Assignment - 1

Speech Signal Processing

2020102004, Anjali Singh

Q1. **Speech** is the expression or the ability, that can be verbal or non-verbal, to express thoughts, feelings and basically communicating and transferring messages.

Speech signal is a non-stationary signal unlike other stationary or even non-stationary signals. It is non-stationary since it can have different frequency components every interval, that is, the characteristics of the signal are changing every interval.

This means that it can have multiple frequency components, in a given time interval. The time period usually is very short, i.e., 10-30 ms, rather than 200 or 250 ms.

Q2.

- **Relation between Fourier transform and Z-transform:**

  Fourier transform of a sequence $x(n)$ can be defined as:-

  $X(\omega) = \sum_{n=-\inf}^{\inf} x(n)e^{-j\omega n}$

  Z- transform of a sequence $x(n)$ can be defined as: -

  $X(z) = \sum_{n=-\inf}^{\inf} x(n)z^{-n}$

  Here, we can see that if we replace, z with $re^{j\omega}$, where $r = \mod z$, which is 1 in case of a unit circle.

  The relation is: -

  if $z = e^{j\omega}$, $X(z) = X(w)$

- **Coarticulation:** It refers to the changes in the speech articulation of the current speech due to neighboring speech. There is influence of the preceding or following speech sound in the current isolated speech sound. *[Cited by Wikipedia]*

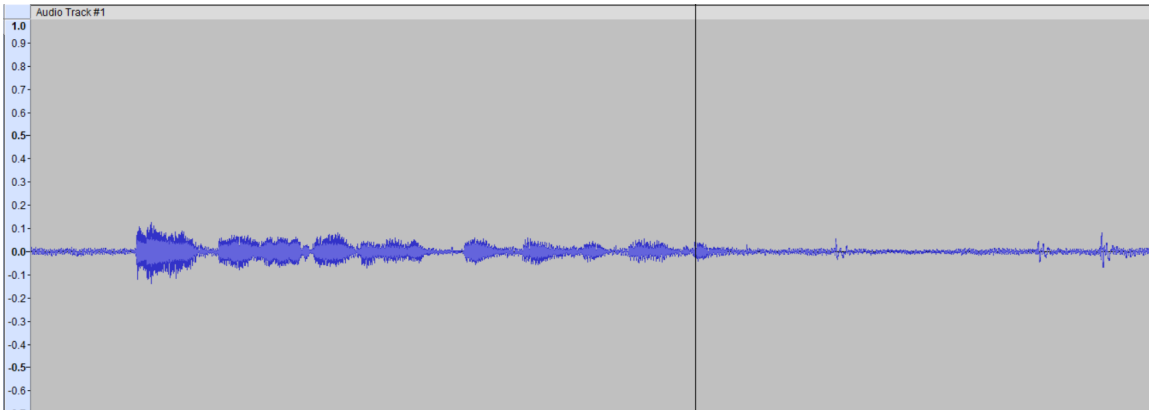
- **Fundamental Frequency:** The lowest base frequency that's generated by any particular instrument is known as fundamental frequency. It works as the supply frequency. The frequency range of that particular instrument is the integral multiple of the fundamental frequency.
- **Formant:** It is the most prominent/dominated frequency from a human vocal tract.

Q3. "Female pitch is more when compared to male pitch."

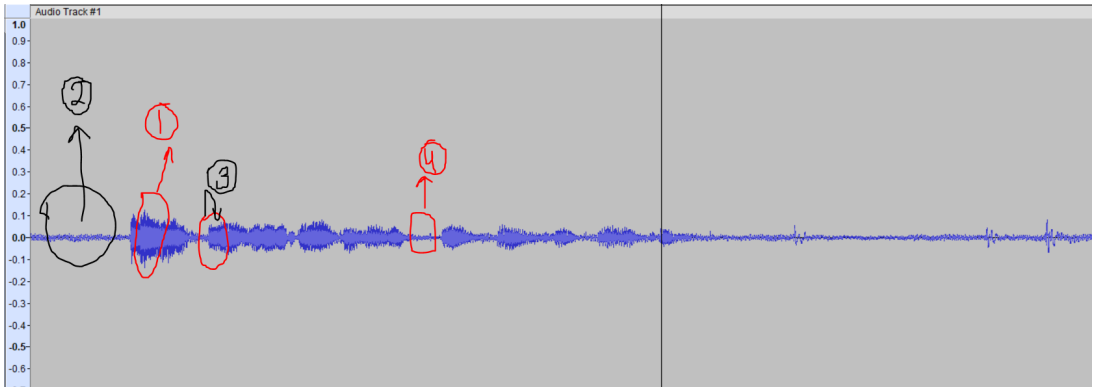
The above statement is true. This is because women have shorter and lighter vocal chords, which results in higher frequency than men. Males have a lower voice because males have longer vocal chords and that's why their voice becomes deeper in pitch.

