

5. Ton und Klang

- 5.1 Ton: Physikalische und physiologische Aspekte
- 5.2 Kompression von Audio-Signalen: MPEG-Audio
- 5.3 Audio-Datenformate: Übersicht
- 5.4 Klangerzeugung und MIDI



Weiterführende Literatur:

Arne Heyda, Marc Briede, Ulrich Schmidt: Datenformate im Medienbereich, Fachbuchverlag Leipzig 2003, Kapitel 5

John Watkinson: MPEG Handbook, 2nd ed., Butterworth-Heinemann 2004

Pulse Code Modulation (PCM)

- Klassische Digitalisierung:
 - Aufzeichnung des analogen Signalwertes zu festgelegten Zeitpunkten mit festgelegter Auflösung
- G.711 (für Telefonie):
 - 8 kHz Abtastfrequenz für 4 kHz breites Teilband (Sprache)
 - Auflösung 8 bit
 - 64 kbit/s Bandbreite = Breite eines ISDN „B-Kanals“
- Viele weitere Anwendungen
 - z.B. digitale Tonaufzeichnung auf Videoband (PCM-1630)

- Kompression von Audiodaten
 - Verlustfreie Kompression nur wenig wirksam
 - Generell relativ niedrige Kompressionsraten erreichbar

Verlustbehaftete Audio-Kompressionsverfahren

- Verlustbehaftete Audiokompression
 - Basiert auf psychoakustischem Modell der Tonwahrnehmung
 - Wichtigster Effekt:
Maskierte Bestandteile des Audio-Signals werden nicht codiert
 - Bekanntester Standard: MPEG Audio Layer III (MP3)
- MPEG = Moving Picture Expert Group
 - Standardisierungsgremium von ISO (International Standards Organization) und IEC (International Electrotechnical Commission)
 - Arbeitet seit 1988 an Video- und Audio-Kompression
 - » Untergruppe MPEG/Audio
 - MPEG-Audio-Standards werden z.B. verwendet bei
 - » DAB (Digital Audio Broadcast)
 - » DVB (Digital Video Broadcast) incl. terrestrischer Variante DVB-T
 - » DVD-Video

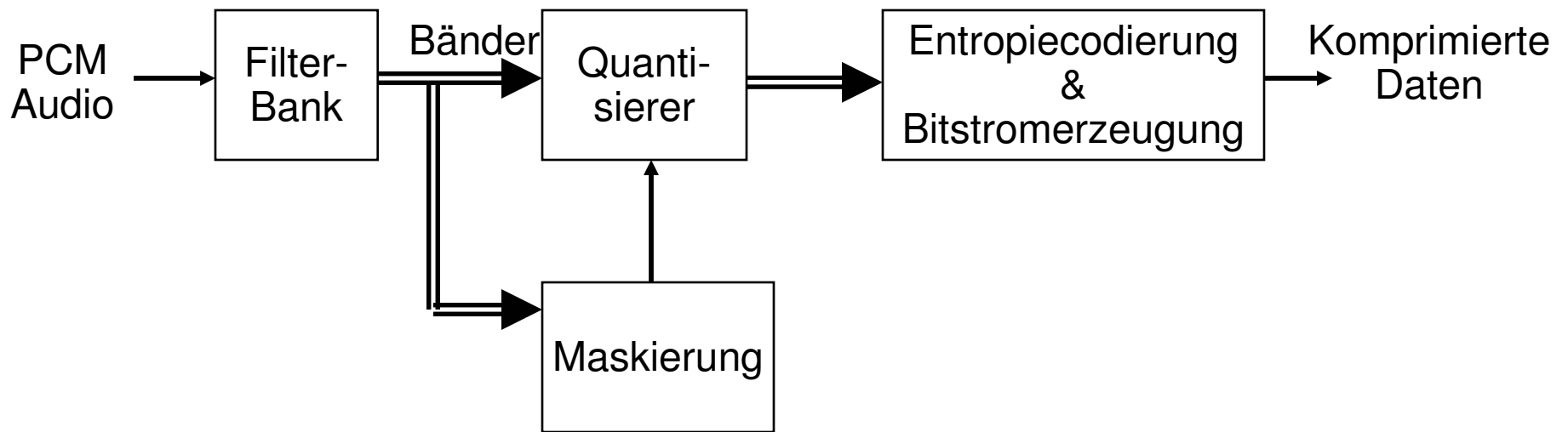
MPEG Audio: Geschichte

- EU-gefördertes "Eureka"-Projekt Nr. 147 (CCETT(F), IRT(D), Philips(NL))
 - MUSICAM (Masking pattern adapted universal sub-band integrated coding and multiplexing)
 - Ziel: DAB-Standard
- Parallelentwicklung (AT&T, Thomson, Fraunhofer, CNET):
 - ASPEC (Adaptive Spectral Perceptual Entropy Coding)
 - Ziel hochwertiges Audio über ISDN
- Juli 1990: Ausführliche Tests beim Schwedischen Rundfunk, anschließend Kombination der beiden Verfahren in die 3 MPEG-Layer.
 - Layer I: vereinfachtes MUSICAM, schwache Kompression, preisgünstig
 - Layer II: = MUSICAM, für DAB und Audio in DVB
 - Layer III: Kombination der Stärken von ASPEC und MUSICAM, hohe Kompression über Telekommunikationsverbindungen

Audio-Codierung in MPEG

- MPEG-1 Audio:
 - PCM mit 32, 44.1 oder 48 kHz
 - max. Datenrate 448 kbit/s
- MPEG-2 Audio:
 - PCM mit 16, 22.05, 24, 32, 44.1 oder 48 kHz
 - max. 5 Kanäle
 - max. Datenrate 384 kbit/s
- Einteilung der Audio-Kompressionsverfahren in drei „Layer“ (I, II, III) verschiedener Kompressionsstärke
 - Unabhängig von Wahl des Standards MPEG-1 bzw. MPEG-2 !
 - „MP3“ = MPEG Layer III (Kompression ca. 11:1)
 - » MP3 patentrechtlich geschützt, Fraunhofer IIS Erlangen
- Inzwischen wesentliche Weiterentwicklungen:
 - z.B. AAC, MPEG-4 Audio (siehe später)
 - Ogg-Vorbis

