**NIRMA UNIVERSITY**

**INSTITUTE OF TECHNOLOGY**

**Digital Signal Processing**

******

**SPEAKER PITCH ESTIMATION USING HOMOMORPHIC FILTER**

By:

**KRISH DESAI**

**PREFACE**

For any DSP Engineer, to design and develop a signal analysis system, prerequisite knowledge of signals and systems, various time to frequency transform techniques, MATLAB functions, and filter designing are essential.

The content of this report is based on the application of above given topics. This report summarizes the Experiment we performed using homomorphic filters along with the knowledge of speech processing. This report will help to understand the complete flow of signal analysis using filters and signal processing to help us find various parameters involved and to develop application based on those parameters. In this experiment we are using speech signal and pitch as parameters.

**INDEX**

|  |  |  |
| --- | --- | --- |
| **S.R. No.** | **Topic** | **Page No.** |
| **1** | Introduction | 4 |
| **2** | Problem Description | 5 |
| **3** | Computational Considerations | 8 |
| **4** | Algorithm | 9 |
| **5** | Flowchart | 9 |
| **6** | MATLAB Code | 11 |
| **7** | Experimental Result | 12 |
| **8** | Observation | 16 |
| **9** | Conclusion | 16 |
| **10** | References | 17 |

**Introduction**

Speech can be seen as the output of a linear time-varying system slowly varying its properties with time. Small segments of speech can be interpreted as a time invariant system.

**Pitch: -** Pitch is the fundamental frequency of the signal. The fundamental frequency is defined as the lowest frequency of a Periodic signal. Every speech segment can be formed out of it.

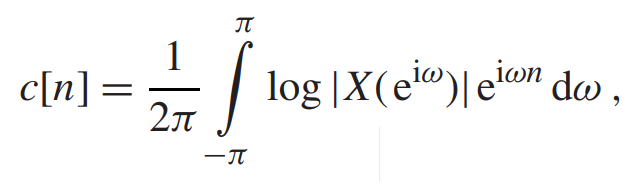
For analysis of speech signals, we use Homomorphic filters.

Before understanding concept and applications of Homomorphic filters we must define some new terminologies: -

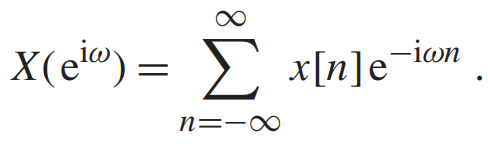
* **Cepstrum: -**

Scientists Bogert, Healy, and Tukey observed that the logarithm of the power spectrum of a signal consists of the logarithm of the signal spectrum plus a periodic component due to the noise(echo). A periodic component can help us understand new parameters of a signal. Hence, cepstrum is introduced.

“The cepstrum of a signal is the inverse discrete time Fourier transform (IDTFT) of the logarithm of the magnitude of the DTFT of the signal.”



Where,



* **Homomorphic: -**

From correlation of the vector space theory and signal operations it was observed by Oppenheim that classes of nonlinear systems could be defined on the generalized principle of superposition and such systems were called homomorphic. The principle of superposition simply states that if an input signal is composed of a linear combination of elementary signals, then the output is a linear combination of corresponding outputs.

**Problem Description**

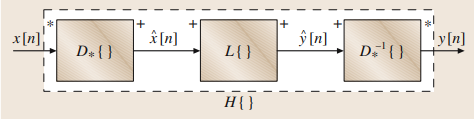
**Why don’t we use simple IIR and FIR filters for pitch detection: -**

* Simple filters are less sensitive towards additive noise which can contribute to pitch estimation and give errors in pitch estimation. Meanwhile homomorphic filters remove the undesired part, giving accurate pitch.
* Simple filters are designed for a particular range if the Pitch falls in the restricted region, it can cause loss in pitch. A speech signal always varies in pitch making it difficult to recognize the actual pitch.
* Pitch is harmonic of fundamental frequency. Sometimes simple filter cannot distinguish between pitch and frequency.

