3 Pulse-Code-Modulation, digitale Signalverarbeitung und Audio-Codierung

3.1 Einführung

Die *Pulse-Code-Modulation* (PCM) ist das Standardverfahren zur Digitalisierung von Sprachsignalen in der Telefonie. Die Grundlagen liefern Überlegungen zum Abtasttheorem und zur Quantisierung. Die wirtschaftlichen Voraussetzungen für den Einsatz von PCM in der Telefonie ist das ISDN (Integrated Services Digital Network) und umgekehrt. Durch die Digitalisierung fügt sich die Sprachübertragung ins ISDN ein. Früher nach Daten- und Sprachkommunikation getrennte Netze können kostengünstiger auf einer gemeinsamen Systemplattform realisiert werden. Als logischer nächster Schritt folgte die Telefonie über das Internet, Voice over Internet Protocol (VoIP) genannt, und die komprimierende Audiocodierung, wie die bekannte MP3-Codierung.

Zu Beginn dieses Abschnitts werden die Schritte vom analogen zum digitalen Signal im Überblick vorgestellt. Das Abtasttheorem und die Quantisierung werden behandelt. Von besonderem Interesse ist die Digitalisierung der Sprache in der Telefonie. Es wird die Frage gestellt und beantwortet, welche Bitrate zur Übertragung eines Telefonsprachsignals notwendig ist.

Die Digitalisierung von Sprach- und Audiosignalen und von Bildern ermöglicht im Verbund mit den heute verfügbaren, leistungsfähigen Mikroprozessoren Anwendungen der digitalen Signalverarbeitung, die mit dem Begriff Multimedia umschrieben werden. Eine Darstellung der Multimediatechnik würde den hier abgesteckten Rahmen sprengen, jedoch sollen mit einer kurzen Vorstellung der schnelle Fourier-Transformation und der digitalen Filter Grundlagen gelegt werden. Abgerundet wird das Thema durch einen Blick auf die Audio-Codierung nach dem Standard MPEG-1 Layer III.

Anmerkung: Weiterführende Literatur zum Thema Quantisierung und A/D- und D/A-Umsetzer z. B. [TiSc02], [VHH98] und [Zöl05]; und zur digitalen Signalverarbeitung z. B. [BSH08], [KaKr06], [MeHo04], [Wer08b] und [Wer08c].

3.2 Digitalisierung analoger Signale

Die prinzipiellen Verarbeitungsschritte zur *Digitalisierung* eines analogen Basisbandsignals zeigt Bild 3-1. Der erste Schritt, die Tiefpassfilterung mit der Grenzfrequenz f_g, kann unterbleiben, wenn das Eingangssignal bereits passend bandbegrenzt ist.

Die eigentliche Digitalisierung geschieht in drei Schritten: der zeitlichen und der wertmäßigen Diskretisierung und der Codierung.

Zunächst werden bei der zeitlichen Diskretisierung, auch (ideale) *Abtastung* genannt, jeweils alle Abtastintervalle T_a Abtastwerte $x[n] = x(nT_a)$ als Momentanwerte aus dem analogen Signal entnommen. Die zeitdiskrete Abtastfolge $x[n]$ besitzt wertkontinuierliche Amplituden. Bei der *Quantisierung* werden den Amplituden über die *Quantisierungskennlinie* Werte aus einem diskreten Zeichenvorrat zugewiesen, so dass das digitale Signal $[x[n]]_Q$ entsteht. Im *Encoder*

3.3 Abtasttheorem

Eine sinnvolle zeitliche Diskretisierung liegt vor, wenn das zeitkontinuierliche Signal durch die Abtastfolge gut wiedergegeben wird. Bild 3-2 veranschaulicht beispielhaft, dass ein Signal ausreichend dicht abgetastet werden muss, damit es aus der Abtastfolge durch eine, im Beispiel, lineare *Interpolation* hinreichend genau wieder gewonnen werden kann. Diese grundsätzlichen Überlegungen werden im *Abtasttheorem* präzisiert.

