COM-202: Signal Processing

Chapter 10: MP3 compression

Overview:

- Timeline of audio formats
- The human auditory system
- MP3 compression

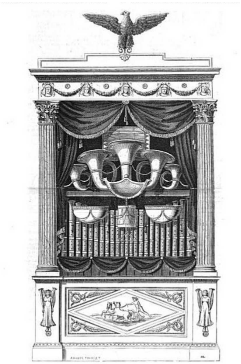
Timeline of audio formats

An audio format could refer to:

- recording media
  - physical method for storage
  - paper, wax, vinyl, magnetic tape, etc.
- digital file format
  - encoding and decoding standards
  - wav, mp3, mpeg-4, etc.

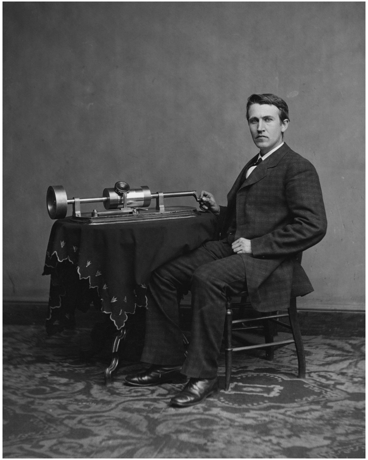
Early analogue devices and storage

Panharmonicon



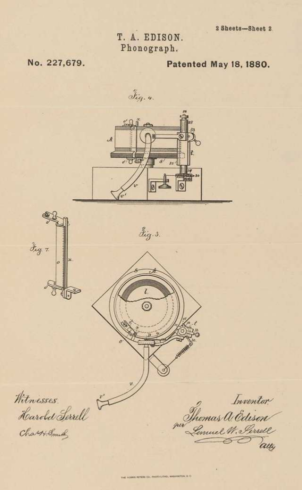
- invented in 1805 by Johann Nepomuk Maelzel
- earliest known automated music player
- directed by rotating cylinders

Phonograph

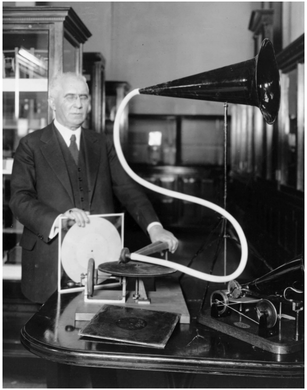


- invented in 1877 by Thomas Edison
- earliest commercial audio device
- music recorded on tin foil and wax cylinders

Phonograph


2 Sheets—Sheet 2
T. A. EDISON.
Phonograph.
No. 227,679.
Patented May 18, 1880.
Fig. 4.
Fig. 3.
Fig. 7.
Witnesses
Harold Serrell
Chas H Smith
Inventor
Thomas A Edison
per Lemuel W. Serrell
atty

Gramophone

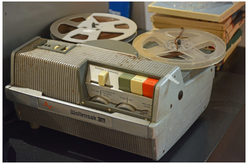


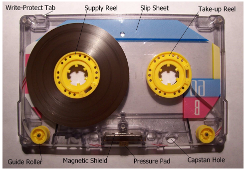
- Emile Berliner transitions to flat discs in 1890s
- discs surpassed cylinders by 1912
- many variations throughout 20s century
  - materials (shellac, vinyl)
  - diameter (12-inch, 10-inch, 7-inch)
  - rotational speed (33.3 rpm, 45 rpm, 78 rpm)

LP vinyl records

- introduced in 1948
- stands for 'long play'
  - each side played 25 mins
  - stared the 'album era' in rock music
- niche resurgence in the 21st century

Electromagnetic storage





- can store analogue and digital signals
  - used for audio, video, data backup
- compact cassette
  - capacity - up to 60 mins per side
  - developed in the 1960s
  - overtook LPs in the 1980s
  - where overtaken by compact discs in the 1990s

Analogue vs digital audio recording

analogue storage:

- continuous-time signals
- highly specialized analogue devices
- physical media and encoding format are inseparable

digital storage:

- discrete-time signals
- bits are universal currency
- bits and storage medium are separate entities

Two possible questions:

- How to improve the physical storage to get more bits?
- How to encode relevant information with given bits more efficiently?

