5.9. Video Compression (1)

- **Basics:**
  - video := time sequence of single images
  - frequent point of view:
    video compression = image compression with a temporal component
  - assumption:
    successive images of a video sequence are similar, e.g. directly adjacent images contain almost the same information that has to be carried only once
  - wanted:
    strategies to exploit temporal redundancy and irrelevance!
    → motion prediction/estimation, motion compensation, block matching
    → intraframe and interframe coding
  - video compression algorithms and standards are distinguished according to the peculiar conditions, e.g. videoconferencing, applications such as broadcast video

Video Compression (2)

- Simple approach:
  - **M-JPEG** compression of a time sequence of images based on the JPEG standard
  - unfortunately, not standardized!
  - makes use of the baseline system of JPEG, intraframe coding, color subsampling 4:1:1, 6 bit quantizer
  - temporal redundancy is not used!
  - applicable for compression ratios from 5:1 to 20:1, higher rates call for interframe coding
  - possibility to synchronize audio data is not provided
  - direct access to full images at every time position
  - application in proprietary consumer video cutting software and hardware solutions

Video Compression (3)

- **Motion prediction and compensation:**
  - kinds of motion:
    - change of color values / change of position of picture elements
    - translation, rotation, scaling, deformation of objects
    - change of lights and shadows
    - translation, rotation, zoom of camera
  - kinds of motion prediction techniques:
    - *prediction of pixels or ranges of pixels*
      neighbouring but no semantic relations
    - *model based prediction*
      grid model with model parameters describing the motion, e.g. head-shoulder-arrangements
    - *object or region based prediction*
      extraction of (video) objects, processing of geometric and texture information,
      e.g. MPEG-4

Technische Universität Kaiserslautern

ICSY Integrated Communication Systems

Video Compression (4)

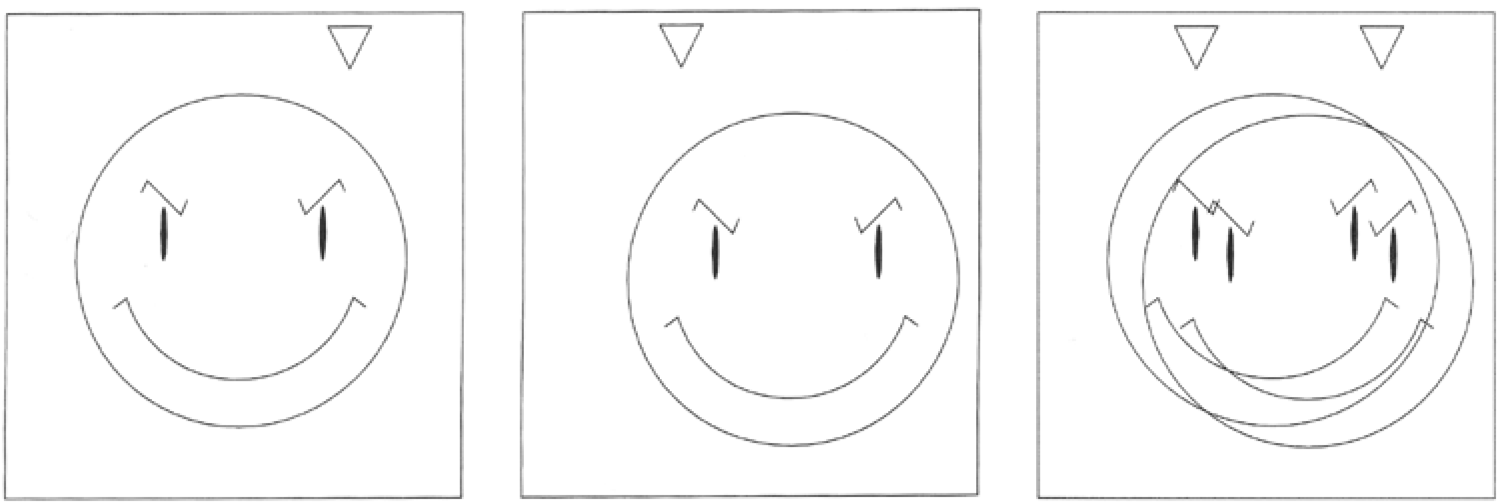
- **Motion prediction and compensation: (cont.)**
  - Block matching:
    - assumptions:
      - little change in the contents of the image from one frame to the next
      - even if, significant portions of the image do not change from one frame to the next

  → use a previous frame to generate prediction for the next
  - Solution:
    Predict the value of a pixel by the value of the pixel at the same location in a previous frame and encode the difference (differential encoding)?

  → No!
    - Objects tend to move between frames!
    - We would increase the amount of information that needs to be transmitted.

Video Compression (5)

- **Motion prediction and compensation: (cont.)**
  - Block matching: (cont.)

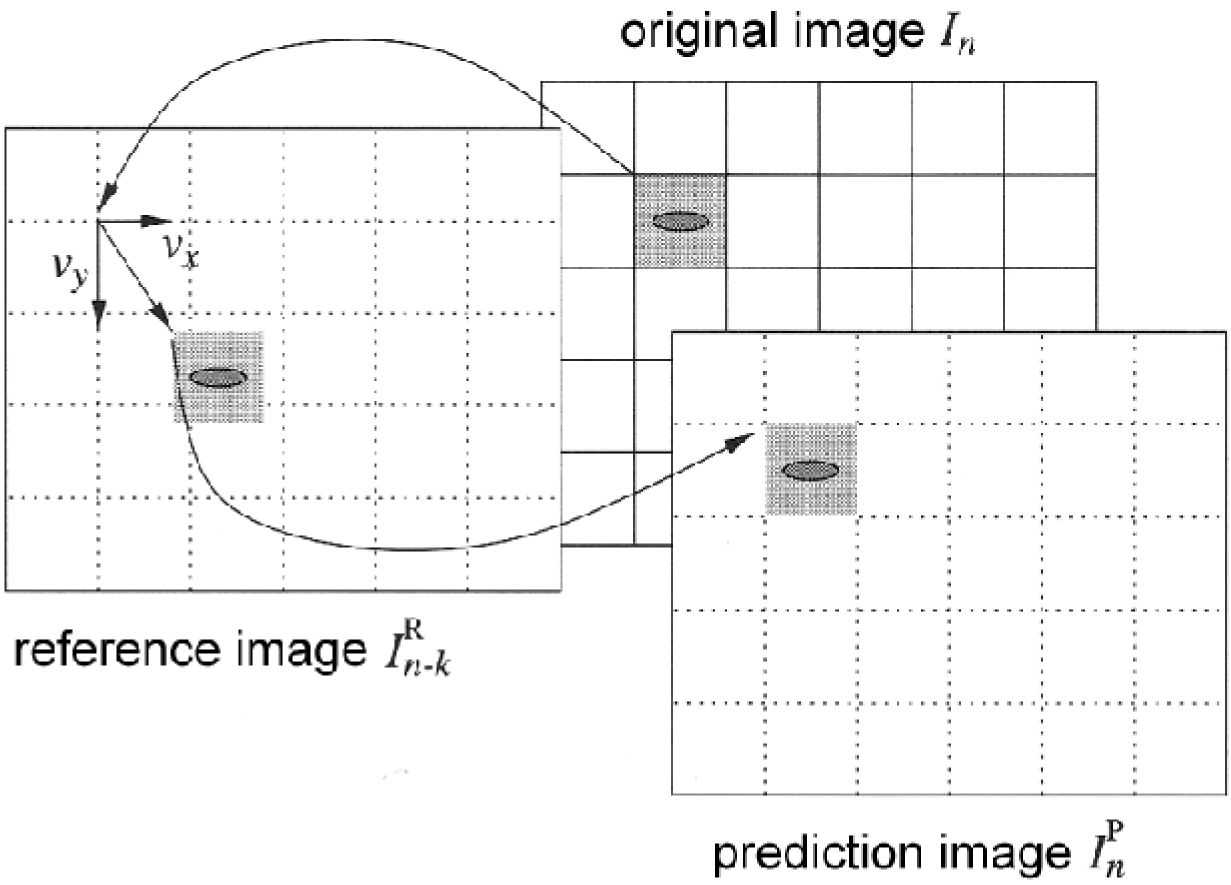


Video Compression (6)

- **Motion prediction and compensation: (cont.)**
- Block matching: (cont.)
  - Forward prediction:
    - divide the frame being encoded into blocks of equal size MxM
    - for each block: search a previous reconstructed frame for the block of size MxM that most closely matches the block being encoded
    - measure?
    - If the distance of the block being encoded to the closest block in a previous reconstructed frame (reference image) is greater than some predefined threshold, the block is declared uncompensable and is encoded without the benefit of prediction.
      If the distance is below the threshold, the difference of the block being encoded to its closest block is encoded and transmitted. Additionally a motion vector is transmitted.