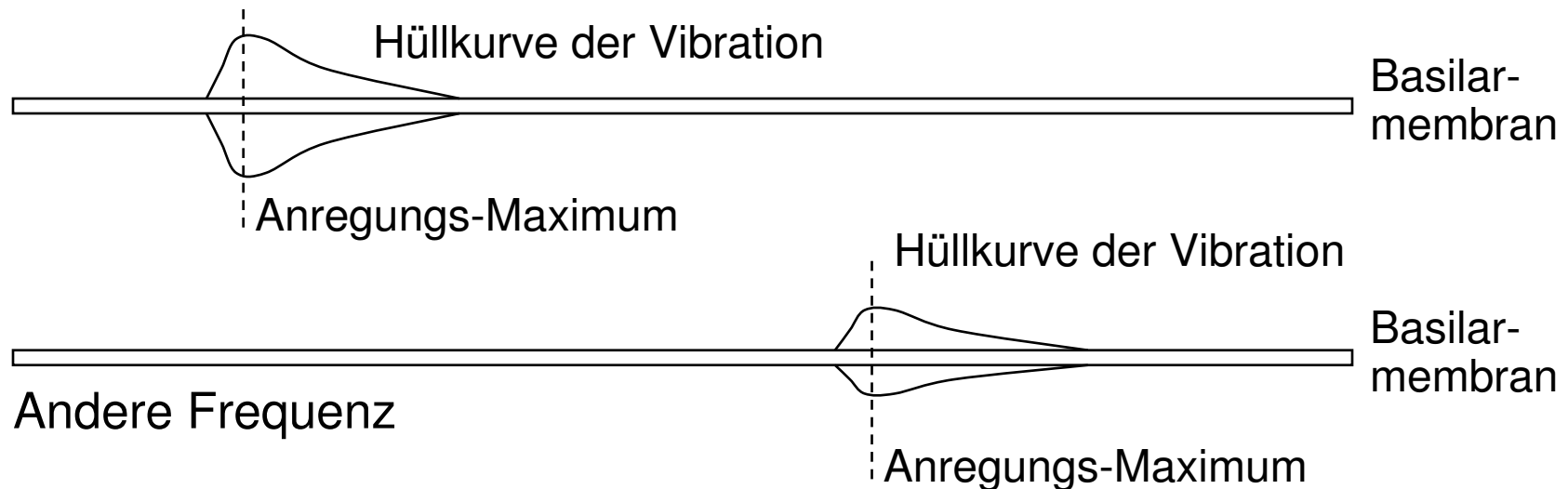
MPEG-Audio Encoder: Grundlegender Aufbau



- Hinweis: Der MPEG-Standard definiert nicht den Aufbau eines Encoders, sondern nur die Decodierung!
- Signal wird in Frequenzbänder aufgeteilt
- Maskierung auf der Basis der Bänder mit einem psychoakustischen Modell

Maskierung und Basilarmembran

- Der Maskierungseffekt erklärt sich physikalisch durch die Anregung der Basilarmembran
 - Frequenz entspricht Ort der Anregung auf der Basilarmembran
 - Genaue Wahrnehmung des Maximums der Anregung (Auflösung ca. 1/12 Halbton, bestimmt durch Abstand der Haarzellen)
 - Anregungen in direkter Frequenz-Nähe sind bis zu einer bestimmten Amplitude nicht wahrnehmbar

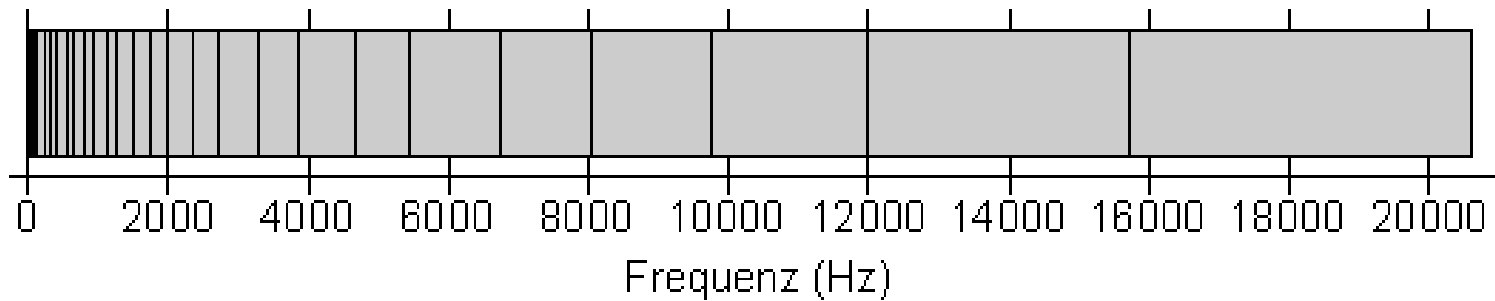


Kritische Bänder

- Einteilung des Hörbereichs in *kritische Bänder*
 - Breite (d.h. Bandbreite im Frequenzspektrum) der Vibrations-Hüllkurve auf der Basilarmembran
 - Breite der Bänder vergrößert sich mit der mittleren Bandfrequenz
- Der Grad der Maskierung einer bestimmten Frequenz ist lediglich abhängig von der Signalintensität im kritischen Band dieser Frequenz.
- "Bark-Skala":
 - Eine Einteilung des Frequenzspektrums entsprechend der Breite kritischer Bänder
 - Benannt nach dem Bremer/Dresdner Physiker Heinrich Barkhausen.

