

Proseminar

Kompressionsalgorithmen am Beispiel Mp3



Von S. Kortler und E. Yueksekdag bei Dr. Gerhard Schrott

22. Mai 2003 TUM Sommersemester 2003

Gliederung

1. Wozu eigentlich Audiokompression
2. Verschiedene Arten der Audiokompression
3. Was ist Mp3
4. Verschiedene Encoder Routinen
5. Datenreduktionsverfahren
 - 5.1 Kurzbeschreibung
 - 5.2 Vorarbeit
 - 5.3 Hörschwellenmaskierung
 - 5.4 Frequenzmaskierung
 - 5.5 Zeitmaskierung
 - 5.6 Stereoredundanz
 - 5.7 Quantisierung

1. Wozu eigentlich Audiokompression

- Der Idee der Komprimierung von Audiodateien lag der Wunsch zugrunde, bessere **Sprachqualität** über Telefonleitungen aus Kupfer zu übertragen. Dies erübrigte sich bald durch die rapide Entwicklung des Glasfasernetzes und neuen Anschlussmöglichkeiten für den Privatbenutzer.
- Technologien wie ISDN brachte die Sprachübertragung entscheidend voran. Durch diese Entwicklung erhielt man neue, visionäre Ansatzpunkte für Einsatzgebiete der Audiocodierung. Erlanger Wissenschaftler wollten nicht nur die Übertragung von Sprache, sondern auch von **Musik** über Telefonleitung oder

anderer Kanäle, bei denen man besonders sparsam mit der verfügbaren Datenrate umgehen muss, ermöglichen denn:

- Speichern von Audiodateien in CD-Qualität benötigt sehr viel Platz. Um 1 Minute mit 44,1 KHz, 16 Bit und Stereo (CD-Qualität) zu archivieren benötigen wir ca. 10 MB. Um das ganze dann noch in Echtzeit über z.B. das Netz zu übertragen benötigen wir eine Leitung die 2.7 Megabit pro Sekunde bringt. Damit können wir streaming für „normale“ Anwender nicht ermöglichen.
- Um nun trotzdem Audiodaten bequem über das Netz verschicken zu können müssten wir ganz empfindlich weit mit der Qualität herunter gehen um riesengroße Wartezeiten zu vermeiden oder aber die Daten komprimieren. Am Fraunhofer Institut entwickelten daraufhin die Wissenschaftler drei Jahre weiter, bis sie 1989 ein Patent anmeldeten welches 1992 von der ISO anerkannt wurde. Und in die MPEG Spezifikation integriert wurde.
- Der Schlüssel zum Erfolg liegt allerdings in der anfänglichen Freigabe, der Nutzung und Entwicklung der Technologie.
- Mit der Entwicklung der ersten MP3 Player 1997 (AMP Advanced Multimedia Products) von Tomislav Ucelak erlangte das Format seinen endgültigen Durchbruch.

2. Verschiedene Arten der Audiokompression

- Es gibt **verlustlose** Audiokompression (Datenkomprimierung) und **verlustbehaftete** Audiokompression (Datenreduktion)
- Für die verlustlose Audiokompression gilt das gleiche wie für ganz normale Datenkomprimierung. Umgesetzt wird das ganze mit Verfahren wie **LZW** oder **Huffmann-Kodierung**, wie sie von **Zip**, **Lha** oder ähnlichen benutzt werden. Nach dem

Dekomprimieren erhalten wir eine nicht vom Original zu unterscheidende Datei. Diese Verfahren sind uns auch aus der Vorlesung bekannt. Dies ist der große Vorteil von dieser verlustlosen Komprimierung. Jedoch sind diese Verfahren nicht extra für Audiokompression entwickelt, und es lässt sich damit auch kein besseres Ergebnis als eine Komprimierung auf die Hälfte der ursprünglichen Größe erzielen. Darauf werde ich nicht weiter eingehen.

- Bei der Datenreduktion hingegen werden deutlich bessere Ergebnisse erzielt. Eine Kompression von etwa **12 bis 14 zu 1**. Dafür jedoch gehen ganz offensichtlich jede Menge Daten verloren. Das ist unerheblich, solange einfach die Daten verloren gehen, die wir mit unserem menschlichem Ohr sowieso nicht gehört hätten.

