

HOCHSCHULE FÜR FILM UND FERNSEHEN
POTSDAM BABELSBERG

STUDIENGANG TON

VORDIPLOMARBEIT

FUNKTIONSWEISE UND VERGLEICH NEUER
AUDIODATENREDUKTIONSVORFAHREN

VORGELEGT VON

IMMO TRÜMPELMANN
5. SEMESTER, MÄRZ 2002
MATRIKELNUMMER 3756

Zielstellung

Die derzeitige Entwicklung von Hörgewohnheiten zu immer mehr Audiokanälen in immer besserer Qualität geht einher mit einer enorm ansteigenden Audiodatenflut, der CDs, DVDs, Speicherchips in portablen Playern etc. ohne leistungsfähige Datenreduktionsverfahren nicht mehr gewachsen sind. Auch der häufige Wunsch, Audiomaterial in hoher Qualität bei geringem Speicherplatzbedarf zu archivieren oder gar per DSL oder ISDN durch das Internet zu versenden und auszutauschen, verlangt nach Audiodatenreduktionsverfahren, die über höchste Effizienz und Qualität bei geringen Bitraten verfügen. Bei der Vielfalt an Verfahren und den in immer kürzeren Abständen auftauchenden Versionen und Schlagwörtern rund um Audiokompression ist es oft schwer, noch einen einigermaßen zusammenhängenden Überblick über die neuen Verfahren zu erlangen. Zudem ist der Kenntnisstand über die Wirkungsweise und Möglichkeiten der jeweiligen Kompressionen meist gering, was oft zu großen Unsicherheiten in ihrer Anwendung führt. Diese vorliegende Arbeit möchte gezielt über die aktuellen Audiokompressionsverfahren informieren, wobei der Schwerpunkt auf der Funktionsweise, den technischen Parametern und, wo es sinnvoll erscheint, auf dem Vergleich dieser Verfahren liegt. Genauer untersucht werden:

- der mittlerweile sehr weit verbreitete **MP3-Standard** und dessen Weiterentwicklung **MP3-Pro**
- der neue, hochkomplexe **MPEG-4** Standard und speziell seine wichtigsten Kompressionsverfahren **AAC** und **TwinVQ**
- die aktuellen Versionen von Sonys **ATRAC**-Kompression
- Verfahren, die ursprünglich aus dem Mehrkanal-Film Bereich stammen und nun durch die DVD für den Heimbereich interessant wurden – namentlich **AC3** und **DTS**
- die alternative Open-Source¹ **OGG**-Kompression
- die beiden verlustfreien Kompressionen **MLP** und **DST**.

Vorkenntnisse in den Gebieten Psychoakustik, Digitalem Audio und Fourier-Analyse werden für das Verständnis dieser Arbeit vorausgesetzt. Als Anlage zu dieser Arbeit gehört eine Audio-CD mit einigen Hörbeispielen zu den hier vorgestellten Verfahren.

¹ Open Source bedeutet, dass der Programmquellcode für Programmierer für Weiterentwicklungen offen liegt und keine Patent oder Lizenzgebühren zu entrichten sind.

INHALTSVERZEICHNIS

1 ALLGEMEINE GRUNDLAGEN DER WAHRNEHMUNGSBEZOGENEN KODIERUNG... 5	5
2 MPEG AUDIOKOMPRESSION..... 7	7
2.1 DIE VERSCHIEDENEN MPEG VERSIONEN..... 7	7
2.2 DIE VERSCHIEDENEN MPEG-1 LAYER..... 7	7
3 MPEG-1 LAYER 3 (MP3)..... 8	8
3.1 ENTWICKLER UND ERSCHEINUNGSDATUM 8	8
3.2 ÜBER ENCODER UND DECODER..... 8	8
3.3 EINSATZGEBIET UND QUALITÄT..... 8	8
3.4 EIGENSCHAFTEN DER MP3-KOMPRESSION 10	10
<i>Nachteile und Kompromisse</i> 10	10
<i>Vorteile</i> 10	10
3.5 ARBEITSWEISE DER MP3 KOMPRESSION..... 10	10
4 MP3-PRO..... 12	12
4.1 ENTWICKLER UND ERSCHEINUNGSDATUM 12	12
4.2 ÜBER ENCODER UND DECODER..... 12	12
4.3 EIGENSCHAFTEN, EINSATZGEBIET UND QUALITÄT VON MP3-PRO 12	12
4.4 ARBEITSWEISE DER MP3-PRO KOMPRESSION..... 13	13
5 MPEG-2..... 13	13
5.1 NEUERUNGEN UND UNTERSCHIEDE ZU MPEG-1 13	13
6 AAC (ADVANCED AUDIO CODING) 14	14
6.1 ENTWICKLER UND ERSCHEINUNGSDATUM 14	14
6.2 EINSATZGEBIET UND QUALITÄT..... 14	14
6.3 EIGENSCHAFTEN DER AAC KOMPRESSION 14	14
6.4 ARBEITSWEISE DER AAC KOMPRESSION 15	15
7 TWIN-VQ (TRANSFORM-DOMAIN WEIGHTED INTERLEAVE VECTOR QUANTIZATION)..... 16	16
7.1 ENTWICKLER 16	16
7.2 EINSATZGEBIET UND QUALITÄT..... 16	16
7.3 EIGENSCHAFTEN DER TWIN-VQ KOMPRESSION 16	16
8 MPEG-4..... 16	16
8.1 ENTWICKLER UND ERSCHEINUNGSDATUM 16	16
8.2 ZUM VERSTÄNDNIS VON MPEG-4 UND DESSEN EINSATZGEBIET 17	17
8.3 NEUE EIGENSCHAFTEN VON MPEG-4..... 17	17
8.4 QUALITÄT UND EFFIZIENZ VON MPEG-4..... 18	18
8.5 DIE EINZELNEN VERFAHREN DER MPEG-4 AUDIOKOMPRESSION 18	18
8.6 AUDIOPROFILE UND KOMPLEXITÄTSSTUFEN (LEVEL) BEI MPEG-4..... 19	19
9 ATRAC (ADAPTIVE TRANSFORM ACOUSTIC CODING) 21	21
9.1 ENTWICKLER UND ERSCHEINUNGSDATUM 21	21
9.2 EINSATZGEBIET UND QUALITÄT..... 21	21
9.3 EIGENSCHAFTEN DER ATRAC KOMPRESSION 22	22
9.4 ARBEITSWEISE DER ATRAC KOMPRESSION 22	22
<i>ATRAC1</i> 22	22
<i>Besonderheiten bei der Neuen ATRAC3-Kompression</i> 22	22

10 AC3 (DOLBY DIGITAL)	23
10.1 ENTWICKLER UND ERSCHEINUNGSDATUM	23
10.2 EINSATZGEBIET UND QUALITÄT	23
10.3 EIGENSCHAFTEN DER AC3 KOMPRESSION	23
10.4 ARBEITSWEISE DER AC3 KOMPRESSION	24
10.5 ÜBER DEN DECODER	24
11 DTS (DIGITAL THEATER SYSTEMS)	25
11.1 ENTWICKLER UND ERSCHEINUNGSDATUM	25
11.2 EINSATZGEBIET UND QUALITÄT	25
11.3 EIGENSCHAFTEN DER DTS KOMPRESSION	25
11.4 ARBEITSWEISE DER DTS KOMPRESSION	26
11.5 ANDERE DTS STANDARDS	26
12 OGG VORBIS	26
12.1 ÜBER DIE OGG KOMPRESSION	26
12.2 EIGENSCHAFTEN DER OGG KOMPRESSION	26
12.3 ARBEITSWEISE DER OGG KOMPRESSION	26
13 DIE VERLUSTLOSEN KOMPRESSIENEN MLP UND DST	27
13.1 ÜBER MLP (MERIDIAN LOSSLESS PACKING)	27
13.2 ÜBER DST (DIGITAL STREAM TRANSFER)	27
13.3 ÜBER DIE ARBEITSWEISE VON MLP UND DST	27
14 TABELLARISCHE ÜBERSICHTEN UND VERGLEICH	28
14.1 ÜBERSICHT DVD/KINO-MEHRKANALCODECS	28
14.2 ÜBERSICHT AUDIOCODECS AUS DEM COMPUTER-/INTERNETBEREICH UND ATRAC	29
15 SCHLUSSBETRACHTUNG	29
16 ANLAGE	30
17 GLOSSAR	32
LITERATURVERZEICHNIS UND LINKS ZUM THEMA	33
ALLGEMEINE QUELLEN:	33
MPEG-KOMPRESSION:	34
ATRAC-KOMPRESSION:	34
AC3-KOMPRESSION:	35
DTS-KOMPRESSION:	35
OGG-KOMPRESSION	35
MLP UND DST-KOMPRESSION	35

1 ALLGEMEINE GRUNDLAGEN DER WAHRNEHMUNGSBEZOGENEN KODIERUNG

Die grundlegende Arbeitsweise von Datenreduktionsverfahren für Audiomaterial ist trotz diverser Standards und Bezeichnungen wie z.B. AC3, ATRAC, MP3 mit ihren speziellen Anwendungsgebieten und technischen Feinheiten fast immer dieselbe. Letztendlich geht es stets um die geschickte Weglassung von Irrelevanzen und Redundanzen und die damit verbundene Datenreduktion des Tonmaterials.

Die Reduktion von Redundanzen erfolgt verlustfrei wie bei Packprogrammen aus der Computerwelt (ZIP, RAR, ACE usw.) unter Weglassung von identischen und mehrfach im Material enthaltenen Daten. Redundant sind also Teile der Information, die auch aus anderen Informationsteilen generiert werden können. So kann beispielsweise eine Sinuswelle entweder aus einer Vielzahl von Funktionswerten ermittelt werden, genauso kann sie aber auch eindeutig aus lediglich vier Angaben, nämlich Frequenz, Amplitude, Phase und Dauer generiert werden. Besonders häufig wird bei der Audiokompression zur Redundanzbeseitigung eine spezialisierte Variante der sogenannten *Huffman-Kodierung* verwendet. Dieser Kompressionsalgorithmus wird schon länger in der Text- und Bildkompression eingesetzt. Hierbei werden bestimmte häufige Zeichenfolgen durch kürzere Folgen ersetzt. Die Huffman-Kodierung benutzt zur Kodierung eine variable Bit Länge, das heißt, die am häufigsten verwendeten Zeichen oder Zeichenfolgen bekommen die geringste Anzahl an Bits zugeordnet. Ein Beispiel: Jedes ASCII²-Zeichen ist fest mit 7 Bit kodiert. Der Huffman-Code überprüft, welche Buchstaben am häufigsten verwendet werden, z.B. das „E“, und weist dem Buchstaben E, wo immer er auftritt, nur noch eine sehr kurze Bitfolge zu (< 7 Bit). Zu jeder komprimierten Datei oder jedem komprimiertem Abschnitt gibt es deshalb eine Huffman-Tabelle, in der festgelegt ist, welches Bit welches Zeichen oder welche Zeichenfolge kodiert. Ist diese Tabelle erst einmal definiert, können die Redundanzen in einer Zeichenfolge effektiv herausgefiltert und weggelassen werden. Geschickte Weglassung von Redundanzen führt meist zu einer Datenreduktion so um den Faktor zwei.

Die Reduktion von Irrelevanzen erfolgt verlustbehaftet, doch sind die Weglassungen im Tonmaterial bei guter Kompression und höheren Bitraten für den Menschen nicht hörbar. Basierend auf Studien über das menschliche Gehör entscheidet der Encoder, welche Informationen wichtig sind und welche nicht. Bevor der Mensch etwas hören kann, filtern Ohr und Gehirn unwichtige Daten weg. Audiokompression schiebt diesen Schritt lediglich weiter nach vorne. Das menschliche Hören basiert nicht auf Samples aus den weitverbreiteten PCM (Pulse Code Modulation)-Wandlern, sondern auf den sogenannten Frequenzgruppen³. Es gibt drei Basis-Techniken zur irrelevanten Audiokompression:

- **Predictive Coding:** Das Wissen über das vergangene Signal wird für die Vorhersage des nächsten Samples benutzt. Die Kompression besteht darin, dass man nur noch die Differenz zwischen dem Signal und der Vorhersage abspeichern muss, die weniger Bits beansprucht.
- **Sub-Band Coding:** Das Audiospektrum wird in verschiedene Frequenzbänder aufgeteilt. Dabei kann ausgenutzt werden, dass fast alle Bänder viel weniger (oder weniger wichtiges/hörbares) Signal beinhalten als das lauteste Band. Die Kompression beruht darauf, dass für die wichtigen Bänder mehr Bits aufgewendet werden, als für die unwichtigen. In manchen Fällen kann man sich die Übertragung auch ganz sparen. Erklärung: Laute Signale verdecken benachbarte Signale in anderen Frequenzbändern, sowohl zeitlich (Vor- und Nachverdeckung) wie auch spektral (Herausbildung von Maskierungsschwellen⁴ um das laute

² Abkürzung für "American Standard Code for Information Interchange". Durch definierte 7 Bit-Wörter werden 128 Zeichen erfasst, die das einfache Alphabet und einige Steuerungs-codes z.B. für Drucker enthalten.

³ Die Frequenzgruppe (engl.: critical band) ist das Frequenzband, innerhalb dessen das menschliche Gehör Intensitäten zusammenfasst. Im Frequenzbereich von 20 Hz bis 16 kHz gibt es etwa 24 Frequenzgruppen. Mit steigender Frequenz nimmt ihre Bandbreite zu (von 100 Hz für tiefe Frequenzen bis 4 kHz für Höhen).

⁴ Maskierung: Überdecken eines Signals durch ein anderes, z.B. maskiert ein Preßlufthammer ein Handy-Telefonat. Das alte Signal ist noch vorhanden, es wird aber nicht mehr wahrgenommen.

Signal herum in der Frequenzdarstellung). Nicht wahrzunehmende Frequenzgruppen werden nicht mitkodiert. Die Bits werden dynamisch auf die Bänder verteilt und dem Encoder fällt die aufwendige Selektionsarbeit zu, was noch zu hören ist und was nicht, für die ein sogenanntes psychoakustisches Modell benutzt wird. Der Encoder muss neben den Audiodaten auch Informationen über die Bitverteilung übertragen. Es werden immer nur soviel Bits pro Frequenzgruppe angewendet, dass das entstehende Quantisierungsrauschen⁵ noch gerade unter der Maskierungskurve liegt. Bleiben Bits nach der Kodierung noch übrig, legt der Encoder sie in ein Bitreservoir. Fehlen ausreichend Bits, so muss der Encoder Quantisierungsrauschen über der Maskierungsschwelle einfügen. Der Decoder muss das Signal nur synthetisieren.

- **Spectral oder Transform Coding:** Über die Wellenform wird periodisch eine Fourier- oder diskrete Kosinus-Transformation berechnet. Somit erhält man das Frequenzspektrum einer Schwingung über einen gewissen kurzen Zeitraum. Weil sich die transformierte Darstellung eines Signals nur langsam ändert, kann die Übertragung seltener erfolgen. Transform-Encoder benutzen gewöhnlich eine hohe Anzahl an Subbändern und betrachten von der Frequenz her benachbarte Samples zusammen. Auch hier wird ein psychoakustisches Modell auf das Audiomaterial angewendet und eine dynamische Bitverteilung vorgenommen, um Irrelevanzen nicht zu übertragen. Eine Filterbank ist das Hardware-Äquivalent von einer Fourier- oder Cosinus-Transformation und besteht aus vielen parallelen Bandpass-Filtern. Ein Decoder wandelt durch inverse Transformation die Frequenzwerte wieder in Zeitwerte um und fügt die umgewandelten Blöcke sinnvoll aneinander.

