Seat No.: ________ Enrolment No.______________

GUJARAT TECHNOLOGICAL UNIVERSITY

M.E –IIst SEMESTER–EXAMINATION – JULY- 2012

Subject code: 1724105 **Date: 12/07/2012**

Subject Name: Speech Signal Processing and Applications

Time: 10:30 am – 13:00 pm **Total Marks: 70**

Instructions:

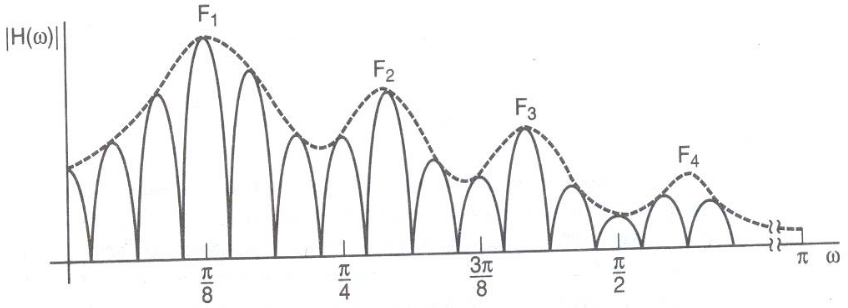
1. **Attempt all questions.**
2. **Make suitable assumptions wherever necessary.**
3. **Figures to the right indicate full marks.**

Q.1 **(a)** List the three major organs in speech production process and briefly describe their functions. **07**

(b) Explain the difference between nasal consonants and fricative consonants in terms of how they are produced and their acoustic characteristics. **07**

Q.2 **(a)** Following figure represents the magnitude of the discrete-time Fourier transform of a steady-state vowel segment which has been extracted using a rectangular window. The envelope of the spectral magnitude is sketched with a dashed line. Note that four formants are assumed and that only the main lobe of the window Fourier transform is depicted. **07**

1. Suppose that the sampling rate is 12000 samples/s, set to meet the Nyquist rate. What is the first formant frequency (F1) in Hz? How long is rectangular window in milliseconds? How long is the window in time samples?
2. Suppose F1 = 600 Hz and you are not given the sampling rate. What is the pitch in Hertz? What is the pitch period in milliseconds? What is the pitch period in time samples?



(b) What is a spectrogram? Compare wideband spectrogram and narrowband spectrogram. **07**

OR

(b) Let $x[n]$ be a signal with a single sinusoid component. The signal $x[n]$ is windowed with an L-point Hamming widow $w[n]$ to obtain $v_1[n]$ before computing $V_1(e^{j\omega})$. The signal $x[n]$ is then windowed with an L-point rectangular window to obtain $v_2[n]$, which is used to compute $V_2(e^{j\omega})$. Will the peaks in $V_2(e^{j\omega})$ and $V_1(e^{j\omega})$ have the same height? If so, justify your answer. If not, which should have a larger peak? **07**

Q.3 **(a)** Consider the periodic impulse train: $x[m] = \sum_{r=-\infty}^{\infty} \delta[m - rP]$ **07**

Using the definition of the short-time autocorrelation as:

$$R_n[k] = \sum_{r=-\infty}^{\infty} [x[n+m]w'[m]][x[n+m+k]w'[k+m]]$$

with $w'[m]$ a rectangular window whose length, N, satisfies:

$QP < N - 1 < (Q+1)P$, where Q is an integer.

Find and sketch $R_n[k]$ for $0 \le k \le N-1$. How would the result change if the window is a Hamming window of the same length?

(b) Write expressions of discrete-time STFT and discrete STFT. **07**

Using these show that $X(n, \omega_0) = e^{-j\omega_0 n}\left(x[n] * w[n]e^{j\omega_0 n}\right)$

Also show that discrete STFT can be expressed as the outputs of a set of analysis filters.

OR

Q.3 (a) Define short-time energy and short-time average zero crossing rate. Write an algorithm to classify speech signal into voiced and unvoiced parts using these two time-domain parameters. **07**

(b) Consider the periodic impulse train $x[n]=\sum_{l=-\infty}^{\infty}\delta[n-lP]$ and analysis window $w[n]$ a triangle of length *P*. Compute its discrete-time STFT and discrete STFT with frequency sampling factor *N*=*P*. Compare both the results. **07**

Q.4 (a) Consider the first order predictor: $\hat{x}[n]=\alpha x[n-1]$. Find α using least square error minimization. Also find $E_{\min}$. **07**

(b) In the short-term LP analysis of speech, comment on the differences in the waveforms as well as magnitude spectra of the LPC residual obtained when the input is a vowel, unvoiced fricative or nasal sound. **07**

OR

Q.4 (a) Draw the block diagram of simplified model of speech production. Explain the concept of linear predictive analysis and derive the expression of prediction error filter and its transfer function. Why is it called an inverse filter? **07**

Q.4 (b) Consider following FIR filter which is the most common form of pre-emphasis: **07**

$$y[n]=x[n]-\alpha x[n-1]$$

where $x[n]$ is the input speech signal and $y[n]$ is the output pre-emphasized speech and α is an adjustable parameter.

1. Determine analytical expressions for the impulse response and system function of this filter.
2. What type (high-pass or low-pass) of filter is it?
3. How can pre-emphasis help speech analysis?

Q.5 (a) Compare linear prediction and homomorphic filtering methods. **07**

(b) Compute the complex cepstrum of $X(z)=\dfrac{1}{1-az^{-1}}$. Assume $|a|<1$. **07**

OR

Q.5 (a) Consider an all-pole model of the vocal tract transfer function of the form **07**

$$V(z)=\frac{1}{\prod_{k=1}^{q}\left(1-c_k z^{-1}\right)\left(1-c_k^{*} z^{-1}\right)}$$

where $c_k=r_k e^{j\theta_k}$.

Show that the corresponding cepstrum is

$$\hat{v}(n)=2\sum_{k=1}^{q}\frac{(r_k)^n}{n}\cos(\theta_k n)$$

(b) Consider a sequence $x(n)$ with complex cepstrum $\hat{x}(n)$. The z-transform of $\hat{x}(n)$ is **07**

$$\hat{X}(z)=\log[X(z)]=\sum_{m=-\infty}^{\infty}\hat{x}(m)z^{-m}$$

where $X(z)$ is the z-transform of $x(n)$. The z-transform $\hat{X}(z)$ is sampled at N equally spaced points on the unit circle, to obtain

$$\hat{X}_p(k)=\hat{X}\left(e^{j\frac{2\pi}{N}k}\right)\quad 0\le k\le N-1$$

Using the inverse DFT, we compute

$$\hat{x}_p(n)=\frac{1}{N}\sum_{k=0}^{N-1}\hat{X}_p(k)e^{j\frac{2\pi}{N}kn}\quad 0\le n\le N-1$$

which serves as an approximation to the complex cepstrum. Express $\hat{X}_p(k)$ in terms of the true complex cepstrum $\hat{x}(m)$. Also show that $\hat{x}_p(n)=\sum_{r=-\infty}^{\infty}\hat{x}(n+rN)$.

****