

HEINRICH-HERTZ-INSTITUT — BERLIN-CHARLOTTENBURG

Technischer Bericht Nr. 164

Sprachübertragung mit adaptiven
Differenz-PCM-Verfahren

von

Dr.-Ing. Peter Noll



Berlin

1 9 7 3

HEINRICH - HERTZ - INSTITUT

Berlin - Charlottenburg



Technischer Bericht Nr. 164

Sprachübertragung mit adaptiven Differenz - PCM - Verfahren

von

Dr.-Ing. P. Noll

Berlin
1973

Sprachübertragung mit adaptiven Differenz - PCM - Verfahren

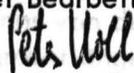
Zusammenfassung:

Die Redundanz von Sprachsignalen lässt sich durch Differenz-PCM-Verfahren eliminieren; der verbleibende, relevante Anteil kann dann mit verringerter Bitrate digital übertragen werden.

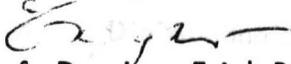
In der vorliegenden Arbeit wird gezeigt, welche Bitratenreduktion durch adaptive Verfahren erreichbar sind, bei denen sich der Codierer an die Änderungen der statistischen Kenn-
daten von Sprachsignalen anpasst. Verschiedene, auch für Hardware-Realisierungen in-
teressante Algorithmen werden dazu miteinander verglichen. Die Untersuchungen zeigen,
dass es möglich ist, bei der Digitalisierung eines jeden Abtastwertes des Sprachsignals
3 bis 4bit einzusparen. Damit kann die Zahl der über PCM -Strecken geführten Telefonver-
bindungen nahezu verdoppelt werden.

Heinrich-Hertz-Institut

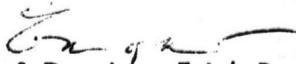
Der Bearbeiter


(Dr.-Ing. P. Noll)

Der Abteilungsleiter


(Prof. Dr.-Ing. Erich R. Berger)

Der Institutsdirektor


(Prof.Dr.-Ing. Erich R. Berger)

Berlin-Charlottenburg, den 25. Mai 1973

Inhaltsverzeichnis

	Seite
1. <u>Einleitung</u>	1
2. <u>Wichtige Grundlagen der nichtadaptiven DPCM</u>	1
2.1 Das Prinzip der DPCM	1
2.2 Die Optimierung des Prädiktors	4
2.3 Ergebnisse für eine mittlere Sprach-AKF	6
2.4 Zusammenfassung	11
3. <u>ADPCM - Strukturen</u>	11
3.1 Segmentierung des Sprachsignals	12
3.2 ADPCM - Struktur bei vorlaufender Analyse	15
3.3 ADPCM - Struktur bei nachlaufender Analyse	16
4. <u>ADPCM- Codierung bei vorlaufender Analyse</u>	18
4.1 <u>Bestimmung der Prädiktorkoeffizienten</u>	18
4.1.1 Deterministische Formulierung des Problems	18
4.1.2 Vocoderformulierung des Problems	24
4.1.3 Statistische Formulierung des Problems	26
4.2 <u>Simulationsergebnisse</u>	28
4.2.1 Die Abhängigkeit des Gewinns von dem Koeffizienten-Bestimmungsverfahren und den Segmentlängen	28
4.2.2 Die Abhängigkeit des Gewinns von den Eigenschaften des Quantisierers	31
4.2.3 Die Abhängigkeit des Gewinns von der Zahl der Prädiktorkoeffizienten	32
4.2.4 Die Abhängigkeit des Gewinns von der Sprachprobe	33
4.3 <u>ADPCM mit gesteuertem Quantisierer</u>	35
4.3.1 ADPCM ohne Entropiecodierung	35
4.3.2 ADPCM mit Entropiecodierung	38

	Seite
4.4	<u>Die Bestimmung der Prädiktorkoeffizienten aus den AKF-Werten</u> 38
4.5	<u>Verfahren der vereinfachten AKF-Bestimmung</u> 41
4.5.1	Einseitige Polaritätskorrelationsfolge 42
4.5.2	Zweiseitige Polaritätskorrelationsfolge 43
4.6	<u>Das Problem der Übertragung der Information über die Prädiktoreinstellung</u> 44
4.6.1	Übertragung der AKF-Werte 47
4.6.2	Übertragung der PARCOR-Koeffizienten 47
4.6.3	Übertragung der Lage der Pole 48
4.6.4	Übertragung der Prädiktorkoeffizienten 48
5.	<u>ADPCM - Codierung bei nachlaufender Analyse (Gradientenverfahren)</u> 54
5.1	Beschreibung des Prinzips der Gradientenverfahren 54
5.2	Die Anwendung des Verfahrens auf eine ADPCM-Sprachcodierung 58
6.	<u>Segmentanalysen</u> 60
6.1	<u>Die Interpretation des Prädiktionsverfahrens im Frequenzbereich</u> 61
6.2	<u>Analysen einzelner Segmente</u> 64
6.2.1	Codierung eines stimmhaften Lautes 65
6.2.2	Codierung eines stimmlosen Lautes 73
6.2.3	Grobe Quantisierung 75
6.3	<u>Fortlaufende Segmentanalysen</u> 75
7.	<u>Der Zusammenhang zwischen der ADPCM - Codierung und dem Vocoderprinzip</u> 81
7.1	Vocoder-Prinzipien 81
7.2	ADPCM - Codierung und Prädiktions-Vocoder 82
7.3	Der ADPCM - Vocoder 83
8.	<u>Zusammenfassende Bewertung des ADPCM - Verfahrens</u> 84
9.	<u>Literatur</u> 86

1. Einleitung

Bei dem Verfahren der nichtadaptiven Differenz-Pulscodemodulation (DPCM) wird ein stationärer Prozess einem Prädiktionsverfahren mit einer Quantisierung des Differenzsignals unterworfen. Aus der Autokorrelationsfunktion (AKF) des Prozesses kann näherungsweise derjenige Prädiktor bestimmt werden, der die beste Schätzung des zu codierenden Signals ermöglicht, so dass die Varianz des Differenzsignals zum Minimum wird (siehe Abschnitt 2.2). Da die im System entstehende Varianz des Quantisierungsfehlers der Varianz des Differenzsignals proportional ist, sollte gewährleistet sein, dass der Prädiktor immer optimal arbeitet. Ist der zu codierende Prozess entgegen der getroffenen Annahme nichtstationär, d. h. sind die statistischen Kenndaten zeitlichen Schwankungen unterworfen, so liefert der Prädiktor falsche Schätzwerte. Eine solche Situation ergibt sich bei der DPCM-Codierung von Sprachsignalen: Die Annahme, die Sprache sei ein stationärer Prozess, und dementsprechend könne der Prädiktor als zeitinvariantes Netzwerk dimensioniert werden, führt auf eine nichtoptimale Prädiktion. Eine Verbesserung ergibt sich dadurch, dass die Sprache in Segmente aufgeteilt wird, innerhalb derer die statistischen Eigenschaften relativ konstant sind, und dass dann der Prädiktor für jedes Segment neu optimiert wird. Dieses Verfahren der Quellencodierung nennen wir im folgenden adaptive Differenz-Pulscodemodulation (ADPCM). Ein adaptiver Algorithmus hat nicht nur den Vorteil eines erhöhten Signal-Rausch-Abstandes bei der Codierung von Sprachsignalen, sondern führt zusätzlich zu einer Verringerung des Ruhegeräusches gegenüber einer nichtadaptiven DPCM-Sprachcodierung; zudem ist das Verfahren durch die Anpassung an die statistischen Kenndaten des Eingangssignals datenkompatibel. Bei einer nichtadaptiven DPCM ist Datenkompatibilität nur erreichbar, wenn recht einschneidende Kompromisslösungen verwendet werden /1/.

2. Wichtige Grundlagen der nichtadaptiven DPCM

Im folgenden wird Sprache als stationärer Prozess vorausgesetzt; $s(k)$ sei eine zeitdiskrete Musterfolge dieses Prozesses mit der Varianz S .

2.1 Das Prinzip der DPCM /2,3,4/

In einem DPCM-System wird von jedem Eingangswert $s(k)$ ein Schätzwert $\hat{s}(k)$ subtrahiert; die Differenz $d(k) = s(k) - \hat{s}(k)$ wird quantisiert, so dass ein quantisierter Wert $\hat{u}(k)$ entsteht. Der Index der jeweils gewählten Quantisierungsstufe wird übertragen. Der Empfänger addiert zu

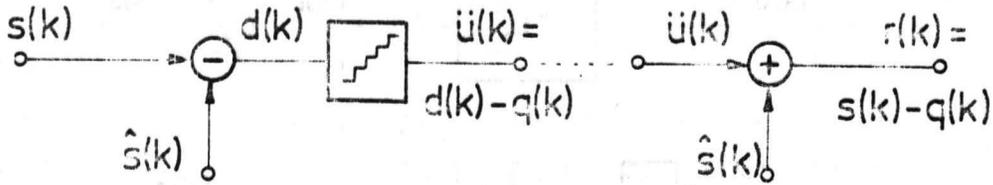


Bild 1 : Prinzip der Prädiktionsverfahren

dem übertragenen Wert $\tilde{u}(k)$ den gleichen Schätzwert $\hat{s}(k)$, der beim Sender vom Eingangswert subtrahiert wurde (Bild 1). Es entsteht das rekonstruierte Signal $r(k)$, das sich vom Eingangswert um den Quantisierungsfehler $q(k)$ unterscheidet, d. h. der Gesamtfehler ist wie bei einem PCM-System mit dem Quantisierungsfehler identisch:

$$q(k) = d(k) - \tilde{u}(k) = s(k) - r(k) \quad (2.1)$$

Wenn sich die Amplitudendichteverteilungen vom Eingangs- und Differenzsignal nicht sehr unterscheiden, dann wird die Leistung des Quantisierungsrauschens von der Leistung S_D des Differenzsignals bestimmt. Es lässt sich zeigen /4/, dass bei der DPCM-Quellencodierung von Sprachsignalen ein Gewinn

$$G = S / S_D \quad (2.2)$$

im Signal-Rausch-Verhältnis gegenüber einem vergleichbaren PCM-System entsteht. Die Höhe des Gewinns hängt von der Fähigkeit des Systems ab, einen möglichst guten Schätzwert $\hat{s}(k)$ für jeden einlaufenden Wert $s(k)$ zu finden. Dem Empfänger stehen zur Schätzung nur rekonstruierte Werte zur Verfügung. Es kommt zu einer Fehlerakkumulation, wenn die Schätzwerte vom Codierer (Sender) und Decodierer (Empfänger) nicht übereinstimmen. Um diese Übereinstimmung zu erreichen, wird auch beim Sender die Schätzung aus rekonstruierten Werten $r(j)$; $j = k - 1, k - 2, k - 3, \dots, k - N$, vorgenommen; dazu wird auch beim Sender aus der Folge der Werte $\tilde{u}(k)$ eine Folge rekonstruierter Werte $r(k)$ erzeugt (Bild 2). Dieses Verfahren führt dann zu der DPCM-Struktur gemäss Bild 3.

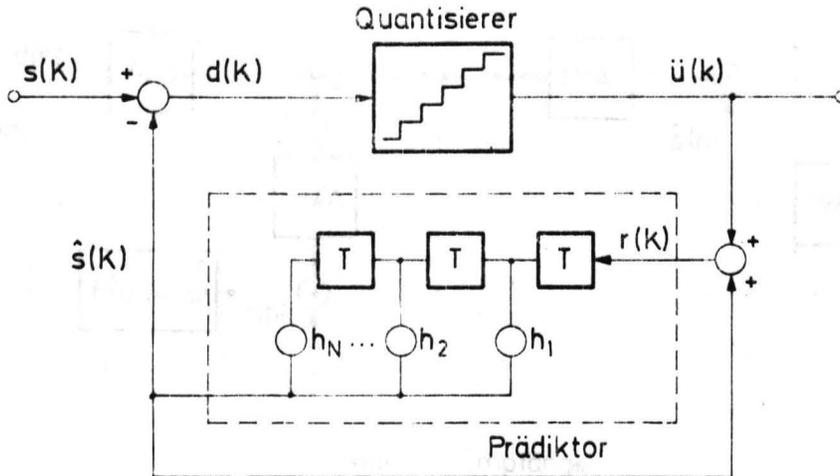


Bild 2 : Das Prinzip des DPCM - Codierers

Der Schätzwert $\hat{s}(\bullet)$ wird mit Hilfe eines nichtrekursiven Filters als Linearkombination aus rekonstruierten Werten $r(\bullet)$ gewonnen.

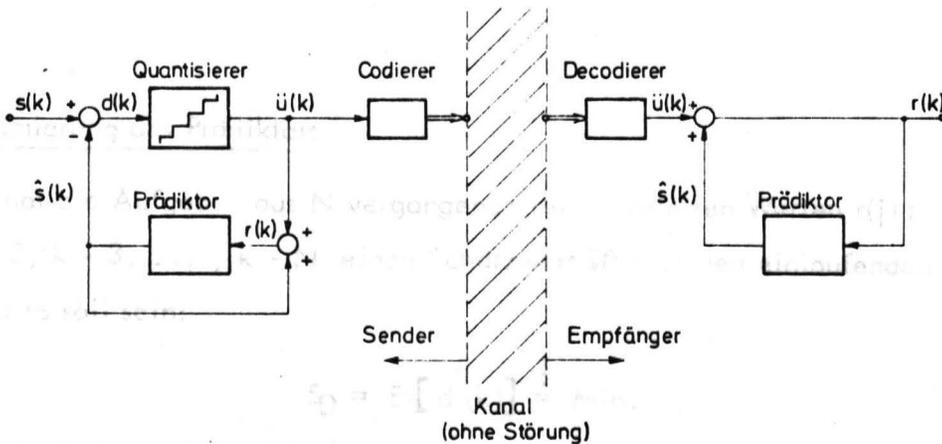


Bild 3 : DPCM - Signalübertragung

Die Schätzwerte $\hat{s}(\bullet)$ werden beim Sender und beim Empfänger aus rekonstruierten Werten $r(\bullet)$ ermittelt.

Die in dieser Darstellung vorgenommene Trennung von Quantisierer und Codierer tritt bei einer Realisierung des DPCM-Systems nicht auf. Quantisierer und Codierer sind dann zu einem A/D - Wandler zusammengefasst, der zu jedem einlaufenden Differenzwert $d(k)$ den zugehörigen Index der Quantisierungsstufe als Binärzahl ausgibt. Bild 4 zeigt die entsprechende Struktur.

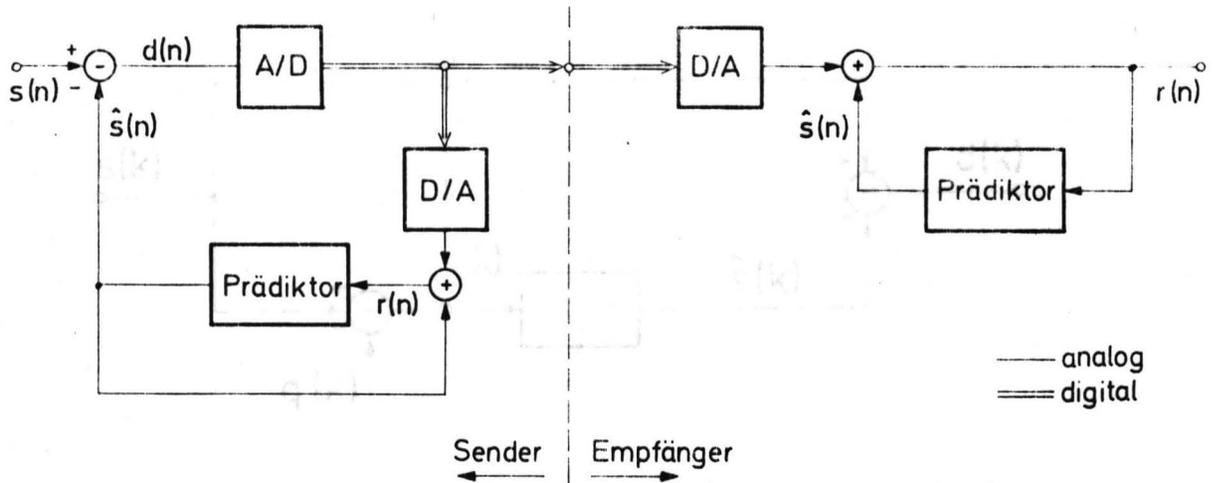


Bild 4: Realisierung einer DPCM - Signalübertragung

Die D/A-Wandlung beim Sender kann auch auf die $\hat{s}(\cdot)$ - Werte angewendet werden (vergl. Bild 10).

2.2 Die Optimierung des Prädiktors

Der Prädiktor hat die Aufgabe, aus N vergangenen rekonstruierten Werten $r(j)$; $j = k - 1, k - 2, k - 3, \dots, k - N$ einen Schätzwert $\hat{s}(k)$ für den einlaufenden Wert $s(k)$ zu bestimmen; es soll sein:

$$S_D = E [d^2(k)] \stackrel{!}{=} \text{Min.} \quad (2.3)$$

Es handelt sich bei diesem Minimierungsproblem um die Aufgabenstellung der gestörten Prädiktion aus der Theorie der Wiener-Kolmogoroff-Filterung /5/: $s(k)$ ist das gewünschte Signal, $r(k) = s(k) - q(k)$ das vorgegebene Signal; das Prädiktionsfilter ist ein nichtrekursives Filter der Länge N . Die Impulsantwort des Filters, $h(j) = h_j$; $j = 1, 2, 3, \dots, N$, ist so zu bestimmen, dass das Differenzsignal zum Minimum wird (Bild 5). Die optimale Bestimmung der Prädiktor-Impulsantwort (Koeffizientenvektor \underline{h} mit den Elementen $h(j) = h_j$) wird dadurch sehr erschwert, dass die Varianz Q_D des Messrauschens (also des Quantisierungsfehlers) nicht vorgegeben ist, sondern von der Varianz S_D des zu minimierenden Differenzsignals abhängt.

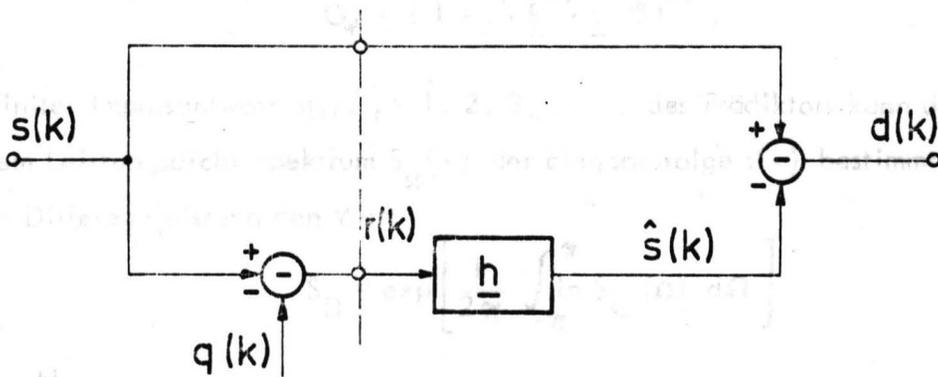


Bild 5 : Die Optimierung des Prädiktors als Aufgabenstellung der Wiener-Kolmogoroff-Filterung

Wird der Einfluss des Messrauschens vernachlässigt ($Q_D \ll S_D$), so ergibt sich ein optimaler Koeffizientenvektor \underline{h}_* aus dem Gleichungssystem

$$\underline{r} = \underline{R} \cdot \underline{h}_* \tag{2.4}$$

bzw.

$$\begin{bmatrix} R_{ss}(0) & R_{ss}(1) & R_{ss}(2) & \dots & R_{ss}(N-1) \\ R_{ss}(1) & R_{ss}(0) & R_{ss}(1) & \dots & R_{ss}(N-2) \\ R_{ss}(2) & R_{ss}(1) & R_{ss}(0) & \dots & R_{ss}(N-3) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ R_{ss}(N-1) & R_{ss}(N-2) & R_{ss}(N-3) & \dots & R_{ss}(0) \end{bmatrix} \times \begin{bmatrix} h_{1*} \\ h_{2*} \\ h_{3*} \\ \vdots \\ h_{N*} \end{bmatrix} = \begin{bmatrix} R_{ss}(1) \\ R_{ss}(2) \\ R_{ss}(3) \\ \vdots \\ R_{ss}(N) \end{bmatrix}$$

In dieser Gleichung ist \underline{R} die Kovarianzmatrix der Eingangsfolge, d. h., das Element (i, j) hat den AKF-Wert $R_{ss}(i-j) = E[s(i) \cdot s(j)]$, und der Vektor \underline{r} hat die Komponenten $R_{ss}(j)$; $j = 1, 2, 3, \dots, N$. Die AKF-Matrix ist streifensymmetrisch (Töplitz-Form); sie ist positiv definit, wenn die AKF-Werte linear unabhängig sind. Damit kann der suboptimale Koeffizientensatz \underline{h}_* berechnet werden:

$$\underline{h}_* = \underline{R}^{-1} \cdot \underline{r} \tag{2.5}$$

Der maximal erreichbare Gewinn ergibt sich zu

$$G_* = (1 - \underline{r}^T \cdot \underline{R}^{-1} \cdot \underline{r} / S)^{-1} \quad (2.6)$$

Bei semi-infiniten Impulsantwort $h(j)$; $j = 1, 2, 3, \dots$ des Prädiktors kann der Gewinn auch direkt aus dem Leistungsdichtespektrum $S_{ss}(\cdot)$ der Eingangsfolge $s(\cdot)$ bestimmt werden, da die minimale Differenzleistung den Wert

$$S_D = \exp \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln S_{ss}(\Omega) d\Omega \right] \quad (2.7)$$

annimmt /5/. Mit

$$S = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_{ss}(\Omega) d\Omega \quad (2.8)$$

ergibt sich also

$$G_* = \frac{\frac{1}{2\pi} \int_{-\pi}^{\pi} S_{ss}(\Omega) d\Omega}{\exp \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln S_{ss}(\Omega) d\Omega \right]} \quad (2.9)$$

Diese Beziehung gilt auch bei einem Prädiktor mit einer Impulsantwort der Länge N , wenn der stationäre Prozess $\{s(n)\}$ ein Markoff-Prozess der Ordnung $\leq N$ ist /5/.

2.3 Ergebnisse für eine mittlere Sprach-AKF

Der Gewinn G_* wird von einem DPCM-System nicht erreicht, da die Prädiktion mit einem Messrauschen behaftet ist; die Abnahme des Gewinns hängt von der Varianz Q_D dieses Rauschens und damit vor allem von der Zahl der Quantisierungsstufen ab. Dieser Einfluss soll an einem

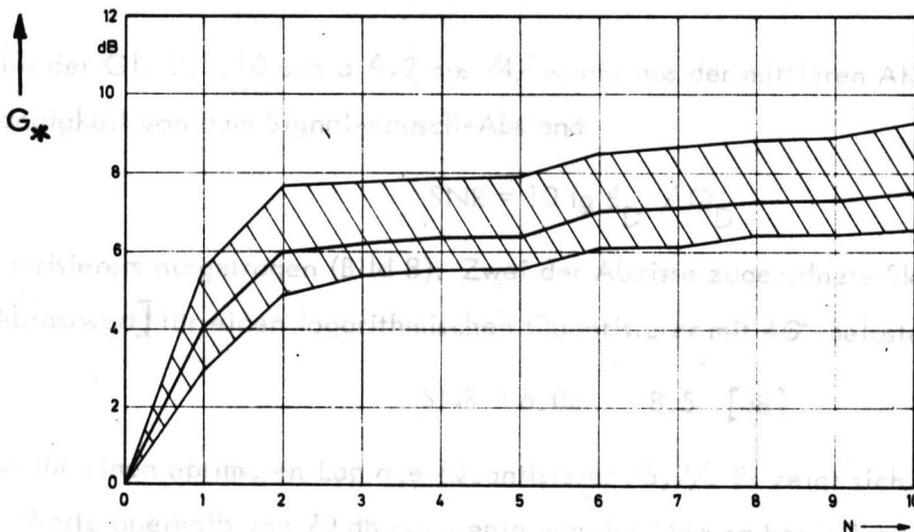


Bild 6: Die Abhängigkeit des Gewinn G_* von der Zahl N der Prädiktorkoeffizienten

Beispiel erläutert werden ; in Bild 6 ist der Gewinn G_* in Abhängigkeit von der Länge N der Impulsantwort des Prädiktors aufgetragen. Der schraffierte Bereich zeigt die Schwankungen für eine grössere Zahl von Sprachproben verschiedener Sprecher (Frequenzbereich der Sprachproben 300 - 3400 Hz = BP-Sprache). Der in dem schraffierten Bereich liegende Linienzug ist der Gewinnverlauf für eine mittlere Autokorrelationsfolge bei BP-Sprache ; Bild 7 zeigt dazu die normierten AKF-Verläufe.

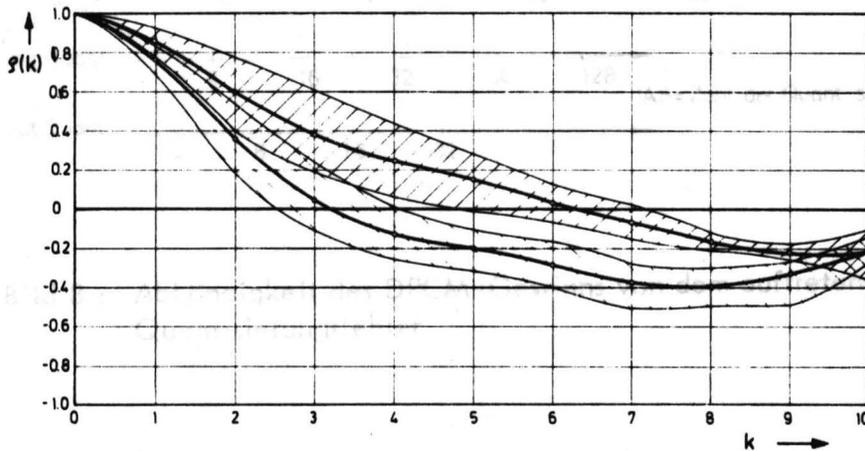


Bild 7 : Verlauf der normierten AKF $g(k) = R_{ss}(k) / R_{ss}(0)$

Oberer schraffierter Bereich : TP-Sprache (0-3400 Hz)

Unterer schraffierter Bereich : BP-Sprache (300-3400 Hz)

Mit Hilfe der Gl. 3.3.10 und 3.4.2 aus /4/ wurde aus der mittleren AKF der Gewinn G in Abhängigkeit von dem Signal-Rausch-Abstand

$$SNR = 10 \lg S_D / Q_D$$

des Quantisierers aufgetragen (Bild 8). Zwei der Abzisse zugeordnete Skalen zeigen die Zahl n [bit/Abtastwert] für einen logarithmischen Quantisierer mit 4 σ -Belastung, für den

$$SNR = 6.02 - 8.5 \text{ [dB]} \quad (2.10)$$

gilt, und für einen optimalen Laplace-Quantisierer /6,7/. Es zeigt sich, dass die Prädiktion für SNR-Werte oberhalb von 20 dB nur wenig von der Störung beeinflusst wird ; weiterhin ist erkennbar, dass die Empfindlichkeit gegenüber Messrauschen mit wachsender Länge N der Im-

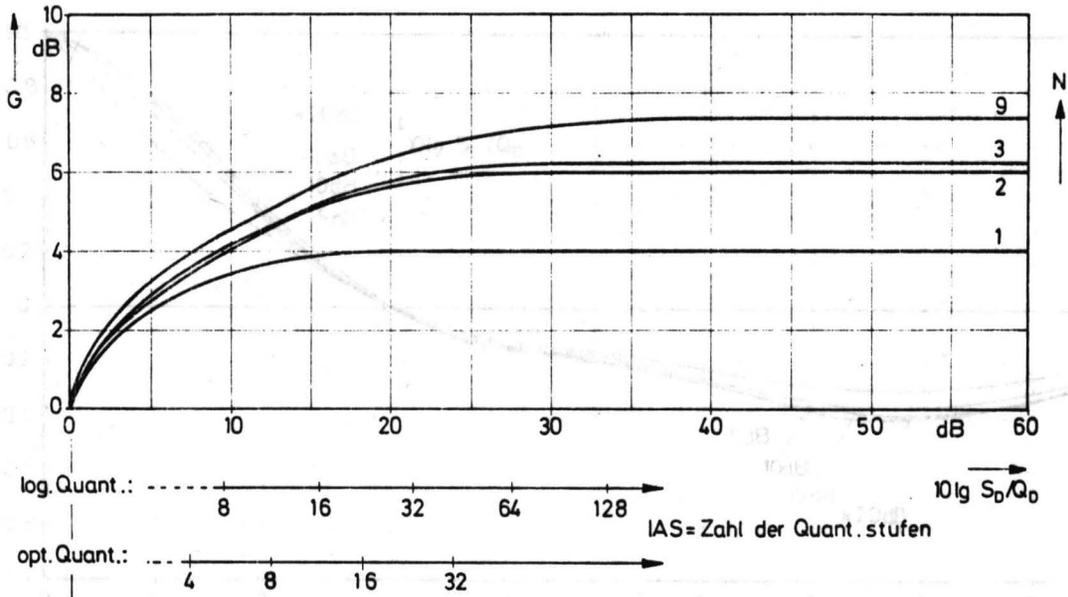


Bild 8 : Abhängigkeit des DPCM-Gewinns von dem auftretenden Quantisierungsfehler

pulsantwort des Prädiktors zunimmt. Die Gewinnabnahme kann auch so gedeutet werden, dass eine ungestörte Prädiktion vorgenommen wird, wobei aber nur noch ein stochastischer Prozess mit reduzierten AKF-Werten zur Verfügung steht ; Bild 9 zeigt diese Reduktion der AKF für verschiedene SNR - Werte. Untersuchungen haben gezeigt, dass der Gewinn nur geringfügig erhöht werden kann, wenn statt des Koeffizientenvektors h_* ein Vektor h verwendet wird, bei dem die Rauscheinflüsse mit berücksichtigt sind. Die mögliche Gewinnerhöhung ΔG lässt sich durch

$$0 \leq \Delta G \leq G_* + 10 \lg \sum_{i=1}^N h_{i*}^2 - \text{SNR} \quad [\text{dB}] \quad (2.11)$$

eingrenzen/8/.

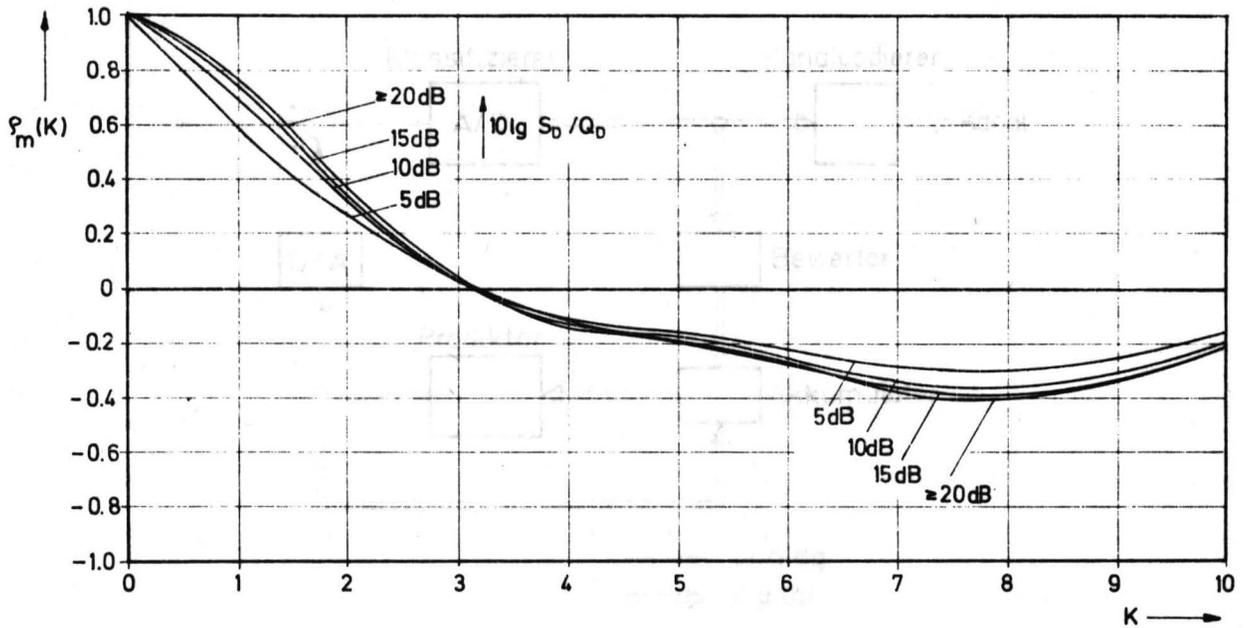


Bild 9 : Abhängigkeit der modifizierten Autokorrelationsfolge $\rho_m(\cdot)$ von dem Signal-Rausch-Abstand des in der DPCM-Schleife verwendeten Quantisierers.

