

Psychoakustisch komprimiert

Vor 30 Jahren wurde eine MPEG-Kodierung – besser bekannt als MP3 – veröffentlicht.

Jörn Donges

Zwischen 1988 und 1993 arbeitete das internationale Gremium MPEG (Moving Pictures Expert Group) daran, die digitale Audio- und Videoübertragung zu standardisieren. Als Arbeitsgruppe der länderübergreifenden Normen-Organisation ISO kamen ihre Mitglieder aus über 200 Unternehmen und Forschungseinrichtungen in mehr als 20 Ländern. Unter zahlreichen Vorschlägen wählte sie für den Audibereich drei Kodierungsverfahren aus. Das effizienteste hieß offiziell „ISO MPEG 1 Layer III“ – und wurde unter dem Namen MP3 so bekannt, dass sich diese Abkürzung heute als Synonym für digitale Musikformate etabliert hat. Das Verfahren entwickelt haben Forschende des Fraunhofer Instituts für integrierte Schaltungen (IIS) und der Uni Erlangen-Nürnberg. Seine Veröffentlichung erfolgte im Dezember 1991.

Der MP3-Standard erlaubt es, Musikdateien auf bis zu ein Zehntel ihrer Größe zu komprimieren, ohne die Klangqualität nennenswert zu verringern. Im jungen und wachsenden Internet der 1990er-Jahre fiel das auf fruchtbaren Boden. Den Inhalt einer CD unkomprimiert mit einem analogen 56k-Modem zu übertragen, hätte mehr als 30 Stunden gedauert – unbezahlbar bei einer Abrechnung im Minutentakt. Die neue Technologie verkürzte diese Zeit auf einen Nachmittag; einzelne Stücke standen in wenigen Minuten zur Verfügung. So begann der weltweite Austausch digitaler Musik.

Die enorme Kompressionsrate bringt durchaus Verluste mit sich. Die größte Innovation des Standards besteht darin, genau die Teile des ursprünglichen Signals herauszufiltern, die das menschliche Gehör ohnehin nicht wahrnimmt, und so die Klangqualität weitgehend zu erhalten. Das Verfahren des „Perceptual Audio Coding“ beruht auf Ergebnissen der Psychoakustik, welche die subjektive Wahrnehmung von Schallereignissen untersucht. Bereits bei einem konstanten Sinuston hängt die empfundene Lautstärke nicht nur von der bestimmenden physikalischen Größe – dem Schalldruck – ab, sondern auch

ändert sich durch andere vorliegende Töne. Ein Signal macht das Gehörssystem bei den Nachbarfrequenzen taub, wenn die Töne hier leiser sind. Diese Anteile unterhalb der Wahrnehmungsschwelle entfallen im komprimierten Signal.

Der Grund dafür liegt im physikalischen Aufbau des Gehörs: Ein Ton verursacht charakteristische Schwingungsmuster der Basiliarmembran. Diese dünne Haut befindet sich in der Gehörschnecke; verschiedene Frequenzen lenken sie an unterschiedlichen Stellen aus. Ein etwas leiserer Ton mit einer geringfügig anderen Frequenz regt die schwingende Membran nicht weiter an und bleibt unhörbar. Je lauter ein Ton ausfällt, desto breitbandiger hebt er die Hörschwelle an (Abb. 1). Der Effekt sorgt im Alltag zum Beispiel dafür, das in einer grö-



Adobe Stock, julio77; Csák István / Pixabay, herbert2512

von der Empfindlichkeit des Gehörs bei unterschiedlichen Frequenzen. Generell nehmen wir Frequenzen von etwa 20 Hz bis 20 kHz wahr, wobei die Wahrnehmungsschwelle bei hohen und tiefen Frequenzen stark ansteigt (Abb. 1). Zudem ist die Hörschwelle keine statische Größe, sondern ver-

ßeren Gruppe die Gespräche unter den einzelnen Mitgliedern stetig an Lautstärke zunehmen.

Darüber hinaus treten zeitliche Maskierungseffekte auf: Nach dem Abklingen eines Signals bleibt die Hörschwelle für weitere 200 Millisekunden erhöht. Daran sind die inneren Haarsinneszellen in der Gehörschnecke schuld. Sie übermitteln den Schwingungszustand der Basiliarmembran an die Nervenzellen und müssen sich neu ausrichten, bevor sie neue Reize empfangen und weitergeben können.

Geschicktes Aussortieren

Für den MP3-Encoder (**Abb. 2**) gilt es zunächst, aus dem zeitlichen Verlauf des Signals das Frequenzspektrum zu bestimmen. Beides lässt sich eindeutig ineinander überführen, solange die Bandbreite ausreichend groß gewählt ist. Eine Filterbank zerlegt das Spektrum und teilt es in schmale Frequenzbänder auf. Außerdem trifft das Signal auf einen psychoakustischen Modellierer, das Herzstück des MP3-Encoders. Dieser berechnet den zeit- und frequenzabhängigen Verlauf der Wahrnehmungsschwelle für das gegebene Spektrum. Dazu sucht er nach den oben beschriebenen Maskierungseffekten: Die lauteste Frequenz dominiert das Signal. Fallen benachbarte Bereiche unter die erhöhte Wahrnehmungsschwelle, werden sie im weiteren Verlauf nicht beachtet. Außerdem passt ein Algorithmus die Kodierungs-Bitrate an die erhöhte Wahrnehmungsschwelle an.

