

Thema: Datenreduktion und Datenkompression

Warum Datenreduktion?

Die Frage läßt sich ganz einfach beantworten: weil man Daten reduzieren will oder muß. Mit geringerer Datenmenge erreicht man eine effektivere Speicherung z.B. für mobile Anwendung, wo Speicher knapp und teuer ist. In schmalen Kanälen wird Musikübertragung in Echtzeit möglich (z.B. im Internet über Modem oder ISDN), und Surroundsignale können über bestehende Stereokanäle (z.B. AES-Leitung) übertragen werden.

Um Daten zu reduzieren bieten sich im Prinzip zwei Verfahren an: die verlustfreie Datenkompression (im Bild z.B. GIF, ZLW, bei der DVD-Audio: MLP) oder die verlustbehaftete Datenreduktion (im Bild z.B. jpeg, bei Audio mp3). Eine verlustfreie Kompression kann meist nur wenig Sparen, erzielt wird eine Rate bis ca. 3:1 die zudem noch vom Inhalt abhängig und deshalb nicht ganz vorhersagbar ist, erhält aber die ursprüngliche Qualität.

Besser ist da die verlustbehaftete Datenreduktion, sie erzielt zwischen 1:12 (gute Qualität) und 1:30 (recht schlecht) beachtliche Einsparungen. Voraussetzung ist aber ein perzeptives Modell, das die Daten nach inhaltlichen Kriterien bearbeitet. Zur Verdeutlichung sei eine kurze Textpassage gezeigt, erst schematisch um 40% reduziert, dann aber so, daß der Sinn leicht zu erkennen ist.

Originaltext:

Musik ist eine universale Sprache, mit der man sich über alle Kulturgrenzen hinweg versteht, heißt es oft. Wie weit es damit her ist, können in diesem Jahr die 30 spanischen und deutschen Nachwuchstalente testen, die sich zwischen 18. und 27. Mai zur "12. Woche der Begegnung junger Musiker" treffen.

Reduktion um 40%:

Musxx ist xxne uxxverxxle Sxxachx, xit dxx man xxch üxxr alxx Kulxxrgrxxzen xxnwex xersxxht, hxxbt ex xft. Wxx weix xs daxxt hex xst, kxxnen xx diexxm Jxxr dix x0 spxxiscxxn unx xeutxxhen xxchwxxhstxxentx xestxx, die xxch zxxschxx 18. xxd 27 xxi zu x2. Woxxe dex xegexxung xxngex xusixxr" trxxfen.

Reduktion um 40% inhaltlich angepaßt:

Musik xxx xxxxx universxxx Sprache, xxx xxx man sich über xxxx Kulturgrenzxx hinweg verstxxx, xxxxx xx xxx. Wie weit es xxxxx xxx xxx, könnxx xx diesem Jahr xxx 30 spanxxxxxx und deutxxxxxx Nachwxxxxtalentx testxx, xxx xxxx zwischen 18. xxx 27. Mai zur xxxxxx xxx Begegnung jungxxx Musikerx treffen.

verlustfreie Komprimierungen:

Die verlustfreie Kompression ist reversibel, d.h. es gehen keine Informationen verloren. Sie ist völlig transparent, es können sogar kritische Computer-Programme komprimiert werden (z.B. im zip-Format), ohne Fehlverhalten zu befürchten. Dazu stehen einige Verfahren zur Verfügung, die aber nicht alle gleich gut für Audio geeignet sind:

- run-length coding (Laufweitenkodierung) - ab 4 gleichen aufeinanderfolgenden Werten kann eine Kennung und die Anzahl der Werte übertragen werden, z.B. wird aus „ABCCCCCCCCDEFGGG“ - „ABC!8DEFGGG“. Bei Audio betrifft dies hauptsächlich die fast immer vorhandenen führenden Nullen, der Wert wird dann ohne sie aber mit einem Skalierungsfaktor übertragen.
- lexikalische Verfahren (bei Bild z.B. GIF, bei Audio wenig sinnvoll) - Bei begrenztem Wertebereich kann man eine (kleinere) Tabelle anlegen, mittels derer die eigentlichen Werte nachgeschlagen werden können.
- delta modulation (Differenzkodierung) - Es wird nur die Differenz zum vorherigen Wert, oder zu einem festgelegten Vorhersagealgorithmus übertragen.

