QUELLENCODIERUNG

Manuskript einer Vorlesung
gehalten an der
Universität Hannover
von Prof. Dr.–Ing. Hans Georg Musmann

Inhaltsverzeichnis

1. Einleitung **1**

2. Informationstheoretische Grundlagen **3**
2.1 Quellenmodelle 3
2.1.1 Diskrete gedächtnislose Quelle 3
2.1.2 Diskrete gedächtnisbehaftete Quelle 9
2.2 Redundanzreduzierende Codierung 13
2.2.1 Codierung diskreter gedächtnisloser Quellen 14
2.2.2 Theorem der Quellencodierung 17
2.2.3 Codierungsalgorithmus von Huffman 20
2.2.4 Codierung mit fester Codewortlänge 22
2.2.5 Arithmetische Codierung 24
2.2.6 Codierung diskreter gedächtnisbehafteter Quellen 30
2.2.7 Universelle Codierung 32
2.3 Anwendungsbeispiele 34
2.3.1 Lauflängencodierung 34
2.3.2 Codierung von Faksimilesignalen 36
2.3.2.1 Eindimensionale Faksimile Codierung 36
2.3.2.2 Zweidimensionale Faksimile Codierung 38
2.3.2.3 Codierung einer binären Markoff–Quelle 40
2.4 Irrelevanzreduzierende Codierung 45
2.4.1 Rate Distortion Theorie 45
2.4.2 Berechnung der Rate–Distortion–Funktion diskreter Quellen 53
2.4.3 Berechnung der Rate–Distortion–Funktion kontinuierlicher Quellen 56
2.4.4 Berechnung der Rate–Distortion–Funktion einer Gaußschen Quelle mit dem Leistungsdichtespektrum 61
2.5 Anwendungsbeispiele 67
2.5.1 Amplitudenquantisierung mit minimaler Fehlervarianz 67
2.5.1.1 Quantisierung mit großer Stufenzahl K 67
2.5.1.2 Quantisierung mit beliebiger Stufenzahl K 71
2.5.2 Quantisierung mit minimaler Ausgangsentropie 76
2.5.3 Logarithmische Quantisierung für Sprachsignale 79
2.5.4 Vektorquantisierung 83

2.5.4.1 Codebuchentwurf 87
2.5.4.2 Gleichförmige Vektorquantisierung 91

3. Wesentliche Codierungsverfahren 99
3.1 Pulscodemodulation (PCM) 99
3.2 Prädiktive Codierung 102
3.2.1 Verlustlose prädiktive Codierung 102
3.2.2 Differenz–Pulscodemodulation (DPCM) 106
3.3 Transformations–Codierung 110
3.3.1 Karhunen–Loeve–Transformation 113
3.3.2 Diskrete Fourier Transformation 114
3.3.3 Diskrete Cosinus Transformation 116
3.3.4 Walsh–Hadamard Transformation 117
3.3.5 Zweidimensionale Transformationen 120
3.3.6 Vergleich von KLT, DFT, DCT, WHT 122
3.3.7 Bitzuweisung und Codierungsgewinn 124
3.4 Teilband–Codierung 127
3.4.1 Systeme zur Zwei–Band–Zerlegung 129
3.4.2 Äquivalenzbeziehungen zwischen Teilbandzerlegung und Transformation 132
3.4.3 Bitzuweisung und Codierungsgewinn 135
3.5 Bewegungskompensierte Prädiktion 137
3.5.1 Verfahren zur Bewegungsschätzung 139
3.5.2 Codierungsgewinn 147

4. Modelle psychovisueller und psychoakustischer Wahrnehmung für die Irrelevanzreduktion 151
4.1 Eigenschaften der visuellen Wahrnehmung 152
4.1.1 Ein psychovisuelles Modell 156
4.2 Eigenschaften der auditiven Wahrnehmung 158
4.2.1 Ein psychoakustisches Modell 162

5. Anwendungen 167
5.1 Codierung von Bildsignalen 167
5.1.1 Der JPEG–Standard zur Codierung von Standbildern 167
5.1.2 Der H261–Standard zur Codierung von Bildfernsprechsignalen . . 172
5.1.3 Der MPEG–1–Standard zur Codierung von Videosignalen im CIF–Format 175

5.1.4 Der MPEG–2–Standard zur Codierung von Fernsehsignalen 179
5.2 Der MPEG–1–Standard zur Codierung von Audiosignalen 180
5.2.1 Das grundlegende Codierungskonzept 181
5.2.2 Die drei Layer des MPEG–Audio–Codierungsstandards 182
5.2.2.1 Layer I 182
5.2.2.2 Layer II 184
5.2.2.3 Layer III 186
5.2.3 Ergebnisse der subjektiven Hörtests 188

6. Literaturverzeichnis 189

1. Einleitung

2. Informationstheoretische Grundlagen

2.1 Quellenmodelle

2.1.1 Diskrete gedächtnislose Quelle

Als Quellenmodelle der Informationstheorie dienen Zufallsprozesse. Das einfachste Modell ist die diskrete, gedächtnislose Quelle. Diese Quelle verfügt über einen Symbolvorrat von K Symbolen $a_1,\ ...,\ a_k,\ ...,\ a_K$, die mit den Wahrscheinlichkeiten $P(a_1),\ ...,\ P(a_K)$ auftreten. Sie erzeugt eine Nachricht, indem sie aus dem Symbolvorrat zeitlich sequentiell Symbole auswählt und zu einer Symbolfolge $u_1,\ ...,\ u_n,\ ...,\ u_N$ zusammenstellt. Jedes u_n dieser Folge steht für ein beliebiges Symbol des Symbolvorrats $a_1,\ ...,\ a_K$. Im Fall der gedächtnislosen Quelle sind die Symbole einer Symbolfolge statistisch unabhängig, d.h. es gilt

$$P(u_N \,|\, u_1,\ ...,\ u_{N-1}) = P(u_N)\ . \tag{2.1}$$

Der Informationsgehalt eines Symbols einer Nachricht wird in der hier betrachteten statistischen Informationstheorie quantitativ an der Ungewißheit des Auftretens eines Symbols gemessen. Ein Symbol besitzt nur geringen Informationsgehalt, wenn sein Auftreten vorhersagbar ist; je größer die Ungewißheit bezüglich des Auftretens eines Symbols ist, um so größer ist sein Informationsgehalt. Mit Bekanntwerden des Symbols wird die Ungewißheit beseitigt. Aufgrund dieses Zusammenhanges erscheint es sinnvoll, den Informationsgehalt $I(a_k)$ eines Symbols a_k anhand der Wahrscheinlichkeit $P(a_k)$ des Auftretens desselben Symbols zu messen:

$$I\,(a_k) = -\log_2 P(a_k)\ . \tag{2.2}$$

Mit einer solchen statistischen Größe als Maß für den Informationsgehalt wird der semantische Inhalt der Nachricht nicht erfaßt; beispielsweise sagt dieses Maß nichts darüber aus, ob die Nachricht gut oder schlecht ist. Dieses Maß der Information ermöglicht es aber, Informationserzeugung, Informationsflüsse und Informationsempfang quantitativ zu erfassen; es führt zu einer statistischen Informationstheorie.

Wird der Informationsgehalt $I(a_k)$ nach Gl. (2.2) mit dem Logarithmus zur Basis 2 bezeichnet, so ergibt sich die Einheit in "bit". Nach dieser Definition besitzt beispielsweise ein Symbol a_k mit der Wahrscheinlichkeit $P(a_k) = 1$ den Informationsgehalt $I(a_k) = 0$ bit. Es besitzt also keinen Informationsgehalt, da es mit Sicherheit

auftritt und keine Ungewißheit bezüglich des Auftretens von a_k besteht. In Bild 2.1 ist die Beziehung (2.2) grafisch dargestellt.

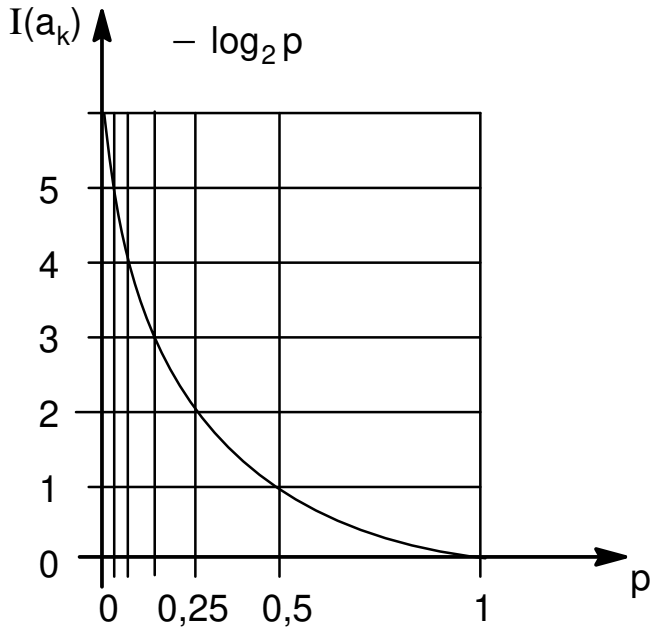


Bild 2.1 Informationsgehalt $I(a_k)$ eines Symbols a_k der Auftretenswahrscheinlichkeit $P(a_k) = p$

Der mittlere Informationsgehalt H in bit je Symbol der von einer diskreten, gedächtnislosen Quelle erzeugten Symbole beträgt demzufolge

$$H = \sum_{k=1}^{K} P(a_k) \cdot I(a_k) = -\sum_{k=1}^{K} P(a_k) \cdot \log_2 P(a_k) \tag{2.3}$$

und wird als Entropie bezeichnet. Bild 2.2 zeigt den von einem Symbol a_k mit der Wahrscheinlichkeit $P(a_k) = p$ zur Entropie im Mittel beigetragenen Informationsgehalt. Multipliziert man H mit der Anzahl der je Zeiteinheit von der Quelle gesendeten Symbole, so erhält man H', die von der Quelle je Zeiteinheit erzeugte Entropie.

Es wird mit dem Theorem der Quellencodierung in Abschnitt 2.2.2 gezeigt, daß der durch Gl. (2.3) definierte mittlere Informationsgehalt in bit je Symbol eine untere Grenze der zur Codierung eines Symbols erforderlichen Anzahl von Binärzeichen darstellt und somit eine sinnvolle Definition ist.

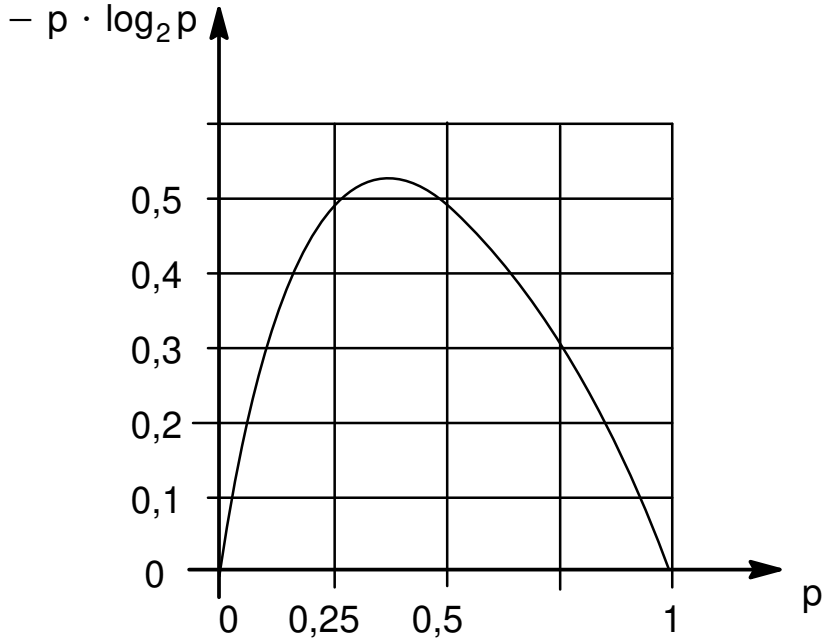


Bild 2.2 Beitrag $P(a_k)\cdot I(a_k)$ eines Symbols a_k der Auftretenswahrscheinlichkeit $P(a_k) = p$ zum mittleren Informationsgehalt H

Beispiel: Eine binäre, gedächtnislose Quelle besitzt einen Symbolvorrat, der nur aus $K = 2$ Symbolen 0 und 1 mit den Wahrscheinlichkeiten

$$P(a_1 = 0) = p \qquad \text{und} \qquad P(a_2 = 1) = 1 - p$$

besteht.

Aus Gl. (2.3) folgt dann

$$H = -\left\{p \cdot \log_2(p) + (1 - p) \cdot \log_2(1 - p)\right\} .$$

Der Verlauf der Entropie $H(p)$ als Funktion von p ist in Bild 2.3 gezeigt.–

Bezeichnet u ein beliebiges Symbol des Symbolvorrates $a_1,\ a_2,\ ...,\ a_K$ und $P(u)$ die Wahrscheinlichkeit des Symbols u, so bilden Symbole und Wahrscheinlichkeiten ein sogenanntes Nachrichtenensemble U, das die Nachrichtenquelle charakterisiert.

Die Entropie dieser Quelle kann dann anstatt in der Form von Gl. (2.3) auch in der vereinfachten Form

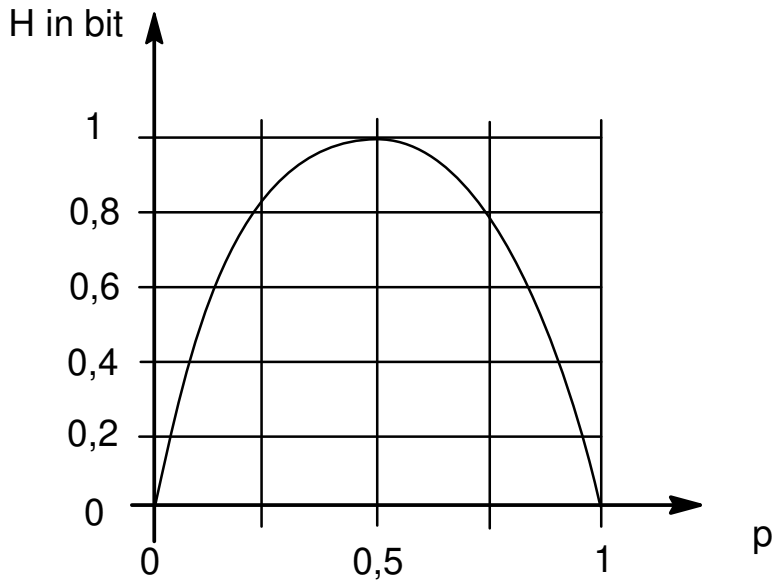


Bild 2.3 Entropie H je Symbol einer diskreten, binären, gedächtnislosen Quelle mit $P(0) = p$ und $P(1) = 1 - p$

$$\begin{aligned} H &= -\sum_{k=1}^{K} P(a_k) \cdot \log_2 P(a_k) \\ &= -\sum_{u} P(u) \cdot \log_2 P(u) \\ &= H(U) \end{aligned} \tag{2.4}$$

geschrieben werden.

Zur Erläuterung dieses Begriffs sollen einige Eigenschaften der Entropie betrachtet werden.

Theorem 2.1:

Die Entropie $H(U)$ ist stets positiv, d.h.

$$H(U) = -\sum_{u} P(u) \cdot \log_2 P(u) \geq 0\,, \tag{2.5}$$

da jeder Summand positiv ist. Nur in einem Sonderfall wird sie Null, wenn ein Symbol mit der Wahrscheinlichkeit $P(a_1) = 1$ gesendet wird und alle anderen die Wahrscheinlichkeit Null aufweisen. In diesem Fall erzeugt die Quelle die

Symbolfolge ..., a_1, a_1, a_1, ... ; die Folge dieser Symbole hat keinen Informationsgehalt.–

Theorem 2.2:

Die Entropie $H(U)$ ist stets kleiner oder gleich $\log_2 K$, d.h.

$$H(U) \leq \log_2 K \,. \tag{2.6}$$

Der Beweis soll mit Hilfe folgender Beziehungen

$$\begin{aligned} &\ln z < z - 1 \;\text{ für }\; z > 0, \;\; z \neq 1 \\ &\ln z = z - 1 \;\text{ für }\; z = 1 \;, \end{aligned} \tag{2.7}$$

sowie der Gleichung

$$\log z = \ln z \cdot \log e \tag{2.8}$$

erbracht werden.

Die Beziehungen (2.7) sind in Bild 2.4 grafisch dargestellt.

Anstelle der Ungleichung (2.6) wird die entsprechende Ungleichung

$$H(U) - \log_2 K \leq 0 \tag{2.9}$$

bewiesen. Es gilt mit (2.8)

$$\begin{aligned} H(U) - \log_2 K &= \sum_u P(u) \cdot \log_2 \frac{1}{P(u)} - \sum_u P(u) \cdot \log_2 K \\ &= \log_2 e \cdot \sum_u P(u) \cdot \ln \frac{1}{P(u) \cdot K} \;\; . \end{aligned} \tag{2.10}$$

Wird die Summe über alle u mit $P(u) \neq 0$ gebildet und werden für jeden Summanden die Beziehungen (2.7) angewendet, so folgt der Beweis

$$H(U) - \log_2 K \leq \log_2 e \cdot \sum_u P(u) \left[\frac{1}{P(u) \cdot K} - 1 \right]$$

$$H(U) - \log_2 K \leq 0 \;\; .$$

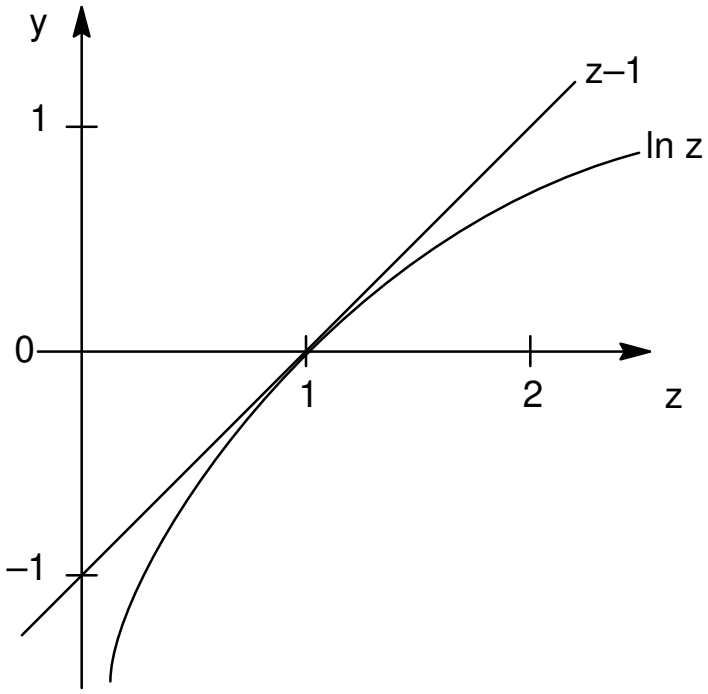


Bild 2.4 Skizze der Funktionen $y = z - 1$ und $y = \ln z$

Das Gleichheitszeichen dieser Ungleichung gilt aufgrund von (2.7) dann und nur dann, wenn für alle u

$$\frac{1}{P(u) \cdot K} = 1 \tag{2.11}$$

gilt. Somit besagt die Gl. (2.6), daß die Entropie je Symbol eines Nachrichtenensembles am größten ist, wenn alle Symbole des Symbolvorrats gleich wahrscheinlich sind. Dieser Maximalwert wird auch als Entscheidungsgehalt H_0 bezeichnet.

$$H_{\max}(U) = -\sum_{k=1}^{K} \frac{1}{K} \cdot \log_2 \frac{1}{K} = \log_2 K = H_0 \tag{2.12}$$

Wie im folgenden gezeigt wird, kann man die Auswahl eines Symbols aus einem Symbolvorrat von K Symbolen in einen Auswahlprozess von H_0 Binärentscheidungen zerlegen. Das Verzweigungsdiagramm nennt man Entscheidungsbaum oder auch Codebaum.

Jedes Symbol des in Bild 2.5 als Beispiel angeführten Symbolvorrats läßt sich durch drei Binärentscheidungen auswählen und entsprechend durch drei Binärzeichen eindeutig kennzeichnen. Die Reihenfolge der Binärzeichen ergibt sich aus der Folge der Binärzeichen entlang der Verzweigungen des Codebaumes, beginnend bei der

Wurzel bis zum ausgewählten Symbol a_k. Allgemein erfordert die Auswahl eines Symbols aus K Symbolen maximal $H_0 = \log_2 K$ Binärentscheidungen.

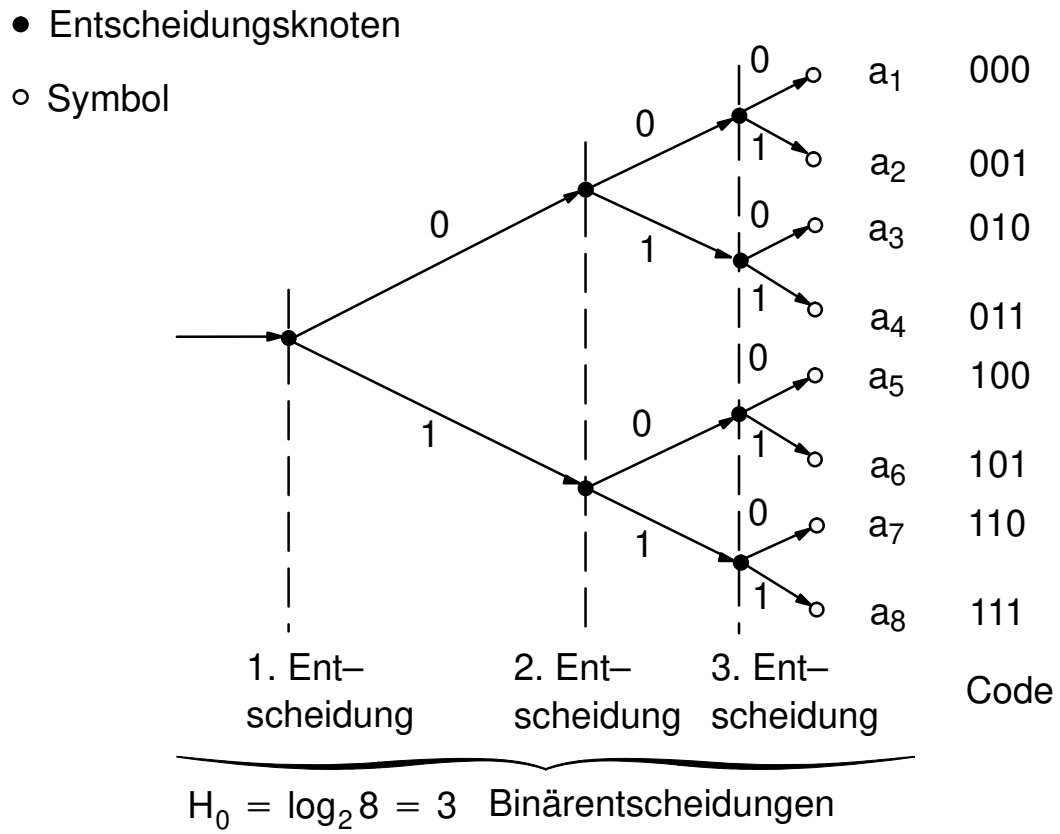


Bild 2.5 Codebaum für 8 Symbole

2.1.2 Diskrete gedächtnisbehaftete Quelle

Reale Nachrichtenquellen sind im allgemeinen gedächtnisbehaftet. Ein sehr nützliches Modell der gedächtnisbehafteten Quelle ist die diskrete, gedächtnisbehaftete Markoff–Quelle. Diese Quelle erzeugt durch zeitlich sequentielle Auswahl der Symbole u aus dem Symbolvorrat $a_1,\ ...,\ a_K$ einen Zufallsprozeß $u_1,\ ...,\ u_N$ statistisch abhängiger Symbole. Dabei wird angenommen, daß das Symbol statistisch abhängig ist, d.h.

$$P(u_N|\, u_1,\ ...,\ u_{N-1}) = P(u_N|\, u_{-m},\ ...,\ u_1,\ ...,\ u_{N-1})\,. \tag{2.13}$$

Werden die $N-1$ vorausgehenden Symbole als bekannt vorausgesetzt, so beträgt der sog. bedingte Informationsgehalt des N–ten Symbols dieser speziellen gedächtnisbehafteten Quelle

$$I(u_N|\, u_1,\ ...,\ u_{N-1}) = -\log_2 P(u_N|\, u_1,\ ...,\ u_{N-1})\ . \tag{2.14}$$

Der mittlere bedingte Informationsgehalt ergibt sich aus dem statistischen Mittelwert für $I(u_N|\, u_1,\ ...,\ u_{N-1})$. Der Mittelwert wird als bedingte Entropie $H(U_N|\, U_1,\ ...,\ U_{N-1})$ bezeichnet.

$$H(U_N|\, U_1,\ ...,\ U_{N-1}) = -\sum_{u_1} ... \sum_{u_N} P(u_1,\ ...,\ u_N) \log_2 P(u_N|\, u_1,\ ...,\ u_{N-1})\ . \tag{2.15}$$

Im Sonderfall statistisch unabhängiger Symbole folgt mit

$$P(u_N|\, u_1,\ ...,\ u_{N-1}) = P(u_N) \tag{2.16}$$

bzw.
$$P(u_1,\ ...,\ u_N) = P(u_1) \cdot P(u_2) \ldots P(u_N) \tag{2.17}$$

aus Gleichung (2.15)

$$H(U_N|\, U_1,\ ...,\ U_{N-1}) = -\sum_{u} P(u_N) \cdot \log_2 P(u_N) = H(U_N)\ , \tag{2.18}$$

d.h. die bedingte Entropie geht in die bereits bekannte Form der Entropie nach Gleichung (2.4) über. Damit entartet die gedächtnisbehaftete Quelle zu einer gedächtnislosen Quelle.

Theorem 2.3:

Die bedingte Entropie genügt der Ungleichung

$$H(U_N|U_1,\ ...,\ U_{N-1}) \le H(U_N) \tag{2.19}$$

wobei das Gleichheitszeichen dann und nur dann gilt, wenn die Ensembles U_N und $U_1,\ ...,\ U_{N-1}$ statistisch unabhängig sind.

Beweis:

Der Beweis wird mit Hilfe der Beziehungen (2.7) und (2.8) erbracht. Dazu wird die Differenz

$$H(U_N|U_1,\ ...,\ U_{N-1}) - H(U_N) = \sum_{u_1} ... \sum_{u_N} P(u_1,\ ...,\ u_N) \cdot \log_2 \frac{P(u_N)}{P(u_N|u_1,\ ...,\ u_{N-1})} \tag{2.20}$$

betrachtet. Indem für jeden Summanden (2.7) und (2.8) angewendet wird, folgt

$$H(U_N|\ U_1,\ ...,\ U_{N-1}) - H(U_N) \leq$$

$$\log_2 e \cdot \sum_{u_1} ... \sum_{u_N} P(u_1,\ ...,\ u_N) \cdot \left[\frac{P(u_N)}{P(u_N|u_1,\ ...,\ u_{N-1})} - 1 \right]$$

$$H(U_N|\ U_1,\ ....,\ U_{N-1}) - H(U_N) \leq$$

$$\log_2 e \cdot \left[\sum_{u_1} ... \sum_{N} P(u_N) \cdot P(u_1,\ ...,\ u_{N-1}) - 1 \right] = 0\ . \tag{2.21}$$

Das Gleichheitszeichen gilt dann und nur dann, wenn $P(u_N) = P(u_N|u_1,\ ...,\ u_{N-1})$ gilt, d.h. die Ensembles U_N und $U_1,\ ...,\ U_{N-1}$ statistisch unabhängig sind. Damit wird auch der Fall eingeschlossen, daß alle Ensembles $U_1,\ U_2,\ ...,\ U_N$ statistisch unabhängig sind.–

Aufgrund von Theorem 2.3 ist der mittlere Informationsgehalt $H(U_N|U_1,\ ...,\ U_{N-1})$ einer physikalischen Quelle, berechnet mit dem Modell der gedächtnisbehafteten Quelle, stets kleiner oder gleich dem mittleren Informationsgehalt $H(U_N)$, berechnet mit dem Modell der gedächtnislosen Quelle.

Betrachtet man ein Verbundereignis $u_1,\ ...,\ u_N$ von N Symbolen, die sowohl gleichzeitig als auch zeitlich nacheinander auftreten können, als ein Ereignis, so kann die Verbundentropie

$$H(U_1,\ ...,\ U_N) = -\sum ... \sum P(u_1,\ ...,\ u_N) \cdot \log_2 P(u_1,\ ...,\ u_N) \tag{2.22}$$

als der mittlere Informationsgehalt eines Verbundereignisses interpretiert werden.

Mit Hilfe der Beziehung

$$P(u_1,\ ...,\ u_N) = P(u_N|u_1,\ ...,\ u_{N-1}) \cdot P(u_1,\ ...,\ u_N) \tag{2.23}$$

kann folgender mathematischer Zusammenhang zwischen den definierten Entropien abgeleitet werden:

$$H(U_1,\ ...,\ U_N) = H(U_N|U_1,\ ...,\ U_{N-1}) + H(U_1,\ ...,\ U_{N-1})\,. \tag{2.24}$$

Die wiederholte Anwendung dieser Beziehung auf die Verbundentropie der rechten Seite der Gleichung (2.24) führt auf die sogenannte Kettenregel:

$$H(U_1,\ ...,\ U_N) = H(U_1) + H(U_2|U_1) + ... + H(U_N|U_1,\ ...,\ U_{N-1})\ . \tag{2.25}$$

Da für jeden Summanden das Theorem 2.3 gilt, erhält man weiterhin die Ungleichung

$$H(U_1,\ ...,\ U_N) \leq \sum_{n=1}^{N} H(U_n)\,, \tag{2.26}$$

in der das Gleichheitszeichen dann und nur dann gilt, wenn alle Ensembles $U_1,\ U_2,\ ...,\ U_N$ statistisch unabhängig sind.

Theorem 2.4:

Es sei

$$H_N(U) = \frac{1}{N}\, H(U_1,\ ...,\ U_N) \tag{2.27}$$

dann gilt

a) $H\,(\,U_N|\,U_1,\ ...,\ U_{N-1})$ verringert sich mit wachsendem N

b) $H_N\,(U) \geq H\,(U_N|\,U_1,\ ...,\ U_{N-1})$ (2.28)

c) $H_N\,(U)$ verringert sich mit wachsendem N

d) $\lim\limits_{N\to\infty} H_N\,(U) = \lim\limits_{N\to\infty} H\,(U_N|\,U_1,\ ...,\ U_{N-1})\,.$ (2.29) –

Beweis:

a) Da eine zusätzliche Bedingung die Entropie verkleinert, folgt aus

$$H\,(U_N|U_1,\ ...,\ U_{N-1}) \leq H\,(U_N|U_2,\ ...,\ U_{N-1})$$

$$H\,(U_N|\,U_1,\ ...,\ U_{N-1}) \leq H\,(U_{N-1}|\,U_1,\ ...,\ U_{N-2}) \tag{2.30}$$

der Beweis für (a).

b) Mit der Kettenregel Gl. (2.25) wird

$$H_N(U) = \frac{1}{N}[\,H\,(U_1) + H\,(U_2|\,U_1) + ...\ H\,(U_N|\,U_1,\ ...,\ U_{N-1})] \tag{2.31}$$

Aufgrund von (a) sind alle Terme der Summe größer als der letzte Term, d.h.

$$H_N(U) \geq \frac{1}{N}[\,N \cdot H(U_N|\,U_1,\ ...,\ U_{N-1})]$$

woraus (b) folgt.

c) $H_N(U)$ kann in der Form

$$H_N(U) = \frac{1}{N}\, H(U_1,\ ...,\ U_{N-1}) + \frac{1}{N}\, H(U_N|\,U_1,\ ...,\ U_{N-1})$$

$$H_N(U) = \frac{N-1}{N}\, H_{N-1}(U) + \frac{1}{N}\, H(U_N|\,U_1,\ ...,\ U_{N-1}) \tag{2.32}$$

geschrieben werden. Aufgrund von (b) gilt

$$H_N(U) \leq \frac{N-1}{N} H_{N-1}(U) + \frac{1}{N} H_N(U)$$

$$H_N(U) \leq H_{N-1}(U)$$

womit (c) bewiesen ist.

d) Unter Verwendung der Kettenregel Gl. (2.25) und Gl. (2.31) wird

$$\begin{aligned} H_{N+j}(U) &= \frac{1}{N+j} H(U_1, \ ..., \ U_{N-1}) \\ &\quad + \frac{1}{N+j} [\, H(U_N | \, U_1, \ ..., \ U_{N-1}) + H(U_{N+1} | \, U_1, \ ..., \ U_N) \\ &\qquad\qquad ... + H(U_{N+j} | \, U_1, \ ..., \ U_{N+j-1})] \\ &\leq \frac{1}{N+j} H(U_1, \ ..., \ U_{N-1}) + \frac{j+1}{N+j} H\,(U_N | \, U_1, \ ..., \ U_{N-1}) \end{aligned} \tag{2.33}$$

Mit $j \rightarrow \infty$ ergibt sich

$$H_\infty(U) \leq H\,(U_N | U_1, \ ..., \ U_{N-1}) \tag{2.34}$$

für alle N, d.h.

$$H_\infty(U) \leq \lim_{N \rightarrow \infty} H\,(U_N | U_1, \ ..., \ U_{N-1}) \tag{2.35}$$

Aus Gl. (2.35) und (b) folgt der Beweis für (d).–

2.2 Redundanzreduzierende Codierung

In diesem Abschnitt werden Möglichkeiten der verlustlosen Codierung diskreter, gedächtnisloser und diskreter gedächtnisbehafteter Quellen behandelt. Es wird nachgewiesen, daß zur eindeutigen Codierung und Decodierung eines Symbols im Mittel H bit je Quellensymbol benötigt werden, wobei H der mittlere Informationsgehalt je Quellensymbol ist. Im Fall, daß die Symbole einer Quelle mit $H_0 = \log_2 k$ bit dargestellt werden, kann eine noch zu definierende Redundanz R in bit je Symbol mittels einer Quellencodierung eliminiert werden, ohne daß ein Informationsverlust auftritt. Diese Codierung wird daher als redundanzmindernde Codierung bezeichnet.

2.2.1 Codierung diskreter gedächtnisloser Quellen

Es sei angenommen, daß eine diskrete, gedächtnislose Quelle eine zeitliche Folge u_1, ..., u_N statistisch unabhängiger Quellensymbole erzeugt, wobei jedes Symbol u_n aus einem Symbolvorrat von K Quellensymbolen a_1, ..., a_k, ..., a_K ausgewählt wird. Die Quellensymbole treten mit den Wahrscheinlichkeiten $P(a_1)$, ..., $P(a_k)$, ..., $P(a_K)$ auf. Jedes Quellensymbol soll durch ein Codewort dargestellt werden, das aus mehreren Codeelementen bestehen kann. Die Codeelemente werden aus einem Codesymbolvorrat von D Codesymbolen ausgewählt. Dabei werde das einzelne Quellensymbol a_k stets durch ein Codewort mit n_k Codeelementen codiert. Die Codewörter der Quellensymbole können unterschiedlich viele Codeelemente aufweisen.

Im folgenden soll die untere Grenze für die mittlere Codewortlänge $\overline{n}$ ermittelt werden, wobei

$$\overline{n} = \sum_{k=1}^{K} P(a_k) \cdot n_k \tag{2.36}$$

gilt.

Tabelle 2.6 zeigt einige Beispiele für binäre Codes mit Codewörtern unterschiedlicher Länge. Wenn eine Codierung keinen Informationsverlust verursachen darf, muß die Codierung eindeutig umkehrbar sein.

Quellen–symbol u	Wahr-schein–lichkeit P(u)	Code I	Code II	Code III	Code IV
a_1	0,5	0	0	0	0
a_2	0,25	0	1	1 0	0 1
a_3	0,125	1	0 0	1 1 0	0 1 1
a_4	0,125	1 0	1 1	1 1 1	0 1 1 1

Bild 2.6 Beispiel für binäre Codes mit ungleichlangen Codewörtern, *K=4, D=2*

Anhand der Beispiele soll gezeigt werden, daß für eine eindeutig umkehrbare Codierung gewisse Nebenbedingungen vom Code erfüllt werden müssen.

Code I ist nicht eindeutig zu decodieren, da für a_1 und a_2 das gleiche Codesymbol 0 verwendet wird. Dieser Code erfüllt somit nicht die gestellten Bedingungen.

Code II ist ebenfalls nicht eindeutig zu decodieren, da für die Folge a_1a_1 und das Quellensymbol a_3 die gleiche Folge von Codesymbolen 00 verwendet wird. Auch dieser Code erfüllt somit nicht die gestellten Bedingungen.

Damit auch eine Folge von Codewörtern eindeutig decodiert werden kann, müssen Anfang oder Ende der Codewörter erkennbar sein. Diese Forderung erfüllen sog. Präfixcodes. Ein Präfixcode ist ein Code mit der Eigenschaft, daß kein Codewort mit dem ersten Teil (Präfix) eines anderen Codewortes identisch ist.

Eine anschauliche grafische Darstellung für Codes ist der Codebaum. Bild 2.7 zeigt den Codebaum für den Präfixcode entsprechend Code III nach Tabelle 2.6.

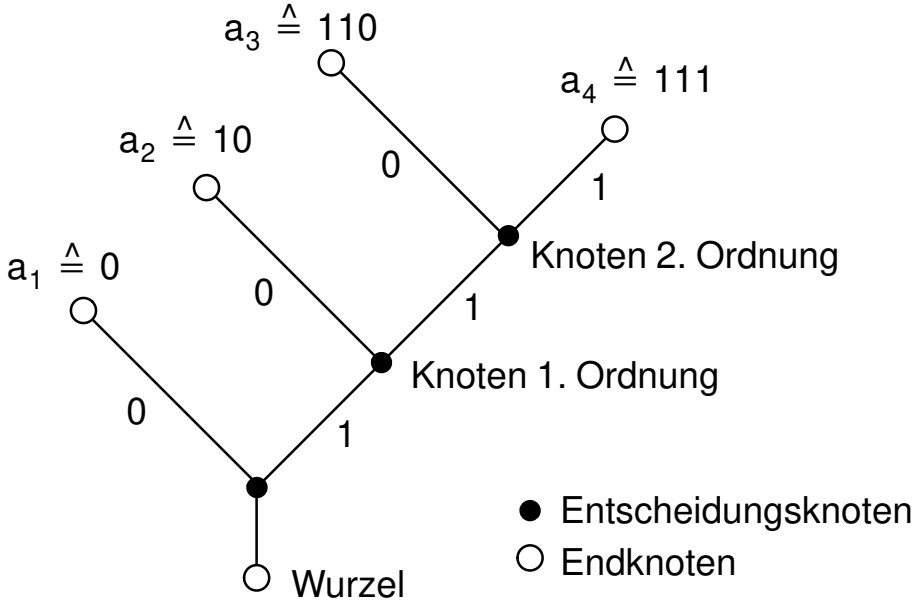


Bild 2.7 Codebaum des Präfixcode III von Tabelle 2.6

Ein Präfixcode ist dadurch gekennzeichnet, daß den Quellensymbolen nur Endknoten und keine Entscheidungsknoten zugeordnet sind. Die Zuordnung der Codewörter zu den Quellensymbolen bzw. zu den Endknoten ergibt sich aus der Aneinanderreihung der Codesymbole, die am Verbindungsweg von der Wurzel bis zum Endknoten liegen.

Code III ist ein Präfixcode. Die zufällige Folge 0111100 kann eindeutig in die Symbolfolge $a_1a_4a_2a_1$ decodiert werden, da das Ende der Codewörter aufgrund der Präfixeigenschaften stets eindeutig erkennbar ist.

Code IV ist ebenfalls eindeutig zu decodieren, obwohl er kein Präfixcode ist. Bei diesem Code ist der Anfang eines Codewortes jeweils durch Auftreten des Symbols 0 zu erkennen.

Um eine eindeutig umkehrbare Codierung zu gewährleisten, werden zunächst nur Präfixcodes betrachtet. Die zur Konstruktion von Präfixcodes erforderlichen Voraussetzungen sollen im folgenden noch näher erörtert werden. Um den K verschiedenen Quellensymbolen je ein Codewort und je einen Endknoten zuordnen zu können, müssen die Codewortlängen $n_1,\ n_2,\ ...,\ n_K$ die sog. Kraft'sche Ungleichung erfüllen:

Theorem 2.5:

Eine notwendige Bedingung für die Konstruktion eines Präfixcodes mit D Codesymbolen ist, daß die Codewortlängen $n_1,\ ...,\ n_k,\ ...,\ n_K$ die Ungleichung

$$\sum_{k=1}^{K} D^{-n_k} \leq 1 \tag{2.37}$$

erfüllen.–

Beweis:

Jeder Knoten des Codebaumes eines Präfixcodes mit D Codesymbolen besitzt D Zweige. Ist n_K die größte Codewortlänge, so gibt es maximal D^{n_K} Endknoten. Besitzt ein Präfixcode einen Endknoten der Ordnung n_K, wobei $n_k \leq n_K$, so vermindert dieser die Anzahl der verfügbaren Endknoten n_K–ter Ordnung um $D^{n_K - n_k}$, wie es in Bild 2.8 dargestellt ist.

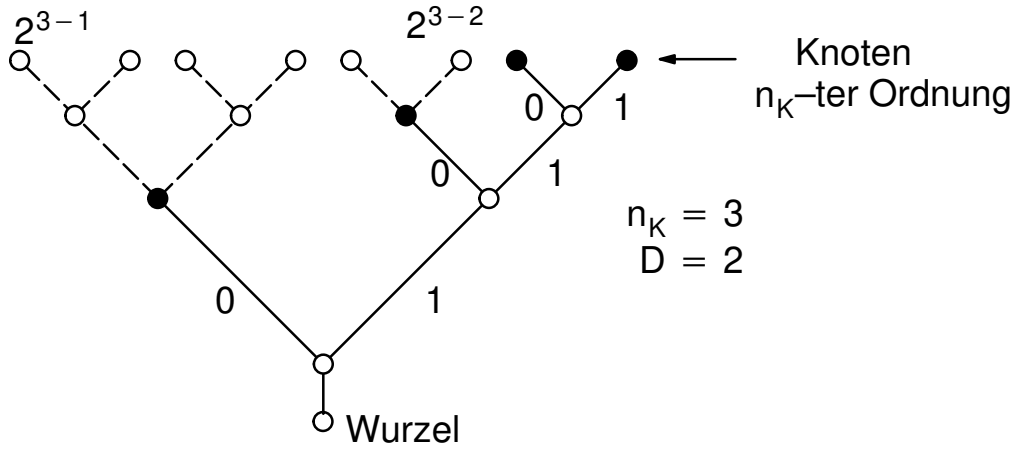


Bild 2.8 Ergänzter Codebaum des Code III

Die Summe aller nicht verfügbaren Endknoten n_K–ter Ordnung und verwendeter Endknoten n_K–ter Ordnung muß immer kleiner oder gleich D^{n_K} sein, d.h.

$$\sum_{k=1}^{K} D^{n_K - n_k} \leq D^{n_K} . \tag{2.38}$$

Wird diese Ungleichung durch D^{n_K} dividiert, so erhält man Theorem 2.5.–

Es läßt sich nachweisen, daß jeder eindeutig umkehrbare Code die Kraft'sche Ungleichung erfüllen muß. Somit ist jeder eindeutig umkehrbare Code auch durch einen Präfixcode darstellbar. Das folgende Theorem der Quellencodierung gilt somit allgemein für alle eindeutig umkehrbaren Codes wie auch für die besondere Gruppe der Präfixcodes.

2.2.2 Theorem der Quellencodierung

Theorem 2.6:

Gegeben sind eine diskrete, gedächtnislose Quelle U mit der Entropie $H(U)$ und ein Codesymbolvorrat von D Codesymbolen. Es ist möglich, den Quellensymbolen a_1, ..., a_K derart Codewörter eines Präfixcodes zuzuordnen, daß die mittlere Codewortlänge $\overline{n}$ je Quellensymbol der Bedingung

$$\frac{H(U)}{\log_2 D} \leq \overline{n} < \frac{H(U)}{\log_2 D} + 1 \tag{2.39}$$

genügt.–

Beweis:

Es seien $P(a_1)$, ..., $P(a_K)$ die Wahrscheinlichkeiten der Quellensymbole a_1, ..., a_K und n_1, ..., n_K die zugehörigen Codewortlängen. Es wird zunächst die linke Ungleichung, d.h. die untere Grenze des Theorems 2.6 bewiesen:

$$H(U) - \overline{n} \log_2 D \leq 0 \tag{2.40}$$

$$H(U) - \overline{n} \log_2 D = \sum_{k=1}^{K} P(a_k) \log_2 \frac{1}{P(a_k)} - \sum_{k=1}^{K} P(a_k)\, n_k \log_2 D \tag{2.41}$$

$$H(U) - \overline{n} \log_2 D = \sum_{k=1}^{K} P(a_k) \log_2 \frac{D^{-n_k}}{P(a_k)} \tag{2.42}$$

Mit den Beziehungen (2.7) und (2.8) erhält man:

$$H(U) - \overline{n}\log_2 D \leq \sum_{k=1}^{K} P(a_k)\left[\frac{D^{-n_k}}{P(a_k)} - 1\right]\log e \tag{2.43}$$

$$H(U) - \overline{n}\log_2 D \leq \left[\sum_{k=1}^{K} D^{-n_k} - 1\right]\log e \leq 0 \tag{2.44}$$

Die Ungleichung (2.44) folgt aus der Kraft'schen Ungleichung (2.37). In Ungleichung (2.44) gelten die Gleichheitszeichen dann und nur dann, wenn für alle k

$$P(a_k) = D^{-n_k} \quad ; \quad 1 \leq k \leq K \tag{2.45}$$

gilt. Mit Gleichung (2.45) ist die Codewortlänge für ein Quellensymbol der Wahrscheinlichkeit $P(a_k)$ festgelegt:

$$n_k = -\frac{\log_2 P(a_k)}{\log_2 D}\,. \tag{2.46}$$

Die häufigen Quellensymbole werden demzufolge in kurze und die seltenen Quellensymbole in lange Codeworte codiert. Aufgrund der Ungleichung (2.40) liefert die Codierung mit Codewortlängen gemäß Gl. (2.46) den Code mit der kleinsten mittleren Codewortlänge.

Da nur Codierungen mit ganzzahligen Codewortlängen n_k realisiert werden können, ist die Codierungsvorschrift (2.46) nur anwendbar, wenn alle $P(a_k)$ in ganzzahligen Potenzen von D vorliegen. In allen anderen Fällen kann die Codierungsvorschrift (2.46) nur näherungsweise erfüllt werden, indem n_k als ganze Zahl entsprechend

$$-\frac{\log_2 P(a_k)}{\log_2 D} \leq n_k < -\frac{\log_2 P(a_k)}{\log_2 D} + 1 \quad , \quad 1 \leq k \leq K \tag{2.47}$$

gewählt wird. Durch Multiplikation mit $P(a_k)$ und Aufsummierung über k erhält man aus der Ungleichung (2.47) den Beweis für das Theorem 2.6. Da die Kraft'sche Ungleichung durch die linke Ungleichung des Theorems erfüllt wird, gibt es immer einen Präfixcode, der das Theorem 2.6. erfüllt.–

Als Redundanz R einer Quelle wird die Differenz zwischen Entscheidungsgehalt H_0 und Entropie H definiert. Diese Definition ist nur für diskrete Quellen anwendbar, da der Entscheidungsgehalt nur für diskrete Quellen definiert ist.

Die Redundanz R eines Symbols einer diskreten Quelle beträgt somit im Mittel

$$R = H_0 - H\,, \tag{2.48}$$

wobei $H_0 = \log_2 K$ und H der mittlere Informationsgehalt je Quellensymbol der diskreten gedächtnislosen oder gedächtnisbehafteten Quelle ist. Die Redundanz ist stets positiv und aufgrund von Theorem 2.3 für die gedächtnisbehafteten Quelle stets größer oder gleich der einer gedächtnislosen Quelle.

Bei einer Codierung mit gleichlangen Codewörtern der Länge n und dem Umfang D des Codesymbolvorrats ist der Entscheidungsgehalt der Codewörter $H_0 = n \cdot \log_2 D$ und die Redundanz

$$R = n \cdot \log_2 D - H\,. \tag{2.49}$$

Bei einer Codierung mit ungleichlangen Codewörtern der mittleren Länge $\overline{n}$ wird

$$R = \overline{n} \cdot \log_2 D - H\,. \tag{2.50}$$

Dabei kann $\overline{n} \cdot \log_2 D$ als mittlerer Entscheidungsgehalt der Codewörter interpretiert werden.

Der Code III in Tab. 2.6 ist gemäß Gl. (2.50) ein optimaler, redundanzfreier Code. Es liegt der Sonderfall vor, daß die Codierungsvorschrift (2.46) ganzzahlige n_k liefert, weil die Wahrscheinlichkeiten der Quellensymbole zufällig ganzzahlige Potenzen von $D = 2$ sind. Aus Gl. (2.46) folgt mit $\log_2 D = 1$

$$\overline{n} = \sum_{k=1}^{K} P(a_k)\; n_k = -\sum_{k=1}^{K} P(a_k)\; \log_2 P(a_k) = H(U)\;. \tag{2.51}$$

Damit erreicht der Code III die untere Grenze $\overline{n} = H(U) = 1.75$ bit. Bezogen auf eine Codierung mit gleichlangen Codewörtern gemäß Gl. (2.49) eliminiert er eine Redundanz von

$$R = n \log_2 D - H = 2 - 1.75 = 0.25 \text{ bit} \tag{2.52}$$

je Symbol.

Redundanzmindernde Präfixcodes, die das Theorem 2.6 erfüllen, wurden von Shannon [15], Fano [5] und Huffman [10] angegeben. Unter diesen Codes führt der von Huffman vorgeschlagene Codierungsalgorithmus nachweisbar auf den Code mit der geringsten Redundanz bzw. geringsten mittleren Codewortlänge $\overline{n}$.

2.2.3 Codierungsalgorithmus von Huffman

Der von Huffman [10] entwickelte Codierungsalgorithmus zur Berechnung von Codes mit minimaler Redundanz lautet im Fall einer Binärcodierung:

1. Schritt: Die K Quellensymbole werden in der Reihenfolge abnehmender Wahrscheinlichkeit geordnet, wie es im Beispiel von Bild 2.9 gezeigt wird.
2. Schritt: Man fasse die zwei kleinstwahrscheinlichen Quellensymbole zu einem Hilfssymbol zusammen und berechne dessen Wahrscheinlichkeit.
3. Schritt: Man ordne das Hilfssymbol und die übrigen Quellensymbole wieder nach abnehmender Wahrscheinlichkeit, fasse dann die zwei kleinstwahrscheinlichen Symbole wieder zu einem neuen Hilfssymbol zusammen und berechne dessen Wahrscheinlichkeit.
4. Schritt: Der 3. Schritt ist so oft zu wiederholen, bis ein Symbol mit der Wahrscheinlichkeit 1 auftritt.
5. Schritt: Kennzeichnet man die Zusammenfassungen von Symbolen wie in Bild 2.9 durch Linien, so erhält man den Codebaum eines Präfixcodes, dessen K Endknoten den K Quellensymbolen zugeordnet sind. Jeder Knoten besitzt zwei Zweige. Jedem Zweig eines Knotens ist eines der zwei Codesymbole 0 oder 1 des Codesymbolvorrats zugeordnet; aus dieser Zuordnung ergibt sich der Code für die Quellensymbole.

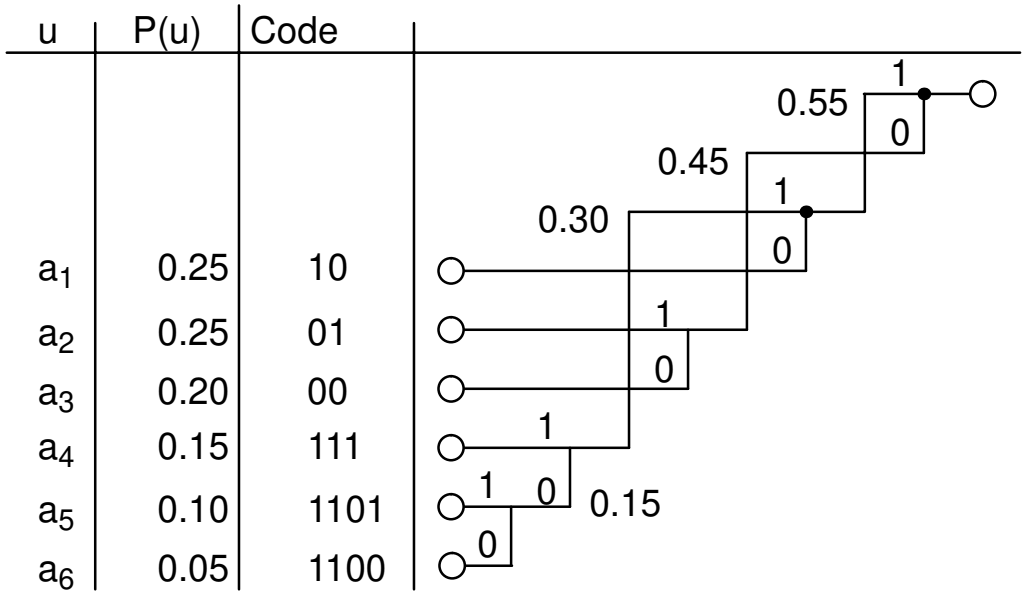


Bild 2.9 Beispiel für die Berechnung eines binären Huffman–Code

Beispiel: Für die in Bild 2.9 angegebene Quelle und Codierung berechnet sich

$$H(U) = -[0,25 \cdot \log_2(0,25) + 0,25 \cdot \log_2(0,25) + 0,20 \cdot \log_2(0,20) + 0,15 \cdot \log_2(0,15) + 0,10 \cdot \log_2(0,10) + 0,05 \cdot \log_2(0,05)]$$

$$H(U) = 2,42 \text{ bit/Symbol}$$

$$\overline{n} = 0,25 \cdot 2 + 0.25 \cdot 2 + 0,20 \cdot 2 + 0,15 \cdot 3 + 0,10 \cdot 4 + 0,05 \cdot 4$$

$$\overline{n} = 2,45 \text{ bit/Symbol.}$$

Die verbleibende Redundanz der Codierung beträgt somit

$$R = \overline{n} - H(U) = 0,03 \text{ bit/Symbol.}$$

Ohne redundanzmindernde Codierung hätte die Redundanz bei einer Codierung mit 3 bit je Symbol

$$R = 3 - 2,42 = 0,58 \text{ bit/Symbol}$$

betragen.–

Die obere Grenze des Theorems der Quellencodierung kann noch eingeschränkt werden, indem nicht einzelne Quellensymbole, sondern jeweils Blöcke von L Quellensymbolen einer Symbolfolge mit einem Codewort codiert werden. Das Theorem der Quellencodierung geht dann in die Form

$$\frac{L \cdot H(U)}{\log_2 D} \leq L \cdot \overline{n} < \frac{L \cdot H(U)}{\log_2 D} + 1$$

bzw.

$$\frac{H(U)}{\log_2 D} \leq \overline{n} < \frac{H(U)}{\log_2 D} + \frac{1}{L} \tag{2.53}$$

über. Für hinreichend große L nähert sich damit die obere Grenze des Theorems 2.6 der unteren Grenze. Damit geht die je Symbol verbleibende Redundanz dieser redundanzmindernden Blockcodierung gegen Null.

2.2.4 Codierung mit fester Codewortlänge

Eine diskrete, gedächtnislose Quelle mit einem Symbolvorrat von K Quellensymbolen habe den mittleren Informationsgehalt $H(U)$ je Quellensymbol.

Zur Codierung mit fester Codewortlänge soll jeweils ein Block von L aufeinander folgenden Quellensymbolen $u_1,\ ...,\ u_L$ durch ein Codewort dargestellt werden.

Es gibt insgesamt K^L mögliche verschiedene Blöcke mit den Wahrscheinlichkeiten $P(u_1,\ ...,\ u_L) = P(u_1) \cdot ... \cdot P(u_L)$ von denen eine Untermenge mit der Eigenschaft

$$\left| -\frac{\log_2 P(u_1,\ ...,\ u_L)}{L} - H(U) \right| \leq \delta \tag{2.54}$$

bzw.

$$\left| -\frac{\sum_{l=1}^{L} \log_2 P(u_l)}{L} - H(U) \right| \leq \delta \tag{2.55}$$

betrachtet wird. Aufgrund des "schwachen Gesetzes großer Zahlen" gilt für beliebige $\delta > 0$.

$$P\left[\left| -\frac{\sum_{l=1}^{L} \log_2 P(u_l)}{L} - H(U) \right| > \delta \right] \leq \varepsilon(L,\delta) \quad \text{und} \quad \lim_{L \to \infty} \varepsilon(L,\delta) = 0\,, \tag{2.56}$$

d.h. der arithmetische Mittelwert konvergiert gegen den statistischen Mittelwert. Vergleiche dazu auch Theorem 2.4 d.

Gleichung (2.56) zufolge wird die Gesamtheit der möglichen Blöcke in zwei Gruppen unterteilt, siehe Bild 2.10.

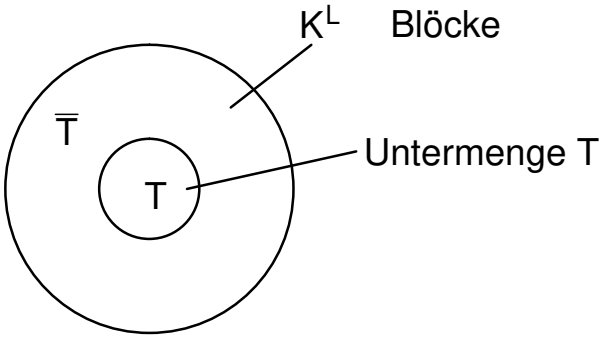


Bild 2.10 Unterteilung der K^L möglichen Blöcke in die Untermengen T und $\overline{T}$

Die Wahrscheinlichkeit, daß ein Block der Untermenge T auftritt, ist

$$P(T) = 1 - \varepsilon(L, \delta) \tag{2.57}$$

und die Wahrscheinlichkeit für einen Block der dazu komplementären Menge $\overline{T}$

$$P(\overline{T}) = \varepsilon(L, \delta) \ . \tag{2.58}$$

Gl. (2.57) besagt, daß mit wachsender Blocklänge $L \rightarrow \infty$ nur noch Blöcke der Untermenge T auftreten.

Aus Gl. (2.54) ergeben sich für die Wahrscheinlichkeit $P(u_1, \ ..., \ u_L)$ eines speziellen Blockes aus T die Grenzen

$$2^{-L\,(H(U)-\delta)} \geq P(u_1, \ ..., \ u_L) \geq 2^{-L\,(H(U)+\delta)} ; \tag{2.59}$$

da δ beliebig klein gewählt werden kann, folgt aus Gl. (2.59) für den Grenzwert $L \rightarrow \infty$, daß jeder Block die gleiche Wahrscheinlichkeit

$$P(u_1, \ ..., \ u_L) = 2^{-LH(U)} \tag{2.60}$$

und die Untermenge T dann umgerechnet M_T

$$M_T = 2^{LH(U)} \tag{2.61}$$

Blöcke enthält.

Zur Codierung mit einem Codesymbolvorrat von D Codesymbolen kann demzufolge ein Code mit fester Codewortlänge N angewendet werden.

Theorem 2.7:

Gegeben sei eine diskrete, gedächtnislose Quelle U mit der Entropie $H(U)$ und ein Blockcode, der L Quellensymbolen ein N–stelliges Codewort eines Codesymbolvorrats von D Codesymbolen entsprechend

$$N \log_2 D \geq L\, H(U) \tag{2.62}$$

zuordnet. Dabei kann die Wahrscheinlichkeit $P(\overline{T})$, daß für einen Block kein Codewort zur Verfügung steht, beliebig klein gehalten werden, indem L hinreichend groß gewählt wird.–

2.2.5 Arithmetische Codierung

Die arithmetische Codierung ist eine redundanzmindernde Blockcodierung variabler Codewortlänge, die sich einer Veränderung der Statistik der Quellensymbole anpaßt. Zur arithmetischen Codierung werden die Quellensymbole $a_1,\ ...,\ a_k$ einer diskreten gedächtnislosen Quelle in beliebiger Reihenfolge geordnet.
Es sei

$$P_k = P(a_1) + P(a_2) + ... + P(a_{k-1}) \tag{2.63}$$

die akkumulierte Wahrscheinlichkeit P_k der Symbole a_1 bis a_{k-1}, die aber $P(a_k)$ nicht enthält.

Zur Berechnung des Codewortes $\vec{X}(a_k)$ für das Quellensymbol a_k wird P_k als Dualzahl entwickelt und die Entwicklung nach n_k Stellen gemäß

$$-\log_2 P(a_k) \leq n_k < -\log_2 P(a_k) + 1 \tag{2.64}$$

abgebrochen.

Durch Gl. (2.64) wird gewährleistet, daß die Codierung das Theorem der Quellencodierung Gl. (2.39) einhält. Quellensymbole großer Wahrscheinlichkeit werden durch kurze und solche kleiner Wahrscheinlichkeit durch lange Codewörter dargestellt.

Die Entwicklung von P_k als Dualzahl ergibt

$$C'(a_k) = \sum_{i=1}^{n_k} c_i \cdot 2^{-i}\,. \tag{2.65}$$

Aufgrund des Abbruchs der Entwicklung ist $C' \leq P_k$. Zur Bildung des Codewortes $\vec{X}(a_k)$ wird C' aufgerundet gemäß

$$\vec{X}(a_k) = \begin{cases} C'(a_k) & \text{für } C'(a_k) = P_k \\ C'(a_k) + 2^{-n_k} & \text{für } C'(a_k) < P_k \end{cases} \tag{2.66}$$

ohne die Codewortlänge n_k zu verlängern.

Beispiel: Gegeben sei eine diskrete, gedächtnislose Quelle mit folgenden Eigenschaften:

a_k	$P(a_k)$	P_k	n_k	$C'(a_k)$	$\vec{X}(a_k)$
a_1	0.4	0.0	2	00	00
a_2	0.3	0.4	2	01	10
a_3	0.2	0.7	3	101	110
a_4	0.1	0.9	4	1110	1111

Das Codewort für a_3 ergibt sich aus der Entwicklung von $P_3 = P(a_1) + P(a_2) = 0,7$ mit $P_3 = 0,10110...$. Aus der Bedingung Gl. (2.64) folgt mit $P(a_3) = 0,2$ der Abbruch mit $2,32 \leq n_3 < 3,32$, d.h. $n_3 = 3$. Aus Gl. (2.66) folgt $\vec{X}(a_3) = C'(a_3) + 2^{-n_3} = 110.$–

Der Code kann in Form einer Intervall–Unterteilung auf einer Ordinate der akkumulierten Wahrscheinlichkeiten P_k dargestellt werden, siehe Bild 2.11. Jedem Quellensymbol ist entlang der Ordinate ein Intervall zugeordnet. Das entsprechende Codewort als Dualzahl interpretiert liegt aufgrund der Aufrundung nach Gl. (2.66) stets im zugehörigen Intervall oder an dessen unterer Grenze.

Die arithmetische Codierung wird auf einen Block von L aufeinanderfolgenden Quellensymbolen $u_1, \ ..., \ u_L$ angewendet. Jedem Quellensymbol der Folge wird eine Ordinate mit gleicher Intervallunterteilung zugeordnet. Die Skalierung der Ordinate ergibt sich für das l–te Quellensymbol aus dem möglichen Wertebereich der Akkumulation der Verbundwahrscheinlichkeiten

$$P(u_1, \ ..., \ u_l) = P(u_1) \cdot P(u_2) \cdot ... \cdot P(u_l) \ , \tag{2.67}$$

siehe Bild 2.12. Zur Vereinfachung der Berechnung des Codewortes wird bei fortlaufenden Symbolfolgen für jedes Symbol u_l die bereits ausgeführte Akkumulation C_{l-1} vom vorangegangenen Symbol u_{l-1} übernommen und nur die zusätzlich auszuführende Akkumulation berechnet, d.h.

$$C_l = C_{l-1} + P(u_1, ..., u_{l-1}) \cdot P_k^{(l)} \quad \text{für } \ l = 2, ..., L \ . \tag{2.68}$$

Dabei ist im ersten Schritt $C_1 = P_k^{(1)}$ und nachfolgend unter der Annahme $u_l = a_k$

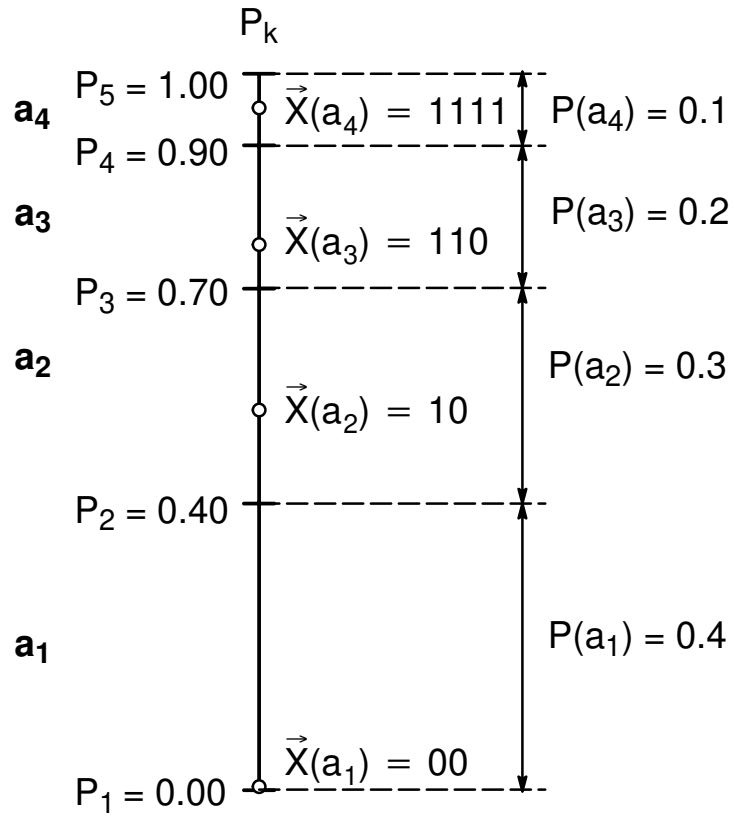


Bild 2.11 Darstellung einer arithmetischen Codierung für Einzelsymbole

$$P_k^{(l)} = P^{(l)}(a_1) + P^{(l)}(a_2) + \ldots + P^{(l)}(a_{k-1}) \,. \tag{2.69}$$

Am Ende des Blockes wird zur Berechnung des Codewortes $\vec{X}(u_1, \ldots, u_L)$ C'_L in eine Dualzahl entwickelt und die Entwicklung nach n Stellen gemäß

$$-\log P(u_1, \ldots, u_L) \leq n < -\log P(u_1, \ldots, u_L) + 1 \tag{2.70}$$

abgebrochen. Die abgebrochene Dualzahl C'_L wird anschließend gemäß Gl. (2.66) aufgerundet.

$$\vec{X}(u_1, \ldots, u_L) = \begin{cases} C'_L & \text{für} \quad C'_L = C_L \\ C'_L + 2^{-n} & \text{für} \quad C'_L < C_L \end{cases} \tag{2.71}$$

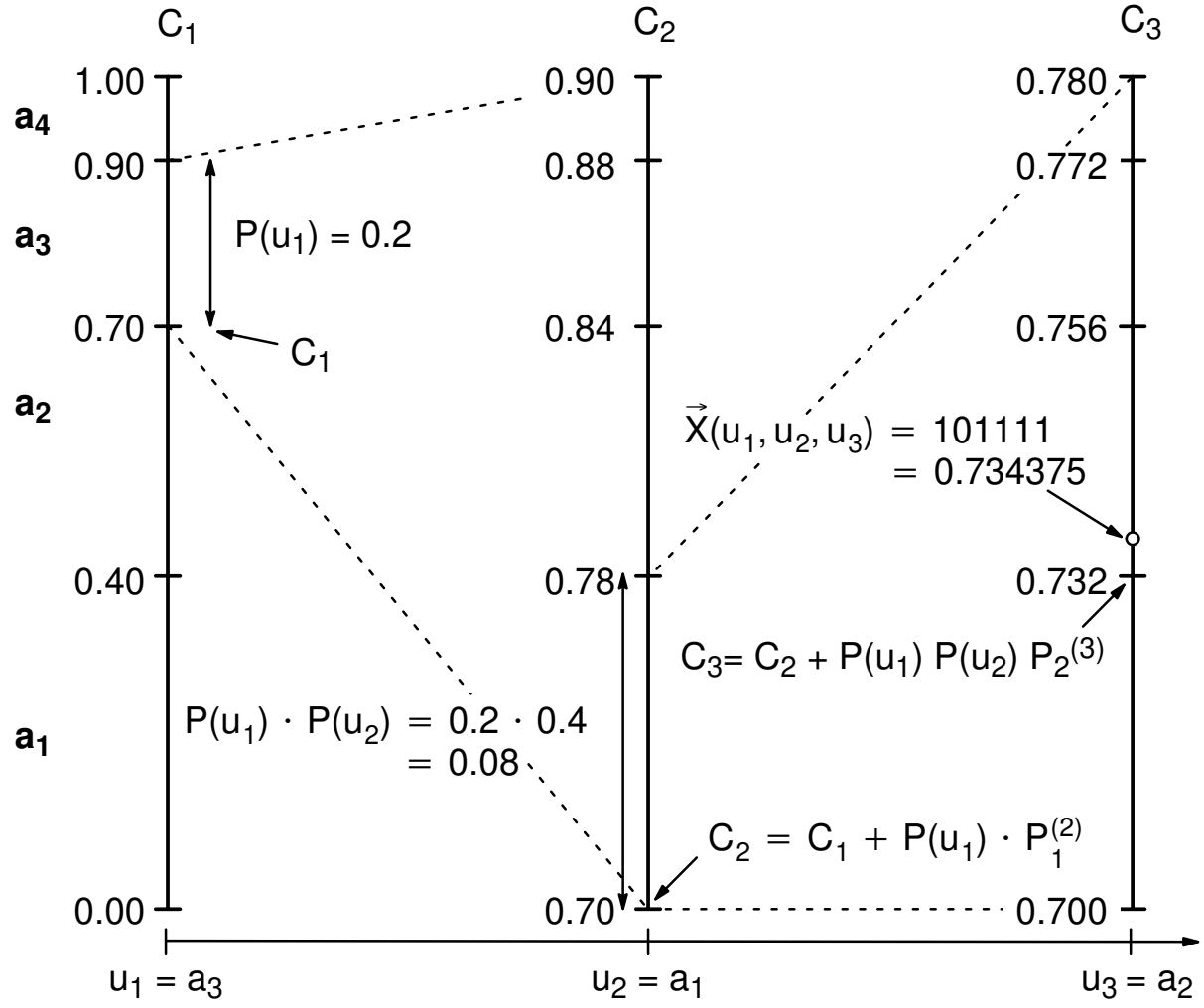


Bild 2.12 Darstellung einer arithmetischen Codierung mit *L=3* für einen Block *(u_1=a_3, u_2=a_1, u_3=a_2)*

Beispiel: Berechnung einer arithmetischen Codierung für die Symbolfolge $u_1 = a_3$, $u_2 = a_1$, $u_3 = a_2$:

l	u_l	$P(u_l)$	$P_k^{(l)}$	C_l	C'_3	$\vec{X}(a_3 a_1 a_2)$
1	$u_1=a_3$	0,2	0,7	0,7		
2	$u_2=a_1$	0,4	0,0	0,7		
3	$u_3=a_2$	0,3	0,4	0,732	101110	101111

Die Codewortlänge folgt aus Gl. (2.70)

$$-\log_2 P(u_1)\cdot P(u_2)\cdot P(u_3) \le n < -\log_2 P(u_1)\cdot P(u_2)\cdot P(u_3) + 1$$

$$-\log_2 0,2\cdot 0,4\cdot 0,3 \le n < -\log_2 0,2\cdot 0,4\cdot 0,3 + 1$$

$$-\log_2 0,024 \le n < -\log_2 0,024 + 1$$

$$5,38 \le n < 6,38$$

d.h. $n = 6$

Das Codewort berechnet sich mit Gl. (2.68) zu

$$\begin{aligned} C_3 &= C_2 + P(u_1 = a_3\,,\ u_2 = a_1)\cdot P_2^{(3)} \\ &= 0,70 + 0,2\cdot 0,4\cdot 0,4 \\ &= 0,732 \end{aligned}$$

Die Dualzahlenentwicklung von C_3 ergibt

$$\begin{aligned} C'_3 &= 101110 \\ &= 1\cdot 2^{-1} + 0\cdot 2^{-2} + 1\cdot 2^{-3} + 1\cdot 2^{-4} + 1\cdot 2^{-5} + 0\cdot 2^{-6} \\ &= 0,50000 + 0,00000 + 0,12500 + 0,06250 + 0,03125 + 0,00000 \\ &= 0,71875 \end{aligned}$$

Da $C'_3 < C_3$ ist, folgt mit Gl. (2.71)

$$\vec{X}(u_1 = a_3\ ,\ u_2 = a_1\ ,\ u_3 = a_2) = 101111\ -$$

Die schrittweise Codierung von $u_1,\ u_2,\ u_3$ ist in Bild 2.12 dargestellt.

