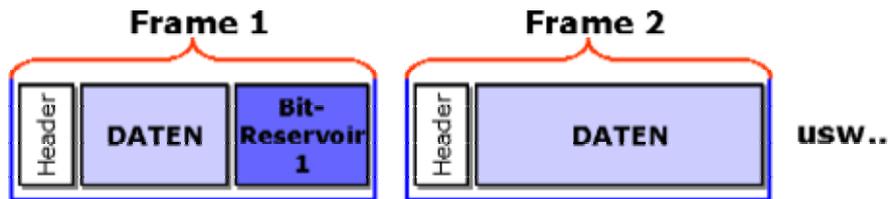


6. MP3

6.1 Aufbau

MP3s werden in kleinere Einheiten, die als Frames bezeichnet werden geteilt. Jedes Frame besteht aus einem Header und den Daten. Im Header werden die Bitrate sowie andere Eigenschaften angegeben.



Am Anfang oder Ende des MP3s können im sogenannten ID3-Tag Informationen zum Namen, Autor und anderen musikspezifischen Daten angegeben werden.

Einige Player können bereits diese ID3 Tags während dem Stream verwenden, daher werden Informationen z.B. bei einer Radiopräsentation immer aktuell gehalten.

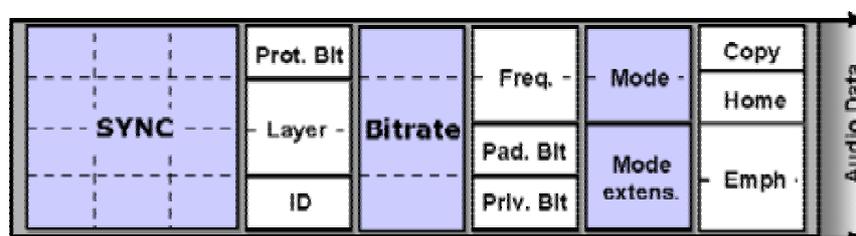
Jedes Frame enthält eine feste Anzahl von Abtastwerten (Samples). Bei Level 3 der MPEG Kompression sind dies 1152 Abtastwerte pro Frame (32 Subbänder x 36 Samples).

Optional kann man nach dem Header eine Fehlererkennung basierend auf den Cyclic Redundancy Codes (CRC) mit Größe 16 Bit kodieren. Diese schützt die wichtigsten Informationen im Header und in der Seiteninformation, ohne die das ganze Frame auf jeden Fall nicht ausgelesen werden kann.

Genauso wie es für Filme einen festgelegten Standard gibt, wie viele Frames pro Sekunde gespielt werden sollen, gibt es auch bei MP3-Spezifikation eine Richtlinie. Egal wie groß die Bitrate ist, ein Frame enthält immer 26ms Audiodaten.

Der Header:

Der Anfang eines MP3 Frames kennzeichnet ein 32Bit langer Header, der mit einem 11Bit sync Block beginnt. Dieser "sync block" ist wichtig, um beispielsweise Daten am Anfang zu ignorieren, oder während der Wiedergabe in der Aufnahme vor- bzw. zurückspringen zu können.



Nach dem Sync Block folgt das ID Bit, welches definiert, ob das Frame in MPEG-1 oder MPEG-2 codiert wurde. Die nächsten 2 Bit legen fest, in welchem Layer sich die Daten befinden. Ist das Protection Bit nicht gesetzt, folgt dem Header eine 16-Bit lange Checksumme. Das Bitrate-Feld enthält Angaben zur Bitrate, mit der das entsprechende Frame encoded wurde. Das Mode Field legt den stereo/mono Status des Frames fest. Sollte Joint-Stereo eingestellt sein, befinden sich weitere Informationen dazu in der Mode-Extension. Das Copyright Bit legt fest ob es "offiziell" erlaubt ist die Datei zu kopieren. Das Private Bit ist nicht weiter definiert, und kann von Applikationen für eigene Zwecke verwendet werden.

