ADVANCED DIGITAL SIGNAL PROCESSING(ELL-720)

ASSIGNMENT –2

PITCH DETECTION

SUBMITTED TO:

Prof. SESHAN SRIRANGARAJAN

Submitted by:

Anil kumar (2023JTM2567)

Pitch is the fundamental period of the speech signal. It the perceptual correlate of fundamental frequency. It represents the vibration frequency of the vocal cords during the sound productions (like vowels, for example). It is generally stated that pitch is the fundamental frequency of the signal.

Pitch or fundamental frequency is the lowest frequency component of a signal. The pitch period is inverse of fundamental frequency, is the smallest repeating unit of a signal. One such period describes the periodic signal (voiced part of the speech) completely.

Extracting Pitch in Time domain by Autocorrelation method

Most of the time-domain pitch period estimation techniques use auto-correlation function (ACF).

Auto-correlation function:

The basic idea of correlation-based pitch tracking is that the correlation signal will have a peak of large magnitude at a lag corresponding to the pitch period.

ACF for a signal s[n] is computed as:

$$R[k] = \sum_{m=0}^{N-k-1} s[m]s[m+k]$$

Where:

N - total number of samples in a window
K -Lag index
N- should be as small as possible to show time variation
& N should be large enough to cover at least 2 periods so that periodicity can be captured by R[k].

Properties of R[k]:

- Same periodicity as of s[m]
- Maximum value at k=0 and R[0] is equal to energy of deterministic signal.
- If s[m] is periodic with period of P samples, R[k] has maximum at k= 0, $\pm$P ,$\pm$2P ,...

Steps involved in pitch detection by correlation method-

1. **Load Audio**: Load the speech signal s1.wav and its corresponding pitch vector p1.mat.
2. **Calculate Energy:** Compute the energy of the signal to estimate its level.
3. **Thresholding:** Set a threshold to remove low-amplitude parts of the signal. Parts with amplitudes below the threshold are set to zero.
4. **Frame Processing**: Divide the signal into frames of length 80 samples.
5. **Autocorrelation**: For each frame, calculate the autocorrelation function. This function measures the similarity between a signal and a time-shifted version of itself.

6. **Peak Detection:** Find the peak in the autocorrelation function, which corresponds to the fundamental period of the signal. Due to noise and other factors, the peak might not be perfectly at the period, so some adjustments are made to account for this.

7. **Pitch Calculation**: Estimate the pitch as the inverse of the period index, multiplied by the sampling frequency.

8. **Pitch Range Limitation:** Limit the pitch values to a desired range, such as for female/male voices.

Auto Correlation Method:

S1.wav (Female voice sample)- S1 file is a voice sample of a female, and it has the range of frequencies 200Hz to 600 Hz.

