



Roll No.

**PRESIDENCY UNIVERSITY
BENGALURU**

SCHOOL OF ENGINEERING

TEST 1

Sem A Y: Odd Sem 2019-20

Date: 01.10.2019

Course Code: SPEECH SIGNAL PROCESSING

Time: 2.30PM to 3.30PM

Course Name: ECE 306

Max Marks: 40

Programme & Sem: B.Tech (ECE) V

Weightage: 20%

Instructions:

- (i) Read the question properly and answer accordingly.
- (ii) Question paper consists of 3 parts.
- (iii) Scientific and Non-programmable calculators are permitted.

PART A [Memory Recall Questions]

Answer all the Questions. Each Question carries five marks.

(2Qx5M=10M)

1. Define Vowels, Semivowels, Diphthongs, Nasals, and Fricatives with an example
(C.O.NO.1)[Knowledge]
2. Give the classification of sounds of speech with an example. (C.O.NO.1)[Knowledge]

PART B [Thought Provoking Questions]

Answer both the Questions. Each Question carries five marks.

(2Qx5M=10M)

3. Explain how Zero-crossing rate method is used to classify the speech signals into voiced, unvoiced signals. If Frequency of the sine wave is 100 Hz and Sampling rate $F_s = 10\text{KHz}$, what is the average zero crossing rate Z_n ? (C.O.NO.1)[Comprehension]
4. Give a general representation of Short time analysis principle for Energy and Magnitude (C.O.NO.1)[Comprehension]

PART C [Problem Solving Questions]

Answer both the Questions. Each Question carries ten marks.

(2Qx10M=20M)

5. With a schematic diagram of Vocal-apparatus, explain the mechanism of speech production. (C.O.NO.1)[Comprehension]
6. Explain the concept of Speech Vs silence Discrimination using Energy and Zero Crossings. (C.O.NO.1)[Application]



SCHOOL OF ENGINEERING

Semester: ODD

Course Code: ECE 306

Course Name: Speech Signal Processing

Branch & Sem: ECE & 5th

Date: 1st October 2019

Time: 1 Hour

Max Marks: 40

Weightage: 20%

Extract of question distribution [outcome wise & level wise]

Q.NO	C.O.NO	Unit/Module Number/Unit /Module Title	Memory recall type [Marks allotted] Bloom's Levels		Thought provoking type [Marks allotted] Bloom's Levels		Problem Solving type [Marks allotted]		Total Marks
			K		C		A		
1	C.O.1	1/Phonemes	5						5
2	C.O.1	2/ Speech Sound Classification	5						5
3	C.O.2	2/ Short time analysis using ZCR			4		1		5
4	C.O.2	2/ Short time analysis			5				5
5	C.O.1	1/ Human Speech production			10				10
6	C.O.2	1/ American Phonemes					10		10
	Total Marks		10		19		11		40

Note: While setting all types of questions the general guideline is that about 60%

Of the questions must be such that even a below average students must be able to attempt, About 20% of the questions must be such that only above average students must be able to attempt and finally 20% of the questions must be such that only the bright students must be able to attempt.

[I hereby certify that All the questions are set as per the above guide lines. Aruna M]

Reviewers' Comments (i) Marks should be distributed evenly.
 (ii) Answer schemes steps to be given and Explanations to be given (1, 2, 5, 6)
 (iii) Question papers little bit lengthy

Jayvee
 Dr. M. Laxya

Annexure- II: Format of Answer Scheme



SCHOOL OF ENGINEERING

SOLUTION

Semester: ODD

Course Code: ECE 306

Course Name: Speech Signal Processing
 Branch & Sem: ECE & 5th

Date: 1st October 2019

Time: 1 Hour

Max Marks: 40

Weightage: 20%

Part A

(2Q x5 M =10 Marks)

Q No	Solution	Scheme of Marking	Max. Time required for each Question
1.	Vowels, Semivowels, Diphthongs, Nasals, and Fricatives with an example	1*5= 5M	7 min
2.	Voiced, unvoiced and Stops	1.5+1.5+2=5M	7 min

Part B

(2Q x5 M = 10 Marks)