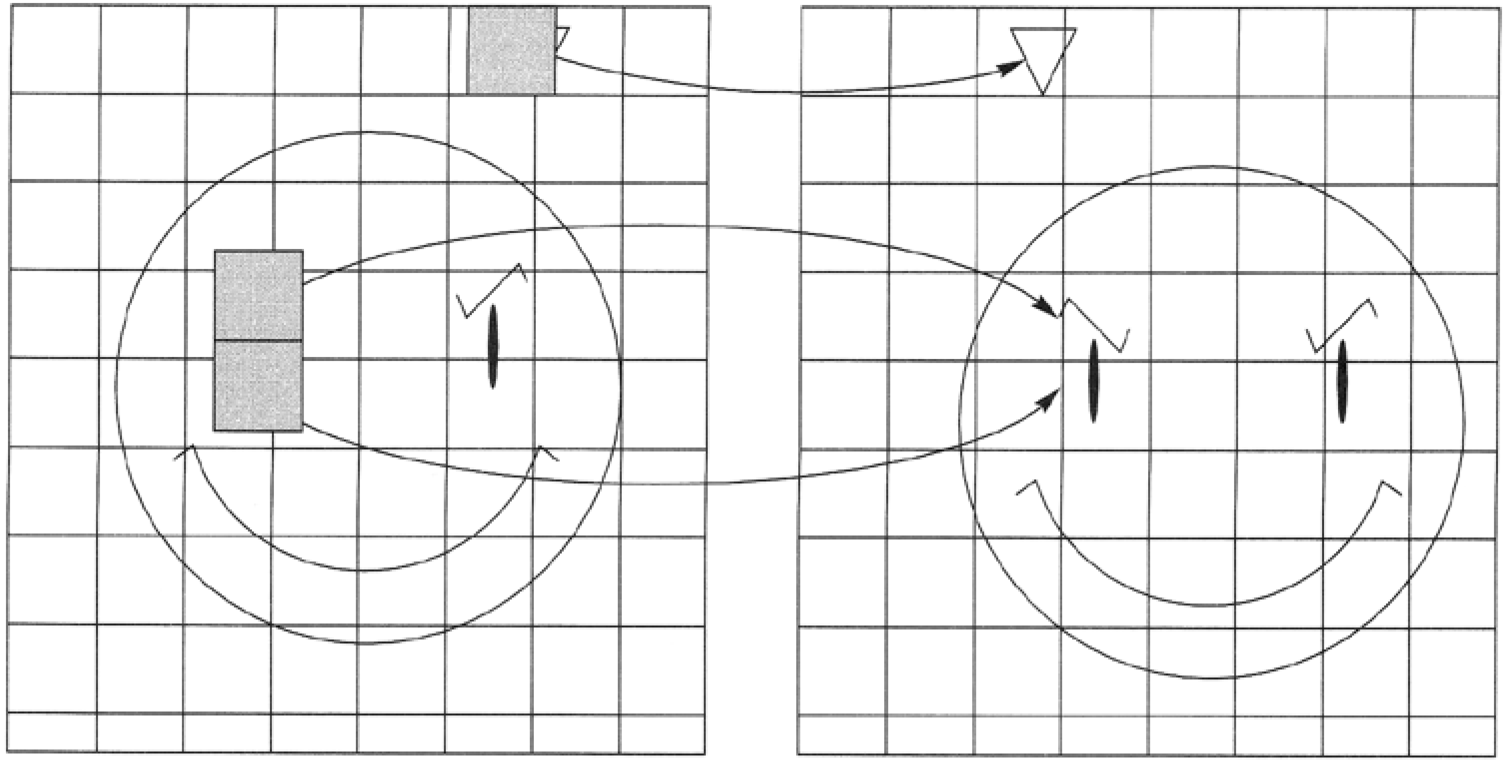
Video Compression (7)

- **Motion prediction and compensation: (cont.)**
  - Block matching: (cont.)



Video Compression (8)

- **Motion prediction and compensation: (cont.)**
  - Block matching: (cont.)

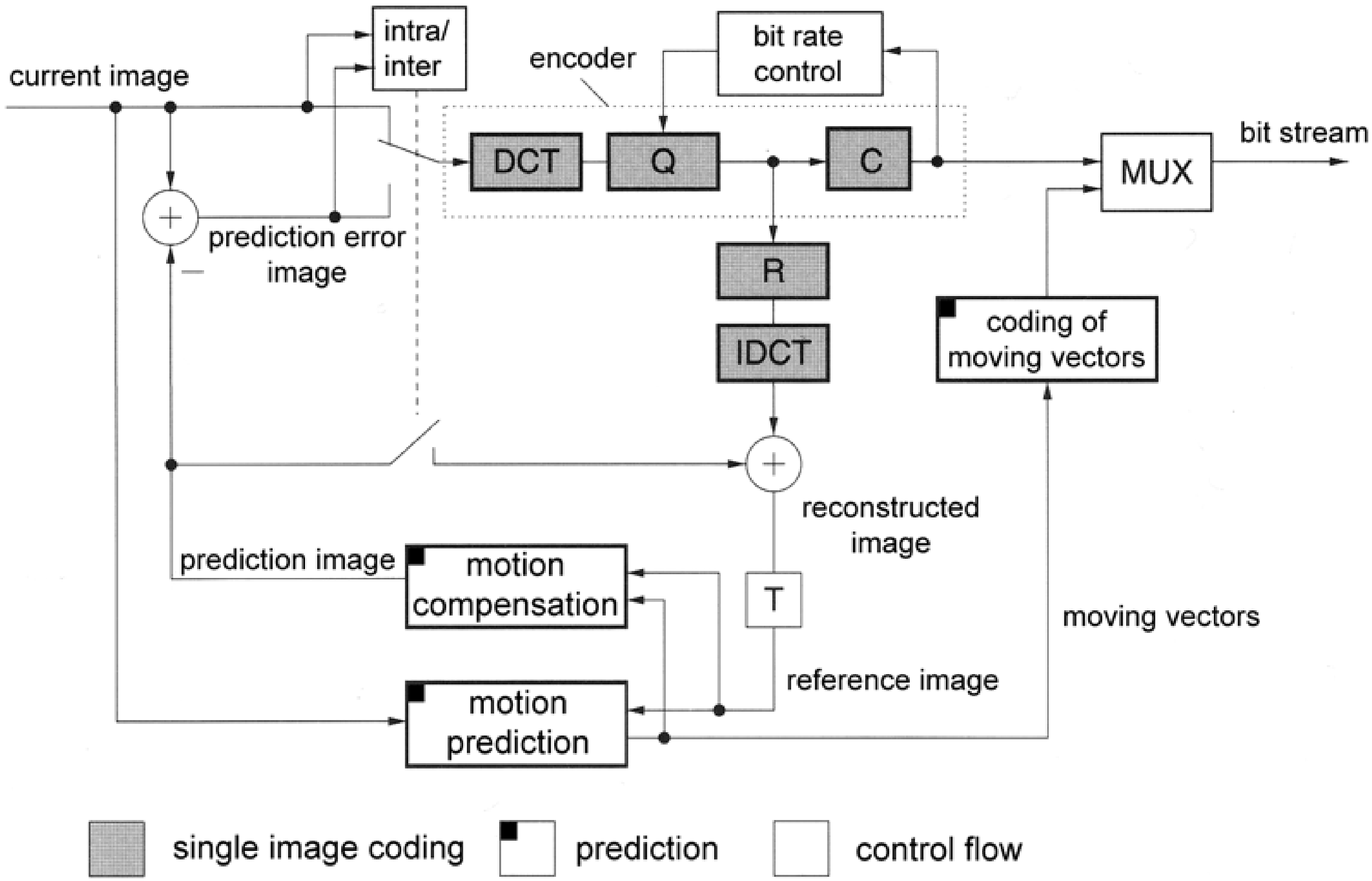


Video Compression (9)

