

**Digital Audio: Abtasttheorem, A/D-  
und D/A-Wandlung, Audioformate  
(WAV, CD, DAT, Minidisk),  
Audioplayer**

Andreas Kellermeyer und Andreas Wimmer

# Inhaltsverzeichnis

<i>Inhaltsverzeichnis</i> .....	2
<b>1. Technik: Analog – Digital</b> .....	3
<b>1.1 Problematik Analog – Digital</b> .....	3
<b>1.2 Abtasttheorem</b> .....	3
<b>1.3 A/D-Wandlung</b> .....	6
1.3.1 Prinzip der Pulskodemodulation (PCM) .....	6
1.3.2 PCM im Detail.....	7
1.3.2.1 Abtastung.....	7
1.3.2.2 Quantisierung und Kodierung.....	7
1.3.2.3 Kompondierung.....	9
1.3.2.4 Praktische Umsetzung der Kodierung.....	10
<b>1.4 D/A-Wandlung</b> .....	11
<b>2. Audioformate</b> .....	12
<b>2.1 WAV-Format</b> .....	12
<b>2.2 CD-Format</b> .....	14
<b>2.3 DAT-Format</b> .....	16
<b>2.4 Minidisc</b> .....	17
<b>3. Audioplayer</b> .....	20
<b>4. Ausblick</b> .....	20

# 1. Technik: Analog – Digital

## 1.1 Problematik Analog – Digital

Um Schwingungen, egal ob es sich um einfach Schwingungen mit bestimmter Frequenz zum Beispiel vom Sinusgenerator oder um ganze Frequenzbänder, wie es bei der Sprache und der Musik der Fall ist, übertragen zu können, müssen diese niederfrequenten Signale entsprechend in elektrische Nachrichtensignale, etwa durch Mikrofone, umgewandelt werden. Diese im Niederfrequenz-Bereich liegenden Nachrichtensignale, d.h. im Hörbereich (unter 20 kHz) können im Allgemeinen in analoge und digitale Signale unterschieden werden. Der Unterschied zwischen analogen und digitalen Signalen besteht im Wesentlichen darin, dass die digitalen Signale auf einen endlichen (binären) Zeichenvorrat beschränkt sind. Die Informationen lassen sich somit als Zahlen speichern – sie liegen also digital vor. Der Vorteil hierbei liegt klar auf der Hand: Es besteht nun die Möglichkeit, die Folgen von 0er und 1er zu bearbeiten, zu kopieren oder sogar verschiedene Algorithmen auf sie anzuwenden, um die Daten beispielsweise zu komprimieren. Die digitalen Daten können nun auch über moderne Medien wie das Internet oder das Telefon übertragen werden. Hierbei spielen zwei Punkte eine Rolle:

- Übertragungsbandbreite, d.h. desto weniger Daten übertragen werden müssen, desto besser
- Vermeidung äußerer Störeinflüsse

Beide Punkte können mit der Digitalisierung von Signalen besser gehandhabt werden als analoge Signale. Die Bandbreite ist gegenüber der analogen Technik sowieso geringer, da man sich bei den digitalen Signalen auf einen endlichen Zeichenvorrat beschränken.

Durch die Verwendung von zum Beispiel fehlerkorrigierenden Codes durch Paritätsbits können auch Störeinflüsse behoben werden.

Doch auch einen wesentlichen Nachteil der Wandlung analoger in digitale Signale darf nicht vergessen werden: Da man sich bei den digitalen Signalen auf einen endlichen Zeichenvorrat beschränkt, gehen natürlich Informationen verloren. Infolgedessen muss man ein ausgewogenes Verhältnis zwischen der Größe des Zeichenvorrats und der Qualität der digitalisierten Signale finden.

## 1.2 Abtasttheorem

Ein einfaches analoges Signal lässt sich durch z.B. eine einfache Sinusschwingung darstellen:

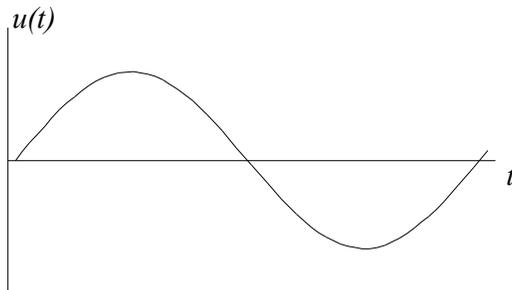
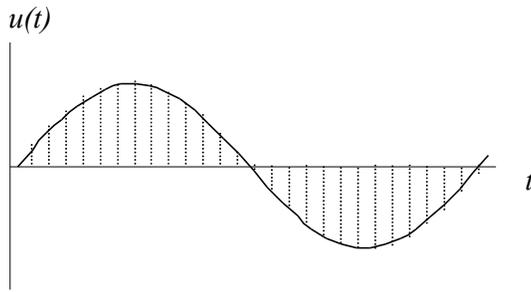


Abbildung 1: Analoges Eingangssignal

Dieses Eingangssignal gilt es, in ein digitales Signal umzuwandeln:



**Abbildung 2: Abgetastetes Signal**

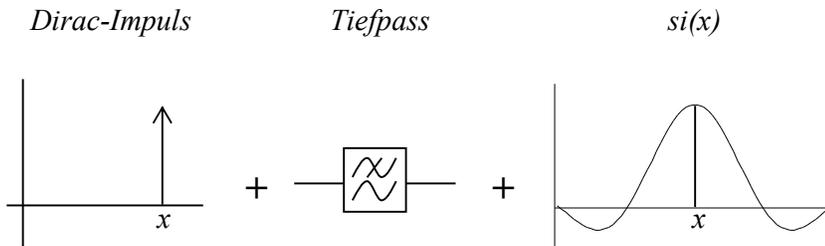
Hierbei tasten die gestrichelten Linien das analoge Eingangssignal ab – z.B. durch eine parallel zum Eingangssignal geschaltete Rechteckschwingung. Es entsteht ein PAM-Signal (Pulsamplitudenmodulation). Man sieht hierbei, dass bei weniger Abtastungen die Qualität abnimmt, da weniger Informationen erhalten werden und umgekehrt mehr Abtastungen liefern eine bessere Qualität. Das abgetastete Signal ist eine Folge von gewichteten Dirac-Impulsen<sup>1</sup>.

Jedes analoge Signal lässt sich durch die Frequenz und die Amplitude darstellen. Wir werden später sehen, dass sich zwar alle analogen Signale digitalisieren lassen, allerdings nicht immer lassen sich die digitalisierten Signale wieder zum ursprünglichen analogen Eingangssignal rekonstruieren. Dieser Sachverhalt hängt von der Frequenz des Eingangssignals ab. Um ein digitales Signal zu erhalten, das wir wieder rekonstruieren können, wählen wir unsere Abtastfrequenz  $f_A$  (d.h. die Anzahl der Abtastungen pro Sekunde) fest und fordern von unserem Eingangssignal, dass es die Maximalfrequenz  $f_{M_{\max}}$  besitzt. Um nun die Maximalfrequenz  $f_{M_{\max}}$  in Abhängigkeit zur Abtastfrequenz  $f_A$  zu bestimmen, ist es sinnvoll zu wissen, wie das digitalisierte Signal rekonstruiert wird.

