Nadar Saraswathi College of Engineering & Technology

NSCET E-LEARNING PRESENTATION

LISTEN … LEARN… LEAD…

ELECTRONICS & COMMUNICATION ENGINEERING

IV YEAR / VIII th SEMESTER

EC 6018 – MULTIMEDIA COMPRESSION & COMMUNICATION

S.PRATHAP M.E.,

Assistant professor

Nadar Saraswathi College of & Technology,

Vadapudupatti, Annanji (po), Theni – 625531.

UNIT-3

Audio and Video Compression

Digital Audio

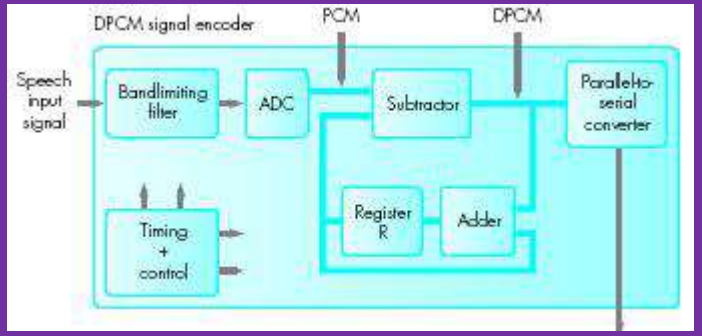
Human auditory system is much more sensitive to quality degradation then is the human visual system

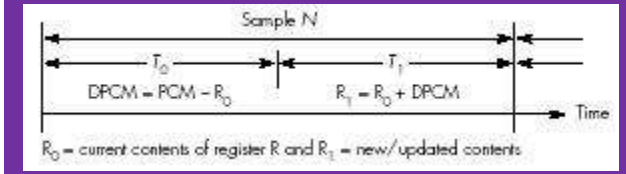
- redundancy is relatively small
- can achieve lower compression ration than for digital video

Digital audio stream

- 2 dimensions: amplitude x time
- c.f.) video stream: 3 dimensions

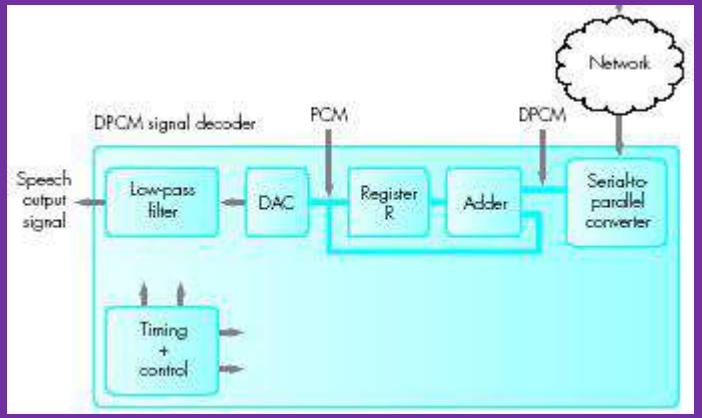
Differential PCM





Third order predictive coding

- Compute difference from last 3 predictive value
- Predictor coeffciet value: 0.5, 0.25, 0.25
- 32kbps



Variants of PCM

Basic compression in digital telephony: (G.711)

- A-law transformation: 13 bits 8 bits (Europe)
- -law transformation: 14 bits 8 bits (North-America)
- Non-liner quantization
- greater accuracy in low-amplitude samples

Differential PCM

- 8bit 6bit

ADPCM (G.721)

- adapts to the characteristics of signals, by
- changing quantization step size
- prediction
- 8bits 4 bits

Adaptive predictive coding (ADC)

- Predictor coefficients are continuously changed

Linear predictive coding (LPC)

- Exploit perceptual features
- Pitch: ear is more sensitive to 2-5kHz
- Period: duration of signal
- Loudness: energy of signal