Die Prädiktion ist sehr unempfindlich gegenüber kleinen Änderungen der Koeffizienten, die ja auch als Rauschüberlagerung gedeutet werden können. Diese Unempfindlichkeit ermöglicht die Verwendung von Prädiktor-Koeffizienten, die aus einer mittleren AKF bestimmt wurden, so dass der DPCM - Codierer sprecherunabhängig dimensioniert werden kann. Die Unempfindlichkeit ermöglicht weiterhin eine ziemlich grobe Quantisierung der Koeffizienten bei einer digitalen Realisierung des Prädiktors /3/; Bild 10 zeigt eine Ausführungsform. Auch hier gilt, dass die Empfindlichkeit mit wachsendem N zunimmt. Bild 11 zeigt den Gewinn in Abhängigkeit von der Genauigkeit der Darstellung der Prädiktorkoeffizienten (der Gewinn wurde für eine Sprachprobe gemessen; Bild 11 ist eine korrigierte Fassung von Bild 14 in der Veröffentlichung /3/.)

11: Gewinn in Abhängigkeit von der Genauigkeit der Darstellung der Prädiktorkoeffizienten

Die Anzahl an Nachkommastellen für die Darstellung der Prädiktorkoeffizienten wurde von 16 auf 10 reduziert. Zusätzlich werden 3 Vorkommalstellen benötigt.

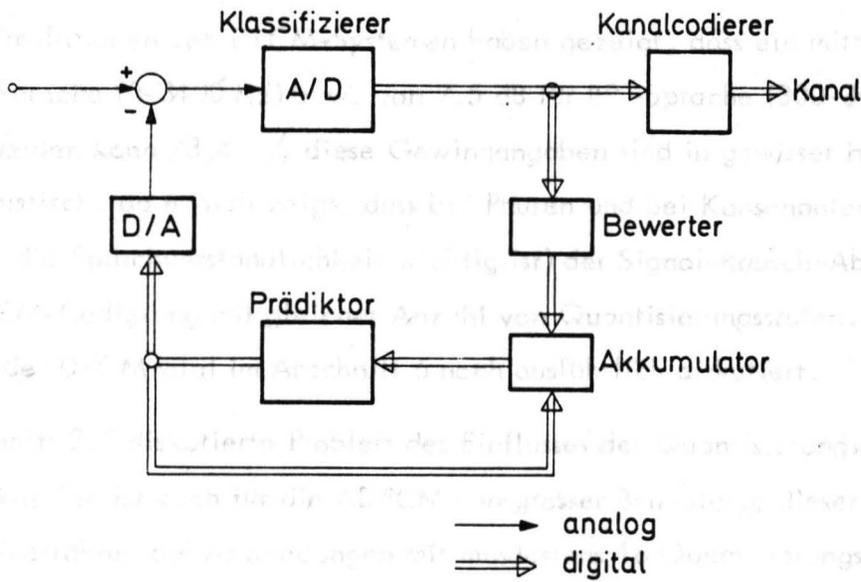


Bild 10 : Digitale Realisierung der Prädiktion

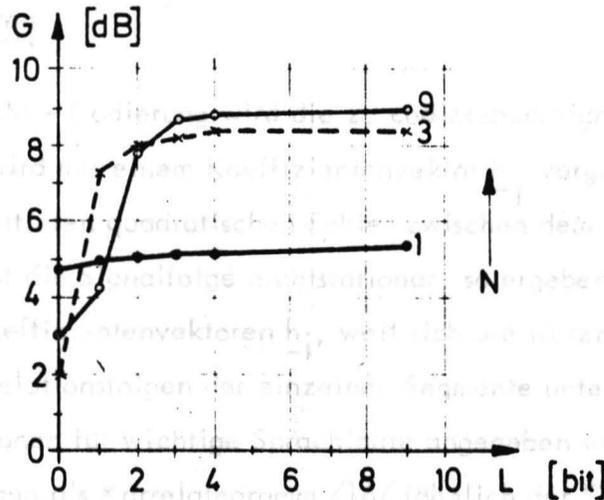


Bild 11: Gewinn in Abhängigkeit von der Genauigkeit der Darstellung des Prädiktorkoeffizienten

L = Anzahl der Nachkomma-bit. Zusätzlich werden 3 Vorkomma-bit benötigt.

2.4 Zusammenfassung

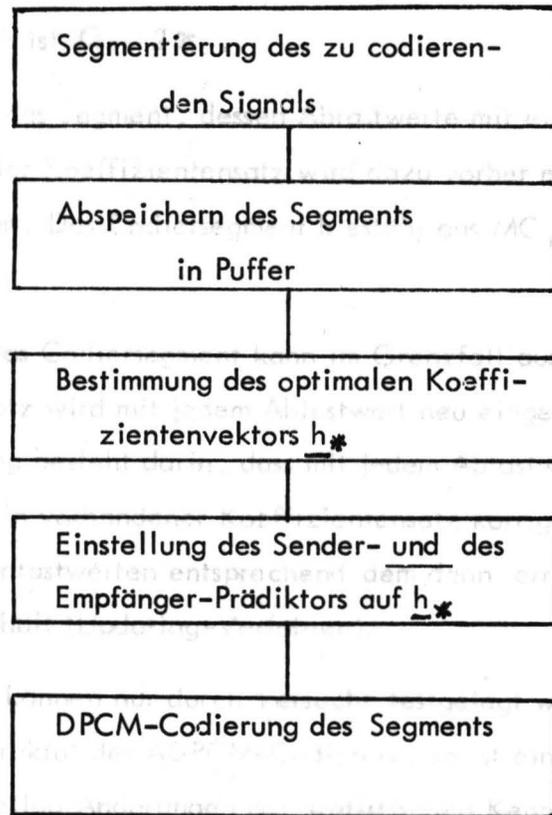
Ausführliche Simulationen von DPCM-Systemen haben gezeigt, dass ein mittlerer Gewinn von 8.5 dB für TP-Sprache (0-3400 Hz) bzw. von 7.5 dB für BP-Sprache (300-3400 Hz) nicht überschritten werden kann /3,4 /; diese Gewinnangaben sind in gewisser Hinsicht noch ein wenig zu optimistisch, da es sich zeigt, dass bei Pausen und bei Konsonanten (deren genaue Wiedergabe für die Sprachverständlichkeit wichtig ist) der Signal-Rausch-Abstand geringer ist als bei einer PCM-Codierung mit gleicher Anzahl von Quantisierungsstufen. Diese unerwünschte Eigenschaft der DPCM wird im Abschnitt 6 noch ausführlich diskutiert.

Das im Abschnitt 2.3 diskutierte Problem des Einflusses des Quantisierungsrauschens auf die Güte der Prädiktion ist auch für die ADPCM von grosser Bedeutung; dieser Einfluss begrenzt das ADPCM-Verfahren auf Anwendungen mit mindestens 16 Quantisierungsstufen. Die gezeigte Unempfindlichkeit gegenüber Ungenauigkeiten der Prädiktorkoeffizienten-Darstellung gilt auch für die ADPCM-Codierung; damit ist eine effiziente digitale Übertragung der Koeffizientenvektoren zum Empfänger gewährleistet.

3. ADPCM - Strukturen

Bei der adaptiven DPCM - Codierung wird die zu codierende Signalfolge in Segmente eingeteilt; die Prädiktion wird mit einem Koeffizientenvektor \underline{h}_j vorgenommen, der für das Segment Nr. j den kleinsten mittleren quadratischen Fehler zwischen dem wahren Wert und dem Schätzwert liefert /9-14/. Ist die Signalfolge nichtstationär, so ergeben sich für die einzelnen Segmente unterschiedliche Koeffizientenvektoren \underline{h}_j , weil sich die Kurzzeit-Leistungsdichtespektren bzw. die Kurzzeit-Autokorrelationsfolgen der einzelnen Segmente unterscheiden. In /15/ sind bereits Autokorrelationsfunktionen für wichtige Sprachlaute angegeben worden; die starke Zeitveränderlichkeit der AKF kann als Korrelatogramm /16/ (ähnlich der "Visible Speech" für Leistungsdichten) oder mit Hilfe eines Films /17/ aufgezeichnet werden. Eine andere Darstellung zeigt Bild 40.

Der für jedes Segment optimale Koeffizientenvektor kann näherungsweise nach verschiedenen Methoden bestimmt werden; sie werden im 4. Abschnitt dieser Arbeit diskutiert. Der ADPCM-Codierer hat die folgenden Aufgaben zu erfüllen:



Für die hierangedeuteten Aufgaben gibt es verschiedene Lösungsmöglichkeiten, die im folgenden diskutiert werden sollen.

3.1 Segmentierung des Sprachsignals

Es gibt verschiedene Möglichkeiten, eine Segmentierung des Sprachsignals vorzunehmen. Ziel muss es sein, die Segmente so zu wählen, dass mit den Segmentdaten ein Koeffizientenvektor h_* berechnet werden kann, der eine möglichst gute Vorhersage der Segmentdaten ermöglicht. Ist das Segment zu lang, so können sich die statistischen Eigenschaften bereits zu stark geändert haben; ist das Segment zu kurz, so wird das Verfahren dann ineffektiv, wenn der Koeffizientensatz zusätzlich zu dem codierten Differenzsignal zum Empfänger übertragen werden muss.

Es müssen zwei Arten von Segmenten unterschieden werden:

- a) das Analysesegment ist das Segment, das zur Bestimmung des optimalen Koeffizientenvektors herangezogen wird; es besteht aus MA Abtastwerten, hat also die Länge MA - 1. Der Abtastabstand wird grundsätzlich zu $T = 1$ angesetzt; die Kreisfrequenz wird mit Ω bezeichnet;

der Periodizitätsbereich ist $\Omega_0 = 2\pi$.

b) das Codiersegment ist das Segment, dessen Abtastwerte mit einem festen Koeffizientensatz codiert werden; der Koeffizientensatz wird dazu vorher mit Hilfe des Analysesegments optimal bestimmt. Das Codiersegment b) besteht aus MC Abtastwerten, hat also die Länge MC - 1.

Es gilt stets $MA \geq MC$. Das Codiersegment kann im Grenzfall aus einem Abtastwert bestehen, d. h. der Koeffizientensatz wird mit jedem Abtastwert neu eingestellt. Eine weitere Möglichkeit der Segmentierung besteht darin, dass mit jedem Abtastwert ein neuer Koeffizientensatz berechnet bzw. ein vorhandener Koeffizientensatz korrigiert wird, dass aber der Prädiktor erst nach MC Abtastwerten entsprechend dem dann erreichten Koeffizientensatz eine neue Einstellung erhält (Updating-Verfahren).

Die Grössen MA und MC können nur durch Versuche festgelegt werden; sie stehen in engem Zusammenhang mit der Struktur des ADPCM-Codierers; so ist einerseits ein kleiner MA-Wert erwünscht, um auch schnellen Änderungen der statistischen Kenndaten folgen zu können, andererseits sollte MC (und damit wegen $MA \geq MC$ auch MA) möglichst gross sein, wenn die geänderten Koeffizientensätze h jeweils dem Empfänger mitgeteilt werden müssen.

Aus Gründen der praktischen Realisierung ist es sinnvoll, MA und MC konstant zu lassen. Es zeigt sich auch, dass es sich nicht lohnt, z. B. bei stimmhaften Lauten eine Segmentierung gemäss der Sprachgrundfrequenz vorzunehmen (pitchsynchrone Segmentierung); der Gewinn nimmt nur unwesentlich zu, wenn der optimale Koeffizientenvektor aus den Daten des Segmentes (bei Ausschluss der Pitchanregung) bestimmt wird, und die Zunahme steht in keinem vernünftigen Verhältnis zu dem erheblichen Mehraufwand, der durch die zusätzliche Pitchbestimmung entsteht (vgl. aber Abschnitt 7). Eine grössere Gewinnzunahme ist allerdings dann möglich, wenn die pitchsynchrone Segmentierung bei dem ADPCM-Verfahren dazu benutzt wird, eine zusätzliche Prädiktion so vorzunehmen, dass die einen Pitchabstand zurückliegende Signalfolge zusätzlich zur Vorhersage herangezogen wird /9,10/. Ein Vergleich mit den in dieser Arbeit genannten Gewinnen ist nicht möglich, da in den zitierten Arbeiten nur subjektive Gewinne genannt wurden.

3.2 ADPCM-Struktur bei vorlaufender Analyse

Die Lage des Codiersegmentes bezüglich des Analysesegments ist für die Struktur des ADPCM-Codierers massgebend.

Überdeckt das Analysesegment das Codiersegment, so wird eine vorlaufende Analyse vorgenommen. Das Codiersegment ist gegenüber dem Analysesegment um $MV - 1$ Einheiten verschoben; die Segmente müssen nicht symmetrisch zueinander liegen. Ist $MA = MC$, so gilt:

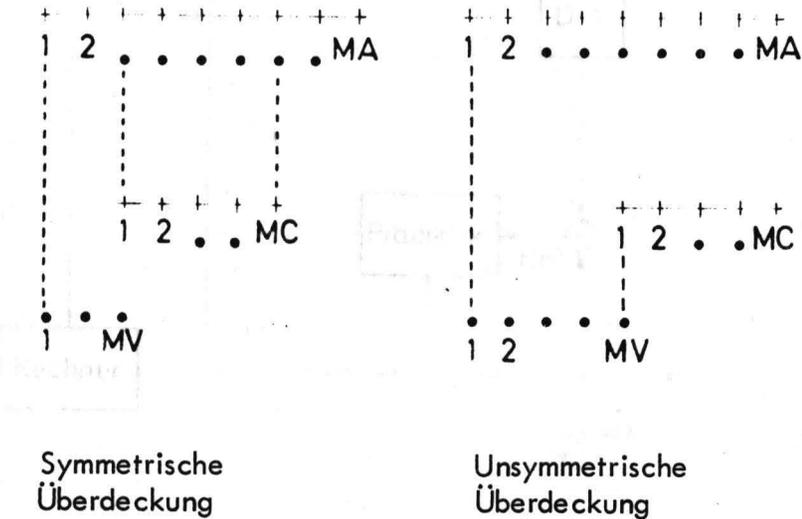


Bild 12 : Lage des Codiersegments bei symmetrischer und unsymmetrischer Überdeckung (vorlaufende Analyse).

$MV = 1$. Wesentlich bei der Verwendung der überdeckenden Segmente ist die Tatsache, dass zur Bestimmung des optimalen Koeffizientenvektors die zu codierenden Abtastwerte und eventuell auch noch zukünftige Abtastwerte bekannt sein müssen. Eine solche Kenntnis ist durch eine Pufferspeicherung erreichbar: die Abtastwerte laufen in einen Durchlaufspeicher (Puffer); sind MA Werte des Segments Nr. j gespeichert, so wird der Vektor \underline{h}_{j*} berechnet, der Prädiktor entsprechend eingestellt und die Codierung der Abtastwerte des Codiersegments vorgenommen. Wird die Länge des Durchlaufspeichers grösser gewählt als die des Analysesegments, so kann erreicht werden, dass genügend Zeit zur Bestimmung der optimalen Koeffizienten zur Verfügung steht. Die Länge des Durchlaufspeichers ist durch die für das Telefonsystem maximal zulässige Verzögerung begrenzt (für Satellitenverbindungen werden max. etwa 400 ms zugelassen). Es ergibt sich damit die Struktur nach Bild 13 und Bild 14.

Die Funktion wird direkt aus den Blockschaltbildern ersichtlich: die den Puffer verlassenden Abtastwerte werden DPCM-codiert, wobei der Prädiktor denjenigen Koeffizientensatz \underline{h}_{j*} verwendet, der für das entsprechende Segment von Abtastwerten optimal ist. Da \underline{h}_{j*} aus den

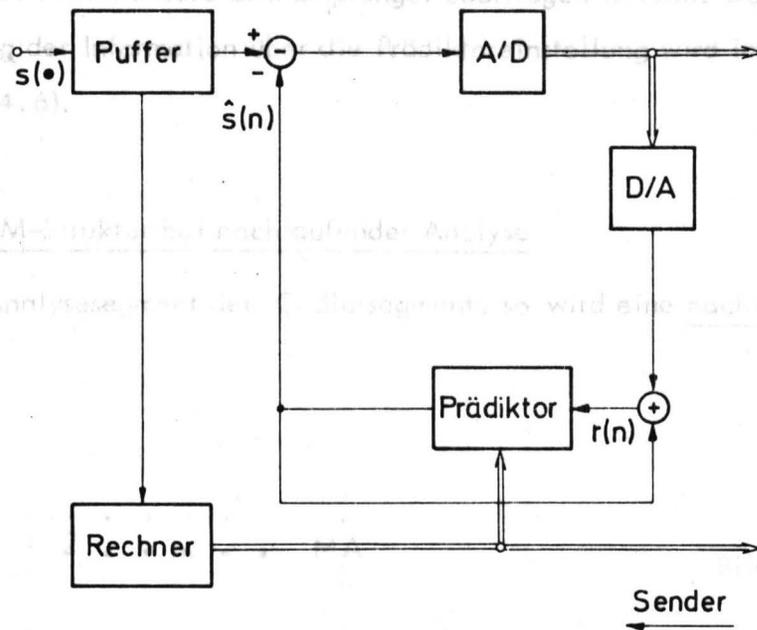


Bild 13 : ADPCM - Codierer bei vorlaufender Analyse

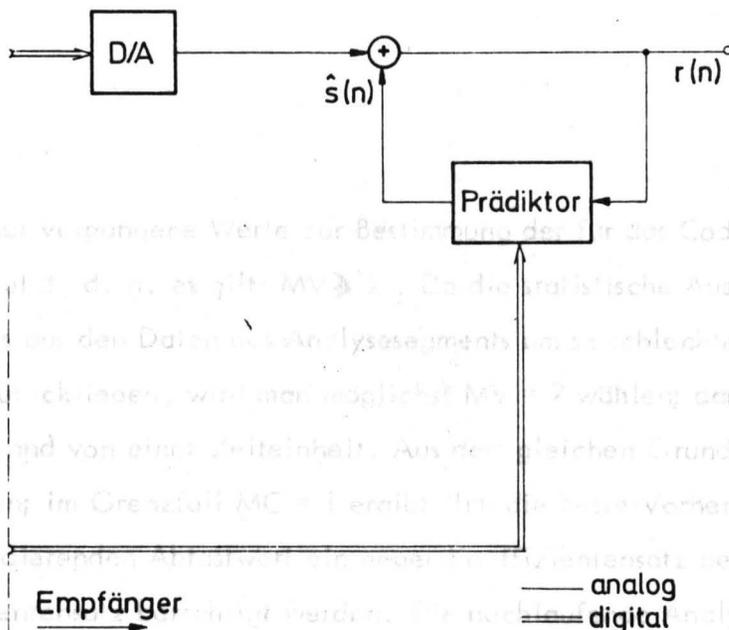


Bild 14: ADPCM - Decodierer bei vorlaufender Analyse

Daten des Eingangssignals gewonnen wird, die dem Empfänger nicht zur Verfügung stehen, muss der Koeffizientensatz zum Empfänger übertragen werden. Das Problem der digitalen Übertragung der Information über die Prädiktoreinstellung wird im folgenden noch diskutiert (Abschnitt 4.6).

3.3 ADPCM-Struktur bei nachlaufender Analyse

Folgt das Analysesegment dem Codiersegment, so wird eine nachlaufende Analyse vorgenommen:

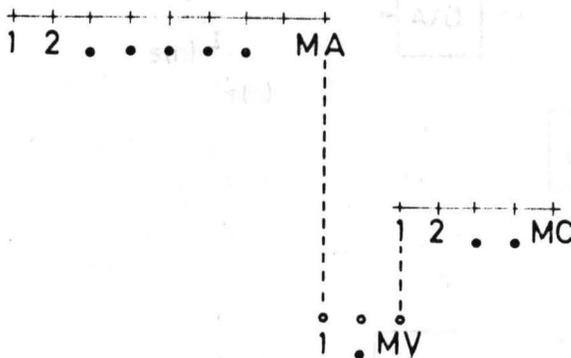


Bild 15: Lage des Analyse- und Codiersegments bei nachlaufender Analyse

Es werden nur vergangene Werte zur Bestimmung der für das Codiersegment optimalen Koeffizienten benutzt, d. h. es gilt: $MV \geq 2$. Da die statistische Aussage über die Daten des Codiersegments aus den Daten des Analysesegments um so schlechter sein wird, je weiter die Analysedaten zurückliegen, wird man möglichst $MV = 2$ wählen; damit liegt zwischen den Segmenten ein Abstand von einer Zeiteinheit. Aus dem gleichen Grund sollte MC nicht zu gross gewählt werden; im Grenzfall $MC = 1$ ergibt sich die beste Vorhersage; dazu muss allerdings für jeden zu codierenden Abtastwert ein neuer Koeffizientensatz berechnet bzw. der vorhergehende Koeffizientensatz berichtigt werden. Die nachlaufende Analyse kann ohne Verzögerung des zu codierenden Signals durchgeführt werden.

Zu den Verfahren der nachlaufenden Analyse gehören die verschiedenen Gradientenverfahren, wie sie aus der Technik der adaptiven Kanalentzerrung /18,19/ und der adaptiven Echokompensation /20/ bekannt sind. Die Struktur bei der nachlaufenden Analyse ist dadurch ausge-

zeichnet, dass der Koeffizientensatz h nicht zum Empfänger übertragen werden muss, da er aus den rekonstruierten Werten $r(n)$ am Eingang des Prädiktors gewonnen wird, die - bei störungsfreier Übertragung - beim Sender und Empfänger gleichermassen zur Verfügung stehen. Es ist daher bei dieser Struktur möglich, den Koeffizientensatz mit jedem Abtastwert zu ändern.

Für die nachlaufende Analyse ergibt sich somit die Struktur nach Bild 16 und Bild 17:

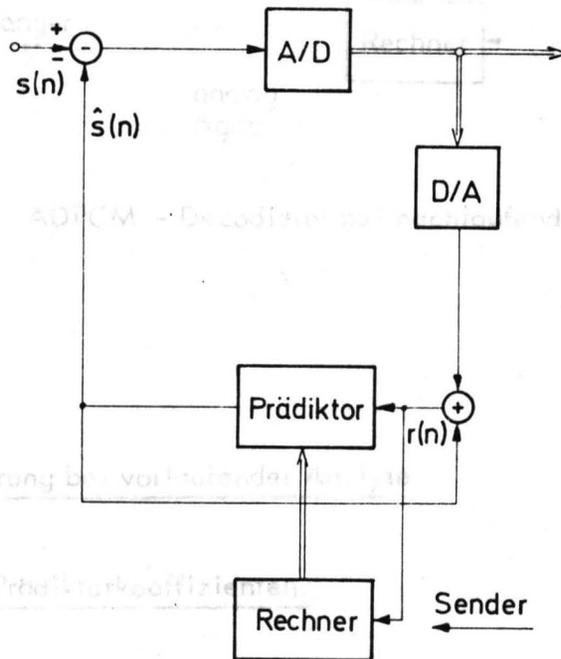


Bild 16 : ADPCM - Codierer bei nachlaufender Analyse

$$\hat{s}(n) = \sum_{i=1}^N h_i s(n-i)$$

Gesucht sind die Prädiktorkoeffizienten $h_i, i = 1, 2, 3, \dots, N$, die die Energie des Differenzsignals, das bei der Prädiktion der Abtastwerte des Codiersegments entsteht, zum Mini-

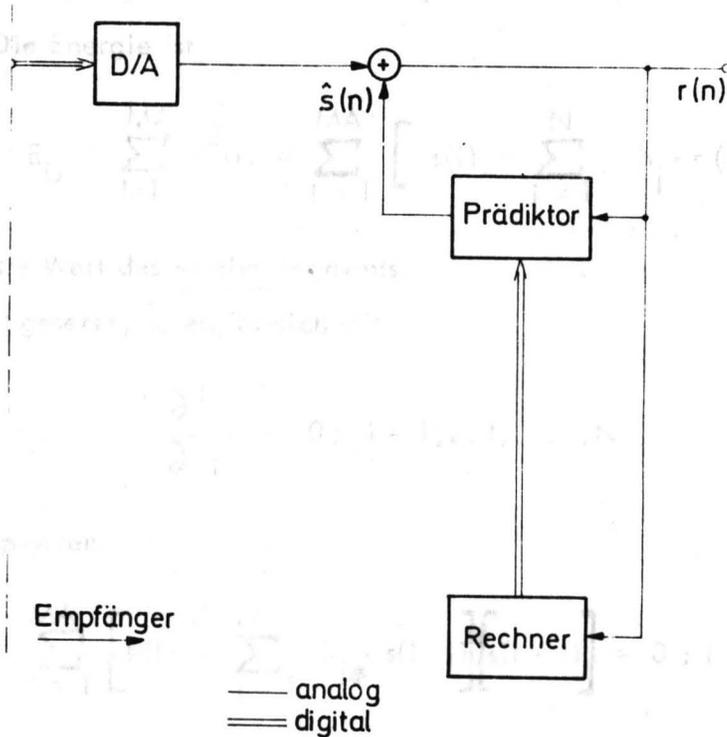


Bild 17: ADPCM - Decodierer bei nachlaufender Analyse

4. ADPCM - Codierung bei vorlaufender Analyse

4.1 Bestimmung der Prädiktorkoeffizienten

4.1.1 Deterministische Formulierung des Problems

Für jedes Segment Nr. k muss der optimale Koeffizientenvektor \underline{h}_k aus MA Werten $s(n)$ der Eingangsfolge bestimmt werden. Wie im Abschnitt 2.2 bereits gezeigt wurde, ist der suboptimale Koeffizientenvektor \underline{h}_{k*} , der ohne Berücksichtigung des Quantisierungsrauschens bestimmt wird, für die Prädiktion ausreichend genau. Im folgenden wird der die Segmentnummer bezeichnende Index weggelassen. Der Schätzwert $s(n)$ ergibt sich als Linearkombination

$$\hat{s}(n) = \sum_{i=1}^N h_i \cdot r(n-i)$$

Gesucht sind die Prädiktorkoeffizienten h_i ; $i=1,2,3,\dots,N$, die die Energie des Differenzsignals, das bei der Prädiktion der Abtastwerte des Codiersegments entsteht, zum Mini-

mum macht. Die Energie ist

$$E_D = \sum_{l=1}^{MA} d^2(l) = \sum_{l=1}^{MA} \left[s(l) - \sum_{i=1}^N h_i \cdot r(l-i) \right]^2 \quad (4.1)$$

$s(1)$ ist der erste Wert des Analysesegments.

Wird $r(\bullet) \approx s(\bullet)$ gesetzt, so ergibt sich mit

$$\frac{\partial E_D}{\partial h_i} = 0; \quad i = 1, 2, 3, \dots, N \quad (4.2)$$

das Gleichungssystem

$$\sum_{l=1}^{MA} \left[s(l) - \sum_{i=1}^N h_{i*} \cdot s(l-i) \right] \left[s(l-i) \right] = 0; \quad i = 1, 2, 3, \dots, N$$

Der bei optimaler Prädiktion entstehende Fehler $d(l)$ ist also orthogonal zu den zur Prädiktion herangezogenen Daten $s(l-i); i = 1, 2, 3, \dots, N$.

Das Gleichungssystem führt auf

$$\left\| \sum_{i=1}^N h_{i*} \cdot R(i, j) = R(0, j); \quad j = 1, 2, 3, \dots, N \right. \quad (4.3)$$

wobei

$$\left\| R(i, j) = \sum_{l=1}^{MA} s(l-i) \cdot s(l-j) \right. \quad (4.4)$$

gesetzt wurde.

Die bei optimaler Prädiktion auftretende Energie ist

$$\begin{aligned} E_{D*} &= \sum_{l=1}^{MA} \left[s(l) - \sum_{i=1}^N h_{i*} \cdot s(l-i) \right]^2 \\ &= \sum_{l=1}^{MA} s^2(l) - 2 \sum_{i=1}^N h_{i*} \cdot \sum_{l=1}^{MA} s(l) \cdot s(l-i) \\ &\quad + \sum_{i=1}^N h_{i*} \cdot \sum_{i=1}^N h_{i*} \cdot \sum_{l=1}^{MA} s(l-i) \cdot s(l-i) \\ &= R(0, 0) - 2 \sum_{i=1}^N h_{i*} \cdot R(0, i) + \sum_{i=1}^N h_{i*} \cdot \overbrace{\sum_{i=1}^N h_{i*} \cdot R(i, i)}^{R(0, j)} \end{aligned} \quad (4.6)$$

$$E_{D*} = R(0,0) - \sum_{i=1}^N h_{i*} \cdot R(0,i) \quad (4.5)$$

Ist $MA = MC$, so ergibt sich für das Segment ein Prädiktionsgewinn

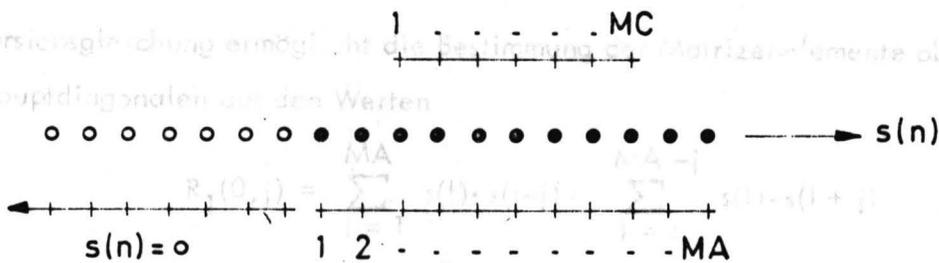
$$G_* = \frac{\sum_{l=1}^{MA} s^2(l)}{E_{D*}} = \frac{R(0,0)}{E_{D*}} = \frac{1}{1 - \sum_{i=1}^N h_{i*} \cdot R(0,i) / R(0,0)} \quad (4.7)$$

Die Grösse der für die Minimierung der Energie der Differenzsignalfolge ausschliesslich massgebenden Funktion $R(\cdot, \cdot)$ ist von der Problemstellung abhängig.

Möglichkeit 1:

Für die h_{i*} - Berechnung stehen ausschliesslich die MA Daten des Analysesegments zur Verfügung: $s(n); n = 1, 2, 3, \dots, MA$. Weiter zurückliegende Daten $s(n); n \leq 0$ sind unbekannt und werden für die $R(\cdot, \cdot)$ - Bestimmung zu Null gesetzt.

$$R_1(i+1, j+1) = R_1(i, j) - s(MA-i) \cdot s(MA-j) \quad (4.8)$$



Aus der Bestimmungsgleichung

Die Matrix mit den Elementen $R_1(i, j); i, j = 1, 2, 3, \dots, N$ ist also symmetrisch, besitzt nur Zeile-Töplitz-Form

$$\sum_{i=1}^N h_{i*} \cdot R_1(i, j) = R_1(0, j) \quad i = 1, 2, 3, \dots, N \quad (4.6)$$

lassen sich die Koeffizienten $h_{i*}; i = 1, 2, 3, \dots, N$ bestimmen.

Für die R_1 - Werte gilt mit $s(n) = 0; n \leq 0$:

$$\begin{aligned} R_1(i, j) &= \sum_{l=1}^{MA} s(l-i) \cdot s(l-j) ; i, j = 0, 1, 2, \dots, N \\ &= \sum_{l=1}^{MA-i} s(l) \cdot s(l+i-j) \end{aligned} \quad (4.7)$$

mit folgenden Eigenschaften:

Symmetrie : wie aus der Bestimmungsgleichung abzulesen ist, gilt:

$$R_1(i, j) = R_1(j, i)$$

Rekursivität:

$$\begin{aligned} R_1(i+1, j+1) &= \sum_{l=1}^{MA} s(l-i-1) \cdot s(l-j-1) \\ &= R_1(i, j) + s(-i) \cdot s(-j) - s(MA-i) \cdot s(MA-i) \end{aligned}$$

Das Produkt $s(-i) \cdot s(-j)$ ist nicht bekannt und wird daher zu Null gesetzt. Damit gilt:

$$R_1(i+1, j+1) = R_1(i, j) - s(MA-i) \cdot s(MA-i) \quad (4.8)$$

Diese Rekursionsgleichung ermöglicht die Bestimmung der Matrizenelemente ober- und unterhalb der Hauptdiagonalen aus den Werten

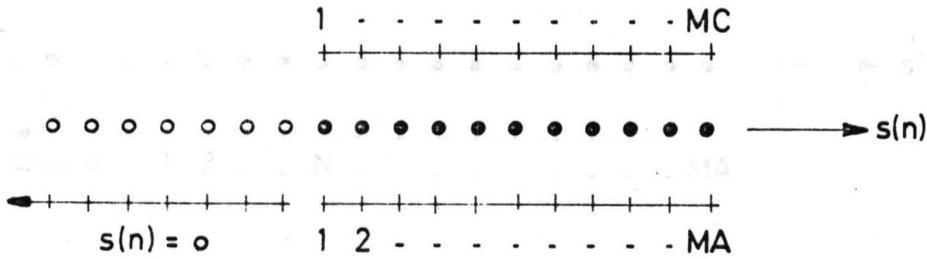
$$R_1(0, j) = \sum_{l=1}^{MA} s(l) \cdot s(l-j) = \sum_{l=1}^{MA-j} s(l) \cdot s(l+j)$$

und der Werte der Hauptdiagonalen aus dem ersten Element

$$R_1(1, 1) = \sum_{l=1}^{MA} s^2(l-1) = \sum_{l=1}^{MA-1} s^2(l) = R_1(0, 0) - s^2(MA)$$

Die Matrix mit den Elementen $R_1(i, j); i, j = 1, 2, 3, \dots, N$ ist also symmetrisch, besitzt aber keine Töplitz-Form.