- huffman coding - ein Kodierverfahren, das den häufigsten Werten die kleinsten Zahlen zuweist.

Die Datenreduktion MLP

Die verlustfreie Reduktion der Audiodaten - Meridian Lossless Packaging - ist Bestandteil der DVD-Audio Spezifikationen. Sie bedient sich mehrerer der obigen Methoden, um den Datenstrom zu reduzieren; wichtigste ist die Skalierung: da Audiodaten selten die signifikanten Bits nutzen, werden führende Nullen nicht mit kodiert, die dazugehörigen Skalierungsfaktoren werden blockweise übertragen.

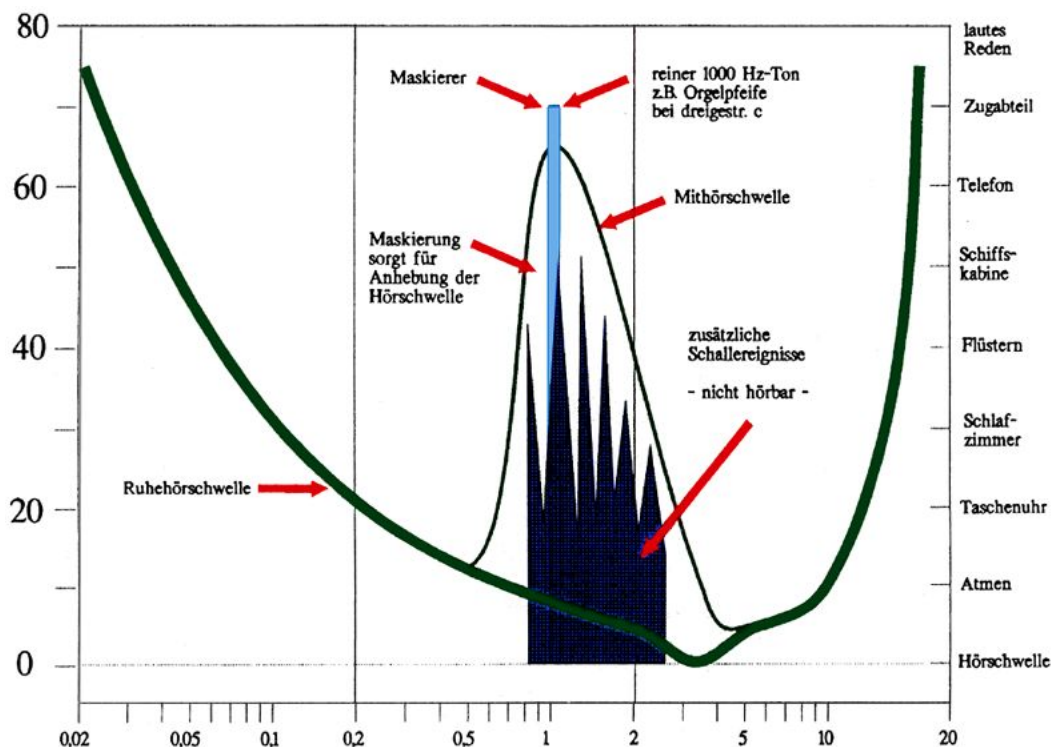
Mit einem intelligenten Vorhersagefilter („prediction filter“) kann vor allem bei periodischem Material (was ja in der Natur der Musik liegt) weiter gespart werden, kodiert wird dann nur noch die Differenz zu der Vorhersage. Bei der Dekodierung muß die gleiche Vorhersage errechnet werden, um mittels der Differenzinformation wieder das ursprüngliche Sample herzustellen. „Prediction“ kann sehr leistungsfähig sein, allerdings müssen die Algorithmen sorgfältig ausgeklügelt und erprobt sein. Es versagt allerdings bei andersartigem Klangmaterial (z.B. stark geräuschhaftem)

Als weitere Methode verwendet man eine Matrizierung für Mehrkanal-Ton. Gleiche Anteile werden gemeinsam kodiert und dann nur noch die Differenzen zwischen den Kanälen, auch wieder mit ausgeklügelten Algorithmen. Mit gleicher Methode aber etwas mehr Daten arbeitet die substream-Technik, es werden mehrere Teildatenströme übertragen. So enthält z.B. substream 0 nur den Stereodownmix, und substream 1 die Differenzen zu den 6 Kanälen. Ein Stereo DVD-Walkman kann dann z.B. nur den stream 0 dekodieren (dies spart Bandbreite und Rechenleistung).

