

6. Ton und Klang

- 6.1 Ton: Physikalische und physiologische Aspekte
- 6.2 Kompression von Audio-Signalen: MPEG-Audio 
- 6.3 Audio-Datenformate: Übersicht
- 6.4 Klangerzeugung und MIDI

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

Einfache 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)

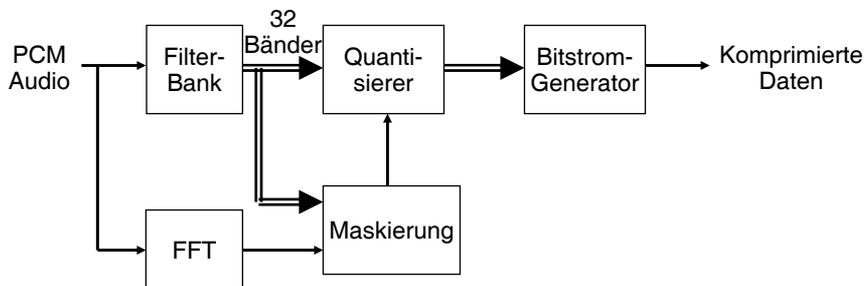
Einfache MPEG-Audio-Kompressionsverfahren

- Verlustbehaftete Audiokompression
 - basiert auf psychoakustischem Modell der Tonwahrnehmung
- 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 in drei „Layer“ (I, II, III) verschiedener Kompressionsstärke
 - Unabhängig von Wahl 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, Ogg-Vorbis (siehe später)

Kritische Bänder

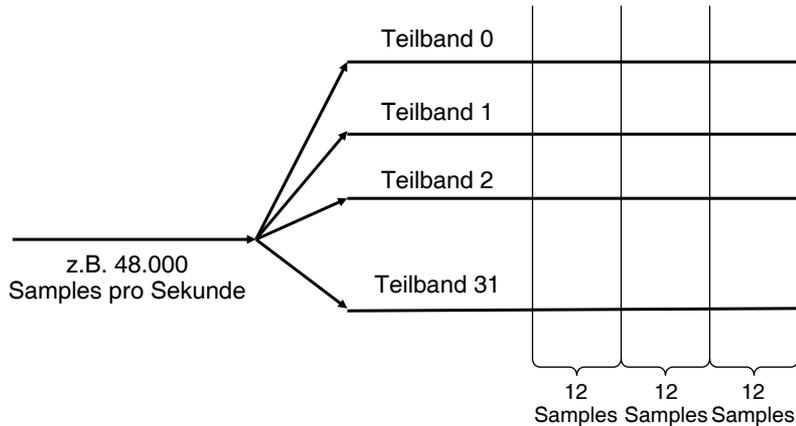
- Heinrich Barkhausen (1881–1956, Physiker, Dresden):
 - Einteilung des Hörbereichs in *kritische Bänder*
 - Breite der Bänder vergrößert sich mit der mittleren Bandfrequenz
 - Einheit für Bandbreite: *bark*
- Definition der bark-Breite:
 - Wenn $f < 500$: $1 \text{ bark} = f / 100$
 - Wenn $f \geq 500$: $1 \text{ bark} = 9 + 4 \log(f / 1000)$

Aufbau eines MPEG-Layer I Encoders



- Signal wird in 32 Frequenzbänder aufgeteilt
- Für die Maskierung mit einem psychoakustischen Modell:
 - Audiosignal durch Fast-Fourier-Transformation (FFT) genauer aufgelöst (512 Frequenzen bei Layer I, 1024 bei Layer II/III)

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 = 364 Samples

Maskierung

- Aus dem Originalsignal und dem psychoakustischen Modell ergibt sich eine Hörbarkeitsschwelle für jedes Band
- Signale in den einzelnen Bändern werden nur codiert, wenn sie oberhalb der Maskierungsschwelle liegen, und dann nur mit einer möglichst geringen Bitanzahl (1 Bit Ungenauigkeit entspricht 6 dB Rauschen)
- 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

