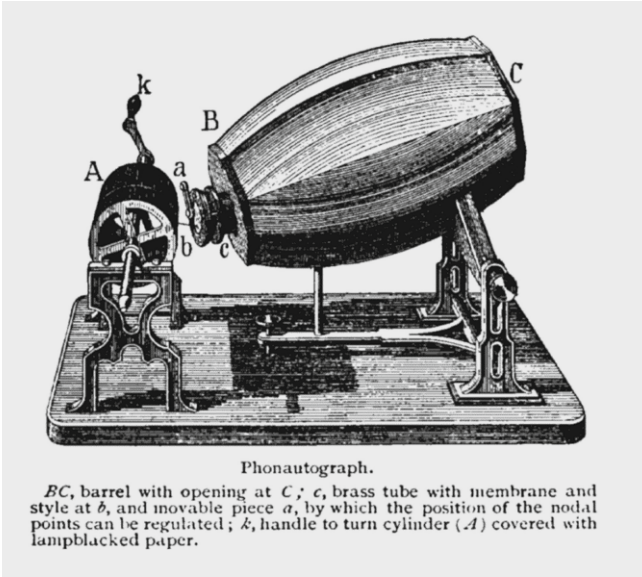
Digitale Tonaufzeichnung

Spagat zwischen Datenvolumen und Informationsverlusten

Teil 1

Hören ist ein höchst subjektiver Vorgang, bei dem das Gehör die Schallwellen analysiert und dabei Unwichtiges von Wichtigem trennt. Mit der elektronischen Nachbildung eines psychoakustischen Modells lässt sich dies bereits bei der Audioaufnahme vorwegnehmen und so die zu speichernde Datenmenge verringern.

Die Geschichte der Aufzeichnung und dauerhaften Speicherung von Schallereignissen (Geräusche, Töne, Musik und Sprache) begann 1860 mit der Anmeldung eines französischen Patents durch Édouard-Léon Scott de Martinville, das ihm die geistige Urheberschaft am „Phonautograph" beurkundet [1]. Sein Phonautograph besteht aus einem großen, trommelartigen Trichter, der die das Schallereignis repräsentierenden Luftdruckschwankungen auf eine dünne Membran leitet (Bild 1). Eine daran befestigte Schweineborste zeichnet diese im wahrsten Wortsinn in die Oberfläche einer rotierenden, rußgeschwärzten Walze. Damit gelang zwar die Aufzeichnung, eine Wiedergabe war freilich erst 148 Jahre später mit moderner Computertechnik möglich, die den Schall aus seinem optischen Abbild rekonstruierte.



Phonautograph.

BC, barrel with opening at *C*; *c*, brass tube with membrane and style at *b*, and movable piece *a*, by which the position of the nodal points can be regulated; *k*, handle to turn cylinder (*A*) covered with lampblacked paper.

Bild 1: Mit dem Phonautograph konnte schon 1860 Schall im echten Wortsinn „aufgezeichnet", aber nicht wiedergegeben werden.

Dennoch finden sich im Scott'schen Phonautographen alle Grundelemente eines Tonaufnahmegeräts: Die Schalltrommel mit Membran und Schweineborste fungiert als Mikrofon, die geschwärzte rotierende Walze ist das Speichermedium. Erst 1878 allerdings, als Edison seinen Phonographen vorstellte, wo die Schallschwingungen in mechanisch abtastbare Veränderungen der Trommeloberfläche (Gravuren) umgesetzt wurden, war die Aufnahme-Wiedergabe-Kette vollständig (Bild 2) [2].

Bild 2: Mit Edisons Phonograph ließ sich Schall speichern und reproduzieren.

Analog – mobil – digital

Die Technik von Tonaufnahme und -wiedergabe für jedermann war viele Jahrzehnte durch Schallplatte und Magnetband analog geprägt. Den ersten Schritt in das Zeitalter des mobilen Audiokonsums machte Sony Mitte 1979 mit

seinem Magnetbandkassetten-basierten „Walkman“ (Bild 3). Er entwickelte sich schnell zum Kultobjekt und Statussymbol der jungen Generation, weil er dank seiner Kompaktheit erstmals die hochwertige Musikwiedergabe über Kopfhörer zu jeder Zeit an jedem Ort ermöglichte. In zahllosen Varianten wurde dieser Gerätetyp von nahezu jedem Unterhaltungselektronikhersteller weltweit gut 30 Jahre lang hergestellt.

Mit ersten Versuchen Anfang der 1970er Jahre, elektrische Audiosignale zu digitalisieren, begann das Zeitalter der Aufzeichnung und Wiedergabe auf der Grundlage von Bits und Bytes. In schneller Folge verdrängten CD-Player und DAT-Rekorder (Digital Audio Tape), beide mit bewegtem Speichermedium (CD und Magnetband) den analogen Kassettenrekorder. Durch den Siegeszug der Festkörperspeicherchips sind heute Audioaufzeichnungs- und -wiedergabegeräte vollständig frei von Mechanik.

Weil digitale Datenspeicher in der Anfangszeit bei hohem Preis geringe Kapazitäten hatten, dachte man über eine optimale Ausnutzung des wertvollen Guts Speicherplatz nach. Die Lösung bestand in effizienten Codierverfahren, von denen MP3 (noch) das verbreitetste ist. Heute, wo Speicher in Gestalt von Flash-EEPROMS bei enorm gestiegener Kapazität zu einem Bruchteil früherer Preise verfügbar ist, kann man aber durchaus den MP3-Kompromiss zwischen verlustbehafteter Codierqualität und erhöhter Speichereffizienz aufgeben und verlustfreie Codierverfahren zur Anwendung bringen.

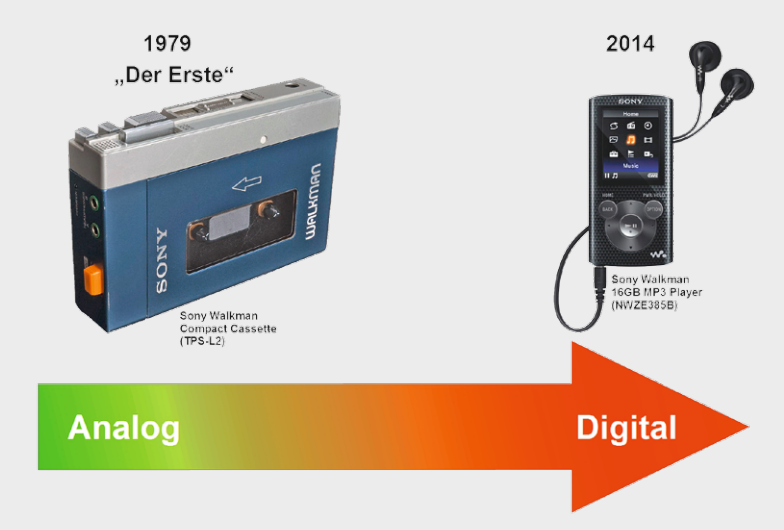


Bild 3: Mit Sonys Walkman begann 1979 die Ära der Herstellung von mobilen Tonkonserven und ihres Konsums auf der Grundlage der Magnetbandkassette. Heute dienen reine Halbleiterchips der Speicherung – Mechanik hat ausgedient.

Die digitale Übertragungskette

Reale Töne haben als Luftdruckschwankungen, die am Beginn der digitalen Übertragungs- und Verarbeitungskette die Mikrofonmembran und an deren Ende das Trommelfell im Innenohr zum Schwingen bringen sollen, eine analoge Natur. In Bild 4 sind die grundlegenden Schritte bei der digitalen Signalverarbeitungskette dargestellt. Wir sehen eine Aneinanderreihung von Baugruppen, die vom zu verarbeitenden analogen Eingangssignal $s_E(t)$ von links gespeist wird und an deren Ausgang auf der rechten Seite ein wiederum analoges Signal $s_A(t)$ zur Ansteuerung eines Kopfhörers oder Lautsprechers zur Verfügung steht.

