

MPEG

Projektarbeit am Institut für elektronische Musik
von Johann Steinecker

Betreuung: Univ. Ass. DI Alois Sontacchi

Inhaltsverzeichnis

1. Der MPEG-Standard	4
2. Geschichtliche Entwicklung	4
3. MPEG-Formate	6
3.1 MPEG-1.....	6
3.1.1 MPEG-1-Video.....	6
3.1.2 MPEG-Audio Layer 1.....	6
3.1.2.1 Frames.....	6
3.1.2.2 Polyphase Filterbank.....	7
3.1.2.3 Encoder.....	9
3.1.2.4 Implementierung.....	11
3.1.2.5 Stereo- Kodierung.....	12
3.1.2.6 Formatierung des Datenstroms.....	12
3.1.2.7 Decoder.....	13
3.1.2.8 Verwendung von MPEG-1.....	13
3.1.3 MPEG-Audio Layer 2.....	14
3.1.4 MPEG-Audio Layer 3 (MP3).....	15
3.1.4.1 Lizenzen.....	15
3.1.4.2 Die Struktur von MP3.....	16
3.1.4.3 Frames.....	17
3.1.4.4 Header.....	18
3.1.4.5 Grösse bzw. Länge eines MP3-Frames.....	21
3.1.4.6 Konstante bzw. variable Bitrate.....	22
3.1.4.7 Implementierung.....	22
3.1.4.8 Filterbank.....	23
3.1.4.9 MDCT.....	24
3.1.4.10 Scalefactor Bands.....	26
3.1.4.11 Quantisierung und Codierung.....	26
3.1.4.12 Formatierung des Bitstroms.....	27
3.1.4.13 Stereo- Kodierung.....	27
3.1.4.14 Decoder.....	27
3.1.5 Psychoakustisches Modell.....	28
3.1.5.1 Zeitsynchronisation.....	28
3.1.5.2 Zeit-Frequenz-Transformation.....	29
3.1.5.3 Frequenzanteile gruppieren.....	29
3.1.5.4 Verteilungsfunktion.....	29
3.1.5.5 Tonanalyse.....	29
3.1.5.6 Untere Grenze der Schwellwerte.....	31
3.1.5.7 Schwellwertsuche für jedes Band.....	31
3.1.5.8 Berechnen der SMR.....	31
3.1.5.9 Abbildungen.....	32

3.2 MPEG-2.....	34
3.2.1 MPEG-2 Video.....	34
3.2.2 MPEG-2 Audio (ISO/IEC 13818-3).....	34
3.2.3 MPEG-2 AAC.....	35
3.2.3.1 Neuerungen.....	35
4. MPEG- Video.....	38
4.1 Verfahren der Videokompression.....	38
4.2 MPEG-1 Video.....	39
4.3 MPEG-2 Video.....	40
4.3.1 Profiles + Levels.....	40
5. MPEG- 3.....	40
6. MPEG- 7.....	40
7. MPEG- 21.....	41
8. MPEG- 4.....	41
8.1 Aufgaben von MPEG-4.....	41
8.2 Eigenschaften von MPEG-4.....	41
8.3 Die MPEG-4 Szene.....	43
8.4 Audio.....	44
8.5 Video.....	44
8.6 Erweiterung in MPEG-4 V2.....	45
9. Literaturverzeichnis.....	47
10. Anhang.....	49

In dieser Projektarbeit ist es das Ziel, alle MPEG-Formate zu beschreiben. Darüber hinaus werden die Formate MP3 (MPEG1/2-Layer III) und MPEG-4 detaillierter betrachtet. Es ist zusätzlich eine DVD entstanden, die Audiobeispiele für die unterschiedlichen Kompressionsformate enthält. (Einschließlich neuer Formate: ogg, vqf, mp3pro, aac, wma)

MPEG

1. Der MPEG-Standard

MPEG selbst ist nicht ein einziger Standard, sondern der Überbegriff einer Familie, die von der „Moving Picture Experts Group“ (MPEG), einem Standard für die Video, Audio und Systemkomprimierung und zugehörig zur International Standards Organisation (ISO), definiert wurde, um zu verhindern, dass viele einzelne zueinander inkompatible Systeme entstehen. MPEG wird in die *Hauptklassen* MPEG-1, MPEG-2 und MPEG-4 unterteilt. (Neu sind nun das MPEG-7 und MPEG-21 Format, Siehe Kapitel 6 und 7)

2. Geschichtliche Entwicklung von MPEG

Bereits Anfang der 80er Jahre wurde an der Erlanger Friedrich-Alexander-Universität mit der Komprimierung digitalisierter Klänge begonnen. Ziel war es damals, eine verbesserte Sprachqualität über Telefonleitungen zu erreichen. Erst später ging man dazu über, auch Musik (bzw. Bilder, Video) zu übertragen.

Die Verfahren wurden 1987 am Fraunhofer-Institut IIS-A (Institut für Integrierte Schaltungen) weiter entwickelt.

2.1 MPEG-1

Die Entwicklung von MPEG-1 begann 1988.

1989 wurde ein Patent angemeldet und erteilt (an das Fraunhofer-Institut in Deutschland und an die Thomson Multimedia SA in Frankreich), 1992 wurde es von der ISO in die Spezifikation der Moving Picture Experts Group integriert. (ISO/IEC IS 11172)

Der Audioteil von MPEG-1 (und MPEG-2) wurde in 3 Betriebsarten (Layer) unterteilt: Layer-1, Layer-2 und Layer-3, wobei Layer-3 die höchste Qualität bei gleicher Bitrate liefert.

Verwendete Abtastraten bei MPEG-1: 32 kHz, 44,1kHz und 48 kHz.

2.2 MPEG-2

Mit der Arbeit an MPEG-2 wurde 1990 begonnen.

MPEG-2 wurde 1994 standardisiert und enthält Neuerungen im Videobereich. (Kapitel 4.1) Im Audibereich gab es keine wesentlichen Erneuerungen (Filterbänke, usw.), es kamen lediglich Ergänzungen hinzu. (Kapitel 3.2 MPEG-2 Audio)

WICHTIG:

Zwischen MPEG-1 Audio und MPEG-2 Audio (MPEG-2,5 Audio) existiert prinzipiell kein Unterschied (keine neuen Filterbänke usw.), nur wurden in MPEG-2 Audio (und MPEG-2,5) zusätzliche Bitraten, Abtastraten, max. Anzahl der Kanäle,... usw. eingeführt! (Daher spricht man von MPEG-1/2 Audio oder kurz MPEG Audio)

Über die Komplexität der Coder gibt nur der Layer Auskunft.

ACHTUNG: MP3 steht nicht für MPEG-3, sondern für MPEG-1/2 Layer-3 !

2.3 MPEG-2.5

MPEG 2.5 ist eine Erweiterung von MPEG-2. Hier kamen nochmals niedrigere Abtastraten (MPEG-2,5 Layer 3 Audio: 8, 11,025 und 12 kHz) hinzu, aber keine wesentlichen Erneuerungen (Filterbänke,...). MPEG-2.5 ist jedoch kein offizieller Standard.

2.4 MPEG2-AAC

Mit der Entwicklung *neuer* Algorithmen für Stereo- und *Mehrkanal*-Sound (bis zu 48 Kanäle) entstand MPEG2-AAC („Advanced Audio Coding“). Es konnte die Bitrate im Vergleich zu MP3 weiter gesenkt werden.(Bei gleicher Qualität um ~ 33%.)

Technische Details zu MPEG2-AAC: siehe Kap. 3.2, S.35

2.5 MPEG-3

Ursprünglich sollte MPEG-3 den nächsten Schritt bei der Video- Codierung darstellen (HDTV, Kapitel 5)- es wurde aber an der Weiterentwicklung von MPEG-2 gearbeitet, die Neuerungen wurden in die MPEG-2 Spezifikation übernommen - und MPEG-3 somit bedeutungslos.

2.6 MPEG-4

Aufgrund der absehbaren Konvergenz von Telekommunikation, Computertechnik, Internet und Film, Video, Audio, TV begann bereits 1994 die Arbeit an MPEG-4.

Erste Veröffentlichung des MPEG-4 Standards: 1998 mit der Version 1, Version 2 folgte Ende 2000.

Im Audibereich kamen keine Neuerungen hinzu, es kann die ganze Bandbreite der bestehenden MPEG-Standards, einschließlich AAC, genutzt werden.

3. MPEG- Formate

3.1 MPEG-1

3.1.1 MPEG-1 Video

Das Ziel liegt darin, Bewegtbilder bis zu einer max. Datenrate von 1,5 Mbit/s zu übertragen bzw. zu speichern. Genaueres zu MPEG- Video: siehe Kapitel 4

3.1.2 MPEG Audio Layer 1: (ISO/IEC 11172-3)

- + Ein oder Zwei Audiokanäle
- + Abtastraten: 32 kHz, 44,1 kHz, 48 kHz
- + Bitraten: 32 kbit/s bis 448 kbit/s (Kilobit pro Sekunde)

Der Audibereich von MPEG-1 (und MPEG-2 bzw. MPEG-2.5) wird in *Unterklassen (Betriebsarten)*, den *Layern*, geteilt. Jeder Layer ist detaillierter als der jeweils darüberliegende und liefert dadurch höhere Qualität bei gleicher Bitrate.

3.1.2.1 Frames

MPEG-Audio Layer-1-Dateien werden in kleinere Einheiten, den *Frames*, unterteilt :

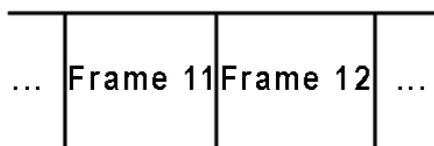


Bild 1: Frames bei MPEG-Audio Layer-1

Genauerer Aufbau eines Frames:

Header (32)	CRC (0,16)	Bit-Zuweisung (128-256)	Skalierungs- faktor (0-384)	Hauptdaten
-----------------------	----------------------	-----------------------------------	---	-------------------

Bild 2: Aufbau eines Frames bei Layer-1, wobei die Zahlen in den Klammern die Bitzahl angibt, die für dieses Feld verwendet werden kann. Beschreibung der jeweiligen Felder: Siehe Kap. Implementierung, S.8

Der Kodierungsvorgang wird nun beschrieben: (Anm. Layer-1 und Layer-2 werden grob beschrieben, genaue Details zu z.B. Psychoakustischem Modell, MDCT,... siehe bei der Beschreibung zu Layer-3)

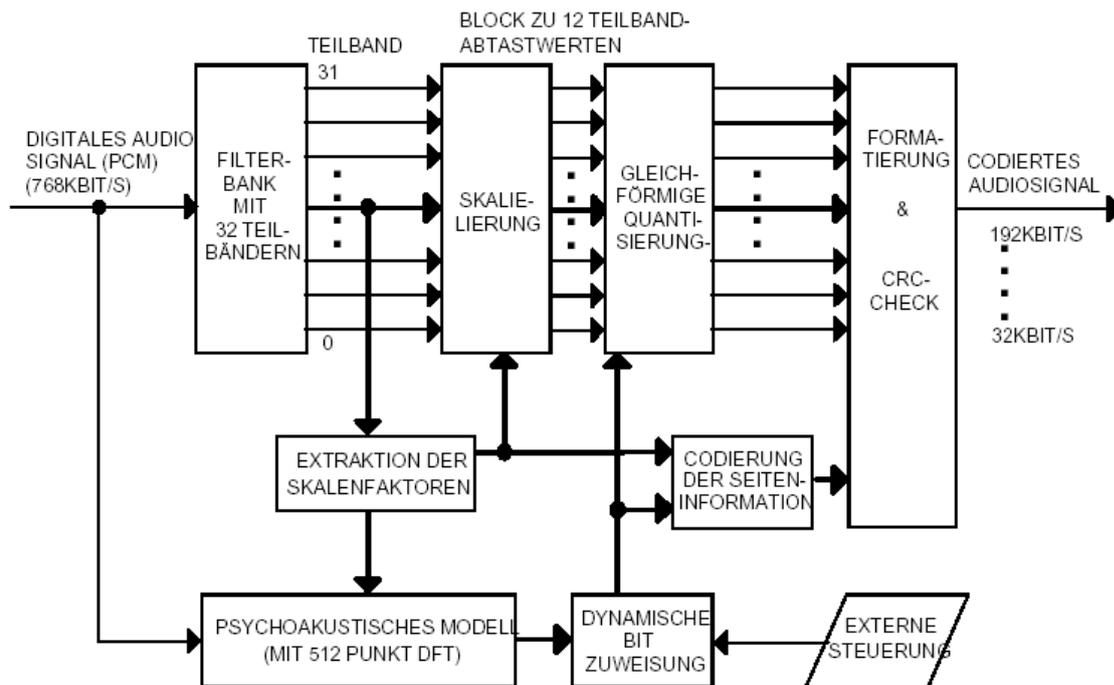


Bild 3: Blockschaltbild des MPEG Audio Layer-1 Coders

3.1.2.2 Polyphase Filterbank

Das digitale Audiosignal wird durch eine Filterbank in 32 Subbänder mit *gleicher Bandbreite* unterteilt (bei einer Abtastrate von 48 kHz : Bandbreite eines Subbandes beträgt 750 Hz, bei einer Abtastrate von 44,1 kHz beträgt die Breite eines Subbandes 625 Hz). Als Filterbank wird ein **Polyphase Filter** verwendet, **diese sind** relativ einfach aufgebaut, **bieten** eine gute Auflösung im Zeitbereich und Frequenzbereich, vergleichbar mit den Wahrnehmungseigenschaften des menschlichen Gehörs.

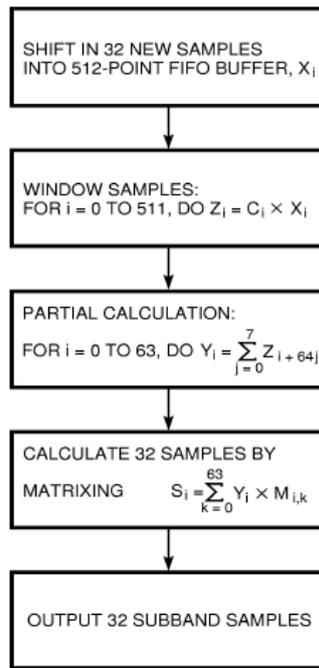


Bild 4: Berechnungsvorschrift der Polyphase Filterbank

Aus Bild 4 lässt sich folgende Formel für die Berechnung herleiten:

$$s_t(i) = \sum_{k=0}^{63} \sum_{j=0}^7 M(i)(k) * [C(k + 64j) * (k + 64j)] \quad \text{.....(3.1)}$$

$$M(i)(k) = \cos\left(\frac{(2i+1) * (k-16)\pi}{64}\right) \quad \text{.....(3.2)}$$

mit:

i: subband index, geht von 0 bis 31

$s_t[i]$: Filterausgangswert für Band i zum Zeitpunkt t; t ist ein Vielfaches von 32 (Samples)

$C[n]$: n-ter Filterkoeffizient für das Analyse Fenster, im Standard definiert; $0 \leq n \leq 511$

$x[n]$: Audio Eingangswert, wird aus einem 512 Werte Buffer gelesen

$M[i][k]$: Koeffizienten der Analyse Matrix

Verständlichere Schreibweise: *mir völlig unklar !*

$$s_t(i) = \sum_{n=0}^{511} x(\tau - n) * H_i(n) \quad \text{.....(3.3)}$$

$$H_i(n) = h(n) * \cos\left(\frac{(2i+1) * (n-16)\pi}{64}\right) \quad \text{.....(3.4)}$$

$$h(n) = \begin{cases} -C(n), & \text{Integerteil von } n/64 \text{ ist ungerade} \\ +C(n), & \text{sonst} \end{cases}$$

mit:

$x[\tau]$: Audio Sample zum Zeitpunkt τ ,

$h[n]$: Impulsantwort des Referenzfilters

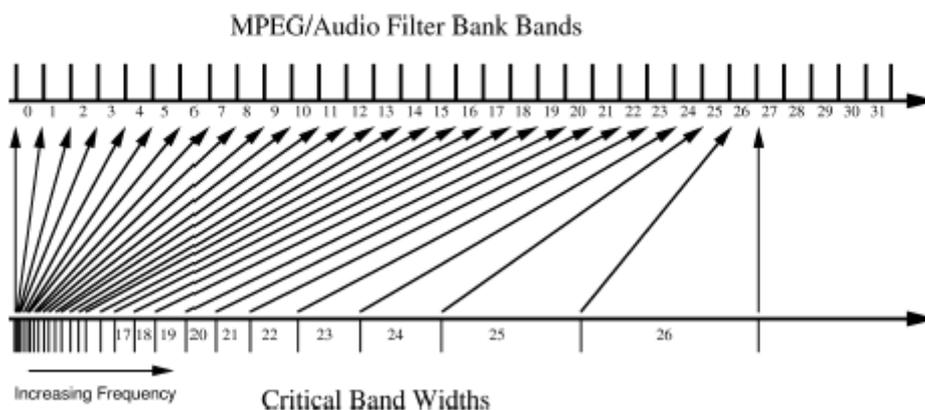
$H_i[n]$: Impulsantwort des zum i . Band verschobenen Referenzfilters

Hier sieht man, daß es sich dabei um FIR- Filter der Länge 512 handelt, wobei immer das selbe Filter verwendet wird - nur in das jeweilige Frequenzband verschoben.