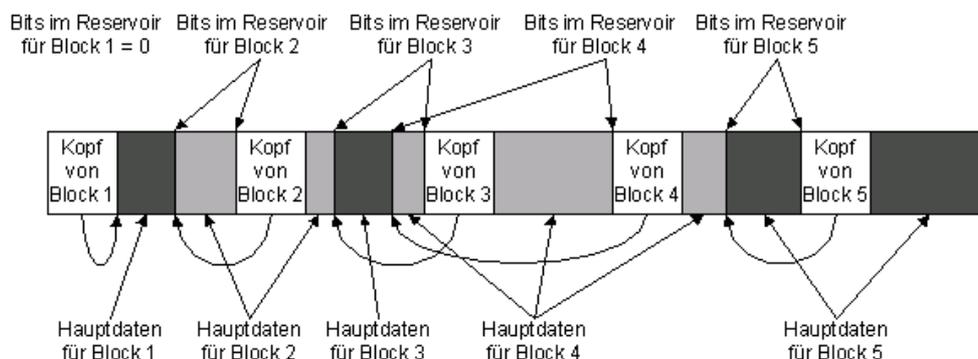
Die Hauptdaten (also die eigentliche Audio Signale) sind Huffman-kodiert und enthalten die Skalierungsfaktoren und die quantisierten Samples. Es besteht die Möglichkeit, zusätzliche Daten in den Datenstrom einzubinden (z.B. zusätzliche Audiokanäle oder Liedtexte), die vom Dekodierer ignoriert werden, wenn er nichts mit ihnen anzufangen weiß.

Für ein Audiosignal, das mit einer Abtastfrequenz von 44,1 kHz aufgezeichnet wurde und durch Kompression auf eine Datenrate von 128 kBit/s reduziert werden soll, ergibt sich pro Datenblock eine Größe von:

$$1152 \text{ (Samples/Block)} \times 128000 \text{ (Bits/s)} / 44100 \text{ (Samples/s)} = 3344 \text{ (Bits/Block)}$$

Benötigt ein Block weniger als diese Anzahl an Bits, so werden die übrigen Bits an das sogenannte "Bit-Reservoir" übergeben. Lässt sich umgekehrt ein Block nicht ohne hörbaren Qualitätsverlust mit der vorgegebenen Blockgröße kodieren, so können Bits aus dem Bit-Reservoir entnommen werden, die zusätzlich zur Kodierung des Blocks verwendet werden. Die Blöcke, die nicht die ganze Blockgröße benötigen, werden mit Daten der nächsten Blöcke aufgefüllt. Es dürfen jedoch zu keiner Zeit mehr Bits entnommen werden als im Bit-Reservoir vorhanden sind.

Ein MP3-Datenstrom beginnt immer mit einem Kopf. Danach folgen die einzelnen Datenblöcke.



Ausschnitt eines MP3-Datenstroms

6.2 Qualitätsstufen

Auf einer Audio-CD kommen rund 10 MB auf eine Minute Musik, MP3-Dateien in guter Qualität beanspruchen rund ein Zehntel dieses Speichers; man kann also mit MP3 etwa 10 Stunden Musik auf einer CD unterbringen.

MP3 Dateien können in verschiedenen Qualitätsstufen erzeugt werden, diese reichen von 8 kbit/s bis 320 kbit/s. Je höher dieser Wert ist, desto mehr Information steht zur Verfügung und umso besser kann der ursprüngliche Klang wiedergegeben werden.

32 kbit/s entsprechen ungefähr dem Hörerlebnis eines Mittelwellenradios, während 96 kbit/s bereits dem Klang eines UKW Radios übertreffen. Ab 128 kbit/s spricht man von CD naher Qualität, während ab 192 kbit/s kaum ein Unterschied zum Original auszumachen ist.

