Linear Predictive Coding and Cepstral Analysis for Telugu Speech Recognition

P. Jeethendra[#1], M. Chandrashekar[#2]

(1)Research Scholar (PP ECE 0125), Subject : ECE (Speech Signal Processing)
Rayalaseema University, Kurnool (AP) India
(2)Centre for Advanced Systems, SFD-BDL,
Kanchanbagh, Hyderabad, Talangana, India

Abstract :-*This research work focused on feature extraction of in speech signal applied to Telugu Language processing. Telugu is a third largest spoken Indian Language which is widely spoken in Southern Indian States of India Talangana and Andhra Pradesh. Telugu is spoken in different accents even in the Telugu speaking geographic area.*

The feature extraction becomes more challenging when it is for a speaker independent speech recognition in nature. Every languages having different speaking styles called as accents or dialects. Identification of the accent before the speech recognition can improve performance of Speech recognition system. If the number of accents are more, then this becomes a crucial part of the study.

If we can understand the different sources of variability in the signal accent then we can begin to approach the problem by separating them out in subsequent analysis stages Speech signal is analyzed in two ways – signal Processing and linguistic processing.

During linguistic processing, signals are cut into chunks of varying degrees of abstraction such as acoustic-phonetic segments(APS), allophones, phonemes, morphophonemic, etc, which will be ultimately correlated with the letters in the script of a language by computational technique.

Among the various techniques presently available in speech processing technology such as Fast Fourier Transforms, Linear Predictive Coding, Mel Frequency Cepstral Coefficients, Cepstral Analysis, Discrete Wavelet Transforms, Wavelet Packet Transforms, Hybrid Algorithm DWPD and their applications in speech processing, we have studied, Out of these LPC and Cepstral Analysis in this research work.

Keywords : *Feature extraction, Speaker Independent, Linear Predictive Coding (LPC), Cepstral Analysis, Telugu, Acoustic Phonetic Segments (APS)*

I. INTRODUCTION

Speech propagates as a longitudinal wave in a medium. It is common to plot the *amplitude* of air pressure variation corresponding to a speech signal as a function of time; this kind of plot is known as a speech pressure waveform or just a speech waveform.

A plot of waveform is shown below of spoken word of ‘‘ARTIFICIAL’’

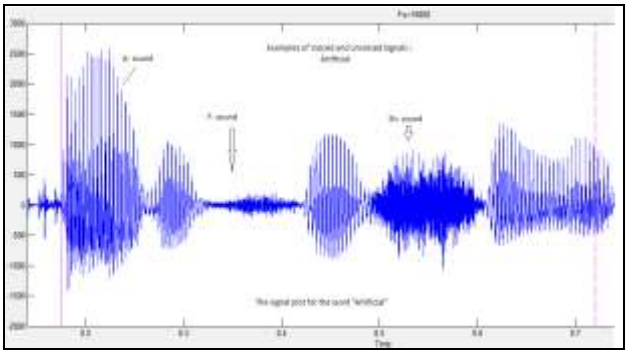

Fig. 1 Plotted waveform sample of spoken word ‘ARTIFICIAL’

The above plot shows the change of acoustic air pressure changes with respect to the time. This corresponds to a *periodic* signal and these signals make up the voiced speech sounds such as vowels and voiced consonants. Another kind of sound is irregular or *aperiodic* and so appears to be just random variations in air pressure.

1. SPECTROGRAM

An amplitude spectrum of a complex waveform shows the amplitude of each frequency component or sinusoid. The phase spectrum shows the phases of these components but is very seldom used in Speech Analysis. In the Speech Analysis , Spectra are used extensively to discern the characteristic shapes associated with different classes of speech sound. Any periodic signal showing a pattern of repetition in the time waveform which corresponds to primary rate of vibration of the signal, is known as Fundamental Frequency. This corresponds to the lowest major frequency component in the signal and, in case of voiced speech it equates to the frequency of vibration of the vocal folds. A vibration source like the vocal folds also produces certain integer multiple range of frequencies of the fundamental frequency called as Harmonics. In spectrum of such signal, the harmonics are represented with spikes.

The fundamental frequency can also be measured from a speech waveform by looking for the period of oscillation of the signal around the zero axis. Estimates can also be made from the spectrum since it shows a large peak at this frequency and at each multiple due to harmonics. Measurements can be made of the frequency of the major peak or of the distance between harmonic peaks

A Spectrogram shows the frequency content of a short section of a waveform, it shows how the spectrum changes with time. This is a two dimensional plot of frequency against time, where the amplitude at each frequency is represented by the darkness of the corresponding point in the display. An example with a utterance of word "STAMP" is shown in the figure below, The dark space corresponds to the high frequency energy in the /s/ phoneme, the vertical striations in the /A/ phoneme corresponding to the fact that the vowel is a periodic signal (Voiced) and the sudden onset of the relatively light area corresponds to the release of the /p/ at the end of the word.

