

Ton

Physikalische Grundlagen

Digital Audio

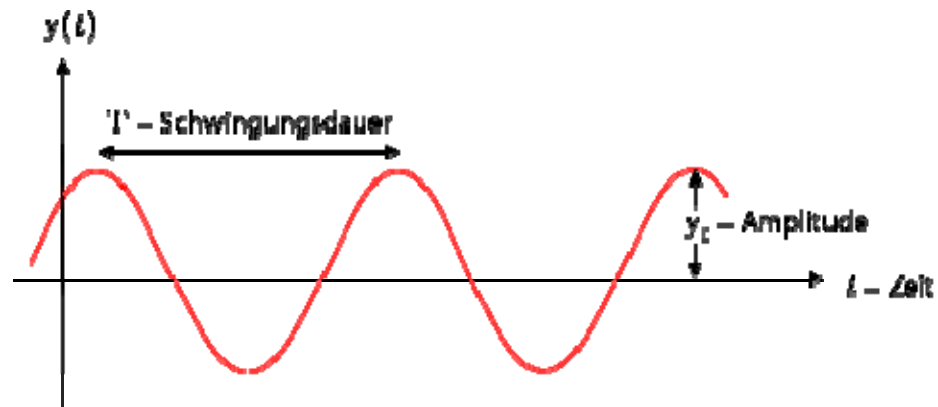
Grundlagen Audiodatenkompression

Grundlagen MPEG

Audiokompression: mp3

1. Physikalische Grundlagen

- Schallwellen haben wie jede andere Welle **vier** Eigenschaften:
 1. **Geschwindigkeit**
 2. **Amplitude** (Schwingsweite, Lautstärke)
 3. **Wellenlänge** (Länge, Schwingungsdauer)
 4. **Frequenz** (Die Häufigkeit der Schwingungen in einem bestimmten Zeitabschnitt)
Die Frequenz ist kein eigenständiges Attribut. Sie ist abhängig von der Wellenlänge.



1. *Geschwindigkeit von Schallwellen*

- Die **Geschwindigkeit** von Schallwellen ist im Wesentlichen abhängig von dem Medium, durch welche sie sich verbreiten.
- In Luft, auf Meereshöhe (=1 Atmosphäre Druck) und bei 20 °C beträgt die Schallgeschwindigkeit **343,8 Meter pro Sekunde**.

2. *Amplituden von Schallwellen*

- Die **Schwingungsweite** einer Schallwelle (= die Höhe des Wellenberges) wird als Lautstärke wahrgenommen.
- In Bewegung befindliche Moleküle der Luft treffen auf das Trommelfell und geben Druck an dieses weiter.

3. *Wellenlänge*

- = der kleinste Abstand zweier Punkte gleicher Phase einer Welle (z.B. Abstand zweier Wellenberge).
- Bei 22.000 Hz ist die Welle 1,56 cm lang, bei 20 Hz 17,19 m!

4. *Frequenz*

- Die **Frequenz** einer Welle gibt die **Anzahl der Schwingungen in einem Zeitintervall** an.
- Die Masseinheit der Frequenz ist **Hertz** (Hz). Sie misst Ereignis/Zeitraum.
- 1 Hz entspricht einer Schwingung (=einem Wellenberg **plus** Wellental) pro Sekunde: 1Hz = 1/s.
- Das menschliche Ohr ist empfindlich genug, um (in Abhängigkeit von Alter und Gesundheit) Töne zwischen 20 Hz und 22.000 Hz (22 kHz) zu unterscheiden (→ Bereich der hörbaren Frequenzen).
- Eine **Verkleinerung der Wellenlänge** (=Vergrößerung der Frequenz, „da so mehr Wellen in den gleichen Zeitintervall passen“) empfindet das menschliche Ohr als **Steigen der Tonhöhe**.
- Bsp.: Bei 22.000 Hz ist die Welle 1,56m lang, bei 20 Hz 17,19m.

Lautstärke

- die Schallintensität wird als Leistung pro Fläche definiert (W/m²)
- Bei der Messung von Lautstärke wird der Druck der Schallwelle gemessen und in elektrische Spannung umgewandelt, welche durch einen Zahlenwert dargestellt werden kann.
- Das menschliche Ohr ist sehr sensibel für Unterschiede in der Lautstärke. Das Lautstärkeempfinden steigt dabei nicht linear sondern **logarithmisch**
- Angaben über die Lautstärke werden auf einer logarithmischen Skala (zur Basis 10) wiedergegeben.
- Die Einheit auf dieser Skala heißt **Dezibel** (dB) und gibt das Verhältnis zweier Schallintensitäten an (Schalldruckpegel in dB):
dB = Logarithmus des Quotienten aus Intensität der Schallwelle und Intensität an der Hörschwelle:

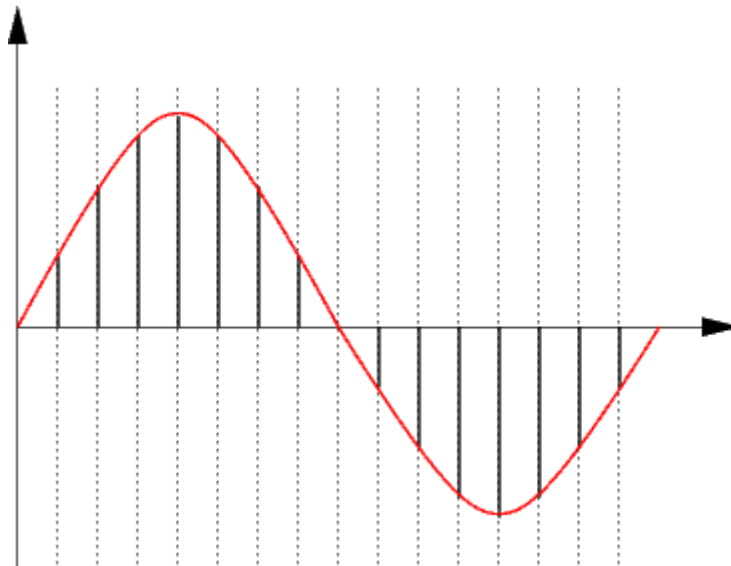
$$L_p = 20 \log_{10} \left(\frac{p_1}{p_0} \right) \text{ dB} = 10 \log_{10} \left(\frac{p_1^2}{p_0^2} \right) \text{ dB}$$

W/m ²	dB	Beispiel
10 ⁻¹²	0	Hörschwelle
10 ⁻¹⁰	20	Ruhige Wohnung
10 ⁻⁶	60	Unterhaltung
10 ⁻²	100	Laute Musik
10 ⁰	120	Donner, Flugzeug, Rennwagen- Cockpit
10 ²	140	Absolute Schmerzgrenze

2. Digital Audio

2.1 Sampling

- Spielt man Töne in ein Mikrophon, wird der Schalldruck in elektrische Spannung verwandelt, die **kontinuierlich** in der Zeit variiert. Diese Repräsentation von Schall durch elektrische Spannung heißt **analog**.
- Zur Digitalisierung von Ton müssen die kontinuierlichen Angaben **diskretisiert** werden, d.h. es werden elektrische Spannungswerte an vielen Zeitpunkten gemessen. Dieser Prozess heißt **Sampling** (ein Messpunkt= 1 Sample)



2.1.1 Sampling Raten (Abtastraten)

- Je häufiger die Schallwelle auf ihren Druck abgetastet wird, desto besser wird der originale Ton digital repräsentiert.
- Zu wenige Abtastpunkte führen zu geraden Linien, zu langen Wellen oder zu gezackten Kurven.