3.1.2.3 Encoder

Es wurden 3 Kompromisse eingegangen:

1.) Die 32 gleich breiten Frequenzbänder stimmen nicht mit den kritischen Frequenzbändern des Gehörs zusammen, siehe Bild 5. Die Bandbreite der Subbänder ist bei niedrigen Frequenzen zu groß. Es kann dadurch nicht auf jedes kritische Band eingegangen werden, sondern wenn mehrere kritische Bänder in einem Subband liegen, wird zur Berechnung der notwendigen Quantisiererbits (siehe unten) das kritische Band mit der höchsten Rauschempfindlichkeit verwendet.



Band Nummer	Frequenz [Hz]	Band Nummer	Frequenz [Hz]
0	50	14	1970
1	95	15	2340
2	140	16	2720
3	235	17	3280
4	330	18	3840
5	420	19	4690
6	560	20	5440
7	660	21	6375
8	800	22	7690
9	940	23	9375
10	1125	24	11625
11	1265	25	15375
12	1500	26	20250
13	1735		

Band Nummer	Frequenz [Hz]
0	625
1	1250
2	1875
3	2500
4	3125
5	3750
6	4375
7	5000
8	5625
9	6250
10	6875
11	7500
.	.
.	.

Bild 5: Gegenüberstellung : Filterbreiten – kritische Bänder

2.) Die Filterbank und ihre Inverse sind keine verlustlosen Transformationen- der dadurch entstehende Fehler ist allerdings klein und praktisch nicht hörbar. Bild 6 zeigt den Frequenzgang von der Analyse Filterbank und der anschließenden Synthese Filterbank ohne Quantisierung.

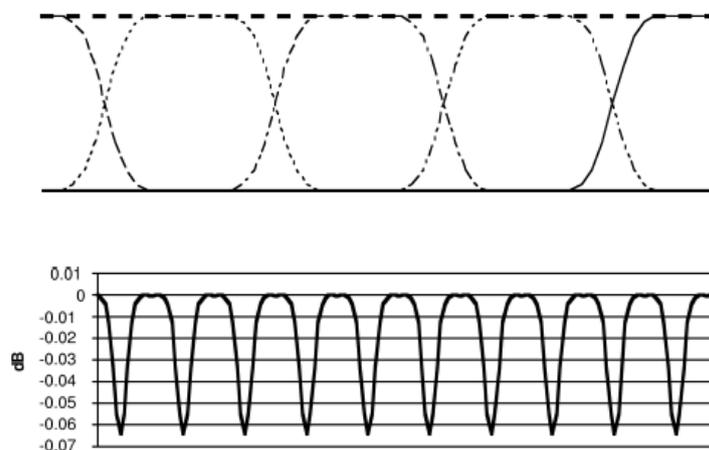


Bild 6: oben: Zusammengesetzter Frequenzgang für 3 Bändereiner Polyphase Filterbank; unten Normalisierter zusammengesetzter Frequenzgang der Filterbank in dB

3.) Benachbarte Frequenzbänder *überlappen* sich, d.h. ein Ton bei einer bestimmten Frequenz kann 2 benachbarte Subbänder beeinflussen- siehe Bild 7. Das führt zu einem geringeren Kompressionsfaktor, da ein Ton in zwei Bändern berücksichtigt wird.

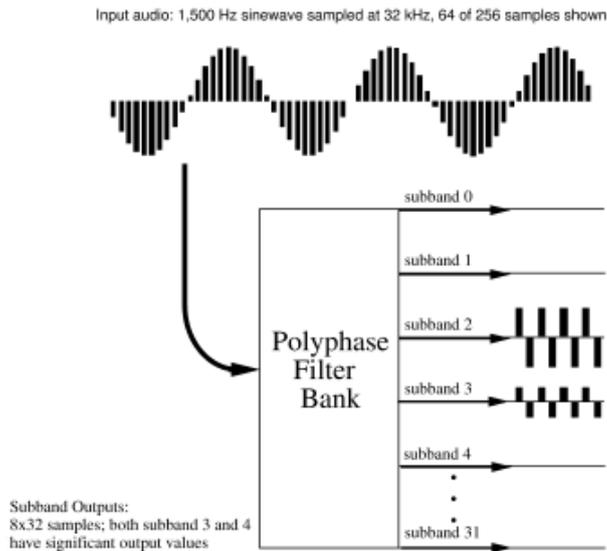


Bild 7: Aliasing Effekt: Reiner Sinuston beeinflusst 2 Subbänder

3.1.2.4 Implementierung:

Die Framegröße bei MPEG-1/2 Layer-1 beträgt 384 Samples pro Frame. Die Filterbank produziert alle 32 Eingangssamples 32 Frequenzsamples, für jedes Band einen. In jedem Band werden 12 Samples ($384 \text{ Samples}/32$) zu einer Gruppe zusammengefasst (siehe Bild 8).

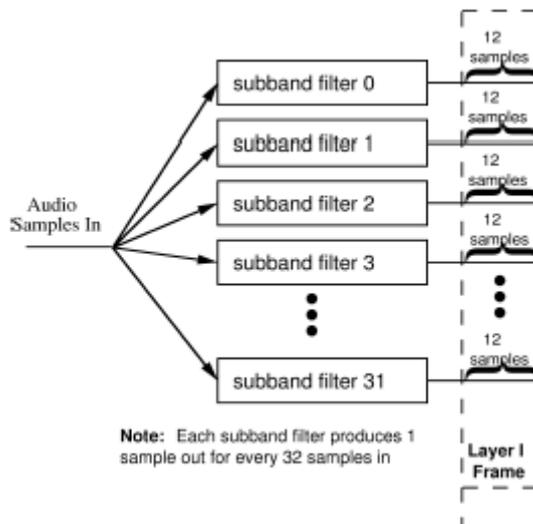


Bild 8: Zusammenfassen der Subband-Samples

Aus diesen wird der maximale Absolutwert bestimmt und mit 6 Bit kodiert, dieser Wert wird als Skalierungsfaktor bezeichnet. Jeder dieser Gruppen wird mit einer bestimmten Anzahl an Bits, entsprechend der vom Psychoakustischen Modell (siehe Kapitel 3.1.5) gesteuerten dynamischen Bitzuweisung quantisiert, und ein Skalierungsfaktor zugeteilt.

Für Layer 1 u. 2 wird das Mask To Noise Ratio folgendermassen berechnet:

$$\text{MNR}_{\text{dB}} = \text{SNR}_{\text{dB}} - \text{SMR}_{\text{dB}}$$

mit MNR = Mask To Noise Ratio, SNR = Signal To Noise Ratio, SMR = Signal To Mask Ratio.

Im MPEG/Audio Standard sind Richtwerte für das SNR bei bestimmten Quantisiererstufen enthalten und das SMR wird vom Psychoakustischen Modell geliefert. Wenn für jedes Subband das SMR berechnet wurde, wird dem Band mit dem niedrigsten MNR eine bestimmte Menge an Codebits zugeteilt.

Anschließend wird das neue SNR geschätzt, das MNR neu berechnet und der Vorgang so lange wiederholt, bis alle zur Verfügung stehenden Bits aufgebraucht sind.

Die Bitzuweisungsinformation (Bit Allocation) wird mit 4 bit codiert. Bei der Quantisierung wird je nach Anzahl der zugewiesenen Bits einer von 14 verschiedenen gleichförmigen Quantisierern mit $2^{(n-1)}$ Quantisierungsstufen benutzt, wobei n zwischen 2 und 15 liegt.

Der Skalierungsfaktor wird verwendet, um den gesamten Bereich des Quantisierers auszunützen. (Mit geringerer Anzahl an Bits quantisierte Signale würden zu leise abgespielt werden.)

3.1.2.5 Stereo Kodierung

Einfache Stereokodierung ist möglich, indem jeder Kanal für sich kodiert wird. Es gibt in Layer-1 und Layer-2 aber eine zusätzliche Möglichkeit, die die Datenredundanz und – Irrelevanz zwischen linkem und rechtem Kanal ausnützt: Intensity Stereo Coding. (Middle/Side (MS) Stereo- Coding wird erst ab Layer-3 unterstützt)

Intensity Stereo- Coding

Es wird die Eigenschaft des Menschlichen Gehörs ausgenützt, das es sich für die Stereo-Wahrnehmung bei Frequenzen über 2 kHz auf die Hüllkurve der Signale, nicht aber auf die genaue zeitliche Struktur der Signale stützt. Dies wird ausgenützt, indem bei tiefen Frequenzen jeder Kanal getrennt betrachtet wird, und bei hohen Frequenzen nur das Summensignal von links und rechts kodiert wird- jedoch mit einem eigenen Skalierungsfaktor für jeden Kanal.

3.1.2.6 Formatierung des Datenstroms:

Nach der Quantisierung findet die Formatierung statt: Es wird ein bitserieller Datenstrom erzeugt, der aus den quantisierten, codierten Werten und Zusatzinformationen wie Bitzuteilung, ... besteht. Zusätzlich werden noch der Header vorangestellt, und eventuell anfallende Daten zur Fehlerkorrektur (CRC, Cyclic Redundancy Check: Ein Prüfsummenverfahren, bei dem Datenfehler zu einer fehlerhaften Prüfsumme führen. Bei einem auftretenden Fehler wird der Datenblock erneut angefordert.) hinzugefügt. Die

Seiteninformation (Bitzuweisung, Skalierungsfaktoren) ist im Bezug auf Übertragungsfehler der sensibelste Teil der codierten Daten und setzt sich hier aus der Bitzuweisungsinformation aller Teilbänder (d.h. 32 mit 4 bit codierte Werte) und den Skalenfaktoren für alle Teilbänder mit von Null verschiedener Bitzuweisung (d.h. bis zu 32 mit je 6 bit codierte Werte) zusammen. (Siehe Bild 2, S.6)

3.1.2.7 Decoder

Im Decoder werden die decodierten Quantisiererausgaben mit dem Skalierungsfaktor multipliziert, um den ursprünglichen Subband-Wert wieder herzustellen.

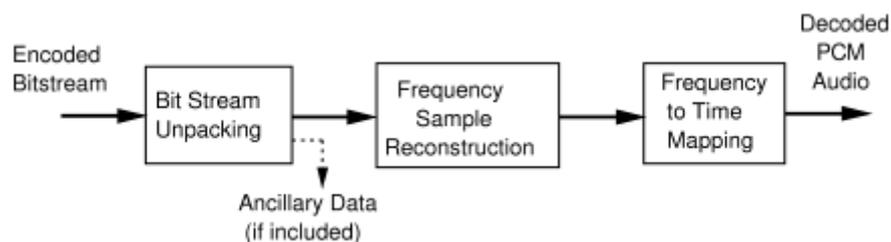


Bild 9: Blockschaltbild des MPEG-Audio Decoders

Das codierte Ausgangssignal bewegt sich zwischen 32 kbps (Kilobit per second) und 448 kbps, mit 192kbps pro Kanal (also 384 kbps in Stereo) wird vergleichbare CD-Qualität erreicht.

3.1.2.8 Verwendung von MPEG-1:

Bis heute wird MPEG-1 Video verwendet :

- +Als Standard bei der Video-CD und CD-I. Audio-Standard bei MPEG-1 ist MPEG-1 Layer-1 oder Layer-2
- +Bei News- Systemen, um Redakteuren an ihrem Arbeitsplatz die Möglichkeit zu geben, über ein 100Mbit-Ethernet-Netzwerk Videoclips aus dem Online-Archiv in verminderter Qualität zu sichten und evtl. einen Vorschnitt zu tätigen, zu vertonen,..
- +Bei der DCC (Digital Compact Cassette): Leicht modifizierter MPEG 1 Layer-1 mit 2x192kbps

3.1.3 MPEG AUDIO LAYER-2

Die Struktur von Layer-2 ist ähnlich von Layer-1, die in Layer-1 nicht benutzten Blöcke sind im Blockschaltbild dunkelgrau hinterlegt, die modifizierten Blöcke sind hellgrau hinterlegt.

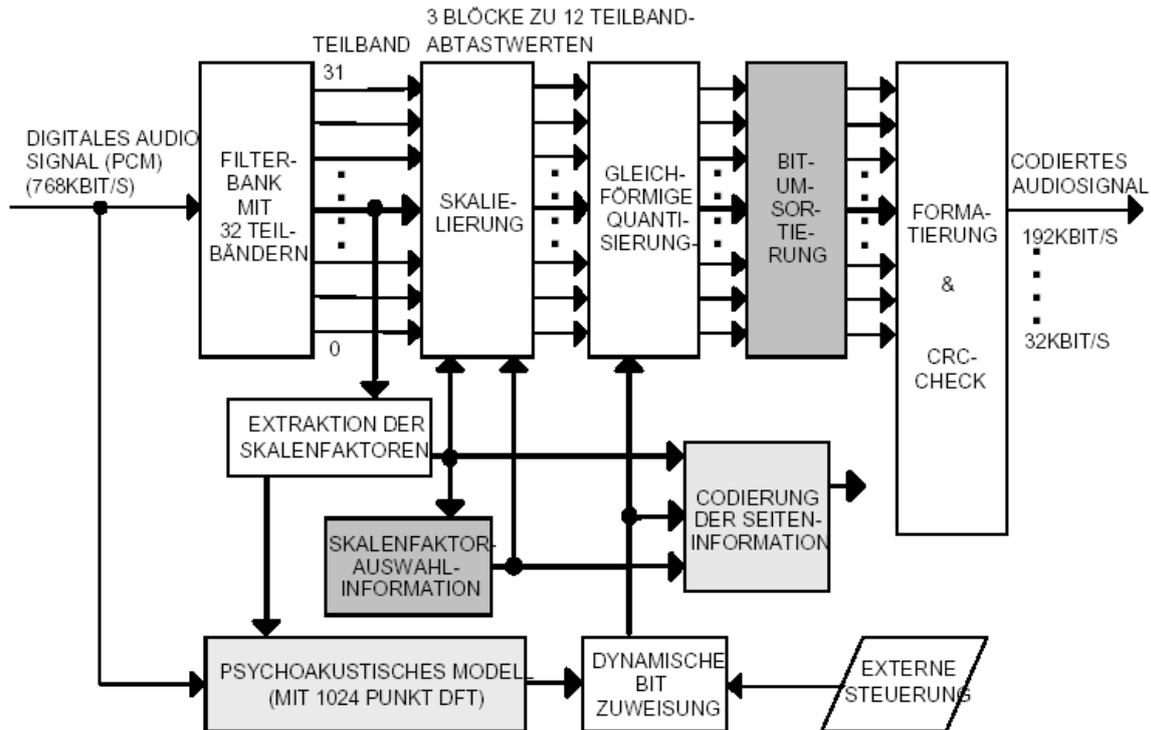


Bild 10: Blockschaltbild des MPEG-Audio Layer-II Coder

Die Framegröße beträgt im Gegensatz zu Layer-1 1152 Samples pro Frame.