To tackle the above listed problems, we use a homomorphic filter for pitch detection. The problem of speech analysis is to estimate the parameters of the speech model and to measure their variations with time. Since the excitation and impulse response of a linear time-invariant system are combined in a convolutional manner, the problem of speech analysis can also be viewed as a problem in separating the components of a convolution. This problem is often called "deconvolution".

To solve this problem homomorphic filters are used.

A **homomorphic filter** is a homomorphic system that lets the desired signal unaltered and undesired signal completely removed. Where the input signal is a convoluted signal and the output is also a convoluted signal. In our case it basically implies the above cepstrum function. Here the linear function is the logarithmic of the dft of the signal.



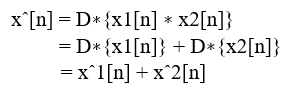
**figure 1.** basic structure of Homomorphic filter. (1)

where, D\* represents the convoluted signal

L represents the linear operation between two signals

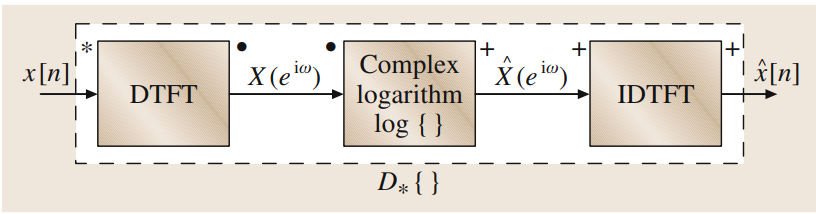
D-\* represents the inverse of D\*

This characteristic system is defined by the property that when x[n] = x1[n] ∗ x2[n], the corresponding output is



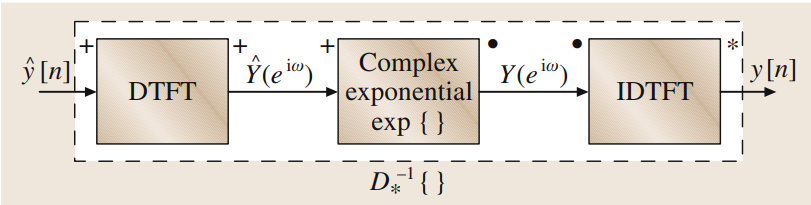
That is, the characteristic system transforms a combination by convolution into a corresponding combination by addition. The middle system in Fig. 2 is the system denoted L{}, which is an ordinary linear system satisfying the principle of superposition with addition as both the input and output operation for signal combination. Finally, the inverse characteristic system must transform a sum into a convolution so that the overall system transformation satisfies





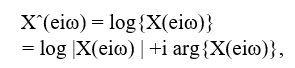
**figure 2.** DTFT representation of the characteristic system for convolution (1).

The inverse of the characteristic system for convolution is depicted in Fig.3. It is obtained by simply using the complex exponential to invert the effect of the complex logarithm.

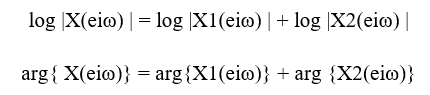


**figure 3.** DTFT representation of the inverse characteristic system for convolution (1).

The output of the characteristic system for convolution is the complex cepstrum. This is not because the complex cepstrum is complex; indeed, if x[n] is real, then xˆ[n] will also be real. Rather, the modifier complex is used to imply that the complex logarithm is used in the computation of the complex cepstrum. The complex logarithm is defined as

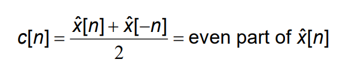


where if X(eiω) = X1(eiω) · X2(eiω), as for the convolution x[n] = x1[n] ∗ x2[n], then the following must hold:



The relationship between the cepstrum and complex cepstrum can be obtained by noting that, if x[n] is real, then log |X(eiω) | is a real and even function of ω, while arg{X(eiω)}, the imaginary part of Xˆ(eiω), is a real and odd function of ω.