- **Motion prediction and compensation: (cont.)**
  - Block matching: (cont.)
  - Search strategies:
    "find closely matching block" requires large amount of computation!
    - increase size of the blocks:
      - more computations per comparison but fewer blocks per frame
      - major drawback: with increasing block size the probability of covering objects moving in different positions increases
      - can result in poor prediction
    - reduce search space:
      - restrict region to search for matching block
      - increases the probability to miss a match

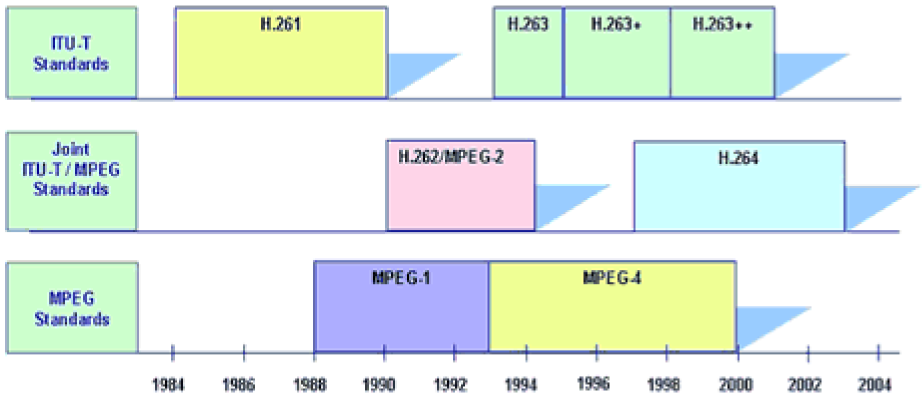
Video Compression (10)

Block diagram: general video compression / encoding procedure



Video Compression Overview

- **Two organizations**
  - ITU: H.26x series
  - ISO: MPEG-x series
  - both define sub-standards: e.g. "H.263.1" or "MPEG-4 part 2"
  - early standards (H.261 / MPEG-1) were developed for specific purposes
  - new standards (MPEG-4 / H.264) were developed for wider range of application



Video Compression - H.261 / H.263 (1)

- **Video compression for video conferences**
  - Compression in real time, symmetric procedure
  - Targeted to ISDN
  - Compressed data stream: p* 64 Kbit/s, p= 1, ..., 30
- **The data stream structure contains more information than just video:**
  - error correction information
  - image sequence number (H.261: 5 bit / H.263: 8 bit)
  - control commands from encoder to decoder
    - start / stop playing video
    - freeze a still image + timeout for automatic restart of video play
    - ...

Video Compression - H.261 / H.263 (2)

- **Image preparation:**
  - Video input must provide 29,97 images per second YUV representation with color sampling ratio of 4:2:0
- **Image sizes:**
  - CIF (Common Intermediate Format)
  - QCIF (Quarter-CIF)
  - 352 x 288
    176 x 144
- **H.263 only:**
  - SQCIF (Sub Quarter CIF)
  - 4CIF (4 time CIF)
  - 16CIF (16 times CIF)
  - 128 x 96
    704 x 576
    1408 x 1152
- Decoders must implement QCIF only, the rest is optional

Video Compression - H.261 / H.263 (3)

- **Structures:**
  - block = 8 x 8 pixel of Y U or V plane
  - macro block = 4 blocks of Y plane + 1 block of U and V plane
  - group of blocks (GOB) = 33 macro blocks
- **Intraframe coding:**
  - apply DCT to blocks
  - quantization with fixed Q-factors
    - for DC coefficient (= coefficient at position 0,0)
    - for AC coefficients (= all other coefficients)
  - between macro blocks within a GOB the Q-factor may be changed
- **Interframe coding (more options for H.263):**
  - there is a moving vector for each macro block
    - the moving vector may be null
    - otherwise the macro block contains the difference only
    - there is a special code if the difference is null

Video Compression - MPEG (1)

- **International standard of ISO 1993**
  - MPEG = Moving Pictures Expert Group
    - Today the first standard is often called MPEG-1
  - Definition of
    - Compression algorithms
    - Formats for storing or transmitting compressed media
  - Includes Video and Audio coding
- **Properties**
  - Asymmetric codec, the encoder is more complex than the decoder
  - The resulting bitrate is limited
    - 1.5 Mbit/s for video
    - 448 Kbit/s for audio

Video Compression - MPEG (2)

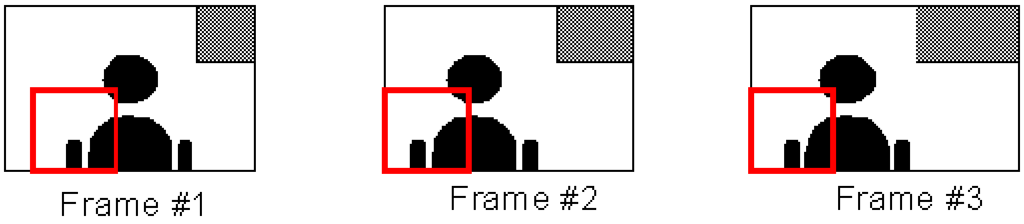
- **Image preparation**
  - Separation of luminance and chrominance (similar to YUV using 4:2:0)
- **More attributes required than for JPEG**
  - Relation of width and height of the pixels to support different TV formats; example:
    - 4:3 for PAL
    - 16:9 for HDTV (Europe and USA)
    - 1:1
  - Picture frequency
    - 25 Hz or 30 Hz for interlaced PAL or NTSC
    - 50 Hz and 60 Hz for non-interlaced Video
    - 23,976 Hz is the lowest possible frequency

Video Compression - MPEG (3)

1. Sequence Layer
   - Controls buffering of video data; each sequence starts with the constant bitrate and amount of memory required for the following sequence
2. Group of Pictures Layer
   - A set of pictures where at least one (usually the first) is encoded without references to other pictures (I-frame). The order of pictures within this layer may be different from the presentation order
3. Picture Layer
   - Contains one picture
4. Slice Layer
   - A set of macroblocks using the same DCT-Quantization scaling
5. Macroblock Layer
   - A block of 16x16 pixel build of four 8x8 blocks of the luminance plane and of two 8x8 blocks of the chrominance planes (one block of each chrominance plane)
6. Block Layer
   - One 8x8 block compressed using DCT

Video Compression - MPEG (4)

1. During image preparation the chrominance precision is reduced
2. Each 8x8 block is transformed using the DCT, quantizised, run-length encoded, and huffman encoded
3. Temporal correlations between video frames are used to reduce the size of 16x16 pixel macro blocks



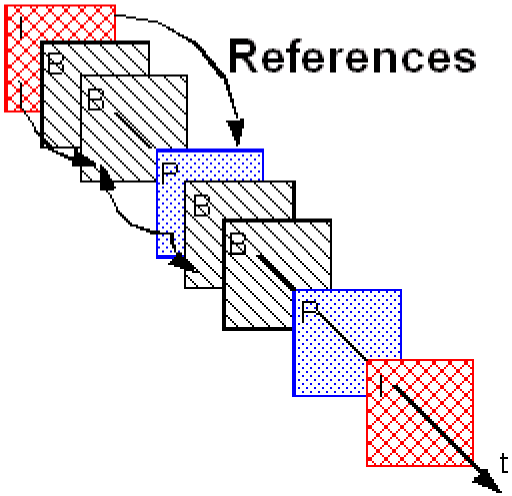
- Parts of frame #1 which change only slightly, may be encoded as references to frame #1 in frame #2 and frame #3
- Moving vectors are references to a 16x16 pixel area in prior or following frames describing temporal correlations

Video Compression - MPEG (5)

- **While encoding a macro block the encoder seeks for a similar 16x16 area in other frames. If such an area has been found the macro block is encoded by:**
  - The moving vector, describing the frame and the position of the similar area
  - Calculate the difference of the actual 16x16 pixel and the similar area (using luminance and chrominance planes)
  - Applying the DCT + quantization + entropy encoding to the difference only. A special value is used to describe a complete empty 8x8 block
- **MPEG defines how to describe a moving vector only**
  - MPEG does not define an algorithm for finding such a vector
    - "good" encoders find "good" matches in prior or following frames
    - "medium" encoders find some matches in other frames
    - "bad" encoders do not even try to find moving vectors
  - Usually only the luminance values are used to find similar areas
  - two adjacent macro blocks often have similar moving vectors

Video Compression - MPEG (6)

