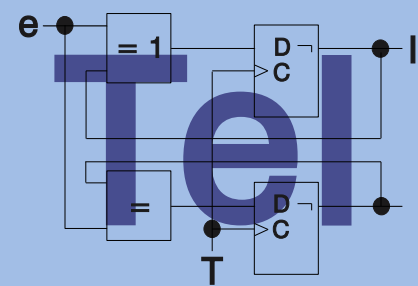


Funktionsweise von Hardware-Encoder und Decoder für MP3

Jan Schirok

s7642618@inf.tu-dresden.de



➤ Einführung

- Motivation
- Akustik
- Leistungen und Grenzen des Gehörs
- Maskierungseffekte

➤ MP3 – Verfahren

- Gesamtüberblick
- Wichtige Teilprozesse
- Umsetzung der Psychoakustik

➤ Umsetzung in Hardware

- Architektur
- Realisierung - Theorie und Praxis
- Ausblick

➤ Datenraten

- PCM 44100 Hz, Stereo, 16 bit: 1378 kBit/s
- MP3, hohe Qualität: 192 kBit/s (Faktor 7)
- MP3, fast CD-Qualität: 128 kBit/s (Faktor 11)

➤ Platz auf einer CD (700MB)

- PCM: 1 Stunde 10 Minuten
- 192er MP3: 8 Stunden 18 Minuten
- 128er MP3: 12 Stunden 25 Minuten

➤ Datenrate Video im Vergleich

- MPEG1 (relativ schlechte Qualität): 1535 kBit/s
- MPEG2 (sehr gute Qualität): 1-15 Mbit/s

➤ Kompression erforderlich

➤ Wichtige physikalische Größen der Akustik

- Schalldruck p [Pa] - Effektivwert der Schwankung (Ursache: Schall) des Luftdrucks um seinen Mittelwert
- Schallintensität I [Watt/m²] - Gesamtleistung durch Fläche von 1m² senkrecht zur Ausbreitungsrichtung

$$I = \frac{p^2}{\rho \cdot c}$$

- Schallpegel L [dB] - logarithmiertes Verhältnis zu einem Referenzdruck / einer Referenzintensität (meist Hörschwelle)

$$L_p = 20 \log \left(\frac{p}{p_0} \right) [dB] \quad L_I = 10 \log \left(\frac{I}{I_0} \right) [dB]$$

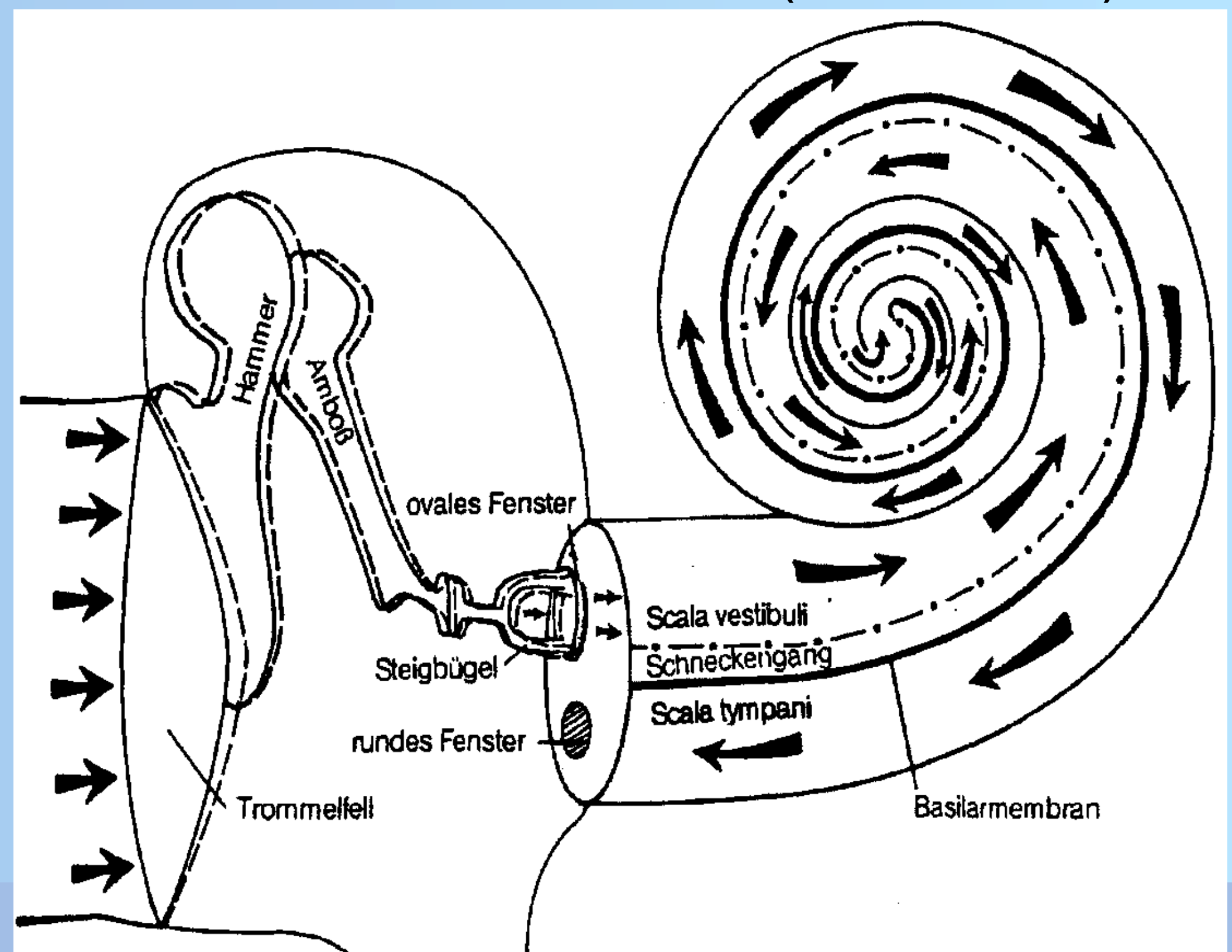
- Bezogen auf 20 μ Pa: Einheit dB SPL (sound pressure level)
- Pegellautstärke L_N [Phon] Töne mit 30 Phon werden so laut empfunden wie ein 1kHz Sinuston mit 30 dB
- Lautheit N [sone] - subjektiv empfundene Lautstärke

