

Versuch i-37-b

Soundverarbeitung

Steve Moser, Markus Ganzenmüller

19. Dezember 2002

Inhaltsverzeichnis

1	Einleitung	3
2	Erzeugung und Bearbeitung der Samples	4
2.1	Creative Wave Studio	4
2.1.1	Bedienung	4
2.2	Erstellung der Samples	5
2.2.1	Musikstück	5
2.2.2	Guten Tag	5
2.3	CoolEditPro	6
2.4	MIDI-Editor Infinity 2.0	7
3	Begriffserklärungen	7
3.1	Bandbreite	7
3.2	MIDI	8
3.3	Hall/Echo	9
3.4	Fade	10
3.5	WAVE	11
3.6	AVI	12
3.7	MPEG	13

3.7.1	MPEG1	13
3.7.2	MPEG2	13
3.7.3	MPEG2 Kompression mit Hilfe von I- P- und B-Frames	14
3.7.4	MPEG3	14
3.7.5	MPEG4	14
4	Fazit	15

1 Einleitung

Bereits Mitte des 17. Jahrhunderts fanden die ersten Versuche im Zusammenhang mit elektronischer Klangerzeugung statt. Die damals von experimentierfreudigen Männern konstruierten Instrumente, waren aber alles andere als alltagsgebräuchlich, sondern sind allenfalls als erster Schritt einer langen und glanzvollen Karriere dieser Geräte zu betrachten, welche seitdem bis in unsere heutige Zeit andauert. Zum Ende der dreißiger Jahre des 20. Jahrhunderts wurden die ersten richtigen Synthesizer gebaut. Und Mitte der achtziger Jahre erreichte die Verwendung von Synthesizern einen Höhepunkt. Zu dieser Zeit wurde das MIDI-Interface erfunden, welches endlich die Kommunikation verschiedenster Geräte der mittlerweile zahlreich gewordenen Hersteller miteinander standardisierte und den Synthesizern zum endgültigen Durchbruch verhalf. Das Interesse der Musikindustrie wuchs immens und führte sogar zur Gründung einer neuen Musikrichtung: dem "Synthipop". Mit dem Einzug der Personalcomputer in die Haushalte begann auch der Siegeszug der elektronischen Erstellung und Bearbeitung von Musikstücken am Computer, an welche die Synthesizer angeschlossen werden konnten. Mit einfachen Mitteln ließen sich durch geeignete Software komplette Musikstücke komponieren und wahlweise mit Hilfe einer Soundkarte im Computer oder eines daran angeschlossenen Synthesizers abspielen. Aber nicht nur Musik sondern auch die menschliche Sprache und alle anderen Arten von Tönen und Geräuschen lassen sich mittlerweile aufnehmen oder künstlich am Computer erzeugen und bearbeiten.

Unser fünfter Praktikumsversuch beschäftigte sich mit der Aufnahme und Bearbeitung verschiedener Samples mit Hilfe von "Creative Wave-Studio". Ferner sollten wir, wie im c't Heft 18 des Jahrgangs 2002 beschrieben, verschiedene Sprachproben aufnehmen und daraus Teilbereiche ausschneiden, um die Worte "Guten Tag" zusammenzuschneiden. Desweiteren sollten Programmfeatures, wie z.B. *Echo* und *Hall* oder *Fading* getestet werden. Zusätzlich sollten die Demoversion eines professionellen Soundbearbeitungsprogramms, wie z.B. CoolEditPro und ein MIDI-Editor ausprobiert werden.

Während des Kolloquiums sollten die erstellten Samples vorgeführt und die Bedienung der Programme und einiger Features geschildert werden. Ein weiteres Thema sollten die heute gängigen Musik- und Bildformate MIDI (eigentlich kein Format, sondern eine Schnittstelle), WAVE, AVI und MPEG sein.

Das nächste Kapitel befaßt sich zunächst mit der grundlegenden Bedienung des Programms "Creative Wave-Studio" und einer kurzen Beschreibung, wie wir unsere Samples erstellt und bearbeitet haben. Der zweite Teil unseres Berichts schildert die oben genannten Bild und Soundformate.

2 Erzeugung und Bearbeitung der Samples

2.1 Creative Wave Studio

Unsere Samples haben wir mit Creative Wave Studio erstellt und bearbeitet. Dieses Programm befindet sich im Softwareumfang der meisten Creative Soundkarten, so auch bei der Soundblaster PCI 128, welche sich im verwendeten Rechner befand. Es lassen sich jedoch nur Sounddateien im wav-Format damit bearbeiten, bzw. erzeugen. Das Programm hat keine große Funktionsvielfalt, ist aber für Anfänger ausreichend, um einfache Beispiele aufzunehmen und zu bearbeiten.

