1. What is a phoneme? Minimal unit of sound

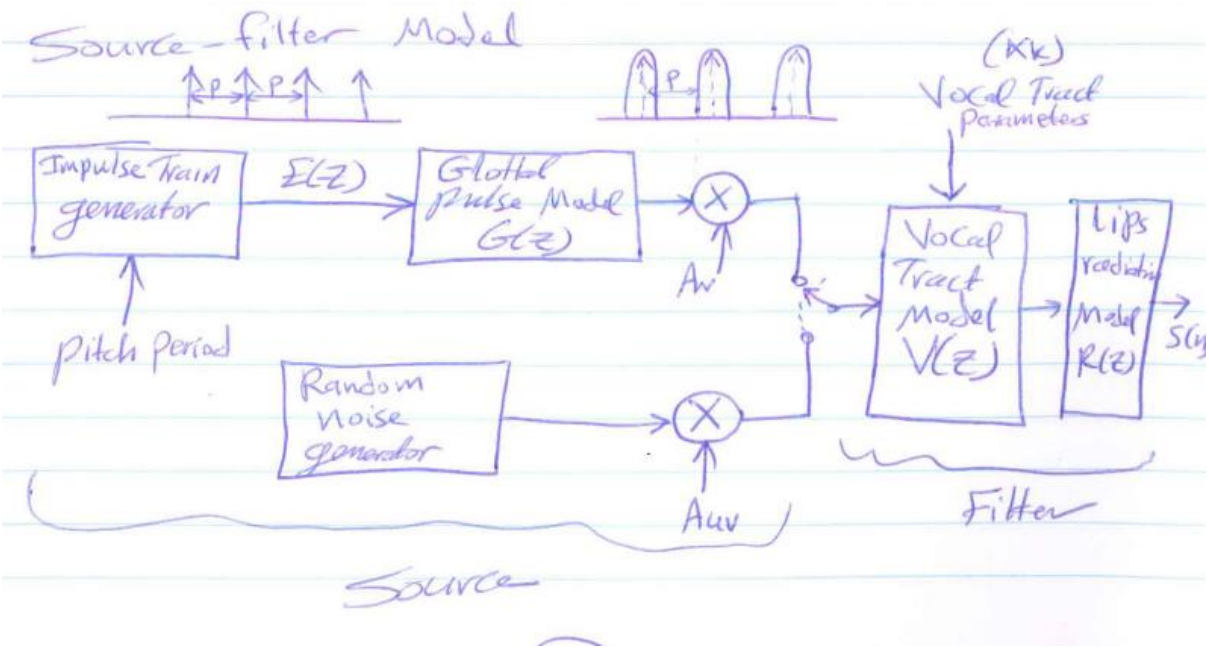
2. What is the difference between the phonemes /b/ and /p/? 'b' is voiced, 'p' is unvoiced

3. What is the name given to the class of phonemes which includes /s/, /f/ and /sh/ (as in /sheet/)
unvoiced, fricatives, consonants

4. What is the name given to the class of phonemes which includes /p/, /t/ and /k/?
unvoiced, stops, consonants

5. What is the source-filter model of speech production? Draw a diagram.

(a) Draw a block diagram for source-filter model of speech production and explain its correspondence to the human speech production process. [3pts]



* Impulse Train generator corresponds to Vocal folds Vibration with frequency of f_0 (fundamental) For Voiced Sounds

* Glottal Model: Corresponds to "opening/closing" of the Vocal folds in the larynx which shapes the impulses to be like pulses (puffs)

* For un-Voiced Sounds (eg. /sh/) the output of the larynx is noise-like signal. This is modeled by a random noise generator.

* In general, Voiced Sounds have greater amplitude than un-Voiced. This is represented by Multiplying output of larynx, by two constants A_V (Voiced) and A_{UV} (unvoiced)
$A_V > A_{UV}$.

* Vocal Tract is modeled by a filter with sufficient number of poles (All-pole filter). Usually 10-14 poles for both Voiced and un-Voiced Sounds.

* lip radiation is modeled by a high-pass filter of a transfer function $R(z)$.

(b) Give an equation for each block and for the overall speech production model in z-domain. [3pts]

$$E(z) = \frac{1}{1 - z^{-p}} \qquad G(z) = \frac{1}{(1 - z^{-1})^2} \qquad \text{F}$$

Vocal Tract model:

$$V(z) = \begin{cases} \dfrac{1}{1 + \sum_{k=1}^{P} a_k z^{-k}}, & \text{for Voiced Sound} \\ \dfrac{1}{1 + \sum_{k=1}^{P+2L} a_k z^{-k}}, & \text{for un-Voiced Sound} \end{cases}$$

P = no. of Poles
L = no. of Zeros.

OR

$$V(z) = \frac{1}{1 + \sum_{k=1}^{p} a_k' z^{-k}}, \quad p \tilde{=} 10-14$$ * For Both Voiced and Un-voiced Speech.

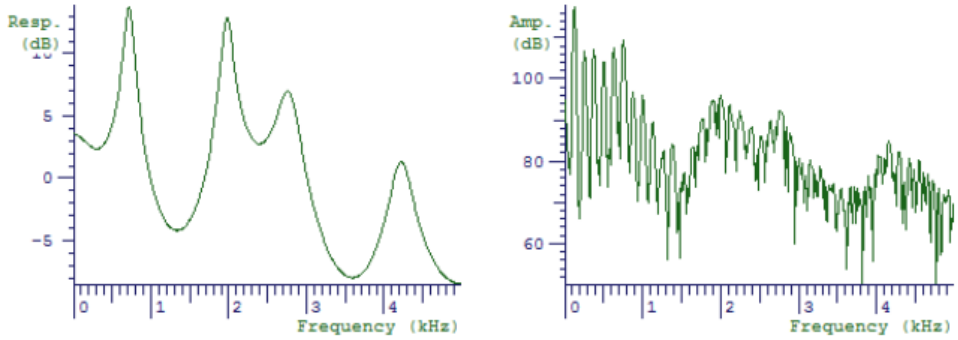
Lips radiation:

$$R(z) = 1 - z^{-1}, \text{ first-order HPF.}$$

* overall

$$\frac{S(z)}{E(z)} = \frac{G}{1 + \sum_{k=1}^{q} a_k' z^{-k}}, \quad 10 \le q \le 14$$

(c) The figure below shows the 'filter response' [left] and 'output spectrum' [right] for a short speech frame. What are these two graphs? Explain how the right-hand graph is related to the left-hand graph? Is this voiced or un-voiced frame? Explain? [4pts]



(c) Right-hand graph is spectrum of speech frame
Left-hand graph is the response of the vocal tract

* Left-hand graph (Vocal tract) is the envelop of the speech spectrum, or low frequency components of the speech spectrum.