Klangqualität	Modus	Bitrate	Kompressionsrate
Telefon	mono	8 KBit/s	96:1
besser als KW-Radio	mono	16 KBit/s	48:1
besser als MW-Radio	mono	32 KBit/s	24:1
ähnlich UKW-Radio	stereo	56 KBit/s	26 bis 24:1
annähernd CD	stereo	96 KBit/s	16:1
CD-Qualität	stereo	112 bis 128 KBit/s	14 bis 12:1

Die Qualitätsstufen von MP3 im Vergleich

6.3 Vorteile des MP3-Formats

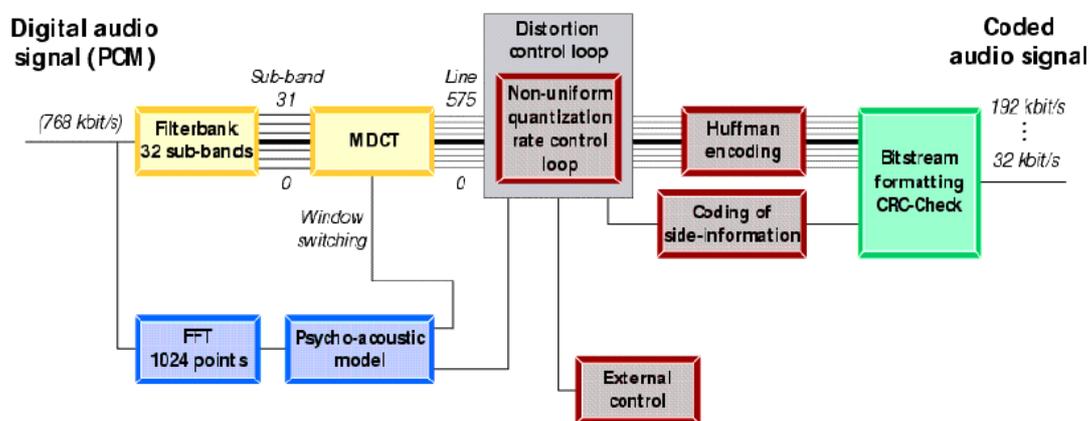
Das MP3 Format hat viele Vorteile:

- Die winzige Dateigröße. Man hat die Möglichkeit, ein sehr großes Musikarchiv auf der Festplatte zu haben.
- MP3 stellt die absolute Verfügbarkeit von Musik dar. Mit einem Mausklick kann man genau das hören, was man will. Lästiges Suchen nach der richtigen CD entfällt.
- Einfache Handhabung und Austausch von MP3s, zum Beispiel die Weitergabe an einem Freund, denn dies ist ja noch im legalen Rahmen.
- Neue Verkaufsstrategien sind möglich, zum Beispiel Verkauf von aktuellen Tracks direkt über das Internet, was für den Nutzer natürlich sehr bequem ist.

7. Kompression

Bei der Erstellung einer MP3-Datei kommen zwei verschiedene Verfahren zum Einsatz, um die hohe Kompression zu erreichen. Im ersten Abschnitt werden die Audiodaten mit Hilfe des psychoakustischen Modelles vereinfacht, in Phase Zwei findet eine Komprimierung nach den Parametern der Huffman-Codierung statt. Die grundlegenden Arbeitsschritte während des Codierungsvorganges sind folgende:

- Die Datei wird in kleinere Einheiten, die als Frames bezeichnet werden, aufgeteilt.
- Der zweite Schritt ist „die Quantizierung und Generierung der Datei“. Es wird determiniert, wie die Bits im Spektrum am geeignetsten zu verteilen sind, um die bestmögliche Soundqualität zu gewährleisten. Während dieses Prozess werden die Daten in Sub-Bänder unterteilt, die separat verarbeitet.
- Der Bitstream wird mittels Huffman-Coding komprimiert. Dieses Verfahren ist verlustfrei.
- Die Daten werden in Frames ausgegeben. Jedem Frame wird ein Header vorangestellt, der die wichtigen Informationen über das Frame enthält.

