

5 Datenkompression

Autor: Gernot Meyer-Schwarzenberger

5.1 Audiodatenkompression

5.1.1 Grundsätzliches

Den Kern der Datenreduktion von digitalen Audio- und Videosignalen bilden hochkomplexe mathematische Verfahren, die auf speziellen Chips (ICs) in den entsprechenden Geräten (Encodern und Decodern) implementiert sind. Immer mehr Personal in den Funkhäusern und Studios kommt in die Lage, sich zumindest ansatzweise mit diesen Verfahren befassen zu müssen, sei es z. B. um sich über die notwendige Reduktion der Daten eines Korrespondentenbeitrags über ISDN klar zu werden, um die Folgen einer Kaskadierung von unterschiedlichen Aufzeichnungsformaten im Videobereich besser einschätzen oder um ganz konkrete Einstellungen von Parametern vornehmen zu können (durch die bedienende Person, den so genannten Compressionist), wie z. B. an einem MPEG-2-Encoder die Quantisierung der Koeffizienten, die GOP-Struktur im Hinblick auf die geforderte Datenrate u. Ä. m.

Ausgangspunkt der erwähnten grundlegenden Verfahren ist *erstens* die „DFT“ („Diskrete Fouriertransformation“), die im Audiobereich als „FFT“ („Fast Fourier Transformation“) und im Videobereich als „DCT“ („Diskrete Cosinustransformation“) in Form superschneller Rechenverfahren („Algorithmen“) verifiziert ist. Bei Audio werden z. B. mit der „1024-Punkte-FFT“ während eines Zeitfensters von knapp 24 ms 1024 „Samples“ (Abtastwerte) in 1024 „Koeffizienten“ (Spektrallinien bzw. „Frequenzkomponenten“) umgerechnet. Bei Video werden z. B. in einem Vollbild die Signalwerte von $8 \cdot 8$ Pixeln, die ein „Ortsfenster“ darstellen, mit der DCT in $8 \cdot 8$ Koeffizienten (zweidimensionale „Ortsfrequenzkomponenten“) umgerechnet. Eine solche Umrechnung ist dann jeweils die Voraussetzung für eine erhebliche Datenreduktion („Datenkompression“) bei Tönen bzw. Einzelbildern. (Näheres zur DFT findet man – evtl. auch als Ergänzung zu den folgenden Ausführungen – unter „DVB-T“ im Abschnitt „Digitales Fernsehen.“) *Zweitens* sind als weitere wesentliche Elemente der Datenreduktion z. B. bei Audio (MPEG-1 und -2) noch die „Teilbandcodierung“ und bei MPEG-Video die „DPCM“ (Differenz-PCM) mit einem „Prädiktor“ als zentralem Bestandteil zu nennen.

Bild 5.1 dient der Veranschaulichung einer FFT, und zwar speziell der „1024-Punkte-FFT“: Jeweils 1024 aufeinander folgende Samples werden in „Koeffizienten“ umgerechnet, die die 1024 Sinustöne dieses Tonabschnitts verkörpern („Tonfrequenzanalyse“ mittels der FFT) und als senkrechte Spektrallinien dargestellt sind. Die herausragenden Töne unter ihnen sind die „Maskierer“, die nun zur Errechnung des „Verdeckungseffektes“ dienen. Die errechneten Verdeckungskurven sind durch die zeldachartig geschwungenen Linien angedeutet. Die eingezeichnete wannenförmige Kurve darunter ist die „Ruhehörschwelle“, unterhalb derer das menschliche Ohr nichts

