GUJARAT TECHNOLOGICAL UNIVERSITY

Master of Engineering
Subject Code: 3724119
Semester – II
Subject Name: Audio Processing

Type of course: Program Specific Elective

Prerequisite: Basic knowledge of Signal and Systems and Digital Signal Processing

Rationale: For humans, speech is a natural way of communicating the ideas. This course is a fundamental course on how to process digital speech and audio signal to extract useful information. The course builds upon the theory of digital signal processing and extends the concepts applied to audio signal in particular. The course also discusses the applications of audio signal processing.

Teaching and Examination Scheme:

Teaching Scheme			Credits	Examination Marks				Total Marks
L	T	P	C	Theory Marks		Practical Marks		
				ESE (E)	PA (M)	ESE (V)	PA (I)	
3	0	2	4	70	30	30	20	150

Content:

Sr. No.	Content	Total Hrs
1.	Principle characteristics of speech: linguistic information, speech and hearing, speech production mechanism, articulatory phonetics: manner of articulation and place of articulation, acoustic phonetics: spectrograms, wide-band and narrow-band spectrograms, acoustic characteristic of speech, statistical characteristics of speech, speech production models	09
2.	Speech analysis and synthesis systems: Digitization, sampling, quantization and coding, spectral analysis, spectral structure of speech, short-time autocorrelation, average magnitude difference function, short time Fourier transform, window function, sound spectrogram, homomorphic processing, properties of complex cepstrum, complex cepstrum of voiced and unvoiced speech, mel-frequency cepstral coefficients, filter bank and zero crossing analysis, analysis –by-synthesis, pitch extraction.	09
3.	Linear predictive analysis: Basic principles of linear predictive analysis, autocorrelation method and covariance method, computation of gain for the model, prediction error signal, frequency domain interpretation of LP analysis, frequency domain interpretation of mean-squared prediction error, applications of LPC parameters	09
4.	Speech coding: reversible coding, irreversible coding and information rate distortion theory, coding in time domain: PCM, ADPCM, adaptive predictive coding, coding in frequency domain: sub band coding, adaptive transform coding, vector quantization, code excited linear predictive coding (CELP).	09

5.	Speech Recognition: Principles of speech recognition, speech period detection, spectral distance measure, structure of word recognition system, dynamic time warping (DTW), theory and implementation of hidden Markov model (HMM).	06
6.	Speaker recognition: Human and computer speaker recognition principles text dependent and text independent speaker recognition systems. applications of speech processing	06
Total		48

Suggested Specification table with Marks (Theory): (For BE only)

Distribution of Theory Marks (100)					
R Level	U Level	A Level	N Level	E Level	C Level
10	**30**	**30**	**20**	**5**	**5**

Legends: R: Remembrance; U: Understanding; A: Application, N: Analyze and E: Evaluate C: Create and above Levels (Revised Bloom's Taxonomy)

Note: This specification table shall be treated as a general guideline for students and teachers. The actual distribution of marks in the question paper may vary slightly from above table.

Reference Books:

[1] Sadaoki Furui, "Digital Speech Processing, Synthesis and Recognition" 2nd Edition, Taylor & Francis, 2000.
[2] Rabiner and Schafer, "Digital Processing of Speech Signals", Pearson Education, 1979.

Course Outcomes:

Sr. No.	CO statement	Marks % weightage
CO-1	Understand different characteristics of Speech.	40%
CO-2	Identify and analyse different speech analysis system.	30%
CO-3	Write algorithms for Recognition of speech.	30%

List of Experiments / Assignments:

1. To study the effects of windowing.
2. To understand the difference between stationary and non-stationary signals.

3. To extract a slice of speech signal and compute its spectrum for different window length.
4. To simulate periodic glottal pulse train.
5. To synthesize vowel using source filter model.
6. To compute wideband and narrowband spectrogram of a given speech signal.
7. To compute short-time energy and ZCR of a given speech signal.
8. To compute short-time autocorrelation function and plot pitch contour for given utterance.
9. To compute short-time AMDF and plot pitch contour for given utterance.
10. To detect pitch using harmonic product spectrum.
11. To study LPC and cepstral analysis method.
12. Develop an algorithm to segment a given speech signal into speech and silence parts.
13. For a given speech signal, classify speech segments into two parts: voiced and unvoiced speech segments
14. Given a speech signal, determine whether it contains an adult voice or a child voice.
15. Discriminate between speech and music signal.
16. Determine the locations of vowels in the given speech signal.

Major Equipment:

Computational lab or facility with the following software or their equivalent:

(1) MATLAB signal processing toolbox
(2) Python

List of Open Source Software/learning website:

(1) NPTEL Video lectures
(2) SciLab