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

DEFINITION AND TERMINOLGY

Department	**Computer Science and Engineering (AI and ML)**				
Course Title	**Image and Speech Processing**				
Course Code	ACAC05				
Program	B.Tech				
Semester	V				
Course Type	Core				
Regulation	UG-20				
Course Structure	Theory			Practical	
	Lecture	Tutorials	Credits	Laboratory	Credits
	3	-	3	-	-
Course Coordinator	Dr Shaik Jakeer Hussain, Associate Professor				

COURSE OBJECTIVES:

The students will try to learn:

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

COURSE OUTCOMES:

After successful completion of the course, students should be able to:

CO 1	**Make use of** image transform techniques for analyzing images in transformation domain for image pre-processing.	Understand
CO 2	**List** the lossy and lossless compression models for achieving image compression.	Analyze
CO 3	**Illustate** the difference between acoustic phonetics and articulatory phonetics for speech processing	Understand
CO 4	**Utilize** digital model designed by sampled speech signal for speech processing applications like speech recognition, speech synthesis and verification.	Apply

CO 5	**Analyze** methods to estimate pitch period to design vocoders,aritificial intelligence voice-controlled assistants like Alexa	Analyze
CO 6	**Apply** linear predictive coding for speech synthesis, compression and spectrographic displays	Apply

DEFINITION AND TERMINOLOGY:

S.No	DEFINITION	CO's
	MODULE I	
	ROCKET DYNAMICS	
1	**Define Pixel of a digital image**	CO 1
	Pixel is the smallest element of an image .	
2	**Define picture element**	CO 1
	Picture element is the smallest discrete component of an image.	
3	**What is an image.**	CO 1
	An image is an array, or a matrix, of square pixels arranged in columns and rows.	
4	**What is image processing**	CO 1
	Image processing is a method to perform some operations on an image.	
5	**Define region in image processing**	CO 1
	A region in an image is a group of connected pixels with similar properties.	
6	**Define Image acquisition**	CO 1
	Image acquisition is the creation of a digitally encoded representation of the visual characteristics of an object	
7	**Define Dynamic Range**	CO 1
	The range of values spanned by the gray scale is called dynamic range of an image .	
8	**What are 2D and 3D images?**	CO 1
	2D is defined as having two axes to plot, usually the x-axis and y-axis., 3D indicates three aspects, which are plotted on the x-axis, y-axis and the z-axis..	
9	**How do you define high contrast of a digital image**	CO 1
	When in an Image anappreciable number of pixels exhibit high dynamic range, the image will have high contrast. .	
10	**Define Grid of digital image**	CO 1
	The sampling points are ordered in the Plane and their relation is called a Grid.	

11	**How is contrast defined in terms of intensity**	CO 1
	It is defined as the difference in intensity between the highest and lowest intensity levels in an image	
12	**What is meanby Gray level?**	CO 1
	Gray level refers to a scalar measure of Intensity that ranges from black to white.	
13	**Define SensorStrips?**	CO 1
	The sensors for image acquisition/Sensor strips are commonly used for in-Line arrangement in imaging geometry.	
14	**What do you meant by Color model**	CO 1
	A Color model is a specification of 3D- Coordinates system and a subspace within that system where each color is represented by a single point.	
15	**Give an application of filter in image processing.**	CO 2
	Filtering is a technique for modifying or enhancing an image	
16	**How is scanner used in image acquistion.**	CO 1
	Scanner is a device that optically scans images, printed text, handwriting or an object and converts it to a digital image.	
17	**What is digitization process?**	CO 1
	The digitization process i.e. the digital image has M rows and N columns, requires decisions about values for M, N, and for the number, L, of gray levels allowed for each pixel.	
18	**Define analog image with examples**	CO 1
	Analog image is a continuous variation Examples of analog images are television images, photographs, paintings, and medical images.	
19	**What is a booster stage?**	CO 1
	The first stage lifts off the entire rocket vehicle system, therefore it is the most powerful stage and is known as the booster stage.	
20	**Specify what is interpretation**	CO 1
	The interpretation is called the assigning to recognize object.	
21	**Give the role of a sensor in image processing**	CO 1
	A sensor is a device that detects and responds to some type of input from the physical environment.	
22	**What is image acquisition**	CO 1
	It is defined as the action of retrieving an image from some source.	

