

HANSER

Leseprobe

Michael Steppat

Audioprogrammierung

Klangsynthese, Bearbeitung, Sounddesign

Herausgegeben von Ulrich Schmidt

ISBN (Buch): 978-3-446-43222-2

ISBN (E-Book): 978-3-446-44198-9

Weitere Informationen oder Bestellungen unter

<http://www.hanser-fachbuch.de/978-3-446-43222-2>

sowie im Buchhandel.

2

Audiosignale



Fragen, die dieses Kapitel beantwortet:

- Was ist Schall?
- Was versteht man unter einem Signal?
- Wie wird ein akustisches Signal in ein elektrisches Signal umgewandelt?
- Wie ist ein Signal zusammengesetzt?
- Was versteht man unter einem Ton?
- Warum können analoge Signale nicht in Computern gespeichert werden?
- Wie wird ein analoges Signal in eine digitale Zahlenfolge umgewandelt?
- Was versteht man unter Abtastung und wie entstehen Abtastfehler?
- Was ist Quantisierung?
- Wie ist eine Audiodatei im WAV-Format aufgebaut? Welche Formate gibt es?
- Wie lässt sich eine Datei mit einem Ton erzeugen?
- Wie kann die Amplitude von Audiodaten bearbeitet werden?
- Wie werden Audiodatenströme gemischt?
- Was versteht man unter Normalisierung und wie wird diese durchgeführt?
- Was unterscheidet eine AIFF-Datei von einer WAV-Datei?

2.1 Akustische Signale

Mit **Schall** werden alle Ereignisse bezeichnet, die durch das Gehör wahrgenommen werden. Physikalisch gesehen ist er eine Abfolge von Luftdruck- und Dichteschwankungen, welche sich in Form von Wellen von einer Schallquelle her ausbreiten [Kutt04]. Diese Form der Wellenausbreitung lässt sich zum Beispiel am Ufer eines möglichst ruhigen Gewässers veranschaulichen, indem man einen Stein in das Wasser wirft. Die Wellen breiten sich dann von der Einwurfstelle ringförmig aus [Vorl08]. Ursache für eine Schallentstehung

können mechanische Schwingungen sein, wie z.B. das angeschlagene Fell einer Trommel, eine angezupfte Saite oder sich plötzlich verändernde Luftvolumen wie z.B. das Knallen eines Sektkorkens oder eines Feuerwerkskörpers. Auch der auf einen Blitz folgende Donner entsteht durch eine Volumenveränderung. Die beim Blitz entstehende Hitze dehnt die Luft schlagartig aus und erzeugt eine Druckwelle. Analog zu den mit der Schallausbreitung verbundenen Luftdruckschwankungen und der Dichte ändert sich bei allen Schallereignissen in der Welle auch die Lage der Luftmoleküle. Sie bewegen sich mit einer Wechselgeschwindigkeit um ihre Ausgangslage. Diese Größe wird Schallschnelle genannt und steht im Zusammenhang mit dem Schalldruck.

Wird bei einem Schallereignis eine Tonhöhe wahrgenommen, spricht man umgangssprachlich von einem **Ton**. Dieser kann gesungen oder auf einem Musikinstrument gespielt werden. Jede Tonhöhe hat eine der Oktave und Tonleiter entsprechende Bezeichnung. Eine C-Dur-Tonleiter besteht z.B. aus den Tonhöhen: c, d, e, f, g, a, h, welche den weißen Tasten auf dem Klavier entsprechen. Wird keine Tonhöhe wahrgenommen, so spricht man von einem Geräusch. Beide sind im technischen Sinn Klänge, welche sich aus einzelnen Tönen zusammensetzen. Ein **Klang** entsteht durch die Überlagerung einzelner Töne.

Der Begriff des Tones ist im technischen Sinne anders zu verstehen. Als **Ton** bezeichnet man dort eine periodisch ablaufende sinusförmige Schwingung. Diese lässt sich z.B. mit einer Stimmgabel erzeugen. Sie besteht aus einer positiven Flanke (Auslenkung > 0) mit einem Maximum, welcher nach einem Nulldurchgang (Auslenkung = 0) in eine negative Flanke (Auslenkung < 0) mit einem Minimum übergeht und lässt sich mit den drei Parametern Amplitude, Frequenz und Phasenlage beschreiben: Die **Amplitude** gibt den Betrag der maximalen Auslenkung in positiver und in negativer Richtung an [Webe85]. Je größer die Amplitude, desto lauter wird der Sinuston wahrgenommen. In Bild 2.1 ist eine Sinusschwingung mit den Amplitudenwerten 0,25; 0,5 und 1,0 dargestellt.

