

MP3-Grundlagen: Psychoakustik

Verschiedene Schallereignisse sind vom menschlichen Hörsinn („Psycho-Ohr“) nicht wahrnehmbar, weil sie von der Physik des Ohrs nicht zu einer entsprechenden Reizweiterleitung führen oder durch die Verarbeitung im Gehirn durch andere (Schall-)ereignisse verdeckt werden.

Das „Psychoakustische Modell“ beschreibt nun, welche akustischen Ereignisse in welcher Form von unserem Wahrnehmungsapparat empfunden werden. Für die medientechnische Datenreduktion bedeutet dies, auf welche „irrelevanten“ Bestandteile eines Audiosignals – obwohl messtechnisch und im Signalverlauf mess- und nachweisbar vorhanden – wir getrost verzichten können, ohne dass die Wiedergabebetreue und Wiedergabequalität dadurch beeinträchtigt würde.

Im Wesentlichen nutzen wir dabei folgende psychoakustischen Phänomene:

- Frequenzmaskierung
- Zeitliche Maskierung
- Räumliche Maskierung

(Phänomen Maskierung = Überdeckung)

(Bei der Encodierung nach MPEG nutzen wir anschließend an die psychoakustische Irrelevanzreduktion noch alle Möglichkeiten der verlustfreien Redundanzreduktion wie Run Length Encoding oder die Huffman-Entropiecodierung.)

Das Hörfeld

Mit unseren Aufnahmemöglichkeiten (Mikrofone, Tonträger) können wir auch Schallereignisse aufnehmen, bearbeiten und wiedergeben, die außerhalb unseres Hörfeldes liegen:

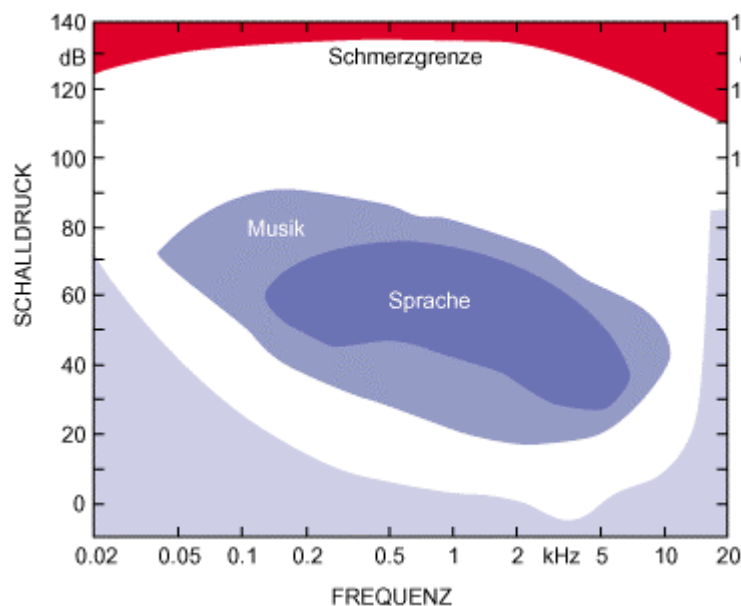


Bild 1:

In diesem Bereich "arbeitet" das menschliche Gehör (Hörfeld). Unterhalb der Ruheshwelle und oberhalb der Schmerzgrenze nehmen wir nichts wahr.

Im Schalldruck-Bereich könnten wir also getrost alle Lautstärken eliminieren, die jenseits der Hörschwelle oder der Schmerzgrenze liegen. Da sich der Schalldruck jedoch auf die dem Ohr angebotene Lautstärke bezieht, können wir hier nur wenig ausrichten, da wir nicht wissen, in welcher Stellung des „Lautstärkereglers“ eine Aufnahme beim Rezipienten wiedergegeben wird.

Entscheidend ist hier der Dynamikumfang, also der Unterschied zwischen der leisesten und der lautesten Stelle in einer akustischen Darbietung.

Im Frequenzbereich können wir Frequenzen außerhalb des menschlichen Hörvermögens getrost eliminieren, allerdings sind wir auch hier abhängig von der insgesamt dargebotenen Lautstärke.

Aus künstlerischen oder wissenschaftlichen Zwecken wollen wir evtl. auch eine Wahrnehmung außerhalb unserer Hörgrenzen ermöglichen und können deshalb das aufgenommene Material – zumindest das Ursprungssignal – nicht verlustbehaftet einschränken.

Frequenzmaskierung

Ein Schallsignal kann ein anderes maskieren, wenn sein Energieniveau höher als das des anderen ist.

Begriffe:

Ruhehörschwelle:

sie ist die untere Begrenzung, die gerade noch hörbare Töne von unhörbaren trennt.

Mithörschwelle:

Anhebung des Nutzsignals über das Störsignal, damit es hörbar wird.

