Hamming Generalized Low Time Liftered Cepstral Analysis

S. Bedi  
Mtech Student, Department ECE,  
Shri SAI College of Engineering Technology,  
Badhani, Punjab, India

R. Singh  
Associate Prof., Department of Engg Technology,  
Shri SAI College of Engineering Technology,  
Badhani, Punjab, India

DOI: 10.23956/ijarcsse/V7I6/0165

Abstract—The main aim of this experiment is to separate the excitation of vocal tract components of the given speech signal by cepstral analysis. The first step is to convert the speech into short-term segments of size 15-20 ms. Then multiplication of each frame is carried out with the help of Hamming Window. The representation of short term speech in cepstral analysis is computed by finding the IDFT of log magnitude spectrum. In order to extract the deconvolved vocal tract and excitation component from the cepstrum liftering has to be done.

Keywords—Hamming, low time, Cepstral, Speech, Vowel.

I. INTRODUCTION

Speech is composed of excitation source and vocal tract system components. In order to analyse and model the excitation and system components of the speech independently and also use that in various speech processing applications, these two components have to be separated from the speech [1]. The objective of cepstral analysis is to separate the speech into its source and system components without having any prior knowledge about the source and the system. The cepstrum is the common transform which is used to gain information from the person speech signal. It can be used in the separation of excitation signal which consists of words, the pitch and the transfer function which contains the quality of voice. By truncating the cepstrum at different quefrency values allow us to preserve different types of spectra [2-4].

Bogert et al developed the cepstrum perspective to find out the echo arrival times in composite signal by decomposing the non-additive constituents. Since, the delayed echos in the logarithmic spectrum of the input data sequence x(n) appears as ripples [5-8]. The cepstrum analysis is perturbed with the de-convolution of two types of signals: The one of fundamental (basic) wavelet and the other one which consists of train of impulses (excitation function). Therefore, the representation of the composite signal in terms of power, complex and phase spectra. The cepstrum analysis is very much familiar with the data which consists of wavelets [9]. However, prior to analysis, this is very true even the shapes of the respective wavelets are not known. It has many uses in military and navy applications in which the power cepstrum was successfully applied in the radar analysis where the arrival time was determined by reducing the interference of the main wavelet, and in marine exploration where source depth was determined and ocean depth was mapped with the help of cepstrum analysis [9-12]. Therefore, the power cepstra is the productive tool inspite of this that the frequencies of the excitation function and the basic wavelet donot overlap with each other. Cepstrum mainly consists of two types: Power cepstrum and the Complex cepstrum. The power cepstrum of the signal is defined as the square of the inverse Z-transform of the logarithm of magnitude squared of the z-transform of the data sequence.

$$x_{pc}(nT) = \left[ z^{-1} \log |x(z^2)| \right]$$

Inspite of the fact that the power cepstrum is used to detect echoes, since the phase information is lost it cannot be used for detecting wavelet recovery [13-14]. So, to overcome this problem complex cepstrum is used. The complex cepstrum of data sequence can be defined as the inverse z-transform of the complex logarithm of z-transform of data sequence. The word cepstrum was coined by reversing the first syllable in the word “Spectrum” [15-18]. The cepstrum exists in the domain referred to as “quefrency” which means that the reversal of first syllable in the frequency which has units of time. The cepstrum is the homographic transformation that allows us to perform such operation:
Cepstral analysis is based on observation that:
\[ x[n] = x_1[n] * x_2[n] \Rightarrow X(z) = X_1(z) \cdot X_2(z) \]

By taking the Log of \( X(z) \)
\[ \log \{X(z)\} = \log \{X_1(z)\} + \log \{X_2(z)\} = \tilde{X}(z) \]

If the complex log is unique and the Z-transform is valid then, by applying \( Z^{-1} \)
\[ \hat{x}[n] = \hat{x}_1(n) + \hat{x}_2(n) \]

The two convolved signals are now additive. The real cepstrum is defined as:
\[
    c_x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left[ X \left( e^{jw} \right) \right] e^{jwn} dw
\]

Its magnitude is real and non-negative.

