(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Veröffentlichungsnummer: **0 449 043 A2**

(12) EUROPÄISCHE PATENTANMELDUNG

(21) Anmeldenummer: **91103907.1**

(51) Int. Cl.5: **G10L 9/14**

(22) Anmeldetag: **14.03.91**

(30) Priorität: **22.03.90 CH 956/90**

(43) Veröffentlichungstag der Anmeldung:
02.10.91 Patentblatt 91/40

(84) Benannte Vertragsstaaten:
AT BE DE DK ES FR GB IT NL SE

(71) Anmelder: **ASCOM ZELCOM AG**
Eichtalstrasse
CH-8634 Hombrechtikon(CH)

(72) Erfinder: **Schaub, Arthur**
Sunnenbergstrasse 20
CH-8633 Wolfhausen(CH)

(74) Vertreter: **Dittrich, Horst**
Zellweger Uster AG Patentabteilung
Wilstrasse 11
CH-8610 Uster(CH)

(54) Verfahren und Vorrichtung zur Sprachdigitalisierung.

(57) Die Sprachdigitalisierung erfolgt unter Anwendung sowohl der Signalform- wie auch der Quellencodierung, mit einem Codierer zur Digitalisierung und einem Decodierer zur Rekonstruktion des Sprachsignals. Im Codierer wird das Sprachsignal in Segmente unterteilt und in einem Teil der Segmente unter möglichst genauer Annäherung der Abtastwerte verarbeitet, wobei anhand von bekannten Abtastwerten eine Berechnung eines Schätzwertes für bevorstehende Abtastwerte erfolgt. Im anderen Teil der Segmente werden nur Parameter für eine Sprachnachbildung im Sinne der Quellencodierung abgeleitet. Die einzelnen Signalsegmente werden mit variabler Bitrate verarbeitet, wobei diese Bitraten verschiedenen Betriebsarten zugeordnet sind, und jedes Signalsegment wird in eine der Betriebsarten klassiert. Dadurch werden die einzelnen Sprachsegmente je nach Erfordernis mit mehr oder weniger Bit codiert, und man erhält ein hybrides Codierverfahren, welches die Quellencodierung und die Signalformcodierung vereinigt. Dies führt zusammen mit der Signalquantisierung vor- und nachgelagerten Signalverarbeitungsschritten zu einer durchschnittlichen Bitrate von 6 kBrt/s und einer Sprachqualität, die 100% derjenigen bei Telefonieübertragung entspricht.

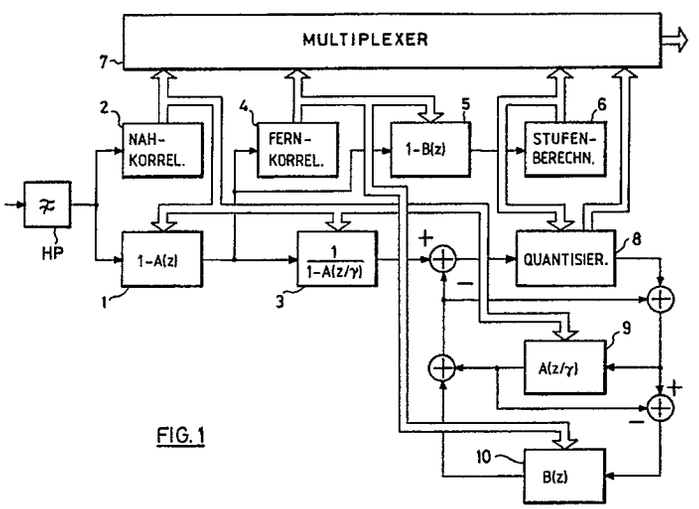

FIG. 1

EP 0 449 043 A2

Wegen der Beschränktheit der von den üblichen Sprachkanälen zugelassenen Datenübertragungsgeschwindigkeiten sind seit langem Bestrebungen zur Reduzierung der Bitrate durch entsprechende Sprachcodierung im Gange. Wenn man die Sprachqualität, das ist die Summe aus Verständlichkeit, Sprechererkennung und natürlichem Klang, in Relation zur Bitrate setzt, dann entspricht die 100 prozentige Qualität (= Telefoniequalität) der bekannten logarithmischen Pulscodemodulation mit einer Bitrate von 64 Kilobit pro Sekunde, welche am oberen Ende des für Funk und Telefonie bedeutenden Bereichs von 2,4 bis 64 kBit pro Sekunde liegt.

Die logarithmische Pulscodemodulation gehört zur Klasse der sogenannten Signalform- oder Kurvenformcodierer, deren Prinzip darin besteht, jeden einzelnen Abtastwert möglichst genau anzunähern. Die Codierung der Abtastwerte kann dabei auf unterschiedliche Arten erfolgen, nämlich so, dass die Codierung vom vorhergehenden Abtastwert abhängt, oder von Parametern, die von den vorhergehenden Abtastwerten abgeleitet wurden, so dass man Vorteile aus einer etwaigen Charakteristik der Sprachsignale ziehen kann und die Möglichkeit besteht, auf diese Weise die Wirksamkeit des Verarbeitungsverfahrens zu verbessern und die Bitgeschwindigkeit zu erniedrigen. Wenn man die Korrelationsfunktion eines Sprachsignalabschnitts kennt, kann man ein optimales Filter berechnen, das die besten Schätzwerte für die Vorhersage eines Abtastwertes aus vorhergehenden Abtastwerten liefert. Dieses Filter wird in einer Rückführschleife eingesetzt, um ein Quantisierungsgeräusch mit flachem Spektrum, das heisst ohne Sprachmodulation, zu erhalten.

Im Unterschied zur Signalformcodierung steht die sogenannte Quellencodierung, die im Englischen in Verbindung mit der Sprachcodierung als Vocoding bezeichnet wird. Hier geht es nur darum, bei der Wiedergabe ein Signal zu erzeugen, das möglichst ähnlich klingt wie das Original, bei dem aber der Signalverlauf selbst, also die einzelnen Abtastwerte, vom Original sehr verschieden sein kann. Es wird unter Benutzung einer Nachbildung der Spracherzeugung das Signal analysiert, um Parameter für eine Sprachnachbildung abzuleiten. Diese Parameter werden digital zur Empfangsseite übertragen, wo sie zur Steuerung einer Syntheseeinrichtung dienen, die der verwendeten Nachbildung der Analyse entspricht.

Die Quellencodierung erzeugt bei 2,4 Kilobit pro Sekunde bereits 60 bis 75% der vollen Sprachqualität, kann aber diese auch bei beliebiger Erhöhung der Bitrate nicht über den Sättigungswert von 75% steigern. Diese reduzierte Qualität macht sich hauptsächlich in einem nicht ganz natürlichen Klang und in erschwerter Sprechererkennung bemerkbar. Der Grund dafür liegt im zu einfachen Modell zur Sprachsynthese.

Bei der Signalformcodierung kann bei Aufrechterhaltung der vollen Sprachqualität die Bitrate von 64 Kilobit bis auf etwa 12 Kilobit pro Sekunde verkleinert werden, wobei allerdings die Komplexität der Codier-Algorithmen entsprechend zunimmt. Unterhalb von 12 Kilobit pro Sekunde nimmt die Sprachqualität der Signalformcodierung rasch ab.

Die vorliegende Erfindung betrifft nun ein Verfahren zur Sprachdigitalisierung unter Anwendung der Signalformcodierung, mit einem Codierer zur Digitalisierung und einem Decodierer zur Rekonstruktion des Sprachsignals, bei welchem im Codierer das Sprachsignal in Segmente unterteilt und unter möglichst genauer Annäherung der Abtastwerte verarbeitet wird, wobei anhand von bekannten Abtastwerten eine Berechnung eines Schätzwertes für bevorstehende, neue Abtastwerte erfolgt.

Durch die Erfindung soll die Lücke zwischen Signalform- und Quellencodierung im Bereich von etwa 3,6 bis 12 Kilobit pro Sekunde geschlossen, oder mit anderen Worten, es soll ein Codierungsverfahren angegeben werden, bei dessen Anwendung die Sprachqualität ab etwa 6 Kilobit/s 100% beträgt, wobei zu deren Erreichung der für Signalformcodierung übliche massvolle Rechenaufwand genügt.

Diese Aufgabe wird erfindungsgemäss dadurch gelöst, dass die Berechnung des Schätzwertes nur in einem Teil der Segmente erfolgt und im anderen Teil der Segmente nur Parameter für eine Sprachnachbildung im Sinn der Quellencodierung abgeleitet werden, und dass die einzelnen Signalsegmente mit variabler Bitrate verarbeitet werden, wobei diese Bitraten verschiedenen Betriebsarten zugeordnet sind und jedes Signalsegment in eine der Betriebsarten klassiert wird.

Dadurch werden die einzelnen Sprachsegmente je nach Erfordernis mit mehr oder weniger Bit codiert, und man erhält ein hybrides Codierverfahren, bei welchem die Methoden der Quellencodierung und der Signalformcodierung vereinigt sind. Die segmentweise Verarbeitung mit unterschiedlicher Bitrate führt zusammen mit den der Signalquantisierung vor- und nachgelagerten Signalverarbeitungsschritten zu einer durchschnittlichen Bitrate von etwa 6 Kilobit pro Sekunde und zu einer Sprachqualität, die 100% derjenigen bei der Telefonieübertragung beträgt. Die entsprechende Abtastrate beträgt 7200 Hz, die Bandbreite 3400 Hz. Die Länge der Sprachsegemente beträgt 20 Millisekunden, so dass ein Segment 144 Abtastwerte umfasst. Die Erfindung betrifft weiter eine Vorrichtung zur Durchführung des genannten Verfahrens mit einem Codierer und einem Decodierer.

Die erfindungsgemässe Vorrichtung ist dadurch gekennzeichnet, dass im Codierer ein adaptives Nah-Prädiktionsfilter zur Berechnung des Schätzwertes für den unmittelbar bevorstehenden, neuen Abtastwert in

dem einen Teil der Segmente, ein adaptives Fern-Prädiktionsfilter für den Einsatz in stimmhaften Signalsegmenten und Mittel zur Untersuchung der Signalsegemente und zu deren Zuordnung zu den einzelnen Betriebsarten vorgesehen sind.

Der Aufbau des erfindungsgemässen Sprachcodierers mit variabler Bitrate basiert somit einerseits auf dem Prinzip der adaptiv-prädiktiven Codierung (APC) und andererseits auf jenem der linearen prädiktiven Codierung des klassischen LPC-Vocoders mit einer Bitrate von 2,4 Kilobit pro Sekunde.

