



UNIVERSITETET  
I OSLO

# MPEG-1 lag 1, 2 og lag 3

Sverre Holm





# MPEG audiokoding

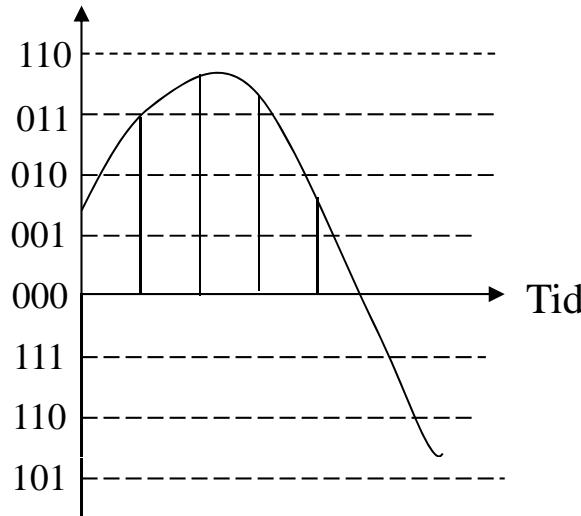
- Motivasjon for de fleste kapitlene i Ambardar, Digital signal processing: A Modern Introduction, Thomson, 2007.





# Digital representation of Sounds

## Pulse Coded Modulation (PCM)

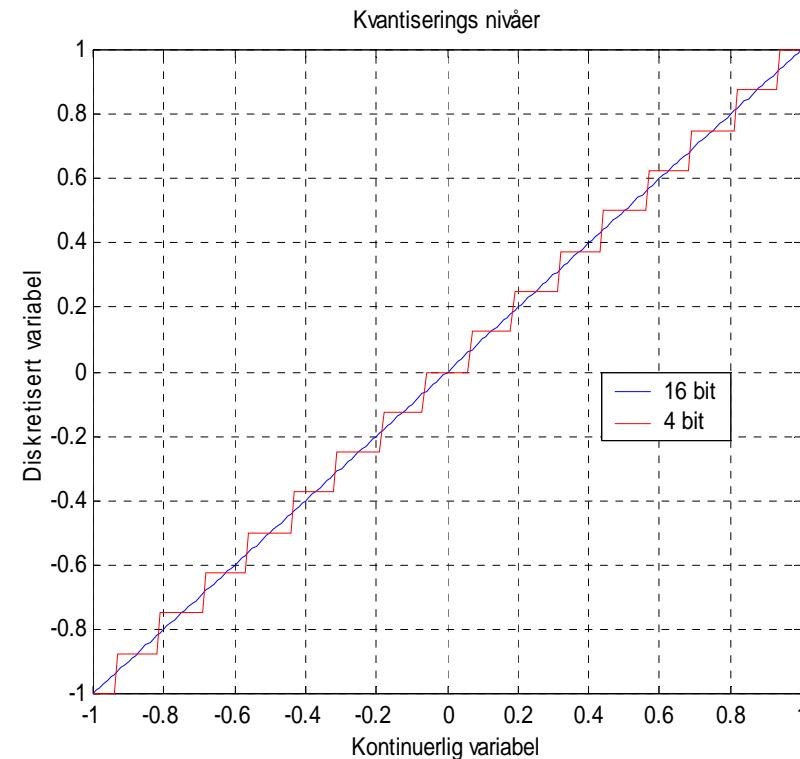


16 bit kvantisering gir

$$2 \cdot 2 \cdot 2 \cdot \dots \cdot 2 = 2^{16} = 65\,536 \text{ nivåer}$$

Ved 44100 samples per sek, blir bitraten:

$$16 \cdot 44100 = 705\,600 \text{ bits/s} = \text{halv CD-rate}$$



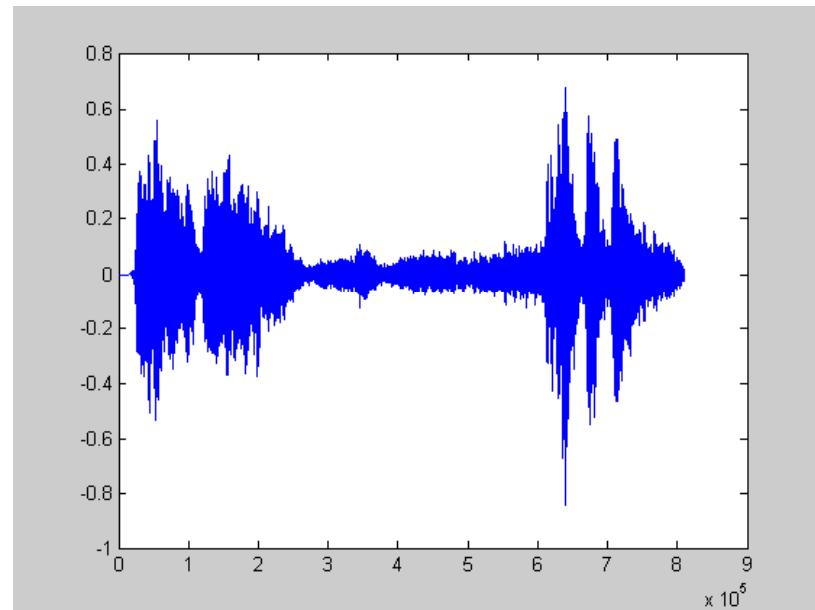
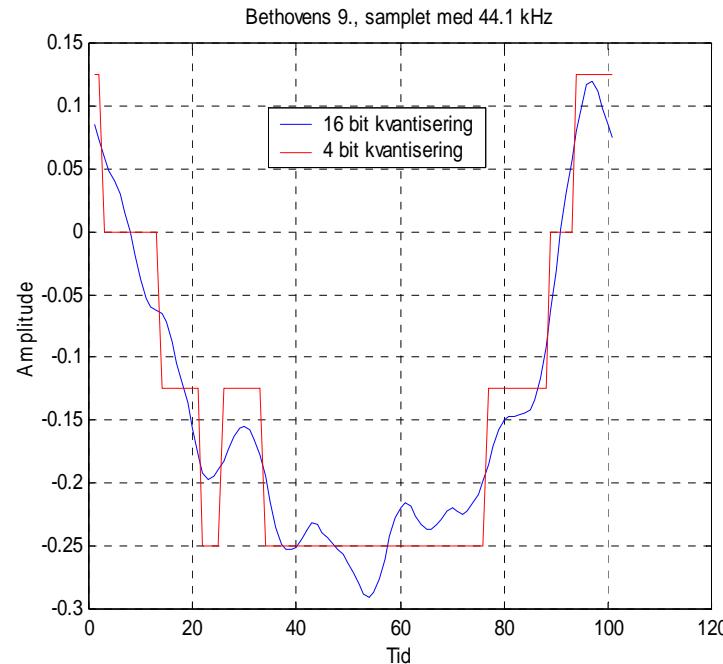


# Beethovens 5. symfoni

16 bit kvantisering  
 $2^{16} = 65536$  nivåer



4 bits kvantisering  
 $2^4 = 16$  nivåer





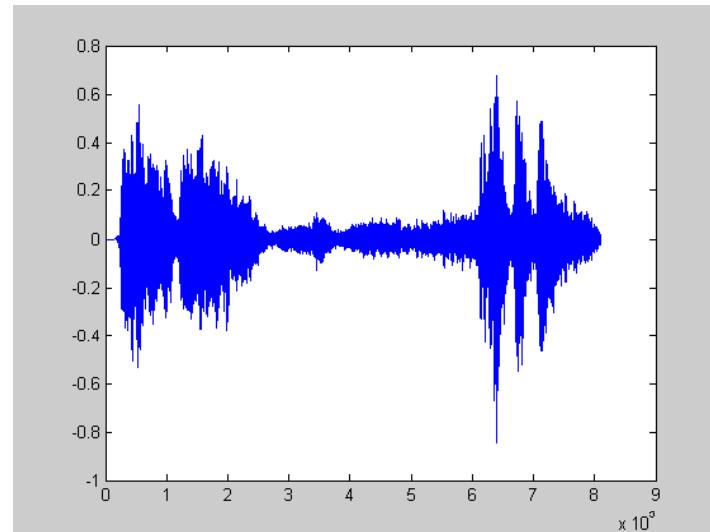
# Hvorfor høres det så ille ut?

