MPEG-1 lag 1, 2 og lag 3

Sverre Holm

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

MPEG audiokoding

- Motivasjon for de fleste kapitlene i Ambardar, Digital signal processing: A Modern Introduction, Thomson, 2007.

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?

Direktesampling (PCM)

KILDE

Unngå aliasing
i samplingen

LP-
FILTER

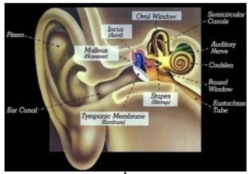
Bruk nok bit til å få
liten nok
kvantiseringsfeil

A/D

BITSTRØM

D/A

LP-
FILTER

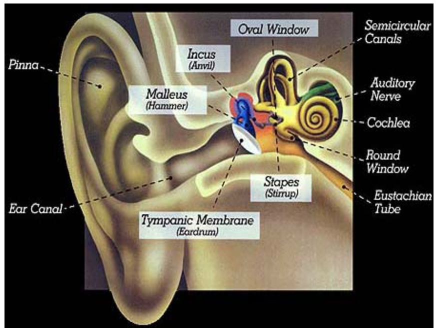


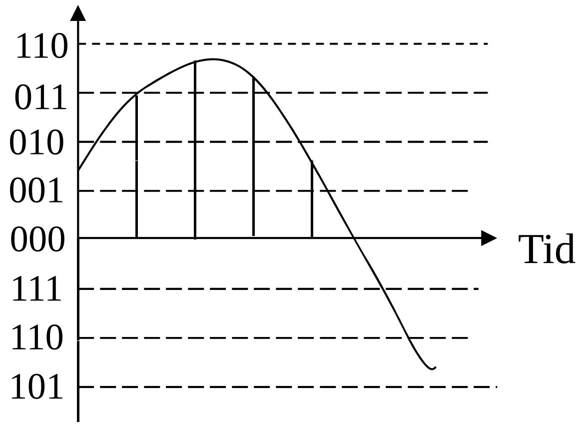
KODER

DEKODER

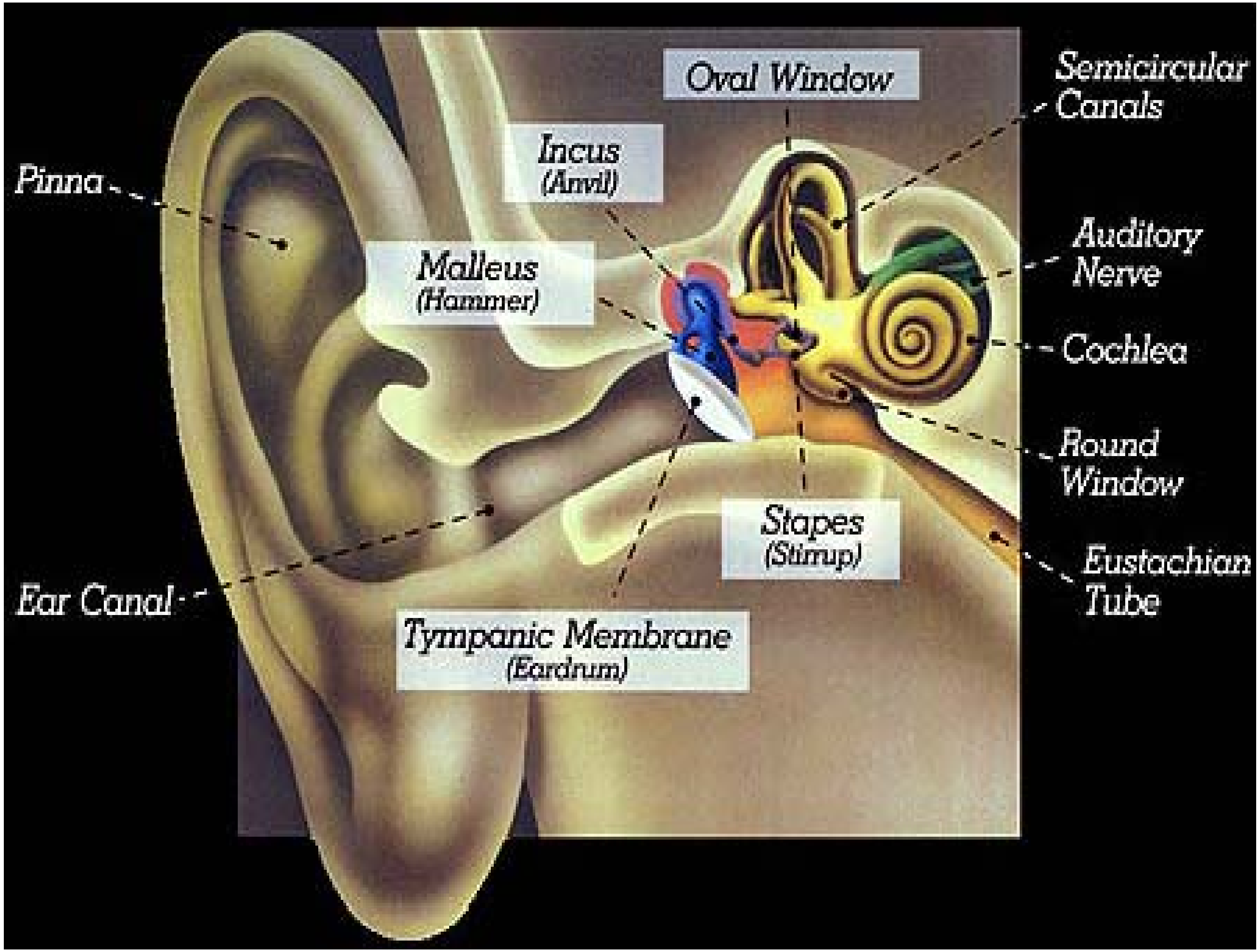
MPEG-1 Audio
Psychoacoustics in sound compression

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





The Ear

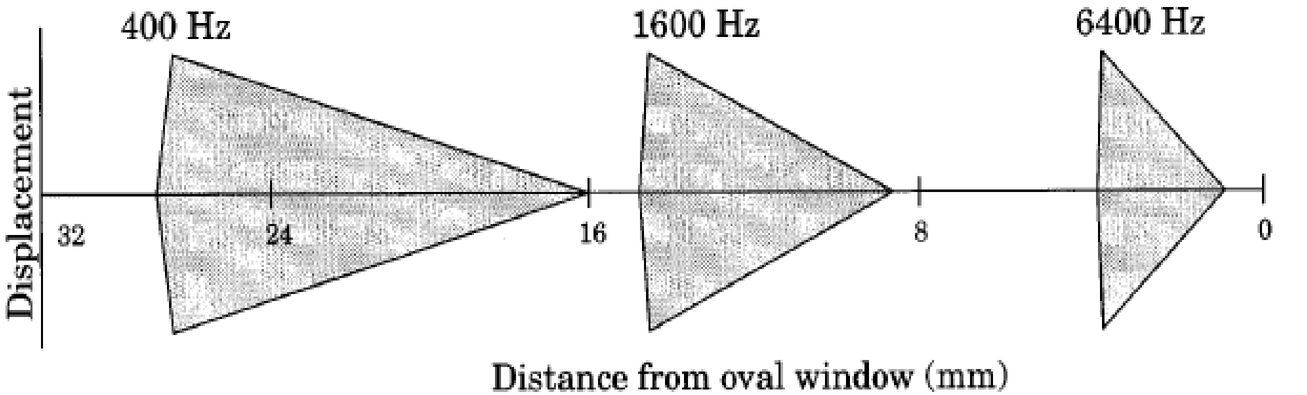

Oval Window
Semicircular Canals
Incus (Anvil)
Pinna
Auditory Nerve
Malleus (Hammer)
Cochlea
Round Window
Stapes (Stirrup)
Eustachian Tube
Ear Canal
Tympanic Membrane (Eardrum)

