

# Proseminar Multimedia Datenbanken & Retrieval

## Audio Coding

Paul Kobold  
Universität Koblenz-Landau  
D-56070 Koblenz, Universitätsstraße 1  
pskobold@uni-koblenz.de

15. Februar 2010

**Zusammenfassung.** Die digitale Codierung von Audiosignalen mit geringer Bitrate, auch Audiokompression genannt, ist seit Ende der 80er zu einem wichtigen Gebiet der Audiosignalverarbeitung geworden. In dieser Ausarbeitung möchte ich einen allgemeinen Einstieg in die Thematik geben. Nach der Einführung stelle ich kurz das menschliche Hören vor und wie analoge Signale digital codiert bzw. repräsentiert werden. Dann folgt eine Übersicht zum Thema Audiokompression, die Vorstellung von ein paar Basistechniken, um dann anhand des MP3 Formates die Vorgehensweise eines konkreten Audiocodecs darzustellen. Anschließend möchte ich kurz ein paar Einsatzgebiete von Audiocodecs abseits des PCs vorstellen, sowie ein wenig die aktuelle Bedeutung von Audiokompression diskutieren.

## 1 Einleitung

Die unkomprimierte Darstellung digitaler Audiodaten in idealer Form (d.h. Abbildung des menschlichen Hörspektrums, Dynamikumfang von 100 dB), <wie z.B. auf einer CD, benötigt eine Datenrate von rund 700 kbit/s pro Kanal, was zu einer Datenrate einer Stereo Aufnahme von rund 10 Megabyte pro Minute führt.

Die Speicherung und Übertragung größerer Mengen solcher Audiodaten ist nicht nur wirtschaftlich unsinnig, sondern für gewisse Anwendungszwecke auch technisch unmöglich.

So würde beispielsweise eine 5.1 Aufnahme, wie z.B. auf einer Film-DVD, fast 2 Gigabyte pro Stunde beanspruchen, so dass für die Videodaten kaum mehr Platz vorhanden wäre. Vor dem Aufkommen von Breitbandanschlüssen bzw. -netzwerken war auch die Übertragung und das Abspielen in Echtzeit solcher Datenmengen, sogenanntes Streaming, unmöglich. Aber auch heute wären Massenübertragungen, von unkomprimierten Audiodaten wie bei Musik- oder Videoportalen im Internet, nicht möglich oder rentabel.

Um eine wirtschaftlich, wie technisch vertretbare Übertragungsrate bei hoher Übertragungsqualität zu erreichen, wurden also zusätzliche Codierungen notwendig. Viele Anwendungen, wie digitaler Rundfunk, digitales Fernsehen oder Musik über das Internet, sind erst durch moderne Datenreduktionsverfahren möglich geworden. Das gleiche gilt natürlich sinngemäß auch für Videodaten.

Aber nicht nur bei der Übertragung, sondern auch im privaten Anwenderbereich spielen moderne Codierungsverfahren eine Rolle. Durch den Erfolg der MP3 Codierung und immer steigender Kapazität der Speichermedien, können mittlerweile tausende CDs umfassende Musiksammlungen in der Hosentasche transportiert und mit einer Vielzahl an Geräten in verschiedenen Bereichen (im Auto, am Computer, unterwegs usw.) abgespielt werden.

## **2 Von Analog zu Digital**

Zuallererst wollen wir uns mit den Eigenschaften des menschlichen Hörens beschäftigen und danach zeigen wie analoge Audiosignale in digitale umgewandelt werden können.

Der Mensch verarbeitet Schallwellen mit Frequenzen von ungefähr 16 bis 20000 Hz, wobei gerade im hohen Frequenzbereich dieser Wert von Mensch zu Mensch verschieden ist. Auch verschlechtert sich dieser Wert im Laufe des Alters, im allgemeinen geht man davon aus, dass die obere Hörschwelle im mittleren Alter bereits nur noch ungefähr 16 kHz beträgt und ab 60 bereits um 10 kHz gemindert sein kann. Die größte Hörempfindlichkeit besitzt der Mensch im Bereich zwischen 1000 und 5000 Hz.

Das menschliche Innenohr misst die Schallwellen und gibt entsprechende Signale an das Gehirn weiter. Frequenzen beschreiben dabei die Tonlage, während der Druck, die Intensität der Welle, in Form von Lautstärke interpretiert wird. Zu hoher Druck und damit Lautstärke kann dabei bleibende Schäden im Ohr verursachen.

Eine Annäherung an dieses menschliche Hörempfinden ist der sogenannte Schalldruck-pegel, gemessen in Dezibel (dB), wobei dies eine logarithmische Größe ist und eine lineare Erhöhung nicht mit einer linearen Erhöhung der wahrgenommenen Lautstärke einhergeht. Im allgemeinen sagt man, dass eine Erhöhung des Schalldruckpegels um 10 dB eine Verdoppelung der wahrgenommenen Lautstärke ist.