➤ Gehörgang

- Hörkanal, Trommelfell, Gehörknöchelchen
- Ohrschnecke (Cochlea) mit Basilarmembran
 - schneckenförmiger Aufbau
 - Sitz der Haarzellen (Hörrezeptoren)
 - Sinustöne führen zu lokaler Resonanz in der Schnecke (Ortstheorie)

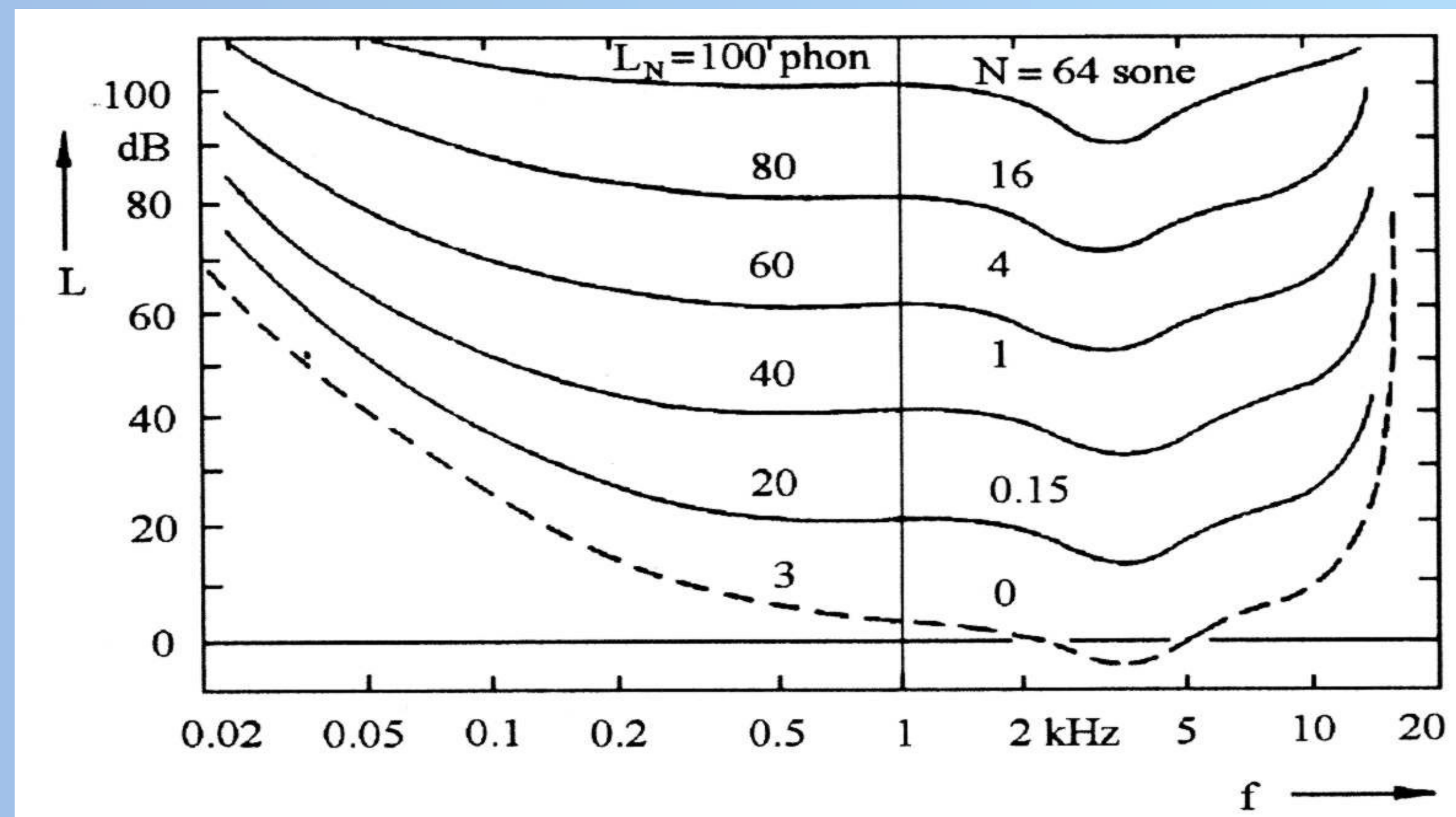
➤ Allgemeine Fakten

- Hörschwelle $p_0 = 20 \mu\text{Pa}$
bzw. $I_0 = 10 \text{pW/m}^2$
- Schmerzgrenze 1W/m^2