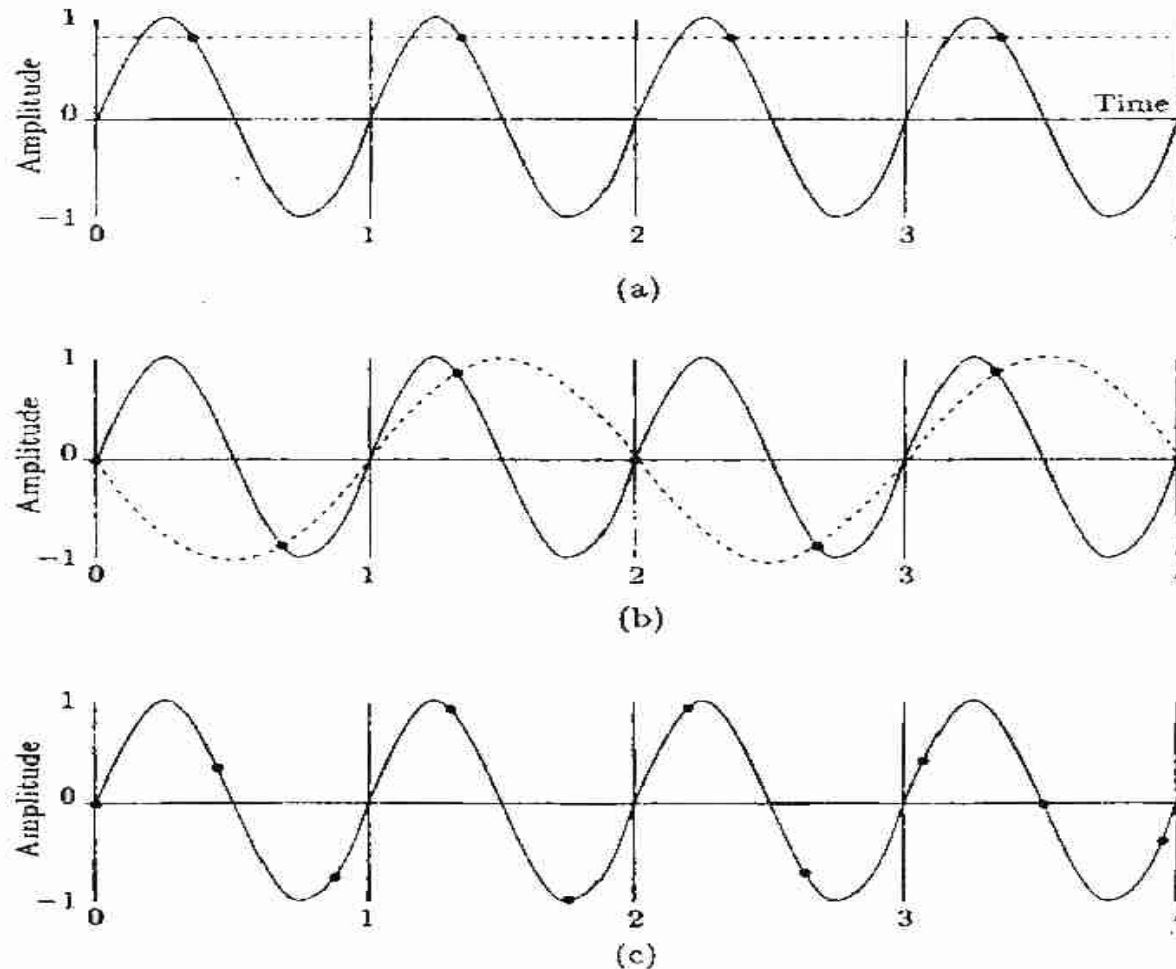


Figure 6.3: Sampling a Sound Wave.

- Um ungenaues Abtasten der Schallwellen zu vermeiden, ist es notwendig, die Töne mit etwas mehr als der **Nyquist-Rate** abzutasten, die dem Doppelten der maximalen enthaltenen Frequenz entspricht.

$$f_{\text{abtast}} > 2 f_{\text{max}}$$

- Da menschliche Ohren Schallwellen von maximal 22.000 Hz (= 22 kHz) wahrnehmen können, liegt die erforderliche Nyquist-Rate bei 44.000 Hz.
- Qualitativ hochwertige Audio-Digitalisate haben deswegen eine Abtastrate von 44.100 Hz.
- Beispiele:
 - Analoge Telefone sampeln Schallwellen mit 8.000 Hz.
 - MW Radioqualität entspricht einer Abtastrate von 11.025 Hz.
 - UKW Radioqualität wird mit einer Abtastrate von 22.050 Hz erzielt.
 - DAT (Digital Audio Tape) hat eine Abtastrate von 48 kHz.
 - Professionelle Audiogeräte sampeln mit einer Rate von 96 kHz

2.2 Digitalisierungsschritte

Die Digitalisierung von analogen Audio-Signalen läuft grob in drei Schritten ab:

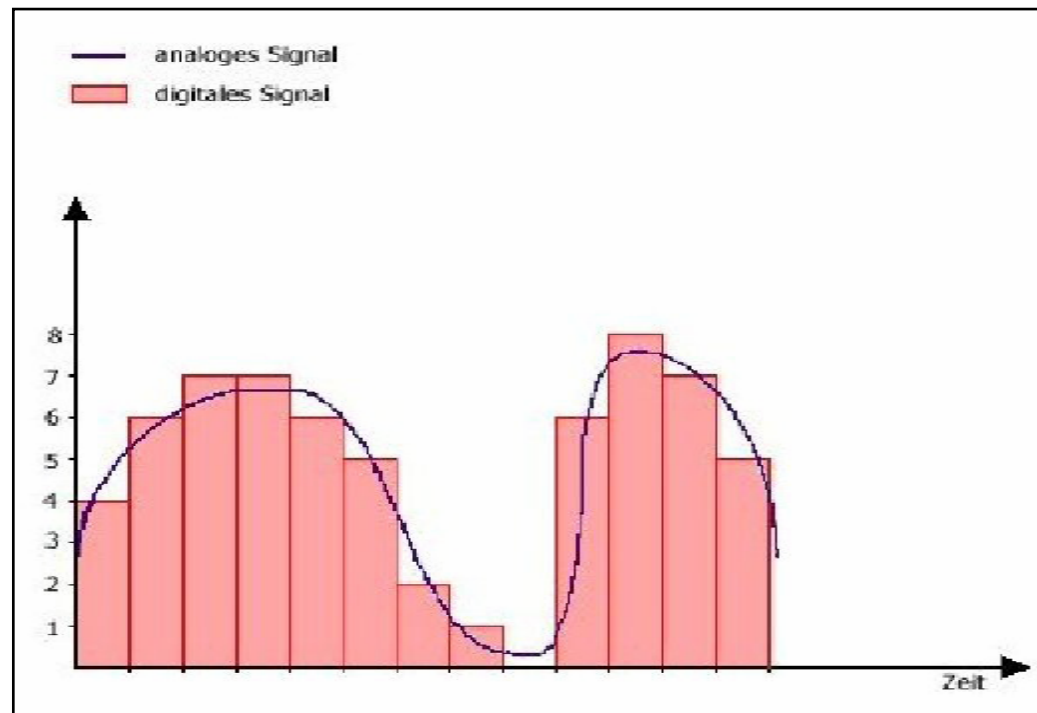


2.2.1 Abtastung

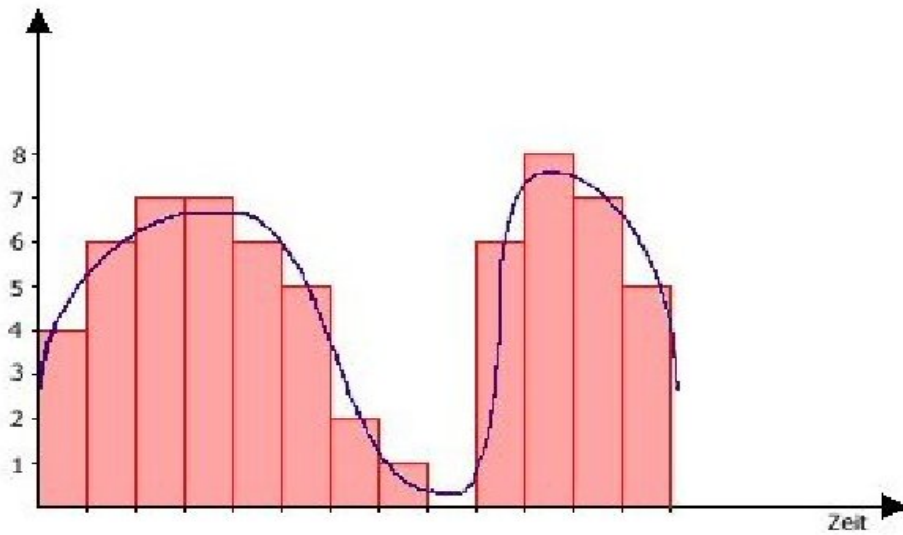
- Um digitale Signale aufzuzeichnen gibt es verschiedene Verfahren. Das bekannteste (und einfachste) ist *Pulse Code Modulation (PCM)*.
- Dabei wird für jeden Abtastzeitpunkt ein Signalwert gemessen.
- Bekannte Audio-Formate die PCM verwenden sind *WAV und AIFF Files*.
- Neben anderen verlustfreien Verfahren (z.B. PAM, PWM) wird beim verlustbehafteten **DPCM** (Differential Pulse Code Modulation) - Verfahren, nur die Differenz zum vorhergehenden Wert gespeichert. Man erreicht so eine Datenkompression.

2.2.2 Quantisierung

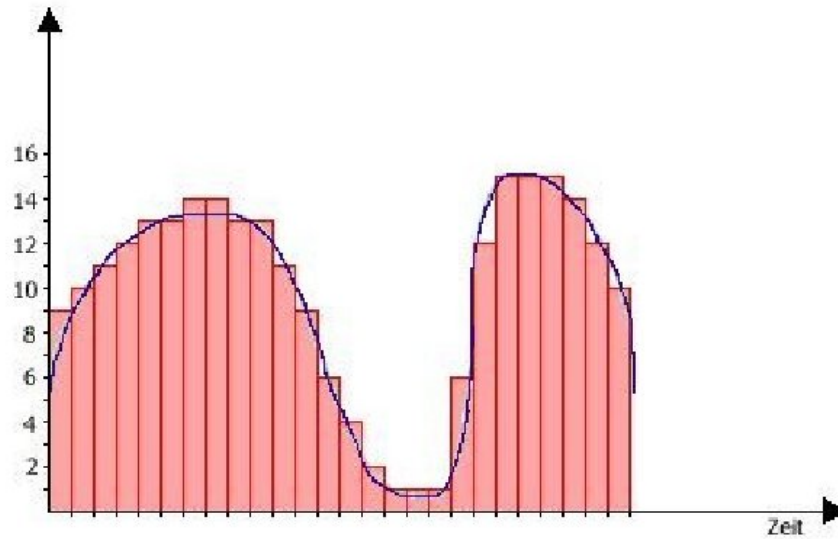
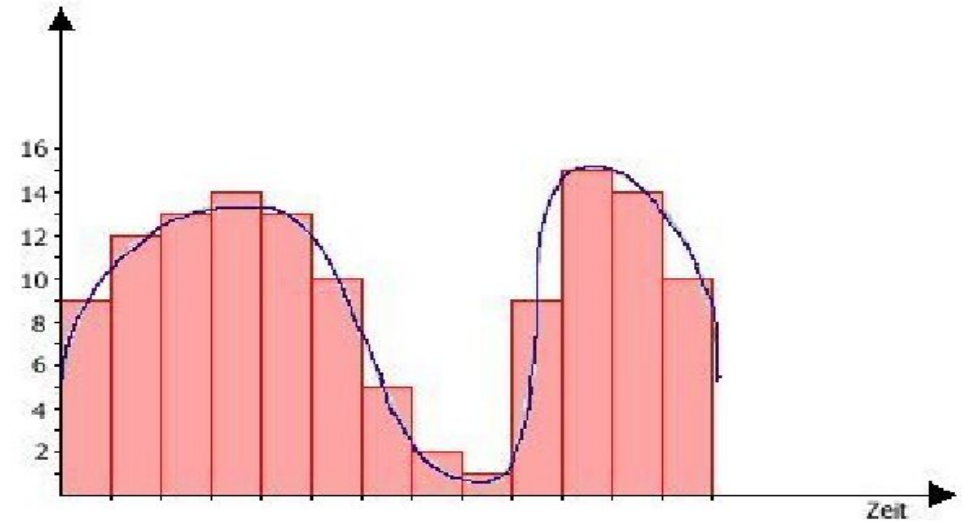
- Bei der **Quantisierung** wird für die einzelnen abgetasteten Signale ein Zahlenwerte ermittelt.
- Für die **Genauigkeit** des Wertes spielt dabei die **Abtasttiefe** (sample size) eine Rolle
(→ Anzahl der Bits, die bei der Speicherung des Werts verwendet werden).
Sie bestimmt, in wievielen **Abstufungen** das analoge Signal repräsentiert werden kann.
- **Quantisierungsfehler**: Da ein analoges Signal nicht mit absoluter Genauigkeit gemessen werden kann, treten Rundungsfehler auf.
Diese **Rundungsfehler** können für Störgeräusche bei der Wiedergabe einer digitalen Aufnahme sorgen, dem **Quantisierungsrauschen**.