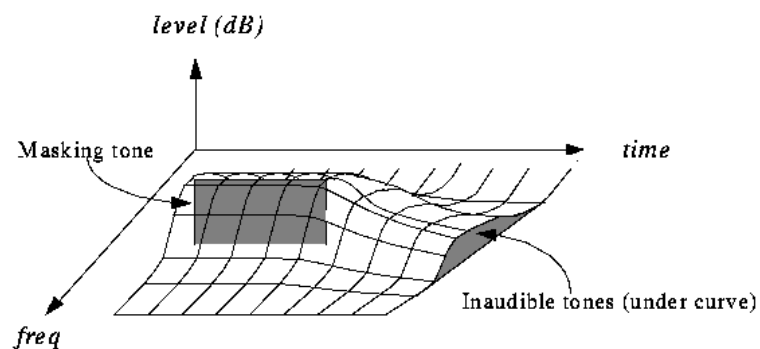
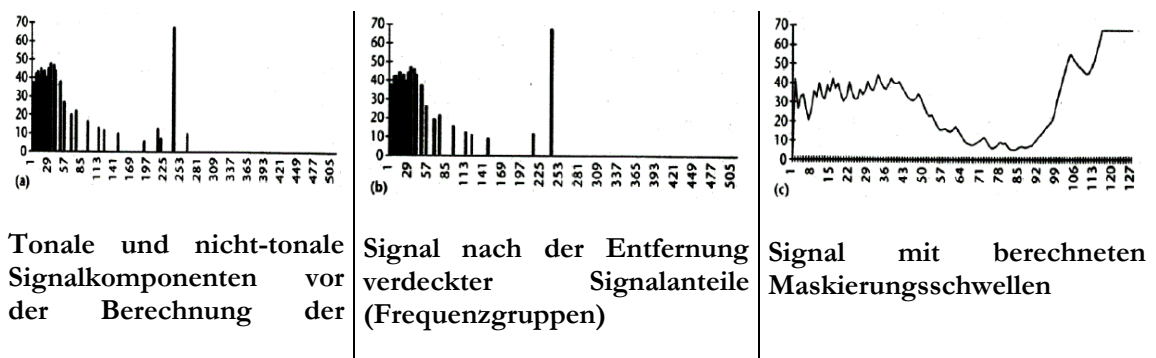


Abb. 1: Zeitliche und frequenzabhängige Verdeckung durch Sinustöne,
Quelle: Steinmetz and Nahrstedt (1996): Audio Compression



⁵ Quantisierungsrauschen: Durch die beschränkte Auflösung/Quantisierung von Samples und Frequenzgruppen entstehen kantige Übergänge im Signal, die bei der Digital/Analog-Wandlung ein Rauschen erzeugen.

Maskierungsschwellen

Abb. 2:

Bei Mehrkanalsignalen können zusätzlich auch Redundanzen zwischen den Kanälen genutzt werden, um die Audiodaten zu reduzieren. Mehr dazu jedoch bei den einzelnen Verfahren selbst. Durch die geschickte Entfernung von Irrelevanzen wird heutzutage ein Kompressionsgrad um den Faktor fünf bis sechs erreicht.

2 MPEG AUDIOKOMPRESSION

2.1 DIE VERSCHIEDENEN MPEG VERSIONEN

Die **Motion Pictures Expert Group** arbeitet seit ihrer Gründung im Jahr 1988 an der Standardisierung von digitalen Kompressionsverfahren für Video und Audio. Unter den Mitgliedern dieser Organisation befinden sich neben vielen anderen AT&T, AOL Time Warner, IBM, Sony, NEC, NTT, Dolby Labs, Deutsche Telekom, Thomson, Panasonic, die Universität Hannover und das Fraunhofer Institut für integrierte Schaltungen Erlangen.

Zur Geschichte der von der MPEG-Organisation entwickelten Standards:

- **MPEG-1** betrifft die Kodierung von Bewegtbild und dazugehörigem Audio für digitale Speichermedien bis zu einer Datenrate von 1,5 MBit/s, International Standard ISO/IEC-11172, abgeschlossen Okt. 92 (Verwendung bei z.B. Video CD, MP3 etc.)
- **MPEG-2** betrifft die generische Kodierung von Bewegtbild und dazugehörigem Audio, u.a. Definition der Advanced Audio Kodierung (AAC), International Standard ISO/IEC-13818, abgeschlossen Nov. 94 (Verwendung bei z.B. Set Top Box beim Digitalen Fernsehen, DVD)
- **MPEG-3** nicht existent, wurde in MPEG-2 integriert
- **MPEG-4** betrifft die komplexe Kodierung von multimedialen Inhalten bei sehr geringen Bitraten, International Standard ISO/IEC 14496, abgeschlossen Feb. 99 (zu erwartende Verwendung bei z.B. Digital Fernsehen, Transport von multimedialen Inhalten im Internet, Mobilanwendungen)
- **MPEG-7** betrifft die Standardisierung zur Beschreibung und Suche von audiovisuellen Inhalten, formeller Name: "Multimedia Content Description Interface"; MPEG-7 soll die Suche, die Filterung, den Überblick bei multimedialen Inhalten im Netz durch z.B. Metadaten und spezielle standardisierte Hilfsmittel verbessern, befindet sich noch in der Entwicklung
- **MPEG-21** betrifft den Begriff „Multimedia Framework“ und soll das Zusammenspiel und die gemeinsame Entwicklung verschiedener Elemente in multimedialen Inhalten standardisieren und optimieren, seit Juni 2000 in der Entwicklung

2.2 DIE VERSCHIEDENEN MPEG-1 LAYER

Für MPEG-1 Audio sind drei verschiedene Kompressionsalgorithmen definiert, die als Layer 1, 2 und 3 bezeichnet werden und die Audiodaten unterschiedlich stark komprimieren. Die höheren Codecs⁶

⁶ Ein Codec ist ein Programm, das die Komprimierung einer Multimedia-Datei - z.B. einem Video oder Musikstück - sowie dessen Wiedergabe durch einen Media Player ermöglicht

schließen dabei die jeweils niedrigeren mit ein, d. h. ein Codec für Layer 3 muss auch Layer 1 und 2 beherrschen. Die Codecs arbeiten mit einer breiten Palette von Abtastfrequenzen und Bitraten bei maximal zwei Kanälen:

Verfahren	Kompressionsrate	Datenmenge bei CD-naher Qualität	Anwendungsbereiche
Originaldaten	keine, 1:1	1408 Kbit/s	Audio-CD oder DAT mit 16 Bit, 44,1 KHz
MPEG-1 Layer 1	1:4	384 Kbit/s	Digital Compact Cassette (DCC)
MPEG-1 Layer 2	1:6 bis 1:8	256 – 192 Kbit/s	Digital Audio Broadcast (DAB), CD-I
MPEG-1 Layer 3	1:10 – 1:12	128 – 112 Kbit/s	Internet Audio, Zweikanal ISDN, Satelliten Radio

Quelle Folie: Folie von Ingo Kock

3 MPEG-1 LAYER 3 (MP3)

3.1 Entwickler und Erscheinungsdatum

Die MP3 Spezifikation wird zwar von dem bereits erwähnten MPEG-Konsortium verwaltet, das eigentliche Kompressionsverfahren wurde aber nicht von der MPEG-Arbeitsgruppe entwickelt. Maßgeblich an der Entwicklung der MP3-Kompressionstechnik beteiligt waren und sind Mitarbeiter des Fraunhofer Instituts für Integrierte Schaltungen (IIS) in Erlangen, Coding Technologies und Thomson Multimedia. Das Entstehungsjahr ist 1992, weiterentwickelt wurde MP3 bis 1994.

3.2 Über Encoder und Decoder

Obwohl die Decodierung nach festgelegten Regeln zu geschehen hat, gibt die MP3-Spezifikation der ISO keine exakten Vorschriften, wie die Encodierung zu geschehen hat. Vorgegeben werden lediglich Richtlinien, nach denen die Entwickler ein gewisses einheitliches Level zu halten haben, um sicherzustellen, dass ein Encoder Dateien produziert, die plattformunabhängig auf beliebigen Decodern abgespielt werden können. Die Definition von MP3 liefert nur eine Basis; die einzelnen psychoakustischen Modelle und Filterungsverfahren bleiben der Entwicklung der Softwarehersteller überlassen, um eine möglichst hohe Produktvielfalt auf einem freien Markt zu schaffen. Jedoch bezahlen viele Hersteller lieber Lizenzgebühren für das Modell des Fraunhofer Institutes, um sich den Aufwand und die Kosten eigener Entwicklungen zu ersparen. Die Hauptziele der Entwickler sind hohe Geschwindigkeit und Qualität der Datenreduktion. Obwohl es zwischen den einzelnen Decodern fast keinen hörbaren Qualitätsunterschied gibt, sind die Differenzen bei der maximalen Geschwindigkeit und Qualität der Encoder enorm.

3.3 Einsatzgebiet und Qualität

MP3 ist zur Zeit der Internetstandard für die Übertragung, den Austausch und die platzsparende Archivierung von Musik. Durch beliebte Musiktaschbörsen wie Napster und Gnutella und kostenlose En- und Decoder erfuhr das Format eine explosionsartige Verbreitung. Immer beliebter wird das MP3-Format auch im Bereich tragbarer Unterhaltungselektronik wie portablen CD-Playern, Mini-CD Playern und speziellen MP3 Playern mit Speicherchips. Hunderte von Radiosendern senden live im Internet per

MP3-Stream⁷, der jedoch in diesem Sektor mit größtenteils geringer Klangqualität mit Streams im RealAudio-Format⁸ konkurriert.

Übersicht MPEG-1 Layer 3

Streaming-/Bitrate	Kompressionsrate	Samplingrate	Dateigröße (1 min)	Qualität
1408 Kbit/s (CD)	1:1	44 100 KHz Stereo	10 337 KByte	CD
8 Kbit/s	175:1	11 025 KHz Mono	59 Kbyte	wie Telefon, 2,5 KHz Bandbreite
20 Kbit/s	71:1	11 025 KHz Stereo 11 025 KHz Mono	146 Kbyte	wie Kurzwele, 4,5 KHz Bandbreite
32 Kbit/s	44:1	11 025 KHz Stereo 22 050 KHz Mono	233 Kbyte	wie Mittelwelle, 7,5 KHz Bandbreite
56 Kbit/s	25:1	22 050 KHz Stereo 44 100 KHz Mono	408 Kbyte	besser als Mittelwelle, 11 KHz Bandbreite
64 Kbit/s	22:1	22 050 KHz Stereo 44 100 KHz Mono	467 Kbyte	
96 Kbit/s	15:1	44 100 KHz Stereo	702 Kbyte	ähnlich UKW, 15 KHz Bandbreite
112 Kbit/s	13:1	44 100 KHz Stereo	818 Kbyte	> 15 KHz Bandbreite
128 Kbit/s	11:1	44 100 KHz Stereo	935 Kbyte	fast CD Qualität
160 Kbit/s	9:1	44 100 KHz Stereo	1 170 Kbyte	CD-Qualität
192 Kbit/s	7:1	44 100 KHz Stereo	1 404 Kbyte	
256 Kbit/s	6:1	44 100 KHz Stereo	1 873 Kbyte	Studioqualität
320 Kbit/s	4:1	44 100 KHz Stereo	2 341 Kbyte	

blau = unkomprimierte PCM-Wav-Datei, zum Vergleich

rot = fast CD-Qualität für Laienhörer, Internetstandard

Info: Weitere ungebräuchliche Bitraten: 16, 18, 24, 48, 224 Kbit/s (Fraunhofer IIS MPEG Layer 3 Codec professional)

⁷ Ein Audio-Format, das es ermöglicht, schon während der Übertragung abgespielt zu werden. Dadurch wird Übertragung von Radio-Sendungen erst möglich, da die MP3 Datei nicht erst vorher komplett heruntergeladen werden muss.

⁸ Von der Firma Progressive Networks entwickeltes Protokoll, mit dem sich Audio- oder Videodaten mit sehr geringer Bitrate in Echtzeit über das Internet übertragen lassen.

3.4 EIGENSCHAFTEN DER MP3-KOMPRESSION

Wie aus der vorhergehenden Tabelle ersichtlich ist, unterstützt die MP3 Kompression Audiomaterial mit 16 Bit, maximal zwei Kanälen und Samplingraten von 11, 22, 44 und 48 KHz und erzeugt je nach angestrebter Qualität Bitraten von 8 Kbit/s bis 320 Kbit/s, wobei sich als guter Kompromiss zwischen Qualität und geringer Bitrate ein Internetstandard von 128 Kbit/s bei CD naher Qualität⁹ etabliert hat.

NACHTEILE UND KOMPROMISSE

Bei niedrigeren Bitraten (≤ 64 Kbit/s) fehlen dem kodierten Material verstärkt die Hochtonanteile, denn der Datenstrom wird zu gering, um Audio in voller Bandbreite zu kodieren. Die MP3 Entwickler hatten in der Entstehungsphase von MP3 bei niedrigen Bitraten die Wahl zwischen hörbaren Artefakten¹⁰ oder eingengerter Bandbreite und entschieden sich dann für Bandbreiteneinengung mit geringem Artefaktanteil. Ein weiterer Kompromiss besteht zwischen der Qualität der Filter für die Zerlegung des Tonmaterials in 32 Frequenzgruppen und der Arbeitsgeschwindigkeit des MP3-Codecs. Hinzu kommt, dass das menschliche Gehör gemäß den Forschungsergebnissen der Psychoakustik im Gegensatz zur Filterbank des Mp3-Codecs komplexe Schallereignisse in exponentiell breiter werdende Frequenzgruppen einteilt. In der Einteilung in gleichgroße Frequenzgruppen besitzt die MP3-Kodierung also keine bestmögliche Effizienz. Auch haben die verwendeten Filter nur eine beschränkte Flankensteilheit und es kommt zu Bereichen, in denen sich die Frequenzbänder überlagern. Schallereignisse müssen durch dieses Übersprechen der Frequenzbänder eventuell doppelt ausgewertet werden, was letztendlich ein höheres Datenvolumen bedeutet. Des weiteren widerspricht die nicht-lineare Arbeitsweise des menschlichen Gehörs bei unterschiedlichen Wiedergabelautstärken¹¹ der technologischen Notwendigkeit bei der Kompression eine definierte Abhörlautstärke festzulegen, um einen Ansatzpunkt für die zur Datenreduktion angewandte psychoakustische Signalbewertung zu finden. Die Konsequenz einer MP3-Kompression bestünde theoretisch darin, für bestmöglichen Klanggenuss einer kodierten Aufnahme stets eine fest definierte Abhörlautstärke einzustellen, um die Aufnahme genau so zu hören, wie es der Encoder „meinte“. Kodiertes Audiomaterial verhält sich außerdem sehr kritisch in der Nachbearbeitung mit Dynamikprozessoren, Equalizern etc., denn viel „rauszuholen“ aus dem Audiomaterial gibt es nach der Reduktion redundanter und irrelevanter Anteile nicht mehr. Datenreduziertes Audiomaterial richtet sich an den Endverbraucher, nicht an den Musik- oder Tonproduzenten.

VORTEILE

MP3 Kompression bietet auch nach dem derzeitigem Stand der Forschung und vielen Konkurrenzverfahren einen hohen Kompressionsgrad bei guter Qualität, entsprechend der frei einstellbaren Bitrate. Man hat die Wahl zwischen diversen variablen Bitraten (VBR), mit denen man optimale Qualität je nach Anspruch des Audiomaterials erhält und konstanten Bitraten (CBR), die sich für Datenübertragungen mit festgelegter Bandbreite und Echtzeitübertragungen (Streaming) eignet. Das MP3 Format besitzt derzeit eine sehr große Verbreitung, ein umfassendes Softwareangebot und es ist plattformunabhängig. Eine MP3-Datei besitzt einen sehr nützlichen sogenannten MP3-Tag, der Metadaten über Titel, Komponist, Stil, Copyright usw. beinhalten kann.