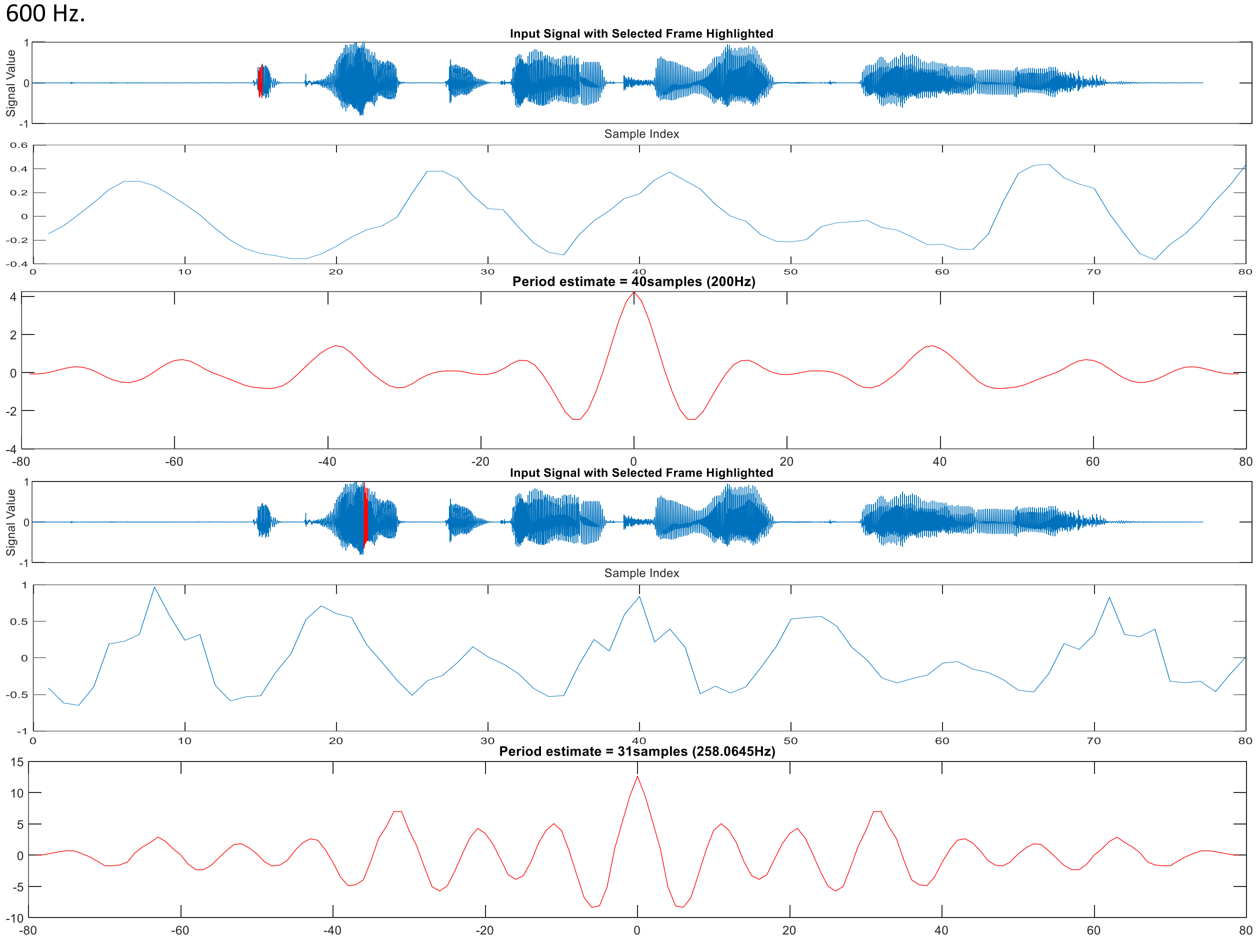


Figure 1- Selected frame of audio signal and its pitch calculation by correlation

