Need for audio coding

- Without data reduction, digital audio signals typically consist of 16 bit samples recorded at a sampling rate more than twice the actual audio bandwidth
- Thus, we end up with more than 1.4 Mbit to represent just one second of stereo music in CD quality
  - 44,1 ksampel/s * 16bits/sampel * 2 channels = 1,4112 Mbit/s
- Using MPEG audio coding, you may shrink down the original sound data from a CD by a factor of 12, without losing sound quality. Factors of 24 and even more still maintain a sound quality that is significantly better than the average listener can hear
  - Commonly 128 kbit/s
- This is realized by *perceptual coding* techniques addressing the perception of sound waves by the human ear.

Compression Approaches

- Delta coding
  - Encode differences only
- Predictive coding
  - Predict the next sample
- Linear Predictive Coding (LPC) - mostly for speech
  - Describe fundamental frequencies + 'error'
  - CELP, RPE, cell-phone standards
- Variable Rate Encoding
  - Don't encode silences
  - regular signal=few bits, variable signal=many bits
- Subband coding
  - Split into frequency bands each encoded separately + efficiently
- Psycho-acoustical coding
  - drop bits where you can't hear it

Compression methods/standards

PCM (Pulse Code Modulation)
u-LAW (Mu-law – logarithmic coding)

LPC-10E (Linear Predictive Coding 2.4kb/s)
CELP 4.8Kb/s – code excited LPC builds on LPC
GSM (European Cell Phones, RPE-LPC)
    1625 bytes/sec (at 8000 samples/sec)
    20 ms blocks of 260 bits each
ADPCM (adaptive, delta PCM, 24/32/40 kbps)
MPEG Audio Layers (builds on ADPCM)
    Layer-2: From 32 kbps to 384 kbps - target bit rate of 128 kbps
    Layer-3: From 32 kbps to 320 kbps - target bit rate of 64 kbps
Complex compression, using perceptual models
RealAudio, Windows Media Formats (builds on above, proprietary)

Audio and MP3

- In 1987, the Fraunhofer IIS-A started to work on perceptual audio coding.
- In a joint cooperation with the University of Erlangen, the Fraunhofer IIS-A devised a very powerful algorithm that is standardized as ISO-MPEG Audio Layer-3
- MPEG Audio Layer-3 is a enhancement of MPEG and is called MP3
- The MPEG compression system includes a subsystem to compress sound called MP3

Typical Data Reduction in MPEG audio

- 1:4 by Layer 1 (corresponds with 384 kbps for a stereo signal)
- 1:6 - 1:8 by Layer 2 (corresponds with 256..192 kbps for a stereo signal)
- 1:10 - 1:12 by Layer 3 (corresponds with 128..112 kbps for a stereo signal)

MPEG Audio Layer-3

- MPEG Layer-3 is the most powerful member of the MPEG audio coding family. For a given sound quality level, it requires the lowest bit rate - or for a given bit rate, it achieves the highest sound quality.

- In all international listening tests, MPEG Layer-3 impressively proved its superior performance, maintaining the original sound quality at a data reduction of 1:12

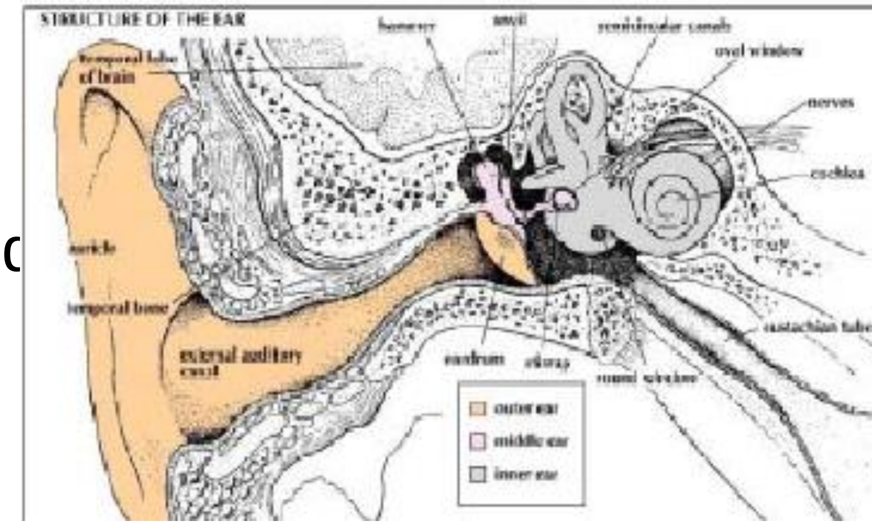
Human Auditory System

1. **Outer Ear:**

• **Pinna:** Collects sound

•**Ear Canal**:Amplifies the sound

•**Ear Drum**:Converts sound to mechanical vibrations.

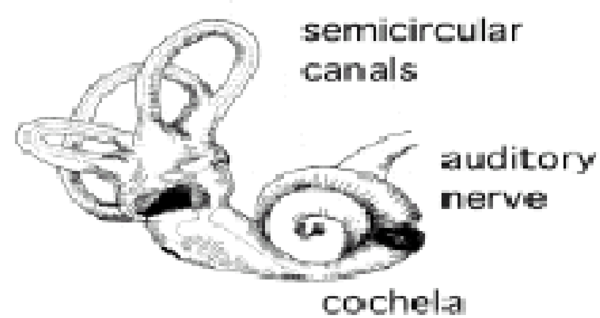


2. **Middle Ear**

• **Hammer, Anvil and stirrup**:

(i) Match outer ear to inner ear.

(ii) Low Pass Filter the sound.



3. Innear ear

- **Oval window**: amplifies sound 15-20 times.
- **Basilar Membrane**: spectrum analyzer

1. If peak frequency are close together it can't distinguish them " **Simultaneous Masking**".
2. Strong peaks cause the membrane not to return to equilibrium until several millisecond. " **Temporal Masking"**

- **Corti organ**: Contains IHCs
- **IHCs**: deliver the vibrations to the brain.