Im folgenden wird die Erfindung anhand eines in den Figuren dargestellten Ausführungsbeispiels näher erläutert; es zeigen:

Fig. 1 ein Blockschaltbild eines Codierers,
Fig. 2 ein Blockschaltbild eines Decodierers,
Fig. 3 das Flussdiagramm des Codierers,
Fig. 4 die Struktur eines ersten Filters,
Fig. 5, 6 Darstellung von Datenformaten; und
Fig. 7-9 Strukturen weiterer Filter.

Die typische Datenraten der Quellencodierung ermöglichen für viele Signalsegmente eine qualitativ ausreichende Wiedergabe. Das gilt zunächst einmal für die deutlich wahrnehmbaren Sprechpausen zwischen Wörtern und Sätzen, aber auch für die kurzen Sprechpausen vor Plosivlauten (p, t, k, b, d und g). Letzteres sind Pausen innerhalb einzelner Wörter, beispielsweise beim Wort "Vater" zwischen a und t. Solche Signalintervalle werden nachfolgend als leise Segmente bezeichnet und einer ersten Betriebsart, Modus I, zugeordnet. Sie werden mit 24 Bit codiert, was eine Datenrate von 1200 Bit/s ergibt.

Auch die Zischlaute (s, f und sch), sowie Atemgeräusche zwischen dem Sprechen, können mit einer geringen Datenrate von vorzugsweise 2400 Bit/s ausreichend wiedergegeben werden. Diese Laute haben die gemeinsame Eigenschaft, dass von der Lunge ein kontinuierlicher Luftstrom durch die Luftröhre, Rachen- und Mundhöhle strömt, und dass an einer bestimmten Stelle durch eine Verengung eine Luftturbulenz entsteht, wobei sich die verschiedenen Zischlaute durch den Ort dieser Verengung unterscheiden: Beim s ist es die Verengung zwischen oberer und unterer Zahnreihe, beim f diejenige zwischen oberer Zahnreihe und Unterlippe und beim sch diejenige zwischen Zungenspitze und Gaumen. In jedem Fall handelt es sich um ein Rauschen, das entsprechend der geometrischen Anordnung der Sprechorgane eine etwas verschiedene spektrale Färbung erfährt. Die entsprechenden Signalintervalle werden nachfolgend als frikative Segmente bezeichnet und einer zweiten Betriebsart, Modus II, zugeordnet. Sie werden mit 48 Bit codiert, was die schon erwähnte Datenrate von 2400 Bit/s ergibt.

Eine weitere Art von Signalintervallen, die nachfolgend als normal bezeichneten Segmente, weist keine Signaleigenschaften auf, die eine besonders sparsame Codierung zulassen würden, so wie die leisen und die frikativen Segmente. Die normalen Segmente zeigen aber auch keine Besonderheit, die einen zusätzlichen Codieraufwand erfordert, wie die stimmhaften Segmente, die als letzte Betriebsart gleich anschliessend erklärt werden. Die normalen Segmente werden einer dritten Betriebsart, Modus III, zugeordnet und mit 192 Bit codiert, was eine Datenrate von 9600 Bit/s ergibt.

Die stimmhaften Laute schliesslich umfassen alle Vokale (a, e, i, o, u, ä, ö, ü und y) und Diphtonge (au, ei und eu) sowie die Nasallaute (m, n, und ng). Ihre gemeinsame Eigenschaft besteht in der Aktivität der Stimmbänder, welche den Luftstrom aus den Lungen durch Abgabe periodischer Luftstösse modulieren. Damit ergibt sich eine quasi-periodische Signalform. Die verschiedenen stimmhaften Laute zeichnen sich durch unterschiedliche geometrische Anordnungen der Sprechorgane aus, was zu unterschiedlichen spektralen Färbungen führt. Eine qualitativ ausreichende Wiedergabe der stimmhaften Laute ist nur möglich, wenn zusätzlich zum Codierverfahren für die normalen Segmente die annähernde Periodizität mitberücksichtigt wird. Dadurch ergibt sich für die einer vierten Betriebsart, Modus IV, zugeordneten stimmhaften Laute eine auf 216 Bit pro Segment erhöhte Datenmenge und daraus eine Datenrate von 10800 Bit/s.

Die verschiedenen Betriebsarten und ihre Datenmengen sind in der nachfolgenden Tabelle 1 zusammengefasst:

Betriebsart	Bezeichnung	Datenmenge	Datenrate
Modus I	leise	24 Bit/Segm.	1200 Bit/s
Modus II	frikative	48 Bit/Segm.	2400 Bit/s
Modus III	normal	192 Bit/Segm.	9600 Bit/s
Modus IV	stimmhaft	216 Bit/Segm.	10800 Bit/s

Tabelle 1

Voraussetzung für die Anwendung der verschiedenen Betriebsarten mit der jeweiligen Datenrate ist eine Signalanalyse, welche jedes Signalsegment in eine der Betriebsarten Modus I bis Modus IV klassiert und die passende Signalverarbeitung einleitet.

Nachfolgend sollen nun anhand der Figuren 1 und 2 der Codierer und der Decodierer erläutert werden. Grundsätzlich basiert der Aufbau des Codierers mit variabler Bitrate einerseits auf dem Prinzip der adaptiv-prädiktiven Codierung (APC) und andererseits auf jenem des klassischen 2,4 kBit/s LPC-Vocoders. Eine ausführliche Beschreibung der adaptiv-prädiktiven Codierung findet sich im Buch "Digital Coding of Waveforms" von N.S. Jayant und P. Noll, Prentice Hall, Inc., Englewodd Cliffs, New Jersey 1984; Kapitel 6: Differential PCM, S. 252-350; Kapitel 7: Noice Feedback Coding, S. 351-371. Die dem LPC-Vocoder zugrunde liegenden Ideen sind in "Digital Processing of Speech Signals" von L.R. Rabiner und R.W. Schafer, Prentice Hall, Inc., Englewodd Cliffs, New Jersey, 1978; Kapitel 8: Linear Predictive Coding of Speech, S. 396-461, beschrieben.

Gemäss Fig. 1 enthält der Codierer an seinem Eingang ein Hochpassfilter HP und an dessen Ausgang ein adaptives Filter 1 mit der Transferfunktion 1-A(z) sowie eine mit Nah-Korrelation bezeichnete Stufe 2. Vom Ausgang des Filters 1 führt der Signalpfad zu einem adaptiven Vorfilter 3 mit der Transferfunktion $1/(1-A(z/\gamma))$, zu einer mit Fernkorrelation bezeichneten Stufe 4 und zu einem adaptiven Filter 5 mit der Transferfunktion 1-B(z), an welches eine mit Stufen-Berechnung bezeichnete Stufe 6 anschliesst. Ausserdem enthält die Schaltung einen Multiplexer 7, vier Summationspunkte, einen adaptiven Quantisierer 8, ein Filter 9 mit der Transferfunktion $A(z/\gamma)$ und ein Filter 10 mit der Transferfunktion B(z).

Der Decodierer enthält gemäss Fig. 2 einen Demultiplexer 11, einen Decodierer/Quantisierer 12, eine Rauschquelle 13, drei Summationspunkte, ein Filter 9 mit der Transferfunktion $A(z/\gamma)$ ein Filter 10 mit der Transferfunktion B(z) und ein adaptives Nachfilter 14 mit der Transferfunktion $(1-A(z/\alpha))/(1-A(z))$.

In der folgenden Tabelle 2 ist angegeben, welche algorithmischen Grundelemente der Codierer enthält, und von welchen der in den Fig. 1 und 2 dargestellten Schaltungselemente diese Funktionen wahrgenommen werden:

Algorithmisches Grundelement	Schaltungselement
Adaptives Nah-Prädiktionsfilter (LPC-Prädiktor)	Filter 9 $A(z/\gamma)$
Adaptives Fern-Prädiktionsfilter (Pitch-Prädiktor)	Filter 10 $B(z)$
Adaptiver Quantisierer	Quantisierer 8, 12
Hochpassfilter zur teilweisen Kompensation der spektralen Schiefe nicht-frikativer Sprachsegmente	Filter HP
Adaptive Vorfilter zur spektralen Formung des Quantisierungsrauschens	Filter 1 und 3 $1-A(z)$, $1/(1-A(z/\gamma))$
Adaptives Nachfilter zur spektralen Formung des Sprachsignals	Filter 14 $(1-A(z/\alpha))/(1-A(z))$
Rauschquelle zur Erzeugung eines synthetischen Anregungssignals	Rauschquelle 13

Tabelle 2

Das auch als Prädiktor zur linearen prädiktiven Codierung (LPC-Prädiktor) bezeichnete Nah-Prädiktionsfilter berechnet, ausgehend von einigen wenigen bereits bekannten Abtastwerten, einen Schätzwert für den unmittelbar bevorstehenden, neuen Abtastwert. Die Transferfunktion des Nah-Prädiktionsfilters wird üblicherweise mit A(z) bezeichnet. Das Filter arbeitet segmentweise mit einer anderen, dem Signalverlauf angepassten Transferfunktion; denn da sich die Signalform eines Sprachsignals fortwährend ändert, sind für jedes Signalelement neue Filterkoeffizienten zu berechnen. Diese Berechnung erfolgt in der mit 2 bezeichneten Stufe Nah-Korrelation.

Werden die vom Prädiktionsfilter berechneten Schätzwerte von den einzelnen Signalwerten subtrahiert, so ergibt sich ein Restsignal, das aus den linear nicht vorhersagbaren Signalanteilen besteht. Die Transferfunktion dieser Filterung ist 1-A(z). Das Restsignal hat aufgrund der Unvorhersagbarkeit Eigenschaften eines Zufallsprozesses, die sich in seinem näherungsweise flachen Spektrum zeigen. Somit hat das adaptive Filter 1 die bemerkenswerte Eigenschaft, die spezifischen Resonanzen, das sind die sogenannten Formanten, eines Lautes flachzuglätten.

