MP3 Workshop

Vortrag: *Andreas Bogk*

Bericht: *Chris Vogel* <c.vogel@link-goe.de>

Der Workshop führte in die Technik der Komprimierung mittels MPEG ein und mündete in eine Diskussion um die Auswirkungen der Technologie auf die zukünftige Entwicklung der Vermarktung von Musik im und außerhalb des Internet.

Die Technik

Musik besteht aus Schallwellen, die mittels eines Mikrofons in elektrische Schwingungen umgewandelt werden können. Um diese analogen Signale, die auf einem Diagramm der Signalstärke, abgetragen gegen die Zeit, komplizierte Wellen ergeben, digital speichern zu können, werden die analogen Axen des Diagramms in kleine Abstände unterteilt (quantisiert).

Die Unterteilung der Zeitachse ergibt eine zeitliche Auflösung, die mit der Sampling-Rate angegeben wird. Die Sampling-Rate gibt in Samples pro Sekunde an, wie häufig ein Meßwert (der Signalpegel) pro Sekunde erfaßt wird.

Die Quantisierung der Amplitude (Signalstärke) ergibt eine maximale Anzahl von Werten, in die die zu digitalisierenden Signalpegel einsortiert werden müssen. Diese Anzahl der möglichen Werte ergibt die Audioauflösung - meist angegeben in Bit.

Auf einer handelsüblichen Musik-CD wird z.B. eine Samplingrate von 44kHz (also 44000 Meßwerte pro Sekunde) und eine Audioauflösung von 16 Bit (65536 mögliche Signalpegel) verwendet.

Speist man die auf diesem Weg erhaltenen Daten in ein normales Komprimierungsprogramm (z.B. ARJ oder ZIP) ein, so stellt man fest, daß die Komprimierung nur sehr gering ist. Solche Komprimierungsverfahren versuchen Regelmäßigkeiten in den zu komprimierenden Daten zu finden, die in digitalisierten Audiodaten kaum vorkommen.

Die erste Idee könnte nun sein, einfach die Audioauflösung zu verringern und somit Daten einzusparen. Leider führt dies zu einem Qualitätsverlust: Auf je weniger mögliche Werte die analogen Daten abgebildet werden, um so größer werden die dabei auftretenden Rundungsfehler. Diese Rundungsfehler erhöhen maßgeblich den Rauschanteil. Für eine möglichst verlustfreie Komprimierung von Audio-Daten ist dieses Verfahren damit unbefriedigend.

Biologen gehen davon aus, daß vom menschlichen Ohr zum Gehirn ein Datenstrom von ca. 2000-3000 bit/s fließt. Die Schätzung geht zurück auf die Annahme, daß Informationen mit ca. 60-100km/h in den Nerven übertragen werden und der Nervenstrang vom Ohr zum Gehirn keinen größere Datenmenge transportieren kann. Eine CD beinhaltet einen Datenstrom von ca. 1.000.000 bit/s. Im Gehirn kommen von den Informationen, die auf einer Audio-CD gespeichert sind, also nur ca. 0,25% an.

Die Psychoakustik beschäftigt sich mit der Frage, welcher Anteil der Informationen, die als Schallwellen auf unser Ohr treffen, tatsächlich im Gehirn ankommen. Als Antwort auf diese Frage wurde bisher keine einheitliche Formel gefunden. Verschiedene Forschungsinstitutionen haben in sehr vielen Versuchsreihen mit Testhörern herausgefunden, was in Audiosignalen weggelassen werden kann, ohne daß ein Unterschied für die Testpersonen hörbar war.

Das Ohr ist träge, und diese Trägheit sorgt dafür, daß leise Töne nach einem lauten Ton nur sehr schlecht oder gar nicht wahrgenommen werden - ebenso wie vorhergehende leise Töne für einen kürzeren Zeitraum überschattet werden. Ähnliches gilt für einen intensiven Ton auf einer Frequenz, der Töne auf unmittelbaren Nachbarfrequenzen überdeckt. Hohe Töne werden eher wahrgenommen als tiefe Töne.

In Hinblick auf diese Erkenntnisse wurden empirische Daten ausgewertet und genaue mathematische Regeln erstellt, die beschreiben, welche Informationen im Audiosignal eingespart werden können, ohne daß die Qualität für den Hörer merklich sinkt.

