

12

EUROPÄISCHE PATENTANMELDUNG

21 Anmeldenummer: 82810391.1

51 Int. Cl.³: **G 10 L 1/08**

22 Anmeldetag: 20.09.82

30 Priorität: 24.09.81 CH 6168/81

43 Veröffentlichungstag der Anmeldung:
06.04.83 Patentblatt 83/14

84 Benannte Vertragsstaaten:
AT CH DE FR GB IT LI NL SE

71 Anmelder: **GRETAG Aktiengesellschaft**
Althardstrasse 70
CH-8105 Regensdorf(CH)

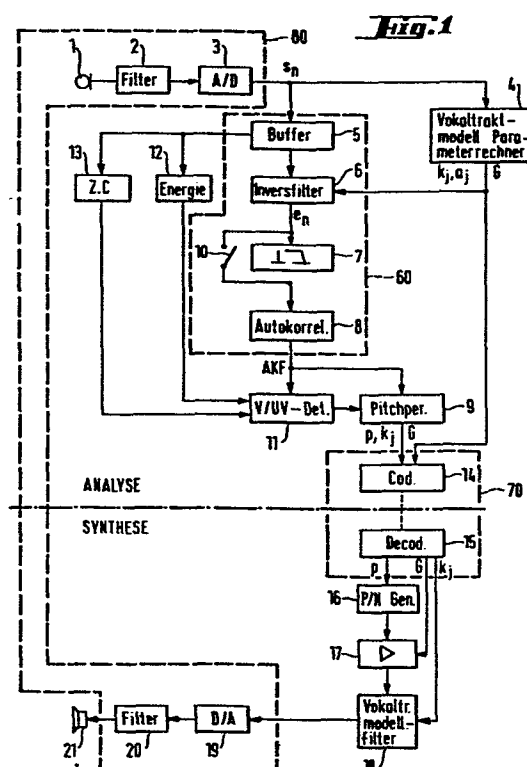
72 Erfinder: **Horvath, Stephan, Dr.**
Ferdinand Hodlerstrasse 16
CH-8049 Zürich(CH)

72 Erfinder: **Bernasconi, Carlo**
Winterthurerstrasse 139
CH-8057 Zürich(CH)

74 Vertreter: **Pirner, Wilhelm et al,**
Patentabteilung der CIBA-GEIGY AG Postfach
CH-4002 Basel(CH)

54 Verfahren und Vorrichtung zur redundanzvermindernden digitalen Sprachverarbeitung.

57 Das Sprachsignal wird nach Digitalisierung in Abschnitte eingeteilt und jeder Abschnitt wird nach den Methoden der linearen Prädiktion analysiert, wobei die Koeffizienten eines Klangbildungsmodellfilters, ein Lautstärkeparameter, eine Information über die stimmhafte oder stimmlose Anregung und im ersteren Falle die Periode der Stimmbandgrundfrequenz ermittelt werden. Zur Verbesserung der Sprachqualität ohne Datenratenerhöhung wird die Anzahl der Sprachabschnitte pro Sekunde erhöht, dafür aber gleichzeitig eine besondere, redundanzvermindernde Codierung der Sprachparameter vorgenommen. Die Codierung der Sprachparameter erfolgt blockweise für jeweils zwei oder drei benachbarte Sprachabschnitte, und zwar in unterschiedlicher Weise je nach dem, ob der betreffende Sprachabschnittsblock mit einem stimmhaften oder einem stimmlosen Abschnitt beginnt. Die Parameter der jeweils ersten Sprachabschnitte werden in vollständiger Form codiert, die der übrigen Sprachabschnitte in differentieller Form oder teilweise überhaupt nicht. Der auf diese Weise verminderte mittlere Bitbedarf pro Sprachabschnitt kompensiert die erhöhte Abschnittsrate, sodass insgesamt die Datenrate nicht erhöht wird.



9-13565

Verfahren und Vorrichtung zur redundanzvermindernden digitalen
Sprachverarbeitung.

Die Erfindung betrifft ein nach der Methode der linearen Prädiktion arbeitendes Verfahren und eine entsprechende Vorrichtung zur redundanzvermindernden digitalen Sprachverarbeitung gemäss dem Oberbegriff von Patentanspruch 1 bzw. Patentanspruch 13.