Ist $MA = MC$ (Analysesegment = Codiersegment), so kann der Prädiktor bei der Codierung der ersten Abtastwerte des Segments keine Linearkombination über N Schritte vornehmen, da der Prädiktor erst nach und nach voll ausgenutzt wird. Der erste Wert eines jeden Seg-



ments wird also PCM-codiert, der zweite Wert wird mittels des ersten vorhergesagt, etc. Erst der $(N + 1)$ te Wert kann mit einer Linearkombination von N im Prädiktor gespeicherten Werten vorhergesagt werden. Die aus der Bestimmungsgleichung errechneten Koeffizienten berücksichtigen diese anfangs unvollständige Prädiktion bereits. Ein solcher Koeffizientensatz ist dann optimal, wenn eine "Reinigung" des Prädiktors vorgenommen wird, d. h. wenn die gespeicherten Rekonstruktionswerte $r(j)$; $j = 0, -1, -2, \dots, -N+1$; zu Beginn eines neuen Segments auf Null gesetzt werden. Eine solche Reinigung kann nützlich sein, um Sender und Empfänger nach Störungen im Übertragungskanal wieder zu synchronisieren. Wie die Untersuchungen zeigten, sollte eine solche Reinigung aber nur dann vorgenommen werden, wenn der zu erwartende Segmentgewinn klein ist, da sonst der Gesamtgewinn zu stark reduziert wird.

Möglichkeit 2:

Für die h_{*} -Bestimmung stehen ausser den MC Daten des Codiersegments zusätzlich wenigstens N vorhergehende Daten zur Verfügung, d. h., es gilt: $MA \geq MC + N$. Es handelt sich um eine unsymmetrische Überdeckung. Da auf jeden Fall, d. h. auch für $MA = MC + N$ die letzten N Daten des vorhergesagten Segments bekannt sind, können sie bereits zu den Zeitpunkten $n = 1, 2, 3, \dots, N$ zur Prädiktion mitverwendet werden; der Prädiktor bildet auch für die Schätzung der ersten Werte des Segments bereits eine Linearkombination mit N Werten. Die hier skizzierte Möglichkeit ergibt sich bei der ADPCM als die normale Art der Prä-

Die Matrixelemente oberhalb und unterhalb der Hauptdiagonalen können wiederum rekursiv aus den Werten $R_2(0, j); j = 1, 2, 3, \dots, N$ berechnet werden:

$$R_2(0, j) = \sum_{l=1}^{MA} s(l) \cdot s(l-j) = \sum_{l=1-j}^{MA-j} s(l) \cdot s(l+j) \quad (4.12)$$

Die Werte der Hauptdiagonalen ergeben sich aus dem ersten Element:

$$R_2(1, 1) = \sum_{l=1}^{MA} s^2(l-1) = \sum_{l=0}^{MA-1} s^2(l) \quad (4.13)$$

Die Matrix mit den Elementen $R_2(i, j); i, j = 1, 2, 3, \dots, N$ ist symmetrisch, besitzt aber keine Töplitz-Form.

4.1.2 Vocoderformulierung des Problems

Zusätzlich:

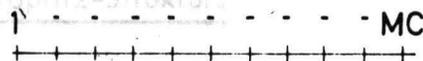
Die hier skizzierte Formulierung unterscheidet sich von den bisher beschriebenen Möglichkeiten der h_x Bestimmung. Sind MC Daten $s(n); n = 1, 2, 3, \dots, MC$ bekannt, so wird wiederum

$$E_D = \sum_{l=1}^{MA} d^2(l)$$

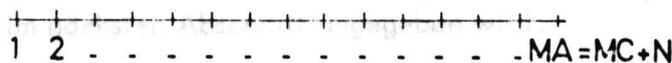
minimiert (s. Gl. 4.1). Im Unterschied zu den bisherigen Betrachtungen wird jetzt die ganze Differenzsignalfolge $d(n); n = 1, 2, 3, \dots, MC + N$ bei der Minimierung berücksichtigt,

$$R_2(i+1, i+1) = R_2(i, i) \quad (4.14)$$

Dabei hat die Matrix eine Töplitz-Struktur:



Die hier skizzierte Formulierung ergibt sich vor allem bei der Analyse und Synthese von Sprache mit Hilfe des Prädiktorsystems /21-24/. Bei der ADPCM-Codierung wird jedoch nur verlangt, dass die Energie der Differenzfolge in dem zu codierenden Segment minimal ist. Danach hat die Lösung die gleiche Bedeutung, da



die sich ergibt, wenn die endliche Signalfolge $s(n)$; $n = 1, 2, 3, \dots, MC$ mit einem Filter (dem Prädiktor) gefiltert wird, dessen Impulsantwort die Länge N hat. Das Gleichungssystem 4.3 wird jetzt

$$\sum_{i=1}^N h_i \cdot R_3(i, j) = R_3(0, j); \quad i = 1, 2, 3, \dots, N \quad (4.12)$$

mit

$$R_3(i, j) = \sum_{l=1}^{MA-i} s(l) \cdot s(l + i - j) \quad (4.13)$$

Symmetrie:

$$R_3(i, j) = R_3(j, i)$$

Rekursivität:

$$\begin{aligned} R_3(i+1, j+1) &= \sum_{l=1}^{MA-i-1} s(l) \cdot s(l + i - j) \\ &= R_3(i, j) - s(MA-i) \cdot s(MA-j) \end{aligned} \quad (4.16)$$

Die Werte $s(MA - i)$ sind für $i = 1, 2, 3, \dots, N$ Null, so dass gilt:

$$R_3(i+1, j+1) = R_3(i, j) \quad (4.14)$$

Damit hat die Matrix eine Töplitz-Struktur.

Die hier skizzierte Formulierung ergibt sich vor allem bei der Analyse und Synthese von Sprache mit Hilfe des Prädiktorsystems /21-24/. Bei der ADPCM-Codierung wird jedoch nur verlangt, dass die Energie der Differenzfolge in dem zu codierenden Segment minimal ist. Dennoch hat die Lösung dieses Abschnittes auch für die ADPCM-Codierung ihre Bedeutung, da sie der Lösung gleicht, die im nächsten Abschnitt angegeben wird.

4.1.3 Statistische Formulierung des Problems

Es wird angenommen, dass die zu codierenden Abtastwerte des Segments einen Ausschnitt aus einem quasi - stationären Prozess darstellen. Aufeinanderfolgende Segmente bilden eine Folge von näherungsweise stationären Zuständen. Der optimale Koeffizientenvektor \underline{h} kann dann aus dem Gleichungssystem Gl. 2.5, also aus

$$\left\| \begin{aligned} \sum_{j=1}^N h_j \cdot R_{ss}(i-j) &= R_{ss}(i) & i = 1, 2, 3, \dots, N \end{aligned} \right. \quad (4.15)$$

bestimmt werden. Dabei sind die R_{ss} - Werte die Werte der AKF des stationären Prozesses $\{s(n)\}$. Für die AKF - Werte gelten die bekannten Eigenschaften:

Symmetrie:

Die Schätzwerte $\hat{R}_{ss}(k)$ (bzw. $\tilde{R}_{ss}(k)$) sind symmetrisch, vor allem dann, wenn die Schätzwerte klein genug sind. Wenn nämlich die Schätzwerte $\hat{R}_{ss}(k)$ (bzw. $\tilde{R}_{ss}(k)$) groß sind (starke Korrelation), so sind die Schätzwerte $\hat{R}_{ss}(k)$ (bzw. $\tilde{R}_{ss}(k)$) stark miteinander korreliert. Die Schätzwerte klugen dann nicht genügend schnell mit wachsenden k ab. Das

$$R_{ss}(i-j) = R_{ss}(j-i)$$

Rekursivität:

Es gilt:

$$R_{ss}(i+1-j-1) = R_{ss}(i-j) \quad (4.16)$$

Die Matrizenzelemente der Hauptdiagonalen und aller anderen parallel liegenden Diagonalen sind also gleich (Streifensymmetrie), so dass alle Elemente aus den N AKF - Werten $R_{ss}(j) = E[s(n) \cdot s(n+j)]$; $j = 0, 1, 2, \dots, N-1$ bestimmt werden können (Töplitz-Matrix). Da die AKF - Werte $R_{ss}(\cdot)$ nicht bekannt sind, müssen Schätzwerte bestimmt werden, wobei die MA Werte des Analysesegments zur Berechnung verwendet werden. Die Frage nach "besten" Schätzwerten wird im nächsten Abschnitt behandelt. Zwei wichtige Schätzfunktionen der AKF sind:

$$\left\| \hat{R}_{ss}(k) = \frac{1}{MA} \sum_{n=1}^{MA-k} s(n) \cdot s(n+k); \quad k = 0, 1, 2, \dots, MA-1 \quad (4.17) \right.$$

und

$$\left\| \tilde{R}_{ss}(k) = \frac{1}{MA-k} \sum_{n=1}^{MA-k} s(n) \cdot s(n+k); \quad k = 0, 1, 2, \dots, MA-1 \quad (4.18) \right.$$

mit dem Zusammenhang

$$\hat{R}_{ss}(k) = \left(1 - \frac{k}{MA}\right) \cdot \tilde{R}_{ss}(k) \quad (4.19)$$

$\tilde{R}_{ss}(k)$ ist erwartungstreu, während $\hat{R}_{ss}(k)$ nur asymptotischerwartungstreu ist. Für beide Schätzfunktionen verringert sich die Varianz der Schätzung proportional $1/MA$. Obwohl $\tilde{R}_{ss}(k)$, da erwartungstreu, die kleinste Abweichung der AKF $R_{ss}(k)$ liefert, muss für die Zwecke der ADPCM-Codierung die Schätzfunktion $\hat{R}_{ss}(k)$ verwendet werden: der mit der Schätzfunktion $\hat{R}_{ss}(k)$ berechnete Koeffizientenvektor \underline{h}_* führt immer zu einem stabilen rekursiven Filter (dem Decodierer). Wie man leicht zeigen kann, ist die dazu notwendige und hinreichende Bedingung, dass die Autokovarianzmatrix \hat{R} der Schätz-AKF $\hat{R}_{ss}(\cdot)$ positiv-semidefinit ist, erfüllt:

$$\underline{a}^T \cdot \hat{R} \cdot \underline{a} > 0 \quad \forall \underline{a}$$

Die Schätz-AKF $\tilde{R}_{ss}(k)$ liefert z. T. instabile inverse Filter (Decodierer), vor allem dann, wenn die Prädiktionsgewinne hoch sind. Wenn nämlich die exakten $R_{ss}(\cdot)$ Werte gross sind (starke lineare Abhängigkeiten zwischen den Segmentdaten), dann erweist es sich, dass benachbarte Schätzwerte $\hat{R}_{ss}(k)$, $\hat{R}_{ss}(k+1)$, ... oder auch $\tilde{R}_{ss}(k)$, $\tilde{R}_{ss}(k+1)$, ... stark miteinander korreliert sind. Die Schätzwerte klingen dann nicht genügend schnell mit wachsenden k ab. Das gilt nicht für die Schätzfunktion $\hat{R}_{ss}(\cdot)$, da deren AKF-Werte gemäss Gl. 4.19 gegenüber der Schätzfunktion $\tilde{R}_{ss}(\cdot)$ durch ein Dreieck-Fenster bewertet sind.

Um genaue AKF-Werte zu erhalten, ist es günstig, das Analysesegment möglichst gross zu machen, jedoch ist die Grösse dadurch beschränkt, dass sich die statistischen Kenndaten für weitab liegende Bereiche der Eingangsmusterfolge bereits geändert haben. Die optimalen Grössen MA und MC können nur durch Versuche ermittelt werden. Anhaltspunkte gewinnt man aus der Kenntnis, dass Übergänge von einem Phonem zu einem anderen - entsprechend der Änderungsgeschwindigkeit des artikulatorischen Apparates - innerhalb von etwa 10 - 50 ms erfolgen. Bei einer 2ms-Segmentierung von Sprache ändern sich die AKF-Verläufe aufeinanderfolgender Segmente kaum /17/.

Die statistische Formulierung hat auch eine deterministische Interpretation: das Gleichungssystem Gl. 4.15 ergibt sich auch, wenn die Energie des Differenzsignals $d(n)$ über die ganze Folge $d(n)$ minimiert wird (siehe Abschnitt 4.1.2), denn es ist $R_3(i, j) = MA \cdot \hat{R}_{ss}(i-j)$.

Auch eine Formulierung nach dem Maximum-Likelihood-Prinzip liefert die gleichen Ergebnisse, jedoch nur unter der Voraussetzung, dass die Fehlerwerte $d(n)$; $n = 1, 2, 3, \dots, MA + N$ gauss verteilt sind /25/.

4.2 Simulationsergebnisse

Im folgenden werden vorwiegend Langzeitgewinne der adaptiven Verfahren genannt und mit den erreichbaren Gewinnen bei nichtadaptiven DPCM- und PCM- Verfahren verglichen. Weitere wichtige Ergebnisse über die Eigenschaften der ADPCM-codierten Sprache und anschauliche Deutungen dieser Ergebnisse liefern Segmentanalysen; sie werden im Abschnitt 6 ausführlich dargestellt.

4.2.1 Die Abhängigkeit des Gewinns von dem Koeffizienten-Bestimmungsverfahren und den Segmentlängen

Um das günstigste Koeffizienten-Bestimmungsverfahren und die günstigste Segmentlänge festlegen zu können, wurde eine 5 sec-Sprachprobe verwendet und der Gewinn für eine ADPCM-Codierung mit $N = 10$ Prädiktorkoeffizienten und einer 5 bit - Quantisierung (logarithmisch, nichtadaptiv) bestimmt. Vier Verfahren wurden untersucht:

Verfahren A : Bestimmung der AKF $\hat{R}_{ss}(\cdot)$ gemäss Gl. 4.17.

Verfahren B : wie A, jedoch Bewertung der $s(n)$ -Werte des Analysesegments mit einem Hammingfenster.

Verfahren C : Bestimmung der AKF $R_1(\cdot, \cdot)$ gemäss Gl. 4.7.

Verfahren D : Bestimmung der AKF $R_2(\cdot, \cdot)$ gemäss Gl. 4.10.

Bild 18 zeigt, wie sich die Analysesegmente bei einer Überdeckung und bei einer Bewertung der Abtastwerte mit einem Hammingfenster überlappen. Eine Bewertung mit einer Fensterfunktion ist nur bei dem Verfahren A der Schätz-AKF-Bestimmung verwendbar; bei den anderen Verfahren widerspricht eine solche Bewertung der Problemformulierung und führt deshalb i. a. auf instabile Filter.

Überdeckungen $MA/MC > 1$ erhöhen den Gewinn nur für $MC \leq 32$. Solch kurze Codiersegmente sind aber uninteressant, da der Koeffizientenvektor \hat{h}_n zu häufig übertragen werden muss.

Bei den Verfahren B, C und D sind nur für $MC \leq 120$ günstiger als das Verfahren A.

Für $MC \leq 120$ unterscheiden sich die Verfahren kaum.

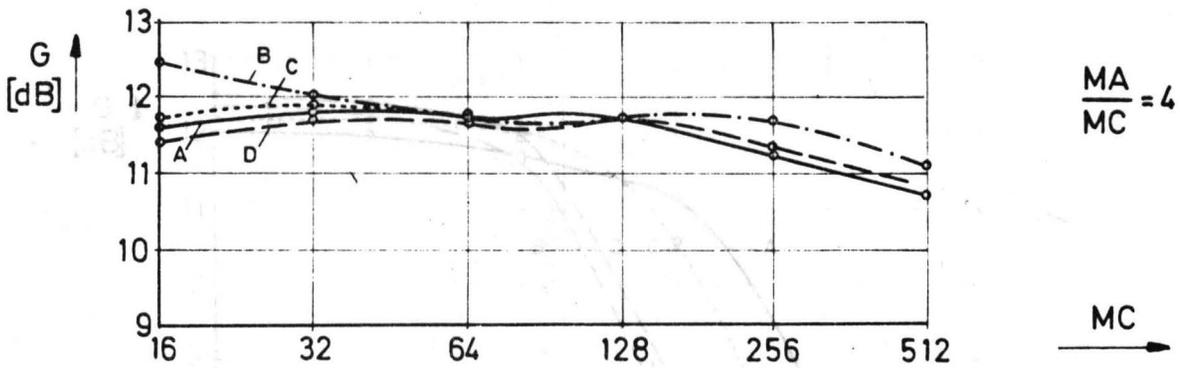
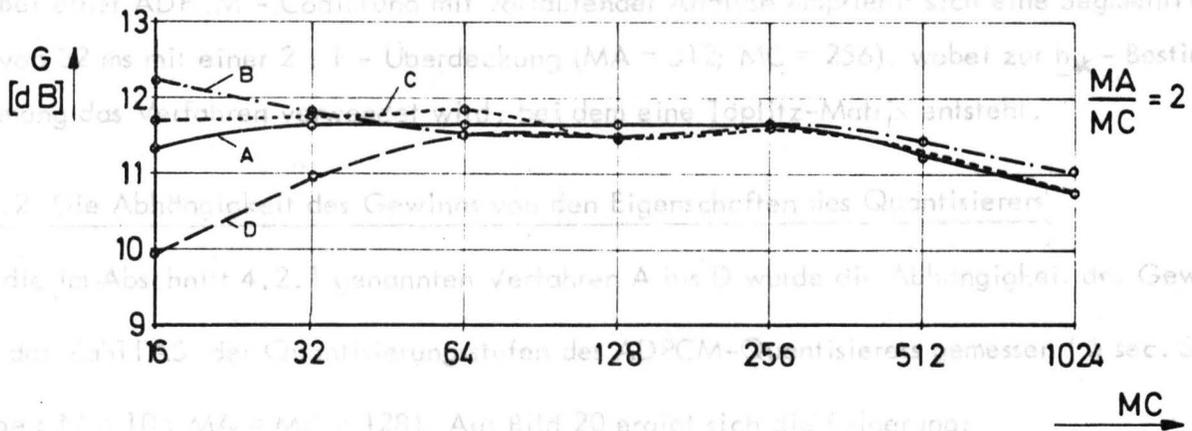
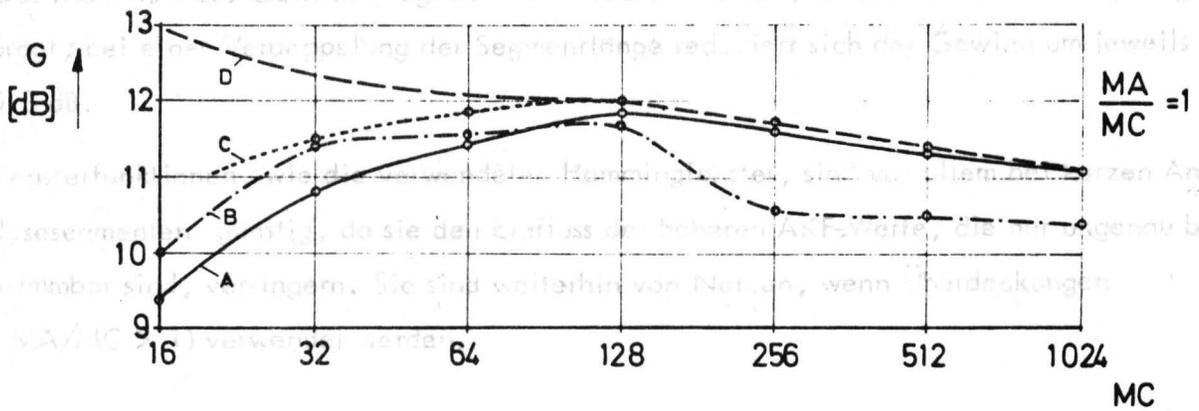


Bild 19 : Der Gewinn in Abhängigkeit von der Länge des Codiersegments und von dem Grad der Überdeckung für die vier auf Seite 28 beschriebenen Verfahren der Koeffizientenbestimmung

- d) Das Maximum des Gewinns liegt bei $MC = 128$ ($\hat{=} 16$ ms), ist jedoch recht wenig ausgeprägt; bei einer Verdoppelung der Segmentlänge reduziert sich der Gewinn um jeweils 0.3 dB.
- e) Fensterfunktionen, wie die verwendeten Hammingfenster, sind vor allem bei kurzen Analysesegmenten günstig, da sie den Einfluss der höheren AKF-Werte, die nur ungenau bestimmbar sind, verringern. Sie sind weiterhin von Nutzen, wenn Überdeckungen ($MA/MC > 1$) verwendet werden.

Da das Verfahren A die einfachste h_{*} -Bestimmung zulässt (vergl. Abschnitt 4.4), ergibt sich aus den Punkten a) bis d) die Folgerung:

Bei einer ADPCM - Codierung mit vorlaufender Analyse empfiehlt sich eine Segmentlänge von 32 ms mit einer 2 : 1 - Überdeckung ($MA = 512$; $MC = 256$), wobei zur h_{*} -Bestimmung das Verfahren verwendet wird, bei dem eine Töplitz-Matrix entsteht.

4.2.2 Die Abhängigkeit des Gewinns von den Eigenschaften des Quantisierers

Für die im Abschnitt 4.2.1 genannten Verfahren A bis D wurde die Abhängigkeit des Gewinns von der Zahl IAS der Quantisierungsstufen des ADPCM-Quantisierers gemessen (5 sec. Sprechprobe; $N = 10$; $MA = MC = 128$). Aus Bild 20 ergibt sich die Folgerung:

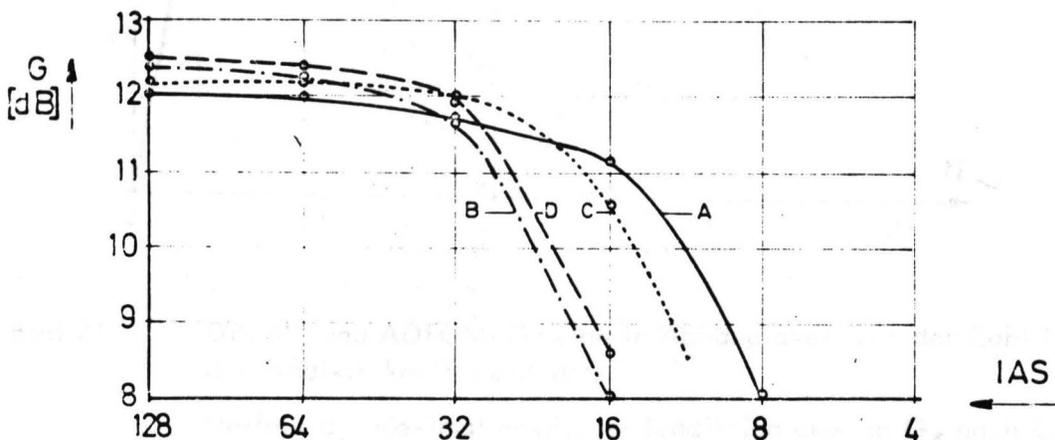


Bild 20: Der Gewinn in Abhängigkeit von der Zahl der Quantisierungsstufen für die vier auf Seite 28 beschriebenen Verfahren der Koeffizientenbestimmung

- a) Bei einer ADPCM - Codierung ist das Verfahren A (Bestimmung der Töplitzmatrix) am unempfindlichsten gegenüber dem Einfluss des Quantisierungsrauschens.
- b) ADPCM - Verfahren sind nur brauchbar, wenn wenigstens $IAS = 16$ Quantisierungsstufen zur Verfügung stehen (siehe dazu Abschnitt 6).
- c) Für $IAS > 32$ kann der Gewinn gegenüber dem Verfahren A um etwa 0.5 dB erhöht werden, wenn das aufwendigere Verfahren D angewendet wird.

4.2.3 Die Abhängigkeit des Gewinns von der Zahl der Prädiktorkoeffizienten

Die ADPCM benötigt eine sehr viel grössere Anzahl von Prädiktorkoeffizienten, als die DPCM, um ihren maximalen Gewinn zu erreichen. Bild 21 zeigt die Abhängigkeit des Gewinns G von der Zahl N der Prädiktorkoeffizienten. Der Verlauf a stellt den maximal möglichen DPCM-Gewinn dar, wie er aus der Langzeit-AKF einer Sprachprobe gemäss Gl. 2.5

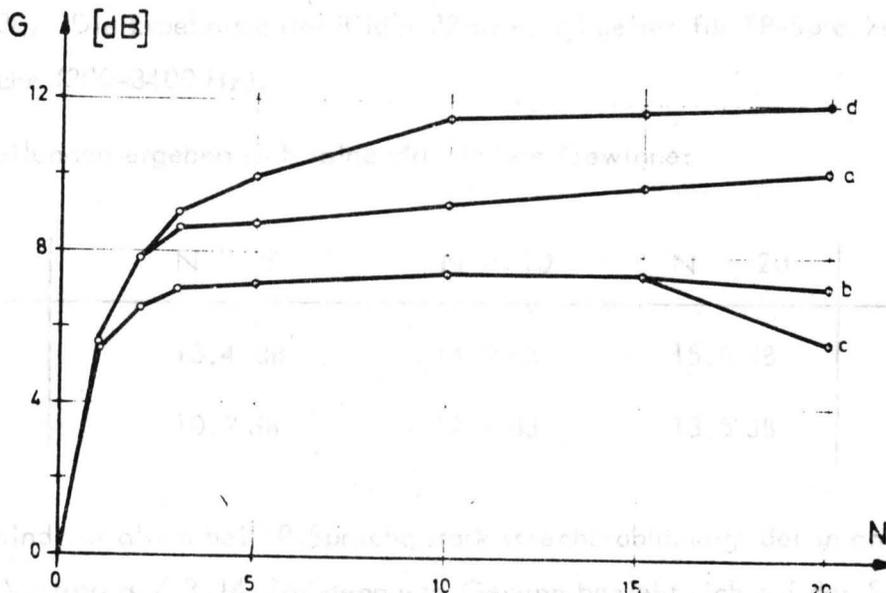


Bild 21: DPCM- und ADPCM-Gewinn in Abhängigkeit von der Zahl N der Prädiktorkoeffizienten
Verlauf a : maximal möglicher Prädiktionsgewinn G_* nach Gl. 2.6
Verlauf b : 5bit-DPCM-Gewinn bei sprecheroptimalem Prädiktor
Verlauf c : 5bit-DPCM-Gewinn bei einem gemäss einer mittleren AKF dimensionierten Prädiktor
Verlauf d : 5bit - ADPCM - Gewinn

berechnet werden kann. Der durch die Simulationen erreichbare Gewinn ist wegen des Einflusses des Quantisierungsrauschens geringer (Verlauf b). Eine weitere Gewinnreduktion entsteht dadurch, dass der DPCM-Prädiktor sprecherunabhängig eingestellt werden muss, d. h. der Koeffizientenvektor h_k wird aus einer über viele Sprachproben gemittelten AKF bestimmt (Verlauf c; AKF-Verlauf: siehe Bild 7). Die DPCM-Codierung ist, wie sich auch hier zeigt, recht unempfindlich gegenüber geringen Änderungen des optimalen Koeffizientenvektors (vgl. auch Abschnitt 2.3). Der Verlauf d zeigt schliesslich, dass durch eine ADPCM-Codierung zusätzliche Gewinne (etwa 4 dB Zunahme) erzielbar sind /3,14/. Bei den DPCM- und ADPCM-Verläufen wurde eine Quantisierung mit 5 bit (log. Quantisierer, $\mu = 100$) vorgenommen; bei den ADPCM-Codierungen war $MA = MC = 128$.

4.2.4 Die Abhängigkeit des Gewinns von der Sprachprobe

Die Bilder 22 und 23 zeigen ADPCM-Gewinne für drei Sprecher in Abhängigkeit von der Zahl N der Prädiktorkoeffizienten (5sec Sprache, 5 bit-Quantisierung, log., $\mu = 100$, $MA = MC = 128$). Die Ergebnisse der Bilder 22 bzw. 23 gelten für TP-Sprache (0-3400 Hz) bzw. BP-Sprache (300-3400 Hz).

Aus den Darstellungen ergeben sich folgende mittlere Gewinne:

	N = 5	N = 10	N = 20	maximale Abweichungen
TP - Sprache	13.4 dB	14.9 dB	15.8 dB	+ 1.8/- 3.5 dB
BP - Sprache	10.7 dB	12.8 dB	13.5 dB	+ 1.8/- 1.6 dB

Die Gewinne sind vor allem bei TP-Sprache stark sprecherabhängig; der in anderen Veröffentlichungen des Verfassers / 3,14,26/ genannte Gewinn bezieht sich auf den Sprecher 13; aus den Bildern 22 und 23 wird deutlich, dass der Sprecher 13 die ungünstigsten Gewinnwerte liefert. Die AKF-Werte dieses Sprechers sind relativ unabhängig von dem tieffrequenten Bereich 0 - 300 Hz, daher unterscheiden sich die Gewinne von TP-Sprache und BP-Sprache nicht sehr. Im Gegensatz dazu entstehen die recht hohen Korrelationen in der Sprachprobe der Sprecherin 15 vor allem durch die tieffrequenten Anteile von 0 - 300 Hz. Bei BP-Sprache ergibt sich daher die starke Gewinnreduktion um fast 5 dB.

4.3 ADPCM mit gesteuertem Quantisierer

4.3.1 ADPCM ohne Entropiecodierung

Bei der Codierung von Sprechsignalen ist die Verwendung eines Quantisierers mit logarithmischer Kennlinie zwingend notwendig, da die Varianz der Signaleignisse stark sprecherabhängig ist, der logarithmisch gesteuerte Quantisierer liefert auch bei unterschiedlichen Pegeln ein weitgehend konstantes Signal-Rausch-Verhältnis. Sprechsignale haben allerdings eine Laplace-Approximationseigenschaft, die hier nicht die logarithmische Abtastung nicht optimal. Bei Verwendung einer Quantisiererkennlinie $1/6,7$ ergibt sich ein Bit-Rausch-Verhältnis von Optimalquantisierung zu setzen zu können wird ein einfaches adaptives Verfahren verwendet (Bilder 24 und 25), die Folge der Prädiktorkoeffizienten wird so verändert, ihre Energie bestimmt wird mit einem gesteuerten Verstärker V an den Quantisierer angepasst $1/26,2$. Die Energie des Differenzsignals ist eigentlich erst bekannt, wenn die Codierung stattgefunden

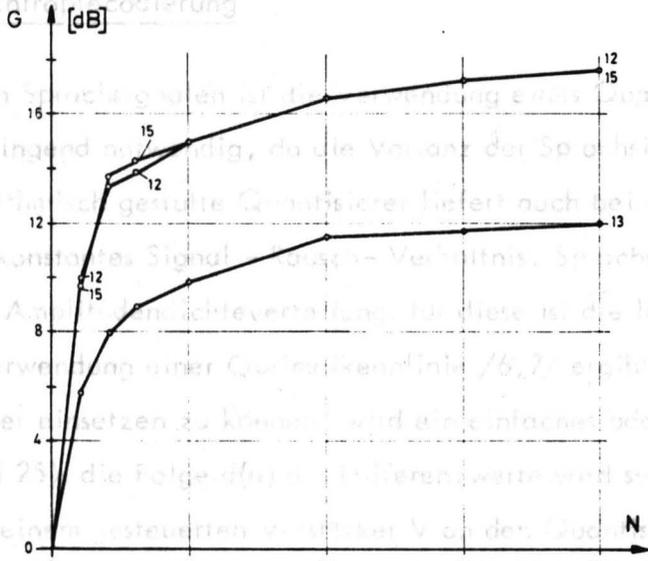


Bild 22 : ADPCM-Gewinne von drei Sprechern (Nr. 12, 13, 15) in Abhängigkeit von der Zahl N der Prädiktorkoeffizienten TP-Sprache (0 - 3400 Hz)

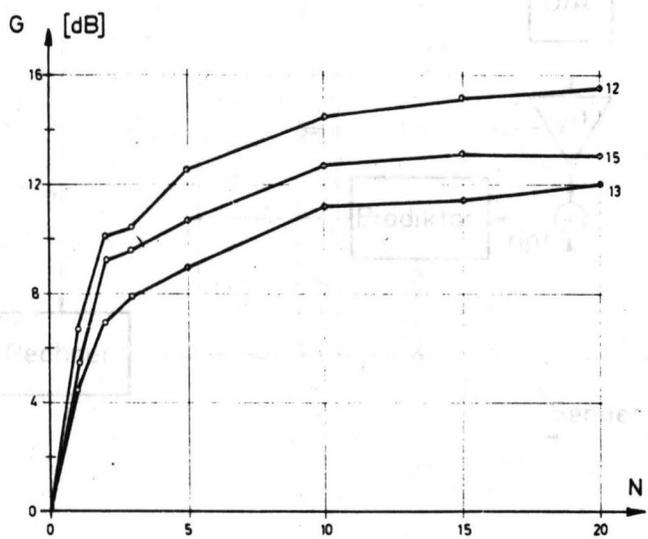


Bild 23 : ADPCM-Gewinne von drei Sprechern (Nr. 12, 13, 15) in Abhängigkeit von der Zahl N der Prädiktorkoeffizienten BP-Sprache (300 - 3400 Hz)

4.3 ADPCM mit gesteuertem Quantisierer

4.3.1 ADPCM ohne Entropiecodierung

Bei der Codierung von Sprachsignalen ist die Verwendung eines Quantisierers mit logarithmischer Kennlinie zwingend notwendig, da die Varianz der Sprachsignale stark sprecherabhängig ist; der logarithmisch gestufte Quantisierer liefert auch bei unterschiedlichen Pegeln ein weitgehend konstantes Signal - Rausch- Verhältnis. Sprachsignale haben näherungsweise eine Laplace - Amplitudendichteverteilung; für diese ist die logarithmische Abstufung nicht optimal. Bei Verwendung einer Optimal Kennlinie [6,7] ergibt sich ein höheres SNR; um Optimalquantisierer einsetzen zu können, wird ein einfaches adaptives Verfahren verwendet (Bilder 24 und 25): die Folge $d(n)$ der Differenzwerte wird segmentiert, ihre Energie bestimmt und mit einem gesteuerten Verstärker V an den Quantisierer angepasst [26,27]. Die Energie des Differenzsignals ist eigentlich erst bekannt, wenn die Codierung stattgefunden

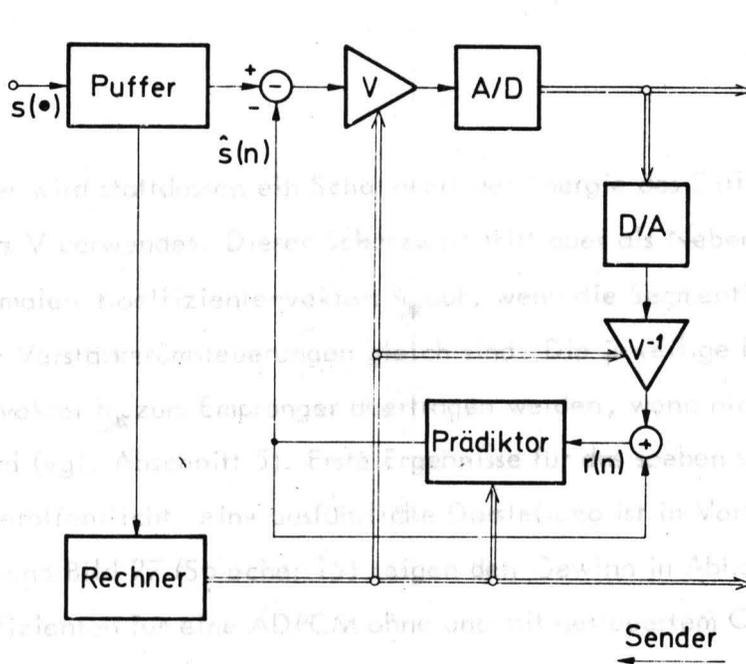


Bild 24: ADPCM-Codierer mit gesteuertem Quantisierer (vorläufige Analyse)

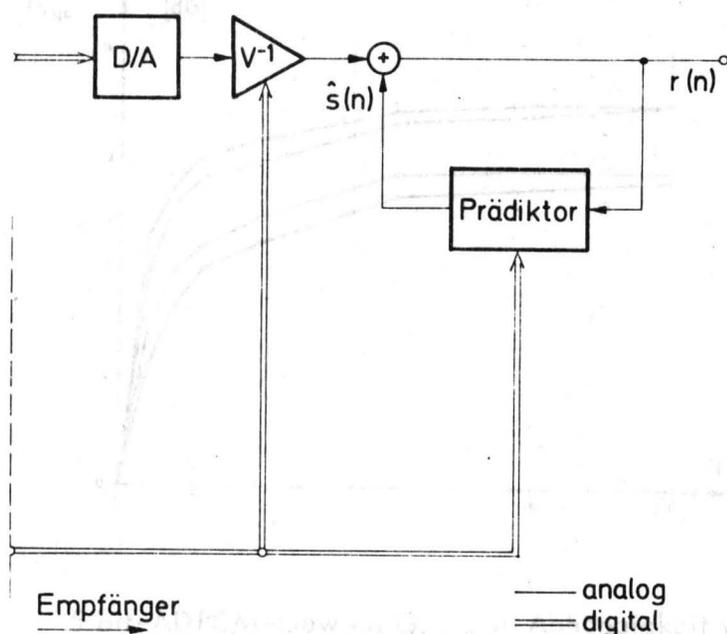


Bild 25 : ADPCM-Decodierer bei einer Codierung mit gesteuertem Quantisierer (vorlaufende Analyse)

den hat. Daher wird stattdessen ein Schätzwert der Energie des Differenzsignals zur Einstellung des Verstärkers V verwendet. Dieser Schätzwert tritt aber als Nebenergebnis bei der Bestimmung des optimalen Koeffizientenvektors \underline{h}_* auf, wenn die Segmentierungen für die Koeffizienten - und die Verstärkerumsteuerungen gleich sind. Die jeweilige Einstellung V muss mit dem Koeffizientenvektor \underline{h}_* zum Empfänger übertragen werden, wenn nicht eine Rückwärtsregelung verwendet wird (vgl. Abschnitt 5). Erste Ergebnisse für das soeben skizzierte Verfahren wurden in /26/ veröffentlicht, eine ausführliche Darstellung ist in Vorbereitung /27/. Bild 26 (Sprecher 13) und Bild 27 (Sprecher 15) zeigen den Gewinn in Abhängigkeit von der Zahl der Prädiktorkoeffizienten für eine ADPCM ohne und mit gesteuertem Quantisierer.