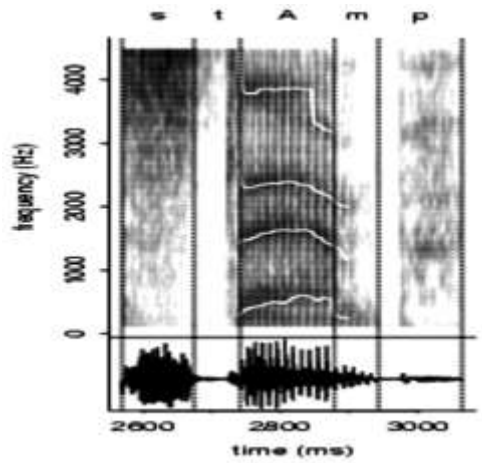

Fig. 2 Spectrogram of the utterance of word "STAMP"

The vertical striations in the vowel of utterance of an example word "STAMP" give another way of measuring the Fundamental frequency of the signal. From the wideband spectrogram as shown in fig. 2, each vertical striation corresponds to one open cycle of the glottis, and by measuring the time between striations tells us the period of oscillation and hence the fundamental frequency of the signal.(1)

1.a. Discrete Fourier Transform and Fast Fourier Transform

The Fourier transform transforms a time domain signal into a frequency domain representation of that signal. This means that it generates a description of the distribution of the energy in the signal as a function of frequency. in DFT the number of calculations required are (N x N) This is normally displayed as a plot of frequency (x-axis) against amplitude (y-axis) called a *spectrum*.

In digital signal processing the Fourier transform is almost always performed using an algorithm called the Fast Fourier Transform or FFT. Which is a , a quick way of performing this transform where it requires only (N log N) calculations and it gives the same results as the slower Discrete Fourier Transform (DFT) would. One consequence of using the FFT algorithm is that the length of the signal being analyzed must be a power of two, eg. 128, 256, 512, 1024 and so on

Applying the FFT algorithm to a 512 point window of data taken from a speech signal results in a vector of 512 numbers corresponding to the energy at 512 frequencies spanning the range from 0Hz to the sampling frequency. From Fourier's theorem we

know that adding together 512 sinusoids at these frequencies will exactly reconstruct the original signal. From Nyquist's theorem we know that the largest frequency component in the original signal must be half the sampling frequency, all frequencies above this are aliased onto lower frequencies. It follows that the second half of the spectrum is the mirror image of the first half:

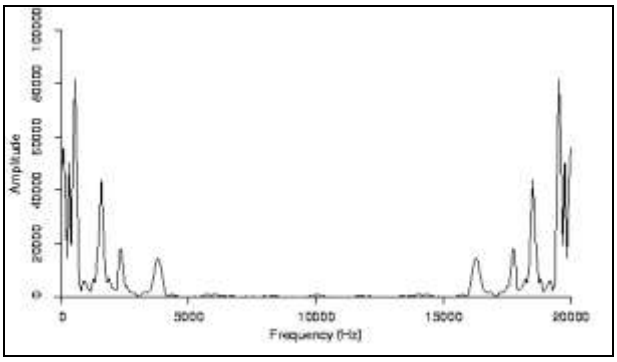

Fig. 3 Plot of Spectrogram with FFT

In fact, the first point in the spectrum isn't reflected and the frequency range from 0 to half the sampling frequency is covered by (N/2)+1 points where N is the size of the window. So from a 512 point fft of speech sampled at 10000Hz we get 257 unique spectral points covering the range 0 to 5000Hz.

1.b. Linguistic Processing

During linguistic processing, signals are cut into chunks of varying degrees of abstraction such as acoustic-phonetic segments, allophones, phonemes, morphophonemes, etc, which will be ultimately correlated with the letters in the script of a language. The relevance of phonetics for speech technology, specifically text-to-speech synthesis is well known (van Santen 2005). As one of the major purposes of speech technology is building systems that convert speech to text or generate speech from text, it may depend heavily upon finding out the correlates of groups of APSs with allophonic variants. After the APSs are identified, further processing of the segment stream should be performed by the linguistic processing module. In this module, a string of APSs (consisting of one or more such segments) are associated with the 'allophones' of the language. Each of these allophones are recognized as members of classes called 'phonemes'. This results in a string of phonemes. Each string of phonemes (consisting of one or more phonemes) is then associated with another underlying unit called 'morphophoneme'. Then certain chunks of morphophonemes in the string are matched with an underlying string that correlates with the mental lexicon or mental grammatical elements that are stored in the memory. These strings are either words in the dictionary or various grammatical items such as suffixes etc. Once this process is completed, the resulting string of distinct words is then transformed into the corresponding string of code-points which gets transformed onto way the string is written in the concerned script.

