



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

**PRESIDENCY UNIVERSITY  
BENGALURU  
SCHOOL OF ENGINEERING**

**TEST - 1**

**Winter Semester:** 2021 - 22

**Course Code:** Speech Signal Processing

**Course Name:** ECE 306

**Programme & Sem:** B.Tech ECE & VI Sem

**Date:** 26/04/2022

**Time:** 1.30 PM to 2.30 PM

**Max Marks:** 30

**Weightage:** 15%

**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**

**Answer both the Questions. Each question carries 5 marks.**

**(2Q x 5M =10M)**

1. Speech signals are composed of sequence of sounds. These sounds serves as a symbolic representation of information. Define Vowels, Semivowels, Diphthongs, Nasals, and Fricatives with an example (CO1) [Knowledge]
2. Voice or voicing is a term used in phonetics and phonology to characterize speech sounds. Speech sounds can be described as either voiceless or voiced. Give the classification of sounds of speech with an example. (CO1) [Knowledge]

**PART B**

**Answer both the Questions. Each question carries 5 marks.**

**(2Q x 5M = 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  $Z_n$  ? (CO2) [Comprehension]
4. The energy associated with speech is time varying in nature. Short-Time Energy and short time magnitude serves to differentiate voiced and unvoiced sounds in speech from silence. Give a general representation of Short time analysis principle for Energy and Magnitude (CO2) [Comprehension]

**PART C**

**Answer the Question. The question carries 10 marks.**

**(1Q x10M =10M)**

5. Speech is a sequence of ever changing sounds, sound properties are highly dependent on context (i.e., the sounds which occur before and after the current sound) With a schematic diagram of Vocal-apparatus, explain the mechanism of speech production. (CO1) [Comprehension]



Roll No.

**PRESIDENCY UNIVERSITY  
BENGALURU**

**SCHOOL OF ENGINEERING**

**TEST - 2**

**Even Semester:** 2021-22

**Course Code:** Speech Signal Processing

**Course Name:** ECE 306

**Programme & Sem:** ECE , VI Sem

**Date:** 01/06/2022

**Time:** 1.30 pm to 2:30pm

**Max Marks:** 30

**Weightage:** 15%

**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**

**Answer all the Questions. Each question carries TWO marks. (5Qx2M=10M)**

1. The problem of locating the beginning and end of a speech utterance in the presence of background noise is very much important in many areas of speech processing. This scheme is based on  
a. Energy and ZCR      b. Energy alone      c. ZCR alone      d. both a & b  
[2] (CO2) [Knowledge]
2. Homomorphic filtering is a generalized technique for signal and image processing, involving a nonlinear mapping to a different domain in which linear filter techniques are applied, followed by mapping back to the original domain. Homomorphic system satisfies the principle of  
a. Linearity      b. Non linearity      c. both a and b      d. None of the above  
[2] (CO4) [Knowledge]
3. Pitch period estimation is one of the most important problems in speech processing. A pitch detection scheme proposed by gold and rabiner has \_\_\_\_\_ pitch period estimators working in \_\_\_\_\_  
a. 7 and series      b. 8 and series      c. 6 and parallel      d. 6 and series  
[2] (CO2) [Knowledge]
4. STFT is a function of  
a. two variables, the time index,  $n$  which discrete and the frequency variable  $\omega$  which is continuous  
b. two variables, the time index,  $n$  which continuous and the frequency variable  $\omega$  which is discrete  
c. time index  $n$  only  
d. frequency variable  $w$  only  
[2] (CO3) [Knowledge]

5. Window plays a key role in STFT. The role of window function in STFT is

- a. Chooses portion of the signal to be analyzed
- b. Window shape determines the nature of the  $X(n, \omega)$
- i. a only
- ii. b only
- iii. both a and b
- iv. none of the above

[2] (CO3) [Knowledge]

### PART B [Thought Provoking Questions]

**Answer all the Questions. Each question carries 5 marks.(2Q x 5M = 10 marks)**

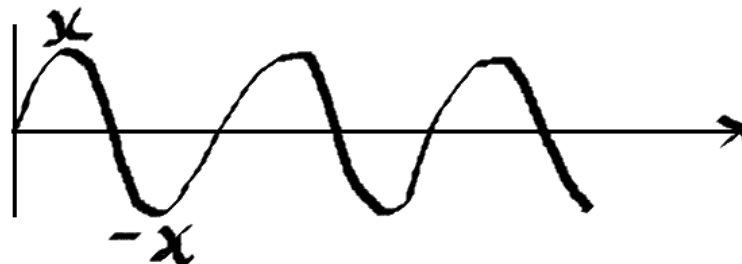
6. Need to be sample STFT in both time and frequency to produce an unaliased representation from which  $x(n)$  can be exactly recovered. The total sampling rate for the STFT is the product of the sampling rates in time and frequency, hence prove the sampling rate of STFT is  $2CFs$  samples/second. 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. [5] (CO3) [Comprehension]

7. Homomorphic filtering is a generalized technique for signal and image processing, involving a nonlinear mapping to a different domain in which linear filter techniques are applied, followed by mapping back to the original domain. Identify different Process involved in building of the Canonic Form for Homomorphic Deconvolution [5] (CO4) [Comprehension]

### PART C [Problem Solving Questions]

**Answer the Question. The question carries 10 marks. (1Q x10M =10 marks)**

8. Pitch period estimation is one of the most important problems in speech processing. Consider the following speech input which has peak amplitude as  $+x$  and valley amplitude a  $-x$ , write down the output of 6 pitch period estimators and also demonstrate the concept of identifying pitch period using a parallel processing approach..



