E4.14
SO16
ISE4.17

IMPERIAL COLLEGE LONDON

DEPARTMENT OF ELECTRICAL AND ELECTRONIC ENGINEERING
EXAMINATIONS 2005

MSc and EEE/ISE PART IV: MEng and ACGI

SPEECH PROCESSING

Monday, 25 April 10:00 am

Time allowed: 3:00 hours

There are SIX questions on this paper.

Figure 1 is on separate sheet

Answer FOUR questions.

All questions carry equal marks

Figure 1 is on a separate sheet

Corrected Copy

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

Examiners responsible	First Marker(s) :	P.A. Naylor
	Second Marker(s) :	D.M. Brookes

© University of London 2005

Special Instructions for Invigilators: None

Information for Candidates:

Numbers in square brackets against the right margin of the following pages are a guide to the marking scheme.

A spectrogram $P(k,m)$ can be obtained by computing

$$P(k,m) = \left| \sum_{i=0}^{N-1} w(i)x(m-i) \exp\left(-\frac{2\pi j}{N} k(m-i) \right) \right|^2$$

where $x(n)$ is a speech signal and $w(i)$ is a window of length N.

(a) State the significance of the variables k and m. Hence give a brief description of the nature of the information captured in a spectrogram and the purpose of a spectrogram in speech processing. [4]

(b) Discuss two ways in which the bandwidth of the window $w(n)$ is typically measured. Choose one commonly used window in speech processing and give the values of both these measurements. Draw a labelled sketch showing the significant features of the window. [4]

(c) [5]

(i) Define frequency resolution and time resolution in spectrograms.

(ii) Give formulae for each in terms of the sampling frequency f_s and the bandwidth of the window as defined in part (b) above.

(iii) State the relationship between frequency resolution and time resolution.

(d) Consider Figure 1 which shows a waveform of male speech and two spectrograms computed from the speech data. [7]

(i) Comment on and explain in detail the differences between the two spectrograms.

(ii) Explain the relationship between the first two formant frequencies and the position and shape of the tongue within the mouth.

(iii) Observe the spectrograms at 1.96 s and 2.18 s and estimate the frequencies in Hz of the first two formants. Give a reasoned argument as to whether or not each of the phonemes in the following list corresponds to the spectrogram at the two time instants.

i	as in "bead"
ʒ (Z)	as in "vision"
ɛ (E)	as in "bed"
ʃ (S)	as in "she"
ɒ (o)	as in "body"

2 (a) Explain why speech synthesisers commonly concatenate diphones rather than individual phoneme segments to generate speech. [4]

(b) Consider a segment $s(n)$ of a voiced speech signal with pitch of 100 Hz containing 2400 samples with sampling frequency 8 kHz. The segment $s(n)$ is to be transformed into $y(n)$ using pitch-synchronous overlap-add procedure (PSOLA) such that $y(n)$ is identical in amplitude and duration to $s(n)$ but the pitch of $y(n)$ is to be modified such that it varies as specified in Figure 2. The first pitch mark in $s(n)$ and $y(n)$ occurs at $n = 0$.

(i) Show that pitch marks in $s(n)$ occur at $n = 2000$, 2080 and 2160. [2]

(ii) Show that pitch marks in $y(n)$ occur at $n = 2056$ and 2120. [2]

(iii) Describe the main elements of the method by which PSOLA generates $y(n)$ in the region close to $n = 2090$. [6]

(iv) Derive an expression for $y(n)$ at $n = 2090$. [6]

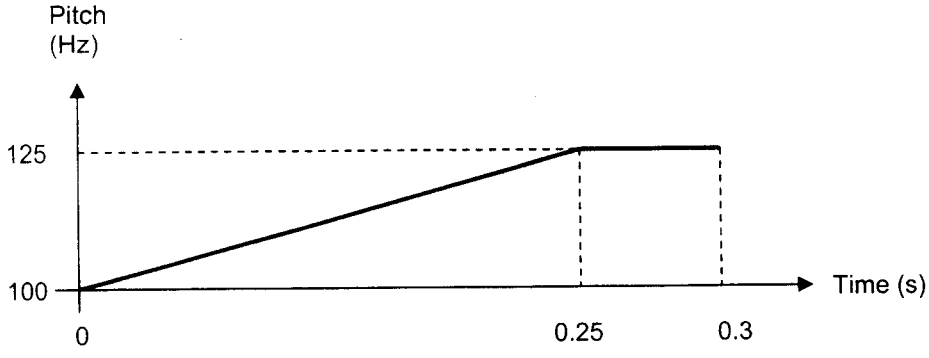


Figure 2

3. (a) Consider a general quantizer and assume that, for any given quantization bin, the input signal is uniformly distributed over the amplitude range of the bin. Derive an expression for the RMS quantization error in terms of the bin width w. [3]