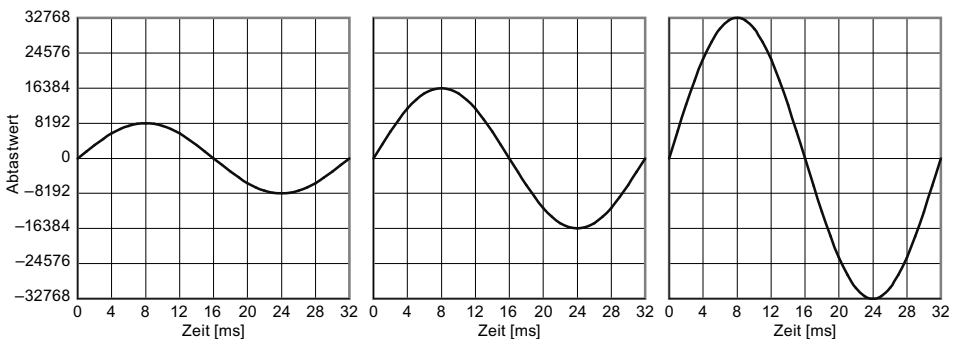


BILD 2.1 Sinusschwingung mit verschiedenen Amplituden

Bei Sinustönen hat die Frequenz einen Einfluss auf die wahrgenommene Tonhöhe. Mit der **Frequenz** wird die Anzahl der Periodendurchläufe pro Sekunde angegeben. Je höher die Frequenz, desto höher auch der wahrgenommene Ton. Das menschliche Gehör nimmt im Idealfall im sogenannten Hörbereich Frequenzen von 20 bis 20000 Hz wahr. Als **Infraschall** bezeichnete Töne unter 20 Hz werden als pulsierende Luftdruckschwankungen wahrgenommen, welche ab 16 bis 20 Hz in eine kontinuierlich wahrgenommene Tonhöhe übergehen. Diese Grenze wird als untere **Hörgrenze** bezeichnet. Die obere Hörgrenze, welche bei Kleinkindern noch bis zu 20000 Hz beträgt, nimmt mit zunehmendem Lebensalter ab

und kann individuell sehr verschieden sein [Meye94]. Bei jungen Erwachsenen geht man von einer oberen Hörgrenze von 16000 Hz aus, welche sich pro Lebensjahrzehnt durchschnittlich um ca. 1000 bis 2000 Hz nach unten verschiebt. Das Bild 2.2 zeigt drei Sinusschwingungen mit den Frequenzen 31,25; 62,5 und 73,75 Hz. Der abgebildete Zeitraum beträgt immer 32 ms.

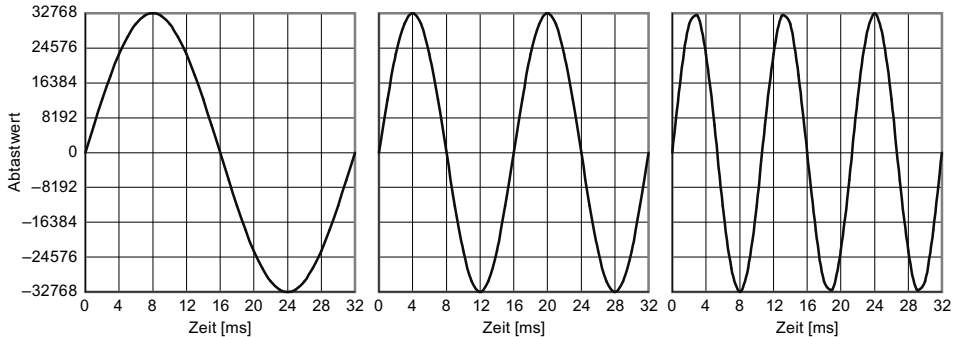


BILD 2.2 Sinusschwingung mit unterschiedlichen Frequenzen

Mit der **Phase** als drittem Parameter wird die Anfangsposition der Auslenkung zu einem bestimmten Zeitpunkt bezeichnet. Da man sich die Sinusfunktion als Auslenkung eines rotierenden Zeigers vorstellen kann, dessen Position mit einem Winkel relativ zu einem Bezugspunkt angegeben werden kann, wird auch die Phase in einem Winkel angegeben. Hierbei geht man von einem linksdrehenden Zeiger aus, dessen Anfangsposition 0° der Position 3 Uhr entspricht. Einer Vierteldrehung nach links entspricht 90° , einer Umdrehung 360° [Veit74].