- Problem: Bare noen få kvantiseringsnivåer => stor avrundingsfeil. Ofte at signalet settes til null da nivået var lavere enn laveste kvantiseringssnivå
- Kvantisering og sampling:  
**Kap 7: Digital behandling av analoge signaler**
- Mulig løsning: Skaler blokker av data slik at maximumsverdien alltid utnytter hele dynamikkområdet til kvantisereren
- Kostnad: Må sende over skalafaktorer
  - NICAM, Near Instantaneous Companded Audio Multiplex: format for digital lyd over analog TV.
  - Blokklengde 32 samples, 3 bit pr blokk sideinfo. Stereo kodes med 10 av 14 bit ved samplingsrate 32 kHz => 728 kbit/s.
  - Variant av adaptiv differensiell puls kode modulasjon



# Stasjonæritet - tidsinvarians

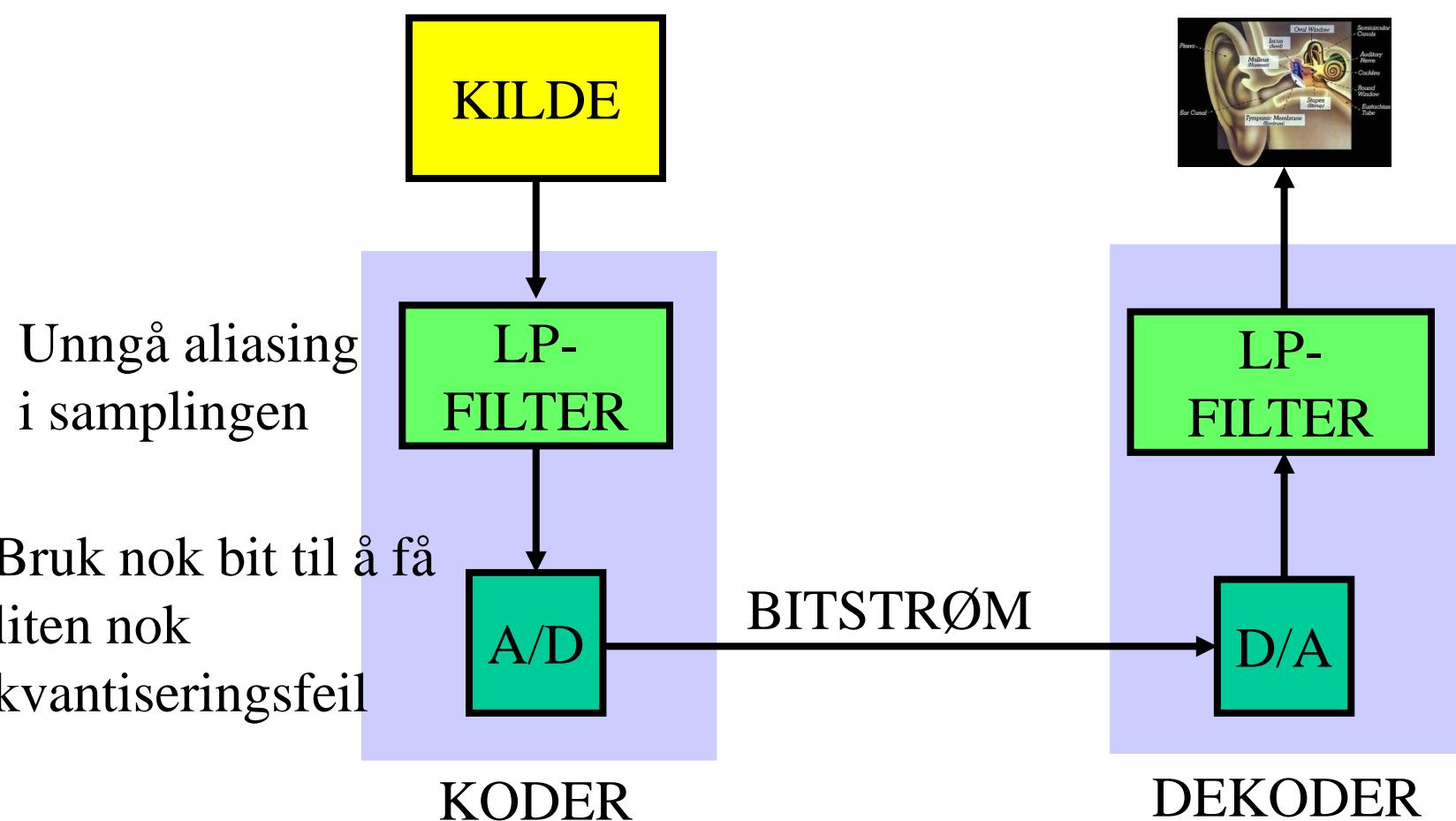
- = Egenskaper varierer ikke med tiden, **kap 3: Tids-domene analyse**
- Forutsettes i det meste av analyser
- Tale er korttids stasjonær, dvs bare over ca 20 ms,
  - Endres  $1/20e-3=50$  ganger pr sekund



Ca 18 sek =  $900 * 20$  ms



# Direktesampling (PCM)





## Bitrater

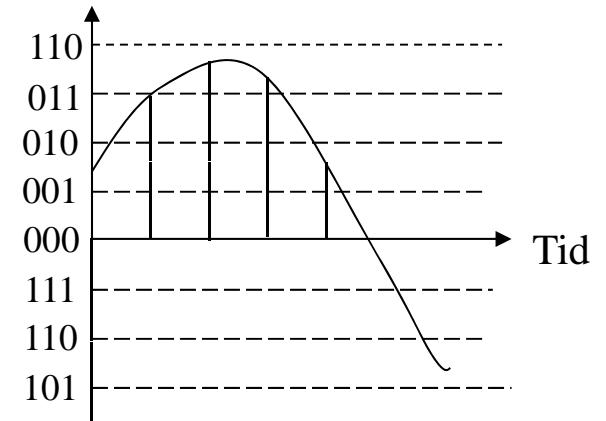
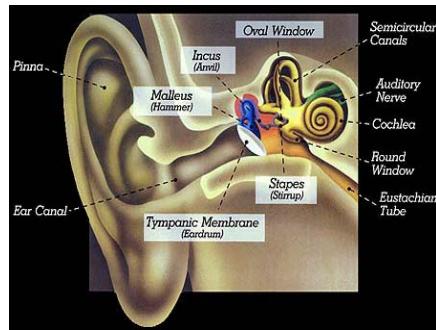
- CD:  $44.1 * 2 * 16 = 1.411 \text{ Mbit/s}$ 
  - 4 bit: 25% => 350 kbit/s låter forferdelig
- MP3, AAC etc: 128 kbit/s ~ CD/12
- Hva er det lure trikset?



# MPEG-1 Audio

## Psychoacoustics in sound compression

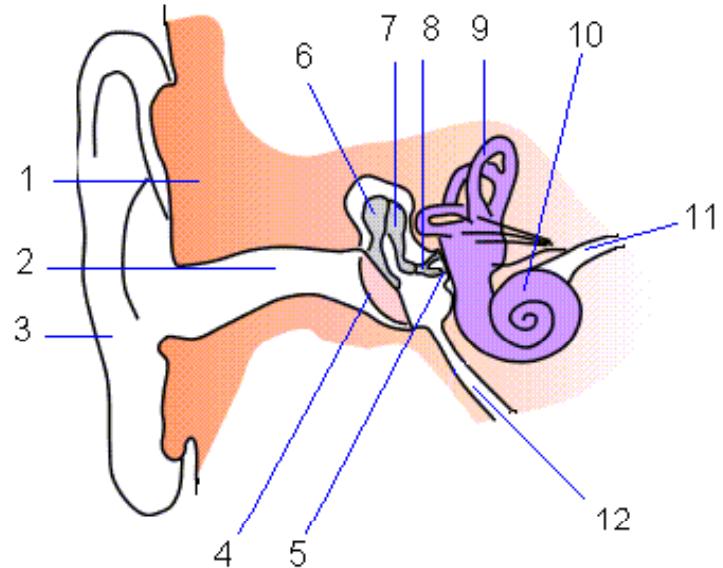
- How do we hear?
- Digital representation of sounds
- Sound compression
- Psychoacoustics
  - Masking
  - Adaptive quantization
  - Bit allocation
- Filterbanks





# Øret

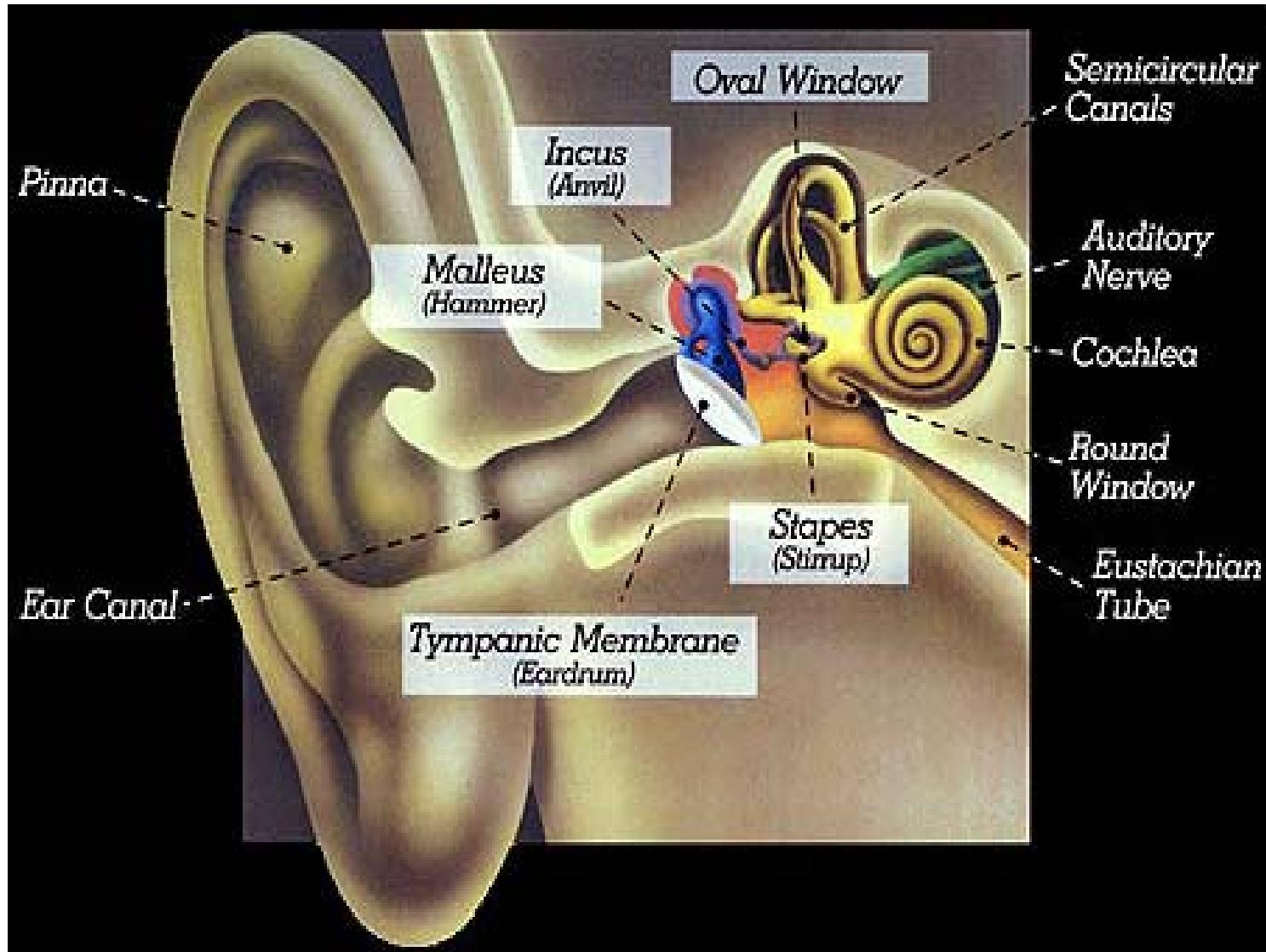
1. Tinning
2. Øregang
3. Ytre øre
4. Trommehinne
5. Ovale vindu
6. Hammeren
7. Ambolt
8. Stigbøylen
9. Bueganger
10. Sneglehuset
11. Hørselnerve
12. Øretrompeten