Digital audio encoding

- first widespread digital audio encodings on compact discs
- used relatively naive quantization and sampling methods
  - pulse-code modulation with $44,100$ Hz sampling frequency
  - 16-bit uniform quantization

Digital audio encoding - mp3

- uses lossy compression
- compact discs - normally stores 80 minutes of music
- compact discs - 5-8 hours of music in mp3s

Lossy data compression

- goal: reduce number of bits to represent original signal $x[n]$
- lossy compression, $x[n] \neq y[n]$
- hide compression distortion in “blind spots” of human hearing

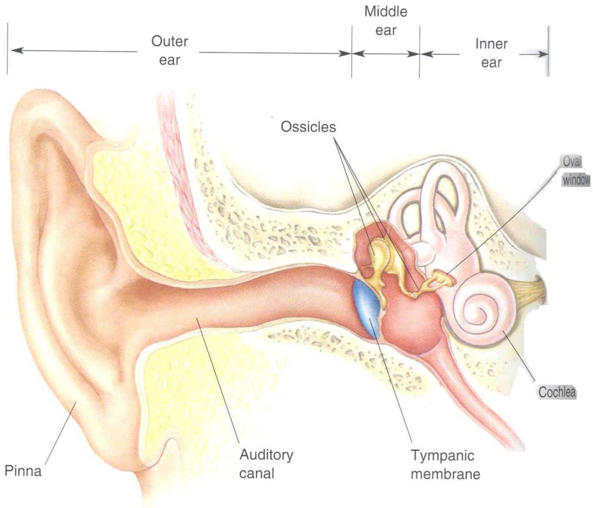
The human auditory system

The amazing human hearing system

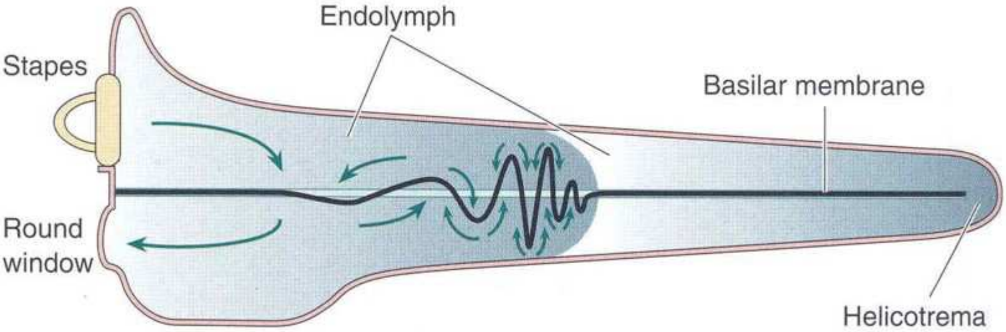
in short: it's insane!

- frequency range from 20 Hz to 20 kHz: ten octaves
- frequency resolution: 0.5%
- dynamic range in excess of 120 dB: a ratio of 1 trillion
- amplitude resolution: 0.6 dB (about 1%)
- time resolution: 10 μs
- ... and all this in the volume of a sugar cube!

The ear


Middle ear
Outer ear
Inner ear
Ossicles
Oval window
Cochlea
Pinna
Auditory canal
Tympanic membrane

The cochlea



Endolymph
Stapes
Basilar membrane
Round
window
Helicotrema

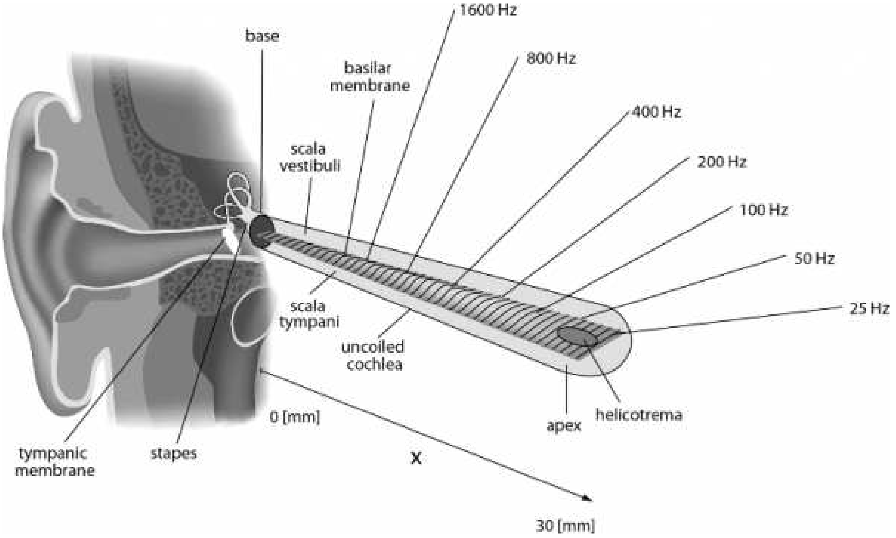
Did you know?