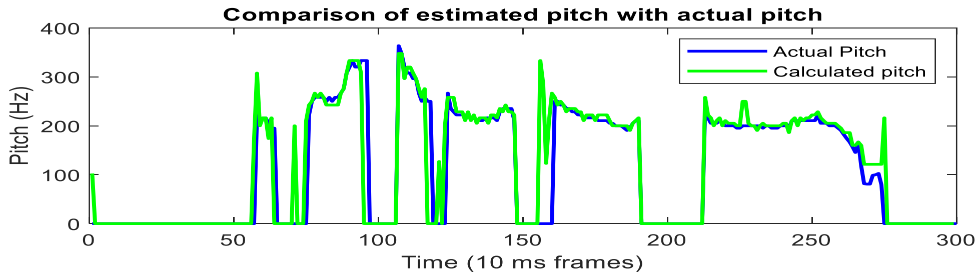
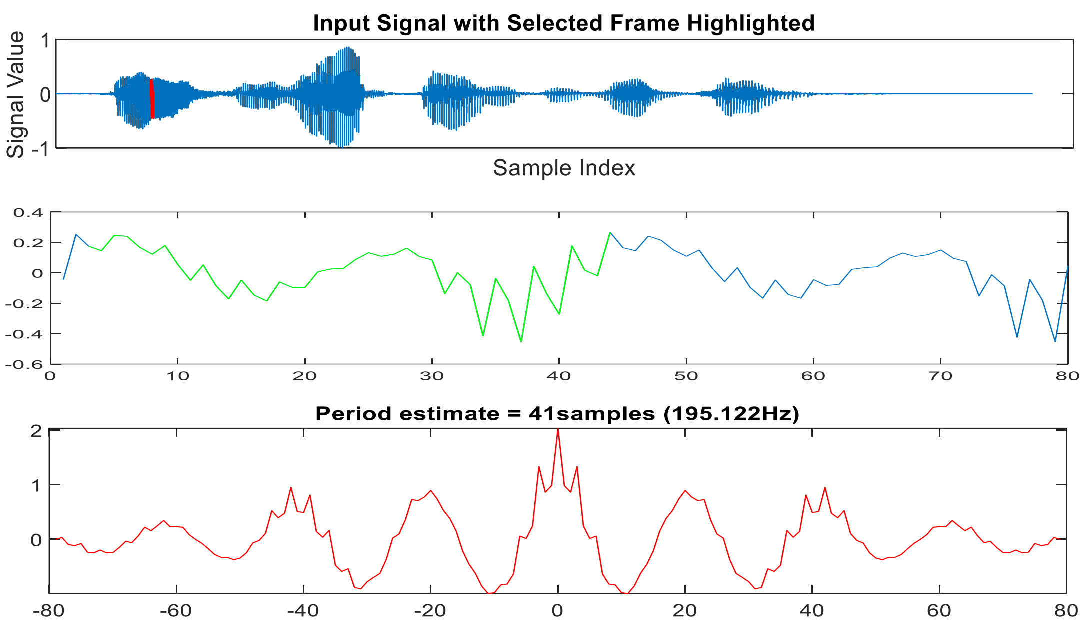
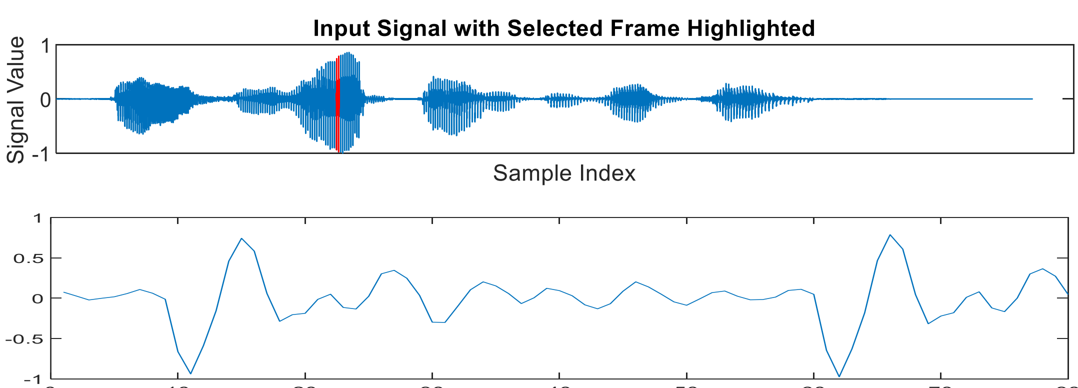


Figure 2- Comparison of estimated pitch with actual pitch

For s2.wav (Male voice sample)- S2 file is a voice sample of a male and it has the range of frequencies 50Hz to 350 Hz.





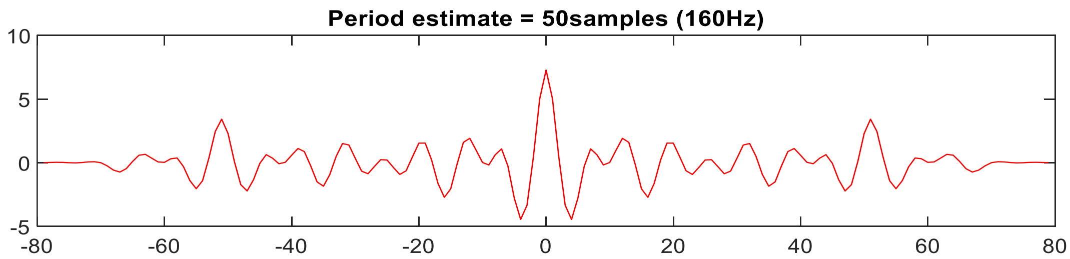


Figure 3- Selected frame of audio signal and its pitch calculation by correlation

