IMPERIAL COLLEGE LONDON

**EE4-14**
**EE9-SO16**

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2013

MSc and EEE/ISE PART IV: MEng and ACGI

Corrected Copy

# SPEECH PROCESSING

Wednesday, 15 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) Show that the transfer function of the lossless tube model of the speech production system can be written

$$V(z) = \frac{0.5(1+r_G)\prod_{k=1}^{N}(1+r_k)z^{-N/2}}{D(z)}$$

with reflection coefficients $r_k$ and denominator

$$D(z) = \begin{bmatrix} 1 & -r_G \end{bmatrix} \begin{bmatrix} 1 & -r_1 \\ -r_1 z^{-1} & z^{-1} \end{bmatrix} \cdots \begin{bmatrix} 1 & -r_N \\ -r_N z^{-1} & z^{-1} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

[ 6 ]

b) For the case when $r_G = 1$, $D(z)$ can be written $D(z) = \mathbf{P}_N \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, where vectors and matrices are written in bold typeface. Find a recursive expression for $\mathbf{P}_k$ in terms of $\mathbf{P}_{k-1}$ for $k = 1, \ldots, N$. Hence deduce a recursive expression for $D_k(z)$ in terms of $D_{k-1}(z)$. Exploit the recursive expression to write out $D_1(z)$, $D_2(z)$ and $D_3(z)$. [ 7 ]

c) Draw a labelled signal flow graph corresponding to the lossless tube model of order 3 using (i) delays corresponding only to half a sampling period and (ii) delays corresponding to one sampling period. Latency can be ignored. [ 4 ]

d) Using the recursion of part (b) in reverse order, find an expression for $D_{k-1}(z)$ in terms of $D_k(z)$, clearly indicating the range of $k$. [ 3 ]

2. Consider a model of the vocal tract filter comprising an all-pole filter $V(z) = \frac{1}{A(z)}$ of order $p$ and characterized by a set of parameters.

a) Write down general expressions for $A(z)$ in terms of the following parameters.

i) Reflection coefficients, $r_k$. (Assume that the reflection coefficient at the glottis = 1).

ii) Predictor coefficients, $a_k$. [ 2 ]

b) Hence write down expressions for the equivalent parameter sets of

i) Log area ratios, $g_k$.

ii) Cepstral coefficients, $c_k$. [ 2 ]

c) Explain the desirable properties of parametric representations of the vocal tract filter for (i) speech recognition and (ii) speech coding and, hence, rank order these 4 parameter sets in terms of their suitability for both these two applications. [ 4 ]

d) For the case of $p = 2$, find expressions for the log area ratios and the cepstral coefficients in terms of the predictor coefficients. [ 4 ]

e) Also for the case of $p = 2$, consider the terms $\alpha = (z + z^{-1})$ and $\beta = (z^2 + z^{-2})$.

i) Derive an expression for the gain of the $A(z)$ in dB in terms of $\alpha$ and $\beta$. Find at what frequency $\alpha = 1$ and $\beta = -1$ and determine the gain at that frequency given $a_1 = 0.98$ and $a_2 = -a_1^2$. [ 4 ]

ii) Now consider more specifically two alternative parameter sets characterising the vocal tract filter: predictor coefficients $\{a_1, a_2\}$, and log area ratios $\{g_1, g_2\}$. For the frequency obtained above, calculate the change in gain due to -1% change in $a_2$ (with all other factors unchanged) and calculate the change in gain due to -1% change in $g_2$ (with all other factors unchanged).

