

HEINRICH-HERTZ-INSTITUT — BERLIN-CHARLOTTENBURG

Technischer Bericht Nr. 181

Adaptive Blockquantisierung von Sprachsignalen

von

Dipl.-Ing. Rainer Zelinski

Dr.-Ing. Peter Noll

Berlin

1 9 7 5

1 BERLIN 10 EINSTEINUFER 37

HEINRICH-HERTZ-INSTITUT

Berlin-Charlottenburg

Technischer Bericht Nr. 181

Adaptive Blockquantisierung von Sprachsignalen

von

Dipl.-Ing. R. Zelinski

Dr.-Ing. P. Noll



Berlin

1975

Adaptive Blockquantisierung von Sprachsignalen

Zusammenfassung:

Bei der digitalen Sprachübertragung kann der Nachrichtenfluss erheblich reduziert werden, wenn das Codiersystem an die sich ändernde Signalstatistik angepasst wird.

In der vorliegenden Arbeit werden verschiedene Verfahren der adaptiven Blockquantisierung untersucht. Dabei wird nicht nur eine Anpassung an den momentanen Signalpegel, sondern auch an das sich zeitlich ändernde Kurzzeitspektrum des Sprachsignals vorgenommen.

Die Ergebnisse zeigen, dass gegenüber einer PCM-Codierung mit logarithmischem Quantisierer eine Vergrößerung des Signal-Rauschabstandes um 17 ... 23 dB erreichbar ist. Die Übertragungsrate kann von 64 kbit/s bei PCM-Codierung auf 12 ... 16 kbit/s bei adaptiver Blockquantisierung verringert werden, ohne dass subjektiv eine Verschlechterung der Sprachqualität feststellbar ist.

Heinrich-Hertz-Institut

Die Bearbeiter

R. Zelinski

(Dipl.-Ing. R. Zelinski)

Der Abteilungsleiter

i. V. Noll

(i. V. Dr.-Ing. P. Noll)

P. Noll

(Dr.-Ing. P. Noll)

Der Institutsdirektor

H. Ohnsorge

(Dr.-Ing. H. Ohnsorge)

Berlin-Charlottenburg, den 28. Mai 1975



Inhaltsverzeichnis

	<u>Seite</u>
1. <u>Einleitung</u>	1
2. <u>Blockquantisierung bei stationären Signalquellen</u>	3
2.1 <u>Basisbeschränkte Transformationssysteme</u>	4
2.1.1 Statistische Kenngrößen im Zeit- und Bildbereich	6
2.1.2 Bestimmung der optimalen Bitzuordnung	8
2.1.3 Theoretischer Gewinn gegenüber einer PCM-Codierung	12
2.2 <u>Eigenschaften spezieller Transformationen</u>	13
2.2.1 Die Loève-Karhunen-Transformation	13
2.2.2 Andere Transformationen	14
2.2.3 Darstellung der Basisvektoren	19
2.3 <u>Vergleich der verschiedenen Transformationen</u>	21
2.3.1 Zusammenhang mit dem Gewinn bei einer Prädiktionscodierung	21
2.3.2 Theoretische Gewinne für einen autoregressiven Prozess endlicher Ordnung	22
2.3.3 Vergleich der Spektren im Bildbereich	26
2.3.4 Rechenaufwand zur Durchführung der Transformationen	28
2.4 <u>Realisierung der Bitzuordnung bei zusätzlichen Randbedingungen</u>	29
3. <u>Codierung von Sprachsignalen mit nichtadaptiver Blockquantisierung</u>	32
3.1 Beschreibung des Simulationssystems	32
3.2 Ergebnisse	33
3.3 ADV der Spektralkoeffizienten	35

	<u>Seite</u>
4. <u>Blockquantisierung mit adaptiven Quantisierern</u>	37
4.1 Beschreibung des Codiersystems	37
4.2 ADV der Spektralkoeffizienten	38
4.3 Ergebnisse	40
4.4 Quantisierung des Verstärkungsfaktors V zur Pegelsteuerung	42
5. <u>Blockquantisierung mit adaptiven Quantisierern und adaptiver Bitzuordnung</u>	44
5.1 <u>Vorbetrachtungen an Modellquellen</u>	45
5.1.1 Theoretische Gewinne für verschiedene Sprachlaute	45
5.1.2 Suboptimaler Gewinn bei kleinen Bitraten	48
5.2 <u>Beschränkt adaptive Blockquantisierung</u>	52
5.2.1 Beschreibung des Codiersystems	52
5.2.2 Ergebnisse	55
5.3 <u>Fehlanpassungen in der Signalstatistik</u>	56
5.4 <u>Volladaptive Blockquantisierung</u>	58
5.4.1 Beschreibung des Codiersystems	59
5.4.2 Ergebnisse	64
5.4.3 ADV der Spektralkoeffizienten	67
5.4.4 Quantisierung der Zusatzinformation über das Kurzzeitspektrum	69
5.4.5 Algorithmus für kleine Blocklängen	72
6. <u>Codierung von Sprachsignalen mit nichtadaptiver und adaptiver Blockquantisierung und Vergleich mit anderen Verfahren</u>	75
6.1 Signal-Rausch-Abstand bei verschiedenen Übertragungsraten	76
6.2 Zeitlicher Verlauf des Segment-SNR	81
6.3 Subjektive Bewertung der Codierverfahren	85
6.4 Vereinfachungen für die Realisierung der volladaptiven Blockquantisierung	86

	<u>Seite</u>
6.4.1 Einsatz von linearen Quantisierern	86
6.4.2 Begrenzung der maximalen Bitzahl	88
6.5 Vergleich des Rechenaufwandes bei voll- adaptiver Blockquantisierung und bei ADPCM-AQ	91
7. <u>Sprachübertragung mit volladaptiver Block- quantisierung bei sehr kleinen Übertra- gungsraten</u>	93
7.1 Bitzuordnung nach einem objektiven Kriterium	93
7.2 Bitzuordnung nach subjektiven Kriterien	96
7.3 Vergleich mit anderen adaptiven Transforma- tionscodierungen	99
8. <u>Einfluss von Kanalstörungen bei der voll- adaptiven Blockquantisierung</u>	102
8.1 Einfluss von Kanalstörungen bei Systemen mit und ohne Fehlerkorrektur	102
8.2 Vergleich mit anderen Codierverfahren	106
9. <u>Zusammenfassende Bewertung der volladap- tiven Blockquantisierung</u>	107
10. <u>Dank</u>	109
11. <u>Literaturverzeichnis</u>	110

1. Einleitung

Für die digitale Übertragung von Sprachsignalen kann die für eine bestimmte Qualität notwendige Übertragungsrate reduziert werden, wenn die im Sprachsignal enthaltene Redundanz eliminiert wird. Im Gegensatz zu Prädiktionsverfahren, die im Zeitbereich arbeiten /1,2/, wird bei der Blockquantisierung die Signalstatistik im Frequenz- bzw. ganz allgemein im Bildbereich ausgenutzt. Dazu wird die Folge der Abtastwerte des Sprachsignals segmentiert und jeweils ein Datenblock durch eine lineare Transformation in einen Bildbereich überführt. Die dort entstehenden Spektralkoeffizienten werden dann unabhängig voneinander quantisiert und in digitaler Form übertragen. Durch eine inverse lineare Transformation entsteht am Empfänger ein rekonstruiertes Signal, das bis auf den Quantisierungsfehler mit dem Ausgangssignal übereinstimmt. Die statistischen Abhängigkeiten des zu verarbeitenden Signals können z.B. dadurch ausgenutzt werden, dass Spektrallinien geringerer Varianz nicht übertragen bzw. nur sehr grob quantisiert werden. Die Übertragungsrate kann dadurch verringert werden, wobei die Verzerrung nur unwesentlich zunimmt. Zur Sprachübertragung mit nichtadaptiver Blockquantisierung liegen bereits einige Untersuchungsergebnisse vor /3,4,5/.

Da die Statistik des Sprachsignals sich zeitlich ändert, arbeitet der Codierer in der Regel nicht optimal. Durch adaptive Codieralgorithmen kann objektiv und vor allem subjektiv eine erhebliche Verbesserung erzielt werden. Diese Adaption wird in zwei Schritten durchgeführt:

- 1.) Der momentane Sprachpegel zeigt erhebliche Schwankungen. Durch eine pegelabhängige Aussteuerung und Einsatz eines Optimal-Quantisierers kann der Signal-Rausch-Abstand vergrößert werden. Dies gilt sowohl für Codierungen im Zeitbereich /6,7/ als auch im Frequenzbereich (Blockquantisierung /8/). Entsprechende

Verfahren wurden auch für lineare Transformationen zur Bildcodierung eingesetzt /9/.

- 2.) Das Kurzzeitspektrum von Sprache ändert sich zeitlich. Dies kann in ungünstigen Fällen dazu führen, dass bei einer Blockquantisierung mit zeitinvarianter Bitzuordnung der Quantisierer zu den einzelnen Spektrallinien gerade die Spektrallinien grob quantisiert oder nicht übertragen werden, die momentan eine hohe Varianz aufweisen. Eine Verbesserung wird durch eine adaptive Bitzuordnung erreicht, die in Abhängigkeit von dem vorliegenden Kurzzeitspektrum optimal gewählt wird. Vergleichbare Verfahren sind für die Bildcodierung untersucht worden /9,10,11/, für die Sprachcodierung liegen bisher nur wenige Ergebnisse vor /5,12/.

In der vorliegenden Arbeit wird untersucht, welche Verbesserungen durch verschiedene Verfahren einer adaptiven Blockquantisierung erreichbar sind. Die Motivation dazu ergibt sich aus folgenden Gesichtspunkten:

- * Verfahren, die sowohl eine Pegelsteuerung als auch eine adaptive Bitzuordnung aufweisen, sind bisher noch nicht eingesetzt worden.
- * In letzter Zeit sind zwei schnelle Transformationen bekannt geworden, die sich durch geringen Rechenaufwand (Diskrete Slant-Transformation /13/) bzw. durch einen hohen Gewinn (Diskrete Cosinus-Transformation /14/) auszeichnen. Über die Brauchbarkeit dieser Transformationen für die Sprachcodierung sind noch keine Ergebnisse bekannt.
- * Es sollen auch Algorithmen zur adaptiven Bitzuordnung eingesetzt werden, die bisher noch nicht für die Sprachcodierung untersucht worden sind.

- * Ein entsprechendes Codierverfahren im Zeitbereich ist die ADPCM mit adaptiver Prädiktion und adaptivem Quantisierer /1,2/. Die Ergebnisse bei der Codierung mit adaptiver Blockquantisierung und ADPCM sollen miteinander verglichen werden, um eine bessere Bewertung der adaptiven Blockquantisierung zu ermöglichen.

Alle Codierverfahren wurden auf dem Digitalrechner simuliert und auf mehrere Sprachproben angewendet. Die untersuchten Algorithmen beruhen alle auf einer Vorwärtsregelung: Die Parameter über die momentane Signalstatistik werden aus dem nichtcodierten Eingangssignal gewonnen und müssen daher zusätzlich zum Empfänger übertragen werden. Dadurch vergrößert sich der Nachrichtenfluss geringfügig; der Vorteil ist jedoch eine erheblich geringere Empfindlichkeit gegenüber Kanalstörungen. Besonderen Wert wurde dabei auf Algorithmen gelegt, die auch in Hardware mit nicht zu grossem Aufwand realisierbar sind. Eine pitchsynchrone Steuerung, die eine weitere Reduktion der Übertragungsrate ermöglicht /15/, ist daher nicht berücksichtigt worden.

2. Blockquantisierung bei stationären Signalquellen

Zur Codierung einer stationären Quelle arbeitet eine nicht-adaptive Blockquantisierung optimal. Für diesen Fall werden wichtige Grundlagen, wie Bestimmung der optimalen Bitzuordnung und theoretischer Gewinn gegenüber einer PCM-Codierung, angegeben. Die für die späteren Simulationen eingesetzten linearen Transformationen werden beschrieben und einige Eigenschaften diskutiert. Die gewonnenen

Ergebnisse bilden den Ausgangspunkt für eine adaptive Blockquantisierung, da das Sprachsignal als quasistationär betrachtet werden kann: Kurze Ausschnitte aus dem Signalverlauf können näherungsweise durch eine stationäre Modellquelle beschrieben werden.

2.1 Basisbeschränkte Transformationssysteme

Das zu codierende Signal besteht aus einer zeitdiskreten Zahlenfolge x_n , die z.B. durch äquidistante Abtastung eines Sprachsignals entsteht. Jeweils ein Block von Daten x_n der Länge N wird zu einem Vektor \underline{x} zusammengefasst. Das Codiersystem ist im Bild 2.1 gezeigt.

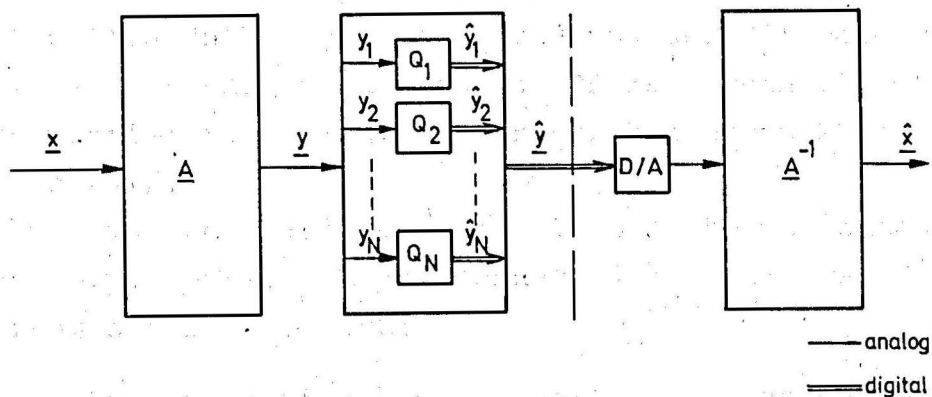


Bild 2.1 : Schema der Blockquantisierung.

Der Vektor \underline{x} wird durch die lineare Transformation

$$\underline{y} = \underline{A} \cdot \underline{x} \quad (2.1.1)$$

in den Bildbereich überführt. Die Matrix \underline{A} sei orthogonal; es gilt dann

$$\underline{A}^{-1} = \underline{A}^T \quad (2.1.2)$$

Die Komponenten von \underline{y} werden mit den N Quantisierern $Q_1 \dots Q_N$ quantisiert und in digitaler Form zum Empfänger übertragen. Der dort empfangene Vektor $\hat{\underline{y}}$ wird durch die Rücktransformation

$$\hat{\underline{x}} = \underline{A}^{-1} \cdot \hat{\underline{y}} \quad (2.1.3)$$

in den Zeitbereich abgebildet und stimmt bis auf den Quantisierungsfehler mit dem Ausgangsvektor \underline{x} überein.

Die Rücktransformation muss nicht notwendig invers zur Sendertransformation sein. Es läßt sich aber zeigen, daß die Rücktransformation nach Gl. 2.1.3 zumindest für Gaußquellen unabhängig von \underline{A} und von der Bitzuordnung optimal im Sinne des kleinsten MQF ist, wenn für $Q_1 \dots Q_N$ Max-Quantisierer (Quantisierer für kleinsten MQF) eingesetzt werden /16/.

Das Schema nach Bild 2.1 wird als basisbeschränktes Transformationssystem bezeichnet /17/, weil zwei Voraussetzungen gelten sollen:

1. Die Transformationsbasis \underline{A} soll zeitinvariant sein (unabhängig von der Eingangsstatistik).
2. Die Quantisierer Q_i arbeiten unabhängig voneinander.

Beide Bedingungen dienen im wesentlichen zur Vereinfachung bei der Optimierung und Realisierung des Codiersystems. Die erste Voraussetzung muss erfüllt sein, wenn von vornherein eine feste Transformation \underline{A} vorgegeben ist, die sich z.B. durch einen geringen Rechenaufwand auszeichnet. Die zweite Bedingung vereinfacht die Optimierung der

Quantisierer, und vor allem die Quantisierung selbst, ganz erheblich. In den nächsten Abschnitten wird deutlich werden, dass diese Einschränkung bei einer Reihe von Transformationen nur zu einer geringen Gewinnabnahme gegenüber einem ideal optimierten System führt.

2.1.1 Statistische Kenngrößen im Zeit- und Bildbereich

Das Eingangssignal x_n sei stationär mit dem Mittelwert

$$E[x_n] = 0 \quad (2.1.4)$$

und der Varianz

$$\sigma_x^2 = E[x_n^2] \quad (2.1.5)$$

Die Korrelationen des Eingangssignals sind durch die Autokovarianzmatrix

$$\underline{R}_{xx} = E[\underline{x} \cdot \underline{x}^T] \quad (2.1.6)$$

gekennzeichnet. Die Spektralkoeffizienten y_i im Bildbereich sind durch

$$y_i = \sum_{j=1}^N a_{ij} x_j \quad (2.1.7)$$

gegeben; a_{ij} sind die Elemente des i -ten Zeilenvektors von A. Für den Mittelwert $E[y_i]$ gilt mit Gl. 2.1.4

$$E[y_i] = \sum_{j=1}^N a_{ij} E[x_j] = 0 \quad (2.1.8)$$

Für die Autokovarianzmatrix \underline{R}_{yy} im Bildbereich gilt mit Gl. 2.1.1 und Gl. 2.1.6

$$\begin{aligned}\underline{R}_{yy} &= E[\underline{y} \cdot \underline{y}^T] \\ &= E[\underline{A} \cdot \underline{x} \cdot \underline{x}^T \cdot \underline{A}^T] \\ \underline{R}_{yy} &= \underline{A} \cdot \underline{R}_{xx} \cdot \underline{A}^T\end{aligned}\quad (2.1.9)$$

Die Varianzen der Spektralkoeffizienten

$$\sigma_i^2 = E[y_i^2] \quad (2.1.10)$$

sind daher als Elemente der Hauptdiagonalen von \underline{R}_{yy} durch Gleichung 2.1.9 festgelegt.

Als mittlere Varianz im Bildbereich wird

$$\sigma_y^2 = \frac{1}{N} \sum_{i=1}^N \sigma_i^2 \quad (2.1.11)$$

definiert. Für σ_y^2 erhält man mit Gleichung 2.1.2 (Orthogonalität von \underline{A})

$$\begin{aligned}\sigma_y^2 &= \frac{1}{N} E[\underline{y}^T \cdot \underline{y}] = \frac{1}{N} E[\underline{x}^T \cdot \underline{A}^T \cdot \underline{A} \cdot \underline{x}] \\ \sigma_y^2 &= \frac{1}{N} E[\underline{x}^T \cdot \underline{x}] = \sigma_x^2\end{aligned}\quad (2.1.12)$$

Die mittlere Varianz im Zeit- und im Bildbereich ist also gleich gross und soll für die weiteren Betrachtungen mit

$$\sigma^2 = \sigma_x^2 = \sigma_y^2 \quad (2.1.13)$$

bezeichnet werden.

2.1.2 Bestimmung der optimalen Bitzuordnung

Die Spektralkoeffizienten y werden unabhängig voneinander quantisiert; für jeden Wert y_i entsteht dabei der Quantisierungsfehler

$$\Delta y_i = y_i - \hat{y}_i \quad (2.1.14)$$

mit der Varianz

$$D_i = E[\Delta y_i^2] \quad (2.1.15)$$

Die mittlere Fehlervarianz im Bildbereich wird mit

$$D = \frac{1}{N} \sum_{i=1}^N D_i \quad (2.1.16)$$

bezeichnet. Es sei $\underline{\Delta y}$ der Vektor der Quantisierungsfehler; dann folgt für D

$$\begin{aligned} D &= \frac{1}{N} E[\underline{\Delta y}^T \cdot \underline{\Delta y}] \\ &= \frac{1}{N} E[(\underline{y} - \hat{\underline{y}})^T \cdot (\underline{y} - \hat{\underline{y}})] \\ &= \frac{1}{N} E[(\underline{x} - \hat{\underline{x}})^T \cdot \underline{A}^T \cdot \underline{A} \cdot (\underline{x} - \hat{\underline{x}})] \\ D &= \frac{1}{N} E[(\underline{x} - \hat{\underline{x}})^T \cdot (\underline{x} - \hat{\underline{x}})] \quad (2.1.17) \end{aligned}$$

Der Vektor $\underline{\Delta x} = \underline{x} - \hat{\underline{x}}$ ist der Quantisierungsfehler im

Zeitbereich und damit der Gesamtfehler des Codiersystems. Die Gleichung 2.1.17 zeigt, dass die mittleren Fehlervarianzen im Zeitbereich und im Bildbereich gleich gross sind. Diese Aussage vereinfacht die nun folgende Optimierungsaufgabe wesentlich. Ziel ist es, den Quantisierungsfehler im Zeitbereich zu minimisieren. Da sowohl die mittleren Signalvarianzen als auch die mittleren Fehlervarianzen im Zeit- und im Bildbereich gleich gross sind, kann diese Aufgabe allein im Bildbereich gelöst werden.

Gegeben sei eine mittlere Übertragungsrate \bar{R} in bit pro Abtastwert. Zur Codierung eines Datenblocks der Länge N stehen damit $N \cdot \bar{R}$ bit zur Verfügung. Jeder Spektralkoeffizient y_i wird mit der Bitzahl R_i quantisiert, wobei

$$\sum_{i=1}^N R_i = N \cdot \bar{R} \quad (2.1.18)$$

Da ein basisbeschränktes Transformationssystem vorausgesetzt wurde, reduziert sich die Optimierungsaufgabe auf das Problem, jedem Spektralkoeffizienten y_i eine solche Bitzahl R_i zuzuordnen, bei der der mittlere Quantisierungsfehler D minimal wird.

Die Optimierungsaufgabe lautet also:

$$D = \frac{1}{N} \sum_{i=1}^N D_i \stackrel{!}{=} \min. \quad (2.1.19)$$

mit der Nebenbedingung

$$\bar{R} = \frac{1}{N} \sum_{i=1}^N R_i = \text{const.} \quad (2.1.20)$$

Der Zusammenhang zwischen der Bitzahl R_i und der Verzerrung D_i bei der Quantisierung des Spektralkoeffizienten y_i (Varianz σ_i^2) wird durch folgende Näherung beschrieben:

$$R_i = \delta + \frac{1}{2} \lg \frac{\sigma_i^2}{D_i} \quad (2.1.21)$$

Der Wert δ wird durch den Quantisierertyp und die Amplitudendichteverteilung (ADV) von y_i bestimmt und ist nur für grössere Werte von R_i näherungsweise konstant /18/. Es wird weiterhin vorausgesetzt, dass die ADV für alle Koeffizienten y_i gleich ist und der gleiche Quantisierertyp verwendet wird, so dass die Gl. 2.1.21 für alle Koeffizienten y_i gültig ist.

Für die Verzerrung D_i ergibt sich aus Gl. 2.1.21

$$D_i = \sigma_i^2 \cdot 2^{2\delta} \cdot 2^{-2R_i} \quad (2.1.22)$$

Löst man die Aufgabe nach Gl. 2.1.19 und 2.1.20 mit der Methode des Lagrangeschen Multiplikators /16/, so erhält man die Gleichungen

$$\frac{\partial}{\partial R_k} \left[\frac{1}{N} 2^{2\delta} \sum_{i=1}^N \sigma_i^2 2^{-2R_i} + \lambda \left(\frac{1}{N} \sum_{i=1}^N R_i - \bar{R} \right) \right] = 0$$

$$\text{mit } k = 1, \dots, N \quad (2.1.23)$$

Daraus ergibt sich

$$R_k = \delta + \frac{1}{2} \lg \frac{\sigma_k^2 \cdot 2 \ln 2}{\lambda} \quad (2.1.24)$$

Mit Gl. 2.1.20 und Gl. 2.1.24 erhält man für λ

$$\frac{1}{2} \text{ld } \lambda = \delta - \bar{R} + \frac{1}{2} \text{ld} \left[2 \ln 2 \left(\prod_{i=1}^N \sigma_i^2 \right)^{\frac{1}{N}} \right] \quad (2.1.25)$$

Mit Gl. 2.1.24 wird die Bitzahl für den k-ten Spektralkoeffizient bestimmt zu

$$R_k = \bar{R} + \frac{1}{2} \text{ld} \frac{\sigma_k^2}{\left[\prod_{i=1}^N \sigma_i^2 \right]^{\frac{1}{N}}} \quad (2.1.26)$$

Die Bitzuordnung für die einzelnen Spektralkoeffizienten wird also allein aus der Verteilung der Varianzen im Bildbereich ermittelt.

Der Verzerrungsbeitrag eines einzelnen Spektralkoeffizienten ist mit Gleichung 2.1.22 und 2.1.26

$$D_k = 2^{2\delta} 2^{-2\bar{R}} \left[\prod_{i=1}^N \sigma_i^2 \right]^{\frac{1}{N}} \quad (2.1.27)$$

Gl. 2.1.27 zeigt, dass die gestellte Optimierungsaufgabe auf eine Bitzuordnung führt, bei der alle Spektralkoeffizienten einen gleich grossen Verzerrungsbeitrag $D = D_k$ aufweisen. Die angegebenen Gleichungen gelten nur, wenn die Bedingung

$$D \leq \min. \{ \sigma_k^2 \} \quad \forall k \quad (2.1.28)$$

eingehalten werden kann. Bei grösseren Verzerrungen würden sonst negative Werte für R_i berechnet werden. Dieser Sonderfall wird in Abschnitt 2.4 noch ausführlicher dis-

kutiert.

2.1.3 Theoretischer Gewinn gegenüber einer PCM-Codierung

Es soll nun untersucht werden, welche Verbesserung die Blockquantisierung gegenüber einer PCM-Codierung bewirkt. Für beide Codiersysteme werde eine gleich grosse mittlere Übertragungsrate \bar{R} angenommen. Gilt für den Zeitbereich das gleiche Quantisierergesetz nach Gl. 2.1.21 wie für den Bildbereich, so ist die mittlere Verzerrung bei einer PCM-Codierung

$$D_{\text{PCM}} = \frac{2\delta}{2 \cdot 2} \cdot \sigma^2, \quad (2.1.29)$$

denn alle Abtastwerte x_n weisen die Varianz σ^2 auf.

Als Gewinn wird das Verhältnis der Verzerrungen bei PCM und bei einer Blockquantisierung definiert, mit Gl. 2.1.27 ergibt sich dann:

$$G_{\text{BQ}} = 10 \lg \frac{D_{\text{PCM}}}{D} = 10 \lg \frac{\sigma^2}{\left[\sum_{i=1}^N \sigma_i^2 \right] \frac{1}{N}} \text{ in dB} \quad (2.1.30)$$

Gleichung 2.1.30 zeigt, dass der Gewinn bei einer Blockquantisierung durch das Verhältnis vom arithmetischen Mittel zum geometrischen Mittel der Varianzen der Spektralkoeffizienten bestimmt wird. Je ungleicher also die Varianzen im Bildbereich verteilt sind, desto grösser ist der Gewinn. Es sei noch darauf hingewiesen, dass die Gewinnangabe gemäss Gl. 2.1.30 nur gilt, wenn in den Gln. 2.1.27 und 2.1.29 der gleiche δ -Koeffizient benutzt wer-

den kann. Das ist dann richtig, wenn sich die ADV durch die lineare Transformation nicht oder nur wenig geändert hat.

2.2. Eigenschaften spezieller Transformationen

2.2.1 Die Loeve-Karhunen-Transformation

Über die Wahl der Transformationsmatrix \underline{A} sind bisher ausser der Forderung nach Orthonormalität noch keine weiteren Aussagen getroffen worden. Soll der Gewinn bei der Blockquantisierung maximal werden, so muss die Matrix \underline{A} in Abhängigkeit von der Statistik des Eingangssignals gewählt werden. Es lässt sich zeigen, dass der Gewinn maximal wird, wenn die Transformationsmatrix aus den Eigenvektoren der Autokovarianzmatrix \underline{R}_{xx} gebildet wird /16/. Die Matrix der Eigenvektoren von \underline{R}_{xx} werde mit \underline{U} bezeichnet; die Eigenvektoren sind dabei die Spaltenvektoren von \underline{U} . Als Transformationsmatrix wird

$$\underline{A} = \underline{U}^T \quad (2.2.1)$$

gewählt, so dass die jedem Spektralkoeffizienten zugeordneten Basisvektoren der Transformation \underline{A} (Zeilenvektoren, vgl. Gl. 2.1.7) mit den Eigenvektoren von \underline{U} identisch sind. Für die Autokovarianzmatrix im Bildbereich ergibt sich mit Gl. 2.1.9

$$\underline{R}_{yy} = \underline{U}^T \cdot \underline{R}_{xx} \cdot \underline{U} \quad (2.2.2)$$

Da andererseits \underline{U} gerade die Matrix der Eigenvektoren von \underline{R}_{xx} ist, ist \underline{R}_{yy} eine Diagonalmatrix mit

$$\underline{R}_{yy} = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \dots & \\ & 0 & & \lambda_i & \dots \\ & & & & \dots & \lambda_N \end{bmatrix} \quad (2.2.3)$$

Die Werte λ_i sind die Eigenwerte der Autokovarianzmatrix \underline{R}_{xx} . Das bedeutet, dass die Varianzen der Spektralkoeffizienten mit den Eigenwerten von \underline{R}_{xx} identisch sind. Das geometrische Mittel der Varianzen ist bei dieser Transformation kleiner als bei jeder anderen Transformation; daher liefert diese Transformation den grössten Gewinn (s. Gl. 2.1.30). \underline{R}_{yy} ist eine Diagonalmatrix; die Spektralkoeffizienten sind nicht miteinander korreliert. Daher führt die Einschränkung eines basisbeschränkten Transformationssystems, bei dem alle Spektralkoeffizienten unabhängig voneinander quantisiert werden, in diesem Fall nicht zu einer suboptimalen Lösung.

Die angegebene Transformation $\underline{A} = \underline{U}^T$ ist unter dem Namen Loeve-Karhunen-Transformation (LKT) bekannt. Obwohl die LKT den maximalen Gewinn ergibt, wird sie oft nur zu Vergleichszwecken verwendet, da die notwendige Zahl von Rechenoperationen zur Durchführung der Transformation recht hoch ist (etwa N^2 Multiplikationen und Additionen).

2.2.2 Andere Transformationen

Ausser der LKT sind eine Reihe von anderen Transformationen bekannt, die sich dadurch auszeichnen, dass zu ihrer Durchführung erheblich weniger Rechenoperationen als bei der LKT notwendig sind. Diese "schnellen" Transformationen

sind dann nicht mehr an die Statistik des Eingangssignals angepasst, so dass der erreichbare Gewinn in der Regel kleiner als bei der LKT sein wird. Im folgenden werden vier Transformationen beschrieben, die für die Simulationen an Sprachsignalen untersucht werden sollen:

1) Die Walsh-Hadamard-Transformation (WHT)

Für die WHT sind die Elemente von \underline{A} alle dem Betrage nach gleich gross und unterscheiden sich nur durch das Vorzeichen, so dass nur Additionen ausgeführt werden müssen. Durch einen schnellen Algorithmus ist die WHT mit $N \text{ ld } N$ Additionen durchführbar /12/. Für die Hadamard-Transformation ist die Transformationsmatrix \underline{A}_N (Blocklänge N) durch die Rekursionsbeziehungen /12/

$$\underline{A}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad (2.2.4)$$

und

$$\underline{A}_{2N} = \frac{1}{\sqrt{2}} \begin{pmatrix} \underline{A}_N & \underline{A}_N \\ \underline{A}_N & -\underline{A}_N \end{pmatrix} \quad (2.2.5)$$

gegeben. Die Walsh-Transformation unterscheidet sich von der Hadamard-Transformation nur dadurch, dass die Zeilenvektoren von \underline{A} nun nach wachsender Zahl von Vorzeichenwechseln im Zeilenvektor umsortiert wurden. Der Gewinn verändert sich aber gem. Gl. 2.1.30 dadurch nicht, so dass beide Transformationen durch die Bezeichnung Walsh-Hadamard-Transformation gekennzeichnet werden sollen.

2.) Die Diskrete Slant-Transformation (DST)

Die DST wurde für den Einsatz in Bildcodierungssystemen entwickelt; sie arbeitet dort fast so gut wie die LKT /13/. Die Zeilenvektoren dieser Transformation weisen einen etwa sägezahnförmigen Verlauf auf und sind damit dem Betrage auch nicht einfach konstant wie bei der WHT. Die Transformation kann mit $2N - 4$ Multiplikationen und $N \text{ ld } N + N/2 - 2$ Additionen durchgeführt werden.

3.) Die Diskrete Fourier-Transformation (DFT)

Die DFT einer Zahlenfolge x_n der Länge N ist durch

$$v_k = \sum_{n=0}^{N-1} x_n \cdot e^{-j \frac{2\pi}{N} n \cdot k} \quad k = 0, \dots, N-1 \quad (2.2.6)$$

gegeben. Die Größen v_k sind die Spektralwerte. Für eine reelle Zahlenfolge x_n und N gerade gilt:

$$\text{Im} \{v_0\} = 0$$

$$\text{Im} \left\{v_{\frac{N}{2}}\right\} = 0$$

$$\text{Re} \{v_k\} = \text{Re} \{v_{N-k}\} \quad ; \quad k = 1, \dots, \frac{N}{2} - 1$$

$$\text{Im} \{v_k\} = -\text{Im} \{v_{N-k}\} \quad ; \quad k = 1, \dots, \frac{N}{2} - 1$$

Es brauchen daher nur N Anteile aus den konjugiert komplexen Spektralwerten v_k übertragen zu werden. Zur Bildung des reellen Vektors \underline{y} wurde folgende Zuordnung gewählt:

$$\begin{aligned}
 y_1 &= \frac{1}{\sqrt{N}} \operatorname{Re} \{v_0\} = \frac{1}{\sqrt{N}} v_0 \\
 y_2 &= \sqrt{\frac{2}{N}} \operatorname{Im} \{v_1\} \\
 y_3 &= \sqrt{\frac{2}{N}} \operatorname{Re} \{v_1\} \\
 y_4 &= \sqrt{\frac{2}{N}} \operatorname{Im} \{v_2\} \\
 y_5 &= \sqrt{\frac{2}{N}} \operatorname{Re} \{v_2\} \\
 &\vdots \\
 y_{N-2} &= \sqrt{\frac{2}{N}} \operatorname{Im} \left\{ v_{\frac{N}{2}-1} \right\} \\
 y_{N-1} &= \sqrt{\frac{2}{N}} \operatorname{Re} \left\{ v_{\frac{N}{2}-1} \right\} \\
 y_N &= \frac{1}{\sqrt{N}} \operatorname{Re} \left\{ v_{\frac{N}{2}} \right\} = \frac{1}{\sqrt{N}} v_{\frac{N}{2}}
 \end{aligned} \tag{2.2.7}$$

Wird der Vektor \underline{x} mittels der DFT nach Gl. 2.2.7 in den Vektor \underline{y} abgebildet, so entspricht dies einer orthonormalen Transformation: In /19/ wird gezeigt, dass die dem Gleichungssystem 2.2.7 zugeordnete reelle Fouriermatrix orthogonale Basisvektoren aufweist. Die Koeffizienten $\frac{1}{\sqrt{N}}$ und $\sqrt{\frac{2}{N}}$ dienen dazu, um auch die Orthonormalität zu gewährleisten. Diese war eine wesentliche Voraussetzung für die Betrachtungen aus Abschnitt 2.1.

