EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

MATHEMATISCH-
NATURWISSENSCHAFTLICHE FAKULTÄT

Medieninformatik / Human-Computer Interaction

Grundlagen der Multimediatechnik

Audiokompression

14.01.2022, Prof. Dr. Enkelejda Kasneci

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Termine und Themen

22.10.2021	Einführung
29.10.2021	Menschliche Wahrnehmung – visuell, akustisch, haptisch, …
05.11.2021	Informationstheorie, Textcodierung und -komprimierung
12.11.2021	Bildverbesserung
19.11.2021	Bildanalyse
26.11.2021	Grundlagen der Signalverarbeitung
03.12.2021	Bildkomprimierung
10.12.2021	Bildkomprimierung
17.12.2022	Videokomprimierung Teil I
14.01.2022	Videokomprimierung Teil 2 + Audiokomprimierung
21.01.2022	Videoanalyse
28.01.2022	Dynamic Time Warping
04.02.2022	Gestenanalyse
11.02.2022	FAQ mit den Tutoren
17.02.2022	Klausur, 14-16 Uhr, N10+N11

Audiokompression

- **Audiodaten nur schwer korrelierbar**
- **Keine erkennbaren Muster**
  → Wörterbuch-Kompression nicht erfolgversprechend
- **Datenwerte gleichverteilt**
  → Huffman-Kodierung nicht erfolgversprechend
- Domänenwandlung grundsätzlich machbar (DCT, FFT)
  → aber zu viele Koeffizienten, Datei wird größer...
- **Gesucht**: Kodierung mit ähnlichen Eigenschaften wie DCT
  - Verteilung der „Energie“ auf wenige Koeffizienten

Verlustbehaftete Audio-Kompressionsverfahren

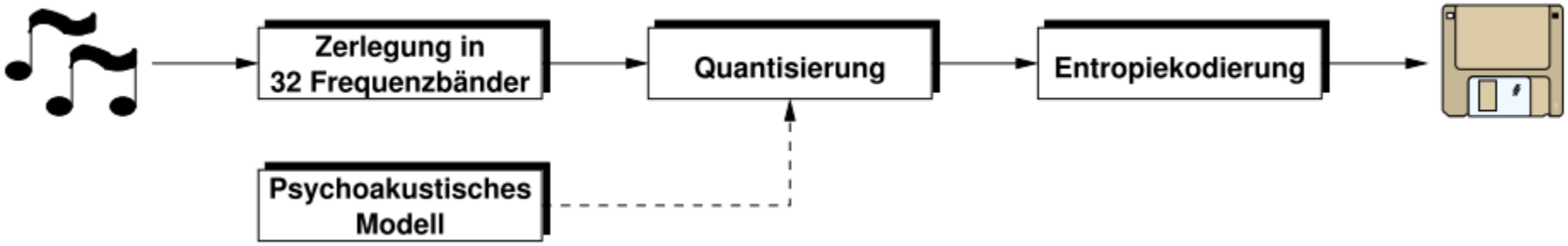
- **Verlustbehaftete Audiokompression**
  - Basiert auf psychoakustischem Modell der Tonwahrnehmung
  - Wichtigster Effekt: **Maskierte Bestandteile des Audio-Signals werden nicht kodiert**
  - Bekanntester Standard: MPEG Audio Layer III (MP3)
  - Moving Picture Expert Group, Untergruppe MPEG/Audio

MPEG-Audiokompression und erzielbare Kompressionsfaktoren

Verfahren	Bandbreite in kBit/s	Kompressions-faktor	Mbyte für 1 min. Audio
Audio CD	1400	1:1	10,58
MPEG-1-Layer I	384	3,6:1	2,88
MPEG-1-Layer II	256	5,5:1	1,92
MPEG-1-Layer III	128	11:1	0,962
MPEG-1-Layer III	64	22:1	0,481
MPEG-1-Layer III	16	88:1	0,120

- Verschiedene Qualitätsparameter einstellbar:
  - **CBR (Constant Bit Rate)** bei variable Qualität
  - **ABR (Average Bit Rate)** bei begrenzte Bandbreite
  - **VBR (Variable Bit Rate)** bei konstanter Qualität

Audiokodierung: Anwendungsbeispiel MP3



- **Zerlegung des Datenstroms** in Frames
- **Aufteilung des Frequenzbereichs** in 32 Subbänder
  - Layer I: gleiche Breite (625 Hz), nur Frequenzmaskierung
  - Layer II: gleiche Breite, Betrachtung von drei Frames (Zeitmaskierung)
  - Layer III: variable Breite
- **Lauteste Frequenzanteile verringern benötigte Auflösung**
- Differenz zwischen linkem und rechtem Kanal
- **Quantisierung gemäß psychoakustischem Modell**
- **Huffman-Kodierung**

MP3: Subband-Kodierung

- Annahme:
  Unterschiedliche Wichtigkeit von Frequenzbereichen
  - Isophone
  - Maskierung

