Audio and Video compression

- Reduction of information in the audio signal
  - Differential pulse code modulation (DPCM)
  - Adaptive differential
  - Adaptive predictive coding (APC)
  - Linear predictive coding (LPC)
  - Perceptual features
  - Human ear
  - MPEG audio coders

- Video
  - I-frame and P-frame
  - Motion estimation
  - MPEG
Audio continuous signals

- After digitization comprise a continuous stream of digital values
- Each value representing the amplitude of a sample of the analog signal taken at repetitive time intervals

Reduction of information in the audio signal

- Either the audio signal is sampled at a lower rate (fewer bits per sample)
  - Quality of the decoded signal is lost
- Or compression algorithm is used
Differential pulse code modulation (DPCM)

- Most audio signals, the range of the differences and amplitude between successive samples of the audio waveform is less than the range of the actual sample altitudes
  - Only the digitized difference signal (residual) is used to encode the waveform
  - Additional savings by varying number of bits used for the difference signal
    - Fewer bits to encode smaller difference values than for larger values

Adaptive differential

- Additive technique known as subband coding
- The audio input signal is passed through two filters
  
  - Dividing into two separate equal-bandwidth signals
    - Lower subband signal
    - And upper subband signal
Lower subband signal (50Hz-3.5kHz)  
And upper subband signal (3.5kHz-7kHz)  
Each is than sampled and encoded independently  

**Sampling frequency**  
The sampling frequency or sampling rate $f_s$ is defined as the number of samples obtained in one second, or $f_s = 1 / T$  
- (T sampling interval)  
The sampling rate is measured in Hertz or in samples per second  
The sampling frequency should be more than twice the maximum frequency component of the signal  
Otherwise, the original signal's information may not be completely recoverable from the sampled signal
Lower subband (50Hz-3.5Hz)
- signal sampled at 8 kHz
- (kHz=kspks=k samples/sec)

Upper subband signal (3.5KHz-7kHz)
- sampled at 16 kHz (kspks) to allow the presence of higher frequency components

Bit rate
- The use of two subbands has the advantage that different bit rates can be used for each
  - Bit rate per = sampling rate/s * bits per sample

- In general the frequency components that are present in the lower subband signal have a higher perceptual importance than those in the higher subband
Multiplexer

- The two bit streams are then multiplexed (merged) together in such a way that the decoder is able to divide them back.
Adaptive predictive coding (APC)

- Additional savings by varying number of bits used for the difference signal
- Fewer bits to encode smaller difference values than for larger values
Linear predictive coding (LPC)

- Analyzing the audio waveform to determine a selection of the perceptual features
- These are then quantized
- For decoding a sound synthesizer regenerates the sound

Perceptual features

- Pitch: closely related to the frequency of the signal and is important because the ear is more sensitive to frequencies in the range 2-5 kHz than to the frequencies that a higher or lower than these (diapasão)
  - span from the highest to the lowest note a particular voice
- Period: duration of the signal
- Loudness
- Vocal tract extraction parameters
  - origins of the sound
  - basic model of the vocal tract
Once these have been obtained from the source waveform, it is possible to use them together with a suitable model to generate a synthesized version of the original speech signal.
Perceptual encoders

- Perceptual encoders have been designed for the compression of general audio.
- They use a psychoacoustic model, its role is to exploit a number of the limitations of the human ear.
- Only these features that are perceptible to the ear are transmitted.
Human ear

- Human ear is sensitive to signals in the range 15 Hz through 20 kHz, the level of sensitivity to each signal is non-linear.

- The ear is more sensitive to some signals than others.

Sensitivity of the ear

- Maximal minimal amplitude of the signal is measured in decibel (dB)
  - Dynamic range of the sound a ear can hear
  - Ratio from the loudest to the quietest sound

- The region is around 96dB, however the sensitivity of the ear varies with the frequency of the signal.
Frequency masking

- When an audio sound consists of multiple frequency signals...
  - The sensitivity of the ear changes and varies with relative amplitude of the signals
  - The sensitivity of the ear changes in the vicinity of a loud signal

- Signal B is larger in amplitude than signal A
- Basic sensitivity curve of the ear is distorted in the region of the signal B
- Signal A will be not heard
The frequency masking varies with the frequencies.
The width of each curve is known as the critical band.
Temporal masking

- If a ear hears a loud sound it takes further short time before it can hear a quieter sound

![Graph showing temporal masking]

MPEG audio coders

- Motion Picture Expert Group (MPEG) was formed to formulate a set of standards relating to a range of multimedia applications
- The coders associated with the audio compression are known as MPEG audio coders
- They use the perceptual coding
- All the signal processing operations associated with the perceptual coder are carried out digitally
The signal is quantized, the sampling rate and the number of bits per sample being determined by the specific application.

The bandwidth that is available is divided into a number of frequency subbands using a bank of analysis filters (critical-band filters).

Each frequency subband is equal width, maps 32 samples into a set of 32 frequency samples one per subband.
- In basic encoder the time duration of each sampled segment of the audio input signal is equal to the time to accumulate 12 successive sets of 32 PCM (pulse code modulation, quantization)
In addition to the filtering the input samples into separate frequency subbands, the analysis filter bank also determines the maximum amplitude of the 12 subband samples in each subband.

