EE6424 Digital Audio Signal Processing (Part II) Assignment

Submission guideline:

1. The weightage of this assignment is 30%. There are 4 questions in total.
2. The answer script should be typed (preferred) or written clearly (no pencil please), with your name and matriculation number on the first page.
3. Keep your assignment report within 16 pages including all figures.
4. Please submit your assignment report (hard copy only) to me in my office (S2-B2a-14) by 6:30pm on Thursday of week 15 (30 April 2020).
5. Please keep your own copy of the assignment report since your submission will not be returned.
6. Please be informed that any form of cheating will be severely penalised.

Question 1

Consider the design of a high quality digital speech processing system. It is required that the SQNR is at least 80 dB with a loading factor of 4. The useful signal bandwidth must be at least 8 k Hz and the signal is assumed to contain some higher frequency components.

i. Draw a block diagram of the basic components needed for A/D and D/A conversion. How many bits are required in the A/D and D/A converters?

ii. Assume that an analog Butterworth filter is used as the anti-aliasing filter. For a sampling frequency of F_s , at least 40 dB attenuation is required at $F = F_s/2$. Design the analog Butterworth filter with an appropriate order and also the sampling rate you will use. Explain your design briefly.

iii. If the objective of the digital speech system is now just to maintain a telephone quality representation of speech, i.e., useful signal bandwidth up to 3.5 k Hz only and SQNR being at least 60 dB with a loading factor of 4, what is the number of bits required? Assume that the sampling rate is 8 k Hz. What is the order of the analog Butterworth filter you will use as the anti-aliasing filter?

(30 marks)

Question 2

The Linde-Buzo-Gray (LBG) algorithm, also known as k-means algorithm, is commonly used to find the optimal vector quantization (VQ) codebook. For this question, the distance of two $N \times 1$ vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ is defined as:

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sum_{n=1}^{N} (x_n - y_n)^2$$

where x_n and y_n are the n-th elements of $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively.

i. Referring to the lecture notes, using Matlab to implement a simplified version of the LBG algorithm (Step 1 to Step 3) for this task:

The following training vectors (2 element vector) are used to train a vector quantizer. Using an initial codebook of (-2, 2) and (-1, 2), find the updated codebook after ONE pass of the LBG algorithm:

(-2, 2) ,(-1, 2), (-1, 1), (0, 0) ,(3, 0), (1, -2)

Include the the updated codebook and the Matlab code in your assignment solution.

ii. Referring to the lecture notes, using Matlab to implement the second version of the LBG algorithm (now Step 1 to Step 4) for this second task:

The following new training data are adopted and the new initial codebook consists of these three vectors: (-2 ,2), (-1, 2), (12,-1). The new training data are:

(-2, 2), (-1, 2), (-1, 1), (0, 0), (3, 0), (1, -2), (0, 2), (4, 0),

(4 -1), (4 -3), (7, -3), (6, -5), (9, 1), (10, 5), (12, -1).

Run the second version of the LBG algorithm with five iterations. For each iteration, record the iteration number, the total quantization error Err, the updated codebook and put them clearly in a table. Include the Matlab code in your assignment solution.

iii. Referring to the lecture notes, using Matlab to implement the final version of the LBG algorithm (now Step 1 to Step 5) for this final task. The training data and the initial codebook are the same as in part ii. The threshold for the change in Err is given as 0.3 in Step 5 of the LBG algorithm. Run the final version of the LBG algorithm until it stops (or stops after five iterations). Record, for each iteration until it stops, the iteration number, the change in Err, the updated codebook and put them clearly in a table. Include the Matlab code in your assignment solution.

(30 Marks)

Question 3

Consider a speech signal given by the name of the file test.wav in NTULearn. Write a MATLAB file to obtain the spectrogram of the given speech signal. You may make good use of existing Matlab functions and you need to include the m.file as part of the assignment solution (in a single file submission). Your tasks include:

i. Determine the sampling rate for the signal in the test.wav file.

ii. Implement three window functions (one at a time): rectangular window; Hanning window; and Hamming window.

iii. Obtain both wideband and narrowband spectrogram (as images).

iv. Give the numbers of samples of the speech signal you use for wideband and narrowband spectrogram, respectively.

v. Briefly discuss what you have done in parts i to iv above.

(20 Marks)

Question 4

A bi-Gaussian model, as shown below, is used for Voice Activity Detection (VAD):

$$p(e)=\sum_{i=1}^{2} w_i \frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left(-\frac{(e-\mu_i)^2}{2\sigma_i^2}\right)$$
$$=0.3\frac{1}{\sqrt{200\pi}}\exp\left(-\frac{(e^2-140e+4900)}{200}\right)+0.7\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(e^2-68e+1156)}{2}\right)$$

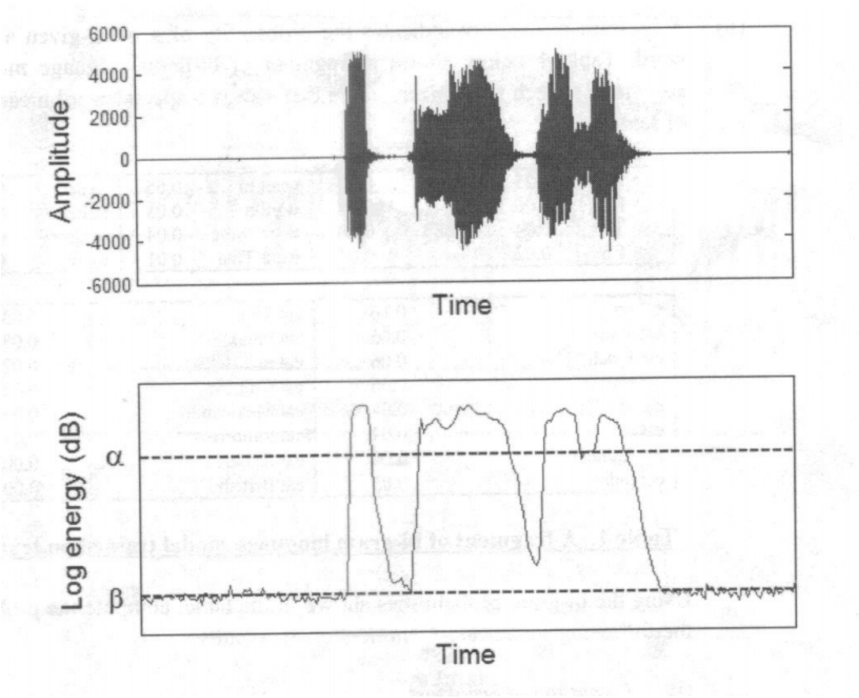
The model is fitted onto the log-energy e of 30-second frames of an input speech signal. The figure below shows the speech signal in the time domain (upper panel) and the corresponding log-energy contour (lower panel). The two dotted lines denote the mean values of the speech and non-speech log-energy. There are just two classes of acoustic events: speech and non-speech.

i. Determine the values for α and β as shown in the figure.

ii. Indicate the percentage of frames containing speech from the given bi-Gaussian model.

iii. Given a frame with log-energy e = 50, determine whether this frame is a speech or a non-speech frame. State your justification.

(20 Marks)



End of assignment