This implies that

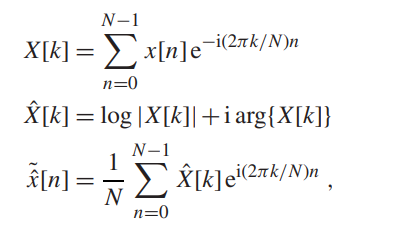


That is, as we have defined them, the cepstrum is the even part of the complex cepstrum.

**Computational considerations**

The mathematical representations of the characteristic system and its inverse depicted in Figs. 2 and 3, respectively, suggest a means for implementing homomorphic systems for convolution. The input signal is taken for a short duration of time around 10-15 ms. We perform Fourier transformations for individual signals and observe the pitch of the cepstrum. Cepstrum is the inverse Fourier transform of logarithm of Fourier transform of input signal. We know that the Fourier transform of a signal gives complex values which consist of undesirable components. To eliminate that we take the logarithmic function of the signal. The even part of the complex cepstrum is cepstrum as mentioned above. So, we perform ifft for the cepstrum and measure the pitch of it. Pitch is the quefrency corresponding to the maximum peak of the cepstrum plot.

The following steps are required to be followed to get the cepstrum of the signal and estimate its pitch.



X[k] represents the discrete Fourier transform of the input signal x[n], X^[k] represents that the input signal’s DFT is complex in nature. We take logarithm to separate the phase and magnitude components. As we know that the real part of a complex cepstrum is equal to the cepstrum, we perform an inverse Fourier transform to get the cepstrum.

**Pseudo Code**

R = audiorecorder() %record live audio

y = getaudiodata(r); % Store audio data in variable X

audiowrite('sample.wav', y, Fs); % Save audio as 'sample.wav'

N = 1:100; % Length of audio

Signal = y(N); % Extract the small length of the signal

ho = fft(Signal);

homomorphic\_signal = log(abs(ho));

cepstrum=ifft(homomorphic\_signal);

[pks, locs] = findpeaks()

% Convert peak locations to pitch frequencies

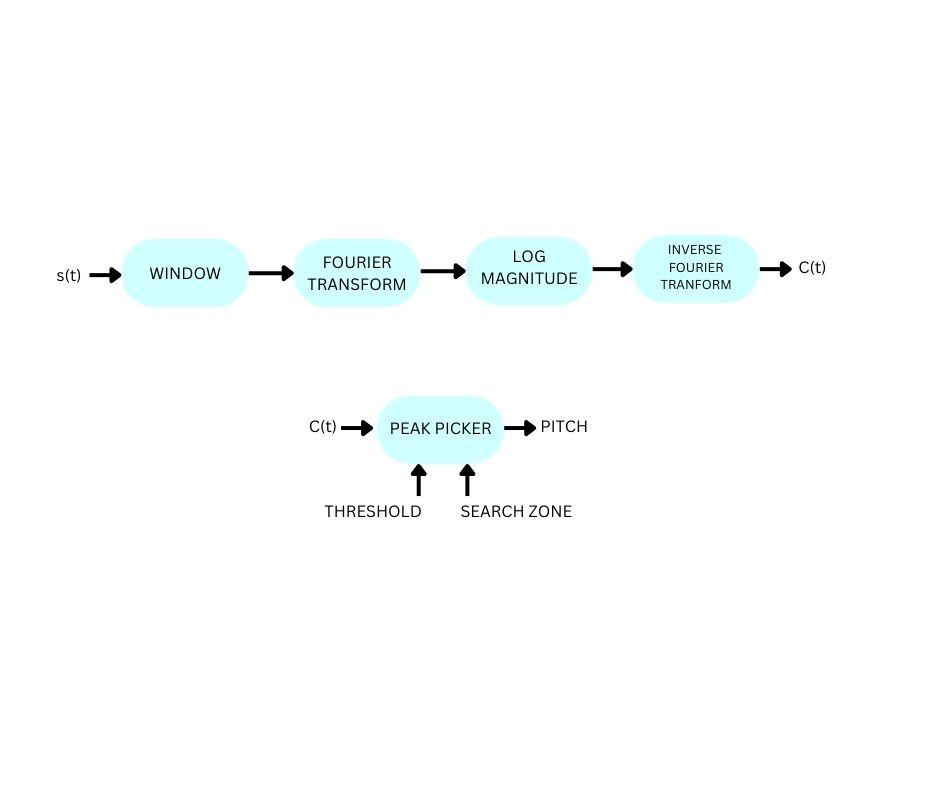
pitch\_frequencies = Fs ./ (locs - 1);

pitch\_hz = max(pitch\_frequencies);