Derartige Sprachverarbeitungssysteme, sogenannte LPC-Vocoder, erlauben eine erhebliche Redundanzreduktion bei der digitalen Uebertragung von Sprachsignalen. Sie gewinnen heute immer mehr an Bedeutung und sind Gegenstand zahlreicher Veröffentlichungen und Patente, von denen hier nur einige repräsentative rein beispielsweise angeführt sind:

B.S. Atal und S.L. Hanauer, Journal Acoust. Soc. Am., 50, S. 637-655, 1971

R.W. Schafer und L.R. Rabiner, Proc. IEEE Vol 63 No.4, S. 662-667, 1975

L.R. Rabiner et al., Trans. Acoustics, Speech and Signal Proc., Vol 24 No. 5, S. 399-418, 1976

B. Gold, Proc. IEEE Vol. 65 No. 12, S. 1636-1658, 1977

A. Kurematsu et al, Proc. IEEE, ICASSP, Washington 1979, S. 69-72

S. Horvath, "LPC-Vocoder, Entwicklungsstand und Perspektiven", Sammelband Kolloquiumsvorträge "Krieg im Aether" XVII. Folge, Bern 1978

US-PS 3 624 302

US-PS 3 361 520

US-PS 3 909 533

US-PS 4 230 906

Die heute bekannten und erhältlichen LPC-Vocoder arbeiten noch nicht voll zufriedenstellend. Zwar ist die nach der Analyse wieder synthetisierte Sprache meistens noch relativ verständlich, jedoch ist sie verzerrt und tönt künstlich. Eine der Ursachen dafür liegt u.a. in der Schwierigkeit, den Entscheid, ob ein stimmhafter oder ein stimmloser Sprachabschnitt vorliegt, mit ausreichender Sicherheit zu treffen. Weitere Ursachen sind mangelhafte Bestimmung der Pitchperiode und ungenaue Bestimmung der Klangbildungsfilterparameter.

Neben diesen grundsätzlichen Schwierigkeiten ergibt sich eine weitere wesentliche Schwierigkeit daraus, dass die Datenrate in vielen Fällen auf einen relativ niedrigen Wert begrenzt sein muss. Sie beträgt z.B. bei Telefonnetzen vorzugsweise nur 2,4 kbit/sec. Bei einem LPC-Vocoder ist die Datenrate durch die Anzahl der in jedem Sprachabschnitt analysierten Sprachparameter, durch die Anzahl der für diese Parameter benötigten Bits und durch die sog. Frame-Rate, d.h. die Anzahl Sprachabschnitte pro Sekunde gegeben. Bei den heute gebräuchlichen Systemen werden, damit überhaupt eine einigermaßen brauchbare Sprachwiedergabe möglich ist, pro Sprachabschnitt minimal etwas über 50 Bit benötigt. Damit ist die maximale Frame-Rate automatisch festgelegt, bei einem 2,4 kbit/sec-System z.B. auf rund 45/sec. Die Sprachqualität bei diesen relativ geringen Frame-Raten ist auch entsprechend schlecht. Eine Erhöhung der Frame-Rate, die sich zur Verbesserung der Sprachqualität an sich anböte, ist nicht möglich, da dadurch die festgelegte Datenrate überschritten würde. Für die Erniedrigung der Anzahl der pro Frame benötigten Bits wäre andererseits eine Verminderung der Anzahl der verwendeten Parameter bzw. eine Vergrößerung von deren Quantisierung nötig, was jedoch automatisch wieder auf eine Verschlechterung der Sprachwiedergabequalität hinauslaufen würde.

Die vorliegende Erfindung befasst sich nun vornehmlich mit diesen durch vorgegebene Datenraten bedingten Schwierigkeiten und hat insbesondere zum Ziel, ein Verfahren bzw. eine Vorrichtung der eingangs definierten Art hinsichtlich der Sprachwiedergabequalität zu verbessern, ohne dabei die Datenraten zu erhöhen.

- 3 -

Das erfindungsgemässe Verfahren und die erfindungsgemässe Vorrichtung sind in den Patentansprüchen 1 und 13 beschrieben. Bevorzugte Ausführungsformen ergeben sich aus den abhängigen Ansprüchen.

Der Grundgedanke der Erfindung besteht also darin, durch eine verbesserte Codierung der Sprachparameter Bits einzusparen, sodass die Frame-Rate erhöht werden kann. Andererseits besteht aber auch insofern eine Wechselbeziehung zwischen der Codierung der Parameter und der Frame-Rate, als eine weniger bit-intensive, redundanzvermindernde Codierung erst bei höheren Frame-Raten möglich bzw. sinnvoll ist. Dies rührt u.a. daher, dass die erfindungsgemässe Codierung der Parameter auf der Ausnützung der Korrelation zwischen benachbarten stimmhaften Sprachabschnitten (Interframe-Korrelation) basiert, welche mit zunehmender Frame-Rate natürlich immer stärker wird.