→ **Anpassung von Auflösung, Quantisierung, Datenrate**

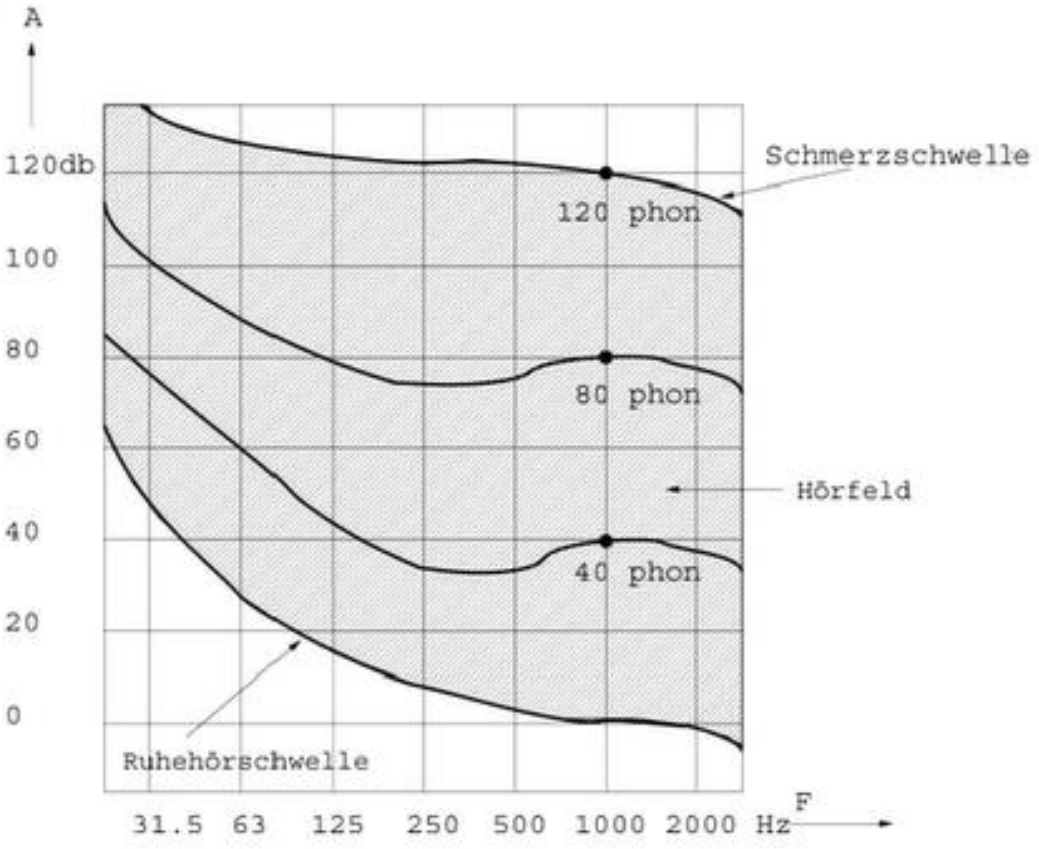
- **Beispiel Sprache**
  - Bass-Bereich und Höhen nicht maßgeblich für Verständlichkeit
  - Mittenbereich (300Hz-1200Hz) wesentlich

  → Zerlegung und Restaurierung des Signals über Vocoder

  → Telefonie-Codecs
- **Beispiel Musik**
  - Zerlegung des Audiosignals in diskrete Frequenzbereiche
  - Bewertung der Frequenzbereiche anhand Isophone und Quantisierung gemäß psychoakustischem Modell

  → Bestandteil der MP3-Kodierung

Menschliches Hörfeld: ca. 20-20.000 Hz bei 0 dB – 120 dB



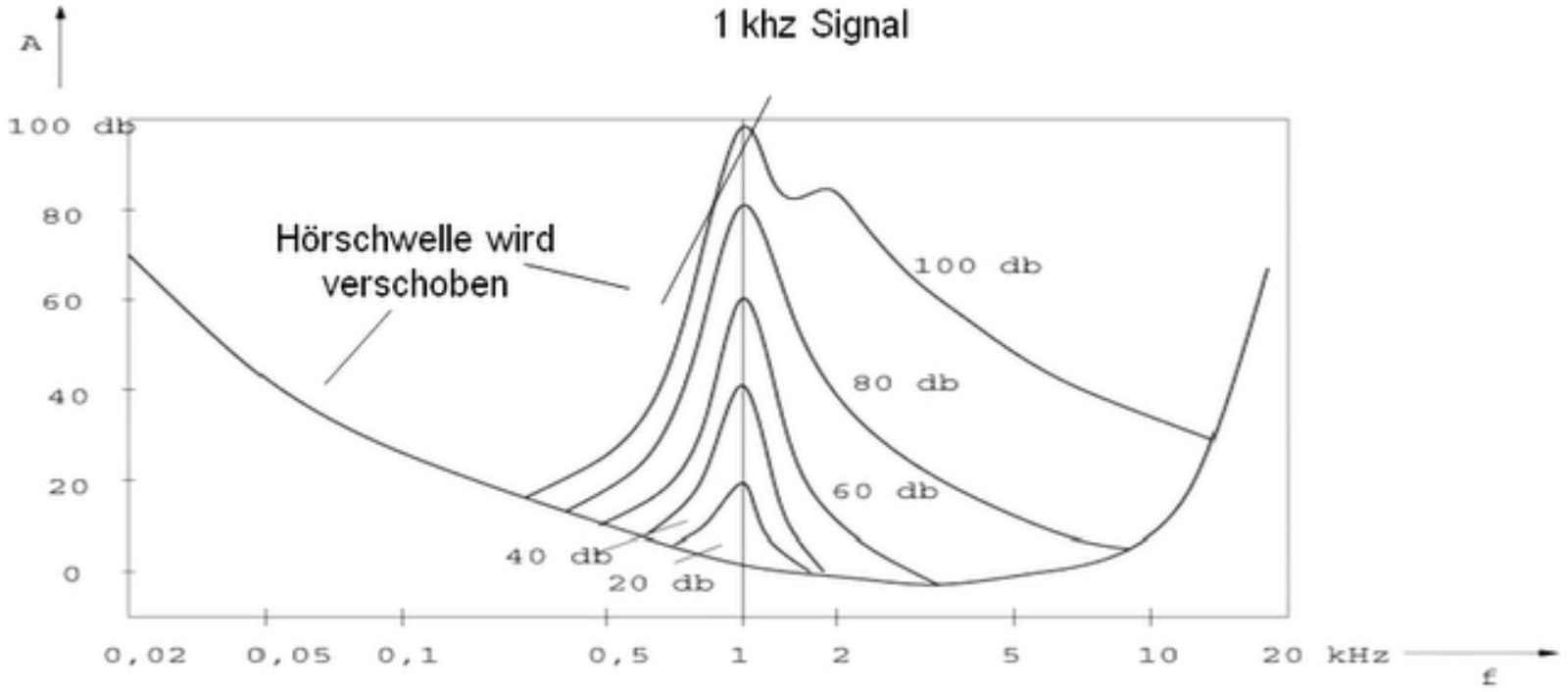
Kodiere nur menschliche Signale im Hörfeld