Da die resultierende Datenrate (wie bei allen verlustfreien Kompressionen) nicht verlässlich vorhergesagt werden kann, bedient man sich eines sog. FIFO Puffers um die Datenrate zu stabilisieren. Die Daten werden mit schwankender Datenrate in den Puffer geschrieben, dann aber kontinuierlich ausgelesen. Ähnlich dem Erschütterungsschutz bei einem CD Walkman versagt diese Technik nur bei anhaltend hoher Rate, die aber so gut wie nie vorkommt.

verlustbehaftete Datenreduktion:

Sehr viel mehr läßt sich das Signal reduzieren, wenn man nur die signifikante Anteile überträgt. Dazu muß man die Eigenschaften des Gehörs kennen und Signalanteile identifizieren die wenig zum Höreindruck beitragen. Dies geschieht anhand sog. psychoakustische Modelle (perceptual coding techniques), die auf erforschte psychoakustische Phänomene aufbauen.¹



Die wichtigste Voraussetzung ist die mögliche Maskierung nebeneinander liegender Töne, die im Pegel unterschiedlich sind (die Mithörschwelle). Das Gehör kann die Töne nicht eigenständig wahrnehmen, wenn sie unterhalb dieses Schwellwertes liegen, sie brauchen dann auch nicht mit übertragen zu werden. Ebenso braucht man nicht die volle Dynamik, sondern nur bis zur Mithörschwelle, da ev. Rauschanteile ebenso unhörbar werden, man kann sich eine temporär angepaßte Hörschwelle vorstellen.

Eine weitere Maskierung ist die zeitliche Nachverdeckung: Signale die kurz nach einem lauten Ton erklingen, werden nicht wahrgenommen, wenn sie unter dieser Schwelle liegen. Ebenso wie die Mithörschwelle ist dieses Phänomen von der Frequenzverteilung abhängig, je näher die Frequenzen beieinander liegen, um so mehr kann verdeckt werden.

Das Grundprinzip der Datenreduktion ist:

- Die Kodierung nicht hörbarer Signale ist redundant
- hörbare Signale können ev. mit reduzierter Auflösung kodiert werden
- Die Stärke der Reduktion kann durch sukzessives Weglassen bei verminderter Qualität beliebig festgesetzt werden

Da die Qualität aber nicht ohne weiteres meßtechnisch erfaßt werden kann, erfolgt eine Evaluation mittels Hörtests. Je nach Einsatzzweck (CD-Qualität, Sprache etc.) sind andere Ansprüche zu stellen. Diesbezügliche Forschungen gibt es schon länger z.B. im Fraunhofer IIS (Institut für Integrierte Schaltungen) seit 1993.

Des Weiteren gilt aber:

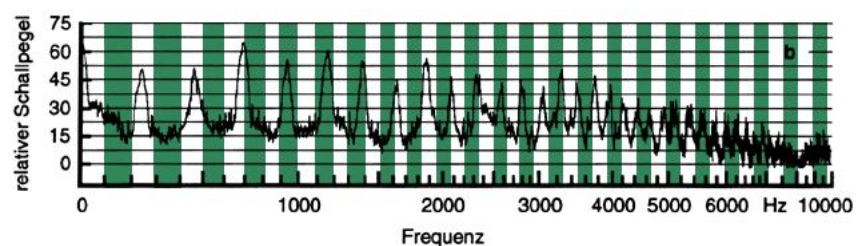
- Das Ausgangssignal nicht mehr identisch mit dem Eingang! Es findet ein (nicht unbedingt hörbarer) Qualitätsverlust statt.
- Die psychoakustische Modelle funktionieren nur bei gesundem Hörsinn! So kann es sein, daß manche Menschen stärker Störungen wahrnehmen, als andere.

