

SSP Assignment-1

• Srihari Bandarupalli

• 2021112006

Question 1

a)

- **Sampling** is the process of discretizing the time axis(x-axis) of the signal. Since we can't store a continuous signal we discretize it by taking sample after certain interval.
- **Quantization** is the process of discretizing the magnitude (y-axis) . Since we can store a value using only limited number of bits, we convert the continuous magnitude values to a finite set of discrete values.

b) Generally in speech analysis , the frame shift is smaller than the frame duration. This is done in order to ensure that we get overlapping windows. For example, when we use Hann window individually the middle part contributes more and the edges contribute less. When we use overlapping windows we can ensure that the entire signal contributes equally in the spectrogram.

c) In rectangular window there will be sudden transitions in the window which leads to distortion in the frequency domain. By using Hann or Hamming windows we avoid this distortion as transition between adjacent frames will be smooth and the frequency contents of signal are preserved

d)

- We take small window length(10-30ms) and we assume that signal is stationary in that region.
- We assume that the Vocal Track system is linear in given window and we design a Source - Filter model which is an LTI system.
- We use Linear Prediction Analysis

e)

- Let's consider a sinusoid signal with a phase shift $\pi/4$ and fundamental Frequency of $1/F$.

$$x(t) = \sin(2\pi F t + \frac{\pi}{2})$$

- Now if we sample this signal with a sampling frequency of $T \Rightarrow$ we get a Impulse train

$$p(t) = x(t) \cdot \sum_{n=-\infty}^{n=\infty} \delta(t - nT) \Rightarrow p(t) = \sum_{n=-\infty}^{n=\infty} \sin(2\pi nFT + \frac{\pi}{2})$$

Question 2

a) **Co-articulation** is the phenomenon where the adjacent phones influence each other's articulation which changes the acoustic properties of the phones. Ex:- \hari\ . Here the acoustic properties of \a\ depend both on \h\ and \r\.

b) **Epochs** are the air impulses which hit the vocal track system when vocal folds open. They are used in pitch estimation and synthesis.

c) **Pitch** is defined as the rate of vocal fold vibration. It is the perceptual quality of sound that corresponds to the fundamental frequency of a sound wave.

d) **Formants** are defined as the dominant vocal track resonances which have relatively high energy and represents the different speech sounds and vowels.

