Assianment-6

1. Which of the following are the natural features of speech signal?

A. Cepstrum, linear predictive coefficients

B. Place and manner of articulation

C. Vocal tract area function

D. MFCC and delta MFCC

2. 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 8^{th} filter

A. 150 Hz

B. 300 Hz

C. 100 Hz

D. 50 Hz

[Ans. The critical band is a band pass filter, adjusted around the center frequency. Below 1 kHz critical bands are placed linear around 100, 200, ... 1000 Hz. Above 1 kHz these bands are placed with the mel-scale.
So 8^{th} filter bandwidth will be 100 Hz]

3. Complex cepstrum $\hat{s}(n)$ is the inverse Fourier transforms of the complex log spectrum of a speech segment s*[n]*. Real Cepstrum *c[n]* defines as the inverse Fourier transform of the log magnitude of the speech segment s*[n]*. *The relation between c[n] and* $\hat{s}(n)$ can be define as

A. $c[n] = \frac{\hat{s}[n] + \hat{s}[-n]}{2}$

B. $c[n] = \frac{\hat{s}[n] - \hat{s}[-n]}{2}$

C. $c[n] = \frac{\hat{s}[-n] + \hat{s}[-n]}{2}$

4. Homomorphic systems can be expressed as a cascade of three homomorphic sub-systems. Figure-1 represents the Homomorphic Systems for Convolution

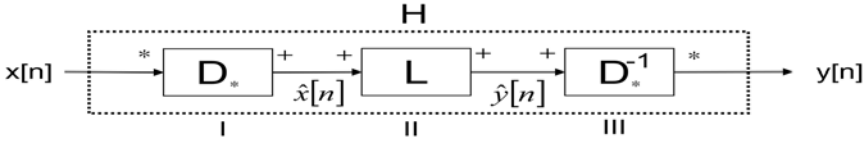
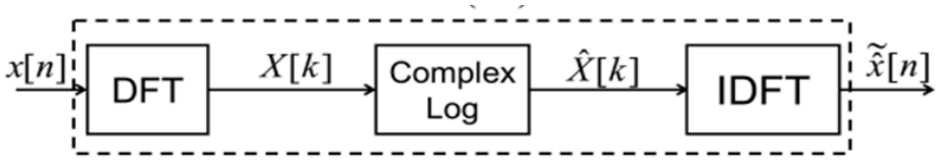
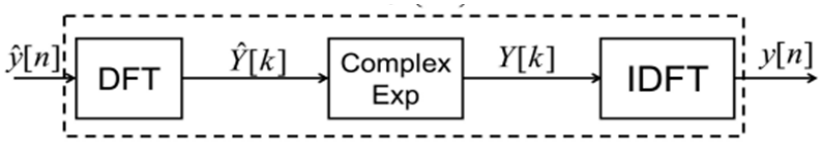


Figure-1

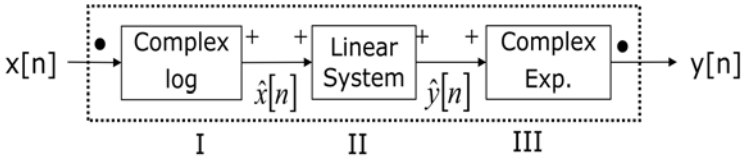
Which one of the following *figure represents* the first homomorphic sub-systems (I) (D_*)?