1.b.a. Structure of Linguistic Processing

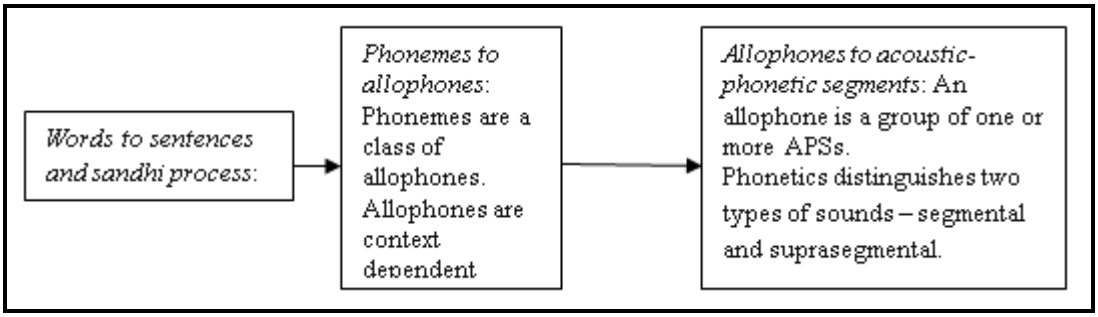

Fig. 4. Structure of Linguistic Processing

Phonemes are a class of allophones. Allophones are context dependent realizations of a phoneme. Most of the words in a language are made of strings of phonemes and the surrounding phonemes have an influence on the selection of an allophone of a given phoneme. Sometimes the context of influence may extend beyond left or right adjacency.

1.b.b. Features of Telugu Language

Telugu (□□□□□□), an Indian Dravidian language spoken by about 75 million people mainly in the southern Indian states of Andhra Pradesh & Telangana, where it is the official language. It is also spoken in such neighboring states as Karnataka, Tamil Nadu, Orissa, Maharashtra and Chattisgarh, and is one of the scheduled languages of India

- Type of writing system: syllabic alphabet in which all consonants have an inherent vowel. Diacritics, which can appear above, below, before or after the consonant they belong to, are used to change the inherent vowel.
- When they appear the beginning of a syllable, vowels are written as independent letters.
- When certain consonants occur together, special conjunct symbols are used which combine the essential parts of each letter.
- Direction of writing: left to right in horizontal lines

In Telugu, there is interaction between phonation type and vowel duration. Vowel duration is shortest when it occurs before voiceless aspirated and longest when it occurs before voiced unaspirated consonants. Shorter vowel duration is noted before consonant sequences (including germinates) as seen in most languages (Nagamma Reddy, 1999).

Prabhavathi Devi (1990) reported that, the duration of a long vowel is approximately twice the duration of the corresponding short vowel. The ratio between short and long vowel is 1:2.

Sreenivasa Rao, Suryakanth, Gangashetty & Yegnanarayana (2001) in their study of durational analysis of Telugu language, reported that, duration and intonation are two most important features responsible for quality of synthesized speech (Huang, Acero & Hon, 2001).

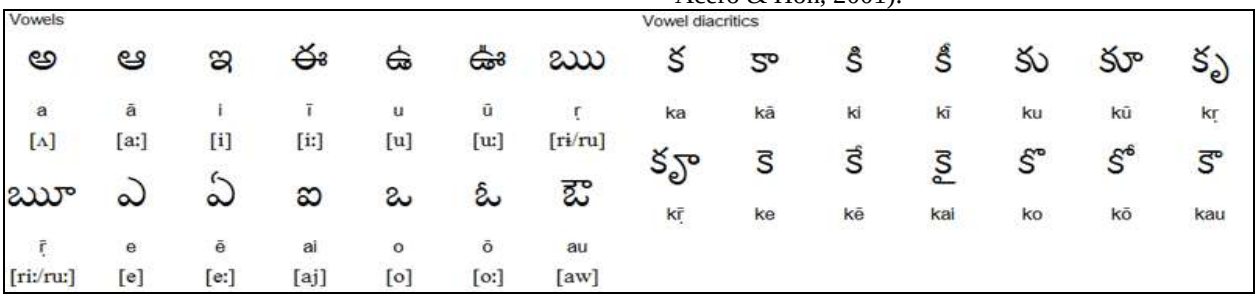

Fig. 5. Table of Telugu Vowel Fonts

- The script system of Telugu languages consists of variety of characters. Each of these characters correspond to one or more phonemes. Although most of the characters and phonemes have a one-to-one equation, some of them have one-to-many or many-to-one equation. As shown below (characters are marked by {} and phonemes by//)
- one character = one phoneme: Telugu, {k}↔/k/
- one character = two phonemes: Telugu: {e:} = /e:/, /æ:/
- two characters = one phoneme: Telugu: {th}, {dh}↔/dh/