Source $e(n)$ → Vocal Tract $v(n)$ → Speech signal $s(n)$

$$s(n) = e(n) * v(n)$$
$$S(z) = E(z) \cdot V(z)$$

* This looks like a voiced sound because the power is higher in the low-frequency components i.e. power at F_1 > power at F_2 > ...

(d) Mention two benefits of applying discrete cosine transform (DCT) in the Mel-frequency Cepstral coefficient (MFCC) feature extraction technique? [2pts]

1- **Uncorrelated features**
2- **Separate source signal from the vocal tract filter response by truncating DCT coefficients**

Question3:[10 marks]

(a) Given the following windowed speech segment,

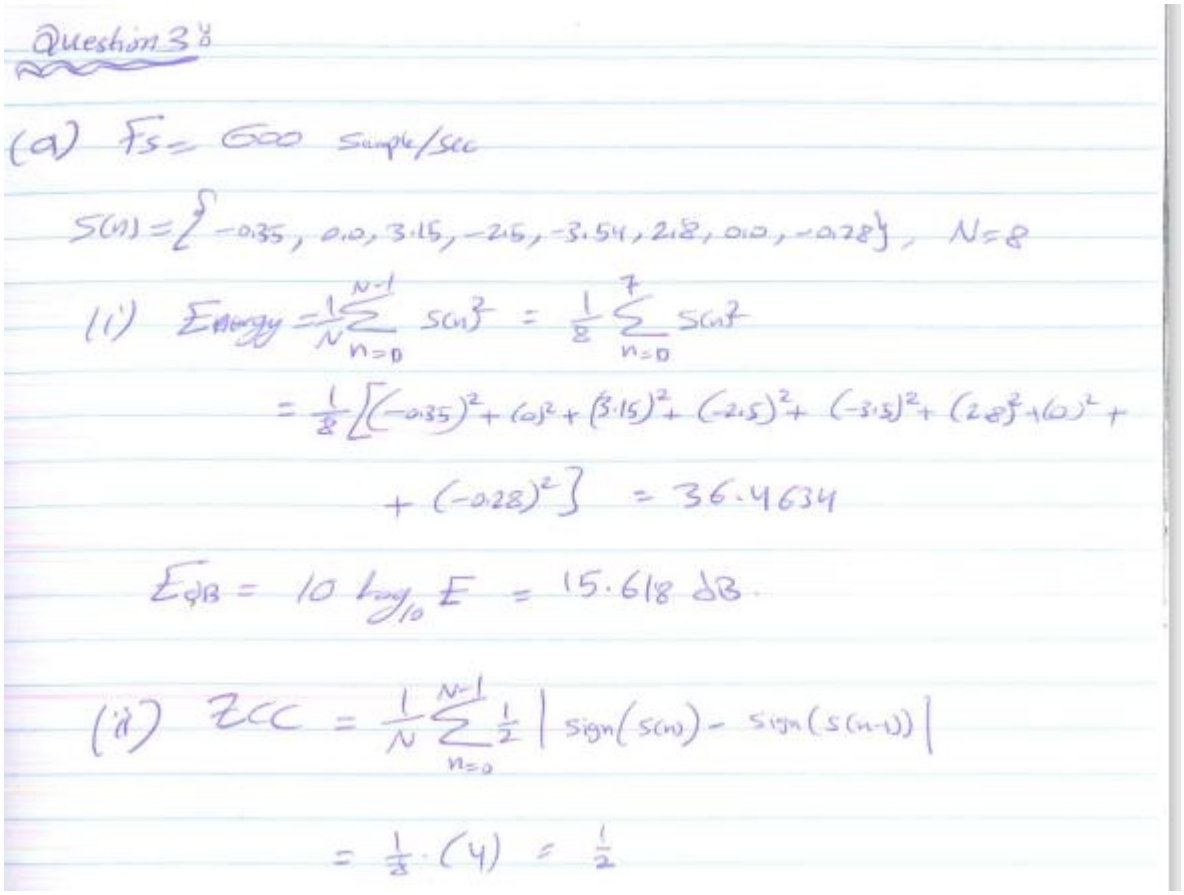
S (n) = [-0.35, 0.0, 3.15, -2.5, -3.54, 2.8, 0.0, -0.28], with sampling frequency of **600** sample/sec.

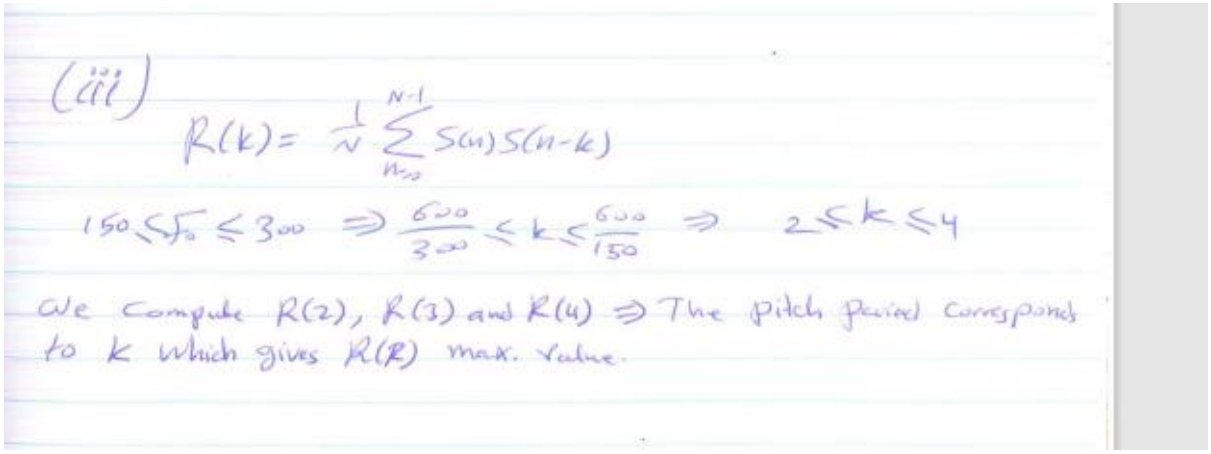
Find the following basic features (show the equation for calculating each one):