Die wahrgenommene Lautstärke ist auch immer abhängig von Frequenz und Druck, sodass bestimmte Klänge unterschiedlicher Frequenz lauter oder leiser als andere wahrgenommen werden, obwohl sie physikalisch den gleichen Schalldruckpegel aufweisen.

Um nun analoge Signale, sei es in bereits vorliegender elektrischer Form durch eine

Quelle, z.B. in Form eines CD-Abspielgeräts über eine analoge Verbindung, oder durch Wandlung von Schallwellen in elektrische Signale, z.B. durch ein Mikrofon, durch einen Computer bearbeiten zu können, müssen sie digitalisiert werden.

Für Audiosignale benutzt man dazu gewöhnlich die Pulse Code Modulation (PCM). Dabei wird n-mal pro Sekunde das Signal gemessen und in ein Datenwort quantisiert, wobei sich hier Wortbreiten(auch Bittiefen genannt) von 16 Bit durchgesetzt haben. Ein einzelner gemessener Wert wird als Sample bezeichnet. Die Häufigkeit mit der das Signal pro Sekunde gemessen wird, bezeichnet man als Abtastrate oder auch Samplerate und wird ebenso in Hz angegeben.

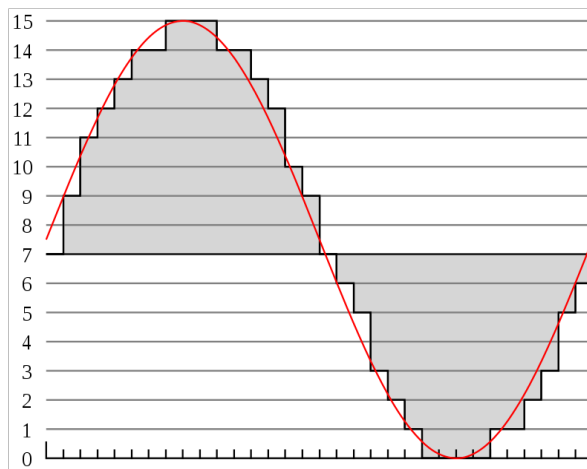


Abbildung 1: Abtastung und Quantisierung eines Signales

Nach dem Shannon-Nyquist Theorem muss die Samplerate mindestens das zweifache der höchsten abzubildenden Frequenz des abzutastenden Signales sein, damit eine Rekonstruktion des Ursprungssignal ohne Informationsverlust möglich ist. Der Aufwand für eine vollständig korrekte mathematische Rekonstruktion ist zwar unendlich groß, lässt sich aber ziemlich gut annähern, so dass man leicht aus den digitalen Werten wieder ein analoges Signal erstellen kann, das dem Ursprungssignal sehr nahe kommt (Abb. 2).

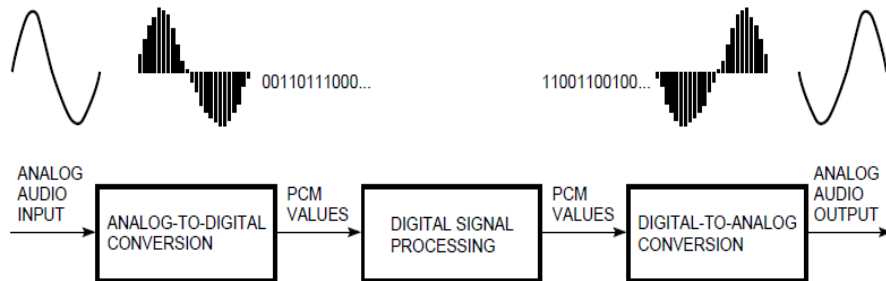


Abbildung 2: Analog-Digital-Analog-Verarbeitungskette [1]

Die CD benutzt z.B. eine Samplerate von 44,1 kHz bei 16 Bit Wortbreite mit 2 Kanälen. So ergeben sich auch die in der Einführung erwähnten Datenraten ( $44100 * 16 * 2 = 1,4 \text{ Mbit/s}$ ). Das abgebildete Hörspektrum umfasst dabei den Bereich 5 Hz bis 20 kHz. Nach dem oben erfassten Theorem sollte dafür zwar eine Abtastrate von 40 kHz reichen, aber da man mittels Tiefpassfilter unerwünschte höhere Frequenzen abschneidet und es keinen idealen Tiefpassfilter gibt, erlauben höhere Abtastraten besser mit Alias-Signalen (Stör- oder Pseudosignale) umzugehen, die sich als störende Frequenzanteile bemerkbar machen würden. So werden auch weitaus höhere Abtastfrequenzen verwendet, z.B. bis zu 192 kHz bei der Audio-DVD.

Quellen [1, 2, 6, 7]

### 3 Kompression von Audiodaten

#### 3.1 Allgemeine Grundlagen

Wie wir bereits in der Einführung gesehen haben, besitzen PCM codierte Signale in guter Qualität eine sehr hohe Datenrate, die zu sehr großen Dateien führt. Diese möchte man natürlich gerne verkleinern und das möglichst ohne Qualitätsverlust. Klassische verlustfreie Datenkompressionsformate, wie z.B. „Zip“ oder „RAR“, verstehen sich hauptsächlich auf die Beseitigung von Redundanzen bzw. impliziert gegebenen Informationen. Da diese bei Audiodaten aber eher selten vorhanden sind, erzielt man mit diesen Techniken auch kaum eine Verringerung der Dateigröße, daher müssen andere Ansätze her.