23	**Interpret what is 4 adjacency** Two pixels p and q with values from V are 4-adjacent if q is in the set N4(p).	CO 1
24	**How is an image identified as digital image** Digital image can be defined by a two-dimensional array specifically arranged in rows and columns.	CO 1
25	**What is meant by brightness** Brightness refers to the overall lightness or darkness of the image.	CO 2
26	**Define gray level resolution** Gray Level Resolution can be defined as the total number of pixels in an image.	CO 1
27	**What is 8 adjacency in digital image** Two pixels p and q with values from V are 8-adjacent if q is in the set N8(p).	CO 1
28	**What are the steps involved in DIP?** 1. Image Acquisition 2. Preprocessing 3.Segmentation 4. Representation and Description 5. Recognition and Interpretation	CO 1
29	**What is sampling?** Sampling is the process of converting continuous time signal into a discrete time signal.	CO 1
30	**What is quantization?** To convert a continuous sensed data into Digital form.	CO 2
31	**Define encoding of a sampled signal** Encoding is the process of converting data from one form to another	CO 1
32	**What is mean by coordinates?** To convert a continuous sensed data in to Digital form.	CO 1
33	**What is quantization?** An image may be continuous in the x-and y-coordinates or in amplitude, or in both.	CO 1
34	**Define adjacency of pixels** Two pixels are connected if they are. Neighbors and their gray levels satisfy some specified criterion of similarity is called adjacency.	CO 1
35	**What is intensity value of a pixel** A pixel is a small block that represents the amount of gray intensity to be displayed for that particular portion of the image.	CO 1
36	**Write the difference between sampling and quantization** Sampling: It determines the spatial resolution of the digitized images. Quantization: It determines the number of grey levels in the digitized images.	CO 1
37	**What is mean by pixel connectivity** Pixel connectivity is the way in which pixels in 2-dimensional 3-dimensional images relate to their neighbors.	CO 1

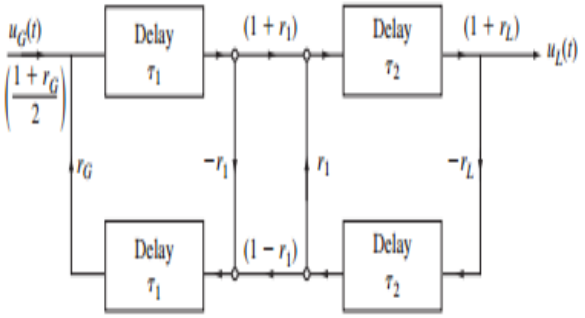
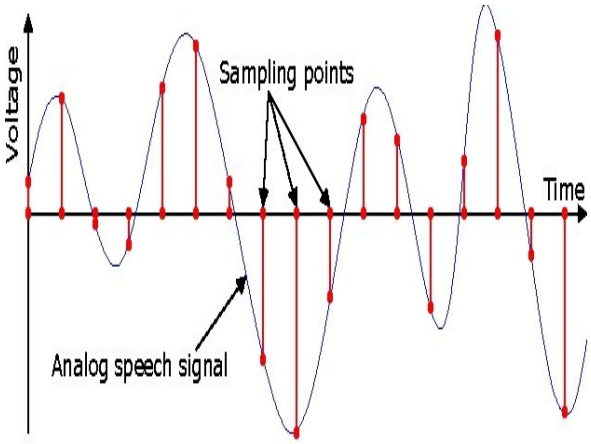
38	**What is scanner** A scanner is a device that captures images from photographic prints, posters, magazine pages, and similar sources for computer editing and display	CO 1
39	**Define m- adjacency** m-adjacency is a combination of 4 and 8 adjacency	CO 1
40	**Define image geometry** The geometric shape which appears after a transformation has been applied to the pre image.	CO 1
MODULE II		
Image Compression		
1	**What is the expanded form ofJPEG?** Image compression is familiar (perhaps inadvertently) to most users of computers in the form of image file extensions, such as the jpg file extension used in the JPEG (Joint Photographic Experts Group) image compression standard.	CO 2
2	**What is image compression?**	CO 2
	Image compression refers to the process of redundancy amount of data required to represent the given quantity of information for digital image.	
3	**What are two main types of Data compression?**	CO 2
	Lossless compression can recoverthe exact original data after compression. It is used mainly for compressing database records, spreadsheets or word processingFiles, where exact replication of the original is essential.	
4	**What is the need for compression?** In terms of storage, the capacity of aStorage device can be effectively increased with methods that compress a body of data on its way to a storage device and decompress it when it is retrieved.	CO 2
5	**What is Data compression?** Data compression requires the identification and extraction of source Redundancy. In other words, data compression seeks to reduce the number of bits used to store or transmit information.	CO 2
6	**What are different Compression Methods?** Run Length Encoding (RLE) Arithmetic coding Huffman coding and Transform coding	CO 2
7	**Define is coding redundancy** If the gray level of an image is coded in a way that uses more code words than necessary to represent each gray level, then the resulting image is said to contain coding redundancy.	CO 2