- “cochlea” is Greek for snail
- as big as a pea and 3cm long if unrolled (don’t unroll it!!)
- first organ to develop in a foetus
- surrounding bone is hardest in whole body

Did you know?

- about 20 thousand frequency-sensitive hair cells in the cochlea
- hair cells can *not* regenerate
- auditory nerve contains about 30 thousand fibers
- capacity of auditory nerve is about 6 Mb/s
- 150 million neurons in brain auditory cortex

Critical bands and tonotopic mapping on basilar membrane



The key to audio compression

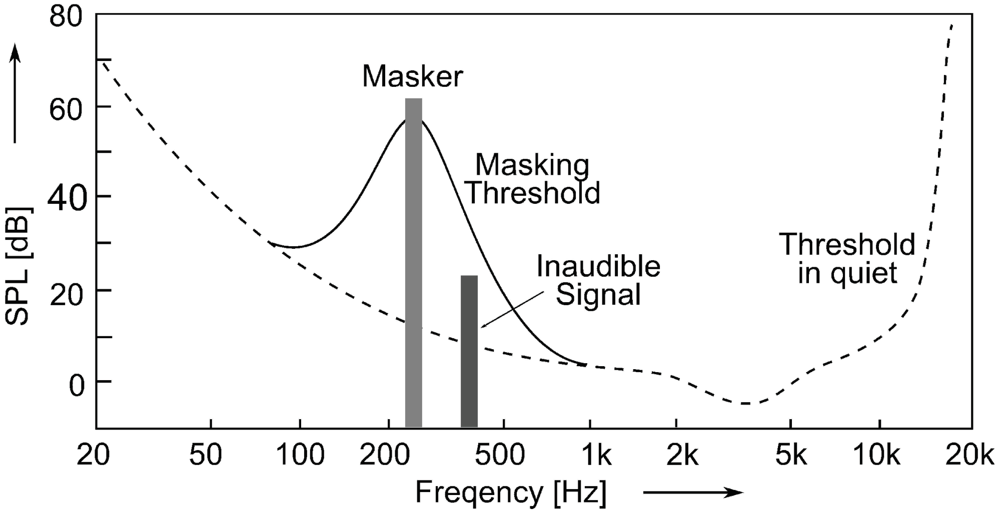
Masking

- noise can “drown out” other sounds
- a loud tone can mask softer tones close in frequency
- frequency-dependent thresholds of hearing

lossy compression discards the data that can’t be heard (well) anyway

to do this, we need an algorithmic model of the hearing system

Masking threshold for one masker

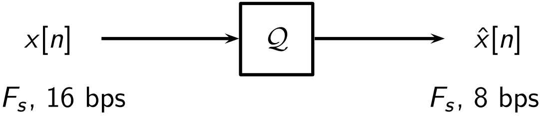


MP3 compression

Fun MP3 facts

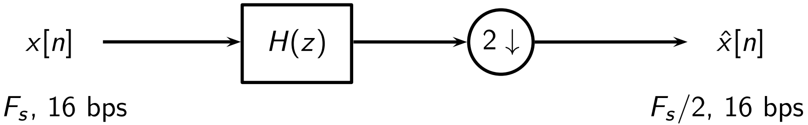
- developed in the late 1980s by Erlangen-Nuremberg University and the Fraunhofer Institute for Integrated Circuits
- first version published in 1992
- the MP3 name was adopted in 1995
- all patents expired in 2017; completely free today

Goal: cut bit rate in half. First idea:



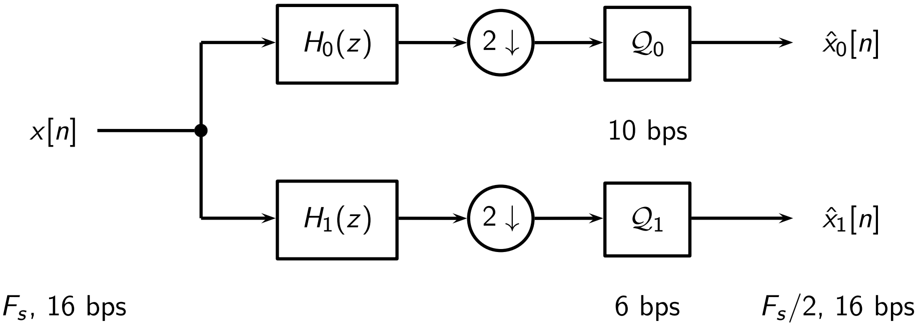
- drop the last 8 bits in each sample
- equivalent to a noise source at 48 dB