Die Folge der Dirac-Impulse wird dazu durch einen so genannten Tiefpass<sup>2</sup> geschickt. Das besondere hierbei ist, dass wir für jeden einzelnen Dirac-Impuls die Funktion

$$\text{si}(x) := \frac{\sin(x)}{x}$$

erhalten:



**Abbildung 3: Rekonstruktion eines Dirac-Impulses**

<sup>1</sup> Dirac-Impuls: Impuls mit Breite  $\rightarrow 0$  und der Höhe 1. In der Praxis ist dies nicht möglich und ist daher ein Impuls mit möglichst geringer Breite. In der Theorie durch einen Strich dargestellt

<sup>2</sup> Tiefpass: Lässt Frequenzen bis zu einer bestimmten Frequenz passieren, höhere Frequenzen werden abgeschnitten

Die si-Funktion hat ihr Maximum an der Stelle und in Richtung des entsprechenden Dirac-Impulses. Die Höhe des Maximums entspricht der des Dirac-Impulses. Das besondere an der si-Funktion ist, dass die Nullstellen immer den Abstand  $\pi$  haben:

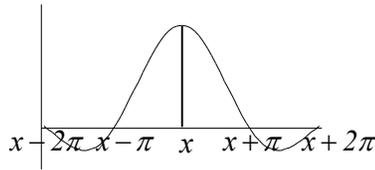


Abbildung 4: Nulldurchgänge der si-Funktion

Für eine feste Abtastfrequenz  $f_A$  und der zugehörigen Maximalfrequenz  $f_{M_{\max}}$  des Eingangssignals ergibt sich folgende Darstellung

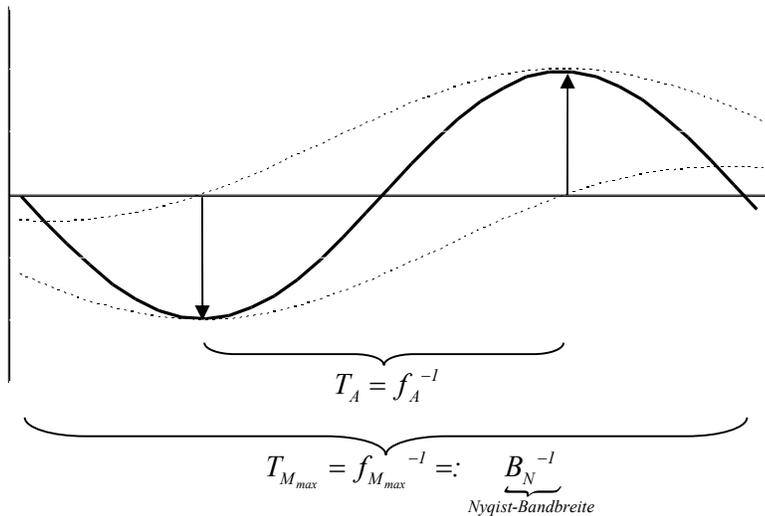


Abbildung 5: Zusammenhang Abtastfrequenz - Maximalfrequenz

Die gestrichelten Linien stellen die jeweiligen si-Funktionen dar. Aufgrund der Eigenschaft der si-Funktion schneiden sich alle si-Funktionen in den Nullpunkten an allen Dirac-Impulsen. Die Aufsummierung aller si-Funktionen liefert wieder das Eingangssignal. Abbildung 5 stellt den Grenzfall dar, d.h. falls die Abtastfrequenz  $f_A$  noch niedriger gewählt wird (also die Abstände der Dirac-Impulse zunehmen), sind die Maxima des Eingangssignals nicht mehr vollständig rekonstruierbar. Im schlechtesten Fall liegen die Dirac-Impulse in den Nulldurchgängen des Eingangssignals. In diesem Fall erhalten wir gar kein Signal mehr. Eine Erhöhung der Abtastfrequenz  $f_A$  hingegen verbessert die Näherung des digitalen Signals an das Eingangssignal.

Aus der Grafik lässt sich das Abtasttheorem ablesen:

$$T_A \leq \frac{1}{2B_N}$$

oder

$$f_A \geq 2f_{M_{\max}}$$

An der letzten Gleichung ist die Beziehung zwischen Abtastfrequenz  $f_A$  und Maximalfrequenz  $f_{M_{\max}}$  des Eingangssignals dargestellt: Die Abtastfrequenz  $f_A$  muss

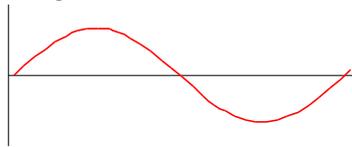
mindestens doppelt so groß sein wie die Maximalfrequenz  $f_{M,max}$  des Eingangssignals, um das Eingangssignal nach Abtastung wieder rekonstruieren zu können.

## 1.3 A/D-Wandlung

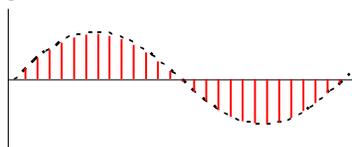
### 1.3.1 Prinzip der Pulskodemodulation (PCM)

Die PCM beschreibt ein Verfahren, ein analoges Signal in ein digitales Signal umzuwandeln. Der erste Schritt besteht darin, das analoge Signal, welches sowohl im Zeit- als auch im Wertebereich kontinuierlich ist, also unendlich viel Information beinhaltet, in ein zeitdiskretes, wertkontinuierliches Signal umzuwandeln. Wir erhalten ein PAM-Signal. Da dieses PAM-Signal noch unendliche Informationen beinhaltet (wertkontinuierlich) muss das Signal im nächsten Schritt quantisiert werden. Man erhält ein zeitdiskretes und wertdiskretes Signal. Dieses Signal kann jetzt durch eine Kodierung auf einen endlichen Zeichenvorrat umgewandelt werden, welcher dann übertragen oder auf verschiedenen Medien gespeichert werden kann. Diese Umwandlung des analogen Signals in ein digitales Signal findet in einem Baustein, dem so genannten ADU (Analog-Digital-Umsetzer) statt. Der nach der Übertragung/Speicherung vorhandene Digital-Analog-Umsetzer (DAU) wandelt das digitale Signal wieder in ein analoges Signal um.

*Analoges Signal:*



*PAM-Signal:*



*Quantisiertes Signal:*

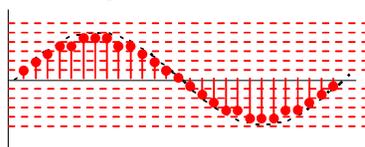


Abbildung 6: Schritte zur PCM

## 1.3.2 PCM im Detail

### 1.3.2.1 Abtastung

Der erste Schritt zur PCM ist die Abtastung des analogen Eingangssignals. Um das Signal rekonstruierbar abtasten zu können, dürfen im analogen Eingangssignal nur Frequenzen bis zu einer bestimmten Maximalfrequenz  $f_{M,max}$  vorkommen. Das Eingangssignal muss somit auf einen bestimmten Frequenzbereich beschränkt werden. Dies wird durch einen Tiefpass erreicht, der so eingestellt ist, dass die entsprechenden Frequenzen gefiltert werden. Dadurch ist das Abtasttheorem anwendbar und das analoge Signal kann abgetastet werden: es ergibt sich das PAM-Signal.

### 1.3.2.2 Quantisierung und Kodierung

Das PAM-Signal wird nun in einen Quantisierungsbereich eingebettet. Der Quantisierungsbereich begrenzt das PAM-Signal in der Amplitude und ist möglichst so zu wählen, dass keine oder nur eine geringe Begrenzung stattfindet. Anschließend wird der Quantisierungsbereich in gleichgroße (lineare) Abstände eingeteilt. Auch eine ungleichmäßige Aufteilung des Quantisierungsbereichs ist denkbar und wird beispielsweise in der Telefonie verwendet. Die entstehenden Stufen nennt man auch Quantisierungsstufen oder Quantisierungsintervalle. Auch hier gilt wie bei der Abtastung: Desto mehr Quantisierungsintervalle, desto näher ist man am Original. Hier findet nach der Abtastung die zweite Diskretisierung, diesmal im Wertebereich statt.

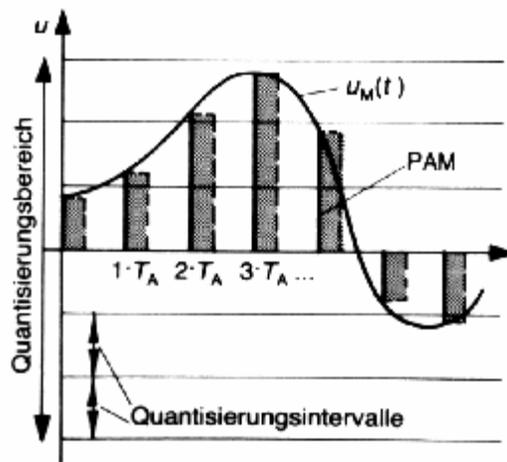


Abbildung 7: Quantisierung des PAM-Signals

Die Grenzen der Quantisierungsintervalle sind die Entscheidungsschwellen für die Kodewörter. Die Dirac-Impulse des PAM-Signals werden nun auf die Kodewörter mit Hilfe der Entscheidungsschwellen auf- bzw. abgerundet. Es entstehen in Abhängigkeit von der Anzahl der Kodewörter unterschiedlich große aber endlich viele Dirac-Impulse.

