Music Compression

- MP3 (MPEG-1 Layer 3)
- AAC (MPEG-2 Advanced Audio Coding)
- Ogg Vorbis (patent free, see *xiph.org*)

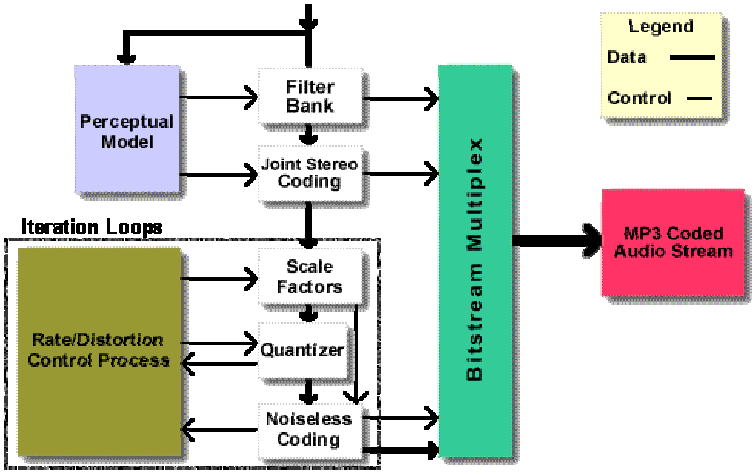
MPEG audio encoding

Layer	Application	Compressed bit rate [kb/s]	Example quality, per channel [kbps]	Example end-to-end delay [ms]
1	Digital audio cassette	32-448	hi-fi: 192	20
2	Digital audio & video broadcasting	32-192	near CD: 128	40
3	low-bit rate audio	64	CD: 64	60

MPEG Moving Pictures Expert Group

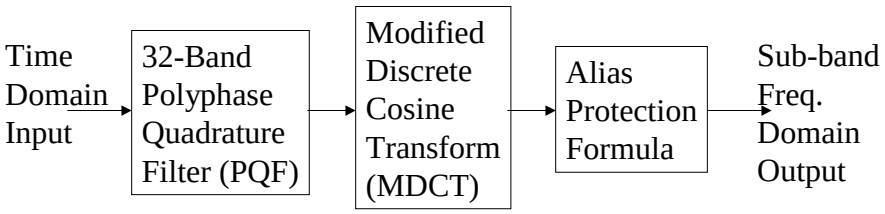
source: "Multimedia Communications" by Fred Halsall, page 189

MP3



Source: Fraunhofer IIS

MP3 Filter bank

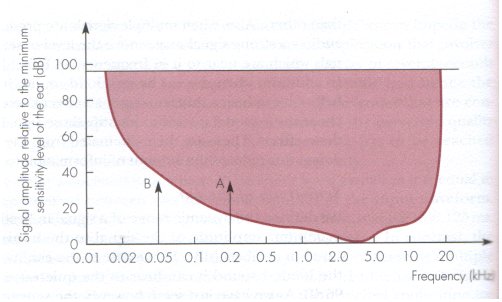


Perceptual coding

- Exploit the properties and limitations of the human ear
- Psycho-acoustic model:
  - sensitivity of ear
  - frequency masking
  - temporal masking
- High or low quality encodings possible
  - trade-off: quality vs. size

Amplitude sensitivity

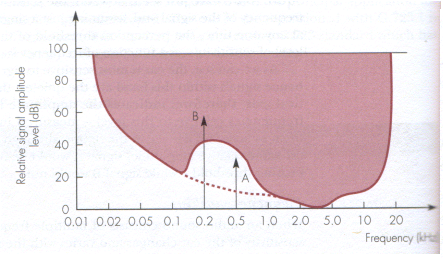
- Dynamic range is a function of frequency
- $D = 20\log(A_{max}/A_{min})$ dB
- Some frequencies are only heard if they are of a certain amplitude:
  - A is heard, B is not heard