(b) Discuss the important differences between uniform and non-uniform quantization of speech signals. [7]

A particular speech signal s has probability density function $p(s)$. Consider the design of a non-uniform quantizer for s such that input amplitudes in the interval $[a_{i-1}, a_i]$ are quantized to s_i for $i = 1, 2, \ldots, N$. Find an expression for the quantization levels s_i, in terms of $p(s)$, a_{i-1} and a_i that give minimum mean squared quantization error.

(c) For a speech signal, consider the 8-bit μ-law quantization scheme with bin centres at [5]

$$\pm\{(m+16½)\ 2^e - 16½\}.$$

(i) Give a brief description of this scheme with specific reference to quantization noise.
(ii) State the maximum signal amplitude that can be represented using this scheme assuming that 4 bits are used for the mantissa.
(iii) Deduce the error amplitudes in quantizing input values of 27 and 1027.

(d) In the diagram of Figure 3, the uniform 5-level quantizer labelled Q has outputs from the set $\{-2, -1, 0, 1, 2\}$. After each sample is quantized, the factor $k(n)$ is updated as follows: [5]

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

Given an input signal $u(n) = \{1, 1, 1, 10, 10, 10, 1, 1\}$ for $n = 0, 1, \ldots, 7$, construct a table showing the corresponding values of $k(n)$ and $x(n)$. Assume that k is initialized to $k = 1.8$.

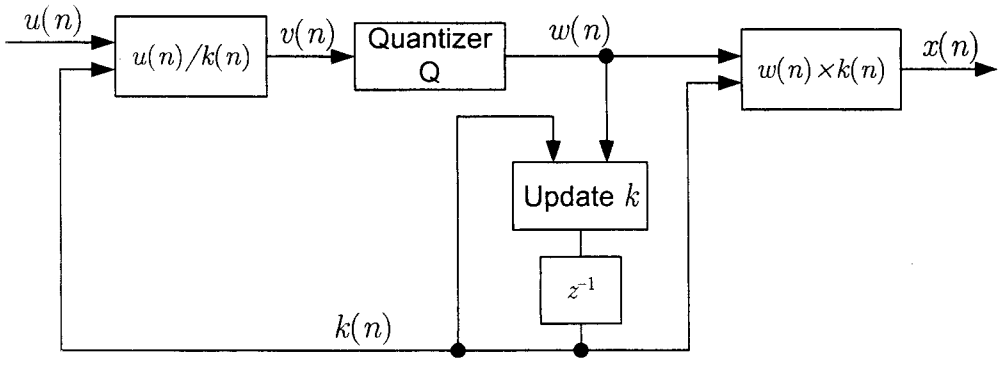


Figure 3

This page is intentionally blank.

4. (a) A speech recogniser employing Hidden Markov models uses a pre-processor that generates feature vectors from the speech every 10 ms. Each feature vector contains 30 elements. Give the number of independent parameters per state needed to specify the observation vector probability distributions when the distributions are: [6]

(i) gaussian with a full covariance matrix,

(ii) gaussian with a diagonal covariance matrix,

(iii) a mixture of 5 diagonal covariance matrix gaussians.

Explain why most speech recognition systems use a mixture of diagonal covariance matrix gaussians.

(b) Part of a Hidden Markov speech model is illustrated in Figure 4. In this model, transitions are possible from state A to state B via any one of K intermediate states numbered 1 to K. The diagram shows only states 1 and K. The other intermediates states are indicated by the dashed lines. The transition probability from state A to intermediate state j is w_j. The transition probability from any of the intermediate states to state B is $(1-p)$. The transition probability from intermediate state i to intermediate state j is pw_j, independent of i. The feature vector is of length F and the output probability density of an observation vector $\mathbf{x}$ in state i is given by [9]

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

where $\mathbf{m}_i$ and $\mathbf{C}_i$ are the mean vector and covariance matrix for state i. Given that $\mathbf{x}_0$ is in state A, derive an expression for the total probability density that the model generates the frames $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_{T+1}$ from a state sequence having $\mathbf{x}_{T+1}$ in state B.

Show that the probability remains unchanged if the intermediate states 1,2,..., K are replaced by a single state with an appropriate gaussian mixture output distribution.

