ISSN 2347 - 3983

International Journal of Emerging Trends in Engineering Research (IJETER), Vol. 3 No.6, Pages : 229- 234 (2015)
Special Issue of NCTET 2K15 - Held on June 13, 2015 in SV College of Engineering, Tirupati
http://warse.org/IJETER/static/pdf/Issue/NCTET2015sp45.pdf

Novel Formant Estimation Techniques for Speech Processing

Mahaveer Chougala[1], **Shridhar K** [2]
[1] Dept. of Electronics and Communication, Basaveshwar Engineering College, Bagalkot
, India, mchougala91@gmail.com
[2] Dept. of Electronics and Communication, Basaveshwar Engineering College, Bagalkot
, India, shridhar.ece@gmail.com

ABSTRACT

The formant frequencies are very important parameters in speech processing and recognition. Hence the accuracy of measurement of formants frequencies is essential. This paper introduces the three methods of estimation of formant frequencies namely cepstral analysis method, log spectrum and Linear predictive cepstral coefficient (LPCC) method. All the three methods have been implemented in matlab for the estimation of first three lower values of formants. Here the formants are calculated for 5 vowels of different male and female speakers and their mean values of formant, coefficient of deviation between male and female speakers and the standard deviation of formant values are calculated from al the three method.

Keywords: Formant frequencies, cepstral analyses, Log spectrum and linear predictive cepstral coefficient (LPCC).

1. INTRODUCTION

The vocal tract shape can be considered as a filter that filters the excitation to produce the speech signal. The frequency response of the filter has different spectral characteristics depending on the shape of the vocal tract.

The formants are the spectral peaks present in the speech spectrum and they are the resonances of the vocal track. Formants carry some of the information about the speech signal so that they are used as an important feature of speech signal in speech processing and recognition system**.** Hence the qualities of systems such as speech recognition and speech identification are dependent on how accurate the formant frequencies are determined.

The purpose of this paper is to estimate the formant frequencies of vowels of 5 different male and female speakers using the cepstral analyses method, log cepstrum and LPCC analyses method.

The outline of the paper is as follows – The section 2 introduces the implementation of three methods or algorithms used for the formant frequencies estimation. In section 3 experiment and results are described and the comparison of the all the three methods are showed and in the last section highlights and conclusion is stated.

2. PROPOSED WORK

2.1 Cepstrum based formant estimation

According to the source filter theory of speech production the speech signal is composed of the excitation component e(t) And vocal track resonance h(t), that is the speech signal is convoluted form of excitation and vocal track resonance components. These components are separated by using the cepstrum analyses method. Cepstrum sometimes referred to as “homomorphic decomposition”. This technique is designed to separate the convoluted signal components by transforming the speech signal s (t) into one domain where the convolution has become a simple summation.

$$S\ (t) = e\ (t) \otimes h\ (t) \qquad (1)$$

Where e(t) and h(t) are the contribution of excitation and the vocal track respectively. Cepstrum is used to separate them.

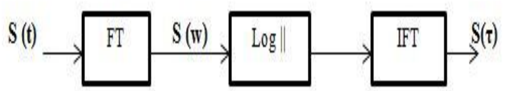

Figure 1: Homomorophic decomposition

The figure 1 shows the block diagram of homomorophic decomposition. The time domain speech signal is transformed into frequency domain by taking the Fourier transform on the both the sides of equation (1), we get

$$S\ (w) = E\ (w) \times H\ (w) \qquad (2)$$

The Log magnitude of the spectrum of the signal is written as

$$\text{Log} \mid S\ (W) \mid = \text{Log} \mid E\ (w) \mid + \text{Log} \mid Y\ (w) \mid \qquad (3)$$

When we take the logarithm of magnitude of the FT output, the periodic excitation is seen as a rapidly varying function and the vocal track response, which appears as the envelope of the spectrum, is a slowly varying function. In the last step Inverse Fourier transform is applied to the output of the Log spectrum to separate the ‘e’ and ‘h’ components and the output of the IFT is called cepstrum.

$$\text{IFT}\{\text{Log}|S(W)|\} = \text{IFT}\{\text{Log}\ |E(W)|\} + \text{IFT}\{|H(W)|\} \qquad (4)$$

The IFT transforms the function back into the time domain, but it is not the same as the time of the original signal. This domain is called the cepstrum, and the time axis is often referred to as the “quefrency” axis.

