



MPEG-1 lag 1, 2 og lag 3

Sverre Holm

Basert på presentasjon laget av
Torbjörn Ekman, 2005 (nå på NTNU)



Bitrater

- CD: $44.1 * 2 * 16 = 1.411$ Mbit/s
– 4 bit: 25% => 350 kbit/s låter forferdelig
- MP3, AAC etc: 128 kbit/s ~ CD/12
- Hva er det lure trikset?



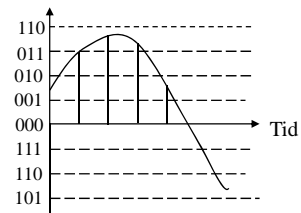
MPEG-1 Audio

Psychoacoustics in sound compression

- How do we hear?
- Digital representation of sounds
- Sound compression
- Psychoacoustics
 - Masking
 - Adaptive quantization
 - Bit allocation
- Filterbanks



INSTITUTT FOR INFORMATIKK

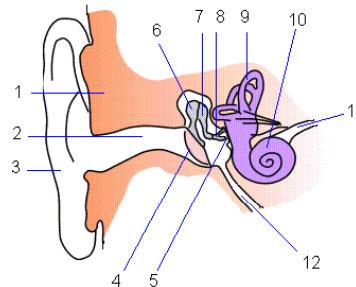


August 2009 - 9



Øret

1. Tinning
2. Øregang
3. Ytre øre
4. Trommehinne
5. Ovale vindu
6. Hammeren
7. Ambolt
8. Stigbøyle
9. Bueganger
10. Sneglehuset
11. Hørselnerve
12. Øretrompeten



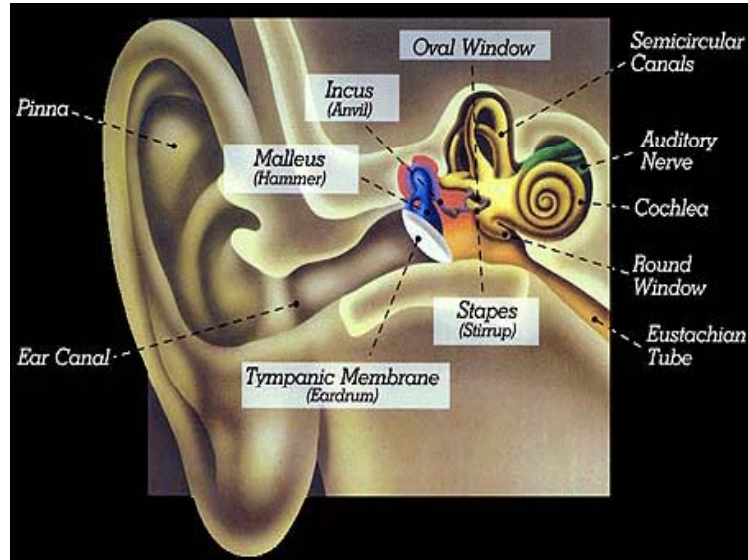
Wikimedia Commons

INSTITUTT FOR INFORMATIKK

August 2009 - 10

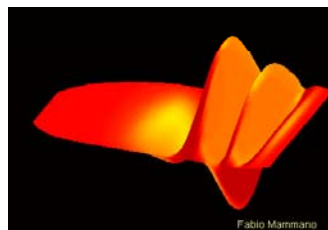
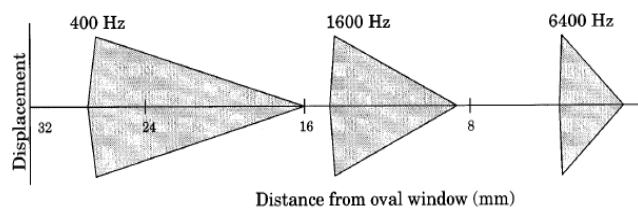