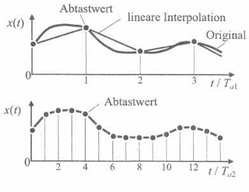

Bild 3-2 Abtastung und (lineare) Interpolation

Anmerkung: Der mathematische Grundgedanke des Abtasttheorems lässt sich auf J. L. Lagrange zurückführen. Ausführliche mathematische Darstellungen lieferten J. Whittaker (1915) und E. T. Whittaker (1935). Wichtige Beiträge zu technischen Anwendungen der Abtastung stammen aus der ersten Hälfte des 20. Jahrhunderts: H. Nyquist (1928), V. A. Kotelnikov (1933), A. Raabe (1939) [Bro04], [Lük99]. In der Literatur wird von der Nyquist-Abtastung (Nyquist Sampling), der Nyquist-Abtastrate (Nyquist (sampling) rate) und dem WKS-Abtasttheorem gesprochen. Durch C. E. Shannon wurde 1948 das Abtasttheorem einem größeren Kreis bekannt, weshalb gelegentlich auch die Bezeichnung Shannon-Abtasttheorem zu finden ist.

Die Wirkung der si-Interpolation zeigt Bild 3-4. Die zur Interpolation verwendeten si-Impulse entsprechen im Frequenzbereich einem idealen Tiefpass mit der Grenzfrequenz f_g. Eine Interpolation mit einem idealen Tiefpass liefert wieder das ursprüngliche zeitkontinuierliche Signal. Die praktische Anwendung des Abtasttheorems geschieht in *Analog-Digital-* bzw. *Digital-Analog-Umsetzern*.

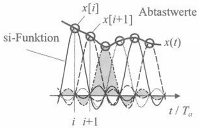

Bild 3-3 si-Interpolation

Abtasttheorem Eine Funktion $x(t)$, deren Spektrum für $|f| \geq f_g$ null ist, wird durch die Abtastwerte $x(t = nT_a)$ vollständig beschrieben, wenn das *Abtastintervall* T_a bzw. die *Abtastfrequenz* f_a so gewählt wird, dass

$$T_a = \frac{1}{f_a} \leq \frac{1}{2f_g} \tag{3.1}$$

Die Funktion kann dann durch die si-*Interpolation* fehlerfrei rekonstruiert werden.

$$x(t) = \sum_{n=-\infty}^{\infty} x(nT_a) \cdot \mathrm{si}\left(f_a \pi [t - nT_a]\right) \tag{3.2}$$

Anmerkungen: (i) Man beachte die Definition der Grenzfrequenz f_g, die eine Spektralkomponente bei eben dieser Frequenz im abgetasteten Signal ausschließt. (ii) Mathematisch gesehen handelt es sich bei der si-Interpolation in (3.2) um eine orthogonale Reihendarstellung ähnlich der Fourier-Reihe in (2.29), wobei die Abtastwerte die Rolle der Entwicklungskoeffizienten übernehmen.

Die Forderung nach strikter Bandbegrenzung (3.1) wird in Bild 3-4 anhand zweier Kosinussignale mit den Frequenzen $f_1 = 1$ kHz und $f_2 = 7$ kHz veranschaulicht. Bei der Abtastfrequenz $f_a = 8$ kHz erhält man in beiden Fällen die gleichen Abtastwerte. Offensichtlich tritt eine Mehrdeutigkeit auf, die nur durch die Bandbegrenzung des zeitkontinuierlichen Signals aufgelöst werden kann.

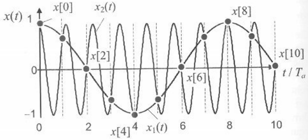

Bild 3-4 Abtastung zweier Kosinussignale $x_1(t)$ und $x_2(t)$ mit den Frequenzen $f_1 = 1$ kHz bzw. $f_2 = 7$ kHz bei einer Abtastfrequenz von $f_a = 8$ kHz

Eine wichtige Anwendung der Abtastung findet sich in der Telefonie mit der auf 300 Hz bis 3,4 kHz bandbegrenzten Telefonsprache. Nach (3.1) ist eine Abtastfrequenz von mindestens 6,8 kHz erforderlich. Tatsächlich wird mit 8 kHz das Signal überabgetastet, um bei der Übertragung mit der Trägerfrequenz-Technik einfachere Filter mit geringerer Flankensteilheit verwenden zu können.

