# INSTITUTE OF AERONAUTICAL ENGINEERING

(Autonomous)

Dundigal - 500 043, Hyderabad, Telangana

## COURSE CONTENT

| SPEECH AND IMAGE PROCESSING | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **VIII Semester: ECE** | | | | | | | | |
| **Course Code** | **Category** | **Hours / Week** | | | **Credits** | **Maximum Marks** | | |
| | | **L** | **T** | **P** | **C** | **CIA** | **SEE** | **Total** |
| **AECD63** | **Elective** | 3 | - | - | 3 | 40 | 60 | 100 |
| **Contact Classes: 48** | **Tutorial Classes: Nil** | **Practical Classes: Nil** | | | | **Total Classes: 48** | | |
| **Prerequisite: Signals and Systems** | | | | | | | | |

### I. COURSE OVERVIEW:

This course introduces the fundamental concepts speech and video processing. It deals with speech processing that ranges from the basic nature of the speech signal, through a variety of methods of representing speech in digital form, to applications in voice communication and automatic synthesis and recognition of speech. Speech processing in widely used is applications like google voice search, artificial intelligence voice-controlled assistant like Alexa. It also includes the basics of video processing with emphasis on the Image Formation techniques and motion estimation algorithms. These topics serves as the pre-processing blocks in most of the video processing applications.

### II. COURSES OBJECTIVES:

**The students will try to learn**

I. Methods and systems for efficient quantization and coding of speech signals.
II. The concepts speech production and linear predictive analysis (LPC) for speech synthesis.
III. The fundamental concepts of image formation in video processing.
IV. The algorithms to solve video processing problems and meet design specifications for industry, medicine and defense applications.

### III. COURSE OUTCOMES:

**At the end of the course students should be able to:**

CO 1 Illustrate the difference between acoustic phonetics and articulatory phonetics for speech processing.

CO 2 Utilize digital model designed by sampled speech signal for speech processing applications like speech recognition, speech synthesis and verification.

CO 3 Analyze methods to estimate pitch period to design vocoders, artificial intelligence voice-controlled assistants like Alexa.

CO 4 Apply linear predictive coding for speech synthesis, compression and spectrographic displays. CO 5 Understand the basic concepts of image formation techniques for Video Processing applications.

CO 6 Acquire the Mathematical and computational skills needed to understand the principal of 2-D Motion Estimation.

## IV. COURSE CONTENT:

**MODULE – I: FUNDAMENTALS OF HUMAN SPEECH PRODUCTION (10)**
The Process of Speech Production, Short-Time Fourier Representation of Speech, The Acoustic Theory of Speech Production. Lossless Tube Models of the Vocal Tract, Digital Models for Sampled Speech Signals.

**MODULE – IV: TIME DOMAIN MODELS AND SHORT TIME FOURIER ANALYSIS FOR SPEECH PROCESSING (10)**
Time dependent processing of speech, short time energy and average magnitude, short time average zero crossing rate, discrimination between speech& silence, pitch period estimation using parallel processing, short time autocorrelation function and AMDF, pitch period estimation using autocorrelation function.

**MODULE - III: SHORT TIME FOURIER ANALYSIS FOR SPEECH PROCESSING (10)**
Fourier transform and linear filtering interpretations, Sampling rates, Spectrographic displays, Pitch and formant extraction, Analysis by Synthesis: Phase vocode, Channel Vocoder, Homomorphic speech analysis: Cepstral analysis of Speech, Formant and Pitch Estimation, Homomorphic Vocoders;

Basic principles of linear predictive analysis, the autocorrelation method, solution of LPC equations for auto correlation method, Application of LPC parameters: Pitch detection, formant analysis, CELP.

**MODULE -IV**: **VIDEO PROCESSING (09)**
Analog Video, Digital Video. Time-Varying Image Formation models: Three-dimensional Motion Models, Geometric Image Formation, Photometric Image Formation, Sampling of Video signals, Filtering operations.

**MODULE –V: MOTION ESTIMATION (09)**
Optical flow, General Methodologies, Pixel Based Motion Estimation Block Matching Algorithm, Mesh based Motion Estimation, Global Motion Estimation, Region based Motion Estimation, Multi resolution motion estimation, Waveform based coding, Block based transform coding, Predictive coding, Application of motion estimation in Video coding.

## V. TEXT BOOKS:


1. Lawrence R. Rabiner, Ronald W. Schafer, “Digital Processing of Speech Signals,” Pearson Education, 2012
2. Video Processing and Communication – Yao Wang, Joem Ostermann and Ya–quin Zhang. 1st Ed., PH Int.


## VI. REFRENCE BOOKS:


1. R. L. Rabiner, R.W. Schafer, “Digital Processing of speech signals”, Pearson Education.
2. B. Gold and Nelson Morgon, “Speech and audio signal processing”, Wiley India Edition,2006.
3. Dan Jurafsky and James H. Martin, “Speech and Language Processing”, 2nd Edition, Pearson, 2017
4. Digital Video Processing – M. Tekalp, Prentice Hall International.
5. Multidimentional Signal, Image and Video Processing and Coding – John Woods, 2nd Ed, Elsevier.
6. Video Demystified – A Hand Book for the Digital Engineer – Keith Jack, 5th Ed., Elsevier.


## VII. ELECTRONICS RESOURCES:

1. https://archive.nptel.ac.in/courses/117/105/117105145/

## VIII. MATERIALS ONLINE

1. Course template
2. Tutorial question bank
3. Definition and terminology
4. Tech-talk topics

5. Assignments
6. Model question paper - I
7. Model question paper - II
8. Lecture notes
9. Early learning readiness videos (ELRV)
10. Power point presentations