— analoges Signal
■ digitales Signal



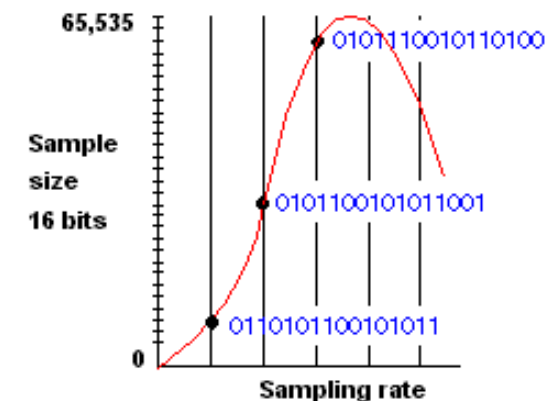
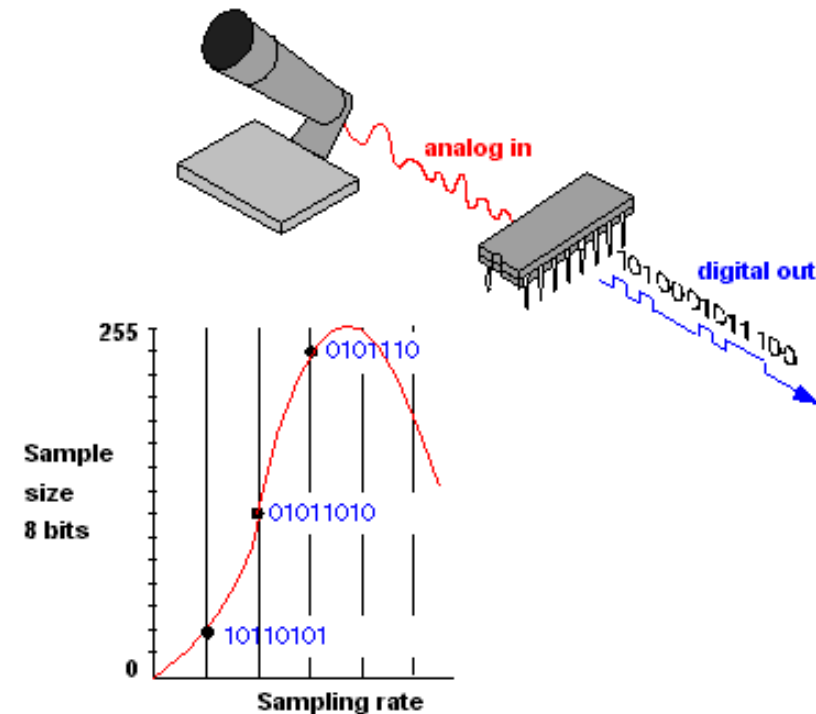
— analoges Signal
■ digitales Signal



Abtasttiefe (Sample size)

From Computer Desktop Encyclopedia
© 1998 The Computer Language Co. Inc.

- Angenommen, die elektrische Spannung, die einen digitalisierten Ton repräsentieren kann, beträgt 1 Volt, dann können mit 8 Bit 1/256 Volt (ca. 0,004 Volt) unterschieden werden.
- Diejenigen Töne, die weniger als 2 mv an Spannung erzeugen, würden als Null gesampelt und als Stille wiedergegeben.
- Ein 16-Bit-Sample könnte Töne bis 1/65.535 (ca. 15 μ v) unterscheiden.
- Wir können uns daher die Abtasttiefe als eine Quantisierung der originalen Audiodaten vorstellen. 8-Bit-Samples werden ungenauer quantisiert als 16-Bit-Samples. Daraus ergibt sich eine bessere Kompression, aber eine schlechtere Wiedergabe.



2.2.3 Kodierung

- Die quantisierten Signale werden zuletzt in ein einer Bitfolge kodiert.
- *Analog-Digital-Wandler (Analog-Digital-Converter oder ADC)* heißen die Geräte, die die Kodierung (und Quantisierung) übernehmen.
- Die Umwandlung der digitalen Daten zurück in elektrische Spannungen übernimmt der *Digital-Analog-Wandler (DAC)*.

2.2.3 Kodierung

- Die quantisierten Signale werden zuletzt in ein einer Bitfolge kodiert.
- *Analog-Digital-Wandler (Analog-Digital-Converter oder ADC)* heißen die Geräte, die die Kodierung (und Quantisierung) übernehmen.
- Die Umwandlung der digitalen Daten zurück in elektrische Spannungen übernimmt der *Digital-Analog-Wandler (DAC)*.

2.3 Datenmengen bei der Audiodigitalisierung

Wie groß ist der Speicherplatzbedarf einer klassischen Aufnahme in CD-Qualität (Stereo) mit 60-minütiger Laufzeit?

2.2.3 Kodierung

- Die quantisierten Signale werden zuletzt in ein einer Bitfolge kodiert.
- *Analog-Digital-Wandler (Analog-Digital-Converter oder ADC)* heißen die Geräte, die die Kodierung (und Quantisierung) übernehmen.
- Die Umwandlung der digitalen Daten zurück in elektrische Spannungen übernimmt der *Digital-Analog-Wandler (DAC)*.

2.3 Datenmengen bei der Audiodigitalisierung

Wie groß ist der Speicherplatzbedarf einer klassischen Aufnahme in CD-Qualität (Stereo) mit 60-minütiger Laufzeit?

Das Audiosignal wird standardmäßig mit 44,1 KHz pro Kanal abgetastet.
Die Abtasttiefe liegt bei 16 bit.

2.2.3 Kodierung

- Die quantisierten Signale werden zuletzt in ein einer Bitfolge kodiert.
- *Analog-Digital-Wandler (Analog-Digital-Converter oder ADC)* heißen die Geräte, die die Kodierung (und Quantisierung) übernehmen.
- Die Umwandlung der digitalen Daten zurück in elektrische Spannungen übernimmt der *Digital-Analog-Wandler (DAC)*.

2.3 Datenmengen bei der Audiodigitalisierung

Wie groß ist der Speicherplatzbedarf einer klassischen Aufnahme in CD-Qualität (Stereo) mit 60-minütiger Laufzeit?

Das Audiosignal wird standardmäßig mit 44,1 KHz abgetastet.

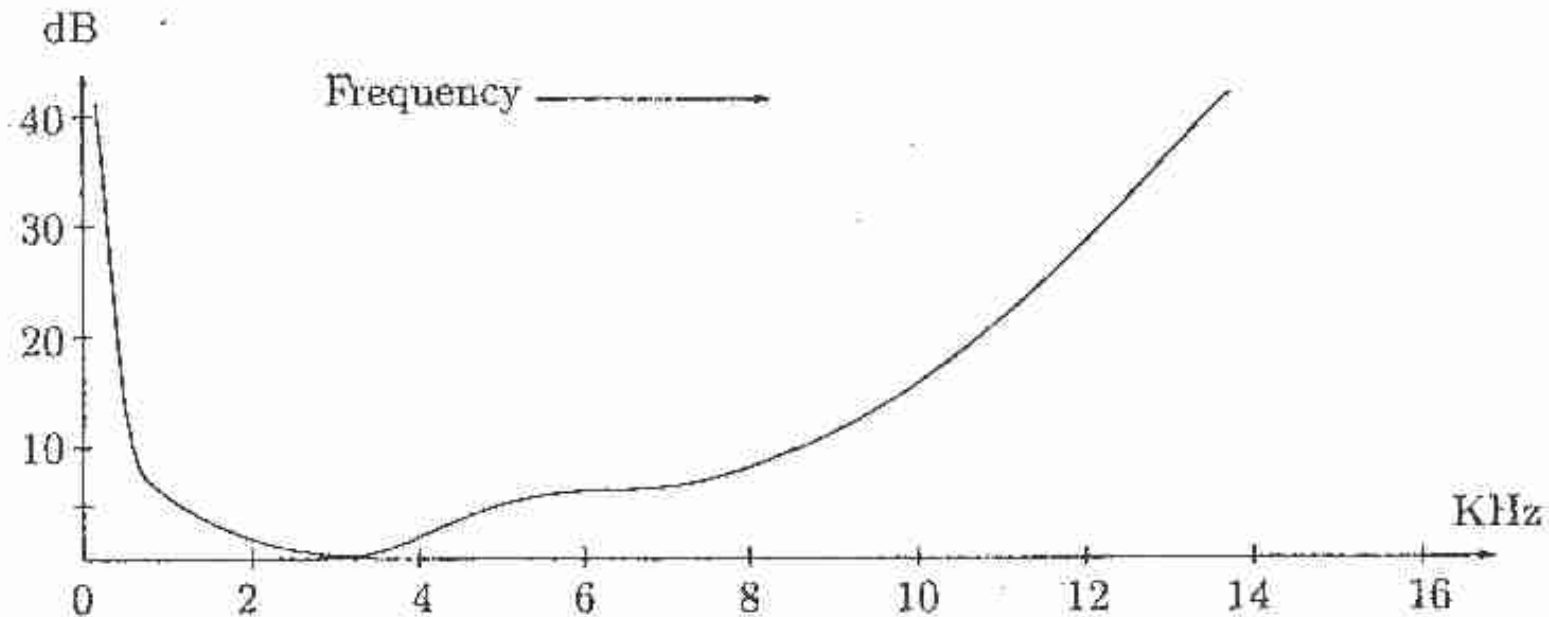
Die Abtasttiefe liegt bei 16 bit.