Der durch den optimalen Laplace-Quantisierer zusätzlich entstehende Gewinn liegt zwischen 3 und 4 dB. Als Gewinn ist hier der SNR-Abstand zu einer 5 bit - PCM - Codierung mit logarithmischen Quantisierer ($\mu = 100$; 16 σ -Belastung) verwendet worden, die ein SNR von 20.7 dB liefert /26/. Ein optimaler Laplace-Quantisierer müsste theoretisch 23.8 dB liefern /6/, also einen zusätzlichen Gewinn von 3.1 dB.

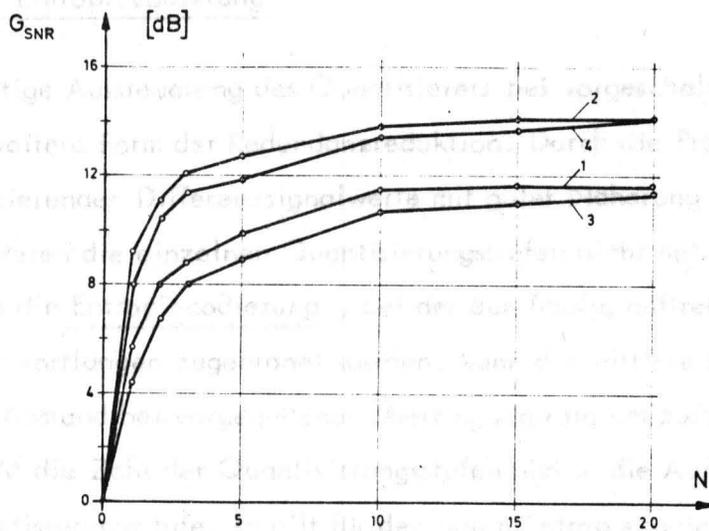


Bild 26 : 5 bit-ADPCM-Gewinn G_{SNR} in Abhängigkeit von der Prädiktorlänge N ; Sprecher 13

G_{SNR} = Gewinn gegenüber einem 5 bit - PCM - System

- Kurve 1 : TP-Sprache, log. Quantisierer
- Kurve 2 : TP-Sprache, gesteuerter Quantisierer
- Kurve 3 : BP-Sprache, log. Quantisierer
- Kurve 4 : BP-Sprache, gesteuerter Quantisierer

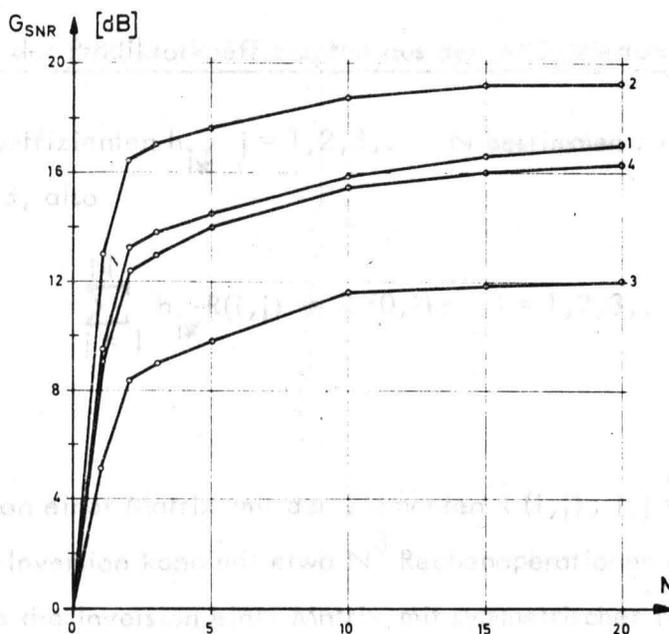


Bild 27 : wie Bild 26, aber Sprecher 15

4.3.2 ADPCM mit Entropiecodierung

Die immer gleichartige Aussteuerung des Quantisierers bei vorgeschalteter Verstärkungsregelung erlaubt eine weitere Form der Redundanzreduktion. Durch die Prädiktionscodierung sind zwar die zu quantisierenden Differenzsignalwerte mit guter Näherung voneinander linear unabhängig, jedoch treten die einzelnen Quantisierungsstufen nicht mit gleicher Wahrscheinlichkeit auf. Durch die Entropiecodierung, bei der den häufig auftretenden Quantisierungsstufen kürzere Codewortlängen zugeordnet werden, kann die mittlere Bitrate verringert bzw. der Signal-Rausch-Abstand bei vorgegebener Übertragungsrate um zusätzlich G_{EC} dB angehoben werden. Ist M die Zahl der Quantisierungsstufen und p_i die Auftretenswahrscheinlichkeit der i -ten Quantisierungsstufe, so gilt für den durch Entropiecodierung möglichen Gewinn:

$$G_{EC} = 6.02 \left(\log_2 M + \sum_{i=1}^M p_i \log_2 p_i \right) \quad [\text{dB}]$$

Bei einer Huffman-Codierung eines jeden Symbols tritt eine gewisse Restredundanz auf. Messungen mit Gaußsignalen haben gezeigt, dass diese Restredundanz vernachlässigbar ist /28/. Bei Verwendung eines Laplace-Optimalquantisierers ergibt sich aus den Simulationen ein zusätzlicher Gewinn G_{EC} von (im Mittel) 4.2 dB /26/.

4.4 Die Bestimmung der Prädiktorkoeffizienten aus den AKF-Werten

Um die optimalen Koeffizienten h_{i*} ; $i = 1, 2, 3, \dots, N$ bestimmen zu können, muss das Gleichungssystem Gl. 4.3, also

$$\sum_{j=1}^N h_{j*} \cdot R(i, j) = R(0, i); \quad i = 1, 2, 3, \dots, N$$

aufgelöst werden.

Dazu ist eine Inversion einer Matrix mit den Elementen $R(i, j)$; $i, j = 1, 2, 3, \dots, N$ erforderlich; eine solche Inversion kann mit etwa N^3 Rechenoperationen durchgeführt werden. Eine Ausnahme bildet die Inversion einer Matrix mit symmetrischer Töplitzstruktur, wie sie beim Verfahren nach Abschnitt 4.1.3 (Gl. 4.15) auftritt. Da dieses Verfahren zudem immer zu stabilen Lösungen führt, ist es den anderen Verfahren nach Abschnitt 4.1.1 vorzuziehen.

Die Inversion der Töplitz-Matrix kann mit einem rekursiven Algorithmus durchgeführt werden /5/, der nur etwa N^2 Rechenoperationen erfordert und eine einfachere Hardware-Realisierung zulässt. Bild 28 zeigt einen Vergleich von Rechenzeiten für die Inversion einer $N \times N$ - Matrix. Die Aufgabe der Inversion einer Matrix kann dadurch umgangen werden, dass ein Gradientenverfahren verwendet wird, bei dem der Koeffizientenvektor \underline{h} schrittweise geändert (verbessert) wird; im Abschnitt 5 wird ein solches Verfahren beschrieben.

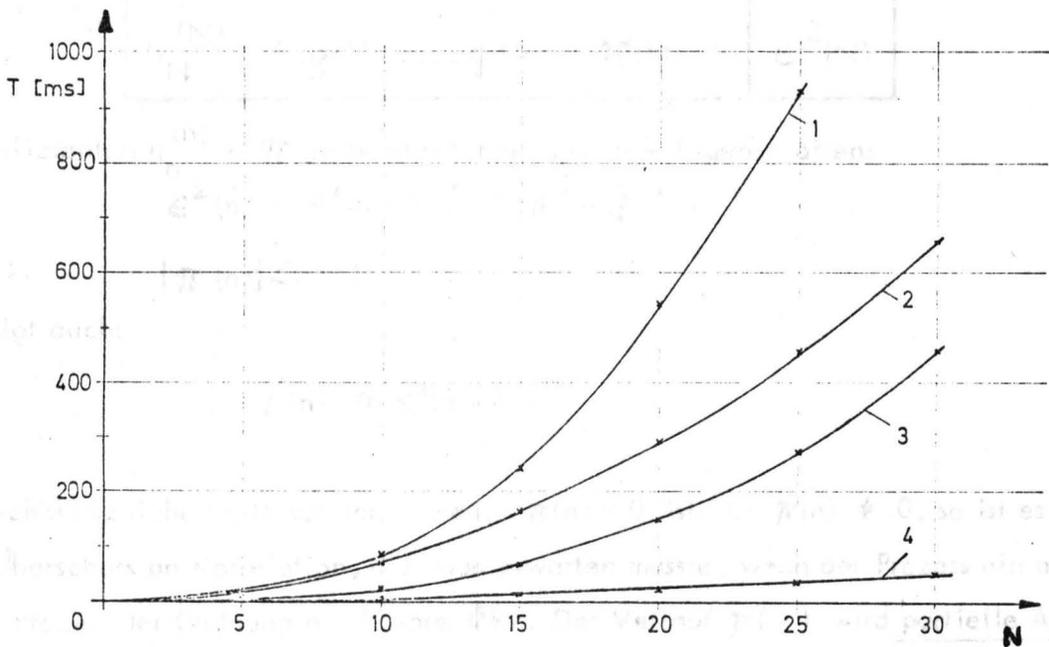


Bild 28 : CPU-Rechenzeiten zur Auflösung von Gleichungssystemen
 Programmiersprache: FORTRAN IV, Rechner: IBM 360/67

N = Zahl der Koeffizienten des Prädiktors

Verlauf 1 : Gauss-Jordan-Verfahren

Verlauf 2 : Cholesky-Dreieckszerlegung

Verlauf 3 : Gauss-Elimination

Verlauf 4 : Rekursives Verfahren für Töplitz-Matrizen

Das auf der rekursiven Inversion einer Töplitz-Matrix beruhende Verfahren liefert sukzessive die optimalen Koeffizienten für Prädiktoren zunehmender Ordnung n bis zur festgelegten Ordnung N und die damit jeweils erreichbare Fehlerleistung $\epsilon^2(n)$; $n = 1, 2, 3, \dots, N$. Sind die Schätzwerte $\hat{\mathbf{g}}(k) = \hat{\mathbf{R}}_{ss}(k) / \hat{\mathbf{R}}_{ss}(0)$ bekannt, so ergibt sich das folgende Rekursionsschema:

<u>1. Schritt :</u>	$h_1^{(1)} = g(1) = \pi(1)$	$\epsilon^2(1)$
<u>2. Schritt :</u>	$h_1^{(2)} = f(g(1), g(2))$ $h_2^{(2)} = f(g(1), g(2)) = \pi(2)$	$\epsilon^2(2)$
<u>N. Schritt :</u>	$h_1^{(N)} = f(g(1), \dots, g(N))$ $h_2^{(N)} = f(g(1), \dots, g(N))$ \vdots $h_N^{(N)} = f(g(1), \dots, g(N)) = \pi(N)$	$\epsilon^2(N)$

Die Koeffizienten $h_n^{(n)} = \pi(n)$ besitzen zwei wichtige Eigenschaften:

$$\epsilon^2(n) = \epsilon^2(n-1) \cdot [1 - \pi^2(n)] \quad (4.20)$$

$$|\pi(n)| \leq 1 \quad (4.21)$$

Damit folgt auch:

$$\epsilon^2(n) \leq \epsilon^2(n-1)$$

Das Gleichheitszeichen gilt nur dann, wenn $\pi(n) = 0$ ist. Ist $\pi(n) \neq 0$, so ist es ein Mass für den Überschuss an Korrelation, die man erwarten müsste, wenn der Prozess ein autoregressiver Prozess der Ordnung $n-1$ wäre [29]. Der Verlauf $\pi(\cdot)$ wird partielle Autokorrelationsfolge genannt, die einzelnen Werte der Folge heissen auch PARCOR - Koeffizienten.

Das im Abschnitt 4.1.3 beschriebene Verfahren der AKF-Bestimmung liefert nur dann stabile inverse (Decodier-) Filter, wenn die Elemente des \underline{h} -Vektors genau bestimmt wurden. Wurde der \underline{h} -Vektor quantisiert (um ihn digital übertragen zu können), so muss das inverse Filter durchaus nicht stabil sein. Das skizzierte rekursive Verfahren liefert eine Kontrolle der Stabilität: wird im Laufe der Berechnung der \underline{h} -Vektoren für Prädiktoren zunehmender Ordnung ein PARCOR-Koeffizient $\pi(j)$ betragsmässig grösser als 1, so ist das Filter der Ordnung j instabil. In diesem Fall wird ein verkürztes Prädiktionsfilter der Ordnung $j-1$ verwendet (vergl. Abschnitt 4.6).

Die Untersuchungen zeigten allerdings, dass Instabilitäten bei nicht zu grober Quantisierung durchaus zugelassen werden können: die Instabilitäten treten nur bei grossen Gewinnen auf, bei denen dann das Differenzsignal eine sehr kleine Varianz hat. Durch die Instabilität wächst dann das Differenzsignal im Segment an, aber ohne dass der Aussteuerbereich des Quantisierers überschritten wird.

4.5 Verfahren der vereinfachten AKF - Bestimmung

Bild 29 zeigt, dass die Bestimmung der AKF - Werte $\hat{R}_{ss}(k) = \sum_{n=1}^{MA-k} s(n) \cdot s(n+k);$

$k = 0, 1, 2, \dots, MA - 1$ recht hohe Rechenzeiten erfordert. Es wurde untersucht, ob es mög-

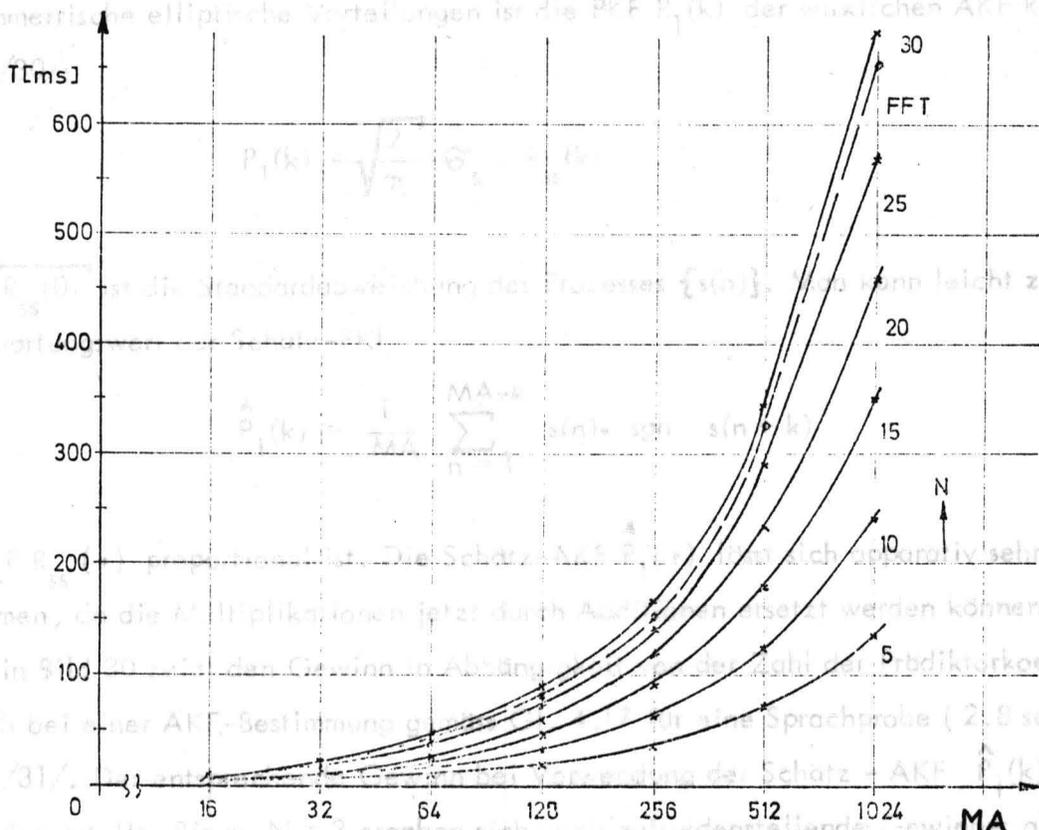


Bild 29 : CPU-Rechenzeiten zur Bestimmung von Korrelationswerten
 Programmiersprache: FORTRAN IV, Rechner : IBM 360/67
 MA = Anzahl der Funktionswerte
 N = Anzahl der benötigten Korrelationswerte
 FFT = Rechenzeiten bei Verwendung des FFT-Algorithmus

lich ist, vereinfachte AKF - Bestimmungen zu verwenden, durch die die Rechenzeiten verringert und Hardware - Realisierungen vereinfacht werden.

4.5.1 Einseitige Polaritätskorrelationsfolge

Die einseitige Polaritätskorrelationsfolge (PKF) wird definiert durch:

$$P_1(k) = E [s(n) \cdot \text{sgn } s(n+k)] \quad (4.22)$$

Für symmetrische elliptische Verteilungen ist die PKF $P_1(k)$ der wirklichen AKF $R_{ss}(k)$ proportional /30/:

$$P_1(k) = \sqrt{\frac{2}{\pi}} \cdot \sigma_s \cdot R_{ss}(k) \quad (4.23)$$

$\sigma_s = \sqrt{R_{ss}(0)}$ ist die Standardabweichung des Prozesses $\{s(n)\}$. Man kann leicht zeigen, dass der Erwartungswert der Schätz-PKF

$$\hat{P}_1(k) = \frac{1}{MA} \sum_{n=1}^{MA-k} s(n) \cdot \text{sgn } s(n+k) \quad (4.24)$$

der AKF $R_{ss}(\bullet)$ proportional ist. Die Schätz-AKF $\hat{P}_1(\bullet)$ lässt sich apparativ sehr einfach bestimmen, da die Multiplikationen jetzt durch Additionen ersetzt werden können. Der Verlauf a in Bild 30 zeigt den Gewinn in Abhängigkeit von der Zahl der Prädiktorkoeffizienten, der sich bei einer AKF-Bestimmung gemäss Gl. 4.17 für eine Sprachprobe (2.8 sec Dauer) ergibt /31/. Der entsprechende Gewinn bei Verwendung der Schätz - AKF $\hat{P}_1(k)$ ist als Verlauf b dargestellt. Bis zu $N = 3$ ergeben sich noch zufriedenstellende Gewinne; ab $N = 3$ nimmt der Gewinn mit wachsenden N ab, da die inversen Prädiktionsfilter nicht mehr stabil sind. Es mussten daher verkürzte Filter verwendet werden; die gestrichelte Kurve zeigt den bei diesem Verfahren entstehenden proportionalen Anteil von verkürzten Filtern.

4.5.2 Zweiseitige Polaritätskorrelationsfolge

Die zweiseitige Polaritätskorrelationsfolge wird definiert durch

$$P_2(k) = E [\text{sgn } s(n) \cdot \text{sgn } s(n+k)] \quad (4.25)$$

Für symmetrische elliptische Verteilungen ergibt sich ein nichtlinearer Zusammenhang /32/ zwischen der PKF $P_2(k)$ und der variierten AKF $R_{ss}(k) / \sigma_s^2$.

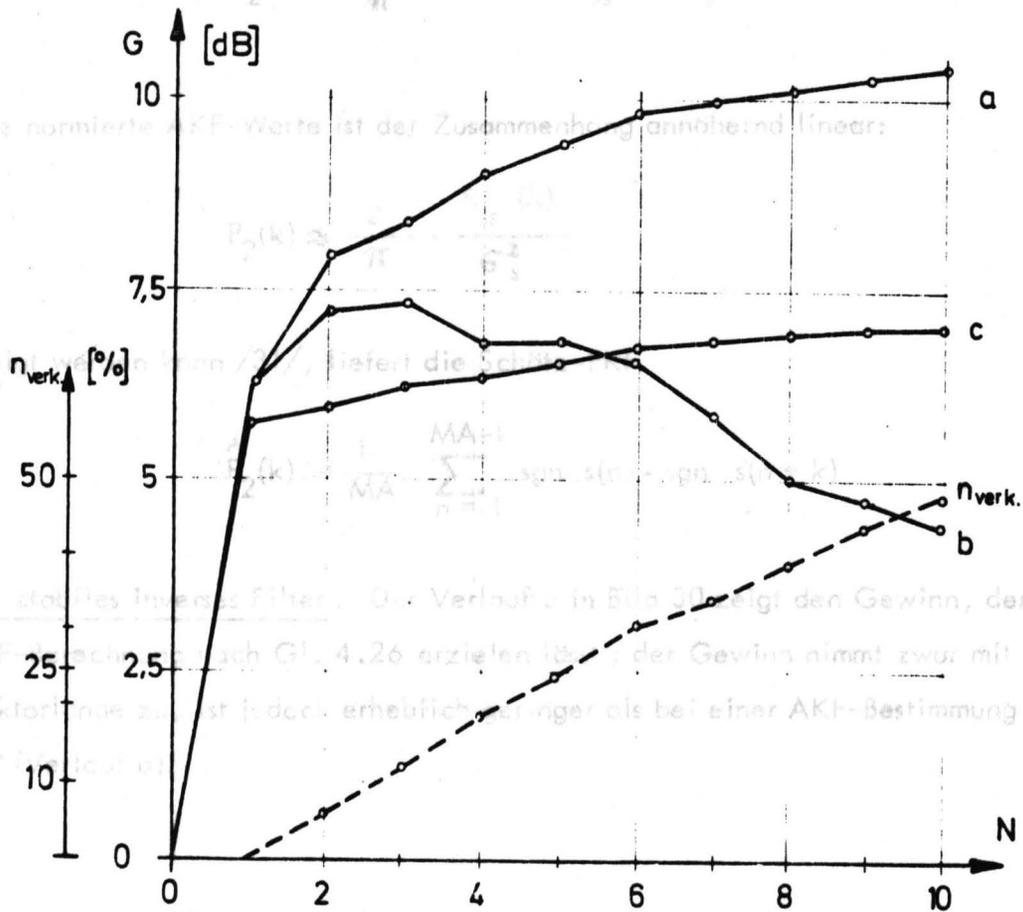


Bild 30 : ADPCM-Gewinn in Abhängigkeit von der Zahl N der Prädiktorkoeffizienten ; Vergleich verschiedener AKF-Schätzverfahren /31/

Kurve a : \hat{h} wurde mit \hat{R} (•) bestimmt (Gleichung 4.17)

Kurve b : \hat{h} wurde mit \hat{P}_1^{SS} (•) bestimmt (Gleichung 4.24)

Kurve c : \hat{h} wurde mit \hat{P}_2 (•) bestimmt (Gleichung 4.26)

Zur Kurve b ist (gestrichelt) der Anteil an verkürzten Filtern aufgetragen.

4.5.2 Zweiseitige Polaritätskorrelationsfolge

Die zweiseitige Polaritätskorrelationsfolge wird definiert durch

$$P_2(k) = E [\text{sgn } s(n) \cdot \text{sgn } s(n+k)] \quad (4.25)$$

Für symmetrische elliptische Verteilungen ergibt sich ein nichtlinearer Zusammenhang /32/ zwischen der PKF $P_2(k)$ und der normierten AKF $R_{ss}(k) / \sigma_s^2$:

$$P_2(k) = \frac{2}{\pi} \cdot \arcsin (R_{ss}(k) / \sigma_s^2)$$

Für kleine normierte AKF-Werte ist der Zusammenhang annähernd linear:

$$P_2(k) \approx \frac{2}{\pi} \cdot \frac{R_{ss}(k)}{\sigma_s^2}$$

Wie gezeigt werden kann /31/, liefert die Schätz-PKF

$$\hat{P}_2(k) = \frac{1}{MA} \sum_{n=1}^{MA-k} \text{sgn } s(n) \cdot \text{sgn } s(n+k) \quad (4.26)$$

immer ein stabiles inverses Filter. Der Verlauf c in Bild 30 zeigt den Gewinn, der sich bei einer AKF-Berechnung nach Gl. 4.26 erzielen lässt; der Gewinn nimmt zwar mit wachsender Prädiktorlänge zu, ist jedoch erheblich geringer als bei einer AKF-Bestimmung nach Gl. 4.17 (Verlauf a).

4.6 Das Problem der Übertragung der Information über die Prädiktoreinstellung

Um die Information über N Prädiktorkoeffizienten (oder daraus abzuleitende Größen) zum Empfänger übertragen zu können, müssen für jeden Koeffizienten r_i ; $i = 1, 2, 3, \dots, N$ bit zur Verfügung stehen. Um den Gleichlauf von Codierer und Decodierer zu gewährleisten, müssen auch die Koeffizienten des Codierers mit jeweils r_i bit/Koeffizient dargestellt werden. Für jedes Codiersegment mit MC Werten müssen $\sum_{i=1}^N (r_i)$ bit Information zusätzlich zu dem digitalen Datenfluss des Differenzsignals übertragen werden. Damit verringert sich der tatsächliche Gewinn gegenüber dem bisher angegebenen SNR-Gewinn um

$$\Delta G = \frac{6.02}{MC} \cdot \sum_{i=1}^N r_i \quad [\text{dB}] \quad (4.27)$$

Diese Gewinnabnahme ist in Bild 31 graphisch dargestellt.

Soll der Gewinn des ADPCM - Verfahrens möglichst hoch sein, so muss die zu übertragene Information über die Prädiktorkoeffizienten möglichst grob quantisiert werden. Damit weichen

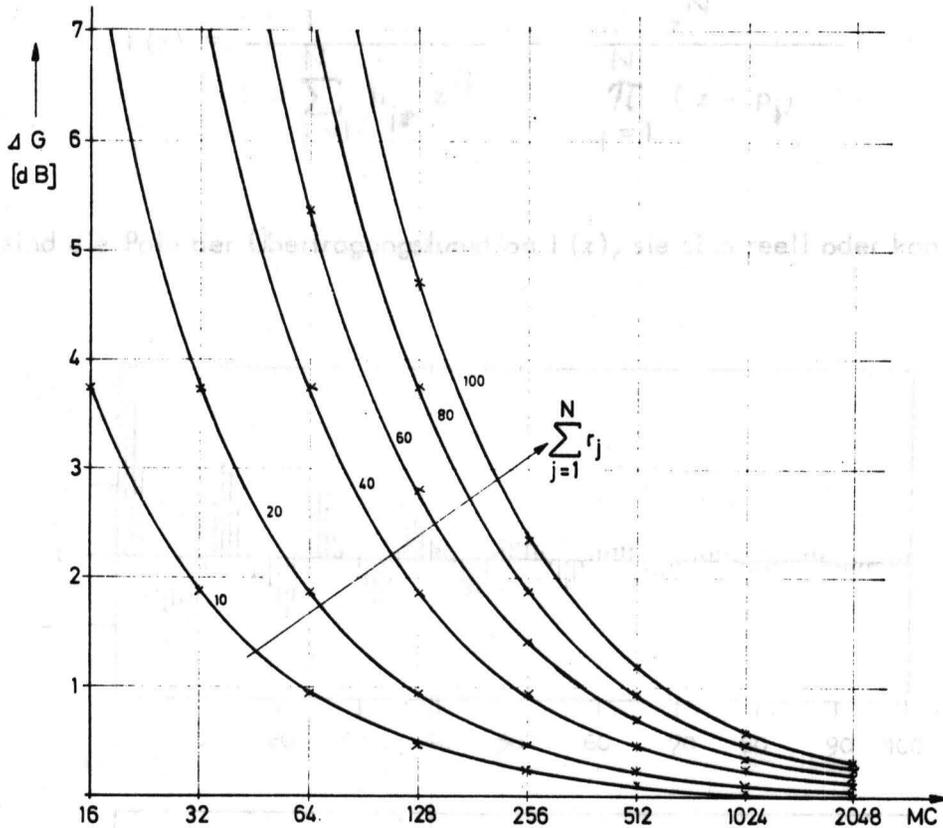


Bild 31 : Gewinnabnahme, hervorgerufen durch die zusätzlich erforderliche Übertragung der \underline{h} - Koeffizientensätze, in Abhängigkeit von der Länge des Codiersegments
Parameter : Anzahl der erforderlichen bit/Segment

die Prädiktorkoeffizienten \underline{h}_i ; $i = 1, 2, 3, \dots, N$ von ihren genauen Werten ab. Zwei Konsequenzen ergeben sich aus dieser ungenauen Darstellung:

- a) die Prädiktion wird schlechter, die Varianz des Differenzsignals vergrößert sich.
- b) das rekursive Filter des Decodierers kann instabil werden.

