MultiMediaPraktikum

Universität Tübingen Wilhelm-Schickard-Institut für Informatik

Wintersemester 2002/ 2003

Aufgabe 3 - MP3-Audio

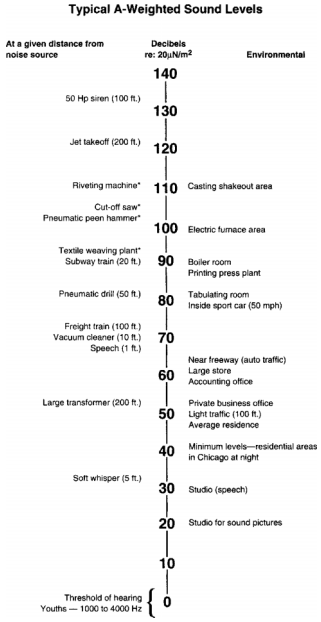
Marc-Oliver Pahl
Ulrike Schaal
David Eißler

Inhaltsverzeichnis

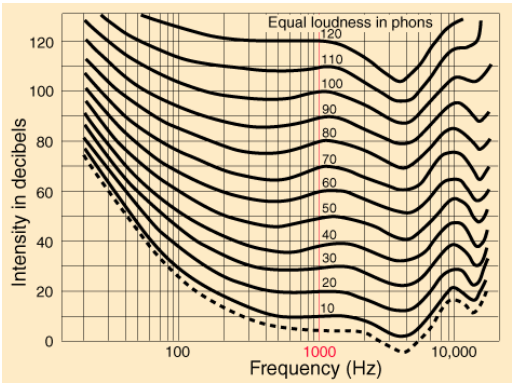
Theorieteil

Phon

60 phons means „as loud as a 60 dB, 1000 Hz tone“



Die Einheit Phon orientiert sich am Lautstärkeempfinden des Menschen. Wir nehmen manche Töne (z.B. 4400Hz) früher (bei geringerem Schalldruck) wahr, als andere (z.B. 40Hz). Das wird bei der Skala berücksichtigt:



Ein tiefer Ton mit wesentlich höherem Schalldruck hat also die gleiche Lautstärke in Phon, wie der -absolut viel leisere- höhere Ton.

Dezibel

Der Schalldruck - gemessen in dB - gibt die absolute, messbare Lautstärke an. Die Schritte in der Einheit sind logarithmisch, d.h.

[dB] = 20 x $_{10}$log (po/ 0,0002 pa) (dezibel = Schalldruckpegel)

po = BezugsSchalldruck (Hörswchwelle bei 100Hz)

[pa] = 1N/ m^2 (pascal = Einheit für Schalldruck)

Hörgrenze bei 0dB Schmerzgrenze bei 140dB.

SPL = SoundPressureLevel

Rauschabstand

Der Signal/ Rauschabstand gibt an, wie groß der Abstand zwischen Nutzsignal und Rauschen ist. Je höher der Wert, desto besser.

Ab einer Schwelle ist kein Signal mehr auszumachen. MP3 verringert den Signal/ Rauschabstand im nicht hörbaren Bereich dynamisch, wenn die Bandbreite sonst überschritten würde.

Das bedeutet, der Encoder prüft iterativ durch Vergleichen von Input und decodiertem Output, inwieweit sich das Quantisierungsrauschen bemerkbar macht und versucht dieses in der geforderten Bandbreite möglichst gering zu halten.

andere Codecs

Kompressionsname	Einsatzgebiet	CD-nah bei...
AAC	digitales Radio, im QuickTime 6	96 kbps
Microsoft WMA	WMP	128 kbps
Ogg Vorbis	soll mp3 ablösen	128 kbps
ATRAC	MD-Player von Sony	292 kbps
TrueSpeech	digitales Telefonieren, DSP-Chips	- (max. 8,5 kbps)

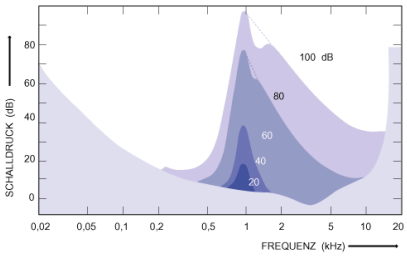
perceptual audio coding

„Perceptual audio coding“ bedeutet, dass sich der Encoder den Daten anpasst, d.h. nicht alles gleich codiert sondern je nach aktuellem Signal und gewünschter Bandbreite anders vorgeht. Die Kodierung ist dabei dem menschlichen Gehör angepasst.

Dazu wird das Signal gefiltert und in Teildatenströme zerlegt (mp3: 32*18=576 sonst üblich: 32), die dann mithilfe einer Maskierungsdatenbank (gleichzeitige Maskierung, zeitliche Maskierung, Stereo-Bearbeitung) den zur Verfügung stehenden Bits zugewiesen werden.

Maskierung

Maskierung beschreibt nichts anderes als Überdeckung zweier Signale und das Unvermögen des menschlichen Gehöres, diese zu differenzieren - ein lautes Signal maskiert „darunterliegende“ leise ->



Die Maksierung eines 1 kHz Tons mit der angegebenen Lautstärke

Es gibt folgende Maskierungsarten:

(a) simultane (auditory) Maskierung

Leise Töne mit ähnlichen Frequenzen wie ein lautes Signal werden ab einer bestimmten Grenze verdeckt und daher weggelassen (rechts).

