Music Compression

Sound in general

- Sound in general: continuous wave through gas, liquid. Esp. air.
- Longitudinal wave.
- Made up of pressure differences
- Detected by measuring pressure levels at a location: tympanum, ear-drum.
- Microphone changes analog sound pressure to analog voltage levels.
- Energy is traveling through the air.

Human audio

- Audible sound frequency: 16-20Hz to 20-22 kHz
- Speed at 1 atm pressure, 20C temperature: 343.8 m/s.
- Wavelength: 20 000Hz$\rightarrow$ 1.72 cm, 20Hz$\rightarrow$ 17.2 m.
- Displacement of air molecules: about 10^{-6} cm, at most 10^{-3} cm.
- Loudness: pressure, energy.
- Dynamics of human ear: 10^{13}. The ration of hearing threshold and pain level at 1000Hz.
- Relative logarithmic scale: P is power, p is pressure

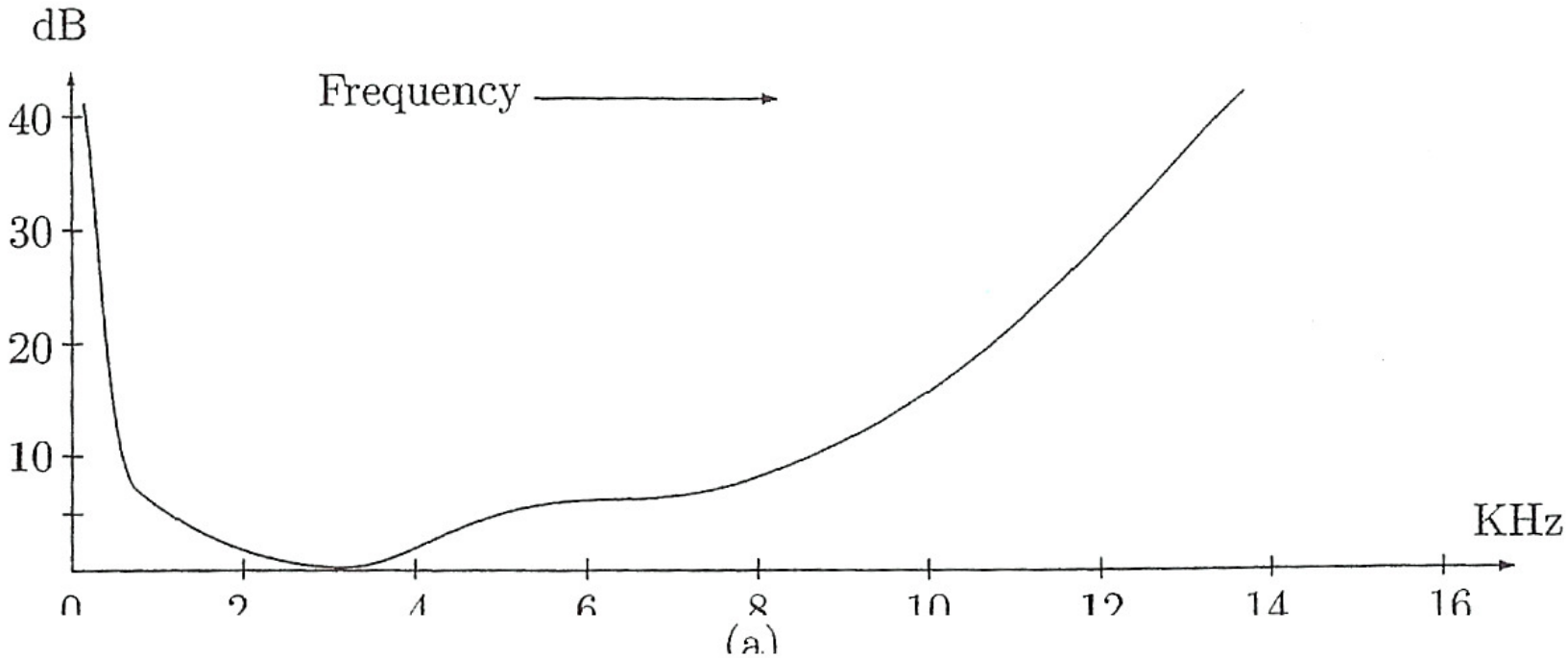
$$10 \log_{10} \frac{P_1}{P_2} = 20 \log_{10} \frac{p_1}{p_2} \, dB \quad (decibel).$$

Reference: hearing threshold at 1000Hz, i.e. $10^{-12}\, W/m^2$.