… auch innerhalb des Hörfelds müssen nicht alle Signale kodiert werden.

- **Simultane Verdeckung**:
  - starkes (lautes) Signal verdeckt (maskiert) gleichzeitiges schwaches (leises) Signal
- **Temporäre Verdeckung**:
  - starkes Signal verdeckt schwaches Signal nicht nur zeitgleich, sondern wirkt …
    - … gewisse Zeit nach (bis 200 ms)
    - … sogar einige Zeit vor (bis 50 ms, Ursache ist Trägheit des Hörvorganges)

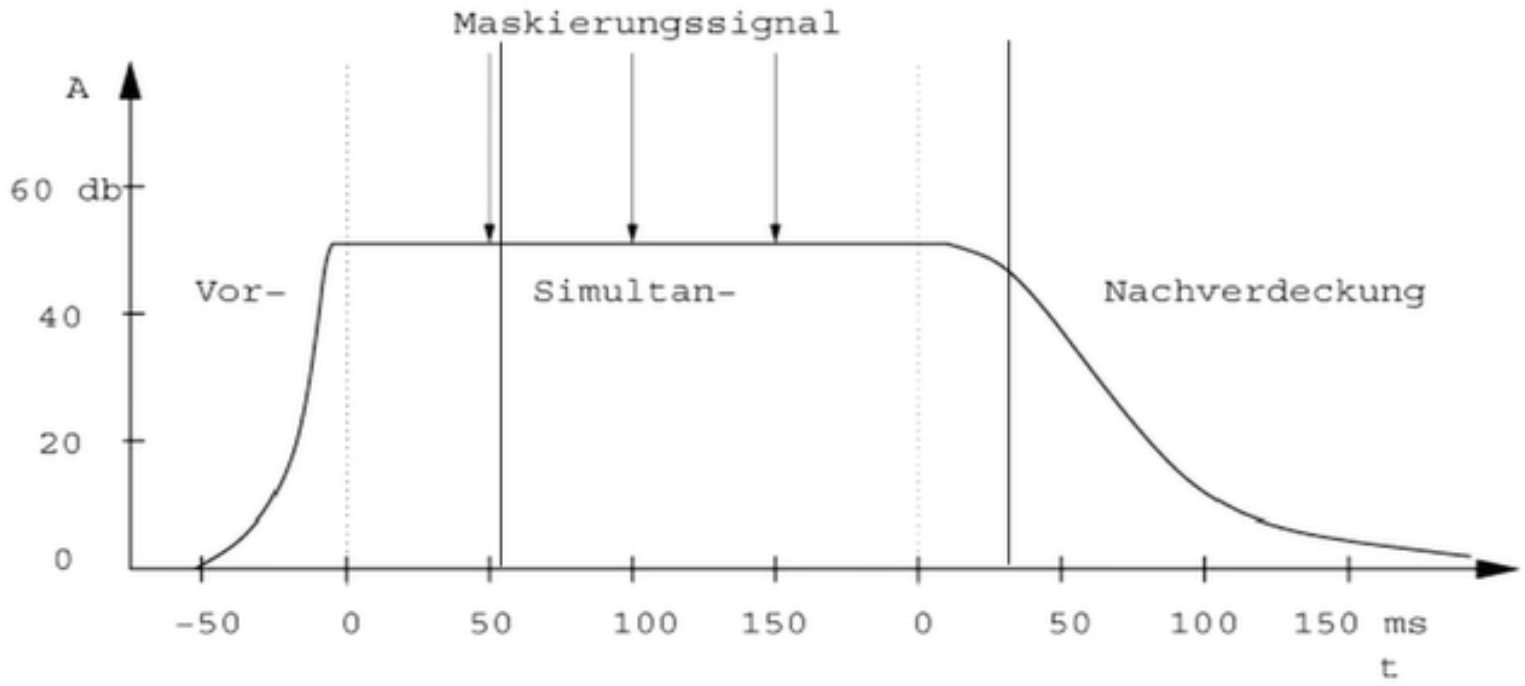
Simultane Verdeckung:

- starkes (lautes) Signal verdeckt (maskiert) gleichzeitiges schwaches (leises) Signal