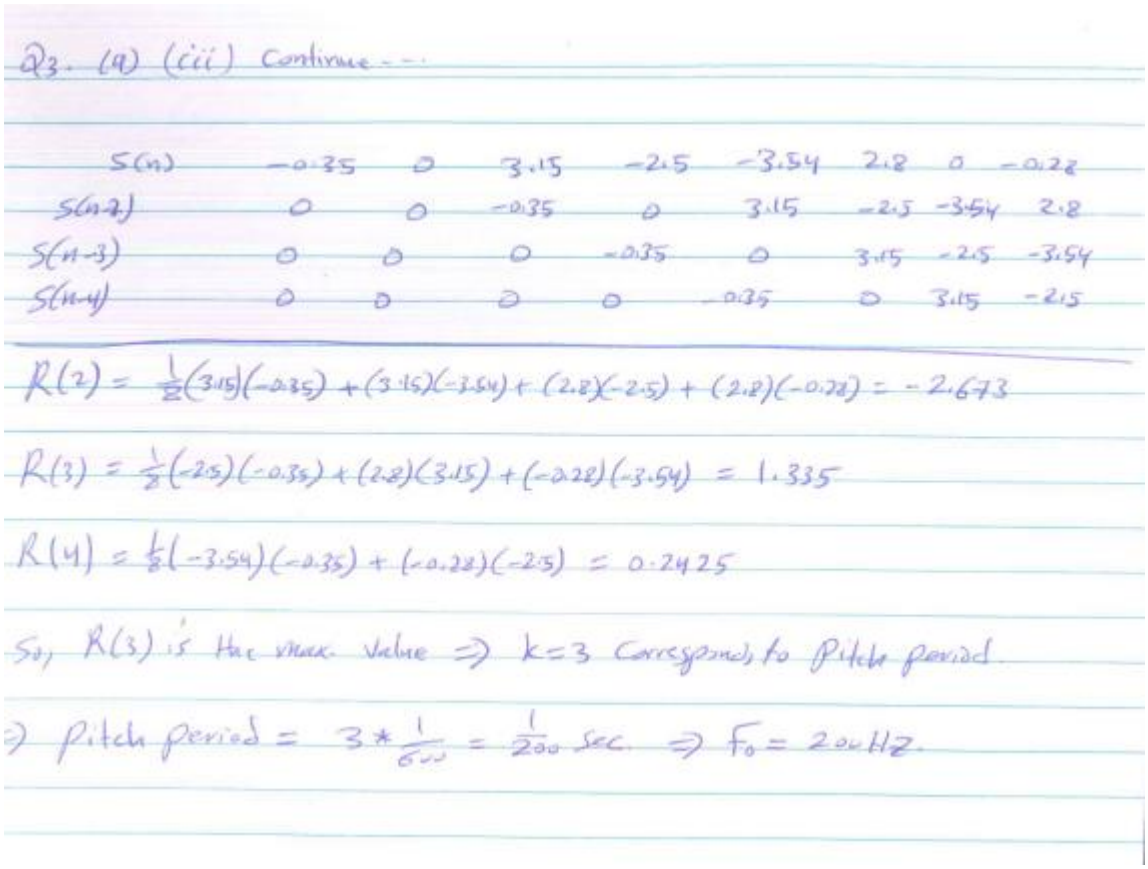
(i) Energy in decibels (dB). [1pt]

(ii) Zero-crossing rate [1pt]

(iii) Use autocorrelation method to find Pitch period **T**, if we assume the fundamental frequency (F_0) is in the range 150-300Hz. [2pts]







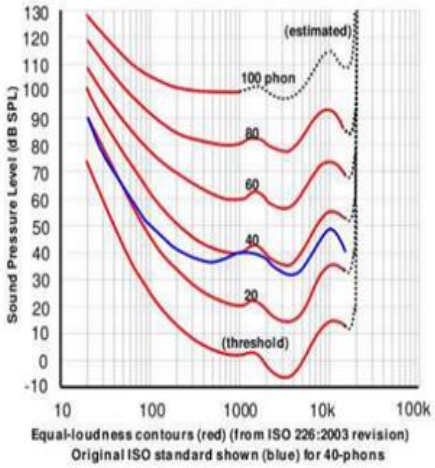
(b) Explain how the basic features, in part (b), can be used for voiced/unvoiced classification of the speech segment? Is the speech segment given in (b) above voiced or unvoiced? [3pts]

If the energy was high, the ZCR is low and there is f0 → **voiced**

If the energy was low, the ZCR is high and f0 = zero → **unvoiced**

This means the signal above is voiced

(c) Use the equal loudness curves shown in the following figure to answer the following questions. The line labeled "threshold" is the line below which humans typically cannot hear.



(1) Can humans typically hear a 100 Hz sound at 10 DB SPL? [1pt]

no because the min. hearing at 100Hz = 25dB

(2) Which sounds louder: a tone of 100 Hz at 25 DB SPL or a tone of 1000 Hz at 20 DB SPL? [1pt]

1000hz at 20dB, because human ear is more sensitve to 1000hz than 100hz

(3) According to the graph above, which frequency range best corresponds to the range at which humans are most sensitive to loudness? [1pt]

A. 3kHz to 10kHz
B. 2kHz to 5kHz
C. 500 Hz to 1000 Hz

B.

6. What is a formant?

Formants: are number of resonant frequencies that forms the vocal tract response.

Ideal vocal tract has 3-4 formants.

(c) What is spectrogram? Compare between narrow and wide –band spectrograms in term of window length, frequency resolution and temporal resolution? [4pts]

(c) Spectrogram: Time-frequency representation of speech signal
It is a Concatenated spectrums of short frames (10-20 ms) and shows the energy distribution over frequency and Time scale.

	Window Length	Freq. resolution	Temporal Resolution
Narrow-Band	Wide	Fine	Poor
Wide-Band	Narrow	Poor	Fine

Question 1: [11 marks]

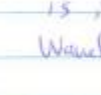
(a) Two of the criteria used for classification of speech sounds are 'manner of articulation' (i.e. how the sound is made) and 'type of excitation'. Give the category name to which the phonemes **/K/, /SH/, /AH/**, and **/M/** belong to according to each criteria. [4pts]

Question 1:
(a)
/K/ plosive, un-Voiced
/SH/ Fricative, un-Voiced
/AH/ Vowel, Voiced
/M/ Nasal, Voiced

(b) Describe, in general term, what the waveforms look like for speech sounds belonging to the same category as phonemes **/K/, /SH/** and **/AH/.** [3pts]

(b)

* plosives (stops) e.g /k/ : Very little power (or no power) followed by sudden explosion and aspiration when contriction is released.

Waveform e.g. 

* Fricatives (e.g. /sh/): Non-periodic sound (friction) - random energy across wide freq. range

Waveform e.g. 