Die gewählte Reihenfolge in Gl. 2.2.7 bezüglich der Real- und Imaginärteile von v_k wurde nach wachsender Zahl von Vorzeichenwechseln der dieser Transformation zugeordneten reellen Basisvektoren vorgenommen. Für die DFT ist bekannt, dass sie asymptotisch optimal ist. Seien G_{LKT} und G_{DFT} die Gewinne nach Gl. 2.1.30 für eine LKT bzw. DFT, so gilt /20/

$$G_{\text{LKT}} - G_{\text{DFT}} \sim \frac{1}{\sqrt{N}} \tag{2.2.8}$$

Mit wachsender Blocklänge N erreicht die DFT den LKT-Gewinn, die Korrelationen zwischen den Spektralkoeffizienten nehmen dabei immer stärker ab. Zur Durchführung der DFT sind eine Reihe von schnellen Algorithmen bekannt /21/. Die Zahl der komplexen Multiplikationen bewegt sich je nach Algorithmus zwischen $N \cdot \lg N$ und $\frac{N}{2} \cdot \lg \frac{N}{2}$.

4.) Die Diskrete Cosinus-Transformation (DCT)

Die DCT beruht im Gegensatz zur DFT nur auf einem System von Cosinus-Funktionen. Abweichend zu /14/ soll die DCT hier so definiert werden, dass die zugeordnete Transformationsmatrix A orthonormal ist:

$$y_1 = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x_n$$
$$y_k = \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} x_n \cos \frac{(2n+1) \cdot (k-1) \pi}{2N} ; \quad k = 2, \dots, N$$

(2.2.9)

Die DCT arbeitet auch bei kleinen Blocklängen nahezu genauso gut wie die LKT, wenn die Autokovarianzmatrix R_{xx} näherungsweise einem autoregressiven Prozess 1. Ordnung mit einem sehr hohen positiven Korrelationskoeffizienten zugeordnet werden kann /14/. Die DCT kann über die DFT berechnet werden, wenn der Eingangsvektor \underline{x} durch Anfügen von Nullen auf die Länge $N^* = 2N$ vergrößert wird. Die DFT muss dann über eine Blocklänge N^* durchgeführt werden.

Mit wachsender Blocklänge N erreicht die DFT den LKT-Gewinn, die Korrelationen zwischen den Spektralkoeffizienten nehmen dabei immer stärker ab. Zur Durchführung der DFT sind eine Reihe von schnellen Algorithmen bekannt /21/. Die Zahl der komplexen Multiplikationen bewegt sich je nach Algorithmus zwischen $N \cdot \lg N$ und $\frac{N}{2} \cdot \lg \frac{N}{2}$.

4.) Die Diskrete Cosinus-Transformation (DCT)

Die DCT beruht im Gegensatz zur DFT nur auf einem System von Cosinus-Funktionen. Abweichend zu /14/ soll die DCT hier so definiert werden, dass die zugeordnete Transformationsmatrix A orthonormal ist:

$$y_1 = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x_n$$

$$y_k = \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} x_n \cos \frac{(2n+1) \cdot (k-1) \pi}{2N} ; \quad k = 2, \dots, N$$

(2.2.9)

Die DCT arbeitet auch bei kleinen Blocklängen nahezu genauso gut wie die LKT, wenn die Autokovarianzmatrix R_{xx} näherungsweise einem autoregressiven Prozess 1. Ordnung mit einem sehr hohen positiven Korrelationskoeffizienten zugeordnet werden kann /14/. Die DCT kann über die DFT berechnet werden, wenn der Eingangsvektor \underline{x} durch Anfügen von Nullen auf die Länge $N^* = 2N$ vergrößert wird. Die DFT muss dann über eine Blocklänge N^* durchgeführt werden.

2.2.3 Darstellung der Basisvektoren

Die Basisvektoren der Transformation sind die Zeilenvektoren von \underline{A} . In Bild 2.2 sind die Basisvektoren für die beschriebenen Transformationen bei einer Blocklänge $N=8$ dargestellt. Für die LKT sind die Basisvektoren von der Autokovarianzmatrix \underline{R}_{xx} abhängig. Der in Bild 2.2 zugrundegelegte Prozess entspricht dem Langzeit-Leistungsdichtespektrum von Sprache (vgl. dazu Abschnitt 2.3.2)

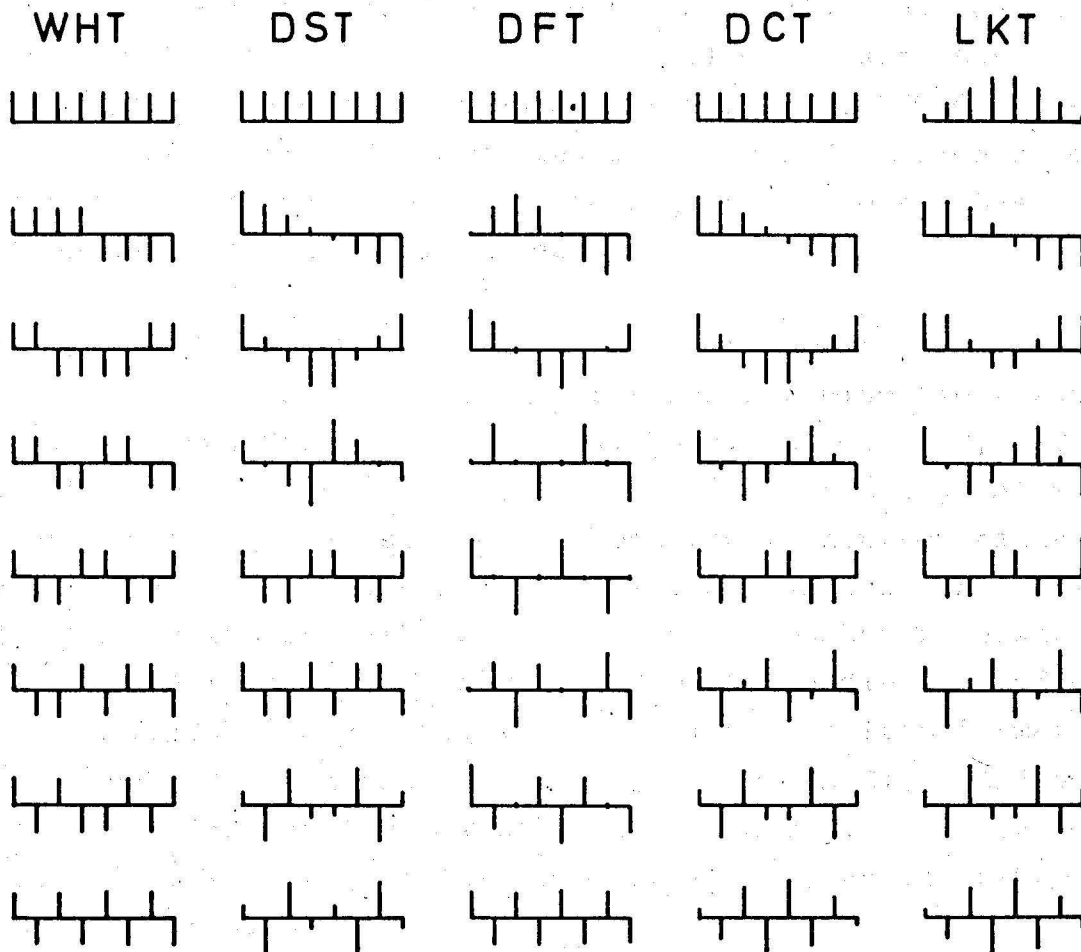


Bild 2.2 : Darstellung der Basisvektoren für verschiedene orthonormale Transformationen der Blocklänge $N=8$.

Es ist ersichtlich, dass für alle Transformationen ein Satz von Basisvektoren zunehmender "Frequenz" existiert. Die Frequenz wird gekennzeichnet durch die Zahl der Vorzeichenwechsel in einem Basisvektor; diese Zahl wird als Sequenz bezeichnet.

Die WHT, DST und DCT besitzen die Sequenzeigenschaft; das bedeutet, jeder Basisvektor weist eine andere Sequenz auf. In Bild 2.2 trifft dies zwar auch für die LKT zu, aber bei grösseren Blocklängen treten in der Regel mehrerer Basisvektoren gleicher Sequenz auf. Für die DFT ist die Sequenzeigenschaft in strengem Sinne nicht erfüllt, da bei einigen Basisvektoren verschiedene Elemente Null sind und dadurch sich die Sequenz nicht eindeutig festlegen lässt. Aus Bild 2.2 ist auch zu entnehmen, warum für die DFT die Zuordnung nach Gl. 2.2.7 gewählt wurde. Es ergibt sich so eine weitgehende Entsprechung zu den sequenzsortierten Basisvektoren anderer Transformationen.

Für die DST, LKT und die Hadamard-Transformation sind die Basisvektoren in Bild 2.2 nach ihrer Sequenz umsortiert worden, um einen Vergleich zu ermöglichen; für die DFT, DCT und Walsh-Transformation ist dies nicht mehr nötig. Alle Transformationen ausser der LKT besitzen einen konstanten Basisvektor. Besonders auffällig ist die grosse Ähnlichkeit der Basisvektoren bei DCT und LKT. Es ist daher zu erwarten, dass die DCT nahezu den gleichen Gewinn wie die LKT aufweisen wird. Für die Blocklänge $N = 2$ sind alle Transformationen identisch. Die dazugehörige Matrix A_2 ist in Gl. 2.2.4 angegeben. Die geometrische Interpretation der Gleichung zeigt, dass der Vektor \underline{x} (ein Paar benachbarter Abtastwerte) um 45° gedreht wird.

2.3 Vergleich der verschiedenen Transformationen

2.3.1 Zusammenhang mit dem Gewinn bei einer Prädiktionscodierung (DPCM)

Zwischen dem Gewinn bei der LKT und dem Gewinn bei einer Prädiktionscodierung kann ein Zusammenhang angegeben werden. Für die Prädiktionscodierung ist der Gewinn durch das Verhältnis von Signalvarianz σ^2 zur Prädiktionsfehlervarianz ϵ_k^2 gegeben:

$$G_{p(k)} = 10 \lg \frac{\sigma^2}{\epsilon_k^2} \quad \text{in dB. (2.3.1)}$$

Die Grösse ϵ_k^2 bezieht sich dabei auf einen optimalen Prädiktor mit k Filterkoeffizienten. Für die LKT ist der Gewinn nach Gl. 2.1.30

$$G_{LKT} = 10 \lg \frac{\sigma^2}{\left[\prod_{i=1}^N \lambda_i \right]^{\frac{1}{N}}} \quad \text{in dB. (2.3.2)}$$

Für die Eigenwerte der Autokovarianzmatrix \underline{R}_{xx} gilt

$$\prod_{i=1}^N \lambda_i = \det \underline{R}_{xx} \quad (2.3.3)$$

Die Determinante von \underline{R}_{xx} kann andererseits gerade durch das Produkt der Prädiktionsfehlervarianzen beschrieben werden/22/:

$$\det \underline{R}_{xx} = \prod_{i=0}^{N-1} \epsilon_i^2 \quad (2.3.4)$$

Damit folgt

$$G_{\text{LKT}} = 10 \lg \frac{\sigma^2}{\left[\prod_{i=0}^{N-1} \varepsilon_i^2 \right]^{\frac{1}{N}}} = \frac{1}{N} \sum_{i=0}^{N-1} G_p(i) \quad \text{in dB. (2.3.5)}$$

Vergleicht man eine LKT der Blocklänge N mit einer Prädiktionscodierung bei einem N -stufigen Prädiktor, so gilt

$$G_{\text{LKT}} \leq G_p(N) \quad (2.3.6)$$

Das Gleichheitszeichen gilt nur für $N \rightarrow \infty$. Auch wenn ein autoregressiver Prozess endlicher Ordnung M vorliegt, muss bei der LKT für die Blocklänge die Bedingung $N \gg M$ erfüllt sein, um den maximalen Gewinn $G_p(\infty)$ zu erreichen, während bei der Prädiktionscodierung ein Prädiktor der Länge M ausreicht. Die angegebenen Beziehungen gelten aber nur, wenn unendlich lange Musterfolgen des zu codierenden Prozesses verarbeitet werden. Liegt nur ein Ausschnitt von Daten der Länge N vor, so stehen dem Prädiktor am Anfang nur wenige Schätzdaten zur Verfügung und der maximale Prädiktionsgewinn wird dadurch nicht erreicht. In diesem Fall ist die LKT der Prädiktionscodierung überlegen /23/.

2.3.2 Theoretische Gewinne für einen autoregressiven Prozess endlicher Ordnung

In diesem Abschnitt sollen die theoretischen Gewinne der fünf angeführten Transformationen mit dem Gewinn bei einer Prädiktionscodierung für ein Beispiel, einen autoregressiven Prozess endlicher Ordnung, verglichen werden. Aus dem Langzeit-Leistungsdichtespektrum von Sprache bzw. den dazugehö-

rigen ersten M AKF-Werten wurde ein autoregressives Modell der Ordnung M=10 berechnet (Abtastfrequenz 8 kHz). Aus den M Filterkoeffizienten des Modells kann dann die Autokovarianzmatrix \underline{R}_{xx} beliebiger Ordnung bestimmt werden /22/.

Das Matrizenprodukt $\underline{A} \cdot \underline{R}_{xx} \cdot \underline{A}^T$ führt auf die Varianzen der Spektralkoeffizienten, die nach Gl. 2.1.30 den Gewinn G_{BQ} bei einer Blockquantisierung festlegen.

Die Ergebnisse der folgenden Betrachtung sind dann auch für die nichtadaptive Blockquantisierung von Sprachsignalen von Bedeutung, da der Modell-Prozess an das Langzeit-Leistungsdichtespektrum von Sprache angepasst wurde.

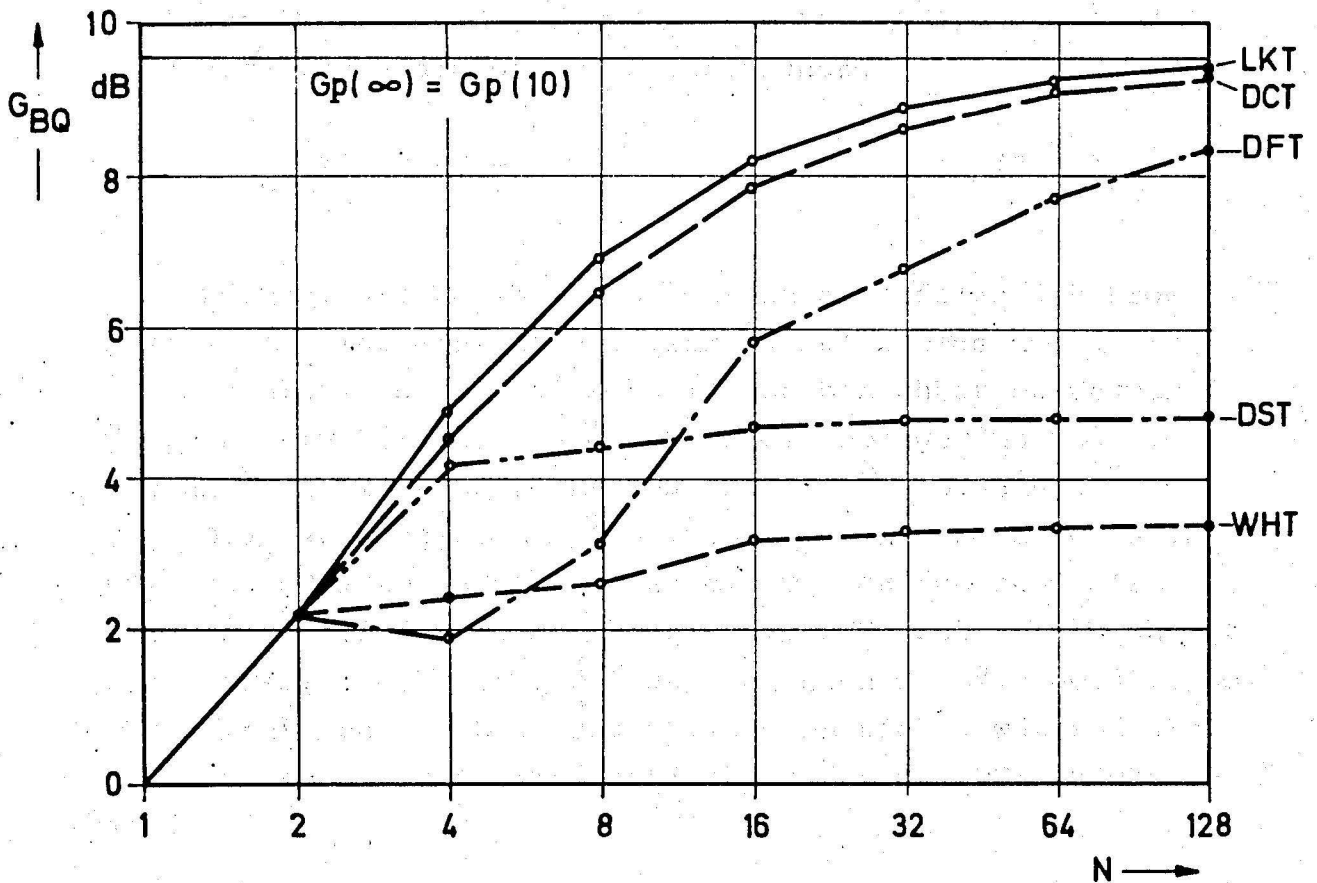


Bild 2.3 : Theoretische Gewinne für einen autoregressiven Prozeß der Ordnung M=10 (Modell aus dem Langzeit-LDS von Sprache).

Bild 2.3 zeigt die theoretischen Gewinne in Abhängigkeit von der Blocklänge N . Zum Vergleich ist der Prädiktionsgewinn $G_p(\infty) = G_p(10)$ eingezeichnet. Die LKT erreicht erst bei grossen Blocklängen den maximalen Gewinn $G_p(\infty)$. Interessant ist, dass die DCT höchstens um etwa 0.5 dB schlechter als die LKT arbeitet und im Gewinn der asymptotisch optimalen DFT deutlich überlegen ist.

Die DST und die WHT erreichen nur einen Gewinn von etwa 5 bzw. 3 dB, der sich auch bei grösseren Blocklängen nicht mehr erhöht. Für $N=2$ haben alle Transformationen den gleichen Arbeitspunkt in Bild 2.3, da jeweils die gleiche Transformationsmatrix A_2 zugrunde liegt. In Abschnitt 2.2.2 ist schon verdeutlicht worden, dass die DCT sehr gut für autoregressive Prozesse (AR-Prozesse) 1. Ordnung mit starker positiver Korrelation geeignet ist. Der AR-Prozess 1. Ordnung wird durch die Differenzgleichung

$$x_n = \rho x_{n-1} + \varepsilon_n \quad (2.3.7)$$

beschrieben. Die Folge x_n ist die Ausgangsfolge des Prozesses, ε_n ist eine unkorrelierte Zahlenfolge und ρ ist der Korrelationskoeffizient zwischen benachbarten Werten der Folge x_n . In Bild 2.4 sind der Prädiktionsgewinn und die Gewinne bei einer LKT, DCT und DFT für $\rho = 0.95$ und

$\rho = -0.95$ angegeben. Ist $\rho = +0.95$, so unterscheiden sich LKT- und DCT-Gewinn um weniger als 0.1 dB, während die DFT bis zu 1.5 dB schlechter als die LKT arbeitet. Wechselt der Korrelationskoeffizient das Vorzeichen ($\rho = -0.95$), so verändern sich der LKT- und DFT-Gewinn nicht, während die DCT dagegen nun erheblich schlechter als die DFT abschneidet.

Aus Bild 2.4 lässt sich daher zusammenfassend ablesen, dass die DFT gleichermassen gut für stark positiv oder negativ korrelierte Prozesse arbeitet. Die DCT ist besonders für

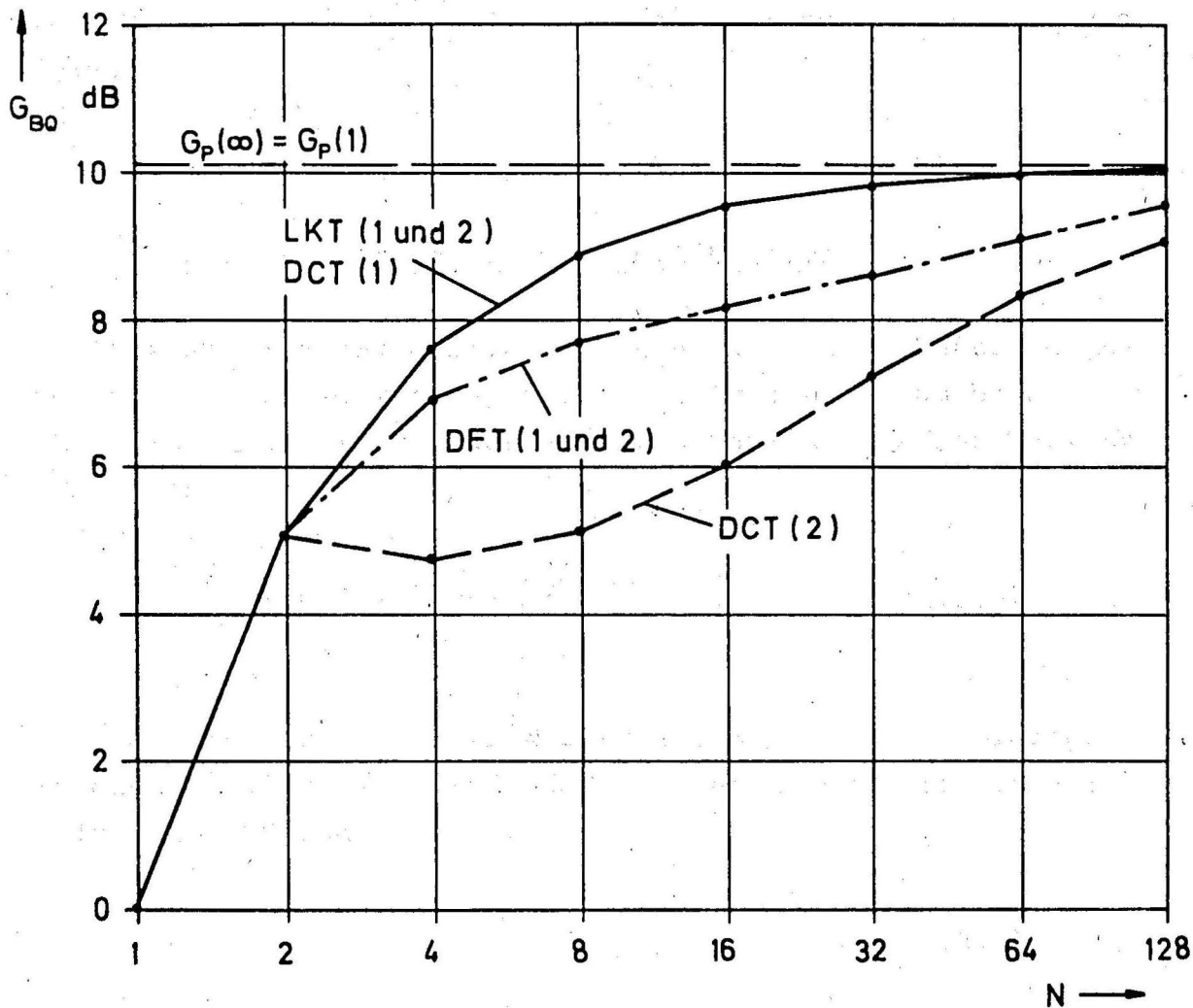


Bild 2.4 : Theoretischer Gewinn G_{BQ} als Funktion der Blocklänge N für einen AR-Prozeß 1. Ordnung.

- 1 : $\rho = +0.95$
- 2 : $\rho = -0.95$

positiv korrelierte Prozesse geeignet und kann daher bezüglich des Gewinns die LKT ersetzen. Für grosse Blocklängen N aber arbeiten alle 3 Transformationen etwa gleich gut, unabhängig von den Korrelationen im zu codierenden Prozess.

2.3.3 Vergleich der Spektren im Bildbereich

Besitzt das autoregressive Prozess-Modell die Übertragungsfunktion $H(z)$, so ist das Leistungsdichtespektrum beim Einspeisen mit einer unkorrelierten Zahlenfolge der Varianz σ_{ϵ}^2 durch

$$S_{xx}(e^{j\Omega}) = |H(e^{j\Omega})|^2 \cdot \sigma_{\epsilon}^2 \quad (2.3.8)$$

gegeben. Interessieren nur N diskrete Werte der normierten Frequenz Ω , so durchläuft Ω den Bereich von Null bis zur halben Abtastfrequenz, wenn

$$\Omega_i = \pi \frac{i-1}{N-1}, \quad i = 1, \dots, N \quad (2.3.9)$$

Die Werte $S_{xx}(e^{j\Omega_i})$ bilden das Sample-LDS des Prozesses und können, wenn sie auf ihren Mittelwert normiert werden, sehr gut mit den Basisspektren der orthonormalen Transformationen verglichen werden. Diese Basisspektren sind durch die Varianzen der Spektralkoeffizienten festgelegt.

In Bild 2.5 a und 2.5b sind die verschiedenen Spektren für das AR-Modell 10. Ordnung, das dem Langzeit-LDS von Sprache angepasst ist, dargestellt. Die Basisvektoren der Transformationen sind dabei wieder nach ihrer Sequenz sortiert worden.

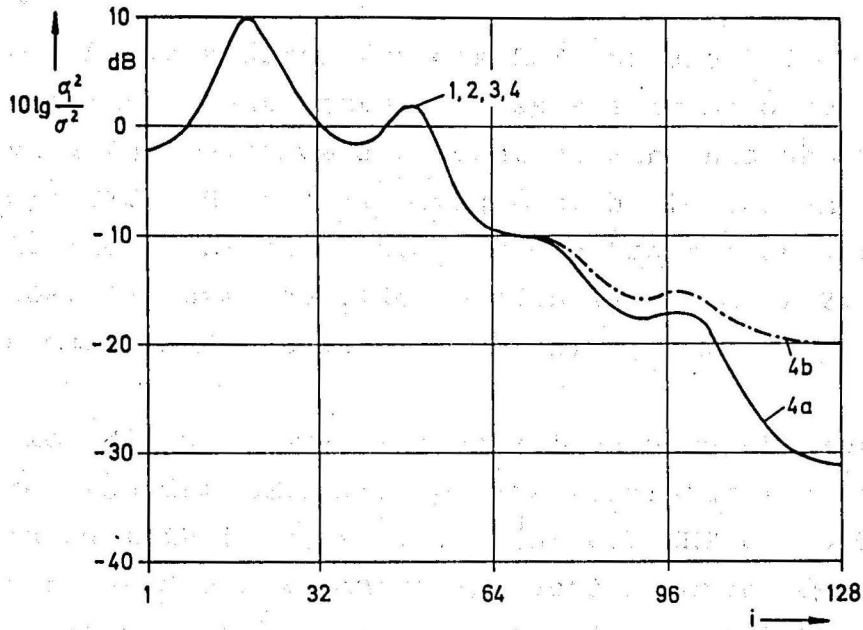


Bild 2.5a

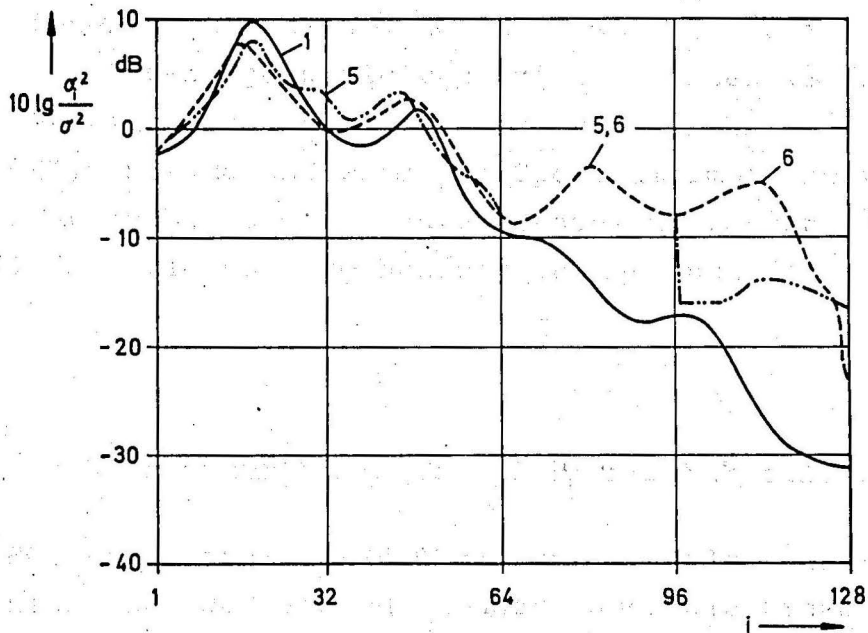


Bild 2.5b

Bild 2.5 : Varianzen der Spektrallinien in Abhängigkeit von der Ordnungszahl i für das AR-Modell des Sprach-LDS (Blocklänge $N=128$).

- 1 : Spektralwerte des LDS nach Gl. 2.3.8
- 2 : Spektralwerte der LKT
- 3 : Spektralwerte der DCT
- 4a: Spektralwerte der DFT (Imaginärteile)
- 4b: Spektralwerte der DFT (Realteile)
- 5 : Spektralwerte der DST
- 6 : Spektralwerte der WHT

Die Spektralwerte für das LDS und die LKT, DCT und DFT stimmen so gut überein, dass sie im Bild nur als eine Kurve dargestellt werden konnten. Eine Ausnahme bildet aber die DFT: Für $i > 64$ bewegen sich die Varianzen der Spektralkoeffizienten, die aus den Realteilen der Fourierkoeffizienten gebildet wurden (vgl. Gl. 2.2.7) auf einem separaten Kurvenast (Kurve 4b).

Die WHT- und DST-Spektren zeigen noch stärkere Abweichungen von dem LDS. Die Spektralwerte für $i = 65, \dots, 96$ (allgemein $i = \frac{N}{2} + 1, \dots, \frac{3}{4}N$) bei WHT und DST stimmen überein, da die dazugehörigen Basisvektoren identisch sind (s. Bild 2.2). Aus den Bildern 2.5a und 2.5b wird deutlich, wodurch die Gewinnunterschiede der einzelnen Transformationen zu erklären sind. Nach Gl. 2.1.30 ist der Gewinn um so grösser, je ungleicher die Varianzen im Spektralbereich verteilt sind. Beim Übergang von der LKT und DCT auf die DFT, DST und WHT wird das Spektrum immer schwächer aufgelöst (d.h., die Varianzunterschiede werden geringer) und in der gleichen Reihenfolge nimmt auch der Gewinn ab.

2.3.4 Rechenaufwand zur Durchführung der Transformationen

Für eine Hardware-Realisierung der Blockquantisierung ist nicht nur der Gewinn, sondern auch die Frage des Rechenaufwandes zur Durchführung der Transformation entscheidend. Die "schnellen" Transformationen (WHT, DST, DFT und DCT) besitzen mit dem Vorteil der verringerten Zahl von Rechenoperationen auch die Eigenschaft, mit einer kleineren Digitalwortlänge als bei der LKT auszukommen. Denn durch die kleinere Zahl von Rechenoperationen verringert sich gleichzeitig der Einfluss des Rundungsrauschens /24/. In der Tabelle 2.1 ist die Zahl der reellen Rechenoperationen pro Abtastwert bei der Blocklänge N angegeben.

Transformation	WHT	DST	DFT	DCT	LKT
Zahl der Additionen	1d N	1d N	$\frac{1}{2}$ 1d N	1d N	N
Zahl der Multiplikationen	-	2	1d N	2 1d N	N

Tabelle 2.1: Zahl der reellen Rechenoperationen bei der Blocklänge N pro Abtastwert.

Dabei ist nur die Grössenordnung berücksichtigt worden. Für die DFT und DCT wurde angenommen, dass eine DFT der Länge N mit $\frac{N}{2}$ 1d $\frac{N}{2}$ komplexen Multiplikationen und Additionen durchgeführt werden kann und ausserdem 2 reelle Datenvektoren gleichzeitig transformiert werden. Vergleicht man den Rechenaufwand bei den verschiedenen Transformationen mit dem Gewinn nach Bild 2.3, so zeigt sich, dass ein grösserer Gewinn auch mit einer erhöhten Zahl von Rechenoperationen erkauft werden muss. Am günstigsten schneidet die DCT ab, die bezüglich des Gewinns für alle Blocklängen nahezu der LKT gleichwertig ist, aber mit sehr viel weniger Rechenaufwand auskommt.

2.4 Realisierung der Bitzuordnung bei zusätzlichen Randbedingungen

In Abschnitt 2.1.2 ist die optimale Bitzuordnung bei Vorgabe der Varianzen σ_i^2 der Spektralkoeffizienten und der mittleren Übertragungsrate \bar{R} hergeleitet worden. Für eine praktische Realisierung treten aber zusätzliche Randbedingungen auf:

- a) Die Bitzahl R_i für jeden Koeffizienten darf nicht negativ sein: $R_i \geq 0$.
- b) Die Bitzahl R_i darf einen Maximalwert nicht überschreiten, wenn ein Quantisierer mit maximal R_{\max} bit zur Verfügung steht: $R_i \leq R_{\max}$.
- c) Die Bitzahl R_i muss ganzzahlig sein.