Q No	Solution	Scheme of Marking	Max. Time required for each Question
3.	Short time Zero Crossing rate: Equation, block diagram, Theory	2+1+1+1=5M	8 min

A zero-crossing occurs if successive samples have different algebraic signs
 It is a measure of the frequency
 Definition

$$Z = \sum_{n=0}^{N-1} \text{sgn}(x(n)) \text{sgn}(x(n-1))$$

where

$$\text{sgn}(x(n)) = \begin{cases} 1 & \text{if } x(n) > 0 \\ -1 & \text{if } x(n) < 0 \end{cases}$$

and

$$x(n) = \sum_{k=0}^{N-1} x(k) \delta(n-k)$$

$$Z_n = 2(F_0 / F_s) = 200 / 10000 = 0.02 \text{ crossings/Sample}$$

4.	<p>i) Short time Energy: Equation, block diagram ii) Short time average magnitude: Equation, block diagram</p>	2.5*2=5M	8 min
----	---	----------	-------

Part C

(2Q x 10M =20 Marks)

Q No	Solution	Scheme of Marking	Max. Time required for each Question
5.	<p>Schematic of Human Vocal Apparatus. Theory</p>	5+5=10 M	15 min

	<ul style="list-style-type: none"> • muscle force pushes air out of the lungs (like a piston pushing air us within a cylinder through bronchi and trachea) • if vocal cords are tensed, air flow is forced to vibrate, producing voiced or quasi-periodic waveforms (musical notes) • if vocal cords are relaxed, air flow continues through vocal tract until it hits a constriction in the tract, causing it to become turbulent, thereby producing unvoiced sounds (like /s/, /z/, /t/, /d/), or it hits a point of total closure in the vocal tract, building up pressure until the closure is opened and the pressure is suddenly released, thereby causing a brief transient sound, like at the beginning of /p/, /t/, or /k/ 		
6.	Speech Vs silence Discrimination using Energy and Zero Crossings, Waveforms, Algorithm, Theory	2+2+3+3=10 M	15 min



Roll No.																			
----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

**PRESIDENCY UNIVERSITY
BENGALURU
SCHOOL OF ENGINEERING**

TEST - 2

Sem & AY: Odd Sem 2019-20

Course Code: ECE 306

Course Name: SPEECH SIGNAL PROCESSING

Program & Sem: B.Tech ECE & V

Date: 19.11.2019

Time: 2.30 PM to 3.30 PM

Max Marks: 40

Weightage: 20%

Instructions:

- (i) Read the question properly and answer accordingly.
- (ii) Question paper consists of 3 parts.
- (iii) Scientific and Non-programmable calculators are permitted.

Part A [Memory Recall Questions]

Answer both the Questions. Each Question carries five marks. (2Qx5M=10M)

1. Define STFT of a speech signal for LPF modulation form and BPF demodulation form
[5] (CO3) [Knowledge]
2. Define [5](CO4) [Knowledge]
 - i) Cepstrum of speech
 - ii) Homomorphic System for Convolution

Part B [Thought Provoking Questions]

Answer both the Questions. Each Question carries seven marks. (2Qx7M=14M)

3. Explain how the sampling rate of STFT is obtained as $2CF_s$? What is the oversampling rate? If $F_s=10000$ Hz, $L=100$ What is Bandwidth B? What is total sampling rate SR? Assume Hamming window.
[7] (CO3) [Application]
4. With a neat block diagram explain Linear filtering interpretation of STFT.
[7] (CO3) [Comprehension]

Part C [Problem Solving Questions]

Answer the Question. The Question carry sixteen marks. (1Qx16M=16M)

5. With related blockdiagram and waveforms explain parallel processing time-domain pitch detection.
[16] (CO2) [Application]



SCHOOL OF ENGINEERING

Semester: ODD

Course Code: ECE 306

Course Name: Speech Signal Processing

Branch & Sem: ECE & 5th

Date: 19th November 2019

Time: 1 Hour

Max Marks: 40

Weightage: 20%

Extract of question distribution [outcome wise & level wise]