– Vibrate at strongest frequency in local domain called bark. " **Simultaneous Masking**".

– Recover between fringes depending on the strength of the delivered pulse " **Temporal Masking**".

Sound compression

A simple lossless compression method is to record the length of a period of silence

– no need to record 44,100 samples of value zero for each second of silence no need to record 44,100 samples of value zero for each second of silence

– form of run-length encoding

– in reality this is not lossless, as "silence" rarely corresponds to sample values of exactly zero; rather some threshold value is applied

• Difference between how we perceive sounds and images results in different lossy compression techniques for the two media

– high spatial frequencies can be discarded in images

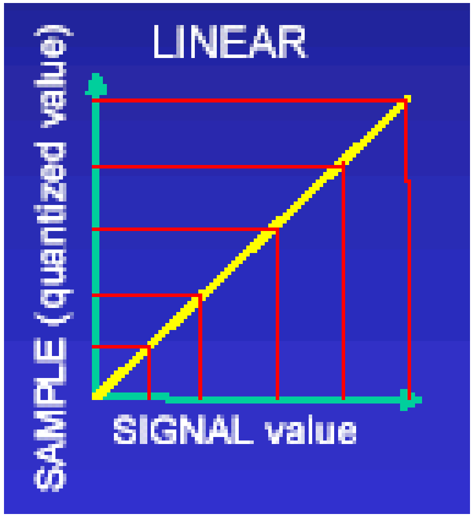
– high sound frequencies, however, are highly significant

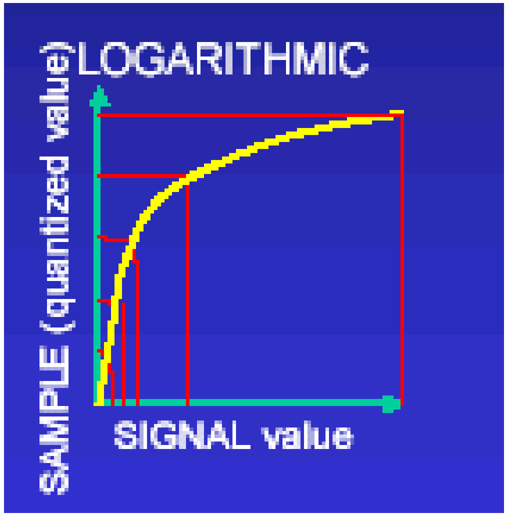
• So what can we discard from sound?

Sound compression

•Our perception of loudness is essentially logarithmic in the amplitude of a sound

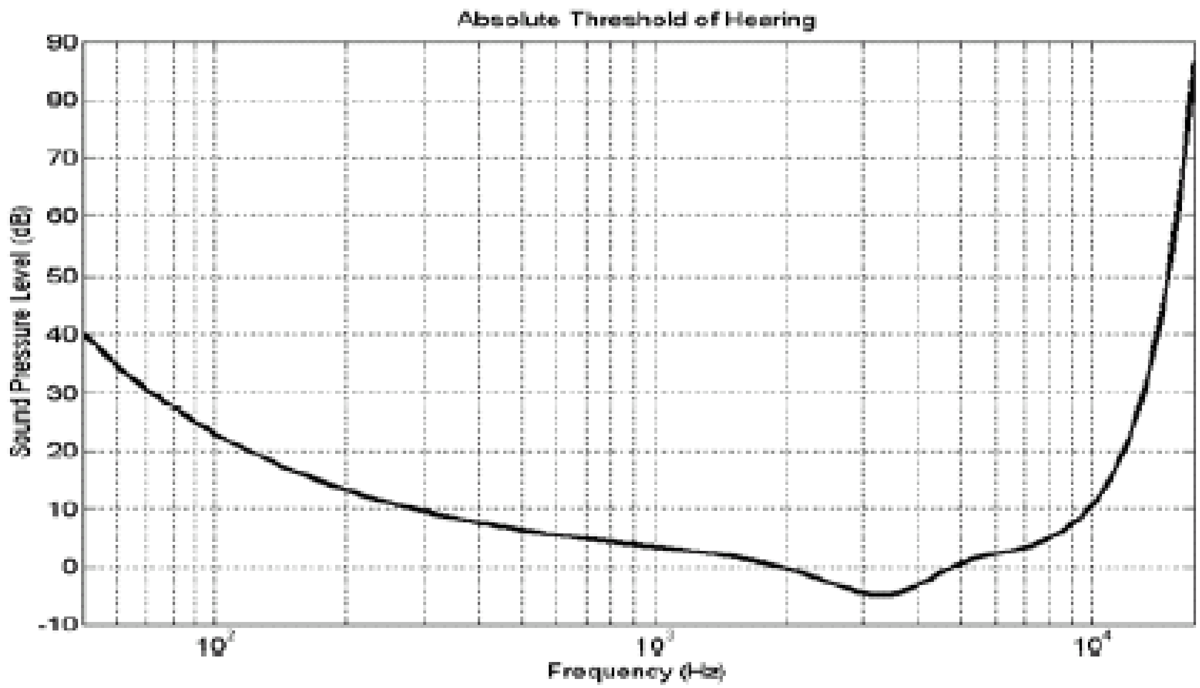
• Non-linear quantization techniques provide compression by requiring a smaller sample size to cover the full range of input than a linear quantization technique would





Principles of perceptual coding

Absolute threshold of hearing: amount of energy needed in a pure tone to be detected by a listener