**Algorithm**

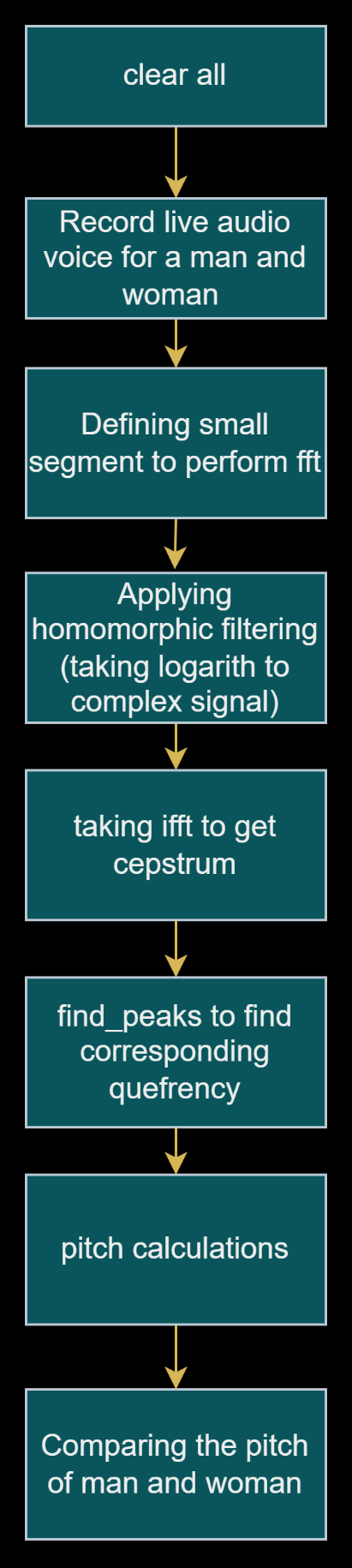
* Clear all the previous data.
* Record the audio from the surroundings and store it as a variable.
* Define a small segment for performing FFT on the speech signal.
* Apply a homomorphic filter, firstly do FFT of the given signal(X1), then do logarithm(X2) of X1, and then find the IFFT(X3) of X2.
* Find the peaks of the X3 along with its locations to find pitch frequencies.
* Divide sampling frequency by locations to find pitch frequencies and max of pitch frequencies to find pitch.
* Display all the plots and pitch.

**Block diagram**



**figure 4.** Block diagram of Homomorphic Pitch detector (2)

**Flow chart**



**Figure 5.**  flowchart of pitch detection using homomorphic filter

**MATLAB code**

clc;

clear all;

close all;

Fs = 4000;

channels = 1;

bits = 16;

r = audiorecorder(Fs, bits, channels); % Audio recording function

duration = 5;

disp('Recording Started');

recordblocking(r, duration); % Record 5 seconds of audio

disp('Recording stopped');

y = getaudiodata(r); % Store audio data in variable X

audiowrite('sample.wav', y, Fs); % Save audio as 'sample.wav'

N = 1:100; % Length of audio

Signal = y(N); % Extract the small length of the signal

ho = fft(Signal);

homomorphic\_signal = log(abs(ho));

cepstrum=ifft(homomorphic\_signal);

[pks, locs] = findpeaks(cepstrum, 'MinPeakDistance', round(Fs / 100)); % define MinPeakDistance