The Ear



The frequency filters of the ear: Mapping frequency to a location

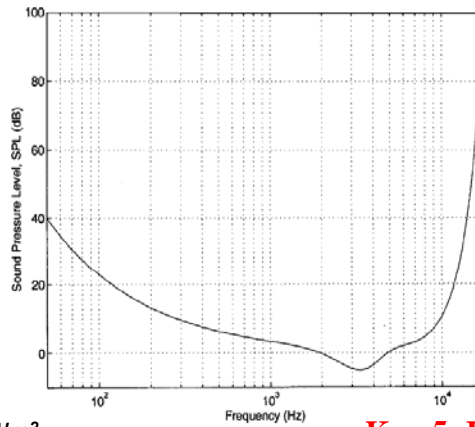
Unwound
cochlea



Kap 5: Frekvensanalyse



Threshold for audible sounds

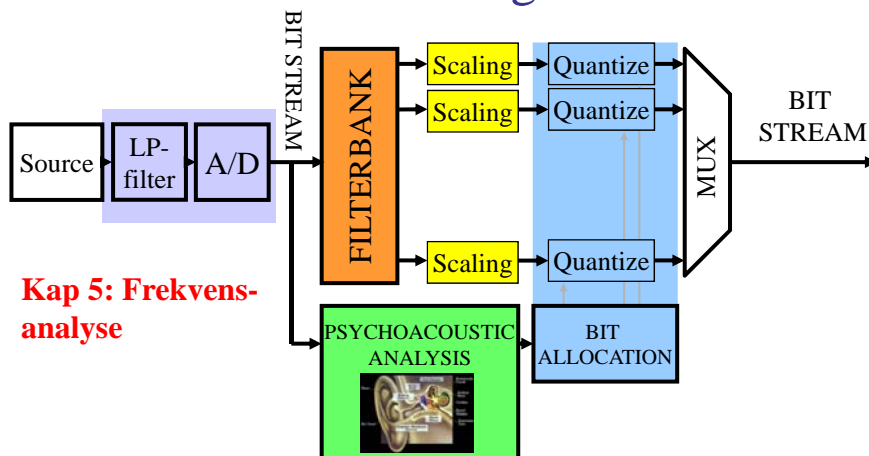


Reference 0 dB:
 $20 \mu\text{Pa} = 2 \cdot 10^{-5} \text{ N/m}^2$

Kap 5: Frekvensanalyse



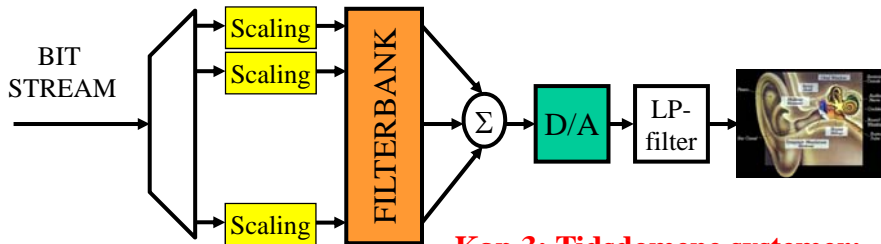
Filterbank Approach Encoding



**Kap 5: Frekvens-
analyse**



Decoding is much simpler

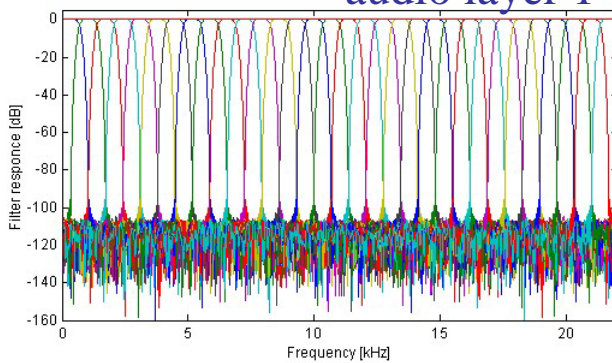


**Kap 3: Tidsdomene systemer:
linearitet**

Kap 3: Inverse systemer



Filterbanks in MPEG-1 audio layer 1-3



**Kap 5: Frekvens-
analyse av systemer**

Kap 6: Digitale filtre

**Kap 10: FIR
Filterdesign**

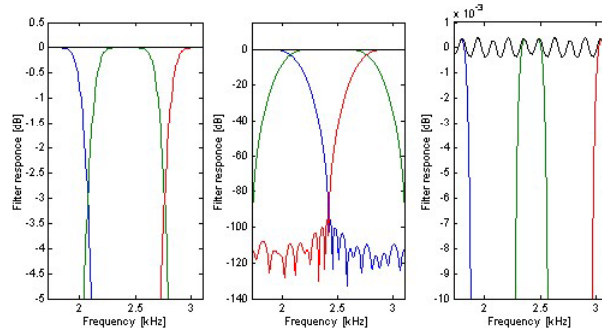
- Polyphase filterbank
- 32 subbands, e.g. bw
 $44100/2/32 = 689$ Hz
- 512 tap FIR-filters
- 80 adds and mults per output

- Equal width
- Not perfect reconstruction
- Frequency overlap

Kap 4: z-transform





A closer look



- The subbands overlap at 3 dB point with the adjacent bands.
- The leakage to the other bands is small.
- The total response almost adds up to one (0 dB).