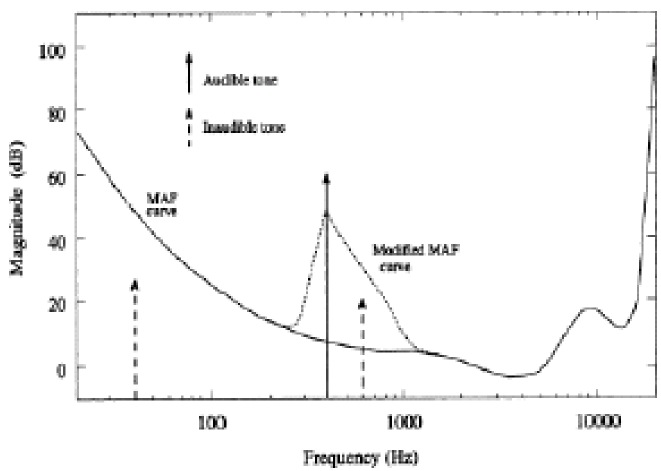


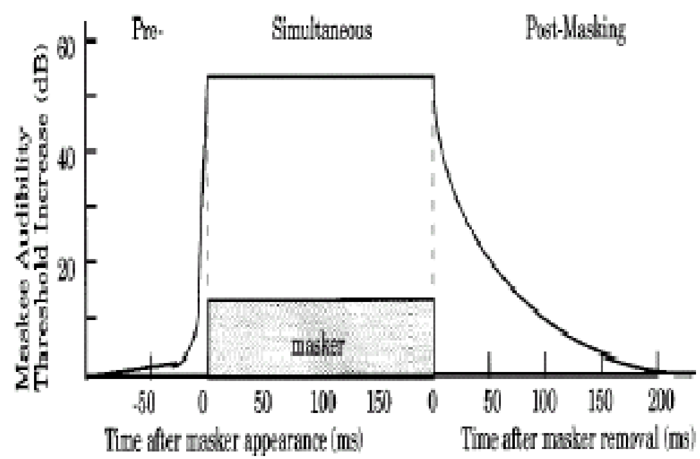
$$L_{SPL} = 20\log_{10}(p/p_0) \quad dB,\ p_0 = 2\times10^{-5}\, N/m^2$$

Principles of perceptual coding

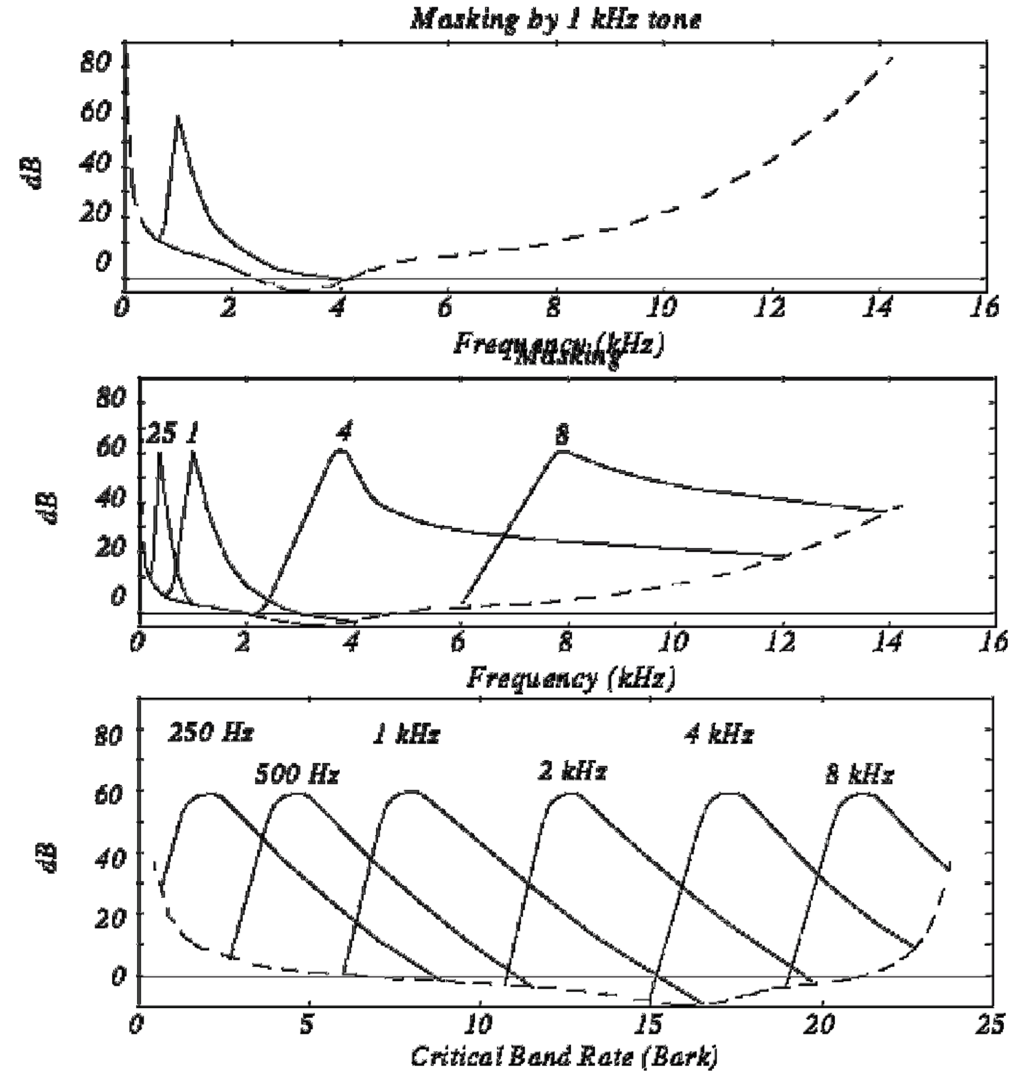
Frequency masking: A strong frequency peak renders nearyby frequencies non audible

