Comparison of Formant Estimation Techniques

Safya S. Bhore **Milind S. Shah**

***Abstract*—Formant frequencies are an important parameter in speech processing and analysis systems. Hence accurate estimation of formants is necessary. This paper, presents a comparison of two formant extraction techniques namely cepstral based and Linear Prediction Cepstral Coefficient (LPCC) based formant estimation. These two methods have been implemented in MATLAB for estimation of lowest three formants which is compared with the values of formants obtained from PRAAT software. These results have been tabulated. It was observed that, in general, the LPCC technique is more accurate than Cepstral analysis technique.**

***Index Terms*—Formant frequencies, Cepstral Analysis, Linear prediction based Cepstral coefficients.**

I. INTRODUCTION

The estimation of formant frequencies has always received considerable attention since it is an important parameter of the acoustic model of speech production [1]. In fact the first three formants carry considerable information about the speech signal. They help in perception of speech sounds, determining the phonetic content of speech, and used widely in recognition systems [2].For example, the spacing between F2 and F3 help to distinguish between glides in the syllable initial position. In fact some studies also relate these resonance frequencies to the age of speakers and even their heights which can be helpful for forensic purposes [3-4].

Historical approaches of detecting formant frequencies include the analysis-by-synthesis approach and the Linear Predictive Coding (LPC) based approach [5]. Various techniques with several modifications have emerged from them [6]. In this paper the Cepstral and the Linear Prediction Cepstral coefficients (LPCC) based techniques for formant estimation are implemented in MATLAB software and the results obtained are compared with the values obtained by PRAAT Software. The root mean square error (RMSE) is thus computed.

The following section deals with the two algorithms implemented for formant estimation. In section III the results of the experiment are described. Section IV highlights the conclusion of this study.

II. IMPLEMENTATION

A. *Cepstral Analysis based formant estimation*

According to the source filter theory of speech production, speech $s(t)$ is composed of the excitation signal $e(t)$ and the vocal tract components $h(t)$[7]. Cepstral analysis aims to make use of this fact for separating the signal into its components in a simplistic manner. So signal is viewed as a linear combination of

these two components [8]. For this purpose, it is required to transform signal into frequency domain $S(\omega)$ and then perform the log magnitude as given in equation (2). A transformation back does not lead to the time domain but to what is called as the cepstral domain and the resultant spectrum obtained is called as the cepstrum $C(n)$.

$$|S(\omega)| = |E(\omega)|\,|H(\omega)| \quad (1)$$

And,

$$C(n) = IDFT\ (\log|S(\omega)|) \quad (2)$$

Where,

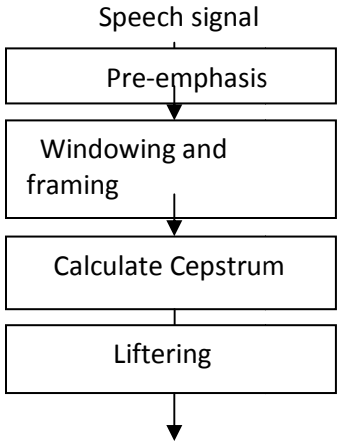
$|S(\omega)|$= speech spectrum

$|E(\omega)|$= excitation signal spectrum

$|H(\omega)|$= vocal tract signal spectrum

$C(n)$ = cepstral coefficients

The implementation algorithm is explained in the following Figure 1. Firstly the pre-emphasis operation is carried out on the speech signal. The pre-emphasized signal is then windowed and framed. For our implementation, a hamming window of 10msec duration is used. Next, the cepstrum is calculated. As the lower part of the cepstrum corresponds to the vocal tract information, this part of the cepstrum is retained with a low time liftering window whose cut off frequency is chosen to be between 15 to 40 samples. The liftering operation is explained in [2]. Cepstral coefficients are estimated for each frame. The smoothed spectrum is calculated to obtain the vocal tract signal [8]. The smoothening is done by the discrete Fourier transform operation performed on the signal. The formants which correspond to peaks in the smoothed spectrum are detected by a peak picking algorithm [2]. The entire process of estimating formants from smoothed spectrum is explained in [9].