- **Several frame types with different temporal correlations:**
  - full frames are transmitted periodically (I-Frames)
  - frames that depend only on preceding frames (P-Frames) are used to simplify the calculation of B-Frames
  - temporal correlations may depend on frames of the "past" and frames of the "future" (B-Frames)
- **Sequence of I-, P-, and B-Frames:**



I-Frames (Intracoded)

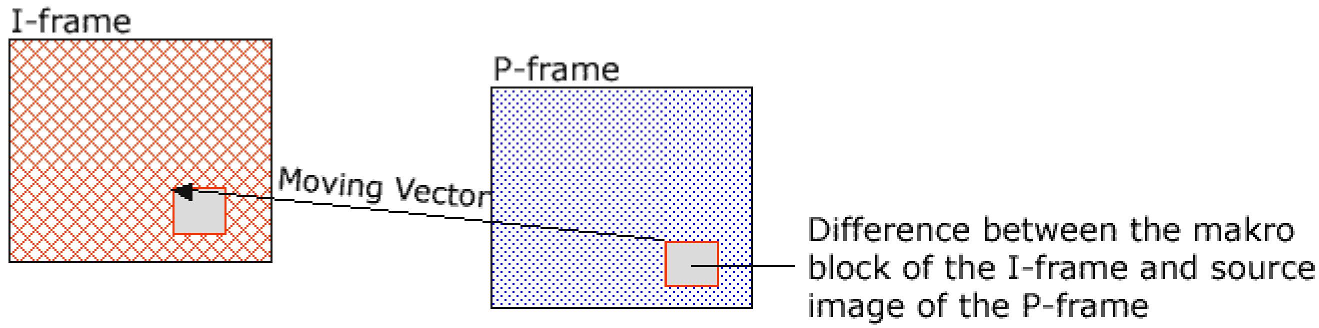
P-Frames (Predictive Coded)

B-Frames (Bidirectionally Coded)

D-Frames (DC Coded)

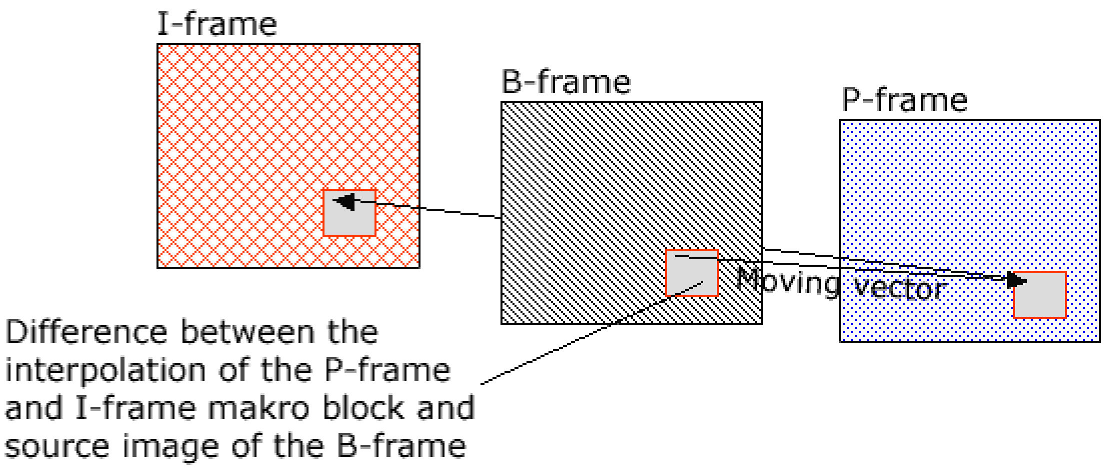
Video Compression - MPEG (7)

- **I-frames**
  - Coding without relation to other images
  - Apply DCT to 8 x 8 pixel blocks + spectral quantization + entropy coding ~ JPEG
- **P-frames**
  - Coding with relation to prior P or I frames
  - Macro blocks of P-frames may be encoded like macro blocks of I-frames or moving vectors may be used, so that only the difference to a prior frame must be encoded



Video Compression - MPEG (8)

- **B-frames**
  - Coding with relation to prior and following I and P frames
  - Macro blocks of B-frames may be encoded like macro blocks of P-frames or may be described as an interpolation of I or P frame macro blocks and its differences to the macro block to be coded

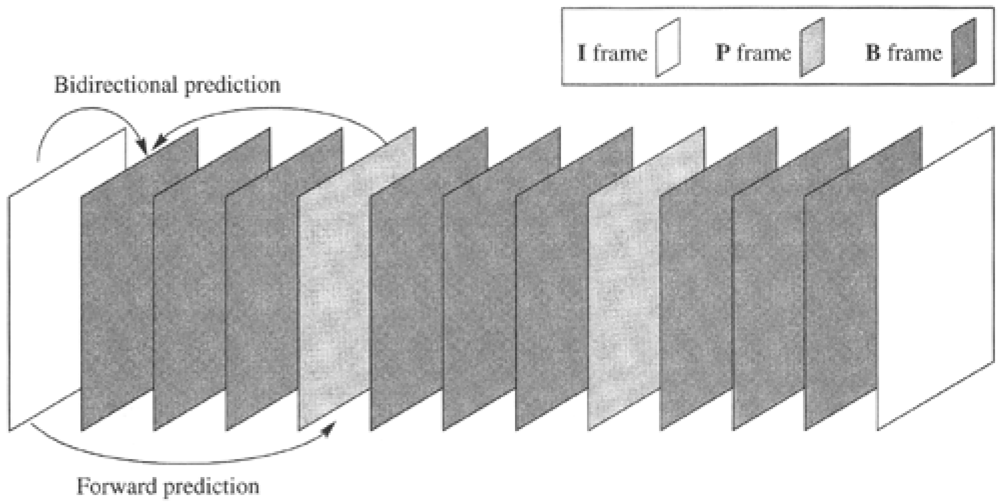


Video Compression - MPEG (9)

- **D-frames**
  - contain low frequency information only, i.e. the DC coefficient only
  - used to support fast forward
  - not necessary when periodic I-frames are used

Video Compression - MPEG (10)

- GOP-structure



  - display order

I	B	B	B	P	B	B	B	P	B	B	B	I
1	2	3	4	5	6	7	8	9	10	11	12	13

  - bitstream order

I	P	B	B	B	P	B	B	B	I	B	B	B
1	5	2	3	4	9	6	7	8	13	10	11	12

Video Compression - MPEG-2 (1)

- **MPEG-2 is an extension to MPEG-1 for high quality video**
- **The standard is flexible**
  - Profiles are defined for application classes
  - Different levels of qualities per profile
- **MPEG-2 defines scaling capabilities, i.e. a decoder can select the required scale**
  - Spatial; the horizontal and vertical resolution can be adapted by the receiver
    - Each picture is coded in different sizes, whereby the coding of size n contains only the differences to the image of size n-1
  - Rate; the number of frames per second may be defined by the receiver
    - Adequate placing of I-frames and (optionally) D-frames enables the playback with different frame rates

Video Compression - MPEG-2 (2)

	Simple Profile	Main Profile	SNR scalable Profile	Spatial scalable Profile	High Profile
attributes	No B-frames	B-frames			
	4:2:0				4:2:0 or 4:2:2
	Not scalable		SNR scalable	SNR or spatial scalable	
High Level 1920x1152		≤ 80 Mbit/s			≤ 100 Mbit/s
High-1440 Level 1440x1152		≤ 60 Mbit/s		≤ 60 Mbit/s	≤ 80 Mbit/s
Main Level 720x572	≤ 15 Mbit/s	≤ 15 Mbit/s	≤ 15 Mbit/s		≤ 20 Mbit/s
Low Level 352x288		≤ 4 Mbit/s	≤ 4 Mbit/s		

Video Compression - MPEG-4 (1)

- **Work on the standard started 1993 with these requirements:**
  - Motivation: instead of rebuilding the service of analog video, use the full flexibility of digitized data, i.e. describe a scene of objects instead of a 2D-video plane + audio
  - requires to efficiently represent a number of data types:
    - Video from very low bitrates to very high quality conditions
    - Music and speech data for a very wide bitrate range
    - Generic dynamic 3-D objects as well as specific objects such as human faces and bodies (MPEG-4 face animation example: http://www.cs.ualberta.ca/~anup/MPEG4/Demo.html)
    - Speech and music to be synthesized by the decoder, including support for 3-D audio spaces
    - Text and graphics

