

Universität
Rostock



Traditio et Innovatio

Projektarbeit zum Thema

"Formate zur Speicherung Digitaler Audiodaten"

zum Abschluß des Studienkurses

Audio & Videotechnik

Betreuer:

Prof. Dr.-Ing. habil. Rainer Kohlschmidt

vorgelegt dem

Institut für Nachrichtentechnik

der

Universität Rostock

Maik Gotzmann

21. März 2010

Inhaltsverzeichnis

1	Einleitung	1
1.1	Motivation	1
1.2	Aufbau der Projektarbeit	2
2	Formate, zur verlustfreien und unkromprimierten Speicherung	3
2.1	Grundlagen	3
2.2	Das VOC Format	4
2.3	Das WAV Format	8
2.4	Weitere Formate	10
3	Formate, zur verlustfreien und kromprimierten Speicherung	11
3.1	Grundlagen	11
3.2	Das FLAC Format	11
3.3	Weitere Formate	14
4	Formate zur verlustbehafteten und kromprimierten Speicherung	16
4.1	Grundlagen	16
4.2	Das MP3 Format	19
4.3	Das OGG Vorbis Format	22
4.4	Das WMA Format	23
4.5	Weitere Formate	24
5	Zusammenfassung und Ausblick	26
6	Erklärung über die selbständige Anfertigung der Arbeit	28

Abbildungsverzeichnis

1.1	Klassifizierung von Audio-Formaten	2
2.1	Waveform einer Audio-Datei	4
2.2	Bringt dem PC Töne bei: Soundblaster 2.0 von Creative Technologies .	5
2.3	Exemplarischer Aufbau eines VOC Files	6
2.4	Exemplarischer Aufbau eines WAV Files	9
3.1	Mid-Side-Coding	12
4.1	Wahrnehmbarer Frequenzbereich	17
4.2	Wirkungsweise von Maskierungseffekten	17
4.3	Der erste tragbare MP3-Player	21

Tabellenverzeichnis

2.1	Datenblocktypen des VOC Formates	7
2.2	Codecs des VOC Formates	8
2.3	Auszug der Codecliste des WAV Formates	8
4.1	Typische Bitraten für MPEG-1 Audiostreams	19
4.2	Meta-Datenblock nach ID3v1	20

Abkürzungsverzeichnis

ASF	:	Advanced Systems Format
ASPEC	:	Adaptive Spectral Perceptual Entropy Coding
CRC	:	Cyclic Redundancy Check
DAB	:	Digital Audio Broadcasting
DRM	:	Digital Rights Management
DSP	:	Digital Signal Processor
MDCT	:	Modified Discrete Cosine Transformation
MIDI	:	Musical Instrument Digital Interface
MPEG	:	Moving Picture Experts Group
PCM	:	Pulse Code Modulation
RIFF	:	Resource Interchange File Format

1 Einleitung

1.1 Motivation

Dateiformate dienen der Speicherung von bestimmten Daten in einer spezifizierten Reihenfolge, dem Format. Sie stellen damit eine bidirektionale Abbildung der Informationen zu einem eindimensionalen binären Speicher dar. Damit die innerhalb einer Datei gespeicherten Informationen von entsprechenden Anwendungen interpretiert werden können, ist die Kenntnis über das Format von essentieller Bedeutung.

Bereits Ende der 50er Jahre des letzten Jahrhunderts wurden die ersten Formate definiert, die der Speicherung von Audiodaten dienten, die im Zuge einer Analog/Digital-Wandlung von niederfrequenten Schwingungen entstanden. Ein grundlegendes Problem dieser Zeit war allerdings die hohe Datenrate, die bei der Aufzeichnung von Tönen auftritt. Die anfänglich nur im professionellen Umfeld, wie Tonstudios und Radiosender, eingesetzten digitalen Speicherformen erlebten mit der Weiterentwicklung der Computertechnik und der Einführung leistungsfähiger, und vor allem soundfähiger Heimcomputer einen regelrechten Boom, der zur Entwicklung einer Vielzahl unterschiedlicher Formate führte, die von nun an der Untermalung von Computerspielen, zur Archivierung eigener Musikarchive oder als Bestandteil von Animationen und Videos dienen. Der Personal Computer als multimedialer Alleskönner tritt dabei seit Anfang der 90er Jahre das Erbe von teurer Spezialhardware im Tonbereich als auch das, der privaten Arbeits- und Unterhaltungsmaschine an, der sich von nun an um die Verarbeitung und Speicherung qualitativ hochwertiger Audiodaten kümmert.

Bei der Betrachtung von Audiodaten muss dabei grundlegend eine Unterscheidung zwischen dem Dateiformat und dem verwendeten Verfahren zur Codierung der Audiodaten (Codec) vorgenommen werden. Bei der Konzeption eines Dateiformates erfolgt zwar meist eine Fixierung auf eine bestimmte Codierungsform, aber im Laufe der Weiterentwicklung und Softwarepflege kommen weitere hinzu, welches zu den Containerformaten führt. In ihnen können Daten in unterschiedlicher Codierungsform vorliegen, eine eindeutige Identifizierung allein anhand der Dateiendung ist dabei aber nicht mehr gegeben, so geschehen zum Beispiel bei Dateien mit der Endung `.wav`, `.mp4` oder `.ogg`.

Diese Projektarbeit soll daher eine kurze, aber keineswegs vollständige Übersicht über heute noch gebräuchliche Dateiformate zur Speicherung von Audiodaten geben, mit dem Ziel, ihre Entstehung, Struktur, Eigenschaften und Verwendung zu durchleuchten.

1.2 Aufbau der Projektarbeit

Die in dieser Projektarbeit behandelten Dateiformate zur Speicherung von Audiodaten werden in drei Klassen gem. der Abbildung 1.1 eingeteilt. Jeder Klasse ist ein eigenes Kapitel gewidmet, welches eine oder mehrere populäre Vertreter dieser etwas näher beschreibt. Auf einen allgemeinen Grundlagenteil, der Kenngrößen und Eigenschaften der Signalverarbeitung und Nachrichtentechnik beschreibt, wird hier bewusst verzichtet. Erklärungen, die über das allgemeine Grundverständnis zur Bearbeitung von Audiodaten hinausgehen, werden in den entsprechenden Kapiteln unter dem Abschnitt „Grundlagen“ gegeben, wenn auch nur auf eher oberflächlicher Art. Ein paar Vertreter der weniger gebräuchlichen Dateiformate werden im jeweiligen Kapitel unter „Weitere Formate“ in aller Kürze beschrieben.

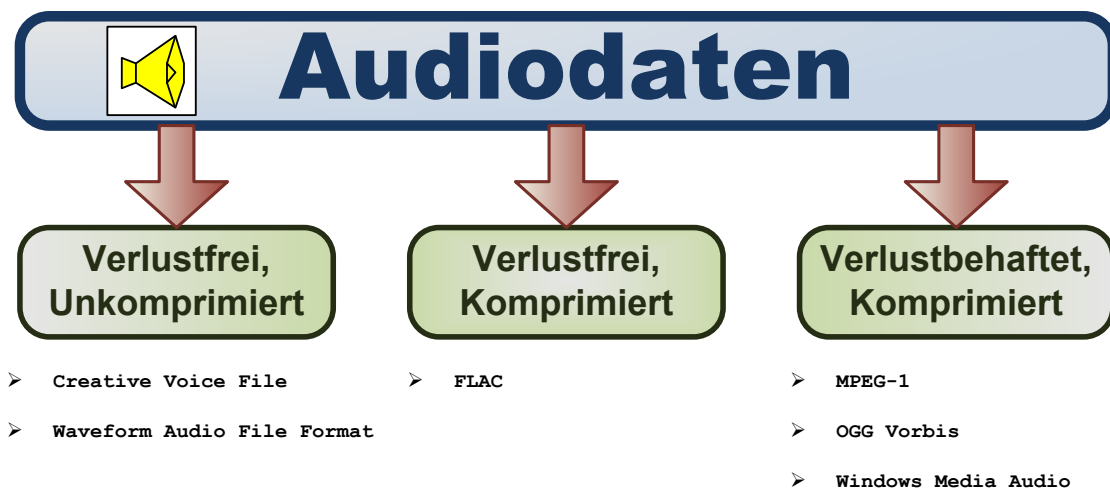


Abbildung 1.1: Klassifizierung von Audio-Formaten

2 Formate, zur verlustfreien und unkromprimierten Speicherung

2.1 Grundlagen

Audioformate, die nach dem Verfahren der verlustfreien und unkromprimierten Speicherung von Audiodaten arbeiten, enthalten als Nutzdaten die PCM (Pulse Code Modulation)-Rohdaten, die durch die Abtastung während einer Analog/Digitalwandlung entstehen. D.h., sie enthalten die Werte (Samples), die eine Auslenkung des Audiosignals zu einem bestimmten Zeitpunkt darstellen. Da eine vorangegangene A/D-Wandlung im festen Zeitintervall, also zeitdiskret, stattfindet, ist es ausreichend, aufeinanderfolgend die quantisierten Werte zu speichern. Mit dem Wissen des Wertebereiches und der Abtastrate ist damit eine Weiterverarbeitung oder eine direkte Ausgabe des Audiosignals mittels Digital/Analog-Wandlung möglich. Kennzeichnend für ein Audio-File, welches verlustfrei und unkromprimiert gespeichert wurde, ist also lediglich die Abtastfrequenz f_T und die Anzahl der maximalen Quantisierungsstufen (Bit-Länge eines Samples). Beide Größen können auch als Maß für die Qualität des so digital gespeicherten Audiosignals aufgefasst werden, die allerdings nicht zuletzt auch das notwendige Datenvolumen bestimmen. Die resultierende Dateilänge ergibt sich dann nach folgender Formel:

$$\text{Dateigröße} = \text{Abtastfrequenz/Hz} * \text{Samplegröße/Byte} * \text{Dauer/s} * \text{Kanäle} \quad (2.1)$$