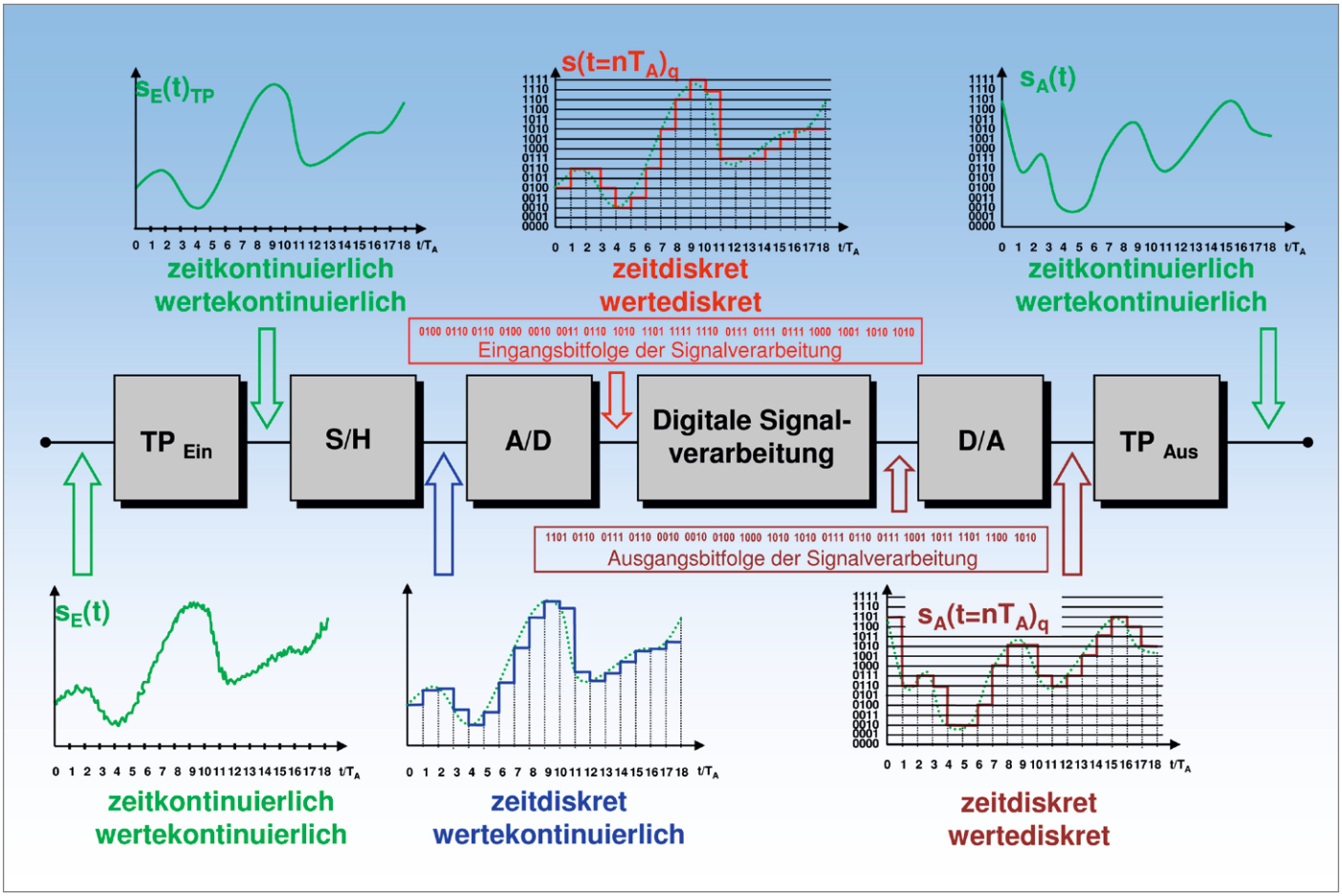


Bild 4: Die Glieder der digitalen Übertragungskette

Das analoge Eingangssignal $s_E(t)$ wird bei Audioanwendungen von einem Mikrofon erzeugt. Es enthält zu jedem Zeitpunkt differierende Werte und wird deshalb als zeit- und wertekontinuierlich bezeichnet. Ein Eingangstiefpass befreit es von höherfrequenten Anteilen, um die Verletzung des Abtasttheorems (Nyquist-Shannon-Theorem) zu vermeiden. Zur Erinnerung: Jedes Signal lässt sich mittels Fourier-Transformation in eine Summe aus harmonischen Schwingungen zerlegen. Durch eine Bandbegrenzung per Tiefpass wird sichergestellt, dass der höchsten verbleibenden Frequenzkomponente mindestens zweimal pro Periode ein Wert entnommen werden kann (Abtastung), ohne dass Doppeldeutigkeiten (Alias-Effekte) beim abgetasteten Signal entstehen.

Das bandbegrenzte Signal wird in der S/H-Stufe (Sample and Hold = Abtasten und Halten) zu den Zeitpunkten nT_A (T_A: Abtastperiode, $0 \leq n \leq n_{max}$) abgetastet (Sample = Werteentnahme) und steht dem darauf folgenden A/D-Wandler bis zum nächsten Abtastwert zur Verfügung (Hold = Halten). Daher ist das Signal vor der A/D-Wandlung (A/D: Analog → Digital) bereits stufig, aber mit unendlich fein aufgelöster Stufenhöhe. Man spricht von einem zeitdiskreten, aber wertekontinuierlichen Verlauf. Die A/D-Wandlung ordnet den wertekontinuierlichen Eingangsfolgewerten mehr oder weniger genaue digitale Zahlenäquivalente zu. Je länger die entsprechenden Bitfolgen, umso genauer ist die Abbildung. Am Ausgang des A/D-Wandlers steht nun eine Bitfolge zur Verfügung, welche das Eingangssignal zeit- und wertediskret digital angenähert repräsentiert. In dieser Form kann es transportiert und verarbeitet werden, um danach von einem D/A-Wandler (D/A: Digital → Analog) in eine zeit- und wertediskrete analoge Treppenkurve übersetzt zu werden. Ein Ausgangstiefpass (Rekonstruktionstiefpass) „verschleift" die Stufen und stellt an seinem Ausgang ein analoges, wiederum zeit- und wertekontinuierliches elektrisches Signal bereit.

Typische Digitalisierungsfehler

Allein bei der Digitalisierung eines Analogsignals entstehen charakteristische Fehler, die aber durch eine Verkleinerung der Abtastperiode und Erhöhung der Auflösung des A/D-Wandlers zumindest theoretisch beliebig minimiert werden können. In der Praxis verbleibt aber eine Granularität durch den Abtastfehler und den Quantisierungsfehler, die den Charakter eines überlagerten Rauschsignals hat.

Das Beispiel einer exponentiell gedämpften Sinusschwingung verdeutlicht das Gesagte. In Bild 5 werden 2,5 Perioden des analogen Eingangssignals 100-mal abgetastet und mit einer Wortlänge von 8 Bit quantisiert. Man erkennt, dass besonders an den Stellen, wo das Analogsignal zwischen zwei Abtastungen starke zeitliche Veränderungen aufweist, der Quantisierungsfehler ansteigt.

Entnimmt man dem gleichen Signalabschnitt 5000 Proben und quantisiert diese mit 32 Bit, sind die Fehler vernachlässigbar und quasi nur unter der Lupe zu sehen (Bild 6).