27 Kritische Bänder

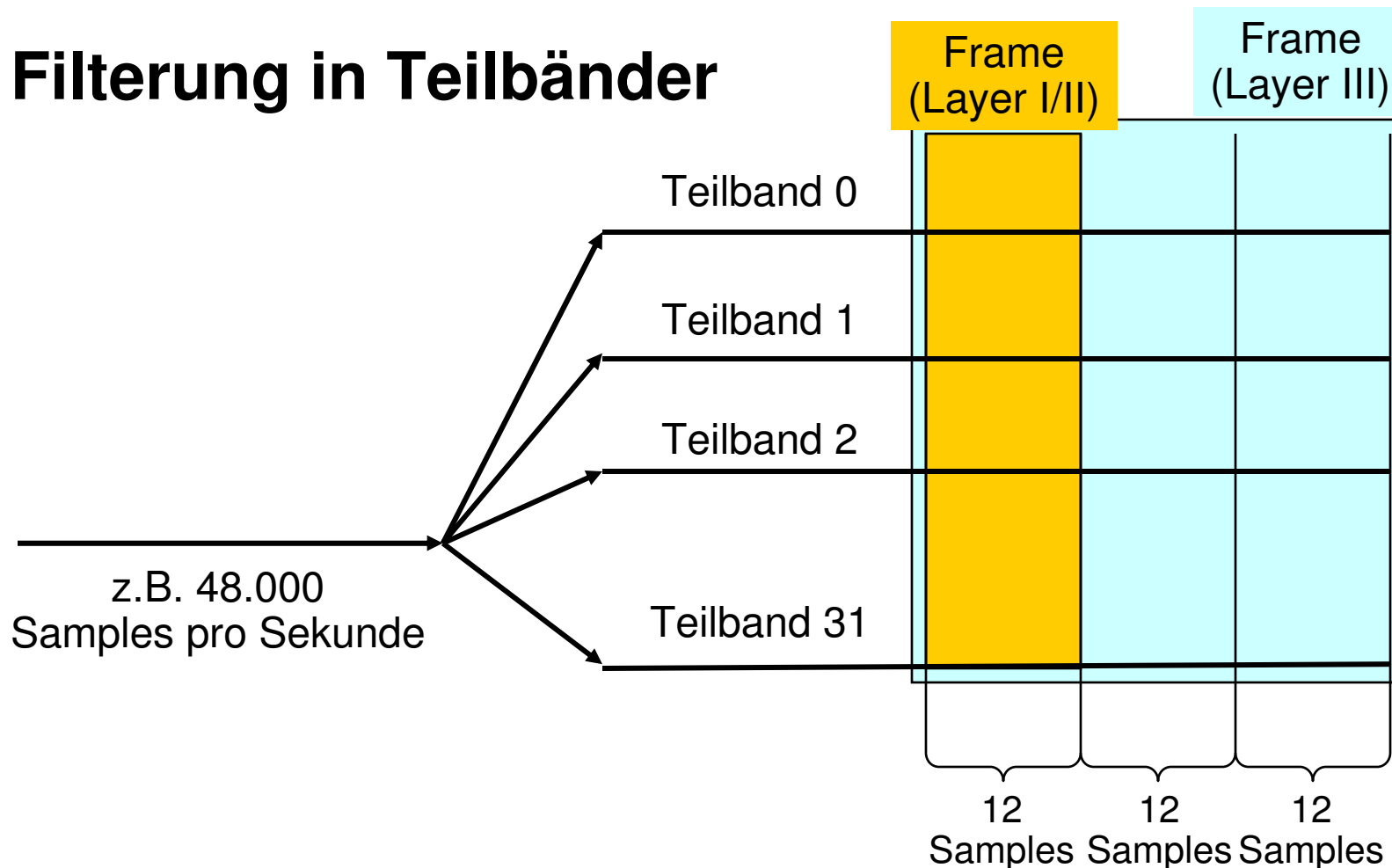
0 – 50	800 – 940	3280 – 3840
50 – 95	940 – 1125	3840 – 4690
95 – 140	1125 – 1265	4690 – 5440
140 – 235	1265 – 1500	5440 – 6375
235 – 330	1500 – 1735	6375 – 7690
330 – 420	1735 – 1970	7690 – 9375
420 – 560	1970 – 2340	9375 – 11625
560 – 660	2340 – 2720	11625 – 15375
660 – 800	2720 – 3280	15375 - 20250