2.1.1 Bedienung

Nach dem Start von Wave Studio findet der Benutzer ein leeres Fenster und diverse Menüs in der Kopfzeile vor. Neben den üblichen Oberpunkten File, Edit, View, Options, Window und Help welche in vielen Programmen ähnlich sind und sich in ihrer Funktion nicht wesentlich unterscheiden, ist eigentlich nur der Oberpunkt "Special" von größerer Bedeutung, weil sich mit den vorhandenen Menüunterpunkten verschiedene Effekte erzeugen lassen. Häufig benutzte Menüpunkte sind unterhalb der Textmenüs auch als grafische Buttons wiederzufinden, um die Bedienung des Programms einfacher zu machen. Zusätzlich befinden sich in der grafischen Menüleiste vier Anzeigen, welche dazu dienen, Anfangspunkt, Endpunkt, Länge eines ausgewählten Bereichs der Sounddatei und dessen Größe zu bestimmen.

Mit den Funktionen, die dem Punkt "Audio" untergeordnet sind, lassen sich Sounddateien aufnehmen und abspielen. Die Aufnahme eigener Sounds ist sehr einfach. Man benötigt lediglich ein Mikrofon, welches an den Mikrofon-Eingang der Soundkarte angeschlossen wird. Ein Klick auf den Aufnahmebutton genügt, um selbige zu starten. Ein weiterer Klick auf den Stop-Button reicht aus, um die Aufnahme zu beenden.

Nachdem man entweder über das Textmenü oder per Klick auf den Open-Button eine wav-Datei geladen oder auch selbst eine Aufnahme gemacht hat, präsentiert sich diese in einem geteilten Fenster, welches in seiner Größe und Position frei veränderbar ist. Der untere Teil des Fensters zeigt die komplette Sounddatei, wobei der rechte und linke Kanal bei Stereoaufnahmen getrennt und in roter, bzw. blauer Farbe angezeigt werden. Der obere Teil des Fensters dient bei der Bearbeitung als Lupe und zeigt, je nach Auswahl, bestimmte Bereiche des Frequenzverlaufs stark vergrößert an, was vor allem bei genauem Arbeiten sehr hilfreich ist. Vergrößert man nun durch Markieren mit der Maus und mit Hilfe des Menüs einen bestimmten Bereich, wird dieser durch ein kleines Fenster im unteren Teil, in dem die vollständige Datei angezeigt wird, markiert. Dieses Fenster wiederum läßt sich mit der Maus nach links und rechts verschieben, oder nach Wunsch vergrößern und verkleinern. Man kann nun per Copy&Paste Dateien verändern, Bereiche

einfügen, ausschneiden, oder durch das Menü "Special" gewünschte Effekte erzeugen, wie z.B. das Umkehren eines Teilstücks, Fading, also Lautsärkenübergänge, oder auch Echoeffekte.

Unsere Aufgabe war unter anderem, verschiedene Samples selbst aufzunehmen und zu bearbeiten, was im folgenden Abschnitt kurz beschrieben werden soll.

2.2 Erstellung der Samples

2.2.1 Musikstück

In unserem ersten Versuch haben wir einen Teil eines Musikstücks aufgenommen, um diesen dann zu verändern. Auf dem beigelegten Blatt Nummer 1 ist ein Screenshot des Creative Wave Studio zu sehen, welcher links oben die Originaldatei zeigt. Markiert ist ein Teil der Datei, an der der Sänger des Stücks das Wort "walking" singt. Diesen Teil haben wir kopiert und mehrere Male hintereinander eingefügt, was im rechten oberen Fenster zu sehen ist.

Im linken unteren Fenster ist wiederum das komplette Teilstück zu sehen, aus der wir aber das "walking" ausgeschnitten haben, was sich beim Abspielen kaum bemerkbar macht. Schon an dieser Stelle wurde uns klar, mit welchen einfachen Mitteln sich Sounddateien komplett verändern, bzw. manipulieren lassen.

