Paper-1

Subject: DIGITAL SPEECH PROCESSING

Time: 3:00 Hours **Full Marks =100**

PART-A

1. Which formant frequency is related to tongue height and which formant related to tongue position?
2. A uniform tube is closed in both end find out the $2^{nd}$and 4^{th} formant frequency if the tube is excited at one end.
3. What are the different techniques for speech synthesis?
4. Why the women speech has high F_0 and formant frequencies compare to men speech?
5. Write the manner of articulation of the phonemes /ʃ/, /ʤʰ/.
6. An audio signal is recorded using the following format.

   F_S = 8 kHz, encoded with16 bit and recorded in MONO
   To store 3s signal in PCM WAV format how much memory is required?
7. Number of zero crossing is extracted from 20ms speech segment of a fricative sound and 20ms speech segment of a voiced sound which one has higher number of zero crossing and why?
8. Figure-1 represents the magnitude of the discrete-time Fourier transform of a steady-state vowel segment. The envelope of the spectral magnitude is sketched with a dashed line. Suppose that the sampling rate is 16 kHz meet the Nyquist rate. Determine the value of the first two formant frequency.

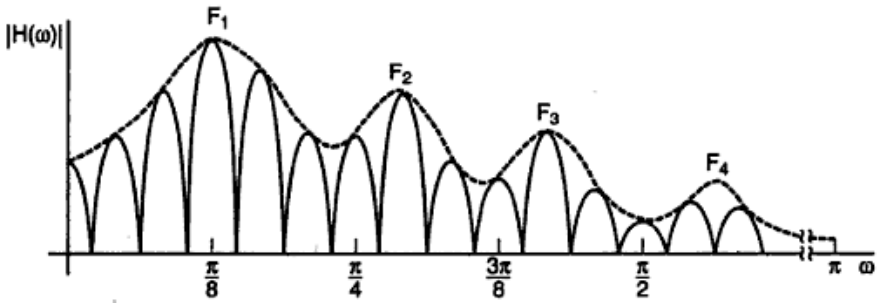


Figure-1

9. What are the Supra-segmental Features of Speech control the speech prosody
10. Which of the following is true for narrow band spectra?

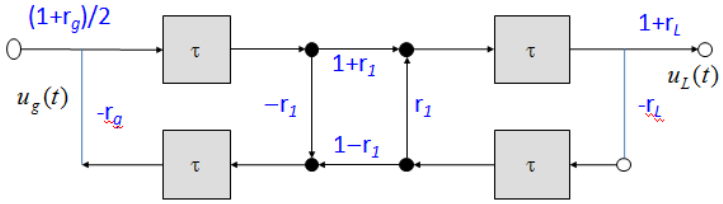
(I) Time resolution low, frequency resolution high

(II) Time resolution high, frequency resolution low

(III) None of the above.

PART-B

1. Consider a two-tube lossless vocal tract model including radiation and glottal boundary condition. The flow diagram of the system is as given below



(a) Show that the transfer function for the above two-tube model is as given in the equation below.

$$V(s) = \frac{0.5(1+r_g)(1+r_1)(1+r_L)e^{-2s\tau}}{1+(r_1r_L+r_1r_G)e^{-2s\tau}+r_Lr_Ge^{-4s\tau}}$$

(b) Sketch an extension of the flow diagram for three concatenated tubes.

2. (a) Show that a zero inside the unit circle can be expressed as an infinite product of pole inside the unit circle

(b) Length of a vocal tract is 17.5 cm and the speed of sound c=350 m/s. Determine the number of tube sections required to produce a voice of 5 kHz bandwidth.

(c) If the above voiced signal is modeled with all-pole model how many complex conjugate poles will be there.

3. (a) A speech signal frame has energy E_n^0 using the autocorrelation method the frame is analyzed and 3 PARCOR coefficient *{k₁, k₂, k₃}* are computed. Find the energy of the liner prediction residual $E_n^3 = \sum_m e_n^2[m]$ that would obtain by inverse filtering the speech signal frame. The inverse filter is designed using the above 3 PARCOR coefficient. Where $k_1 = 0.52; \quad k_2 = -0.25; \quad k_3 = 0.36$

(b) Derive the expression for LPC model gain of a voiced speech segment. If the order of the LPC analysis is 3 and LPC coefficients are *{α₁, α₂, α₃}* compute the model gain for a signal *x[n] = {1,2,1,-1,2}* where $\alpha_1 = 0.52; \quad \alpha_2 = -0.25; \quad \alpha_3 = 0.36$

4. (a)Draw the MFCC feature extraction block diagram and explain the requirement male scale frequency warping in MFCC feature extraction

(b) Why the delta and double delta MFCC is useful for speech signal classification.

(c) MFCC features are extracted from a speech signal if the speech signal is sampled at 16 kHz and initial filter bandwidth is 100Hz what will be the bandwidth of 13th filter.

5. What are the different techniques for speech synthesis? Draw a functional block diagram of concatinative speech synthesis system and describe the function of Grapheme to Phoneme conversion block. Write the name of three approaches of Grapheme to Phoneme conversion

6. (a) Show that cepstrum c[n] define as the inverse Fourier transform of the log magnitude is the even part of $\hat{x}(n)$ where $\hat{x}(n)$ of a digital signal *x[n]* is the inverse Fourier transform of the complex log spectrum.

(b) Draw the Cepstral Transform Coefficients (CC) feature extraction block diagram. 15