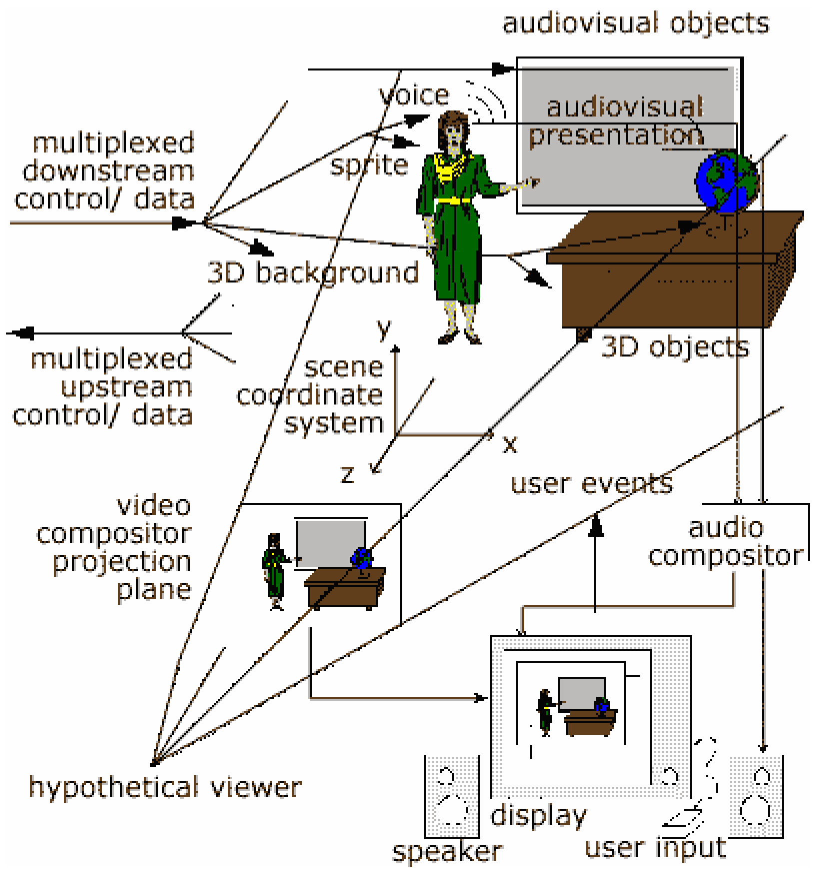
Video Compression - MPEG-4 (2)

- **Further requirements:**
  - Provide in the encoding layer resilience to residual errors for the various data types and for difficult channel conditions such as mobile ones
  - Independently represent the various objects in the scene, thus allowing independent access for their manipulation and re-use
  - Compose audio and visual as well as natural and synthetic objects into one audiovisual scene
  - Describe the objects and the events in the scene
  - Provide interaction and hyperlinking capabilities
  - Manage and protect intellectual property on audiovisual content and algorithms, so that only authorized users have access
  - Provide a delivery media that is independent from representation format to transparently cross the borders of different delivery environments

MPEG-4 Standards Overview (3)

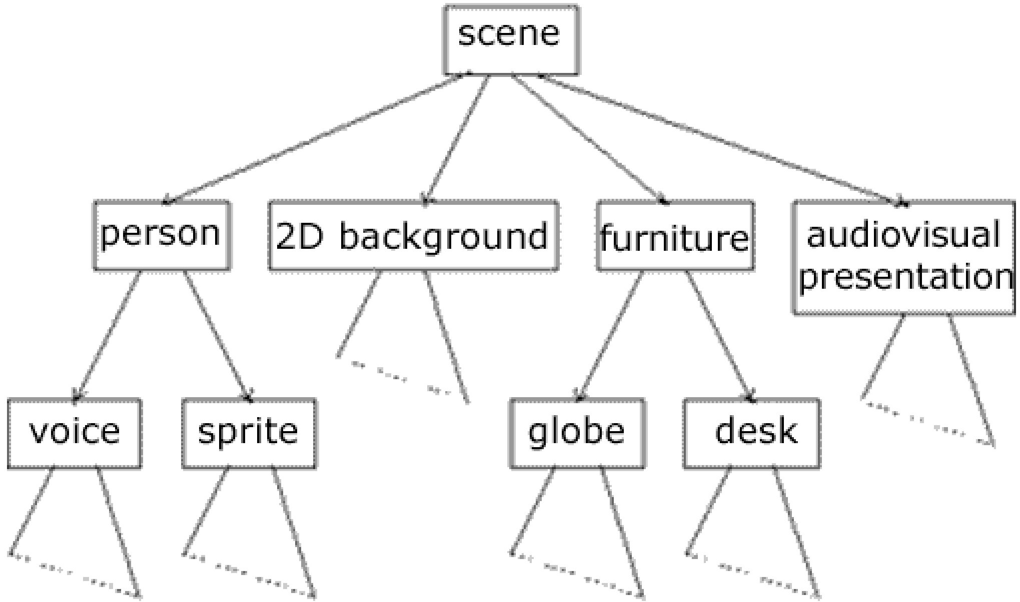
- ISO/IEC 14496: Information technology -- Coding of audio-visual objects
  - **ISO/IEC 14496-1:2004 Part 1 Systems (first Version published: 1998)**
  - **ISO/IEC 14496-2:2004 Part 2: Visual**
  - ISO/IEC 14496-3:2001 Part 3: Audio
  - ISO/IEC 14496-6:2000 Part 6: Delivery Multimedia Integration Framework (DMIF)
  - ISO/IEC 14496-8:2004 Part 8: Carriage of ISO/IEC 14496 contents over IP networks
  - **ISO/IEC 14496-10:2005 Part 10: Advanced Video Coding**
  - ISO/IEC 14496-11:2005 Part 11: Scene description and application engine
  - ISO/IEC 14496-13:2004 Part 13: Intellectual Property Management and Protection (IPMP) extensions
  - ISO/IEC 14496-14:2003 Part 14: MP4 file format
  - ISO/IEC 14496-15:2004 Part 15: Advanced Video Coding (AVC) file format
  - total of 78 standards subdivided to 22 Parts

Video Compression - MPEG-4 (4)



Video Compression - MPEG-4 (4)

- A scene is described by an object tree and coordinates for each AVO
- Each AVO is encoded by a special CoDec, e.g. wavelet for textures



MPEG-4 Visual

- A video scene is made up of video object planes (VOP)
  - VOPs may be used to encode foreground and background objects separately
  - A VOP is an arbitrary 2D-rectangles, which consists of
    - a texture
    - a bit mask defining which pixels are visible/unvisible or an alpha channel
  - still textures can be encoded using wavelets
  - changing textures are encoded similarly to MPEG-2 using I, P, and B-VOPs
  - encoding of bitmasks: arithmetic encoding (specialized variant) plus motion compensation
- Some additional features of MPEG-4 Part2: Visual
  - Motion compensation may use 16x16 blocks (as MPEG-2) or 8x8 blocks
  - Motion vectors use half pixel resolution to improve precision of matchings
  - Coefficient predicition: DC koeffizient (0,0) and coeffizients of the first row or first colum may be predicted by adjacent blocks (only upper or left block)
- DivX and Xvid are codecs that use MPEG-4 Visual
- MPEG-4 Natural Video Coding - An overview

Video Compression - MPEG-4 (5)

- **The standard is very flexible, it may be extended by**
  - new compression techniques
  - data types
  - extensions to the standard are described by the "MPEG-4 Syntactic Description Language" (MSDL)
- **The scene description of MPEG-4 is based on VRML**
- **Compression efficiency:**
  - Modelling of a scene adds complexity to the encoder and decoder
  - Distinction between foreground and background will lead to a better compression (compared to MPEG-1 or 2) in most cases
- MPEG-4 face animation example: http://www.cs.ualberta.ca/~anup/MPEG4/Demo.html

MPEG-4 AVC

- Advanced Video Codec (AVC) is identical to H.264
  - a design goal was to be very scalable
    - enable very low bandwidth video (e.g. for video conference)
    - high quality video (AVC is a mandatory codec for HDTV video on blu-ray discs)
- AVC is a very flexible format
  - The motion compensation may reference to multiple previous or following pictures (MPEG-2 only allows references to one picture only). Thus the display and reference order of pictures is decoupled.
  - Weighted prediction by motion compensation (higher efficiencies of fading pictures)
  - Macro block size may vary between 16x16 down to 4x4 pixel
  - Motion vectors use quarter pixel resolution to improve precision of matchings
  - Motion vectors may reference outside of picture boundaries
  - Deblocking filters (reduce artifacts, so far a common post processing step) are already taken into account during motion compensation and thus become mandatory
  - Flexible macro block ordering, enables arbitrary ordering of macro blocks, could be used to enhance robustness of the video stream (update important regions more often)
  - Context adaptive entropy encoding (run-length and arithmetic coding), improve efficiency
- The high degree of flexibility enables high compression rates but requires much more CPU resources than previous MPEG video codecs

Video Compression - MPEG-x

- **MPEG-7:**
  - activities that focus on the development of a multimedia content description interface
  - development of compact descriptions of information
  - efficient identification of multimedia content due to standardized description
  - (semi-)automatic search capabilities, filtering and selection of contents
  - automatic processing, intelligent applications
  - call for proposals in October 1998
  - MPEG-7 Overview
- **MPEG-21:**
  - Multimedia framework
  - includes: financial services, directory services, content providers, aggregators, consumer devices, regulatory services, technology services, delivery services,...