Q.NO	C.O.NO	Unit/Module Number/Unit /Module Title	Memory recall type [Marks allotted] Bloom's Levels			Thought provoking type [Marks allotted] Bloom's Levels			Problem Solving type [Marks allotted]			Total Marks
			K			C			A			
1	C.O.3	STFT	2.5	2.5								5
2	C.O.4	Homomorphic Speech Processing	2.5	2.5								5
3	C.O.3	Sampling rate of STFT				5			2			7
4	C.O.3	Linear Filtering Interpretation of STFT				7						7
5	C.O.2	Pitch period estimation	5			7			4			16
	Total Marks		10	5		19			6			40

K = Knowledge Level C = Comprehension Level, A = Application Level

Note: While setting all types of questions the general guideline is that about 60%

Of the questions must be such that even a below average students must be able to attempt, About 20% of the questions must be such that only above average students must be able to attempt and finally 20% of the questions must be such that only the bright students must be able to attempt.

St

Annexure- II: Format of Answer Scheme



SCHOOL OF ENGINEERING

SOLUTION

Semester: ODD

Course Code: ECE 306

Course Name: Speech Signal Processing

Branch & Sem: ECE & 5th

Date: 19th November 2019

Time: 1 Hour

Max Marks: 40

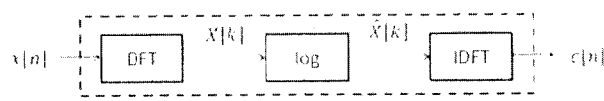
Weightage: 20%

Part A

(2Q x5 M =10 Marks)

Q No	Solution	Scheme of Marking	Max. Time required for each Question
1	<p>i.i</p> <p>1. modulation-lowpass filter $\Rightarrow X_n(e^{j\hat{\omega}}) = w(n) * [x(n)e^{-j\hat{\omega}n}]$, $\hat{n} = n$ variable; $\hat{\omega}$ fixed $X_n(e^{j\hat{\omega}}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} W(e^{j\theta}) X(e^{j(\theta+\hat{\omega})}) e^{j\hat{\omega}n} d\theta$</p> <p>2. bandpass filter-demodulation $\Rightarrow X_n(e^{j\hat{\omega}}) = e^{-j\hat{\omega}n} [(w(n)e^{j\hat{\omega}n}) * x(n)]$, $\hat{n} = n$ variable, $\hat{\omega}$ fixed</p>	<p>2.5*2=5M</p> <p>2.5</p> <p>2.5</p>	8 min

↓
5M
or 10M

2	<p>- The cepstrum is defined as the inverse DFT of the log magnitude of the DFT of a signal</p> $c[n] = F^{-1}[\log F\{x[n]\}]$ <p>where F is the DFT and F^{-1} is the IDFT</p> <p>- For a windowed frame of speech $x[n]$, the cepstrum is</p> $c[n] = \sum_{k=0}^{N-1} \log\left(\left \sum_{n=0}^{N-1} x[n]e^{-j\frac{2\pi}{N}kn}\right \right) e^{j\frac{2\pi}{N}kn}$ <div style="text-align: center;">  </div>	2.5*2=5M	8 min
	<p>- Homomorphic filtering is a generalized technique involving (1) a nonlinear mapping to a different domain where (2) linear filters are applied, followed by (3) mapping back to the original domain</p> <p>- Consider the transformation defined by $y(n) = L[x(n)]$</p> <ul style="list-style-type: none"> If L is a linear system, it will satisfy the principle of superposition: $L[x_1(n) + x_2(n)] = L[x_1(n)] + L[x_2(n)]$ <p>- By analogy, we define a class of systems that obey a generalized principle of superposition where addition is replaced by convolution</p> $H[x(n)] = H[x_1(n) * x_2(n)] = H[x_1(n)] * H[x_2(n)]$ <ul style="list-style-type: none"> Systems having this property are known as homomorphic systems for convolution, and can be depicted as shown below 	2.5	

Total 5/2

Part B

(2Q x 7 M = 14 Marks)