Das rechte untere Fenster zeigt das Frequenzmuster, nachdem wir über das Menü "Special" den Effekt Fading-Out auf beide Kanäle angewendet und danach den rechten Kanal umgedreht haben. Fading bedeutet nichts anderes als Lautstärkenübergang, bzw. Lautstärkenregulierung. Im Beispiel wird die Lautstärke des linken Kanals langsam reduziert und die des rechten Kanals angehoben. Spielt man die Datei ab, wird die Musik in der linken Box immer leiser, bis sie letztlich verstummt, in der rechten Box jedoch ist sie zuerst gar nicht zu hören, schwillt dann langsam an, bis sie die volle Lautstärke erreicht hat. Vom Prinzip her lassen sich auf diese Weise auch einfache "Raumklänge" erzeugen. Würde man das Geräusch eines Motorrads aufnehmen und den Fading Effekt darauf anwenden, hätte man bei geschlossenen Augen den Eindruck, daß das Motorrad von links nach rechts vorbeifährt.

2.2.2 Guten Tag

Nachdem wir nun mit der Handhabung des Programms einigermaßen vertraut waren, begannen wir damit den Versuch der PC Zeitschrift c't (Heft 18, 2002) nachzuvollziehen. Hier war es die Aufgabe, acht Lauteinheiten, sogenannte "Logatome", aufzunehmen und daraus Laute zu extrahieren, um aus diesen die Worte "Guten Tag" zusammenzuschneiden.

Die Lauteinheiten waren im einzelnen:

- [g]aterei
- di[gu]traí
- dib[ut]raí
- gla[tn]
- ryk[nt]rag
- di[ta]traí
- dib[ak]raí
- deska[h]

Die letzte Lauteinheit sollte laut c't eigentlich derka[k] heißen, wurde aber vermutlich aus versehen (oder absichtlich ?) falsch in den Text der Aufgabenstellung übernommen. An dieser Stelle ist ein Blick zum Screenshot auf dem beigelegten Blatt Nummer zwei hilfreich, welcher im oberen Fenster das gesprochene "deska[h]" zeigt und im unteren Bereich die Worte "Guten Tag". Aus dem Logatom "deska[h]" haben wir das langgezogene *a* herausgeschnitten, um beim Wort "Tag" das *a* ein wenig in die Länge zu ziehen. Die beiden Pfeile auf dem Screenshot markieren diesen Arbeitsschritt. Auf diese Art und Weise sind wir vorgegangen, um nacheinander aus allen Probeaufnahmen die richtigen Teile auszuschneiden und die Worte "Guten Tag" zusammenzufügen, was recht gut geklappt hat. Erwähnenswert ist eventuell noch, dass die Aufnahmen nicht nur von einem einzigen, sondern von vier verschiedenen Sprechern gemacht wurden.

2.3 CoolEditPro

Mit diesem Editor haben wir uns aus Zeitgründen nur kurz beschäftigt. Außerdem ließen sich unsere Ergebnisse leider nicht abspeichern, weil die Zeitbeschränkung dieser Shareware leider schon abgelaufen war.

CoolEditPro ist ein professionelles Programm zur Bearbeitung verschiedener Soundformate, wobei der Funktionsumfang anfangs nahezu unüberschaubar ist. Es lassen sich verschiedene Kanäle parallel bearbeiten und zahlreiche Effekte erzeugen. Um die Software zu testen haben wir mittels drag&drop ein kurzes Musikstück aus wav-Dateien zusammengebastelt, welche wir auf dem Laborcomputer gefunden hatten. Auf einem zweiten Kanal haben wir das Stück dann mit Trommeleffekten und kurzen Stimmproben unterlegt.

2.4 MIDI-Editor Infinity 2.0

Auch diesen Editor haben wir nur während der Abnahme kurz angetestet. Dazu haben wir von der Internetseite <http://www.squest.com> eine Demoversion heruntergeladen und installiert. Beispielsamples waren leider nicht enthalten, sodaß wir kurzerhand wahllos ein MIDI-Musikstück aus dem Internet heruntergeladen und ausprobiert haben. Wahlweise lassen sich im Programm verschiedene Kanäle stumm schalten, was es dem Benutzer ermöglicht, das Stück ablaufen zu lassen und sozusagen selbst auf einem gewählten Instrument mitzuspielen. Das haben wir erfolgreich getestet, wobei das Ergebnis aufgrund mangelnden Talentes zu wünschen übrig ließ.

Grundsätzlich ist es für Anfänger empfehlenswert, sich zuerst im Umgang mit einfacherer Software zu üben, bevor man auf professionelle Programme umsteigt, da diese wesentlich schwerer zu bedienen und wesentlich umfangreicher sind. Zudem sollte man sich vor dem Kauf mit der Demoversion des entsprechenden Programms auseinandersetzen, um einem Fehlkauf vorzubeugen. Hat man sich einmal in die Umgebung eingearbeitet, stehen einem mächtige Werkzeuge zur Verfügung.