- Dynamik $D = 120 \text{dB}$
(12 Zehnerpotenzen)
- Frequenzbereich von
20 Hz bis 20 kHz



➤ Weitere Charakteristika

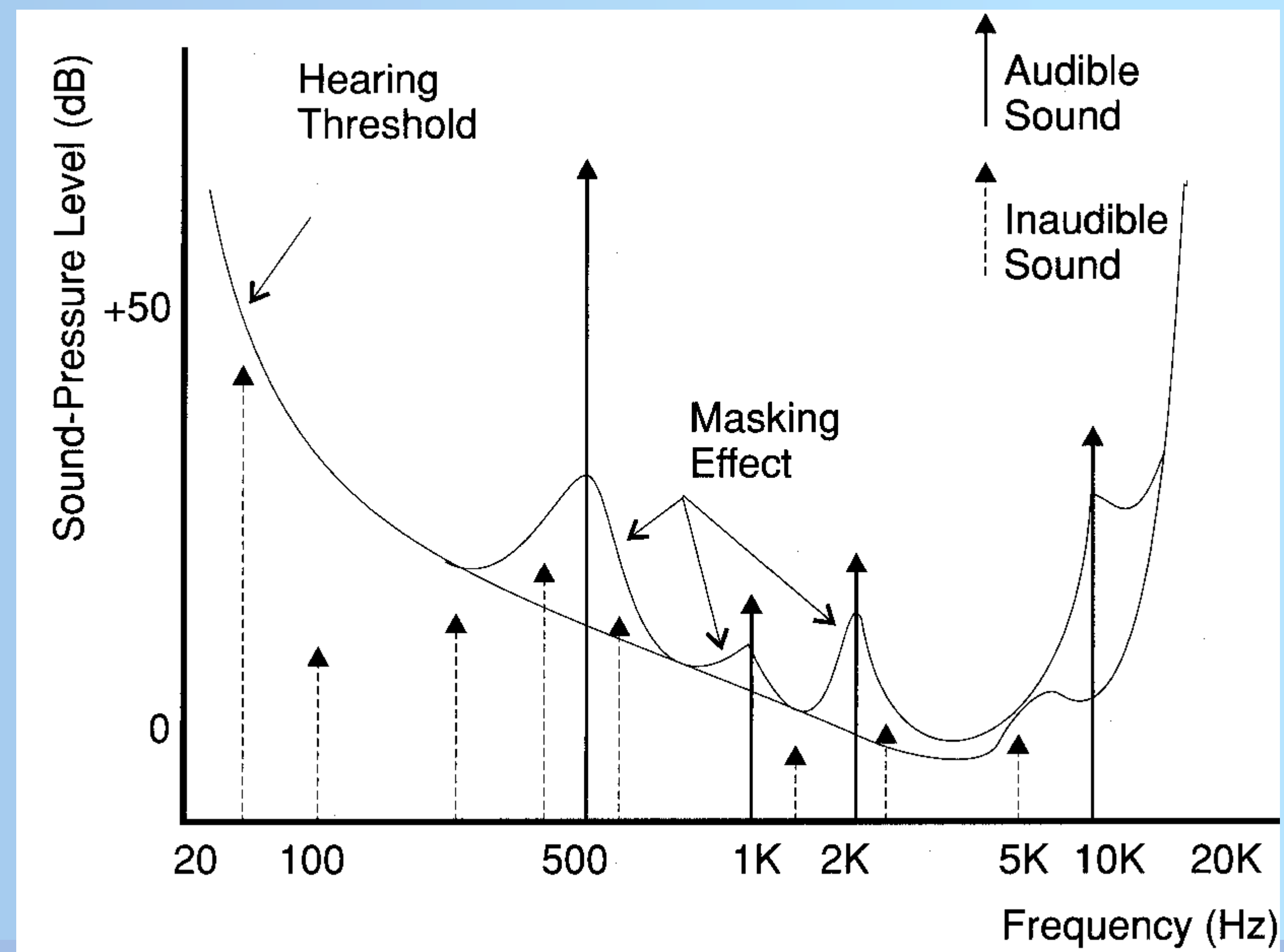
- Hörschwelle frequenzabhängig (zwischen 2 und 5 kHz am geringsten)
- subjektiv gleich laute Töne haben unterschiedliche Schallpegel
- Zerteilung komplexer Schalle in 24 Frequenzgruppen
- Frequenzabhängigkeit der Ortsauflösung