In der Literatur wird für die Winkelangabe fast ausschließlich das Bogenmaß verwendet. Dies ist das Verhältnis eines Kreisbogens zu dessen Radius. Dabei entsprechen 2π einer ganzen Umdrehung. Eine Umrechnung vom Gradmaß in das Bogenmaß b erfolgt dadurch, indem man das gegebene Gradmaß β mit π multipliziert und dann durch 180 teilt: $b = \beta * \pi / 180$. Um die Konstante 2π nicht immer mit angeben zu müssen, wird bei der Frequenzangabe in der Literatur fast ausschließlich die Kreisfrequenz ω verwendet. Diese ist die mit dem Faktor 2π multiplizierte Frequenz [Bart11].

Mit diesen drei Parametern sind die Eigenschaften einer Sinusschwingung hinreichend beschrieben. In Bild 2.3 werden die unterschiedlichen Phasenlagen veranschaulicht.

Ein **akustisches Signal** ist eine zeitveränderliche messbare physikalische Größe. Die Schwingungen der Stimmgabel drücken die umgebenden Luftmoleküle periodisch zusammen und ziehen sie auseinander. Dies führt zu einer Schwankung des Luftdrucks, welcher in der Einheit Pascal gemessen wird.

Zeichnet man diese Luftdruckschwankungen auf, erhält man eine Darstellung des Schalldrucks in Abhängigkeit von der Zeit. Diese Funktion wird **Zeitfunktion** genannt und ist die in Audioeditoren gebräuchlichste Darstellung zur Bearbeitung von Audioinhalten. In ihr lässt sich auf einen Blick erkennen, an welchen Passagen Schallereignisse stattfinden, welche Aussteuerung (Amplitude) sie haben und wo sich Pausen befinden. Das Auffinden der Pausen bzw. der Nulldurchgänge ist für die Schnittbearbeitung von Bedeutung, da nur in den Nulldurchgängen geschnitten werden darf, um Klickgeräusche zu vermeiden, die

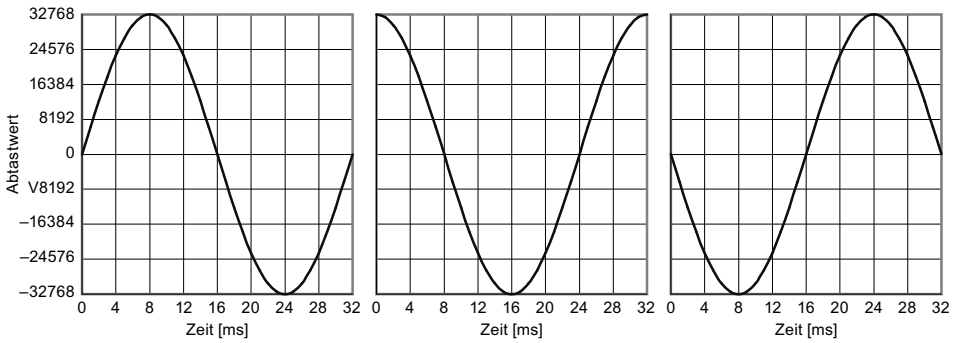


BILD 2.3 Sinusschwingungen mit den Phasenlagen 0, 90, 180° (im Bogenmaß: 0, $\frac{\pi}{2}$, π)

durch plötzliche Signalsprünge hervorgerrufen werden. Die Zeitfunktion kann auch mit dem Oszilloskop sichtbar gemacht werden [Brau97].

■ 2.2 Schallwandlung

Um akustische Signale aufzunehmen und mit Audiosoftware bearbeiten zu können, müssen diese in mehreren Schritten gewandelt werden. Als Luftschall vorliegende Schallergebnisse werden aufgenommen, in elektrische Signale gewandelt und daraus Messdaten erzeugt, welche auf dem Rechner als Zahlenfolgen gespeichert und weiterverarbeitet werden können.