Durch die Verdreifachung der Framegröße ergeben sich nun 36 statt bisher 12 Teilbandabtastwerte. Diese werden in drei Blöcke mit je 12 Abtastwerte gruppiert, für jeden Block wird entsprechend Layer-1 ein Skalierungsfaktor bestimmt, somit ergeben sich drei Skalierungsfaktoren pro Teilband.

Die Komprimierungsstrategie sieht zur Datenreduktion eine Variation der Anzahl der Skalierungsfaktoren vor. Beim Kodieren wird die Anzahl der tatsächlich verwendeten Skalierungsfaktoren pro Teilband bestimmt: Wenn z.B. die Skalierungsfaktoren ähnlich sind, oder der Kodierer erkennt, dass die zeitliche Maskierung des Gehörs die auftretenden Fehler unhörbar macht, werden nur 2 bzw. 1 Skalierungsfaktor verwendet.

Im Gegensatz zu Layer-1 lassen sich noch die Quantisierer anführen, denn in Layer-2 wird im unteren Frequenzbereich ein Satz mit 15, im mittleren mit 7 und im oberen mit 3 verschiedenen gleichförmigen Quantisierern verwendet. Je nach Satzgröße werden 4, 3 oder 2 Bit zur Kodierung der Bitzuweisungsinformation benötigt.

Da für bestimmte Quantisierungsstufen ({3, 5, 7, 9, 15, 31, 63, ..., 65535}) keine effiziente Kodierung mit 4, 3 oder 2 Bit möglich ist, werden in diesen Fällen drei aufeinanderfolgende Teilband-Abtastwerte zu einer Gruppe zusammengefasst, dadurch ist auch eine zusätzliche Bitumsortierung von Nöten.

Die Seiteninformation in Layer-2 setzt sich aus der Bitzuweisungsinformation für alle 32 Teilbänder (d.h. 32 mit je 4, 3 oder 2 Bit codierte Werte), der Skalenfaktorauswahl und den Skalenfaktoren für alle Teilbänder mit von Null verschiedener Bitzuweisung (d.h. bis zu 32 mit je 2 Bit codierte Werte bzw. 3x32 mit je 6 Bit codierte Werte). Ansonsten gilt für die Seiteninformation das Gleiche wie in Layer-1.

Header (32)	CRC (0,16)	Bit-Zuweisung (26-188)	SCFSI (0-60)	Skalierungs- faktor (0-1080)	Haupt- daten
----------------	---------------	---------------------------	-----------------	------------------------------------	-----------------

Bild 11: Aufbau eines Frames bei Layer-2, im Feld SCFSI (Scale Factor Selection Information) steht, wie viele Skalierungsfaktoren verwendet werden und wie diese unter den 3 Blocks aufgeteilt werden

3.1.4 MPEG AUDIO LAYER-3 (MP3)

Die Abkürzung MP3 steht nicht für MPEG-3, sondern für MPEG-1/2 Layer-3, und wird nur für Audiodateien verwendet.

Durch die Unterstützung der ISO und MPEG wurde die Technologie als Industriestandard schnell akzeptiert, wichtig für den Erfolg von MP3 war aber auch, dass das Fraunhofer-Institut anfangs die freie Nutzung und Entwicklung zuließ.

Der Durchbruch von mp3 begann mit dem Aufkommen erster Codier- und Decodiersoftware. 1997 wurde der erste MP3-Player, „AMP“ genannt, entwickelt- dieser diente als Vorlage für viele Windows- und Mac-Anwendungen. (WinAmp, MacAmp)

3.1.4.1 Lizenzen

Das Fraunhofer-Institut und die Thomson Multimedia SA besitzen ein Patent auf den MP3-Codec, bzw. auf die Erstellung eines mp3-kompatiblen Bitstreams. Es werden relativ hohe Lizenzgebühren von den Herstellern der MP3-Encoder verlangt . Dabei wird unterschieden: Die Preise für die Verwendung bei Anwendungen, Spiele, Hardware,... usw. sind unterschiedlich und liegen zw. 2500\$ (Spiele) bis 60000\$ (Anwendungen). Eine genaue Tabelle gibt es im Anhang.

3.1.4.2 Die Struktur von MP3

+ID3-Tag

Es können genaue Informationen über das MP3-kodierte Musikstück (z.B Autor, Titel,...) in die MP3-Datei geschrieben werden. Diese Informationen stehen im ID3-Tag. Dieser ist kein offizieller Standard, sondern eine Erfindung von Internetusern und Programmierern von MP3-Codecs.

Es gibt 3 Versionen:

ID3v1-Tag

Das ID3v1-Tag hängt am Ende eines Files und ist 128 Bytes groß.

Die ersten 3 Bytes enthalten „TAG“ als Kennung. Danach stehen:

- *30 Bytes für den Titel des Stücks
- *30 Bytes für den Künstler
- *30 Bytes für den Titel der CD
- *4 Bytes für das Jahr der Veröffentlichung
- *30 Bytes für ein Kommentar
- *1 Byte für die Genre-Kennung

ID3v1.1-Tag

Die letzten beiden Bytes des Kommentars sind für die Track-Kennung (=Titel des Tracks auf der CD) reserviert.

ID3v2-Tag

Das ID3v2-Tag fügt weitere Funktionen hinzu: Bilder für die Darstellung von CD-Covern, Songtexte,..

Das ID3v2-Tag steht am Anfang des MP3-Files-der Vorteil besteht darin, dass Informationen bei Streaming sofort angezeigt werden können- es muß nicht vorher das gesamte File downgeloadet werden, um am Ende des Files den ID3-Tag zu lesen.

Das Problem liegt darin, dass die meisten MP3-Player nicht kompatibel mit v2 sind und das gesamte MP3-File nicht abspielen können, da sie das am Anfang liegende ID3v2-Tag nicht lesen können.

Bislang konnten sich nur ID3v1 und in geringem Maße ID3v1.1 durchsetzen.

3.1.4.3 Frames

MP3-Dateien werden in kleinere Einheiten, den *Frames (Rahmen)*, unterteilt :

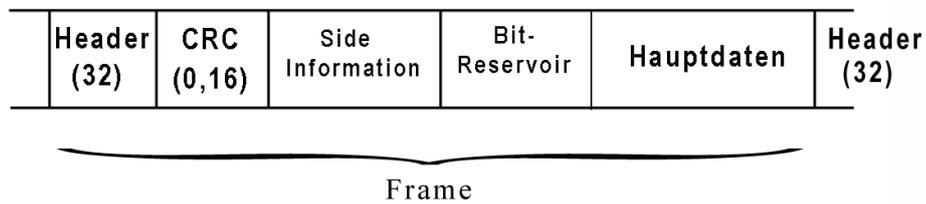


Bild 12: Frames bei MP3

Ein MP3-Datenstrom beginnt mit dem *Header*, in dem z.B. Informationen über den benutzten Layer (Layer 1 bis 3), die Datenrate,... usw. stehen. (Details der Bitbelegung des Headers: siehe Seite 18)

Der *CRC-Block* ist optional, im Header wird mittels des Protection Bit angegeben, ob er benutzt wird oder nicht.

Im *Side Information – Block* steht unter anderem, mit welchen Huffman-Bäumen die Hauptdaten kodiert sind und ein Zeiger auf den tatsächlichen Beginn der zum Block gehörenden Daten.

Das *Bit-Reservoir* (nicht in Layer-1 und 2 enthalten) entsteht, wenn so komprimiert werden kann, dass nicht die volle theoretische Bit-Anzahl eines Frames benötigt wird. Bei anderen Frames, wo zusätzliche Bits benötigt werden würden (z.B. bei komplexen Musikstellen), kann dieses Reservoir genutzt werden. (Siehe „Konstante bzw. Variable Bitrate“, S.22)

Die Hauptdaten sind Huffman- kodiert und enthalten die Skalierungsfaktoren und die quantisierten Samples. Wegen des Bit-Reservoirs gehören die Daten nicht zwangsläufig zum aktuellen Datenblock, wie Bild 13 zeigt:

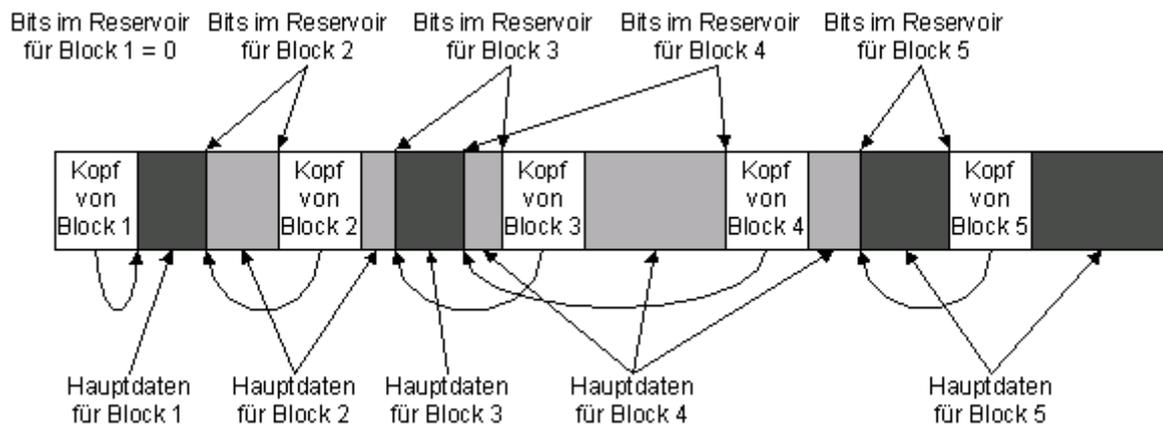


Bild 13: Ausschnitt eines MP3-Datenstroms

Frames können jeweils nur als ganzes abgespielt werden, nicht z.B. nur die erste Hälfte eines Frames.

3.1.4.4 Header

Der Header steht jedem Frame voran und enthält 32 Bit Daten:

Position		Länge [Bit]
A	Frame-Sync	11
B	MPEG Audioversion (ID) (MPEG-1, 2, etc.)	2
C	MPEG Layer (Layer I, II, III, etc.)	2
D	Protection (wenn aktiv: Checksumme nach Header)	1
E	Bitrate-Index	4
F	Frequenz der Samplingrate (44.1kHz, etc.,)	2
G	Padding Bit (kompensiert unvollständige Belegung)	1
H	Private Bit (Applikations-spezifische Trigger)	1
I	Channelmode (Stereo, Joint-Stereo)	2
J	Mode-Extension bei Verwendung von Joint Stereo)	2
K	Copyright	1
L	Original ("0", wenn Kopie, "1" wenn Original)	1
M	Emphasis (veraltet)	2

Tabelle 3.2 : Der Header bei MP3

+Sync-Block

Der Sync-Block dient dazu, um beispielsweise Daten am Anfang zu überspringen, bzw. in der Wiedergabe nach vor bzw. zurück zu springen. Bei Streaming ist der Sync-Block wichtig, um bei einem evtl. auftretendem Packetloss (Fehlen der Daten, z.B. bei Übertragungsfehlern, zu langsamen Verbindungen,...) resyncen zu können.

+ID Bit

Das ID-Bit gibt an, ob in MPEG-1 oder MPEG-2 codiert wurde:

Bitbelegung:

- 00 -MPEG-2.5 (ist kein offizieller Standard)
- 01 -Reserviert
- 10 -MPEG-2
- 11 -MPEG-1

+MPEG-Layer

Angabe des Layers.

Bitbelegung:

- 00 – Reserviert
- 01 – Layer-III
- 10 – Layer-II
- 11 – Layer-I

+Protection Bit

Ist das Protection Bit nicht gesetzt, folgt dem Header eine 16 Bit lange Checksumme (CRC).

Bitbelegung:

- 0 – Protected (16 Bit Checksumme folgt dem Header (CRC))
- 1 – Not Protected

+Bitrate Index

Angabe zur Bitrate (in kbit/s), mit der das Frame codiert wurde.

Bits	MPEG-1,L-1	MPEG-1,L-2	MPEG-1,L-3	MPEG-2,L-1	MPEG-2,L-2&3
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	Nicht mögl.				

Tabelle 3.3 : Tabelle des Bitrate-Index

* - „free“ bedeutet, dass eine freie Bitrate gewählt werden kann, die unter der maximal möglichen (448 kBit/sec) liegen und konstant bleiben muss. Decoder müssen diese Bitrate allerdings nicht zwingend unterstützen.

+Sampling-Frequenz

Bitbelegung:

• bits	MPEG1	MPEG2	MPEG2.5
00	44100	22050	11025
01	48000	24000	12000
10	32000	16000	8000
11	reserv.	reserv.	reserv.

Tabelle 3.4 : Bitbelegung bei untersch. Samplingfrequenzen

+Padding Bit

Es kann vorkommen, dass die Framelänge schwankt (z.B. bei einer Bitrate von 128 kBit, 44,1 kHz Samplerate: manche Frames enthalten 417 Bytes und manche 418 Bytes). Um eine genaue Bitrate von 128 kbit/s einzuhalten, wird das Padding Bit gesetzt. Beim Setzen des Padding-Bits wird das Frame um einen „Slot“ ergänzt. Dieser hat bei Layer 1 die Länge von 32 Bit und in Layer-2 bzw. 3 die Länge von 8 Bit.

Bitbelegung:

- 0 – kein Padding
- 1 – Padding mit einem extra Slot

+Private Bit

(Muß nicht codiert werden)
Dient nur zur Information.

+Channel- Mode

Bitbelegung:

- 00 – Stereo
- 01 – Joint Stereo
- 10 – Dual Channel
- 11 – Mono

+Mode Extension

Mode Extension kommt nur zu tragen, wenn der Joint-Stereo-Modus aktiviert ist. In Layer-1 und Layer-2 bestimmen die beiden Bits, in welchen Frequenzbändern Intensitäts-Stereofonie verwendet wird. (MS Stereo ist in Layer-1 und Layer 2 nicht möglich)

In Layer-3 wird bestimmt, welcher Typ von Joint Stereo verwendet wird. (Intensität oder MS)

• Layer I und II	Layer III
Wert	Intensity stereo MS stereo
00 bands 4 to 31	off off
01 bands 8 to 31	on off
10 bands 12 to 31	off on
11 bands 16 to 31	on on

Tabelle 3.5 : Mode Extension

+Copyright Bit

(Muß nicht codiert werden)

Bitbelegung:

0 – kein Copyright vorhanden

1 – Copyright vorhanden

+Original Bit

(Muß nicht codiert werden)

Bitbelegung:

0 – Kopie

1 – Original

+Emphasis

Filter (Anhebung von hohen Frequenzen) vor dem Codieren. (Dieses wird beim Abspielen automatisch rückgängig gemacht). Wird sehr selten benutzt.

3.1.4.5 Größe bzw. Länge eines MPEG-1/2-Frames

Unterschied: Framegröße (Framesize) <-> Framelänge (Framelenght)

+Framegröße ist die Anzahl an Samples pro Frame und ist immer konstant: 384 Samples für Layer-1 und 1152 Samples für Layer-2 und Layer-3. (Nach der

Aufteilung des Audiosignals in 32 Frequenzbänder, die jeweils 12*3 Samples enthalten ergibt sich: $32*12*3=1152$)

+Framelänge ist die Anzahl an Bytes eines komprimierten Frames. Berechnung:

Für Layer 1:

$$\text{Framelänge} = (12 * \text{BitRate} / \text{SampleRate} + \text{Padding}) * 4 \quad \dots \text{in Bytes}$$

Für Layer-2 und Layer-3:

$$\text{Framelänge} = (144 * \text{BitRate} / \text{SampleRate}) + \text{Padding} \quad \dots \text{in Bytes}$$

z.B. für 320kbit/s, 48kHz:

$$144*320\,000/48\,000 = 960 \text{ Bytes/frame}$$

3.1.4.6 Konstante bzw. Variable Bitrate

Mp3 Dateien können mit konstanter (CBR) oder variabler Bitrate (VBR) kodiert werden. Der Vorteil von variabler Bitrate liegt darin, dass nur an Stellen, wo mehr Bits benötigt werden (z.B. laute, klangfarbenreiche Musik,...) mit höherer Bitrate und an weniger komplexen (z.B. leisen) Stellen mit geringerer Bitrate codiert wird.