Subband-Kodierung

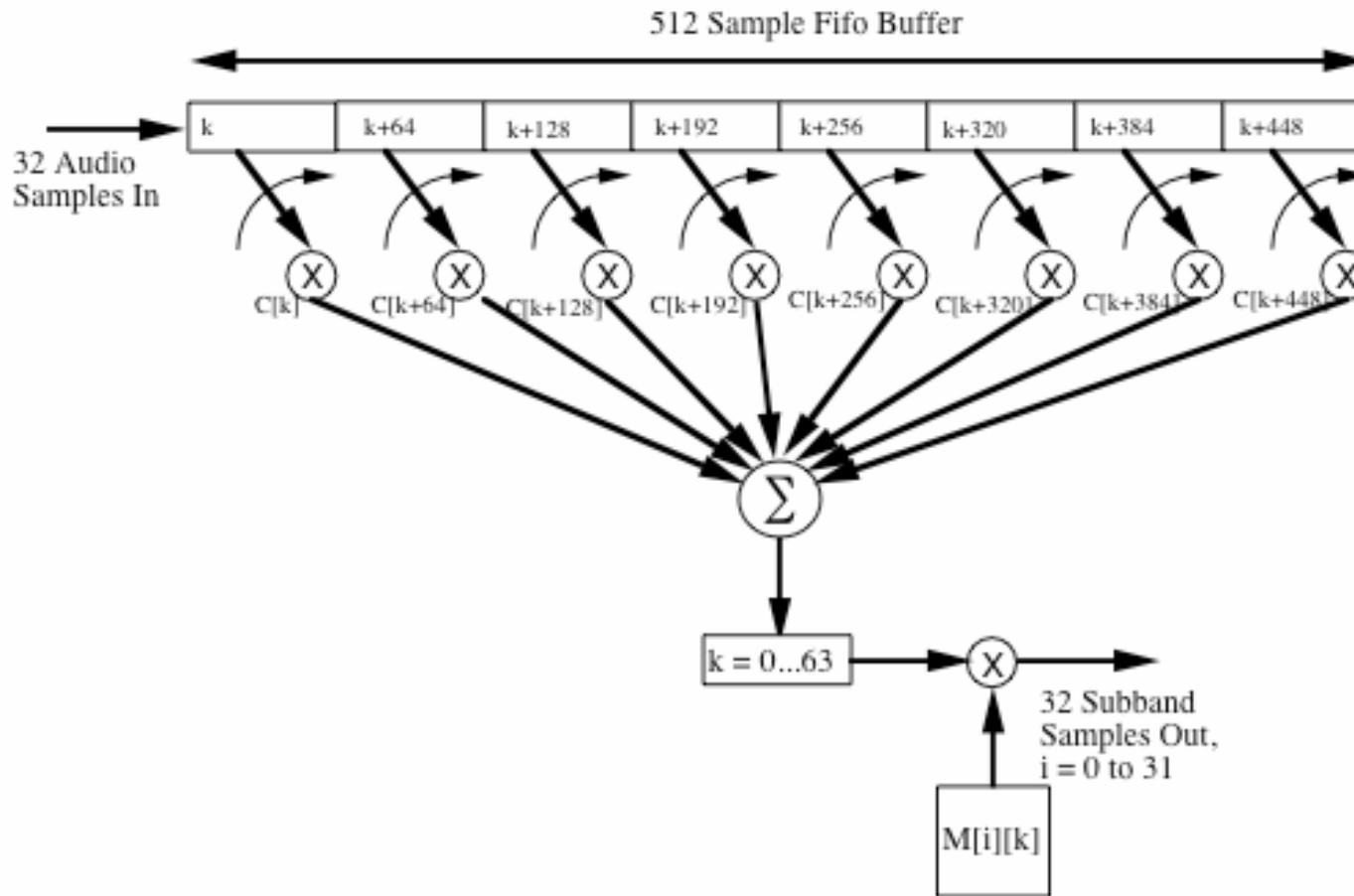
- Energie eines Tonsignals ist meist nicht gleichmäßig auf das Frequenzspektrum verteilt
- Idee:
 - Aufteilen des Signals in Teil-Frequenzbänder
 - Ermittlung des Signalpegels für jedes Teilband
 - Einzel-Codierung der Teilbänder mit jeweils angemessener Bitanzahl
 - » z.B. nicht belegtes Teilband: 0 Bit
 - Funktioniert optimal, wenn Teilbänder an kritische Bänder des Gehörs angepasst

Filterung in Teilbänder



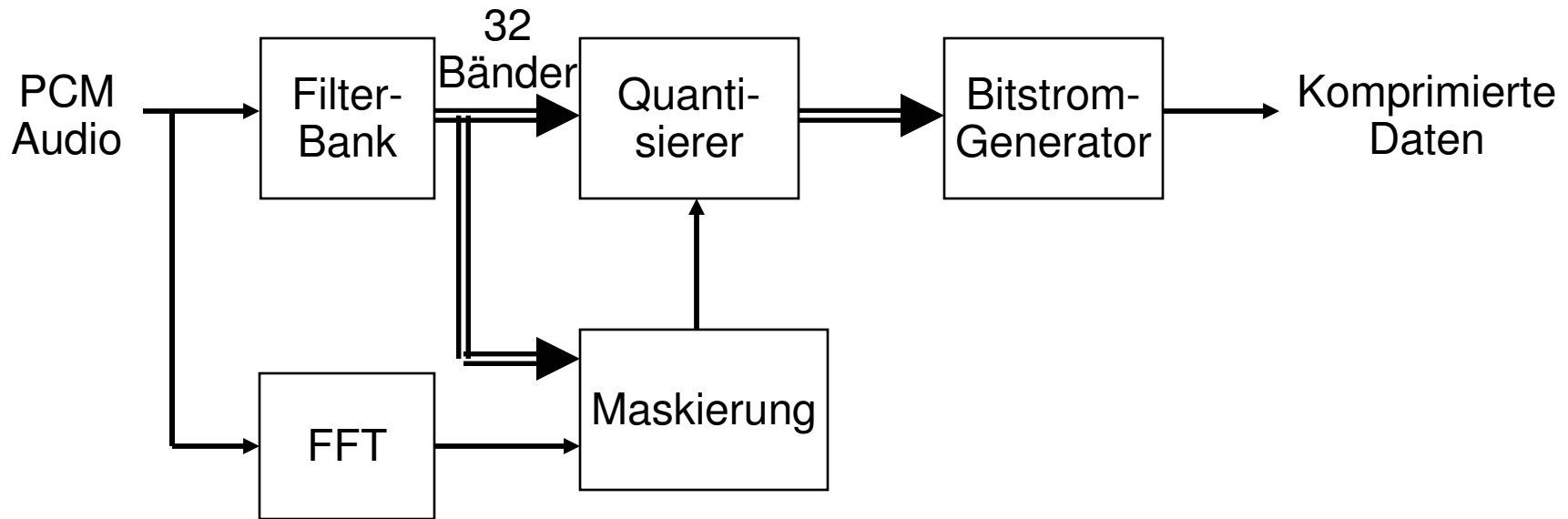
- 12 Samples entsprechen bei 48 kHz ca. 8 ms
- Ein Block von Samples in einem Teilband wird manchmal *bin* genannt
- *Frame*: Gesamtheit der Samples in allen Teilbändern
12 x 32 = 384 Samples in Layer I/II, 3 x 12 x 32 = 1152 Samples in Layer III