Die Filterung 1-A(z) mit dem am Eingang des Codierers angeordneten Filter 1 findet in jeder der vier Betriebsarten (Tabelle 1) statt. Für die verschiedenen Betriebsarten gelangen unterschiedliche Filterordnungen zum Einsatz; für die leisen Segmente (Modus I) hat das Filter die Ordnung drei und für die anderen Betriebsarten die Ordnung acht. Die Prädiktionskoeffizienten werden im weiteren Verlauf der Codierung auch noch für die Filter 3 und 9 benötigt, was in Fig. 1 durch die breiten, den Datenfluss kennzeichnenden Pfeile symbolisiert ist. Ebenso werden die Prädiktionskoeffizienten beim Decodieren von Fig. 2 für die Filter 9 und 14 gebraucht. Da aber das Prädiktionsfilter nur im Codierer bei vorliegendem Signalsegment berechnet werden kann, müssen die berechneten Koeffizienten codiert und zusammen mit weiteren digitalen Informationen abgespeichert werden, damit der Decodierer das Signal rekonstruieren kann. Diese Codierung der Koeffizienten ist in Fig. 1 als ein Bestandteil der Nah-Korrelationsstufe 2 gedacht. Ihre Abspeicherung ist durch den Datenpfeil zum Multiplexer 7 symbolisiert. Vom Demultiplexer 10 in Fig. 2 gelangen die Prädiktionskoeffizienten dann entlang den eingezeichneten Datenpfeilen zu den Filtern 9 und 14.

Das adaptive Fern-Prädiktionsfilter wird entsprechend der englischen Bezeichnung für die Grundfrequenz des in stimmhaften Lauten vorhandenen, periodischen Anregungssignals auch als Pitch-Prädiktor bezeichnet. Sein Einsatz ist nur in stimmhaften Segmenten sinnvoll (Modus IV), und der eigentlichen

Filterung geht in jedem Fall eine Signalanalyse voraus, die für oder gegen ihren Einsatz entscheidet. Diese Analyse findet in der Fern-Korrelationsstufe 4 statt. Weitere Aufgaben dieser Stufe sind die Berechnung und Codierung der Koeffizienten des Fern-Prädiktionsfilters, welche so wie jene des Nah-Prädiktionsfilters als Teil der digitalen Information gespeichert werden müssen, damit der Decodierer den Signalverlauf in stimmhaften Segmenten rekonstruieren kann.

Die Transferfunktion des Fern-Prädiktionsfilters ist mit B(z) bezeichnet. Es ist als Transversalfilter implementiert; seine Filterordnung beträgt drei. Im Unterschied zum Nah-Prädiktionsfilter arbeitet es nicht auf den unmittelbar vorangegangenen Signalwerten, sondern auf solchen im Abstand einer Grundperiode M des periodischen Anregungssignals. Die Bestimmung dieser auch als Pitch-Periode bezeichneten Grösse M ist eine weitere Aufgabe der Fern-Korrelationsstufe 4.

Der adaptive Quantisierer (Tabelle 2) setzt sich aus der Stufenberechnung 6 und dem Quantisierer 8 zusammen. Seine Arbeitsweise ähnelt derjenigen eines gewöhnlichen Analog/Digital-Wandlers, mit dem Unterschied, dass der adaptive Quantisierer nicht mit einer konstanten maximalen Signalamplitude arbeitet, sondern einen variablen Wert benützt, der in der Stufenberechnung 6 periodisch neu ermittelt wird.

Die Stufenberechnung, die in allen Betriebsarten erfolgt, unterteilt jedes Signalsegment in Teilsegmente und berechnet für jedes Teilsegment einen neuen, dem Signalverlauf angepassten Stufenwert. Leise Segmente werden in zwei, die übrigen in drei Teilsegmente unterteilt. Die Stufenwerte werden ebenfalls codiert und abgespeichert.

Die Quantisierung und Codierung der einzelnen Signalwerte findet im Quantisierer 8 statt und erfolgt mit nur einem einzigen Bit pro Signalwert, wobei ein positiver Signalwert mit 1 und ein negativer Signalwert mit 0 codiert wird. Damit haben diese Daten die Bedeutung von Vorzeichenbits. Die Signalwerte am Ausgang des Quantisierers 8 sind der positive aktuelle Stufenwert für den Code 1 und der negative aktuelle Stufenwert für das Codewort 0. Die Quantisierung der einzelnen Signalwerte findet nur in den normalen und stimmhaften Segmenten statt. Dieser Sachverhalt führt zu den auffallend geringen Datenraten der leisen und frikativen Signalelemente.

Der Decodierer/Quantisierer 12 erhält die Vorzeichenbits zur Rekonstruktion der einzelnen Signalwerte nur in den normalen und stimmhaften Segmenten. In den leisen und frikativen Segmenten ist die Rauschquelle 13 aktiv, welche ein pseudo-zufälliges Signal konstanter Leistung liefert, dessen Werte mit dem aktuellen Stufenwert multipliziert werden. Dieses lokal erzeugte Signal ermöglicht eine qualitativ ausreichende Wiedergabe der leisen und fikativen Segmente.

Die Signalpfade in Fig. 1 mit dem Quantisierer 8, den Prädiktoren 9 und 10, sowie den vier Summationspunkten werden zusammen als ΔPCM-Schleife bezeichnet. Im gewöhnlichen APC-Schema gelangt das eintreffende Sprachsignal direkt zur ΔPCM-Schleife, also ohne die Filter 1 und 3 zu durchlaufen, und es gelangt in der ΔPCM-Schleife anstelle des Filters 9 mit der Transferfunktion $A(z/\gamma)$ das Nah-Prädiktionsfilter A(z) zum Einsatz.

Gemäss Fig. 1 wird vom Signalwert am Ausgang des Hochpassfilters HP ein Prädiktionswert subtrahiert, der sich in stimmhaften Segementen aus dem Nah- und dem Fernprädiktionswert zusammensetzt. In nicht stimmhaften Segmenten liefert das Fern-Prädiktionsfilter keinen Beitrag. Der Differenzwert wird in beiden Fällen quantisiert, und am Ausgang des Quantisierers 8 wird der Prädiktionswert zum quantisierten Differenzwert addiert. Diese Addition ergibt einen quantisierter Sprachsignalwert, der den in die ΔPCM-Schleife eingespeisten und nicht quantisierten Sprachsignalwert approximiert. Im Decodierer von Fig. 2 wird mit Hilfe der abgespeicherten digitalen Informationen genau dieser Näherungswert rekonstruiert. Beim gewöhnlichen APC- Schema gelangt nun das quantisierte Sprachsignal direkt zum Lautsprecher, ohne das Filter 14 zu durchlaufen.

Das Besondere am APC-Schema liegt darin, dass den Prädiktoren das quantisierte Sprachsignal als Eingangssignal dient, und dass die Prädiktoren in einer Rückführschleife angeordnet sind. Aus Fig. 1 ist auch ersichtlich, dass die beiden Prädiktoren in Serie arbeiten, so dass das Ausgangssignal des Nah-Prädiktionsfilters vom quantisierten Sprachsignal subtrahiert wird und diese Differenz in das Fern-Prädiktionsfilter gelangt.

Bei genügend feiner Quantisierung mit mehreren Bit pro Signalwert unterscheidet sich der quantisierte Differenzwert vom nicht quantisierten durch einen geringfügigen Rundungsfehler. Das Signal der aufeinanderfolgenden Rundungsfehler ist in diesem Fall unkorreliert und zeigt ein flaches Spektrum. Dieses sogenannte Quantisierungsrauschen ist im quantisierten Sprachsignal additiv enthalten. Sein Spektrum setzt sich also aus dem Spektrum des ursprünglichen, nicht quantisierten Sprachsignals und dem flachen Spektrum des Quantisierungsrauschens zusammen. Bei feiner Quantisierung ist der Signal/Rausch-Abstand so gross, dass man das Quantisierungsrauschen gar nicht oder nur leise wahrnimmt.

Bei grober Quantisierung mit einem oder zwei Bit pro Signalwert ist dagegen der Rauschabstand so klein, dass das Quantisierungsrauschen als störend wahrgenommen wird. Im Frequenzbereich zeigt sich,

dass das Quantisierungsrauschen Teile des Sprachsignalspektrums überdeckt, wobei es sich um Frequenzintervalle zwischen den Formanten handelt. Die Formanten selbst ragen wie Bergspitzen aus dem Quantisierungsrauschen heraus.

Um das Quantisierungsrauschen zwischen den Formanten unter das Niveau des Sprachsignalspektrums absinken zu lassen, wird das Sprachsignal vor der ΔPCM-Schleife so bearbeitet, dass die Formanten weniger ausgeprägt vorhanden sind. Vor der Wiedergabe muss dann das quantisierte Signal einer inversen Formung unterzogen werden, damit es wieder den ursprünglichen Klang annimmt. Im Vergleich mit dem gewöhnlichen APC-Schema erhöht sich dann das Quantisierungsrauschen in den mit Formanten besetzten Frequenzintervallen; es findet also eine Umlagerung des Quantisierungsrauschens innerhalb einzelner Frequenzintervalle statt. Daher wird die beschriebene Formung als spektrale Formung des Quantisierungsrauschens (Tabelle 2) bezeichnet.

Aufgrund von physiologischen Gegebenheiten des menschlichen Wahrnehmungsapparates darf der Rauschabstand in den Formanten im Vergleich zu den Verhältnissen bei APC wohl etwas verkleinert werden, aber nur massvoll. Der ideale Kompromiss ist dann gegeben, wenn das Quantisierungsrauschen zwischen den Formanten knapp unter das Niveau des Sprachsignals gelangt und in den Formanten immer noch deutlich unter dem Signalspektrum bleibt. In diesem Fall wird das quantisiete Sprachsignal als praktisch störungsfrei wahrgenommen (sogenannter Maskierungseffekt).

Ein wesentlicher Teil der Erfindung besteht nun darin, dass es gelungen ist, diesen Maskierungseffekt bei der bereits beschriebenen spärlichen Quantisierung überhaupt zu erreichen. Das gelingt durch den kombinierten Einsatz

- einer festen Hochpass-Vorfilterung (f_{3db} = 700..1000 Hz) zusammen mit
- der adaptiven spektralen Formung des Quantisierungsrauschens, realisiert durch die adaptive Vorfilterung (1-A(z))/1-A(z/γ)) und die adaptive Nachfilterung (1-A(z/α))/(1-A(z)), die ihrerseits
- infolge α kleiner γ durch eine Formanten-Ueberbetonung die Maskierung verstärkt und zugleich die durch das feste Hochpassfilter veränderte Klangfarbe der Vokale weitgehend kompensiert, wogegen andererseits
- einer zu starken Ueberbetonung der frikativen Segmente durch diese vorgängigen Verarbeitungsschritte mittels einer reduzierten Signalleistung der synthetischen Rauschquelle entgegengewirkt wird.