[1], S.65

➤ Maskierung im Frequenzbereich

- Überdeckung leiser Töne in Umgebung von lauten
- hohe Schallintensität erzeugt breiteren Resonanzbereich in der Ohrschnecke
- Zusammenführung der Maskierungen = Untergrenze des Hörbaren
- Unterscheidung nicht-tonaler und tonaler Geräusche
- Hörbeispiel



➤ Zeitliche Maskierung

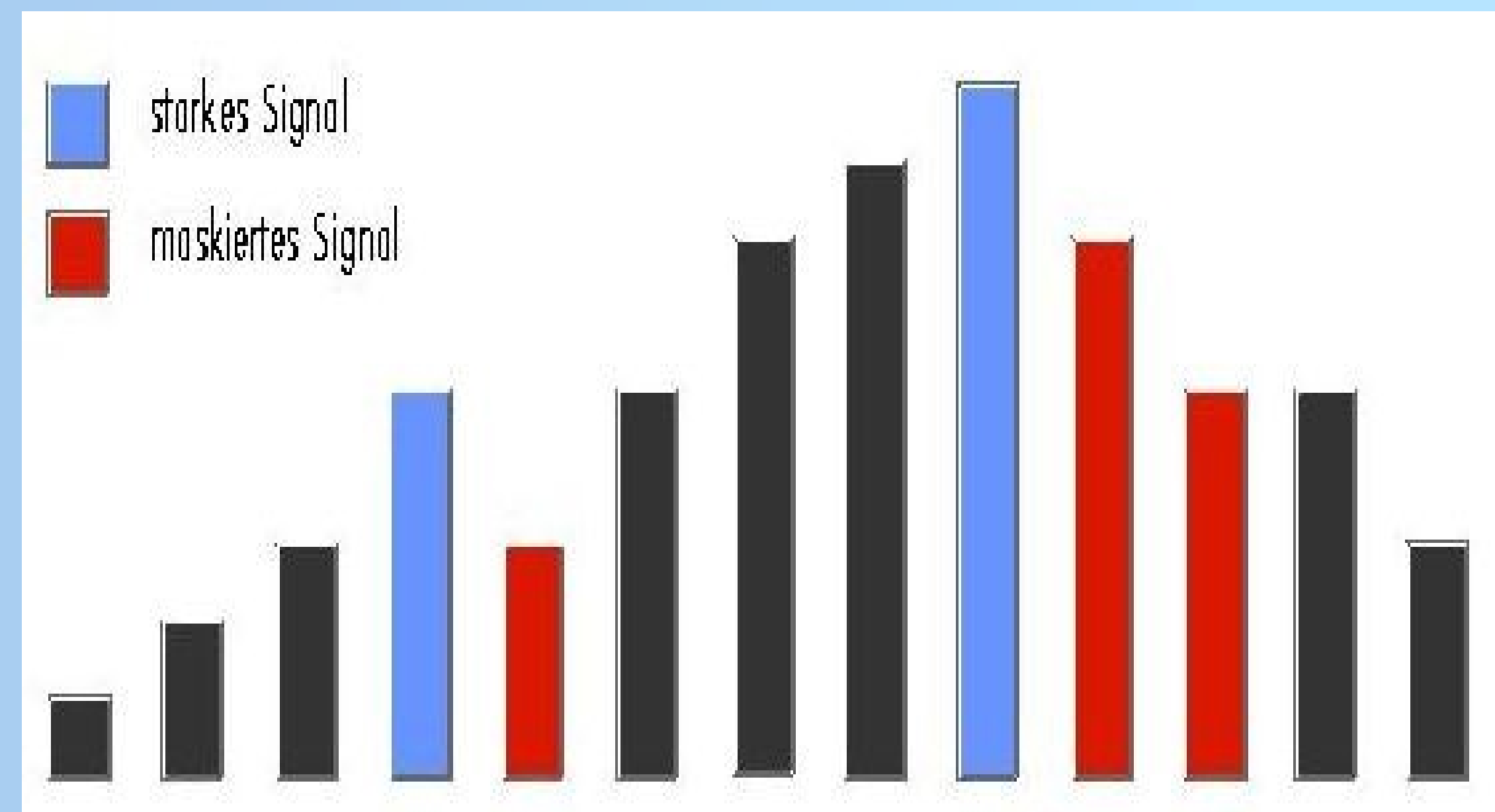
- Vorwärtsmaskierung von folgenden leiseren Tönen
- Rückwärtsmaskierung möglich, jedoch wesentlich kürzer
- Hörbeispiel

➤ Ansätze für Kompression von Musik

- Hörschwelle
- Nichtlinearität
- Frequenzmaskierung
- zeitliche Maskierung

➤ Umsetzung im Encoder

- Psychoakustisches Modell basierend auf empirischen Daten



[3]