Mit Hilfe eines Beispiels wird die erforderliche Datenmenge zur Speicherung eines Audio-Files etwas deutlicher. Angenommen sei hier eine Abtastfrequenz von $f_T = 22050 \text{ Hz}$, 256 Quantisierungsstufen, welches einer Samplegröße von genau 8 Bit = 1 Byte entspräche, die Länge des Audiosignals betrage $1\text{min} = 60\text{s}$ und es möge sich um ein Stereo-Signal handeln, es gibt also 2 Kanäle:

$$\text{Dateigröße} = 220500\text{Hz} * 1\text{Byte} * 60\text{s} * 2 = 2646000\text{Bytes} = 2584\text{KB} = 2,5\text{MB} \quad (2.2)$$

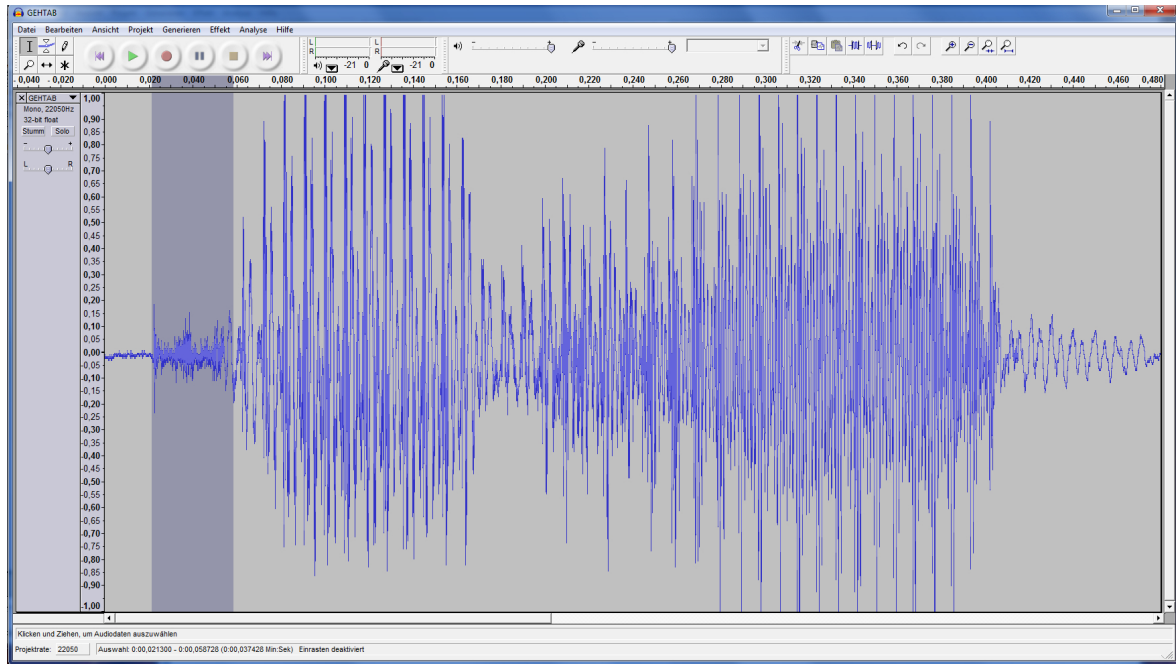


Abbildung 2.1: Waveform einer Audio-Datei, zur Bearbeitung geöffnet mit dem Open-Source Tool Audacity

Dabei wird ersichtlich, dass die erforderliche Datenmenge schon recht beträchtlich ist. Auf einer HD-Diskette mit einer Gesamtkapazität von 1,44 MB fände dieses Audiofile bereits keinen Platz mehr, und dass bei einer Qualität, die nur ein viertel der einer Audio-CD ($f_T = 44100\text{Hz}$, 16Bit Samplelänge) beträgt. Daher eignet sich die Speicherung von Audio-Files in dieser Formatform eher nur zum Zwischenspeichern während der Bearbeitung. Die Archivierung von Audiodaten verbraucht dabei zu viel Speicherplatz.

2.2 Das VOC Format

Bis zum Ende der 80er Jahre des letzten Jahrhunderts führte der Personal Computer, welcher bereits 1981 durch die Firma IBM vorgestellt und vertrieben wurde, ein Schattendasein im Grafik- und Soundbereich. Als reine Arbeitsmaschine konzipiert, blieb ihm aufgrund nur durchschnittlicher Rechenleistung, fehlender Grafik- und Soundfähigkeit, dem mangelndem Bedienkomfort und dem hohen Anschaffungspreis der Einsatz als Privatrechner im Heimgebrauch verwehrt. Die Mitte der 80er Jahre erscheinende neue Generation von 16-Bit Heimcomputern, beispielsweise aus den Häusern Atari und Commodore Amiga, feierten hingegen den Einzug in Millionen von privaten Haushalten als neues Unterhaltungs- und Arbeitsmedium. Ihren Siegeszug verdankten sie nicht nur allein der auf 16/32-Bit basierten Rechnerplattform, den günstigen Anschaffungspreis,

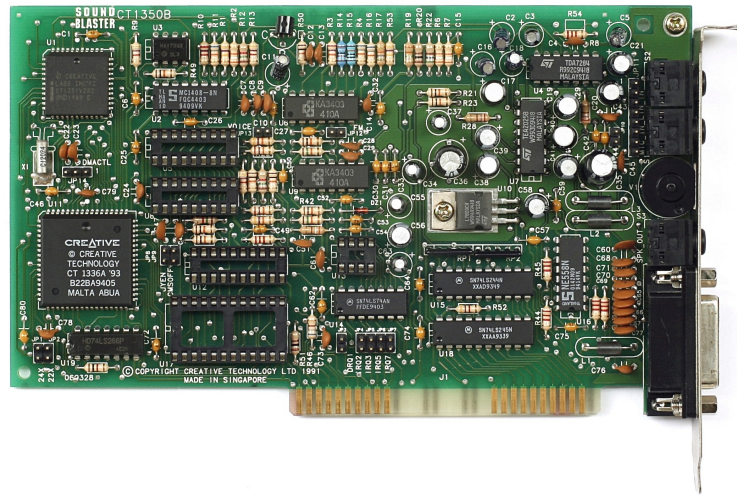


Abbildung 2.2: Bringt dem PC Töne bei: Soundblaster 2.0 von Creative Technologies

der grafischen Benutzeroberfläche mit Mausbedienung, sondern auch der hörenswerten Soundqualität, die diese Rechner von Haus aus mitbrachten. So sind beispielsweise die Rechner der Atari ST Familie standardmäßig mit einer MIDI-Schnittstelle (Musical Instrument Digital Interface – Schnittstelle zur Ansteuerung digitaler Musikinstrumente wie Synthesizer, Drumcomputer etc.) ausgestattet, was ihnen eine weite Verbreitung in Tonstudios bescherte. Der Commodore Amiga 500 enthielt als erster Heimcomputer einen DSP (Digital Signal Processor) zur Aufzeichnung und Wiedergabe von digitalen Audiofiles. Andere Computersysteme der Firmen Apple und Acorn wiesen ähnliche Fähigkeiten auf. Um dem PC der damaligen Zeit Töne zu entlocken, musste man sich des intern verbauten Systemlautsprechers bedienen. Dabei kann von Klang keine Rede sein, es handelt sich in diesem Falle eher um eine Aneinanderreihung von nervigen Piepstönen. Abhilfe schafften die erstmals 1987 auftauchenden Erweiterungskarten der Firmen Roland und Adlib, die dem PC mit einer Soundkarte bereicherten und den PC um die Soundfähigkeit erweiterten. Der Durchbruch blieb beiden Anbietern an Soundkarten allerdings verwehrt, da eine Roland-Karte soviel kostete, wie der gesamte PC selbst, und Erweiterungskarten der Firma Adlib beschränkten sich auf den Verbau von eher billigen Synthesizer-Chips, welche dem PC damit zum Klangbild einer billigen Heimorgel verhalfen. Die Möglichkeit zur digitalen Aufzeichnung bzw. Wiedergabe von Samples war mit der ersten Generation der Adlib-Soundkarte nicht möglich. Erst 1989 kam mit dem Erscheinen der Soundblaster 2.0 Erweiterungskarte der späte, aber rettende Durchbruch, um den PC im Audibereich so einigermaßen konkurrenzfähig gegenüber den Heimcomputern zu machen.