The cepstral window is designed and passes the output of the IFT block though the cepstral window allowing the formant information and finally the peaks are picked from the smoothed cepstral envelope and the smoothed cepstral envelope of the vocal track can be obtained easily by the following scheme.

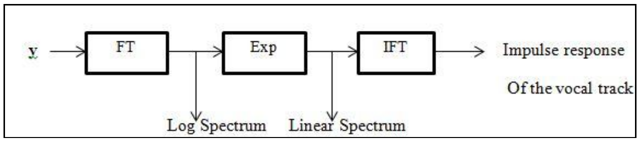

Figure 2: Block diagram for Smoothed spectrum transformation

After calculating the smoothed spectrum, we can afterward extract amplitudes corresponding to the vocal track resonance and this can be obtained by localizing the spectral maxima from the frequency bands corresponding to the first four formants F1, F2, F3 and F4.

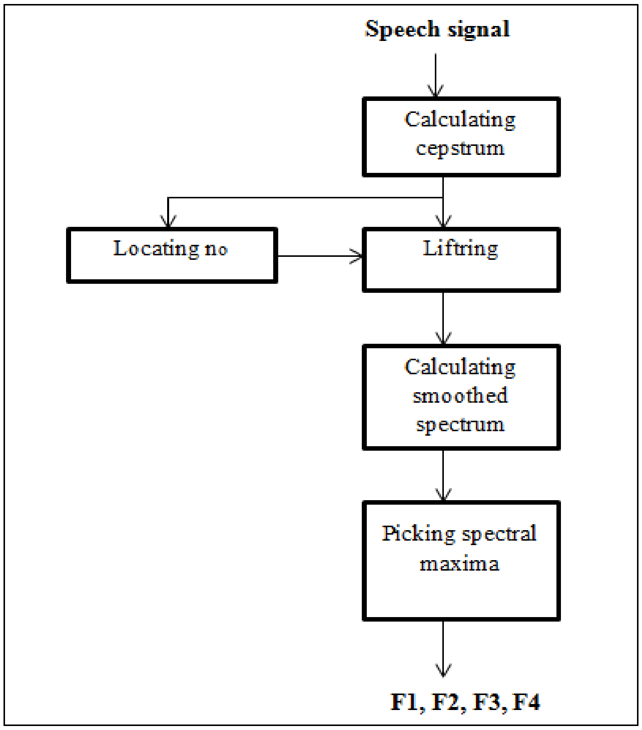

Figure 3: Formant estimation algorithm using cepstrum method.

2.2 Log Spectrum based formant estimation

The formants can also be tracked from the Log spectrum. The speech signal given as input to the system consists of a periodic excitation convolved with the impulse response of a vocal track which is a slowly varying function. When we take the log magnitude, the periodic excitation is rapidly varying and the vocal track response, which is envelope, is slowly varying. The formant information lies in the envelope of the log spectrum. The block schematic of formant estimation using log spectrum is shown in the fig 4.

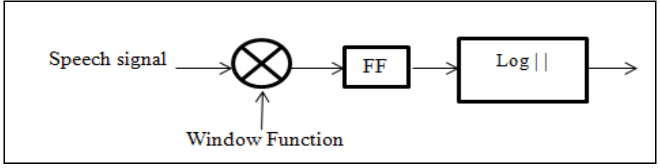

Figure 4: The block schematic for formants estimation using Log spectrum.

2.3 LPCC based Formant Estimation

The LPC method estimates the formant frequencies from the all pole model of the vocal track transfer function. The formants are estimated by two ways that is by solving the roots of the transfer function of the vocal track filter by factorizing the denominator polynomial and in the second method the formants are estimated by spectrum calculation and picking the peaks from the smoothed spectrum.

The general procedure for formant estimation using LPCC method is shown in fig 5.