Der Rauschabstand im für das Gehör nicht wahrnehmbaren Bereich (maskierfähiger Bereich) wird verringert.

(b) zeitliche Maskierung (erst vollständig bei MPEG-2 AAC)

Leise Signale, die nahe bei (vor oder nach) lauten Signalen liegen, können nicht wahrgenommen werden. Bei der zeitlichen Maskierung werden diese daher ausmaskiert.

(c) joined stereo coding

midle/ side stereo coding

Der Stereoeindruck entsteht durch Phasen- und Pegeldifferenz zwischen linkem und rechtem Kanal(Lautstärke und Zeitpunkt des Eintreffens am linken und rechten Ohr). Das Stereosignal wird zerlegt in ein Mitten- (L+R) und ein Seitensignal (L-R). Dabei enthält das Seitensignal sehr viel weniger Informtion als ein vollwertiger Stereokanal. Zusätzlich wird die räumliche Information reduziert . Diese Datenreduktion ist verlustlos reversibel.

(d) Intesity Stereo Coding

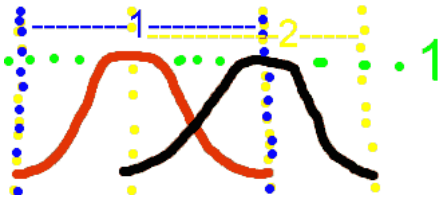
intensity stereo coding

Ab einer bestimmten Frequenz werden nur noch ein Summenkanal und Richtungsinformation übertragen. Allerding bewirkt diese Datenreduzierung schon deutliche Unterschiede zum Original.

pre echo

Bei der zeitlichen Maskierung mithilfe von Frames tritt das Phänomen des pre echos auf. Bei der MDCT überlappen die aufeinanderfolgenden Transformationsfenster um 50%. Zusätzlich wird das Signal mit einer Gauß-glockenkurvenähnlichen Funktion (Hanning-Kurve) gefiltert.

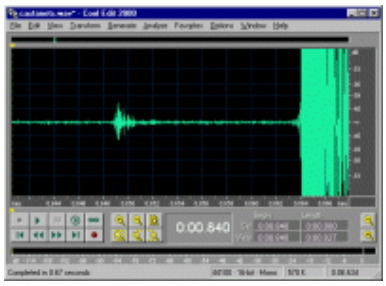
Das Originalsignal erhält man dann wieder durch Zusammensetzen der sich überlappenden Kurven:



An der Stelle, an der sich die beiden Frames überlagern kann das Original nur mithilfe beider Frames gewonnen werden. Findet nun in einem - aufgrund eines starken (lauten) Signals- eine starke Quantisierung statt, so wirkt sich diese als pre echo schon vor dem Signal aus (hörbar zu schlechte SNR).

Um dies zu verhindern, ist die Blockgröße die in der MDCT verwendet wird variabel und wird an das Signal angepasst (bei den Kastagnetten z.B. funktioniert das aber nicht).

Kommt also in einem späteren Frame ein lautes kurzes Signal auf, so wird es aufgrund des psychoakustischen Modells auf das gesamte Frame verteilt (Maskierung; vorher und nachher hören wir sonst nichts; Anhebung des Rauschens). Da die Frames sich aber überlappen, wird ein eigentlich noch nicht vorhandenes Signal im vorhergehenden Frame berücksichtigt und das Signal hat ein vorher liegendes Echo.


Referenz


„normales“ pre echo


„extremes“ pre echo

recovery time

Als recovery time wird die Zeit bezeichnet, die das Gehör sowohl bei lauten als auch leisen Geräuschen benötigt, bis es wieder voll funktionstüchtig ist. Die recovery time erstreckt sich sowohl vor (<5 ms) als auch hinter (<300 ms) das intensive Signal.

psychoakustisches Modell

Die Psychoakustik untersucht, was wir hören, indem sie den objektiven technischen Reiz mit dem subjektiven Hörempfinden des Menschen vergleicht.

Ein psychoakustisches Modell beschreibt die Wahrnehmung des menschlichen Gehörs.

Da „das Hören“ nicht objektiv messbar ist, wird solch ein Modell dadurch entwickelt, dass Maskierung und Artefakte (Verfälschung aufgrund von Reduktion) von Testpersonen angehört und subjektiv beurteilt werden.

Dabei sind genormte und jederzeit nachvollziehbare Rahmenbedingungen Voraussetzung für einen sinnvollen Hörtest:

„triple stimulus, hidden reference“ (3 Hörbeispiele in Folge, A ist immer das Original, B und C wechseln zufällig zwischen Original und datenreduziertem Signal), gleicher Frequenzgang und identische Lautstärke der Audiosignale.

VBR

VBR steht für variable bit rate. Das heißt die Bitrate passt sich dem Signal an. Man stellt eine maximale Bandbreite ein und der Encoder nutzt diese nicht immer voll aus sondern codiert das Signal nur so gut wie nötig. Das heißt, bei maximaler Bitrate von 192 kB/s kann es gut sein, dass das Signal zeitweise nur mit 96 kB/s encodiert wird, weil die Bandbreite des Signals entsprechend gering ist.