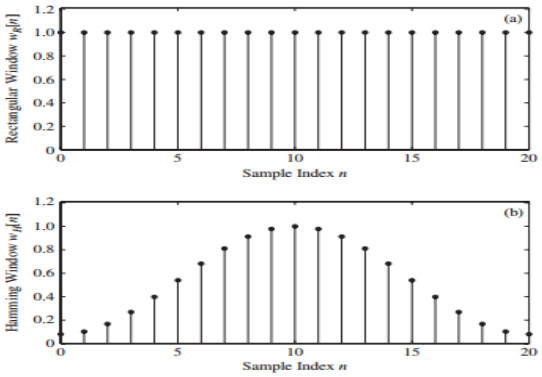
8	**Define inter pixel redundancy**	CO 2
	The value of any given pixel can be predicted from the values of its Neighbors. The information carried by is small. Therefore the visual contribution of a single pixel to an image is redundant. Otherwise called as spatial redundant geometric redundant or inter pixel redundant. Eg: Run length coding	
9	**What is run length coding?**	CO 2
	Run-length Encoding, or RLE is a Technique used to reduce the size of a repeating string of characters. This repeating string is called a run; typically RLE encodes a run of symbols into two bytes, a count and a symbol.	
10	**Define compression ratio.**	CO 2
	Compression Ratio = original size /compressed size	
11	**Define psychovisual redundancy**	CO 2
	In normal visual processing certain information has less importance than other information. So this information is said to be psycho visual redundant..	
12	**Give the types of encoder and specify its uses**	CO 2
	Source encoder is responsible for Removing the coding and interpixel redundancy and psycho visual redundancy. There are twocomponentsA) Source Encoder B) Channel Encoder	
13	**Operation of source encoder**	CO 2
	Source encoder performs three operations 1) Mapper -this transformstheInput data into non-visual format. It reduces the interpixel redundancy.	
14	**What is channel encoder**	CO 2
	The channel encoder reduces the Impact of the channel noise by inserting redundant bits into the source encoded data.	
15	**What are the types of decoder?**	CO 2
	ASource decoder- has two components a) Symbol decoder- b)Inverse mapping- c) Channel decoder	
16	**What operations are performed by error free compression?**	CO 2
	Devising an alternative representation of the image in which its inter pixel redundant are reduced. Coding the representation to eliminate coding redundancy	
17	**What is Variable Length Coding?**	CO 2
	Variable Length Coding is the simplestapproach to error free compression. It reduces only the coding redundancy. It assigns the shortest possible codeword to the most probable gray levels.	
18	**Define Huffman coding and mention its limitation**	CO 2
	Huffman coding is a popular technique for removing coding redundancy.	

