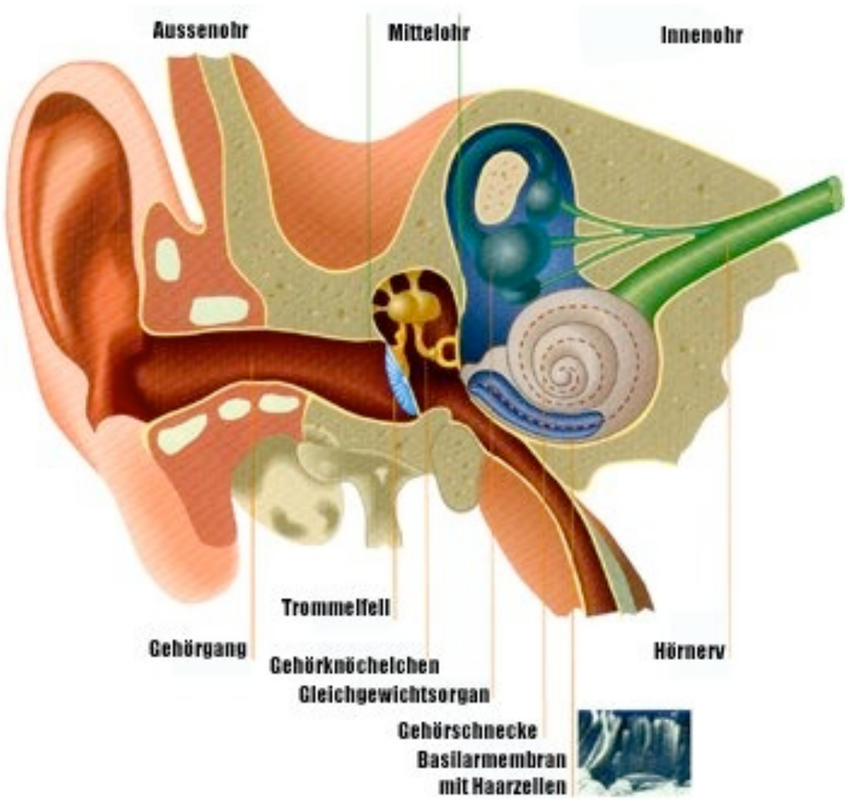
# Geschichte und Technik Digitaler Medien

## 08 – Audiokompression

Prof. Dr. Jochen Koubek

# Hörphysiologie

# Ohr

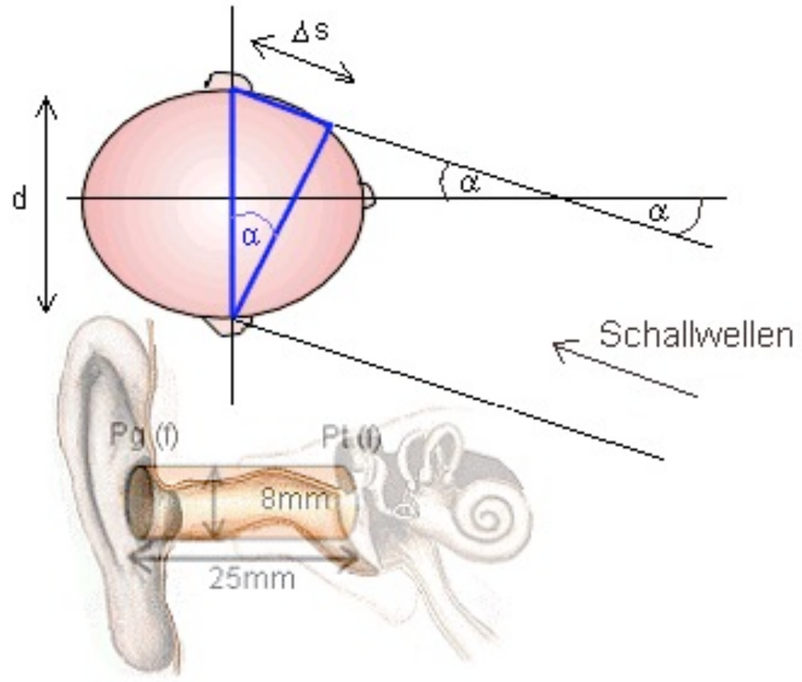



Aussenohr

Mittelohr

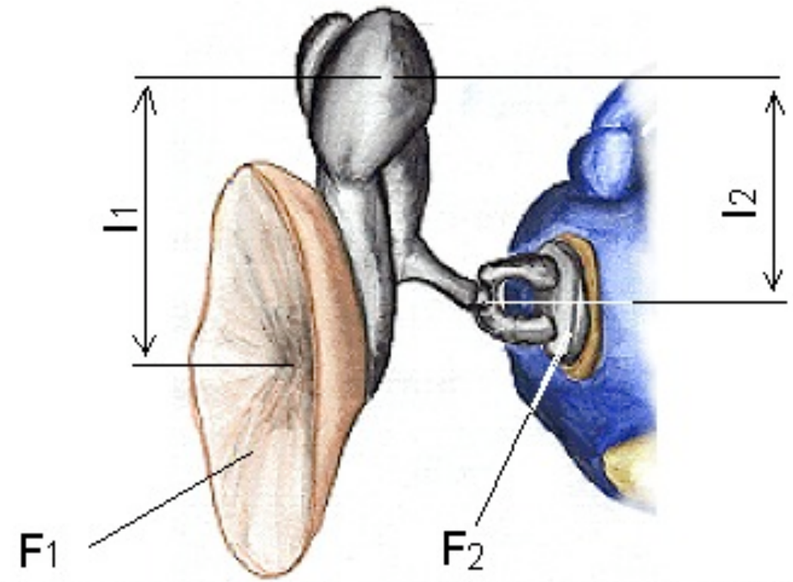
Innenohr

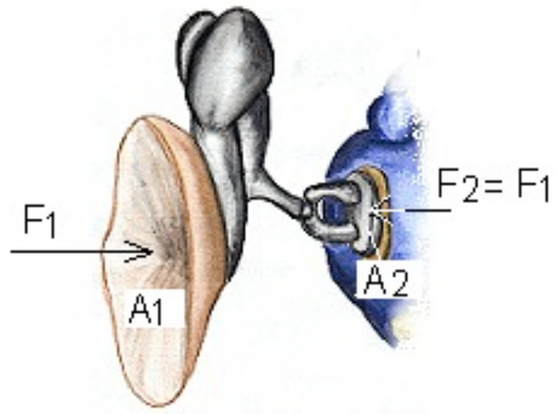
# Aussenohr



# Mittelohr

l1 , l2 : Wirksame Länge der Gehörknöchelchen

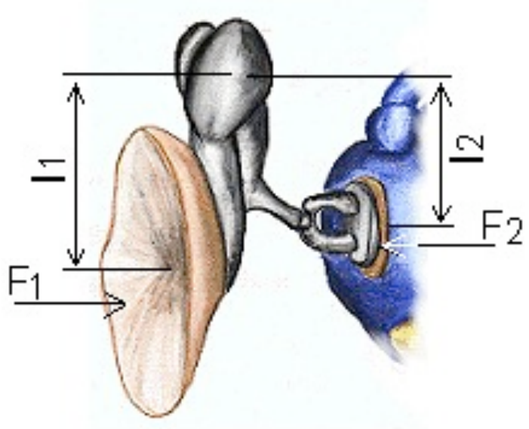



$p_1 = F_1 / A_1$ $p_2 = F_2 / A_2$

A1 = 55mm² A2 = 3,2mm²

vF=55/3,2 = 17



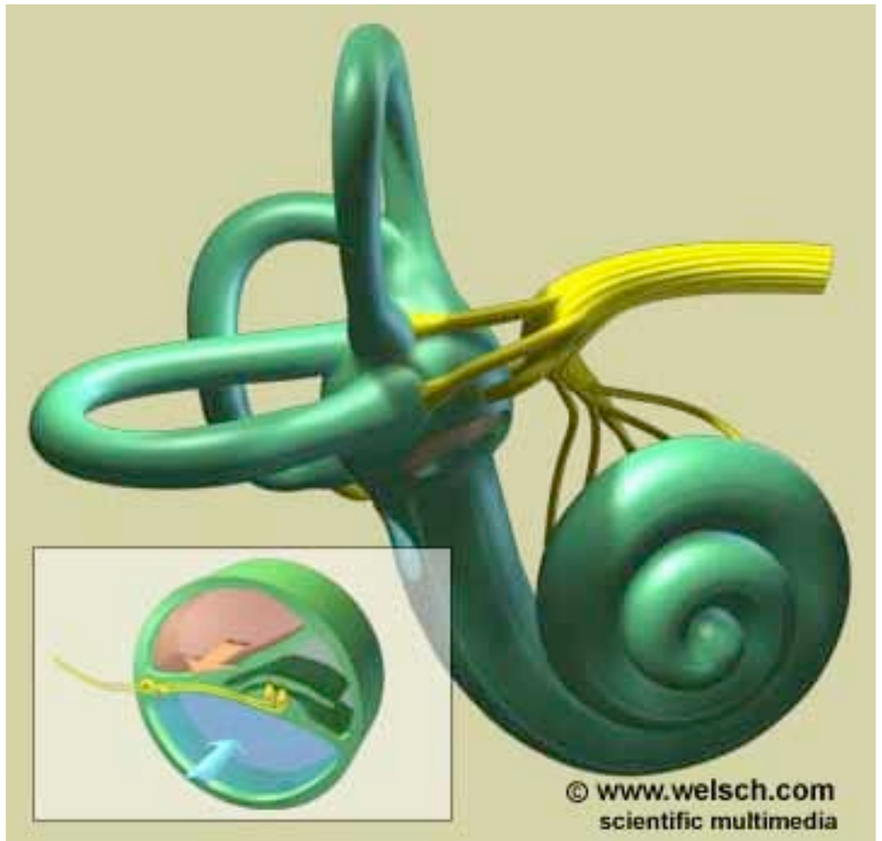

vH = 1,3

$v = vF \cdot vH$
$v = 17 \cdot 1,3$
$v = 22$

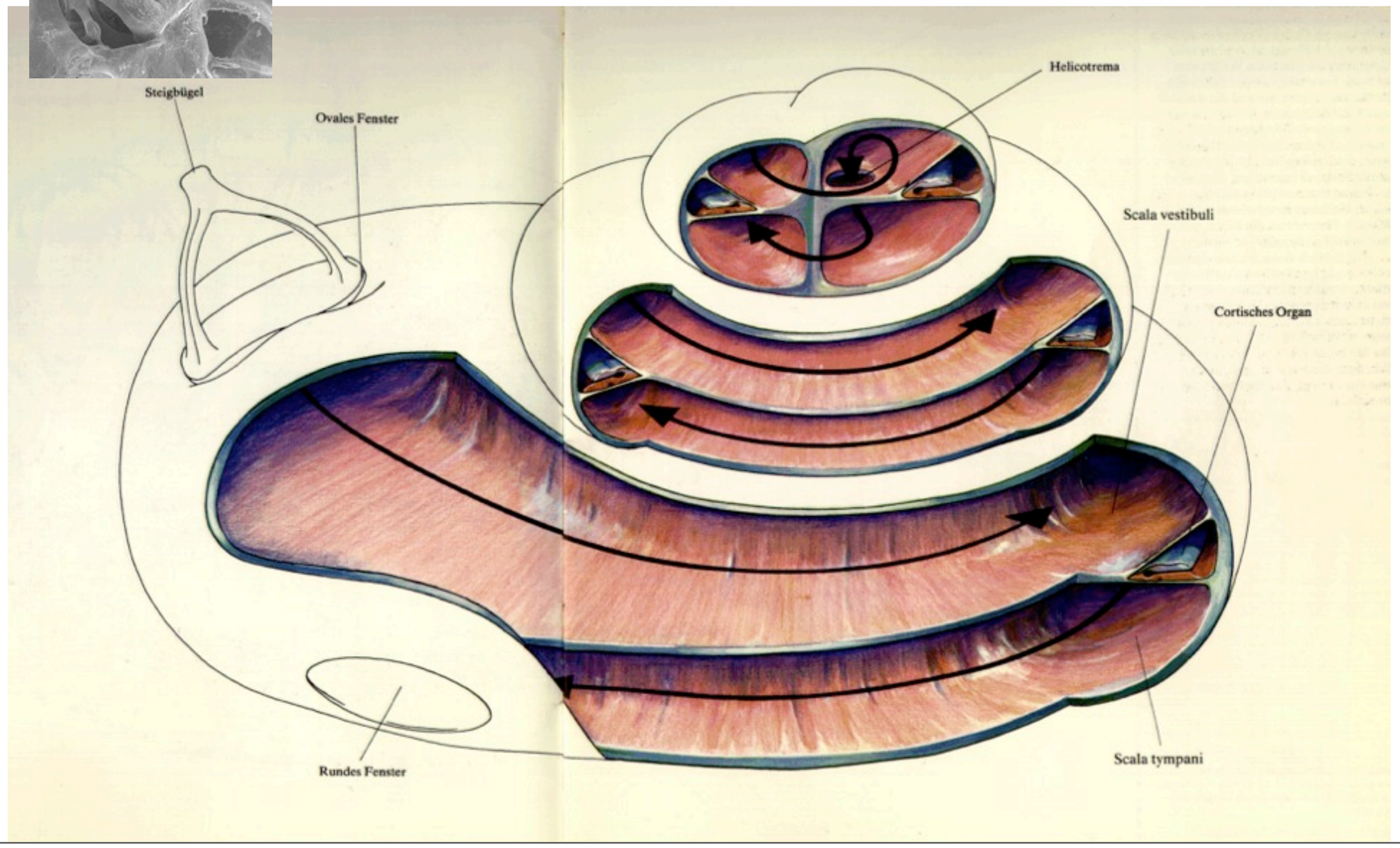
Aufgrund der Längenverhältnisse der wirksamen Hebelarme der Gehörknöchelchen und ihrer Lage zueinander, wird die Auslenkung der Steigbügelfußplatte gegenüber der Auslenkung am Trommelfell um den Faktor 1,3 vermindert.
Unter Anwendung des Hebelgesetzes
$F_1 \cdot l_1 = F_2 \cdot l_2$
bedeutet dies, daß die Kraft auf die Steigbügelfußplatte um den gleichen Faktor verstärkt wird.

http://www.dasp.uni-wuppertal.de/ars_auditus/physiologie/mittelohr12.htm

# Innenohr

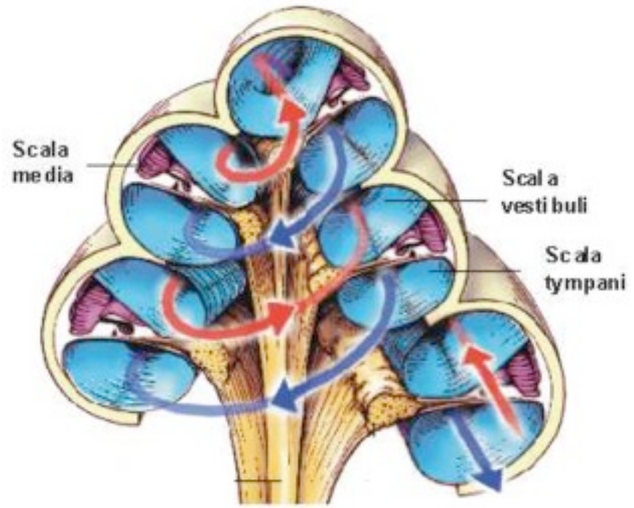


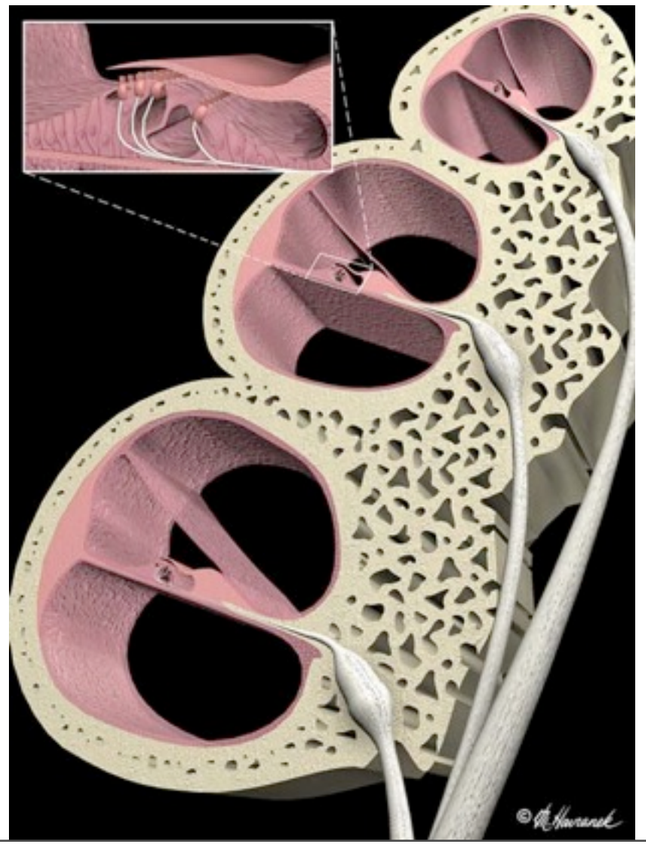
# Schnecke

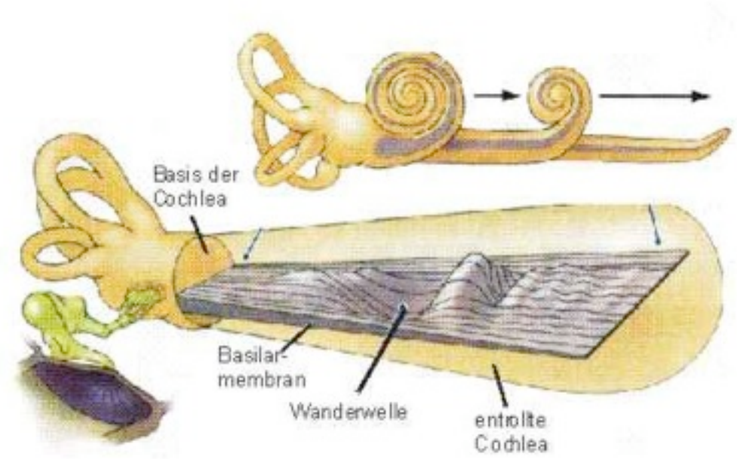

Steigbügel
Ovales Fenster
Helicotrema
Scala vestibuli
Cortisches Organ
Rundes Fenster
Scala tympani