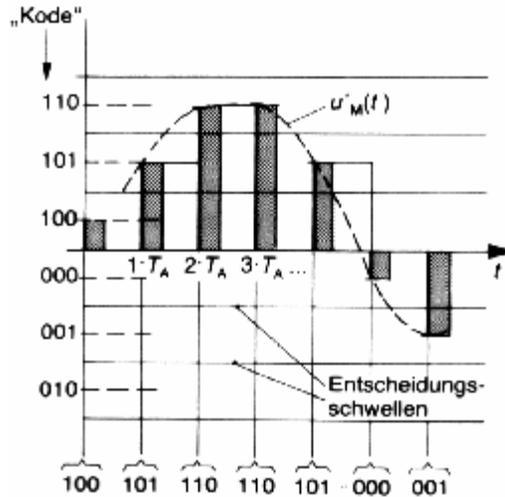


Abbildung 8: Kodierung für eine Kodewortlänge von 3 Bit

Die Kodewörter werden binär kodiert. Die Anzahl  $s$  der Kodewörter ergibt sich demnach als

$$s = 2^n$$

bei einer Kodewortlänge von  $n$  Bit.

Aufgrund dieser Information ist es möglich, die Bitrate (d.h. die Anzahl der Bits, die es pro Sekunde zu übertragen gilt) für unterschiedliche Abtastraten und Kodewortlängen zu berechnen.

Dies soll im Folgenden am Beispiel der CD erläutert werden:

Abtastrate (Samplerate): 44,1 kHz

Kodewortlänge: 16 Bit (14 Bit für Kodewortlänge  $2^{14} = 16384$  Quantisierungsintervallen)

Anzahl der Kanäle: Stereo (2-Kanal-Ton)

Die Bitrate  $B$  berechnet sich wie folgt:

$$B = \underbrace{2}_{\text{Stereo}} \cdot 16 \text{Bit} \cdot 44100 \text{Hz} = 1411,2 \frac{\text{kBit}}{\text{s}} = 176,4 \frac{\text{kB}}{\text{s}}$$

Dies lässt sich mit verschiedenen anderen Medien auch berechnen. Folgende Tabelle zeigt einen kleinen Überblick:

Qualität	Abtastrate [kHz]	Auflösung [Bit]	Mono/Stereo	Datenrate [kB/s]
Telefon	8	8	Mono	8
AM Radio	11	8	Mono	11
FM Radio	22	16	Stereo	88
CD	44	16	Stereo	176
DAT	48	16	Stereo	196

Tabelle 1: Übersicht verschiedener Medien

Falls das quantisierte Signal durch einen DAU wieder in ein analoges Signal umgewandelt werden würde, entsteht ein Signal, das dem ursprünglichen Eingangssignal sehr ähnlich ist, aber eben nur ähnlich:

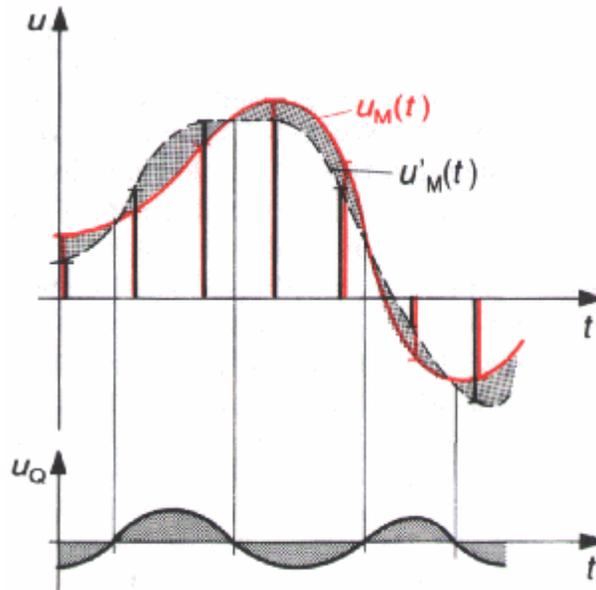


Abbildung 9: Quantisierungsgeräusch

Die rote Kurve entspricht dem ursprünglichen Eingangssignal, die gestrichelte dem analogen Signal nach der Quantisierung. Die darunter dargestellte Funktion ist die Differenz der beiden oberen Signale und wird als Fehlerfunktion bezeichnet. Die Amplitude der Fehlerfunktion ist maximal  $\pm$  einer halben Quantisierungsstufe. Bei vielen Quantisierungsstufen wird also das als Rauschen empfundene Quantisierungsgeräusch gering.

Der Quantisierungsgeräuschabstand stellt ein Maß für die Qualität der Quantisierung dar:

$$a_{q_{\max}} = 20 \cdot \lg(s) \text{ dB}$$

Bei der CD beträgt dieser Quantisierungsgeräuschabstand beispielsweise 84,3 dB, beim Telefon 48,2 dB.

### 1.3.2.3 Kompondierung

Das menschliche Gehör reagiert auf die relative Größe der Störung, d.h. wenn die Amplitude sehr groß ist, dann wird das Quantisierungsgeräusch kaum oder gar nicht wahrgenommen - bei kleiner Amplitude hingegen schon. Wenn man also an der Datenmenge sparen will (z. B. bei der Telefontechnik), ist es sinnvoll, wenn man diese Tatsache in die Quantisierung mit einbezieht. Die technische Umsetzung hierfür heißt Kompondierung, zusammengesetzt aus den beiden Wörtern „Kompression“ und „Expandierung“. Ziel der Kompondierung ist es den Störabstand der kleinen Momentanwerte eines Signals auf Kosten der Großen zu verbessern. Dazu gibt es drei Methoden:

- Momentanwertkompondierung
- digitale Kompondierung
- 13-Segment-Kompondierkennlinie.

Bei der Momentanwertkompondierung werden die großen Momentanwerte vor der Quantisierung gepresst, und die kleineren aus dem Kanalrauschen herausgehoben („Presser“), um damit einen Störabstandsgewinn erhalten. Im Empfänger wird dieses Verfahren im so genannten „Dehner“ wieder umgekehrt, womit das ursprüngliche Signal wieder erhalten wird.

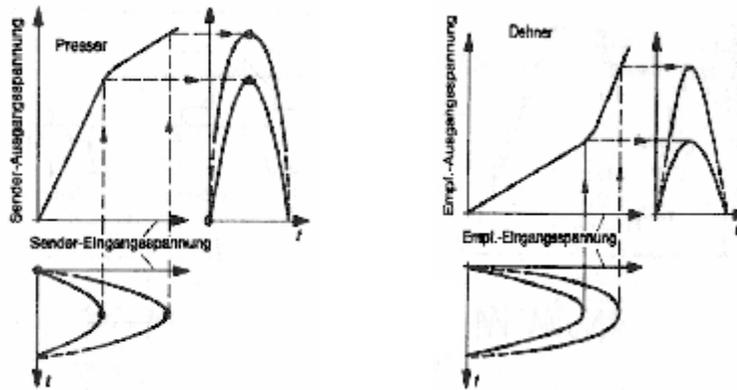


Abbildung 10: Presser und Dehner

Bei der digitalen Kompanidierung wird das analoge Signal im Gegensatz zur Momentanwertkompanidierung nicht verändert. Die Quantisierung findet nicht mehr linear statt, sondern wird nichtlinear durchgeführt. Bei kleineren Signalspannungswerten wird in kleineren Intervallen quantisiert als bei großen Signalspannungswerten. Bei der technischen Umsetzung gibt es dazu zwei Möglichkeiten: Entweder arbeitet der A/D-Umsetzer mit unterschiedlichen Stufenhöhen, oder der A/D-Wandler ist linear mit durchgehend kleineren Quantisierungsintervallen und es erfolgt anschließend eine Umrechnung.