➤ Entstehung

- Grundlagen über Musikkompression seit 1970er
- Zusammenarbeit von
 - Fraunhofer Institut für integrierte Schaltungen Erlangen (ab 1985)
 - u.a. Erlangen-Nürnberg-Universität, AT&T, Thomson
- seit 1992 Teil des MPEG-1 Standard als höchster Audio-Layer
- Festlegung des Namens „MP3“ durch interne Umfrage 1995
- 1998 erste tragbare MP3-Player (u.a. Rio PMP300)

➤ MPEG

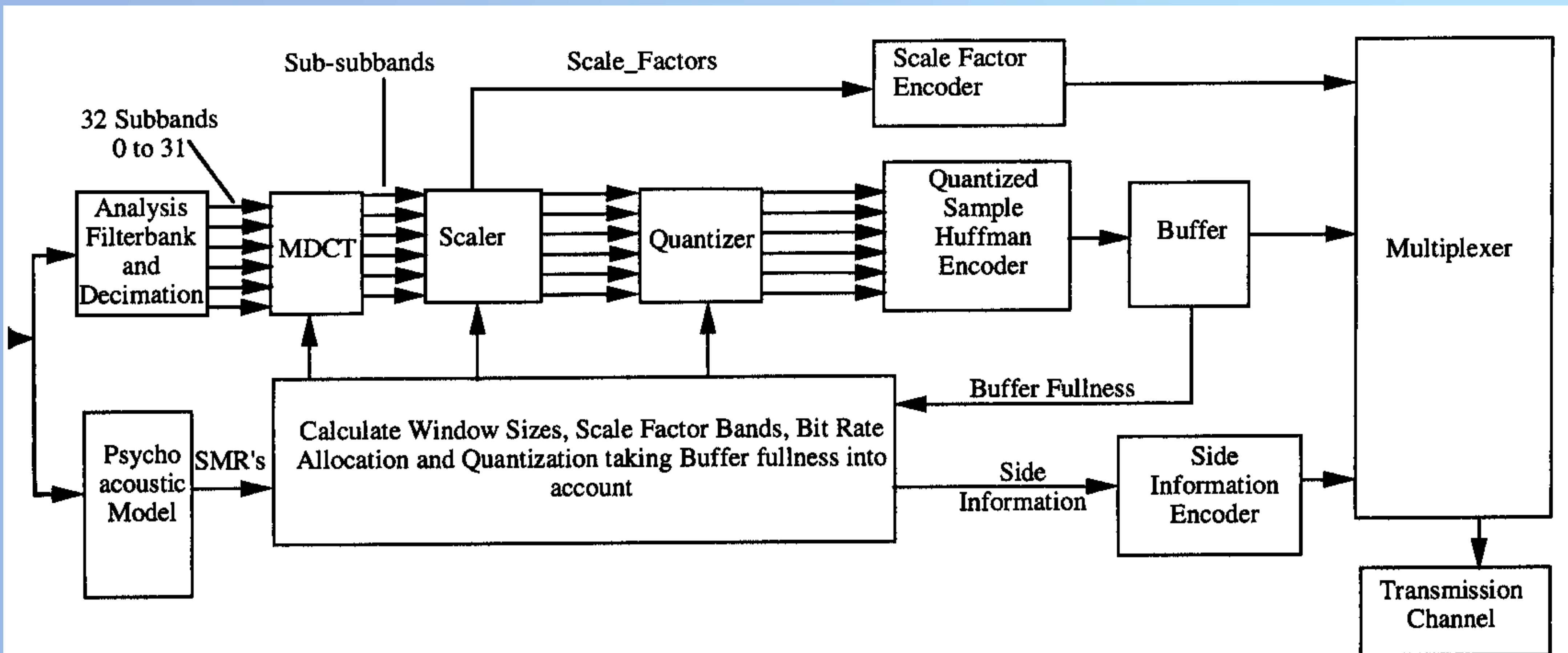
- Moving Picture Experts Group
- Standardisierung des Audio **Decoders** und des Formats

➤ Gebühr pro Endgerät

- 0.75 \$ Decoder, 1.25 \$ Encoder

➤ Gesamtblockdiagramm Encoder

- nur beispielhaftes Schema, nicht im Standard spezifiziert
- Decoding folgt umgekehrtem Weg (ohne Psychoakustik)