Daraus folgt: $44,1\text{kHz} * 2 * 16\text{bit} * 60 * 60 \text{ s} = 635 \text{ Megabyte}$

3. Grundlagen der Audiodatenkompression

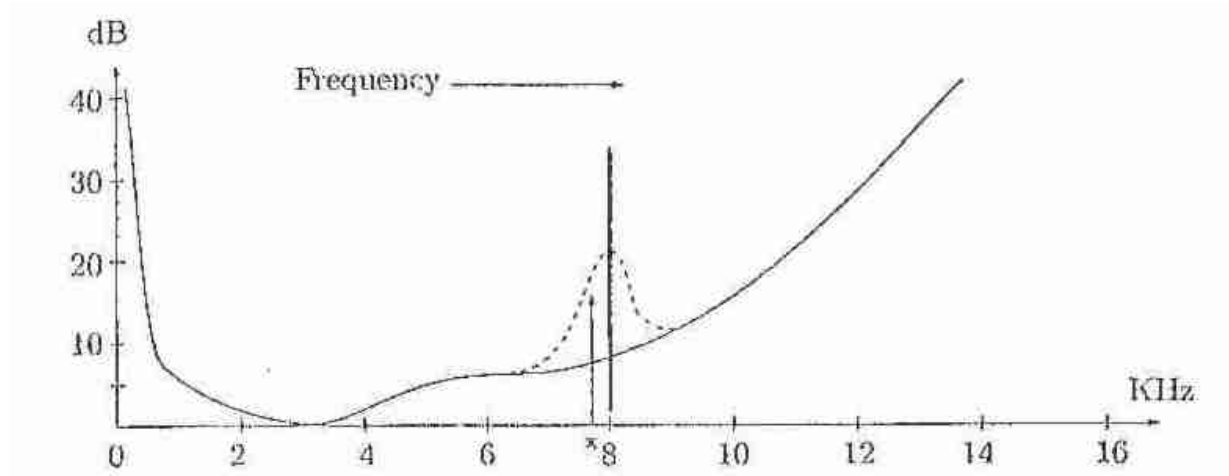
3.1 Einflüsse des menschlichen Hörvermögens: Hörschwelle

- Das menschliche Ohr ist zwar in der Lage, Tonhöhen von 20 Hz bis 22 kHz wahrzunehmen, doch diese Fähigkeit ist nicht gleichmäßig ausgeprägt.
- Die Empfindlichkeit des Ohres ist im Bereich von 2 kHz bis 4 kHz am größten (→ es reicht eine geringe Lautstärke, um den Ton zu hören).
- Das Vorhandensein dieser Hörschwellen ist ein Ansatzpunkt für die Datenkompression.



3.2 Frequenz-Maskierung

- Gemeinsam erklingende Töne ähnlicher Frequenz beeinflussen die Hörschwelle des jeweils anderen Tons.
 - Der lautere Ton maskiert den leiseren Ton, indem die Hörschwelle für den leiseren Ton erhöht wird.
 - Selbst wenn der leisere Ton ein so starkes Signal erzeugt, dass er normalerweise über der Hörschwelle liegen würde, kann dieser (maskierte) Ton eventuell doch nicht gehört werden.
 - Gute Kompressionsmethoden sollten diese Fälle erkennen und das Signal löschen.
- **Ein Beispiel** ist das Ticken eines Weckers. Bei Stille ist es gut hörbar, klingelt der Wecker, kann man es nicht mehr wahrnehmen, obwohl der Wecker weitertickt. Das Ticken wird also durch das Klingeln maskiert.



3.3 Temporäre Maskierung

- Auch zeitliche Abstände zwischen Tönen haben einen Einfluß auf die Hörbarkeit.
- Geht einem starken Ton A ein schwächerer Ton B in ähnlicher oder gleicher Frequenz voraus oder folgt ihm, so gibt es zeitliche Abstände von bestimmter Dauer, in denen Ton B nicht zu hören ist.

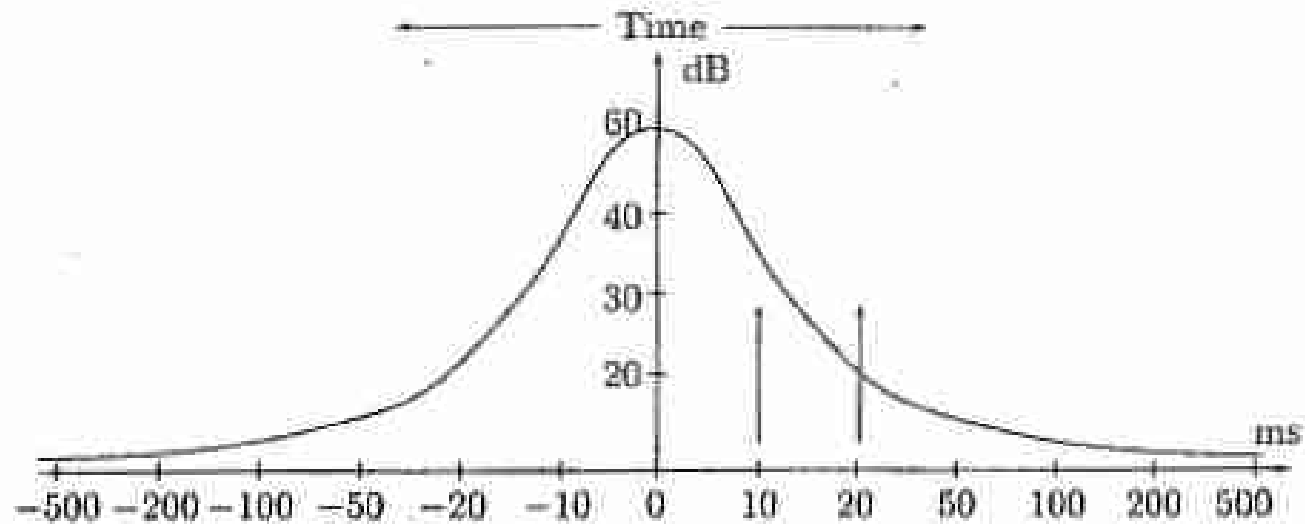


Figure 6.6: Threshold and Masking of Sound.