Beispiele:

Gespräch --> Flugzeuglärm

Disco 110 dB, kein Gespräch möglich

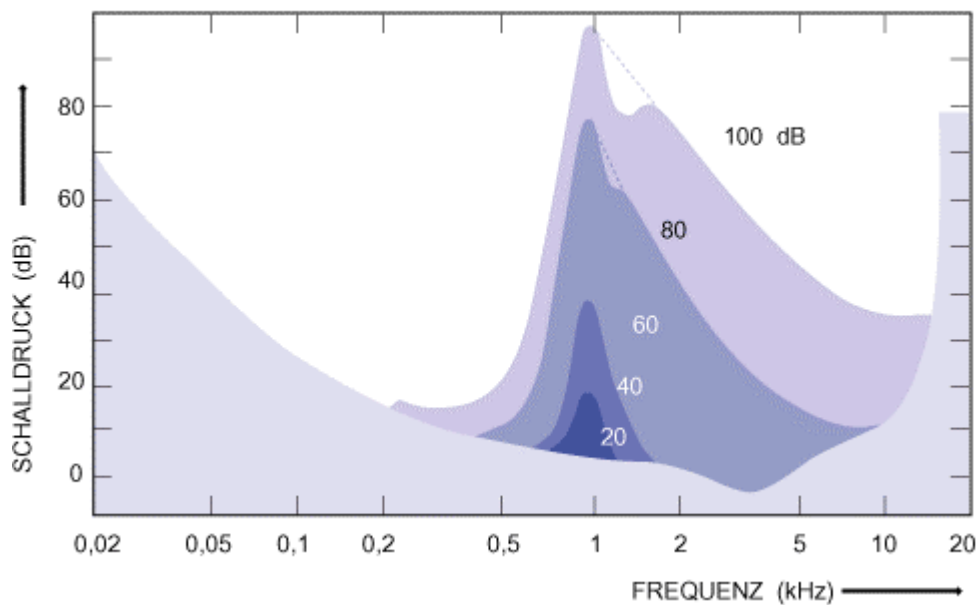


Bild 2:

Die Grafik zeigt die Mithörschwellen bei Tönen, die von Schmalbandrauschen der Mittenfrequenz 1 kHz verdeckt werden, in Abhängigkeit von der Frequenz.

Ausnutzung der Überdeckung. Angenommen, Sie haben einen Ton mit 1 kHz. Ein weiterer Ton liegt bei 1,1 kHz und ist um zirka 18 dB leiser. Der Ton bei 1,1 kHz ist vollkommen vom ersten Ton verdeckt, da er unterhalb der Mithörschwelle liegt. Folglich wären auch andere schwächere Töne in unmittelbarer Nähe des ersten Tones maskiert. Kommt ein dritter Ton mit 2 kHz hinzu, und ist dieser ebenfalls 18 dB leiser als der Erste, wäre er jedoch hörbar, da er über der Mithörschwelle liegt.

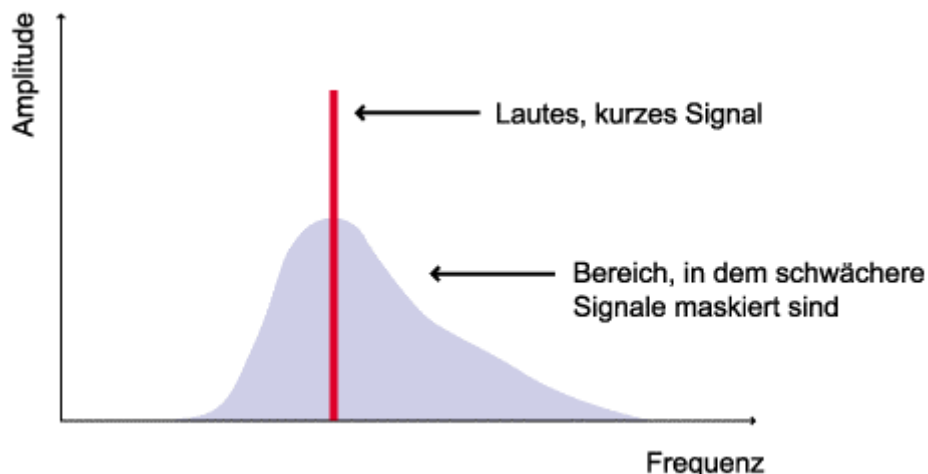


Bild 3:

Starke Signale mit einer geringen Bandbreite sind stärker verdeckend als breitbandige schwache Signale

Psychoakustischer Effekt:

Ein kurzes, lautes Signal (tonal) ist stärker verdeckend als geräuschhafte, breitbandige Signale (non-tonal). Bild 3 zeigt das Spektrum eines lauten, kurzen Sinussignals und die sich daraus ergebende Verdeckungskurve.

Zeitliche Maskierung

Ein weiterer psychoakustischer Effekt: forward-, backwardmasking (Vor- und Nachmaskierung)

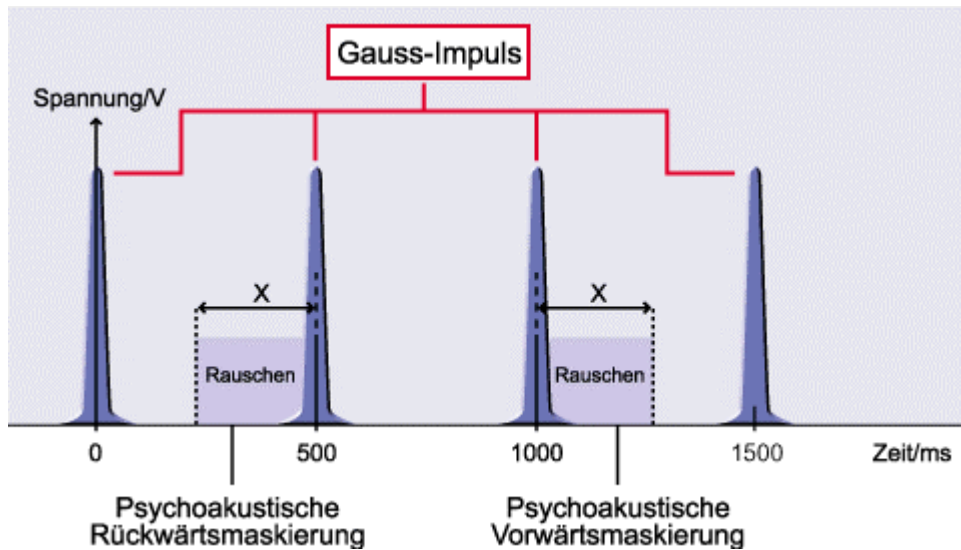


Bild 4:

Das Spektrum eines Sinustons fällt beiderseits gaußförmig ab. Schwächere Signale vor und nach dem Ton werden überdeckt.

