

Advanced Audio Coding - AAC

Datenkompression Pro-Seminar

Dozent: Dr.-Ing The Anh Vuong

Graphische Daten Verarbeitung, Informatik Institut

Johann Wolfgang Goethe Universität , Frankfurt am Main

Projekttitel	Advanced Audio Coding - AAC
Projektautoren	Johannes Kaepfel / 3516096
Erstellt am	20.06.2011
Zuletzt geändert	06.07.2011
Status	x in Arbeit <ul style="list-style-type: none">- vorgelegt- fertig gestellt- Abgenommen
Dokumentablage	

History

Version	Datum	Name	Status	Bearbeitung
0.1	20.06.2011	Johannes Kaepfel	la	erstellen

Status: **ia**: in Arbeit, **v**: vorgelegt, **f**: fertig, **ab**: abgenommen

Inhaltsverzeichnis

1 Thema	1
1.1 Einführung	1
1.2 Warum spezielle Audiokompression?	2
1.3 Warum ist MP3 noch so verbreitet?	2
2 Theoretische Grundlagen.....	3
2.1 Psychoakustik	3
2.2 Abtastung	5
2.3 Quantisierung / Scaling	6
2.4 Modifizierte Diskrete Kosinus Transformation (MDCT)	6
2.5 TNS	7
2.6 Gain Control	7
2.7 Intensity coding	7
2.8 Prediction	8
2.9 M/S	8
2.10 Variable Length coding - VLc	8
3 Beschreibung des Verfahrens	9
3.1 Daten- und Kontrollfluss	9
3.2 Der Ablauf	10
4 Anwendungsgebiet	11
4.1 Wann sollte AAC eingesetzt werden?	11
4.2 Wann sollte AAC nicht eingesetzt werden?	11
5 Qualitätsbewertung	12
6 Literatur / Referenzen	15

1 Thema

Diese Seminararbeit befasst sich mit dem im MPEG-4 Standard beschriebenen Kompressionsverfahren AAC. Es wird vermittelt auf welchen Grundlagen AAC basiert, wie das Kompressionsverfahren arbeitet und welche Vor- und Nachteile dieses Verfahren hat.

1.1 Einführung

Advanced Audio Coding (AAC) ist ein mächtiger Kompressionsstandard, welcher seinen Vorgänger MP3 sowohl in Effizienz als auch in der Vielzahl seiner Anwendungsbereiche übertrifft. Entwickelt wurde AAC vom Fraunhofer Institut für Integrierte Schaltungen, Dolby, AT&T, Sony und Nokia. Im Jahr 1997 wurde AAC im MPEG-2 Layer 7 als ISO13818-7 definiert und ersetzte damit das MP3-Format.

Die Größe von Audiodateien lässt sich durch AAC erheblich reduzieren, und obwohl es sich um ein verlustbehaftetes Verfahren handelt, besticht es, vor allem bei geringen Bitraten, durch eine sehr gute Wiedergabequalität. Das Verfahren ist nicht abwärtskompatibel, da sonst die hohe Qualität in dieser Form nicht möglich wäre. Wenn man die Unterschiede zwischen den beiden Verfahren betrachtet, wird verständlich, warum ein MP3-Decoder kein AAC dekodieren kann (siehe Tabelle 1).

Obwohl MP3 nach wie vor sehr verbreitet ist und einen großen Stellenwert hat, beginnt AAC an Popularität zu gewinnen. Das Unternehmen Apple betreibt den iTunes Store, über den Musik digital verkauft wird. Als Standard-Kompressionsverfahren wird dort AAC mit einer Bitrate von 128kbps genutzt. Außerdem entschied sich Apple bei der Entwicklung seiner mobilen Multimediageräte (iPhone, iPad, iPod), die Dekodierung von AAC in Hardware zu unterstützen. Allgemein ist zu beobachten, dass die Verbreitung von Smartphones zunimmt und dedizierte mobile Multimediageräte wie MP3-Player durch sie ersetzt werden. Die Anbindung an das Internet erfolgt in der Regel über das Mobilfunknetz, wobei die Netzabdeckung Probleme bereitet. In Ballungszentren ist die Netzabdeckung von Hochgeschwindigkeitsnetzen ausreichend gegeben, doch in dünner besiedelten Gebieten muss auf Netze mit eingeschränkter Bandbreite zurückgegriffen werden. Um dennoch bei Echtzeitübertragungen wie etwa Audiostreams die Ladezeiten zu reduzieren oder gar zu vermeiden, bedarf es einer effizienten Kompression der Daten. Sowohl die Technik in den Geräten als auch die der Kopfhörer ermöglichen hochqualitative Audiowiedergabe. Daher reicht es nicht, die Daten auf ein Minimum zu komprimieren. Vielmehr soll dem Anwender nicht auffallen, dass eine Kompression stattgefunden hat. Für diesen Anwendungsbereich ist AAC eine ideale Lösung.

Dass es ein zukunftsweisender Standard ist haben bereits mehrere Unternehmen erkannt und so hat die japanische ARIB die Audiostreams für alle digitalen Rundfunksysteme Japans auf AAC umgestellt. Weitere Unternehmen die auf AAC setzen sind XM Radio, DRM Consortium, Digital Video Broadcasting, Internet Streaming Alliance, 3GPP und die Bluetooth Special Interest Group. Von etablierten Halbleiterherstellern wie beispielsweise National Semiconductor sind inzwischen Dekodier-Chips auf dem Markt, so dass immer mehr moderne Multimediageräte AAC dekodieren können.

1.2 Warum spezielle Audiokompression?

Es ist möglich, akustische Informationen mit weit verbreiteten Methoden, wie zum Beispiel Zip oder Rar, zu komprimieren. Aufgrund der universellen Anwendbarkeit dieser Kompressionsverfahren gehören sie zu der Kategorie der verlustfreien Kompressionen. Maßgabe für verlustfreies Komprimieren ist, dass die Originaldaten zu 100% wiederhergestellt werden können. Ein häufiger Fall, der verlustfreie Kompression verlangt, ist die Kompression von Software, da durch minimale Veränderungen die Software unbrauchbar würde. Um dies sicherzustellen müssen enorme Einbußen im Kompressionsgrad hingenommen werden.