Obwohl diese, von der Firma Creative Technologies (sie existiert noch heute unter dem Namen Creative Labs) entwickelte Soundkarte lediglich Samples mit einer Bittiefe von 8 Bit verarbeiten kann und sich auf einen einzigen Kanal (Mono) beschränkt, verhalf sie

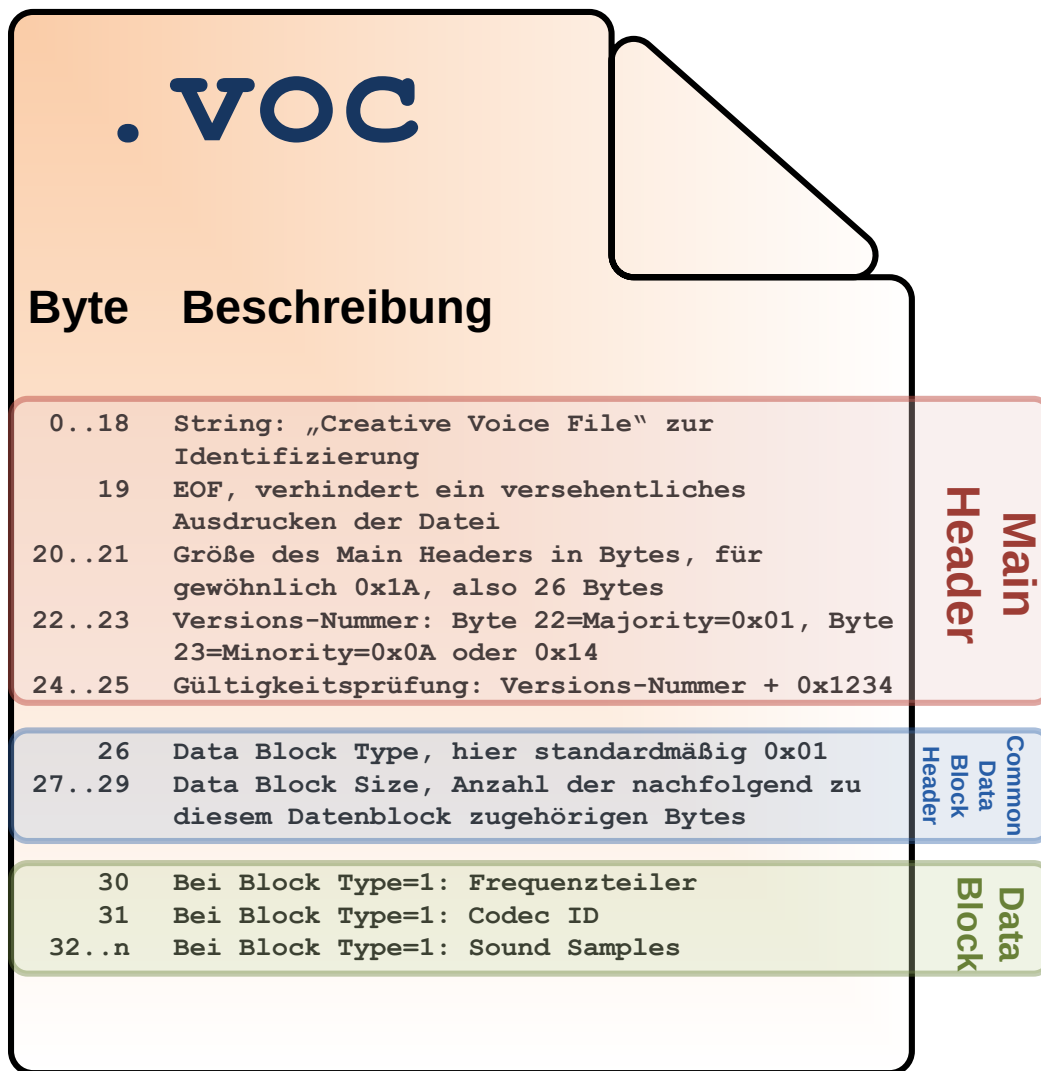


Abbildung 2.3: Exemplarischer Aufbau eines VOC Files

dem PC zum Durchbruch als multimedialen Alleskönner. Denn sie war relativ preiswert in der Anschaffung (Kaufpreis 1990 ca. 200 DM) und verfügte über einen Line-In Eingang mit A/D-Wandler. Fördernd kam hinzu, dass massiv preiswerte PC-Nachbauten den Markt überfluteten und die gut ein Dutzend unterschiedlichen und untereinander inkompatiblen Heimcomputersysteme vom Markt verdrängten. Produzenten von Spiel- und Lernsoftware sind nun plötzlich an einer Umsetzung ihrer Software für den PC interessiert. Somit etablieren sich die Soundblaster-Soundkarten als quasi Standard für die PC-Systeme. Auch wenn in diesen bewegten Zeiten weitere Hersteller, wie beispielsweise Gravis, mit neuen und besseren Soundkarten den Markt betreten, können die sich aufgrund der fehlenden Kompatibilität nicht behaupten. Einhergehend mit diesem Hardware-Standard für Soundkarten, etabliert sich auch ein Dateiformat zur Speicherung von digitalen Audiodaten als quasi Standard auf MS-DOS basierten Rechnern: Creative Voice File mit der Dateiendung `.voc`.

Data Block Type	Bedeutung
0	Terminator , signalisiert das Ende der Datei
1	Sound Data , eigenständiger Datenblock mit Samples
2	Sound Data Continuation , Datenblock mit Samples, Codec und Abtastfrequenz werden vom zuvor wiedergegebenen Datenblock übernommen
3	Silence , Datenblock enthält keine Samples, nur Angaben über Länge der Stille (2 Bytes in Samples) und Frequenzteiler
4	Marker , vorgesehen zur Synchronisation mit Animationen
5	Text , String Daten
6	Repeat Start , der nachfolgende Datenblock wird x-mal wiederholt, wobei $x = 0xFFFF$ eine Endlosschleife bedeutet
7	Repeat End , leerer Datenblock, der eine wiederholte Wiedergabe wieder aufhebt
8	Extra Info , ersetzt die Eigenschaften des vorangegangenen Datenblocks mit neuen Werten zur Frequenz und Codec
9	Sound Data (new Format), ersetzt den Standard Blocktyp 1 und erweitert deren Eigenschaften, erst ab V1.20 verfügbar, um die Fähigkeiten der Soundblaster16 zu unterstützen

Tabelle 2.1: Datenblocktypen des VOC Formates

Jedes VOC File beginnt dabei mit einem Mainheader, der alle notwendigen Informationen zur Identifizierung dieser Datei enthält [6]. Darauf folgen einer oder mehrere Datenblöcke, welche die eigentlichen Samples enthalten. Somit ist es auch möglich, innerhalb eines VOC Files mehrere, in ihren Eigenschaften völlig unterschiedliche, Audioaufnahmen zu speichern. Dazu enthält jeder Datenblock einen 4 Byte großen allgemeinen Data Block Header, der Angaben zur Art des Datenblocks und seine Länge, abzüglich dieser 4 Bytes, enthält. Eine Übersicht der möglichen Blocktypen kann der Tabelle 2.1 entnommen werden. Bei Verwendung der Blocktypes 1 oder 9 folgen nun noch Informationen zur Abtastfrequenz und zum Format, in dem die Samples vorliegen (Codec Information gem. Tabelle 2.2). Abbildung 2.3 zeigt exemplarisch den Aufbau eines VOC Files für den Datenblocktyp 1.

Der Wert des Frequenzteilers lässt sich dabei wie folgt bestimmen:

$$\text{Frequenzteiler} = 256 - \frac{1000000}{\text{reale Abtastrate in Hz}} \quad (2.3)$$

Auch wenn das VOC Format heutzutage als obsolet gelten möge, so ist es doch das erste standardisierte Dateiformat, das sich auf dem PC etablierte. Professionelle Software zur Bearbeitung von Audiofiles, wie zum Beispiel Cool Edit Pro, unterstützen dieses Format nachwievor.

Codec ID	Bedeutung
0	8 Bit PCM, vorzeichenlos
1	4 zu 8 Bit Creative ADPCM, in Hardware implementierte Komprimierung auf Soundblaster Karten
2	3 zu 8 Bit Creative ADPCM
3	2 zu 8 Bit Creative ADPCM
4	16 Bit PCM, vorzeichenbehaftet
6	A-Law
7	μ -Law
512	4 zu 16 Bit Creative ADPCM, nur in Verbindung mit Data Block Type 9

Tabelle 2.2: Codecs des VOC Formates

2.3 Das WAV Format

Mit dem Siegeszug von Microsofts Windows als Standard-Betriebssystem für PCs, etablierte sich auch ein neues Format zur Speicherung von Audiodaten als Standard. Das „Waveform Audio File Format“, kurz Wave oder WAV genannt, wurde von Microsoft und IBM entwickelt und wird heute selbst auf UNIX- Rechnern und Computern der Firma Apple benutzt.

Formattag	Codec
0x0001	PCM
0x0002	MS ADPCM
0x0006	A-Law
0x0007	μ -Law
0x0055	MPEG-1 Layer III (MP3)
0xFFFF	WAVE Format Erweiterung

Tabelle 2.3: Auszug der Codecliste des WAV Formates

Das WAV Format ist ein Containerformat und setzt auf die RIFF-Dateistruktur (Resource Interchange File Format) auf. Eine RIFF-Datei besteht dabei aus mehreren ineinander gekapselten Abschnitten (engl.: chunks). Für das WAV Format sind mindestens drei Abschnitte verpflichtend: An erster Stelle steht der Master RIFF chunk, der die Datei als RIFF-Datei identifiziert, die Gesamtlänge der Datei in Bytes angibt und sie als WAV Datei spezifiziert. Es folgt nun der Formatabschnitt, eingeleitet mit der chunkID=“fmt“. Dieser enthält nun Angaben, in welchem Format die gespeicherten Audiodaten vorliegen. Ihm kann man unter anderem die Abtastfrequenz, Bittiefe eines Samples, Anzahl der Kanäle und Aufzeichnungsart entnehmen. Eingeleitet mit chunkID=“data“ folgt nun der Datenabschnitt, der die Samples enthält. Ist die Gesamtanzahl der Samples ungerade, wird dem Datenabschnitt noch ein leeres Byte

(Paddingbyte) hinzugefügt, so dass die Abschnittlänge wieder eine gerade Zahl in Bytes ergibt. Exemplarisch kann der Aufbau einer WAV Datei mit PCM Daten der Abbildung 2.4 entnommen werden.