Variable Bitrate (VBR) kann auf 2 Arten erzeugt werden:

+Bitrate switching

Jedes Frame wird mit anderer Bitrate codiert. Dieses Verfahren muß in Layer-3 kompatiblen Playern funktionieren, Layer-1 oder Layer-2-kompatible Player müssen dies nicht unbedingt unterstützen.

+Bit Reservoir

Bits können von Bitreservoirs voriger Frames gelesen werden. Problem: Die Frames sind voneinander abhängig, schneiden der mp3-Dateien nur bedingt möglich. Das Verfahren wird nur in Layer-3 unterstützt.

3.1.4.7 Implementierung

Obwohl die Decodierung nach festgelegten Regeln zu erfolgen hat, gibt die ISO keine exakten Vorschriften, wie die Encodierung zu geschehen hat. Es werden lediglich Richtlinien vorgegeben, eine gewisse Struktur zu halten, damit der Encoder Dateien produziert, die Plattformunabhängig auf verschiedenen Decodern abgespielt werden können.

Die Definition von MP3 liefert nur eine Basis, die einzelnen Modelle und Filterungsverfahren bleiben den Softwareentwicklern überlassen.

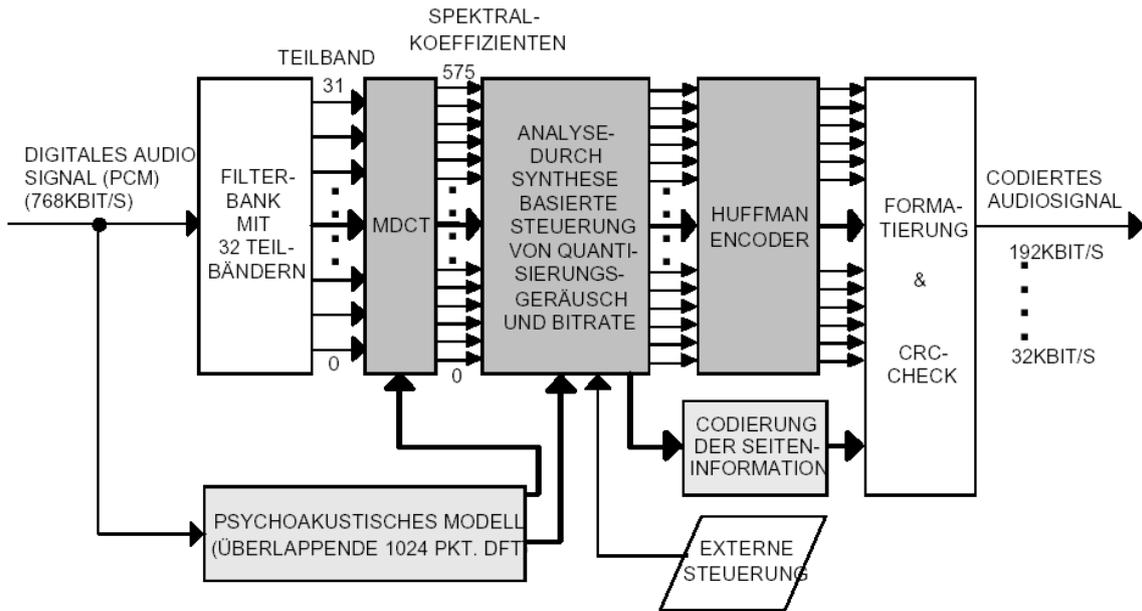


Bild 14: Blockdiagramm eines MP3-Kodierers Die Struktur von Layer-3 ist komplexer als die von Layer-2, die in Layer-2 nicht benutzten Blöcke sind im Blockschaltbild dunkelgrau hinterlegt, die modifizierten Blöcke sind hellgrau hinterlegt.

3.1.4.8 Filterbank

Im Unterschied zu Layer 1 und 2 verwendet Layer 3 eine Hybrid-Filterbank, bestehend aus einer Kaskadierung der Polyphasenfilterbank mit einer MDCT (Modified Discrete Cosine Transformation), mit adaptiver Fensterumschaltung.

Das digitale Eingangssignal wird zunächst in 32 Frequenzbänder (Subbands) gleicher Breite zu je 625Hz (@44,1 kHz) oder 750 Hz (@48kHz) transformiert. Als Filterbank wird hier ein Polyphase Filter verwendet (Siehe Layer-1, S.7). Dieser Schritt der Zerteilung in 32 Subbänder findet aus Kompatibilitätsgründen zu den Layern 1 und 2 statt.

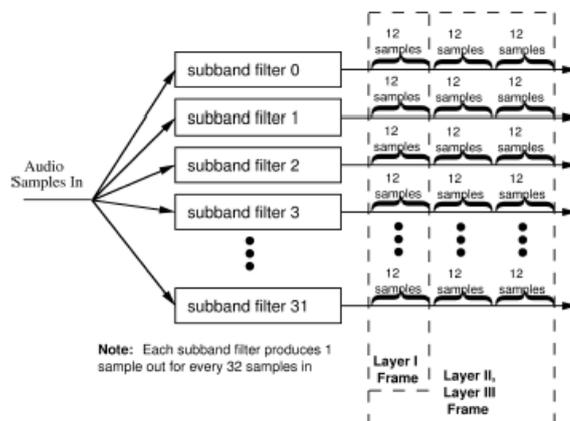


Bild 15: Zusammenfassen der Subband-Samples

3.1.4.9 MDCT

Ergänzend werden die Abtastwerte am Ausgang eines jeden der 32 Teilbänder einer MDCT (Modified Discrete Cosine Transformation) zugeführt, die jedes Subband noch einmal aufteilt, in 18 Teile ($32 \times 18 = 576$ Frequenzbänder gesamt) oder 6 Teile ($32 \times 6 = 192$ Frequenzbänder gesamt) – und zwar adaptiv: Zunächst werden die Subbänder in 18 Teile geteilt, es wird dann zwar eine bessere Frequenzauflösung erreicht, aber die Zeitauflösung ist dementsprechend schlecht. Bei schnellen Lautstärkewechseln (Siehe Bild 17, S. 25, bzw. TNS, Seite 35) kann dies evtl. zu Problemen führen: Das Quantisierungsrauschen kann während sehr leisen Stellen hörbar werden, Pre-Echo (siehe TNS, Seite 35) kann entstehen. Der Encoder versucht das nun mittels Psychoakustischem Modell (S.28) zu erkennen und gegebenenfalls zu beseitigen: Eine Möglichkeit wäre, sich Bits aus dem Bit-Reservoir zu leihen, oder das Subband in 6 statt 18 Teile zu zerlegen, um eine bessere Zeitauflösung zu erreichen.

Um bei der Cosinus Transformation die Randeffekte zu verringern, werden die Eingangswerte durch ein Fenster gefiltert. Da eine Erhöhung der Auflösung im Frequenzbereich eine Verringerung der Auflösung im Zeitbereich mit sich bringt, stehen für die MDCT zwei verschiedene Blocklängen zur Verfügung: 36 Samples oder 12 Samples. Da sich die MDCT zu 50% überlappt, sind 4 verschiedene Fenster notwendig: 1 kurzes, 1 langes und 2 für den Übergang von lang auf kurz und umgekehrt. (siehe Bild 16):

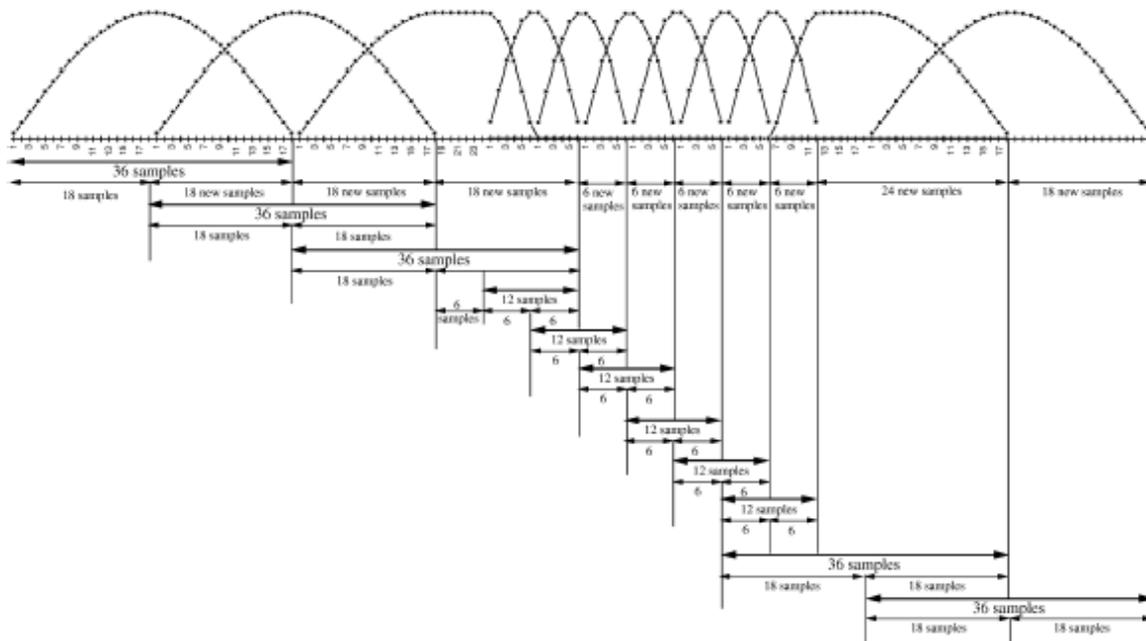


Bild 16 : Fenster der MDCT

Der kürzere Block erhöht die Auflösung im Zeitbereich und wird bei transienten Anteilen im Audiosignal verwendet. Es sind in Layer-3 drei verschiedene Betriebsarten möglich: Alle Bänder mit langer MDCT, alle Bänder mit kurzer MDCT und ein gemischter Modus, indem die 2 niedrigsten Bänder mit langer, und die übrigen Bänder mit kurzer MDCT arbeiten.

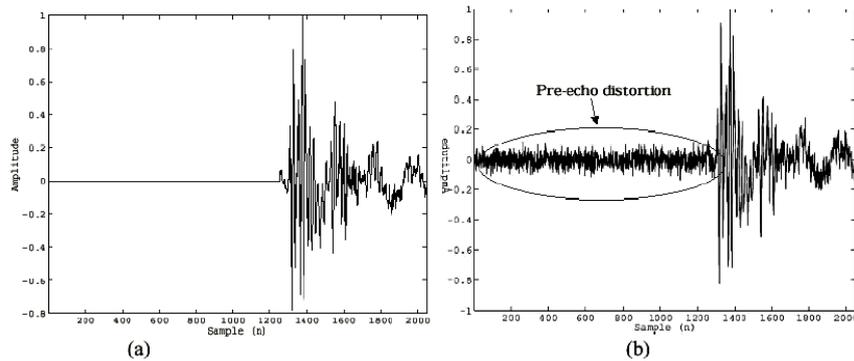


Fig. 14. Pre-Echo Example: (a) Uncoded Castanets. (b) Transform Coded Castanets, 2048-Point Block Size [Painter Spanias]

- Problem mit plötzlichen lauten Signalen nach einer Blockgrenze
 - vorheriger Block hat sehr wenig Bits kodiert
 - Rauschen des lauten Signals wird nicht mehr maskiert
- => Situation erkennen und behandeln (kürzere MDCT Blöcke)

Bild 17: Pre- Echo

Die MDCT und ihre inverse, die IMDCT wird folgendermaßen berechnet:

$$x_t(m) = \sum_{k=0}^{n-1} f(k)x_t(k) \cos\left(\frac{\pi}{2n} \left(2k + 1 + \frac{n}{2}\right)(2m + 1)\right), m = 0.. \frac{n}{2} - 1 \quad \dots(3.5)$$

$$y_t(p) = f(p) \frac{n}{4} \sum_{m=0}^{\frac{n}{2}-1} x_t(m) \cos\left(\frac{\pi}{2n} \left(2p + 1 + \frac{n}{2}\right)(2m + 1)\right), p = 0..n - 1 \quad \dots(3.6)$$

$x(k)$ sind die Samples im Zeitbereich. $x_t(k)$, $k=0..n-1$ sind die Samples die verwendet werden, um die Werte $X_t(k)$, $k=0..n/2-1$ des t -ten Blocks im Frequenzbereich zu berechnen. $f(k)$ sind die Werte des jeweiligen Fensters, das verwendet wird.

Zusätzlich wird durch eine Überlappung und Addition eine Unterdrückung der Aliasing Terme erreicht:

$$\bar{x}_t(q) = y_{t-1}\left(q + \frac{n}{2}\right) + y_t(q), q = 0.. \frac{n}{2} - 1 \quad \dots(3.7)$$

Es gibt für die MDCT bereits viele verschiedene Algorithmen, die sich in Punkten wie Geschwindigkeit, Speicherverbrauch und Anfälligkeit gegenüber Rundungsfehler unterscheiden.

3.1.4.10 Scalefaktor Bands:

Bei Layer 1 u. 2 kann jedes Band einen eigenen Skalierungsfaktor haben.

Bei Layer-3 werden die Ausgänge der MDCT zu Bändern zusammengefasst, die ungefähr den kritischen Bändern des Ohres entspricht, und für das neu entstandene Band nur *ein* Skalierungsfaktor verwendet. Diese werden bei der Rausch-Zuteilung berechnet (s.u.) und ermöglichen so ein buntes Rauschen, welches sich den unterschiedlichen Maskierungsschwellwerten anpasst.

Die Skalierungsfaktoren werden bei der Dekodierung mit den quantisierten Samples multipliziert.

3.1.4.11 Quantisierung und Codierung

Prinzipielles Verfahren: Wenn die Signalintensität eines Subbandes unter dessen Maskierung (vorgegeben vom Psychoakustischen Modell, s.u.) liegt, braucht es nicht kodiert zu werden. Sonst werden die Samples mit gerade so vielen Bits quantisiert, dass das entstehende Quantisierungsrauschen durch die Maskierung gerade noch unhörbar bleibt (Noise Allocation). Zusätzlich darf aber die maximal erlaubte Bitrate nicht überschritten werden, dies wird folgendermaßen erreicht:

2 verschachtelte Schleifen: Ziel ist es einerseits, das Quantisierungsrauschen immer unter der Hörschwelle zu halten (Noise Control/Distortion Loop), und andererseits die maximal erlaubte Bitrate nicht zu überschreiten (Rate Loop).

+Innere Schleife (Rate Loop):

Es wird die Bitrate des Coders so lange modifiziert, bis die maximal erlaubte Bitrate (z.B. 128 kbps) nicht überschritten wird. Dies geschieht folgendermaßen:

Die quantisierten Samples werden Huffman-kodiert (Siehe unten). Sollte die daraus resultierende Bitrate für einen zu kodierenden Block die erlaubte Bitrate übersteigen, wird so entgegengewirkt, dass der globale Grad der Verstärkung des Signals verändert wird. Daraus resultiert eine größere (oder kleinere) Quantisierschrittbreite (und somit weniger (oder mehr) Quantisierungsschritte). Diese Schleife wird so lange mit verschiedenen Quantisierschrittbreiten durchgeführt, bis die Bitrate klein genug ist.

+Äußere Schleife (Noise Control/ Distortion Loop):

Hier wird versucht, das Quantisierungsrauschen (bzw. Verzerrungen) für jedes Subband unter die vom Psychoakustischen Modell mitgeteilte erlaubte Lautstärke zu senken:

Jedem Scalefactor Band wird ein Skalierungsfaktor (Scalefactor Bands, siehe oben) zugeteilt. Liegt in einem Subband das Rauschen über dem erlaubten Wert, wird der Faktor des Bandes erhöht, um das Rauschen (bzw. Verzerrungen) zu reduzieren- dadurch müssen dem Band mehr Bits zur Verfügung gestellt werden. Dadurch erhöht sich aber die Bitrate. So muß jedes Mal, wenn neue Faktoren vergeben werden, die innere Schleife angewendet werde, um die Bitrate zu senken. Die äußere Schleife wird so lange wiederholt, bis in jedem Subband das Quantisierungsrauschen unter der maximal erlaubten Lautstärke liegt.

