This exam consists of four questions. Please show all steps leading to your answers. Just an answer without the explanation will not receive full credit.

1. (10 points)
In Figure 1(a) and 1(b), the short time energy and the short time zero crossing rates of a speech are plotted, respectively.
(a) For what purposes would you use these plots?
(b) Show the usage by marking on the plots. Give explanations for your answers

Answer:
1(a): Short-time energy and zero crossing rate can be used for the classification of voiced sounds and unvoiced sounds because the short-time energy is high for voiced sounds and low for unvoiced sounds where as the zero crossing rate is low for voiced
sounds and high for unvoiced sounds. In addition the short-time energy can be used for the detection of word ending or end point detection.

1(b): see the plots above

2. (15 points)

\[ x(0) = 6, \ x(1) = 3, \ x(2) = -2 \text{ and } x(3) = 2 \]

(a) If you are asked to estimate \( H(z) \) using the autocorrelation method what is the maximum order of poles of \( H(z) \) for the above given \( x(n) \) values?
(b) Estimate \( H(z) \) for one order less than the maximum
(c) If the input \( x(n) = \delta(n) \) what is the output \( y(n) \)?

Answers:

2(a): For the given \( x(n) \) values, the autocorrelation of lag up to 3 i.e., \( R(3) \) can be calculated. Hence, the maximum order of poles of \( H(z) \) is three.

2(b): The order poles of \( H(z) \) one less than the maximum is 2. That means we need to estimate \( a_1, a_2 \) and \( G \). The estimation of these values are as follows:

\[
\begin{align*}
R(0) &= 6^2+3^2+(-2)^2+2^2 = 53 \\
R(1) &= 6\times3+3\times-3+-2\times2 = 8 \\
R(2) &= 3\times-2+-2\times2 = -10
\end{align*}
\]
\[
\begin{bmatrix}
R(0) & R(1) \\
R(1) & R(0)
\end{bmatrix}
\begin{bmatrix}
a_1 \\
a_2
\end{bmatrix}
= 
\begin{bmatrix}
R(1) \\
R(2)
\end{bmatrix}
\]  
substituting for R’s we get two sets of equations with two unknowns. These two set of linear equations can be solved for \(a_1, a_2\). The estimated values that are obtained after solving the two equations are:

\[
a_1 = \frac{(53 \times 8) - (8 \times 10)}{53^2 - 8^2} = 0.1253 \quad \text{and} \quad a_2 = \frac{(8 \times 8) - (53 \times -10)}{8^2 - 53^2} = -0.2164.
\]

\[
G = \sqrt{E_{\text{min}}} = \sqrt{R(0) - a_1 R(1) - a_2 R(2)} = \sqrt{53 - 0.1253 \times 8 - (-0.2164) \times (-10)} = 7.0603
\]

2(c) For the given \(x(n), y(n)\) = impulse response of the LTI system. That is \(Y(z) = H(z)\) in the Z-domain. \(y(n)\) can be obtained by taking the inverse ZT which is as shown below.

\[
Y(z) = H(z) = \frac{G}{1 - \sum_{k=1}^{p} a_k z^{-k}}
\]

\[
= \frac{7.0603}{1 - 0.1253 z^{-1} + 0.2164 z^{-2}} \text{ using the estimated values in 2(b)}
\]

\[
Y(z) \times \left(1 - 0.1253 z^{-1} + 0.2164 z^{-2}\right) = 7.0603
\]

By applying the inverse ZT for the above equation we get:

\[
y(n) = 7.0603 + 0.1253 y(n-1) - 0.2164 y(n-2)
\]

3. (15 points)
The windowed DFT is computed using a rectangular window and a Hamming window and they are plotted for a speech sound in Figures 2(a) and 2(b) below.

(a) Which window is used in Figure 2(a) and 2(b)? Give explanations for your answer.