(c) Suppose that F=1 so that $\mathbf{x}_i$, $\mathbf{m}_i$ and $\mathbf{C}_i$ become scalars x_i, m_i and C_i. If $x_1, x_2, \ldots, x_T$ are assumed to be generated by some sequence of the states 1, 2, ..., K, then m_i, C_i and w_i can be iteratively reestimated using the Baum-Welch formulae. [5]
Give the Baum-Welch reestimation formulae for this case in terms of

$$A_{i,t} = \frac{d_i(x_t)}{\sum_{k=1}^{K} d_k(x_t)}$$

and give a descriptive interpretation of the significance of $A_{i,t}$ is these formulae.

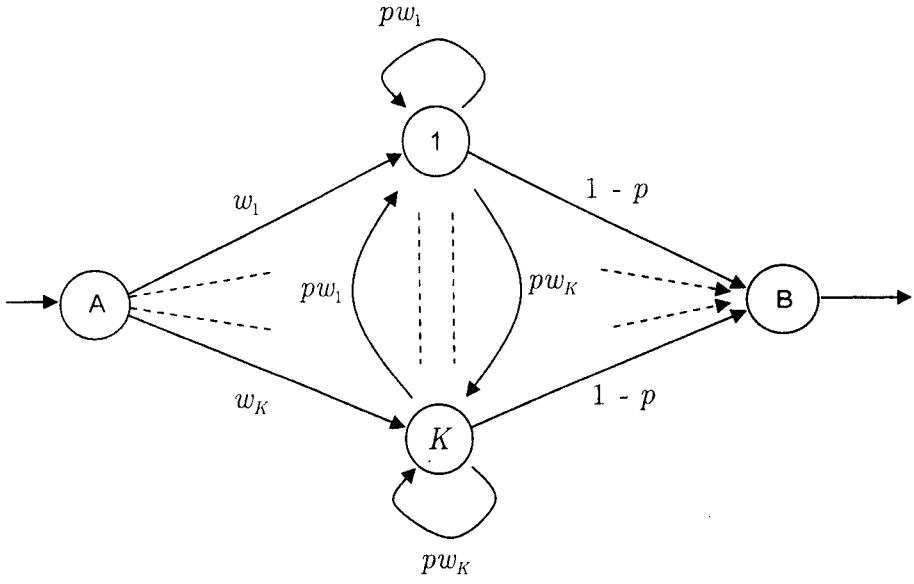


Figure 4

5. (a) Describe the lossless tube model of the vocal tract. Illustrate your description with a labelled sketch. Write down an expression for the total acoustic pressure at a point in the model and state any assumptions made. You may use the relationship that the pressure is inversely proportional to the cross-sectional area. [6]

(b) Explain why the lossless tube model is not a good model for certain speech sounds and give examples. [2]

(c) For some phonemes, the vocal tract may be represented, as in Figure 5, by a tube which includes separate branches for the nose and mouth cavities. In such phonemes the mouth branch is normally closed as shown in the figure.

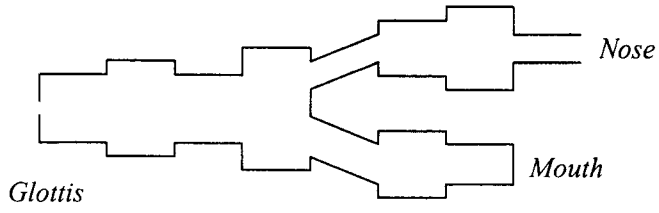


Figure 5

An enlarged view of the point at which the tube splits into two branches is shown in Figure 6. The cross sectional areas of the three tube sections are P, Q and R as indicated. The quantities B, C, D, E, F and G represent the z-transforms of the volume flow rates of the acoustic waves either side of the junction. The propagation direction of each wave is indicated by an arrow.

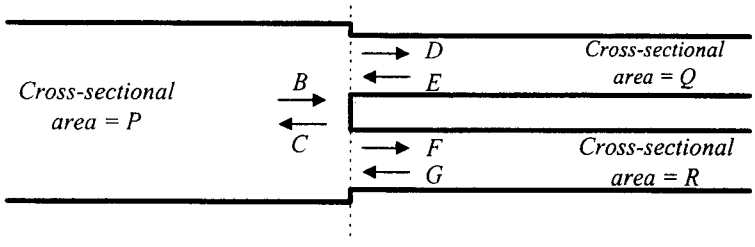


Figure 6

(i) Write down the equations relating the flows B, ..., G that are imposed by equality of pressure at the junction of the three tube sections and by conservation of mass. [4]

(ii) Derive a matrix equation for $\begin{pmatrix} D \\ E \end{pmatrix}$ in terms of $\begin{pmatrix} B \\ C \end{pmatrix}$, the cross sectional areas of the tube sections, and the transfer function $H(z)$ of the closed branch of the tube where $H = G / F$. [8]

6. (a) Consider the speech signal $s(n)$, $n = 0, 1, \ldots, N\text{–}1$ and a p^{th} order linear predictor with prediction coefficients a_k.