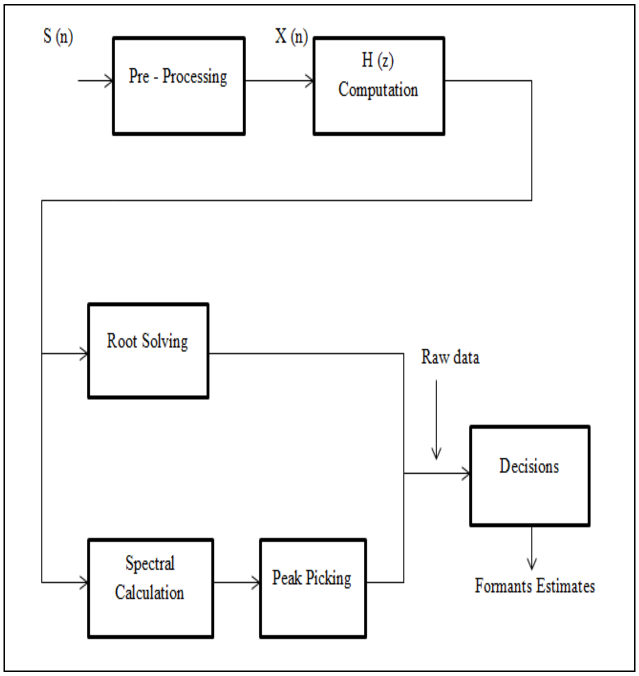

Figure 5: The block diagram shows the general procedure for formants estimation using LPCC method.

As shown in the figure 5, the Pre-emphasis and windowing is carried out on the speech signal and find the frequency response of the filter using the LPC coefficients and roots by solving the denominator of the transfer function of the filter. The cepstral coefficients are calculated by using LP coefficients for LPC order “n”and the formants are calculated by picking the peaks from the smoothed spectrum which is obtained by using the LPC coefficients. The algorithm to estimate the cepstral coefficients using LPCC analyses method is shown in fig 6.

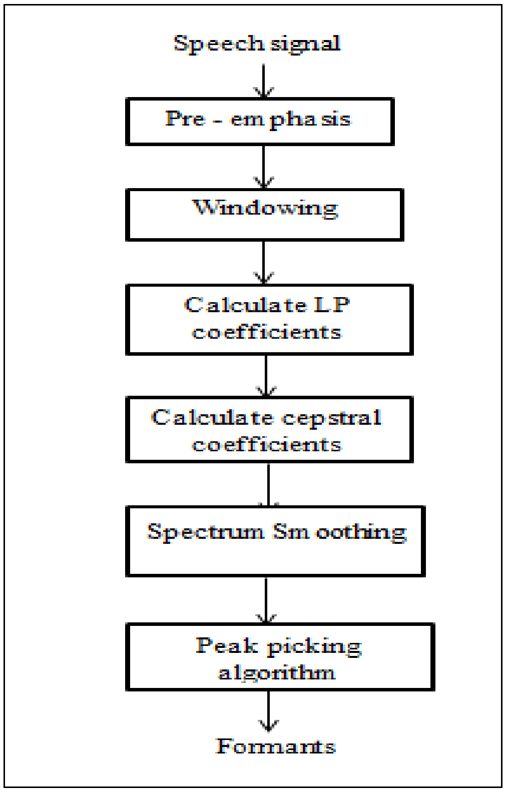

Figure 6: Algorithm for formant estimation using LPC analyses method.

III. EXPERIMENT AND RESULTS

The speech database of 16 KHz sampling frequency is used in this experiment. The TIMIT data base is used for speech analyses algorithm because it justified by the fact that it contains labeled and segmented speech from the great number of speakers.

In the experiment 5 vowels are taken from five different male and female speakers. All speakers read the same vowels. These vowels are [‘ah’, ‘eh’, ‘ih’, ‘oo’ and ‘uh’].

In this experiment the formant estimation algorithms are tested and extracted the first three formant frequencies mean values of formant frequencies for English vowels by male and female speakers, coefficient of deviation between the corresponding male and female speakers and also standard deviation for English vowels by male and female speakers are calculated from al the three formant estimation methods to confirm the accuracy of the algorithms.

3.1 Cepstrum based formant estimation method

The cepstrum of vowel “aa” spoken by male and female speakers are computed by equation (2).