Da bei Audio- und Videowiedergabe das menschliche Wahrnehmungsvermögen einen natürlichen Flaschenhals bietet, ist eine verlustfreie Kompression nicht nötig, um das wiederhergestellte Signal dem Originalen gleich erscheinen zu lassen. Es wird versucht, die entstandenen Verluste in den Bereich von Schwächezonen unserer Wahrnehmung zu legen. Durch diese Täuschung kann der Kompressionsgrad signifikant erhöht werden. Da unsere akustische und visuelle Wahrnehmung sehr unterschiedliche Grenzen hat, lässt sich die Effizienz der Kompression steigern, indem man sowohl für Bildkompressionen als auch für Audiokompressionen dedizierte Methoden verwendet.

1.3 Warum ist MP3 noch so verbreitet?

MP3 wurde 1993 als ISO/IEC 11172-3 im MPEG1 Layer3 definiert. Obwohl AAC nur 4 Jahre darauf folgte, ist MP3 immer noch das am weitesten verbreitete Audiokompressionsverfahren weltweit. Eine große Anzahl an Wiedergabegeräten unterstützt den MP3 Standard, ebenso sind sehr schnelle Kompressions- und Dekompressionsalgorithmen frei verfügbar und leicht zu finden. AAC kämpft nach wie vor mit einer zu geringen nativen Unterstützung. Freie Kompressions- und Dekompressionsalgorithmen sind zwar verfügbar, aber die Verwendung ist meistens nicht so einfach wie bei MP3. Microsoft Windows ist das am weitesten verbreitete Betriebssystem und unterstützt AAC nicht ohne Zusatzsoftware. Der Windows Media Player ist mit einem Plug-In in der Lage, AAC zu dekodieren, jedoch erweist sich die Installation eines AAC-fähigen Drittanbieter Mediaplayers wie Foobar oder VLC für einen Laien leichter. AAC wird unter Linux aufgrund der Opensource-Gemeinde schon lange unterstützt und auch Apple unterstützte AAC schon früh in seinen Produkten. Die Unterstützung des AAC Standards nimmt zwar bei neuen Wiedergabegeräten zu und damit wird auch dessen Verbreitung vorangetrieben, doch sind noch sehr viele Geräte ohne AAC-Unterstützung auf dem Markt. Stellt man die Eigenschaften von MP3 und AAC gegenüber so wird klar, dass MP3 technisch AAC unterlegen ist. Das AAC-Signal ist dynamischer und bei gleicher Qualität ist die Bitrate ~30% kleiner als die einer MP3.

Bis AAC das verbreitetste Audiokompressionsverfahren ist, wird es mit Sicherheit noch einige Jahre dauern, doch aufgrund der technischen Unterlegenheit seiner Konkurrenz wird sich AAC auf Dauer durchsetzen.

	MP3	AAC
Blockgröße stationär / dynamisch	576 / 192 Samples	1024 / 128 Samples
Abtastrate	8KHz – 48KHz	8KHz – 192KHz
Filterbank	Hybride Filterbank	Nur MDCT
Audiokanäle	Maximal 5	Maximal 48
Bitrate	8 – 320 kbps	8 – 529 kbps

Tabelle 1, Unterschiede zw. MP3 und AAC

2 Theoretische Grundlagen

2.1 Psychoakustik

Die Psychoakustik befasst sich mit der Erforschung der Eigenschaften des menschlichen Gehörsinnes. Dabei wird das Gehör als Blackbox, welche ein Schallsignal in eine Empfindung übersetzt, betrachtet. Mittels vieler Experimente, in denen geschulte Testpersonen die Hörbarkeit von Audiosignalen bewerten, kann ein theoretisches Modell, das die speziellen Eigenschaften des menschlichen Gehörs beschreibt, ermittelt werden. Interessant an diesem Modell ist unter welchen Bedingungen Schallwellen nicht, oder nur reduziert, in eine Empfindung übersetzt werden. Betrachtet man diese Informationen im Kontext der Datenkompression so erkennt man, dass Daten, welche nicht übersetzt werden, nicht gespeichert werden müssen. Bei Daten, welche nur schwach in eine Empfindung übersetzt werden kann eine ungenaue, Speicher sparende Darstellung gewählt werden. Jeder Mensch hat einen unterschiedlich ausgeprägten Gehörsinn, daher kann dieses theoretische Modell nicht den Gehörsinn jedes Menschen exakt beschreiben. Jedoch gibt es bei dem Gehörsinn jedes Menschen Gemeinsamkeiten auf die die Kompression setzen kann. Die dominierenden Techniken, welche in der Audiokompression Anwendung finden, sind nachfolgend beschrieben.

- **Hörschwellenmaskierung**

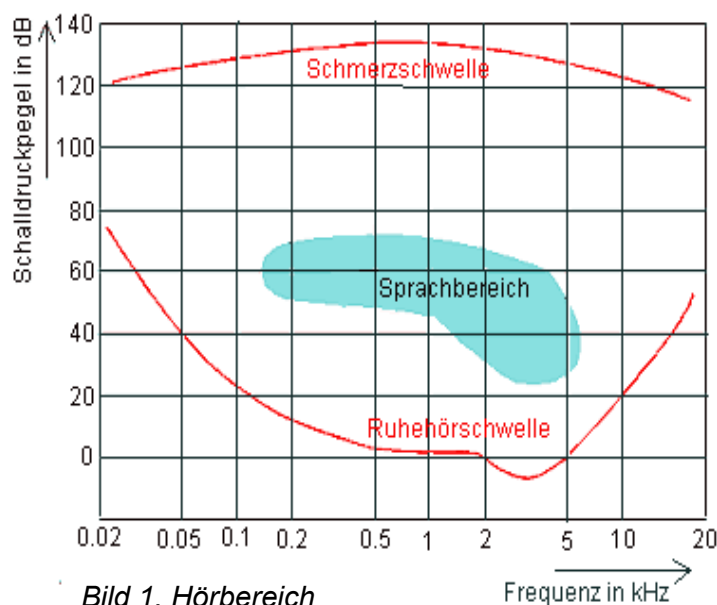
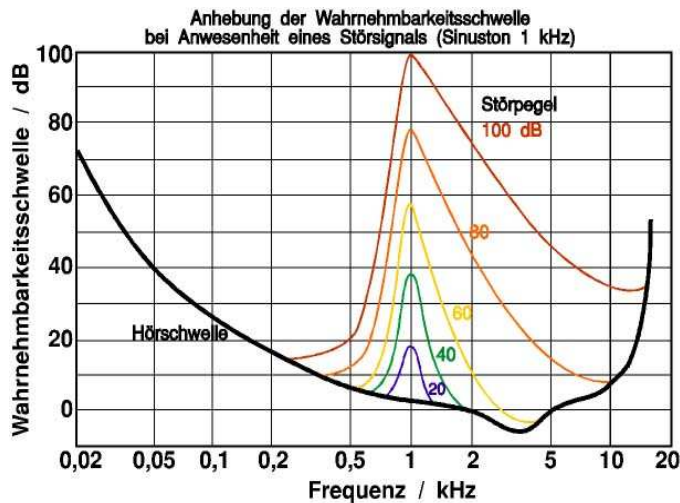