Aufnahmen für die Audio-CD erfassen den Frequenzbereich von 20 Hz bis 20 kHz bei einer Abtastfrequenz von 44,1 kHz. Darüber hinaus sind der Audiotechnik sind auch die Abtastfrequenzen von 48 und 96 kHz gebräuchlich; und in der Videotechnik wird beispielsweise das Luminanzsignal mit der Helligkeitsinformation mit 13,5 MHz abgetastet.

3.4 Quantisierung

Das Prinzip der Digitalisierung wird anhand des Beispiels in Bild 3-5 erläutert. Das analoge Signal $x(t)$ sei auf den *Quantisierungsbereich* $[-1, 1]$ begrenzt. Falls nicht, wird das Signal mit seinem Betragsmaximum normiert. Im Weiteren wird stets von einem Quantisierungsbereich von -1 bis $+1$ ausgegangen. Die Amplituden der Abtastwerte sollen mit je 3 Bits dargestellt werden. Man spricht dann von einer *Wortlänge* von 3 Bits und schreibt kurz $w = 3$. Mit 3 Bits können genau $2^3 = 8$ *Quantisierungsintervalle* oder Quantisierungsstufen unterschieden werden.

Bei der *gleichförmigen Quantisierung* teilt man den Quantisierungsbereich in 2^w Intervalle mit der *Quantisierungsintervallbreite* oder Quantisierungsstufenhöhe.

$$Q = 2^{-(w-1)} \tag{3.3}$$

Im Beispiel ergibt sich $Q = 1/4$. Dementsprechend ist die Ordinate in Bild 3-5 in 8 gleichgroße Intervalle eingeteilt. Den Quantisierungsintervallen werden eindeutige *Codenummer* zugewiesen. Im Beispiel sind das die Nummern 0 bis 7.

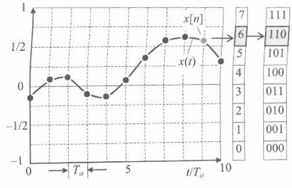

Bild 3-5 Gleichförmige Quantisierung mit 3 Bit Wortlänge: Analoges Signal $x(t)$, Abtastwerte $x[n]$, Codenummern 0 bis 7 und Codetabelle 000 bis 111

Jetzt kann die Quantisierung für jeden Abtastwert durchgeführt werden. Im Bild sind dazu entsprechend dem vorgegebenen Abtastintervall T_a die Abtastwerte als Kreise markiert. Zu jedem Abtastwert bestimmt man das Quantisierungsintervall und ordnet die entsprechende Codenummer zu. Im Beispiel des Abtastwertes für $t = 9 \cdot T_a$ ist das die Codenummer 6.

Jeder Codenummer wird bei der späteren Digital-Analog-Umsetzung genau ein diskreter Amplitudenwert, der *Repräsentant*, zugeordnet. Bei der gleichförmigen Quantisierung liegt dieser in der Intervallmitte, so dass der Abstand zwischen Abtastwert und Repräsentant die halbe Quantisierungsintervallbreite nicht überschreitet, siehe Bild 3-6. Die Repräsentanten sind im Bild als Quadrate kenntlich gemacht. Es ergibt sich eine interpolierende Treppenkurve die meist durch einen nachfolgenden Tiefpass noch geglättet wird.