3 Begriffserklärungen

Die letzte Aufgabe war es, Informationen über verschiedene Bild- und Soundformate zu sammeln. Verlangt war die Erklärung der Begriffe MIDI, AVI, WAVE, MPEG, ECHO/HALL, Fade und Bandbreite.

3.1 Bandbreite

Der Begriff Bandbreite entspricht auch in der elektronischen Musikerzeugung und -bearbeitung der Menge an Informationen, die in einem bestimmten Zeitabschnitt übertragen werden kann.

Die Audiobandbreite für Stereoklang bei einer Samplingrate von 44,1 kHz entspricht 224 KBit/s, das heißt es werden mindestens 224 KBit an Daten pro Sekunde benötigt, um einen Klang dieser Qualität zu erzeugen. Quelle: [1]

Im Zusammenhang mit dem "Hören" hat der Begriff Bandbreite jedoch noch eine andere Bedeutung. So beschreibt er auch die Breite des z.B. in einem Musikstück verwendeten Frequenzbandes. Dies ist wichtig im Zusammenhang mit der Kompression von Audiodaten. Da im Internet nur eine sehr begrenzte Bandbreite vorhanden ist, müssen Audiodateien stark komprimiert werden, um sie in erträglicher Zeit zu versenden. Dabei werden die Schwächen des menschlichen Ohrs ausgenutzt und Frequenzbereiche, die für

das menschliche Ohr nicht mehr hörbar sind einfach herausgefiltert und somit die Menge an Daten stark verringert. So sind mit dem MPEG Layer3 Format, kurz MP3, Kompressionsraten möglich, die ein Musikstück auf ein Zehntel der ursprünglichen Größe zusammenschrumpfen lassen. Die Tonqualität eines Telefons zum Beispiel ist sehr niedrig und entspricht einer Bandbreite von ca. 2,5 kHz, was einer Bitrate von 8kbps und einem Kompressionsfaktor von 96:1, sowie Mono-Sound entspricht. Quelle: [2]

3.2 MIDI

Der Begriff MIDI bedeutet Musical Instrumental Digital Interface und definiert die Schnittstelle zur Datenübertragung zwischen Musiker und einem elektronischen Klangerzeuger. Die in der Mitte der achtziger Jahre erfundene Schnittstelle hatte, wie in der Einleitung bereits beschrieben den großen Vorteil, daß sie die Kommunikation von Geräten verschiedener Hersteller standardisierte.

Bei einer normalen Gitarre werden mittels physischer Kraftausübung die einzelnen Saiten in Schwingungen versetzt. Hierbei erfolgt die Übertragung der Information direkt und in analoger Form. Bei einem Synthesizer funktioniert dies mit Hilfe der MIDI-Schnittstelle in digitaler Form. Der Anschlag auf der Tastatur des Instruments wird je nach Anschlagstärke und Dauer in MIDI-Bytes codiert und elektronisch zum Klangerzeuger übertragen, welcher die Bytes interpretiert und die entsprechenden Klänge gemäß den empfangenen Daten abspielt. Dabei können Tastatur und Klangerzeuger auch weiter voneinander entfernt sein. Anfangs waren der Menge der Übertragenen Informationen durch die Art des Übertragungsprotokolls gewisse Grenzen gesetzt. Es konnten maximal 31250 Baud/s übertragen werden. Mittlerweile wurde das "Media-Link-Protokoll" in den MIDI-Standard integriert, welches wesentlich höhere Übertragungsraten und größere Entfernungen ermöglicht.

Seitens der Hardware besteht ein MIDI-Stecker aus 5 Pins, von denen jedoch lediglich 3 benötigt werden. Bei der Informationsübertragung repräsentiert eine Spannung von 5 Volt die logische "0" und eine Spannung von 0 Volt eine logische "1". Besonderheit bei teuren Steckern ist eine optoelektronische Einheit, welche störende Brummschleifen oder Rauschen verhindert. Dabei wird je nach Spannung eine kleine Diode zum Leuchten gebracht, welche einen optischen Sensor beleuchtet, der dies wiederum interpretieren kann. Der Vorteil dieser Art der Übertragung liegt darin, das kein direkter Strom fließt, welcher gegebenenfalls zu unangenehmen Störeffekten führen kann.