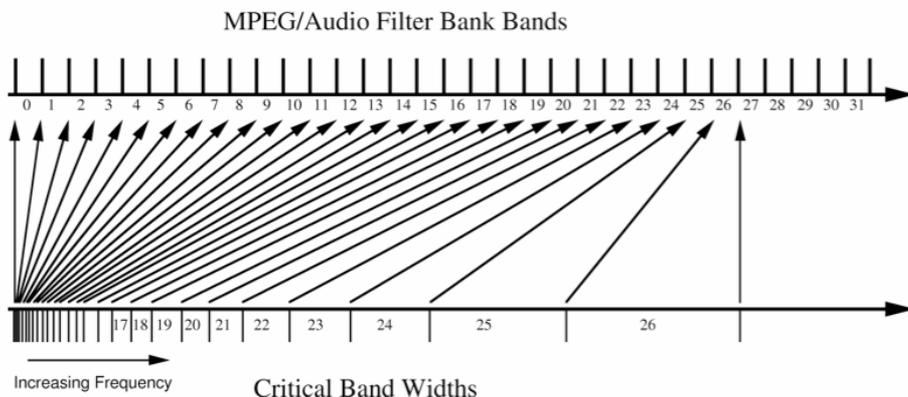
- Pegel in Band 8 ist 60 dB
-> Maskierung von 12 dB in Band 9 und 15 dB in Band 7
- Pegel in Band 7 ist 10 dB --> Weglassen!
- Pegel in Band 9 ist 35 dB --> Codieren!
wegen Maskierung 12 dB Ungenauigkeit zulässig, d.h. mit zwei Bit weniger codierbar

Unterschiede der MPEG Layer

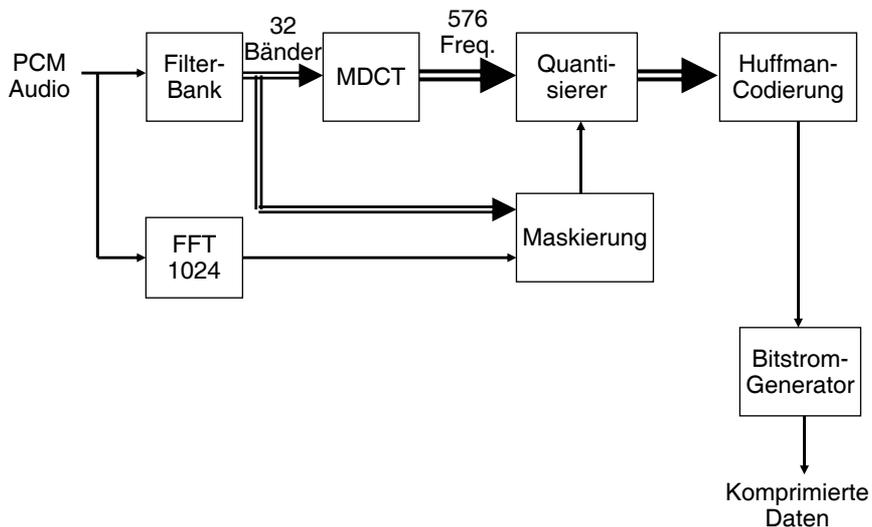
- Layer I:
 - 32 gleichbreite Teilbänder
 - Betrachtung nur eines Frames
 - Psychoakustisches Modell benutzt nur Frequenzmaskierung
- Layer II:
 - 32 gleichbreite Teilbänder
 - FFT mit 1024 Punkten
 - Betrachtung von drei Frames (jetzt, vorher, nachher)
 - Einfache Zeitmaskierung, verfeinerte Bittiefenzuweisung
- Layer III:
 - Teilbänder verschiedener Breite, ähnlich zu den kritischen Bändern
 - (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:



Aufbau eines MPEG-Layer III Encoders



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:
 - Es gibt Experimente, nach denen räumliches Hören z.T. die MPEG-Kompressionsverluste wahrnehmbar machen kann

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

Bedeutung von MPEG

- Audio-Standard MP3
 - Extrem hohe Verbreitung
 - Tragbare MP3-Player
- Rundfunk
 - Digital Audio Broadcast (DAB) und Digital Video Broadcast (DVB) basieren auf MPEG(-2)
- DVD-Video
 - Filmcodierung nach MPEG-2

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)