Comment on the significance of the result for the choice of parameters for speech coding. [ 4 ]

3. a) In the context of speech enhancement, consider noise $v(n)$, clean speech $s(n)$ and noisy speech $y(n)$.

i) Formulate and describe the minimum statistics approach for noise estimation given noisy speech. Include relevant mathematical formulations in your description. [ 6 ]

ii) Obtain an expression for the filter $H(l,k)$, for time frame $l$ and frequency index $k$, that implements amplitude spectral subtraction exploiting a noise estimate. [ 4 ]

b) Consider a cascade formant synthesizer used to synthesize speech sounds at a sampling frequency of 8 kHz.

i) Describe the cascade formant synthesizer and give an illustrative diagram. [ 2 ]

ii) A cascade formant synthesizer is used to synthesize speech containing a formant of bandwidth 50 Hz with a centre frequency of 650 Hz. Explain how this formant would be modelled. Determine all relevant parameters. [ 3 ]

c) In a parallel formant synthesizer, the excitation signal input to each formant filter contains the weighted sum of a periodic signal component and a random signal component. Figure 3.1 shows the ratio of these two signal components for each of the first four formants, F1 to F4, as a function of a Voicing Control parameter.

i) Explain how this type of voicing control would be integrated into a parallel formant synthesizer and illustrate your explanation with a block diagram for the specific example with the value of the Voicing Control parameter = 0.5. [ 3 ]

ii) State an example phoneme for which the Voicing Control parameter would be close to 0.5 and, for this example, comment on the level of voicing in F1 and F4. [ 2 ]

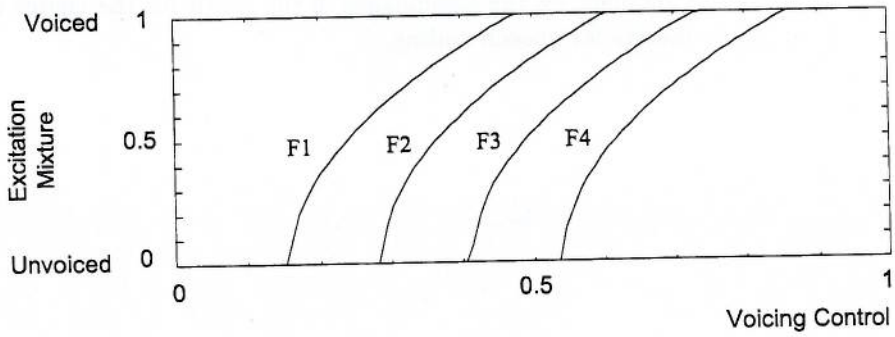



Figure 3.1 Voicing Control

4. In a speech recognition task, a female speech utterance has been segmented into $T = 5$ frames, each with a corresponding time duration of 20 ms, and denoted $\mathbf{x}_t$ for $t = 1, 2, \ldots, T$. The speech frames are compared with a hidden Markov model with $S = 4$ states having output log probability densities for frame $t$ in state $i$ of $\log(d_i(\mathbf{x}_t))$ as shown in Table 1. The transition probability from state $i$ to state $j$ of the model is $a_{ij}$. The transition probabilities are indicated on the state diagram of the model in Fig. 4.1.

| | Input Frame | | | | |
|---|---|---|---|---|---|
| State | $\mathbf{x}_1$ | $\mathbf{x}_2$ | $\mathbf{x}_3$ | $\mathbf{x}_4$ | $\mathbf{x}_5$ |
| 1 | -5 | -8 | -5 | -5 | -7 |
| 2 | -4 | -5 | -7 | -7 | -6 |
| 3 | -5 | -3 | -3 | -8 | -6 |
| 4 | -6 | -7 | -8 | -4 | -8 |

Table 1 Output log probability densities

The maximum probability density that the model generates frames $\mathbf{x}_1, \ldots, \mathbf{x}_t$ from any sequence of states for which frame 1 is in state 1 and frame $t$ is in state $s$ is defined at $B(t,s)$.