White noise

- The white noise run through the filterbank.
- The samples from each band are played in the order of the subbands. 
- The reconstructed sequence 
 - The reconstruction error is -84 dB.

**Kap 7: Digital behandling av analoge signaler;
multirate signalbehandling**



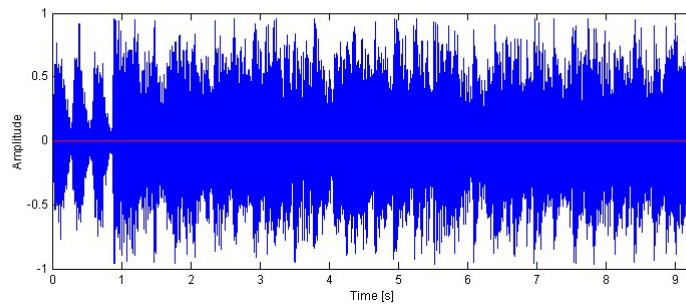
Reconstruction Using Nonideal Filterbanks


$$Y(e^{j\omega}) = X(e^{j\omega}) \underbrace{\frac{1}{M} \sum_{k=0}^{M-1} H_k^R(e^{j\omega}) H_k^A(e^{j\omega})}_{\approx 1} + \sum_{n=1}^{M-1} X(e^{j(\omega - \frac{2\pi n}{M})}) \underbrace{\frac{1}{M} \sum_{k=0}^{M-1} H_k^R(e^{j\omega}) H_k^A(e^{j(\omega - \frac{2\pi n}{M})})}_{\approx 0}$$

- In a perfect filterbank the first part is the only part.
- The second part consists of the aliasing terms.
- The filterbank is designed so that the aliasing is small.



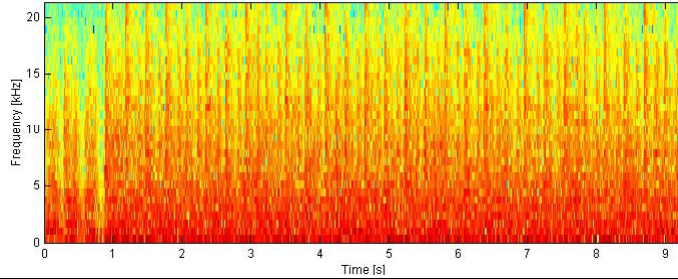
Tubthumper, a time domain view



The red line is the reconstruction error after splitting the signal in subbands, down sampling and applying the synthesis filterbank. The reconstruction error is -84 dB and sounds like 