Huffman- Kodierung:

Die 576 Koeffizienten (32 Subbänder x 18 MDCT- Koeffizienten) werden nach aufsteigenden Frequenzen sortiert. Anschließend werden sie in 3 Bereiche aufgeteilt, wovon jeder mithilfe einer speziellen, optimierten Huffman- Tabelle codiert wird. Dabei werden häufig vorkommenden Werten kurze Codewörter und selten vorkommenden Werten lange Codewörter zugeteilt.

3.1.4.12 Formatierung des Bitstreams:

Nachdem die Komprimierung abgeschlossen wurde, wird ein bitserieller Datenstrom erzeugt. Dieser besteht aus den quantisierten, codierten Werten und Zusatzinformationen wie Bitzuteilung, Bit- Reservoir, ... Zusätzlich werden noch der Header vorangestellt, und eventuell anfallende Daten zur Fehlerkorrektur hinzugefügt.

3.1.4.13 Stereo Kodierung bei Layer-3

Verbesserungen gegenüber Layer-1 und 2 treten auch beim Stereo- Coding auf: Neben Intensity-Stereo Coding gibt es auch das Middle/Side (M/S) Stereo- Coding:

Middle/Side Stereo- Coding:

Es wird bei bestimmten Frequenzen ein Summsignal (Middle) und ein Differenzsignal (Side) des linken und rechten Kanal gebildet. Für das Side (Seiten)- Signal werden dann spezielle auf das Stereo- Hören abgestimmte psychoakustische Modelle verwendet. Dadurch kann die Kompressionsrate zusätzlich erhöht werden, weil das Seitensignal weniger Information enthält, als ein Monosignal.

Das Verfahren für die Stereo- Codierung verwendet wird, nennt sich „Joint Stereo“. (Kombination aus Intensity- und M/S Stereo- Coding).

Zusätzlich zu Layer-1 und Layer-2 besitzt Layer-3 noch andere Erweiterungen:

- + Alias- Reduction: Durch die Überlappung der Bänder der Filterbank und der darauffolgenden MDCT entsteht aliasing, welches durch verbesserte Algorithmen verringert werden kann.
- + Nichtlineare Quantisierung: Der Quantisierer des Decoders potenziert das Eingangssignal mit $\frac{3}{4}$, um das Signal to Noise Ratio (SNR) über den gesamten Quantisierungsbereich konstant zu halten.

3.1.4.14 Decoder

Der Aufwand zur Dekodierung eines MP3-Datenstroms ist verglichen mit dem der Kodierung gering. Es müssen lediglich die kodierten Frequenzsamples rekonstruiert werden, die dann

durch eine inverse modifizierte Cosinus-Transformation (IMDCT) und eine inverse Filterbank wieder in den Zeitbereich zurückgewandelt werden.

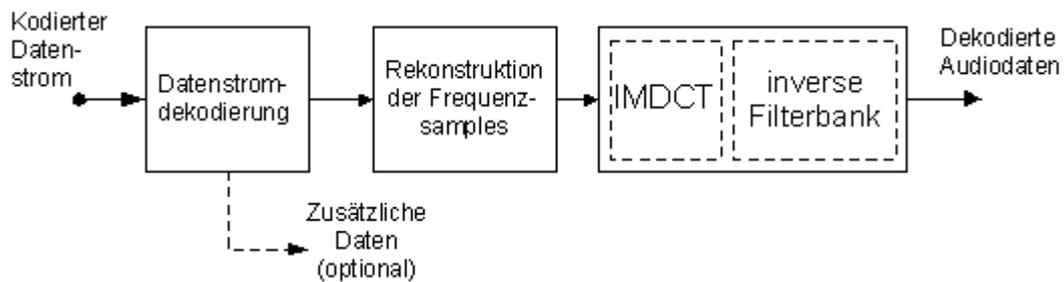


Bild 18: Blockdiagramm eines MP3-Decoders

Für diese Schritte wird nur eine geringe Rechenleistung benötigt.

3.1.5 Psychoakustisches Modell

Das Psychoakustische Modell ist notwendig, um das Audiosignal in den Wahrnehmungsbereich des Ohres zu transformieren und zu entscheiden, welche Bänder mit welcher Auflösung quantisiert werden müssen. Dazu wird das Audiosignal analysiert und der Pegel des Rauschens (vom Ohr nicht hörbar) als Funktion der Frequenz berechnet. Es gibt zwei verschiedene psychoakustische Modelle. Modell 1 ist einfacher als Modell 2 aber liefert nicht so gute Ergebnisse. Jedes Modell ist für jeden Layer verwendbar, jedoch nur Modell 2 besitzt spezielle Anpassungen an Layer-3.

Es folgt nun ein Überblick der Schritte, die für ein psychoakustisches Modell notwendig sind. Wenn nicht anders erwähnt, gelten diese für beide Modelle:

3.1.5.1 Zeitsynchronisation

Das Modell wird für jeden Frame einmal berechnet. Die Daten die das Modell bekommt müssen synchron mit den Daten sein, die gerade quantisiert werden. D. h. es muß die Verzögerung der Filterbank berücksichtigt werden. Zusätzlich muß ein Offset so einberechnet werden, daß die relevanten Daten in der Mitte des Analyse Fenster des Modells liegen. Wenn z. B. Layer-1 und Modell 1 verwendet werden, ist die Verzögerung durch die Filterbank 256 Samples und der Offset, um die 384 Samples eines Layer-1 Frame in der Mitte des 512 Punkte Analyse Fenster von Layer-1 zu platzieren, $(512-384)/2=64$ Samples. Dadurch ergibt sich ein notwendiger Offset von 320 Samples, um das psychoakustische Modell mit den Ausgängen der Filterbank zu synchronisieren.

3.1.5.2 Zeit - Frequenz Transformation

Die beiden Modellparameter brauchen eine eigene Zeit – Frequenz Transformation, da die Frequenzauflösung der Filterbank nicht hoch genug ist, um ein gutes Hörmodell zu erstellen. Dafür wird eine FFT mit Hann- Fenster verwendet.

Modell 1 verwendet eine 512 Punkte FFT für Layer-1 und eine 1024 Punkte FFT für Layer-2 und 3.

512 Punkte sind genug um die 384 Samples von Layer-1 zu erfassen. 1024 Punkte sind ein bisschen zu wenig um die 1152 Samples von Layer-2 und 3 zu erfassen, es ist aber ein Kompromiß, um Rechenleistung zu sparen. (Dabei werden die ersten und letzten 64 Samples einfach abgeschnitten und nicht berücksichtigt)

Modell 2 verwendet eine 1024 Punkte FFT für jeden Layer. Die 384 Samples von Layer-1 werden wie bei Modell 1 im Fenster der FFT zentriert. Bei Layer-2 u. 3 werden die 1152 Samples in 2 gleich große Teile zerlegt und die FFT zweimal durchgeführt. Einmal wird die erste Hälfte der Eingangssamples und dann die zweite Hälfte im FFT-Fenster zentriert. Es wird dann die Berechnung genommen, die die niedrigere Maskierungsschwelle ergibt. Die Maskierungsschwelle ist jener Pegel, den das Quantisierungsrauschen nicht überschreiten darf, da es sonst vom Ohr wahrgenommen wird.

3.1.5.3 Frequenzanteile gruppieren

Die spektralen Anteile werden zu Gruppen zusammengefasst, die mit den kritischen Bändern des Gehörs übereinstimmen. Dies erleichtert die Berechnung des akustischen Modells.

3.1.5.4 Verteilungsfunktion

Jedes Signal eines Bandes beeinflusst auch die Maskierungsschwelle umliegender Bänder. Darum wird eine Verteilungsfunktion verwendet, die den Maskierungseffekt auch auf die angrenzenden Bänder überträgt. Bild 20 zeigt die Wirkungsweise der Verteilungsfunktion (Spreading Function). (Angewendet auf das in den Wahrnehmungsbereich des Ohres transformierte Signal aus Bild 19)

3.1.5.5 Tonanalyse

Beide Modelle entscheiden, ob sich bei den spektralen Anteilen um Töne oder Geräusche handelt. (Bild 21) Dies ist wichtig, da sich diese unterschiedlich auf die Hörschwelle des Rauschens auswirken.

Die Berechnung unterscheidet sich in den beiden Modellen.

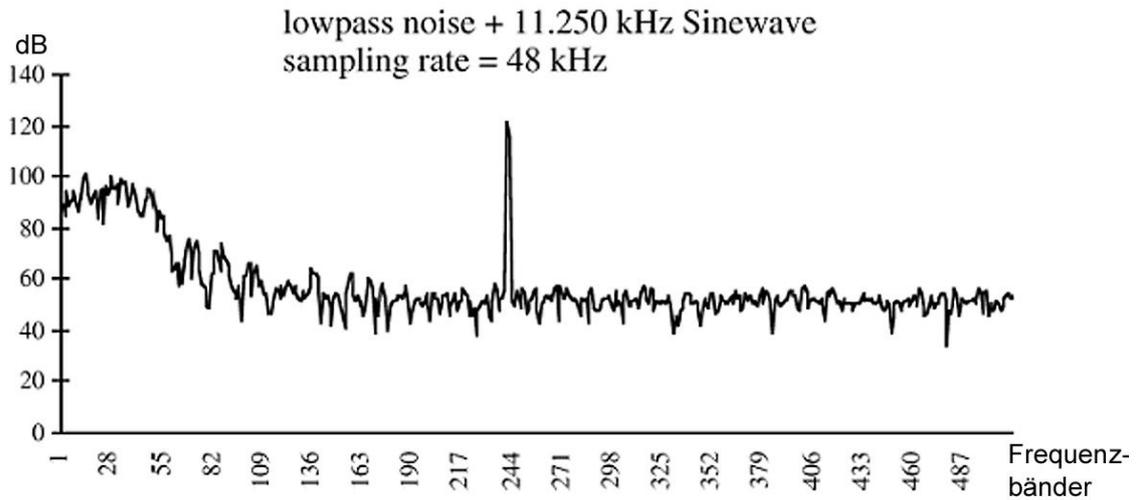


Bild 19: Eingangssignal

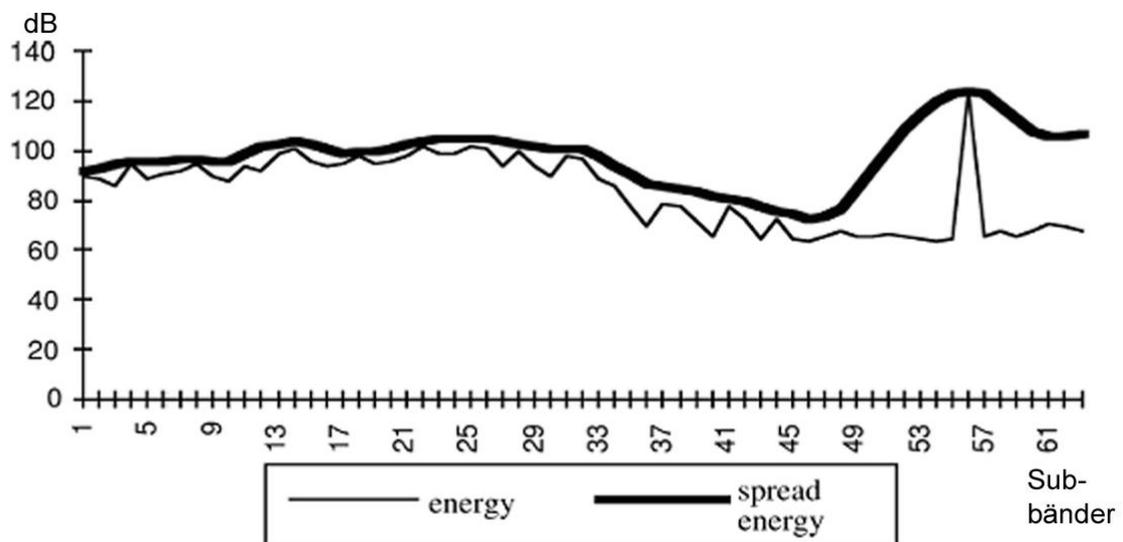


Bild 20: Audioenergie und verteilte Energie im Wahrnehmungsbereich des Gehörs

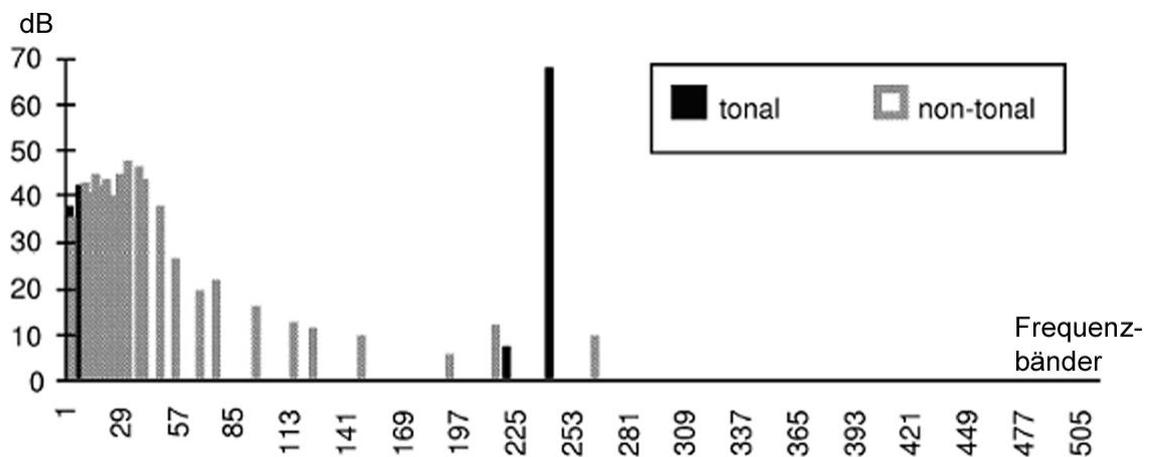


Bild 21: Ton-Geräuschidentifikation bei Layer-1 und 2

In Modell 2 wird nicht strikt in ton- und nicht-tonartige Signale getrennt, sondern es werden die Signale nach einer Tonähnlichkeits- Skala bewertet. D. h. es wird bewertet, ob es sich sicher eher um einen Ton oder eher um ein Geräusch handelt (s. Bild 22). Als Maß wird dafür die Vorhersagbarkeit einer Spektralkomponente hergenommen. Das Modell merkt sich die letzten 2 Werte jeder Frequenz und extrapoliert diese, um den nächsten zu erraten. Je größer die Abweichung zum richtigen Wert ist, desto eher handelt es sich um ein Geräusch und nicht um einen Ton.