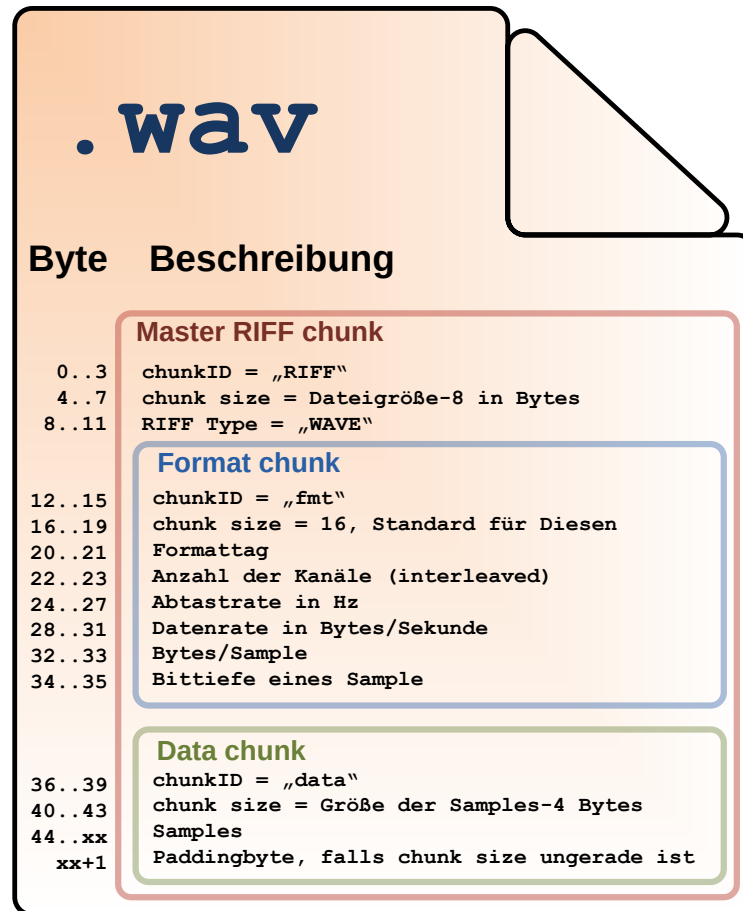


Abbildung 2.4: Exemplarischer Aufbau eines WAV Files

Mit dem WAV Format lassen sich aber nicht nur reine PCM-Daten speichern, sondern es steht eine Vielzahl unterschiedlicher Kompressionsmethoden zur Verfügung. Eine entsprechende Auswahl dazu wird im Formatabschnitt mittels der Angabe zum Formattag getätigt. Tabelle 2.3 enthält einen Auszug verwendbarer PCM und nicht-PCM Formate. Bei Verwendung eines der Letzgenannten muss ein weiterer Abschnitt mit der chunkID=“fact“ hinzugefügt werden. Um das WAV Format auch für zukünftige Kompressionsmethoden aktuell zu halten, kann mit Setzen des Formattags auf 0xFFFE ein eigenes Format definiert werden.

Durch die einfache Handhabung dieses Dateiformates für PCM Daten und den flexiblen Einstellmöglichkeiten des Formates betreffend, findet das WAV Format eine weite Verbreitung und das nicht nur um Audiodaten zu speichern. Das CAD Programm LTSpice verwendet beispielsweise das WAV Format zur Speicherung von Waveforms, die im Zuge der Schaltungssimulation entstehen.

2.4 Weitere Formate

Jedes neu erscheinende Computersystem/Betriebssystem, welches in der Lage war, Audiodateien mit PCM Inhalten zu verarbeiten, brachte seiner Zeit ein eigenes Audioformat mit. Da die Liste Derer beinahe unendlich lang wäre, seien im Folgendem nur ein paar Wenige erwähnt, die einem vielleicht hier oder dort noch einmal über dem Weg laufen könnten [1]:

IFF (Interchange File Format) Dieses Containerformat wurde 1985 von dem Spielgiganten Electronic Arts in Zusammenarbeit mit Commodore Amiga entwickelt, um den Datenaustausch zwischen den verschiedenen, in den 80er Jahren weit verbreiteten Heimcomputersystemen sicherzustellen. Da in diesem Format auch Bilder (ILBM) gespeichert werden, hat sich der Begriff **IFF-8SVX** bzw. **IFF-16SVX** geprägt, der die entsprechenden Dateien als Audiodateien näher spezifizieren. Das IFF Format stand Pate für das RIFF, und damit auch für das WAV Format und wurde in seinen strukturellen Grundzügen von Microsoft übernommen.

AIFF (Audio Interchange File Format) Dieses Format wurde 1988 in Kooperation mit Apple entwickelt und basiert strukturell auf das IFF Format und stellt dessen Weiterentwicklung dar. Es wurde erstmals auf den Macintosh Rechnern der Firma Apple eingesetzt.

AU Dieses eher schlichte Format wurde erstmals von der Firma Sun Microsystems eingeführt und war auch auf Rechnern der Firma NeXT weit verbreitet, obwohl die Rechner selbst es nicht waren.

3 Formate, zur verlustfreien und kromprimierten Speicherung

3.1 Grundlagen

Wie in dem bereits vorangegangenen Kapitel deutlich wird, ist die verlustfreie und unkomprimierte Speicherung von Audiodaten enorm speicheraufwendig. Um aber über Generationen hinweg derartige Daten digital zu archivieren, ist eine verlustfreie Speicherung unabdingbar. Um den Verbrauch von Festwertspeicher zu minimieren, liegt es nahe, bekannte Datenkompressionsverfahren wie ZIP oder RAR zu verwenden. In der Praxis stellte sich jedoch heraus, dass die Effizienz der Komprimierung von Audiodaten mit diesen Verfahren stark zu wünschen übrig lässt. Kein Wunder, handelt es sich im Falle von Audiodaten doch um quantisierte zeitdiskrete Werte einer Schwingung. Vorangegangener, aktueller und nachfolgender Wert eines Samples sind in den meisten Fällen grundverschieden. Herkömmliche Kompressionsverfahren arbeiten aber nach dem Prinzip des Zusammenfassens und Kodierung gleicher Werte. Daher werden in den letzten Jahren große Anstrengungen unternommen, Komprimierungsalgorithmen zu finden, die, bezogen auf Audiodaten, eine höhere Effizienz aufweisen. Ein Format, welches zur Speicherung von Audiodaten eine gute Effizienz aufweist, ist das FLAC (Free Lossless Audio Codec) Format.

3.2 Das FLAC Format

Die Entwicklung des FLAC Formates begann im Jahre 2000. 2003 wurde dieses Format in das OGG Containerformat für Audiodaten durch die Xiph.Org Foundation integriert, wodurch es eine weite Verbreitung durch viele unterstützende Player fand [2]. Zum Erfolg dieses Formates trägt auch die Tatsache bei, dass es keinen patentrechtlichen Verpflichtungen unterliegt. Der Codec selbst ist Open Source, und damit frei verfügbar und verwendbar. Darüberhinaus ist er auf vielen Betriebs- und Computersystemen anwendbar, egal ob MS-DOS, MS-Windows, Unix, Linux oder gar AmigaOS.

Die Komprimierung einer Audiodatei, in der bereits die PCM Daten vorliegen, erfolgt dabei in den nachfolgend aufgeführten fünf Stufen:

Blocking Im ersten Schritt werden die vorliegenden Samples in Blöcke unterteilt. Dabei sollten je Block 2 bis 6 tausend Samples enthalten sein. In der Standardeinstellung werden 4096 Samples einem Block zugeordnet, wobei die Blockgröße innerhalb der Audiodatei variieren darf.

Interchannel Decorrelation In den meisten Fällen dürfte es in den zur Komprimierung vorliegenden Audiodaten um Stereo-Signale handeln, sie enthalten also 2 Kanäle. Und in den meisten dieser Fälle, ist der rechte Audiokanal dem Linken recht ähnlich. Somit wird nun aus beiden separaten Kanälen ein Mittelkanal, welcher zu dem jeweiligen Zeitpunkt den Mittelwert beider Kanäle darstellt, berechnet. Die Abweichung der originalen Stereokanäle von diesem Mittelwert wird in einem weiteren Kanal, dem Seitenkanal gespeichert, um sie im Zuge der Decodierung vollständig rekonstruieren zu können. Abbildung 3.1 zeigt die Mid-Side-Codierung anhand eines in Waveform vorliegenden Stereosignals. Die Berechnung erfolgt anhand folgender Formeln:

$$\text{Mittelkanal} = \frac{\text{Linker Kanal} + \text{Rechter Kanal}}{2} \quad (3.1)$$

$$\text{Seitenkanal} = \text{Linker Kanal} - \text{Rechter Kanal} \quad (3.2)$$

Eine Codierung auf Basis des linken/rechten Kanals statt des Mittelkanals zzgl. Seitenkanals ist dabei separat für jeden Block ebenso möglich.

