EE4-14 ✓
EE9-SO16 ✓

IMPERIAL COLLEGE LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2014

MSc and EEE/EIE PART IV: MEng and ACGI

Corrected Copy

SPEECH PROCESSING

Wednesday, 14 May 10:00 am

Time allowed: 3:00 hours

There are FOUR questions on this paper.

Answer ALL questions.

All questions carry equal marks

Any special instructions for invigilators and information for candidates are on page 1.

Examiners responsible First Marker(s) : P.A. Naylor
Second Marker(s) : W. Dai

1. a) A model of speech production is illustrated in Figure 1.1. Explain the meaning and give the definition of each of the signals, parameters and operations in this model. Discuss the rate at which the parameters of the model should be updated in a speech coding application. [4]

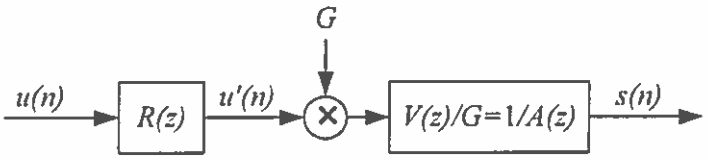

Figure 1.1 A model of speech production.

b) Consider a speech signal $s(n)$ over a frame of samples $\{F\}$ and the speech covariance matrix Φ with elements ϕ_{ij}. Consider also the prediction of $s(n)$ using LPC with predictor coefficients denoted a_i, for $i = 0, 1, 2, \ldots$.

i) Show how Φ and the elements ϕ_{ij} are formed in terms of $s(n)$. [2]

ii) Formulate an expression for the prediction error in terms of the prediction coefficients a_i. [2]

iii) Write down an expression for the sum squared prediction error over the frame of speech samples $\{F\}$. [1]

iv) Derive an expression in terms of $s(n)$ for the predictor coefficients that minimize the sum squared prediction error and show that this expression can be written in terms of ϕ_{ij}. [5]

v) How is $\{F\}$ chosen in the case of autocorrelation LPC? State any consequences of this choice on the computation of a_i. [2]

vi) Let the normalized power spectrum of the prediction error signal $e(n)$ be defined as

$$P_E(e^{j\omega}) = \frac{|E(e^{j\omega})|^2}{Q_E}$$

$$Q_E = \frac{1}{2\pi}\int_{\omega=0}^{2\pi} |E(e^{j\omega})|^2 d\omega$$

where $E(z)$ is the z-transform of $e(n)$.

Next let the spectral roughness be defined

$$R_E = \frac{1}{2\pi}\int_{\omega=0}^{2\pi} \left(P_E(e^{j\omega}) - 1 - \log\left(P_E(e^{j\omega})\right)\right) d\omega.$$

By considering $E(z) = A(z)S(z)$, show that minimizing Q_E is equivalent to minimizing the spectral roughness of the prediction error. [4]

2. a) Consider a signal being quantized using a quantization process employing a set of quantization bins. Each bin covers an amplitude range w and any particular amplitude is contained in exactly one bin. Now consider an input signal such that the distribution of signal amplitude in any bin is uniform over the amplitude range of the bin.

i) Briefly explain what is meant by 'one least significant bit' in this context. [1]

ii) For a quantization bin spanning the range $-w/2$ to $+w/2$, within which all input values are quantized to zero, derive an expression for the RMS quantization error in this bin in terms of the bin width. [3]

b) i) Consider a speech signal $s(n)$ at time index n with probability density function $p(s)$. Further consider a nonuniform quantizer such that the input signal amplitudes in the range $[a_{i-1}, a_i]$ are quantized to output amplitude values s_i. Find an appropriate expression for the quantized amplitudes s_i in terms of a_i, a_{i-1}, and $p(s)$ such that the quantization error is minimum in the mean square. [4]

ii) Give an example of a probability density function $p(s)$ for which nonuniform quantization would be preferable compared to uniform quantization and explain your reasoning. [2]

iii) A speech signal $s(n)$ is represented using PCM with a precision of 16 bits per sample. It is now intended to encode $s(n)$ using μ-law encoding in which σ, e and m denote the sign, exponent and mantissa bits respectively, and the quantization scheme has bin centres at

$$\pm\{(m+16.5)2^e-16.5\}.$$

For some particular $n=n_1$, it is found that $s(n_1)=-1793$. Determine the bit values used to μ-law encode $s(n_1)$ and state the amplitude of the error introduced by μ-law coding of this sample. [5]

c) The quantizer labelled Q in Figure 2.1 is a uniform 5-level quantizer with outputs such that

$$w(n) \in \{-2,-1,0,1,2\}.$$

The block labelled 'Update k' modifies the input value of k at time index n according to the following rule:

$$k(n+1) = \begin{cases} 3k(n) & \text{when} \quad w(n) = \pm 2 \\ 1.1k(n) & \text{when} \quad w(n) = \pm 1 \\ 0.9k(n) & \text{when} \quad w(n) = 0. \end{cases}$$

