International Journal of Scientific Engineering and Research (IJSER)
www.ijser.in
ISSN (Online): 2347-3878, Impact Factor (2015): 3.791

Comparative Analysis of MFCC, LFCC, RASTA -PLP

P. Prithvi[1], Dr. T. Kishore Kumar[2]

[1]National Institute of Technology, Warangal, Telangana – 506 004, India
[2]National Institute of Technology, Warangal, Telangana – 506 004, India

Abstract: ***A human's voice has various parameters that convey vital information. Speech feature extraction follows preprocessing of the speech signal. This process makes certain that the speech feature extraction contains true and accurate information that reveals the emotions of the speaker. In this paper, we present a study and comparison of feature extraction methods like Mel-Frequency Cepstral Co-efficient (MFCC), Linear Predictive Cepstral Co-efficient (LPCC), and Relative Spectral Analysis Perceptual Linear Prediction (RASTA-PLP). These techniques will be analyzed for their suitability and usage in recognition of the speaker. The experimental results show that the better recognition rate is obtained for MFCC as compared to LPCC and RASTA-PLP.***

Keywords: MFCC, LPCC, RASTA-PLP, Pre-processing

1. Introduction

Speech is one of the natural means of communication between human beings and several machines have been developed to analyse, recognize, and produce speech. Speech technology is rapidly evolving and a number of tools have been developed for improved implementation.

Speech recognition is an area that involves developing systems that recognize spoken words and allows a computer to convert the captured acoustic speech signal to word(s). Automatic speech recognition (ASR) is one of the fastest growing fields. ASR allows the computer to convert the speech signal into text or commands through the process of identification and understanding. Speech recognition is connected to many fields of physiology, psychology, linguistics, computer science and signal processing, and is even linked to the person's body language, and its objective is to achieve natural language communication between human and computer. ASR finds numerous applications such as automatic call processing in telephone networks, and query based information systems that provide updated travel information, stock price quotations, weather reports, data entry, voice dictation, access to information: travel, banking, avionics, automobile portal, speech transcription, supermarket, railway reservations etc. [1][2]

2. Speech Pre-Processing

The common steps involved to prepare speech for feature extraction are [6]:

- Sampling
- Pre-emphasis
- Framing
- Windowing

2.1 Pre-emphasis

Pre-emphasis of the speech signal has become a standard pre-processing step at high frequencies. Pre-emphasis reduces the dynamic range of the speech spectrum, enabling to estimate the parameters more accurately. At the synthesis stage, speech synthesised from the parameters representing the pre-emphasised speech is deemphasised. This step processes the passing of signal through a filter which emphasizes higher frequencies. This increases the energy of signal at higher frequency.

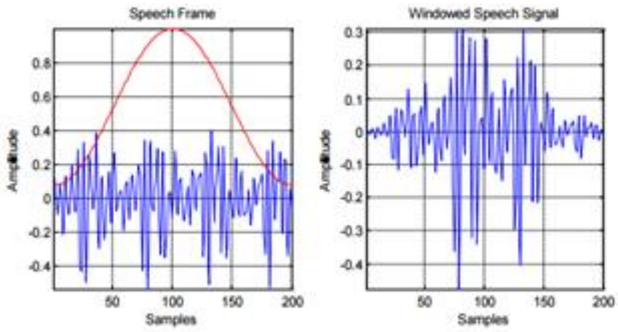

Figure 1: Framing and Windowing

2.2 SNR estimation

First order High-pass filter (FIR) is used to flatten the speech spectrum and compensate for the unwanted high frequency part of the speech signal. The following equation describes the transfer function of FIR filter in z-domain.