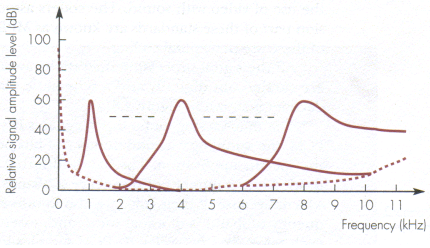


source: "Multimedia Communications" by Fred Halsall, page 184

Frequency masking

- Loud sounds can hide softer sounds:
  - dynamic range "distorted" around a loud sound
- Masking effect is frequency dependent
  - critical bandwidth
  - e.g. 1KHz, 4KHz, and 8KHz

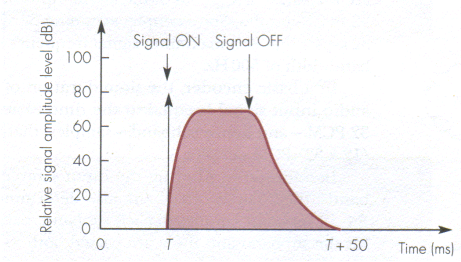




source: "Multimedia Communications" by Fred Halsall, pages 184, 185

Temporal masking

- A loud sound leaves an "audio image" that takes some time to fade
- This hides other sounds that occur soon after it



source: "Multimedia Communications" by Fred Halsall, page 185

Quantisation and Coding

- Quantisation: power-law (large values coded with less accuracy)
- Coding: Huffman coding

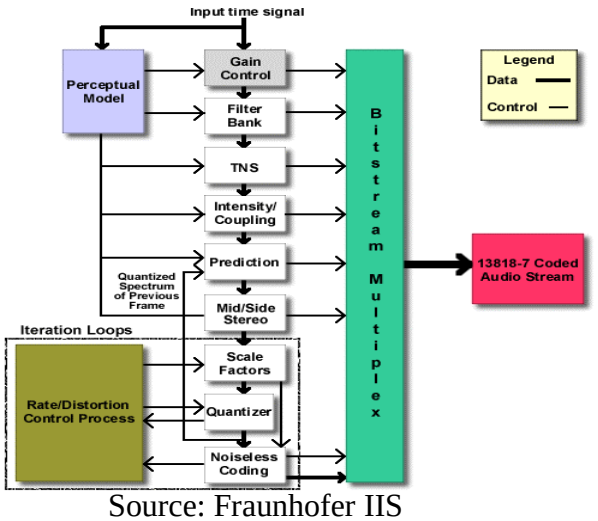
Scale factors

- **Inner iteration loop (rate loop)**
  - If the number of bits resulting from the Huffman coding operation exceeds the number of bits available to code a given block of data, this can be corrected by adjusting the global gain to result in a larger quantization step size, leading to smaller quantized values.
  - This operation is repeated with different quantization step sizes until the resulting bit demand for Huffman coding is small enough.
- **Outer iteration loop (noise control/distortion loop)**
  - To shape the quantization noise according to the masking threshold, scalefactors are applied to each scalefactor band. (The systems starts with a default factor of 1.0 for each band).
  - If the quantization noise in a given band is found to exceed the masking threshold (allowed noise) as supplied by the perceptual model, the scalefactor for this band is adjusted to reduce the quantization noise.
  - Since achieving a smaller quantization noise requires a larger number of quantization steps and thus a higher bitrate, the rate adjustment loop has to be repeated every time new scalefactors are used. (In other words, the rate loop is nested within the noise control loop).
  - The outer (noise control) loop is executed until the actual noise (computed from the difference of the original spectral values minus the quantized spectral values) is below the masking threshold for every scalefactor band (i.e. critical band).

MPEG-2 Advanced Audio Coding (AAC)

- Efficiently encodes 5.1 surround sound (up to 48 channels!) and sample-rates up to 96kHz
- Uses MDCT (no PQF/alias protection)
- Temporal Noise Shaping (TNS): shapes the distribution of quantization noise in time by prediction in the frequency domain. (Good for voice signals)
- Prediction: A technique commonly established in the area of speech coding systems

AAC



Source: Fraunhofer IIS