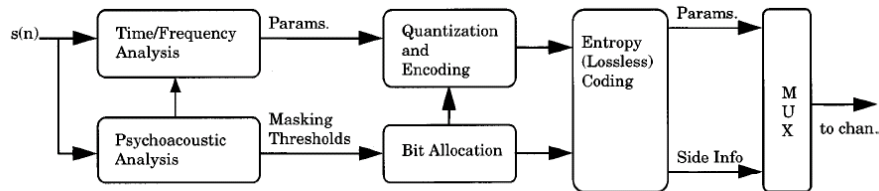
Tubthumper, frequency view



Subband	1	2	4	8	16	32
Center frequency [kHz]	0.3	1.0	2.4	5.2	10.7	21.7
No subsampling						
Subsampled 32 times						



What is this Psychoacoustics that is used in the Encoder ?



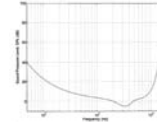
Kap 8: Diskret Fourier Transform; Estimering av effektspektrum



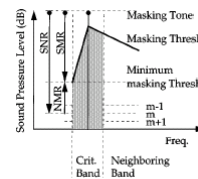
Masking

We do not hear all sounds.

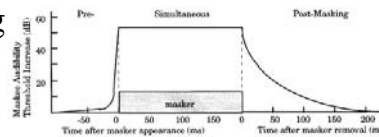
1. Absolute threshold of hearing: →
2. Masking: One sound is inaudible in the presence of another sound.



1. Simultaneous masking
 - Noise Masking Tone
 - Tone Masking Noise
 - Noise Masking Noise

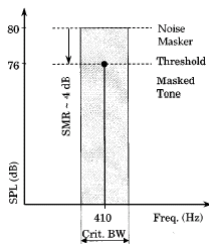


2. Nonsimultaneous masking
 - Pre masking (2 ms)
 - Post masking (100 ms)



Noise Masking Tone

Filtered Noise Center 410 Hz Width 111 Hz	Tone 1, 820 Hz 5 dB below noise	Tone 2, 410 Hz 5 dB below noise	Noise + Tone 1	Noise + Tone 2
			Not masked 	Masked



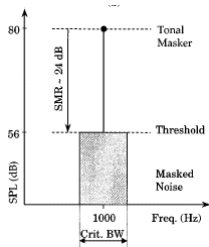
You can not hear a sinusoid that lies in the same critical band as a filtered noise if the sound pressure level is below a certain threshold.

This effect also stretches out beyond the critical band.



Tone Masking Noise

Filtered Noise Center 1 kHz Width 162 Hz 15 dB below	Tone 1, 2 kHz	Tone 2, 1 kHz	Noise + Tone 1	Noise + Tone 2
			Not masked 	Masked



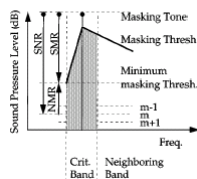
You can not hear a filtered noise that lies in the same critical band as a sinusoid if the sound pressure level is below a certain threshold.

This effect also stretches out beyond the critical band.



Exploit Masking

- If a sound is masked we can't hear it.

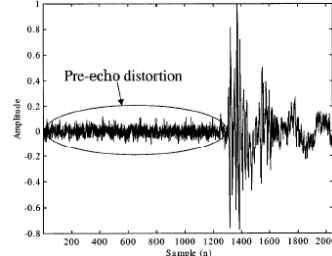
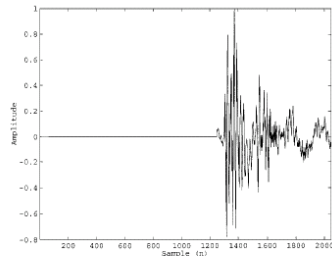


- Make a frequency analysis of the signal and find the masking threshold.
- Put the quantization noise under the masking threshold and we won't hear the quantization.

**Kap 8: DFT, Fast Fourier transform,
Estimering av effektspektrum**



Pre echo distortion



- The original sound of a castanet.
- The abruptness in time domain result in all frequencies being involved.
- The data is split in two windows of finite length.
- The quantization noise is spread over a whole window.
- This makes the castanets sound less distinct.
- Audible effects can be avoided with shorter windows, exploiting premasking.