Q4. Software used: Audacity

- Display the waveform
  - This is the time-domain waveform

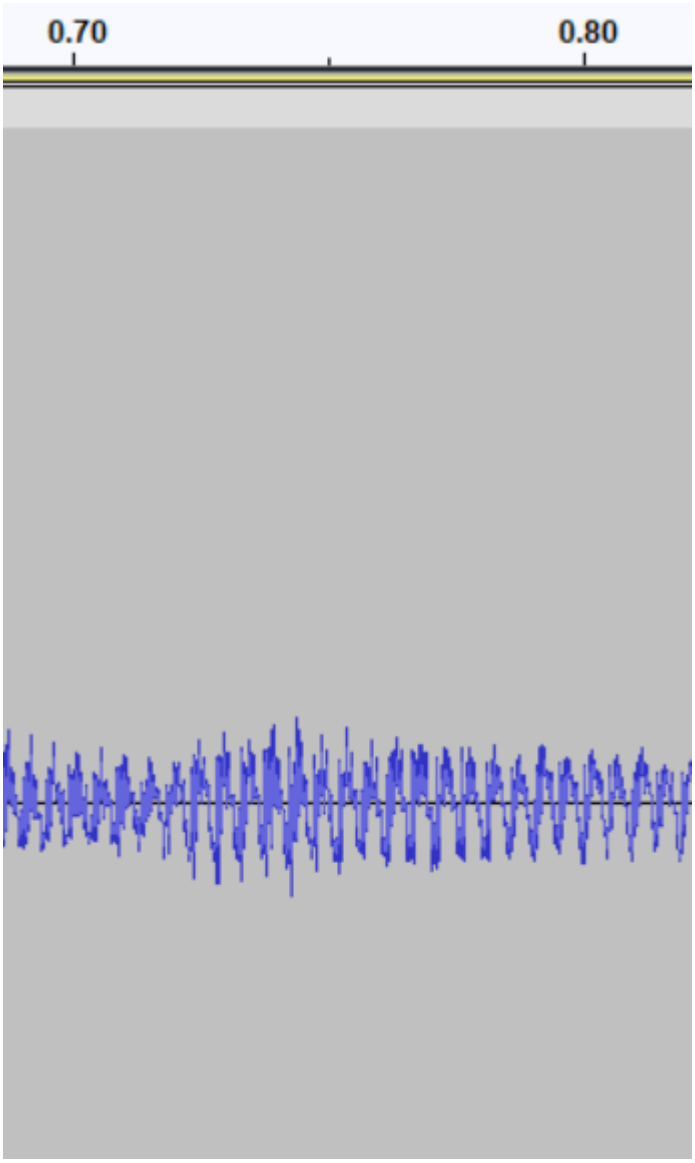


- Identify and mark the voiced, unvoiced, silence and plosive regions.
  - With the complete waveform, marked as 1. voiced 2. unvoiced 3. plosive 4. silence