Datenreduktion

Das psychoakustische Modell erreicht eine Datenreduktion von 1:5 (? Es fließt nahezu überall ein ?).

Es fließt zusätzlich sowohl bei der Diskretisierung (MDCT) ein (je nach Signal wird in größeren oder kleiner Blöcken diskretisiert [geringeres oder höheres Datenvolumen]),

als auch anschließend bei der Quantisierung der Frequenz (Welche Frequenzen gewichte ich wie?) ein.

Anschließend wird noch die Amplitude nichtlinear quantisiert, was bedeutet, dass kleinere Werte (leise) genauer, größere (laut) ungenauer behandelt werden.

Am Ende wird noch die Huffman-Codierung (mit 32 festen Codetabellen) eingesetzt (20% weniger Daten).

Verluste

Die Qualitätsverluste treten beim MP3 also durch das perceptual audio coding (zeitliche/ räumliche Maskierung und Quantisierung entsprechend der gewünschten Datenrate [Rundungsfehler/ pre-echo/ Aliasing]) und das intensity stereo coding auf.

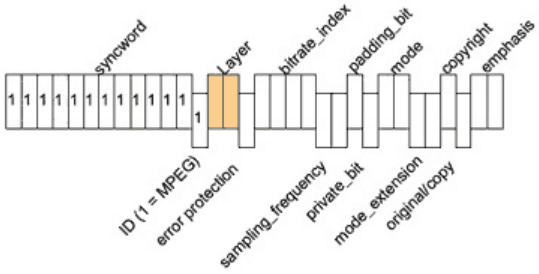
Dabei gehen nicht nur Daten verloren sondern es entstehen sogar falsche neue in Form von Quantisierungsrauschen.

MP3 Frames

http://goethe.ira.uka.de/seminare/rftk/mp3/

Die MP3-Datei beginnt und endet optional mit dem ID3-Tag in dem die Daten zu der Datei (Titel/ Künstler etc.) stehen. Dazwischen befinden sich die mp3-frames, die die eigentlichen Daten enthalten.

Jedes Frame beginnt mit folgendem Header:

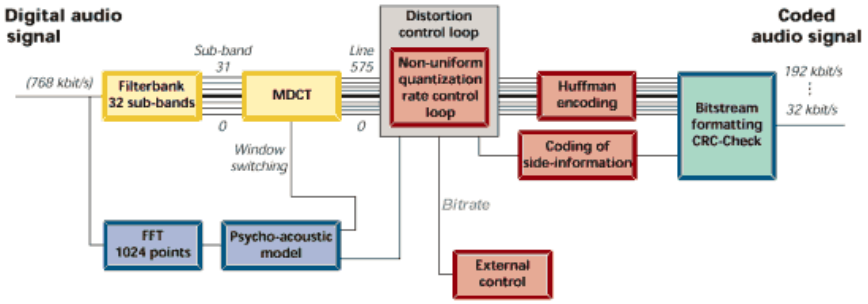


Position	Aufgabe	Länge [Bit]
A	**Frame-Sync**	12
B	**MPEG Audioversion** (MPEG-1, 2, etc.)	1
C	**MPEG Layer** (Layer I, II, III, etc.)	2
D	**Protection** (wenn aktiv: Checksumme nach Header)	1
E	**Bitrate-Index**	4
F	**Frequenz der Samplingrate** (44.1kHz, etc.,)	2
G	**Padding Bit** (kompensiert unvollständige Belegung)	1
H	**Private Bit** (Applikations-spezifische Trigger)	1
I	**Channelmode** (Stereo, Joint-Stereo)	2
J	**Mode-Extension** bei Verwendung von Joint Stereo)	2
K	**Copyright**	1
L	**Original** ("0", wenn Kopie, "1" wenn Orignial	1
M	**Emphasis** (veraltet)	2

Dann folgen 1152 Samples (~26ms):

18 (MDCT) * 2 (MDCT 50% overlap) * 32 (Subbänder) Frequenzwerte

Gespeichert werden die benutzten Huffman-Tabellen, Skalierungsfaktoren, Quantisierungsschrittgröße und die eingesetzten MDCT-Transformationsblöcke. Im Anschluss an die Seiteninformationen kommen die Hauptdaten, d.h. die kodierten Spektrallinien und danach optional noch ein benutzerdefinierter Datenblock, in dem z.B. bei Digital Audio Broadcasting (eingesetzt bei live Radiostreams) der aktuelle Musiktitel übermittelt werden kann.