Goal: cut bit rate in half. Second idea:



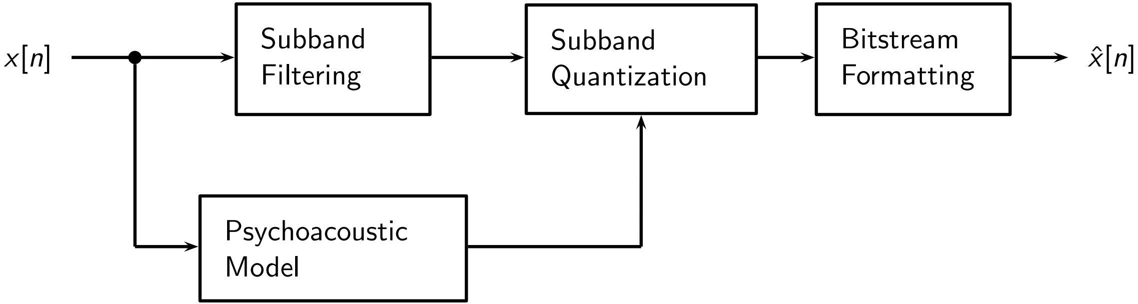
- $H(z)$ lowpass with cutoff $\pi/2$
- we are brutally discarding the high frequency content

Goal: cut bit rate in half. Combine both ideas:

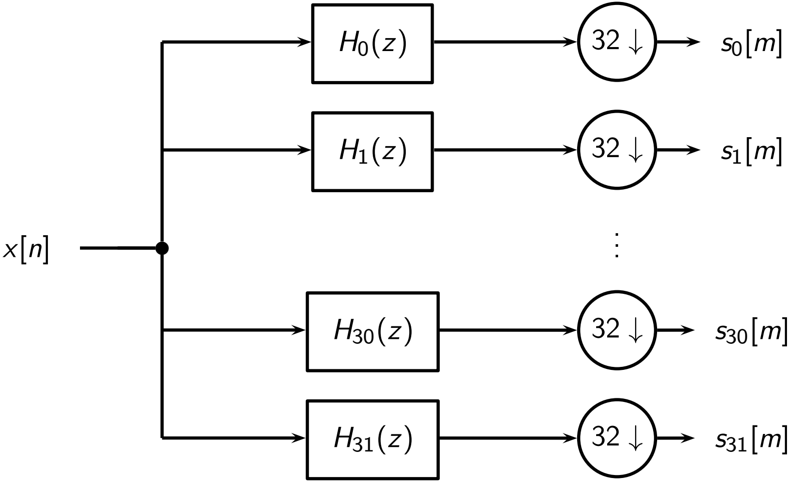


- $H_0(z)$ lowpass with cutoff $\pi/2$, $H_1(z)$ complementary highpass
- rationale: quantization noise less audible at high frequencies