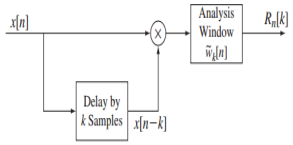
19	**Define Block code** Each source symbol is mapped into fixed sequence of code symbols or code words. So it is called as block code	CO 2
20	**Define instantaneous code** A code word that is not a combination of any other codeword is said to be uniquely decodable code.	CO 2
21	**Define uniquely decodable code** A code word that is not a combination of any other codeword is said to be uniquely decodable code	CO 2
22	**Define B2 code** Each code word is made up of continuation bit c and information bit which are binary numbers. This is called B2 code or B code. This is called B2 code because two information bits are used for continuation bits	CO 2
23	**Define the procedure for Huffman shift coding** List all the source symbols along with its probabilities in descending order. Divide the total number of symbols into block of equal size. Sum the probabilities of all the source symbols outside the reference block.	CO 2
24	**Define arithmetic coding** In arithmetic coding one to one corresponds between source symbols and code word does't exists where as the single arithmetic code word assigned for a sequence of source symbols. A code word defines an interval of number Between 0 and 1.	CO 2
25	**What is bit plane Decomposition?** An effective technique for reducing an image's inter pixel redundancies is to process the image's bit plane individually	CO 2
26	**How effectiveness of quantization can be improved?** 1. 1. Introducing an enlarged quantization Interval around zero, called a dead zero. 2.Adapting the size of the quantization intervals from scale to scale. In either case, the selected quantization intervals must be transmitted to the decoder with the encoded image bitstream	CO 2
27	**What are the coding systems in JPEG?** 1. A lossy baseline coding system,which is based on the DCT and is adequate for most compression application. 2. An extended coding system for greater compression, higher precision or Progressive reconstruction applications. 3. A lossless independent coding system for reversible compression.	CO 2
28	**What is JPEG?** The acronym is expanded as "Jointhotographic Expert Group". It is an international standard in 1992.	CO 2

29	**What are the basic steps in JPEG?**	CO 2
	The Major Steps in JPEG Coding involve: 1. DCT 2. Quantization 3. Zigzag Scan 4. DPCM on DC component 5. RLE on AC Components 6. Entropy Coding	
30	**What is MPEG?**	CO 2
	The acronym is expanded as "Moving Picture Expert Group". It is an international standard in 1992. It perfectly Works with video and also used in teleconferencing	
31	**Define I-frame**	CO 2
	I-frame is Intraframe or Independentframe. An I-frame is compressed independently of all frames. It resembles a JPEG encoded image.	
32	**Define P-frame**	CO 2
	P-frame is called predictive frame. A P-frame is the compressed difference between the current frame and a prediction of it based on the previous I or P-frame	
33	**Define B-frame**	CO 2
	B-frame is the bidirectional frame. AB-frame is the compressed difference between the current frame and a prediction of it based on the previous I or P-frame or next P-frame.	
34	**What is shift code?**	CO 2
	The two variable length codes are referred to as shift codes. A shift code is generated by i)Arranging probabilities of the source symbols are monotonically decreasing.	
35	**What are the types of redundancy?**	CO 2
	1. Coding Redundancy 2. Interpixel Redundancy 3. Psychovisual Redundancy	
36	**Define Psychovisual redundancy.**	CO 2
	Certain information which has lessrelative importance than other information in normal visual processing are said to be psychovisually redundant information	
37	**What is image compression?**	CO 2
	Image compression refers to the process of redundancy amount of data required to represent the given quantity of information for digital image.	
38	**What is Data Compression?**	CO 2
	Data compression requires the Identification and extraction of source redundancy. In other words, data compression seeks to reduce the number of bits used to store or transmit information	
39	**What are two main types of Data compression?**	CO 2
	1. Lossless compression 2. Lossy compression	

