CCE5223_ Speech Processing and Coding

June 2012

Answer ANY THREE questions

1 (a) Figure 1 shows an acoustic time waveform of a word in Maltese.

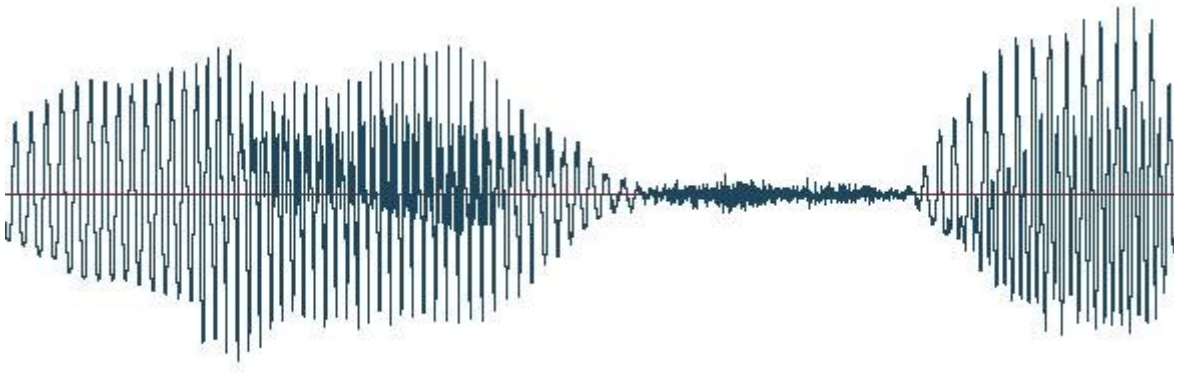

Figure 1

Which one of the following words is it:

(i)mela (ii) ilma (iii) nafu (iv) kesa

Give reasons for your answer. (6 marks)

(b) (i) Distinguish between a labial fricative and an alveolar fricative. (5 marks)

(ii) Do they exhibit the same average energy when used between two vowels? Give reasons for your answer. (3 marks)

(c) For the phonemes /a/, /s/, /g/ and an equal time span (ie same number of samples) rate the three phonemes starting from the highest to the lowest in terms of:

(i) energy;(ii) zerocrossings; (iii) peak frequency of first formant

Give reasons for your choice.

(9 marks)

(d) (i) What is pitch in voiced speech? (3 marks)

(ii) Describe the physiological mechanism that produces voicing in speech? (4 marks)

(iii) How does a real glottal pulse differ from the model of a sample at a pitch period? (3 marks)

2 (a) Describe a method to obtain:
(i) a voiced/unvoiced decision for a speech segment (6 marks)
(ii) end-point detection (6 marks)

(b) (i) Outline the procedure to obtain a set of mel frequency cepstral coefficients from a frame of speech.. (6 marks)
(ii)Why are lifters used in obtaining cepstral coefficients for speech recognition. (4 marks)

(c)Consider a vowel with formants at 500, 1500, 2500 Hz lowpass filtered to 4000 Hz and then sampled at the Nyquist rate.
(i) Draw a detailed block diagram of a system to generate a good version of this already sampled signal at 12,000 samples per second. (5 marks)
(ii) For the system sampled at 12,000 samples per second, and within the range $|\omega| < \pi$, where do the three formants appear in terms of digital frequency range? (6 marks)

3. (a) (i) Describe a DPCM based speech waveform model. (5 marks)
(ii) Why are less bits/frame required for such a system than for PCM.at the same quality (3 marks)
(iii) How can the bits /frame used for DPCM be further reduced but still maintaining same quality. (4 marks)
(iv) What is the major problem in using a differential system when coding speech? (3 marks)

(b) Two improvements to the basic LPC model are
(i) a perceptual weighting filter
(ii) a long term predictor filter
Describe the improvements obtained to the basic LPC model when they are used. (8 marks)

(c) (i) What is the residual error in LPC? (4 marks)
(ii) For what type of speech is the residual error large? Give reasons (3 marks)
(iii) When the error signal is transmitted with limited bandwidth, which frequency range is sent? (3 marks)

4 (a) Describe briefly the 'top-down' and 'bottom-up' models for speech recognition using an HMM phoneme model and a word language model, highlighting the advantages and disadvantages of each type. (6 marks)

(b) An HMM is specified by A, B and JI where A is a state transition matrix, B is a discrete observation probability matrix, and JI is the initial vector. The state transition matrix elements are a_{ij} denoting the probability of being in state i and j is the next state. The elements of B are b_{kj} denoting the probability of the observation k in state j. There are 4 states from 1 to 4, and six discrete observations from 1 to 6.

The matrix A is given by and matrix B is given by

Matrix A

0.6	0.3	0.1	0.0
0.0	0.4	0.4	0.2
0.0	0.0	0.5	0.5
0.0	0.0	0.2	0.8

Matrix B

0.2	0.3	0.4	0.25
0.2	0.2	0.0	0.25
0.2	0.0	0.0	0.1
0.1	0.3	0.2	0.1
0.2	0.1	0.2	0.1
0.1	0.1	0.2	0.2

The initial state Π is state 1, with observation 1. The system has the following series of observations:

1 3 4 2

Work out the most probable state sequence for this series of observations.

(12 marks)

(c) Figure 2 is a language state assignment model for a set of words used in speech recognition. The system receives the following sequence of probable words from the

Language State Diagram

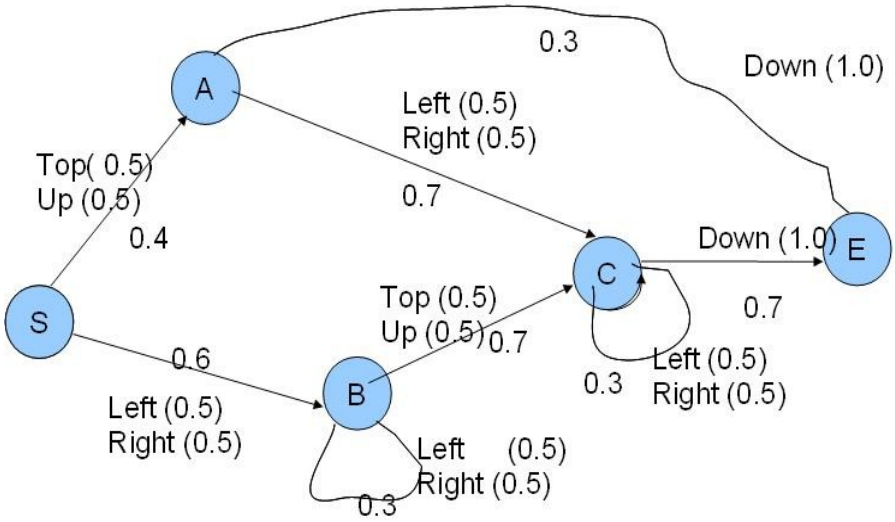

Figure 2

HMM level:

left *right* *top* *down*

Calculate the probability of this word sequence from the LSD of Figure 2.

(10 marks)

(d) Why are several Gaussian mixtures used when developing an HMM phoneme model using continuous probability density functions for the acoustic paramters.?

(5 marks)