"Recovery Time":

Das Gehör benötigt sowohl bei lauten als auch bei leisen Geräuschen eine Erholungszeit, bis es wieder voll funktionstüchtig ist.

Bsp: Gauß-Impulse (Knacken) → Ohr.

Der Effekt: Ein vor dem Impuls liegendes, leises Rauschen ist ebenso unhörbar wie eines danach.

Nachmaskierung - 200 ms

Vormaskierung - 20 ms

Der laute Impuls wird vom Hirn schneller verarbeitet, als das leise Hintergrundrauschen.

Der kurze, laute Impuls "überholt" quasi die schwächeren Signale in der Verarbeitung.

Bisher haben wir einen Komprimierungsgrad 5 bis 6 erreicht.

Räumliche Maskierung

Weitere Ausnutzung der Psychoakustik: Reduzierung der Stereo-Informationen.

Das menschliche Ohr:

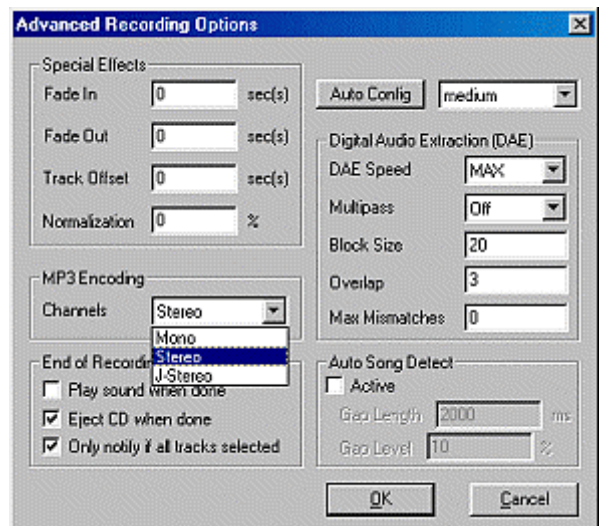
- achtet oberhalb von 2 kHz mehr auf den zeitlichen Wechsel, als auf die hohe Auflösung des Audiosignals.
- kennt keine Richtungsabhängigkeit unterhalb 100 Hz und keine ausgeprägte räumliche Differenzierung unterhalb 500 Hz. (Phänomen Subwoofer, --> kann irgendwo stehen)
- reagiert zwar empfindlicher auf Laufzeitunterschiede als auf Lautstärkeunterschiede bei der Richtungsordnung, begnügt sich aber bei fehlenden Laufzeitunterschieden mit der Lautstärkedifferenz zwischen den beiden Kanälen (Ohren).

Stereosignal:

- intensity (LR) stereo coding → nur 2 Seitensignale (LR: Links, Rechts)
- Middle/Side (MS) stereo coding → Mittensignal (Rechts+Links) und Seitensignal (Rechts-Links)
- Joint Stereo → kodiert nur einige hochfrequente Subbands zu einem Summenkanal (anstatt alle 32 Subbands) → für Datenraten < 64kbps

Bild 5:

Um noch höhere Kompressionsraten zu erzielen, kann man beispielsweise die Stereoinformationen reduzieren.





MP3-Grundlagen: Aufbau und Funktion

Eckdaten (CD):

44,1 kHz Samplingfrequenz, 16 Bit/Sample, Stereo (2 Kanäle)=176 kByte/s oder 1,4 Mbit/s

Irrelevanzreduktion

Kernpunkt bei MP3 ist ein Kompressionsverfahren, das überflüssige Informationen herausfiltert.

Ziel: Daten reduzieren, die das menschliche Gehör nicht, oder kaum wahrnimmt.

Die Grundlage dafür bildet das Psychoakustische Modell des Menschen.

Für MPEG-Audio-Encoder bedeutet dies: Töne, die überdeckt sind, können weggelassen werden, da sie nicht hörbar sind und somit Redundanz bedeuten.

Theoretisch erarbeitet von der „Motion Pictures Expert Group“.

(Siehe Extra-Skript: Was ist MPEG?)