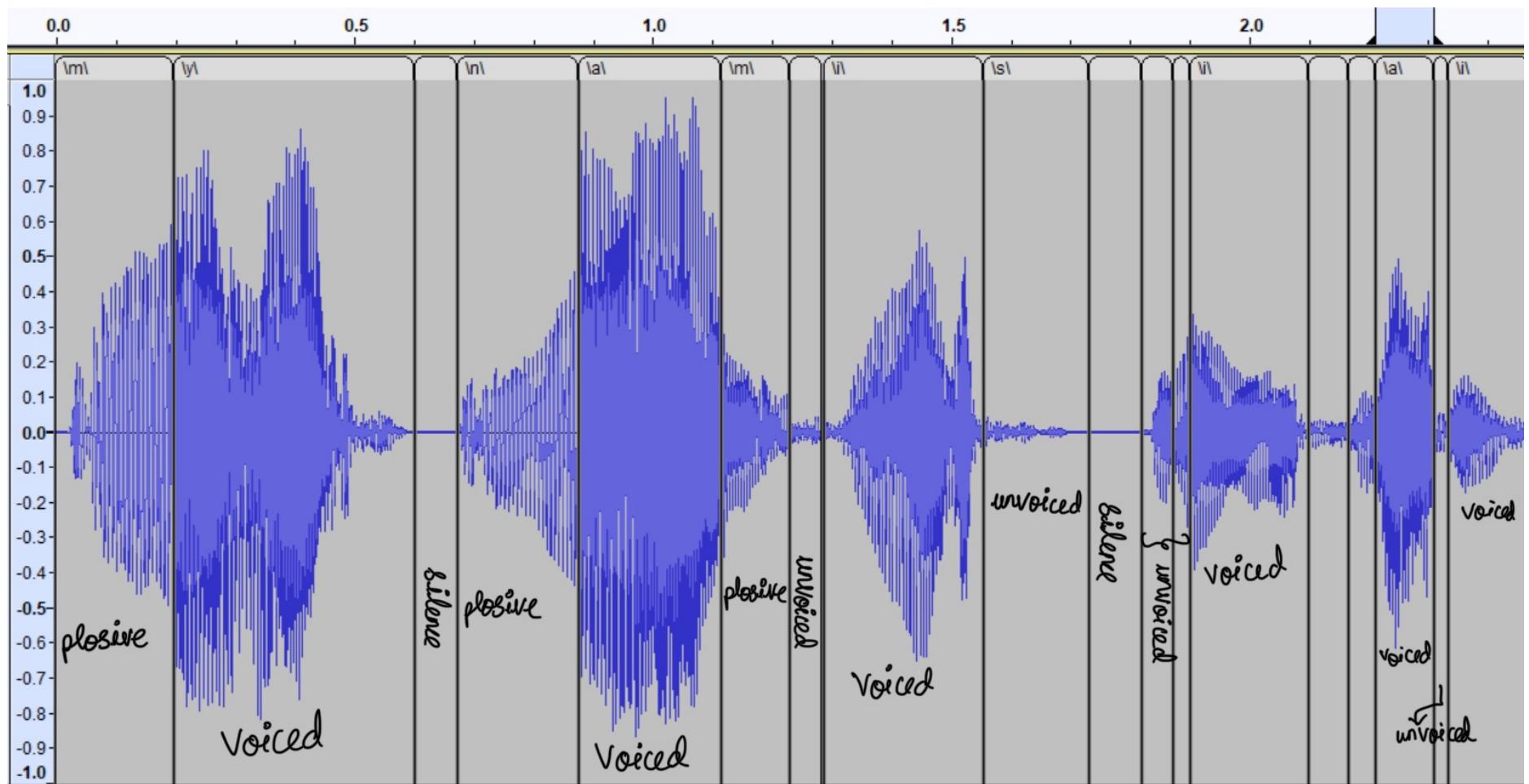
Question 3

- Yes, female pitch is more when compared to male pitch.
- This is because men have longer, thicker vocal chords that vibrate at low frequencies, giving them a deeper voice, whereas women have shorter, thinner vocal cords that vibrate more quickly and produce higher frequency, leading to high pitched voice.