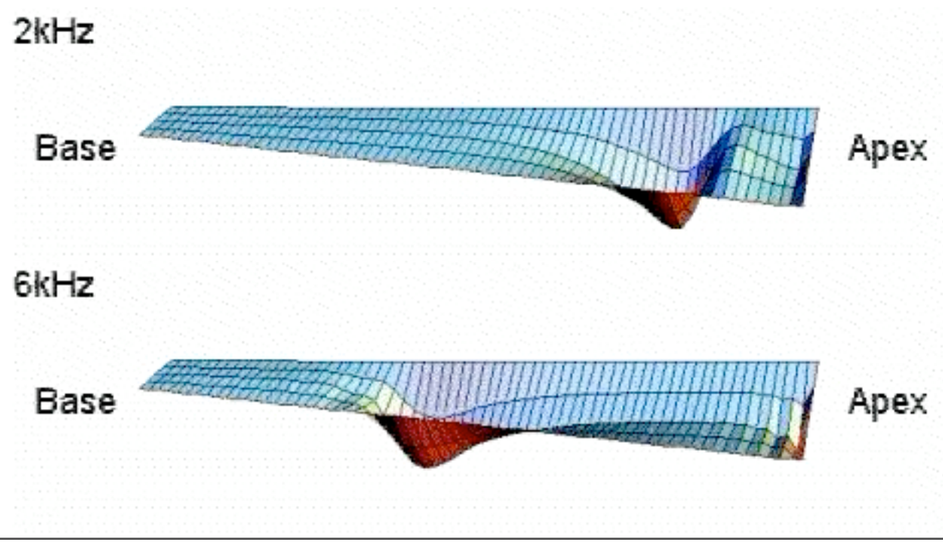
# Basilarmembran

Länge: 35 mm

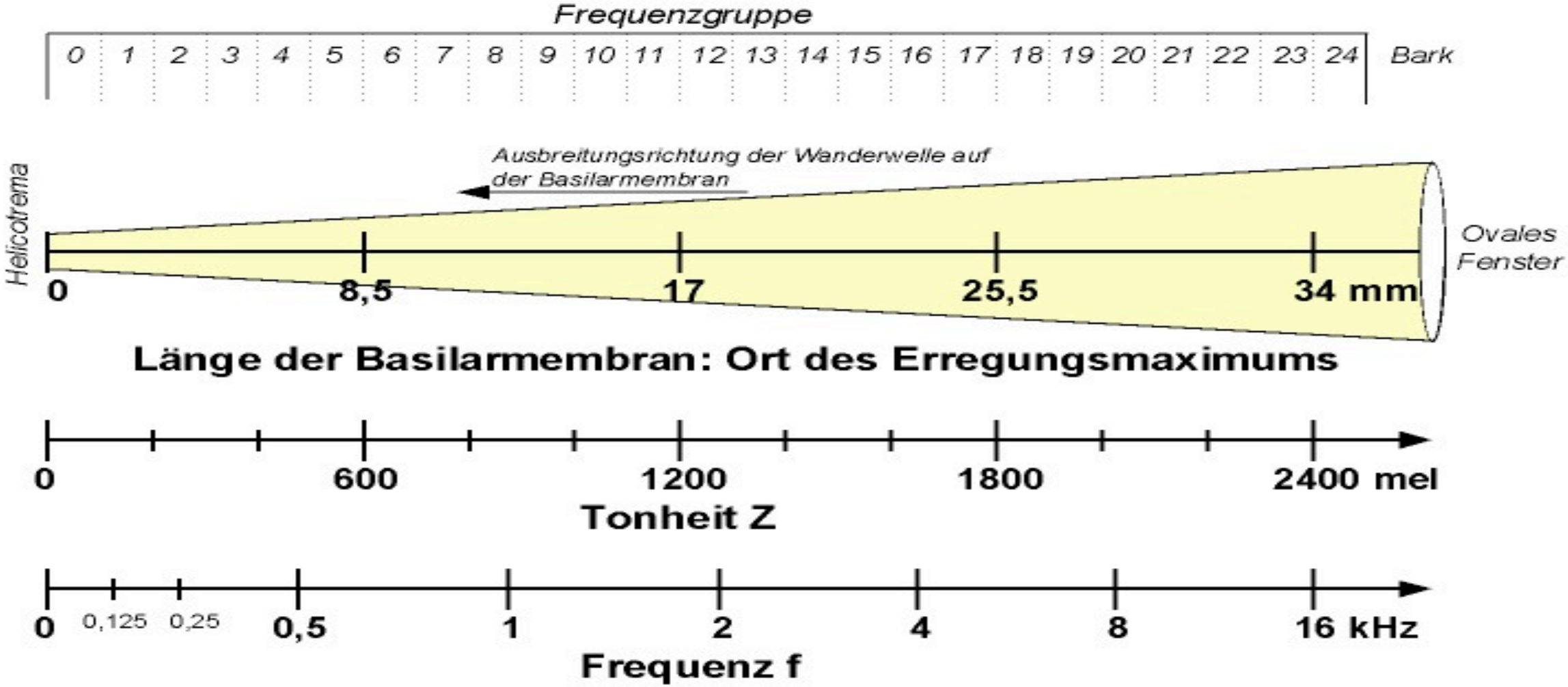




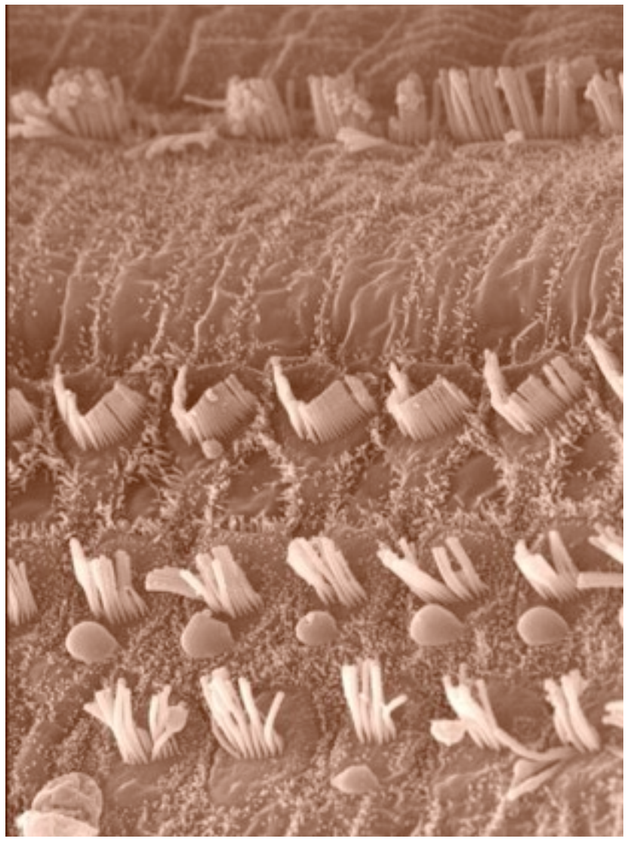



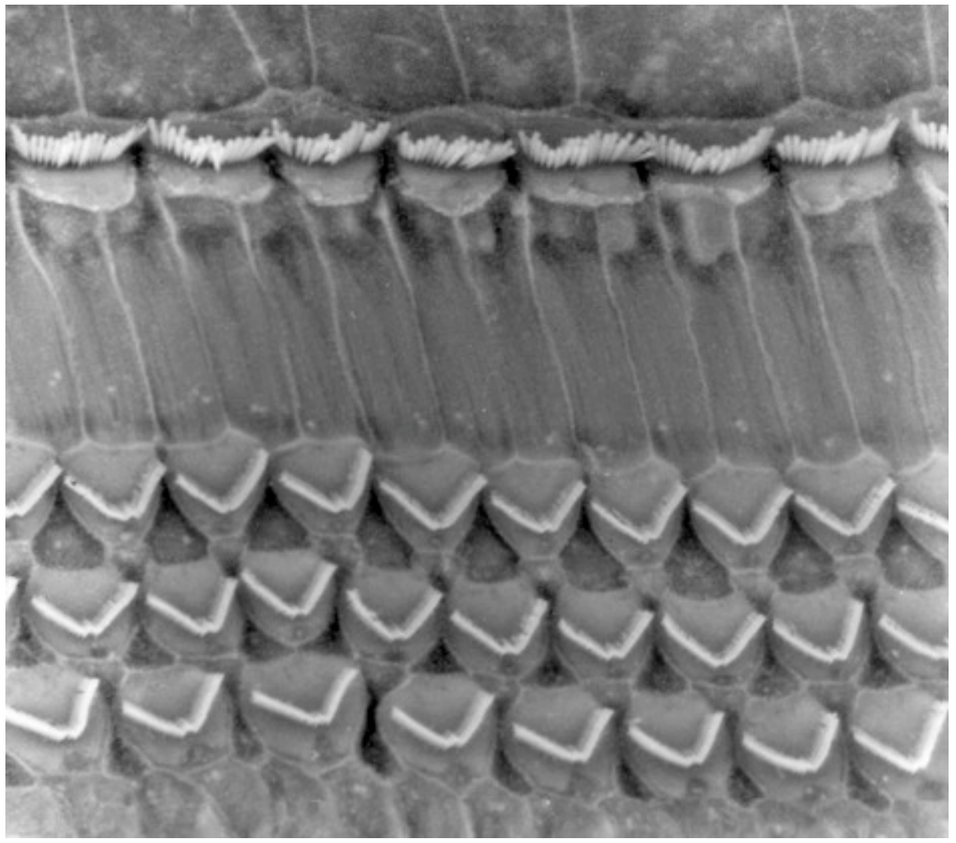
# Barkhausen-Bänder




1 Bark = 1,3 mm Basilarmembran-Länge

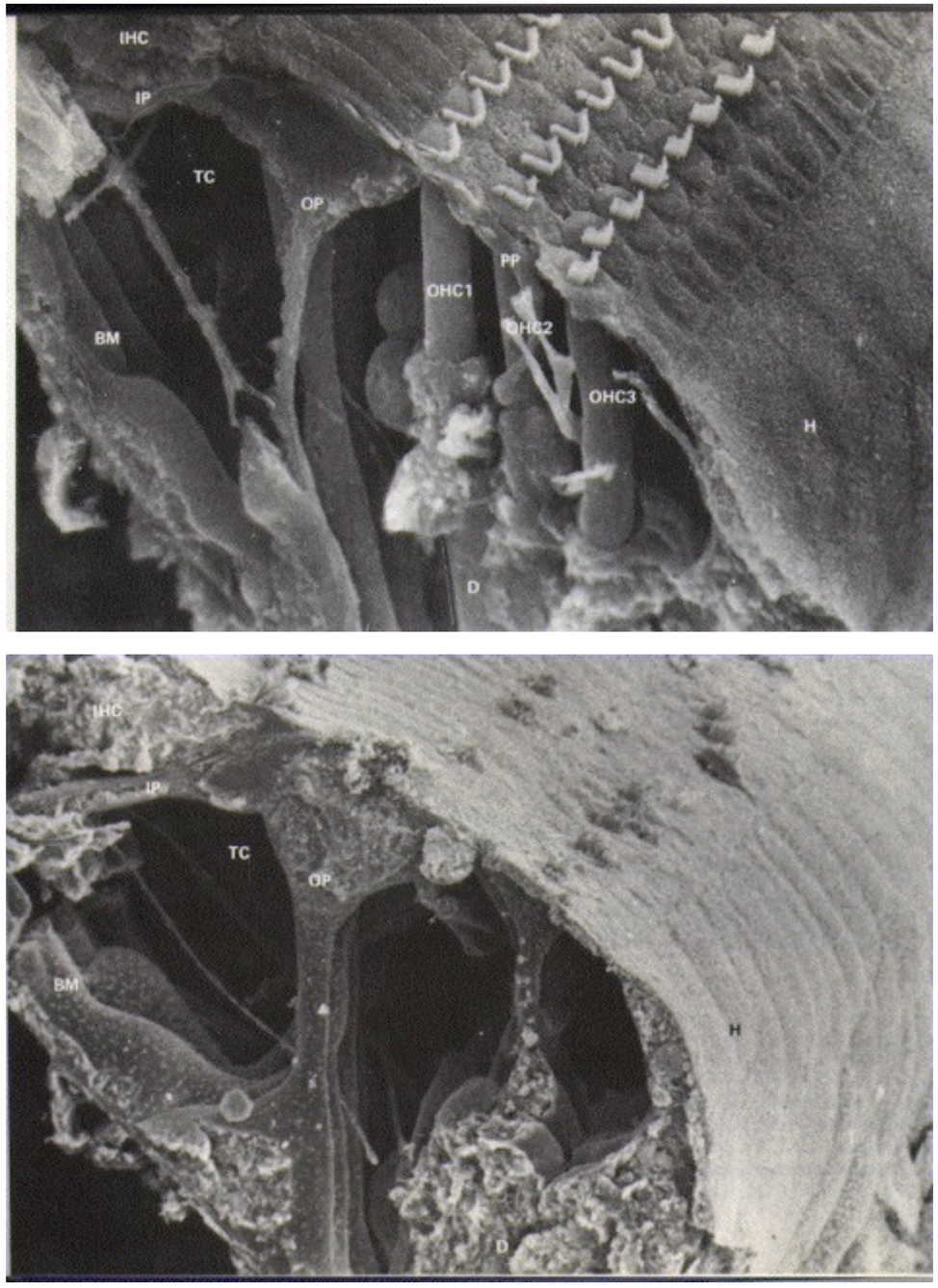
# Hahrzellen





3 Äußere, 1 Innere Reihe, 15.000 Zellen

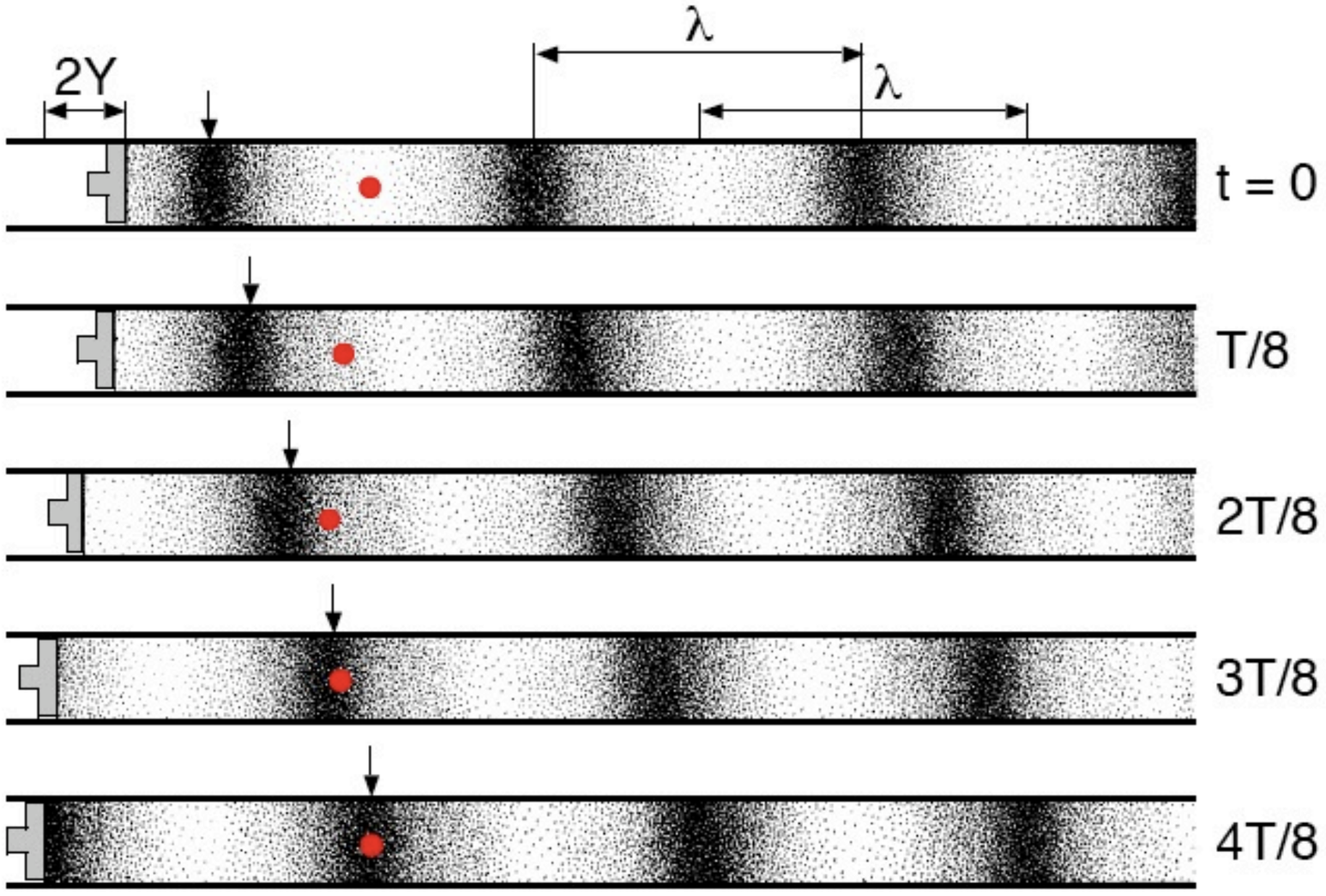
# Hörschäden

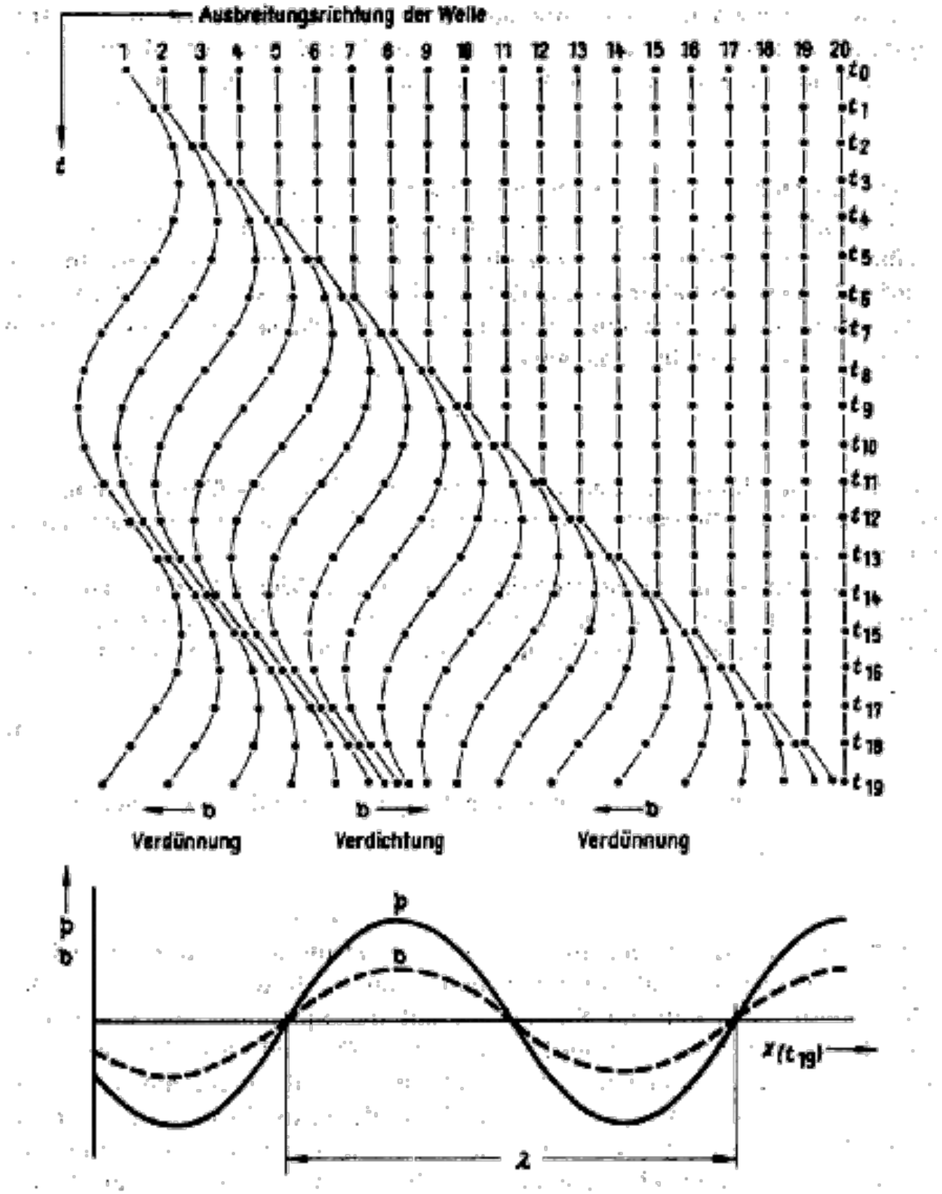



Intakte Cochlea (oben), zerstört nach Überstimulation (unten)