Bei der spektralen Formung des Quantisierungsrauschens geht es also darum, die Formanten des Sprachsignals vor der Einspeisung in die ΔPCM-Schleife massvoll zurückzubilden und im Anschluss an die Decodierung wieder im gleichen Mass zu verstärken. Im Codierer geschieht dies durch die aufeinanderfolgenden Filter 1 und 3, in der ΔPCM-Schleife gelangt das Prädiktionsfilter 9 zum Einsatz, da seine Transferfuntkion auf das spektral geformte Signal abgestimmt ist. Es wurde schon erwähnt, dass das Filter 1 die in einem Signalsegement vorhandenen Formanten glättet; das inverse Filter mit der Transferfunktion 1/(1-A(z)) ist folglich in der Lage, einem flachen Spektrum die entsprechenden Formanten wieder einzuprägen, wobei ein einzelner Filterparameter , welcher zwischen null und eins liegt, genügt, um die Formanten in kontrollierter Weise schwächer auszubilden.

Das Filter 14 zur inversen spektralen Formung im Decodierer müsste eigentlich die Transferfunktion (1-A(z/γ)/(1-A(z)) aufweisen, besitzt aber anstelle von γ den Filterparameter α, welcher zwischen null und γ liegt, wodurch die Frequenzintervalle mit besserem Rauschabstand gegenüber jenen mit schlechterem Abstand etwas verstärkt werden. Das Filter 1-A(z/α) glättet das quantisierte Signal nicht vollständig flach, und das nachfolgende Filter 1/(A(z)) prägt einem Signal mit flachem Spektrum die Formanten in vollem Ausmass ein. Da beim Eingangssignal des letzteren Filters die Formanten ansatzweise vorhanden sind, werden sie durch die Filterung im Vergleich mit dem nicht quantisierten Sprachsignal wie gewünscht überbetont. Mit g ist eine adaptive Lautstärkensteuerung bezeichnet (siehe auch Fig. 9), die sich aus den k-Werten des Filters berechnet und die zum Ausgleich von Lautstärkeschwankungen dient, die durch die unterschiedlichen Filterkoeffizienten α und γ verursacht werden.

Die Filter 1, 3 zur spektralen Formung im Codierer und 14 im Decodierer sind in allen Betriebsarten aktiv, wobei diese für die subjektiv empfundene Sprachqualität wesentlichen Massnahmen keine zusätzlichen Daten zur Abspeicherung verursachen. Die einmal gewählten Werte für die Filterparameter γ und α bleiben bei der Anwendung konstant.

Die einzelnen Signalverarbeitungsschritte im Codierer sind aus dem Flussdiagramm von Figur 3 ersichtlich. Dieses zeigt die Anordnung von Verarbeitungssequenzen und Entscheidungen, wobei die Entscheidungen eine allmähliche Aufgliederung in die separate Verarbeitung der vier verschiedenen Betriebsarten (Tabelle 1) bewirken.

Darstellungsgemäss beginnt die Verarbeitung mit der Berechnung der Autokorrelationskoeffizienten; die nachfolgende Entscheidung trennt die Verarbeitung der leisen von jener der übrigen Segmente.

Der Autokorrelationskoeffizient r(0) dient als Mass für die in einem Segment enthaltene Energie, wobei

der Entscheid, ob es sich um ein leises Segment handelt, im Vergleich mit einer adaptiv nachgeführten Schwelle Θ erfolgt. Wenn ein Bruchteil des Autokorrelationskoeffizienten die Schwelle übertrifft, dann wird die Schwelle auf den Wert dieses Bruchteils angehoben. Der Entscheid für ein leises Segement fällt, wenn die Signalleistung kleiner als die momentane Schwelle wird.

Die Verarbeitung der leisen Segmente umfasst die Berechnung und Codierung der Koeffizienten des Nah-Prädiktionsfilters, die Filterung 1-A(z) durch das Filter 1 (Fig. 1) und die Berechnung und Codierung der Quantisierungsstufen.

Das in Fig. 4 dargestellte Filter 1 ist als sogenanntes Gitterfilter (englisch: Lattice-Filter) implementiert, dessen Koeffizienten die sogenannten Reflektionskoeffizienten k_1,.... k_m sind. Struktur und Eigenschaften der Lattice-Filter sind im Buch "Adaptive Filters" von C.F.N. Cowan und P.M. Grant, Prentice Hall, Inc., Englewodd Cliffs, New Jersey, 1985, Kapitel 5: Recursive Least-Squares Estimation and Lattice Filters, S. 91-144, beschrieben. Da in den leisen Segmenten die Filterordnung drei beträgt, werden nur drei Reflektionskoeffizienten berechnet und die übrigen null gesezt.

Die Berechnung erfolgt ausgehend von den bereits ermittelten Autokorrelationskoeffizienten, wobei irgendeines der bekannten Verfahren (Durbin-Levinson, Schur, Le Roux - Gueguen) angewendet werden kann. Von praktischer Bedeutung ist dabei, dass eine Ueberwachung der Filterstabilität miteingeschlossen ist: Wenn die Berechnung für einen Reflektionskoeffizienten dem Betrag nach einen Wert grösser eins liefert, dann wird dieser und alle Koeffizienten höherer Ordnung null gesetzt.

Nach der Berechnung der Reflektionskoeffizienten folgen verschiedene Schritte zu ihrer Quantisierung und Codierung. In einem ersten Schritt werden die berechneten Werte auf in der Praxis relevante Wertebereiche reduziert, welche Intervalle darstellen, in die 99% aller Werte einer umfangreichen Sprachprobe entfielen. Wenn ein berechneter Koeffizient den minimalen oder maximalen Wert überschreitet, dann wird an seiner Stelle der tabellierte Extremwert weiterverarbeitet. Diese Begrenzung ist im Flussdiagramm von Fig. 3 nicht ausgewiesen, sie bewirkt aber einen effizienteren Einsatz der zur Codierung der Koeffizienten zur Verfügung stehenden Bits.

Die weiteren Schritte umfassen die Berechnung der sogenannten Log Area Ratio und die lineare Quantisierung/Codierung dieser Werte. Diese beiden Schritte bewirken, dass die infolge der Codierung möglichen endlich vielen, diskreten Werte für jeden Reflektionskoeffizienten so sinnvoll auf die genannten Wertebereiche verteilt werden, dass sich die Rundungsfehler, die sich beim Quantisieren der Koeffizienten ergeben, im Wiedergabesignal möglichst wenig bemerkbar machen. Im Codierer und Decodierer gelangen die quantisierten Filterkoeffizienten, und damit identische Filter, zum Einsatz, was für eine hohe Signalqualität unerlässlich ist.

Im Anschluss an die Filterung 1-A(z) werden für die leisen Segmente zwei Quantisierungsstufen berechnet, wobei die erste Stufe für die ersten 10 ms und die zweite Stufe für die zweiten 10 ms des Segments gilt, welches total 144 Abtastwerte aufweist. Die Quantisierungsstufen ergeben sich als mittlere Absolutwerte der Signalwerte in den Teilsegmenten. Für die Codierung stehen für jede Stufe vier Bit zur Verfügung. Es kommt eine quadratwurzelförmige Quantisierungskennlinie zum Einsatz, welche für schwache Signale eine feinere Auflösung ergibt als für die lauteren Signalelemente.

In Fig. 5 ist das Datenformat illustriert, mit dem die Parameter eines leisen Segments abgespeichert werden. Der Hintergrund ist mit Streifen belegt, deren Breite einem Bit entspricht. Zur Bezeichnung der aktuellen Betriebsart sind zwei Bit erforderlich, die Log Area Ratio des ersten und des zweiten Reflektionskoeffizienten k_1 und k_2 sind mit je fünf Bit codiert, jene des dritten Reflektionskoeffizienten k_3 mit vier Bit. Die beiden Quantisierungsstufen q_1 und q_2 sind ebenfalls mit je vier Bit codiert, so dass sich die gesamte Datenmenge auf 24 Bit beläuft. Die Datenformate der übrigen Segmente sind als ganzzahlige Vielfache von 24 gewählt; es handelt sich dabei um eine Anpassung an die Wortbreite des Motorola Signalprozessors DSP 56000.

Aus dem Flussdiagramm des Codierers (Fig. 3) ist ersichtlich, dass die Verarbeitungssequenz vom Schur-Algorithmus bis zur Filterung 1-A(z) für die leisen und die übrigen Segmente auf den ersten Blick übereinstimmen. Der Unterschied besteht, wie schon erwähnt, nur in der Filterordnung.

Für die nicht leisen Sprachsegmente führt die Signalverarbeitung anschliessend an die Filterung 1-A(z) zunächst zum gemeinsamen Block Pitch-Untersuchung. Erst anschliessend an diese erfolgen die Verzweigungen, die zu einer Aufgliederung in die verbleibenden drei Betriebsarten führen. Bevor nun die Verarbeitung der stimmhaften Segmente erläutert wird, wird zuerst diejenige der frikativen und der normalen Segmente beschrieben.

Zur Verarbeitung der frikativen Segmente gelangt man, wenn die Pitch-Untersuchung im Anschluss an die Filterung 1-A(z) keinen stimmhaften Signalverlauf detektiert und der Autokorrelationskoeffizienten r(1) kleiner als null ist. Diese letztere Bedingung bedeutet nämlich, dass im höherfrequenten Teil des Kurzzeit-Spektrums mehr Energie ist als im Teil mit den tieferen Frequenzen, und das bedeutet wiederum, dass es

sich um einen Zischlaut oder um Atemgeräusche handelt.

Die Verarbeitung der frikativen Segmente unterscheidet sich von derjenigen der leisen Segmente in zwei Punkten: Einerseits weist das Filter 1-A(z) eine höhere Filterordnung auf, und zwar beträgt diese wie bei den normalen und stimmhaften Segmenten acht. Und andererseits beträgt die Anzahl der Quantisierungsstufen bei der adaptiven Quantisierung, ebenfalls in Uebereinstimmung mit den Verhältnissen bei den normalen und stimmhaften Segmenten, drei.

Die Verarbeitung der acht Reflektionskoeffizienten umfasst die bereits für die leisen Segmente erläuterten Schritte: Begrenzung der Wertebereiche, Berechung der Log Area Ratio, Quantisierung mit linearer Kennlinie und Rückrechnung. Ein Unterschied zu den leisen Segmenten besteht darin, dass die ersten drei Koeffizienten mit höherer Auflösung codiert werden. Dann erfolgt die Berechnung der drei Quantisierungsstufen; ihre Codierung erfolgt gleich wie bei den leisen Segmenten.