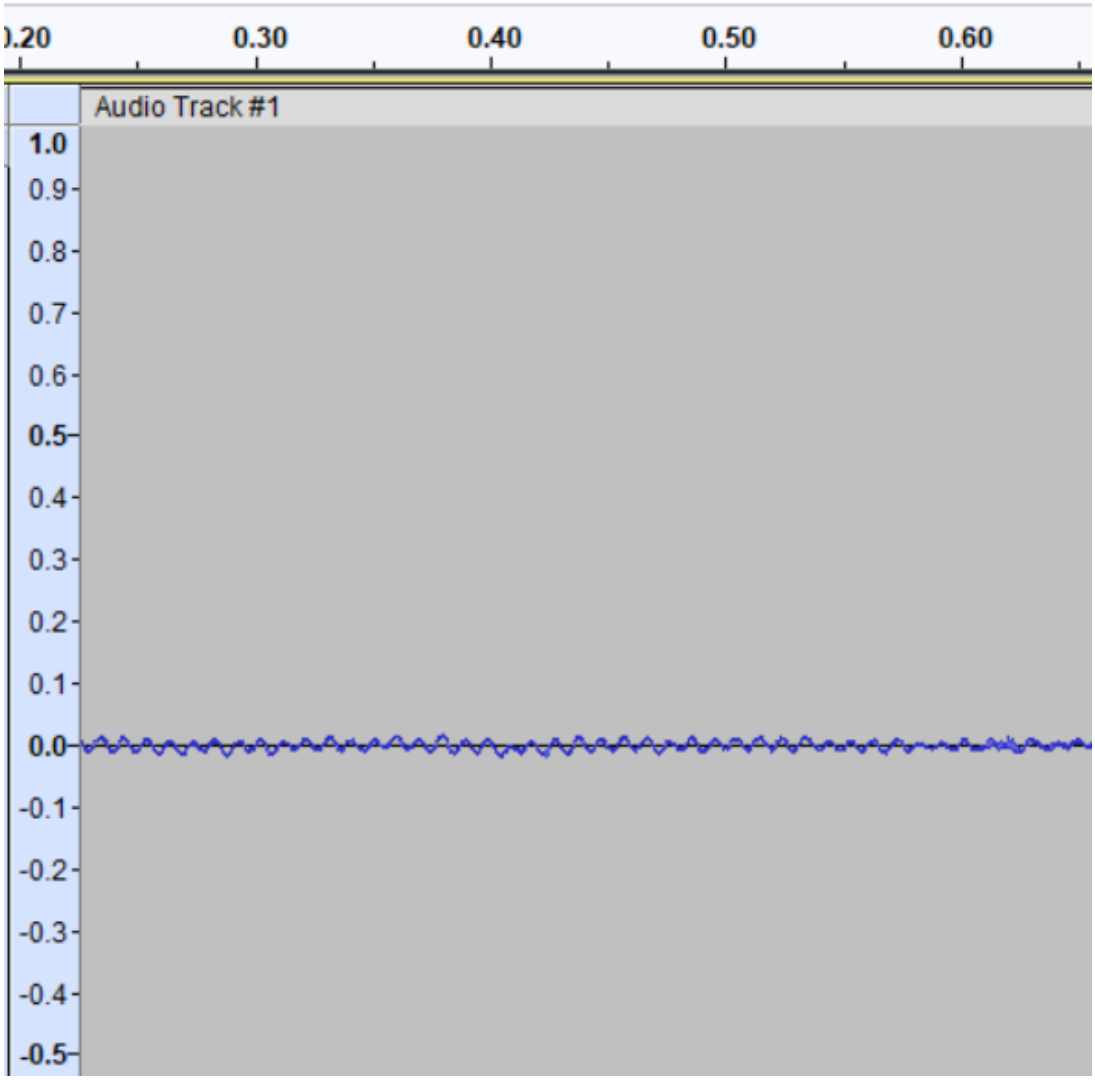


All the regions with visible high amplitudes are considered as voiced regions

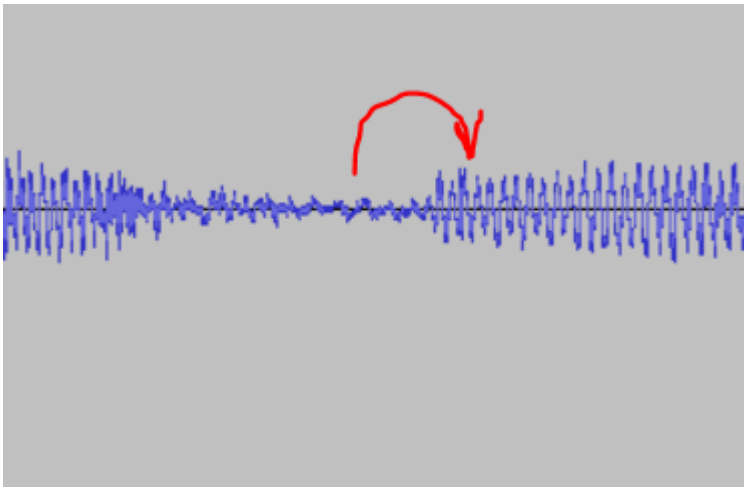
- Voiced (zoomed-in)