Journal Impact Factor-2.02
Indexed by- ProQuest, DOAJ and many International Bodies.

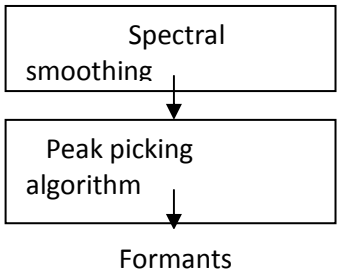


Fig.1. Algorithm for formant estimation using Cepstral Analysis [2].

B. LPCC based Formant Estimation

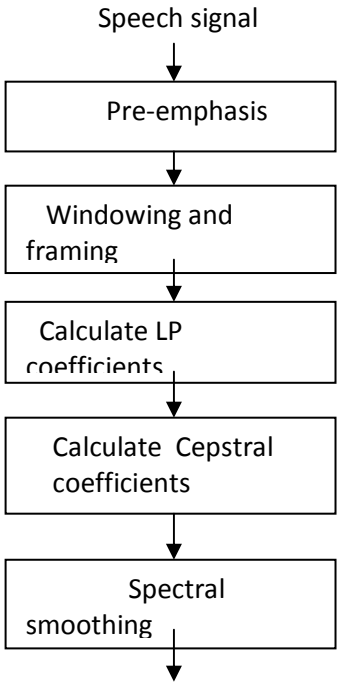
The implemented algorithm is shown in Figure 2. As in the previous method, pre-emphasis and windowing is carried out on the speech signal. And then the Linear Prediction Coefficients (LPC), for lpc order p, is computed. Cepstral coefficients from LPC are calculated using equations (3) to (5). This method of cepstral coefficients computation avoids taking the fourier transforms. So, cepstral coefficients derived from linear prediction parameters (LPCC) are calculated by the following set of recursive equations, from an autoregressive modelling of order p of a signal and estimated in a frame of the signal [6].

$$c_0 = log_e P \quad (3)$$

$$c_m = -\alpha_m + \frac{1}{m}\sum_{k=1}^{m-1}[-(m-k)\, c_{m-k}\alpha_k],\ 1 \le m \le p \quad (4)$$

$$c_m = \sum_{k=1}^{p}\left[\frac{-(m-k)}{m}\ \alpha_k\ c_{(m-k)}\right],\ p < m < n \quad (5)$$

Where,
c_k are the linear prediction based cepstral coefficients
α_k are the linear prediction coefficients
P is the prediction error power
n is the number of cepstral coefficients



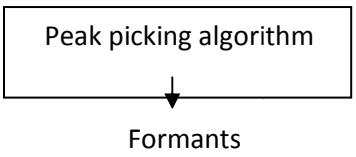


Fig. 2. Algorithm for formant frequencies estimation using LPCC Analysis. [2]

III. RESULTS AND DISCUSSION

A. Speech Material

Synthesized speech for vowel sounds /a/, /i/ and /u/ at a sampling rate of 11025 Hz was used. For natural speech, three female and three male within an age group of 19 to 55 were chosen. Natural speech was recorded using a microphone and sampled at a rate of 11025 Hz, and the vowel sounds /a/, /i/, /u/ were recorded by the speakers.

B. Results

For formant estimation by cepstral method, a 10 msec window is used. For calculating the cepstrum, FFT length equal to the number of samples is used. The obtained formant values were compared with the formant values obtained from PRAAT software. Formant values are obtained for every frame. And hence from this, the average root mean square error (RMSE) of the synthesized and natural speech of three male speakers this technique is as follows:

TABLE I
Root Mean Square Error for synthesized and natural male speech for Cepstral technique

vowel s	Synthesized speech			Natural speech		
	F1	*F2*	*F3*	*F1*	*F2*	*F3*
/a /	19.50	9.74	3.055	13.09	4.01	3.87
/i /	7.87	4.86	3.652	51.63	8.02	9.85
/u /	8.01	3.81	4.73	24.48	26.76	19.74