(i) Show that the prediction error E can be written $E = \sum_{n=p}^{N-1}\left(s(n) - \sum_{k=1}^{p} a_k s(n-k)\right)^2$ and explain why the first summation begins at $n = p$. [2]

(ii) Show that the prediction coefficients that minimize E satisfy $\mathbf{Ra} = \mathbf{b}$ where the $(i,j)^{\text{th}}$ element of $\mathbf{R}$ is given by $r_{i,j} = \sum_{n=p}^{N-1} s(n-i)s(n-j)$, the i^{th} element of $\mathbf{b}$ is given by $b_i = r_{i,0}$ and $\mathbf{a} = [a_1 \quad a_2 \quad \cdots \quad a_p]^T$. [7]

(b)

(i) Discuss the differences between covariance LPC and autocorrelation LPC and state their relative advantages. [4]

(ii) A signal sampled at 8 kHz consists of three cosine waves at 800 Hz, 1 kHz and 1.2 kHz of relative amplitudes 1, 1 and 0.2 respectively, plus white noise. [7]

Four different 10th order LPC analyses are performed on the signal. These are, in no particular order:

- Covariance LPC with a frame length of 3.5 ms
- Covariance LPC with a frame length of 80 ms
- Autocorrelation LPC using a Hamming window of length 3.5 ms
- Autocorrelation LPC using a Hamming window of length 80 ms.

Identify which plot in Figure 7 corresponds with each of the four cases. Give reasons for your choice and explain the factors that cause the differences between the plots.