Im folgenden wird die Erfindung anhand der Zeichnungen näher erläutert. Es zeigen:

- Fig. 1 ein stark vereinfachtes Blockschaltbild eines LPC-Vocoders,
- Fig. 2 ein Blockschaltbild eines entsprechenden Multi-Prozessor-Systems und
- Fig. 3 und 4 ein Flussschema für ein Programm zur Durchführung einer Variante der erfindungsgemässen Codierung.

Der allgemeine Aufbau und die Funktionsweise der erfindungsgemässen Sprachverarbeitungsvorrichtung gehen aus Fig. 1 hervor. Das von irgendeiner Quelle, z.B. einem Mikrophon 1 stammende analoge Sprachsignal wird in einem Filter 2 bandbegrenzt und dann in einem A/D-Wandler 3 abgetastet und digitalisiert. Die Abtastrate beträgt dabei etwa 6 bis 16 kHz, vorzugsweise etwa 8 kHz.

Die Auflösung ist etwa 8 bis 12 bit. Der Durchlassbereich des Filters 2 erstreckt sich bei sog. Breitbandsprache gewöhnlich von ca. 80 Hz bis etwa 3,1-3,4 kHz, bei Telefonsprache von etwa 300 Hz bis 3,1-3,4 kHz.

Für die nun folgende digitale Verarbeitung des Sprachsignals wird dieses in aufeinanderfolgende, vorzugsweise überlappende Sprachabschnitte, sog. Frames, eingeteilt. Die Sprachabschnittslänge beträgt etwa 10 bis 30 msec, vorzugsweise etwa 20 msec. Die Frame-Rate, d.h. die Anzahl von Frames pro Sekunde, beträgt etwa 30-100, vorzugsweise etwa 50 bis 70. Im Interesse hoher Auflösung und damit Sprachqualität bei der Synthetisierung sind möglichst kurze Abschnitte und entsprechend hohe Frame-Raten erstrebenswert, jedoch stehen dem einerseits bei Echtzeit-Verarbeitung das begrenzte Leistungsvermögen des eingesetzten Computers und andererseits die Forderung möglichst niedriger Bitraten bei der Uebertragung entgegen.

Für jeden dieser Sprachabschnitte erfolgt nun eine Analyse des Sprachsignals nach den Prinzipien der linearen Prädiktion, wie sie z.B. in den eingangs erwähnten Publikationen beschrieben sind. Grundlage der linearen Prädiktion ist ein parametrisches Modell der Spracherzeugung. Ein zeitdiskretes Allpol-Digitalfilter modelliert die Klangformung durch Hals- und Mundtrakt (Vokaltrakt). Bei stimmhaften Lauten ist die Anregung dieses Filters eine periodische Pulsfolge, deren Frequenz, die sog. Pitchfrequenz, die periodische Anregung durch die Stimmbänder idealisiert. Bei stimmlosen Lauten ist die Anregung weisses Rauschen, idealisierend für die Luftturbulenz im Hals bei nicht angeregten Stimmbändern. Ein Verstärkungsfaktor schliesslich kontrolliert die Lautstärke. Auf der Grundlage dieses Modells ist somit das Sprachsignal durch die folgenden Parameter vollständig bestimmt:

1. Die Information, ob der zu synthetisierende Laut stimmhaft oder stimmlos ist,
2. die Pitch-Periode (bzw. die Pitch Frequenz) bei stimmhaften Lauten (bei stimmlosen ist die Pitchperiode per def. gleich 0)
3. die Koeffizienten des zugrundegelegten Allpol-Digitalfilters (Vokaltraktmodells) und

4. der Verstärkungsfaktor.

Die Analyse gliedert sich demnach im wesentlichen in zwei Hauptprozeduren, und zwar zum einen in die Berechnung des Verstärkungsfaktors bzw. Lautstärkeparameters sowie der Koeffizienten bzw. Filterparameter des zugrundeliegenden Vokaltrakt-Modellfilters und zum anderen in den Stimmhaft-Stimmlos-Entscheid und in die Ermittlung der Pitch-Periode im stimmhaften Falle.