Vindus-svitsjing: 1.1 og 1.3 (= MP3)

- Blokkstørrelse i transform og delbåndskodere:
 - Små blokker: god transientgjengivelse, dårlig koding pga mye overhead
 - Store blokker: god kodingsgevinst; gir pre-ekko
- Vindus-svitsjing mellom $N=64$ og 1024 blokkstørrelse
 - Små blokker ved ikke-stasjonæritet
 - Ellers store blokker



Scale factors and Quantization

- When the dynamics change over time, only a small subset of the quantization steps are used in regions with low magnitudes.
- Use scale factors instead:
 - Take a window of data.
 - Find the max magnitude in this window.
 - Use the next larger scale factor from a table.
 - Normalize with the scale factor.
 - Quantize.
 - Now the whole dynamic range of the quantizer is used.
 - Send scale factor and quantized samples.

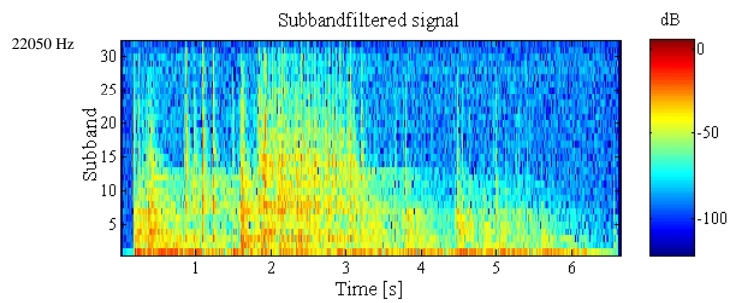
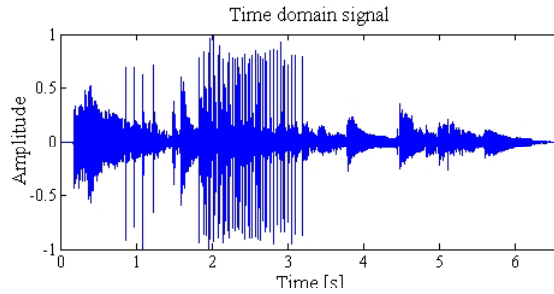


Bit Allocation and Masking

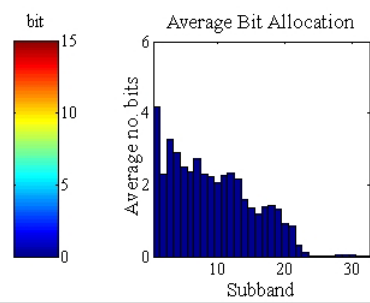
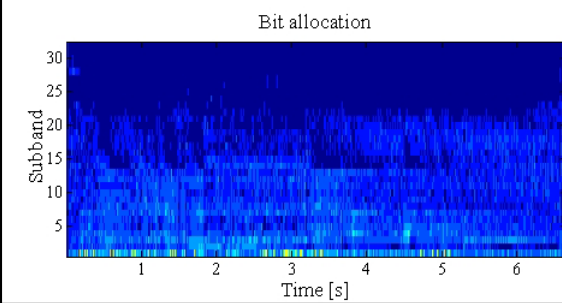
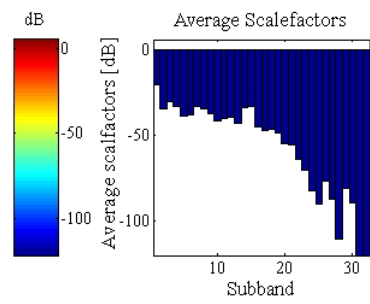
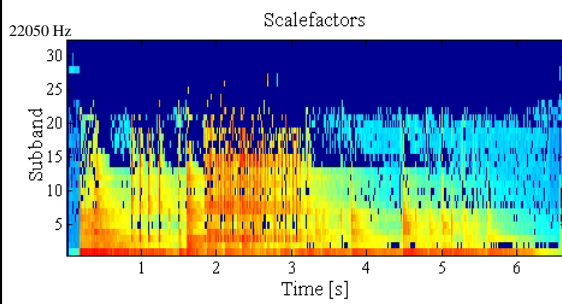
- The masking threshold in each subband gives the Just Noticeable Distortion (JND) limit for that band.
- Bits are assigned to subbands so that the quantization noise falls below or as little over the JND as possible.
- Then the Signal to Quantization noise Ratio (SQR) falls below JND