MPEG Audio: encoder's block diagram

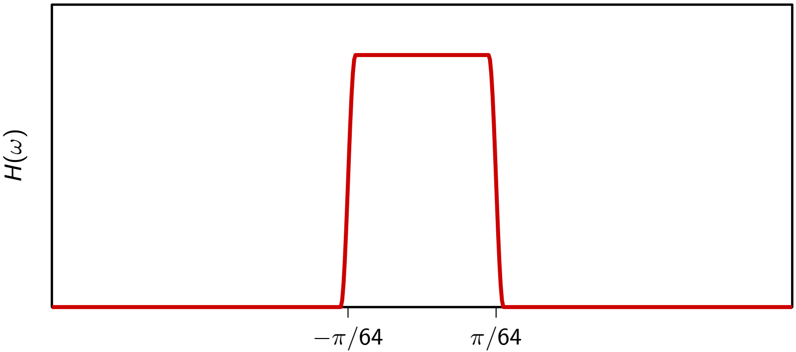


32-band filterbank

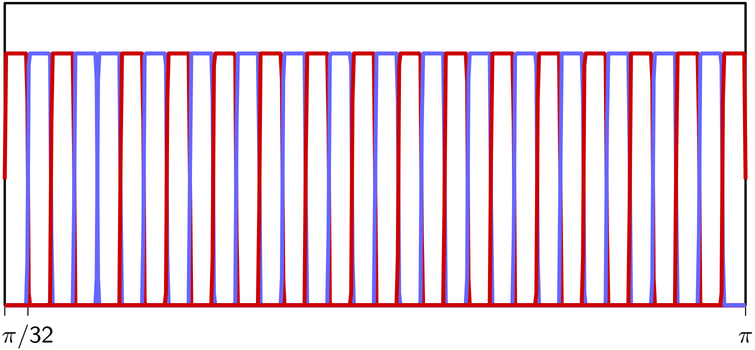


512-tap FIR subband filters

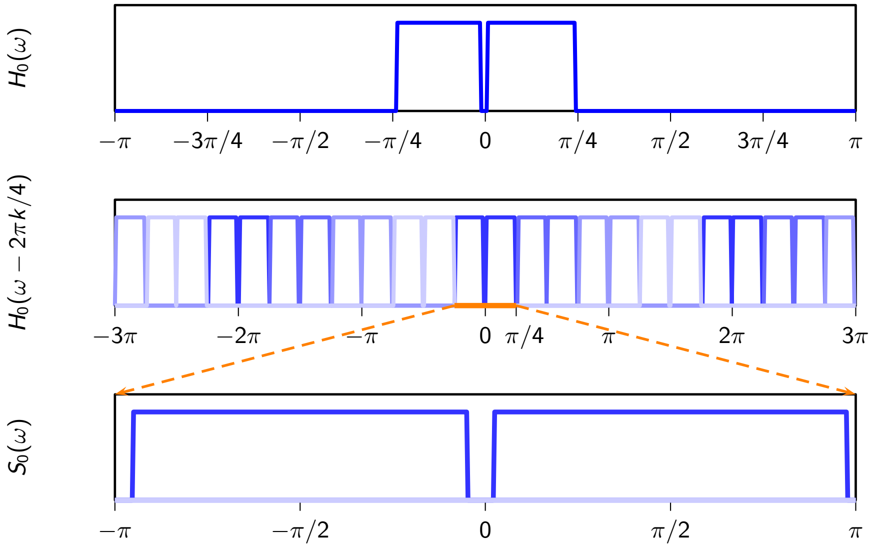
Prototype filter: frequency response



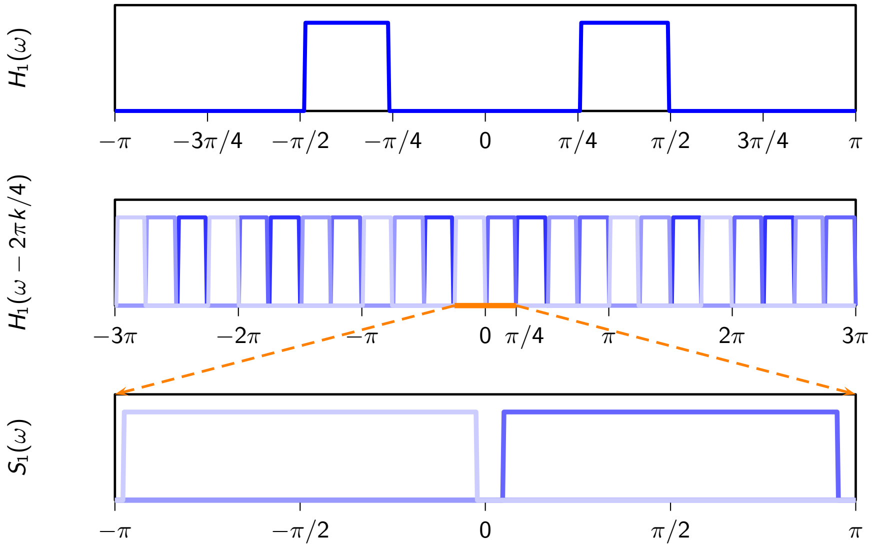
32-filter bank



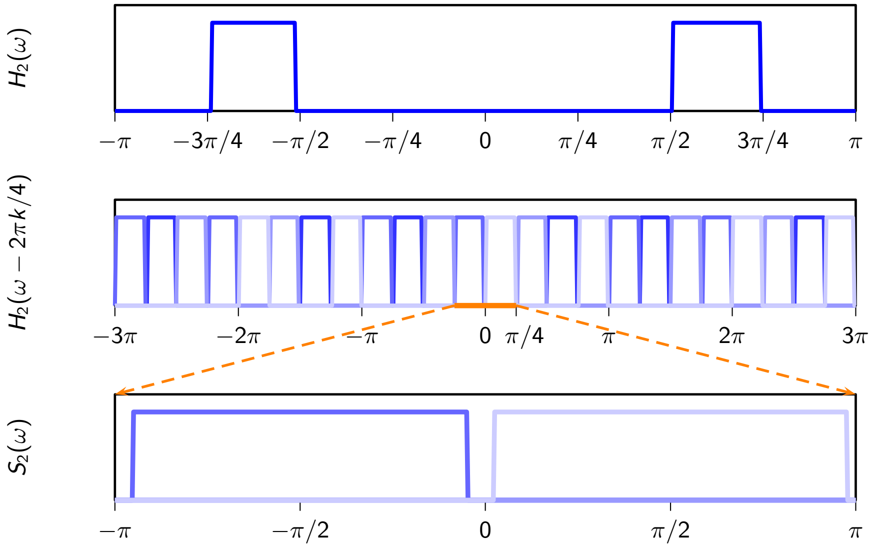
Critically sampled filterbank, 4 bands



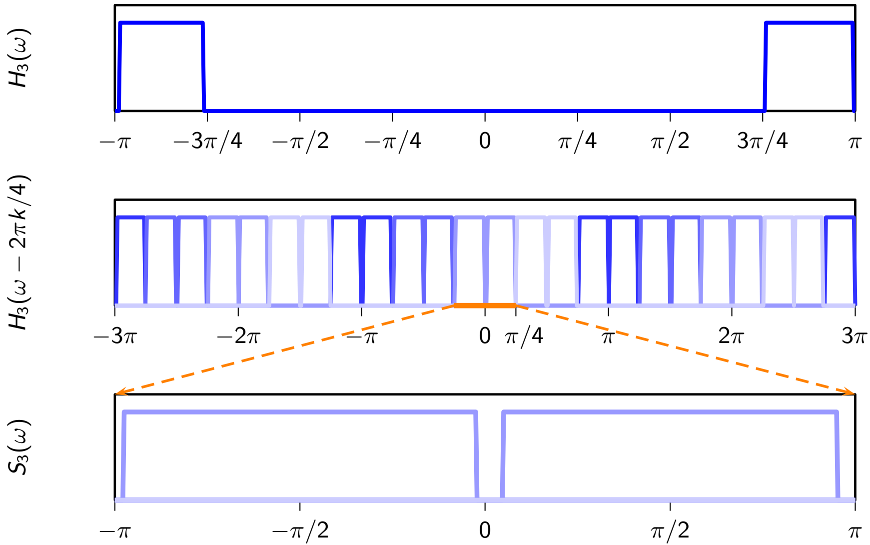
Critically sampled filterbank, 4 bands



Critically sampled filterbank, 4 bands



Critically sampled filterbank, 4 bands