3. Was ist MP3

- Der Name **Mp3** stammt von der Gruppe **MPEG**. Moving Picture Experts! Ein Synonym für eine Gruppe von Standards, die das digitale kodieren von Filmen, Videos und Musik ermöglicht.
- Es gibt mehrere sogenannte **Layer** für die Audiokomprimierung. Layer 1 bis 3. Alle benutzen die wahrnehmungsbezogene Kodierung, die jedoch von Layer 1 bis 3 stark verbessert wurde.
- Layer 1 benutzt ausschließlich die Frequenzmaskierung auf die ich später noch eingehen werde. Kurz gesagt es werden die für das menschliche Ohr unhörbaren Frequenzen weggelassen.
- Layer 2 arbeitet schon komplexer. Er „weiß“ zum Teil welche Informationen weggelassen werden können. Ab einer Samplingrate von 256 kbps ist selbst für geschulte Ohren kaum ein Unterschied zum Original festzustellen.

- Layer 3 benutzt zusätzlich die Huffman-Kodierung. Die uns aus der Vorlesung bekannt ist, und zu der später noch was gesagt wird. Es ist das beste und komplexeste Audiomodell was derzeit auf dem Markt ist.
- Der Name kommt also von MPEG Layer 3 kurz Mp3.

4. Verschiedene Encoder Routinen

- Zunächst gibt es nur zwei verschiedene Encoder-Algorithmen. Die der **ISO** und die **ACM** (Audio Compression Manager) Algorithmen des **FIIS** (Fraunhofer Institut für Integrierte Schaltungen).
- Für die Layer der ISO gibt es freie Quellcodes. Die ACM-Algorithmen müssen jedoch beim FIIS lizenziert werden. Das hat auch seine Begründung denn:
- Die ISO-Layer 3 Codes bieten gute Qualität bei Bitraten über 128 kbps, sind jedoch immer noch schlechter als die Layer 2 der ACM.
- Die tolle Verbreitung des MP3 Formates ist auf den Layer 3 des Fraunhofer Instituts zurückzuführen. Durch die schon erwähnte anfängliche **Freigabe** der Technologie zum Experimentieren und Entwickeln, haben viele abhängige und unabhängige Wissenschaftler mit Hilfe der Spezifikation neue Software entwickelt. Die den Durchbruch ermöglichten.
- Ein guter Algorithmus allein ist noch nicht viel. Die Leute müssen auch damit **arbeiten**.
- Denn heute gibt es schon deutlich **effizientere** Routinen. Algorithmen die bessere Qualität bei weniger Speicherplatz

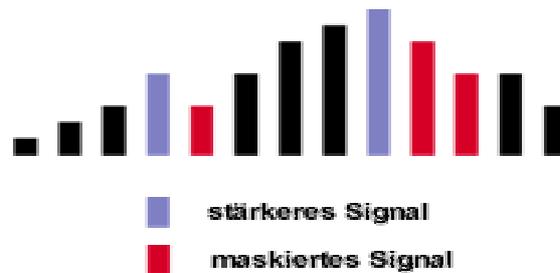
liefern. Z.B. Advanced Audio Compression (AAC), Perceptual Audio Coder (PAC) oder Ogg Vorbis. Aber das soll nicht Thema dieses Vortrags sein.

1. Datenreduktionsverfahren

5.1 Kurzbeschreibung

- Wie werden diese tollen Kompressionsraten erreicht, ohne dass ein hörbarer Unterschied auftritt?
- Es werden einfach die Signale in einem Song weggelassen, die von einem anderen Signal übertönt (das nennt man maskiert) werden.
- Beispiel: In der Disko haben wir 110 dB. Es ist kein Gespräch möglich. Es wäre unsinnig dieses mit aufzuzeichnen. Man sagt. Die laute Musik in der Disko überdeckt (**maskiert**) das Gespräch.
- Das passiert genau dann, wenn das Energieniveau eines Geräuschs höher ist als das eines Anderen (Frequenzmaskierung). Andere Signale in Songs sind sowieso für uns nicht hörbar, da sie unter unserer Hörschwelle liegen (Hörschwellenmaskierung). Diese können einfach ausgespart werden.
- Der Encoder entscheidet, welche Signale für uns verzichtbar sind und welche nicht.
- Dabei stellt er sich sehr geschickt an, denn die Signale die der Encoder auslöscht, wären später von unserem Gehirn sowieso ignoriert worden.