Überblick: MPEG

Audiokompression: mp3

1. Allgemeines zu MPEG

- **MPEG** (Moving Picture Experts Group) ist eine Arbeitsgruppe von ISO (International Standards Organization) und IEC (International Electro-Technical Commission) .
- Die Arbeitsgruppe wurde 1988 in Ottawa, CA gegründet.
- Ihre Aufgabe ist die Entwicklung von internationalen Standards zur Komprimierung, Dekomprimierung, Verarbeitung, Kodierung und Dekodierung von bewegten Bildern und assoziiertem Audio. Der offizielle Titel lautet: „Coding of moving pictures and audio“.
- Die Expertengruppe trifft sich in turnusmäßigen Abständen, um Arbeitsergebnisse zu diskutieren und neue Arbeitsbereiche abzustecken. Das Ergebnis sind verschiedene Standards: MPEG-1, MPEG-2, MPEG-4, MPEG-7 und MPEG-21 (in Arbeit).

1.1 MPEG-1

- MPEG-1 ist seit 10/92 als Standard ISO/IEC 11172 definiert.
- Praktisch bedeutete dies, einen Standard für die effiziente Speicherung und Wiedergabe von audio-visuellen Daten auf CDs zu erarbeiten, der eine max. Datenübertragungsrate von 1,5 Mbit/s unterstützt.
„Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s“

1.1.2 MPEG-1 Definitionen

- MPEG-1 besteht aus 2 großen Teilen:
 - Video
 - **Audio**
- Wir beschäftigen uns hier zunächst mit dem Audio-Teil.

1.1.3 Anwendungen von MPEG-1

- **VCD** (Video CD) , Digital Compact Cassette (DCC)
- **MPEG-1, Teil 3, Layer III** wird populär als **mp3** bezeichnet.
- Zum Erzeugen/Abspielen/Darstellen von MPEG-Dateien braucht man sogenannte **Codecs**.
Codec ist die Abkürzung für **Coder/Decoder**, also die Soft- oder Hardwarebestandteile, die die Video- und Audiodaten im entsprechenden Format kodieren oder dekodieren.

MPEG Audio:

- Der "Audio"-Teil beschreibt die Kodierung von Audiodaten (Mono/Stereo).
- In diesem Teil des Standards werden drei verschiedene Methoden (layer) der Datenkompression definiert:
- Die drei Methoden werden mit I, II und III bezeichnet und zeichnen sich durch **steigende Komplexität** und **Leistungsfähigkeit** aus.
- Die Layer sind hierarchisch kompatibel, d.h. die Funktionen der "niedrigeren" Methoden sind den "höheren" bekannt, die höheren Layer verfügen allerdings über zusätzliche Eigenschaften, die in besserer Kompression resultieren.
 - Layer I: ermöglicht eine Datenreduktion von 1:4. Dies entspricht 384 kBit/s für ein Stereosignal.
 - Layer II: ermöglicht eine Datenreduktion von 1:6...1:8. Dies entspricht 256...192 kBit/s für ein Stereosignal.
 - Layer III: ermöglicht eine Datenreduktion von 1:10...1:12. Dies entspricht 128...112 kBit/s für ein Stereosignal.

1.2 MPEG-2

- Die nächste Entwicklung der Arbeitsgruppe war MPEG-2. Der offizielle Titel lautet: „Generic Coding of Moving Pictures and Associated Audio“.
- MPEG-2 ist als Standard ISO/IEC 13818 definiert.
- Wurde überwiegend für die Speicherung und Übertragung von Videosignalen vorgesehen (digitales Fernsehen). Übertragungsraten: 1,5 - 15 Mbit/s.

1.2.2 Anwendungen von MPEG-2

- Video-Teil von DVDs
- AAC ist von Japan als Standard für nationales digitales Fernsehen gewählt worden.
- Digitale Fernsehübertragungen

1.3 MPEG-4, MPEG-7, MPEG-21

- Eine weitere Entwicklung der Arbeitsgruppe ist **MPEG-4**. Der offizielle Titel lautet: „Coding of audio-visual objects“.
- **MPEG-4** ist als Standard ISO/IEC 14496 definiert. Die Version 1 wurde 10/98 verabschiedet, Version 2 12/99.
- Während MPEG-1 und -2 vor allem für Audio- und Video-Kompression benutzt werden, ist das **Ziel von MPEG-4** einen **Standard für Multimediaanwendungen** zu schaffen mit Hilfe dessen Computer, Fernsehen und Telekommunikation zusammenwachsen können (Internet übers Handy, interaktives Fernsehen, Video on demand, Fernseh-Liveübertragungen über Internet etc.);
- **MPEG-7** fokussiert im Gegensatz zu MPEG-4 nicht auf die Kodierung und Darstellung von audio-visuellen Objekten, sondern auf die Beschreibung von Multimedia-Inhalten (Multimedia Content Description Interface) Ziel: schnelle Suche, Identifizierung und Verarbeitung von multimedialen Inhalten.
- **MPEG-21** ist die jüngste Entwicklung von MPEG und noch in Arbeit. Die jüngste Entwicklung der Arbeitsgruppe ist MPEG-21. Der offizielle Titel lautet: *Multimedia Framework*.
- **MPEG-21** zielt ab auf eine zukünftige Umgebung, in der Inhalte unterschiedlichster Art von Usern verschiedener Kategorien über multiple Anwendungen und Domänen bewegt werden können.

2. Audiokompression: MPEG/ mp3

2.1 Charakteristik

- **mp3** ist eine andere (populäre) Bezeichnung für eine Realisierung der Spezifikation des MPEG-1, Audio Layer III, vom Fraunhofer Institut Anfang der 90er Jahre entwickelt.
- Der Kodierer (encoder) ist extrem komplex, dafür aber extrem leistungsfähig in Bezug auf die Kompressionsraten.
- Der Dekodierer dagegen ist viel einfacher aufgebaut und zudem abwärtskompatibel.

2.2 Grundsätzliche Funktionsweise von MPEG-1 Audio Layer III

- Das Kompressionsverfahren orientiert sich generell an den Gegebenheiten des menschlichen Gehörs, d.h. dessen Wahrnehmungsmöglichkeiten von Tönen (*perceptual audio coder*). Ziel ist es also, Datenreduktion unter Beibehaltung der subjektiven Klangqualität zu erreichen.
- Das psychoakustische Phänomen der Tonmaskierung wird dabei besonders ausgenutzt.

- **Exkurs: Das Gehör**

- Das **Gehörorgan**, auch Cochlearorgan genannt, befindet sich im Innenohr. Dort wird eine **Frequenzanalyse** der einfallenden Schallwellen vorgenommen: Verschiedene Bereiche des Gehörs werden durch verschiedene Frequenzen angeregt. Das Gehör kann so Frequenzen von ungefähr 20Hz bis 20kHz erfassen (mit zunehmendem Alter oft nur bis 16kHz).
- Um also die durch Maskierung unhörbaren Daten zu erkennen, muss man als ersten Schritt das Audiosignal ebenfalls in den Frequenzbereich transformieren.