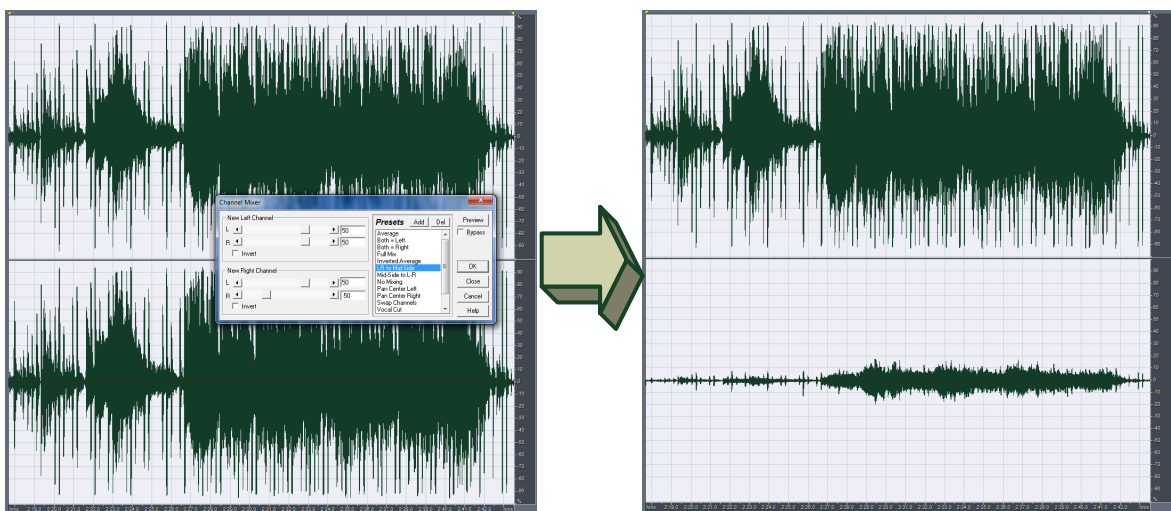


Abbildung 3.1: Mid-Side-Coding: aus einem Stereosignal (links) entstehen Mittelkanal (rechts oben) und Seitenkanal (rechts unten)

Prediction Nun erfolgt die Approximation des jeweiligen Audiokanals: der Verlauf

des Signals wird als eine Funktion beschrieben, in der versucht wird, sich dem Verlauf der tatsächlichen Schwingung anzugleichen. Dies kann zum einen durch Beschreibung der Schwingung mittels eines Polynoms geschehen, welches weniger rechenintensiv, aber fehleranfälliger ist, oder aber mit dem Verfahren des Linear Predictive Coding. Der Vorteil dieses Vorgehens liegt dabei auf der Hand: anstatt Werte jedes einzelnen Samples speichern zu müssen, genügt die Speicherung deutlich weniger Parameter für die Vorhersage des Signalverlaufs. Für eine verlustfreie Rekonstruktion ist es aber auch hier wieder notwendig, die Fehlerwerte (Differenz der vorhergesagten Werte von den Tatsächlichen) separat zu speichern. Dabei verlässt man sich auf die Wahrscheinlichkeit, dass die in den beiden Fehlerkanälen enthaltenen Werte möglichst klein sind, was der noch folgenden Codierung zu Gute kommt.

Residual Coding Die Werte der nun noch verbleibenden beiden Fehlerkanäle werden nun mittels dem Verfahren von Rice codiert. Durch die Annahme, dass beide Kanäle nur kleine Werte enthalten, eignet sich diese Codierung besonders gut, da kleinen Werten automatisch die kleinste Wortlänge zugewiesen wird. Dies spart Rechenzeit, da weder Auftrittswahrscheinlichkeit noch Codierungstabelle angelegt werden müssen. Unter diesen Bedingungen erreicht man mit der Rice-Codierung eine ähnlich gute Datenreduktion, wie das deutlich rechenintensivere Huffman-Coding. Zur Steigerung der Komprimierungseffizienz ist es möglich, den Datenblock in mehrere Partitionen zu zerlegen und die Codierung separat auf jene anzuwenden.

Framing Im letzten Schritt wird das FLAC Bitstreamfile erstellt. Innerhalb der ein oder mehreren Frames werden alle für den Decoder relevanten Daten, wie zum Beispiel Abtastfrequenz, Bittiefe eines Samples, Anzahl der Kanäle und Komprimierungseinstellungen gespeichert. Ein Frame wird jeweils mit Header und Footer begrenzt, in denen Synchronisationsdaten und eine CRC16 (16 Bit Cyclic Redundancy Check) Prüfsumme gespeichert werden. Letztgenannte, um fehlerhaft gespeicherte Daten zu erkennen. Die codierten Audiodaten sind in den Subframes enthalten. Ebenso werden bis zu 128 verschiedene Arten von Metadatenblöcke unterstützt, in denen nähere Angaben (beispielsweise Interpret, Name des Albums, Bild des Covers, etc.) zum gespeicherten Audiofile getätigt werden können.

Zu den lobenswerten Eigenschaften des FLAC Formates zählt auch der Umstand, dass sämtliche Berechnungen zur Codierung und Decodierung mittels Ganzzahlarithmetik erfolgen, was vielen portablen Abspielgeräten aufgrund der beschränkten Rechenleistung zu Gute kommt. Auch sind die Eigenschaften der zu speichernden Audiodaten

sehr variable einstellbar: Es werden bis zu acht separate Kanäle mit einer Abtastrate von 1 bis 655 350 Hz (in Schritten zu 1 Hz einstellbar) unterstützt. Die Wortlänge eines Samples kann dabei zwischen 4 und 32 Bit betragen. Bei der Verwendung von PCM Daten, die den Spezifikationen einer Audio-CD entsprechen, erreicht dieses Format eine Datenreduktion auf 50%-60% gegenüber den Ausgangsdaten. Zum Vergleich: mittels Datenkomprimierung via ZIP ist eine allgemeine Reduzierung von 10%-20% gegenüber der Originalgröße zu erwarten.

3.3 Weitere Formate

Ob es nun darum geht, über das Internet Musiktitel in unverfälschter CD-Qualität zu vertreiben, oder den Datenaustausch von Audiodaten zwischen Verleger und Rundfunksendern sicherzustellen: um die vorhandene Bandbreite und Speicherplatz auf Datenträgern möglichst effizient auszunutzen, dafür hat sich die Verwendung von verlustfrei komprimierten Audiodaten bewährt. So ist es nicht verwunderlich, dass in den letzten 10 Jahren gut ein Dutzend verschiedene Formate entstanden, die nun um die Gunst des Kunden und des Nutzers buhlen. Ein paar Wenige seien hier noch erwähnt:

Apple Lossless Dieses Audioformat wurde 2004 als Bestandteil von QuickTime von der Firma Apple eingeführt. Die Speicherung der komprimierten Daten geschieht dabei im MPEG-4 Container, und die Datei trägt meist die Endung `.m4a` bzw. `.mp4`. Wie bei dem FLAC Format, basiert das Verfahren auf eine lineare Vorhersage mit anschließender Codierung des Restsignals mittels Colomb-Rice-Code. Aber im Unterschied zu FLAC, ist Apple Lossless ein proprietäres Format und wird so standardmäßig nur von Apple-Software, wie beispielsweise iTunes, und Apple-Geräten wie dem iPod unterstützt. Findige Programmierer unterzogen diesem Format jedoch eines Reverse-Engineerings, somit steht seit 2008 ein freier Codec zur Verfügung, der eine Konvertierung in eines der allgemein gebräuchlichen Formate erledigt.

Windows Media Audio Lossless Auch Microsoft bereichert sein WMA (Windows Media Audio) Codec-Paket seit der Version 9 um ein weiteres Format, um verlustfrei Audiodaten in komprimierter Form zu speichern. Wie nicht anders zu erwarten, ist aber auch dieses Format rein proprietär und nur unter MS Windows und diversen Abspielgeräten, implementiert in Hardware, zu benutzen. Bereits seit der ersten Version wird Raumklang mit 6 getrennten Kanälen unterstützt.

Monkey's Audio Ebenso wie FLAC, ist dieses Format ebenfalls frei erhältlich. Auch die Codierung/Decodierung erfolgt nach dem gleichem Schema. Dabei werden

etwas bessere Komprimierungsraten erzielt, allerdings zu Lasten der Rechengeschwindigkeit: Codieren und Decodieren benötigen etwas mehr Rechenzeit gegenüber dem FLAC Format. Die Dateiendung der Audiodaten lautet `.ape`, die der Metadaten ist `.apl`.

4 Formate zur verlustbehafteten und kromprimierten Speicherung

4.1 Grundlagen

Wie in dem Kapitel zuvor zu sehen war, ist mit ausgeklügelten Komprimierungsalgorithmen lediglich eine Datenreduzierung von 50% zu erreichen. In vielen Bereichen, in denen die Bandbreite und Speicherkapazität sehr begrenzt ist, ist aber eine noch viel höhere Datenreduktion wünschenswert. Dies kann aber nur erreicht werden, wenn weniger relevante Anteile des Signals mit minderer Präzision gespeichert bzw. vollständig verworfen werden (Irrelevanzkodierung). Dabei bedient man sich der Eigenschaft des menschlichen Gehörs, dass nicht alle Frequenzen und Amplituden mit gleicher Intensität wahrgenommen werden können. Diese Effekte, die die menschliche Wahrnehmung in Bezug des Hörens beeinflussen, werden in dem psychoakustischem Modell beschrieben [9].

Die wichtigsten Effekte, die eine deutliche Datenreduktion ermöglichen, seien hier in aller Kürze erwähnt:

Hörfläche Dies ist der Bereich, in denen Frequenzen mit bestimmten Schalldruckpegel (allgemein: ein Maß für die Lautstärke) vom menschlichen Gehör wahrgenommen werden können. Dabei werden die Frequenzen, so wie in Abbildung 4.1 zu sehen ist, im Grenzbereich dieser Fläche deutlich schlechter wahrgenommen als beispielsweise Frequenzen um 2 bis 5 KHz (Sprachbereich) und lassen sich mit geringerer Präzision speichern, ohne dass die Wahrnehmung darunter leidet. Frequenzen unterhalb von 20 Hz und Frequenzen über 20 KHz werden von dem Normalhörenden erst gar nicht mehr wahrgenommen.