% Convert peak locations to pitch frequencies

pitch\_frequencies = Fs ./ (locs - 1);

pitch\_hz = max(pitch\_frequencies);

figure;

plot(Signal);

title('Original Audio Signal');

xlabel('Time');

ylabel('Amplitude');

figure;

plot(abs(ho));

title('FFT of Audio signal');

xlabel('Frequency');

ylabel('Amplitude');

figure;

plot(homomorphic\_signal);

xlabel('Sample');

ylabel('Pitch Frequency (Hz)');

title('logarithmic of Audio signal');

figure;

plot(cepstrum);

xlabel('Quefrency');

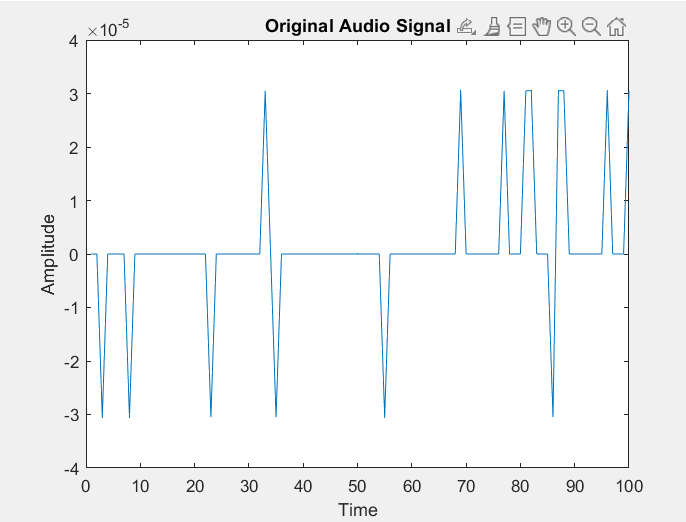
ylabel('Amplitude');

title('Cepstrum');

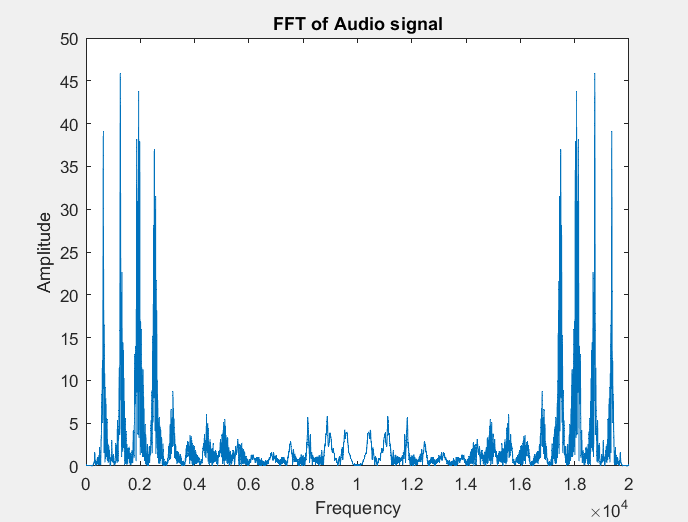
fprintf('Pitch: %.2f Hz\n', pitch\_hz)