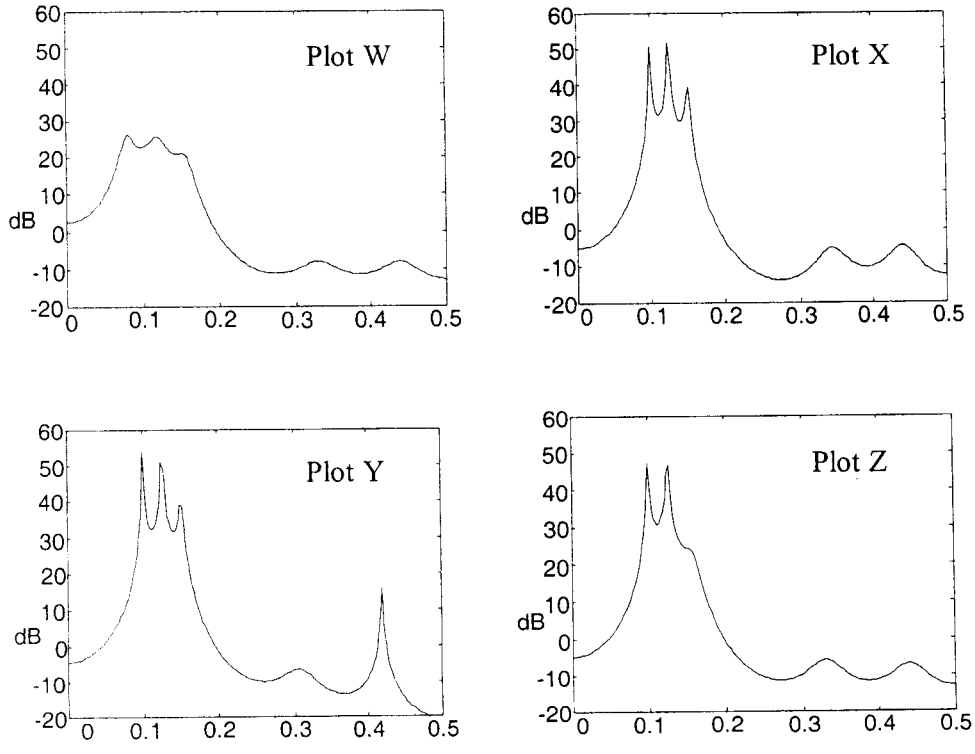
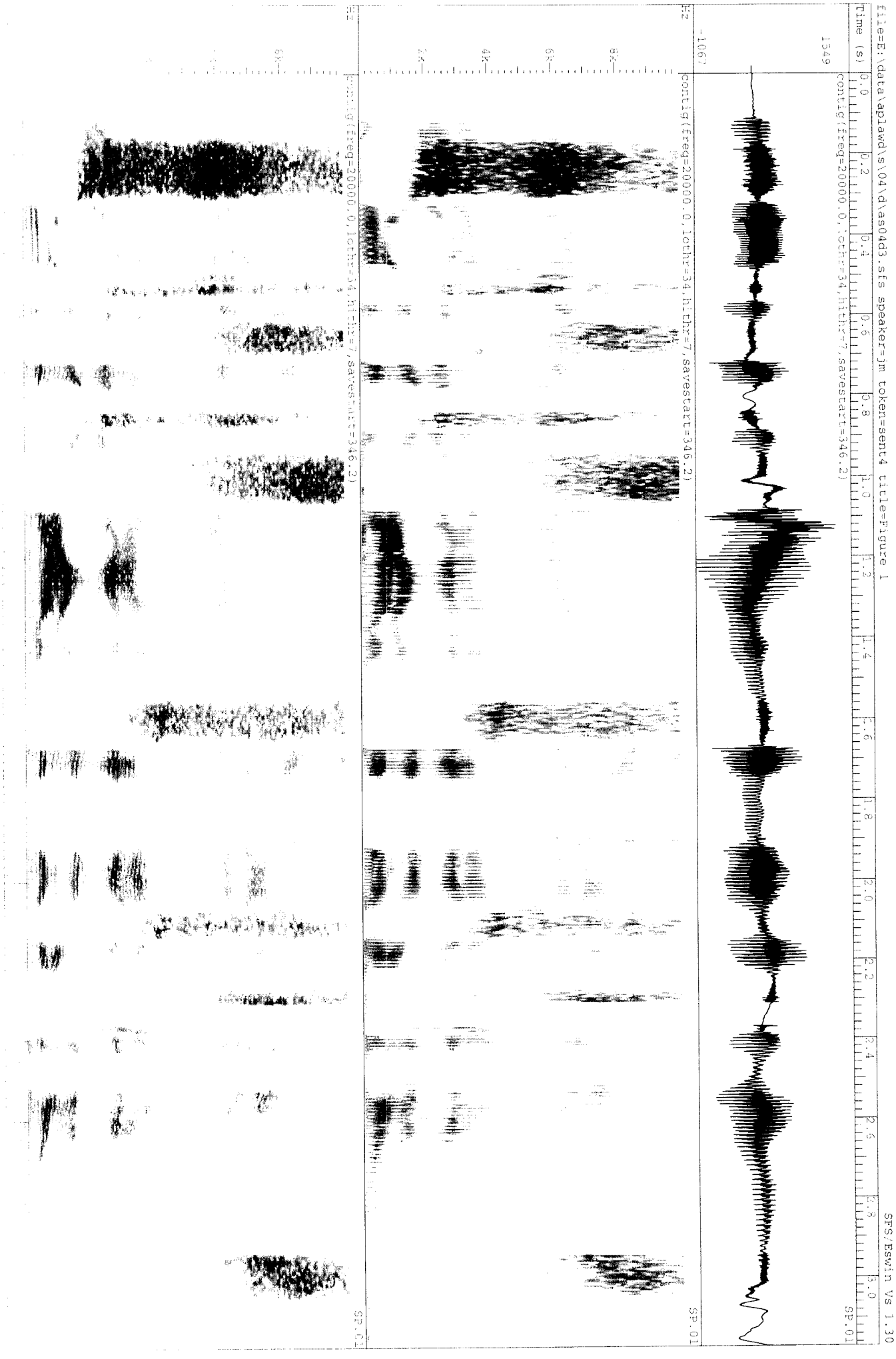


Figure 7

Figure 1


file=E:\data\aplawds\04\das04d3.sfs speaker=jm token=sent4 title=Figure 1
Time (s)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4
2.6
2.8
3.0
1549
-1067
contig(freq=20000.0,ithr=34,mthr=7,savestart=346.2)
SP.01
Hz
8k
6k
4k
2k
contig(freq=20000.0,ithr=34,mthr=7,savestart=346.2)
SP.02
SFS/Eswin Vs 1.30

Master 15/3/05

SPEECH PROCESSING 2005 - SOLUTIONS

1.