Bild 1, Hörbereich

Das Hörempfinden des Menschen reagiert nur auf Frequenzen zwischen ca. 20Hz bis 20KHz, jedoch ist je nach Frequenz ein Mindestschalldruckpegel nötig damit das Geräusch wahrnehmbar wird. Exakt auf der Ruheshwelle liegende Signale erscheinen dem Menschen gleich laut, obwohl sie unterschiedliche Schalldruckpegel aufweisen. Besonders empfindlich reagiert unser Gehör auf Frequenzen zwischen ca. 2KHz und 5KHz. Der Bereich zwischen der Ruheshwelle und der Schmerzschwelle wird Hörfläche genannt. Nur in diesem Bereich sind Informationen für unser Gehör relevant.

Je höher die Ruheshwelle, desto niedriger kann die Qualität des Signals sein ohne eine Veränderung der Empfindung zu erzeugen. In der Hörschwellenmaskierung werden Teile des Audiosignals entsprechend ihrer Wahrnehmbarkeit markiert, so dass bei der Quantisierung bei weniger relevanten Signalabschnitten eine geringere Auflösung gewählt werden kann. Bereiche außerhalb der Hörfläche werden entsprechend markiert und bei der Quantisierung komplett ignoriert.

- **Frequenzmaskierung**



Je nach Schalldruckpegel kann ein Signal andere Frequenzen überdecken, man spricht von Maskieren. Die überdeckten Frequenzen sind unhörbar und können komplett ignoriert werden. Je höher der Schalldruckpegel, desto größer ist der maskierte Bereich (siehe Bild 2). Ähnlich verhält es sich bei der Frequenz, je höher die Frequenz, desto breiter ist der maskierte Bereich (siehe Bild 3). Mit dieser Erkenntnis lassen sich sehr viele Informationen und damit Speicherplatz einsparen.

Bild 2, Überlagerung bzgl. Schalldruckpegel

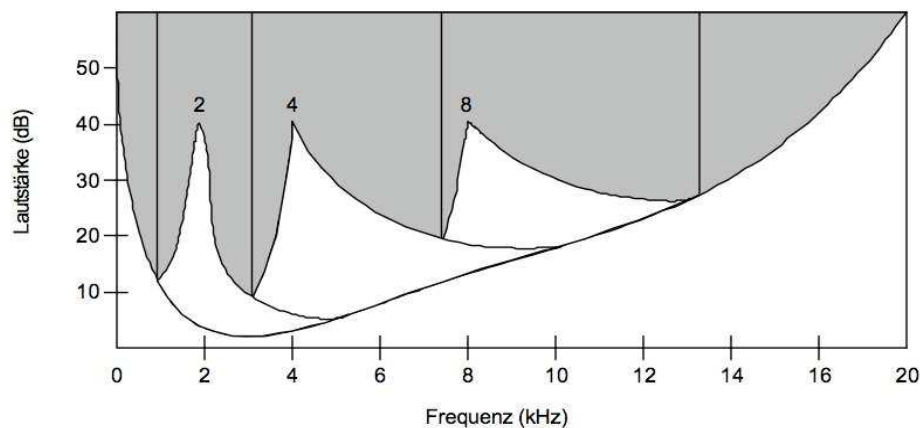


Bild 3, Überlagerung bzgl. Frequenz

- **Zeitmaskierung**

Die Zeitmaskierung beschreibt den Effekt, dass vor und nach einem Signal eine Überdeckung stattfindet (siehe Bild 4).

Wenn ein Signal einen gewissen Schalldruckpegel aufweist, so dauert es nach dem Verstummen einen Augenblick bis von diesem Signal vorher überdeckte Frequenzen wahrgenommen werden. Ebenfalls wird ein kurzer Bereich vor dem Signal unhörbar. Diese Signalabschnitte werden beim Abtasten nicht berücksichtigt, da sie keine für das Hörvermögen verwertbaren Informationen enthalten.

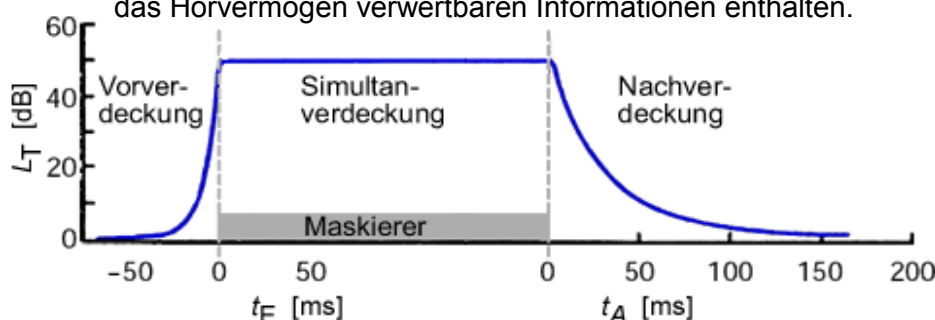


Bild 4, Zeitmaskierung

- **räumliche Redundanz**

Die räumliche Redundanz beschreibt den Effekt, dass Schallquellen mit sehr niedrigen Frequenzen ortonabhängig erscheinen. Die menschliche Wahrnehmung ist außer Stande, die Richtung, aus der Schall unter ca. 400Hz kommt, zu bestimmen. Dies kann zur Datenkompression verwendet werden, da diese Frequenzen nur in einem Kanal, also Mono, gespeichert werden müssen.