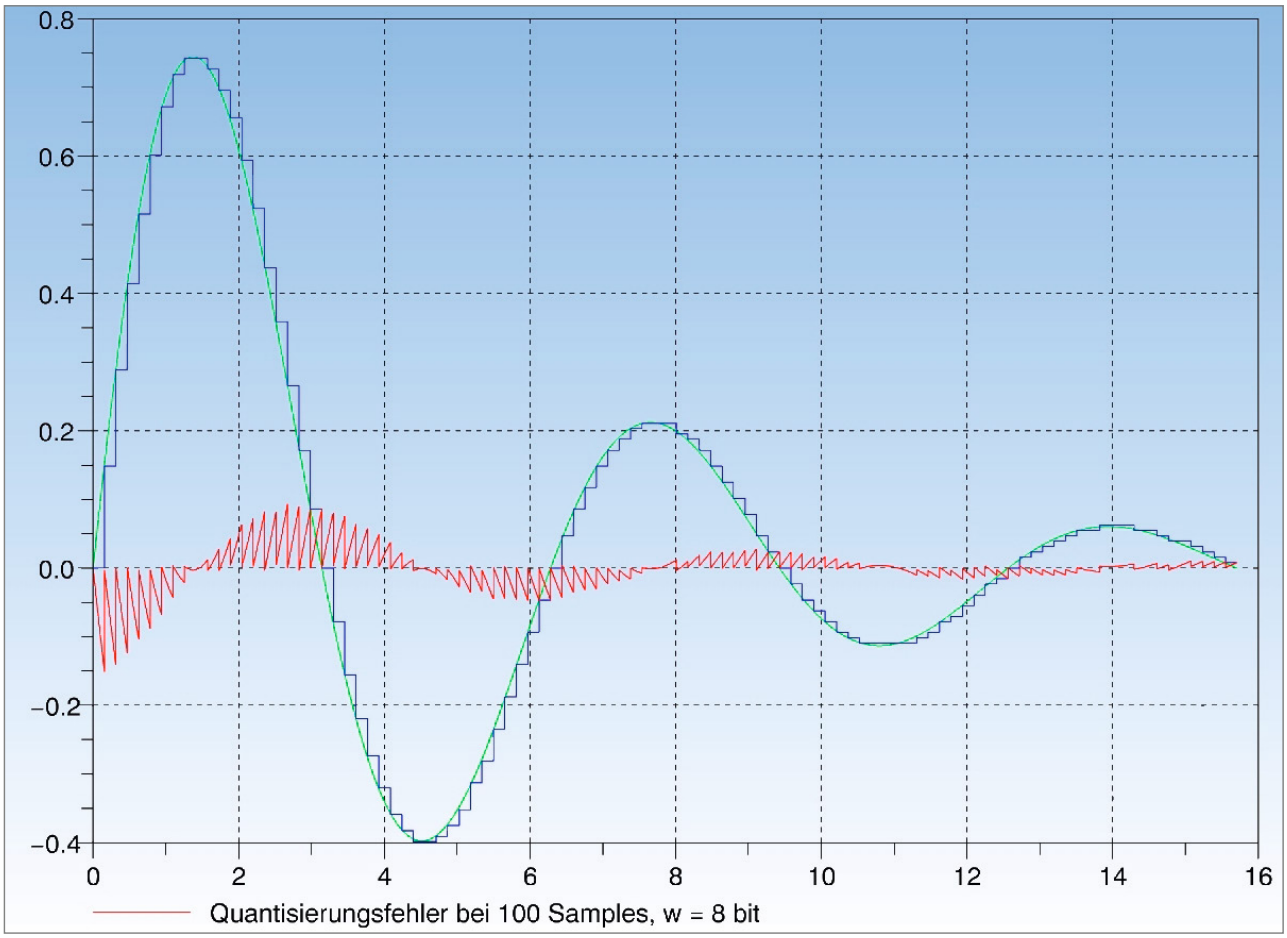


Bild 5: Die Häufigkeit und Genauigkeit ...

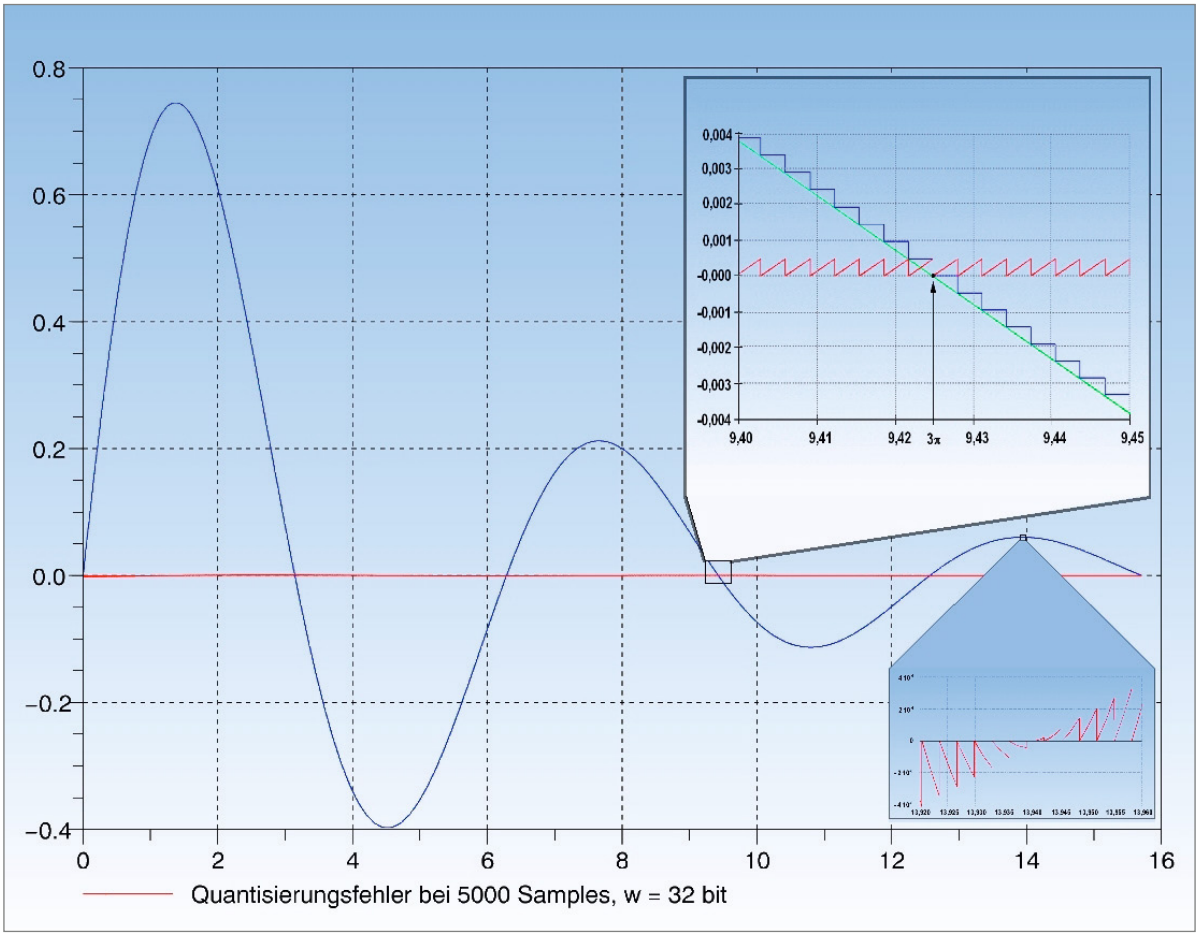


Bild 6: ... der Abtastung bestimmt, wie nah die digitale Repräsentation an das analoge Original herankommt.

Klirrfaktor und Signal-Rausch-Abstand

Bei der Umwandlung analoger Signale in Digitalwerte spielt, sofern das Nyquist-Kriterium eingehalten ist, die Auflösung (Wortlänge) des A/D-Wandlers die wichtigste Rolle. Je höher sie ist, umso kleiner sind der Quantisierungsfehler und die damit verbundenen Klirr- und Rauschanteile im Digitalsignal. Die nachfolgend angegebenen Formeln für Klirrfaktor und Geräuschabstand gelten unter der Voraussetzung eines statistisch gleichverteilten, mit der Nyquist-Frequenz abgetasteten Eingangssignals (keine Überabtastung) und vollständig genutzter, gleich hoher Quantisierungsstufen (lineare Quantisierung und Vollaussteuerung) [3].

Unter diesen Voraussetzungen berechnet sich der minimale Klirrfaktor aus der Zahl der Quantisierungsstufen q gemäß Gleichung (1).

$$\frac{k}{\%} = \frac{1}{\sqrt{q^2-1}} \quad \textit{minimaler Klirrfaktor eines quantisierten Signals} \qquad (1)$$

Fehler im Analogkreis und die Tatsache, dass der A/D-Wandler oft nicht voll ausgesteuert ist (was einer Reduzierung von q entspricht), vergrößern den Klirrfaktor in der Praxis regelmäßig.

Ebenso hängt der Signal-Geräusch-Abstand s (manchmal auch als Fremdspannungsabstand bezeichnet) von der Anzahl der Quantisierungsstufen q gemäß Gleichung (2) ab.

$$\frac{s}{dB} = -20\log(k) \quad \textit{Signalgeräuschabstand eines quantisierten Signals} \qquad 2)$$

Einige repräsentative Werte für k und s aus den Gleichungen 1 und 2 zeigt die Tabelle in Bild 7: Die Faustregel „1 Bit mehr erhöht den Signal-Geräusch-Abstand um 6 dB" sieht man hier schön bestätigt. In Rot sind die Werte einer CD eingetragen. Klirrfaktoren von 0,0015 % werden selbst von guten CD-Playern und Audioendstufen kaum erreicht.