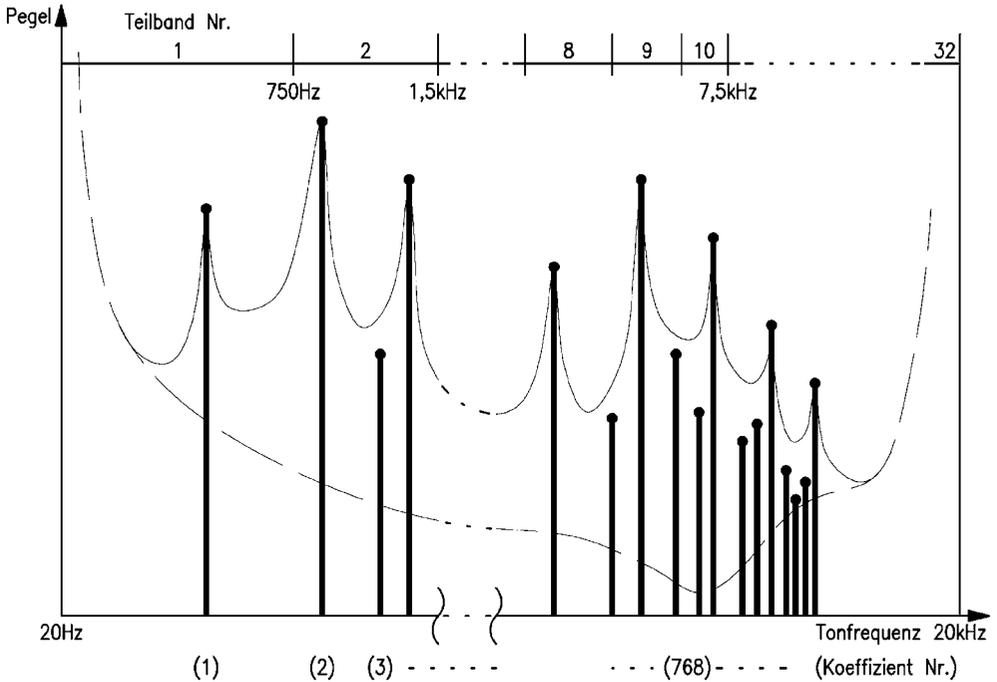


Bild 5.1: Schematische Darstellung einer 1024-Punkte-FFT

mehr wahrnimmt. Ebenfalls angedeutet sind die 32 gleich breiten „Teilbänder“, die durch einen anderen Filterungsprozess (s. u.) parallel zur FFT erzeugt werden, auf die die eigentliche Datenreduktion angewendet wird, und die hier nur wegen der logarithmischen Einteilung der Frequenzachse immer schmäler zu werden scheinen.

Was jetzt immer noch reichlich unanschaulich erscheinen mag, wird trotzdem leicht verständlich und in unterhaltsamer Weise auf einer von der srt Schule für Rundfunktechnik Nürnberg entwickelten, in dieser Form konkurrenzlosen **CD-ROM** dargestellt (als interaktives CBT – Computer Based Training): „**MPEG, JPEG & Co., Wege der Datenreduktion**“ (Autor: Gernot Meyer-Schwarzenberger), Hüthig-Verlag 1997, ISBN 3-7785-2632-4.

Die anschließenden Darstellungen basieren zwar auch auf der genannten Quelle, es wurde aber Wert darauf gelegt, dass der Leser/die Leserin ihnen auch ohne detaillierte Vorkenntnisse folgen kann.

Die *Gründe* für die Reduktion von Daten („die Datenkompression“) liegen auf der Hand. Man will (Kosten) einsparen: bei der Speicherkapazität – auf Servern, Discs etc. – und bei der Übertragungskapazität – auf Satelliten, Kabeln, terrestrischen Sendern und Netzen. Ein Beispiel aus dem Audiobereich: Das Herunterladen eines vierminütigen Musiktitels von 42 MB unkomprimierter Größe über eine ISDN-Leitung würde knapp 90 Minuten dauern. Beim Einsatz eines MP3-Encoders dauert das Herunter-

laden nur noch ca. 7 Minuten; über *zwei* ISDN-Leitungen kann der Titel sogar in Echtzeit übertragen und dabei abgehört werden.

5.1.2 MPEG-1 und Layer

Ausgangspunkt der Datenreduktion sind die nach dem „AES/EBU“-Standard pulscodierten („PCM“-)Tonsignale, die digitale Studioqualität erreichen, in ihrer hochwertigsten Form mit 48 kHz Abtastfrequenz und 16, 20, 22 oder 24 bit Auflösung (s. Ausbildungshandbuch 2).

Der Standard „MPEG-1“ (MPEG = Moving Pictures Experts Group = die Experten-Gruppe für Bewegtbilder, die den Standard entwickelt hat) erlaubt Abtastfrequenzen von 48, 44,1 und 32 kHz. Im codierten Audiodatenstrom sind vier Übertragungsarten („Audio Modes“) spezifiziert: Einkanal („Single Channel Mode“), Stereo („Stereo Mode“), Zweikanal („Dual Channel Mode“) und „Joint Stereo Mode“ (s. gleich). Die erlaubten Bitraten im Single Channel Mode sind 32, 48, 56, 64, 80, 96, 112, 128, 160 und 192 kbit/s, in den anderen Modi das Doppelte.

MPEG-1 unterscheidet hinsichtlich der Komplexität und Effizienz der Reduktionsalgorithmen zwischen Layer I, II und III, die aufwärtskompatibel sind (d. h. z. B., Layer III und II können I „abspielen“, aber nicht umgekehrt) und die auch Bestandteile von MPEG-Video-Datenströmen sein können.

Layer I

Mit diesem Verfahren erreicht man einen Kompressionsfaktor von etwa 4 : 1 (192 kbit/s pro Monokanal, theoretische Verzögerung > 19 ms). Es entspricht einem vereinfachten MUSICAM-Verfahren (s. u.) und ist weitgehend transparent. Ein Anwendungsgebiet ist der Programmaustausch.