Consonents

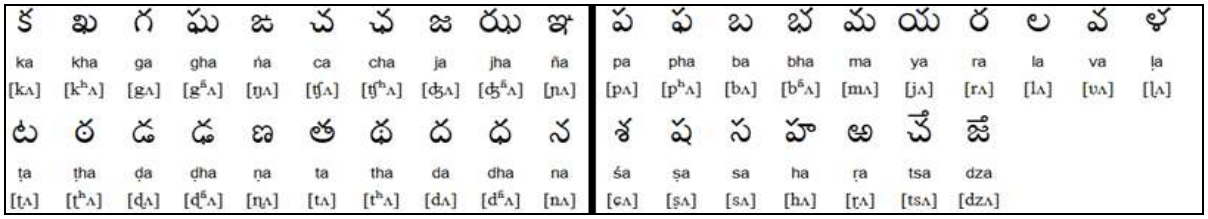

Fig. 6. Telugu Consonents Font

Numerals

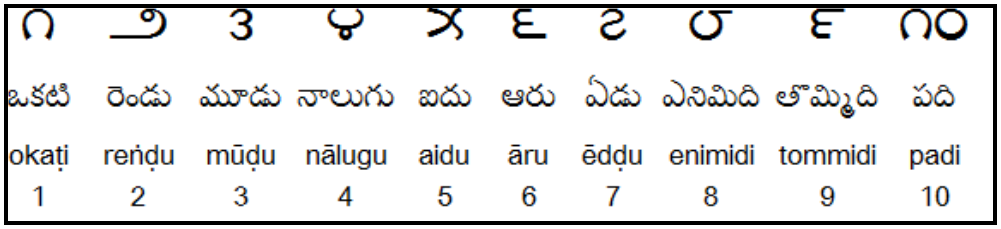

Fig.7. Telugu Numerals

A List of examples are shown below :
1. For realization of letter {m}in Telugu into its phonemic equivalent /m/ and the allophonic realizations of /m/:
1. {m}→/m/
2a. /m/→[M]/ V___V, #
2b. /m/→[m]/ elsewhere
where:
{m} is the script letter
/m/ is a phoneme
[M] is a labiodental nasal allophone
[m] a bilabial nasal allophone.

Allophone assignment rules will locate the phoneme /m/ and check its left and right contexts.
If they find a vowel (V=any vowel) on the left, and on the right if a V or word boundary (#)
is found, then the /m/ is replaced by its allophone [M]. If any part of the above context is not
satisfied, /m/ is replaced by [m].
2. An word in Telugu " Adi Tey " having utterance meaning 'Bring it!'

(a) Script input: {adi te:}
(b) Pre-*sandhi* phoneme string output: //adite://
(c) Post-*sandhi* phoneme string output after applying *sandhi* rules: /atte:/
(d) Allophonic string output: [5tte:]

Explanation: //adi te:// is converted into /atte:/ by the application of two *sandhi* rules: (i) deletion
of /i/ between two homorganic consonants (/d/ and /t/ in this case) followed by assimilation of the two homorganic consonants (/d/ and /t/ here becoming /tt/). /a/ is converted into the appropriate [5] allophone because of the presence of a low vowel (/e:/) in the next syllable.
This shows that Telugu Language employs a stronger assimilation process.
These examples altogether show that the *sandhi* rule sets differ from language to language. (Other language details not mentioned here) They also show what looks similar to the same phoneme across two languages may (and will) have different allophonic outputs.

***Allophones to acoustic-phonetic segments*:** An allophone is a group of one or more APSs. Phonetics distinguishes two types of sounds – segmental and suprasegmental. [p], [b], [s], [S], [v], [r] are some examples of segmental sounds. Vowel length, consonant length, stress, pitch of a vowel or other voiced continuants (such as [m], [l]) are suprasegmental sounds. All segmental sounds are composed of one or more APSs arranged sequentially in time. Among suprasegmentals, vowel and consonant length can also be viewed as a time-wise sequential arrangement. Pitch variations that contribute to the intonation or tone of a segment has to be viewed as an APS that is superimposed on another sound. Derivation of the string of allophones from the underlying string of letters is mostly in the realm of linguistic processing. Matching the allophone string with a string of APSs is at the interface between linguistic processing and the actual speech signal.

1.b.c. Vowel analysis in Telugu Language

Further with reference to the Vowel sounds of Telugu language it is discussed as Vowel Quantity and Vowel Quality. In the Voice Quantity for an example a short vowel might become extremely short – sometimes just having two pitch pulses – say, at 100 Hz fundamental, its length will be 20 msec. and a long vowel phoneme may have an extra long allophone when it occurs in the grammatical position of a conjoiner: e.g.,: /va:d.-u:adi:vacce:ru/ 'He and she came'. Here, /u:/ and /i:/ function as conjoiners (denoting 'and'). In this function, these two vowels have extra-long allophones (which may be tentatively represented as [u::] and [i::], respectively).and regarding the Vowel Quality the quality of a vowel in a syllable is contextually decided by the quality of the vowel that occurs in the next syllable. This is shown in table below:

	A			B		
	Meaning	Phonemic transcription	Phonetic transcription	Meaning	Phonemic transcription	Phonetic transcription
1	'cat'	/pilli/	[pilli]	'girl'	/pilla/	[pɪlla]
2	'nail'	/me:ku/	[me:ku]	'goat'	/me:ka/	[mɛ:ka]
3	'tie'	/kaṭṭu/	[kaṭṭu]	'bundle'	/kaṭṭa/	[kɜṭṭa]
4	'rust'	/tuppu/	[tuppu]	'bush'	/tuppa/	[tʊppa]
5	'a core'	/ko:ṭi/	[ko:ṭi]	'fort'	/ko:ṭa/	[kɔ:ṭa]