Zur Decodierung wird die gleiche Intervallunterteilung wie bei der Codierung verwendet, siehe Bild 2.13.

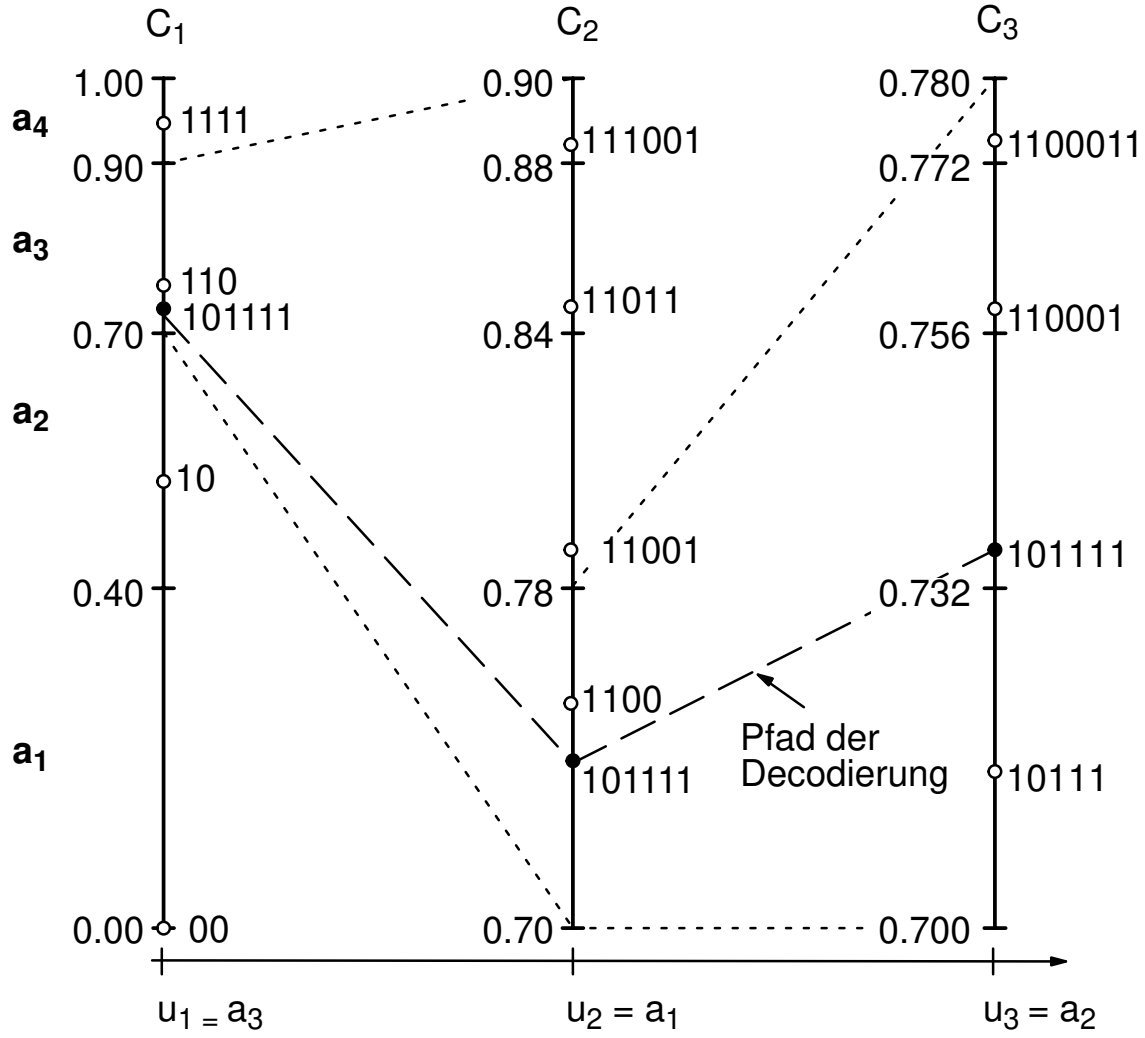


Bild 2.13 Darstellung der arithmetischen Decodierung einer empfangenen Codesymbolfolge 101111 in *(u_1=a_3, u_2=a_1, u_3=a_2)*

Eine empfangene Codesymbolfolge wird mit den Intervallgrenzen verglichen und dabei festgestellt, in welchem Intervall die Codesymbolfolge liegt. Im Beispiel der empfangenen Codesymbolfolge 101111 kann nach 5 empfangenen binären Codesymbolen 10111 bereits entschieden werden, daß die Codesymbolfolge im Intervall des Quellensymbols a_3 liegt. Somit kann $u_1 = a_3$ decodiert werden. Im nächsten Schritt liegt die Codesymbolfolge 10111 im Intervall von a_1, d.h. $u_2 = a_1$. Die Decodierung ist im behandelten Beispiel aufgrund von $L = 3$ im dritten Schritt beendet. Dieser liefert $u_3 = a_2$, da 101111 im Intervall von a_2 liegt.

Bild 2.13 enthält neben den Codewörtern des Blockcode für $L = 3$ entlang der Ordinate C_3 auch die der Blockcodes für $L = 1$ und $L = 2$. Im letzten Schritt einer

Block – Decodierung muß C_L mit einem der eingetragenen möglichen Codewörter übereinstimmen. Da die arithmetische Codierung keine Präfixeigenschaften besitzt, wird zur Erkennung des Blockendes ein spezielles a_k als End of Block–Symbol verwendet.

Die der arithmetischen Codierung zugrundeliegende Blockcodierung mit fortlaufender Codierung und Decodierung hat den Vorteil, daß lediglich K Intervallgrenzen, nicht aber eine Codetabelle mit K^L Codewörtern gespeichert werden muß.

Die K Intervallgrenzen ergeben sich aus der Wahrscheinlichkeit der K Quellensymbole. Indem die Häufigkeiten des Auftretens der Quellensymbole $a_1,\ ...,\ a_K$ in K Zählerständen fortlaufend überprüft werden, kann eine beobachtete veränderte Häufigkeitsverteilung dazu benutzt werden, die Intervall–Unterteilung zu ändern und damit den Code adaptiv zu gestalten. Da das Codewort erst im L –ten Schritt aus der dann vorliegenden Intervall–Unterteilung berechnet wird, brauchen die verwendeten verschiedenen Codes auch in diesem Fall nicht gespeichert zu werden. Darin liegt ein besonderer Realisierungsvorteil der arithmetischen Codierung.

2.2.6 Codierung diskreter gedächtnisbehafteter Quellen

Es sei angenommen, daß eine diskrete, gedächtnisbehaftete Quelle eine zeitliche Folge von statistisch abhängigen Quellensymbolen $u_1,\ ...,\ u_N$ erzeugt, die alle einem Symbolvorrat von K Symbolen $a_1,\ ...,\ a_K$ entstammen. Es sei ferner angenommen, daß die betrachtete gedächtnisbehaftete Quelle eine diskrete Markoff–Quelle sei, bei der ein beliebiges Symbol u_N einer Folge von jeweils nur $N-1$ vorausgehenden Symbolen $u_1,\ ...,\ u_{N-1}$ statistisch abhängig ist. Jede Kombination von $N-1$ vorausgehenden Symbolen $u_1,\ ...,\ u_{N-1}$ kennzeichnet einen Zustand s_j der Quelle. Mit der Erzeugung des Symbols u_N geht die Quelle in einen neuen Zustand über. Der bedingte mittlere Informationsgehalt je Quellensymbol sei $H(U_N|U_1,\ ...,\ U_{N-1})$.

Auf diese Markoff–Quelle können die bereits für gedächtnislose Quellen gewonnenen Ergebnisse angewendet werden, indem man jeden einzelnen Zustand s_j der Quelle für sich betrachtet. Zur Codierung der möglichen Symbole u_N wird in jedem Zustand s_j ein spezieller Präfixcode mit den Codewortlängen n_j gemäß

$$\frac{-\ log_2\ P(u_N|s_j)}{\log_2 D} \leq n_j < \frac{-\ log_2\ P(u_N|s_j)}{\log_2 D} + 1 \tag{2.72}$$

angewendet. Die mittlere Codewortlänge $\overline{n}_j$ je Quellensymbol für den im Zustand s_j angewandten Code ergibt sich dann aus dem Mittel über alle u_N mit der Grenze

$$\frac{H(U_N \mid s_j)}{\log_2 D} \leq \overline{n}_j < \frac{H(U_N \mid s_j)}{\log_2 D} + 1 \,. \tag{2.73}$$

Mittelt man weiterhin über alle Zustände s_j, d.h. über alle Folgen $u_1,\ ...,\ u_{N-1}$, so erhält man für die Markoff–Quelle mit dem bedingten mittleren Informationsgehalt $H(U_N|U_1,\ ...,\ U_{N-1})$ je Symbol eine mittlere Codewortlänge $\overline{n}$ von

$$\frac{H(U_N \mid U_1,\ ...,\ U_{N-1})}{\log_2 D} \leq \overline{n} < \frac{H(U_N \mid U_1,\ ...,\ U_{N-1})}{\log_2 D} + 1 \,. \tag{2.74}$$

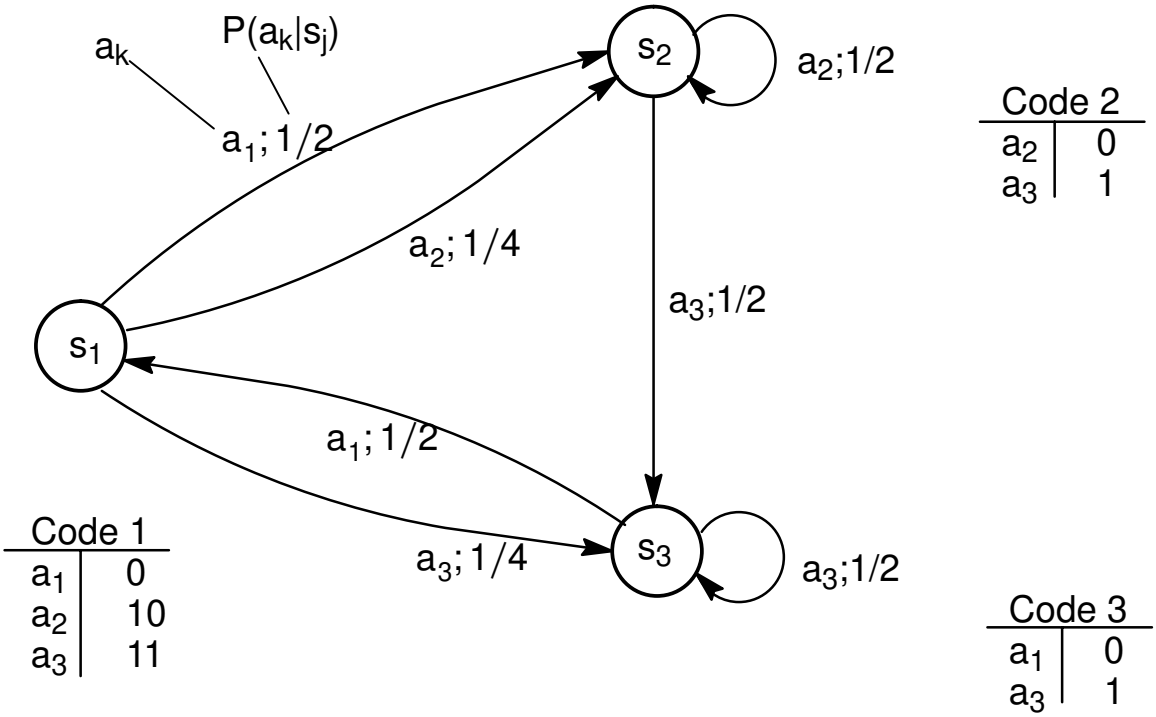


Bild 2.14 Graphische Darstellung einer Markoff–Quelle und ihrer Codierung

Ein derartiger Code ist redundanzfrei, wenn die mittlere Codewortlänge $\overline{n} = H(U_N|U_1,\ ...,\ U_{N-1})/\log_2 D$ ist, d.h. in (2.74) das Gleichheitszeichen gilt.

Beispiel: Bild 2.14 zeigt das Zustandsdiagramm einer Markoff–Quelle mit 3 Zuständen $s_1,\ s_2,\ s_3$ der Symbolwahrscheinlichkeit $P(U_N \mid s_j)$ der Zustände und die zugehörigen Huffman–Codes.

Da die Markoff–Quelle bis zu $J = K^{N-1}$ verschiedene Zustände besitzt, werden zur Realisierung der erläuterten Codierung von Markoff–Prozessen bis zu J verschiedene Präfixcodes benötigt. Die Realisierung dieser Codierung ist wegen des

damit verbundenen großen technischen Aufwandes nur in Sonderfällen durchführbar.

Die obere Grenze der Ungleichung (2.74) kann wiederum eingeschränkt werden, indem nicht einzelne Symbole u_N, sondern ganze Folgen $u_N,\ u_{N+1},\ ...,\ u_{N+L-1}$ von jeweils L Symbolen als ein Ereignis codiert werden. In Analogie zu Theorem 2.6 und der Blockcodierung (2.53) erhält man hier für die mittlere Codewortlänge die Grenzen

$$\frac{H(U_N \mid U_1,\ ...,\ U_{N-1})}{\log_2 D} \leq \overline{n} < \frac{H(U_N \mid U_1,\ ...,\ U_{N-1})}{\log_2 D} + \frac{1}{L}\,. \tag{2.75}$$

2.2.7 Universelle Codierung

Auch für jene Fälle, daß die Statistik der Quellensymbole unbekannt ist, sind Codierungsverfahren vorgeschlagen worden, die als universelle Codierungsalgorithmen bezeichnet werden. Sie versuchen, die Statistik während der Codierung fortlaufend zu messen, um anschließend durch Anpassung die Codierungseffizienz zu verbessern.

Als Beispiel einer universellen Codierung soll hier der Lempel–Ziv–Welch–Algorithmus beschrieben werden. In diesem Fall wird die Codierung mit Hilfe einer sogenannten String–Tabelle durchgeführt. Es werden Strings von Quellensymbolen codiert. Die String–Tabelle wird zu Anfang mit den ersten K Strings besetzt, die den K Quellensymbolen entsprechen.

Die Codierung beginnt mit dem aktuellen String S, der dem ersten Quellensymbol u_1 einer zu codierenden Folge $\vec{u}$ von Quellensymbolen entspricht. Dieser String $S = u_1$ ist bereits in der Tabelle enthalten.

Anschließend wird folgender Codierungsalgorithmus angewendet:

1. Falls keine weiteren Quellensymbole folgen, codiere S mit einem Codewort, das seinem Index j in der String–Tabelle entspricht. Andernfalls
2. Erweitere den String S um das nächste Quellensymbol P.
3. Falls SP in der String–Tabelle enthalten ist, bildet SP den neuen String S und es wird fortgefahren mit Schritt 1. Andernfalls
4. Codiere den zu S gehörenden Index j.
5. Erweitere die String–Tabelle um den String SP.
6. Setze $S = P$ und fahre fort mit Schritt 1.

Zur effektiven Codierung der Strings wird die Statistik der Strings fortlaufend überprüft und die Codewortlänge den Wahrscheinlichkeiten der Strings angepaßt. Die Tabelle enthält eine vereinbarte Anzahl von J Strings.

Beispiel: Codierung einer binären Quelle mit dem Lempel–Ziv–Welch–Algorithmus, siehe Bild 2.15:

Die zu codierende Folge der Quellensymbole $\vec{u} = u_1, u_2, ..., u_n$ sei 101110101... Der erste String S entspricht dem ersten Quellensymbol $S = u_1 = 1$. Da Schritt 1 nicht zutrifft, wird der String gemäß Schritt 2 erweitert zu $SP = u_1u_2 = 10$. Dieser String ist nicht in der Tabelle enthalten. Somit muß gemäß Schritt 4 der Index von $S = u_1 = 1$ d.h. $j = 2$ codiert und $SP = u_1u_2 = 10$ in der String–Tabelle unter dem Index $j = 3$ gespeichert werden.

Entsprechend Schritt 6 ist nun $S = u_2 = 0$ zu setzen und mit Schritt 1 fortzufahren. Gemäß Schritt 2 wird $S = u_2u_3 = 01$. Da dieser String nicht in der String–Tabelle enthalten ist, wird $j = 1$ codiert und $S = u_2u_3 = 01$ unter $j = 4$ gespeichert.

Entsprechend Schritt 6 ist nun $S = u_3 = 1$ zu setzen und mit Schritt 1 fortzufahren. Gemäß Schritt 2 wird $S = u_3u_4 = 11$, $j = 2$ codiert und $S = u_3u_4 = 11$ unter $j = 5$ gespeichert.

Entsprechend Schritt 6 ist nun $S = u_4 = 1$ zu setzen und mit Schritt 1 fortzufahren. Gemäß Schritt 2 wird $S = u_4u_5 = 11$. Da dieser String in der String–Tabelle enthalten ist, wird gemäß Schritt 2 $S = u_4u_5u_6 = 110$ gesetzt. Dieser String ist nicht in der String–Tabelle enthalten. Daher wird $j = 5$ codiert und $S = u_4u_5u_6 = 110$ unter $j = 6$ gespeichert.

Entsprechend Schritt 6 ist nun $S = u_6 = 0$ zu setzen und mit Schritt 1 fortzufahren. Gemäß Schritt 2 wird $S = u_6u_7 = 01$. Da dieser String in der String–Tabelle enthalten ist, wird gemäß Schritt 2 $S = u_6u_7u_8 = 010$ gesetzt. Dieser String ist nicht in der String–Tabelle enthalten. Daher wird $j = 4$ codiert und $S = u_6u_7u_8 = 010$ unter $j = 7$ gespeichert u.s.w.

Zu codierende Quellensymbole	u_1 1	u_2 0	u_3 1	u_4 1	u_5 1	u_6 0	u_7 1	u_8 0	u_9 1	....
Codierter Index	2	1	2	5		4				
S	1	0	1	1		0		0		
SP	10	01	11	11		01		01		
S				11		01		⋮		
SP				110		010				
⋮										
Speicherung 0 1	10	01	11	110		010				String–Tabelle
Speicheradresse j 1 2	3	4	5	6		7				

Bild 2.15 Darstellung einer universellen Codierung mit String–Tabelle nach Lempel–Ziv–Welch

2.3 Anwendungsbeispiele

2.3.1 Lauflängencodierung

Die Lauflängencodierung dient zur Codierung binärer Quellen. Bei der Lauflängencodierung werden Folgen u_1, ..., u_N der Quellensymbole a_0 bzw. a_1 in Abschnitte oder sog. Lauflängen gleicher Quellensymbole unterteilt und jede Lauflänge durch ein Codewort dargestellt. Die Lauflängencodierung soll hier am Beispiel der Faksimile–Übertragung erläutert werden.

Für eine digitale Faksimile–Übertragung von Schwarz–Weiß–Bildvorlagen wird das durch zeilenweise Abtastung gewonnene Bildsignal in zwei Amplitudenstufen quantisiert. Jeder Abtastwert stellt einen schwarzen oder weißen Bildpunkt dar, der einen Entscheidungsgehalt von $H_0 = 1$ bit besitzt. Um die Übertragungszeit mittels einer redundanzreduzierenden Quellencodierung zu verkürzen, werden die Bildpunktfolgen in Abschnitte aufeinanderfolgender Bildpunkte gleicher Amplitude aufgeteilt und für jeden Abschnitt die sog. Lauflänge in codierter Form übertragen, siehe Bild 2.16. Dabei werden weiße (w) und schwarze (s) Lauflängen mit unterschiedlichen Codes codiert.

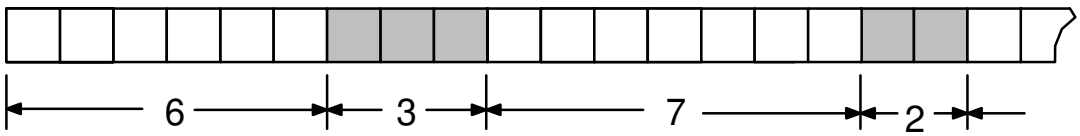


Bild 2.16 Lauflängencodierung (Run–length coding)

Legt man dieser Lauflängencodierung das Modell der gedächtnislosen Quelle zugrunde und geht dazu von der Annahme statistisch unabhängiger Lauflängen aus, so beträgt der mittlere Informationsgehalt H_w und H_s der weißen und schwarzen Lauflängen

$$\begin{aligned} H_w &= -\sum_{r=1}^{r_{\max}} P_w(r) \cdot \log_2 P_w(r) \\ H_s &= -\sum_{r=1}^{r_{\max}} P_s(r) \cdot \log_2 P_s(r)\,, \end{aligned} \tag{2.76}$$

wobei r die jeweilige Lauflänge und $P(r)$ die Wahrscheinlichkeit der Lauflänge ist. Mit den mittleren Lauflängen $\overline{r}_w$ und $\overline{r}_s$ gemäß

$$\begin{aligned} \overline{r}_w &= \sum_{r=1}^{r_{\max}} P_w(r) \cdot r \\ \overline{r}_s &= \sum_{r=1}^{r_{\max}} P_s\,(r) \cdot r \end{aligned} \tag{2.77}$$

folgt für die Entropie $H_{Bildpkt}$ eines Bildpunktes dieser Lauflängencodierung

$$H_{Bildpkt} = \frac{H_w + H_s}{\overline{r}_w + \overline{r}_s}\,. \tag{2.78}$$

In Abschnitt 2.3.2.1 wird eine Codierung der Lauflängen r_n, r_1 mittels zweier Huffman–Codes vorgestellt.

2.3.2 Codierung von Faksimilesignalen

Auflösung :	Vertikal	3,85 Zeilen/mm	(standard)
		7,7 Zeilen/mm	(optional)
	Horizontal	1728pel/215mm	(standard)
		2048pel/255mm	(optional)
		2432pel/303mm	(optional)
Papiergröße:	ISO A4–Größe (minimum)		
Codierung:	eindimensional:	MH–Code	
	zweidimensional:	MR–Code	
Protokoll:	Rec.T.30 (binär)		
Netzwerk:	PSTN		
Modulation und	V.27 ter	(4800 bit/s, 2400 bit/s)	
Demodulation:	V.29	(9600 bit/s, 7200 bit/s)	

Bild 2.17 Spezifikation der Faksimilegeräte der Gruppe 3

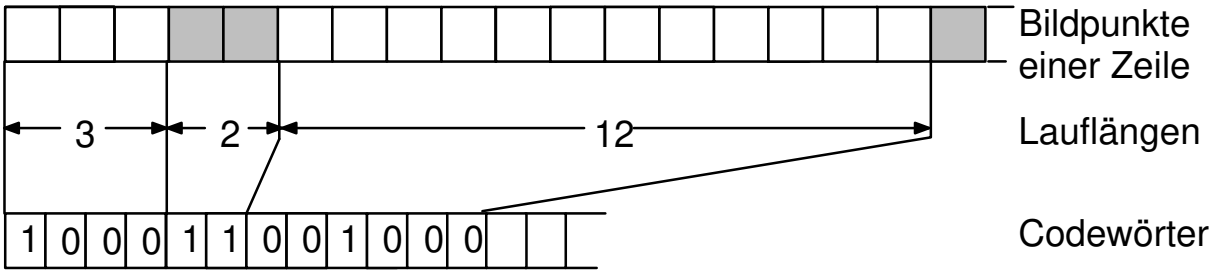
2.3.2.1 Eindimensionale Faksimile Codierung

Zur Codierung der weißen und schwarzen Lauflängen wendet man zwei Huffman–Codes an, deren Codewortlängen den Wahrscheinlichkeiten $P_w(r)$ und $P_s(r)$ angepaßt sind. Zur Verminderung des Realisierungsaufwandes werden jeweils nur die ersten 63 Lauflängen nach Huffman codiert und größere Lauflängen durch zwei Codewörter dargestellt, wobei das erste Codewort eine Lauflänge von $n \cdot 64$ und das zweite Codewort den verbleibenden Teil der Lauflänge darstellt. Die verbleibende Lauflänge kann auch Null sein.

Bild 2.18 zeigt den international standardisierten sog. Modified Huffman–Code für den Telefax–Dienst. Eine Lauflänge von $r_w = 130$ wird demzufolge mittels der

Binärzeichen 10010 0111 codiert. Die Decodierung ist aufgrund der Präfixeigenschaften der Codewörter eindeutig .

a)

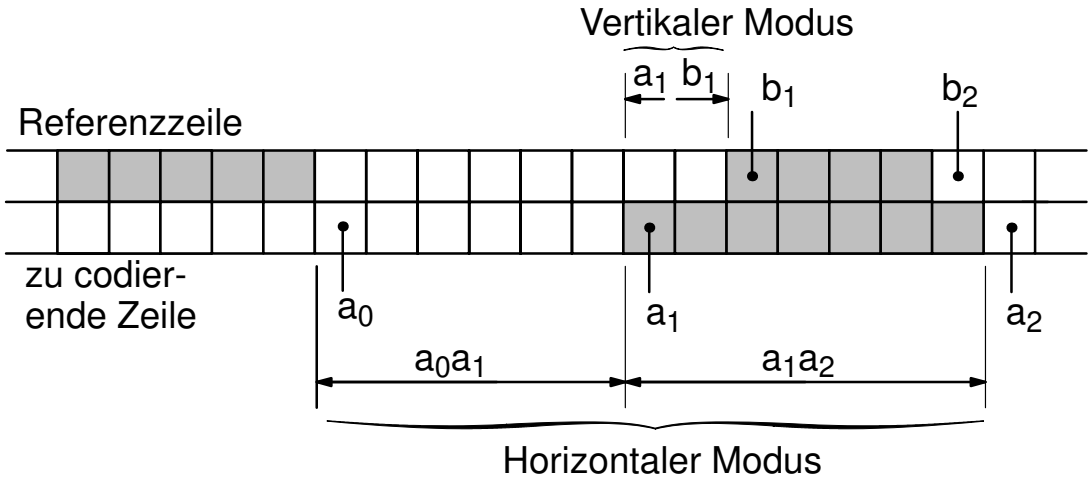


b)

Lauflänge	Codewörter	
	Weiße Lauflänge	Schwarze Lauflänge
0	00110101	0000110111
1	000111	010
2	0111	11
3	1000	10
.	.	.
.	.	.
63	00110100	000001100111
64	11011	0000001111
128	10010	000011001000
192	010111	000011001001
.	.	.
.	.	.
1728	010011011	0000001100101

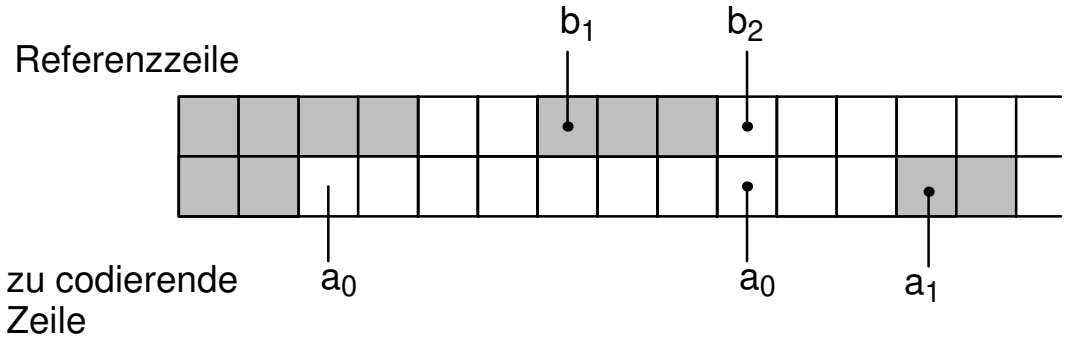
Bild 2.18 Modified Huffman–Code
a) Beispiel einer Codierung
b) Tabelle des Modified Huffman–Code

2.3.2.2 Zweidimensionale Faksimile Codierung



Vertikaler Modus: Wenn die relative Distanz $|a_1 b_1| \leq 3$, dann wird $a_1 b_1$ durch den vertikalen Modus codiert

Horizontaler Modus: Wenn die relative Distanz $|a_1 b_1| > 3$, dann werden $a_0 a_1$ und $a_1 a_2$ durch den horizontalen Modus codiert.



Pass–Modus: Wenn b_2 links von a_1 liegt, dann wird im Pass–Modus codiert.

Bild 2.19 Zweidimensionale Faksimile Codierung

Die zweidimensionale Faksimile Codierung nutzt neben der horizontalen statistischen Abhängigkeit der Bildpunktamplituden zusätzlich die vertikale statistische Abhängigkeit zur Verbesserung der Codierungseffizienz aus.

Eine zu codierende Bildzeile wird unter Bezug auf die vorausgehende, bereits übertragene Zeile codiert, siehe Bild 2.19. Dabei werden drei verschiedene Codierungs–

Modi angewendet, deren Codierungstabellen in Bild 2.20 zusammengefaßt sind. Im sogenannten horizontalen Modus wird dabei die Tabelle des eindimensionalen Modified–Huffman–Code angewendet.

Modus	zu codierende Elemente		Notation	Codewort
Pass	b_1,b_2		P	0001
Horizontal	a_0a_1, a_1a_2		H	001+M(a_0a_1)+M(a_1a_2)
Vertikal	a_1 unter b_1	a_1b_1=0	V(0)	1
	a_1 rechts von b_1	a_1b_1=1	V_R(1)	011
		a_1b_1=2	V_R(2)	000011
		a_1b_1=3	V_R(3)	0000011
	a_1 links von b_1	a_1b_1=1	V_L(1)	010
		a_1b_1=2	V_L(2)	000010
		a_1b_1=3	V_L(3)	0000010
Erweiterung				0000001xxx

Bild 2.20 Modified Reed–Code
(*M(r)* ist ein Codewort des Modified Huffman–Code)

Weitere zur Codierung eines binären Faksimilesignals anwendbare Quellenmodelle sind in Bild 2.21 zusammen mit den zugehörigen Entropien je Bildpunkt dargestellt. Da die Entropien den unteren Grenzwerten für die erreichbare Übertragungsrate entsprechen, erlauben sie auch eine Bewertung unterschiedlicher Quellenmodelle.

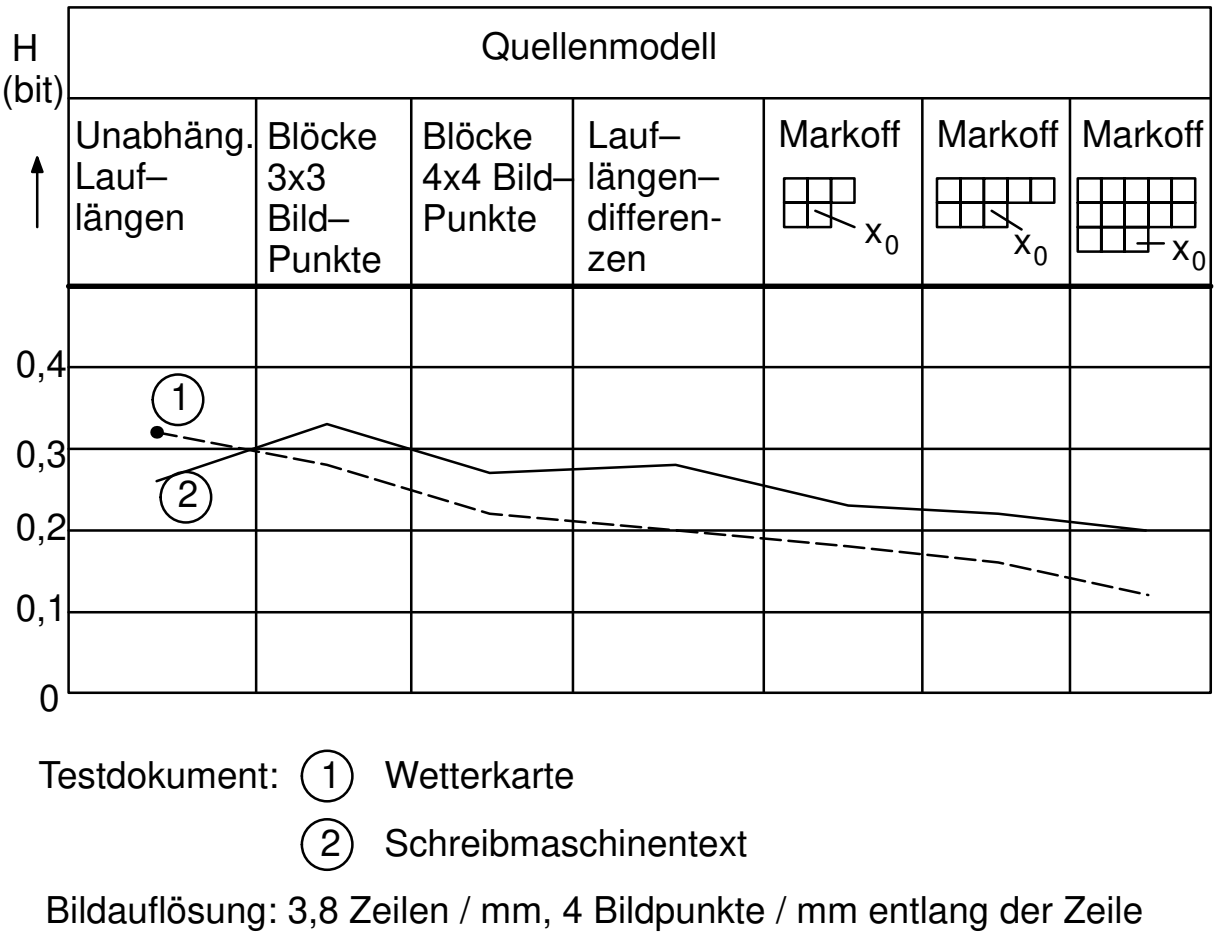


Bild 2.21 Gemessene Entropie H je Bildpunkt für verschiedene Quellenmodelle

2.3.2.3 Codierung einer binären Markoff–Quelle

Zur Anwendung des Modells einer binären Markoff–Quelle mit $K = 2$ und dem mittleren Informationsgehalt $H(U_N|U_1,\ ...,\ U_{N-1})$ müßte entsprechend Abschnitt 2.2.6 in jedem Zustand s_j ein Huffman–Code mit 2 Codewörtern angewendet werden. Dieser Huffman–Code hätte die Codewörter 0 und 1 und ist demzufolge für die redundanzmindernde Codierung von schwarz–weißen Faksimilesignalen mit mittlerem Informationsgehalt unter 1 bit je Bildpunkt unbrauchbar.

Daher wird zunächst eine zweidimensionale zustandsabhängige Prädiktion angewendet. Ein zu codierender Bildpunkt u_N wird mit seinem Prädiktionswert $\hat{u}_N$ verglichen, der aus den Umgebungspunkten $u_1,\ ...,\ u_{N-1}$ des Markoff–Modells gewonnen wird. Wenn der Prädiktionswert mit u_N übereinstimmt, wird an der Stelle

u_N im sogenannten Prädiktionsfehlerbild der Bildpunkt $e_N = 0$, ist die Prädiktion falsch, wird $e_N = 1$ gesetzt, d.h.

$$e_N = u_N \oplus \hat{u}_N. \tag{2.79}$$

Das entstehende Prädiktionsfehlerbild ist eine umkehrbar eindeutige Transformation des Originalbildes. Mit Hilfe der Prädiktionsvorschrift läßt sich aus dem Prädiktionsfehlerbild das Originalbild exakt rekonstruieren. Zur Rekonstruktion des Originalbildes wird im Decoder ebenfalls $\hat{u}_N$ berechnet und mit der übertragenen Information e_N gemäß

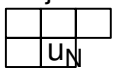
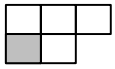
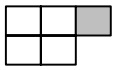
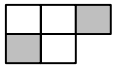
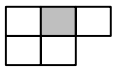
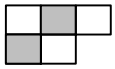
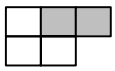
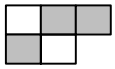
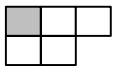
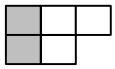
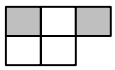
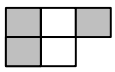
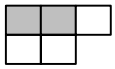
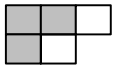
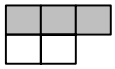
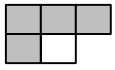
$$u_N = \hat{u}_N \oplus e_N \tag{2.80}$$

$\hat{u}_N$ rekonstruiert. Demzufolge kann das Prädiktionsfehlerbild anstelle des Originalbildes in codierter Form übertragen werden.

Beispiel: Bild 2.22 zeigt, wie der Prädikationswert $\hat{u}_N$ einer binären Markoff–Quelle mit $N - 1 = 4$ gewonnen werden kann. Jede Kombination der zur Prädiktion verwendeten Bildpunkte $u_1,\ u_2,\ u_3,\ u_4$ definiert einen Zustand s_j der Quelle. Es gibt $2^4 = 16$ verschiedene Zustände. Der aktuelle Zustand der Quelle ist bei Coder und Decoder bekannt, da er durch bereits übertragene Bildpunkte definiert ist. Bei gegebenem Zustand s_j wird als Prädiktionswert $\hat{u}_N$ der jeweils wahrscheinlichere Wert $\hat{u}_N = 0$ oder 1 verwendet, je nachdem, ob $P(u_N = 0|s_j)$ oder $P(u_N = 1|s_j)$ größer ist, d.h.

$$\hat{u}_N = \begin{cases} 0 & , \text{ falls } P(u_N = 0|s_j) \geq P(u_N = 1|s_j) \\ 1 & , \text{ falls } P(u_N = 1|s_j) > P(u_N = 0|s_j)\ . \end{cases} \tag{2.81}$$

Beispielsweise ist gemäß der Prädiktionstabelle von Bild 2.22 im Zustand s_2 die bedingte Wahrscheinlichkeit für einen "weißen" Bildpunkt größer und daher $\hat{u}_N = 0$. Befindet sich u_N im Randbereich, werden nicht definierte Bildpunkte $u_1,\ ...,\ u_{N-1}$ als weiß, d.h. $u_n = 0$ angenommen.

Zustand s_j	$p(u_N \mid s_j)$ $u_N = 0$	$u_N = 1$	Prädiktionswert $\hat{u}_N$
s_0 	0,996	0,004	0
s_1 	0,499	0,501	1
s_2 	0,810	0,190	0
s_3 	0,048	0,952	1
s_4 	0,115	0,885	1
s_5 	0,080	0,920	1
s_6 	0,468	0,532	1
s_7 	0,009	0,991	1
s_8 	0,989	0,012	0
s_9 	0,553	0,447	0
s_{10} 	0,912	0,088	0
s_{11} 	0,337	0,663	1
s_{12} 	0,391	0,609	1
s_{13} 	0,232	0,768	1
s_{14} 	0,616	0,384	0
s_{15} 	0,089	0,911	1

0 = weißer Bildpunkt 1 = schwarzer Bildpunkt

Bild 2.22 Bedingte Wahrscheinlichkeiten und Prädiktionswerte für eine Markoff–Quelle mit 16 Zuständen

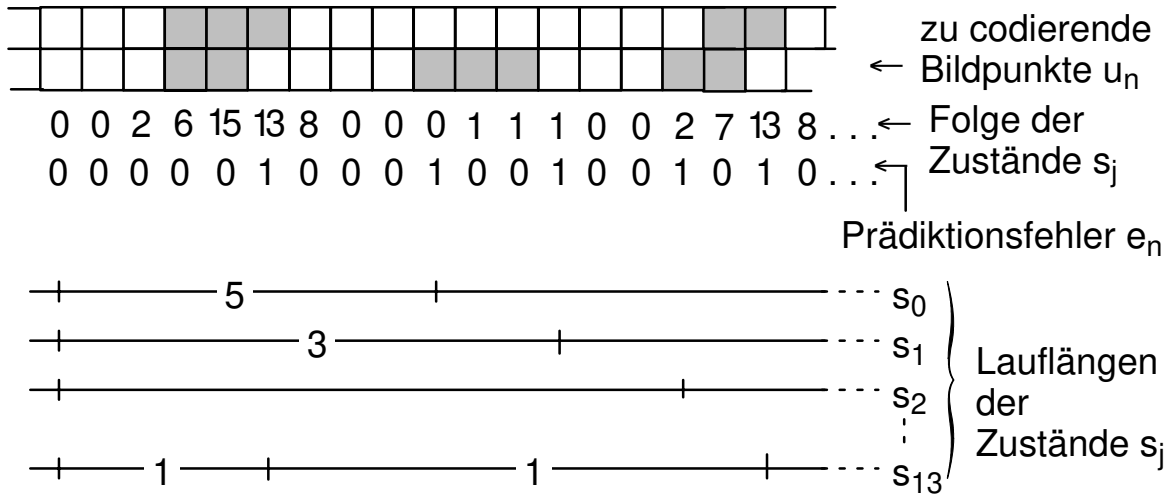


Bild 2.23 Darstellung einer nach Zuständen getrennten Lauflängencodierung von Prädiktionsfehlern

Wird die Prädiktion gemäß Bild 2.22 auf die in Bild 2.23 dargestellten Bildpunkte eines Faksimile–Ausschnittes angewendet, dann durchläuft die Quelle die in Dezimalzahlenfolge angegebenen Zustände. An den Stellen $e_N = 1$ treten dabei Prädiktionsfehler auf. Die Prädiktionsfehler werden nun sortiert nach Zuständen codiert. Im gezeigten Beispiel befindet sich die Quelle 5 mal im Zustand s_0 bis der erste Prädiktionsbefehl im Zustand s_0 auftritt. Für den Zustand s_0 wird demgemäß die Lauflänge 5 codiert und übertragen, Prädiktionsfehler anderer Zustände werden überschlagen. Nachdem die Lauflängen des Zustandes s_0 codiert sind, wiederholt sich der gleiche Vorgang für die Zustände s_1, s_2 usw.

In jedem Zustand s_j wird ein Lauflängen–Code verwendet, der den Lauflängenwahrscheinlichkeiten $P_j(r)$ des Zustandes s_j individuell angepaßt ist. Zur Decodierung werden die Lauflängen der Zustände in der Reihenfolge s_0, s_2, s_6, s_{15}, s_{13}, ... usw. benötigt.

Für diese Codierung einer Markoff–Quelle soll nachfolgend die mittlere Codewortlänge $\overline{n}$ je Bildpunkt berechnet werden. Dazu wird zunächst der Zustand s_j betrachtet. Die Wahrscheinlichkeit $P_j(r)$ der Lauflänge r im Zustand s_j berechnet sich mit $P(u_N = \hat{u}_N|s_j) = q_j$ und $P(u_N \neq \hat{u}_N|s_j) = p_j$

$$P_j(r) = q_j^{r-1} \cdot p_j \tag{2.82}$$

wobei $q_j = 1 - p_j$.

Unter der Annahme einer redundanzfreien Codierung der Lauflängen mit der Codewortlänge

$$n_j(r) = -\log_2 P_j(r) \tag{2.83}$$

erhält man für die mittlere Codewortlänge je Bildpunkt im Zustand s_j

$$\overline{n}_j = -\frac{1}{\overline{r}_j}\left[\sum_{r=1}^{\infty} P_j(r)\log_2 P_j(r)\right], \tag{2.84}$$

wobei $\overline{r}_j$ die mittlere Lauflänge

$$\begin{aligned}\overline{r}_j &= \sum_{r=1}^{\infty} r\, P_j(r) = \sum_{r=1}^{\infty} r\, q_j^{r-1}\, p_j \\ &= \frac{1}{(1-q_j)^2}\, p_j = \frac{1}{p_j^2}\cdot p_j = \frac{1}{p_j}\end{aligned} \tag{2.85}$$

ist. Durch Einsetzen von $P_j(r)$ und $\overline{r}_j$ in $\overline{n}_j$ erhält man

$$\overline{n}_j = -p_j\left[\sum_{r=1}^{\infty} P_j(r)\log_2 (q_j^{r-1}\cdot p_j)\right]$$

$$\overline{n}_j = -p_j\left[\sum_{r=1}^{\infty}(r-1)P_j(r)\log_2 q_j + \sum_{r=1}^{\infty} P_j(r)\log_2 p_j\right]$$

$$\overline{n}_j = -p_j\left[\sum_{r=1}^{\infty} r\, P_j(r)\log_2 q_j - \log_2 q_j + \log_2 p_j\right]$$

$$\overline{n}_j = -p_j\left[\frac{1}{p_j}\log_2 q_j - \log_2 q_j + \log_2 p_j\right]$$

$$\overline{n}_j = -\left[\log_2 q_j - p_j\log_2 q_j + p_j\log_2 p_j\right]$$

$$\overline{n}_j = -\left[q_j\log q_j + p_j\log p_j\right] \tag{2.86}$$

bzw.

$$\overline{n}_j = H(U_N|s_j)\,. \tag{2.87}$$

Die Mittelwertbildung über alle Zustände s_j liefert

$$\overline{n} = H(U_N|U_1,\ \ldots,\ U_{N-1})\,. \tag{2.88}$$

Damit ist nachgewiesen, daß diese von D. Preuß [14] vorgeschlagene Codierung die untere Grenze des Theorems der Quellencodierung einhält.

2.4 Irrelevanzreduzierende Codierung

2.4.1 Rate Distortion Theorie

Das Ziel der folgenden Betrachtungen ist die Berechnung einer unteren Grenze der Übertragungsbitrate für eine gegebene Nachrichtenquelle, die eine Rekonstruktion der Nachricht unter Einhaltung eines vorgegebenen Gütekriteriums gewährleistet. Die theoretischen Grundlagen dieser Berechnungen liefert die Rate–Distortion–Theorie.

Mit Hilfe der Rate–Distortion–Theorie können neue und bekannte Quellencodierungen wie beispielsweise die Amplitudenquantisierung im Vergleich zur berechneten unteren Grenze der Übertragungsrate hinsichtlich ihrer Effektivität bewertet werden. Eine Schwierigkeit der Rate–Distorsion–Theorie besteht in der Definition und Anwendung geeigneter Gütekriterien.

Es wird zunächst das Beispiel einer diskreten, gedächtnislosen Quelle U mit einem Symbolvorrat $(0,\ 1,\ ...,\ k,\ ...,\ K-1)$ betrachtet. Dabei sei K endlich und $Q(k)$ die Wahrscheinlichkeit der Quellensymbole k. Das Ausgangssignal dieser Quelle ist eine Symbolfolge $u_1,\ u_2,\ ...$ von statistisch unabhängigen Symbolen, wobei jedes u für ein beliebiges Symbol des Symbolvorrats $(0,\ 1,\ ...,\ K-1)$ steht. Nach Codierung, fehlerhafter Übertragung und Decodierung möge der Sinke eine Symbolfolge $v_1,\ v_2,\ ...$ zugeführt werden, wobei jedes v für ein Symbol des Symbolvorrats $(0,\ 1,\ ...,\ j,\ ...,\ J-1)$ mit $J < \infty$ steht.

Mit Hilfe des Verzerrungsmaßes $d(k;\ j)$, das für alle $0 \le k \le K-1$ und $0 \le j \le J-1$ definiert ist, wird jeder möglichen Darstellung eines gesendeten Symbols k durch ein empfangenes Symbol j ein numerischer Wert als Maß für die Verzerrung oder Verfälschung beigemessen. Aufgrund der angenommen fehlerfreien Übertragung ist diese Verzerrung allein durch die Quellencodierung verursacht. Es sei vorausgesetzt, daß $d(k;\ j) \ge 0$ für alle $k,\ j$ gilt und für jedes k mindestens ein j existiert, für das $d(k;\ j) = 0$ gilt. Nach dieser Definition bedeutet ein großes $d(k;\ j)$ eine große Verfälschung des Symbols k. Die Verzerrungsmaße $d(k;\ j)$ können in einer $(K,\ J)$–Matrix zusammengestellt werden.

Zur Erläuterung seien zwei Beispiele angeführt:

Beispiel: 1. Beispiel: Annahme: $J = K$

$$d(k;\ j) = 0 \quad \text{für} \quad j = k$$
$$d(k;\ j) = 1 \quad \text{für} \quad j \neq k$$

Dieses Verzerrungsmaß gewichtet jede Verfälschung eines Symbols k gleich stark.

2. Beispiel: Annahme: $J = K$

$$d(k;\ j) = (j - k)^2 \quad \text{für alle } j \text{ und } k$$

Dieses Verzerrungsmaß ist für Symbole geeignet, die Amplitudenwerte darstellen und entspricht dann dem quadratischen Fehler.–

Zur Berechnung der unteren Grenze der Übertragungsrate werden Quellencoder und Quellendecoder zusammen als ein gestörter Kanal betrachtet, wobei nur der Quellencoder Störungen erzeugt. Demzufolge kann die erforderliche Übertragungsrate am Ausgang des Quellencoders mittels der Transinformation $F(Q;\ P)$ berechnet werden. Mit den Übergangswahrscheinlichkeiten $P(j|k)$ und den Wahrscheinlichkeiten $Q(k)$ der Quellensymbole ergibt sich für die mittlere Transinformation $F(Q;\ P) = H(U) - H(U|V)$

$$F(Q;\ P) = \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} Q(k) \cdot P(j|k) \cdot \log_2 \frac{P(j|k)}{\sum_{i=0}^{K-1} Q(i) \cdot P(j|i)} \tag{2.89}$$

und die mittlere Verzerrung $\overline{d}$

$$\overline{d} = \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} Q(k) \cdot P(j|k) \cdot d(k;\ j)\ . \tag{2.90}$$

Während bei der Berechnung der Transinformation eines Übertragungskanals die Übertragungswahrscheinlichkeiten $P(\,j|k)$ durch die physikalischen Eigenschaften des Kanals vorgegeben sind, können diese im vorliegenden Fall mit der Quellencodierung so gewählt werden, daß sich eine möglichst kleine Übertragungsrate ergibt. Dabei muß die als Gütekriterium vorgegebene maximale mittlere Verzerrung d^* eingehalten werden. Das Ergebnis dieser Optimierung ist die Rate–Distortion–Funktion $R(d^*)$

$$R(d^*) = \min_{P:\overline{d} \le d^*} F(Q;\ P)\ . \tag{2.91}$$

Es kann nachgewiesen werden, daß eine Quelle stets mit einer Übertragungsbitrate entsprechend $R(d^*)$ codiert werden kann, ohne daß die vorgegebene zulässige Verzerrung d^* um mehr als ein beliebig kleines δ überschritten wird. Umgekehrt läßt sich zeigen, daß eine kleinere Übertragungsrate als $R(d^*)$ nur durch Überschreiten der zulässigen Verzerrung d^* möglich ist. $R(d^*)$ kann deshalb als untere Grenze der

für eine Quelle erforderlichen Übertragungsrate bei vorgegebenem Gütekriterium d^* angesehen werden.

Nachfolgend sollen einige Eigenschaften der Rate–Distortion–Funktion $R(d^*)$ erörtert werden.

Theorem 2.8:

$R(d^*)$ ist stets positiv.– (2.92)

Beweis:

Die Transinformation $F(Q;P)$ ist stets größer als Null.–

Theorem 2.9:

$R(d^*)$ ist eine mit wachsendem d^* fallende Funktion. – (2.93)

Beweis:

Die Menge der erlaubten Wahrscheinlichkeitsverteilungen $P(j|k)$ ist begrenzt und wächst mit d^*. Somit kann sich $R(d^*)$ mit wachsendem d^* nur vermindern.–

Theorem 2.10:

$R(d^*)$ ist eine konvexe U–Funktion, d.h. für jedes Paar von Verzerrungswerten d_1^* und d_2^* und beliebigen Wert $\lambda, 0 < \lambda < 1$, gilt die Ungleichung

$$R[\lambda \cdot d_1^* + (1-\lambda) \cdot d_2^*] \leq \lambda \cdot R(d_1^*) + (1-\lambda) \cdot R(d_2^*).– \quad (2.94)$$

Dieser Zusammenhang wird in Bild 2.24 veranschaulicht und bedeutet, daß $R(d^*)$ zwischen den Abszissenwerten d_1^* und d_2^* immer unter der Sehne liegt.

Für die Sehne gilt dabei die Geradengleichung

$$y(d^*) = R(d_2^*) + \frac{R(d_1^*) - R(d_2^*)}{d_1^* - d_2^*} \cdot \lambda \cdot (d_1^* - d_2^*)$$

$$y(d^*) = \lambda \cdot R(d_1^*) + (1 - \lambda) \cdot R(d_2^*) \quad . \quad (2.95)$$

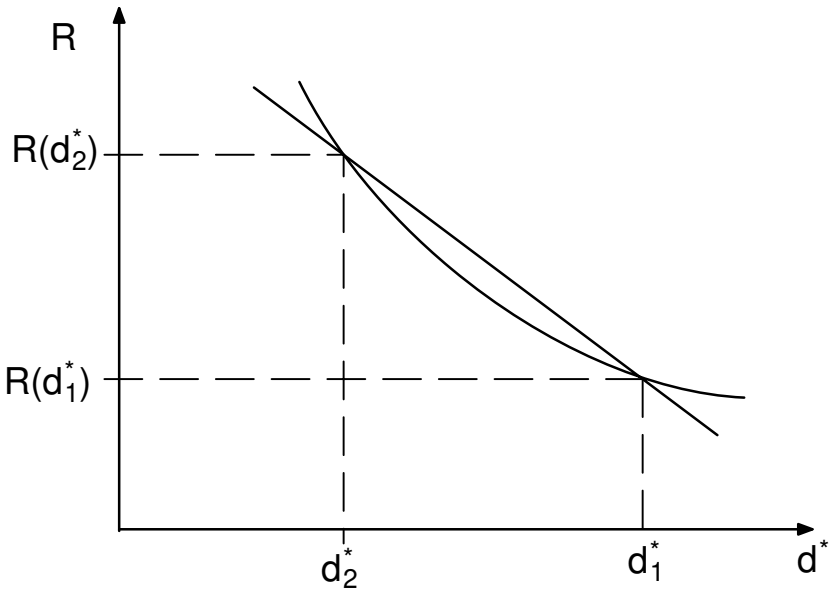


Bild 2.24 Eine konvexe *U*–Funktion

Beweis:

Beweis der Ungleichung (2.94) :
Es seien $P_1(j|k)$ und $P_2(j|k)$ die zu d_1^* und d_2^* gehörenden Verteilungen der Übergangswahrscheinlichkeiten, die Gl. (2.91) erfüllen, und $P'(j|k)$ eine Hilfsgröße, definiert durch

$$P'(j|k) = \lambda \cdot P_1(j|k) + (1 - \lambda) \cdot P_2(j|k). \tag{2.96}$$

Da $\overline{d}$ eine lineare Funktion von $P(j|k)$ ist, folgt

$$\begin{aligned} \overline{d}\,[\,P'(j|k)] &= \overline{d}\,[\,\lambda \cdot P_1(j|k) + (1 - \lambda) \cdot P_2(j|k)] \\ &= \overline{d}\,[\,\lambda \cdot P_1(j|k)] + [\overline{d}\,(1 - \lambda) \cdot P_2(j|k)] \\ &\lambda \cdot d_1^* + (1 - \lambda) \cdot d_2^*\,. \end{aligned} \tag{2.97}$$

Ferner gilt aufgrund von Gl. (2.91) und (2.97)

$$R[\lambda \cdot d_1^* + (1 - \lambda) \cdot d_2^*] \leq F(Q; P'). \tag{2.98}$$

Durch Einsetzen von Gl. (2.96) in $F(Q; P)$ und Erweiterung der Argumente der logarithmischen Funktionen erhält man

$$F(Q;P') = \lambda \sum_k \sum_j Q(k) \cdot P_1(j|k) \cdot \log_2\left[\frac{P'(j|k)}{P'(j)} \cdot \frac{P_1(j|k) \cdot P_1(j)}{P_1(j|k) \cdot P_1(j)}\right]$$
$$+ (1-\lambda) \sum_k \sum_j Q(k) \cdot P_2(j|k) \cdot \log_2\left[\frac{P'(j|k)}{P'(j)} \cdot \frac{P_2(j|k) \cdot P_2(j)}{P_2(j|k) \cdot P_2(j)}\right]. \quad (2.99)$$

Durch Zusammenfassen der Gleichungen (2.7), (2.8) erhält man

$$\log_2(x) \leq (x-1) \cdot \log_2(e). \quad (2.100)$$

Damit kann Gl. (2.99) umgeformt werden in

$$F(Q;P') \leq \lambda \sum_k \sum_j Q(k) \cdot P_1(j|k) \cdot \left[\log_2 \frac{P_1(j|k)}{P_1(j)} + \left(\frac{P'(j|k) \cdot P_1(j)}{P'(j) \cdot P_1(j|k)} - 1\right) \cdot \log_2(e)\right]$$
$$+ (1-\lambda) \sum_k \sum_j Q(k) \cdot P_2(j|k) \cdot \left[\log_2 \frac{P_2(j|k)}{P_2(j)} + \left(\frac{P'(j|k) \cdot P_2(j)}{P'(j) \cdot P_2(j|k)} - 1\right) \cdot \log_2(e)\right] \quad (2.101)$$

$$F(Q;P) \leq \lambda\left[F(Q;P_1) + \left[\sum_j \frac{P_1(j)}{P'(j)} \sum_k Q(k) \cdot P'(j|k) - 1\right] \cdot \log_2(e)\right]$$
$$+ (1-\lambda)\left[F(Q;P_2) + \left[\sum_j \frac{P_2(j)}{P'(j)} \sum_k Q(k) \cdot P'(j|k) - 1\right] \cdot \log_2(e)\right]. \quad (2.102)$$

Da die Summation $\sum_k Q(k) \cdot P'(j|k) = P'(j)$ ergibt, folgt aus Gl. (2.102)

$$F(Q,\ P') \leq \lambda \cdot F(Q,\ P_1) + (1-\lambda) \cdot F(Q,\ P_2)$$

$$F(Q,\ P') \leq \lambda \cdot R(d_1^*) + (1-\lambda) \cdot R(d_2^*)\,. \quad (2.103)$$

Die Kombination von Gl. (2.103) und (2.98) liefert den Beweis des Theorems 2.10.–

Bild 2.25 zeigt eine Skizze der Rate–Distortion–Funktion $R(d^*)$.

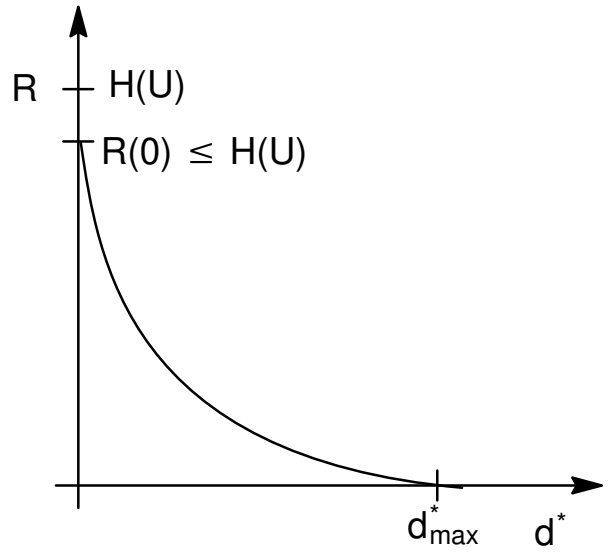


Bild 2.25 Typischer Verlauf der Rate–Distortion–Funktion $R(d^*)$ einer diskreten, gedächtnislosen Quelle

Der kleinste Wert der Verzerrung ist $d^* = 0$. In diesem Fall dürfen nur Symbolpaare $k,\ j$ auftreten, für die $d(k;\ j) = 0$ gilt. Der Maximalwert für $R(d^* = 0)$ ist $H(U)$. Es sei $d^*_{\max}$ als der kleinste Wert d^* definiert, für den $R(d^*) = 0$ wird, d.h.

$$d^*_{\max} = \min \sum_k \sum_j Q(k) \cdot P(j|k) \cdot d(k;\ j)\ . \tag{2.104}$$

Die Nebenbedingung $R(d^*_{\max}) = 0$ bedeutet, daß die Transinformation Null ist bzw. die Symbole k und j statistisch unabhängig sind; somit folgt für Gl. (2.104)

$$d^*_{\max} = \min \sum_j P(j) \sum_k Q(k) \cdot d(k;\ j)\ . \tag{2.105}$$

Das Minimum wird erzielt, indem die Summen $\sum_k Q(k) \cdot d(k;\ j)$ für alle j nacheinander überprüft werden und für das j mit der kleinsten Summe $P(j) = 1$ gewählt wird und Null in allen anderen Fällen. Die minimale Verzerrung $d^*_{\max}$ wird somit

$$d^*_{\max} = \min_j \sum_k Q(k) \cdot d(k;\ j). \tag{2.106}$$

Um für eine Übertragungsrate $R(d^*)$ eine möglichst geringe mittlere Verzerrung $\overline{d} \le d^*$ zu erreichen, muß die Quellencodierung der Quellenstatistik und dem

Verzerrungsmaß $d*$ angepaßt werden. Eine optimale Quellencodierung erfüllt Gl. (2.91).

Zur Codierung der diskreten, gedächtnislosen Quelle U wird eine Blockcodierung mit M Codewörtern angewendet, die eine Folge von L Quellensymbolen $\vec{u} = u_1,\ u_2,\ ...,\ u_L$ in eine Folge $\vec{v} = v_1,\ v_2,\ ...,\ v_L$ von Symbolen der Sinke abbildet, die hier als Codewort bezeichnet wird. Bild 2.26 zeigt ein Beispiel.

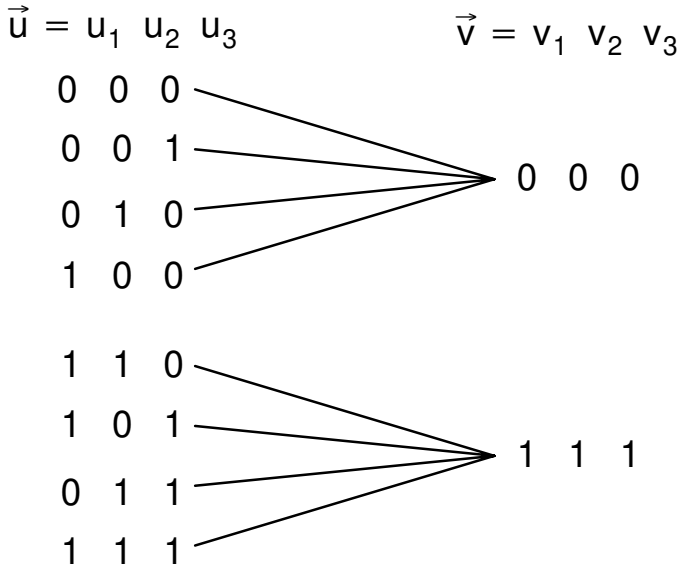


Bild 2.26 Codetabelle für eine Quellencodierung mit *L=3* und *M=2*

Bei gegebenen Verzerrungsmaß $d(k;\ j)$ kann die mittlere Verzerrung je Symbol berechnet werden zu

$$\bar{d}_L = \frac{1}{L}\sum_{\vec{u}} Q(\vec{u}) \cdot D\left[\vec{u};\ \vec{v}(\vec{u})\right], \tag{2.107}$$

wobei $\vec{v}(\vec{u})$ das $\vec{u}$ entsprechende Codewort ist und

$$D\left[\vec{u};\ \vec{v}(\vec{u})\right] = \sum_{l=1}^{L} d(u_1;\ v_1) \tag{2.108}$$

gilt. Da $\bar{d}_L$ die auf ein Symbol bezogene mittlere Verzerrung ist, kann Gl. (2.107) auch in der bereits bekannten Form

$$\bar{d}_L = \sum_j \sum_k Q(k) \cdot P(j|k) \cdot d(k;\ j)\,. \tag{2.109}$$

geschrieben werden.

Jedes der M Codewörter hat einen Informationsgehalt von $\log_2 M$ bit. Die Übertragungsbitrate je Codeelement beträgt dann $1/L \cdot \log_2 M$ und kann über einen Kanal der Kapazität $C = 1/L \cdot \log_2 M$ bit übertragen werden. Das zu lösende Problem besteht nun darin, bei vorgegebenen Werten für L und M einer Codierung die kleinstmögliche Verzerrung $\overline{d}_L$ zu erreichen.
Die Lösung dieses Problems ist mit den beiden folgenden Theoremen gegeben, die hier ohne Beweis angeführt werden.

Theorem der Quellencodierung bei Berücksichtigung eines Gütekriteriums

Theorem 2.11:

Es sei $R(d^*)$ die Rate–Distortion–Funktion einer diskreten, gedächtnislosen Quelle U. Dann existiert für jedes $d^* \geq 0,\ \delta > 0$ und hinreichend groß gewählter Blocklänge L ein Code mit M Codewörtern und der Übertragungsrate $1/L \cdot \log_2 M$, wobei

$$\frac{1}{L}\log_2 M \leq R(d^*) + \delta\ , \tag{2.110}$$

der die Ungleichung

$$\overline{d} \geq d^* + \delta \qquad \text{erfüllt.–} \tag{2.111}$$

Umgekehrtes Theorem der Quellencodierung bei Berücksichtigung eines Gütekriteriums

Theorem 2.12:

Ist $R(d^*)$ die Rate–Distortion–Funktion einer diskreten, gedächtislosen Quelle U, dann existiert kein Code, der das Gütekriterium $\overline{d} \leq d^*$ einhält und eine Übertragungsrate

$$\frac{1}{L}\log_2 M < R(d^*) \qquad \text{aufweist.–} \tag{2.112}$$

Nach diesen Überlegungen ist ein Übertragungssystem dann optimal genutzt, wenn es aus einem Quellencoder besteht, der die Übertragungsrate von $R(d^*)$ bit je Quellensymbol liefert, und einen Kanalcoder aufweist, der diese Übertragungsrate bis auf eine beliebig kleine Restfehlerwahrscheinlichkeit gegen Übertragungsfehler des Kanals einer Kanalkapazität $C > R(d^*)$ sichert.
Die für eine optimale Codierung mit der Übertragungsrate $R(d^*)$ erforderlichen Übertragungswahrscheinlichkeiten $P(j|k)$ werden im Quellencoder durch eine

Blockcodierung realisiert. Um eine eindeutige Decodierung zu gewährleisten, muß die Zuordnung zwischen den Symbolfolgen $\vec{u}$ und den Codewörtern $\vec{v}$ deterministisch sein, d.h. $P(\vec{v} \mid \vec{u}) = 0$ oder 1. Um gleichzeitig die gewünschten Übergangswahrscheinlichkeiten $P(j \mid k)$ zu erreichen, müssen die Zuordnungen dabei so gewählt werden. daß sich für die einzelnen Symbole die Übergangswahrscheinlichkeiten $P(j \mid k)$ des optimalen Code einstellen. Eine Codierung von Einzelsymbolen müßte immer deterministisch sein und kann beide Bedingungen daher nur in Sonderfällen erfüllen.