Struktur der Kodierer

All coders use the same basic structure. The coding scheme can be described as perceptual noise shaping or perceptual subband / transform coding. The encoder analyzes the spectral components of the audio signal by calculating a filterbank transform and applies a psychoacoustics model to estimate the just noticeable noise-level. In its quantization and coding stage, the encoder allocates data bits in a way to meet both the bitrate and masking requirements. The decoder is much less complex. Its only task is to synthesize an audio signal out of the coded spectral components. (Fraunhofer IIS)

Alle Kodierer arbeiten mehr oder weniger nach dem gleichen Prinzip:

- In einer hybriden Filterbank wird eine Aufteilung in Subbänder vorgenommen (mind. 32, auch flexibel).
- Das psychoakustische Modell bestimmt die Maskierungsschwellen und das max. zulässige Rauschen pro Subband
- Quantisierung und Skalierung ("bit allocation" - wie viele bit und welcher Faktor den Werten zugeordnet wird) erfolgt gemeinsam pro Frame (36 Sample bei mp3)
- Die Option „Joint Stereo“ verringert die notwendige Datenrate durch gemeinsame Kodierung der Stereokanäle.
- Am Schluß folgt noch eine Huffman Kodierung (noiseless coding)



Bei der Dekodierung ist eine Fehlerkorrektur nötig, diese kann je nach Wichtigkeit der Daten aber unterschiedlich ausfallen (die Skalierungsfaktoren sind z.B. anfälliger, da ein Fehler stärker hörbar wird).

Einige Datenraten:

	Datenrate	Speicherbedarf	Übertragung für 1 min
CD 44,1kHz 16 bit	= 1411,2 kbit/s	= ca. 10 MB / min	21 min ISDN
MPEG 1/Layer 2	= 256 kbit/s	= ca. 1,9 MB / min	4 min
mp3 / aac	= 128 kbit/s	= ca. 960 kB / min	2 min
mp3 / aac	= 64 kbit/s	= ca. 480 kB / min	1 min ISDN (Echtzeit)
mp3 / aac	= 48 kbit/s	= ca. 320 kB / min	45 sec ISDN

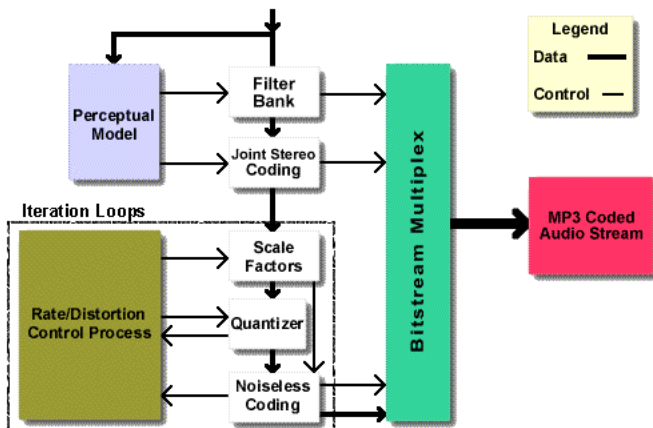
MPEG 1

MPEG steht für „Motion Picture Experts Group“, eine Expertengruppe die sich voran-
gig Film- und Videoübertragung widmet. Für die Tonkodierung gibt es die Layer 1 - 3
(aufwärtskompatibel):

Reduktion Layer 1 - 1:4 (verwendet bei DCC - heute obsolet)

Reduktion Layer 2 - 1:8 (192 - 256 kbit/s - DVD, DAB, DIGAS)

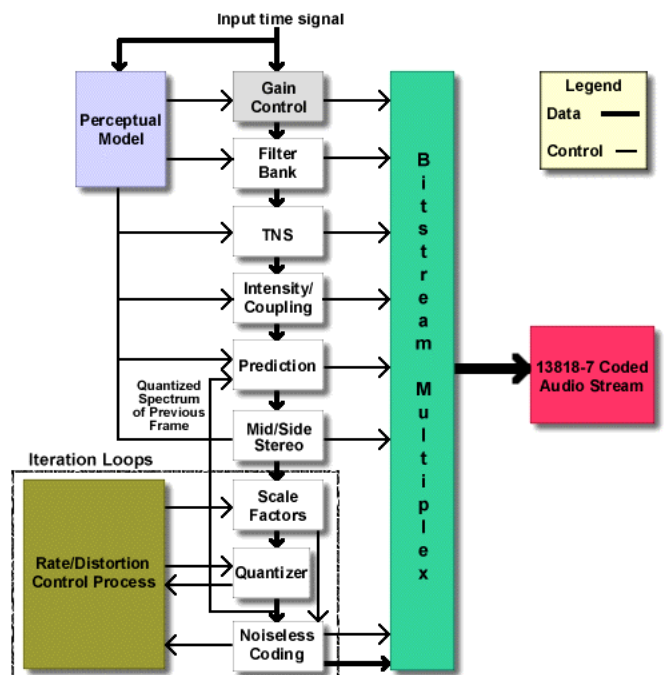
Reduktion Layer 3 - 1:12 (bekannt als mp3 - Internet)