Maskierungseffekt Durch diesen Effekt können Töne der einen Frequenz, Töne einer anderen Frequenz überdecken und ihn so nicht mehr wahrnehmbar machen. Zum Beispiel ist bei Auftreten eines sehr lauten tiefen Tons, ein leiser Ton mittlerer Frequenz nicht mehr wahrnehmbar. Entscheidend dabei ist der jeweilige Schall-

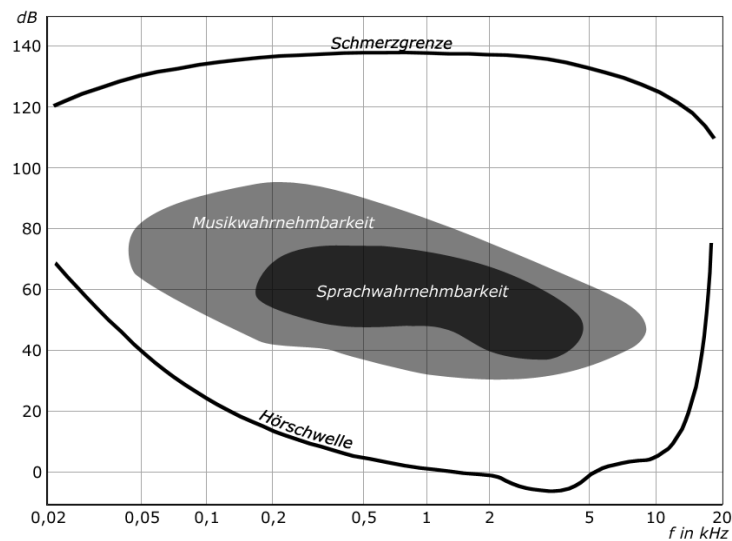


Abbildung 4.1: Wahrnehmbarer Frequenzbereich in Abhängigkeit vom Schalldruckpegel, Quelle: <http://de.wikipedia.org/wiki/Hörbereich>

druckpegel und der Frequenzabstand beider Töne zueinander. Die Maskierungsbereiche können der Abbildung 4.2 entnommen werden. So verdeckt ein Ton mit der Frequenz $f = 1$ KHz mit einer Lautstärke von 80 dB einen weiteren Ton mit $f = 2$ KHz und der Lautstärke von 40 dB. Dieser kann nicht mehr wahrgenommen werden und muss folglich auch nicht gespeichert werden. Tritt der zweite Ton dagegen mit einem Schalldruckpegel von 60 dB auf, so ist er zwar wahrnehmbar, aber es ist ausreichend, ihn in minderer Qualität zu speichern.

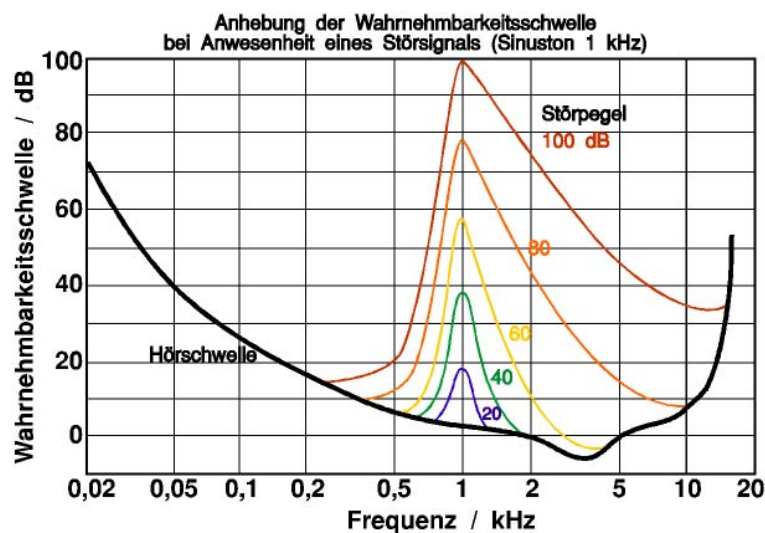


Abbildung 4.2: Wirkungsweise von Maskierungseffekten, Quelle: <http://de.wikipedia.org/wiki/Maskierungseffekt>

Da das Codiervorgang die irrelevanten Anteile des Eingangssignals von den Relevan-

ten unterscheiden muss, ist eine genaue Analyse des zu komprimierenden Datenstroms notwendig. Die Ausnutzung der datenreduzierenden Effekte aus dem psychoakustischen Modell werden dabei im Frequenzbereich durchgeführt. Die Transformation der Eingangsauidodaten ist daher im Vorgang der Codierung sehr bedeutend. Bei den meisten Verfahren wird im ersten Schritt das Eingangssignal, durch das Anwenden von verschiedenen Bandpässen, in Frequenzbänder gleicher Breite unterteilt. Diese einzelnen Frequenzbänder, welche in ihrer Anzahl von ein paar wenigen bis hin zu ein paar tausend variieren können, werden dann einer Modifizierten Diskreten Kosinustransformation (MDCT) unterzogen. Anschließend wird ein Jenes durch die durch das psychoakustische Modell gegebenen Bedingungen unterschiedlich quantisiert. Dadurch wird eine Datenreduzierung von 1:10 bis 1:50 gegenüber dem Eingangsdatenstrom erreicht, mehr oder weniger auf Kosten der Qualität.

Durch die aufwendige Neuquantisierung der Audiodaten unter Berücksichtigung des menschlichen Wahrnehmungsvermögens kann es aber auch zu Verfälschungen kommen:

Generationsverlust Die Codierung von Audiodaten in ein verlustbehaftetes komprimiertes Audiosignal sollte immer erst im letzten Schritt geschehen, da eine Wiederherstellung gegenüber dem Originalzustand ausgeschlossen ist. Eine Konvertierung von einem verlustbehafteten Audiosignal in ein anderes Verlustbehaftetes (Transcodierung) führt in der Regel dazu, dass weitere Informationen der originalen Audiodaten verloren gehen und die Qualität dadurch enorm leidet.

Vorecho Dieses Codierungsartefakt tritt vor allem beim Wechsel sehr ruhiger Passagen gegenüber lauterem, hochfrequenten Passagen, wie etwa dem Einsetzen von Schlagzeug, auf. Es ist dadurch gekennzeichnet, dass man das Einsetzen der folgenden Passage bereits im Vorfeld als Echo wahrnimmt. Zurückführen lässt sich dies auf die Wahl eines zu großen Frequenzblockes, so wird das eher große Quantisierungsrauschen der stillen Passage auf die abrupt wechselnde Passage mit extrem kurzer Einschwingphase übertragen.

Unterschiedliches Quantisierungsrauschen Hervorgerufen durch unterschiedliche Quantisierungen der einzelnen Frequenzblöcke, kann das damit verbundenen Quantisierungsrauschen den Eindruck der Rauheit einer Audiodatei hervorrufen, der sich in Brummen und Schnarren äußert.

4.2 Das MP3 Format

Bereits seit Ende der 70er Jahre des letzten Jahrhunderts bestehen Bemühungen, ein geeignetes digitales Verfahren zur Übertragung von Sprache und Musik zu finden, um eine qualitativ bessere Übertragungsqualität über die neuen Kommunikationswege ISDN und Glasfaserkabel zu realisieren. Zu diesem Zweck gehen im Jahr 1987 die Universität Erlangen-Nürnberg und das Fraunhofer-Institut für Integrierte Schaltungen IIS ein Forschungsbündnis ein, um im Rahmen des von der EU geförderten „EUREKA“-Projektes für Digital Audio Broadcasting (DAB) einen geeigneten Algorithmus zur verlustbehafteten Komprimierung, mit der Vorgabe der maximalen Übertragungsqualität bei minimaler Datenrate, zu entwickeln. Im Rahmen einer Ausschreibung der 1988 von der internationalen Standardisierungsorganisation ISO gegründeten „Moving Picture Experts Group“ (MPEG), welche fortan verantwortlich für die Entwicklung von Audio- und Videokompressionsstandards sein sollte, wurde unter anderem auch der vom Fraunhofer-Institut entwickelte ASPEC (Adaptive Spectral Perceptual Entropy Coding) Codec als Vorschlag eingereicht. 1991 fällt dann die Entscheidung durch MPEG, drei der insgesamt 14 eingereichten Vorschläge als neuen Standard für die Audiodatenkompression zu etablieren. 1993 beschließt MPEG dann letztlich, die erste Codec-Familie für die Speicherung von Audiodaten zu standardisieren [7].

In dieser enthält der **Layer-1** den Codec zur Speicherung von Audiodaten nach dem MUSICAM Verfahren, welcher bereits bei der Digital Compact Cassette (einer 1992 von Philips und Matsushita entwickelten Alternative zur analogen Musikkassette) eingesetzt wurde und für die noch erscheinende Video CD vorgesehen war. Beide Medien hatten nur wenig Erfolg und existieren heute nicht mehr. Bei der Realisierung des digitalen Rundfunks DAB entschied man sich für die Verwendung des **Layer-2**, einem etwas komplexeren MUSICAM Codecs, der bei gleicher Qualität mit geringerer Bandbreite auskommt. Ebenso kommt er bei der Ausstrahlung des digitalen Fernsehens DVB zum Einsatz. Das aufwendige, aber in Hinblick auf den Speicherplatz sehr effiziente Verfahren ASPEC, wird in dem **Layer-3** spezifiziert. Zur einheitlichen Dateierweiterung MP3 kam es dagegen erst 1995. Eine Übersicht über typische Datenraten der einzelnen Standards kann der Tabelle 4.1 entnommen werden.