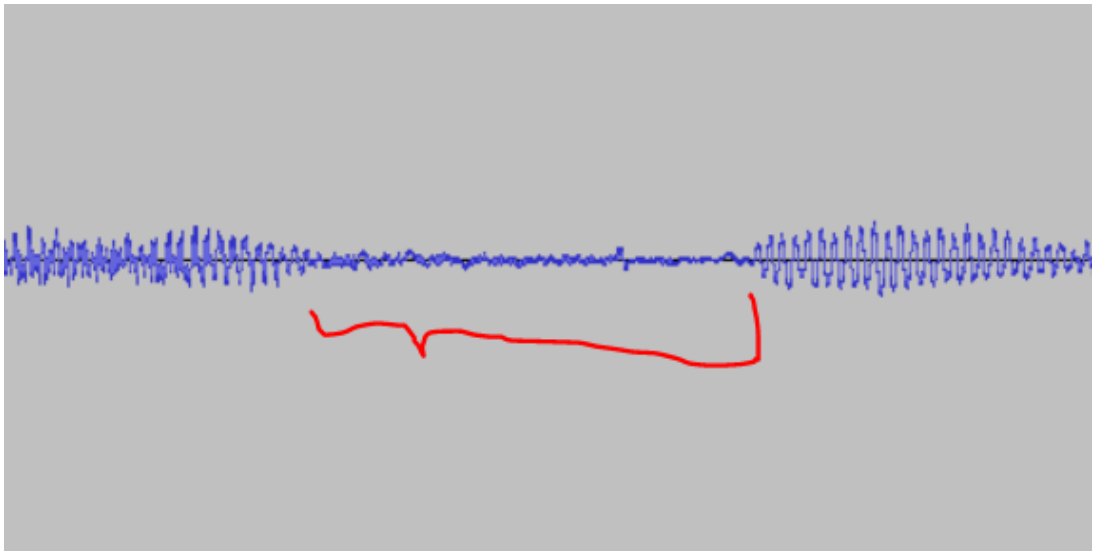
- Unvoiced (Zoomed-in)



- Plosive region (zoomed-in)



- Silence region



- Acoustic-phonetic description of the regions (MOA and POA)
  - The statement is “I am daughter of Mrs. Ratna Singh.”
  - Breaking down the words here as: -

    I/am/daughter/of/Mrs./Ratna/Singh

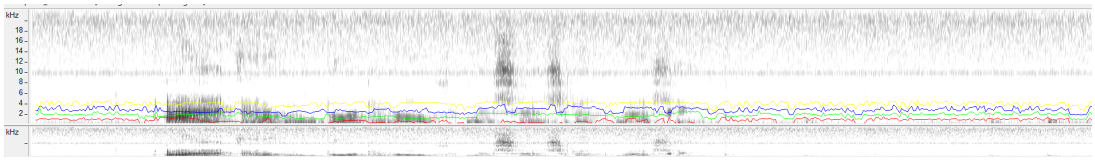
    I/a/m/d/a/u/g/h/t/e/r/o/f/m/i/s/s/e/s/r/a/t/n/A/s/i/n/g/h

Vowels/Consonants	Breakdown	Description
i	/ai/	vowel
Am	/a/, /m/	vowel followed by bilabial nasal stop
Daughter	/d/, /au/, /t/, /e/, /r/	alveolar, voiced, unaspirated followed by dental stop followed by alveolar, unvoiced, unaspirated followed by vowel followed by alveolar semi-vowel stop
of	/au/, /ph/	vowel followed by unvoiced aspirated bilabial stop
Mrs.	/m/, /i/, /s/, /s/, /e/, /s/	nasal followed by front vowel followed by voiced unaspirated stop.
Ratna	/r/, /A/, /t/, /n/, /a/	Unvoiced un-aspirational bilabial stop followed by middle vowel
Singh	/s/, /i/, /n/, /gh/	fricative followed by vowel

- Time-varying system description
  - I - /ai/ - tongue hump at alveolar ridge with wide opening
  - am
    - /a/ - vocal chords vibration
    - /m/ - closure at lips
  - daughter
    - /d/ - lips open and constraint near teeth, tongue bent down in the middle
    - /au/ - constrain near teeth released, vocal chords vibration
    - /t/ - tongue tip rise to the top
    - /e/ - constraints released
    - /r/ - mid tongue region touches the alveolar region
  - of
    - /au/ - constraints released, vocal chords vibration
    - /ph/ - tongue hump is less
  - Mrs.
    - /mi/ - lip movement going inward
    - /ss/ - tongue rolling
    - /es/ - tongue tip touching bottom
  - Ratna
    - /r/ - tongue touches top teeth side
    - /a/ - mouth opening
    - /t/ - tongue tip rise to the top
    - /n/ - tongue tip movement from up to down
    - /a/ - vocal chords vibration
  - Singh
    - /si/ - contraints released
    - /n/ - tongue tip movement from up to down