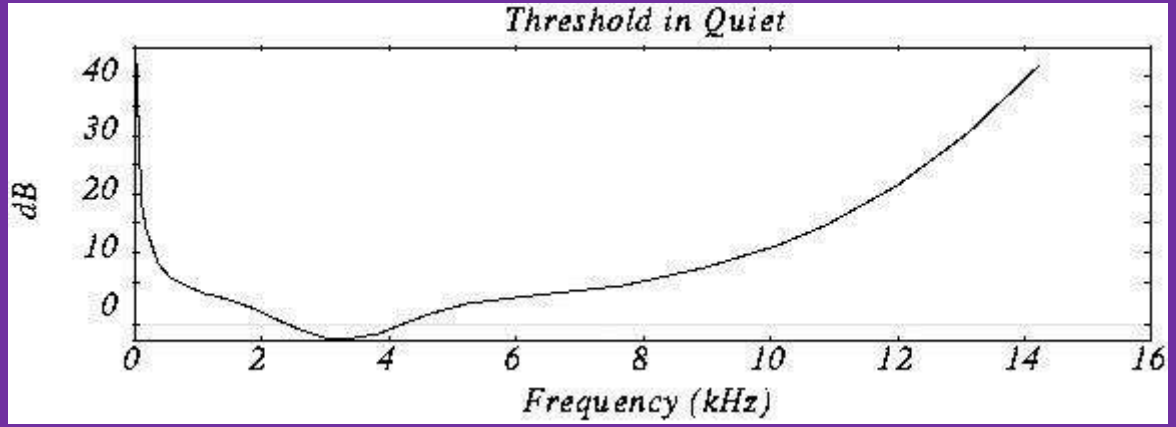
Code-excited LPC (CELP)

- Use waveform template
- G.728 16kbps 0.625ms low bit rate telephony
- G.729 8kbps 25ms cellular network
- G.729(A) 8kpbs 67.5ms digital simultaneous voice and data
- G.723.1 5.3/6.3kbps 67.5ms video/Intenet telephony

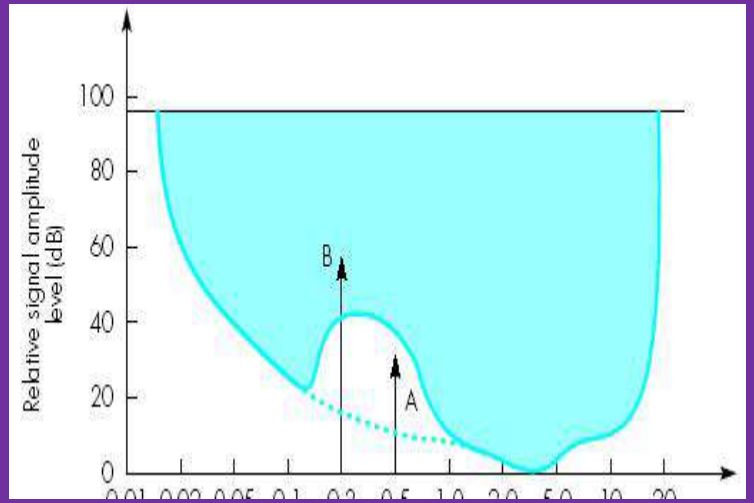
Psycho-Acoustics

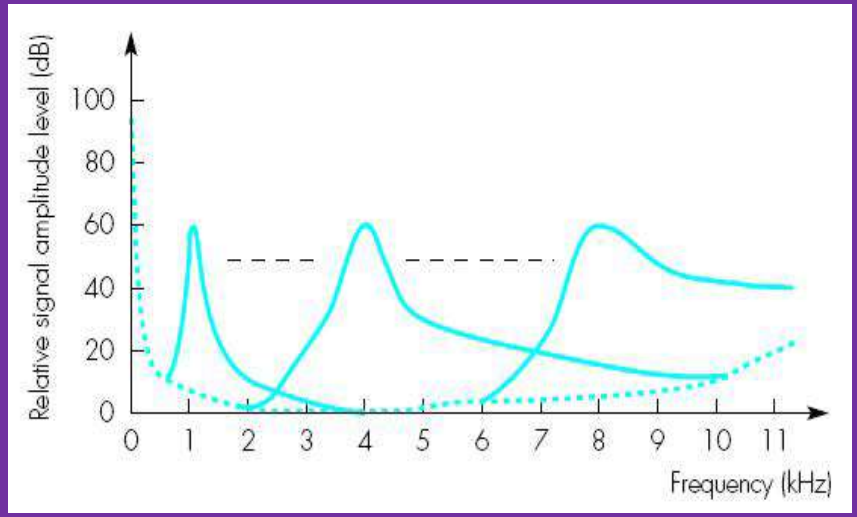
- Human hearing and voice
  - Audible tone: 15 Hz - 20 KHz
  - Most sensitive at 2KHz to 5 KHz
  - Dynamic range (quietest to loudest) is about 96 dB
  - Normal voice range is about 500 Hz to 2KHz
- Low frequencies are vowels and bass
- High frequencies are consonants