- Diese irrelevanten Informationen brauchen natürlich auch nicht unsere Festplatte zu belegen oder die Internetleitung zu blockieren.
- Der Hauptanteil von irrelevanten Informationen verschwindet wie schon angesprochen durch das Maskieren



- Der Encoder erkennt also welche Signale noch wahrgenommen werden und welche maskiert werden. Doch woher weiß der Encoder was wahrgenommen wird und was nicht?

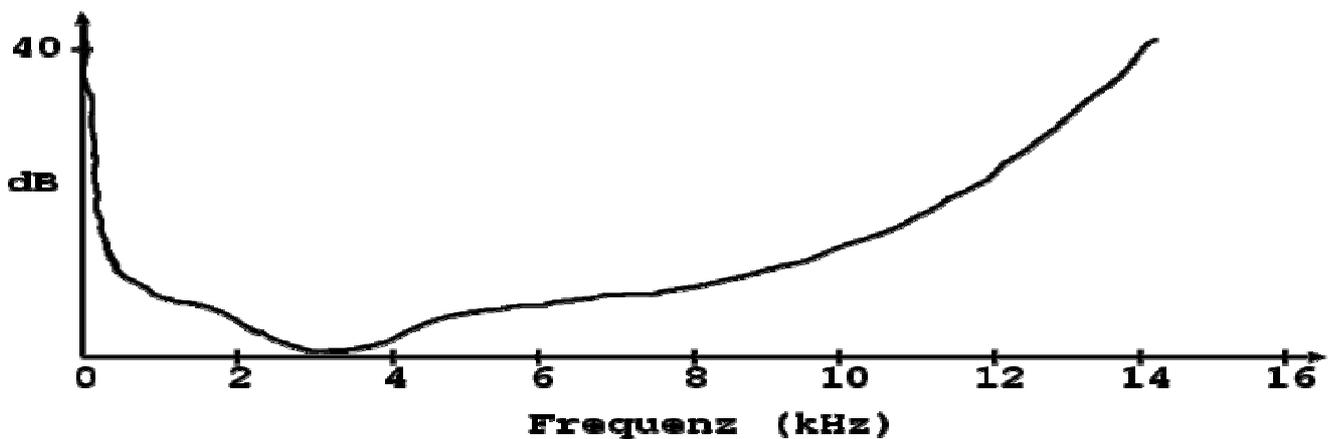
5.2 Vorarbeit

- Als erstes wird durch die Fast Fourier Transformation (**FFT**) oder der Diskreten Kosinus Transformation (**DCT**) das Frequenzspektrum der Audiodaten bestimmt. Dazu später im Vortrag genaueres.
- Dann wird das gesamte Spektrum in viele Intervalle eingeteilt. Auf jedes dieser Intervalle passt dann ein speziell angepasstes psychoakustisches Modell des Ohres.
- Nun wird die Zeitachse des Audiosignals in Intervalle unterteilt, um das Audiosignal stückchenweise zu kodieren.

5.3 Hörschwellenmaskierung

- Jetzt werden wie am Anfang angesprochen, in jedem einzelmem Intervall die Signale aus dem gesamten Spektrum entfernt, die unter der Hörschwelle des menschlichen Ohres liegen und dadurch für uns nicht notwendig sind.

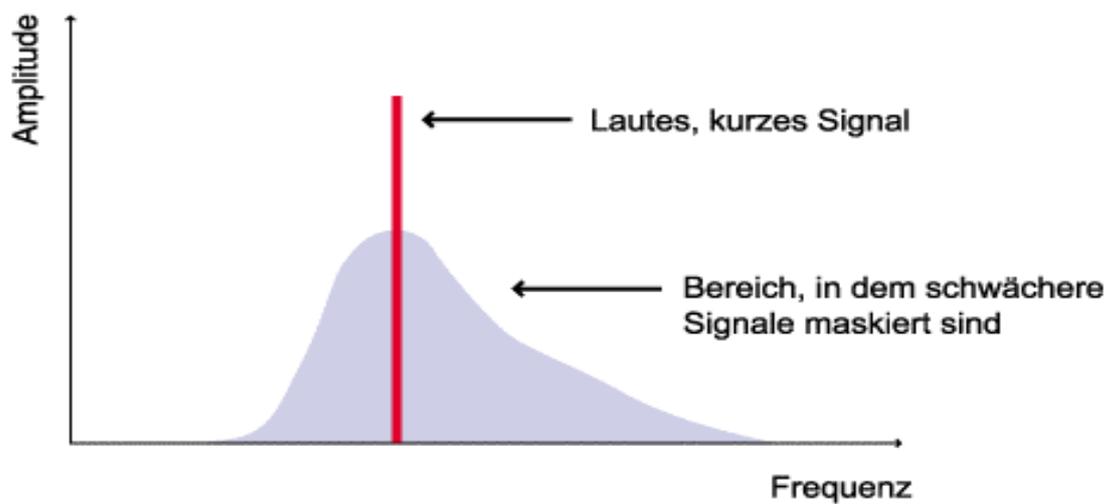
Abbildung: Hörschwelle des menschlichen Gehöres