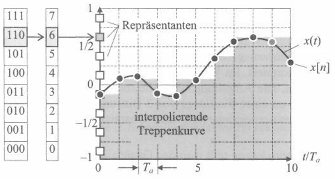

Bild 3-6 Rekonstruktion eines analogen Signals durch die interpolierende Treppenkurve

Entsprechend der *Codetabelle* werden die Codenummern zur binären Übertragung in ein Codewort umgewertet. Im Beispiel werden die Codenummern von 0 bis 7 nach dem *BCD-Code* (Binary Coded Decimal) durch die Codeworte 000 bis 111 ersetzt. Es kann der zugehörige Bitstrom abgelesen werden.

$$b_n = \{011,100,100,011,011,100,101,110,110,110,101,\ldots\}$$

Die Quantisierung beschreibt die *Quantisierungskennlinie*. Letztere definiert die Abbildung der kontinuierlichen Abtastwerte auf die zur Signalrekonstruktion verwendeten Repräsentanten. Die dem Beispiel zugrunde liegende Quantisierungskennlinie ist in Bild 3-7 links angegeben.

An der linken Quantisierungskennlinie lassen sich die beiden grundsätzlichen Probleme der Quantisierung erkennen:

- Eine *Übersteuerung* tritt auf, wenn das Eingangssignal außerhalb des vorgesehenen Aussteuerungsbereichs liegt. In der Regel tritt dann die *Sättigung* ein und es wird der Maximalwert bzw. der Minimalwert ausgegeben (*Sättigungskennlinie*).
- Eine *Untersteuerung* liegt vor, wenn das Eingangssignal (fast) immer viel kleiner als der Aussteuerungsbereich ist. Im Extremfall entsteht *granulares Rauschen* bei dem das quantisierte Signal scheinbar regellos zwischen den beiden Repräsentanten um die Null herum wechselt.

> Bei der Quantisierung ist auf die richtige Aussteuerung des Eingangssignals zu achten. Übersteuerungen und Untersteuerungen sind zu vermeiden.

In der digitalen Signalverarbeitung sind auch andere Quantisierungskennlinien gebräuchlich. Bild 3-7 rechts zeigt die Kennlinie der Quantisierung im *2er-Komplement-Format*. Der Wert 0 wird explizit dargestellt. Beim Einsatz von Festkomma-Signalprozessoren wird meist das 2er-Komplement-Format bei einer Wortlänge von 16 oder 32 Bits verwendet. Es werden die Zahlen im Bereich von −1 bis +1 dargestellt.

$$x = -a_0 2^0 + \sum_{i=1}^{w-1} a_i 2^{-i} \quad \text{mit } a_i \in \{0,1\} \text{ und } -1 \le x \le 1 - 2^{-w+1} \tag{3.4}$$

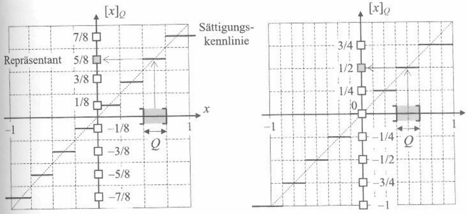

Bild 3-7 Quantisierungskennlinie der gleichförmigen Quantisierung mit $w = 3$ (links mit Sprung bei null und rechts mit der Darstellung von null)

Die negativen Zahlen berechnen sich vorteilhaft durch Komplementbildung und Addition eines Bits mit geringster Wertigkeit, dem LSB (*Least Significant Bit*).

$$-x = -\overline{a}_0 2^0 + \sum_{i=1}^{w-1} \overline{a}_i 2^{-i} + 2^{-w+1} \quad \text{mit } a_i \in \{0,1\} \text{ und } -1 \le x \le 1-2^{-w+1} \tag{3.5}$$

Beispiel Zahlendarstellung im 2er-Komplement-Format mit der Wortlänge von 8 Bits

$$2^{-2} + 2^{-4} + 2^{-6} = 0{,}328125_d = 0010\ 1010_{2c}$$

$$-0{,}328125_d = 1101\ 0110_{2c}$$

Anmerkungen: (i) Das 2er-Komplement-Format beinhaltet die Zahl −1. Oft wird jedoch aus Symmetriegründen auf sie verzichtet. (ii) Das 2er-Komplement-Format ermöglicht relativ einfache Schaltungen zur Addition von positiven und negativen Zahlen. (iii) Bei aufwändigeren Signalprozessoren und auf PCs kommt häufig das Gleitkomma-Format nach IEEE 754-1985 zum Einsatz. Das Gleitkomma-Format besteht aus Exponent und Mantisse, so dass ein größerer Zahlenbereich dargestellt werden kann [Wer08c].