Eine optimale Bitzuordnung, die die angegebenen Randbedingungen erfüllt, wird auf folgendem Wege erhalten:

- 1) Bestimmung der R_i mit $i=1, \dots, N$ nach Gl. 2.1.26.
- 2) Bildung von zwei Index-Mengen:
 - I : Indexmenge für alle i mit $R_i \geq 0$.
 - I^* : Indexmenge für alle i mit $R_i < 0$.

Ist I^* die leere Menge, d. h. , existiert kein negativer Wert für R_i , so wird auf Schritt 3 gesprungen. Anderenfalls enthält I^* L Elemente ($L > 0$); es ist dann folgende Korrektur durchzuführen:

$$\Delta R = \frac{1}{N-L} \sum_{i \in I^*} |R_i|$$

$$R_i^{(\text{neu})} = R_i^{(\text{alt})} - \Delta R \quad \text{für } i \in I$$

$$R_i^{(\text{neu})} = 0 \quad \text{für } i \in I^*$$

Begründung: Die negativen Werte R_i müssen zu Null gesetzt werden. Damit sind nur noch die Variablen R_i mit $i \in I$ frei wählbar, über diese Variablen kann die Optimierung nach Gl. 2.1.19 und 2.1.20 erneut durchgeführt werden /25/. Es ist leicht nachzuprüfen, dass die ursprünglichen Werte R_i mit $i \in I$ dann alle um den gleichen Wert ΔR vermindert werden müssen. Der Wert ΔR kann sehr einfach bestimmt werden (s.o.); es ist keinesfalls eine Neuberechnung der R_i ent-

sprechend Gl. 2.1.26 notwendig. Es erfolgt nun ein Rücksprung auf den Anfang von Schritt 2.

3) Bildung von zwei Indexmengen:

I : Indexmenge für alle i mit $R_i \leq R_{\max}$.

I^* : Indexmenge für alle i mit $R_i > R_{\max}$.

Es sei L die Zahl der Elemente in I^* . Ist L gleich Null, so wird auf Schritt 4 gesprungen, anderenfalls ist folgende Korrektur durchzuführen:

$$\Delta R = \frac{1}{N-L} \sum_{i \in I^*} (R_i - R_{\max})$$

$$R_i^{(\text{neu})} = R_{\max} \quad \text{für } i \in I^*$$

$$R_i^{(\text{neu})} = R_i^{(\text{alt})} + \Delta R \quad \text{für } i \in I$$

Begründung: Die Werte R_i mit $i \in I^*$ müssen gleich R_{\max} gesetzt werden. Eine Reoptimierung über die verbleibenden R_i mit $i \in I$ führt analog zu Schritt 2 auf die angegebenen Gleichungen. Es erfolgt nun ein Rücksprung auf den Anfang von Schritt 3.

4) Die berechneten Werte R_i werden nun nach kleinsten Rundungsfehlern auf- oder abgerundet ($R_i \rightarrow R_i^*$). Dabei wird die Summe der jetzt ganzzahligen R_i^* in der Regel nicht mehr mit dem vorgegebenen Wert $N \cdot \bar{R}$ übereinstimmen. Entsprechend der Zahl der überzähligen oder fehlenden bit wird bei den Koeffizienten R_i^* mit den grössten Auf- bzw. Abrundungsfehlern je eine Korrektur um ein bit vorgenommen, bis

$$\sum_i R_i^* = N \cdot \bar{R}$$

erfüllt ist. Es lässt sich zeigen, dass diese Rundungsprozedur die kleinstmögliche Zusatzverzerrung gegenüber den nichtgerundeten Werten von R_1 liefert /8,26/. Der angegebene Algorithmus führt auf eine Bitzuordnung, bei der nun alle Randbedingungen erfüllt sind. Die Auswirkung der beiden ersten Randbedingungen a und b wird in den nächsten Abschnitten noch ausführlicher diskutiert werden.

3. Codierung von Sprachsignalen mit nichtadaptiver Blockquantisierung

In diesem Abschnitt werden die Simulationsergebnisse mit Sprachsignalen für eine nichtadaptive Blockquantisierung angegeben. Die Ergebnisse sollen einerseits die theoretischen Betrachtungen aus Abschnitt 2 bestätigen und andererseits eine Vergleichsgrundlage geben, aus der die Verbesserungen durch adaptive Algorithmen zu entnehmen sind.

3.1 Beschreibung des Simulationssystems

Sprachproben:

Als Sprachsignale standen Sprachproben von 2 männlichen Sprechern zur Verfügung. Der eine Sprecher soll mit Spr.13 bezeichnet werden; er ist mit dem Spr. 13 aus der Veröffentlichung /1/ identisch. Der andere Sprecher wird mit Spr. 40 bezeichnet. Die Sprachsignale wurden mit 8 kHz abgetastet und sind wahlweise Tiefpass - (TP) oder Bandpass-(BP)- gefiltert mit dem Frequenzbereich 0 ...3.4 kHz bzw. 0.3 ...3.4 kHz. Für die folgenden Simulationsläufe wird in der Regel der Sprecher 13, BP mit einer Zeitdauer von $T = 1.6$ s verwendet. In dem abschliessenden Vergleich in Abschnitt 6 wird auf längere Sprachproben von Spr. 13 und Spr. 40 und auch auf BP- und TP-gefilterte Signale eingegangen.

Codiersystem:

Das Codiersystem ist aus Bild 2.1 zu entnehmen. Die Quantisierer für die Spektralkoeffizienten sind logarithmische Quantisierer mit $\mu = 100$ und 16 σ -Belastung (Aussteuergrenze = 16-facher Effektivwert des zu quantisierenden Signals). Ein Quantisierer mit logarithmischer Kennlinie ist infolge der starken Pegelschwankungen des Sprachsignals unbedingt erforderlich. Die Varianzen der Spektrallinien, die die Bitzuordnung und die Aussteuergrenzen der Quantisierer bestimmen, wurden vor der eigentlichen Simulation in einem Vorlauf gemessen. Die nichtadaptive Blockquantisierung mit logarithmischem Quantisierer trägt im folgenden die Bezeichnung BQ-LOG.

Messgrößen:

SNR_{BQ-LOG} : Der gemessene Signal-Rausch-Abstand bei der Blockquantisierung.

$SNR_{PCM-LOG}$: Der gemessene Signal-Rausch-Abstand bei einer PCM-Codierung mit dem gleichen log. Quantisierer ($\mu = 100$, 16 σ -Belastung) und der gleichen mittleren Bitrate \bar{R} .

Als Gewinn im Signal-Rausch-Abstand wird definiert:

$$G_{SNR} = SNR_{BQ-LOG} - SNR_{PCM-LOG} \quad (3.1.1)$$

3.2 Ergebnisse

Die Simulationen wurden mit einer mittleren Übertragungsrate von $\bar{R} = 4$ bit/Abtastwert durchgeführt. Bei einer Abtastfrequenz von 8 kHz entspricht 1 bit/Abtastwert einer Rate von 8 kbit/s; in diesem Fall ergibt sich also eine Übertragungsrate $R = 32$ kbit/s.

Der Sprecher ist Spr. 13-BP mit einer Zeit $T = 1.6$ s. Für die PCM-Codierung ergibt sich ein SNR von

$$\text{SNR}_{\text{PCM-LOG}} = 14.91 \text{ dB} \quad (3.2.1)$$

Die gemessenen SNR-Gewinne für die Blockquantisierung sind in Bild 3.1 angegeben. Diese können mit den theoretischen Gewinnen aus Bild 2.3 verglichen werden, da das AR-Modell 10. Ordnung genau aus dieser Sprachprobe gewonnen wurde. Es ergibt sich, dass die gemessenen Gewinne bis zu 1 dB unterhalb der theoretischen Werte liegen. Die Abwei-

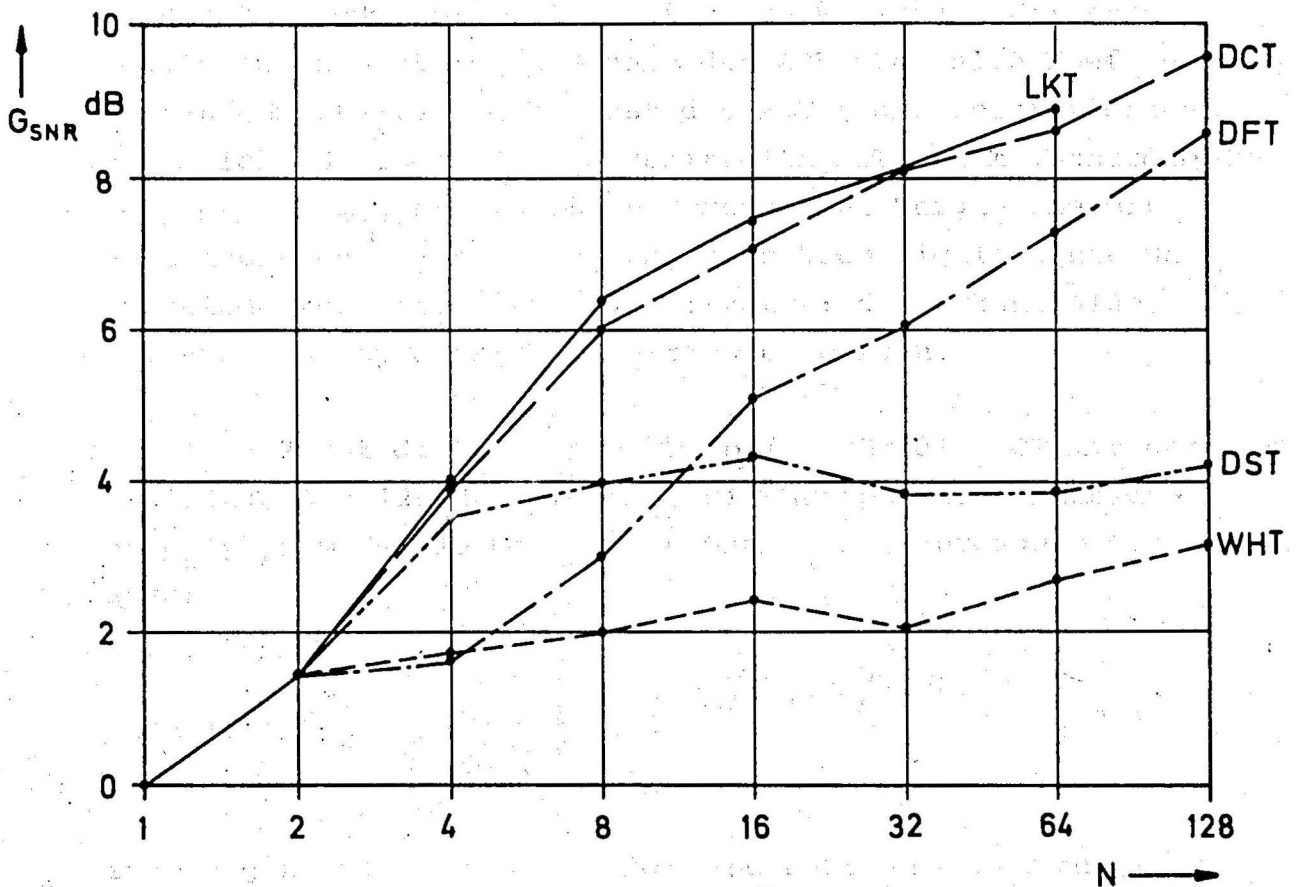


Bild 3.1 : Gemessener Gewinn G_{SNR} bei Blockquantisierung BQ-LOG in Abhängigkeit von der Blocklänge N .
($R=32$ kbit/s , Spr. 13-BP , $T=1.6$ s)

chung ist auf die nichtoptimale Bitzuordnung bei der Simulation zurückzuführen, da das Quantisierungsgesetz nach Gl. 2.1.21 nur näherungsweise gültig ist und ausserdem die berechneten Werte für R_i noch gerundet werden mussten. Insgesamt gesehen wird aber das Ergebnis der theoretischen Betrachtungen bestätigt. Es zeigt sich auch hier, dass die DCT -zumindest für das Langzeit-LDS von Sprache- die LKT ersetzen kann.

3.3 ADV der Spektralkoeffizienten

Zur ADV-Messung der Spektralkoeffizienten wurde zunächst die ADV für jeden einzelnen auf seine Varianz normierten Koeffizienten bestimmt und dann die ADV über alle Koeffizienten gemittelt. Dadurch hat die ADV jedes Koeffizienten das gleiche Gewicht; die so gemittelte ADV zeigt ausserdem erheblich kleinere statistische Streuungen. Vergleiche mit den Einzel-ADV's haben ergeben, dass kaum signifikante Unterschiede zur mittleren ADV auftreten; diese kann also durchaus als repräsentativ betrachtet werden.

In Bild 3.2 ist die mittlere ADV $p_m(y)$ für die DCT mit der Blocklänge $N = 128$ dargestellt. Die Messpunkte kennzeichnen $p_m(y)$; zum Vergleich ist die Gamma-ADV (durchgezogene Kurve)

$$p_G(y) = \frac{1}{2} \frac{\sqrt[4]{3}}{\sqrt{2\pi\sigma^2 |y|}} e^{-\sqrt{3} |y| / (2\sigma)} \quad (3.3.1)$$

eingetragen. Die Gamma-ADV ist eine sehr gute Näherung für die Sprach-ADV /27/. Bild 3.2 zeigt, dass die "Spektral-ADV" eine noch ausgeprägtere Spitze als die Gamma-ADV bzw. Sprach-ADV aufweist. Dies ist darauf zurückzuführen, dass zusätzlich zu den Pegelschwankungen des Sprachsignals, die sich proportional in den Frequenzbereich übertragen (Gl.2.1.13),

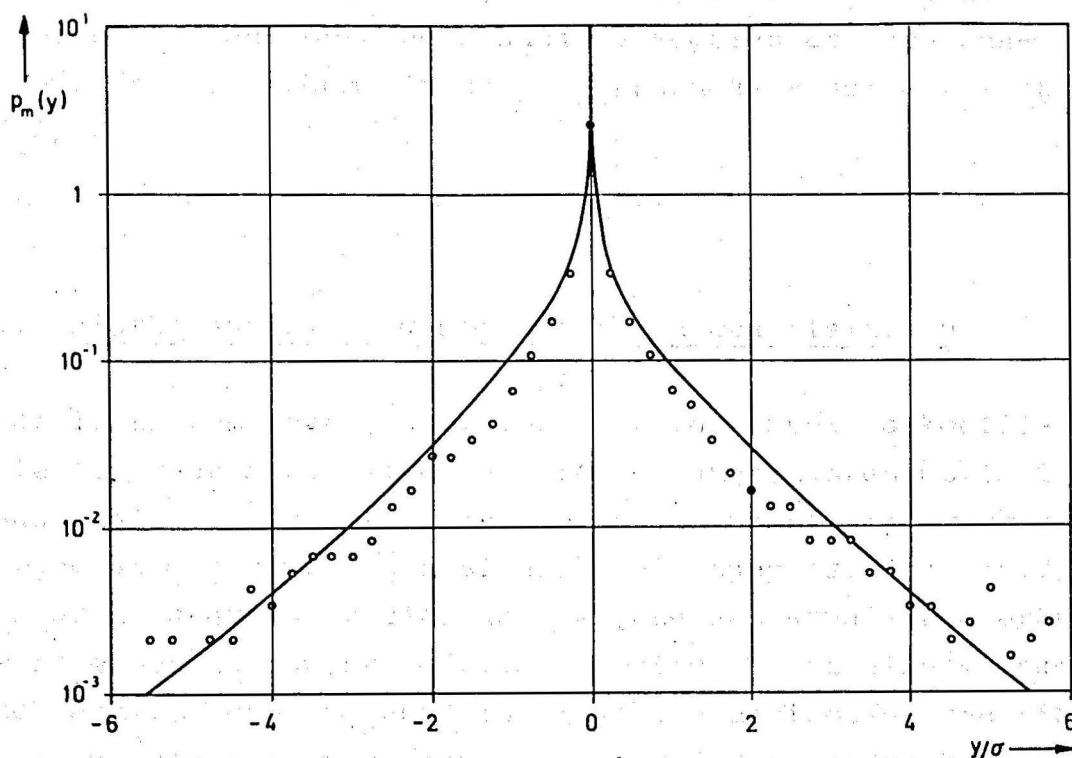


Bild 3.2 : Mittlere ADV der Spektralkoeffizienten $p_m(y)$
(Meßpunkte) und Vergleich mit einer Gamma-ADV.
(BQ-LOG , DCT , Blocklänge N=128)

noch Streuungen auswirken, die durch das sich zeitlich ändernde Kurzzeitspektrum verursacht werden. Je nach den Korrelationen in dem Datenblock kann der eine Spektralkoeffizient mal eine hohe, mal eine kleine Varianz aufweisen, auch wenn in beiden Fällen das Eingangssignal den gleichen Effektivwert aufweist!

Geht man auf andere Transformationen über oder variiert man die Blocklänge, so zeigt sich stets ein ähnlicher Verlauf der mittleren "Spektral-ADV" wie in Bild 3.2. Es ergibt sich daher folgende Konsequenz: Für eine nichtadaptive Blockquantisierung muss stets ein Quantisierer mit logarithmischer Kennlinie eingesetzt werden. Der Einsatz eines Opti-

malquantisierers ist wenig sinnvoll, weil er viel zu empfindlich auf sprecherabhängige Pegeländerungen reagiert und ausserdem ein zu grosses Ruhegeräusch aufweist /18/.

4. Blockquantisierung mit adaptiven Quantisierern

Der logarithmische Quantisierer für die Spektralkoeffizienten kann durch einen an die ADV angepassten Optimalquantisierer ersetzt werden, wenn die Quantisiereraussteuerung an den momentanen Signalpegel angepasst wird /6,7/. Dadurch vergrössert sich das SNR, und das Kanal-Ruhegeräusch wird erheblich verringert. Zusätzlich tritt eine starke subjektive Verbesserung ein, da die niedrigpegeligen Signalsegmente nun erheblich genauer codiert werden können. Die Quantisiereraussteuerung kann mit Vorwärts- oder Rückwärtsregelung bestimmt werden. In den folgenden Untersuchungen wird nur eine Vorwärtsregelung betrachtet, bei der die Quantisierer -Steuerparameter aus dem nicht codierten Signalsegment bestimmt werden und als zusätzlicher Datenstrom zum Empfänger zu übertragen sind. Die Bitzuordnung zu den einzelnen Spektralkoeffizienten bleibt dabei zeitinvariant und wird wie in Abschnitt 3 in einem Vorlauf aus der Signalstatistik bestimmt.

4.1. Beschreibung des Codiersystems

Das verwendete Codiersystem ist in Bild 4.1 gezeigt. Die Stufenhöhen der Quantisierer werden gemäss den in einem Vorlauf bestimmten Varianzen einmalig fest eingestellt. Der skalare Verstärkungsfaktor V bewirkt dann gemeinsam für alle Quantisierer die Anpassung an den aktuellen Signalpegel.

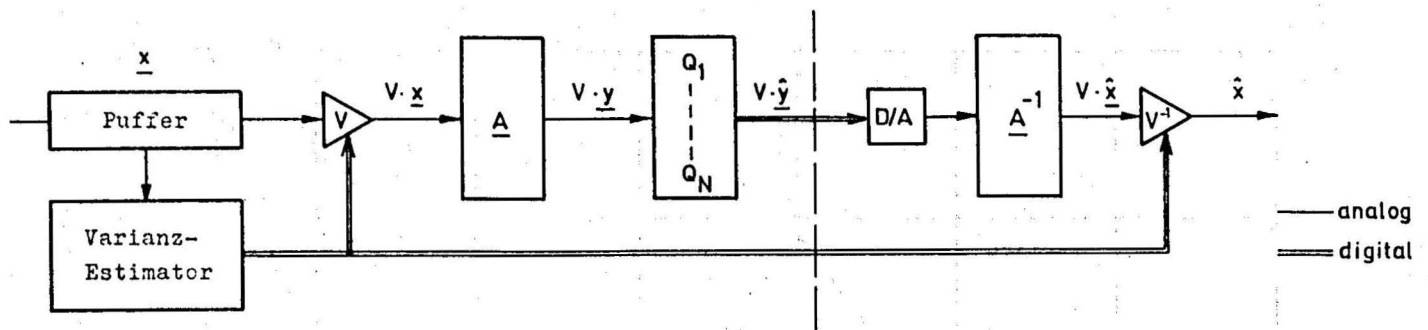


Bild 4.1 : Blockquantisierung mit adaptiven Quantisierern.

Der Wert V wird aus dem zu codierenden Vektor \underline{x} bestimmt:

$$V = \hat{\sigma}^{-2}, \text{ wobei} \quad (4.1.1)$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^N x_i^2$$

Die Grösse V wird anschliessend quantisiert und zum Decodierer übertragen. Über die erforderliche Genauigkeit der Quantisierung wird noch in Abschnitt 4.4 berichtet werden. Die Blockquantisierung mit adaptivem Quantisierer trägt im folgenden die Bezeichnung BQ-AQ.

4.2 ADV der Spektralkoeffizienten

Durch die pegelgesteuerte Sprache ergibt sich eine geänderte ADV der Spektralkoeffizienten. Es wurde wiederum die mittlere ADV $P_m(y)$ bestimmt (vgl. Abschnitt 3.3), die im Bild 4.2 für die DCT mit der Blocklänge $N=128$ dargestellt ist. Die durchgezogene Kurve im Bild ist die Laplace-ADV,

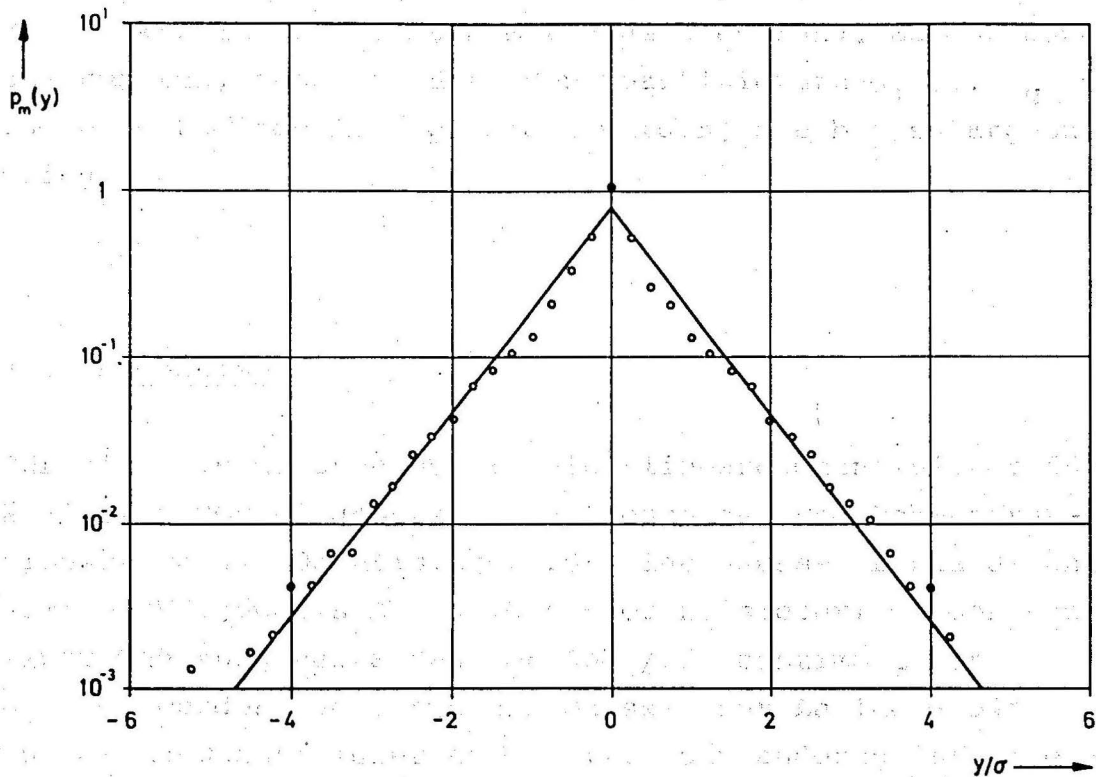


Bild 4.2 : Mittlere ADV der Spektralkoeffizienten $p_m(y)$ bei Pegelsteuerung (Meßpunkte) und Vergleich mit einer Laplace-ADV.

(BQ-AQ , DCT , Blocklänge $N=128$)

die durch

$$P_L(y) = \frac{1}{\sqrt{2} \sigma} e^{-\sqrt{2}|y|/\sigma} \quad (4.2.1)$$

beschrieben wird. Es zeigt sich, dass durch die Pegelsteuerung die ADV-Spitze abgebaut wird (vgl. Bild 3.2). Es entsteht aber keine Gauss-ADV, da immer noch starke Varianzschwankungen der Spektralkoeffizienten auftreten, die durch das zeitabhängige Kurzzeit-LDS der Sprache hervorgerufen werden (die Quantisierer werden nur an die mittlere Segment-

varianz bzw. an den Mittelwert der Spektralvarianzen σ_i^2 angepasst. Auch für andere Blocklängen und andere Transformationen ist die Laplace-ADV eine gute Modell-ADV für die Spektralkoeffizienten mit Pegelsteuerung. Es ist daher zu erwarten, dass ein Satz von Quantisierern $Q_1 \dots Q_N$, die für eine Laplace-ADV optimiert wurden, die besten Ergebnisse bringt.

4.3 Ergebnisse

Für die Simulationen wurden nichtlineare Quantisierer für kleinsten MQF eingesetzt. Die Übergangs- und Rekonstruktionswerte der Quantisierer für eine Gauss-, Laplace- und Gamma-ADV (jeweils 1 ... 10 bit-Quantisierer) wurden nach einem Näherungsverfahren von Roe /28/ bestimmt. Für $Q_1 \dots Q_N$ wurden zum einen nur Gauss-, nur Laplace- oder nur Gamma-Quantisierer eingesetzt; zum anderen ist untersucht worden, welche Verbesserung sich erreichen lässt, wenn für jede Spektrallinie der optimale Quantisierertyp aus den 3 angegebenen Möglichkeiten verwendet wird. Diese Zuordnung ist in einem Vorlauf bestimmt worden, in dem jede Spektrallinie mit einem Gauss-, Laplace- und Gamma-Quantisierer codiert wird und dann derjenige Quantisierer verwendet wird, der für die Spektrallinie die kleinste Verzerrung hervorruft ("indirekte ADV-Messung").

Als Gewinn G_{AQ} durch adaptive Quantisierung wird die SNR-Verbesserung gegenüber einer nicht adaptiven Blockquantisierung definiert:

$$G_{AQ} = \text{SNR}_{BQ-AQ} - \text{SNR}_{BQ-LOG} \quad (4.3.1)$$

Die Ergebnisse sind in Bild 4.3 für die DCT und DFT angegeben. Von den Möglichkeiten 2 ... 4 zeigt der Laplace-Quantisierer, wie zu erwarten, den grössten Gewinn. Das Ver-

fahren 1 (für jede Spektrallinie der optimale Quantisierer) kann das SNR nur noch um weniger als 1 dB verbessern und lohnt daher wegen des erhöhten Aufwandes nicht, zumal zu bedenken ist, dass die gefundene Zuordnung des Quantisierertyps bei dem geringen Stichprobenumfang wohl nur für die spezielle Sprachprobe gültig ist.

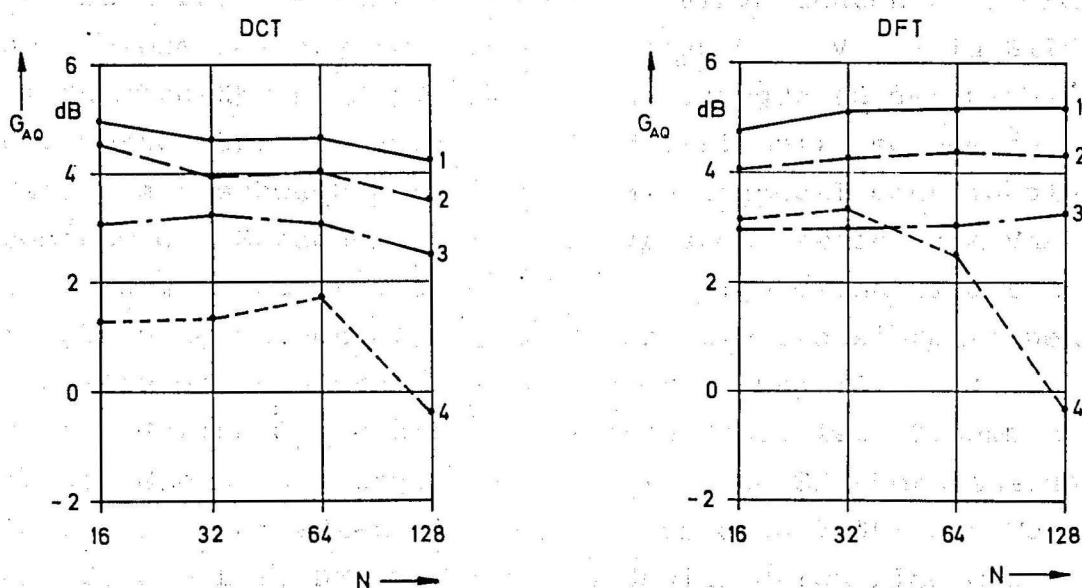


Bild 4.3 : Gemessener Gewinn G_{AQ} durch Verwendung gesteuerter Optimalquantisierer statt eines logarithmischen Quantisierers bei der Blocklänge N .
($R=32$ kbit/s , Spr. 13-BP , $T=1.6$ s)

- 1 : Für jede Spektrallinie der optimale Quantisierer
- 2 : Laplace-Quantisierer
- 3 : Gamma-Quantisierer
- 4 : Gauß-Quantisierer

Entsprechende Untersuchungen auch für andere Transformationen haben ergeben, dass generell mit einem gesteuerten Laplace-Quantisierer etwa ein Zusatzgewinn

$G_{AQ} = 4$ dB erreichbar ist.

Anmerkung:

Ein wesentlicher Hinweis ist notwendig zur Messung der Signalstatistik für die Bestimmung der zeitinvarianten Bitzuordnung. Diese Messung ist mit dem festen Wert $V = 1$ in Bild 4.1 durchzuführen. Wird die Messung dagegen an der pegelgesteuerten Sprache vorgenommen (V variabel), so beeinflussen die schwach korrelierten niedrigpegeligen Signalsegmente der Sprache mit viel zu grossem Gewicht die Varianzaufteilung im Spektralbereich. Als Folge davon wird eine Bitzuordnung berechnet, bei der für alle Transformationen bezüglich der spektralen Ungleichverteilung (Gl. 2.1.30) nur ein Gewinn $G_{BQ} = 2 \dots 3$ dB erreichbar ist. Zusammen mit dem Gewinn G_{AQ} ergibt sich damit eine SNR-Verbesserung gegenüber einer PCM-LOG-Codierung von etwa 7 dB; ein Wert, der für die LKT, DFT und WHT als Gewinn durch Blockquantisierung mit AQ in der Untersuchung /8/ angegeben wurde. Verwendet man dagegen die Signalstatistik aus der nicht pegelgesteuerten Sprache, so ist - je nach Transformation - eine SNR-Verbesserung um $G_{BQ} + G_{AQ}$ gegenüber PCM-LOG bis zu 14 dB erreichbar.

4.4 Quantisierung des Verstärkungsfaktors V zur Pegelsteuerung

Beim Einsatz von adaptiven Quantisierern mit Vorwärtsregelung muss die Information über den Verstärkungsgrad V des Verstärkers vor dem Quantisierereingang zusätzlich zum Decodierer übertragen werden. Dazu ist eine Quantisierung des zunächst analogen Wertes V notwendig. Die folgende Un-

tersuchung gibt eine Aussage über die erforderliche Genauigkeit bei der Quantisierung des Wertes V . Als Dynamikbereich wird ein fester Bereich von 80 dB vorgegeben. Die Verstärkungswerte V werden logarithmisch quantisiert, so dass der maximale relative Quantisierungsfehler $\Delta V/V$ innerhalb des zulässigen Dynamikbereiches stets gleich gross ist.

Bild 4.4 zeigt die Ergebnisse für eine Simulation mit BQ - AQ, wenn pro Block n_V bit zur Quantisierung von V aufgewendet werden. Es ist ersichtlich, dass für $n_V = 5$ noch kein signifikanter SNR-Rückgang auftritt (bei dieser Simulation ergab sich durch statistische Streuungen sogar ein SNR-Anstieg gegenüber $n_V \rightarrow \infty$). Der maximale Quantisierungsfehler $\Delta V/V$ beträgt dabei 1.25 dB, d.h., Fehlaussteuerungen der Optimalquantisierer bis etwa 1 dB machen sich überhaupt nicht bemerkbar.