2.2 Abtastung

Beim Abtasten wird in regelmäßigen Abständen die Signalstärke gemessen und als so genanntes Sample gespeichert. Je kleiner diese Zeitabstände (je höher die Samplerate), desto genauer kann anhand der gespeicherten Werte das Signal rekonstruiert werden. Das in Bild 5 dargestellte analoge Signal kann nach der Abtastung zu einem Signal wie in Bild 6 rekonstruiert werden.

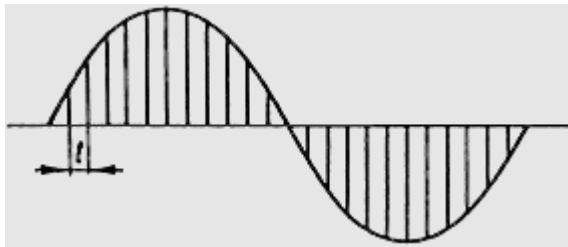


Bild 5, Analogsignal

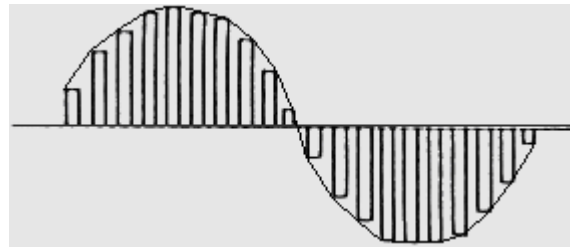


Bild 6, abgetastetes Signal

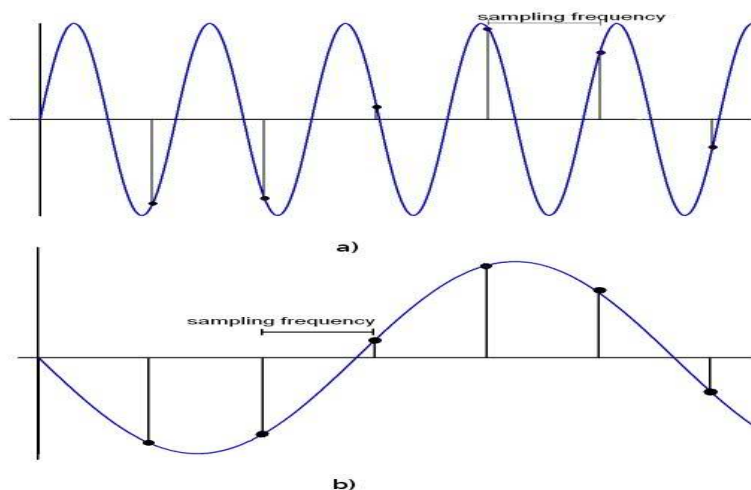


Bild 7, der aliasing Effekt

Wird die Samplerate, also die Anzahl der Abtastungen pro Sekunde, zu gering gewählt, so entsteht nicht nur das Problem, dass das rekonstruierte Signal zunehmend an Details verliert. Das viel größere Problem ist, dass Veränderungen zwischen den Samples nicht registriert werden. Im Extremfall kann eine Signalverfälschung auftreten wie sie in Bild 7 dargestellt wird. Bei diesem Phänomen spricht man von einem aliasing Effekt.

Um diesen Problemen entgegenzuwirken, müssen mindestens 2 Samples pro Amplitudenwechsel erzeugt werden. Wählt man die maximale vom Menschen hörbare Frequenz (20KHz), so benötigt man eine Abtastung von 40KHz (40000 Samples pro Sekunde) um alle Signale, die der Mensch hören kann, fehlerfrei zu rekonstruieren. Die am häufigsten verwendete Samplingrate ist 44.1KHz.

Um weitere aliasing Effekte zu vermeiden, wird das Signal vor der Abtastung durch einen Filter von allen Frequenzen befreit, die oberhalb 20KHz liegen (siehe Psychoakustik). Das nach der Abtastung aus den Samples bestehende Signal wird auch PCM-Signal genannt. PCM steht dabei für Pulse-Code-Modulation.

2.3 Quantisierung / Scaling

Die Quantisierung macht sich eine Ungenauigkeit des menschlichen Gehörs zunutze. Wenn ein Signal sich nur minimal verändert, so erscheint es dem Menschen als bliebe es konstant. Erst wenn die Veränderung stark genug ist, wird sie wahrgenommen. In Experimenten wurde ermittelt, dass unser Gehör weniger als 65536 Abstufungen wahrnehmen kann. Diese Anzahl entspricht 2Byte, bzw. 16Bit. Für jedes Bit weniger entsteht ein Quantisierungsrauschen von 6db.

Bei der Quantisierung wird jedes Sample aus der Abtastung mit den ermittelten Abstufungen verglichen und auf den Wert der ähnlichsten gerundet. Anhand des psychoakustischen Modells werden verschiedene Frequenzen unterschiedlich fein abgestuft (Scaling). Es würden deutlich mehr als 16Bit pro Sample benötigt, wenn jedes Sample den exakten Wert behalten würde. Da nach der Quantisierung nur noch maximal 16Bit pro Sample belegt werden, erweist sie sich als der Teil der Audiodatenkompression, der am meisten Einsparpotential bietet.

2.4 Modifizierte Diskrete Kosinus Transformation (MDCT)

Die MDCT transformiert ein zeitdiskretes Signal in ein frequenzdiskretes. Dies ist notwendig, da unser Gehör auf Basis von Frequenzbereichen arbeitet. Zusätzlich lässt sich das Signal nach der Transformation kompakter darstellen und ist so ein weiterer Kompressionsfaktor. Die MDCT erhält verschieden groß gewählte Frames, die nacheinander transformiert werden. Dabei unterscheidet sie sich von der DCT (Diskrete Kosinus Transformation) durch eine Überlappung der Fenster (siehe Bild 8). Dies ist nötig um sicherzustellen, dass die inverse Transformation zum Ursprungssignal zurückführt. Zusätzlich werden aliasing Effekte zwischen zwei Frames reduziert.