Der erste Schritt ist die Wandlung des akustischen Signals in ein elektrisches Signal über ein Mikrofon, welches eine zu dem Schalldruck proportionale elektrische Spannung erzeugt. Diese wird im Audioadapter oder in einem vorgeschalteten Mischpult verstärkt und zu einem **Analog-Digital-Wandler** (AD-Wandler) geleitet, welcher das in kontinuierlicher Form vorliegende Signal abtastet und in zeitdiskrete Werte wandelt. Die Abtastwerte werden digital entweder als 8, 16, 24, 32 oder 64-bit Werte ausgegeben und können nun in einem digitalen Signalverarbeitungssystem weiterverarbeitet werden [Dick08].

Der Abtastvorgang erfolgt in regelmäßigen zeitlichen Abständen in einer festgelegten **Abtastrate**. Gebräuchlich sind auf modernen Systemen die Abtastraten 44100, 48000, 96000 und 192000 Hz. Neben jeder theoretisch beliebigen Abtastrate waren früher auch die Formate: 8000, 11025, 16000, 22050 und 32000 Hz gebräuchlich. Da diese von der Bandbreite nicht den gesamten Hörbereich abdecken, geht dies auf Kosten der Klangqualität. Die Verwendung dieser Formate beschränkte sich daher in der Regel auf Sprachaufnahmen und diente hauptsächlich dazu, Speicherplatz auf der Festplatte einzusparen. Mit zunehmender Speicherkapazität der Festplatten haben diese Formate an Bedeutung verloren.

Die Abtastung entspricht der zeitlichen Momentaufnahme einer Schwingung. Veranschaulichen lässt sich dies mit einem kleinen Gedankenexperiment, in dem man die Auslenkung eines einfachen Schwingers betrachtet, welcher aus einem kleinen Gewicht besteht, das an einer Feder hängt. Zieht man an dem Gewicht und lässt es anschließend los, so pendelt das Gewicht in einer bestimmten Frequenz um seine Ausgangslage herum [Boru84].

Befestigt man an dem Gewicht einen kleinen Stift, der auf ein Lineal zeigt, kann man die Auslenkung messen. Die einzelnen Werte lassen sich erfassen, wenn die Bewegung mit einer Videokamera aufgezeichnet wird. Jedes Einzelbild liefert einen Messwert zu einem bestimmten Zeitpunkt. Durch die festgelegte Bildwechselfrequenz findet die Messung in konstanten Zeitabständen statt.

Nach dem gleichen Prinzip erfolgt die Abtastung der elektrischen Spannungswerte im AD-Wandler. In regelmäßigen Abständen wird die Spannung gemessen in einen Zahlenwert mit einer festgelegten Bitbreite übertragen und gespeichert. Bei der Abtastung ist zu beachten, dass die Abtastfrequenz immer höher als das Zweifache der höchsten zu übertragenden Frequenz ist. Bei Nichtbeachtung treten sogenannte Alias-Signale in Form von Spiegelfrequenzen auf. Die **Spiegelfrequenz** ist die Differenz von Abtastfrequenz und Signalfrequenz. Wird zum Beispiel ein 1000-Hz-Signal abgetastet, so muss die Abtastfrequenz größer als 2000 Hz sein.

Eine zu niedrige Abtastung mit nur 1500 Hz würde dazu führen, dass eine Spiegelfrequenz von 500 Hz entsteht. Um sicherzustellen, dass das Signal keine Anteile enthält, deren Frequenzen größer/gleich der halben Abtastfrequenz sind, wird das Signal vor der Wandlung mit einem Tiefpassfilter gefiltert, welches diese Anteile abschneidet. Für den Hörbereich des menschlichen Gehörs bedeutet dies, dass die Abtastrate größer als 40000 Hz sein muss, um alle hörbaren Frequenzen zu übertragen. Das Tiefpassfilter wird als Antialiasing-Filter bezeichnet [Broe99].

Da Tiefpassfilter nicht exakt ab einer bestimmten Frequenz alle höherliegenden Frequenzen abschneiden, sondern über einen Übergangsbereich verfügen, gilt für praktische Anwendungen, dass die Abtastfrequenz größer als das 2,2-fache der maximal zu übertragenden Frequenz sein muss. Daher beträgt die Abtastrate für Audio-CDs z.B. 44100 Hz.