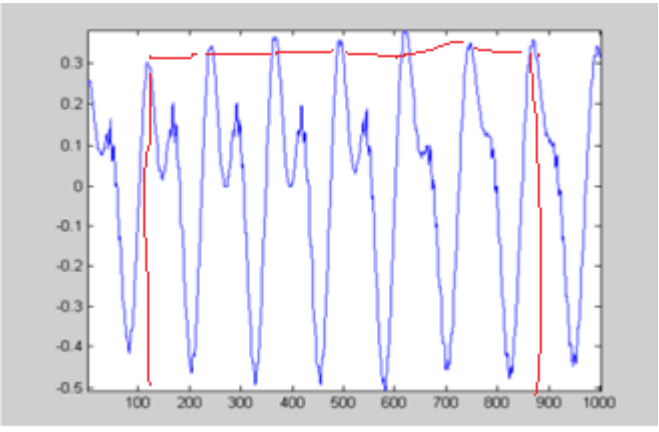
* Vowels (e.g /AH/): periodic waveform with fundamental frequency corresponding to vibration rate of vocal-folds. Vocal Tract enhances some frequencies (Formants) and suppresses others. Energy of vowels sounds is relativily higher than other sounds.

Waveform e.g. 

8. Given a segment of speech waveform, how can you estimate the fundamental frequency?

It is hard to find the fundamental frequency if the speech was unvoiced, if it was voiced speech (periodic), then I can count on how many complete periods are in the speech then get the find the interval, divide the interval by the period counts, this gives you pitch period → f0 = 1/T.

This minimize the error.



(880-120)/6 = T

1/T = f

Question 1: [20 marks]

(a) Choose the correct answer: [12pts]

1- A speech signal was recorded at 16 kHz. The length of the window for acoustic analysis is set to 30ms. The length of the window in samples is

A – 160 samples

B – 440 samples

C- 480 samples

D- 320 samples

2. Which of the following digital signal processing techniques is used to estimate the pitch frequency (F0)?

A. Average Magnitude Difference Function (AMDF).

B. Linear Predictive Coding (LPC)

C. Cross-correlation

D. Fast Fourier Transform (FFT)

3. Which of the following digital signal processing techniques is used to construct spectrograms?

A. RMS Amplitude

B. Short-time energy.

C. Autocorrelation

D. Fast Fourier Transform (FFT)

4. When conducting Fast Fourier Transforms, longer window sizes, as compared to shorter window sizes, result in

A. discontinuous breaks at the edges of windows.

B. better amplitude resolution, but worse time resolution.

C. better time resolution, but worse frequency resolution.

D. better frequency resolution, but worse time resolution.

5. Sound A is a tone played at 1200 Hz, sound B is played at 2000Hz, sound C is played at 12kHz, sound D is played at 13kHz. All are played at 60dB. Which sounds like a greater increase in pitch?

A. The increase from Sound A to Sound B.

B. The increase from Sound C to Sound D.

C. They pitch increase from Sound A to Sound B is perceived as the same as the pitch increase from Sound C to Sound D.

6. What kinds of speech sounds can easily be identified by the shapes of their formants?

A. Consonants.

B. Vowels.

C. Syllables.

D. All phonemes.

7. Which is NOT an example of variability in the speech signal?

A. The lack of boundaries between words and phonemes

B. People with different sized/shaped vocal tracts

C. People talking in noisy environments

D. Dialect /accent differences

8. From the point of view of speech PERCEPTION, coarticulation of speech sounds helps us to:

A. figure out words quickly

B. speak faster

C. create phonemic categories quickly

D. perceive F1 and F2 quickly

9. A whispered sound may be said to be _____ .

A. voiced

B. unvoiced

C. stopped

D. a fricative

10. Normal speech has an intensity of around _____.

A. 30-35 dB

B. 40-50 dB

C. 55-65 dB

D. 75-85 dB

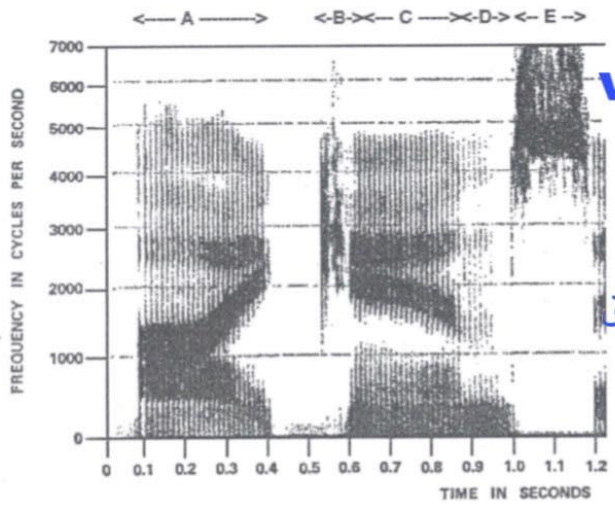
Question 1 answers:

voiced غامق من تحت فاتح من فوق

unvoiced غير هيك

Question 2: [14 marks]

The figure below shows a spectrogram od a speech segment.



vowel ڤويسد عريض

nasal ڤويسد رفيع

فاضي من تحت غامق من فوق

consonent

(i) Estimate the bandwidth of this signal in Hz, as accurate as you can? [4pts]

from the spectrogram:
B.W is 7000Hz or 7kHz

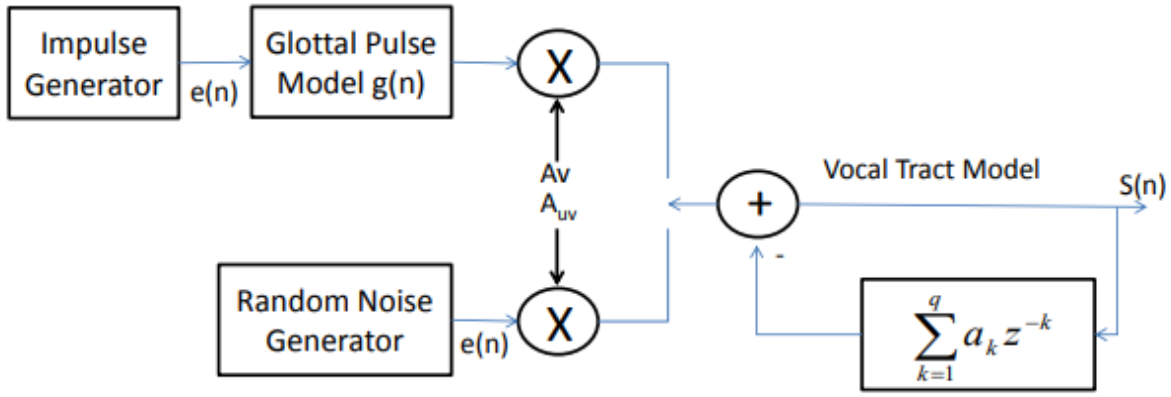
(ii) for each of the identified areas in the figure, A, B, C, and D as (voiced/unvoiced), and as (consonent, vowel or nasal):

	Voiced/unvoiced	consonant/vowel/nasal
(i) Area A:	Voiced	Vowel
(ii) Area B:	Unvoiced	Consonent
(iii) Area C:	Voiced	Vowel
(iv) Area D:	Voiced	Nasal
(v) Area E:	Unvoiced	Consonent

9. What is autocorrelation function and what we used it for?

R = 1/N Sum (0 to N-1) (s(n) * s(n-1)) , it is used to calculate the fundamental frequency of a given sample.