$$y[n] = x[n] - A.x[n-1] \qquad (1)$$

where x[n]:input speech signal

x[n-1]:previous speech signal
A:pre-emphasis factor, which is chosen as 0.975

2.3 Framing and windowing

In order to ensure the smoothing transition of estimated parameters from frame to frame, pre-emphasized signal y[n] is blocked into 200 samples with 25 ms frame long and 10 ms frame shift. In addition to that hamming window as shown in equation was selected and applied on each frame in order to minimize the signal discontinuities at the beginning and the end of each frame as shown in equation:

$$w(n) = 0.54 - 0.46\left(\frac{2\pi n}{N} - 1\right), 0 \leq n \ll N \qquad (2)$$

Licensed Under Creative Commons Attribution CC BY

3. Speech Feature Extraction

Feature extraction is used to convert the acoustic signal into a sequence of acoustic feature vectors that carry a good representation of input speech signal. These features are then used to classify and predict new words. To increase the feature evidence of dynamic coefficients, delta and delta can be devoted by adding the first and second derivative approximation to feature parameters. In this project, several conventional and hybrid feature extraction techniques were simulated and tested using MATLAB software to generate parameter coefficients.

3.1 Mel Frequency Cepstrum Coefficients (MFCC)

MFCC is the most dominant method used to extract spectral features. MFCCs analysis is started by applying Fast Fourier Transform (FFT) on the frame sequence in order to obtain certain parameters, converting the power spectrum to a Melfrequency spectrum, taking the logarithm of that spectrum, and computing its inverse Fourier transform as shown in figure [3]

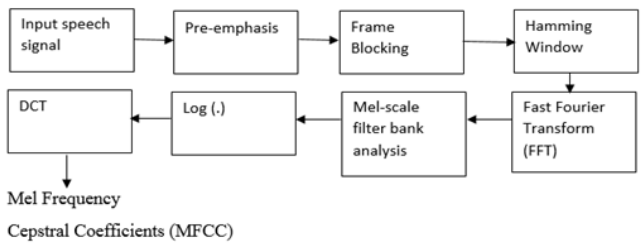

Figure 2: MFCC block diagram

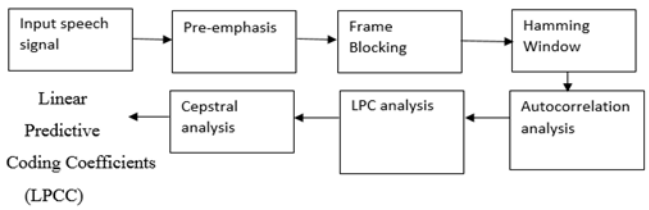

Figure 3: LPCC block diagram

3.2 Linear Predictive Cepstrum Coefficients (LPCC)

LPCC is one of the earliest algorithms that worked at low bitrate and represented an attempt to mimic the human speech and was derived using auto-correlation method. Autocorrelation technique is almost an exclusively used method to find the correlation between the signal and itself by auto-correlating each frame of the windowed signal using following equation as shown in figure. [7]

$$R[i] = \sum_{n=1}^{N_w-1} s_w(n).s_w(n-1), 0 \leq i \leq p \quad (3)$$

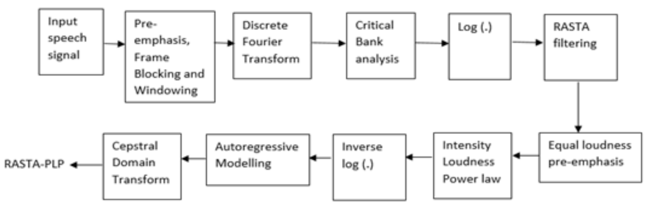

Figure 4: RASTA-PLP block diagram

3.3 Relative Spectral Analysis - Perceptual Linear Prediction (RASTA-PLP)

A special band-pass filter was added to each frequency sub-band in traditional PLP algorithm in order to smooth out short-term noise variations and to remove any constant offset in the speech channel. The following figure shows the most processes involved in RASTA-PLP which include calculating the critical-band power spectrum as in PLP, transforming spectral amplitude through a compressing static nonlinear transformation, filtering the time trajectory of each transformed spectral component by the band pass filter using equation as given below, transforming the filtered speech via expanding static nonlinear transformations, simulating the power law of hearing, and finally computing an all-pole model of the spectrum. [4][10]

$$H(z) = (0.1)\left(\frac{2+z^{-1}-z^{-3}-2z^{-4}}{z^{-4}(1-0.98z^{-1})}\right) \quad (4)$$

4 Results and Conclusions

In this paper, we analyzed various feature extraction techniques like MFCC, LPCC, RASTA-PLP for speech recognition. Different stages of speech pre-processing were also studied and implemented.