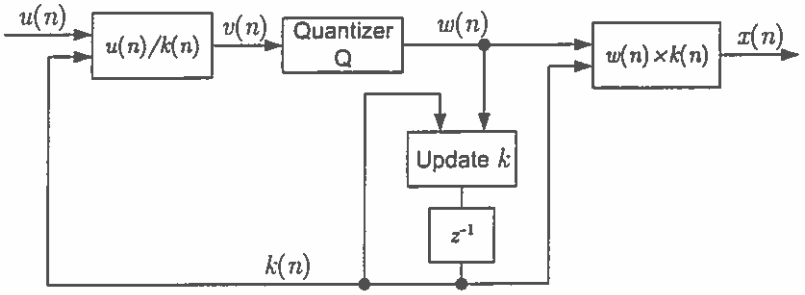

Figure 2.1 Quantizer

An input signal $u(n) = \{1,2,2,8,10,10\}$ is applied. Determine the resulting values of $k(n)$ and the quantization error $(u(n) - x(n))$ for $n = 0,1,\ldots,5$, given that k is initialized to 1. | 5 |

3. a) Draw and label a general block diagram of a single channel noise reduction system for speech enhancement. [5]

b) Consider a specific case of a frequency domain-based single channel noise reduction scheme employing power spectrum subtraction. The noisy signal in time frame l and at frequency bin k is denoted $Z(l,k)$.

i) Describe the operation of this noise reduction scheme and explain what information is needed by the scheme, in addition to the noisy signal. [2]

ii) Show that the noise reduction scheme can be viewed as a filtering operation such that the output, $Z_o(l,k)$, after processing by the scheme can be written

$$Z_o(l,k) = H(l,k)Z(l,k)$$

and then determine the expression for $H(l,k)$ in this case. [5]

c) Now consider a system with two microphones, separated physically by a small distance. Each of the two microphones receives the same speech signal $s(n)$ corrupted by independent additive noise sources $e_1(n)$ and $e_2(n)$ respectively. The signals at the two microphones $x_1(n)$ and $x_2(n)$ are given by

$$x_1(n) = s(n) + e_1(n)$$
$$x_2(n) = s(n) + e_2(n).$$

With the expectation operator denoted $E[\cdot]$, the noise sources have the following properties in terms of their means and variances:

$$E[e_1(n)] = E[e_2(n)] = 0$$
$$E[e_1^2(n)] = \sigma_1^2$$
$$E[e_2^2(n)] = \sigma_2^2.$$

Now consider a weighted sum of the microphone signals

$$z(n) = ax_1(n) + (1-a)x_2(n)$$

where a is a scalar constant.

i) If $z(n)$ is written as

$$z(n) = s(n) + e_3(n),$$

find $e_3(n)$ in terms of $e_1(n)$ and $e_2(n)$. [2]

ii) Next, find the optimal value of a that minimizes the variance of $e_3(n)$. [4]

iii) Finally, determine the minimum variance of $e_3(n)$ in terms of σ_1^2 and σ_2^2. [2]

4. Consider a speech signal that has been segmented into time frames. A feature vector $\mathbf{x}_t$ is computed from each frame $t = 1,2,\ldots,T$ of the speech signal, where T is the number of frames. The set of feature vectors representing the speech signal is then denoted $\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_T\}$. Next consider a speech recognition system based on a hidden Markov model containing S states $\{s_1,s_2,\ldots,s_S\}$.

a) Explain with the use of any appropriate diagrams how the hidden Markov model can be used to recognize speech. Include a clear list and explanation of the parameters involved. Also include an explanation of an *alignment* in this context. Further include the definitions of the output probability density and transition probability in the hidden Markov model. [6]

b) Let $P(t,s)$ denote the total probability density of all the possible alignments of $\{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_t\}$ given that $\mathbf{x}_1$ is aligned to state s_1 and $\mathbf{x}_t$ is aligned to state s. Then let $Q(t,s)$ denote the total probability density of all the possible alignments of $\{\mathbf{x}_{t+1},\mathbf{x}_{t+2},\ldots,\mathbf{x}_T\}$ given that $\mathbf{x}_t$ is aligned to state s and $\mathbf{x}_T$ is aligned to state S.

Now show how $P(t,s)$ and $Q(t,s)$ could be computed recursively and explain your reasoning. Clarify your approach by expressing $P(t,s)$ in terms of $P(t-1,k)$ for $k = 1,2,\ldots,S$, and also by expressing $Q(t,s)$ in terms of $Q(t+1,k)$ for $k = 1,2,\ldots,S$. Also state the initial conditions for P and Q for recursive computation. [6]

c) A hidden Markov model with 4 states is going to be used for a simplified speech recognition task.

i) Sketch the state diagram of a hidden Markov model with 4 states. Label the diagram and the state transitions such that the model is constrained to be a *'left-to-right, no skips'* model. [2]

ii) The state transition probabilities are

$$a_{12} = 0.1,\ a_{23} = 0.5,\ a_{34} = 0.8$$

and the exit probability, denoted $a_{4=}$, has the value $a_{4=} = 0.5$.

Label these transition probabilities on the state diagram.

The output probability densities for each of 6 observed feature vectors are shown in Table 1. Determine the total probability of all alignments of the observation with the model for which frame 3 is in state 2 given that frame $\mathbf{x}_1$ is in state s_1 and frame $\mathbf{x}_6$ is in state s_4. Show your answer to 6 decimal places. [6]

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$	$\mathbf{x}_6$
s_1	0.5	0.3	0.5	0.2	0.1	0.1
s_2	0.4	0.5	0.8	0.6	0.3	0.8
s_3	0.2	0.8	0.2	0.8	0.2	0.2
s_4	0.5	0.4	0.5	0.2	0.5	0.8

Table 1 Output probability densities.