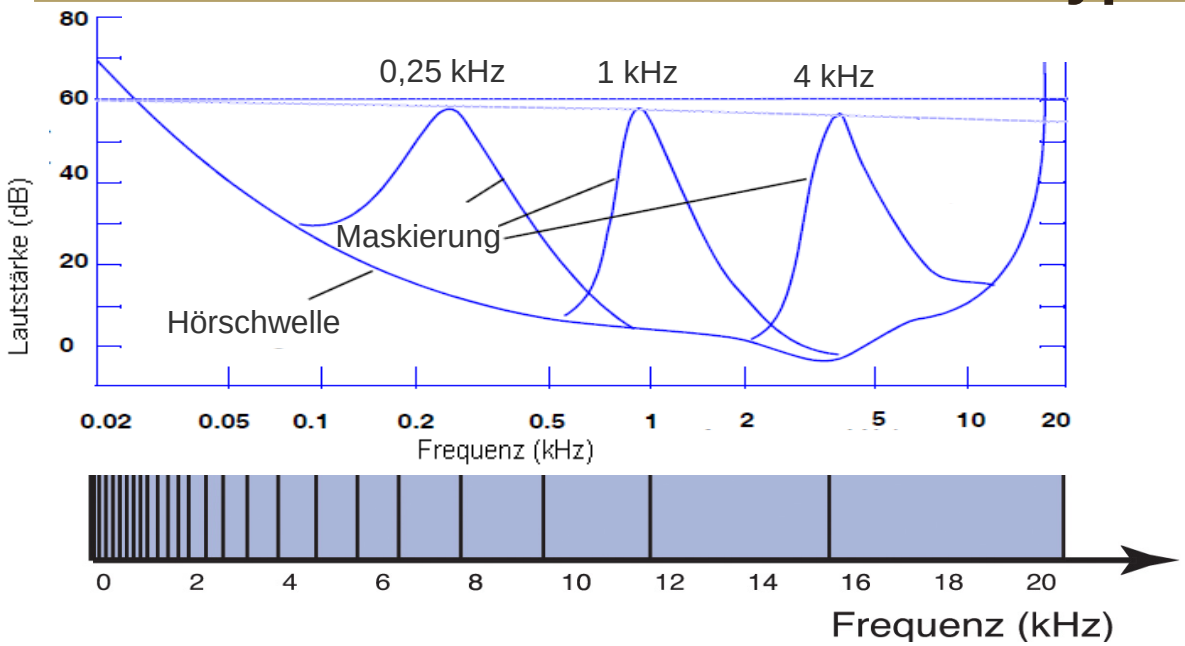


Temporäre Verdeckung:

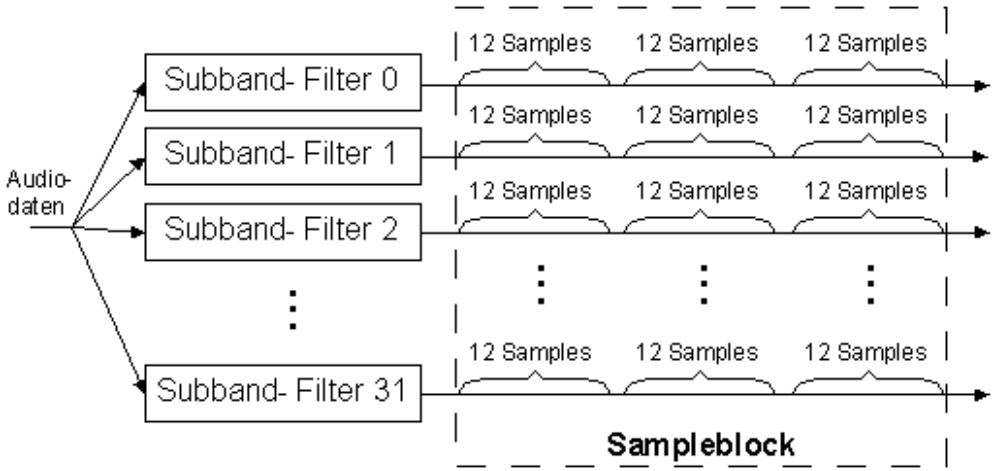
- starkes Signal verdeckt schwaches Signal nicht nur zeitgleich, sondern wirkt nach bzw. sogar vor



MP3: Frequenzbänder und Polyphasen-Filterbank



1. Teilung des Signals in 32 Subbänder
2. Überführung in den Frequenzbereich
3. Quantisierung nach entsprechenden Maskierungsmethoden
4. Reduktion der Information



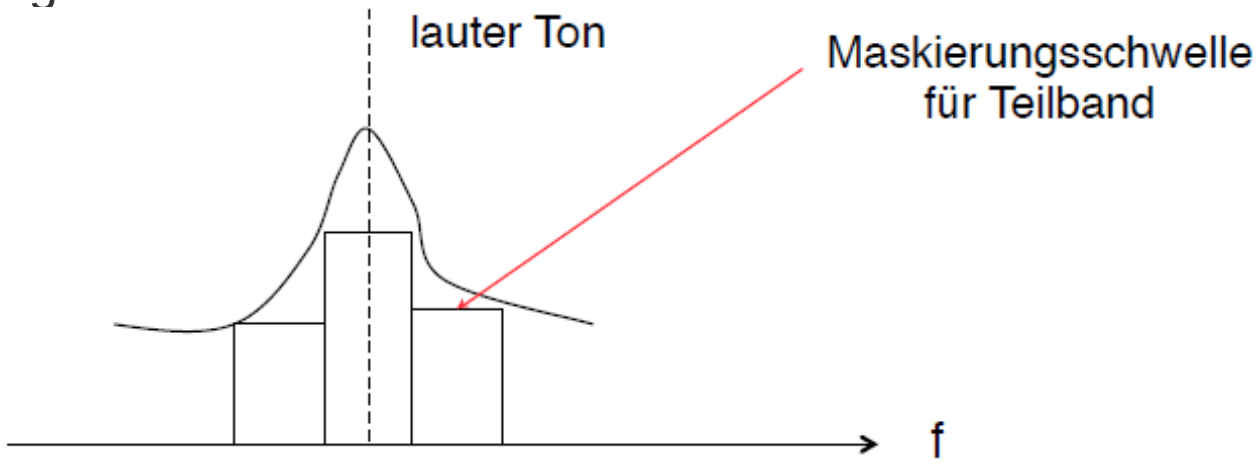
- Polyphasenfilterbank erlaubt **keine vollständige Rekonstruktion** (auch ohne Quantisierung)