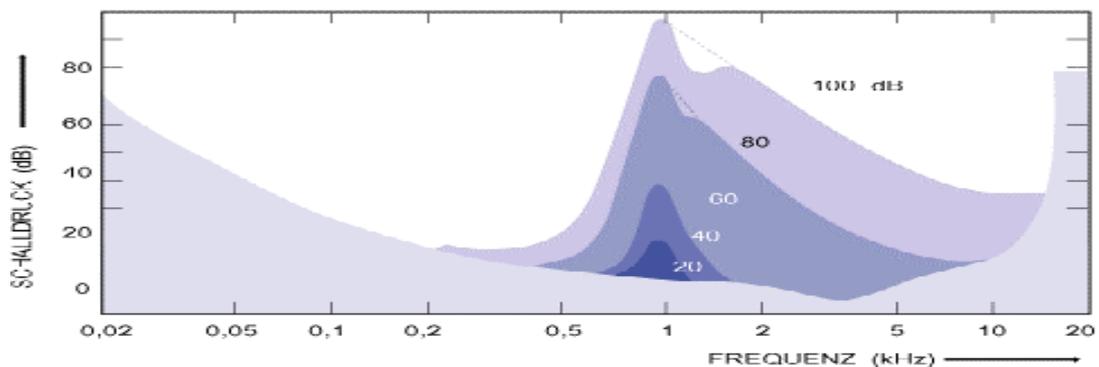
- Wie in Abbildung 1 ersichtlich, würde z.B. ein Ton mit einer Frequenz von 12 kHz und einer Lautstärke von 5 dB entfernt werden, da das menschliche Gehör ihn nicht wahrnimmt.

5.4 Frequenzmaskierung

Nach der Hörschwellenmaskierung folgt eine Frequenzmaskierung. Bei dieser Maskierung werden genau die Signale entfernt, die von Signalen mit einer anderen Frequenz übertönt werden.



Wir haben einen Ton bei 1 kHz und einen bei 1,1 kHz. Dieser ist ca. 18 dB leiser. Er wird vom ersten Ton vollkommen verdeckt, da er unterhalb der Mithörschwelle liegt. Also werden auch andere schwächere Töne in unmittelbarer Nähe des ersten Tones maskiert. Ein anderer Ton mit 2 kHz wäre jedoch hörbar, da er in einem anderen Frequenzband liegt.



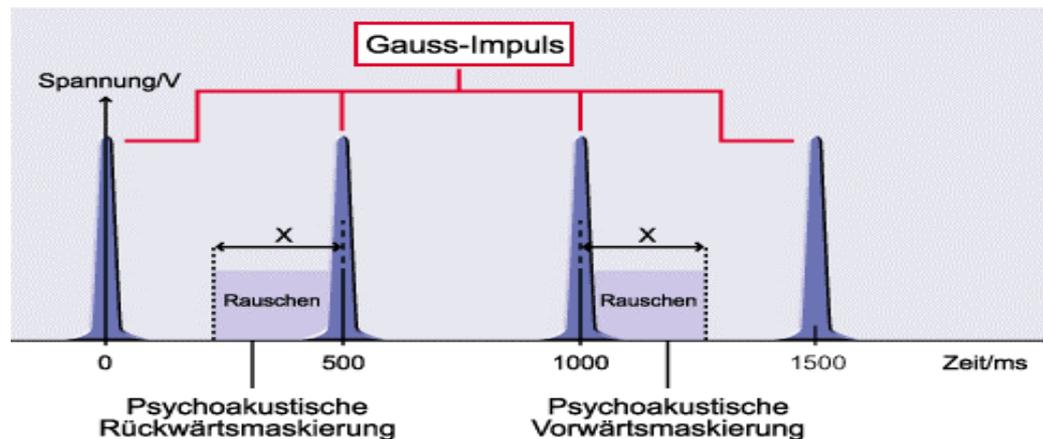
Die Grafik zeigt die Mithörschwellen bei Tönen, die von Schmalbandrauschen der Mittenfrequenz 1 kHz verdeckt werden, in Abhängigkeit von der Frequenz.