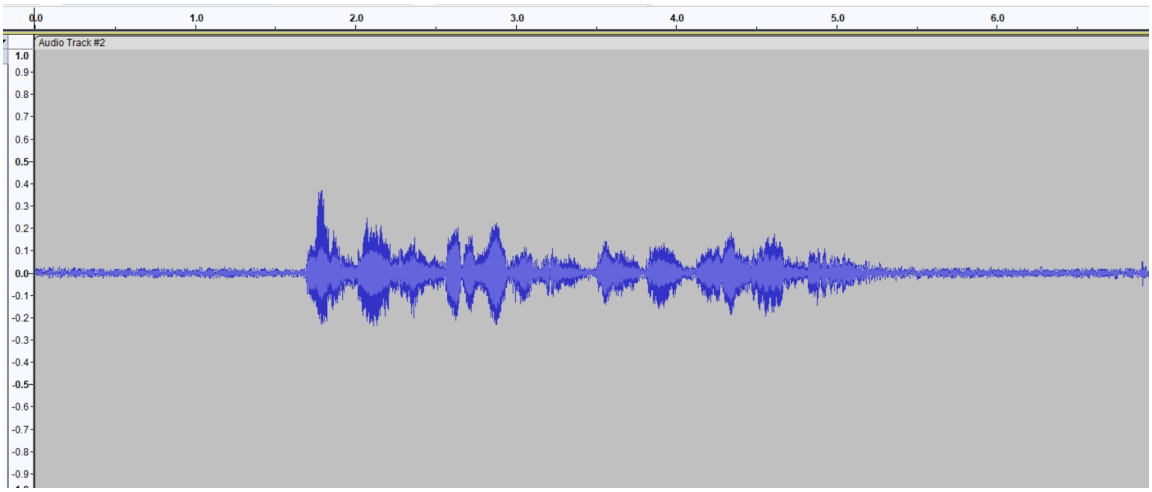
    - /gh/ - velum closed completely

- Spectral details for sound units present in the waveform
    - The spectrogram is as follows: (Software used: Wavesurfer)