Wikimedia Commons



# The Ear

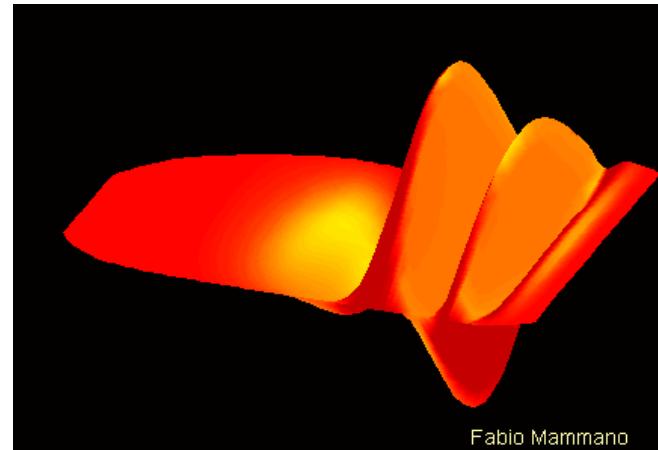
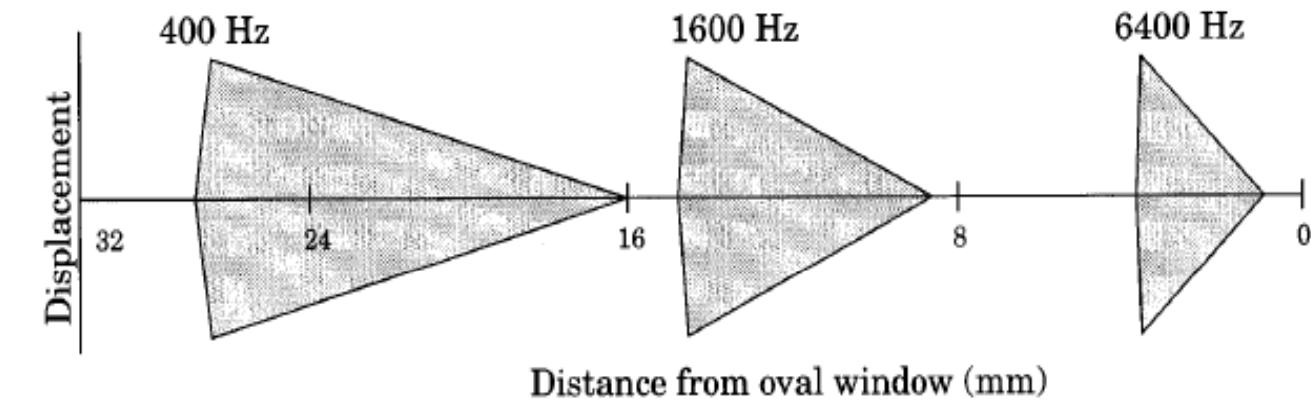




UNIVERSITETET  
I OSLO

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

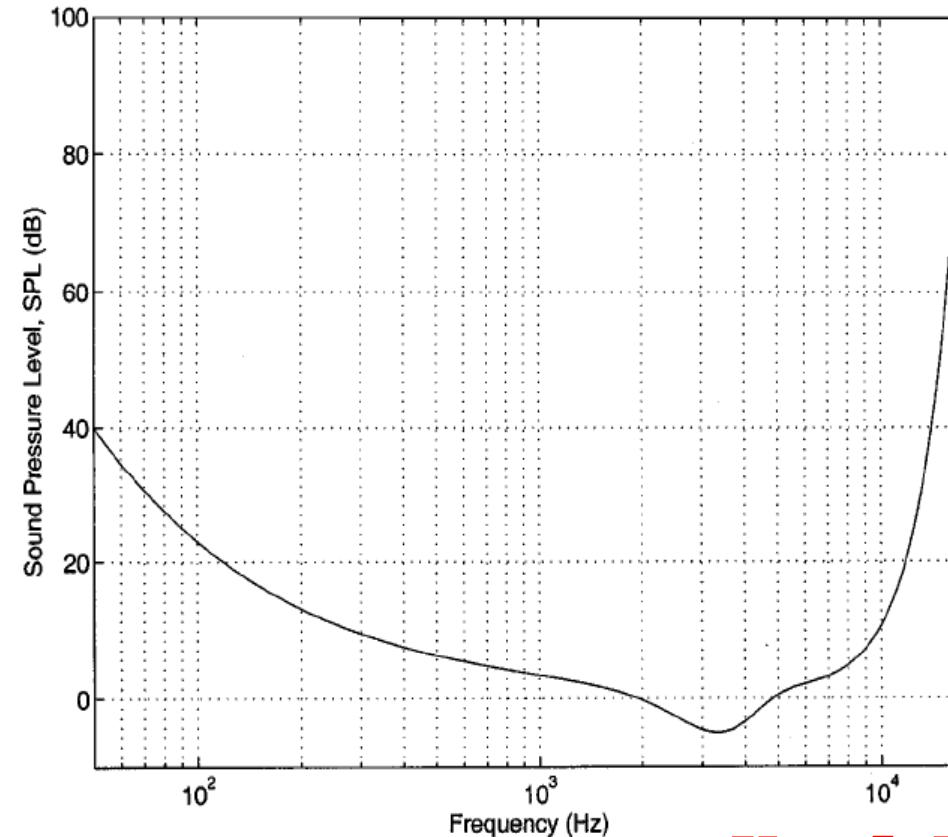
Unwound  
cochlea



**Kap 5: Frekvensanalyse**



# Threshold for audible sounds



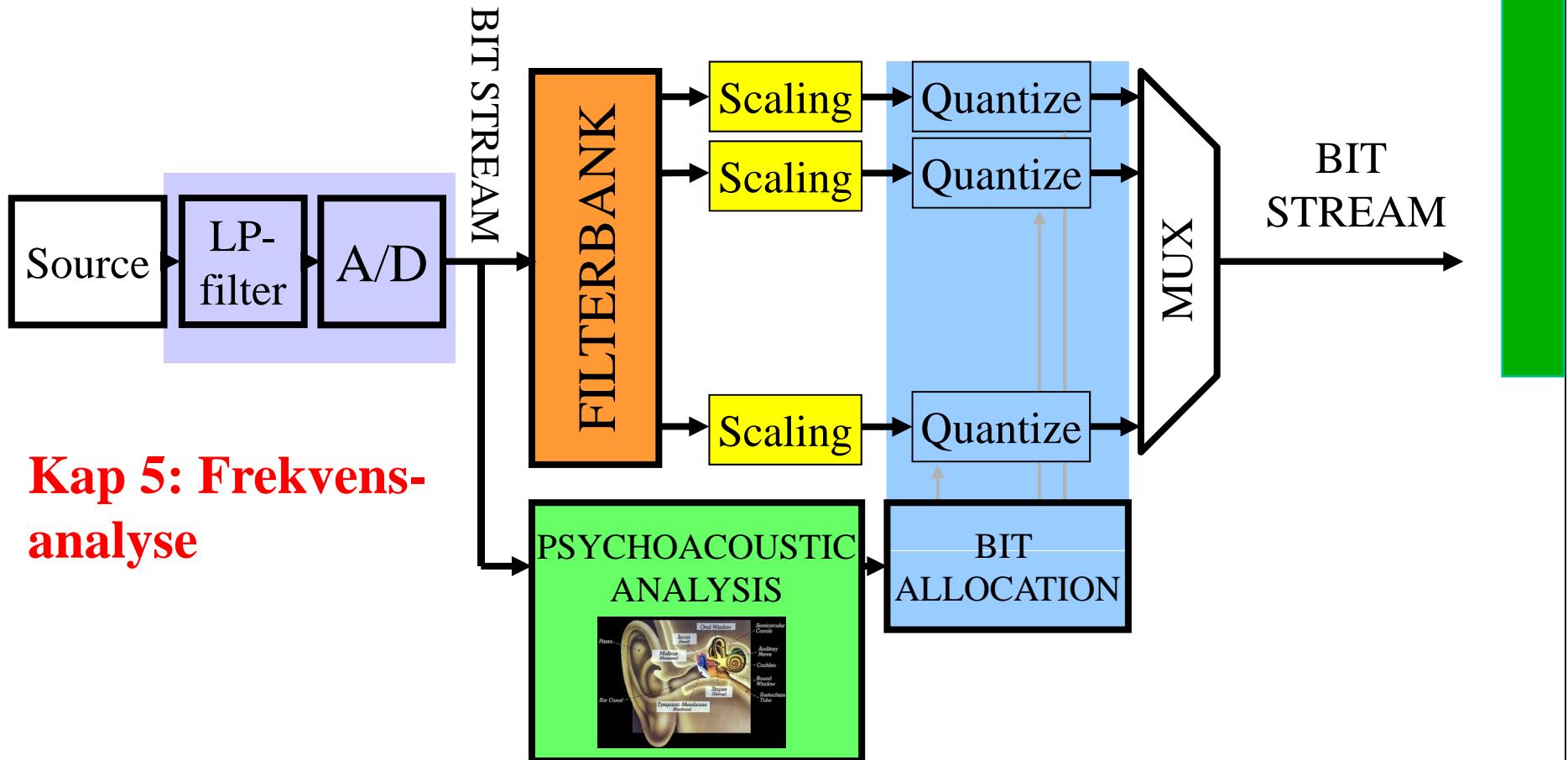
**Reference 0 dB:**  
 $20 \mu\text{Pa} = 2 \cdot 10^{-5} \text{ N/m}^2$