Realisierung einer Filterbank



- Ca. 80 Multiplikationen und 80 Additionen pro Ausgabewert

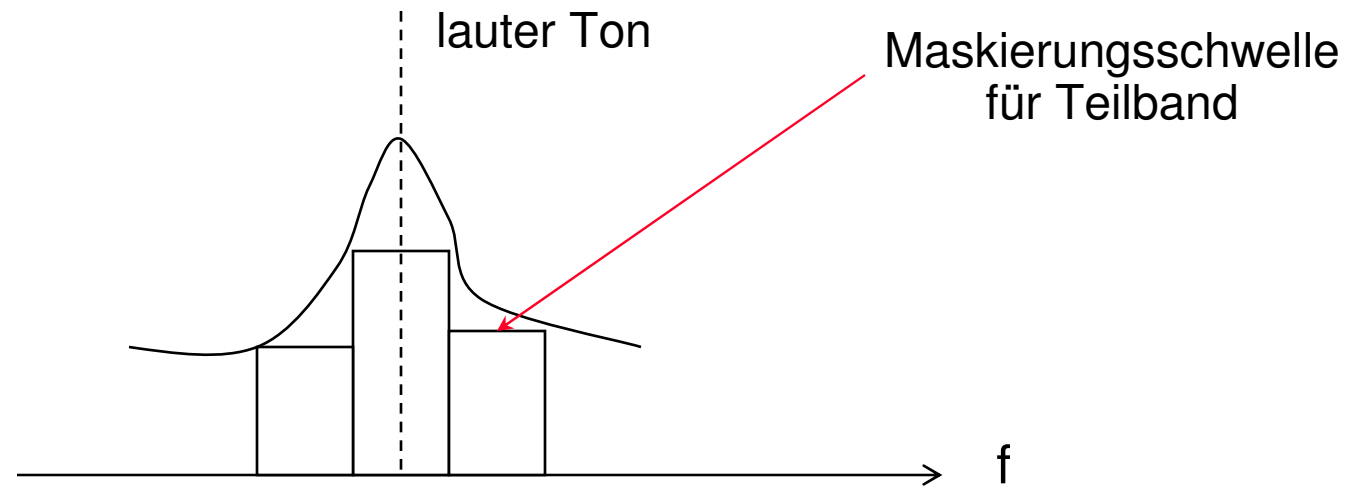
Aufbau eines MPEG-Layer I/II Encoders



- Signal wird in 32 *gleich breite* Frequenzbänder aufgeteilt
 - Effektive Bandfilter funktionieren nur für gleich breite Teilbänder
 - Breite der Teilbänder bei Layer I/II: 750 Hz
 - „Unterabtastung“ der Subbänder: Keine zusätzliche Bandbreite benötigt
- Wegen der Eigenschaften des menschlichen Gehörs sind die Teilbänder ungeeignet für Maskierung
 - Zu breit bei niedrigen und zu schmal bei hohen Frequenzen
 - Einsatz einer zusätzlichen Frequenzanalyse (Fast Fourier Transform, FFT)

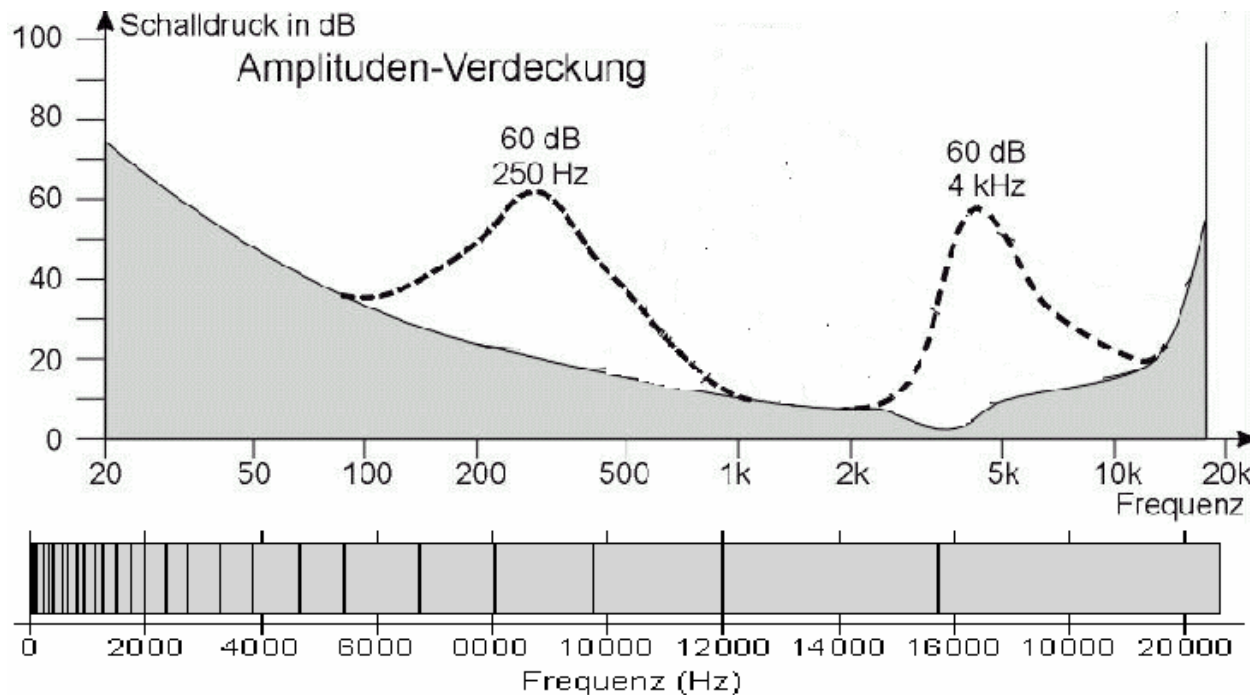
FFT zur Berechnung der Maskierungsschwelle

