CCE5223_ Speech Processing and Coding

Answer ANY THREE questions

1   (a) Distinguish between an allophone and a phoneme.   (4 marks)
     (b) For the Maltese word ' qaleb', work out the phonemes and in each case
     mention the type of phoneme.   (6 marks)
     (c)What is a speech spectrogram? Distinguish between a wideband and a
     narrowband spectrogram   (5 marks)
   (d) Mention THREE time waveform acoustic parameters that are used in
        classifying and distinguishing phonemes. For each type mention the type of
        phonemes for which they are suitable to use.   (6 marks)
   (e) Consider the phoneme sequence /ame/
     (i)   Describe how the formants below 3kHz would behave on a wideband
              spectrogram of this utterance   (4 marks)
     (ii)   How would the formant bandwidths change during the utterance?
                                                                  (4 marks)
     (iii)   When and where do spectral zeroes appear in this utterance?
                                                                  (4 marks)

2   (a) Consider a pitch detection scheme that lowpass filters a piece of male
     speech to 900Hz and then calculates an autocorrelation function.
              (i)   Why is the speech first lowpass filtered?   (4 marks)
              (ii)   How is the autocorrelation function used to generate a pitch
                       estimate.   (6 marks)
              (iii)   For a sampling rate of 16,000 samples per minute suggest, with
                       reasons, a suitable range of values of displacement τ in samples
                       to obtain the pitch value   (5 marks)

     (b)   (i) Describe the physiological mechanisms that contribute to speech
     production and hence relate the mechanism to a discrete-time speech
     production model.   (8 marks)
              (ii) What are the approximations of the model to the real speech
     mechanism?   (4 marks)
     (c)   What is the Mel Frequency Cepstral Coefficient, (MFCC). Describe,
              with diagrams a practical way of obtaining these coefficients for use in
              speech recognition?   (6 marks)

3.   (a) In linear predictive coding, codebooks are used.
              (i)   Distinguish between an excitation codebook and an error
                       codebook.   (4 marks)
              (ii)   Outline briefly the mechanism for the generation of a
                       codebook.   (6 marks)
              (iii)   Describe the way that error codebook information is used
                       within a speech transmission system.   (4 marks)

(b) For a linear predictive system, that predicts the current input x(n) from a set of p previous inputs x(n-1) to x(n-p), the error is given by

$$e(n) = x(n) - \hat{x}(n)$$

$$where \quad \hat{x}(n) = \sum_{i=1}^{p} a_i x(n-i)$$

is the predicted input.

Derive the set of Yule-Walker equations that obtain the values of the a_i ,(i=1 to p), to minimize the error e(n)

(12 marks)

(c)   Describe the multiband excitation vocoder and compare it to the Code Excited LPC (CELP) vocoder givng typical values of the bits/second required in each case and the quality of the transmitted speech. State the assumptions being taken.   (7 marks)

4   (a) (i) Describe TWO methods that are used to enhance the speech, and reduce the noise, from speech that has noise added to it.

(6 marks)

(ii) Can these techniques be used to filter out, (distinguish) a particular voice from speech that has more than one speaker. Give reasons.   (5 marks)

(b)   Define the following terms with respect to a Hidden Markov Model
          (i)   The state probability matrix
          (ii)   The output probability matrix
          (iii)   The Baum-Welch algorithm

(6 marks)

(c)   A simple discrete output system uses 8 centroids. The HMM also uses
          4 states.
     The state transition matrix, A, is given by

0.6	0.4	0.0	0.0
0.0	0.6	0.4	0.0
0.0	0.0	0.6	0.4
0.0	0.0	0.0	1.0

The probability matrix, B, governing an output when entering a state is given by

0.125	0.125	0	0.2
0.125	0.125	0	0.1
0.125	0.125	0.25	0
0.125	0.125	0	0.1
0.125	0.125	0.25	0.2
0.125	0.125	0.25	0.1
0.125	0.125	0	0.1
0.125	0.125	0.25	0.2

where the entry bij is for the ith row and j^{th} column, i, from 0 to 7, j from 0 to 3.

(i)   The clip presented to an HMM has 6 frames. Assume that the
       movement through the HMM is as shown below

STATE   0   0   1   1   2   2   3

OUTPUT   1   4   3   5   2   4   6

What is the value of the forward probability at the end?

(12 marks)

(ii) If in going from state 2 to state 3, the output is 6, what will the value of the
forward probability be? What does this mean?   (4 marks)