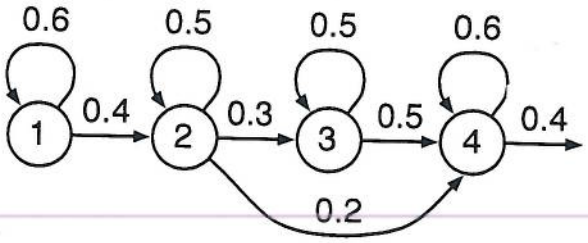



Figure 4.1 Hidden Markov model

a) Explain any advantages obtained by speech recognition systems from the use of log probability densities instead of probability densities. [ 2 ]

b) For the case where

$$d_i(\mathbf{x}) = (2\pi)^{-1/2P}|\mathbf{C}_i|^{-1/2}\exp\left(-1/2(\mathbf{x}-\mathbf{m}_i)^T\mathbf{C}_i^{-1}(\mathbf{x}-\mathbf{m}_i)\right)$$

write a simplified expression for $\log(d_i(\mathbf{x}))$. Explain any advantageous features in the computation involved in calculating $\log(d_i(\mathbf{x}))$. [ 4 ]

c) For the speech recognition task described above, determine $\log(B(5,4))$ and the most likely state sequence. [ 12 ]

d) Briefly discuss the advantages and disadvantages of increasing the time duration corresponding to each frame by a factor of 10. [ 2 ]

1. a) Show that the transfer function of the lossless tube model of the speech production system can be written

$$V(z) = \frac{0.5(1+r_G)\prod_{k=1}^{N}(1+r_k)z^{-N/2}}{D(z)}$$

with reflection coefficients $r_k$ and denominator

$$D(z) = \begin{bmatrix} 1 & -r_G \end{bmatrix} \begin{bmatrix} 1 & -r_1 \\ -r_1 z^{-1} & z^{-1} \end{bmatrix} \cdots \begin{bmatrix} 1 & -r_N \\ -r_N z^{-1} & z^{-1} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

[ 6 ]

***Solution:***
*The length of each segment is the distance travelled by propagation of sound in 0.5 samples.*
*Segment delays are modelled by*

$$\begin{bmatrix} U \\ V \end{bmatrix} = z^{+1/2} \begin{bmatrix} 1 & 0 \\ 0 & z^{-1} \end{bmatrix} \begin{bmatrix} W \\ X \end{bmatrix}.$$

*Segment junctions are modelled by*

$$\begin{bmatrix} U \\ V \end{bmatrix} = \frac{1}{1+r} \begin{bmatrix} 1 & -r \\ -r & 1 \end{bmatrix} \begin{bmatrix} W \\ X \end{bmatrix}.$$

*Combining these in cascade gives rise to a product and the given expression follows.*

b) For the case when $r_G = 1$, $D(z)$ can be written $D(z) = \mathbf{P}_N \begin{bmatrix} 1 \\ 0 \end{bmatrix}$, where vectors and matrices are written in bold typeface. Find a recursive expression for $\mathbf{P}_k$ in terms of $\mathbf{P}_{k-1}$ for $k = 1, \ldots, N$. Hence deduce a recursive expression for $D_k(z)$ in terms of $D_{k-1}(z)$. Exploit the recursive expression to write out $D_1(z)$, $D_2(z)$ and $D_3(z)$. [ 7 ]

***Solution***

*Because we are seeking a recursive solution, and using $r_G = 1$, let us start with*

$$\mathbf{P}_1 = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & -r_1 \\ -r_1 z^{-1} & z^{-1} \end{bmatrix} = \begin{bmatrix} (1+r_1 z^{-1}), & -(r_1 + z^{-1}) \end{bmatrix}.$$

*It follows that*

$$\mathbf{P}_2 = \mathbf{P}_1 \begin{bmatrix} 1 & -r_2 \\ -r_2 z^{-1} & z^{-1} \end{bmatrix}.$$

*By induction*

$$\mathbf{P}_k = \mathbf{P}_{k-1} \begin{bmatrix} 1 & -r_k \\ -r_k z^{-1} & z^{-1} \end{bmatrix} \quad k = 1, \ldots, N.$$