- FFT = Fast Fourier Transform
- Umsetzung des Amplitudensignals in Frequenzspektrum
 - Angewandt auf die Länge eines Frames (12 Samples)
- Ergebnis:
 - Aufteilung des aktuellen Signals auf viele (Layer I 512, Layer II 1024) Frequenzanteile
- Weiterverarbeitung:
 - Berechnung der aktuellen Kurve für die (frequenzabhängige) Maskierungsschwelle



Psychoakustisches Modell

- Aus dem aktuellen Signalspektrum ergibt sich eine aktuelle Hörbarkeitskurve (wird berechnet)
 - Insbesondere: Für jedes Frequenzband eine Maskierungsschwelle, unter der der Ton nicht mehr hörbar ist
 - Details: z.B. tonale vs. geräuschartige Anteile verschieden behandelt



Maskierung: Beispiel

- 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

- Annahme: Psychoakustisches Modell liefert, 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 Maskierung 12 dB Ungenauigkeit (Rauschen) zulässig, d.h. mit zwei Bit weniger codierbar

1 Bit der Codierung =
doppelter Amplitudenumfang =
6 dB Genauigkeit !

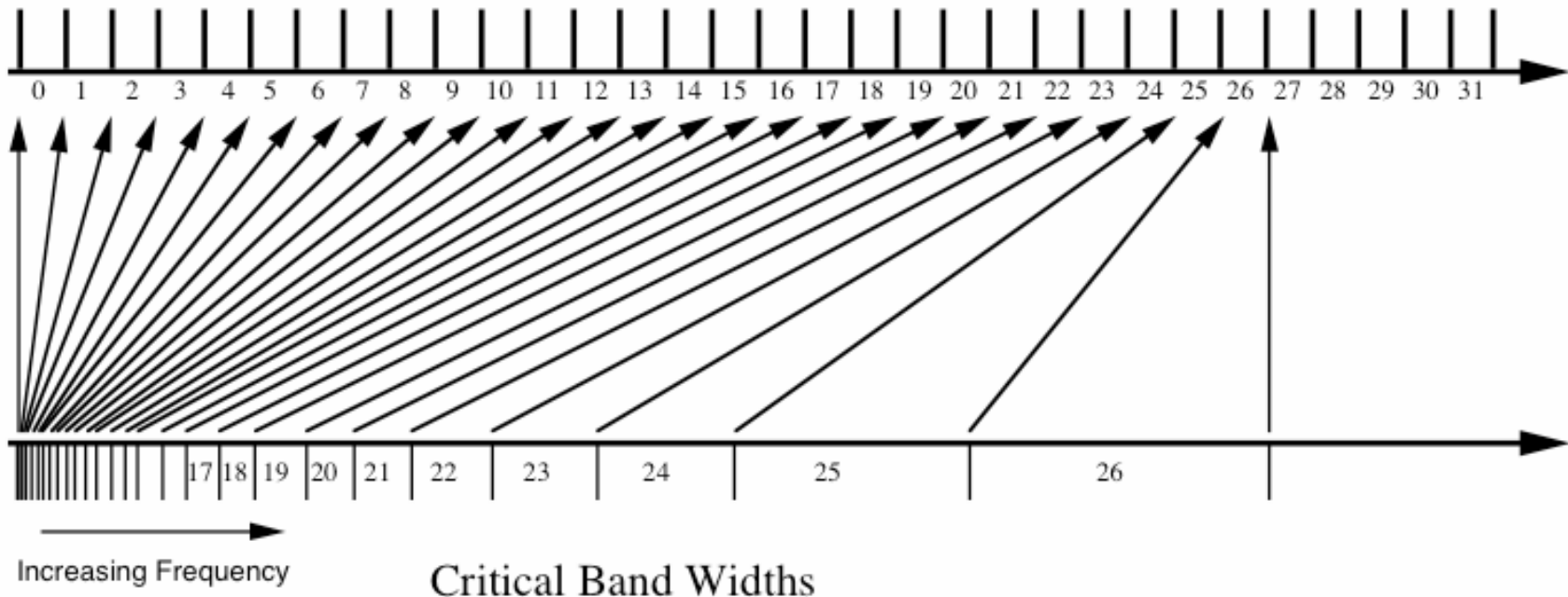
Unterschiede der MPEG Layer

- Layer I:
 - 32 gleichbreite Teilbänder
 - FFT mit 512 Punkten
 - Betrachtung nur eines Frames
 - Psychoakustisches Modell benutzt nur Frequenzmaskierung
- Layer II:
 - 32 gleichbreite Teilbänder
 - FFT mit 1024 Punkten
 - Betrachtung von drei Frames (jetzt, vorher, nachher)
 - Einfache Zeitmaskierung, verfeinerte Bittiefenzuweisung
- Layer III:
 - Teilbänder verschiedener Breite, ähnlich zu den kritischen Bändern
 - Größere Frames (36 Samples)
 - (Modified) DCT der Teilbänder
(in überlappenden „Fenstern“ variierender Breite)
 - Zusätzliche Entropiecodierung (Huffman)
 - Behandlung von Stereo-Redundanzen