2.4.2 Berechnung der Rate–Distortion–Funktion $R(d^*)$ diskreter Quellen

Im folgenden soll ein Berechnungsverfahren zur Bestimmung der optimalen Übergangswahrscheinlichkeiten $P(j|k)$ abgeleitet werden, die auf eine Übertragungsrate $R(d^*)$ führen. Die Berechnung der Gl. (2.91)

$$R(d^*) = \min_{P(j|k)} F\,[\,P(j|k)] = \min_{P(j|k)} \sum_k \sum_j Q(k) \cdot P(j|k) \cdot \log \frac{P(j|k)}{P(j)} \tag{2.113}$$

mit

$$P(\,j) = \sum_i Q(i) \cdot P(j|i) \tag{2.114}$$

führt auf ein Extremwertproblem. Es ist das Minimum der Funktion $F[P(j|k)]$ unter den Nebenbedingungen $\overline{d} = d^*$, d.h.

$$\sum_k \sum_j Q(k) \cdot P(j \mid k) \cdot d(k;\ j) - d^* = 0 \tag{2.115}$$

und

$$\sum_j P(j|k) - 1 = 0 \qquad \text{für alle } k \tag{2.116}$$

bei positiven Wahrscheinlichkeiten $P(j|k)$ zu suchen. Die Nebenbedingung (2.115) folgt aus der Forderung $\overline{d} \le d^*$. Aufgrund der Eigenschaft (2.93) muß das Minimum von $R(d^*)$ bei $\overline{d} = d^*$ liegen.

Zur Lösung des Extremwertproblems wird das Lagrange'sche Verfahren herangezogen. Dazu werden die Nebenbedingungen zusammen mit Lagrange'schen Multiplikatoren μ_k, s zur minimierenden Funktion $F[P(j|k)]$ addiert und das Minimum der Hilfsfunktion $\Phi[P(j|k), \mu_k, s]$ berechnet, wobei

$$\Phi\Big[P(j|k),\ \mu_k,\ s\Big] = F\Big[P(j|k)\Big] + \sum_k \mu_k \cdot \left[\sum_j P(j|k) - 1\right] + s \cdot \left[\sum_k \sum_j Q(k) \cdot P(j|k) \cdot d(k;\ j) - d^*\right]. \tag{2.117}$$

Zur Vereinfachung der Berechnung wird der in $F[P(j|k)]$ enthaltene Logarithmus zur Basis e gerechnet, d.h. $R(d^{*)}$ in der Einheit "nat" berechnet.

$$\Phi\Big[P(j|k),\ \mu_k,\ s\Big] = + \sum_k \sum_j Q(k) \cdot P(j|k) \cdot \ln \frac{P(j|k)}{\sum_i Q(i) \cdot P(j|i)} + \sum_k \mu_k \cdot \left[\sum_j P(j|k) \cdot - 1\right] + s \cdot \left[\sum_k \sum_j Q(k) \cdot P(j|k) \cdot d(k;\ j) - d^*\right]. \tag{2.118}$$

Die Extremwertberechnung von $\Phi[P(j|k),\ \mu_k,\ s]$ liefert

$$\frac{\partial \Phi}{\partial P(j|k)} = 0 \qquad \text{für alle} \quad j,\ k \tag{2.119}$$

zunächst $K \cdot J$ Bestimmungsgleichungen und mit den Nebenbedingungen (2.115) und (2.116) zusätzlich $1 + K$ weitere Bestimmungsgleichungen zur Berechnung der insgesamt $K \cdot J + K + 1$ Unbekannten $P(j|k)$ und s.
Die Differentiation von GL. (2.118) ergibt für jedes spezielle Wertepaar $j,\ k$:

$$\frac{\partial \Phi}{\partial P(j|k)} = Q(k) \cdot \ln \frac{P(j|k)}{\sum_i Q(i) \cdot P(j|i)} + Q(k) - Q(k) \cdot P(j|k) \cdot \frac{Q(k)}{\sum_i Q(i) \cdot P(j|i)} - \frac{\sum_{i \neq k} Q(i) \cdot P(j|i) \cdot Q(k)}{\sum_i Q(i) \cdot P(j|i)} + \mu_k + s \cdot Q(k) \cdot d(k;\ j)\,. \tag{2.120}$$

Nach Zusammenfassung der Summanden folgt

$$\frac{\partial \Phi}{\partial P(j|k)} = Q(k) \cdot \left[\ln \frac{P(j|k)}{P(j)} + 1 - \frac{\sum_i Q(i) \cdot P(j|i)}{\sum_i Q(i) \cdot P(j|i)} \right] + \mu_k + s \cdot Q(k) \cdot d(k;\ j)\,. \tag{2.121}$$

Wird zur Vereinfachung der Rechnung ferner die Substitution

$$\mu_k = -\, Q(k) \cdot \ln \frac{f_k}{Q(k)} \tag{2.122}$$

eingeführt, so folgt für (2.121) mit (2.122)

$$\frac{\partial \Phi}{\partial P(j|k)} = Q(k) \cdot \ln \left[\frac{P(j|k) \cdot Q(k)}{P(j) \cdot f_k} + s \cdot d(k;\ j) \right] = 0\,. \tag{2.123}$$

Die Auflösung nach $P(j|k)$ liefert

$$P(j|k) = f_k \cdot \frac{P(j)}{Q(k)} \cdot e^{-s \cdot d(k;\, j)} \tag{2.124}$$

für alle Wertepaare $j,\ k$. Dadurch, daß sich die optimalen $P(j|k)$ in Gl. (2.124) als Produkt von $f_k \cdot P(j)$ ergeben, werden anstelle von $K \cdot J + K + 1$ Bestimmungsgleichungen nur noch $J + K + 1$ Bestimmungsgleichungen benötigt, die mit (2.124), (2.115) und (2.116) berechnet werden können. Die Unbekannten sind $P(j)$, f_k und s.
Werden beide Seiten der Gl. (2.124) mit $Q(k)$ multipliziert und über k aufsummiert, erhält man

$$1 = \sum_k f_k \cdot e^{-s \cdot d(k;\, j)} \tag{2.125}$$

für alle j. Mit (2.124) und der Nebenbedingung $\sum_j P(j|k) = 1$ folgt

$$1 = \frac{f_k}{Q(k)} \sum_j P(j) \cdot e^{-s \cdot d(k;\, j)} \tag{2.126}$$

für alle k. Aus der Nebenbedingung $\overline{d} = d^*$ ergibt sich mit der Lösung (2.124) die mittlere Verzerrung

$$d^* = \sum_k \sum_j f_k \cdot P(j) \cdot e^{-s \cdot d(k;\, j)} \cdot d(k;\ j) \tag{2.127}$$

als Bestimmungsgleichung für s. Für die Rate–Distortion–Funktion folgt

$$\begin{aligned} R(d^*) &= \sum_k \sum_j Q(k) \cdot P(j|k) \cdot \ln \frac{P(j|k)}{P(j)} \\ &= \sum_k \sum_j Q(k) \cdot P(j|k) \cdot \left[\ln \frac{f_k}{Q(k)} - s \cdot d(k;\ j) \right] \\ &= \sum_k Q(k) \cdot \ln \frac{f_k}{Q(k)} - s \cdot d* \\ &= \sum_k Q(k) \cdot \log_2 \frac{f_k}{Q(k)} - s \cdot d* \cdot \log_2 e \end{aligned} \tag{2.128}$$

Da $R(d^*)$ eine konvexe U–Funktion ist, muß das durch Gl. (2.128) dargestellte Extremum ein Minimum sein. Die Lösungen (2.127) und (2.128) werden auch als "Parameter–Form der Rate–Distortion–Funktion bezeichnet. Sie gelten nur für Bereiche mit $P(j|k) \geq 0$.

Es kann nachgewiesen werden, daß für die Ableitung $R(d^*)$ nach d^*, vgl. dazu T. Berger [1]

$$\frac{d[R(d^*)]}{d(d^*)} = -s \tag{2.129}$$

gilt. Somit kann $(-s)$ als das Steigungsmaß der Funktion interpretiert werden. Dabei ist s selbst noch eine Funktion von d^*.

2.4.3 Berechnung der Rate–Distortion–Funktion $R(d^*)$ kontinuierlicher Quellen

Die Quellensignale zeit– und wertkontinuierlicher Quellen, wie Sprachquellen und Bildquellen, sind im allgemeinen bandbegrenzt und können gemäß dem Abtasttheorem durch äquidistante Abtastwerte des Quellensignals dargestellt werden. Die Folge der Abtastwerte kann als ein zeitdiskretes, wertkontinuierliches Signal angesehen werden. Die folgenden Betrachtungen zur Rate–Distortion–Funktion kontinuierlicher Quellen beschränken sich auf diese für die Praxis sehr bedeutenden zeitdiskreten, wertkontinuierlichen Quellen.

Es sei $u_1,\ u_2,\ ...,\ u_L$ eine Folge von Quellensymbolen bzw. Abtastwerten, wobei jedes u für ein beliebiges Symbol x einer zeitdiskreten, wertkontinuierlichen Quelle steht. Die Quellensymbole x entstammen einem kontinuierlichen Wertevorrat

$-\infty \leq x \leq +\infty$. Es sei $q(u)$ die Wahrscheinlichkeitsdichte der Quellensymbole u. Durch eine Quellencodierung möge die ursprüngliche Folge der Quellensymbole in eine Folge empfangener Codesymbole $v_1,\ v_2,\ ...,\ v_L$ abgebildet werden, wobei jedes v für ein beliebiges Symbol y steht.

Die Verzerrung wird mit dem Verzerrungsmaß $d(u;\ v)$ gemessen, das jedem möglichen Wertepaar $u,\ v$ von Quellensymbol und Codesymbol einen numerischen positiven Zahlenwert zuordnet. Zu jedem u möge es mindestens einen Wert v geben, gewöhnlich $u = v$, für den $d(u;\ v) = 0$ gilt. Als Gütekriterium soll hier vor allem die quadratische Abweichung $d(u;\ v) = (v-u)^2$ betrachtet werden.

Die Rate–Distortion–Funktion wird für die zeitdiskrete, wertkontinuierliche Quelle in gleicher Weise definiert, wie für die diskrete Quelle. Anstelle der Summen treten hier lediglich Integrale auf.

$$R(d^*) = \min_{p(v|u)} \int_u \int_v q(u) \cdot p(v|u) \cdot \log_2 \frac{p(v|u)}{p(v)}\, du \cdot dv \tag{2.130}$$

Die Lösung der Gl. (2.130) unter der Nebenbedingung $\overline{d} \leq d^*$, wobei

$$\overline{d} = \int_u \int_v q(u) \cdot p(v|u) \cdot d(u;\ v)\, du \cdot dv \tag{2.131}$$

und der Nebenbedingung

$$\int_v p(v|u)\, dv = 1 \tag{2.132}$$

stellt ein Variationsproblem dar. Es kann gezeigt werden, daß $R(d^{*})$ auch in diesem Fall eine konvexe Funktion ist. Im Unterschied zur diskreten Quelle kann

$$\lim_{d^* \to 0} R(d^*) = \infty$$

sein, wenn der Wertevorrat der Quellensymbole unendlich groß ist. Ferner kann nachgewiesen werden, vergl. dazu R. G. Gallager [6], S. 471 ff., daß die Codierungstheoreme der diskreten Quelle auch für die zeitdiskrete, wertkontinuierliche Quelle gelten.

Die Berechnung von $R(d^*)$ führt auf folgende Lösungen, die hier an die Stelle der Gl. (2.124) bis (2.128) der diskreten Quelle treten:

$$p(v|u) = \frac{f(u) \cdot p(v)}{q(u)} \cdot e^{-s \cdot d(u;\ v)}, \tag{2.133}$$

wobei $f(u)$ die Bedingung

$$1 = \int_u f(u) \cdot e^{-s \cdot d(u;\, v)}\, du \tag{2.134}$$

für alle v erfüllt. Aus der Nebenbedingung (2.132) folgt hier

$$1 = \frac{f(u)}{q(u)} \int_v p(v) \cdot e^{-s \cdot d(u;\, v)} dv \tag{2.135}$$

für alle u. Für die Parameter–Darstellung der Rate–Distortion–Funktion ergibt sich mit diesen Lösungen

$$d^* = \int_u \int_v f(u) \cdot p(v) \cdot e^{-s \cdot d(u;\, v)} \cdot d(u;\ v)\, du\, dv \tag{2.136}$$

und

$$\begin{aligned} R(d^*) &= \int_u q(u) \cdot \ln \frac{f(u)}{q(u)}\, du - s \cdot d^* && \text{[nat]} \\ R(d^*) &= \int_u q(u) \cdot \log_2 \frac{f(u)}{q(u)}\, du - s \cdot d^* \cdot \log_2 e && \text{[bit]}\ . \end{aligned} \tag{2.137}$$

Mit der Wahl eines Gütekriteriums $d(u; v)$ kann $R(d^*)$ noch näher spezifiziert werden. Für den Sonderfall

$$d(u;\ v) = (v - u)^2 \tag{2.138}$$

d.h. der quadratischen Abweichung als Verzerrungsmaß folgt aus (2.134)

$$1 = \int_{-\infty}^{+\infty} f(u) \cdot \exp[-s \cdot (v - u)^2]\, du \tag{2.139}$$

für alle v. Über die Substitution $z = u - v$ im Integral

$$\int f(z + v) \cdot \exp[-s \cdot z^2]\, dz$$

erkennt man, daß die Forderung (2.139) nur eingehalten werden kann, wenn $f(u)$ eine Konstante ist, gegeben durch

$$f(u) = \sqrt{\frac{s}{\pi}} \; . \tag{2.140}$$

Durch Einsetzen von (2.140) in (2.133) erhält man

$$\frac{p(u|v) \cdot q(u)}{p(v)} = \sqrt{\frac{s}{\pi}} \cdot e^{-s \cdot (v-u)^2} = p(u|v) \; . \tag{2.141}$$

Die Konstante s wird mit Hilfe der Nebenbedingung (2.136) und der Substitution $z = u - v$ bestimmt. Daraus folgt

$$\begin{aligned} d^* &= \int_u \int_v \sqrt{\frac{s}{\pi}} \cdot p(v) \cdot e^{-s \cdot (v-u)^2} \cdot (v-u)^2 \, du \, dv \\ &= \int_v p(v) \, dv \int_z \sqrt{\frac{s}{\pi}} \cdot e^{-s \cdot z^2} \cdot z^2 \, dz = \frac{1}{2s} \; . \end{aligned} \tag{2.142}$$

Durch Umformung von Gl. (2.137) erhält man für $R(d^*)$

$$\begin{aligned} R(d^*) &= H(U) + \log_2 f(u) - s \cdot d^* \cdot \log_2 (e) \\ &= H(U) + \frac{1}{2} \cdot \log_2 \frac{s}{\pi} - s \cdot d^* \cdot \log_2 e \end{aligned} \tag{2.143}$$

und nach Einsetzen der Gl. (2.142)

$$R(d^*) = H(U) - \frac{1}{2} \cdot \log_2 (2\pi \cdot e \cdot d^*) \; . \tag{2.144}$$

Diese Beziehung für $R(d^*)$ gilt nur für das spezielle Gütekriterium $d(u; \; v) = (v-u)^2$. Der zweite Term auf der rechten Seite der Gl. (2.144) kann als die mögliche Entropiereduktion bei vorgegebenem Gütekriterium d^* interpretiert werden. Um $R(d^*)$ zu erreichen, muß die Verzerrung $v - u$ eine Gaußsche Wahrscheinlichkeitsdichte mit der Varianz d^* aufweisen.

Für eine zeitdiskrete, wertkontinuierliche, gedächtnislose Gaußschen Quelle mit der Entropie $H_G(U) = \frac{1}{2} \log_2 (2\pi e \sigma_u{}^2)$ und der Varianz $\sigma_u^2 = \sigma_x^2$ gilt

$$\begin{aligned} R_G(d^*) &= \frac{1}{2} \log_2 (2\pi e \sigma_x{}^2) - \frac{1}{2} \log_2 (2\pi e d^*) \\ R_G(d^*) &= \frac{1}{2} \log_2 \frac{\sigma_x{}^2}{d^*} \end{aligned} \tag{2.145}$$

in bit je Quellensymbol bzw. Abtastwert.

Im Grenzfall, daß alle u in $v = 0$ abgebildet werden, beträgt die mittlere Verzerrung $\overline{d} = \sigma^2$. Daraus folgt

$$R(d^*) = \frac{1}{2}\log_2 \frac{\sigma^2}{d^*} \qquad d^* < \sigma^2$$

$$R(d^*) = 0 \qquad\qquad d^* \geq \sigma^2$$

Eine Umrechnung des Signal–Rauschleistungs–Verhältnisses ${\sigma_x}^2/d^*$ in die Einheit "Dezibel" liefert

$$R_G(d^*) = \frac{1}{20} \cdot 10 \cdot \log_{10} \frac{\sigma_x^2}{d^*} \cdot \log_2 (10) = 0,166 \cdot \left(\frac{\sigma_x^2}{d^*}\right)_{\text{dB}} .$$

Diese Funktion ist in Bild 2.27 dargestellt. Das Signal–Rauschleistungs–Verhältnis wird auch mit "Signal–to–Noise–Ratio" (SNR) bezeichnet. Damit kann Gl. (2.145) auch in der Form

$$SNR_{\text{dB}} = 6,02 \cdot R_G(d^*) \tag{2.146}$$

geschrieben werden. Aus Gl. (2.146) ist zu entnehmen, daß jede Vergrößerung der Übertragungsrate um 1 bit je Abtastwert eine Verbesserung des Signal–Rauschleistungs–Verhältnisses von etwa 6 dB ergibt.

Bei nicht Gaußschen gedächtnislosen Quellen mit gegebener Varianz ${\sigma_x}^2$ ist $H(U)$ kleiner als die Entropie $H_G(U)$ der Gaußschen Quelle. Nach Gl. (2.145) muß die Rate Distortion–Funktion dieser Quelle unter der von Bild 2.27 liegen. Gl. (2.145) stellt somit eine obere Grenze für $R(d^*)$ dar.

Zum Vergleich sei eine zeitdiskrete, gedächtnislose Quelle mit konstanter Wahrscheinlichkeitsdichte und gleicher Varianz ${\sigma_x}^2$ betrachtet. Eine solche Quelle besitzt die Entropie

$$H(U) = \frac{1}{2}\log_2 (12{\sigma_x}^2) \tag{2.147}$$

und liegt damit um die Differenz

$$\frac{1}{2}\log_2 (2\pi e{\sigma_x}^2) - \frac{1}{2}\log_2 (12{\sigma_x}^2) = 0,25 \text{ bit}$$

unter der Grenze für $R(d^*)$ gegeben durch die Gaußsche Quelle gemäß (2.145). Der relativ geringe Abstand von der Grenze deutet darauf hin, daß auch die übrigen Quellen gleicher Varianz relativ nahe an der Grenze liegen.

Bei 2W Abtastwerten je Sekunde einer kontinuierlichen, bandbegrenzten Gaußschen Quelle mit den Bandgrenzen ergibt sich aus Gl. (2.145) eine Übertragungsbitrate von

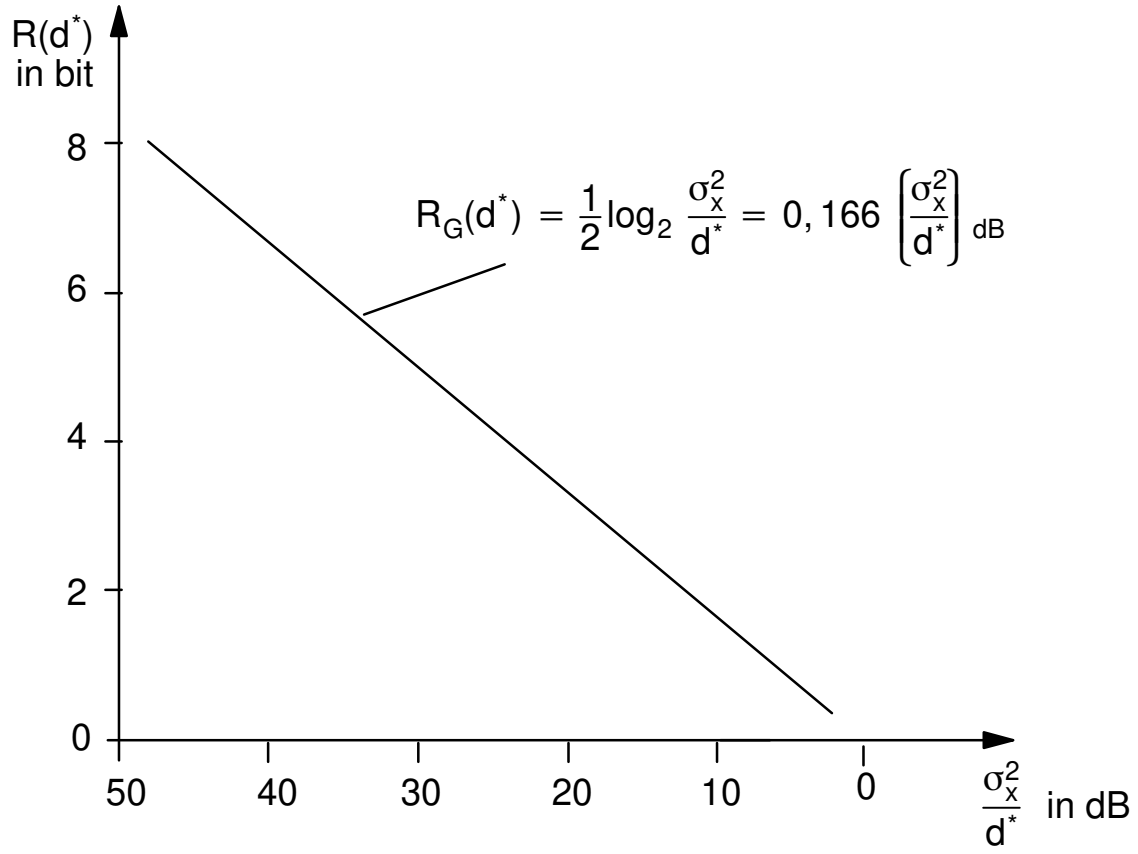


Bild 2.27 $R_G(d^*)$für eine gedächtnislose Gaußsche Quelle mit quadratischer Abweichung als Verzerrungsmaß

$$R(d^*) = W \cdot \log_2 \frac{{\sigma_x}^2}{d^*} \tag{2.148}$$

in der Einheit bit/s.

2.4.4 Berechnung der Rate–Distortion–Funktion einer Gaußschen Quelle mit dem Leistungsdichtespektrum *S*(*f*)

Für die Behandlung von Sprachsignal– und Bildsignalquellen ist das Modell der gedächtnisbehafteten Quelle von großer Bedeutung. Im folgenden soll für gedächtnisbehaftete Quellen, deren Leistungsdichtespektrum bekannt ist, eine obere Grenze für $R(d^*)$ ermittelt werden. Dazu wird von einer gedächtnisbehafteten Gaußschen Quelle mit der spektralen Leistungsdichte $S(f)$ ausgegangen. Diese gedächtnisbehaftete Gaußsche Quelle kann man in eine Summe von gedächtnislosen Gaußschen

Quellen mit jeweils weißem Spektrum und der Varianz $S(f)\,df$, wie in Bild 2.28 gezeigt, zerlegen.

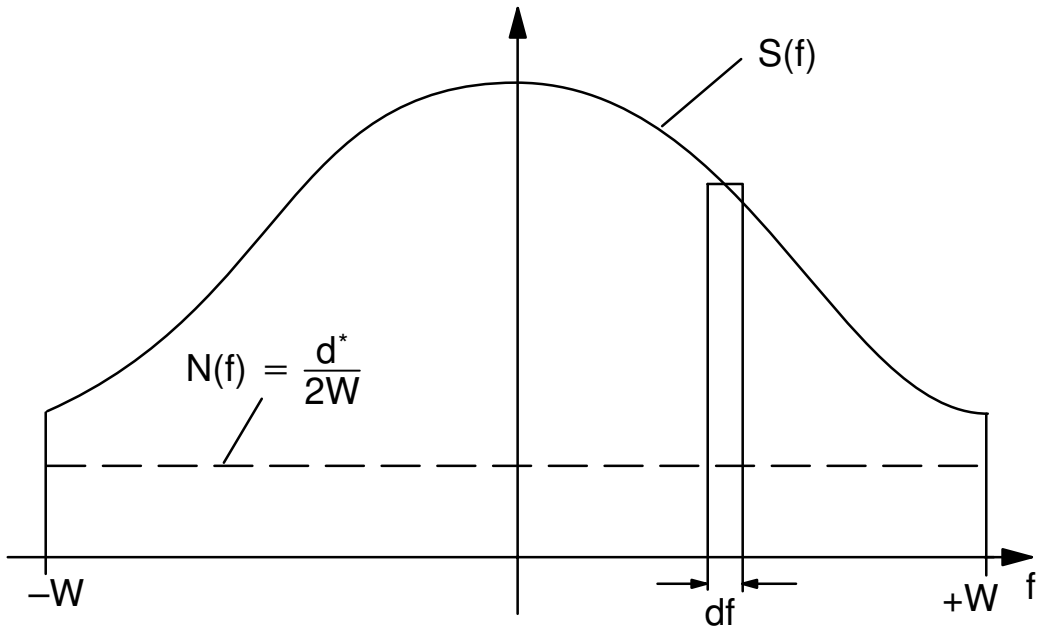


Bild 2.28 Zerlegung einer gedächtnisbehafteten Gaußschen Quelle in eine Summe gedächtnisloser Gaußscher Quellen

Ist $N(f)\,df$ die Verzerrung der einzelnen gedächtnislosen Gaußschen Quelle, so kann man mit Hilfe von Gl. (2.145) unter der Voraussetzung $S(f) > N(f)$ den Mittelwert der Übertragungsrate für die gedächtnislosen Gaußschen Quellen ermitteln.

$$R(d^*) = \frac{1}{2W} \int_{-W}^{+W} \frac{1}{2} \log_2 \frac{S(f)\,df}{N(f)\,df}\,df$$

$$R(d^*) = \frac{1}{2W} \int_{-W}^{+W} \frac{1}{2} \log_2 \frac{S(f)}{N(f)}\,df \tag{2.149}$$

in bit je Quellensymbol bzw. Abtastwert. Zur Einhaltung des Gütekriteriums muß dabei

$$\int_{-W}^{+W} N(f)\,df = d^* \tag{2.150}$$

sein. Ferner muß die Verzerrung $N(f)$ eine konstante Leistungsdichte aufweisen, d.h. es muß jede einzelne Quelle gleich viel Verzerrung zur Gesamtverzerrung beitragen, damit bei vorgegebener Verzerrung d^* die Übertragungsrate $R(d^*)$ das Minimum entsprechend Gl. (2.130) erreicht.

Beweis:

Die Minimierung von $R(d^*)$ als Funktion von $N(f)$ unter der Nebenbedingung (2.150) führt auf eine Variationsrechnung und Minimierung von

$$\int_{-W}^{+W} \left[\log_2 \frac{S(f)}{N(f)} + \lambda \cdot N(f) \right] df \,. \tag{2.151}$$

Die Bedingung dafür ist

$$-\frac{1}{N(f)} \log_2 e + \lambda = 0 \,,$$

d.h. $N(f)$ ist eine Konstante gemäß

$$d^* = 2W \cdot N(f) = 2W \cdot \Theta \,. - \tag{2.152}$$

Damit erhält man für $R(d^*)$

$$R(d^*) = \frac{1}{2W} \int_{-W}^{+W} \frac{1}{2} \log_2 \frac{S(f) \cdot 2W}{d^*} \, df \,. \tag{2.153}$$

Die Rate–Distortion–Funktion gemäß Gl. (2.153) stellt für nicht Gaußsche, gedächtnisbehaftete Quellen mit gleicher spektraler Leistungsdichte eine obere Grenze für $R(d^*)$ dar.

Beispiel: Berechnung der oberen Grenze der Rate–Distortion–Funktion für Fernsprech– und Fernsehsignale. Die Leistungsdichtespektren von Fernsprech– und Fernsehluminanzsignalen können durch die Näherung

$$S(f) = A \left[1 + \left(\frac{f}{f_0} \right)^n \right]^{-1} \tag{2.154}$$

approximiert werden, vergl. dazu Bild 2.29, wobei im Fall von Fernsprechsignalen die Bandbreite $W = 3.4\,\text{kHz}$ mit den Parametern $f_0 = 0,5\,\text{kHz}$, $n = 4$

und im Fall von Fernsehsignalen die Bandbreite $W = 5$ MHz und die Parameter $f_0 = 0,015$ MHz, $n = 2$, angenommen werden.

Die Auswertung der Gleichungen (2.153) und (2.152) liefert mit

$$\int_{-W}^{+W} S(f)\, df = {\sigma_x}^2 \tag{2.155}$$

für Fernsprechsignale

$$R_S(d^*) = \frac{1}{2}\log_2 \frac{{\sigma_x}^2}{d^*} - 1,8 \text{ bit} \tag{2.156}$$

und für Fernsehsignale

$$R_{TV}(d^*) = \frac{1}{2}\log_2 \frac{{\sigma_x}^2}{d^*} - 3,0 \text{ bit}\,. \tag{2.157}$$

Diese Rate–Distortion–Funktionen sind in Bild 2.30 im Vergleich zur Grenze gedächtnisloser Quellen dargestellt. Dabei ist das Signal–Rauschleistungs–Verhältnis unter Verwendung von Gl. (2.146) in die Einheit "Dezibel" umgerechnet worden.–

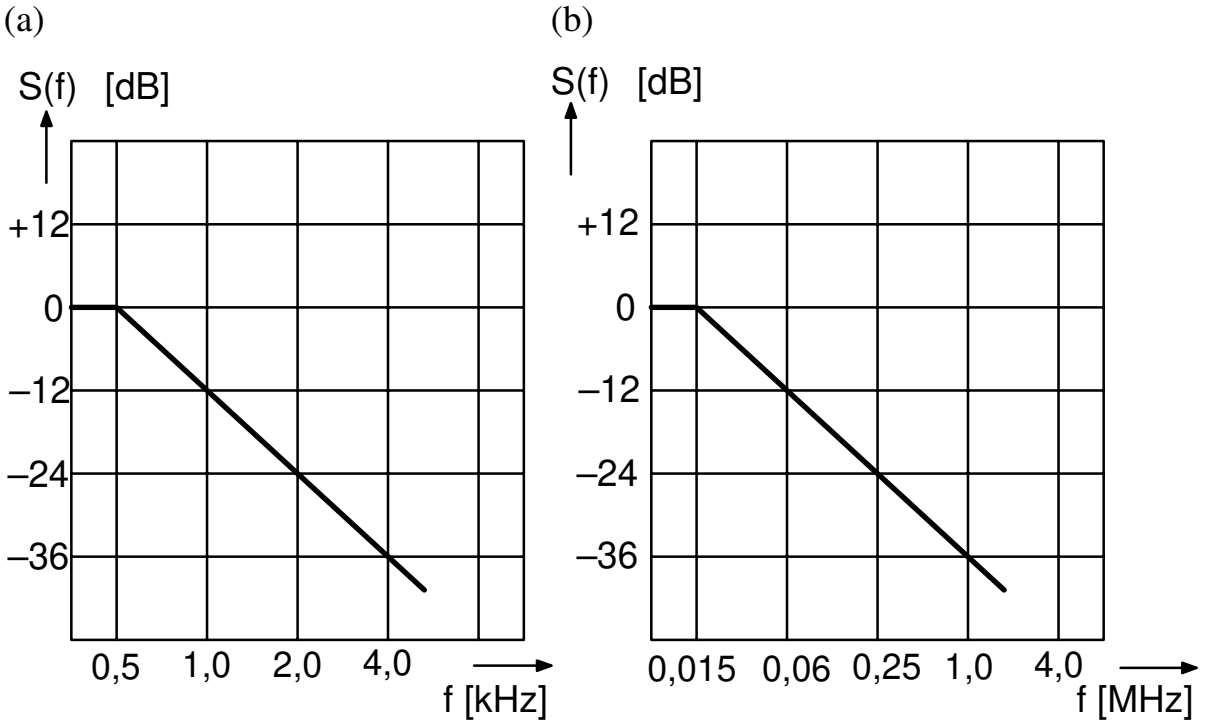


Bild 2.29 Leistungsdichtespektrum von Fernsprechsignalen (a) und Fernsehluminanzsignalen (b)

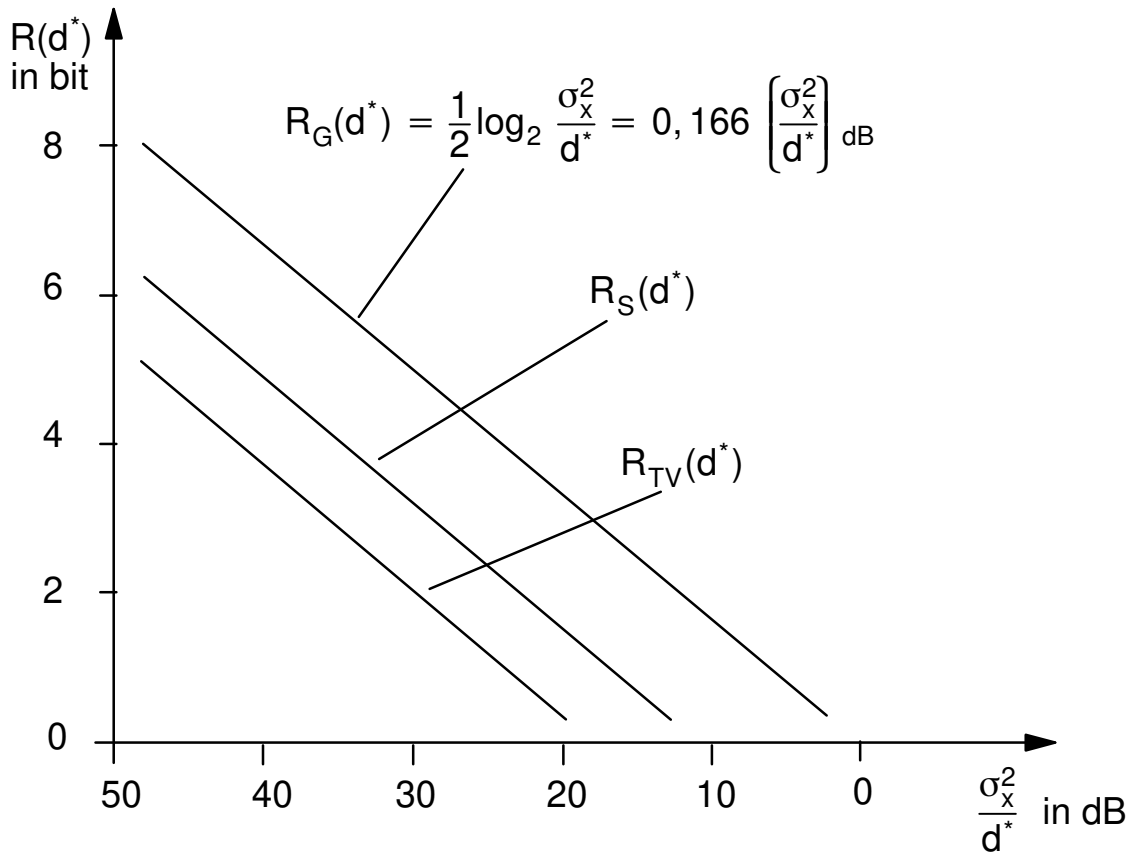


Bild 2.30 Obere Grenze der Übertragungsraten für eine gedächtnislose Quelle $R(d^*)$ sowie für Fernsprech– und Fernsehsignale $R_S(d^*)$, $R_{TV}(d^*)$

Im Vergleich zur gedächtnislosen Quelle ergibt sich für gedächtnisbehaftete Quellen eine geringere obere Grenze, die vom Verlauf der spektralen Leistungsdichte abhängt. Diese Verminderung der Übertragungsrate ist auf im Signal enthaltene Korrelation zurückzuführen, denn ein nicht weißes Leistungsdichtespektrum bedeutet gemäß dem Wiener–Khintchineschen Theorem, daß die Abtastwerte korreliert sind. Mit zunehmend weißer spektraler Leistungsdichte vermindert sich die Korrelation und $R(d^*)$ nähert sich der Grenze der Übertragungsrate für gedächtnislose Quellen. Es gilt

$$R(d^*) = \frac{1}{2W} \int_{-W}^{+W} \frac{1}{2}\log_2 \frac{S(f) \cdot 2W}{d^*}\, df = \frac{1}{2}\log_2 \frac{S(f) \cdot 2W}{d^*} = \frac{1}{2}\log_2 \frac{{\sigma_x}^2}{d^*}, \qquad (2.158)$$

da bei konstanter Leistungsdichte

$$S(f) \cdot 2W = {\sigma_x}^2 \qquad \text{ist.}$$

Die Verzerrung d^* wird im Codierungssystem durch den Quantisierungsfehler q bzw. dessen Varianz σ_q^2 bestimmt. Dabei gilt für eine gleichförmige Quantisierung mit der Stufenbreite Δx näherungsweise

$$d^* = 2W \cdot \theta = \sigma_q^2 \approx \frac{\Delta x^2}{12} \tag{2.159}$$

Bei Quantisierung mit geringer Stufenzahl K, d.h. mit großer Stufenbreite Δx kann θ größer als $S(f)$ werden (vergleiche Bild 2.31). In diesen Frequenzbereichen wird das Ausgangssignal $v = 0$ gesetzt. Auf diese Weise entsteht eine Verzerrung, die der spektralen Leistungsdichte $S(f)$ entspricht. Diese Frequenzbereiche tragen nicht zur Übertragungsrate bei, d.h.

$$d^* = \int\limits_{f:\, S(f) > \theta} \theta \, df + \int\limits_{f:\, S(f) \le \theta} S(f) \, df \,, \tag{2.160}$$

$$R(d^*) = \frac{1}{2W} \int\limits_{f:S(f) > \theta} \frac{1}{2} \log_2 \frac{S(f)}{\theta} \, df \tag{2.161}$$

in bit je Abtastwert.

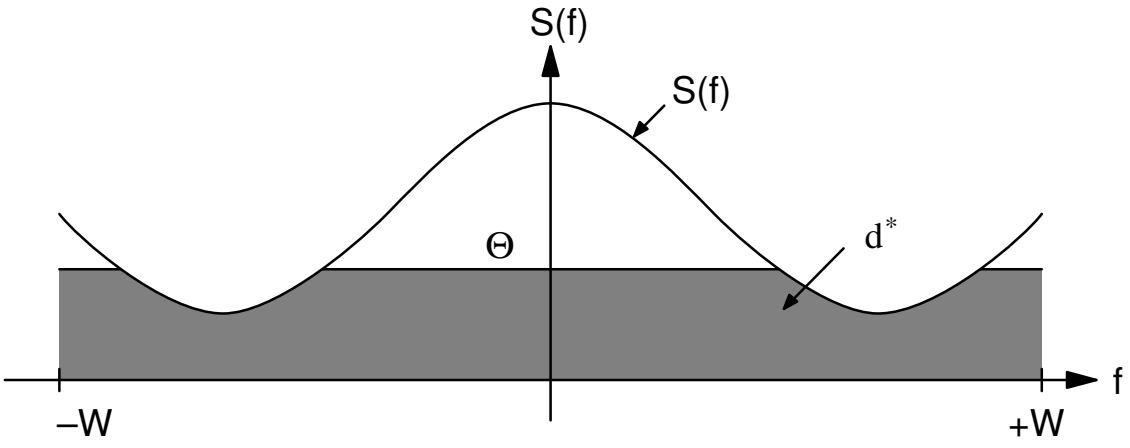


Bild 2.31 Grafische Darstellung der Integrale d^* und $R(d^*)$

2.5 Anwendungsbeispiele

2.5.1 Amplitudenquantisierung mit minimaler Fehlervarianz

2.5.1.1 Quantisierung mit großer Stufenzahl K

Eine zeitdiskrete, wertkontinuierliche Quelle X soll durch Quantisierung der Quellensymbole x in eine diskrete Quelle umgewandelt werden. Die Quellensymbole mögen einen begrenzten kontinuierlichen Wertevorrat $-1 \leq x \leq +1$ und eine Wahrscheinlichkeitsdichte $p(x)$ aufweisen. Dieser kontinuierliche Wertevorrat soll in K Quantisierungsstufen der Breite Δ_k mit einem jeweiligen Repräsentativwert x_k quantisiert werden. Es sei angenommen, daß die Quantisierung so fein ist, daß $p(x)$ als näherungsweise konstant im Intervall Δ_k angesehen werden kann, und daß der Repräsentativwert x_k jeweils in der Mitte des Intervalls Δ_k liegt. Die Quantisierung soll nun so gewählt werden, daß der mittlere quadratische Quantisierungsfehler $E[q^2]$ als Gütekriterium ein Minimum erreicht.

Der mittlere quadratische Quantisierungsfehler der k–ten Quantisierungsstufe beträgt

$$E_k[q^2] = \int_{x_k-\frac{1}{2}\Delta_k}^{x_k+\frac{1}{2}\Delta_k} (x - x_k)^2 \, p(x) \, dx \ , \tag{2.162}$$

wenn jeder in das Intervall Δ_k fallende Wert durch x_k dargestellt wird und somit $q = x - x_k$ der Quantisierungsfehler ist.

Mit der Annahme

$$p(x) \simeq p(x_k) \quad \text{für} \quad x_k - \frac{1}{2} \cdot \Delta_k \leq x \leq x_k + \frac{1}{2} \cdot \Delta_k \tag{2.163}$$

wird $p(x)$ durch eine Treppenkurve dargestellt und es folgt

$$E_k[q^2] = \int_{x_k-\frac{1}{2}\Delta_k}^{x_k+\frac{1}{2}\Delta_k} (x - x_k)^2 \, dx = \frac{1}{12} \, p(x_k) \, \Delta_k{}^3 \ . \tag{2.164}$$

Die Wahrscheinlichkeit, daß ein Wert x in das Intervall Δ_k fällt, ist

$$P_k = p(x_k) \cdot \Delta_k \tag{2.165}$$

und für (2.164) folgt

$$E_k[q^2] = \frac{1}{12} P_k \Delta_k^{\,2} , \tag{2.166}$$

sowie für den gesamten mittleren quadratischen Quantisierungsfehler $E[q^2]$

$$E[q^2] = \sum_{k=1}^{K} E_k[q^2] = \frac{1}{12} \sum_{k=1}^{K} P_k \Delta_k^{\,2} . \tag{2.167}$$

Zur Minimierung von $E[q^2]$ muß im allgemeinen eine Quantisierung mit ungleich breiten Quantisierungsstufen angewendet werden. Eine solche ungleichförmige Quantisierung wird zur Veranschaulichung durch eine sog. Kompressorkennlinie dargestellt.

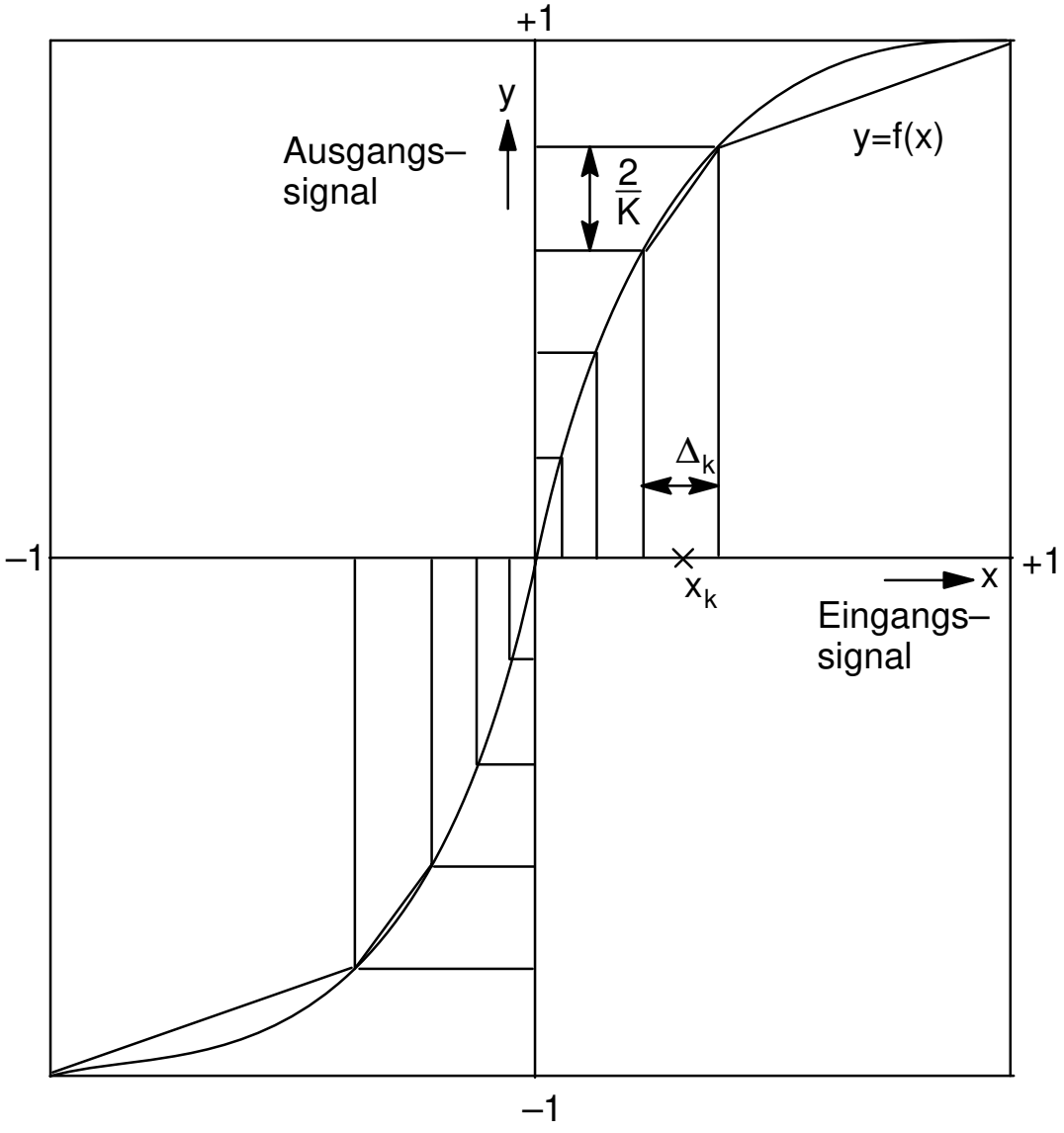


Bild 2.32 Beispiel einer Kompressorkennlinie

Im Diagramm des Bildes 2.32 wird die ungleichförmige Quantisierung der Eingangswerte x über die Kompressorkennlinie $y = f(x)$ auf eine gleichförmige Quantisierung der Ausgangswerte y abgebildet. Die ungleichförmige Quantisierung kann demgemäß gedanklich mit Hilfe eines Kompressors der Übertragungsfunktion $y = f(x)$ und einem gleichförmigen Quantisierer realisiert werden. Zur Rückwandlung wird dann ein Expander mit einer zur Kompression inversen Kennlinie erforderlich. Das Blockschaltbild eines solchen Systems ist in Bild 2.33 dargestellt.

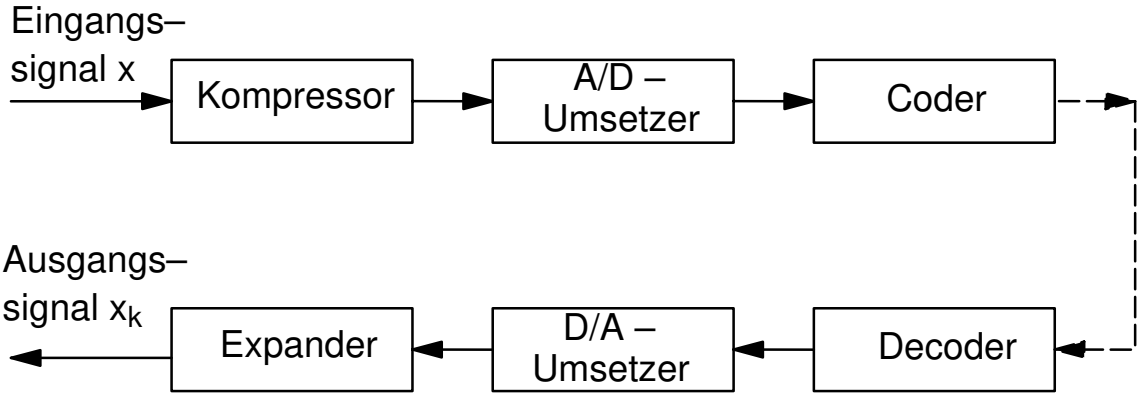


Bild 2.33 Übertragungssystem mit Kompandierung

Ersetzt man die Kompressorkennlinie durch eine Segmentkennlinie der Sehnen, wie in Bild 2.32 gezeigt, so gilt

$$\left(\frac{dx}{dy}\right)_{x=x_k} = \frac{\Delta_k}{2} K . \tag{2.168}$$

Damit erhält man für $E[q^2]$ die Form

$$E[q^2] = \frac{1}{12}\sum_{k=1}^{K} P_k \Delta_k{}^2 = \frac{1}{3K^2}\sum_{k=1}^{K} P_k \left(\frac{dx}{dy}\right)^2_{x=x_k} . \tag{2.169}$$

Bei hinreichend feiner Quantisierung kann die Summe durch das Integral

$$E[q^2] = \frac{1}{3K^2}\int_{-1}^{+1} p(x)\left(\frac{dx}{dy}\right)^2 dx \tag{2.170}$$

dargestellt werden.

Die folgende Minimierung des mittleren quadratischen Quantisierungsfehlers $E[q^2]$ ist ein Problem der Variationsrechnung. Gesucht wird die Extremale $y' = dy/dx$ mit der Nebenbedingung

$$\int_{-1}^{+1} y' \, dx = 2 \, . \tag{2.171}$$

Die Variationsrechnung führt auf die Beziehung

$$\frac{d}{dy'}\left[p(x) \cdot y'^{-2} + \lambda \cdot y'\right] = -2 \cdot p(x) \cdot y'^{-3} + \lambda = 0$$

$$y'_{opt} = \left(\frac{2}{\lambda}\right)^{1/3} p(x)^{1/3} \quad , \tag{2.172}$$

wobei der Lagrangesche Multiplikator λ mit Hilfe der Nebenbedingung (2.171) wie folgt bestimmt werden kann:

$$\left(\frac{2}{\lambda}\right)^{1/3} = \frac{2}{\int_{-1}^{+1} p(x)^{1/3} \, dx} = \frac{1}{C} \, . \tag{2.173}$$

Dabei ist C eine Konstante, die durch die Wahrscheinlichkeitsdichte der Quellensymbole bestimmt ist. Die gesuchte Kompressorkennlinie ergibt sich mit (2.172) und (2.173) zu

$$y'_{opt} = \frac{1}{C} \cdot p(x)^{1/3} \tag{2.174}$$

$$y_{opt} = \frac{1}{C} \cdot \int_{0}^{x} p(x)^{1/3} \, dx \, . \tag{2.175}$$

Der minimale mittlere quadratische Quantisierungsfehler beträgt mit (2.170) und (2.174)

$$E[q^2]_{\min} = \frac{1}{3K^2} C^2 \int_{-1}^{+1} p(x)^{1/3} \, dx = \frac{2}{3} \, \frac{C^3}{K^2} \, . \tag{2.176}$$

Bei optimaler Kompressorkennlinie zeigt sich, daß der mittlere quadratische Quantisierungsfehler der einzelnen Quantisierungsstufen (2.164) gleich groß ist.

$$E_k[q^2] = \frac{1}{12} p(x_k) \, \Delta_k^{\,3} = \frac{1}{12} p(x_k) \, \frac{8}{K^3} \left(\frac{dx}{dy} \right)^3_{x = x_k}$$

$$E_k[q^2] = \frac{2}{3} \frac{C^3}{K^3} \tag{2.177}$$

Im Sonderfall konstanter Wahrscheinlichkeitsdichte $p(x)$ folgt aus Gl. (2.175), daß die Kompressorkennlinie eine Gerade ist. Somit ergibt sich eine gleichförmige Quantisierung der Eingangswerte x. Alle Quantisierungsstufen sind gleich groß, $\Delta_k = \Delta = 2/K$. Aus der Beziehung (2.167) folgt

$$E[q^2] = \frac{1}{12} \, \Delta_k^{\,2} \cdot \sum_{k=1}^{K} P_k = \frac{1}{12} \, \Delta_k^{\,2} = \frac{1}{3K^2} \; . \tag{2.178}$$

Diese gleichförmige Quantisierung hat die besondere Eigenschaft, daß der mittlere quadratische Quantisierungsfehler auch bei nicht konstanter Wahrscheinlichkeitsdichte $p(x)$ des Eingangssignals stets den gleichen Wert liefert.

Die hier durchgeführte Berechnung des minimalen Quantisierungsfehlers geht auf Panter und Dite [13] zurück und beruht auf den eingangs erwähnten einschränkenden Annahmen, daß $p(x) \simeq p(x_k)$ im Intervall Δ_k und daß x_k in der Mitte von Δ_k liegt.

2.5.1.2 Quantisierung mit beliebiger Stufenzahl K

Die Berechnung eines Quantisierers mit minimaler, mittlerer Verzerrung und beliebiger Stufenzahl K geht auf J. Max [12] zurück. Im Gegensatz zu dem Verfahren von Panter und Dite [13] sind keine einschränkenden Annahmen erforderlich.

Zur Bestimmung des Quantisierers werden folgende Annahmen getroffen. Der Wertebereich der Quellensymbole sei $-\infty \leq x_k \leq +\infty$ und die Wahrscheinlichkeitsdichte $p(x)$. Jeder Repräsentativwert x_k des K–stufigen Quantisierers liege zwischen zwei Entscheidungswerten gemäß $d_k \leq x_k \leq d_{k+1}$. Die Entscheidungswerte $d_1 = -\infty$ und $d_K = +\infty$ liegen an den Grenzen des Wertebereichs, s. Bild 2.34. Unter diesen Annahmen ergibt sich eine mittlere Verzerrung

$$D = \sum_{k=1}^{K} \int_{d_k}^{d_{k+1}} f(x - x_k) \; p(x) \; dx \tag{2.179}$$

wobei $f(x - x_k)$ das Verzerrungsmaß ist.

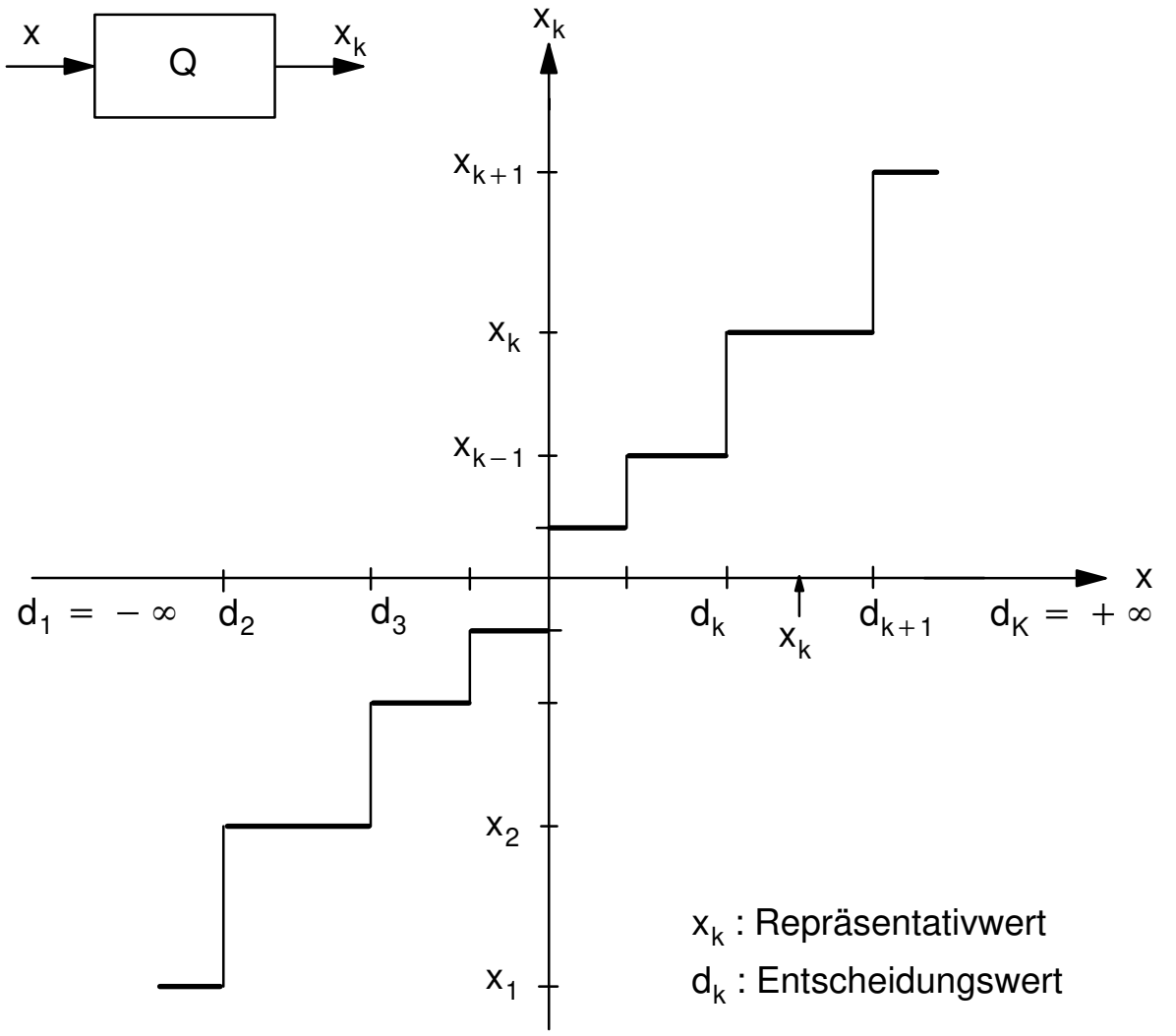


Bild 2.34 Charakteristik eines Quantisierers

Der gesuchte Quantisierer ist durch die Entscheidungs– und Repräsentativwerte $(d_{K,}\ x_k)$ definiert. Die Berechnung des gesuchten Quantisierers mit minimaler Verzerrung D führt auf ein Extremwertproblem. Zur Bestimmung der gesuchten Größen $d_k,\ x_k$ wird die Verzerrung D partiell nach den Unbekannten abgeleitet und die Ableitungen Null gesetzt, d.h.

$$\frac{\partial D}{\partial d_k} = f(d_k - x_{k-1})\ p\ (d_k) - f(d_k - x_k)\ p\ (d_k) = 0 \tag{2.180}$$

für $k = 2,\ ...,\ K$

und

$$\frac{\partial D}{\partial x_k} = -\int_{d_k}^{d_{k+1}} f'(x - x_k)\; p\,(x)\, dx = 0 \tag{2.181}$$

für $k = 1, \ldots, K$.

Zur weiteren Auswertung sei alss Verzerrungsmaß der mittlere quadratische Quantisierungsfehler angenommen, d.h.

$$f(x - x_k) = (x - x_k)^2 \tag{2.182}$$

Dann folgt aus Gl. (2.180)

$$(d_k - x_{k-1})^2 = (d_k - x_k)^2$$

und mit $d_k < x_k$

$$d_k - x_{k-1} = -(d_k - x_k)$$

bzw.

$$x_k = 2d_k - x_{k-1} \tag{2.183}$$

für $k = 2, \ldots, K$

sowie aus Gl. (2.181)

$$\int_{d_k}^{d_{k+1}} (x - x_k)\, p(x)\, dx = 0$$

$$x_k = \frac{\displaystyle\int_{d_k}^{d_{k+1}} x\, p(x)\, dx}{\displaystyle\int_{d_k}^{d_{k+1}} p(x)\, dx} \tag{2.184}$$

für $k = 1, \ldots, K$.

Dieses Gleichungssystem ist in der Regel nicht geschlossen lösbar. Nachfolgend werden drei mögliche Lösungsansätze kurz erläutert.

1. Lösungsansatz: Iterative Lösung

1. Wähle x_1 und berechne d_2 mit Gl. (2.184)

$$\int_{d_1=-\infty}^{d_2} (x - x_1) \cdot p(x)\, dx = 0\,, \tag{2.185}$$

   so daß x_1 Schwerpunkt der Fläche von p(x) zwischen d_1 und d_2 ist.
2. Unter Verwendung von x_1 und d_2 berechne x_2 mit Gl. (2.183).

$$x_2 = 2d_2 - x_1\,, \tag{2.186}$$

   usw.
3. Überprüfe, ob x_K Schwerpunkt zwischen d_K und $+\infty$ ist, andernfalls muß mit einem anderen x_1 die Iteration wiederholt werden.

2. Lösungsansatz: LBG–Algorithmus

Der LBG–Algorithmus ist ein Verfahren zum Entwurf von Vektorquantisierern, das auf Linde, Buzo und Gray [11] zurückgeht und im Kapitel 2.5.4 ausführlich dargestellt wird. Es ermöglicht auch den Entwurf von skalaren Quantisierern, wie sie hier betrachtet werden.

3. Lösungsansatz: Geschlossene Lösung

Für Quantisierer mit sehr kleiner Stufenzahl K kann das Gleichungssystem (2.183), (2.184) zum Teil auch geschlossen gelöst werden. Vergleiche dazu folgendes Beispiel:

Beispiel:

Quantisierung mit $K = 2$ Stufen.
Die Quelle habe die normierte Gaußsche Wahrscheinlichkeitsdichte

$$p(x) = p_G(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}\,, \quad \sigma_x^2 = 1\,, \quad m_x = 0\,.$$

Aufgrund der symmetrischen Wahrscheinlichkeitsdichte wird
$d_1 = -\infty\,,\ d_2 = 0\,,\ d_3 = +\infty$
$x_1 = -x_2$, siehe Bild 2.35
Mit Gl. (2.184) kann x_1 gemäß

$$x_1 = \frac{\int_{-\infty}^{0} x\, p(x)\, dx}{\int_{-\infty}^{0} p(x)\, dx} = 2 \int_{-\infty}^{0} x\, p(x)\, dx$$

berechnet werden. Daraus folgt

$$x_1 = -0,798 \;,\; D = 0,363$$

$$SNR = 10 \log_{10} \frac{\sigma_x^2}{D} = 4,40 \text{ dB} \;.–$$

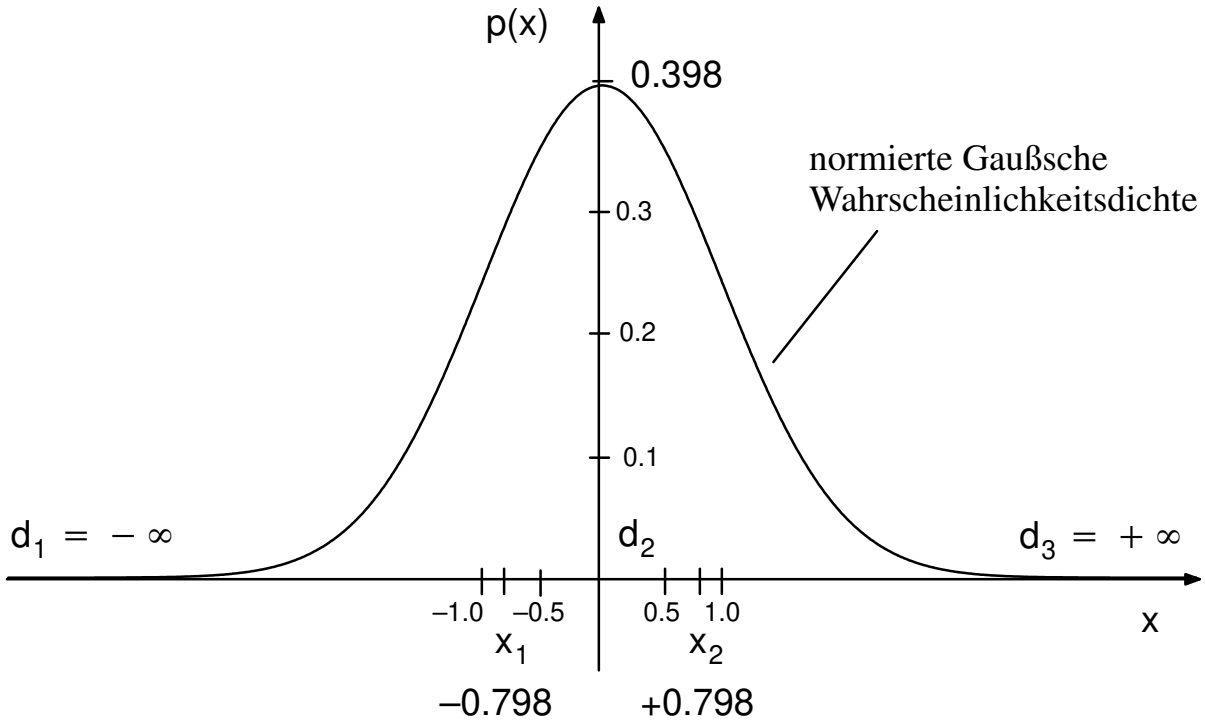


Bild 2.35 Beispiel einer Quantisierung mit $K = 2$ Quantisierungsstufen

2.5.2 Quantisierung mit minimaler Ausgangsentropie

Die Entropie $H_Q(X)$ der Repräsentativwerte x_k eines beliebigen Quantisierers beträgt

$$\begin{aligned} H_Q(X) &= -\sum_{k=1}^{K} p(x_k)\,\Delta_k\,\log_2\bigl(p(x_k)\,\Delta_k\bigr) \\ &= -\int_{-\infty}^{+\infty} p(x)\,\log_2\,p(x)\,dx - \sum_{k=1}^{K} p(x_k)\,\Delta_k\cdot\log_2\,\Delta_k \\ &= H(X) - \sum_{k=1}^{K} p(x_k)\,\Delta_k\cdot\log_2\,\Delta_k\,, \end{aligned} \tag{2.187}$$

wobei Δ_k die Stufenbreite der k–ten Quantisiererstufe ist. Der mittlere quadratische Quantisierungsfehler q dieses Quantisierers ergibt sich bei einer Quantisierung mit hinreichend großer Stufenzahl K nach Gl. (2.167) zu

$$E_Q[q^2] = \sum_{k=1}^{K} p(x_k)\cdot\Delta_k\cdot\frac{\Delta_k^2}{12}\;. \tag{2.188}$$

Zum Vergleich wird ein Quantisierer mit gleichförmiger Quantisierung betrachtet. Mit

$$\Delta_k = \Delta \qquad \text{für alle } k \tag{2.189}$$

erhält man für die gleichförmige Quantisierung

$$H_{GQ}(X) = H(X) - \log_2\Delta \tag{2.190}$$

$$E_{GQ}[q^2] = \frac{\Delta^2}{12}\;. \tag{2.191}$$

Es soll nun gezeigt werden, daß der gleichförmige Quantisierer bei gleichem mittleren Quantisierungsfehler die geringere Entropie hat. Dazu wird die Differenz der Entropie berechnet

$$H_{GQ}(X) - H_Q(X) = \sum_{k=1}^{K} p(x_k)\Delta_k \cdot \log_2 \frac{\Delta_k}{\Delta}$$
$$= \frac{1}{2}\sum_{k=1}^{K} p(x_k)\Delta_k \log_2\left(\frac{\Delta_k}{\Delta}\right)^2 . \tag{2.192}$$

Mit der Hilfsbeziehung $\ln z \leq z-1$ für $z > 0$ und $\ln z = z-1$ für $z = 1$ sowie $\log_2 z = \ln z \cdot \log_2 e$ wird

$$H_{GQ}(X) - H_Q(X) \leq \frac{1}{2}\log_2 e \sum_{k=1}^{K} p(x_k)\,\Delta_k\left[\left(\frac{\Delta_k}{\Delta}\right)^2 - 1\right]$$
$$\leq \frac{1}{2}\log_2 e\left[\frac{E_Q[q^2]}{E_{GQ}[q^2]} - 1\right] . \tag{2.193}$$

Daraus folgt für

$$E_Q[q^2] \leq E_{GQ}[q^2] , \tag{2.194}$$

daß

$$H_{GQ}(X) - H_Q(X) \leq 0 . \tag{2.195}$$

Demzufolge gibt es unter den Quantisierern mit einem mittleren quadratischen Quantisierungsfehler, der kleiner oder gleich dem des gleichförmigen Quantisierers ist, keinen, dessen Entropie kleiner als die des gleichförmigen Quantisierers ist.

Bild 2.36 zeigt die Übertragungsbitrate als Funktion des mittleren Quantisierungsfehlers eines nach J. Max berechneten Quantisierers im Vergleich zu der eines gleichförmigen Quantisierers und zur Rate–Distortion–Funktion der gedächtnislosen Gaußschen Quelle.