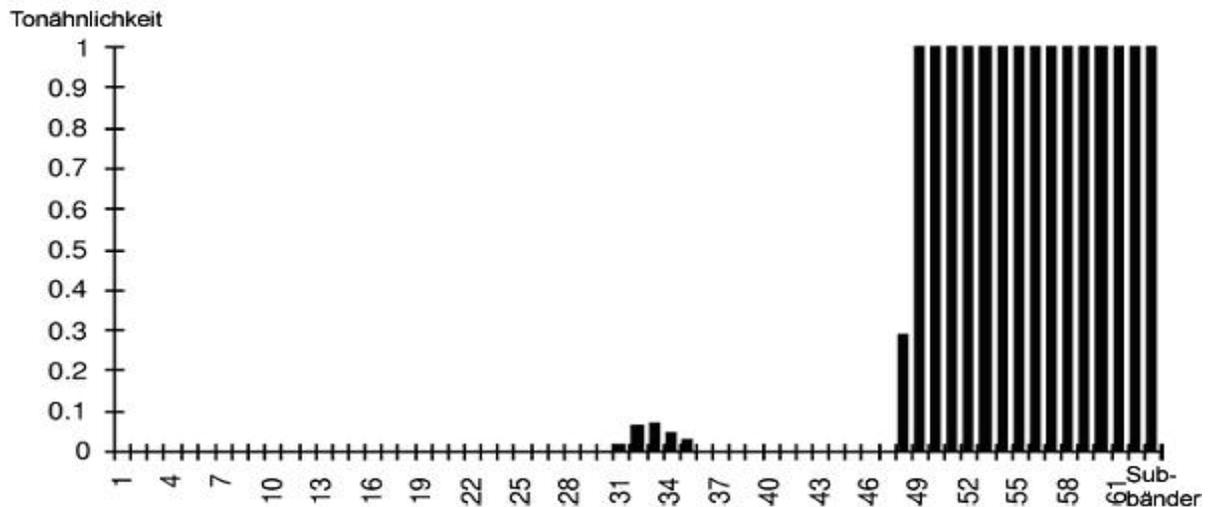


Bild 22: Tonähnlichkeitsbewertung bei Layer-3: 0..Geräuschhaft
1..Tonhaft

3.1.5.6 Untere Grenze der Schwellwerte

Bei beiden Modellen gibt es einen unteren Grenzwert für die Maskierungsschwelle, die Ruhehörschwelle.

3.1.5.7 Schwellwertsuche für jedes Band

Da die Frequenz- Auflösung des psychoakustischen Modells besser als die der Filterbank ist, muß aus einer Vielzahl an berechneten Schwellwerten in diesem Band einer gewählt bzw. berechnet werden. Bei niedrigen Frequenzen, wo die Filterbandbreite groß im Verhältnis zur kritischen Bandbreite ist, wird der minimale Schwellwert verwendet. Bei hohen Frequenzen wird jedoch der Mittelwert der Schwellwerte in dem jeweiligen Subband verwendet.

3.1.5.8 Berechnung der Signal To Mask ratio (SMR)

Das SMR wird für jedes Subband als Verhältnis der Signalenergie in diesem Band zum minimalen Schwellwert in diesem Band berechnet. Der berechnete Wert wird anschließend zur Bit- bzw. Rauschzuteilungseinheit (Bit Allocation, Noise Allocation) weitergereicht.(Bild 25)

3.1.5.9 Abbildungen zum Psychoakustischen Modell

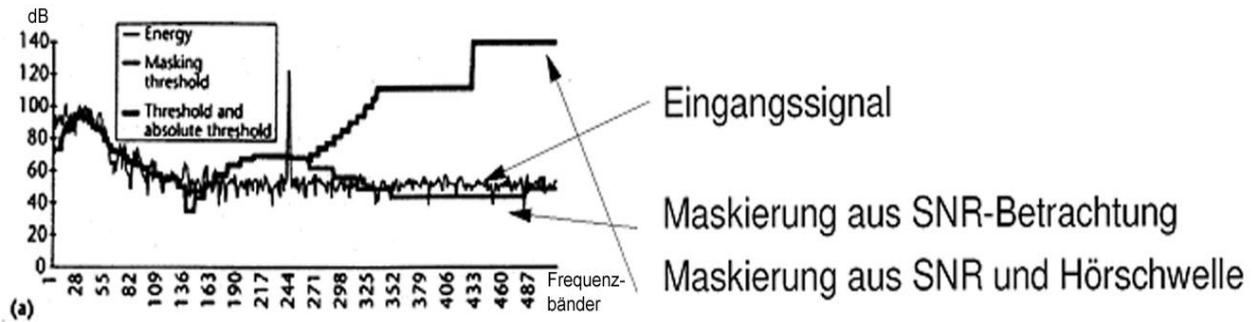


Bild 23: Maskierung bei dem Eingangssignal von Bild 19

Kodierung nach Modell 1:

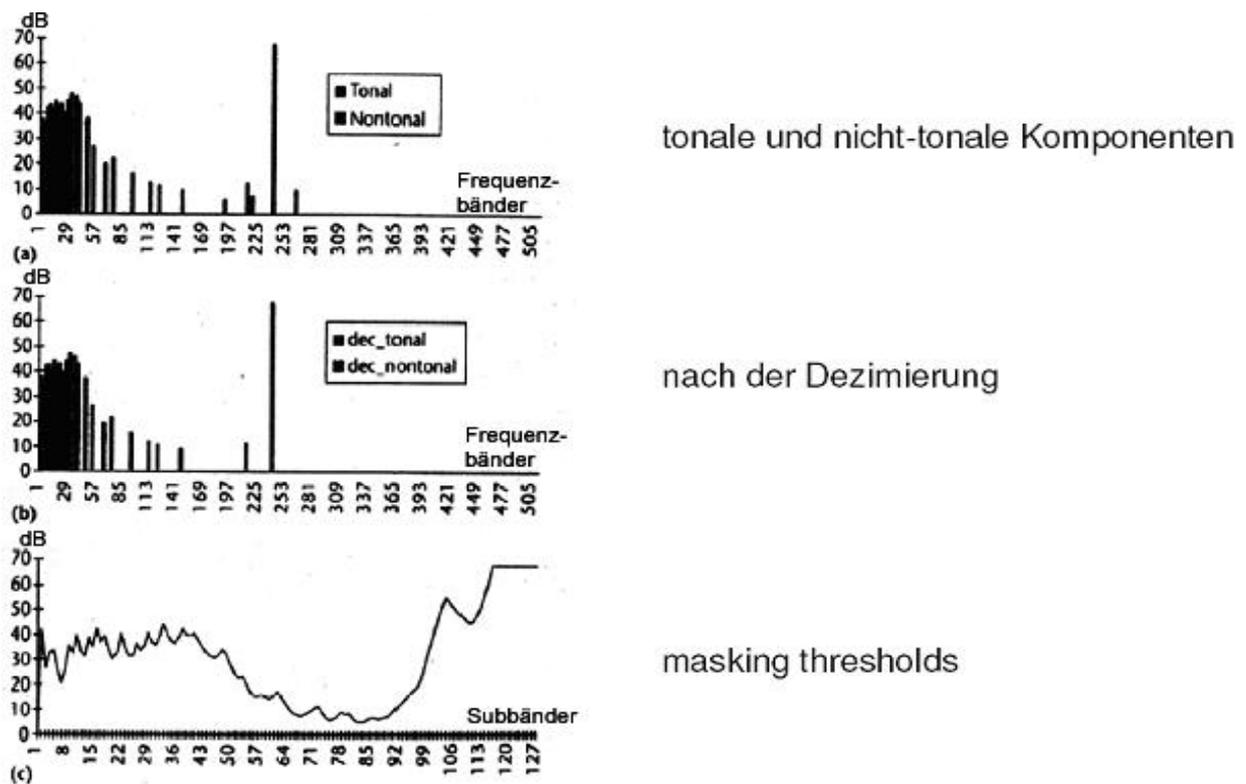
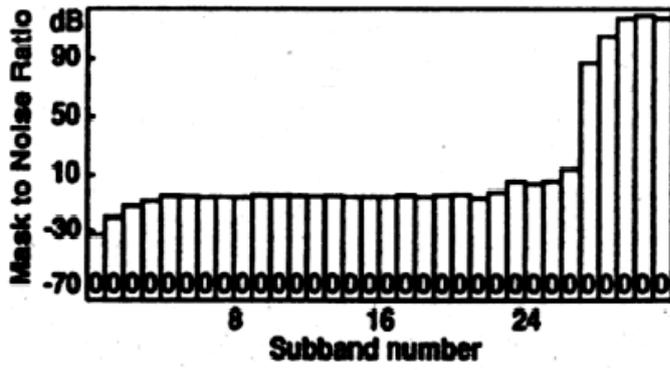
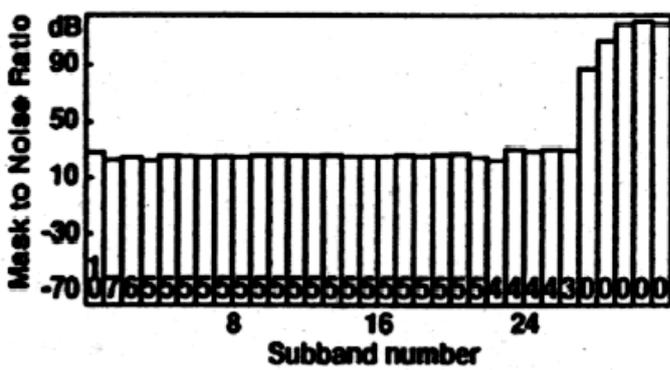


Bild 24: Bestimmung der tonalen Komponenten und anschließende Dezimierung. Unten ist nochmals die Maskierungsschwelle dargestellt



Mask-to-noise ratio

- beim Start der bit-allocation



- nach der bit-allocation
- Anzahl bits / Band

Bild 25: MNR zu Beginn und nach der bit- allocation: die für das jeweilige Subband verwendete Anzahl an Bits (nach der bit- allocation) ist im jeweiligen Subband eingetragen. (Die obersten 5 Subbänder liegen unterhalb der Maskierungsschwelle und somit werden diesen auch keine Bits zugeteilt)

Hier nun das Ergebnis der Kodierung nach Modell 1:

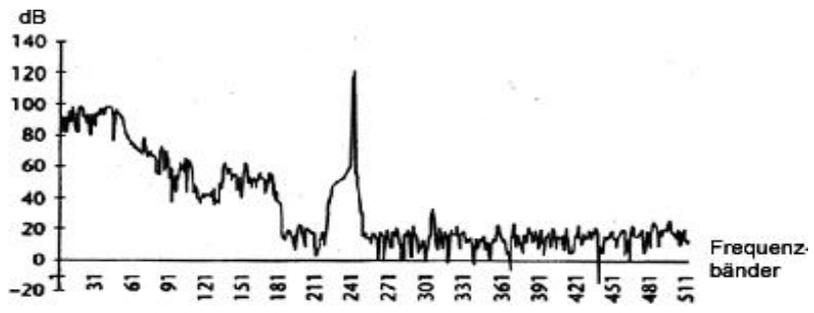


Bild 26. Ergebnis Psychoakustisches Modell 1

Im Vergleich zu Modell 2 ist Modell 1 teilweise deutlich schlechter (s.u.)

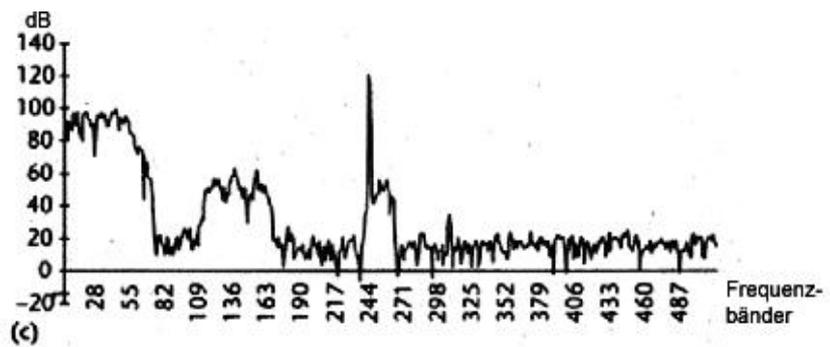


Bild 27. Ergebnis Psychoakustisches Modell 2

Weitere Informationen und nützliche Programme sind auf <http://lame.sourceforge.net/> zu finden. Weiters sind hier Sourcecodes und Tipps für die Programmierung zu finden.

3.2 MPEG 2

3.2.1 MPEG-2 Video

Bei diesem Standard sollten Fernsehbilder in Sendequalität bei Datenraten bis zu 10 Mbit/sec übertragen bzw. gespeichert werden können.
Genaueres zu MPEG- Video: siehe Kapitel 4

3.2.2 MPEG-2 Audio: (ISO/IEC 13818-3)

Es kamen von MPEG-1 Audio lediglich Erweiterungen hinzu:

MPEG-2 Layer 3 Audio:

- 1.)Vor- und Rückwärtskompatibles Multikanal- Encoding (z.B. für 5.1 Surround-Anwendungen), Max. Anzahl der Kanäle : 7.1
- 2.)Niedrigere Abtastraten: 16 kHz; 22,05 kHz und 24 kHz

Unterteilung in MPEG-2 / LSF und MPEG-2/Multichannel:

MPEG-2 / LSF:

- +Ein oder Zwei Audiokanäle
- +Abtastraten: 16 kHz, 20,05 kHz, 24 kHz
- +Bitraten: 8 kbit/s bis 256 kbit/s

MPEG-2 / Multichannel:

- +Bis zu 7 Audiokanäle plus einem LFE (Low Frequency Enhancement)
- +Sampleraten 32 kHz, 44,1 kHz und 48 kHz
- +Bitraten: bis 912 kbit/s bei 7.1

Verwendung:

MPEG-2 ist der Video-Standard bei der SVCD (Super Video CD) und DVD. Audio-Standard in MPEG-2 sind MPEG-1/2 Layer-1 bis 3 und auch MPEG2-AAC.

MPEG-2 wird bei DVB (Digital Video Broadcast) verwendet.

Zusätzlich gibt es Camcorder im professionellen Bereich, die in MPEG-2 aufzeichnen. (Betacam SX)

3.2.3 MPEG-2 AAC

Der Encodingprozess unterscheidet sich recht deutlich von MPEG 1/2 Layer-3. Es wird zwar ebenfalls eine Filterbank im Eingang, eine Iterations-Schleife sowie eine Huffman-Codierung angewendet, jedoch kommen im Detail einige Neuerungen hinzu, speziell im Bereich des Kodierens von Audiosignalen bei sehr geringen Bitraten. AAC ist deshalb nicht rückwärtskompatibel zu Layer-3.

Es sind bis zu 48 Audiokanäle plus max. 16 LFE Kanäle möglich (gesamt 64 Kanäle). Die mögliche Bitrate reicht von 8 bis 512 kbps pro Kanal.

3.2.3.1 Neuerungen:

- + höhere Frequenzauflösung (1024 Frequenzlinien anstatt 576)
- + eine optionale Rückwärts- Prädiktion vermeidet Probleme beim Codieren von sehr tonalen Signalen
- + verbesserte Joint-Stereo-Kodierung gestattet niedrigere Bitraten. Sie ist pro Frequenzband anwendbar und nicht mehr nur für den ganzen Audio-Frame.
- + verbesserte Huffman-Kodierung
- + Anstatt der kaskadierten Filterbank bei Layer-3 wird bei ACC eine „Switched MDCT“ (Modified Discrete Cosine Transformation) mit 8x128 Punkten mit einer Impulsantwortzeit von 5,3 ms bei einer Abtastrate von 48 kHz verwendet. Dies ist deutlich kürzer als bei Layer-3 (18,6 ms) und reduziert den Anteil von Vorechos.
- +Es wurde Temporal Noise Shaping (TNS) eingeführt, womit die Qualität von transienten Signalen (Sprache, Kastagnetten,...) bei sehr niedrigen Bitraten verbessert wird. TNS: Zusätzlicher Schritt zwischen der Filterbank und dem Quantisierer/Codierer : „Pre-Echos“ können nur dann maskiert werden, wenn kein signifikanter Anteil an Quantisierungsrauschen länger als ~2ms vor dem Einsetzen des Signals vorhanden ist. Sonst wird dieses Rauschen als Pre-Echo wahrgenommen, ein kurzes , rauschähnliches Signal kurz vor Beginn des eigentlichen Audiosignals. Entstehung: Quantisierungsfehler, die im Frequenzbereich entstehen, sind nach der Zusammensetzung des Signals im Zeitbereich über längere Zeit vorhanden. (z.B bei 1024 Linien MDCT, bei einer Abtastrate von 48 kHz: Dauer 40ms). TNS erlaubt dem Coder eine gewisse Kontrolle über die zeitliche Feinstruktur des Quantisierungsrauschens sogar innerhalb jeder Filterbank, und kann das Quantisierungsrauschen so regeln, dass Pre- Echo (praktisch) eliminiert werden kann.