**Kap 5: Frekvensanalyse**



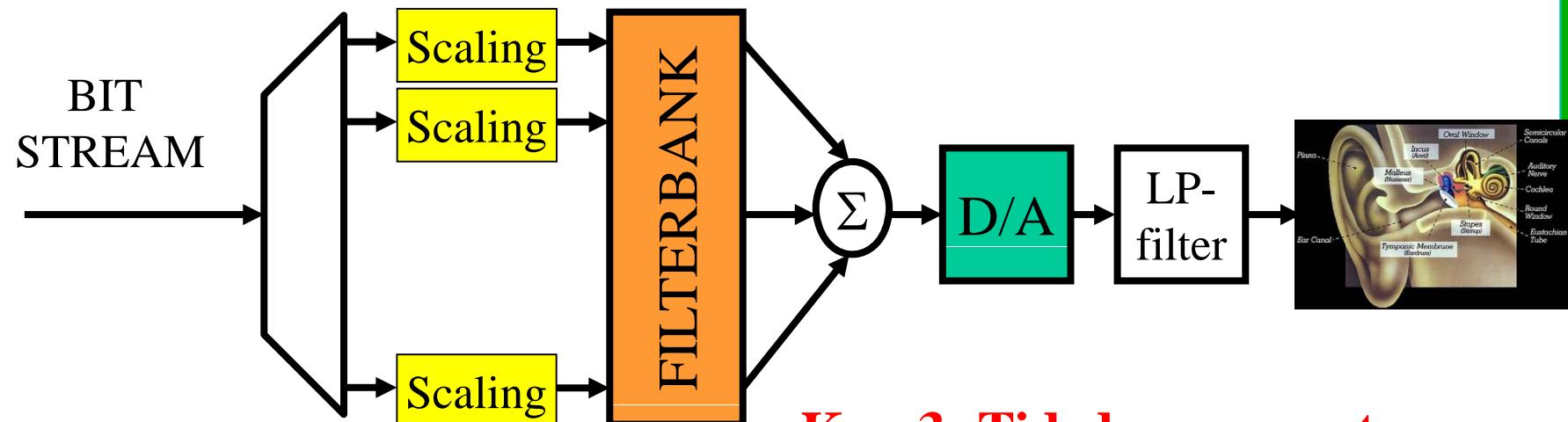
# Filterbank Approach

## Encoding





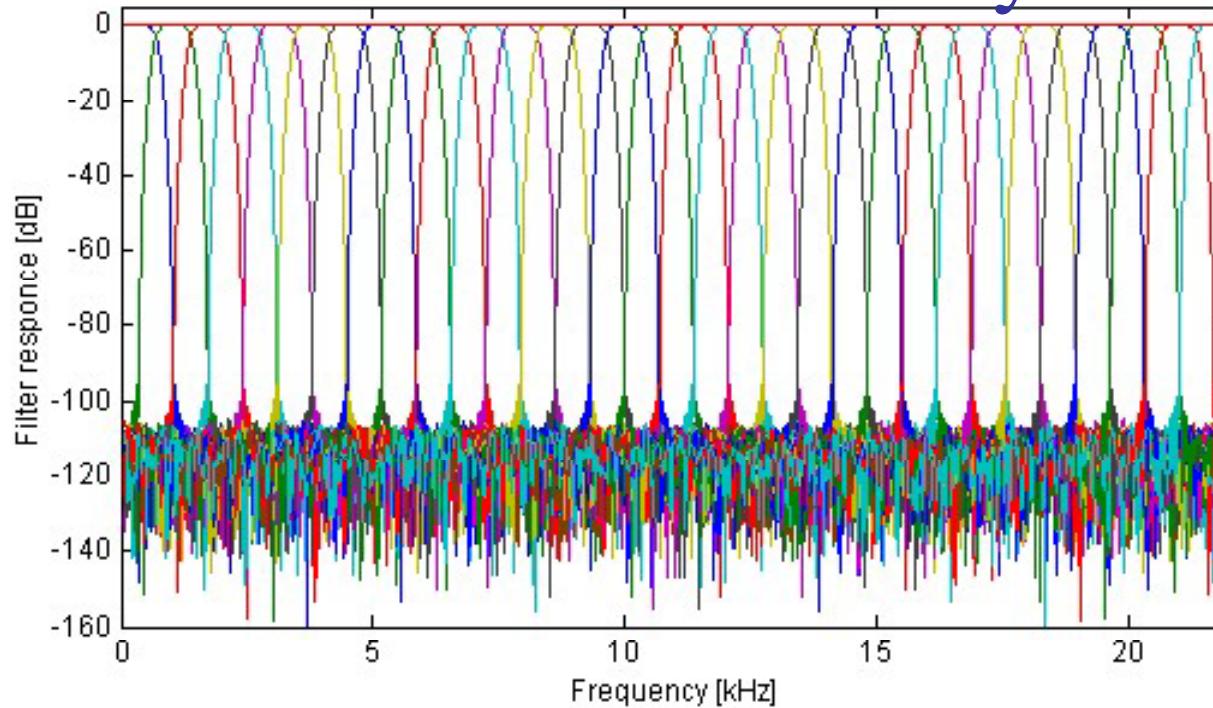
# Decoding is much simpler



**Kap 3: Tidsdomene systemer:  
linearitet**  
**Kap 3: Inverse systemer**



# Filterbanks in MPEG-1 audio layer 1-3



- 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**

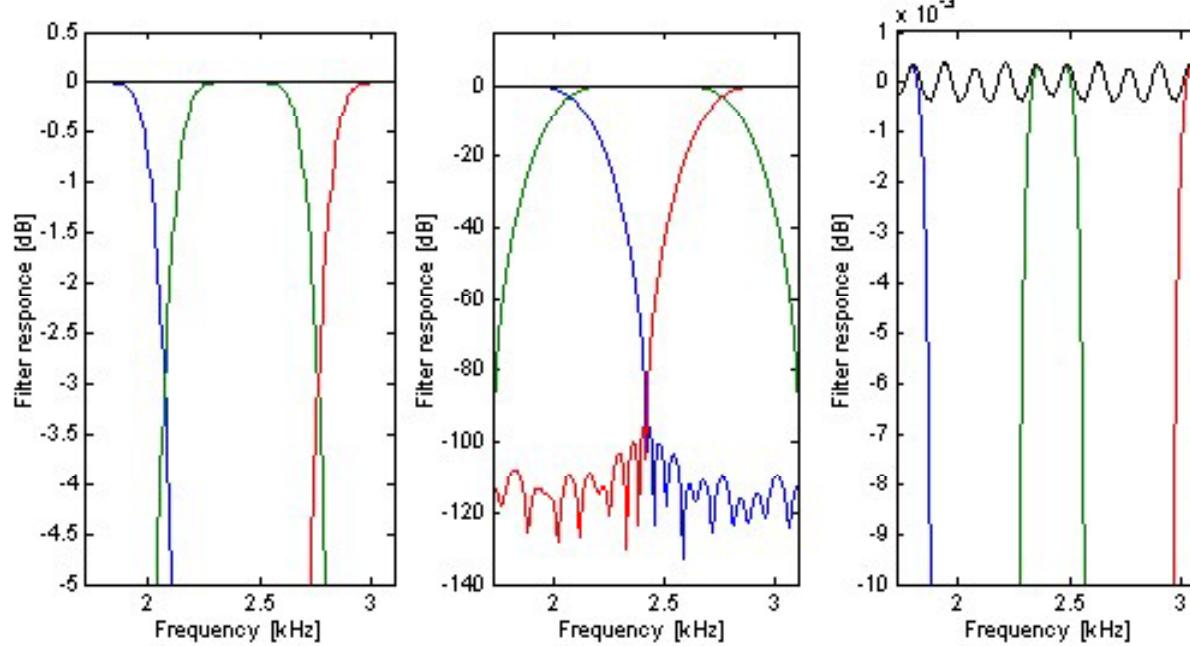
**Kap 5: Frekvens-analyse av systemer**

**Kap 6: Digitale filtre**

**Kap 10: FIR Filterdesign**



# 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**



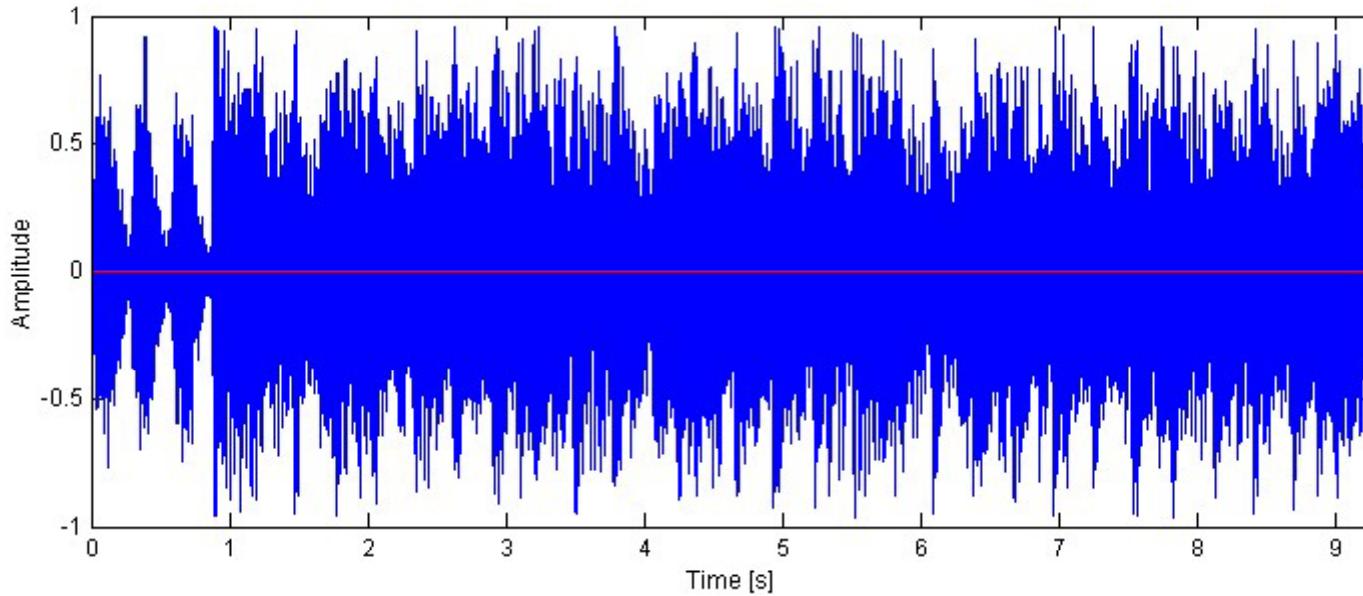
# Reconstruction Using Nonideal Filterbanks

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