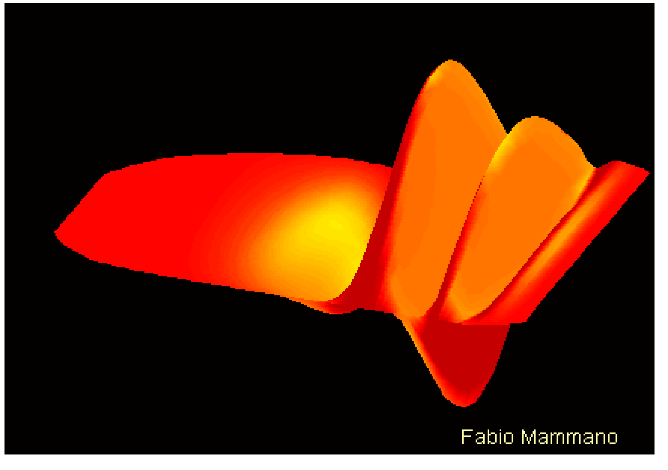
UNIVERSITETET
I OSLO

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

Unwound

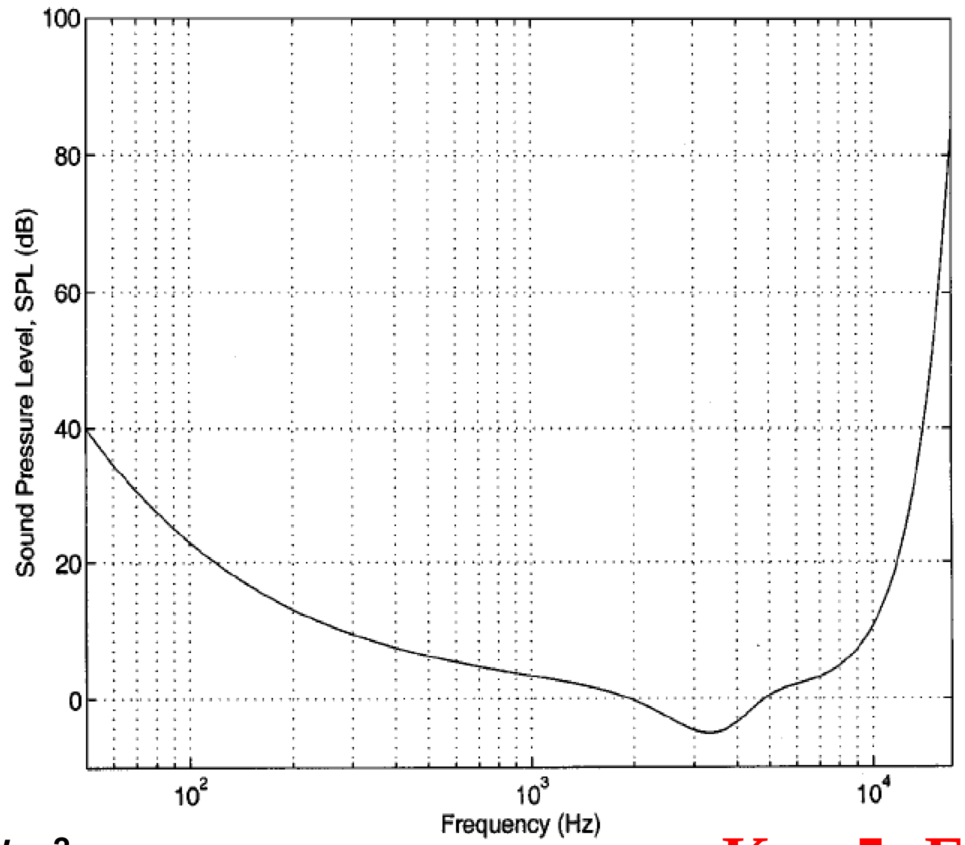
cochlea





Kap 5: Frekvensanalyse

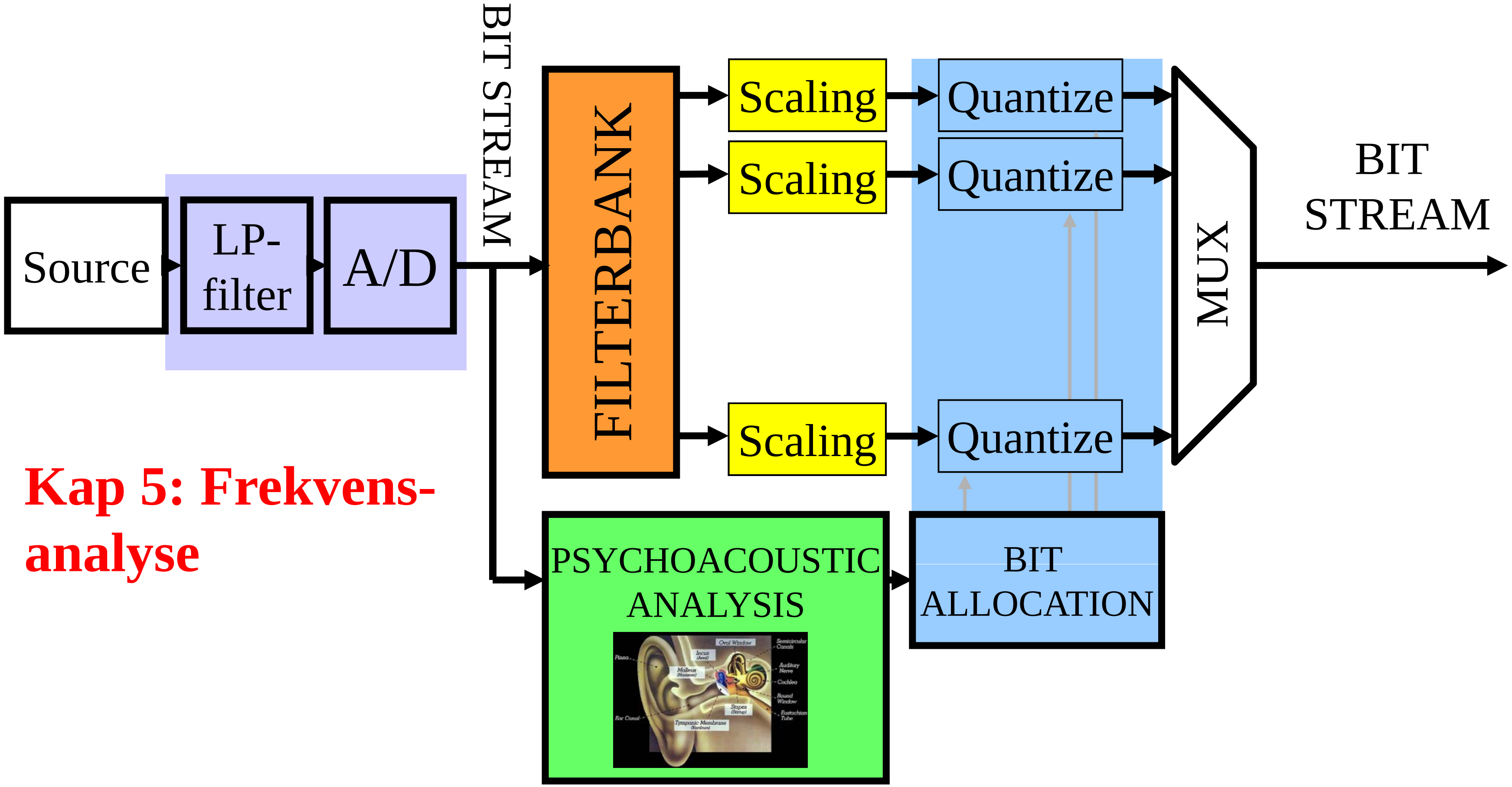
Threshold for audible sounds



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

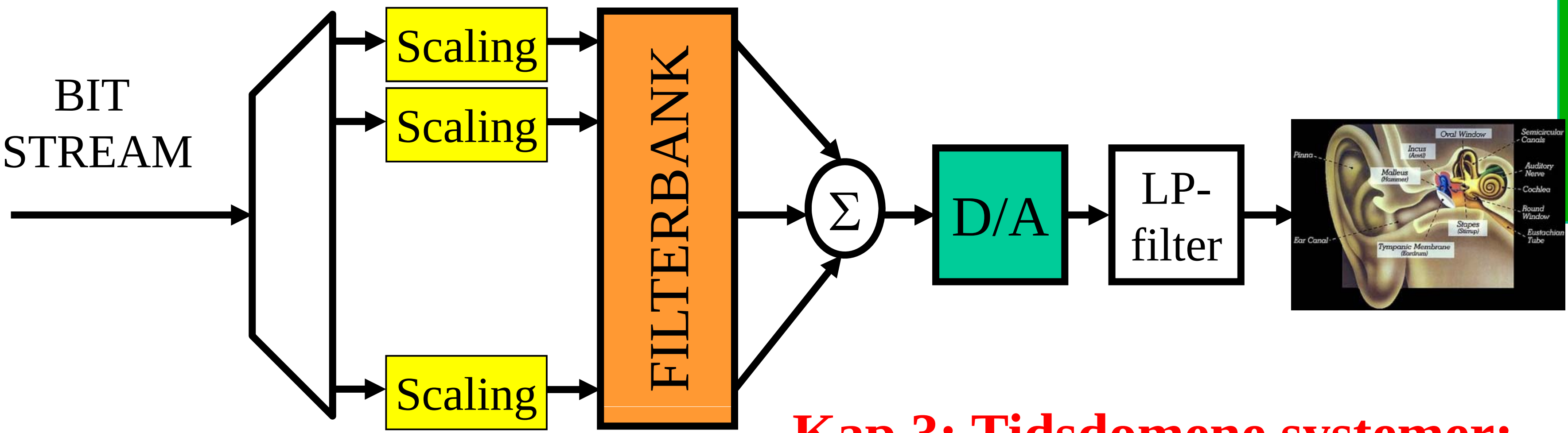