MPEG 1

MP3 spricht: MPEG 1 - Audio Layer 3

Codierung asymmetrisch (Aufwand beim Codieren wesentlich höher als beim Decodieren)

Eckdaten:

Die Qualitätsstufen von MPEG 1 im Überblick (gemäß Fraunhofer Institut)				
Klangqualität	Modus	Bitrate	Kompressionsrate	Bandbreite
Layer 1				
Dig. Compact Cassette	stereo	384 kbit/s		20 kHz
Layer 2				
DAB, Musicam	stereo	192 bis 256 kbit/s		20 kHz
Layer 3 (MP3)				
Telefon	mono	8 kbit/s (0,3-3,4 KHz Bandbreite)	96:1	2,5 kHz
besser als KW-Radio	mono	16 kbit/s	48:1	4,5 kHz
besser als MW-Radio	mono	32 kbit/s	24:1	7,5 kHz
ähnlich UKW-Radio	stereo	56 bis 64 kbit/s	26 bis 24:1	11 kHz
annähernd CD	stereo	96 kbit/s	16:1	15 kHz
CD-Qualität	stereo	112 bis 128 kbit/s	14 bis 12:1	> 15 kHz

MPEG 2

MPEG 3

MPEG 4

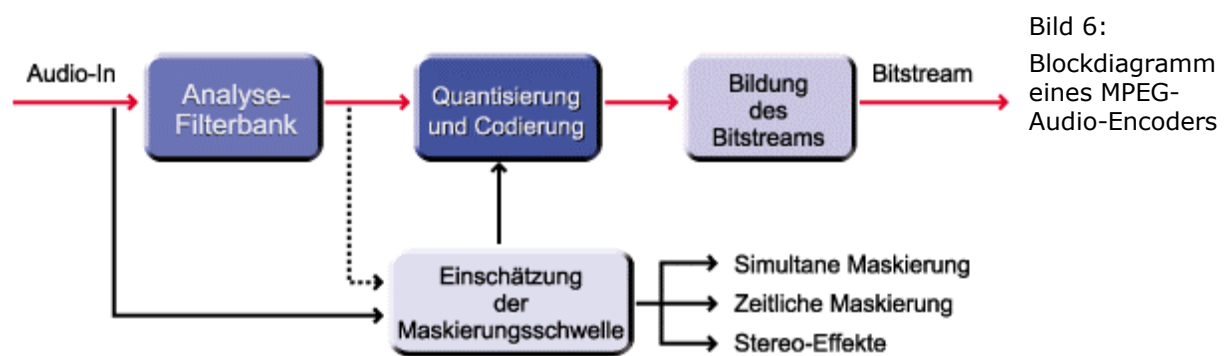
(DivX, Xvid)

MP3 – Prinzip

Grundsätzliche Schritte:

1. Teilung in 32 Subbänder (fest Layer 1, 2; variabel Layer 3)
2. Quantisierung entspr. der Maskierungsmethoden (frequenz-, zeit-abhängig)
3. Reduzierung überflüssiger Stereoinformation
4. RLC
5. Huffmann

Bild 6 zeigt den prinzipiellen Aufbau eines MP3-Encoders. Ein Audiosignal passiert eine Filterbank, die das Signal in einzelne Bereiche (Subbands) aufteilt. Zeitgleich durchläuft das Audiosignal das Psychoakustische Modell. Hier wird die Maskierungsschwelle mit Hilfe der Diskreten Fourier-Transformation für jede Komponente ermittelt. (Die Psychoakustik ist lediglich ein Teil der MP3-Codierung. Das Audiosignal durchläuft weit mehr Stationen.) Das Psychoakustische Modell gibt die maximal erlaubten Quantisierungsfehler an, mit der noch kodiert werden darf, ohne dass diese das menschliche Gehör wahrnimmt --> Anzahl der Quantisierungsstufen (Bitanzahl). Letzter Schritt --> Formatierung des Bitstreams.



Zerlegung des Audiosignals in 32 Subbänder

32 Frequenzbänder (Subbands) mit einer Bandbreite von 625 Hz.

Beim Layer 3 sind die Subbänder an die Eigenschaften des menschlichen Gehörs angepasst --> unterschiedliche Bandbreite.

Der MP3-Encoder untersucht jedes Subband einzeln gemäß dem Psychoakustischen Modell auf verzichtbare Frequenzen. Hierbei erfolgt eine Ermittlung der Maskierungsschwelle.

Psychoakustischer Effekt:

Variable Quantisierung: Im Überdeckungsbereich kann mit geringerer Auflösung digitalisiert (quantisiert) werden.

Beim Digitalisieren entsteht immer ein Quantisierungsrauschen. Dabei werden analoge Signale in wert- und zeitdiskrete (bestimmte) Signale zerlegt.

Die Sampling-Raten:

8 Bit	256 Stufen
12 Bit	4096 Stufen
16 Bit	65536 Stufen

Bei MPEG-Audio wird der hörbare Frequenzbereich in 32 Subbands unterteilt. Bsp.:

1. Im achten Subband befindet sich zum Beispiel ein Ton mit 1 kHz und 60 dB.
2. Der MPEG-Audio-Encoder berechnet nun den Maskierungseffekt und erkennt, dass die Maskierungsschwelle 36 dB weiter unterhalb liegt.
3. Der Encoder berechnet nun den akzeptablen Signalrauschabstand. (60 dB – 36 dB, also 24 dB. 24 dB entsprechen einer Auflösung von 4 Bit.

Beachte: 1 Bit weniger --> 6 dB mehr Rauschen. CD-Audio --> 16 Bit, 16*6dB --> 96 dB Dynamik

Im ersten Schritt fallen die Subbänder weg, deren Pegel unterhalb dieser Verdeckungsfunktion liegt. Jedes Subband erfährt nun eine Abtastung, deren Resultat 16-Bit-große Samples sind.