Zunächst sollte man zwischen Datenkompression und reiner Datenreduktion unterscheiden.

Eine Datenreduktion kann man z.B. durch eine niedrigere Samplerate, Bittiefen oder Kanalanzahl erreichen. Codiert man z.B. die Daten einer Audio-CD neu in Mono, 8 Bit und einer Samplerate von 8 kHz, so ist die resultierende Datengröße gerade mal

ein 22tel der Ursprungsgröße. Allerdings hat man dafür auch große Qualitätseinbußen. So hat das genannte Resultat die Qualität eines Gesprächs über das analoge Telefonnetz, das eine Bandbreite von 3,1 kHz für Gespräche hat. Das so reduzierte Audiodaten zwar anders bzw. schlechter, aber immer noch ähnlich wie das Ursprungsmaterial klingen, liegt daran, dass (natürliche) Töne nicht nur aus einzelnen Frequenzen bestehen, sondern sich aus dem Grundton und den sogenannten Obertönen, die Vielfache der Grundtonfrequenz sind, zusammensetzen. Fallen dann Frequenzen weg, ändert sich der Klang, aber die grundsätzliche Melodie/Sprache bleibt erhalten. Durch weniger Samples und einer geringeren Bittiefe ergibt sich auch ein höheres Quantisierungsrauschen, dies führt zu einem niedrigeren Signal-Rausch-Abstand, was letztendlich zu einem stärker wahrgenommen hörbaren Rauschen führt.

Da man diesen Qualitätsverlust natürlich nur in bestimmten Anwendungsfällen akzeptiert bzw. als Kompromiss einget, entwickelten sich seit den 80ern verschiedene Verfahren zur Komprimierung von digitalen Audiodaten, so dass heute eine Vielzahl an unterschiedlichen Codierungsformaten und Verfahren (engl. Codecs) mit verschiedenen Ansätzen und Einsatzzwecken existieren.

Dabei geht es nicht nur um die höchst mögliche Effizienz, also die kleinste Dateigröße bei gleich bleibender Qualität, sondern die Anforderung an Encoder und Decoder hinsichtlich der Komplexität und dem Ressourcenbedarf unterscheidet sich je nach Anwendungsgebiet. Z.B. bei Codecs die zum direkten Abspielen der komprimierten Daten gedacht sind, muss die Decodierung auch möglichst effizient in Software oder Low-Cost Hardware umsetzbar sein, während der Encoder nicht echtzeitfähig sein muss. Bei direkter Aufnahme, Broadcasting und Wiedergabe, z.B. beim digitalen Funk, müssen dagegen Encoder und Decoder echtzeitfähig sein, während zur reinen Archivierung beide nicht echtzeitfähig sein müssen.

Weitere Anforderungsprofile und Verfahrensunterschiede für Codecs sind z.B. Streaming-Unterstützung, Unterstützung von Mehrkanal-Signalen, Anspringen beliebiger Positionen in einem Audiostrom, Soft- und Hardwareunterstützung, Fehlertoleranz und Lizenzarten.

Auch die Art der Tonquelle kann eine Rolle für ein bestimmtes Verfahren spielen. So versucht man z.B. bei der Sprachkompression die Tonquelle zu modellieren. Generell stellt man an die Komprimierung von Sprache, auf Grund ihrer physikalischen Charakteristik und der Wahrnehmung des Menschen, einen geringeren Anspruch an die reine digitale Qualität (Samplerate, Bitrate). So gibt es Codecs, die mit sehr geringen Datenraten von 2 - 5 kbit/s (also ungefähr 1:300 gegenüber der CD-Aufnahme) noch brauchbare Ergebnisse erzielen.

Für generelle Audiokompression möchte man aber keine Annahmen über die Tonquelle treffen, sondern man setzt am anderen Ende an – dem Gehör. Man unterscheidet zunächst zwischen verlustfreien (engl. lossless) und verlustbehafteten (engl. lossy) Codecs.

Verlustfrei bedeutet, dass aus der codierten Datei sich eine bit-genaue Rekonstruktion des Ursprungsmaterials erstellen lässt.

Heutige Verfahren erreichen dabei eine Kompressionsrate zwischen 25 bis 70 % bei Ausgangsmaterial in CD Qualität, abhängig von dessen Komplexität.

Angewendet werden sie vor allem im professionellem Tonbearbeitungsbereich, neueren Datenträgern wie der SACD oder DVD-Audio, aber auch im Privatbereich finden sie unter qualitätsbewussten Hörern immer mehr Einsatz.