Kap 5: Frekvensanalyse

Filterbank Approach Encoding



Kap 5: Frekvens-analyse

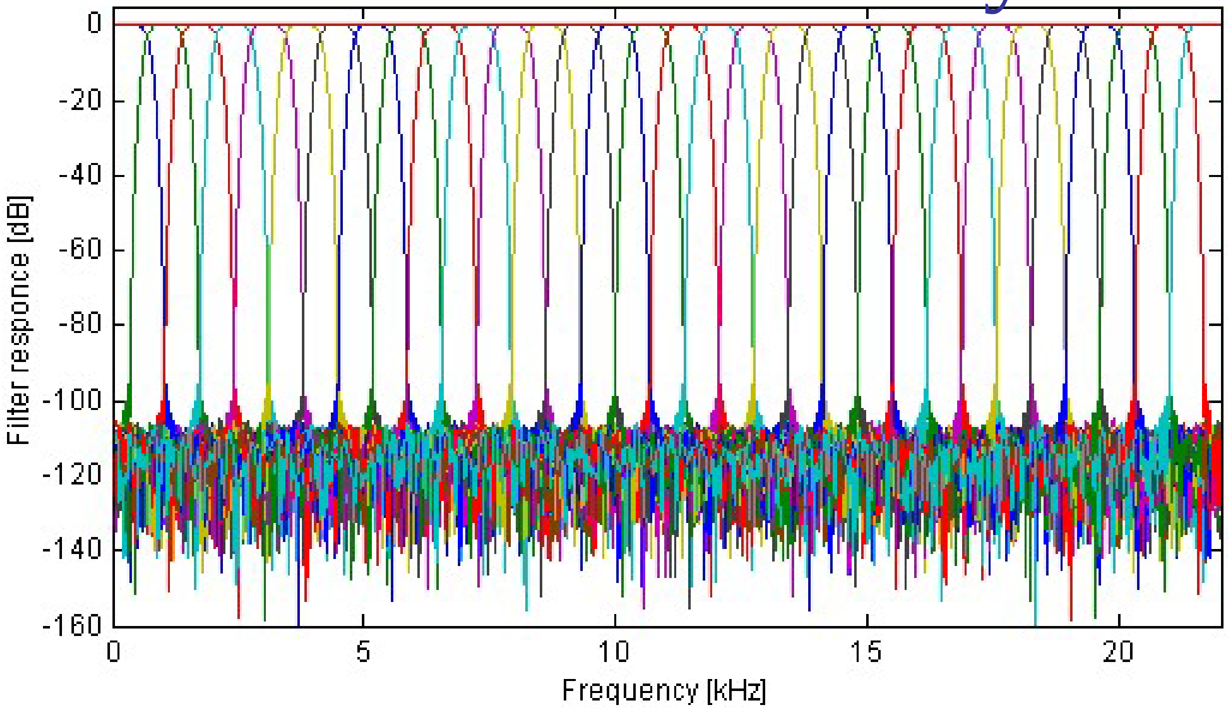
Decoding is much simpler



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

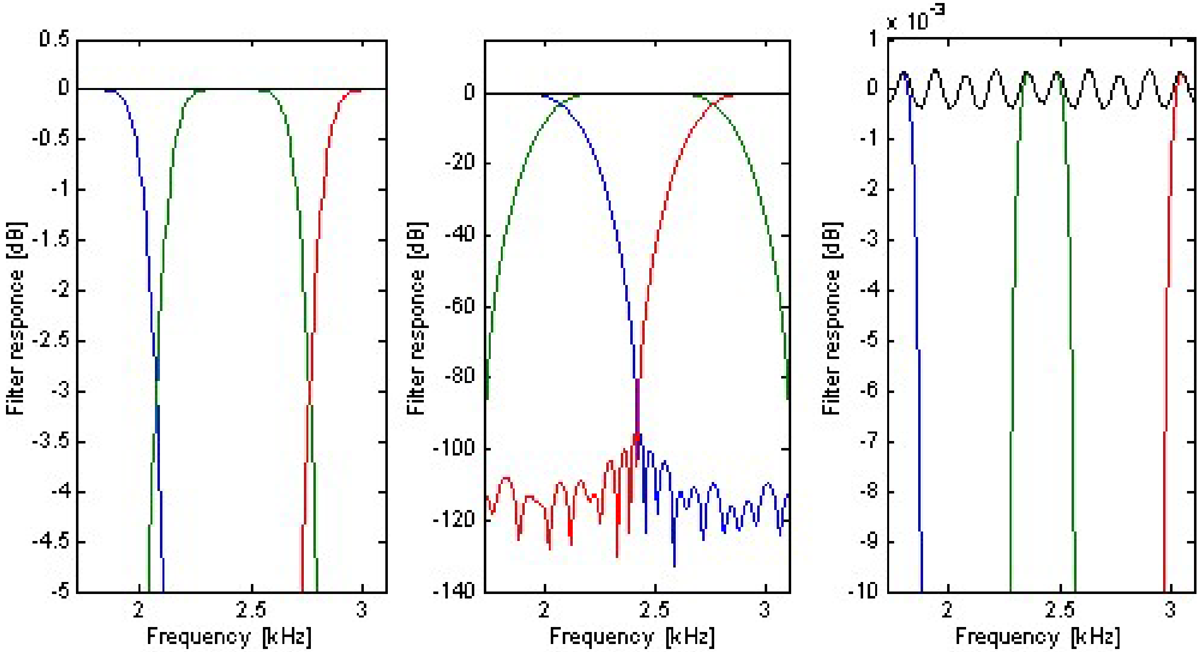
- Polyphase filterbank
- 32 subbands, e.g. bw 44100/2/32 = 689 Hz
- 512 tap FIR-filters
- 80 adds and mults per output