The processing associated with both frequency and temporal masking is carried out by the psychoacoustic model which is performed currently with the filtering operations.

The 12 sets of 32 PCM samples are transformed into equivalent set of frequency components using discrete Fourier transform (DFT).

Then using the known hearing thresholds and masking properties of each subbands the model determines the various masking effects of this set of signals.

The output of the model indicate those frequency components whose amplitude is below the related audible threshold.
- Header contains information such as the sampling frequency
- SBS = subband samples

<table>
<thead>
<tr>
<th>Header</th>
<th>SBS format</th>
<th>$12 \times 32$ subband samples (SBS)</th>
<th>Ancillary data</th>
</tr>
</thead>
</table>

- There are three levels of processing known as layers 1, 2, 3
  - For example layer 1 does not include temporal masking, but is present in layer 2,3
  - MPEG Audio is used for compression of audio, and audio associated with video

- MPEG-1 Audio Layer 3, more commonly referred to as MP3, is a popular digital audio encoding
Video

- Video is also referred to as moving pictures and the term **frame** and picture are used interchangeably.

- Apply JPEG to each picture known as moving JPEG or MJPG.

In practice additional to the spatial redundancy in each frame considerable redundancy is often present between a set of frames since general only small portion of each frame is involved with any motion.

- A typical scene in a movie has about 3 seconds, frame refreshing rate of 60 frames per second, a scan would have 180 frames.

- Hence by sending only information relating to these segments of *each frame that have movement associated with them*, considerable additional saving can be achieved.
Instead of sending the source video as a set of individually compressed frames, just a selection is sent in this form and for the remaining frames only the difference between the actually frame contents and the predicted contents are sent.

The accuracy of the prediction is determined by how well any movement between successive frames is estimated.

- **Motion estimation**
  - since this process is not exact, additional information must also be sent to indicate small difference between predicted and actual positions of the moving segment involved
  - Motion compensation
Frame types

- There are two basic types
  - Those that are encoded independently and those that are predicted
  - Intracoded frames I-frames
    - I-frames are encoded without any reference to any other frame
    - Each frame is treated as a separate picture and encoded (JPEG)

P-frames

- Predicted frames predictive or P-frames
- The encoding of P-frame is relative to the contents of either a preceding I-frame or preceding P-frame
- Error of the first P-frame will propagate to the next
- The number of frames between P-frame and the immediately preceding I- or P-frame is called prediction span Represented by \( M \), value 1 to 3
The number of frame between successive I-frames is known as group of pictures GOP typical value from 3 to 12 symbol N

Motion estimation

- Target frame are compared on a pixel-by-pixel basis with the contents of the corresponding macro block in the preceding I- or P-frame (reference frame)

- Digitized contents of the Y-Matrix associated with each frame are first divided into a two-dimensional matrix of 16*16 pixels known as macro blocks
- If a close match is found, then only the address of the macro block is encoded
- If a match is not found, the search is extended to cover an area around the macro block
- Normally only the contents of the Y matrix are used in the search and a match is found if the mean of the absolute errors in the pixel position in the difference macro block is less than a given threshold.

- If a close match is found, two parameters are encoded known as the motion vector.
The second parameter is known as the prediction error and comprises three matrices each contains the difference values between those target macro blocks and the set of pixels in the search area that produce the close match.

Motion vectors can have large values, most moving objects a larger than a macro block. Several macro blocks are affected in a similar way. The motion vectors are encoded using differential encoding, the resulting codeword are then Huffman encoded.
- It is possible for a segment to have moved outside of the search region
- To allow this possibility, in addition to P-frames B-frames are used
H.261

- H.261 just I and P frames, three P frames between each pair of I-frame
- 8*8 pixel blocks

- H.261 is an 1990 ITU video coding standard originally designed for transmission over ISDN lines on which data rates are multiples of 64 kbit/s. The data rate of the coding algorithm was designed to be able to operate between 40 kbit/s and 2 Mbit/s.
H.263

- H.263 I, P, B-Frames
- H.263 was developed as an evolutionary improvement based on experience from H.261, the previous ITU-T standard for video compression, and the MPEG-1 and MPEG-2 standards.

MPEG (pronounced EM-peg) has standardized the following compression formats and ancillary standards:
- MPEG-1: Initial video and audio compression standard. Later used as the standard for Video CD, and includes the popular Layer 3 (MP3) audio compression format.
- MPEG-2: Transport, video and audio standards for broadcast-quality television. Used for over-the-air digital television and (with slight modifications) for DVDs
- MPEG-3: Originally designed for HDTV, but abandoned when it was discovered that MPEG-2 was sufficient for HDTV.
- MPEG-4: Expands MPEG-1 to support video/audio "objects," 3D content, low bitrate encoding and support for Digital Rights Management
  - MPEG-7: A formal system for describing multimedia content
  - MPEG-21: MPEG describes this future standard as a multimedia framework
- Reduction of information in the audio signal
  - Differential pulse code modulation (DPCM)
  - Adaptive differential
  - Adaptive predictive coding (APC)
  - Linear predictive coding (LPC)
  - Perceptual features
  - Human ear
  - MPEG audio coders

- Video
  - I-frame and P-frame
  - Motion estimation
  - MPEG