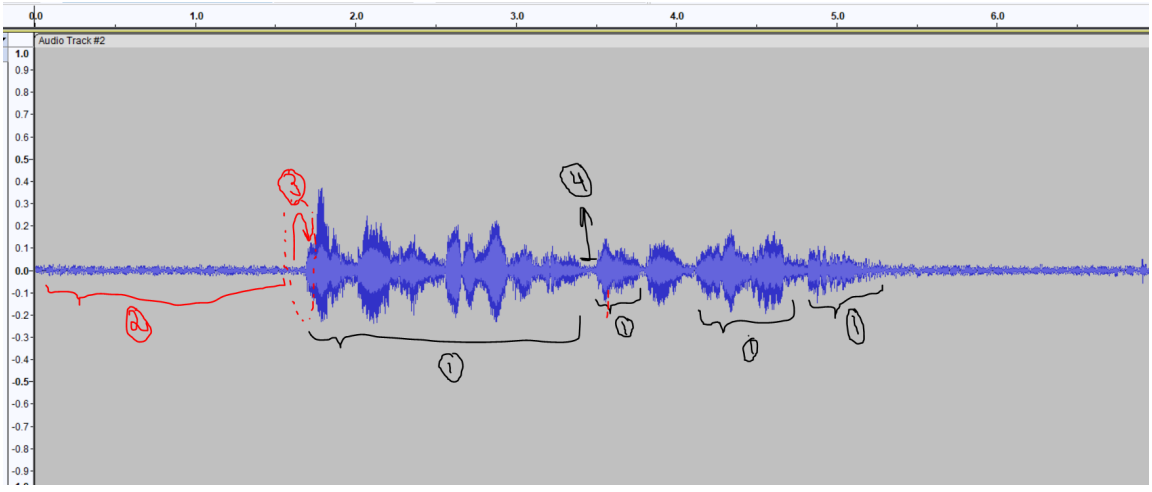


Q5. Software used: Audacity

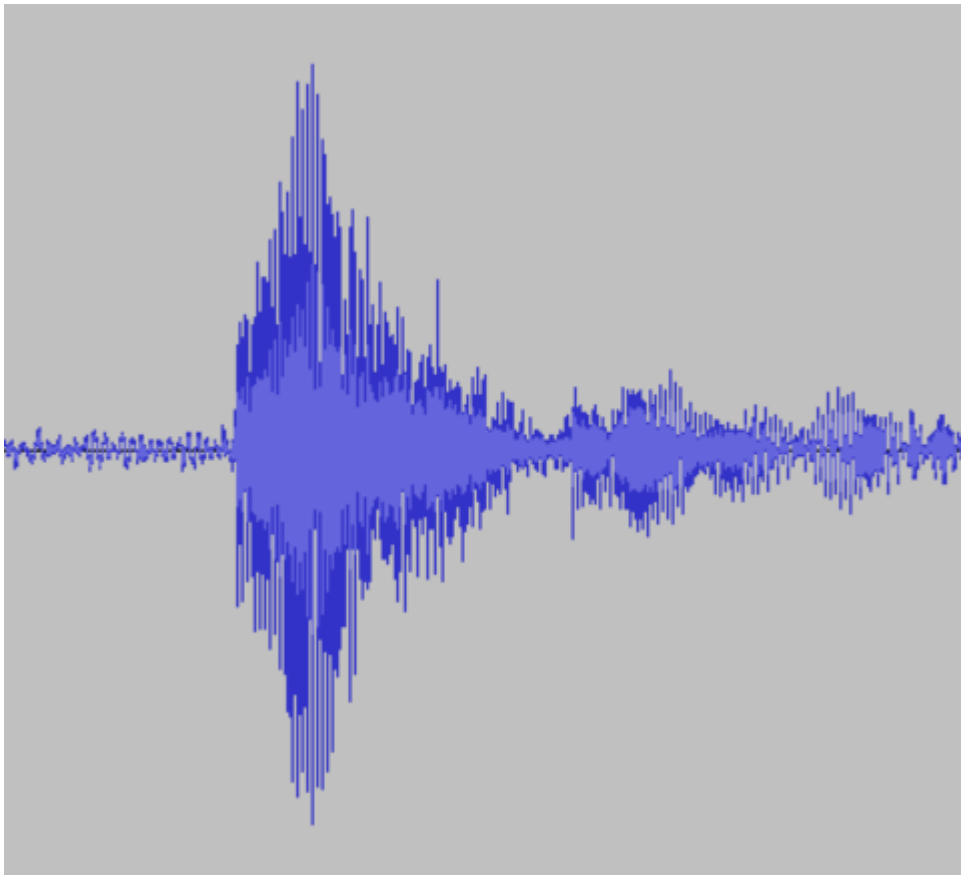
- Display the waveform



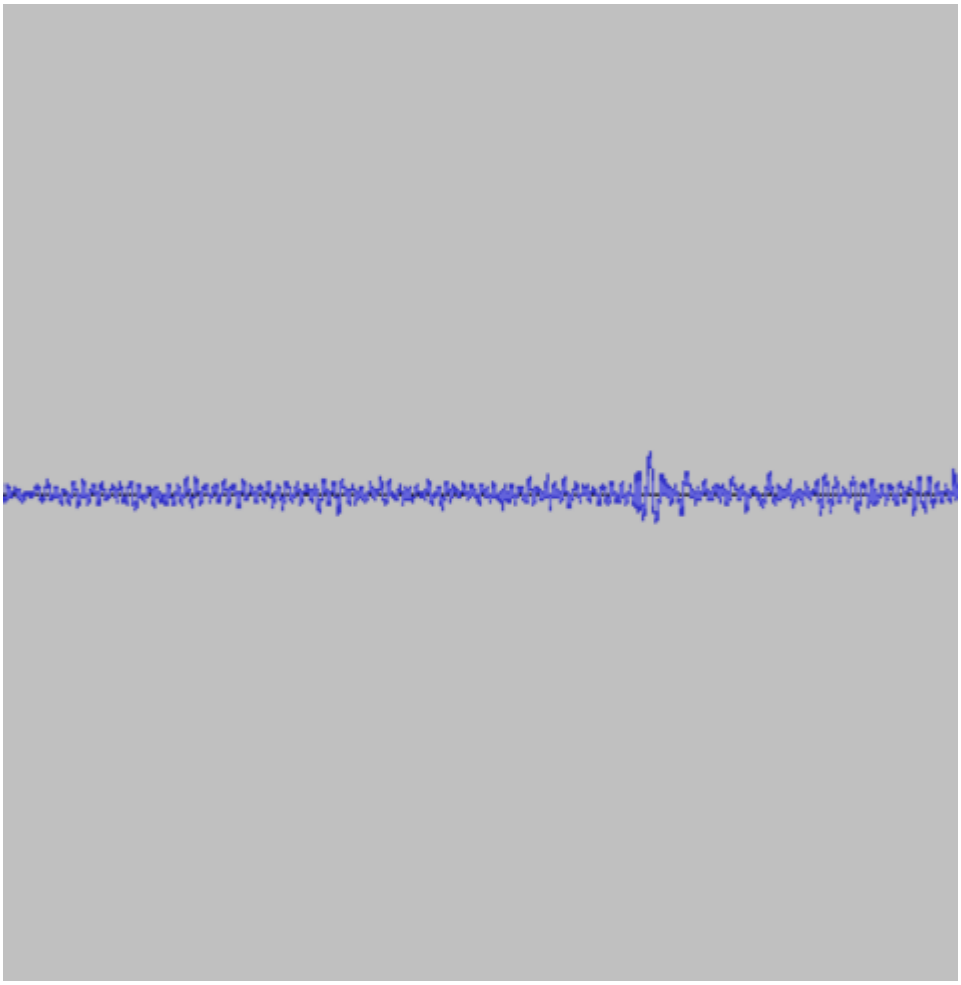
- Identify and mark the 1. voiced, 2. unvoiced, 3. plosive and 4. silence regions.