Bei einer Standard-Audio-CD (CD-A) werden die Audiodaten mit 44,1 kHz abgetastet und als 16-Bit-Werte aufgezeichnet. Für zwei Stereokanäle ergibt dies rechnerisch eine Datenrate von 1.411.200 Bit/s, die aber wegen zusätzlich aufgezeichneter Metadaten geringfügig höher ist.

Das menschliche Gehör

Die Bandbreite des menschlichen Gehörs liegt zwischen 16 Hz und 20 kHz, wobei die obere Frequenzgrenze nur für Säuglinge gilt. Mit zunehmendem Alter sinkt die Hörfrequenzgrenze und dürfte bei den meisten Erwachsenen unter 12 kHz liegen. Dem trägt die CD-A mit einem Frequenzumfang von 5 Hz bis 20 kHz und einem Dynamikumfang von 96 dB voll Rechnung.

Es sei darauf hingewiesen, dass bei der Audio-CD (CD-A) keine Kompressionsverfahren zur Erhöhung des Dynamikumfangs oder Reduzierung der Bitrate zum Einsatz kommen (lineare PCM). Die damit verbundenen Randeffekte treten nicht auf.

Alle elektronischen Audiocodierverfahren machen sich spezifische Eigenheiten des menschlichen Gehörs zunutze, um die entstehende Datenrate so klein wie möglich bei gleichzeitig nicht oder zumindest nur minimal wahrnehmbaren Wiedergabeverlusten zu halten.

Eine wesentliche Eigenart des Gehörs ist, dass es durch die physikalischen Eigenschaften der Basilarmembran im Innenohr (Cochlea) Schallereignisse in 24 Frequenzgruppen (critical bands) wahrnimmt, deren Inhalte vom Gehirn jeweils gemeinsam ausgewertet werden [2].

n (Anzahl der Bits)	1	4	8	12	16	20	24	28	32
q (Anzahl der Quantisierungsstufen = 2^n)	2	16	256	4.096	65.536	1.048.576	16.777.216	268.435.456	4.294.967.296
k (Klirrfaktor)	58 %	6,3 %	0,39 %	0,024 %	0,0015 %	0,00010 %	0,0000060 %	0,00000037 %	0,00000002 %
s (Signal-Geräusch-Abstand)	4,77 dB	24,1 dB	48,2 dB	72,2 dB	96,3 dB	120 dB	144 dB	169 dB	193 dB

Bild 7: Je mehr Quantisierungsstufen, umso geringer ist das Klirrgeräusch und umso höher der Signal-Rausch-Abstand.

Bandnummer	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Bandmittenfrequenz / Hz	50	150	250	350	450	570	700	840	1000	1170	1370	1600	1850	2150	2500	2900	3400	4000	4800	5800	7000	8500	10500	13500
Bandbreite / Hz	100	102	104	109	114	122	133	145	162	182	207	239	277	325	386	460	559	685	867	1111	1426	1851	2463	3468
Berechnete obere Bandgrenzen / Hz	100	200	300	400	510	630	770	910	1080	1260	1470	1720	1990	2310	2690	3130	3680	4340	5230	6360	7710	9430	11730	15230
Publizierte obere Bandgrenzen / Hz	100	200	300	400	510	630	770	920	1080	1270	1480	1720	2000	2320	2700	3150	3700	4400	5300	6400	7700	9500	12000	15500

Bild 8: Die 24 Subbänder des menschlichen Gehörs nach Zwicker

Das Gehör fasst also die Signalleistungen aller Schallereignisse in einer Frequenzgruppe zu einem subjektiven Lautstärkeeindruck zusammen. Weil in jedem der „Critical Bands" eine getrennte Schallanalyse vorgenommen wird, ist, vereinfacht gesagt, ein „Critical Band" der Audiofrequenzbereich, in dem ein zweiter leiserer Ton gleichzeitig mit der Wahrnehmung eines ersten lauteren Tons zu einem kombinierten Ton verschmilzt und nicht mehr als eigener Ton in Erscheinung tritt. Er wird quasi „verdeckt" (spektrale Maskierung). Erst wenn der zweite Ton in einem anderen „Critical Band" liegt, kann er als eigenständiger Ton erkannt werden. Das ist physiologisch begründet durch den Ort auf der Cochlea, an dem die dort angesiedelten Haarzellen als Wandler von mechanischen Schwingungen in elektrische, zum Gehirn geleitete Nervensignale durch eine bestimmte Frequenz maximal angeregt werden. So wird die Verarbeitung akustischer Ereignisse im Gehirn beschleunigt und damit die Reaktionsgeschwindigkeit des Hörers auf Kosten einer detaillierteren Analyse des in das Ohr einfallenden Schalls erhöht. Zudem wird das räumliche Hören (Richtungsbestimmung) erleichtert [4].

Zusammenfassend ist festzustellen: Zur Zerlegung des Hörfrequenzbereichs in 24 Frequenzgruppen bildet das Gehör eine Filterbank nach. Aus den Mittenfrequenzen f_C der Filterbankfrequenzgruppen berechnen sich nach der aus Versuchsreihen abgeleiteten Gleichung (3) deren Bandbreiten Δf_G.

$$\frac{\Delta f_G}{Hz} = 25 + 75\left[1 + 1{,}4\left(\frac{f_C}{kHz}\right)^2\right]^{0{,}69} \qquad (3)$$

Die berechneten und daraus abgeleiteten oberen Bandgrenzen sind der Tabelle in Bild 8 zu entnehmen.

Man erkennt, dass die Bänder umso schmalbandiger sind, je niedriger ihre Bandmittenfrequenzen liegen. Bis zu einer Mittenfrequenz von 500 Hz ist die Frequenzgruppenbreite näherungsweise konstant etwa