Hearing threshold of human hear



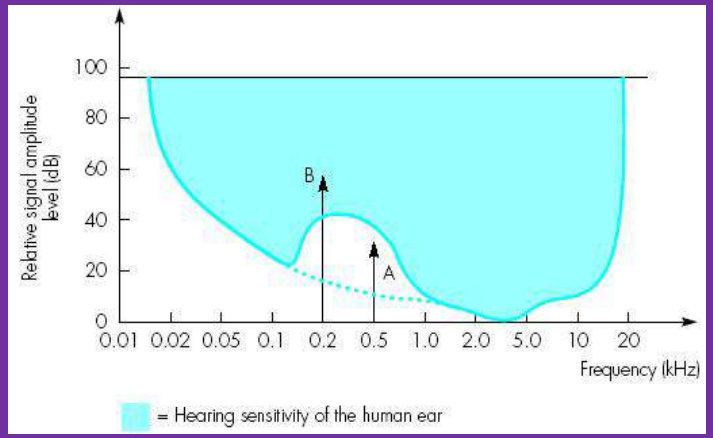
Frequency Masking





Frequency Masking

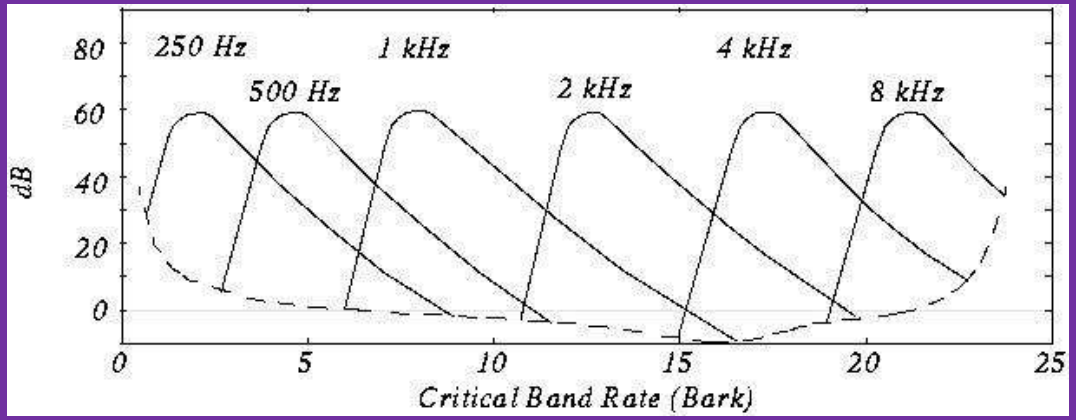
- A strong audio signal makes a spectral neighborhood of weaker audio signals imperceptible
- The threshold at any given frequency is dependent on the signal activity within a critical band of that frequency



Critical Bands

- Perceptually uniform measure of frequency
- Bark
- □ new unit for frequency
- □ 1 Bark = width of one critical band
- □ for frequency < 500 Hz, 1 Bark = freq/100
- □ for frequency > 500 Hz, 1 Bark = 9 + 4log(freq/1000)