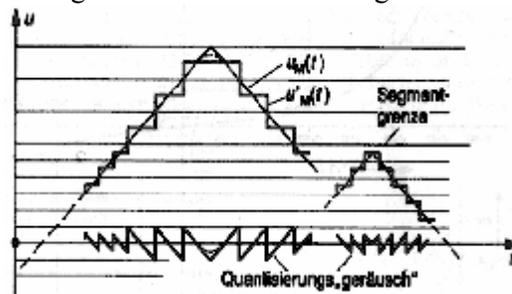


Abbildung 11: Digitale Kompanidierung – nichtlineare Quantisierung

Die 13-Segment-Kompanidierkennlinie ist ein Verfahren, welches bei der PCM-Telefonie eingesetzt wird. Der Quantisierungsbereich wird hier in 13 Segmente unterteilt, wobei jedes Segment unterschiedlich große Quantisierungsintervalle hat. Innerhalb eines Segments sind die Stufenhöhen allerdings gleich. Die Stufenhöhe unterscheidet sich von Segment zu Segment um den Faktor 2. Dadurch wird bei den in der Telefonie wichtigen kleinen Signalspannungen sehr genau quantisiert, und man erreicht somit eine Wortlänge von 8 Bit. Bei gleichmäßiger Quantisierung wäre eine Wortlänge von 12 Bit nötig gewesen, um die gleiche Qualität zu erreichen.

### 1.3.2.4 Praktische Umsetzung der Kodierung

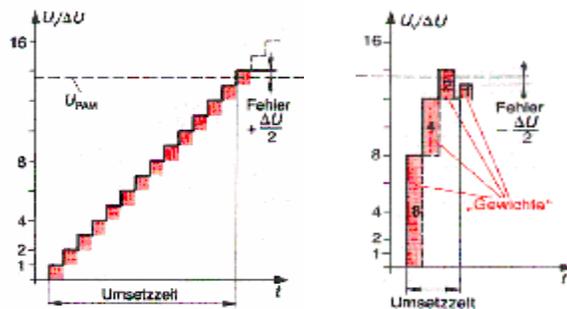
Die Kodierung bezeichnet die Umwertung der Analogwerte in Digitalwerte, und kann im engeren Sinne als A/D-Wandlung bezeichnet werden, wobei aber auch die Quantisierung und Kompanidierung im weiteren Sinne dazu gehören. Nachdem zuvor das Signal zeitdiskret und wertdiskret vorbereitet wurde, erzeugt die Kodierung nun daraus den Binärkode. Hierfür gibt es im Wesentlichen drei Verfahren:

- Zählverfahren (schrittweise Annäherung, engl. „step-at-a-time“)
- Wägeverfahren (ADU mit sukzessiver Annäherung, engl. „bit-at-a-time“)

- Parallelverfahren (auch Momentanwertkodierer genannt, engl. „word-at-a-time“)

Beim Zählverfahren wird der zu kodierende Momentanwert mit einer Treppenspannung verglichen, d.h. es existiert ein Normal das der Höhe einer Quantisierungsstufe entspricht, welches solange aufaddiert wird, bis die aufaddierten Normalen den momentanen Abtastwert erreicht haben. Bei jedem Durchgang wird ein Binärzähler hoch gezählt, der am Ende der Kodierung eines Wertes den Binärkode ausgibt. Anschließend wird dieser Binärzähler wieder auf Null gesetzt, um für den nächsten Wert wieder von neuem beginnen zu können. Bei diesem Verfahren sind somit sehr viele Schritte notwendig ( $2^n - 1$  Schritte bei einem Kodewort der Länge  $n$ ), womit er nicht für zu große Abtastfrequenzen oder Quantisierungsstufen geeignet ist. Allerdings ist seine Umsetzung sehr einfach.

Beim Wägeverfahren verwendet man unterschiedliche große Normale. Jedem Bit des Kodewortes wird einem ihm entsprechend großes Normal zugeordnet und dann, beginnend beim höchstwertigen Bit, mit dem Signalwert verglichen. Wenn das Normal größer als der Signalwert ist, wird das entsprechende Bit auf null gesetzt, andernfalls auf eins, das nächste Normal aufaddiert und wieder verglichen, bis alle Bits gesetzt sind. Bei diesem Verfahren ist der Aufwand nur noch bei  $n$  Vergleichen bei Kodewörtern der Länge  $n$ .

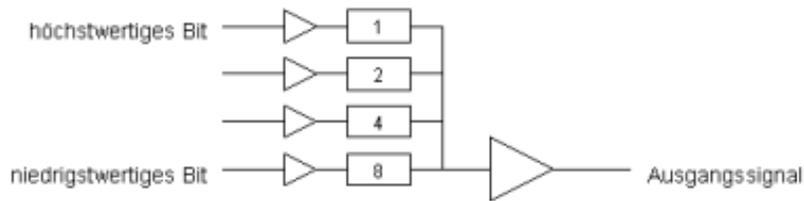


**Abbildung 12: Vergleich Zählverfahren - Wägeverfahren**

Das Parallelverfahren ist das technisch aufwendigste, da so viele Normale benötigt werden, wie es Quantisierungsstufen gibt. Dadurch ist aber ein, wie der Name schon sagt, paralleles Vergleichen mit allen Normalen möglich, dessen Ergebnis anschließend von einer Logik ausgewertet wird. Somit ist nur noch ein Schritt für einen abgetasteten Wert notwendig. Bei Audiosignalen ist diese Geschwindigkeit allerdings normal nicht notwendig, weshalb diese Methode eher dort eingesetzt wird, wo man es mit sehr hochfrequenten Signalen zu tun hat (z.B. Fernsehsignale).

## 1.4 D/A-Wandlung

Das digitale Signal liegt uns als Binärwörter mit jeweils gleicher Kodewortlänge von  $n$  Bit vor. Um das Signal zu dekodieren und wieder in ein analoges Signal umzuwandeln, müssen diese  $n$  Bit mit Hilfe von Widerständen auf den jeweiligen Spannungswert gebracht werden. Dazu wird jedes einzelne Bit eines Codes zuerst durch einen Verstärker geschickt, so dass das Bit 0 auch wirklich eine Spannung von 0 V erhält und das Bit 1 eine Spannung von z.B. 5V erhält. Anschließend wird das höchstwertigste Bit (jetzt eben mit Spannung 0V oder 5V) eines Codes durch einen Widerstand bestimmter Größe geschickt. Das nächste Bit geht durch einen doppelt so großen Widerstand usw. bis zum niederwertigsten Bit. An der Grafik eines 4-Bit-Wandlers sieht das wie folgt aus:



**Abbildung 13: 4-Bit-D/A-Wandlung**

Widerstände, an denen nun die gleiche Spannung (also 0V oder 5V) anliegt, können als parallel geschaltet betrachtet werden. Der Ersatzwiderstand zweier parallel geschalteter Widerstände ergibt sich aus der Formel

$$R_E = \frac{R_1 \cdot R_2}{R_1 + R_2}$$

Diese Widerstände arbeiten also als Spannungsteiler. Die Ausgangsspannung eines Spannungsteilers ist durch die Formel

$$U_a = \frac{U_1 R_2 - U_2 R_1}{R_1 + R_2}$$

gegeben. Somit lässt sich das Ausgangssignal nach vorgeschaltetem Verstärker bestimmen. Dieser Vorgang muss nun für alle Codes wiederholt werden. Bei einer Abtastrate von 44,1 kHz werden also 44100 D/A-Wandlungen pro Sekunde durchgeführt. Wichtig ist hierbei, dass die Anzahl der D/A-Wandlungen pro Sekunde mit der Abtastrate bei der Wandlung des analogen Eingangssignals zum PAM-Signal übereinstimmt. Andernfalls wird das Ausgangssignal gegenüber dem Eingangssignal gestaucht bzw. gestreckt.

## 2. Audioformate

Die Wandlung eines analogen Signals in ein digitales Signal wird durch das PCM-Verfahren erreicht. Doch um die digitalen Informationen auch nutzen zu können, müssen sie auf Medien gespeichert werden. Im Folgenden werden verschiedene Formate beschrieben, wie das PCM-Signal gespeichert werden kann.

### 2.1 WAV-Format

Das WAV-Format wurde ursprünglich von Electronic Arts (EA) entwickelt. Microsoft entwickelte basierend auf dem EA-IFF-85-Standard sein eigenes WAV-Format. Der Unterschied zwischen Electronic Arts und Microsoft bestand hauptsächlich in der Anordnung der Bytes: Microsoft verwendet die Intel-Byteanordnung (Little Endian). Da WAV von Microsoft in der heutigen Form entwickelt wurde, findet es hauptsächlich in Windows-Umgebungen Anwendung, verliert aber zunehmend an Bedeutung. WAV ist allerdings nur ein Teil eines Standards, dem so genannten RIFF-Standard (Resource-Interchange-File-Format). Dieser Standard erlaubt es neben PCM-Daten (WAV) auch Bitmaps (Bilder), Farbpaletten und MIDI-Daten zu speichern.