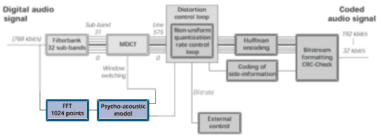


Encodierung

FFT

Für die Analyse mithilfe des psychoakustischen Modells wird das Signal mit einer 1024bit- und vier 256bit-FastFourierTransformation in den Ortsraum transformiert.(Dadurch werden zwar nicht alle 1152 später entstehenden Bänder analysiert, die Transformation ist dadurch, dass sie auf Zweierpotenzen operiert aber um den Faktor 100 schneller.)

psychoakustik -> MDCT



Die gewonnenen Erkenntnisse bezüglich der Signalart fließen zum Einen in die Modified Discreet Cosinus Transformation (MDCT) ein. Diese heißt unter Anderem modified, weil sich die Blöcke um 50% überlappen, was zum pre echo (<-) führen kann. Um dies zu verhindern sind zwei unterschiedliche Längen von MDCT-Blöcken spezifiziert:

- lange Blöcke (long block) bestehend aus 36 Samples (höhere Auflösung im Frequenzbereich)

- 3 kurze Blöcke (short block) mit 12 Samples (höhere zeitliche Auflösung)

Die Modi zur Verwendung dieser Blöcke sind:

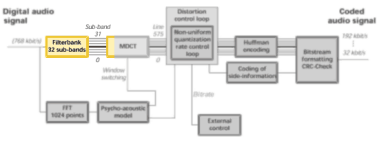
- short-block oder long-block mode: nur kurze oder lange Blöcke

- mixed-block mode: lange Blöcke für die zwei untersten Teilbänder (wo eine höhere Auflösung im Frequenzbereich benötigt wird), kurze Blöcke für die restlichen 30 Bänder

psychoakustik -> Maskierung

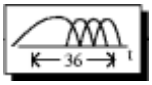
Zum Anderen wird ermittelt, welche Frequenzen wann und wo maskiert sind und folglich höher quantisiert werden können.

SubBandZerlegung

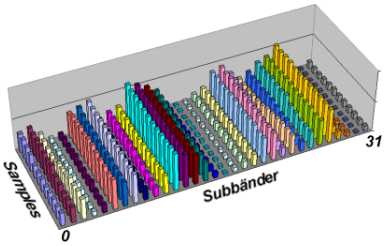


Bevor das Signal in die MDCT geht wird es in 32 teilweise überlappende Subbänder zerlegt. Dabei werden einfach 32 Bandpassfilter zum Zerteilen eingesetzt.

Ein Frame besteht aus 1152 Samples, das heißt, für ein Frame entstehen 36*32 Subbänder (36 Zeiten mal die 32 Subbänder).



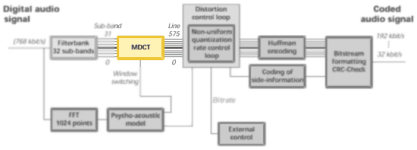
Die Zerteilung auf der Zeitachse erfolgt durch Multiplikation mit einer Henning-Kurve (ähnlich Gaußsche Glockenkurve). Die Samples überlagern sich dabei jeweils zur Hälfte, um „Anschlussprobleme“ an den Blockgrenzen zu vermeiden.



Die Zerlegung in die Subbänder unterteilt das gesamte Spektrum und zwar entsprechend unserem Gehör. Das heißt in Bereichen in denen wir besser hören sind die Subbänder schmaler, in anderen breiter.

Somit findet schon hier eine erste Quantisierung statt, denn im Anschluss wird jeweils ein solches Subband mithilfe der MDCT vom Zeit in den Frequenzraum transformiert.

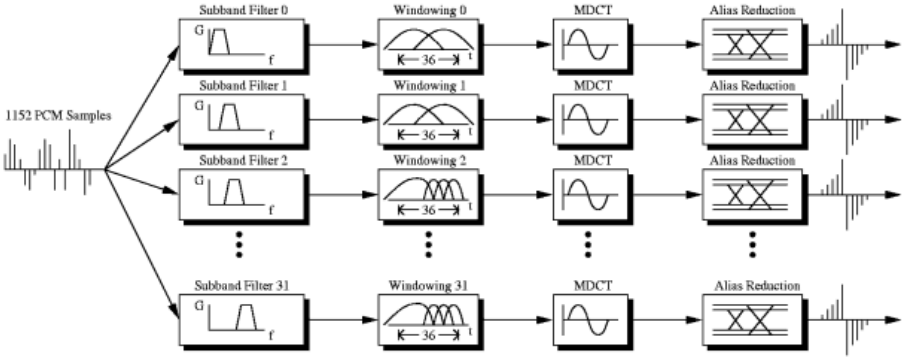
MDCT



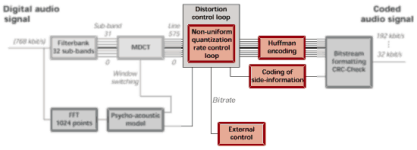
Im Gegensatz zur DCT bei JPG operiert die MDCT nicht auf zweidimensionalen sondern auf eindimensionalen Daten. Sie zerlegt jedes der 32 Subbänder jeweils nochmal in 18 Frequenzen, was der Auflösung im Frequenzbereich natürlich zugute kommt. Gleichzeitig wird aber durch die Eindimensionalität die Zeitinformation innerhalb des Blocks verworfen (-> schlechtere zeitliche Auflösung --> pre echo). Außerdem werden auch nur die Inneren 18 (von 36) Sampewerte verwendet, da sich die Blöcke ja sowieso um 50% überlappen.

Dadurch, dass die 32 Subbänder sich überlagern, kann es natürlich sein, dass sich ein und dieselbe Frequenz jetzt nach der MDCT in mehreren Subbändern wiederfindet. Da die Subbänder aber getrennt voneinander quantisiert werden, würde dies zu ungewünschtem Aliasing führen.

Deshalb müssen diese Frequenzen in einem Anti-Aliasing-Schritt nach der MDCT noch herausgefiltert werden.



