange Commission's = 32ms. Each frame will be increased with a hamming window to keep the continuity of the first along with the last points in the mode. MATLAB ® also provides that command for generating the curve on the Hamming window. There is FFT performed to choose the magnitude frequency response of each frame that's assumed of periodic within mode. The triangular band pass filters are utilized to extract an envelope enjoy features. The multiple the magnitude frequency response by the set of triangular band pass filters to find the log energy of each triangular band pass filter that can give nonlinear perception for several tones or pitch of words signal. The Mel frequency M(F) in connection with the common linear frequency f is by way of the following equation:

M(F) = 1125 * ln (1 + free p / 700) … … …. [8]

Then Discrete Cosine Transform (DCT) is applied relating to the log energy to have several Mel-scale Cepstral coefficients. The DCT converts the signal from frequency domain to a time domain. Because, the features act like cepstrum, it is referred to as being the Mel-scale Cepstral coefficients. MFCC can be installed as the feature for special message recognition. For better performance can generated with the addition of the log energy and complete delta operation. As new options in MFCC, Delta cepstrum may be generated which has advantage inside time derivatives of energy with signal. It can use for simply finding the velocity and acceleration of electrical power with MFCC. MFCC

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 10, October 2016

base speaker recognition process in MATLAB® can significantly improve the accuracy rate of training together with recognition, and reduce the info required by calculation at better recognition rate [9].

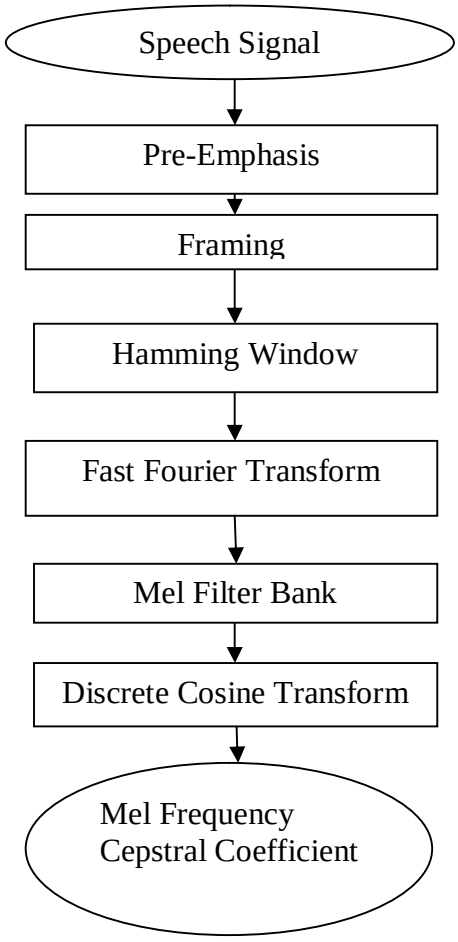

Fig 1: Steps of MFCC Feature Extraction Method

Feature extraction using LPC in Details

As shown in Fig 2, there are actually basic four steps of LPC model, Pre-emphasis where the digitized speech signal is flatten to produce less susceptible to finite exquisitely detailed effects signal processing. In minute step of frame blocking, that output signal is blocked inside frames of N samples, using adjacent frames being separated just by M samples. In Windowing, there does exist to window each individual frame so that it will minimize the signal discontinuities in the starting and ending of just about every frame, same as in MFCC. Auto corrections analysis will auto correlate just about every frame of windowed signal so as to give highest auto correction value. With final step of LPC Test converts each frame of k + 1 auto correction into LPC parameter set by employing Durbin's method.

In equation, each sample in the signal x^ ( ) is expressed for a linear combination of the old samples x(n − i) it can be called linear predictive coding. [14] These, ai are the predictor coefficients.

^( ) = ai x(n − i) … … … [14]

LPC and MFCCs coefficients combined incorporate the use of for dynamic or runtime attribute extraction. These both combine incorporate the use of as feature vector for attachments of speaker identified like mad, boredom, neutral,

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 10, October 2016

happy and sorry. [15] The Hindi alphabet is usually done for emotion identification using syllables occur inside pattern Consonant Vowel Consonant (CO3VCO3). [16].