3.5 Quantisierungsgeräusch

Aus den Repräsentanten kann das ursprüngliche Signal bis auf künstliche Spezialfälle nicht mehr fehlerfrei rekonstruiert werden. Wie im Beispiel deutlich wurde, wird der Quantisierungsfehler durch die Wortlänge kontrolliert. Je größer die Wortlänge, desto kleiner ist der Quantisierungsfehler. Mit wachsender Wortlänge nimmt jedoch auch die Zahl der zu übertragenden bzw. zu speichernden Bits zu. Je nach Anwendung ist zwischen der Qualität und dem Aufwand abzuwägen.

Für die PCM in der Telefonie soll nun beispielhaft die Frage beantwortet werden: Wie viele Bits werden zur Darstellung eines Abtastwertes benötigt?

Um die Frage zu beantworten, muss zunächst die Qualität quantitativ messbar sein. Dazu verwendet man das Modell der additiven Störung mit dem *Quantisierungsgeräusch* in Bild 3-8.

Anmerkung: Der Einfachheit halber werden zeitkontinuierliche Signale betrachtet, da die entsprechenden Zusammenhänge der digitalen Signalverarbeitung nicht als bekannt vorausgesetzt werden. Dies ist auch ohne Komplikationen möglich, weil die Quantisierung für jeden Momentanwert und damit auch für jeden Abtastwert unabhängig von der Zeit gilt.

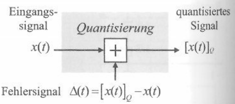

Bild 3-8 Ersatzmodell für die gleichförmige Quantisierung mit Fehlersignal $\Delta(t)$

Ein übersichtliches Beispiel liefert die Quantisierung des periodischen dreieckförmigen Signals $x(t)$ in Bild 3-9. Im unteren Bild ist das entstehende *Fehlersignal* $\Delta(t)$, das *Quantisierungsgeräusch*, aufgetragen.

Anmerkung: Von der Telefonie und Audiotechnik her kommend wird traditionell vom (Quantisierungs-)Geräusch gesprochen, da bei geringer Wortlänge die Quantisierungsfehler hörbar sind. Bei der Digitalisierung von Bildern, z. B. durch eine Digitalkamera, spricht man vom (Quantisierungs-)Bildrauschen.

Betrachtet man den Zeitpunkt $t = 0$, so ist $x(0) = 0$ und $[x(0)]_Q = Q / 2$. Mit wachsender Zeit steigt das Eingangssignal zunächst linear an und nähert sich dem Wert des Repräsentanten. Der Fehler wird kleiner und ist für $t = T_0 / 16$ gleich null. Danach ist das Eingangssignal größer als der zugewiesene Repräsentant. Das Fehlersignal ist negativ, bis das Quantisierungsintervall wechselt. Beim Übergang in das neue Quantisierungsintervall springt das Fehlersignal von $-Q / 2$ auf $Q / 2$. Entsprechendes kann für die anderen Signalabschnitte überlegt werden.

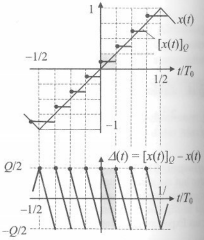

Bild 3-9 Quantisierung eines periodischen dreieckförmigen Signals (oben) und das dabei entstehende Fehlersignal $\Delta(t)$ (unten)