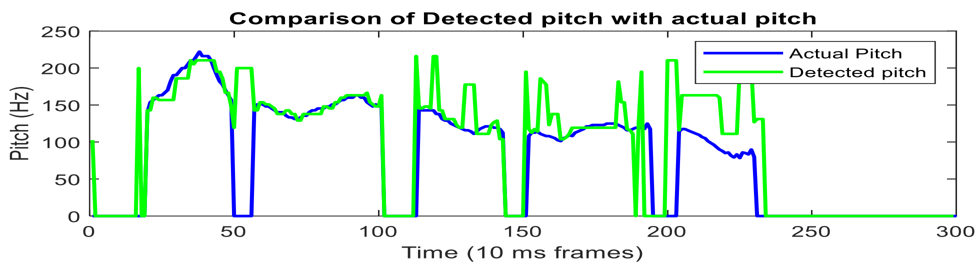
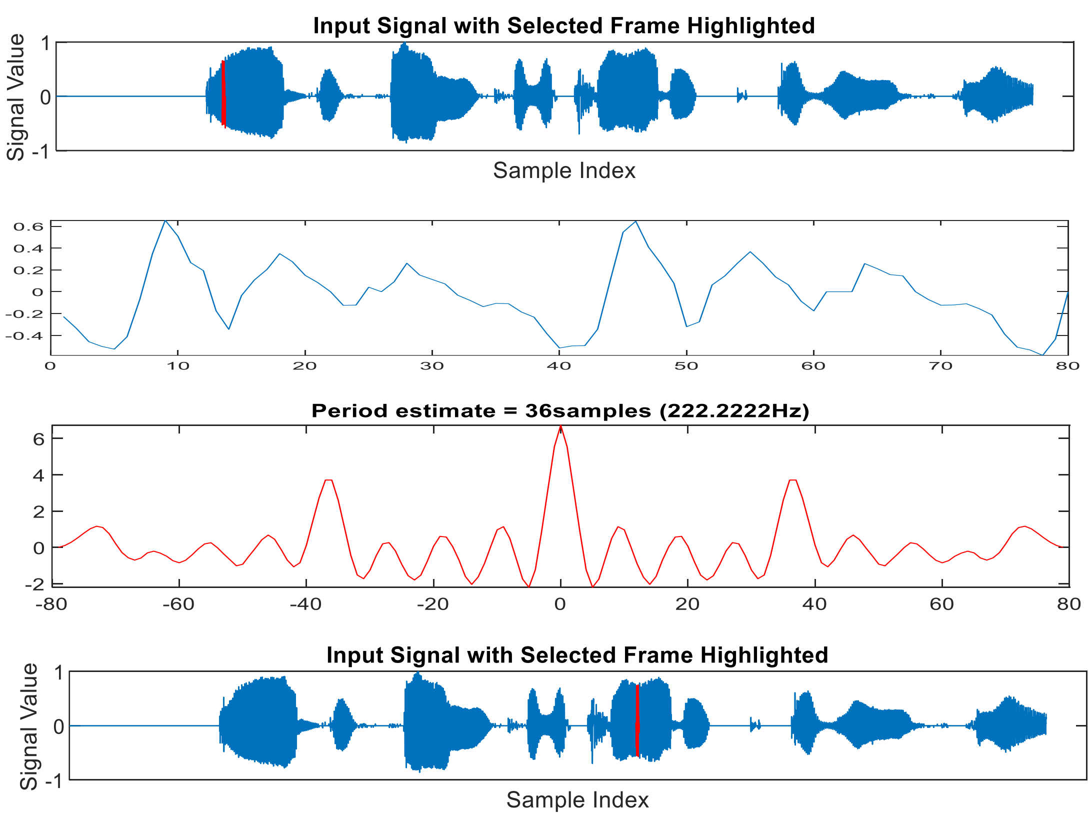


Figure 4- Comparison of estimated pitch with actual pitch

<u>For s3.wav (Female voice sample)-</u>



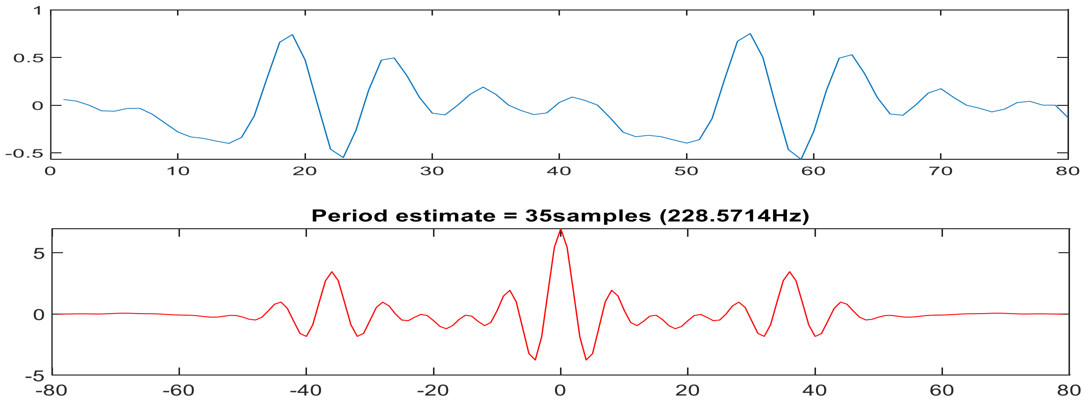


Figure 5- Selected frame of audio signal and its pitch calculation by correlation