MPEG 1 Layer 3 (**mp3**) ist eine Weiterentwicklung des Layer 2, die auf niedrige Datenraten getrimmt ist. Mp3 verwendet angepasste Filter, das bedeutet gleichzeitige eine Dezimierung und Filterung der Abtastwerte. Durch die erstmal mögliche Übertragung von Musik in CD-naher Qualität im Internet, hat sich mp3 zum Quasistandard etabliert.

AAC - „Advanced Audio Coding“ ist ein raffinierteres Verfahren, mit dem Ziel bessere Qualität bei niedrigen Datenraten zu erzielen. Es wurde 1997 für MPEG 2 MCA (Multichannel Audio) entwickelt und ist Bestandteil des neuen MPEG 4 Standard.

Bei der Kodierung hat man eine Vorhersage („prediction“) und eine Pausen-Erkennung (um nicht Rauschen zu kodieren), Noise-Shaping (das Quantisierungsrauschen wird in maskierte Frequenzbereiche transformiert) sowie einige weitere Optimierungen hinzugefügt. Außerdem erfolgt eine Analyse des Signals aufgrund bekannter Probleme (s.u.), so daß bereits mit



128kbit/s CD-Qualität erreicht wird. Die Samplingfrequenz kann zw. 8 - 96 kHz liegen bei 1 - 48 Kanäle (die Bandbreite steigt dann mit der Wurzel der Kanäle also z.B. 128kbit/s $\sqrt{5.1} = 290\text{kbit/s}$).

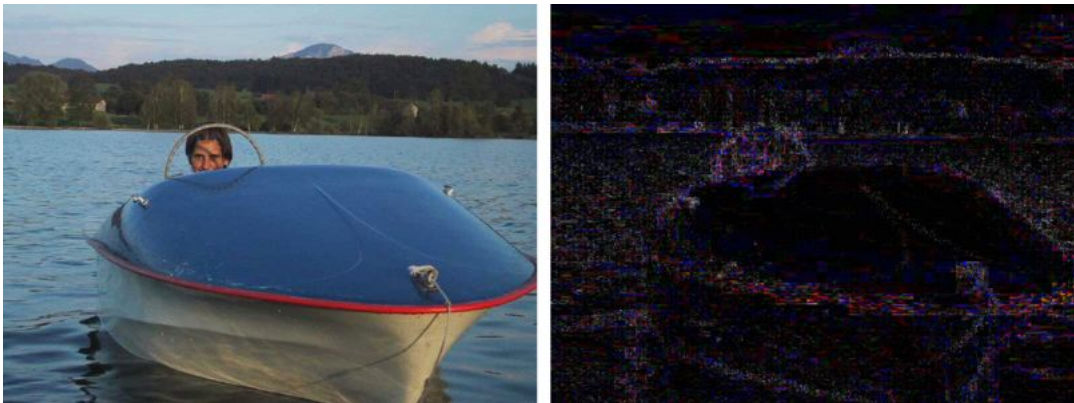
Zu den Unterschieden:

Der AAC setzt ausschließlich eine MDCT („Modified Discrete Cosine Transformation“) mit zwei verschiedenen Blocklängen ein. Bei "long blocks" werden 2048 Samples, bei "short blocks" 256 Samples verarbeitet, wodurch gegenüber MP3 sowohl eine höhere Frequenz- als auch Zeitauflösung erreicht wird. Der "Joint Stereo"-Mode ist gegenüber MP3 flexibler, da er unabhängig für Frequenzbereiche geschaltet werden kann und nicht, wie bei MP3, für den gesamten Frequenzbereich. Der AAC-Encoder besitzt außerdem einige Tools, wie z.B. LTP (long term prediction), PNS (perceptual noise substitution) und TNS (temporal noise shaping), die die Bitrate weiter absenken können. (Antje Barth)

Vergleichen Sie mal mit 64kbit/s kodierte Musik in mp3 und AAC, die Verbesserungen sind enorm (ein kostenloses Tool dafür ist z.B. „iTunes“ von apple, auch als Windowsversion erhältlich²).

