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

PART-A

1. The frequency response of a uniform tube is as given in the following equation (1). The length of the tube ***l=17.5*** cm and speed of sound ***c=350m/s***. Draw the volume velocity vs. Frequency curve for first 3 root.

$$\frac{U(l,\Omega)}{U_g(\Omega)} = V_a(\Omega) = \frac{1}{\cos(\Omega l / c)} \quad .....(1)$$

2. An audio signal is recorded using the following format.
F_S = 16 kHz, encoded with ***16 bit*** and recorded in **MONO**. To store ***2 sec*** signal in PCM WAV format calculate the memory requirement for store the signal?

3. ***3 kHz*** sinusoid signal is sampled at ***10 kHz*** determine the number of zero crossing in ***30 ms*** segment

4. A signal is sampled at ***16 KHz, 16 bit***, encoded with ***16th order LPC***. Each of the LPC coefficients is encoded with ***2 byte, Gain in 2 byte***. Voiced unvoiced F_0 information is encoded using ***1 byte***. Calculate the compression ratio if frame rate is ***100 frame /sec***?

5. Figure-1 represent the LPC Spectrum of a speech segment determine the order of the LPC analysis. If 2 poles are used for radiation and 2 poles are used glottal pulse modeling

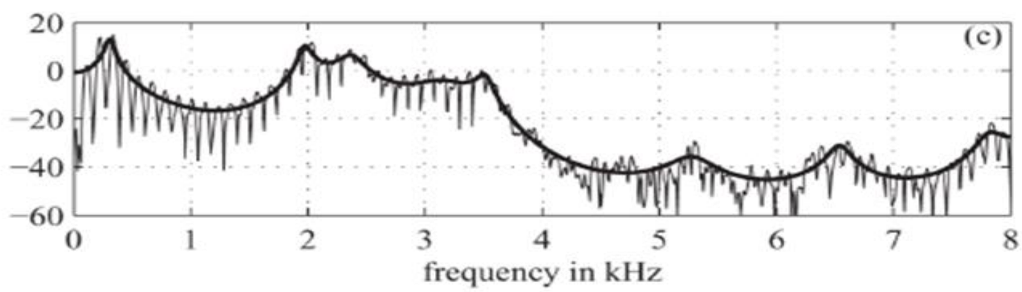

Figure-1

6. STFT analysis of a speech segment is required for noise reduction if the STFT analysis is done based on the Hamming Window of length ***20 ms*** determine the maximum possible temporal decimation factor so that signal is completely invertible . Where sampling frequency ***F_S=16 kHz***

7. Write the manner of articulation of the phonemes **/g/, /t^h/.**

8. Uniform Filter Banks analysis is used to extract the parameters of a speech segment, if the bandwidth of each filter is ***100 Hz*** and speech signal is recorded with sampling frequency ***16 KHz*** determine the required number filter to cover the entire spectrum of the speech segment

9. Write two advantages of having two Ears for sound perception

10. Draw the Schematic representation of the physiological mechanism of speech production

PART-B

1. A speech signal frame has energy $E_n^0 = 3000$using the autocorrelation method the frame is analyzed and 3 PARCOR coefficients $k_1 = 0.5; \quad k_2 = -0.2; \quad k_3 = 0.3$ are extracted. **[5+7+4]**

(a) Determine the energy of the liner prediction residual that would obtain by inverse filtering the speech signal frame. The inverse filter is designed using the above ***3 PARCOR*** coefficients.