- **Exkurs: Kritische (Frequenz-)Bänder**

- Empirische Untersuchungen haben ergeben, dass die bestimmten Frequenzen zugeordneten Bereiche nicht fließend ineinander übergehen, sondern dass es sogenannte **kritische Bänder** gibt, die jeweils bestimmte **Frequenzbereiche** abdecken. Frequenzen innerhalb eines kritischen Bandes werden zusammen wahrgenommen, deshalb sind dort **Maskierungseffekte** besonders ausgeprägt. Ein lauter Ton in einem kritischen Band bewirkt, dass man andere Frequenzen im gleichen Band nur sehr schlecht hören kann.
- Die kritischen Bänder sind nicht gleichbreit, sondern werden mit zunehmender Frequenz breiter. Damit verbunden wird auch die **Auflösung für höhere Frequenzen schlechter**. Am dichtesten liegen die kritischen Bänder im Bereich von 2-5kHz, was ungefähr der Frequenz der menschlichen Sprache entspricht.
- Das heißt: Kritische Bänder sollten bei der Maskierung zusammen behandelt werden und das Frequenzspektrum sollte möglichst nach der Lage der kritischen Bänder unterteilt werden.

2.3 Teilschritte der mp3-Kodierung

- Das Audiosignal wird in bestimmten Zeitintervallen abgetastet. Dabei wird das Signal auf **Frequenzbänder** abgebildet.
- Anhand eines **psychoakustischen Modells** wird für jedes dieser **Frequenzbänder** getestet, ob Signale innerhalb der jeweiligen Frequenzgruppe durch Maskierungseffekte überlagert werden. Hierfür ermittelt man für einen Zeitpunkt **t** für ein Signal dessen **Maskierungskurve**.
- Alle Signale, die zum gleichen Zeitpunkt **t** unterhalb dieser Kurve liegen, werden entfernt, da das Gehör die betroffenen Signale nicht wahrnehmen kann.
- Anschließend werden die so ermittelten Frequenzdaten **quantisiert**, d.h. entsprechend der Abtasttiefe (8-Bit, 16 Bit...) gerundet und gespeichert.
- Die Anzahl der verwendeten Bits wird durch die vorgegebene **Bitrate** beschränkt:
 - Die Bitrate oder Datenübertragungsrate bezeichnet die Menge der Daten, die innerhalb einer Zeiteinheit übertragen werden kann (z. B. kBit/s). Die Bitrate hängt bei MP3 vom Kompressionsfaktor ab: Je geringer die Datenkompression, desto mehr Daten können übertragen werden. MP3-Dateien mit einer hohen Bitrate weisen eine entsprechend geringe Kompression auf. Daraus ergibt sich, dass bei niedriger Kompression die Dateien größer werden, die Bitrate und die Qualität aber steigen.

2.4 mp3-Kodierer

- **mp3** verwendet für die Frequenztransformation eine sog. Filterbank. Das Signal wird dabei in 32 gleich lange Subbänder aufgeteilt.
- Da sich die Subbänder überlappen kann ein Ton einer best. Frequenz zwei Bänder beeinflussen. Deshalb wird zusätzlich eine modifizierte diskrete Kosinustransformation (MDCT) durchgeführt. Dieses mathematische Verfahren optimiert die Übergänge zwischen Frequenzbändern, so dass beim Übergang keine Klangstörungen entstehen.
- Im psychoakustischen Modell, das der mp3-Kodierer verwendet, werden einige Verfahren benutzt, die für die hohe Leistungsfähigkeit maßgeblich verantwortlich sind: z.B. die Erkennung von tonalen und nicht-tonalen Komponenten; Basis sind experimentelle Untersuchungen der Wahrnehmungsfähigkeiten des Gehörs.
- **Quantisierung**
- Die quantisierten Signale werden in sog. *frames* gepackt, wobei Huffman-Kodierung verwendet wird.

2.4.1 Weitere Komprimierungsstrategien

Bit Reservoir

- In mp3 wurde außerdem ein sogenanntes Bitreservoir implementiert. In dieses können die bei der Kodierung eines Frames nicht benötigten Bits gepackt werden. Im darauf folgenden Frame steht das Bitreservoir dann zusätzlich zum Kodieren zur Verfügung. Dies Reservoir ist allerdings nicht sehr groß, es umfasst nur 4088 Bit. Um wirklich adaptiv auf das Eingangssignal reagieren zu können, ist die variable Bitrate besser geeignet.

Variable Bit Rate (VBR)

- Im Gegensatz zur Kodierung mit konstanter Datenrate (<--> constant bit rate, **CBR**) werden hierbei je nach Komplexität des zugrundeliegenden Materials einzelne Zeitabschnitte unterschiedlich stark komprimiert (d.h. mit unterschiedlicher Bitrate gespeichert), um gleich bleibende Qualität bei möglichst niedrigem Datenvolumen erzielen zu können.
(→ leise Stellen vor „Hidden Tracks“)

Joint Stereo coding

- Joint Stereo coding bezeichnet verschiedene Tools, die die Datenmenge weiter reduzieren:
 - Intensity Stereo
 - Mid/Side (M/S) stereo

Intensity Stereo:

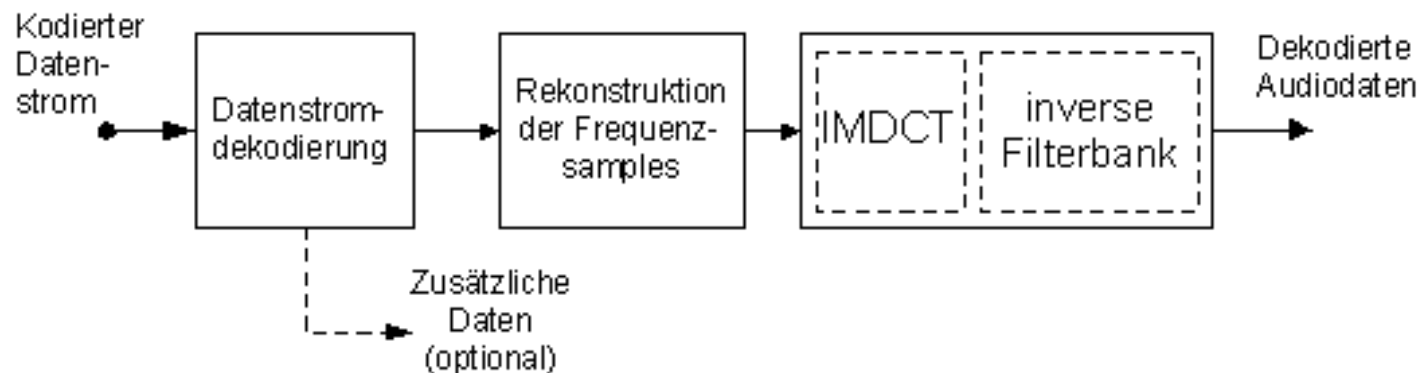
- Hier wird die Schwäche des menschlichen Ohrs ausgenutzt, bei besonders hohen und tiefen Tönen keine genaue Lokalisierung mehr vornehmen zu können.
- Die Daten werden in diesen Fällen nicht als zwei Signale (stereo) abgelegt, sondern als ein Signal (mono) mit ein paar Zusatzinformationen, um ein Minimum an "räumlicher" Information zu rekonstruieren.

Middle/Side (M/S) stereo:

- Wenn die Stereo-Informationen auf beiden Kanälen einander sehr ähnlich sind, werden die Informationen nicht als zwei getrennte Signale gespeichert, sondern als "Mitte" und "Seite".
- Mitte bedeutet, dass die Kanalinformationen addiert werden ($L+R$), Seite nimmt die Differenz der Informationen ($L-R$) auf.

