Speech and Signal Processing

Assignment 1:

Name:Ajay Ray

Roll:2021102032

1.a. Difference between sampling and quantization are below:

- In sampling, the values on the *y-axis*, usually amplitude, are continuous but the time or x-axis is discretized. While In quantization, time or *x-axis* is continuous and the *y-axis* or amplitude is discretized.
- Sampling is done prior to the quantization process.

1.b. Speech is quasi-stationary . So in speech analysis we divide the signal into smaller segments called frames. Each frame is analyzed individually to capture the characteristics of the speech signal over time. The frame shift refers to how much the starting point of consecutive frames are shifted along the time axis.

The frame shift is smaller than the frame duration in speech analysis to capture rapid changes. It also maintains stationary nature and enables smooth transition.

But it comes at a cost of more computation.

1c. The Hamming or Hann window is preferred over a rectangular window in speech signal processing because of the sharp shape of the rectangular window without any tapering at the edges.

- Because of that shape it abruptly cuts off the signal leading to discontinuity and hence creating unwanted side lobes and energy leakage in other frequencies other than main frequency.Hamming and Hann windows have gradually decreasing values towards the edges, which reduces the high-frequency content introduced by abrupt changes.

- The use of Hamming or Hann windows results in smoother spectrograms . Adjacent frames overlap more naturally . This overlap helps in tracking changes in speech characteristics across the frames.

- Hamming and Hann windows also help in frequency resolution having a wider bandwidth.
  It helps in distinguishing the closely separated frequencies.

1d. There are some assumptions to be considered for analyzing the speech. We have to assume

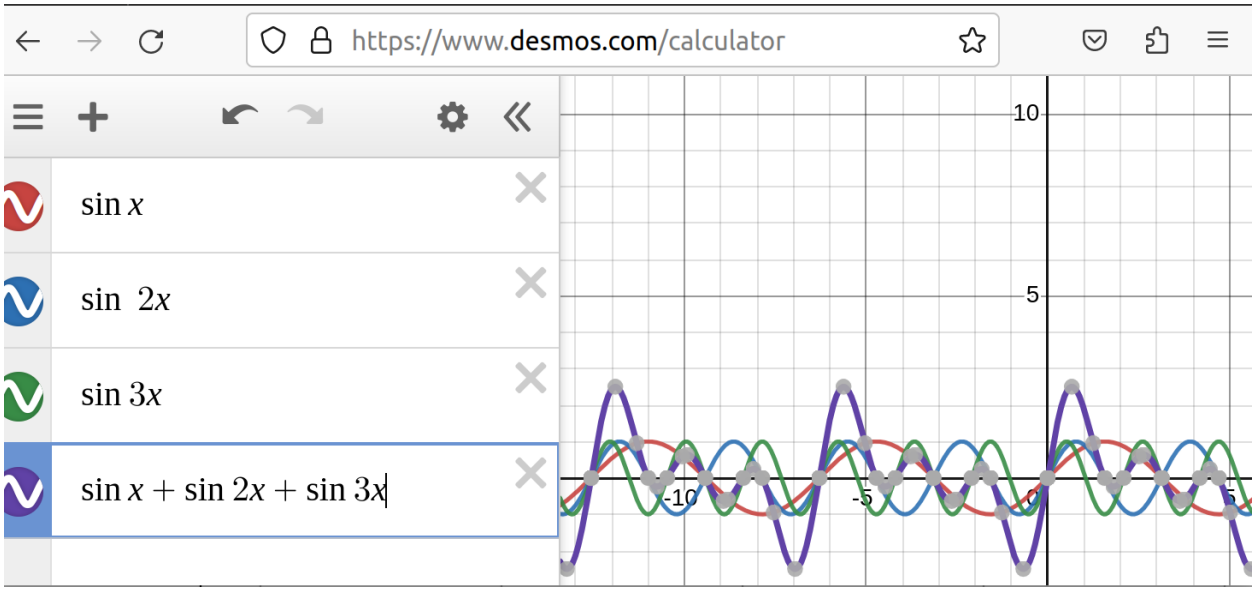
that speech is stationary and linear time invariant for a short interval of time.

Many aspects of speech, such as pitch and formants, are quasi-periodic in nature as it helps

in further analyzing the signal.

1e. An impulse train with a fundamental frequency FF can be generated by summing up sine waves at harmonically related frequencies. The harmonics are integer multiples of the fundamental frequency .

Visualization using desmos__



Each sine wave contributes its own oscillation pattern, and when combined, these patterns create a waveform that resembles a repeating sequence of spikes, forming the characteristic shape of an impulse train. The more harmonics you add, the closer the waveform gets to the idealized impulse train.

2.

a. <u>Co-articulation:</u>

Coarticulation refers to changes in speech articulation of the current speech segment (phoneme) due to neighboring speech.In other words, the way we produce one sound is affected by the sounds that come before or after it in a sequence of speech sounds.

A common example is The quality of a vowel sound can change slightly depending on the neighboring consonant sounds.

b. Epochs

Epoch is the instant of significant excitation of the vocal-tract system during production of speech.In speech processing, epochs can be used to analyze specific phonemes or speech sounds. By isolating individual speech sounds, researchers can study their acoustic properties more effectively.