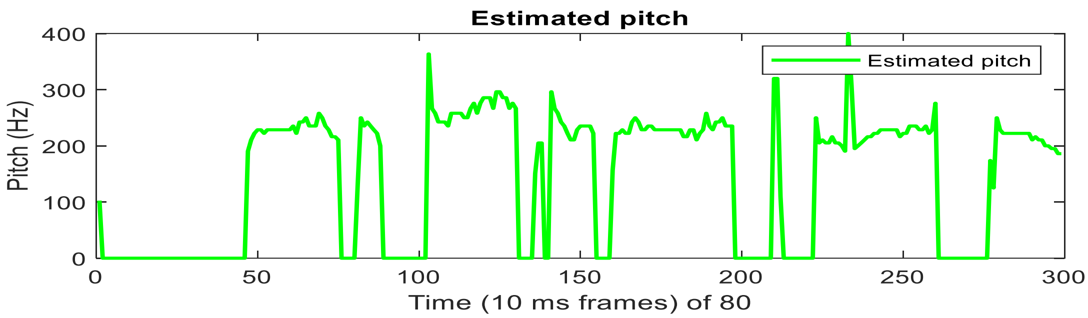
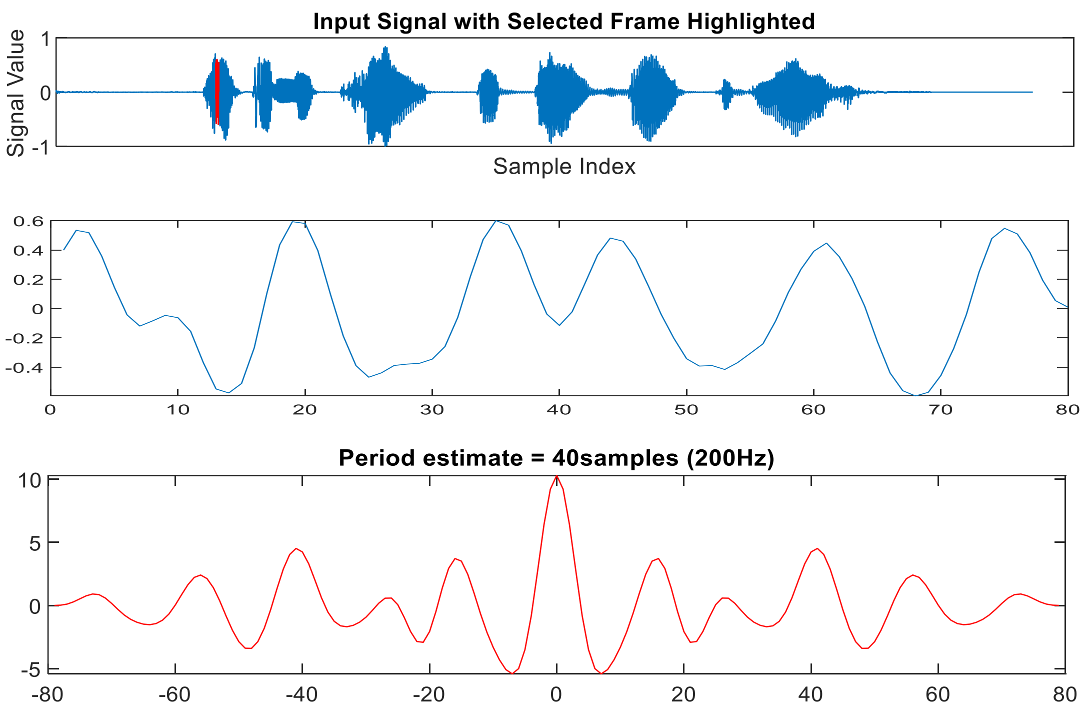