Die Ausgabe der MDCT sind einzelne Blöcke von Frequenzbändern, die durch andere Methoden abschnittsweise kodiert werden können. Ein besonderes Augenmerk fällt auf die Wahl der Fenstergröße. Ist ein Signalabschnitt gleichmäßig, so wird eine lange Fenstergröße gewählt, ist es unstetig, so wird eine kleine Fenstergröße gewählt. Ein langes Fenster besteht aus 8 kleinen Fenstern. Dies bewirkt eine weitere Reduzierung von aliasing Effekten und eine Erhöhung der Kompression.

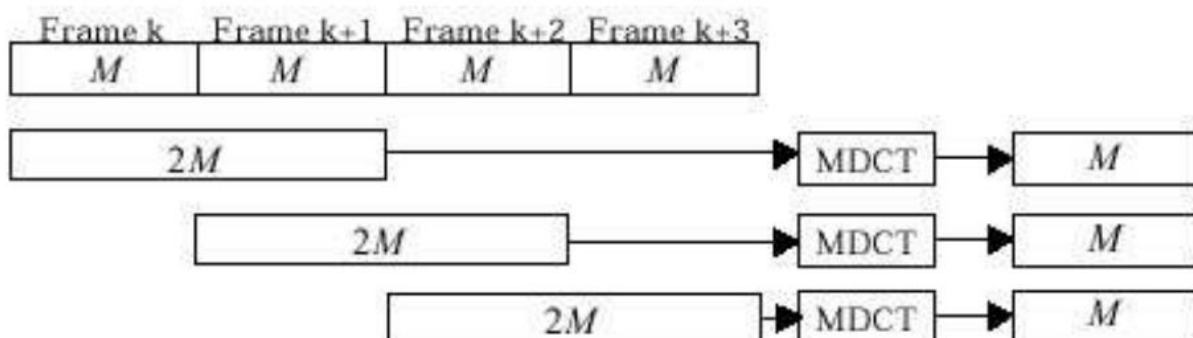


Bild 8, MDCT

2.5 TNS

Temporal Noise Shaping (TNS) ist ein Verfahren, um die Fehler, welche durch die MDCT entstehen können, weiter zu reduzieren. Wie weiter oben beschrieben, wählt die MDCT je nach Stetigkeit des Signals ein großes oder ein kleines Frame. Wenn das Signal jedoch weder eindeutig stetig, noch eindeutig unstetig verläuft, kann am Ende eines langen Frames noch ein unstetiger, impulsiver Signalverlauf folgen. Dies wird im transformierten Signal als ein Rauschen (Pre-Echo), welches unter Umständen hörbar ist, bemerkbar. Die Abstufungen werden immer zwischen dem Minimum und dem Maximum der Signalstärke eines Fensters gewählt und die ruhigen Signale werden daher, bei den erwähnten Umständen, zu grob abgestuft. Der Name Pre-Echo beschreibt das Störmuster, bei dem vor dem impulsiven Signalausbruch ein Rauschen entsteht (siehe Bild 9). Sozusagen ein sehr leises Echo vor dem Signal. Um dem entgegen zu wirken, werden beim TNS in ruhigen Abschnitten die Signale in einer hohen Auflösung kodiert, in impulsiven Abschnitten wird die Auflösung zurückgenommen. Damit wird das Rauschen auf den impulsiven Abschnitt verschoben, allerdings wird das Rauschen hier aufgrund des impulsiven Signals maskiert (siehe Frequenzmaskierung).

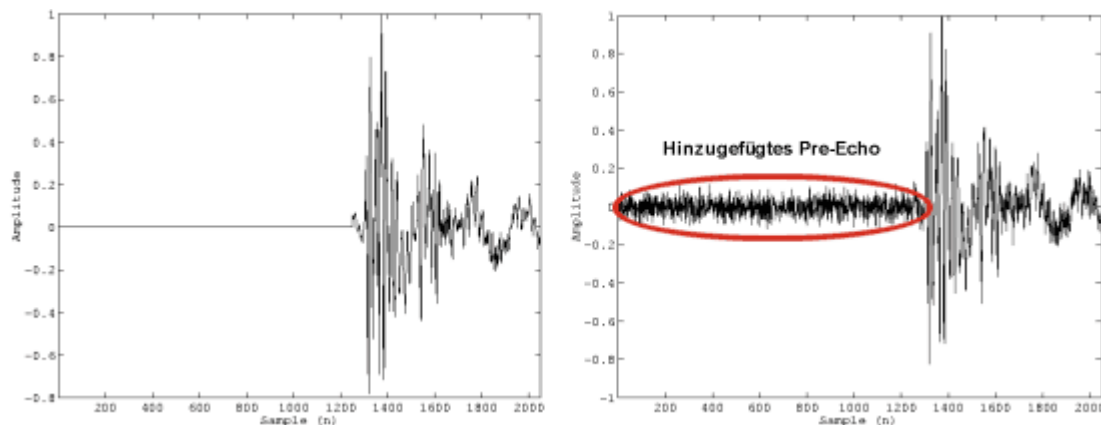


Bild 9, Pre-Echo

2.6 Gain Control

Beim Gain Control wird das analoge Signal in 4 gleich große Frequenzbereiche geteilt, auf die später unterschiedliche Abtastraten angewandt werden können. Zusätzlich werden die 4 Signale, falls nötig, durch Verstärkung in eine Form gebracht, welche weniger anfällig für den Pre-Echo-Effekt ist.

2.7 Intensity coding

Das Intensity coding vergleicht verschiedene Kanäle miteinander und speichert, falls möglich, hochfrequente Signale nur in einem Kanal. Der Mensch kann in hochfrequenten Bereichen kleine Unterschiede zwischen den Signalen nicht erkennen, daher reicht es, bei geringer Abweichung, einen aus den Signalen gewonnenen Mittelwert zu speichern und diesen auf allen beteiligten Kanälen wiederzugeben.

2.8 Prediction

Kommt es vor, dass eine Signalfolge sich periodisch wiederholt, so wird beim Prediction-Verfahren diese Signalfolge nur einmal gespeichert. Bei der Wiedergabe wird dieses eine Signal dann, entsprechend der Periode, immer wieder erneut abgespielt.