# Akustik

# Schallwellen



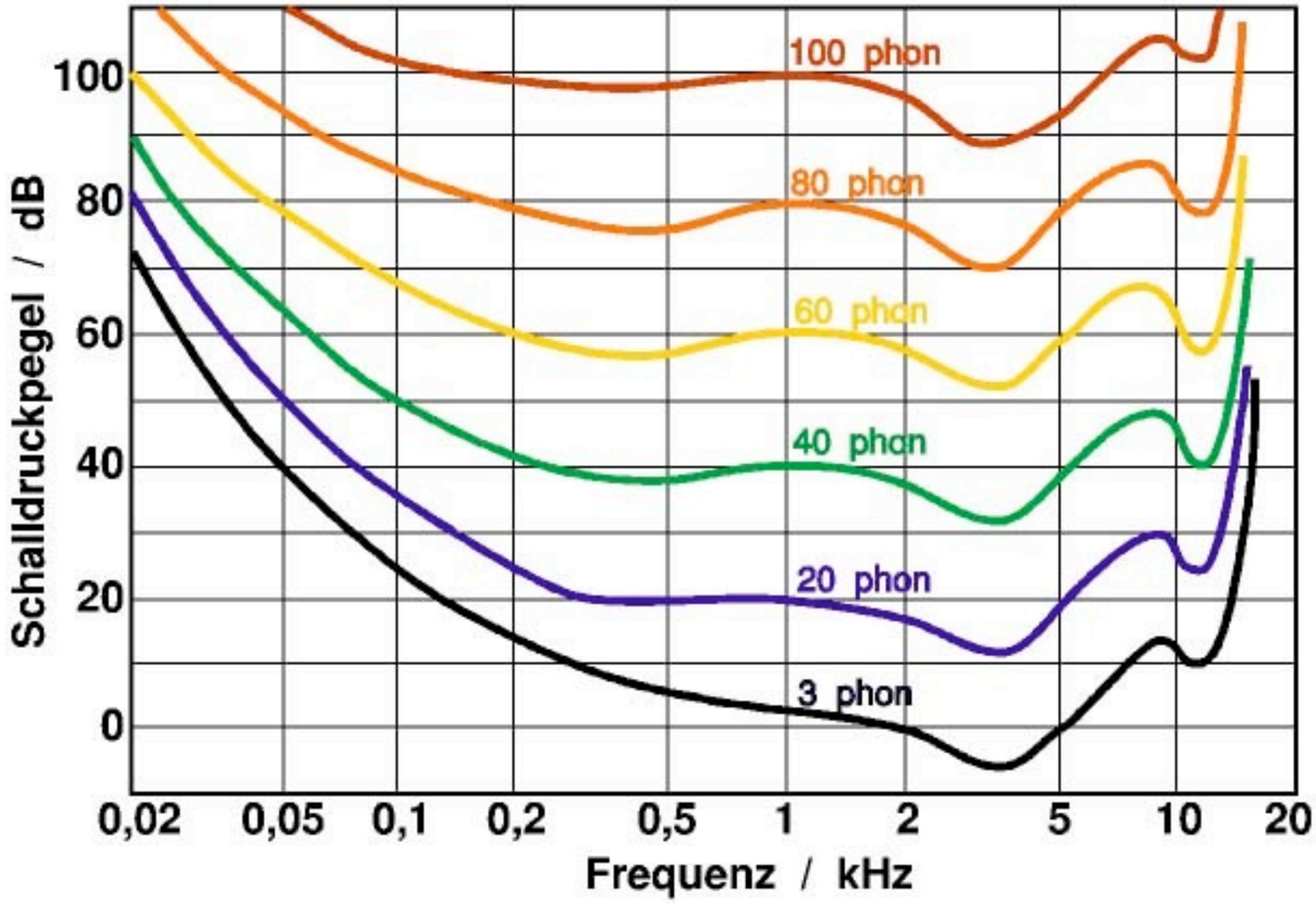



# Schalldruck

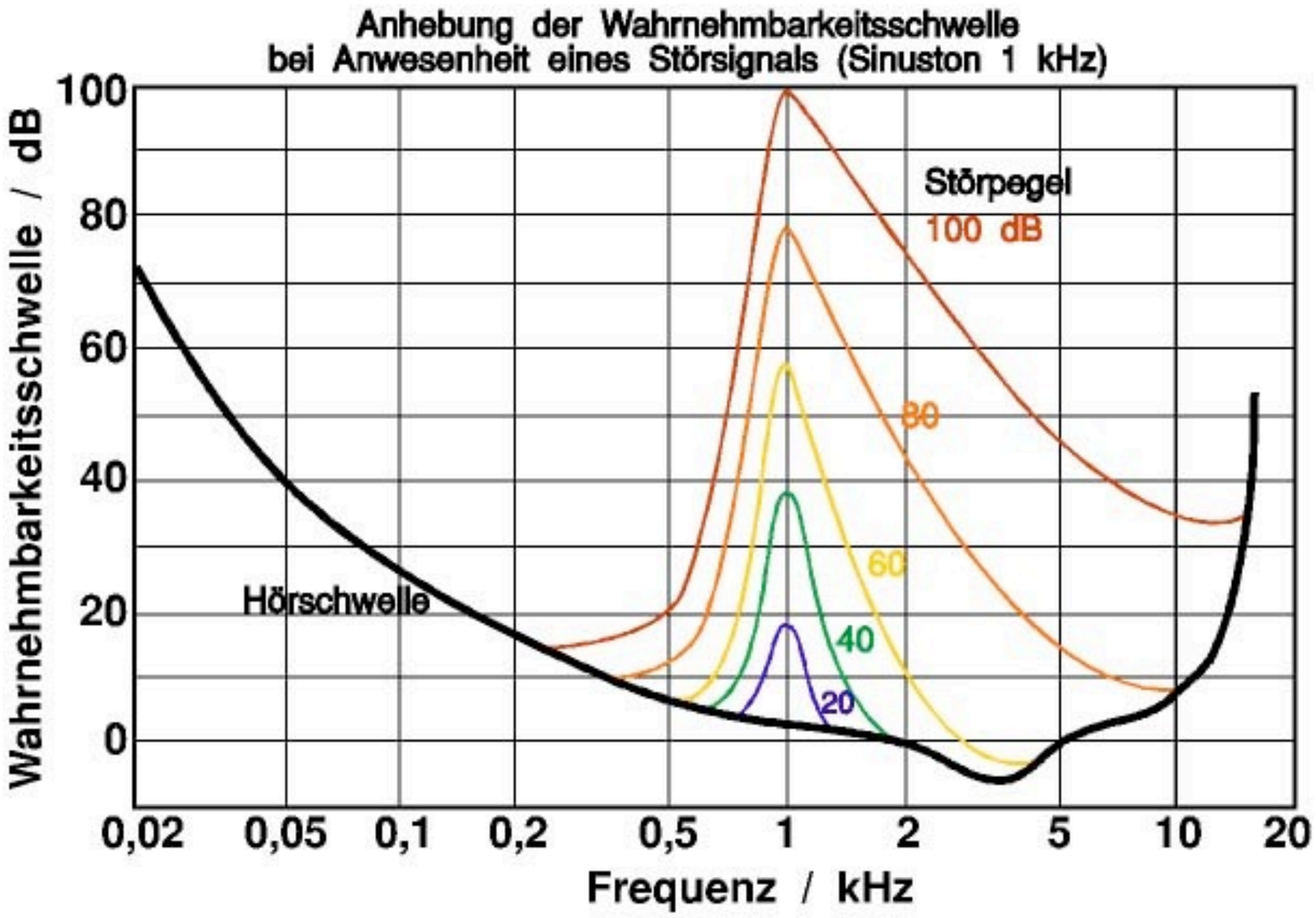
Druck ist das Verhältnis von Kraft zu Fläche, auf die diese Kraft senkrecht wirkt.
($Pa = N/m^2$)

Durch Schwingungen der Masseteilchen entstehen in der Schallwelle Druckschwankungen. Den Effektivwert dieses Druckes bezeichnet man als **Schalldruck p**. Er wird gemessen in $N/m^2$ bzw. in Mikrobar [µb]
(1 µb = 0,1 $N/m^2$).

# Hörschwelle

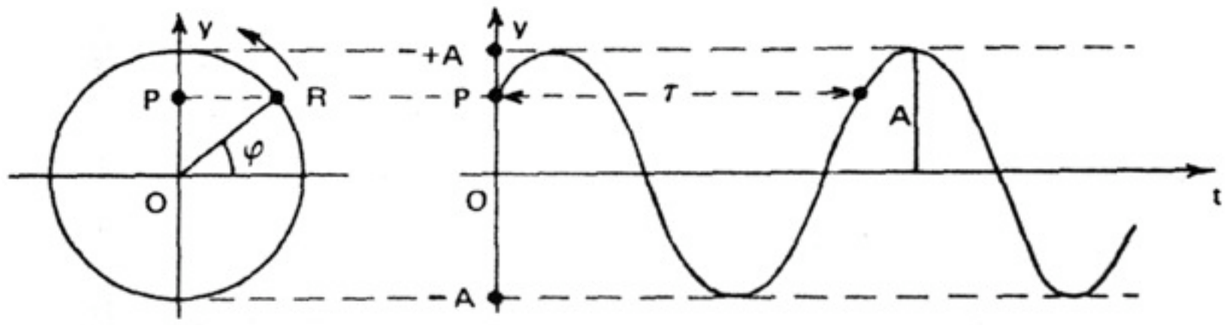


# Maskierung


Anhebung der Wahrnehmbarkeitsschwelle
bei Anwesenheit eines Störsignals (Sinuston 1 kHz)
Wahrnehmbarkeitsschwelle / dB
100
80
60
40
20
0
Störpegel
100 dB
80
60
40
20
Hörschwelle
0,02
0,05
0,1
0,2
0,5
1
2
5
10
20
Frequenz / kHz

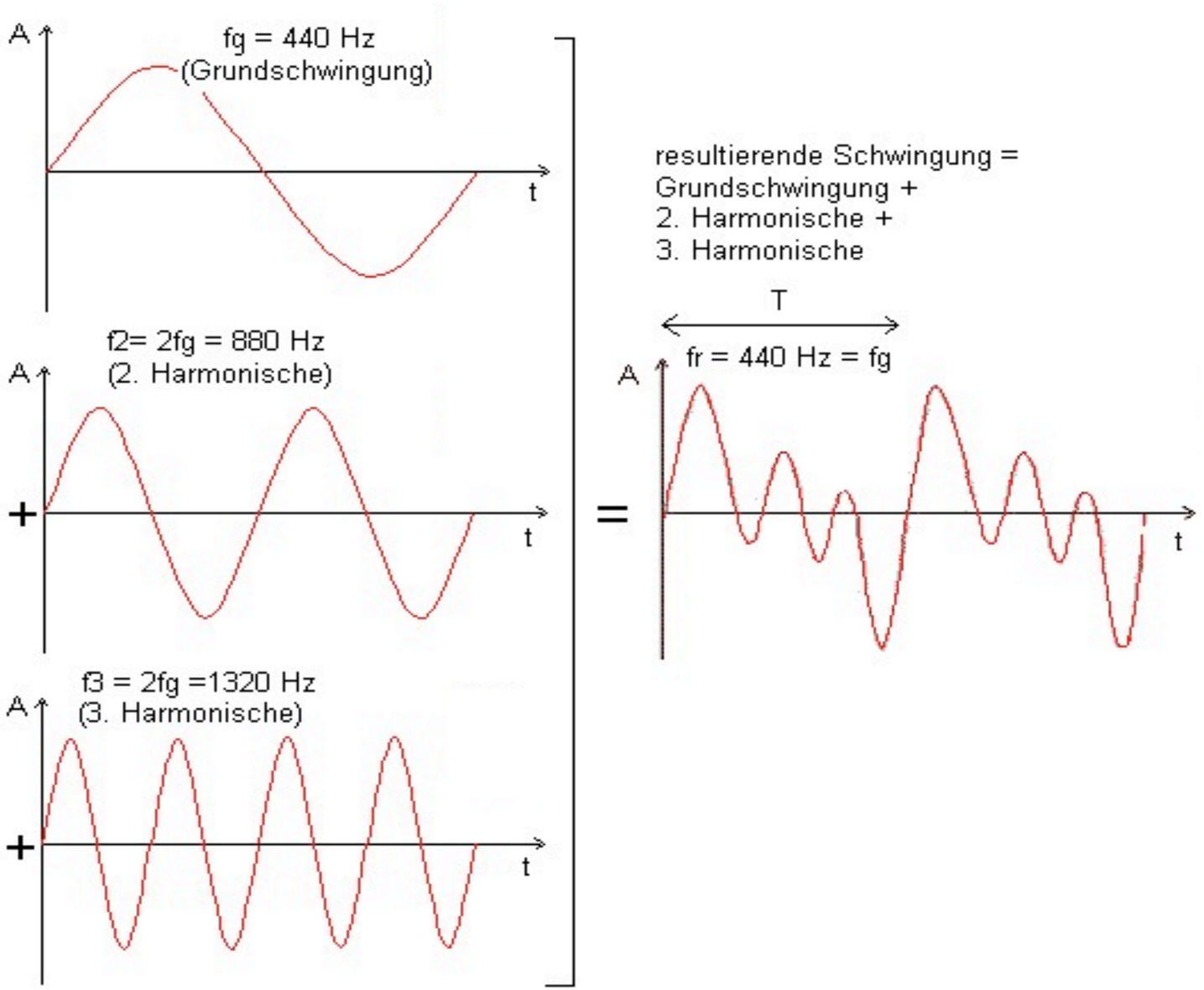
# Ton

Ein **Ton** ist eine Schallwelle mit einer einzigen Frequenz.



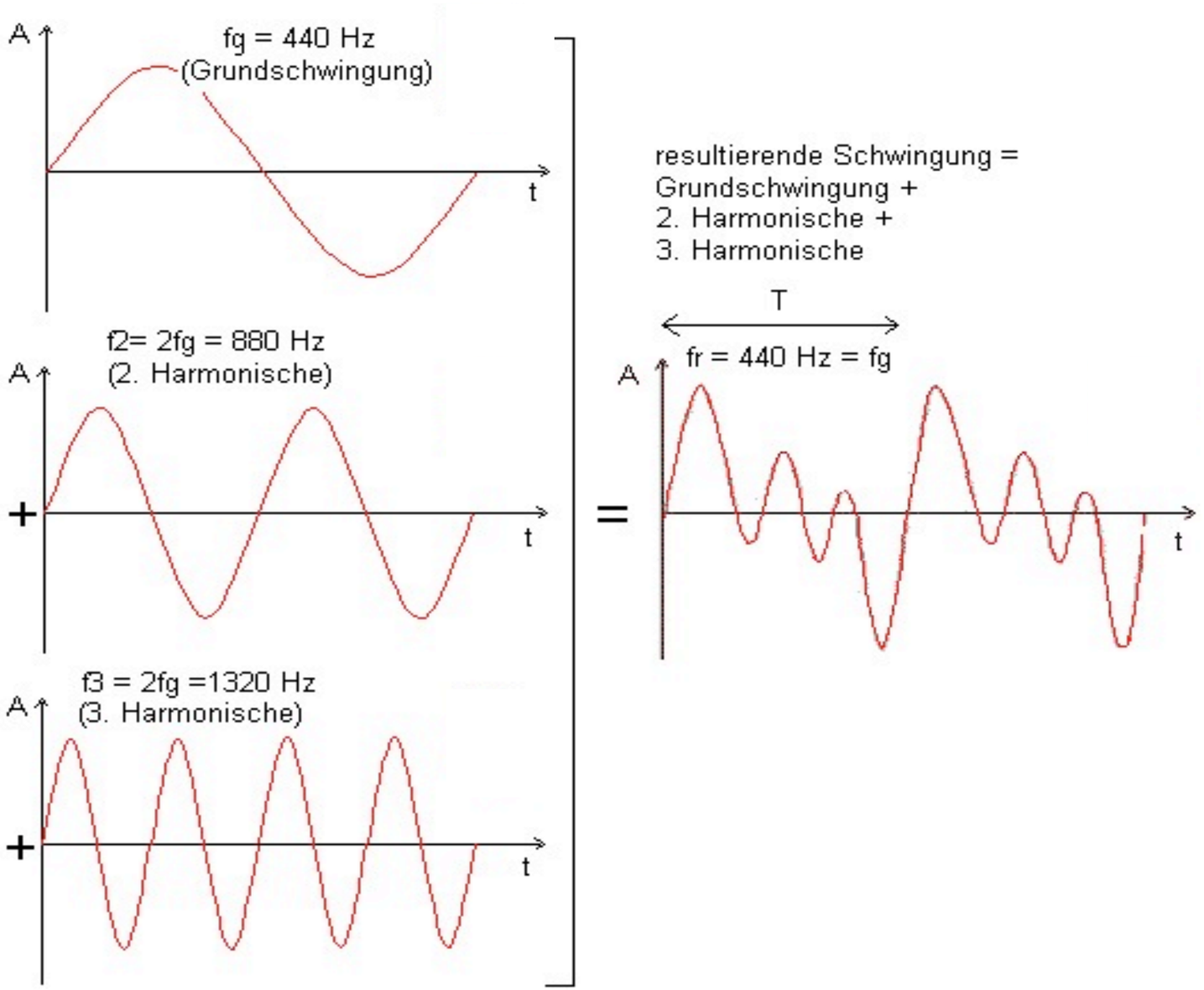

Schalldruck *p* eines Sinustons
$p(t) = A\sin(2 \pi f t)$

# Klang



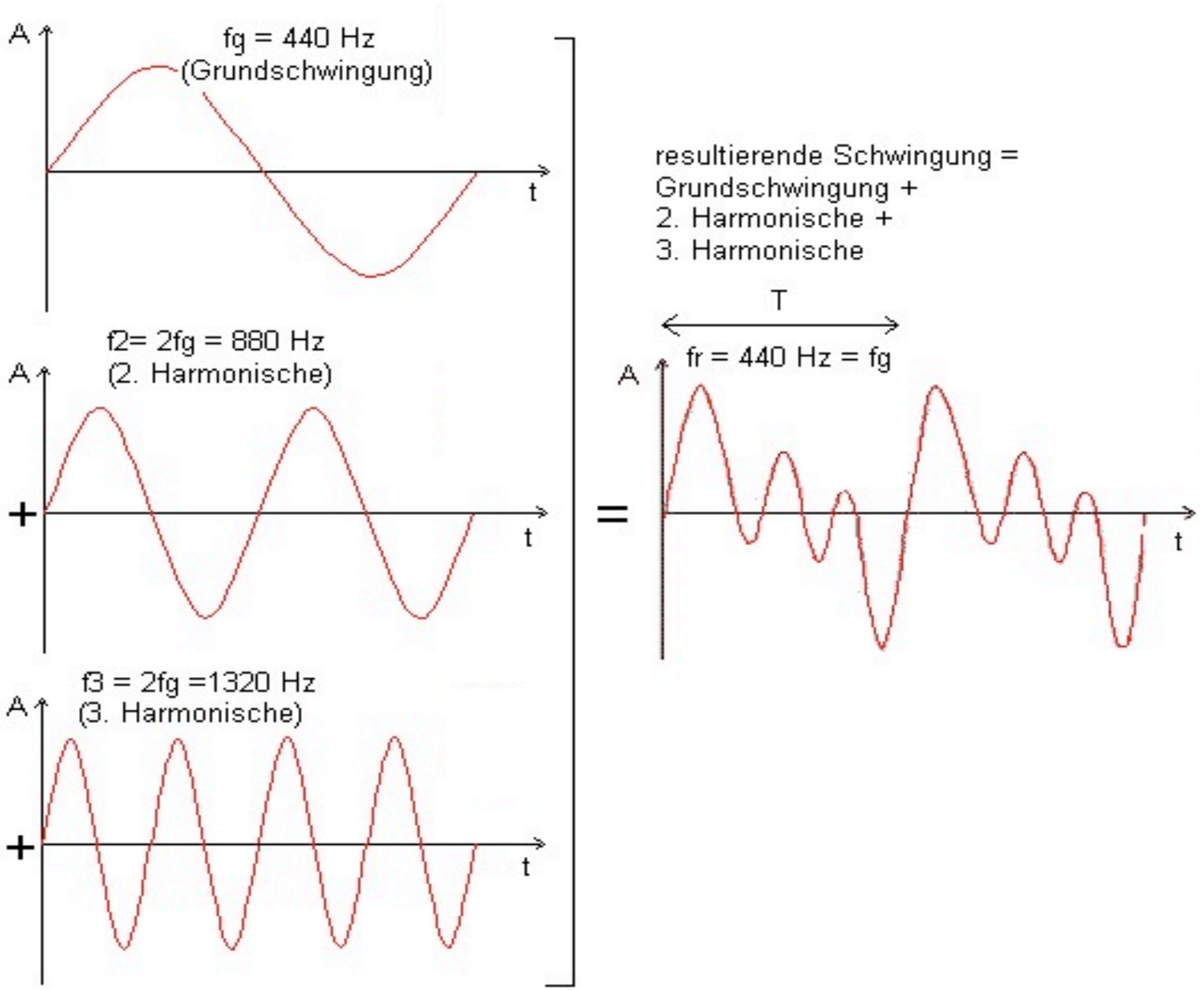

Ein Klang besteht aus einem Grundton und einer Anzahl von Obertönen, welche die Klangfarbe bestimmen.

# Klang



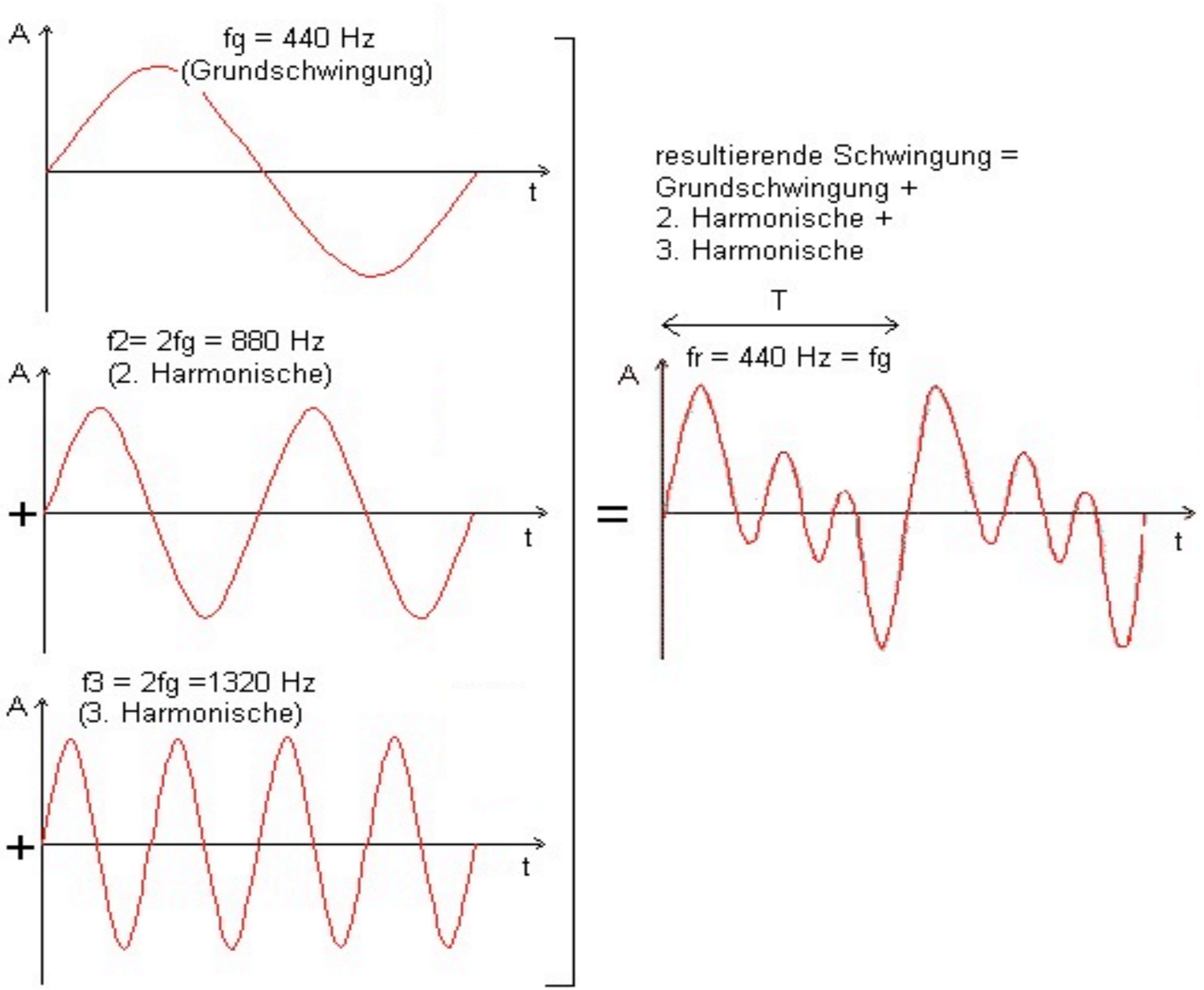

Ein Klang besteht aus einem Grundton und einer Anzahl von Obertönen, welche die Klangfarbe bestimmen.