Eine RIFF-Datei ist in so genannte Chunks eingeteilt. Jeder Chunk stellt einen abgeschlossenen Abschnitt dar. Um eine WAV Format zu erzeugen, werden drei Chunks benötigt:

- Wave-Chunk
- Format-Chunk
- Data-Chunk

Der Wave-Chunk gibt an, dass es sich bei den folgenden Chunks um WAV-Informationen handelt. Der Format-Chunk enthält Informationen zu den PCM-Daten, d.h. Abtastrate, Anzahl der Kanäle usw. Erst im Data-Chunk sind die Informationen zum PCM-Signal abgelegt.

Der Wave-Chunk hat folgenden Aufbau:

Anzahl Bytes	Daten
4	„RIFF“ (Hex: 0x52494646)
4	Anzahl der verbleibenden Bytes bis zum Dateieende
4	„WAVE“ (Hex: 0x57415645)
char[]	Format-Chunk und Data-Chunk

**Tabelle 2: Aufbau Wave-Chunk**

Der Format-Chunk enthält die Informationen zum PCM-Signal:

Anzahl Bytes	Daten
4	„fmt “ (Hex: 0x666D7420)
4	Anzahl der verbleibenden Bytes im aktuellen Chunk
2	Angabe, wie die Audiodaten zu interpretieren sind
2	Anzahl der Kanäle
4	Abtastrate in Hz
4	Samplerate in Bit/s
2	Bytes pro Sample
2	Kodewortlänge in Bit

**Tabelle 3: Aufbau Format-Chunk**

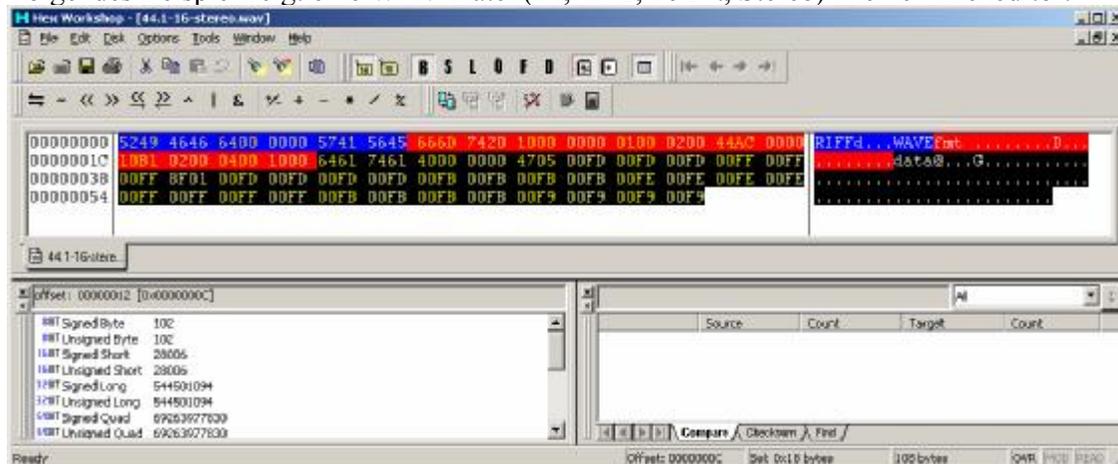
Die Daten des PCM-Signals werden im Data-Chunk bereitgestellt:

Anzahl Bytes	Daten
4	„data“ (Hex: 0x64617461)
4	Anzahl der verbleibenden Bytes bis zum Dateieende
n	Raw-Data

**Tabelle 4: Aufbau Data-Chunk**

Da die Angabe „Anzahl der verbleibenden Bytes bis zum Dateieende“ im Wave-Chunk bzw. im Data-Chunk auf 4 Byte beschränkt ist, ergibt sich eine Maximalgröße von 4GB für eine WAV-Datei.

Folgendes Beispiel zeigt eine WAV-Datei (44,1kHz, 16 Bit, Stereo) in einem Hexeditor:



**Abbildung 14: WAV-Datei im Hexeditor**

Der blaue Bereich zeigt den Wave-Chunk, der rote Bereich den Format-Chunk und im schwarzen Bereich ist das PCM-Signal gespeichert (Data-Chunk).

## 2.2 CD-Format

Die CD wurde gemeinsam von Phillips und Sony entwickelt und sollte die LP durch bessere Qualität und Handhabung ersetzen. 1980 wurde der CD-Standard im so genannten „Red Book“ veröffentlicht. 1983 wurden die ersten CD-Player verkauft. Im Laufe der Jahre wurden noch mehrere Books veröffentlicht, die allerdings den Audio-CD-Standard nicht beeinflussten. Sie hielten die Bestimmungen für CD-Rom, CD-I, Video-CD und CD-R fest.

Auf der Audio-CD wird das unkomprimierte PCM-Signal abgespeichert. Die Abtastrate beträgt, wie bereits erwähnt, 44,1 kHz und die Wortlänge ist 16 Bit. Die Binärdaten werden in einer spiralförmigen Spur, die von innen nach außen verläuft gespeichert. In dieser Spur sind kleine Vertiefungen, die so genannten Pits - die Fläche zwischen den einzelnen Pits heißt Land. Bei der Herstellung werden die Pits in das Trägermaterial gepresst. Da die Daten mittels eines Lasers ausgelesen werden, wird im nächsten Schritt eine dünne Metallschicht (meist Aluminium) über die Lands und Pits gezogen, die das Laserlicht reflektiert. Die nächsten Schichten sind die Schutzschicht, die, wie der Name schon sagt, die Daten vor mechanischen Einwirkungen schützen soll, und eine Schicht für die Druckfarbe des Labels.

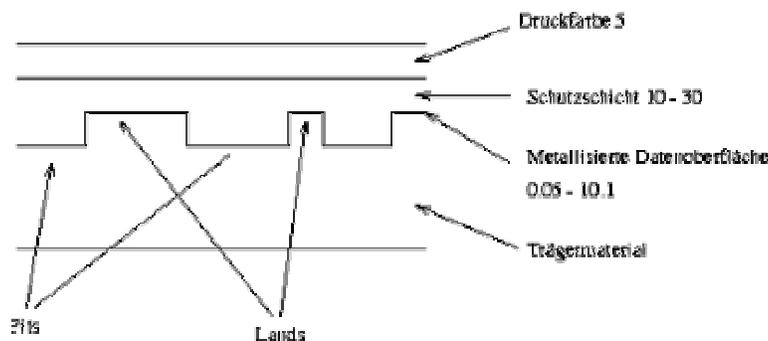


Abbildung 15: Aufbau einer CD

Der Aufzeichnungsbereich der CD ist in drei Teile unterteilt:

- Lead-In (Eingangsbereich)
- Datenbereich
- Lead-Out (Ausgangsbereich)

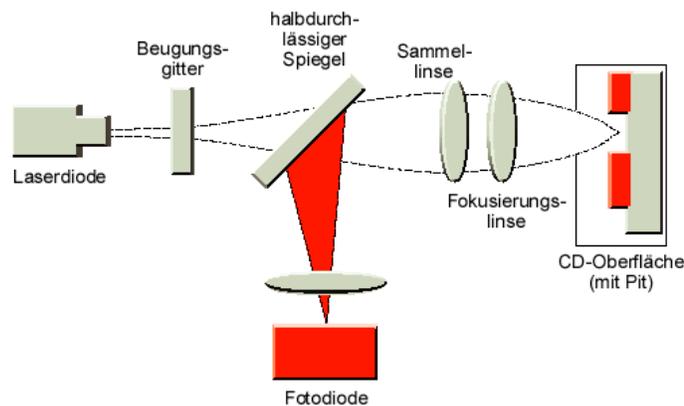
Im Lead-In ist das Inhaltsverzeichnis, auch TOC (Table of Content) genannt, welches der CD-Player als erstes ausliest. Das Lead-Out markiert das Ende der Aufzeichnung und enthält keine Daten.

Die CD kann bis zu 99 Tracks enthalten, die jeweils in Sektoren mit einer Länge  $\frac{1}{75}$  Sekunden unterteilt sind, wodurch die Adressierung (Minute:Sekunde:Sektor) ermöglicht wird.