Die angewendeten Techniken machen sich dabei Ähnlichkeiten von Kanälen (s. Abschnitt 3.2), Abhängigkeiten von aufeinander folgenden Abtastwerten, sowie der Entropie der so gewonnenen bzw. transformierten Daten zunutze.

Verlustbehaftete Komprimierung versucht dagegen, neben der Verwendung der Techniken zur verlustfreien Komprimierung, durch Beseitigung von Irrelevanzen eine Datenreduktion ohne hörbaren Qualitätsverlust vorzunehmen. Hierbei macht man sich die Eigenschaften des menschlichen Hörens durch ein psychoakustisches Modell zu Nutzen, um Informationen, die der Hörer kaum bzw. gar nicht wahrnehmen würde, mit geringerer Präzision zu speichern oder ganz wegzulassen.

So lässt sich bereits eine Kompression von Audiodaten in CD Qualität um den Faktor 1:12 ohne einen hörbaren Qualitätsverlust durchführen. Je nach geforderter Qualität lässt sich die Kompressionsrate weiter steigern, dann allerdings mit wahrnehmbaren Verlusten. Als Indikator für die Qualität der komprimierten Daten werden daher meist die Bitraten angegeben (z.B. Bei mp3 Dateien entsprechen 192 kbit/s CD-, 96 kbit/s Radio-, 32 kbit/s Telefonqualität).

Es gibt auch Hybrid Codierungen, die eine verlustbehaftete Kompression durchführen, aber die Korrekturinformationen, um dennoch eine verlustfreie Rekonstruktion ermöglichen, innerhalb der Datei speichern. Das „mp3HD“ Format z.B. ermöglicht dadurch Abwärtskompatibilität, in dem es von Abspielgeräten, die das Format nicht unterstützen, als normale MP3 Datei interpretiert und abgespielt wird.[8]

### **3.2 Techniken der Audiodatenkompression**

An dieser Stelle wollen wir ein paar Grundtechniken näher betrachten.

#### **3.2.1 Joint-Stereo-Verfahren**

Als Joint-Stereo Verfahren werden 2 Algorithmen beschrieben, die sich allerdings grundsätzlich unterscheiden, Intensity Stereo und Mid/Side Stereo (MS-Stereo).

Im Audiocoding-Bereich ist aber im Zusammenhang mit Joint-Stereo oft nur MS-Stereo gemeint.

Beim Intensity Stereo Verfahren, einem Verfahren aus der Intensitätsstereofonie, wird ein Monokanal, sowie die Richtungsinformationen zur Rekonstruktion der Stereodaten gespeichert. Dieses Verfahren ist allerdings verlustbehaftet, da dabei Phaseninformationen verloren gehen. Die so eventuell auftretende Phasenverschiebung ist allerdings oberhalb einer gewissen Frequenz nicht mehr wahrnehmbar, so dass sich dieses Verfahren für niedrigere Datenraten, bei denen man dennoch eine Stereoinformation haben möchte, dennoch eignet.

Beim MS-Stereo Verfahren speichert man den Mittelwert des linken und rechten Kanals, sowie die Differenzinformation der beiden Kanäle jeweils in einem eigenen Kanal, Mittel- und Seitenkanal genannt,

$$MK = \frac{LK + RK}{2}$$
$$SK = \frac{LK - RK}{2}$$

so dass sich die ursprünglichen Kanäle wieder verlustfrei rekonstruieren lassen:

$$LK = MK + SK$$
$$RK = MK - SK$$

Der entscheidende Vorteil ist, dass der Differenzwert oft sehr klein oder 0 ist und sich dadurch sehr gut komprimieren lässt bzw. man dem Mittelkanal eine höhere Datenrate zuweisen kann. Dadurch lässt sich eine höhere Kompressionsrate bei gleich bleibender Hörqualität und umgekehrt erreichen. Außerdem können Monoabspielgeräte auch einfach nur den Mittelkanal abspielen ohne das es zu Klangverfärbungen kommt.

Das Verfahren wird z.B. auch beim FM-Stereo im Rundfunk verwendet.

### 3.2.2 Linear Predictive Coding

Linear Predictive Coding (LPC) ist ein Verfahren, das nach der Methode der linearen Vorhersage, einem mathematischen Verfahren der Zeitreihenanalyse, vorgeht.

Dabei wird basierend auf der linearen Verknüpfung vorhergehender Abtastwerte der nächste Abtastwert geschätzt und nur die Differenz dazu, sowie bei adaptiven Verfahren noch Parameter zur Beschreibung des verwendeten Vorhersagefilters übertragen. Je nach Qualität dieser Vorhersage bleibt ein Restsignal mit kleineren Werten zurück, die mit Entropiekodierungsverfahren effizient verpackt oder auch für weitere Datenreduktion quantisiert werden können.

Dieses Verfahren wird vor allem bei Sprach- und verlustfreien Kompressionen eingesetzt.