*Now concerning $D$, starting with $D_1(z) = 1 + r_1 z^{-1}$ it can be seen that*

$$\mathbf{P}_1 = \begin{bmatrix} D_1(z), & -z^{-1} D_1(z^{-1}) \end{bmatrix}$$

*and by induction similarly*

$$\mathbf{P}_k = \begin{bmatrix} D_k(z), & -z^{-k} D_k(z^{-1}) \end{bmatrix}.$$

*Substituting for $\mathbf{P}_2$, we find*

$$\mathbf{P}_2 = \begin{bmatrix} D_1(z) + r_2 z^{-2} D_1(z^{-1}), & -r_2 D_1(z) - z^{-2} D_1(z^{-1}) \end{bmatrix}$$

*which can be written as*

$$\mathbf{P}_2 = \begin{bmatrix} D_2(z) & -z^{-2} D_2(z^{-1}) \end{bmatrix}$$

*with*

$$D_2(z) = D_1(z) + r_2 z^{-2} D_1(z^{-1}).$$

*So we finally obtain*

$$D_k(z) = D_{k-1}(z) + r_k z^{-k} D_{k-1}(z^{-1}).$$

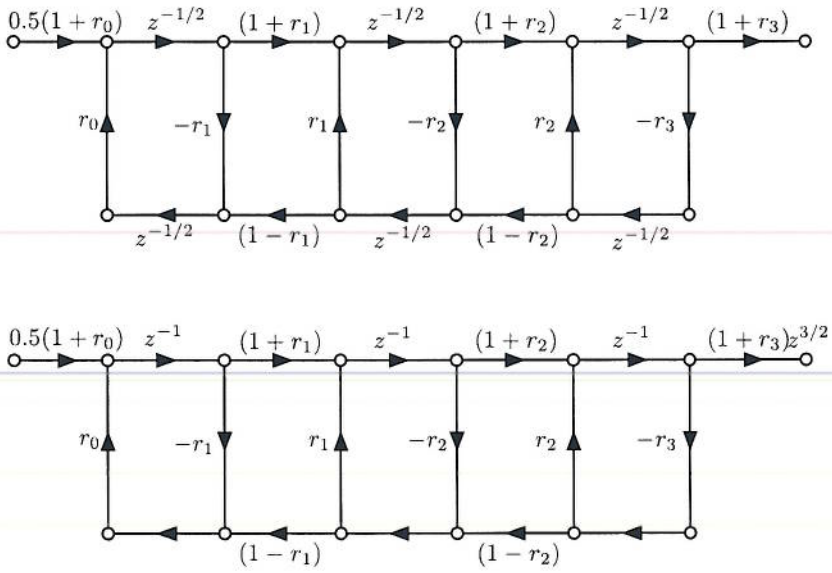
This leads to

$$\begin{aligned} D_1(z) &= 1 + r_1 z^{-1} \\ D_2(z) &= 1 + (r_1 + r_1 r_2) z^{-1} + r_2 z^{-2} \\ D_3(z) &= 1 + (r_1 + r_1 r_2 + r_2 r_3) z^{-1} + (r_2 + r_1 r_3 + r_1 r_2 r_3) z^{-2} + r_3 z^{-3}. \end{aligned}$$

c) Draw a labelled signal flow graph corresponding to the lossless tube model of order 3 using (i) delays corresponding only to half a sampling period

and (ii) delays corresponding to one sampling period. Latency can be ignored. [ 4 ]

***Solution***




d) Using the recursion of part (b) in reverse order, find an expression for $D_{k-1}(z)$ in terms of $D_k(z)$, clearly indicating the range of $k$. [ 3 ]

***Solution***

$$D_{k-1}(z) = \frac{D_k(z) - r_k D_k(z^{-1}) z^{-k}}{1 - r_k^2} \qquad k = N, N-1, \ldots, 2.$$

2. Consider a model of the vocal tract filter comprising an all-pole filter $V(z) = \frac{1}{A(z)}$ of order $p$ and characterized by a set of parameters.

a) Write down general expressions for $A(z)$ in terms of the following parameters.

i) Reflection coefficients, $r_k$. (Assume that the reflection coefficient at the glottis = 1).

ii) Predictor coefficients, $a_k$. [ 2 ]