Die CD wird mittels eines Lasers ausgelesen, was den Vorteil hat, dass das Medium berührungs- und damit verschleißfrei ausgelesen wird. Somit wird eine Verschlechterung der Qualität durch mehrmaliges Abspielen, wie es bei der LP der Fall war, verhindert.

Der Laserstrahl passiert zuerst ein Beugungsgitter, welches den Strahl ausweitet. Anschließend geht er durch einen halbdurchlässigen Spiegel und wird schließlich durch ein Fokussierungs-Linsensystem so verengt, dass er auf der CD-Oberfläche 0,8 mm breit ist. Durch die Brechungseigenschaften des Trägermaterials wird der Strahl bis zum Auftreffen auf die Datenspur auf 1,7  $\mu\text{m}$  gebündelt. Durch diese Methode wirken sich kleine Kratzer oder

Staubpartikel kaum aus. Der Laserstrahl muss nun bei einem Pit eine längere Strecke zurücklegen als bei einem Land, wodurch das Licht unterschiedlich reflektiert wird und so unterschiedlich interpretiert wird. Der Übergang von einem Pit zu einem Land oder umgekehrt wird als eine 1 im Binärkode interpretiert, alles andere als 0. Der Strahl wird durch die Metallschicht reflektiert und durch den halbdurchlässigen Spiegel auf eine Fotodiode geworfen, die den Strahl interpretiert.



**Abbildung 16: Auslesen der Informationen von einer CD**

Aufgrund von möglichen mechanischen Schäden der CD und aus lesetechnischen Gründen ist es nötig, dass man mehrere Verfahren anwendet, um Lesefehler zu vermeiden. Die komplette Fehlerkorrektur einer CD ist teilweise sehr kompliziert und umfangreich, weshalb hier nur auf die grundlegenden Methoden eingegangen wird:

- EFM-Modulation (Eight to Fourteen Modulation):  
Wie schon erwähnt, wird ein Übergang von einem Pit zu einem Land oder umgekehrt als 1 interpretiert. Diese Information nennt man Channel-Bit und ist die kleinste Informationseinheit auf einer CD. Die übrigen Stellen sind Null-Channel-Bits. Wenn die Pits oder Lands kürzer als 3 bzw. länger als 11 Channel-Bits sind, werden die Übergänge nicht mehr richtig erkannt, weshalb nicht jede Bitkombination möglich ist. Die EFM-Modulation wandelt deshalb 8 Bit in 14 Bit nach einer festen Tabelle um, womit dieses Problem umgangen wird.
- Parity Bits:  
Um die fehlenden Informationen, die womöglich durch Kratzer entstanden sind, auszugleichen benötigt man noch Zusatzinformationen, um den kompletten Kode wieder rekonstruieren zu können. Hierfür verwendet man ein Paritätsbit, das z.B. zu einem Byte zusätzlich angehängt wird. Das Bit wird auf 0 oder 1 gesetzt, je nachdem ob die Anzahl der Einsen im Byte gerade oder ungerade ist. Wenn man noch mehr Bits anfügt lässt sich so der Fehler noch lokalisieren und beheben (Reed-Solomon-Code).
- Interleaving:  
Da bei einer CD oft etwas größere Kratzer entstehen, die damit viele hintereinander folgende Informationen zerstören (Burst), benötigt man noch ein weiteres Korrekturverfahren. Beim Interleaving werden hintereinander folgende Daten niemals benachbart abgespeichert, sondern so umverteilt, dass bei einem Kratzer nur Teile eines Blockes zerstört werden, dessen weitere Informationen an einer anderen Stelle auf der CD sind und so mit Hilfe der Paritätsbits der Block wieder rekonstruiert werden kann. Die Verfeinerung dieser Technik nennt man Cross Interleaving Hierbei werden die Daten mehrere Male in kürzeren als auch in längeren Intervallen verschachtelt.

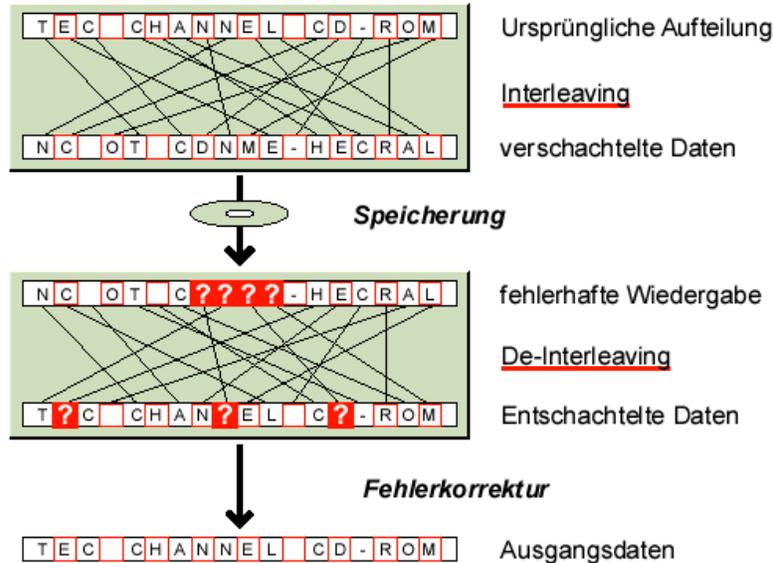


Abbildung 17: Interleaving

Die Kombination aus Paritätsbits und Cross-Interleaving nennt man CIRC (Cross Interleave Reed-Solomon-Code).

## 2.3 DAT-Format

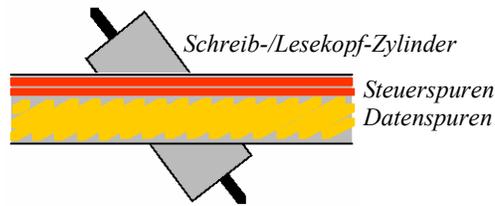
Ein damaliges Hauptproblem der CD bestand in der fehlenden Aufzeichnungsmöglichkeit für Heimanwender auf CD-Medien. Die entsprechende Technik von Sony war noch nicht ausgereift genug, um sie auf den Markt zu bringen. Darum arbeitete man an einer Alternativlösung. 1987 wurde als Ergebnis DAT (Digital Audio Tape) eingeführt. Das DAT sieht einer herkömmlichen Kassette sehr ähnlich, sie ist nur kleiner und dicker. Außerdem besitzt das DAT einen Bandschutzmechanismus, den nur das Laufwerk öffnen kann. Der Vorteil des DAT lag vor allem im variablen Format: es ist möglich, mit einer Abtastfrequenz von 32kHz bei 12 Bit aufzuzeichnen, aber auch bei 44kHz mit 16 Bit und 48kHz mit ebenfalls 16 Bit. Durch letzteres Format wurde sogar die Qualität der CD übertroffen. Dies ist der Grund, warum DAT als Vorlage für eine CD, dem so genannten Mastering verwendet wurde. Ansonsten ist DAT der CD sehr ähnlich: die PCM-Daten werden ebenfalls unkomprimiert gespeichert und es besteht auch die Möglichkeit eines Kopierschutzes.

DAT verliert heute allerdings zunehmend an Bedeutung. Die Zugriffszeit beträgt immerhin 30 Sekunden. Moderne Festplatten erreichen mittlerweile Zugriffszeiten zwischen 8 und 12ms. Zudem bieten heutige Festplatten mehr Kapazität und können besser verwaltet werden.

Genau diese Kapazitätsfrage stellte anfangs eine Aufgabe für DAT an das herkömmliche Tape. Die Technik der linearen Speicherung der Daten auf das Band brachte gerade einmal drei Minuten PCM-Daten in CD-Qualität auf ein normales Band von 60 Metern. Durch eine Übernahme einer Technik namens Helial Scan aus den VHS-Rekordern wurden die Daten schräg auf das Band gespeichert. Dadurch werden die Daten dichter auf das Band gepackt:



Abbildung 18: Lineare Speicherung der Daten



**Abbildung 19: Schräge Speicherung der Daten**

Erreicht wird diese schräge Speicherung der Daten durch schräg zum Band liegende Zylinder, die die Schreib- und Leseköpfe beinhalten. Der Zylinder dreht sich gegen die Laufrichtung des Bandes.