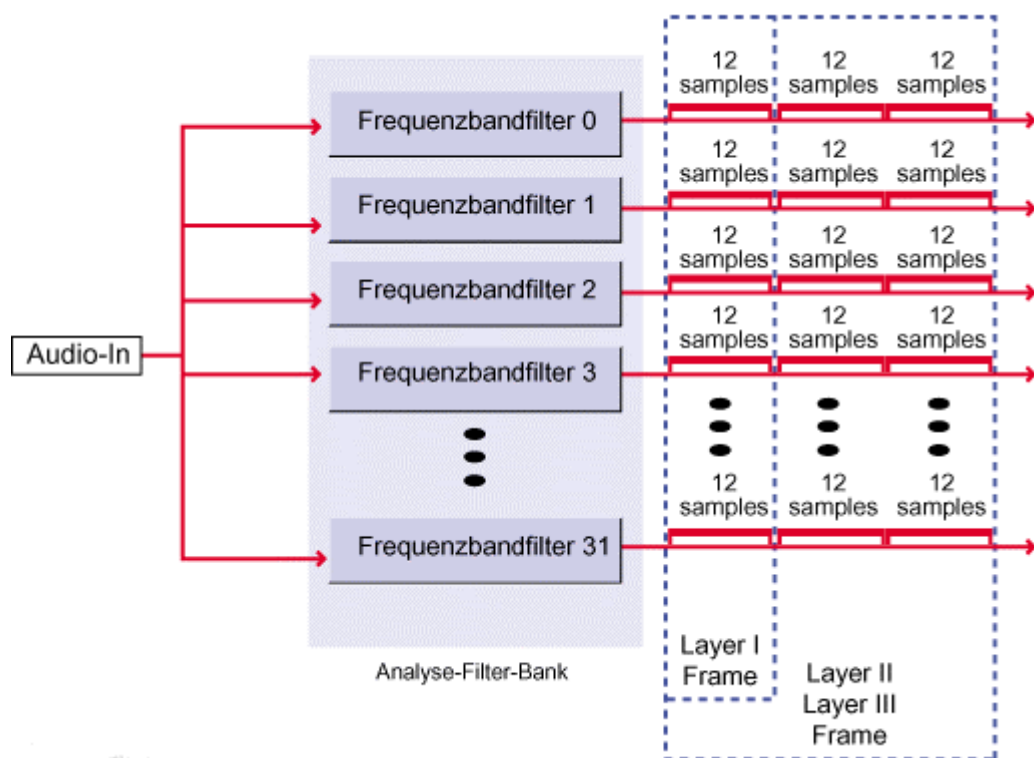


Bild 7:
Über eine Filterbank wird das Audiosignal in 32 Frequenzbänder unterteilt.

Jedes Sample besteht zwar aus 16 Bit, jedoch sind nicht unbedingt alle 16 nötig, um den Pegel darzustellen.

So können beispielsweise die führenden Nullen eines 16-Bit-Samples entfallen. Ergibt sich bei einem

Sample 0000011101010101, --> 11101010101 RLC

+2 Angaben:

Skalierungsfaktor: gibt an, an welcher Stelle die verbliebenen Bits ursprünglich gestanden haben

Bitallocation: Angabe, wie viele Bits verblieben sind

--> Würde man nun diese Informationen für jedes Sample einzeln ablegen, wäre nicht viel gewonnen. Daher teilen sich je zwölf Samples diese Werte.

Kompression Layer 3 (variable Subbänder)

Bereits Layer 1 und 2 unterteilten in Subbänder, Layer 3 geht jedoch einen Schritt weiter.

Veranschaulicht an einem Sinuston von 1500 Hz lässt sich erkennen, dass hierbei zwei Subbänder Werte aufweisen. Hier zeigt sich ein Nachteil der Filterbank, die 32 Subbänder überlappen sich, wodurch ein Ton einer festen Frequenz zwei Bänder beeinflussen kann.

Um diese Folgen der Filterbank zu beseitigen und die Spektralauflösung zu erhöhen, erfolgt beim Layer 3 eine modifizierte diskrete Cosinus-Transformation, die sich um die Zeit-/Frequenztransformation kümmert.

Geht es um die Bearbeitung von Sound, lassen sich grundsätzlich zwei Perspektiven unterscheiden. Entweder das Audiosignal wird als eine Serie von Samples betrachtet, die einem analogen Signal entsprechen. Oder 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.