Drückt der Musiker eine Taste auf seinem Synthesizer, werden die Informationen in sogenannte MIDI-Bytes umgewandelt. Ein MIDI-Byte enthält 10 Bit, wobei darin nur 8 echte Datenbits enthalten sind. Das erste und letzte Bit dient als Start- bzw. Stoppbit. Desweiteren wird zwischen Status- und Datenbytes unterschieden. Auf ein Statusbyte folgen entweder kein, ein oder zwei Datenbytes. Das Statusbyte übermittelt Informationen wie den

Befehlstyp und Adresse des Gerätes, an das die Datenbytes gerichtet sind. Ein Datenbyte dient zur Übermittlung von Werten eines bestimmten Parameters, also zum Beispiel, Dauer des Tons, Anschlagstärke oder Controller-Daten. Falls ein Datenbyte nicht ausreicht, wird die Information in maximal zwei Datenbytes codiert. Quelle: [3]

3.3 Hall/Echo

Hall:

Hall entsteht dadurch, dass Schall von Wänden oder sonstigen Hindernissen reflektiert wird und auf Umwegen nochmals beim Hörer ankommt. In Räumen wird der Schall von allen Seiten reflektiert und dadurch recht oft zum Hörer zurückgeworfen, bis seine Energie aufgebraucht ist und er verebbt. Man spricht in diesem Zusammenhang von Direktschall (von der Quelle direkt zum Hörer) und Diffusschall (Reflexionen/Hall). Das Verhältnis von Direkt- zu Diffusschall bestimmt, wie weit entfernt uns ein Klang erscheint. Ein Klang ohne Hall drängt sich in den Vordergrund, während ein höherer Hallanteil eine größere Entfernung der Schallquelle suggeriert. Wichtig für die Richtungswahrnehmung sind die ersten Reflektionen (Early Reflections). Die Schallquelle wird in der Richtung geortet, wo die ersten Reflektionen herkommen. Quelle: [7]

Hall-Algorithmen

Oft kann man in Soundeditoren zwischen einigen unterschiedlich klingenden Hall-Algorithmen auswählen, die Namen tragen wie Medium Room, Plate Hall, oder ähnliche.

Nachhallzeit

Die Zeit vom Einsetzen bis zum Verklingen des Halls (in ms). Sie ist abhängig von der Raumgröße und davon, wie viel reflektierende Flächen im Raum vorhanden sind.

Delay

Je größer der Raum, desto länger dauert es bis der Schall von den Wänden zum Hörer zurückgeworfen wird und um so größer ist die Verzögerung des Halls gegenüber dem Originalsignal. Die Verzögerungen sind oft gar nicht bewusst wahrnehmbar und beeinflussen trotzdem den Raumeindruck und die Richtungswahrnehmung. Es lohnt sich deshalb die Verzögerung richtig einzustellen (in ms).

High-Cut

Wenn Schall vielfach reflektiert wird oder auf dämmende Materialien trifft, verliert er als erstes hohe Frequenzen. Tiefe Frequenzen haben mehr Energie, können sich besser durchsetzen und weitere Strecken zurücklegen, so dass sie auch längere Nachhallzeiten erreichen. Dadurch wird der Hall meistens dumpfer als das Originalsignal, was mit einem High-Cut Filter für den Hall eingestellt werden kann. (Cut-Off-Frequenz in Hz)

Echo:

Obwohl Echos in freier Wildbahn eher selten vorkommen, sind die Effektgeräte zur künstlichen Herstellung von Echos um so verbreiteter. Sie kommen vor allem bei der Musikproduktion zum Einsatz. Es wird dabei das Signal mit einer Verzögerung ein oder mehrmals wiederholt.

Einstellungsmöglichkeiten:

Mix:

Lautstärkenverhältnis zwischen Originalsignal und den Wiederholungen.

Delay Time:

Zeit, nach der das Signal wiederholt wird. (in ms)

Feedback:

Bestimmt die Anzahl der Wiederholungen. (in %)

High-Cut:

Wie beim Hall, verliert auch beim Echo der Schall an Energie, was als erstes die hohen Frequenzen betrifft. Um ein natürliches Echo herzustellen, lassen sich die hohen Frequenzen pro Wiederholung Dämpfen. (Cut-Off-Frequenz in Hz)

3.4 Fade

Das Wort Fade kommt aus dem englischen und bedeutet soviel wie "verblassen". Es ist ein allgemeiner Ausdruck für eine Lautstärkeveränderung. Zum Beispiel beim Ausblenden aus einem Song. Mit einem Fade kann man auch alle Spuren oder einzelne Instrumente einblenden. Es wird entweder manuell oder mit Hilfe einer programmierten Computervoreinstellung vorgenommen. Quelle: [6]

Fade-In/Out

Engl.: Ein-/Ausblenden eines Signals (Audio oder Video). Diese Funktion, die bei einigen Consumergeräten (Camcorder, Minidisc-Recorder) fest eingebaut ist, wird normalerweise erst beim Schlussmix (der definitiven Abmischung) eingesetzt. Sie kann manuell z. B. über die Masterfader des Mischpultes erfolgen oder automatisch (bei den meisten digitalen Mischpulten und MIDI-Sequenzern programmierbar).