Das vorgestellte einfache Modell ermöglicht, die Qualität der Quantisierung quantitativ zu erfassen. Als Qualitätsmaß wird das Verhältnis der Leistungen des Eingangssignals und des Quantisierungsgeräusches, das *Signal-Quantisierungsgeräusch-Verhältnis*, kurz *SNR*, zugrunde gelegt. Im Beispiel ergibt sich für das normierte Signal bei Vollaussteuerung die mittlere Signalleistung (2.9)

$$S = \frac{1}{T_0} \int_0^{T_0} |x(t)|^2 \, dt = \frac{2}{T_0} \int_0^{T_0/2} \left(\frac{t}{T_0/2} \right)^2 dt = \frac{1}{3} \tag{3.6}$$

Die mittlere Leistung des Quantisierungsgeräusches kann ebenso berechnet werden. Das Fehlersignal ist wie das Eingangssignal abschnittsweise linear, siehe Bild 3-9. Nur die Werte sind jetzt auf das Intervall $[-Q/2, Q/2[$ beschränkt. Die mittlere Leistung ist demnach

$$N = \frac{1}{3} \cdot \frac{Q^2}{4} = \frac{Q^2}{12} \tag{3.7}$$

Für das SNR im Beispiel folgt

$$\frac{S}{N} = \frac{1/3}{Q^2/12} = 2^{2w} \tag{3.8}$$

wobei die Quantisierungsintervallbreite durch die Wortlänge (3.3) ersetzt wurde. Im logarithmischen Maß resultiert das SNR

$$\left(\frac{S}{N}\right)_{\text{dB}} = 10 \cdot \log_{10} 2^{2w} \text{ dB} = 20 \cdot w \cdot \log_{10} 2 \text{ dB} \approx 6 \cdot w \text{ dB} \tag{3.9}$$

Das SNR verbessert sich um etwa 6 dB pro Bit Wortlänge.

Im Allgemeinen hängt das SNR von der Art des Signals ab. Ein periodischer Rechteckimpulszug der zwischen zwei Repräsentanten wechselt, wird fehlerfrei quantisiert. Einem Sinussignal wird wiederum ein anderes Fehlersignal zugeordnet. Während das SNR für derartige deterministische Signale prinzipiell wie oben berechnet werden kann, wird für stochastische Signale, wie die Telefonsprache, die Verteilung der Signalamplituden zur Berechnung des SNR benötigt.

Anmerkung: Eine einfache Approximation für die Verteilung der Sprachsignalamplituden liefert die zweiseitige Exponentialverteilung. Bei vorgegebener Verteilung, z. B. durch eine Messung bestimmt, und Wortlänge kann die Lage der Quantisierungsintervalle und Repräsentanten so bestimmt werden, dass das SNR maximiert wird. Derartige Quantisierer sind in der Literatur unter den Bezeichnungen Optimal-Quantisierer und Max-Lloyd-Quantisierer zu finden [Pro01].

Das für spezielle Modellannahmen gefundene Ergebnis liefert jedoch eine brauchbare Näherung für die weiteren Überlegungen.

6dB-pro-Bit-Regel Für eine symmetrische gleichförmige Quantisierung mit hinreichender Wortlänge w in Bits und Vollaussteuerung gilt für das SNR

$$\left(\frac{S}{N}\right)_{\text{dB}} \approx 6 \cdot w \text{ dB} \tag{3.10}$$

Eine hinreichende Wortlänge liegt erfahrungsgemäß vor, wenn das Signal mehrere Quantisierungsintervalle durchläuft.

Den Einfluss einer ungenügenden Aussteuerung schätzt man schnell ab. Halbiert man die Aussteuerung, reduziert sich die Signalleistung um 6 dB und die effektive Wortlänge um 1 Bit. Nur noch die Hälfte der Quantisierungsintervalle wird tatsächlich benützt.

Anmerkung: Die Frage der Genauigkeit der Quantisierung relativiert sich vor dem Hintergrund der prinzipiell begrenzten Messgenauigkeit physikalischer Größen, wie bei der Spannungsmessung mit einem Voltmeter einer bestimmten Güteklasse. Geht man weiter davon aus, dass dem zu quantisierenden Signal eine

– wenn auch kleine – Störung überlagert ist, so ist auch nur eine entsprechend begrenzte Darstellung der Abtastwerte erforderlich.