- 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

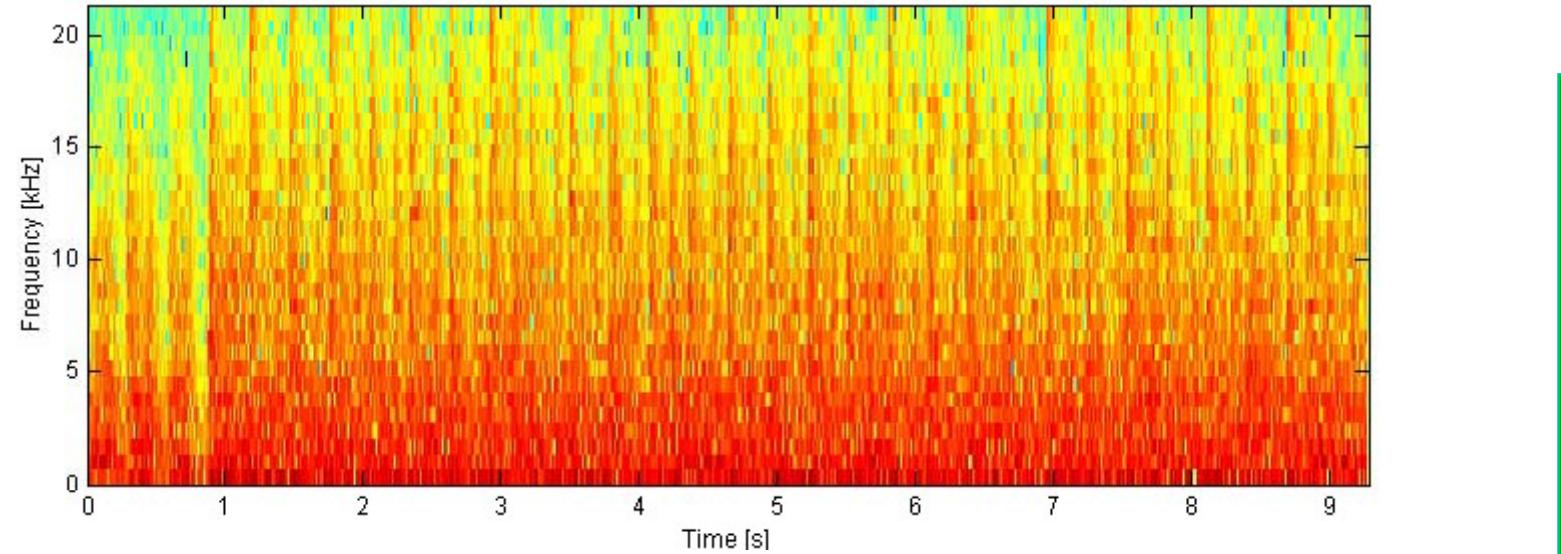


The red line is the reconstruction error after splitting the signal in subbands, down sampling and applying the synthesis filterbank. The reconstruction error is  $-84$  dB and sounds like





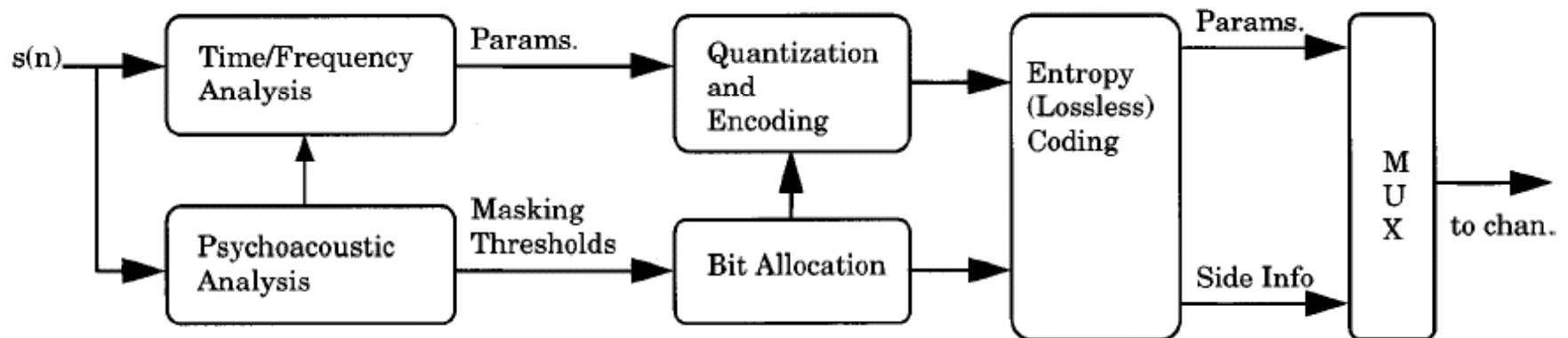
# Tubthumper, frequency view



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.

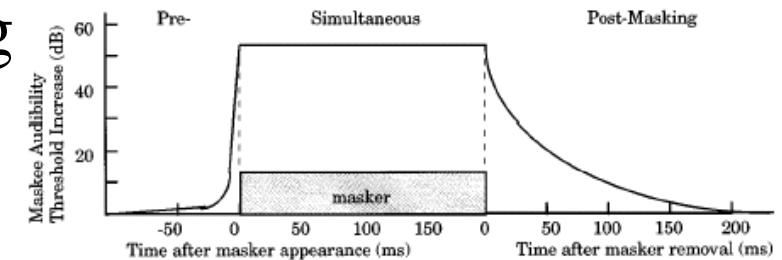
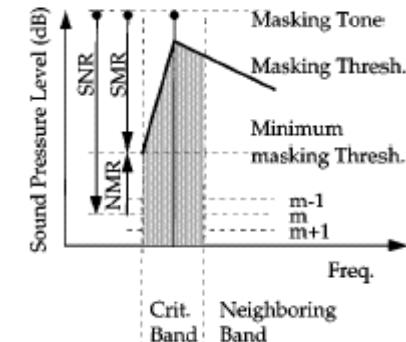
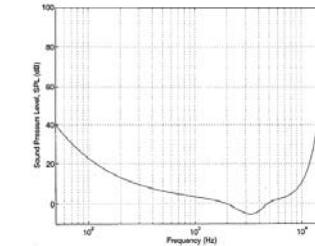
1. Absolute threshold of hearing: →
2. Masking: One sound is inaudible in the presence of another sound.