Fader

Engl.: Schieberegler, Flachbahnregler.

a) Bei analogen Mischpulten veränderbare Widerstände in Flachbauweise, die das durchfließende Signal abschwächen.

b) Bei digitalen Geräten ist der Effekt zwar derselbe, doch wird nicht das Signal selbst direkt beeinflusst, sondern mit der Faderbewegung eine Rechenoperation bewirkt. Diese Fader können auch zur Veränderung von Parameterwerten eingesetzt werden.

c) In Software findet man virtuelle Fader als Teil der graphischen Benutzeroberfläche. Sie werden mit der Maus bedient. Da dies für anhaltendes Arbeiten zu mühsam ist, wurden inzwischen Hardware-Fader entwickelt, mit denen diese virtuellen Software-Fader bedient werden können.

3.5 WAVE

Das Wave-Format ist ein Teil des RIFF Formats (Resource Interchange File Format) und jede Datei ist aus sogenannten "Chunks" aufgebaut, welche immer die gleiche Anordnung haben. Dieses Format ist, im Gegensatz zu komprimierten Audioformaten wie z.B. MP3, nicht verlustbehaftet und daher immer noch das Arbeitsformat bei der Bearbeitung von Audiodaten. Der Speicherbedarf von Wave-Dateien ist allerdings sehr groß, ca. 10 Megabyte pro Minute einer Sounddatei, bei Standardparametern (Auflösung 16 Bit, 44,1 kHz Samplingrate, Stereo). Durch Variation dieser Werte variieren auch Dateigröße und Qualität. Quelle: [8]

Struktur:

Eine Wave-Datei besteht aus mindestens aus einem WaveChunk, welcher genau einen FormatChunk und dahinter genau einen DataChunk enthält. Der WaveChunk oder RIFF-Chunk gibt an, dass es sich um eine Wave-Datei handelt. Der Formatchunk gibt, wie sein Name schon erahnen läßt, an, welches Format die Wave-Datei hat. Im Datachunk sind die eigentlichen Audiodaten enthalten. Optional vorhanden sind Pad-Chunk, welcher Füllwerte enthält, ein Cmd-Chunk (Command-Chunk), welcher evtl. Kommandozeilen an das aufgerufene Programm übergibt, Peak-Chunk, welcher den Wert für den Spitzenschalldruck in Pascal angibt und der Info-Chunk, welcher zusätzliche Informationen enthalten kann.

Folgende Tabelle gibt den Aufbau des Fmt-Chunks, also des Format-Chunks an, welcher insgesamt 16 Bytes lang ist.

Offset	Typ	Bezeichnung	Wert
0	unsigned short	CODE	Kodierung der Werte
2	unsigned short	CHANNELS	Anzahl der Kanäle
4	unsigned short	SAMPLERATE	Abtastrate in Hz
8	unsigned short	BYTESPERSECOND	Durchschnittliche Datenrate in Bytes/s
12	unsigned short	BYTESPERSTEP	Anzahl der Bytes per Zeitabschnitt
14	unsigned short	BYTESPERSAMPLE	Auflösung der Quantisierung

