

michael STEPPAT



# AUDIO- PROGRAMMIERUNG

// KLANGSYNTHESE

// BEARBEITUNG

// SOUNDDESIGN

HANSER



**Im Internet:** Aktuelle Ergänzungen  
und weitere Informationen

# Inhalt

<b>1</b>	<b>Einleitung</b> .....	<b>13</b>
1.1	Was ist Audioprogrammierung? .....	13
1.2	An wen richtet sich dieses Buch? .....	17
1.3	Wie kann man mit diesem Buch arbeiten? .....	17
<b>2</b>	<b>Audiosignale</b> .....	<b>19</b>
2.1	Akustische Signale .....	19
2.2	Schallwandlung .....	22
2.3	Quantisierung .....	23
2.4	Speicherung von Audioinhalten .....	24
2.5	Aufbau und Programmierung einer Windows-Wave-Datei .....	26
2.6	Aufbau und Programmierung einer AIFF-Datei .....	32
2.7	Digitale Verstärkung und Dezibelwerte .....	34
2.8	Normalisierung von Audiodateien .....	38
2.9	Mischen von Audiodatenströmen .....	39
2.10	Zusammenfassung .....	40
2.11	Übungsaufgaben .....	40
<b>3</b>	<b>Datenreduktion und Kompressionsalgorithmen</b> .....	<b>42</b>
3.1	Verfahren zur Datenreduktion .....	42
3.2	Verlustlose Datenkompression .....	43
3.3	Verlustbehaftete Verfahren .....	46
3.3.1	Deltacodierung beim Format DPCM .....	46
3.3.2	Codierung im Frequenzbereich .....	46
3.3.3	Gehörphysiologische Grundlagen .....	48
3.3.4	Aufbau von MP3-Dateien .....	49
3.4	Programmierung von Codecs .....	51
3.5	Audio Compression Manager .....	53

3.6	LAME-Codec .....	56
3.7	DirectShow .....	59
3.8	Übungsaufgaben .....	64
<b>4</b>	<b>Bearbeitungswerkzeuge für Audiodaten .....</b>	<b>65</b>
4.1	Programme für Audibearbeitung .....	65
4.1.1	Wavosaur .....	66
4.1.2	Goldwave.....	67
4.1.3	Audacity .....	67
4.1.4	Kommerzielle Software für den semiprofessionellen und professionellen Einsatz.....	68
4.2	Virtual Studio Technologie (VST).....	69
4.2.1	Gemeinsamkeiten und Unterschiede von Java und C++ .....	70
4.2.2	VST-SDK von Steinberg.....	72
4.2.3	VST-Hostanwendung .....	74
4.2.4	Delayeffekte mit der Beispielanwendung ADelay .....	74
4.2.5	Die Benutzerschnittstelle VSTGUI.....	77
4.3	Klangfilter.....	77
4.3.1	Tief- und Hochpassfilter, FIR-Filter erster Ordnung .....	83
4.3.2	Bandpass und Bandsperre als FIR-Filter zweiter Ordnung.....	84
4.3.3	IIR-Filter mit unendlicher Impulsantwort.....	85
4.3.4	Regelbare Klangfilter .....	87
4.4	Übungsaufgaben .....	90
<b>5</b>	<b>Audioanalyse.....</b>	<b>91</b>
5.1	Visualisierung von Audiodaten .....	91
5.1.1	Zeitfunktion.....	92
5.1.2	Messung von Aussteuerung und Lautheit .....	99
5.2	Verbindung zum Audioadapter .....	103
5.3	Fouriertransformation .....	103
5.3.1	Zusammenhang mit den Korrelationsfunktionen.....	106
5.3.2	Fast Fourier transform (FFT) .....	107
5.3.3	Berechnung mit Tabellenkalkulation.....	112
5.3.4	Darstellung des Betrags- und des Phasenspektrums .....	113
5.4	Spektrumanalyser .....	114
5.4.1	Oktavbandanalyser .....	115
5.4.2	Terzbandanalyser .....	115
5.5	Oktavsiebanalyse .....	115
5.6	Übungsaufgaben .....	116