Video Compression - MPEG Overview

- MPEG-1:
  - the standard on which such products as Video CD and MP3 are based;
- MPEG-2:
  - the standard which such products as Digital Television set top boxes and DVD are based on;
- MPEG-4:
  - the standard for multimedia for the fixed and mobile web;
- MPEG-7:
  - the standard for description and search of audio and visual content.
  - MPEG-7 Overview
- MPEG-21:
  - Work on the new standard MPEG-21 "Multimedia Framework" has started in June 2000;
  - includes: financial services, directory services, content providers, aggregators, consumer devices, regulatory services, technology services, delivery services,...

5.10. Audio Compression (1)

- Conventional, lossless compression methods, such as Run Length Coding, can be used to compress sound file, but the results depend heavily on the specific sound
- Some lossy sound compression methods, like silence compression and companding (Compressing Expanding), taking advantage of our perception of sound; can get better quality
  - Silence compression: some small samples are treated as if they were silence (as samples of zero), since some people have less sensitive hearing
  - Companding: ear requires more precise samples at low amplitudes than high amplitudes. Companding uses a nonlinear formula to reduce the number of bits per sample

µ-Law and A-Law

- **µ-Law and A-Law: encoder**
  - International standard, logarithm-based, to encode digitized audio samples for ISDN digital telephony services
  - Experiments indicate that low amplitudes of speech signals contain more information than high amplitudes
  - Encoder receiving 14-bit input sample x, normalizing it to , it is within the interval [-1,+1], G.711 standard recommends µ=255

$$x' = \operatorname{sgn}(x)\frac{\ln(1+\mu|x|)}{\ln(1+\mu)}, \qquad \text{where } \operatorname{sgn}(x) = \begin{cases} +1, & \text{for } x > 0 \\ 0, & \text{for } x = 0 \\ -1, & \text{for } x < 0 \end{cases}$$

  - Since logarithms are complex, in practice the encoder performs much simpler calculations that produce an approximation

Audio Compression (2)

- µ-Law Codeword Format

P	S2	S1	S0	Q3	Q2	Q1	Q0

- Algorithm:
  1. Adding 33 to the absolute value of the input sample
  2. Determining the bit position of the most significant 1-bit among bits 5 through 12 of the input
  3. Encode S2-S0: subtracting 5 from that position -> Segment Code
  4. Encode Q3-Q0: the 4-bits quantization code is set of the four bits following the bit position determined in step 2

- Example, encode : –656
  1. |-656| + 33 =

				Q3	Q2	Q1	Q0					
0	0	0	1	0	1	0	1	1	0	0	0	1
12	11	10	9	8	7	6	5	4	3	2	1	0

  2. the most significant 1-bit is #9
  3. 9 - 5 = 4 in binary representation S2 S1 S0 = 1 0 0
  4. Q3 Q2 Q1 Q0 = bits at position 8-5 = 0 1 0 1

Audio Compression (3)

- µ-Law and A-Law: decoder example

P	S2	S1	S0	Q3	Q2	Q1	Q0
1	1	0	0	0	1	0	1

1. Multiply the quantization code by 2 and add 33:
   Binary 0101 equals decimal 5, so 5 x 2 + 33 = 43
2. Multiply the result by 2 raised to the power of the segment code:
   Binary 100- equals decimal 4, so 43 x 2^4 = 688
3. Decrement the result by 33: 688 – 33 = 655
4. Use bit P to determine the sign: -655

MPEG Audio (1)

- **MPEG-1: (1992, ISO/IEC 11172-3)**
  - signal-channel (mono) and two-channel (stereo) coding of digitized sound signal
  - used for high quality audio at 32, 44.1, and 48 kHz sampling rate
  - 3 coding methods are defined: Layer-1, Layer-2, Layer-3
    - Predefined bit rates range from 32 to 448 Kbit/s for layer-1, from 32 to 384 Kbit/s for Layer-2, and from 32 to 320 Kbit/s for Layer-3
- **MPEG-2 BC: (1994, ISO/IEC 13818-3)**
  - A backwards compatible multichannel extension to MPEG-1; up to 5 main channels plus a low frequent enhancement" channel (5.1 channels) can be coded
  - Low sample rate extension, sampling frequencies at 16, 22.05, and 24 kHz
  - Lower bit rate extension, bit rates down from 32 to 256 Kbit/s for Layer-1, and from 8 to 160 Kbit/s for Layer-2 and Layer-3

MPEG Audio (2)

- **MPEG-2 AAC (Advanced Audio Coding): (1997, ISO/IEC 13818-7)**
  - A high quality audio coding standard for 1 to 48 channels at sampling rates of 8 to 96 kHz
  - Not backward compatible, cannot be read and interpreted by an MPEG-1 audio decoder
- **MPEG-4: (1999, ISO/IEC 14496-3)**
  - Object-based coding

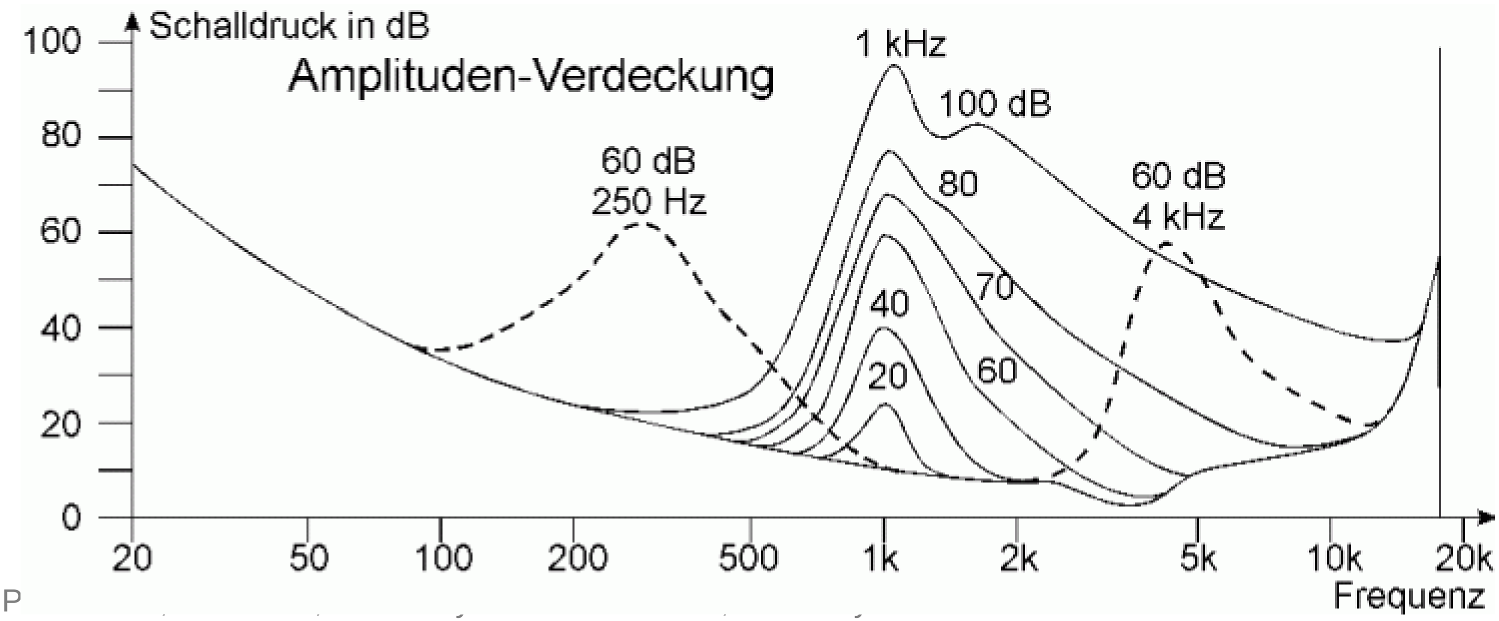
Layer

- Both in MPEG-1 and MPEG-2 BC three layers are defined. These Layers represent a family of coding algorithms. Basic model is the same, but codec complexity, performance and delay increase with each layer