***Solution***

$$A(z) = \begin{pmatrix} 1 & -1 \end{pmatrix} \prod_{k=1}^{p} \begin{pmatrix} 1 & -r_k \\ -r_k z^{-1} & z^{-1} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$A(z) = 1 - \sum_{k=1}^{p} a_k z^{-k}$$

b) Hence write down expressions for the equivalent parameter sets of

i) Log area ratios, $g_k$.

ii) Cepstral coefficients, $c_k$. [ 2 ]

***Solution***

$$g_k = \log \left( \frac{1 + r_k}{1 - r_k} \right)$$

$$c_k = a_k + \frac{1}{k} \sum_{j=1}^{k-1} (k - j) c_{k-j} a_j$$

c) Explain the desirable properties of parametric representations of the vocal tract filter for (i) speech recognition and (ii) speech coding and, hence, rank order these 4 parameter sets in terms of their suitability for both these two applications. [ 4 ]

***Solution***

*Recognition: We want parameters that are good at distinguishing between different speech sounds. i.e. different speech sounds should give different parameter values whereas multiple examples of any particular speech sounds should give similar values. Best: Cepstral coefficients. Worst: Predictor coefficients.*

*Coding: Quantisation errors must not affect the spectrum too much. It is best if parameters have a natural order (unlike for example the pole positions) as this means the transmission order conveys useful information and a particular parameter will have less variability. Want easy stability check. Want to interpolate between frames. Best: reflection coefficients or log area ratios. Worst: Predictor coefficients.*

d) For the case of $p = 2$, find expressions for the log area ratios and the cepstral coefficients in terms of the predictor coefficients. [ 4 ]

***Solution***

$$g_1 = \log\left(\frac{1-a_1-a_2}{1+a_1-a_2}\right)$$
$$g_2 = \log\left(\frac{1-a_2}{1+a_2}\right)$$
$$c_1 = a_1$$
$$c_2 = a_2 + 0.5a_1^2$$

e) Also for the case of $p = 2$, consider the terms $\alpha = (z + z^{-1})$ and $\beta = (z^2 + z^{-2})$.

i) Derive an expression for the gain of the $A(z)$ in dB in terms of $\alpha$ and $\beta$. Find at what frequency $\alpha = 1$ and $\beta = -1$ and determine the gain at that frequency given $a_1 = 0.98$ and $a_2 = -a_1^2$. [ 4 ]

***Solution***

$$-10\log_{10}\left((1-a_1z^{-1}-a_2z^{-2})(1-a_1z-a_2z^2)\right)$$
$$-10\log_{10}\left(1-a_1^2+a_2^2+a_1(a_2-1)(z+z^{-1})-a_2(z^2+z^{-2})\right)$$

The conditions are found from solving $e^{j\omega} + e^{-j\omega} = 1, \cos\omega = 1/2$ to obtain a normalized frequency of 1/6 Hz. This gives a gain of 29.3 dB.

ii) Now consider more specifically two alternative parameter sets characterising the vocal tract filter: predictor coefficients $\{a_1, a_2\}$, and log area ratios $\{g_1, g_2\}$. For the frequency obtained above, calculate the change in gain due to -1% change in $a_2$ (with all other factors unchanged) and calculate the change in gain due to -1% change in $g_2$ (with all other factors unchanged).

Comment on the significance of the result for the choice of parameters for speech coding. [ 4 ]

***Solution***

| | Original | a2-1% | g2-1% |
|---|---|---|---|
| a1 | 0.980 | 0.980 | 0.979 |
| a2 | -0.960 | -0.951 | -0.959 |
| g1 | -1.098 | | -1.098 |
| g2 | 3.902 | | 3.863 |
| Gain dB | 29.26 | 27.35 | 28.96 |

It can be seen that the vocal tract filter gain is more robust to small changes in log area ratios. Hence these would be a better choice for quantization in a coder.