40	**What is the need for Compression?** In terms of communications, the bandwidth of a digital communication link can be effectively increased by compressing data at the sending end and decompressing data at the receiving end.	CO 2
MODULE III		
Fundamentals of Speech Processing		
1	**What are phonetics in a language** Phonetics is the study and classification of speech sounds	CO 3
2	**Identify are the parts of vocal tract** Vocal tract consists of pharynx and mouth or oral cavity	CO 3
3	**Define the range of length of vocal tract in males.** Vocal tract in male varies from 17-17.5 cm	CO 3
4	**Specify the part of human body are involved in sound production?** Mouth, tongue, velum,epiglotis,esophagus,glottis and trachea are the parts which are involved in speech production	CO 3
5	**How do you define speech sound?** Speech sound is defined as ever changing sounds	CO 3
6	**How are the voiced signals produced.** Voiced sounds are produced by forcing air through the glottis with the tension of the vocal cords adjusted so that they vibrate in a relaxation oscillation, thereby producing quasi-periodic pulses of air that excite the vocal tract, leading to a quasi-periodic waveform	CO 5
7	**When are the unvoiced signals produced?** Unvoiced or fricative sounds are generated by forming a partial constriction at some point in the vocal tract (usually toward the mouth end), and forcing air through the constriction at a high enough velocity to produce turbulence.	CO 3
8	**How is the silence sound identified?** silence or background sounds are identified by their lack of the characteristics of either voiced or unvoiced sounds and usually occur at the beginning and end of speech utterances	CO 3 CO 3
9	**What is the phonetic representation of 'should we chase'** This example 'shpould we chase' is phonetically represented as /SH UH D - W IY - CH EY S/.	CO 3
10	**If excitation signal is e(t) and vocal tract impulse response is v(t) then what is the speech waveform?** The speech wavefoorm is defined as the convolution of the impulse response and excitation signal in time domain and multiplication in frequency domain.	CO 3

11	**How many phenomees are existing in the English language?**	CO 3
	In english language, there are 48 phonetics which can be further divided into 5 broad classes.	
12	**How many vowels, dipthongs exists in english language**	CO 3
	The english language has 11 vowels and 4 dipthongs.	
13	**What do you mean by formant frequency?**	CO 3
	Formant frequency is defined as the distinctive frequency components of acoustic signal produced by speech.	
14	**Draw the complete flow graph of two-tube model**	CO 3
	 FIGURE 5.30 Complete flow diagram of a two-tube model.	
MODULE IV		
Time Domain Models for speech processing		
1	**Represent sampling of a speech signal by waveforms**	CO 5
		

2	**Illustrate the rectangular window and Hamming window through graphs**	CO 5
	 FIGURE 6.6 Plots of the time responses of a 21-point (a) rectangular window; (b) Hamming window.	
3	**How do you define energy of a discrete time signal?**	CO 5
	Energy of discre time signal is given by $$E = \sum_{m=-\infty}^{\infty} (x[m])^2.$$	
4	**Explain why do you require automatic gain control.**	CO 5
	The purpose of an AGC is to keep the signal amplitude as large as possible without saturating or overflowing the allowable dynamic range of a digital representation of the speech samples	
5	**Give the mathematical representation of exponentail window**	CO 5
	$$\tilde{w}[n] = (1-\alpha)\alpha^{n-1}u[n-1] = \begin{cases} (1-\alpha)\alpha^{n-1} & n \geq 1 \\ 0 & n < 1, \end{cases}$$.	
6	**How do you define short time zero crossing?**	CO 5
	It is defined as the number of times the speech signal changes sign within given time window.	
7	**How is short time magnitude better than shprt time energy**	CO 5
	For the short-time magnitude computation dynamic range (ratio of maximum to minimum) is approximately the square-root of the dynamic range for the standard energy computation. Thus the differences in level between voiced and unvoiced regions are not as pronounced as for the short-time energy	