Table 1. Showing the Telugu Harmony (6)

In each of the words (in both the columns), we have the same vowel phoneme in the first syllable in each of the five rows. They are /i/ /e:/ /a/ /u/ /o/, respectively, in rows 1 to 5. InColumn A, the words have a close vowel in the second syllable (/i/ or /u/), whereas in Column B, the corresponding words have an opener vowel in the second syllable (/a/). The 'closeness' or 'openness' of a the vowel in the second syllable controls the 'closeness' or 'openness' of the allophone of the vowel in the first syllable. Thus in Column A, we get the 'close' allophones [i], [e:], [a], [u], [o:], whereas in Column B we get their 'open' counterparts [I], [E:], [3], [U], [O:]. Selection of the appropriate allophone of the vowel is highly important for a clean synthesis of these vowels in Telugu.

II. PROCESS OF FEATURE EXTRACTION

Since no two utterances of the same word or sentence are likely to give rise to the same digitial signal this obvious point underlies the difficulty in speech recognition but also means that we may be able to extract more than just a sequence of words from the signal. If we can understand the different sources of variability in the signal then we can begin to approach the problem of either separating them out for subsequent analysis stages.

The factors which could cause two random speech samples to differ from one another:

- *Phonetic identity*: the two samples might correspond to different phonetic segments, eg. a vowel and a fricative. Another source of variability is coarticulation between phonemes.
- *Pitch*: pitch and other source features such as breathiness and amplitude can be varied independently.
- *Speaker*: different speakers have different vocal tracts and source physiology. Speakers also get colds, get emotional and do other things to modify their voice properties.
- *Microphone*: and other properties of the transmission channel (eg. fixed vs. mobile telephone).
- *Environment*: background noise, room acoustics, distance from microphone.

Clearly the kind of variability we want to preserve in our signal model for speech recognition is that due to phonetic identity, which is largely due to the vocal tract configuration. All of the other sources might be considered noise to be removed from the signal as far as possible; however, there might be situations when some of these could provide useful information. The observation helps in that these sources of variability are largely independent from each other and so might be amenable to some kind of de-convolution operation. This is what we do to separate the source and filter components in Linear Predictive Coding and .Cepstral Analysis

Fig. 8 Location of Feature Extraction in ASR

In speaker independent speech recognition, a premium is placed on extracting features that are

somewhat invariant to changes in the speaker. So feature extraction involves analysis of speech siganl. Broadly the feature extraction techniques are classified as temporal analysis and spectral analysis technique. In temporal analysis the speech waveform itself is used for analysis. In spectral analysis spectral representation of speech signal is used for analysis Critical Band Filter Bank Analysis is one of the most fundamental concepts in speech processing. although it is a crude model in the initial stages os transduction of Human Auditory system , but it helps in building the motivation for development of subsequent high end feature extraction technique. We analyze two techniques LPC and Cepstral Analysis for Telugu Language Feature Extraction.

Starting with the Signal Analysis , which requires Spectrographic analysis of speech which is one of the most widely used techniques for studying the acoustic-phonetic characteristics of different phonemes in a language. It is an extension of the short-term spectral analysis, and primarily involves representation of the 3-D spectral information obtained by computing the magnitude spectrum over short overlapped window segments, i.e., 2-D spectral content varying with respect to time. The 3-D spectral information is represented on a 2-D plane with the X-axis representing time, Y-axis representing frequency, and the third dimension denoting the log-magnitude of the sinusoidal frequency components is converted to a proportional intensity or gray value. The resulting representation is referred to as a *spectrogram*.

Two popular spectrographic representations (11) used for analysis are *wideband spectrogram* and *narrowband spectrogram*, depending on the spectral and temporal resolution preserved in the final representation in the frequency domain. In wideband (WB) spectrograms, the spectral information is averaged over frequency windows of bandwidths 100 to 200 Hz. The corresponding time window chosen is 10 to 5 ms, respectively, so as to maintain unit time-bandwidth product. An example of a wideband spectrogram computed with a time domain window size of 5 ms and a shift of 2.5 ms is shown in Fig.9

- Wideband spectrogram provides higher temporal resolution at the cost of spectral resolution.
- Narrowband spectrogram provides higher spectral resolution at the cost of spectral resolution.
- Wideband spectrogram uses a short window size (typically 5 ms) for analysis, which results in a broader main lobe of the spectrum of the window signal. This smears or smoothens the spectral features of the speech segment thereby reducing the spectral resolution.
- Pitch or periodicity information of voiced sounds is reflected as vertical striations in WB spectrogram, while the pitch harmonics manifest as horizontal striations in NB spectrogram.