Das Datenformat der frikativen Segmente ist in Fig. 6 abgebildet. Die Codierung der ersten vier Reflektionskoeffizienten k_1 bis k_4 erfolgt mit sieben, sechs, fünf und vier Bit, jene der letzten vier k_5 bis k_8 mit je drei Bit. Zusammen mit dem Codewort für die Betriebsart und mit den drei Quantisierungsstufen ergibt das eine Datenmenge von 48 Bit.

Zur Verarbeitung der normalen Segmente gelangt man ebenfalls erst im Anschluss an eine Pitch-Untersuchung, die keinen stimmhaften Signalverlauf erkennen konnte. Die Klasse der normalen Segmente umfasst dann alle jene Segmente, welche die Bedingung r(1) kleiner null für ein frikatives Segment nicht erfüllen.

Die Verarbeitung der normalen Segmente unterscheidet sich von jener der frikativen Segmente dadurch, dass in der ΔPCM-Schleife die Vorzeichenbits der einzelnen Signalwerte ermittelt und abgespeichert werden. Dazu muss vorgängig die spektrale Formung des Eingangssignals mit der Filterung $1/(1-A(z/\gamma))$ (Filter 3, Fig. 1) vervollständigt werden. Das Filter 3 (Fig. 7) ist wiederum ein Gitterfilter, aber mit der zum Filter 1 (Fig. 4) komplementären Struktur, wobei der Filterparameter γ jedem Verzögerungsglied z^{-1} multiplikativ vorangestellt ist.

Fig. 8 zeigt die Struktur des Nah-Prädiktionsfilters 9 (Fig. 1) in der ΔPCM-Schleife. Es handelt sich wieder um ein Gitterfilter mit einer dem Filter 1 (Fig. 5) ähnlichen Struktur. Beim Filter 1 gelangt das Eingangssignal auf dem oberen Signalpfad ohne Verzögerung und ohne Skalierung zum Ausgang, womit also der Anteil A(z) der Summe der vom unteren zum oberen Signalpfad gelangenden Teilsignal entspricht. Auf genau diese Weise bildet das Prädiktionsfilter von Fig. 8 die Schätzwerte. Die Implementierung des Filterparameters γ erfolgt wieder als Multiplikator vor jedem Verzögerungsglied z^{-1}.

Das Datenformat der normalen Segmente ergibt sich als Erweiterung des Datenformats der frikativen Segmente, wobei als zusätzliche Daten die in der ΔPCM-Schleife ermittelten Vorzeichenbits dazukommen. Entsprechend der Unterteilung der Segmente in drei Teilsegemente sind diese in drei Gruppen zu je 48 Bits zusammengefasst, woraus sich eine Gesamtdatenmenge von 192 Bits ergibt.

Ausgangspunkt für die Detektion der stimmhaften Segmente ist die Berechnung der Korrelationskoeffizienten (Pitch-Untersuchung, Fig. 3), wobei ρ^2 berechnet wird, damit im Signalprozessor auf das Wurzelziehen verzichtet werden kann. Die möglichen Pitch-Perioden sind auf 14 bis 141 Abtastintervalle, also auf 128 mögliche Werte, beschränkt, was zu einem 7 Bit Codewort für die Pitch-Periode führt.

Die Entscheidung für ein stimmhaftes Segment hängt von drei Bedingungen ab: Einmal muss der Quadratwert des grössten Korrelationskoeffizienten

$$\rho^2_{min} = 0,16$$

übersteigen, dann muss es sich um eine positive Korrelation handeln, und schliesslich darf der dem Koeffizienten eines Prädiktionsfilters erster Ordnung entsprechende Quotient einen bestimmten Maximalwert von 1,3 nicht übersteigen. Diese Bedingung verhindert den Einsatz eines Prädiktionsfilters mit sehr grosser Verstärkung, was sich gelegentlich in anklingenden stimmhaften Segmenten ergibt, und bewahrt dadurch den Codieralgorithmus vor möglicher Instabilität.

Der auf die beschriebene Weise getroffene Entscheid für ein stimmhaftes Segment ist erst vorläufig und bedeutet, dass im nächsten Schritt die Prädiktionskoeffizienten β_{-1}, β_0 und β_{+1} für ein transversales Pitch-Filter B(z) berechnet werden. Im Anschluss an die Berechnung der Filterkoeffizienten fällt dann der definitive Entscheid für oder gegen die Verarbeitung als stimmhaftes Segment.

Bei der Berechnung der Koeffizienten des Fern-Prädiktionsfilters oder Pitch-Prädiktors wird vorausgesetzt, dass die Grundperiode M der quasi-periodischen Anregung stimmhafter Laute aus der Pitch-Untersuchung bereits bekannt ist. Die gesuchten Filterkoeffizienten ergeben sich dann als Lösung einer

gewohnten Optimierungsaufgabe, bei der die Summe der Fehlerquadrate minimiert wird. Infolge der symmetrischen Struktur der in der Gleichung auftretenden Matrix, kann die Lösung effizient mit der sogenannten Cholesky-Zerlegung berechnet werden. Die Quantisierung der Filterkoeffizienten erfolgt mit den nach Tabelle 3 vorgängigen Umrechnungen, Extremwertbegrenzungen und Auflösung. Im Ausnahmefall, wenn die Summe der drei Filterkoeffizienten kleiner als der tabellierte Mindestwert von 0,1 ausfällt, wird die bisherige Entscheidung zugunsten eines stimmhaften Segmentes fallengelassen, andernfalls aber definitiv bestätigt.

$0,1 \leq$	Σ_{-0+}	:	β_{-1}	+	β_0	+	β_{+1}	$\leq$	0,5	7 Bit
$-0,1 \leq$	Σ_{-+}	:	β_{-1}	+	β_{+1}			$\leq$	0,3	5 Bit
$-0,3 \leq$	Δ_{-+}	:	β_{-1}	-	β_{+1}			$\leq$	0,3	5 Bit

Tabelle 3

Die Verarbeitung der stimmhaften Segmente unterscheidet sich von derjenigen der normalen Segmente durch den zusätzlichen Einsatz des Fern-Prädiktionsfilters in der ΔPCM-Schleife. Bei der Berechnung der Quantisierungsstufe muss die Wirkung des zusätzlichen Prädiktors angemessen berücksichtigt werden, was durch die vorgängige Filterung 1-B(z) des sonst direkt zur Berechnung herangezogenen Signals erfolgt. Die Berechnung der Quantisierungsstufen erfolgt auf die im Flussdiagramm von Fig. 3 angegebenen Art, ihre Codierung erfolgt wie bei den übrigen Segementen. Die Codierung der Pitch-Periode und der Koeffizienten des Fern-Prädiktionsfilters ergibt zur Datenmenge der normalen Segmente zusätzliche 24 Bits.

Der Decodierer (Fig. 2) enthält neben Teilen, welche der Codierer funktionsmässig mitenthält, zwei besondere Elemente, die im Codierer nicht vorkommen, es sind das die Rauschquelle 13 und das Filter 14.

Bei der Rauschquelle handelt es sich um ein 24 Bit lineares, rückgekoppeltes Schieberegister, dass eine Maximallängensequenz der Länge 2^{24} -1 erzeugt, in welcher die einzelnen Bits in pseudo-zufälliger Reihenfolge erscheinen. Die Definition des Schieberegisters, das heisst die Anordnung der XOR-Rückführung ist dem Buch "Error-Correcting Codes" von W.W. Peterson, E.J. Weldon, MIT Press, Cambridge, Massachusetts, 1972; Appendix C: Tables of Irreducible Polynomials over GF(2), S. 472-492, entnommen.

Es werden je vier aufeinanderfolgende Bits zu einer Zufallszahl zusammengefasst, die als binäre Bruchzahl interpretiert wird. Diese Zufallszahlen werden mit einem festen fünften, rechts anschliessenden (LSB)Bit = 1 symmetrisch um null angeordnet. Das Zufallssignal der Rauschquelle das sich aus den aufeinanderfolgenden Zufallszahlen zusammensetzt, wird mit der für jedes Teilsegement codierten Quantisierungsstufe multipliziert. Auf diese Weise ergibt sich das sogenannte synthetische Anregungssignal in den leisen und frikativen Segmenten.

Der mittlere Absolutwert der aufeinanderfolgenden Zufallszahlen beträgt $\frac{1}{2}$. Durch Multiplikation mit der Quantisierungsstufe, die ihrerseits als mittlerer Absolutwert errechnet wurde, ergibt sich damit ein systematisch um 6 dB zu leises synthetisches Anregungssignal, womit die für frikative Segmente zweifach verstärkenden Wirkungen von festen Hochpass-Vorfilter und adaptiver Formanten-Ueberbetonung sinnvoll kompensiert werden. Im weiteren wird diese Absenkung der Signalleistung in den leisen Segmenten subjektiv als qualitätssteigernd empfunden.

Das adaptive Filter 14 dessen Struktur in Fig. 9 dargestellt ist, dient der inversen spektralen Formung und Ueberbetonung der Formanten. Es handelt sich um eine Serieschaltung der beiden in den Fig. 4 und 7 dargestellten Filter-Strukturen. Gibt man α im ersten Teilfilter einen etwas kleineren Wert als dem Parameter γ im Codierer, so werden die im decodierten Sprachsignal teilweise vorhandenen Formanten nicht ganz flach geglättet. Das nachfolgende zweite Teilfilter vermag einem Signal mit flachem Spektrum die im ursprünglichen Signal enthaltenen Formanten in voller Stärke einzuprägen. Seine Anwendung auf das Signal mit nicht vollständig flachem Spektrum bewirkt die gewünschte Ueberbetonung der dominanten Signalanteile.

Patentansprüche

1. Verfahren zur Sprachdigitalisierung unter Anwendung der Signalformcodierung, mit einem Codierer zur Digitalisierung und mit einem Decodierer zur Rekonstruktion des Sprachsignals, bei welchem im Codierer das Sprachsignal in Segmente unterteilt und unter möglichst genauer Annäherung der Abtastwerte verarbeitet wird, wobei anhand von bekannten Abtastwerten eine Berechnung eines Schätzwertes für bevorstehende, neue Abtastwerte erfolgt, dadurch gekennzeichnet, dass die Berechnung des Schätzwertes nur in einem Teil der Segmente erfolgt und im anderen Teil der Segmente nur Parameter für eine Sprachnachbildung im Sinn der Quellencodierung abgeleitet werden, und dass die einzelnen Sprachsegmente mit variabler Bitrate verarbeitet werden, wobei diese Bitraten verschiedenen Betriebsarten zugeordnet sind und jedes Signalelement in eine der Betriebsarten klassiert wird.