A



B



C

5. Removal of unwanted components can be attempted in the cepstral domain. This kinds of technique is called

A. Filtering

B. Quefrency

C. Liftering

6. Which one of the following is the time domain parameter of speech signal?

A. Delta Mel Frequency Cepstral Coefficients

B. Cepstral Transfer Coefficient (CC)

C. Formant Parameters

D. Linear Prediction Coefficient

7. 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

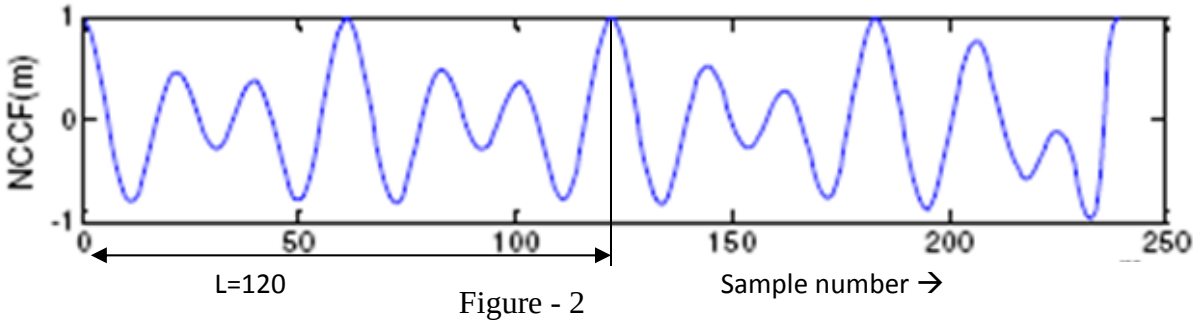


Figure - 2

A. 266 Hz

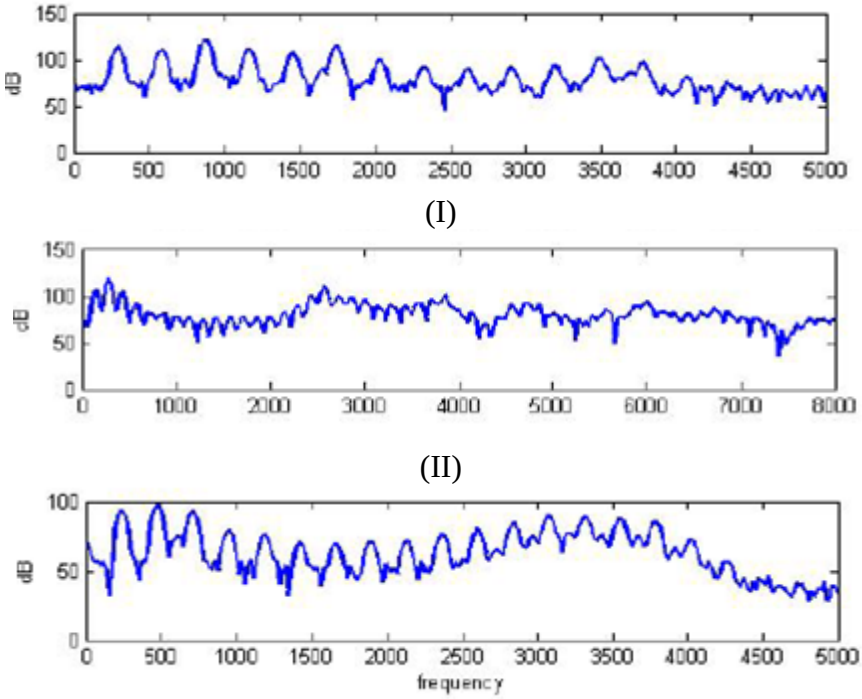
B. 120 Hz

C. 60 Hz

D. 150 Hz

[$T_0 = 120/2 = 60$ samples; $f_0 = 16\times10^3/60 = 266$ Hz]

8. Figure 3 shows plots of 3 voiced speech segment short-time log magnitude spectra as obtained using a Hanning window of an appropriate length. Which of the voiced speech spectra most likely come from an adult male?



(I)

(II)

(III)

A. (III)

B. (I)

C. (II)

9. Equation 1 is used to compute the average magnitude difference between the signal and a time-shifted version of itself. Then F_0 value of the signal is detected based on

$$D_x[k] = \frac{1}{N} \sum_{n=0}^{N-1-k} \left| x(k) - x(n+k) \right|, \quad (1) \qquad 0 \le k \le K_0$$

A. Valleys in the average magnitude difference function

B. Peaks in the average magnitude difference function

C flat portion in the average magnitude difference function

10. Uniform Filter Banks analysis is used to extract the parameters of a speech segment, if the bandwidth of each filter is 100Hz and speech signal is recorded with sampling frequency 12 KHz determine the required number filter to cover the entire spectrum of the speech segment

A. 30

B. 60

C. 120

D. 100

[No. of filter = Fs/2×100 = $12\times10^3/200$ = 60 Nos.]