(b) Sampling frequency \(F_s = 8000\ \text{Hz}\), estimate \(F_1\) and \(F_0\) for the spectrum plotted in Figures 2(a) and 2(b)

(c) Identify the gender of the speaker of Figure 2(a) and 2(b)

(d) Calculate the window length used in Figure 2(a) and 2(b)
Answer:
3(a) The type of window used in figure 2 (a) is rectangular because the peaks of each main lobe is sharper than in figure 2(b) and the effect of side lobe is predominant which can be seen clearly as compared to figure 2(b) with the spread of pulses within the formant frequencies. The type of window in figure 2(b) is Hamming because the side lobes are much below the main lobe (so the effect of side lobe is less) and the main lobe peaks are not as sharp as in Figure 2(a).

comments: This question is almost impossible to make it right. So as long as your answer has major characteristics of these two windows, you are credited. The most important property, which is the side-lobe comparison of these windows, however, has to be specified to get full credit.

3(b) In Figure 2 (a) $F_1$ corresponds to approximately 1100 Hz as indicated with a black line from the first peak in the plot. The first peak is identified with a red arrow.
There are four pulses or four main lobes within the first formant. So $F_i = 4F_0$.

Therefore $F_0 = \frac{F_i}{4} = \frac{1100}{4} = 275Hz$. In Figure 2 (b) $F_1$ corresponds to approximately 750 Hz as indicated with a black line from the first peak in the plot. The first peak is identified with a red arrow. There are six pulses or six main lobes within the first formant. So $F_i = 6F_0$. Therefore $F_0 = \frac{F_i}{6} = \frac{750}{6} = 125Hz$.

3 (c) Since the pitch frequency of figure 2 (a) is 275 Hz which is in the range of pitch frequency of female, therefore the gender of figure 2 (a) is female. The pitch frequency of figure 2 (b) is within the range of male, therefore the gender of figure 2 (b) is male.

3 (d) For figure 2(a)

$$F_1 = \frac{1100}{4000} \pi \text{ in radians.}$$

$F_1 = 4F_0$ Therefore $F_0$ in radians is : $F_0 = \frac{11}{40 \times 4} \pi = \frac{4\pi}{N}$ because a rectangular window is used. $N = \frac{4 \times 40 \times 4}{11} = 58.18$. Which is approximately equal to 64 samples. 

For DFT computation window length used will be in powers of two for efficient computation. Since the sampling frequency is 8000, 64 samples corresponds to 64/8000 = 0.0080 seconds

For figure 2(b)

$$F_1 = \frac{750}{4000} \pi \text{ in radians.}$$

$F_1 = 6F_0$ Therefore $F_0$ in radians is : $F_0 = \frac{750}{4000 \times 6} \pi = \frac{8\pi}{N}$ because a Hamming window is used. $N = \frac{4000 \times 6 \times 8}{750} = 256$ samples.

For DFT computation window length used will be in powers of two for efficient computation. Since the sampling frequency is 8000, 256 samples corresponds to 256/8000 = 0.032 seconds

comment: (d) is not graded, since the correct distinguishing of different windows is vague.

4. (10 points)
Two spectrograms of the same speech signal are plotted in Figures 3(a) and 3(b).
(a) Identify the type of spectrogram in 3(a) & 3(b). What causes them to look different even though they are for the same speech signal? Give explanations.

Answer: The spectrogram types of figure 3(a) and 3(b) are: Narrowband and wideband, respectively. They are obtained by using longer and shorter window length in the time domain, respectively. Due to the time – frequency resolution trade-off the Short-time FT these two plots look differently even though they are for the same speech signal. In the case of plot 3(a), the longer window length is used which results in better frequency resolution which in turn helps in resolving harmonics. The harmonics corresponds to horizontal striations which in turn corresponds to pitch frequency. In the case of plot 3(b) shorter window length is used which results in better time resolution which in turn helps in resolving the pitch periods which appear as vertical striations.