Die Datenrate R_V kennzeichne den Informationsfluss für die Verstärkungswerte V . Bei einer Blocklänge $N = 128$, Abtastfrequenz $f_A = 8$ kHz und $n_V = 5$ ergibt sich

$$R_V = 0.3125 \text{ kbit/s} \quad (4.4.1)$$

Die Datenrate R soll auch für die zukünftigen Betrachtungen stets den alleinigen Datenstrom aus den Quantisiererentscheidungen für die Spektralkoeffizienten kennzeichnen. Die gesamte Übertragungsrate R_{ges} ergibt sich jeweils aus R und der Zusatzinformation:

$$R_{\text{ges}} = R + R_V \quad (4.4.2)$$

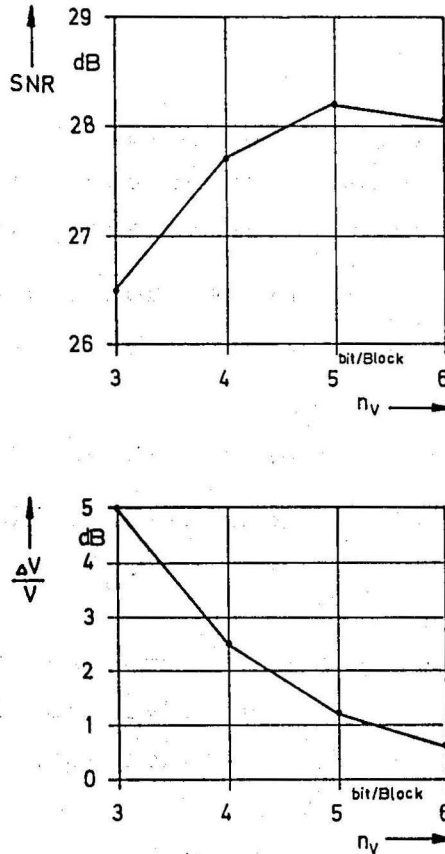


Bild 4.4 : SNR und maximaler Quantisierungsfehler $\Delta V/V$ ohne Übersteuerung in Abhängigkeit von der Bitzahl n_V (bit/Block), mit der der Verstärkungsfaktor V quantisiert wird.
(BQ-AQ , DCT , $N=128$, Laplace-Quant. ,
 $R=32$ kbit/s , Spr. 13-BP , $T=1.6$ s)

5. Blockquantisierung mit adaptiven Quantisierern und adaptiver Bitzuordnung

Eine weitergehende Anpassung des Codiersystems an die Signalstatistik der Sprache wird erst erreicht, wenn auch das sich zeitlich ändernde Kurzzeitspektrum des Sprachsignals berücksichtigt wird. Zum Erreichen des maximal möglichen Signal-Rausch-Abstandes wären dabei folgende Schritte notwendig, die für jeden Datenblock x erneut durchzuführen

ren sind:

1. Bestimmung der optimalen Transformation \underline{A} (LKT) aus der Statistik des zu codierenden Vektors \underline{x} und Durchführung der Transformation.
2. Bestimmung der normierten Varianzen der Spektrallinien σ_i^2 / σ^2 bei der gewählten Transformation, daraus ergeben sich:
 - a) Die optimale Bitzuordnung für die Quantisierer Q_1, \dots, Q_N .
 - b) Die individuellen Aussteuerungen für jeden Quantisierer $Q_1 \dots Q_N$.
3. Bestimmung des Segmentpegels σ^2 zur Steuerung des skalaren Verstärkungsfaktors V , der gemeinsam für alle Quantisierer die Anpassung an den momentanen Signalpegel durchführt.

Der Schritt 1 ist besonders aufwendig, da die Durchführung und vor allem die Bestimmung der optimalen LKT ausserordentlich rechenintensiv ist. In den folgenden Abschnitten wird daher untersucht, welche Vereinfachungen möglich sind.

Diese Vereinfachungen sollen einerseits nur zu einer geringen Gewinnabnahme gegenüber der optimalen Lösung führen, andererseits aber noch gut in Realtime-Hardware realisierbar sein.

5.1. Vorbetrachtungen an Modellquellen

5.1.1 Theoretische Gewinne für verschiedene Sprachlaute

Bei der bisherigen Betrachtung der theoretischen Gewinne verschiedener Transformationen ist bisher stets das Langzeit-LDS der Sprache zugrunde gelegt worden. Es soll nun untersucht werden, wie sich die Transformationen bei verschiedenen typischen Sprachlauten verhalten. Dazu wurde

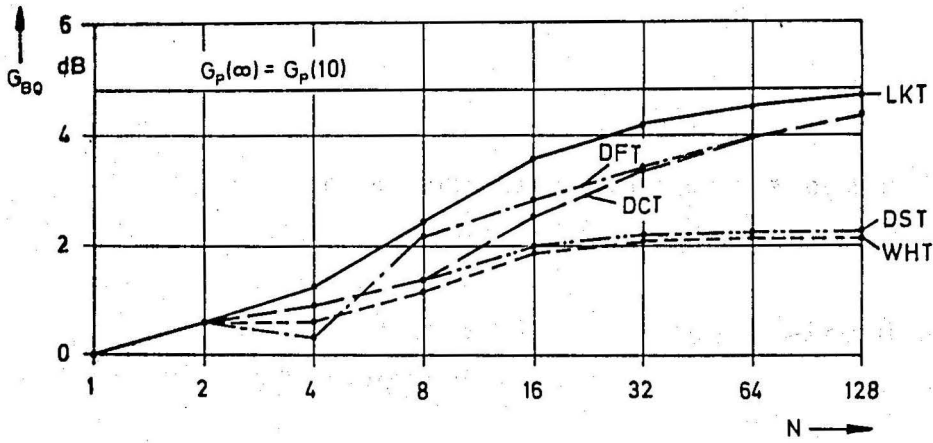
eine Reihe von Sprachlauten analysiert, für jeden Laut die ersten $M = 10$ AKF-Werte bestimmt und daraus das dazugehörige AR-Modell der Ordnung M berechnet. Aus dem AR-Modell können sofort, wie im Abschnitt 2.3.2 beschrieben, die theoretischen Gewinne G_{BQ} ermittelt werden. Bild 5.1 zeigt einen repräsentativen Querschnitt für Laute mit kleiner ("sch"), mittlerer ("ü") und hoher Korrelation ("m") im Signalverlauf.

Einige Ergebnisse sind bemerkenswert:

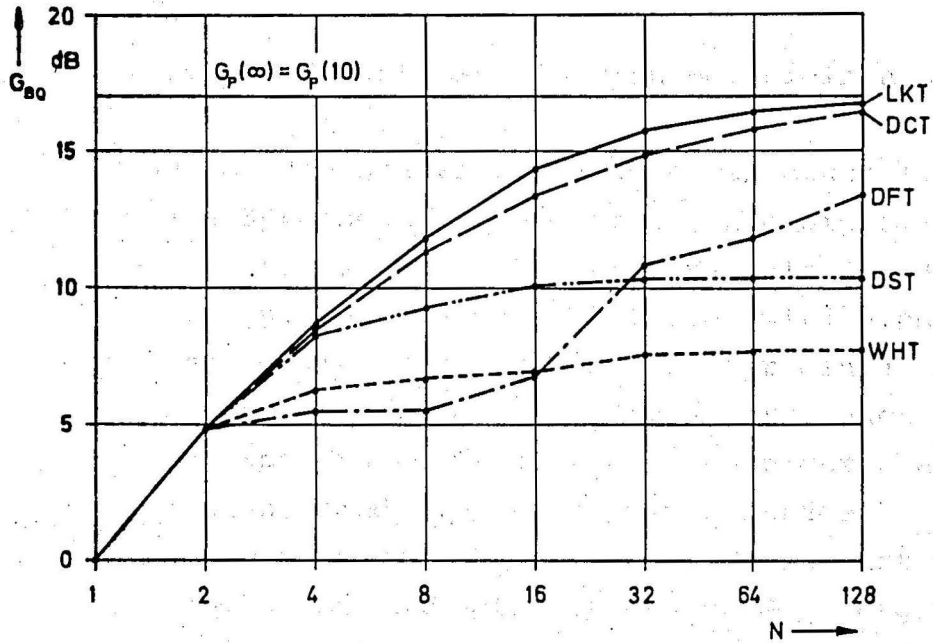
- * Die DCT ist der LKT im Gewinn nur um höchstens 2 dB, oft aber weniger als 1 dB, unterlegen.
- * Die DFT arbeitet, obwohl sie asymptotisch optimal ist, bei den untersuchten Blocklängen bis zu 13 dB schlechter als die DCT und wird in Einzelfällen sogar von der weniger rechenintensiven DST erheblich übertroffen.
- * Die Frage, welche Transformation geeignet ist, hängt ausser bei LKT und DCT sehr stark von den gewählten Blocklängen und der Statistik des zu codierenden Prozesses ab.

Es zeigt sich also, dass die DCT für die Sprachcodierung als "schnelle LKT" aufgefasst werden kann. Durch Wahl der DCT als zeitinvariante Transformation A sind zwei Probleme gelöst:

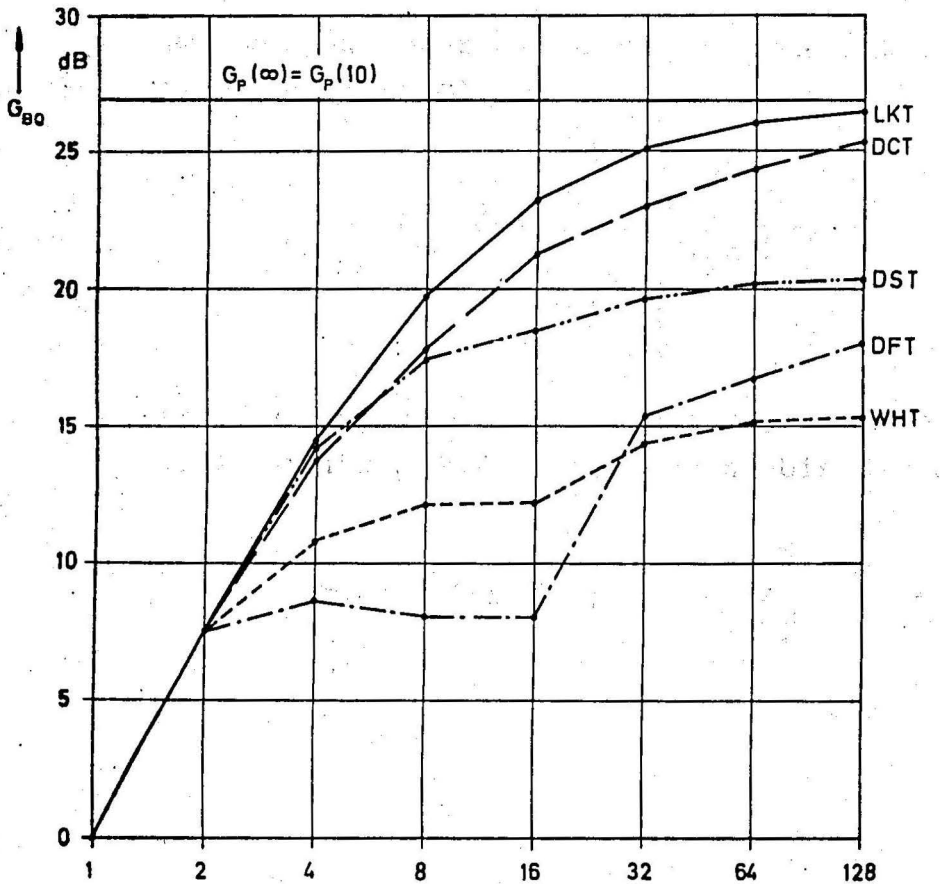
1. Es ist nicht notwendig (bzw. wenig lohnend), in Abhängigkeit von der Statistik des Eingangssignals jeweils erneut eine optimale LKT zu berechnen oder aus einer vorgegebenen Klasse von verschiedenen LKT-Matrizen eine Transformation auszuwählen.



Laut "sch"



Laut "ü"



Laut "m"

Bild 5.1 :
Theoret. Gewinn G_{BQ}
bei der Blocklänge N
für verschiedene AR-
Prozesse 10. Ordnung.

2. Die Transformation kann mit einem schnellen Algorithmus durchgeführt werden.

Für die weiteren Untersuchungen wird daher bevorzugt die DCT angewendet werden.

5.1.2 Suboptimaler Gewinn bei kleinen Bitraten

Durch die adaptive Blockquantisierung ist es möglich, für die Sprachcodierung ausreichende Signal-Rausch-Abstände bereits bei kleineren Bitraten als $R = 32$ kbit/s zu erzielen. Bei kleineren Bitraten wird nicht der theoretische Gewinn G_{BQ} erreicht, da für einige Spektralkoeffizienten negative Bitzahlen R_i errechnet werden. Da diese Spektralkoeffizienten dann nicht übertragen werden können, sind die Verzerrungsbeiträge D_i im Spektralbereich nicht mehr gleich gross, der Gewinn ist dadurch kleiner geworden. Der suboptimale Gewinn G_{sub} ergibt sich aus folgender Betrachtung:

Bei vorgegebener mittlerer Bitrate \bar{R} werden zunächst die Werte R_i nach Gl. 2.1.26 errechnet:

$$R_i = \bar{R} + \frac{1}{2} \text{ld} \frac{\sigma_i^2}{\left[\prod_{j=1}^N \sigma_j^2 \right]^{\frac{1}{N}}} \quad (2.1.26)$$

Es wird nun eine Bitumverteilung entsprechend dem Verfahren in Abschnitt 2.4 vorgenommen, bis die erste Randbedingung

$$R_i \geq 0 \quad \text{für } \forall i, \text{ wobei } \sum_{i=1}^N R_i = N \cdot \bar{R} \quad (5.1.1.)$$

erfüllt ist. Es werden nun zwei Index-Mengen gebildet:

I : Indexmenge für alle i mit $R_i > 0$.

I^* : Indexmenge für alle i mit $R_i = 0$.

Die Spektralkoeffizienten mit $R_i = 0$ werden nicht übertragen; am Empfänger wird dafür stets der Rekonstruktionswert Null gemeldet. Es entstehen dann folgende Verzerrungen:

$$D_i = \sigma_i^2 \cdot 2^{2\delta} \cdot 2^{-2R_i} \quad , \quad i \in I \quad . \quad (2.1.22)$$

$$D_i = \sigma_i^2 \quad , \quad i \in I^* \quad . \quad (5.1.2)$$

Damit wird die mittlere Gesamtverzerrung

$$D = \frac{1}{N} \sum_{i=1}^N D_i = \frac{1}{N} \left[\sum_{i \in I} D_i + \sum_{i \in I^*} \sigma_i^2 \right] \quad . \quad (5.1.3)$$

Der Vergleich mit der PCM-Verzerrung

$$D_{\text{PCM}} = 2^{2\delta} \cdot 2^{-2\bar{R}} \cdot \sigma^2 \quad (2.1.29)$$

ergibt den suboptimalen Gewinn

$$G_{\text{sub}} = 10 \lg \frac{D_{\text{PCM}}}{D} = 10 \lg \frac{2^{2\delta} \cdot 2^{-2\bar{R}} \cdot \sigma^2}{\frac{1}{N} \left[\sum_{i \in I} 2^{2\delta} \sigma_i^2 \cdot 2^{-2R_i} + \sum_{i \in I^*} \sigma_i^2 \right]} \quad (5.1.4)$$

Für die folgende Abschätzung wird der Wert δ gleich Null gesetzt, d. h., als Quantisierungsgesetz wird die Rate-Distortion-Funktion einer gedächtnisfreien Gaussquelle /26/ angenommen:

$$G_{\text{sub}} = 10 \lg \frac{2^{-2\bar{R}} \cdot \sigma^2}{\frac{1}{N} \left[\sum_{i \in I} 2^{-2R_i} \cdot \sigma_i^2 + \sum_{i \in I^*} \sigma_i^2 \right]} \quad (5.1.5)$$

In Bild 5.2 sind die Werte für G_{sub} in Abhängigkeit von der Blocklänge N und der Übertragungsrate R angegeben. Die Beispiele sind die schon in den früheren Abschnitten beschriebenen AR-Prozesse. Als Transformation wurde die DCT verwendet, die wegen ihres grossen Gewinns (und damit der grössten spektralen Ungleichverteilung) eine obere Schranke des Gewinnrückgangs $G_{\text{BQ}} - G_{\text{sub}}$ für die anderen schnellen Transformationen angibt.

Die Ergebnisse zeigen:

- * Bei Übertragungsraten $R \geq 32$ kbit/s tritt kein Gewinnrückgang ein.
- * Auch wenn die Übertragungsrate auf $R = 16$ kbit/s reduziert wird, geht der Gewinn um weniger als 1 dB zurück.
- * Bei einer Übertragungsrate von $R = 8$ kbit/s geht der Gewinn zwar um etwa 3 dB zurück, kann aber durchaus noch Werte über 22 dB erreichen.
- * Der Gewinnrückgang durch Verminderung der Übertragungsrate ist relativ unabhängig davon, wie stark das Signal korreliert ist. Dies ist ein erheblicher Vorteil gegenüber (A)DPCM-Systemen. Die Rückkopplung des Quantisierungsfehlers in diesen Codiersystemen kann bei stark korrelierten Signalen und kleinen Bitraten zu einem extremen Gewinnrückgang führen (vgl. Abschnitt 6.1)

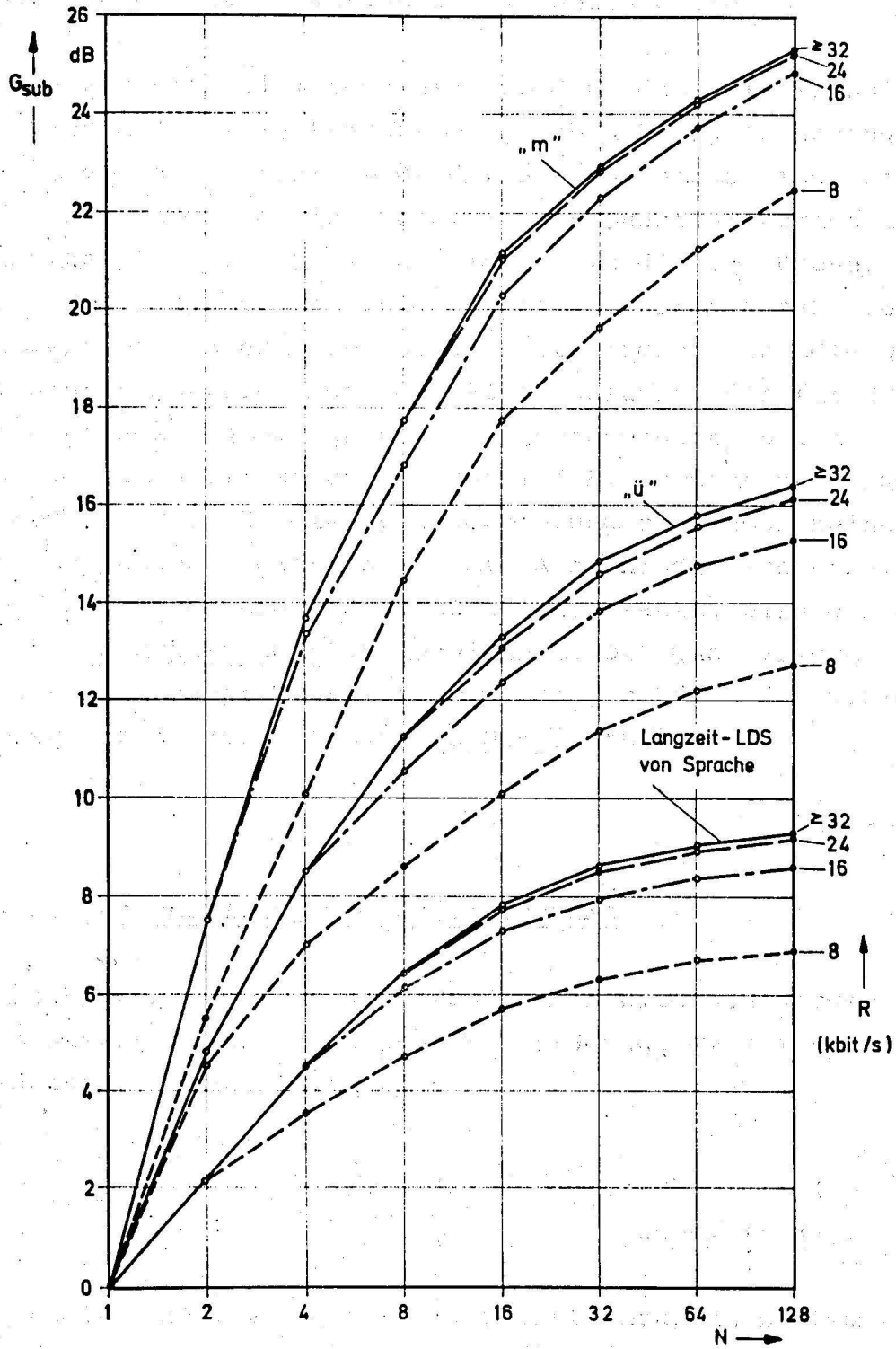


Bild 5.2 : Suboptimaler Gewinn G_{sub} bei kleinen Bitraten R (kbit/s) für verschiedene AR-Prozesse 10. Ordnung. (Transformation: DCT)

5.2 Beschränkt adaptive Blockquantisierung

In Abschnitt 5.1 ist gezeigt worden, dass die Transformation A (DCT) zeitinvariant bleiben kann. In Abhängigkeit von dem Kurzzeit-LDS des Sprachsignals sind nur noch die Bitzuordnung und die einzelnen Quantisiereraussteuerungen für $Q_1 \dots Q_N$ zu bestimmen. In diesem Abschnitt wird ein Codiersystem untersucht, das aus einer Anzahl fest vorgegebener Mustersätze für die Bitzuordnung und die Quantisiereraussteuerungen den Satz aussucht, der für das zu codierende Signalsegment x am geeignetsten ist. Ein vergleichbares Verfahren ist für die Bildcodierung eingesetzt worden /10/; für die Sprachcodierung sind noch keine Ergebnisse bekannt geworden. Da die Auswahl der Mustersätze sich nur auf einen endlichen Vorrat beschränkt, wird der Codierer als beschränkt adaptiv bezeichnet. Die Quantisierer sollen weiterhin adaptiv gesteuert werden, so dass das Codiersystem die Kurzbezeichnung BA-BQ-AQ erhält.

5.2.1 Beschreibung des Codiersystems

Das verwendete Codiersystem ist in Bild 5.3 dargestellt. Es stehen L Auswahlätze zur Verfügung. Zu dem k-ten Auswahlatz gehört der Vektor

$$\underline{\alpha}^{(k)} = (\alpha_1^{(k)}, \dots, \alpha_N^{(k)})^T, k \in I \quad (5.2.1)$$

wobei $I = \{1, 2, \dots, L\}$

der die einzelnen Quantisiereraussteuerungen bestimmt und der Vektor

$$\underline{R}^{(k)} = (R_1^{(k)}, \dots, R_N^{(k)})^T, k \in I \quad (5.2.2)$$

der die Bitzuordnung festlegt ($R_i^{(k)}$ ist die Bitzahl der

i -ten Spektrallinie in der k -ten Klasse). Der Klassifikator ermittelt unter Kenntnis des zu codierenden Vektors \underline{x} die Klasse k , mit der die kleinste Codierverzerrung zu erwarten ist. Zusätzlich zum Verstärkungswert V muss dann auch der Klassenindex k mit $\text{ld } L$ bit über den digitalen Kanal übertragen werden, um dem Empfänger die richtige Decodierung zu ermöglichen.

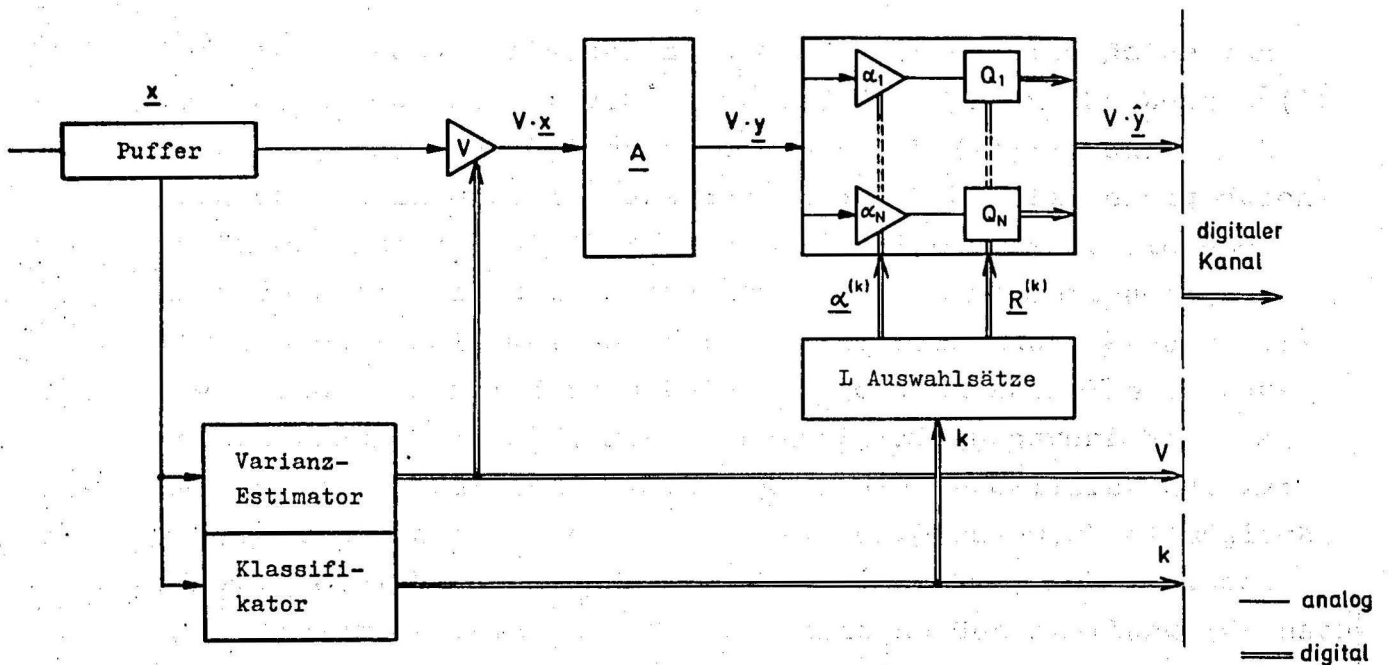


Bild 5.3 : Blockquantisierung mit beschränkt adaptiver Bitzuordnung und adaptiven Quantisierern.

(BA-BQ-AQ)

Die Vektoren $\underline{\alpha}^{(k)}$ und $\underline{R}^{(k)}$; $k = 1, \dots, L$ werden in einem Vorlauf bestimmt ("Lernphase"). Dazu werden für jede Klasse getrennt die Varianzen $\sigma_{i,k}^2$ (k : = k -te Klasse) gemessen und dann die Bitzuordnungen $\underline{R}^{(k)}$ bestimmt. Der Vektor $\underline{\alpha}^{(k)}$ wird dann durch die Reziprokwerte der normierten Standardabweichungen der Spektralkoeffizienten festgelegt:

$$\underline{\alpha}^{(k)} = \left(\frac{\sigma}{\sigma_{1,k}}, \dots, \frac{\sigma}{\sigma_{N,k}} \right)^T \quad (5.2.3)$$

Die Stufenhöhen der Quantisierer werden dann für die Varianz $\sigma^2 = 1$ ausgelegt und bleiben unverändert.

Zur Einrichtung des Klassifikators gibt es eine Reihe von Möglichkeiten. Es ist z.B. denkbar, den ersten AKF-Wert $\rho(1)$ zu bestimmen, den möglichen Wertebereich $(-1, +1)$ dann in L Bereiche zu unterteilen und danach die Klassifizierung durchzuführen. Für die Simulationen wurde eine Klassifizierung nach der Zahl der Nulldurchgänge im Vektor \underline{x} vorgenommen. Für Sprachsignale weisen stark korrelierte Laute eine geringere Zahl von Nulldurchgängen als schwach korrelierte Laute und Zischlaute auf. Dieses Klassifizierungsverfahren hat den Vorteil einer sehr einfachen Hardware-Realisierung. Ist N die Blocklänge, so sind $0 \dots N-1$ Vorzeichenwechsel möglich. Dieser Bereich wurde in L Klassen unterteilt, wobei einmal eine lineare, einmal eine logarithmische Unterteilung (Wachsen der Klassenbreite mit zunehmender Zahl der Vorzeichenwechsel) vorgenommen wurde. Mit der logarithmischen Klassenunterteilung sind höhere Gewinne zu erwarten, da dann stärker zwischen den tieffrequenten stark korrelierten Lauten differenziert werden kann.

5.2.2 Ergebnisse

Als Gewinn durch die beschränkt adaptive Bitzuordnung wird

$$G_{BA} = \text{SNR}_{BA-BQ-AQ} - \text{SNR}_{BQ-AQ} \quad (5.2.4)$$

definiert. Bild 5.4 zeigt die Ergebnisse für eine DCT mit einer Blocklänge $N = 128$ in Abhängigkeit von der Klassenzahl L . Für $L < 8$ ergibt die logarithmische Klassenunterteilung tatsächlich einen höheren Gewinn. Eine wesentliche Vergrößerung von L über 8 hinaus scheint wenig sinnvoll, weil dann die Statistik ($\underline{\alpha}^{(k)}$ und $\underline{R}^{(k)}$) zu sehr auf die Sprachprobe der Lernphase zugeschnitten wird und bei anderen Sprechern dann grössere Fehlanpassungen zu erwarten sind.

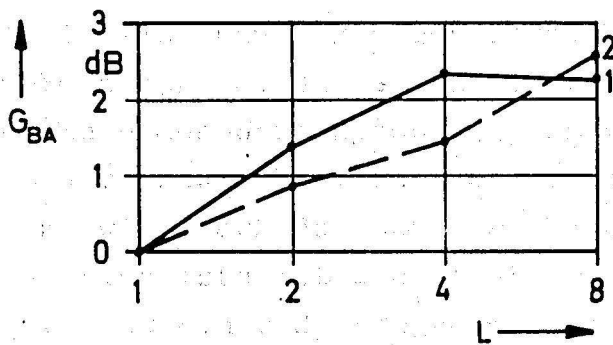


Bild 5.4 : Gewinn G_{BA} durch beschränkt adaptive Bitzuordnung (BA-BQ-AQ) in Abhängigkeit von der Klassenzahl L . (DCT , $N=128$, Laplace-Quant. , $R=32$ kbit/s , Spr. 13-BP , $T=1.6$ s)

- 1 : log. Klassenunterteilung
- 2 : lin. Klassenunterteilung

Wesentliche Ergebnisse sind:

- * Der Zusatzgewinn im SNR beträgt bei einer beschränkt-adaptiven Bitzuordnung mit 8 Klassen etwa 2.5 dB.
- * Der Laplace-Quantisierer bringt wie bei BQ-AQ die besten Ergebnisse.
- * Ausser dem objektiven SNR-Anstieg ergibt sich auch ein starker subjektiver Gewinn, da nun für Sprachlaute, die stark von der Langzeitstatistik abweichen, eine erheblich günstigere Bitzuordnung eingesetzt wird.

5.3 Fehlanpassungen in der Signalstatistik

Für alle bisher betrachteten Codiersysteme war es notwendig, zumindest einige Parameter (Bitzuordnung und Quantisiereraussteuerungen) aus der Langzeit-Signalstatistik zu ermitteln, bevor mit der eigentlichen Codierung begonnen werden konnte. Die Signalstatistik wurde dabei aus der zu codierenden Sprachprobe gewonnen. Dieser Idealfall wird aber in der Praxis kaum auftreten. Es soll daher untersucht werden, wie stark das SNR zurückgeht, wenn Fehlanpassungen in der Signalstatistik eintreten. Der Einfluss der Fehlanpassung wurde an einer Sprachprobe (Spr. 13-BP) ermittelt. Das Sprachsignal für $t = 0 \dots 1.6$ s wird codiert. Die Signalstatistik wird ermittelt aus dem gleichen Sprachsignal, aber:

- 1) Optimal : Statistik aus $t = 0 \dots 1.6$ s
- 2) Fehl-1 : " " $t = 0.8 \dots 2.4$ s
- 3) Fehl-2 : " " $t = 1.6 \dots 3.2$ s

Diese drei Fälle kennzeichnen vollständige, halbseitige und keine Überlappung von Codier- und Statistiksignal. Die bei der Simulation erhaltenen SNR-Werte sind in Tabelle 5.1 an-

gegeben:

Codierung	SNR in dB		
	Optimal	Fehl-1	Fehl-2
BQ-LOG	24.51	19.44	19.46
BQ-AQ (Laplace-Q.)	28.03	23.93	24.08
BA-BQ-AQ (Laplace-Q. 8 Klassen)	30.25	26.74	26.89

Tabelle 5.1: SNR-Werte bei Fehlanpassung
(DCT, N = 128, R = 32 kbit/s)

Es zeigt sich, dass schon bei geringeren Abweichungen in der Signalstatistik (halbseitige Überlappung: Fehl-1) ein SNR-Rückgang von 3...5 dB auftritt, der sich dann aber bei grösseren Abweichungen kaum noch vergrössert. Die Empfindlichkeit gegenüber Abweichungen ist offenbar darauf zurückzuführen, dass eine sehr grosse Zahl von Parametern (bei N = 128 sind dies 128 Werte für die Quantisiereraussteuerung) aus der Signalstatistik bestimmt werden muss. Die SNR-Unterschiede und damit die Gewinne der einzelnen Blockquantisierungsverfahren bleiben aber etwa erhalten. Lediglich die absoluten SNR-Werte, die bisher stets bei optimaler Signalstatistik bestimmt wurden, sind als obere Schranken anzusehen.

5.4 Volladaptive Blockquantisierung

Eine weitere Erhöhung des Signal-Rausch-Abstandes ist erreichbar, wenn die Bitzuordnung optimal an den zu codierenden Datenblock \underline{x} angepasst wird und nicht aus einer Klasse von vorgegebenen Mustersätzen gewählt werden muss. Dadurch wird eine vollständige Unabhängigkeit von sonst vorausgehenden Statistikmessungen erreicht; der Codierer arbeitet dann stets für jeden Sprecher optimal. Die Aufgabe besteht also darin, die Bitzuordnung und die Aussteuerungen für die Quantisierer $Q_1 \dots Q_N$ unter Kenntnis des Datenvektors \underline{x} optimal zu wählen. Es wäre daher folgender Weg denkbar (ideales, aber kaum realisierbares Verfahren):

1. Aus dem Datenblock \underline{x} werden die ersten M AKF-Werte $\hat{\mathcal{P}}(1) \dots \hat{\mathcal{P}}(M)$ geschätzt (z. B. $M = 10$).
2. Aus $\hat{\mathcal{P}}(1) \dots \hat{\mathcal{P}}(M)$ werden die PARCOR-Koeffizienten $\hat{\pi}(1) \dots \hat{\pi}(M)$ berechnet /22/, quantisiert ($\hat{\pi}(1), \hat{\pi}(2), \dots, \hat{\pi}(M)$) und zum Decodierer übertragen. Die PARCOR-Koeffizienten sind zur Übertragung besonders geeignet, da sie sehr grob quantisiert werden können /22/.
3. Codierer und Decodierer berechnen aus den quantisierten Werten $\hat{\pi}(1) \dots \hat{\pi}(M)$ wieder die dazugehörigen AKF-Werte $\hat{\mathcal{P}}(1) \dots \hat{\mathcal{P}}(M)$ und verlängern daraus die AKF bis zum Wert $\hat{\mathcal{P}}(N-1)$. Die Verlängerung wird so erreicht, dass den AKF-Werten $\hat{\mathcal{P}}(1) \dots \hat{\mathcal{P}}(M)$ ein AR-Modell der Ordnung M zugeordnet wird und dann die fehlenden AKF-Werte aus diesem theoretischen Modell berechnet werden /22/.
4. Damit liegt die geschätzte Autokovarianzmatrix $\hat{\underline{R}}_{xx}$ vor. Aus $\underline{A} \cdot \hat{\underline{R}}_{xx} \cdot \underline{A}^T$ (Gl. 2.1.9) werden die geschätzten Varianzen der Spektralkoeffizienten $\hat{\sigma}_i^2$ entnommen. Mit den Werten $\hat{\sigma}_i^2$ werden die optimale Bitzuordnung und die Quantisiereraussteuerungen bestimmt.