5.5 Zeitmaskierung

- Hier werden wie bei der Frequenzmaskierung auch Signale entfernt, die von anderen Signalen übertönt werden, jedoch

zusätzlich wird die zeitliche Abfolge der verschiedenen Signale beachtet.

- Das Gehör braucht bei allen Tönen eine kurze Erholungszeit bis es wieder aufnahmefähig ist.



- Beim Gauß-Impulse wird vor und nach dem Knacken das leise Rauschen überhört, **Nachmaskierung ca. 200 ms** und **Vormaskierung ca. 20 ms**
- Der laute Impuls wird vom Hirn schneller verarbeitet, als das leise Rauschen. Nach der Zeitmaskierung haben wir bereits um den Grad 5 bis 6 komprimiert!
- Das Spektrum eines Sinustons fällt beiderseits Gaußförmig ab. Schwächere Signale, vor und nach den Ton werden, überdeckt.

5.6 Stereoredundanz

- In einem Stereosignal sind meist auf beiden Seiten **gleiche** Anteile. Diese müssen natürlich **nicht doppelt** gespeichert

werden. Zusätzlich sind aber natürlich Elemente vorhanden, die nicht auf beiden Kanälen gleich sind. Daher kann man nicht einfach beide Kanäle zusammenlegen. Stattdessen gibt es verschiedene Verfahren den Stereoeffekt zu erhalten und trotzdem weiteren Platz zu sparen. Für jedes einzelne Intervall wird deshalb entweder das **Intensity-Stereo-**, das **Mid/Side-Stereo-** oder kein Redundanzverfahren verwendet.

- Bei Intensity-Stereo (**IS-Stereo**) wird die Summe der beiden Kanäle genommen und dann für jede Seite ein Skalierungsfaktor bestimmt. Dadurch geht natürlich in diesem Intervall der Stereoeffekt verloren. Es wird aber auch nur angewandt wenn nur Frequenzen unter 80 Hz vorhanden sind, da man bei diesen nicht orten kann aus welcher Richtung sie kommen.
- Bei Mid/Side-Stereo (**MS-Stereo**) wird ein Summen- und ein Differenzsignal gebildet, dabei ist zu beachten, dass später dem Differenzsignal weniger Speicherplatz zur Verfügung steht. Diese Kodierung spart zwar nicht ganz so viel Speicherplatz, dafür bleibt aber die Phaseninformation erhalten, die bei der vorigen Intensity-Stereo-Variante verloren gehen. So können auch Dolby-Surround Signale kodiert werden.

5.7 Quantisierung

- Das ist die Umsetzung **analoger in diskrete** Werte. Es bedeutet eine Unterteilung des Amplitudenbereichs eines kontinuierlich verlaufenden Signals in eine endliche Anzahl kleiner Teilbereiche.
- Die Quantifizierung der Samples ist der nächste Arbeitsschritt zur Datenreduktion. Jedes Sample besteht zwar aus 16 Bit, jedoch sind nicht alle 16 nötig, um das Signal darzustellen. Die führenden Nullen eines 16-Bit-Samples könnten entfallen. Ergibt sich bei einem Sample der Wert 0000011101010101, werden die

führenden Nullen abgeschnitten.

- Um daraus wieder die 16 Bit rekonstruieren zu können, benötigt der Decoder den Skalierungsfaktor sowie die Bit-Allocation.
- Der Skalierungsfaktor gibt an, an welcher Stelle die verbliebenen Bits des Samples gewesen sind.
- Die Bit Allocation enthält die Information, wie viele Bits im Sample verblieben sind, da nicht mehr mit einer festen Zahl von 16 Bit gerechnet werden kann.
- Diese Informationen teilen sich je zwölf Samples, um nochmals Platz zu sparen.
- Es macht sich ein sogenannter **Quantisierungsfehler** bemerkbar (ca. 6 dB Rauschen pro weggelassenem Bit). Hierbei wird wieder darauf geachtet, dass der Fehler unter den oben beschriebenen Maskierungsschwellen bleibt, damit das Gehör diesen Fehler nicht bemerkt.

This document was created with Win2PDF available at <http://www.daneprairie.com>.
The unregistered version of Win2PDF is for evaluation or non-commercial use only.