# Klang



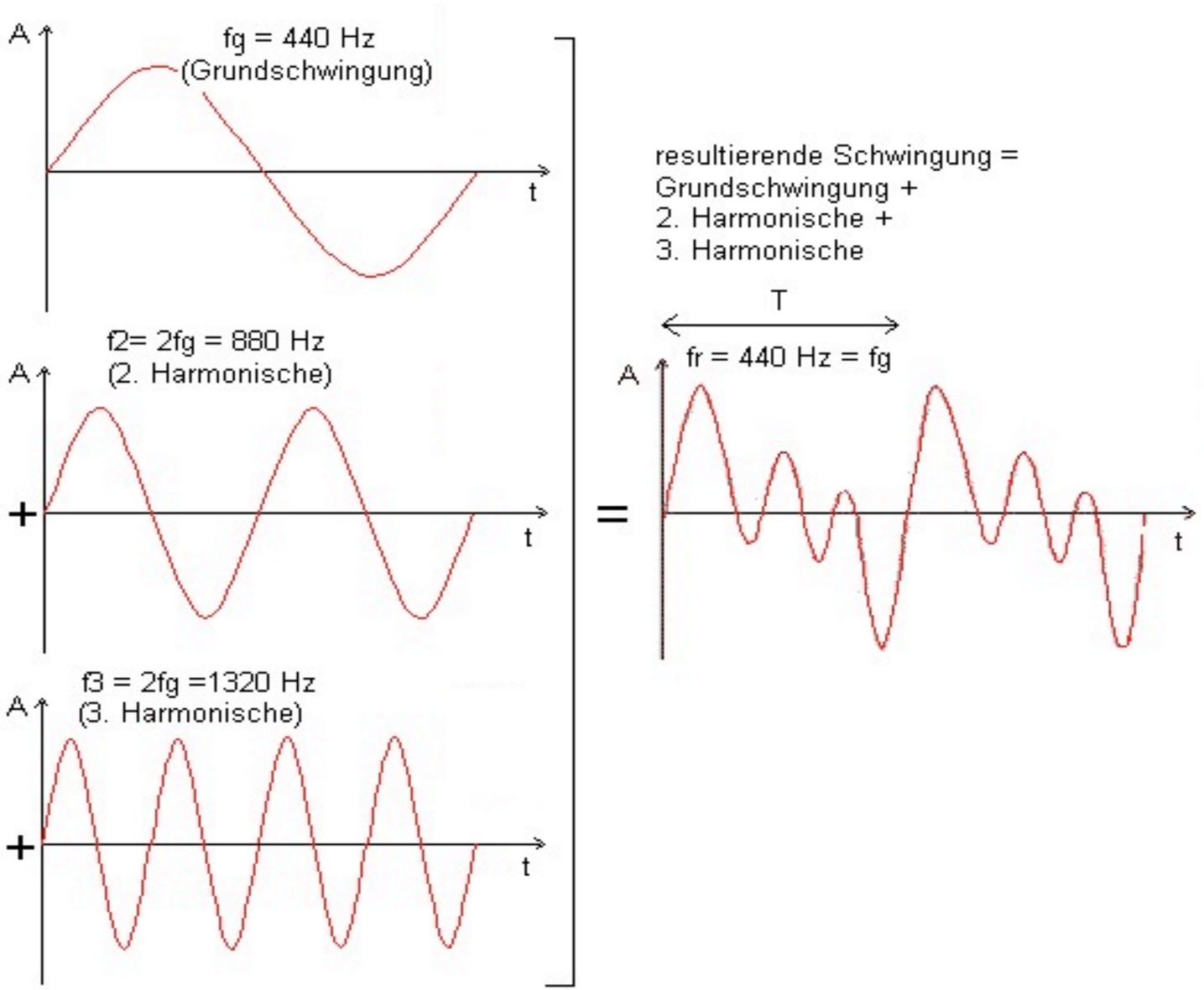

Ein Klang besteht aus einem Grundton und einer Anzahl von Obertönen, welche die Klangfarbe bestimmen.

# Klang




Ein Klang besteht aus einem Grundton und einer Anzahl von Obertönen, welche die Klangfarbe bestimmen.

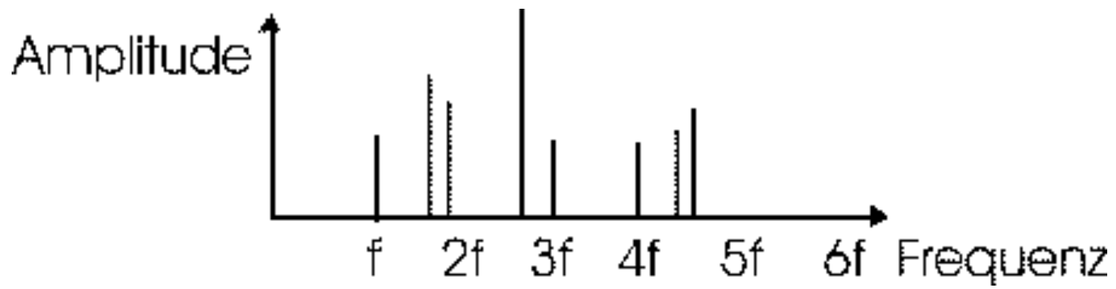
# Klang

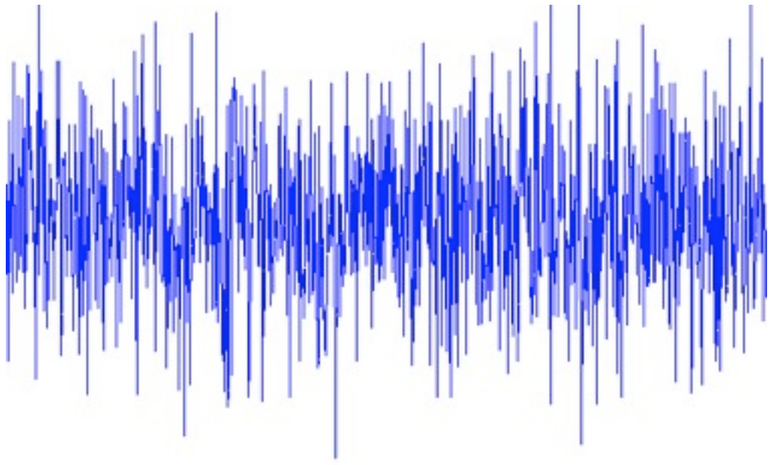



Ein Klang besteht aus einem Grundton und einer Anzahl von Obertönen, welche die Klangfarbe bestimmen.

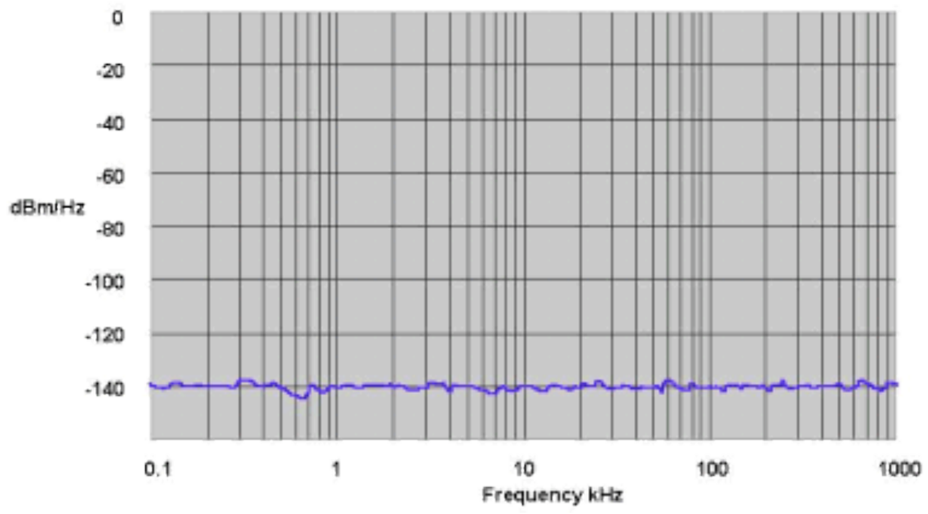
# Geräusch

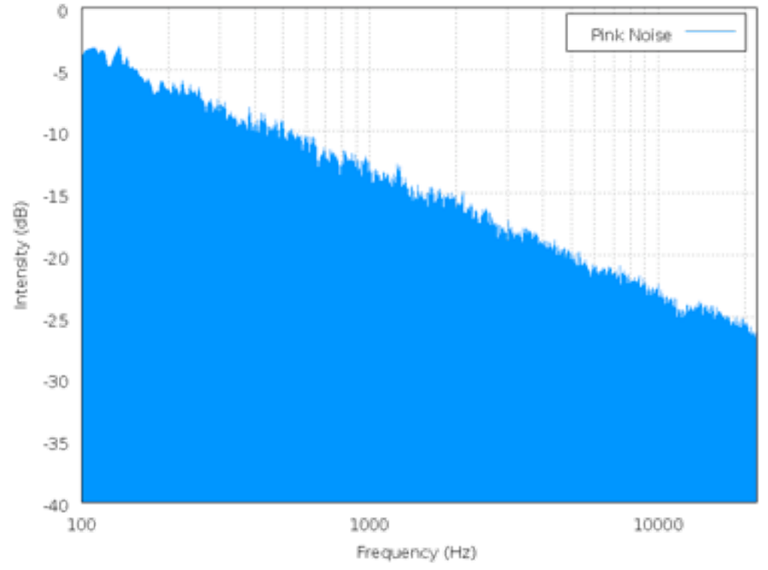
Ein **Geräusch** ist ein beliebiges Frequenzgemisch ohne erkennbare Obertonstruktur.

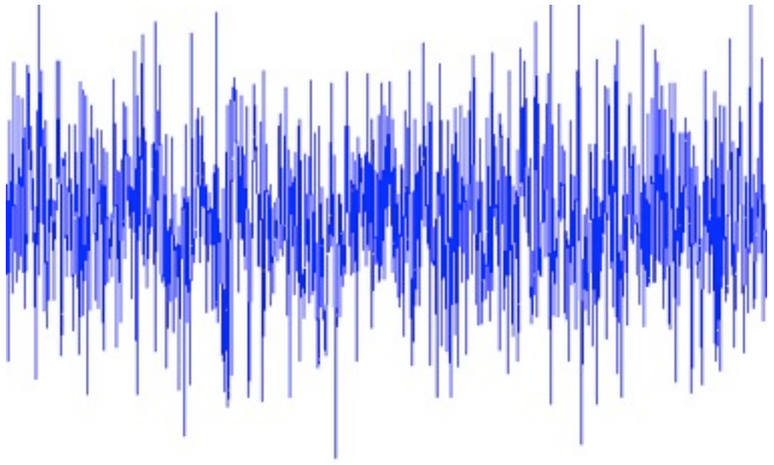




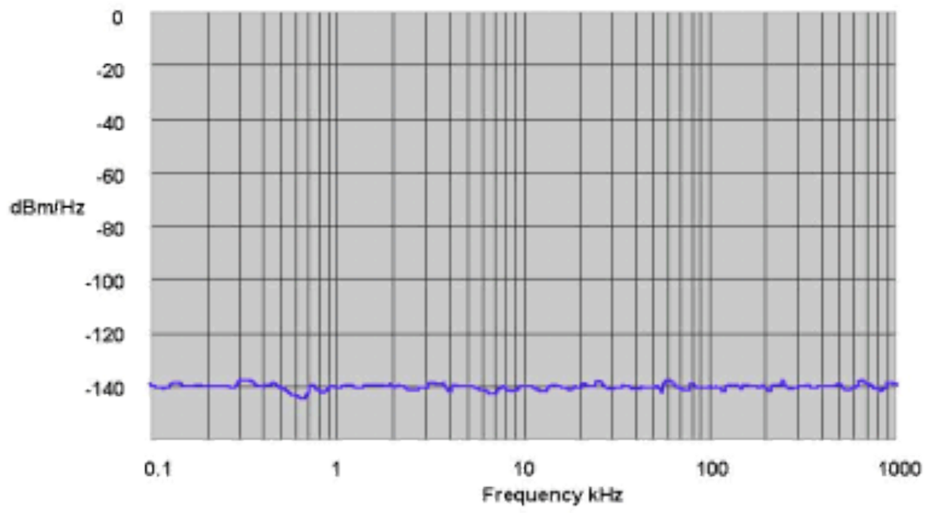
# Rauschen

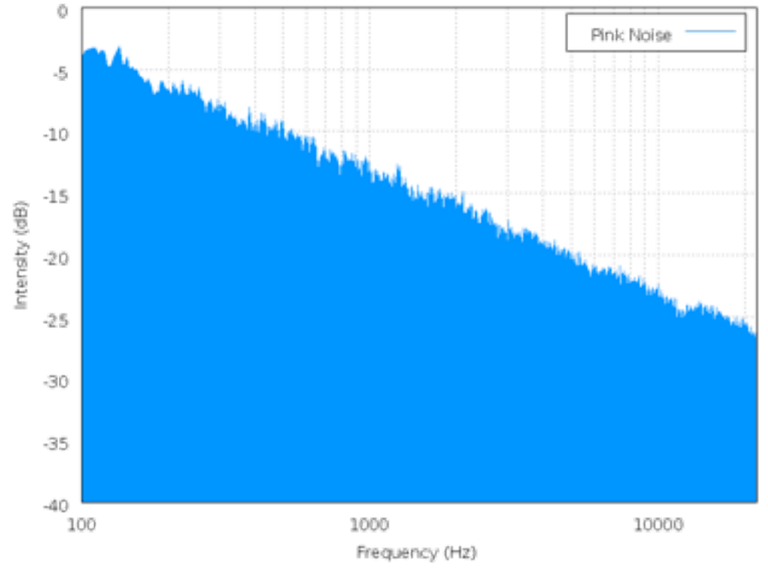

0
-20
-40
-60
-80
-100
-120
-140
dBm/Hz
0.1
1
10
100
1000
Frequency kHz

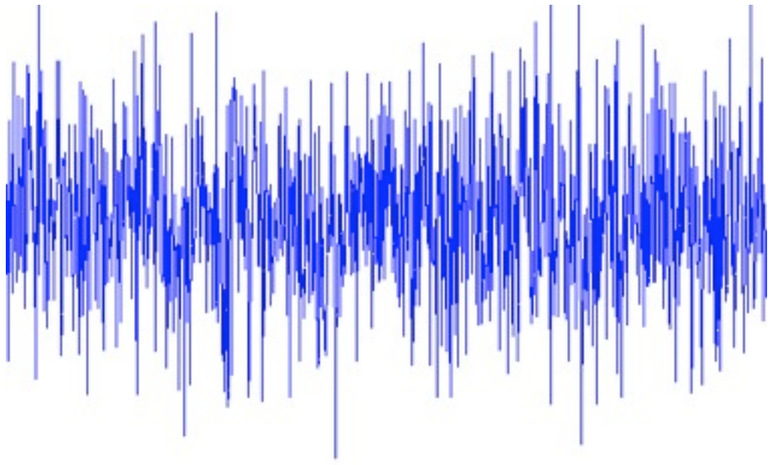

Pink Noise
0
-5
-10
-15
-20
-25
-30
-35
-40
Intensity (dB)
100
1000
10000
Frequency (Hz)



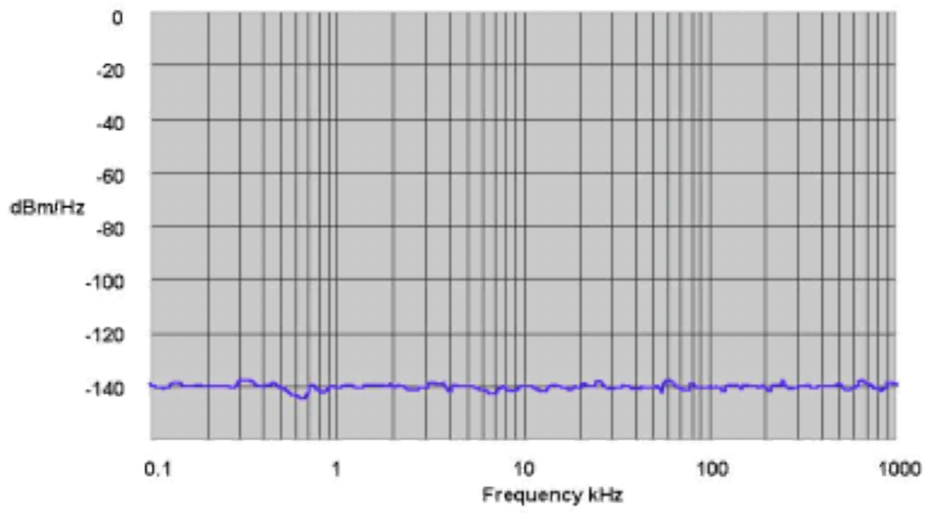
# Rauschen

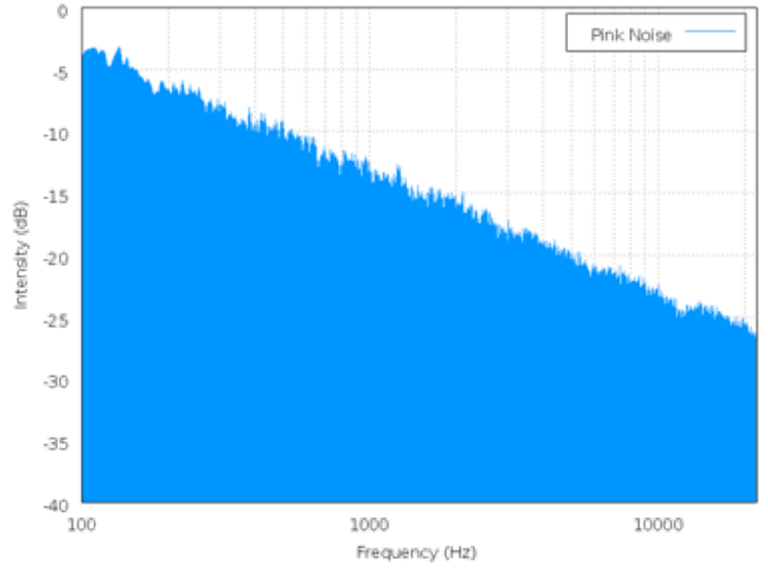

0
-20
-40
-60
-80
-100
-120
-140
dBm/Hz
0.1
1
10
100
1000
Frequency kHz


Pink Noise
0
-5
-10
-15
-20
-25
-30
-35
-40
Intensity (dB)
100
1000
10000
Frequency (Hz)



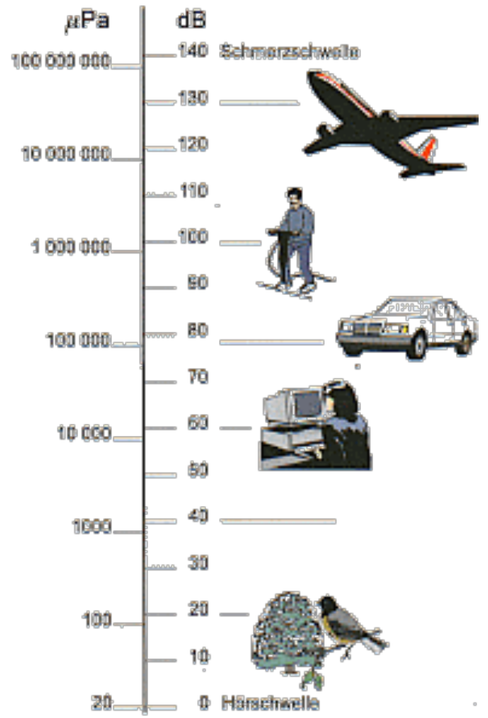
# Rauschen


0
-20
-40
-60
-80
-100
-120
-140
dBm/Hz
0.1
1
10
100
1000
Frequency kHz


Pink Noise
0
-5
-10
-15
-20
-25
-30
-35
-40
Intensity (dB)
100
1000
10000
Frequency (Hz)

# Hören

| Lärmquelle | Lärmpegel in dB-SPL |
|---|---|
| Schmerzgrenze | 130 |
| startendes Düsenflugzeug | 120 |
| Presslufthammer | 97 |
| Fabriken und Werkstätten | 50-75 |
| starker Straßenverkehr | 68 |
| normale Unterhaltung in 1 m Abstand | 65 |
| Bahnhofshalle, Stadion | 55-65 |
| großes Büro | 60-65 |
| Ladengeschäft | 45-60 |
| Autowerkstatt | 55 |
| Auto bei 80 km/h | 45-50 |
| Hotel, Zimmer | 42 |
| Wohnung in einer Großstadt | 40 |
| Wohnung auf dem Land | 30 |
| leerer Hörsaal | 25-35 |
| leere Kirche | 30 |
| Klassenzimmer ohne Schüler | 30 |
| Fernsehstudio ohne Publikum | 25-35 |
| Aufnahmestudio | 20-30 |
| Flüstern | 10-20 |
| Hörschwelle | 0-5 |