Quantisierung



Nun kommt die eigentliche Quantisierung der Daten.

Durch die nun vorliegende Repräsentation im Frequenzraum können gezielt die maskierten Frequenzen (psychoakustisches Modell) entfernt oder zumindest recht hoch quantisiert (durch einen hohen Koeffizienten geteilt) werden.

Es wird nicht linear quantisiert sondern so, dass kleinere Werte schwächer, größere aber stärker quantisiert werden. Das entspricht auch wieder unserem Hörempfinden: Ist eine Fequenz sehr stark vorhanden, so hören wir kleinere Unterschiede (also ähnliche Frequenz vorhanden) sowieso kaum oder gar nicht.

In die Quantisierung gehen nicht die 32 Subbänder oder die 576 Frequenzlinien der MDCT, sondern diese werden in 24 Gruppen unterteilt, die den Haupthörbändern unseres Ohres entsprechen.

Die Quantisierung findet in zwei Schleifen statt:

Die innere Schleife „ratio control loop“ codiert das Signal mit der aktuellen Quantisierungsstufe, macht eventuell JointStereo oder IntensityStereo, und packt das Ganze anschließend mit einem der 32 Huffmann-Bäume. Danach wird überprüft, ob die Bitrate überschritten wurde. Wenn ja, muss erneut mit höherer Quantisierung gepackt werden.

Die äußere Schleife kontrolliert die Qualität des Signals [„distortion control loop“ (Verzerrungskontrollschleife) oder „noise control loop“ (Geräuschkontrollschleife)]. Dazu decodiert sie das Signal und vergleicht es mit dem Original. An den Stellen, an denen „hörbares“ Quantisierungs-

rauschen auftritt setzt sie die Faktoren für die Quantisierug herunter, wenn dies die eingestellte Bitrate zulässt.

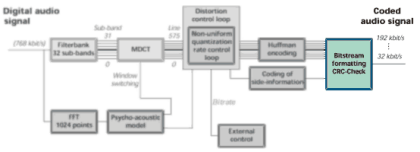
Die innere Schleife konvergiert immer, da die Koeffizienten einfach so lange erhöht werden, bis wir innerhalb der Bitrate sind.

Die äußere Schleife wird dagegen ihr Ziel nicht immer erreichen, weil die Datenreduktion einfach nicht beliebig ohne hörbare Verschlechterung des Signals erfolgen kann.

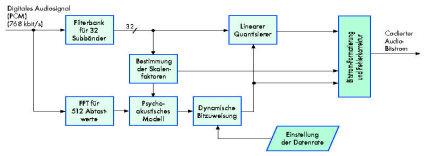
Verluste treten vor allem bei der Dynamik innerhalb der betrachteten Bänder des Signls auf. „Die Signale ebnen sich ein“.

Bitstream erstellen

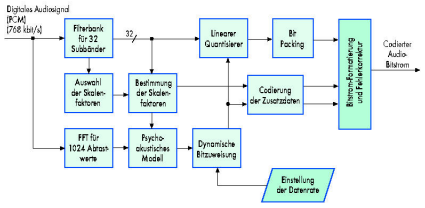
Zum Abschluss wird das Frame noch entsprechend dem vorher beschriebenen mit Header und eventuell einer Prüfsumme versehen.



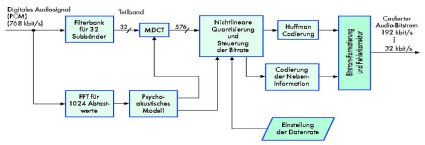
Layer 3 : Layer 1, Layer 2



Layer 1



Layer 2



Layer 3

Layer 1, 2 und 3 sind die Audio-Formate des MPEG-Standards. Mit steigender Nummer steigt auch der Aufwand, den die Encodierung benötigt.

Die drei größten Neuerungen an Layer 3 im Vergleich zu Layer 1 und 2 sind die MDCT, die Huffman-Codierung und die nichtlineae Quantisierung.

Die MDCT führt zu einer dutlichen Steigerung der Frequenzauflösung um den Faktor 18.

Die Huffman-Codierung mit den 32 festen Bäumen reduziert das Datenvolumen um 20%.

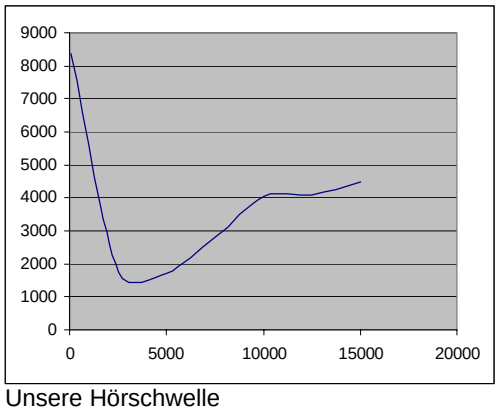
Die nichtlineare Quantisierung verbessert das Signal indem in Bereichen in denen wir besser hören weniger stark und in den anderen dafür stärker quantisiert wird.