For formant estimation by LPCC technique, the LPC order was chosen to be between 10 and 20, and the number of derived cepstral coefficients was chosen to vary between 40 and 100. For vowels which have close formants (like F1 and F2 in /a/), the number of derived cepstral coefficients has to be kept high so as to resolve the formant peaks. Some other vowel sounds like /u/, in addition, demand a higher value of the lpc order. The average root mean square error (RMSE) of the synthesized and natural speech for this technique for three male speakers is as follows:

TABLE II
Root Mean Square Error for synthesized and natural male speech for LPCC technique

vowels	Synthesized speech			Natural speech		
	F1	*F2*	*F3*	*F1*	*F2*	*F3*
/a /	61.76	6.376	46.101	61.44	18.31	25.02

Journal Impact Factor-2.02
Indexed by- ProQuest, DOAJ and many International Bodies.

/i/	0.953	9.7834	11.2	15.214	14.579	14.57
/u /	9.04	50.9	14.41	9.24	11.23	16.5

For F1, mostly the technique based on Cepstral analysis shows more accuracy than that of linear prediction based Cepstral analysis. The second formants are more acceptable using LPCC technique. For F3, the results obtained by LPCC are more accurate than those obtained by Cepstral analysis. We can deduce that there is a wide range in the estimated values of formant frequencies. However, it is observed that the error increases with the order of the formant. In case of synthesized speech, as can be observed the error is significantly less and more close to the actual values of PRAAT. This is because there are more variations in natural speech than in synthesized speech and hence there exists more spurious peaks around the actual formant value, which makes peak detection process difficult.

CONCLUSION

Two techniques for the estimation of speech formant frequencies based on cepstral analysis and linear prediction based cepstral analysis were compared. The purpose of this study was to evaluate the performance of these methods. We compared the formant frequencies of natural and synthesized vowels detected by the two methods with the values obtained from PRAAT software. It was observed that overall LPCC based technique showed more accuracy than Cepstral based technique.

REFERENCES

[1] R.W. Schafer and L.R. Rabiner, "System for automatic formant analysis of voiced speech," *J. Acoust. Soc. Am.*, vol. 47, no. 2, pp. 635-648, 1969.

[2] G. Gargouri, M. A. Kamoun M. A. Zerzri and A. B. Hamida, "Cepstral method evaluation in speech formant frequencies estimation*," ICIT*, vol. 3, pp. 1612-1616, 2004.

[3] P. Busby and P. L. Plant, "Formant frequency values produced by pre-adoloscent boys and girls*," J. Acoust. Soc. Am.*, vol. 97, no. 4, pp. 2603-2606, 1995.

[4] H. Cao, Y. Wang and J. Kong, "Correlations between body heights and formant frequencies in young male speakers," *ISCSLP*, pp. 536-540, 2014.

[5] J. Hogberg, "Prediction of formant frequencies from linear combinations of filterbank and cepstral coefficients," *TMH-QPSR*, vol. 4, pp. 41-48, 1997.

[6] G. Gargouri, M. A. Kamoun, M. A. Zerzri and A. B. Hamida, "Formant estimation techniques for speech analysis," *International Conference on Machine Intelligence*, pp. 96-100, 2005.

[7] Douglas O' Shaughnessy, *Speech Communications: Human and Machine*, 2nd ed., Hoboken: New Jersey, John Wiley & Sons, 2000.

[8] G. Gargouri, M. A. Kamoun, M. A. Zerzri and A. B. Hamida, "Cepstrum vs LPC: A comparative study for speech formant frequencies estimation," *GESTS Trans. Intl.Comm. Signal Proces.,* vol. 9, no. 1, pp. 87-102, 2006.

[9] R.W. Schafer and L. R. Rabiner, "System for automatic formant analysis of voiced speech," *J. Acoust. Soc. Am.*, vol.47, no. 2, pp. 64-648, 1969.