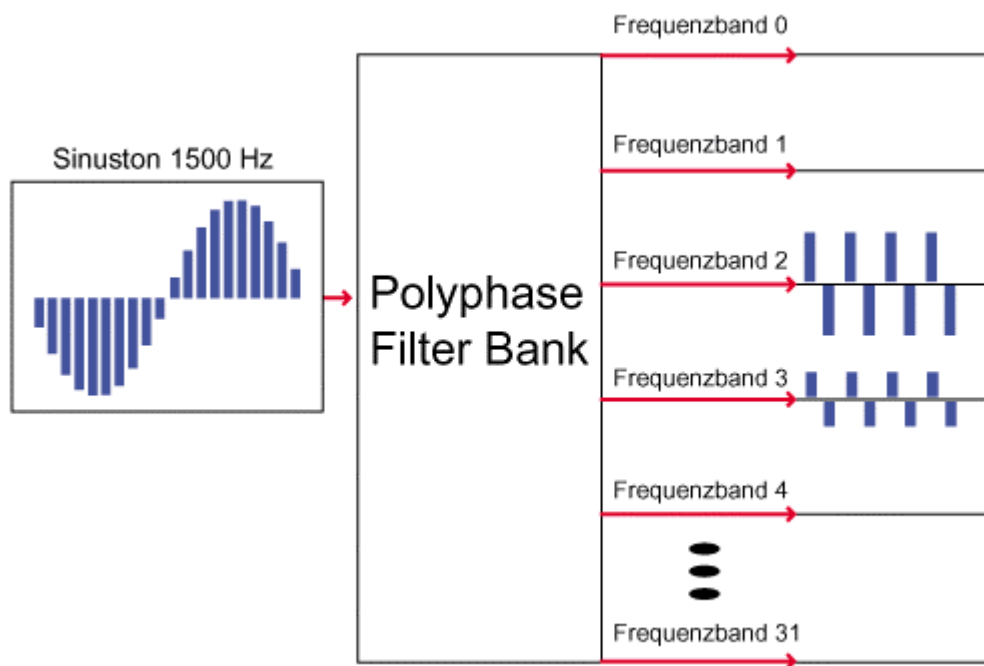


Bild 8:
Ein Sinuston von 1500 Hz schlägt sich nach Durchlaufen der Filterbank in zwei Frequenzbändern nieder.

Um die Daten von einer in die andere Perspektive zu konvertieren, bedient man sich der Fourier-Transformation. Um ein Optimum zwischen Zeit- und Frequenzunterscheidung zu erreichen, bildet Layer 3 zwei verschiedene Blocklängen: eine lange mit 36 Samples und eine kurze mit zwölf Samples. In der Praxis heißt dies: Bei den niedrigen Frequenzen kommen lange Blöcke zum Einsatz. Die langen Blöcke würden jedoch bei höheren Frequenzen keine ausreichende Auflösung erlauben, hier finden die kurzen Blöcke Verwendung. Im so genannten Mixed Block Mode kommen für die beiden Frequenzbänder mit den niedrigsten Frequenzen lange Blöcke zum Einsatz. Für die 30 verbliebenen Frequenzbänder sind die kurzen Blöcke an der Reihe.
Fazit: variable Blocklängen

Vor der Huffman-Kodierung steht die Anordnung der Subbänder. Die Frequenz gibt die Reihenfolge der Bänder vor. Die Subbänder mit niedrigeren Frequenzen enthalten üblicherweise deutlich mehr Werte, als diejenigen der hohen Frequenzen, wo in der Regel 0 oder Werte in der Nähe von 0 erscheinen. Der Encoder unterteilt die Subbänder in drei Bereiche. Jeder Bereich erhält einen eigenen Huffman-Baum, um den optimalen Kompressionsfaktor zu erreichen.

Zunächst klammert der Encoder die Frequenzbänder mit den hohen Frequenzen aus, eine Kodierung ist hier nicht notwendig, da sich ihre Größe aus denen der anderen zwei Regionen ableiten lässt. Die zweite Region enthält Frequenzbänder, in denen die Werte von -1 bis 1 häufig auftauchen. Der dritte Bereich enthält alle verbleibenden Werte und wird ein weiteres Mal in drei Regionen unterteilt, von denen jede einen eigenen Huffman-Baum zugewiesen bekommt. Welcher Huffman-Baum zum Einsatz kommt, wird innerhalb der MP3-Datei abgelegt.

Aufbau Layer-3-Bitstream

Nach der Aufspaltung des Audiosignals und der bisherigen Codierung, wird das Ergebnis in so genannten Frames zusammengefasst. Jedes Frame enthält eine feste Anzahl von Abtastwerten (Samples). Bei Level 3 sind dies 1152 Abtastwerte pro Frame (32 Subbänder x 36 Samples). Ein Frame besteht aus einem Header, einem Prüfsummencheck, den eigentlichen Audiodaten sowie unter Umständen einem so genannten Bit-Reservoir. Ein solches Reservoir entsteht, wenn sich die Samples innerhalb des Frames so komprimieren lassen, dass nicht die komplette theoretische Bit-Anzahl eines Frames benötigt wird.

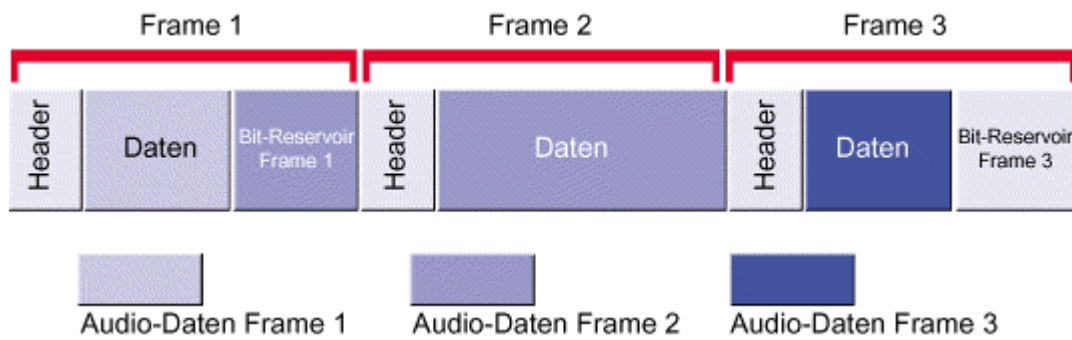


Bild 9: Der Datenstrom des Layer 3 besteht aus so genannten Frames, von denen jedes einen eigenen Header besitzt.

Auf diese Reservoirs kann der Encoder zurückgreifen, wenn bei einem späteren Frame die vorhandenen Bits nicht ausreichen. Hierbei muss man zwei Begriffe unterscheiden: Framegröße und Framelänge.