[10] (co2) [Application]



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

**PRESIDENCY UNIVERSITY  
BENGALURU  
SCHOOL OF ENGINEERING**

**END TERM FINAL EXAMINATION**

**Semester :** Even Semester: 2021 - 22

**Date:** 30<sup>th</sup> June 2022

**Course Code:** ECE 306

**Time:** 9:30 AM to 12.30PM

**Course Name:** SPEECH SIGNAL PROCESSING

**Max Marks:** 100

**Program & Sem:** B.Tech (ECE) & VI Sem

**Weightage:** 50 %

**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.**

**(15Qx 2M= 30M)**

- 1) Zero crossing rate is a measure of frequency content of the signal. Zero crossing is said to occur if successive samples have \_\_\_\_\_ sign. The rate at which zero crossings occur is a simple measure of \_\_\_\_\_ content of a signal.[2] (CO2) [Knowledge]
- 2) The vocal tract is the area from the nose and the nasal cavity down to the vocal cords deep in the throat. The nasal tract begins at \_\_\_\_\_ and ends at \_\_\_\_\_[2] (CO1) [Knowledge]
- 3) Pitch period estimation is one of the most important problems in speech processing. During blanking interval  $\tau(n)$ 
  - a. No pulse will be detected
  - b. No pulse will be recognized
  - c. All the pulses are detected
  - d. Both a and b are true [2] (CO2) [Knowledge]
- 4) The oversampling rate of STFT 2C indicates the ratio of
  - a. SR/Fs
  - b. 2B/Sr
  - c. CFs
  - d. 2BL[2] (CO3) [Knowledge]
- 5) If 2C is the oversampling rate of STFT then
  - a. For hamming window oversampling rate will be 4 and for rectangular window it is 2
  - b. For hamming window oversampling rate will be 2 and for rectangular window it is 4
  - c. For hamming window oversampling rate will be 8 and for rectangular window it is 4
  - d. For hamming window oversampling rate will be 4 and for rectangular window it is 8[2] (CO3) [Knowledge]

6) Any homomorphic system can be represented as a cascade of three systems. The system that takes inputs combined by convolution and transforms them into additive outputs is known as

- a. Characteristic System
  - b. Inverse Characteristic system
  - c. Linear System
  - d. Nonlinear system
- [2] (CO4) [Knowledge]

7) If  $F_s=10000$  Hz,  $L=100$  What is Bandwidth B? What is total sampling rate SR? Assume Hamming window.

- a.  $B=200$ Hz and Sampling Rate= 40000Samples/Sec
  - b.  $B=200$ Hz and Sampling Rate= 20000Samples/Sec
  - c.  $B=100$ Hz and Sampling Rate= 40000Samples/Sec
  - d.  $B=400$ Hz and Sampling Rate= 40000Samples/Sec
- [2] (CO3) [Knowledge]

8) Speech sounds are classified into \_\_\_\_\_ distinct classes according to their mode of excitation, they are \_\_\_\_\_, \_\_\_\_\_ and \_\_\_\_\_ sounds

- a. 4, voiced, unvoiced, plosives
  - b. 3, voiced, unvoiced, plosives
  - c. 3, voiced, diphthongs, plosives
  - d. None of the above
- [2] (CO1) [Knowledge]

9) Voice response systems are specialized technologies designed for providing callers with verbal and faxed answers to inquiries without assistance from a person. In multiline digital voice Response system \_\_\_\_\_ number of users can be assisted at the same time.

[2] (CO4) [Knowledge]

10) The problem of locating the beginning and end of a speech utterance in the presence of background noise is very much important in many areas of speech processing. This scheme is based on

- a. Energy and ZCR
  - b. Energy alone
  - c. ZCR alone
  - d. None of the above
- [2] (CO2) [Knowledge]

11) The double threshold procedure in locating starting point and end point of a speech signal ensures that dip in average magnitude do not falsely signal the end point. Hence \_\_\_\_\_ Number of frames are limited preceding N1 to check for ZCR with respect to IZCT.

- a. 35
  - b. 18
  - c. 25
  - d. 55
- [2] (CO2) [Knowledge]

12) Pitch period estimation is one of the most important problems in speech processing. A pitch detection scheme proposed by gold and rabiner has \_\_\_\_\_ pitch period estimators working in \_\_\_\_\_

- a. 7 and series
  - b. 8 and series
  - c. 6 and parallel
  - d. 6 and series
- [2] (CO2) [Knowledge]

13) In speech production, the resonance frequencies of the vocal tract tube are called as \_\_\_\_\_ which depends on the shape and dimension of the \_\_\_\_\_.