**Experimental results**

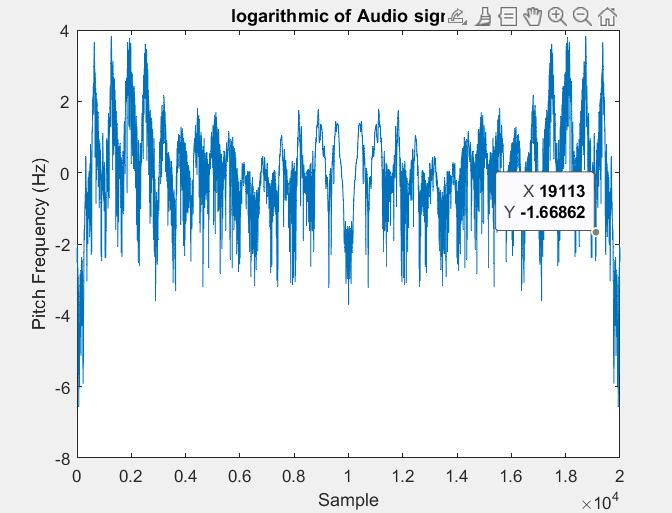
1. *Input audio signal of man*

**

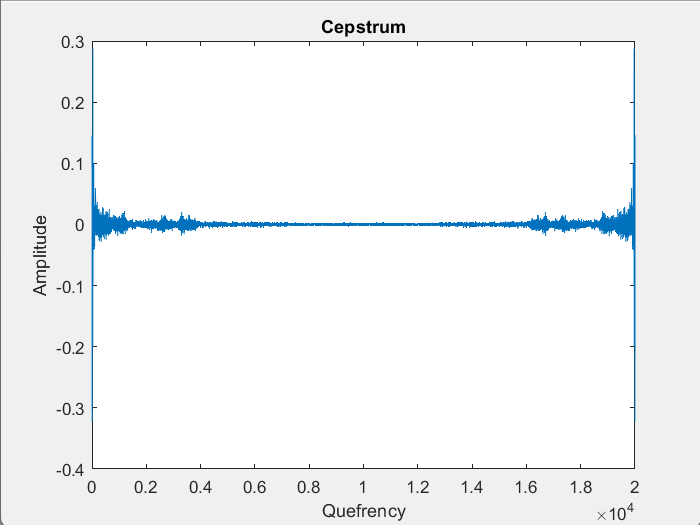
**Figure 6** Input speech signal of male