Gerade dieser zweite Punkt erweist sich als recht problematisch. Bild 32 zeigt die Impulsantwort, die Leistungsübertragungsfunktion und die Polstellenlage des rekursiven Filters für den Vokal a /33/. Für einen Koeffizientensatz \underline{h}_* ist die Übertragungsfunktion des inversen Filters :

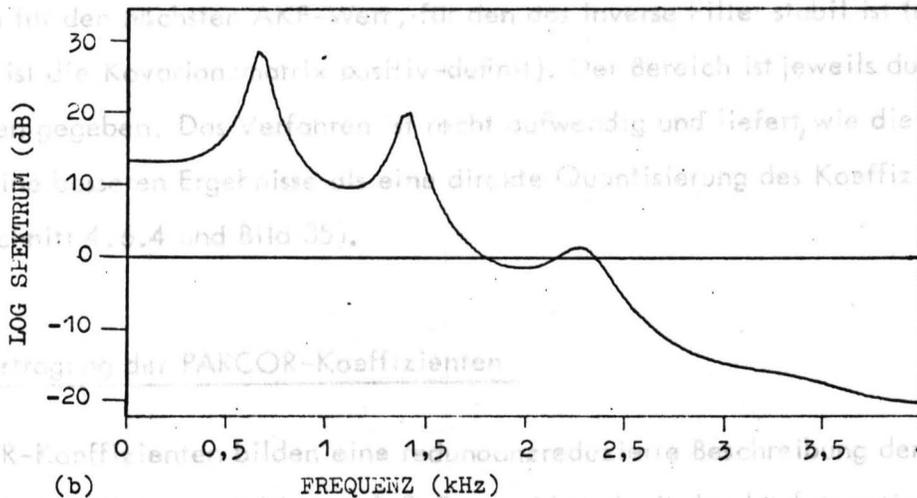
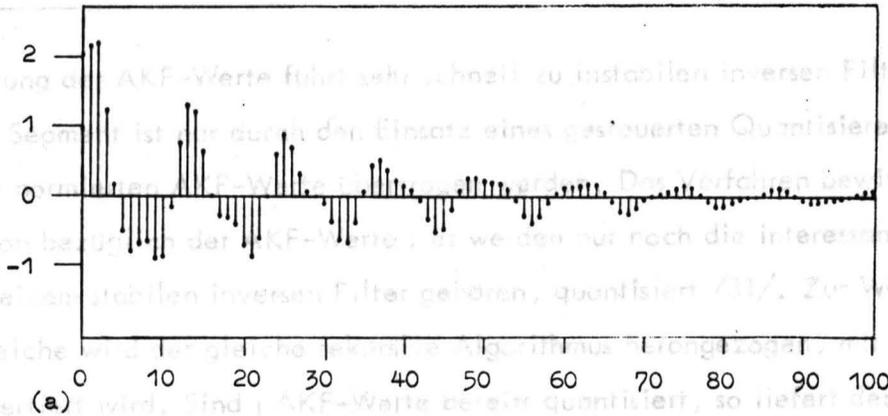
Es wird deutlich, dass eine Pole nicht am Einheitskreis liegt z^N und damit durch eine zu große Quantisierung auch über den Einheitskreis hinausgeschoben werden können. Diese Gefahr besteht vor allem bei N vollläufigen h_{ij} und h_{ij} (vergleiche Abschnitt 4.6.1).

$$I(z) = \frac{1}{1 - \sum_{i=1}^N h_{ij} z^{-i}} = \frac{z^N}{\prod_{i=1}^N (z - p_i)}$$

Die Werte p_i sind die Pole der Übertragungsfunktion $I(z)$, sie sind reell oder konjugiert komplex.

4.6.1 Übertragung der AKF-Werte

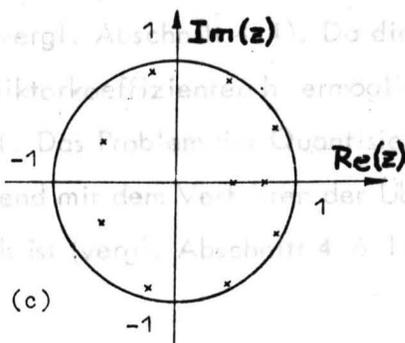
Eine Quantisierung der AKF-Werte führt sehr schnell zu instabilen inversen Filtern. Eine geringe Bitrate (e. Sekunde) ist nur durch den Einsatz eines gesteuerten Quantisierers erreichbar [31], wenn die interessanten Bereiche der AKF-Werte übergeben werden. Das Verfahren bewirkt eine Redundanzprodukt an den Stellen der AKF-Werte, in denen nur noch die interessanten AKF-Bereiche, die zu einem stabilen inversen Filter gehören, quantisiert [31]. Zur Wahl dieser interessanten Bereiche ist ein gleiches rekursives Algorithmus (Bsp. 4.6.1) notwendig, dem die Topologie-Matrix invertiert wird. Sind J AKF-Werte bereits quantisiert, so liefert der Algorithmus den Bereich für den nächsten AKF-Wert, für den das inverse Filter stabil ist (d. h. für diese AKF-Werte ist die Kovarianzmatrix positiv-definit). Der Bereich ist jeweils durch die PARCOR-Koeffizienten gegeben. Das Verfahren ist recht aufwendig und liefert, wie die Simulationen zeigten, kein besseres Ergebnis als eine direkte Quantisierung des Koeffizientenvektors a_k (vergl. Abschnitt 4.4 und Bild 32).



4.6.2 Übertragung der PARCOR-Koeffizienten

Die PARCOR-Koeffizienten bilden eine neue, reduzierte Beschreibung der AKF-Werte. Die PARCOR-Koeffizienten $\pi(j)$, $j = 1, 2, 3, \dots, N$ sind mit den höchstwertigen Prädiktorkoeffizienten der Prädiktoren der Länge j identisch. Sie beschreiben gleichzeitig die Abweichung der AKF von dem Wert, der bei einem autoregressiven Prozess der Ordnung $j-1$ auftreten würde (vergl. Abschnitt 4.4). Da die PARCOR-Koeffizienten, sind sie zur Übertragung der Prädiktorinformation geeignet. Das Problem der Quantisierung der $\pi(j)$ -Werte wurde nicht untersucht, da es weitgehend mit dem Problem der Übertragung der AKF-Werte mit gesteuertem Quantisierer identisch ist (vergl. Abschnitt 4.6.1).

Bild 32 : Impulsantwort, Spektrum und Lage der Polstellen für den Vokal "a"/33/



Es wird deutlich, dass einige Pole dicht am Einheitskreis liegen und damit durch eine zu grobe Quantisierung auch über den Einheitskreis hinausgeschoben werden können. Diese Gefahr besteht vor allem bei Nasallauten wie "m" und "n" (vergl. Abschnitt 6).

Verschiedene Verfahren der Informationsübertragung bieten sich an:

4.6.1 Übertragung der AKF-Werte

Eine Quantisierung der AKF-Werte führt sehr schnell zu instabilen inversen Filtern. Eine geringe Bitrate je Segment ist nur durch den Einsatz eines gesteuerten Quantisierers erreichbar /31/, wenn die normierten AKF-Werte übertragen werden. Das Verfahren bewirkt eine Redundanzreduktion bezüglich der AKF-Werte; es werden nur noch die interessanten AKF-Bereiche, die zu einem stabilen inversen Filter gehören, quantisiert /31/. Zur Wahl dieser interessanten Bereiche wird der gleiche rekursive Algorithmus herangezogen, mit dem die Töplitz-Matrix invertiert wird. Sind j AKF-Werte bereits quantisiert, so liefert der Algorithmus den Bereich für den nächsten AKF-Wert, für den das inverse Filter stabil ist (d. h. für diese AKF-Werte ist die Kovarianzmatrix positiv-definit). Der Bereich ist jeweils durch die PARCOR-Koeffizienten gegeben. Das Verfahren ist recht aufwendig und liefert, wie die Simulationen zeigten, keine besseren Ergebnisse als eine direkte Quantisierung des Koeffizientenvektors \underline{h}_x . (vergl. Abschnitt 4.6.4 und Bild 35).

4.6.2 Übertragung der PARCOR-Koeffizienten

Die PARCOR-Koeffizienten bilden eine redundanzreduzierte Beschreibung der AKF-Werte. Die PARCOR-Koeffizienten $\pi(j)$; $j = 1, 2, 3, \dots, N$ sind mit den höchstwertigen Prädiktorkoeffizienten der Prädiktoren der Länge j identisch. Sie beschreiben gleichzeitig die Abweichung der AKF von dem Wert, der bei einem autoregressiven Prozess der Ordnung $j - 1$ auftreten müsste (vergl. Abschnitt 4.4). Da die PARCOR-Koeffizienten eine sukzessive Berechnung der Prädiktorkoeffizienten \underline{h} ermöglichen, sind sie zur Übertragung der Prädiktorinformation geeignet. Das Problem der Quantisierung der $\pi(j)$ -Werte wurde nicht untersucht, da es weitgehend mit dem Verfahren der Übertragung der AKF-Werte mit gesteuertem Quantisierer identisch ist (vergl. Abschnitt 4.6.1).

4.6.3 Übertragung der Lage der Pole

Die Nullstellen des Polynoms $\sum_{i=1}^N h_{i*} z^{-i}$ (also der Übertragungsfunktion des optimal eingestellten Prädiktors) sind die Pole der Übertragungsfunktion des Decodierers. Um die Pole zu übertragen, müssten sie beim Sender aus dem Koeffizientenvektor \underline{h}_{i*} berechnet werden. Beim Empfänger müsste jeweils dazu der Vektor \underline{h}_{i*} aus den Angaben über die Lage der Pole bestimmt werden. Dieses Verfahren ist sehr aufwendig und wurde daher nicht untersucht. Nimmt man den Aufwand in Kauf, so bietet dieses Verfahren jedoch einige Vorteile:

- * Der Einfluss der Quantisierung ist überschaubar : durch die Quantisierung wird direkt die Lage der Resonanzfrequenzen und die Bandbreite der Resonanzen geändert.
- * Es ist eine einfache Prüfung möglich, ob das Filter stabil ist ($|p_i| < 1$).
- * Das nichtrekursive Filter kann sehr einfach in Kaskadenschaltung dimensioniert werden ; eine solche Struktur vermindert den Effekt der Fehlerakkumulation.
- * Vermutlich ist eine recht grobe Quantisierung zulässig, wenn der Abstand der Pole von dem Nullpunkt der z-Ebene nichtlinear quantisiert wird (feine Quantisierung in der Nähe des Einheitskreises) /35/.

4.6.4 Übertragung der Prädiktorkoeffizienten

Dieses Verfahren ist anzustreben, da der übertragene, quantisierte Koeffizientenvektor \underline{h}_{i*} direkt zur Einstellung des Decodierers verwendet werden kann. Es zeigt sich, dass der Vektor \underline{h}_{i*} relativ unempfindlich gegenüber einer Quantisierung ist (vergl. Abschnitt 2.2). Bild 33 zeigt die Auftretenswahrscheinlichkeiten für die Filterkoeffizienten h_{j*} ; $j = 1, 2, 3, \dots, N$, gemessen an einer 2.8 s - Sprachprobe (/31/, Sprecher 13, TP-Sprache, MA = MC = 128). Es wird deutlich, dass die einzelnen Koeffizienten nur gewisse Amplitudenbereiche überdecken. Die Quantisierer für die einzelnen Koeffizienten können an diese Bereiche angepasst werden. Bild 34 zeigt dazu die Mittelwerte und Standardabweichungen der einzelnen Koeffizienten.

Sprecher 13, TP Sprache, MA = MC = 128

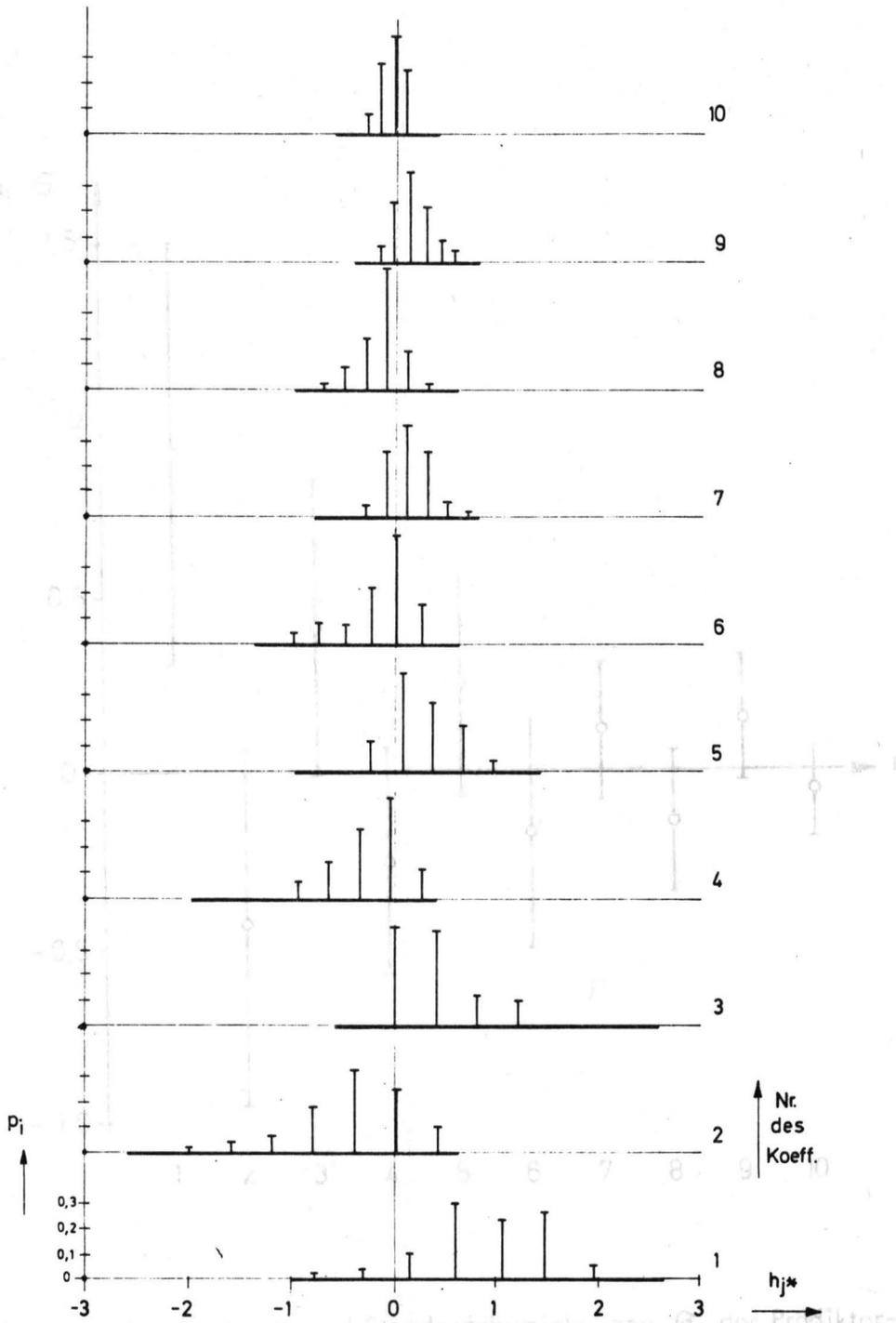


Bild 33 : Auftretenswahrscheinlichkeiten der Prädiktorkoeffizienten h_{j^*} ; $j = 1, 2, 3, \dots, 10$; /31/ *

Sprecher 13, TP-Sprache, MA = MC = 128

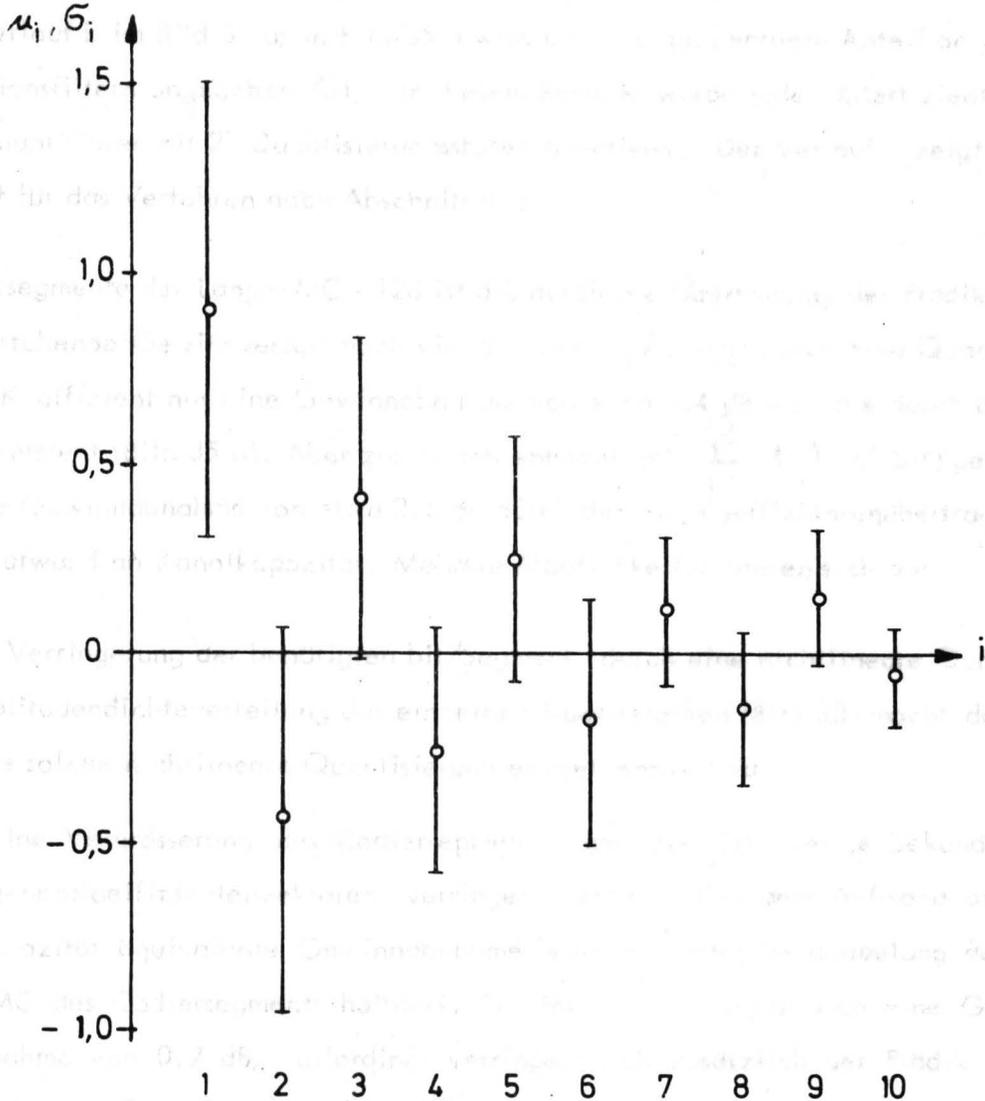


Bild 34: Mittelwerte μ_i und Standardabweichungen σ_i der Prädiktorkoeffizienten h_{i*} ; $i = 1, 2, \dots, 10$

ADPCM, $N = 10$, Sprecher 13, TP-Sprache, $MA = MC = 128$

Werden die Quantisierer für die einzelnen Koeffizienten h_j ; $j = 1, 2, 3, \dots, N$ optimal an diese Amplitudenbereiche angepasst, d. h. bei Berücksichtigung des Mittelwertes und des Aussteuerbereiches, so ergibt sich der Gewinn in Abhängigkeit von der Bitzahl / Segment als Verlauf b im Bild 35 a; in Bild 35 b wird dazu der prozentuale Anteil an verkürzten Prädiktionsfiltern angegeben /31/. In diesem Beispiel wurde jeder Koeffizient mit einem linearen Quantisierer mit 2^r Quantisierungsstufen quantisiert. Der Verlauf c zeigt den Gewinnverlauf für das Verfahren nach Abschnitt 4.6.1.

Für Codiersegmente der Länge $MC = 128$ ist die durch die Übertragung der Prädiktorkoeffizienten entstehende Gewinnverlust noch viel zu hoch. Zwar tritt durch eine Quantisierung mit 6 bit / Koeffizient nur eine Gewinnabnahme von etwa 0.4 dB ein, die durch ungenaue Prädiktion entsteht (Bild 35 a). Aber zusätzlich entsteht (mit $\sum_{j=1}^{10} (r_j) = 60$ bit) gemäss Bild 31 eine Gewinnabnahme von etwa 2.8 dB durch den zur Koeffizientenübertragung notwendigen Aufwand an Kanalkapazität. Mehrere Möglichkeiten bieten sich an:

- * Weitere Verringerung der benötigten bit/Segment durch eine nichtlineare Quantisierung. Die Amplitudendichteverteilung der einzelnen Koeffizienten (Bild 33) macht deutlich, dass eine solche nichtlineare Quantisierung empfehlenswert ist.
- * Durch eine Vergrößerung des Codiersegments kann die Zahl der je Sekunde zu übertragenen Koeffizientenvektoren verringert werden. Die dem Aufwand an Kanalkapazität äquivalente Gewinnabnahme wird mit jeder Verdoppelung der Länge MC des Codiersegments halbiert. Für $MC = 512$ ergibt sich eine Gewinnabnahme von 0.7 dB; allerdings verringert sich zusätzlich der Prädiktionsgewinn um 0.6 dB (gegenüber $MC = 128$), so daß insgesamt eine Abnahme von 1.3 dB entsteht.
- * Eine weitere Verringerung des Datenflusses des Koeffizientenvektors ist mit einem redundanzreduzierenden Codierungsverfahren möglich. Hier bieten sich die Loeve-Karhunen-Transformation oder die einfacher zu realisierende Differenz-PCM an. Bild 36 macht jedoch deutlich, dass durch eine DPCM - Codierung nur bescheidene Gewinne möglich sind (einem 6 dB-Gewinn entspricht eine Einsparung von 1 bit/Koeffizient). Die hier gezeigte DPCM-Codierung wurde unter Verwendung eines Prädiktors mit einem Prädiktorkoeffizienten durchgeführt. Es zeigte sich, dass der Gewinn nur sehr wenig zunimmt, wenn eine Prädiktion mit mehreren Koeffizienten vorgenommen wird, d. h., die den Koeffizientenvektor \underline{h} erzeugende Quelle ist mit recht guter Näherung eine Markoff-

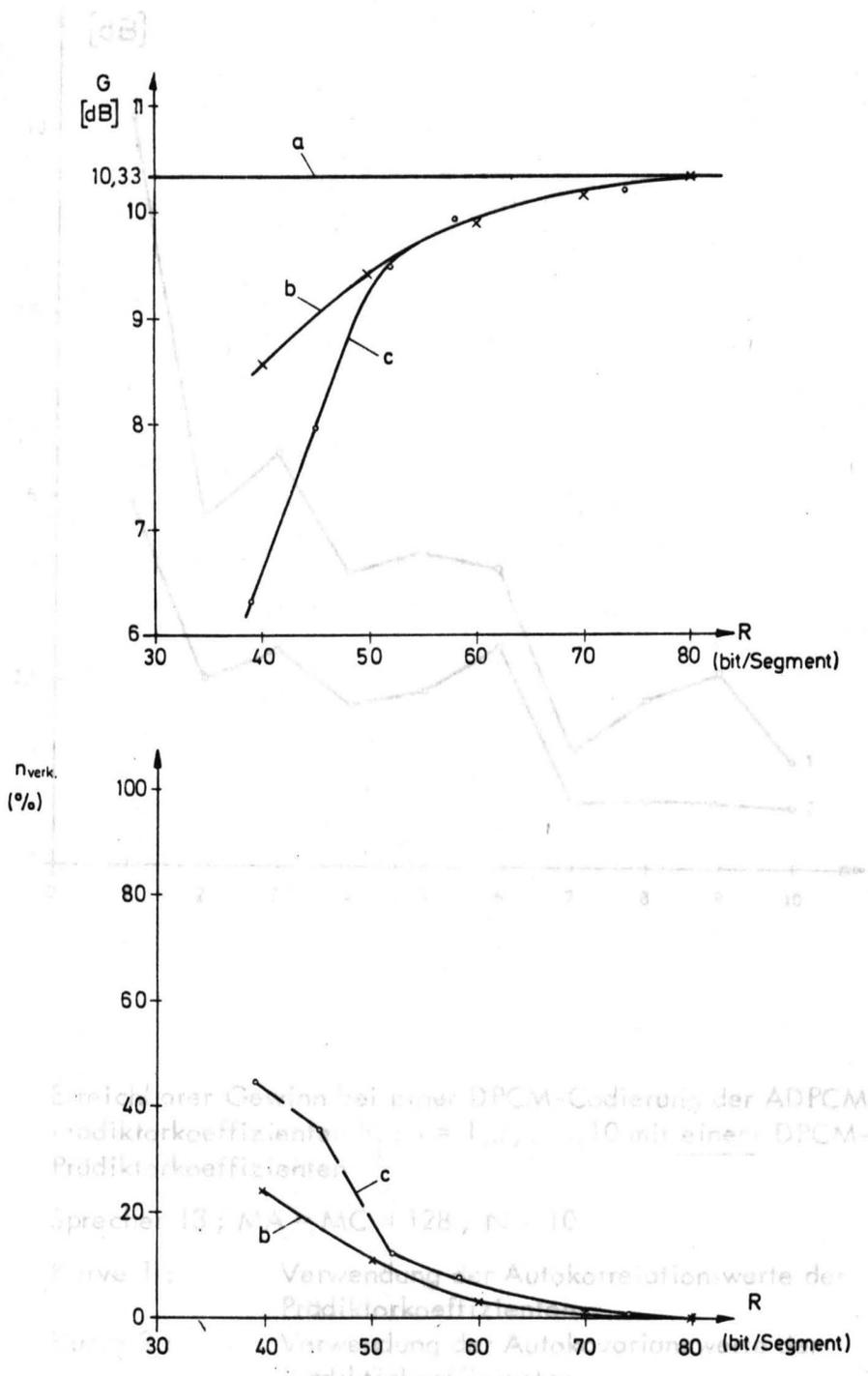


Bild 35: a) Erreichbarer Gewinn bei einer DPCM-Codierung der ADPCM-Prädiktorkoeffizienten h_{i*} ($N = 10$) mit einem DPCM-Prädiktorkoeffizienten $\beta = 1/2$. Sprecher 13; $MA = MC = 128$, $N = 10$.
 Kurve a: Grenzwert ohne Quantisierung des h_{i*} -Vektors
 Kurve b: Lineare Quantisierung jedes Koeffizienten h_{i*} mit $R/10$ bit
 Kurve c: Optimale Quantisierung der AKF-Werte

Bild 35 : a) Gewinn in Abhängigkeit von der zur h_{i*} -Übertragung zur Verfügung stehenden Anzahl R bit/Segment

ADPCM, $N = 10$, Sprecher 13 ; $MA = MC = 128$

Kurve a: Grenzwert ohne Quantisierung des h_{i*} -Vektors

Kurve b: Lineare Quantisierung jedes Koeffizienten h_{i*} mit $R/10$ bit

Kurve c: Optimale Quantisierung der AKF-Werte

b) Anteil der verkürzten Filter

Übertragung (PCM-Übertragung). Eine solche PCM-Übertragung wird immer dann vorgenommen, wenn das Signal eine geringe Energie hat,

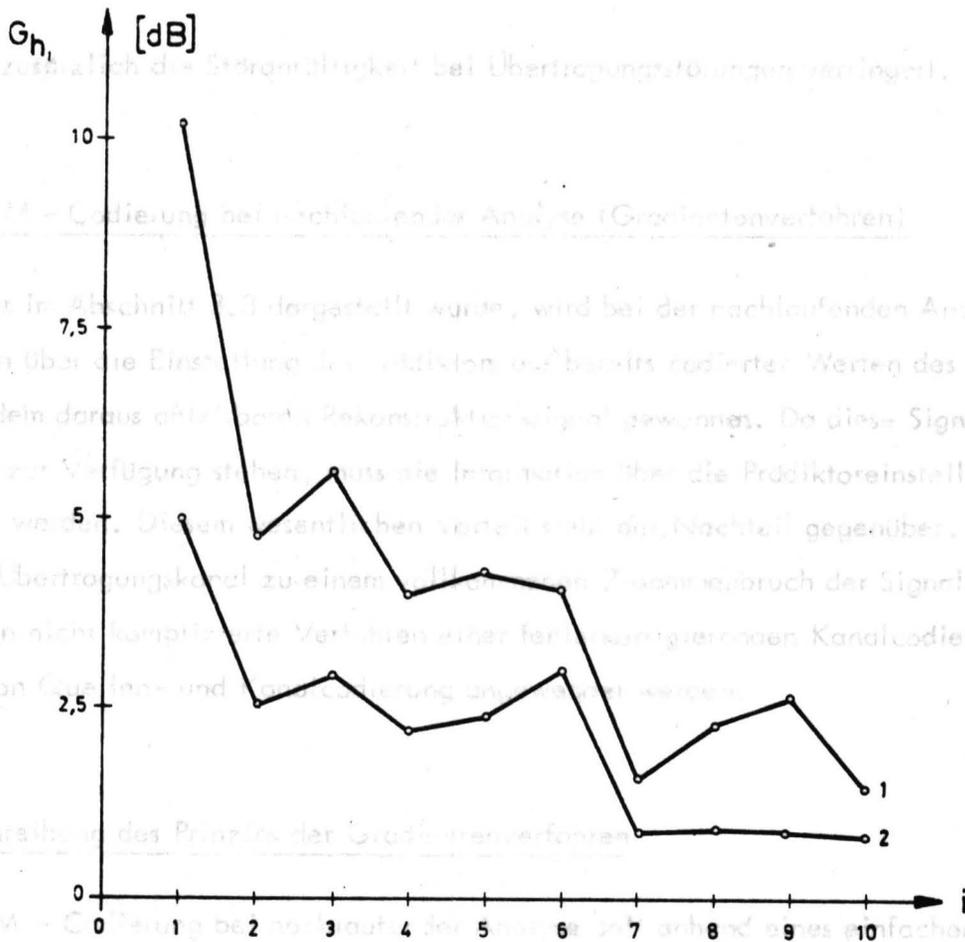


Bild 36 : Erreichbarer Gewinn bei einer DPCM-Codierung der ADPCM-Prädiktorkoeffizienten h_i ; $i = 1, 2, \dots, 10$ mit einem DPCM-Prädiktorkoeffizienten

Sprecher 13 ; MA = MC = 128 ; N = 10

Kurve 1 : Verwendung der Autokorrelationswerte der Prädiktorkoeffizienten

Kurve 2 : Verwendung der Autokovarianzwerte der Prädiktorkoeffizienten

quelle erster Ordnung.

Eine weitere Möglichkeit der Verringerung der zur Informationsübertragung der Prädiktoreinstellung notwendigen Kanalkapazität besteht darin, nur noch eine endliche Zahl von vorgegebenen Prädiktorkoeffizientensätzen \underline{h} zu übertragen, d. h. es muss dann nur noch der Index der nummerierten Sätze \underline{h} übertragen werden. Um den passenden Satz zu finden, muss eine Mustererkennung vorgenommen werden. Nicht passenden Sätzen kann dann der Vektor $\underline{h} = \underline{0}$ zugeordnet werden (PCM-Übertragung). Wird eine solche PCM-Zuordnung auch immer dann vorgenommen, wenn das Signal eine geringe Energie hat,

so wird zusätzlich die Störanfälligkeit bei Übertragungsstörungen verringert.

5. ADPCM - Codierung bei nachlaufender Analyse (Gradientenverfahren)

Wie bereits im Abschnitt 3.3 dargestellt wurde, wird bei der nachlaufenden Analyse die Information über die Einstellung des Prädiktors aus bereits codierten Werten des Differenzsignals bzw. aus dem daraus ableitbaren Rekonstruktionssignal gewonnen. Da diese Signale auch dem Empfänger zur Verfügung stehen, muss die Information über die Prädiktoreinstellung nicht mehr übertragen werden. Diesem wesentlichen Vorteil steht der Nachteil gegenüber, dass jeder Fehler im Übertragungskanal zu einem vollkommenen Zusammenbruch der Signalübertragung führt, wenn nicht komplizierte Verfahren einer fehlerkorrigierenden Kanalcodierung oder Kombination von Quellen- und Kanalcodierung angewendet werden.

5.1. Beschreibung des Prinzips der Gradientenverfahren

Die ADPCM - Codierung bei nachlaufender Analyse soll anhand eines einfachen Gradientenverfahrens gezeigt werden /36/. Die Bilder 16 und 17 zeigen die Struktur eines Systems mit nachlaufender Analyse. Dem Rechner stehen im Zeitpunkt n Rekonstruktionsswerte $r(n-j) = r_{n-j}$; $j = 1, 2, 3, \dots, N$ zur Bestimmung der optimalen Prädiktorkoeffizienten h_j ; $j = 1, 2, 3, \dots, N$ zur Verfügung. Der mittlere quadratische Fehler zwischen der Eingangsfolge $s(n) = s_n$ und der Prädiktor-Ausgangsfolge $\hat{s}(n) = \hat{s}_n$ soll zum Minimum gemacht werden. Bild 37 zeigt die Struktur des Prädiktors.