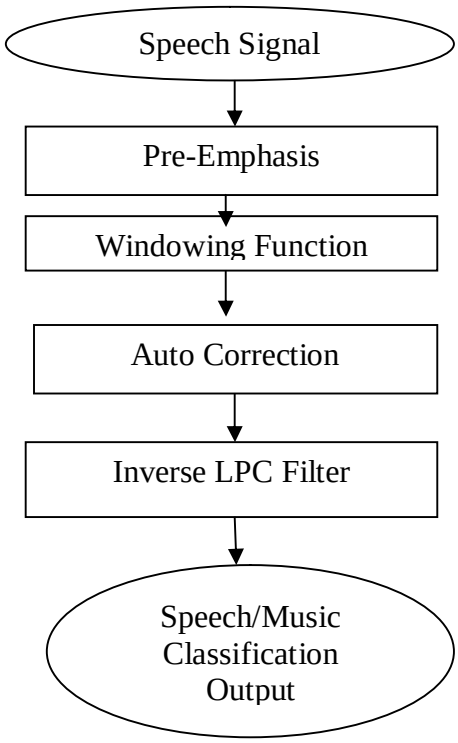

Fig 2: Steps of LPC Feature Extraction Method

V. CONCLUSION

The proposed approach may be to implement for isolated speech popularity system in Gujarati. The MFCC and LPC are utilized as speech feature extractor parameter. The identical testing helps to conclude that MFCC is usually more accurate feature extractor special message signals. The present work was tied to phonemes of Gujarati only.

REFERENCES

[1] A. A. M. Abushariah and T. S. Gunawan, "English Digits Speech Recognition System Based on Hidden Markov Models," *Proceeding Int. Conf. Comput. Commun. Eng. (ICCCE 2010)*, no. May, pp. 11–13, 2010.
[2] R. Applications, O. Cheng, W. Abdulla, Z. Salcic, and S. Member, "Hardware – Software Codesign of Automatic Speech Recognition System for Embedded," *Ieee Trans. Ind. Electron.*, vol. 58, no. 3, pp. 850–859, 2011.
[3] J. Melorose, R. Perroy, and S. Careas, "No Title No Title," *Statew. Agric. L. Use Baseline 2015*, vol. 1, pp. 682–686, 2015.
[4] Q. Fan and D. Wang, "A statistical speech recognition of ningbo dialect monosyllables," *Proc. 2010 IEEE Int. Conf. Intell. Syst. Knowl. Eng. ISKE 2010*, pp. 266–269, 2010.
[5] T. N. Sainath, B. Ramabhadran, M. Picheny, D. Nahamoo, and D. Kanevsky, "Exemplar-based sparse representation features: From TIMIT to LVCSR," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 19, no. 8, pp. 2598–2613, 2011.
[6] S. T. Pan and X. Y. Li, "An FPGA-based embedded robust speech recognition system designed by combining empirical mode decomposition and a genetic algorithm," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 9, pp. 2560–2572, 2012.
[7] U. N. Wisesty, T. H. Liong, and Adiwijaya, "Indonesian speech recognition system using Discriminant Feature Extraction - Neural Predictive Coding (DFE-NPC) and Probabilistic Neural Network," *Proceeding - 2012 IEEE Int. Conf. Comput. Intell. Cybern. Cybern. 2012*, pp. 158–162, 2012.
[8] W. Zhang, D. Wu, Z. Xue, T. Zhang, D. Wang, and C. Hou, "Design and implementation of voice signal processing system based on DSP and FPGA," *2013 IEEE Third Int. Conf. Inf. Sci. Technol.*, pp. 1602–1605, 2013.

ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 10, October 2016

[9] A. Ai-ghanim, N. Al-oboud, R. Ai-haidary, S. A. S. Altammami, and H. A. Mahmoud, "I See What You Say (ISWYS): Arabic Lip Reading System," pp. 11–17, 2013.
[10] C. Lee, "APPLICATIONS OF DYNAMIC PROGRAMMING TO SPEECH Dynamic programming (DP)
[11] N. Morgan and H. A. Bourlard, "Neural networks for statistical recognition of continuous speech," *Proc. IEEE*, vol. 83, no. 5, pp. 742–770, 1995.
[12] [12]H. K. Sze, S. Saileh, and A. Zuri, "New developed speech recognition education software," pp. 1701–1704, 2002.