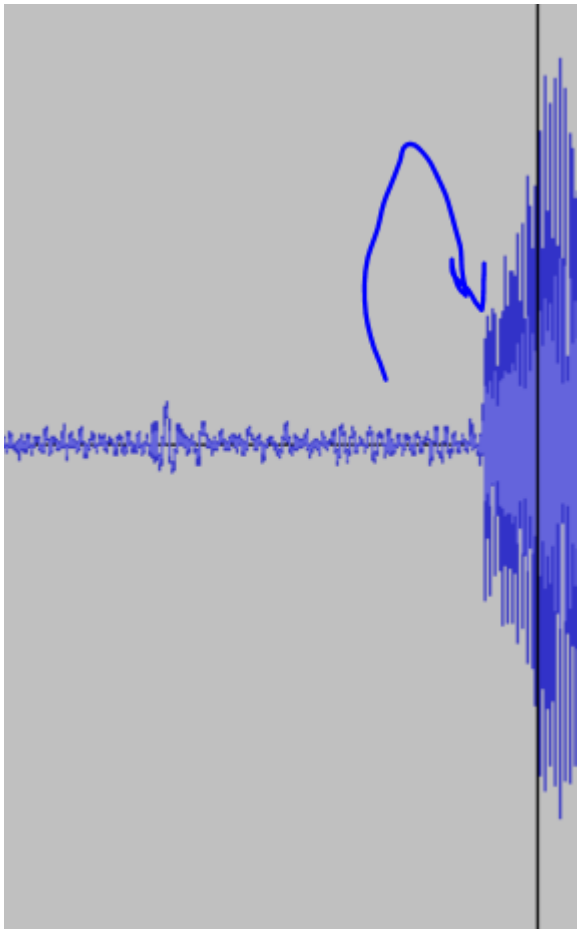
- Voiced (zoomed-in)



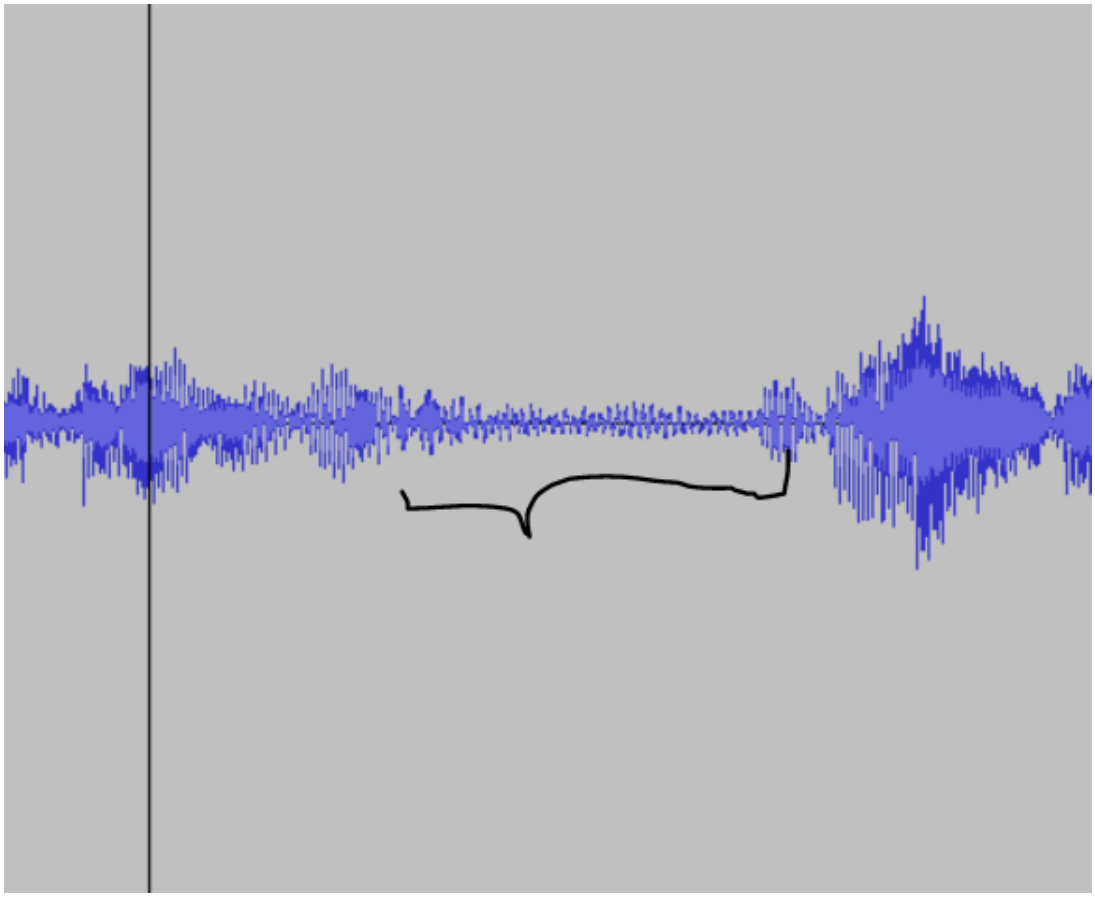
- Unvoiced (zoomed-in)