Die mit gleichförmiger Quantisierung und anschließender redundanzmindernder oder auch sog. Entropie–Codierung erreichbaren Übertragungsbitraten müssen um 0,25 bit über der Rate–Distortion–Funktion liegen, da der Quantisierungsfehler näherungsweise eine konstante Wahrscheinlichkeitsdichte anstelle einer Gaußschen Wahrscheinlichkeitsdichte aufweist.

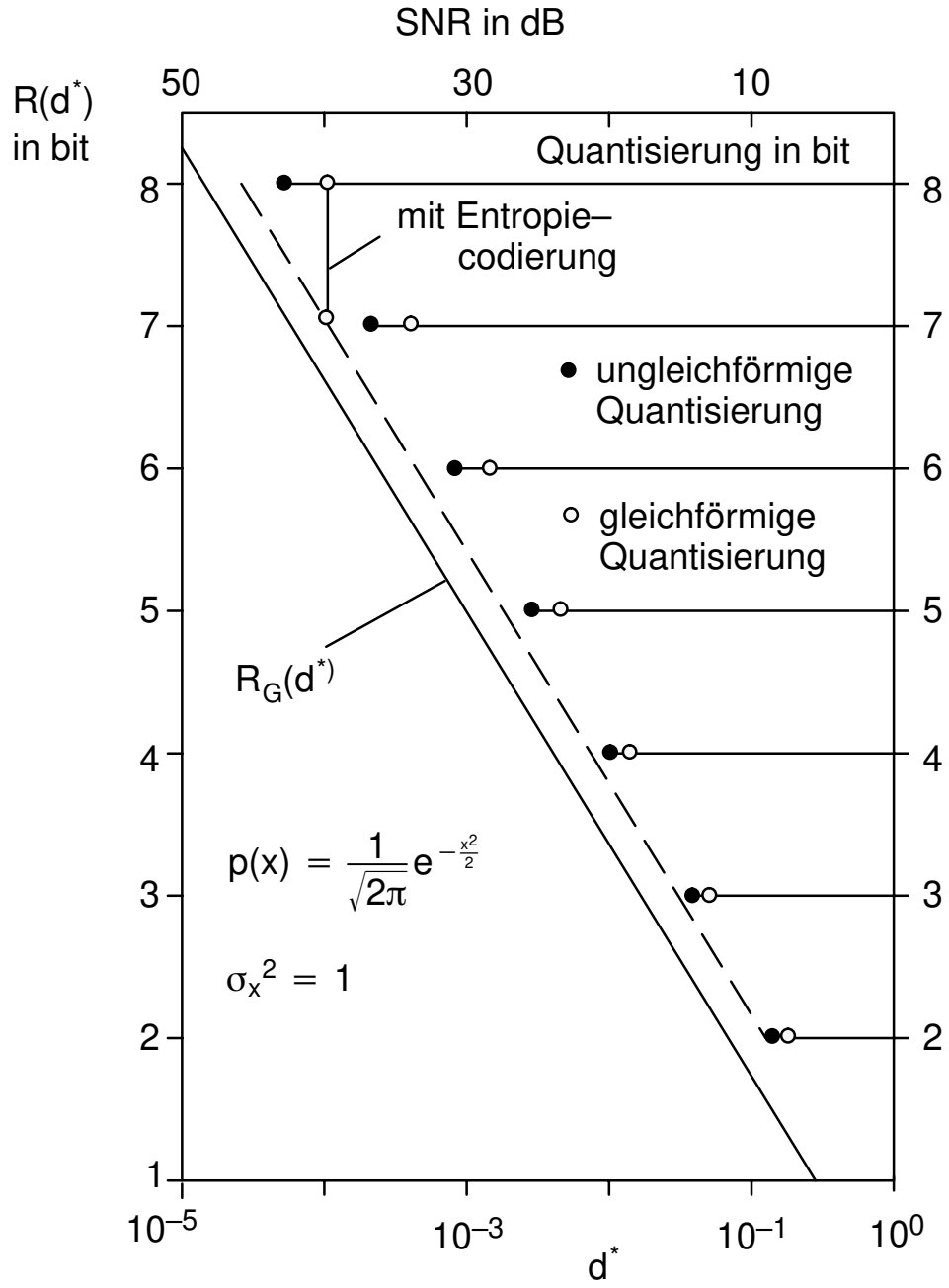


Bild 2.36 Übertragungsbitrate als Funktion des mittleren Quantisierungsfehlers d^* verschiedener Quantisierer im Vergleich zur Rate–Distortion–Funktion (RDF) der gedächtnislosen Gaußschen Quelle

2.5.3 Logarithmische Quantisierung für Sprachsignale

Bei der Fernsprechübertragung wird ein Signal–Rauschleistungs–Verhältnis angestrebt, das möglichst unabhängig ist von der Spracheigenschaft des Sprechers und der Zuleitungsdämpfung, d.h. möglichst unabhängig ist von der Wahrscheinlichkeitsdichte $p(x)$ der Abtastwerte und ihrer Leistung ${\sigma_x}^2$. Das Signal–Quantisierungsrauschleistungs–Verhältnis beträgt mit Gl. (2.170)

$$\frac{P}{E[q^2]} = 3K^2 \frac{\int p(x)\, x^2\, dx}{\int p(x) \left(\frac{dx}{dy}\right)^2 dx} \quad . \tag{2.196}$$

Es ist leicht einzusehen, daß jede Kompressorkennlinie $y = f(x)$, die der Bedingung

$$\frac{dx}{dy} = B \cdot x\,, \qquad \text{B = Konstante} \tag{2.197}$$

genügt, zu einem Signal–Rauschleistungs–Verhältnis gemäß

$$\frac{P}{E} = \frac{3K^2}{B^2} \tag{2.198}$$

führt, das die geforderte Eigenschaft besitzt. Die zugehörige Kompressorkennlinie erhält man durch Lösung der Differentialgleichung (2.197)

$$y = 1 + \frac{1}{B} \ln x \;, \tag{2.199}$$

wobei die Integrationskonstante aus der Anfangsbedingung $y = 1$, wenn $x = 1$, bestimmt wird. Da der Ausdruck $\ln x$ in Gl. (2.199) für $x \to 0$ unbestimmt wird, verwendet man bei der PCM–Sprachcodierung eine Approximation der Lösung (2.197). Für positive x gilt

$$\begin{aligned} y &= \frac{1 + \ln(A \cdot x)}{1 + \ln(A)} \quad \text{für } 1 \geq x \geq \frac{1}{A} \\ y &= \frac{A \cdot x}{1 + \ln(A)} \qquad \text{für } \frac{1}{A} \geq x \\ A &= 87.56 \qquad\qquad\qquad\qquad , \end{aligned} \tag{2.200}$$

d.h. für kleine x wird y durch eine Gerade dargestellt. Diese sog. A–Kompressorkennlinie ist eine ungerade Funktion.

Aus Gründen der Schaltungsrealisierung wird sie nicht exakt, sondern in Form einer 13–Segment–Kennlinie realisiert. Die 13–Segment–Kennlinie berührt die A–Kennlinie an den Stellen $x = 1/2^n$ mit $n = 1, 2, 3, 4, 5, 6$, wie in Bild 2.37 dargestellt.

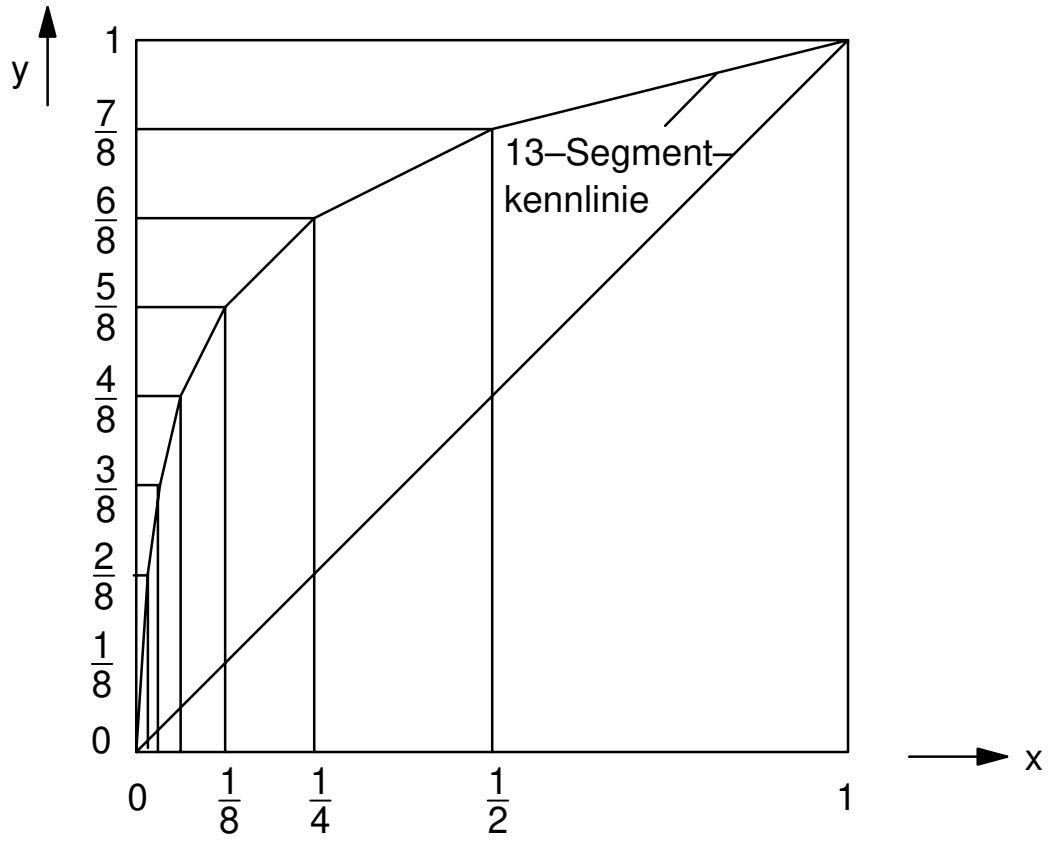


Bild 2.37 13–Segmentkennlinie der PCM–Sprachkompandierung

Das Signal–Rauschleistungs–Verhältnis für eine gleichförmige Quantisierung ergibt sich mit Gl. (2.178) zu

$$\left[\frac{P}{E}\right]_{\mathrm{dB}} = 10\log_{10}\frac{P_0 \cdot P}{E \cdot P_0} = 10\log_{10}\left(\frac{3}{2}\,K^2\,\frac{P}{P_0}\right), \tag{2.201}$$

wobei $P_0 = 1/2$ ein Bezugspegel der Eingangsleistung P ist, der einer Vollaussteuerung der Quantisierungskennlinie mit einer Sinusspannung der Amplitude 1 entspricht. Unter Verwendung der Beziehungen $\log_{10} x = \log_2 x \cdot \log_{10} 2$ und $\log_2 K = R$ wird

$$\left[\frac{P}{E}\right]_{\mathrm{dB}} = 6[R]_{\mathrm{bit}} + 1.8 + \left[\frac{P}{P_0}\right]_{\mathrm{dB}}. \tag{2.202}$$

Dieser Zusammenhang zwischen Signal–Rauschleistungs–Verhältnis und Eingangspegel einer gleichförmigen Quantisierung entspricht einer Geraden. Sie ist in Bild 2.38 für das Beispiel einer Quantisierung mit $R = 8$ bit entsprechend $K = 256$ Quantisierungsstufen dargestellt.

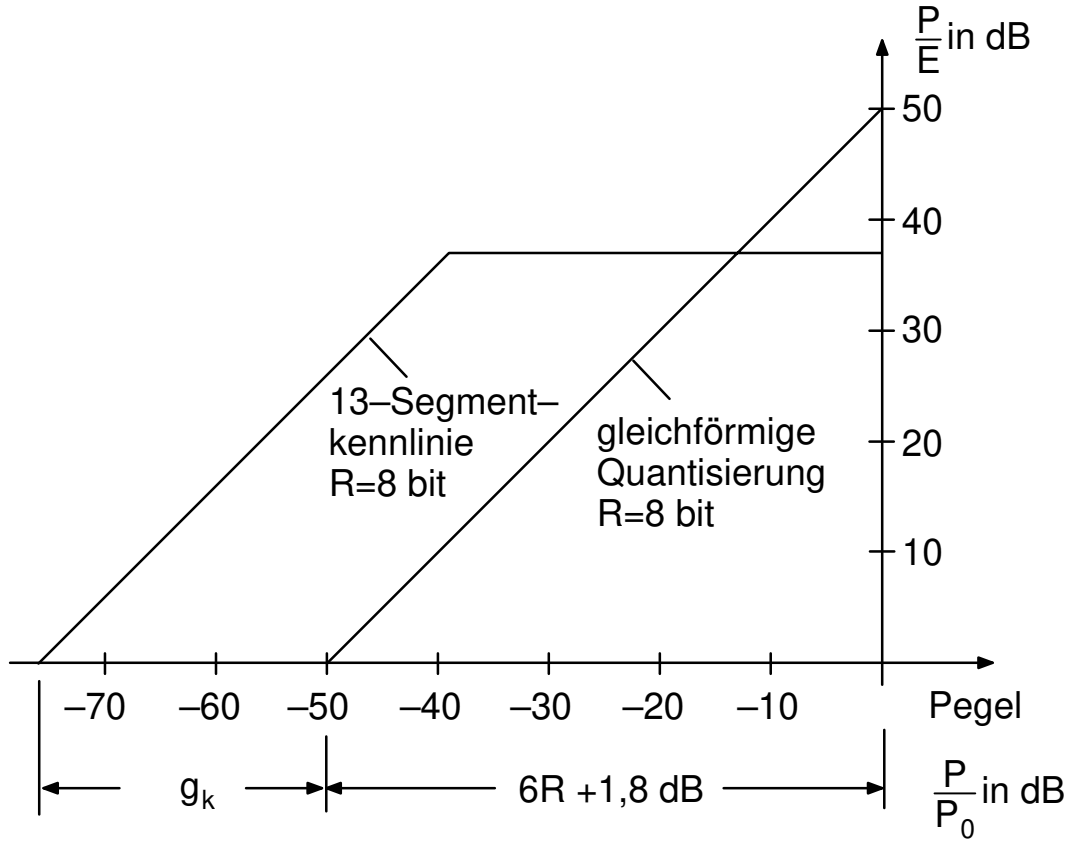


Bild 2.38 Signal–Rauschleistungs–Verhältnis *P/E* als Funktion der normierten Eingangsleistung P/P_0 für eine lineare und eine logarithmische Kompressorkennlinie

Das Signal–Rauschleistungs–Verhältnis der 13–Segment–Kennlinie wird für die beiden Bereiche der Kennlinie (2.200) getrennt ermittelt. Für $x \geq 1/A$ gilt mit (2.198)

$$\begin{aligned}\left[\frac{P}{E}\right]_{\mathrm{dB}} &= 10\log_{10}\frac{3K^2}{B^2} = 10\log_{10}\frac{3K^2}{[1+ln(A)]^2}\\ &= 10\log_{10}K^2 + 10\log_{10}\frac{3}{30} = 6\log_2 K - 10\\ &= 6[R]_{\mathrm{bit}} - 10\,.\end{aligned} \tag{2.203}$$

In diesem Bereich ist P/E unabhängig von der Eingangsleistung P. Für $x \leq 1/A$ ist die Kompressorkennlinie eine Gerade. Im Vergleich zur gleichförmigen Quantisierung in Bild 2.37 ist die Steilheit dieser Geraden jedoch um den Betrag $A/[1 + ln(A)]$ größer. Daraus resultieren kleinere Quantisierungsstufen und ein geringeres Quantisierungsrauschen. Die Verminderung des Quantisierungsrauschens wird als Kompandergewinn g_k

$$g_k = 20\log\frac{A}{1+ln(A)} = 24.1\ \mathrm{dB} \tag{2.204}$$

bezeichnet. Das Signal–Rauschleistungs–Verhältnis als Funktion der normierten Eingangsleistung P/P_0 ergibt sich somit für $x \leq 1/A$ aus dem der gleichförmigen Quantisierung zuzüglich Kompandergewinn (2.202)

$$\left[\frac{P}{E}\right]_{\mathrm{dB}} = 6[R] + 1.8 + \left[\frac{P}{P_0}\right]_{\mathrm{dB}} + g_k\,. \tag{2.205}$$

Der Gesamtverlauf des Signal–Rauschleistungs–Verhältnisses als Funktion von P/P_0 ist für die 13–Segment–Kennlinie ebenfalls in Bild 2.38 dargestellt. Man erkennt, daß das konstante Signal–Rauschleistungs–Verhältnis über einen Dynamikbereich der Eingangsleistung P/P_0 von 35 dB erzielt werden kann. Die Kompandierung bringt im Vergleich zur gleichförmigen Quantisierung einen Gewinn an P/E bei kleinem Eingangspegel P/P_0 und dafür einen Verlust an P/E bei großem Pegel.

2.5.4 Vektorquantisierung

Bei der Vektorquantisierung werden N zu quantisierende, wertkontinuierliche Abtastwerte x_n zu einem N–dimensionalen Vektor $\vec{x} = (x_1,\ x_2,\ ...,\ x_n,\ ...,\ x_N)^T$ zusammengefaßt und auf einen wertdiskreten Repräsentativvektor $\vec{y} = Q(\vec{x})$ abgebildet. Der Vektorquantisierer verfügt dazu über ein Codebuch C mit J Repräsentativvektoren $C = \left\{\vec{y}_1,\ \vec{y}_2,\ ...,\ \vec{y}_j,\ ...,\ \vec{y}_J\right\}$.
Die Abbildung $Q:\ \vec{x} \rightarrow \vec{y}_j$ erzeugt eine Partition des Vektorraumes $\vec{x}$ in disjunkte Quantisierungszellen S_j .Jeder Vektor $\vec{x}$ aus S_j wird auf den zugehörigen Repräsentativvektor $\vec{y}_j$ abgebildet.

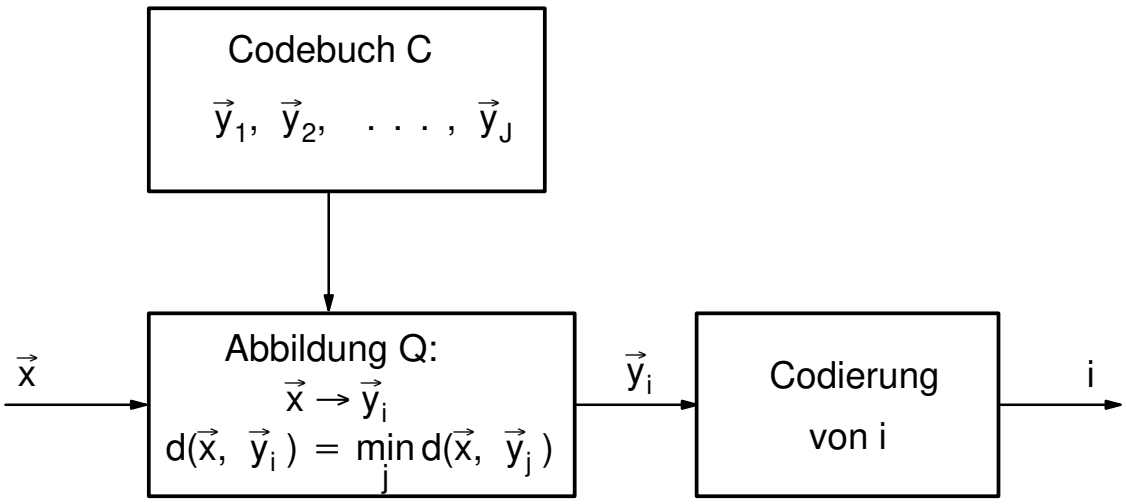


Bild 2.39 Struktur einer Vektorquantisierung mit Codierung

Betrachtet man die Quantisierung zusammen mit der Codierung des Repräsentativvektors, siehe Bild 2.39, so kann die Vektorquantisierung (VQ) und entsprechende inverse Vektorquantisierung (IVQ) als Verkettung folgender Abbildungen angesehen werden

$$\begin{aligned} \text{VQ} &: \vec{x} \rightarrow i \\ \text{IVQ} &: i \rightarrow \vec{y}_i\ . \end{aligned} \tag{2.206}$$

Bei binärer Codierung mit fester Codewortlänge $n = \log_2 J$ ergibt sich die Datenrate R je Abtastwert

$$R = \frac{1}{N}\log_2 J\ . \tag{2.207}$$

Die Vektorquantisierung ist durch das Codebuch C und die Partition S des Vektorraumes vollständig beschrieben. Bei vorgegebenem Verzerrungsmaß $d(\vec{x},\ \vec{y})$ und vorgegebenem Codebuch C minimieren die Quantisierungszellen

$$S_i^{opt} = \left\{ \vec{x} \,|\, d(\vec{x},\ \vec{y}_i) = \min_j d(\vec{x},\ \vec{y}_j) \right\} \tag{2.208}$$

der Partition den Quantisierungsfehler. Im Fall des mittleren quadratischen Fehlers als Verzerrungsmaß d

$$d = (\vec{x} - \vec{y}_i)^T \cdot (\vec{x} - \vec{y}_i)/N \tag{2.209}$$

führt die Zuordnung nach kürzestem Abstand auf eine sogenannte Voronoische Partition, siehe Bild 2.40, bei der die Grenzflächen der Partition senkrecht auf der Verbindungsgeraden zwischen benachbarten Repräsentativvektoren stehen.

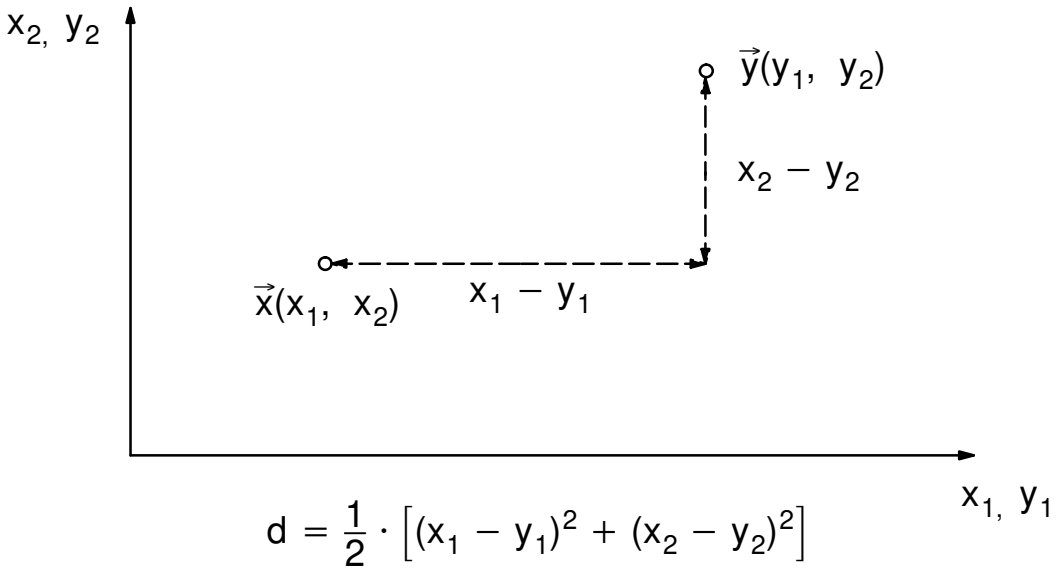


Bild 2.40 Darstellung des mittleren quadratischen Fehlers als geometrischer Abstand im Vektorraum

Die Wirkungsweise der Vektorquantisierung soll anhand von zwei in Bild 2.41 dargestellten Beispielen veranschaulicht werden.

Im Fall, daß die zu quantisierenden Abtastwerte statistisch unabhängig sind und eine normalverteilte Wahrscheinlichkeitsdichte $p(x)$ bzw. $p(\vec{x}) = p(x_1) \cdot p(x_2)$ entsprechend Bild 2.42 aufweisen, erzeugt die Partition der Vektorquantisierung gemäß Bild 2.41 (b) einen geringeren Quantisierungsfehler als die Partition einer gleichförmigen Quantisierung entsprechend Bild 2.41 (a), da die Quantisierungszellen im Bereich großer Wahrscheinlichkeitsdichte kleiner sind.

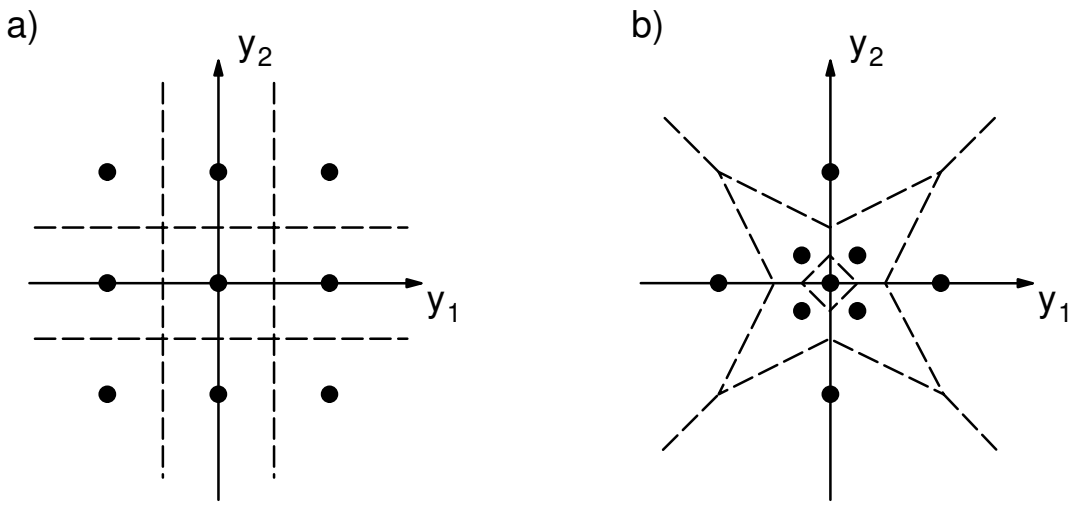


Bild 2.41 Beispiele der Partition eines gleichförmigen skalaren Quantisierers (a) und eines Vektorquantisierers (b) mit $N = 2$ und $J = 9$

Bei der skalaren Quantisierung entsprechend Partition (a) werden insbesondere die auf den Diagonalen liegenden Repräsentativvektoren schlecht genutzt, wie ein Vergleich von Bild 2.41 und Bild 2.42 zeigt.

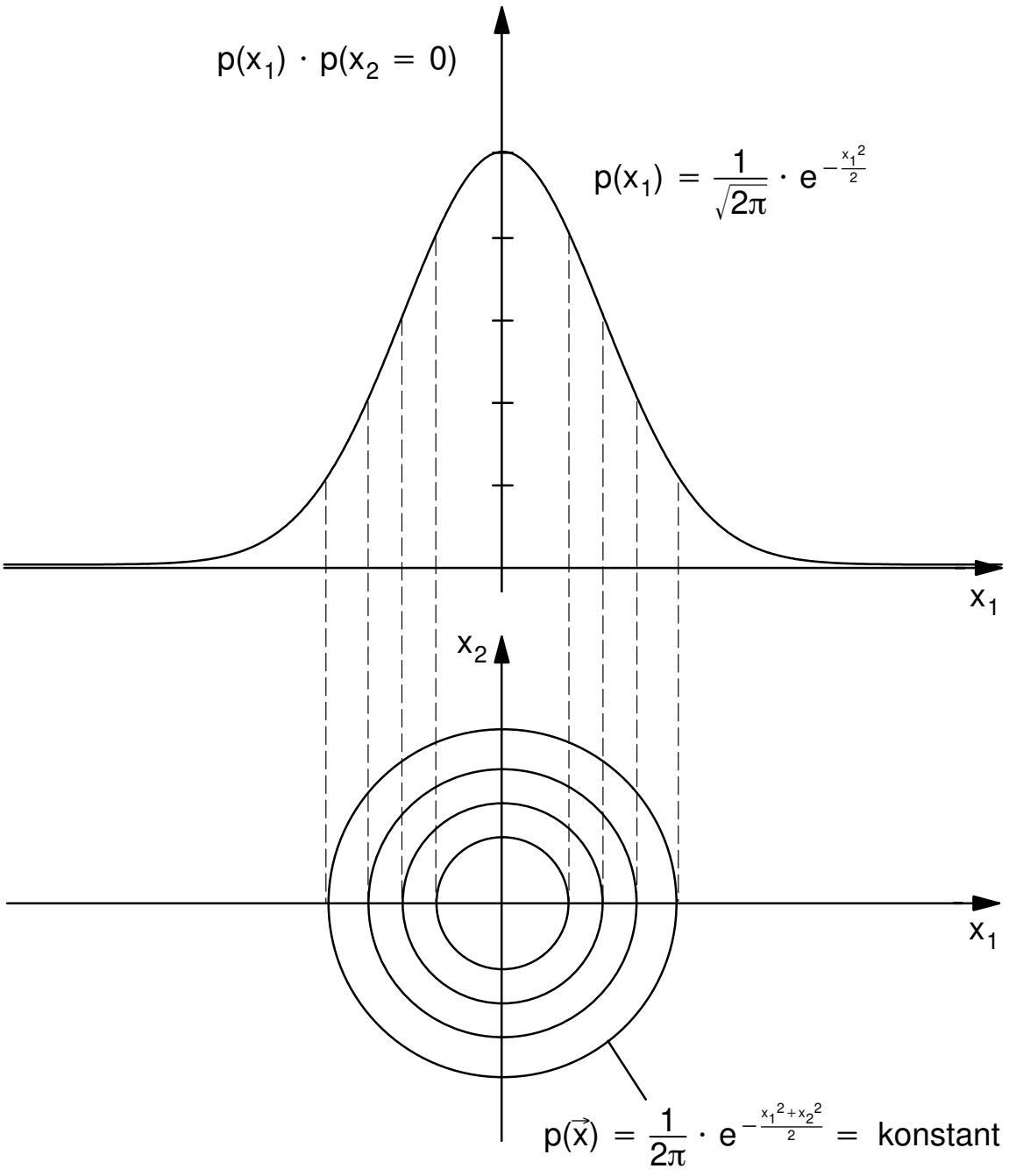


Bild 2.42 Höhenlinien konstanter Wahrscheinlichkeitsdichte
$p(\vec{x}) = p(x_1,\ x_2) = p(x_1) \cdot p(x_2)$
zweier normalverteilter Zufallsvariablen

2.5.4.1 Codebuchentwurf

Die Minimierung des mittleren quadratischen Quantisierungsfehlers

$$D = E[d(\vec{x},\ Q(\vec{x}))] = \int_{\vec{x}} d(\vec{x},\ Q(\vec{x}))\, p(\vec{x})\, d\vec{x} \tag{2.210}$$

in Abhängigkeit von der gegebenen Wahrscheinlichkeitsdichte $p(\vec{x})$ setzt eine optimale Anordnung der Repräsentativvektoren $\vec{y}_j$, $j = 1,\ ...,\ J$ voraus. Bild 2.43 veranschaulicht anhand eines Beispiels, daß die Abstände zwischen den Repräsentativvektoren dort klein gewählt werden müssen, wo $p(\vec{x})$ große Werte annimmt. Der Entwurf des optimalen Codebuchs erfolgt iterativ durch wiederholte Ausführung der beiden Anweisungen:

1. Wähle für ein gegebenes Codebuch C eine optimale Partition S_{opt}.
2. Wähle für eine gegebene Partition S ein optimales Codebuch C_{opt}.

Die erste Anweisung erzeugt bei vorgegebenem Codebuch eine Voronoische Partition S, indem die Vektoren $\vec{x}$ einer Trainingsfolge bestehend aus T Trainingsvektoren jeweils der Quantisierungszelle S_i mit dem Repräsentativvektor $\vec{y}_i$ zugeordnet werden, die der Bedingung

$$d(\vec{x},\ \vec{y}_i) \le d(\vec{x},\ \vec{y}_j) \qquad \text{für alle } j \neq i \tag{2.211}$$

genügt. Sind alle Trainingsvektoren zugeordnet, werden mit der dann erhaltenen Partition S die Repräsentativvektoren entsprechend der zweiten Anweisung aktualisiert. Dazu wird innerhalb jeder Quantisierungszelle S_j der Schwerpunktvektor cent S_j ermittelt und als aktualisierter Repräsentativvektor $\vec{y}_j$ verwendet. Er genügt der Bedingung

$$E[\, d(\vec{x},\ \vec{y}_j) \,|\, \vec{x} \in S_j] = \min_{\vec{u}} E[\, d(\vec{x},\ \vec{u}) \,|\, \vec{x} \in S_j]. \tag{2.212}$$

Bei Anwendung des mittleren quadratischen Fehlers als Verzerrungsmaß $d(\vec{x},\ \vec{y})$ folgt aus dieser Bedingung

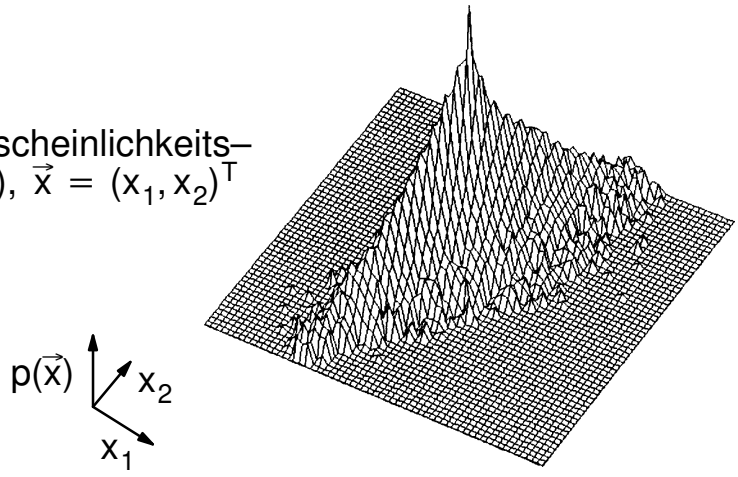
$$\frac{\partial}{\partial \vec{u}} E[(\vec{x} - \vec{u})^T (\vec{x} - \vec{u}) \,|\, \vec{x} \in S_j]\big|_{\vec{u}=\vec{y}_j} = \vec{0}$$

$$E[-2(\vec{x} - \vec{u}) \,|\, \vec{x} \in S_j]\big|_{\vec{u}=\vec{y}_j} = \vec{0}$$

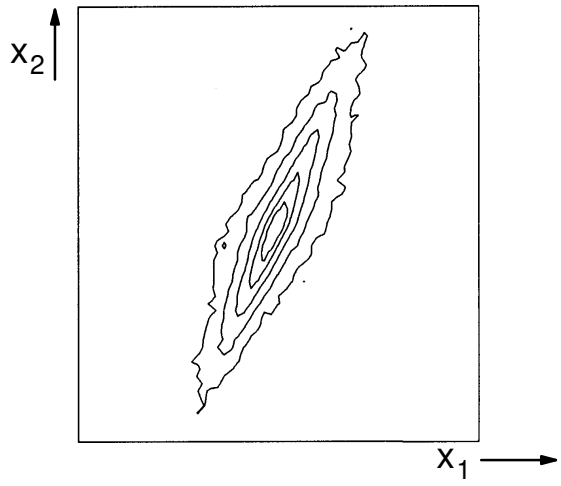
$$E\Big[\vec{x} \,|\, \vec{x} \in S_j\Big] - \vec{y}_j = \vec{0}$$

$$\vec{y}_j = \text{cent}\, S_j = E[\,\vec{x} \,|\, \vec{x} \in S_j\,] \qquad j = 1,\ ...,\ J\ . \tag{2.213}$$

a) Gemessene Wahrscheinlichkeits–dichtefunktion $p(\vec{x})$, $\vec{x} = (x_1, x_2)^T$



b) Höhenlinien $p(\vec{x})$ = konstant der Wahrscheinlichkeits–dichtefunktion



c) Optimierte Repräsentativvektoren und Quantisierungszellen der zweidimensionalen VQ

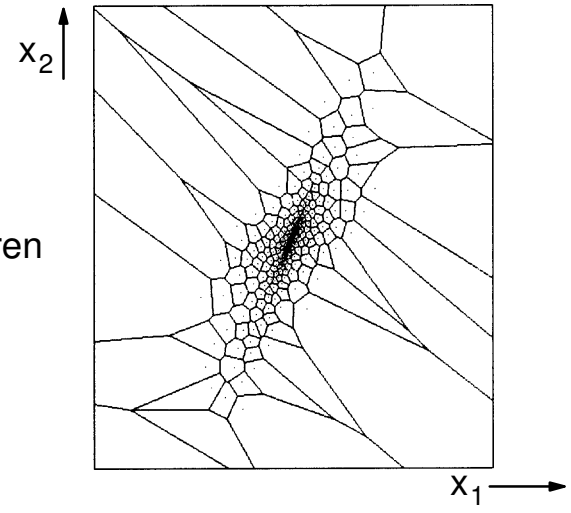


Bild 2.43 Wahrscheinlichkeitsdichtefunktion und optimiertes Codebuch einer deutschen Sprachsignalprobe

Mit jedem Iterationsschritt nimmt der mittlere Quantisierungsfehler ab oder bleibt im Grenzfall konstant.

Für den Sonderfall $N = 1$ entartet die Vektorquantisierung in eine skalare Quantisierung. Die Berechnung der Partition und Repräsentativvektoren entspricht dann der Berechnung der Schwellwerte und Repräsentativwerte des Max–Quantisierers.

Der Codebuchentwurf der VQ erfordert bei gegebener Wahrscheinlichkeitsdichte $p(\vec{x})$ die Auswertung von Mehrfachintegralen zur Bestimmung der obengenannten Erwartungswerte. Der beim praktischen Codebuchentwurf angewandte LBG–Algorithmus [11], benannt nach den Autoren Y. Linde, A. Buzo und R. M. Gray, umgeht dieses Problem, indem die gesuchten Erwartungswerte durch Zeitmittelwerte ersetzt werden. Diese werden anhand einer ausreichend langen Trainingsfolge ermittelt.

Zur Erzeugung des Startcodebuchs kann das sog. Splittingverfahren angewendet werden. Ausgehend vom Schwerpunktvektor der gesamten Trainingsfolge wird durch wiederholte Aufspaltung und Optimierung sukzessive die Auswahl der Repräsentativvektoren verdoppelt, bis schließlich die gewünschte Anzahl vorliegt. Die Optimierung erfolgt, wie beschrieben, iterativ, wobei in jedem Schritt zunächst die Trainingsfolge quantisiert und anschließend die Schätzung des optimalen Codebuchs aktualisiert wird. Dazu ersetzt man jeden Repräsentativvektor $\vec{y}_j$ durch den Schwerpunktvektor

$$\vec{y}_j = \frac{\sum_{t=1}^{T} \vec{x}_t \cdot 1(\vec{x}_t \in S_j)}{\sum_{t=1}^{T} 1\,(\vec{x}_t \in S_j)} \tag{2.214}$$

der entsprechenden Quantisierungszelle. Hierbei ist $1(\vec{x}_t \in S_j)$ die Indikatorfunktion

$$1(\vec{x}_t \in S_j) = \begin{cases} 1 & \text{für } \vec{x}_t \in S_j \\ 0 & \text{in allen anderen Fällen.} \end{cases} \tag{2.215}$$

Die Iteration wird so lange fortgesetzt, bis keine nennenswerte Verbesserung mehr zu verzeichnen ist.

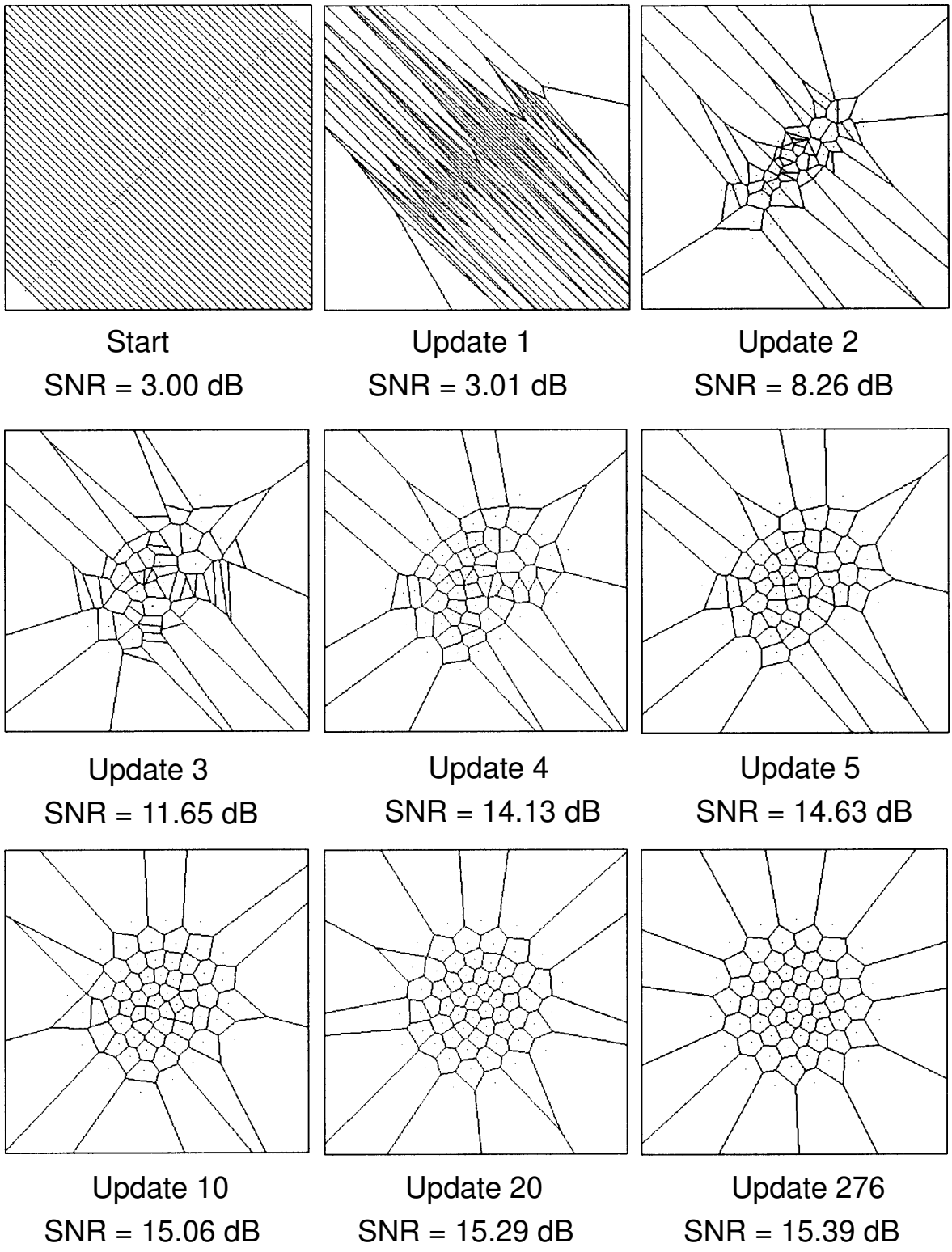


Bild 2.44 Partition und Repräsentativvektoren nach unterschiedlich vielen Iterationszyklen für eine unkorrelierte normalverteilte Zufallsfolge

Bild 2.44 veranschaulicht das Konvergenzverhalten des LBG–Algorithmus am Beispiel zweier unkorrelierter normalverteilter Zufallsvariablen. Die Simulationsergebnisse zeigen den beim Codebuchentwurf erzielten Signal–Rauschabstand in Abhängigkeit der Iterationsschritte für den Entwurf von $J = 64$ zweidimensionalen Repräsentativvektoren mit einer Trainingsfolge von $T = 800000$ Zufallszahlen. Als Startcodebuch dienen in diesem Beispiel Repräsentativvektoren, die in gleichmäßigen Abständen auf der Geraden $x_1 = x_2$ angeordnet sind.

2.5.4.2 Gleichförmige Vektorquantisierung

Der Rechenaufwand für den Codebuchentwurf läßt sich verkürzen, wenn der Entwurf auf Vektorquantisierer mit gleichförmiger geometrischer Struktur beschränkt wird, die auch als Lattice–VQ bezeichnet werden. Bild 2.45 zeigt 3 Beispiele gleichförmiger VQ. Dabei entspricht die Vektorquantisierung gemäß Bild 2.45(a) einer gleichmäßigen Skalarquantisierung.

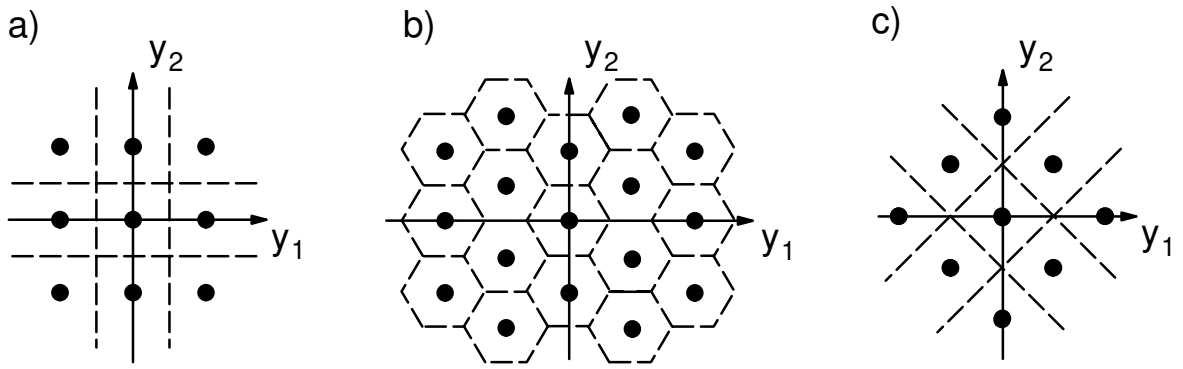


Bild 2.45 Lattice–Strukturen für *N=2*
a) Z_2–Lattice
b) hexagonales A_2–Lattice
c) D_2–Lattice

Im folgenden soll die zweidimensionale gleichförmige VQ mit hexagonalen Quantisierungszellen gemäß Bild 2.45(b) betrachtet werden. Diese Zellenform zeichnet sich dadurch aus, daß sie für ein gleichverteiltes Signal den mittleren Quantisierungsfehler minimiert.

Die hexagonale Zellenform wird mit einem VQ entsprechend Bild 2.46 erzeugt. Der Eingangsvektor $\vec{x} = (x_1,\ x_2)^T$ wird mittels der orthogonalen Transformation $\vec{z} = T^T\vec{x}$ mit

$$T^T = \begin{bmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{bmatrix} \tag{2.216}$$

um den Winkel β gedreht. Diese Drehung wird am Ausgang des Quantisierers durch Multiplikation mit T wieder kompensiert und soll zunächst unberücksichtigt bleiben ($\beta = 0$).

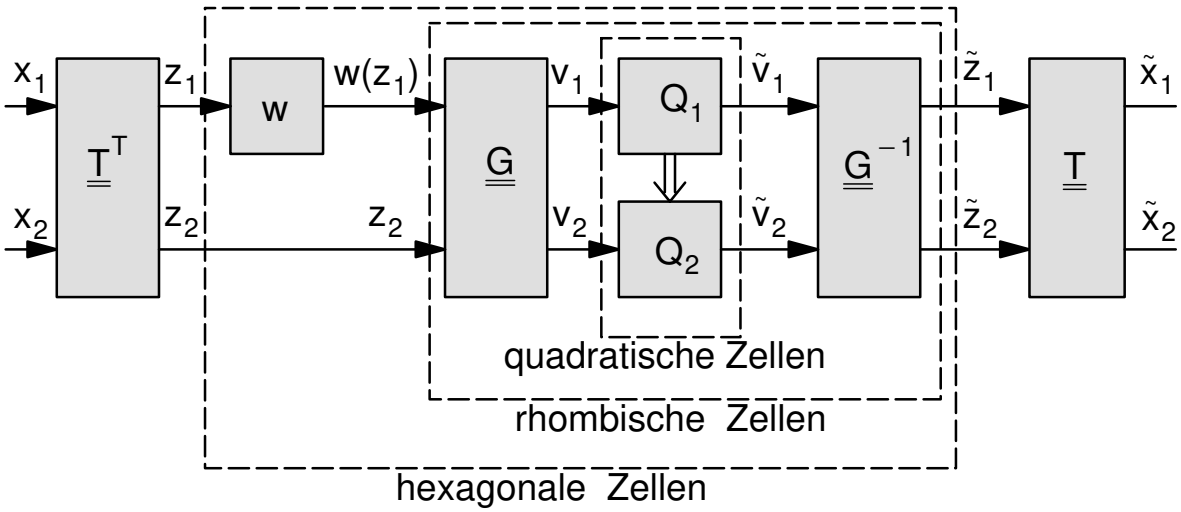


Bild 2.46 Blockschaltbild des zweidimensionalen Quantisierers mit hexagonalen Quantisierungszellen

Die erste Komponente des Vektors $\vec{z} = (z_1, z_2)^T$ wird nach der Vorschrift

$$w(z_1) = \begin{cases} n\dfrac{\Delta s}{2} & \text{für} \quad n\dfrac{\Delta s}{2} - \dfrac{\Delta s}{6} \le z_1 \le n\dfrac{\Delta s}{2} + \dfrac{\Delta s}{6} \\ 3z_1 - (2n+1)\dfrac{\Delta s}{2} & \text{für} \quad n\dfrac{\Delta s}{2} + \dfrac{\Delta s}{6} \le z_1 \le (n+1)\dfrac{\Delta s}{2} - \dfrac{\Delta s}{6} \end{cases} \tag{2.217}$$

nichtlinear abgebildet. Hierbei bezeichnet n eine ganze Zahl und Δs die Stufenhöhe der skalaren Quantisierer Q_1 und Q_2. Die Kennlinie dieser Abbildung ist in Bild 2.47 normiert dargestellt. Ihre Aufgabe ist es, die hexagonalen Quantisiererzellen in rhombische Form zu überführen.

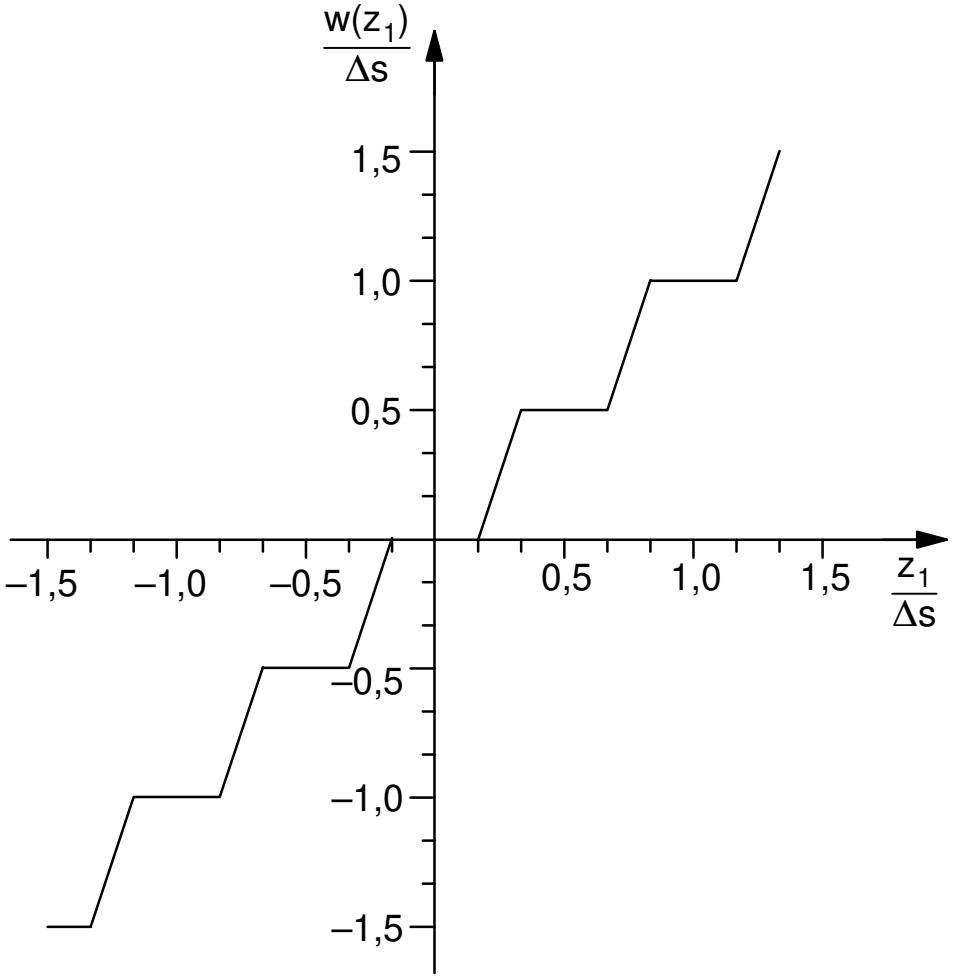


Bild 2.47 Kennlinie der nichtlinearen Abbildung $w(z_1)$

Die Wirkung der nichtlinearen Abbildung *w* auf die Eingangsvektoren veranschaulicht Bild 2.48. Alle Vektoren, die in Bild 2.48(a) in dunkel dargestellten Gebieten liegen, werden auf die Mittelsenkrechte der entsprechenden rhombischen Zelle in Bild 2.48(b) abgebildet. Außerdem werden die hellen Gebiete in Bild 2.48(a) in z_1–Richtung gestreckt. Infolgedessen liegen alle Vektoren, die vorher in einem der Hexagone enthalten waren, nach der Abbildung vollständig in einem der Rhomben. Die Lage der Repräsentativvektoren ändert sich dabei nicht. Man beachte, daß die Partition in Bild 2.48(b) keine Voronoische Partition ist, da die Grenzflächen der Quantisierungszellen nicht senkrecht auf der Verbindungsgeraden zwischen benachbarten Repräsentativvektoren stehen.

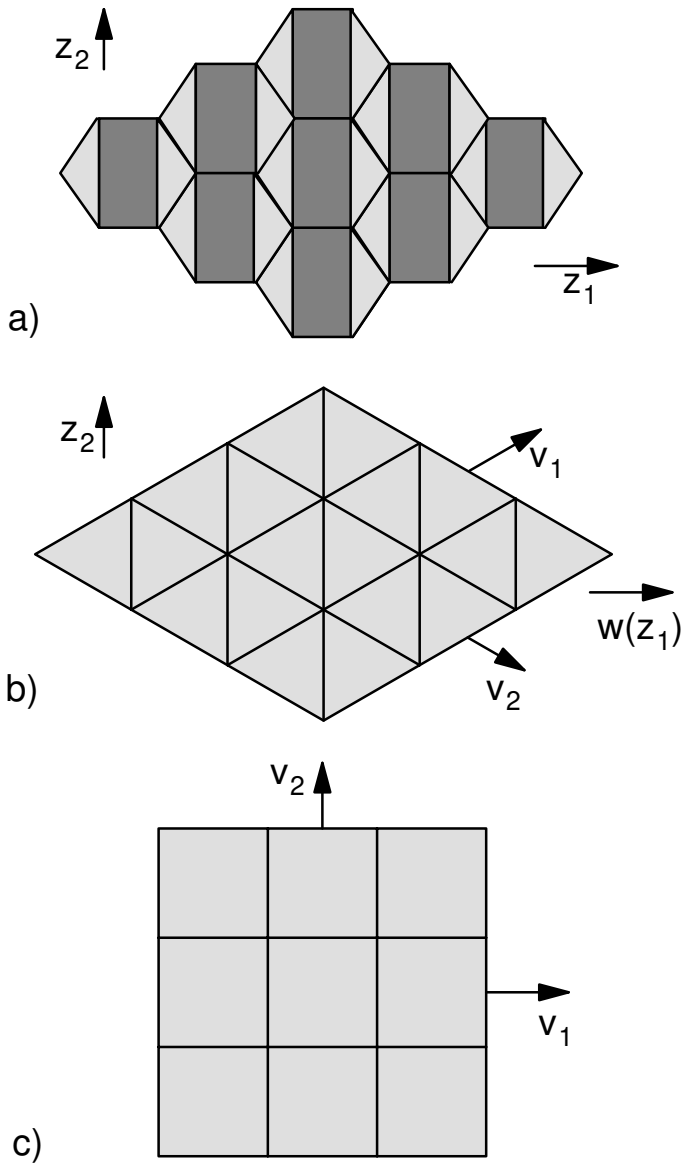


Bild 2.48 Übergang von hexagonalen auf quadratische Quantisierungszellen

Durch die lineare Transformation mit

$$G = \begin{bmatrix} 1 & \sqrt{3} \\ 1 & -\sqrt{3} \end{bmatrix} \tag{2.218}$$

werden die rhombischen Zellen in quadratische Zellen gemäß Bild 2.48(c) überführt. Diese VQ mit quadratischen Zellen realisiert man durch komponentenweise, gleichförmige Quantisierung mit den skalaren Quantisierern Q_1 und Q_2. Die Stufenhöhe Δs der skalaren Quantisierer bestimmt dabei nicht nur die Seitenlänge der

quadratischen Zellen, sondern auch die der Hexagone. Im Anschluß an die komponentenweise Quantisierung ist eine Rücktransformation der quantisierten Vektoren $\vec{\hat{v}}$ mit

$$G^{-1} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2\sqrt{3}} & -\frac{1}{2\sqrt{3}} \end{bmatrix} \tag{2.219}$$

erforderlich, um den Einfluß von G auf die Lage der Repräsentativvektoren rückgängig zu machen. Da $w(z_1)$ die Lage der Repräsentativvektoren nicht verändert, ist eine dazu inverse Operation nicht erforderlich.

Bei einer vorgegebenen endlichen Anzahl von J Repräsentativvektoren und unendlich ausgedehnter Wahrscheinlichkeitsdichte $p(\vec{x})$ gibt es immer einen Übersteuerungsbereich des gleichförmigen Vektorquantisierers. Beim Entwurf des Codebuches muß der durch Übersteuerung verursachte Fehlerzuwachs berücksichtigt werden, indem der Aussteuerungsbereich in Größe, Form und Lage der Wahrscheinlichkeitsdichte $p(\vec{x})$ angepaßt wird. Dazu muß der Rand des Aussteuerungsbereiches mit einer Höhenlinie $p(\vec{x}) = \text{konstant}$ zusammentreffen.

Bei der erläuterten Struktur des Vektorquantisierers kann die Größe des Aussteuerungsbereichs mit der Stufengröße Δs der Quantisierer Q_1 und Q_2 eingestellt werden. Die Form des Aussteuerungsbereichs wird durch die Aussteuerungsgrenzen der skalaren Quantisierer Q_1 und Q_2 festgelegt. Durch Anwendung einer zusätzlichen Transformation T kann die Lage des Aussteuerungsbereiches noch um den Winkel β gedreht werden.

Beispiel: Zwei mittelwertfreie, normalverteilte Zufallsvariablen x_1, x_2 mit dem Korrelationskoeffizienten ϱ haben die Verbundwahrscheinlichkeitsdichte

$$p(x_1,\ x_2) = \frac{1}{2\pi(1-\varrho^2)^{1/2}} \exp\left[-\frac{x_1^2 - 2\varrho x_1 x_2 + x_2^2}{2(1-\varrho^2)}\right].$$

Somit bilden die Höhenlinien konstanter Wahrscheinlichkeitsdichte $p(\vec{x}) = p(x_1,\ x_2)$ abhängig von ϱ Ellipsen oder Kreise. Zur Anpassung des Aussteuerungsbereichs des gleichmäßigen Vektorquantisierers wird durch Änderung des Winkels β die v_1–Achse in Bild 2.48(b), die dem skalaren Quantisierer Q_1 zugeordnet ist, mit der Hauptachse der Ellipsen zur Deckung gebracht. Bild 2.49 zeigt die optimale Anordnung von jeweils 256 hexagonalen Quantisierungszellen für verschiedene Korrelationskoeffizienten ϱ. Der Drehwinkel der Transformation T beträgt hier $\pi/12$. Er vergrößert den Winkel zwi-

schen v_1–Achse und Abszisse aus Bild 2.48(b) von $\pi/6$ auf $\pi/4$ und bringt so die v_1–Achse mit der Hauptachse, der in Bild 2.49 dargestellten Höhenlinien zur Deckung.

Bild 2.50 zeigt im Vergleich die Ergebnisse einer ungleichförmigen Vektorquantisierung für die gleiche Quelle. Das Signal–Rauschleistungs–Verhältnis der ungleichförmigen VQ ist abhängig von ϱ etwa 1–2 dB besser.–

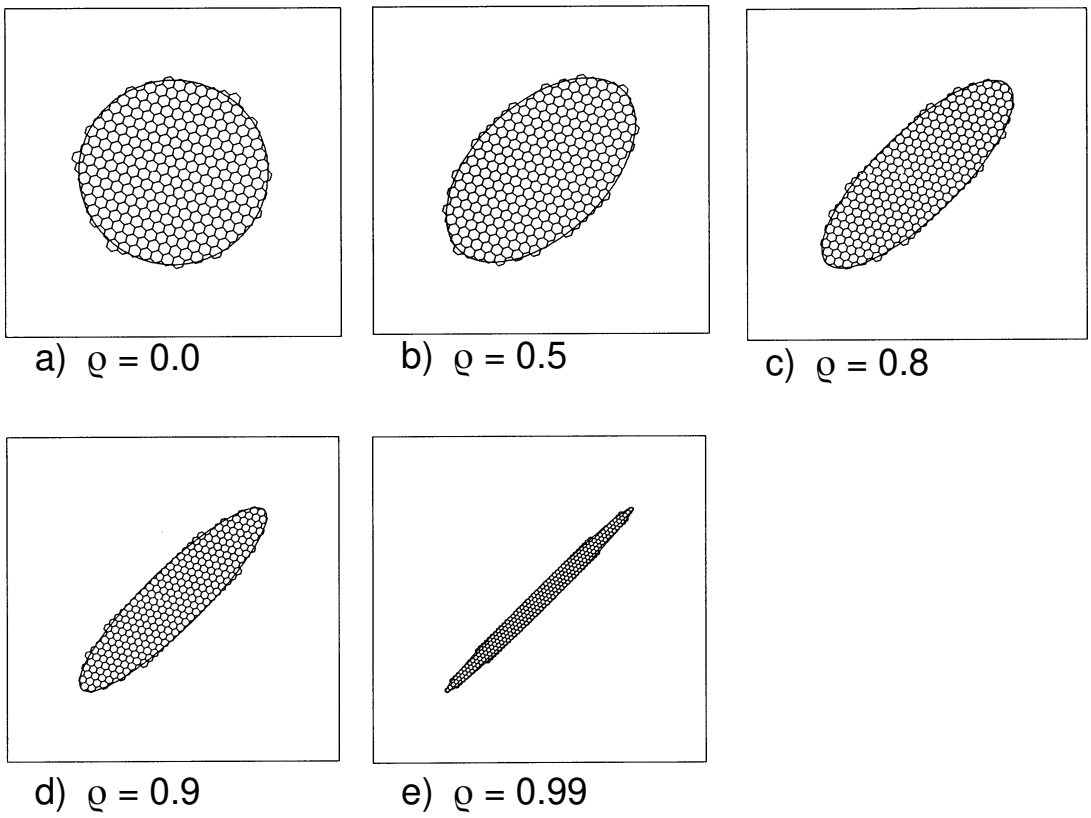


Bild 2.49 Optimale Anordnung und Größe von jeweils 256 Hexagonen für Gauss–Markoff Signale der gleichen Varianz mit unterschiedlichen Korrelationskoeffizienten ϱ (Drehwinkel $\beta = \pi/12$)

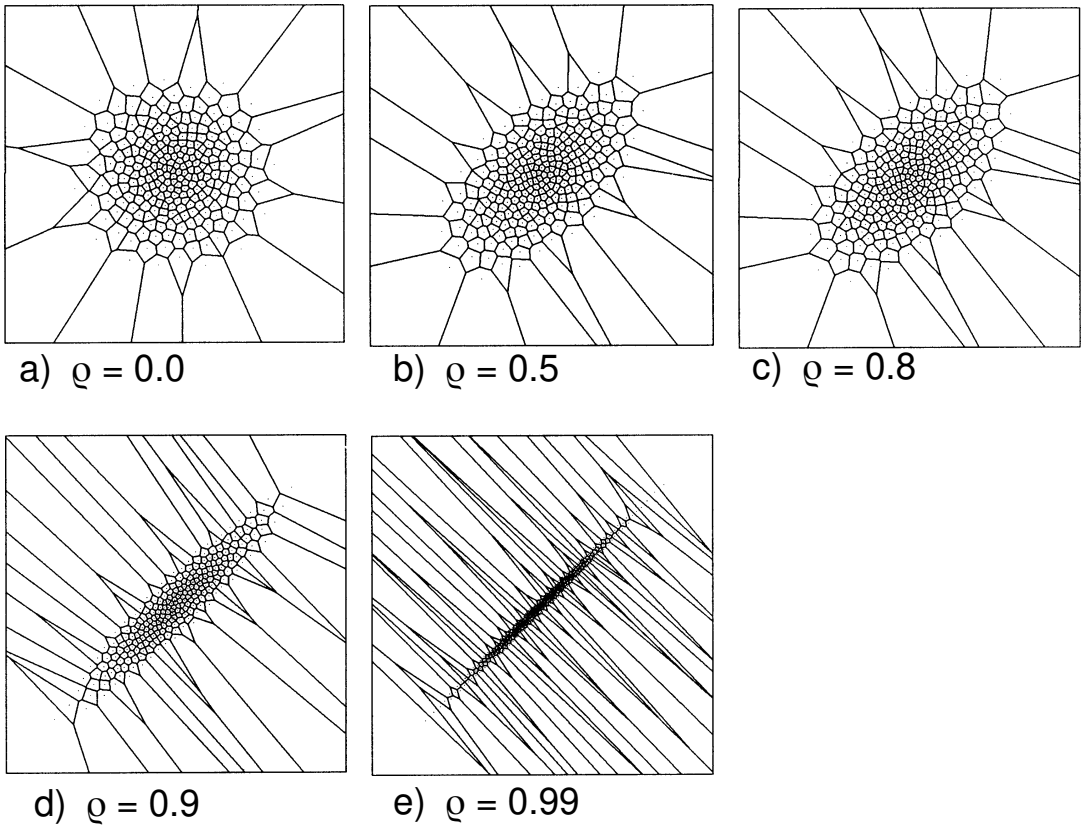


Bild 2.50 Zweidimensionale VQ mit jeweils 256 Repräsentativvektoren und Quantisierungszellen für Gauss–Markoff Signale der gleichen Varianz mit unterschiedlichen Korrelationskoeffizienten

3. Wesentliche Codierungsverfahren

Die zur Codierung von Sprach–, Ton– und Bildsignalen entwickelten und standardisierten Codierungsverfahren bestehen im allgemeinen aus einer Kombination mehrerer Codierungsverfahren. Die erste Stufe bildet dabei die Pulscodemodulation.

3.1 Pulscodemodulation (PCM)

Mit Hilfe der Pulscodemodulation können zeit– und wertkontinuierliche Quellensignale in eine binäre Darstellungsform umgewandelt werden. Die Pulscodemodulation kann somit als eine spezielle Quellencodierung angesehen werden. Ein PCM–Coder bzw. PCM–Decoder besteht dabei aus den in Bild 3.1 gezeigten Komponenten.

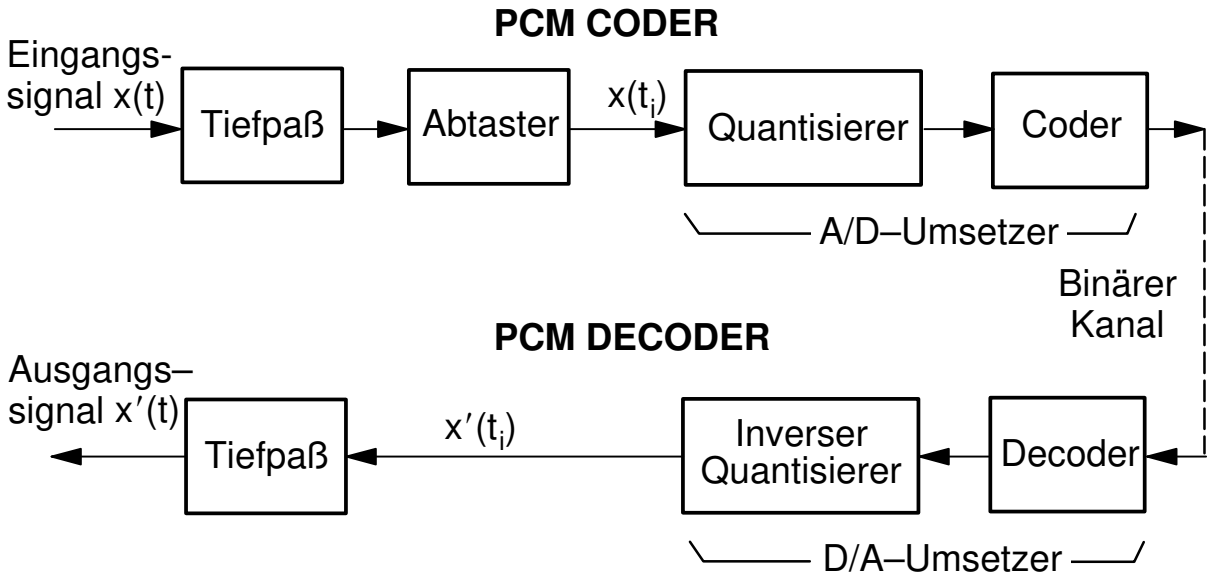


Bild 3.1 Komponenten eines PCM–Coders und –Decoders

Das kontinuierliche Eingangssignal $x(t)$ wird mittels eines Tiefpasses der Bandbreite $\pm\ W$ bandbegrenzt und anschließend mit einer Abtastfrequenz $f_A = 2W$ abgetastet. Der Tiefpaß dient zur Vermeidung von Aliasingfehlern, die durch Spektralanteile im Frequenzbereich $f > W$ entstehen können; vergleiche dazu Bild 3.2.