2.5 mp3-Dekodierer

- Der **Dekodierer** ist weit einfacher konstruiert als der Kodierer: Psychoakustisches Modell und Bit-Allokation werden bei der Dekodierung nicht benötigt.
- Zur Dekodierung eines MP3-Datenstroms müssen lediglich die kodierten Frequenzsamples rekonstruiert werden, die dann durch eine inverse modifizierte Cosinus-Transformation (IMDCT) und eine inverse Filterbank wieder in den **Zeitbereich*** zurückverwandelt werden.



* Geht es um die Bearbeitung von Sound, lassen sich grundsätzlich **zwei Perspektiven** unterscheiden:

- 1) Entweder das Audiosignal wird als eine Serie von Samples betrachtet, die einem analogen Signal entsprechen. Oder
- 2) die Unterscheidung erfolgt nicht über die Zeit, sondern über die Frequenzen. Je nach Art der gewünschten Manipulation eignet sich die eine oder andere Perspektive besser. Das Mischen von Signalen oder das Erhöhen der Amplitude geschieht leichter in einer zeitorientierten Basis. Frequenzmanipulationen fallen naturgemäß bei der nach Frequenzen aufgeschlüsselten Sichtweise leichter.

Um die Daten von einer in die andere Perspektive zu konvertieren, bedient man sich der **Fourier-Transformation**

2.6 *mp3-Files*

- Der vom Encoder erzeugte Datenstrom wird in eine mp3-Datei geschrieben. Dieser Datenstrom besteht in der Grundstruktur aus *frames* (jeder MPEG-Datenstrom, nicht nur mp3).

2.6.1 *mp3 Format*

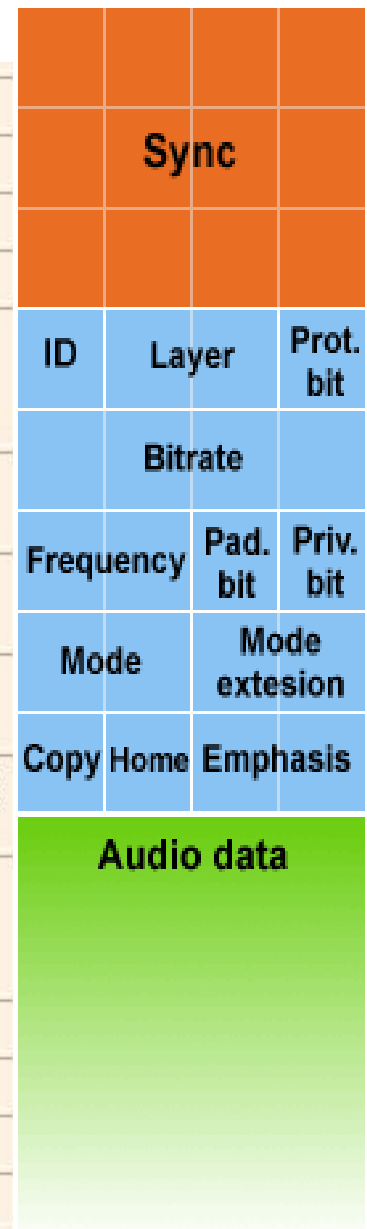
- Jedes Frame nimmt 36 Signale pro Subband (→ 32 Stück), also 1152 Samples auf.

- Ein mp3-Audio-Frame wiederum ist aufgebaut aus
 - Header
 - Kontrollblock (CRC = Cyclic Redundancy Code, optional)
 - Eigentliche Audiodaten
 - Hilfsdaten („side information“) → Infos zur Huffman-Kodierung

Header (32)	CRC (0,16)	Side Information (136,256)	Main Data	Ancillary Data
-----------------------	----------------------	--------------------------------------	------------------	-----------------------

MPEG Audioframe

Feld	Bit	Bemerkungen
syncword	12	Wert immer 1111 1111 1111
ID	1	1 für MPEG-Audio, 0 ist reserviert
Layer	2	'11' für Layer I; '10' für Layer II; '01' für Layer III; '00' ist reserviert
protection_bit	1	zeigt an, ob Redundanz in den Audio Bitstream eingefügt wurden, um Fehlererkennung und -beseitigung zu ermöglichen; '1' = keine Redundanz hinzugefügt, '0' Redundanz hinzugefügt.
bit_rate_index	4	zeigt die Bitrate unabhängig vom Aufnahmemodus an; die Werte werden in einer speziellen Tabelle definiert.
sampling_frequency	2	zeigt die Sampling-Frequenz an; '00' für 44.1 kHz, '01' für 48 kHz, '10' für 32 kHz, '11' ist reserviert
padding_bit	1	wird auf '1' gesetzt, wenn Bits aufgefüllt werden mussten, um die Bitrate an die Sampling-Frequenz anzupassen.
private_bit	1	für den privaten Gebrauch; das Bit wird in der Zukunft von ISO nicht belegt.
mode	2	zeigt den Aufnahmemodus an; '00' für stereo, '01' für joint_stereo (intensity_stereo and/or ms_stereo), '10' für dual_channel, '11' für single_channel
mode_extension	2	als Spezifikation der Mode-Angaben
copyright	1	'1' für Copyright-Schutz vorhanden, '0' ohne Copyright-Schutz
original/home	1	'1' für Original, '0' für Kopie
emphasis	2	zeigt die Art des De-emphasis an, die verwendet werden sollte



Bitraten

- '0000' zeigt in allen Layern an, dass ein "freies" Format benutzt wird, welches nicht in der Liste der Bitraten spezifiziert sein muss.
- In Layer II sind nicht alle Kombinationen von Bitraten und Modi erlaubt.
- Die Angaben sind jeweils in kBit/s.

Bitrate value	MPEG-1, layer I	MPEG-1, layer II	MPEG-1, layer III	MPEG-2, layer I	MPEG-2, layer II und III	
0 0 0 0						
0 0 0 1	32	32	32	32	8	
0 0 1 0	64	48	40	48	16	
0 0 1 1	96	56	48	56	24	
0 1 0 0	128	64	56	64	32	
0 1 0 1	160	80	64	80	40	
0 1 1 0	192	96	80	96	48	
0 1 1 1	224	112	96	112	56	
1 0 0 0	256	128	112	128	64	
1 0 0 1	288	160	128	144	80	
1 0 1 0	320	192	160	160	96	
1 0 1 1	352	224	192	176	112	
1 1 0 0	384	256	224	192	128	
1 1 0 1	416	320	256	224	144	
1 1 1 0	448	384	320	256	160	
1 1 1 1	bad	bad	bad	bad	bad	bad

2.6.2 ID3 Tag

- Zur Angabe von Metadaten ist ein zusätzlicher Standard entwickelt worden, ID3 (aktuell ID3, version 2.4.0).
- ID3 tag richtet sich in erster Linie an MPEG-1- und -2-Dateien, kann aber auch mit anderen Audioformaten oder eigenständig verwendet werden.
- Funktionsprinzip: ID3 ist wie MPEG aus frames aufgebaut, die in die MPEG-Datei geschrieben werden können.
- Mit ID3v2 lassen Informationen mit bis zu 256 Mb Kapazität aufnehmen, neben Text auch Bilder oder sogar Files.

Aufgaben

- 1) **Wiederholen** Sie den Stoff dieser Sitzung **bis zur nächsten Sitzung** (siehe dazu den Link zur Sitzung auf der HKI-Homepage). Informieren Sie sich zusätzlich durch eigene Literaturrecherche!
- 2) Beantworten Sie die Fragen aus der Sammlung „**beispielhafte Klausurfragen**“ zum Bereich Ton (soweit in dieser Sitzung behandelt).