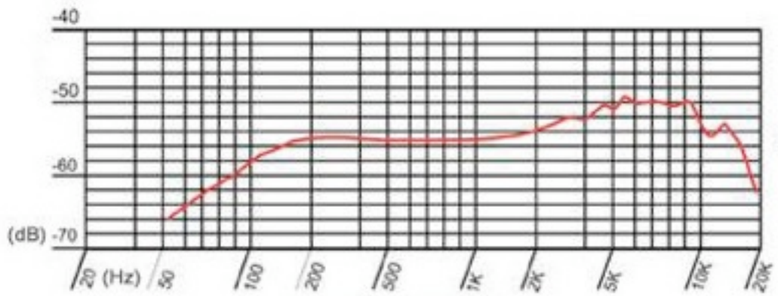



Lärm:

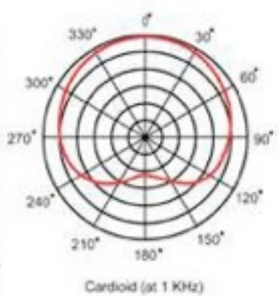
Zwischen Hörgrenze und Schmerzgrenze

# Aufnahme und Kompression

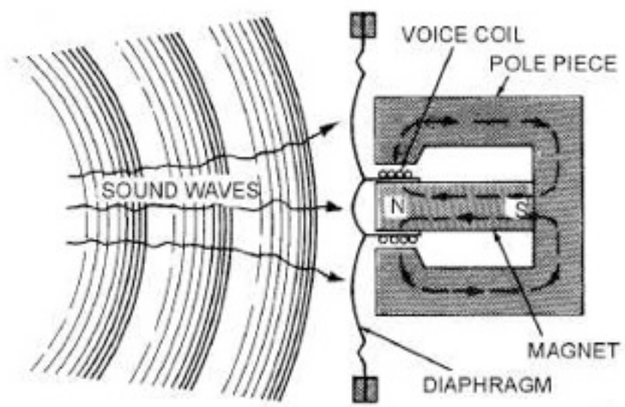
# Aufnahme: Mikrofone

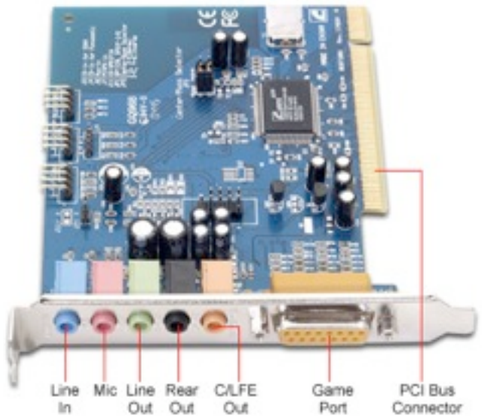


Frequency Response Graph



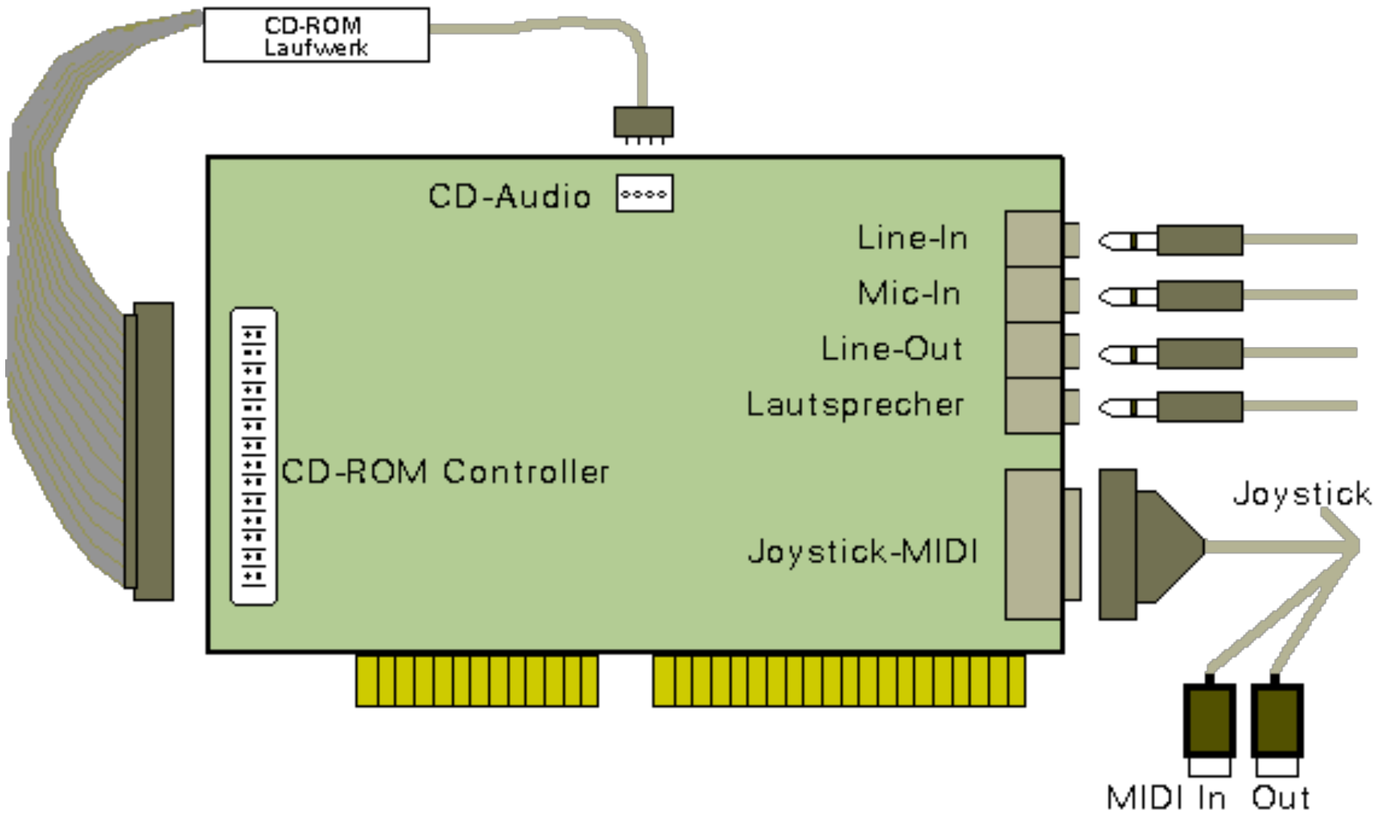
Cardioid (at 1 KHz)





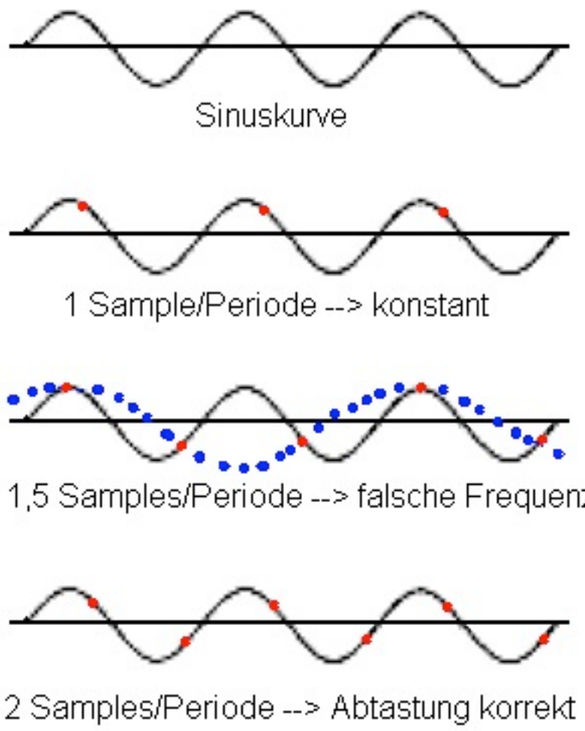

# Digitalisieren: Soundkarte

**Schematische Darstellung einer Soundkarte:**




Line: 500 mV - 2 V
Mic: 5 mV - 50 mV

# Nyquist-Theorem

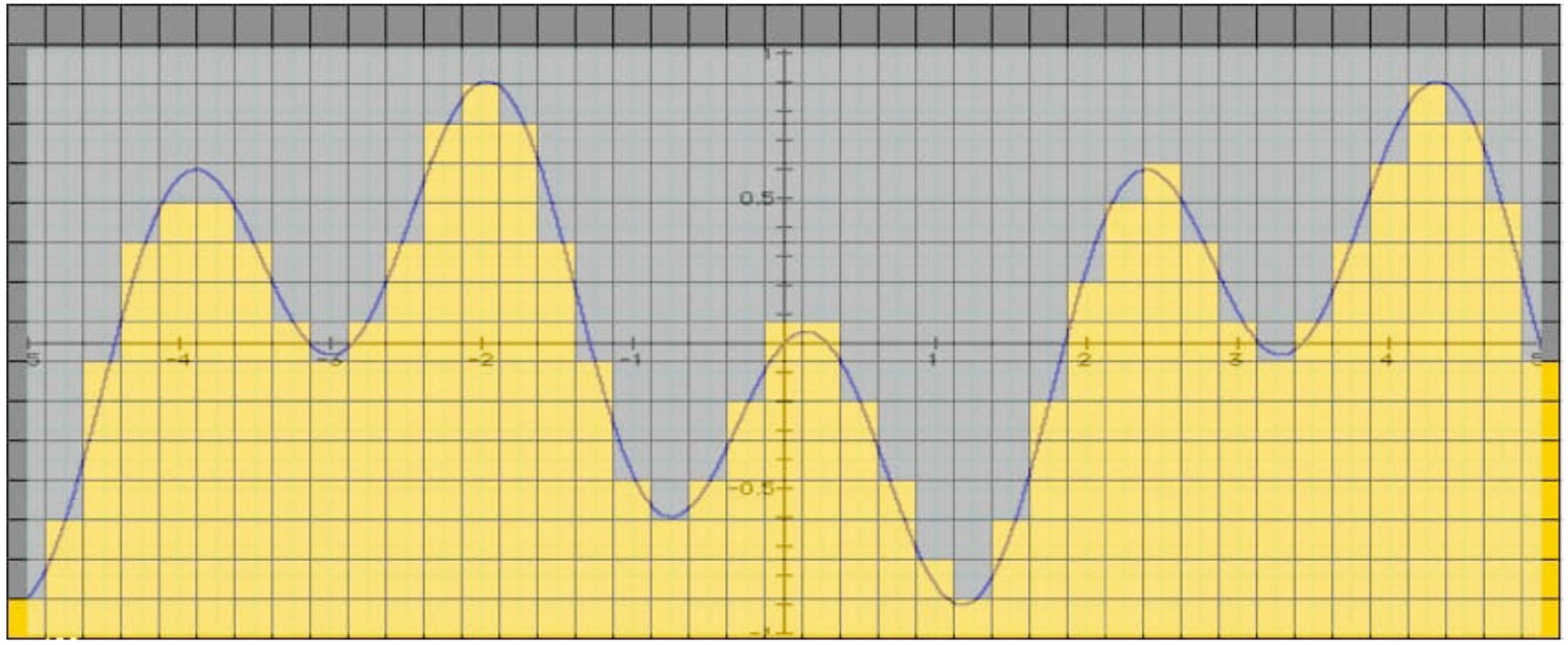



## Abtasttheorem

Ein Signal der Maximalfrequenz f muss mit mehr als 2*f abgetastet werden, damit aus den Samples das Ausgangssignal vollständig rekonstruiert werden kann.

Liegen Samples mit der Abtastfrequenz g vor, dann lassen sich daraus lediglich Signalfrequenzen bis maximal g/2 rekonstruieren.

# Pulskode: Amplitudenmodulation




1011 1011 1010 1000 0111 1000 1010 1101 1110 1101 1010 0111 0100 0111 0100 0110 1000 1000 0110 0100 0010 0001 0011 0110 1001 1011 1100 1010 1000 0111 0100 1010 1100

11 11 10 08 07 08 10 13 14 13 10 07 04 03 04 06 08 08 06 04 02 01 03 06 09 11 12 10 08 07 08 10 12

# Digitale Tonformate

Audio CD
.IFF, .AIF, .AIFF Audio Interchange File Format

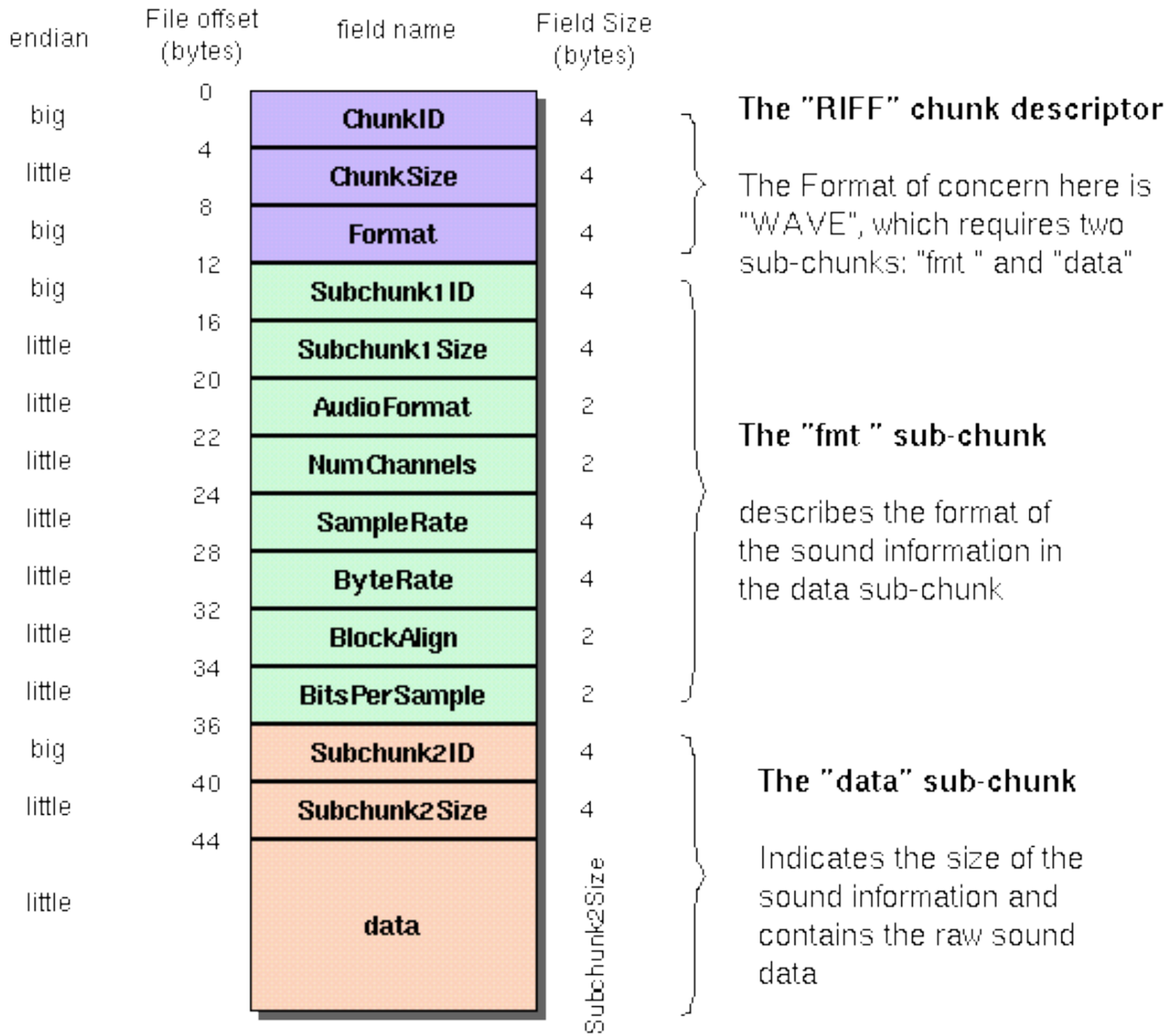
Proprietäre Formate
.WAV Windows
.RAM Real Player
.AU SUN/Unix

Verlustfreie Audiokompression
Flac - Free Lossless Audio Codec

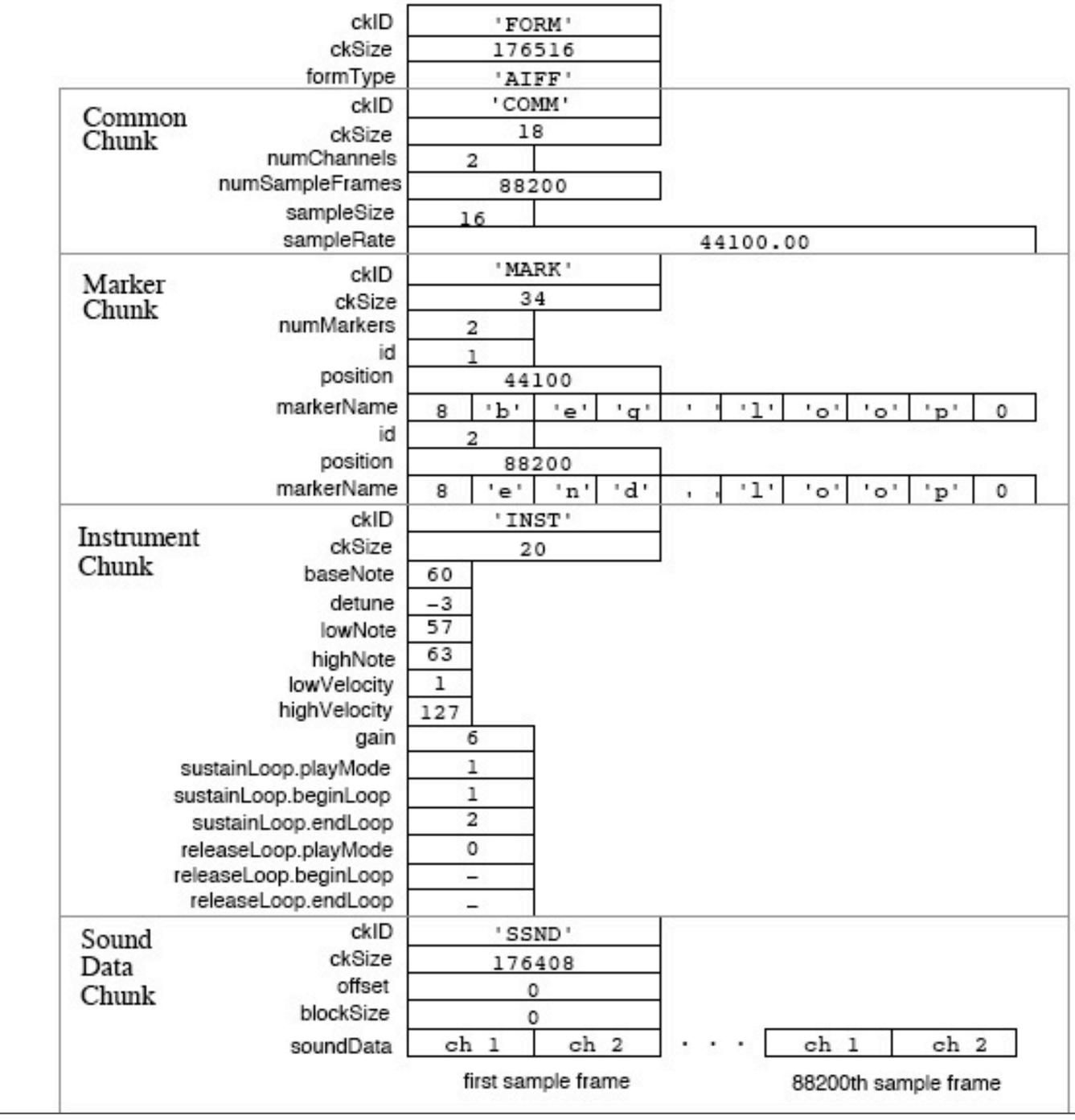
Offene Kompressor-Standards
.MP3 MPEG-1, Layer III (1988/1994)
.AAC MPEG-2 Advanced Audio Coding (1997
.ogg Ogg-Vorbis (1993/2002)

Digitale Notation
.MID, .MIDI Musical Instruments Digital Interface

# *The Canonical WAVE file format*



# AIFF




Contains information about the sampled sound such as the sampling rate and sample size.