Es sei $S(f)$ das Leistungsdichtespektrum von $x(t)$. Dann hat das abgetastete Signal $x(t_i) = x_i$ das in Bild 3.2.b dargestellte periodische Leistungsdichtespektrum. Das

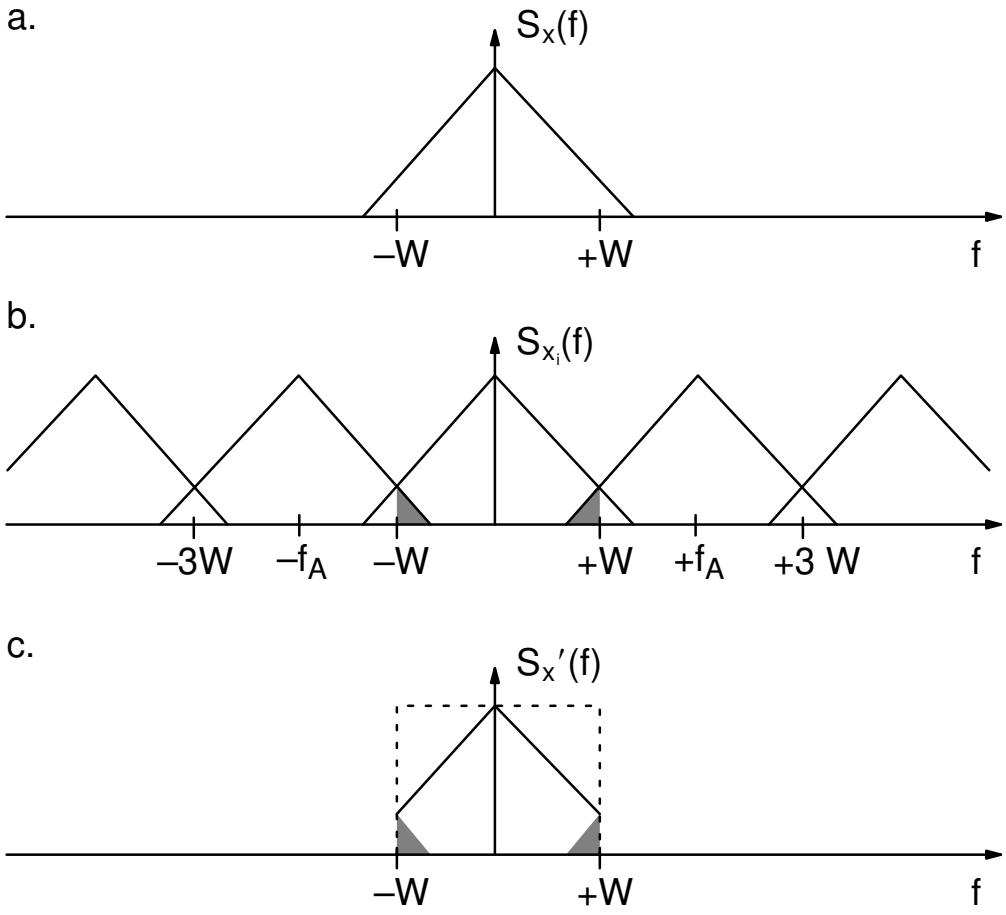


Bild 3.2 Darstellung der Entstehung von Aliasing–Fehlern ($f_A = 2W$)
a. Leistungsdichtespektrum des Eingangssignals
b. Leistungsdichtespektrum des abgetasteten Signals
c. Leistungsdichtespektrum des Ausgangssignals
(Aliasingfehler sind grau unterlegt)

abgetastete Signal x_i wird quantisiert und die Repräsentativwerte in binär codierter Form übertragen. Bei K Quantisierungsstufen und Codierung mit fester Codewortlänge beträgt die Übertragungsbitrate

$$R = 2W \cdot \log_2 K \quad \text{bit/s} \ . \tag{3.1}$$

Auf der Empfangsseite werden die binären Codewörter mittels des Decoders und des inversen Quantisierers wieder in Abtastwerte x_i' umgesetzt. Sie sind mit dem Quantisierungsfehler q_i behaftet.

$$x_i' = x_i + q_i \tag{3.2}$$

Zur Rekonstruktion des Eingangssignals $x'(t)$ müssen die quantisierten Abtastwerte x_i' mittels eines weiteren Tiefpasses der Bandbreite $\pm W$ interpoliert werden. Bild 3.2.c veranschaulicht, daß im entstehenden Leistungsdichtespektrum $S'(f)$ schraffiert gezeichnete Aliasingfehler enthalten sind, wenn das Leistungsdichtespektrum $S(f)$ auf der Sendeseite nicht hinreichend bandbegrenzt wird. Die Varianz σ_a^2 des Aliasingfehlers berechnet sich gemäß

$$\sigma_a^2 = 2 \int_W^\infty S(f)\, df \ . \tag{3.3}$$

Tabelle 3.1 und 3.2 zeigt standardisierte PCM–Formate für Ton– und Bildsignale. In den Beschreibungen der Standards sind auch Empfehlungen für die Übertragungsfunktion der zugehörigen Tiefpaßfilter enthalten.

PCM–Format	DVD–Audio (typ.)	DAT (typ.)	CD	Fern–sprechen (7 kHz)	Fern–sprechen ITU–T G.711
Abtastfrequenz	96 kHz	48 kHz	44,1 kHz	16 kHz	8 kHz
Quantisierung der Abtastwerte	24 bit gleichförmig	16 bit gleichförmig	16 bit gleichförmig	16 bit gleichförmig	8 bit ungleichförmig
Kanalanzahl	2	2	2	1	1
Bitrate	2·2,3 Mbit/s	2·768 kbit/s	2·706 kbit/s	256 kbit/s	256 kbit/s

Tabelle 3.1: PCM–Formate für Ton– und Sprachsignale

Im Fall einer Gaußschen Quelle und redundanzfreier Codierung ergibt sich die untere Grenze der Übertragungsbitrate mit Gl. (2.145) zu

$$R_{PCM} = \frac{1}{2}\log_2 \frac{{\sigma_x}^2}{{\sigma_q}^2} \tag{3.4}$$

in bit je Abtastwert. Der Signal–Rauschabstand (SNR) beträgt entsprechend Gl.(2.146)

$$SNR_{PCM} = 6,02\, R_{PCM} \text{ in dB.} \tag{3.5}$$

Mit gleichförmiger Quantisierung und anschließender redundanzmindernder Codierung erreicht man eine Datenrate, die um 0,25 bit je Abtastwert über dem Grenzwert liegt.

PCM–Format	HDTV–I (16:9)	ITU–R BT.601 (4:3)	CIF/SIF (4:3)	QCIF (4:3)
Auflösung in Pel Luminanz(Y) Chrominanz(U,V)	 1920·1152 960·1152	 720·576 360·576	 352·288 176·144	 176·144 88·72
Bildfolge–frequenz	50 Hz interlaced	50 Hz interlaced	29,97 Hz / 25 Hz non–interlaced	10 Hz (typ.) non–interlaced
Quantisierung d. Abtastwerte	8 bit	8 bit	8 bit	8 bit
Bitrate	885 Mbit/s	166 Mbit/s	36,5 Mbit/s / 30,5 Mbit/s	3 Mbit/s

Tabelle 3.2: PCM–Formate für Bewegtbildsignale

3.2 Prädiktive Codierung

3.2.1 Verlustlose prädiktive Codierung

Die prädiktive Codierung nutzt die statistische Abhängigkeit der gedächtnisbehafteten Markoff–Quelle zur Datenreduktion aus. Anstelle von u_N wird die Differenz $u_N - \hat{u}_N$ zwischen dem Abtastwert u_N und einem Prädiktionswert $\hat{u}_N$ codiert und übertragen, wie es in Bild 3.3 dargestellt ist. Der Prädiktionswert wird im Coder und im Decoder aus $N - 1$ vorausgehenden übertragenen Abtastwerten gebildet. Der mittlere zu übertragende Informationsgehalt ändert sich dabei nicht, d.h.

$$H(U_N - \hat{U}_N \mid U_1,\ ...,\ U_{N-1}) = H(U_N \mid U_1,\ ...,\ U_{N-1})\ . \tag{3.6}$$

Der Vorzug der prädiktiven Codierung im Vergleich zur zustandsabhängigen Codierung der Markoff–Quelle besteht darin, daß die bedingte Wahrscheinlichkeit des Differenzsignals

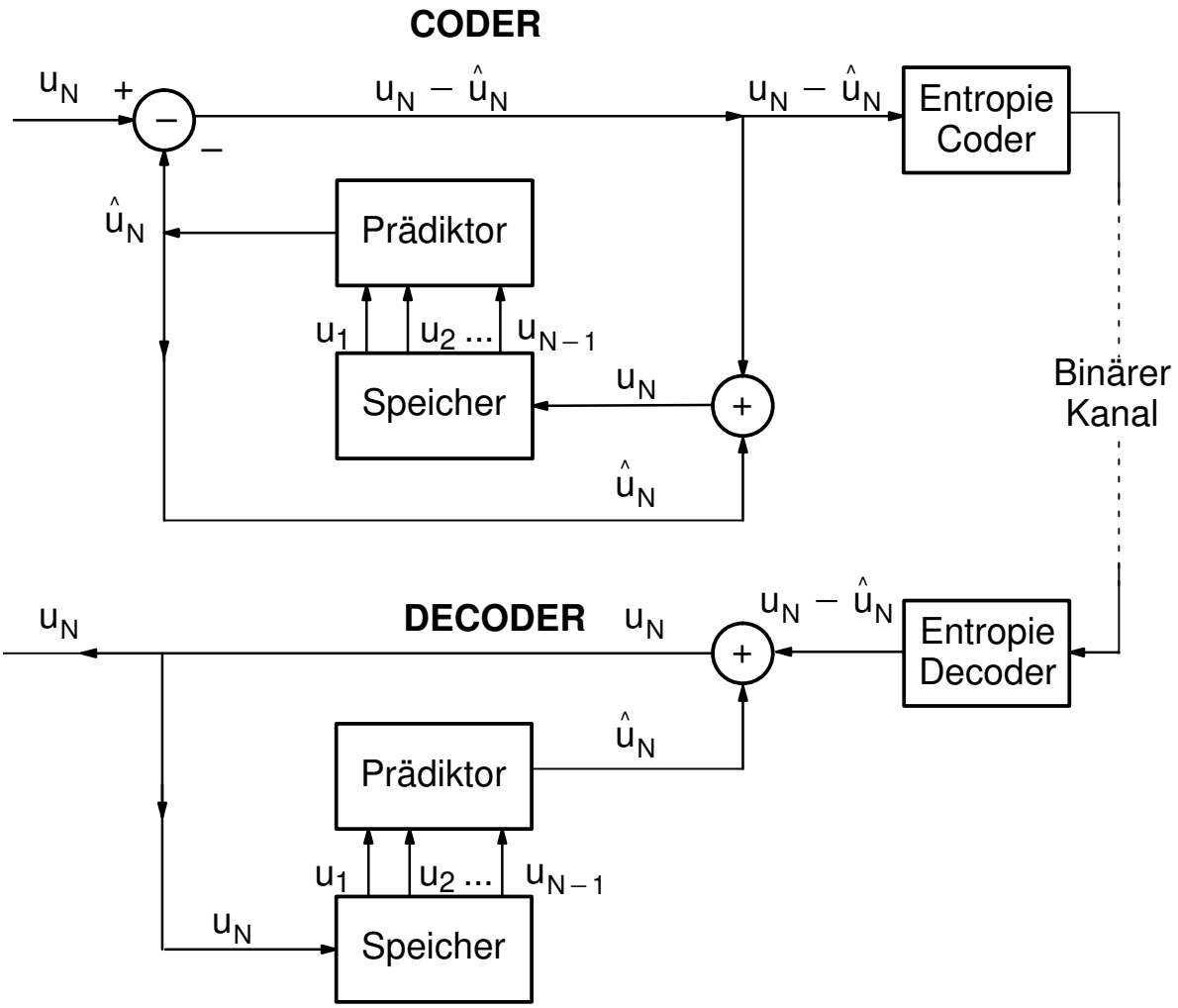


Bild 3.3 Blockschaltbild einer prädiktiven Codierung

$$P(u_N - \hat{u}_N \mid u_1,\ ...,\ u_{N-1}) \tag{3.7}$$

für viele Zustände $S_j = u_1,\ ...,\ u_{N-1}$ ähnlich wird und damit die Anzahl der Zustände für die prädiktive Codierung geringer ist. In den meisten Fällen wird sogar nur noch ein Zustand verwendet, d.h. es wird nur noch ein Huffman–Code benötigt, der an die Wahrscheinlichkeitsverteilung

$$P(u_N - \hat{u}_N) \tag{3.8}$$

angepaßt ist.

Das wesentliche Ziel der prädiktiven Codierung ist es, ein Differenzsignal zu erzeugen, dessen Entropie $H(U_N - \hat{U}_N)$ möglichst klein ist. Demzufolge sollte die Wahrscheinlichkeitsverteilung $P(u_N - \hat{u}_N)$ ein Ereignis mit sehr großer und viele mit möglichst kleiner Wahrscheinlichkeit aufweisen.

1. Lösungsansatz: Maximum–Likelihood–Prädiktor

Der Maximum–Likelihood–Prädiktor löst dieses Problem, indem er in jedem Zustand $S_j = u_1, \ ..., \ u_{N-1}$ das Symbol a_i mit der größten bedingten Wahrscheinlichkeit

$$P(\hat{u}_N = a_i \,|\, u_1, \ ..., \ u_{N-1}) \geq P(u_N = a_k \,|\, u_1, \ ..., \ u_{N-1}) \tag{3.9}$$

für alle k als Prädiktionswert verwendet. Als Ergebnis hat die Differenz

$$u_N - \hat{u}_N = 0 \tag{3.10}$$

die größte Wahrscheinlichkeit.

2. Lösungsansatz: Minimierung des mittleren quadratischen Fehlers

Häufig wird auch ein Prädiktor verwendet, der den bedingten mittleren quadratischen Fehler

$$E\left[(u_N - \hat{u}_N \,|\, u_1, \ ..., \ u_{N-1})^2\right] \tag{3.11}$$

minimiert. Dieser Prädiktor muß der Bedingung genügen

$$\frac{d}{d\,\hat{u}_N} E\left[(u_N - \hat{u}_N \,|\, u_1, \ ..., \ u_{N-1})^2\right] = 0$$

$$\frac{d}{d\,\hat{u}_N}\left\{ E\left[(u_N \,|\, u_1, \ ..., \ u_{N-1})^2\right] - 2\ \hat{u}_N E\left[(u_N \,|\, u_1, \ ..., \ u_{N-1})\right] + \hat{u}_N^2 \right\} = 0$$

$$-\ 2\, E\left[(u_N \,|\, u_1, \ ..., \ u_{N-1})\right] + 2\ \hat{u}_N = 0\ . \tag{3.12}$$

Daraus folgt für den optimalen Prädiktor

$$\hat{u}_N = E\left[(u_N \,|\, u_1, \ ..., \ u_{N-1})\right] = \sum_{u_N} u_N P\,(u_N \,|\, u_1, \ ..., \ u_{N-1}) \tag{3.13}$$

in jedem Zustand $S_j = u_1, \ ..., \ u_{N-1}$. Dieser Prädiktor verwendet den linearen Mittelwert als Prädiktionswert.

Die Optimierung der beiden behandelten Prädiktoren setzt die Kenntnis der bedingten Wahrscheinlichkeit $P\,(u_N \,|\, u_1, \ ..., \ u_{N-1})$ voraus. Diese Voraussetzung ist nicht immer gegeben.

3. Lösungsansatz: Linearer Prädiktor

Der lineare Prädiktor

$$\hat{u}_N = \sum_{n=1}^{N-1} b_n \, u_n \tag{3.14}$$

bildet den Prädiktionswert mittels einer Linearkombination der bereits übertragenen Abtastwerte $u_1, \ ..., \ u_{N-1}$ unter Verwendung von Prädiktorkoeffizienten b_n als Gewichtsfaktor. Die Minimierung des mittleren quadratischen Fehlers führt in diesem Fall auf die Bedingung

$$\begin{aligned}
&\frac{\partial}{\partial b_n}\left\{E\left[(u_N - \hat{u}_N)^2\right]\right\} = 0 \\
&\frac{\partial}{\partial b_n}\left\{E\left[\left[u_N - \sum_{m=1}^{N-1} b_m \, u_m\right]^2\right]\right\} = 0 \\
&-2 \, E\left[\left[u_N - \sum_{m=1}^{N-1} b_m \, u_m\right] \cdot u_n\right] = 0 \qquad \text{für} \quad n = 1, \ ..., \ N-1 \, .
\end{aligned} \tag{3.15}$$

Bildet man mit $d_n = E[u_N \cdot u_n]$ und b_n die Spaltenvektoren $\vec{d} = (d_1, \ ..., \ d_{N-1})^T$ und $\vec{b} = (b_1, \ ..., \ b_{N-1})^T$ sowie mit den Elementen $r_{nm} = E[u_n \cdot u_m]$ die Korrelationsmatrix R, so folgt aus Gl. (3.15)

$$\vec{d} - R \, \vec{b}_{opt} = \vec{0} \ , \tag{3.16}$$

und die optimalen Prädiktorkoeffizienten ergeben sich gemäß

$$\vec{b}_{opt} = R^{-1} \, \vec{d} \ . \tag{3.17}$$

Zur Bestimmung des optimalen linearen Prädiktors müssen die Elemente der Korrelationsmatrix R und des Korrelationsvektors $\vec{d}$ bekannt sein. Für den Fall eines stationären Prozesses genügt dabei die Kenntnis der Autokorrelationsfunktion bzw. des Leistungsdichtespektrums.

3.2.2 Differenz–Pulscodemodulation (DPCM)

Die Differenz–Pulscodemodulation ist eine prädiktive Codierung, die im Vergleich zur PCM neben der durch Quantisierung verursachten Irrelevanzreduktion eine Redundanzreduktion bewirkt und auf diese Weise bei gleicher mittlerer Verzerrung mit geringerer Übertragungsrate auskommt. Die zusätzliche Redundanzreduktion wird mittels eines linearen Prädiktors erzielt, der die Korrelation der Abtastwerte ausnutzt. Die Wirkungsweise einer DPCM soll anhand von Bild 3.4 erläutert werden.

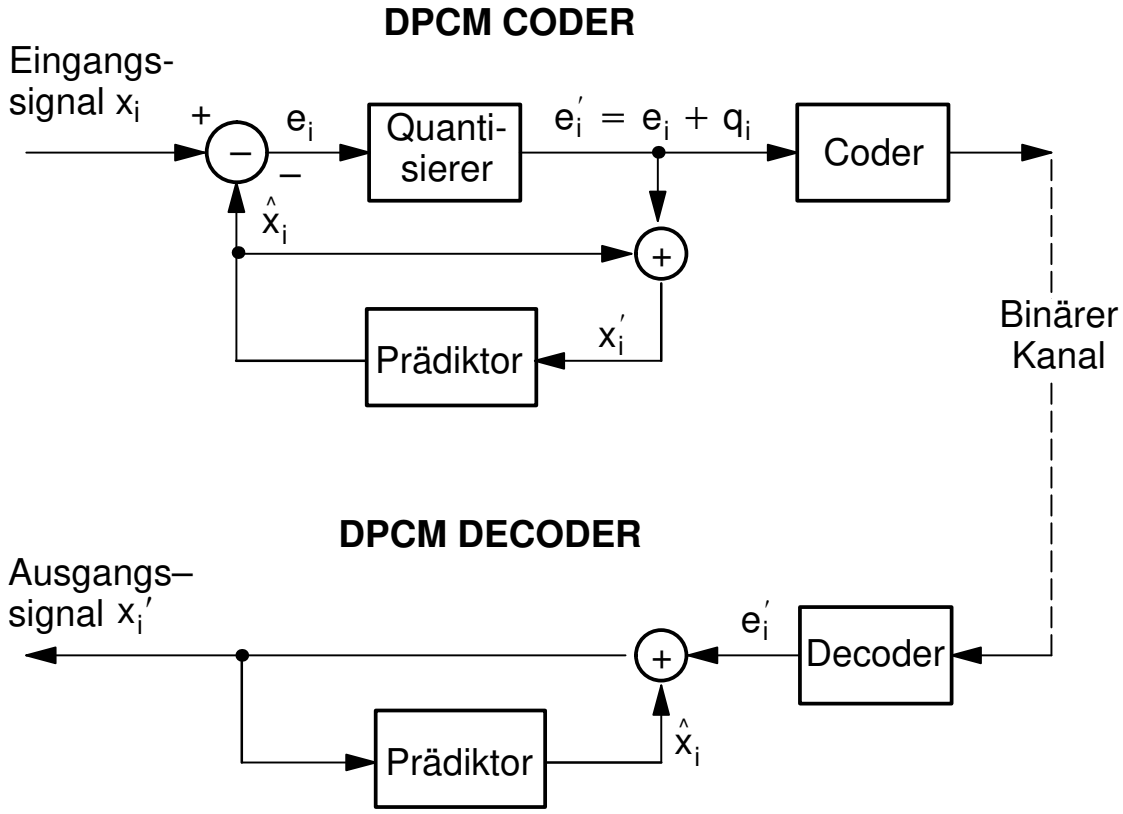


Bild 3.4 Komponenten eines DPCM–Coders und –Decoders

In einem DPCM–System wird von jedem Eingangsabtastwert $x_i = x(t_i)$ zum Zeitpunkt t_i, der mittels linearer Prädiktion gebildete Vorhersagewert $\hat{x}_i$ subtrahiert. Der Vorhersagefehler oder auch sog. Prädiktionsfehler e_i

$$e_i = x_i - \hat{x}_i \tag{3.18}$$

wird quantisiert und der mit dem Quantisierungsfehler q_i behaftete Wert e'_i

$$e'_i = e_i + q_i \tag{3.19}$$

in binär codierter Form übertragen. Beim Empfänger wie auch im Sender wird zu e'_i der Vorhersagewert $\hat{x}_i$ addiert. Die Summe

$$x'_i = e'_i + \hat{x}_i \tag{3.20}$$

ergibt das Ausgangssignal x'_i. Bei störungsfreier Übertragung ist der Quantisierungsfehler $x'_i - x_i$ des rekonstruierten Abtastwerts x'_i

$$x'_i - x_i = e'_i + \hat{x}_i - x_i = e'_i - e_i = q_i \tag{3.21}$$

gleich dem Quantisierungsfehler des entsprechenden Vorhersagefehlers e'_i.
Der Vorhersagewert $\hat{x}_i$ des linearen Prädiktors wird aus einer Linearkombination von $N - 1$ vorangegangenen rekonstruierten Abtastwerten x'_{i-n} gemäß der Beziehung

$$\hat{x}_i = \sum_{n=1}^{N-1} a_n \cdot x'_{i-n} \tag{3.22}$$

gebildet. Um den mittleren quadratischen Prädiktionsfehler $E[e_i]^2$ zu minimieren, werden die Prädiktorkoeffizienten a_n gemäß Gl. (3.17) gewählt. Sie genügen der Bedingung Gl. (3.15).

$$\begin{aligned} \frac{\partial}{\partial a_n}\left\{E\left[\left[x_i - \sum_{m=1}^{N-1} a_m\, x'_{i-m}\right]^2\right]\right\} &= 0 \\ -\,2\,E\left[\left[x_i - \sum_{m=1}^{N-1} a_m\, x'_{i-m}\right] x'_{i-n}\right] &= 0 \\ E\left[e_i \cdot x'_{i-n}\right] &= 0 \end{aligned} \tag{3.23}$$

für $n = 1, \ldots, N - 1$. Daraus folgt, daß der Prädiktionsfehler e_i und der Vorhersagewert orthogonal sind.

$$E\left[e_i \cdot \hat{x}_i\right] = 0 \ . \tag{3.24}$$

Der minimierte mittlere quadratische Prädiktionsfehler ergibt sich mit der Näherung

$$x'_{i-n} \approx x_{i-n} \tag{3.25}$$

und Gl.(3.24) zu

$$
\begin{aligned}
E[e_i^2] &= E\left[e_i(x_i - \hat{x}_i)\right] \\
&= E\left[e_i\, x_i\right] - E\left[e_i \cdot \hat{x}_i\right] \\
&= E\left[(x_i - \hat{x}_i)\, x_i\right] \\
&= E\left[x_i^2 - \hat{x}_i\, x_i\right] \\
&= E\left[x_i^2\right] - a_1\, E\left[x_{i-1}\, x_i\right] - a_2\, E\left[x_{i-2}\, x_i\right] \ldots \\
&\qquad \ldots - a_{N-1}\, E\left[x_{i-N+1}\, x_i\right] \,. \qquad (3.26)
\end{aligned}
$$

Aus Gl. (3.26) wird erkennbar, daß der Beitrag eines jeden Prädiktorkoeffizienten a_n zur Verringerung des mittleren quadratischen Prädiktionsfehlers $E[e_i]^2$ von den Korrelationen der vergangenen Abtastwerte zum aktuellen Abtastwert x_i abhängt. Unter der Annahme

$$E\,[\,x_i\,] = E\,[\,e_i\,] = 0 \qquad (3.27)$$

folgt aus Gl.(3.26)

$$
\begin{aligned}
\sigma_e{}^2 &= \sigma_x^2\,[1 - a_1\varrho_1 - a_2\varrho_2 \ldots - a_{N-1}\varrho_{N-1}] \\
&= \sigma_x^2\,[1 - \vec{\varrho}^{\,T}\,\vec{a}\,] \,, \qquad . \qquad (3.28)
\end{aligned}
$$

wobei $\vec{a}$ der Spaltenvektor der Prädiktionskoeffizienten a_n und $\vec{\varrho}$ ein Spaltenvektor der Korrelationskoeffizienten

$$\varrho_n = \frac{E\,[x_i\, x_{i-n}]}{\sigma_x^2} \qquad n = 1,\ \ldots,\ N-1 \qquad (3.29)$$

ist.

Indem man in Gl. (3.17) R und $\vec{d}$ durch die Varianz σ_x^2 teilt und die Reihenfolge der Elemente vertauscht, erhält man

$$\vec{a}_{opt} = \Lambda^{-1}\vec{\varrho}\,, \qquad (3.30)$$

wobei Λ die zur Korrelationsmatrix R gehörende Kovarianzmatrix mit den Elementen

$$\lambda_{nm} = \frac{E[x_{i-n}\, x_{i-m}]}{\sigma_x^2} \qquad n,\ m = 1,\ \ldots,\ N-1 \qquad (3.31)$$

ist.

Aus Gl. (3.30) folgt für Gl. (3.28)

$$\sigma_e^2 = \sigma_x^2[\, 1 - \vec{\varrho}^{\,T} \Lambda^{-1} \vec{\varrho}\,] \ . \tag{3.32}$$

Der mit der linearen Prädiktion einer DPCM erzielbare Gewinn in dB

$$G_{DPCM} = 10 \log_{10} \frac{\sigma_x^2}{\sigma_e^2} = 10 \log_{10} (\, 1 - \vec{\varrho}^{\,T} \Lambda^{-1} \vec{\varrho}\,)^{-1} \tag{3.33}$$

ist ein Maß für die reduzierte Signalleistung am Eingang des Quantisierers, aber nicht immer ein Maß für die Reduktion der Datenrate; insbesondere dann nicht, wenn die Wahrscheinlichkeitsdichte von x und e sich sehr unterscheiden. Bild 3.5 zeigt die Anwendung der Gl.(3.33) für Fernsprechsignale in Abhängigkeit von der Anzahl $N - 1$ der Prädiktionskoeffizienten.

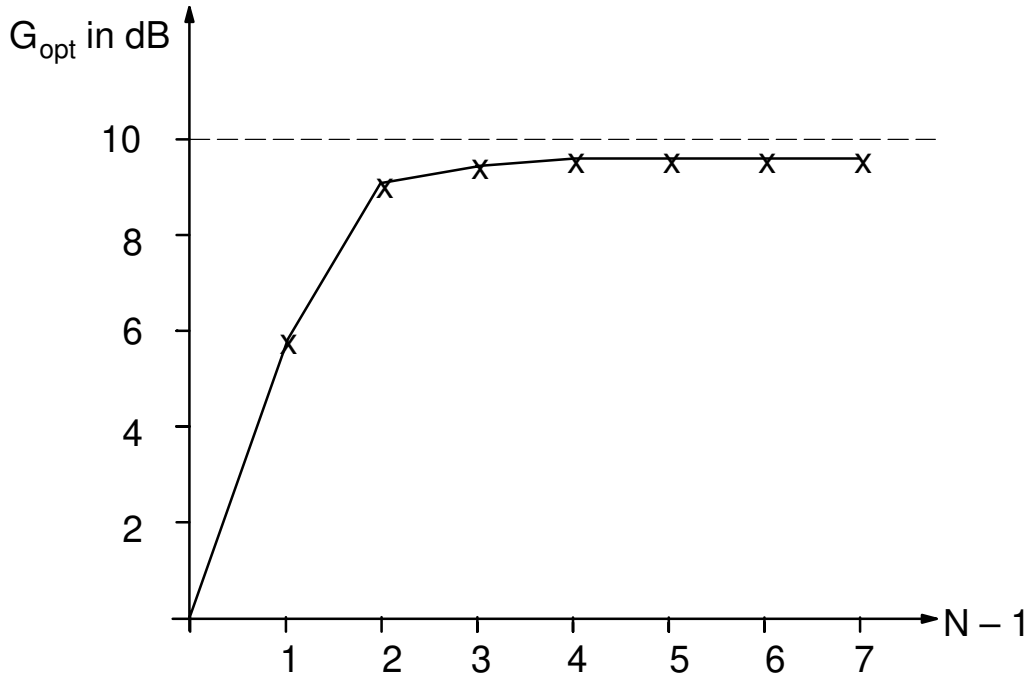


Bild 3.5 DPCM–Gewinn als Funktion der Anzahl N–1 der Prädiktionskoeffizienten für Fernsprechsignale der Frequenzbandbreite W = 3,4 kHz und Abtastfrequenz f_A = 8 kHz

Im Fall einer Gaußschen Quelle haben x_i und e_i je eine Gaußsche Wahrscheinlichkeitsdichte, die sich nur bezüglich der Varianzen σ_x^2 und σ_e^2 unterscheiden. Vergleicht man die DPCM mit der PCM bei festem SNR, d.h.

$$SNR_{DPCM} = SNR_{PCM}\,,$$

so gilt mit Gl. (2.145) für die Datenrate

$$
\begin{aligned}
R_{DPCM} &= \frac{1}{2}\log_2 \frac{\sigma_e^2}{\sigma_q^2} \\
&= \frac{1}{2}\log_2 \frac{\sigma_x^2}{\sigma_q^2} \cdot \frac{\sigma_e^2}{\sigma_x^2} \\
&= \frac{1}{2}\log_2 \frac{\sigma_x^2}{\sigma_q^2} - \frac{1}{2}\log_2 \frac{\sigma_x^2}{\sigma_e^2} \\
&= R_{PCM} - \frac{1}{6,02} \cdot G_{DPCM} \ ,
\end{aligned}
\tag{3.34}
$$

d.h. die Datenrate der DPCM verringert sich im Vergleich zur Datenrate der PCM um den Betrag $G_{DPCM} \ / \ 6,02$ in bit je Abtastwert.
Umgekehrt gilt bei unveränderter Datenrate

$$
\begin{aligned}
R_{DPCM} &= R_{PCM} \\
\frac{1}{2}\log_2 \frac{\sigma_e^2}{\sigma_{q_D}^2} &= \frac{1}{2}\log_2 \frac{\sigma_x^2}{\sigma_{q_P}^2} \\
\frac{\sigma_e^2}{\sigma_{q_D}^2} &= \frac{\sigma_x^2}{\sigma_{q_P}^2} \\
10\log_{10} \frac{\sigma_x^2}{\sigma_{q_D}^2} &= 10\log_{10} \frac{\sigma_x^2}{\sigma_{q_P}^2} \cdot \frac{\sigma_x^2}{\sigma_e^2}
\end{aligned}
$$

$$
SNR_{DPCM} = SNR_{PCM} + G_{DPCM} \ , \tag{3.35}
$$

d.h. der Signal–Rauschabstand der DPCM verbessert sich im Vergleich zur PCM um G_{DPCM} in dB.

3.3 Transformations–Codierung

Bei dieser Codierung werden N in PCM–Darstellung vorliegende Abtastwerte x_i zu einem Block bzw. Spaltenvektor $\vec{x} = (x_1, \ ..., \ x_n, \ ..., \ x_N)^T$ zusammengefaßt und vor der Übertragung einer linearen Transformation T mit den Elementen t_{mn} unterworfen, wie es in Bild 3.6 dargestellt ist.
Die N Komponenten des Ergebnisvektors $\vec{c} = (c_1, \ ..., \ c_N)^T$

$$
\vec{c} = T\,\vec{x} \tag{3.36}
$$

werden als Transformationskoeffizienten bezeichnet.

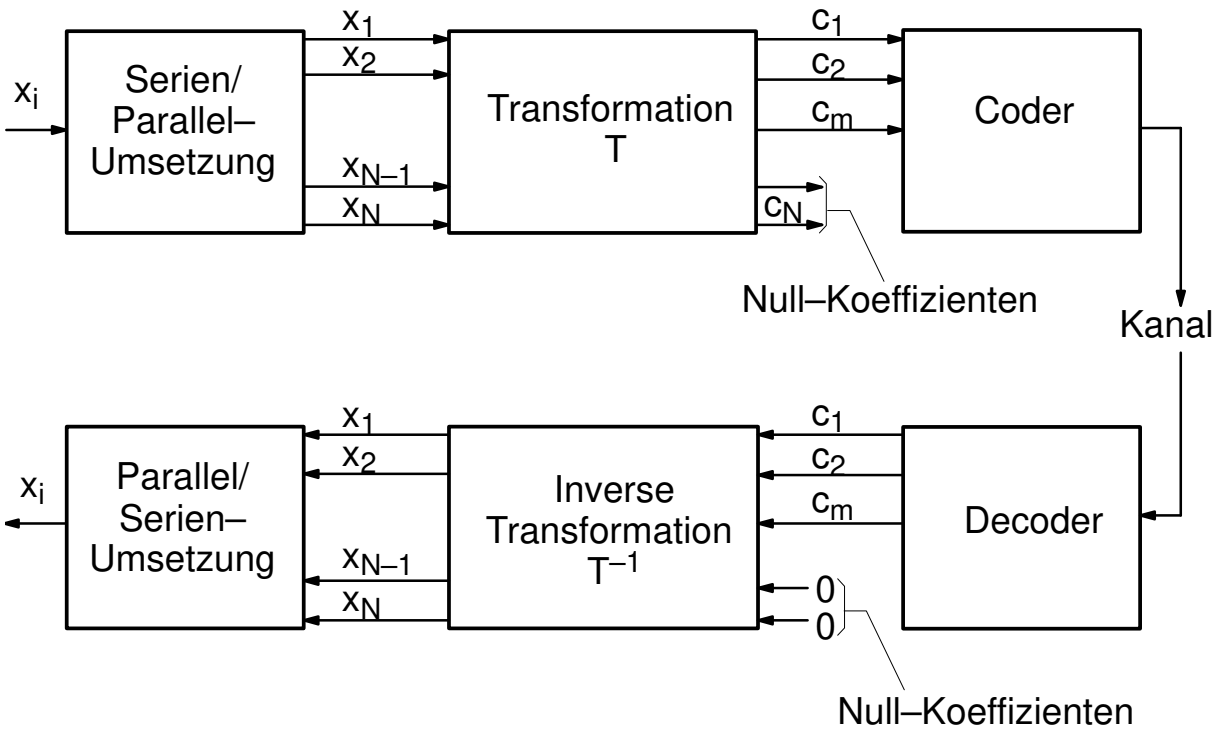


Bild 3.6 Blockschaltbild einer Transformationscodierung

Aus ihnen kann der Eingangsvektor $\vec{x}$ mittels einer inversen Transformation T^{-1} zurückgewonnen werden

$$\vec{x} = T^{-1}\,\vec{c}. \tag{3.37}$$

Für unitäre Transformationen gilt dabei zusätzlich

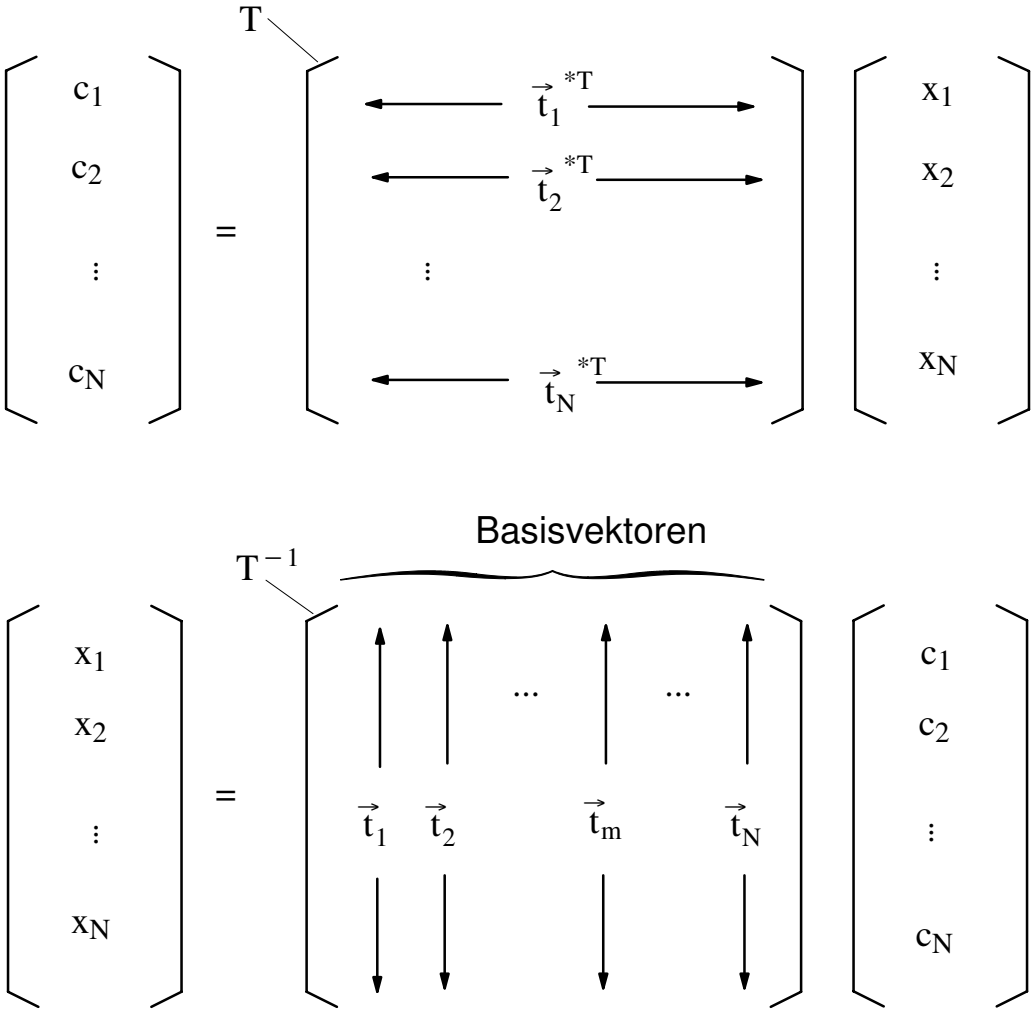
$$T^{-1} = T^{*T}, \tag{3.38}$$

wobei T^{*T} die zu T transponierte konjugiert komplexe Matrix ist. Eine unitäre Transformation mit reeller Matrix T wird als orthogonale Transformation bezeichnet, wobei $T^{-1} = T^{T}$ gilt.

Es seien $\vec{t}_m$ die Spaltenvektoren von T^{-1}. Dann kann Gl. (3.37) auch in der Form

$$\vec{x} = \sum_{m=1}^{N} c_m\,\vec{t}_m \tag{3.39}$$

geschrieben werden. Demzufolge ist $\vec{x}$ als Linearkombination der Spaltenvektoren oder auch sog. Basisvektoren von T^{-1} darstellbar, wobei jeder Basisvektor $\vec{t}_m$ durch den zugehörigen Transformationskoeffizienten c_m gewichtet wird, vergl. dazu Bild 3.7.



$$\vec{x} = \vec{t}_1 c_1 + \vec{t}_2 c_2 + \ldots + \vec{t}_N c_N$$

Bild 3.7 Basisvektoren unitärer Transformationen

Da der Informationsgehalt von $\vec{x}$ sich bei linearer umkehrbarer Transformation nicht ändert, gilt

$$H(\vec{X}) = H(\vec{C}) \tag{3.40}$$

in bit je Block.

Es ist das Ziel der Transformationscodierung, die statistische Abhängigkeit und die damit verbundene Redundanz der Transformationskoeffizienten zu verringern.

Der Vorteil einer Transformation, die statistische Abhängigkeiten der Transformationskoeffizienten $\{c_1, \ldots, c_N\}$ eliminiert, liegt darin, daß infolge

$$H(\vec{C}) = \sum_{m=1}^{N} H(c_m) \tag{3.41}$$

die Transformationskoeffizienten nicht im Block, sondern einzeln mit viel geringerem Aufwand codiert werden können.

3.3.1 Karhunen–Loeve–Transformation

Bisher ist nur eine orthogonale Transformation bekannt, mit denen unkorrelierte Transformationskoeffizienten c_m erzeugt werden können. Sie wird als Karhunen–Loeve–Transformation (KLT) bezeichnet und erfüllt die Bedingung

$$E[(c_n - m_n)(c_m - m_m)] = \lambda_m \delta_{nm} \qquad n,\ m = 1,\ ...,\ N \tag{3.42}$$

wobei $m_n = E[c_n],\ m_m = E[c_m]$ und λ_m die Varianz von c_m ist.
Zur Erläuterung der KLT sei

$$E[\vec{x}] = E[\vec{c}] = \vec{0} \tag{3.43}$$

angenommen. Dann gilt für die $N \times N$ Korrelationsmatrix R

$$\begin{aligned} R &= E[\,\vec{x}\ \vec{x}^T] \\ &= E[T^T\ \vec{c}\ \vec{c}^T T] \\ &= T^T E\,[\vec{c}\ c^T] T \end{aligned} \tag{3.44}$$

und mit der Gl. (3.38)

$$R\,T^T = T^T\,E\,[\,\vec{c}\,\vec{c}^T] \ . \tag{3.45}$$

Aufgrund von Gl. (3.42) ist im Fall der KLT $E[\vec{c}\,\vec{c}^T]$ eine N x N Diagonalmatrix, deren Elemente λ_m entlang der Diagonalen ($n = m$) die Varianzen von c_m sind. Mit den Spaltenvektoren $\vec{t}_m$ von T^T wird demzufolge

$$R\,\vec{t}_m = \lambda_m \vec{t}_m \qquad m = 1,\ ...,\ N\ , \tag{3.46}$$

d.h. die Spaltenvektoren $\vec{t}_m$ der KLT sind die Eigenvektoren der Korrelationsmatrix R und die Varianzen λ_m der KLT Koeffizienten die Eigenwerte von R. Da R symmetrisch und positiv definit ist, sind alle Eigenwerte reell und positiv. Mit dem Gleichungssystem Gl. (3.46) können die Elemente der KLT bestimmt werden.

Die KLT hat die zusätzlichen vorteilhaften Eigenschaften, daß die Spaltenvektoren so angeordnet werden können, daß die Varianzen λ_m mit zunehmendem Index fallen, d.h. $\lambda_1 > \lambda_2,\ ...,\ \lambda_N$. Werden Koeffizienten $c_1,\ ...,\ c_N$, deren Varianz kleiner als ein vorgegebener Schwellwert ist, auf Null gesetzt, so ergibt sich für den Fall, daß nur die ersten pN Koeffizienten größer als die Schwelle sind, ein Ausgangsvektor

$$\vec{c}' = (c_1,\ c_2,\ ...,\ c_{pN},\ 0,\ 0,\ ...,\ 0)^T\ , \tag{3.47}$$

dessen Null–Koeffizienten mittels eines Lauflängen–Codewortes sehr effizient codiert werden können.
Werden stets nur pN Koeffizienten übertragen, wobei $0 < p < 1$, so entsteht ein mittlerer quadratischer Fehler

$$E\left[\frac{1}{N}\sum_{n=1}^{N}(x_n - x_n')^2\right] = \frac{1}{N}\sum_{m=pN+1}^{N}\lambda_m\ . \tag{3.48}$$

3.3.2 Diskrete Fourier Transformation

Neben der KLT werden eine Reihe anderer linearer orthonormierter Transformationen zur Codierung verwendet, die unterschiedliche Vorzüge aufweisen.
So kann gezeigt werden, daß die unitäre Diskrete Fourier Transformation (DFT) mit den Elementen

$$t_{mn} = \frac{1}{\sqrt{N}}\exp\left[-j\frac{2\pi}{N}(n-1)(m-1)\right] \quad n,\ m = 1,\ ...,\ N \tag{3.49}$$

für stationäre Prozesse und große N ebenfalls unkorellierte Koeffizienten $c_1,\ ...,\ c_N$ liefert.
Bild 3.8 zeigt die Transformationsmatrix T der DFT für $N = 8$. Da $\vec{x}$ reell ist, sind die komplexen Koeffizienten von $\vec{c}$ konjugiert symmetrisch, d.h.

$$c_{2+p} = c^*_{N-p} \qquad 0 \le p \le N-2\ . \tag{3.50}$$

Demzufolge müssen zur Rekonstruktion von $\vec{x}$ nur N reelle Werte übertragen werden.
Die Rücktransformation der Koeffizienten c_m liefert ein periodisch erweitertes Signal $\vec{x}$ mit Diskontinuitäten an den Blockgrenzen, s. Bild 3.9.a. Die Diskontinuitäten erzeugen unerwünschte energiereiche Koeffizienten in hochfrequenten Bereichen. Sie können durch Anwendungen einer Diskreten Cosinus Transformation vermindert werden.

a.

$$\mathrm{Re(T)} = \frac{1}{\sqrt{8}} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & \sqrt{1/2} & 0 & -\sqrt{1/2} & -1 & -\sqrt{1/2} & 0 & \sqrt{1/2} \\ 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 \\ 1 & -\sqrt{1/2} & 0 & \sqrt{1/2} & -1 & \sqrt{1/2} & 0 & -\sqrt{1/2} \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & -\sqrt{1/2} & 0 & \sqrt{1/2} & -1 & \sqrt{1/2} & 0 & -\sqrt{1/2} \\ 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 \\ 1 & \sqrt{1/2} & 0 & -\sqrt{1/2} & -1 & -\sqrt{1/2} & 0 & \sqrt{1/2} \end{bmatrix}$$

b.

$$\mathrm{Im(T)} = \frac{1}{\sqrt{8}} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\sqrt{1/2} & -1 & -\sqrt{1/2} & 0 & \sqrt{1/2} & 1 & \sqrt{1/2} \\ 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 \\ 0 & -\sqrt{1/2} & 1 & -\sqrt{1/2} & 0 & \sqrt{1/2} & -1 & \sqrt{1/2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sqrt{1/2} & -1 & \sqrt{1/2} & 0 & -\sqrt{1/2} & 1 & -\sqrt{1/2} \\ 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 \\ 0 & \sqrt{1/2} & 1 & \sqrt{1/2} & 0 & -\sqrt{1/2} & -1 & -\sqrt{1/2} \end{bmatrix}$$

Bild 3.8 Matrix der Diskreten Fourier Transformation für $N=8$,
a. Realteil, b. Imaginärteil

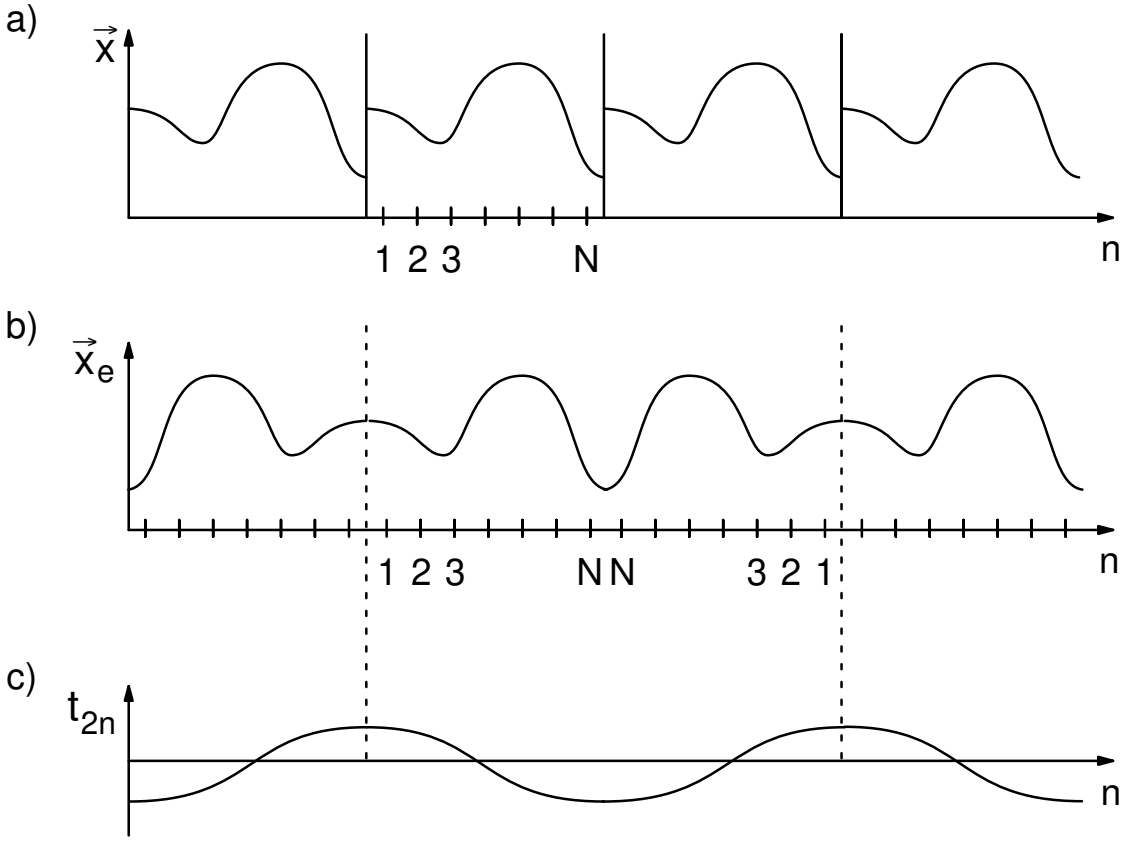


Bild 3.9 Illustration der periodisch erweiterten Signale einer DFT (a), einer DCT (b) und einer entsprechenden Basisfunktion (c)

3.3.3 Diskrete Cosinus Transformation

Bei der Diskreten Cosinus Transformation (DCT) wird der Eingangsvektor $\vec{x}$ durch Spiegelung zu einem Vektor $\vec{x}_e$ mit $2N$ Abtastwerten erweitert.

$$\vec{x}_e = (x_1,\ x_2,\ ...,\ x_N,\ x_N,\ ...,\ x_2,\ x_1)^T \tag{3.51}$$

Diesen Vektor $\vec{x}_e$ kann man mit einer zeitverschobenen DFT der Länge $2N$ transformieren, deren Symmetriepunkt bei $n = \frac{1}{2}$ liegt, so daß $\vec{x}_e$ bezüglich des Symmetriepunktes eine gerade Funktion darstellt, s. Bild 3.9.b. Die zeitverschobene DFT hat die Elemente

$$t_{mn} = \frac{1}{\sqrt{2N}} \exp\left[-j \frac{2\pi}{2N} \left(n - \frac{1}{2} \right) (m - 1) \right] . \tag{3.52}$$

Infolge dessen liefert die Transformation $2N$ Koeffizienten, deren Imaginärteile Null sind. Ferner sind die Koeffizienten paarweise gleich, so daß nur N berechnet

werden müssen. Nach der Orthonormierung auf N erhält man die Elemente der sogenannten Diskreten Cosinus Transformation

$$t_{mn} = \sqrt{\frac{2 - \delta_{mn}}{N}} \cos\left[\frac{\pi}{N}\left(n - \frac{1}{2}\right)(m - 1)\right] \qquad n,\ m = 1,\ ...,\ N. \tag{3.53}$$

Vergleiche dazu Bild 3.9.c.

3.3.4 Walsh–Hadamard Transformation

Die Walsh–Hadamard Transformation (WHT) hat den Vorzug einer besonders einfachen Realisierung, da nur Additionen und Subtraktionen auszuführen sind. Die symmetrische Matrix T der Walsh–Hadamard Transformation wird nach folgendem Bildungsgesetz aufgestellt.

Für $N = 2^n$ mit $n = 0,\ 1,\ 2,\ ...$ sei

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{3.54}$$

und allgemein

$$H_{2N} = \begin{bmatrix} H_N & H_N \\ H_N & -H_N \end{bmatrix} . \tag{3.55}$$

Dann lautet die orthonormierte WHT

$$T = \frac{1}{\sqrt{N}} H_N = T^T . \tag{3.56}$$

Bild 3.10 zeigt die Transformationsmatritzen der WHT für $N = 4$ und $N = 8$.

(a) $$T_4 = \frac{1}{\sqrt{4}} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

(b) $$T_8 = \frac{1}{\sqrt{8}} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix}$$

Bild 3.10 Matrix T der WHT für (a) $N = 4$ und (b) $N = 8$

Zum Vergleich mit anderen Transformationen werden die Basisvektoren auch in Form von Schwingungsverläufen dargestellt. Ordnet man diese anschließend nach wachsender Frequenz, erhält man eine Darstellung entsprechend Bild 3.11. Die Frequenz kann anhand der Anzahl der Nulldurchgänge geteilt durch zwei überprüft werden. Sie wird in diesem Zusammenhang auch als Sequency bezeichnet.

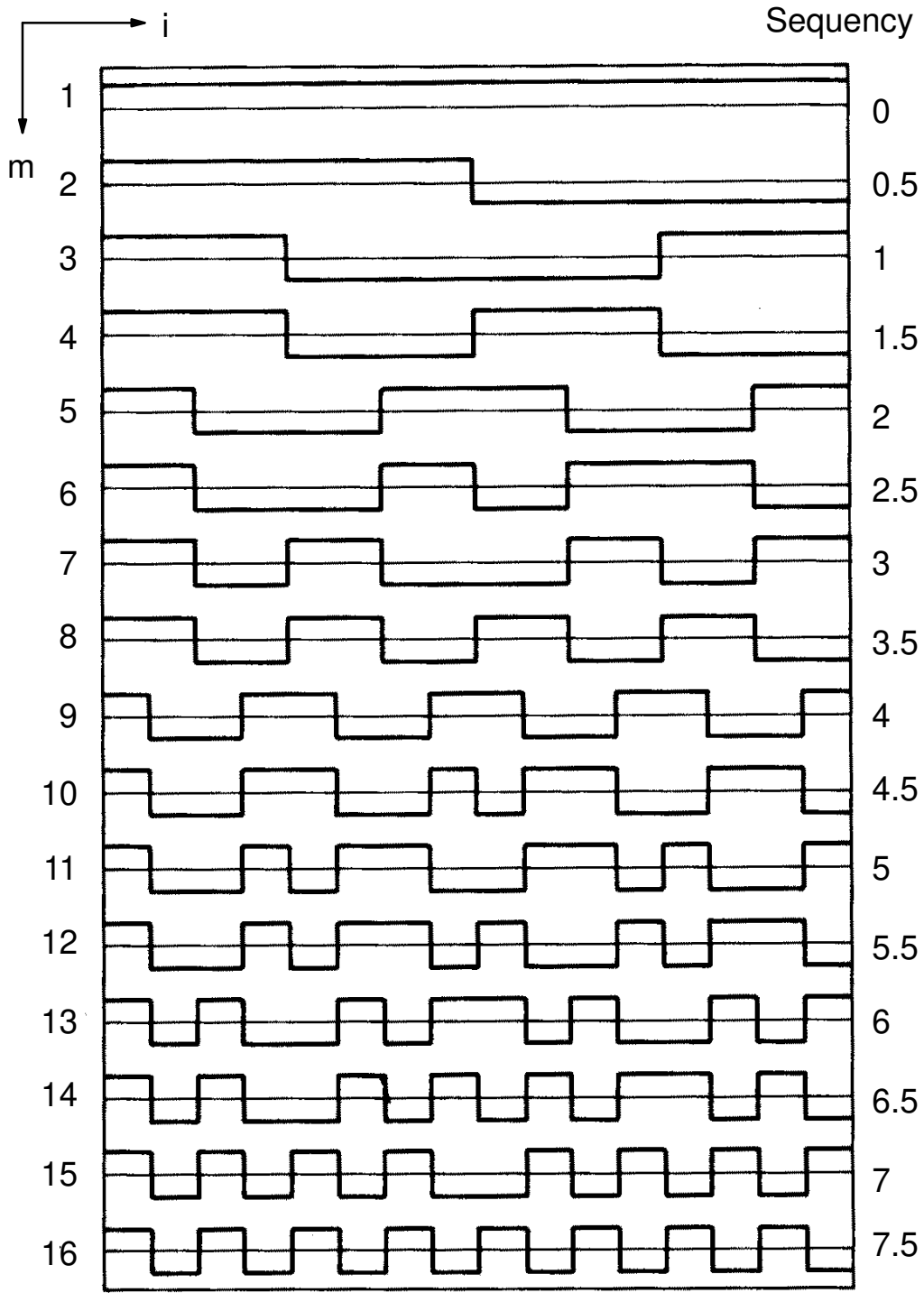


Bild 3.11 Nach Frequenz (Sequency) geordnete Basisvektoren der Walsh–Hadamard Transformation für $N = 16$

3.3.5 Zweidimensionale Transformationen

Für die Bildcodierung werden Transformationen benötigt, die sowohl die horizontalen als auch die vertikalen Korrelationen der Bildpunkte zur Codierung ausnutzen. Dazu werden Blöcke von $N \times N$ Bildpunkten gebildet. Jeder Block wird durch eine Matrix

$$X = [\, x_{ij} \,] \tag{3.57}$$

dargestellt.

Bei Anwendung separierbarer zweidimensionaler Transformationen kann die Transformation des Blockes X in zwei Schritten ausgeführt werden. In einem ersten Schritt werden die Zeilen von X der Länge N transformiert, um die horizontalen Korrelationen auszunutzen. Anschließend werden die Spalten transformiert, um die vertikalen Korrelationen auszunutzen. Die beiden Operationen können in der Form

$$c_{mn} = \sum_{i=1}^{N} \sum_{j=1}^{N} x_{ij}\, t_{nj}\, t_{mi} \tag{3.58}$$

geschrieben werden, wobei t_{nj} und t_{mi} Elemente der $N \times N$ Transformationsmatrix T sind. Unter der Annahme, daß die eindimensionalen Basisvektoren $\vec{t}_m$ nach wachsender Frequenz geordnet sind, kennzeichnet der Index m die vertikalen räumlichen Frequenzen und der Index n die horizantalen.

Zwischen den Matrizen C, T und X besteht folgender Zusammenhang

$$C = T\, X\, T^T . \tag{3.59}$$

Für orthogonale Transformationen kann die Rücktransformation

$$X = T^T\, C\, T \tag{3.60}$$

auch in der Form

$$X = \sum_{m=1}^{N} \sum_{n=1}^{N} c_{mn}\, \vec{t}_m\, \vec{t}_n{}^T \tag{3.61}$$

geschrieben werden. Jedes dyadische Produkt $\vec{t}_m\, \vec{t}_n{}^T$ stellt eine Matrix oder auch ein sogenanntes Basisbild dar. Der Block X kann gemäß Gl. (3.61) als eine gewichtete Summe der N^2 Basisbilder dargestellt werden, s. Bild 3.12. Das Gewicht eines Basisbildes wird durch die Größe des zugehörigen Transformationskoeffizienten c_{mn} bestimmt. Bild 3.13 zeigt als Beispiel die Basisbilder einer WHT.

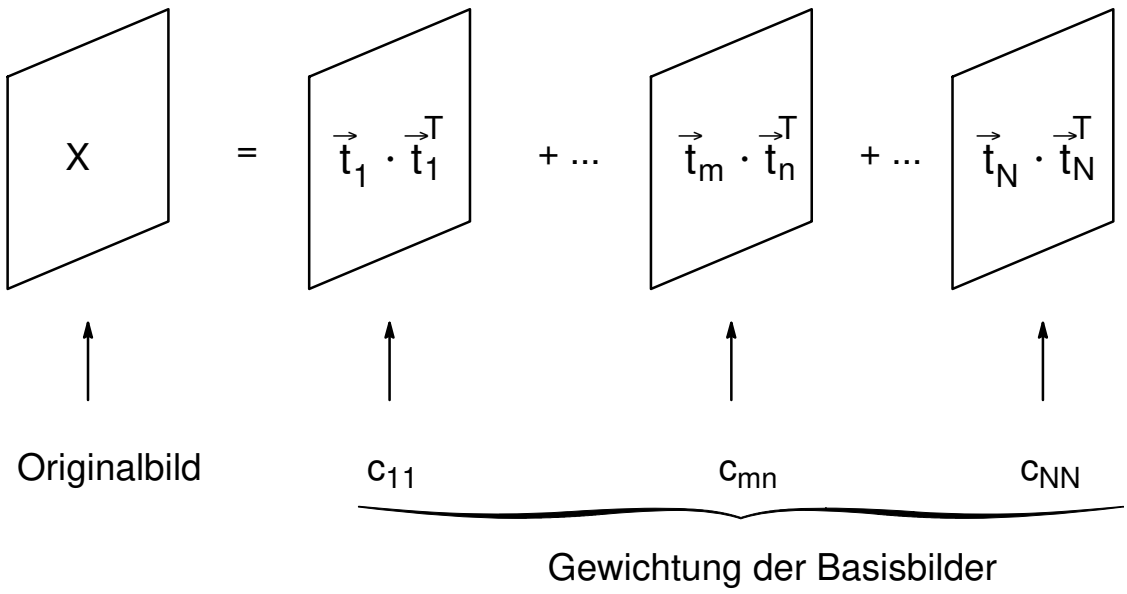


Bild 3.12 Darstellung der Rücktransformation durch eine Linearkombination von Basisbildern

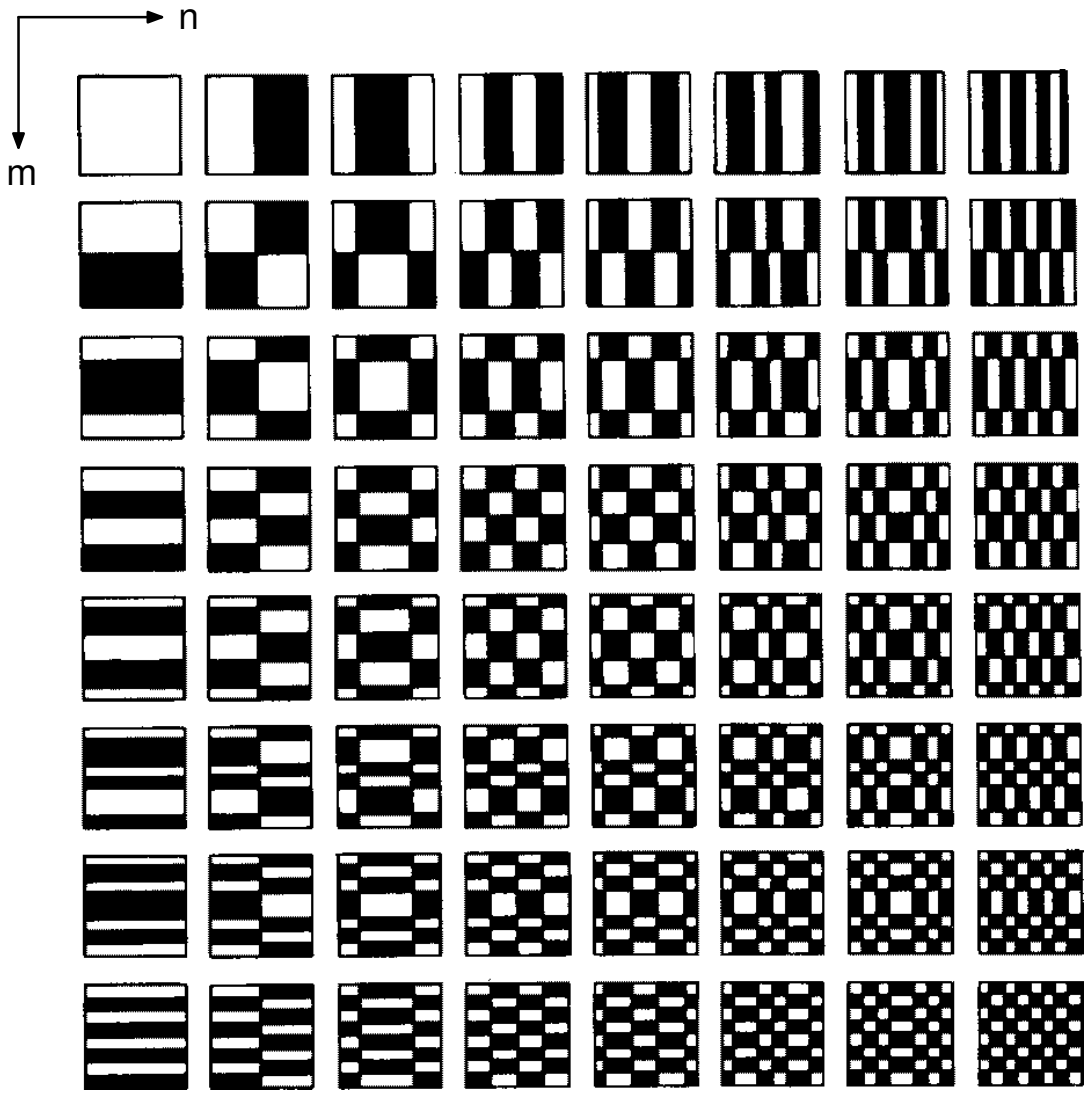


Bild 3.13 Basisbilder einer 8x8 Walsh–Hadamard Transformation

3.3.6 Vergleich von KLT, DFT, DCT, WHT

Für einen Vergleich der behandelten Transformationen soll der erzielbare Signal–Rauschabstand als Funktion der Blockgröße N betrachtet werden.

Die orthonormierten Transformationen DFT, WHT, KLT und DCT erhalten die Energie, d.h.

$$\frac{1}{N}\sum_{n=1}^{N} E\left[x_n^2\right] = \frac{1}{N}\sum_{m=1}^{N} E\left[c_m^2\right] \ . \tag{3.62}$$

Falls nur ein fester Anteil $p < 1$ von Transformationskoeffizienten je Block übertragen und zur Rekonstruktion von $\vec{x}$ verwendet wird, beträgt der mittlere quadratische Fehler (MSE)

$$\begin{aligned} MSE &= \frac{1}{N}\sum_{n=1}^{N} E\Big[(x_n - x_n{}')^2\Big] \\ &= \frac{1}{N}\sum_{m=1}^{N} E\Big[(c_m - c_m{}')^2\Big] \end{aligned} \tag{3.63}$$

wobei

$$c_m{}' = \begin{cases} c_m & 1 \le m \le pN \\ 0 & m > pN \ . \end{cases}$$

Gl. (3.63) berücksichtigt nur das Auslassen von Transformationskoeffizienten, noch nicht deren Quantisierung. Aus Gl. (3.63) folgt

$$MSE = \frac{1}{N}\sum_{m>pN}^{N} E\Big[c_m^2\Big] \ . \tag{3.64}$$

Bei einer 8–bit PCM Darstellung liegt der Wertebereich der Abtastwerte zwischen 0 und 255. Das Verhältnis zwischen dem möglichen Maximalwert der Signalleistung und dem Mittelwert des quadratischen Fehlers ergibt den sogenannten Peak–Signal–to–Noise–Ratio (PSNR) in dB gemäß

$$PSNR = 10\log_{10}\left(\frac{255^2}{MSE}\right)\text{dB} \tag{3.65}$$

Im Fall zweidimensionaler Transformationen ist bei der Berechnung von MSE N durch N^2 zu ersetzen.

Bild 3.14 zeigt den anhand eines Testbildes ermittelten PSNR als Funktion der Blockgröße $N \times N = N^2$ für verschiedene zweidimensionale Transformationen und $p = 0,6$. Die Transformationsgewinne wachsen mit der Blockgröße $N \times N$. Doch ist der Zuwachs für Blöcke größer 8×8 nur noch klein. Im allgemeinen sind die Transformationsgewinne noch von der Bildstatistik abhängig, jedoch zeigt Bild 3.14, daß die Gewinne der DCT sehr dicht bei der KLT liegen. Da die DCT eine separierbare Transformation ist und die Transformation T nicht erst berechnet zu werden braucht, wird in der Bildcodierung die DCT bevorzugt angewendet.