### 3.2.3 Perceptual Coding

Wie bereits erwähnt, versuchen die meisten verlustbehafteten Kompressionstechniken die subjektive Wahrnehmung des Menschen bzw. die verschiedenen Eigenschaften des menschlichen Gehörs wie Frequenzgruppenbildung, Hörbereichsgrenzen, Maskierungs-effekte und Signalverarbeitung des Innenohrs durch ein psychoakustisches Modell auszunutzen.

Da der Mensch nicht in allen Frequenzbereichen äquivalente Höreigenschaften besitzt, teilt man den Frequenzbereich z.B. in Subbänder auf, die unterschiedlich gewichtet werden. Auch macht man sich verschiedene Maskierungseffekte zu Nutzen. So verdecken laute Töne leisere, wie z.B. ein Presslufthammer die Umgebungsgeräusche. Das muss auch nicht exakt gleichzeitig geschehen, sondern frequenzabhängig kann es vor (bis ca. 50 ms) oder nach (bis ca. 200 ms) lauten Tönen zur Maskierung von leiseren kommen. Diese Informationen, die der Mensch nicht wahrnimmt, werden aber dennoch aufgezeichnet, so dass man hier ansetzt und diese versucht herauszurechnen.

Zur Analyse und Verarbeitung des psychoakustischen Modells und den Erkenntnissen der Signaltheorie, wird das Signal mit Hilfe von mathematischen Transformationen wie der Fourier Transformation oder der Diskreten Cosinus Transformation in den Frequenzraum transformiert und verarbeitet, wobei man im digitalen Audibereich optimierte Algorithmen wie die Fast Fourier Transformation (FFT) bzw. die modifizierte diskrete Cosinus Transformation (MDCT) verwendet.

Bei der anschließenden Quantisierung und Codierung der Werte ist dann das Ziel, so wenig Bits wie möglich zu benötigen, aber genug um das Quantisierungsrauschen unterhalb der Maskierungsschwelle zu halten, so dass es nicht wahrgenommen wird.

Anschließend können dann noch weitere Entropiekodierungsverfahren wie z.B. die Huffman-Codierung angewendet werden.

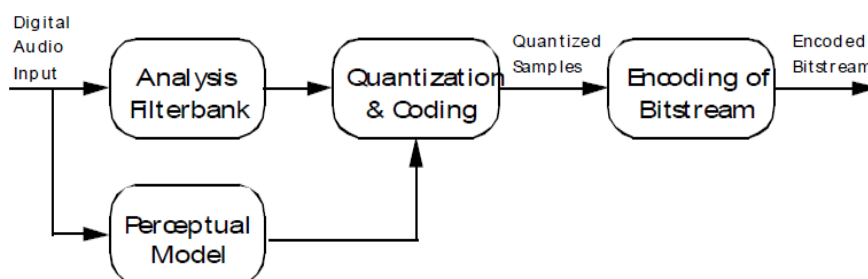


Abbildung 3: Blockdiagramm eines perzeptionellen Codierungssystems [3]

Quellen [1, 2, 3, 6, 7]



## 4 Ein Codierungsverfahren im Detail: MP3

Das MP3 Format, dessen Namen durch die offizielle Dateierweiterung für so codierte Audiodateien „mp3“ gegeben ist, ist ein Teilverfahren zur Audiocodierung im MPEG (Moving Picture Experts Group) 1 Standard und wird vollständig korrekt als ISO MPEG-1 Audio Layer 3 bezeichnet. Es basiert auf den Techniken der perzeptionellen Codierung und wurde in den 1980ern von einer Gruppe um Karlheinz Brandenburg und Harald Popp am Fraunhofer-Institut für Integrierte Schaltungen (FhG-IIS) in Erlangen, sowie an der Friedrich-Alexander-Universität Erlangen-Nürnberg in Zusammenarbeit mit AT&T Bell Labs und Thomson entwickelt. Seit 1992 ist es ISO (s. ISO/IEC IS 11172-3) standardisiert. Dies ist ein offener Standard, allerdings sind viele Techniken durch Patente geschützt und das Verwenden des Formates kostet Lizenzgebühren.

Allerdings ist nur der Decoder standardisiert, während der Encoder und damit das zugrundeliegende psychoakustische Modell dies nicht ist. Dessen Forschung und Entwicklung ist sehr aufwendig und kostenintensiv und man wollte es u.a. deswegen nicht offen legen, um so andere Lizenzmodelle gegenüber dem ISO Standard verwenden zu können.

Der entscheidende technische Vorteil daran ist, dass Weiterentwicklungen und Verbesserungen des psychoakustischen Modells unabhängig vom Decoder und damit den Abspielplattformen sind. So ergeben sich auch Klangunterschiede bei der Verwendung unterschiedlicher Encoder, bei ansonsten gleichen Einstellungen.

Der MPEG-1 Audiostandard umfasst drei Operationsmodi, als Layer bezeichnet. Layer 1, Layer 2 und Layer 3 unterscheiden sich dabei in Komplexität und Performance, wobei Layer 3 dabei die beste Qualität liefert, aber auch die höchste Rechenleistung benötigt.