Die Framegröße wird durch die Anzahl der Samples bestimmt und ist innerhalb eines Layers konstant. Im Format Layer 1 sind dies stets 384 Samples pro Frame, bei Layer 2 und 3 kommen 1152 auf ein Frame. Die Länge des Frames kann sich jedoch bei Layer 3 durch das Wechseln der Bitrate oder das nicht aufgefüllte Bit-Reservoir unterscheiden. Vor Layer 3 existierte die Möglichkeit eines Bit-Reservoirs nicht. Theoretisch ist auch eine zerstückelte MPEG-Sounddatei abspielbar. Dies gilt jedoch nicht für nach Layer 3 kodierte Dateien, da hier die Frames auf Grund des Bit-Reservoirs verschachtelt sind. Ebenfalls im Frame enthalten sind die vorher erwähnten Informationen bezüglich des Skalierungsfaktors und der Bit Allocation, um alle Samples wieder rekonstruieren zu können.

Frameheader

Jeder Frame enthält einen Header sowie Audiodaten. Ein Datei-Header, wie er von anderen Dateiformaten her bekannt, existiert bei einer MP3-Datei nicht. Der Frame Header belegt die ersten 32 Bit eines Frames. Die ersten 12 Bit des Headers bilden das Frame Sync.

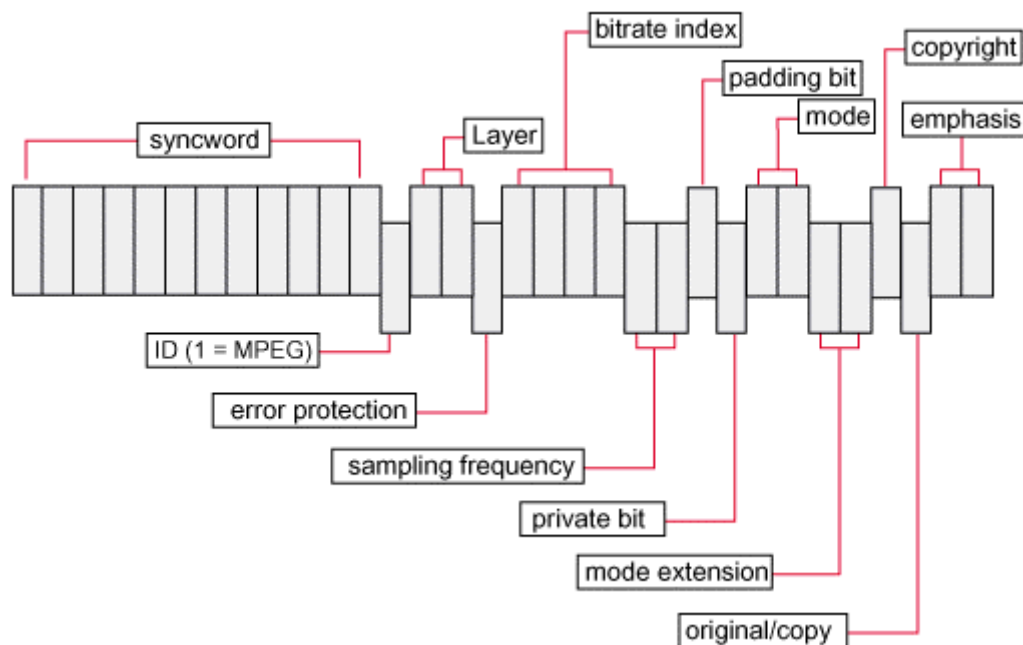


Bild 10: Im Frameheader (32 Bit) sind alle wichtigen Informationen zu der Qualität der MP3-Datei hinterlegt.

Syncword (12 Bit) ID (1 Bit):

Offiziell ist ID-Signatur nur ein Bit groß, um die verwendete MPEG-Version zu kennzeichnen. In der Urfassung steht der Wert 1 für MPEG 1 und der Wert 0 für MPEG 2. Mit dem inoffiziellen Standard MPEG 2.5 reicht ein Bit als Kennung jedoch nicht mehr aus. Anwendungen, die diesen Standard bereits unterstützen, verwenden das zwölfte Bit des Syncwords als zusätzliche MPEG-ID. Ist dieses auf 0 gesetzt, steht dies für MPEG 2.5. Standardmäßig sind alle Bits des Syncwords gesetzt.

Layer (2 Bit):

Gibt an, nach welchem Layer die Codierung erfolgt ist.

Wert	Bedeutung
00	reserviert
01	Layer 3
10	Layer 2
11	Layer 1

error protection (1 Bit):

Wenn das Error Protection Bit gesetzt ist, folgt nach dem Frameheader eine 16-Bit-CRC-Prüfsumme.

bitrate_index (4 Bit):

Hier steht, in welcher Bitrate die Datei codiert wurde.

Wert	MPEG1 Layer I	MPEG1 Layer II	MPEG1 Layer III
0000	frei	frei	frei
0001	32 KBit/s	32 KBit/s	32 KBit/s
0010	64 KBit/s	48 KBit/s	40 KBit/s
0011	96 KBit/s	56 KBit/s	48 KBit/s
0100	128 KBit/s	64 KBit/s	56 KBit/s
0101	160 KBit/s	80 KBit/s	64 KBit/s
0110	192 KBit/s	96 KBit/s	80 KBit/s
0111	224 KBit/s	112 KBit/s	96 KBit/s
1000	256 KBit/s	128 KBit/s	112 KBit/s
1001	288 KBit/s	160 KBit/s	128 KBit/s
1010	320 KBit/s	192 KBit/s	160 KBit/s
1011	352 KBit/s	224 KBit/s	192 KBit/s
1100	384 KBit/s	256 KBit/s	224 KBit/s
1101	416 KBit/s	320 KBit/s	256 KBit/s
1110	448 KBit/s	384 KBit/s	320 KBit/s
1111	reserviert	reserviert	reserviert

sampling_frequency (2 Bit):

Hier wird die Sampling Rate in Hz festgelegt. Richtet sich unter anderem nach der in der MPEG-ID festgelegten MPEG-Version.