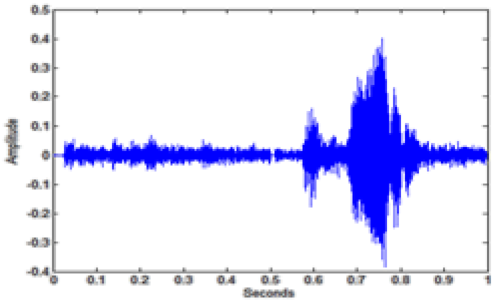

Figure 5: Input speech signal for the word HELLO

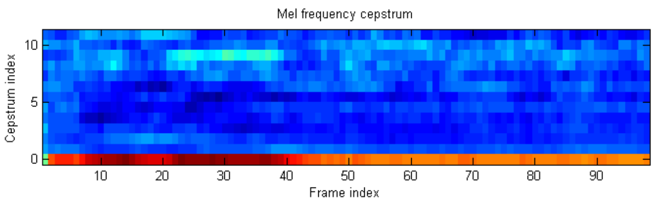

Figure 6: MFCC feature extraction.

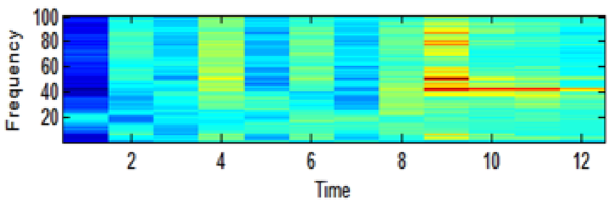

Figure 7: LPCC feature extraction

Licensed Under Creative Commons Attribution CC BY

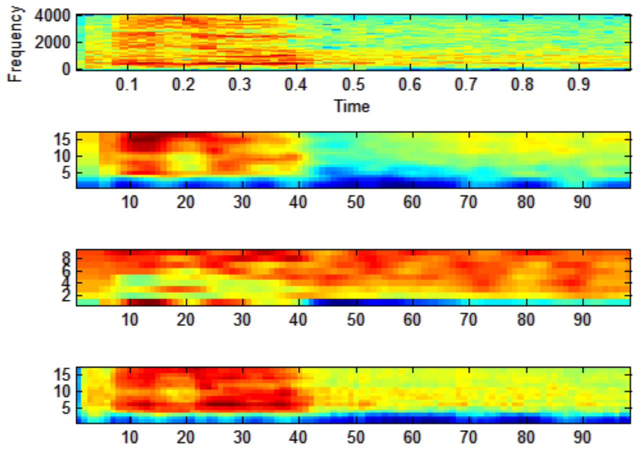

Figure 8: RASTA-PLP feature extraction

The two most important factors that affect the speech recognition process are the quality of speech and the size of the codebook. Varying the order of LPCC affects the speech recognition process. Increasing the order improves the rate of speech recognition. However, when higher coefficients are used, the speech recognition rate decreases. On the other hand, varying the order does not have significant impact on speech recognition rate in case of MFCC. The recognition rate increases and remains constant. MFCC yields better performance than LPCC and RASTA-PLP when the codebook size is small. Also, MFCC is robust to noise. To conclude, MFCC has better recognition rate and is robust to noise and spectral estimation errors when higher order coefficients are used. [5][8][9]

The results show that MFCC performs better than LPCC, RASTA-PLP in speech recognition. Also, MFCC gives consistent results and is robust to noise due to the fact that it is based on human perception of speech. [11][12].