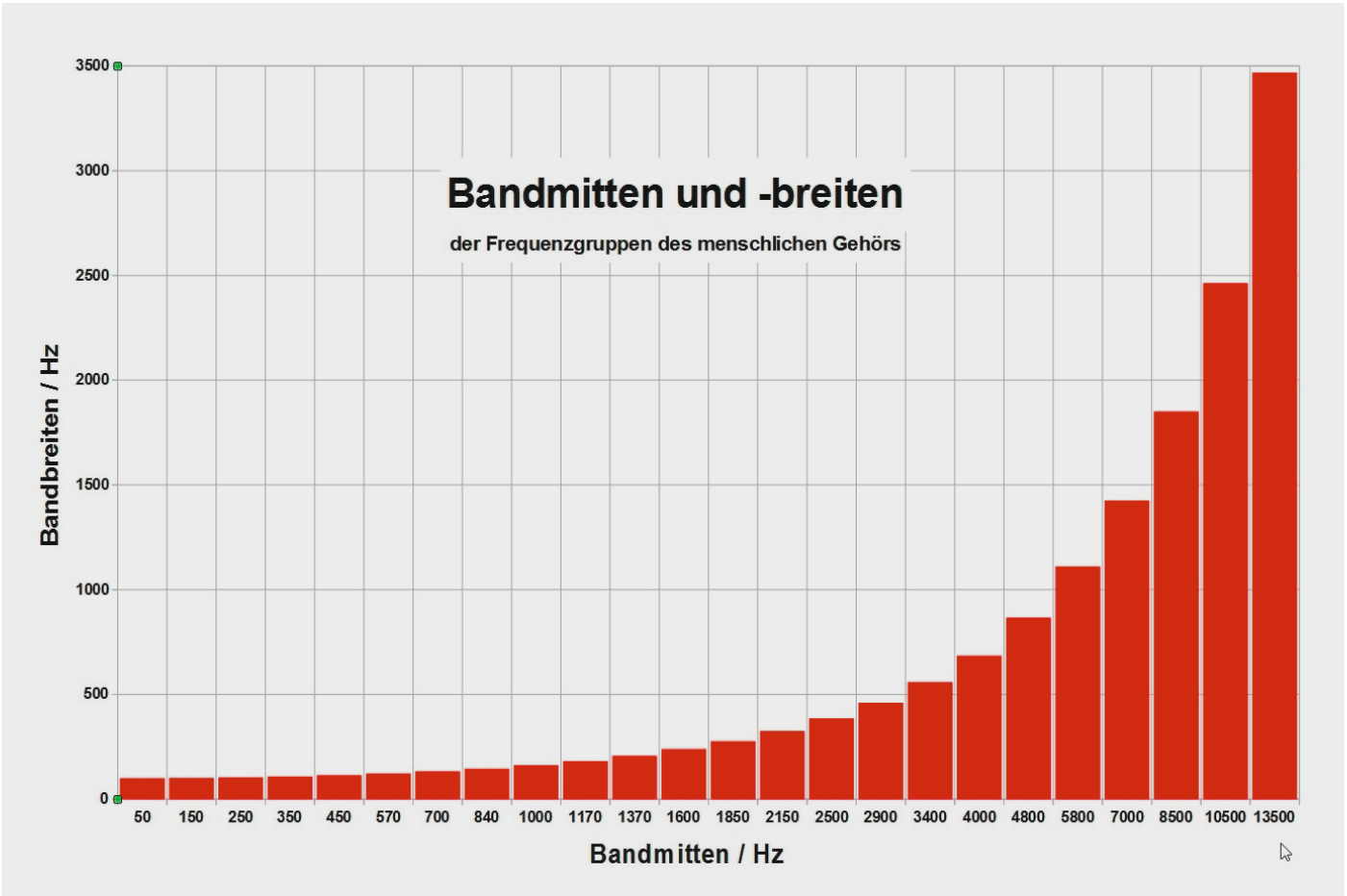


Bild 9: Je niedriger die Mittenfrequenz eines Subbandes, umso schmaler ist es. Die Frequenzauflösung des Gehörs ist also bei tiefen Tönen besser als bei hohen.

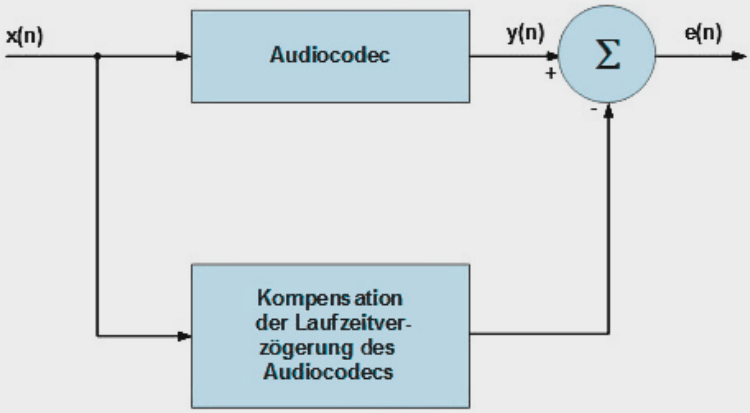


Bild 10: Wenn die Encoder-Decoder-Kette (Codec) einen Codierfehler liefert, der unter der Mithörschwelle liegt, ist er nicht wahrnehmbar.

100 Hz, darüber nimmt sie auf über 20 % der Mittenfrequenz zu. Eine Visualisierung der tabellarischen Daten zeigt Bild 9.

Aus dem Gesagten kann man schließen, dass bei niedrigen Frequenzen Töne mit geringeren absoluten Tonhöhendifferenzen unterschieden werden können als bei höheren Frequenzen.

Ergänzend sei noch bemerkt, dass kurz vor und nach einem sehr lauten Geräusch, z. B. einem Knall, leisere Geräusche nicht wahrgenommen werden (temporäre Maskierung).

Audiocodecs

Das Wort Codec beschreibt eine Anordnung aus einem Audiosignal-**Co**dierer (dessen Ausgangssignale weiterverarbeitet oder gespeichert werden) und einem **Dec**odierer (der wieder ein Audiosignal erzeugt). Ist ein solcher Codec verlustfrei, ist sein Ausgangssignal y(n) identisch mit dem Eingangssignal x(n). Ist er verlustbehaftet, produziert er einen Codierungsfehler e(n) als Differenz zwischen dem uncodierten Signal x(n) und dem decodierten y(n) (Bild 10). Bei der Qualitätsbeurteilung eines Codecs gilt es nun zu bewerten, in welchem Maß das Fehlersignal e(n) zusätzlich zum Originalsignal x(n) hörbar ist. Grundsätzlich wird e(n) nur dann nicht wahrgenommen, wenn es eine vom Originalsignal abhängige sogenannte Mithör- oder Maskierungsschwelle unterschreitet.

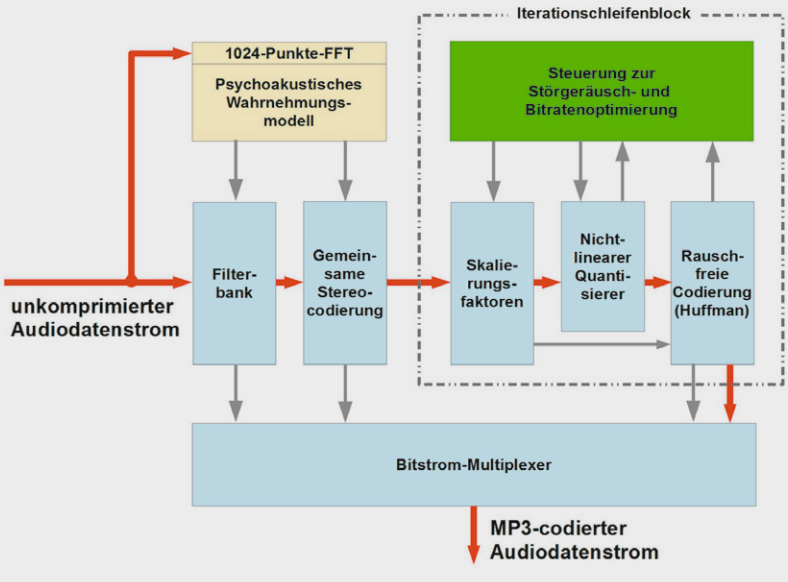


Bild 11: Die wesentlichen Funktionsblöcke eines MP3-Coders

MP3 – Klassiker unter den Audio-Kompressionsverfahren

Entwickelt wurde das Format MP3 ab 1982 unter der Leitung von Hans-Georg Musmann von einer Gruppe um Karlheinz Brandenburg am Fraunhofer-Institut für Integrierte Schaltungen (IIS) in Erlangen sowie an der Friedrich-Alexander-Universität Erlangen-Nürnberg in Zusammenarbeit mit AT&T Bell Labs und Thomson. MP3 wurde von der International Organization for Standardization (ISO) als ISO-MPEG Audio Layer-3 standardisiert [5].

Bereits Mitte der 1990er Jahre konnte man auf dem PC MP3-Dateien erzeugen und abspielen, erste MP3-Player kamen 1998 in den Handel. Weil ein MP3-Coder psychoakustische Effekte wie vorstehend beschrieben nutzt, indem er irrelevante Anteile des Audiosignals entfernt und dadurch die Bitrate reduziert, gehört die MP3-Encodierung zu den verlustbehafteten Kompressionsverfahren. Die entfernten irrelevanten, d. h. nicht oder kaum wahrnehmbaren Signalanteile sind unwiederbringlich verloren und lassen sich aus den MP3-Daten nicht wieder rekonstruieren.

Anhand von Bild 11 soll der Codierungsablauf knapp beschrieben werden. Der unkomprimierte Datenstrom wird von dem Funktionsblock „Psychoakustisches Wahrnehmungsmodell" analysiert und stellt die Steuerparameter für eine Filterbank (Polyphasenfilterbank und modifizierte diskrete Cosinus-Transformation) bereit. Das Wahrnehmungsmodell ist hauptsächlich für die Codierqualität ausschlaggebend. Es liefert die Maskierungsschwellenwerte (unter denen das Quantisierungsrauschen liegen muss, um unhörbar zu sein) und die zulässigen Geräuschpegelabstände jeder Frequenzgruppe [6]. Die von der Filterbank gewonnenen Frequenzgruppen werden einer gemeinsamen Stereocodierung unterworfen und in einen Funktionsblock mit zwei verschachtelten Iterationsschleifen eingespeist.

Die innere Schleife dient dem Finden einer optimalen Bitrate. Der nichtlineare Quantisierer hat eine Potenzkennlinie, wodurch große Signalwerte weniger genau (bitsparend) abgebildet werden. Der Huffman-Codierer weist den häufiger auftretenden kleineren Quantisierungswerten kürzere Codewörter zu. Weil er dies verlustfrei tut, entsteht kein weiteres Rauschen. Durch eine Vergröberung der Quantisierungsschritte des Quantisierers wird die Bitrate so lange verringert, bis sie klein genug ist. Der Quantisierer bewirkt im Wesentlichen die Codierverluste, trägt aber auch den größten Anteil an der Bitratenreduktion.