Figure 6- Estimated pitch of s3.wav

<u>For s4.wav (male voice sample)-</u>



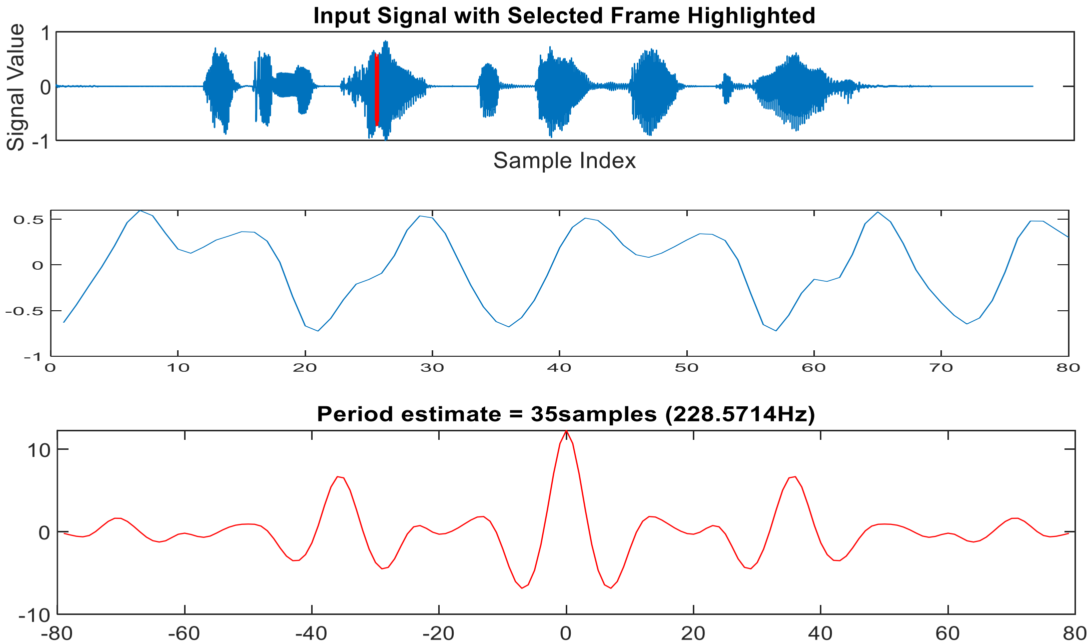


Figure 7- Selected frame of audio signal and its pitch calculation by correlation

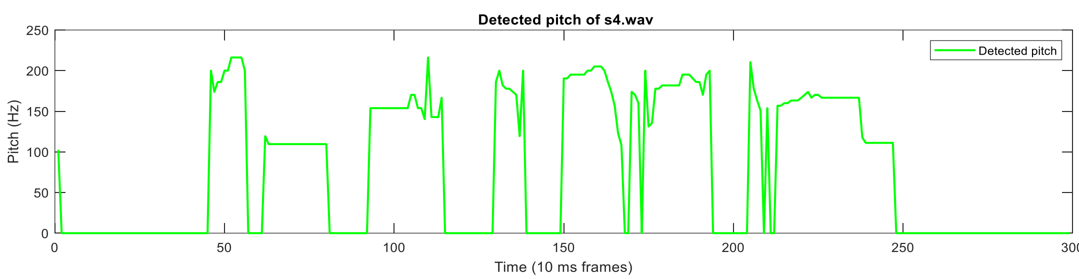


Figure 8- Estimated pitch of s4.wav

Extracting Pitch in frequency domain by Cepstral method

Pitch Detection using Cepstral Method:

Pitch detection often utilizes the Cepstral domain due to its ability to represent the frequency content of a signal in a logarithmic magnitude spectrum. The Cepstrum is created by applying either the FFT or IFFT to the log magnitude spectrum of a signal. These transformations are interchangeable, as each will provide a valid representation for our purposes.

In the Cepstral domain, pitch can be estimated by identifying peaks within a specific range of the resulting signal. The Cepstrum introduces the concept of "quefrency," which is essentially a measure of pitch lag. The peak with the highest energy corresponds to the dominant frequency in the log magnitude spectrum, thus indicating the pitch.

However, there are some considerations to keep in mind. Firstly, pitch and fundamental frequency are not identical, so the algorithm's choice of peak may yield the fundamental frequency (F0) or one of the formants (Fl). Secondly, the Cepstrum is sensitive to time shifts, meaning that precise alignment of time domain windows over voiced speech segments is crucial. This can be challenging, as Voice Activity Detection (VAD) systems often introduce errors, leading to phase ambiguity in the Cepstrum.

Steps involved while calculation of pitch:

1. Load the WAV file.
2. Define the frame length, number of samples, and sampling frequency.
3. Compute the FFT of each frame.
4. Calculate the magnitude spectrum of each frame.
5. Compute the log magnitude of the spectrum.
6. Compute the IFFT of the log-magnitude spectrum to obtain the cepstrum.
7. Since the cepstrum is symmetric, consider only the first half of the array.
8. Apply a high-time liftering process to emphasize the pitch frequency.
9. Obtain the high-time liftered cepstrum.
10. Perform matrix multiplication of the half-cepstrum and the liftering window.