Castanets and Guitar

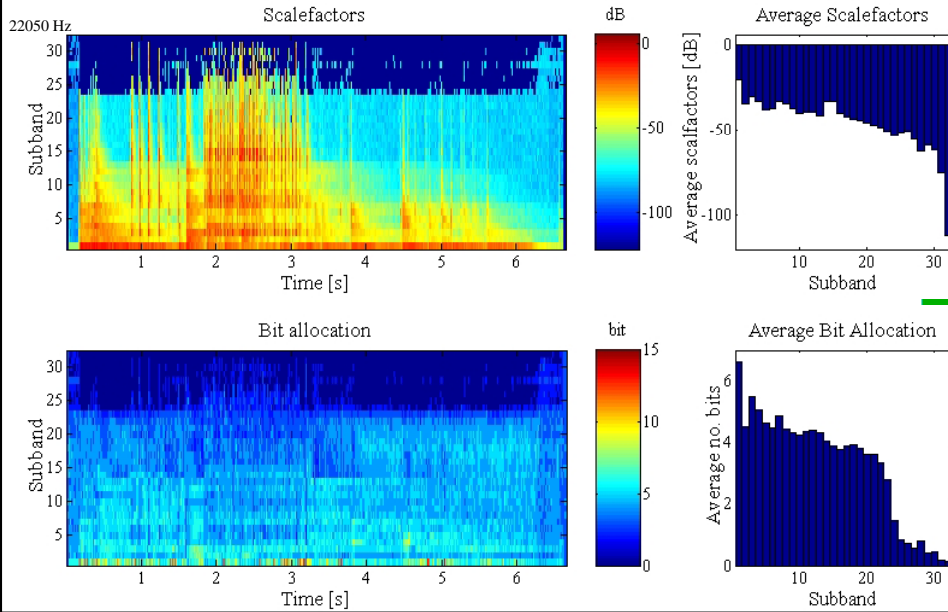


Bit allocation with 2 bits per sample

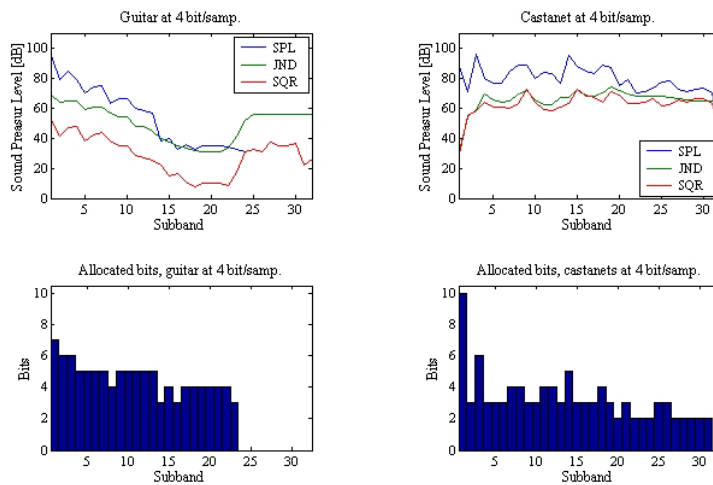




Bit allocation with 4 bits per sample



Signal to Quantization Noise Ratio and the Just Noticeable Distortion
















Frame at $t=0.6$ s

Frame at $t=1.1$ s

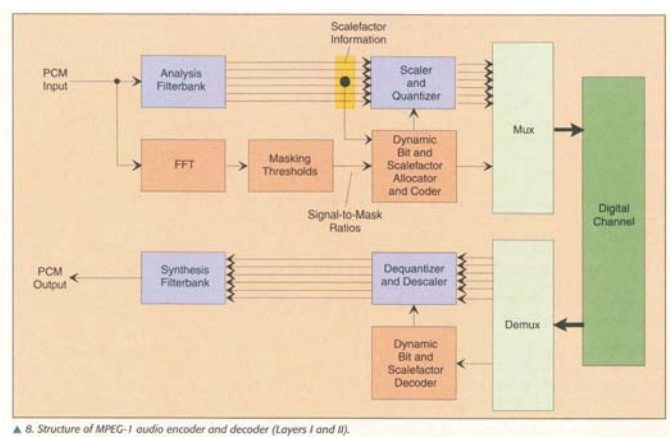


Examples on compression

Compression	2	4	8
MP1		4 bit 	2 bit 
MP1 error (SQR)		22 dB 	11 dB 
Direct Quantization	8 bit 	4bit 	2 bit 
Direct Quantization Error (SQR)	31 dB 	7.8 dB 	1.1 dB 
Downsampling to 22 kHz bandwidth and quantization	16 bit 	8 bit 	4 bit 



MPEG-1 layers I and II

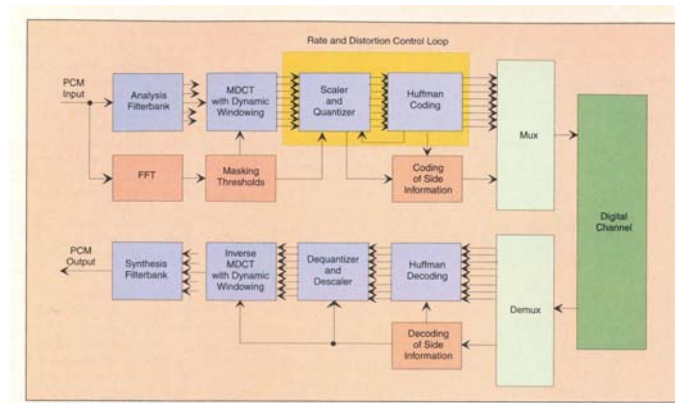


▲ 8. Structure of MPEG-1 audio encoder and decoder (Layers I and II).

P. Noll, MPEG digital audio coding, IEEE Sign. Proc. Mag., Sep 1997



MPEG-1 layer III = MP3



▲ 14. Structure of MPEG-1 audio encoder and decoder (Layer III).

P. Noll, MPEG digital audio coding, IEEE Sign. Proc. Mag., Sep 1997



MPEG-1 audio (ca 1990)