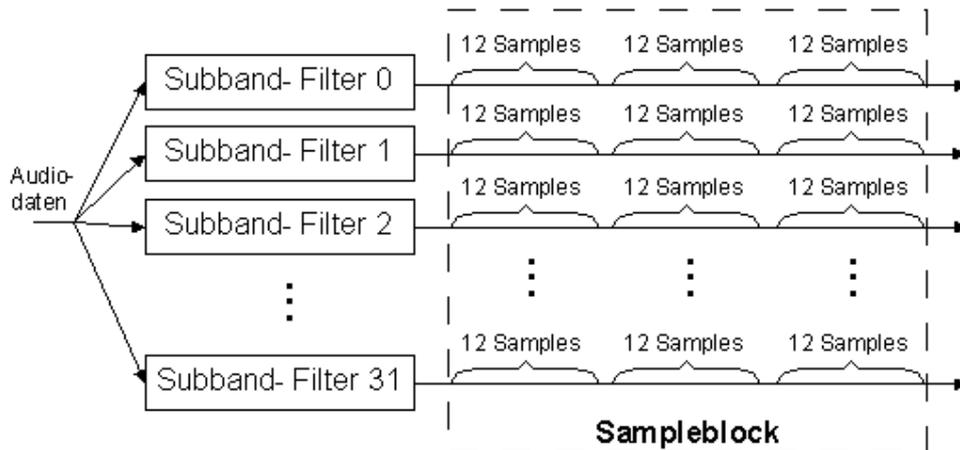


Aufbau eines MPEG Layer 3-Encoders

In der Grafik ist der Kompressionsprozess nochmals in einer anderen Form dargestellt, um die einzelnen Abschnitte im Detail zu veranschaulichen.

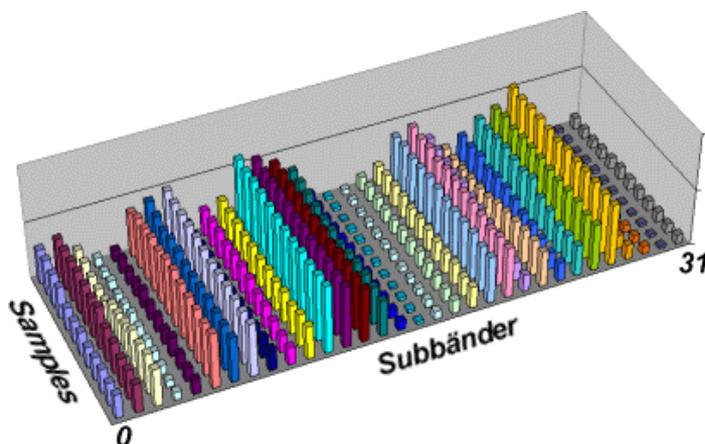
7.1 Filterbank

Die Filterbank transformiert das zu kodierende Audiosignal in 32 gleichbreite Subbänder, die zu jeweils einem Filter gehören. Diese Subbänder sollen die kritischen Bänder approximieren.



Für 32 eingelesene Samples wird pro Subband ein Sample ausgegeben. Eine unerwünschte Eigenschaft der Filterbank ist das sogenannte "Aliasing": Die 32 Subbänder überlappen sich, und so kann ein Ton einer festen Frequenz zwei Bänder beeinflussen.

Die Subbänder werden durch eine modifizierte diskrete Cosinus-Transformation (MDCT) jeweils nochmals in 18 Teilbereiche unterteilt. Dadurch ergibt sich eine höhere Spektralauflösung, außerdem ist es nun möglich, den Aliasingeffekt teilweise zu entfernen.