<b>6</b>	<b>Audiosynthese</b> .....	<b>117</b>
6.1	Elektronische Klangerzeugung .....	117
6.2	Spektren mathematischer Funktionen .....	120
6.3	Skriptbasierte Synthese von Klängen .....	123
6.3.1	Steuerung des zeitlichen Hüllkurvenverlaufs .....	123
6.3.2	Klanggenerierung durch additive Klangersynthese .....	126
6.4	Synthese von Orgelklängen .....	127
6.5	Bestimmung von Soundparametern .....	128
6.6	Klangstruktur gezupfter und geschlagener Saiteninstrumente .....	128
6.6.1	Bass .....	129
6.6.2	Gitarre .....	130
6.7	Klangstruktur verschiedener Orchesterinstrumente .....	130
6.7.1	Holzbläser .....	131
6.7.2	Blechbläser .....	133
6.7.3	Streicher .....	135
6.8	Übungsaufgaben .....	136
<b>7</b>	<b>Komponieren und MIDI</b> .....	<b>137</b>
7.1	Aufbau einer MIDI-Datei .....	137
7.2	Software für MIDI-Verarbeitung und Notensatz .....	144
7.2.1	Finale .....	144
7.2.2	Capella .....	144
7.2.3	Cubase und Nuendo .....	145
7.3	Komposition .....	147
7.3.1	Melodie .....	148
7.3.2	Begleitung und Harmonik .....	149
7.3.3	Mehrstimmigkeit .....	150
7.3.4	Fuge als mehrstimmige Kompositionsform .....	150
7.3.5	Zwölftontechnik .....	151
7.4	Übungsaufgaben .....	152
<b>8</b>	<b>Sonderstellung mit visueller Programmierung</b> .....	<b>155</b>
8.1	Einleitung .....	155
8.2	Reaktor von Native Instruments .....	156
8.2.1	Ebenen der Klangersynthese .....	156
8.2.2	Ein einfacher Drumsampler .....	159
8.3	GraphEdit von DirectShow .....	162
8.3.1	Kategorien von Filtern .....	162
8.3.2	Anwendungsbeispiele .....	163

8.4	Max MSP .....	166
8.4.1	Patches als Signalfussgraphen .....	166
8.4.2	Erstellen von Patches .....	167
8.5	Übungsaufgaben .....	170
<b>9</b>	<b>Physikalische Klangmodellierung .....</b>	<b>171</b>
9.1	Klangmodellierung .....	171
9.2	Numerische Lösung von Differentialgleichungen .....	172
9.2.1	Einmassenschwinger .....	173
9.2.2	Saiten .....	178
9.2.3	Schwingungen von Stäben und Balken .....	188
9.2.4	Platten .....	194
9.3	Zusammenfassung .....	204
9.4	Übungsaufgaben .....	204
<b>10</b>	<b>Designing Sound .....</b>	<b>205</b>
10.1	Motivation .....	205
10.2	Funktionen von Filmmusik .....	209
10.3	Anlegen der Filmmusik .....	210
10.4	Kompositionstechniken .....	213
10.4.1	Die Bedeutung von Zeit, Raum und Klang .....	213
10.4.2	Improvisation .....	213
10.4.3	Variation .....	214
10.5	Analyse audiovisueller Musikdarbietungen .....	215
10.5.1	Analyse des Tanzes „Eairth“ von Domenico Strazzeri .....	217
10.5.2	Verfeinerung der Analyse .....	219
10.6	Sounddesign bei „Star Wars“ .....	220
10.6.1	Erstellung der Soundeffekte .....	220
10.6.2	Die Klangvielfalt von R2D2 .....	221
10.6.3	Effektbearbeitung .....	225
10.7	Die Filmmusik der James Bond-Filme .....	227
10.7.1	Stille als dramaturgisches Gestaltungselement .....	227
10.7.2	„Moonraker – Streng geheim“ .....	228
10.8	Übungsaufgaben .....	232
<b>11</b>	<b>Schlussbemerkungen .....</b>	<b>234</b>
11.1	Zusammenfassung .....	234
11.2	Ausblicke .....	234

<b>A Beispielprojekte</b> .....	<b>236</b>
A.1 Übersicht über die Programmbeispiele .....	236
A.2 Kapitel 2 Audiodateien .....	236
A.2.1 SinusWav und SinusAiff .....	237
A.2.2 AmplifyWav .....	237
A.2.3 NormalizeWav .....	238
A.2.4 MixWav .....	238
A.3 Kapitel 3 MP3Konverter .....	239
A.4 Kapitel 4 Delay .....	239
A.5 Kapitel 4 FourBandEQ .....	239
A.6 Kapitel 5 Aussteuerungsmessung .....	239
A.7 Kapitel 5 Visualisierung .....	240
A.8 Kapitel 6 Klangsynthese .....	240
A.9 Kapitel 7 MidiFileMaker .....	240
A.10 Kapitel 9 Modellierung .....	241
A.10.1 Masse-Feder-Dämpfer-System .....	241
A.10.2 Saitenschwingung .....	242
A.10.3 Modellierung von Balkenschwingungen .....	242
A.10.4 Schwingungen einer Platte .....	242
A.11 Kapitel 10 Sounddesign .....	242
A.11.1 R2D2 als Polizeisirene .....	243
A.11.2 Schnelle Tonfolge von R2D2 .....	243
A.11.3 R2D2 pfeift .....	243
<b>Literatur</b> .....	<b>245</b>
Bildnachweise .....	248
<b>Index</b> .....	<b>249</b>