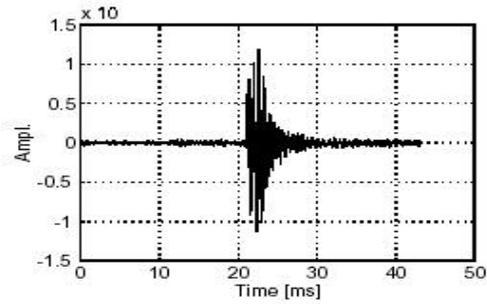


Figure 3: Transient signal (castanets, uncoded).

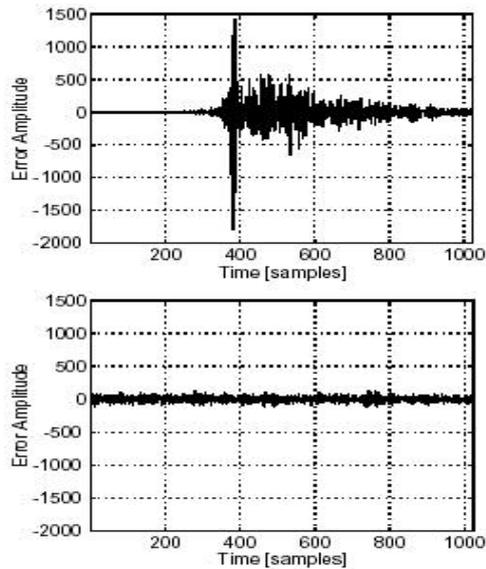


Bild 28: Temporal Noise Shaping: oben das Originalsignal, in der Mitte das Fehlersignal ohne TNS, unten das Fehlersignal mit TNS, was eine deutliche Verbesserung mit sich bringt.

+Einführen von 3 Profilen:

- +Main Profile mit vollem Umfang des AAC-Algorithmus
- +LC Profile (Low Complexity) ohne Prädiktion und begrenztes TNS, der Kodiervorgang geschieht dadurch deutlich schneller
- + SSR Profile (Scaleable Sampling Rate) Die Filterbänke wurden abgeändert, um eine Wiedergabe bei frei einstellbaren Sampleraten zu ermöglichen.

Durch die Verbesserungen wird die gleiche Qualität wie bei MPEG 1/2 Layer-3 erreicht, bei nur noch ~ 66 % der Bitrate.

Input Sampling Rates	8, 11.025, 12, 16, 22.05, 24, 32, 44.1, 48, 64, 88.1 and 96 (kHz)
Output Data Rates	Fixed between 8 and 320 kb/s per channel, or variable between 8 and 320 kb/s per channel+ VBR presets
Input file format	16-bit PCM raw data
AAC Profiles	MPEG-2 AAC LC, Main; MPEG-4 AAC LC, Main, LTP
AAC Bitstream Formats	ADTS, ADIF, Raw Data Transport
Channel Configurations	Mono, Stereo, 3.0, 3.1, 5, 5.1, 7, 7.1, up to 48 Channels
Encoding Speed	Up to 2.0 times faster than real-time on Pentium III 850 for stereo 44.1 kHz encoding at 128 kbits/s

Tabelle 3.6: Tabelle der möglichen Abtastfrequenzen und Bitraten.
 LC...Low Complexity (Prediction, PNS tools und Gain Control werden bei LC nicht verwendet)

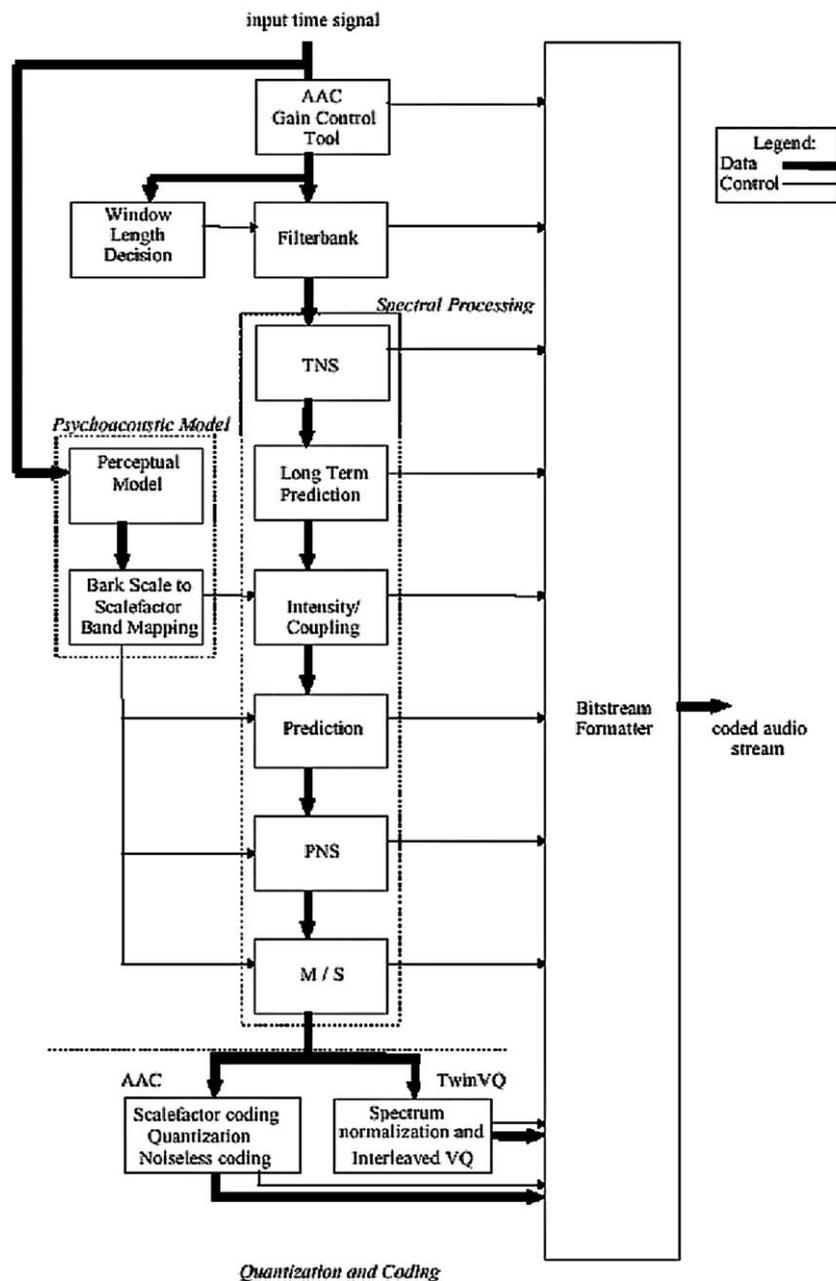


Bild 29: Blockschaltbild des AAC Audio-Encoder-Algorithmus

Das folgende Bild gibt eine Aufteilung der CPU-Rechenleistung für die jeweiligen Encodier-Schritte :

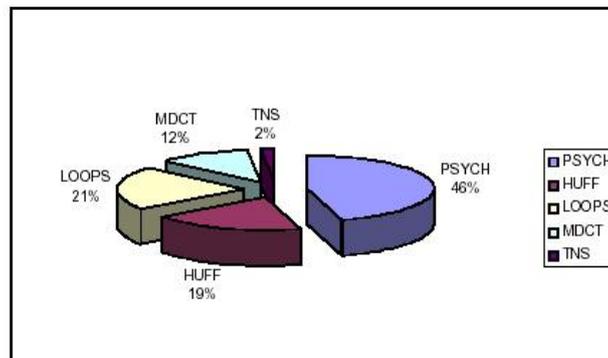


Bild 30: Aufteilung der Rechenleistung

Lizenzmodell und Preise : Siehe Anhang

Im Moment hat das mp3-Format durch seine weite Verbreitung eine sehr gefestigte Position. AAC ist mp3 allerdings deutlich überlegen, wird von Programmen allerdings noch kaum unterstützt. In der Zukunft wird AAC allerdings mehr und mehr an Bedeutung zulegen, da es im MPEG-4 Standard enthalten ist.

4. MPEG Video

4.1 Verfahren der Videokompression

Im einfachsten Fall reicht es, die hintereinanderfolgenden Einzelbilder (Frames) einzeln zu koprimieren, z.B. im Format gif, pcx, ... Wenn die Einzelbilder im jpg – Verfahren komprimiert wurden, spricht man von MJPEG, das in Studios oft verwendet wird. Die dadurch erreichbare Datenrate liegt bei etwa 25 Mbit/s (Megabit pro Sekunde), die für die Verwendung bei DVD- Video oder DVB (Digital Video Broadcast) zu hoch ist. Auf einer DVD 5 (4,7 GByte Speicher) passen so ~ 22 min Video plus Ton im PCM-Format, ein Satellitentransponder, der üblicherweise 35 Mbit/s tragen kann, wäre mit einem Fernsehprogramm fast ausgelastet. Auf DVD und bei DVB werden Datenraten zw. 4 Mbit/s

und 10 Mbit/s erlaubt. Um eine höhere Kompressionsrate zu erzielen, werden nicht nur Einzelbilder betrachtet, sondern es werden die Einzelbilder zueinander in Beziehung gebracht.

Der einfachste Fall hierbei ist die Deltaframe- Kompression:

Es werden nur die Differenzen zum jeweils letzten Bild übertragen, es ist aber nötig, alle 5 bis 15 Einzelbilder ein vollständiges Schlüsselbild zu übertragen, das „Keyframe“. Ebenfalls bei Szenenwechsel wird ein Keyframe übertragen. Dieses Verfahren ist einfach und der Berechnungsaufwand bei Kodierung / Dekodierung ist gering (lediglich Addition bzw. Subtraktion). Dadurch wurde es in einigen Codecs (z.B. Cinepak, Intel Indeo ...) verwendet, und kommt bis heute noch in den Einsatz.

Bei MPEG- Video geht das Verfahren jedoch weit über die Deltaframe -Kompression hinaus: Es gibt 3 Frametypen: I, B, P- Frames.

I- Frames (Intraframes): dabei wird einfach ein Vollbild kodiert (jpg). Dies ist der Startpunkt einer (Im Normalfall) 15 Bilder fassenden Bildfolge, GOP (Group of Pictures) genannt:

GOP : IBB P BB P BB P BB P BB

P- Frames (Predicted Frames): Hier werden die Differenzen zum jeweils vorhergehenden I- oder P – Frame kodiert, im Grunde handelt es sich also um ein Deltaframe- Verfahren. Allerdings sind die Möglichkeiten vielfältiger: Es werden Makroblöcke (16x16 Pixel) eingeführt, die auf 2 Arten kodiert werden können: einerseits mit dem Deltaframe- Verfahren (wobei hier die DCT- Koeffizienten der Differenzen abgelegt werden) oder mit dem „Motion-Compensaton“- Verfahren: Hierbei wird die Delta-Information um einen Bewegungsvektor ergänzt. D.h. der Encoder muss nicht immer den exakt gleichen Bildausschnitt bearbeiten, sondern kann um ausscheren, z.B. bei Kamerafahrten usw. bringt dieses Verfahren eine deutliche Erhöhung des Kompressionsfaktors.

B- Frames (Bidirectional Frames): Hier werden nicht nur die Differenzen zum vorhergehenden, sondern auch zum darauffolgenden I oder P -Frame kodiert.

4.2 MPEG-1 Video

Das Ziel liegt darin, Bewegtbilder bis zu einer max. Datenrate von 1,5 Mbit/s zu übertragen bzw. zu speichern. Der Standard beschreibt, wie das Format aufgebaut ist und legt die Wiedergabe fest. Wie die Kompression erfolgen muss, um ein optimales Ergebnis zu erzielen, bleibt offen.

MPEG-1 erlaubt Videowiedergabe in ungefähr gleicher Qualität wie VHS. Der Standard arbeitet mit voller Bildwiederholrate, allerdings werden Halbbilder *nicht* berücksichtigt!

Die Bildauflösung wird um ein Viertel reduziert:

+PAL-Format: 25 Vollbilder/sec bei einer Auflösung von 352*288

+NTSC-Format:30 Vollbilder/sec bei einer Auflösung von 352*240

(Diese Formate werden CIF (Common Interface Format) genannt)

4.3 MPEG-2 Video

Bei diesem Standard sollten Fernsehbilder in Sendequalität bei Datenraten bis zu 10 Mbit/sec übertragen bzw. gespeichert werden können. Die Bandbreitenvorgabe ist hier also viel höher. Es werden volle Bildauflösung (PAL: 720*576, NTSC: 720*480) und volle Bildwiederholraten unterstützt.

Hier werden erstmals auch Halbbilder unterstützt. (PAL: 50 Halbbilder, NTSC 60 Halbbilder)

MPEG-2 ist der Videostandard für DVD und DVB.

4.3.1 Profiles & Levels

MPEG-2 wurde, im Gegensatz zu MPEG-1, skalierbar ausgelegt. Erreicht werden sollte damit, dass Geräte nicht unbedingt den gesamten Umfang der MPEG-2 Spezifikation, sondern nur eine Untermenge erfüllen mussten.

Es wurden zu diesem Zwecke sog. „Profiles“ für die Beschreibung der zu erfüllenden Funktionalität und „Levels“ für die Auflösung eingeführt.

5. MPEG-3

Ursprünglich dachte man an MPEG-3 als Standard für extrem hohe Datenraten und Auflösungen, wie dies z.B. bei HDTV notwendig wäre. (HDTV: „High Definition Television“: verbesserter Fernsehstandard, der in Japan und den USA bereits gesendet wird. Auflösungen sind skalierbar: von den Standard- Auflösungen (768x576 (PAL) oder NTSC (640x480) bis zu 1920*1080 Pixel. Zusätzlich sind das Halbbild- Verfahren (interlaced) oder Vollbildverfahren (progressive) vorgesehen. Auch sind mehrere Bildwiederholraten möglich: 24p, 25p, 30p, 50i, 60i.

Diese Spezifikationen wurden dann aber in MPEG-2 übernommen, und MPEG-3 fallengelassen.

6. MPEG-7

Die Entwicklung von neuen Standards geht weiter- bei MPEG-7 wird die Beschreibung verschiedener Arten von Multimedia-Informationen standardisiert. Das soll schnelles und effizientes Suchen nach Kriterien ermöglichen. MPEG-7 ist auch als „Multimedia Content Description Interface“ bekannt. Es wird aber nicht die Funktion einer Suchmaschine integriert.

7. MPEG-21

MPEG-21 beschreibt eine weitere Arbeitsgruppe. Diese hat die Aufgabe, die verschiedenen Elemente, die zum Aufbau einer Infrastruktur zum Verteilen und Darstellen von Multimedia-Inhalten notwendig sind, zu analysieren, Lücken bezüglich existierender Standards aufzuzeigen und entsprechende Ergänzungen bzw. neue Standards vorzuschlagen.

8. MPEG-4

Vorgabe für MPEG-4 war, eine Codierung und ein flexibles Format zu entwickeln, um Audio und visuelle Daten aller Art für Multimedia-Anwendungen (z.B. Interaktives, digitales Fernsehen) zu übertragen, bzw. zu speichern. Speziell das Erreichen (sehr) niedriger Datenraten (niedriger als MPEG-1), die für Internetanwendungen (bzw. Telekommunikation,..) brauchbar sind, standen zunächst im Mittelpunkt.

MPEG-4 ersetzt aber weder MPEG-1 noch MPEG-2. Es geht vielmehr um einen neuen Aufgabenbereich, der über die bloße Kompression von Audio- und Videodaten hinausreicht. Es wird standardisiert, wie verschiedenste Arten an multimedialen Informationen, einschließlich Echtzeit-kritischer Elemente wie Audio oder Video aufbereitet, übertragen und dargestellt werden.

8.1 Aufgaben von MPEG-4

MPEG-4 umfasst die *Produktion, Aufbereitung, Übertragung und Darstellung* von multimedialen Inhalten.

Für die Entwickler von Inhalten ist MPEG-4 interessant, weil es die Wiederverwendung von Inhalt erleichtert und mehr Flexibilität bzw. neue Möglichkeiten- durch die interaktive Verbindung von WWW, animierten Grafiken, digitalem Fernsehen,...- bietet. In MPEG-4 ist ein Mechanismus zum Schutz von Urheberrechten vorgesehen.