3.5 ARBEITSWEISE DER MP3 KOMPRESSION

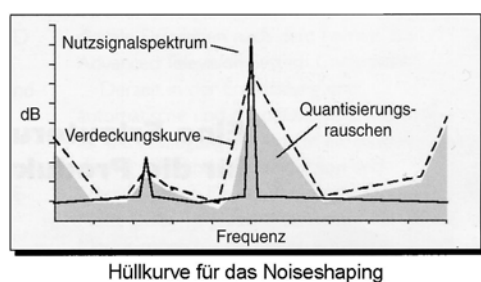
Eine Filterbank teilt das Signal vor der Transformation in 32 gleich breite Frequenzgruppen auf. Eine Form der MDCT (modifizierte diskrete Cosinus Transformation) zerlegt jedes einzelne Frequenzband in 18 noch feiner aufgelöste spektrale Komponenten; insgesamt ergeben sich somit 576 Frequenzbänder.

⁹ Qualität hängt stark vom verwendeten Encoder ab, das Qualitätsempfinden variiert je nach audiophilem Anspruch des Zuhörers

¹⁰ Artefakte sind durch die Kompression entstandenen hörbaren Verzerrungen und Störgeräusche, sie resultieren aus schlechten Codecs oder zu geringen Bitraten

¹¹ Entsprechend den sogenannten Fletcher-Munson Kurven

Die MDCT arbeitet mit Blocklängen¹² von 16 Samples oder 8 Samples bei einer Überlappung¹³ von 50% und eine folgenden Fenstergröße¹⁴ von 36 oder 12 Samples. Die lange Blockgröße erlaubt eine bessere Frequenzauflösung für stationäre Audiosignale, während die kurze Blocklänge bessere Zeitauflösung für transiente Signale bietet. Eine zusätzliche Fast Fourier Transformation mit 1024 Stützstellen erzeugt ein spektrales Abbild des Audiomaterials, welches nur als Vorlage für die Anwendung des psychoakustischen Modells zur Datenreduktion dient. Die gewonnenen Werte werden in tonale und geräuschhafte Komponenten eingeteilt, da Töne anders maskieren als Geräusche. Für jedes Subband wird die Maskierungsschwelle in Hinblick auf die benachbarten Subbänder berechnet. Nach diesen Transformationen und Maskierungsberechnungen erfolgt eine Reduktion der irrelevanten Signalinformation nach Bewertung mit einem abgespeicherten gehörpsychologischen Modell unter Ausnutzung zeitlicher und spektraler Verdeckung, wie auch der Ortbarkeit im Stereobild von Schallereignissen. Je nach Hörbarkeit und Gewichtung werden die Subbänder dann mit einer gewissen Anzahl an Bits quantisiert, und zwar gerade mit soviel Bits wie notwendig sind, um das Quantisierungsrauschen unter die Maskierungsschwelle zu setzen. Das Quantisierungsrauschen wird also aktiv geformt und unter Bereiche „verschoben“, für die das menschliche Ohr relativ unempfindlich ist oder die gerade durch lautere Schallereignisse verdeckt sind.



Diesen Prozess nennt man Noise Shaping. Bleiben Bits nach der Quantisierung übrig, so wandern sie bei der CBR-Kodierung¹⁵ in ein Bit-Reservoir, um bei schwierigeren Stellen Bit bereitzustellen. Dieses Reservoir ist allerdings nicht sonderlich groß, da es viel Speicherkapazität beim Encoder erfordern und somit Hardware teuer machen würde. Ein kleines Reservoir für kurze Audioabschnitte hat zudem den Vorteil, beim „Vorspulen“ an relativ beliebigen Stellen des Tonmaterials mit der Wiedergabe einsetzen zu können. Der

Encoder darf nur Bits entnehmen, die er vorher „reingelegt hat“ – Kredit wird nicht vergeben.

Für die Entfernung von Redundanzen zweier Kanäle (Stereo) stehen zwei Verfahren zur Datenreduktion bei MP3 zur Verfügung: der Intensitätsstereo-Modus (IS) und Mitten-/Seitensignalstereo-Modus (MS). Beide Verfahren laufen unter dem Oberbegriff Joint Stereo Coding.

- **MS Modus:** Durch Matrizierung entsteht ein Mitte/Seite-Signal aus einem Links/Rechts Signal. Wenn nun das Audiomaterial sehr mittenlastig ist, wie z.B. Sprache, können viele Bits für das S Signal eingespart werden. In der Regel werden etwa 30% der Bits für das S-Signal genutzt. MS Kodierung hat günstige Auswirkungen auf das Quantisierungsrauschen bei der Dematrizierung, denn es liegt genau in der Mitte und wird dort besser verdeckt als bei zweikanaliger, diskreter Kompression.
- **IS Modus:** Es wird ein Hauptsignal in Mono erzeugt. Informationen über die abweichende Richtungsverteilung der einzelnen Frequenzgruppen werden mitkodiert, die Phasenverteilung fällt dadurch weg. Da in den hohen Frequenzbereichen ab 2 kHz das menschliche Gehör keine Phasenunterschiede sondern nur noch Intensitätsunterschiede auswerten kann, stört dieses Verfahren höchstens in den tieferfrequenten Bereichen. Das IS Verfahren ist für geringere

¹² Die Blocklänge gibt die Stützstellenanzahl an, die bei der Transformation von der zeitlichen in die spektrale Darstellung einer Signals betrachtet wird. Die Stützstellenanzahl entspricht der Anzahl der durch die Transformation erhaltenen Spektralwerte.

¹³ Die einzelnen Blöcke überlappen sich meist um einen gewissen Prozentsatz, um die zeitliche Auflösung zu erhöhen und die Übergänge zwischen einzelnen Blöcken harmonischer zu gestalten.

¹⁴ Blöcke werden mit einer Fensterung versehen, die meist an den Enden des Blockes die Abtastwerte Richtung Nulllinie energetisch abrunden, um die scharf abgeschnittenen Kanten der abgetasteten Welle zu glätten. Die gewonnenen Spektralwerte haben durch geschickte Fensterung geringere sogenannte Nebenkeulen. Das sind Fehlerstreuungen um die Zielfrequenz herum (nicht vorhandene Nebenfrequenzen).

¹⁵ Siehe Abschnitt „Vorteile der MP3-Kodierung“ (konstante Bitrate)

Bitraten geeignet, ansonsten erweist sich MS als qualitativ besser. Geschulte Ohren können IS Stereophonie durchaus erkennen.

Nach der Entfernung der Irrelevanzen wird ein speziell für die MPEG Datenreduktion optimierter Huffman-Code angewendet.

4 MP3-PRO

4.1 Entwickler und Erscheinungsdatum

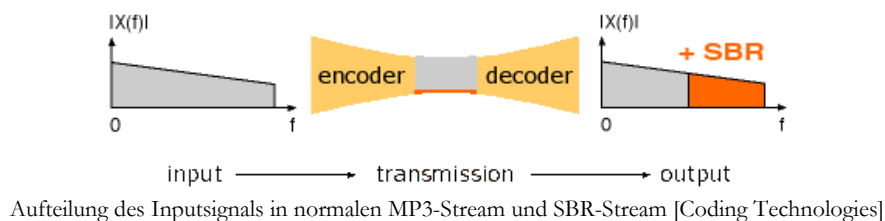


Entwickelt wurde MP3-Pro von Coding Technologies, Thomson Multimedia und dem Fraunhofer Institut. Der erste gut funktionierende Encoder von Thomson kam Mitte 2001 auf den Markt.

4.2 Über Encoder und Decoder



Encoder und Decoder basieren bei MP3-Pro stark auf den technischen Prinzipien von MP3. Der Encoder besitzt die Neuerung, das Tonmaterial vor der Kodierung in einen hochfrequenten und einen tieffrequenten Anteil aufzuteilen, um dann beide gefilterten Signalanteile verschiedenen Kodierungsalgorithmen zuzuführen. Das Niederfrequenzband wird ganz normal nach MP3-Definition komprimiert und kann aufgrund des Fehlens hoher Signalanteile mit niedriger Samplingfrequenz und demzufolge niedriger Zielbitrate (z.B. 64 Kbit/s) verarbeitet werden. Das Hochfrequenzband wird auf seine Inhalte analysiert und mit Hilfe der Spectral-Band-Replication¹⁶ (SBR) Technik als Hochfrequenzschlüssel mit geringer Bitrate in den MP3 Strom eingebettet. Der MP3-Pro Decoder liest den ersten sowie den zweiten Teil einer MP3-Pro Datei aus, rekonstruiert mit Hilfe des Schlüssels die hohen Frequenzanteile und fügt beide Bänder wie zusammen. MP3-Pro Dateien besitzen Kompatibilität zu normalen MP3 Decodern, die jedoch den eingebetteten SBR-Hochfrequenzschlüssel ignorieren und nur das bandbreitenreduzierte MP3-kodierte Signal wiedergeben.



4.3 Eigenschaften, Einsatzgebiet und Qualität von MP3-Pro

Unterstützte Bit-/Streamingraten und Qualitätsbewertung (Stand 20.10.01)

Bit-/Streamingrate	Kanäle	anvisierter Einsatzbereich
24 Kbit/s	Mono	Internetradio, Audiostreams, Vorhören von Musik vor eigentlichem Download in besserer Qualität, mobile Kommunikation (Speech Codec), etc.,
32 Kbit/s	Mono / Stereo	
40 Kbit/s	Mono / Stereo	
48 Kbit/s	Mono / Stereo	
56 Kbit/s	Stereo	

¹⁶ Mehr zu SBR unter „Arbeitsweise der MP3-Pro Kompression“

64 Kbit/s	Stereo	Pendant zu MP3 mit über 100 Kbit/s; Speicherung in portablen Playern, möglicher Standart bei Internet-Musiktauschbörsen etc.
80 Kbit/s	Stereo	
96 Kbit/s	Stereo	High Quality Audio Speicherung, Archivierung, Übertragung etc.

Die in den MP3-Stream eingebettete SBR-Komponente beseitigt den sonst bei niedrigen Bitraten und reduzierten Samplingfrequenzen (11/22 KHz) auftretenden dumpfen Klang normaler MP3 Dateien. SBR erweitert den Frequenzbereich gut hörbar auf bis zu 16 KHz. Die aufgrund der niedrigen Bitraten zusätzlich entstehende Artefaktbildung der MP3-Kodierung vermag die SBR-Technik nicht zu kompensieren. MP3-Pro eignet sich meiner Meinung nach nicht zur Erzeugung hochkomprimierter Audiostreams < 96 Kbit/s in Studio- oder CD-Qualität. Der Klangeindruck sehr hochfrequenter Anteile wie z.B. Hi-Hats und auch die Attackphase hoher Anteile sind im direkten Vergleich zwischen Originalmaterial und MP3-Pro Datei auffällig unterschiedlich. Im Bereich niedriger Bitraten und bei noch akzeptabler bis CD-naher Qualität setzt es jedoch neue Maßstäbe in Hinblick auf die Reproduktion des ursprünglichen Frequenzganges, der normalerweise stets der Samplingfrequenz unterworfen ist. MP3 Dateien klingen unter 100 KBit/s wirklich schlechter als MP3-Pro Dateien, klanglich darf man jedoch wie gesagt keine qualitativen Wunder erwarten.

4.4 Arbeitsweise der MP3-Pro Kompression

Die Funktionsweise der verwendeten MP3 Kompression ist im Abschnitt „Arbeitsweise der MP3 Kompression“ schon ausführlich beschrieben worden und soll hier nicht weiter erläutert werden. Neu an MP3-Pro ist lediglich die SBR (Spectral Band Replication) Technik. SBR überträgt nicht das komprimierte hohe Spektrum, sondern rekonstruiert, oder besser synthetisiert die hohen Anteile gemäß der Analyse des tieffrequenten Audiospektrums aus dem unterliegenden Algorithmus, bei MP3-Pro ist das MP3. Um eine akkurate Rekonstruktion des Spektrums zu gewährleisten, werden nur SBR-Steuerdaten mit einer sehr geringen Datenrate in den kodierten Audiodatenstrom eingebettet. Laut Coding Technologies gelingt diese Rekonstruktion für harmonische wie auch geräuschhafte Anteile sehr gut. Im Durchschnitt erreiche man mit SBR im Vergleich mit gängigen wahrnehmungsbasierten Kompressionsalgorithmen wie MP3 eine um 30% bessere Effizienz.

5 MPEG-2

5.1 NEUERUNGEN UND UNTERSCHIEDE ZU MPEG-1

MPEG-2 ist genau wie MPEG-1 für die Kodierung von Audiosignalen hoher Qualität entwickelt worden. Es gibt jedoch Neuerungen im Bereich geringer Bitraten. MPEG-1 hat noch strenge eingeschränkte Spezifikationen, die durch MPEG-2 umfassend erweitert werden. Es werden zusätzliche Samplingfrequenzen für die Kompression erlaubt: 8, 11, 16, 22 und 24 KHz. Zusammen mit der Definition neuer Bitraten gestaltet sich die Kodierung von Signalen aus digitalen Quellen mit bereits eingengerter Bandbreite wie ISDN oder G.722 mit MPEG-2 weitaus unproblematischer. Ein großer Fortschritt von MPEG-2 ist mit der Kodierung von bis zu fünf Kanälen plus einem Tieftonkanal bis 100 Hz erreicht worden, weshalb die MPEG-Technologie auch auf DVD's verwendet werden kann und sich endlich für 5.1 Surround und multilinguale Anwendungen öffnet. Auf DVDs ist im Tonbereich jedoch mittlerweile Dolby Digital Standard, es erscheinen keine Filme mehr mit MPEG-2. Tatsächlich wurde MPEG-2 bislang nur auf einigen wenigen DVDs eingesetzt. Die variable Audio-Datenrate (bis zu 640 kbit/s) mit der ebenfalls variablen Datenrate für das Bild zu kombinieren ist aufwendiger und auch kostenintensiver. Daher wurde der Ton auf den bisherigen MPEG-MC DVDs wie bei Dolby Digital mit konstanten 384 kbit/s kodiert. Mit MPEG-2 wird ein variables Bit-pro-Sample Verhältnis (VBR) je nach Bedarf des Audiomaterials definiert, was ursprünglich bei MP3 nicht festgelegt worden war. Zu den bei MPEG-1 definierten Methoden der Audiokompression sind einige zusätzliche hinzugekommen:

- Intensity Stereo Coding (ISC) ¹⁷
- Phantom Coding of Center (PCC) – der Mittenkanal lässt sich durch Matrizierung auf den linken und rechten Kanal verteilen
- Dynamic Transmission Channel Switching
- Dynamic Cross Talk
- Adaptive Multi-Channel Prediction – die adaptive Vorhersage des Signalverlaufes (nützlich für die Anpassung der Kompressionsparameter) kann separat für verschiedene Kanäle verlaufen, da z.B. einige Kanäle für Sprache reserviert sein könnten

Neu hinzugekommen ist auch die Definition des Advanced Audio Coding (AAC)

6 AAC (ADVANCED AUDIO CODING)

6.1 ENTWICKLER UND ERSCHEINUNGSDATUM

AAC wurde hauptsächlich vom Fraunhofer Institut, Dolby, Sony und AT&T entwickelt und ist fester Bestandteil des MPEG-2 (AAC Definitionen seit ca. 1997) und MPEG-4 Standards (definiert 1999).