Psychoacoustic model

- Hearing threshold
- Frequency masking
- Temporal masking
- Critical bands: Bark scale
- Nonlinear quantization

Sensitivity of the ear



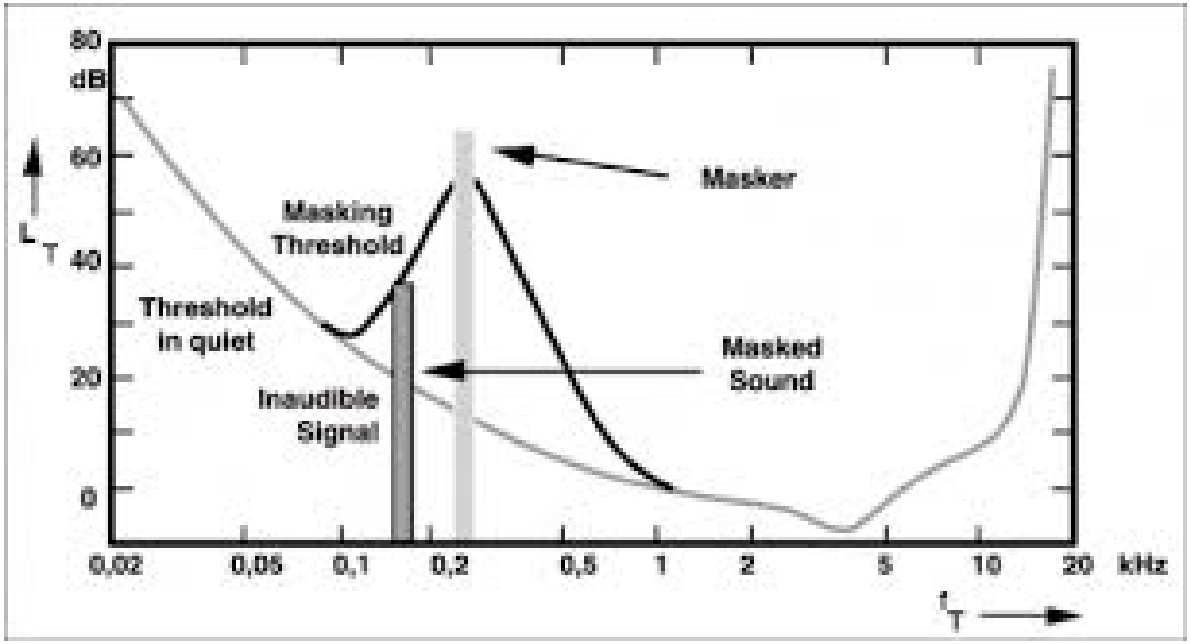
Hearing threshold

Expressed as relative signal amplitude (db).

Frequencies only heard if they exceed a sensitivity threshold.

Sensitivity of the ear is dependent on frequency. Most sensitive in range of 2-5KHz