Kritische Bänder und Filterbänder

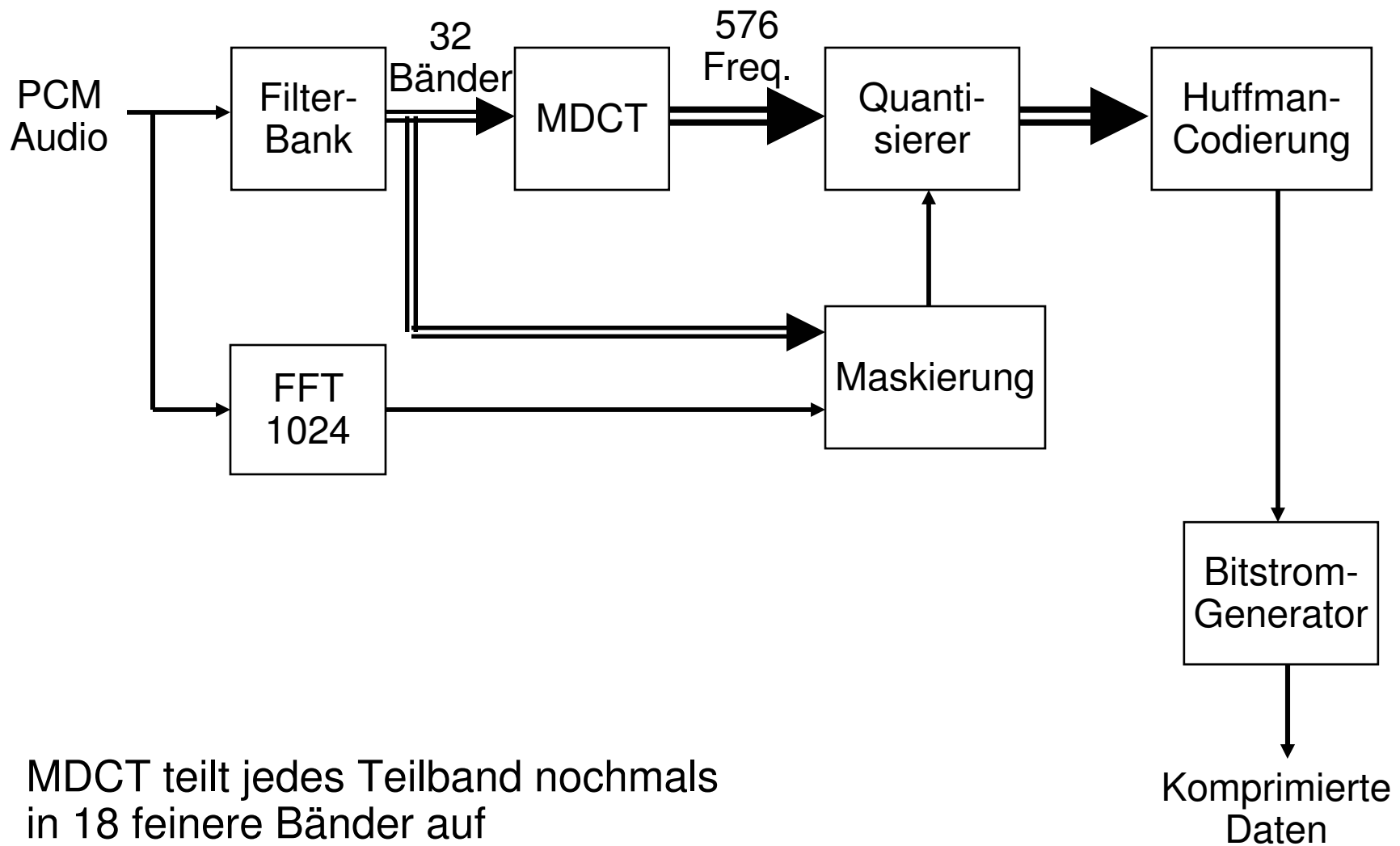
- Situation in MPEG Layer I/II:

MPEG/Audio Filter Bank Bands



Ziel: bessere Anpassung an die Bandbreite der kritischen Bänder
Aber: Nicht durch Filterbank realisierbar

Aufbau eines MPEG-Layer III Encoders



DCT: Diskrete Cosinus-Transformation

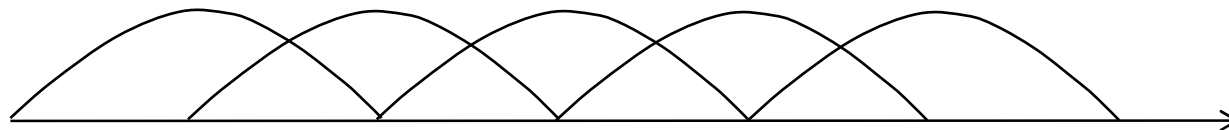
- Ähnlich zur Fourier-Transformation:
 - Gegebenes Signal wird durch Anteile bestimmter Grundfrequenzen beschrieben
- Diskrete Transformation:
 - n Messwerte werden in n Anteilswerte (*Koeffizienten*) umgerechnet
 - Lineare Transformation (Matrixmultiplikation)
 - » D.h. sehr effizient zu berechnen
- Vorteile der Cosinus-Transformation
 - Besser geeignet für Kompression (Filtern von Frequenzen)
 - Bessere „Kompaktheits“-Eigenschaften (Energie auf wenige Grundfrequenzen konzentriert)

$$f_j = \sum_{k=0}^{n-1} x_k \cos \left[\frac{\pi}{n} (j + 1/2)(k + 1/2) \right]$$

Modified Discrete Cosine Transform MDCT (1)

- DCT
 - entspricht kleineren Teilbändern bei der Maskierungsanalyse
 - bei Audio Probleme mit Artefakten an Blockgrenzen
- Modified DCT
 - Überlappung der Cosinusfunktionen um 50%
 - Damit Vermeidung von Artefakten durch Blockgrenzen
 - Doppelt einbezogene Werte heben sich gegenseitig auf
 - Adaption der „Fenstergröße“ an Signalverlauf möglich