Die äußere Iterationsschleife ist für das Minimieren des Störgeräuschs zuständig. Wenn die Signalintensität eines Subbandes bereits unter dessen Maskierungsschwelle liegt, ist das Signal bereits unhörbar und braucht nicht codiert zu werden. Andernfalls werden die Skalierungsfaktoren so ermittelt, dass das Störgeräusch in jedem Subband gerade unter die vom Wahrnehmungsmodell vorgegebene Maskierungsschwelle zu liegen kommt. Für jeden Iterationsschritt der äußeren Schleife muss die innere Schleife neu durchlaufen werden, damit die zulässige Bitrate nicht überschritten wird.

Bits	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Binary	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
Hex	F		F		F									B			A						0				4				0		
Meaning	MP3 Sync Word												Version	Layer		Error Protection	Bit Rate				Frequency		Pad. Bit	Priv. Bit	Mode		Mode Extension (Used With Joint Stereo)		Copy	Original	Emphasis	
Value	Sync Word												1 = MPEG	01 = Layer 3		1 = No	1010 = 160				00 = 44100 Hz		0 = Frame is not padded	Unknown	01 = Joint Stereo		0 = Intensity Stereo Off	0 = MS Stereo Off	0 = Not Copy-righted	0 = Copy Of Original Media	00 = None	

Bild 12: Im „Kopf" (header) eines MP3-Frames stehen alle wichtigen Informationen für den Decoder.

Nach erfolgter Komprimierung wird ein serieller Bitstrom mit Signal- und Metadaten (Informationen für den Decoder) erzeugt. Dieser besteht aus einer Folge von Datenblöcken (frames) mit eigenem Kopf (header) und den Audioinformationen.

MP3-Frameheader (Bild 12) sind 32 Bit (4 Byte) lang, von denen die ersten zwölf immer auf 1 gesetzt sind und zur Synchronisierung dienen. Es folgen 3 Bit mit Informationen über die MPEG-Audio-Version und den Layer (1, 2 oder 3). Ein weiteres Bit zeigt den Einsatz von CRC (Cyclic Redundancy Check, ein Prüfwertverfahren zum Erkennen von Übertragungsfehlern) an. Ist es 0, folgt unmittelbar auf den Header ein 16-Bit-CRC-Wert. Andernfalls wird kein CRC verwendet. Die nächsten 4 Bit informieren über Layerversion und Bitrate (z. B. 1001 bei MPEG 1 Layer-3 und 128 Kbit/s Bitrate) und so geht es weiter mit Abtastrate, Kanalmodus (Mono, Stereo), Copyright, Originalbit und Emphase.

Um die Dauer und Größe eines Frames zu berechnen, benötigt man Bit-, Frame- und Samplingrate [6]. Ein MP3-File mit einer Samplingrate von 44.100 Samples pro s enthält zum Beispiel 1.152 Samples pro Frame. Bei einer Bitrate von 128.000 Bit/s ergeben sich Dauer und Größe eines Frames nach Gleichung 4.

$$Framedauer = \frac{1152\frac{Samples}{Frame}}{Samplingfrequenz} = \frac{1152\frac{Samples}{Frame}}{44100\frac{Samples}{s}} = 26{,}1\,\frac{ms}{Frame} \qquad (4)$$

$$Framegröße = \frac{1152\frac{Samples}{Frame}\cdot Bitrate}{8\,\frac{Bit}{Byte}\cdot Samplingrate} = \frac{1152\frac{Samples}{Frame}\cdot 128000\frac{Bit}{s}}{8\,\frac{Bit}{Byte}\cdot 44100\frac{Samples}{s}} = 417{,}96\,\frac{Bytes}{Frame}$$

Wie man sieht, ist in diesem Beispiel die Zahl der Byte/Frame nicht ganzzahlig und muss deshalb mit einem zusätzlichen Byte aufgerundet werden (Padding). Ob dieser Fall vorliegt, wird im Header durch Bit Nr. 23 „Pad. Bit" angezeigt. Zu beachten ist, dass die relativ lan-

```
G:\Fachartikel\elv\Verlustfreie Audiocodierung\Testaudiodateien\WAV>lame -?
LAME 32bits version 3.99.5 (http://lame.sf.net)

usage: lame [options] <infile> [outfile]

    <infile> and/or <outfile> can be "-", which means stdin/stdout.

RECOMMENDED:
    lame -V2 input.wav output.mp3

OPTIONS:
  Input options:
    --scale <arg>   scale input (multiply PCM data) by <arg>
    --scale-l <arg> scale channel 0 (left) input (multiply PCM data) by <arg>
    --scale-r <arg> scale channel 1 (right) input (multiply PCM data) by <arg>
    --mp1input      input file is a MPEG Layer I   file
    --mp2input      input file is a MPEG Layer II  file
    --mp3input      input file is a MPEG Layer III file
    --nogap <file1> <file2> <...>
                    gapless encoding for a set of contiguous files
    --nogapout <dir>
                    output dir for gapless encoding (must precede --nogap)
    --nogaptags     allow the use of VBR tags in gapless encoding

  Input options for RAW PCM:
    -r              input is raw pcm
    -x              force byte-swapping of input
    -s sfreq        sampling frequency of input file (kHz) - default 44.1 kHz
    --bitwidth w    input bit width is w (default 16)
    --signed        input is signed (default)
    --unsigned      input is unsigned
    --little-endian input is little-endian (default)
    --big-endian    input is big-endian


  Operational options:
    -a              downmix from stereo to mono file for mono encoding
    -m <mode>       (j)oint, (s)imple, (f)orce, (d)ual-mono, (m)ono (l)eft (r)ig
ht
                    default is (j) or (s) depending on bitrate
                    joint  = joins the best possible of MS and LR stereo
                    simple = force LR stereo on all frames
                    force  = force MS stereo on all frames.
    --preset type   type must be "medium", "standard", "extreme", "insane",
                    or a value for an average desired bitrate and depending
                    on the value specified, appropriate quality settings will
                    be used.
                    "--preset help" gives more info on these
    --comp  <arg>   choose bitrate to achieve a compression ratio of <arg>
    --replaygain-fast   compute RG fast but slightly inaccurately (default)
    --replaygain-accurate   compute RG more accurately and find the peak sample
    --noreplaygain  disable ReplayGain analysis
    --clipdetect    enable --replaygain-accurate and print a message whether
                    clipping occurs and how far the waveform is from full scale
    --flush         flush output stream as soon as possible
    --freeformat    produce a free format bitstream
    --decode        input=mp3 file, output=wav
    --swap-channel  swap L/R channels
    -t              disable writing wav header when using --decode


  Verbosity:
    --disptime <arg>print progress report every arg seconds
    -S              don't print progress report, VBR histograms
    --nohist        disable VBR histogram display
    --quiet         don't print anything on screen
    --silent        don't print anything on screen, but fatal errors
    --brief         print more useful information
    --verbose       print a lot of useful information

  Noise shaping & psycho acoustic algorithms:
    -q <arg>        <arg> = 0...9.  Default  -q 5
                    -q 0:  Highest quality, very slow
                    -q 9:  Poor quality, but fast
    -h              Same as -q 2.   Recommended.
    -f              Same as -q 7.   Fast, ok quality


  CBR (constant bitrate, the default) options:
    -b <bitrate>    set the bitrate in kbps, default 128 kbps
    --cbr           enforce use of constant bitrate

  ABR options:
    --abr <bitrate> specify average bitrate desired (instead of quality)

  VBR options:
    -V n            quality setting for VBR.  default n=4
                    0=high quality,bigger files. 9=smaller files
    -v              the same as -V 4
    --vbr-old       use old variable bitrate (VBR) routine
    --vbr-new       use new variable bitrate (VBR) routine (default)
    -Y              lets LAME ignore noise in sfb21, like in CBR
    -b <bitrate>    specify minimum allowed bitrate, default  32 kbps
    -B <bitrate>    specify maximum allowed bitrate, default 320 kbps
    -F              strictly enforce the -b option, for use with players that
                    do not support low bitrate mp3
    -t              disable writing LAME Tag
    -T              enable and force writing LAME Tag


  MP3 header/stream options:
    -e <emp>        de-emphasis n/5/c  (obsolete)
    -c              mark as copyright
    -o              mark as non-original
    -p              error protection.  adds 16 bit checksum to every frame
                    (the checksum is computed correctly)
    --nores         disable the bit reservoir
    --strictly-enforce-ISO   comply as much as possible to ISO MPEG spec
    --buffer-constraint <constraint> available values for constraint:
                                    default, strict, maximum

  Filter options:
    --lowpass <freq>        frequency(kHz), lowpass filter cutoff above freq
    --lowpass-width <freq>  frequency(kHz) - default 15% of lowpass freq
    --highpass <freq>       frequency(kHz), highpass filter cutoff below freq
    --highpass-width <freq> frequency(kHz) - default 15% of highpass freq
    --resample <sfreq>  sampling frequency of output file(kHz)- default=automatic


  ID3 tag options:
    --tt <title>    audio/song title (max 30 chars for version 1 tag)
    --ta <artist>   audio/song artist (max 30 chars for version 1 tag)
    --tl <album>    audio/song album (max 30 chars for version 1 tag)
    --ty <year>     audio/song year of issue (1 to 9999)
    --tc <comment>  user-defined text (max 30 chars for v1 tag, 28 for v1.1)
    --tn <track[/total]>    audio/song track number and (optionally) the total
                            number of tracks on the original recording. (track
                            and total each 1 to 255. just the track number
                            creates v1.1 tag, providing a total forces v2.0).
    --tg <genre>    audio/song genre (name or number in list)
    --ti <file>     audio/song albumArt (jpeg/png/gif file, v2.3 tag)
    --tv <id=value> user-defined frame specified by id and value (v2.3 tag)
    --add-id3v2     force addition of version 2 tag
    --id3v1-only    add only a version 1 tag
    --id3v2-only    add only a version 2 tag
    --id3v2-utf16   add following options in unicode text encoding
    --id3v2-latin1  add following options in latin-1 text encoding
    --space-id3v1   pad version 1 tag with spaces instead of nulls
    --pad-id3v2     same as '--pad-id3v2-size 128'
    --pad-id3v2-size <value> adds version 2 tag, pad with extra <value> bytes
    --genre-list    print alphabetically sorted ID3 genre list and exit
    --ignore-tag-errors  ignore errors in values passed for tags

    Note: A version 2 tag will NOT be added unless one of the input fields
    won't fit in a version 1 tag (e.g. the title string is longer than 30
    characters), or the '--add-id3v2' or '--id3v2-only' options are used.
    or output is redirected to stdout.

MS-Windows-specific options:
    --priority <type>  sets the process priority:
                            0,1 = Low priority (IDLE_PRIORITY_CLASS)
                            2 = normal priority (NORMAL_PRIORITY_CLASS, default)
                            3,4 = High priority (HIGH_PRIORITY_CLASS))
    Note: Calling '--priority' without a parameter will select priority 0.

Misc:
    --license       print License information


  Platform specific:
    --noasm <instructions> disable assembly optimizations for mmx/3dnow/sse


MPEG-1   layer III sample frequencies (kHz):  32  48  44.1
bitrates (kbps): 32 40 48 56 64 80 96 112 128 160 192 224 256 320

MPEG-2   layer III sample frequencies (kHz):  16  24  22.05
bitrates (kbps):  8 16 24 32 40 48 56 64 80 96 112 128 144 160
```