Die Anforderungsprofile von unterschiedlichen Layern liegen in der damaligen Rechenleistung der Hardwarearchitekturen begründet.

Der Ansatz bei MPEG Audio ist nicht ein gegebenes Signal auf möglichst wenig Platz abzulegen, sondern eine gegebene Bandbreite optimal für das Signal auszunutzen.

Layer 3 zeigt dabei bereits zu damaligen Zeiten sehr gute Resulte und man erreicht bei einer Bitrate von 128 kbit/s für ein Stereo Signal bereits annähernd CD Qualität. Ab 192 kbit/s können die meisten Menschen in der Regel schon keinen Unterschied zur ursprünglichen Aufnahme ausmachen, wobei dies natürlich noch vom Abspielequipment abhängig ist.

Unterstützt werden Bitraten von 8 bis 320 kbit/s und Sampleraten von 32 kHz, 44.1 kHz und 48 kHz. Außerdem werden zwei Kanäle mit bis zu vier Optionen unterstützt:

- Single Channel – mono
- Dual Channel – 2-Kanalaufnahmen, zum Beispiel für unterschiedliche

Sprachen

- Stereo
- Joint-Stereo – sowohl Mid/Side- als auch Intensity Stereo werden unterstützt

Weiterhin werden sowohl konstante wie variable Bitraten unterstützt.

Unabhängig von der eingestellten Bitraten-Methode verwendet das Format ein Bit-Reservoir, also einen Puffer, der nicht notwendige Bits bei Passagen, die nicht die volle Bitrate ausnutzen würde, reserviert und sie bei Passagen verwendet, in denen die Bitrate nicht ausreicht. Allerdings ist dieses Reservoir nicht sehr groß und darf auch nur für eine sehr kurze Zeit verwendet werden.

Bei MP3 wird das Audiosignal in Frames kodiert, die jeweils für sich eine eigene Bitrate haben können. Für jeden Frame geht der Encoder folgendermaßen vor:

Jeder Frame besteht aus 1152 Samples des Audiosignals, die durch eine polyphase Filterbank laufen und in 32 Subbänder aufgeteilt werden. Parallel wird das Signal durch eine 1024-wertige FFT für die Bearbeitung durch das psychoakustische Modell vorbereitet. Das Modell analysiert die Daten und berechnet die Werte für die Maskierungskurven, unterhalb denen die Codierung von Daten nicht notwendig sind.

Die 32 Subbänder werden nun durch die Modifizierte Diskrete Cosinus-Transformation (MDCT), mit Hilfe der Werte des psychoakustischen Modells um vor allem Pre-Echos zu vermeiden, jeweils in weitere 18 Subbänder zerlegt, so dass sich insgesamt 576 Subbänder ergeben. Im nächsten Schritt folgt dann die Quantisierung und die eigentliche Codierung. Als Parameter dienen die von der MDCT konstruierten Subbänder, das psychoakustische Modell, sowie die vom Anwender eingestellten Parameter. Anschließend findet noch eine Huffman-Codierung statt und die notwendigen Informationen zur Dekompression des Frames werden in einen Header gepackt. Beides zusammen ergibt dann den fertigen Frame.

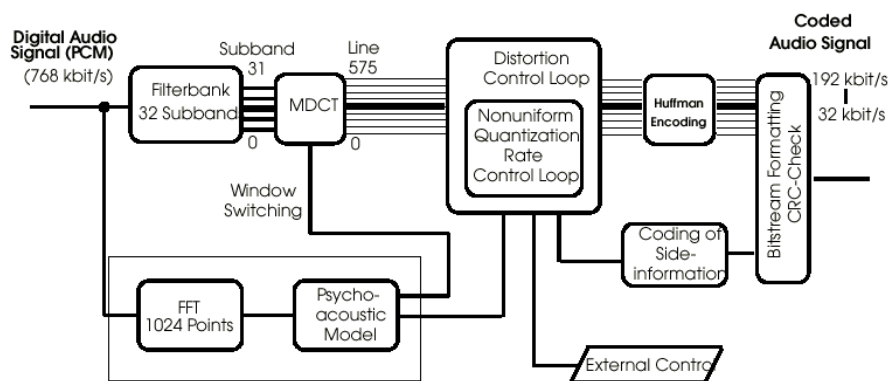


Abbildung 4: Blockdiagramm eines MP3 Encoders [3]

Die MDCT erhöht die Frequenzauflösung drastisch.[2] Durch diese Erhöhung kann Layer 3 auch die Aliasing-Effekte, die durch die Frequenzüberlappung benachbarter Bänder entstehen, rückgängig rechnen.

Layer 3 spezifiziert zwei MDCT-Blocklängen: 18 und 6 Samples. Nachfolgende Transformations-Fenster überlappen sich zu 50%, so dass die Fenstergröße 36 und 12 beträgt. Die lange Blockgröße erlaubt eine bessere Frequenzauflösung für stationäre Audiosignale, während die kurze Blocklänge bessere Zeitauflösung für transiente Signale bietet. Layer 3 schreibt definiert Varianten, in welcher Kombination kurze und lange Blöcke vorkommen können. Der Wechsel zwischen Blocklängen funktioniert nicht unmittelbar, sondern wird durch einen speziellen langen Block eingeleitet.