Durch diese Technik können anstatt drei Minuten nun 120 Minuten PC-Daten in CD-Qualität auf 60 Meter Band gespeichert werden.

Durch die nun vorhandene Möglichkeit, digitale Kopien zu erstellen, legte die AHRA (Audio Home Recording Act) 1992 fest, neue DAT-Rekorder mit einem Kopierschutzmechanismus, dem SCMS (Serial Copy Management System) auszustatten. Dabei wurde der PCM-Header um das „ID-6“-Feld erweitert. Es gab nun drei Möglichkeiten für dieses neue Feld:

- 00: unendlich viele digitale Kopien
- 10: keine digitalen Kopien
- 11: eine digitale Kopie

Falls nun ein DAT mit dem SCMS-Kode „11“ kodiert war, musste ein DAT-Rekorder in der digitalen Kopie die 10 durch eine 11 ersetzen. Somit konnte man von der Kopie keine weitere Kopie anfertigen. Analoge Kopien waren durchaus weiterhin möglich.

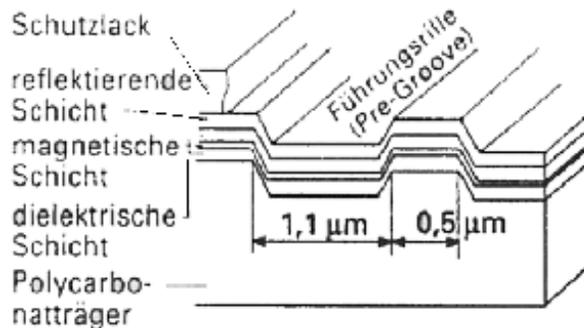
Der heutige CD-Brenner fällt übrigens nicht unter die AHRA, da er nicht primär zur Tonspeicherung entwickelt wurde.

## **2.4 Minidisc**

Die Minidisc (MD) wurde von Sony entwickelt und veröffentlicht und sollte die Audio-Kassette ablösen, indem man die Vorteile von Kassette und CD verbindet. Das neue Medium sollte wie die Kassette aufnahmefähig und gleichzeitig kompakt und robust sein und damit portabel. Außerdem sollte es auch die Vorteile und Qualität der CD haben, so dass man wahlfreien Zugriff auf alle Tracks hat und kein langes Umspulen nötig ist. Erreicht wurde dies durch die MO-Technologie (magneto-optisch) das Sony bereits 1988 entwickelt hatte. Durch diese Technik ist es möglich, mit einem Gerät, das zugleich kompakt gebaut werden kann, die Minidisc unkompliziert zu bespielen und zu lesen. Des Weiteren wurde, um das kleine Format zu verwirklichen und dieselbe Kapazität wie bei der CD unterzubringen, eine verlustbehaftete Kodierung entwickelt, die aber durch das Ausnutzen psychoakustischer Merkmale des menschlichen Gehörs kaum hörbare Unterschiede zur CD hervorbringt. Die Minidisc wurde 1992, etwa zeitgleich mit der DCC (Digital Compact Disc) von Phillips, auf den Markt gebracht. Allerdings war das Produkt von Phillips dem Marketing und den Vorteilen der Minidisc nicht lange gewachsen, weshalb die Produktion der DCC etwa vier Jahre nach Erscheinung eingestellt wurde.

Die Daten der MD wurden auf einer der CD ähnlichen Scheibe untergebracht, die allerdings in einer Cartridge (7cm x 6,75cm x 0,5cm) eingebaut wurde. Der Zugang zum Lesen und Beschreiben wird durch einen Shutter ermöglicht, der beim Einlegen in das Gerät automatisch geöffnet wird. Der Speicher der Minidisc beträgt 160MB, auf dem durch ATRAC 74 Minuten passen. Die beschreibbaren und die unbeschreibbaren Medien sind in ihren Aufbau unterschiedlich. Während die unbeschreibbare Minidisc bis auf Cartridge und Maße der CD entspricht, ist die beschreibbare Minidisc nach der MO-Technologie aufgebaut. Der

Grundkörper ist wie bei der CD aus Polycarbonat, jedoch ist in ihm schon eine fertige Rillenstruktur die Spiralförmig von innen nach außen geht und somit für die Führung des Lasers dient. Auf diesem Grundkörper sind weitere Schichten. Die magnetisch wirksame Schicht ist in Hilfsschichten eingebettet, die eine hohe Leistungsfähigkeit bei geringem Leistungsbedarf gewährleisten und die Reflexion des Laserlichts sichern. Mit diesem Aufbau sind eine Million Schreibzyklen ohne Qualitätsverlust und guter Langzeitstabilität möglich.



**Abbildung 20: Aufbau einer Minidisc**

Die Lasermodulation, wie sie bei optischen Speichern der Computer eingesetzt wird, wollte man bei der CD nicht verwenden, da hier Löschen und Beschreiben zwei getrennte Verfahren sind. Wegen entscheidenden Vorteilen in der Handhabung entwickelte man bei der Minidisc die Magnetfeldmodulation. Bei dieser Technik wird mit einer Kombination aus einem Laser und einem Magnetfeld gearbeitet. Der Laser markiert dabei die Stelle der Minidisc, auf der geschrieben werden soll und erhitzt diese auf ca. 180°C, während das Magnetfeld von der anderen Seite der Disc einwirkt. Der Laser ist so immer während der Aufnahme eingeschaltet und das Magnetfeld wird dem Aufzeichnungssignal entsprechend umgepolt.

Das Inhaltsverzeichnis (UTOC = User Table of Content) der beschreibbaren Minidisc ist im Vergleich zum TOC bei der unbeschreibbaren Minidisc variabel, d.h. es muss nicht den exakten Verlauf der Daten auf der MD enthalten. Es können somit Tracks in ihrer Reihenfolge einfach verändert werden, indem das UTOC verändert wird. Somit wird das Erstellen und Löschen von Tracks sehr komfortabel.

Da die Minidisc für den portablen Einsatz entwickelt wurde, war es notwendig, Störungen durch Erschütterungen und dem damit verbundenen Springen des Abtasters zu vermeiden. Erreicht wurde dies mit Hilfe eines Zwischenspeichers, der in der Lage ist, die Musiksignale für eine genügend lange Zeit zwischenzuspeichern. Die Datenkompression durch ATRAC ist hier von Vorteil, da der Minidisc-Player die Daten in der selben Geschwindigkeit ausliest wie bei der CD (1,4 Mbit/s), für den kontinuierlichen Datenfluss aber nur 0,3 Mbit/s benötigt werden. Wenn nun der Abtaster durch eine Erschütterung gestört wird, entleert sich der Speicher langsam. Nachdem die Störung vorbei ist kann der Speicher aufgrund der höheren Geschwindigkeit des Auslesens wieder aufgefüllt werden, ohne dass der Datenfluss dabei gestört wird.

Da die Minidisc aufgrund ihrer geringen Abmessung nur ein fünftel an Daten gegenüber einer CD speichern kann, wurde ein spezielles Datenkompressionsverfahren entwickelt, genannt ATRAC (Adaptive Transform Acoustic Coding). Damit sich die Qualität nicht sonderlich von der CD unterscheidet und trotzdem 75 min Spielzeit auf die MD passen, wurden bestimmte psychoakustische Eigenschaften des menschlichen Gehörs ausgenutzt, um so unnütze Daten wegzulassen.

Die Empfindlichkeit des Ohres ist von der Frequenz des Tones abhängig, d.h. ein Ton der in einer bestimmten Frequenz gerade noch wahrgenommen wird, kann bei gleicher Intensität aber anderer Frequenz unhörbar sein. Das Maximum der Ohrempfindlichkeit liegt um 4kHz.

Somit kann auch ein Ton bei Anwesenheit eines lautereren unhörbar werden, was mit Verdeckung oder Maskierung bezeichnet wird. Je enger die Töne in ihrer Frequenz sind, desto stärker zeichnet sich dieser Effekt aus.

