Speech Signal Processing

Assignment -2

1.a. Autocorrelation: The idea of autocorrelation is to provide a measure of similarity between a signal and itself at a given lag.Thus, the autocorrelation is the correlation of a signal with itself. Mathematically is

$$R(\tau) = \int x(t)x*(t-\tau)dt, -\infty < t < \infty$$

1.b. Zero Crossing Rate(ZCR): The zero-crossing rate is the rate at which a signal changes its polarity. It provides information about the rapid changes in the signal .

1.c.Mel spectrogram:A spectrogram with the Mel Scale as its *y* axis is defined as Mel spectogram. This Mel Scale is constructed such that sounds of equal distance from each other on the Mel Scale, also “sound” to humans as they are equal in distance from one another.
In contrast to Hz scale, where the difference between 500 and 750 Hz is obvious, whereas the difference between 8250 and 8500 Hz is barely noticeable.

1.d. LP spectrum is a representation of a signal's spectral content obtained by modeling the signal as a linear combination of its past samples.In LP analysis, the goal is to estimate a set of coefficients that best represent the linear relationships among the signal samples.

2.

Voiced speech occurs when there is vibration of vocal cords as air passes through them. This vibration produces a periodic waveform characterized by regular patterns in the speech signal.

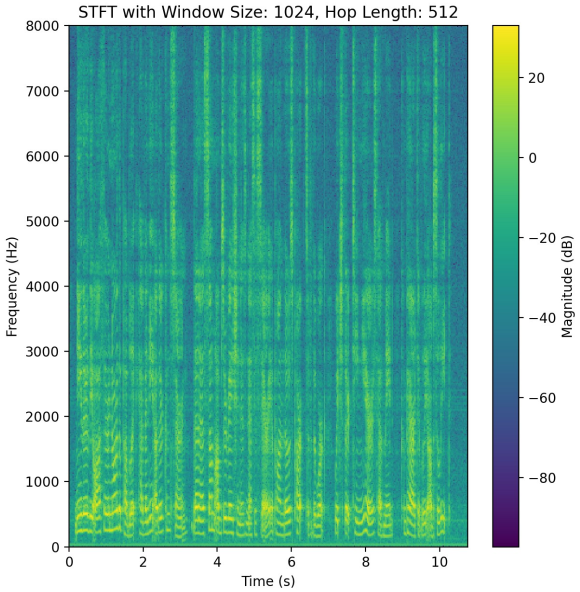
Unvoiced speech, on the other hand, occurs when the vocal cords remain apart and do not vibrate. The resulting speech signal is characterized by noise-like, turbulent patterns.

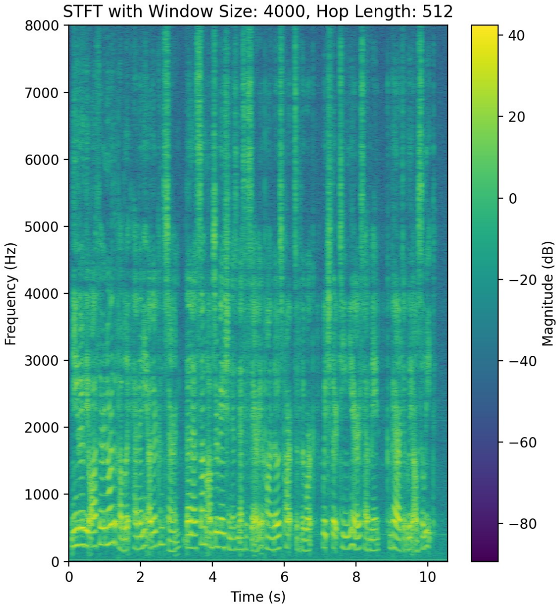
Three methods to distinguish voiced from unvoiced speech:

1. Short Time Zero Crossing Rate:unvoiced speech tends to have a higher zero crossing rate due to the noise-like nature of the signal whereas voiced speech tends to have a lower zero crossing rate due to the periodic nature of the vocal cord vibrations.
2. Short Time Energy:Voiced speech typically contains energy in the lower frequencies due to the periodic vibrations of the vocal cords. This results in a relatively higher energy concentration in the lower frequency range. While Unvoiced speech, being noise-like, doesn't exhibit as much energy in the lower frequency range. Its energy tends to be spread across a broader spectrum.
3. Normalized Error in LP Analysis:On modelling with linear prediction model, voiced region will give lower normalized error compared to unvoiced region. This is because the periodicity of voiced speech can be captured by the linear prediction model.

3.a. Effect of Window Size:

There is an inherent trade-off between frequency resolution and time resolution when we are changing the window size. on applying larger window frequency resolution increased and so we can distinguish closely spaced frequencies but time resolution is decreased and spectogram is appearing more smoother compared to the spectogram of smaller window.





b.effect of window shape:

The rectangular window has the poorest frequency resolution and the highest spectral leakage.
The triangular window offers better frequency resolution than the rectangular window but still has noticeable spectral leakage.
The Hann window provides the best compromise between time and frequency resolution, with reduced spectral leakage and well-defined main lobes.

Plots are attached in the polt folder.

4.b. No of frames is 19. I took window size of 1024samples and 512 samples for hop size.

4.e. on reconstructing back into time domain the voiced frame yield a signal with a clear pitch and harmonic structure, resembling a musical tone.
The unvoiced frame yield a signal that is noisy and lacks a clear pitch.