

**End Semester Examination 2012**

Subject: INTRODUCTION TO DIGITAL SPEECH PROCESSING

Code: ET60007

Time: 3:00 Hours

PART-A:-  $8 \times 2 = 16$

PART-B:-  $6 \times 14 = 84$

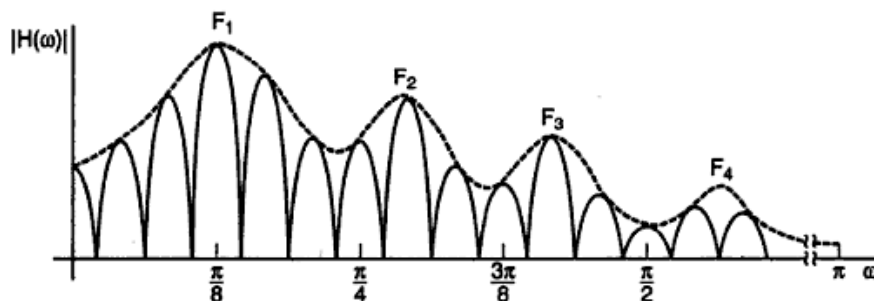
**Full Marks = 100**

**Answer all the questions of PART-A, and any Six questions from PART-B**

*(Please enclose the Annexure-1 along with the answer script)*

**PART-A**

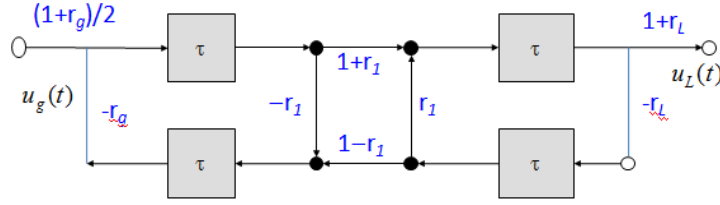
1. If cross section area of a uniform tube is  $A$ , and average atmospheric density is  $\rho$  write down the expression of acoustic capacitance and acoustic inductance
2. Write the phonetic transcription of the last word of your surname and syllabify it.
3. What is the meaning of just noticeable difference of Fundamental frequency ( $F_0$ ) of speech signal is 3-4%.
4. What are the Supra-segmental Features of Speech control the speech prosody
5. Which of the following pair of tones is perceived as louder tone and why?  
(a) 20dB level at 500Hz and 20 db at 200 Hz (b) 5dB level at 1 KHz and 5dB level at 8 KHz
6. Figure-1 shows a spectrogram of a CVC segment (In annexure-1). Where C represents consonant and V represents the vowel. Segment the following region  
i) Consonant vowel transition, ii) vowel consonant transition and iii) steady-state vowel
7. A signal  $X_n[e^{j\omega}]$  is the STFT of a signal  $x_n[n]$  using hamming window  $w[n]$  and  $w[n]$  has the effective bandwidth is  $B$  Hz determine the sampling rate in time where length of the window  $L=100$ , and  $F_s=10000\text{Hz}$ .
8. Figure-2 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 12 kHz meet the Nyquist rate. Determine the value of the first formant frequency.



**Figure-2**

## 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.

$$V(s) = \frac{0.5(1+r_g)(1+r_1)(1+r_L)e^{-2s\tau}}{1 + (r_1r_L + r_1r_g)e^{-2s\tau} + r_Lr_g e^{-4s\tau}} \quad 9$$

- (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 4

- (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. 7

- (c) If the above voiced signal is modeled with all-pole model how many complex conjugate poles will be there. 3

3. Complex cepstrum  $\hat{x}(n)$  of a digital signal  $x[n]$  is the inverse Fourier transform of the complex log spectrum.