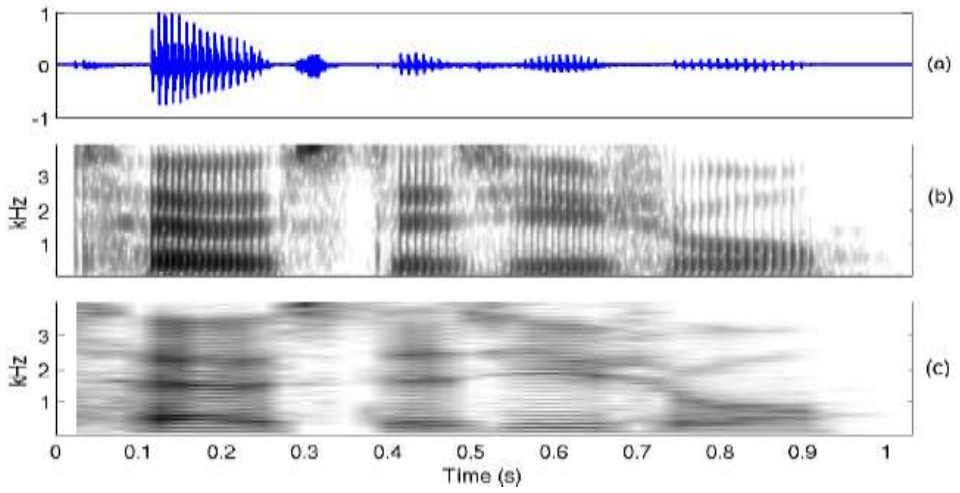

Fig. 9 . Spectrographic analysis of speech for an utterance "toast as usual". (a) Speech waveform. (b) Wideband spectrogram. (c) Narrowband spectrogram.

2.a. Identifying the oiced/Unvoiced/Plosive/Silence regions using spectrogram:

1. Voiced: In the case of vowels a regular formant structure (3 to 4 formant frequencies) and pitch harmonics (vertical striations in the case of wideband spectrogram) are used for identifying the voiced regions, where as nasals and voiced stops low frequency regions and pitch harmonics are used as clues.
2. Unvoiced: Energy at high frequency regions and no regular formant structure
3. Plosive: A silence bar followed by energy at high frequency regions.
4. Silence: no frequency components (white region)

III. LINEAR PREDICTIVE CODING :

LPC is a tool which is used for speech processing. LPC is based on an assumption: In a series of speech samples, we can make a prediction of the nth sample which can be represented by summing up the target signal's previous samples (k). The production of an inverse filter should be done so that is corresponds to the formant regions of the speech samples. Thus the application of these filters into the samples is the LPC process.[7] LPC is a way of encoding the information in a speech signal into a smaller space for transmission over a restricted channell. LPC encodes a signal by finding a set of weights on earlier signal values that can predict the next signal value:

$$y[n] = a[1]y[n-1] + a[2]y[n-1] + a[3]y[n-3] + e[n]$$

If values for a[1..3] can be found such that e[n] is very small for a stretch of speech (say one analysis window), then we can transmit only a[1..3] instead of the signal values in the window. The speech frame can be reconstructed at the other end by using a default e[n] signal and predicting subsequent values from earlier ones. Clearly this relies on being able to find these values of a[1..k] but there are a couple of algorithms which can do this (one is covered in the book). The result of LPC analysis then is a set of coefficients a[1..k] and an error signal e[n], the error signal will be as small as possible and represents the difference between the predicted signal and the original.

There is an obvious parallel between the LPC equation and that of a recursive filter (y*a = x):

$$y[n] = -a[1]y[n-1] - a[2]y[n-1] - a[3]y[n-3] + ... + x[n]$$

where we have rearranged the terms as in Equation 8.9 in the text. The LPC coefficients correspond to those of a recursive filter and the error signal corresponds to a source signal. Moreover, the conditions under which the error signal is minimized in LPC analysis mean that the error signal will have a flat spectrum and hence that the error signal will approximate either an impulse train or a white noise signal. This is a very close match to our source filter model of speech production where we excite a vocal tract filter with either a voiced signal (which looks like a series of impulses) or a noise source. So, LPC analysis has the wonderful property of finding the coefficients of a filter which will convert either noise or an impulse train into the original frame of speech.

The filter coefficients derived by LPC analysis contain information about the glottal source filter, the lip radiation/preemphasis filter and the vocal tract itself. However since these are much less variable than the vocal tract filter we can factor them out in practice (eg. by preemphasis before LPC analysis).

The LPC is implemented in the following methods

1. Covariance method
2. Autocorrelation method
3. Lattice method
4. Inverse filter formulation
5. Spectral estimation formulation
6. Maximum likelihood method
7. Inner product method

The characteristics of LPC are

1. Provides autoregression based speech features.[12]
2. Is a formant estimation technique
3. It is a static technique.[14]
4. The residual sound is very close to the vocal tract input signal.

Advantages of LPC

1. It is a reliable, accurate and robust technique for providing parameters which describe the time varying linear system which represent the vocal tract. [13]
2. Computation speed of LPC is good and provides with accurate parameters of speech.
3. Useful for encoding speech at low bit rate.