■ 2.3 Quantisierung

Im AD-Wandler wird die Spannung des Abtastwerts mit einer Vergleichsspannung über einen Komparator verglichen. Dieser liefert einen H-Pegel, wenn beide Spannungen übereinstimmen. Die Vergleichsspannung wird stufenweise erzeugt. Bei einer 8-bit-Wandlung besteht sie aus 256 Stufen, bei einer 16-bit-Wandlung aus 65536 Stufen. Die Verwendung dieser stufenweise erzeugten Menge an Werten wird Quantisierung genannt. In der Audio-technik sind Quantisierungen mit 8, 16 und 24-bit üblich. Da das Signal durch die Quantisierung nicht mehr wertekontinuierlich ist, entstehen bei der Wiedergabe Signalverzerrungen, welche als Quantisierungsfehler bezeichnet werden. Diese werden als Rauschen hörbar und daher auch als **Quantisierungsrauschen** bezeichnet. Je niedriger die Quantisierung, desto größer wird die Abweichung vom Originalsignal, die zu einem Anstieg des Rauschanteils führt. Das Quantisierungsrauschen wird mit dem Signal-Rauschabstand SNR (signal to noise ratio) in Dezibel angegeben. Er ist von der Anzahl N der Bits abhängig [Broe99].

$$SNR = N \cdot 6,02dB + 1,76dB \quad (2.1)$$

Diese Formel kann als Faustformel vereinfacht werden: Mit jedem weiteren Bit nimmt der Signal-Rauschabstand um 6 dB zu und beträgt somit bei 8-bit-Aufnahmen ca. 50 dB, bei

16-bit-Aufnahmen ca. 98 dB und bei 24-bit-Aufnahmen ca. 146 dB. Welchen Einfluss die Reduktion einer Aufnahme um 1-bit hat, kann man sich anhand des folgenden Beispiels verdeutlichen: Angenommen eine Abtastfolge besteht aus den Ganzzahlwerten: 0, 4, 7, 8, 9, 11. Teilt man diese durch 2, was einer Reduktion um 1-bit entspricht, erhält man die Folge: 0, 2, 3, 4, 4, 5. Bei der anschließenden Multiplikation mit 2, was einer Erhöhung um 1 bit entspricht, lässt sich die ursprüngliche Folge nicht mehr gänzlich rekonstruieren. Alle ungeradzahligen Werte sind um 1 reduziert. Als Folge erhält man: 0, 4, 6, 8, 8, 10. Zwischen der ursprünglichen Abtastfolge und der rekonstruierten Folge besteht die Differenz: 0, 0, 1, 0, 1, 1. Diese Differenz entspricht dem durch die Reduktion um 1-bit entstandenen Quantisierungsrauschen.

Der Quantisierungsfehler lässt sich auch veranschaulichen, wenn man sich vorstellt, dass z.B. der Wertebereich von 1–12 auf einer vierstufigen Skala quantisiert werden soll, was einer 2-bit-Quantisierung entspricht. Dabei werden die Werte 1, 2, 3 der ersten Stufe, die Werte 4, 5, 6 der zweiten Stufe, die Werte 7, 8, 9 der dritten und die Werte 10, 11, 12 der vierten zugeordnet werden. Hierdurch entsteht eine Mehrdeutigkeit, die nicht mehr exakt rekonstruiert werden kann. Eine Quantisierungsstufe stellt also immer einen Wertebereich dar und keinen absolut zugeordneten Wert.

■ 2.4 Speicherung von Audioinhalten

Die Speicherung digitaler Audioinhalte erfolgt in Dateien mit unterschiedlichen Formaten. In der Datei werden neben den Audiodaten, welche in komprimierter oder unkomprimierter Form vorliegen, auch mindestens die Formatdaten gespeichert. Auch Copyrightinformationen, Markierer, Kommentare und Gerätedaten lassen sich in diesen Dateien speichern. Die einzelnen Datentypen werden in Blöcken angeordnet. Derartige Dateien werden als **Containerdatei** bezeichnet. Auch andere Multimediadateien wie Text-, Bild- und Videodateien werden in diesem Format abgespeichert.

Um eine Audiodatei abspielen zu können, müssen dem Programm und auch der Soundkarte verschiedene Formatdaten übergeben werden, welche in Tabelle 2.1 aufgelistet sind.