Es gilt:

$$\hat{s}_n = \sum_{j=1}^N h_j \cdot r_{n-j} = \underline{h}^T \cdot \underline{r}_n \quad (5.1)$$

mit $\underline{h}^T = (h_1, h_2, h_3, \dots, h_n)$ = Koeffizientenvektor

$\underline{r}_n^T = (r_{n-1}, r_{n-2}, \dots, r_{n-N})$ = Vektor des Speicherzustandes des Prädiktors zum Zeitpunkt n

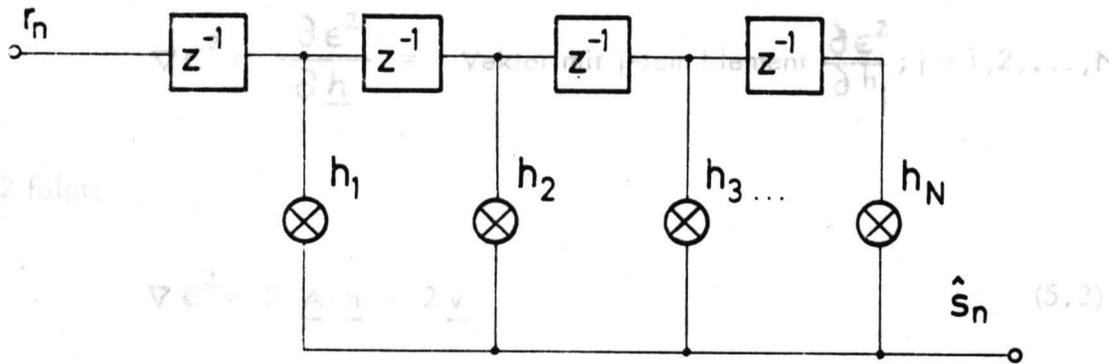


Bild 37: Struktur des Prädiktors

Der Fehler (das Differenzsignal) zum Zeitpunkt n ist $d(n) = d_n = s_n - \hat{s}_n$. Der mittlere quadratische Fehler, berechnet über die letzten $L + 1$ Werte der Lernphase, ist

$$\begin{aligned} \epsilon^2 &= \sum_{k=n-L}^n (s_k - \hat{s}_k)^2 = \sum_{k=n-L}^n (s_k - \underline{h}^T \cdot \underline{r}_k)^2 \\ &= \underline{h}^T \cdot \underline{A} \cdot \underline{h} - 2 \underline{h}^T \cdot \underline{v} + \sum_{k=n-L}^n s_k^2 \end{aligned} \quad (5.2)$$

mit $\underline{A} = \sum_{k=n-L}^n \underline{r}_k \cdot \underline{r}_k^T =$ Autokorrelationsmatrix der Folge der Rekonstruktionswerte r_{k-j} ; $j = 1, 2, 3, \dots, N$

$\underline{v} = \sum_{k=n-L}^n s_k \cdot \underline{r}_k =$ Vektor der Kreuzkorrelationswerte zwischen den Eingangswerten und vorhergehenden Rekonstruktionswerten

ϵ^2 ist eine quadratische Funktion des Koeffizientenvektors \underline{h} und hat damit ein eindeutiges Minimum. Zur Fehlerbestimmung werden die letzten $L + 1$ Werte herangezogen (Lernphase).

$$\epsilon_{*}^2 = \sum_{k=n-L}^n v^2 - \underline{v}^T \underline{A}^{-1} \underline{v} \quad (5.3)$$

Der Gradient des Fehlers ϵ^2 ist

$$\nabla \epsilon^2 = \frac{\partial \epsilon^2}{\partial \underline{h}} = \text{Vektor mit } j\text{-tem Element } \frac{\partial \epsilon^2}{\partial h_j}; j = 1, 2, \dots, N$$

Aus Gl. 5.2 folgt:

$$\nabla \epsilon^2 = 2 \cdot \underline{A} \cdot \underline{h} - 2 \underline{v} \quad (5.3)$$

Bei optimaler Einstellung ist

$$\nabla \epsilon^2 = 0$$

und damit folgt für den optimalen Koeffizientenvektor

$$\underline{h}_* = \underline{A}^{-1} \cdot \underline{v} \quad (5.4)$$

Dieser \underline{h}_* -Vektor, der aus den Daten der Lernphase bestimmt wurde und für diese Lernphase optimal ist, wird zur Codierung des (oder der) nächsten Signalwerte verwendet. Dieses Verfahren entspricht somit dem Prinzip für adaptive Prädiktoren (Balakrishnan /37/). Gleichung 5.4 entspricht der Gleichung 2.5 für stationäre Prozesse, die AKF ist hier jedoch durch Schätzwerte dieser AKF ersetzt worden.

\underline{A} muss positiv-definit sein, d. h. : $\underline{h}^T \cdot \underline{A} \cdot \underline{h} \geq 0$ für alle von Null verschiedenen Vektoren \underline{h} .

Diese Bedingung ist aber erfüllt, da

$$\underline{h}^T \cdot \underline{A} \cdot \underline{h} = \sum_{k=n-L}^n \underline{h}^T \cdot \underline{r}_k \underline{r}_k^T \cdot \underline{h} = \sum_{k=n-L}^n (\underline{h}^T \cdot \underline{r}_k)^2$$

$$= \sum_{k=n-L}^n s_k^2 \geq 0$$

Mit Gleichung 5.4 ergibt sich der minimale Fehler

$$\epsilon_*^2 = (\underline{A}^{-1} \cdot \underline{v})^T \cdot \underline{A} \cdot \underline{A}^{-1} \cdot \underline{v} - 2(\underline{A}^{-1} \cdot \underline{v})^T \underline{v} + \sum_{k=n-L}^n s_k^2$$

$$\epsilon_*^2 = \sum_{k=n-L}^n s_k^2 - \underline{v}^T \cdot \underline{A}^{-1} \cdot \underline{v} \quad (5.5)$$

Mit Hilfe des minimalen mittleren quadratischen Fehlers ϵ_*^2 lässt sich der mittlere quadratische Fehler ϵ^2 (Gleichung 5.2) wie folgt schreiben:

$$\begin{aligned} \epsilon^2 &= \underline{h}^T \cdot \underline{A} \cdot \underline{h} - 2 \underline{h}^T \cdot \underline{v} + \sum_{k=n-L}^n s_k^2 \\ &= \underline{h}^T \cdot \underline{A} \cdot \underline{h} - \underline{h}^T \cdot \underline{v} - \underline{v}^T \cdot \underline{h} + \underline{v}^T \cdot \underline{A}^{-1} \cdot \underline{v} + \epsilon_*^2 \\ &= \underline{h}^T \cdot \underline{A} \cdot \underline{h} - \underline{h}^T \cdot \underline{A} \cdot \underline{h}_* - (\underline{A} \underline{h}_*)^T \cdot \underline{h} + (\underline{A} \underline{h}_*)^T \underline{A}^{-1} \underline{A} \underline{h}_* + \epsilon_*^2 \\ &= \epsilon_*^2 + \underline{h}^T \cdot \underline{A} \cdot \underline{h} - \underline{h}^T \cdot \underline{A} \cdot \underline{h}_* - \underline{h}_*^T \cdot \underline{A} \cdot \underline{h} + \underline{h}_*^T \cdot \underline{A} \cdot \underline{h}_* \end{aligned}$$

Mit $\underline{A} = \underline{A}^T$ folgt:

$$\epsilon^2 = \epsilon_*^2 + (\underline{h} - \underline{h}_*)^T \cdot \underline{A} \cdot (\underline{h} - \underline{h}_*) \quad (5.6)$$

Für $\underline{h} = \underline{h}_*$ wird die quadratische Form zu Null, es ergibt sich das eindeutige Minimum $\epsilon^2 = \epsilon_*^2$.

Das Gradientenverfahren besteht darin, dass der Koeffizientenvektor \underline{h} entweder mit jedem Takt der Abtastfolgen $s(n)$ etc. oder nach einem Vielfachen dieses Taktes geändert wird. Nach der i -ten Änderung ergibt sich ein Fehler:

$$\epsilon_i^2 = \epsilon_*^2 + (\underline{h}_i - \underline{h}_*)^T \cdot \underline{A} \cdot (\underline{h}_i - \underline{h}_*) \quad (5.7)$$

\underline{h}_i ist der Vektor der Prädiktorkoeffizienten nach der i -ten Einstellung des Prädiktors. Die Korrekturen des Vektors \underline{h} werden in Richtung der stärksten Verminderung des mittleren quadratischen Fehlers vorgenommen.

Der Gradientenalgorithmus lautet:

$$\begin{aligned} \underline{h}_{i+1} &= \underline{h}_i - \frac{1}{2} \alpha_i \cdot \nabla \epsilon^2(\underline{h}_i) & ; \quad \alpha_i \geq 0, & \quad (5.8) \\ &= \underline{h}_i - \alpha_i \cdot (\underline{A} \underline{h}_i - \underline{v}) \end{aligned}$$

Der Gradientenkoeffizient α_i bestimmt die Grösse der Änderung in Richtung des Gradienten; er muss so gewählt werden, dass mit möglichst wenig Schritten der minimale Fehler ϵ_*^2

erreicht wird. Um die Korrektur $-\frac{1}{2} \alpha_i \nabla \epsilon^2$ des i -ten Koeffizientenvektors \underline{h}_i vornehmen zu können, muss der Gradient $\nabla \epsilon^2$ bestimmt werden:

$$\begin{aligned} \nabla \epsilon^2 &= \nabla \sum_{k=n-L}^n d_k^2 \\ &= 2 \sum_{k=n-L}^n d_k \cdot \nabla d_k = 2 \sum_{k=n-L}^n d_k \cdot \nabla (s_k - \hat{s}_k) \end{aligned}$$

Mit $\nabla s_k = 0$, Gleichung 5.1 und $\nabla \underline{h}_k^T \underline{r}_k = \underline{r}_k$ wird

(5.10)

$$\nabla \epsilon^2 = -2 \sum_{k=n-L}^n d_k \underline{r}_k$$

bzw.

$$\frac{\partial \epsilon^2}{\partial h_j} = -2 \sum_{k=n-L}^n d_k \cdot r_{k-j} \quad j=1,2,3,\dots,N$$

Der Gradientenvektor kann also als Kreuzkorrelation zwischen der Folge der Differenzwerte und der Folge der Rekonstruktionswerte bestimmt werden.

$$\underline{h}_{i+1} = \underline{h}_i + \alpha_i \sum_{k=n-L}^n d_k \cdot \underline{r}_k \quad (5.9)$$

Da für den Gradienten nur Schätzwerte bestimmt werden können, heisst das Verfahren auch stochastische Approximation.

5.2 Die Anwendung des Verfahrens auf eine ADPCM - Sprachcodierung

Um das Verfahren für die Sprachcodierung und- übertragung verwenden zu können, muss es noch modifiziert werden:

* Da der Algorithmus zeitlichen Schwankungen der statistischen Kenndaten folgen soll, darf der Gradientenkoeffizient α_i nicht, wie bei stationären Prozessen üblich, mit zunehmender Zahl von Approximationsschritten verringert werden. Die einfachste Hardware-Realisierung wird erreicht, wenn der Koeffizient konstant gelassen wird: $\alpha_i = \alpha; \forall_i$.

* Da dem Empfänger die Folge d_n der Differenzwerte nicht bekannt ist, muss diese Folge für die Gradientenbestimmung durch die Folge $\hat{u}(n) = \hat{u}_n$ der quantisierten Differenzwerte ersetzt werden. Beide Folgen unterscheiden sich nur um den Quantisierungsfehler. Sind die Eingangssignalfolge und die Quantisierungsfehlerfolge nicht miteinander korreliert, so ist die Verwendung der Folge \hat{u}_n zulässig.

Damit ergibt sich der Gradientenalgorithmus

$$\hat{h}_{i+1} = \hat{h}_i + \alpha \sum_{k=n-L}^n \hat{u}_k \cdot r_k \quad (5.10)$$

Die Realisierung des Codierers ist im Bild 38 dargestellt.

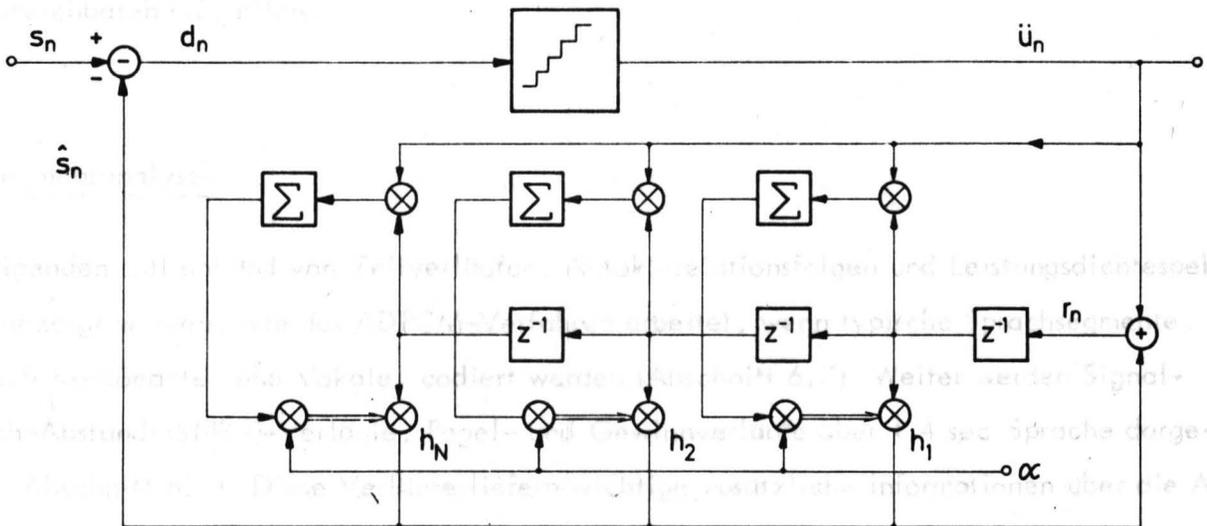


Bild 38 : ADPCM - Codierer mit einer Realisierung des Gradientenalgorithmus nach Gleichung 5.10

6.1 Die Interaktion des Prädiktionsverfahrens im Frequenzbereich
Der mit dem Algorithmus nach Gleichung 5.10 erreichbare Gewinn des ADPCM-Codierers hängt von dem Takt ab, mit dem ein neuer Koeffizientenvektor \underline{h} bestimmt wird; ausserdem beeinflussen die Wahl des Gradientenkoeffizienten α und der Lernphase L die Güte des Systems. Es erwies sich als günstig, die Lernphase grösser zu wählen als den Takt der \underline{h} -Änderung.

Bei einer \underline{h} -Änderung nach jedem 10. Abtastwert und einer Lernphase $L = 64$ ergibt sich ein Gewinn, der um 1 dB unter dem Gewinn liegt, der bei einer ADPCM-Codierung mit vorlaufender Analyse erreichbar ist. Da bei dem letzteren Verfahren ein Verlust von etwa 1 dB durch die zusätzliche Übertragung des Koeffizientenvektors entsteht, sind damit beide Verfahren bezüglich des Gewinns gleichwertig.

Das Gradientenverfahren besitzt eine grosse Zahl von Varianten, von denen einige für eine Hardware-Realisierung sehr interessant sind (so z.B. das Verfahren, bei dem nur noch das Vorzeichen des Gradienten bestimmt wird). Der oben genannte Gewinn dürfte das Maximum des Erreichbaren darstellen.

6. Segmentanalysen

Im folgenden soll anhand von Zeitverläufen, Autokorrelationsfolgen und Leistungsdichtespektren gezeigt werden, wie das ADPCM-Verfahren arbeitet, wenn typische Sprachsegmente, nämlich Konsonanten und Vokale, codiert werden (Abschnitt 6.2). Weiter werden Signal-Rausch-Abstand-(SNR-)-Verläufe, Pegel- und Gewinnverläufe über 1.4 sec Sprache dargestellt (Abschnitt 6.3). Diese Verläufe liefern wichtige zusätzliche Informationen über die Arbeitsweise und die Güte der ADPCM; die alleinige, über ein längeres Sprachsegment gemittelte Gewinnangabe stellt eine zu pauschale Aussage dar. Um den Zusammenhang zwischen den Zeitverläufen und ihren Spektren zu verdeutlichen, werden im vorausgestellten Abschnitt 6.1 die wichtigsten Beziehungen der Prädiktionsverfahren, wie sie sich im Frequenzbereich darstellen, zusammengestellt:

Die gezeigten Verläufe können mit einem am Heinrich-Hertz-Institut entwickelten Programmsystem für Sprachcodierungsverfahren am Display dargestellt werden. Mit dem Programmsystem können unterschiedliche Codierungsverfahren (PCM, DPCM, ADPCM, nicht-adaptive und adaptive Deltamodulation) aufgerufen werden.

6.1 Die Interpretation des Prädikationsverfahrens im Frequenzbereich

Zur Vereinfachung der Diskussion wird bei dieser Interpretation der Einfluss des Quantisierungsrauschens vernachlässigt. Dieser Einfluss kann tatsächlich gesondert betrachtet werden; die Analyse in /4/ zeigt, dass der Quantisierungsfehler am Ausgang des DPCM-Systems ein weisses Spektrum besitzt.

Bleibt also der Quantisierungsfehler unberücksichtigt, so ergibt sich das Prädiktionssystem gemäss Bild 39:

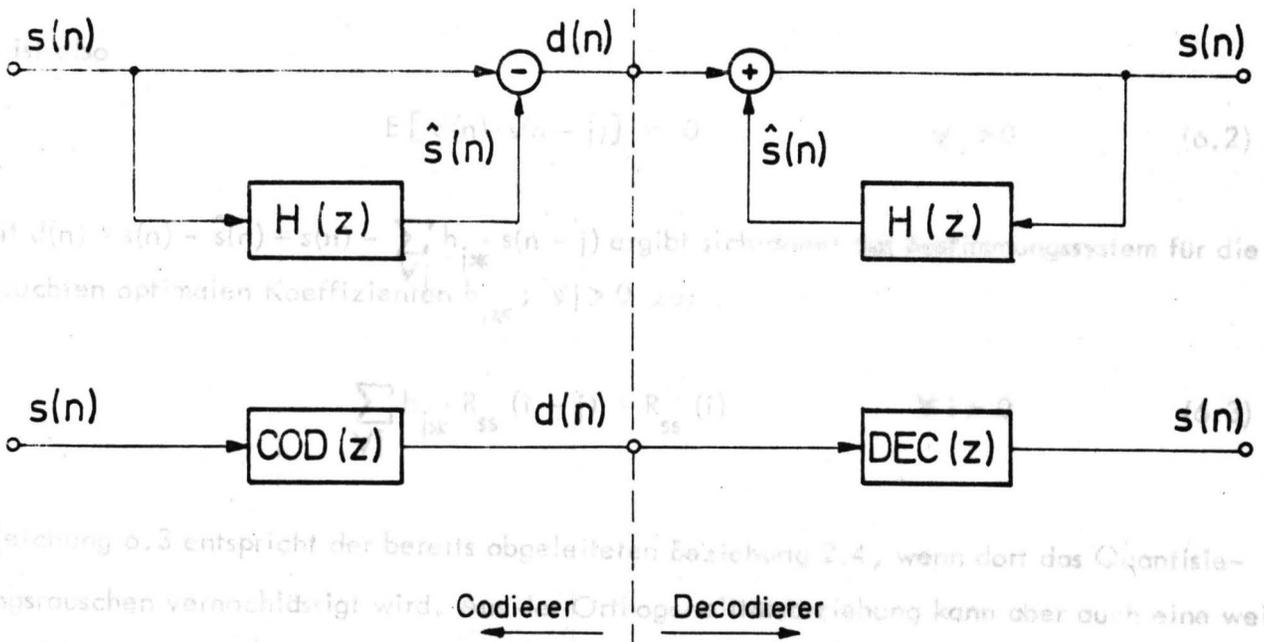


Bild 39: Vereinfachtes Prädiktionssystem (ohne Berücksichtigung des Quantisierers) und sein Ersatzbild

$H(z)$ = Übertragungsfunktion des Prädiktors

$COD(z)$ = Übertragungsfunktion des Prädiktionsfehlerfilters

$DEC(z)$ = Übertragungsfunktion des inversen Prädiktionsfehlerfilters

Der auf der rechten Seite stehende Erwartungswert ist aber wiederum Null (vgl. Gl. 6.2).

Damit ist das Differenzsignal bei optimaler Prädiktion dekorreliert:

$H(z)$ ist die Übertragungsfunktion des Prädiktors:

$$H(z) = \sum_{\forall i > 0} h_i \cdot z^{-i} \quad (6.1)$$

Die Zahl der Prädiktorkoeffizienten wird zuerst als unbeschränkt angenommen. Ist der Prädiktor optimal eingestellt, so gilt das Orthogonalitätsprinzip /5/:

Bei optimaler Filterung ist das Differenzsignal $d(n)$ orthogonal zu allen zur Schätzung von $s(n)$ herangezogenen Beobachtungsdaten, d. h. zu allen im Prädiktor gespeicherten Werten $s(n-i)$; $\forall i > 0$.

Es ist also

$$E [d(n) \cdot s(n - i)] = 0 \quad \forall i > 0 \quad (6.2)$$

Mit $d(n) = s(n) - \hat{s}(n) = s(n) - \sum_{\forall i} h_{i*} \cdot s(n - i)$ ergibt sich damit das Bestimmungssystem für die gesuchten optimalen Koeffizienten h_{i*} ; $\forall i > 0$ zu:

$$\sum_{\forall i} h_{i*} \cdot R_{ss}(i - j) = R_{ss}(j) \quad \forall i > 0 \quad (6.3)$$

Gleichung 6.3 entspricht der bereits abgeleiteten Beziehung 2.4, wenn dort das Quantisierungsrauschen vernachlässigt wird. Aus der Orthogonalitätsbeziehung kann aber auch eine weitere Aussage gewonnen werden: wird in Gleichung 6.2 die Beziehung

$$\begin{aligned} s(n - j) &= d(n - j) + \hat{s}(n - j) \\ &= d(n - j) + \sum_{\forall i} h_{i*} \cdot s(n - i) \end{aligned}$$

eingesetzt, so gilt:

$$E [d(n) \cdot d(n-j)] = - \sum_{\forall i} h_{i*} \cdot E [d(n) \cdot s(n - i)]$$

Der auf der rechten Seite stehende Erwartungswert ist aber wiederum Null (vergl. Gl. 6.2).

Damit ist das Differenzsignal bei optimaler Prädiktion dekorreliert:

ter Sprachsegments bestimmt werden. Diese Bestimmung hat einige Vorteile gegenüber anderen Verfahren (z.B. E [d(n) · d(n-j)] = 0 $\forall j > 0$) (6.4)

Das Differenzsignal hat also ein weisses Spektrum

$$S_{dd}(\Omega) = S_D \tag{6.5}$$

mit $S_D = E [d^2(n)]$

Beim Empfänger wird aus dem Differenzsignal d(·) das ursprüngliche Signal s(·) erzeugt, wobei das Decodiernetzwerk gemäss Bild 39 verwendet wird.

Seine Übertragungsfunktion DEC(z) ist invers zu der Übertragungsfunktion COD(z) des Codierers:

$$DEC(z) = \frac{1}{1 - H(z)} = COD^{-1}(z) \tag{6.6}$$

Da nun dem Decodiernetzwerk ein weisses Rauschen angeboten wird, muss seine Leistungsübertragungsfunktion $|DEC(z)|^2$ mit dem Leistungsdichtespektrum $S_{ss}(z)$ der Eingangsfolge s(n) übereinstimmen.

Bisher wurde nichts über die Zahl der Prädiktorkoeffizienten $h_i; i = 1, 2, 3, \dots, ?$ gesagt. Hier ist nun von Bedeutung, dass Sprachsignalfolgen s(n) näherungsweise als Musterfolgen eines autoregressiven Prozesses endlicher Ordnung aufgefasst werden können. Eine jede dieser Musterfolgen entsteht als Ausgangsfolge eines rekursiven Netzwerkes, wenn dieses mit weissem Rauschen angeregt wird. Hat das Netzwerk K Koeffizienten (autoregressiver Prozess K-ter Ordnung), so wird eine vollkommene Dekorrelation des Differenzsignals d(n) dann erreicht, wenn der Prädiktor H(z) (vergl. Gl. 6.1) $N \gg K$ Koeffizienten h_i besitzt /5/. Bei einer 8 kHz-Abtastung liegt K für Sprachsignale zwischen 8 und 12.

Den bisherigen Betrachtungen lagen stationäre Signalfolgen zugrunde; eine Erweiterung auf nichtstationäre Folgen ist sehr einfach möglich, wenn diese Folgen segmentiert werden und die Betrachtungen auf die aus diesen Segmenten berechenbaren Schätzwerte der AKF und der Leistungsdichtespektren ausgedehnt werden. Aus dem jeweils bestimmten optimalen Koeffizientenvektor h_k kann mit Gleichung 6.6 die Leistungsübertragungsfunktion $|DEC(z)|^2$ des analysier-

ten Sprachsegments bestimmt werden. Diese Bestimmung hat einige Vorteile gegenüber anderen Verfahren (z.B. der homomorphen Filterung):

- * sie ist einfacher durchzuführen, die Rechenzeiten sind kurz;
- * sie liefert ein geglättetes Hüllspektrum ;
- * sie liefert eine ausgeprägte Darstellung der Formanten (der Pole), die für die Sprache von besonderer Bedeutung sind.

Dieser dritte Punkt ergibt sich einfach daraus, dass mit der Bestimmung der $|DEC(z)|^2$ - Leistungsübertragungsfunktion versucht wird, ein Nur - Pole - Modell des Spektrums (also einem autoregressiven Prozess) zu approximieren./5/.

6.2. Analysen einzelner Segmente

Das ADPCM-Verfahren beruht darauf, dass Segmentanalysen durchgeführt werden. Die Prädiktorkoeffizienten h_* beinhalten - wie der letzte Abschnitt gezeigt hat - die Information über das geglättete Kurzzeitspektrum, das Differenzsignal liefert die Information, die notwendig ist, um das Sprachsignal möglichst genau wieder rekonstruieren zu können (bei Vocodern wird diese Zusatzinformation vernachlässigt). Die Koeffizientenvektoren h_* werden aus der AKF bestimmt; Bild 40 zeigt anhand eines Sprachausschnittes, wie sich diese normierte Kurzzeit-AKF von Segment zu Segment ändert (MC = 64, MA = 256, Überdeckung also 4 : 1).

Es wird deutlich, dass sehr unterschiedliche Autokorrelationsfolgen auftreten können. Der anfangs gemessene, fast periodische AKF-Verlauf weist auf hohe Gewinne, aber auch auf Probleme der Stabilität des Decodier-Netzwerkes hin. Für einen Vokal ist die Überlagerung mit einem höher-"frequenten" Anteil typisch ; er weist auf einen ausgeprägten, in höheren Frequenzbereich liegenden Formanten hin. Das ADPCM-Verfahren nutzt diese unterschiedlichen AKF-Verläufe für die Prädiktion aus, während die DPCM stattdessen eine mittlere AKF verwendet. Diese Unterschiede zwischen den Verfahren sollen anhand von zwei Beispielen genauer dargestellt werden :

Es wurde ein Laut ausgewählt, dessen spektrale Eigenschaften in dem verwendeten 32 ms Segment den Langzeit-Eigenschaften von Sprache ähnlich sind. Bild 41 zeigt den Zeitverlauf des Spektrums und die AKF dieses Segmentes. Das Spektrum setzt sich multiplikativ aus den Spektren der Anregungsfunktion und der Vokaltrahales zusammen (vgl. Abschn. 7) und enthält zusätzlich Einflüsse der Segmentbegrenzung. Die logarithmische Darstellung macht das Spektrum der Anregungsfunktion zu einer additiven Störung. Eine Glättung wird

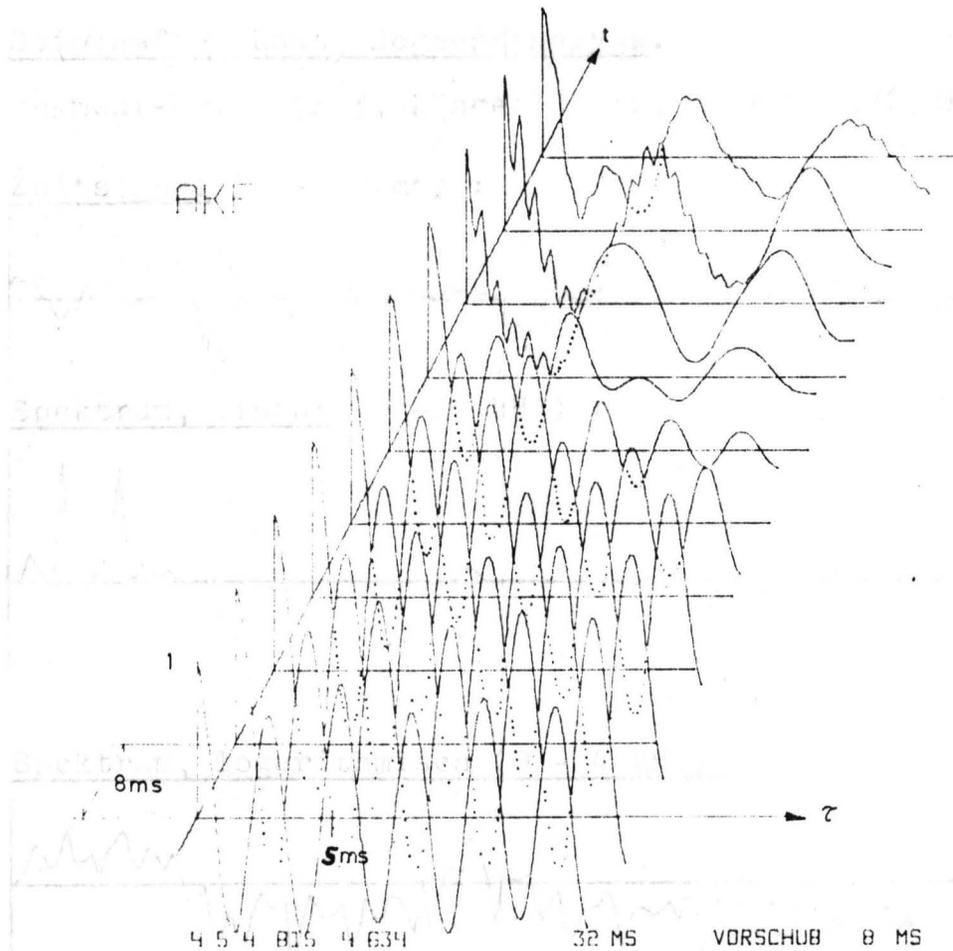


Bild 40 : Darstellung der Änderungen der Kurzzeit-Autokorrelationsfunktionen innerhalb eines 80 ms - Sprachausschnittes.
MC = 64 , MA = 256. Der Verlauf der AKF ist über jeweils 16 ms gezeigt.

6.2.1 Codierung eines stimmhaften Lautes

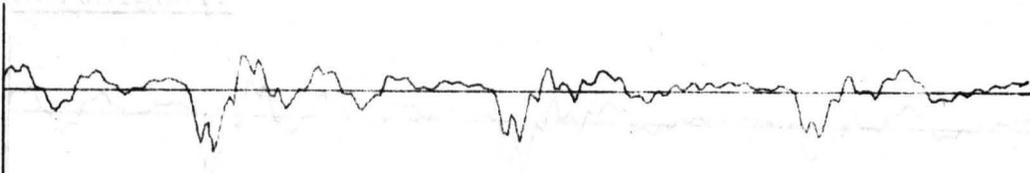
Es wurde ein Laut ausgewählt, dessen spektrale Eigenschaften in dem verwendeten 32 ms-Segment den Langzeit-Eigenschaften von Sprache ähnlich sind. Bild 41 zeigt den Zeitverlauf, das Spektrum und die AKF dieses Segmentes. Das Spektrum setzt sich multiplikativ aus den Spektren der Anregungsfunktion und des Vokaltraktes zusammen (vergl. Abschn. 7) und enthält zusätzlich Einflüsse der Segmentbegrenzung. Die logarithmische Darstellung macht das Spektrum der Anregungsfunktion zu einer additiven Störung. Eine Glättung wird

Bild 41 :

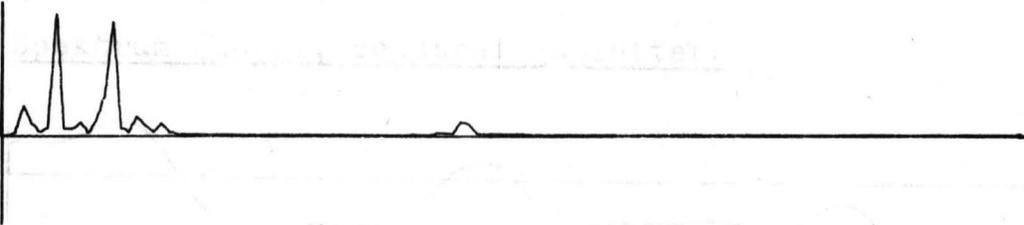
Stimmhafter Laut, Segmentanalyse.

Segment-Nr. 221/13, Länge: 32 ms, Pegel: 0.32 dB

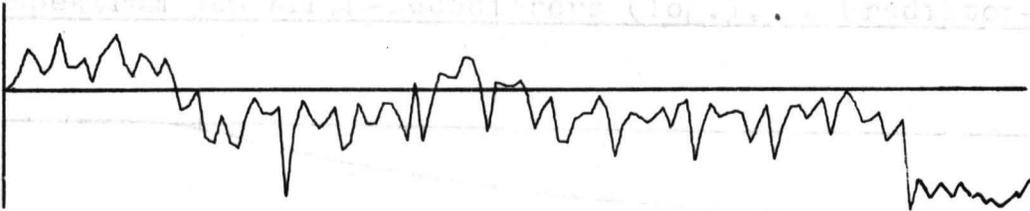
Zeitsignal (0 - 32 ms) :



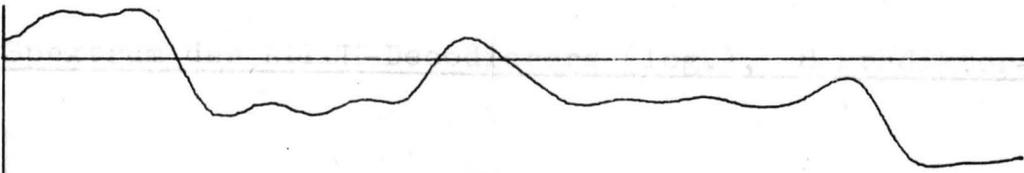
Spektrum, linear (0 - 4 kHz) :



Spektrum, logarithmisch (0 - 4 kHz) :



Spektrum, log., cepstral geglättet (0 - 4 kHz) :



Autokorrelationsfunktion (0 - 16 ms) :

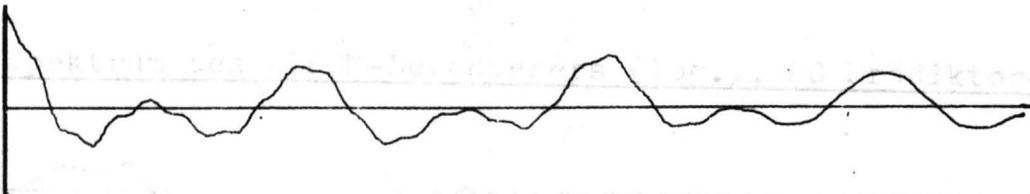


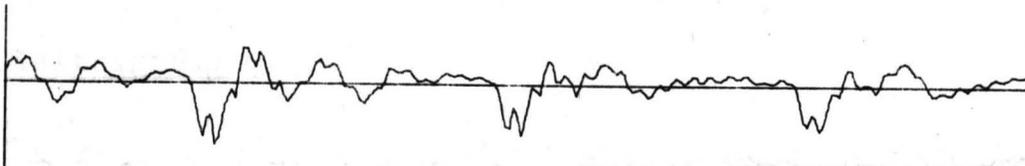
Bild 42 :

Stimmhafter Laut, DPCM-Spektren.

