SPEECH AND AUDIO PROCESSING
(ECEN 3133)

Time Allotted: 3 hrs **Full Marks : 70**

Figures out of the right margin indicate full marks.

Candidates are required to answer Group A and any 5 (five) from Group B to E, taking at least one from each group.

Candidates are required to give answer in their own words as far as practicable.

Group – A
(Multiple Choice Type Questions)

1. Choose the correct alternative for the following: **10 × 1 = 10**

(i) Articulators move in response to the neural signals to perform a sequence of gestures, the end result of which is an _____waveform which contains the information in the original message.
(a) acoustic (b) transient
(c) aperiodic (d) periodic.

(ii) In CELP coders, the sampling rate and quantization rate in bits/sample are:
(a) 4000 and 13 (b) 8000 and 13
(c) 4000 and 20 (d) 8000 and 20.

(iii) The expression "r/A" is called acoustic_____, where r is the density of air in a tube simulating the vocal tract, and A(x,t) is the cross-sectional area normal to the x axis of the tube, as a function of the distance along the tube and as a function of time t.
(a) Capacitance (b) Inductance
(c) Resistance (d) Conductance.

(iv) LSF is important as it indicates:
(a) the frequency wise power distribution
(b) the frequencies present
(c) the maximum power present
(d) the operation of the transmitter.

(v) Main effect of friction and thermal conduction losses on a simulated vocal tube is that the formants bandwidth is _____.
(a) decreased (b) increased
(c) unchanged (d) halved.

(vi) The number of phonemes in spoken English language is:
(a) 40-50 (b) 40-100
(c) 40-60 (d) 400-600.

(vii) Prosody is composed of :
(a) intonation (b) stress signals
(c) both intonation and stress signals (d) gender information.

(viii) The vocal tract acts as:
(a) a time-varying filter (b) a time independent filter
(c) band pass filter (d) a low pass filter.

(ix) Children have voice pitch frequency in the range:
(a)100-300 (b) 200-400
(c) 200-600 (d) 300-600.

(x) Message information, a person needs to speak out, is first converted to a set of neural signals which control the _____ mechanism.
(a) tower (b) articulatory
(c) human (d) sensory.

Group – B

2. (a) For a uniform acoustic tube behaving identically to a lossless uniform transmission line terminated in a short circuit at one end and excited by a current source at the other end, derive the expressions for "Acoustic inductance" and "Acoustic capacitance". Assume plane wave propagation and no losses at tube walls.

(b) Loss less tube simulating a vocal tract can be characterized by a set of resonances but effect of losses needs to be considered. The bandwidth of the two lowest resonances is affected by what type of losses? The bandwidth of the higher resonant frequencies depends on what types of losses?

6 + (3 + 3) = 12

3. (a) What are the several types of losses in a vocal tract and what is its effect?

(b) Analyze the effects of losses in vocal tract by wall vibration?

6 + 6 = 12

Group – C

4. (a) Define 'Pitch' in relation to human voice. What information is contained in the pitch signal? Name and explain at least 4 of them.

(b) How does gender and age affect speech? Explain with specific data.

(2 + 5)+ 5 = 12

5. (a) Draw the block diagram for the speech-filter model for speech production. What are the functions of Glottal model and voice tract models?

(b) Express mathematically the linear prediction model of a speech. Convert it into frequency domain.

(2 + 4) + 6 = 12

Group – D

6. (a) What is speech synthesis?

(b) Describe the VODER and the VOCODER?

(c) Under Speech communication system where the Signal is transmitted, stored and processed in many ways, what are the primary sources of concern?

2 + (3 + 3) + 4 = 12

7. (a) Highlight the main components of an ASR (Automatic Speech Recognition) system, and list them in order.

(b) What is meant by "feature extraction" in ASR? What is the function of a local match module in ASR?

6 + (3 + 3) = 12

Group – E

8. (a) Why is uniform filter-bank using discrete Fourier Transform not ideal for human voice? What is the solution proposed?

(b) What is tree structured AS FB solution? Explain. Describe briefly an ASFB with a block diagram.

(2 + 2) + (4 + 4) = 12

9. (a) What is the expanded form of CELP? What is the principle of operation of a CELP encoder?

(b) Draw the typical block diagram for a CELP coder and explain the functions of the blocks.

(1 + 5) + 6 = 12

Department & Section	Submission Link
ECE	https://classroom.google.com/w/MTM4NDQzODI5ODgy/tc/Mjg3MDQ5NTM0NDkx