Format	Datenrate in KBit/s	Anwendung
Audio CD	1411	Compact Disc Digital Audio
MPEG-1 Layer-1	384	Digital Compact Cassette, Video CD
MPEG-1 Layer-2	256	DAB, DVB
MPEG-1 Layer-3	128	MP3, DIVX

Tabelle 4.1: Typische Bitraten für MPEG-1 Audiostreams

In der ersten Phase der Codierung werden die in PCM vorliegenden Audiodaten durch eine Mehrphasen-Filterbank in 32 Subbänder verschiedener Breite aufgeteilt. Nun wird jedes Subband in den Frequenzbereich mittels der MDCT transformiert. Anschließend erfolgt die Kodierung anhand der Maskierungskurven aus dem psychoakustischen Modell. Durch Festlegung der Bitrate, kann der Benutzer selbst die Qualität des zu erzeugenden Audiofiles beeinflussen, denn die zur Verfügung stehenden Bits bestimmen das Signal to Mask Ratio (Verhältnis von Signal zu maskiertem Bereich). Der nicht maskierte Bereich der Signale wird dann mit den noch zur Verfügung stehenden Bits nichtlinear quantisiert. Die noch verbleibenden Daten werden nun einer Huffman-Codierung unterzogen. Die codierten Signale werden in Frames gespeichert, wobei ein Jeder von einem 32 Bit umfassenden Header eingeleitet wird. Dieser dient zum einen dem Hin- und Herspringen in der MP3-Datei (also zum Vor- und Zurückspulen), und zum anderen sind in ihm jeweils Informationen zu Abtastfrequenz und Bitrate enthalten. Da diese von Frame zu Frame verschieden sein können, ermöglicht es diesem Format die Nutzung einer variablen Bitrate bei der Codierung von Audiodaten. Dem jeweiligen Header folgen nun Seiteninformationen, die zur Dekomprimierung benötigt werden. Diese setzen sich aus den Huffman-Tabellen, der Quantisierungsschrittgröße und den MDCT-Blockgrößen zusammen. Danach folgen die codierten Frequenzen, also die eigentlichen Audiodaten. Die Decodierung erfolgt dann in umgekehrter Reihenfolge, wobei dies mit deutlich weniger Rechenaufwand verbunden ist.

Offset	Länge in Bytes	Bedeutung
0	3	String „TAG“
3	30	Songtitel
33	30	Künstler/Interpret
63	30	Album
93	4	Jahr der Veröffentlichung
97	30	Kommentar
127	1	Genre

Tabelle 4.2: Meta-Datenblock nach ID3v1

Die Spezifikation des MP3-Formates sieht keine Art der Speicherung von Meta-Daten vor. Trotzdem wurde mit ID3 (**I**dentify an **MP3**) ein informeller Standard geschaffen, um detaillierte Informationen zu den vorliegenden Audiodaten zu speichern. Die erste Version ID3v1 besteht aus einem festen 128 Byte umfassenden Datenblock, der einer MP3-Datei am Ende angefügt wird. Der Aufbau des ID3v1-Datenblocks kann der Tabelle 4.2 entnommen werden. Das aktuelle Meta-Datensystem in der Version 2.4 ist dabei deutlich flexibler und erlaubt unter anderem auch die Speicherung von Bildern (bis hin zu Slideshows).

Mit dem Ausbau der breitbandigen Datenautobahnen und dem dadurch resultierenden Boom des Internets Mitte der 90er Jahre, erlangte das MP3 Format den Status des Standardformates. Aufgrund der geringen Größe und der guten Qualität der Audiodaten, entstanden, zum Leidwesen der Musikindustrie, diverse Tauschbörsen wie Napster oder Kazaa, über die dann einzelne Musikstücke, Alben oder sogar ganze Musikarchive kostenlos über den gesamten Globus verteilt wurden. Aber auch die Spieleindustrie wusste das MP3 Format zu schätzen und setzte es verstärkt zur akustischen Untermalung ihrer Spiele auf den immer leistungsfähiger werdenden PCs ein. Das Aufkommen der ersten tragbaren MP3-Player im Jahre 1998 verdrängte in kürzester Zeit sämtliche portable Kontrahenten, deren Speicherung auf Kassette, CD bzw. MiniDisc beruhte. Für Radio- und TV-Sendungen über das Internet (Webradion und IPTV) bildet das MP3 Format ebenso die Grundlage in Form von Streams.



Abbildung 4.3: 1998 erschien der erste tragbare MP3-Player, der „mpman“ von SaeHan mit unglaublichen 32 MByte Flashspeicher

Im Zuge der Weiterentwicklung wurde aus dem ehemals externen Codec (einmal auf dem System installiert, kann jede Software frei darauf zugreifen) ein interner Codec, der nur von offiziell lizenzierter Software genutzt werden konnte. Dies hätte das Aus für die vielen sehr guten Open-Source Programme bedeutet, die die Bearbeitung von Audiodaten erledigen. Aber auch dies hinderte Niemanden, das MP3 Format zu nutzen. Schon nach sehr kurzer Zeit erschien von der Cracker-Gruppe Radium eine „überarbeitete“ Version des Fraunhofer Codecs, der nicht nur die neuen Restriktionen wieder aufhob, sondern auch diverse Fehlerbereinigungen und Optimierungen mitbrachte [8]. Um zumindest im privaten Umfeld nicht auf den lizenzpflichtigen MP3 Codec zugreifen zu müssen, ohne dabei in die Illegalität durch Benutzung des Radium MP3 Codecs gedrängt zu werden, bietet sich heute die Verwendung des LAME Codecs an. Dieses

Open Source Projekt widmet sich seit 1998 der Entwicklung eines freien MP3 Codecs auf Basis eines von der ISO veröffentlichten Beispielquelltextes, und erreicht bei ähnlicher Qualität und Speichergröße hundertprozentige Kompatibilität gegenüber dem lizenzpflichtigen Fraunhofer Format.

4.3 Das OGG Vorbis Format

Von der Ankündigung durch das Fraunhofer Institut, ihren MP3 Codec unter lizenzrechtlichen Bestimmungen zu stellen, waren viele Open Source Projekte betroffen. Daher startete 1998 ein Projekt namens Vorbis zur Entwicklung eines völlig freien Codecs, welcher leistungstechnisch dem MP3 Format in nichts nachstehen sollte. Im Laufe seiner Entwicklung trat der Entwickler des Vorbis Codecs der Xiph.Org Foundation bei, und der Codec wurde in das Containerformat OGG miteingebunden, welches zur Aufnahme einer Vielzahl von freien Audio- und Videostreams dient. 2002 wurde die Version 1.0 dieses Codecs veröffentlicht und steht seitdem in ständiger Konkurrenz zu MP3 und anderen kommerziellen Audioformaten, die der verlustbehafteten komprimierten Speicherung dienen [10].

Das OGG Vorbis Format unterstützt Audiodaten mit bis zu 255 unterschiedlichen Kanälen. Der Schwerpunkt bei der Codierung der Audiodaten liegt dabei auf Verwendung einer variablen Bitrate, bei der sich schnell ändernde Samples innerhalb eines Frames besser quantisiert werden als die in ruhigeren Passagen. Durch die Einbindung des Vorbis Codecs in das OGG Containerformat, wird auch durch die Benutzung von deren umfangreichen und im Vergleich zu ID3 des MP3 Formates weniger restriktiven Meta-Daten Systems names Vorbis Comment profitiert. Trotz sehr guter Codierungseigenschaften in Hinsicht auf Qualität bei geringem Verbrauch an Speicherplatz, ist das OGG Vorbis Format deutlich weniger verbreitet, als es das MP3 Format ist. Die Standardimplementierung des Vorbis Decoders arbeitet hauptsächlich mit Gleitkommazahlen. Doch den meisten portablen Abspielgeräten steht keine separate Einheit zur Berechnung von Gleitkommazahlen (FPU) zur Verfügung, was einer weiten Verbreitung dieses Formates lange Zeit im Wege stand. Abhilfe schafft hier der 2002, ebenfalls im Open Source Bereich freigegebene Tremor Decoder, der die Decodierung des Audiodatenstroms mit Ganzzahlarithmetik erledigt.

4.4 Das WMA Format

Mit „Windows Media Audio“ wird das Containerformat zur Speicherung von Audiodaten und der Codec zur verlustbehafteten komprimierten Codierung/Decodierung gleichermaßen bezeichnet. Das WMA Format wurde erstmals 1999 als Microsofts neue Multimedia-Plattform „Windows Media Technologies 4“ veröffentlicht, dessen Ziel es war, die Audioanteile in Webinhalten (Webradio, IPTV etc) auf ein der Audio CD ähnliches Niveau zu heben, bei gleichzeitiger Reduzierung der dafür nötigen Bandbreite [3]. Erst kurze Zeit darauf wurde das Format unter den bis heute gültigen Namen WMA veröffentlicht. Mit dem 2003 erscheinenden Codec-Paket WMA9 hielten noch drei weitere Codecs Einzug [4], die aber alle inkompatibel zueinander sind:

- Windows Media Audio 9 Professional (Mehrkanalunterstützung)
- Windows Media Audio 9 Lossless (siehe Kap. 3.3)
- Windows Media Audio 9 Voice (Sprachkodierung)

Die mit WMA codierten Audiodaten werden dabei in Microsofts proprietärem ASF Format (Advanced Systems Format, früher: Advanced Streaming Format bzw. Active Streaming Format) gespeichert, welches die Dateiendung `.asf`, `.wma` oder `.wmv` besitzt. Das Abspeichern von Meta-Daten wird von diesem Format ebenso unterstützt, vergleichbar mit dem ID3 des MP3-Formates. Von dem originalen WMA-Codec werden allerdings nur Audiodaten mit maximal 2 Kanälen bei einer Abtastfrequenz von maximal $f = 48$ KHz unterstützt. Die Möglichkeit der Codierung/Decodierung mit variabler Bitrate wurde dabei erst mit erscheinen des WMA9 Codecs eingeführt. Die Codierung/Decodierung der Audiodaten erfolgt dabei nach dem gleichen Schema, wie es schon im MP3 Format vonstatten geht.