Artefakte

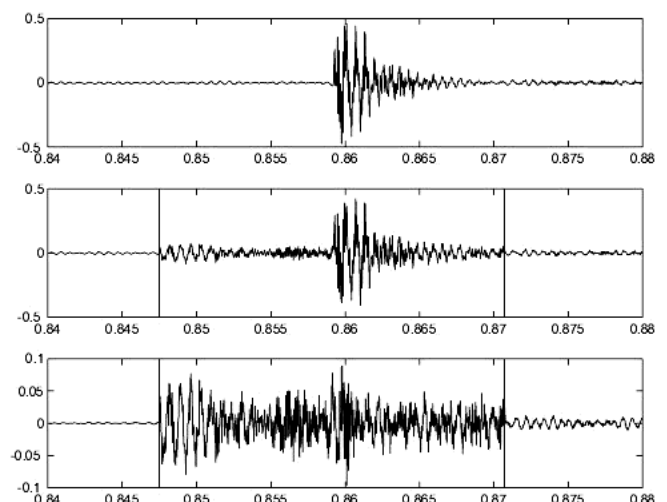
Da die Datenreduktion Signalanteile wegläßt, kann man durch Differenzbildung die resultierenden Störungen sichtbar machen, wie an diesem jpeg-Bild:



Solche Störungen nennt man „Artefakte“, sie sind prinzipiell bei jeder verlustbehafteten Kodierung vorhanden, wenn auch meist nicht wahrnehmbar.

Im Zuge der Verbreitung von mp3 und AAC hat die AES eine Untersuchung der Artefakte angestrengt³. Die wichtigsten Artefakte seien hier erwähnt:

- Antialiasing - es entsteht bei jeder A/D-Wandlung und stört nur wenn nicht ausreichend gefiltert wurde (dann entstehen Spiegelfrequenzen)
- Birdies (Zwitschern) - entsteht durch die Anpassung der Skalierung in höheren Frequenzbereichen, ist bei AAC weitgehend unterdrückt
- Vorechos - durch die blockweise Kodierung werden die Signale verschliffen (ist vor allem bei festen Blockgrößen wie mp3 ein Problem) Bsp. Kastagnetten auf der AES-CD⁴:
- Die Stereobreite kann durch Joint-Stereo Kodierung eingeengt werden, eher ein geringfügiges Problem (hörbar vor allem bei Beifall)



Entwicklungen:

Mit MPEG4, einem objektorientiertem Containerformat, hat auch die leistungsfähige AAC-Kodierung Einzug gefunden. Das Hauptproblem ist dabei noch die relativ große Latenzzeit von ca. 250 ms (Zeit, mit der das Signal durch die Bearbeitung verzögert wird), die eine Echtzeitanwendung erschwert. Deshalb wurde das AAC-LD (low-delay audio coding) entwickelt, daß vorzugsweise <30 ms Latenz verursacht (z.B. für Telefon, live Anwendungen etc.).

Eine weitere Forschung geht in Richtung skalierbarer Audio-Streams, wie sie z.B. schon bei dts 24/96 genutzt wird (siehe VL Surround). Dabei wird das Signal in mehrere Teilströme aufgeteilt, die sich auch ergänzen können. So kann jeder Empfänger nur den Teilstrom lesen, der für ihn angebracht ist.