Alle drei Verfahren beruhen auf einem psychoakustischen Modell zur Datenreduktion. In Layer 1 wird eine nur 512bit-wertige FFT zur Ermittlung der Maskierungen verwendet. Bei Layer 2 wird schon auf 1024bit analysiert; außerdem fließen bei Layer 2 auch noch die Informationen der 32 Subbänder in das psychoakustische Modell mit ein, was sie bei Layer 3 indirekt durch die MDCT tun.

Die 32 Subbänder, in die das Signal zu allererst aufgeteilt wird, sind bei Layer 1 und 2 jeweils 625Hz breit. Bei Layer 3 sind die Bänder unterschiedlich breit, dem unterschiedlichen Hörvermögen innerhalb verschiedener Frequenzbereiche angepasst.

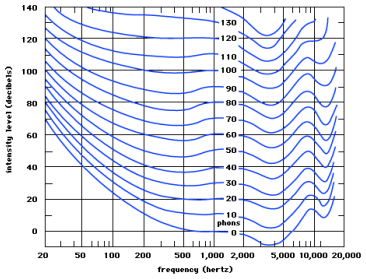
Praxisteil

Hörschwellenversuch



Unsere Hörschwelle

Um die Hörschwelle zu erkunden haben wir Signale von 50Hz, 2500Hz, 5000Hz, 7500Hz, 10000Hz, 12500Hz und 15000Hz so leise gestellt, dass wir sie gerade noch wahrnehmen konnten und dann in die Kurve links die Lautstärken eingetragen.



Wenn man unsere Hörschwelle mit der „Referenz“ (Fletcher/Munson Graphik rechts) vergleicht und dabei beachtet, dass unsere Skala nicht logarithmisch ist, dann sind die Kurven sehr ähnlich und wir haben super gemessen ;-)

Verdeckungsversuch

In diesem Versuch geht es um die Verdeckung von in der Frequenz benachbarten Signalen.

Dazu haben wir zu einem 200Hz Signal ein 210Hz Signal angelegt. Zu hören war vor allem die Schwebung (durch Überlagerung von zwei Schwingungen entsteht eine tieferfrequente Schwingung).

Bis zu einer Lautstärke von 2500 war das Vorhandensein des zweiten Signals durch die Schwebung, die in einem wesentlich tieferen Frequenzbereich liegt, hörbar.

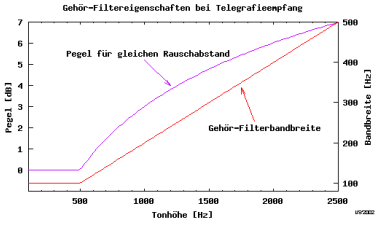
Der Ton B alleine kam erst ab 2600 in den für uns hörbaren Bereich.

Wenn wir nicht auf die Schwebung hören, so ist es natürlich so, dass der lautere Ton den ganz nahe liegenden leiseren sofort maskiert...

Den 300Hz Ton haben wir daher natürlich auch schon früher gehört (bzw. bei gleicher Lautstärke viel intensiver), weil die Maskierungskurve des 200Hz Signals bei 300Hz schon deutlich niedriger liegt...

Den ganzen Versuch haben wir dann nochmal mit 5000Hz, 5100Hz und 7500Hz durchgeführt und dabei festgestellt, dass es sich hier fast genauso verhält, obwohl die Frequenzen hier viel weiter auseiander liegen.

Das lässt sich dadurch erklären, dass wir bei tieferen Frequenzen genauer hören, also feinere Unterschiede wahrnehmen.

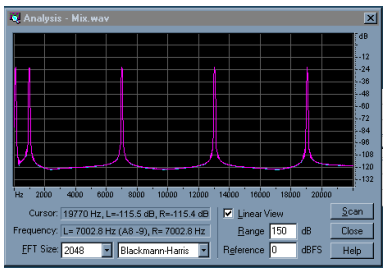


Zeitliche Maskierung

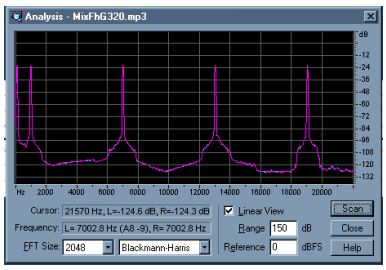
Der Versuch sollte die Wirkung der Maskierung im Zeitbereich zeigen. Dazu hatten wir ein Sample mit einem Schlag und dahinter Rauschen. Das Rauschen sollten wir so weit abschneiden, bis wir es nicht mehr wahrnehmen konnten.

Der Versuch ergab, dass das Rauschen durch den Schlag 62ms lang maskiert wird (Recovery Time).

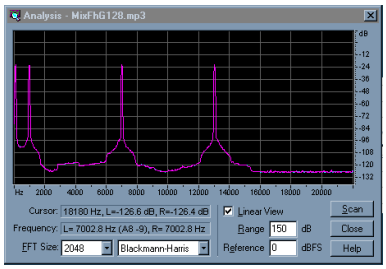
Kodierverhalten einzelner Frequenzen



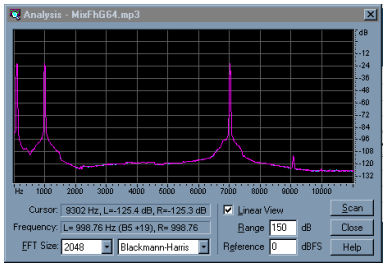
Original



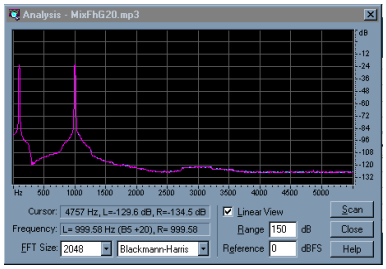
FHG 320kb/s



FHG 128kb/s



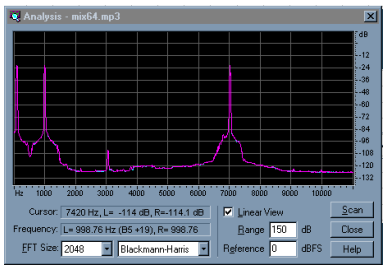
FHG 64kb/s (! andere Skala !)