2.9 M/S

Das Mid/Side Stereo Verfahren (M/S) vergleicht 2 Kanäle miteinander und prüft, ob die Signale stark voneinander abweichen. Falls die Signale recht ähnlich sind, so werden die Summe und die Differenz der Signale gespeichert. Andernfalls werden die Signale beider Kanäle getrennt gespeichert. Da das Differenzsignal ähnlicher Signale sehr wenig Dynamik hat, belegt es weniger Speicher.

2.10 Variable Length coding - VLc

Wird eine Reihe von Werten gespeichert, so würde ohne Variable Length coding jeder Wert extra gespeichert. Wird Variable Length coding angewandt, so werden alle Werte bezüglich ihrer Häufigkeit gezählt und mit neuen Werten substituiert. Zum Beispiel würde das Wort „HALLO“ zu dieser Kodierung führen: 110 111 0 0 10

Dabei wurden die Buchstaben wie folgt substituiert:

H	-> 110
A	-> 111
L	-> 0
O	-> 10

Der am häufigsten vorkommende Buchstabe ist L und dieser wurde mit dem kleinsten Wert substituiert. Das Resultat der Kodierung ist 10 Bit lang. Ohne Kodierung hätten für jeden Buchstaben 5 Bit und damit insgesamt 20 Bit gewählt werden müssen, da viermal einer von 26 Großbuchstaben möglich wären und 26 in der Binär Darstellung 5 Bit (11010) benötigt.

Eine der Verbreitetsten VLc Verfahren ist die Huffman Kodierung.

3 Beschreibung des Verfahrens

AAC ist ein international definierter Standard, daher gibt es viele unterschiedliche, mit dem Standard konforme, Implementierungen. Zusätzlich existieren AAC-Profile, die AAC auf die benötigte Anwendung genauer zuschneiden. Auf diese Profile wird hier jedoch nicht genauer eingegangen. Die Beschreibung des Verfahrens ist in zwei Abschnitte geteilt. Der erste Abschnitt veranschaulicht den Fluss der Daten, so wie den Fluss der Kontrollparameter. Dabei werden bewusst einige Elemente ausgelassen um ein leichter zu überblickendes Modell zu erzeugen. Im zweiten Abschnitt wird, auf den ersten Abschnitt aufbauend, das Verfahren komplett erläutert. Der Fokus liegt auf der Kodierung des Signals. Da die Dekodierung sehr ähnlich invers abläuft wird sie am Schluss nur kurz erläutert.

3.1 Daten- und Kontrollfluss

Ein grober Kodierablauf lässt sich mit dem Diagramm aus Bild 10 veranschaulichen. Das analoge Signal wird abgetastet. Der Strom an Samples wird vom psychoakustischen Modell analysiert, um die Parameter der weiteren verarbeitenden Methoden zu ermitteln. Um das Signal aus der zeitdiskreten in die frequenzdiskrete Darstellung zu bringen, wird die MDCT auf den Samplestrom angewandt. Dabei liefert das psychoakustische Modell die Parameter, anhand deren die optimale Fenstergröße ermittelt wird. Da es trotz der Fensterüberlappungen der MDCT noch zu aliasing Effekten kommen kann, wird der Strom der nun frequenzdiskreten Samples durch das TNS-Verfahren dynamischer, differenzierter abgestuft. Die Parameter des psychoakustischen Modells bestimmen die optimale Auflösung der Abstufungen. Der durch die TNS bereinigte Strom an Samples wird nun quantisiert. Hier stellt das psychoakustische Modell die Parameter, anhand deren die minimal nötige Auflösung für die Quantisierung gewählt wird (die maximal nötige Auflösung beträgt in der Regel 16Bit). Das Signal liegt nun in einem Format vor, das in einen Bitstream konvertiert werden kann. Der Bitstream ist ein Strom von Bits, die von einem Decoder zurück in Audiosignale übersetzt werden können. Betrachtet man das ganze Verfahren genauer, so fehlen in dieser Darstellung einige Elemente.