Technisch wird hierfür das Audiosignal in 32 Frequenzbänder geteilt (z.B. 100Hz, 200Hz, ... , 2kHz, 4kHz...). Dabei teilt sich ohne Komprimierung die Anzahl der möglichen Werte (Audioauflösung) durch die Anzahl der Frequenzbänder, und jedes Frequenzband erhält einen eigenen, entsprechend kleineren Wertebereich.

Für die Komprimierung nach den psychoakustischen Kriterien wird nun die Anzahl der möglichen Werte für durch benachbarte laute Frequenzen überdeckte Frequenzen gesenkt, da diese Frequenzen schlechter wahrgenommen werden. Dadurch können Daten eingespart werden.

Ein ähnliches Verfahren wird auf der Zeitachse des Audiosignals verwendet, um Signalen, die einem sehr viel lauterem Signal vorausgehen oder folgen, ebenfalls einen kleineren Wertebereich zuzuweisen.

Die entstehenden Ungenauigkeiten (Rauschanteil) bei der verlustbehafteten Komprimierung werden somit auf Signalanteile verteilt, die über den Hörapparat das Gehirn nicht oder nur sehr schwach erreichen. Der Qualitätsverlust ist im Verhältnis zur eingesparten Datenmenge sehr gering. Bei einer Komprimierung von 1:12 ist die Qualität für die Wiedergabe von Musik mit der Qualität von Radio vergleichbar.

In der Weiterentwicklung des umrissenen Verfahrens soll 'guessing' - der Versuch die Werte eines Folge-Samples zu raten - zum Einsatz kommen und die Komprimierungsrate nocheinmal erhöhen. Auch hierfür werden Versuchsreihen mit Testhörern durchgeführt, die zeigen sollen, bis zu welchem Maß das Verfahren geeignet ist, und an welchen Stellen die Abweichung in den geratenen Samples durch Checksummen korrigiert werden muß.

Die technischen Verfahren zur Dekomprimierung (!) solcher Datenströme wurden in ISO-Normen von der Motion Picture Encoding Group (MPEG) normiert. Zur Norm gehören nicht die Kodierungsverfahren, deren Qualität maßgeblich von der Forschungsarbeit der einzelnen Anbieter auf dem Gebiet der Psychoakustik abhängt. In den Normen der MPEG wird jedoch - wie der Name vermuten läßt - nicht nur Audiodekomprimierung festgelegt, sondern auch die Verfahren zur Videodekomprimierung. Zielsetzung der Gruppe ist es, Normen festzulegen, nach denen Bild und Ton von digitalen Datenträgern (CD, DVD) oder aus digitalen Datenströmen (Fernsehen, Internet) dekodiert und wiedergegeben werden können.

Der erste verabschiedete Standard war MPEG1. In MPEG1 (ISO11172) wurde festgeschrieben, wie von einer normalen CD mit 1-facher Abspielgeschwindigkeit Video und Audio wiedergegeben werden kann. Die Norm teilt den Datenstrom in drei Layer: Auf Layer 1 werden Daten für die Zusammensetzung von Audio und Video Datenstrom transportiert (System Stream), auf Layer 2 werden die Video-Daten als halbes PAL-Bild übertragen und auf Layer 3 letztendlich die Audio-Daten.

Da MPEG1 für die CD entwickelt wurde, sind flexible Bandbreiten für die einzelnen Kanäle nicht vorgesehen.

Das technische Verfahren zur Dekomprimierung von Audiodaten im anfänglich beschriebenen Format findet sich in der Nachfolgenorm MPEG2 Layer 3. Ebenso wie bei MPEG1 ist MPEG2 (ISO13848) in drei Layer für Steuerdaten, Video und Audio getrennt. Mittels der neuen Norm können Datenströme mit beliebiger Qualität dekodiert werden und die Verfahren wurden optimiert und erweitert um eine Übertragung des Datenstroms über eine verlustbehaftete Verbindung (z.B. einen Fernsehkanal) zu unterstützen. In einer Erweiterung (MPEG2 Layer3 ACR), die bisher noch nicht zum Einsatz kommt, ist schon festgeschrieben, wie mehrere Audiokanäle für z.B. Surround Sound übertragen werden.