3. a) In the context of speech enhancement, consider noise $v(n)$, clean speech $s(n)$ and noisy speech $y(n)$.

i) Formulate and describe the minimum statistics approach for noise estimation given noisy speech. Include relevant mathematical formulations in your description. [ 6 ]

***Solution***

*This technique is based on the assumption that, at some frequency, during a speech pause, or within brief periods between words and even syllables, the speech energy is close to zero. As a result, a short-term power spectrum estimate of the noisy signal, even during speech activity, decays frequently due to the noise power. Thus, by tracking the temporal spectral minimum without distinguishing between speech presence and speech absence, the noise power in a specific frequency band can be estimated.*

*The noisy speech is given by* $y(n) = s(n) + v(n)$. *Then the variance is obtain as*

$$\hat{\phi}_y(l,k) = \alpha\hat{\phi}_y(l-1,k) + (1-\alpha)|Y(l,k)|^2$$

*The noise can then be estimated as*

$$\hat{\phi}_v(l,k) = \min\left\{\hat{\phi}_y(l,k), \hat{\phi}_y(l-1,k), \ldots, \hat{\phi}_y(l,D+1,k)\right\}$$

ii) Obtain an expression for the filter $H(l,k)$, for time frame $l$ and frequency index $k$, that implements amplitude spectral subtraction exploiting a noise estimate. [ 4 ]

***Solution***

$A_z(l,k) = A_y(l,k) - \sqrt{\hat{\phi}_v(k,k)}$

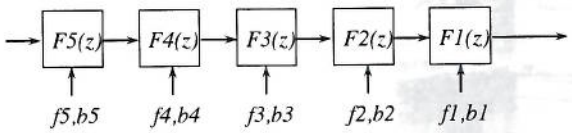
$Z(l,k) = H(l,k)Y(l,k)$ *with* $H(l,k) = 1 - \sqrt{\frac{\hat{\phi}_v(l,k)}{|Y(l,k)|^2}}$.

b) Consider a cascade formant synthesizer used to synthesize speech sounds at a sampling frequency of 8 kHz.

i) Describe the cascade formant synthesizer and give an illustrative diagram. [ 2 ]

***Solution:***

$$\frac{1}{(1-e^{-\pi b+2j\pi f}z^{-1})(1-e^{-\pi b-2j\pi f}z^{-1})} = \frac{1}{1-2e^{-\pi b}\cos(2\pi f)z^{-1}+e^{-2\pi b}z^{-2}}$$




ii) A cascade formant synthesizer is used to synthesize speech containing a formant of bandwidth 50 Hz with a centre frequency of 650 Hz. Explain how this formant would be modelled. Determine all relevant parameters. [ 3 ]

***Solution***
*Modelled by a pair of complex conjugate poles such that*

$$G(z) = \frac{1}{(1-\exp(-\pi b + j2\pi f)z^{-1})(1-\exp(-\pi b - j2\pi f)z^{-1})}$$
$$= \frac{1}{1-2\exp(-\pi b)\cos(2\pi f)z^{-1}+\exp(-2\pi b)z^{-2}}$$

*Now for the case with $f = 650$ Hz and bandwidth $b = 50$ we obtain the coefficient of $z^{-1}$ in the denominator as 1.7111 and the coefficient of $z^{-2}$ as 0.9615, hence*

$$G(z) = \frac{1}{1-1.7111z^{-1}+0.9615z^{-2}}$$

c) In a parallel formant synthesizer, the excitation signal input to each formant filter contains the weighted sum of a periodic signal component and a random signal component. Figure 3.1 shows the ratio of these two signal components for each of the first four formants, F1 to F4, as a function of a Voicing Control parameter.

i) Explain how this type of voicing control would be integrated into a parallel formant synthesizer and illustrate your explanation with a block diagram for the specific example with the value of the Voicing Control parameter = 0.5. [ 3 ]

***Solution***
*We would expect to integrate the voicing control to set the relative levels of voiced and unvoiced excitation into each formant resonator. The unvoiced excitation (turbulent excitation) originates along the vocal tract whereas the periodic excitation originates at the larynx. In this case, a different transfer function should be applied to the two signals. In practice, the proportion of larynx excitation is greater for low frequency formants than for high frequency formants.*

ii) State an example phoneme for which the Voicing Control parameter would be close to 0.5 and, for this example, comment on the level of voicing in F1 and F4. [ 2 ]