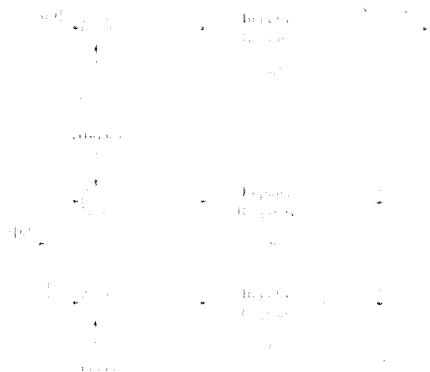
Q No	Solution	Scheme of Marking	Max. Time required for each Question
3.	<ul style="list-style-type: none"> the "total" sampling rate for the STFT is the product of the sampling rates in time and frequency, i.e., $SR = SR(\text{time}) \times SR(\text{frequency})$ $= 2B \times L \text{ samples/sec}$ $B = \text{frequency bandwidth of window (Hz)}$ $L = \text{time width of window (samples)}$ for most windows of interest, B is a multiple of F_s/L, i.e., $B = C F_s/L \text{ (Hz)}$ $C=1$ for Rectangular Window $C=2$ for Hamming Window $SR = 2C F_s \text{ samples/second}$ can define an 'oversampling rate' of $SR/F_s = 2C = \text{oversampling rate of STFT as compared to conventional sampling representation of } x(n)$ for RW, $2C=2$; for HW $2C=4 \Rightarrow$ range of oversampling is 2-4 this <u>oversampling</u> gives a <u>very flexible representation</u> of the speech signal <p>$B = C F_s / L = 2 * 10000 / 100 = 200 \text{ Hz};$ $SR = 2BL = 2 * 200 * 100 = 40000 \text{ samples/sec}$</p>	<p>5+1+1=7M</p> <p style="text-align: center;">5</p> <p style="text-align: center;">1</p> <p style="text-align: center;">1</p>	12 min

Linear Filtering Interpretation

3.5*2=7M

12 min

LPF



1. modulation-lowpass filter form:

$$X_n(e^{j\hat{\omega}}) = \sum_{m=-\infty}^{\infty} x(m)e^{-j\hat{\omega}m}w(n-m),$$

n variable, $\hat{\omega}$ fixed

$$= (x(n)e^{-j\hat{\omega}n}) * w(n)$$

$$= (x(n)\cos(\hat{\omega}n)) * w(n)$$

$$+ j(x(n)\sin(\hat{\omega}n)) * w(n)$$

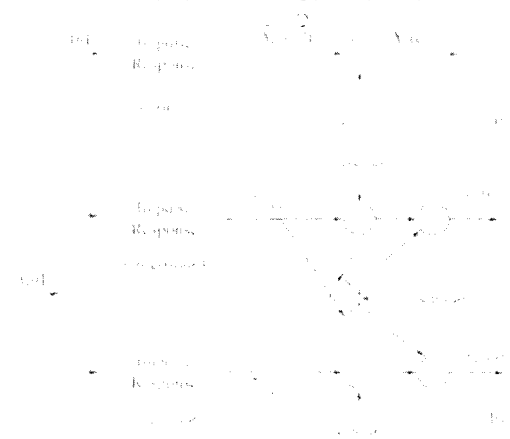
$a_r(\hat{\omega}) \quad j b_i(\hat{\omega})$

3.5

BPF

4.

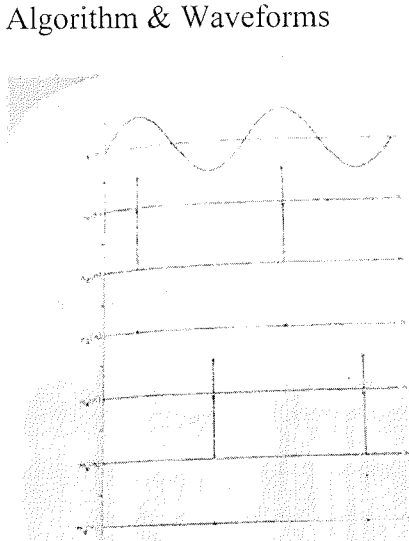
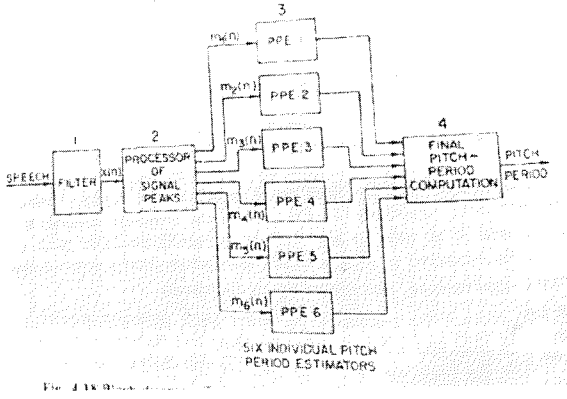
$$X_n(e^{j\hat{\omega}}) = e^{-j\hat{\omega}n} \left[(w(n)e^{j\hat{\omega}n}) * x(n) \right], \quad n \text{ variable, } \hat{\omega} \text{ fixed}$$