10. How do the glottal source spectrum and vocal tract transfer function combine to produce speech?



$$G(z) = \frac{1}{(1-z^{-1})^2} \qquad \text{F}$$

$$\frac{S(z)}{U(z)} = \frac{G}{1+\sum_{k=1}^{q} a_k z^{-k}}$$

11. What is short-term cepstrum? What information is carried by the lower cepstral coefficients?

- The **short-time spectrum at time *t***

  – From the perspective of the human speech production, it tells us about the shape of the vocal tract at time *t*

  – From the perspective of human speech perception, we know that a similar analysis is performed in the cochlea in the initial stages of human speech perception

12 MFCC features

13. What is a critical bandwidth?

- For a given frequency, the critical band is the smallest band of frequencies around it which activate the same part of the BM
  - Critical bandwidths correspond to about 1.5 mm spacing along the BM

14. What is the mel frequency scale?

Mel-scale

- Mel-scale is approximately linear below 1 kHz and logarithmic above 1 kHz

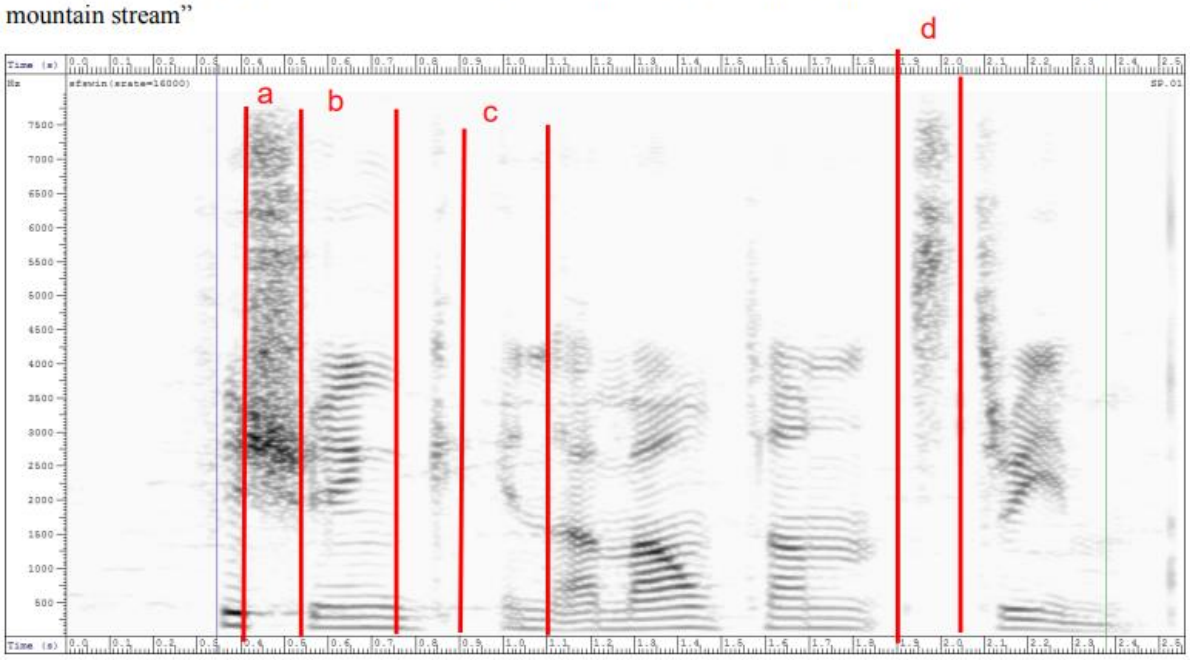
$$Mel(f) = 2595 \log_{10}(1 + \frac{f}{700})$$

15. How many mels equal 1kHz?

=2595 ln (1 + 1000/700) / ln(10)
=2595 ln (2.428) / ln (10)

= 999.998 mels
= 1000 mels

16. The figure below shows an SFS display of a speech spectrogram for an example of a phrase "fishing in a mountain stream"



(a) for each of the following start times and end times, identify the voicing classification (voice/unvoice) and the manner of articulation classification (fricatives, vowel and plosives).

(i) start time 0.4s, end time 0.54s fricative, unvoiced.

(ii) start time0.54s , end time 0.76s vowel, unvoiced

(iii) start time 0.9s, end time 1.1s plosive, voiced

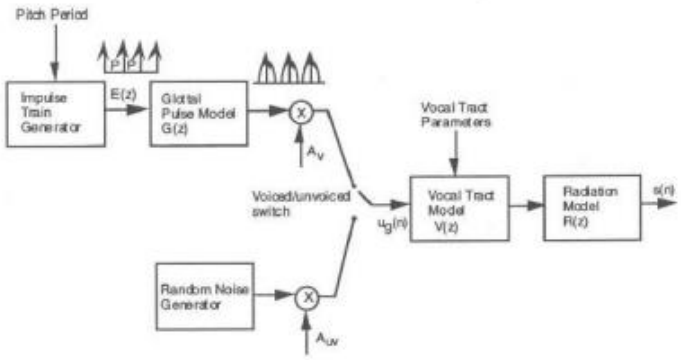
(iV) strst time 1.9s, end time 2.05s fricative, unvoiced

(b) What is meant by 'source-filter' model of speech production? Your answer should include a diagram.