3.6 AVI

Das AVI-Format wurde von Microsoft entwickelt und steht für Audio-Video-Interleaving, was übersetzt bedeutet, das Audio- und Videosignale miteinander "verzahnt" abgespeichert werden. Die erste Definition des AVI-Formats ist ziemlich alt und wurde von Microsoft als einheitliches Format zur Wiedergabe von kurzen Videoclips geschaffen. Die ursprüngliche Festlegung sah 15 Bilder pro Sekunde bei einer maximalen Auflösung von 160 mal 120 Pixeln vor, was in Anbetracht heutiger Spezifikationen fast schon vorsintflutlich erscheint. Im Gegensatz zu anderen damals üblichen Animationsformaten wurde bei AVI die sogenannte Keyframe-Technik eingesetzt. Dabei wird lediglich jedes 12. bis 17. Bild (abhängig vom Bildinhalt) als Vollbild gespeichert. Für die dazwischen liegenden Frames werden nur die Unterschiede zum jeweils vorhergehenden Bild angegeben. Trotz der vermeintlichen Schwächen des Formats, verbreitete es sich sehr schnell, da es im Funktionsumfang von Windows enthalten war und die Treiber sowie Abspielsoftware kostenlos mitgeliefert wurden. Ein großes Problem bei AVI ist, dass es sich um ein echtes Dateiformat handelt und ein Videoclip nur abgespielt werden kann, wenn er vollständig vorhanden ist. Dadurch ist zum Beispiel kein Video-Streaming möglich, wie es heute oft im Internet verwendet wird. Dies wird auch zum Problem, wenn man etwa mehrere kurze Szenen aneinander fügen will. Abspielen läßt sich der Clip nur, wenn er zuvor als komplett intakte AVI-Datei abgespeichert wurde. Ein weiteres Problem ist, daß AVI ein offener Standard ist, der von verschiedenen Herstellern an eigene Bedürfnisse angepasst und modifiziert werden kann. Das hat zur Folge, dass schon geringe Abweichungen vom ursprünglichen Standard dafür sorgen können, daß eine AVI-Datei nur mit Hilfe eines bestimmten Video-Players abspielbar ist. Quelle: [4]

3.7 MPEG

MPEG bedeutet "Motion Pictures Expert Group" und ist ein genormtes Kompressionsverfahren. Der wesentliche Lösungsansatz dieses Formats liegt in der Erkenntnis, dass verschiedene aufeinanderfolgende Bilder redundante Informationen enthalten, die lokalisiert und entfernt werden können. Es werden im Bezug auf ein festgelegtes Referenzbild immer nur die Veränderungen abgespeichert. Ferner gliedert sich das MPEG-Format in vier verschiedene einzelne Formate auf, welche mit MPEG1 bis MPEG4 gekennzeichnet sind. Quellen: [4] und [5]

3.7.1 MPEG1

MPEG1 wurde für flüssige Video-Wiedergaben entworfen und weist daher die größten Kompressionsraten auf. Die MPEG-1-Komprimierung bzw. Dekomprimierung war ursprünglich ein hardwareabhängiges Verfahren. Es ist allerdings mittlerweile, dank der schnellen Prozessoren, auch eine Komprimierung mittels Software möglich.

Die Spezifikationen:

- 25 Frames, keine Information zwischen den Halbbildern
- Auflösung von 352 x 288 (SIF-Auflösung)
- 1,2 bis 3 MBits/s (1,2 MBits/s bei einer Video CD)

3.7.2 MPEG2

Der wesentliche Unterschied zwischen MPEG-1 und MPEG-2 besteht darin, daß MPEG-2 besser mit dem beim Fernsehen eingesetzten Zeilensprungverfahren (Interlace) umgehen kann. Das Geheimnis von MPEG-2 liegt in der Kompression auf höchster Qualitätsstufe, so daß Filmmaterial nahezu 1 zu 1 in Studioqualität bearbeitet und editiert werden kann. Konsequenterweise etablierte sich MPEG-2 zu einem Broadcaststandard. Bei einer reinen I-Framecodierung läßt sich MPEG-2 sogar im Schnittbetrieb einsetzen.

Die Spezifikationen:

- 250 Fields / 25 Frames
- 2High-Auflösung 1920 x 1152 bis zu 80 MBits/s (High Definition TV - HDTV)
- 2High-Auflösung 1440 x 1440 bis zu 60 MBits/s (HDTV)
- 2Main-Auflösung 720 x 576 bis zu 15 MBits/s (digitales TV und DVD-Video)
- 2Low-Auflösung 352 x 288 bis zu 4 MBits/s (S-VHS, SIF)

3.7.3 MPEG2 Kompression mit Hilfe von I- P- und B-Frames

Wie bereits angesprochen, ist der Grundgedanke der Kompression, redundante Informationen zu lokalisieren und zu entfernen. Bei der MPEG2 Kompression werden in gewissen Abständen Referenzbilder, sogenannte I-Frames erfaßt und nur die Änderungen relativ zu diesen I-Frames gespeichert. Dazu wird über das Bild ein Raster gelegt (das Bild wird in Makroblöcke mit 16x16 Pixeln unterteilt). Kommt ein Makroblock in beiden Bildern vor, wird nur ein Bewegungsvektor gespeichert. Kommt der Block nicht in dem I-Frame vor (neues Bildelement), dann wird dieser Block zusätzlich gespeichert. Das leisten die sogenannten P-Frames (predictive). P-Frames speichern also Veränderung und Neues. Aber auch P-Frames enthalten noch gewisse Redundanzen. Ein B-Frame kann dieses Problem lösen. Es speichert eine Referenz zu einem zukünftigen Bild (der Teil der vorher verdeckt war). Die dazu erforderliche Analyse braucht also einen Blick auf zukünftige sowie auf vergangene Bilder, also bidirektional. B-Bilder bringen die stärkste Kompression, also niedrigen Speicherbedarf und geringen Datenstrom.