Bild 13: Die Kommandozeilenparameter des LAME-Encoders sind äußerst umfangreich. Man kann sie durch Eingabe der Kommandozeile LAME -? zu Gesicht bekommen.

ge Frame-Dauer von ca. 26 ms die Exaktheit eines Schnittes begrenzt.

Informationen über Titel, Künstler, Album, Jahr, Genre sowie ein optionaler Kommentar können in einem sogenannten ID3-Tag untergebracht werden. Dieser liegt in der Version ID3 V1.0 128 Byte am Ende der Datei. Die ersten 3 Byte enthalten die Buchstaben TAG, wodurch eine Erkennung des Tags möglich ist. Folgeversionen enthalten weitere Informationen. So ist bei ID3 V2.2 der Tag bis zu 256 MByte lang und am Dateianfang angeordnet. Er speichert neben den üblichen Daten auch den Songtext oder ein eingebettetes Bild, z. B. vom Cover des Albums.

Bei der Dekompression werden die Schritte der Kompression in umgekehrter Reihenfolge ausgeführt. Nach der Huffman-Decodierung werden die Daten invers quantisiert und der Inversen Modifizierten Diskreten Cosinustransformation (IMDCT) unterworfen und einer inversen Filterbank zugeführt. Deren Ausgangsdaten werden in die ursprünglichen Samples rückgerechnet. Sie sind jetzt allerdings verlustbehaftet, aber typischerweise um den Faktor 6 bis 12 „geschrumpft".

MP3-Codierer

Wie bereits erwähnt, hat die Fraunhofergesellschaft großen Anteil an der Entstehung des MP3-Audiokompressionsverfahrens. Zusammen mit Thomson besitzt sie 18 Patente auf einige der verwendeten Teilverfahren. Ab 1998 verlangen die beiden Firmen Lizenzgebühren für den Einsatz ihrer MP3-Codecs. Deshalb sind freie, d. h. kostenlose FhG/Thomson-basierte Encoder nicht aus legalen Internetquellen zu beziehen.

Dieser Missstand hat zur Entwicklung des Open-Source-Encoders LAME geführt (**L**AME **A**in't an **M**P3 **E**ncoder = LAME ist kein MP3-Encoder, gemeint ist der FhG/Thomson-Encoder) [7]. Durch ständige Optimierungen ist LAME einer der qualitativ hochwertigsten kostenlos verfügbaren MP3-Encoder.

LAME erwartet als Input reine WAV-Dateien. WAV steht für WAVe (Welle) und ist ein Microsoft- und IBM-Dateiformat zum unkomprimierten, verlustlosen Speichern von Audio-Bitstreams auf einem PC. Die CDA-Files (CD Audio) einer Audio-CD müssen also zunächst ins WAV-Format gerippt werden, bevor sie mit LAME ins MP3-Format überführt werden können. Das geht natürlich mit Programmen wie CDex (aktuelle Version 1.72) einfacher, dafür ist die Feinheit der Parametrisierung des Encoders nicht gegeben. Details zum Encodiervorgang und einen Encodervergleich findet man im Internet [8], [9].

Als äußerst flexibles Kommandozeilenprogramm wird LAME beim Aufruf über eine Unmenge von hinter dem Programmnamen angehängten Parametern eingestellt (Bild 13). Durch die Kommandozeile

```
LAME -V2
Me_And_Bobby_McGee_---_Janis_Joplin.wav
Me_And_Bobby_McGee_---_Janis_Joplin.MP3
```

encodiert LAME z. B. die WAV-Datei „Me_And_Bobby_McGee_---_Janis_Joplin.wav" in die MP3-Datei „Me_And_Bobby_McGee_---_Janis_Joplin.mp3" (Bild 14). In diesem Beispiel muss die Datei LAME.EXE im gleichen Verzeichnis wie die zu wandelnde Datei stehen, in das auch die gewandelte Datei geschrieben wird. Das Größenverhältnis von komprimierter zu unkomprimierter Datei ist mit 5,84 : 45,32 = 0,13 beachtlich. Die MP3-Datei hat nur knapp ein Achtel der Größe des WAV-Originals.

Die meisten Anwender empfinden die Kommandozeilenbedienung von LAME als altmodisch, kompliziert und unbequem. Deshalb verstecken viele Programme, welche die Datei „LAME.exe" nutzen, diese