- Equal width
- Not perfect reconstruction
- Frequency overlap

Kap 4: z-transform

UNIVERSITETET
I OSLO

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

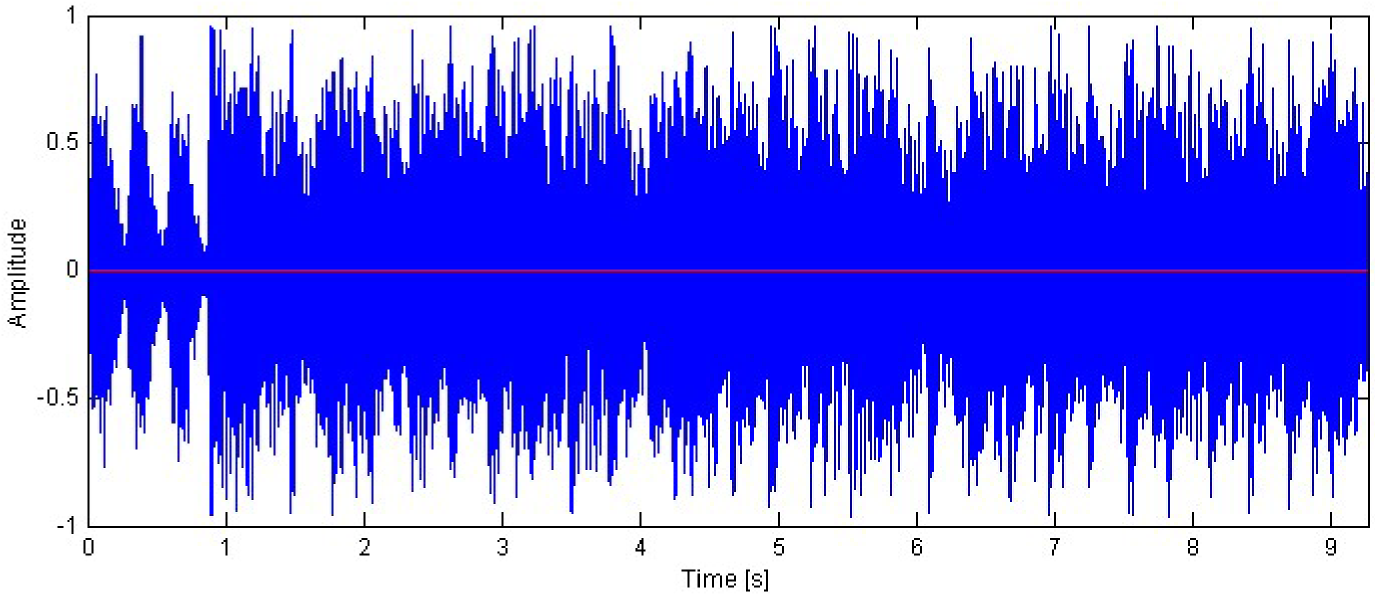
Reconstruction Using Nonideal Filterbanks

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

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

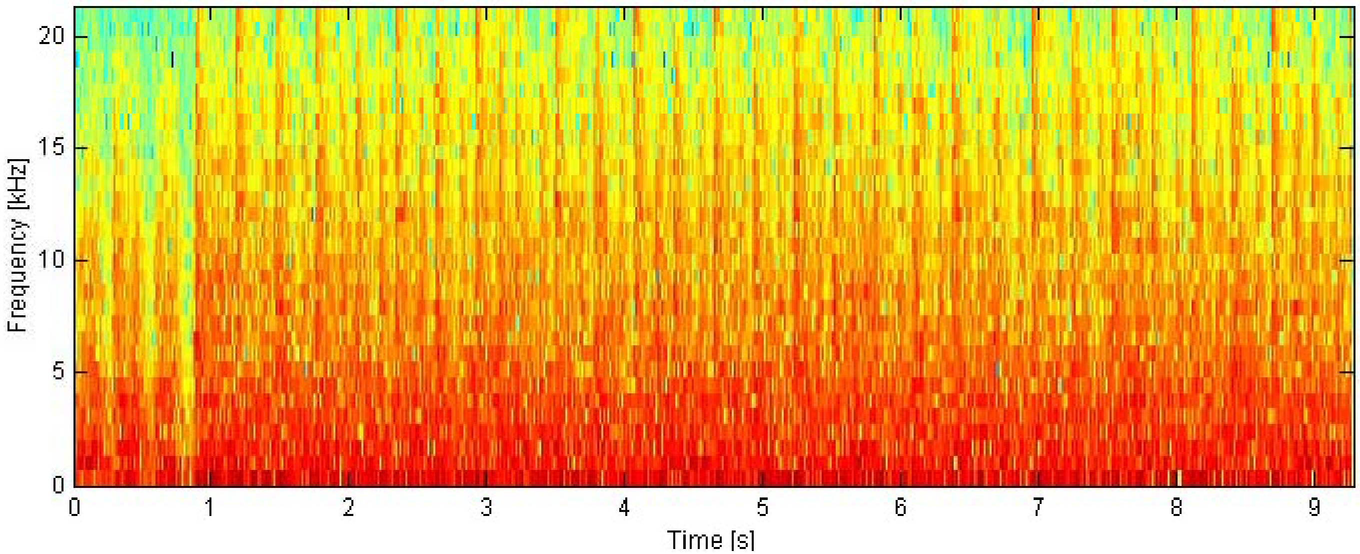
- In a perfect filterbank the first part is the only part.
- The second part consists of the aliasing terms.
- The filterbank is designed so that the aliasing is small.

Tubthumper, a time domain view



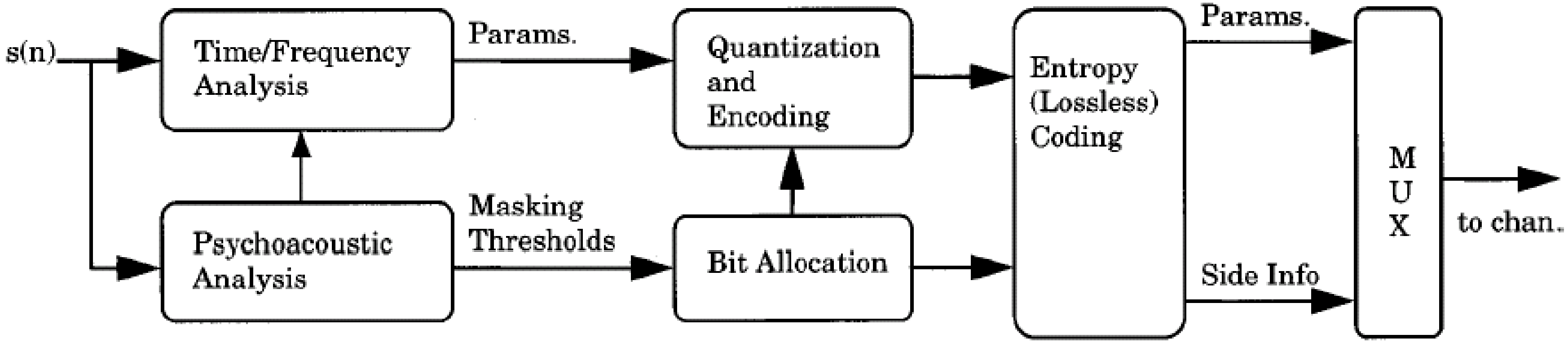
The red line is the reconstruction error after splitting the signal in subbands, down sampling and applying the synthesis filterbank. The reconstruction error is –84 dB and sounds like

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 ?