(a) The variables k and m represent the frequency bin index and the time frame index respectively. A spectrogram captures a time-frequency representation of the speech signal and facilitates the visualization of the time variations and frequency variations, importantly including formant tracks.

(b) The two common measures are a = -3 dB bandwidth and b = -∞ dB bandwidth. For example:
Rectangular Window: a=1.21, b=2
Hanning Window: a=1.65, b=4

The significant features of the window are best shown in the frequency domain. The level of the sidelobes should be labelled numerically.

(c) Frequency resolution: Equal amplitude frequency components with this separation will give distinct peaks
$= f_s a / N$

Time resolution: Amplitude variations with this period will be attenuated by 6 dB
$= 2N / af_s$

Time resolution x Frequency resolution = 2.

(d)

(i) These are wideband and narrowband spectrograms obtained from different choices of window size N.
Large N -> narrow band, good frequency resolution
Small N -> wide band, good time resolution.

(ii) Tongue is lifted to partially divide vocal track into two. Ratio of lengths of subdivided portions affects ratio of F1 to F2. Rear cavity determines F1: F1 decreases as the tongue hump moves forwards or is raised higher. Front cavity determines F2: F2 increases as the tongue hump moves forwards.

(iii) 1.96 s: First 2 formants at approximately 800 Hz, 1700 Hz
Relatively high frequency of F2 indicates Front placement.
Speech sound is
ɛ (E) as in "bed" F1=550 F2=1770

2.18 s: First 2 formats at approximately 600 Hz, 1000 Hz
Relatively low frequency of F2 indicates Read placement.
Speech sound is
ɒ (o) as in "body" F1=500 F2=900

Students are not expected to be able to 'read' the spectrogram accurately but marks will be awarded for well reasoned suggestions

D.M. Rich
Patrick Naylor

2. (a) The use of diphone segments means that segment junctions occur in the centre of each phoneme rather than at the phoneme bondaries. This means that we use a recorded example of each phoneme-to-phoneme transition rather than making an abrupt change from one to the next.

The centre of a phoneme is much less variable than the phoneme boundary and so switching from one recorded example to another at this point causes less of a discontinuity.

(b) (i)
For s(n) they occur every 80 samples: these include samples 2000, 2080 and 2160.

(ii)
We need to consider the phase difference of the pitch marks for the two cases: constant 100Hz pitch, pitch variation as specified.

Phase is given by integral of frequency.
(i) For constant 100 Hz pitch:

$$2\pi \int_{t=0}^{0.25} f_{pitch} dt = 2\pi \int_{t=0}^{0.25} (100)\, dt = 2\pi [100t]_0^{0.25} = 2\pi \times 25$$

(ii) For pitch variation:

$$2\pi \int_{t=0}^{0.25} f_{pitch} dt = 2\pi \int_{t=0}^{0.25} (100 + 100t)\, dt = 2\pi \left[100t + 50t^2\right]_0^{0.25} = 2\pi \times 28.125$$

The ratio is given by 28.125/25 = 1.125. The pitch marker therefore occurs at n = 2000 + 0.875*8000/125 = 2056 with the next following 64 samples later at 2120.

(iii) The output, y(n), around sample 2090 therefore consists of superposed waveform segments centred on the new pitch marks at 2056 and 2120 and taken from the nearest pitch segments in the original waveform, i.e. from pitch marks 2080 and 2160 (or 2080 since they are equidistant) respectively. The waveform segments must be windowed by multiplying by a Hanning window whose length is twice the minimum of the original pitch period (80 samples) and the new pitch period (64 samples).

(iv) At sample 2090, we are 34 samples after the pitch mark at 2056 and 30 samples before the one at 2120. The output is therefore:

$$y(2090) = \frac{x(2080+34)w(34/64) + x(2160-30)w(30/64)}{\sqrt{w^2(34/64) + w^2(30/64)}}$$
$$= 0.635x(2114) + 0.773x(2130)$$

where the window is defined by
$w(x) = \frac{1}{2}(1 + \cos(\pi x)) = \cos^2(\frac{1}{2}\pi x)$
The normalising factor in the denominator corrects for the windowing under the assumption that the phases in the segments are uncorrelated and that their powers will therefore add.