2.c. Pitch

In simple terms pitch is a perceptual property of sounds that makes it possible to judge sounds as "higher" and "lower" in the sense associated with musical melodies.
Pitch is closely related to the frequency of a sound wave. Sounds with higher frequencies are perceived as higher in pitch, while sounds with lower frequencies are perceived as lower in Pitch

2.d Formants

Formants are resonant frequencies that define the characteristic sound of vowel and consonant speech sounds.They arise from the resonant properties of the vocal tract (mouth, throat, and nasal passages) during speech production.By analyzing the pattern of formants, our auditory system is able to identify and differentiate various speech sounds.

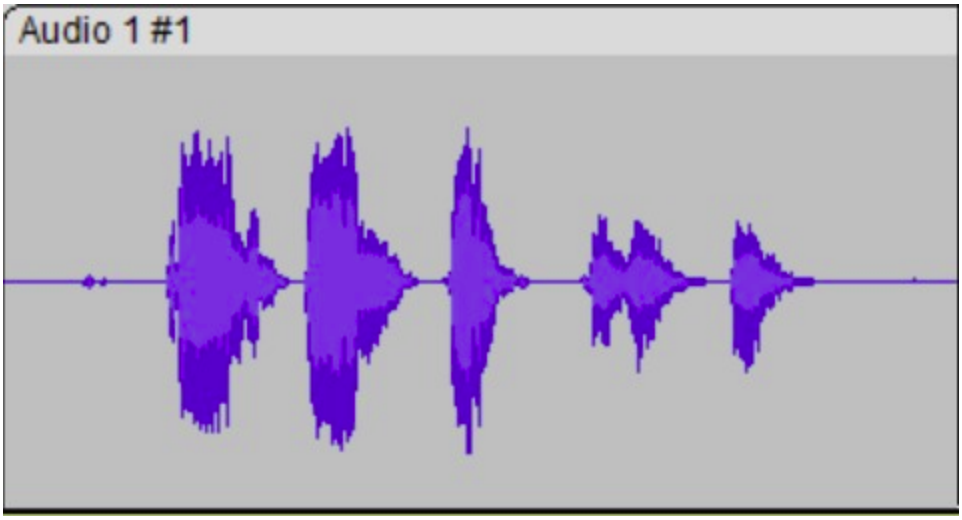
Q3.

True and the reason is below:

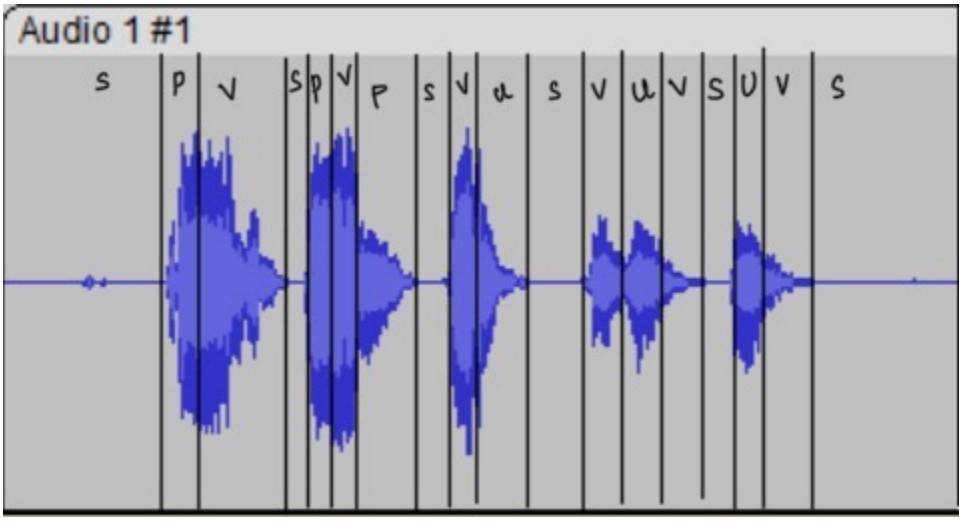
This is due to biological reasons. The size of our vocal cords determines our voice's pitch.
The vocal tract acts as a resonator for sound produced by the vocal folds. A shorter vocal tract females tend to amplify higher frequencies, contributing to a higher-pitched voice.

Q4.

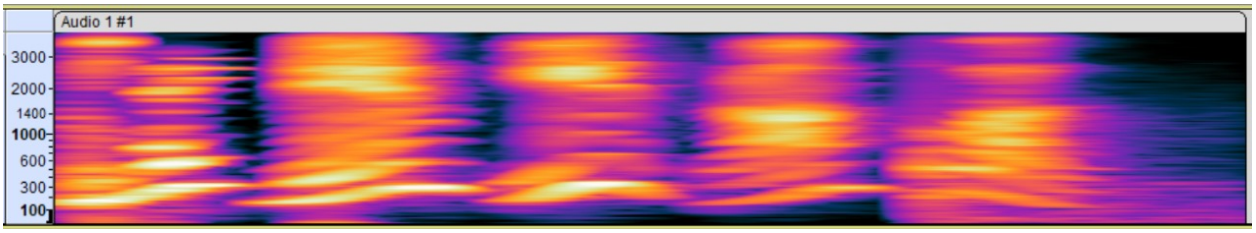
4a.



4.b. P:plosive s:silent v:voices u:unvoiced



4.d



4.c.

phenome	MoA and PoA	Formants
my		
/m/	nasal, bilabial	

/y/	semi vowel, palatal	
name		
/n/	nasal, dental	
/a/	voiced	384,754,1896
/m/	nasal, bilabial	
/e/	Medium Front Vowel	147,389,1098
is		
/i/	front high vowel	382,586,17854
/s/	fricative, dental	
Ajay		
/a/	central low vowel	324,584,1487
/j/	nasal, bilabial	