	Compression ratio	Delay	Application
Layer-1	4:1	19 to 50 ms	Digital compact Cassette (DCC) 384 Kbit/s, Stereo
Layer-2	6:1 to 8:1	35 to 100 ms	DVD Digital audio Broadcast 192-256 Kbit/s
Layer-3 (MP3)	10:1 to 12:1	59 to 150 ms	Internet Audio 112-128 Kbit/s

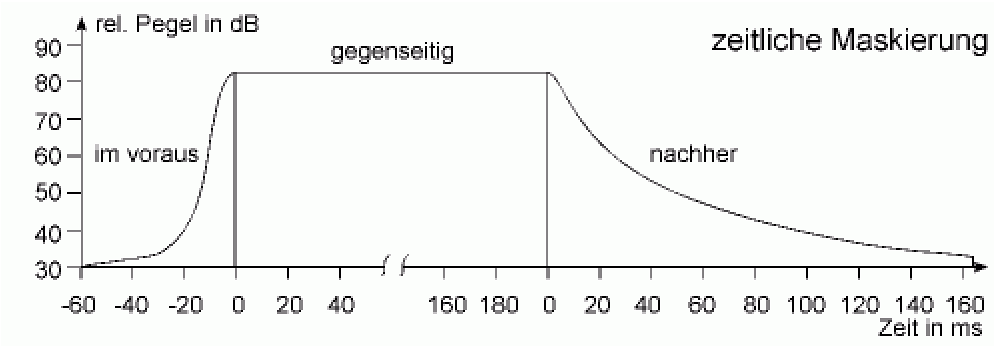
Psychoacoustic (1)

- MPEG's modern perceptual audio coding techniques exploit the psychoacoustic principles. The term psychoacoustic describes the characteristic of the human auditory system
- Sensitivity of human hearing
  - Range is about 20 Hz to 20 kHz, most sensitive at 2 to 4 kHz
- Spectral masking effect
  - Critical band: narrow at low audible frequencies, wide at high frequencies

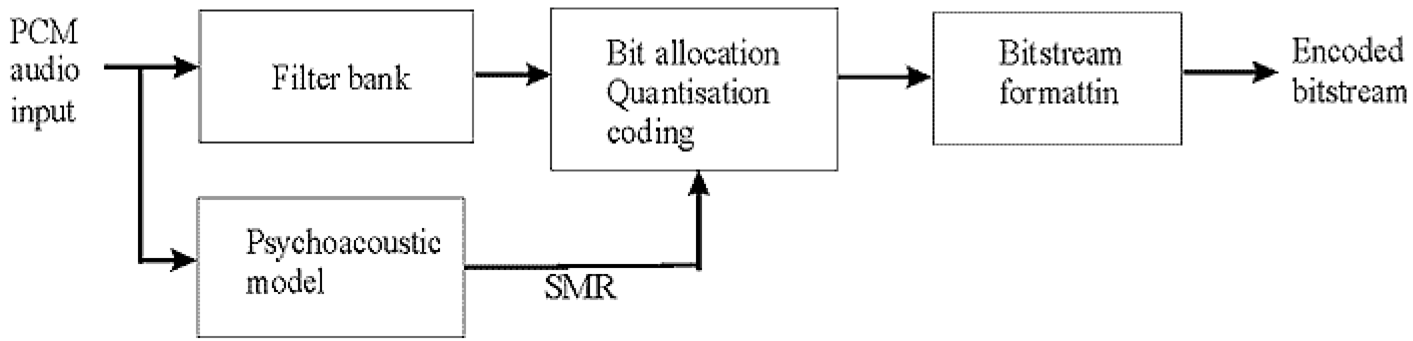


Psychoacoustic (2)

- Temporal masking effect
  - Masking effect before and after a strong sound
  - Pre masking is about 2 to 5 ms, the postmasking can be up to 100 ms



MPEG Audio encoder



- The input audio stream passes through a filter bank that divides the input into 32 subbands of frequency (the audio samples are transformed from time domain to frequency domain). The subband samples are packaged into frames containing 384 samples in Layer-1 (12 samples/subband * 32) and 1152 samples in Layer-2/3 (36 * 32)
- The input audio stream simultaneously passes through a psychoacoustic model that determines the masking threshold (SMR, Signal-to-Mask Ratio) for each subban

Frame Format of MPEG Audio Bitstreams Layer-1 Layer-2

Header (32)	CRC (0,16)	Bit Allocation (128-256)	Scalefactors (0-384)	Samples	Ancillary data

- The header of each frame contains general information such as the synchronization information, MPEG layer, the sampling frequency, the number of channels, the bit-rate, and coding mode
- Bit allocation describes the number of bits per sample in each subband (samples in different subbands could be represented by different numbers of bits), for Layer-1 this allocation can be 0 to 15 bits per subband
- The 12 signals of each subband are scaled such that the largest one becomes very close to one but not greater than one, each scale factor has 6 bits
- 1152 samples in Layer-2/3 (36 * 32) in each frame
- A frame is divided into 3 parts, each resembles a Layer-1 frame
- SCFSI (scale factor selection information) has 2 bits, it indicates whether 1, 2 or 3 scale factors per subband are written in the frame

Frame Format of MPEG Audio Bitstreams Layer-3

- For Layer-3 variable bit rate is allowed
  - Encoder can borrow bits donated from past frames
  - Side information includes a 9 bits point, "main_data_begin",

Header	CRC	Side information	Samples
32	(0,16)	(136,256)	

Header Side Info Frame 1	1	2	Herder Side Info Frame 2	2	3	4	Header Side Info Frame 3	4	Header Side Info Frame 4	4	5

Bit Allocation (1)

- The psychoacoustic model and the bit allocation algorithm are invoked to determine the number of bits allocated to the quantization of each scaled sample in each subband
- The bit allocation algorithm works in this way, that the quantization noise (difference between the original spectral values and the quantized spectral values) is below the masking threshold, while using not more than the available bits over a frame.
  - Since sample rate, samples per frame and bit rate are known, the available bits per frame is fixed
- For the most efficient compression, each subband should be allocated no more bits than necessary to make the quantization noise inaudible

Bit Allocation (2)

- **For Layer-1 and Layer-2:**
  - The process repeats until no more code bits can be allocated
- **For Layer-3:**
  - The process is done by two nested iteration loops
  - The quantized value are coded by Hoffman Coding
  - Inner iteration loop (rate loop)
  - If the number of bits exceeds the number of bits available ->
  - leading to larger quantization
  - Outer iteration loop
  - Achieving a smaller quantization noise requires a larger number of bits

MPEG-2 AAC (Advanced Audio Coding)

- MPEG-2 AAC is the state of the art audio coding scheme for generic coding of stereo multichannel signals; it gives up backwards compatibility to MPEG-1
- AAC follows the same basic coding paradigm as Layer-3 (high frequency resolution filter bank, non-uniform quantization, Huffman coding, iteration loop structure), but improves on Layer-3 in a lot of details and uses new coding tools for improved quality at low bit-rates
- AAC on the contrary standards only the format of the encoded audio data.
- The only audio coding scheme used within the MPEG-4 standard
- Application
  - Digital broadcasting system
  - Audio via Internet

5.11. Multimedia File Formats

images	raster image	e.g. BMP, TIFF, JPEG, PNG
	vector image	e.g. SVG
audio	wave audio	e.g. WAV, MP3
	audio synthesis	e.g. MIDI
video		in practice not relevant
combined formats/ container formats		e.g. AVI, OGG