Frequency masking



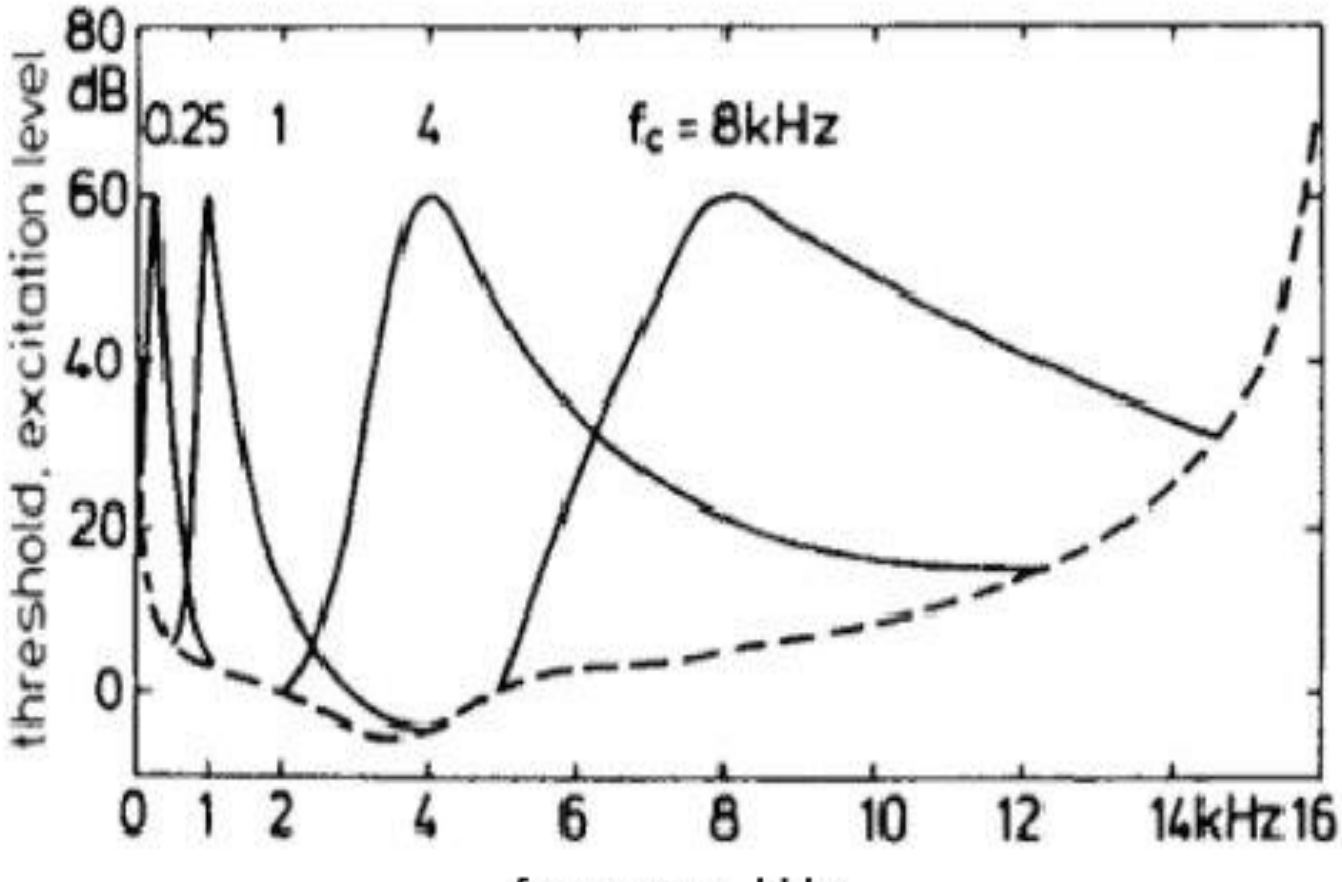
Frequency masking

Frequency masking

- The sensitivity threshold curve is distorted by the presence of loud sounds.
- The threshold is elevated by the loud sound in its frequency neighborhood:
  Frequencies just above and below the frequency of a loud sound need to be louder than the normal minimum amplitude before they can be heard.
- The characteristics of the frequency masking effect depends on the frequency range.