TABLE I
COMPARISON OF MFCC, LPCC, RASTA-PLP

Method	Property	Remarks
MFCC	Filterbank coefficients: It is the result of short-term energy spectrum and expressed on Mel-scale which is linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz.	The MFCC reduces the frequency information of the speech signal into a small number of coefficients. It is easy and relatively fast to compute. More information about lower frequencies than higher frequencies due to mel spaced filter banks, hence behaves more like a human ear as compared to other techniques, based on STFT which has fixed time-frequency resolution.
LPCC	Uses a bank of equal bandwidth filters with linear spacing central frequencies; Modelled by all pole model.	The equal bandwidth of all filters renders unnecessary the effort for normalization of the area under each filter; Gives smoother spectral envelope and stable representation.
RASTA-PLP	Applies a band pass filter to each spectral component in the critical-band spectrum estimate.	These features are best used when there is a mismatch in the analog input channel between the development and fielded systems. Lower order analysis results in better estimates of recognition parameters training data.

References

[1] Lawrence Rabiner, Ronald Schafer, Introduction to digital speech processing, Prentice Hall.
[2] L. Rabiner and B. H. Juang, Fundamentals of Speech Recognition, Upper Saddle River, NJ: Prentice-Hall, 1993.
[3] Taabish Gulzar, Anand Singh, Sandeep Sharma,Comparative Analysis of MFCC, LPCC and BFCC for the Recognition of Hindi Words using Artificial Neural Networks, International Journal of Computer Applications (0975 8887), September 2014, Volume 101 No. 12.
[4] Hermansky H.,Perceptual linear predictive (PLP) analysis of speech, Journal of Acoustical Society of America 87, Apr. 1990, pp. 1738-1752.
[5] Yucesoy E., Nabiyev V. V., Comparison of MFCC, LPCC and PLP features for the determination of a speakers gender, IEEE Signal Processing and Communications Applications Conference, 2014.
[6] J. R. Deller, J. H. L. Hansen, J. G. Proakis,Discrete-Time Processing of Speech Signals, IEEE Press, Piscataway (N.J.), 2000.
[7] Makhoul, J.,Linear Prediction: A Tutorial Review, Proc. IEEE, 1975, Vol.63, pp. 561-580.
[8] Achmad F. Abka, Hilman F. Pardede, Speech Recognition Features: Comparison Studies on Robustness against Environmental Distortions, 2015 International Conference on Computer, Control, Informatics and Its Applications.
[9] M.A.Anusuya, S.K.Katti,Comparison of Different Speech Feature Extraction Techniques with and without

Licensed Under Creative Commons Attribution CC BY

Wavelet Transform to Kannada Speech Recognition, International Journal of Computer Applications (09758887), July 2011, Volume 26 No.4.

[10] Hermansky H., RASTA-PLP Speech Analysis, International Computer Science Institute, 1947.

[11] Pratik K. Kurzekar, Ratnadeep R. Deshmukh, Vishal B. Waghmare, Pukhraj P. Shrishrimal, A Comparative Study of Feature Extraction Techniques for Speech Recognition System, International Journal of Innovative Research in Science, Engineering and Technology, December 2014, Vol. 3, Issue 12.

[12] Shreya Narang, Ms. Divya Gupta, Speech Feature Extraction Techniques: A Review, International Journal of Computer Science and Mobile Computing, March 2015, Vol. 4, Issue. 3, pg. 107 114.

Author Profile

Smt. P. Prithvi is an Assistant Professor at the Dept, of ECE, National Institute of Technology, Warangal. Her major areas of interest include Speech Processing, Electronic Instrumentation.

Dr. T. Kishore Kumar is an Associate Professor at the Dept. of ECE, National Institute of Technology, Warangal. His areas of interest include Speech Processing, Radar Signal Processing, Real Time Embedded Systems, Advanced Digital System Design, VLSI systems, Real Time Processing using Embedded Systems.

Licensed Under Creative Commons Attribution CC BY