2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass die Signalquantisierung mit einem oder zwei Bit pro Abtastwert erfolgt, und dass dieser Signalverarbeitungsschritte vor- und nachgelagert werden, durch welche das Quantisierungsgeräusch weitgehend der Wahrnehmung entzogen und der natürliche Klang der Sprache bewahrt wird.

3. Verfahren nach Anspruch 2, dadurch gekennzeichnet, dass die vor- und nachgelagerten Signalverarbeitungsschritte eine feste Hochpassfilterung (HP), eine adaptive Vorfilterung (1, 3), eine adaptive Nachfilterung (14) und eine Abschwächung des Anregungssignals in Segmenten mit Zischlauten umfasst.

4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, dass vier Betriebsarten festgelegt werden, eine erste für nachfolgend als leise Segmente bezeichnete Sprechpausen, eine zweite für nachfolgend als frikative Segmente bezeichnete Zischlaute, eine dritte für nachfolgend als normale Segmente bezeichnete Signalsegmente, die keine besonders sparsame Codierung zulassen, und eine vierte für nachfolgend als stimmhafte Segmente bezeichnete stimmhafte Laute.

5. Verfahren nach Anspruch 4, dadurch gekennzeichnet, dass die leisen Segmente mit 24 Bit pro Segment, die frikativen Segmente mit 48 Bit pro Segment, die normalen Segmente mit 192 Bit pro Segment und die stimmhaften Segmente mit 216 Bit pro Segment codiert werden.

6. Verfahren nach Anspruch 5, dadurch gekennzeichnet, dass als Mass für die in einem Segment enthaltene Energie der Autokorrelationskoeffizient (r(0)) verwendet wird, und dass der Entscheid, ob es sich um ein leises Segment handelt, durch einen Vergleich des Autokorrelationskoeffizienten mit einer adaptiv nachgeführten Schwelle erfolgt.

7. Verfahren nach Anspruch 6, dadurch gekennzeichnet, dass die nicht als leise bewerteten Segmente einer sogenannten Pitch-Untersuchung zur Detektion eines stimmhaften Signalverlaufs unterzogen werden, und dass von den als nicht stimmhaft bewerteten Signalelementen diejenigen mit einem Autokorrelationskoeffizienten (r(1)) kleiner null als frikative Segmente bewertet werden.

8. Verfahren nach Anspruch 7, dadurch gekennzeichnet, dass in allen Betriebsarten eine Stufenberechnung durchgeführt wird, bei welcher jedes Signalsegment in Teilsegmente unterteilt und für jedes Teilsegment ein neuer, dem Signalverlauf angepasster Stufenwert berechnet wird, wobei leise Segmente in zwei und die übrigen Segmente in drei Teilsegmente unterteilt werden.

9. Verfahren nach Anspruch 8, dadurch gekennzeichnet, dass bei den normalen und stimmhaften Segmenten nach der Stufenberechnung eine adaptive Quantisierung der einzelnen Signalwerte mit dem bei der Stufenberechnung ermittelten, dem Signalverlauf angepassten Stufenwert erfolgt.

10. Vorrichtung zur Durchführung des Verfahrens nach Anspruch 1, mit einem Codierer und einem Decodierer, dadurch gekennzeichnet, dass im Codierer ein adaptives Nah-Prädiktionsfilter (9) zur Berechnung des Schätzwertes für den unmittelbar bevorstehenden, neuen Abtastwert in dem einen Teil der Segmente, ein adaptives Fern- Prädiktionsfilter (10) für den Einsatz in stimmhaften Signalelementen und Mittel zur Untersuchung der Signalelemente und zu deren Zuordnung zu den einzelnen Betriebsarten vorgesehen sind.

11. Vorrichtung nach Anspruch 10, gekennzeichnet durch einen adaptiven Quantisierer (6, 8) zur Unterteilung der Signalsegmente in Teilsegmente und zur Berechnung eines dem Signalverlauf angepassten

Stufenwertes.

12. Vorrichtung nach Anspruch 11, dadurch gekennzeichnet, dass der Quantisierer (8) und die beiden Prädiktionsfilter (9, 10) zusammen mit vier Summationspunkten eine sogenannte ΔPCM-Schleife bilden, dass das Eingangssignal des Nah-Prädiktionsfilters (9) durch das quantisierte Sprachsignal gebildet ist, und dass das Ausgangssignal des Nah-Prädiktionsfilters vom quantisierten Sprachsignal subtrahiert wird und diese Differenz in das Fern-Prädiktionsfilter (10) gelangt.

13. Vorrichtung nach Anspruch 12, dadurch gekennzeichnet, dass der PCM-Schleife eine festes Hochpassfilter (HP) und adaptive Vorfilter (1, 3) vorgeschaltet sind, dass im Decodierer eine synthetische Rauschquelle (13) und ein adaptives Nachfilter (14) angeordnet sind, dessen Filterkoeffizient (α) kleiner ist als derjenige (γ) des einen Vorfilters (3), und dass die synthetische Rauschquelle mit einer reduzierten Signalleistung betrieben wird.

14. Vorrichtung nach Anspruch 13, dadurch gekennzeichnet, dass die beiden Vorfilter (1, 3) und das Nachfilter (14) durch sogenannte Gitter-Filter gebildet sind.

15. Vorrichtung nach Anspruch 12, dadurch gekennzeichnet, dass das Fern-Prädiktionsfilter (10) durch ein Transversalfilter gebildet ist.

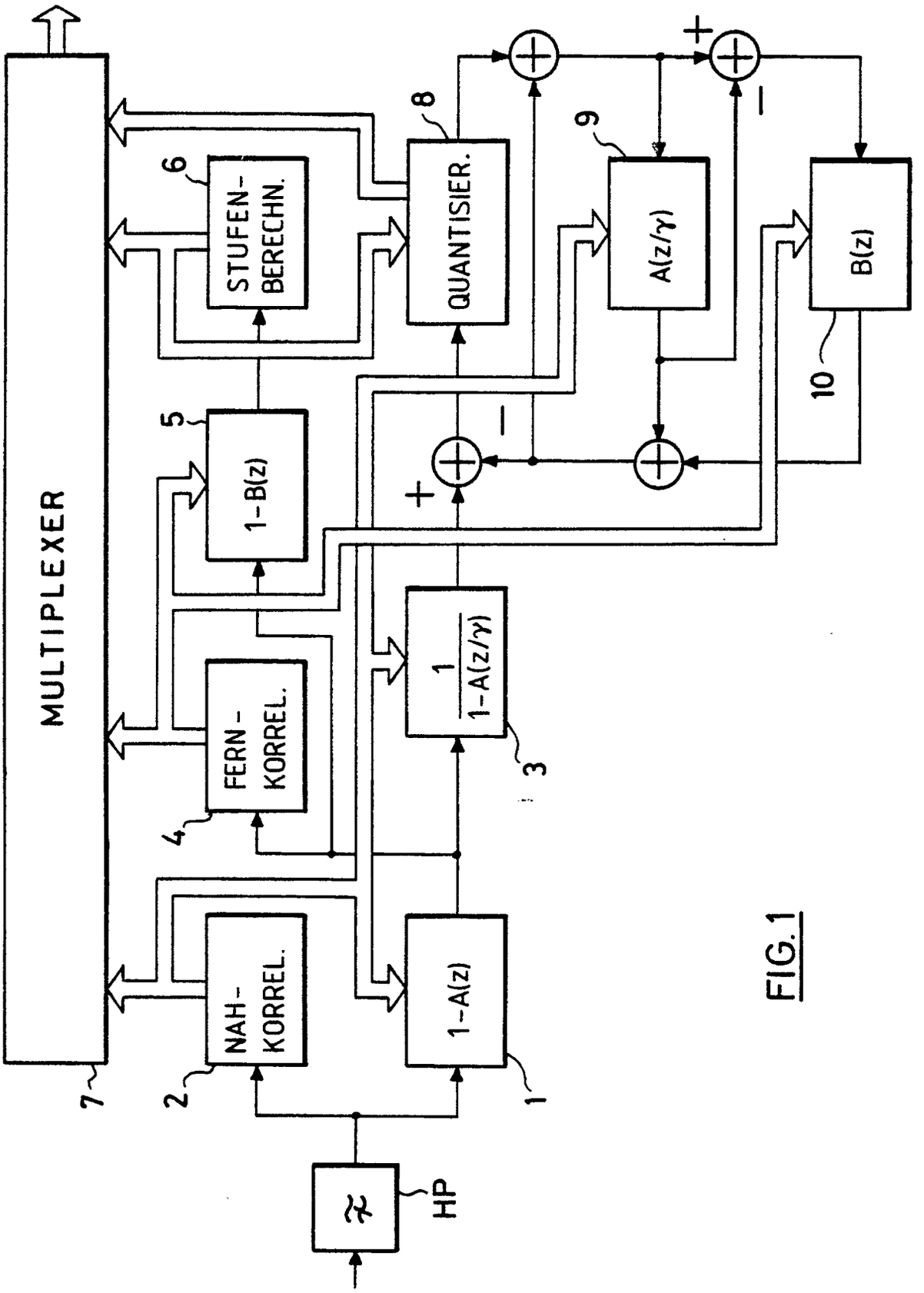
MULTIPLEXER
7
NAH-KORREL.
2
FERN-KORREL.
4
STUFEN-BERECHN.
6
1-B(z)
5
1-A(z)
1
1/(1-A(z/γ))
3
QUANTISIER.
8
A(z/γ)
9
B(z)
10
HP
+
−
+
−