Temporal masking: A strong peak render nearby frequencies in time inaudible



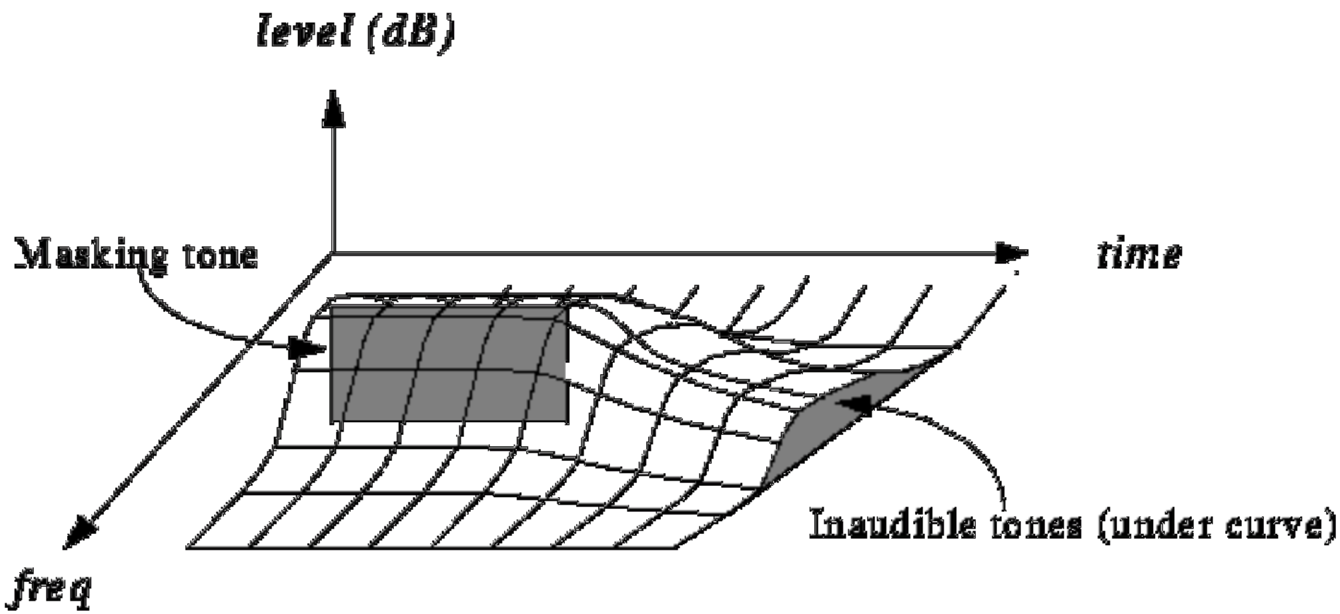


Principles of perceptual coding



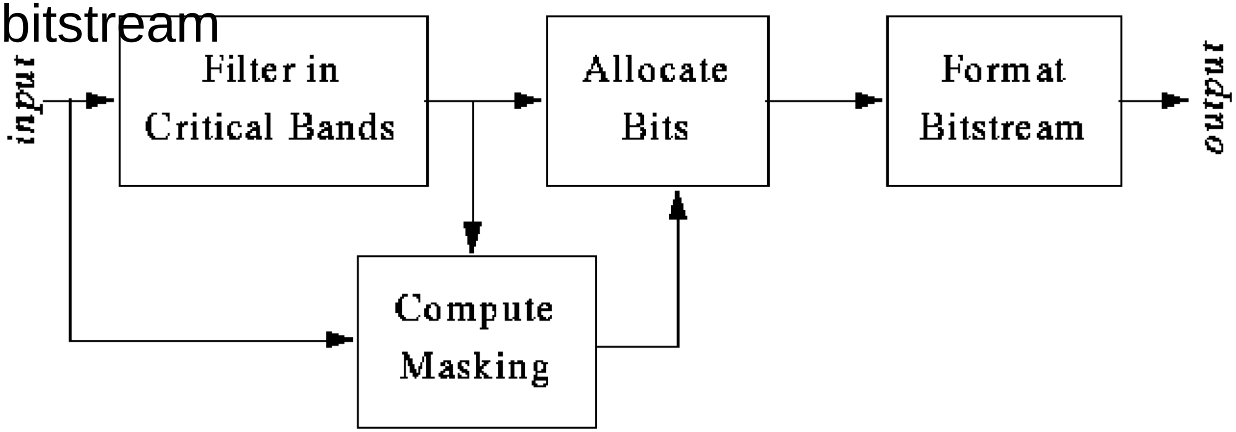
Principles of perceptual coding

3D view of frequency / temporal masking



Steps of perceptual coding

- Use convolution filters to divide the audio signal into 32 frequency sub-bands
- Determine the amount of masking for each band caused by nearby band using the *psycho-acoustic model*
- If the power in a band is below the masking threshold, do not encode it
- Otherwise, determine the number of bits needed to represent the coefficient so that noise introduced by quantization is below masking effect
- Format bitstream



Perceptual coding: Example

•After analysis, the first levels of 16 of the 32 bands are these:

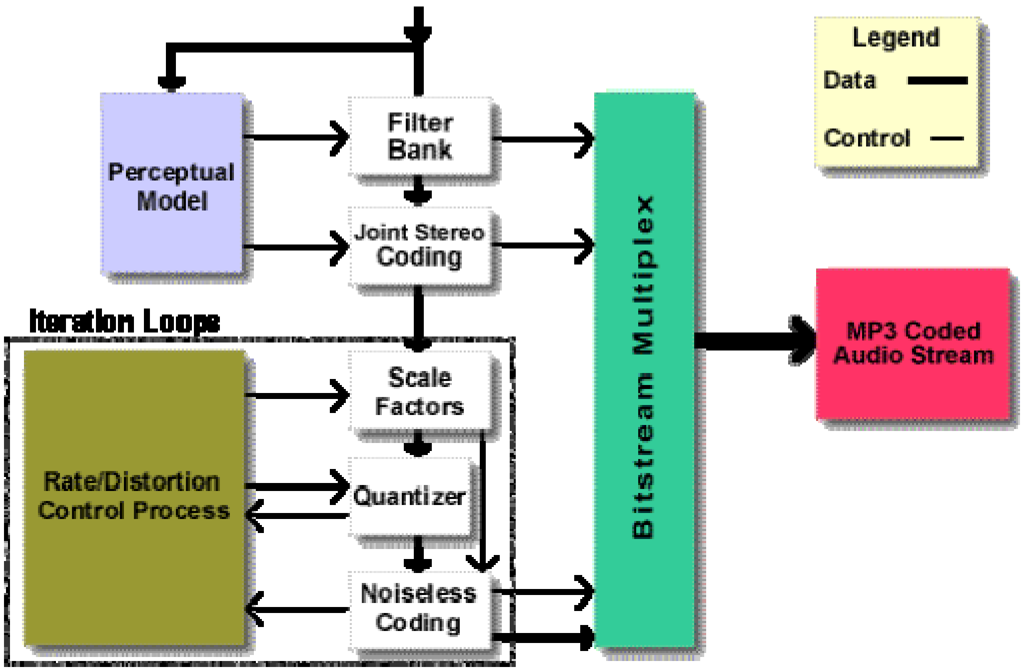
Band	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Level (db)	0	8	12	10	6	2	10	60	35	20	15	2	3	5	3	1