Das Verfahren zur Komprimierung von Videosignalen funktioniert ähnlich wie die Komprimierung der Audiodaten (jedoch ist in der Norm auch für Video nur die Dekomprimierung für Video genormt): Das Videobild wird in Quadrate von 8x8 Pixeln zerteilt. Betrachtet man die 64 Pixel eines solchen Ausschnitts aneinandergereiht und fährt mit einer konstanten Geschwindigkeit an ihnen entlang, so ergibt sich eine Signal aus Helligkeitwerten, das sich genauso wie ein Audiosignal in Samples zerlegen läßt. Diese Samples können wieder in Frequenzbänder geteilt und nach ähnlichen Kriterien wie das Audiosignal komprimiert werden. Die Komprimierung der Videobilder entspricht technisch dem JPEG-Verfahren für Einzelbilder.

Zusätzlich wird betrachtet, ob sich ein bestimmter Ausschnitt aus 8x8 Pixeln im Folgebild an einer anderen Stelle wiedergefunden werden kann und somit nur dessen Bewegungsvektor und nicht die vollständige Bildinformation übertragen werden muß. Ein Fernsehbild läßt sich so komprimiert in einen

6MBit Datenstrom verpacken.

Konkurrierende Verfahren wie z.B. RealAudio, RealVideo oder LiquidAudio weichen vom technischen Verfahren von MPEG2 nur in der Umsetzung ab. Die Grundlagen sind dieselben. Es läßt sich nicht feststellen, daß eines dieser Verfahren qualitativ besser wäre als die normierten.

Für die Komprimierung von Audiodaten nach MPEG2 Layer 3 in Echtzeit wird ein PentiumII 300MHz benötigt. Eine günstige Hardwarelösung ist von Thomson für ca. 300,- DM angekündigt. Zum Abspielen reicht hingegen jeder Pentium-PC. Von zwei IC-Herstellern (ITT und Thomson) gibt es Ein-Chip-Lösungen für die Dekomprimierung, die in entsprechender Stückzahl für ca. 15$ verkauft wird.

Folgen, Utopien

Weltweit gibt es bisher 15 Hersteller, die Taschengeräte zum Abspielen von MPEG2 Layer 3 kodierten Musikdaten anbieten. Der Speicher der Geräte liegt zwischen 32MB und 64MB und reicht somit für 30-60min Musik in Radioqualität.

Im Internet werden Dateien mit Musikstücken verteilt und getauscht, die über eine Computerschnittstelle in die meisten Abspieler überspielt werden können. Diese Entwicklung wird von den großen Musikproduzenten gebremst, kann jedoch nicht gestoppt werden.

In Zukunft könnte die Stellung der Produktionsfirmen ins Wanken kommen, da der Vertrieb über das Internet theoretisch direkt von den Künstlern zu den Konsumenten möglich ist. Neben allen damit verbundenen Ideen, Idealen und Utopien wurde jedoch nicht aufgezeigt, daß es neben einer Untergrundszene, die illegal Musik kopiert und tauscht, ernstzunehmende Ansätze für die Bedienung eines öffentlichen Massenmarktes gibt.

Den einzige, von den Anwesenden abgelehnten Vorstoß in den elektronischen Vertrieb von Audiodaten unternimmt ein Internet-Anbieter in Deutschland: Über WWW-Seiten können Titel ausgewählt und gegen eine Gebühr von 0,10DM in Ausschnitten angehört werden. Wird ein Titel zum Kauf gewählt, so wird die Verbindung zum Internet-Anbieter getrennt und ein Server ruft via ISDN den Kunden an und liefert die Datei aus. Die Musikdatei ist allerdings verschlüsselt und kann nur mit dem Programm der Empfängerin gehört werden. Für die private Nutzung gibt es kostenlos eine Zweitlizenz für eine zweite Installation. Bei einem Preis von 3-5 DM pro Musiktitel wird an der Akzeptanz gegenüber dieses Verfahrens zu Recht gezweifelt.

Einigkeit bestand in dem Punkt, daß Musik zu teuer ist und nicht nachvollziehbar ist, wo das Geld bleibt. Der von den Produktionsfirmen angestrebte Preis von 50,- DM pro CD dürfte angesichts der aktuellen Entwicklungen jedoch nicht haltbar bleiben.

Ein philosophischer Ansatz für die freie Verfügbarkeit von Musik war, daß sie - einmal gehört - in Fragmenten im Gehirn gespeichert wird. Beim erneuten Hören werden lediglich Erinnerungen erneuert. Unbeantwortet blieb hierbei jedoch die Frage, warum die Dienstleistung der Auffrischung von Erinnerungen nicht vergütet werden sollte.