Für die Speicherung unkomprimierter Audiodaten ist das **Audio Interchange File Format** (mit der Endung: *.aiff) und das **Windows Wave Format** (Endung: *.wav) sehr stark verbreitet. Das Audio Interchange File Format (AIFF) basiert auf dem 1985 von Firma Electronic Arts entwickelten Interchange File Format (IFF). Dieses Format zeichnet sich dadurch aus, das darin alle Datentypen in einer dreigliedrigen Struktur angelegt sind. Diese besteht erstens aus einem Bezeichner, der in Form eines Codes angibt, um welche Art von Datentyp es sich handelt.

TABELLE 2.1 Formatdaten und eine Auswahl möglicher Audioformate

Formatdaten	Mögliche Werte
Format	PCM, ADPCM, MP3, OGG, WMA
Abtastrate	8, 11,025, 16, 22,050, 32, 44,1, 48, 96, 192 kHz
Bitanzahl je Sample	8, 16, 24, 32, 64 bit
Anzahl der Kanäle	Mono, Stereo, Surround, Multichannel

Als zweites Element wird die Länge der gespeicherten oder zu übertragenden Daten angegeben, danach folgen die eigentlichen Daten. Diese Struktur wird **Type-Length-Value** (TLV) genannt und wird auch bei Übertragungen in Netzwerken als Protokoll genutzt. Die Anordnung der TLV-Blöcke in Form sogenannter Chunks kann reihenweise oder verschachtelt (d.h. die Blöcke enthalten selbst untergeordnete TLV-Blöcke als Subchunks) erfolgen. Die Bezeichner haben eine Länge von 4 Bytes und werden daher auch als **Four Character Code** (Abk.: FourCC) bezeichnet. Das AIFF-Format wurde von der Firma Apple als Audiodateiformat für den Macintosh auf der Grundlage des IFF-Formates weiterentwickelt. Es besteht aus einem FORM-Chunk, welcher als Subchunk einen Common Chunk mit der Kennung „COMM“ für die Formatdaten und eine Sound Data Chunk mit der Kennung: „SSND“ für die Audiodaten enthält [Bour96].

Mit der Entwicklung von Windows 3.1 wurde im Jahr 1991 das IFF-Format ebenfalls von den Firmen Microsoft und IBM übernommen und darauf aufbauend das **Resource Interchange File Format** (RIFF) entwickelt. Dieses Format findet Anwendung für Videodateien, Audiodateien und Midi-Dateien. Eine Windows-Wave-Datei enthält mindestens einen RIFF-Chunk und die beiden Subchunks „fmt “ und „data“. Der fmt-Chunk enthält als viertes Byte ein Leerzeichen und enthält die Formatdaten, der data-Chunk die Audiodaten [Demb96]. Eine Übersicht des Aufbaus zeigt Bild 2.4.

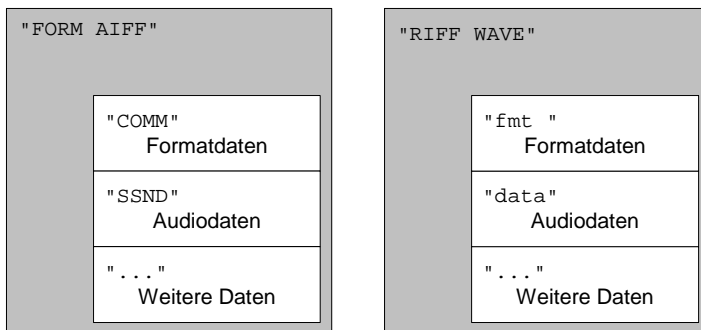


BILD 2.4 Aufbau von AIFF- und Windows-Wave-Dateien

Neben den unterschiedlichen Bezeichnungen der Chunks unterscheidet sich auch die Struktur der in den Chunks enthaltenen Daten. Ein ganz entscheidender Unterschied ist auch die Reihenfolge, in welcher Datentypen mit einer Länge von mehr als einem Byte gespeichert werden. Bei den im Macintosh verwendeten Motorola-Prozessoren der Serie 68000 werden die Bytes mit den höherwertigen (most significant) Bits zuerst gespeichert und mit zunehmender Speicheradresse die Bytes der niederwertigeren (least significant) Bits.