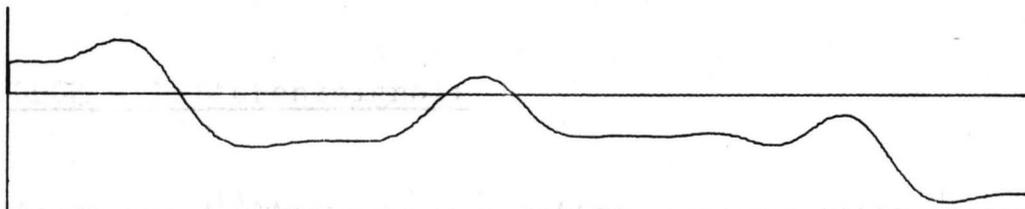
Segment-Nr. 221/13, Länge: 32 ms, Pegel: 0.32 dB.

Darstellung der Spektren: 0 - 4 kHz.

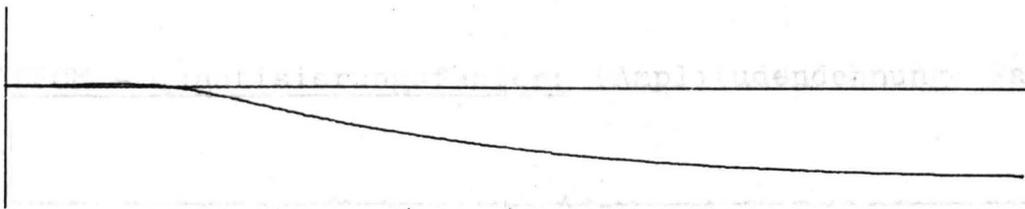
Zeitsignal:



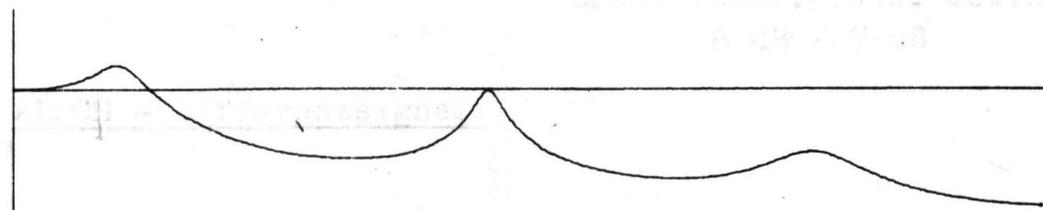
Spektrum (log.), cepstral geglättet:



Spektrum des ADPCM-Decodierers (log.), 2 Prädiktor-Koeff.



Spektrum des ADPCM-Decodierers (log.), 8 Prädiktor-Koeff.



Spektrum des ADPCM-Decodierers (log.), 20 Prädiktor-Koeff.

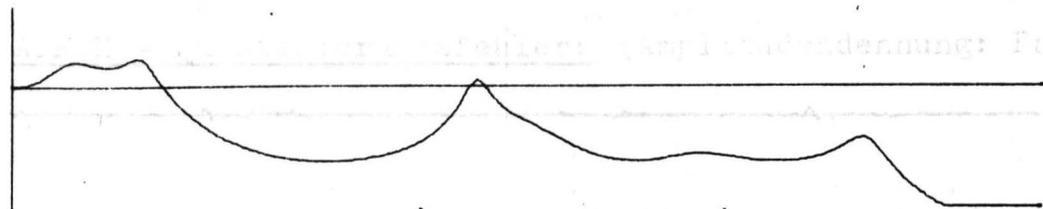


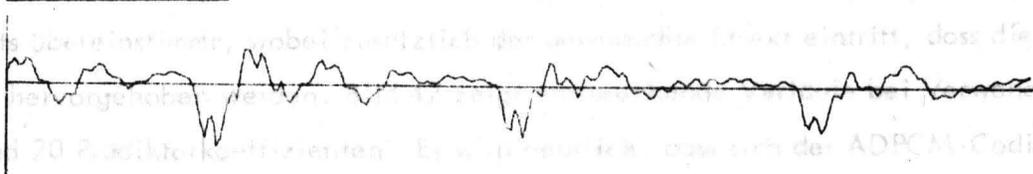
Bild 43 :

Stimmhafter Laut, DPCM-Codierung.

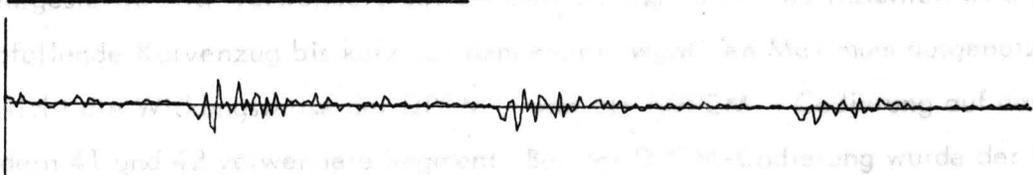
Segment-Nr. 221/13, Länge: 32 ms, Pegel: 0.32 dB

Codierung : Zahl der Quantisiererstufen 32
Zahl der Prädiktorkoeffizienten: 8

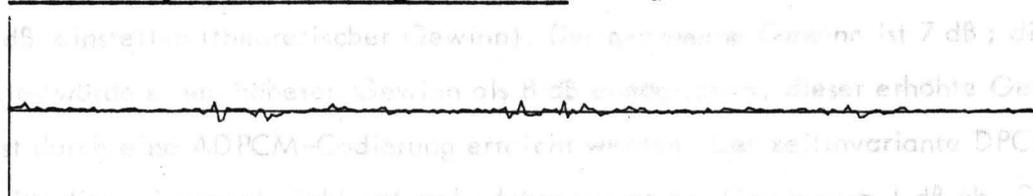
Zeitsignal:



DPCM - Differenzsignal:

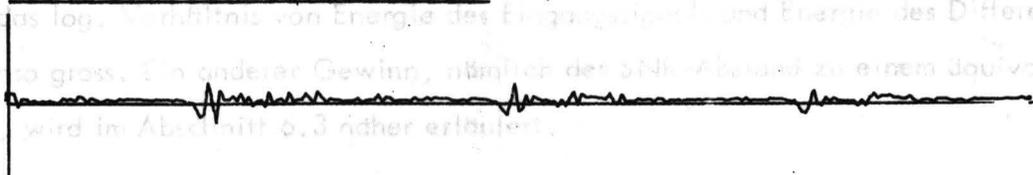


DPCM - Quantisierungsfehler: (Amplitudendehnung: Faktor 5)

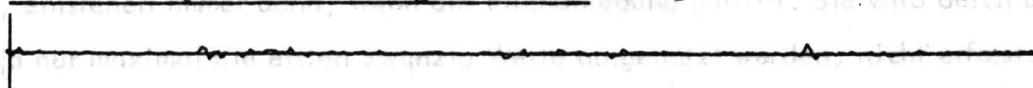


DPCM: theor./tats. Gewinn:
8 dB / 7 dB

ADPCM - Differenzsignal:



ADPCM - Quantisierungsfehler: (Amplitudendehnung: Faktor 5)



ADPCM: theor./tats. Gewinn:
10 dB / 10 dB

durch eine Tiefpassfilterung in Quereffizienzbereich erreicht (Cepstrumglättung, homomorphe Filterung) /38/. Die geglätteten Kurzzeitspektren sind für die Zwecke der Spracherkennung, der Sprachsynthese etc. von grosser Bedeutung. Die Glättung über das Cepstrum ist recht aufwendig, da sie ausser dem Logarithmieren und Entlogarithmieren drei Fourier-Transformationen benötigt. Die Diskussion des Abschnittes 6.1 zeigt nun, dass die Leistungsübertragungsfunktion des ADPCM-Decoderers mit dem geglätteten Kurzzeitspektrum des Sprachsegments übereinstimmt, wobei zusätzlich der gewünschte Effekt eintritt, dass die Pole besonders hervorgehoben werden. Bild 42 zeigt entsprechende Verläufe bei Verwendung von 2, 8 und 20 Prädiktorkoeffizienten. Es wird deutlich, dass sich der ADPCM-Codierer gemäss dem Kurzzeitspektrum des Sprachsegments einstellt, wenn wenigstens 8 Koeffizienten verwendet werden. Zur Bestimmung des optimalen Koeffizientensatzes h_{*} wird die in Bild 41 unten dargestellte AKF verwendet. Bei der Berechnung von 8 Koeffizienten wird dazu nur der erste abfallende Kurvenzug bis kurz vor dem ersten negativen Maximum ausgenutzt. Bild 43 verdeutlicht die Wirkungsweise der DPCM - und der ADPCM - Codierung auf das auch in den Bildern 41 und 42 verwendete Segment. Bei der DPCM-Codierung wurde der Prädiktor der Länge $N = 8$ gemäss der Langzeit-AKF von Sprache eingestellt (vergl. Abschnitt 2). Wäre Sprache ein stationärer Prozess, dann müsste sich für jedes Segment ein Prädiktionsgewinn von etwa 8 dB einstellen (theoretischer Gewinn). Der gemessene Gewinn ist 7 dB; die AKF des Segmentes würde einen höheren Gewinn als 8 dB ermöglichen, dieser erhöhte Gewinn kann aber erst durch eine ADPCM-Codierung erreicht werden. Der zeitinvariante DPCM-Prädiktor ist also für dieses Segment nicht optimal, daher nimmt der Gewinn um 1 dB ab. Der ADPCM-Codierer kann hingegen die mit der Segment-AKF bestimmten linearen Abhängigkeiten voll ausnutzen, da der Prädiktor entsprechend eingestellt wird. Aus der AKF kann ein Gewinn von 10 dB für dieses Segment berechnet werden (theoretischer Gewinn); der tatsächliche Gewinn, das log. Verhältnis von Energie des Eingangssignals und Energie des Differenzsignals, ist ebenso gross. Ein anderer Gewinn, nämlich der SNR-Abstand zu einem äquivalenten PCM-System, wird im Abschnitt 6.3 näher erläutert.

Interessant ist der Zeitverlauf des ADPCM-Differenzsignals: grosse Schätzfehler, also Differenzen, entstehen immer dann, wenn die Pitchanregung auftritt. Sie wird durch die AKF, von denen ja nur maximal die ersten zwanzig Werte ausgenutzt werden, nicht erfasst. Das Differenzsignal ermöglicht auf diese Weise eine Pitch-Extraktion (vergleiche Abschnitt 7).

Bild 44 :

Stimmloser Laut , Segmentanalyse.

Segment-Nr. 211/13, Länge: 32 ms, Pegel: -22.8 dB
Vergrößerung: 20 dB.

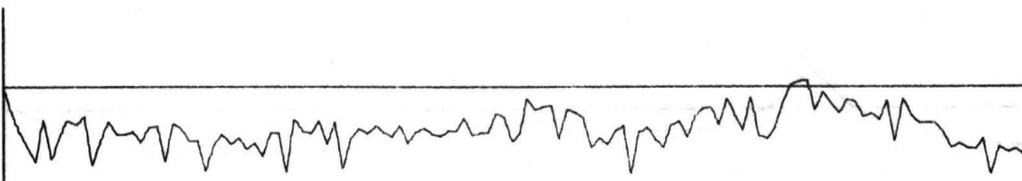
Zeitsignal (0 - 32 ms) :



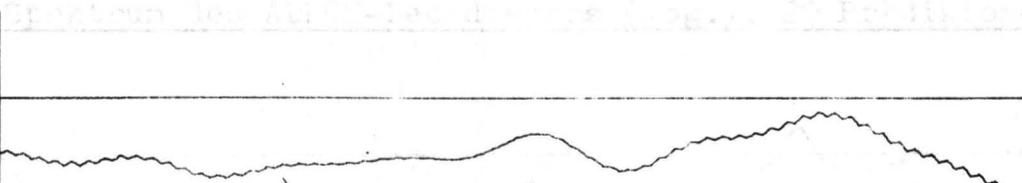
Spektrum, linear (0 - 4 kHz) :



Spektrum, logarithmisch (0 - 4 kHz) : 8 Prädiktor-Koeff.



Spektrum, log., cepstral geglättet (0 - 4 kHz) :



Autokorrelationsfunktion (0 - 16 ms) :

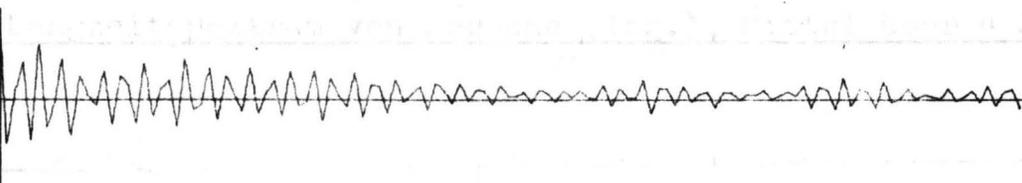


Bild 45 :

Stimmloser Laut, DPCM-Spektren.

Segment-Nr. 211/13, Länge: 32 ms, Pegel: -22.8 dB
Vergrößerung: 20 dB.
Darstellung der Spektren: 0 - 4 kHz.

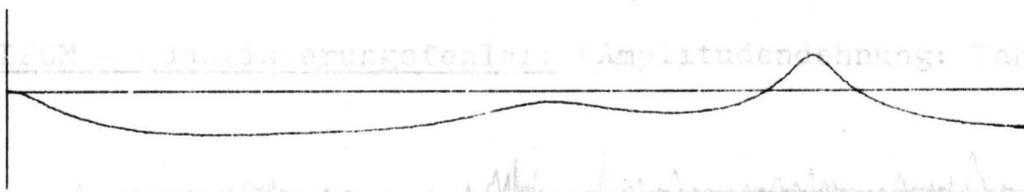
Zeitsignal:



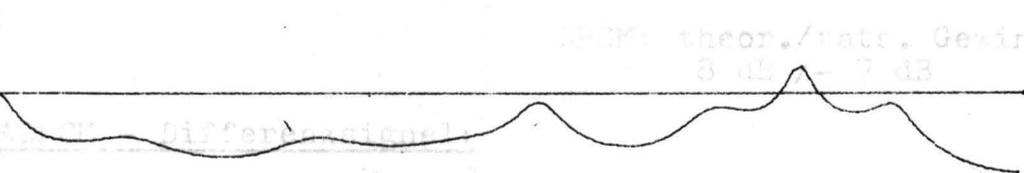
Spektrum (log.), cepstral geglättet:



Spektrum des ADPCM-Decodierers (log.), 8 Prädiktor-Koeff.



Spektrum des ADPCM-Decodierers (log.), 20 Prädiktor-Koeff.



Langzeitspektrum von Sprache (log.), Mittel über 4 Sprecher.



Bild 46 :

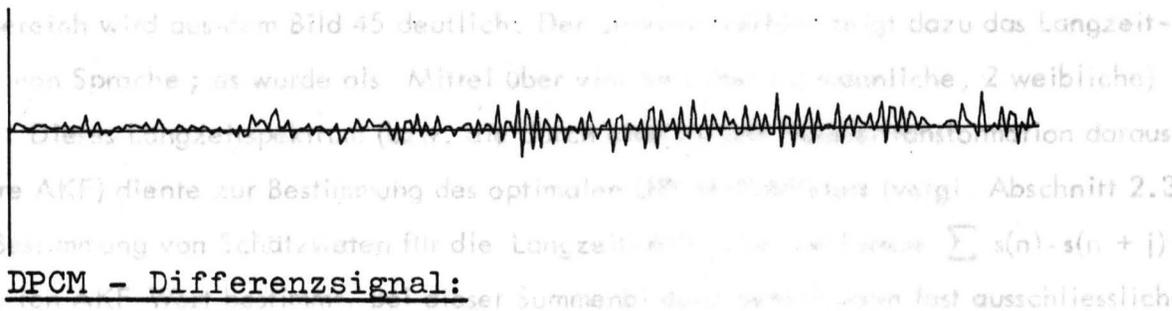
Stimmloser Laut, DPCM-Codierung.

Segment-Nr. 211/13, Länge: 32 ms, Pegel: -22.8 dB

Für den analysierten stimmlosen Laut Vergrößerung: 20 dB.

Codierung : Zahl der Quantisiererstufen: 32
Zahl der Prädiktorkoeffizienten: 8

Zeitsignal:



DPCM - Differenzsignal:



DPCM - Quantisierungsfehler: (Amplitudendehnung: Faktor 5)

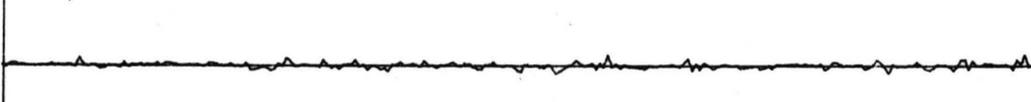


DPCM: theor./tats. Gewinn:
8 dB / - 7 dB

ADPCM - Differenzsignal:



ADPCM - Quantisierungsfehler: (Amplitudendehnung: Faktor 5)



ADPCM: theor./tats. Gewinn:
3 dB / 3 dB

6.2.2 Codierung eines stimmlosen Lautes

Für den analysierten stimmlosen Laut ist kennzeichnend, dass der wesentliche Spektralanteil bei hohen Frequenzen liegt (Bild 44). Ab 3400 Hz tritt dann durch den Einfluss der Tiefpassfilterung eine Absenkung der hohen Frequenzen ein. Die Überlegenheit der Glättung des Kurzzeitspektrums durch Verwendung der Prädiktorkoeffizienten gegenüber der Glättung im cepstralen Bereich wird aus dem Bild 45 deutlich. Der unterste Verlauf zeigt dazu das Langzeitspektrum von Sprache; es wurde als Mittel über vier Sprecher (2 männliche, 2 weibliche) bestimmt. Dieses Langzeitspektrum (bzw. die durch eine inverse Fouriertransformation daraus ableitbare AKF) diente zur Bestimmung des optimalen DPCM-Prädiktors (vergl. Abschnitt 2.3). Bei der Bestimmung von Schätzwerten für die Langzeit-AKF wird die Summe $\sum s(n) \cdot s(n + j)$ für den j -ten AKF-Wert bestimmt. Bei dieser Summenbildung beeinflussen fast ausschliesslich die grossen $s(\cdot)$ - Werte das Ergebnis, kleine $s(\cdot)$ - Werte tragen nur unwesentlich zu der AKF-Bildung ein. Daher ist die Langzeit-AKF im Prinzip eine Aussage über die mittlere AKF von Vokalen, deren Energie sehr viel höher ist als die von Konsonanten und Pausen. Bild 46 zeigt die Konsequenzen dieser Aussage; hier ist das DPCM- und das ADPCM-Differenzsignal für den stimmlosen Laut dargestellt. Bei der DPCM-Codierung ist der Prädiktor gemäss der Langzeit - AKF eingestellt; danach ist er optimal für Spektren mit abnehmenden Frequenzgang (Vergl. Bild 45 unten). Da das Segment jedoch - entgegen der Annahme der Stationarität des Sprachsignals - einen Frequenzanstieg besitzt, ist der Prädiktor falsch eingestellt und führt daher zu einem Differenzsignal, das eine höhere Energie hat als das Eingangssignal. Der Prädiktionsgewinn wird also negativ. Bei der ADPCM-Codierung ist der Prädiktor optimal eingestellt, und es ergibt sich ein positiver Prädiktionsgewinn. Ein negativer Gewinn kann nie auftreten; ist das Leistungsdichtespektrum weiss, so wird der logarithmische Gewinn zu Null ($G_* = 1$, vergl. Gleichung 2.9).

Ein bei einer DPCM-Codierung auftretender negativer Prädiktionsgewinn muss nicht unbedingt zu einer Verschlechterung gegenüber einem vergleichbaren PCM-System führen; dieser Effekt wird im Abschnitt 6.3 näher erläutert.

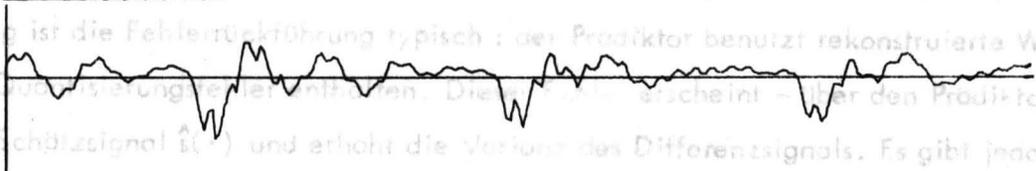
Bild 47 :

Stimmhafter Laut, DPCM-Analyse: Verlauf des Differenzsignals bei grober Quantisierung.

Zahl der Koeff.: 8

Segment-Nr. 221/13, Länge: 32 ms, Pegel: 0.32 dB

Zeitsignal:

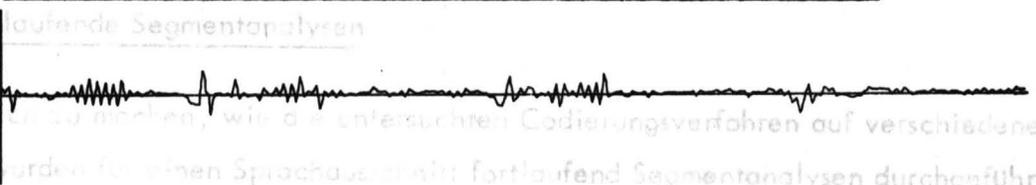


DPCM-Differenzsignal, 8 Quantisiererstufen:



theor./tats. Gewinn: 8 dB/ 2 dB

ADPCM-Differenzsignal, 8 Quantisiererstufen:



theor./tats. Gewinn: 10 dB/ 7 dB

DPCM-Differenzsignal, 4 Quantisiererstufen:



theor./tats. Gewinn: 8 dB/-4 dB

ADPCM-Differenzsignal, 4 Quantisiererstufen:



theor./tats. Gewinn: 10 dB/-1 dB

$E_S = \frac{1}{K} \sum E_S$ = mittlere Energie des Sprachausschnittes, bestehend aus K Segmenten

$10 \lg E_S / E_D$ = Pegel des Differenzsignals

$10 \lg E_S / E_{QP}$ = Signal-Rausch-Abstand (SNR) eines PCM-System

6.2.3 Grobe Quantisierung

Der bei einer ADPCM-Codierung erreichbare Gewinn hängt sehr stark von der Zahl der Quantisierungsstufen ab (vergleiche Abschnitt 4.2.2). Für jede DPCM - bzw. ADPCM-Codierung ist die Fehlerrückführung typisch : der Prädiktor benutzt rekonstruierte Werte, die den Quantisierungsfehler enthalten. Dieser Fehler erscheint - über den Prädiktor bewertet - im Schätzsignal $\hat{s}(\cdot)$ und erhöht die Varianz des Differenzsignals. Es gibt jedoch einen weiteren Grund für die Verringerung des Prädiktionsgewinns bei grober Quantisierung : das DPCM- bzw. ADPCM-Netzwerk kann zu Instabilitäten führen. Diese Instabilitäten werden bei einer 3 bit-Quantisierung schwach sichtbar (Bild 47) ; bei einer 2 bit-Quantisierung sind sie für das Verhalten des Codierverfahrens bestimmend. Das ADPCM-Verfahren ist daher nur für Anwendungen mit genügend feiner Quantisierung sinnvoll ($n \geq 4$ bit/ Abtastwert).

6.3 Fortlaufende Segmentanalysen

Um deutlich zu machen, wie die untersuchten Codierungsverfahren auf verschiedene Laute wirken, wurden für einen Sprachausschnitt fortlaufend Segmentanalysen durchgeführt ; die Zeitabhängigkeit der Kennwerte dieser Analysen wurde graphisch dargestellt. Für jedes der K Segmente wurden die folgenden Grössen gemessen:

E_S = Energie der Eingangsfolge

E_D = Energie der Differenzfolge

E_{QD} = Energie der Quantisierungsfehlerfolge (DPCM, ADPCM)

E_{QP} = Energie der Quantisierungsfehlerfolge (PCM)

	SNR	SNR
DPCM N=3	27.8 dB	7.1 dB
ADPCM N=10	32.2 dB	11.5 dB
ADPCM N=10 (optimaler Laplace-Quantisierer)	34.8 dB	14.1 dB

Aus diesen Messungen wurden die folgenden Daten abgeleitet:

$\overline{E_S} = \frac{1}{K} \sum E_S$ = mittlere Energie des Sprachausschnittes, bestehend aus K Segmenten

$10 \cdot \lg E_S / \overline{E_S}$ = Pegel des Eingangssignals

$10 \cdot \lg E_D / \overline{E_S}$ = Pegel des Differenzsignals

$10 \cdot \lg E_S / E_{QP}$ = Signal-Rausch-Abstand (SNR) eines PCM-Systems

5-bit-Differenz-PCM mit drei Prädiktorkoeff.

$$10 \cdot \lg E_S / E_{QD} = \text{Signal-Rauschabstand (SNR) eines DPCM-bzw. ADPCM-Systems}$$

$$\overline{\text{SNR}} = 10 \lg \sum E_S / \sum E_{QD} = \text{mittlerer SNR - Wert}$$

$$10 \cdot \lg E_S / E_D = \text{Prädiktionsgewinn}$$

$$10 \cdot \lg E_{QP} / E_{QD} = \text{SNR-Gewinn (Gewinn im Signal-Rausch-Abstand gegenüber einem PCM - System)}$$

$$G_{\text{SNR}} = 10 \lg \sum E_{QP} / \sum E_{QD} = \text{mittlerer SNR-Gewinn}$$

— Differenzsignal
.... Eingangssignal

In den folgenden Bildern 48 - 50 werden die wichtigsten Pegel-, SNR- und Gewinnverläufe dargestellt. Der Sprachausschnitt von 1.44 sec Länge umfasst die Worte:

Das Herz Münchens .

SNR-Verläufe:
— DPCM
.... PCM
SNR = 27.8 dB.

Die Segmente sind 16 ms lang (MA = MC = 128); es wurden also K = 90 Segmente erfasst. Vier Verfahren wurden miteinander verglichen:

			$\overline{\text{SNR}}$	G_{SNR}
PCM		log. Quantisierer ($\mu = 100, 5 \text{ bit}$)	20.7 dB	0
DPCM	N = 3	" " " "	27.8 dB	7.1 dB
ADPCM	N = 10	" " " "	32.2 dB	11.5 dB
ADPCM	N = 10	adaptiver Laplace-Quantisierer	34.8 dB	14.1 dB

Um Übersteuerungseffekte zu vermeiden, wurde der logarithmisch abgestufte Quantisierer nur schwach angesteuert (16 σ -Belastung).

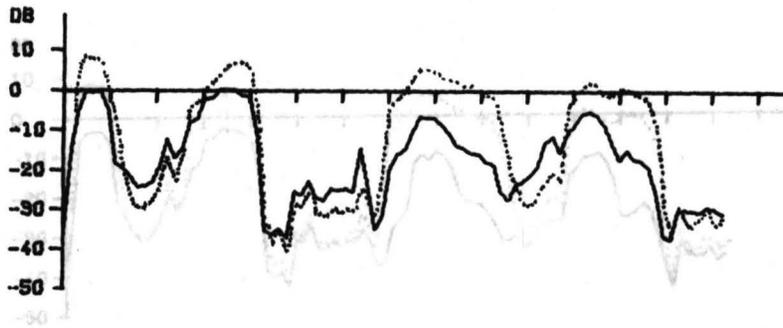
Aus den Abbildungen 48 - 50 lassen sich einige wichtige Ergebnisse ablesen:

PCM (Bild 48):

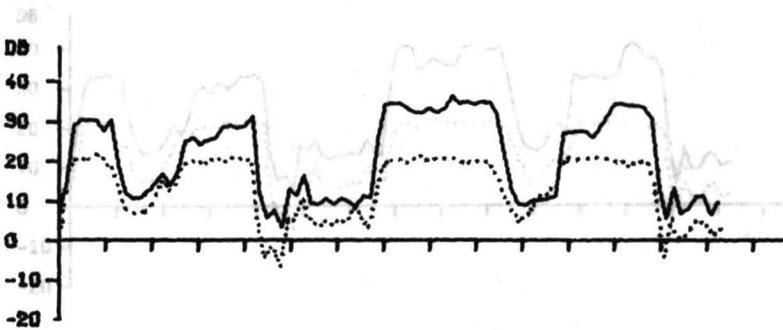
Der SNR-Verlauf zeigt trotz der logarithmischen Kennlinie des Quantisierers starke Einbrüche, wenn Pausen oder Konsonanten auftreten. Der mittlere SNR-Wert $\overline{\text{SNR}}$ wird nur bei den Vokalen erreicht, die eine hohe Energie besitzen.

Bild 48 : 5-bit-Differenz-PCM mit drei Prädiktorkoeff.

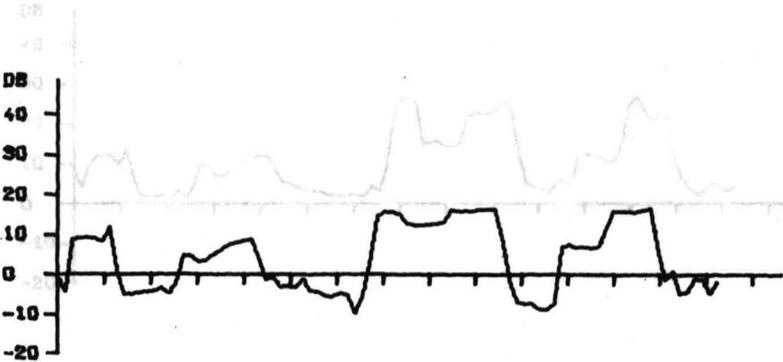
Log. Quantisierer ($\mu=100$, 16 σ -Bel.)



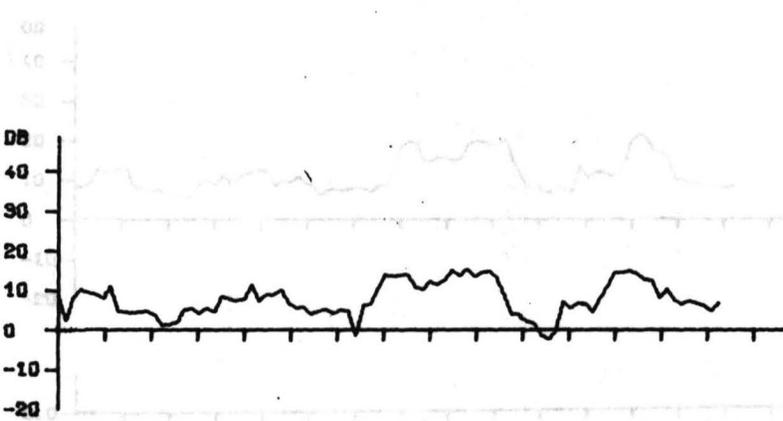
Pegelverläufe:
Pegelverläufe: zsignal
— Differenzsignal
.... Eingangssignal



SNR-Verläufe:
SNR-Verläufe:
— DPCM
.... PCM
SNR= 32.2 dB.
SNR= 27.8 dB.



Prädiktionsgewinn.
Prädiktionsgewinn.



SNR-Gewinn gegen-
über PCM.
SNR-Gewinn gegen-
über PCM.

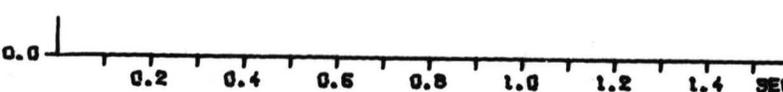
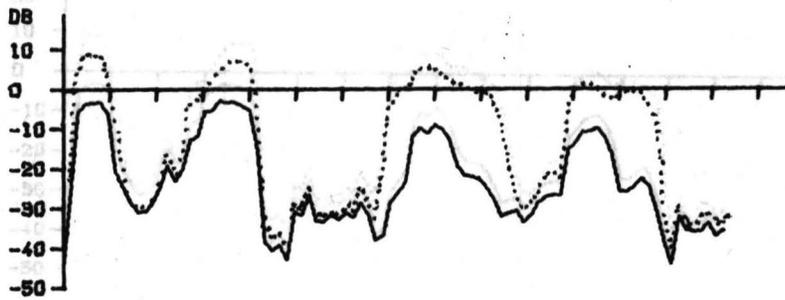


Bild 49: Adaptive Differenz-PCM mit zehn Prädiktorkoeff.