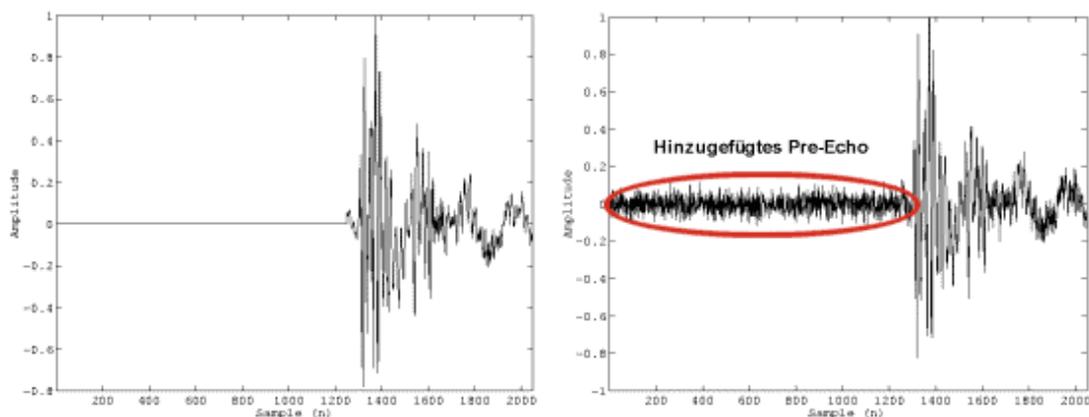
**Ausgabe der Filterbank
32 Subbänder mit Zeitsamples**

Die modifizierte Diskrete Cosinus-Transformation Filter (MDCT):

Die modifizierte diskrete Cosinustransformation (MDCT) ist eine Zeit-Frequenz-Transformation, ähnlich der diskreten Fouriertransformation. Im Gegensatz zur normalen diskreten Cosinustransformation (DCT) wie sie z.B. bei der JPEG-Kompression zum Einsatz kommt, arbeitet die MDCT nicht auf zweidimensionalen, sondern nur auf eindimensionalen Daten.

Die Frequenzauflösung wird durch die MDCT deutlich verbessert, die Zeitauflösung wird aber entsprechend schlechter, da das erhaltene Frequenzspektrum nun die Zeitspanne aller 1152 Eingangssamples repräsentiert.

Die Verschlechterung der Zeitauflösung hat zur Folge, dass ein Pre-Echo-Effekt auftreten kann. Als Pre-Echo bezeichnet man solche Geräusche, die vor dem eigentlichen Signal entstehen. Beobachten kann man es bei solchen Signalen, die mit einer großen, plötzlichen Lautstärkenänderung beginnen, insbesondere wenn ein Bereich mit sehr niedrigen Amplituden vorausgeht.

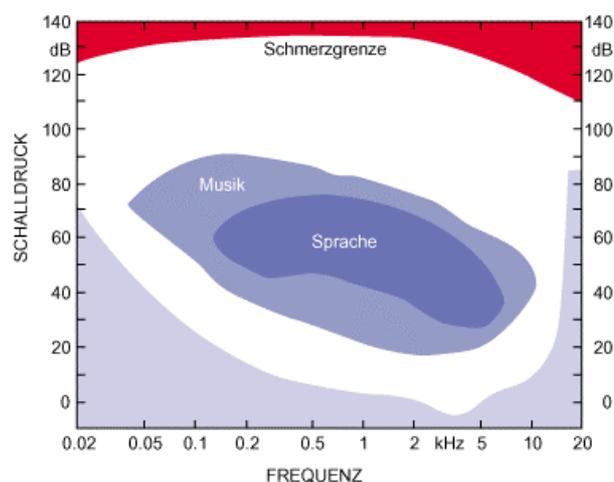


Im linken Bild von Abbildung sieht man das unkodierte Audiosignal. Im rechten Bild wurde dieses mit einem Transform Coder (d.h. basierend auf einer diskreten Zeit-Frequenztransformation wie bei MPEG Layer 3) kodiert. Man kann deutlich die eingefügten Geräusche vor dem eigentlichen Signal erkennen.

Um dem Pre-Echo-Effekt vorzubeugen, werden bei der MDCT unterschiedliche Transformationsblocklängen gewählt. Lange Blöcke werden bei stationären, sich also zeitlich wenig ändernden Signalen eingesetzt, um die Kompression zu erhöhen. Kurze Blöcke werden bei sich schnell ändernden Signalen verwendet, um den Bereich, in dem Pre-Echo auftreten kann zu verkleinern. Die Entscheidung darüber, welcher Typ von Signal vorliegt, liefert das psychoakustische Modell.