•The level in 8th band is 60dB -> masking of 12 dB in 7th band 15 dB in 9th band

•Level in 7th band is 10 dB -> Ignore

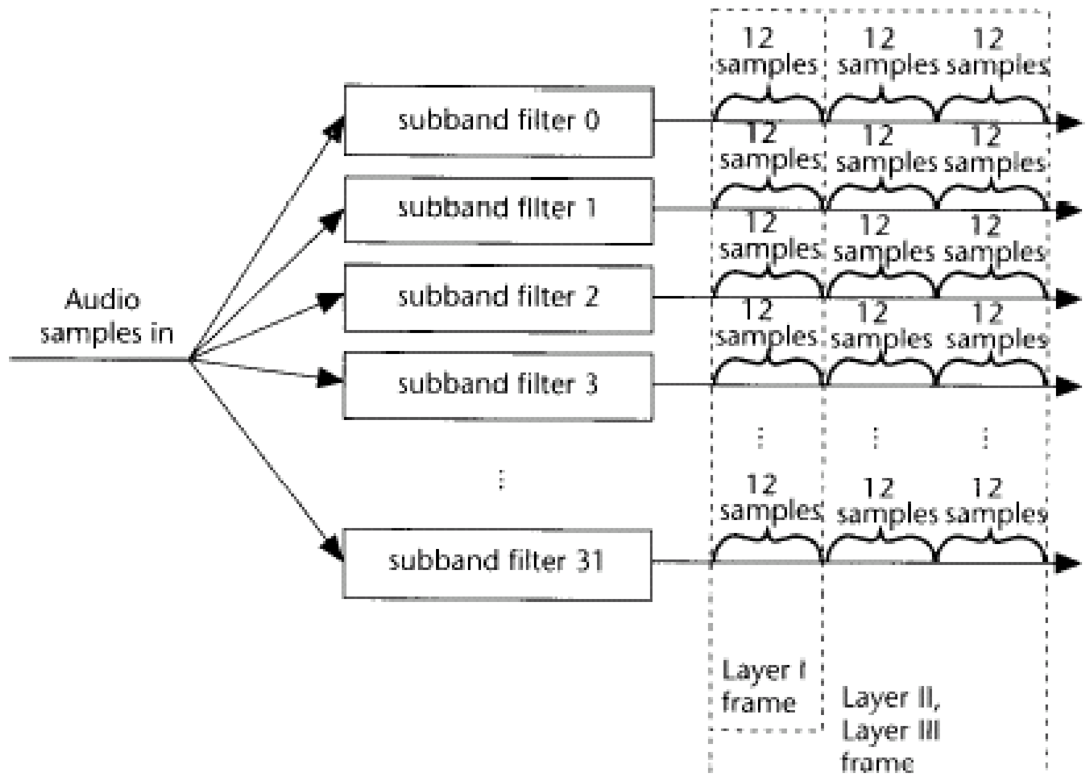
•Level in 9th band is 35 dB, but reduce number of bits according to masking (15 dB -> 2 bits reduce)

MP3 basic diagram

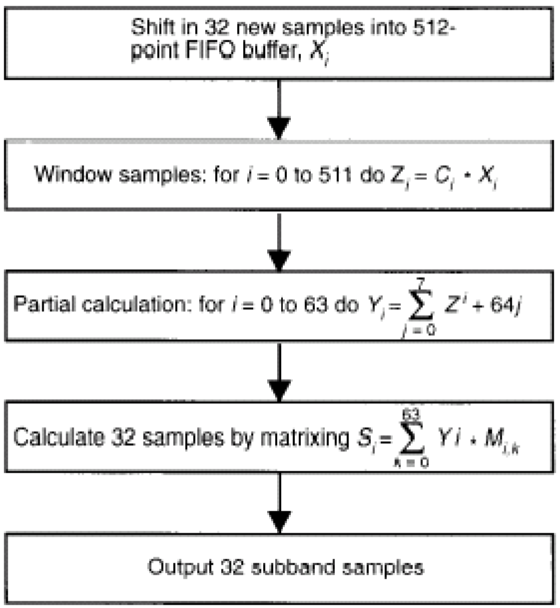


MPEG Audio Layers

•Divides data into frames, each of the contains 384 samples, 12 samples from each of the 32 filtered sub-bands
• L1: DCT type filter with or frame and equal frequency spread per band
• L2: Use three frames in filter (temporal masking)
• L3: Better critical band filter, psycho-acoustic model with temporal masking, stere redundancy, Huffman coder