→ verlustbehaftet

FFT zur Berechnung der Maskierungsschwelle

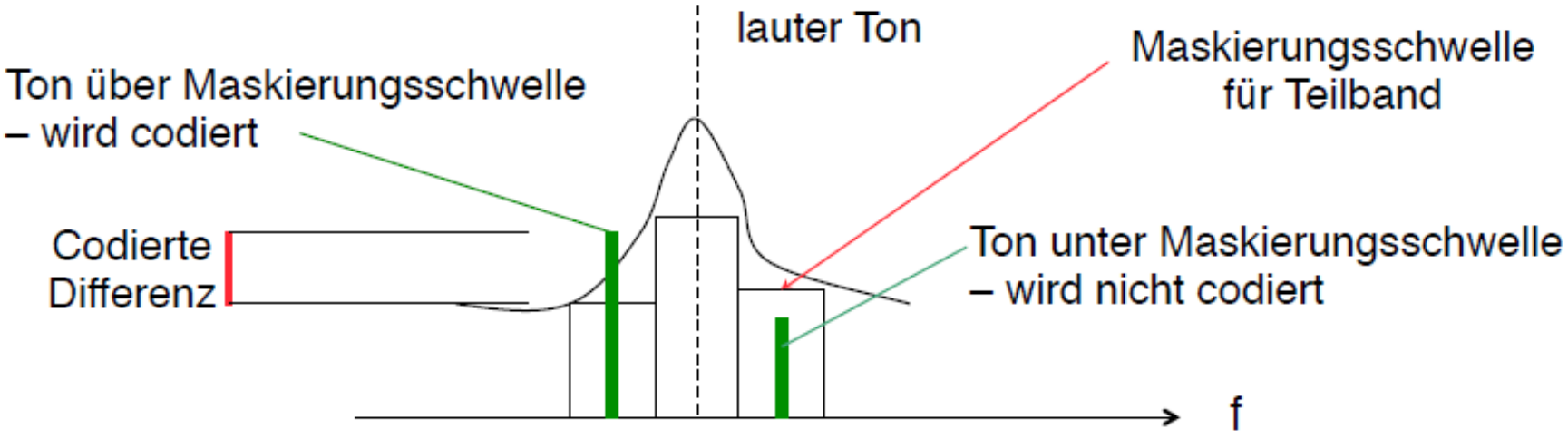
FFT = Fast Fourier Transformation

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



Maskierungsschwelle

- **Maskierungsschwellen** aus dem psychoakustischen Modell **werden mit tatsächlichem Signalpegel** (pro Teilband) **verglichen**
  - Verdeckte Signalanteile werden nicht codiert!
- Es genügt bei teilweiser Maskierung eine geringere Bitauflösung
  - Nur „Differenz“ oberhalb der Maskierungsschwelle wird wahrgenommen!