Disadvantages of LPC

1. Is not able to distinguish the words with similar vowel sounds [15].

2. Cannot represent speech because of the assumption that signals are stationary and hence is not able to analyze the local events accurately.
3. LPC generates residual error as output that means some amount of important speech gets left in the residue resulting in poor speech quality.

. IV. CEPSTRAL ANALYSIS

In a Cepstral Analysis the Signals are analyzed in Cepstral domain. Short time Cepstral analysis was proposed by Schroeder and Nall to determine the pitch of the human speech. The Cepstral coefficients are derived from Cepstral analysis, which can be utilized for applications like speech recognition, speaker verification, etc. Cepstrum of signal can be obtained by taking the Fourier transform of the log spectrum of the signal. Cepstrum is of two types namely power Cepstrum and complex Cepstrum Figures 2 and 3 show the diagram computation of power cepstrum and complex cepstrum for the input speech signal. Figure 4 provides the block diagram for the computation of Cepstral coefficients. The speech segment of size applied as input to the system is passed via windowing function. Generally a hamming window is selected to avoid end effects, since the hamming window spectrum has highest side lobe attenuation than other windowing like Hanning, Block man, Barlett, Kaiser, Lauczos, Tukey. The other windows rather than hamming has broader main lobe and smaller side lobes.(18)

This analysis technique is very useful as it provides methodology for separating the excitation from the vocal tract shape [16]. In the linear acoustic model of speech production, the composite speech spectrum, consist of excitation signal filtered by a time-varying linear filter representing the vocal tract shape as shown in fig.10.

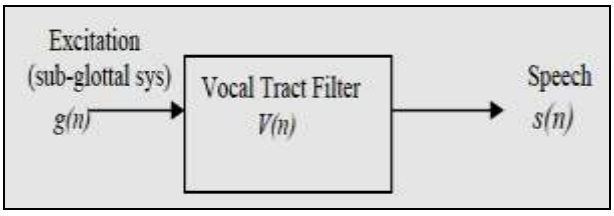

Fig. 10. Showing the Filtering of the Speech Signal with Acoustic Modeling

The speech signal is given as

$s(n) = g(n) * v(n)$

where $v(n)$: vocal tract impulse response and $g(n)$: excitation signal .The frequency domain representation

$S(f) = G(f) \cdot V(f)$

Taking log on both sides

$\log(S(f)) = \log(G(f)) + \log(V(f))$

Hence in log domain the excitation and the vocal tract shape are superimposed, and can be separated. Cepstrum is computed by taking inverse discrete Fourier transform (IDFT) of logarithm of magnitude of discrete Fourier transform finite length input signal as shown in fig.11.

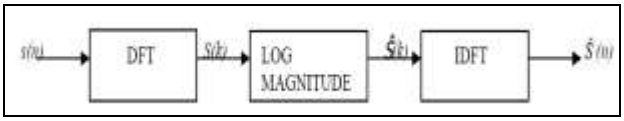

Fig. 11. System for obtaining the Cepstrum

$$S(k) = \sum_{n=0}^{N-1} s(n) \exp(-j2\pi/N)nk$$

$$\hat{S}(k) = \log(S(K))$$

$$\hat{s}(n) = \sum_{k=0}^{N-1} 1/N)\,(k)\,\exp(j2\pi/N)nk$$

$\hat{s}(n)$ is defined as cepstrum. In speech recognition cepstral analysis is used for formant tracking and pitch ($f0$) detection. The samples of (n) in its first 3ms describe $v(n)$ and can be separated from the excitation. The later is viewed as voiced if (n) exhibits sharp periodic pulses. Then the interval between these pulses is considered as pitch period. If no such structure is visible in (n), the speech is considered unvoiced.

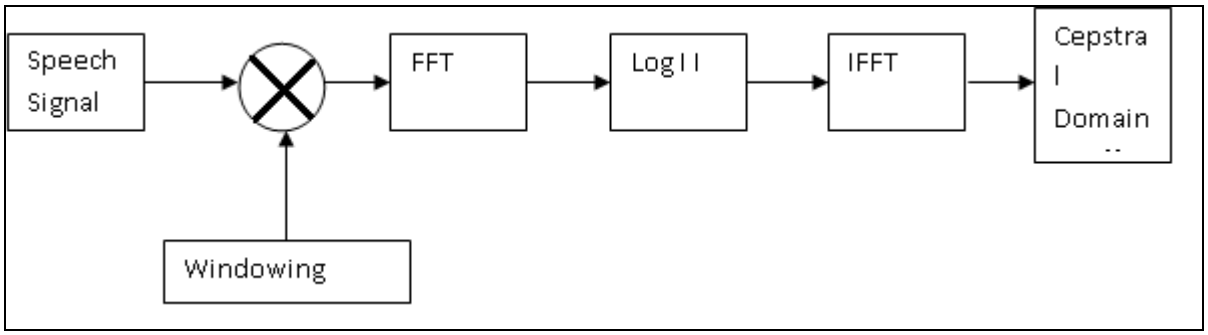

Fig. 12. Cepstral Analysis Processing Path for computation of Cepstral Coefficients

Characterstics of Cepstral Analysis

1. It is a Static feature extraction method.
2. It considers Power spectrum method.
3. It can be used to represent spectral envelope.