Eine angenommene Hexadezimalzahl mit dem Wert 0x1A2B3C4D wird in der Datei oder im Speicher auch in der Reihenfolge: 1A 2B 3C 4D gespeichert. Dies erleichtert das Auslesen der Werte mit einem Hexeditor. Diese Byte-Reihenfolge wird **Big-Endian** genannt, das direkt übersetzt „Groß-Ender“ bedeutet. Da die AIFF-Dateien zunächst überwiegend auf Macintosh-Rechnern verwendet wurden, werden die Datentypen short, int und long in dieser Byte-Reihenfolge gespeichert.

In den von Windows verwendeten Intel-X86-Prozessoren ist die Byte-Reihenfolge umgekehrt, d. h. der Wert 0x1A2B3C4D wird im Speicher wie folgt angelegt: 4D 3C 2B 1A. Das Byte mit den niederwertigsten Bits steht an erster Stelle, das Byte mit den höherwertigen an letzter. Diese Reihenfolge wird dann als **Little Endian** bezeichnet, welches mit „Klein-Ender“ übersetzt werden kann.

Die Bezeichnungen gehen auf Jonathan Swifts Roman „Gullivers Reisen“ zurück. In diesem wird von den Bewohnern des Landes „Liliput“ erzählt, wo zwei rivalisierende Gruppen leben, welche ihre Frühstückseier an unterschiedlichen Enden aufschlagen. Die eine Gruppe favorisiert das kleine spitze Ende (little ender), die andere das dicke große (big ender). Die Reihenfolge „Little Endian“ findet in den RIFF-Dateien ihre Anwendung. Im Little-Endian-Format gespeicherte Daten sind schwerer zu lesen, da immer erst die Reihenfolge der Bytes umgestellt werden muss.

Vorteile bietet das Format bei der Typumwandlung. Muss eine Ein-Byte-Zahl, wie z.B. vom Typ char in eine Zwei-Byte-Zahl gewandelt werden, so kann diese im Speicher an ihrem Platz bleiben. Es muss nur ein nachfolgendes Byte mit dem Wert null angehängt werden. Bei einer Typumwandlung in der Big-Endian-Reihenfolge muss dieses Byte um eine Speicheradresse verschoben werden.

■ 2.5 Aufbau und Programmierung einer Windows-Wave-Datei

Die Windows-Wave-Datei besteht in ihrer minimalen Anordnung aus einem Chunk mit den Formatdaten, welche auch als **Kopfdaten** oder **Header** bezeichnet werden und den Audiodaten in Form von Abtastwerten als Zahlenfolge. Im folgenden Beispiel soll dieses Format mit einer Datei im unkomprimierten PCM-Format mit einer Abtastrate von 44100 Hz, einer Wortbreite von 16-bit und einem Audiokanal vorgestellt werden. Zur Analyse der Struktur von Audiodateien sind Hexeditoren am besten geeignet.

Der RIFF-Chunk beginnt mit der 4 Bytes langen Kennung „RIFF“, welche im Textfeld des Hexeditors lesbar ist. Die folgenden 4 Bytes enthält die im der Little-Endian-Format gespeicherte Länge des RIFF-Chunks. Verwendet wird hierfür der Datentyp int. Bei der Längenangabe eines Chunks werden immer die 8 Bytes, welche für die Kennung und die Längenangabe reserviert sind, nicht mitgezählt. Daher werden diese von der Gesamtlänge des Chunks (in diesem Fall von der gesamten Datei) abgezogen. Somit ist die Länge = Gesamtlänge - 8. Ein RIFF-Chunk enthält noch ein weiteres Datenfeld, welches einen Bezeichner enthält, der angibt, um welchen Datentyp es sich handelt. Der Bezeichner „WAVE“ gibt an, dass es sich bei der RIFF-Datei um eine Audiodatei handelt.

Der Chunk mit den Formatangaben beginnt mit der Kennung „fmt“. Er besteht zwar nur aus drei Buchstaben, ist aber dennoch ein Four Character Code, da das vierte Zeichen ein Leerzeichen ist. Die folgenden 4 Bytes geben als 32-bit-Wert die Länge des fmt-Chunks an. Die Länge des fmt-Chunks ist abhängig vom Format der Audiodaten. Für das unkomprimierte PCM-Format wird ein 16 Byte langer Datenblock verwendet. In den Programmiersprachen C/C++ lassen sich Datenblöcke mit Daten unterschiedlicher Wortlänge in Struk-