Eine geringere Bitrate reduziert die Zahl der diskreten Werte, um die Signalform digital abzubilden. Das Auf- und Abrunden zum nächsten diskreten Wert führt zu einem Quantisierungsfehler. Dieser entspricht der Differenz aus Signalwert und kodiertem Wert und ist statistisch verteilt. Beim Abspielen ergibt sich ein Rauschen, das die Klangqualität negativ beeinflusst und so gering wie möglich ausfallen sollte.

Bei der Kompression ist es demnach entscheidend, die Bitrate nur so weit zu verringern, dass der Rauschanteil des Signals die angehobene Wahrnehmungsschwelle nicht übersteigt. Dann lässt sich das unvermeid-

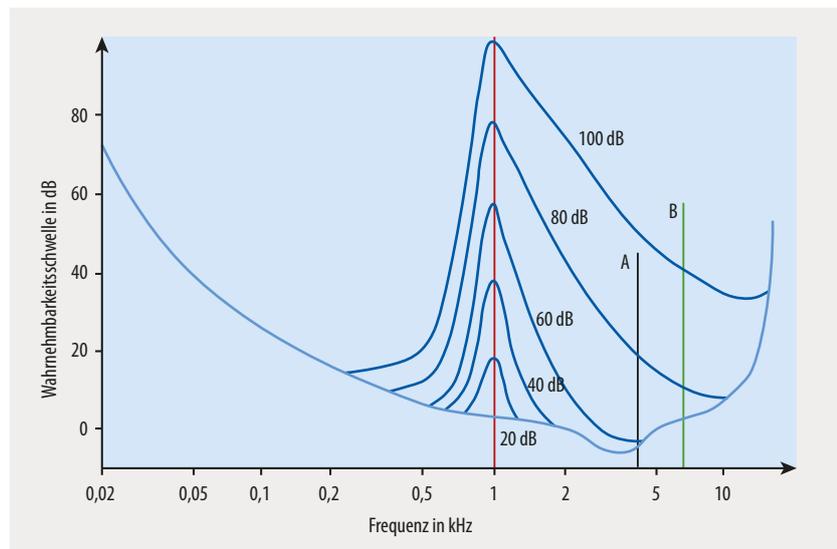


Abb. 1 Störsignale mit 1 kHz Frequenz (rot) heben die Hörschwelle (hellblau) je nach ihrer Lautstärke zwischen 20 und 100 dB unterschiedlich stark an (blau). Bei 100 dB verdeckt das Störsignal den Ton A (schwarz), während der höhere und lautere Ton B (grün) hörbar bleibt.

liche Quantisierungsrauschen im Ausgangssignal trotz starker Datenreduktion nicht wahrnehmen. Anschaulich bedeutet das beispielsweise, dass ein lauter dominanter Einzelton zu einer geringeren Auflösung führt.

Die MP3-Kodierung ist verlustbehaftet: Das ursprüngliche Signal lässt sich mit dem Ausgangssignal nicht vollständig rekonstruieren. Die meisten Menschen bemerken diesen Verlust aber nicht. Nach der Kodierung erfolgt eine weitere Komprimierung, die allerdings verlustfrei abläuft. Der MP3-Standard setzt hier einen sogenannten Huffman-Code ein. In der Bitfolge des Signals treten einzelne Teilsequenzen immer wieder auf. Ein Huffman-Code prüft, wie oft jede Sequenz vorkommt. Je seltener dies der Fall ist, desto länger ist die Bitfolge, die ihr der Code zuordnet – und umgekehrt. Ähnlich dem Morse-Alphabet handelt es sich um präfixfreie Folgen, die Verwechslungen auch ohne Trennzeichen zwischen den einzelnen Sequenzen ausschließen. Im Mittel reduziert dieses Vorgehen die Länge der Bitfolge nochmals deutlich, ohne Information zu verlieren.

Streaming-Dienste haben mittlerweile die klassischen Datenträger wie CDs verdrängt. Wenn die Musik live aus der Cloud kommt, braucht es Verfahren, die bei gleicher Klangqualität noch stärker komprimieren als der MP3-Standard. Zu den bekannteren zählen „Advanced Audio Codec“ und „Ogg Vorbis“. Beim ersteren passt sich beispielsweise das Zeitfenster für die Maskierung an die Signalform an und spart so nochmals an der Datenmenge. Auch wenn der MP3-Standard dadurch langsam an Verbreitung verliert, basieren auch die neuen Methoden noch immer auf seiner Schlüsseltechnologie: der gehörangepassten Audiokodierung, welche die nicht wahrnehmbaren Signalanteile entfernt.

Der Autor

Dr. Jörn Donges, Technischer Redakteur, donges-doku.de

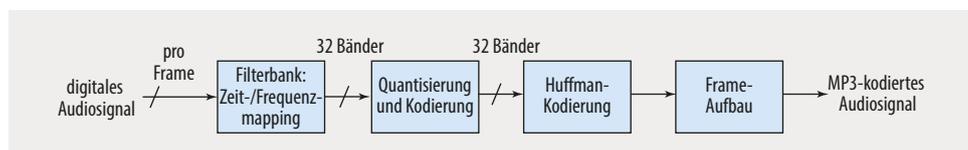


Abb. 2 Das Blockdiagramm zeigt vereinfacht die einzelnen Arbeitsschritte eines MP3-Encoders.