$$\hat{X}(e^{j\omega}) = \log|X(e^{j\omega})| + j \arg[X(e^{j\omega})]$$

- Show that cepstrum  $c[n]$  define as the inverse Fourier transform of the log magnitude is the even part of  $\hat{x}(n)$  i.e. 14

$$c[n] = \frac{\hat{x}[n] + \hat{x}[-n]}{2}$$

4. (a) Derive the FBS constraint as in equation (1) in the frequency domain from equation (2)

$$w[n] \sum_{r=-\infty}^{\infty} \delta[n-rN] = w[0]\delta[n] \quad (1) \quad \sum_{k=0}^{N-1} W(\omega - \frac{2\pi}{N}k) = Nw[0] \quad (2) \quad 6$$

- (b) An unvoiced speech signal segment can be modeled as a segment of a stationary random process of the form:  $x[n] = w[n] + \beta w[n-1]$  8

What system can be used to recover  $w[n]$  from  $x[n]$ ? What is the normalized autocorrelation of  $x[n]$  at a delay of 1 sample?

5. (a) Explain the process of MFCC feature extraction. (b) Why the delta and double delta MFCC is useful for speech signal classification. (c) What are the different frequency domain speech parameters? 8+3+3

6. (a) What are the technological challenges faced in Speech Recognition? (b) Find the minimum edit-distance between the words 'SPEECH' and first word of your name. (c) What are the differences between a spoken language and written language and how does it affect automatic speech recognition? 3+7+4

7. What are the different techniques for speech synthesis? Draw a functional block diagram of part-name based concatenative speech synthesis system and describe it. 5+9

## Annexure-1

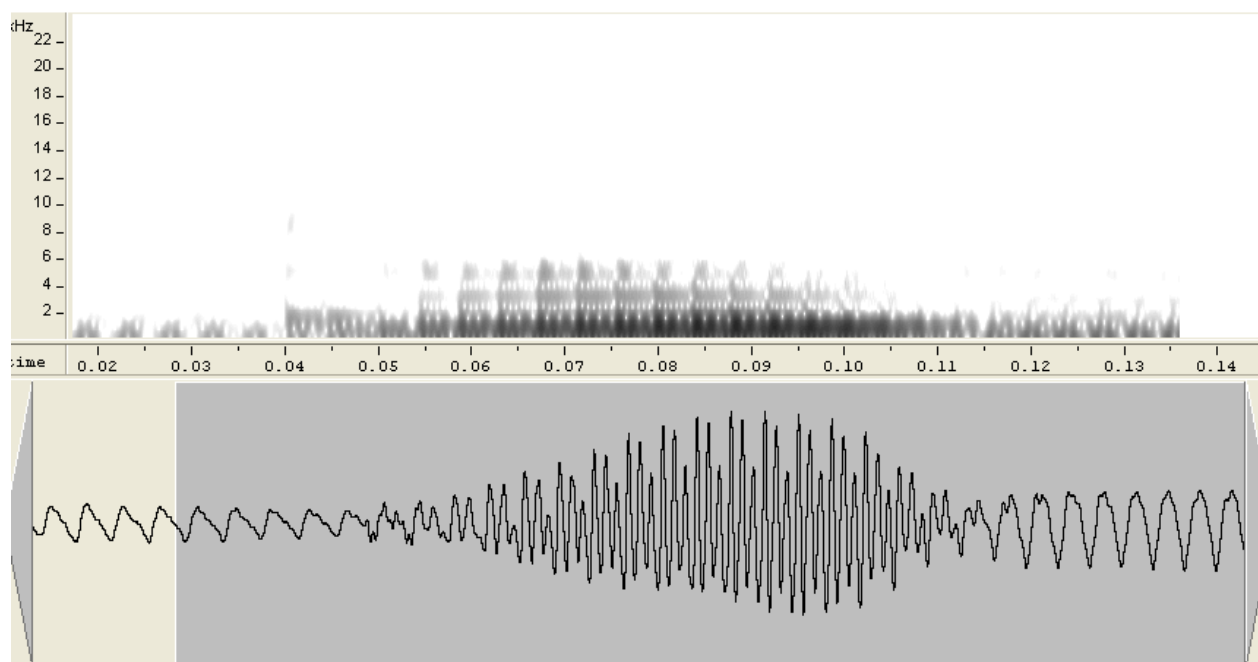


Figure-1