INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)
Dundigal - 500 043, Hyderabad, Telangana

COURSE CONTENT

SPEECH AND VIDEO PROCESSING								
VI Semester: CSE (AI&ML)								
Course Code	**Category**	**Hours / Week**			**Credits**	**Maximum Marks**		
		L	**T**	**P**	**C**	**CIA**	**SEE**	**Total**
ACAD15	**Elective**	3	0	0	3	40	60	100
Contact Classes: 48	**Tutorial Classes: NIL**	**Practical Classes: Nil**				**Total Classes: 48**		
Prerequisite: Machine Learning Algorithms								

I. COURSE OVERVIEW:

Speech processing is the study of speech signals and the processing methods of signals. The signals are usually processed in a digital representation, so speech processing can be regarded as a special case of digital signal processing, applied to speech signals.

II.COURSES OBJECTIVES:

The students will try to learn:

I. The fundamental concepts of speech production, signal modeling, and feature extraction techniques for speech recognition and analysis.
II. The principles of video processing, including motion estimation, image segmentation, and 3D structure reconstruction from visual data.
III. Practical skills in applying computational models and algorithms for analyzing, processing, and interpreting speech and video signals in real-world applications.

III. COURSE OUTCOMES:

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

CO1 Apply Fourier transform properties, filter banks, and linear prediction analysis for effective feature extraction in speech signal processing.
CO2 Apply distance measures, vector quantization, Gaussian Mixture Models (GMM), and Hidden Markov Models (HMM) for effective speech pattern modelling and classification.
CO3 Analyze multi-dimensional signals and systems using appropriate transforms and sampling theories for signal representation and processing.
CO4 Implement nonlinear optimization techniques and 3D motion and structure estimation for advanced image and video analysis tasks.
CO5 Understand image formation and motion models, and apply differential, matching, and transform-domain methods for 2D apparent motion estimation.
CO6 Evaluate the performance of segmentation and matting methods to enhance the accuracy and quality of visual information processing.

IV.COURSE CONTENT:

MODULE – I: SPEECH PROCESSING CONCEPTS (09)

The speech production mechanism, Discrete time speech signals, Pole-Zero modeling of speech, relevant properties of the fast Fourier transform for speech recognition, convolution, linear and nonlinear filter banks, spectral estimation of speech using DFT, Linear Prediction analysis of speech.

MODULE – II: SPEECH RECOGNITION (09)

Feature extraction for speech, static and dynamic feature for speech recognition, MFCC, LPCC, Distance measures, vector quantization models, Gaussian Mixture model, HMM.

MODULE – III: MULTI-DIMENSIONAL SIGNALS AND SYSTEMS (09)

Multi-Dimensional Signals, Multi-Dimensional Transforms, Multi-Dimensional Systems, Multi-Dimensional Sampling Theory, Sampling Structure Conversion Digital Images and Video: Human Visual System and Color, Digital Video.

MODULE – IV: MOTION ESTIMATION (09)

Image Formation, Motion Models, 2D Apparent-Motion Estimation, Differential Methods, Matching Methods, Nonlinear Optimization Methods, Transform-Domain Methods, 3D Motion and Structure Estimation

MODULE – V: VIDEO SEGMENTATION AND TRACKING (09)

Image Segmentation, Change Detection, Motion Segmentation, Motion Tracking, Image and Video Matting, Performance Evaluation.

V. TEXTBOOKS:

1. Fundamentals of Speech recognition – L. Rabiner and B. Juang, Prentice Hall signal processing Series.
2. Digital Video processing, A Murat Tekalp, 2nd edition,Prentice Hall.

VI. REFERENCE BOOKS:

1. Discrete-time speech signal processing: principles and practice, Thomas F. Quatieri, Coth.
2. Video Processing and Communications, Yao Wang, J. Osternann and Qin Zhang, Pearson Education
3. “Speech and Audio Signal Processing”, B.Gold and N. Morgan, Wiley.
4. “Digital image sequence processing, Compression, and analysis”, Todd R. Reed, CRC Press
5. “Handbook of Image and Video processing”, Al Bovik, Academic press, second Edition.

VII. ELECTRONICS RESOURCES

1. https://github.com/weimeng23/speech-recognition-learning-resources
2. https://www.vssut.ac.in/lecture_notes/lecture1423722885.pdf
3. https://www.academia.edu/394319/video_and_image_processing

VIII. MATERIAL ONLINE:

1. Course template
2. Tutorial question bank
3. Tech talk topics
4. Open-ended experiments
5. Definitions and terminology
6. Assignments
7. Model question paper – I
8. Model question paper – II
9. Lecture notes
10. PowerPoint presentation
11. E-Learning Readiness Videos (ELRV)