3.7.4 MPEG3

Der Teil des MPEG-3-Standard, der für High Definition TV-Qualität (HDTV) vorgesehen war, wurde mittlerweile in den MPEG-2-Standard implementiert. MPEG-3 ist nicht mit MP3 zu verwechseln.

3.7.5 MPEG4

MPEG 4 ist eine Weiterentwicklung des MPEG-2-Formats und befindet sich seit 1996 in der Entwicklung. Obwohl MPEG-4 ursprünglich als ein Codierungsstandard für audiovisuelle Daten mit sehr niedriger Bitrate gedacht war, diente die Entwicklung weit mehr Zwecken als lediglich dem Streaming von linearen Mediendaten bei Internet- und drahtlosen Anwendungen. MPEG-4 stellt z.B. effiziente Mechanismen zur Komprimierung und Distribution interaktiver Medieninhalte bereit. Außerdem verfügt MPEG-4 über 3D-Potenziale, um künstliche Intelligenzen (artificial intelligence - AI) zu visualisieren oder Avatare darzustellen, z.B. im Rahmen von Videokonferenzen. Dies geschieht mit Hilfe einer VRML-ähnlichen Skriptsprache. Zusätzlich gibt es eine Java-Schnittstelle zur Ansteuerung externer Hardware.

Die Kompressionsrate als solche ist bei MPEG-4 nicht höher als bei MPEG-2, aber "Sprites" können besser komprimiert werden, weil dem Codier-Mechanismus dafür wesentlich mehr Zeit zur Verfügung steht. Eventuell kann dabei sogar auf Wavelets umgeschaltet

werden. Die Skriptsprache ermöglicht es, in wenigen Bytes Operationen wie "verschieben" wesentlich schneller durchzuführen, als es die digitalisierte komprimierte Form der gleichen Operation ermöglichen würde. (Mit Hilfe dieser "Sprites" können beliebig konturierte Standbilder über sich bewegende Bilder geschoben werden.)

Microsofts erste Implementierung von MPEG-4 ist seit Windows 98 auf jedem Windows-PC installiert und beschränkt die maximale Datenrate auf 256 Kilobit pro Sekunde. Für ein ruckelfreies Video in ansprechender Qualität ist das zu wenig. Durch den DivX-Hack wurde die Erhöhung der Bitrate auf bis zu 6000 Kilobit pro Sekunde möglich. Doch eine so hohe Bitrate ist gar nicht notwendig: 600 Kilobits pro Sekunde sind völlig ausreichend.

4 Fazit

Nach kurzer Einarbeitungszeit hat es uns sehr viel Spaß gemacht, mit Hilfe der Editoren die Samples zu erstellen. Es ist beeindruckend, wie mächtig die Werkzeuge zur elektronischen Bearbeitung von Audiodateien geworden sind und welche Möglichkeiten sie dem Benutzer bieten.

Abschließend möchten wir noch darauf hinweisen, dass zu dem im c't Heft 18 des Jahres 2002 beschriebenen Versuch auch eine Internetseite existiert, welche zwar nicht die komplette Versuchsbeschreibung enthält, zumindest aber die Klangbeispiele, so wie sie bei der Durchführung der Aufgabe erstellt werden sollen. Ferner ist die Webseite eine gute Informationsquelle und enthält ein paar interessante Links im Zusammenhang mit Spracheingaben.

Der Link zur Seite lautet:

<http://heise.de/ct/ftp/projekte/sprachsynthese>

Literatur

- [1] <http://www.uni-karlsruhe.de/RZ/Dienste/GVM/2.multimedia.1.chance.shtml>
- [2] <http://www.scram.de/musik/sonstige/hilfe/wasist/haupt.htm>
- [3] <http://kunstbank.waidhofen.at/mitglied/junker/publikationen/diplomarbeit/diplom1.htm>
- [4] <http://www.glossar.de>
- [5] <http://www.computerbecker.com/hintergrundinfos/mpeg.html>
- [6] http://www.gccarstensen.com/glossar/glossar_f.htm

[7] http://www.movie-college.de/filmschule/ton/hall_echo.htm

[8] <http://www.emt.uni-linz.ac.at/roland/wave/node68.html>