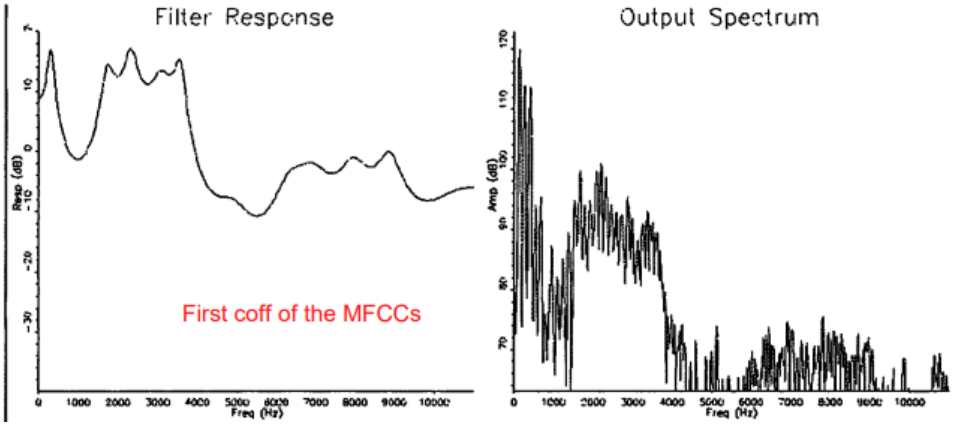
Voiced: starts with train of impulses, then glottal puffs, then multiply with high energy, vocal tract model.

Voiced: starts random noise generator, then multiply with low energy, vocal tract model.

Both of them then go to radiant model then to speech.



(c) the figure below shows the 'filter response' and 'output spectrum' at time 2.2s . what are these two graphs? Explain in detail how the right-hand graph is related to the left-hand grapg.



(c) Right-hand graph is spectrum of speech frame
Left-hand graph is the response of the vocal tract

* Left-hand graph (Vocal tract) is the envelop of the speech spectrum, or low frequency compones of the speech Spectrum.

Source → e(n) → Vocal Tract v(n) → Speech signal s(n)

$s(n) = e(n) * v(n)$

$S(z) = E(z) \cdot V(z)$

* This looks like a voiced sound because the power is highier in the low-frequency components i.e. power at F_1 > power at f_2 > ...

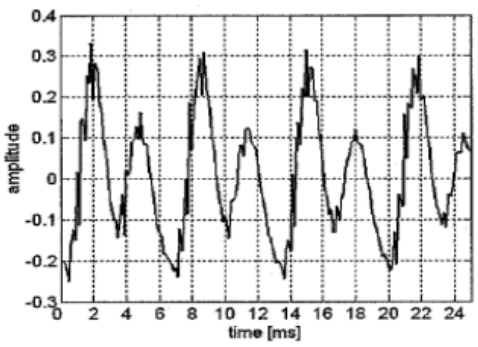
(d) What is the short-term cepstrum? Explain in detial how it can be used to recover the left-hand grapg from the right-hand graph in part (c).

- The **short-time spectrum at time *t***
  - From the perspective of the human speech production, it tells us about the shape of the vocal tract at time *t*

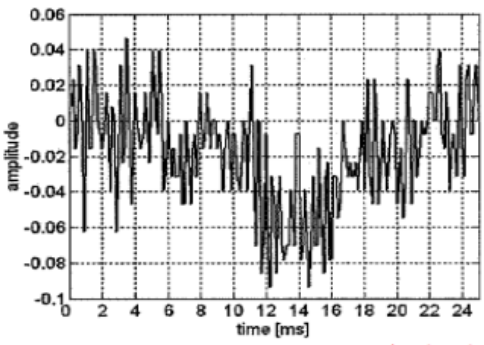
(e) What is meant by the term 'delta cepstrum', and what is the motivation for its use in automatic speech recognition.

- Also, we know that speech signal is not constant (slope of formants, change from stop burst to release).
- So we want to add the changes in features (the slopes).
- We call these **delta** features

(c) Figure 1 depicts the waveform of two speech sounds. The x-axis depicts time in [ms], the y-axis is the amplitude of the signal.



(a)



(b)

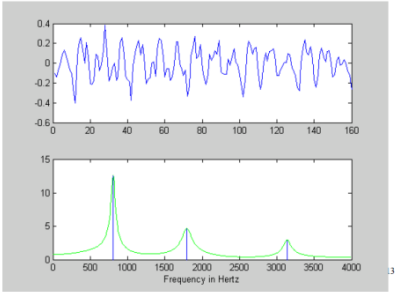
a)voiced,periodic,high power
b)unvoiced,non periodic, low power

(i) Classify the sounds based on the type of excitation (voiced/unvoiced)? Justify your answer.

(ii) Estimate the fundamental frequency of the voiced sound as accurately as you can? Explain your calculations? 15-9 ms = 6ms

(d) Draw a figure illustrating an example of the frequency characteristic of the vocal tract when producing vowel sound? Explain the figure? الفارمنتس بنزلو

Voiced Speech



(e) Consider the IPA classification of speech sounds based on manner of articulation (i.e. how sound is made). Which categories do the speech sounds /m/ and /p/ belongs to? For one of these sounds, explain how the sound is made? Give properties of the sound and give more examples of phonemes belong to the same category?

m=> nassle,consonant ==> n
p=>unvoiced,plosive ==> t

(a) Noise corruption in speech signals can be categorized into two general types. additive,convolutive

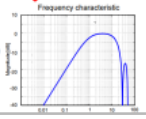
(i) Give their names and give an example illustrating how each type of noise might occur in practice.
conv from channel. Additive from the surrounding env

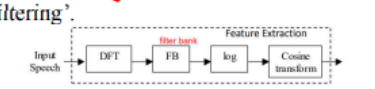
(ii) What is the effect of each type of noise at each stage of the front-end processing producing MFCCs? (your answer should include a block diagram of the front –end processing).

(iii) Describe the techniques called 'Cepstral Mean Subtraction' and '*Rasta filtering*'.

- Cepstral Mean Subtraction (CMS)
  - mean (over a num of frames) subtraction
  - lowpass filtering
  - eliminates communication channel spectral shaping

•Filtering log filter bank output (or equivalently cepstral) temporal trajectories by band pass filter
•Remove slow changes to compensate for the channel effect (≈CMS over 0.5 sec. sliding window)
•Remove fast changes (> 25Hz) likely not caused by speaker with limited ability to quickly change vocal tract configuration





Type of Noise	Time Domain	DFT	Filter Bank	log	Cosine Transform
Additive	+	+	– +	? ~ max	?
Convolutive	* conv	x	– x	– +	– +

(b) Given the following short speech segment,

S (n) = [3.4 -3.5 0.4 -2.2 1.2 -2.4 1.8 3.0 6.8 -0.8], with sampling frequency of **600** sample/sec.

Find the following basic features (with showing the equation for calculating each one):

(i) Energy [4] بتربع كل سامبل وبنجمع

(ii) Zero-crossing count [4] اختلاف بالاشارة

(iii) Pitch period **T**, if we assume the fundamental frequency (F_0) is in the range 100-300Hz. [9] F0 = Fs/P[in samples]

(c) Explain how the basic features, in part (b), can be used for voiced/unvoiced classification of the speech segment? [4] voiced: high energy, low ZCC
Unvoiced : low energy,high ZCC

22.