6.2 EINSATZGEBIET UND QUALITÄT

AAC ist für Mehrkanalanwendungen entwickelt worden und bietet großen Freiraum in Hinblick auf Kanäle, Kompression und Qualität. Durch die Mehrkanalfähigkeit wird AAC auch interessant für die Übertragung mehrsprachiger Inhalte und für DAB¹⁸. Als MPEG-2 AAC ist es schon auf DVDs als 5.1-Kanalton enthalten. Spezifiziert wurde AAC bereits für den japanischen digitalen Fernsehstandard, für den digitalen Rundfunk in den USA und Europa und teilweise schon für kommerzielle Musik-Downloads aus dem Internet. Angekündigt sind portable Musik-Player mit AAC Unterstützung.

Die Effizienz wurde gegenüber MPEG-1 Layer 3 durch Modifikationen im Kompressionscodec und das Einbringen neuer Forschungserkenntnisse leicht verbessert und es wird demzufolge eine etwas höhere Datenkompression bei vergleichbar guter Qualität erreicht. Mit einem hochqualitativen Encoder erreicht man MP3-Qualität bei einer ca. 30% geringeren Bitrate. Aber solche Encoder sind noch relativ selten, billige Freeware Encoder erzeugen teilweise schlechtere Qualität verglichen mit herkömmlicher MP3-Kodierung. Ursprünglich sollte AAC relativ zügig MP3 ablösen, aber die Realität ließ MP3 zum unangefochtenen breiten Internetstandard werden. AAC ist nicht kompatibel mit MP3.

6.3 EIGENSCHAFTEN DER AAC KOMPRESSION

Die neuesten AAC Spezifikationen und geplanten Encoder kodieren bis zu 48 Breitband-Kanäle, 16 Kanäle für tiefe Frequenzen mit eingeschränkter Bandbreite und 15 eingebettete Datenstreams. Erlaubt sind diverse Samplingfrequenzen bis 96 KHz und diverse Bitraten bis maximal 512 Kbit/s pro Kanal bei 96 KHz. Es existieren drei unterschiedliche Komplexitätsstufen von AAC. Am verbreitetsten ist derzeit die Stufe mit der geringsten Komplexität (LC – low complexity)

Vergleich AAC mit MP3 von PsyTEL

	PsyTEL AAC	MPEG Layer III (MP3)
Maximum Channels	48 Main + 16 LFE Channels, 64 channels total	2 (Stereo)
Max. bit-rate per channel	512 kbits/s (at 96 kHz)	160 kbits/s

¹⁷ mehr dazu im Abschnitt „Arbeitsweisen der MP3 Kompression“

¹⁸ Digital Audio Broadcasting (Digitaler Rundfunk)

Sampling Rates Supported	8, 11.025, 12, 16, 22.05, 24, 32, 44.1, 48, 64, 88.1 and 96 kHz	32, 44.1, 48 kHz (MPEG-1) and 16, 22.05 and 24 kHz (MPEG-2)
Frame Prediction	Yes (Long Term Predictor)	No
Stereo @128 kbits/s	Very High Quality	Medium Quality
VBR Support	Native	Optional
Joint Stereo Coding	Fine (per frequency band)	Rough (per frame)
Additional Tools	Yes (TNS coding and PNS)	No
Spectral Resolution @48 kHz (smaller is better)	23.4 Hz	41.6 Hz
Filterbank	MDCT	Hybrid Filterbank
Pre-Echo Control	Good	Weak
Impulse Response @48 kHz	5.3 ms	18.6 ms (pre-echo)
Further Development	Yes (MPEG-4 V2)	No

6.4 ARBEITSWEISE DER AAC KOMPRESSION

Die Arbeitsweise des Encoders ist der Kodierung von MP3 Dateien sehr ähnlich. Ich werde mich hauptsächlich auf die Neuerungen des AAC Encoders beschränken. Das ankommende Signal wird zuerst mit einem sogenannten Polyphase Quadrature Filter (PQF) in vier Subbänder geteilt. Für die Umwandlung der Zeit- in die Frequenzdarstellung des Signals wird eine modifizierte diskrete Kosinus Transformation (MDCT) mit einer Stützstellenzahl von 2048 Samples (bei stationären Klängen) oder auch 256 Samples (bei transienten Elementen) angewendet. Die Überlappung beträgt 50%, die Fensterung erfolgt je nach Material entweder sinusförmig oder nach dem Kaiser-Bessel-Fenster. Das nach der MDCT entstandene spektrale Abbild des Audiomaterials wird in einzelne Frequenzgruppen aufgeteilt, die ungefähr der Bark-Scala und der Aufteilung der Basiliar-Membran des menschlichen Gehörs entsprechen. Den Frequenzgruppen werden nach ihrer Bewertung mit dem psychoakustischem Modell und Maskierungsschwellenberechnung je nach Wichtigkeit und Hörbarkeit quantisiert, d.h. die Bitverteilung konzentriert sich auf die wichtigen Frequenzgruppen. Hohe Frequenzen brauchen in der Regel mehr Bits für die Kodierung als tiefe. Wie bei der MP3-Kompression sind zur Reduktion der Gemeinsamkeiten mehrerer Kanäle MS-Stereo und IS-Stereo vorgesehen.

Aus zwei vorhergehenden Audioblöcken errechnet ein spezieller Vorhersage-Mechanismus (ab MPEG-2) für jedes Frequenzband die wahrscheinliche Weiterentwicklung, was bei stationären Signalen, die gut vorhersagbar sind, die Effizienz erhöht. Nur die Differenzen müssen übertragen werden und werden durch Huffman Code komprimiert.

Neu bei AAC ist das Temporal Noise Shaping (TNS). Das normale Noise Shaping wurde bereits im Kapitel „Arbeitsweise der MP3 Kompression“ beschrieben und hat den Nachteil, dass das geformte Quantisierungsrauschen über den Zeitraum der Audioblocklänge oder Stützstellenzahl konstant ist und auch nur in diesen zeitlichen Abständen geändert werden kann. Diese Abstände können für sich drastisch ändernde Signale bereits zu lang sein und das Quantisierungsrauschen könnte an gewissen Stellen hörbar werden. Hier setzt TNS ein, indem es durch vorausschauende Spektralanalyse und Wahrscheinlichkeitsvorhersagen das Quantisierungsrauschen schneller als die Audioblocklänge anpasst.

Die unter MPEG-4 AAC neu eingeführte Long Time Prediction entfernt effektiv Redundanzen von aufeinanderfolgenden kodierten Audiosegmenten, besonders bei Segmenten mit einer eindeutigen Tonhöhe und wenig klanglicher Veränderung über längere Zeiträume hinweg.

Perceptual Noise Substitution (PNS), eingeführt durch MPEG-4 AAC, basiert auf der Erkenntnis, dass Rauschanteile im Signal klanglich meist sehr ähnlich sind und die Feinstruktur des Rauschens bei der

Wahrnehmung kaum eine Rolle spielt. Aus diesem Grund werden Rauschanteile nicht mehr kodiert und komplett übertragen. Es wird nur noch die Information übertragen, dass ein Rauschanteil mit einer gewissen Energie im Signal vorhanden ist. Der Decoder erzeugt dann gezielt in bestimmten Frequenzbereichen ein Zufallsrauschen, das lediglich von Informationen über den jeweiligen Energieinhalt des zu synthetisierenden Rauschens gesteuert wird. Diese Informationen werden mit einer sehr geringen Datenrate in den Bitstream eingebettet. PNS ist von der Arbeitsweise her stark mit SBR (Spectral Band Replication bei MP3-Pro) verwandt.

7 TWIN-VQ

(TRANSFORM-DOMAIN WEIGHTED INTERLEAVE VECTOR QUANTIZATION)

7.1 ENTWICKLER



TwinVQ wurde hauptsächlich von der Japanischen Telekom NTT entwickelt, wobei z.B. Yamaha an der Encoder/Decoder Entwicklung teilnahm. TwinVQ ist Bestandteil von MPEG-4. Es war einer der ersten ernstzunehmenden Konkurrenten von MP3 bei niedrigen Bitraten, wurde aber lange Zeit nicht weiterentwickelt und hatte einen zu geringen Bekanntheitsgrad. Umgangssprachlich hat sich für TwinVQ der Begriff VQF eingebürgert, was auch der Dateiendung entspricht.

7.2 EINSATZGEBIET UND QUALITÄT

TwinVQ hat seine Stärken bei den niedrigen Bitraten bis 80 KBit/s. Es erreicht keine Studio- oder CD-Qualität, übertrifft jedoch bei Sprachübertragung und kleinen Bitraten die MP3-Kodierung. Die Artefakte von TwinVQ sind anders. Es klingt nicht so, als würde Rauschen dazukommen, sondern als würden gewisse Anteile im Tonsignal verloren gehen. Das Signal klingt sozusagen „weichgespült“. Artefakte entstehen besonders bei Attackphasen im Audiomaterial. Bei Sprachübertragung geht dabei gewöhnlich u.a. der Straßenlärm-Teil des Signals verloren, weshalb TwinVQ für Sprachübertragung und für sehr kleine Bitraten gut geeignet ist. Die Qualität von 40 KBit/s TwinVQ soll mit der Qualität von 64 KBit/s MP3 vergleichbar sein.

7.3 EIGENSCHAFTEN DER TWIN-VQ KOMPRESSION

TwinVQ hat einen anderen Ansatz als beispielsweise MPEG. Aber auch TwinVQ muss irrelevante Daten wegfällen lassen, um solch niedrige Bitraten zu erreichen. Beide Verfahren wandeln das Signal über eine Transformation in die Kurzzeit-Frequenz-Darstellung, nutzen die Bit-Reservoir-Technik und auch Vorhersagefunktionen. TwinVQ kodiert die Samples dann aber nicht wie bei MP3, sondern betrachtet jeweils Muster aus Bits und Samples in Form von Vektoren. Der Encoder hat eine Tabelle von Standard-Mustern, mit denen die Eingabevektoren verglichen werden. Der Index des der Eingabe am nächsten kommenden Vektors wird dann übertragen. Die Bitraten reichen von 6 bis ungefähr 96 KBit/s. Der Rechenaufwand ist höher als bei der MPEG-Kodierung.

Hinweis: Mit einer neueren Version der weitverbreiteten Brennsoftware Nero der Ahead-Software GmbH kann man TwinVQ Dateien erstellen.

8 MPEG-4

8.1 ENTWICKLER UND ERSCHEINUNGSDATUM

Die MPEG-Organisation verabschiedete bereits im Oktober 1998 die erste MPEG-4 Version, die erweiterte zweite Version wurde im Dezember 1999 standardisiert.

8.2 ZUM VERSTÄNDNIS VON MPEG-4 UND DESSEN EINSATZGEBIET

Mit MPEG-4 sollen die bisher getrennten Bereiche der Sprach- und Musikkodierung sowie der Computermusik zusammengeführt werden. Die in MPEG-4 definierten Verfahren zur Audiodatenreduktion wurden nicht mehr nur für eine Art Signale (Musik oder Sprache) oder gar nur eine Bitrate optimiert, sondern als universelles und flexibles System konzipiert. MPEG-4 beschreibt kein einzelnes hocheffizientes Kompressionsverfahren, sondern ein komplettes Sortiment zahlreicher „objekt-orientierter“ Algorithmen für diverse Einsatzzwecke z.B. Sprachkodierung, Sprachsynthese, Musikkodierung, Musiksynthese (a la Midi). Während frühere Verfahren sich vor allem mit der Kompression der Audiodaten beschäftigten, soll MPEG-4 die universelle „Sprache“ für Broadcasting-, Film- und Multimedia-Anwendungen werden. Die Neuerungen in der MPEG-4 Spezifizierung beziehen sich vor allem auf Audio mit sehr geringer Bitrate wie auch synthetisierte Sprache und Musik. Das anvisierte Ziel dieser Standardisierung ist die Vorstellung, eine audiovisuelle Szene in ihre einzelnen Komponenten, die sogenannten Objekte zu zerlegen, die dann mit der für sie am effektivsten funktionierenden Kompression versehen werden. Beispiel: In einer Filmsequenz liegt der Dialog einkanalig in drei Sprachen stark komprimiert vor, die Hintergrundmusik dagegen fünfkanalig in hoher Qualität etc. Andere Anwendungsgebiete der MPEG-4 Kompressionsverfahren sind u.a. die Übertragung von Sprachsignalen mit Bitraten < 10 KBit/s, die Übertragung von Audio über das Internet und der Digitale Hörfunk über Mittel- und Langwellenkanäle.

8.3 NEUE EIGENSCHAFTEN VON MPEG-4

- Inhaltsbezogene Interaktivität: Änderungen der Darstellung sind vom Benutzer auf der Dekodier-Seite möglich. Die Auswahl von Kanälen, die Änderung der Tonhöhe oder die Wiedergabegeschwindigkeit sind somit interaktiv veränderbar. Ein Beispiel für diese Interaktivität ist die Darstellung von Orchestermusik nach Einzelinstrumenten oder Instrumentengruppen.
- Universeller Zugriff: Unter diesem Begriff verbirgt sich die Zugriffsmöglichkeit auf MPEG-4-Daten über verschiedene Netzwerktypen (LAN, ISDN, Mobilfunknetze). Für den Fall einer Übertragung über Mobilkanäle ist korrekte Dekodierung auch bei Übertragungsfehlern notwendig, was bessere Fehlerbehandlungsverfahren erforderlich machte. Damit die verschiedenen Netzwerktypen unterstützt werden, wird die früher entwickelte Methode der Skalierbarkeit verwendet.
- Skalierbarkeit: Schon MPEG-1 und 2 enthalten diese Eigenschaft, damit die Verfahren in einem weiten Bitratenbereich funktionsfähig sind. Darüber hinaus wurde in MPEG-4 das sogenannte „embedded coding“ realisiert, bei dem ein Teil des Bitstroms genügt, um ein Signal geringerer Qualität dekodieren zu können. Damit kann beispielsweise der Zugriff auf Multimedia-Datenbanken über Netzwerke mit unterschiedlicher Bandbreite erfolgen, ohne dass die Daten mehrfach in verschiedenen Datenraten gespeichert vorliegen müssen. Aus Sicht der Wissenschaft bestand die Herausforderung dabei, diese Skalierbarkeit ohne große Verluste in der Kodiereffizienz zu realisieren. Skalierbarkeit für MPEG-4 ist wie folgt definiert: Ein Bitstrom, bei dem eine Teilmenge der Bits ausreicht, um ein Audiosignal dekodieren bzw. erzeugen zu können, wird skalierbarer Bitstrom genannt.
- Unterstützung von Audio-Objekten begrenzter Länge: Eine gewünschte Eigenschaft war außerdem, Audiodaten in Objekte einzuteilen, die dann einzeln manipulierbar wiedergegeben werden können. Damit können zum Beispiel Orchesteraufnahmen neu abgemischt oder einzelne Instrumente ganz weggelassen werden.
- Einstellbarkeit der Gesamtverzögerung: Wie bei allen Kodierverfahren bedingt die Erhöhung der Kodiereffizienz auch eine Erhöhung der Rechenzeit. Für alle Anwendungen, bei denen eine hohe Rechenzeit ungünstig ist (direkte Zweiweg-Kommunikation), wurden spezielle Modi mit geringer Systembelastung entworfen.