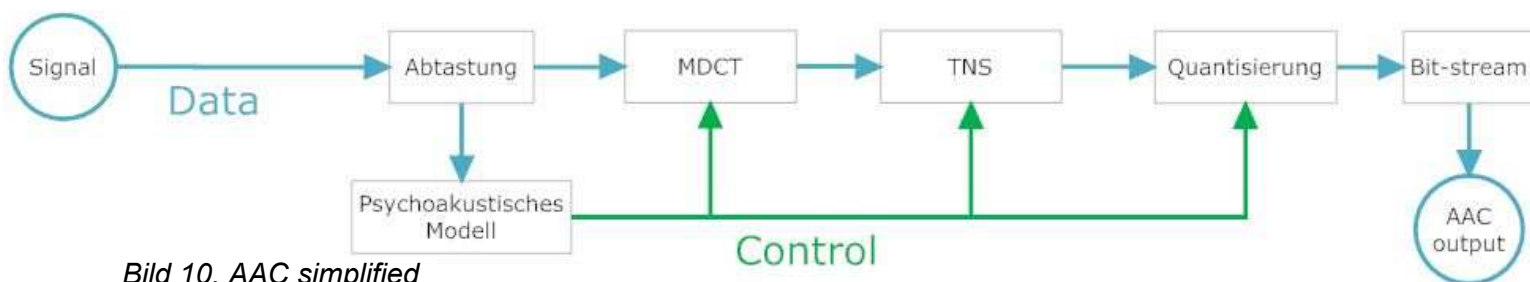


Bild 10, AAC simplified

3.2 Der Ablauf

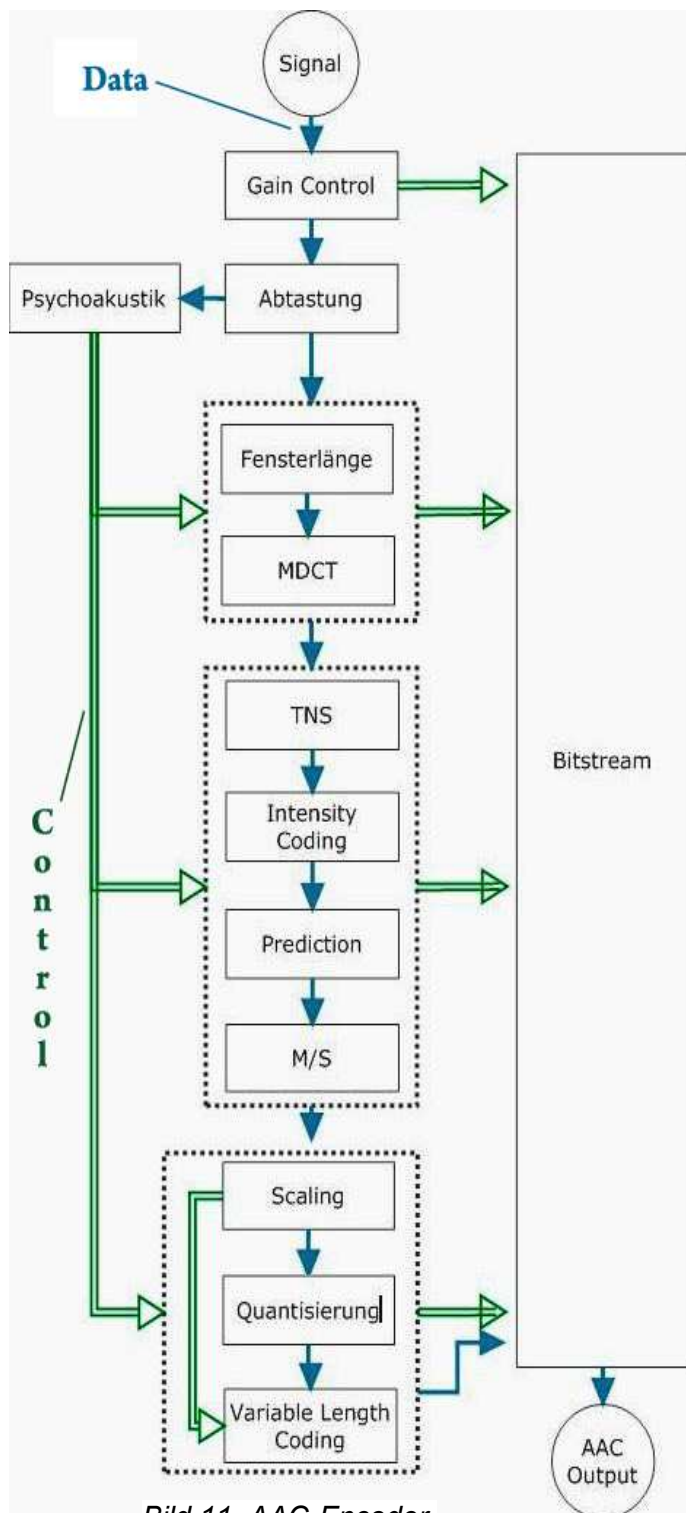


Bild 11, AAC-Encoder

- **Gain Control**

Das analoge Signal wird gemäß des Gain Control Verfahrens bearbeitet. Die Veränderungen am originalen Signal werden im Bitstream hinterlegt, um eine Rekonstruktion zu ermöglichen. Die 4 Frequenzbänder werden zur Abtastung übergeben.

- **Zeit-Frequenz Transformation**

Gemäß dem psychoakustischen Modell wird über die geeignete Fensterlänge für die MDCT entschieden. Diese Entscheidungen werden im Bitstream hinterlegt. Die MDCT transformiert nun alle zeitdiskreten Samples in die frequenzdiskrete Darstellung.

- **Frequenzdiskrete Bearbeitung**

Die von der MDCT bearbeiteten Daten werden durch das TNS zunächst vom Pre-Echo Effekt bereinigt bevor weitere Kompressionsschritte folgen. Das Intensity Coding entfernt hochfrequente Redundanzen. Die Prediction entfernt Redundanzen innerhalb eines Kanals. Das M/S-Verfahren entfernt nun die restlichen Redundanzen zwischen je zwei Kanälen. Die Schritte TNS und Intensity Coding erhalten Kontrollparameter des psychoakustischen Modells. Durch das Entfernen der Redundanzen ist die Datenmenge auf das Wesentliche reduziert worden und kann nun quantisiert werden.

- **Quantisierung**

Gemäß dem psychoakustischen Modell wird im Scaling die Wertetabelle mit den Abstufungen erzeugt. Nach der eigentlichen Quantisierung werden, anhand der im Scaling erzeugten Tabelle, die zur Darstellung der Werte nötigen Daten komprimiert. Hierzu wird das Variable Length coding verwendet. Im Bitstream werden die Werte hinterlegt, mit denen substituiert wurde, da ansonsten eine Rekonstruktion nicht möglich wäre.

- **Bitstream**

Abschließend werden alle Daten entsprechend den bisher übermittelten Parametern in eine Form gebracht, welche ein AAC-Dekoder verarbeiten kann. Dieser Bitstream wird in eine Datei geschrieben.

Bei der Dekodierung werden alle Schritte invers abgelaufen. Dabei entfallen die Methoden der Psychoakustik, da alle zur Rekonstruktion nötigen Daten im Bitstream hinterlegt sind.

4 Anwendungsgebiet

Da die Kompression, ähnlich dem MP3 Standard, Frequenzen vernachlässigt, welche sich außerhalb des menschlichen Hörvermögens befinden, eignet sich AAC besonders für Kompressionen von Musik, Audio in Videos etc.

Die Grenzen der menschlichen Wahrnehmung verhindern, dass die durch die Kompression entstandenen Fehler bemerkt werden.

4.1 Wann sollte AAC eingesetzt werden?

Wenn ein dekodiertes Signal nicht 100% dem Original Signal entsprechen muss bietet sich eine verlustbehaftete Kompression an. So ist AAC sehr geeignet die benötigte Bandbreite von Audiostreams gering zu halten und durch die reduzierte Dateigröße Up- und Downloadzeiten zu verkürzen. Gerade bei Bitraten von <128kbps übertrifft AAC seine Konkurrenten in Qualität und Kompressionsgrad.

4.2 Wann sollte AAC nicht eingesetzt werden?

Ist es von Nöten, dass das dekodierte Signal exakt dem Originalsignal entspricht, so ist eine Kompression mittels AAC oder anderen verlustbehafteten Kompressionsverfahren nicht sinnvoll. Für diese Zwecke gibt es verlustfreie Kompressionsverfahren, welche jedoch bei dem Kompressionsgrad deutliche Einbußen gegenüber AAC haben.