Dieses Verfahren dürfte wohl die besten Ergebnisse bringen. Es ist aber so rechenintensiv (wegen der Matrizenmultiplika-

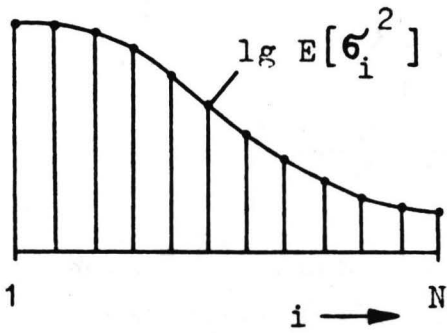
tionen in $\underline{A} \cdot \hat{\underline{R}}_{xx} \cdot \underline{A}^T$), dass bei einer Blocklänge $N = 128$ noch nicht einmal eine Simulation auf dem Digitalrechner sinnvoll ist. Im nächsten Abschnitt wird daher ein Codierverfahren angegeben, das dem geschilderten etwa entspricht, aber nur mit einem Bruchteil der Rechenoperationen auskommt.

5.4.1 Beschreibung des Codiersystems

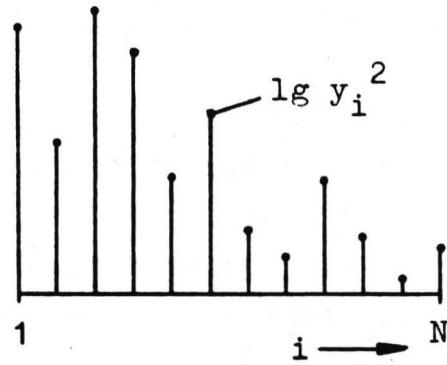
Die Bestimmung der Varianz-Schätzwerte $\hat{\sigma}_i^2$ aus dem Produkt $\underline{A} \cdot \hat{\underline{R}}_{xx} \cdot \underline{A}^T$ ist sehr aufwendig. Da die Transformation $\underline{y} = \underline{A} \cdot \underline{x}$ sowieso durchgeführt werden muss, ist es viel einfacher, die Varianzen direkt aus dem Vektor \underline{y} zu berechnen. Aus \underline{y} werden L Stützwerte des Spektrums ermittelt, quantisiert und (als Zusatzinformation) zum Decodierer übertragen. Das Verfahren zur Gewinnung der spektralen Zusatzinformation ist in Bild 5.5 skizziert. Die obere Zeichnung zeigt den Erwartungswert des Kurzzeitspektrums. Darunter sind die tatsächlichen Quadrate der Spektralkoeffizienten y_i^2 angegeben, die natürlich sehr stark von ihren Erwartungswerten abweichen können. Durch Mittelung von je $M = N/L$ benachbarter Quadrate mit einem Rechteckfenster werden L Stützwerte $\sigma_{s,1}^2$ für das zu schätzende Spektrum gewonnen, die nun eine erheblich kleinere Streuung aufweisen als die ursprünglichen Spektralkoeffizienten. Die Stützwerte werden quantisiert ($\sigma_{s,1}^2 \rightarrow \hat{\sigma}_{s,1}^2$) und zum Decodierer übertragen. Die Interpolation der quantisierten Stützwerte führt auf das vollständige geschätzte Kurzzeitspektrum.

Das Verfahren läuft im einzelnen wie folgt ab:

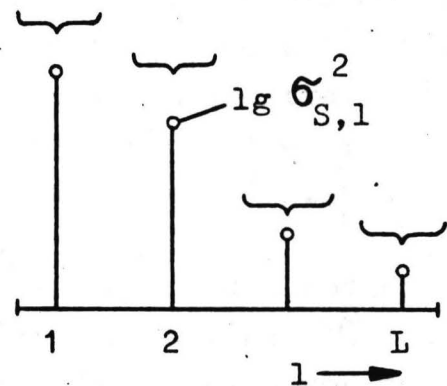
1. Der Datenvektor \underline{x} wird transformiert: $\underline{y} = \underline{A} \cdot \underline{x}$
2. Es werden L Stützwerte $\sigma_{s,1}^2$ ($1 = 1, \dots, L$) des Spektrums aus jeweils $M = \frac{N}{L}$ benachbarten Spektralkoeffizienten y_i gebildet:



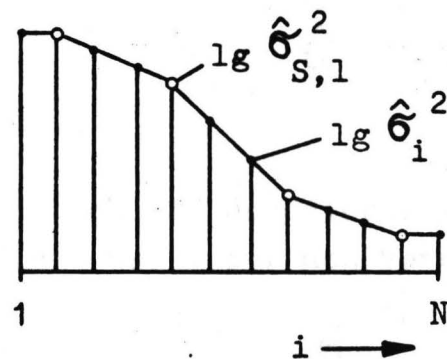
Erwartungswert des
Kurzzeitspektrums



Quadrate der
Spektralkoeffizienten



Stützwerte für das
zu schätzende Spektrum



Interpolation der quanti-
sierten Stützwerte:
vollständiges geschätztes
Spektrum

Bild 5.5 : Gewinnung der spektralen Zusatzinformation und Interpolation für die volladaptive Blockquantisierung.
(Zeichnung für $N=12$, $L=4$ und $M=3$)

$\sigma_{s,1}^2$

$$\sigma_{s,l}^2 = \frac{1}{M} \sum_{j=1}^M y_k^2, \quad l = 1, \dots, L \quad (5.4.1)$$

wobei $k = M \cdot (l-1) + j$

Der Wert $\sigma_{s,l}^2$ kann als Schätzwert für die Varianz σ_m^2 des Spektralkoeffizienten y_m ($m = M \cdot (l-1) + \frac{M+1}{2}$) in der Klassenmitte der benachbarten M y_i -Werte betrachtet werden.

- 3) Die Stützwerte $\sigma_{s,l}^2$ werden quantisiert ($\sigma_{s,l}^2 \rightarrow \hat{\sigma}_{s,l}^2$) und zum Decodierer übertragen.
- 4) Die Schätzwerte für die Varianzen der Spektralkoeffizienten in den L Klassenmitten sind nun bekannt. Codierer und Decodierer berechnen sich anschliessend aus den quantisierten Werten $\hat{\sigma}_{s,l}^2$ die Schätzwerte für die Varianzen der restlichen $N - L$ Spektralkoeffizienten durch eine lineare Interpolation der logarithmierten Varianzen. Diese Schätzwerte werden mit $\hat{\sigma}_i^2$ ($i = 1, \dots, N$) bezeichnet.
- 5) Im Codierer und Decodierer wird die optimale Bitzuordnung R aus den Varianzen $\hat{\sigma}_i^2$ berechnet. Der Verstärkungsgrad α_i des Verstärkers vor dem Quantisierer Q_i ist sofort durch

$$\alpha_i = \hat{\sigma}_i^{-1} \quad (5.4.2)$$

gegeben. Im Decodierer wird der Verstärker für die Spektrallinie Nr. i auf den Verstärkungsgrad

$$\alpha_i^{-1} = \hat{\sigma}_i \quad \forall i$$

eingestellt.

- 6) Der transformierte Vektor y wird quantisiert und zum Empfänger übertragen.

Dieses Verfahren ist nur möglich, wenn:

- a) \underline{R}_{xx} eine Toeplitz-Matrix ist (für einen AR-Prozess, der das Sprachsegment näherungsweise beschreibt, erfüllt) und
- b) die Transformation \underline{A} die Sequenz-Eigenschaft besitzt und die Basisvektoren von \underline{A} sequenz-sortiert sind (z.B. für die DCT erfüllt).

Unter diesen Bedingungen sind die Erwartungswerte der Varianzen benachbarter Spektralkoeffizienten etwa gleich gross, d. h. das Basisspektrum weist einen glatten Verlauf ohne Sprünge auf (vgl. dazu Bild 2.5a, Kurve 3 für die DCT). Die Gewinnung der Zusatzinformation über das Basisspektrum ($\sigma_{s,l}^2$; $l = 1, \dots, L$) entspricht etwa dem Verfahren bei einem Kanalvocoder, wenn die dazugehörige Filterbank auch L Kanäle aufweist. In beiden Fällen wird die spektrale Information aus der Leistung in den L Kanälen bzw. L Klassen gewonnen. Da das geschilderte Codierverfahren sich an jede beliebige Signalstatistik anpassen kann, wird es als volladaptive Blockquantisierung bezeichnet. Die einzelnen Quantisierer werden auch hier adaptiv angesteuert; daraus ergibt sich die Kurzbezeichnung VA-BQ-AQ.

Die Zahl der Rechenoperationen zur Gewinnung der spektralen Zusatzinformation ist sehr gering. Während bei dem Verfahren nach Abschnitt 5.4 zur Produktbildung $\underline{A} \cdot \underline{R}_{xx} \cdot \underline{A}^T$ in der Grössenordnung N^3 Multiplikationen nötig wären, sind es hier bei dem direkten Verfahren nur etwa N Multiplikationen (wenn man von der Interpolation absieht). Für eine Blocklänge $N = 128$ ist dies ein Verhältnis von 16384 : 1 !

Für die Realisierung des Systems ist es günstiger, die auf ihren Mittelwert normierten Stützwerte zu übertragen. Dadurch ist der mögliche Wertebereich stets gleich gross und der Quantisierer für die Stützwerte kann feste Grenzen aufweisen. Der absolute Wert des Pegels wird dann wie bisher separat übertragen (skalarer Verstärkungsfaktor V für alle Quantisierer).

Das komplette Codiersystem ist in Bild 5.6 angegeben. Die spektrale Zusatzinformation ist in dem quantisierten Vektor der Stützwerte $\hat{\sigma}_S$ enthalten. Auf der Decodiererseite werden wie beim Codierer aus $\hat{\sigma}_S$ die Bitzuordnung \underline{R} und die Aussteuerungen der Quantisierer $\underline{\alpha}$ berechnet. Damit ist eine richtige Decodierung (D/A-Wandlung) des digitalen Datenflusses für $V \cdot \hat{\underline{y}}$ möglich. Die Rücktransformation mit \underline{A}^{-1} führt dann auf den quantisierten Vektor $\hat{\underline{x}}$ im Zeitbereich.

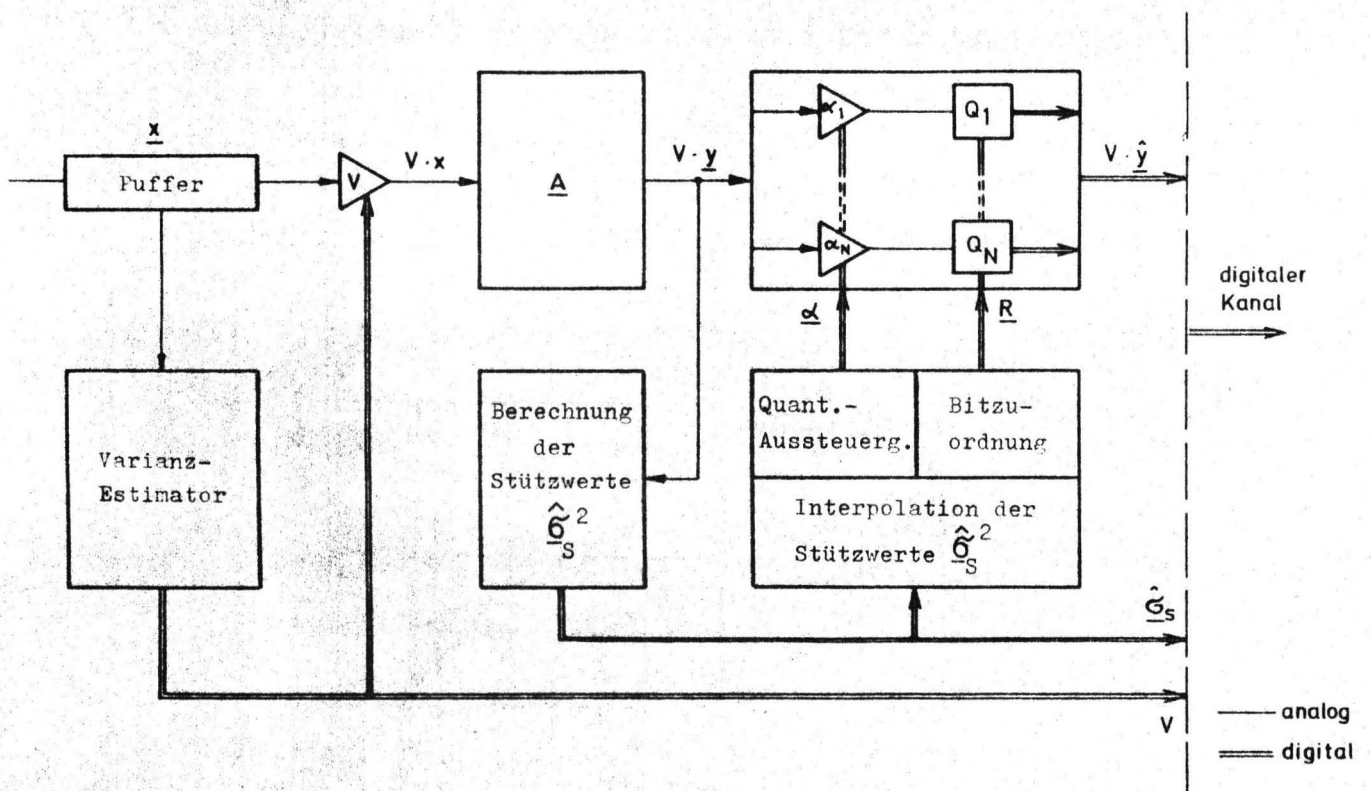


Bild 5.6 : Blockquantisierung mit volladaptiver Bitzuordnung und adaptiven Quantisierern.
(VA-BQ-AQ)

Ein solches volladaptives Codierverfahren ist bisher weder für die Sprach- noch für die Bildcodierung eingesetzt worden. Für die Bildcodierung ist die Verwendung dieses Verfahrens nur sinnvoll, wenn die Autokovarianzmatrix näherungsweise auf eine Toeplitz-Matrix zurückgeführt werden kann. Das bedeutet, die Zuordnung des laufenden Index zu den Bildpunkten eines Blocks müsste so gewählt werden, dass das sequenzsortierte Basisspektrum näherungsweise einen "glatten" Verlauf aufweist.

5.4.2 Ergebnisse

Als Gewinn durch die volladaptive Bitzuordnung wird der SNR-Anstieg gegenüber einer nichtadaptiven Bitzuordnung definiert:

$$G_{VA} = \text{SNR}_{VA-BQ-AQ} - \text{SNR}_{BQ-AQ} \quad (5.4.3)$$

In Bild 5.7 sind die Ergebnisse für die DCT mit der Blocklän-

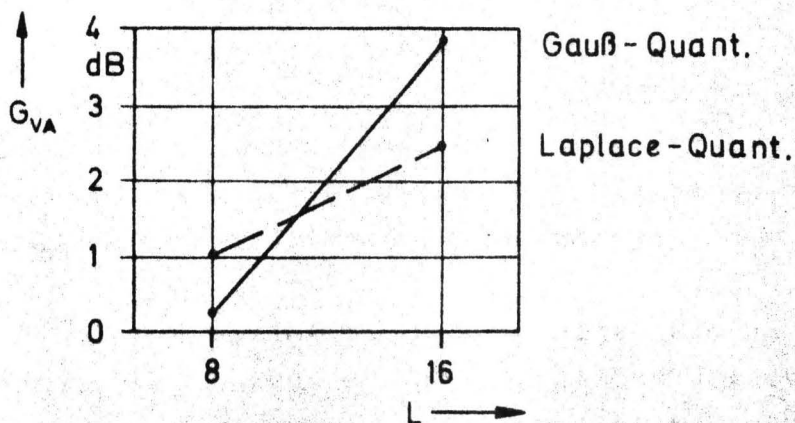


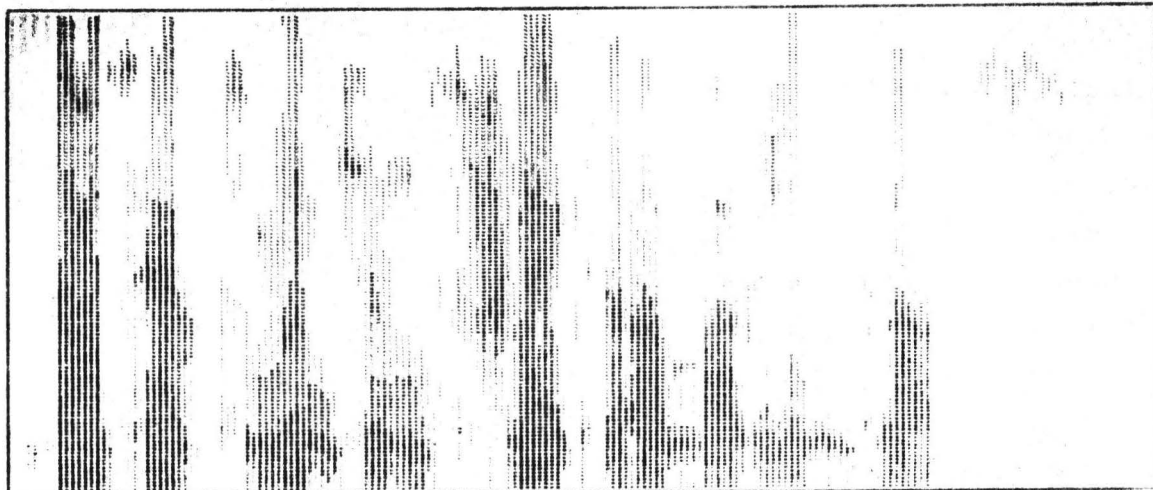
Bild 5.7 : Gewinn G_{VA} durch volladaptive Bitzuordnung (VA-BQ-AQ) in Abhängigkeit von der Zahl der spektralen Stützwerte L.

(DCT , N=128 , R=32 kbit/s , Spr. 13-BP ,
T=1.6 s)

ge $N = 128$ angegeben. Bei einer Stützwertzahl von $L=16$ und einem Gauss-Quantisierer ergibt sich ein Gewinn von $G_{VA} = 4$ dB. Eine weitere Erhöhung von L wird sicher den Gewinn noch vergrössern; der Gewinnzuwachs wird aber durch die vergrösserte Zusatzinformation wieder kompensiert. Der Gesamtgewinn gegenüber einer Codierung mit PCM-LOG beträgt damit 17 dB. Die erreichbaren SNR-Werte mit dem Verfahren VA-BQ-AQ werden noch in Abschnitt 6.1 für verschiedene Bitraten und verschiedene Sprecher ausführlich diskutiert.

Die Wirkungsweise der volladaptiven Blockquantisierung soll noch einmal genauer erläutert werden. In Bild 5.8 ist in der oberen Zeichnung ein Visible-Speech-Diagramm für den Satz "das Herz Münchens schlägt am Marienplatz" (Zeit $t = 0 \dots 3$ s; Frequenzachse $0 \dots 4$ kHz) dargestellt. Darunter ist im gleichen Zeitmaßstab die Folge der Bitzuordnungen bei der Codierung mit VA-BQ-AQ (DCT, $N = 128$, $R = 16$ kbit/s) durch Schwärzungsgrade angegeben. Eine zunehmende Schwärzung bedeutet eine zunehmende Bitzahl; weisse Stellen markieren die nicht übertragenen Spektralkoeffizienten. Die Ordinate entspricht dabei einerseits der Ordnungszahl der Spektralkoeffizienten ($1 \dots 128$), andererseits aber weiterhin der dazugehörigen Frequenz ($0 \dots 4$ kHz), da die DCT die Sequenzeigenschaft besitzt. Entsprechend der Blocklänge $N = 128$ wird alle 16 ms ein neuer Vektor \underline{R} (R_1, \dots, R_{128}) für die Bitzuordnung berechnet, der im unteren Diagramm in vertikaler Richtung durch die Schwärzungsgrade gekennzeichnet wird.

Der Vergleich beider Diagramme zeigt, wie die volladaptive Blockquantisierung sich dem zeitlichen Verlauf des Kurzzeitspektrums der Sprache anpasst. Bei stimmhaften Lauten werden die stärksten Formantfrequenzen mit den grössten Bitzahlen quantisiert (dunkle Stellen im unteren Diagramm), die hohen Frequenzanteile werden überhaupt nicht mehr übertragen (weisse Flächen). In den Sprachpausen oder bei schwach korrelierten Lauten treten keine spektralen Maxima auf; in vertikaler Richtung ist dann im unteren Diagramm eine etwa konstante



" Das Herz Münchens schlägt am Marien...platz... "

Bild 5.8a : Visible-Speech-Diagramm.

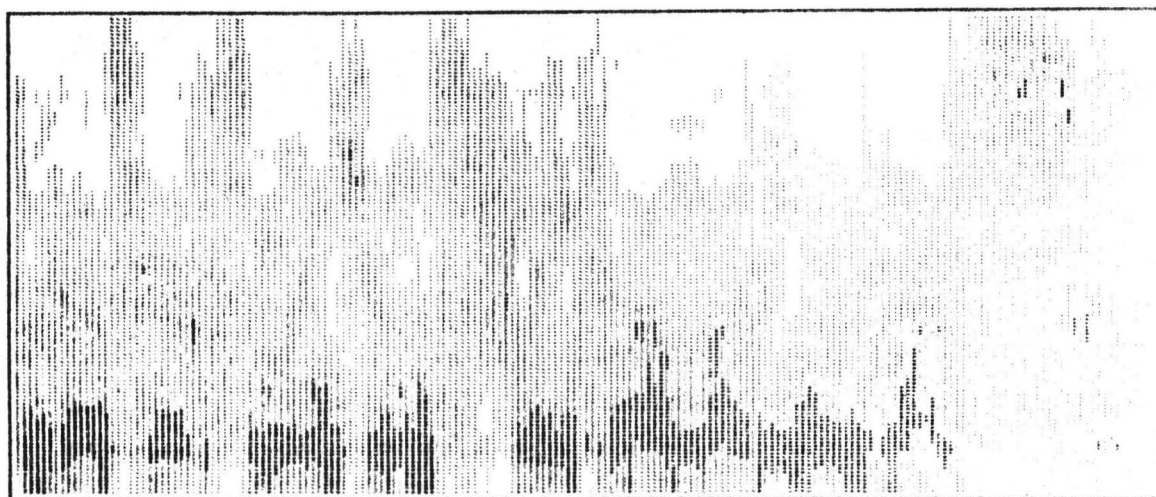


Bild 5.8b : Bitzuordnung bei VA-BQ-AQ mit $R=16$ kbit/s.
(DCT , $N=128$)

Bild 5.8 : Wirkungsweise der volladaptiven Blockquantisierung.
(beide Diagramme im gleichen Zeitmaßstab:
 $t = 0 \dots 3$ s)

Graustufe festzustellen, die aussagt, dass alle Spektralkoeffizienten mit gleicher Bitzahl quantisiert werden. Die volladaptive Blockquantisierung entspricht also in ihrer Wirkungsweise einem Filter, das (hier) alle 16 ms seinen Durchlassbereich verändern kann. Je nach dem vorliegenden Kurzzeitspektrum werden die energiestarken Frequenzbereiche bevorzugt (durch hohe Bitzahlen für die Spektralkoeffizienten) und die energieschwachen Frequenzbereiche z.T. vollständig unterdrückt. Bild 5.8 macht deutlich, dass die volladaptive Blockquantisierung eine Redundanzreduktion im Frequenzbereich dadurch bewirkt, dass zu jedem Zeitpunkt nur noch die relevanten Frequenzanteile übertragen bzw. besonders fein quantisiert werden.

5.4.3 ADV der Spektralkoeffizienten

Durch die volladaptive Bitzuordnung und die damit verbundene individuelle Aussteuerung der einzelnen Quantisierer ändert sich wiederum die ADV der Quantisierereingangsfolgen. Wie im Abschnitt 3.3 wurde die mittlere ADV $p_m(y)$ bestimmt, die in Bild 5.9 für die DCT mit der Blocklänge $N = 128$ (bei $L=16$ Stützwerten) dargestellt ist. Die durchgezogene Kurve im Bild ist die Gauss-ADV. Der Vergleich zeigt, dass die gemessene ADV sehr gut durch die Gauss-ADV beschrieben werden kann. Die Aussteuerung der einzelnen Quantisierer entsprechend dem vorliegenden Kurzzeit-LDS bewirkt, dass die Quantisierer kaum noch über- oder untersteuert werden können. Dadurch werden die extremen Amplitudenschwankungen am Quantisierereingang abgebaut, so dass nicht mehr die Laplace-ADV (bei BQ-AQ) oder sogar die Gamma-ADV (bei BQ-LOG) auftreten kann. Die ADV-Messung steht im Einklang mit dem Simulationsergebnis aus Bild 5.7, bei dem mit einem Gauss-Quantisierer das höchste SNR erzielt wurde.

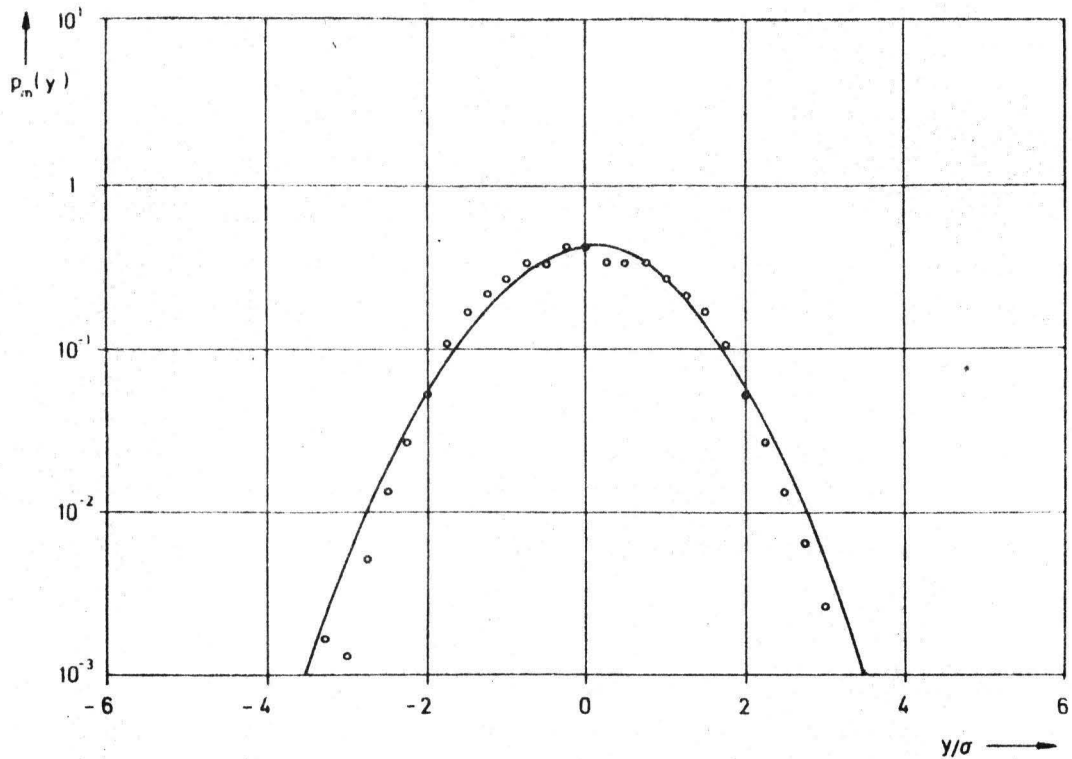


Bild 5.9 : Mittlere ADV der Spektralkoeffizienten $p_m(y)$ (Meßpunkte) und Vergleich mit einer Gauß-ADV. (VA-BQ-AQ: DCT, N=128, L=16 Stützstellen)

Aus früheren Untersuchungen /18,29/ ist bekannt, dass der Ersatz des nichtlinearen Quantisierers für kleinsten MQF durch einen linearen Quantisierer mit Entropiecodierung für Gaussquellen die Übertragungsrate nur um höchstens 0.468 bit/Abtastwert reduzieren kann. In Anbetracht des zusätzlichen Aufwandes (Pufferspeicher und Codeumwandlung) verbunden mit Laufzeitverzögerungen ist eine Entropiecodierung bei der volladaptiven Blockquantisierung im allgemeinen kaum lohnend.

5.4.4 Quantisierung der Zusatzinformation über das Kurzzeitspektrum

Für die volladaptive Blockquantisierung sind die L spektralen Stützwerte $\sigma_{s,1}^2$ zusätzlich zum Decodierer zu übertragen. Um diese Zusatzinformation so gering wie möglich zu halten, wird untersucht, wie grob die Stützwerte quantisiert werden dürfen, ohne dass das SNR merklich absinkt.

Die Stützstellen $\sigma_{s,1}^2$ seien nicht die (pegelabhängigen) Absolutwerte, sondern bereits die auf ihren Mittelwert normierten Größen. Dann kann ein Quantisierungsbereich mit festen Grenzen für $\sigma_{s,1}^2$ gewählt werden. Aus den verschiedenen Untersuchungen an Modellquellen (AR-Prozesse 10. Ordnung, Abschnitt 5.1.1) wurden die minimal und maximal auftretenden Varianzen der Spektralkoeffizienten entnommen und damit der Quantisierungsbereich für alle Stützwerte zu

$$0.005 \leq \sigma_{s,1} \leq 5.0 ; \quad l = 1, \dots, L \quad (5.4.4)$$

festgelegt.

Dieser Bereich wird logarithmisch quantisiert, so dass der maximale relative Quantisierungsfehler unabhängig von der Grösse von $\sigma_{s,1}$ ist. In Bild 5.10 sind die Ergebnisse angegeben, wenn jeder Stützwert mit n_s bit quantisiert wird.

Es ist ersichtlich, dass der SNR-Rückgang gegenüber $n_s \rightarrow \infty$ kleiner als 1 dB ist, wenn $n_s = 3$ gewählt wird. Der Wert n_s kann noch weiter reduziert werden, denn in Gl. 5.4.4 wurde der ungünstigste Fall für den Wertebereich von $\sigma_{s,1}$ zugrundegelegt. Dies ist sofort aus Bild 5.11 ersichtlich, das die Histogramme für die Stützwerte $\sigma_{s,1}$ zeigt. Für $l = 1, \dots, 10$ wird etwa nur der halbe Quantisierungsbereich ausgenutzt, so dass 1 bit eingespart werden kann (bei gleichem maximalen Quantisierungsfehler). Für $l > 10$ könnte eben-

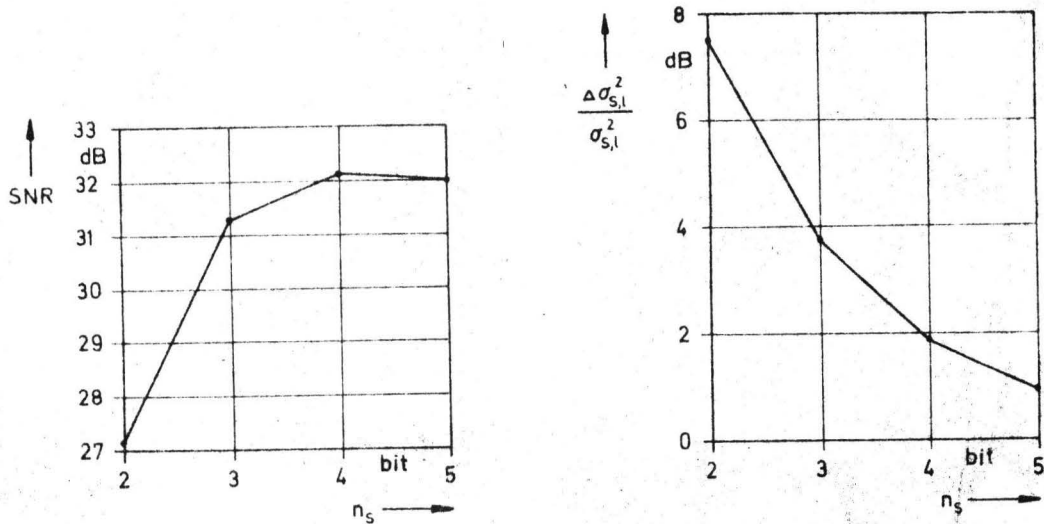


Bild 5.10 : SNR und max. Quantisierungsfehler $\frac{\Delta\sigma_{s,1}^2}{\sigma_{s,1}^2}$ ohne Übersteuerung in Abhängigkeit von der Bitzahl n_s , mit der jeder Stützwert $\sigma_{s,1}^2$ quantisiert wird.

(VA-BQ-AQ: DCT, N=128, L=16, Gauß-Quant.,
R=32 kbit/s, Spr. 13-BP, T=1.6 s)

falls 1 bit eingespart werden, da bei diesen Stützwerten wegen ihrer kleinen Varianz sich ein grösserer Quantisierungsfehler auf das Gesamt-SNR nur wenig auswirkt.