- Lag I: Delbåndskoding i 32 like frekvensbånd, 512 koeffisienters polyfase kvadratur speilfiltre og psykoakustisk modell som bestemmer adaptiv bit-tilordning, rammelengde 8 ms
 - ~192 kbit/s pr kanal for CD-kvalitet, ~384 kbit/s for stereo
- Lag II: Rammelengde 24 ms
 - 92 kbit/s pr kanal, 192 kbit/s for stereo
- Lag III: kaskadekople en 6 eller 18 punkts (dynamisk vindus-svitsjing) MDCT med lag I's filterbank
 - => $32 * 18 = 576$ frekvensbånd =>
 - 64 kbit/s pr kanal (variabel) (128 kbit/s for stereo)
- MPEG-1, layer III = MP3



DAB = MPEG-1, lag 2

- Digital Audio Broadcasting
 - Koder fra slutten av 80-tallet, standard fra 1991
 - Bitrate stereo: (112), 128, 160 (~FM), 192 kbit/s
 - Mono: ned til 56 kbit/s (nyheter)
 - England: Mange musikkstasjoner i mono, f.eks 96 kbit/s
- DAB og lyd kvalitet:
 - I praksis aldri bedre enn god FM, ofte dårligere:
<http://heim.ifi.uio.no/~sverre/DAB.shtml>
 - S. Holm, "Audio quality on the air in DAB digital radio in Norway," in Proc. 31st Audio Engineering Society International Conference, London, UK, 2007.
 - De fleste land velger nå DAB+, også Danmark, som er kommet mye lenger enn Norge med innføring av DAB