3.
(d)

u	1	1	1	10	10	10	1	1
k	1.8	1.98	2.18	1.96	5.88	17.6	19.4	17.5
v=u/k	0.56	0.51	0.46	5.1	1.7	0.57	0.052	0.057
w	1	1	0	2	2	1	0	0
x=w*k	1.8	1.98	0	3.92	11.76	17.6	0	0

(a)

$$\int_{-½w}^{+½w} x^2 \frac{dx}{w} = \left[\frac{x^3}{3w}\right]_{-½w}^{+½w} = \frac{w^2}{12} \quad \Rightarrow \text{rms error} = 0.289w$$

(b)
Uniform: bin centres are uniformly spread over the range of values to be quantized. Quantization error is minimized if signal is uniformly distributed on the input range of the device.

Non-uniform: bin centres are typical chosen to match the (non-uniform) pdf of the input signal. Speech has non-uniform pdf (super-Gaussian) and therefore lower quantization error can be obtained using non-uniform quantization.

Quantization error $q(s) = s_i - s$

Mean square quantization error is $E = \int_{-\infty}^{+\infty} p(s)q^2(s)ds = \sum_{i=1}^{N} \int_{a_{i-1}}^{a_i} p(s)(s_i - s)^2\, ds$

Differentiating gives: $\frac{\partial E}{\partial s_i} = \int_{a_{i-1}}^{a_i} -2p(s)(s - s_i)\, ds = 2s_i \int_{a_{i-1}}^{a_i} p(s)ds - 2\int_{a_{i-1}}^{a_i} sp(s)ds$

Setting to zero gives: $s_i = \frac{\int_{a_{i-1}}^{a_i} sp(s)ds}{\int_{a_{i-1}}^{a_i} p(s)ds}$

c)
Non-uniform scheme. Wider bins at high amplitudes causes more quantization noise power, but this is to some extent balanced by the higher signal power. Hence the Signal to Quantization Noise Ratio is approximately constant with signal amplitude. For signals with sharp (super-gaussian) pdfs, amplitudes are small for most samples (and hence small quantization noise is introduced).
Bin centres at ±{(*m*+16½) 2^e – 16½}
Bin widths: 2^e

Max Quantization level for 8 bit (3-bit exponent, 4 bit mantissa) = (15+16½).2^7 – 16½ = 4015.5.

Nearest quantized value to 27 is 30 giving an error of 3. Nearest quantized value to 1027 is 1039.5 giving and error of 12.5.

4. (a)

(i) 30 means + (30×31)/2 covariance elements = 495 parameters
(ii) 30 means + 30 variances = 60 parameters
(iii) (30 means +30 variances + 1 weight) × 10 mixtures –1 = 609 parameters. The –1 arises because the weights must sum to unity.

The use of gaussians allows a very simple formula for log probability. A mixture of gaussians provides two benefits over a single gaussian: (i) it can model the tails of the true distribution which do not fall to zero as quickly as a gaussian, (ii) it can model the multimodal distribution that arises when a single phoneme can be pronounced in several different ways.

(b)

If frame t–1 in states 1, ..., K then, since the transition probabilities are independent of the initial state, the probability that state t is also in states 1, ..., K is given by $p\sum_{i=1}^{K} w_i d_i(\mathbf{x}_t)$.

Hence the probability for the entire sequence is

$$\sum_{i=1}^{K} w_i d_i(\mathbf{x}_1) \times \prod_{t=2}^{T}\left(p\sum_{i=1}^{K} w_i d_i(\mathbf{x}_t)\right) \times (1-p) d_B(\mathbf{x}_{T+1})$$

$$= p^{T-1}(1-p) d_B(\mathbf{x}_{T+1}) \prod_{t=1}^{T}\sum_{i=1}^{K} w_i d_i(\mathbf{x}_t)$$

$$= p^{T-1}(1-p) d_B(\mathbf{x}_{T+1}) \prod_{t=1}^{T} d(\mathbf{x}_t)$$

where $d(\mathbf{x}_t) = \sum_{i=1}^{K} w_i d_i(\mathbf{x}_t)$ is the gaussian mixture distribution.

(c)

$$m_i \leftarrow \frac{\sum_{t=1}^{T} A_{i,t} x_t}{\sum_{t=1}^{T} A_{i,t}}, \quad C_i \leftarrow \frac{\sum_{t=1}^{T} A_{i,t} x_t^2}{\sum_{t=1}^{T} A_{i,t}} - m_i^2, \quad w_i \leftarrow \frac{\sum_{t=1}^{T} A_{i,t}}{T} \quad \text{where} \quad A_{i,t} = \frac{d_i(x_t)}{\sum_{k=1}^{K} d_k(x_t)}$$

$A_{i,t}$ is the conditional probability that frame t belongs to state i given that it belongs to one of states 1, ..., K. It is the fraction of frame t that is assigned to state i.