Framing

- 384 samples (or 10 milli seconds) per frame (per channel)
- 12 samples per subband per frame
- each set of 12 subband samples normalized to the $[-1, 1]$ interval using one of 64 pre-defined scale factors (cost: 6 bits)

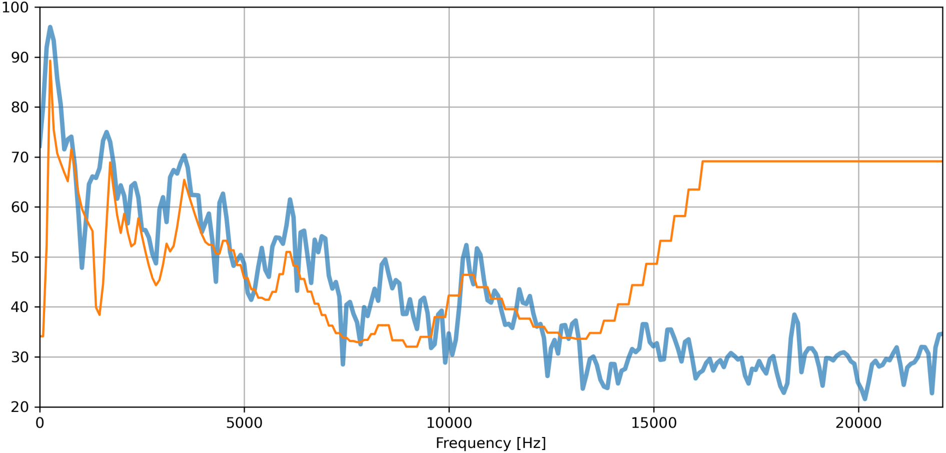
Quantization

- each subband quantized independently
- uniform quantizers with rates from 2 to 15 bps
- quantizer rate for each subband determined iteratively
- goal: push the subband quantization noise below the subband masking threshold
- masking thresholds provided by the psychoacoustic model