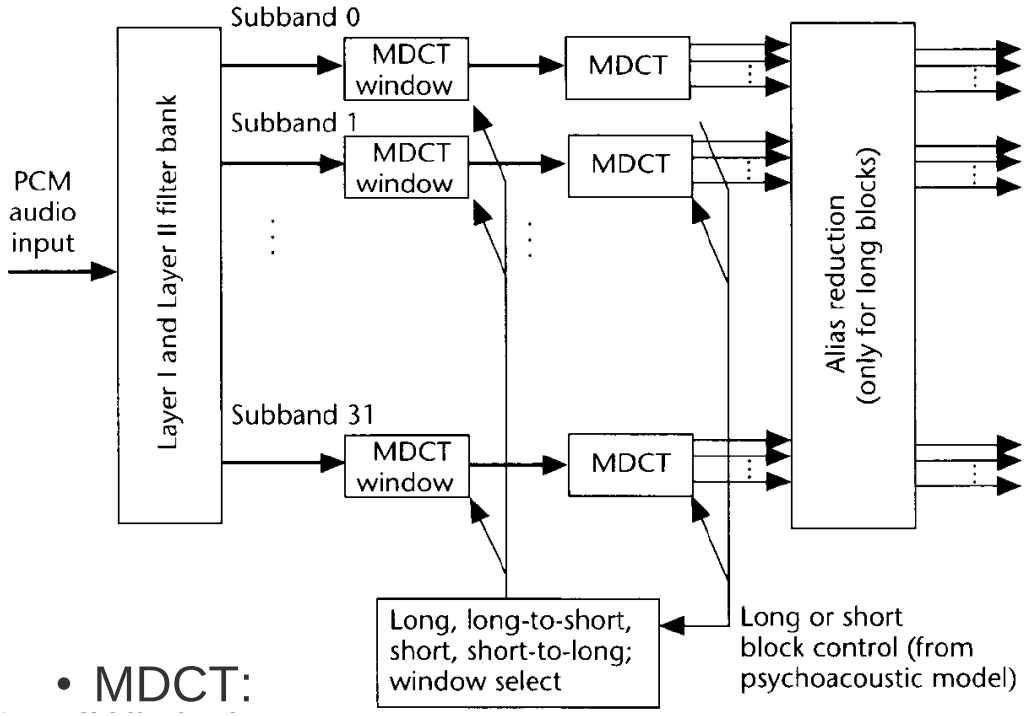
Maskierung: Beispiel

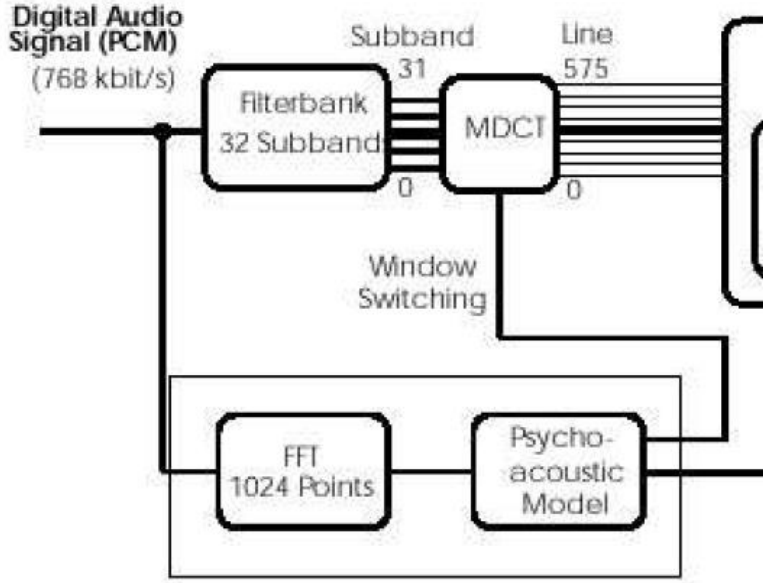
- Ergebnis nach der Analyse der ersten 16 Bänder:

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
  → kodieren
  - Wegen Maskierung 12 dB Ungenauigkeit (Rauschen) zulässig, d.h. mit zwei Bit weniger kodierbar

MP3: Hybrid-Filterbank (Polyphase + MDCT)





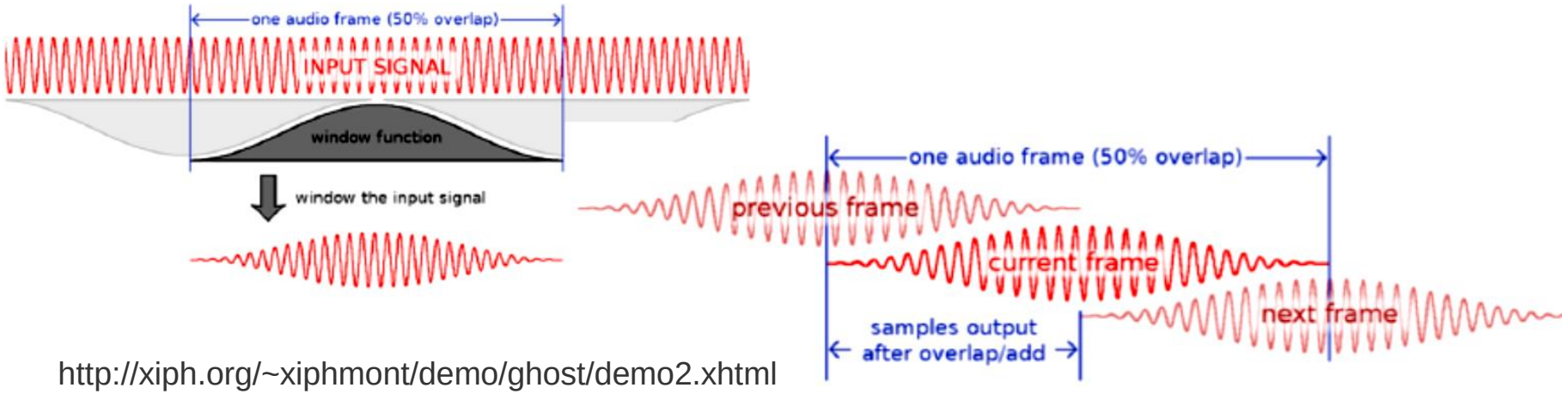
- MDCT:

$$X(m) = \sum_{k=0}^{n-1} f(k)x(k)\cos[\frac{\pi}{2n}(2k+1+\frac{n}{2})(2m+1)],\ m = 0...\frac{n}{2}-1$$

- MP3 spezifiziert zwei unterschiedliche Blocklängen für die MDCT
  - 18 Spektralpunkte oder 6 Spektralpunkte (Grundfrequenzen)
- Analyse der Maskierungseffekte unter Verwendung einer 1024-Punkte-FFT

Modifizierte Diskrete Cosinus Transformation MDCT (I)

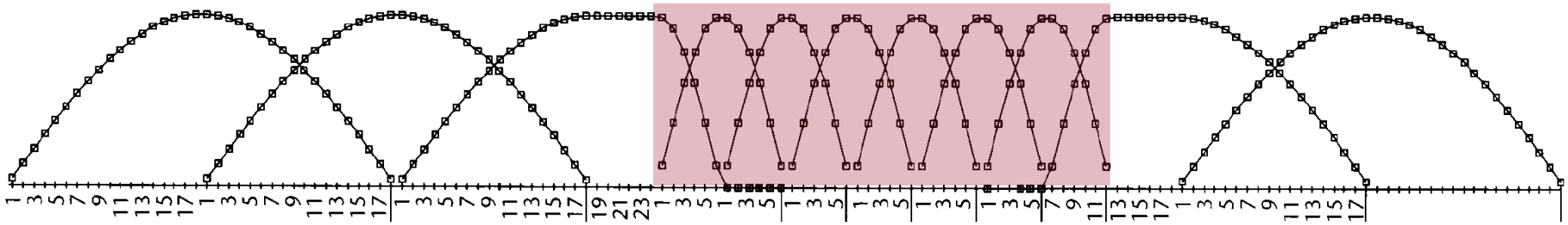
- **DCT**
  - bei Audio Probleme mit Artefakten an Blockgrenzen
  - Block = beliebiger Ausschnitt des Signals, wiederholt
- **Modifizierte DCT (MDCT)** (Princen, Johnson, Bradley 1987)
  - Überlappung der Cosinus-Funktionen um 50%
  - Vermeidung von Artefakten durch Blockgrenzen
  - Doppelte Signalanteile heben sich gegenseitig auf
    → Time-Domain Aliasing Cancelation (TDAC)



http://xiph.org/~xiphmont/demo/ghost/demo2.xhtml

Modifizierte Diskrete Cosinus Transformation MDCT (II)