Masking Thresholds on critical band scale



Temporal Masking

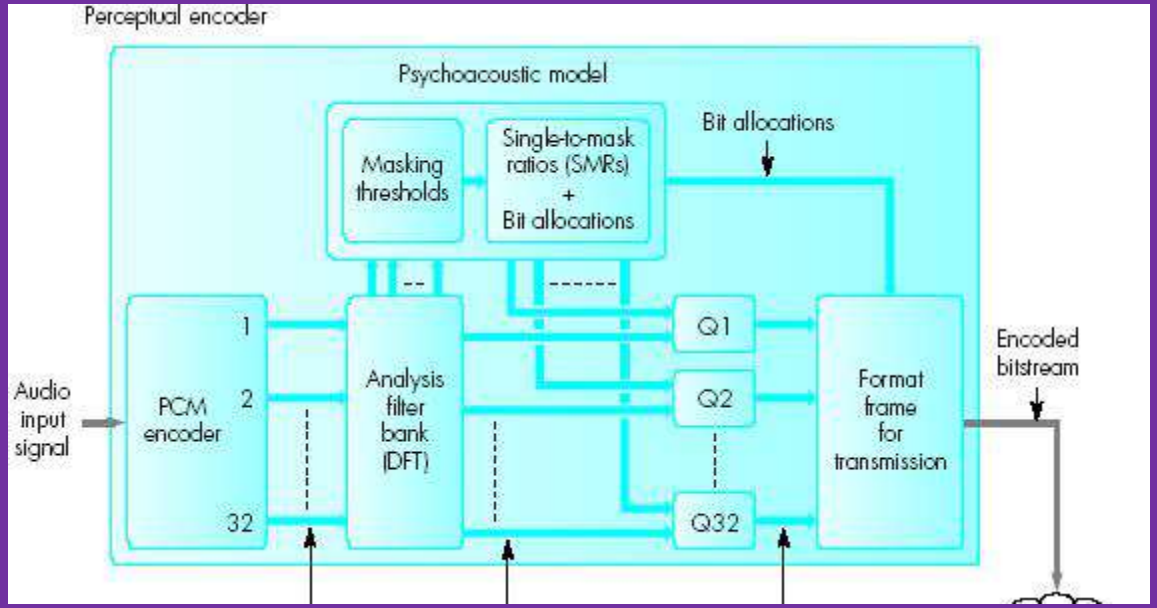
- If we hear a loud sound, then it stops, it takes a little while until we can hear a soft tone nearby.

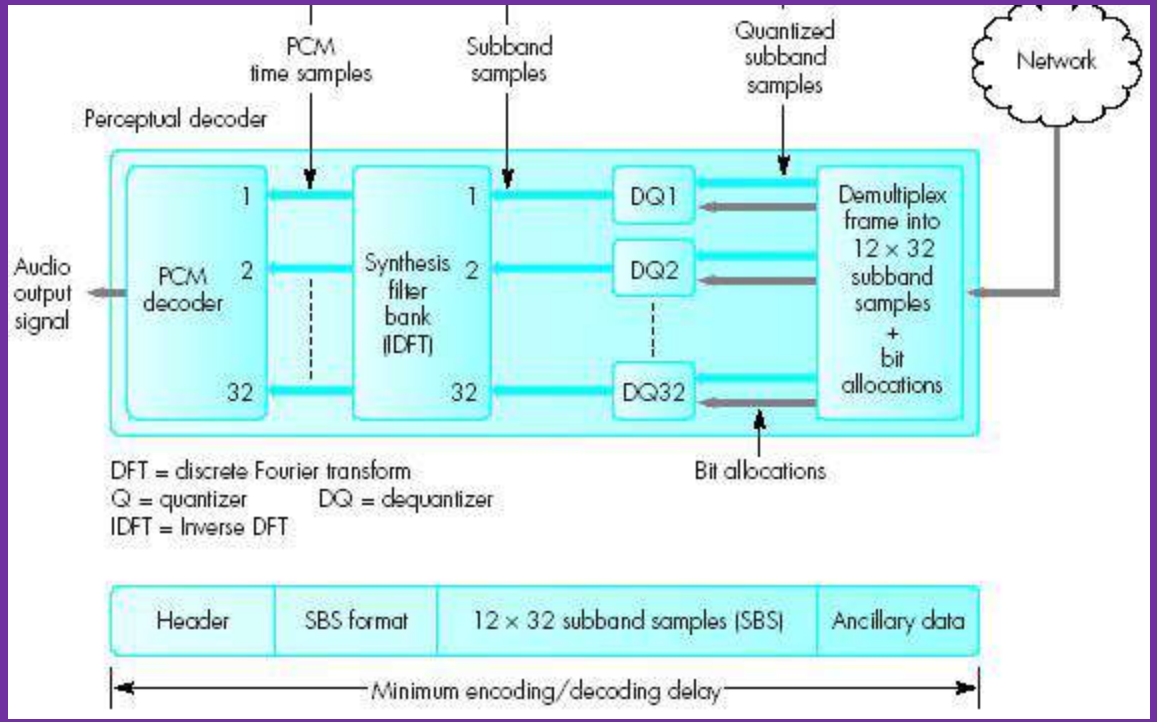
MPEG Audio

- Lossy compression but perceptually lossless
- Identify and remove perceptually irrelevant parts of the audio signals
- exploiting limitation of human audible system
  - hearing threshold, auditory masking
- MPEG-1 Audio
  - supports 1 or 2 audio channels
- MPEG-2 Audio
  up to 5 channels (left, right, center, 2 surrounding)
- a low-frequency enhancement channel
- up to 7 commentary/multilingual channels
- backward compatible with MPEG-1