8.2 Eigenschaften von MPEG-4

+Kompression:

Es sollen noch erkennbare Audio-Videoinformationen bei extrem niederen Datenraten von 5-64 kbit/s geboten werden. Nach oben hin liegt die Grenze für Video/Film/TV bei 10 Mbit/s.

+Interaktivität:

MPEG-1 und MPEG-2 sind reine Distributionsformate für Audio/Video. MPEG-4 bietet Interaktivität, erlaubt den Beobachter in den Abspielvorgang einzugreifen: lokale Manipulation der MPEG-4 Szene und auch Datenaustausch zum Sender über einen Rückkanal. Dabei unterstützt MPEG-4 Client-Server-Anwendungen (Netzwerk-basiert oder auch Broadcasting) und Massenspeicher-Anwendungen (z.B. DVD).

+Objektorientierter Ansatz:

Neben komprimierten Audio- bzw. Videodaten können andere Informationen aufgenommen werden: Grundlage sind audiovisuelle Inhalte, Media- Objekte (siehe unten) genannt, die Audio, Video, Text, Grafik,... enthalten können. Eine Szenenbeschreibung –angelehnt an VRML (Virtual Reality Modeling Language: Damit ist es möglich, z.B. 3D-Modelle im Internet zu betrachten, um diese herumzufahren,...) - legt fest, wie diese in Beziehung zueinander stehen und die Darstellung beim Betrachter erfolgen soll.

+Übertragungsweg:

Bei MPEG-1 und MPEG-2 ging man von einer zuverlässigen Übertragung mit hoher Bandbreite aus. Diese Voraussetzung ist aber derzeit für Internet- oder UMTS (Universal Mobile Telecommunications: Mobilfunkstandard der 3. Generation: es werden deutlich höhere Datenübertragungsraten möglich (theoretisch bis 2Mbit/s), Mobiler Internetzugang, mobile multimediale Anwendungen,...) Anwendungen nicht gegeben- das wird bei MPEG-4 berücksichtigt. Die Daten werden ggf. in verschiedenen Qualitäten angeboten (Skalierbarkeit), um Bandbreiten optimal nutzen zu können. MPEG-4 legt auch fest, wie Daten der Media- Objekte übertragen werden, und wie sichergestellt wird, dass sie im gewünschten zeitlichen Bezug wiedergegeben werden.(„Synchronizing“ :Synchronisation der Daten der Media-Objekte)

Media- Objekte:

Audiovisuelle Szenen bestehen aus Media- Objekten, die hierarchisch organisiert sind.(Baumdiagramm)

MPEG-4 standardisiert eine Reihe von Media- Objekten, wie z.B.:

„Primitive“ Media- Objekte:

- +Standbilder (z.B. für Hintergrund)
- +Video-Objekte (z.B. Sprecher im Vordergrund)
- +Audio-Objekte (z.B. Sprache)
- +...

Es sind auch Text, Grafiken, symbolische Angaben (z.B. Phoneme, um einen künstlichen Gesprächspartner sprechen zu lassen) oder 2D/3D-Informationen denkbar.

Die Szenenbeschreibung baut auf Konzepten von VRML („Virtual Reality Modeling Language“) auf, geht aber über VRML hinaus -> das bei MPEG-4 verwendete Format nennt sich BIFS („Binary Format for Scenes“)

8.4 Audio

Es wird Sprache sowie Mono- bis Mehrkanal- Audio in hoher Qualität unterstützt.

Sprache:

Die Sprachkodierung nutzt Bitraten von 2 bis 24 kbit/s (bei Verwendung von CBR (Konstante Bitrate)) bzw. bis herunter zu 1,2 kbit/s bei Verwendung von VBR (Variabler Bitrate) . Bei Verwendung von synthetischer Sprache ist es möglich, die Bitrate auf 200 bps bis 1,2 kbit zu reduzieren. Dies funktioniert entweder auf reiner Text-Phonem-Basis oder zusammen mit Zusatzinformationen wie Pitch Contour, Phonem-Dauer,..., um der Stimme Persönlichkeit zu verleihen. Zur Funktionalität gehören : Lip-Sync (Lippen-Synchronisation zum Ton) über Phonem-Information, Abspielfunktionalität (pause, vor, zurück,...), Erhaltung von Spracheigenschaften (Geschlecht, Alter,...) und die Übermittlung von Parametern zur Gesichtsinformation eines synthetischen Sprechers.

Musik:

Musik wird ab einer Bitrate von 6 kbit/s (Mono, Bandbreite 4kHz) bis Mehrkanal-Audio unterstützt.

8.5 Video

Video-Bitraten liegen zwischen 5kbit/s bis 10 Mbit/s. Es werden sowohl Halb- als auch Vollbildmodi unterstützt. Auflösungen reichen bis HDTV (Siehe Kapitel 5).

Erneuerung zu MPEG-2:

+Bildinhalt- abhängige Codierung und Kompression:

Es ist möglich, das Bild in verschiedene Bestandteile zu zerlegen (z.B. Sprecher im Vordergrund, unbewegter Hintergrund,..), und diese *einzel*n zu behandeln. (z. B. könnte der Hintergrund höher komprimiert und mit einer geringen Bildwiederholrate codiert werden,...)

+“Shape-Coding“:

Shape: Der „Umfang“ eines Bildinhaltes:

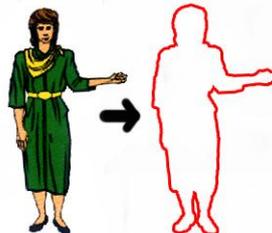


Bild 29: Shapes

Einzelne Objekte vor einem Hintergrund werden mit „Shapes“ dargestellt, und bei (z.B. kontinuierlicher Bewegung) nur die Shapes übertragen. (Der Inhalt des Shapes muß nur einmal übertragen werden)

+Face Animation:

Hier wird festgelegt, wie die Parameter zur Kalibrierung und Animation von computergenerierten Gesichtern aufgebaut sind. Dazu gehören: Definition und Codierung der Parameter zur Gesichtsanimation, Textur (Textur = eine Art „Foto“, die auf die Oberfläche eines 3D-Objektes gelegt wird) eines Gesichtes, Lippenkonfiguration ähnlich zu Phonemen der Sprache, „Visemes“ genannt, 3D-Geometrie- Beschreibungen.

Als Beispiel: Über eine Wireless- Verbindung soll Bildkommunikation realisiert werden. Da die Bandbreite nicht ausreicht, wird mit synthetischen Köpfen gearbeitet, die entsprechend der Sprache animiert werden (Lip-Sync). Zu Beginn der Übertragung wird ein Bild von beiden Gesprächspartnern digitalisiert, und als Textur auf die 3D-Kopf-Geometrie gelegt. So müssen nur Sprache und Parameter für die Steuerung der Gesichtsanimation übertragen werden. Noch weniger Bandbreite würde bei Verwendung synthetischer Sprache gebraucht werden. Diese Ideen sind nicht neu, werden aber erstmals in MPEG-4 standardisiert.

8.6 Erweiterung in MPEG-4 Version 2:

+Advanced BIFS (Binary Format for Scenes):

Umgebungsmodelle für Audio und Parameter für Body-Animation (Ganzkörperanimation).

+Festlegung eines MPEG-4 Dateiformates:

Es basiert auf dem Quicktime- Format von Apple, eignet sich für Speicherung und Streaming. Kleinste Einheiten werden „Atome“ genannt, die durch einen ID-Tag und ihre Länge definiert werden.

+MPEG-J:

MPEG-J beschreibt eine API (Programmierschnittstelle) für die Programmierung von MPEG-4-Media-Playern mit der Programmiersprache Java. Die gemeinsame Übertragung von MPEG-4-Media und ausführbarem Java-Code soll auch komplexe Anwendungen, z.B. Spiele erlauben.

+Übertragung mehrerer Ansichten

Es ist möglich, eine Szene aus verschiedenen Positionen zu betrachten (falls angeboten). Auch könnten Stereophone Bilder/Videos übertragen werden.

+Video

Verbesserte Algorithmen bringen eine Einsparung von 50% der Bandbreite:

+GMC („Global Motion Compensation“):

Die Bewegungs-Erkennung bzw. -Kompensation wurde verbessert: Die Bewegung von Objekten kann mit wenigen Parametern beschrieben werden.(Vektoren,...)

+Einführung der SA-DCT (Shape Adaptive Discrete Cosinus Transformation):
Maskeninformationen werden in der DCT berücksichtigt.
+Standbilder werden nicht mehr im JPEG- Format, sondern per Wavelet-
Transformation komprimiert. Dies bringt bei geringen Datenraten ein besseres Bild,
und hinterlässt weniger Artefakte .(z.B. Blockbildung) Zudem ist während des Ladens
eine niedrig aufgelöste Voransicht möglich.

Mit MPEG-4 wird versucht, eine bislang von vielen unterschiedlichen Formaten bestimmte
Multimedia-Umgebung, zu standardisieren. Erste Geräte, die MPEG-4 benutzen, sind bereits
auf dem Markt. (Realtime- Encoder, Internet- Viewcams,...)
MPEG-4 soll auch in Verbindung mit der DVD verwendet werden- dies bedingt allerdings
eine komplett neue Abspieltechnologie.

9. Literaturverzeichnis

MPEG Audio Layer-1 bis 3:

Davis Pan: A Tutorial on MPEG/Audio Compression, IEEE Multimedia Journal Summer 1995

Davis Pan: Digital Audio Compression, Digital Technical Journal Vol. 5 No. 2, Spring 1993

T. Sporer, K. Brandenburg, B. Edler: The use of multirate filter banks for coding of high quality digital audio, 6th European Signal Processing Conference (EUSIPCO), Amsterdam, June 1992, Vol. 1, pages 211-214

K. Brandenburg, G. Stoll: The ISO/MPEG-Audio Codec: A Generic Standard for Coding of High Quality Digital Audio, Audio Engineering Society Preprint 3336 [2TM1.02]

Chiariglione, Leonardo: Multimedia Communication (MPEG-II), Brussels 1992

c't (Kürzel ku): MPEG-2-Standard festgeklopft, Heft 6, 1993

Gadegast Frank: Offene Dokumentverarbeitung mit multimedialen Standards, Referat, Jan. 1993

Gadegast Frank, Jürgen Meyer: Integrität von Multimedialen Daten am Beispiel MPEG, Semesterarbeit, Jun. 1993

Internet: www.mp3.com, www.mpeg.org, www.mpeg1.de, www.mpeg2.de, www.itu.int, www.iso.ch, www.mp3tech.org.

Sourcecodes, Software, Hintergründe: <http://lame.sourceforge.net/>

AAC:

Davis Yen Pan: A Tutorial on MPEG/Audio Compression, IEEE Multimedia Journal, Vol. 2, No. 2, S. 60 ff. (Als PostScript-Datei: <http://www.bok.net/~pan/>)

Davis Yen Pan: Digital Audio Compression, Digital Technical Journal, Vol. 5, No. 2, 1993 (Als PostScript-Datei: ftp://ftp.digital.com/pub/Digital/info/DTJ/v5n2/Digital_Audio_Compression_01oct1993DTJA03P8.ps)

Matthias Carstens: Musik kompakt - Audio-Kompression mit MPEG Layer-3, c't-Magazin, Heft 21, 1998, S. 242 ff.

Production Partner: AAC-Advanced Audio Coding, Heft 9, 2000

Björn Wesén: A DSP-based decompressor unit for high-fidelity MPEG-Audio over TCP/IP networks, <http://www.sparta.lu.se/~bjorn/whitney/compress.htm>

Homepage des Fraunhofer-Instituts für integrierte Schaltungen: <http://www.iis.fhg.de/amm/>

Transformationen:

Wavelet-Transformation:

http://i31www.ira.uka.de/docs/semin94/03_Wavelet/node7.html

<http://www.ondelette.com/indexen.html>

<http://engineering.rowan.edu/~polikar/WAVELETS/WTtutorial.html>

<http://www.public.iastate.edu/~rpolikar/WAVELETS/waveletindex.html>

DCT:

<http://www.cs.cf.ac.uk/Dave/Multimedia/node231.html>

<http://documents.wolfram.com/applications/digitalimage/UsersGuide/8.4.html>

K. R. Rao and P. Yip, *Discrete Cosine Transform: Algorithms, Advantages, Applications* (Academic Press, Boston, 1990).

A. V. Oppenheim, R. W. Schaffer, and J. R. Buck, *Discrete-Time Signal Processing*, second edition (Prentice-Hall, New Jersey, 1999).

DFT:

<http://ccrma-www.stanford.edu/~jos/mdft/mdft.html>

<http://www.spd.eee.strath.ac.uk/~interact/fourier/dft.html>

A. V. Oppenheim, R. W. Schaffer, and J. R. Buck, *Discrete-Time Signal Processing*, second edition (Prentice-Hall, New Jersey, 1999).

MPEG-4

http://leonardo.telecomitalia.com/icjfiles/mpeg-4_si/5-BIFS_paper/5-BIFS_paper.htm

<http://www.tele.ntnu.no/users/andrew/Papers/LS-MPEG-4/MPEG-4%20BIFS.pdf>

Production Partner: MPEG 4, Heft 9, 2000

www.mpeg4.net

<http://www.media.mit.edu/~eds/mpeg4>

10. ANHANG

10.1 Preismodell bei MP3

Overview

		PC Software Applications
mp3	Decoder	- US\$ 0.75 per unit or US\$ 50 000.00 - US\$ 60 000.00 one-time paid-up
	Encoder / Codec	- US\$ 2.50 - US\$ 5.00 per unit
mp3PRO	Decoder	- US\$ 1.25 per unit or US\$ 90 000.00 one-time paid-up
	Encoder / Codec	- US\$ 5.00 per unit
		Hardware Products
mp3	Decoder	- US\$ 0.75 per unit
	Encoder / Codec	- US\$ 2.50 - US\$ 5.00 per unit
mp3PRO	Decoder	- US\$ 1.25 per unit
	Encoder / Codec	- US\$ 5.00 per unit
		ICs / DSPs
For available software, supported platforms, porting and licensing options, please contact us at info@mp3licensing.com .		
		Games
mp3		- US\$ 2 500.00 per title
mp3PRO		- US\$ 3 750.00 per title
		Electronic Music Distribution / Broadcasting / Streaming
mp3		- 2.0 % of related revenue
mp3PRO		- 3.0 % of related revenue

Quelle: <http://www.mp3licensing.com/royalty/index.html>

Reine Decoder/Player sind billiger als Encoder oder komplette Codecs, mp3pro ist teurer als mp3. Auf der Webseite kann man auch eine Liste der Lizenznehmer einsehen, wie Philips, Microsoft, ...

10.2 MPEG-2 AAC

Lizenzmodell

Licensing Terms

Standard Rates:

Volume (per channel / quarterly reset)	Consumer Decoder or Encoder Channels	Consumer Codec Channels	Professional Decoder Channels	Professional Encoder Channels
	Implementations			
	\$ 0.00	\$ 0.00	\$ 0.00	\$ 0.00
	Products			
Flat Rate	n/a	n/a	\$ 2.00	\$ 20.00
1 to 100,000	\$ 0.50	\$ 1.00	-	-
100,001 to 500,000	\$ 0.37	\$ 0.74	-	-
500,001 to 1,000,000	\$ 0.27	\$ 0.54	-	-

1,000,001 to 5,000,000	\$ 0.22	\$ 0.44	-	-
5,000,001 to 10,000,000	\$ 0.17	\$ 0.34	-	-
10,000,001 or more	\$ 0.12	\$ 0.24	-	-

PC-Based Software Pricing: (for private non-commercial use)

Volume (per channel / annual reset)	Consumer Decoder Channels	Consumer Encoder Channels	Consumer Codecs Channels
1 to 100,000	\$ 0.25	\$ 0.50	\$ 0.75
100,001 to 500,000	n/a	\$ 0.37	\$ 0.62
500,001 to 1,000,000	n/a	\$ 0.27	\$ 0.52
Maximum annual payment	\$ 25,000	\$ 250,000	\$ 275,000

Quelle: <http://www.aac-audio.com/Licensing/>