- a. Formant frequency, vocal tract                      b. formants, nostrils  
c. vocal frequency, mouth                                d. None of the above                      [2] (CO1) [Knowledge]

14) STFT can be represented in different forms, one such form to represent  $X_n(e^{j\omega})$  is in terms of real and imaginary parts only as

- a.  $X_n(e^{j\omega}) = a_n(\omega) - j b_n(\omega)$     b.  $X_n(e^{j\omega}) = a_n(\omega) + j b_n(\omega)$   
c.  $X_n(e^{j\omega}) = a_n(\omega) - b_n(\omega)$     d.  $X_n(e^{j\omega}) = b_n(\omega) - j a_n(\omega)$                       [2] (CO3) [Knowledge]

15) The problem of locating the beginning and end of a speech utterance in the presence of background noise is very much important in many areas of speech processing. The fundamental assumption which is used in an algorithm to locate beginning and end of speech signal assumes

- a. The first 100ms interval contains noise                      b. The last 100ms interval contains noise  
c. The first 120ms interval contains noise                      d. The first 10ms interval contains noise  
[2] (CO2) [Knowledge]

### Part B [Thought Provoking Questions]

**Answer all the Questions. Each Question carries 10 marks. (4Qx10M=40M)**

16. Homomorphic filtering is a generalized technique for signal and image processing, involving a nonlinear mapping to a different domain in which linear filter techniques are applied, followed by mapping back to the original domain. Identify different Process involved in building of the Canonic Form for Homomorphic Deconvolution, characteristic System and Inverse Characteristic system [10] (CO4) [Comprehension]

17. Cepstrum Analysis is a tool for the detection of periodicity in a frequency spectrum, have been used mainly in speech analysis for voice pitch determination. Infer with analyzer and synthesizer of Homomorphic vocoder, how low frequency and high frequency components of a speech spectrum aids in producing a very high natural sounding speech. [10] (CO4) [Application]

18. a) The energy associated with speech is time varying in nature. Short-Time Energy and short time magnitude serves to differentiate voiced and unvoiced sounds in speech from silence. Give a general representation of Short time analysis principle for Energy and Magnitude. [5] (CO2) [Comprehension]

b). The Zero-Crossing Rate (ZCR) of an audio frame is the rate of sign-changes of the signal during the frame. Interpret system involved in identifying noise present in a speech signal through the concept of Zero crossing rate [5] (CO2) [Application]

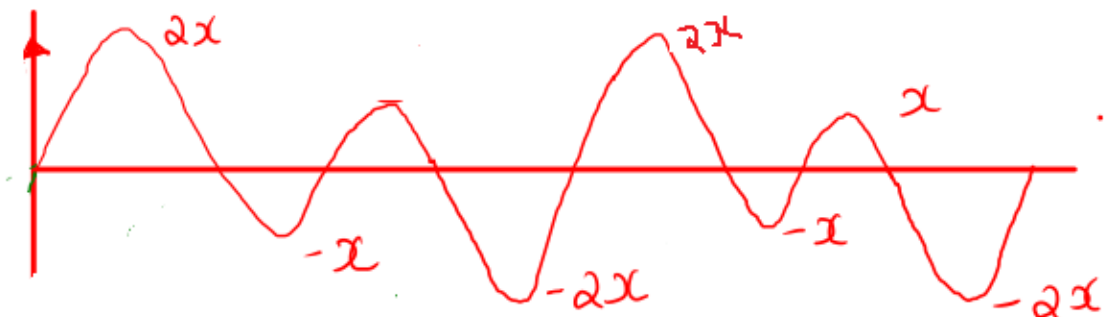
19. Voice or voicing is a term used in phonetics and phonology to characterize speech sounds. Speech sounds can be described as either voiceless or voiced. With a neat human vocal apparatus demonstrate the mechanism of human speech production.[10](CO1)[Comprehension]

**Part C [Problem Solving Questions]**

**Answer all the Questions. Each Question carries TEN marks.**

**(3Qx10M=30M)**

20. Pitch period estimation is one of the most important problems in speech processing. Consider the following speech input which has peak amplitude as  $+2x$  and  $-2x$ . Write down the output of 6 pitch period estimators and also explain the concept of identifying pitch period using 6 parallel pitch period estimators with a neat block diagram.



[10] (CO2)

[Application]

21. a) Need to be sample STFT in both time and frequency to produce an unaliased representation from which  $x(n)$  can be exactly recovered. The total sampling rate for the STFT is the product of the sampling rates in time and frequency, hence prove the sampling rate of STFT is  $2CF_s$  samples/second. 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.

[5] (CO3)

[Comprehension]

b). Illustrate the working principle of telephone directory assistance system for assisting different requests for a end user through Voice response system [5] (CO4) [Application]

22. Speaker recognition can be classified into speaker identification and speaker verification system. Interpret the process of accepting or rejecting the identity claimed by a speaker using speaker verification system. [10] (CO4) [Application]