FIG. 1

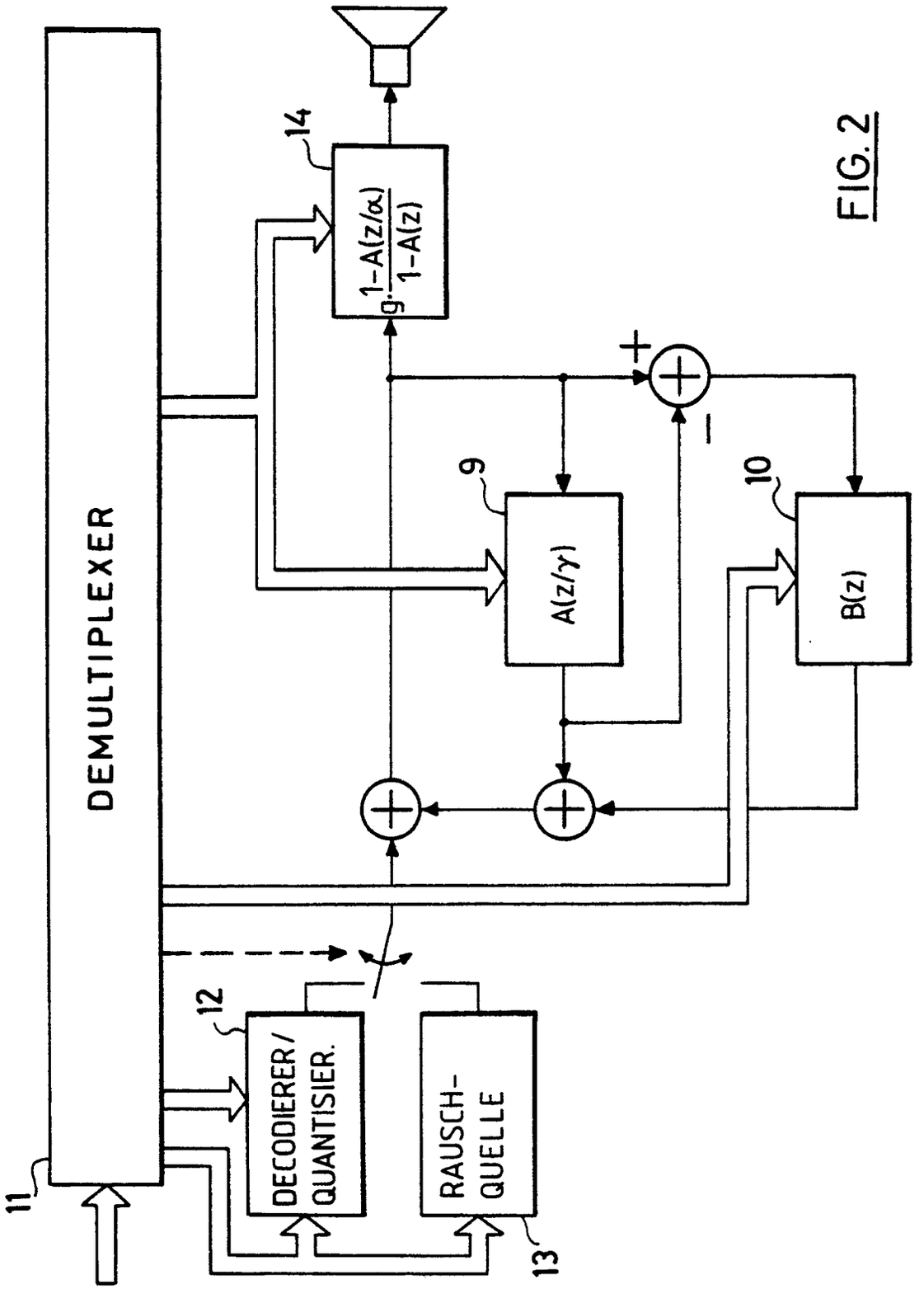
DEMULTIPLEXER
11
DECODIERER/ QUANTISIER.
12
RAUSCH- QUELLE
13
A(z/γ)
9
B(z)
10
1-A(z/α) / 1-A(z)
14
g.
+
−

FIG. 2

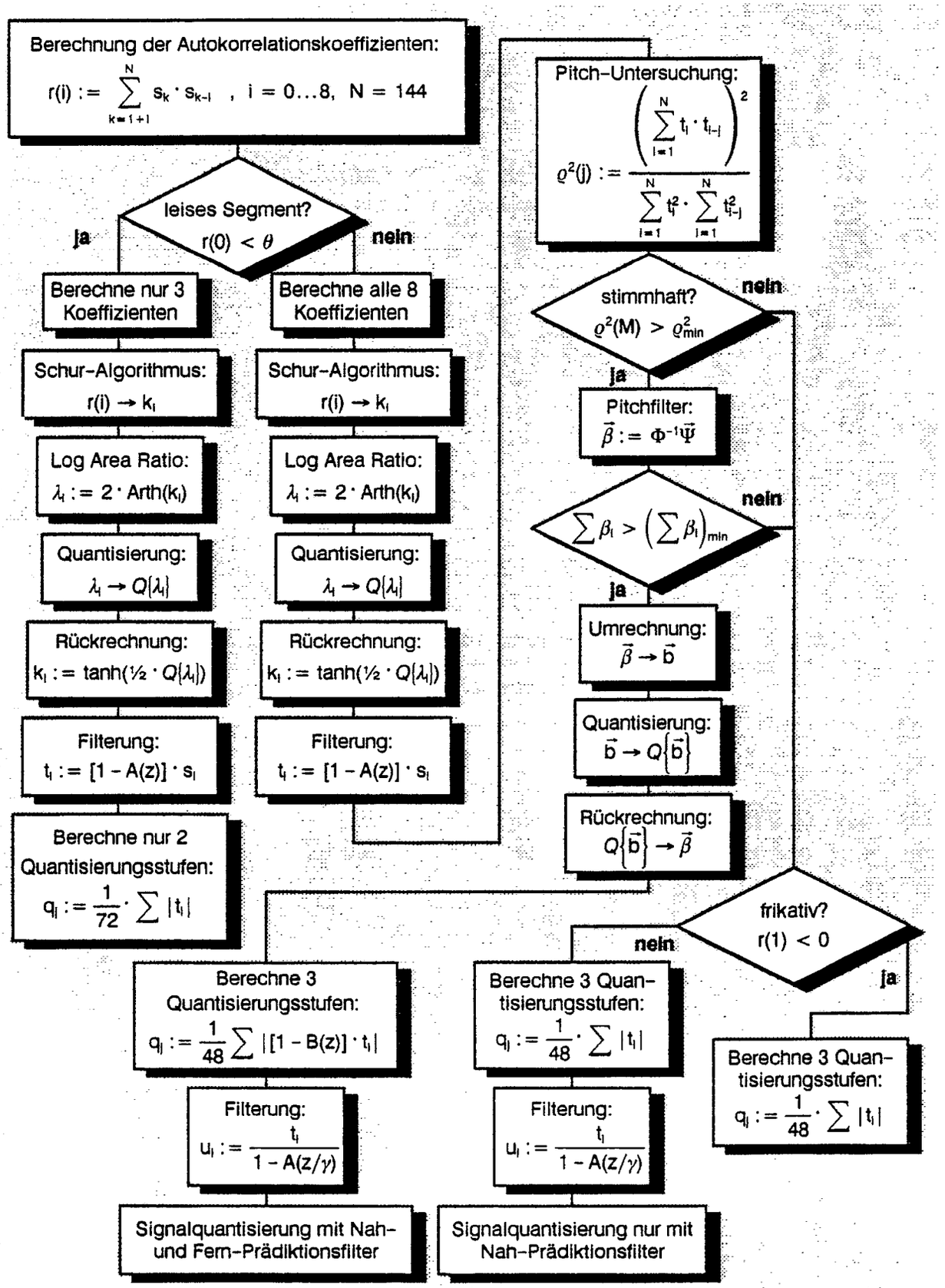
Berechnung der Autokorrelationskoeffizienten:
$r(i) := \sum_{k=1+i}^{N} s_k \cdot s_{k-i}$, i = 0...8, N = 144
leises Segment?
r(0) < θ
ja
nein
Berechne nur 3 Koeffizienten
Berechne alle 8 Koeffizienten
Schur-Algorithmus:
r(i) → k_i
Schur-Algorithmus:
r(i) → k_i
Log Area Ratio:
λ_i := 2 · Arth(k_i)
Log Area Ratio:
λ_i := 2 · Arth(k_i)
Quantisierung:
λ_i → Q{λ_i}
Quantisierung:
λ_i → Q{λ_i}
Rückrechnung:
k_i := tanh(½ · Q{λ_i})
Rückrechnung:
k_i := tanh(½ · Q{λ_i})
Filterung:
t_i := [1 − A(z)] · s_i
Filterung:
t_i := [1 − A(z)] · s_i
Berechne nur 2 Quantisierungsstufen:
$q_j := \frac{1}{72} \cdot \sum |t_i|$
Pitch-Untersuchung:
$\varrho^2(j) := \frac{\left(\sum_{i=1}^{N} t_i \cdot t_{i-j}\right)^2}{\sum_{i=1}^{N} t_i^2 \cdot \sum_{i=1}^{N} t_{i-j}^2}$
stimmhaft?
$\varrho^2(M) > \varrho^2_{min}$
nein
ja
Pitchfilter:
$\vec{\beta} := \Phi^{-1}\vec{\Psi}$
$\sum \beta_i > \left(\sum \beta_i\right)_{min}$
nein
ja
Umrechnung:
$\vec{\beta} \rightarrow \vec{b}$
Quantisierung:
$\vec{b} \rightarrow Q\{\vec{b}\}$
Rückrechnung:
$Q\{\vec{b}\} \rightarrow \vec{\beta}$
frikativ?
r(1) < 0
nein
ja
Berechne 3 Quantisierungsstufen:
$q_j := \frac{1}{48} \sum |[1 - B(z)] \cdot t_i|$
Berechne 3 Quantisierungsstufen:
$q_j := \frac{1}{48} \cdot \sum |t_i|$
Berechne 3 Quantisierungsstufen:
$q_j := \frac{1}{48} \cdot \sum |t_i|$
Filterung:
$u_i := \frac{t_i}{1 - A(z/\gamma)}$
Filterung:
$u_i := \frac{t_i}{1 - A(z/\gamma)}$
Signalquantisierung mit Nah- und Fern-Prädiktionsfilter
Signalquantisierung nur mit Nah-Prädiktionsfilter