Zum Beispiel wäre eine Nutzung von AAC nicht sinnvoll wenn Frequenzbereiche außerhalb des menschlichen Hörvermögens ausgewertet werden müssen.

5 Qualitätsbewertung

Wann immer eine verlustbehaftete Audiokomprimierung nötig ist, so scheint zurzeit AAC die beste Wahl zu sein. Wenn für die Kompression die richtigen Parameter gewählt werden, so wird eine außerordentlich gute Kompression und Qualität erzielt. Die Gründe für die schleppend laufende Verbreitung des AAC-Standards liegen eindeutig nicht in dem Verfahren und daher ist anzunehmen, dass AAC in Zukunft das verbreitetste Audiokompressionsverfahren wird. Da sich viele der in AAC verwendeten Methoden bewährt haben, werden potentielle Weiterentwicklungen auf dem AAC-Standard basieren und daher vermutlich abwärtskompatibel sein. Dies befürwortet die Entscheidung moderne Multimediageräte MPEG-4- und damit auch AAC-Konform zu entwerfen, da ein ähnlicher Siegeszug wie bei MP3 zu erwarten ist.

Zusammenfassend lässt sich feststellen, dass AAC ein ausgezeichnetes Kompressionsverfahren mit viel Zukunft ist.

6 Glossar

3

3GPP

3rd Generation Partnership Project, eine weltweite Kooperation von Standardisierungsgremien für die Standardisierung im Mobilfunk

A

AAC

Advanced Audio Coding, Standard zur Kompression von Audiodaten

ARIB

Association of Radio Industries and Businesses, Standardisiert die Übermittlung von Funk- und Bildübertragungen Japanischer Medienkonzerne

Audiostream

Live-Übertragung von Audiodaten

B

Bandbreite

max. Menge an Informationen die pro Zeiteinheit übertragen werden kann

Bitrate

Menge an Informationen in Bit pro Zeit, siehe kbps

Blackbox

ein Objekt, dessen innere Funktionsweise unbekannt ist. Man kennt lediglich das Eingangs-, sowie Ausgangssignal

Bluetooth Special Interest Group

C

Chips

Integrierte Schaltkreise auf Halbleiterbasis

D

Decoder

erzeugt ein Audiosignal aus den komprimierten Daten

Digital Video Broadcasting

standardisierte Verfahren zur Übertragung von digitalen Inhalten (Fernsehen, Radio, Mehrkanalton, Raumklang, interaktive Dienste wie MHP, EPG und Teletext und weitere Zusatzdienste) durch digitale Technik.

DRM Consortium

Digital Radio Mondiale Consortium, Standardisiert den DRM30 Standard für weltweiten Digitalrundfunk

F

Frame

ein Abschnitt des Signals, bei dem die Positionen der enthaltenen Signalverläufe in Relation zum Frame und nicht zum gesamten Signal bezogen werden.

I

IEC

International Electrotechnical Commission, eine internationale Normungsorganisation für Normen im Bereich der Elektrotechnik und Elektronik. Einige Normen werden gemeinsam mit ISO entwickelt. (siehe ISO)

Internet Streaming Alliance

eine internationale Vereinigung zur Integration von Multimedia IPTV Standards

ISO

International Organization for Standardization, eine internationale Vereinigung von Normungsorganisationen und erarbeitet internationale Normen in allen Bereichen mit Ausnahme der Elektrik und der Elektronik

K

kbps

Kilobyte, die zum Speichern einer Sekunde belegt werden

Kompressionsgrad

beschreibt den Faktor um den eine Datei verkleinert werden kann

M

MP3

MPEG-1 Layer3, Vorgänger von AAC, siehe AAC

MPEG

Moving Picture Experts Group, ist eine Arbeitsgruppe, die sich mit der Standardisierung von Videokompression und den dazugehörigen Bereichen, wie Audiodatenkompression oder Containerformaten, beschäftigt. Die zusammengeführten Standards wurden MPEG-1, MPEG-2 etc. benannt.

P

Plug-In

Funktionserweiterung eines Programms

X

XM Radio

ein kommerzieller Anbieter von Satellitenradio in Nordamerika

7 Literatur / Referenzen

[vuo] The Anh Vuong, Proseminar Vorlage, Universität Frankfurt , 2011

Brandenburg, Karlheinz, „MP3 and AAC explained“, Fraunhofer IIS, 1999

ISO13818-3 - International Organization for Standardization, 1998

Jsc digital media, AAC Audio and the MP4 Media Format [Stand: 30.06.2011]
<http://www.jiscdigitalmedia.ac.uk/audio/advice/aac-audio-and-the-mp4-media-format/>

Wikipedia AAC [Stand: 30.06.2011]
http://de.wikipedia.org/wiki/Advanced_Audio_Coding

Hydrogen Audio - AAC [Stand: 30.06.2011]
<http://wiki.hydrogenaudio.org/index.php?title=AAC>

Die Psychoakustik – Universität Wuppertal [Bild 1] [Stand: 30.06.2011]
http://www.dasp.uni-wuppertal.de/ars_auditus/psychoak/psychoak0.htm

Akademic.ru [Bild 2] [Stand: 30.06.2011]
http://de.academic.ru/pictures/dewiki/65/Akustik_Mithoerschwelle2.JPG

EncycloSPACE.org [Bild 3] [Stand: 30.06.2011]
http://www.encycloSPACE.org/Musik_informatik_1/Bilder/bild58b3.jpg

Laermorama.ch [Bild 4] [Stand: 30.06.2011]
http://www.laermorama.ch/laermorama/modul_hoeren/bilder/Maskierung_Zeit.gif

Uni Hamburg [Bild 5, 6] [Stand: 30.06.2011]
http://www1.uni-hamburg.de/DigiLit/hiebler/03_analog_digital.html

Jsc digital media [Bild 7] [Stand: 30.06.2011]
<http://www.jiscdigitalmedia.ac.uk/audio/advice/an-introduction-to-digital-audio>

Joebert S. Jacaba, „Audio Compression using MDCT“, University Philippine, 2001 [Bild 8]

Mp3Encoding.de [Bild 9] [Stand: 30.06.2011]
http://mp3encoding.de/index_10.html

Schaude, Jonas & Beuter, Florian, „MP3 und AAC – wie funktioniert das?“, Universität Ulm, 2011 [Tabelle 1]