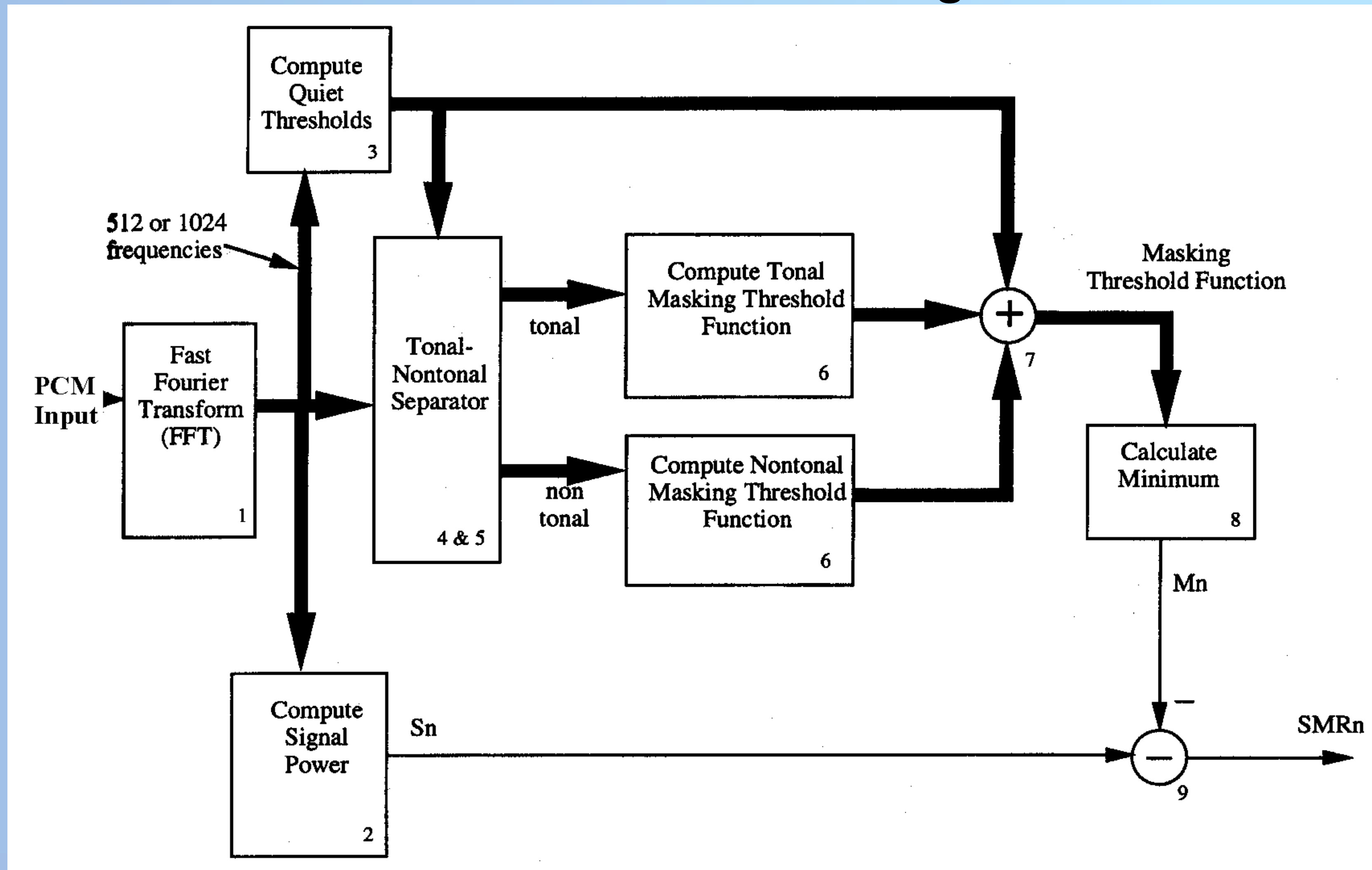
➤ Teilprozesse

- Filterbänke - Zerteilung in Frequenzbänder
- MDCT – Modifizierte diskrete Cosinus Transformation
- Psychoakustisches Modell – Berechnung der Maskierungsgrenze
- Skalierer – Normalisierungsfaktoren in Bändern
- Quantisierer – inkl. Potenzierung mit 0,75
- Huffman Encoder – aus fester Tabelle
- Puffer mit Bit-Reserve
- Multiplexer – Zusammenführung der Informationen und Daten

Umsetzung der Psychoakustik

➤ Psychoakustik

- Umsetzung sowohl als einzelner Block, als auch im gesamten Verfahren
- möglicher Aufbau des „Psycho-acoustic models“:



Architektur: DSP oder diskrete Schaltung?

- **Generelle Architekturmerkmale**
 - blockweise Abarbeitung (Frames)
 - getrennte Baublöcke
 - Iteration bei Pufferfüllung im Encoder
- **Pipelining**
 - Verarbeitung in einer diskreten Schaltung einzelner Baublöcke
 - weiterreichen unter den Baublöcken
 - höhere Latenz: negativer Einfluss auf Empfindung unwahrscheinlich – da Anwendungsbereich Musik
- **Betrachtung der Operationen**
 - viele Gleit- bzw Festkommaoperationen, Multiplikationen und Summen, einige Verzweigungen => DSP Kern
 - Filterbänke, u.a. – „SIMD-freundlich“

Architektur: DSP oder diskrete Schaltung?