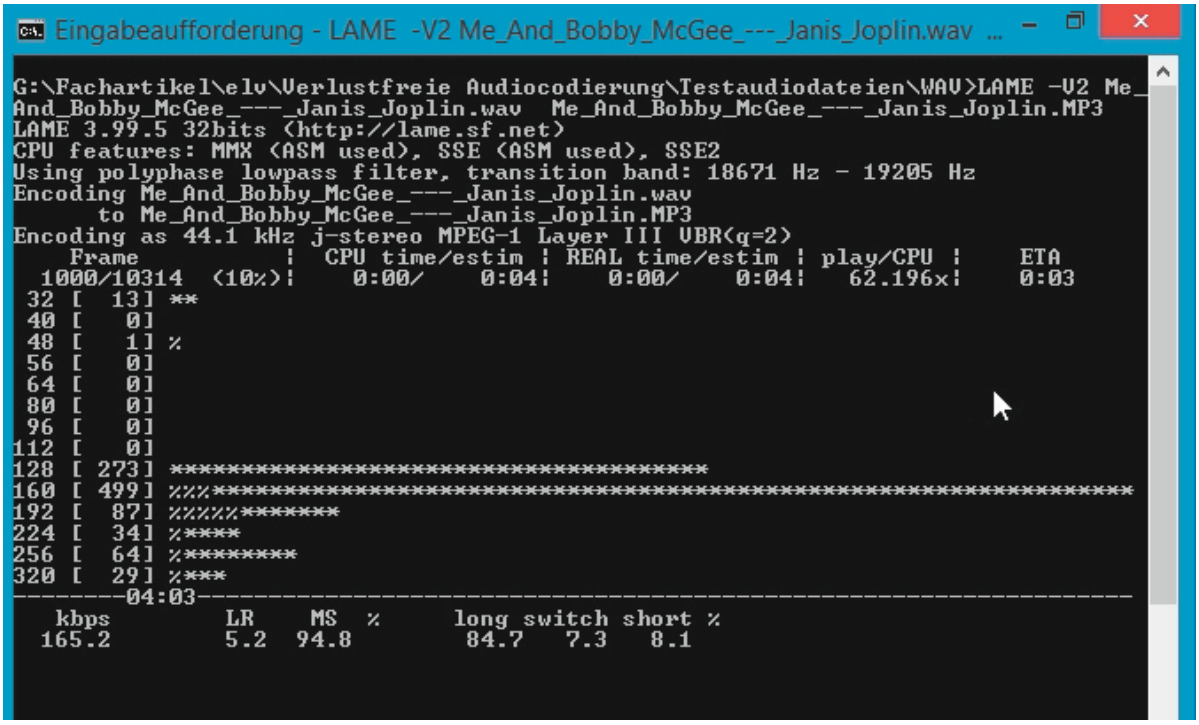


Bild 14: Das bekommt man im Eingabeaufforderungsfenster von Windows zu Gesicht, wenn man den im Text beschriebenen LAME-Befehl zum Wandeln einer Datei aus dem WAV- ins MP3-Format eingibt.

hinter einer Windows-Oberfläche, so dass der Encoder überhaupt nicht in Erscheinung tritt. Eine rein grafische Bedienoberfläche für LAME ohne weitere Funktionalität ist RazorLame (Bild 15), die auf der Homepage des Autors [10] erhältlich ist. Damit lässt sich gut die Vielfalt der Einstellungsmöglichkeiten und ihre Auswirkungen auf die Klangqualität studieren.

Ein besonderer Vorteil von LAME liegt in der Möglichkeit zur Erzeugung von MP3-Dateien mit variabler Bitrate (VBR: Variable Bit Rate). Dabei wird die Bitrate der Komplexität der zu codierenden Audiopassage dynamisch angepasst, was die Relation von Codierqualität zu Speicherbedarf optimiert. VBR liefert deshalb eine höhere Qualität bei insgesamt geringerem Speicherplatzverbrauch, verglichen mit einer Kompression mit konstanter Bitrate (CBR: Constant Bit Rate).

Zusammengefasst

LAME ist ein ausgezeichneter MP3-Encoder, der bereits bei einer Abtastrate von 128 kSamples/s ein derart transparentes Klangbild liefert, dass es dem Autor beim verwendeten Testmaterial nicht sicher möglich war, das Original von der komprimierten Kopie zu unterscheiden.

Aber Hi-Fi-Enthusiasten haben häufig ein extrem feines Gehör und wollen sich nicht mit dem MP3-typischen Entfernen von Detailinformationen abfinden oder ihre Musikschätze ohne jeden Verlust archivieren.

Sie können im zweiten Teil dieses Artikels etwas über verlustfreie Codierverfahren, das fehlerfreie Rippen von Audio-CDs und die zum Abspielen erforderliche Hardware erfahren.

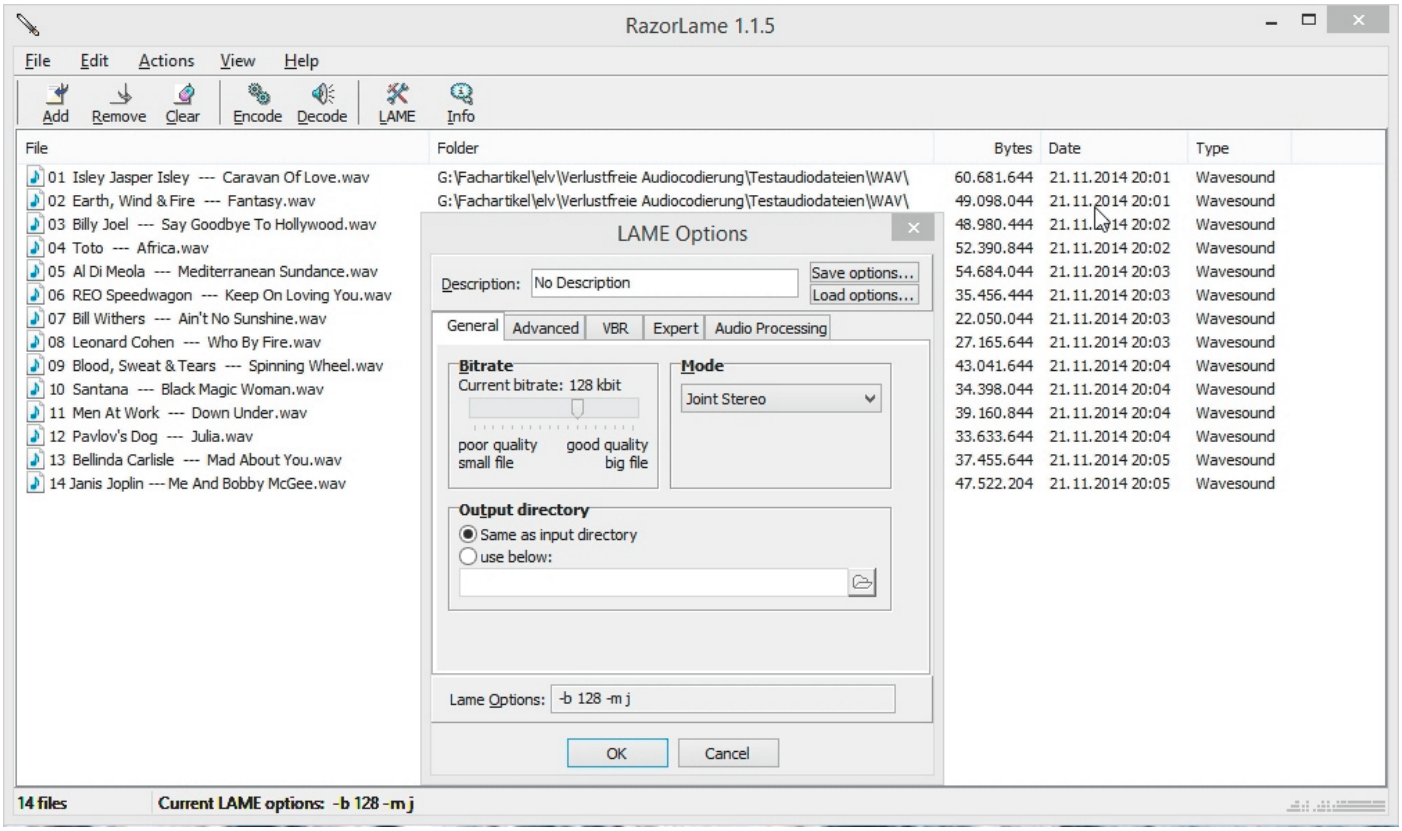


Bild 15: RazorLame ist eine grafische Bedienoberfläche für LAME, wie sie der normale PC-Nutzer gewohnt ist.

Literaturverweise:

[1] http://de.wikipedia.org/wiki/Phonautograph
[2] http://de.wikipedia.org/wiki/Phonograph
[3] https://rrzk.uni-koeln.de/fileadmin/zustaendigkeiten/multimedia/Digitale_Audiobearbeitung.pdf
[4] Zwicker E., „Unterteilung des hörbaren Frequenzbereichs in Frequenzgruppen“, Acustica 10, 185 (1960)
[5] http://de.wikipedia.org/wiki/MP3
[6] http://de.wikipedia.org/wiki/Frequenzgruppe
[7] www.compuphase.com/mp3/sta013.htm#MP3FRAMEHDR
[8] http://wiki.hydrogenaud.io/index.php?
[9] http://de.wikipedia.org/wiki/MP3-Encoder
[10] www.dors.de/razorlame/index.php