Weitere Formate seien nur kurz erwähnt:

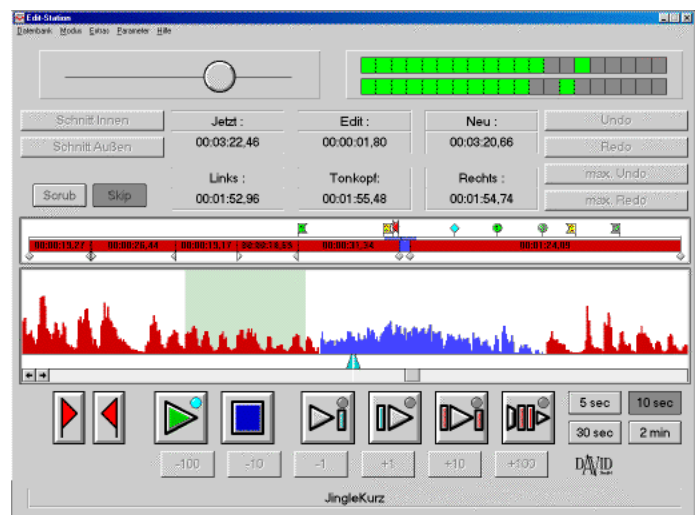
- Real Audio - ein Internetformat hauptsächlich fürs Streaming (Echtzeitübertragung), weniger Datenreduziert aber Dropout-sicher durch Interleaving über 4 s.
- WMA (Windows Media Audio) ähnlich wie mp3
- Ogg Vorbis (ein open source Coder - lizenzfreie Alternative zu mp3), unterstützt 44,1 oder 48 kHz und 16 bis 128 kbit/s - teilweise etwas besser als mp3

Schnitt/Bearbeitung:

Eine Schnitt oder Bearbeitung direkt im kodierten Signal ist nicht möglich, da bei einer Veränderung der Struktur die psychoakustischen Voraussetzungen nicht mehr gelten. Deshalb muß in der Regel das Signal vor der Bearbeitung gewandelt werden und hinterher wieder zurück. Eine solche Kaskadierung (mehrfache Wandlung) verschlechtert bei verlustbehafteten Verfahren aber das Signal, vor allem, wenn der Frequenzgang verändert wird, können bisher verdeckte Artefakte zum Vorschein kommen!

Die Alternative ist ein Schnitt innerhalb der Framegrenzen des kodierten Signals, dies reduziert jedoch die Schnittgenauigkeit, reicht aber für aktuelle Wort-Produktionen aus. Das im Rundfunk verwendete DIGAS⁵-System geht diesen Weg. Es verwendet MUSICAM (MPEG1 / Layer2) und ist vor allem wegen der handhabbaren Datenraten und der flexiblen Speicherung, Sendung und Archivierung sehr verbreitet.

Das Schnittsystem Sequoia⁶ kann DIGAS-Material direkt einbinden und in Echtzeit wandeln, d.h hier kann datenreduziertes mit nicht reduzierten Material kombiniert und z.B. auf CD gebrannt werden.

**Übertragung / Sendung**

Als Anwendung wurden Kodierer entwickelt, die datenreduziertes Audio über eine ISDN-Verbindung übertragen können (zwei B-Kanäle = 128 kbit/s). ISDN ist preiswert und fast überall verfügbar, und daher für den aktuellen Einsatz ideal. Der Verteran und Motor dieser Entwicklung war das „Musiktaxi“ von Dialog 4, inzwischen abgelöst von moderneren Produkten wie „Centauri“ von Mayah⁷.

Links:

zu mp3:

<http://substream.org/><http://www.tecchannel.de/multimedia/57/index.html><http://www.stud.uni-hannover.de/user/73884/audiocoder.html>

Fraunhofer IIS:

<http://www.iis.fraunhofer.de/amm/techinf/index.html>

Ogg Vorbis:

<http://www.xiph.org/ogg/vorbis/>

MLP:

http://www.hometheaterhifi.com/volume_8_4/dvd-benchmark-part-6-dvd-audio-11-2001.html

Quellen:

¹ Zwicker, E. und Feldtkeller, R.: Das Ohr als Nachrichtenempfänger. S. Hirzel Verlag, Stuttgart, 1967.

² <http://www.apple.com/de/itunes/>

³ AES CD (mit CD-ROM): Perceptual Audio Coders: „What to listen for“ ©2001 AES

⁴ dito

⁵ <http://www.david-gmbh.de/index2.htm>

⁶ http://www.sekd.de/product_main/sequoia_info.htm

⁷ <http://www.mayah.com/> und <http://www.broadcastisdn.net/dialog4.htm>