V. CONCLUSION

Speech conveys both Linguistic and Non Linguistic Information. There is no as such a single method that is best for the accurate Speech Recognition and Analysis. Speech Analysis is the front end of the Feature Extraction Process. LPC and Cepstral Analysis are widely used for the Speech feature extraction. Telugu Language employs a stronger assimilation process and hence development of a technique for an efficient and fast ASR system for regional languages like Telugu is need of the hour. In this paper the linguistic features of Telugu Language has been analyzed , its vowel strengths have been verified . The work implemented in the paper is a step towards the development of such type of systems. LPC analysis provides compact representation of vocal tract configuration by relatively simple computation compared to cepstral analysis. To minimize analysis complexity it assumes all pole model for speech production system. But speech has zeros due to nasals so in these cases the result are not as good as in case of vowels but still reasonably acceptable if order of model is sufficiently high, hence LPC model improves accuracy in Telugu Language The analysis may further be extended to continuous speech recognition. the accuracy of the system is a challenging area to work upon. There has been a lot of research in the field of speech recognition but still the speech recognition systems till date are not a hundred percent accurate. The systems developed so far have limitations: there are a limited number of vocabularies in the current systems and we need to work towards expanding this vocabulary, in regional language like Telugu. By implementing these coefficients of LPC with the Cepstral Analysis, The accuracy is increased.

VI. REFERENCES

1. Steve Cassidy - Speech Recognition Department of Computing, Macquarie University, Sydney,Australia , Steve.Cassidy@mq.edu.au (2002)

2. L. R. Rabiner and R. W. Schafer. Digital Processing of Speech Signals. Prentice Hall, Englewood Cliffs, New Jersey, 1978.

3. Douglas O'Shaugnessy. Speech Communication Human and Machine. Addison Wesley Books, 1978.

4. M. M. Sondhi. New Methods of Pitch Extraction. IEEE Trans. Audio and Electroacoustics, Vol. AU-16, No. 2, pp. 262-266, June 1968.

5. Harshita Gupa, Divya Gupta Department of Computer science and engineering, Amity University Uttar Pradesh, Noida, India, LPC and LPCC method of feature extraction in Speech Recognition System - Cloud System and Big Data Engineering (Confluence), 2016 6th International Conference, 14-15 Jan. 2016, IEEE *Xplore*: 11 July 2016

6. PERI BHASKARARAO, Salient phonetic features of Indian languages in speech technology, Tokyo University of Foreign Studies, Tokyo, Japan, published in *Sa¯dhana¯* Vol. 36, Part 5, October 2011, pp. 587–599._c Indian Academy of Sciences.

7. Manish P. Kesarkar, FEATURE EXTRACTION FOR SPEECH RECOGNITON, M.Tech. Credit Seminar Report, Electronic Systems Group, EE. Dept, IIT Bombay, Submitted November2003

8. H. Hermansky, B. A. Hanson, and H. Wakita, "Perceptually based processing in automatic speech recognition," *Proc. IEEE Int. Conf. on Acoustic, speech, and Signal Processing,"* pp. 1971-1974, Apr.1986.

9. Omniglot online encyclopedia Telugu%20alphabet,% 20 pronunciation% 20and%20 language.html

10. Thomas F. Quatieri , *Discrete-Time Speech Signal Processing*, , Chapter 7

11. Indexie .html

12. Tomyslav Sledevic,˘ Artu¯ras Serackis, Gintautas Tamulevici˘ us, Dalius Navakauskas, International Journal of Electrical,Computer, Electronics and Communication on Evaluation of Features Extraction Algorithms for a Real-Time Isolated Word Recognition System Vol:7 No:12, 2013

13. Shanthi Therese Chelpa Lingam, International Journal of Scientific Engineering and Technology (ISSN : 2277-1581) a Review of Feature Extraction Techniques in Automatic Speech Recognition, Volume No.2, Issue No.6, pp : 479-484 1 June 2013

14. Santosh K.Gaikwad and Pravin Yannawar, A Review, International Journal of Computer Applications ***A*** Review on Speech Recognition Technique Volume 10– No.3, November 2010

15. Navnath S Nehel and Raghunath S Holambe Journal on Audio, Speech, and Music Processing, on DWT and LPC based feature extraction methods for isolated word recognition, 2012

16. Rybach, D.; C. Gollan; G. Heigold; B. Hoffmeister; J. Lööf; R. Schlüter; H. Ney (September 2009). "The RWTH Aachen University Open Source Speech Recognition System". Interspeech-2009: 2111–2114.

17. Shreya Narang *et al*, International Journal of Computer Science and Mobile Computing, Vol.4 Issue.3, March-2015, pg. 107-114

18. A. Shiva Prasad et al Speech Features Extraction Techniques for Robust Emotional Speech Analysis/Recognition Indian Journal of Science and Technology, Vol 10(3), DOI: 10.17485/ijst/2017/v10i3/110571, January 2017