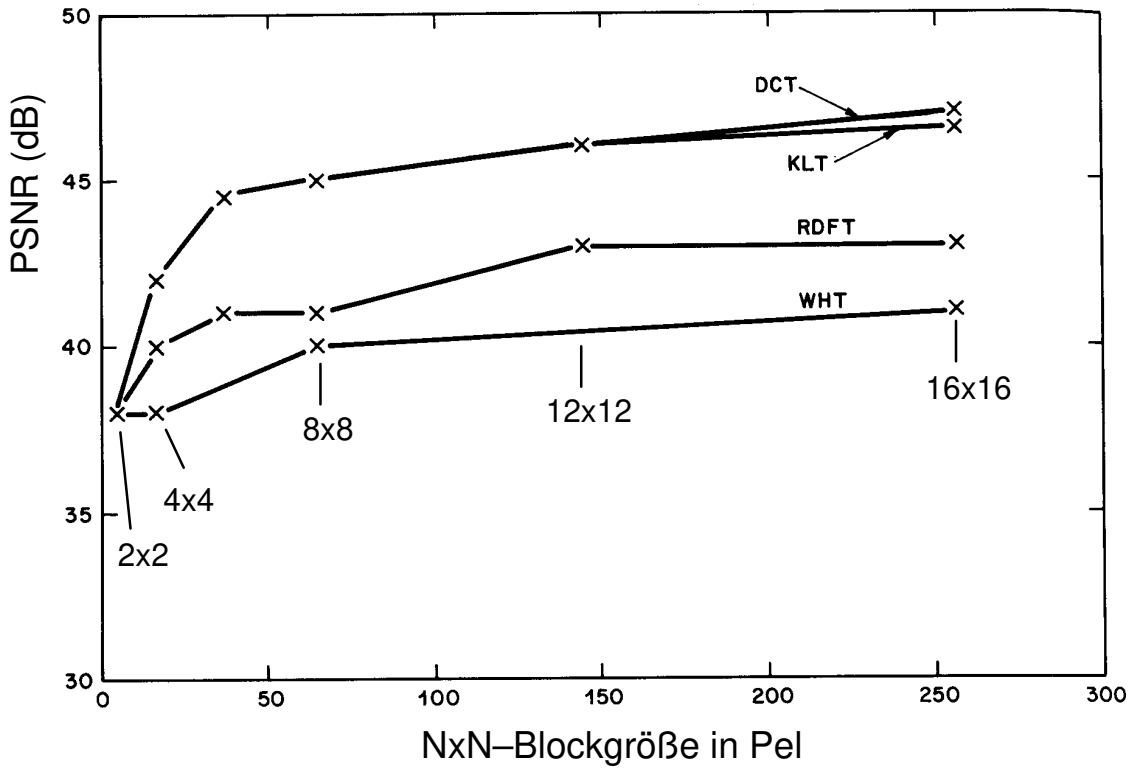


Bild 3.14 PSNR als Funktion der Blockgröße $N \times N$ für $N = 2, 4, 6, 8, 12, 16$ wenn 60% der Koeffizienten übertragen werden ($p = 0.6$)

3.3.7 Bitzuweisung und Codierungsgewinn

Die Transformationskoeffizienten aufeinanderfolgender Blöcke werden als Abtastwerte von N parallelen gedächtnislosen Quellen interpretiert. Die Quelle C_m erzeugt die Koeffizienten c_m mit der Varianz σ_m^2. Infolge einer Quantisierung von c_m möge ein Quantisierungsfehler q_m mit der Varianz $\sigma_{q_m}^2$ auftreten.

Unter der Annahme, daß die Koeffizienten c_m eine Gaußsche Wahrscheinlichkeitsdichte aufweisen, wird entsprechend Gl. (2.145) eine Übertragungsrate von mindestens

$$R_m = \frac{1}{2}\log_2 \frac{\sigma_m^2}{\sigma_{q_m}^2} \tag{3.66}$$

benötigt. Nach den Ergebnissen von Kapitel 2.4.4 müssen die Varianzen der Quantisierungsfehler $\sigma_{q_m}^2$ alle gleich groß gewählt werden, d.h.

$$\sigma_{q_m}^2 = \sigma_q^2 \tag{3.67}$$

damit sich eine minimale mittlere Gesamtdatenrate R

$$R = \frac{1}{N} \sum_{m=1}^{N} R_m \tag{3.68}$$

einstellt. Ist die zur Verfügung stehende Übertragungsrate R vorgegeben, kann die Bitzuweisung R_m für die einzelne Quelle C_m mit Hilfe von Gl. (3.66) und Gl. (3.68) bestimmt werden. Es gilt

$$\begin{aligned} R &= \frac{1}{N} \sum_{m=1}^{N} \frac{1}{2} \log_2 \frac{\sigma_m^2}{\sigma_q^2} = \frac{1}{N} \sum_{m=1}^{N} \frac{1}{2} \log_2 \sigma_m^2 - \frac{1}{2} \log_2 \sigma_q^2 \\ &= \frac{1}{N} \frac{1}{2} \log_2 \left[\prod_{m=1}^{N} \sigma_m^2 \right] - \frac{1}{2} \log_2 \sigma_q^2 \end{aligned} \tag{3.69}$$

und damit

$$\begin{aligned} R_m - R &= \frac{1}{2} \log_2 \sigma_m^2 - \frac{1}{2} \log_2 \left[\prod_{m=1}^{N} \sigma_m^2 \right]^{1/N} \\ R_m &= R + \frac{1}{2} \log_2 \frac{\sigma_m^2}{\left[\prod_{m=1}^{N} \sigma_m^2 \right]^{1/N}} \end{aligned} \tag{3.70}$$

in der Einheit bit je Koeffizient.

Aufgrund von Gl. (3.70) werden einem Transformationskoeffizienten mit großer Varianz σ_m^2 mehr Bits zugewiesen als einem mit kleiner Varianz. Bei Anwendung einer gleichförmigen Quantisierung hat jeder Quantisierer die gleiche Stufenbreite Δ und gleiche Varianz $\sigma_q^2 = \frac{\Delta^2}{12}$. Da 2^{R_m} gemäß Gl. (3.66) bis auf eine Konstante proportional zu σ_m ist, wird einem Koeffizienten mit großer Varianz auch ein Quantisierer mit großem Ansteuerungsbereich zugewiesen, siehe Bild 3.15.

Zur Berechnung des mit der Transformation erzielbaren Gewinns G_{TC} wird wie bei der prädiktiven Codierung das Verhältnis der zu codierenden Signalleistungen herangezogen. Sie beträgt ohne Transformation

$$\sigma_x^2 = \frac{1}{N} \sum_{m=1}^{N} \sigma_m^2 \tag{3.71}$$

und mit der Transformation

$$\sigma_t^2 = \left(\prod_{m=1}^{N} \sigma_m^2 \right)^{1/N} \tag{3.72}$$

wie es Gl. (3.69) zu entnehmen ist. Damit ergibt sich der Gewinn der Transformation G_{TC} zu

$$\begin{aligned} G_{TC} &= 10 \log_{10} \frac{\sigma_x^2}{\sigma_t^2} \\ G_{TC} &= 10 \log_{10} \frac{\sigma_x^2}{\left(\prod_{m=1}^{N} \sigma_m^2 \right)^{1/N}} \end{aligned} \tag{3.73}$$

in dB. Im Vergleich zu PCM ergibt sich bei festem SNR die Datenrate R_{TC} der Transformations–Codierung entsprechend Gl. (3.34) zu

$$\begin{aligned} R_{TC} &= \frac{1}{2} \log_2 \frac{\sigma_t^2}{\sigma_q^2} \\ &= \frac{1}{2} \log_2 \frac{\sigma_x^2}{\sigma_q^2} + \frac{1}{2} \log_2 \frac{\sigma_t^2}{\sigma_x^2} \\ &= R_{PCM} - \frac{1}{2} \log_2 \frac{\sigma_x^2}{\sigma_t^2} \\ &= R_{PCM} - \frac{1}{6,02} \cdot G_{TC} \, . \end{aligned} \tag{3.74}$$

Umgekehrt gilt bei unveränderter Datenrate $R_{TC} = R_{PCM}$ für den Signal–Rauschabstand der Transformationscodierung

$$SNR_{TC} = SNR_{PCM} + G_{TC} \, . \tag{3.75}$$

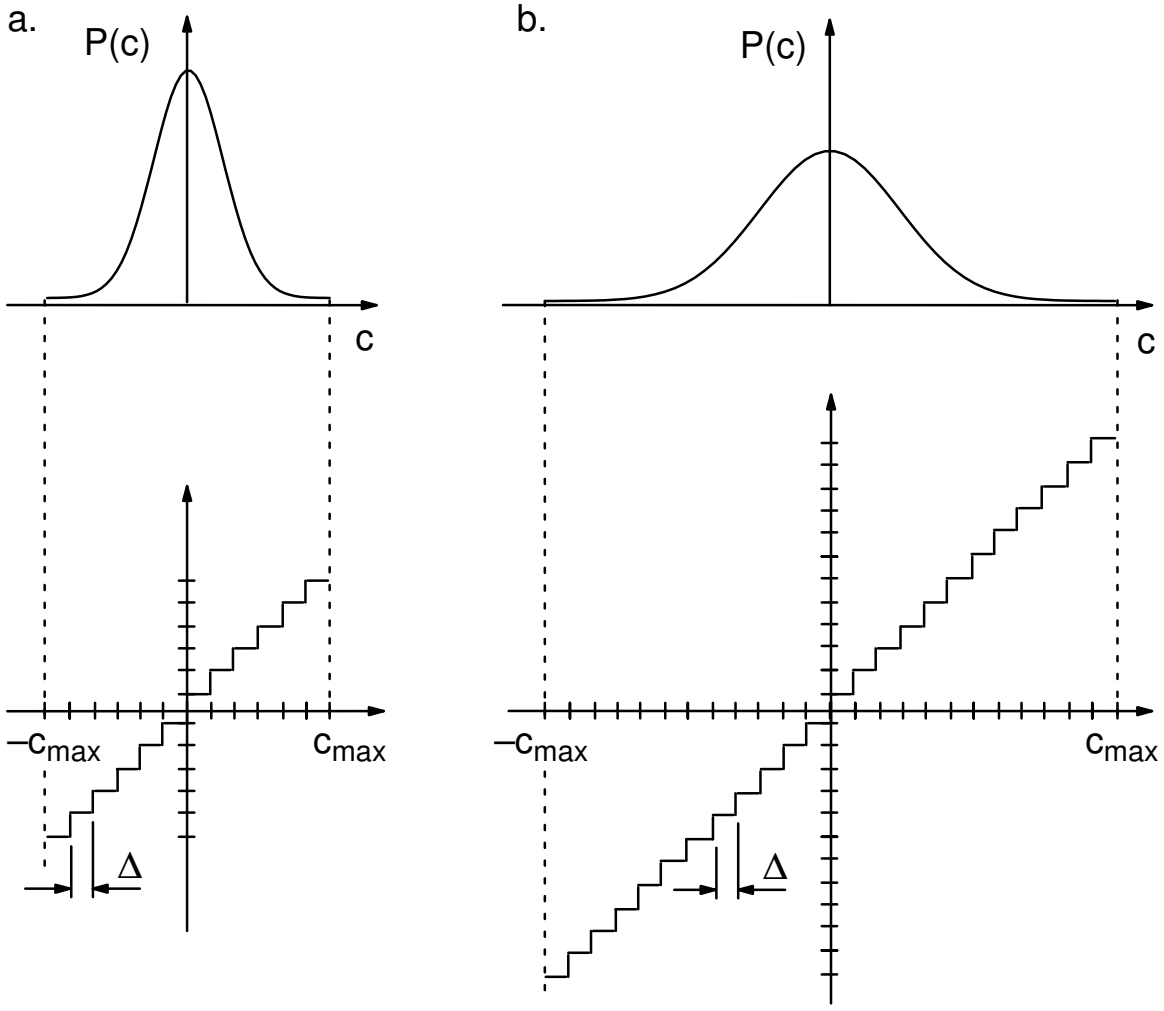


Bild 3.15 Quantisieraussteuerung für Koeffizienten mit kleiner (a) und großer Varianz (b)

3.4 Teilband–Codierung

Bei einer Transformations–Codierung werden die Abtastwerte $x(n)$ mit Hilfe einer Transformation vom Zeitbereich in den Frequenzbereich abgebildet. Diese Abbildung kann auch mittels einer Teilbandzerlegung unter Verwendung einer Filterbank realisiert werden. Eine Teilband–Codierung, wie sie in Bild 3.16 dargestellt ist, verwendet dazu im Coder eine Analysestufe und im Decoder eine Synthesestufe.

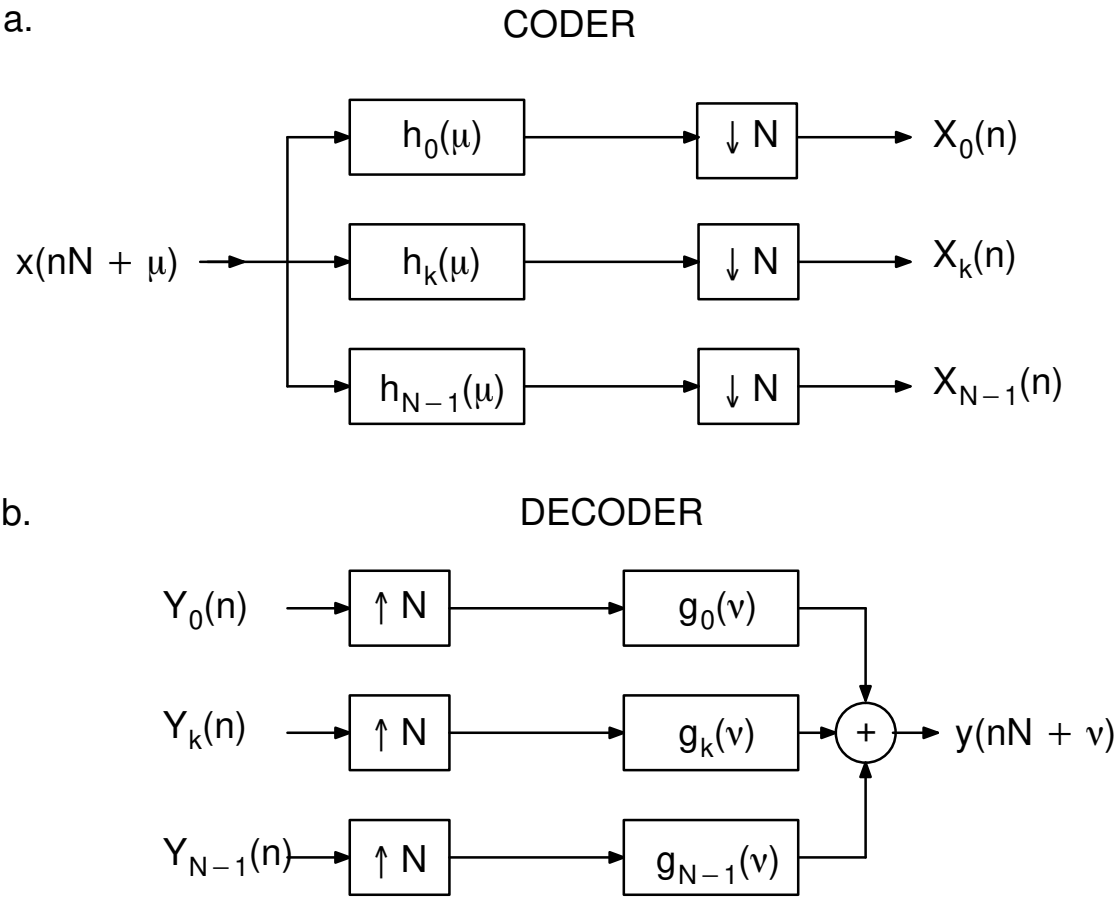


Bild 3.16 Analysestufe (a) und Synthesestufe (b) bei gleichförmiger Bandaufteilung in N Teilbänder und kritischer Abtastung.

Die Analysestufe, bestehend aus einer Filterbank in Verbindung mit einer Reduktion der Abtastfrequenz zerlegt das Eingangsspektrum in Spektralkomponenten $X(n)$. Bei gleichförmiger Bandaufteilung in N Teilbänder gleicher Bandbreite und kritischer Abtastung mit einer um den Faktor N reduzierten Abtastfrequenz, ist die Summe der Abtastfrequenzen der Teilbandsignale gleich der Abtastfrequenz des Eingangssignals. Durch die Reduzierung der Abtastfrequenz können bei Verwendung nicht–idealer Bandpässe Aliasing–Komponenten in den Teilbandsignalen entstehen. Um die Möglichkeit der exakten Rekonstruktion zu erhalten, müssen Filterbanksysteme mit Aliasing–Kompensation angewendet werden. Im Anschluß an die Analysestufe werden die Teilbandsignale quantisiert und in codierter Form übertragen.

Im Decoder müssen die übertragenen, quantisierten Spektralkomponenten $Y(n)$ der Teilbandsignale zunächst decodiert werden. Anschließend werden in der Synthese-

stufe die Abtastfrequenzen der Teilbandsignale wieder um den Faktor N erhöht und am Ausgang der zweiten Filterbank durch additive Überlagerung das Ausgangssignal $y(n)$ erzeugt.

3.4.1 Systeme zur Zwei–Band–Zerlegung

Die Filterbänke zur Zwei–Band–Zerlegung lassen sich durch die Impulsantwort der vier Teilbandfilter in Bild 3.17 beschreiben.

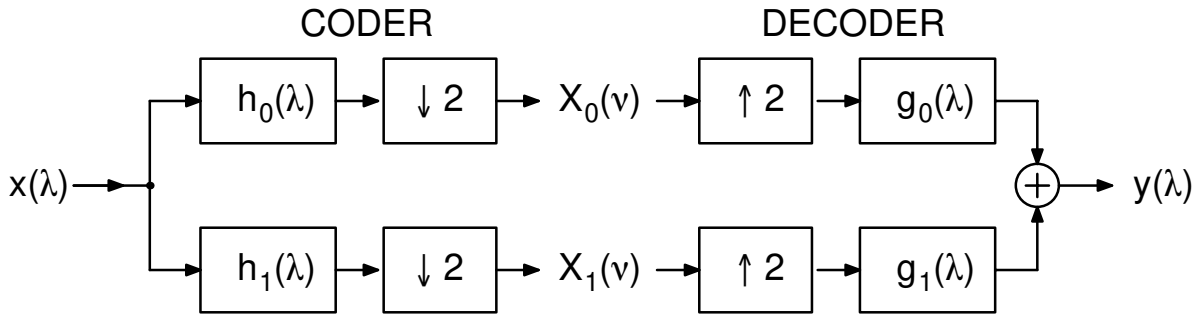


Bild 3.17 Analysestufe und Synthesestufe eines Systems zur Zwei–Band–Zerlegung mit $\nu = \lambda/2$ für gerade λ

Für die Zwei–Band–Zerlegung sollen die Bedingungen für die exakte Rekonstruktion unter Verwendung der Z–Transformation abgeleitet werden. Eine Reduktion der Abtastfrequenz um den Faktor 2 und anschließende Erhöhung der Abtastfrequenz um den Faktor 2 kann gemäß Bild 3.17 zusammengefaßt werden zu einer Multiplikation der Form

$$\{0,\ 1,\ 0,\ 1,\ 0,\ \ldots\} = \frac{1 + (-1)^{\lambda}}{2} \qquad \lambda = 1,\ 2,\ \ldots\ . \tag{3.76}$$

Bei Anwendung der Z–Transformation erhält man mit

$$\begin{aligned} Z\{x(\lambda)\} &= \sum_{\lambda} x(\lambda) z^{-\lambda} = X(z) \\ Z\{(-1)^{\lambda}\, x(\lambda)\} &= \sum_{\lambda} x(\lambda) z^{-\lambda} (-1)^{\lambda} = X(-z) \end{aligned} \tag{3.77}$$

für

$$Z\left\{\frac{1 + (-1)^{\lambda}}{2} x(\lambda)\right\} = \frac{1}{2}[X(z) + X(-z)]\ . \tag{3.78}$$

Die Z–Transformierte des Ausgangssignals $y(\lambda)$ des in Bild 3.18 dargestellten Gesamtsystems beträgt dann

$$Y(z) = \frac{1}{2}\Big[G_0(z) \cdot (H_0(z) \cdot X(z) + H_0(-z) \cdot X(-z)) + G_1(z) \cdot (H_1(z) \cdot X(z) + H_1(-z) \cdot X(-z))\Big] \tag{3.79}$$

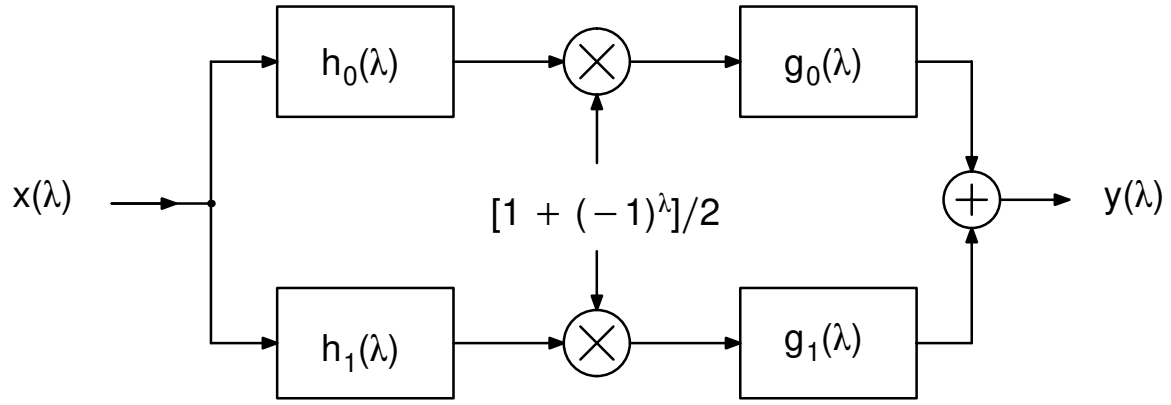


Bild 3.18 Zusammengefaßtes Ersatzschaltbild von Analyse– und Synthese–stufe

Dabei stellen die Terme $X(-z)$ Aliasing–Komponenten dar. Die Trennung dieser Komponenten ergibt

$$Y(z) = \frac{1}{2}\Big[G_0(z) \cdot H_0(z) + G_1(z) \cdot H_1(z)\Big] X(z) + \frac{1}{2}\Big[G_0(z) \cdot H_0(-z) + G_1(z) \cdot H_1(-z)\Big] X(-z)\,. \tag{3.80}$$

Um die Bedingung der exakten Konstruktion zu erfüllen, muß bei Teilbandfiltern der Länge L das Ausgangsignal $y(\lambda)$ bis auf eine Verzögerung von $L-1$ Abtastwerten gleich dem Eingangssignal $x(\lambda)$ sein, d.h.

$$G_0(z) \cdot H_0(z) + G_1(z) \cdot H_1(z) = 2z^{-(L-1)}\,. \tag{3.81}$$

Ferner müssen sich die Aliasing–Komponenten kompensieren, d.h.

$$G_0(z) \cdot H_0(-z) + G_1(z) \cdot H_1(-z) = 0. \tag{3.82}$$

Gl.(3.81) und Gl.(3.82) lassen sich zu einem Gleichungssystem zusammenfassen

$$\begin{bmatrix} H_0(z) & H_1(z) \\ H_0(-z) & H_1(-z) \end{bmatrix} \begin{bmatrix} G_0(z) \\ G_1(z) \end{bmatrix} = \begin{bmatrix} 2z^{-(L-1)} \\ 0 \end{bmatrix} \tag{3.83}$$

mit dem die Synthese– und Analysefilter bestimmt werden können.
Die Lösung des Gleichungssystems lautet

$$\begin{aligned} G_0(z) &= \frac{2z^{-(L-1)} \cdot H_1(-z)}{D(z)} \\ G_1(z) &= -\frac{2z^{-(L-1)} \cdot H_0(-z)}{D(z)} \end{aligned} \tag{3.84}$$

mit $$D(z) = H_0(z) \cdot H_1(-z) - H_0(-z) \cdot H_1(z) \ .$$

Bei Impulsantworten g_0 und g_1 endlicher Länge L kann der Nennerterm in Gl. (3.84) nur eine einfache Potenz von z multipliziert mit einer Konstanten c sein, d.h. $D(z)$ ist ein Verzögerungsglied mit der Verstärkung c.
Da $D(z)$ andererseits ein Filter darstellt, das durch Faltung zweier Impulsantworten der Länge L entsteht, muß diese Verzögerung $L-1$ sein. Die Bedingungen für exakte Rekonstruktionen können somit auch in der Form

$$\begin{aligned} G_0(z) &= \frac{2}{c} H_1(-z) \\ G_1(z) &= -\frac{2}{c} H_0(-z) \\ H_0(z) \cdot H_1(-z) - H_0(-z) \cdot H_1(z) &= c\, z^{-(L-1)} \end{aligned} \tag{3.85}$$

geschrieben werden. Durch inverse Z–Transformation und Substitution von $F(z) = H_0(z) \cdot H_1(-z)$ sowie $F(-z) = H_0(-z) \cdot H_1(z)$ erhält man die Bedingungen für die Impulsantworten der Teilbandfilter

$$\begin{aligned} g_0(\lambda) &= (-1)^\lambda \cdot \frac{2}{c} h_1(\lambda) \\ g_1(\lambda) &= -(-1)^\lambda \cdot \frac{2}{c} h_0(\lambda) \end{aligned}$$

mit

$$\begin{aligned} f(\lambda)\,(1-(-1)^\lambda) &= c\,\delta\,(\lambda - L + 1) \\ f(\lambda) &= h_0(\lambda) * ((-1)^\lambda \cdot h_1(\lambda)) \end{aligned} \tag{3.86}$$

Damit die Gleichung (3.86) für $\lambda = L-1$ erfüllt werden kann, muß λ ungerade bzw. die Filterlänge L gerade sein.
Durch Vorgabe weiterer Bedingungen für die Teilbandfilter ergeben sich verschiedene Klassen von Filterbanken zur Zwei–Band Zerlegung wie z.B. Quadraturspiegelfilter (QMF) oder konjugierte Quadraturfilter (CQF).

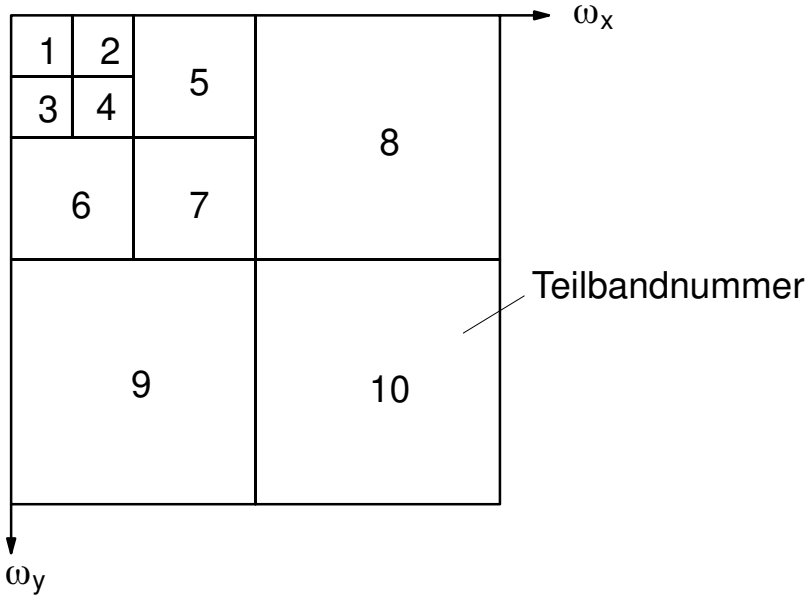


Bild 3.19 Aufteilung der zweidimensionalen Frequenzebene von Bildsignalen in 10 Teilbandsignale mit einer sog. Wavelet–Teilbandzerlegung

Um eine möglichst große Redundanzreduktion zu erreichen, sind mehr als nur zwei Teilbandsignale erforderlich. Die Anzahl der Teilbandsignale kann durch Kaskadierung der in Bild 3.17 gezeigten Filterbänke erhöht werden. Für die Codierung von Bildsignalen können bei dieser Kaskadierung neben horizontalen Filterungen auch vertikale Filterungen verwendet werden, um eine zweidimensionale Aufteilung der Frequenzebene zu erreichen. Dabei sind auch ungleichmäßige Aufteilungen ähnlich Bild 3.19 möglich.

3.4.2 Äquivalenzbeziehungen zwischen Teilbandzerlegung und Transformation

Im folgenden wird gezeigt, daß Teilbandzerlegung und Transformation äquivalente Operationen darstellen, wenn die Matrixelemente der Transformation und die Filterkoeffizienten der Teilbandzerlegung bestimmten Bedingungen genügen. Zu diesem Zweck wird eine Abbildung von L Abtastwerten in N Spektralkoeffizienten eingeführt, die als LINC–Darstellung (L–Into–N–Coding) bezeichnet wird [3].

Dazu wird zunächst die allgemeine Form der Transformation mit überlappenden Blöcken betrachtet. Bei der Hintransformation wird im n–ten Transformationsschritt ein Eingangsblock von $L = P \cdot N$ aufeinanderfolgenden Abtastwerten $X(n\,N - L + \lambda)$ zu den Zeitpunkten $n\,N - L + \lambda$ mit $0 \leq \lambda \leq L - 1$ durch Multi-

plikation mit der Transformationsmatrix $A = \{a_{k\lambda}\}$ der Größe $N \times L$ abgebildet in einen Block von N Spektralkoeffizienten $\{X_k(n)\}$

$$X_k(n) = \sum_{\lambda=0}^{L-1} a_{k,\lambda} \cdot x\,(nN - L + \lambda) \qquad (3.87)$$
$$0 \le k \le N - 1$$

Die Eingangsblöcke aufeinanderfolgender Transformationsschritte sind jeweils um N Abtastwerte gegeneinander versetzt, so daß die Anzahl der je Transformationsschritt erzeugten Spektralkoeffizienten gleich der Anzahl der neu hinzugekommenen Eingangswerte ist, s. Bild 3.20.

Dadurch entsteht eine Transformation mit P überlappenden Eingangsblöcken, d.h. jeder Abtastwert des Eingangssignals ist in den Eingangsblöcken von P aufeinanderfolgenden Transformationsschritten enthalten.

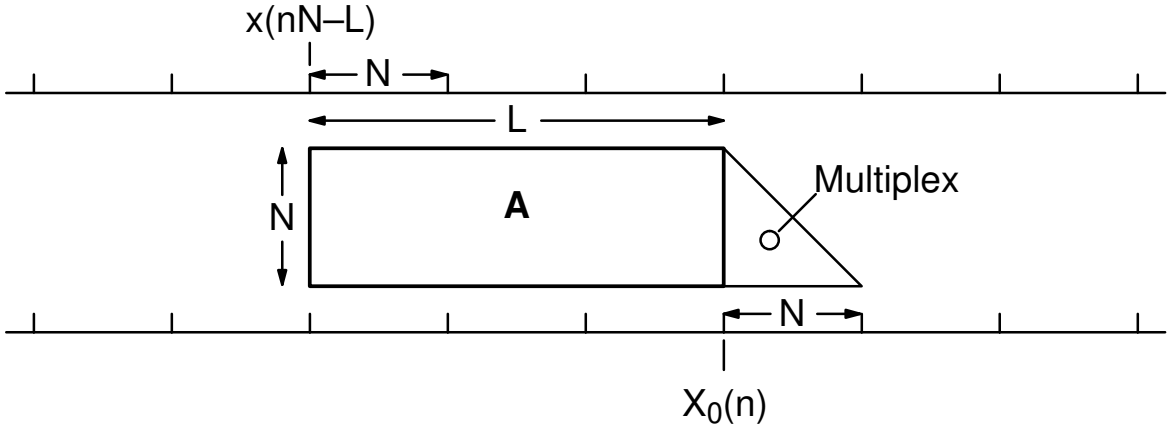


Bild 3.20 Transformation mit *P=3* und Multiplex für die Spektralkoeffizienten

Bei der Rücktransformation werden N Spektralkoeffizienten $\{Y_k(n)\}$ durch Multiplikation mit der Transponierten der Matrix $B = \{b_{k\lambda}\}$ der Größe $N \times L$ in einen Block von L Zwischenwerten abgebildet, s. Bild 3.21.

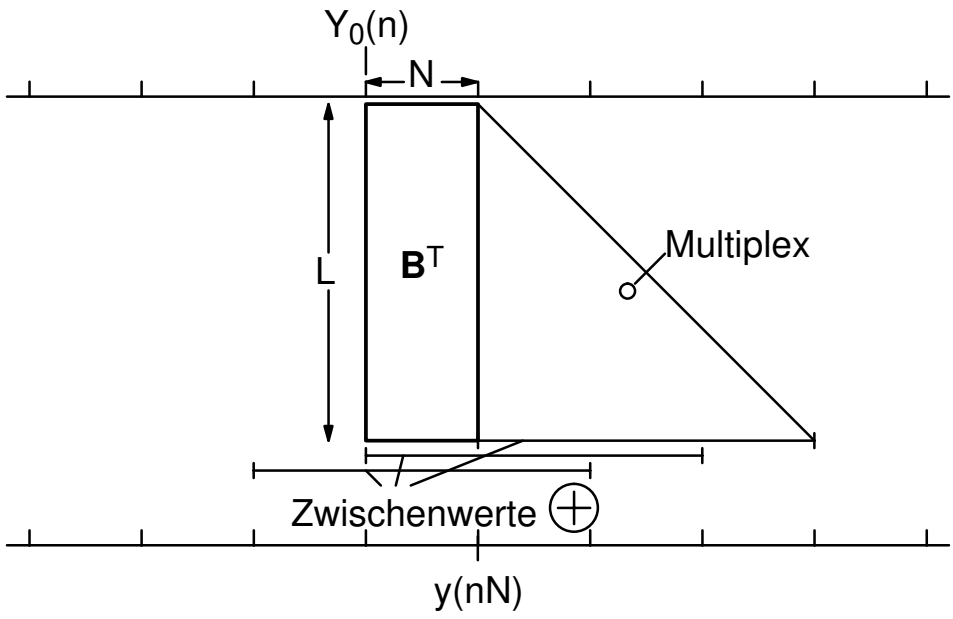


Bild 3.21 Rücktransformation mit *P=3*

Durch Überlagerung und Addition (Overlap–Add) von P um jeweils N Werte versetzter Zwischenblöcke erhält man in jedem Transformationsschritt einen Ausgangsblock von N rekonstruierten Abtastwerten y zu den Zeitpunkten $nN + \nu$

$$y(nN + \nu) = \sum_{p=0}^{P-1} \sum_{k=0}^{N-1} b_{k,pN+\nu} \cdot Y_k(n - p) \qquad (3.88)$$

$$0 \leq \nu \leq N - 1 \, .$$

Dabei müssen die Matritzen A und B^T die Bedingungen zur exakten Rekonstruktion erfüllen.

Demgegenüber erzeugt die Analysestufe einer Teilbandzerlegung mit gleichförmiger Bandaufteilung in N Teilbänder und Teilbandfiltern mit Impulsantwortfunktionen $h_k(\mu)$ der Längen L gemäß Bild 3.16 a Spektralkoeffizienten.

$$X_k(n) = \sum_{\mu=0}^{L-1} h_k(\mu) \cdot X(nN - 1 - \mu) \qquad (3.89)$$

$$0 \leq k \leq N - 1 \, .$$

Das Ausgangssignal der Synthesestufe ergibt sich gemäß Bild 3.16 b zu

$$y(nN+\nu) = \sum_{k=0}^{N-1} \sum_{p=0}^{P-1} g_{k,PN+\nu}\, Y_k(n-p) \qquad 0 \le \nu \le N-1\,. \tag{3.90}$$

Vergleicht man die Spektralkoeffizienten der Transformation Gl. (3.87) mit denen der Teilbandzerlegung Gl. (3.89) und setzt zu diesem Zweck $\lambda = L-1-\mu$ in Gl. (3.87) ein, so erhält man

$$X_k(n) = \sum_{\mu=0}^{L-1} a_{k,L-1-\mu} \cdot x\,(nN-1-\mu)\,. \tag{3.91}$$

Man erkennt, daß bei den Operationen bei gleichem Eingangssignal $X(\mu)$ gleiche Spektralkoeffizienten liefern, wenn

$$h_k(\mu) = a_{k,L-1-\mu} \qquad 0 \le \mu \le L-1 \tag{3.92}$$

gilt. Demzufolge stellen Transformation und Teilbandzerlegung äquivalente Operationen dar, wenn jede Teilbandfilter–Impulsantwort identisch ist mit einer gespiegelten Zeile der Transformationsmatrix A. Entsprechendes gilt für die Rücktransformation und die Synthesestufe einer Teilbandzerlegung, wenn

$$g_k(\mu) = b_{k,\mu} \qquad 0 \le \mu \le L-1 \tag{3.93}$$

gilt, d.h. wenn jede Teilbandfilter–Impulsantwort einer Zeile der Matrix B entspricht.

3.4.3 Bitzuweisung und Codierungsgewinn

Die Bitzuweisung für die Spektralkoeffizienten der Teilbandzerlegung kann in gleicher Weise begründet werden wie die der Transformation im Abschnitt 3.3.7. Demzufolge erhält der Spektralkoeffizient X_k entsprechend Gl. (3.70) die Datenrate

$$R_k = R + \frac{1}{2}\log_2 \frac{\sigma_k^2}{\left[\prod_{k=0}^{N-1} \sigma_k^2\right]^{1/N}} \tag{3.94}$$

in der Einheit bit je Koeffizient.
Dabei berechnet sich die Varianz σ_k^2 gemäß

$$\sigma_k^2 = \frac{1}{2\pi} \int_{-\pi}^{+\pi} | H_k(\omega) |^2 \cdot S_{xx}(\omega)\, d\omega \tag{3.95}$$

wobei $|H_k(\omega)|$ der Betragsfrequenzgang des k–ten Analysefilters und $S_{xx}(\omega)$ das Leistungsdichtespektrum des Eingangssignals ist.

Unter der Annahme, daß die Analyse– und die Synthesestufe jeweils die Energie erhalten, ergibt sich der Codierungsgewinn entsprechend Gl. (3.73)

$$G_{TBC} = 10 \log_{10} \frac{\sigma_x^2}{\left[\prod_{k=0}^{N-1} \sigma_k^2\right]^{1/N}} \tag{3.96}$$

in dB, wobei

$$\sigma_x^2 = \frac{1}{N} \sum_{k=0}^{N-1} \sigma_k^2 \tag{3.97}$$

Die Codierungsgewinne einer Teilbandzerlegung sind größer als die einer Transformation mit nicht überlappenden Blöcken, da die Korrelation von Block zu Block ausgenutzt wird. Bild 3.22 zeigt die maximalen Gewinne einer $N \times N$ Karhunen–Loeve–Transformation und einer $N \times L$ Teilbandzerlegung mit unendlich langen Impulsantwortfunktionen $L \rightarrow \infty$ für einen AR(1) – Prozeß mit $\varrho = 0,95$ bzw. $\varrho = 0,6$. Bei einer Blockgröße von $N = 8$ beträgt der Unterschied etwa 1,0 dB bzw. 0,25 dB.

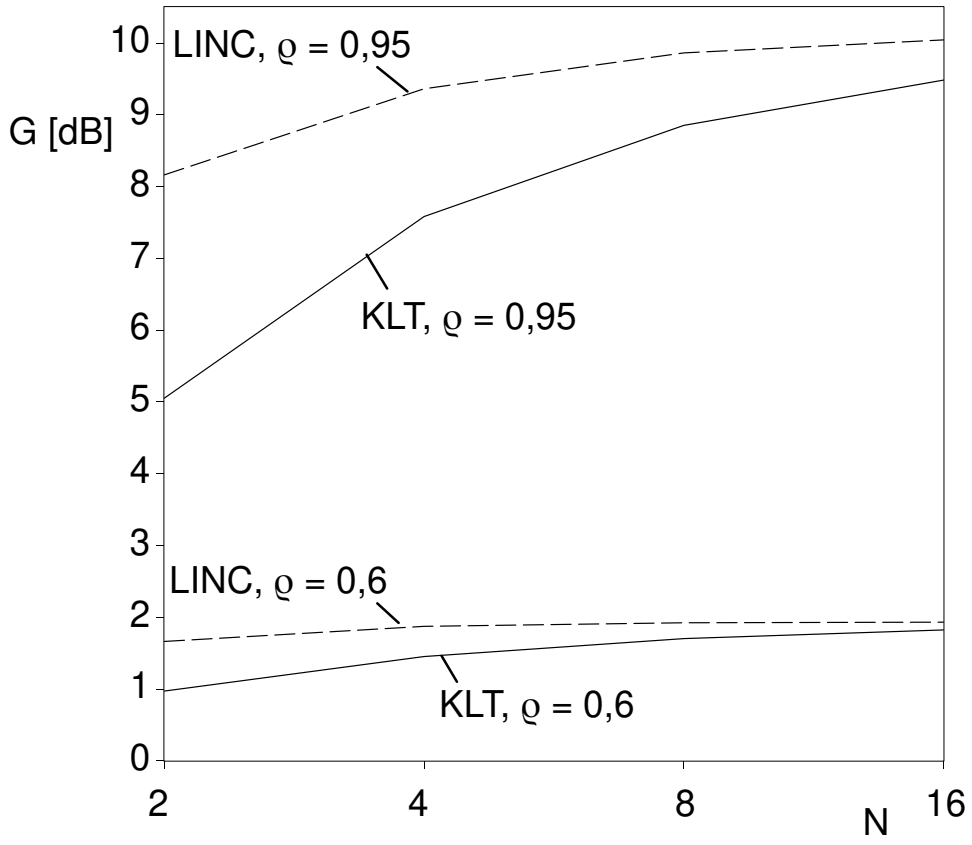


Bild 3.22 Theoretische Grenzen der Gewinne in dB der Karhunen–Loeve–Transformation (KLT) und Teilbandzerlegung (LINC) mit $L \rightarrow \infty$ bei einem AR(1)–Prozeß mit $\varrho = 0{,}95$ bzw. $\varrho = 0{,}6$

3.5 Bewegungskompensierte Prädiktion

Bei der Codierung von Bewegtbildsequenzen kann auch die Bild–zu–Bild–Korrelation zur Prädiktion ausgenutzt werden, wenn die translatorische Bewegung der Objekte in der Bildszene bekannt ist. Dazu wird eine DPCM entsprechend Abschnitt 3.2.2 mit einem Prädiktor angewendet, der Bildpunkte des vorangehenden Bildes verwendet, die entsprechend der translatorischen Bewegung versetzt sind. Zu diesem Zweck wird das aktuelle zu codierende Bild k in Blöcke von beispielsweise $N \times N$ Bildpunkte unterteilt, siehe Bild 3.23.

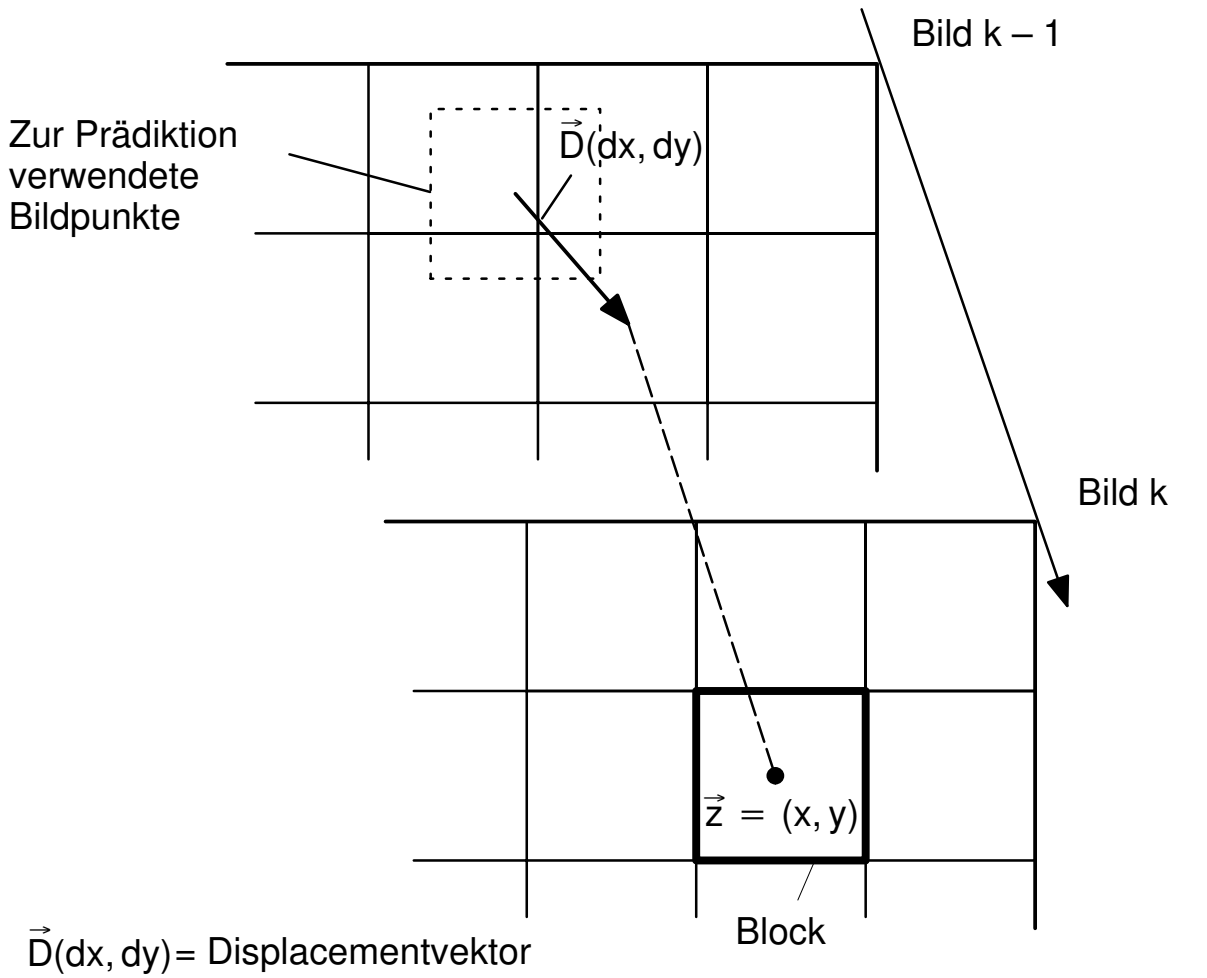


Bild 3.23 Darstellung einer bewegungskompensierten Prädiktion

Für jeden Block wird der sogenannte Displacementvektor $\vec{D}(dx, dy)$ gemessen und zum Empfänger übertragen. Mit Hilfe des Displacementvektors $\vec{D}$ kann nun im Coder wie auch im Decoder eine sog. bewegungskompensierte Prädiktion ausgeführt werden, indem die Luminanz s_k eines Bildpunktes an der Stelle $\vec{z} = (x, y)$ im Bild k mit Hilfe der Luminanz s_{k-1} des Bildpunktes an der Stelle $\vec{z} - \vec{D}$ im Bild $k-1$ vorhergesagt wird, d.h.

$$\hat{s}_k(\vec{z}) = s_{k-1}(\vec{z} - \vec{D}) \quad . \tag{3.98}$$

Es wird der Prädiktionsfehler $e(\vec{z})$

$$\begin{aligned} e(\vec{z}) &= s_k(\vec{z}) - \hat{s}_k(\vec{z}) \\ e(\vec{z}) &= s_k(\vec{z}) - \hat{s}_{k-1}(\vec{z} - \vec{D}) \end{aligned} \tag{3.99}$$

nach einer Quantisierung und Codierung übertragen.

Bei der Bestimmung des Displacementvektors $\vec{D}$ eines Blockes wie auch bei der Prädiktion geht man von der Annahme aus, daß alle Bildpunkte eines Blockes die gleiche translatorische Bewegung haben. Einflüsse von Kamera – Zoom, Objekt – Rotation, Verdeckung von Objekten oder freiwerdender Hintergrund bleiben unberücksichtigt.

3.5.1 Verfahren zur Bewegungsschätzung

Die Displacementvektoren $\vec{D}$ können mit Hilfe sogenannter Blockmatching – Verfahren oder Rekursiven Verfahren geschätzt werden.

Blockmatching – Verfahren

Um das Displacement für einen Bildpunkt $\vec{z}(x = i,\ y = j)$ in Bild k zu bestimmen, wird ein Block von $N \times N$ Bildpunkten mit dem Mittelpunkt $(x = i,\ y = j)$ gebildet. Die Bildinformation sei durch die Luminanzwerte $s_k(m,\ n)$ beschrieben. Unter der Annahme, daß ein maximales Displacement d_m in horizontaler und vertikaler Richtung auftritt, wird in einem Suchbereich SR von

$$SR = (N + 2d_m) \times (N + 2d_m) \tag{3.100}$$

Bildpunkten im Bild $k-1$ unter Verwendung der Bildinformation $s_k(m,\ n)$ des Blockes aus Bild k die Position maximaler Korrelation gesucht, siehe Bild 3.24. Dieses Verfahren wird daher als "Full Search" bezeichnet. Anstelle der Korrelation wird zur Vereinfachung auch häufig für jede Position $x,\ y$ im Suchbereich der Mittelwert der Beträge der Luminanzdifferenzen oder die sog. Mean Absolute Difference (MAD) ausgewertet.

Für die Position $\vec{z}(x,\ y)$ berechnet sich die Mean Absolute Difference gemäß

$$MAD(x,\ y) = \frac{1}{N^2} \sum_{m \geq N/2}^{+N/2} \sum_{n \geq N/2}^{+N/2} \left| s_k(m,\ n) - s_{k-1}(m + x,\ n + y) \right| . \tag{3.101}$$

Im Fall des sog. "Full Search" wird die MAD für alle Positionen $\vec{z}(x,\ y)$ mit

$$-d_m \leq i,\ j \leq +d_m \tag{3.102}$$

bestimmt.

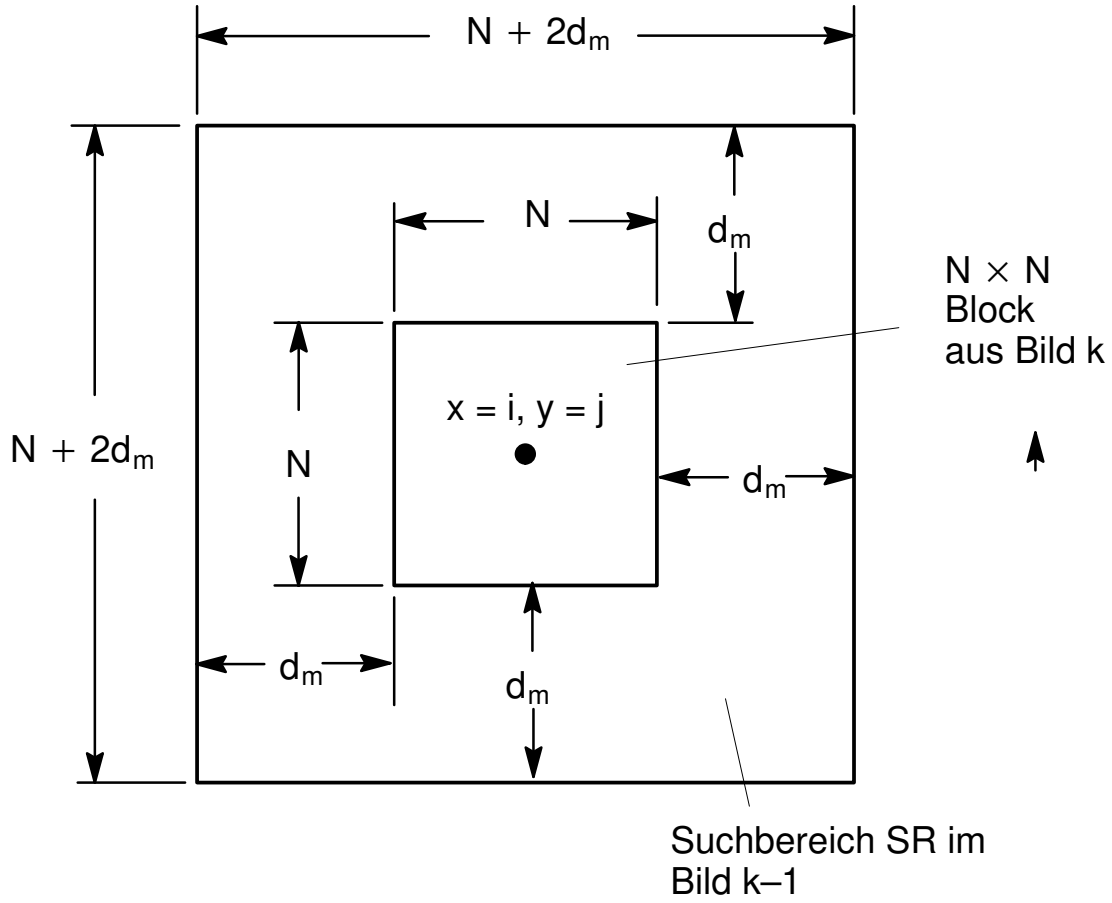


Bild 3.24 Suchbereich SR im Bild *k–1* für Block $N \times N$ aus Bild *k*

Das Berechnungskriterium nach Gl.(3.101) hat den Vorteil, daß keine Multiplikation ausgeführt werden muß. Um darüberhinaus die Anzahl der Rechenoperationen zu reduzieren, werden verschiedene Suchverfahren vorgeschlagen. Als Beispiel soll das sog. "Three Step Search" Verfahren erläutert werden, s. Bild 3.25.

Beim Three Step Search werden im ersten Schritt die Position (i, j) und acht weitere Positionen getestet, die relativ weit auseinander liegen und in Bild 3.25 durch eine "1" gekennzeichnet sind. Die Position $(i + 3,\ j + 3)$ ergibt die minimale MAD. Dann werden im zweiten Schritt acht weitere Positionen getestet, die um das Minimum des ersten Schrittes nun in engerem Abstand angeordnet sind. Sie sind in Bild 3.25 durch eine "2" gekennzeichnet. Als neues Minimum der MAD ergibt sich die Position $(i + 3,\ j + 5)$. Im dritten Schritt werden die mit "3" gekennzeichneten Positionen getestet. Im dargestellten Beispiel hat der gefundene Displacementvektor die Komponenten $\vec{D}(dx = 2,\ dy = 6)$.

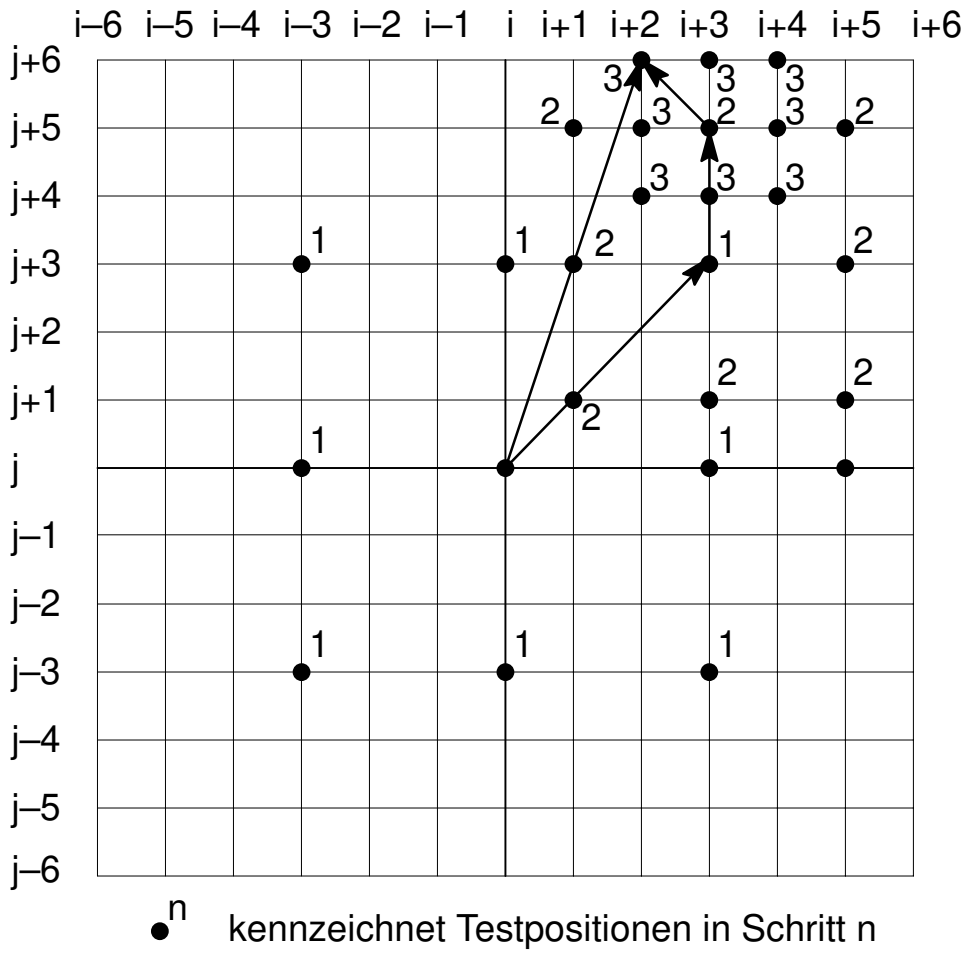


Bild 3.25 Three Step Search Verfahren

Um bei großen Displacements den Displacementvektor zuverlässig zu bestimmen, müssen große Blöcke $N \times N$ zur Suche verwendet werden. Dann ist die Wahrscheinlichkeit einer mehrfachen zufälligen Übereinstimmung von Blockinhalt mit Bildinhalten im Suchbereich gering. Große Blöcke haben andererseits den Nachteil, daß die Displacementvektoren über große Bereiche konstant sein müssen. Um diesen Nachteil zu vermeiden, sind "Hierarchische Displacement–Schätzverfahren" vorgeschlagen worden. Sie verwenden im ersten Schritt große Blöcke und reduzieren die Blockgröße in den nachfolgenden Schritten.

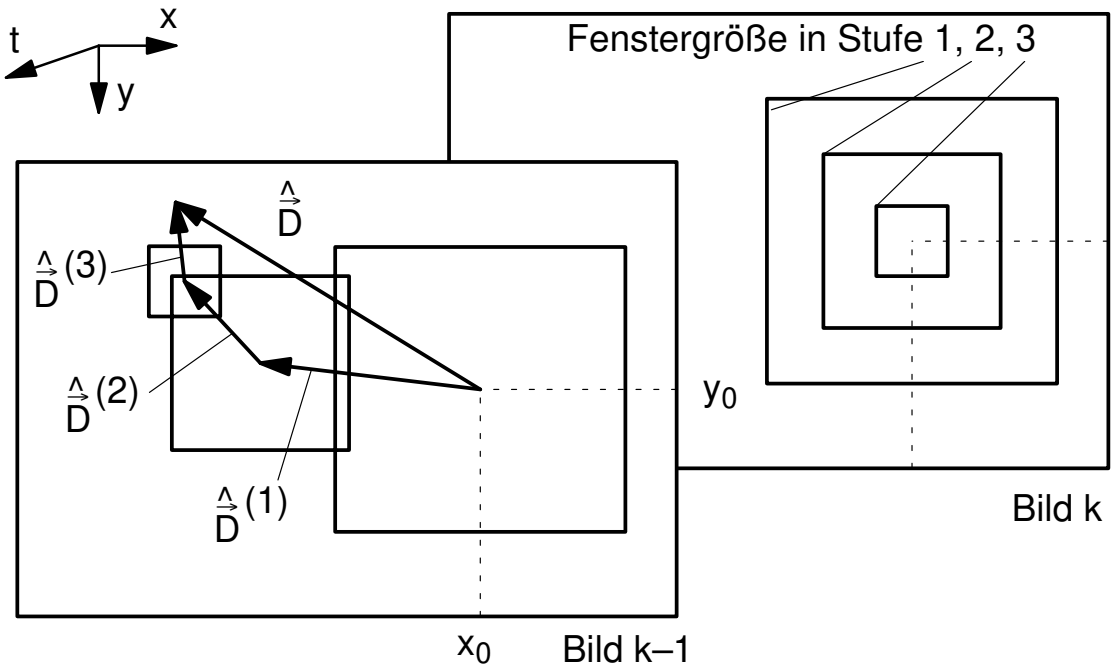


Bild 3.26 Prinzip der hierarchischen Displacementschätzung für ein Beispiel mit drei Hierarchiestufen

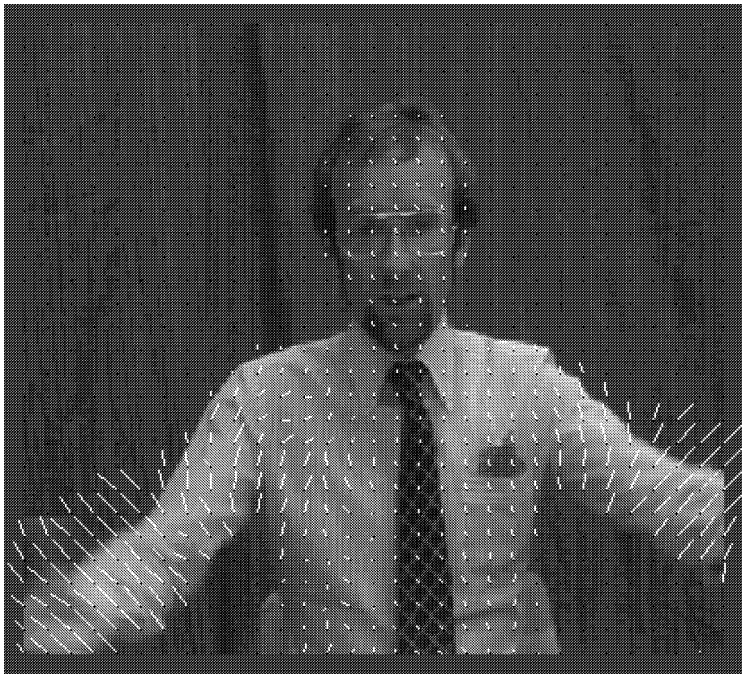
Bild 3.26 veranschaulicht diese Funktionsweise der hierarchischen Suchverfahren. Dabei wird mit jeder Stufe auch eine Tiefpaßfilterung und Reduzierung der Abtastfrequenz vorgenommen, damit die Anzahl der zu korrelierenden Bildpunkte je Block in jeder Stufe konstant ist, siehe Bild 3.27.

Hierarchiestufe	1	2
Maximaldisplacement	± 7	± 3
Meßfenstergröße	64	16
Schrittweite	32	16
Fenstergröße des Mittelwertfilters	5	3
Unterabtastung	8	2

Bild 3.27 Parameter eines vereinfachten hierarchischen Suchverfahrens mit zwei Hierarchiestufen

In Bild 3.28 sind die mit einem Three Step Search und einem hierarchischen Schätzverfahren bestimmten Displacementvektorfelder zum Vergleich gegenübergestellt. Es wird erkennbar, daß das hierarchische Schätzverfahren homogenere Vektorfelder liefert, die besser der wahren Bewegung entsprechen.

a.



b.

Bild 3.28 Vergleich der Displacement–Vektorfelder hierarchischer (a) und nichthierarchischer (b) Displacementschätzung

Rekursive Schätzverfahren

Die rekursiven Displacementschätzverfahren nutzen die durch translatorische Bewegung verursachten örtlichen und zeitlichen Bildsignaländerungen zur Schätzung des Displacementvektors aus. Bild 3.29 veranschaulicht die Zusammenhänge zwischen zeitlicher Bildsignaländerung $FD(x,\ y)$,

$$FD(x,\ y) = s_k(x,\ y) - s_{k-1}(x,\ y) \tag{3.103}$$

örtlicher Bildsignaländerung $\frac{\partial s}{\partial x}$

$$\frac{\partial s_k(x,\ y)}{\partial x} = \frac{1}{2}\big(s_k\,(x + 1,\ y) - s_k\,(x - 1,\ y)\big) \tag{3.104}$$

und Displacement $D(\vec{z})$ für den Fall eines eindimensionalen Displacements $\vec{D} = dx$. Einen einfachen Schätzwert $\hat{D}$ erhält man mit

$$\hat{D} = -\frac{FD(x,\ y)}{\partial s_k(x,\ y)/\partial x}\ . \tag{3.105}$$

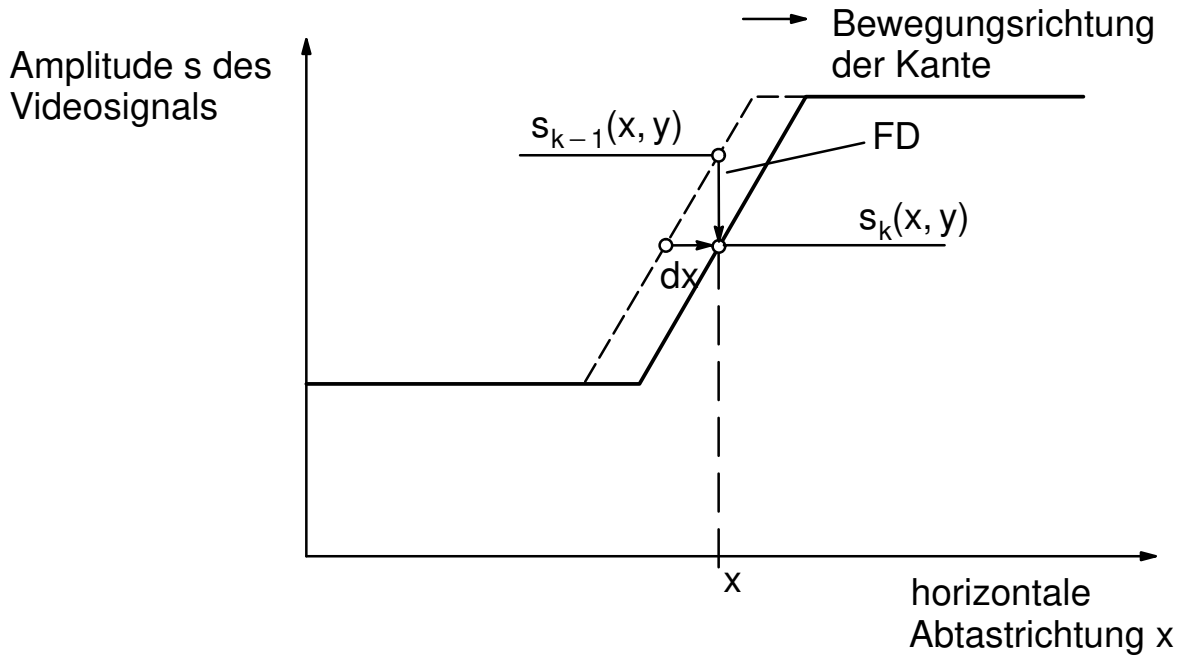


Bild 3.29 Darstellung des Zusammenhanges zwischen Displacement dx, Frame–Difference FD und Gradienten $\partial s/\partial x$

Die rekursiven Schätzverfahren gehen davon aus, daß für einen Bildpunkt ein erster Schätzwert als Startwert vorliegt, der durch Iteration verbessert wird. Als Startwert

kann beispielsweise der Displacementvektor eines benachbarten Bildpunktes oder ein Schätzwert nach Gl.(3.105) verwendet werden. Der im i–ten Iterationsschritt berechnete Displacementvektor $\hat{D}_i$ wird im (i+1)–ten Iterationsschritt durch einen sog. Update Term U_i verbessert, so daß

$$\hat{D}_{i+1} = \hat{D}_i + U_i \ . \tag{3.106}$$

Diese Displacementschätzung wird für jeden Bildpunkt eines bewegten Objektes ausgeführt. Zur Bestimmung des Update Term U_i wird das Displacement zwischen Bildinhalt k und Bildinhalt $k-1$ kompensiert, d.h. die sog. Displaced–Frame–Difference (DFD)

$$DFD(x,\ y,\ \hat{D}_i) = s_k\,(x,\ y) - s_{k-1}(x - \hat{d}x_i,\ y - \hat{d}y_i) \tag{3.107}$$

gebildet und deren Quadrat $[DFD(x,\ y,\ \hat{D}_i)]^2$ als Kostenfunktion des Schätzverfahrens verwendet. Bei Erreichen des wahren Displacements $\hat{D}$ und exakter Kompensation wird die DFD Null. Im praktischen Fall wird die Kostenfunktion $[DFD(x,\ y,\ \hat{D}_i)]^2$ ein Minimum erreichen, wie es in Bild 3.30 dargestellt ist.