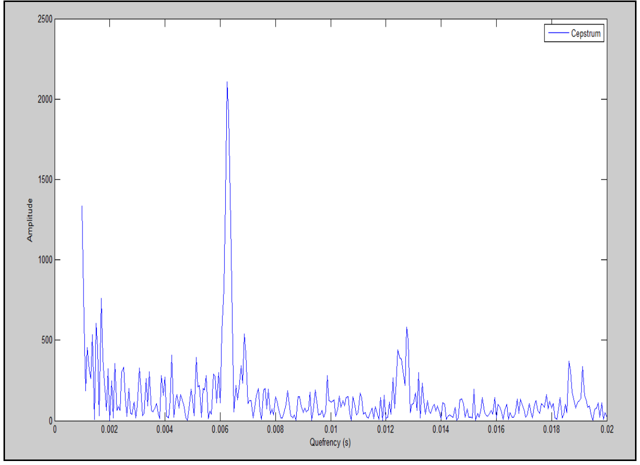

Figure 7: Cepstrum for male speaker

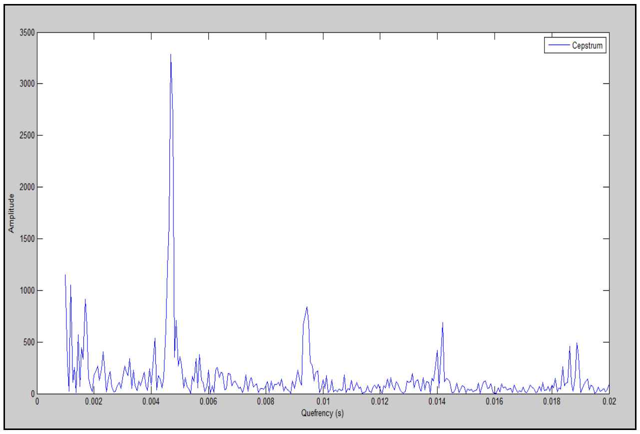

Figure 8: Cepstrum for female speaker

The fundamental frequency is determined by the spike of the highest amplitude which is varies from each gender. From the above figure it is found equal to 160 Hz for male speaker where as it is equal to 213.35 Hz for female speaker.

The cepstral envelope for male speaker is plotted in fig…

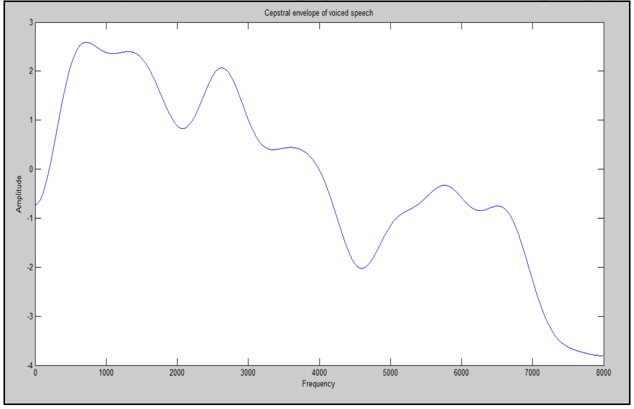

Figure 9: Cepstral envelope for male speak

Table1: Mean values of formant frequencies (in Hz) for vowels by male and female speakers

	F1		F2		F3	
	male	female	male	female	male	female
‘ah’	803.6	71808	1277	1644.8	2588.6	2988.6
‘eh’	559.2	666.4	1741.6	2055.2	2543.2	2987.6
‘ih’	478.2	558	1898.4	2241.8	2581.6	3051.6
‘oo’	495.6	584.4	1342.2	1494.2	2462.6	2823.8
‘uh’	611	659.8	1327	1564.4	2599.2	2899.8

Table2: Coefficient of deviation between male and female Formant

	F1	F2	F3
“ah”	0.95	0.896	0.926
“eh”	0.85	0.87	0.83
“ih”	0.89	0.86	0.83
“oo”	0.91	0.72	0.88
“uh”	0.90	0.80	0.90

Table3: Standard deviation of formant frequencies for vowels by male and female speakers

	F1		F2		F3	
	male	female	male	female	male	female
‘ah’	94.30	40.06	204.04	094.38	128.38	273.19
‘eh’	24.77	38.66	83.20	129.95	112.53	240.21
‘ih’	32.77	24.28	91.81	188.31	118.39	238.20
‘oo’	29.79	5.85	91.16	249.83	156.93	194.28
“uh”	36.92	40.67	167.18	171.49	30.19	262.93

3.2 Log spectrum based formant estimation

Table 4: Mean values of formant frequencies (in Hz) for vowels by male and female speakers

	F1		F2		F3	
	male	female	male	female	male	female
‘ah’	791.4	832.8	1387.6	1548.4	2647	2856.4
‘eh’	566.8	662.6	1773.8	2018	2518.2	3002.2
‘ih’	436	488.4	1900	2205.8	2574.6	3099.4
‘oo’	482.4	488.4	1900	2205.8	2574.6	3099.4
‘uh’	619	680.4	1355	1674	2613	2888.6

Table 5: Coefficient of deviation between male and female formant

	F1	F2	F3
‘ah’	1.119	0.7763	0.8661
‘eh’	08391	08474	0.8512
‘ih’	0.8569	0.8468	0.8459
‘oo’	0.8480	0.8982	0.8720
‘uh’	0.9260	0.8482	0.8963