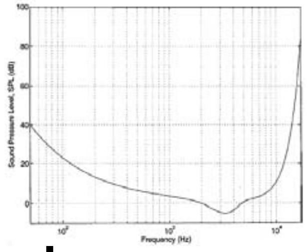


Kap 8: Diskret Fourier Transform; Estimering av effektspektrum

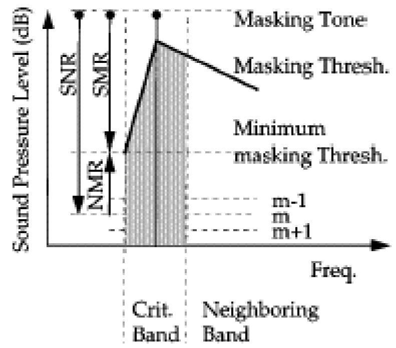
Masking

We do not hear all sounds.

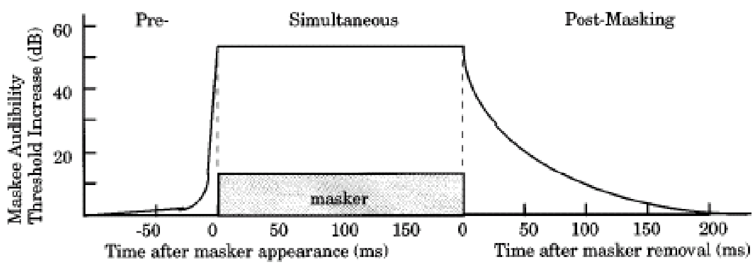
1. Absolute threshold of hearing: →



2. Masking: One sound is inaudible in the presence of another sound.
   1. Simultaneous masking
      - Noise Masking Tone
      - Tone Masking Noise
      - Noise Masking Noise

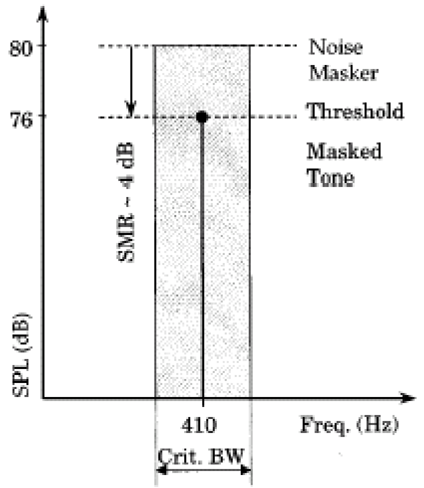


   2. Nonsimultaneous masking
      - Pre masking (2 ms)
      - Post masking (100 ms)



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

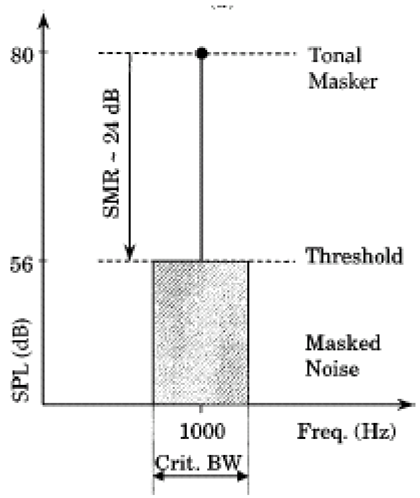


You can not hear a sinusoid that lies in the same critical band as a filtered noise if the sound pressure level is below a certain threshold.

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

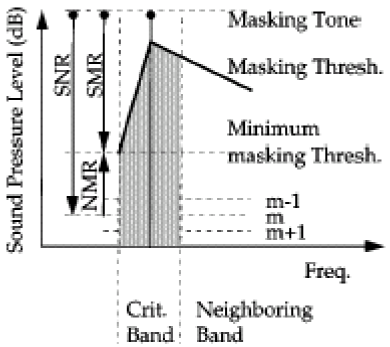


You can not hear a filtered noise that lies in the same critical band as a sinusoid if the sound pressure level is below a certain threshold.

This effect also stretches out beyond the critical band.

Exploit Masking

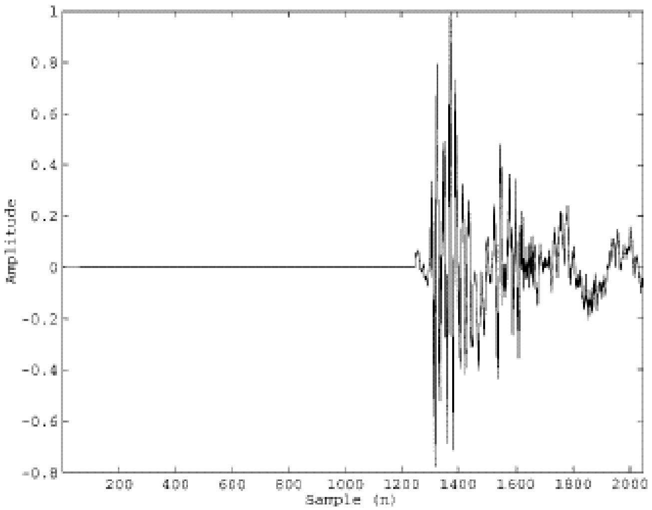
- If a sound is masked we can't hear it.

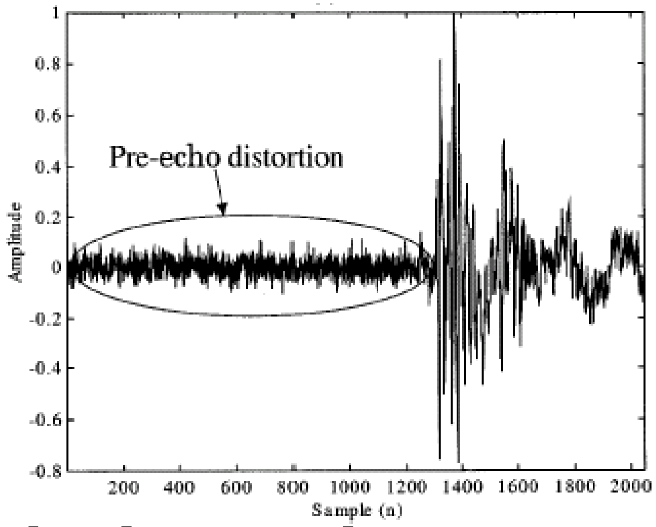


- Make a frequency analysis of the signal and find the masking threshold.
- Put the quantization noise under the masking threshold and we won't hear the quantization.

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

Pre echo distortion





- The original sound of a castanet.
- The abruptness in time domain results 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: l.1 og l.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

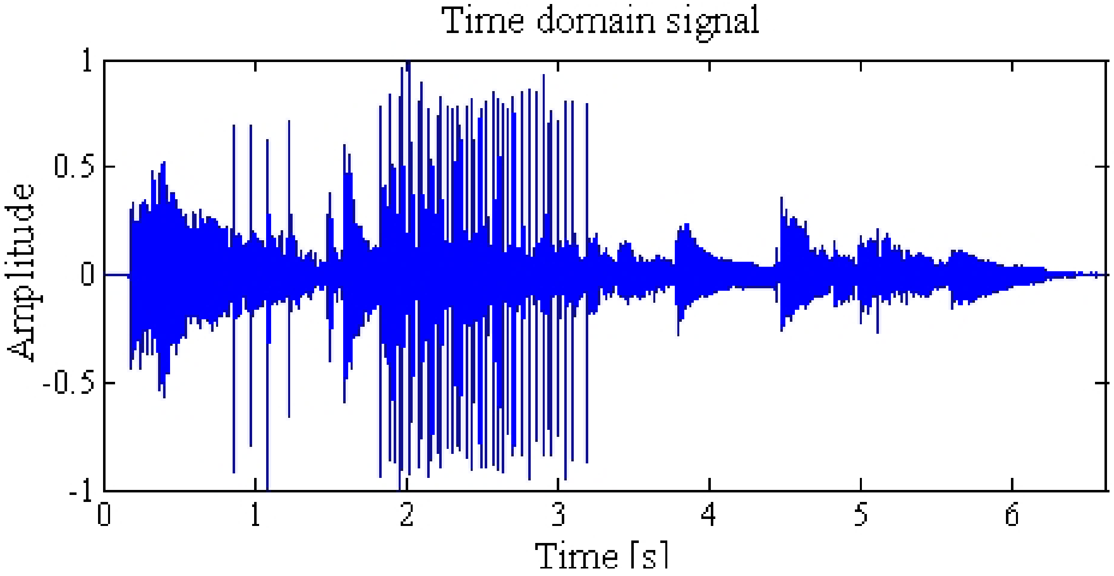
Scale factors and Quantization

- 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.