- Plosive (zoomed-in)



- Silence (zoomed-in)



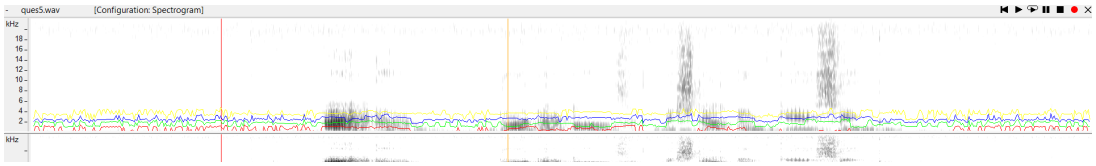
- Acoustic-phonetic description of the regions (MOA and POA)
  - The statement is “My father’s name is Mr. N.K. Singh.”

Vowels/Consonants	Breakdown	Description
my	/m/, /ai/	nasal followed by fricative dipthong
father’s	/ph/, /a/, /th/, /e/, /r/, /z/	unvoiced aspirated bilabial followed by vowel followed by unvoiced aspirated alveolar followed by vowel followed by alveolar semi-vowel stop
Name	/n/, /a/, /m/, /e/	dental nasal followed by vowel stop
is	/i/, /z/	vowel followed by aspirated bilabial
Mr.	/m/, /i/, /s/, /t/, /e/, /r/	aspirated bilabial
N.K.	/n/, /k/	nasal followed by dental
Singh	/s/, /i/, /n/, /gh/	fricative followed by vowel

- Time-varying system description
  - my

        - /m/ - closure at lips
        - /ai/ - constriction at lips released, tongue hump downwards in middle region, velum narrowed
    - father’s
        - /ph/ - constriction near teeth and lips
        - /a/ - constriction released, vocal chords vibration
        - /th/ - constriction applied at dental region
        - /e/ - constriction released, vocal chords vibration
        - /r/ - mid tongue touches the alveolar region
        - /z/ - tongue rise upward vibration
    - name
        - /n/ - constriction at dental region and nasal tract narrowed
        - /a/ - constriction released, vocal chords vibration
        - /m/ - closure at lips
        - /e/ - constriction released
    - is
        - /i/ - vocal chords vibration
        - /z/ - tongue tip touching mouth upwards
    - Mr.
        - /m/ - lip movement released
        - /i/ - vocal chords vibration
        - /s/ - tongue touching upward with a hiss sound
        - /te/ - tongue hump upwards concave
        - /r/ - tongue rolling
    - N.K. Singh
        - /n/ - tongue rise upwards
        - /k/ - tongue touching bottom
        - /si/ - contraints released

        - /n/ - tongue tip movement from up to down
        - /gh/ - velum closed completely
- Spectral details for sounds units present in the waveform
    - The spectrogram is as follows: - (Software used: Wavesurfer)



[Both for Q4 and Q5]

Observations: All the words in a sentence are formed by using words which themselves are formed by vowels and consonants. Every consonants/vowels have their own sound creation which is used in speech analysis which involves the use of various muscular movement. According to different sound production, energy is distributed that can be seen from the plots of frequency vs time plots. All this represents the speech, air distributions frequency-density distributions while the energy is transmitted.