(a) Draw a block diagram for the Mel-frequency Cepstral Coefficients (MFCC) feature extraction? [10]

(b) Describe, briefly, the functionality of each block? [5]

(c) Mention two benefits of using Discrete Cosine Transform (DCT) in the MFCC feature extraction? [5]

-DCT produces highly uncorrelated features
-Since the logpower spectrum is real and symmetric, inverse DFT reduces to DCT(better in computation)

(a) Human ear consists of three main parts, give their names, and name two components in each part.

- The ear divided into three sections:
  - the outer
  - Middle
  - Inner ear

outer: pinna + auditory canal
middle: malleus, staped, incus
inner: colchlea, basilar membrane

(b) What Auditory masking means? What are the two type of masking? Explain briefly what is the difference between two types of masking?

Masking

- A phenomenon whereby the perception of a sound is obscured by the presence of another (i.e., the latter raises the threshold of the former)
- Masking is the major non-linear phenomenon that prevents treating the perception of speech sounds as a summation of responses

Two types of masking phenomena

- Frequency masking
  - A lower frequency sound generally masks a higher frequency one
  - Leads to the concept of critical bands (next)
- Temporal masking
  - Sounds delayed wrt one another can cause masking of either sound
  - Pre-masking tends to last 5ms; post-masking can last up to 50-300ms

(c) What is meant by intra-band and inter-band masking? Support your answer by examples.

- Signal components within a given critical band can be masked by other components within the same critical band.
- This is called **intra-band** masking.

- In addition, sounds on one critical band can mask sounds in different critical bands.
- This is called **inter-band** masking.

(d) what is meant by post-masking and pre-masking? Give examples?

- When the signal proceeds the masker in time, the condition is called post-masking; when the signal follows the masker in time, the condition is pre-masking,

Q: Two methods to find the Pitch frequency.

- The two most commonly used techniques are:
  - Short Time Autocorrelation Function
  - Average Magnitude Difference Function (AMDF)

Pattern Recognition

- Typically we have:
  - A set of **classes** $C_1, \ldots, C_K$, each class characterised by a model
  - A sequence of feature vectors $Y = y_1, \ldots, y_T$
  - The classifier computes probability $P(Y|C_k)$ that the class C_k is the correct explanation of Y
- Classes in our case
  - Speech, Music, (other)
- What model we should use to describe each class?

Basic Parameter Extraction

- There are a number of very basic speech parameters which can be easily calculated for use, in simple applications:
  - Short Time Energy
  - Short Time Zero Cross Count (ZCC)
  - Pitch Period
- All of the above parameters are typically estimated for frames of speech between 10 and 20 ms long

Short Time Energy

- The short-time energy of speech may be computed by dividing the speech signal into frames of N samples and computing the total squared values of the signal samples in each frame.

Note that a recursive lowpass filter H(z) can also be used to calculate the short-time energy:

$$H(z) = \frac{1}{1 - az^{-1}} \qquad 0 < a < 1$$

Uses of Energy and ZCC

- Short Time Energy and ZCC can form the basis for :
  - Automated speech "end point" detection
    - Needs to be able to operate with background noise
    - Needs to be able to ignore "short" background noises and intra-word silences (temporal aspects)
  - Voiced\Unvoiced speech detection
    - High Energy + Low ZCC – Voiced Speech
    - Low Energy + High ZCC – Unvoiced Speech
  - Parameters on which simple speech recognition\speaker verification\identification systems could be based

Short Time Zero Crossing Count

- The Short Time ZCC is calculated for a block of N samples of speech as

$$ZCC_i = \sum_{k=1}^{N-1} 0.5 \,|\, sign(s[k]) - sign(s[k-1]) \,|$$

- The ZCC essentially counts how many times the signal crosses the time axis during the frame
  - It "reflects" the frequency content of the frame of speech
    - High ZCC implies high frequency
- It is essential that any constant DC offset is removed from the signal prior to ZCC calculation

Time Domain Methods

- Since pitch frequency is typically less then 600-700 Hz, the speech signals are first low passed filtered to remove components above this frequency range
- The two most commonly used techniques are:
  - Short Time Autocorrelation Function
  - Average Magnitude Difference Function (AMDF)
- During voiced speech, the speech signal is "quasi-periodic"
- Either technique attempts to determine the period (in samples between "repetitions" of the voiced speech signal

Pre-emphasis

- The high-pass filtering function can be achieved by use of the following difference equation:

$$y(n) = s(n) - as(n-1)$$

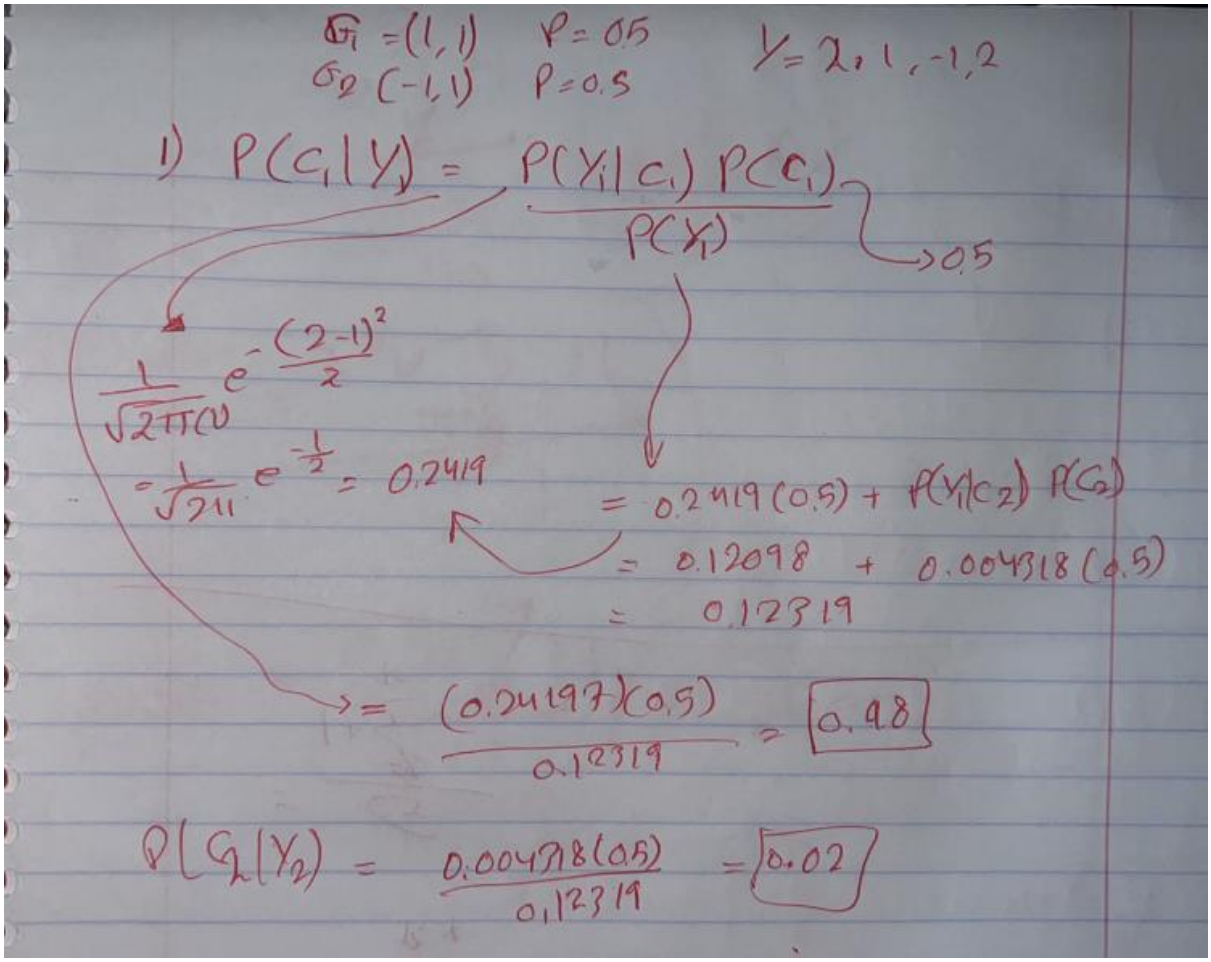
- Normally a is chosen between 0.9 and 1