8.4 QUALITÄT UND EFFIZIENZ VON MPEG-4

Hauptanliegen von MPEG-4 ist es, trotz zusätzlicher Funktionen bei jeder Signalart einen Fortschritt in der Kodiereffizienz gegenüber bisher eingesetzten Standardverfahren zu erreichen. Da es nicht möglich ist, alle Signale und Qualitätsstufen mit demselben Algorithmus effektiv zu kodieren, wurden verschiedene Algorithmen implementiert, die frei konfigurierbar sind. Je nach Anwendungsfall können die verschiedenen Algorithmen auch kombiniert werden. Die folgende Liste zeigt einen Überblick der verschiedenen Signalklassen, welche von MPEG-4 gespeichert werden können:

- Musik in HiFi-Qualität („High Fidelity“), Studioqualität
- Musik in CD-Qualität
- Musik und Sprache in mittlerer Qualität
- Breitbandsprache
- Sprache in Telefonqualität
- Sprache bei sehr niedrigen Bitraten
- Synthetische (künstliche) Sprache
- Synthetische Musik

MPEG-4 stellt für die Entwickler somit ein „Do-it-all“-System für die Zukunft dar, wobei CD-Qualität bei einer Kompressionsrate von 1:20 erreicht werden soll. Dieser neue, offene Multimedia-Standard wurde so konzipiert, dass zukünftige Ideen im Bereich Audio problemlos implementiert werden können. So werden sich in den geplanten Versionen 3 und 4 auch Mechanismen zum Schutz der Urheberrechte befinden. Durch die Unterstützung von Audio-Objekten können wie bereits erwähnt einzelne Instrumente separiert werden, es lassen sich jedoch auch multilinguale Übertragungen in verschiedenen Sprachen übermitteln.

8.5 DIE EINZELNEN VERFAHREN DER MPEG-4 AUDIOKOMPRESSION

MPEG-4 ist im Bereich niedriger Bitraten konzipiert für Streams von 64 KBit/s bis herunter zu 0,2 KBit/s. Gerade in diesem Bereich existiert eine Vielzahl an Kompressionscodecs. Hier nun die verschiedenen in MPEG-4 definierten Verfahren in einer kurzen Übersicht:

- AAC¹⁹, auch General Audio Coding genannt
- Parametrische Kodierer mit Vocoder-ähnlicher Arbeitsweise
- CELP Kodierer für Sprach-Signale > 4 kBits/s (Code Excited Linear Predictive Coding - Vektorquantisierung von Prädiktionsfehlersignalen), frei skalierbare Bandbreite und Bitrate, fein einstellbar je nach Verbindungsqualität zwischen Sender und Empfänger
- HVXC (Harmonic Vector eXcitation Coding) für Sprach Signale von 2 – 4 kBit/s
- TwinVQ²⁰ (Transform-Domain weighted interleave Vector Quantization) – arbeitet effizienter als AAC, hat aber erst bei Bitraten unter 16 Kbit/s bessere Qualität als AAC
- Text to Speech Tools (Sprachsynthese nach Textvorlage)
- Talking Heads (Gesichtsanimationen)
- Structured Audio (Beschreibungssprache für die Synthese von Musik bei geringsten Bitraten!)

Die in der Auflistung erwähnte **CELP-Sprachkompression** arbeitet nach Prinzipien, die auch viele andere Sprach-Codecs zur Datenreduktion verwenden. Die Kompression beruht auf gründlicher Analyse des eingehenden Sprachsignals und dem Versuch, durch komplexe Filterung eines synthetisierten oder zumindest stark komprimierten Grundsignals (source) die Stimme in akzeptabler Qualität zu reproduzieren. Bei einer Hybrid-Sprach-Komprimierung wie CELP wird ein synthetisch erzeugtes Sprachsignal in Anlehnung an ein Filtermodell des menschlichen Stimmtraktes erzeugt, wobei der Klang

¹⁹ mehr dazu im Abschnitt „AAC (Advanced Audio Coding)“

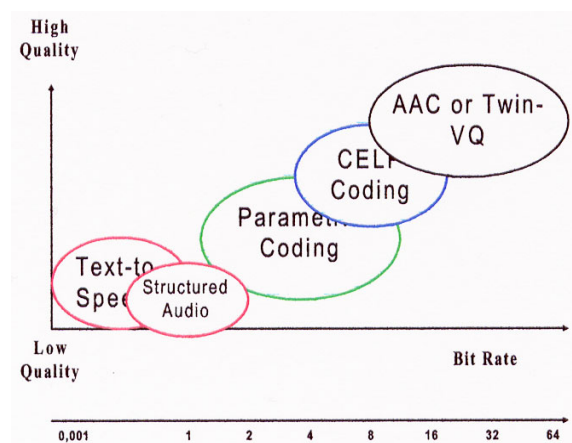
²⁰ mehr zu TwinVQ im Abschnitt „TwinVQ (Transform-Domain weighted interleave Vector Quantization)“

des eingehenden Signals permanent mit dem synthetisierten Klang durch diverse Parameter abgeglichen wird. Zwischen Encoder und Decoder müssen auf diese Weise nur Parameterwerte übertragen werden, ähnlich wie bei Midi.

Structured Audio besteht im Wesentlichen aus zwei verschiedenen Codecs: Text to Speech und Score driven Synthesis.

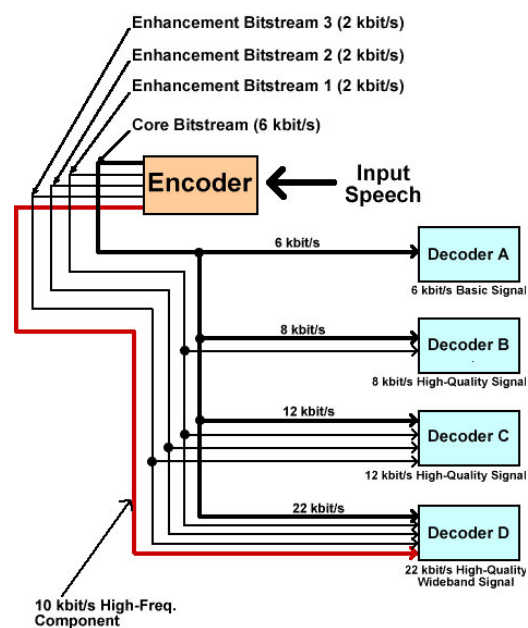
Text to Speech arbeitet mit einer Bitrate von 200 bit/s bis 1,2 kBit/s. Als Vorlage dienen reine Texte oder Texte mit Parametervorgaben wie Tonhöhenverlauf, Phonem/Silben-Dauer usw. Mit diesen Vorgaben wird eine synthetisch erzeugte Stimme gesteuert, die dann bei kleinstem Bitstrom Texte vorlesen kann, ohne dass kodierte Sprachaufnahmen übertragen werden müssen. Weiterhin kann Text to Speech mit Gesichtsanimationsprogrammen (Talking Heads) synchronisiert werden. Text to Speech kann zusätzliche Informationen an andere Komponenten schicken, die synchron zum Text gesteuert werden. MPEG-4 definiert nur das Text To Speech Interface, aber keinen vorgeschriebenen Sprachsynthesizer.

Bei der **Score Driven Synthesis**, einer synthetisch erzeugten Musik, wird der Decoder von einer speziellen Synthese-Sprache, genannt SAOL – Structured Audio Orchestra Language, gesteuert, welche das „Orchester“ mit seinen benutzten „Instrumenten“ festlegt. Die Instrumente selbst werden teilweise heruntergeladen und sind nicht unbedingt im Decoder als Presets fest verankert. Natürlich wird es auch je nach verwendetem Decoder Presets geben. MPEG-4 standardisiert keine einzelne Synthesemethode, sondern nur die Art und Weise wie man verschiedene Syntheseformen wie Wavetable-, FM-, Additive-, Physical-Modeling-, Granularsynthese etc. einbinden kann. Die Sequenzen und Noten werden in einer Sprache, genannt SASL (Structured Audio Score language), übertragen, über die auch feinere Soundmodifikationen möglich sind. Eine einfachere Steuerung über Midi wird auch möglich sein. Für einfachere Anwendungen, die ohne solche speziellen Syntheseformen auskommen, wurde das „wavetable blank format“ standardisiert, bei dem die Samples für die Wavetablesynthese mit Filter, Hall, Choruseinstellungen im Bitstream enthalten sind.



MPEG-4, Einordnung der Verfahren nach Qualität und Bitrate

Grafiken Quelle: [3]



Skalierbarer Audiostream je nach verfügbarer Bandbreite

8.6 AUDIOPROFILE UND KOMPLEXITÄTSSTUFEN (LEVEL) BEI MPEG-4

Für MPEG-4 Audio wurde eine geringe Anzahl an Profilen definiert, die für bestimmte Einsatzzwecke die Anwendung von speziell abgestimmten Algorithmen vorsehen. Selbst in einfachen Profilen befinden sich diverse Typen von Audioobjekten mit verschiedenen Kompressionen. Hier einige Beispiele:

- **Speech Coding Profile:** CELP- und HVXC-Objekte, Text-to-Speech Interface
- **Scalable Profile:** AAC-Objekte geringer Komplexität (Low Complexity, LTP), TwinVQ- und Speech-Objekte, weitere Tools für skalierbares Audio
- **Main Profile:** alle MPEG-4 Kodieralgorithmen für sowohl natürliches wie auch synthetisches Audiomaterial

Das skalierbare Profil wird wahrscheinlich von vielen Anwendungen benutzt werden. Es beinhaltet verschiedene vordefinierte Level. Diese Level stehen für verschiedene Komplexitätsstufen des kodierten Materials, um wenigstens ein bisschen standardisierte Ordnung in das verwirrende Gemisch aus Kanälen, Samplingraten und Bitraten zu bringen:

- Level 1: ein Mono-Objekt, bis 24 kHz Samplingfrequenz
- Level 2: ein Stereo oder zwei Mono-Objekte, bis 24 kHz Samplingfrequenz
- Level 3: ein Stereo oder zwei Mono-Objekte, bis 48 kHz Samplingfrequenz
- Level 4: ein 5.1 Kanal-Objekt oder ähnliches bis 48 kHz Samplingfrequenz

Beim skalierbaren Profil mit seinen verschiedenen Leveln bieten sich für Sprach- und Musikinhalte folgende Kombinationen von Kompressionsalgorithmen an:

- schmalbandiges CELP base layer plus AAC (gute Sprachqualität bei geringen Bitraten; bei geringer Übertragungskapazität reicht das „base layer“ aus, wird zusätzlich noch AAC übertragen, steigt die Qualität und Bandbreite)
- TwinVQ base layer plus AAC (für musikalische Inhalte bei geringer Bitrate in erträglicher Qualität am geeignetsten)
- ausschließlich AAC layers (wenn das Base Layer über 16 kBit/s hinausgehen kann, sind reine AAC Layer die qualitativ beste Wahl)

Coding tool	channels	Total bit rate	Typical subjective quality
AAC	5	320 kb/s	4.6
1995 Backward Compatible MPEG-2 Layer II	5	640 kb/s	4.6
AAC	2	128 kb/s	4.8
AAC	2	96 kb/s	4.4
MPEG-2 Layer II	2	192 kb/s	4.3
MPEG-2 Layer III	2	128 kb/s	4.1
AAC	1	24 kb/s	4.2
Scalable: CELP base and AAC enhancement	1	6 kb/s base, 18 kb/s enh.	3.7
Scalable: Twin VQ base and AAC enhancement	1	6 kb/s base, 18 kb/s enh.	3.6
AAC	1	18 kb/s	3.2
G.723	1	6.3 kb/s	2.8
Wideband CELP	1	18.2 kb/s	2.3
BSAC	2	96 kb/s	4.4
BSAC	2	80 kb/s	3.7
BSAC	2	64 kb/s	3.0
AAC – LD (20 ms one-way delay)	1	64 kb/s	4.4
G.722	1	32 kb/s	4.2
AAC – LD (30 ms one-way delay)	1	32 kb/s	3.4
Narrowband CELP	1	6 kb/s	2.5
Twin VQ	1	6 kb/s	1.8
HILN	1	16 kb/s	2.8
HILN	1	6 kb/s	1.8

Quelle: INTERNATIONAL ORGANISATION FOR STANDARDISATION, ISO/IEC JTC1/SC29/WG11, CODING OF MOVING PICTURES AND AUDIO

9 ATRAC (ADAPTIVE TRANSFORM ACOUSTIC CODING)

9.1 ENTWICKLER UND ERSCHEINUNGSDATUM

Das ATRAC-Verfahren wurde von Sony entwickelt. Seit der ersten ATRAC-Version im Jahr 1991 sind immer wieder Verbesserungen am Kompressionsverfahren vorgenommen worden, was zu neuen ATRAC-Versionen mit besserer Qualität führte. Die neueste ATRAC-Version ist ATRAC3.

9.2 EINSATZGEBIET UND QUALITÄT

ATRAC-Audiokompression ist die Schlüsseltechnologie der ursprünglich von Sony entwickelten Minidisc-Systeme. Minidiscs besitzen eine Kapazität von ungefähr 150 Mbyte und können durch das ATRAC-Verfahren bis zu 75 Minuten Musik in CD-Qualität speichern. ATRAC liefert gute Qualität bei einem nicht mehr ganz zeitgemäßen Kompressionsverhältnis. Seit einiger Zeit existieren auch Portable Musik-Player mit Memory-Stick-Karten, die neben MP3 und meist WMA²¹ auch die neue ATRAC3-Kompression bei geringen Bitraten beherrschen. ATRAC3 besitzt im Longplay-Modus LP 2 ähnlich gute Qualität wie ATRAC1, im Lonplay-Modus LP 4 jedoch hörbare Artefakte.

Übersicht ATRAC-Versionen und deren Qualität:

- **1991 ATRAC1 1.0:** - Frequenzgang bis 15 kHz
- starkes Quantisierungsrauschen
- interne Wortbreite 16 / 18 Bit
- **1992 ATRAC1 2.0:** - Frequenzgang bis 18 kHz
- geringes Quantisierungsrauschen
- **1995 ATRAC1 3.0:** - interne Wortbreite 18 Bit / bei ATRAC 3.5: 20 Bit interne Wortbreite
- Frequenzgang bis 20 kHz
- neue Block Floating Technologie reduziert Rauschen wirkungsvoll
- nur noch marginaler Klangunterschied zur CD
- **1996 ATRAC1 4.0:** - interne Wortbreite 24 Bit
- Frequenzgang > 20 kHz ab ATRAC Version 4.5
- **2000 ATRAC3 1.0:** - arbeitet mit halber Bitrate im Vergleich zu ATRAC1 bei etwa gleicher Qualität, zwei Longplay-Modi, nicht abspielbar auf alten ATRAC Playern

Bei ATRAC schleichen sich bei einer mehrfachen Kompression und Dekompression Artefakte ein. MPEG hingegen hat sich in Hörtests als weitgehend resistent erwiesen.

Übersicht über die Qualität der ATRAC-Versionen nach mehreren Kopie-Generationen

ATRAC 1 zu ATRAC 1	Nach 5 Generationen unakzeptabel, nach 20 Generationen grauenhaft
ATRAC 2 zu ATRAC 2	Nach 5 Generationen kein hörbarer Unterschied, nach 20 Generationen winzige Verzerrungen
ATRAC 3 zu ATRAC 3	nicht viel besser als bei ATRAC 2.
ATRAC 3.5 zu ATRAC 3.5	Im Bezug auf Rauschen etwas besser als ATRAC 3

Quelle: http://audioheat.virtualave.net/tech/m_fa.html#2

²¹ Windows Media Audio – Streaming-Media Format von Microsoft, unterstützt das Digital-Rights-Management-System (DRM)

9.3 EIGENSCHAFTEN DER ATRAC KOMPRESSION

Die ATRAC1-Kompression ist eine Kompression mit variabler Bitrate und braucht deshalb beim Abspielen von einer Minidisc stets einen Zwischenspeicher. Die Bitrate liegt bei Stereo etwa bei 292 KBit/s. Das Kompressionsverhältnis beträgt etwa 1:5, was zum Erscheinungsdatum der Minidisk schon erstaunlich war, heute aber nicht mehr ganz zeitgemäß ist.