- complex bandpass filter output modulated by signal $e^{-j\hat{\omega}n}$
- if $W(e^{j\theta})$ is lowpass, then filter is bandpass around $\theta = \hat{\omega}$
- all real computation for lower half structure

3.5

Q No	Solution	Scheme of Marking	Max. Time required for each Question
5.	<p>Block Diagram</p> <p>Algorithm & Waveforms</p> <p>Theory</p>	<p>5+4+4+3=16 M</p> <p>5</p> <p>4</p> <p>4</p> <p>3</p>	<p>20 min</p>



- Fig. 4 B** Input (sinusoid) and corresponding impulse trains generated from the peaks and valleys.
- $m_1(n)$: An impulse equal to the peak amplitude occurs at the location of each peak.
 - $m_2(n)$: An impulse equal to the difference between the peak amplitude and the preceding valley amplitude occurs at each peak.
 - $m_3(n)$: An impulse equal to the difference between the peak amplitude and the preceding peak amplitude occurs at each peak. (If this difference is negative the impulse is set to zero.)
 - $m_4(n)$: An impulse equal to the negative of the amplitude at a valley occurs at each valley.
 - $m_5(n)$: An impulse equal to the negative of the amplitude at a valley plus the amplitude at the preceding peak occurs at each valley.
 - $m_6(n)$: An impulse equal to the negative of the amplitude at a valley plus the amplitude at the preceding local minimum occurs at each valley. (If this difference is negative the impulse is set to zero.)



Roll No																			
---------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

**PRESIDENCY UNIVERSITY
BENGALURU**

SCHOOL OF ENGINEERING

END TERM FINAL EXAMINATION

Semester: Odd Semester: 2019 - 20

Date: 23 December 2019

Course Code: ECE 306

Time: 9:30 AM to 12:30 PM

Course Name: SPEECH SIGNAL PROCESSING

Max Marks: 80

Program & Sem: B.Tech (ECE) & V (DE-II)

Weightage: 40%

Instructions:

- (i) Read the all questions carefully and answer accordingly.
- (ii) Question paper consists of 3 parts.
- (iii) Scientific and Non-programmable calculators are permitted.

Part A [Memory Recall Questions]

Answer all the Questions. Each Question carries 2 marks.

(10Qx2M=20M)

1.
 - a. Define Speech and Phonetics. (C.O.No.1) [Knowledge]
 - b. In speech production, the resonance frequencies of the vocal tract tube are called as _____ which depends on the shape and dimension of the _____. (C.O.No.1) [Knowledge]
 - c. Speech sounds are classified into _____ distinct classes according to their mode of excitation, they are _____, _____ and _____ sounds. (C.O.No.1) [Knowledge]
 - d. Define Diphthongs, Vowels. (C.O.No.1) [Knowledge]
 - e. Write time domain expression for Short time energy and short time average magnitude. (C.O.No.2) [Comprehension]
 - f. Zero crossing is said to occur if successive samples have _____. The rate at which zero crossings occur is a simple measure of _____ content of a signal. (C.O.No.2) [Knowledge]
 - g. The sampling rate of STFT in time domain is _____ and in frequency domain is _____. (C.O.No.3) [Knowledge]
 - h. Speaker Recognition system is broadly classified into _____ system and _____ system. (C.O.No.4) [Knowledge]