Weil die MDCT bessere Frequenzauflösung bietet, ist die Zeitauflösung entsprechend schlechter. Die MDCT operiert auf 12 oder 36 Filterbank-Samples, also ist das effektive Zeitfenster um den Faktor 12 oder 36 größer. Das Quantisierungsrauschen wird also Fehler erzeugen, die über dieses große Zeitfenster verteilt sind, also ist es wahrscheinlicher, dass man sie hört. Bei einem Signal, bei dem sich laute und leise Abschnitte in schneller Folge abwechseln, verteilt sich das Rauschen nicht nur auf die lauten Stellen (wo man es eher weniger stark hört), sondern auch auf die leisen Stellen. Diese Störungen treten gewöhnlich als sogenannte Pre-Echos auf, d.h. beim Decodieren erhält man ein Rauschsignal vor dem eigentlichen Event, das es verursacht, was dann vom Hörer wahrgenommen werden kann.

Um Pre-Echos zu reduzieren beinhaltet Layer 3 deshalb mehrere Maßnahmen. Das psychoakustische Modell versucht, die Voraussetzungen dafür zu erkennen, und Layer 3 kann auch Bits aus dem Bit-Reservoir borgen, um das Quantisierungsrauschen einzuschränken. Der Encoder kann auch auf eine kleinere MDCT-Blockgröße umschalten, um die effektive Fensterlänge zu kürzen.

Bei der Quantisierung verwendet man üblicherweise 2 verschachtelte Schleifen.

Die Quantisierung ist eine nicht-lineare Quantisierung nach dem *power law* Verfahren, bei dem größere Werte mit geringerer Genauigkeit gespeichert werden.

In der inneren Schleife, dem *rate-loop*, werden die notwendigen Abstände der Quantisierungsstufen ermittelt, damit nach der Huffman-Codierung die Daten auch in der zur Verfügung stehenden Bitrate gespeichert werden können.

In der äußeren Schleife wird dann geprüft ob das Quantisierungsrauschen die Maskierungskurve nicht überschreitet. Falls dies der Fall ist, werden die entsprechenden Bänder mit einem anderen Skalierungsfaktor gewichtet und der *rate-loop* erneut durchlaufen, wobei man hier aufpassen muss nicht in eine Endlosschleife zu geraten.

Für die Dekompression werden die Schritte der Kompression in umgekehrter Reihenfolge ausgeführt. Nach der Huffman-Dekodierung werden die Daten mittels inverser Quantisierung für die inverse modifizierte Cosinustransformation (IMCT) aufbereitet. Diese leitet ihre Daten weiter zu einer inversen Filterbank, welche nun die ursprünglichen Samples berechnet.

Quellen [2, 3, 4, 5]

## 5 Audiocodecs und ihre Einsatzgebiete

Audiocodecs werden natürlich nicht nur zum Abspielen von Musik- oder Filmdateien auf dem PC benötigt, sondern sind die Basis für sämtliche digitale Formate in denen Audiodaten verwendet werden.

Mit Entwicklung digitaler Festnetztelefonie (ISDN) Anfang der 80er wurden simple Codierungsverfahren wie das A-law bzw.  $\mu$ -law-Verfahren in Standards wie dem G.711, einem internationalen Telekommunikationsstandard, festgehalten und seitdem verwendet.

Bei den Sprachkompression ist die primäre Basis die CELP (*Code Excited Linear Prediction*) Codierung geworden, die mittels lineare Vorhersage und einer vordefinierten Tabelle, die spezielle Werte auf Basis der menschlichen Sprache enthält, vorgeht und auch mit sehr niedrigen Datenraten (ab 2 kbit/s) noch gute Ergebnisse liefert. Weiterentwicklungen wie LD-CELP (Low Delay CELP) oder Speex werden nicht nur in der Telekommunikation eingesetzt, wie z.B. bei der VoIP-Technologie, sondern auch in Anwendungen auf dem PC wie z.B. bei Sprachkonferenzprogrammen wie Teamspeak verwendet.

Beim digitalen Fernsehen (DVB, Digital Video Broadcasting) kommen je nach zur Verfügung stehende Bandbreite und der verwendeten Technik verschiedene Codecs zur Anwendung, wie der mittlerweile sehr alte MP2 (MPEG 1 Layer 2) Standard oder dessen Weiterentwicklungen wie MPEG2/MPEG4 AAC (Advanced Audio Coding).

Europaweit setzt der digitale Rundfunk auf den DAB Standard, der mittlerweile relativ alt ist, noch auf MP2 setzt und bereits seit 1995 zum flächendeckenden Einsatz kommt. Ab 2012 soll er europaweit das analoge Rundfunknetz endgültig ersetzen. Einige Länder wollen aber jetzt schon auf andere Codecs umstellen, was z.B. zu einem Problem für die Endverbraucher werden kann, die beim Besuch in anderen Ländern dann z.B. möglicherweise keinen Rundfunkempfang mit ihren Geräten haben.