Table 6: Standard deviation of formant frequencies for vowels by male and female speakers

	F1		F2		F3	
	male	female	male	female	male	female
“ah”	69.11	59.80	65.82	88.53	166.5	197.8
“eh”	67.93	147.27	94.64	133.52	82.15	255.9
“ih”	65.58	40.43	109.3	130.82	120.26	176.1
“oo”	40.26	64.68	77.81	88.37	97.86	248.2
“uh”	50.32	39.29	131.7	114.1	76.30	201.7

3.3 LPCC based formant estimation

Table 7: Mean values of formant frequencies (in Hz) for vowels by male and female speakers

	F1		F2		F3	
	male	female	male	female	male	female
“ah”	719.5	774.4	1167.2	1263	2670	2807
“eh”	555	683.2	1783	2065	2601	3022
“ih”	425.4	511.6	1869	2265	2636	3094
“o”	461.2	535.6	1188.6	1491.8	2691	2796
“uh”	581.4	698	1179	1529	2687	2940

Table 8: Coefficient of deviation between male and female formant

	F1	F2	F3
“ah”	0.9291	0.9235	0.9514
“eh”	0.8123	0.8636	0.8605
“ih”	0.8515	0.8252	0.8519
“oo”	0.8610	0.7967	0.9623
“uh”	0.8415	0.7710	0.9140

Table 9: Standard deviation of formant frequencies for vowels by male and female speakers

	F1		F2		F3	
	male	female	male	female	male	female
“ah”	41.91	59.15	70.20	259.3	62.82	206.24
“eh”	44.45	24.48	89.00	140.38	155.7	239.13
“ih”	46.84	24.79	102.14	165.45	194.06	1244.9
“oo”	27.63	21.40	112.98	221024	221.47	172.97
“uh”	36.34	42.56	140.6	201.42	147.26	350.11

From the above results it is observed that the formant frequencies of male speakers should be lower than those of female, this variation is due to the fact that the female vocal track is 15% briefer than that of a male. Generally the formant frequencies decrease as the length of the vocal track increases. The 70% of this result is justified by the table 2, 5 and 8.
The standard deviation for each vowel of male and female speaker is calculated to confirm the accuracy of the algorithms. From the table 3, 6 and 9 it is shows that the LPCC algorithm has more efficiency and accuracy than the other two formant estimation algorithm.

IV. CONCLUSION

In this paper comparison is made between the three methods of speech formant estimation: the cepstral method, log spectrum method and LPCC method. The Cepstral envelope technique is a frequency method based on localizing and peaks picking from the cepstral smoothed spectrum.
Vowel database is collected for 5 male and female speakers and are analyzed using each formant estimation algorithm and data collected from each gender is compared. It is found that formant frequencies for male speaker are lower than those of female.
For the cepstral and log spectrum method, the result shows that values of formant frequencies obtained present a wide range of measurement. This confirms the limitation of these methods to the estimation of formants especially at high frequencies. However for the LPC method, result shows that a narrow range in the estimated values of formant frequencies has been obtained and this validates the use of LP algorithm for the estimation of formant frequencies.

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

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

[3] J. Hogberg, “Prediction of formant frequencies from linear combinations of filterbank and cepstral coefficients,” TMH-QPSR 1997.

[4] 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.

[5] 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. 87-102, 2006.

[6] O. Schmidbauer, “An algorithm for automatic formant extraction in continuous speech”, In Proc. EUSIPCO-90, Fifth European Signal Processing Conf., vol. 2, pp. 1151-1154, Spain 1990.

[7] L. Welling & H. Ney, “Formant estimation for speech recognition”, IEEE Trans. Speech and Audio Processing, vol. 6, pp. 36-48, 1998.

[8] S. McCandless, “An algorithm for automatic formant extraction using linear prediction spectra”, IEEE Trans. Acoust. Speech and Signal Processing, vol. ASSP-22, pp.135-141, 1974.

[9] R .C. Snell and F. Milinazzo, “Formant location from LPC analysis data”, IEEE Trans. Speech Audio Processing, vol. 1, pp. 129-134, 1993.

[10] M.A. Kammoun, D. Gargouri, M. Frikha, A. Ben Hamida, “Linear Prediction Method Evaluation in Speech Formant Frequencies Estimation”, 3rd Inter. Conf. on Systems, Signals & Devices, Volume III Communication and Signal Processing IEEE, Sousse-Tunisia, March 2005.