Die ATRAC3-Kompression unterstützt verschiedene Longplay Modi, die effizienter und deshalb bei niedrigeren Bitraten als ATRAC1 arbeiten. LP 2 (Longplay 2) speichert bei einer Bitrate von 132 Kbits/s ca. 160 Minuten in Stereo, LP 4 speichert bei 66 KBit/s ca. 320 Minuten in „Joint-Stereo“ auf einer Minidisc.

9.4 ARBEITSWEISE DER ATRAC KOMPRESSION

ATRAC1

Am Beginn des Kompressionsvorganges wird das Spektrum des eingehenden Signals durch ein Quadrature Mirror Filter (QMF) in drei Subbänder aufgeteilt: 0...5,5 kHz; 5,5...11 kHz; 11...22 kHz. Dann werden die Subbänder mittels MDCT in Spektralwerte umgewandelt. Die Transformation von anhaltenden, relativ stationären Signalen geschieht in Audioblöcken von 11,6 ms (entspricht 512 Samples), während die Transformation von kurzen Impulsen in Blöcken von 2,9 ms (128 Samples), im hohen Subband sogar 1,45 ms (64 Samples) erfolgt. Bei längeren Zeitblöcken (11,6 ms) kann man nach der Transformation engere Frequenzbänder für die Analyse verwenden, das bedeutet höhere Qualität. Die kürzeren Audioblöcke bieten eine bessere Zeitauflösung für transiente Signale, gerade auch im Hinblick auf das Noiseshaping. Nach der Transformation wird das entstandene Spektrum zu 52 Frequenzbändern zusammengefasst. Die Einteilung der Bandbreite der 52 Frequenzbänder erfolgt gehörangepasst in Anlehnung an die Frequenzgruppen. Die variable Quantisierung der einzelnen Frequenzgruppen richtet sich nach ihren jeweiligen Maskierungs- und Pegeleigenschaften, d.h. wie bei der MPEG-Kompression entscheidet ein psychoakustisches Modell, welche Frequenzgruppen wichtig sind und mit wie vielen Bits sie quantisiert werden müssen, um das Quantisierungsrauschen unter die Maskierungsschwelle zu drücken. Frequenzen, die im empfindlichen Bereich menschlichen Hörens liegen, werden automatisch mit einer höheren Auflösung quantisiert als unempfindlichere.

Eine spezielle Eigenschaft der ATRAC-Kompression ist seit ATRAC-3 der Block-Floating Algorithmus. Dabei entfernt ein „Bit-Verschiebe“-Algorithmus ungenutzte höherwertige Bits, um sie an anderen Stellen einzusetzen, z.B. bei der Darstellung eines 5 Bit Wertes in einem 8 Bit Wort. Das Wort wird entsprechend auf 5 Bit gekürzt, ohne dass Klangeinbußen hörbar werden und die eingesparten Bits werden an anderen Stellen verwendet. Beim Abspielen einer Minidisc wandelt der ATRAC Decoder das reduzierte Signal, die Frequenzwerte, durch invertierte MDCT wieder in Zeitwerte um.

BESONDERHEITEN BEI DER NEUEN ATRAC3-KOMPRESSION

Das eingehende Signal wird in vier statt in drei Subbänder aufgeteilt. Tonale und nicht-tonale Frequenzkomponenten werden unterschiedlich kodiert, da bei nicht-tonalen Bestandteilen das Quantisierungsrauschen meist lauter sein darf und Geräusche andere Maskierungsschwellen besitzen als klare Töne. Die MDCT benutzt im Vergleich zu ATRAC längere Audioblöcke, um die doppelte Frequenzauflösung zu erhalten. Gain Adjustment wurde eingeführt, um bei percussivem Material mit intensiven Attackphasen, die sogenannten Pre-Echo Artefakte zu reduzieren. Pre-Echos entstehen, wenn das Quantisierungsrauschen bei leisen Passagen kurz vor einem starken Attack des Signals aufgrund der Audioblockbildung hörbar wird.

10 AC3 (DOLBY DIGITAL)

10.1 ENTWICKLER UND ERSCHEINUNGSDATUM



AC3 oder auch Dolby Digital wurde von Dolby Laboratories entwickelt und steht für Audio Coding in der 3. Generation. Die ersten mit AC3 kodierten Spielfilme erschienen 1992. Der Vorgängercodex AC2 arbeitete nur einkanalig, AC3 verarbeitet maximal 5.1 Kanäle.

Dolby spezifiziert wie bei der MPEG-Kompression nur den Decoder und nicht den Encoder. AC3 ist damit genau so „offen“ für andere Anbieter wie der MPEG-Standard. De facto gibt es aber keine nennenswerten anderen Entwickler, die sich an der aufwendigen Erforschung und Entwicklung von AC3 Encodern versuchen würden. Interessant erscheint in diesem Zusammenhang, dass auch Dolby bei MPEG-2 AAC mitgearbeitet hat und dafür sogar MPEG-Lizenz-Ansprechpartner ist.

10.2 EINSATZGEBIET UND QUALITÄT

Dolby Digital ist seit 1992 mit Premiere im Film „Batman Returns“ das Standard 5.1-Kanal Format im Kinobereich. Seit 1995 findet man AC3 als Tonspur auf Laserdiscs auch im Heimbereich. 1996 gab es erste DVDs mit einer Dolby Digital Spur und 5.1 Sound. Im U.S. Fernsehen ist es seit 1997 fest für den HDTV²²-Standard spezifiziert.

Mit einem relativ hohen Kompressionsverhältniss von durchschnittlich 1:12 bei 384 KBit/s soll die Kompression bei sehr anspruchsvollem Material hörbar sein. MPEG und AC3 arbeiten nach sehr ähnlichen Prinzipien. Entscheidend für die Qualität ist bei beiden Verfahren das jeweils angewendete psychoakustische Modell, welches jedoch nicht konkret spezifiziert wurde.

10.3 EIGENSCHAFTEN DER AC3 KOMPRESSION

Die maximale Kanalanzahl von AC3 sind 5.1 Kanäle. Die Bitraten von AC3 sind folgende:

- 192 Kbit/s Standard für Stereo
- 320 Kbit/s Standard für Film im Kino mit 5.1 Kanälen
- 384 Kbit/s für das Standard-Digitalfernsehen, Standard für 5.1 Kanäle auf DVD
- 640 Kbit/s als High Quality Option für DVD, maximale Datenrate

Das Kompressionsverhältnis bei AC3 beträgt bei der Standardbitrate von 384 KBit/s ca. 1:12, da sechs unkomprimierte Signale mit 16 Bit und 48 KHz Samplingfrequenz eine Datenrate von 4,6 Mbit/s benötigen. Der Frequenzgang von AC3 beträgt 20 Hz bis 20 KHz auf den Hauptkanälen und 20 bis 120 Hz auf dem LFE-Kanal²³. Im Dolby Digital Stream ist ein Timecode zur Bildsynchronisation integriert.

Die Standardspurbelegung bei AC3 ist folgende: 1 Left, 2 Right, 3 Center, 4 LFE, 5 Left Surround, 6 Right Surround, 7 Lt (optional), 8 Rt (optional).

AC3 unterstützt digitales Audio mit bis zu 20 Bit Wortlänge und bis 48 KHz Samplingfrequenz.

²² Abkürzung für „High-Definition Television“

²³ LFE – Low Frequency Effect (Subwoofer)

10.4 ARBEITSWEISE DER AC3 KOMPRESSION

Das Eingangssignal wird in Blöcken von 512 Samples mit 50% Überlappung aufgeteilt und mittels MDCT (die AC-3 Dokumente nennen es nur TDAC, was Time Domain Aliasing Cancellation heißt und die wichtigste Eigenschaft von MDCT beschreibt) von der zeitlichen in die Frequenzdarstellung gebracht. Die dadurch entstehenden 256 spektralen Koeffizienten werden zu 50 Frequenzgruppen zusammengefasst, deren Bandbreite der Empfindlichkeitsverteilung entlang der Basiliarmembran im Ohr entspricht. Wie bei den anderen Verfahren, die auf Datenreduktion durch Ausnutzung der Psychoakustik beruhen, werden die einzelnen Bänder durch ein psychoakustisches Modell gemäß ihrer Wichtigkeit und Verdeckungseigenschaften bewertet und dann entsprechend quantisiert. Auch AC3 besitzt eine Transientenerkennung, die bei stark impulshaftem Material die Audioblocklänge reduziert, um das Quantisierungsrauschen stets unter der Maskierungsschwelle zu halten.

AC3 besitzt ein sehr effektives Bitmanagement-System. Der gemeinsame Bitvorrat wird kanalübergreifend vergeben. Wenn ein Kanal für sehr komplexes Material mehr Bits als normal zur Kodierung benötigt, kann er diese auf Kosten anderer weniger „datenhungriger“ Kanäle zugewiesen bekommen. Auch können Signale mit hohem Pegel in einem bestimmten Kanal leisere Signale auf den anderen Kanälen verdecken.

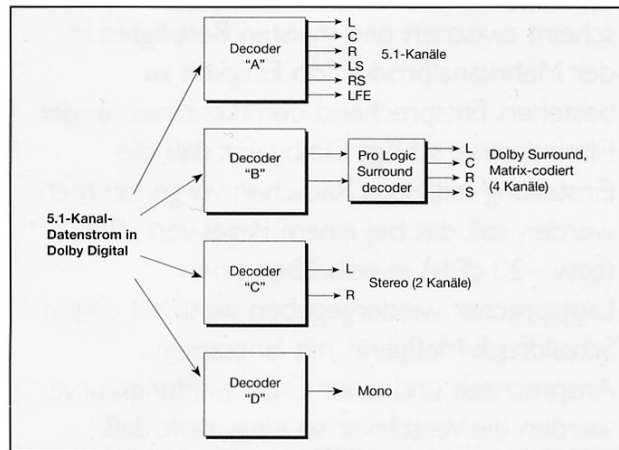
Der Encoder kann bei sehr anspruchsvollem Material das psychoakustische Standardmodell kurzzeitig modifizieren, um Artefaktbildung zu reduzieren. In diesem Fall werden entsprechende Änderungsbits mit Informationen für den Decoder im Bitstrom eingebettet.

Nach der Entfernung der Irrelevanzen werden die Daten mit einem speziellen Packmechanismus komprimiert.

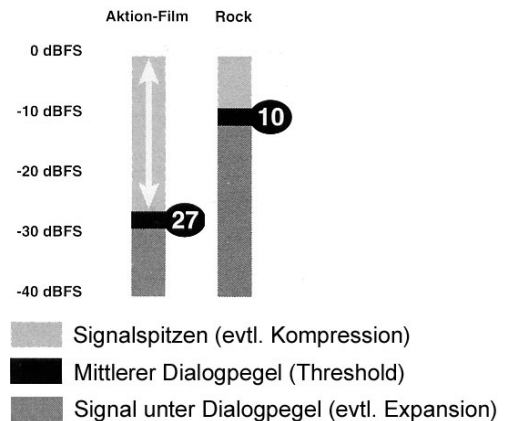
10.5 ÜBER DEN DECODER

Der AC3 Decoder besitzt im Consumerbereich flexible Möglichkeiten des Heruntermischens mehrerer Kanäle auf eine geringere Kanalanzahl. Er ermöglicht eine unkomplizierte Anpassung des Audiodatenstroms an die gewünschte Lautsprecherkonfiguration. Dabei wird schon beim Enkodierprozess das Verhalten des Decoders beim Downmixen festgelegt und in den AC3-Stream integriert.

Ein weiteres Decoderfeature ist die Dialog Normalisierung. Beim Enkodierprozess gibt man den Standardpegel für Dialoge normaler Lautstärke an (die sogenannte „Dialnorm“). Die Dolby Empfehlung liegt bei -31 dB(Fs). Dieser Wert wird in den Metadaten des Codes gespeichert. Der Anwender kann nun je nach Abhörsituation oder zur Anpassung an andere Programme diese Normal-Dialog Lautstärke hochsetzen. In den Metadaten des AC3 Streams sind Informationen zur Steuerung des Dynamikumfangs enthalten. Man kann den Decoder somit als einfachen Kompressor anwenden, um die Wiedergabelautstärke der Umgebung anzupassen. Die Lautstärke des normalen Dialoges (Dialnorm) bildet sozusagen den Schwellwert (Threshold) des Kompressors, über dem je nach Einstellung die Kompression ansetzt.



Der AC-3-Datenstrom eignet sich für verschiedene Wiedergabebedingungen



Grafik-Quelle: [28]

11 DTS (DIGITAL THEATER SYSTEMS)

11.1 ENTWICKLER UND ERSCHEINUNGSDATUM



1993 hatte DTS mit dem Film „Jurassic Park“ seine Weltpremiere. Entwickelt wurde DTS von Digital Theater Systems, Inc.

11.2 EINSATZGEBIET UND QUALITÄT

DTS bietet hochqualitativen 5.1- bis 6.1-Kanal Sound für das Kino und seit 1998 für die DVD, bei den Standard-Bitraten sogar ohne verlustbehaftete Kompression. Mit dem „Coherent Acoustics“ Algorithmus bedient DTS den Heimbereich mit einem flexiblen Codec für Mehrkanalanwendungen mit frei einstellbaren Bitraten und Qualitätsstufen. Seit 1997 wird DTS auch auf Laserdiscs eingesetzt.

11.3 EIGENSCHAFTEN DER DTS KOMPRESSION

Die Standardbitrate von DTS im Kinoton liegt bei 1411 Kbit/s, ähnlich der Bitrate einer normalen Audio-CD. Das Kompressionsverhältnis beträgt bei 16 Bit Material 2,9:1 oder 4,3:1 bei 24 Bit. Bei DVD-Video liegt die Datenrate etwas höher bei 1536 Kbit/s, da mit 48 KHz Samplingfrequenz und bis 20 Bit Wortbreite gearbeitet wird. Für besonders lange Filme und datenintensive Programme ist die halbierte Datenrate von 754 Kbit/s vorgesehen.

Der für die Heimanwendung und DVD konzipierte DTS-Algorithmus mit dem Namen „Coherent Acoustics“ bietet diverse einstellbare Bitraten von 32 Kbit/s bis 4096 Kbit/s. Es sind bis zu acht Kanäle in 24 Bit möglich bei einer Samplingrate bis 192 kHz. Implementiert werden meist 5.1 Kanäle. Es gibt auch eine variable Bitrate für verlustfreie Kompression.

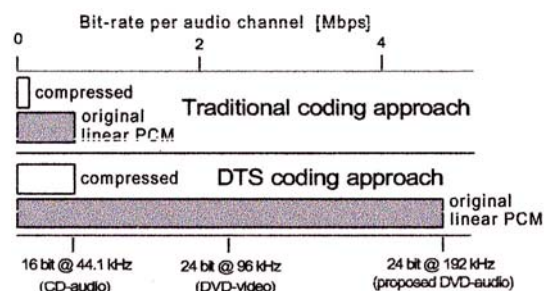
Mögliche Kanalkombinationen sind: 1/0, 2/0, 3/0, 2/1, 2/2 und 3/2, der LFE Kanal ist nach Bedarf zuschaltbar.

Audio data-rate [kbit/s/channel]	Sampling rate	Sample resolution	Quality
8 to 32	Up to 24 kHz	16 bit	Telephony
32 to 96	Up to 48 kHz	20 bit	CD-audio
96 to 256	Up to 96 kHz	24 bit	Studio
256 to 512	Up to 192 kHz	24 bit	> Studio
Variable	Up to 192 kHz	24 bit	> Studio

Quelle: [33]

11.4 ARBEITSWEISE DER DTS KOMPRESSION

DTS verwendet zur Datenreduktion meistens kein „Perceptual Coding“, d.h. bei den höheren Datenraten nutzt es nicht die Anwendung von psychoakustischen Erkenntnissen zur Entfernung von Irrelevanzen. Die Datenreduktion von DTS geschieht bei den Standard-Bitraten verlustfrei durch effektive Weglassung von Redundanzen wie z.B. bei ZIP in der Computerwelt. Bei höheren Kompressionsverhältnissen ab 4:1 wird verstärkt „Perceptual Coding“ zugeschaltet und entsprechend der Psychoakustik Irrelevanzen herausgerechnet.