## 1. Simultaneous masking

- Noise Masking Tone
- Tone Masking Noise
- Noise Masking Noise

## 2. Nonsimultaneous masking

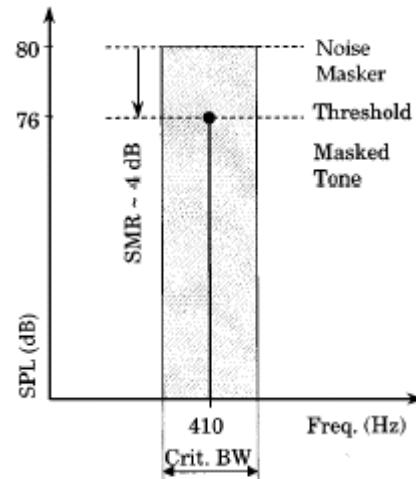
- Pre masking (2 ms)
- Post masking (100 ms)





# 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



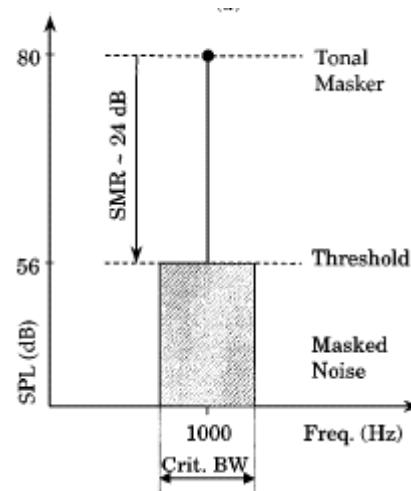
You can not hear a sinusoid that lies in the same critical band as a filtered noise if the sound pressure level is below a certain threshold.

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



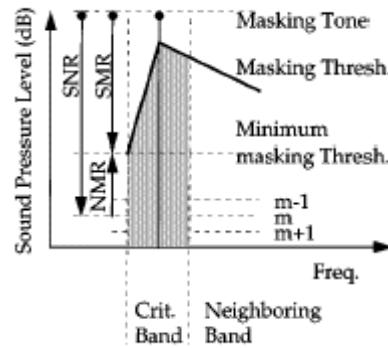
You can not hear a filtered noise that lies in the same critical band as a sinusoid if the sound pressure level is below a certain threshold.

This effect also stretches out beyond the critical band.



# Exploit Masking

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

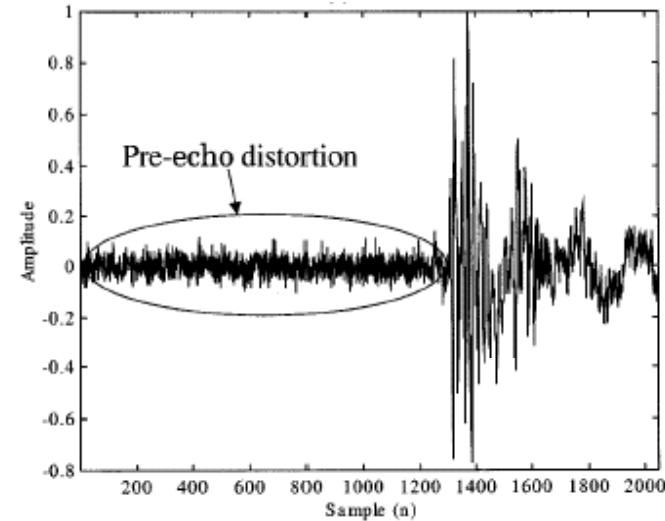
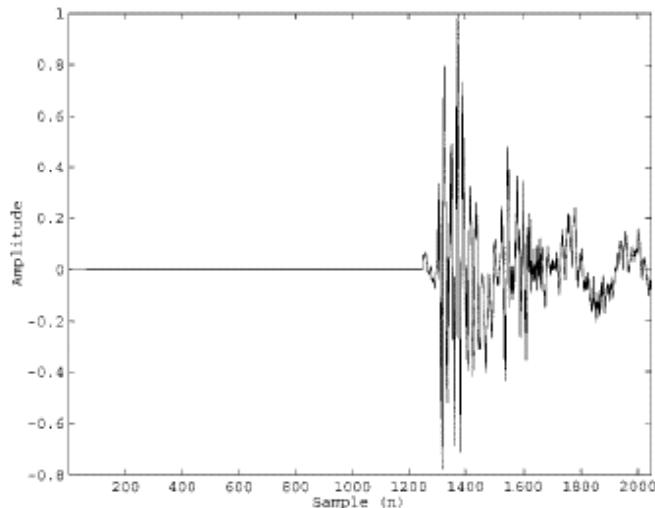


- Make a frequency analysis of the signal and find the masking threshold.
- Put the quantization noise under the masking threshold and we won't hear the quantization.

**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 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

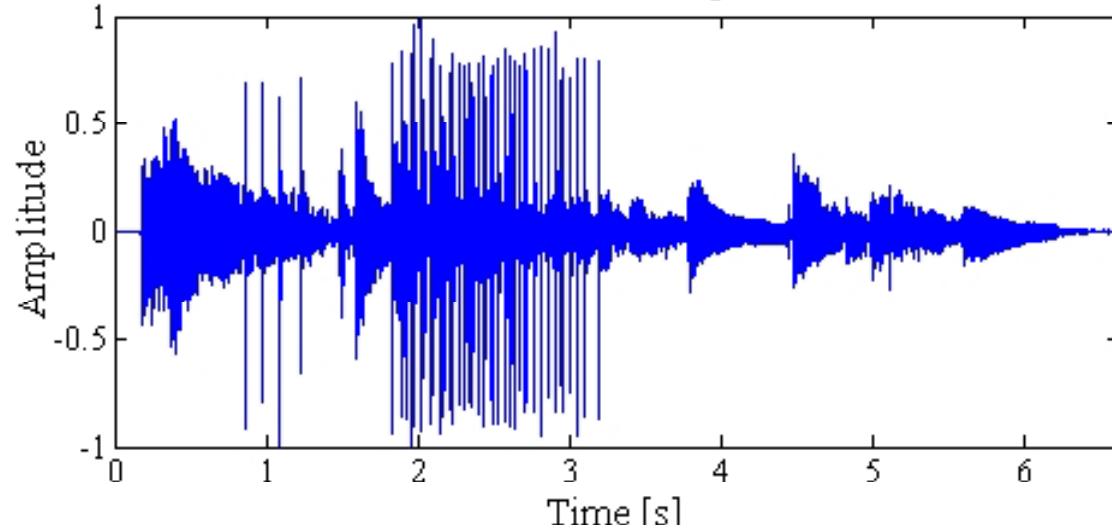
- The masking threshold in each subband gives the Just Noticeable Distortion (JND) limit for that band.
- Bits are assigned to subbands so that the quantization noise falls below or as little over the JND as possible.
- Then the Signal to Quantization noise Ratio (SQR) falls below JND