*Solution*
*An example phoneme with voice control of approximately 0.5 is /z/ (mixed voicing and turbulent excitation). For this case, the formant F1 would be 100% voiced and formant F4 would be 0% voiced.*

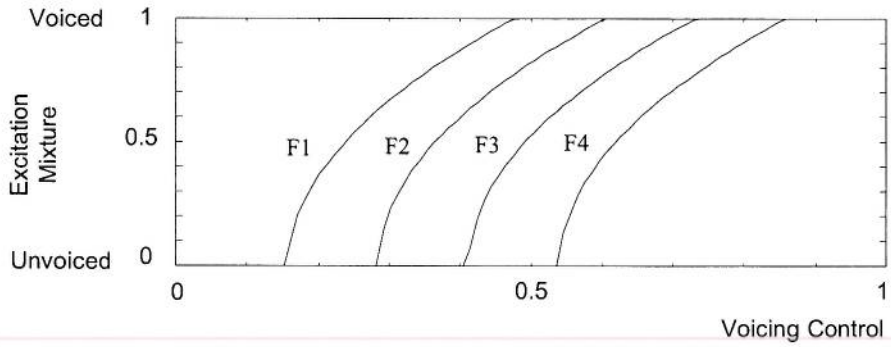



Figure 3.1 Voicing Control

4. In a speech recognition task, a female speech utterance has been segmented into $T = 5$ frames, each with a corresponding time duration of 20 ms, and denoted $\mathbf{x}_t$ for $t = 1, 2, \ldots, T$. The speech frames are compared with a hidden Markov model with $S = 4$ states having output log probability densities for frame $t$ in state $i$ of $\log(d_i(\mathbf{x}_t))$ as shown in Table 1. The transition probability from state $i$ to state $j$ of the model is $a_{ij}$. The transition probabilities are indicated on the state diagram of the model in Fig. 4.1.

| | Input Frame | | | | |
|---|---|---|---|---|---|
| State | $\mathbf{x}_1$ | $\mathbf{x}_2$ | $\mathbf{x}_3$ | $\mathbf{x}_4$ | $\mathbf{x}_5$ |
| 1 | -5 | -8 | -5 | -5 | -7 |
| 2 | -4 | -5 | -7 | -7 | -6 |
| 3 | -5 | -3 | -3 | -8 | -6 |
| 4 | -6 | -7 | -8 | -4 | -8 |

Table 1 Output log probability densities

The maximum probability density that the model generates frames $\mathbf{x}_1, \ldots, \mathbf{x}_t$ from any sequence of states for which frame 1 is in state 1 and frame $t$ is in state $s$ is defined at $B(t, s)$.

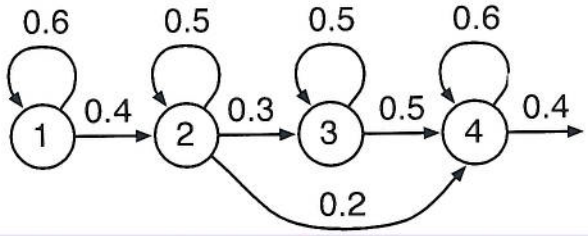



Figure 4.1 Hidden Markov model

a) Explain any advantages obtained by speech recognition systems from the use of log probability densities instead of probability densities. [ 2 ]

***Solutions:***
*When many probabilities are multiplied together, the result may underflow the computer's floating point representation. The use of the log enables a much larger numerical dynamic range to be accommodated.*

b) For the case where

$$d_i(\mathbf{x}) = (2\pi)^{-1/2P} |\mathbf{C}_i|^{-1/2} \exp\left(-1/2(\mathbf{x} - \mathbf{m}_i)^T \mathbf{C}_i^{-1} (\mathbf{x} - \mathbf{m}_i)\right)$$

write a simplified expression for $\log(d_i(\mathbf{x}))$. Explain any advantageous features in the computation involved in calculating $\log(d_i(\mathbf{x}))$. [ 4 ]