Notes on bit allocation

- can work without a psychoacoustic model: at each step choose subband with smaller SNR
- psychoacoustic modeling algorithm not part of MPEG standard: different vendors competed on that
- large asymmetry in computational requirements between encoder and decoder, typical of all lossy compression algorithms

Global masking threshold

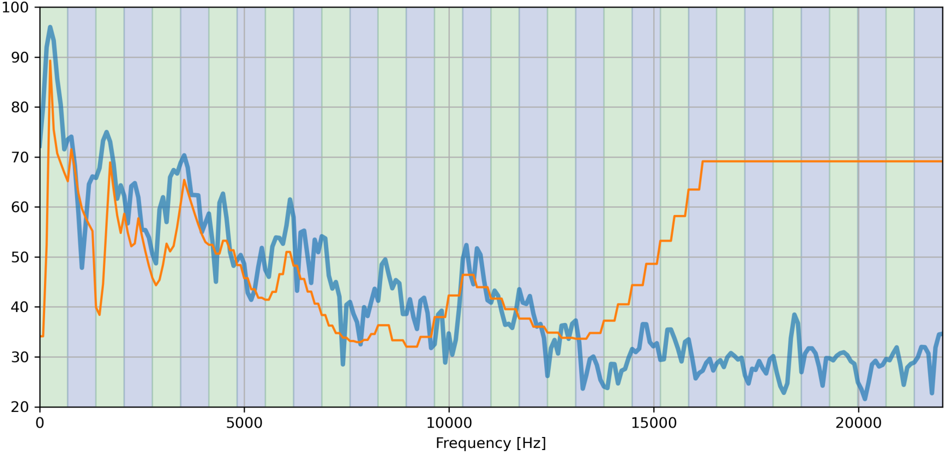


Bit allocation algorithm

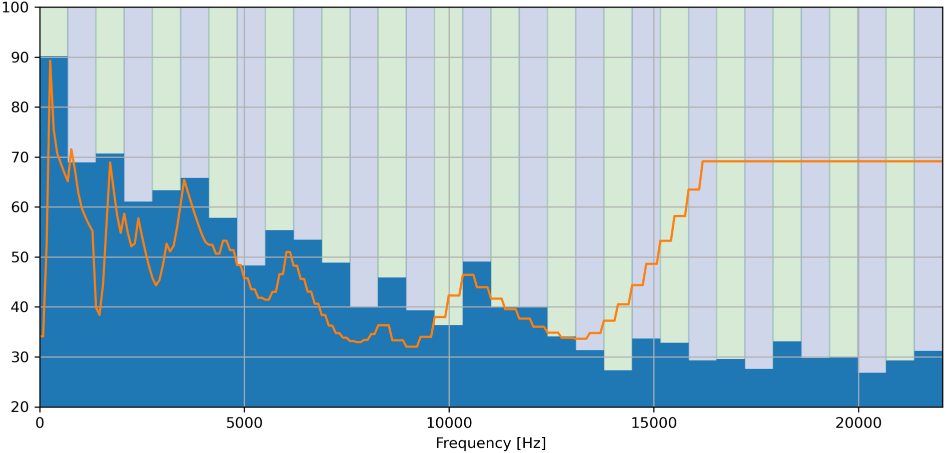
For each subband:

- compute signal power (average power over subband)
- find masking threshold (minimum threshold value over subband)
- compute signal-to-mask ratio (SMR)
- allocate 0 bits $\Rightarrow$ SNR $= 0$ dB