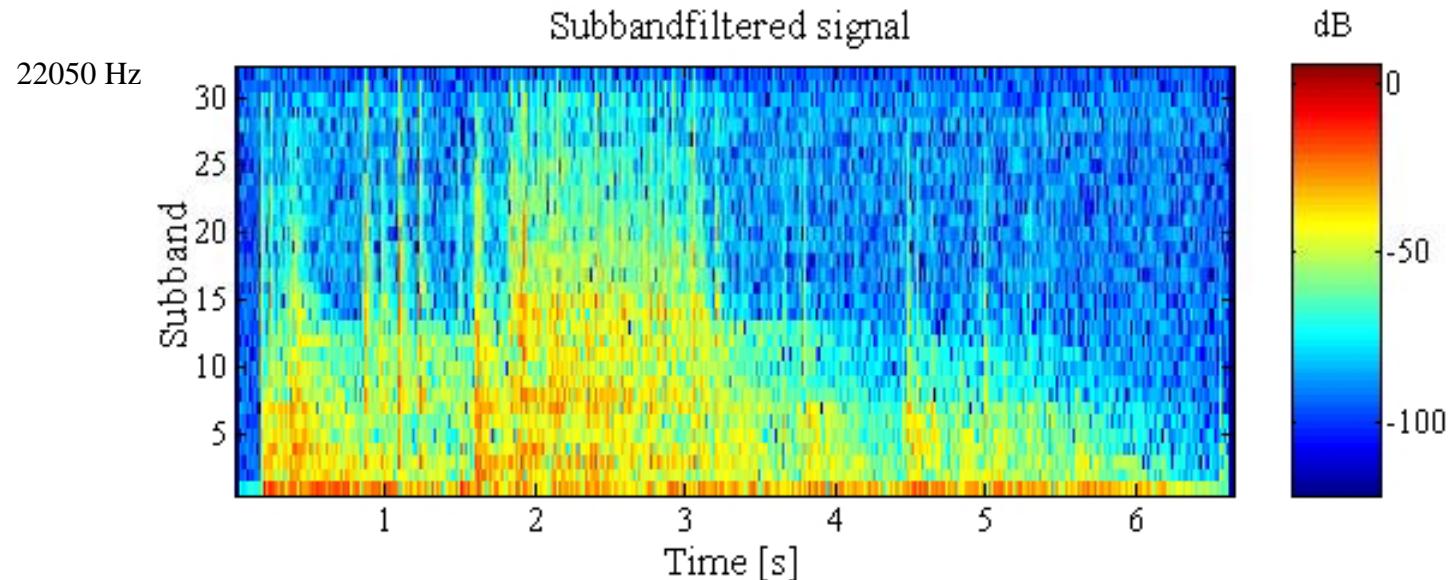
# Castanets and Guitar



Time domain signal

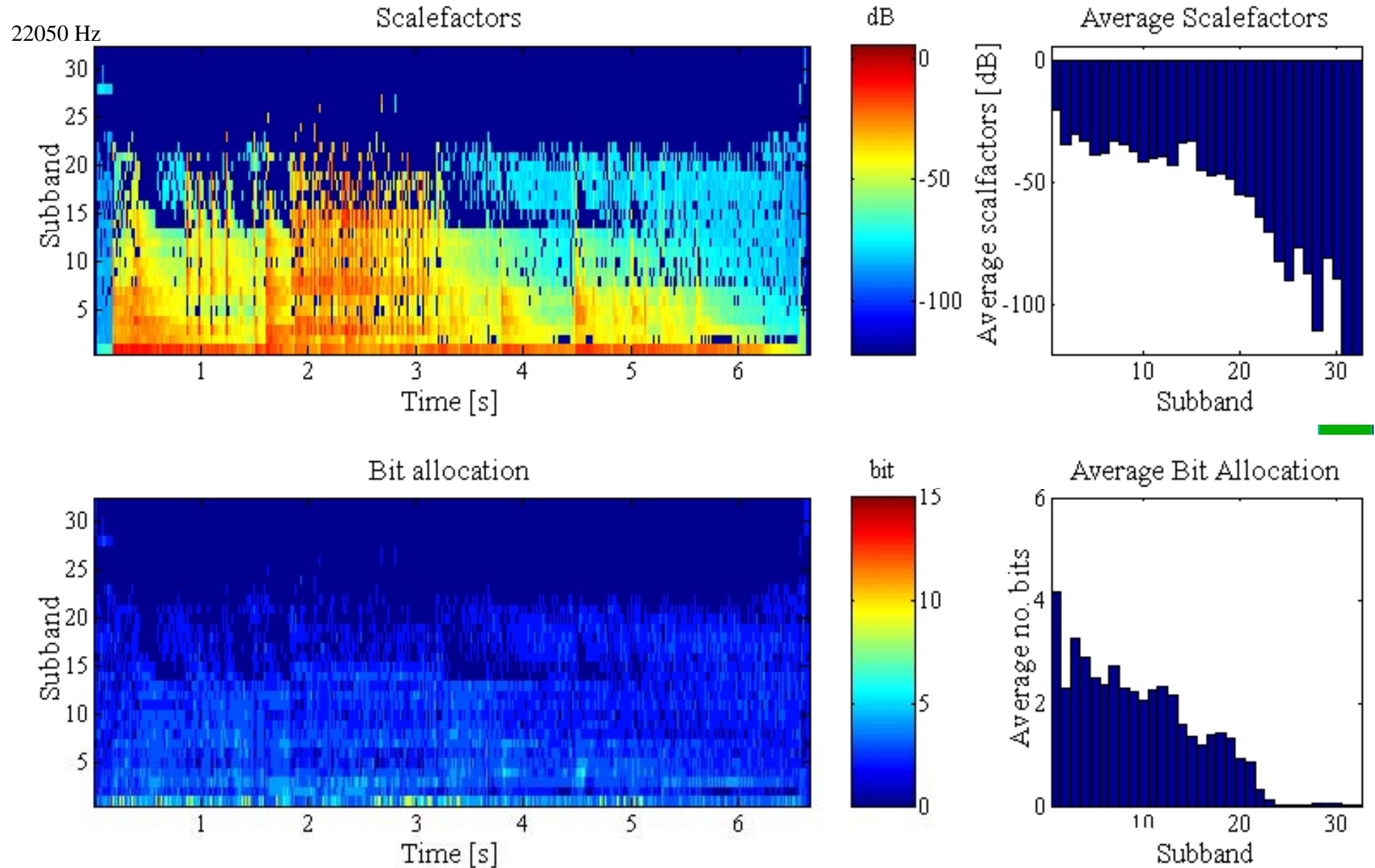


Subbandfiltered signal



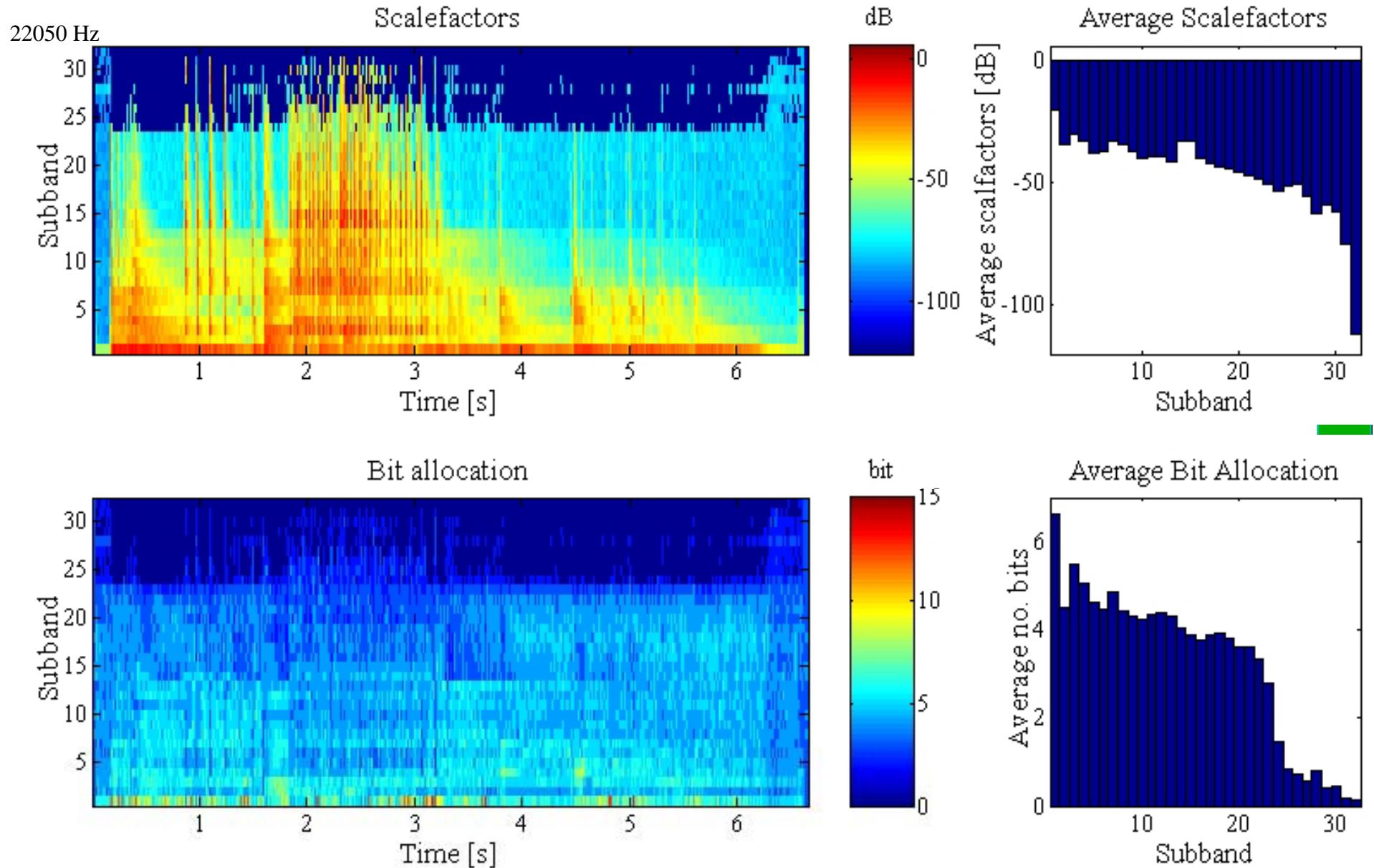


# Bit allocation with 2 bits per sample



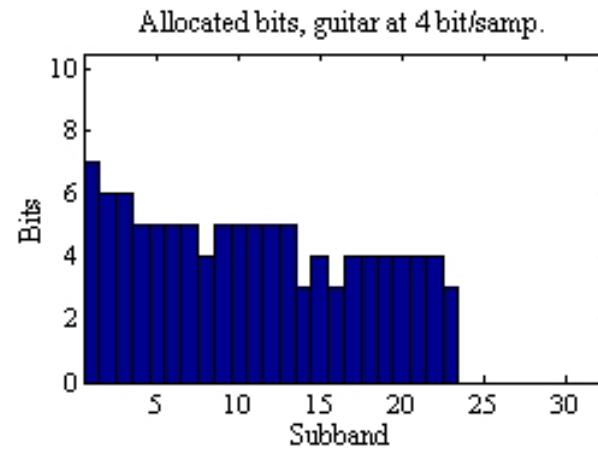
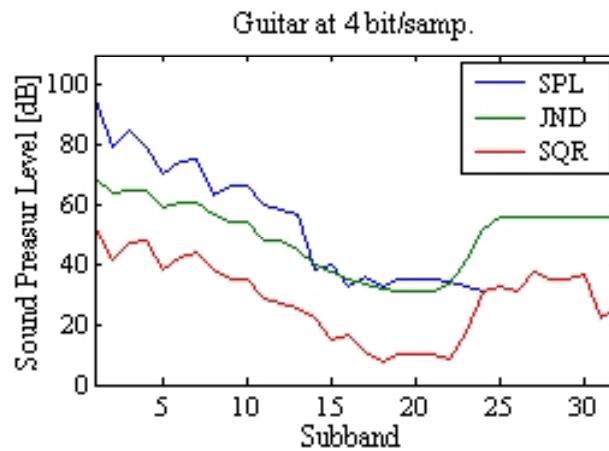