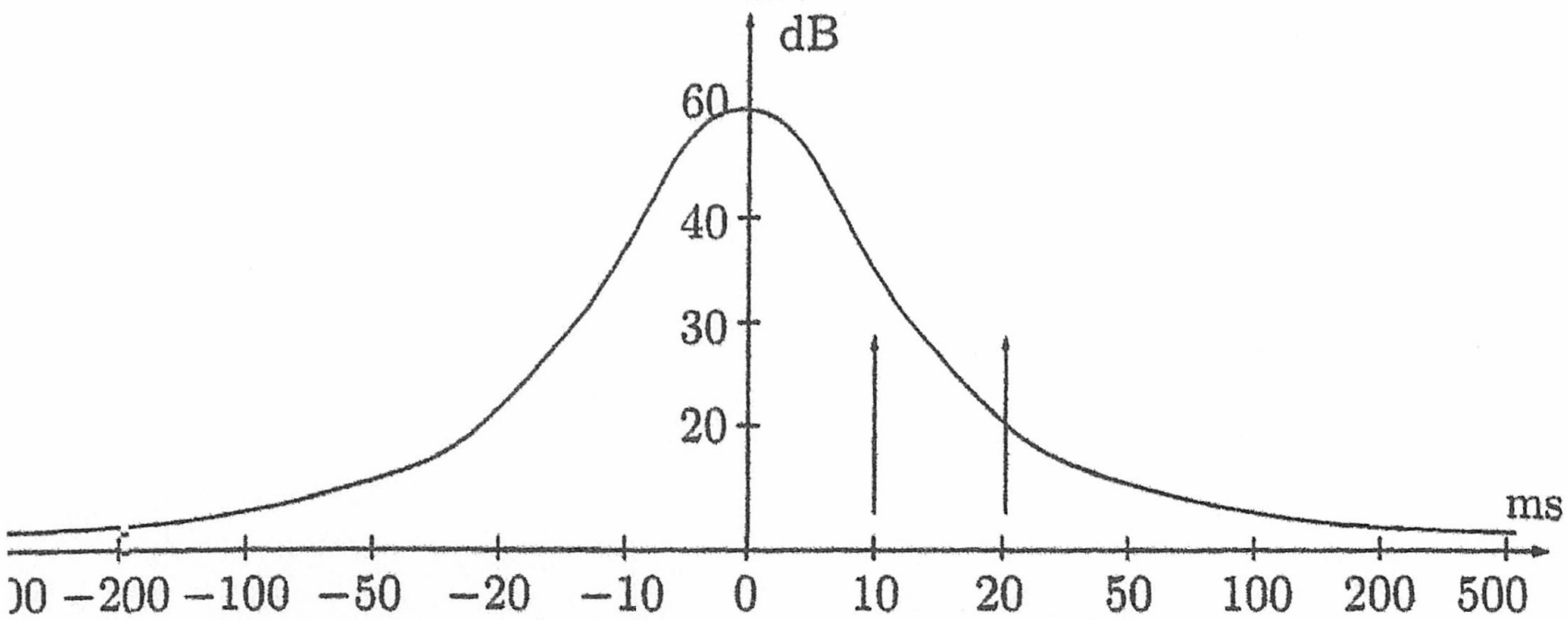
Frequency masking

Frequency masking curves



Temporal masking

After hearing a loud sound, the ear is deaf to quieter sounds in the same frequency range for a short time.



Critical bands: Bark scale

Resolution of auditory system: 27 critical bands called "barks".

The width of the bands change with the index:

- Less than 100 Hz in the lower frequency region
- Greater than 4kHz in the upper frequency region

Formula for the band boundaries (approximation)

$$1\text{Bark} = \begin{cases} \frac{f}{100} & , f < 500\text{Hz}, \\ 9 + 4\log_2 \frac{f}{1000} & , f > 500\text{Hz}. \end{cases}$$

The parameters in the psychoacoustic model (threshold, frequency and temporal masking) change form one critical band to the other one.

Consequence: In order to optimize the parameters the signal must be decomposed according to the critical bands.

Critical bands

Table of empirical values of the critical bands.

Band Number	Frequency (Hz)[1]	Band Number	Frequency (Hz)[1]
0	50	14	1,970
1	95	15	2,340
2	140	16	2,720
3	235	17	3,280
4	330	18	3,840
5	420	19	4,690
6	560	20	5,440
7	660	21	6,375
8	800	22	7,690
9	940	23	9,375
10	1,125	24	11,625
11	1,265	25	15,375
12	1,500	26	20,250
13	1,735		

[1] Frequencies are at the upper end of the band.

Exploitation of psychoacoustic properties

Note that the psychoacoustic properties are expressed in the frequency domain.

Consequence: Discrete Fourier transform is to be performed.

In order to imitate the Bark scale: decomposition in the frequency domain into 32 subbands by means of filter banks.

The main components of music compression in mpeg layers

- hearing threshold
- frequency masking
- temporal masking
- nonlinear quantization, dynamic bit allocation