Bei den Massensfilmmedien wie der DVD, Blu-ray oder der (mittlerweile vom Markt gedrängten) HD-DVD, kommen zwar die Videocodecs der MPEG Standards zum Einsatz (MPEG-2 bzw. MPEG-4 AVC/H.264), bei den Audiocodecs haben sich aber die Formate der Hersteller Dolby bzw. Digital Theater Systems, Inc (DTS), wie Dolby Digital/AC3, Dolby Digital Plus, dts, dts-HD durchgesetzt.

Diese finden sich auch im Kino wieder, dort meist leicht modifiziert gegenüber den Codecs für die Datenträger, um etwas höhere Datenraten und bessere Fehlerkorrektur (z.B. Staub und Kratzer auf Filmrollen) zu erlauben.

Was Multimediacodierungen am PC angeht, existieren mittlerweile zahlreiche Containerformate (z.B. avi, mkv, ogg, wmv, wma, qt, rm), die die Kombination von diversen Video- und Audiocodierungen<sup>1</sup> zulassen.

## 6 Schlussfolgerung und Diskussion

In den letzten 2 Jahrzehnten hat die Entwicklung von Audiocodern enorme Fortschritte gemacht und es lassen sich sehr effizient und qualitativ hochwertig Audiodaten speichern bzw. übertragen.

Die heutige Entwicklung im Bereich der verlustbehafteten Kompression konzentriert sich vor allem auf Streaming und Broadcasting Technologien bzw. der Qualitätsverbesserung bei niedrigen Bitraten.

Im High-Quality Audiodbereich gibt es vor allem im Endverbrauchermarkt dagegen kaum noch Bedarf für neue verlustbehaftete Kompressionsformate. Die bisherigen Technologien gelten als sehr ausgereift und neue Standards lassen sich nur schwer für eine breite Akzeptanz durchsetzen.

Das MP3 Format z.B. ist seit 1992 spezifiziert, und obwohl es zahlreiche Weiterentwicklungen gibt, die bessere Qualität bei entsprechenden Bitraten als MP3 codierte Dateien bieten, ist das MP3 Format immer noch das marktbeherrschende Format für reine Musikdateien. Für den normalen Endverbraucher ist auch nicht ersichtlich warum er andere Codierungen verwenden sollte, die sich dann mit seinen bisher gekauften Abspielgeräten kaum bis gar nicht abspielen lassen. Bei hohen Bitraten ist MP3 auch heute noch konkurrenzfähig und die etwas bessere Platzersparnis anderer Formate ist in Zeiten von billigen Terrabyte-Festplatten und günstigem Flashspeicher in Gigabyte-Größenordnungen nicht mehr relevant.

Wer als Musikliebhaber seine Musik archivieren möchte, sollte am Besten einen verlustfreien Audiocodern verwenden und für den alltäglichen Bedarf ein Format nach Wahl verwenden, das möglichst kompatibel ist. Hier wird die Zeit zeigen ob und wann sich ein anderes Format als Mainstream-Standard durchsetzen wird. Immerhin unterstützen viele MP3 Abspielgeräte auch andere Formate, wie z.B. OGG Vorbis oder Windows Media Audio (WMA). Letzteres vor allem aus DRM (Digitales Rechte Management) Gründen, da WMA dieses unterstützt. Allerdings hat der Markt bzw. die Verbraucher DRM Lösungen auch nur bedingt bis gar nicht akzeptiert, und bei vielen Musikportalen ist man zum MP3 Format zurückgekehrt.

Auch bei den neuen High Definition Formaten, wie Dolby Digital Plus, Dolby TrueHD oder DTS-HD, geht die Industrie den umgekehrten Weg und hat mittlerweile signifikant (je nach Format bis zu dem 50fachen der normalen DolbyDigital Codierung) die Datenraten gesteigert. Faktoren sind dabei verlustfreie Kodierungen, Erhöhung der Anzahl der Kanäle, Abtastfrequenzsteigerungen und Erhöhung der

---

<sup>1</sup> Eine gute Übersicht zu den gängigsten Audiocodern :  
<http://www.mpex.net/info/formate.html>

Bittiefe.

**Quellenverzeichnis:**

- [1] Davis Yen Pan – Digital Audio Compression
- [2] Felix von Leitner: Audio – Kompression
- [3] K. Brandenburg: Audio Coding: basics and state of the art
- [4] K. Brandenburg and H. Popp: An introduction to MPEG Layer-3, EBU  
TECHNICAL REVIEW – June 2000
- [5] <http://www.mp3-tech.org/>
- [6] Wikipedia, Die freie Enzyklopädie – <http://www.wikipedia.de>
- [7] Wikipedia, The free Encyclopedia – <http://www.wikipedia.com>
- [8] <http://www.mpex.net/info/verlustfrei.html#mp3hd>