For s1.wav (Female voice samples)-

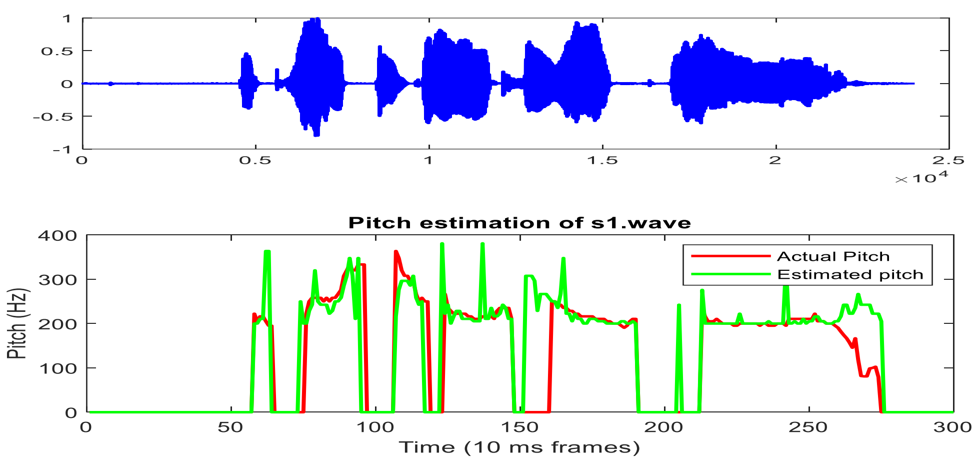


Figure 9- Comparison of estimated pitch with actual pitch

For s2.wav (male voice samples)-

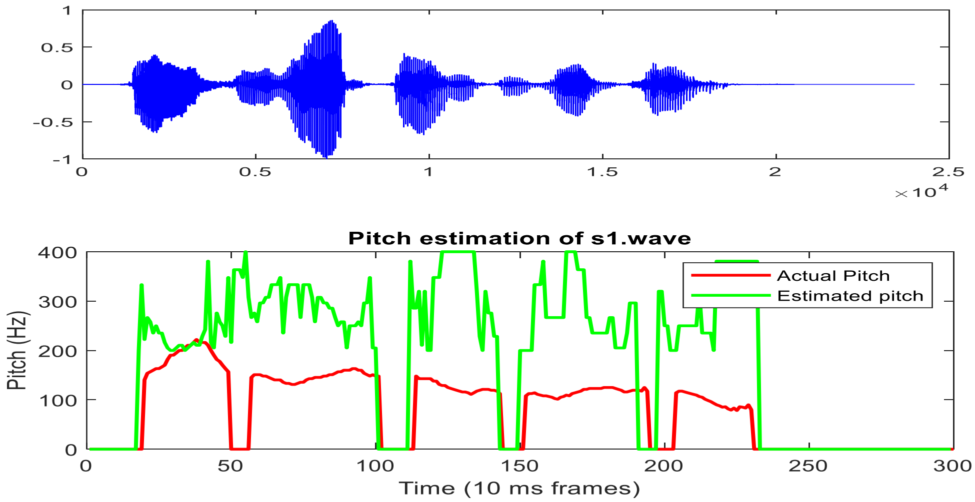


Figure 10- Comparison of estimated pitch with actual pitch

Note -Here we can analyze that cepstral method is not properly estimating pitch for low pitch voice samples.