- i. Homomorphic systems for convolution obeys a generalized principle of _____ and it is expressed by the equation _____. (C.O.No.4) [Knowledge]
- j. The Homomorphic vocoder converts _____ to _____. (C.O.No.4) [Knowledge]

Part B [Thought Provoking Questions]

Answer all the Questions. Each Question carries 10 marks. (3Qx10M=30M)

2. In many speech applications synthetic speech is used instead of natural speech. Explain the technique used in this conversion with a neat block diagram of Analyzer and Synthesizer. (C.O.No.4) [Comprehension]
3. Write the canonic form for Homomorphic convolution, Characteristic and inverse characteristic system along with its neat block diagram. (C.O.No.4) [Application]
- 4.
- a. Explain linear filtering interpretation of speech for complex and real operations. [7M](C.O.No.3) [Comprehension]
- b. In computing STFT the length of the window used is 100, sampling frequency F_s is 10000 Hz, Find the overall Bandwidth and sampling rate. [3M](C.O.No.4) [Application]

Part C [Problem Solving Questions]

Answer all the Questions. Each Question carries 10 marks. (3Qx10M=30M)

5. With a schematic diagram of Vocal-apparatus, explain the mechanism of speech production. (C.O.No.1) [Comprehension]
6. Write a short note on online digital speaker verification system. Also explain with a block diagram of the signal processing aspects of the speaker verification system. (C.O.No.4) [Application]
7. With a neat block diagram explain the isolated digit recognition system. (C.O.No.4) [Application]



SCHOOL OF ENGINEERING

END TERM FINAL EXAMINATION

Extract of question distribution [outcome wise & level wise]

Q.NO	C.O.NO (% age of CO)	Unit/Module Number/Unit /Module Title	Memory recall type	Thought provoking type	Problem Solving type	Total Marks
			[Marks allotted]	[Marks allotted]		
			Bloom's Levels	Bloom's Levels	[Marks allotted]	
			K	C	A	
1	1	1	2			2
2	1	1	2			2
3	1	1	2			2
4	1	1	2			2
5	2	2		2		2
6	2	2	2			2
7	3	3	2			2
8	4	4	2			2
9	4	4	2			2
10	4	4	2			2
11	4	4		10		10
12	4	4		5	5	10
13	3	3		7	3	10
14	1	1		10		10
15	4	4			10	10
16	4	4			10	10
Total Marks			18	34	28	80

K = Knowledge Level C = Comprehension Level, A = Application Level

Note: While setting all types of questions the general guideline is that about 60%

Of the questions must be such that even a below average students must be able to attempt, About 20% of the questions must be such that only above average students must be able to attempt and finally 20% of the questions must be such that only the bright students must be able to attempt.

I hereby certify that all the questions are set as per the above guidelines.

Faculty Signature:

Reviewer Comment:

Format of Answer Scheme



SCHOOL OF ENGINEERING

SOLUTION

Semester: Odd Sem. 2019-20

Course Code: ECE 306

Course Name: SPEECH SIGNAL PROCESSING

Program & Sem: B.TECH & 5th SEMESTER (ECE)

Date: 23.12.2019

Time: 9.30 AM to 12.30 PM

Max Marks: 80

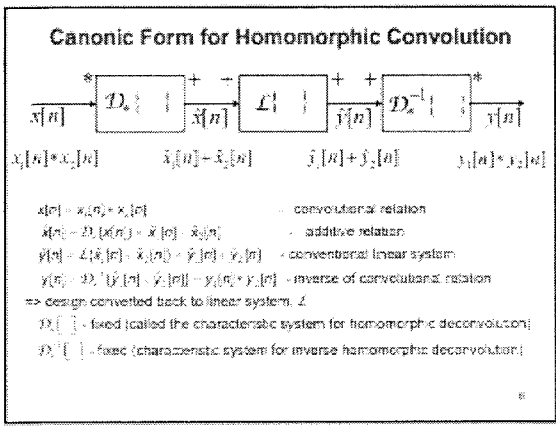
Weightage: 40%

Part A

(10Q x 2M = 20Mark)