Contains markers that point to positions in the sound data

Defines basic parameters that an instrument (such as a sampling keyboard) can use to play back the sound data

Contains the sample frames that comprise the sampled sound

# Verlustfreie Kompression (auch für Tondaten möglich)

| program | % size | compress speed | decompress speed |
|---|---|---|---|
| UNIX compress | 74.0 | 5.1 | 15.0 |
| UNIX pack | 69.8 | 16.1 | 8.0 |
| GNU gzip | 66.0 | 2.2 | 17.2 |
| shorten default (fast) | 42.6 | 13.4 | 16.1 |
| shorten LPC (slow) | 41.7 | 5.6 | 8.0 |
| aifc[de]compress | lossy | 2.3 | 2.2 |

# Free Lossless Audio Codec – FLAC

| Source Available? | Plugins Available? | Hardware Support? | Streamable? | Seekable? | Cost | OS support |
|---|---|---|---|---|---|---|
| YES (OSI approved license) | YES (XMMS, AlsaPlayer, Winamp, MacAmp Lite, dBpowerAMP, Foobar2000, QCD, Apollo) | YES (PhatBox, Kenwood MusicKeg, Rio Receiver, Dell Digital Audio Receiver, Turtle Beach AudioTron) | YES | YES | FREE | Linux, Windows, Mac OS X, *BSD, Solaris, OS/2, BeOS, others |

Reduktion 1:2 bis 1:3, langsam

# Free Lossless Audio Codec

**Blockung:** Die Eingabedaten werden in Blöcke unterschiedlicher Länge, abhängig von Abtastrate, Spektralverteilung in der Zeit u.a. abhängen.

**Kanalabhängigkeiten:** Alle Kanäle sollen verlustfrei erhalten bleiben. Bei Stereo wird typischerweise eine Summe (ein Durchschnittswert) und eine Differenz der Kanäle gebildet.

**Prädiktion:** Es werden feste Kodierer und Dekodierer vereinbart. Das Signal eines Blockes wird mit einem festen parametrisierten Verfahren grob angenähert. Diese Parameter, die im Regelfall deutlich kompakter als das Signal sind, werden übertragen.

**Korrekturkodierung:** Die Differenz zwischen Eingabeblock und seiner Prädiktion wird berechnet und verlustfrei kodiert.

Ziel ist es, dass Prädikation und Korrekturkodierung  kleiner als das Eingabesignal wird.

# Verlustfreie Codecs

| | Encode time | | Decode time | | Compression | |
|---|---|---|---|---|---|---|
| **Codec** | Total | CPU | Total | CPU | Size | **Avg.ratio** |
| La 0.3c | 150:12.30 | 146:54.11 | 144:50.36 | 142:49.41 | 375.76 MB | 49.86% |
| Tak 1.01 (extra high max) | 66:52.15 | 63:48.91 | 7:39.41 | 4:46.01 | 383.70 MB | 50.60% |
| Monkey's Audio 3.99 (insane) | 129:07.86 | 126:05.33 | 135:13.19 | 133:12.71 | 381.79 MB | 50.65% |
| Tak 1.01 (extra high) | 25:14.00 | 21:48.45 | 7:50.46 | 4:54.86 | 384.06 MB | 50.66% |
| Monkey's Audio 3.99 (extra high) | 28:34.03 | 25:12.14 | 29:38.88 | 26:53.34 | 384.55 MB | 50.87% |
| optimFROG 4.21 (mode 4 @ 1x) | 338:34.15 | 335:12.47 | 339:18.57 | 336:25.54 | 386.22 MB | 51.06% |
| optimFROG 4.21 (mode 3 @ 4x) | 58:21.94 | 54:52.62 | 59:34.12 | 56:51.21 | 386.71 MB | 51.15% |
| optimFROG 4.21 (mode 2 @ 4x) | 32:20.46 | 28:36.38 | 33:39.96 | 30:49.89 | 387.93 MB | 51.33% |
| Tak 1.01 (normal) | 11:42.59 | 8:12.00 | 6:36.47 | 3:38.14 | 391.16 MB | 51.39% |
| optimFROG 4.21 (mode 1 @ 4x) | 24:25.04 | 20:37.96 | 25:48.90 | 22:52.76 | 389.04 MB | 51.52% |
| Monkey's Audio 3.99 (high) | 15:26.16 | 12:03.16 | 16:30.00 | 13:40.59 | 389.83 MB | 51.53% |
| Monkey's Audio 3.99 (normal) | 13:20.15 | 9:56.40 | 14:24.02 | 11:33.71 | 393.17 MB | 51.97% |
| WavPack 4.41 (extra high -x) | 27:23.56 | 23:29.26 | 10:58.90 | 7:48.32 | 396.56 MB | 52.22% |
| optimFROG 4.21 (mode 0 @ 4x) | 16:36.98 | 12:51.58 | 17:55.55 | 14:58.99 | 394.69 MB | 52.24% |
| Tak 1.01 (turbo) | 7:25.21 | 3:51.02 | 6:16.82 | 3:10.87 | 399.97 MB | 52.71% |
| WavPack 4.41 (high) | 11:48.47 | 7:45.58 | 9:19.07 | 6:05.35 | 399.90 MB | 52.73% |
| Monkey's Audio 3.99 (fast) | 10:24.29 | 6:58.46 | 11:32.07 | 8:37.81 | 400.57 MB | 53.11% |
| WavPack 4.41 (normal -x) | 14:14.54 | 10:13.33 | 7:23.14 | 4:24.26 | 403.10 MB | 53.19% |
| FLAC 1.2.1 (-8) | 26:46.72 | 23:21.85 | 5:40.25 | 2:36.47 | 404.23 MB | 53.36% |
| WavPack 4.41 (normal) | 9:48.59 | 5:46.91 | 7:37.26 | 4:30.11 | 405.84 MB | 53.56% |
| FLAC 1.2.1 (-5, default) | 10:07.41 | 6:35.68 | 5:23.16 | 2:22.41 | 406.25 MB | 53.67% |
| WavPack 4.41 (fast -x) | 11:54.47 | 7:50.84 | 6:32.50 | 3:25.77 | 411.52 MB | 54.39% |
| FLAC 1.2.1 (-3) | 7:23.77 | 3:47.42 | 5:31.15 | 2:19.07 | 412.42 MB | 54.57% |
| WavPack 4.41 (fast) | 8:52.27 | 4:47.74 | 6:33.73 | 3:28.19 | 415.05 MB | 54.92% |
| Apple Lossless (iTunes 4.5) | 19:53.27 | 19:53.27 | 10:01.86 | 10:01.86 | 414.45 MB | 54.96% |
| Bonk 0.5 | 51:45.58 | 48:32.10 | 42:02.76 | 39:05.43 | 418.65 MB | 55.43% |
| FLAC 1.2.1 (-1) | 6:24.51 | 2:42.93 | 5:26.87 | 2:17.49 | 431.72 MB | 56.97% |
| Shorten 3.2a (-p0 -b256, default) | 10:01.38 | 6:23.40 | 6:38.43 | 3:30.66 | 433.56 MB | 57.29% |
| Shorten 3.2a (-p8 -b2048) | 12:09.68 | 8:31.46 | 7:34.17 | 4:24.09 | 438.86 MB | 58.11% |
| RIFF WAVE | 73:44.94 | | 73:44.94 | | 780.56 MB | 100.00% |

# Audio CD

Audio CDs speichern unkomprimierte Stereosignale auf

2 Tonspuren mit 44,1 KHz Abstastrate, womit ein Tonumfang von etwa 20Hz bis 20000 Hz darstellbar ist.

Der Ursprung für diese Zahlen liegt in der Sony U-Matic:
294 (PAL-Halbbilder) * 50 (pro Sekunde) * 16 (Bit Auflösung) = 1.411.200 Bit/s
/32 (16 Bit Lautstärkeauflösung bei 2 Kanälen) = 44.100 Bit/s

Datenmenge (Bit) = Auflösung (Bit) * Abtastrate (Hz) * Kanäle * Zeit

1 Minute ~ 10,1 MB

Das gespeicherte Signal ist in mehreren Ebenen fehlerkorrigierend gespeichert.

# Tonkompression am Beispiel MPEG 1 Audio Layer III (MP3)

CD audio signals

44,1 kHz Abtastfrequenz

16 bit Amplituden Auflösung

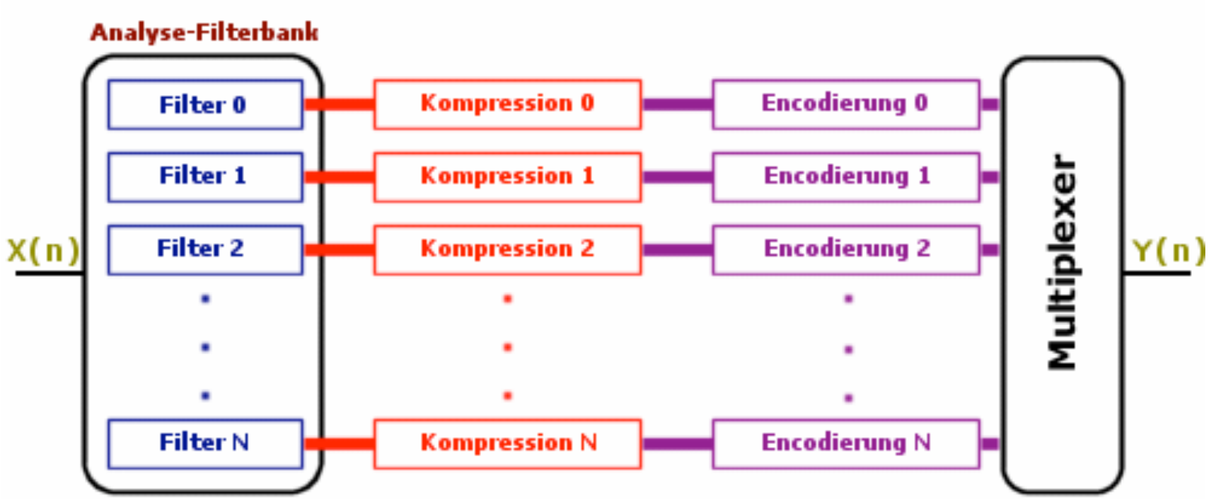
1,4112 Mbit/s Stereosignal

»By using MPEG audio coding, you may shrink down the original sound data from a CD by a factor of 12, without losing sound quality.

Factors of 24 and even more still maintain a sound quality that is significantly better than what you get by just reducing the sampling rate and the resolution of your samples.«

**FhG Institut für Integrierte Schaltungen, Uni Erlangen**

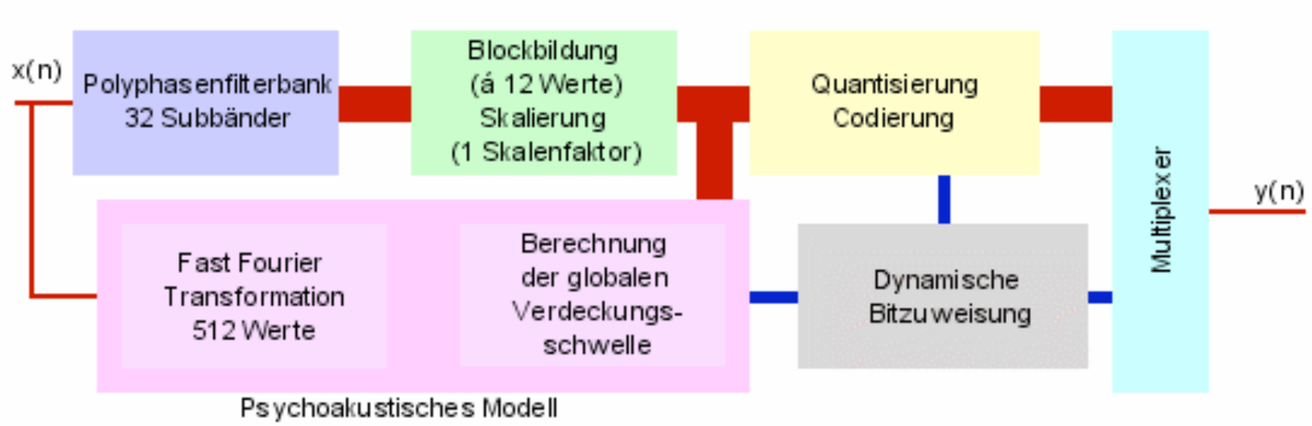
# Subband-Codierung




**Aufbau eines Subband Encoders**

Zunächst durchläuft das Audioeingangssignal eine Analysefilterbank. Diese besteht aus einer bestimmten Anzahl an Bandpassfiltern, die das Eingangssignal X(n) in Frequenzbänder aufspalten. Nach dieser Aufspaltung kann jedes Subband einzeln komprimiert und encodiert werden.

# Audio Layer 1: MP1

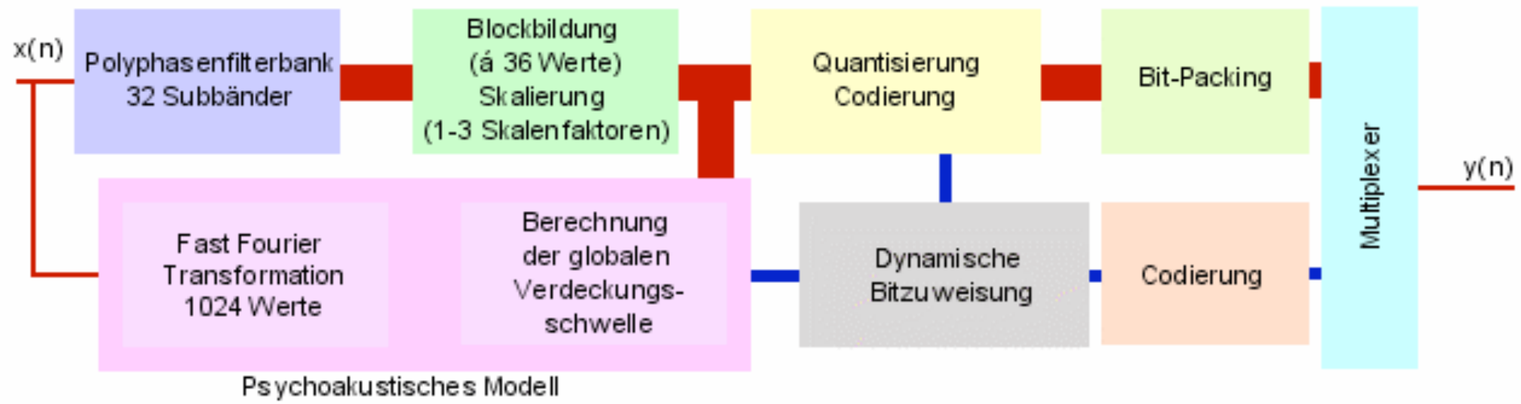



**Blockbild MPEG Audio Layer-1 Encoder**

384 PCM/Frame
384/32=12 PCM / Subband
8 ms bei 48 kHz

Kompression 4:1

# Audio Layer 2: MP2

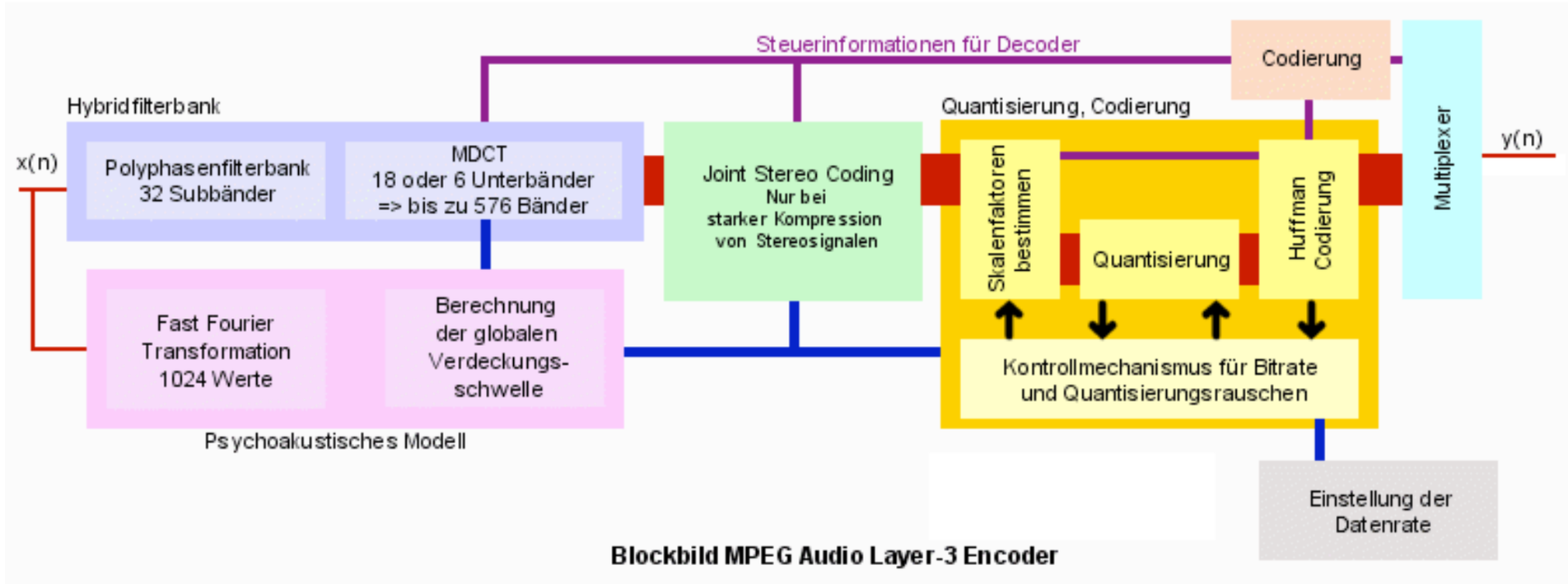



**Blockbild MPEG Audio Layer-1 Encoder**

1152 PCM/Frame
1152/32 = 36 PCM / Subband
24 ms bei 48 kHz

Kompression 6:1

# Audio Layer 3: MP3




**Blockbild MPEG Audio Layer-3 Encoder**

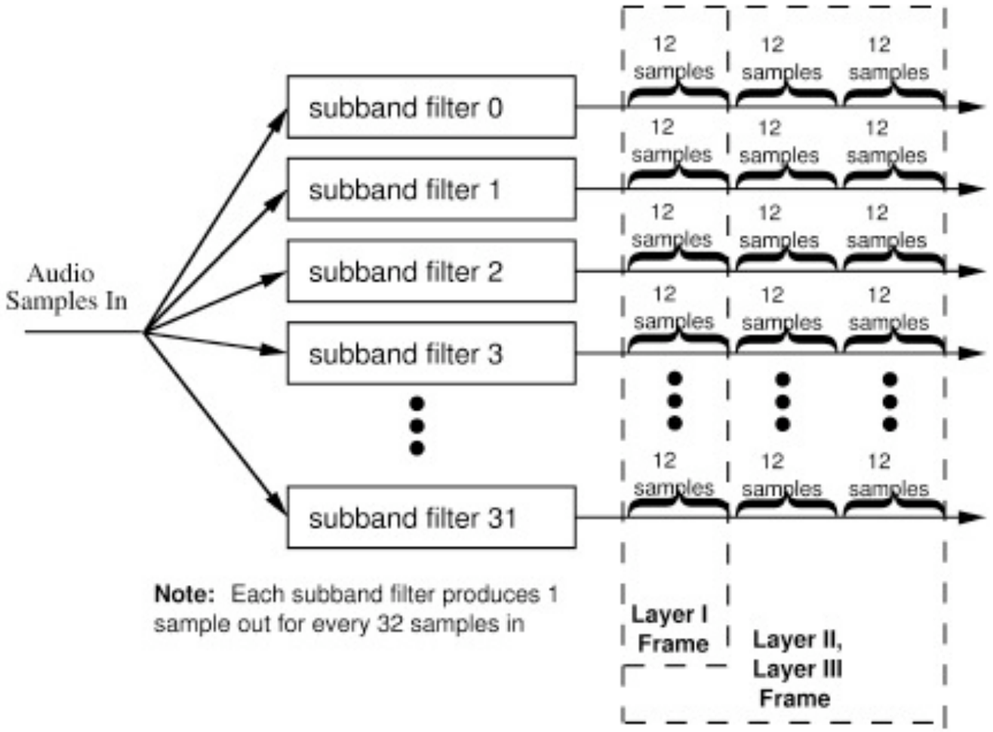
ISO-MPEG 1, Audio Layer-111

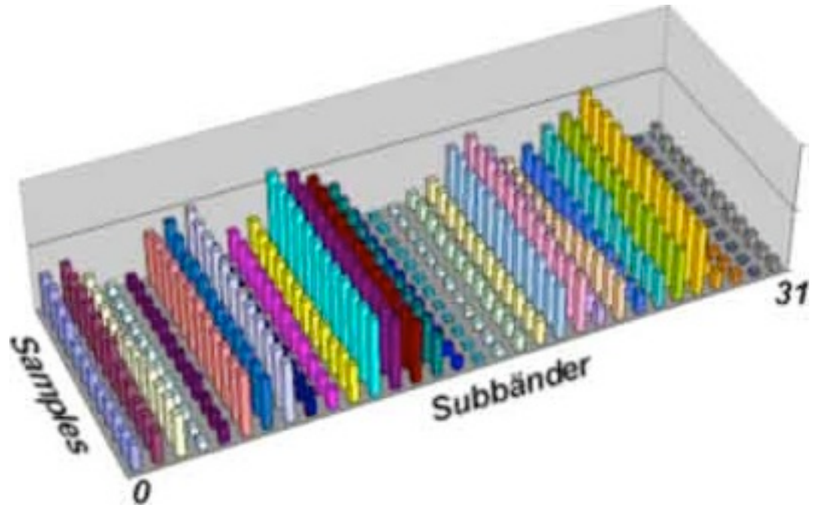
(IS 11172-3 und IS 13818-3)