File Formats: BMP

- **Name:**
  (Microsoft Windows) BitMaP
- **File extension:**
  .bmp
- **Development:**
  Microsoft Corp.
- **Color representation:**
  - Monochrome, grayscale, RGB;
  - indexed colors
- **ICC profiles:**
  - only Windows color management
- **Channels:**
  - 1 or 3 color channels
  - no alpha channels
- **Color depth:**
  - 1-8 bit per channel
  - resulting in 1, 4, 8, 24 bit
- **Compression:**
  - none (mostly)
  - RLE (only at 4 or 8 bit color depth supported)
- **Advantages:**
  - on Windows platform: standard format
- **Disadvantages:**
  - large files
  - rare deployment on non Windows platforms
  - not well suited for the Internet

File Formats: TIFF

- **Name:**
  Tag(ged) Image File Format
- **File extension:**
  .tif, .tiff
- **Development:**
  Aldus Corporation (now part of Adobe Systems, Inc.), Microsoft, HP, as well as other printer and scanner manufacturers
- **Color representation:**
  - monochrome; grayscale; RGB; CMYK
  - indexed colors
  - and others
- **ICC profiles:**
  yes
- **Channels:**
  - 1, 3, or 4 color channels
  - 20 alpha channels
- **Color depth:**
  - 1-16 bit per channel
- **Compression:**
  - None
  - RLE (PackBits)
  - LZW, CCITT Group 3 and 4
  - JPEG
- **Advantages:**
  - multiple pictures per file are supported
  - extendable
  - platform independent
- **Disadvantages:**
  - files are comparatively large, thus not well suited for Internet use
  - different compression methods cause interoperability issues

File Formats: JPG

- **Name:**
  Joint Photographic Expert Group (JPEG) File Interchange Format (JFIF)
- **File extension:**
  .jpg, .jpeg
- **Development:**
  JPEG, ISO
- **Color representation:**
  - grayscale; RGB; CMYK
  - and others
- **ICC profiles:**
  yes
- **Channels:**
  - 3 or 4 color channels
  - no alpha channels
- **Color depth:**
  - 1-16 bit per channel
- **Compression:**
  - lossless (prediction, Huffman; used seldom: not very effective)
  - DCT
- **Advantages:**
  - well suited for colorful images
  - widely supported
  - good compression results
- **Disadvantages:**
  - lossy compression not suited for graphics (unsatisfying results with outlines, missing transparency)
  - no advanced features (see JPEG2000)

File Formats: JPEG2000

- **Name:**
  Joint Photographic Expert Group 2000
- **File extension:**
  .jp2
- **Development:**
  - JPEG, ISO
- **Color representation:**
  - grayscale; RGB; CMYK
  - and others
- **ICC profiles:** yes
- **Channels:**
  - 256
- **Color depth:**
  - 1-16 bit per channel
- **Compression:**
  - Wavelet
- **Advantages:**
  - better compression results at same quality compared to JPEG (no artifacts)
  - errors in file affect images only locally
  - ROI (Region of Interest)
  - many options for progressive image display
  - lossy and lossless compression possible
- **Disadvantages:**
  - more processing needed compared to JPEG
  - still not widely supported

File Formats: GIF

- **Name:**
  Graphics Interchange Format
- **File extension:**
  .gif
- **Development:**
  CompuServe Inc.
- **Color representation:**
  - indexed colors
- **ICC profiles:**
  no
- **Channels:**
  - 1
  - transparent color possible
- **Color depth:**
  - max. 8 bit
- **Compression:**
  - LZW
- **Advantages:**
  - well suited for Internet graphics
  - widely supported
  - interlaced: progressive image display possible
  - animated graphics possible
- **Disadvantages:**
  - license issues until mid 2004
  - not suited for colorful images (only 256 colors, limited to lossless compression)
  - two incompatible versions (GIF87a und GIF89a)

File Formats: PNG

- **Name:**
  Portable Network Graphics
- **File extension:**
  .png
- **Development:**
  Massachusetts Institute of Technology (MIT)
- **Color representation:**
  - grayscale; RGB
  - indexed colors
- **ICC profiles:**
  limited
- **Channels:**
  - 3 channels
  - 1 alpha channel
- **Color depth:**
  - 1-16 bit per channel
- **Compression:**
  - deflate (similar to LZ77)
- **Advantages:**
  - Supported by webbrowsers since IE4 (no transparency!), Netscape 4.04
  - Superset of GIF functionality
  - Improved interlacing
- **Disadvantages:**
  - Still lacking support by webbrowsers

File Formats: EPS

- **Name:**
  Encapsulated PostScript
- **File extension:**
  .eps
- **Development:**
  Adobe Systems, Inc.
- **Color representation:**
  - monochrome; grayscale; RGB; CMYK
  - and others
- **ICC profiles:**
  yes
- **Channels:**
  - 1, 3, or 4 color channels
  - no alpha channels
- **Color depth:**
  - 1-8 bit per channel
- **Format of Preview Images:**
  - PICT (Macs), TIFF
  - Metafile (WMF; Windows), EPSI
- **Compression:**
  - none (usually)
  - JPEG
- **Advantages:**
  - well suited for printing
  - supports vector graphics and raster graphics
- **Disadvantages:**
  - files are comparatively large, thus not suited for Internet use

Audio/Video Formats

Suffix	Format	Website
.mp2	MPEG-1 Layer 1,2	
.mp3	MPEG-1 Layer 3	http://www.mpeg.org/MPEG/MPEG-audio-player.html
.mp4	MPEG-4	http://sound.media.mit.edu/mpeg4
.wav	Microsoft Wave Format	http://www.jmcgowan.com/avitech.html
.avi	Microsoft Audio Video Interleaved	http://www.jmcgowan.com/avitech.html
.ogg	ogg (not an acronym)	http://www.xiph.org
.mov, .MooV	Apple Quicktime	
.au	Sun Audio Format (Java 1.1)	

Microsoft AVI

- AVI Files are a special case of RIFF files.
  RIFF is Resource Interchange File Format and serves as general purpose format for exchanging multimedia data types. It was defined by Microsoft and IBM based on IFF from Electronic Arts.
- RIFF starts with a header followed by chunks and lists.
- Some of the codecs used for video compression in an AVI file:

Cinepak	high quality, good colors, high computational cost
Indeo	Intel, good for pictures without much movement, high computational cost, bad for large uniform areas
RLE	Random Length Encoding, lossless, relatively fast, only 8 bit colors
Video 1	Microsoft, good quality, fast computation, high bitrate, only 16 bit colors
MPEG-1	TV quality, high computational cost

- Cinepak codec creates about 60 MB/min for 384x288 pixels and 25 fps.
  Indeo 5.10 codec creates about 40 MB/min for 384x288 pixels and 25 fps.

Microsoft WMV

- Window Media Video
- Proprietary video codec from Microsoft
- Often encapsulated in ASF (Advanced Streaming Format) files
- Similar to MPEG-4 Video
- High definition format available (WMVHD)
- Supports DRM
- Three versions: WMV1, WMV2, WMV3 (released with Media Player 7-9)
- WMV3 (also called VC-1) available as video codec on Blue-Ray discs

Apple Quicktime

- QuickTime is a method for time-based compression of video and synchronized audio.
- It can also be used for time-based measurement data.
- QuickTime comes with a lot of convenient functions to ease the programmer's task and create a standard look-and-feel for the user.
- It is possible to add new compression algorithms and driver for video codecs.
- Unfortunately, the technical details are hidden. Especially the video compression algorithms are all from Apple and not documented.
- Apple claims that MPEG-4 is based on the QuickTime video format. QuickTime 6 works with MPEG-4.

Storing Multimedia Data

- For storing multimedia data at home, there are currently three popular formats: VCD, SVCD, and DVD.
- The video compact disc (VCD) uses MPEG-1 with 352x288 pixels and can store up to 74 min of video and audio on a single CD (650 MB). According to the standard, there is 1.15 Mbit/s video and 224 Kbit/s audio with 44.1 kHz audio sampling rate.
- The super video compact disc (SVCD) uses MPEG-2 with 480x576 pixels and allows interlaced frames. There is up to 2.6 Mbit/s data including 64-384 Kbit/s audio. 35 to 60 min fit on a single disc (650 MB).
- The digital video disc (DVD) uses MPEG-2 with 720x576 pixels and allows interlaced frames. There is up to 9.8 Mbit/s data including 64-6144 Kbit/s audio. 60 to 180 min fit on a DVD (4.7GB).