➤ Decoder

- sowohl Realisierung in Baublöcken als auch als DSP denkbar

➤ Encoder

- aufgrund höherer Komplexität eher als DSP-Realisierung
- Echtzeitfähigkeit fordert hohe Rechenleistung
- Psychoakustik, Iterationen

➤ Tendenz

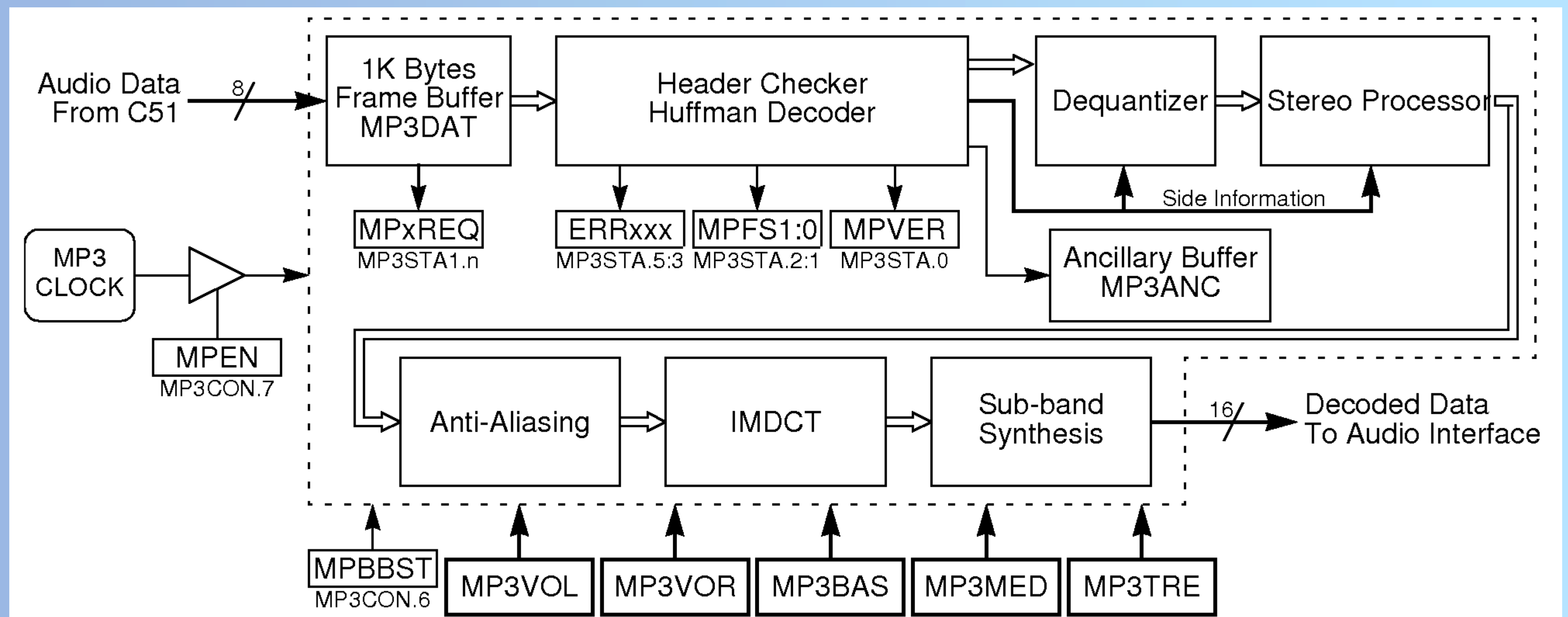
- Spezialisierung auf MP3 mit speziellem Datenpfad ungünstig
- Wechsel zu DSP Architektur
- größere Formatauswahl (WMA, AAC, Ogg Vorbis, ...)
- Encoder: generell mit schnellen, drosselbaren DSPs ausgestattet (z.B. MAS3587)

- Beispiele der Umsetzung in Hardwaredecoder
- Baublock-basierte Lösung
 - Atmel AT89C51SND1 (All-In-One Lösung für MP3-Player)
 - „hardwired Decoder“
 - zweiter Nachfolger AT85C51SND3 bereits mit DSP
- DSP basierte Lösungen
 - überwiegende Anzahl
 - repräsentativ MAS3507, STA013, VS1001K – dedizierte Decoder
 - VS1001K mit DAC und Verstärker
 - MAS3507 & VS1001K mit Platz für eigene Programme