**

**Figure 7** FFT of male voice

**

**Figure 8** logarithmic of FFT of male voice

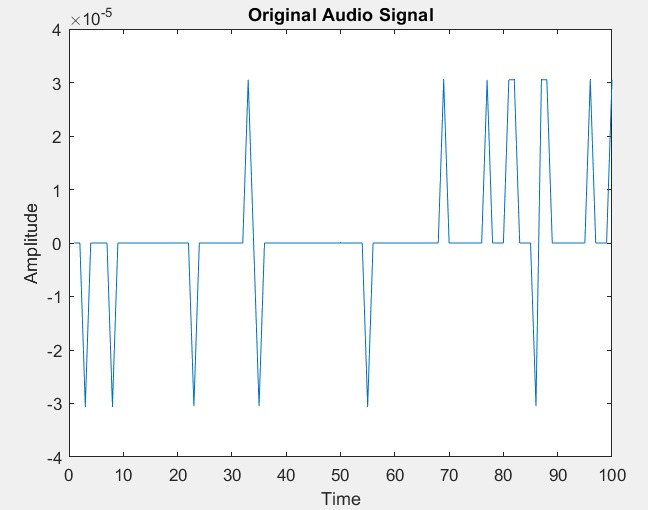
**

**Figure 9** Cepstrum of male voice

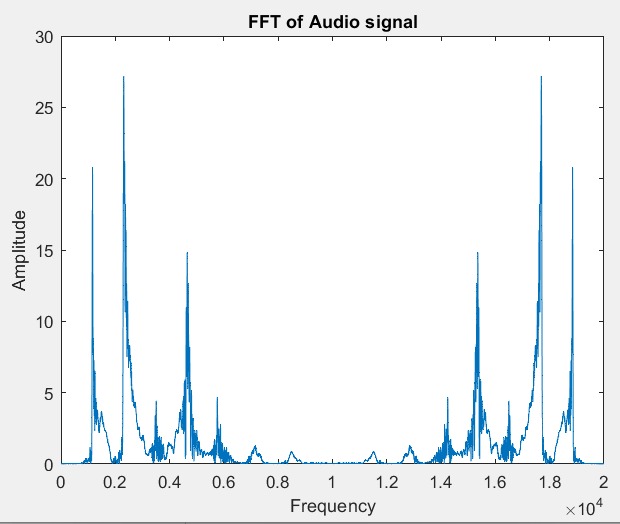
**Pitch of male voice: -**



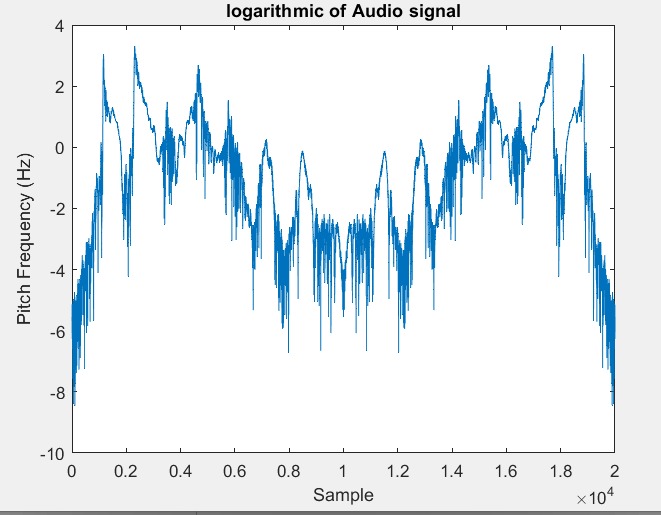
1. *Input audio signal for woman*

**

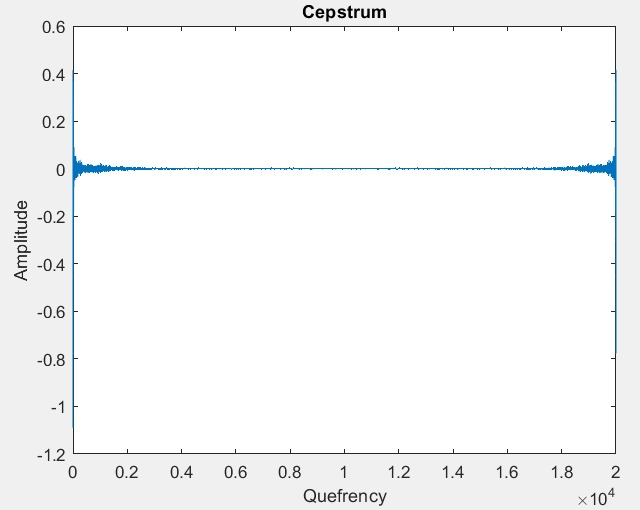
**Figure 10** Input speech signal of female

**

**Figure 11** FFT of female voice



**Figure 12** logarithmic of FFT of female voice

**

**Figure 13** cepstrum of female voice

**Pitch of female voice: -**



**Observation**

* We observe that FFT for a short period of speech signal is an Even function (fig.7 and fig.11).
* The Logarithm is also an Even function (fig.8 and fig.12).
* We observe peaks which are higher than the set Threshold value in the cepstrum graph of both male and female speech signals indicating that they both are voiced signals (fig.9 and fig.13)
* As we know, an adult woman's average range is from 165 to 255 Hz, while a man's is 85 to 155 Hz.

From the above simulation we observed that:

* Pitch of the male voice is 129.03 Hz.
* Pitch of the female voice is 235.29 Hz.

**Conclusion**

* In conclusion, the utilization of homomorphic filtering cepstrum for pitch detection in audio signals represents a significant advancement in the field of audio signal processing.
* This approach has demonstrated its efficacy in accurately and reliably determining the fundamental frequency, or pitch, of audio signals.
* Homomorphic filtering cepstrum takes advantage of the power of cepstral analysis and homomorphic processing to separate the source and filter components of an audio signal, making it particularly robust in the presence of noise and complex harmonic structures.
* Its ability to provide precise pitch estimation is crucial for numerous applications, such as speech recognition, music transcription, and sound analysis

**REFERENCES: -**

1. Springer Handbook of Speech Processing by Jacob Benesty, M. Mohan Sondhi, Yiteng Huang (Eds.).
2. Homomorphic Pitch detection by D.B. Paul.
3. Introduction to Digital Speech Processing Lawrence R. Rabiner and Ronald W. Schafer