For s3.wav (Female voice samples)-

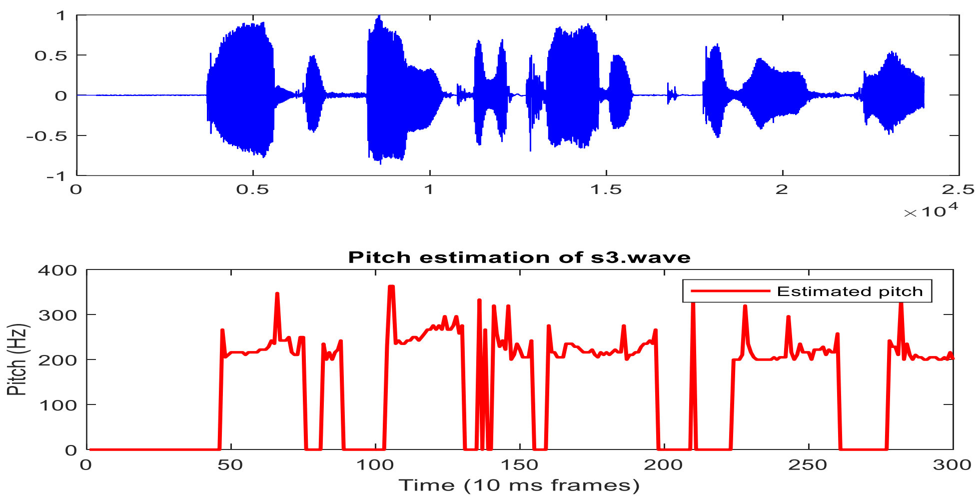
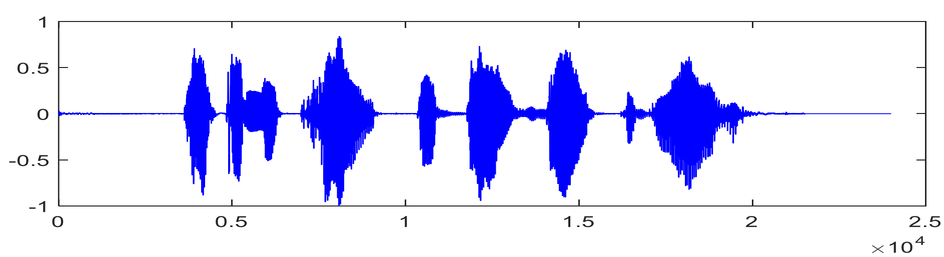


Figure 11- Pitch estimation of s3.wav

For s4.wav (male voice samples)-



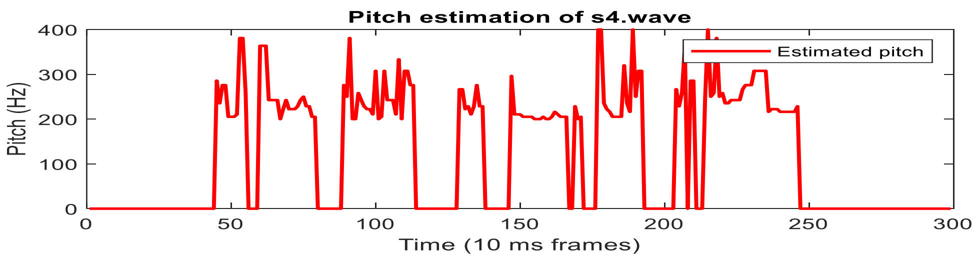


Figure 12- Pitch estimation of s4.wav

Comparison of both methods:

Pitch detection is a common task in digital signal processing that involves identifying the fundamental frequency of a sound signal. There are several methods for pitch detection, including the autocorrelation method and the cepstral method. Here's a brief comparison of these two methods: -

Auto-correlation method:

The auto-correlation method is a technique used to estimate the pitch of a sound signal. It involves computing the auto-correlation function of the signal, which quantifies the similarity between the signal and a time-delayed version of itself. By analyzing the peaks in the auto-correlation function, the pitch can be estimated as the delay that maximizes the function. This method is commonly used in speech processing and audio analysis to determine the fundamental frequency of a sound signal.

Pros:

- The auto-correlation method is simple to implement and computationally efficient for both high and low frequency voices.

Cons:

- The autocorrelation method can be sensitive to noise and harmonics that are not related to the fundamental frequency.
- It may not work well for complex signals with multiple sources and non-harmonic components.

Cepstral method:

The cepstral method offers a distinct approach to estimating the fundamental frequency of a sound signal. It involves transforming the signal into the cepstral domain, which provides a logarithmic representation of the power spectrum. By examining the peaks in this transformed spectrum, the fundamental frequency can be determined.

Advantages of the cepstral method include its resilience to noise and harmonic interference, making it suitable for analyzing complex signals with multiple sources and non-harmonic components.

Drawbacks: It tends to be more computationally intensive compared to the auto-correlation method. Additionally, its effectiveness may **be limited for signals with low-frequency content,** as the method relies on a logarithmic representation.

In **conclusion**, the choice between the auto-correlation and cepstral methods depends on the characteristics of the signal and the specific requirements of the application. The auto-correlation method is preferable for simpler signals with few sources and harmonic components, whereas the cepstral method is better suited for more complex signals with multiple sources and non-harmonic components.