Q No	Solution	Scheme of Marking	Max. Time required for each Question
1	Speech: Speech signals are composed of sequence of sounds which serve as a symbolic representation of information Phonetics: The study and classification of sounds of speech	1M 1M	3 min
2	Formant frequencies and Vocal tract	1M, 1M	3 min
3	Three, Voiced, unvoiced or fricatives, plosives	0.5*4=2M	3 min
4	Diphthongs: Gliding monosyllabic speech item that starts at or near the articulatory position for one vowel and moves to or toward the position for another. Vowel: Vowels are the speech sounds which are produced by exciting a fixed vocal tract with quasi-periodic pulses of air caused by vibration of vocal cords.	1M 1M	3 min

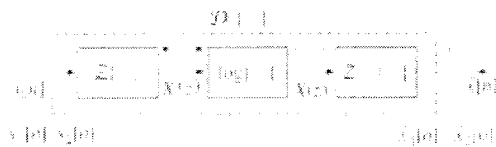
12



4M

20 min

Characteristic System for Deconvolution



3M

$$X(z) = \sum x[n]z^{-n} = |X(z)|e^{j\arg[X(z)]}$$

Inverse Characteristic System for Deconvolution



3M

$$\hat{Y}(z) = \sum \hat{y}[n]z^{-n}$$

$$Y(z) = \exp[\hat{Y}(z)] = \log|Y(z)| + j\arg[Y(z)]$$

13.a. **Linear Filter Interpretation:**

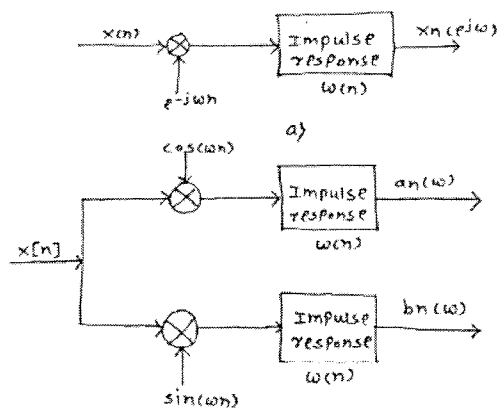
$$x_n(e^{j\omega}) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)e^{-j\omega m}$$

is linear filtering or convolution.

(i) It is also evident from the equation that of every value of $w_n(x_n(e^{j\omega}))$ is nothing but the convolution of the sequence $w(n)$ with the sequence $x(n)e^{-j\omega n}$.

(ii) For any particular given value of $w_n(x_n(e^{j\omega}))$ can be visualized as the output of a system. We can see that the output is complex.

(iii) Now, if $x_n(e^{j\omega})$ is expressed as, $x_n(e^{j\omega}) = a_n(\omega) - jb_n(\omega)$



1M

15 min

3M

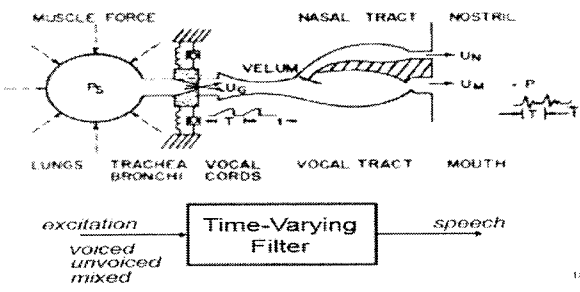
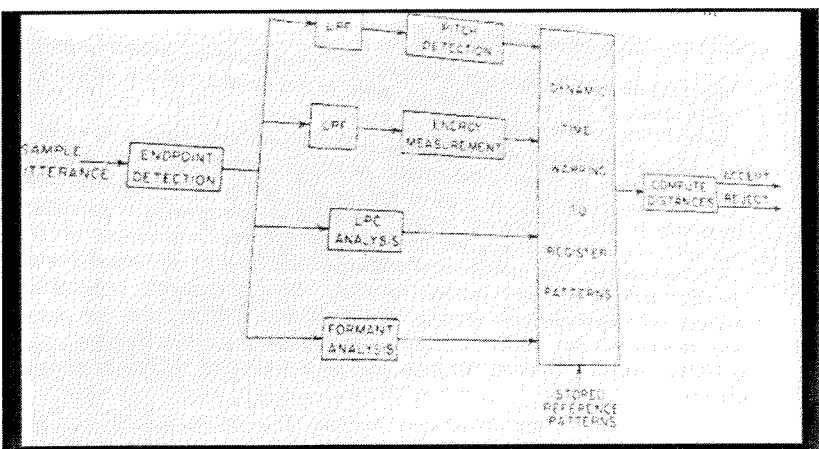
3M