Wert	MPEG1	MPEG2	MPEG2.5
00	44100	22050	11025
01	48000	24000	12000
10	32000	16000	8000
11	reserviert	reserviert	reserviert

padding_bit:

Dieses Bit wird dazu benutzt, um die Bitraten exakt zu bestimmen. Ist es auf 0 gesetzt, ist das Frame nicht aufgefüllt.

private_bit (1 Bit):

Steht zur freien Verfügung. Beispielsweise um in Applikationen ein bestimmtes Ereignis auszulösen.

mode (2 Bit):

Hier erfolgt die Angabe über den Stereomodus. Es wird unterschieden, ob die Datei Stereo, Zweikanal oder Mono codiert ist. Folgende Bitfolgen stehen für die entsprechenden Werte:

Wert	Bedeutung
00	Stereo
01	Joint Stereo
10	Zweikanal
11	Mono

mode extension (2 Bit):

Ist nur von Bedeutung, wenn bei der Kodierung der Joint-Stereo-Modus gewählt wurde. Während bei Layer 1 und 2 hier die Zuordnung der Frequenzbänder an den Joint-Stereo-Modus erfolgt, gilt für Layer-3-Dateien die Festlegung welcher Joint-Stereo-Modus zum Einsatz kommt (intensity stereo oder M/S-Stereo). Beim Layer 3 bestimmt der Dekompressionsalgorithmus die Frequenzbänder.

Wert	intensity stereo	M/S-stereo
00	aus	aus
01	an	aus
10	aus	an
11	an	an

copyright (1 Bit):

Ist das Copyright-Bit gesetzt, ist die MP3-Datei urheberrechtlich geschützt.

original/copy (1 Bit):

Gibt an, ob es sich um ein Original oder eine Kopie handelt.

emphasis:

Hier wird die verwendete Rauschunterdrückung angegeben. Der Einsatz ist allerdings nicht sehr gebräuchlich.

Wert	Rauschunterdrückung
00	keine
01	50/15 ms
10	reserviert
11	CCITT j.17

Der ID3-Tag

Beim ID3-Tag handelt es sich um eine Art Inhaltsangabe der MP3-Datei, ähnlich wie dies CD-Text für Audio-CDs darstellt. So beziehen die MP3-Player (Soft- wie Hardware), die Informationen über das abzuspielende Stück aus dem ID3-Tag. Der ID3-Tag ist 128 Byte groß und steht am Ende einer MP3-Datei.

Länge (Byte)	Beschreibung
3	TAG-Identifizierung, hier muss TAG stehen
30	Titel des Stückes
30	Interpret

30	CD-Titel
4	Erscheinungsjahr
30	Kommentar
1	Genre

Das Genre ist ein numerisches Feld. Es existiert eine Liste, in der bestimmte Nummern den Genres von Acid Punk bis Southern Rock zugeordnet sind. Mittlerweile wird außerhalb der MP3-Spezifikation der ID3-Tag von unabhängigen Interessensgruppen weiterentwickelt. Da der ID3-Tag immer am Ende der MP3-Datei steht, ist es relativ einfach, ihn auszulesen beziehungsweise zu editieren.



Bild 11:

Mittlerweile existieren zahlreiche Programme, die ein nachträgliches Editieren des ID3-Tag erlauben.

In neueren Varianten des ID3-Tags werden Bytes des Kommentarfelds geopfert, um die Original-Track-Nummerierung der CD ebenfalls speichern zu können. Inzwischen kursieren zahlreiche Sharewareprogramme, die sich der Bearbeitung des ID3-Tags annehmen. Beliebte Funktionen sind dabei die automatische Erzeugung des ID3-Tag aus Verzeichnis- und Dateinamen sowie umgekehrt das Benennen von Dateien auf Grund vorliegender ID3-Tag-Informationen.

Audio-Demos

Damit Sie die Effekte jeweils selbst nachvollziehen und Ihr Gehör testen können, stellen wir Ihnen drei Testdateien zur Verfügung. (Mit freundlicher Genehmigung von Dipl.-Ing. Thomas Sporer des Fraunhofer Instituts für integrierte Schaltungen.)

Audio-Demo 1: [Rauschen verdeckt einzelne Töne](#)

Maskierer: Schmalbandrauschen zwischen 900 Hz und 1100 Hz, -20 dBFS

Testsignal: Sinustöne bei 600 Hz, 1000 Hz und 1600 Hz. Der Schalldruck erhöht sich stetig um 10 dB angefangen von -80 dBFS bis -20 dBFS

Audio-Demo 2: [Forward Masking](#)

Maskierer: Gauß-Impuls bei 0 dBFS

Testsignal: Rauschen (durchschnittliches Spektrum wie der Gauß-Impuls), 50 dBFS

Audio-Demo 3: [Backward Masking](#)

Maskierer: Gauß-Impuls bei 0 dBFS

Testsignal: Rauschen (durchschnittliches Spektrum wie der Gauß-Impuls), 50 dBFS