***Solution:***

$$\log(d_i(\mathbf{x})) = -\frac{1}{2}\left(P\log(2\pi) + \log(|\mathbf{C}_i|) + (\mathbf{x} - \mathbf{m}_i)^T \mathbf{C}_i^{-1} (\mathbf{x} - \mathbf{m}_i)\right)$$

*The terms involving the log in the above expression are additive and only need to be calculated once since they do not depend on* $\mathbf{x}$.

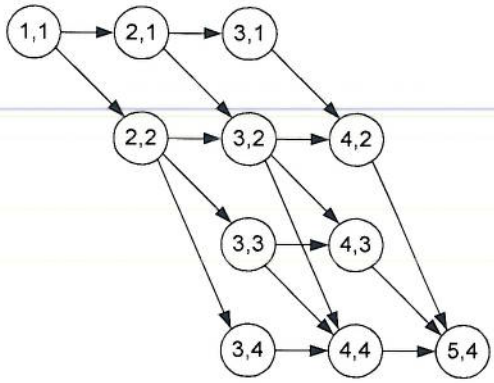
c) For the speech recognition task described above, determine $\log(B(5,4))$ and the most likely state sequence. [ 12 ]

***Solution:***

*The table of transition probabilities in log form is:*

| | 1 | 2 | 3 | 4 | end |
|---|---|---|---|---|---|
| 1 | -0.511 | -0.916 | | | |
| 2 | | -0.693 | -1.204 | -1.609 | |
| 3 | | | -0.693 | -0.693 | |
| 4 | | | | -0.511 | -0.916 |

*The alignment lattice can be constructed as:*




*The table of log probabilities is then given by*

| | | |
|---|---|---|
| B11 | logd(1,1) | **-5** |
| B21 | B11 + loga(1,1) + logd(1,2) = -13.5108 | -13.5108 |
| B22 | B11 + loga(1,2) + logd(2,2) = -10.9163 | **-10.9163** |
| B31 | B21 + loga(1,1) + logd(1,3) = -19.0217 | -19.0217 |
| B32 | B21 + loga(1,2) + logd(2,3)= -21.4271 | -18.6094 |
| | B22 + loga(2,2) + logd(2,3) = **-18.6094** | |
| B33 | B22 + loga(2,3) + logd(3,3) = -15.1203 | **-15.1203** |
| B34 | B22 + loga(2,4) + logd(4,3) = -20.5257 | -20.5257 |
| B42 | B31 + loga(1,2) + logd(2,4) = -26.9379 | -26.3026 |
| | B32 + loga(2,2) + logd(2,4) = **-26.3026** | |
| B43 | B32 + loga(2,3) + logd(3,4) = -27.8134 | -23.8134 |
| | B33 + loga(3,3) + logd(3,4) = **-23.8134** | |
| B44 | B32 + loga(2,4) + logd(4,4) = -24.2189 | **-19.8134** |
| | B33 + loga(3,4) + logd(4,4) = **-19.8134** | |
| | B34 + loga(4,4) + logd(4,4) = -25.0366 | |
| B54 | B42 + loga(2,4) + logd(4,5) = -35.9120 | **-28.3242** |
| | B43 + loga(3,4) + logd(4,5) = -32.5066 | |
| | B44 + loga(4,4) + logd(4,5) = **-28.3242** | |

*The corresponding path is therefore 1,2,3,4,4.*

d) Briefly discuss the advantages and disadvantages of increasing the time duration corresponding to each frame by a factor of 10. [ 2 ]

***Solution:***
*Disadvantage: This would degrade the discriminant capabilities of the classifier as each frame would likely contain more than one phoneme. Therefore the parameterisation of the frame would not be indicative of any particular phoneme but instead a combination of phonetic content.*
*Advantage: Reducing the number of frames reduces the complexity of the alignment task proportionately.*