And the complex cepstrum:
\[
    \hat{s}[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left[ X \left( e^{jw} \right) \right] e^{jwn} dw
\]
\[
    = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \left| X \left( e^{jw} \right) \right| e^{jwn} dw + j \cdot \text{arg} \left( X \left( e^{jw} \right) \right) e^{jwn} dw
\]

Where, \( \text{arg}() \) represents the phase. We call it complex because it uses complex logarithm, not due to sequence, which can also be real [19-20].

II. METHODOLOGY

The research work is divided into three major parts. Speech material was recorded from four speakers (2 male and 2 female, ages 23-27 years). In first part of speaker selection, speech recording and segmentation was done. These required male and female speakers given vowels and consonants in English language, which was recorded using high quality recording system. The second part of this research work is to do recording of words by these male and female speakers. In the third part, the research work is to do recording of sentences by these male and female speakers. After that normalization of speech amplitudes is carried out in range of -1.0 to 1.0. Then segmentation process is carried out in which we take 20ms segment. To calculate the cepstrum of an audio sample before it computes the fourier transform, it applies a Hamming window to the sample. The speech of two male and female speakers was recorded in form of vowels and consonants. After that the segmentation of speech is carried out, since each speaker takes different time for speaking out vowels and consonants. Block diagram of methodology for speech smoothing and pitch estimation is shown in Fig.3. various parameters of vowel /a/ is tabulated in Table I.

![Fig. 3 Detailed representation of experiment](image)

<table>
<thead>
<tr>
<th>Speech Parameters</th>
<th>Sp1</th>
<th>Sp2</th>
<th>Sp3</th>
<th>Sp4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pitch (Hz)</td>
<td>183.14</td>
<td>246.69</td>
<td>136.06</td>
<td>131.57</td>
</tr>
<tr>
<td>Intensity (dB)</td>
<td>74.195</td>
<td>69.608</td>
<td>73.083</td>
<td>77.383</td>
</tr>
<tr>
<td>Formant Frequency (Hz)</td>
<td>912.79</td>
<td>988.44</td>
<td>654.198</td>
<td>794.89</td>
</tr>
</tbody>
</table>

III. RESULTS AND DISCUSSIONS

Research work is carried out to compute cepstrum. Analysis of various stages during the computation of cepstrum is done. The first step is to convert the speech into short-term segments of size 15-20 ms. Then each frame is multiplied by a Hamming window. Then cepstral representation of short-term speech is computed by finding the IDFT of the log magnitude spectrum. In order to extract de convolved vocal tract component and excitation component from the cepstrum liftering has to be done. For extraction the vocal tract component low-time liftering was carried out in various stages of our experiment. From the plots which are shown below we find the vocal tract characteristics of the cepstrum obtained through the low-time liftering. Figure (a)-(d) shows the speech signal, applied hamming window, DFT cepstrum, and low time liftered cepstrum signal of two male and two female speakers for vowel /a/. It can be observed from the low time cepstrum that quefrency is speaker dependent.
Fig. 4 (a)-(d) Speech signal with cepstrum and formant extracted from low time lifting for four speakers.
IV. CONCLUSIONS

The estimation of formant locations is carried out with the help of vocal tract spectral characteristics. The computation is done by the spectral representation of the low-time liftered spectral coefficients. As the DFT of the low-time liftered cepstral coefficients gives the corresponding smoothing in the log magnitude spectrum, the location of formants was carried out by using the simple peak-picking algorithm. The plots of various codes shows the cepstrally smooth log magnitude spectrum of vocal tract components and formant locations corresponds to peaks as shown in the spectrum. The computation is carried out on the basis of low-time liftering. The graph of the low-time liftered cepstrum shows vocal tract characteristics of cepstrum obtained through the process of low-time liftering.

REFERENCES