Baublock-basiert - AT89C51SND1

➤ AT89C51SND1 und SND2

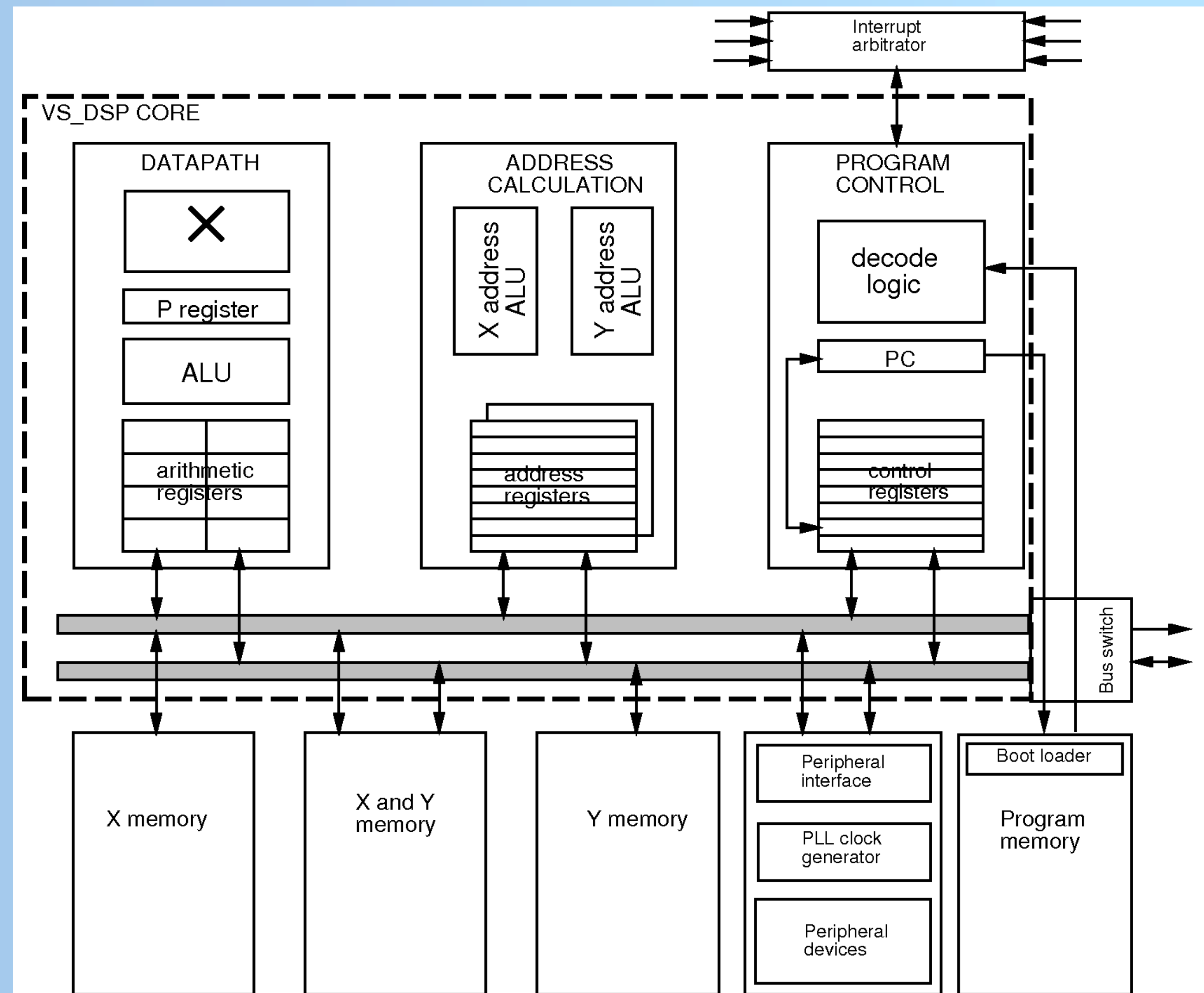
- „hardwired“ Decoder – festverdrahtete Baublöcke
- folgt standardmäßigem Dekoderprozess + Stereo Dekodierung
- gestattet im Prozess Anpassung von Lautstärke, Bass, ...



[5], S. 62

➤ DSP-Core von VLSI Solution Oy

- Trennung in
 - Datenpfad
 - Adressberechnung
 - Programmausführung
- Bearbeitung zweier Operanden gleichzeitig
- Multiplikation und Summation in einem Takt (MAC, MSU)



[6], S. 8

- mehr Formate (WMA, AAC) => DSP
- Audiokompression in allen Lebensbereichen (MP3-Player, Internet-Radio, Musikhandy...)
- Trennung Sprach- und Musikkompression
- Tendenz Hardware- => Softwaredecoder
 - stromsparende, leistungsfähige CPUs
 - Apple IPOD: Dual Core ARM, 90 Mhz
 - PC mit hochqualitativer Soundkarte
- Ablösung von MP3 nicht in Sicht

- [1] Jürgen Hellbrück: Hören – Physiologie, Psychologie und Pathologie, Verlag für Psychologie, Göttingen, 1993
- [2] Roy Hoffmann: Data Compression in Digital Systems, Chapman & Hall, London, 1997
- [3] Felix von Leitner: Die Kunst des Weglassens – Grundlagen der Audio-Kompression, c't – Magazin für Computertechnik, Ausgabe 3/2000, S. 130ff.
- [4] Barry G. Haskell, Atul Puri, Arun N. Netravali: Digital Video: An introduction to MPEG-2, Chapman & Hall, 1997
- [5] Datenblatt des AT89C51SND1:
http://www.atmel.com/dyn/resources/prod_documents/doc4109S.pdf
- [6] Datenblatt des VLSI DSP Cores, enthalten in:
<http://www.vlsi.fi/software/vskit/vskit116.zip>