7.2 Das psychoakustische Modell

Bei der MP3-Kompression werden jegliche Daten, die das Gehör gar nicht wahrnimmt, weggelassen, um eine hohe Kompressionsrate zu erreichen. Das psychoakustische Modell ist eine Art Tabelle, mit deren Hilfe der Encoder jegliche Art vom Maskierung erkennen kann. Somit ist entscheidbar, welche Töne ausgespart werden können. Das psychoakustische Modell ist entscheidend für die Qualität des Encoders. Um es zu entwickeln, sind aufwendige Hörtests notwendig. Da der Encoder von der ISO nicht standardisiert ist (und somit auch das psychoakustische Modell nicht), kommt es zwischen verschiedenen Encodern zu Qualitätsunterschieden.



Die Wahrnehmbarkeitsbereiche des menschlichen Gehöres

Das psychoakustische Modell beschreibt jene akustischen Eigenschaften des menschlichen Gehörs, die es bei der MP3-Codierung ermöglichen, Audiodaten mit einer hohen Rate zu komprimieren. Die Tatsache, dass der menschliche Gehörsinn unter gewissen Umständen Signale bzw. Teile davon nur vermindert bzw. gar nicht wahrnehmen kann, ist die Basis für diese Überlegungen.

Die Eigenschaften für die MP3-Kompression sind beispielsweise, dass von 2 schnell aufeinanderfolgenden Geräuschen, die eine ähnliche Klangfarbe haben, nur das Stärkere wahrgenommen werden kann.

Frequenzen unter 20 Hz und über 16 kHz können nur schwach bzw. gar nicht erfasst werden. Diese und andere Analysen des Gehörs, welche allgemein als Psychoakustik bezeichnet werden, können mathematisch beschrieben und in Tabellen zusammengefasst werden, um ein Psychoakustisches Modell darzustellen.

7.3 Huffman-Kodierung

Eine andere wichtige Teil des MP3-Encoders ist die Huffman-Kodierung. Durch die Quantisierung wird die Auflösung eines Signals verringert. Die nach der Quantisierung erhaltenen Werte werden anschließend mittels Huffman-Verfahren kodiert. Das Huffman-Verfahren hat den Vorteil, das es redundante Daten verlustfrei entfernen kann. Und das ist einer der Gründe, wieso MPEG Layer 3 hohe Audioqualität bei kleinen Bitraten erzielen kann.

7.4 Kompression Layer 3

In Bezug auf MPEG1-Layer III geht die Codierung noch etwas mehr ins Detail und weist somit eine komplexere Struktur auf als Layer I und II. Die Idee dahinter ist, den Datenstrom so gering wie möglich zu halten, was allerdings zu einer längeren Encodierungszeit bei der Erstellung der Datei führt. Vom Algorithmus gesehen wird das Verfahren um folgende Schritte erweitert:

- a. Neben der FFT kommt noch ein andere Filter, die Diskrete Cosinus-Transformation Filter (DCT) zum Einsatz. Mit dieser Methode kann ein eine bessere Performance und mehr Zuverlässigkeit des Psychoakustischen Modells erreicht werden.
- b. In der letzten Phase, in der die Frames vorbereitet werden, ist ein komplexer Algorithmus hinzugefügt, der die Framegröße nach bestimmten Parametern variieren lässt. Es wird auch das sogenannte Reservoir Bit gesetzt, dass eventuell auftretende kritische Verzerrungen bei der Encodierung des Samples löst.