Log. Quantisierer ($\mu=100$), 5 bit, 166 -Bel.

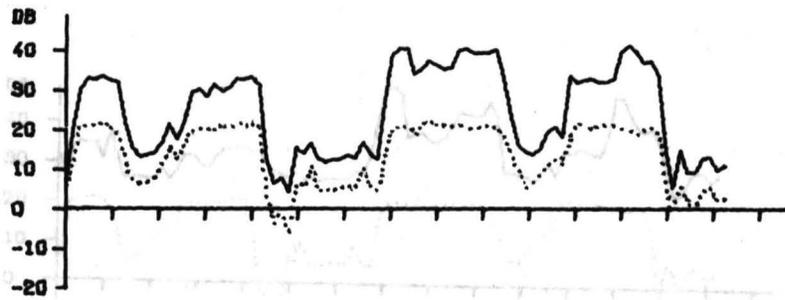


Pegelerläufe:

— Differenzsignal

.... Eingangssignal

.... Eingangssignal



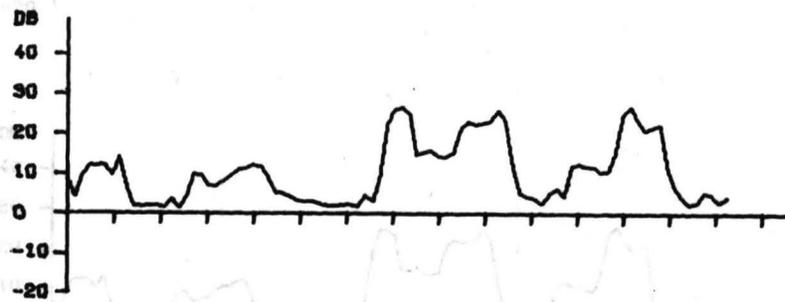
SNR-Verläufe:

— ADPCM

.... PCM

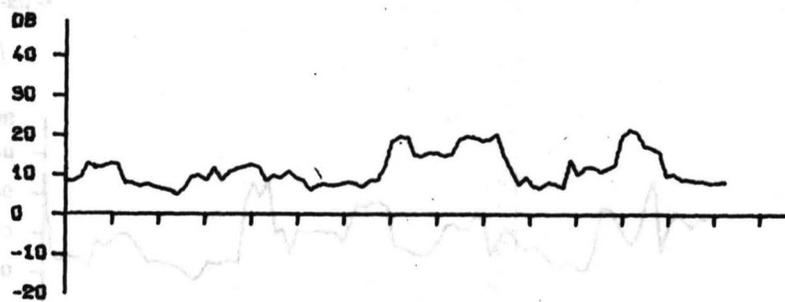
SNR= 32.2 dB.

SNR= 34.8 dB



Prädiktionsgewinn.

Prädiktionsgewinn.



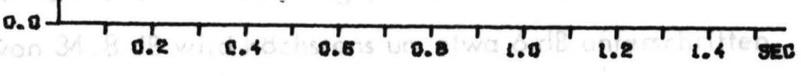
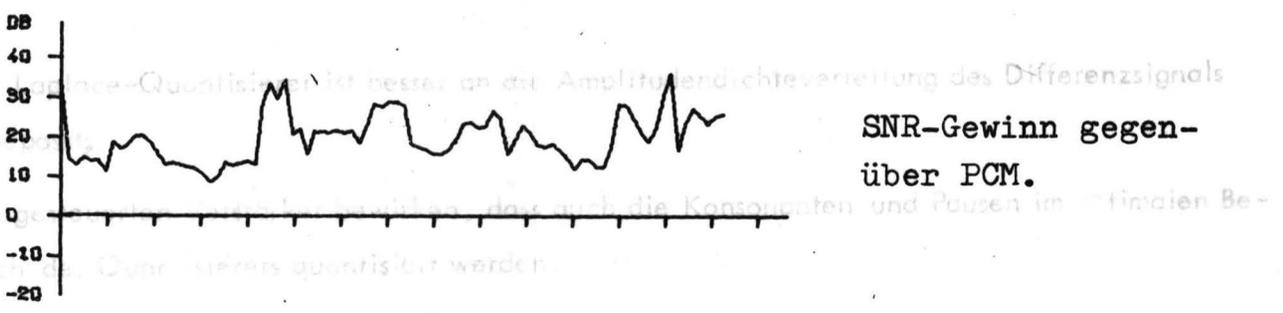
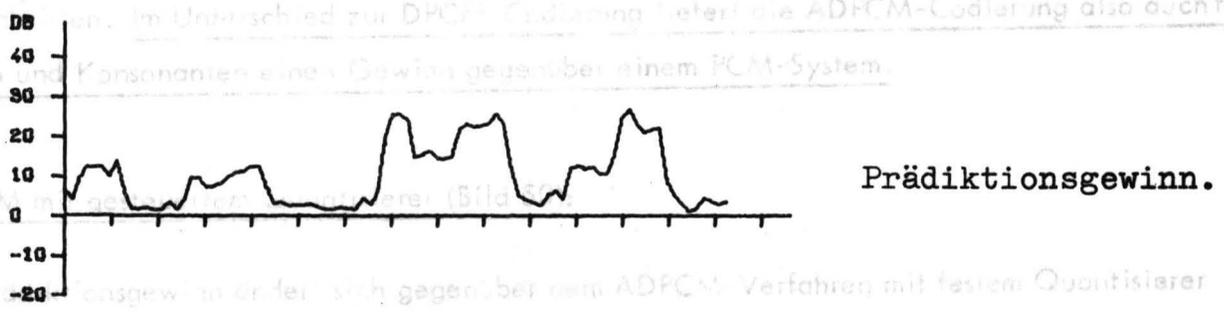
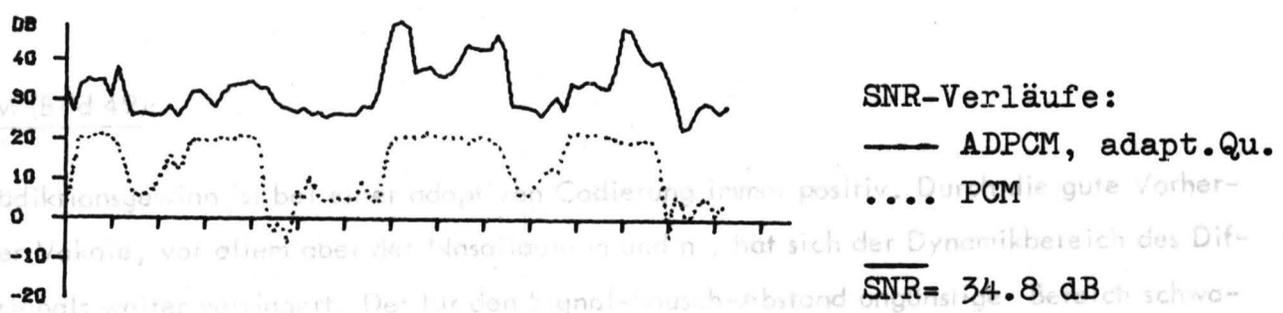
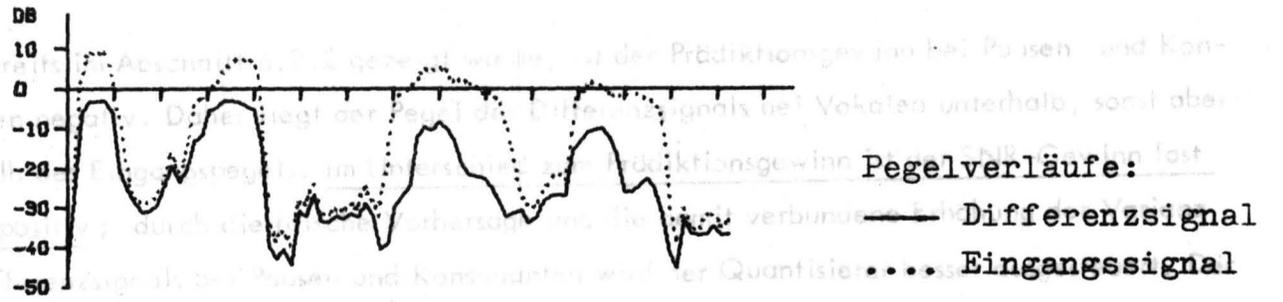
SNR-Gewinn gegen-
über PCM.

SNR-Gewinn gegen-
über PCM.

0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4 SEC

0.2 0.4 0.6 0.8 1.0 1.2 1.4 SEC

Bild 50: Adaptive Differenz-PCM mit zehn Prädiktor-Koeff.
Adaptiver Laplace-Quantisierer, 5 bit.



Wie bereits Abschn. 2.2 gezeigt wurde, ist der Prädiktionsgewinn bei Pausen und Konsonanten gering. Dies liegt der Regel der Laplace-Quantisierung bei Vokalen unterhalb, sonst aber oberhalb der Spracherzeugung im Frequenzbereich der Prädiktionsgewinn fast immer positiv; durch die adaptive Quantisierung des adaptiven Laplace-Quantisierers verbundenen Frequenzbereich des Differenzsignals bei Pausen und Konsonanten wird der Quantisiererrausch vermindert, der Quantisierer arbeitet also nicht mehr in einem ungenutzten Bereich der dynamischen Aussteuerung.

Nur Prädiktionsgewinn bleibt aber adaptiv im Codierbereich positiv. Durch die gute Vorhersage des Signals, vor allem aber der Nasalvokale, hat sich der Dynamikbereich des Differenzsignals weiter vermindert. Der für den Signal-Rausch-Abstand zugewandene Bereich schwächer aussteuert, als zuvor noch benutzt, daher wird ein SNR-Gewinn von etwa 6 dB nicht unterschätzt. Im Unterschied zur DPCM-Codierung liefert die ADPCM-Codierung also auch für Pausen und Konsonanten einen Gewinn gegenüber einem PCM-System.

Der Prädiktionsgewinn erhebt sich gegenüber dem ADPCM-Verfahren mit festem Quantisierer nicht, jedoch tritt eine starke Erhöhung des SNR-Gewinns auf. Dafür sind zwei Gründe massgebend:

1) Der adaptive Laplace-Quantisierer ist besser an die Amplitudenverteilung des Differenzsignals angepasst.

2) Die adaptive Quantisierung des adaptiven Laplace-Quantisierers, das auch die Konsonanten und Pausen im Frequenzbereich des Differenzsignals quantisiert werden.

Der ADPCM-SNR-Gewinn zeigt, was mit diesem Verfahren erreichbar ist: der mittlere SNR-Wert von 34,8 dB ist ein Wert, der bei einer PCM-Codierung sehr viel grössere Unterschreitungen des mittleren SNR-Wertes aufweisen. Stimmlose Laute haben jetzt einen sehr hohen Signal-Rausch-Abstand; er wirkt sich zwar nicht in der Angabe des mittleren SNR-Wertes aus, weil die Energie dieser Laute sehr gering ist, bewirkt aber eine starke

DPCM (Bild 48):

Wie bereits im Abschnitt 6.2.2 gezeigt wurde, ist der Prädiktionsgewinn bei Pausen und Konsonanten negativ. Daher liegt der Pegel des Differenzsignals bei Vokalen unterhalb, sonst aber oberhalb des Eingangspiegels. Im Unterschied zum Prädiktionsgewinn ist der SNR-Gewinn fast immer positiv; durch die falsche Vorhersage und die damit verbundene Erhöhung der Varianz des Differenzsignals bei Pausen und Konsonanten wird der Quantisierer besser angesteuert. Der Dynamikbereich des zu quantisierenden Signals hat sich verringert, der Quantisierer arbeitet also nicht mehr in dem ungünstigen Bereich schwacher Aussteuerung.

ADPCM (Bild 49):

Der Prädiktionsgewinn ist bei einer adaptiven Codierung immer positiv. Durch die gute Vorhersage der Vokale, vor allem aber der Nasallaute m und n, hat sich der Dynamikbereich des Differenzsignals weiter verringert. Der für den Signal-Rausch-Abstand ungünstige Bereich schwacher Aussteuerung wird kaum noch benutzt; daher wird ein SNR-Gewinn von etwa 6 dB nicht unterschritten. Im Unterschied zur DPCM-Codierung liefert die ADPCM-Codierung also auch für Pausen und Konsonanten einen Gewinn gegenüber einem PCM-System.

ADPCM mit gesteuertem Quantisierer (Bild 50):

Der Prädiktionsgewinn ändert sich gegenüber dem ADPCM-Verfahren mit festem Quantisierer nicht, jedoch tritt eine starke Erhöhung des SNR-Gewinns auf. Dafür sind zwei Gründe maßgebend:

- Der Laplace-Quantisierer ist besser an die Amplitudendichteverteilung des Differenzsignals angepasst;
- Die gesteuerten Verstärker bewirken, dass auch die Konsonanten und Pausen im optimalen Bereich des Quantisierers quantisiert werden.

Der ADPCM-SNR-Verlauf zeigt, was mit diesem Verfahren erreichbar ist: der mittlere SNR-Wert von 34.8 dB wird höchstens um etwa 6 dB unterschritten, während bei einer PCM-Codierung sehr viel grössere Unterschreitungen des mittleren SNR-Wertes auftreten. Stimmlose Laute haben jetzt einen sehr hohen Signal-Rausch-Abstand; er wirkt sich zwar nicht in der Angabe des mittleren SNR aus, weil die Energie dieser Laute sehr gering ist, bewirkt aber eine starke

subjektive Verbesserung. Der mittlere SNR-Gewinn $G_{\text{SNR}} = 14.1$ dB gibt ein viel zu pessimistisches Bild von der Güte des Verfahrens. Das zeigt auch der Verlauf des SNR-Gewinns (unterste Kurve in Bild 50): der mittlere SNR-Gewinn $G_{\text{SNR}} = 14.1$ dB wird ganz selten unterschritten, der Segmentgewinn liegt aber sehr oft weit oberhalb des Mittelwertes.

7. Der Zusammenhang zwischen der ADPCM-Codierung und dem Vocoderprinzip

7.1. Vocoder-Prinzipien

Sprache ist ein stochastischer Prozess und kann deshalb durch eine Quelle beschrieben werden, die ein weisses Rauschen abgibt, das durch ein - im einfachsten Fall - lineares System gefärbt wird. Das lineare System kann wahlweise angesetzt werden als

- * ein nichtrekursives Filter : die Abtastwerte $s(n)$ sind dann jeweils eine Linearkombination aus den vergangenen Rauschwerten (moving-average-Prozess);
- * ein rein rekursives Filter : die Abtastwerte $s(n)$ sind dann jeweils eine Linearkombination aus vergangenen Abtastwerten $s(n-1), s(n-2), \dots$ (autoregressiver Prozess);
- * ein gemischt - rekursives Filter.

7.2. ADPCM-Codierung und Prädiktionsvocoder

Der physikalische Entstehungsprozess von Sprache hat nun Parallelen zu der formalen Aufteilung, wie sie eben skizziert wurde: das lineare System wird durch den Vokaltrakt gebildet, der von einem durch den Kehlkopf gesteuerten Luftstrom angeregt wird. Für stimmlose Laute ist diese Anregung ein weisses Rauschen, für stimmhafte Laute ist die Anregung näherungsweise eine Impulsfolge (pitch). Der Impulsabstand ist dabei gross gegenüber dem Abtastabstand der $s(n)$ -Folge (3 - 12 ms pitch-Abstand gegenüber $125 \mu\text{s}$ Abtastabstand).
Bei einer Codierung von Sprache wird die Musterfolge $s(n)$ und damit ihr Spektrum $S(z)$ übertragen. Bei einer Vocoder-Übertragung wird bei der Analyse das Spektrum $S(z)$ in zwei Komponenten $A(z)$ (Anregung) und $V(z)$ (Vokaltrakt) aufgespalten. Die Komponenten $A(z)$ und $V(z)$ werden getrennt übertragen und bei der Synthese zum Spektrum $S(z) = A(z) \cdot V(z)$ zusammengesetzt. Die Übertragungsfunktion $V(z)$ wird i. a. durch das Energiedichtespektrum $|V(z)|^2$ ersetzt; es kann näherungsweise als Hüllkurve des Kurzeitspektrums bestimmt werden. Die Information über das Spektrum kann auf verschiedene Arten übertragen werden /38/:

- * Stützwerte der Hüllkurve (Kanalvocoder).
- * Tiefqufrequenter Anteil des Cepstrums ; (homomorpher Vocoder) .
- * Formanten (Pole) und Antiformanten (Nullstellen) ; (Formant - Vocoder) .
- * AKF als inverse Fouriertransformierte der Hüllkurve des Kurzzeitspektrums ; (Autokorrelations-Vocoder) .
- * Prädiktorkoeffizienten ; (Prädiktions-Vocoder, vergl. Abschnitt 7.2) .

Um die gewünschte Informationsreduktion zu erreichen, wird die Anregungsfunktion $A(z)$ nur für stimmhafte Laute als Impulsfolge übertragen. Für stimmlose Laute ist die Anregungsfunktion $A(z)$ näherungsweise ein weisses Rauschen ; es wird nicht übertragen, sondern bei der Synthese durch ein künstlich erzeugtes Rauschen gleicher Varianz ersetzt. Da die Information über den Vokaltrakt in jedem Fall übertragen wird, kann der Vokaltrakt bei der Sprachsynthese als lineares System nachgebildet werden und durch die pitch oder durch das weisse Rauschen angeregt werden. Im letzteren Fall entsteht am Ausgang des Synthetisators ein Signal, dessen Signalform nicht mit dem analysierten Signal übereinstimmt ; die Leistungsdichtespektren stimmen aber näherungsweise überein.

7.1 Der ADPCM-Vocoder

7.2 ADPCM-Codierung und Prädiktionsvocoder

Bei der ADPCM - Codierung von Sprachsignalen wird der Koeffizientenvektor \underline{h} als Information über das Spektrum des Vokaltraktes übertragen. Wie im Abschnitt 6.2 gezeigt wurde, beschreibt das mit \underline{h} gebildete rein rekursive Decodierfilter die Hüllkurve des Kurzzeitspektrums, wobei die besonders wichtigen Maxima der Hüllkurve sehr genau dargestellt werden /24,25/. Der Koeffizientenvektor \underline{h} enthält zusätzlich aber i. a. auch einige Informationen über die Anregung (reelle Pole, vergl. Bild 32). Zusätzlich wird das Differenzsignal übertragen ; es enthält alle restlichen Informationen über die Anregung des Vokaltraktes, weiterhin aber alle Informationen über die Abweichungen von dem angenommenen Modell der akustischen Quelle (einem autoregressiven Prozess).

Beim Prädiktions-Vocoder /21,38/ wird auch der Koeffizientenvektor \underline{h} als Information über den Vokaltrakt übertragen. Statt des Differenzsignals werden dem Empfänger jedoch nur noch rudimentäre Informationen mitgeteilt: für stimmlose Laute ist dies die Varianz des Differenzsignals,

für stimmhafte Laute sind dies Informationen über die pitch. Es wird damit die Annahme gemacht, dass stimmhafte Laute eine Folge von Impulsantworten des Vokaltraktes sind. Diese für jedes Vocoderprinzip wesentliche Annahme führt auf zwei Probleme:

- * Die pitch muss bei der Analyse aus dem Sprachsignal extrahiert werden; es handelt sich um eine nicht einfache Aufgabe der Mustererkennung. Das Prädiktionsverfahren scheint für diese Aufgabe recht geeignet zu sein /39/; wie Bild 43 zeigt, ist die pitch im Differenzsignal gut erkennbar. Leider ist diese Erkennbarkeit für andere stimmhafte Laute nicht so gut.
- * Die Anregung des Decodierfilters mit einer Impulsfolge kann zu instabilem Verhalten bei der Synthese führen./21/.

Die Verfahren der ADPCM-Codierung und des Prädiktions-Vocoders verwenden das gleiche Verfahren der h - Koeffizientenbestimmung. Die Vocoderformulierung der h - Bestimmung nach Abschnitt 4.1.2 ist mit der Aufgabenstellung identisch, die Wertefolge $s(n)$ innerhalb eines pitch-Rahmens durch die Impulsantwort eines rekursiven Filters zu approximieren /23,40/. Daher sind die im 4. Abschnitt vorgelegten Ergebnisse auch für den Entwurf von Prädiktions-Vocodern von Bedeutung.

7.3 Der ADPCM-Vocoder

Bei der ADPCM-Codierung wird eine hochwertige, aber effiziente Übertragung des Sprachsignals angestrebt; dazu wird eine Redundanzreduktion vorgenommen (Beseitigung der linearen Abhängigkeiten). Bei dem Prädiktionsvocoder muss eine gewisse Informationsreduktion vorgenommen werden, weil die Aufteilung in Anregungsfunktion und Vokaltraktbeschreibung nur unvollkommen gelingt. Problematisch und aufwendig ist vor allem die pitch-Bestimmung.

Hier bietet sich eine Mischform zwischen den genannten Verfahren an: stimmhafte Laute werden ADPCM-codiert, stimmlose Laute und Pausen werden nach dem Vocoderprinzip übertragen. Die Unterscheidung zwischen diesen Zuständen ist sehr einfach durchzuführen:

- a) Ist der Prädiktionsgewinn gering, so liegt ein stimmloser Laut oder eine Pause vor. Das Differenzsignal wird nicht übertragen, es wird beim Empfänger durch ein unkorreliertes Signal gleicher mittlerer Energie ersetzt, so dass das Energiedichtespektrum des rekonstruierten Signals $r(\cdot)$ mit dem des Sprachsignals $s(\cdot)$ übereinstimmt.

b) Ist der Prädiktionsgewinn hoch, so wird eine normale ADPCM-Codierung vorgenommen.

Dieser ADPCM-Vocoder benötigt keine komplizierte Form der Mustererkennung; der als Umschaltkriterium dienende Prädiktionsgewinn wird bei der Bestimmung des Koeffizientenvektors h bereits mitberechnet. Subjektive Tests zeigten, dass der Unterschied zu einer reinen ADPCM-Codierung nicht feststellbar ist, wenn die Umschaltung bei einem Prädiktionsgewinn von etwa 6 dB vorgenommen wird. Dabei kann dann bei einem Drittel der Segmente auf die Übertragung des Differenzsignals verzichtet werden. Bei einer Schranke von 9 dB ergaben sich an zwei Stellen der Sprachprobe deutliche Verzerrungen; sie können sicher durch die Verwendung eines verbesserten Umschaltkriteriums vermieden werden.

Umschaltkriterium	:	$G = 6 \text{ dB}$	$G = 9 \text{ dB}$
Einsparung	:	32 %	54 %
Äquivalenter Zusatzgewinn	:	3.3 dB	6.8 dB

Der ADPCM-Vocoder ermöglicht also eine zusätzliche Einsparung von 1 bit/Abtastwert. Der wesentliche Nachteil des Verfahrens besteht darin, dass ein asynchroner Datenfluss auftritt.

8. Zusammenfassende Bewertung des ADPCM-Verfahrens

Eine nichtadaptive DPCM-Codierung von Sprache liefert einen mittleren Gewinn von 8.5 dB (TP-Sprache : 0 - 3400 Hz) bzw. von 7.5 dB (BP-Sprache : 300 - 3400 Hz). Mit einer ADPCM-Codierung werden nicht nur höhere Gewinne erzielt, sondern es entstehen einige zusätzliche Vorteile gegenüber dem DPCM-Verfahren :

- * die ADPCM-Codierung ist datenkompatibel ;
- * das Ruhegeräusch (Pausen) ist geringer ;
- * Konsonanten werden genauer codiert als bei einem DPCM-Verfahren .

Die gemessenen Gewinne beziehen sich auf eine 5-bit-DPCM- bzw. ADPCM-Quantisierung . Die oft in der Literatur zu findende Angabe des reinen Prädiktionsgewinns (ohne Quantisierer) ist, wie sich zeigt, zu optimistisch, da sich dieser Gewinn durch die Fehlerrückführung ver-

ringert. Bei einer ADPCM-Codierung mit einer 32 ms-Segmentierung und 10 Prädiktor-koeffizienten ergaben sich die folgenden, über drei Sprecher gemittelten Gewinne (vergl. Abschnitt 4.2.4):

TP - Sprache : Gewinn etwa 15 dB

BP - Sprache : Gewinn etwa 13 dB

Gegenüber einer DPCM-Codierung tritt also ein zusätzlicher Gewinn von etwa 6 dB auf.

Die Verwendung eines gesteuerten Quantisierers (Abschnitt 4.3) liefert zusätzliche Gewinne :

Gesteuerter Quantisierer:

ohne Entropiecodierung : Zusatzgewinn etwa 3.5 dB

mit Entropiecodierung : Zusatzgewinn etwa 7.5 dB

Mit der aufwendigen Entropiecodierung ist also fast ein Gesamtgewinn von 24 dB, also eine Einsparung von 4 bit/Abtastwert erreichbar. Für die ungünstigste Sprachprobe ist allerdings auf diese Weise nur eine Einsparung von 3 bit/Abtastwert erzielbar.

Ein weiterer zusätzlicher Gewinn ist durch die im Abschnitt 7.3 skizzierte Anwendung des ADPCM-Vocoders möglich: unabhängig von den bisher genannten Gewinnen kann damit zusätzlich etwa 1 bit/Abtastwert eingespart werden.

9. Literaturverzeichnis

- /1/ O Neal, J. B. ; Stroh, R.W.
Differential PCM for Speech and Data Signals
IEEE Trans. on Communications, Vol. COM-20, No. 5 (1972) S. 900-912.
- /2/ Mc Donald, R.A.
Signal-to-noise and idle channel performance of differential pulse code modulation systems-particular applications to voice signals
Bell System Technical Journal, Vol. 45(1966), S. 1123-1151.
- /3/ Noll, P.
Sprachcodierung mit Differenz-Pulsmodulation
NTG-Fachbericht "PCM-Technik", 1972, VDE-Verlag Berlin, S. 72-81.
- /4/ Noll, P. ; Block, R.
Simulation von Differenz-PCM-Systemen zur Sprachcodierung
Technischer Bericht Nr. 150 des Heinrich-Hertz-Instituts Berlin-Charlottenburg, 1972.
- /5/ Noll, P.
Digitale Wiener-Kolmogoroff-Filterung
In Vorbereitung.
- /6/ Paez, M.D. ; Glisson, T.H.
Minimum mean-squared-error quantization in speech PCM and DPCM systems
IEEE Trans. on Communications Vol. COM - 20(1972) S. 225-230.
- /7/ Noll, P. ; Zelinski, R.
Performance of quantizers for Laplacian-distributed signals
To be published.
- /8/ Noll, P.
Über die Dimensionierung von DPCM-Prädiktornetzwerken
1972 ; Unveröffentlicht.
- /9/ Atal, B.S. ; Schröder, M.R.
Predictive coding of speech signals
WESCON Technical Papers, 1968 ; Paper 8/2.

- /10/ Atal, B.S. ; Schröder, M.R.
Adaptive predictive coding of speech signals
Bell System Techn. Journal 1970, S. 1973 - 1986.
- /11/ Fujisaki, H. ; Ishida, H.
Simulation and evaluation of predictive coding systems for speech transmission
Annual Report of the Engineering Institute; Faculty of Engineering, University of Tokyo, Vol. 29 (1970), S. 195 - 200.
- /12/ Noll, P.
Sprachcodierung mit Differenz-Pulscodemodulation
Vortrag NTG-Tagung "PCM-Technik", München 1971.
- /13/ Dunn, I.G.
An experimental 9600 - bit/s Voice Digitizer employing adaptive Prediction
IEEE COM-19, 6 , Dec. 1971, S. 1021 ff.
- /14/ Noll, P.
Nonadaptive and adaptive differential pulse code modulation of speech signals
Polytechnisch tijdschrift, Den Haag 1972, Heft 19, S. 623 - 629.
- /15/ Stevens, K.N.
Autocorrelation of speech sounds
Journal Acoustic Soc. America Vol. 22 (1950) , S. 769 - 771.
- /16/ Bennet, W.R.
A machine for continuous display of short term correlation - "The Correlograph"
Bell System Techn. Journal, Vol 32 (1953), S. 1173 - 1185.
- /17/ Block, R. ; Noll, P.
Differenz-Pulscodemodulation
Heinrich-Hertz-Institut Berlin-Charlottenburg, 1971, 16 mm - Film .
- /18/ Rudin, H. jr.
Automatic equalization using transversal filters
IEEE spectrum, Vol. 4, S. 53 - 59, 1967.
- /19/ Lucky, R.W. ; Rudin, H.R.
An automatic equalizer for general purpose communication channels
Bell System Techn. Journal, Vol,46 (1967) S. 2179 - 2208.

- /20/ Rosenberger, J.R. ; Thomas, E.J.
Performance of an adaptive echo canceller operating in a noisy, linear, time-invariant environment
Bell System Techn. Journal Vol. 50 (1971), S. 785 - 813.
- /21/ Atal, B.S. ; Hanauer, S.L.
Speech analysis and synthesis by linear prediction of the speech wave
J. Acoust. Soc. America, Vol. 50 (1971) S. 637 - 655.
- /22/ Zue, V.W.
Speech analysis by linear prediction
MIT, Quarterly Progress Report, No. 105 (1972) S. 133 - 142.
- /23/ Portnoff, M.R. ; Zue, V.W. ; Oppenheim, A.V.
Some considerations in the use of linear prediction for speech analysis
MIT, Quarterly Progress Report, No. 106 (1972), S. 141 - 150.
- /24/ Markel, J.
Digital inverse filtering - a new tool for formant trajectory estimation
IEEE Trans. on Audio and Electroacoustics, Vol. AU - 20(1972),
S. 129 - 137.
- /25/ Itakura, F. ; Saito, S.
A statistical method for estimation of speech spectral density and formant frequencies
Electronics and Communication in Japan, Vol. 53, No. 1, 1970, S. 36 ff.
- /26/ Noll, P.
Redundanzreduzierende Sprachcodierung - eine vergleichende Betrachtung verschiedener Verfahren
NTG-Tagung "Signalverarbeitung", 1973, Erlangen.
- /27/ Noll, P.
Sprachcodierung mit segmentweise gesteuertem Quantisierer
In Vorbereitung.
- /28/ Zelinski, R. ; Noll, P.
Untersuchungen zur Quantisierung gedächtnisfreier Gaussquellen
Technischer Bericht Nr. 159 des Heinrich-Hertz-Instituts Berlin-Charlottenburg, 1972.

- /29/ Jenkins, G.M. ; Watts, D.G.
Spectral analysis and its applications
Holden Day, San Francisco, 1968.
- /30/ Lange, F. H.
Signale und Systeme, Bd. 3
VEB Verlag Technik, Berlin, 1971, Abschnitt 3.2 .
- /31/ Zelinski, R.
Analyse eines digitalen rekursiven Filters
Diplomarbeit TU Berlin, 1973.
- /32/ Veltman, B.P.Th. ; Kwakernaak, H.
Theorie und Technik der Polaritätskorrelation für die dynamische Analyse
niedrigerfrequenter Signale und Systeme
Regelungstechnik 9 (1961), S. 357 - 364.
- /33/ Winzer, L.-M.
Formant- und Pitchbestimmung für Sprachsignale
Diplomarbeit, TU Berlin, 1973.
- /34/ Itakura, F. ; Saito , S.
On the optimum quantization of feature parameters in the PARCOR speech
synthesizer
IEEE - AFCRL 1972 International Conference on Speech Communication
and Processing, Boston, Mass.
- /35/ Berger, E.R.
Mündlicher Vorschlag.
- /36/ Noll, P.
Zeitdiskrete Nachrichtenübertragung und Nachrichtenreduktion
Vorlesungsskriptum TU Berlin 1972/73.
- /37/ Balakrishnan, A.V.
An adaptive nonlinear data predictor
Proc. 1962 National Telemetry Conference, Paper 6 - 5 .
- /38/ Flanagan, J. L.
Speech analysis, synthesis, and perception
Springer-Verlag Berlin, 1972 .

/39/

Maksym, J.N.

Real time pitchperiod extraction by adaptive prediction of the speech waveform

IEEE-AFCRL International Conference on Speech Communication and Processing, Boston, 1972.

/40/

Weinstein, C.J. ; Oppenheim, A.V.

Predictive coding in a homomorphic Vocoder

IEEE Trans., Vol. AU - 19, 1971, S. 243-248.

Die Untersuchungen wurden von der Deutschen Forschungsgemeinschaft Bonn gefördert und unterstützt. Der Informatik-forschungsbereich der TU Berlin stellte die notwendige Rechenzeit am System IBM 360/57 zur Verfügung. Beiden Institutionen gilt mein Dank.

IEEE Trans., Vol. AU - 19, 1971, 2, 543-548.
Predictive coding in a nonhomomorphic vocoder

Weinstein, C. T. & Oppenheim, A. V.

Processing, Boston, 1975.

IEEE-AECSE International Conference on Speech Communication and
Waveform

Real time bitstream extraction by adaptive prediction of the speech

Makris, J. M.

D a n k

Mein besonderer D a n k gilt den Herren R. B l o c k , E. G r o ß m a n n ,
B. H e u s e r , R. Z e l i n s k i
für ihre Hilfe bei der Durchführung der Simulationen und für den Aufbau des Pro -
grammsystems zur Sprachverarbeitung.

Die Untersuchungen wurden von der Deutschen Forschungsgemeinschaft Bad Godesberg
unterstützt. Der Informatik-Forschungsbereich der TU Berlin stellte die notwendige
Rechenzeit am System IBM 360/67 zur Verfügung. Beiden Institutionen gilt mein
Dank.