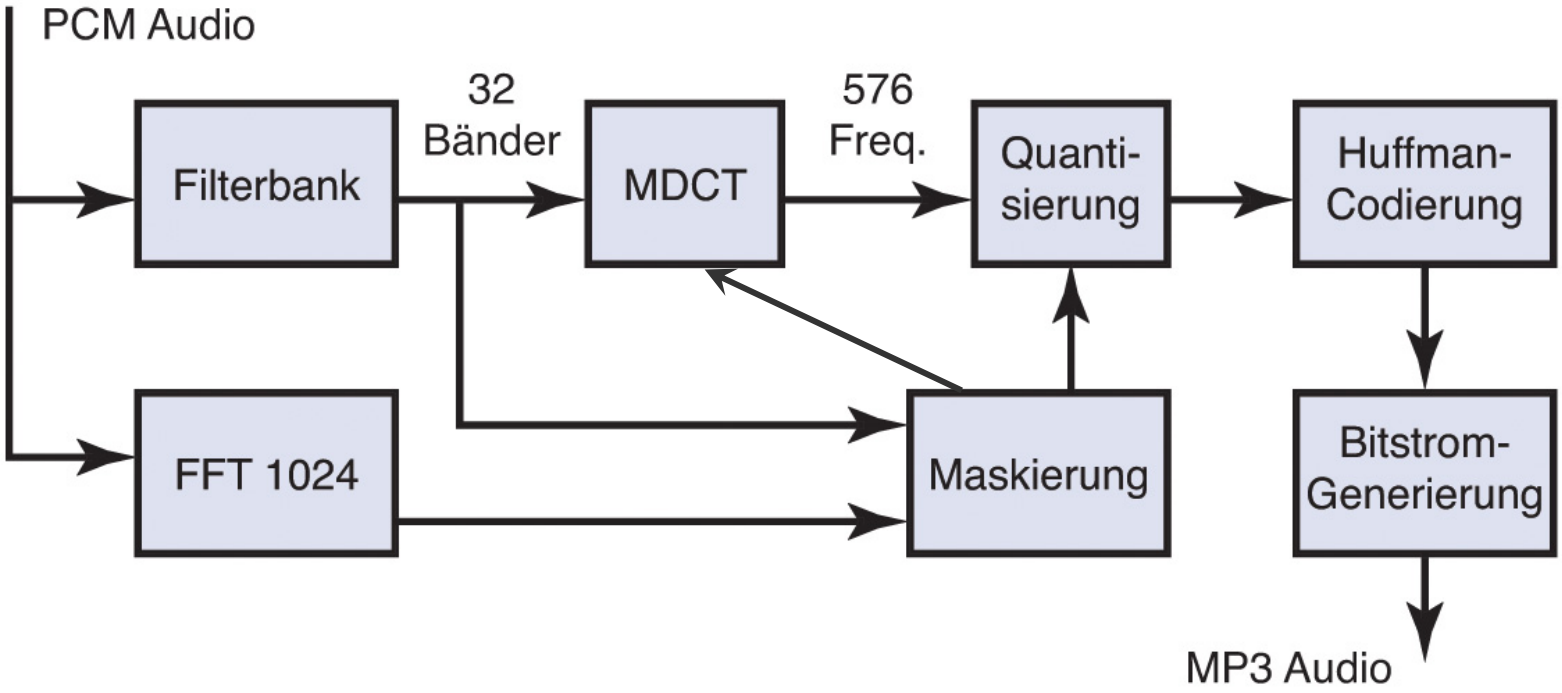
- Modified DCT
  - **Adaption der „Fenstergröße“ an Signalverlauf möglich**



Beispiel überlappender MDCT-Fenster mit unterschiedlicher Größe

- Bei MP3: MDCT wahlweise mit 12-Sample- und 36-Sample-Blöcken
  - 12 Samples → 6 Grundfrequenzen → $32 \cdot 6 = 192$ (3 Teilblöcke)
    Spektralkoeffizienten: **gut für schnelle Änderungen** (Transienten)
  - 36 Samples → 18 Grundfrequenzen → $32 \cdot 18 = 576$
    Spektralkoeffizienten: **gute Frequenzauflösung**
    (wenn Signal relativ stationär)
  - Übergangsblöcke: long-to-short, short-to-long

Aufbau eines MPEG-Layer III Encoders



- MDCT teilt jedes Teilband nochmals in 18 feinere Bänder auf

MPEG-4 Advanced Audio Coding

- **AAC = Advanced Audio Coding**
  - **Verbesserte Fassung des MPEG-2 Standards** im aktuellen Video-/Audio-Standard **MPEG-4**
- **MPEG-4 AAC**
  - alle Vorteile von MPEG-2 AAC
  - Perceptual Noise Substitution: Rauschen-ähnliche Teile des Signals werden beim Dekodieren synthetisiert
  - Long Term Prediction: Verbesserte Prädiktionskodierung
  - „Baukasten“ zur Konstruktion verschiedener Kompressionsverfahren
    (effiziente Sprachcodierung bis hin zu sehr hoher Musikqualität)
  - „Profile“, d.h. feste Kombinationen der Bausteine, Beispiele:
    - Speech Audio Profile, Synthetic Audio Profile, High Quality Audio Profile, Low Delay Audio Profile, Mobile Audio Internetworking Profile