MPEG Audio Filter bank

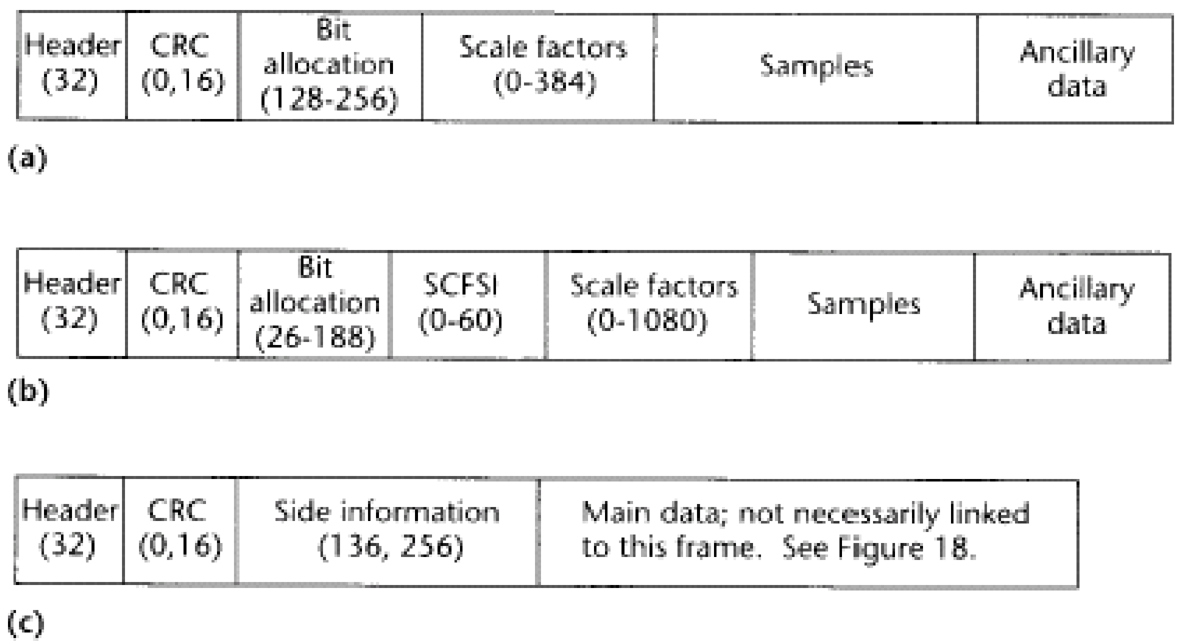


$$s_t[i] =$$
$$\sum_{k=0}^{63} \sum_{j=0}^{7} M[i][k] \times \left(C[k+64j] \times x[k+64j]\right)$$

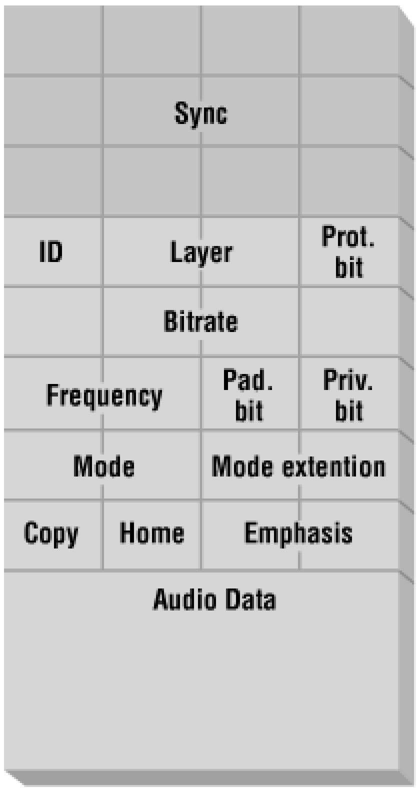
i – sub-band index
$s_t[i]$ – filter output sample at index i
C[n] – coefficient by standard
x[n] – audio input sample
M[i][k] - analysis matrix coefficients

$$M[i][k] = \cos\left[\frac{(2 \times i + 1) \times (k - 16) \times \pi}{64}\right]$$

MPEG Audio Stream format



MP3 header



Position Purpose Length (in Bits)

A Frame sync 11

B MPEG audio version (MPEG-1, 2, etc.) 2

C MPEG layer (Layer I, II, III, etc.) 2

D Protection (if on, then checksum follows header) 1

E Bitrate index (lookup table used to specify bitrate for this MPEG version and layer) 4