(b) If the same speech signal segment is generated using lossless tube modeling and above 3 PARCOR coefficients are used to estimate the vocal tract cross-section area, calculate the value of the cross-section areas of the connected tubes. [Where initial tube cross-section area = ***0.75 cm²***]

(c) Figure-2 represent plot of the Normalized cross correlation Coefficients of speech segment. If the ***L= 120*** sample determine the F_0 of the speech segment. Where sampling frequency ***Fs=16*** KHz

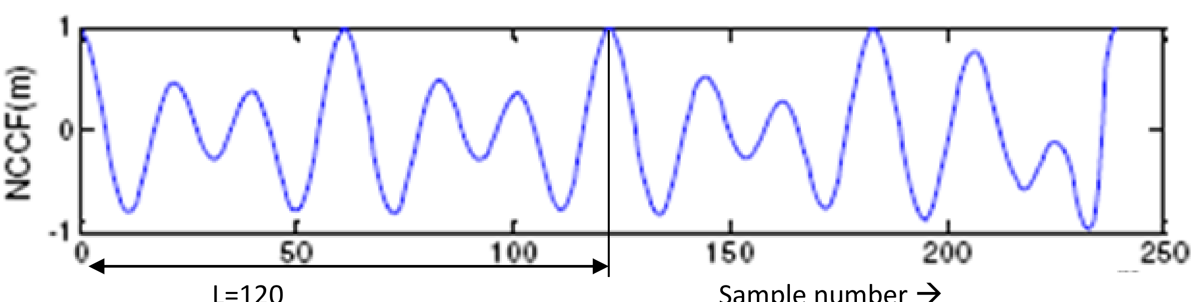

Figure-2

2. 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) **[6+10]**

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

$$H(z) = \frac{A}{1 - \sum_{k=1}^{p} \alpha_k z^{-k}} \quad (1) \qquad A(z) = 1 - \sum_{k=1}^{p} \alpha_k z^{-k} \quad (2)$$

$$e^i[m] = e^{i-1}[m] - k_i b^{i-1}[m-1] \quad (3a) \qquad b^i[m] = b^{i-1}[m-1] - k_i e^{i-1}[m] \quad 3(b)$$

(a) Draw the signal flow diagram of the Filter ***H(z).***

(b) If the signal ***s[n] = {1,0, 1,-1}*** applied in the design error filter ***A(z)*** (as in question no.) calculate the value of the forward prediction error at the output of the third lattice.

Where

$$k_i^{\text{PARCOR}} = \frac{\sum_{m=0}^{L-1+i} e^{i-1}[m]b^{i-1}[m-1]}{\left(\sum_{m=0}^{L-1+i}[e^{i-1}[m]]^2 \sum_{m=0}^{L-1+i}[b^{i-1}[m-1]]^2\right)^{1/2}}$$

3. What are the time domain methods for ***F_0*** extraction? 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. **[4+4+8]**

4. (a) Short-Time Fourier Transform Magnitude $|S(nL,\omega)|$ is compute for a speech signal segment with time decimation rate ***L=128*** sample. If the signal is recover with modify decimation rate ***M=32*** sample. Determine the speed-up of factor. **[4]**

(b) A signal $X_n[k]$ is the STFT of a signal $x_n[n]$ if the length of the DFT used is ***1024*** determine the frequency resolution. Where sampling frequency ***F_s=10 kHz*** **[4]**

(c) Complex cepstrum $\hat{x}(n)$ of a digital signal *x[n]* is the inverse Fourier transform of the complex log spectrum. $\hat{X}(e^{j\omega}) = \log\left|X(e^{j\omega})\right| + j\arg[X(e^{j\omega})]$ **[8]**

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.

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

5. (a) Draw the functional block diagram of Cepstral Coefficients (CC) extraction from a speech signal including the basic signal processing block. **[4]**

(b) Removal of unwanted components can be attempted in the cepstral domain. What is the name of these kinds of technique? **[2]**

(c) MFCC features are extracted from a recorded speech signal of ***2.5 seconds*** with the sampling frequency ***16 KHz***. If the length of the window is ***25 mile seconds*** and frame rate is ***100 frame/sec***. How may frames of MFCC features can be extracted from the above recorded speech signal? **[6]**

(d) Write ***2*** advantages of the delta and double delta MFCC features for speech signal classification **[4]**