Weitere Audiokompressionsverfahren

- **Dolby AC-3** (Audio Code No. 3)
  - Prinzipiell sehr ähnlich zu den MPEG-Verfahren
  - MDCT mit Time-Domain Aliasing Cancellation (TDAC)
- **ATRAC** (Adaptive Transform Acoustic Encoding)
  - Sony-Verfahren, entwickelt für MiniDisc
  - Ebenfalls Aufteilung auf Teilbänder, MDCT, Skalierung
  - Hörbare Verzerrungen bei mehrfachem komprimieren/dekomprimieren
- **Microsoft Windows Media Audio (WMA)**
  - Nicht offengelegtes Verfahren mit recht hoher Kompression (CD-Qualität bei 64 kbit/s)

Free Lossless Audio Codec (FLAC)

- **Freie, verlustfreie Audiokompression**
- **Fokus auf Streaming und Dekompression in Echtzeit**
- **Festkomma-Operation** (vermeidet Rundungsfehler)
- **Flexibel parametrisierbar**
  - Auflösung (4-32 Bit)
  - Sample-Rate (1-655350Hz in 1Hz-Schritten)
  - Kanalanzahl (1-8)
  - Kanalgruppierung (Stereo, Surround) zur Interkanal-Korrelation
  - Rice-Parameter $0 \leq M \leq 16$ → siehe nächste Folie

Free Lossless Audio Codec (FLAC)

- **Kompression**
  - **Blocking (Blockbildung)**
    - FLAC unterteilt die Daten jedes Kanals stets in Blöcke zu je 1000 bis 6000 Samples.
  - **Inter-Channel Dekorrelation**
    - Transformation der Links-Rechts-Kodierung in eine Mid-Side-Kodierung (mid = (left + right) / 2 und side = left – right)
    - Dynamische Auswahl des kleineren Frames.
  - **Modellierung**
    - Annäherung des Werteverlaufs eines Blocks
      - durch eine Polynomfunktion oder
      - mittels Linear Predictive Coding (Schätzung künftiger Werte über lineare Funktionen unter Verwendung eines Quellenfilters → Koeffizienten die Fehlersignal (Residual Energy) minimieren
  - **Residual Coding**
    - Das Fehlersignal (Unterschied zwischen dem tatsächlichen Signal und dem modellierten Signal) wird mittels Rice-Kodierung verlustfrei im Frame gespeichert
  - **Lauflängencodierung** für Blöcke mit identischen Samples (z.B. Stille)

Golomb/Rice-Kodierung

- Code-Variante für die effiziente Kodierung von Lauflängen
- Aufspaltung eines Eingabewerts N in zwei Teile q und r
- Rice-Code ist Untermenge des Golomb-Codes für $M = 2^k$
  - Rest $r = N \bmod M$ mit $M = 2^k, k \in \mathbb{N}$ (= letzte k Binärstellen)
  - Quotient $q = \left\lfloor \frac{N}{M} \right\rfloor = N >> k$ (Rechts-Shift um k Stellen)
- **Repräsentation einer Zahl durch**
  - r: Offset innerhalb des Behälters in verkürzter Binärkodierung (truncated binary coding)
  - q: Position des Behälters (bin) in unärer Kodierung (unary coding)
  - abschließendes Bit
- **Beispiel:**
  - Eingabe: $N = 10_{\text{dezimal}} = 1010_{\text{binär}}$, Rice-Kodierung mit $k = 2$
  - $r = 10_{\text{binär}}$, $q = N >> 2 = 2 \rightarrow 11_{\text{unär}}$, Ausgabe: $10110_{\text{binär}}$

Golomb/Rice-Kodierung

- **Einsatz als Quellenkodierung** zur Prädiktion
  - r liegt typischerweise in geometrischer Verteilung vor, d.h. kleine r sind häufiger als große r
  - Golomb/Rice-Code approximiert Huffman-Code, jedoch ohne Notwendigkeit einer Tabelle
- Beispiel: Rice-Codes für verschiedene Codierungsparameter k

x	binär	$k=0$	$k=1$	$k=2$	$k=3$
0	00000	0	0 0	00 0	000 0
1	00001	10	1 0	01 0	001 0
2	00010	110	0 10	10 0	010 0
3	00011	1110	1 10	11 0	011 0
4	00100	11110	0 110	00 10	100 0
5	00101	111110	1 110	01 10	101 0
6	00110	1111110	0 1110	10 10	110 0
7	00111	11111110	1 1110	11 10	111 0
8	01000	111111110	0 11110	00 110	000 10
9	01001	1111111110	1 11110	01 110	001 10
10	01010	11111111110	0 111110	10 110	010 10
⋮	⋮	⋮	⋮	⋮	⋮

Zusammenfassung

- **Verlustbehaftete Audiokompression (z.B. MP3, AAC)**
  - Psychoakustisches Modell ist fundamentaler Bestandteil
  - Aufteilung in Frequenzbänder
  - Analyse von Maskierungseffekten
  - MDCT weitverbreitet zur Frequenzbandzerlegung unter Vermeidung von Blockartefakten mit variabler Blocklänge
  - Einbringung von Verstärkungs- und Dämpfungsfaktoren zur Quantisierung
  - Huffman-Codierung der quantisierten Daten

- **Verlustfreie Kompressionsverfahren (z.B. FLAC)**
  - Approximation des Werteverlaufs eines Blocks und Speicherung des Differenzsignals zum Ursprungswert zur Fehlervermeidung