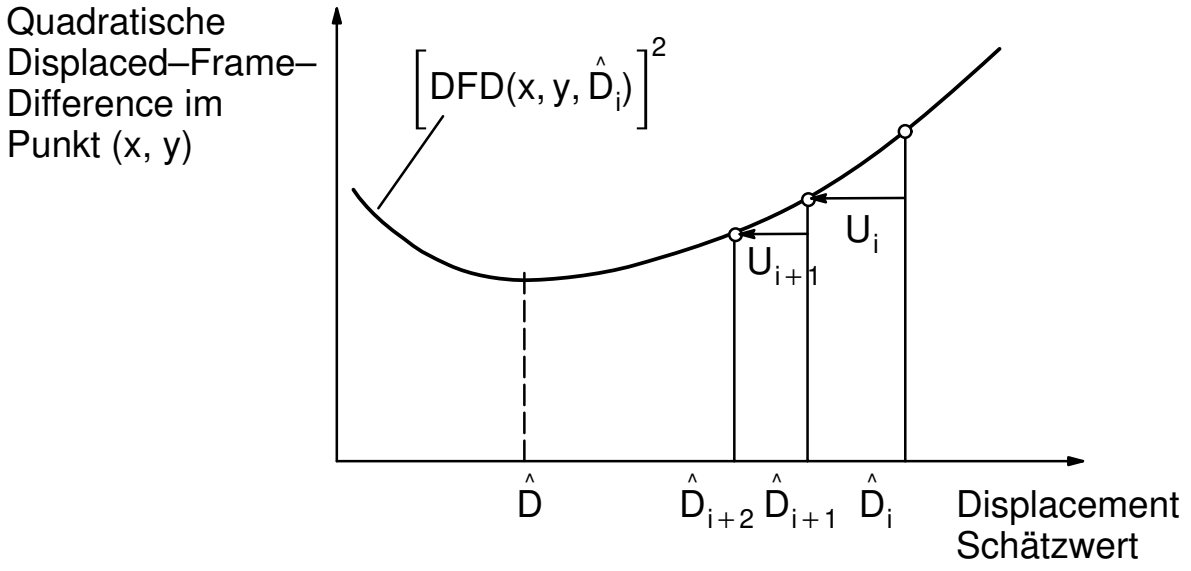


Bild 3.30 Darstellung einer rekursiven Schätzung

Das im Minimum gemessene $\hat{D}_i$ wird als Schätzwert $\hat{D}$ verwendet. Um schrittweise das Minimum der Kostenfunktion zu erreichen, kann das Gradientenverfahren angewendet werden, d.h.

$$\hat{D}_{i+1} = \hat{D}_i - \frac{1}{2}\,\varepsilon\,\nabla_{\hat{D}_i}[\,DFD\,(x,\ y,\ \hat{D}_i)]^2 \tag{3.108}$$

wobei $\nabla_{\hat{D}_i}$ der Gradientenoperator und ε eine positive Konstante ist. Die Auswertung von Gl.(3.108) liefert

$$\hat{D}_{i+1} = \hat{D}_i - \varepsilon \;\; DFD\,(x,\;\; y,\;\; \hat{D}_i)\;\; \nabla_{\hat{D}_i}\;\; [\; DFD(x,\;\; y,\;\; \hat{D}_i)] \tag{3.109}$$

und mit der Beziehung (3.107)

$$\hat{D}_{i+1} = \hat{D}_i - \varepsilon \;\; DFD\,(\;x,\;\; y,\;\; \hat{D}_i\;)\,\nabla\, s_{k-1}(\;x - \hat{dx}_i,\;\; y - \hat{dy}_i\;) \tag{3.110}$$

Dabei ist ∇ der Gradientenoperator bezüglich horizontaler und vertikaler Koordinaten $x,\;\; y$, d.h.

$$\nabla\, s_{k-1}(x - \hat{dx}_i,\;\; y - \hat{dy}_i) = \begin{bmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \end{bmatrix} s_{k-1}(x - \hat{dx}_i,\;\; y - \hat{dy}_i) \tag{3.111}$$

Mit der Konstanten ε kann die Größe des in jedem Schritt angewendeten Update–Terms U_i festgelegt werden. Ein kleines ε führt zu langsamer Konvergenz, erzeugt aber genauere Schätzwerte.

Zur Verbesserung der Konvergenz und Zuverlässigkeit sind inzwischen eine Reihe von Varianten des Gradientenverfahrens vorgeschlagen worden. Ferner können Blockmatching–Verfahren und rekursive Verfahren auch kombiniert angewendet werden. Dabei werden die Startwerte mit den Blockmatching–Verfahren bestimmt und die Vorzüge der genaueren Schätzung der rekursiven Verfahren anschließend genutzt.

3.5.2 Codierungsgewinn

Um den Codierungsgewinn der bewegungskompensierten Prädiktion zu bestimmen, werden die Übertragungsbitraten für eine Bewegtbildcodierung mit und ohne bewegungskompensierte Prädiktion berechnet und verglichen. Dabei wird angenommen, daß die Codierung eine Übertragungsbitrate liefert, die der Rate–Distortion–Funktion entspricht und das Leistungsdichtespektrum $S_{ss}(\omega_x\omega_y)$ des Bewegtbildsignals bekannt ist.

Die Übertragungsbitrate $R_s(d^*)$ einer zweidimensionalen sog. Intraframe–Codierung ($2D$) beträgt mit Gl. (2.149)

$$R_s(d^*) = \frac{1}{4\pi^2} \int_{\omega_x} \int_{\omega_y} \frac{1}{2} \log_2 \frac{S_{ss}(\omega_x\;\; \omega_y)}{d^*}\, d\omega_x\, d\omega_y \tag{3.112}$$

in bit je Abtastwert bzw. Bildpunkt.

Bei Anwendung einer bewegungskompensierten Prädiktion wird der Prädiktionsfehler e codiert und übertragen. Unter der Annahme einer rein translatorischen Bewegung des Bildinhaltes und exakter Displacementschätzung $\hat{D}$ ist der Prädiktionsfehler Null. Ist der Schätzwert $\hat{D}$ jedoch mit einem Schätzfehler $\Delta d = (\Delta x,\ \Delta y)$ behaftet, dessen Wahrscheinlichkeitsdichte $p_{\Delta x \Delta y}(\Delta x,\ \Delta y)$ ist, so ergibt sich ein Prädiktionsfehler e mit dem Leistungsdichtespektrum

$$S_{ee}(\omega_x,\ \omega_y) = S_{ss}(\omega_x,\ \omega_y) \cdot 2 \cdot [1 - Re\{P(\omega_x,\ \omega_y)\}] \tag{3.113}$$

wobei $P(\omega_x,\ \omega_y)$ die zweidimensionale Fouriertransformierte der Wahrscheinlichkeitsdichte $p_{\Delta x \Delta y}(\Delta x,\ \Delta y)$ ist [7]. Mit Gl. (3.112) ergibt sich die Übertragungsbitrate $R_e(d^*)$ bei Anwendung einer bewegungskompensierten Prädiktion bzw. Bewegungskompensation (MC) und Intraframe–Codierung (2D) des Prädiktionsfehlers

$$R_e(d^{*}) = \frac{1}{4\pi} \int_{\omega_x} \int_{\omega_y} \frac{1}{2} \log_2 \frac{S_{ee}(\omega_x,\ \omega_y)}{d^*}\, d\omega_x\, d\omega_y\ . \tag{3.114}$$

Zur numerischen Auswertung der Gl. (3.111) und Gl. (3.113) wird ein Leistungsdichtespektrum gemäß

$$S_{ss}(\omega_x,\ \omega_y) = \frac{2\pi}{\omega_0^2} \left[1 + \frac{\omega_x^2 + \omega_y^2}{\omega_0^2} \right]^{-3/2} \tag{3.115}$$

innerhalb der Bandgrenzen

$$\pi f_{sx} = 42.19\, \omega_0 \qquad \pi f_{sy} = 46.15\, \omega_0 \tag{3.116}$$

und $S_{ss}(\omega_x\ \omega_y) = 0$ außerhalb der Bandgrenzen angenommen. Diese Parameter sind anhand von Videokonferenzsignalen in SIF–Darstellung ermittelt worden.

Ferner wird angenommen, daß der Displacementschätzfehler eine Gaußsche Wahrscheinlichkeitsdichte

$$P_{(\Delta x,\, \Delta y)}(\Delta x,\ \Delta y) = \frac{1}{2\pi\sigma_{\Delta d}^2} \exp\left[- \frac{\Delta d_x^2 + \Delta d_y^2}{2\sigma_{\Delta d}^2} \right] \tag{3.117}$$

mit der Varianz $\sigma_{\Delta d}^2$ aufweist.

Bild 3.31 zeigt das Ergebnis der numerischen Auswertung für ein d^* entsprechend einem Signal–Rauschabstand von $SNR = 30\,\text{dB}$. Die Intra–Frame–Codierung (2D) benötigt für den Signal–Rauschabstand $SNR = 30\,\text{dB}$ eine Übertragungsbitrate von $R_s = 2,75\,\text{bit}$ je Abtastwert. Im Vergleich zu einer gedächtnislosen Quelle mit der

Übertragungsbitrate $SNR_{dB} = 6.02 R_{bit}$ entsprechend Gl. (2.146) bringt die Intraframe–Codierung einen Gewinn von 2.25 bit je Abtastwert.

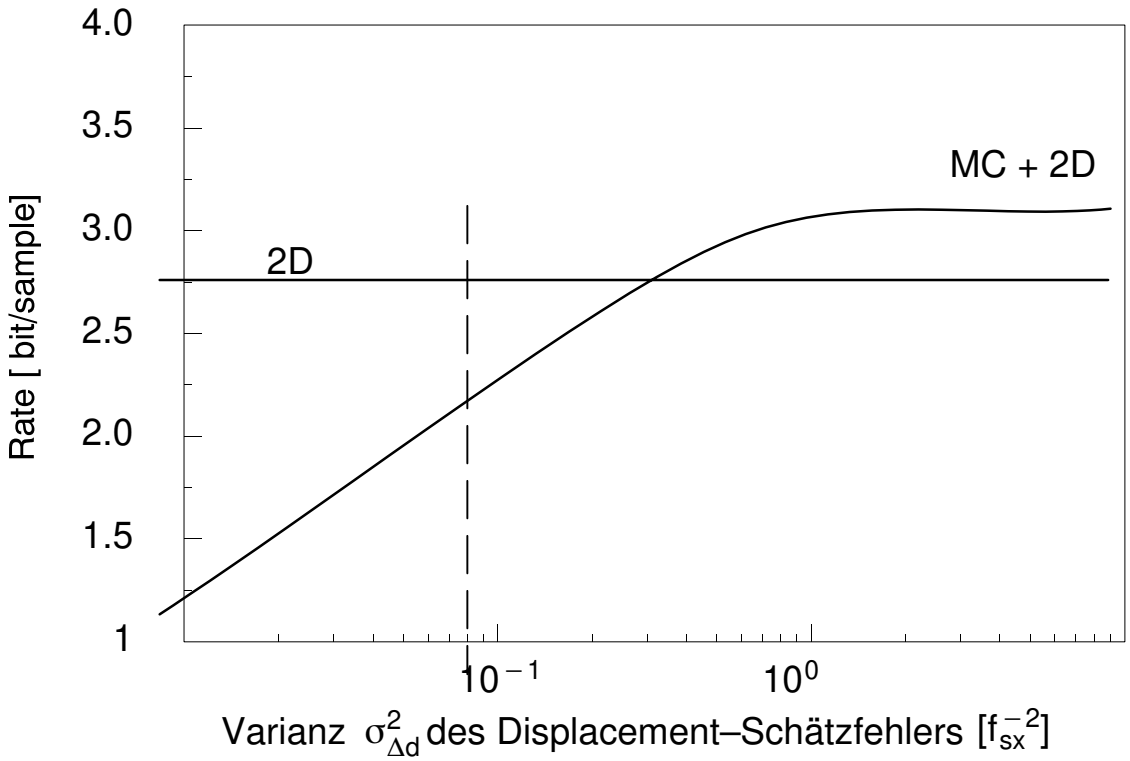


Bild 3.31 Minimale erreichbare Übertragungsrate als Funktion der Varianz des Displacement–Schätzfehlers bei einem Signal–Rauschabstand von SNR=30dB
2D Intraframe Codierung von s
MC+2D Intraframe Codierung des bewegungskompensierten Prädiktionsfehlers e

Die bewegungskompensierte Prädiktion mit Intraframe–Codierung des Prädiktionsfehlers kann den Codierungsgewinn verbessern, wenn die Varianz $\sigma^2_{\Delta d}$ des Displacementschätzfehlers unter $0,33 \cdot f_{sx}^{-2}$ liegt. Dabei ist f_{sx}^{-1} der Abstand zweier Bildpunkte in x–Richtung. Eine Displacementschätzer mit Integer–Pel–Genauigkeit hat eine Varianz von etwa $\sigma^2_{\Delta d} \approx 0,08\, f_{sx}^{-2}$. In diesem Fall liegt der zusätzliche Gewinn der bewegungskompensierten Prädiktion bei etwa $0,75$ bit je Abtastwert.

4. Modelle psychovisueller und psychoakustischer Wahrnehmung für die Irrelevanzreduktion

Bei der Irrelevanzreduktion werden Abweichungen zwischen dem zu codierenden Eingangssignal und dem decodierten bzw. rekonstruierten Ausgangssignal zugelassen, solange die Abweichungen vom menschlichen Auge oder Ohr nicht wahrgenommen werden. Im Fall, daß als Eingangssignal ein PCM–Signal vorliegt, werden die Abweichungen des Ausgangssignals im wesentlichen durch die im Coder enthaltenen Quantisierer und deren Quantisierungsfehler erzeugt. Je größer der zulässige Quantisierungsfehler ist, um so größer ist die mögliche Irrelevanzreduktion. Eine Irrelevanzreduktion ist optimal, wenn die von den Wahrnehmungsschwellen abhängigen zulässigen Quantisierungsfehler voll ausgenutzt werden.

Zu diesem Zweck werden die Quantisierer eines Coders von einem psychovisuellen oder psychoakustischen Modell gesteuert. Das Modell bildet die Wahrnehmung nach und berechnet die Wahrnehmungsschwellen. Aus den Wahrnehmungsschwellen ergeben sich die maximal erlaubten Stufenbreiten der Quantisierer. Da die Wahrnehmungsschwellen vom Eingangssignal abhängig sind, müssen die Quantisierer fortlaufend vom Modell durch Steuersignale angepaßt werden, siehe Bild 4.1. Die Teilsignale des dargestellten Coders können beispielsweise Frequenzkomponenten einer Transformation oder Teilbandzerlegung oder aber auch die Luminanz– und Chrominanzkomponenten eines Fernsehsignals sein.

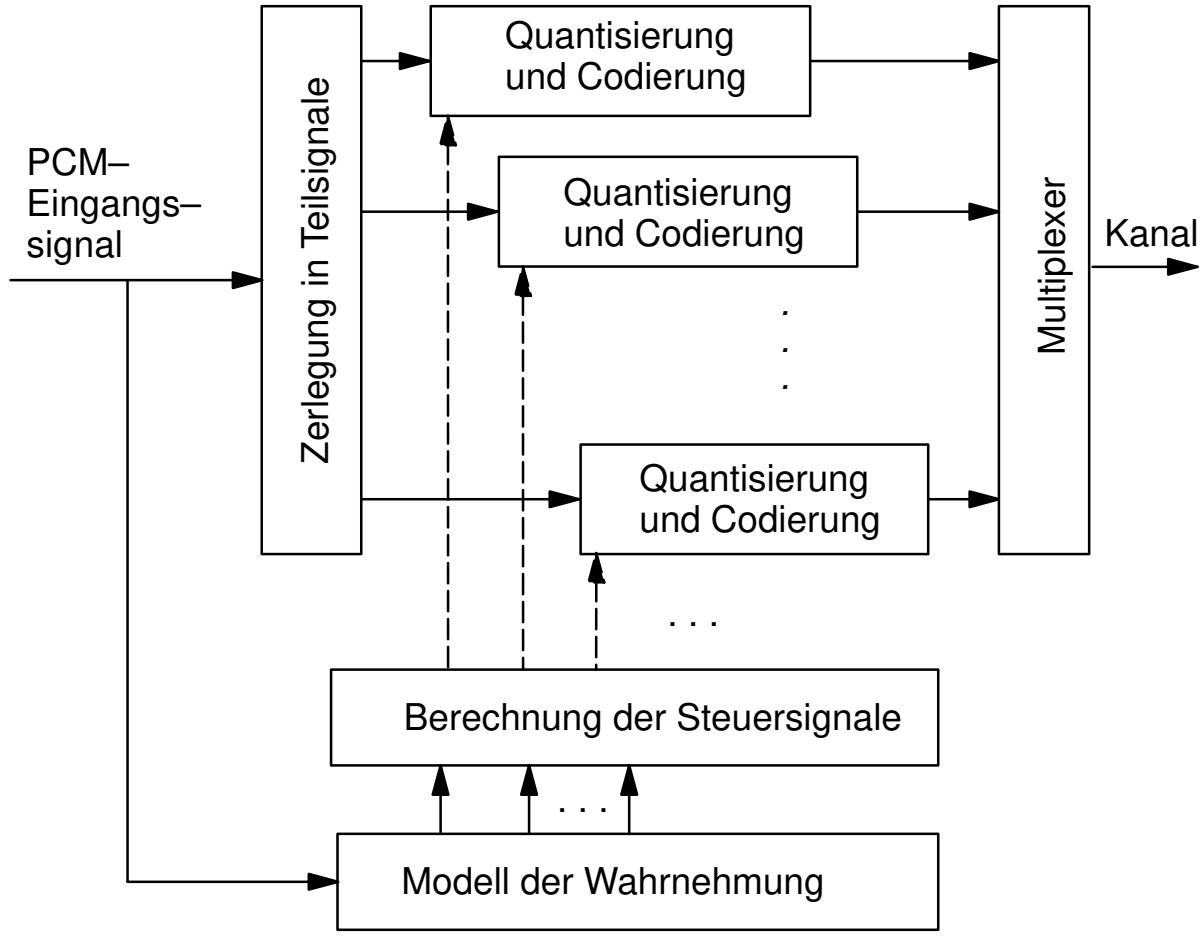


Bild 4.1 Blockschaltbild eines Coders mit Wahrnehmungsmodell für die Quantisierungssteuerung

4.1 Eigenschaften der visuellen Wahrnehmung

Es sollen einige wesentliche Eigenschaften des menschlichen visuellen Systems erläutert werden. Dazu gehören die Modulationsübertragungsfunktion, die Maskierung durch Leuchtdichtesprünge und die Abhängigkeit der Sichtbarkeitsschwelle für Störsignale von der Hintergrundleuchtdichte.

Modulationsübertragungsfunktion

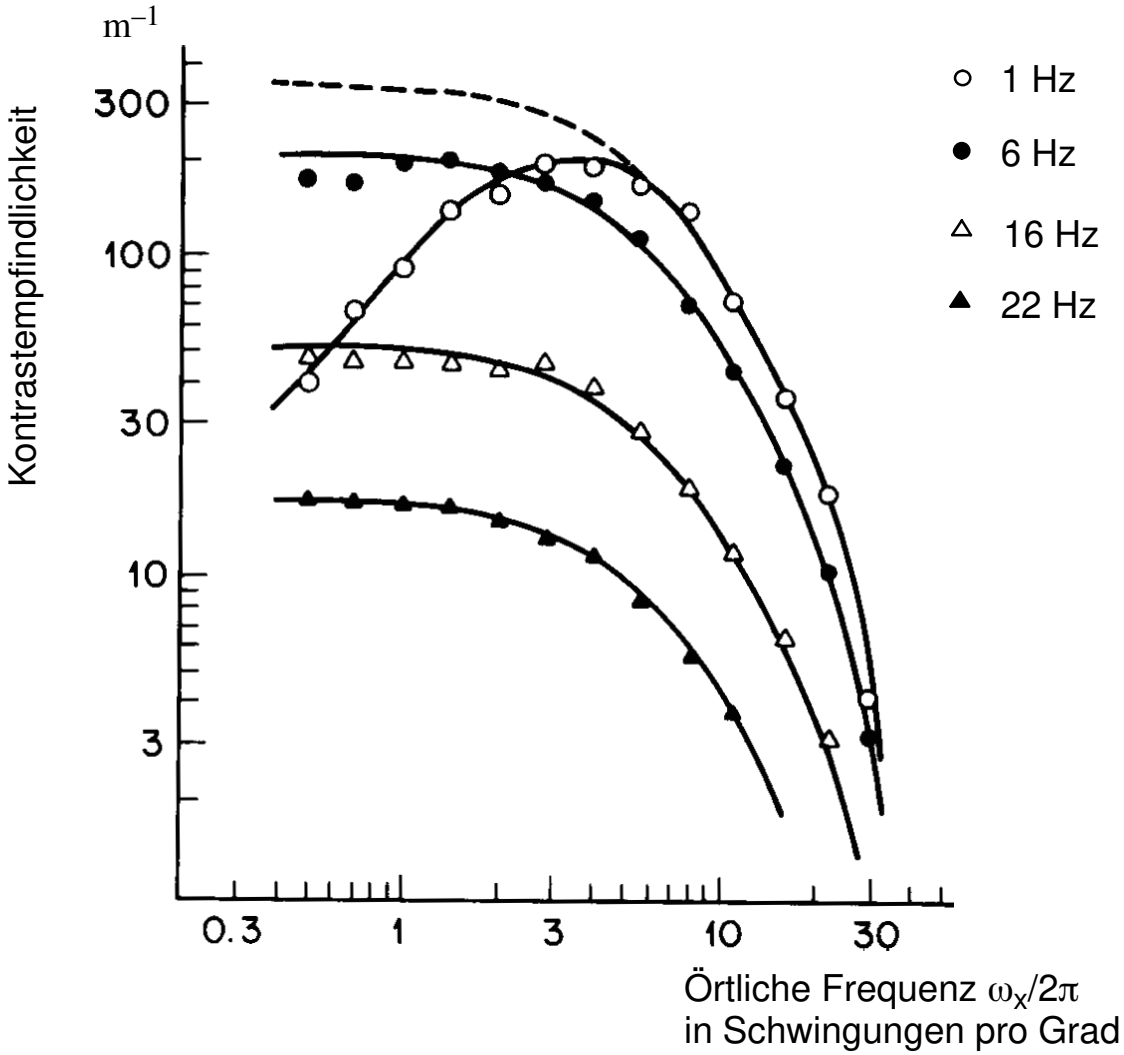


Bild 4.2 Kontrastempfindlichkeitsfunktionen des visuellen Systems
Parameter: unterschiedliche zeitliche Frequenzen $\omega_t/2\pi$

Unter der Modulationsübertragungsfunktion des menschlichen visuellen Systems (modulation transfer function, MTF) versteht man die Kontrastempfindlichkeit als Funktion der Frequenz. Die Modulationsübertragungsfunktion ist eine Funktion der Ortsfrequenzen ω_x, ω_y und der zeitlichen Frequenz ω_t. Sie kann in visuellen Tests mit einer in x–Richtung sinusförmig verlaufenden Leuchtdichte l des Bildschirmes entsprechend

$$l = l_0(1 + m \cos \omega_x x) \tag{4.1}$$

gemessen werden. Die Größe l_0 stellt dabei die Hintergrundleuchtdichte und die Größe m den Modulationsgrad bzw. Kontrast dar. Die Sichtbarkeitsschwellen dieses als Gitter erscheinenden Testsignals sind im wesentlichen nur vom Kontrast m des Gitters abhängig.

Bild 4.2 zeigt die Kontrastempfindlichkeit m^{-1} als Funktion von ω_x. Das Maximum der Empfindlichkeit liegt bei etwa 3 bis 4 Schwingungen je Grad. Indem die Leuchtdichte l zusätzlich zeitlich moduliert wird, kann der Einfluß der zeitlichen Frequenz ω_t erfaßt werden. Für Bildvorlagen mit in 45^o –Richtung modulierter Leuchtdichte ist eine um 3 dB geringere Kontrastempfindlichkeit zu messen.

Zur Irrelevanzreduktion wird vor allem die abfallende Kontrastempfindlichkeit der menschlichen visuellen Wahrnehmung bei hohen Ortsfrequenzen ausgenutzt.

Maskierung durch Leuchtdichtesprünge

Örtliche wie auch zeitliche Leuchtdichtesprünge erhöhen lokal die Sichtbarkeitsschwelle. Dieser Effekt wird als Maskierung eines Testsignals bzw. Störsignals durch ungleichförmige Hintergrundleuchtdichte bezeichnet.

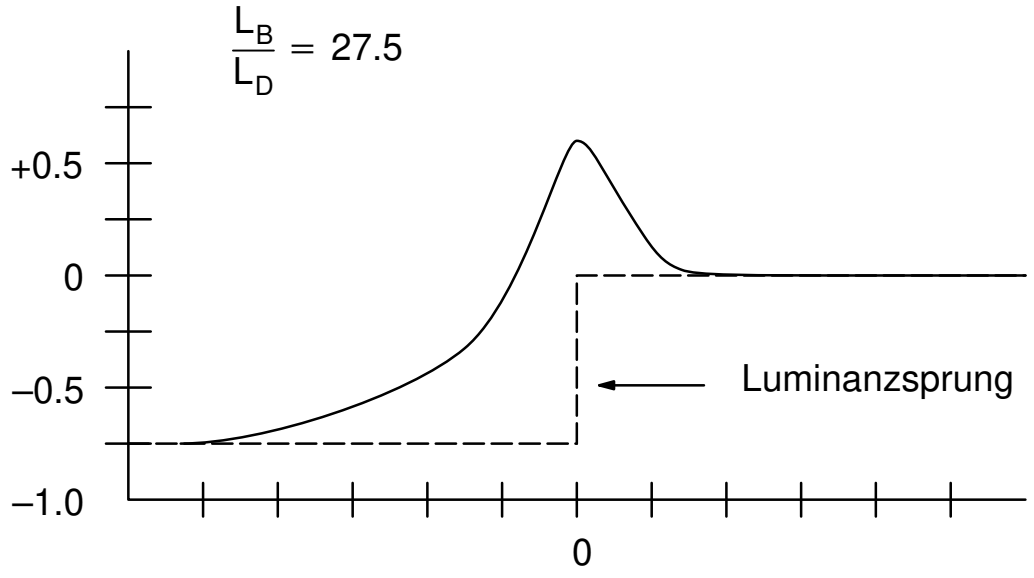


Bild 4.3 Sichtbarkeitsschwelle als Funktion der Distanz zum Luminanzsprung für ein Kontrastverhältnis $L_B/L_D = 27{,}5$ nach Fiorentini

Zur Messung der Maskierung wird einem örtlichen Leuchtdichtesprung in verschiedenen Abständen eine schmale Linie variabler Leuchtdichte als Testsignal überlagert. Die Leuchtdichte, bei der das Testsignal sichtbar wird, bestimmt die Sichtbarkeitsschwelle. Bild 4.3 zeigt die Sichtbarkeitsschwellen für zwei Leuchtdichtesprünge unterschiedlichen Kontrastes L_B/L_D.

Die Maskierung erlaubt größere Quantisierungsfehler in der Nähe von Leuchtdichtesprüngen. Sie wird insbesondere bei der DPCM–Codierung zur Irrelevanzreduktion ausgenutzt.

Hintergrundleuchtdichte

In Bereichen gleichförmiger Hintergrundleuchtdichte hängt die Sichtbarkeitsschwelle für Testsignale von der Größe der Hintergrundleuchtdichte ab. Die Sichtbarkeitsschwelle kann mit einem kreisförmigen Testmuster von $1{,}5^0$ Durchmesser gemessen werden, indem die für Sichtbarkeit erforderliche Leuchtdichtedifferenz Δl zur Hintergrundleuchtdichte l bestimmt wird. Das Ergebnis dieses Tests ist das sog. Weber–Fechnersche Gesetz. Es besagt, daß für große Leuchtdichten das Verhältnis $\Delta l/l$ nahezu konstant ist, d.h.

$$\Delta l/l \;=\; 10^{-2} \;=\; \text{konstant} \;. \tag{4.2}$$

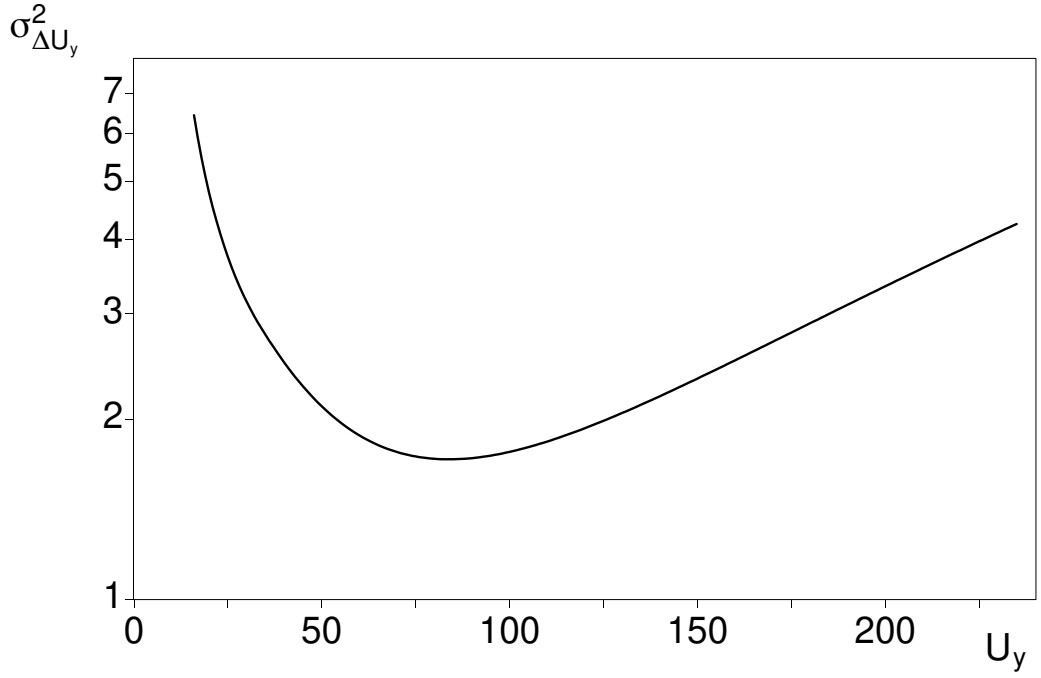


Bild 4.4 Zulässige Rauschleistung $\sigma^2_{\Delta U_y}$ bei konstanter Hintergrundleuchtdichte als Funktion des Luminanzsignals U_y

Diesem Gesetz zufolge sind bei großer Leuchtdichte größere Quantisierungsfehler als bei geringer Leuchtdichte erlaubt. Bei der Auswertung dieses Effektes zur Irrelevanzreduktion von Videosignalen muß aber noch der nichtlineare Zusammenhang zwischen Luminanzsignal U_y und Leuchtdichte l eines Fernsehmonitors, die sog. Gamma–Funktion, berücksichtigt werden. Die Sichtbarkeitsschwelle für die zuläs-

sige Quantisierungsfehlervarianz $\sigma^2_{\Delta Uy}$ des Luminanzsignals hat dann einen Verlauf entsprechend Bild 4.4.

4.1.1 Ein psychovisuelles Modell

Im folgenden soll ein psychovisuelles Modell für die Quantisierersteuerung von Fernsehsignalen behandelt werden. Um den Zusammenhang zwischen einem durch Quantisierung des Luminanzsignals U_y erzeugten Störsignal ΔD_y und der Wahrnehmungsschwelle θ für sichtbare oder nicht sichtbare Störungen herzustellen, wird das in Bild 4.5 dargestellte, aus 4 Stufen bestehende Modell zugrundegelegt.

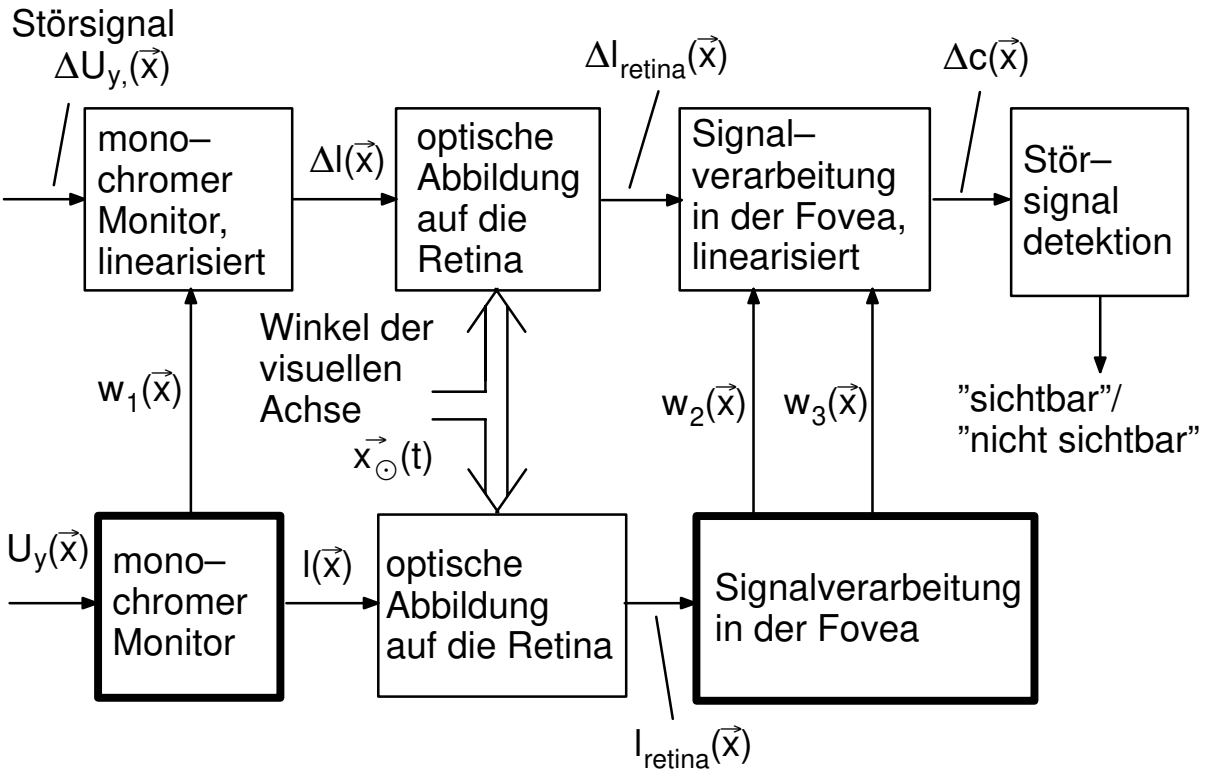


Bild 4.5 Psychovisuelles Modell der menschlichen visuellen Wahrnehmung von monochromen Fernsehsignalen

Der untere Zweig dieses Modells berücksichtigt als sog. Markierungszweig den Einfluß des Bildinhaltes auf die Wahrnehmung von Störsignalen. Dieser Zweig besteht aus einer Modellierung des Monitors, der optischen Abbildung auf die Netzhaut (Retina) sowie der Signalverarbeitung in der Fovea, der Zone des schärfsten Sehens. Dabei sind der Monitor und die Signalverarbeitung in der Fovea durch

nichtlineare Systeme darzustellen. Der obere Zweig beschreibt die Verarbeitung eines im allgemeinen kleinen Störsignals um den vom Nutzsignal vorgegebenen Arbeitspunkt. Diese Aufteilung in eine getrennte Verarbeitung von Nutzsignal und Störsignal hat den Vorteil, daß das Störsignal mittels linearer Systemtheorie behandelt werden kann.

Einfluß des Monitors

Der Faktor $w_1(\vec{z})$ erfaßt den Einfluß der nichtlinearen Bildröhrenkennlinie. Der Zusammenhang zwischen Leuchtdichte $l(\vec{z})$ an der Stelle $\vec{z}$ und Luminanzsignal $S(\vec{z})$ kann durch die Beziehung

$$l(\vec{z}) = L_{MON}\,(\,S(\vec{z}) + S_{MON})^{\gamma} \tag{4.3}$$

beschrieben werden, wobei die Monitor–Parameter bei Einstellung gemäß CCIR 500 und CCIR 601 Empfehlung in etwa

$$\begin{aligned} L_{MON} &= 0,35 \cdot 10^{-3} cd/m^2 \\ S_{MON} &= 15 \\ \gamma &= 2,2 \end{aligned} \tag{4.4}$$

betragen.

Ein Luminanz–Störsignal ΔS_y erzeugt dann ein Leuchtdichte–Störsignal Δl gemäß

$$\Delta l(\vec{z}) = \left[\left(\frac{dl}{dS_y}\right) \cdot \Delta S_y\right] = w_1(\vec{z}) \cdot \Delta S_y \ . \tag{4.5}$$

Einfluß der optischen Abbildung auf die Retina

Aufgrund der begrenzten Güte der optischen Abbildung wird ein Lichtpunkt auf dem Bildschirm nicht genau in einem Punkt auf der Retina abgebildet, sondern in einem kleinen Lichtfleck. Dieser Effekt läßt sich durch Faltung mit einer Point–Spread–Funktion $g_{optic}(\vec{z})$ beschreiben, so daß sich die Leuchtdichte auf der Retina ergibt als

$$l_{Retina}(\vec{z}) = g_{optic}(\vec{z}) * l(\vec{z}) \ . \tag{4.6}$$

Die Form der Point–Spread–Funktion entspricht sehr gut einer isotropen Gaußfunktion

$$g_{optic}(\vec{z}) = \frac{1}{2\pi r_{optic}^2} \exp\left\{\frac{-x^2 - y^2}{2r_{optic}^2}\right\} \tag{4.7}$$

mit $r_{optic} = 1,0\,\text{min}$

Signalverarbeitung in der Fovea

Mit den Faktoren $w_2(\vec{z})$ und $w_3(\vec{z})$ wird der Einfluß der Rezeptorträgheit und der sog. Inhibition berücksichtigt. Die Berechnung dieser Faktoren ist in [8] erläutert.

Störsignaldetektion

Diese Modellstufe faßt alle auf die Signalverarbeitung in der Fovea folgenden Vorgänge zusammen, die schließlich zur Entscheidung "Störsignal sichtbar" oder "Störsignal nicht sichtbar" führen. Das Störsignal $\Delta S(\vec{z})$ erzeugt ein Störsignal $\Delta C(\vec{z})$ der zum Gehirn führenden Gangliensignale. Die Entscheidung der Störsignaldetektion beruht auf der Auswirkung der sog. potentiellen Sichtbarkeit $PSB(\vec{z})$ an der Stelle $\vec{z}$. Dazu wird die lokale Leistung des Störsignals $\Delta C^2(\vec{z})$ berechnet und einem Filter $g_\theta(\vec{z})$

$$g_\theta(\vec{z}) = \frac{1}{2\pi r_\theta^2} \exp\left[\frac{-x^2 - y^2}{2r_\theta^2}\right] \tag{4.8}$$

mit dem Parameter $r_\theta = 13$ min zugeführt. Damit ergibt sich die potentielle Sichtbarkeit gemäß

$$PSB(\vec{z}) = g_\theta(\vec{z}) * \Delta C^2(\vec{z}) \ . \tag{4.9}$$

Als Bedingung für die Nichtsichtbarkeit eines Störsignals muß

$$\underset{z}{Max}\left\{PSB(\vec{z})\right\} \geq \theta \tag{4.10}$$

gefordert werden, wobei θ die Schwelle den Wert $\theta = 1,25 \cdot 10^{-5}$ hat.

Das erläuterte psychovisuelle Modell erfaßt mit den Gewichtsfaktoren $w_2(\vec{z})$ und $w_3(\vec{z})$ u.a. die Einflüsse der Modulationsübertragungsfunktion des menschlichen visuellen Systems, der Maskierung durch Leuchtdichtesprünge und die Abhängigkeit der Sichtbarkeitsschwelle von der Hintergrundleuchtdichte.

4.2 Eigenschaften der auditiven Wahrnehmung

Zu den wesentlichen Eigenschaften der auditiven Wahrnehmung gehören die gerade wahrnehmbare Änderung des Schalldruckpegels, die Simultanverdeckung im Frequenzbereich und die Verdeckung im Zeitbereich. Diese Eigenschaften werden nachfolgend erläutert.

Gerade wahrnehmbare Änderung des Schalldruckpegels

Die Wahrnehmungsschwelle für Änderungen Δp eines Schalldruckpegels p hängt von der Größe der Pegels ab. Sie wird mit einem Sinuston gemessen, der nacheinander mit zwei unterschiedlichen Pegeln angeboten wird. Der Zusammenhang zwischen Δp und p kann durch das Webersche Gesetz approximiert werden, d.h.

$$\frac{\Delta p}{p} = \text{konst.} \tag{4.11}$$

Eine genauere Beschreibung dieses Quotienten, die in einem größeren Pegelbereich gültig ist, wird in [9] mit

$$\frac{\Delta p}{p} = \frac{1}{4}\left(\frac{p}{p_R}\right)^{-1/6} \tag{4.12}$$

angegeben. Dabei ist p_R die sog. Ruhehörschwelle.

Simultanverdeckung im Frequenzbereich

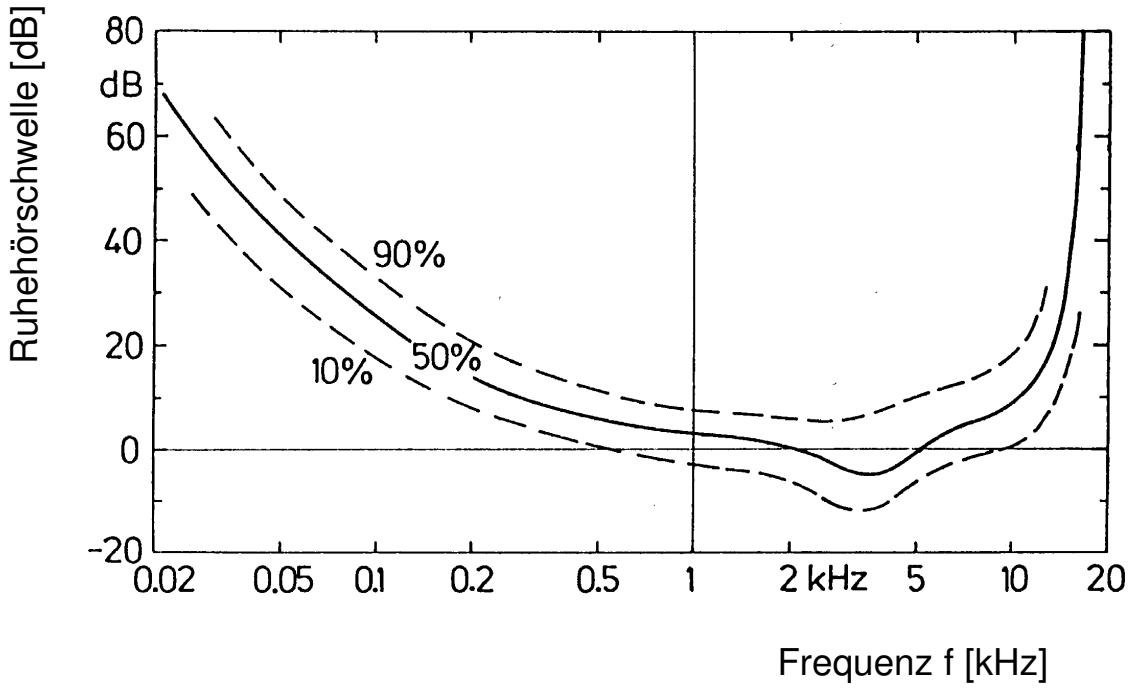


Bild 4.6 Ruhehörschwelle für einen Sinuston. Die Prozentangaben stellen den Anteil der Meßdaten von 100 Testpersonen dar, bei denen der Sinuston unterhalb der angegebenen Schwelle noch wahrnehmbar war. [17]

Ein relativ leises Testsignal kann durch ein gleichzeitig vorhandenes, lauteres Maskierersignal teilweise oder ganz verdeckt werden. Diese Eigenschaft des Ohres wird als Simultanverdeckung bezeichnet.

Für den Fall, daß kein Maskierersignal vorhanden ist, wird der Schalldruckpegel (SPL) an der Wahrnehmungsschwelle eines Testsignals als Ruhehörschwelle bezeichnet, siehe Bild 4.6. Der Schalldruckpegel wird in der Einheit dB angegeben und auf den Referenzdruck $p_0 = 20$ µPa bezogen.

Bei vorhandenem Maskierer erhöht sich die Wahrnehmungsschwelle. Sie wird in diesem Falle als Mithörschwelle bezeichnet.

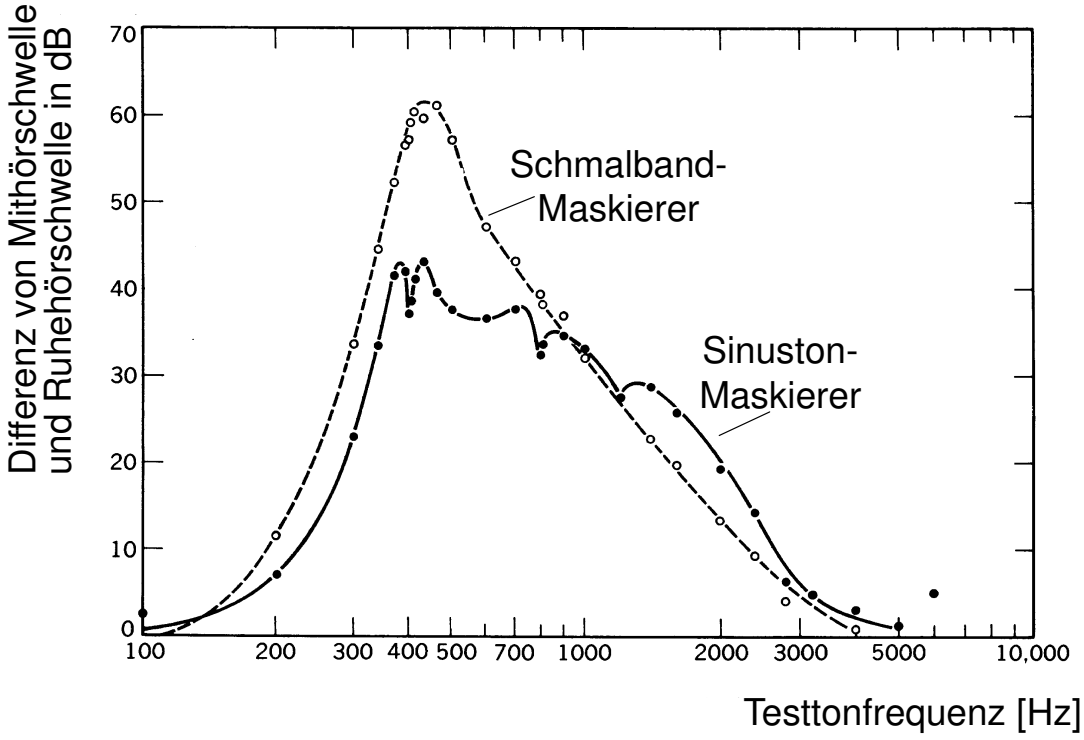


Bild 4.7 Differenz von Mithörschwelle und Ruhehörschwelle für einen Sinus-Testton verdeckt durch ein Schmalbandrauschen (Bandbreite 90 Hz) bzw. Sinuston bei 400 Hz mit einem Schalldruckpegel von 80 dB. [4]

Bild 4.7 zeigt die von einem maskierenden Schmalbandrauschen bzw. Sinuston erzeugten Mithörschwellen für einen Sinus-Testton. Die Schwellen zeigen einen asymmetrischen Verlauf bezüglich der Maskiererfrequenz mit einer deutlich stärkeren Verdeckungswirkung zu höheren Testsignalfrequenzen hin. Nahe der Maskiererfrequenz ist die Mithörschwelle des Sinustons gegenüber dem Schmal-

bandrauschen um bis zu 20 dB reduziert. Diese Reduzierung wird auch bei anderen tonalen Signalen im Unterschied zu rauschartigen Signalen beobachtet.

Der Mithörschwellenverlauf gibt bereits näherungsweise Aufschluß über die Frequenzselektivität des Gehörs. Interpretiert man den Verlauf der Mithörschwelle als Übertragungsfunktion eines Filters und stellt die von der Mittenfrequenz abhängige sog. Frequenzgruppenbreite B_{FG} mit einer gehörangepaßten Frequenzskalierung dar, haben alle Filter eine konstante Bandbreite. Dementsprechend sind auch die Flankensteilheiten bei einer Darstellung der Mithörschwellen auf dieser sog. Tonheitsskala nahezu unabhängig von der Maskiererfrequenz. Bild 4.8 zeigt die Frequenzgruppenbreite B_{FG} in Abhängigkeit von der Filtermittenfrequenz.

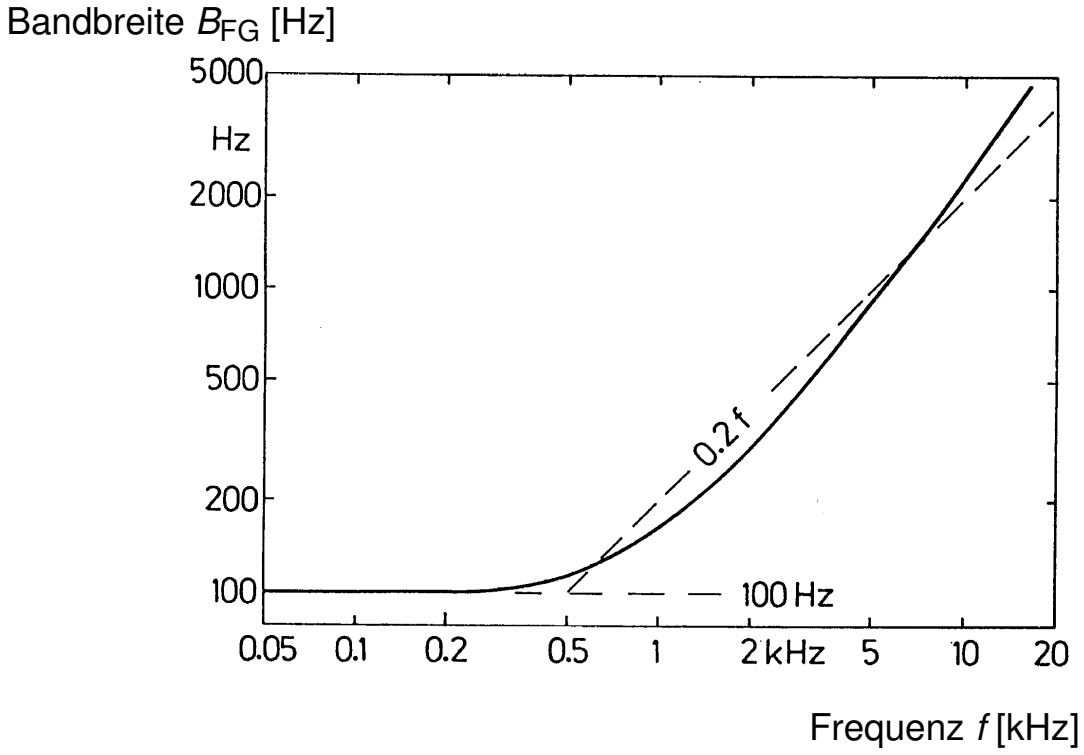


Bild 4.8 Frequenzgruppenbreite B_{FG} in Abhängigkeit von der Filtermittenfrequenz f. [17]

Verdeckung im Zeitbereich

Sehr schnell an- bzw. abklingende Maskierersignale können eine Verdeckung zeitlich vorangehend bzw. nachfolgend dargebotener Testsignale bewirken. Diese als Vor- und Nachverdeckung bezeichneten Effekte haben eine deutlich unterschiedliche Ausprägung, wie anhand der zeitlichen Mithörschwelle bei Verdeckung durch einen Sinusburst in Bild 4.9 ersichtlich wird. Die Vorverdeckung ist nur innerhalb

weniger Millisekunden vor dem Einsetzen des Maskierersignals wirksam, während die Nachverdeckung erst nach ca. 200 ms vollständig abgeklungen ist.

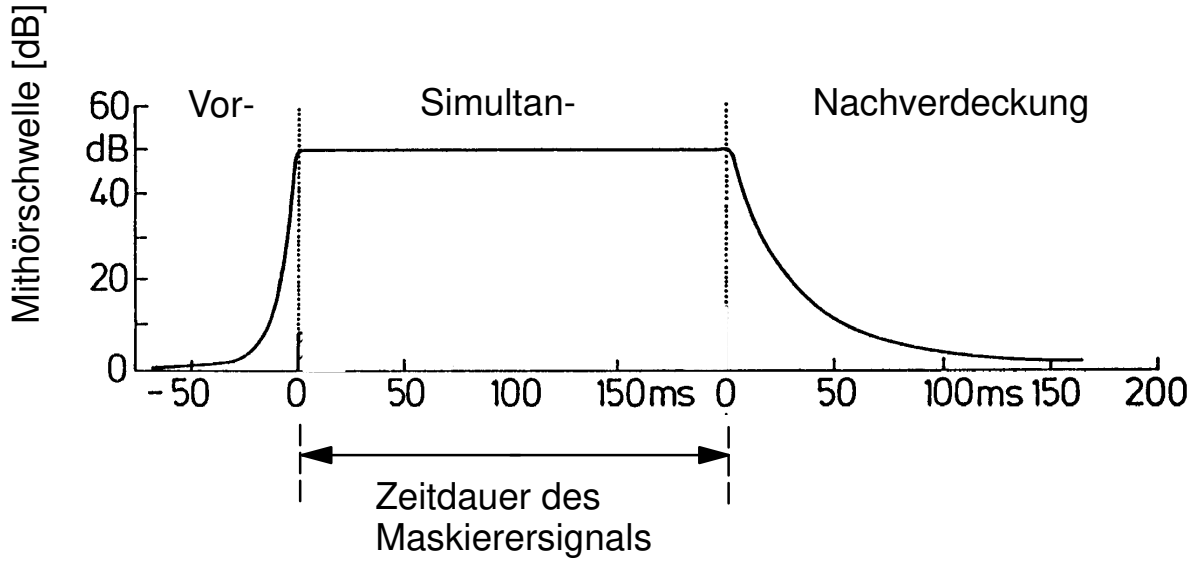


Bild 4.9 Zeitabhängigkeit der von einem 200 ms andauernden Maskierersignal hervorgerufenen Mithörschwelle. Das Maskierersignal ist ein Schmalbandrauschen. [17]

4.2.1 Ein psychoakustisches Modell

Das psychoakustische Modell berechnet jeweils für kurze Zeitintervalle des PCM–Eingangssignals einen Schätzwert für die aktuelle signalabhängige Maskierungsschwelle. Die Maskierungsschwelle gibt für jede Spektralkomponente die gerade noch wahrnehmbare Quantisierungsfehlerleistung an. Diese wird dann zur Generierung der Steuerinformation für die dynamische Bit– bzw. Rauschleistungszuweisung verwendet, indem die aktuell erforderliche Anzahl der Quantisierungsstufen für jede Spektralkomponente und jedes Zeitintervall bestimmt wird.

Im ersten Schritt führt das psychoakustische Modell eine Diskrete Fourier Transformation (DFT) für einen Block von Abtastwerten des PCM–Eingangssignals aus, siehe Bild 4.10. Im allgemeinen könnte dieselbe Spektralzerlegung wie bei der Zeit–Frequenzbereichs–Umsetzung des Coders verwendet werden. Allerdings ist die Frequenzauflösung bzw. die spektrale Darstellung dort nicht in jedem Fall für die Berechnung der Maskierungsschwellen geeignet. Aus der Psychoakustik ist bekannt, daß die Tonalität einer Maskierungskomponente Einfluß auf die Maskie-

rungsschwelle hat. Daher ist es angebracht, zwischen tonalen – z.B. sinusförmigen – und nicht–tonalen – z.B. rauschartigen – Komponenten zu unterscheiden. Dementsprechend werden die relevanten tonalen und nicht–tonalen Maskierer aus den Ausgangswerten der DFT für den aktuellen Abschnitt des Audiosignals bestimmt. Im nächsten Schritt erfolgt eine Anpassung der Frequenzauflösung an die Frequenzgruppen des Gehörs durch geeignete Gruppierung der DFT–Koeffizienten. Anschließend werden die individuellen Maskierungsschwellen für alle Frequenzkomponenten des Eingangssignals jeweils in Abhängigkeit von der Lage im Frequenzbereich, dem Lautstärkepegel und der Tonalität berechnet. Durch Überlagerung aller individuellen Maskierungsschwellen – einschließlich der Ruhehörschwelle – wird die sog. globale Maskierungsschwelle bestimmt. Für jede Spektralkomponente am Ausgang der Zeit–Frequenzbereichs–Umsetzung wird jeweils der Minimalwert dieser Maskierungskurve ermittelt, der wiederum die maximal zulässige Quantisierungsfehlerleistung angibt. Anhand dieser Werte wird dann die Bit– bzw. Rauschleistungszuweisung gesteuert, indem jeweils die erforderliche Quantisierungsstufenanzahl für jede Spektralkomponente ausgewählt wird.

Die Mithörschwelle und Steuersignale der Quantisierer werden jeweils für einen Block von Abtastwerten berechnet. Bei einer kurzzeitigen Signalleistungszunahme am Ende eines Blockes können aufgrund der blockweiten Einstellung der Quantisierer sog. Vorechos hörbar werden. Zur Vermeidung von Vorechos kann in solchen Situationen auf eine kurze Blocklänge umgeschaltet werden, bei der eine bessere Anpassung des zeitlichen Störsignalverlaufs an die Mithörschwelle möglich ist. Bei einer kurzzeitigen Energieabnahme des Audiosignals wird in der Regel kein Nachecho hörbar, weil die Nachverdeckung wesentlich länger andauert als die Vorverdeckung.

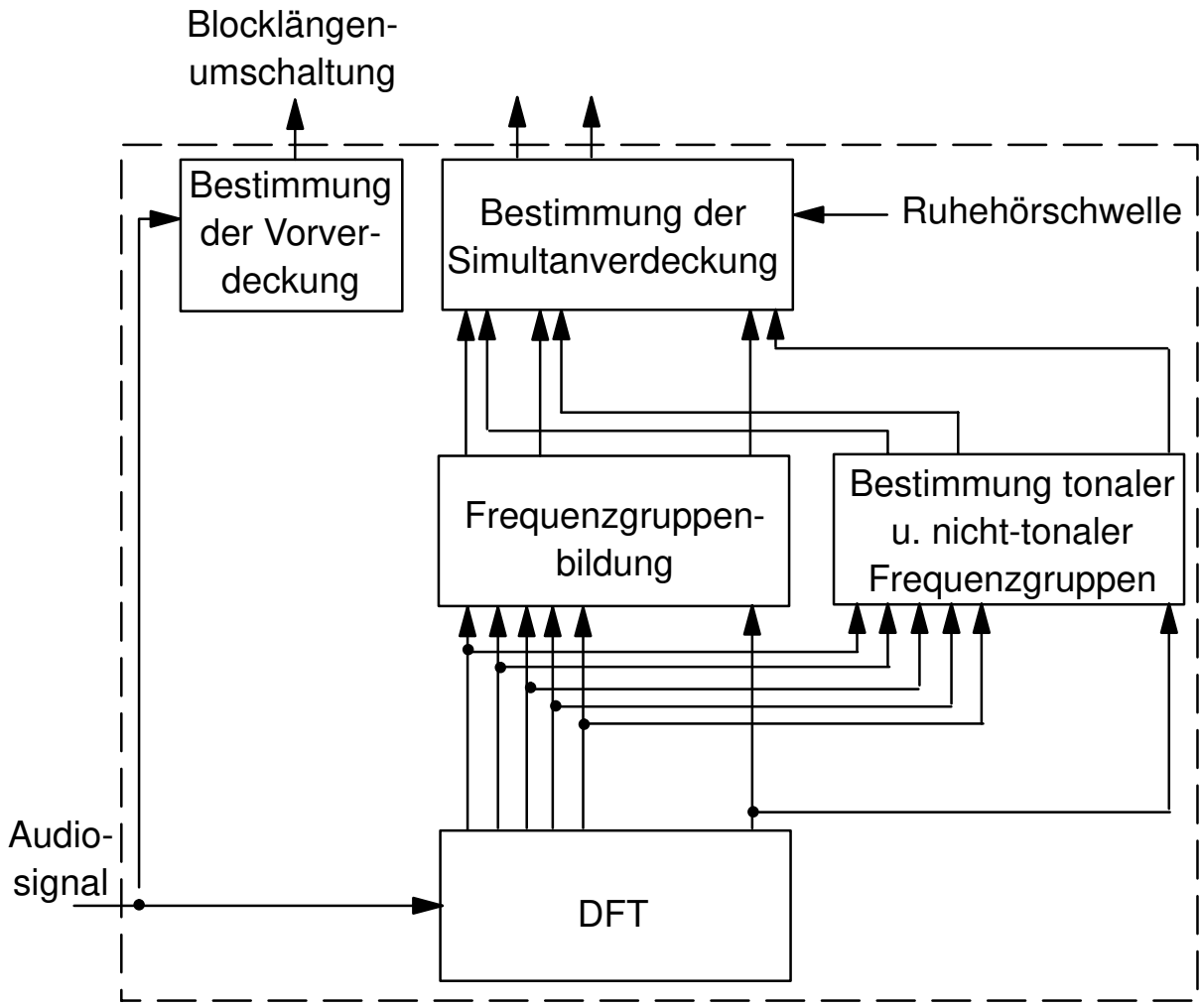


Bild 4.10 Psychoakustisches Modell der auditiven Wahrnehmung

Ein Beispiel für die Berechnung einer Maskierungsschwelle zeigt Bild 4.11. Dargestellt sind die Ruhehörschwelle, die Signalkomponenten des Vokals ”A” und die zugehörige globale Maskierungsschwelle. Die grau hinterlegten Bereiche geben jeweils die durch eine geeignete Bitzuweisung eingestellte Quantisierungsfehlerleistung für die ersten Spektralkomponenten an.

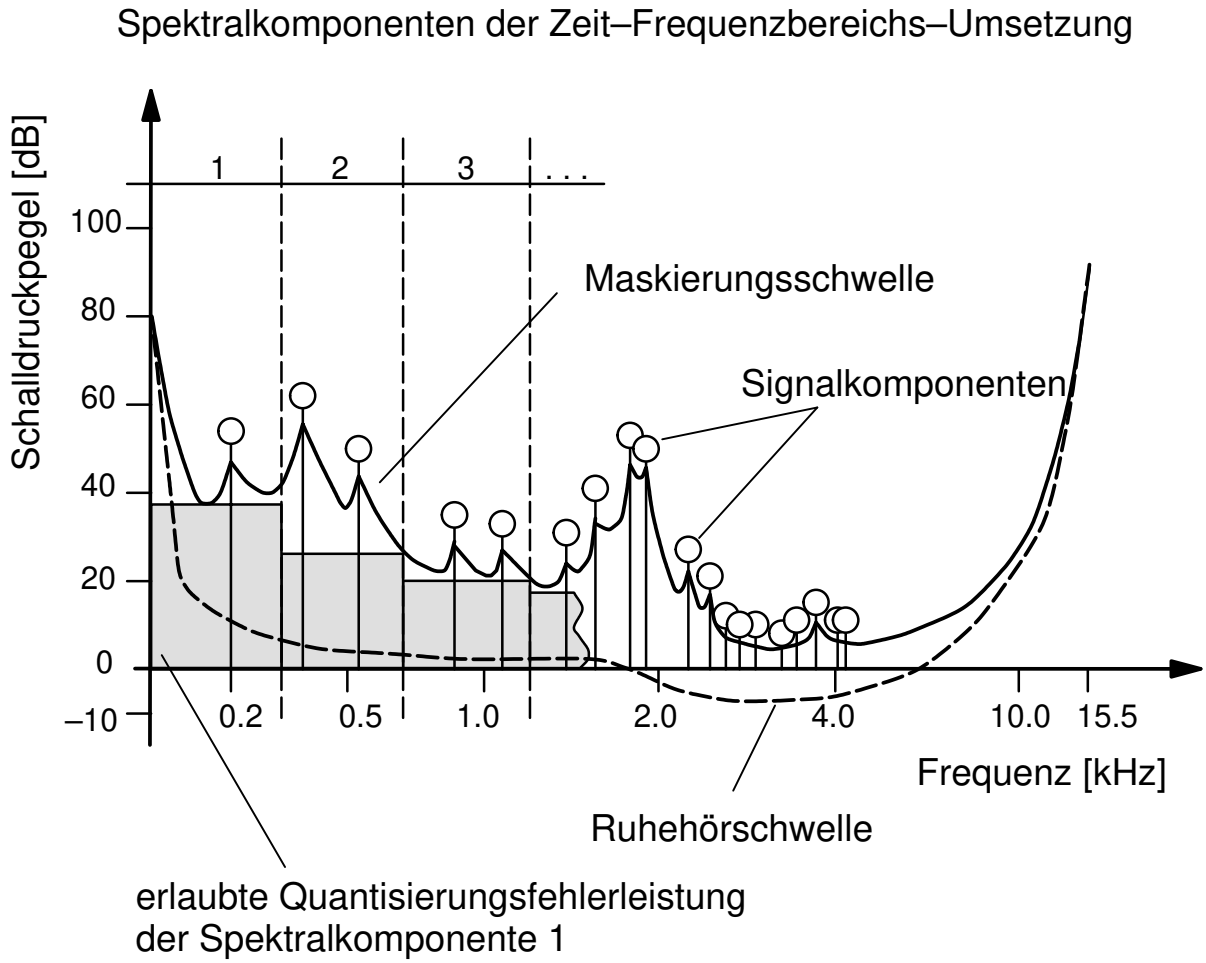


Bild 4.11 Berechnung der Maskierungsschwelle für den Vokal ”A”

5. Anwendungen

Es wird ein Überblick über die Kompressionsalgorithmen zur Speicherung und Übertragung von digitalen Video– und Audiosignalen für Multimedia–Systeme gegeben. Der Überblick beschränkt sich auf Algorithmen, die von internationalen Organisationen der Standardisierung entwickelt wurden. Es werden die Codierungskonzepte beschrieben und Literaturhinweise für detailierte Informationen gegeben.

Aufgrund der verschiedenartigen möglichen Anwendungen, wie z.B.

- interaktive Multimedia–Systeme,
- Bildtelefon,
- Electronic News Gathering,
- Satelliten und terrestrischer TV–Rundfunk,

muß ein Codierungsstandard sehr unterschiedliche Systemanforderungen erfüllen. Die wesentlichen Anforderungen sind

- verschiedenartige Eingangsformate,
- flexible Datenraten,
- Skalierbarkeit des Bitstroms,
- Editierbarkeit des codierten Bitstroms,
- begrenzte Codierungs– und Decodierungsverzögerung.

Der Überblick schließt den ISO–JPEG–Standard zur Codierung von Standbildern, den CCITT–Standard H.261 zur Codierung von Bewegtbildern bei niedrigen Datenraten von $p \times 64$ kbits und die ISO–MPEG–Standards zur Codierung von Videosignalen ein, die für Datenraten von 1,5 Mbit/s und 6,0 Mbit/s ausgelegt sind. Abschließend wird der ISO–MPEG–Audio–Codierungsstandard erläutert.

5.1 Codierung von Bildsignalen

5.1.1 Der JPEG–Standard zur Codierung von Standbildern

Im folgenden wird der JPEG–Basis–Algorithmus skizziert. Erweiterungen, wie z.B. die Optionen zur prädiktiven Codierung oder zur progressiven Standbildübertra-

gung sind zusammen mit weiteren Einzelheiten in [25], [27] angegeben. Der Basis–Algorithmus beschreibt eine Transformationscodierung basierend auf der DCT.

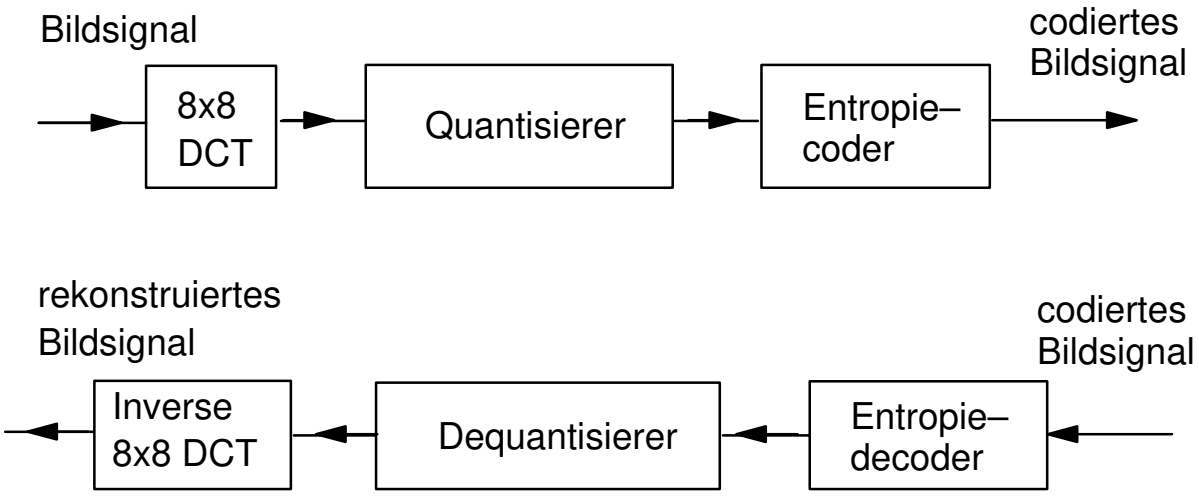


Bild 5.1 JPEG–Blockschaltbild des Basis–Algorithmus

In Bild 5.1 sind die wesentlichen Verarbeitungsschritte auf der Coder– und Decoderseite dargestellt. Zunächst wird ein Standbild blockweise mit einer 8×8 DCT transformiert. Ein Bildsignal mit der PCM–Darstellung von jeweils 8 bit pro Abtastwert erfordert dabei eine Genauigkeit von 11 bit für die DCT–Koeffizienten.

Um die Eigenschaften des menschlichen visuellen Wahrnehmungssystems auszunutzen, wird eine gewichtete Quantisierung angewendet. Sie berücksichtigt, daß im allgemeinen die subjektive Wahrnehmung von Quantisierungsrauschen bei hohen Frequenzen unempfindlicher ist. Deshalb wird eine Quantisierungsmatrix angewendet, die eine gröbere Quantisierung von hochfrequenten DCT–Koeffizienten erzeugt. Die Werte der Quantisierungsmatrix hängen von zahlreichen Parametern ab, wie z.B. dem Betrachtungsabstand, dem Monitor und dem Bildsignal selbst. Aus diesem Grund können bis zu 4 Quantisierungsmatrizen für ein Standbild als zusätzliche Seiteninformation mit den komprimierten Bilddaten übertragen werden. Da die Farbkomponenten des Bildsignals unabhängig voneinander codiert werden, kann jeder Farbkomponente jeweils eine Quantisierungsmatrix zugeordnet werden. Als Beispiel sind in Bild 5.2 zwei Quantisierungsmatrizen für die Luminanz– und Chrominanzkomponenten eines Standbildes im CCIR 601 Format angegeben.