F Sampling rate frequency (44.1kHz, etc., determined by lookup table) 2

G Padding bit (on or off, compensates for unfilled frames) 1

H Private bit (on or off, allows for application-specific triggers) 1

I Channel mode (stereo, joint stereo, dual channel, single channel) 2

J Mode extension (used only with joint stereo, to conjoin channel data) 2

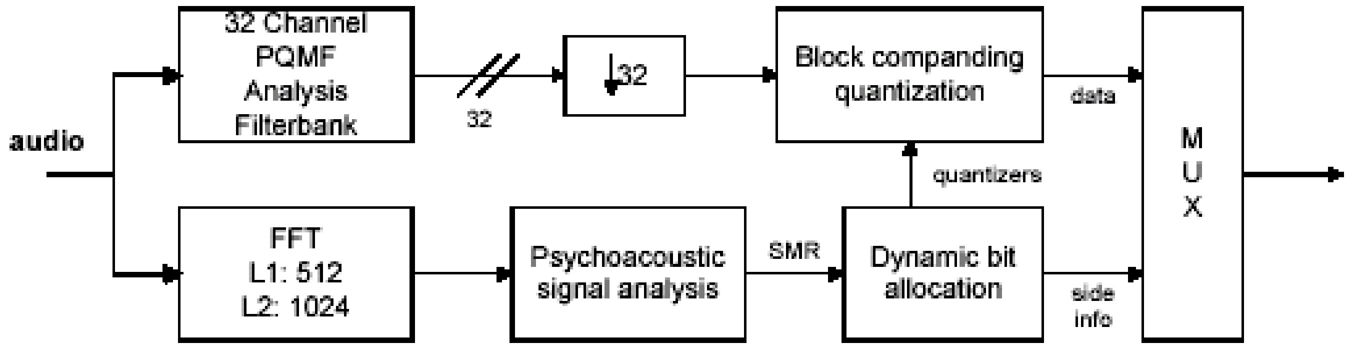
K Copyright (on or off) 1

L Original (off if copy of original, on if original) 1

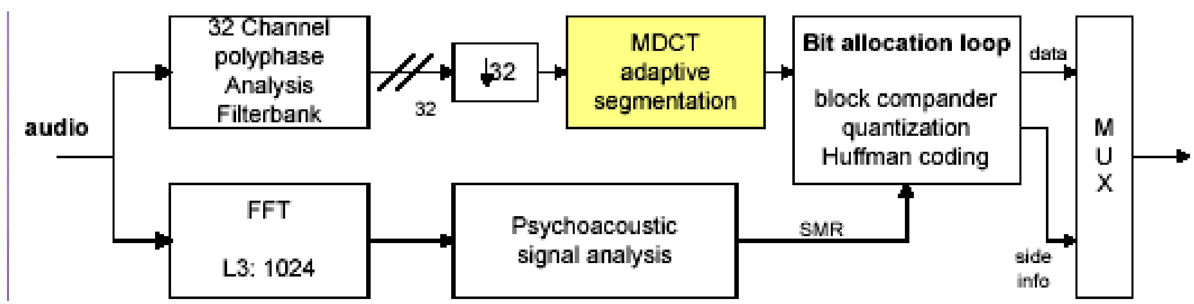
M Emphasis (respects emphasis bit in the original recording; now largely obsolete) 2

MPEG-1 Audio encoder block diagram

Layer I/II



Layer III



Bladeenc:

http://bladeenc.mp3.no

Encoding a chunk of PCM samples:

1. Rebuffer audio stream
2. Perform psychoanalysis on stream (FFT + energy calculations)
3. Perform the polyphase filtering
4. Apply mdct to the polyphase outputs
5. Assign bits to the outputs
6. Format the bitstream

More Details

- The filter bank used in MPEG Layer-3 is a hybrid filter bank which consists of a polyphase filter bank and a Modified Discrete Cosine Transform (MDCT).
  - This hybrid form was chosen for reasons of compatibility to its predecessors, Layer-1 and Layer-2.
- Quantization is done via a power-law quantizer.
  - This way, larger values are automatically coded with less accuracy and some noise shaping is already built into the quantization process.
- The quantized values are coded by Huffman coding.
  - Thus, it is called noiseless coding because no noise is added to the audio signal.

MPEG-2 Audio

- ISO/IEC 13818-3 BC/LSF (1994)
  - BC: backward compatible
  - LSF: low sampling frequency
  - Mono, stereo, support 16, 22.05, 24, 32, 44.1 and 48 kHz
  - Rate: 32-640 kb/s

- ISO/IEC 13818-7 NBC/AAC (1996)
  - NBC/AAC: Non-backward compatible/Advanced audio coding
  - Profiles: Main/ Low complexity (LC)/ Scalable sample rate (SSR)
  - 5-channel: left, right, center, surround left, surround right
  - Support 32, 44.1 and 48 kHz
  - Rate: 8-64 kb/s /channel

MPEG-2 AAC

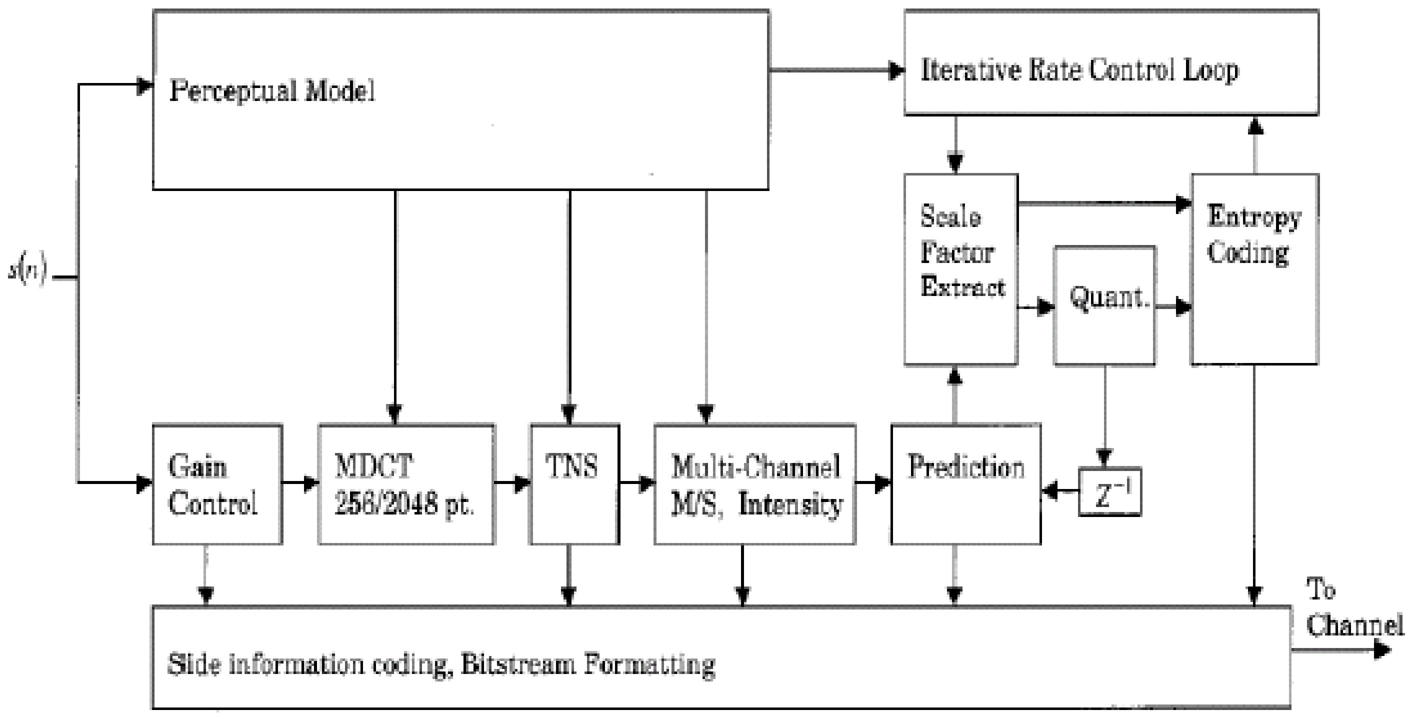
Advanced **a**udio **c**oding

Differences to MPEG Audio Layer 3

- Can handle 48 (full) + 16 (low freq) audio channels
- Pure MDCT filter bank
- Long / short time windows (avoiding pre-echo / transient handling)
- Prediction tool