MPEG-1 Audio

- Divide audio signals into frequency subbands that approximates critical subbands
- For each subband (bandwidth = 500Hz) sample and accumulate 12 successive set of 32 samples
- 12 x 32 PCM samples
- Quantize each subband according to the audibility of quantization noise
- to make quantization noise in audible
- Adaptive bit allocation
- to control the spectral properties of the noise as a function of a signal so that the noise remains
- imperceptible
- Uses Psycho-Acoustic modeling


Perceptual encoder
Psychoacoustic model
Masking thresholds
Single-to-mask ratios (SMRs) + Bit allocations
Bit allocations
Audio input signal
PCM encoder
1
2
32
Analysis filter bank (DFT)
Q1
Q2
Q32
Format frame for transmission
Encoded bitstream


PCM
time samples
Subband
samples
Quantized
subband
samples
Network
Perceptual decoder
1
2
32
PCM
decoder
Audio
output
signal
1
2
32
Synthesis
filter
bank
(IDFT)
DQ1
DQ2
DQ32
Demultiplex
frame into
12 × 32
subband
samples
+
bit
allocations
Bit allocations
DFT = discrete Fourier transform
Q = quantizer
DQ = dequantizer
IDFT = Inverse DFT
Header
SBS format
12 × 32 subband samples (SBS)
Ancillary data
Minimum encoding/decoding delay

MPEG Audio Layers

- Higher layer must be able to decode audio signals of lower layers
- Layer I
  - DCT type filter with one frame and equal frequency spread per band. (32 constant-width frequency bands)
  - Psychoacoustic model only uses frequency masking.
  - Sampling rates used
    - 32ksps for broadcast communication equipment
    - 44.1ksps for CD-quality audio equipment
    - 48ksps for professional sound equipment
- Layer II
  - Use 3 frames in filter. This models temporal masking.
- Layer III
  - Better critical band filter is used(non-equal frequency).
  - Psychoacoustic model : temporal masking effects
  - Stereo redundancy, Huffman coder

Dolby Audio Coders

- Dolby AC-1 (Acoustic coder)
  - Designed for use in satellites to relay FM radio and TV
  - Fix bit allocation
  - 40 subbands at 32ksps, 44.1ksps, 48ksps
  - Typical compressed bit rate: 512kbps for stereo
- Dolby AC-2
  - Designed for PC sound cards
  - Hi-fi quality at a bit rate of 256kbps
- Dolby AC-3
  - Designed for a similar range of MPEG audio standards
  - HDTV standard in North America
  - 32 / 44.1 /48ksps
  - 192kbps for stereo

Video Compression Standards

- JPEG: ISO and ITU-T
  - ➢ for compression of still image
- Moving JPEG (MJPEG)
- H.261: ITU-T SG XV
  - ➢ for audiovisual service at p x 64Kbps
- MPEG-1, 2, 4, 7: ISO
- IEC/JTC1/SC29/WG11
  - ➢ for compression of combined video and audio
- H.263: ITU-T SG XV
  - ➢ for videophone at a bit-rate below 64Kbps
- JBIG: ISO
  - ➢ for compression of bi-level images

History of Video Compression Standards

Year	Standard	Publisher	Popular Implementations
1984	H.120	ITU-T	
1990	H.261	ITU-T	Videoconferencing, Videotelephony
1993	MPEG-1 Part 2	ISO, IEC	Video-CD
1995	H.262/MPEG-2 Part 2	ISO, IEC, ITU-T	DVD Video, Blu-ray, Digital Video Broadcasting, SVCD
1996	H.263	ITU-T	Videoconferencing, Videotelephony, Video on Mobile Phones (3GP)
1999	MPEG-4 Part 2	ISO, IEC	Video on Internet (DivX, Xvid ...)
2003	H.264/MPEG-4 AVC	ISO, IEC, ITU-T	Blu-ray, Digital Video Broadcasting, iPod Video, HD DVD

Compressing Digital Video

- Exploit *spatial redundancy within frames (like JPEG:* transforming, quantizing, variable length coding)
- Exploit *temporal redundancy between frames*
  - Only the sun has changed position between these 2 frames

Thank You