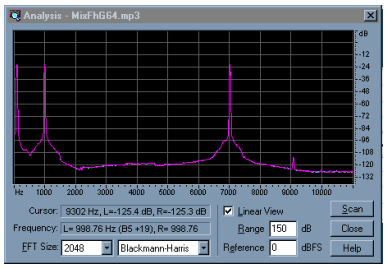
FHG 20kb/s

Für diesen Versuch haben wir auf eine 100Hz Sinusschwingung weitere Sinusschwingungen mit 1000Hz, 7000Hz 13000Hz und 19000Hz aufmoduliert und das ganze anschließend mit verschiedenen Kodiereinstellungen und Encodern kodiert.

Im Vergleich von Lame und Frauenhofer-Codec haben wir gesehen, dass der Lame etwas mehr Quantisierungsrauschen (eingeebneteres Spektrum) hat, vom Frequenzgang waren beide aber sonst ziemlich ähnlich:



Lame 64kb/s



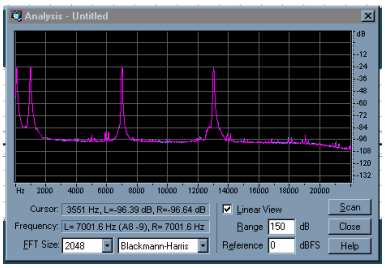
FHG 64 kb/s

Am besten erhalten blieben trotz hoher Kompressionsrate die beiden tiefsten Frequenzen (100 und 1000 Hz). Sie waren auch mit 20kbps bei durchweg allen verwendeten Enodern noch „erkennbar“ (vgl. Gehör).

Die Frequenz 7000Hz wurde bis 64kbps noch korrekt wiedergegeben.

Die beste Kodierung hinsichtlich der Frequenzen lieferte der Encoder ins Real-Format.

Obwohl ein starkes Grundrauschen auftrat, konnte er bis 64kbps auch die Frequenz 13000Hz noch erhalten:



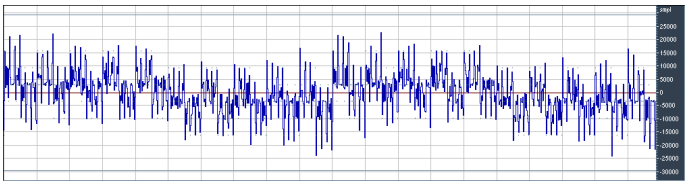
Allerdings erzeugte der Encoder Dateien von fast vierfacher Größe gegenüber den mp3-Encodern.

Interessant fanden wir auch noch, dass die mp3-Dateien beider Encoder gleich groß waren, was aber aufgrund der Spezifikation des Formates logisch ist (feste Framegröße).

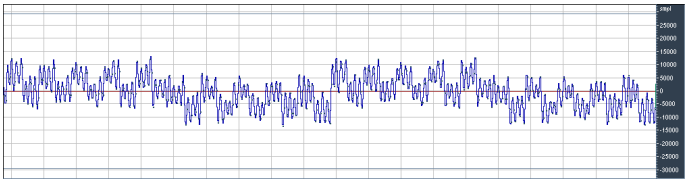
Rechtecksignal

Den gleichen Versuch galt es jetzt noch einmal mit einem Rechtecks-signal anstelle der Sinuskurve durchzuführen.

Wie der Name schon sagt, ist das Rechtecksignal eckig. Wie wir davor ja gesehen haben ebnen die Encoder das Signal ein, das heißt, wir erwarten, dass das signal runder wird...



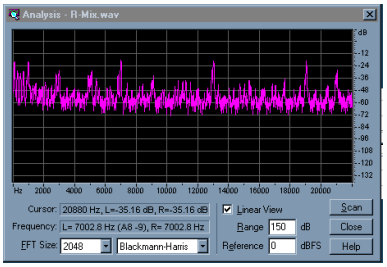
Original



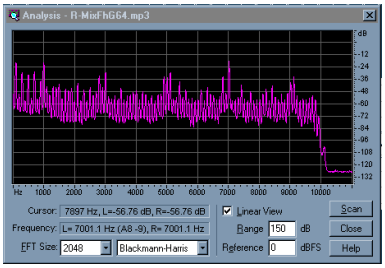
FHG 64kb/s

Unsere Vermutung hat sich bestätigt. Das Signal wurde im Ortsraum „abgerundet“.

