INDIAN INSTITUTE OF TECHNOLOGY KHARAGPUR
Mid-Autumn Semester Examination 2023-24
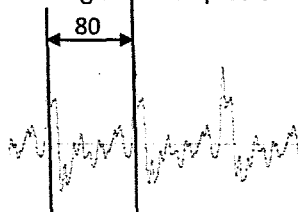
Advanced Technology Development centre

Subject: INTRODUCTION TO DIGITAL SPEECH PROCESSING   **Code: ET60007**

Time: 2:00 Hours   Full Marks [10x2+5x6] = 50
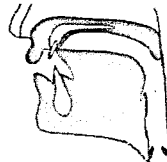
Answer all the questions
PART-A

1. Determine the F_0 of the following signal if the signal is sampled at **16 kHz**

2. Following figure represents the production of a consonant write the **place of articulation** of the consonant

3. First formant frequency of a steady state vowel is $F_1=450Hz$. Consider that the vowel is produced using a single lossless acoustic tube. What will be the length of the vocal tract? Where the speed of sound **c=350m/s**.

4. Intensity (I) of a speech signal is **80dB**; find out the Loudness (L) in Sones. Where $L = 445I^{0.333}$

5. An audio signal is recorded in uncompress format and store it in computer in **PCM WAV format**. The file size of the store signal is **24044 bytes**. If the signal is encoded with **8 bit** and recorded in MONO format determine the sampling frequency of recording.

6. Write the *state of the Glottis* during the pronunciation of the following phoneme?

   $/g^h/, /u/, /t^h/, /p/$

7. In the Source-System model of speech production, which part of the human acts as a sound source and sound filter respectively?

8. Write the name of the **perceptual parameters** of speech.

9. Time varying glottal resistance and inductance-both functions of glottal opening $A_G(t)$. If the glottis is completely open what will be the value of glottal impedance and volume velocity?

10. Why the women speech has high F_0 and *Formant* frequencies compare to men speech?

PART-B

1. Write your first name in *IPA* and syllabify it. Write the place and manner of articulation of each of the phonemes of your first name

2. (a) Figure-1 represents waveform of a VCV speech segment respectively. Where C represent consonant and V represent Vowel. Write the manner of articulation of the consonant represented by the figures
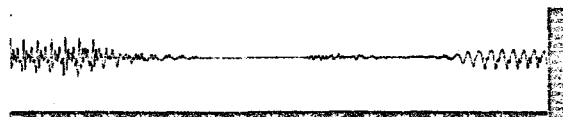

Figure-1

(b) Let length of the vocal tract *l=17cm* and the velocity of sound *c=340m/s* find the number of section required to generate *4 kHz* bandwidth voiced signal. If the voiced signal is modeled with all-pole model how many complex conjugate poles will be there?

3. Consider the two tube lossless vocal tract model is used to produce the vowel sounds. Assume that the termination at glottis and lips are completely lossless. The table-1 represents tube parameters for production of vowel */ae/* and */a/.* Determine the transfer function of the tube for each of the vowel. Determine the first formant frequency of the each vowel when sampling frequency is *Fs=10 kHz.* The transfer function of N connected tube can be express as in equation-1.

$$\frac{1}{V(z)} = \frac{U_G(z)}{U_L(z)} = \frac{2}{1+r_G}[1,-r_G]\left(z^{N/2}\prod_{k=1}^{N}\frac{1}{1+r_k}\right)\left(\prod_{k=1}^{N}\hat{R}_k\right)\begin{bmatrix}1\\0\end{bmatrix} \quad (1)$$

Table-1

Vowel	Tube-1 parameters		Tube-2 parameters	
	Length l_1	Area A_1	Length l_1	Area A_1
/ae/	4 cm	$2\ cm^2$	13 cm	$8\ cm^2$
/a/	9 cm	$1\ cm^2$	8 cm	$7\ cm^2$

4. A voiced operated lift operation is designee using the following words
   A. **STOP**, B. **UP**, C. **DOWN**, D.**FIRST**, E. **SECOND**, F. **THIRD.**

Figure-2 shows wideband spectrograms of one version of each of these words. Using your knowledge of acoustic phonetics, determine which wideband spectrogram corresponds to which word.
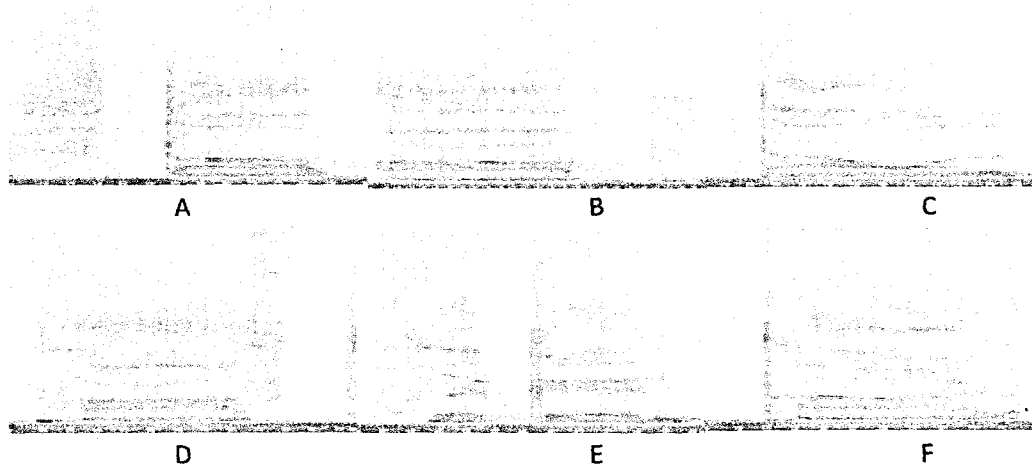

A   B   C

D   E   F

Figure-2

5. (a) A speech segment is produce by multiple lossless tube modeling and the frequency response of the output signal is given in figure 3. **Determine** the number of section required to implement the model.
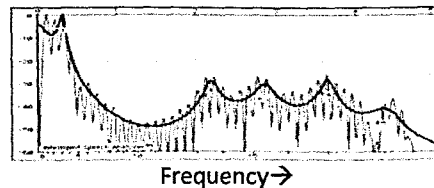

Frequency→

Figure-3

(b) Which formant frequency is related to tongue height, tongue position and lip rounding respectively?