Quantisierungsmatrix Q(u,v) für Luminanzwerte :

Q(u,v)		u							
		0	1	2	3	4	5	6	7
	0	16	11	10	16	24	40	51	61
	1	12	12	14	19	26	58	60	55
	2	14	13	16	24	40	57	69	56
v	3	14	17	22	29	51	87	80	62
	4	18	22	37	56	68	109	103	77
	5	24	35	55	64	81	104	113	92
	6	49	64	78	87	103	121	120	101
	7	72	92	95	98	112	100	103	99

Quantisierungsmatrix Q(u,v) für Chrominanzwerte:

Q(u,v)		u							
		0	1	2	3	4	5	6	7
	0	17	18	24	47	66	99	99	99
	1	18	21	26	66	99	99	99	99
	2	24	26	56	99	99	99	99	99
v	3	47	66	99	99	99	99	99	99
	4	99	99	99	99	99	99	99	99
	5	99	99	99	99	99	99	99	99
	6	99	99	99	99	99	99	99	99
	7	99	99	99	99	99	99	99	99

Bild 5.2 JPEG–Beispiele für Quantisierungsmatritzen zur visuellen Gewichtung

Jedes Element $Q(u,\ v)$ einer Quantisierungsmatrix bezeichnet die Stufenbreite eines gleichförmigen Quantisierers, der auf die DCT–Koeffizienten $F(u,\ v)$ entsprechender Frequenzen $u,\ v$ angewendet wird. Da die Quantisierungsmatrizen dem Decoder bekannt sind, können die quantisierten Koeffizienten $F'(u,\ v)$ durch eine sogenannte Dequantisierung rekonstruiert werden, siehe Bild 5.3.

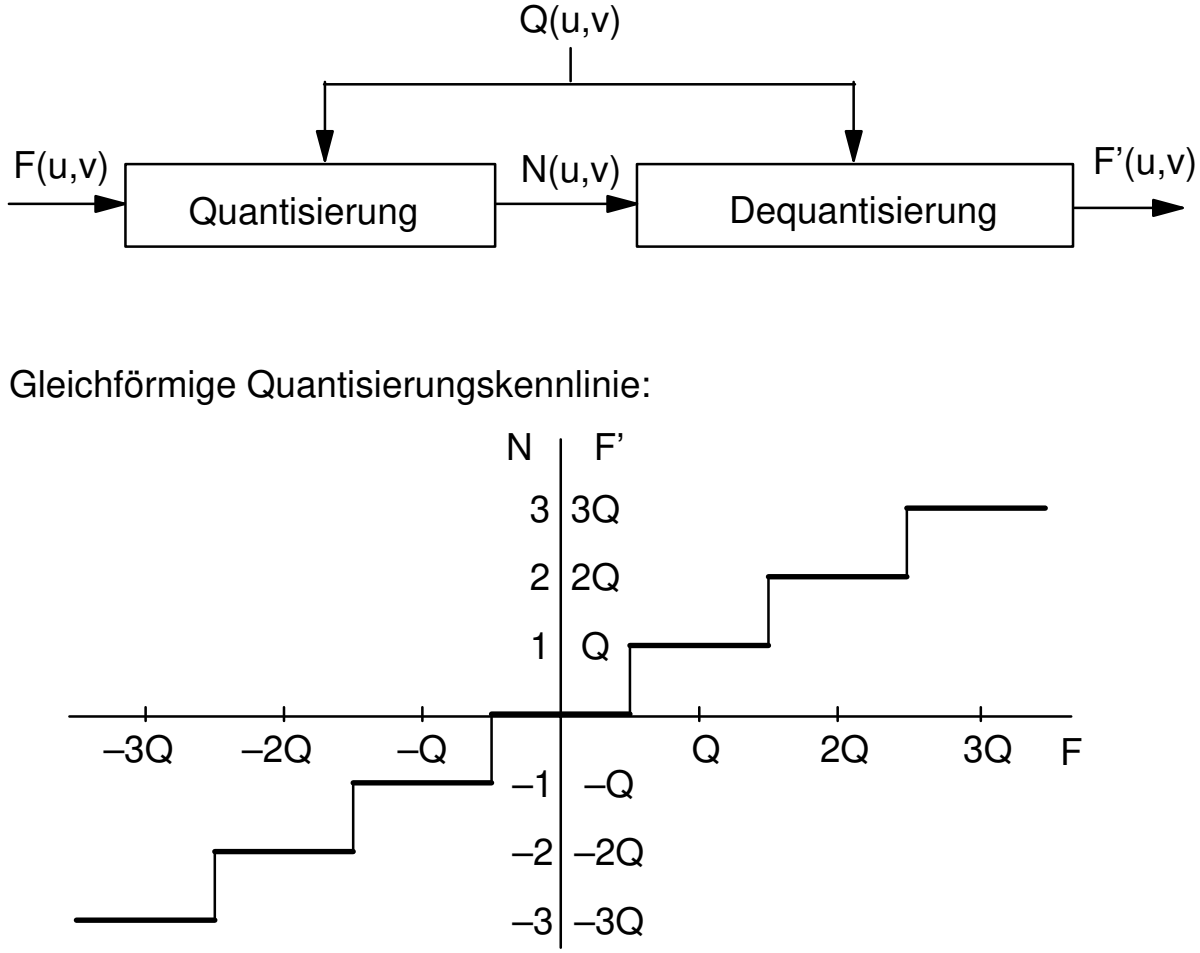


Bild 5.3 JPEG–Quantisierung und Dequantisierung von DCT–Koeffizienten

Nach der Quantisierung werden die DC–Werte ($u = 0,\ v = 0$) mittels einer DPCM codiert. Diese DPCM nutzt die große Korrelation von DC–Werten in benachbarten 8×8 Blöcken. Der verbleibende DC–Prädiktionsfehler wird zusammen mit den AC–Koeffizienten eines Blocks gemäß Bild 5.4 mittels eines Zigzag–Scan sequentiell angeordnet und einer Entropiecodierung zugeführt.

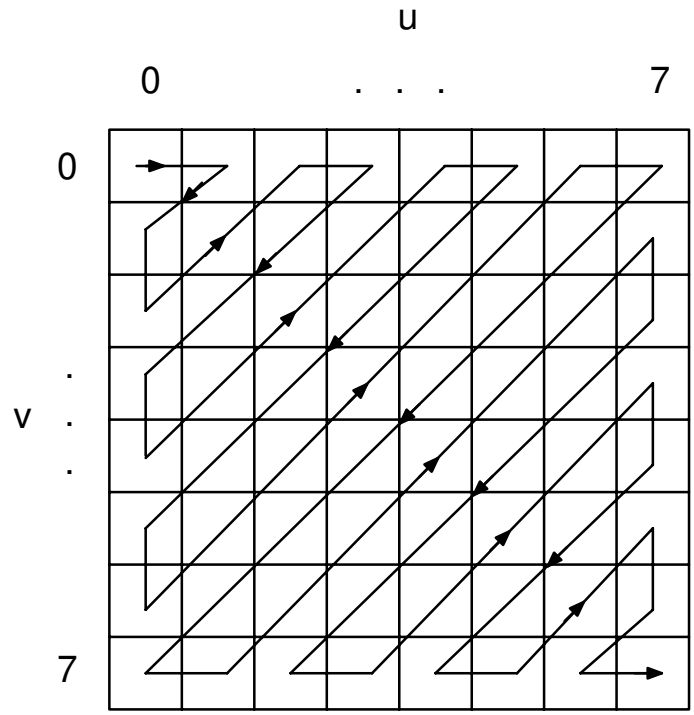


Bild 5.4 ZIGZAG–Scan für Blöcke mit 8x8 DCT–Koeffizienten

Der Zigzag–Scan beginnt mit dem DC–Koeffizienten gefolgt von niederfrequenten AC–Koeffizienten und endet bei hochfrequenten Koeffizienten. Nach einer Quantisierung ergeben sich häufig sehr kleine Werte oder Null für die hochfrequenten AC–Koeffizienten. Nachdem der letzte von Null verschiedene Koeffizient codiert worden ist, kann zur Verbesserung der Codierungseffizienz ein spezielles End Of Block (EOB) Codewort gesendet werden. Einem Decoder wird damit angezeigt, daß die restlichen AC–Koeffizienten in der Reihenfolge des Scans jeweils Null sind.

Für die DC– und die AC–Koeffizienten wird jeweils eine Codewort–Tabelle nach Huffman benutzt. Zwei Sätze von Codewort–Tabellen, die sich an die unterschiedlichen Statistiken von Luminanz– und Chrominanzkomponenten eines Standbildes anpassen, können als zusätzliche Seiteninformationen übertragen werden.

Ein JPEG–Coder arbeitet typischerweise mit Kompressionsverhältnissen im Bereich von 8:1 bis 18:1.

5.1.2 Der H261–Standard zur Codierung von Bildfernsprechsignalen

Der H.261 Standard [28], [29] kann zur Kompression von Bildsequenzen für Videokonferenzen oder für Bildfernsprechen verwendet werden. Der Algorithmus zielt auf Datenraten von $p \times 64$ kbits/s, wobei p im Bereich von 1 – 30 liegt. Der H.261 Coder verarbeitet Bilder, die nicht zeilenverschränkt sind, und verwendet als Eingabedatenformat das sogenannte Commom Intermediate Format (CIF), siehe Tabelle 5.1.

	CCIR	SIF–625	CIF
ANZAHL DER ABTAST-WERTE JE ZEILE LUMINANZ: CHROMINANZ:	720 360	352 176	352 176
ANZAHL DER AKTIVEN ZEILEN JE BILD LUMINANZ: CHROMINANZ:	288 288	288 144	288 144
QUANTISIERUNG IN BIT JE ABTASTWERT:	8	8	8
BILDFOLGE–FREQUENZ:	50 ZEILEN–VERSCHRÄNKT	25 NICHT ZEILEN–VERSCHRÄNKT	29,97 NICHT ZEILEN–VERSCHRÄNKT
BITRATE IN MBIT / S:	166	30,5	36,5

Tabelle 5.1: PCM–Formate

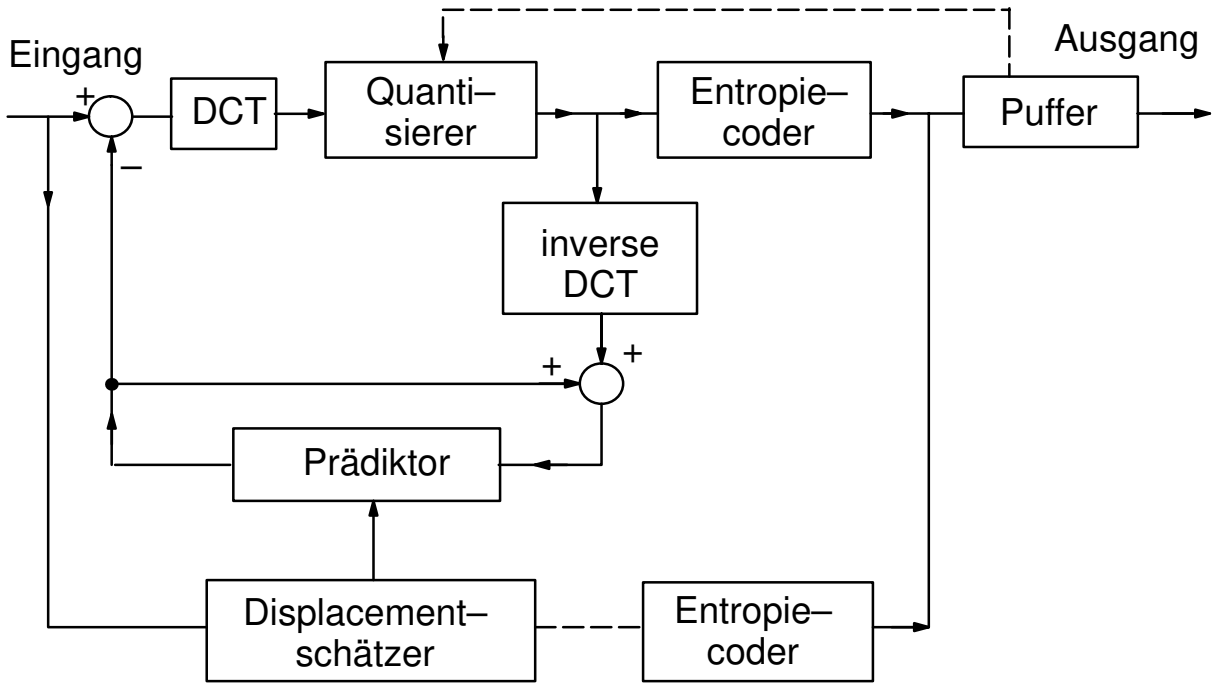


Bild 5.5 Blockschaltbild eines bewegungskompensierenden Hybridcoders

Prädiktive Codierung und Transformationscodierung werden in einem hybriden Codieralgorithmus kombiniert. Das Blockschaltbild eines bewegungskompensierenden hybriden Coders ist in Bild 5.5 dargestellt. Die zeitliche Redundanz wird mit Hilfe der bewegungskompensierenden Prädiktion reduziert. In einer DPCM–Schleife wird der verbleibende Prädiktionsfehler berechnet und anschließend mit der DCT transformiert. Die Transformationscodierung reduziert die örtliche Redundanz. Die quantisierten DCT–Koeffizienten werden dann einer Entropiecodierung unterzogen. Displacementvektoren werden auf der Coderseite geschätzt und nach einer Entropiecodierung als zusätzliche Seiteninformation an den Decoder übertragen. Da Entropiecodierungen Codewörter mit variabler Länge benutzen, wird ein Puffer eingeführt um eine konstante Datenrate am Ausgang zu gewährleisten. Der Pufferfüllstand wird zum Quantisierer zurückgeführt und dient zur Regelung der Quantisierung, um einen Pufferüberlauf zu vermeiden.

Jedes Bild wird in sogenannte Makroblöcke unterteilt. Ein Makroblock besteht aus einem Luminanzblock von 16×16 Bildpunkten und zwei korrespondierenden Chrominanzblöcken jeweils der Größe 8×8. Es gibt grundsätzlich zwei Prädiktionsmodi zwischen denen unabhängig für jeden Makroblock gewählt werden kann. Ein bewegungskompensierender (MC) Modus gestattet, einen Displacementvektor mit Bildpunktgenauigkeit für einen Makroblock zu übertragen, so daß eine bewe-

gungskompensierende Prädiktion durchgeführt werden kann. Falls der Displacementvektor gleich dem Nullvektor ist, wird der nicht bewegungskompensierende (No–MC) Modus gewählt, um statische Bildbereiche zu prädizieren. Der MC– und der No–MC–Modus entsprechen einer Interframecodierung mit der zeitlichen Referenz zu einem Block im vorangehenden, bereits übertragenen Bild. In Fällen, wo die Prädiktion versagt, z.B. verursacht durch einen Szenenschnitt, schaltet ein dritter sogenannter Intra–Modus die Prädiktion ab. Ein Makroblock wird dann mittels der DCT intraframe codiert.

Sämtliche Makroblöcke des ersten Bildes einer Sequenz werden intraframe codiert., da keine zeitliche Referenz zu einem vorangehenden Bild möglich ist. Nach dem ersten Bild sind MC–, No–MC– und Intra–Modus für jeden Makroblock erlaubt. Die Elemente eines Makroblocks werden blockweise mit einer 8×8 DCT transformiert. Ein Prädiktionsfehlersignal mit der PCM–Darstellung von 9 bit pro Abtastwert erfordert eine Genauigkeit von 12 bit für die DCT–Koeffizienten. Ein gleichförmiger Quantisierer mit adaptiver Totzone wird zur Quantisierung der DCT–Koeffizienten benutzt. Die Quantisierstufenbreite kann für jeden Makroblock individuell eingestellt werden. Die gewählte Stufenbreite wird anschließend für sämtliche Koeffizienten der 6 DCT–Blöcke eines Makroblocks angewendet. Eine Quantisierungsmatrix wird nicht benutzt.

Die quantisierten DCT–Koeffizienten eines 8×8 Blocks werden mit dem Zigzag–Scan sequentiell angeordnet und einer Entropiecodierung zugeführt. DC–Koeffizienten in intraframe codierten Makroblöcken werden gesondert behandelt. Diese werden mit einem gleichförmigen Quantisierer fester Stufenbreite quantisiert und mit Codewörtern konstanter Länge codiert. DC–Werte in interframe codierten Makroblöcken und AC–Werte werden mit einem zweidimensionalen Code variabler Länge codiert. Dabei wird ein Wertepaar bestehend aus einer von Null verschiedenen Amplitude und einer Lauflänge von vorangehenden Nullen im Zigzag–Scan als ein Ereignis aufgefaßt und mit einer Codewort–Tabelle nach Huffman codiert. Dem letzten codierten Ereignis eines Blocks folgt ein End Of Block (EOB)–Codewort. Sämtliche Codewort–Tabellen liegen fest und sind in H.261 standardisiert. Es können keine Codewort–Tabellen als zusätzliche Seiteninformation übertragen werden.

Eine Bewegtbildsequenz mit einer örtlichen CIF–Auflösung und einer Bildfolgefrequenz von 10 Hz kann mit einem Kompressionsverhältnis von ca. 190:1 codiert werden, so daß sich eine komprimierte Datenrate von 64 kbit/s einstellt.

5.1.3 Der MPEG–1–Standard zur Codierung von Videosignalen im CIF–Format

In Phase I ihrer Aktivitäten hat die ISO–Moving Picture Experts Group (MPEG) einen Standard zur Codierung von Videosignalen für interaktive Multimediasysteme entwickelt. In digitalen Speichermedien, wie z.B. der CD–ROM, stehen Datenraten von ca. 1,15 Mbit/s für die komprimierten Videodaten zur Verfügung.

Es werden Bildsequenzen ohne Zeilenverschränkung verarbeitet. Durch Auslassen der geraden Halbbilder, örtliche Filterung und Reduzierung der Abtastfrequenz wird eine CCIR 601 Sequenz in eine Source Input Format (SIF) Sequenz konvertiert, siehe Tabelle 5.1.

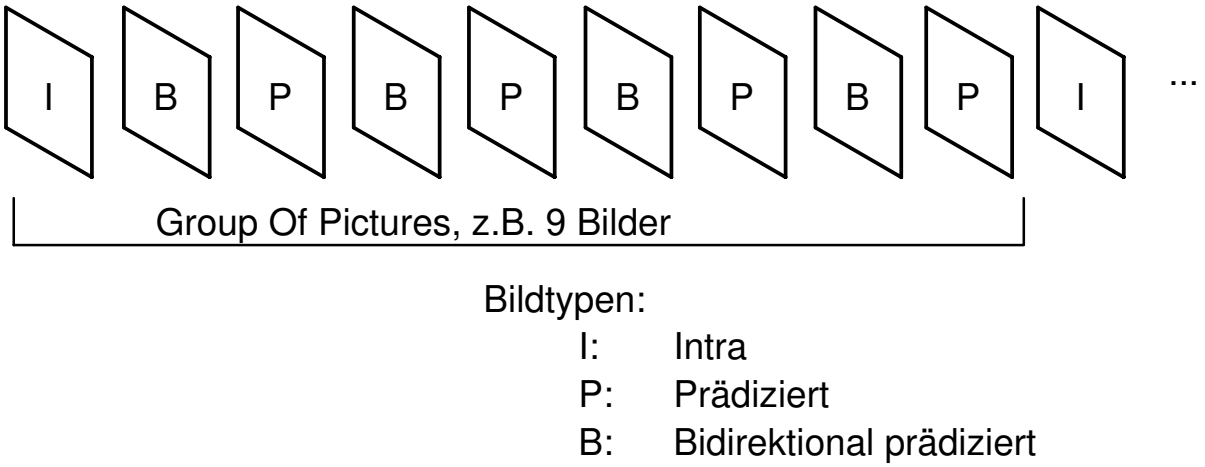


Bild 5.6 MPEG–Group Of Pictures (GOP)

Der MPEG–1 Coder teilt die Bewegtbildsequenz wie in Bild 5.6 dargestellt in sogenannte Group Of Pictures (GOP) ein. Die Bilder eines GOP werden in unterschiedlichen Modi verarbeitet. Durch Abschalten der Prädiktion werden sämtliche Makroblöcke eines I–Bildes durch eine DCT–Infraframecodierung verarbeitet. Während in P–Bildern eine bewegungskompensierende Hybridcodierung durchgeführt wird, werden B–Bilder durch Anwendung einer Hybridcodierung mit bidirektionaler bewegungskompensierender Prädiktion codiert, siehe Bild 5.7. Für jeden Makroblock eines B–Bildes kann die bewegungskompensierende Prädiktion entweder vorwärts, rückwärts oder bidirektional gewählt werden. In P– und B–Bildern kann die Verarbeitung für einzelne Makroblöcke auf den Intra–Modus geschaltet werden, falls die Prädiktion versagt.

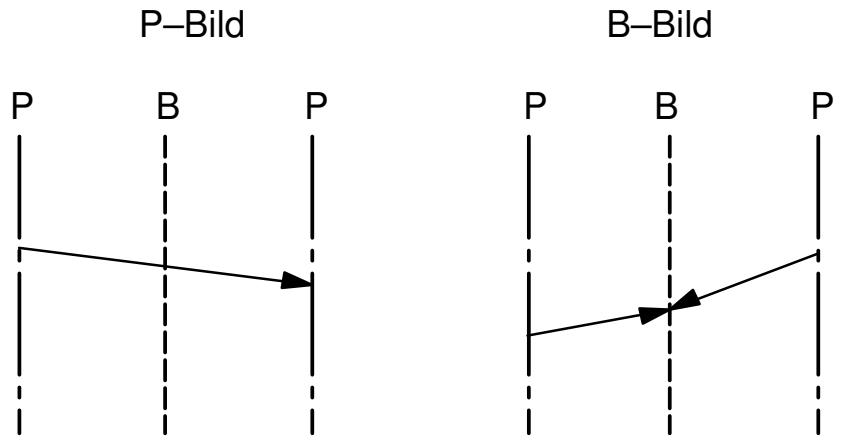


Bild 5.7 MPEG–1–Typen der bewegungskompensierenden Prädiktion

Für die Prädiktion in P–Bildern wird die zeitliche Referenz zu einem vorangehenden I– oder P–Bild gemacht, während B–Bilder ein vorangehendes und ein nachfolgendes Bild in der Reihenfolge der Bildwiedergabe jeweils vom Typ I oder P verwenden. B–Bilder werden nicht als zeitliche Referenz zur Prädiktion benutzt. Aufgrund der Bildtypen I, P, B ist ein Speicher für zwei Bilder im Decoder notwendig. Darüber hinaus ist die Verarbeitungsreihenfolge unterschiedlich zur Reihenfolge der Bildwiedergabe, wenn B–Bilder benutzt werden, z.B. wird nach Verarbeitung des ersten I–Bildes eines GOP in Bild 5.6 zunächst das erste P–Bild übertragen, welches dann von dem ersten B–Bild gefolgt wird.

Da es für eine GOP erforderlich ist, mit einem I–Bild in Verarbeitungsreihenfolge zu beginnen, kann eine Bewegtbildsequenz als eine Folge von individuellen GOP's aufgefaßt werden, wobei jeder GOP seine eigene Länge und Anzahl von I–,P– und B–Bildern besitzt. I–Bilder dienen als Einstiegspunkte für den wahlfreien Zugriff und ermöglichen so Trickmodi. Der Gebrauch von B–Bildern hat eine zusätzliche Verzögerungszeit zwischen Coder und Decoder zur Folge, was insbesondere für Echtzeitanwendungen wie z.B. Videokonferenzen unerwünscht ist. Eine Betriebsart mit geringen Verzögerungszeiten vergleichbar zu H.261 läßt sich erreichen, indem ein GOP mit unendlicher Länge gewählt wird, der nach dem ersten I–Bild nur noch P–Bilder enthält. Eine zu JPEG ähnliche Codierung der Bilddaten ist durch ausschließliche Wahl von I–Bildern möglich. In beiden Fällen wird die Effizienz des Codierungsalgorithmus reduziert, so daß ein Verlust an Bildqualität die Folge ist.

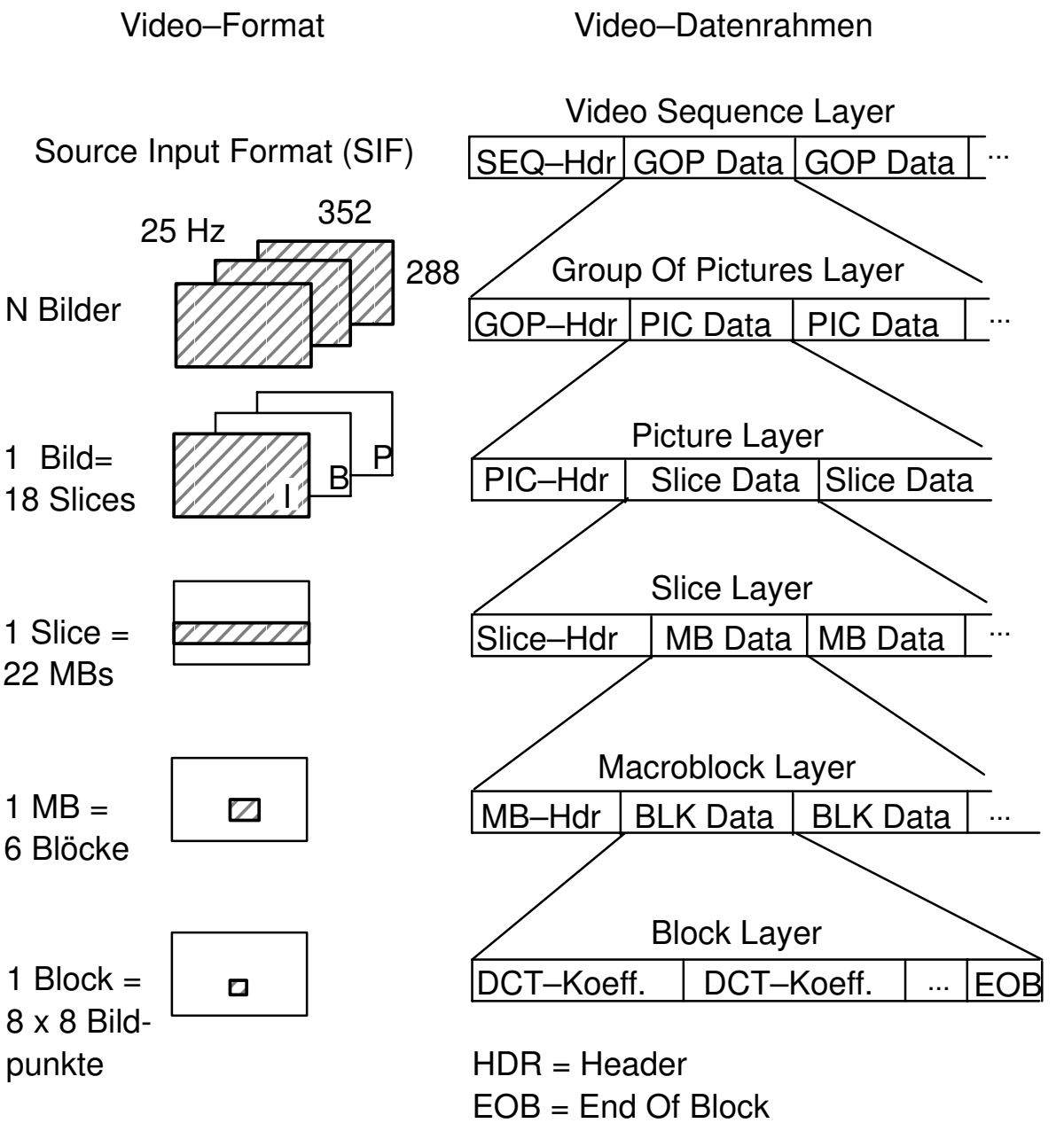


Bild 5.8 MPEG–I–Hierarchisch gegliederter Datenrahmen

Der MPEG–1 Coder erzeugt einen hierarchisch gegliederten Datenrahmen, siehe Bild 5.8. Dem Video Sequence Layer folgen der GOP Layer und der Picture Layer. Der Slice Layer wird zur Resynchronisation in Fällen von Fehlern des Übertragungskanals verwendet. Ein Slice kann, wie in Bild 5.8 dargestellt, aus einer Zeile von Makroblöcken bestehen, im allgemeinen kann jeder Slice seine eigene Länge

besitzen. Auf diese Weise läßt sich die Slicestruktur an die Statistik eines gestörten Kanals anpassen.

Die Prädiktionsmodi werden individuell für jeden Makroblock gewählt und im Macroblock Layer übertragen. Im Falle der unidirektionalen bewegungskompensierenden Prädiktion wird ein Displacementvektor gesendet; falls die bidirektionale bewegungskompensierende Prädiktion gewählt wird, werden zwei Displacementvektoren als zusätzliche Seiteninformation im Macroblock Layer codiert. Die Displacementvektoren können mit einer Genauigkeit von einem halben Bildpunkt angegeben werden. Luminanzwerte zwischen benachbarten Bildpunkten werden durch bilineare Interpolation berechnet.

Die Quantisierung kann dem Bildinhalt eines Makroblocks und der verfügbaren Datenrate durch Wahl einer geeigneten Quantisiererstufenbreite angepaßt werden. Die gewählte Stufenbreite wird als Seiteninformation im Macoblock Layer übertragen. Der MPEG–1–Standard spezifiziert zwei Quantisiererkennlinien für intra und non–intra codierte Makroblöcke. Die Multiplikation der Quantisiererstufenbreite mit den Elementen einer Wichtungsmatrix gestattet die gewichtete Quantisierung von DCT–Koeffizienten entsprechend ihren Frequenzen innerhalb eines 8 x 8 Blocks. Eine Intra– und Non–Intra–Wichtungsmatrix können im Video Sequence Layer übertragen werden. Der Macroblock Layer wird gefolgt von Blockdaten für die 4 Luminanz– und die 2 Chrominanzblöcke eines Makroblocks.

Im Block Layer werden die quantisierten DCT–Koeffizienten in Reihenfolge des Zigzag–Scans für jeden Block codiert. Analog zu H.261 wird eine feste Codeworttabelle zur Codierung der zweidimensionalen Ereignisse (Null–Lauflänge, Amplitude) verwendet. Die Übertragung von Blockdaten wird mit einem EOB–Codewort beendet. DC–Werte von intra–codierten Makroblöcken werden mittels DPCM ähnlich zu JPEG codiert.

Die zugewiesene Datenmenge für jedes Bild in einem GOP kann erheblich schwanken und ist abhängig vom Bildtyp. Im Vergleich zu B–Bildern benötigen I– und P–Bilder eine größere Datenmenge, um eine ausreichende und ausgeglichene Bildqualität in einem GOP zu gewährleisten.

Der MPEG–1 Coder verarbeitet eine SIF–Sequenz typischerweise mit einem Kompressionsverhältnis von ca. 26:1 und erzielt dabei eine Datenrate von 1,15 Mbit/s.

5.1.4 Der MPEG–2–Standard zur Codierung von Fernsehsignalen

In Phase II ihrer Aktivitäten hat die ISO–Motion Picture Exports Group (MPEG) einen Standard zur Codierung von Videosignalen gemäß dem CCIR 601 Format für Fernsehanwendungen entwickelt, siehe Tab. 5.1. Er wird für das digitale Fernsehen sowie zur Aufzeichnung auf der Digital Versatile Disc (DVD) verwendet. In Erweiterung des MPEG–1–Standard berücksichtigt dieser Standard die spezifischen Eigenschaften eines Abtastrasters mit Zeilensprung.

Wie beim MPEG–1–Standard wird eine Bildsequenz in Group of Pictures (GOP) von Vollbildern unterteilt. In P– und B–Vollbildern kann die bewegungskompensierende Prädiktion beim MPEG–2–Standard makroblockweise jeweils im Vollbild– oder im Halbbildraster erfolgen, siehe dazu Bild 5.9.

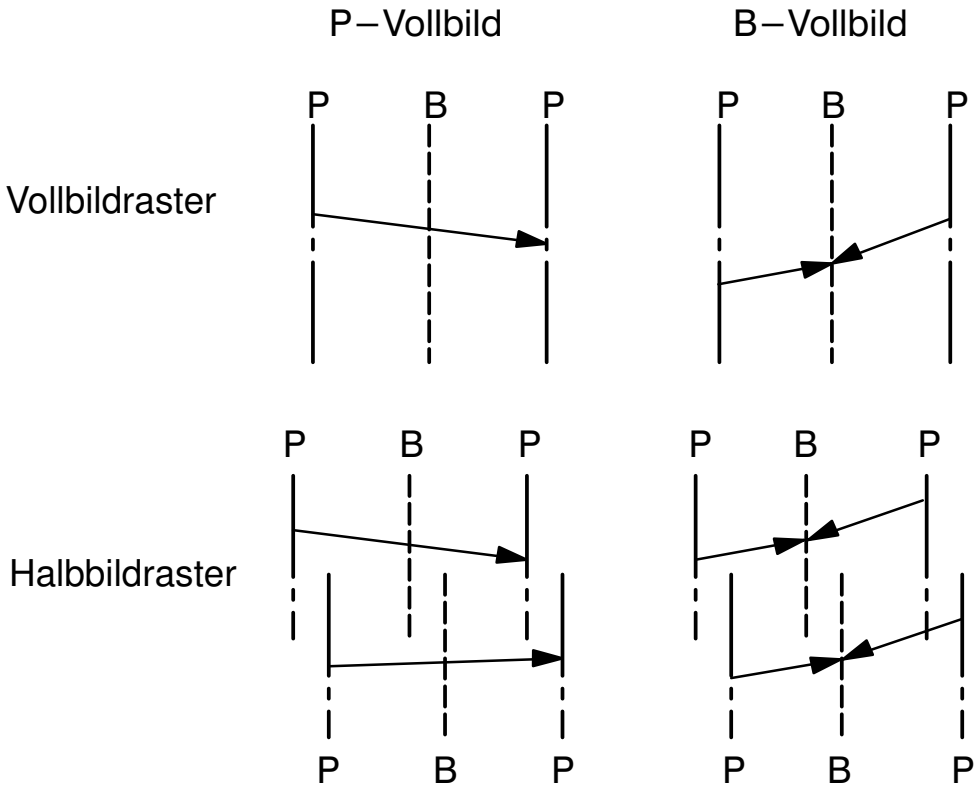


Bild 5.9 MPEG–2–Typen der bewegungskompensierenden Prädiktion im Vollbild– und Halbbildraster

Der MPEG–2–Coder erzielt ein Kompressionsverhältnis von etwa 28:1 und ermöglicht damit die Übertragung eines Videosignals im CCIR–Format mit 6 Mbit/s bei sehr hoher Bildqualität.

5.2 Der MPEG–1–Standard zur Codierung von Audiosignalen

Im Rahmen von ISO–MPEG hat eine internationale Gruppe von Audio–Codierungs–Experten einen Standard für die Codierung von Audiosignalen erarbeitet, der sowohl für audiovisuelle als auch für reine Audioanwendungen eingesetzt werden kann.

Die digitale Darstellung eines stereophonen Audiosignals im Studioformat erfordert eine Abtastfrequenz von 48 kHz und eine gleichförmige Quantisierung entsprechend 16 bit pro Abtastwert [30]. Daraus resultiert eine Datenrate von 768 kbit/s für ein Monosignal und entsprechend 2×768 kbit/s, also etwa 1,5 Mbit/s für ein Stereosignal. Aufgrund der etwas geringeren Abtastfrequenz von 44,1 kHz ergibt sich für das Stereosignal einer Compact Disc eine Datenrate von 2×706 kbit/s, also etwa 1,4 Mbit/s.

Die Hauptzielsetzung sah vor, bei Codierung eines Stereosignals mit einer von 1,5 Mbit/s auf 256 kbit/s reduzierten Datenrate eine der Compact Disc vergleichbare Audioqualität zu erreichen.

Der entwickelte MPEG–Audio–Codierungsstandard besteht aus drei Layern, wobei Komplexität und Codierungseffizienz von Layer I zu Layer II und Layer III jeweils zunimmt.

5.2.1 Das grundlegende Codierungskonzept

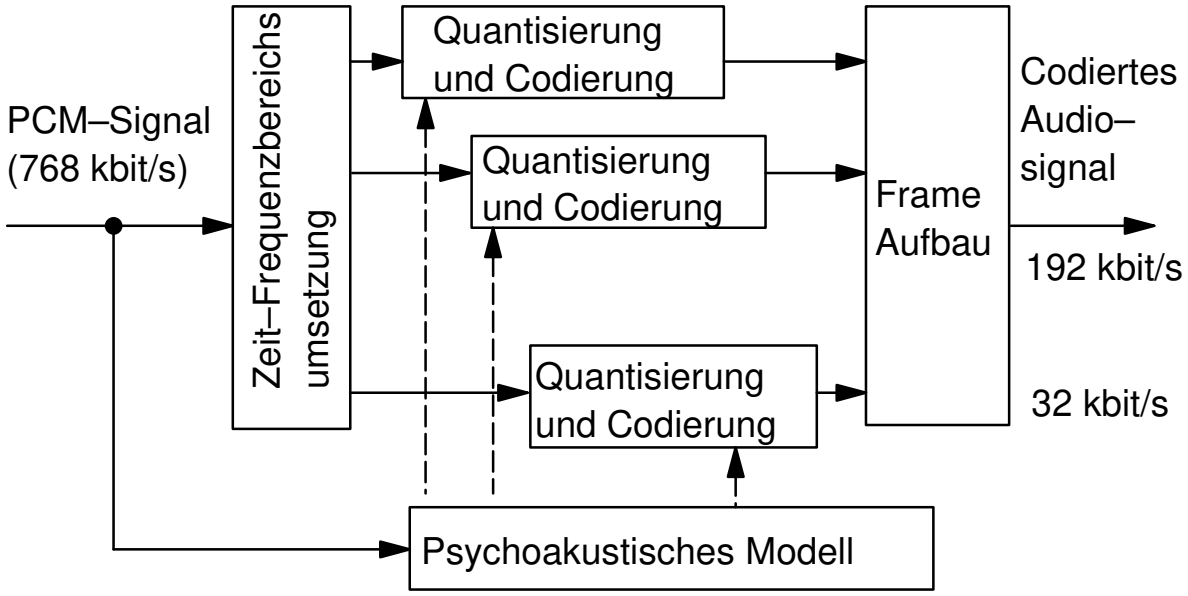


Bild 5.10 Generelles Blockschaltbild eines Audio Coders

Bild 5.10 zeigt das allgemeine Blockschaltbild eines Audio–Coders. Zunächst wird das am Eingang eingespeiste PCM–Audiosignal aus dem Zeitbereich in eine Darstellung im Frequenzbereich umgesetzt, was einer Zerlegung des Eingangssignals in Spektralkomponenten entspricht. Eine ähnliche Zerlegung in die sog. Frequenzgruppen wird auch vom menschlichen Gehör bei der Wahrnehmung von Tonsignalen vorgenommen. Anschließend wird jede aus der Zeit–Frequenzbereichs–Umsetzung hervorgegangene Spektralkomponente individuell entsprechend den aktuellen Maskierungseigenschaften des Gehörs quantisiert und codiert. Die Steuerinformation für die Quantisierer wird dabei aus einer Schätzung der aktuellen signalabhängigen Maskierungsschwelle gewonnen, die parallel zur Zeit–Frequenzbereichs–Umsetzung vom psychoakustischen Modell berechnet wird. Die Maskierungsschwelle gibt die maximal erlaubte Quantisierungsfehlerleistung für jede einzelne Spektralkomponente an. Die Quantisierung ist optimal, wenn jede Spektralkomponente exakt mit der durch die Maskierungsschwelle vorgegebenen Genauigkeit quantisiert und codiert wird. In diesem Fall wird das dem Nutzsignal überlagerte Quantisierungsgeräusch vom menschlichen Gehör gerade noch nicht wahrgenommen.

Die Zeit–Frequnezbereichs–Umsetzung kann entweder mit einer Filterbank oder einer Transformation oder auch einer Kombination von beiden realisiert werden [32], [33]. Zur Gewährleistung einer effizienten Codierung ist es erforderlich, daß die Anzahl der zu codierenden Spektralkomponenten in jedem Zeitintervall nicht größer ist als die Anzahl der Abtastwerte des Eingangssignals. Durch den Einsatz von Filterbänken mit Abtastratenreduktion und Aliasing–Kompensation kann dieses Problem gelöst werden.

Das psychoakustische Modell besitzt die im Abschnitt 4.2.1 beschriebene Struktur. Es berechnet jeweils für kurze Zeitintervalle des PCM–Eingangssignals einen Schätzwert für die aktuelle signalabhängige Maskierungsschwelle. Für jede Spektralkomponente gibt die Maskierungsschwelle im zugehörigen Frequenzintervall die gerade noch wahrnehmbare Quantisierungsfehlerleistung an. Diese wird dann zur Generierung der Steuerinformation für die dynamische Bit– bzw. Rauschleistungszuweisung verwendet, indem die aktuell erforderliche Anzahl der Quantisierungsstufen für jede Spektralkomponente und jedes Zeitintervall bestimmt wird.

5.2.2 Die drei Layer des MPEG–Audio–Codierungsstandards

Im folgenden werden die drei Layer beschrieben. Die vollständige Spezifikation des Standards ist in [26] zu finden.

5.2.2.1 Layer I

Bild 5.11 zeigt das Blockschaltbild des Layer I Coders. Das Eingangssignal wird blockweise in sog. Frames zu je 384 PCM–Abtastwerten verarbeitet, was bei einer Abtastfrequenz von 48 kHz einem Zeitintervall von 8 ms entspricht. Die Zeit–Frequenzbereichs–Umsetzung wird durch eine Polyphasenfilterbank mit Aliasing–Kompensation [34], [35] realisiert, die auf einem Finite Impulse Response (FIR) Prototypfilter der Länge 512 basiert. Das Eingangsignal wird in 32 gleichbreite Teilbänder zerlegt. Bei einer Abtastfrequenz von 48 kHz entspricht dies einer Bandbreite von 750 Hz je Teilband. Da die Filterbank eine geeignte Abtastratenreduktion beinhaltet, werden für jeden Frame mit 384 PCM–Abtastwerten in jedem der 32 Teilbänder jeweils 12 Teilband–Abtastwerte generiert [36], [37]. Jedes Teilbandsignal wird anschließend individuell codiert.

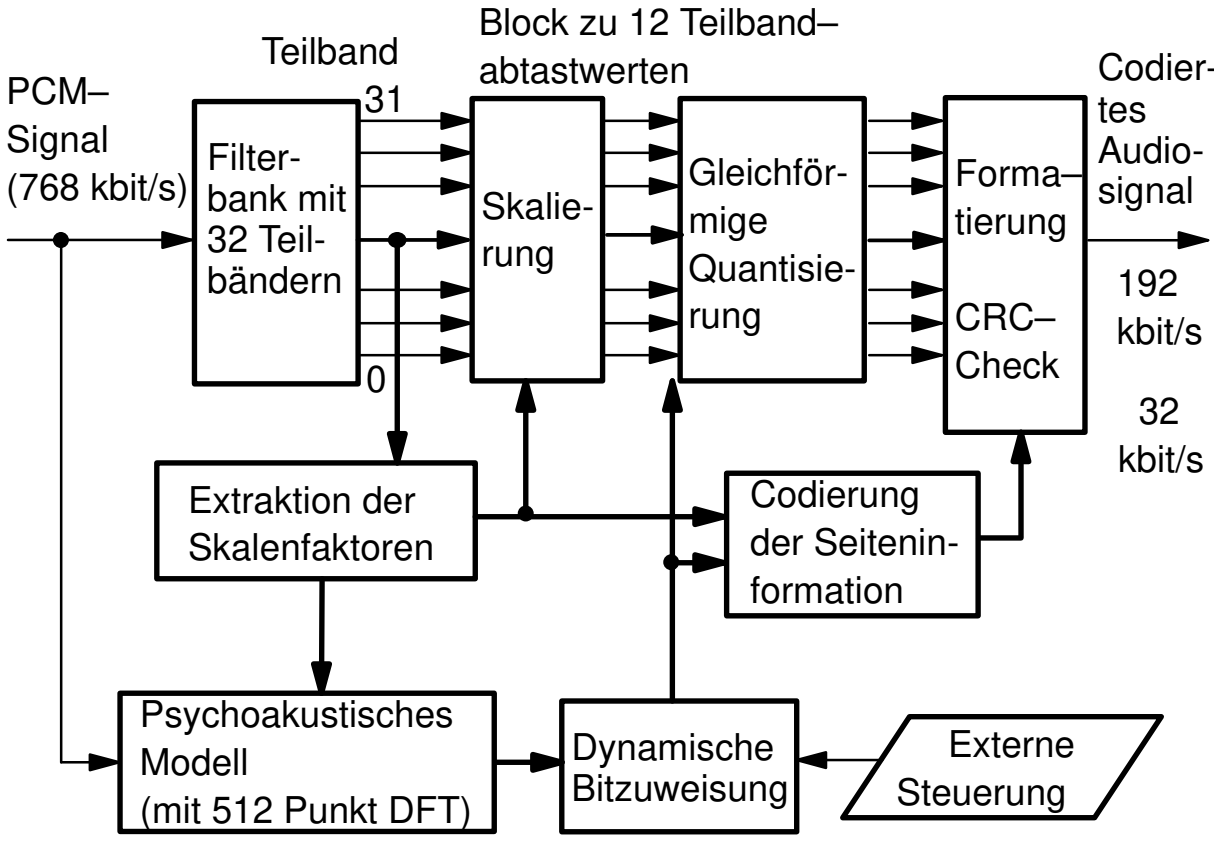


Bild 5.11 Blockschaltbild des ISO–MPEG Layer I Coders

Für die Quantisierung der Teilband–Abtastwerte wird die Technik der Blockkompandierung eingesetzt. Dazu wird für jeden Block von 12 aufeinanderfolgenden Teilband–Abtastwerten ein Skalenfaktor berechnet, indem der maximale Absolutwert dieser 12 Teilband–Abtastwerte bestimmt und mit 6 bit codiert wird. Die Teilband–Abtastwerte werden so skaliert und mit der vom psychoakustischen Modell gesteuertn dynamischen Bitzuweisung quantisiert. Entsprechend der Anzahl der zugewiesenen Bits wird entweder einer von 14 verschiedenen gleichförmigen Quantisierern mit 2^{n-1}, $2 \leq n \leq 15$ Quantisierungsstufen verwendet oder das Teilband wird nicht übertragen falls keine Bits zugewiesen wurden. Innerhalb eines Frames wird die Bitzuweisungsinformation für jedes Teilband mit jeweils 4 bit codiert.

Bei dem in Layer I verwendeten psychoakustischen Modell handelt es sich um eine auf einer 512–Punkt DFT basierende Realisierung mit geringer Komplexität, die zusätzlich auch die Skalenfaktoren zur Steuerung der Bitzuweisung auswertet.

Die für einen Frame zu übertragende Seiteninformation besteht aus der Bitzuweisungsinformation für alle 32 Teilbänder, d.h. 32 mit jeweils 4 bit codierte Werte, und den Skalenfaktoren für alle Teilbänder mit von Null verschiedener Bitzuweisung, d.h. bis zu 32 mit jeweils 6 bit codierte Werte.

5.2.2.2 Layer II

Der Layer II Coder ist verglichen mit dem Layer I komplexer, bietet dafür jedoch eine verbesserte Codierungseffizienz. Die Struktur des Layer II Coders ist der des Layer I sehr ähnlich.

Bild 5.12 zeigt das Blockschaltbild des Layer II Coders. Die grau hinterlegten Blöcke kennzeichnen zusätzliche Blöcke, die in Layer I nicht verwendet werden und die schraffierten Blöcke kennzeichnen modifizierte Module.

Das Eingangsignal wird in Frames zu je 1152 PCM–Abtastwerten verarbeitet, was bei einer Abtastfrequenz von 48 kHz einem Zeitintervall von 24 ms entspricht. Die Zeit–Frequenzbereichs–Umsetzung ist identisch mit der in Layer I, aber entsprechend der verdreifachten Frame–Länge werden jetzt in jedem Teilband 36 Teilband–Abtastwerte je Frame generiert. Auch hier wird wieder jedes Teilbandsignal individuell codiert, wobei allerdings eine verglichen mit Layer I verbesserte, aber auch kompliziertere Codierung der Teilband–Abtastwerte und der Seiteninformation angewendet wird.

Die 36 Teilband–Abtastwerte in jedem Teilband werden in drei Blöcke mit jeweils 12 Abtastwerten gruppiert. Für jeden Block mit 12 Abtastwerten wird – wie bei Layer I beschrieben – ein Skalenfaktor bestimmt, so daß man innerhalb eines Frames drei Skalenfaktoren für jedes Teilband erhält. Innerhalb eines Frames werden die drei aufeinanderfolgenden Skalenfaktoren in jedem Teilband gemeinsam betrachtet und in bestimmte Skalenfaktormuster klassifiziert. In Abhängigkeit vom jeweiligen Muster werden entweder ein, zwei oder drei Skalenfaktoren zusammen mit einer zusätzlichen Skalenfaktor–Auswahl–Information übertragen.

Ähnlich wie in Layer I werden die Abtastwerte in den Teilbändern mit von Null verschiedener Bitzuweisung nach der Skalierung gleichförmig quantisiert. Die Auswahl der Quantisierer wird von der durch das psychoakustische Modell gesteuerten dynamischen Bitzuweisung bestimmt. Das in Layer II verwendete psychoakustische Modell weist die gleiche Grundstruktur wie das des Layer I auf, ist aber komplexer und leistungsfähiger. Es basiert auf einer 1024–Punkt DFT.

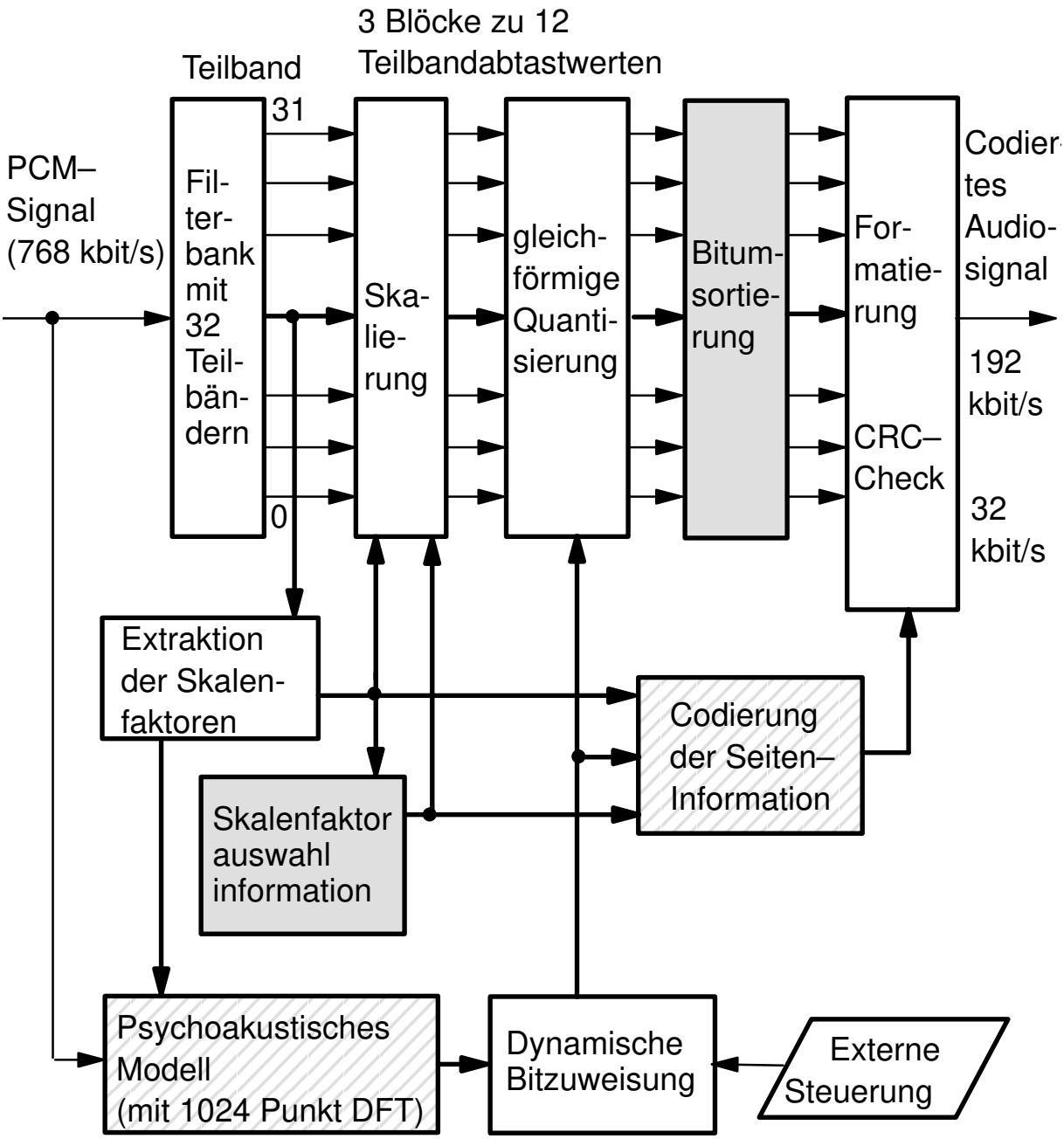


Bild 5.12 Blockschaltbild des ISO–MPEG Layer II Coders

In Layer II werden drei verschiedene Sätze von Quantisierern eingesetzt. Im unteren Frequenzbereich wird ein Satz mit 15, im mittleren Frequenzbereich ein Satz mit 7 und im oberen Frequenzbereich ein Satz mit 3 verschiedenen gleichförmigen Quantisierern verwendet. Entsprechend der Satzgröße werden jeweils 4, 3 oder 2 bit zur Codierung der Bitzuweisungsinformation benötigt. Teilband–Abtastwerte mit ge-

ringer Quantisierungsstufenzahl werden zu einer Gruppe zusammengefaßt, die dann mit einem Codewort codiert wird. Aufgrund der Gruppenbildung ist eine zusätzliche Bitumsortierung erforderlich, bevor das codierte Audiosignal zusammengesetzt wird.

Die für einen Frame zu übertragende Seiteninformation besteht aus der Bitzuweisungsinformation für alle 32 Teilbänder, d.h. 32 mit jeweils 4, 3 oder 2 bit codierte Werte, der Skalenfaktor–Auswahl–Informationen und den Skalenfaktoren für alle Teilbänder mit von Null verschiedener Bitzuweisung.

5.2.2.3 Layer III

Der Layer III Code erreicht die höchste Codierungseffizienz der drei Layer, weist aber auch die größte Komplexität und somit den höchsten Implementierungsaufwand auf. Bild 5.13 zeigt das Blockschaltbild des Layer III Coders. Die grau hinterlegten Blöcke kennzeichnen zusätzliche Blöcke, die in Layer II nicht verwendet werden und die schraffierten Blöcke kennzeichnen modifizierte Module.

Wie in Layer II wird das Eingangsignal in Frames zu je 1152 PCM–Abtastwerten verarbeitet, was bei einer Abtastfrequenz von 48 kHz einem Zeitintervall von 24 ms entspricht. Die Zeit–Frequenzbereichs–Umsetzung wird durch eine Hybrid–Filterbank [38] realisiert, die zusätzlich eine adaptive Fensterumschaltung [41] erlaubt.

Die Hybrid–Filterbank besteht aus einer Kaskadierung der Polyphasenfilterbank, wie sie in Layer I und II verwendet wird, mit einer Modifizierten Diskreten Cosinus Transformation (MDCT) [39]. Diese Hybrid–Filterbank gewährleistet einerseits Kompatibilität mit Layer I und II, bietet andererseits aber Möglichkeiten für eine verbesserte Spektralzerlegung. Die Abtastwerte am Ausgang eines jeden der 32 Teilbänder werden einer MCDT zugeführt, die dann 18 Spektralkoeffizienten für jedes Teilbandsignal erzeugt. Dies führt zu einer spektralen Auflösung von 32 x 18 = 576 Spektralkoeffizienten. Meistens wird diese hohe spektrale Auflösung verwendet, da sie für stationäre Signale den maximalen Codierungsgewinn ermöglicht. Erfordert das Eingangssignal jedoch eine höhere zeitliche Auflösung, z.B. aufgrund von Anschlägen, so kann die spektrale Auflösung der MCDT für ausgewählte oder alle Teilbänder durch Einsatz der adaptiven Fensterumschaltung auf 6 Spektralkoeffizienten reduziert werden. Die Frequenzauflösung wird somit bis auf 32 x 6 = 192 Spektralkoeffizienten reduziert werden, was einer entsprechend höheren zeitlichen Auflösung entspricht.

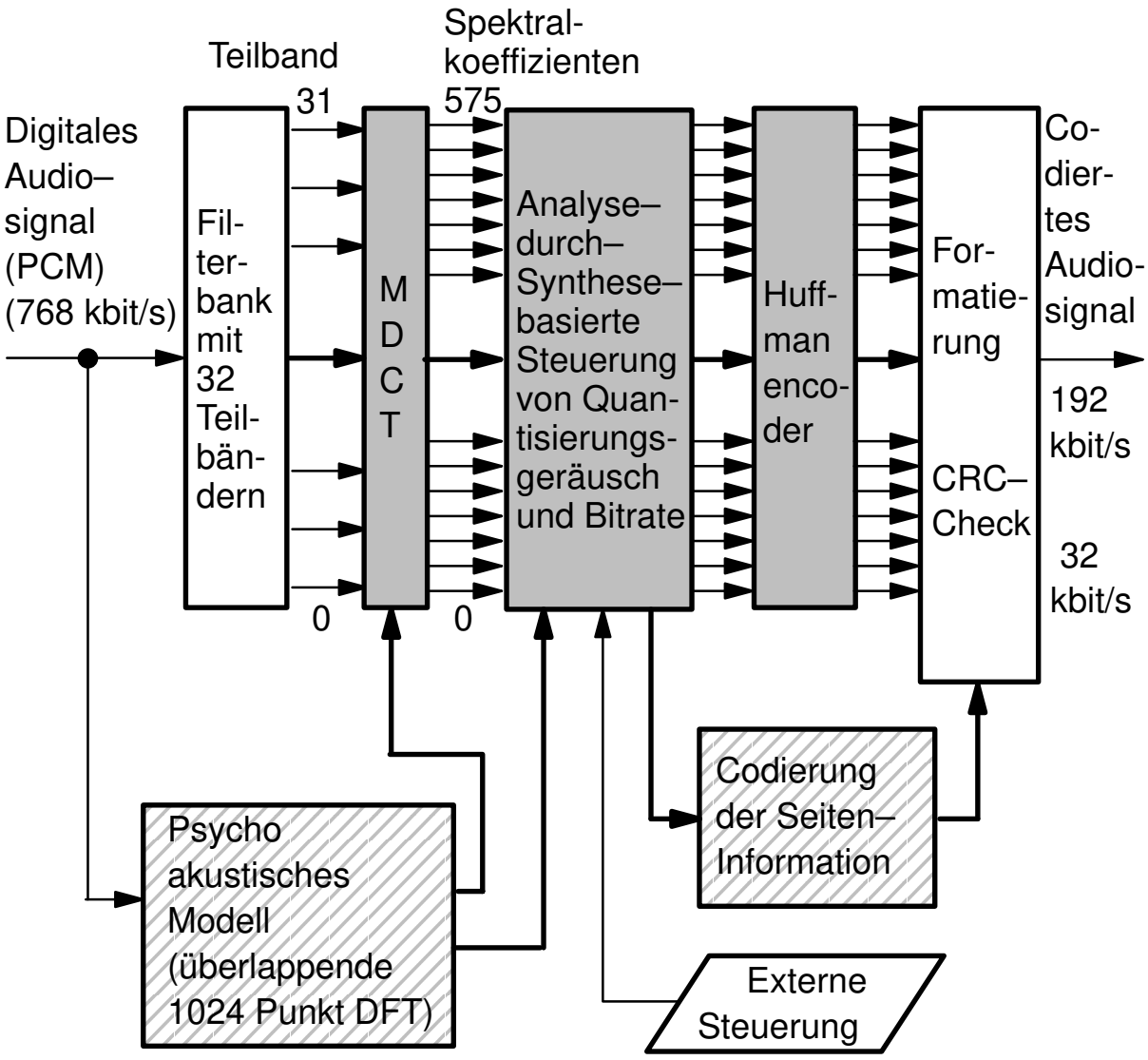


Bild 5.13 Blockschaltbild des ISO–MPEG Layer III Coders

Die Spektralkoeffizienten werden mit Hilfe einer durch das psychoakustische Modell gesteuerten Analyse–durch–Synthese Technik codiert. In der Analyse–durch–Synthese Prozedur wird der aktuelle Quantisierungsfehler mit der vom psychoakustischen Modell berechneten Maskierungsschwelle verglichen und die Quantisierung entsprechend angepaßt. Dabei wird eine ungleichförmige Quantisierung eingesetzt. Die quantisierten Spektralkoeffizienten werden dann einer Huffman–Codierung und einer Lauflängencodierung unterzogen. Zur Verbesserung der Codierungseffizienz wird zusätzlich eine Pufferspeichertechnik, das sog. Bit–Reservoir, verwendet, die eine von Frame zu Frame leicht variierende Datenrate mit

konstantem Langzeitmittelwert erlaubt. Dabei werden Bits aus Frames, in denen das Signal weniger kritisch ist, aufgespart, um sie dann bei der Codierung von Frames zu verwenden, in denen das Signal sehr kritisch ist [40].

Das in Layer III verwendete psychoakustische Modell ist das leistungsfähigste der drei Modelle. Es basiert auf einer überlappenden 1024–Punkt DFT, wobei die Fensterlänge 1024 beträgt, die Berechnung aber für jeweils 576 Abtastwerte synchron zur Hybrid–Filterbank erfolgt. Die separate Berechnung der Frequenzbereichsdarstellung ist notwendig, da die Ausgangswerte der Hybrid–Filterbank nicht direkt in eine für die Berechnung des Tonalitätsmaßes benötigte Betrags–Phasen–Darstellung überführt werden können. Zur Berechnung der Tonalität werden Betrag und Phase eines Spektralkoeffizienten von zwei seitlich vorausgehenden extrapoliert und als Maß für die Tonalität verwendet.

5.2.3 Ergebnisse der subjektiven Hörtests

Für ein stereophones Tonsignal wird eine der Compact Disc vergleichbare Qualität bei einer Datenrate von 384 kbit/s mit einem Layer I Code und bei einer Datenrate von 256 kbit/s mit einem Layer II und Layer III Code erreicht.

Bei einer Datenrate von 128 kbit/s für ein stereophones bzw. 64 kbit/s für ein monophones Tonsignal erreicht der Layer III Code eine Qualität, die der der Compact Disc sehr nahe kommt. Lediglich bei einigen kritischen Testsignalen werden Codierungsfehler wahrgenommen.

6. Literaturverzeichnis

[1] Berger, T., *Rate Distortion Theory, Prentice Hall*, Inc. Englewood Cliffs, New Jersey, (1971)

[2] Chahabadi, D., *Adaption einer Vektorquantisierung für breitbandige Sprachsignale mittels affiner Transformation*, Dissertation, Universität Hannover, VDI–Verlag, (1992)

[3] Edler, B., *Äquivalenz von Transformation und Teilbandzerlegung in der Quellencodierung*, Dissertation, Universität Hannover, (1994)

[4] Egan, J.P., Hake, H.W., *On the masking pattern of a single auditory stimulus*, Acoust. Soc. Am., Vol. 22, 622–630

[5] Fano, R.M., *Informationsübertragung*, R.Oldenburg Verlag, München, Wien 1966

[6] Gallager, R.G., *Information Theory and Reliable Communication*, J. Wiley and Sons, 1968

[7] Girod, B., *The Efficiency of Motion–Compensating Coding Prediction for Hybrid Coding of Video Sequences*, IEEE Journal on Selected Areas in Communication, Vol. SAC–5, No. 7, August 1987

[8] Girod, B., *Ein Modell der menschlichen visuellen Wahrnehmung zur Irrelevanzreduktion von Fernsehluminanzsignalen*, Dissertation, Universität Hannover, VDI–Fortschritt–Berichte Reihe 10, Nr. 84, 1988

[9] Green, D.M., *Profile analysis; auditory intensity discrimination*, New York, Oxford: Oxford University Press, 1988

[10] Huffman, D.A. *A Method for the Construction of Minimum Redundancy Codes*, Proc. IRE, No.40, pp. 1098–1101, (1952)

[11] Linde, Y., Buzo, A. and Gray, R.M. *An Algorithm for VQ Design*, IEEE Transactions on Communications, COM–28, No. 1, pp. 84–95, (1980)

[12] Max J., *Quantizing for Minimum Distortion*, IRE Transactions on Information Theory, Vol. IT–6, pp. 7–12, March 1960

[13] Panter, P. F., Dite, W. *Quantization Distortion in Pulse Count Modulation with nonuniform Spacing of Levels*, Proc. IRE, Vol. 35, pp. 44–48, (1951)

[14] Preuß, D., *Ein Beitrag zur Quellencodierung von zweistufigen Faksimilesignalen*, Dissertation, Universität Hannover, (1976)

[15] Shannon, C.E., *A Mathematical Theory of Communication*, Bell System Tech. J., Vol. 27, pp. 379–423 (Part I), pp. 623–656 (Part II), (1948).
Reprinted in book form by W.Weaver, Univ. of Illinois Press, Urbana, (1949)

[16] Shannon, C.E., *Coding Theorems for a Discrete Source with a Fidelity Criterion,* IRE Nat'l. Conv. Rec., Part 4, pp. 142–163, (1959)

[17] Zwicker,E., Fastl. H., *Psychoacoustics. Facts and models.*, Springer–Verlag (1990)

[18] Musmann, H.G., *Informationstheorie*, Manuskript einer Vorlesung an der Universität Hannover

[19] Jayant, N. S., Noll, P., *Digital Coding of Waveforms*, Prentice–Hall, 1984, TIB: FH 3625

[20] Netravali, A.N., Haskell, B.G., *Digital Pictures*, Plenum Press, 1988, TIB: mat 832 AJ 4348

[21] Pratt, W. K., *Image Transmission Techniques*, Academic Press, 1979, TIB: ZA 1971:Spl(12)

[22] Gray, R. M., *Source Coding Theory*, Kluwer Academic Press, 1990

[23] Bierling, M., *Hierarchische Displacementschätzung zur Bewegungskompensation in digitalen Fernsehbildsequenzen*, Dissertation, Universität Hannover, 1991, VDI–Verlag, TIB: ZN 6951(179)

[24] Werner, O., *Arithmetische Codierung für Bildsignale*, Diplomarbeit am Institut für Theoretische Nachrichtentechnik, 1990

[25] ISO/IEC, *Draft International Standard DIS 10918: Information technology–coded representation of digital continuous–tone still pictures* ISO/IEC JTCI/SC29/WG10, Geneva, Januar 1992

[26] SO/IEC, *Draft International Standard DIS 11172: Information technology–coding of moving pictures and associated audio for digital storage medi at up to about 1.5 Mbit/s*, ISO/IEC JTCI/SC29/WG11, Geneva, April 1992

[27] Wallace, G.K., *The JPEG still picture compression standard*, Comm. ACM, April 1991

[28] CCITT, *Video codec for audio visual services at* $p \times 64$ *kbits/s*, CCITT Recommendation H261. Geneva 1990

[29] CCITT, *Draft revision of Recommendation H261:* Video codec for audio visual services at $p \times 64$ kbits/s, Signal Processing: Image Communication, vol 2, no. 2, pp221–240, August 1990

[30] CCIR, *Source encoding for digital sound signals in broadcasting studios*, CCIR Recommendation 646, Geneva, 1986

[31] Blauert, J., *Spatial hearing*, The MIT Press, Cambridge, Massachusetts, 1983

[32] Brandenburg, K., Edler, B., Eberlein, E.,Herre, J., *Comparison of filter banks for high quality audio coding*, ISCAS '92, San Diego, May 1992

[33] Sporer, T., Brandenburg, K., Edler, B., *The use of multirate filter banks for coding of high quality digital audio*, EUSIPCO '92, Bruxelles, September 1992

[34] Rothweiler, J.H., *Polyphase quadrature filters – a new subband coding technique*, IEEE Proc. ICASSP 1983, pp. 2161–2164, Boston, 1983

[35] Nussbaumer, H.J., Verrerli, M., *Computationally efficient QMF filter banks*, IEEE Proc. ICASSP 1984, pp. 11.3.1–4, 1984

[36] Stoll, G., Wiese, D., *Bit rate reduction of high quality audio signals by modeling the ears masking threasholds*, 89th AES–Convention, Los Angeles, Preprint no 2970, September 1990

[37] Dehery, Y.F., Stol, G., Kerkhof, L.v.d., *MUSICAM source coding for digital sound*, Symposium Record "Broadcast Sessions" of the 17th International Television Symposium, Montreux, pp. 612–617, June 1991

[38] Brandenburg, K., Johnston, J.D., *Second generation of perceptualaudio coding: the hybrid coder*, 88th AES–Convention, Montreux, Preprint no. 2937, June 1990

[39] Princen, J.P., Johnston, J.D., Bradley, A.B., *Subband/transform coding using filter bank designs based on time domain aliasing cancellation*, IEEE Proc. ICASSP 1987, pp. 2161–2164, 1987

[40] Brandenburg, K., Johnston, J.D., et al, *ASPEC: Adaptive spectral perceptual entropy coding of high quality music signals*, 90th AES–Convention, Paris,Preprint no. 3011, Februar 1991

[41] Edler, B., *Codierung von Audiosignalen mit überlappender Transformation und adaptiven Fensterfunktionen*, Frequenz vol, 43, pp. 252–256, 1989