Linear Filtering using a) Complex operations b) Real operations only

	b) Bandwidth= $2F_s/L=(2*10000)/100=200$ Hz Sampling Rate= $2BL=2*200*100=40000$ samples/sec	2M 1M	5 min
--	---	----------	-------

Part C

(3Q x 10M = 30Marks)

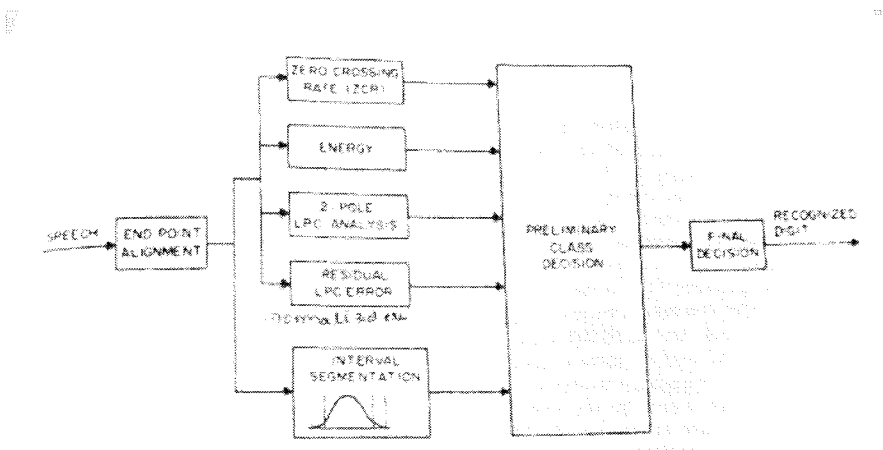
Q No	Solution	Scheme of Marking	Max. Time required for each Question
14	<p>Schematic of Human Vocal Apparatus, Theory</p>  <p>Theory</p> <ul style="list-style-type: none"> • muscle force pushes air out of the lungs (like a piston pushing air up within a cylinder) through bronchi and trachea • if vocal cords are tensed, air flow causes them to vibrate, producing voiced or quasi-periodic speech sounds (musical notes) • if vocal cords are relaxed, air flow continues through vocal tract until it hits a constriction in the tract, causing it to become turbulent, thereby producing unvoiced sounds (like /s/, /sh/), or it hits a point of total closure in the vocal tract, building up pressure until the closure is opened and the pressure is suddenly and abruptly released, causing a brief transient sound, like at the beginning of /p/, /t/, or /k/ 	5M 5M	20 min
15	<p>Block Diagram</p> 	5M	20 min

Theory
 A speaker verification system takes the speech of an unknown speaker with his/her claimed identity, and it determines whether the claimed identity matches the speech. The claimed identity can be fed into the system using various channels such as keyboard, identity card, etc.

Each speaker recognition system has two phases: Enrollment and verification.

5M

16 Block Diagram



5M

Theory

The block diagram of the overall digit recognition system that was implemented is shown above. Following endpoint alignment in which the interval containing the word to be recognized is carefully determined, the speech is analyzed every 10 ms to obtain zero-crossing rate, energy, two-pole model linear-predictive-coding (LPC) coefficients, and the residual LPC estimation error. To aid in making preliminary classification decisions, the speech interval is segmented into three well-defined regions. All the speech information is fed in parallel into a preliminary decision-making algorithm that chooses one of several possible digit classes for the input utterance—e.g., one class contains the digits 1 and 9. A final decision is then made based on the presence or absence of certain key features in the input speech. In this section, we show how the various digits can be characterized in terms of certain acoustic features. Then we discuss some key signal processing functions that are heavily relied on in the decision algorithms and that contribute strongly to making the system speaker independent.

5M

20 min