8	**Give the formula for calculating signal frequency from zero crossing rate**	CO 5
	$$F_e = 0.5F_sZ^{(1)},$$	
9	**Draw the diagram for hort time autorrelation for lag index k**	CO 2
		
10	**How is AMDF used speech processing**	CO 5
	AMDF stands for Average Magnitude Difference Function and it is a variation of autocorrelation analysis for measuring the periodicity of voiced speech and music signals..	
11	**What is pitch of speech signa;**	CO 5
	Pitch, in speech, the relative highness or lowness of a tone as perceived by the ear, which depends on the number of vibrations per second produced by the vocal cords	
12	**Explain why pitch period caluclation is important in speech processing?**	CO 8
	Unless specified, the term 'pitch' refers to the fundamental frequency ' Fo'. Pitch is an important attribute of voiced speech. It contains speaker-specific information. It is also needed for speech coding task. Thus estimation of pitch is one of the important issue in speech processing.	
MODULE V		
Short time fourier analysis and linear predictiv coding		
1	**What are spectrograpgic displays**	CO 6
	A spectrogram is a visual representation of the spectrum of frequencies of a signal as it varies with time. When applied to an audio signal, spectrograms are sometimes called sonographs, voiceprints, or voicegrams.	
2	**How are the spectrograms of speech signal differentiated?**	CO 6
	We use two types of spectrogram for speech study: one which emphasises the frequency aspects by using long signal sections or narrow analysis filters, and one which emphasises the temporal aspects by using short signal sections or wide analysis filters.	

3	**Why is a spectrogram useful??**	CO 5
	A spectrogram is most helpful for vibration analysis in a changing environment. It illustrates the patterns of energy change which may not be visible in an FFT or PSD. In comparison to an FFT, a spectrogram gives a better look into how the vibration changes over time	
4	**What are filter banks in speech processing**	CO 6
	In signal processing, a filter bank (or filterbank) is an array of bandpass filters that separates the input signal into multiple components, each one carrying a single frequency sub-band of the original signal	
5	**List out five reasons why linear preductive analysis is required for speech processing?**	CO 6
	he predominant technique for estimating the parameters of the discrete-time model for speech production (i.e., pitch, formants, short-time spectra, vocal tract area functions) and is widely used for representing speech in low bit rate transmission or storage and for automatic speech and speaker recognition. The importance of this method lies both in its ability to provide accurate estimates of the speech parameters and in its relative ease of computation	
6	**List out different formulaitions of LPC**	CO 6
	LPC formulations are 1.the covariance method 2. the autocorrelation formulation 3. the lattice method 4. the inverse filter formulation 5. the spectral estimation formulation 6. the maximum likelihood formulation 7. the inner product formulation .	
7	**Specify the methods for obtaining predictor coefficients**	CO 6
	The three methods are (1) The Cholesky Decomposition (2) The Levinson–Durbin Algorithm (3)Lattice Formulations and Solutions	
8	**How do you define predictor error signal**	CO 6
	Predictor error signal is defined as $$e[n] = s[n] - \sum_{k=1}^{p} \alpha_k s[n-k],$$	
9	**Define what is LPC analysis**	CO 10
	LPC (linear predictive coefficients) analysis is a technique for estimating the vocal tract transfer function, from which its poles, he formant frequencies, can be analytically calculated.	

10	**Write down what is covariance mthod**	CO 10
	the signal is extended by samples outside the normal range of 0¡m¡L-1 to include p samples occurring prior to m=0.This eliminates the need for a tapering window;.	

Course Coordinator: **HOD CSE (AI&ML)**
Dr Shaik Jakeer Hussain, CSE(AI&ML)