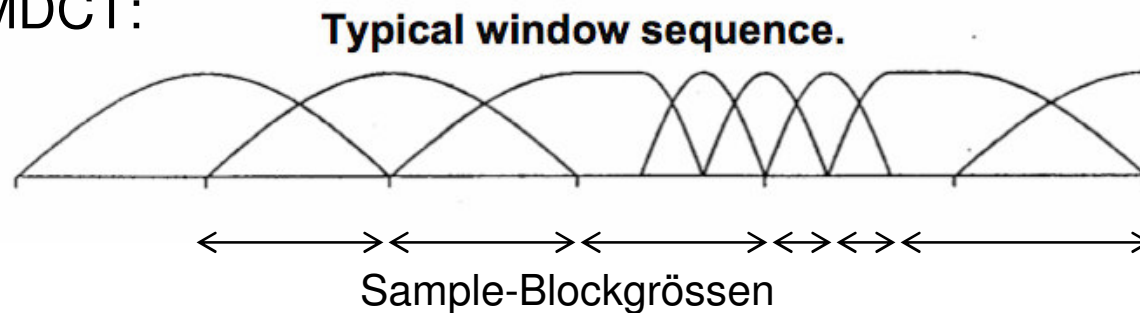
Überlappungen der Fenster bei MDCT:



Modified Discrete Cosine Transform MDCT (2)

- Modified DCT
 - Adaption der „Fenstergröße“ an Signalverlauf möglich

MDCT:



- Bei MP3: 6-Sample-Blöcke (Transienten) und 18-Sample-Blöcke
 - 6 Samples: Gut für schnelle Änderungen (Transienten)
 - 18 Samples: Gute Frequenzauflösung (wenn Signal relativ stationär)

Stereophonie in MPEG-Audio

- Single Channel
 - Monosignale
- Dual Channel
 - Verschiedene Monosignale (z.B. Sprachsynchronisation)
- Stereo Coding
 - Separat codierte Stereosignale
- Joint Stereo Coding
 - Redundanzen im Stereosignal ausgenutzt
 - Linker Kanal und Differenz Links/Rechts
 - Frequenzabhängigkeit der Raumwahrnehmung
 - » Monosignal für tiefe Frequenzen

- Hinweis:
 - Räumliches Hören kann z.T.MPEG-Kompressionsverluste wahrnehmbar machen; spezielle Vorkehrungen nötig

MPEG AAC

- AAC = Advanced Audio Coding
 - Nachträglich zu MPEG-2 standardisiert
 - Verbesserte Fassung in MPEG-4
 - Nicht rückwärtskompatibel
- MPEG-2 AAC:
 - 48 volle Audio-Kanäle
 - Reines MDCT-Filter, keine Filterbank mehr
 - Stark adaptierende Fenstergrößen
 - Prädiktive Kodierung im Frequenzraum (Temporal Noise Shaping TNS)
 - » gute Kodierung für „Transiente“ (zeitweilige Pegelspitzen)
- MPEG-4 AAC:
 - Perceptual Noise Substitution: Rauschen-ähnliche Teile des Signals werden beim Dekodieren synthetisiert
 - Long Term Prediction: Verbesserte Prädiktionskodierung

Weitere Audiokompressionsverfahren

- Dolby AC-3 (Audio Code No. 3)
 - Prinzipiell sehr ähnlich zu den MPEG-Verfahren
 - Time-Domain Aliasing Cancellation (TDAC)
 - » Überlappende Fenster in einer MDCT
 - » Transformation so ausgelegt, dass sich Redundanzen im Folgefenster auslöschen
- ATRAC (Adaptive Transform Acoustic Encoding)
 - Sony-Verfahren, entwickelt für MiniDisc
 - Ebenfalls Aufteilung auf Teilbänder, MDCT, Skalierung
 - Hörbare Verzerrungen bei mehrfachem Kopieren
- Microsoft Windows Media Audio (WMA)
 - Nicht offengelegtes Verfahren mit recht hoher Kompression (CD-Qualität bei 64 kbit/s)

VORBIS

- Meist in Zusammenhang mit dem "Container"-Format (zur Datenspeicherung) *Ogg* benutzt, deshalb auch *Ogg-Vorbis*
- Offenes und kostenloses Audio-Kompressionsverfahren
 - Xiph.org Stiftung, OpenSource-Projekt
 - Reaktion auf Patentansprüche aus MP3
- Ähnlich AAC:
 - Reine MDCT
 - Signal wird in "Basis-Rauschen" und Rest aufgeteilt
 - » Angenehmeres Verhalten bei zu niedriger Bitrate als MP3
 - "Bitrate Peeling":
 - » Vorhandene Dateien in der Bitrate reduzieren

Einfachere verlustbehaftete Verfahren

- Stummunterdrückung (*silence compression*)
 - Ausblenden von Zeitbereichen mit Nullsignal
- μ -Gesetz-Codierung bzw. α -Gesetz-Codierung (u.a. in G.711):
 - Nichtlineare Quantisierung: leise Töne angehoben
 - Ähnlich zu Dynamischer Rauschunterdrückung in Audiosystemen
- Adaptive Differential Pulse Code Modulation (ADPCM)
 - Prädiktives Verfahren
 - Vorhersage des Signalverlaufs durch Mittelung über bisherige Werte
 - Laufende Anpassung der Quantisierungstiefe an Signal
 - Kodierung der Differenzwerte zur Prädiktion
- Linear Predictive Coding (LPC)
 - Vergleicht Sprachsignal mit analytischem Modell der menschlichen Spracherzeugung, codiert Modellparameter und Abweichungen von der Vorhersage (militärische Entwicklung)
 - Nur für Sprache, klingt „blechern“, hohe Kompression
 - Weiterentwicklungen, z.B. Code Excited Linear Predictor (CELP)