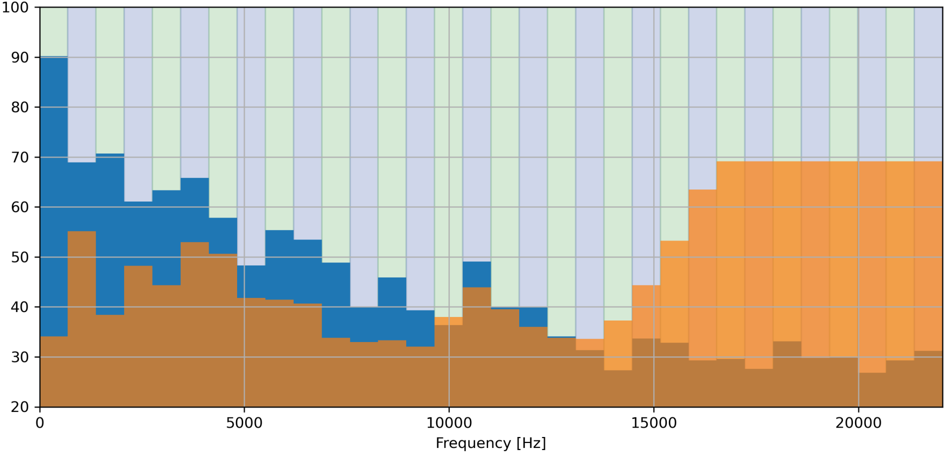
Power and Masking Curves



Subband Power



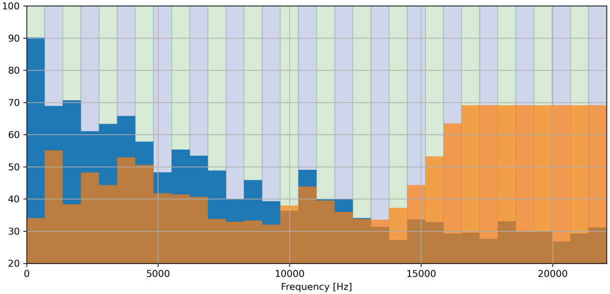
Subband Masking Threshold



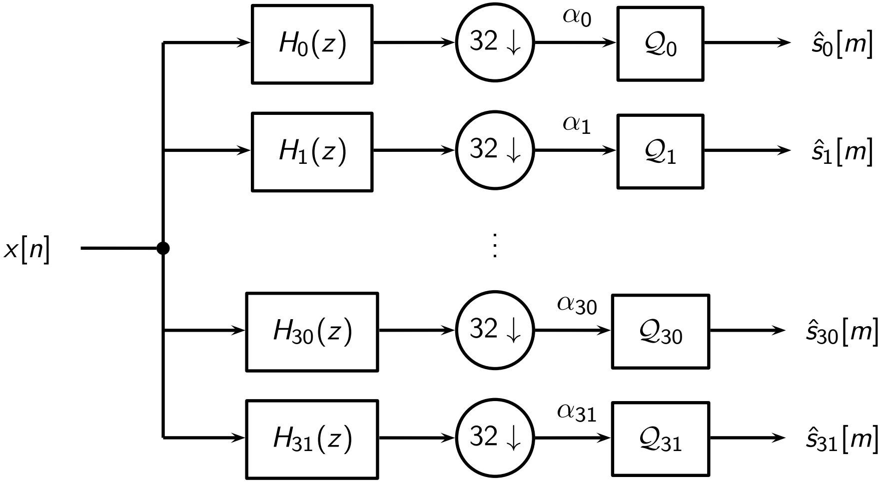
Bit allocation algorithm

Iterative bit allocation for a bit budget B: at step n

- $b_k[n]$ bits allocated to band k; SNR $\approx 6b_k[n]$
- compute mask-to-noise ratio (MNR = SNR - SMR) for each subband
- choose subband with largest MNR and allocate one extra bit
- repeat until all B bits used



Encoding


$x[n]$
$H_0(z)$
$32\downarrow$
α_0
$\mathcal{Q}_0$
$\hat{s}_0[m]$
$H_1(z)$
$32\downarrow$
α_1
$\mathcal{Q}_1$
$\hat{s}_1[m]$
$\vdots$
$H_{30}(z)$
$32\downarrow$
α_{30}
$\mathcal{Q}_{30}$
$\hat{s}_{30}[m]$
$H_{31}(z)$
$32\downarrow$
α_{31}
$\mathcal{Q}_{31}$
$\hat{s}_{31}[m]$

Concluding remarks - mp3 compression

- near universal audio compression format today
- offers very good tradeoffs between file size and music quality
- uses psychoaucustic coding to ‘hide’ distortion where a human cannot hear it
- exact specifications complex, but the key idea uses
  - multi-rate sampling
  - quantization