5.
(a)
Total volume flow = u–v
Total acoustic pressure = (u+v) × \rho × c / A

Assumptions
Sound waves are 1-dimensional: true for frequencies < 3 kHz whose wavelengths are long compared to the tube width
No frictional or wall-vibration energy losses

(b)
The model is less good for if the sound has two points of excitation, (e.g. a voiced fricatives) or if the velum is lowered resulting in a branching tube (e.g. nasalized consonants).

(c)
Conservation of mass: $B - C = D - E + F - G$
Pressure Equality: $(B + C)/P = (D + E)/Q = (F + G)/R$

We have three equations above plus a fourth: $G = FH$ and we want to eliminate both F and G. For convenience we define: $q = Q/P$, $r = R/P$ and $k = r\frac{1-H}{1+H}$.
Using $G = FH$ to eliminate G gives: $B - C = D - E + F(1 - H)$ and also
$B + C = (D + E)/q = (1 + H)F/r$.

The outer components of the last equation give $F = r\frac{B+C}{1+H}$ which we can substitute in the mass equation to give $B - C = D - E + r(B + C)\frac{1-H}{1+H} = D - E + k(B + C)$

We thus have the two equations: $\begin{aligned} D + E &= qB + qC \\ D - E &= (1-k)B - (1+k)C \end{aligned}$ from which we get

$$\begin{pmatrix} D \\ E \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1-k+q & -1-k+q \\ -1+k+q & 1+k+q \end{pmatrix}\begin{pmatrix} B \\ C \end{pmatrix}$$

Substituting in the original quantities gives:

$$\begin{pmatrix} D \\ E \end{pmatrix} = \frac{1}{2}\left\{\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} + \frac{Q}{P}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \frac{R(1-H)}{P(1+H)}\begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix}\right\} \times \begin{pmatrix} B \\ C \end{pmatrix}$$

6. (a)
(i)

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{i=1}^{p} a_i s(n-i)$$

$E = \sum_n e^2(n)$ which for a window length of N-p gives the range of summation from n=p to N-1.

(ii)

$$-\tfrac{1}{2}\frac{\partial E}{\partial a_i} = \sum_n \left(s(n) - \sum_{k=1}^{p} a_k s(n-k) \right) s(n-i)$$

$$= \sum_n s(n-i)s(n) - \sum_{k=1}^{p} a_k \sum_n s(n-i)s(n-k)$$

$$= r_{i,0} - \sum_{k=1}^{p} r_{i,k} a_k$$

Setting these partial derivatives to zero gives:

$\sum_{k=1}^{p} r_{i,k} a_k = r_{i,0}$ for $i = 1, \ldots, p$

or in matrix form **Ra** = **b**.

(b)
(i)
Covariance LPC performs the summation of part (a) over a finite window [0,N–1]. No windowing of the signal is involved so there is no compromise between time and frequency resolution: we can get infinite frequency resolution with small data windows provided there is no noise. The window length, N, must be greater than the filter order, p, and in practice should be twice as long. The resultant filter is not guaranteed to be stable. Solving the equation for a requires order p^3 operations.

For autocorrelation LPC, the input signal is first multiplied by a window, e.g. a Hamming window, that tapers to zero (or near zero) at its ends. This imposes a tradeoff between time and frequency resolution: to obtain adequate frequency resolution for speech (100 Hz), the window length must be at least 20 ms. The summation of part (a) is then performed over n=-inf to inf, although all but a finite number of terms are zero. The resultant R matrix is toeplitz and the coefficients a can be found using the Levinson-Durbin algorithm with order p^2 operations. The filter will always be stable.

(ii)
The 1200 Hz sinewave should have an amplitude 14 dB less than the other two tones.
Plot W has the poorest frequency resolution and broadest spectral peaks and is therefore autocorrelation LPC with a window length of 3.5 ms. This gives a frequency resolution of about 2/3.5 kHz = 571 Hz.
Plot Z is better, but the 1200 Hz sinewave has been almost overwhelmed by the sidelobes of the 1 kHz sinwave. This is therefore autocorrelation LPC with a window length of 80 ms.
Plot Y has accurately found the three sinewaves although their amplitudes are not quite right. It has also generated a spurious peak at around 3300 Hz. This is covariance LPC with a window length of 3.5 ms (28 samples) and the white noise has resulted in the false peak.
Finally Plot X is covariance LPC with a window length of 80 ms. The sinewaves have been accurately modelled and the remaining four poles have been used to create a noise floor at –10 dB.