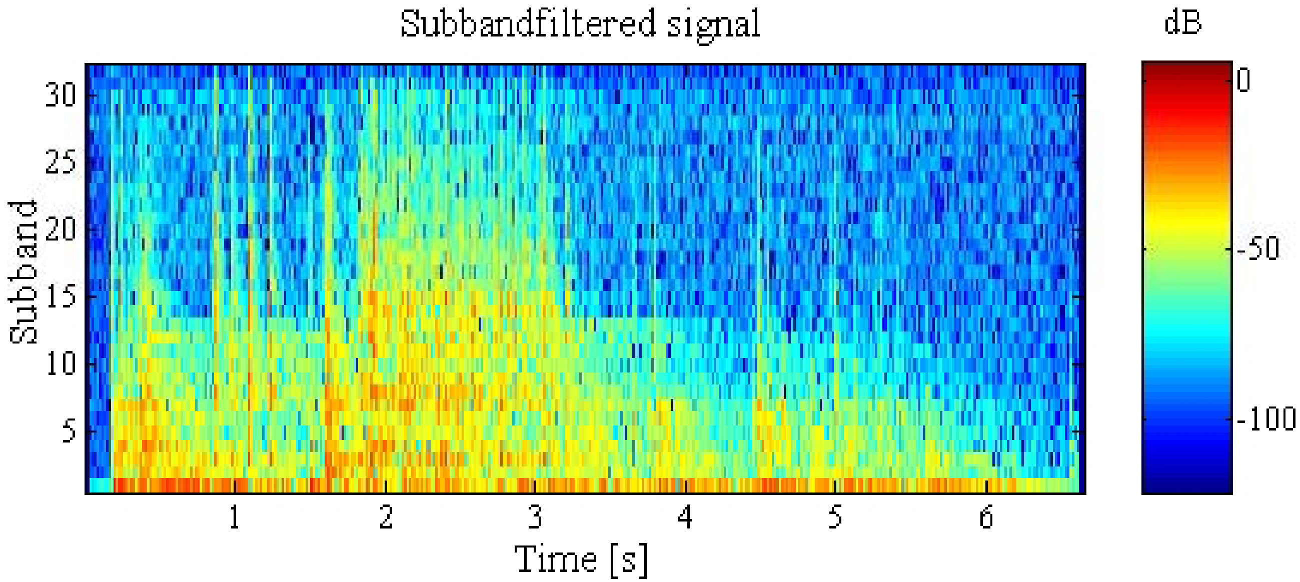
Bit Allocation and Masking

- 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

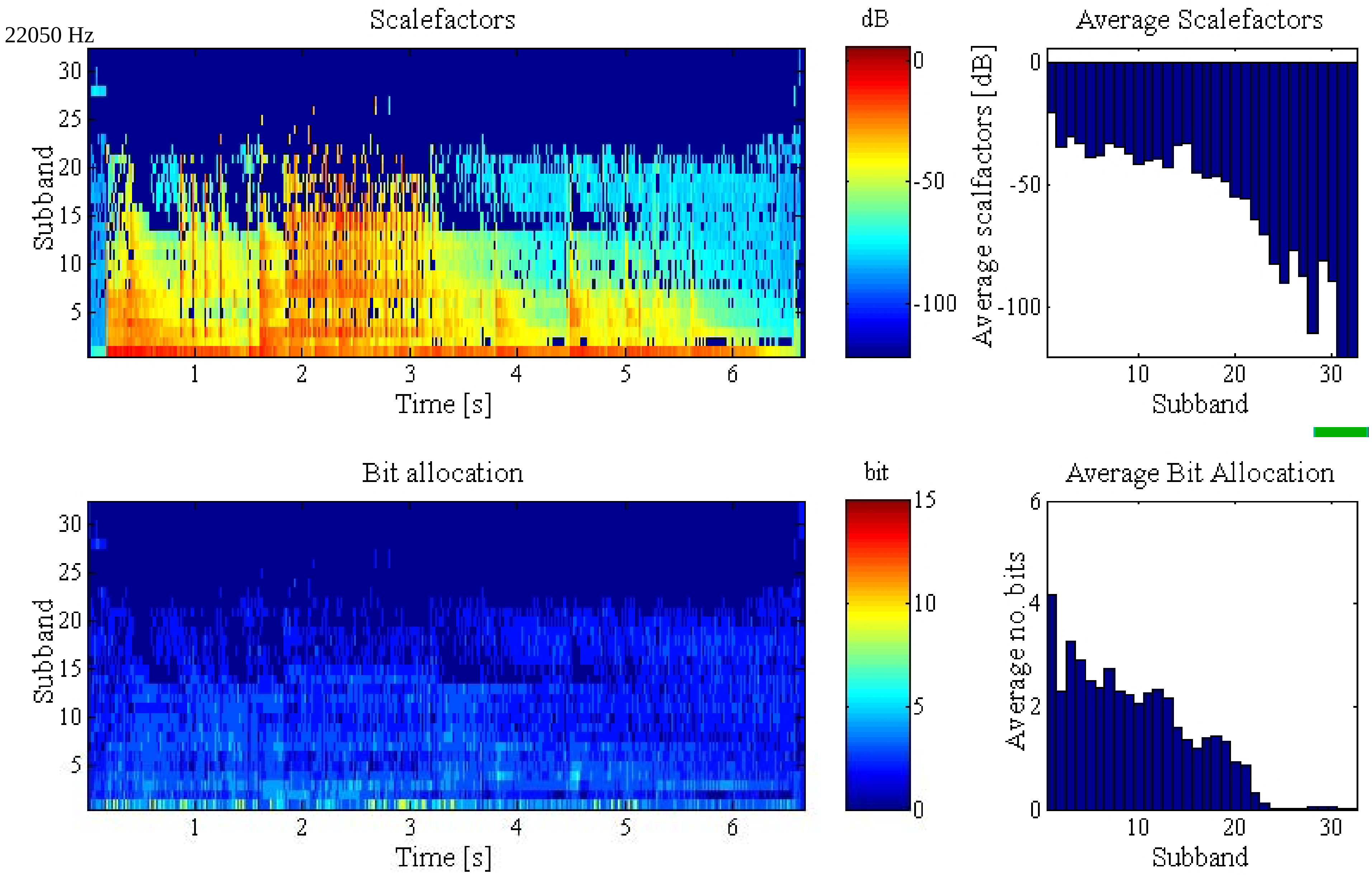


22050 Hz



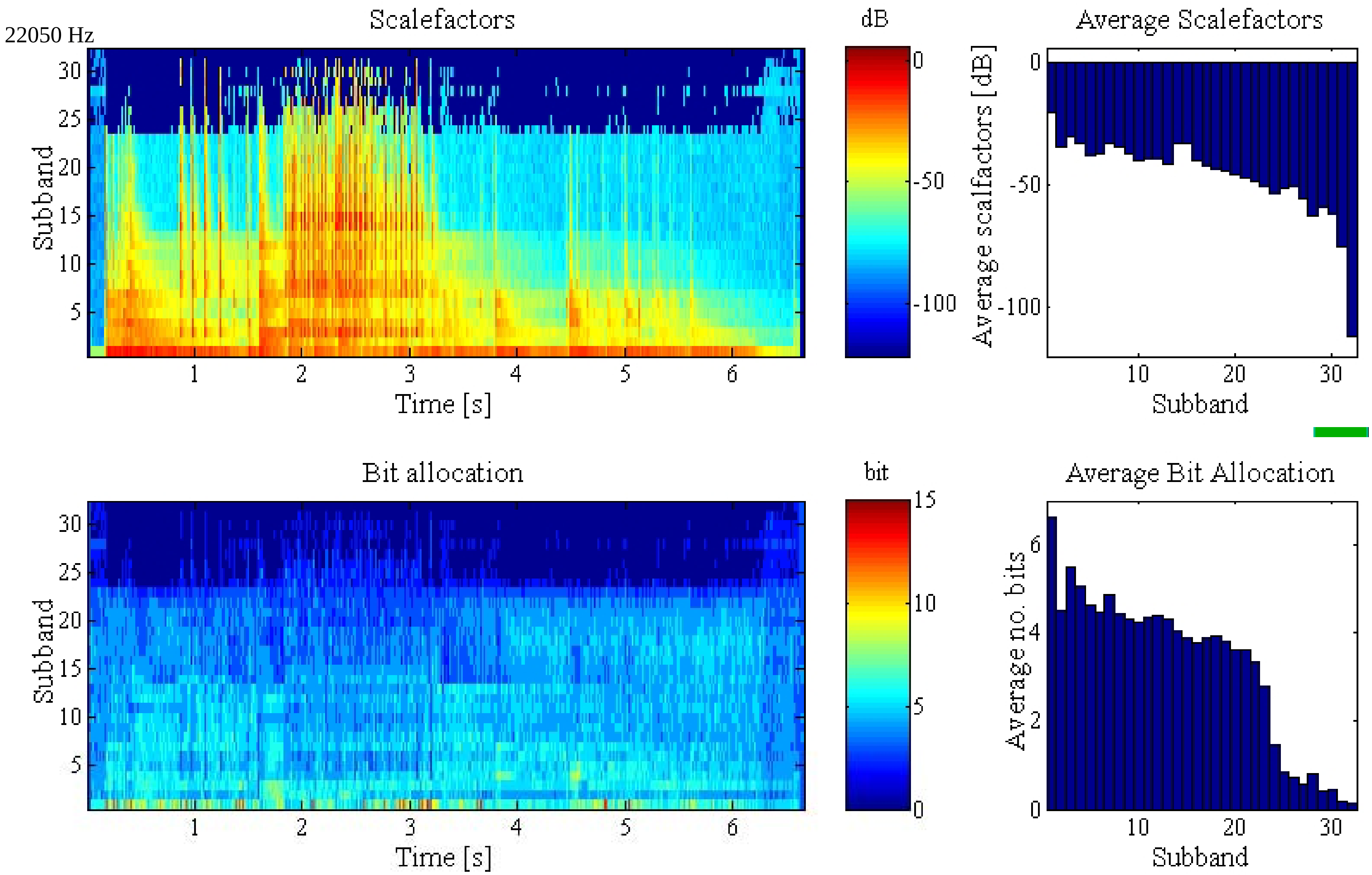
UNIVERSITETET
I OSLO

Bit allocation with 2 bits per sample

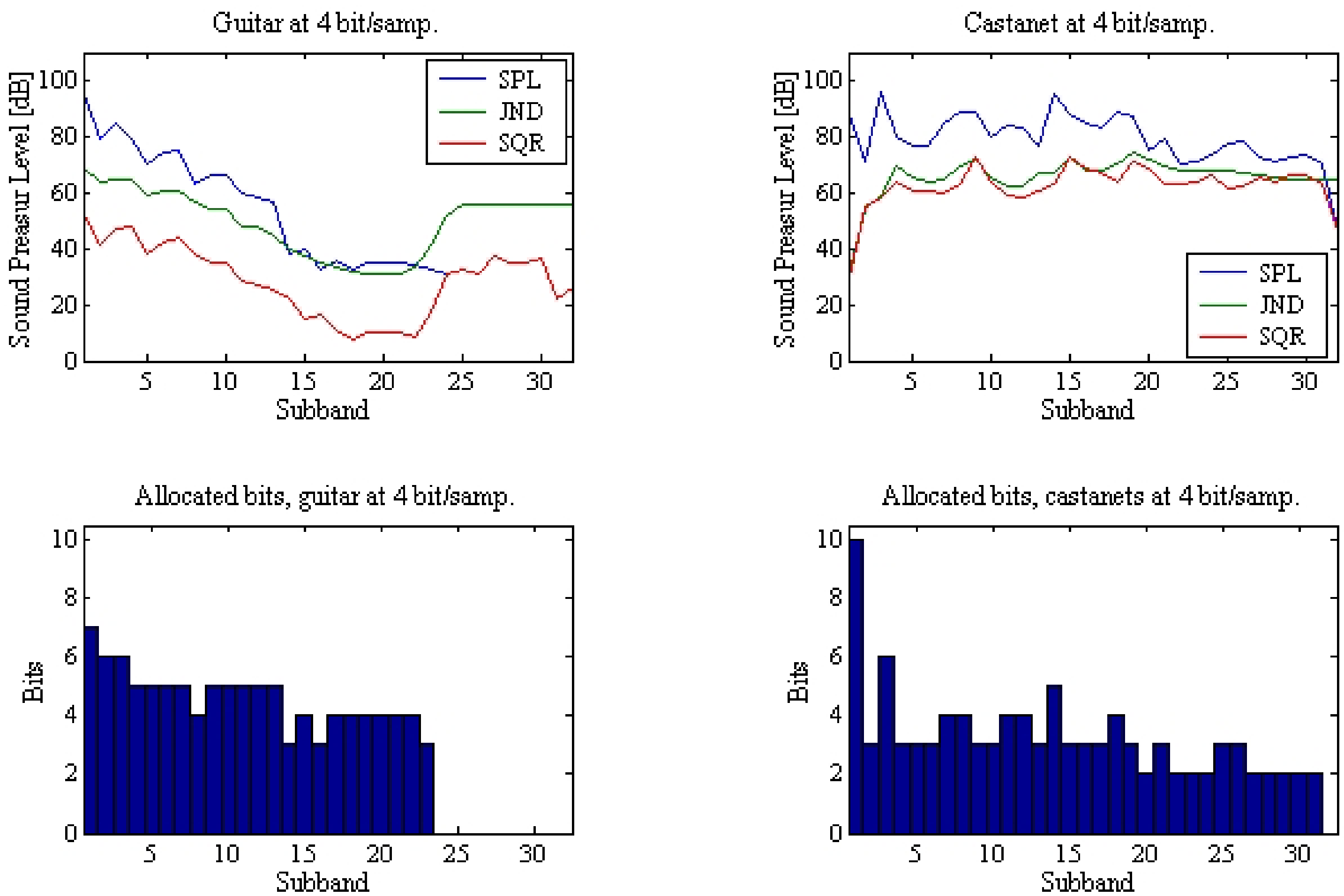


UNIVERSITETET I OSLO

Bit allocation with 4 bits per sample



Signal to Quantization Noise Ratio and the Just Noticeable Distortion



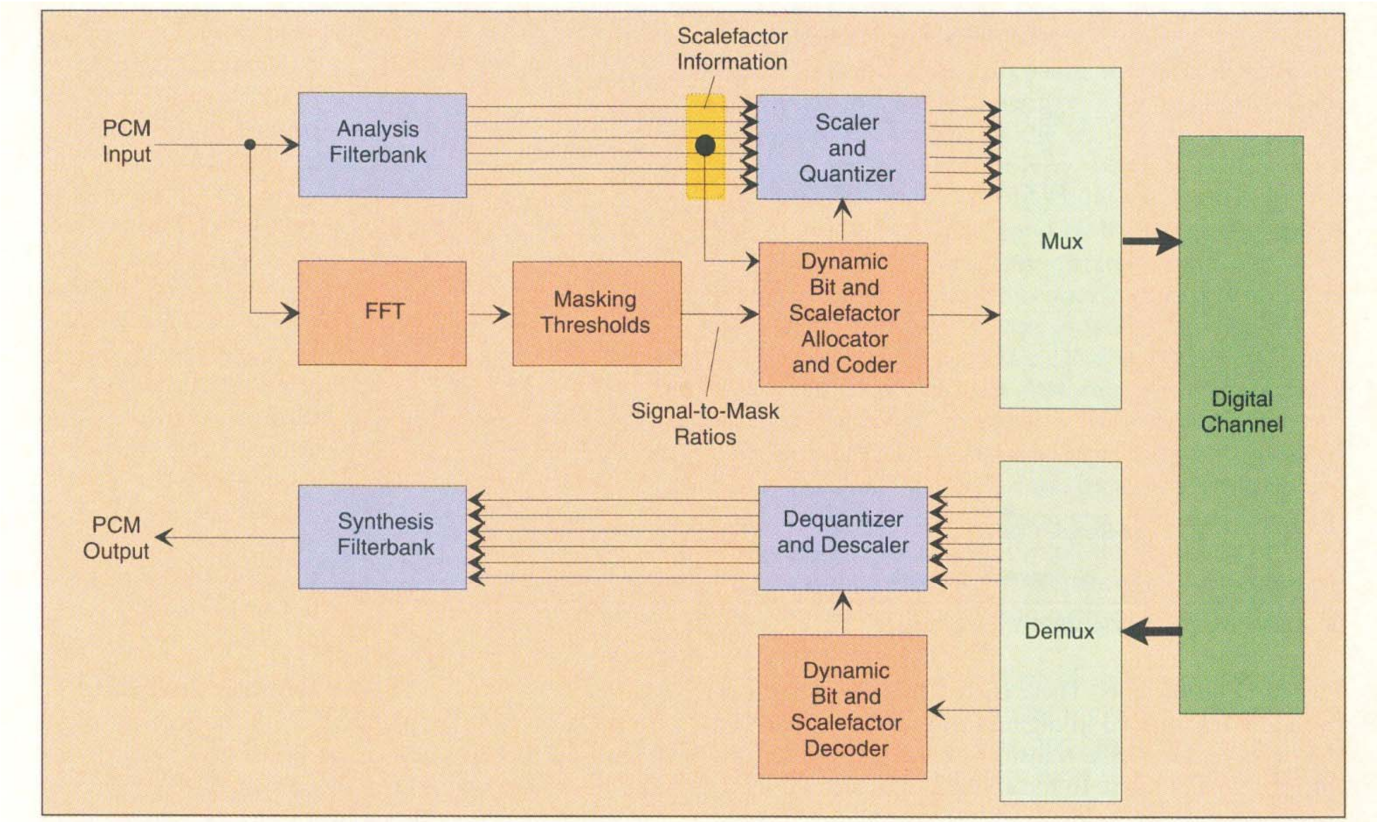
Frame at t=0.6 s

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

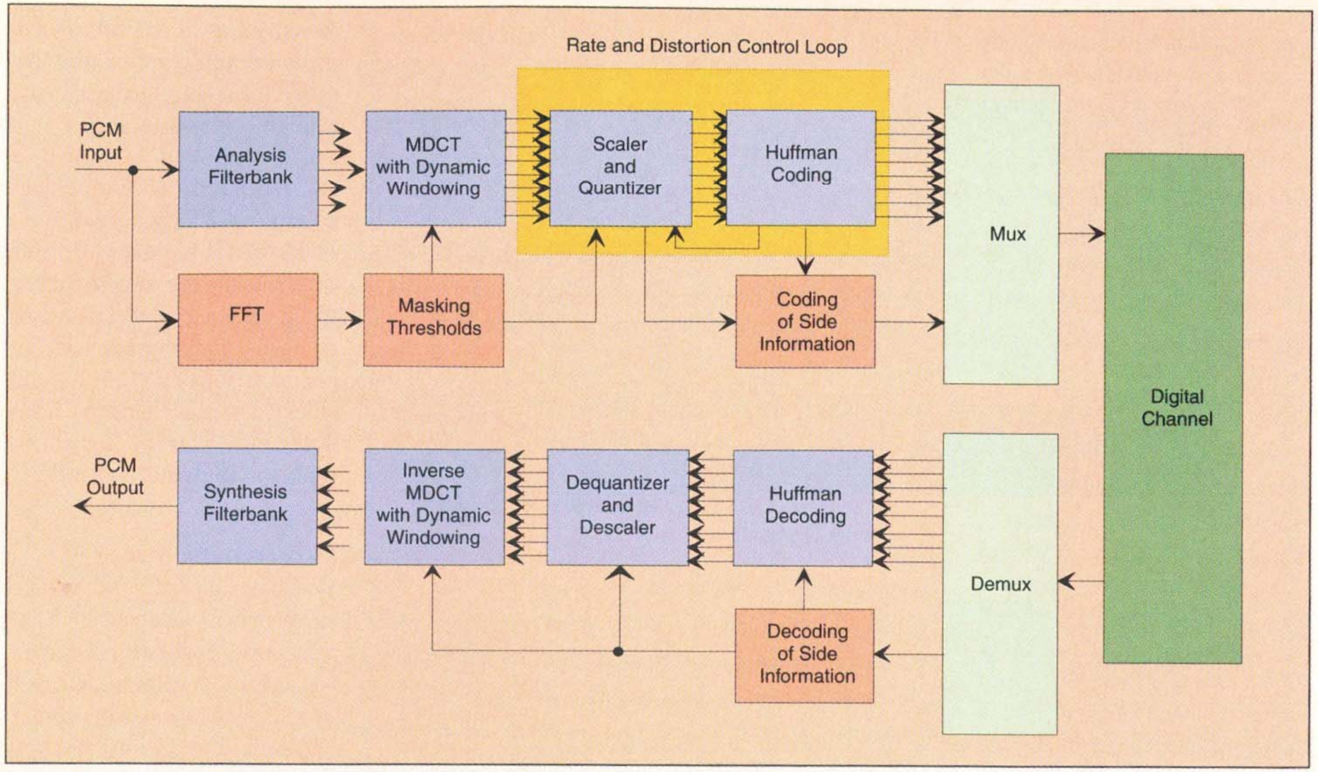
MPEG-1 layers I and II



▲ *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: kaskadekopler en 6 eller 18 punkts (dynamisk vindus-svitsjing) MDCT med lag Is filterbank
  => 32*18 = 576 frekvensbånd =>
  - 64 kbit/s pr kanal (variabel) (128 kbit/s for stereo)
- MPEG-1, layer III = MP3

Kapitler i læreboka – MP3-koding

1	Overview	• Praktiske eksempler som mp1/mp3
2	Discrete Signals	• Musikk og talesignal
3	Time-Domain Analysis	• 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	• Analyse av filter i filterbanken: nøkkel til å få til filterdesign
5	Frequency Domain Analysis	• Frekvensdomene er sentralt i modell av hørsel • Frekvensselektive filtre: båndpassfiltre • Inverse systemer: kan dele i bånd i koder og addere sammen igjen i dekoder
6	Filter Concepts	• Filterstrukturer, hvordan implementere filterbank i koder og dekoder
7	Digital Processing of Analog Signals	• A/D-analyse: kvantiseringsstøy ved direkte sampling • Mutirate 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.	• 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.	• Hvordan finne koeffisienter til bp-filtrene i filterbanken?
11	MATLAB Examples	
A	Useful Concepts from Analog Theory	