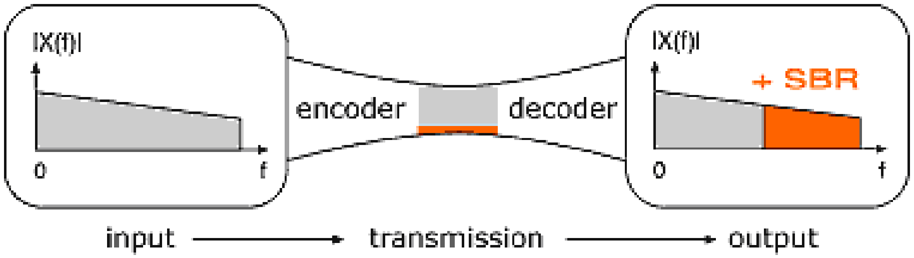
Licenses from FraunHofer-Geschellshaft

MPEG-2 AAC Diagram

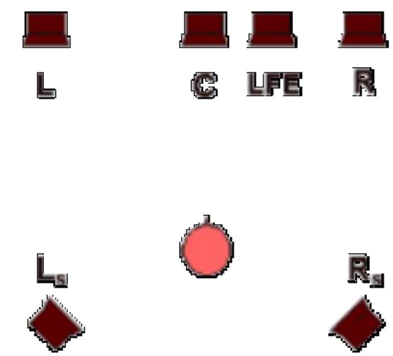


MP3Pro

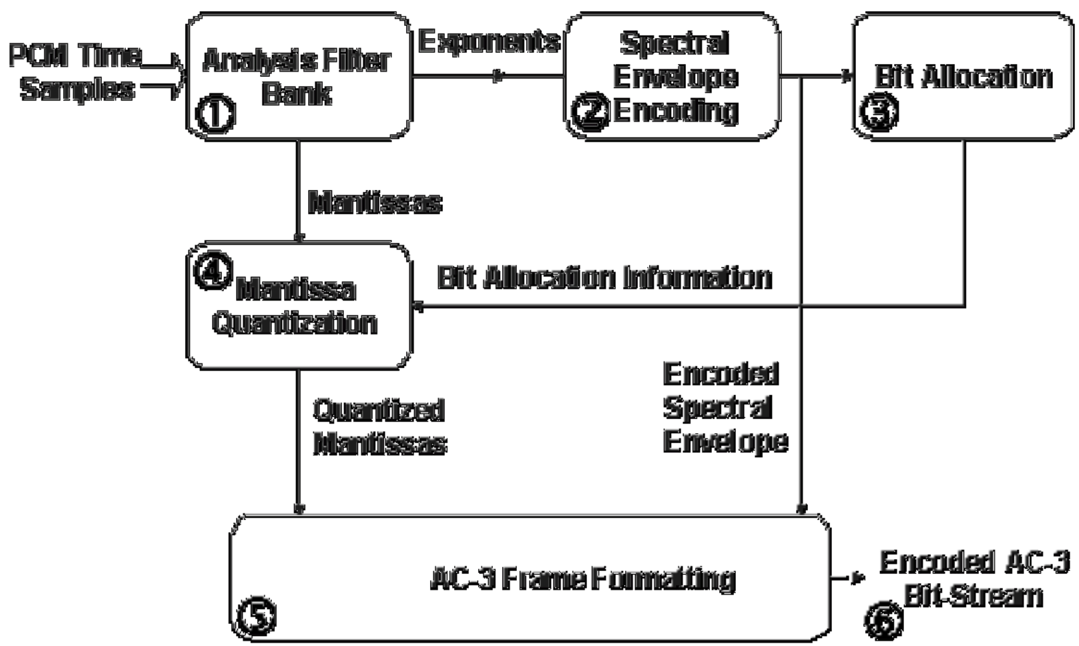
- Target 64 kbit/s
- up to 8kHz encoded as normal MPEG Audio Layer 3
- 8-16 kHz using Spectral Band Replication (SBR)
- ”Borrows” information of the 8-16 kHz band from lower frequency bands (hence ”replication”)



Dolby AC-3



- Also using masking as basic compression technique
- At cinemas (THX quality)



Applications of MPEG Audio

- MPEG-1 layer I: @ 384 kb/s, digital compact cassette (DCC)
- MPEG-1 layer II: @ 224 kb/s, direct broadcast satellite (DBS)
- MPEG-1 layer II: @ 256 kb/s, Eureka 147 digital audio broadcasting (DAB)
- MPEG-1layer III: MP3 music

- MPEG-2 BC/LSF: cinema, DigiTV
- MPEG-2 NBC/AAC: Internet, LiquidAudio, DRM, Xradio.

Coding examples

- Original Wave, PCM, 44100 Hz, 16 bit
- MPEG-1 Audio Layer 3, 32 kbit/s
- MPEG-1 Audio Layer 3, 64 kbit/s
- MPEG-1 Audio Layer 3, 128 kbit/s

Linear predictive coding

• Radical approach to compression of speech
• Uses mathematical model of vocal tract
• Instead of transmitting speech as audio samples, the parameters describing the state of the vocal tract are sent
• At the receiving end these parameters are used to reconstruct the speech by applying them to the same model
•Achieves very low data rates: 2.4kps
• Speech has a machine-like quality
• Suitable for accurate transmission of content, but not faithful rendition of a particular voice
• Similar concept to vector-coding of 2D and 3D graphics