Layer II

Dabei handelt es sich um das eigentliche MUSICAM-Verfahren. Der Kompressionsfaktor beträgt etwa 8 : 1 (128 kbit/s pro Monokanal, theoretische Verzögerung > 35 ms), und das Verfahren ist nur eingeschränkt transparent. Anwendungen liegen z. B. in der Distribution.

Layer III

Auf Layer II wurde ein Teil des so genannten ASPEC-Verfahrens (s. später) aufgesetzt, womit ein Kompressionsfaktor von etwa 12 : 1 bis 14 : 1 (64 kbit/s pro Monokanal, theoretische Verzögerung > 59 ms) erreicht wird. (Man kann sagen, dass Layer III die Vorteile von MUSICAM und ASPEC vereint.) Das Verfahren ist nicht transparent, aber die annähernde CD-Qualität z. B. über ISDN übertragbar. Das Über-spiel- und Austauschformat „MP3“ bedeutet übrigens nichts anderes als die Codierung in MPEG-1 Layer III.

Joint Stereo Codierung

In allen drei Layern lässt sich für eine weitergehende Datenreduktion die so genannte Joint Stereo Codierung einsetzen, die die Gesetzmäßigkeiten der Intensitätsstereofonie, genauer die ihr innewohnenden Redundanzen und Irrelevanzen, zur gemeinsamen (engl. „joint“) und damit datensparenden Codierung der beiden Stereokanäle nutzt. Bei ungenügender Korrelation der Kanäle muss allerdings wieder die „getrennte“ Codierung angewendet werden, weil sich andernfalls bei der Wiedergabe Auslöschungen aufgrund von Gegenphasigkeiten ergeben können.

5.1.3 MPEG-2 Audio

„MPEG-2“ Audio ist eine zu den drei Layern von MPEG-1 Audio kompatible Erweiterung („Extension“). Es erlaubt halbierte Abtastraten, also außer 48, 44,1 und 32 auch 24, 22,05 und 16 kHz (mit Bandbreiten von 7,5 bzw. 10,5 bzw. 11,5 kHz), sowie zusätzlich zur Mono- und Stereo- noch die Mehrkanal- („Multichannel“- „MC“-) Übertragung. Während für Stereo die typischen Datenraten zwischen 128 und 256 kbit/s liegen, erfordert ein Mehrkanalton bei Kinofilmen („Movie Soundtrack“) Datenraten zwischen 320 und 640 kbit/s. MPEG-2 definiert hierfür eine Extension auf fünf Kanäle mit voller Bandbreite (20 Hz bis 20 kHz) zuzüglich eines tieffrequenten Effektkanals („LFE Channel“ = „Low Frequency Enhancement Channel“ zwischen etwa 30 und 120 Hz, auch „Subwoofer“ genannt). Deshalb spricht man von „5.1-Multichannel“-Übertragung, wobei 5.1 = 5+1 bedeutet bzw. „1“ (= „Komma eins“ in deutscher Schreibweise) auch auf die Tatsache hinweist, dass der LFE-Kanal nur etwa 1/10 der Datenmenge eines „vollen“ Kanals beansprucht. Die 5-Kanal-Übertragung entspricht der bekannten „3/2-Stereofonie“ (drei Kanäle „vorn“ und zwei Surroundkanäle „hinten“). Darüber hinaus kann man bei MPEG-2/Layer II noch sieben Sprachkanäle für Mehrsprachigkeit mit wählbaren Abtastraten (für Kommentare oder Teleconferen-

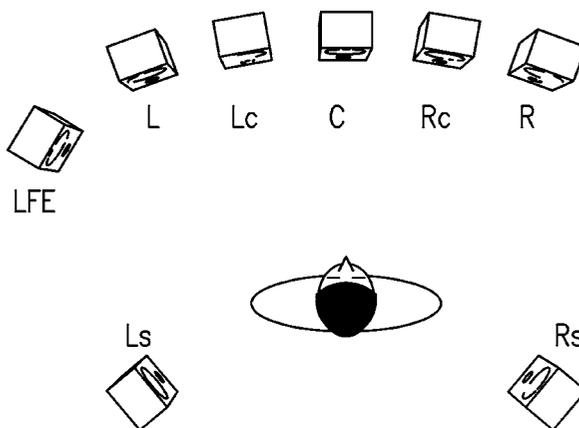


Bild 5.2: Die Anordnung von 7+1 Lautsprecherboxen bei 7.1-Multichannel-Übertragung. Es bedeuten: L = Links (Left), R = Rechts (Right), C = c = Center bzw. central, s = surround (für Rundumklangerzeugung), LFE = Low Frequency Enhancement.