Bei etwa unverändertem SNR sind damit $n_s = 2$ bit pro Stützwert ausreichend. Die Datenrate R_s kennzeichne den Informationsfluss für die Kurzzeitspektren. Bei einer Blocklänge $N = 128$, Abtastfrequenz $f_A = 8$ kHz, Stützstellenzahl $L = 16$ und $n_s = 2$ ergibt sich

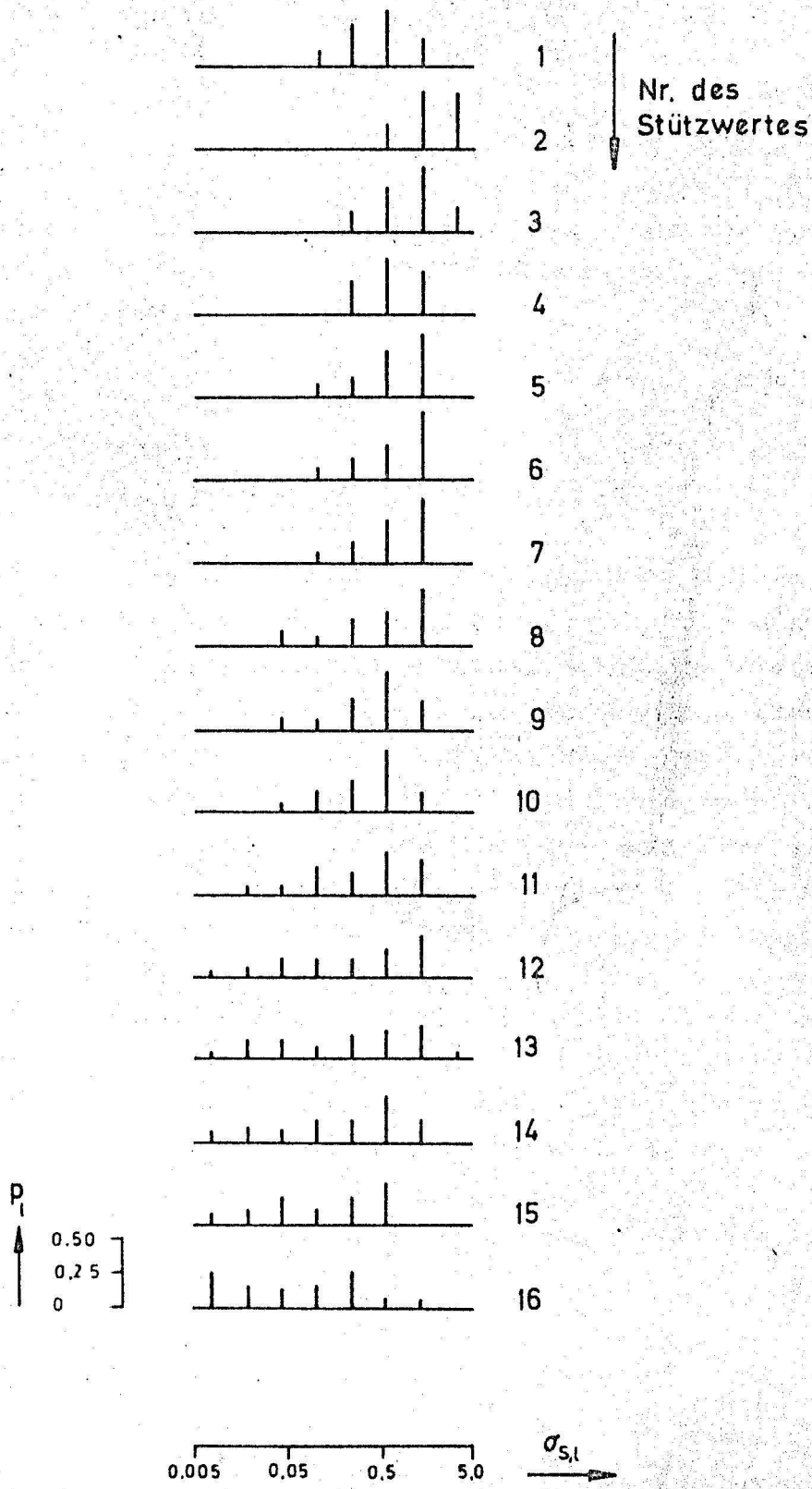


Bild 5.11 : Auftretenswahrscheinlichkeiten der normierten Stützwerte $\sigma_{s,1}$.
(weitere Parameter wie in Bild 5.10)

$$R_s = 2 \text{ kbit/s} \quad (5.4.5)$$

Diese Datenrate könnte noch weiter reduziert werden, wenn die Korrelationen innerhalb der Vektorfolge $\underline{\sigma}_s^2$ ausgenutzt werden.

5.4.5 Algorithmus für kleine Blocklängen

Um den Aufwand bei der Realisierung der volladaptiven Blockquantisierung zu verringern, kann die Blocklänge N verkleinert werden. Dazu muss ein anderer Algorithmus zur Gewinnung der spektralen Zusatzinformation eingesetzt werden, denn die Zahl der Stützstellen L würde sonst dieselbe Größenordnung wie die Blocklänge einnehmen und damit die Zusatzinformation für den Decodierer unzulässig hohe Werte annehmen. Für kleinere Blocklängen soll nun nicht die "örtliche Korrelation" (benachbarte Spektralkoeffizienten des gleichen Blocks haben etwa gleiche Varianz), sondern die "zeitliche Korrelation" (Spektralkoeffizienten gleicher Ordnungszahl in benachbarten Blöcken haben etwa gleiche Varianz) ausgenutzt werden, um die spektrale Zusatzinformation zu gewinnen. Dazu wird das Signal in Segmente der Länge N_{SEG} unterteilt, wobei ein Segment aus $M = N_{SEG}/N$ Blöcken besteht (z.B. $N_{SEG} = 128$ und $N = 16$). Jeder Block (die Blocknummer sei j) wird für sich transformiert:

$$\underline{y}(j) = \underline{A} \cdot \underline{x}(j) \quad ; \quad j = 1, \dots, M \quad (5.4.6)$$

Die Schätzwerte für die Varianzen der Spektralkoeffizienten $\hat{\sigma}_i^2$ werden aus

$$\hat{\sigma}_i^2 = \frac{1}{M} \sum_{j=1}^M y_{i,j}^2 \quad ; \quad i = 1, \dots, N \quad (5.4.7)$$

gebildet.

Die Schätzwerte $\hat{\sigma}_i^2$ werden nun quantisiert und zum Decodierer übertragen. Codierer und Decodierer berechnen sich nun wie in Abschnitt 5.4.1 die Bitzuordnung und die einzelnen Quantisiereraussteuerungen, die für alle Blöcke eines Segments unverändert bleiben. Anschliessend wird die Übertragung der M Blöcke $\underline{y}(j)$ ($j=1, \dots, M$) durchgeführt, dann wird das nächste Segment analysiert und in gleicher Weise codiert usw. Dieser Algorithmus bildet gegenüber dem Verfahren aus Abschnitt 5.4.1 (für grosse Blocklängen) folgende Vorteile:

- * Bei kleineren Blocklängen verringert sich die Zahl der Rechenoperationen pro Abtastwert zur Durchführung der Transformation (vgl. Abschnitt 2.3.4).
- * Es ist keine Interpolation der spektralen Schätzwerte erforderlich, da die Zahl der Schätzwerte gleich der Blocklänge ist.
- * Die Bitzuordnung und die Quantisiereraussteuerungen werden nur für eine kleine Vektorlänge (Blocklänge N) berechnet; die Berechnung ist aber nicht für jeden Block, sondern nur für jedes Segment zu wiederholen. D.h., die Zahl der Rechenoperationen pro Abtastwert zur Verarbeitung der spektralen Zusatzinformation ist erheblich kleiner geworden.

Als Nachteil ist ein kleineres SNR zu erwarten, da der Gewinn mit Verkleinerung der Blocklänge zurückgeht.

Bild 5.12 zeigt die Simulationsergebnisse für Blocklängen $N = 8$ und $N = 16$ bei einer Segmentlänge $N_{SEG} = 128$. Zum Vergleich ist der SNR-Wert bei einer DCT für eine volladaptive Blockquantisierung mit der Blocklänge $N = 128$ eingetragen. Die DCT mit $N = 16$ weist gegenüber diesem Wert ein um weniger als 2 dB verringertes SNR auf, während für die anderen Fälle der SNR-Rückgang erheblich grösser ist.

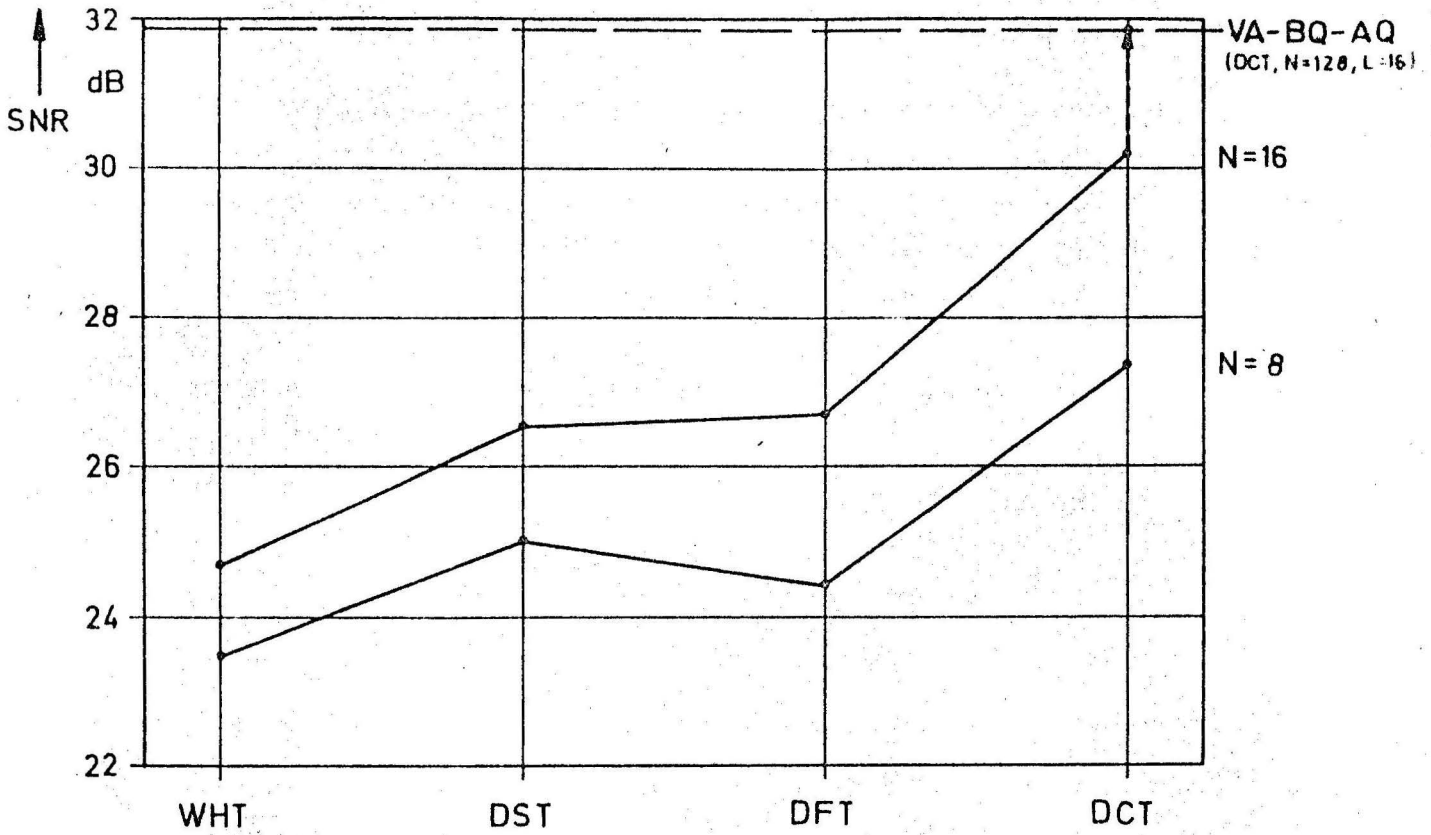


Bild 5.12 : SNR-Werte mit VA-BQ-AQ für kleine Blocklängen N bei einer Übertragungsrate $R=32$ kbit/s .
(NSEG=128, Gauß-Quant., Spr. 13-BP, T=1.6 s)

6. Codierung von Sprachsignalen mit nichtadaptiver und adaptiver Blockquantisierung und Vergleich mit anderen Verfahren

In den bisherigen Untersuchungen wurden nur Simulationsergebnisse für eine Sprachprobe verwendet (Spr. 13-BP). Es sollen nun weitere Sprachproben mit TP- und BP-Filterung hinzugenommen werden und auch ein Vergleich mit einem anderen Codierverfahren durchgeführt werden.

Für die Blockquantisierung werden folgende Codiersysteme betrachtet:

BQ - LOG : DCT, N = 128

BQ-AQ : DCT, N = 128, Laplace-Quant.

BA-BQ-AQ : DCT, N = 128, 8 Klassen, Laplace-Quant.

VA-BQ-AQ : DCT, N = 128, 16 Stützwerte, Gauss-Quant.

(Diese Parametersätze gelten auch für die Betrachtungen im Abschnitt 7 und 8).

Andere Codierverfahren:

1. PCM-LOG:

Die PCM-Codierung mit logarithmischem Quantisierer wird als Bezugscodierung eingesetzt. Aus dem SNR-Wert bei PCM-LOG lassen sich die Gewinne im Signal-Rausch-Abstand für andere Codierverfahren bestimmen.

2. ADPCM-AQ:

ADPCM-AQ ist die Bezeichnung für eine DPCM mit adaptiver Prädiktion und adaptivem Quantisierer /1,2,30/. Die ADPCM-AQ bewirkt eine Redundanzreduktion im Zeitbereich. Sie ist das Äquivalent zur Codierung im Frequenzbereich

mit VA-BQ-AQ und passt sich in entsprechender Weise an das Kurzzeitspektrum des Sprachsignals (durch adaptive Prädiktion) und an den momentanen Sprachpegel (durch einen adaptiven Quantisierer) an. Auch bei der Codierung mit ADPCM-AQ werden im folgenden nur Systeme mit Vorwärtssteuerung betrachtet, d. h., aus je NSEG nicht codierten Abtastwerten wird die Information über den Prädiktor und die Quantisiereraussteuerung gewonnen. Diese Information wird dann zum Decodierer übertragen und anschliessend die Codierung der NSEG Abtastwerte durchgeführt.

Es werden für die ADPCM-AQ folgende Parameter gewählt:
Pufferlänge NSEG = 128; Prädiktor mit 10 Koeffizienten;
Laplace - Quantisierer.

Die Analysesegmentlängen für die beiden volladaptiven Codierverfahren im Zeitbereich (ADPCM-AQ) und im Frequenzbereich (VA-BQ-AQ) sind also gleich gross. Der Informationsfluss R_s für die Zusatzinformation (Kurzzeit-LDS) bei einem SNR-Rückgang von weniger als 1 dB ist für

$$\begin{aligned} \text{ADPCM-AQ} : R_s &\cong 3 \text{ kbit/s} && /1/ \\ \text{VA-BQ-AQ} : R_s &\cong 2 \text{ kbit/s} \end{aligned}$$

und bewegt sich damit auch in der gleichen Grössenordnung, so dass ein Vergleich des erreichbaren Signal-Rausch-Abstandes dieser beiden Verfahren sinnvoll ist.

6.1 Signal-Rausch-Abstand bei verschiedenen Übertragungsraten

In Bild 6.1 a...c sind die erreichbaren SNR-Werte für verschiedene Sprecher und Übertragungsraten von $R = 16 \dots 32$ kbit/s gezeigt. Der Signal-Rausch-Abstand nimmt generell von a nach c zu. Dies hat folgende Ursachen:

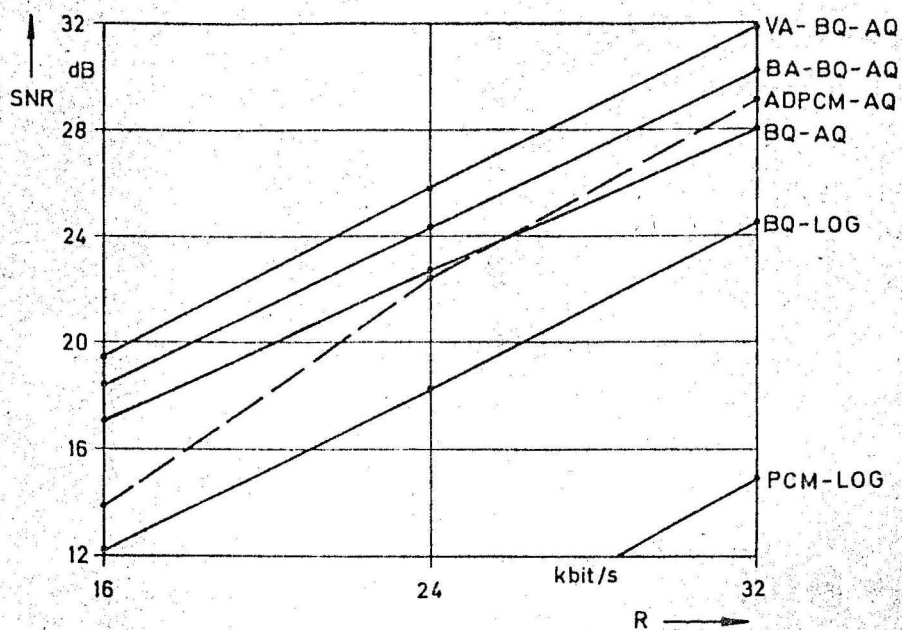


Bild 6.1a :
Spr. 13-BP,
T=1.6 s

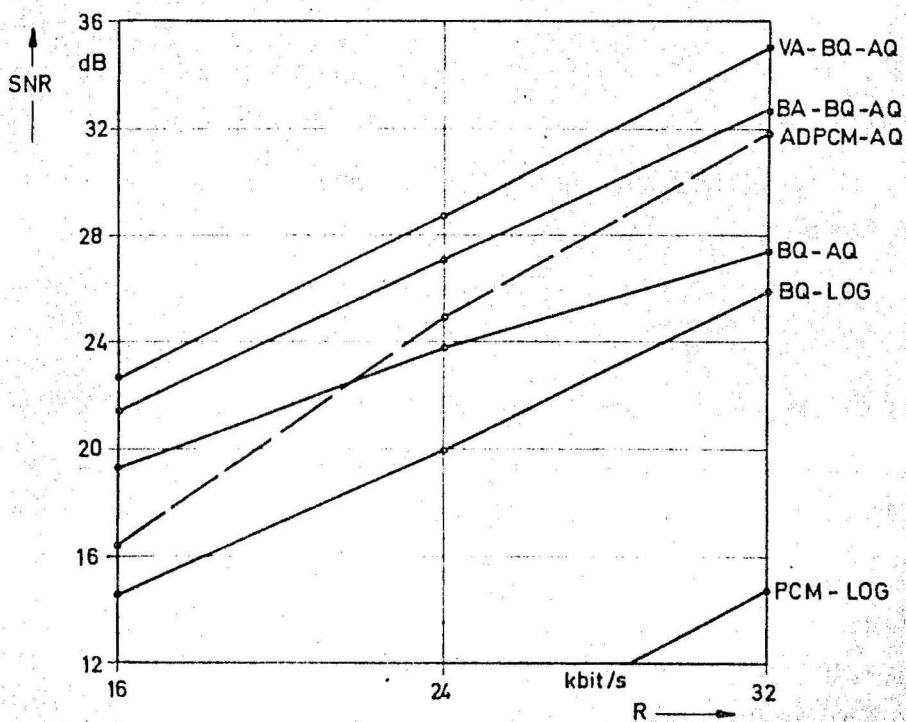


Bild 6.1b :
Spr. 40-BP,
T=3.0 s

Bild 6.1 : Vergleich des erreichbaren Signal-Rausch-Abstandes bei verschiedenen Codierverfahren und verschiedenen Sprechern.

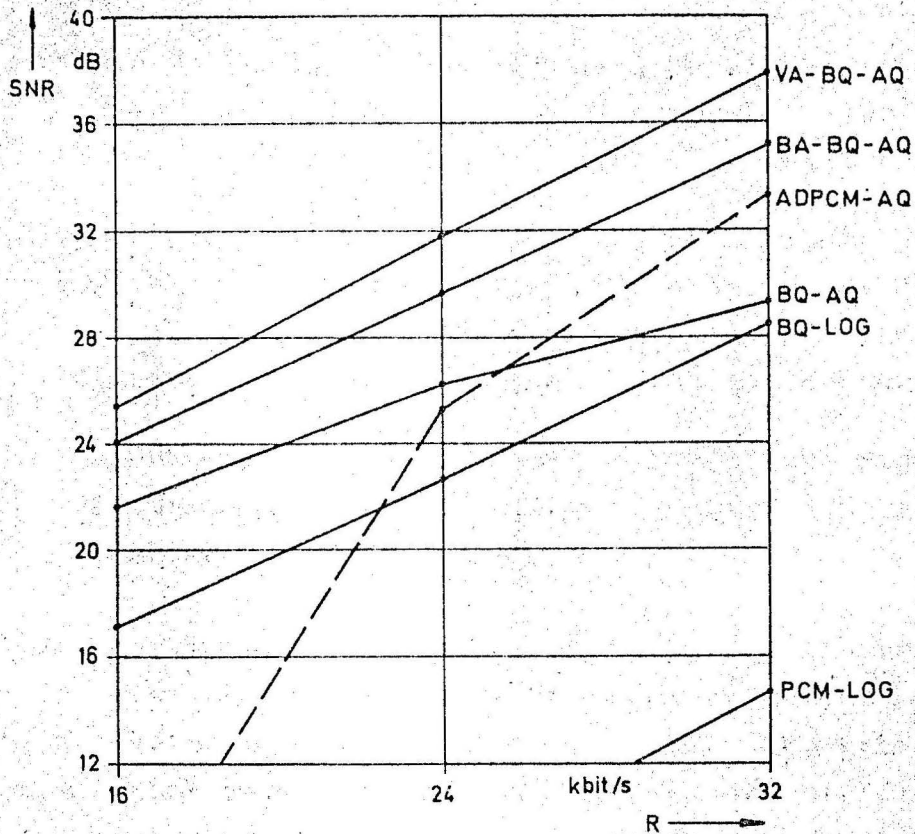


Bild 6.1c :
Spr. 40-TP,
T=3.0 s

- I) TP-gefilterte Sprache enthält noch Frequenzanteile unter 300 Hz, die einen grösseren Gewinn ermöglichen /1,31/.
- II) Spr. 40 hat eine weniger abfallende Autokorrelationsfunktion (Langzeit-Messung) als Spr. 13. Dadurch wird der mittlere Gewinn vergrössert.

Aus Bild 6.1 können folgende Ergebnisse abgelesen werden:

- * Mit der volladaptiven Blockquantisierung kann das SNR gegenüber PCM-LOG um 17...23 dB vergrössert werden.
- * Die volladaptive Blockquantisierung bringt gegenüber der nichtadaptiven BQ-LOG ein Gewinn von mindestens 8...10 dB. (Die SNR-Werte für BQ-LOG und auch für BQ-AQ und BA-BQ-AQ sind nur als obere Schranken anzusehen, da eine optimale Anpassung an die Signalstatistik vorausgesetzt wurde).

- * Die VA-BQ-AQ ist der ADPCM-AQ um mindestens 3 dB überlegen. Der SNR-Unterschied vergrößert sich noch bei Verringerung der Übertragungsrate oder bei stärker korrelierten Sprachsignalen.
- * Die VA-BQ-AQ erreicht auch bei einer Übertragungsrate von $R = 16$ kbit/s noch ein SNR von etwa 20dB und darüber, während die ADPCM-AQ etwa um 6 dB schlechter arbeitet.
- * Für ungünstige Fälle (stark korreliertes Signal und kleine Bitrate; im Bild 6.1c für $R = 16$ kbit/s) versagt die ADPCM-AQ vollkommen.

Der starke SNR-Rückgang der ADPCM-AQ bei kleineren Übertragungsraten ist auf den Einfluss der Fehlerrückkopplung des Quantisierungsfehlers zurückzuführen /1,31/. Der Prädiktor arbeitet dann nicht mehr optimal; die Varianz des Differenzsignals vergrößert sich, so dass der Quantisierer übersteuert wird. Die Fehlerrückkopplung wirkt sich umso mehr aus, je stärker das Signal korreliert ist (s.o.). Auch für die adaptive Blockquantisierung geht der Gewinn für kleine Übertragungsraten durch eine nicht mehr optimale Bitzuordnung zurück (vgl. Abschnitt 5.1.2). Dieser Einfluss ist aber erheblich geringer als bei der ADPCM-AQ und bei den untersuchten Übertragungsraten kaum feststellbar. Dies ist daran zu erkennen, dass der Verlauf $SNR(R)$ in Bild 6.1 für die VA-BQ-AQ sehr gut durch eine Gerade beschrieben werden kann, deren Steigung 6 dB pro bit bzw. 6 dB pro 8kbit/s beträgt.

Auch für grössere Übertragungsraten ist die Codierung mit VA-BQ-AQ der ADPCM-AQ um etwa 3 dB überlegen. Dies ist im wesentlichen darauf zurückzuführen, dass in beiden Fällen ein Quantisierer ohne Entropiecodierung eingesetzt wird; die ADV am Quantisierereingang aber bei VA-BQ-AQ etwa gaussverteilt (vgl. Abschnitt 5.4.3), bei ADPCM-AQ dagegen näherungsweise laplaceverteilt ist /7/. Bei einer Gauss-ADV kann mit einem Quantisierer ohne Entropiecodierung ein grösserer Signal-Rausch-Abstand erreicht werden als bei einer Laplace-ADV /18/.

Die Signal-Rausch-Abstände auch bei sehr kleinen Übertragungsraten bis 4 kbit/s sind für die VA-BQ-AQ in Bild 6.2 angegeben. Der Gewinnrückgang ist in diesem Bereich gut erkennbar; trotzdem ist bei einer Übertragungsrate von 4 kbit/s noch ein SNR von etwa 10 dB erreichbar.

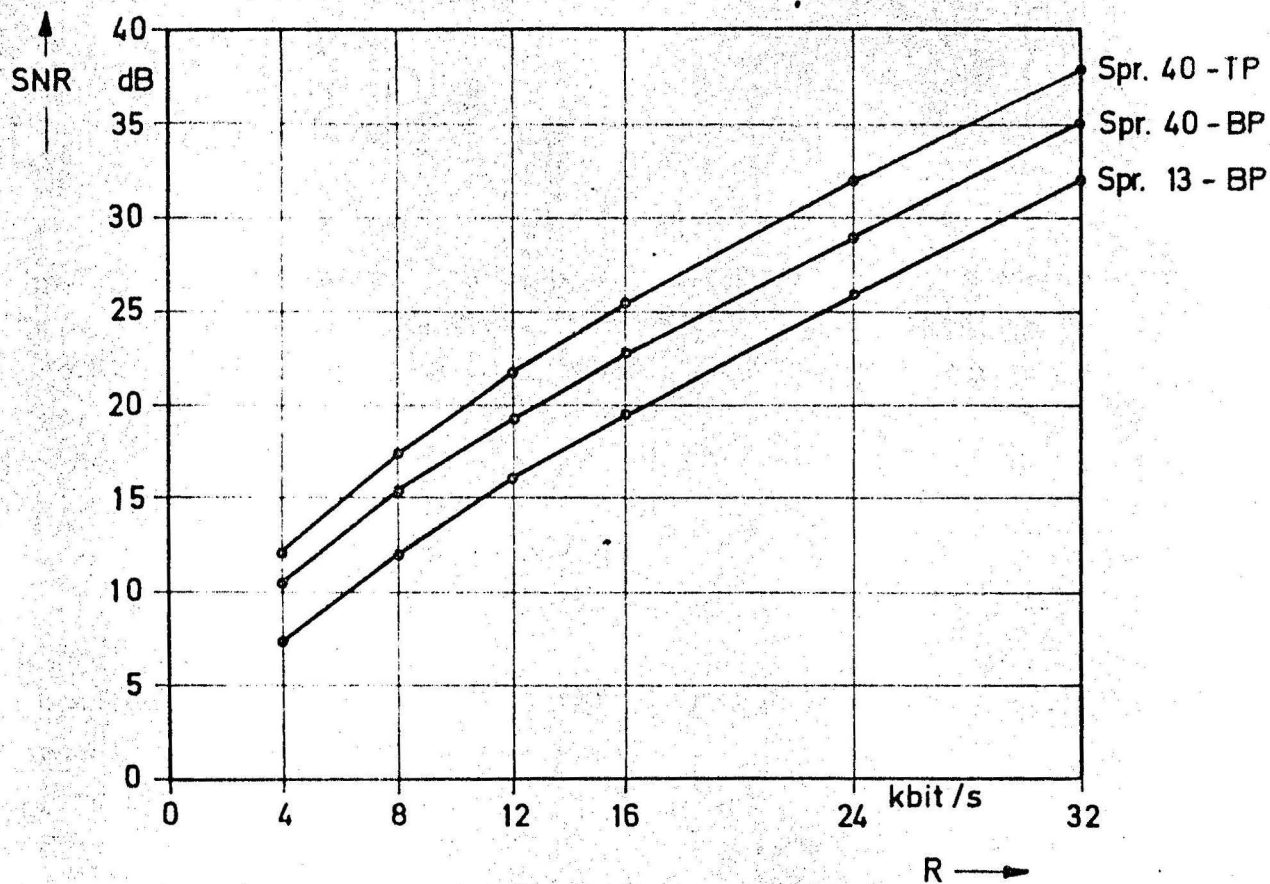


Bild 6.2 : Signal-Rausch-Abstand für die VA-BQ-AQ bei verschiedenen Sprechern.

6.2 Zeitlicher Verlauf des Segment-SNR

Die Angabe des Signal-Rausch-Abstandes kennzeichnet nur ungenügend die Qualität einzelner Codierverfahren. Ein wesentlich besserer Vergleich ist möglich, wenn der zeitliche Verlauf des Signal-Rausch-Abstandes für aufeinanderfolgende Datensegmente angegeben wird. Jedes Datensegment besteht aus $N = 128$ Abtastwerten /7/. Der Signal-Rausch-Abstand in dem k -ten Segment ist definiert als

$$\text{SNR}_S(k) = 10 \lg \left(\frac{\sum_{j=1}^N x_j^2}{\sum_{j=1}^N q_j^2} \right) \quad (6.2.1)$$

wobei q_j^2 die Quantisierungsverzerrungen im Zeitbereich bei der Codierung der Abtastwerte x_j sind. Zusätzlich soll der Verlauf des Segment-Pegels $P_S(k)$ dargestellt werden, der durch

$$P_S(k) = 10 \lg \left(\frac{\sum_{j=1}^N x_j^2}{N \cdot \sigma_{LZ}^2} \right) \quad (6.2.2)$$

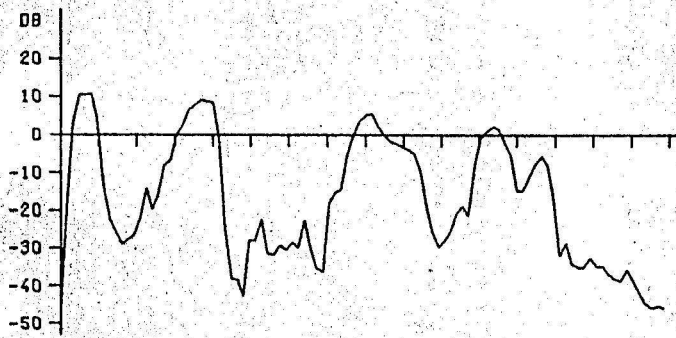
gegeben ist. Der Wert σ_{LZ}^2 ist die Langzeit-Varianz der zu codierenden Sprachprobe.

In den Bildern 6.3 und 6.4 sind die Verläufe von Segment-SNR und Segment-Pegel für verschiedene Codierverfahren bei der Übertragungsrate $R = 24$ kbit/s dargestellt. Die Sprachprobe ist 1.6 s lang und umfasst die Worte " Das Herz Münchens ".