# Bit allocation with 4 bits per sample

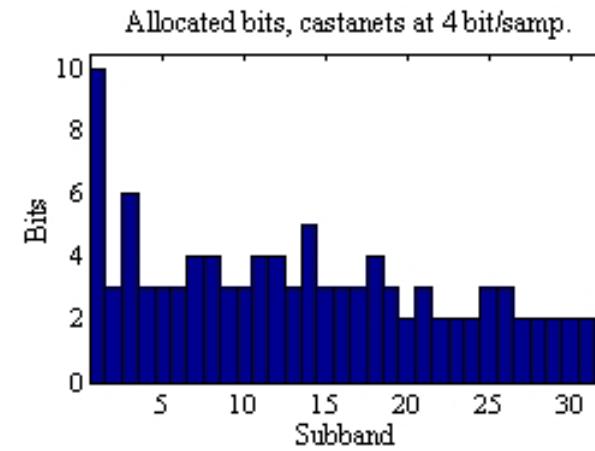
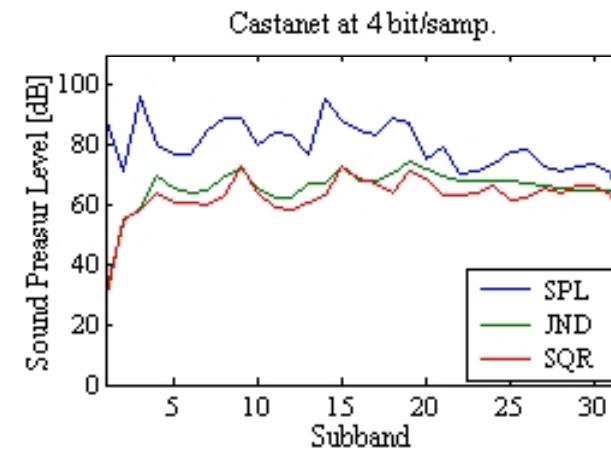




# Signal to Quantization Noise Ratio and the Just Noticeable Distortion



Frame at  $t=0.6$  s



Frame at  $t=0.6$  s





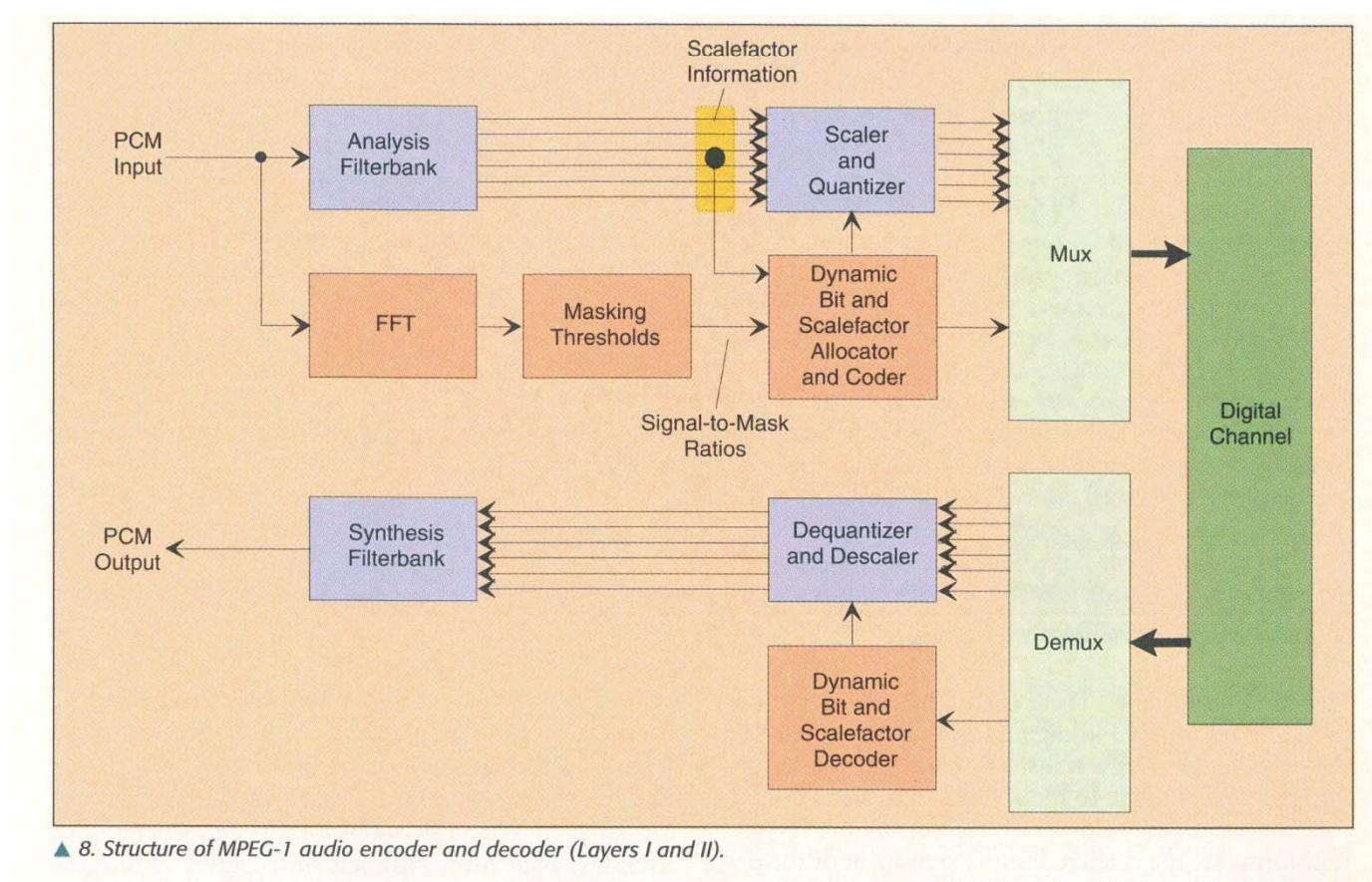
# Examples on compression



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



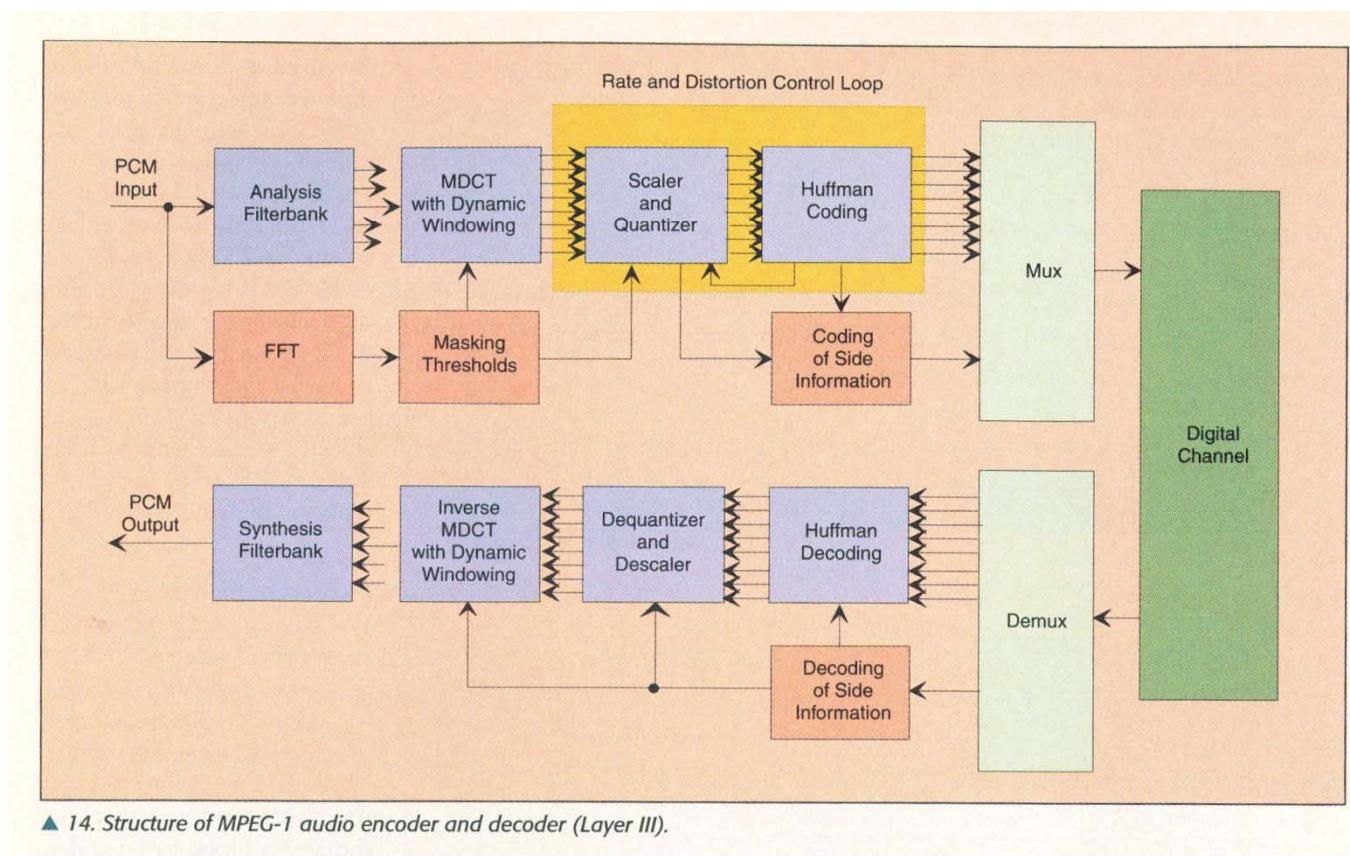
# MPEG-1 layers I and II



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



# MPEG-1 layer III = MP3



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 IIs filterbank  
 $\Rightarrow 32 * 18 = 576$  frekvensbånd  $\Rightarrow$ 
  - 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 lydkvalitet:
  - 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



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



UNIVERSITETET  
I OSLO

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