FIG. 3

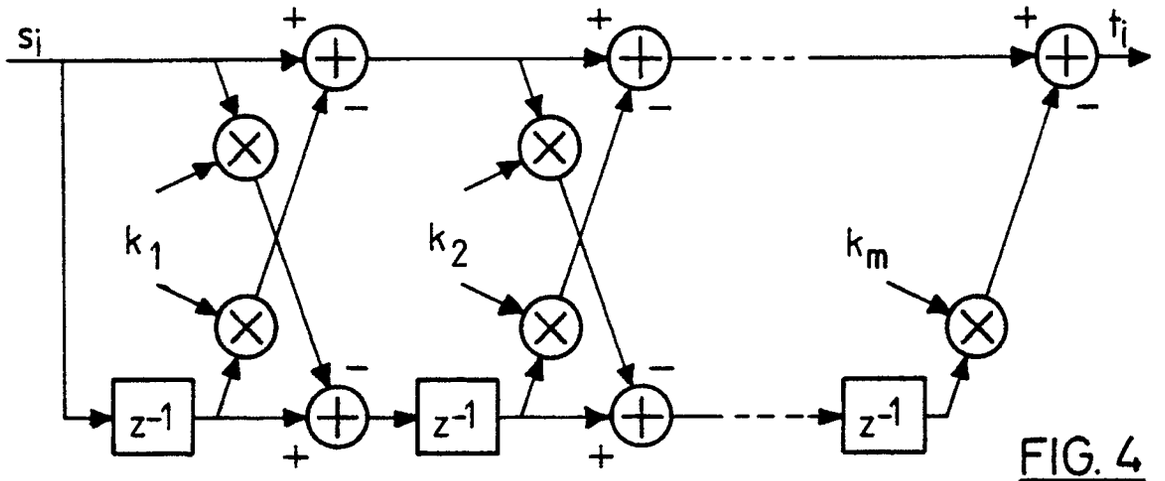

FIG. 4

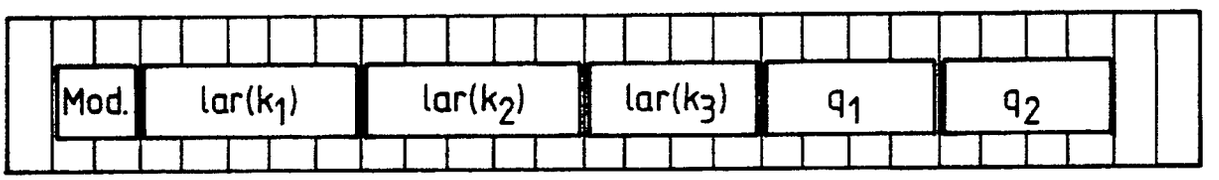

FIG. 5

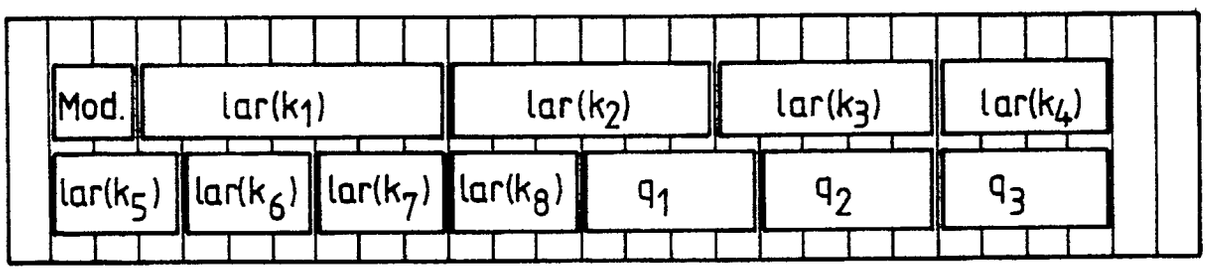

FIG. 6

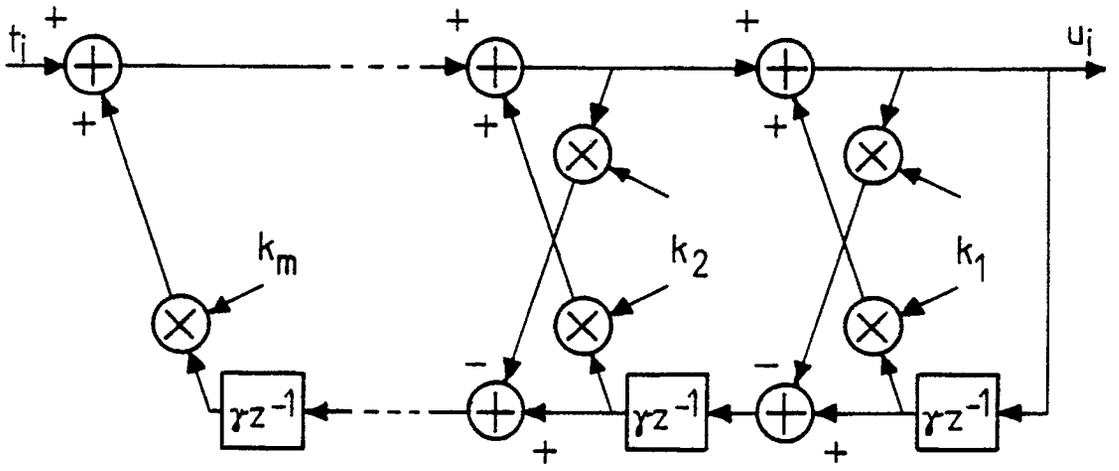
f_i
u_i
k_m
k_2
k_1
γz^{-1}
γz^{-1}
γz^{-1}

FIG. 7

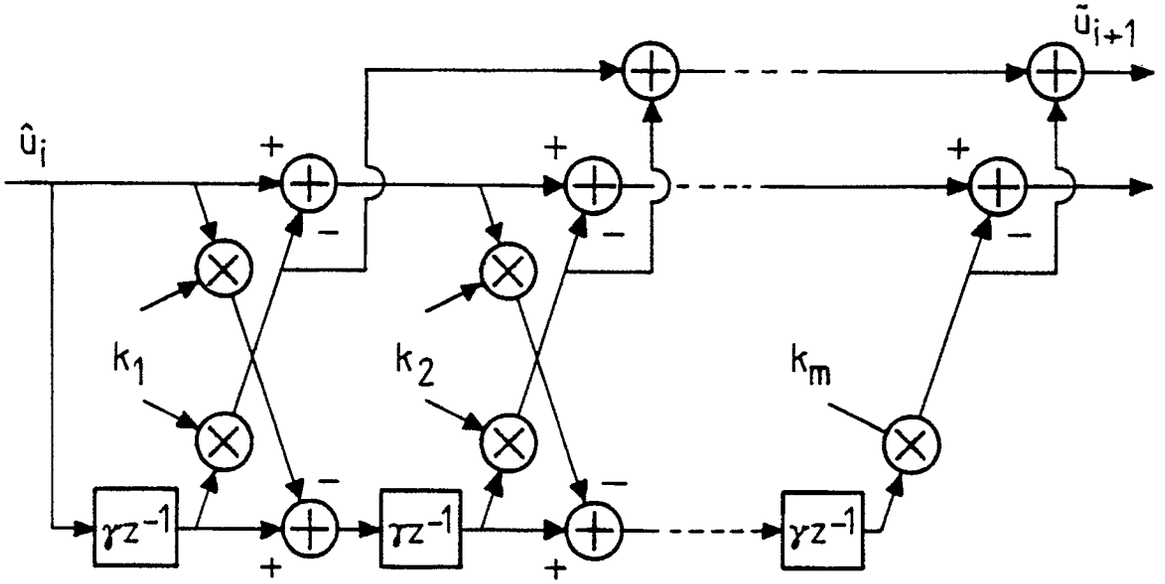
$\hat{u}_i$
$\tilde{u}_{i+1}$
k_1
k_2
k_m
γz^{-1}
γz^{-1}
γz^{-1}

FIG. 8

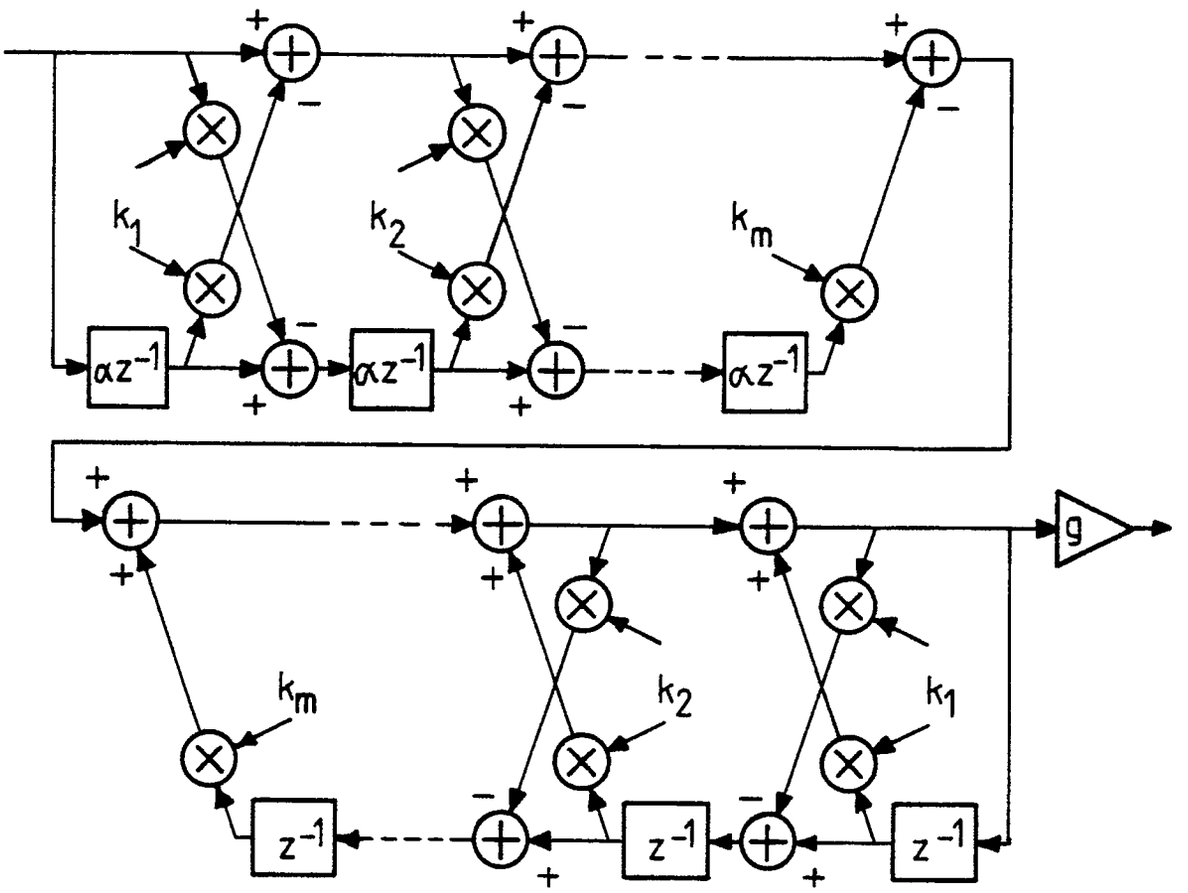

FIG. 9