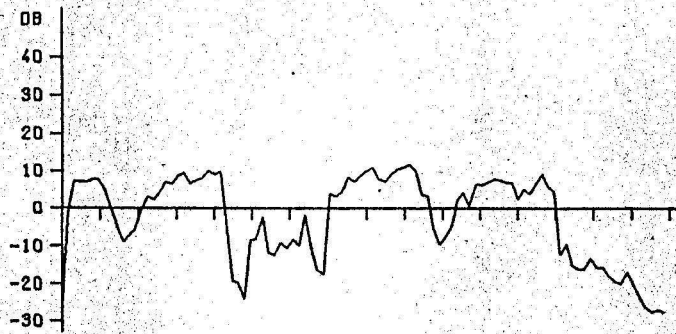
Diskussion der SNR-Verläufe:

PCM-LOG:

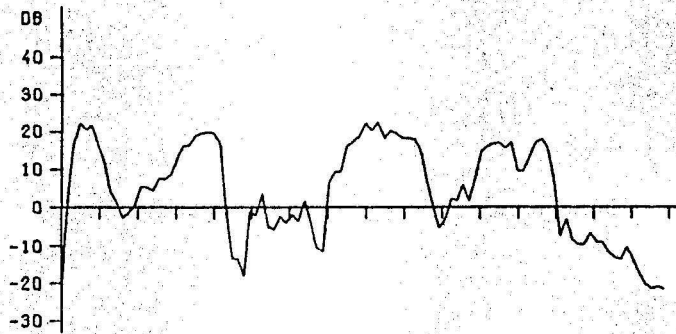
Der Verlauf von Pegel und Segment-SNR ist eng miteinander korreliert. Bei niedrigpegeligen Segmenten geht das SNR



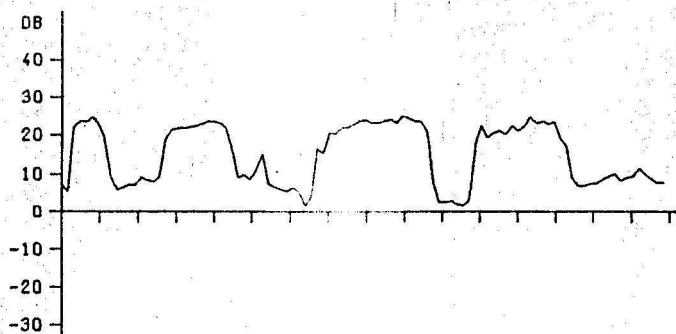
Pegel-Verlauf



SNR-Verlauf
bei PCM-LOG



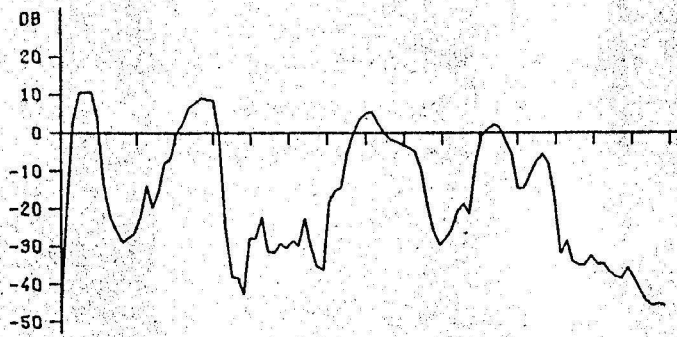
SNR-Verlauf
bei BQ-LOG



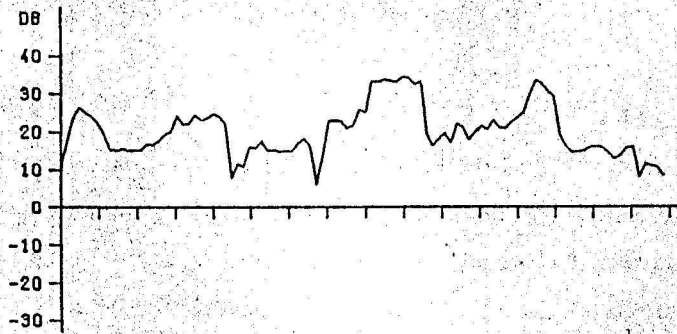
SNR-Verlauf
bei BQ-AQ

0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.8 SEC

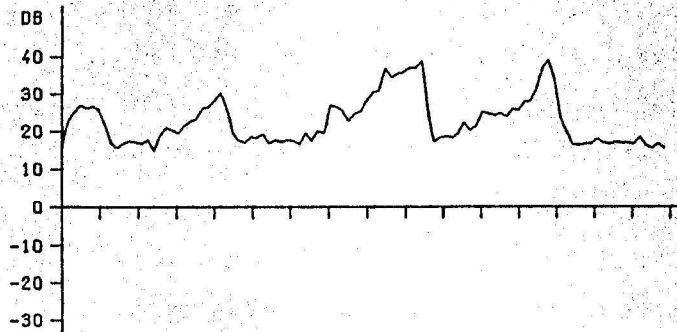
Bild 6.3 : Verlauf von Sprachpegel und Segment-SNR für verschiedene Codierverfahren.
($R=24$ kbit/s)



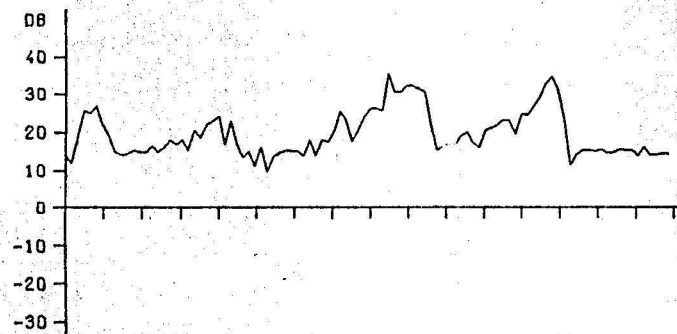
Pegel-Verlauf



SNR-Verlauf
bei BA-BQ-AQ



SNR-Verlauf
bei VA-BQ-AQ



SNR-Verlauf
bei ADPCM-AQ

0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 1.6 SEC

Bild 6.4 : Verlauf von Sprachpegel und Segment-SNR für verschiedene Codierverfahren.
(R=24 kbit/s)

sehr stark in den negativen Bereich, weil der logarithmische Quantisierer untersteuert wird. Daher ist ein grosses Kanal-Ruhegeräusch zu erwarten.

BQ-LOG:

Das SNR wird im Mittel etwa um 10 dB angehoben. Trotzdem treten auch hier negative Signal-Rausch-Abstände auf, wenn Sprachsegmente mit sehr kleinem Pegel vorliegen.

BQ-AQ:

Die maximalen SNR-Werte werden gegenüber BQ-LOG um etwa 4 dB angehoben. Es treten nun keine negativen Signal-Rausch-Abstände mehr auf, so dass das Kanal-Ruhegeräusch erheblich vermindert ist. Durch die feste Bitzuordnung arbeitet das System aber nicht für alle Laute optimal. Besonders bei schwach korrelierten Signalsegmenten geht das SNR fast auf 0 dB zurück (negativer Gewinn), während dagegen bei einer PCM-Codierung mit gesteuertem 3bit-Quantisierer das Segment-SNR stets grösser als 10 dB ist /7/.

BA-BQ-AQ:

Durch die Möglichkeit einer beschränkten Anpassung an die Signalstatistik tritt eine erhebliche Verbesserung ein. Die maximalen SNR-Werte werden gegenüber BQ-AQ um etwa 10 dB angehoben und das SNR geht kaum unter 10 dB herunter. Obwohl das mittlere SNR nur um etwa 2 dB grösser geworden ist, ist subjektiv eine erheblich stärkere Verbesserung zu erwarten.

VA-BQ-AQ:

Dieses Codierverfahren arbeitet für alle Datenssegmente stets optimal. Daher geht das SNR nie unter 15 dB zurück und erreicht Maximalwerte bis 40 dB. Für die VA-BQ-AQ wird ein Gauss-Quantisierer eingesetzt. Ein 3bit-Quantisierer für eine unkorrelierte Gauss-Quelle weist ein SNR von 14.62 dB auf /18/. Daraus ist ersichtlich, dass die volladaptive Blockquantisierung stets einen positiven Gewinn bezüglich der spektralen

Ungleichverteilung ermöglicht.

ADPCM-AQ:

Der SNR-Verlauf ist weitgehend dem SNR-Verlauf bei VA-BQ-AQ ähnlich. Es zeigt sich aber, dass nicht die maximalen SNR-Werte wie bei VA-BQ-AQ erreicht werden; für einige Segmente geht das SNR bis auf 10 dB herunter.

6.3 Subjektive Bewertung der Codierverfahren

Bei kleineren Übertragungsraten ist das Fehlersignal nicht mehr weiss; die Angabe des SNR-Wertes ist dann kein ausreichendes Mass mehr für die Qualität eines Codierverfahrens /32/. Daher wurden die untersuchten Codiersysteme auch durch eine subjektive Bewertung der codierten Sprachsignale überprüft. Die Ergebnisse lassen sich wie folgt zusammenfassen:

- * Systeme mit einem adaptiven Quantisierer (AQ) weisen ein sehr geringes Ruhegeräusch auf. Dies bedeutet zusätzlich zum objektiven SNR-Anstieg gegenüber einem nichtadaptiven Quantisierer eine erhebliche subjektive Verbesserung.
- * Das codierte Sprachsignal bei BQ-AQ und BQ-LOG klingt etwas dumpf, da hohe Frequenzanteile nicht oder ungenügend übertragen werden. Bei BA-BQ-AQ oder VA-BQ-AQ besteht dieser Nachteil nicht, da diese Codiersysteme sich an das jeweils vorliegende Kurzzeitspektrum anpassen können.
- * Die Sprachqualität bei VA-BQ-AQ bis zu Übertragungsraten von 16 kbit/s herunter ist einer 64 kbit/s PCM-LOG etwa gleichwertig.
- * Bei einer Übertragungsrate unter 24 kbit/s ist die VA-BQ-AQ der ADPCM-AQ nicht nur objektiv, sondern auch subjektiv deutlich überlegen.

Aus Untersuchungen anderer Autoren /8,33/ ist bekannt, dass Codiersysteme mit einem multiplikativen Rauschen (Quantisierungsfehlervarianz proportional der Signalvarianz, z.B. mit AQ) bei einem SNR von 20 dB eine Qualität aufweisen, bei der das codierte Signal vom Originalsignal kaum noch zu unterscheiden ist. Unsere Untersuchungen bestätigen diese Aussage, denn die Codierung mit VA-BQ-AQ weist bei einer Übertragungsrate von 12...16 kbit/s gerade ein SNR um 20 dB auf (vgl. Bild 6.2).

6.4 Vereinfachungen für die Realisierung der volladaptiven Blockquantisierung

Für eine Realisierung der diskutierten Codiersysteme ist der nötige Hardware-Aufwand ein weiterer wesentlicher Gesichtspunkt, der zur Auswahl eines Systems führt. Bevor der Vergleich durchgeführt wird, sollen noch mögliche Vereinfachungen für die volladaptive Blockquantisierung angegeben werden.

6.4.1 Einsatz von linearen Quantisierern

Für die volladaptive Blockquantisierung zeigt der Gauss-Quantisierer das beste SNR; bisher wurde stets ein nicht-linearer Quantisierer für kleinsten MQF eingesetzt. Für kleinere Übertragungsraten kann aber auch ein linearer Quantisierer eingesetzt werden, wobei der MQF nur wenig zunimmt /29/. Der Realisierungsaufwand verringert sich dagegen beträchtlich, da eine lineare Kennlinie erheblich einfacher zu verwirklichen ist. In Bild 6.5 ist der SNR-Rückgang ΔSNR bei Verwendung einer linearen Kennlinie angegeben. Der Wert ΔSNR vergrößert sich mit wachsender Übertragungsrate, was auch aus theoretischen Untersuchungen zu erwarten war /29/. Bei

einer Übertragungsrate von 16 kbit/s ist der SNR-Rückgang ΔSNR kleiner als 0.5 dB. Da andererseits bei dieser Übertragungsrate eine sehr gute Sprachqualität erreichbar ist, können also für die volladaptive Blockquantisierung lineare Quantisierer eingesetzt werden, ohne dass das SNR erheblich zurückgeht.

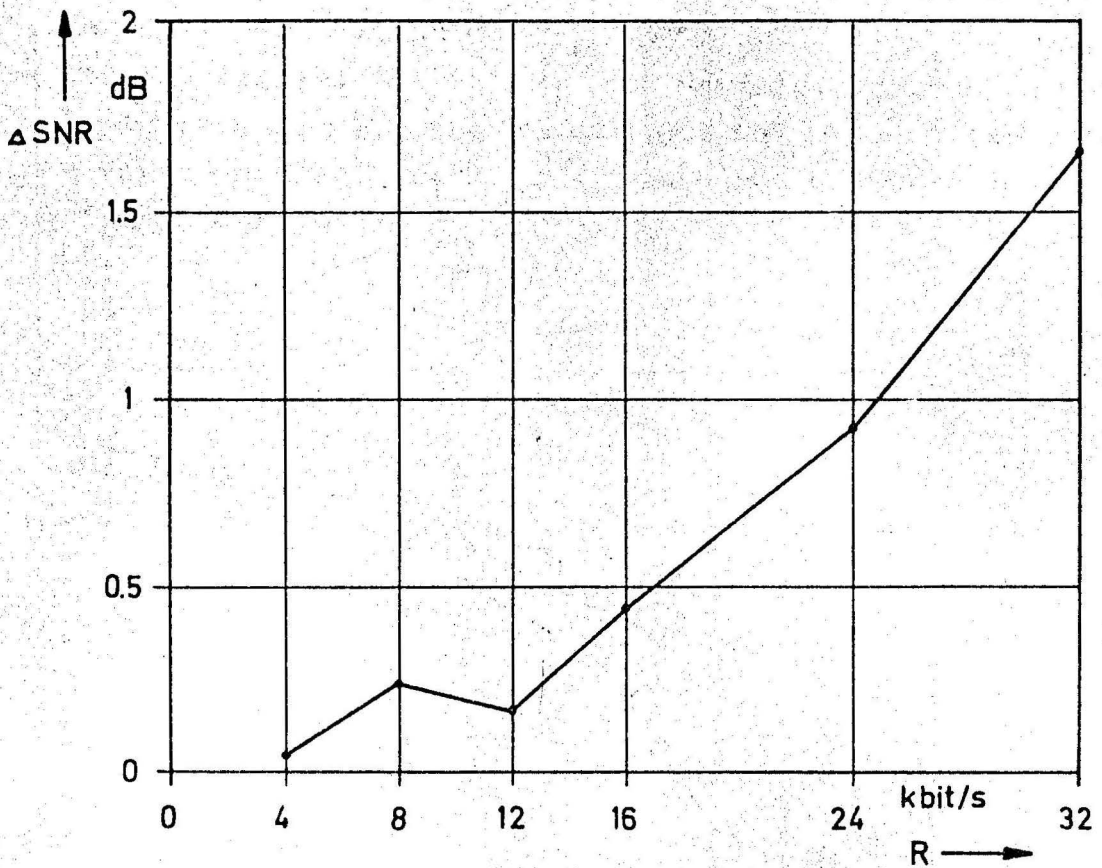


Bild 6.5 : SNR-Rückgang ΔSNR bei Verwendung eines linearen statt eines nichtlinearen Gauß-Quantisierers für die VA-BQ-AQ.
(Spt. 40-BP)

6.4.2 Begrenzung der maximalen Bitzahl

Die Spektralkoeffizienten werden mit unterschiedlichen Bitzahlen n quantisiert. Die bei einer Simulation überhaupt maximal auftretende Bitzahl n_{\max} bestimmt wesentlich den Aufwand für die Realisierung des Quantisierers.

In Bild 6.6 sind die Auftretenswahrscheinlichkeiten p_n der verschiedenen Bitzahlen n bei Übertragungsraten von 4 bis 16 kbit/s angegeben. Es ist ersichtlich, dass sich die dazugehörigen maximalen Bitzahlen n_{\max} von 5 auf 8 bit vergrößern. Die hohen Bitzahlen treten aber mit geringer Wahrscheinlichkeit auf, so dass eine Begrenzung der maximalen Bitzahl auf einen vorzugebenden Wert R_{\max} nur zu einem geringen SNR-Rückgang führen dürfte.

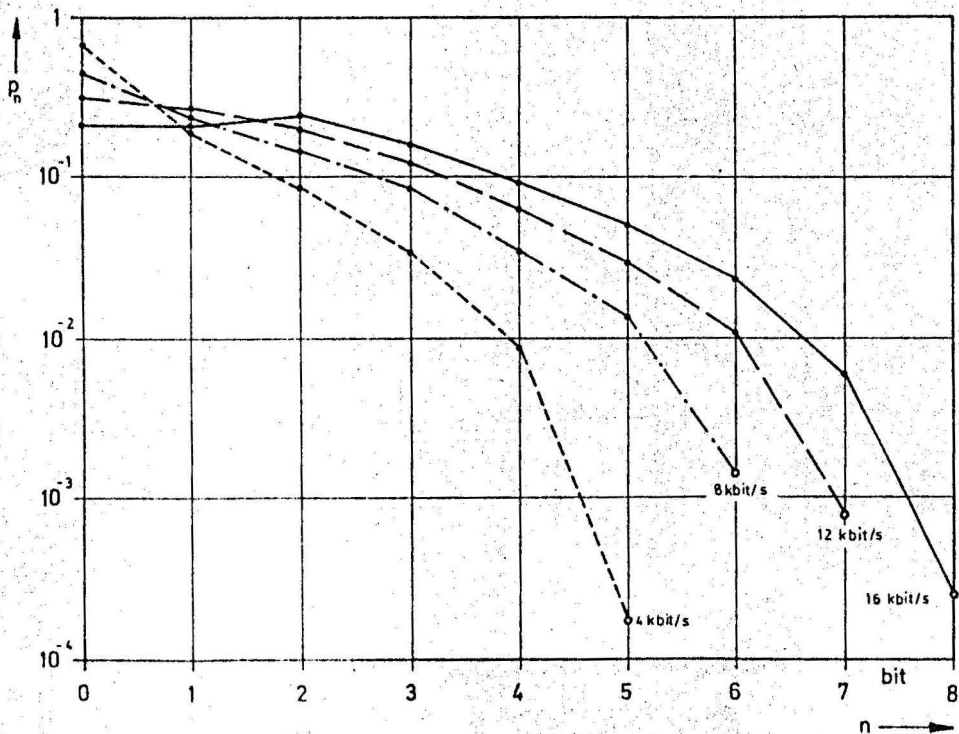


Bild 6.6 : Auftretenswahrscheinlichkeiten p_n der Bitzahlen n , mit denen die Spektralkoeffizienten quantisiert werden, bei Übertragungsraten von 4 bis 16 kbit/s . (VA-BQ-AQ, Spr. 40-BP)

Der SNR-Rückgang ΔSNR ist in Bild 6.7 angegeben. Die zusätzliche Randbedingung durch Vorgabe von R_{max} wurde gemäss dem in Abschnitt 2.4 beschriebenen Verfahren berücksichtigt. Für eine Übertragungsrate von 16 kbit/s zeigt sich bei $R_{\text{max}} = 6$ bit nur ein SNR-Rückgang von weniger als 0.1 dB.

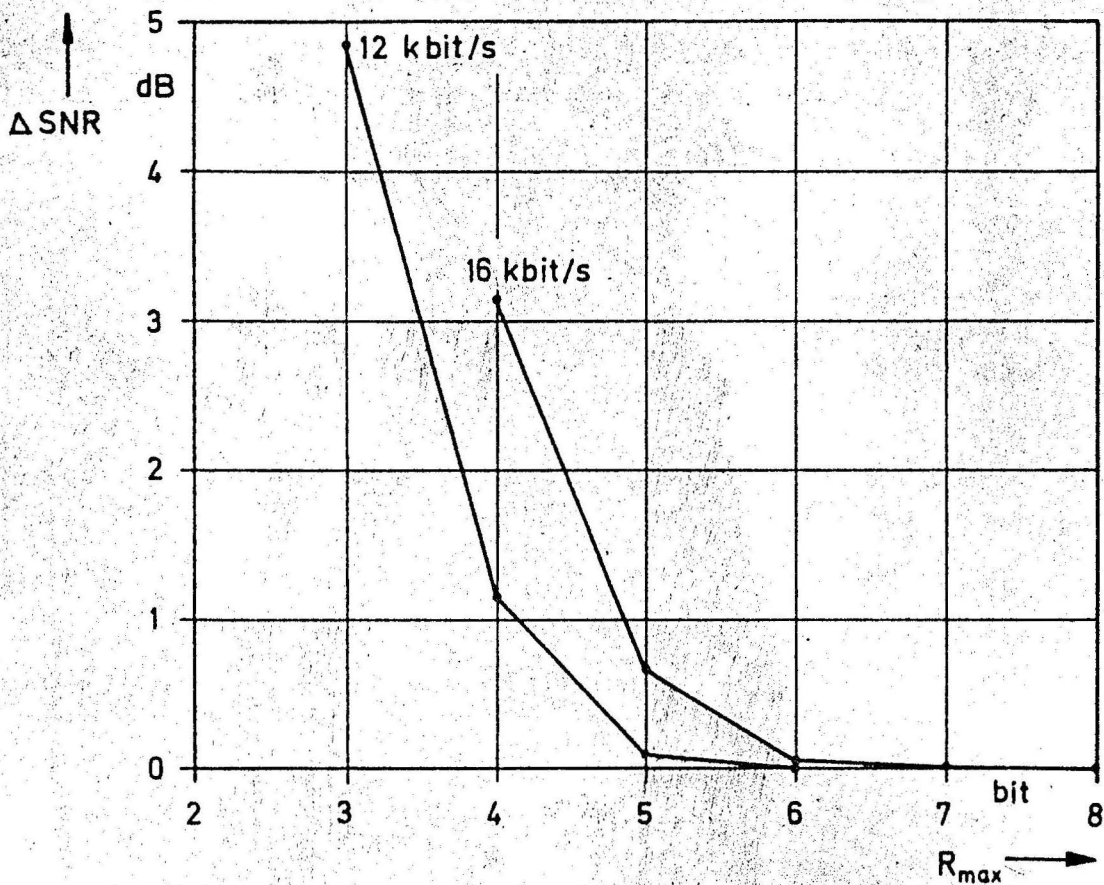


Bild 6.7 : SNR-Rückgang ΔSNR bei Vorgabe der maximalen Bitzahl R_{max} für Übertragungsraten von 12 und 16 kbit/s (VA-BQ-AQ , Spr. 40-BP)

Aus diesen Untersuchungen kann also geschlossen werden, dass für die volladaptive Blockquantisierung mit einer Übertragungsrate von 16 kbit/s ein linearer Quantisierer mit einer maximalen Bitzahl von 6 bit vollkommen ausreichend ist.

Anmerkung:

Durch Vorgabe von R_{\max} wird die minimal erreichbare Verzerrung in einem codierten Datenblock begrenzt:

Die mittlere Verzerrung D_{BQ} bei der Blockquantisierung ist nach Gl. 2.1.22

$$D_{\text{BQ}} = \frac{1}{N} \sum_{i=1}^N D_i = \frac{1}{N} 2^{2\delta} \sum_{i=1}^N \sigma_i^2 \cdot 2^{-2R_i} \quad (6.4.1)$$

R_{\max} sei die maximale Bitzahl; mit

$$R_{\max} \geq R_i ; \quad i = 1, \dots, N \quad (6.4.2)$$

ist folgende Abschätzung möglich:

$$D_{\text{BQ}} \geq \frac{1}{N} 2^{2\delta} \sum_{i=1}^N \sigma_i^2 \cdot 2^{-2R_{\max}} = \sigma^2 \cdot 2^{2\delta} \cdot 2^{-2R_{\max}} \quad (6.4.3)$$

Die rechte Seite in Gl. 6.4.3 ist die Verzerrung $D_{\text{PCM}, R_{\max}}$ bei einer PCM-Codierung des Datenblocks mit R_{\max} bit pro Abtastwert. Damit gilt:

$$D_{\text{BQ}} \geq D_{\text{PCM}, R_{\max}} \quad (6.4.4)$$

Die Verzerrung bei einer Blockquantisierung ist also mindestens so gross wie bei einer PCM-Codierung mit R_{\max} bit.

Für eine Gaussquelle und $R_{\max} = 6$ bit wird das SNR bei PCM /18/ mit einem nichtlinearen Gauss-Quantisierer

$$\text{SNR}_{\text{PCM}, R_{\max}} = 10 \lg \frac{\sigma^2}{D_{\text{PCM}, R_{\max}}} = 31.89 \text{ dB} \quad (6.4.5)$$

Das bedeutet also, der Signal-Rausch-Abstand in einem Datenblock SNR_{BQ} bei Blockquantisierung ist für $R_{\max} = 6$ bit nie grösser als 32 dB. Dieser Wert ist aber für eine gute Sprachqualität bei Codierung mit adaptiven Quantisierern weit mehr als ausreichend.

6.5 Vergleich des Rechenaufwandes bei volladaptiver Blockquantisierung und bei ADPCM-AQ

Es soll nur eine grobe Abschätzung angegeben werden. Als Vergleichskriterium wird die Zahl der Multiplikationen verwendet, die zu den rechenintensivsten Operationen zählen.

Für die Bestimmung der Bitzuordnung bei der Blockquantisierung ist eine Logarithmusbildung erforderlich. Da die notwendige Genauigkeit aber relativ gering ist (vgl. Bild 5.10, Quantisierung der Stützwerte), wird angenommen, dass die Logarithmusfunktion durch einen Polygonzug approximiert werden kann und daher nur eine lineare Interpolation erforderlich ist. Die beiden Codiersysteme weisen folgende Parameter auf:

VA-BQ-AQ:

DCT; Blocklänge $N = 128$; Zahl der Stützwerte $L = 16$.

ADPCM-AQ:

Segmentlänge $N = 128$; Zahl der Prädiktorkoeffizienten $N_K = 10$.

Die Zahl der Multiplikationen pro Block bzw. pro Segment ist in Tabelle 6.1 angegeben.

Operation	ADPCM-AQ	VA-BQ-AQ
Bestimmung des Pegels (für AQ)	N	N
Kennwerte für das Kurzzeit-LDS	AKF: $N \cdot N_K$ Prädiktorkoeffizienten mit rekursiven Verfahren: N_K^2	Stützwerte: N Interpolation der Stützwerte: N Logarithm. der geschätzten Varianzen: N
Codierung des Datenvektors	$N \cdot N_K$	$2 N \cdot \text{ld } N$
Gesamtzahl	$N(2N_K + 1) + N_K^2$	$N(2 \text{ld } N + 4)$
Zahl pro Abtastwert für das Beispiel	22	18

Tabelle 6.1: Zahl der Multiplikationen bei Codierung mit ADPCM-AQ und VA-BQ-AQ (nur Grössenordnung berücksichtigt)

Aus der letzten Zeile in Tabelle 6.1 ist ersichtlich, dass beide Codierverfahren etwa die gleiche Zahl von Multiplikationen benötigen. Das bedeutet, dass auch bezüglich des Aufwandes bei einer Hardware-Realisierung die volladaptive Blockquantisierung ein Konkurrent zur ADPCM-AQ sein kann.

7. Sprachübertragung mit volladaptiver Blockquantisierung bei sehr kleinen Übertragungsraten

7.1 Bitzuordnung nach einem objektiven Kriterium

Mit der volladaptiven Blockquantisierung ist bis zu Übertragungsraten von 12...16 kbit/s herunter eine sehr gute Qualität mit praktisch nicht wahrnehmbaren Verzerrungen erreichbar. Bei weiterer Verringerung der Übertragungsrate bis auf 4 kbit/s treten wahrnehmbare Verzerrungen auf, die sich in folgender Weise bemerkbar machen:

- * Das codierte Sprachsignal klingt dumpf, da die hohen Frequenzanteile fehlen.
- * Es treten eigentümliche Zwitschergeräusche auf, da bei stark korrelierten Lauten oft nur die erste Formantfrequenz übertragen wird.

Die Ursache dieses Verhaltens ist in der Bitzuordnung zu suchen, die streng nach einem objektiven Kriterium vorgenommen wird (kleinster MQF). Mit Reduktion der Übertragungsrate werden dann speziell bei stark korrelierten Lauten mit ausgeprägten Polstellen z.T. nur noch sehr wenige Spektrallinien übertragen.

Die Wirkungsweise der Bitzuordnung nach kleinstem MQF ist in Bild 7.1 demonstriert. In der oberen Darstellung sind die Varianzen der Spektralkoeffizienten (DCT, $N=32$) für den Sprachlaut "n" gezeigt. Die Varianzunterschiede bestimmen die Unterschiede in den Bitzahlen R_i ; ein Varianzverhältnis von 6 dB bewirkt eine Differenz von 1 bit (vgl. Gl. 2.1.26). Das Verfahren der Bitzuordnung arbeitet nun so, dass die "Bitfläche" unter der Kurve $10 \lg \sigma_i^2 / \sigma^2$ bei horizontalem Spiegel solange aufgefüllt wird, bis die Summe der Bitzahlen den vorgegebenen Wert

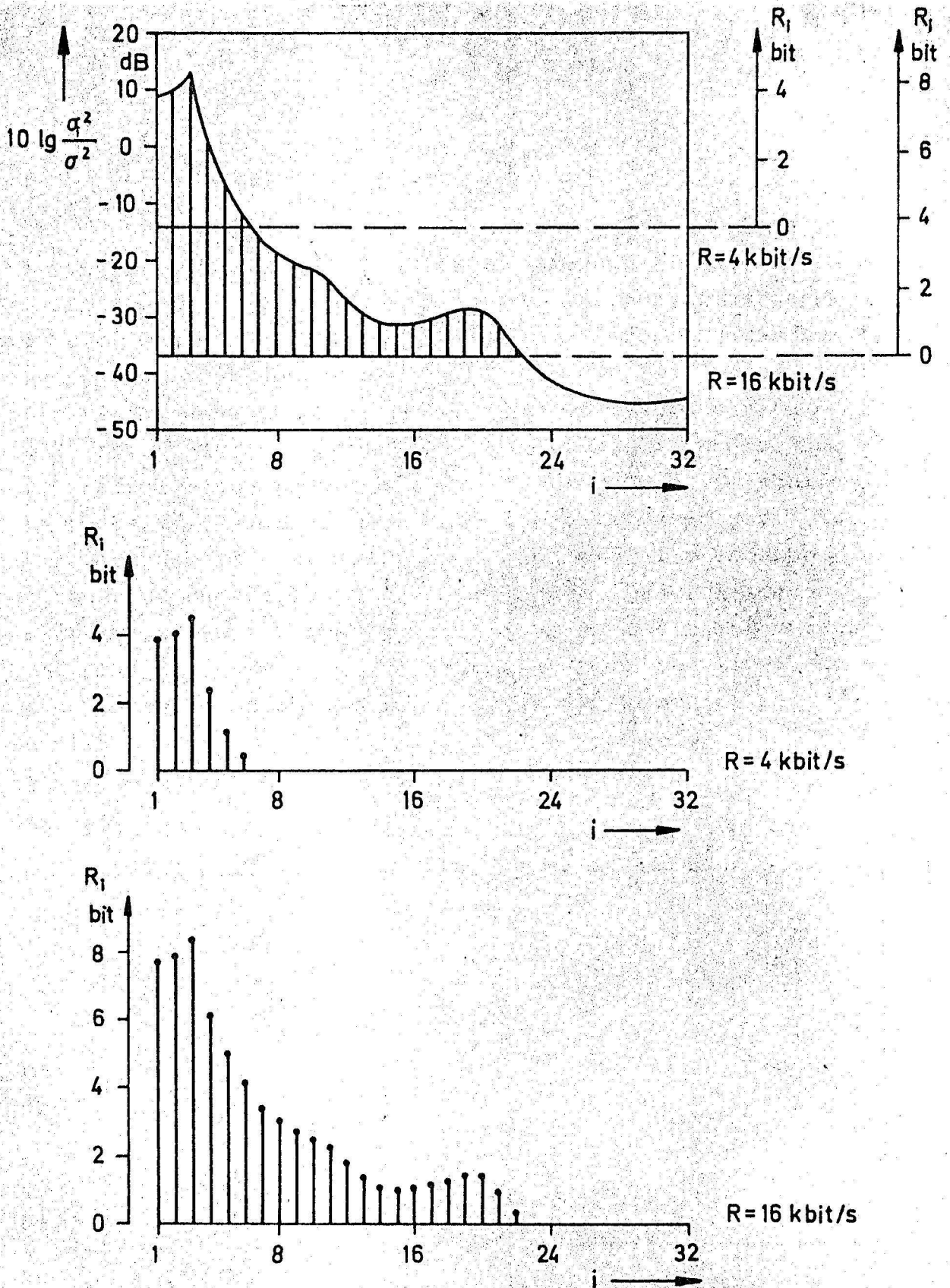


Bild 7.1 : Varianzen der Spektrallinien und Bitzuordnung bei 2 verschiedenen Übertragungsraten R für den Sprachlaut "n" .
(DCT, Blocklänge N=32)

$$\sum_{i=1}^N R_i = N \cdot \bar{R} \quad (7.1.1)$$

erreicht. Hier tritt also wieder das bekannte "water-filling" - Prinzip /25/ auf. Die Fläche unter dem Leistungsdichtespektrum wird wie mit Wasser aufgefüllt, bis eine vorgegebene Verzerrung bzw. hier Übertragungsrate erreicht ist. Es ist ersichtlich, dass mit Verringerung der Übertragungsrate R bzw. \bar{R} immer weniger Spektrallinien übertragen werden. Bei stark korrelierten Lauten mit einer ausgeprägten Polstelle (wie in Bild 7.1) werden dann alle $N \cdot \bar{R}$ verfügbaren bit nur zur Übertragung dieses Pols aufgewendet und alle anderen Formantfrequenzen vollständig unterdrückt. Dadurch treten subjektiv störende Verzerrungen auf. In Bild 7.2 ist entsprechend Bild 5.8b die Folge der Bitzuordnungen bei einer Übertragungsrate $R = 4$ kbit/s für das gleiche Sprachbeispiel dargestellt.

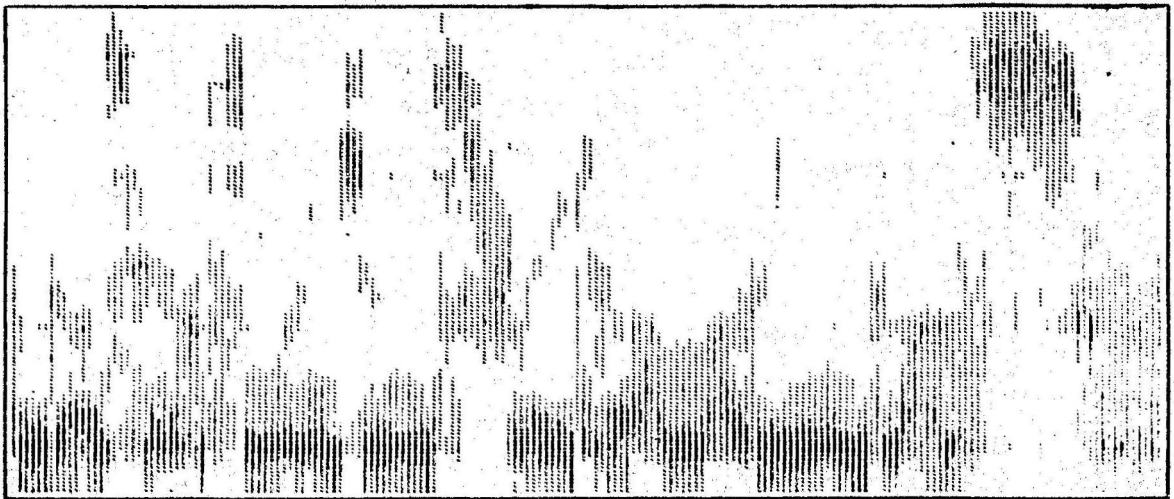


Bild 7.2 : Grafische Darstellung der Bitzuordnung für kleinsten MQF bei VA-BQ-AQ mit einer Übertragungsrate $R=4$ kbit/s .
(DCT, $N=128$)

Der Vergleich zeigt noch einmal deutlich, dass bei der kleinen Übertragungsrate in der Regel nur sehr wenige Spektrallinien pro Block übertragen werden. Dadurch wird das Codiersystem in der Bandbreite sehr schmalbandig. Die Verzerrungen äussern sich daher im Fehlen von höheren Formantfrequenzen, nicht aber in einem superponierten Rauschen wie bei vielen Codierverfahren, die im Zeitbereich arbeiten. Dies ist auch aus Bild 7.2 zu entnehmen: Die schwarzen Flächen im Diagramm kennzeichnen Spektralkoeffizienten mit hohen Bitzahlen. Es werden zwar nur wenige Spektralkoeffizienten übertragen, die aber mit hohen Bitzahlen quantisiert werden. So ist subjektiv ein "weisses Quantisierungsrauschen" nicht feststellbar.