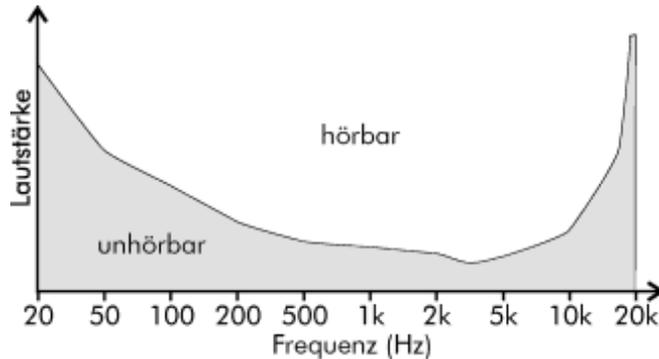


Abbildung 21: Wahrnehmung von Tönen beim menschlichen Gehör

ATRAC zerlegt das Musiksignal in zeitlich begrenzte Blöcke, bestimmt die Empfindlichkeit jedes Frequenzbereiches und zeichnet das Musiksignal aufgrund dessen mehr oder weniger genau auf. Bei der Unterteilung in die verschiedenen Frequenzbereiche wird ebenfalls wieder auf bestimmte Eigenschaften des Ohres geachtet. Deshalb ist der Frequenzbereich in verschiedene Frequenzgruppen (Critical Bands) aufgeteilt, die beim menschlichen Ohr festgestellt wurden, die im unteren Frequenzbereich wesentlich enger zusammen liegen als im oberen (Abbildung 22).

Weiter wird auch noch darauf geachtet, ob ein Signal lebhaft oder eher ruhiger ist. Dies wirkt sich auf die Länge der gebildeten Blöcke aus: bei lebhaften Passagen werden kürzere Zeitblöcke von 1,49 ms oder 2,9 ms gebildet, bei getragenen Teilen bis zu 11,6 ms. Insgesamt lässt sich mit diesen Methoden ein beträchtlicher Teil der Daten reduzieren, bei gleichzeitiger Vermeidung hörbarer Unterschiede.

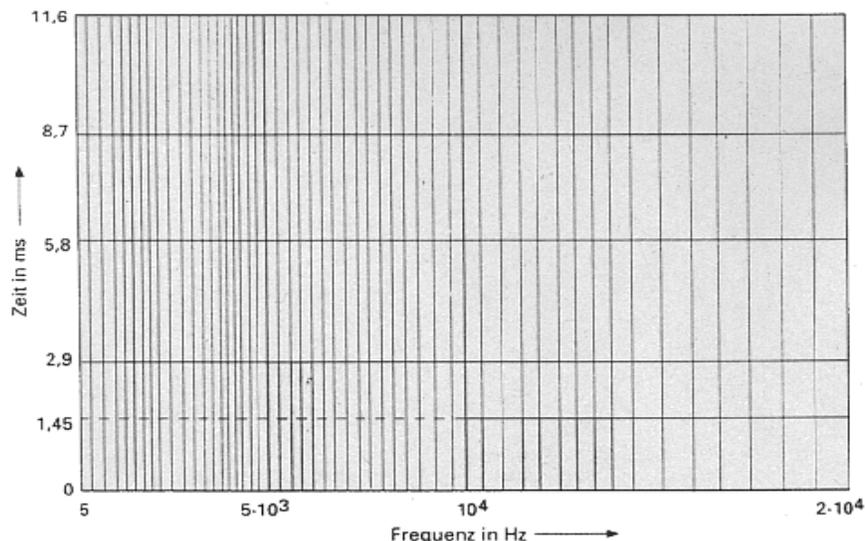


Abbildung 22: Die unterschiedliche Zeit- und Frequenzaufteilung bei ATRAC

### 3. Audioplayer

Audioplayer lassen sich in zwei Gruppen einteilen:

- Softwareplayer
- Hardwareplayer

Beide Gruppen haben Vor- und Nachteile. Da Softwareplayer am Computer programmiert werden, können neue Features erstellt bzw. alte Fehler schnell behoben werden. Softwareplayer sind daher sehr flexibel gestaltbar. Es gibt die Möglichkeit, sie durch eine Plug-In-Schnittstelle auszustatten. Dadurch können z.B. neue Formate in den Audioplayer miteinbezogen werden oder bestimmte Effekte (Visualisierungen, Equalizer,...) verwirklicht werden. Vorteil hierbei ist es, dass diese Plug-Ins nicht an den Hersteller des Audioplayers gebunden sind – jeder kann sich sein eigenes Plug-In schreiben. Der Audioplayer wird dadurch sehr umfangreich, was allerdings eine komplexe Menüführung erfordert. Sie sind deswegen oft umständlich zu handhaben. Ein weiterer Nachteil ist, dass bereits Hardware, sprich ein PC vorhanden sein muss. Im Gegensatz zu Hardwareplayern sind Softwareplayer langsam, da sie allgemein für PCs geschrieben wurden.

Hardwareplayer hingegen sind fest verdrahtet und auf ein bestimmtes Format optimiert. Sie erreichen dadurch optimale Geschwindigkeit. Die Funktionen eines Hardwareplayers kommen meist nicht an die eines Softwareplayers ran, da man bei der Bedienung (z.B. durch eine Fernbedienung) eingeschränkt ist. Zudem sind Hardwareplayer umständlich zu modifizieren. Die einzige Möglichkeit ist hierbei, eine Firmware zu verwenden, die geflasht werden kann. Diese Methode ist aber meist umständlich und wird von den meisten Käufern von Hardwareplayern nicht praktiziert.

### 4. Ausblick

Die hier beschriebenen Formate – bis auf Minidisc – speichern die PCM-Daten unkomprimiert auf das Medium. Um eine bessere Qualität der PCM-Daten zu erreichen, ist es möglich, die Abtastrate, die Kodewortlänge und die Anzahl der verwendeten Kanäle zu erhöhen. Ohne Komprimierung kommt man bei heutigen Medien schnell an die Grenzen. Alleine schon bei der Übertragung z.B. über das Internet empfiehlt es sich, die PCM-Daten zu komprimieren. In Zukunft wird es eine Kombination aus besserer Komprimierung und Erhöhung der Qualität der PCM-Daten geben.

Ein zusätzlicher wichtiger Aspekt ist der Kopierschutz. Bisherige Kopierschutze konnten immer geknackt werden.

Es gibt heute schon neue Techniken, wie DVD-Audio oder die Super Audio CD (SACD) die hauptsächlich die Qualität der PCM-Daten durch Erhöhung der Kanäle verbessern. Komprimierungsalgorithmen wie Mp3Pro oder Ogg Vorbis sind ebenfalls stark im kommen, sowie die Entwicklung neuer Medien mit mehr Kapazität wie die Blue Ray Disc.

## Literaturverzeichnis

STADLER, E.: Modulationsverfahren. 7., überarbeitete und erweiterte Auflage. Vogel  
Buchverlag Würzburg, 1993

JOHANN, J.: Nachrichtentechnik Band22: Modulationsverfahren. 1. Auflage. Springer Verlag  
Heidelberg, 1992

<http://www.nano.physik.uni-muenchen.de/elektronik/nav/k12t0.html>

[http://www-ra.informatik.uni-tuebingen.de/lehre/ss01/pro\\_internet\\_ausarbeitung/proseminar\\_flohr\\_ss2001.pdf](http://www-ra.informatik.uni-tuebingen.de/lehre/ss01/pro_internet_ausarbeitung/proseminar_flohr_ss2001.pdf)

[http://www.i-m.de/home/datenetze/ut\\_digi1.htm](http://www.i-m.de/home/datenetze/ut_digi1.htm)

<http://webster.fhs-hagenberg.ac.at/staff/schaffer/avt2/AVT2-SS03-AD-Wandlung.pdf>

<http://www.members.aol.com/torkraemer/adda/adda.html>

<http://nomac.e-technik.uni-ulm.de/lomiweb/lehre/praktikum/grundlagen4.html>

<http://homepages.fh-regensburg.de/~cuh39305/pe/pek12.pdf>

[http://www6.in.tum.de/lehre/seminare/ps\\_ws0203/bierbaum.pdf](http://www6.in.tum.de/lehre/seminare/ps_ws0203/bierbaum.pdf)

<http://media.mh-luebeck.de/wisspro/module/M-Medium-Audio/kapitel5>

<http://www.chscene.ch/ccc/contrib/netzmafia/multimedia/mm-31c.html>

<http://www.fet.uni-hannover.de/~purnhage/dat/dat.html>

[http://www.minidisc.org/aes\\_atrac.html](http://www.minidisc.org/aes_atrac.html)

<http://www.dvddemystified.com/dvdfaq.html>

<http://www.iar-80.com/page17.html>