Der Coherent-Acoustics-Algorithmus teilt das Signal zur Analyse in 32 Frequenzbänder und benutzt für die anschließende Transformation fünf verschiedene Stützstellenanzahlen je nach Bitrate und Samplingfrequenz (256, 512, 1024, 2048, 4096 Samples). Bei niedrigeren Bitraten und im Kinobereich wird ein älterer apt-X100 Algorithmus verwendet, der mit nur 4 Frequenzbändern arbeitet. Die verwendeten Kompressionsmethoden bei den niedrigeren Bitraten

ähneln stark den unter MPEG oder AC3 ausführlicher beschriebenen Verfahren. Angewandt werden unter anderem: Linear Prediction, ADPCM (Adaptive Quantization), Ausnutzung der Verdeckungseffekte anhand eines psychoakustischen Modells und Variable Length Coding.

11.5 ANDERE DTS STANDARDS

- DTS-ES (Extended Surround) – zusätzlicher Surround Center Kanal, matriziert in den LS und RS Kanälen, entspricht Surround EX
- DTS-ES 6.1 – mit zusätzlichem diskretem Surround Center Kanal
- DTS 96/24 – für bis zu 5.1 Kanäle in 24 Bit, 96 KHz, digitale Ausgabe über SPDIF möglich
- DTS Neo:6 – hochqualitatives Dolby Surround Decoding, wandelt ein matriziertes Lt Rt – Signal in bis zu 6.1 Kanäle um

12 OGG VORBIS

12.1 ÜBER DIE OGG KOMPRESSION



OGG Vorbis ist ein patent- und lizenzfreier Codec (Open Source), der von der Xiphophorus Company entwickelt wurde. Bei den gängigen Bitraten von 96 bis 160 KBit/s ist OGG MP3 qualitativ durchaus ebenbürtig, wenn nicht sogar durch die variable Bitrate qualitativ besser. Im Gegensatz zu den meisten Qualitäts-MP3-Codecs ist OGG außerdem völlig kostenlos.

12.2 EIGENSCHAFTEN DER OGG KOMPRESSION

OGG arbeitet mit Samplingfrequenzen von 8 bis 48 KHz bei 16 Bit Wortlänge. Im jetzigen Entwicklungsstand arbeitet OGG ausschließlich mit variablen Bitraten zwischen 16 bis 350 KBit/s. Bis zu 255 mögliche Kanäle werden zur Zeit noch diskret komprimiert. Es existieren noch keine Joint Stereo Modi. OGG ist für Echtzeitanwendungen wie Internetradio streamingfähig.

12.3 ARBEITSWEISE DER OGG KOMPRESSION

OGG arbeitet in starker Anlehnung an die MPEG-Kompression mit Signalzerlegung in überlappende Audioblöcke, Anwendung einer MDCT, Datenreduktion durch Anwendung eines psychoakustischen Modells der Maskierung, flexibler Quantisierung der Frequenzgruppen und Huffman-Kodierung zur weiteren Datenreduktion.

13 DIE VERLUSTLOSEN KOMPRESSIENEN MLP UND DST

13.1 ÜBER MLP (MERIDIAN LOSSLESS PACKING)



MLP ist ein verlustfreies Kompressionsverfahren von Dolby und wurde im August 1998 als Kompressionsverfahren für DVD-Audio verabschiedet. Die Kompressionsrate von MLP liegt zwischen 38 und 52 %. Ein MLP Decoder kann das Signal bitgenau wiederherstellen. Ohne MLP würden auf eine Single Layer DVD nur 45 Minuten mit 24 Bit / 96 KHz 5.1 Kanal Audio passen, mit MLP sind es mehr als 74 bis 89 Minuten. Unterstützt werden bis zu 63 Kanäle von 32 bis 192 KHz Samplingfrequenz bis 24 Bit Wortbreite. Der MLP – Datenstrom beinhaltet auch einen Kanal für Metadaten für z.B. die Lautsprecherzuordnung der kodierten Kanäle, Kanalhierarchien, Copyrights etc.

Die Arbeitsweise der MLP-Kompression ähnelt der im folgenden beschriebenen DST-Kompression.

13.2 ÜBER DST (DIGITAL STREAM TRANSFER)

DST ist ein verlustfreies Kompressionsverfahren von Philips. Es wird speziell auf den von Philips entwickelten SuperAudio-CDs²⁴ verwendet, um auf einer SA-CD 74 Minuten 5.1 Kanäle in der Direct Stream Digital-Technologie mit 1 Bit und 2,8224 Mhz Abtastrate unterzubringen. Der bei der neu entwickelten DSD-Delta Sigma Modulation entstehende Datenaufwand ist wegen der hohen Oversamplingrate ungefähr viermal so groß wie bei einer Audio-CD (2822,4 Kbit/s pro Kanal). Die durchschnittliche Kompressionsrate liegt bei ungefähr 2:1. Ein DST Decoder kann das Signal bitgenau wiederherstellen.



13.3 ÜBER DIE ARBEITSWEISE VON MLP UND DST

In Echtzeit werden kurze Audioblöcke aus dem Datenstrom entnommen. Dieser Prozess wird als „Framing“ bezeichnet und erfolgt z.B. bei DST mit 75 Frames pro Sekunde. Für jeden Audioblock wird die Kompression, d.h. die Entfernung der Redundanzen, jedes Mal neu optimiert. Es werden Redundanzen zwischen beiden Kanälen und zwischen den Samples eines Kanals analysiert und reduziert. Dieses Verfahren ähnelt der Huffman-Kodierung, allerdings mit einer speziellen adaptiven Methode, um im Audiomaterial möglichst effektiv Redundanzen aufzufinden. Am Ende eines jeden Frames wird eine Substitutionstabelle geschrieben, mit deren Hilfe ein Decoder die komplette Bitfolge wiederherstellen kann. Am Ausgang des Encoders hat das Signal eine variable Bitrate, weshalb es, bevor es auf die Platte oder auf das Band kommt, in einem Zwischenspeicher gepuffert werden muss.

²⁴ bietet Mehrkanalton und besseren Klang als normale Audio-CD, mehr Infos zur SA-CD beispielsweise unter: www.clockwork.de/sacd/sacd.htm

14 TABELLARISCHE ÜBERSICHTEN UND VERGLEICH

14.1 ÜBERSICHT DVD/KINO-MEHRKANALCODECS

			
Systembezeichnung	Dolby Digital (AC-3)	DTS	MPEG-2 Audio
Kompressionsverfahren	AC-3	APT-X 100 (Kino) Coherent Acoustics (Heim)	MPEG-2 Audio coding
Einsatz im Kino	seit 1993	seit 1993	-
Einsatz als Heimsystem	LaserDisc seit 1995 DVD seit 1996 US TV seit 1997	CD seit 1996 LaserDisc seit 1997 DVD seit 1998	DVD seit 1998
Spezifikationen			
Datenrate	32 (mono) - 640 kBit/s (5.1)	max. 4.096 kBit/s	32 (MPEG 1) - 1.060 kBit/s
Datenrate im Kino	320 kBit/s	1.411 kBit/s	-
Datenrate auf LD	384 kBit/s	1.411 kBit/s	-
Datenrate auf DVD	384 kBit/s - 640 kBit/s	1.536 kBit/s	448 kBit/s - 640 kBit/s
Max. Kanalanzahl	5 plus Subwoofer	8 inkl. Subwoofer	7 plus Subwoofer
Frequenzgang Front	20Hz-20kHz	20Hz-20kHz	20Hz-20kHz
Frequenzgang Rear	20Hz-20kHz	80Hz-20kHz	?
Dynamik Front	92dB	90dB	?
Dynamik Rear	92dB	96dB	?
Frequenzbereich Subwoofer	bis 120 Hertz	bis 80 Hz	bis 100 Hertz
Bit-Wortlänge	bis 20 Bit	bis 24 Bit	bis 20 Bit
Max. Abtastfrequenz	48 kHz	192 KHz	48 kHz
Optionen			
Dynamik-Reduzierung	ja (Midnight Mode)	ja	nein
Lautstärke-Angleichung	ja	nein	nein
Variable Datenrate	nein	nein	ja
Abwärts-Kompatible	nein	nein	zu MPEG-1
Verlustfreie Kodierung	nein	ja	nein
Hintergrund			
Entwickelt von	Dolby Labs	DTS Technology Inc.	IRT, Fraunhofer Institut
Unterstützt von	Toshiba und anderen	Dreamworks, Panasonic	Philips, Sony
Kinosäle weltweit	15.700 (Juni 98)	12.000 (Mai 98)	-
Verfügbare Kinofilme	8.700	4.900	-
Lieferbare LDs	387	52	-
Lieferbare DVDs	ca.680 (2.0 und 5.1)	6	8 (nur 5.1, mehr MPEG-1)

Quelle: [30], [31]

14.2 ÜBERSICHT AUDIOCODECS AUS DEM COMPUTER-/INTERNETBEREICH UND ATRAC

	MP3	MP3 Pro	AAC	TwinVQ	ATRAC3	OGG
Entwickler	Fraunhofer IIS, Coding Technologies, Thomson	Fraunhofer IIS, Coding Technologies, Thomson	Fraunhofer IIS, Dolby, Sony, AT&T	NTT (Japanische Telekom)	Sony	Xiphophorus Company
Erscheinungsjahr	1992	2001	ca. 1997	? (seit ca. 1993)	2000	?
Enthalten im Standard	MPEG-1 und MPEG-2 (als Layer 3)	Weiterentwicklung von MP3, nicht in MPEG-1/2 spezifiziert	MPEG-2, MPEG-4	MPEG-4	ATRAC	patentfreier Open Source Codec
Datenrate min, max	8 ... 320 KBit/s für Stereo	24 ... 96 KBit/s für Stereo	ca. 8 ... 512 KBit/s pro Kanal	6 KBit/s (Mono) ... 96 KBit/s für Stereo	66 KBit/s ... 292 KBit/s für Stereo	16 ... 350 KBit/s für Stereo
typische Datenrate bei guter Qualität	128 KBit/s für Stereo Kompression: 1:12	96 KBit/s für Stereo Kompression: 1:15	sehr gut bei 320 KBit/s für 5 Kanäle, 96 KBit/s für Stereo, Kompression 1:15	mäßige Qualität bei 96 KBit/s für Stereo, Kompression 1:15	132 KBit/s für Stereo, Kompression: 1:11	128 KBit/s, Kompression 1:12
max. Kanalanzahl	2	2	48 + 16 LFE	2	2	255
Frequenzgang	je nach Bitrate, ab 96 KBit/s > 15 KHz für Stereo	bis 16 KHz	je nach Bitrate	je nach Bitrate	bis 16 KHz	je nach Bitrate
max. Bit-Wortlänge	16 Bit	16 Bit	?	? (16 Bit)	16 Bit	16 Bit
max. Abtastfrequenz	48 KHz	44 KHz (?)	96 KHz	? (44 KHz)	? (48 KHz)	48 KHz
Besonderheiten	Joint Stereo Modi, variable und konstante Bitraten, streamingfähig	Spectral Bend Replication - Technologie, ansonsten siehe MP3	Temporal Noise Shaping, Long Time Prediction, Perceptual Noise Substitution, Skalierbarkeit, ansonsten siehe MP3	Quantisiert das Audiospektrum mittels festen Mustern von Vektoren	Block-Floating Algorithmus, Joint Stereo Modus	zur Zeit nur mit variabler Bitrate und ohne Joint Stereo Modi
Haupteinsatzgebiet	Internet-Audio, Musikkodierung und Archivierung	Internet-Audio bei geringer Bandbreite, Vorhören von Audioclips	High Quality Audiokodierung, Internet, DAB	Musik- und besonders Sprachkodierung bei sehr geringen Bitraten	Minidisk, portable Musik-Player	Internet

15 SCHLUSSBETRACHTUNG

1992/93 kamen nahezu gleichzeitig die neuen, komplexen Audiodatenreduktionsverfahren MPEG-1, AC3, DTS heraus, nachdem kurz zuvor Sony mit ATRAC neue Maßstäbe in Sachen effektiver und platzsparender Audiospeicherung gesetzt hatte. MPEG-1 Layer 3 und AC3 übertrafen ATRAC bei halber Bitrate in besserer Qualität. MP3 wurde durch das Internet zum ersten massenhaft verbreiteten Kodierstandard. Ohne große Kenntnisse konnte nun fast jeder Anwender Musik bei hoher Qualität auf ein zwölftel der ursprünglichen Größe reduzieren oder komprimiertes Audio mehr oder weniger legal übers Netz herunterladen. Die mehrkanaligen AC3 und DTS-Decoder in DVD-Playern hielten und halten Einzug in die Wohnzimmer und ersetzen nach und nach den Videorecorder samt Dolby Surround Decoder. Seit 1992 hat sich viel im Bereich der Audiokompressionscodecs getan. Bestehende Verfahren wurden optimiert und es entstanden neue Verfahren für mehrere Kanäle und den 24 Bit / 96 KHz Standard. Viele Hersteller meinten, mit ihrem Verfahren innerhalb kurzer Zeit den MP3-Standard ablösen zu können. Bis heute wurde daraus nichts, denn wirkliche Revolutionen in der Audiokompression hat es seit Anfang der neunziger Jahre nicht mehr gegeben. Viele Erkenntnisse von gestern gelten immer noch heute. Durch Anwendung eines psychoakustischen Modells zur Irrelevanzreduktion von Audiodaten wird eine Kompression von ca. 1:6 und durch geschickte Anwendung Huffman-ähnlicher Packalgorithmen zur Irrelevanzreduktion eine Kompression von ca. 1:2 erreicht. Das ergibt zusammengerechnet ein Kompressionsverhältnis von 1:12 bei guter Qualität des kodierten Materials. Heute, im Frühjahr 2002, erreicht man durch die Forschung und Erkenntnisse der letzten zehn Jahre, durch ausgefuchste