(8) Consider a sequence of one-dimensional data Y = y_1, y_2, y_3, y_4 be modeled by a two-component GMM. Starting from the following initial GMM parameters and training data Y = 2,1,-1,-2 .

P(Y|GMM component 1)= N(1,1) , P(Y|GMM component 2)= N(-1,1), P(C1)=P(C2)=0.5.

(i) Fill in the following table by calculating the posterior probabilities of each Gaussian component given training data points Y.

	y_1	y_2	y_3	y_4
$P(C_1\|Y)$				
$P(C_2\|Y)$				



This is two points we continue to do the same for all the table as shown above

	y_1	y_2	y_3	y_4
$P(C_1\|Y)$	0.98	0.88	0.12	0.02
$P(C_2\|Y)$	0.02	0.12	0.88	0.98

(ii) Use EM algorithm to re-estimate the GMM parameters (one iteration). Hint: use posterior probability matrix in part (i) above.

1st gaussian (old values P(W) = 0.5, mean = 1, Variance = 1)

New P(W) = $\frac{(0.98+0.88+0.12+0.02)}{4} = 0.5$

new Mean = $\frac{0.98(2)+0.88(1)+0.12(-1)+0.02(-2)}{0.98+0.88+0.12+0.02} = \frac{2.68}{2} = 1.34$

Variance = $\frac{0.98(2-1.34)^2+0.88(1-1.34)^2+0.12(-1-1.34)^2+0.02(-2-1.34)^2}{0.98+0.88+0.12+0.02} = \frac{1.4087}{2} = 0.7$

2nd gaussian (old values P(W) = 0.5, mean = -1, Variance = 1)

New P(W) = $\frac{(0.02+0.12+0.88+0.98)}{4} = 0.5$

new Mean = $\frac{0.02(2)+0.12(1)+0.88(-1)+0.98(-2)}{0.02+0.12+0.88+0.98} = \frac{-2.68}{2} = -1.34$

Variance = $\frac{0.02(2+1.34)^2+0.12(1+1.34)^2+0.88(-1+1.34)^2+0.98(-2+1.34)^2}{0.02+0.12+0.88+0.98} = \frac{1.4087}{2} = 0.7$

In a simple speaker verification system, a speaker S is represented by a 4-component Gaussian Mixture Model given by:

$$P(y_t) = 0.2G_1(y_t) + 0.3G_2(y_t) + 0.1G_3(y_t) + 0.4G_4(y_t)$$

Where G_1, G_2, G_3, G_4, are all multivariate Gaussian PDFs and y_t is an acoustic vector. Let be $Y = \{y_1, y_2, y_3\}$ be a sequence of acoustic feature vectors which represents a sample of speech which is claimed to have been spoken by speaker S. The probability for each vector and each PDF is given in the following table:

	G_1	G_2	G_3	G_4
y_1	$P(y_1/G_1)=0.03$	$P(y_1/G_2)=0.07$	$P(y_1/G_3)=0.04$	$P(y_1/G_4)=0.12$
y_2	$P(y_2/G_1)=0.02$	$P(y_2/G_2)=0.06$	$P(y_2/G_3)=0.05$	$P(y_2/G_4)=0.09$
y_3	$P(y_3/G_1)=0.03$	$P(y_3/G_2)=0.06$	$P(y_3/G_3)=0.03$	$P(y_3/G_4)=0.11$

(a) Write down the class conditional probability of the data Y given the speaker S.

[3] $P(Y/S) = P(y_1 y_2 y_3/S) = P(y_1/S)P(y_2/S)P(y_3/S)$

(ii) Assuming the prior probability of speaker S is 0.1 and assuming that the system will accept a speaker if the posterior probability of speaker S given the data Y is greater than 0.5, will this speaker be accepted or rejected? Justify your answer.

(1) $P(S/Y) = \frac{P(Y/S)P(S)}{P(Y)}$

(1) $P(Y/S) = P(y_1/S)P(y_2/S)P(y_3/S)$

(1) $P(y_1/S) = 0.2G_1(y_1) + 0.3G_2(y_1) + 0.1G_3(y_1) + 0.4G_4(y_1)$

$= 0.2(0.03) + 0.3(0.07) + 0.1(0.04) + 0.4(0.12) = 0.079$

(1) $P(y_2/S) = 0.2(0.02) + 0.3(0.06) + 0.1(0.05) + 0.4(0.09) = 0.063$

(1) $P(y_3/S) = 0.2(0.03) + 0.3(0.06) + 0.1(0.03) + 0.4(0.11) = 0.071$

(3) $P(S/Y) = (0.079)(0.063)(0.071)(0.1) = 3.53 \times 10^{-5}$

(2) $P(S/Y) <$ Threshold (0.5) $\Rightarrow$ Rejected.