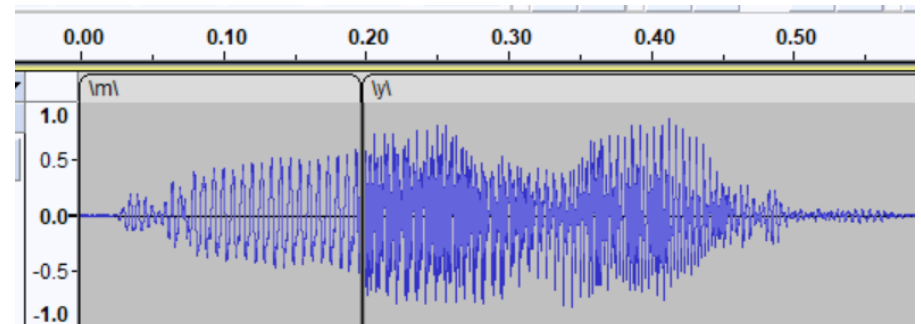
Question 4

a) Waveform

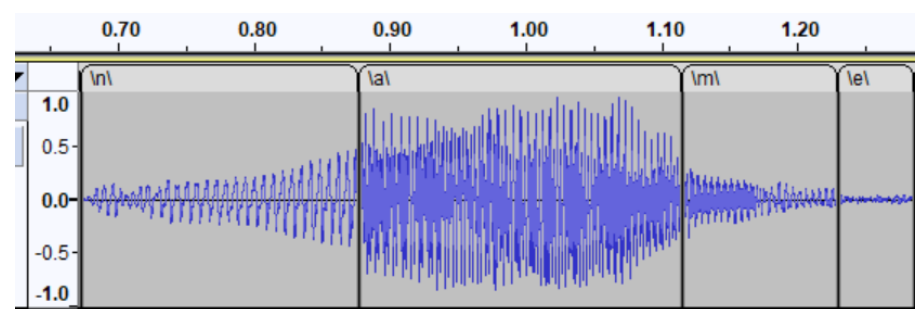
- My name is SRIHARI



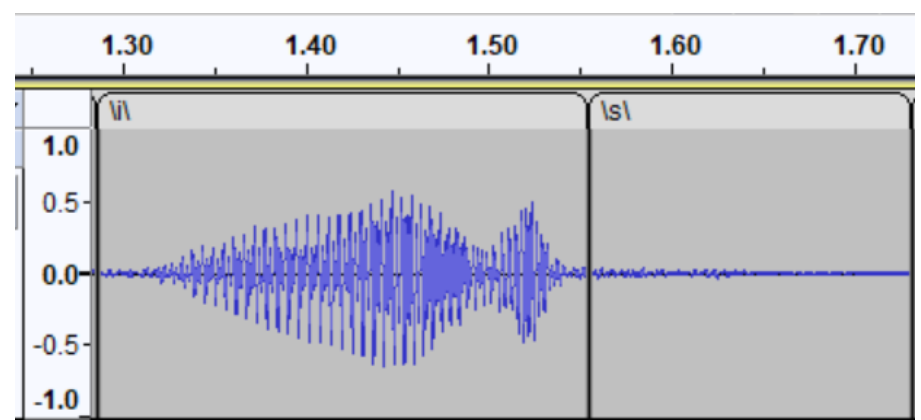
- my



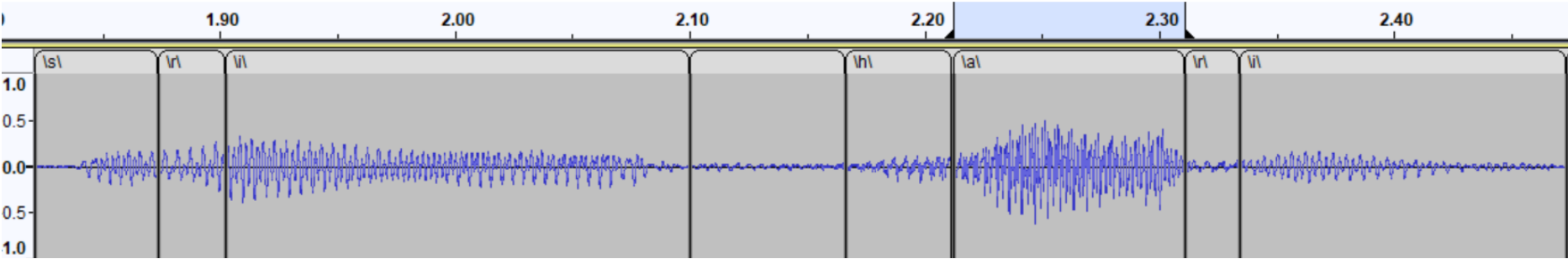
- name



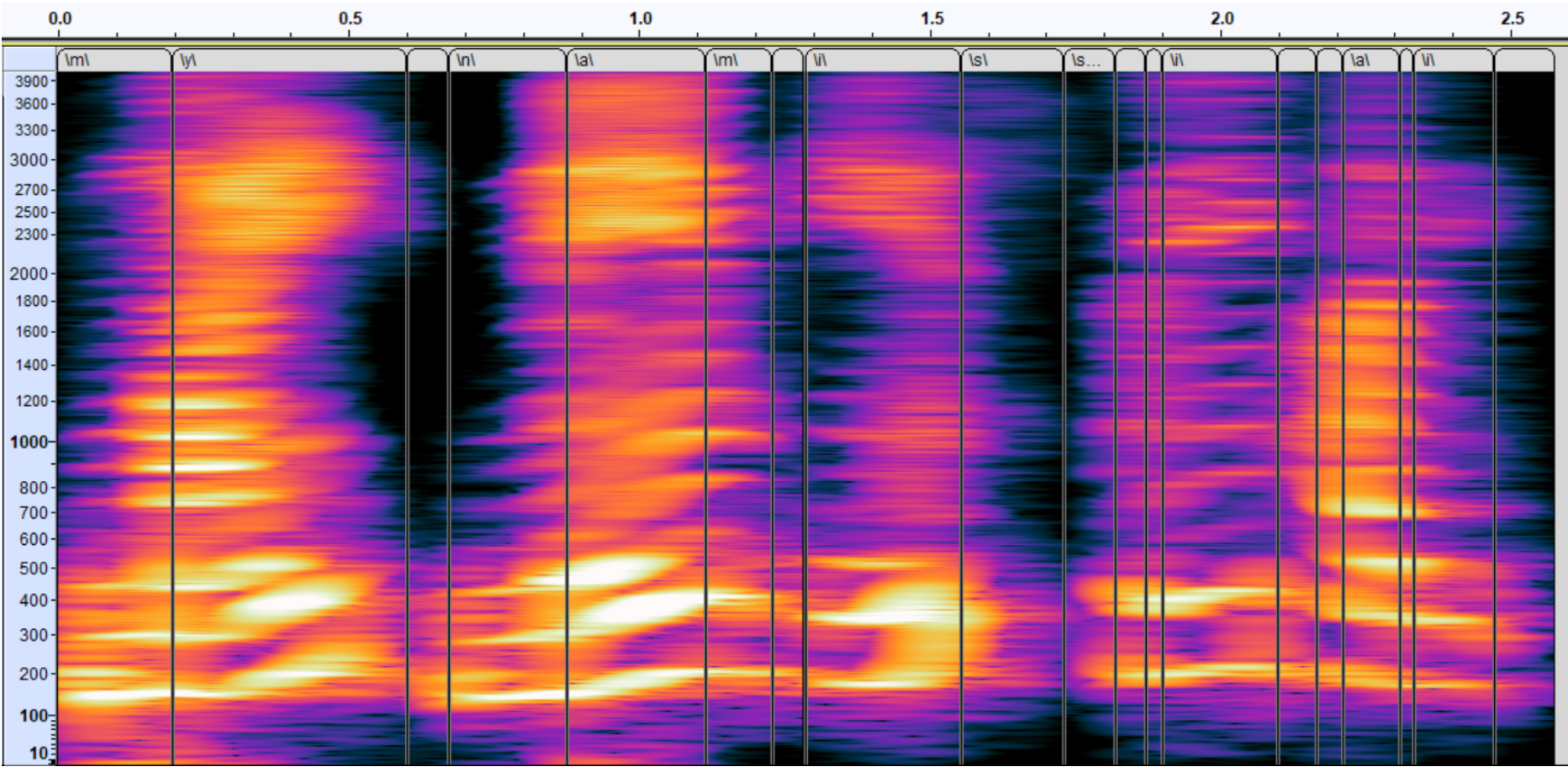
- is



- srihari



Mel-Spectrogram



(b, c, d)

- Manner of Articulation and Place of articulation for all phones
- Formants for all the vowels

phoneme	MoA and PoA	Formants (in Hz)
	my	
/m/	Nasal – Bilabial	
/y/	Semi Vowel - Palatal	
	name	
/n/	Nasal – Dental	
/a/	Central Low Vowel	394,791,1025
/m/	Nasal – Bilabial	
/e/	Medium Front Vowel	173,389,1044
	is	
/i/	Front High Vowel	352,516,1031
/s/	Fricative - Dental	
	srihari	
/s/	Fricative - Dental	
/r/	Semi Vowel - Alveolar	
/i/	Front High Vowel	352,516,1031
/h/	Fricative - velar	
/a/	Central Low vowel	394,791,1025
/r/	Semi vowel - Alveolar	
/i/	Front High Vowel	352,516,1031

Question 5

- in jupyter notebook