7.2. Bitzuordnung nach subjektiven Kriterien

Eine subjektive Qualitätsverbesserung der volladaptiven Blockquantisierung bei kleinen Übertragungsraten ist erreichbar, wenn die Zahl der übertragenen Spektrallinien und damit die Übertragungsbandbreite vergrössert wird. Dazu ist es notwendig, den Algorithmus der Bitzuordnung aus Abschnitt 2.4 nach subjektiven Kriterien zu modifizieren.

Bei einer Blocklänge N und der mittleren Übertragungsrate \bar{R} stehen $N \cdot \bar{R}$ bit zur Codierung eines Datenblocks zur Verfügung. Die Bitzuordnung arbeitet nun in folgender Weise:

1. Den M_1 stärksten Spektralkoeffizienten ($M_1 < N \cdot \bar{R}$) wird je 1 bit zugeordnet.
2. Die restlichen $N \cdot \bar{R} - M_1$ bit werden nach dem Kriterium des kleinsten MQF (Verfahren aus Abschnitt 2.4) auf die Spektralkoeffizienten verteilt.

In Bild 7.3 ist die Folge der Bitzuordnungen für $R = 4\text{ kbit/s}$ und $M_1 = 54$ dargestellt. Der Vergleich mit Bild 7.2 zeigt das Anwachsen der Übertragungsbandbreite, die jetzt nicht kleiner als M_1 Spektrallinien pro Block sein kann.

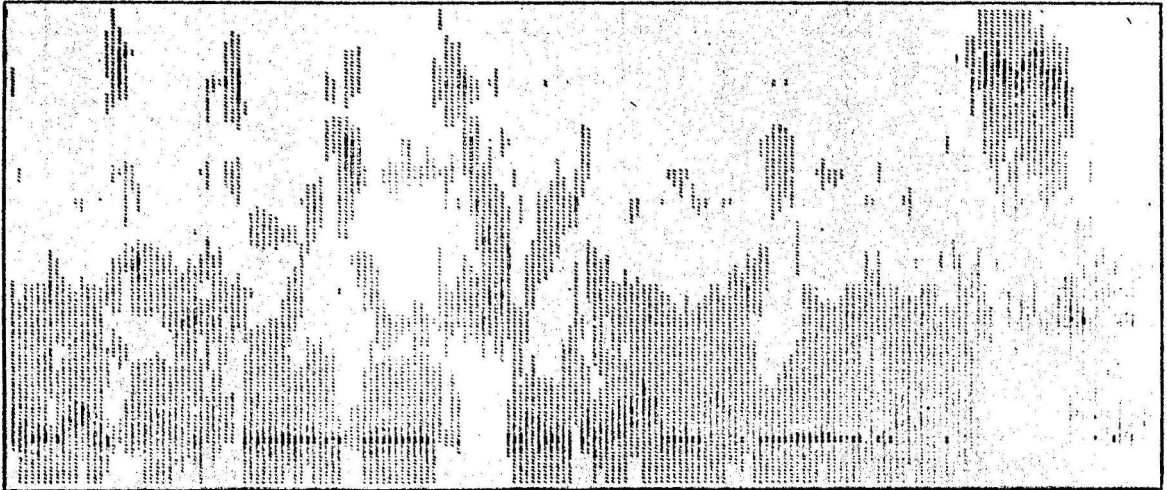


Bild 7.3 : Grafische Darstellung der Bitzuordnung nach subjektiven Kriterien bei VA-BQ-AQ mit einer Übertragungsrate $R=4$ kbit/s .
(DCT, $N=128$)

Zusätzlich wurde untersucht, ob es von der subjektiven Beurteilung her möglich ist, nicht übertragene Spektralkoeffizienten durch Rauscheinspeisung nachzubilden. Dazu werden auf der Empfängerseite für die M_R stärksten nichtübertragenen Spektralkoeffizienten statt des Quantisierer-Rekonstruktionswertes Null nun unkorrelierte gaussverteilte Zufallszahlen eingesetzt. Die Varianzen für die Zufallszahlen werden aus der spektralen Zusatzinformation $\underline{\sigma}_S^2$ ermittelt.

In Tabelle 7.1 sind Parameterkombinationen für verschiedene Übertragungsraten angegeben, die sich nach informatorischen Hörtests als geeignet gezeigt haben.

R kbit/s	$N \cdot \bar{R}$ bit/Block	M_1	M_R	$R_{ges} = R + R_S + R_V$ kbit/s
8	128	70	0	10
4	64	40	30	6
2	32	20	50	4

Tabelle 7.1: Parameter für VA-BQ-AQ bei kleinen Übertragungsraten R. (DCT; N=128)

Zusätzlich zur Datenrate R für die Quantisiererentscheidungen ist die Information über den Pegel R_V (Gl.4.4.1) und über das Kurzzeitspektrum R_S (Gl. 5.4.5) zu übertragen. Für $R_S + R_V$ sind etwa 2kbit/s nötig; die gesamte Datenrate R_{ges} ist in der letzten Spalte von Tabelle 7.1 angegeben.

Eine subjektive Bewertung der Codierungen nach Tabelle 7.1 zeigte folgendes:

- * Auch bei der kleinsten Übertragungsrate $R_{ges} = 4$ kbit/s ist eine gute Verständlichkeit gegeben.
- * Die sprecherspezifischen Merkmale der Sprache bleiben auch bei der kleinsten Übertragungsrate weitgehend erhalten. Das Klangbild ist natürlicher als bei einer Reihe von Vocoderverfahren.
- * Die zusätzliche Rauscheinspeisung beseitigt zwar den Schmalband-Effekt, bewirkt jedoch eine gewisse Rauigkeit der Sprache.

Die Güte des Codierverfahrens könnte sicher noch verbessert werden, wenn durch umfangreiche subjektive Untersuchungen festgestellt wird, welche Manipulationen im Spektralbereich

zu einer stärkeren Irrelevanzreduktion führen. Das bedeutet, die $N \cdot \bar{R}$ bit pro Block müssen optimal so aufgeteilt werden, dass subjektiv die geringsten Verzerrungen entstehen. Die notwendige Information über die Prioritäten im Spektralbereich ist für alle Spektralkoeffizienten bereits durch $\hat{\sigma}_s^2$ gegeben. Im einzelnen wären folgende Möglichkeiten denkbar:

- * Konzentrieren der Bits auf bestimmte Frequenzgruppen (z.B. Formantfrequenzen).
- * Übertragung nur jeder zweiten Spektrallinie innerhalb dieser Gruppen; die fehlenden werden durch Rauschen ersetzt.
- * Verwendung von korrelierter statt unkorrelierter Rausch-einspeisung.
- * Für eine Hardware-Realisierung ist auch die DST interessant, da sie einfacher als die DCT zu verwirklichen ist. Der objektive Gewinn ist dadurch zwar kleiner, aber dies kann eher nützlich als schädlich sein, da dadurch die hohe Bitkonzentration auf wenige Spektrallinien vermieden wird.

7.3 Vergleich mit anderen adaptiven Transformationscodierungen

Eine adaptive Bitzuordnung zu den Spektralkoeffizienten zur Übertragung von Sprache bei kleineren Bitraten ist auch von anderen Autoren untersucht worden. Es wurden dabei die WHT /5,12/ und die DFT /5/ eingesetzt. Die Verfahren arbeiten so, dass für jeden Block (Blocklänge N) die m im Absolutbetrag grössten Spektralkoeffizienten ($m < N$) herausgesucht werden. Nur diese m Koeffizienten werden quantisiert und übertragen. Als Zusatzinformation R_A müssen die Adressen dieser Koeffizienten übertragen werden. Dadurch wird der zusätzliche Datenfluss relativ hoch, wobei dann noch keine Information über die zu erwartenden Varianzen der Spektralkoeffizienten gegeben ist.

Tabelle 7.2 zeigt einige Codierungsbeispiele für eine Blocklänge $N = 64$.

Transformation	m	bit pro Block für		R_A kbit/s	R_{ges} kbit/s	Qualitäts- angabe (subj.)
		Adressie- rung	Quantisie- rung			
WHT	4	24	28	3	6.5	brauchbare Sprachwieder- gabe /12/
WHT	8	48	56	6	13	gute Sprach- wiedergabe /12/
WHT und DFT	12	60	59	7.5	14.875	entspricht 5-6 bit PCM-LOG /5/

Tabelle 7.2: Beispiele für Transformationscodierungen bei einer Blocklänge $N = 64$ nach /5,12/

Aus der Tabelle ist ersichtlich, dass für eine bessere Codierungsqualität ein Zusatzinformationsfluss von $R_A \approx 6 \dots 8$ kbit/s notwendig ist, wobei nur etwa 10...20 % der Spektrallinien eines Blocks übertragen werden können. Diese Zahl ist auf der anderen Seite minimal erforderlich, weil sonst subjektiv zu grosse Verzerrungen auftreten, die die Verständlichkeit erheblich beeinträchtigen.

Das in dieser Arbeit untersuchte Verfahren der volladaptiven Blockquantisierung (VA-BQ-AQ) benötigt dagegen durch Ausnutzen der Existenz eines "Basisspektrum" nur einen zusätzlichen Datenfluss $R_S = 2$ kbit/s, mit dem die Varianzen aller Spektralkoeffizienten gekennzeichnet werden. Aus diesem Datenfluss folgen alle notwendigen Informationen:

1. Bitzuordnung für die Spektralkoeffizienten
2. Aussteuerung der einzelnen Quantisierer.

3. Steuerung der Bitzuordnung bei subjektiven Kriterien.
4. Steuerung einer eventuellen Rauscheinspeisung; die notwendigen Varianzen werden ebenfalls aus R_G gewonnen.

Durch diesen Algorithmus ist es möglich, trotz adaptiver Bitzuordnung, nahezu den gesamten Datenfluss nur für die Quantisierung der Spektralkoeffizienten aufzuwenden. Daher können bei gleicher Übertragungsrate kleinere Verzerrungen erzielt werden bzw. bei kleinen Bitraten erheblich mehr Spektralkoeffizienten codiert werden.

Die volladaptive Blockquantisierung zeigt eine gewisse Ähnlichkeit mit Vocoderverfahren. In beiden Fällen wird segmentweise eine Information über das Kurzzeitspektrum gegeben. Am grössten ist die Entsprechung beim Kanalvocoder: Die spektrale Zusatzinformation besteht wie bei der volladaptiven Blockquantisierung aus den Leistungen in bestimmten Frequenzgruppen. Die volladaptive Blockquantisierung unterscheidet sich jedoch in folgenden Punkten von einem Vocoder:

1. Es ist keine Extraktion der Pitchfrequenz notwendig.
2. Es wird nicht nur ein Parameter (Pitchfrequenz) sondern es werden mehrere Parameter (Spektralkoeffizienten) übertragen.
3. Die Optimierung wird nicht im Analyseteil (beim Vocoder: Pitchextraktion und Kennzeichnung des Kurzzeitspektrums) sondern am Empfänger nach subjektiven Kriterien durchgeführt mit dem Ziel einer weitgehenden Irrelevanzreduktion.

Die volladaptive Blockquantisierung bei kleinen Übertragungsraten ($R_{ges} \approx 4$ kbit/s) stellt somit bezüglich der Übertragungsrate, des Realisierungsaufwandes und der Qualität eine mögliche Alternative zu den Codierungen mit Vocoderverfahren dar.

8. Einfluss von Kanalstörungen bei der volladaptiven

Blockquantisierung

In den bisherigen Betrachtungen ist vorausgesetzt worden, dass der digitale Übertragungskanal störungsfrei arbeitet. Werden bei der Übertragung einzelne bit verfälscht, so treten zusätzliche Verzerrungen auf, da am Decodierer dann falsche Rekonstruktionswerte eingesetzt werden. In den folgenden Abschnitten wird durch Simulation untersucht, welche Signal-Rausch-Abstände bei verschiedenen Fehlerraten erreichbar sind.

8.1 Einfluss von Kanalstörungen bei Systemen mit und ohne

Fehlerkorrektur

In diesem Abschnitt wird nur die volladaptive Blockquantisierung VA-BQ-AQ (DCT, N= 128) betrachtet. Dabei wird vorausgesetzt, dass die Zusatzinformation über den Pegel (V) und über das Kurzzeitspektrum ($\hat{\sigma}_s^2$) störungsfrei übertragen wird. Dies ist durch Hinzufügen von Redundanz zur Fehlerkorrektur erreichbar. Die gesamte Datenrate wird dadurch nur wenig vergrößert, da der Zusatzinformationsfluss selbst relativ gering ist.

Als Kanalstörungen wurden Einzelfehler mit einer Bitfehler-rate P simuliert. Zur Kennzeichnung der Quantisiererrekonstruktionswerte r_i wird ein gefalteter Binärcode (FBC) eingesetzt. Tabelle 8.1 zeigt ein Codierungsbeispiel für den nichtlinearen 3bit-Gauss-Quantisierer.

i	r_i	Code
1	2.152	011
2	1.344	010
3	0.756	001
4	0.245	000
5	-0.245	100
6	-0.756	101
7	-1.344	110
8	-2.152	111

Tabelle 8.1: FBC-Codierung für den 3bit-Gauss-Quantisierer

Wird z.B. bei der Übertragung aus 001 das Wort 101, so wird am Decodierer statt r_3 der Rekonstruktionswert r_6 eingesetzt und damit die Verzerrung vergrößert. Der Hauptdatenstrom bei der Übertragung besteht aus den Digitalworten für die Quantisierer-Rekonstruktionswerte. Bei der Übertragung dieser Daten werden 4 Fälle unterschieden:

1. FK-0 : Die Übertragung wird ohne Fehlerkorrektur durchgeführt.
2. FK-1 : Bei den 64 stärksten Spektrallinien wird das höchstwertige bit (das Vorzeichen-bit) durch Fehlerkorrektur geschützt.
/64
3. FK-2 : Bei den 64 stärksten Spektrallinien werden die beiden höchstwertigen bit (das Vorzeichen-bit und das zweite bit) durch Fehlerkorrektur geschützt.
/64

4. FK-128 (optimale Fehlerkorrektur): Es werden insgesamt 128 bit pro Block durch Fehlerkorrektur geschützt. Bei den stärksten Spektrallinien werden dabei fast alle, bei den schwächeren Spektrallinien nur das Vorzeichen-bit oder gar kein bit geschützt. Die notwendige Zuordnung wird auf folgendem Wege erhalten: Es wird zunächst aus $\hat{\sigma}_s^2$ die Bitzuordnung berechnet, die für eine Codierung mit $R=8$ kbit/s bzw. 128 bit pro Block optimal ist. Die so erhaltenen Bitzahlen R_i geben für jede Spektrallinie mit dem Index i die Zahl der zu schützenden (höchstwertigen) bit an.

Die Instrumentierung der Fehlerkorrektur wurde dabei nicht simuliert, sondern es wurde angenommen, dass alle Fehler korrigiert werden. Daher sind die Ergebnisse für 2)...4) nur als obere Schranken anzusehen.

Die Simulationsergebnisse für verschiedene Bitfehlerraten P sind in Bild 8.1 angegeben.

Der Vergleich der beiden Darstellungen zeigt:

- * Ohne Fehlerkorrektur wird das SNR bei grossen Fehlerraten nur durch den Kanal und nicht durch die Übertragungsrate R bestimmt.
- * Bei hohen Fehlerraten ($P=5\%$) müssen mindestens 128 bit/Block geschützt werden, damit das SNR nur wenig unter 20 dB absinkt.
- * Das Verfahren FK-128 ist dem Verfahren FK-2/64 bei hohen Fehlerraten um 2...3 dB überlegen, obwohl in beiden Fällen 128 bit geschützt werden. Hier macht sich wieder der Vorteil der volladaptiven Blockquantisierung bemerkbar, die eine sehr effektive Prioritätensteuerung im Spektralbereich ermöglicht. Die notwendige Information wird aus den ohnehin zu übermittelnden Daten $\hat{\sigma}_s^2$ gewonnen.

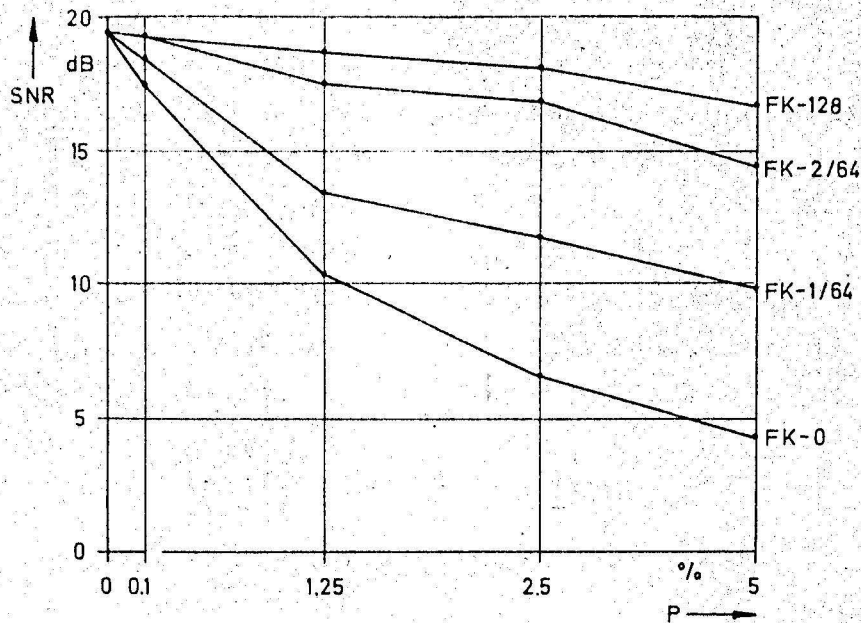


Bild 8.1a :
R=16 kbit/s

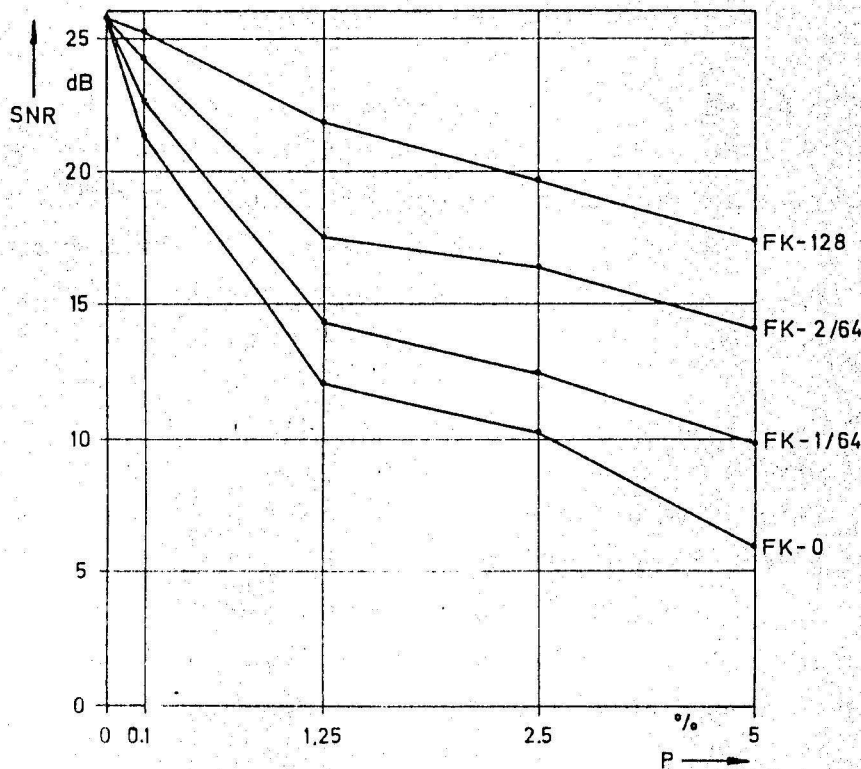


Bild 8.1b :
R=24 kbit/s

Bild 8.1 : Signal-Rausch-Abstand bei verschiedenen Bitfehler-
raten P für die Codierung VA-BQ-AQ .
(Spr. 13-BP)

8.2 Vergleich mit anderen Codierverfahren

Das Verhalten bei Kanalstörungen wurde auch für ein anderes Codierverfahren untersucht, um einen Vergleich zu ermöglichen. Dazu wurde eine PCM-Codierung mit adaptivem Quantisierer (PCM-AQ) eingesetzt. Auch für die PCM-AQ wurde eine Vorwärtsregelung verwendet: Das Signal wird segmentiert (Segmentlänge $N = 128$) und mit einem optimal ausgeregelten Gauss-Quantisierer codiert. Die Zusatzinformation V über den Pegel soll auch hier in jedem Falle fehlergeschützt sein.

Es werden zwei Fälle unterschieden:

- 1) FK-0 : ohne Fehlerkorrektur
- 2) FK-1 : Für jeden Quantisierer-Rekonstruktionswert wird das höchstwertige bit geschützt.

Bild 8.2 zeigt die Ergebnisse für verschiedene Fehlerraten. Bei $P = 5\%$ Fehlerrate sind mit FK-0 etwa 5 dB und mit FK-1 etwa 10 dB Signal-Rausch-Abstand für PCM-AQ mit 32 kbit/s erreichbar. Diese Werte gelten auch für eine Reihe anderer Codierverfahren wie DPCM-AQ und ADPCM-AQ /34/, wenn im ungestörten Fall das SNR grösser als 20 dB ist.

Zusätzlich ist in Bild 8.2 der Signal-Rausch-Abstand für VA-BQ-AQ (Übertragungsrate 16 kbit/s) angegeben. Im ungestörten Fall erreichen VA-BQ-AQ und PCM-AQ beide ein SNR von etwa 20 dB. Bei Zunahme der Kanalstörungen und ohne Fehlerkorrektur ist der SNR-Rückgang bei beiden Codierverfahren etwa gleich. Wesentlich anders sind die Verhältnisse, wenn pro Block bzw. pro Segment 128 bit gegen Fehler geschützt werden. In diesem Fall ist die volladaptive Blockquantisierung mit FK-128 erheblich überlegen; bei $P = 5\%$ ist das SNR gegenüber PCM-AQ mit FK-1 um 8 dB vergrössert. Es zeigt sich also, dass die volladaptive Blockquantisierung erheblich wirksamer gegen Kanalstörungen geschützt werden kann.

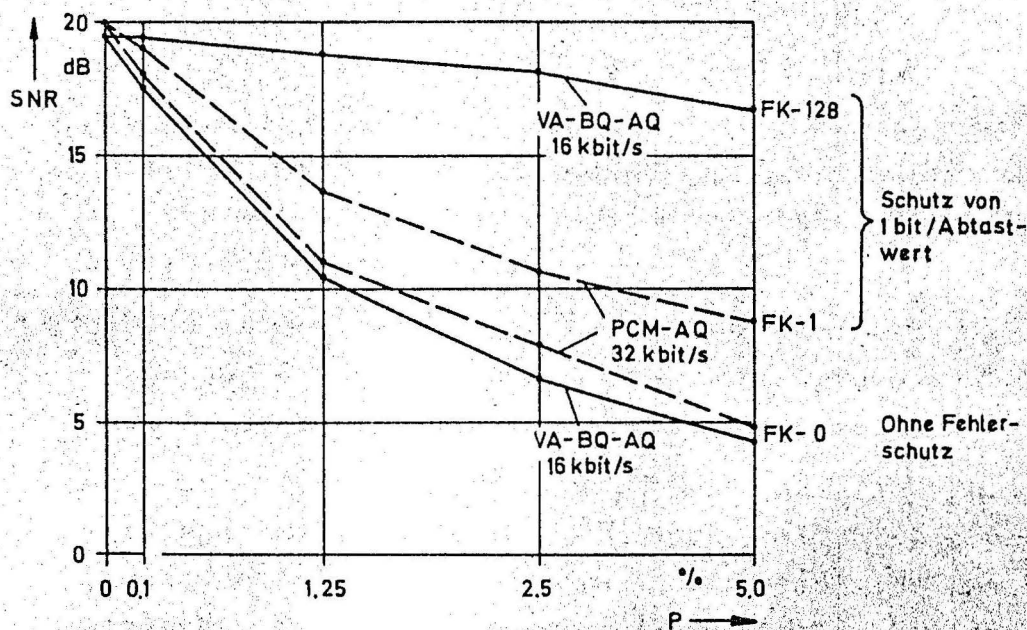


Bild 8.2 : Signal-Rausch-Abstand bei verschiedenen Bitfehler-raten P für VA-BQ-AQ und PCM-AQ.
(Spr. 13-BP)

9. Zusammenfassende Bewertung der volladaptiven Blockquantisierung

Die volladaptive Blockquantisierung (VA-BQ-AQ) arbeitet mit adaptiver Bitzuordnung und adaptiven Quantisierern für die Spektralkoeffizienten, aber ohne die Notwendigkeit einer pitch-, synchronen Steuerung. Bei einer Blocklänge $N=128$ und der Diskreten Cosinus Transformation (DCT) sind folgende Ergebnisse erreichbar:

- * SNR-Anstieg gegenüber einer nichtadaptiven Blockquantisierung mit DCT: 8...12 dB.
- * SNR-Anstieg gegenüber einer PCM-Codierung mit logarithmischem Quantisierer: 17...23 dB.

- * Die Übertragungsrate kann von 64 kbit/s bei PCM-Codierung auf 12...16 kbit/s bei volladaptiver Blockquantisierung verringert werden, ohne dass subjektiv eine Verschlechterung der Sprachqualität feststellbar ist.
- * Die volladaptive Blockquantisierung ist einer ADPCM mit adaptiver Prädiktion und adaptivem Quantisierer um 3 dB, bei kleineren Übertragungsraten (16 kbit/s) um 6dB im Signal-Rausch-Abstand überlegen.
- * Die volladaptive Blockquantisierung kann erheblich effektiver gegen Kanalstörungen geschützt werden als entsprechende Verfahren, die im Zeitbereich arbeiten.
- * Auch bei kleinen Übertragungsraten von 4 kbit/s ist eine gut verständliche Sprachcodierung möglich. Die volladaptive Blockquantisierung schliesst daher bezüglich der Übertragungsrate die Lücke zwischen Vocoderverfahren und Codierungen im Zeitbereich (PCM,DPCM,DM).

Der Realisierungsaufwand für die VA-BQ-AQ bewegt sich etwa in der Grössenordnung wie für eine ADPCM mit 10 Prädiktorkoeffizienten. Bei einer Übertragungsrate von 16 kbit/s können zur Quantisierung der Spektralkoeffizienten lineare Quantisierer eingesetzt werden, wobei eine maximale Bitzahl von 6 bit vollkommen ausreichend ist.

Das Codiersystem arbeitet ohne Entropiecodierung und liefert daher eine synchrone Datenrate. Von der gesamten Übertragungsrate wird nur ein Anteil von etwa 2 kbit/s benötigt, um die Zusatzinformation über Kurzzeitspektrum und Sprachpegel zu übertragen.

10. Dank

Die Untersuchungen wurden von der Deutschen Forschungsgemeinschaft Bad Godesberg finanziert. Der Informatik-Forschungsbereich der TU Berlin stellte die notwendige Rechenzeit am System IBM 370/158 zur Verfügung. Beiden Institutionen gilt unser Dank.

Wir danken besonders den Mitarbeitern E. Grossmann und B. Heuser, die einige nichtadaptive Blockquantisierungsverfahren vor Beginn der Untersuchungen auf dem Digitalrechner implementiert haben.

11. Literatur

- /1/ NOLL, P.
Sprachübertragung mit adaptiven Differenz-PCM-Verfahren.
Technischer Bericht Nr. 164 des Heinrich-Hertz-Instituts, Berlin-Charlottenburg, 1973
- /2/ NOLL, P.
Untersuchungen zur Sprachcodierung mit adaptiven Prädiktionsverfahren.
NTZ (Nachrichtentechn. Zeitschrift), Heft 2, 1974, S. 67-72
- /3/ CAMPANELLA, S.J.; ROBINSON, G.S.
A comparison of orthogonal transformations for digital speech processing.
IEEE Trans. on Commun. Technology, Vol.COM-19, 1971, S. 1045 - 1049
- /4/ NOLL, P.
Redundanzreduzierende Sprachcodierung - eine vergleichende Betrachtung verschiedener Verfahren.
NTG-Tagung "Signalverarbeitung", 1973, Erlangen
- /5/ GETHÖFFER, H.
Transform coding of speech using Walsh functions.
Proc. of the 1973 Symposium on Theory and Applications of Walsh Functions. Naval Research Lab. Washington D.C.
- /6/ CUMMISKEY, P.; JAYANT, N.S.; FLANAGAN, J.L.
Adaptive quantization in differential PCM coding of speech.
Bell Syst. Tech.J., Vol.52, 1973, S.1105-1118
- /7/ NOLL, P.
Adaptive quantizing in speech coding systems.
International Zürich Seminar on Digital Communications, 12th-15th March 1974. Swiss Federal Institute of Technology (ETHZ).

- /8/ MODENA, G; NEBBIA, L; SCAGLIOLA, C.
Bitrate reduction of digital speech transmission
by linear transformations.
NTG-Tagung "Signalverarbeitung", 1973, Erlangen
- /9/ ANDERSON, G.B.; HUANG, T.S.
Piecewise fourier transformation for picture
bandwidth compression.
IEEE Trans. on Commun. Technology, Vol. COM-19,
1971, S. 133-140
- /10/ TASTO, M.; WINTZ, P.A.
Image coding by adaptive block quantization.
IEEE Trans. on Commun. Technology, Vol. COM-19,
1971, S. 957-971
- /11/ WINTZ, P.A.
Transform picture coding.
Proc. of the IEEE, Vol. 60, 1972, S. 809-820
- /12/ SHUM, F.; ELLIOT, A.; BROWN, W.
Speech processing with Walsh-Hadamard transforms.
IEEE Trans. on Audio and Electroacoustics,
Vol. AU-21, 1973, S. 174 - 179
- /13/ PRATT, W.; CHEN, W.; WELCH, L.
Slant transform image coding.
IEEE Trans. on Commun., Vol. COM-22, 1974,
S. 1075 - 1093
- /14/ AHMED, N.; NATARAJAN, T.; RAO, K.
Discrete cosine transform.
IEEE Trans. on Comput., Vol. C-23, 1974, S. 90-93
- /15/ LÄNGLE, D.; PAULUS, E.
Efficiency and limitations of linear transforma-
tions in digital speech transmission.
Proceedings of the Speech Commun. Seminar,
Stockholm, 1. - 3. August 1974
- /16/ HUANG, J.; SCHULTHEISS, P.
Block quantization of correlated gaussian random
Variables.
IEEE Trans. on Commun. Systems, Vol. CS-11, 1963,
S. 289-296

/17/

PEARL, J.

Basis-restricted transformations and performance measures for spectral representations.
IEEE Trans. on Inform. Theory, Vol. IT-17, 1971, S. 751-752

/18/

NOLL, P.; ZELINSKI, R.

Ein Beitrag zur Quantisierung gedächtnisfreier Modellquellen.
Technischer Bericht Nr. 170 des Heinrich-Hertz-Instituts Berlin-Charlottenburg, 1974

/19/

GEORGI, K.H.

Die Orthogonalität der aus den Abtastwerten von sin- und cos-Funktionen gebildeten reellen Fouriermatrix.
Forschungsbericht Nr. 39 der TH Darmstadt, FG Übertragungstechnik, 1972

/20/

PEARL, J.

On coding and filtering stationary signals by discrete fourier transforms.
IEEE Trans. on Inform. Theory, Vol. IT-19, 1973, S. 229 - 232

/21/

SCHÜSSLER, H. W.

Digitale Systeme zur Signalverarbeitung.
Springer Verlag, Berlin - Heidelberg - New York, 1973

/22/

ZELINSKI, R.

Anwendungen der PARCOR-Koeffizienten für die lineare Prädiktion stationärer Prozesse.
Technischer Bericht Nr. 173 des Heinrich-Hertz-Instituts Berlin-Charlottenburg, 1974

/23/

HABIBI, A.; HERSHEL, R.S.

A unified representation of differential pulse-code modulation (DPCM) and transform coding systems.
IEEE Trans. on Commun., Vol. COM-22, 1974, S. 692-696

/24/

OPPENHEIM, A; WEINSTEIN, C.

Effects of finite register length in digital filtering and the fast fourier transform.
Proc. of the IEEE, Vol. 60, 1972, S. 957-976

- /25/ BUNIN, B.
Rate-distortion functions for gaussian markov processes.
Bell Syst. Tech.J., Vol. 48, 1969, S.3059-3074
- /26/ BERGER, T.
Rate distortion theory.
Prentice-Hall, 1971
- /27/ ZELINSKI, R.; NOLL, P.
Optimale Quantisierung gedächtnisfreier Gamma-
quellen und Anwendungen für die PCM-Codierung
von Sprache.
Technischer Bericht Nr. 169 des Heinrich-Hertz-
Instituts Berlin-Charlottenburg, 1974.
- /28/ ROE, G.M.
Quantizing for minimum distortion.
IEEE Trans. on Inform. Theory, Vol.IT-10, 1964,
S. 384-385
- /29/ ZELINSKI, R.; NOLL, P.
Optimale Quantisierung gedächtnisfreier Gauss-
quellen.
AEÜ, Band 27, 1973, S. 497-504
- /30/ NOLL, P.
A comparison of various quantization schemes for
the encoding of speech signals.
Bell Syst. Tech.J. 1975 (zur Veröffentlichung
angenommen).
- /31/ NOLL,P.; BLOCK, R.
Simulation von Differenz-PCM-Systemen zur Sprach-
codierung.
Technischer Bericht Nr. 150 des Heinrich-Hertz-
Instituts Berlin-Charlottenburg, 1972
- /32/ JAYANT, N.S.
Digital coding of speech waveforms: PCM, DPCM,
and DM quantizers.
Proc. of the IEEE, Vol.62, 1974, S. 611-632

/33/

JAYANT, N.S.

Mündliche Mitteilung

/34/

NOLL, P.

Effects of channel errors on the signal-to-noise performance of speech encoding systems. Bell Syst. Tech. J. 1975 (zur Veröffentlichung angenommen).