Der Verlust der hohen Frequenzen war vergleichbar mit den komprimierten Sinusschwingungen aus der vorigen Aufgabe, nur, dass diesmal keine einzelnen Frequenzen im Spektrum hervorstachen, sondern alles ein „großes Rauschen“ war:

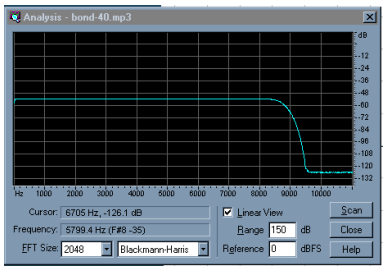


Original



FHG 64kb/s (! andere Skala !)

Alle Frequenzen



Frequenzverlauf bei 22,05kHz Samplingrate

In diesem Versuch sollte untersucht werden, ab welcher Frequenz eine von 10Hz bis 20000Hz aufsteigende Sinusschwingung (also alle Frequenzen dazwischen vorhanden) bei der Komprimierung mit verschiedenen Auflösungsstufen (Sample Rates) abgeschnitten wird.

Zu erwarten war, dass nach dem Abtasttheorem die höchste dargestellte Frequenz der halben Abtastrate (Nyquist-Frequenz) entspricht.

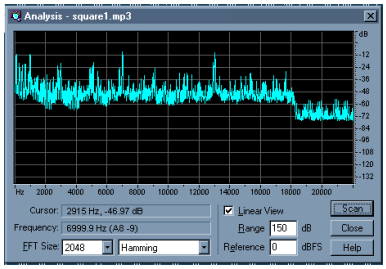
Allerdings fiel die Amplitude der Sinusschwingung schon viel früher als erwartet ab. Zum Beispiel trat bei der Quantisierung mit 22050Hz erhöhtes Quantisierungsrauschen schon ab 9000Hz auf; ab 9500Hz wurde das Signal komplett abgeschnitten.

Grund dafür könnte das psychoakustische Modell sein, bzw. die Zerlegung in die Subbänder, die besagen, dass uns der Verlust der höheren Frequenzen weniger stört, als der der tiefen?

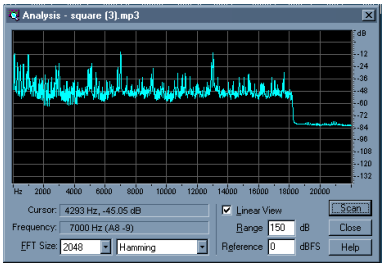
Mehrmalige Kodierung

Grundsätzlich bleibt das Signal auch nach mehrmaliger Kodierung (entgegen unserer Erwartung) erstaunlich gut erhalten - zumindest, was das Hören angeht.

Bei der Frequenzanalyse sieht man schon deutlich, dass nach mehrmaligem Kodieren die Teile, die stark quantisiert werden schlechter geworden sind:

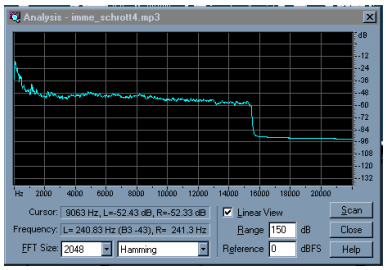


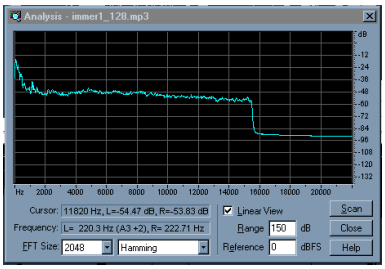
einmal



mehrmals

Bei dem Musikstück war es anders herum. Im Spektrum sieht man fast nichts:





Dafür hört man bei beiden Encodern nach ca. vier Kodierungen ein leichtes „Scheppern“, das wohl auf die stetige Verstärkung des Quantisierungsrauschens zurückzuführen ist, wenn man es auch nicht sieht (ab 10x Rauschen/ Dumpfklingen).

Der Lame Encoder schnitt insgesamt schlechter ab, weil bei ihm die Bassfrequenzen nach einigen Kodierungsvorgängen verloren gingen.

Die Dateigröße bleibt auch nach mehrmaligem Kodieren gleich, weil sie durch die Sample Rate explizit festgelegt wird.

Hörtest

Kodiert wurden hier drei Audiofiles mit unterschiedlichen Bitraten: Ein Ausschnitt aus einem sehr dynamikreichen klassischen Stück (Edvard Grieg, Peer Gynt Suite, Tanz in der Höhle des Bergriesen), ein Ausschnitt aus einem Pop-Stück mit exakt differenzierten Percussionanteilen (Beatbetrieb, Für immer) und dem Anfang eines Konzertes mit Sprache und Applaus (The Corrs, live).

Besonders bei klassischer Musik fällt der Qualitätsverlust bei niedrigeren Bitraten auf. Dabei komprimiert wiederum der Encoder der Fraunhofer Gesellschaft besser als der Lame Encoder. Bei diesem sind insbesondere bei niedrigen Bitraten wie 32kbps deutlicheres Quantisierungsrauschen und „Verwischen“ zu hören.

Auch bei dem Popstück fällt auf, dass der Klang stumpf und arm an klaren Höhen und Tiefen wird.

Das Klatschen wird bei niedrigen Bitraten zu einem undifferenzierten Rauschen.