Redundanzeliminierungen, bessere psychoakustische Modelle, optimale Ausnutzung des Bitvorrates und teilweise Synthetisierung bestimmter Signalanteile eine maximale Verbesserung um 30% gegenüber MP3. Wie der MPEG-4 Standard aufzeigt, hängt die Auswahl des Codecs von der Art des zu kodierenden Materials ab. Hochwertige klassische Musik gibt es auch heute kaum unter einer Kompression von ca. 1:13, wohl aber einkanalige Sprachbeiträge selbst noch bei 1:20. Bei den meisten unabhängigen Hörtests schneidet zur Zeit der MPEG-4 AAC-Codec im Bereich hoher Qualität (128 KBit/s) am besten ab. Auch in Verbindung mit der SBR-Technik, scheint AAC bis runter zu 56 KBit/s am effizientesten zu arbeiten. Trotzdem wird noch viel Zeit vergehen, bis das hochkomplexe MPEG-4 den einfach anzuwendenden und robust funktionierenden MP3-Codec ablösen wird. Umständliche und teure Lizenzierungen und die gebührenpflichtige Verwendung von De- und Encodern bremsen zur Zeit die breite Einführung des MPEG-4 Standards. Sony ist seit der Veröffentlichung von ATRAC3 für Minidiscs und portable Player in Sachen Qualität bei geringer Bitrate wieder auf der Höhe der Zeit. Qualitativ war zwar an ATRAC schon seit der Version ATRAC1 V.3 nichts mehr zu bemängeln, jedoch erschien bis zur Einführung von ATRAC3 das Kompressionsverhältnis im Vergleich zu MPEG nicht mehr dem aktuellen Stand der Entwicklung zu entsprechen. Das alternative open-source Kompressionsverfahren OGG zeigt, dass man für einen qualitativ hochwertigen Codec nicht unbedingt Geld an die MPEG-Organisation zahlen muss. OGG ist MP3 gegenüber als gleichwertig einzuschätzen. Bis zur jetzigen Version ist es jedoch kompliziert zu handhaben (per Befehlszeile oder einen Drop-Objekt) und kaum verbreitet, kann dafür aber wie DTS oder AC3 mehr als nur zwei Kanäle kodieren. MP3-Pro behält bei niedrigen Bitraten um 64 KBit/s den unbeschnittenen Frequenzgang bis 16 KHz, jedoch kann man bei den synthetisierten hohen Anteilen und geringen Artefakten im normalen MP3-Layer nicht mehr von CD-Qualität sprechen. Auch der Exot unter den Codecs, TwinVQ mit seiner ausgefallenen Datenreduktionsmethode, ist nicht für hochqualitative Musikreproduktion geeignet, sondern hat seine Stärken bei Sprachkompression mit geringen Bitraten. Im Videobereich konkurrieren derzeit zwei Mehrkanal-Codecs um die Gunst der Zuschauer – AC3 und DTS. MPEG-2 spielt auf DVDs mittlerweile keine Rolle mehr. DTS hat gegenüber AC3 den Vorteil einer weitaus größeren Bitrate und einer viel kleineren, meist verlustfreien Kompression. AC3 hingegen muss bei einer MP3-ähnlichen Kompression von 1:12 das Audiomaterial stark reduzieren, so dass geschulte Hörer Artefakte hören können. Der DTS Coherent Acoustics Algorithmus bietet auch stärkere Kompressionen als „normales“ DTS, jedoch gibt es selbigen wie auch den AC3-Codec nicht (oder zumindest nur sehr schwer) für den Heimanwender-PC-Bereich, aufgrund dessen ich im Zuge dieser Vordiplomarbeit auch keine qualitativen Bewertungen zu DTS und AC3 machen kann. Es scheint, dass nicht die Musikindustrie, sondern der Video- und Multimediabereich die stärkste Triebkraft für die Entstehung und Weiterentwicklung mehrkanaliger Codecs darstellt. Reine Audioaufnahmen mit 5.1 Kanälen und 24 Bit oder DSD-Qualität sind selbst heutzutage nur einer kleinen Gruppe elitärer Hifi-Freaks vorbehalten, da diese neuen Techniken für den Normalverbraucher selten einen spürbaren Qualitätssprung bedeuten. MLP und DST sind verlustfreie Algorithmen, die gerade für diese reinen Audioformate entwickelt wurden und dort unbemerkt ihre unspektakuläre Kompression verrichten.

16 ANLAGE

Zu dieser Vordiplomarbeit gehört eine Audio-CD mit vielen Hörbeispielen zu einigen der hier vorgestellten Datenreduktionsverfahren. Durch intensive Recherchen für diese Arbeit im Internet stieß ich einige Male auf En- und Decoder der jeweiligen Verfahren zum Herunterladen und Ausprobieren. Die Audio-CD soll es dem Interessenten ermöglichen, sich ein eigenes Urteil über die diversen Codecs zu bilden.

Inhalt der CD: (die Tracks bestehen aus kurzen, prägnanten Abschnitten der jeweiligen Musiken)

- Track 1 bis 14: Beispiel 1 - Pop Song (Annas Song, Gaillard-Studios-Productions 2002)
- Track 15 bis 28: Beispiel 2 – Akustische Gitarre, Solo („Sichia“ von der CD „Passion Grace and Fire“ von John McLaughlin, Al DiMeola and Paco De Lucia)
- Track 29 bis 42: Beispiel 3 – Gesang, Solo („Amazing Grace“ von den Kissingers)
- Track 43 bis 56: Beispiel 4 - Percussion Groove (“Breathe” von Minus 8)

- Track 57 bis 70: Beispiel 5 - Klassik (Antonin Dvorak, „Aus der Neuen Welt“, Satz 3 Scherzo: Molto Vivace)
- Track 71 bis 84: Beispiel 6 - Gitarre und Klavier (Soundtrack zu „Seefuchs“ von Torsten Lüders)

Jedes dieser sechs Beispiele wird zuerst im Original gespielt. Die nachfolgenden Tracks wurden mit den folgenden Verfahren komprimiert, dekomprimiert und schließlich auf diese CD gebrannt:

- MPEG1 Layer 3 96 KBit/s CBR (Fraunhofer IIS MPEG Layer 3 Codec Professional)
- MPEG1 Layer 3 64 KBit/s CBR (Fraunhofer IIS MPEG Layer 3 Codec Professional)
- MP3-Pro 64 KBit/s CBR (IS) (Thomson MP3-Pro Audio Player Demo 1.04)
- MPEG4 AAC 96 KBit/s CBR (Psytel Encoder AACENC 2.11)
- MPEG4 AAC 85 KBit/s VBR (Psytel Encoder AACENC 2.11)
- MPEG4 AAC 96 KBit/s CBR (Freeware AAC Encoder FAAC 1.9)
- TWIN-VQ 96 KBit/s CBR (Yamaha SoundVQ Encoder 2.60)
- TWIN-VQ 96 KBit/s CBR (Nero Burning Rom 5.53)
- TWIN-VQ 80 KBit/s CBR (Nero Burning Rom 5.53)
- ATRAC3 132 KBit/s CBR
- ATRAC3 104 KBit/s CBR
- OGG 96 KBit/s VBR (Vorbis OGG Encoder OGGENC 0.9)
- OGG 64 KBit/s VBR (Vorbis OGG Encoder OGGENC 0.9)

17 GLOSSAR

Joint Stereo: Methode zur Entfernung von Irrelevanzen und Redundanzen zwischen zwei Audiokanälen, gemeint ist entweder die Bildung eines Mitte-/Seitensignals (MS) oder Bildung eines Monosignals mit Angaben zur intensitätsstereophonischen Lautstärkeverteilung (IS) zwischen den Kanälen

CBR: Konstante Bitrate, besonders geeignet für Audioübertragungen mit festgelegter Bandbreite (z.B. via ISDN)

VBR: Variable Bitrate, erzeugt je nach Inhalt des Audiomaterials optimale Qualität durch freie Verfügung über benötigte Bits

SBR: Spectral Band Replication, hauptsächlich bei MP3-Pro angewandte Technologie, um durch Bandbreiteneinengung verloren gegangene hohe Signalanteile synthetisch wiederherzustellen

TNS: Temporal Noise Shaping, das durch die Datenreduktion entstehende Quantisierungsrauschen kann in kürzeren Abständen als die Audioblocklänge der Transformation an das Audiosignal angepasst werden, um hörbare Artefakte zu minimieren

LTP: Long Time Prediction, entfernt Redundanzen zwischen längeren Audiosegmenten speziell bei statischem, tonalen Material

Block Floating Algorithmus: durch gezielte Verkürzung von „Wörtern“ deren höherwertige Bits ungenutzt sind, gewinnt man eben diese Bits, um sie an wichtigeren Stellen einzusetzen

Skalierbarkeit: freie Einstellmöglichkeiten bezüglich der angestrebten Bitrate; nur ein Teil des gesamten Bitstromes ist notwendig, um das Signal mit gewissen qualitativen Einschränkungen zu reproduzieren

DAB: Digital Audio Broadcast, der digitale Rundfunk per Satellit

Artefakte: sind durch Kompression entstandene hörbare Verzerrungen und Störgeräusche wie das Quantisierungsrauschen, sie resultieren aus schlechten Codecs oder zu geringen Bitraten

PCM: Pulse Code Modulation, n-mal pro Sekunde wird eine elektrische Schwingung auf ihren Schwingungszustand hin abgetastet, die Abtastwerte werden bei z.B. 16 Bit in 65 536 mögliche Stufen eingeteilt / quantisiert und abgespeichert

Codec: ein Codec ist ein Programm, das die Komprimierung einer Multimedia-Datei - z.B. einem Video oder Musikstück - sowie dessen Wiedergabe durch einen Media Player ermöglicht

Streaming: bezeichnet das simultane Herunterladen und Wiedergeben von Audio- und Videodateien, Anwendungsbeispiele sind Videokonferenzen und Radiosendungen per Internet

Quantisierungsrauschen: durch die beschränkte Auflösung/Quantisierung von Samples und Frequenzgruppen entstehen kantige Übergänge im Signal, die bei der Digital/Analog Wandlung ein Rauschen erzeugen

Maskierung: Überdecken eines Signals durch ein anderes, das alte Signal ist noch vorhanden, es wird aber nicht mehr wahrgenommen

Frequenzgruppen: die Frequenzgruppe (engl.: critical band) ist das Frequenzband, innerhalb dessen das menschliche Gehör Intensitäten zusammenfasst; im Frequenzbereich von 20 Hz bis 16 kHz gibt es

etwa 24 Frequenzgruppen; mit steigender Frequenz nimmt ihre Bandbreite zu (von 100 Hz für tiefe Frequenzen bis 4 kHz für Höhen)

Open Source: bedeutet, dass der Programmquellcode für Programmierer für Weiterentwicklungen offen liegt und keine Patent oder Lizenzgebühren zu entrichten sind

Blocklänge / Audioblock: Die Blocklänge gibt die Stützstellenanzahl an, die bei der Transformation von der zeitlichen in die spektrale Darstellung eines Signals betrachtet wird; die Stützstellenanzahl entspricht der Anzahl der durch die Transformation erhaltenen Spektralwerte

Überlappung: die einzelnen Audioblöcke überlappen sich meist um einen gewissen Prozentsatz, um die zeitliche Auflösung zu erhöhen und die Übergänge zwischen einzelnen Blöcken harmonischer zu gestalten

Fensterung / Fenstergröße: Blöcke werden mit einer Fensterung versehen, die meist an den Enden des Blockes die Abtastwerte Richtung Nulllinie energetisch abrunden, um die scharf abgeschnittenen Kanten der abgetasteten Welle zu glätten; die gewonnenen Spektralwerte haben durch geschickte Fensterung geringere sogenannte Nebenkeulen, das sind Fehlerstreuungen um die Zielfrequenz herum (nicht vorhandene Nebenfrequenzen)

LFE: Low Frequency Effect (Subwoofer)

LITERATURVERZEICHNIS UND LINKS ZUM THEMA

ALLGEMEINE QUELLEN:

- [1] www.fefe.de/ct/audio.pdf - von Leitner, Felix (1999): Audio Kompression
- [2] Schriber, Pascal (2000): Das neue Audiospeicherformat MP3, Diplomarbeit, Universität Zürich
- [3] Noll, Peter (2000): Vom Hörfunkstandard zum Advanced Audio Coding
- [4] Flohr, Daniel (2001): Elektronische Audioformate im Internet, Proseminarscript
- [5] www.billabong-media.com/compression/codecs.html - Hayward-Warburton, John (1999): The Codec Comparisons
- [6] www.hydrogenaudio.org - vBulletin Portal, Audio Forum und Infoseite
- [7] www.audiocoding.com - MPEG-2 / MPEG-4 Infoseite, Audioforum, Codec-Downloads
- [8] www.filmvorfuehrer.de - die Filmvorfuehrer Homepage, kurze Infos über alle Kinotonformate
- [9] Schumann, Roger (2001): Audio und Internet, Vordiplomsarbeit, HFF

MPEG-KOMPRESSION:

[10] <http://mpeg.telecomitalialab.com> - die MPEG-Homepage, detaillierte Übersichten mit Spezifikationen aller MPEG-Standards

- Chiariglione, Leonardo (1996): Short MPEG-1 description
- Leonardo Chiariglione, Chiariglione (2000): Short MPEG-2 description
- Koenen, Rob (2001): Overview of the MPEG-4 Standard
- Martínez, José (2001): Overview of the MPEG-7 Standard
- Bormans, Jan und Hill, Keith (2001): MPEG-21 Overview v.3

[11] www.crs4.it/~luigi/MPEG - ausführliche MPEG-Informationen, FAQs

[12] Vorlesungsscript: Digitale Audioverarbeitung, Uni Hamburg, WS 2000

[13] <http://substream.org> - Tutorial und Infoseite über MP3 und Streaming

[14] http://www.zdnet.de/technik/artikel/prg/199802/mp3_00-wc.html - Golgath, Andreas und Popp, Harald: MPEG Layer 3 (TechTalk Thema Audiokompression bei ZDNet)

[15] Weigel, Peter: Das Audiodatenreduktionsverfahren MP3, Vordiplomsarbeit, HFF

[16] Rätz, Dominik: Digital Audio Broadcasting, Vordiplomsarbeit, HFF

[17] www.mp3prozone.com - kommerzielle Infoseite zu MP3-Pro und SBR von Coding Technologies

[18] www.codingtechnologies.com - Homepage von Coding Technologies, Infos über MP3 Pro und SBR

[19] <http://www.powerweb.de/phade/diplom/kap234.htm> – Gadegast, Frank (1995): TCP/IP-basierte Dienste zur Speicherung von Multimedia-Daten, Kapitel Audio; Diplomarbeit TU Berlin, Infos über MPEG-1 und 2

[20] http://leonardo.telecomitalialab.com/icjfiles/mpeg-4_si - detaillierte Papiere und Abhandlungen über MPEG-4

ATRAC-KOMPRESSION:

[21] M.Pregler, Vordiplomsarbeit „Die Minidisk als Tonmedium in der professionellen Audiotechnik“

[22] www.minidisc.org - Minidisc Community Portal, ATRAC Infos, FAQs, Codec-Vergleiche

[23] http://minidisc.amulation.com/atrac_versions.html - ATRAC Versionen Liste diverser MD Recorder

[24] Techno World Informationsblätter (2000): ATRAC3 High Quality Audio Encoding Technology

[25] Körner, Dietrich: Das Minidisc-Format – Technische Grundlagen und Anwendung des neuen digitalen Speichermediums, Vordiplomsarbeit, HFF

[26] Pregler, Michael (2001): Die Minidisk als Tonmedium in der professionellen Tontechnik, Vordiplomsarbeit, HFF

AC3-KOMPRESSION:

[27] www.dolby.com - kommerzielle Homepage der Dolby Laboratories, AC3 / Dolby Digital Infos

[28] Mitchell, Doug (1998): Entschlüsselt: Der Dolby Digital Code

[29] Davis, Mark (1993): The AC-3 Multichannel Coder, Technical Papers

[30] http://212.112.179.250/Background/Dolby_Digital.htm - Becker, Jörg: Dolby Digital und Surround Sound

[31] www.digital-movie.de - das Online DVD & Heimkino Magazin, Infos über AC3

DTS-KOMPRESSION:

[32] www.dtsonline.com - kommerzielle Homepage von Digital Theater Systems, Inc.

[33] Smyth, Mike (1999): White Paper – An Overview of the Coherent Acoustics Coding System

OGG-KOMPRESSION

[34] www.xiph.org - Homepage der Xiphophorus Company, den OGG Entwicklern

MLP UND DST-KOMPRESSION

[35] www.meridian-audio.com - kommerzielle Homepage von Meridian Audio Ltd.

[36] www.daisy-laser.com - Digital Audio Industry Supply – Technologien von Philips ; Infos über DTS, SA-CD, DVD etc.

[37] Maderlechner, Titus (2001): Die Super-Audio CD, Technologie und Funktionsweise, Vordiplomsarbeit, HFF