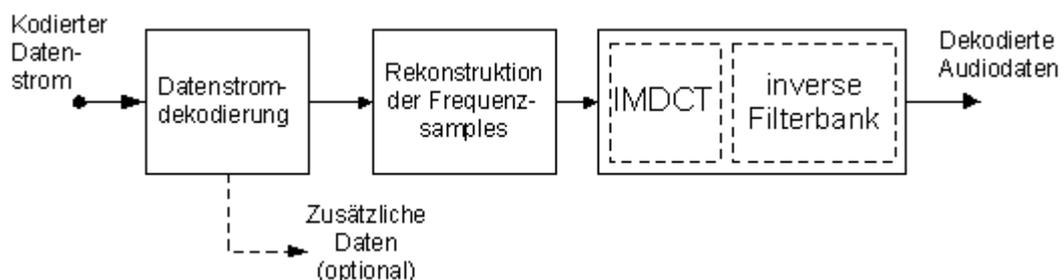
Kompressionsverfahren	Kompressionsverhältnis	Datenrate
CD-Audio (unkomprimiert)	1:1	ca. 1,4 MBit/s
MPEG-1 Layer I	1:4	384 kBit/s
MPEG-1 Layer II	1:6...1:8	256...192 kBit/s
MPEG-1 Layer III	1:10...1:12	128...112 kBit/s

Vergleich der Kompressionsraten

8. Die Player und Decoder

8.1 Decoder

Der Aufbau des Decoders ergibt sich direkt aus dem des Encoders. Im Decoder werden alle Schritte in umgekehrter Reihenfolge ausgeführt. Zuerst wird der MPEG Layer 3-Bitstrom dekodiert und die Seiteninformationen für die inverse Quantisierung gewonnen. Nach der Rückgängigmachung der Quantisierung werden mittels inverser modifizierter diskreter Cosinustransformation (IMDCT) und anschließender Filterbanksynthese wieder die ursprünglichen Samples im Zeitbereich berechnet. Der Decoder kann sehr gut auf Übertragungsfehler reagieren, so dass bei Bit-Fehlern in einzelnen Teilen die gesamte Datei oder der Stream weiterhin lesbar bleibt.



8.2 Software

Um MP3 Dateien zu erzeugen, bedarf es einer Software, die Audiodaten von einer CD ausliest (Ripper) und diese dann ins MP3 Format umwandelt. Hierzu kann es sein, dass die Software den Zwischenschritt über die bekannten Windows Wave Dateien geht oder dass sie den Datenstrom direkt in MP3 Dateien schreibt. Für das Endergebnis ist dies letztendlich unerheblich.

Eine leistungsstarke Software zur Erzeugung und Verwaltung von MP3 Dateien ist Music Match. Die All-In-One Lösung ist Ripper, Encoder, Abspiel- und Brennsoftware in einem. Music Match basiert auf dem MP3 Codec des Fraunhofer Instituts. Aber auch Microsofts WMA kann Music Match abspielen. Weitere Features sind eine Musik Archivverwaltung, Anschluss an eine Internetdatenbank, um Musikstücke automatisch zu benennen, direkter Dateizugriff auf portable MP3 Player uvm.

Ein alternativer Ripper ist das Programm Exact Audio Copy (EAC), welches besonders leistungsfähig im Auslesen von Audio CDs ist. Es unterstützt externe MP3 Codecs, wie z.B. Lame, die leicht eingebunden werden können. Insofern kann die Kombination EAC/Lame eine bessere Qualität erzeugen, als die Music Match Software.

8.3 Player

Zum Abspielen, also dem Dekodieren, werden spezielle Player benötigt. Wie bei den Encodern, gibt es eine Vielzahl an Player. Eine der beliebtesten ist der WinAmp Player. Er bietet neben austauschbaren Oberflächen (sogenannte Skins), auch noch eine spezielle Unterstützung der MMX- und 3D-Now-Technologie.

9. MP3-PRO

MP3 PRO wurde von der Firma Coding Technologies entwickelt. MP3 PRO ist kein eigenständiges Audioformat, sondern eine Kombination aus MP3 und dem SBR (Spectral Band Replication). Es ermöglicht mit kleinem zusätzlichem Datenoverhead die verlorenen hohen Frequenzen zu restaurieren. Diese zusätzlichen Daten können in das bestehende MP3-Format eingebettet werden, so dass ein spezieller MP3 PRO Dekoder die Dateien in besserer Qualität wiedergeben kann, während ein normaler Layer III Dekoder die Zusatzdaten ignoriert. Das Format verspricht gute Audioqualität bei einer Datenrate von 64 kBits/s.