Die 1999 getätigte Beteuerung, dass eine mit WMA codierte Audiodatei mit einer Bitrate von 64 Kbit/s qualitativ mit einer mit 128 Kbit/s codierten MP3 Datei vergleichbar ist, ist man von Seiten Microsofts bis heute schuldig geblieben [5]. Das WMA Format wird dennoch von der breiten Masse von tragbaren Abspielgeräten unterstützt. Bestimmend für diesen Erfolg ist die Einbindung von DRM (Digital Rights Management) in das ASF Format, welches wie ein Segen auf die Musikindustrie wirkte. Die Nutzung dieser Mechanismen sollte die ungebremste Verbreitung von urheberrechtlich geschützten Musiktiteln über das Internet oder austauschbaren Datenträgern in Form der sehr populären MP3 Dateien bremsen und verhindern. Durch die Integration der digitalen Rechteverwaltung in das Format selbst, entdeckte die Musikindustrie den Vertriebsweg ihrer Produkte mittels Online-Shops. Einer der Ersten, die diesen neuen Vertriebsweg sehr erfolgreich nutzten, war im Jahr 2003 der „iTunes Music Sto-

re“. Doch die Einschränkungen in Hinsicht auf Benutzerfreundlichkeit, Archivierung und Interoperabilität der Abspielgeräte für die bezahlte Ware wurde von vielen kritisiert, weshalb heutzutage im kommerziellen Bereich bei den meisten Anbietern auf die Nutzung von DRM verzichtet wird.

4.5 Weitere Formate

MPEG-2/4 AAC Das AAC Format (Advanced Audio Coding) wurde 1997 durch MPEG als Nachfolger des sehr populären MP3 Formates vorgestellt und ist seitdem Teil der MPEG-2 und MPEG-4 Spezifikation zur Codierung von Audiodaten. Die Speicherung erfolgt dabei im Allgemeinen im MPEG-4 Containerformat, welches die Endung **.mp4** bzw. **.m4a** trägt und eine Weiterentwicklung des QuickTime-Formates der Firma Apple darstellt. Mit AAC halten eine Vielzahl von Verbesserungen gegenüber dem MP3 Format Einzug, wie eine Unterstützung von bis zu 48 Kanälen, Verwenden von Abtastfrequenzen im Bereich von 8 bis 96 KHz, sowie diverse Optimierungen, welche die Qualität der komprimierten Audiodaten verbessern bei gleichem Speicheraufwand und den Rechenaufwand für Codierung/Decodierung verringern. Ebenso ist eine Nutzung der Mechanismen des DRM vorgesehen, um eine kommerzielle Vermarktung von Audiodaten zu gewährleisten. Die größte Veränderung gegenüber MP3 stellt aber die Verwendung sogenannter **Tools** während der Codierung dar. Die Verwendung dieser verfolgt ein modulares Konzept zur Verbesserung der Codierungsqualität. Da nicht jeder Codierer/Decodierer alle Tools unterstützten muss, werden Kombinationen derer in **Object Types**, je nach Anwendungszweck, definiert und standardisiert.

mp3PRO Dieses Format stellt eine Weiterentwicklung des MP3 Formates dar, mit dem Ziel, die stark nachlassende Qualität bei Verwendung von Bitraten unterhalb der 96 KBit/s Grenze zu verbessern. Dies wird mit dem proprietären, patentierten Verfahren der Spektralband-Replikation erreicht, durch die die Töne mit hohen Frequenzen mittels Steuersignalen, welche in den Tönen des niederen bis mittleren Frequenzbereiches codiert sind, regeneriert werden, wodurch sich wiederum ein gewisses Maß an Bandbreite einsparen lässt. Das mp3Pro Format ist dabei abwärtskompatibel zu dem herkömmlichen MP3 Format, jedoch werden die hohen Frequenzen innerhalb der Audiodatei nicht korrekt wiedergegeben, sofern der Player dieses Format nicht explizit unterstützt. Dies äußert sich dann in einem eher dumpfen, basslastigen Klangbild der jeweiligen Audiodatei. Aufgrund einer restriktiven Lizenzpolitik, blieb diesem Format der Durchbruch verwehrt, so dass die Weiterentwicklung 2003 eingestellt wurde. Als Alternative bietet sich derzeit

das AAC Format aus der MPEG-4 Familie mit dem Obejct Type: HE-AAC (High Efficiency Advanced Audio Coding) an.

5 Zusammenfassung und Ausblick

Je nach Anwendungsgebiet, hat jede der hier behandelten Klasse von Audioformaten ihre Daseinsberechtigung. Während der Bearbeitung von Audiodaten bietet sich die verlustfreie und unkomprimierte Speicherung an, was allerdings den meisten Festspeicher in Anspruch nimmt. Für die Archivierung von Audiodaten empfiehlt sich die Speicherung in eines der verlustfrei komprimierten Formate, da sie zur Wahrung der Originalität weniger Speicherplatz verbrauchen. Überall dort, wo es statt auf originalgetreuer Wiedergabequalität der Audiodaten auf Nutzung knapper Ressourcen in Form von Bandbreite und Speicherplatz ankommt, bietet sich eine Speicherung in eines der verlustbehafteten komprimierten Verfahren an.

Auch wenn in naher Zukunft Revolutionen im Bereich der Codierung von Audiodaten ausbleiben werden, so sind doch gewisse Trends abzusehen. Bei der Weiterentwicklung der verlustbehafteten komprimierten Formate rückt stärker die Sprachcodierung in den Vordergrund, um mit möglichst kleiner Bandbreite akzeptable Qualität bei der Übertragung von Sprache zu ermöglichen. Ebenso wurde in den vergangenen Jahren sehr viel Wert darauf gelegt, diese bestehenden Formate um die Unterstützung von Mehrkanal-Audiostreams zur Realisierung von Raumklang zu erweitern.

Bei dem Vertrieb von Musiktiteln über das Internet und der Archivierung setzt man verstärkt auf verlustfrei komprimierte Formate. In diesem Segment haben sich vor allem die durch engagierte Open-Source Projekte geschaffenen Formate gegenüber ihren proprietären Konkurrenten durchgesetzt. Ein Trend, der sich übrigens auf beide Klassen der komprimierten Speicherung abzeichnet. Die Akzeptanz der freien Audio-Codecs und ihrer Formate wächst ständig, aufgrund ihrer flexiblen, transparenten Struktur, der hohen Qualität und nicht zuletzt des niedrigen Preises, der für eine kommerzielle Nutzung in Frage kommt. Derweil ist die Musikindustrie noch auf der Suche nach wirksamen Mechanismen des Watermarkings für ihre Produkte, ohne auf das nutzungsbeschränkende und gescheiterte DRM zurückgreifen zu müssen. Und das, obwohl die Vergangenheit lehrt: noch nie hat ein Kopierschutz das gehalten, was er versprochen hat !

Literaturverzeichnis

- [1] Chris Bagwell. *Audio File Formats FAQ*. <http://sox.sourceforge.net/AudioFormats.html>, 1998.
- [2] Josh Coalson. *Flac - Free Lossless Audio Codec*. <http://flac.sourceforge.net/>, 2008.
- [3] Microsoft Corporation. *Windows Media Technologies 4*. <http://www.microsoft.com/presspass/press/1999/Aug99/WM4Lnchpr.msp>, 1999.
- [4] Microsoft Corporation. *Windows Media Audio Codecs*. <http://www.microsoft.com/windows/windowsmedia/forpros/codecs/audio.aspx>, 2010.
- [5] Nicholas Cravotta. *The Internet-audio (r)evolution*. <http://www.edn.com/index.asp?layout=article&articleid=CA46537>, 2000.
- [6] Justin Deltener. *Programming The Soundblaster 16*. <http://www.inversereality.org/tutorials/soundaster16example3.html>.
- [7] Fraunhofer IIS. *Die MP3 Geschichte*. <http://www.iis.fraunhofer.de>, 2009.
- [8] David McCandless. *Die Welt der Raubkopierer: Eine Geschichte von Sammlern und Jägern*. <http://www.heise.de/tp/r4/artikel/9/9171/1.html>, 2001.
- [9] Stefan Weinzierl. *Handbuch der Audiotechnik*. Springer-Verlag Berlin-Heidelberg, 2008.
- [10] Xiph.Org. *Vorbis audio compression*. <http://xiph.org/vorbis/>, 2009.

6 Erklärung über die selbständige Anfertigung der Arbeit

Ich versichere, dass ich die vorliegende Arbeit selbständig und ohne unerlaubte Hilfe Dritter verfasst sowie keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Alle Angaben, die inhaltlich oder wörtlich aus fremden Werken stammen, wurden kenntlich gemacht. Diese Arbeit lag in gleicher oder ähnlicher Weise noch keiner Prüfungsbehörde, Fachhoch- oder Hochschule vor und wurde ebenfalls bisher noch nicht veröffentlicht.

Rostock, 21. März 2010

Maik Gotzmann