Audio - mono & stereo, ISO/IEC 11172-3 encoder

Kompression 12:1

# Filterbank



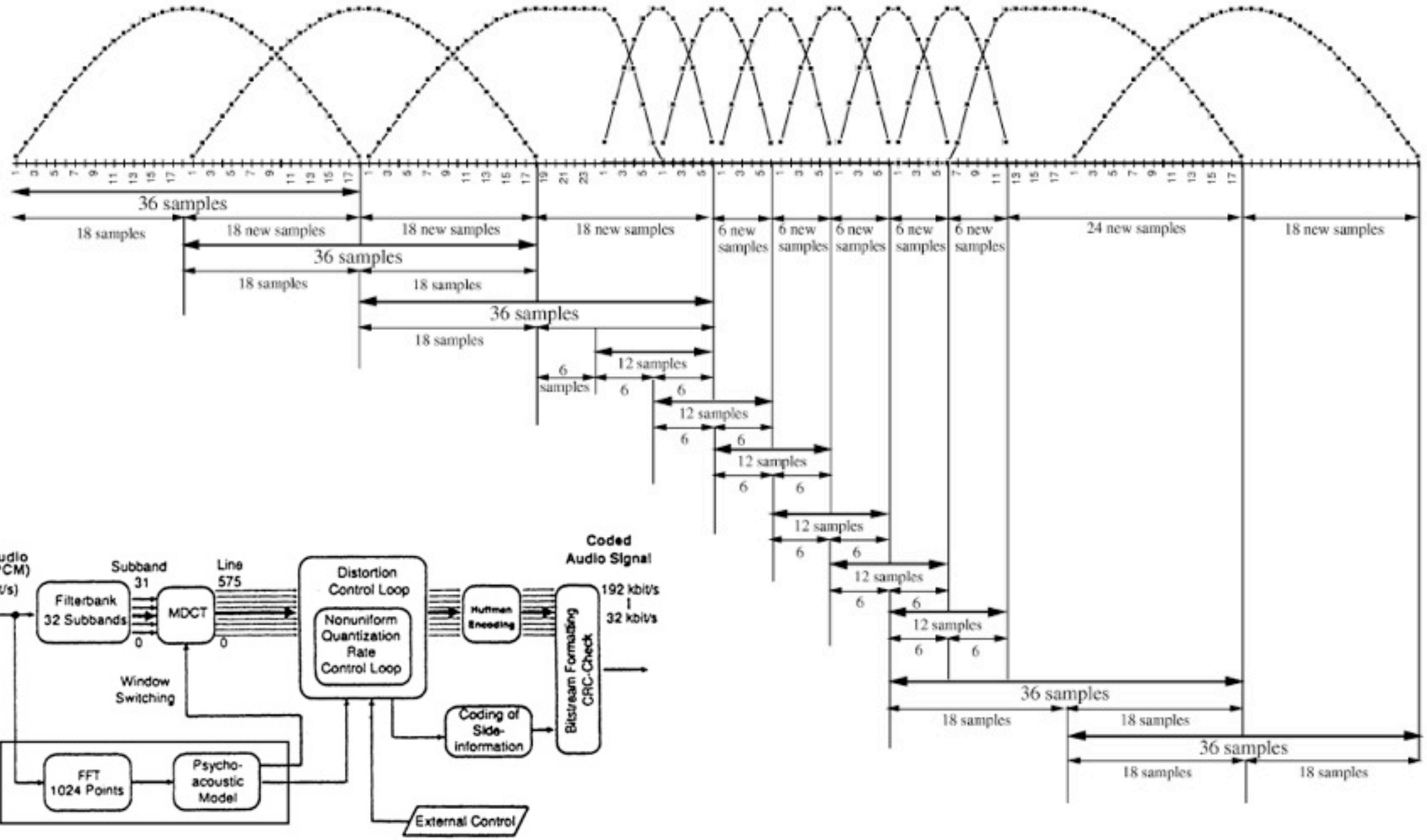




1152=12*3*32 Samples

# MDCT

$$X_k = \sum_{n=0}^{2N-1} x_n \cos\left[\frac{\pi}{N}\left(n + \frac{1}{2} + \frac{N}{2}\right)\left(k + \frac{1}{2}\right)\right]$$

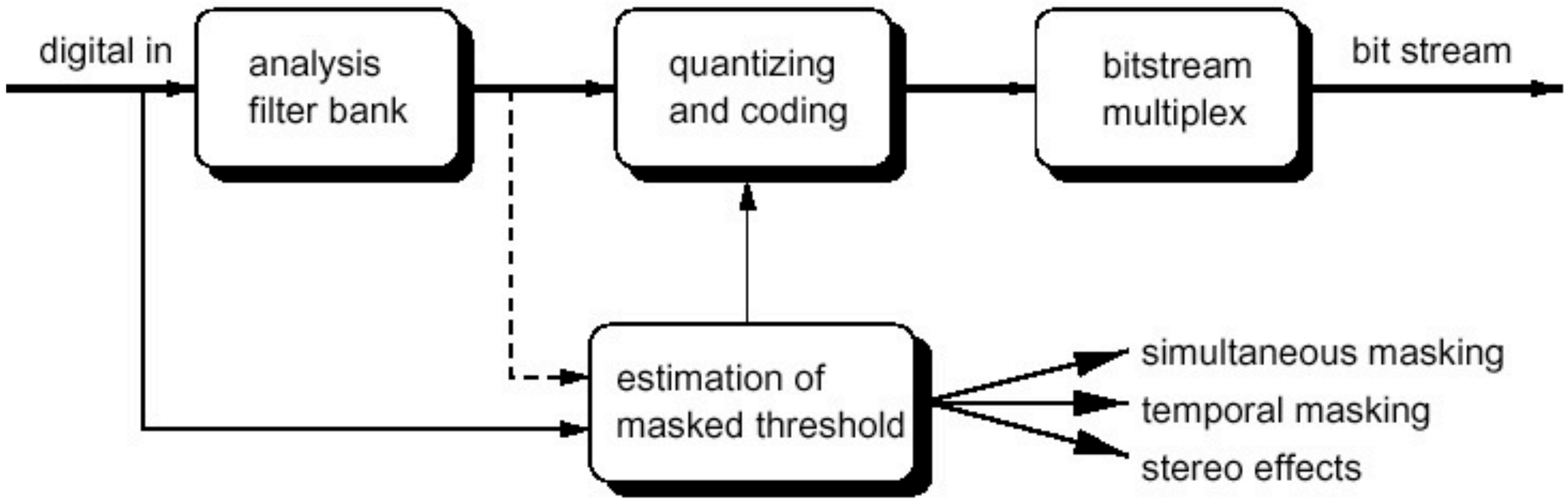
Modifizierte Diskrete Cosinus Transformation



# MP3 Hörmodell

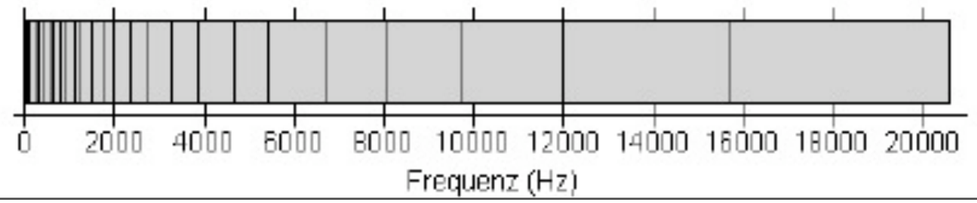
## Perceptual Audio Coding

Audiokodierung mit einem psychoakustischen Modell



# Kritische Bänder (Barkhausen)

| Kritisches | Frequenz | | | Kritisches | Frequenz | | |
|---|---|---|---|---|---|---|---|
| Band | Tief | Hoch | Breite | Band | Tief | Hoch | Breite |
| **0** | 0 | 100 | 100 | **13** | 2000 | 2320 | 320 |
| **1** | 100 | 200 | 100 | **14** | 2320 | 2700 | 380 |
| **2** | 200 | 300 | 100 | **15** | 2700 | 3150 | 450 |
| **3** | 300 | 400 | 100 | **16** | 3150 | 3700 | 550 |
| **4** | 400 | 510 | 110 | **17** | 3700 | 4400 | 700 |
| **5** | 510 | 630 | 120 | **18** | 4400 | 5300 | 900 |
| **6** | 630 | 770 | 140 | **19** | 5300 | 6400 | 1100 |
| **7** | 770 | 920 | 150 | **20** | 6400 | 7700 | 1300 |
| **8** | 920 | 1080 | 160 | **21** | 7700 | 9500 | 1800 |
| **9** | 1080 | 1270 | 190 | **22** | 9500 | 12000 | 2500 |
| **10** | 1270 | 1480 | 210 | **23** | 12000 | 15500 | 3500 |
| **11** | 1480 | 1720 | 240 | **24** | 15500 | 22050 | 6550 |
| **12** | 1720 | 2000 | 280 | | | | |



# MP3-Hörmodell

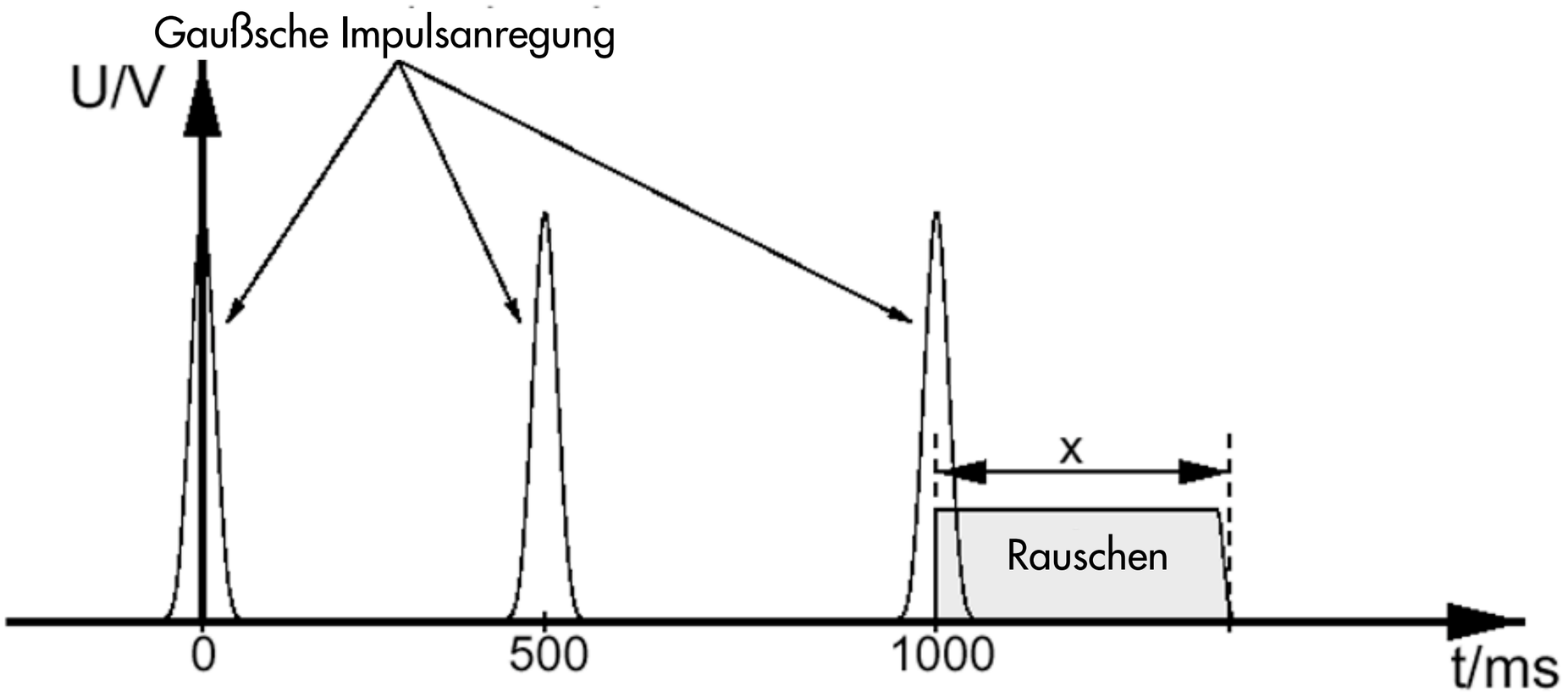
MUSICAM - Masking pattern adapted Universal Subband Integrated Codingand Multiplexing (MUSICAM)

## Psychoakustische Reduktion des Signals

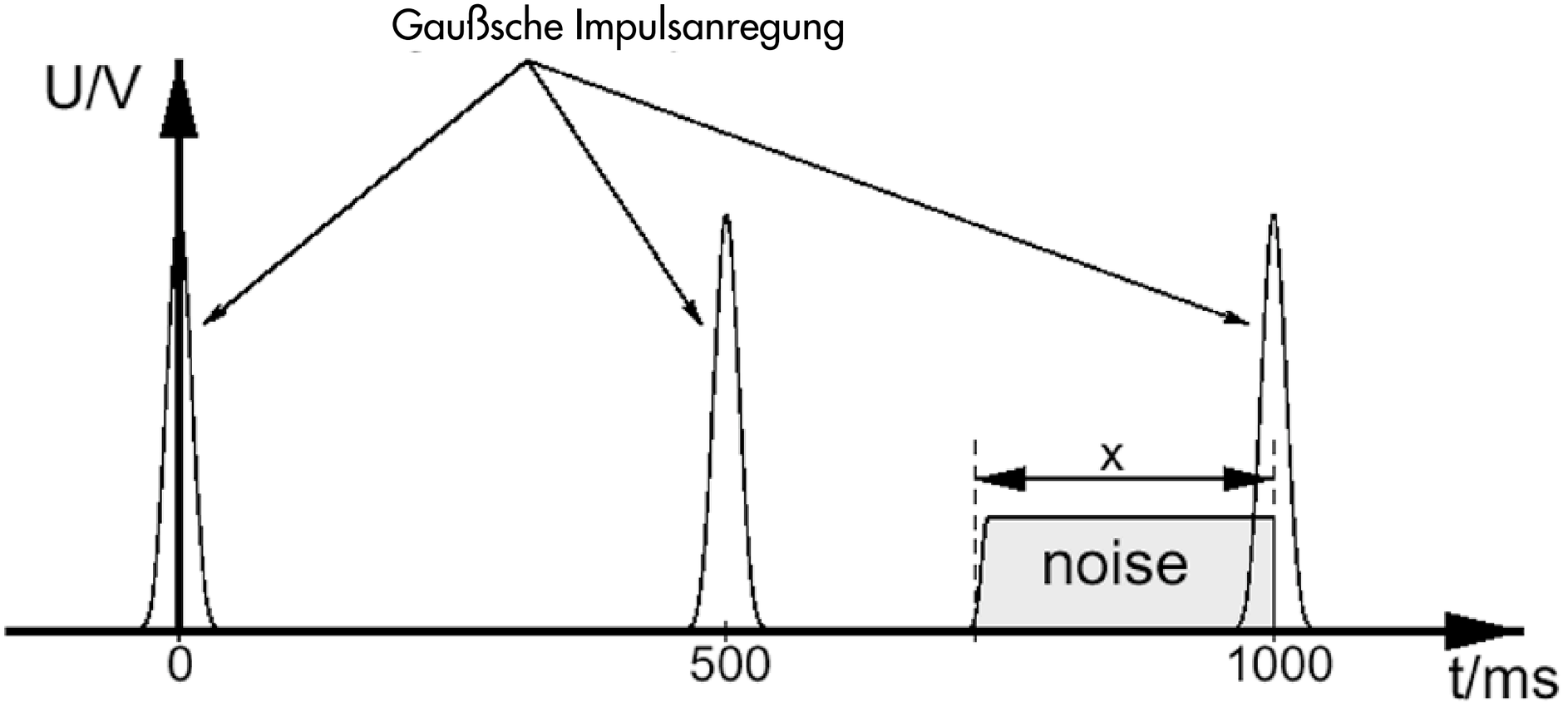
Reduktion wegen Lautstärkemaskierung für 15-20ms

Reduktion räumlicher Information (Stereo im Bassbereich)