# 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?

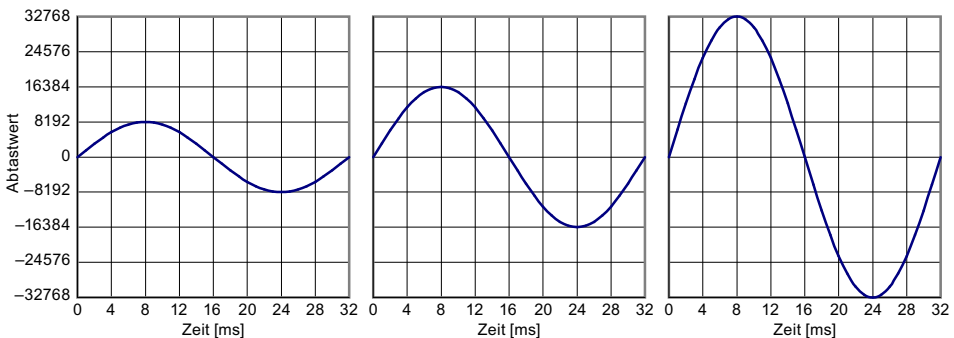
## 2.1 Akustische Signale

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.

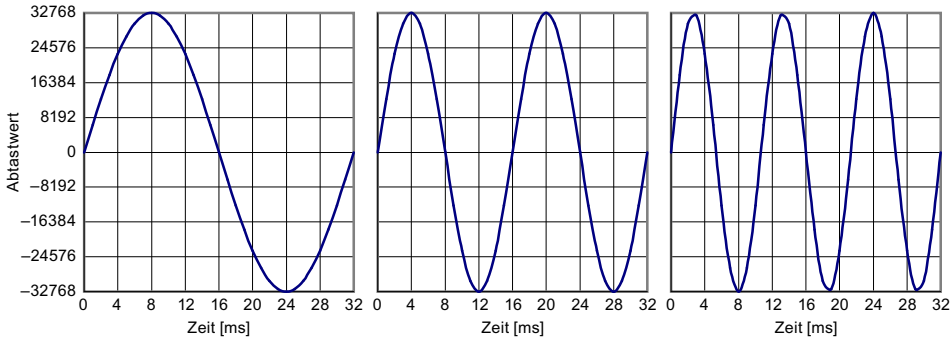


**BILD 2.1** Sinusschwingung mit verschiedenen Amplituden

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

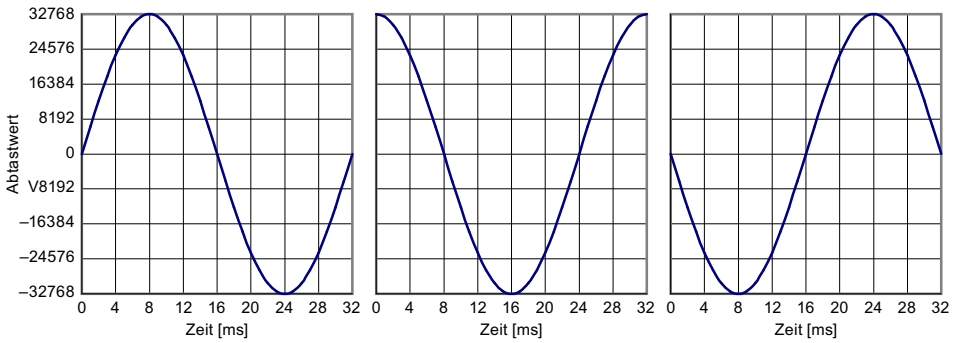
Mit der **Phase** als drittem Parameter wird die Anfangsposition der Auslenkung zu einem bestimmten Zeitpunkt bezeichnet. Da man sich die Sinusfunktion als Auslenkung eines rotierenden Zeigers vorstellen kann, dessen Position mit einem Winkel relativ zu einem Bezugspunkt angegeben werden kann, wird auch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

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

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