MPEG-2 AAC - april 1997

- MPEG-2 AAC (Advanced Audio Coding)
 - Samplingsrater fra 8 til 96 kHz
 - opp til 48 audiokanler
 - 320 kbit/s for 5.1 kanaler høykvalitets audio, kompresjon 12:1 i forhold til CD
 - Høyoppløselig filterbank (1024 punkts mod. cosinustransform),
 - støyforming i tid, prediksjonsteknikker



AAC's improvements over MP3

Advanced Audio Coding is designed to be the successor of the MP3 format and demonstrates greater sound quality and transparency than MP3 files coded at the same bit rate.

- More sample frequencies (from 8 kHz to 96 kHz) than MP3 (16 kHz to 48 kHz)
- Up to 48 channels (MP3 supports up to two channels in MPEG-1 mode and up to 5.1 channels in MPEG-2 mode)
- Arbitrary bit-rates and variable frame length. Standardized constant bit rate with bit reservoir.
- Higher efficiency and simpler filterbank (rather than MP3's hybrid coding, AAC uses a pure MDCT)
- Higher coding efficiency for stationary signals (AAC uses a blocksize of 1024 or 960 samples, allowing more efficient coding than MP3's 576 sample blocks)
- Higher coding accuracy for transient signals (AAC uses a blocksize of 128 or 120 samples, allowing more accurate coding than MP3's 192 sample blocks)
- Can use Kaiser-Bessel derived window function to eliminate spectral leakage at the expense of widening the main lobe
- Much better handling of audio frequencies above 16 kHz
- More flexible joint stereo (different methods can be used in different frequency ranges)
- Adds additional modules (tools) to increase compression efficiency: TNS, Backwards Prediction, PNS etc... These modules can be combined to constitute different encoding profiles.
- Wikipedia



Kapitler i læreboka – MP3-koding

1	Overview	<ul style="list-style-type: none"> • Praktiske eksempler som mp1/mp3
2	Discrete Signals	<ul style="list-style-type: none"> • Musikk og talesignal
3	Time-Domain Analysis	<ul style="list-style-type: none"> • Linearitet: kan dele i frekvensbånd, behandle for seg og så sette sammen igjen • Tidsinvarians gjelder bare over kort tid for musikk og tale. • Differanseligninger: FIR filtre
4	z-Transform Analysis	<ul style="list-style-type: none"> • Analyse av filter i filterbanken: nøkkel til å få til filterdesign
5	Frequency Domain Analysis	<ul style="list-style-type: none"> • Frekvensdomene er sentralt i modell av hørsel • Frekvensselektive filtre: båndpassfiltre • Inverse systemer: kan dele i bånd i koder og addere sammen igjen i dekker
6	Filter Concepts	<ul style="list-style-type: none"> • Filterstrukturer, hvordan implementere filterbank i koder og dekker
7	Digital Processing of Analog Signals	<ul style="list-style-type: none"> • A/D-analyse: kvantiseringsstøy ved direkte sampling • Multirate system: Hvert delfilter kan nedsamples pga bare 1/32 av total båndbredde => trenger bare 1/32 samplerate per filter
8	The Discrete Fourier Transform and Its Applications.	<ul style="list-style-type: none"> • Frekvensanalyse av signaler • FFT brukes i estimering av spektrum i koder. • Må estimere korttidsspektrum for å gjøre adaptiv bittildeling
9	Design of IIR Filters.	
10	Design of FIR Filters.	<ul style="list-style-type: none"> • Hvordan finne koeffisienter til bp-filtrene i filterbanken?
11	MATLAB Examples	
A	Useful Concepts from Analog Theory	