# Vorwärtsmaskierung

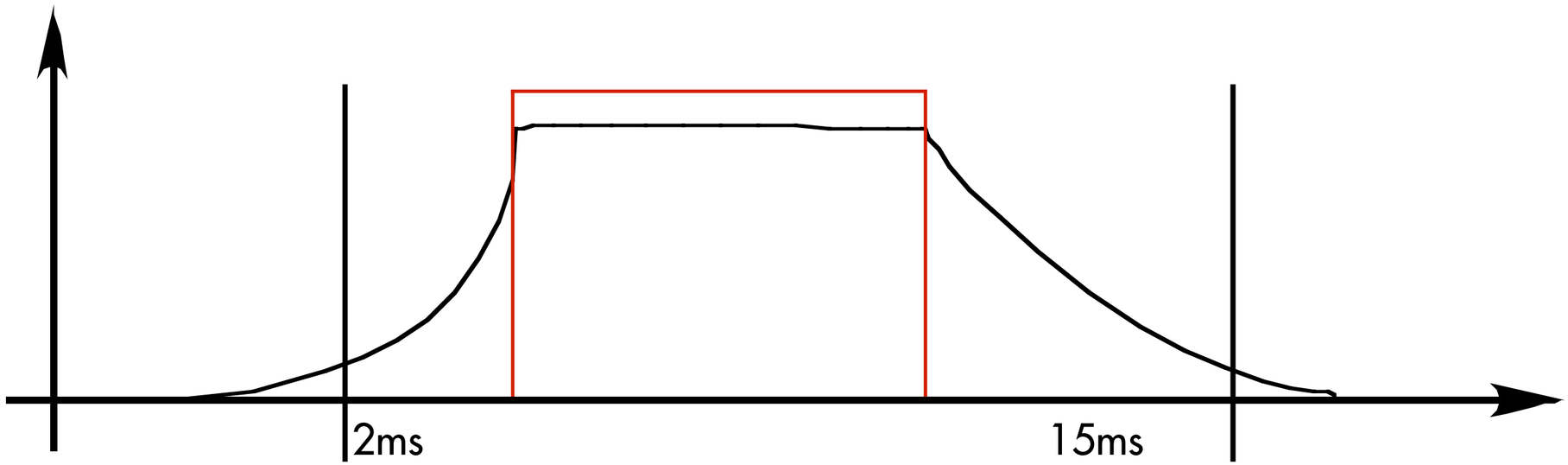


# Rückwärtsmaskierung



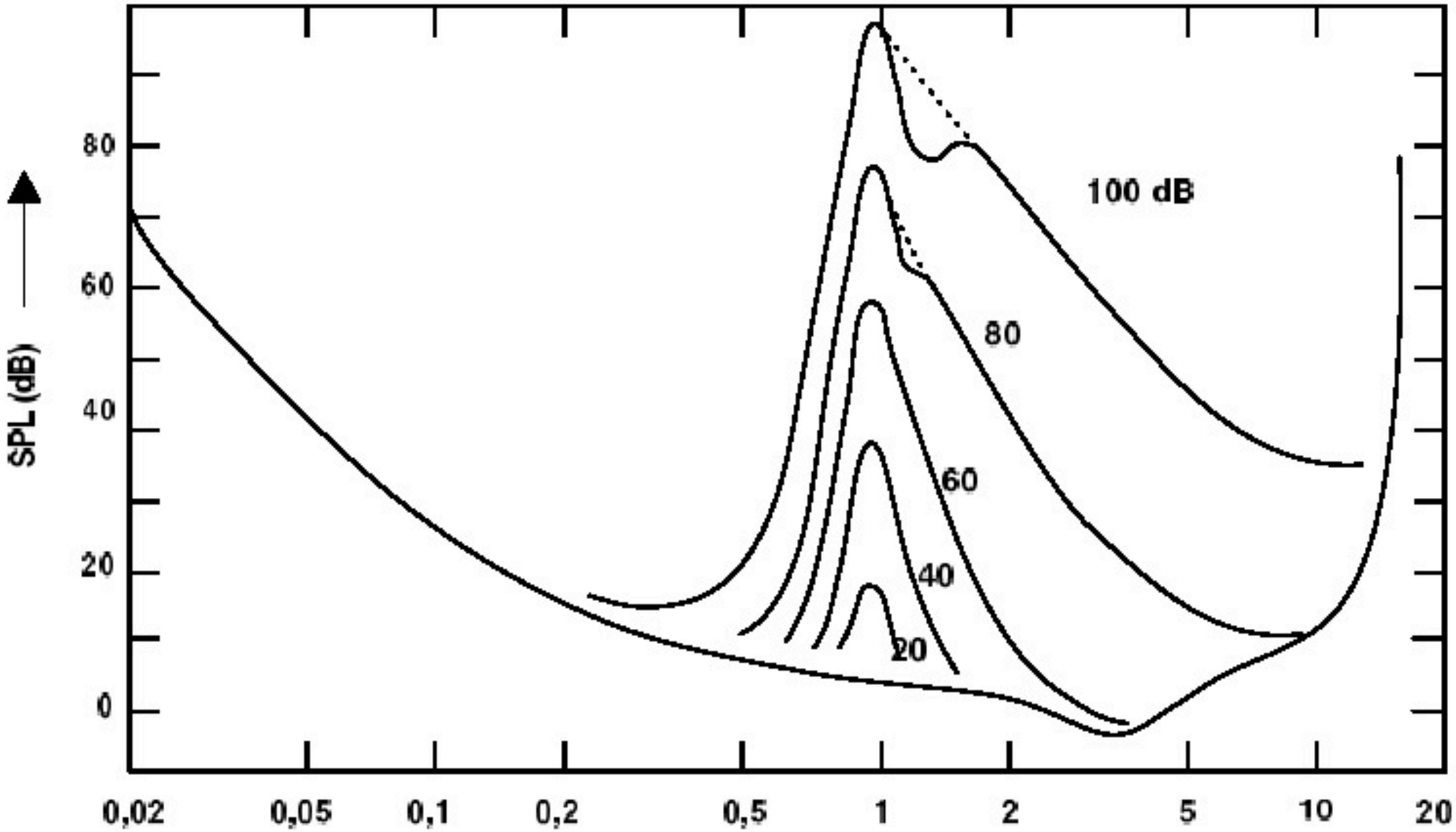

Rückwärtsmaskierung

# MP3 Hörmodell



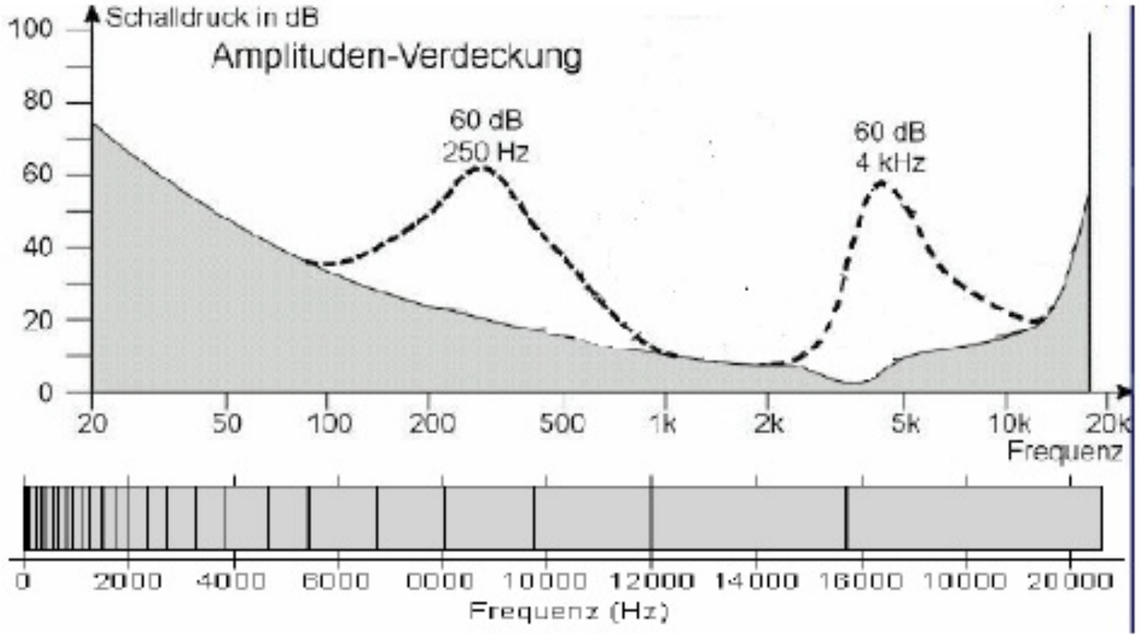

Vorwärts-, Lautstärken- und Rückwärtsmaskierung
am Beispiel eines Rechteckimpulses

# MP3 Hörmodell

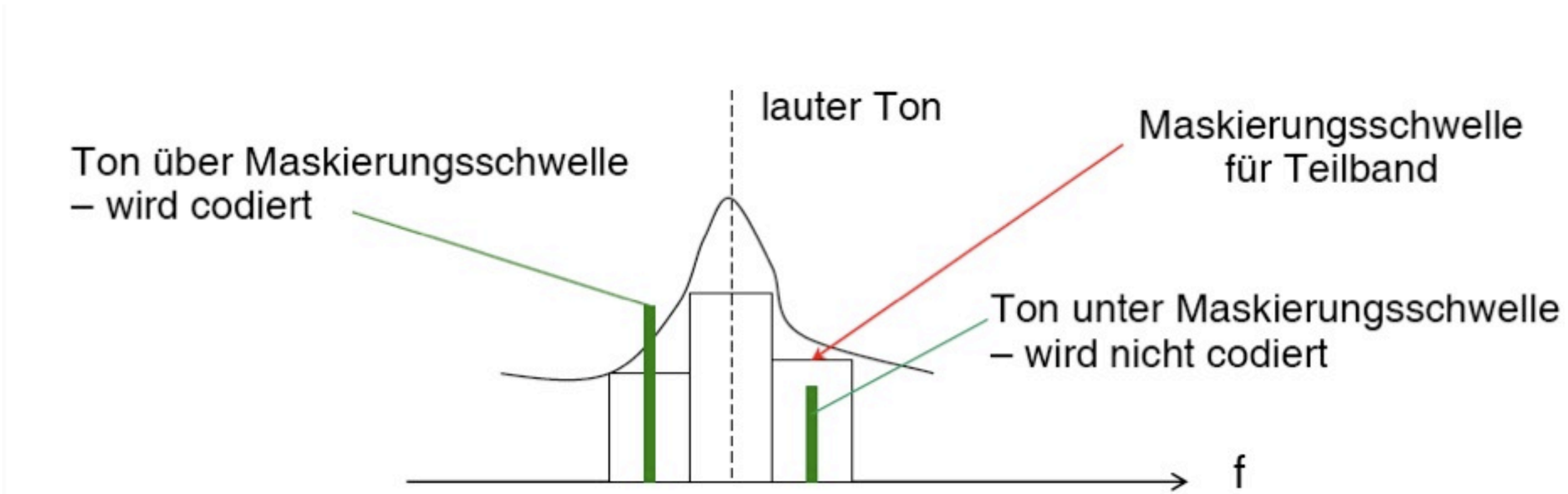



Maskierung durch Lautstärke

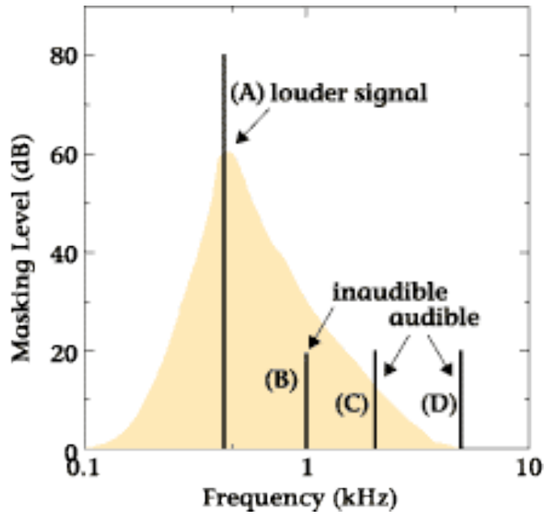
# Subband-Maskierung

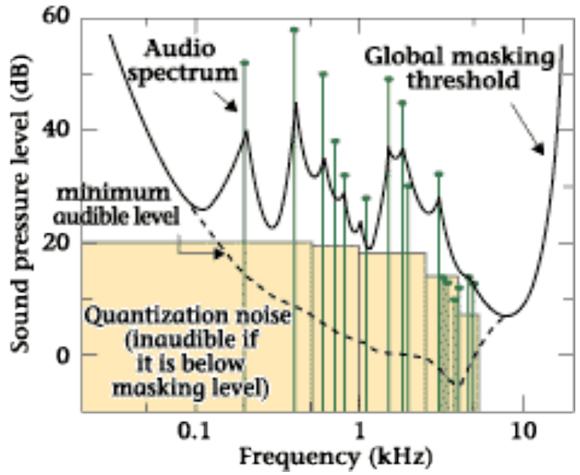


# Maskierungsschwelle



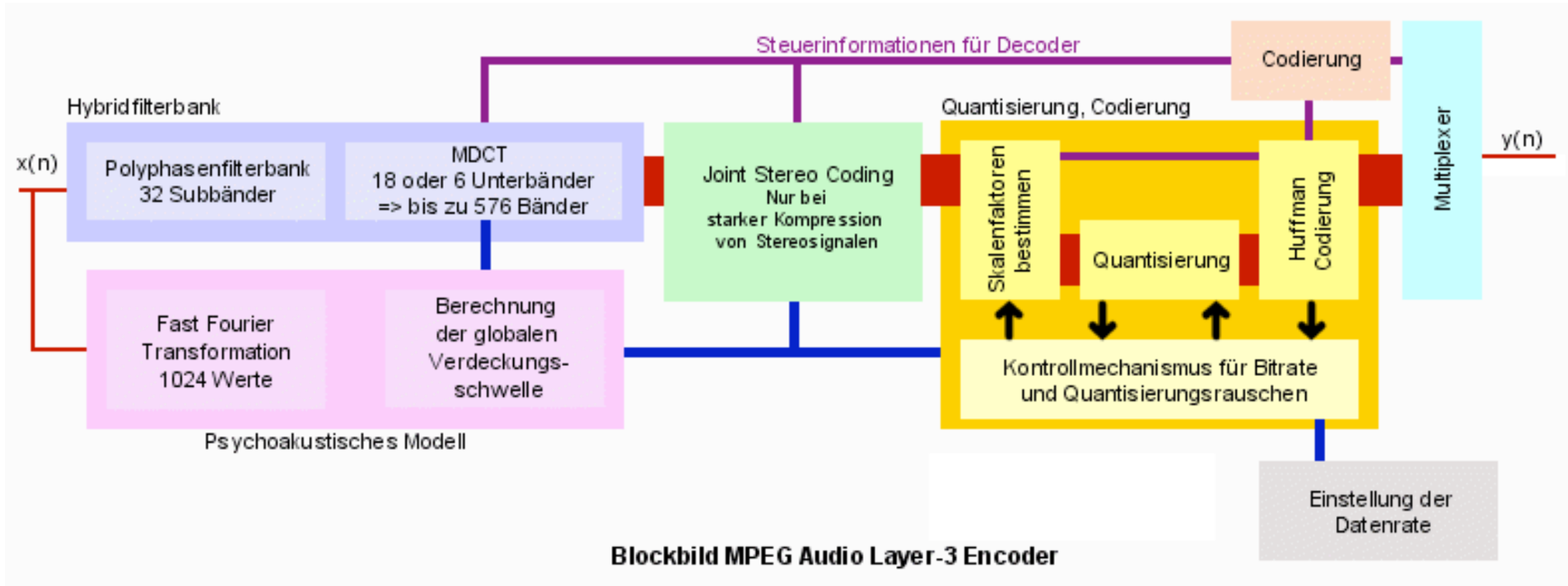
# Globale Maskierungsschwelle



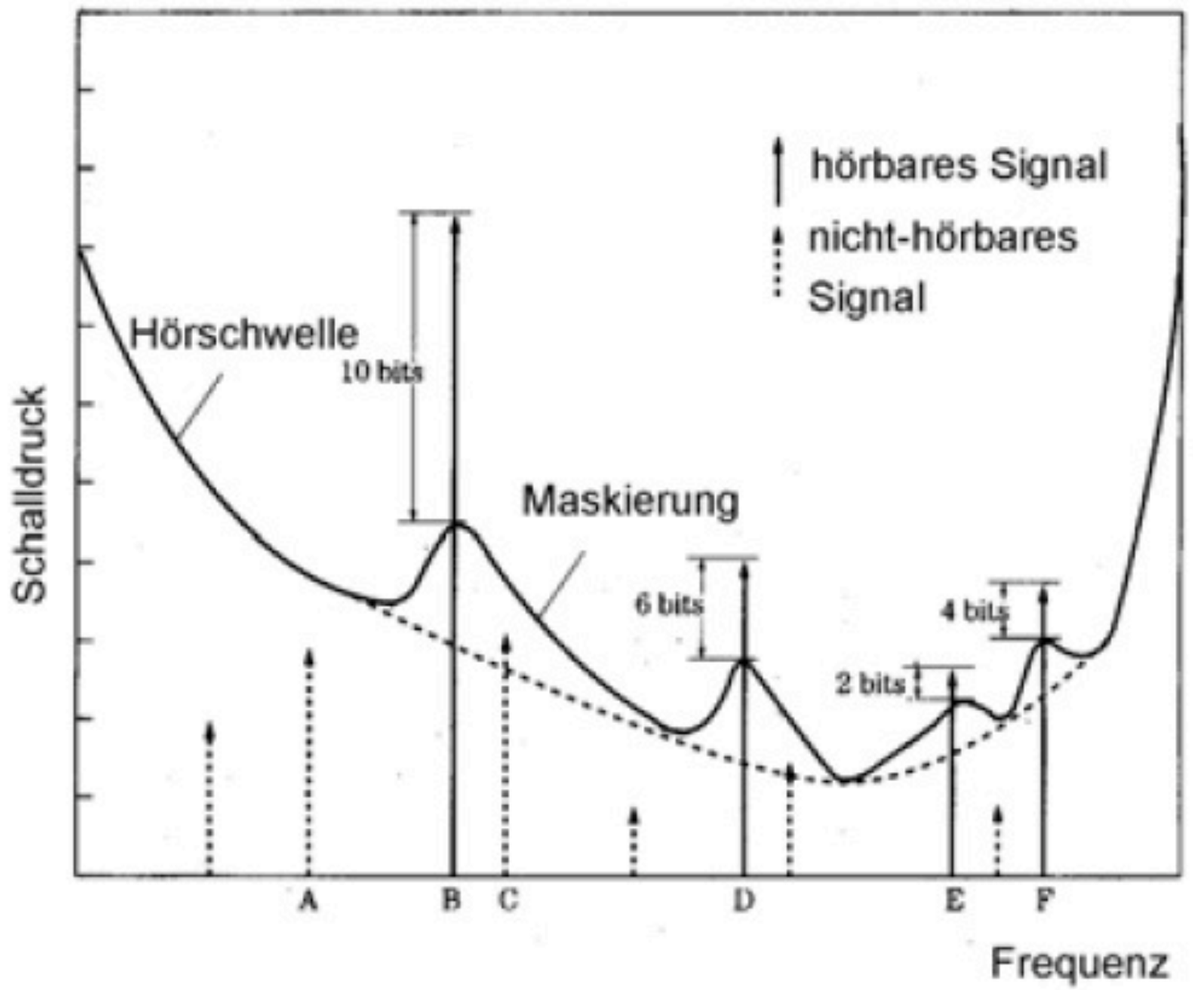

Die globale Maskierungsschwelle wird aus der Kombination der Maskierungsschwellen aller Bänder sowie der Hörschwelle berechnet.

# Quantisierung




**Blockbild MPEG Audio Layer-3 Encoder**

# Maskierungsschwellen



# Maskierung

Ergebnis nach der Analyse der ersten 16 Bänder:

| Band | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pegel (dB) | 0 | 8 | 12 | 10 | 6 | 2 | 10 | 60 | 35 | 20 | 15 | 2 | 3 | 5 | 3 | 1 |

Falls das psychoakustische Modell besagt, dass der Pegel in Band 8 (60 dB) zu folgender Maskierung der Nachbarbänder führt:

Maskierung um 12 dB in Band 9

Maskierung um 15 dB in Band 7

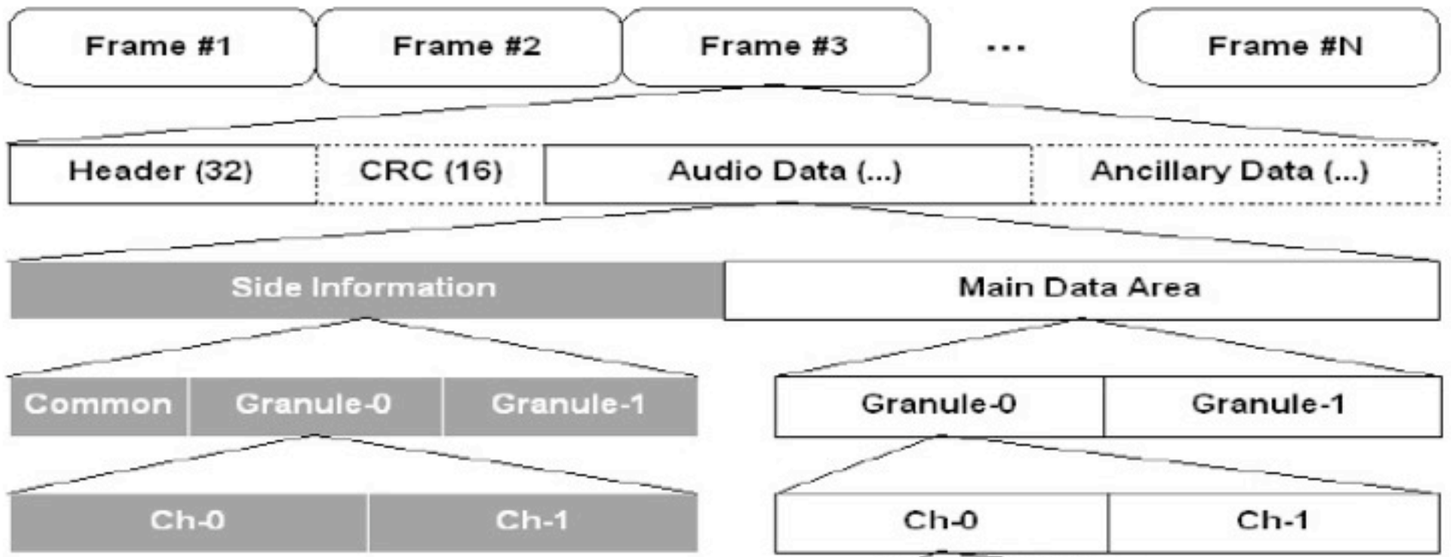
Pegel in Band 7 ist 10 dB —> Weglassen

Pegel in Band 9 ist 35 dB —> Codieren

Wegen der Maskierung ist eine Ungenauigkeit von 12 dB zulässig,
d.h. mit zwei Bit weniger codierbar. Jedes Bit verdoppelt die Amplitudenauflösung, d.h. erhöht sie um 6 dB.

# Steuerinformationen

## Side Information (32 Bytes)




Für jeden Kanal werden für jedes Subband mitgeliefert:

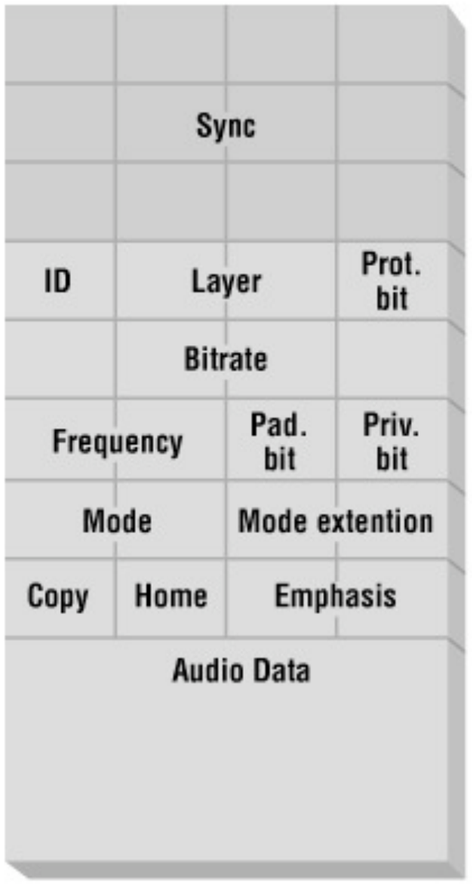
Anzahl der Kodierungsbits

Skalierungsfaktor

Huffmann-Tabelle (Index)

# MP3 – Frame

| Header (32) | CRC (0,16) | Side Information (136,256) | Main Data | Ancillary Data |
|---|---|---|---|---|




**Bitrate**

| bits | V1,L1 | V1,L2 | V1,L3 | V2,L1 | V2, L2 & L3 |
|---|---|---|---|---|---|
| 0000 | free | free | free | free | free |
| 0001 | 32 | 32 | 32 | 32 | 8 |
| 0010 | 64 | 48 | 40 | 48 | 16 |
| 0011 | 96 | 56 | 48 | 56 | 24 |
| 0100 | 128 | 64 | 56 | 64 | 32 |
| 0101 | 160 | 80 | 64 | 80 | 40 |
| 0110 | 192 | 96 | 80 | 96 | 48 |
| 0111 | 224 | 112 | 96 | 112 | 56 |
| 1000 | 256 | 128 | 112 | 128 | 64 |
| 1001 | 288 | 160 | 128 | 144 | 80 |
| 1010 | 320 | 192 | 160 | 160 | 96 |
| 1011 | 352 | 224 | 192 | 176 | 112 |
| 1100 | 384 | 256 | 224 | 192 | 128 |
| 1101 | 416 | 320 | 256 | 224 | 144 |
| 1110 | 448 | 384 | 320 | 256 | 160 |
| 1111 | bad | bad | bad | bad | bad |

**Layer**
00 - reserved
01 - Layer III
10 - Layer II
11 - Layer I

**Mode**
00 - Stereo
01 - Joint stereo (Stereo)
10 - Dual channel (Stereo)
11 - Single channel (Mono)

Copy
0 - Audio is not copyrighted
1 - Audio is copyrighted

Original
0 - Copy of original media
1 - Original media