Indian Institute of Technology, Kharagpur

Centre for Educational Technology

**End Semester Examination 2017**

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

**Time: 3:00 Hours** PART-A:-10\*2=20; PART-B:-5\*16=80 **Full Marks =100**

***Answer all the questions of PART-A and PART-B***

***[Please enclosed the Annexure-1 along with the answer script]***

**PART-A**

1. What are the supra-segmental features of speech to control the speech prosody
2. 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 formant frequency.

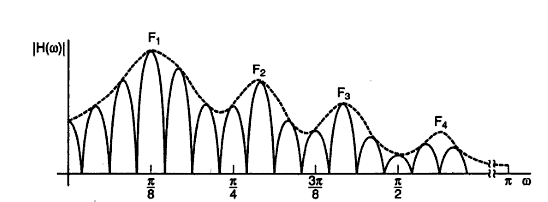


Figure-1

1. Write the three names of time domain methods for ***F0*** extraction?
2. Two source of speech signal producing a vowel **/a/**. The overall intensity of one source is ***20dB*** and other is ***21dB***. If a human being perceives the two sounds as same loudness sound explains why this happen?
3. Length of a uniform tube is ***17 cm*** and the tube is closed at one end find out the value of the pressure wave in open end.
4. First formant frequency of a steady state vowel is ***F1=250Hz***. 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***.
5. Write two advantages of having two Ears for sound perception
6. Acoustic intensity of an audio system is ***10W/m2***. Represent it in *dB* and find out the Loudness (L) in Sones. Where



1. A ***2 sec.*** audio signal is recorded and store in a computer as windows PCM “.wav” format. The size of the store signal is ***32.044 Kbyte***. Calculate the sampling frequency of the store audio signal if it encoded with ***16 bit***.
2. 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?

**PART-B**

1. **(a)** Figure 2 (I-VI) in annexure-1 shows plots of 6 speech segment short-time log magnitude spectra as obtained using a Hanning window of an appropriate length. The set of 6 spectra include vowel and consonant regions by a male, a female and a child talker.
2. Which of the spectra’s correspond to voiced sounds
3. Calculate the ***F0*** of the voiced sounds and commend which of the spectra are most likely to have been uttered by a child?
4. Which of the voiced speech spectra most likely come from an adult male; which from an adult female? [2+6+2]

**(b)** Figure-3 (i) and (ii) in annexure-1 represents waveform and spectrogram of a VCV speech segment where C represent consonant and V represent Vowel. Indicate the occlusion period, burst and VOT part of the consonant. Write the manner of articulation of the consonant represented by the figure-3. [4+2]

1. **(a)** For a given signal ***x[n] = {1, 2, 1,-1, 2} 3rd***order LPC analysis is done based on the following set of LPC analysis equation.
2. Calculate the value of the LPC coefficients *{α1, α2, α3}*
3. Compute the value of model gain. [8+4]



**(b)** A signal is sampled at ***16 KHz, 16 bit***, encoded with ***18th order*** LPC. Each of the LPC coefficients is encoded with ***2 byte***, Gain in ***2 byte***. Voiced unvoiced ***F0*** information is encoded using ***1 byte***. Calculate the compression ratio if frame rate is ***100 frame /sec***? [4]

1. **(a)** Draw the MFCC feature extraction block diagram [4]

**(b)** Complex cepstrum of a digital signal *x[n]* is the inverse Fourier transform of the complex log spectrum. [8]



Show that cepstrum *c[n]* define as the inverse Fourier transform of the log magnitude is the even part of



**(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 ***10th*** filter. [4]

1. **(a)** Draw a functional block diagram of a text to speech conversion system and explain the function of text normalization and grapheme to phoneme conversion block.
2. What are the different techniques for speech synthesis?
3. Name the signal segment that required for synthesized your first name using part name based Concatenative synthesizer. [8+3+5]
4. **(a)** Draw the LPC decoder and encoder block diagram [5]
5. A causal LTI system has system function is given in equation-1. Equation 2 represents the expression of prediction error filter. Lattice Formulations of Linear Prediction as given in equation 3(a) and 3(b) and draw the signal flow diagram of error filter ***A(z).***

Where *e[m]* represents the forward prediction error, *b[m]* represents the backward prediction error and *ki* is the PARCOR coefficient [5]

1. If the signal ***s[n] = {1,-1, 2,-2}*** applied in the above design error filter ***A(z)*** calculate the value of the forward prediction error at the output of the third lattice. [6]





Annexure-1

Name: ……………………………………………………………………….. Roll No……………………….



(VI)

(V)

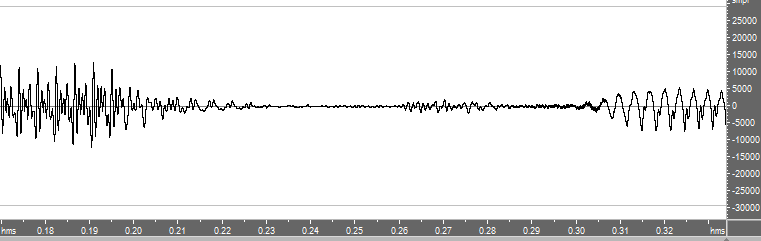
(IV)

(III)

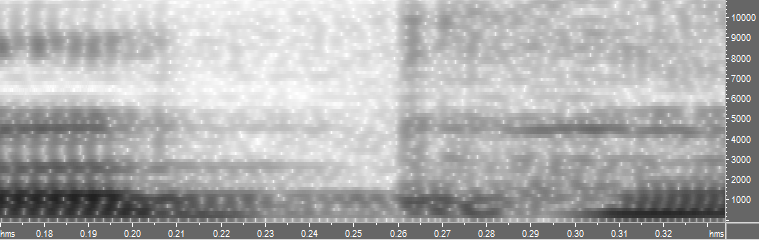
(II)

(I)

Figure-2



(a)



(b)

Figure-3