Die Filterkoeffizienten werden in einem Parameterrechner 4 durch Lösung des Gleichungssystems gewonnen, welches erhalten wird, wenn die Energie des Prädiktionsfehlers, d.h. die Energie der Differenz zwischen den tatsächlichen Abtastwerten und den aufgrund der Modellannahme geschätzten Abtastwerten im betrachteten Intervall (Sprachabschnitt) in Funktion der Koeffizienten minimiert wird. Die Auflösung des Gleichungssystems erfolgt vorzugsweise nach der Autokorrelationsmethode mittels eines Algorithmus' nach Durbin (vgl. z.B. L.B. Rabiner and R.W. Schafer, "Digital Processing of Speech Signals", Prentice-Hall Inc., Englewood Cliffs, N.J., 1978, Seiten 411-413). Dabei ergeben sich neben den Filterkoeffizienten bzw. -parametern (a_j) gleichzeitig auch die sogenannten Reflexionskoeffizienten (k_j), welche auf Quantisierung weniger empfindliche Transformierte der Filterkoeffizienten (a_j) sind. Die Reflexionskoeffizienten sind bei stabilen Filtern dem Betrag nach stets kleiner als 1 und ausserdem nimmt ihr Betrag mit zunehmender Ordnungszahl ab. Wegen dieser Vorteile werden bevorzugt diese Reflexionskoeffizienten (k_j) statt der Filterkoeffizienten (a_j) übertragen. Der Lautstärkeparameter G ergibt sich aus dem Algorithmus als Nebenprodukt.

Zur Auffindung der Pitch-Periode p (Periode der Stimmgrundfrequenz) wird das digitale Sprachsignal s_n in einem Buffer 5 zunächst solange zwischengespeichert, bis die Filterparameter (a_j) berechnet sind. Dann passiert das Signal ein mit den Parametern (a_j) eingestelltes Inversfilter 6, welches eine zur Übertragungsfunktion des Vokaltraktmodellfilters inverse Übertragungsfunktion besitzt. Das Ergebnis dieser

Invers-Filterung ist ein Prädiktionsfehlersignal e_n , welches dem mit dem Verstärkungsfaktor G multiplizierten Anregungssignal x_n ähnlich ist. Dieses Prädiktionsfehlersignal e_n wird nun im Falle von Telefonsprache direkt oder im Falle von Breitbandsprache über ein Tiefpassfilter 7 einer Autokorrelationsstufe 8 zugeführt, welches daraus die auf das Autokorrelationsmaximum nullter Ordnung normierte Autokorrelationsfunktion AKF bildet, anhand welcher in einer Pitchextraktionsstufe 9 die Pitchperiode p ermittelt wird, und zwar in bekannter Weise als Abstand des zweiten Autokorrelationsmaximums R_{XX} vom ersten Maximum (nullter Ordnung), wobei vorzugsweise ein adaptives Suchverfahren angewandt wird.

Die Klassifikation des betrachteten Sprachabschnitts als stimmhaft bzw. stimmlos erfolgt in einer Entscheidungsstufe 11 nach bestimmten Kriterien, welche u.a. auch die Energie des Sprachsignals und die Anzahl der Nulldurchgänge desselben im betrachteten Abschnitt beinhalten. Diese beiden Werte werden in einer Energiebestimmungsstufe 12 und einer Nulldurchgangsbestimmungsstufe 13 ermittelt.

Der vorstehend beschriebene Parameterrechner ermittelt pro Sprachabschnitt (Frame) je einen Satz Filterparameter. Selbstverständlich könnten die Filterparameter auch anders bestimmt werden, beispielsweise laufend mittels einer adaptiven inversen Filtrierung oder eines anderen bekannten Verfahrens, wobei die Filterparameter zwar mit jedem Abtasttakt laufend nachgeregelt, aber nur jeweils zu den durch die Frame-Rate festgelegten Zeitpunkten für die weitere Verarbeitung bzw. Uebertragung bereitgestellt werden. Die Erfindung ist diesbezüglich in keiner Weise eingeschränkt. Wesentlich ist lediglich, dass für jeden Sprachabschnitt ein Satz Filterparameter vorliegt.

Die nach der eben geschilderten Methode gewonnenen Parameter (k_j) , G und p werden dann einer Codierungsstufe 14 zugeführt, wo sie in noch näher zu beschreibender Weise in eine für die Uebertragung geeignete, besonders bit-rationelle Form gebracht (formatiert) und bereitgestellt werden.

Die Rückgewinnung bzw. Synthese des Sprachsignals aus den Parametern erfolgt in bekannter Weise dadurch, dass die zunächst in einem Decoder 15 decodierten Parameter einem Puls-Rausch-Generator 16, einem Verstärker 17 und einem Vokaltraktmodellfilter 18 zugeführt werden und das Ausgangssignal des Modellfilters 18 mittels eines D/A Wandlers 19 in analoge Form gebracht und dann nach der üblichen Filterung 20 durch ein Wiedergabegerät, z.B. einen Lautsprecher 21 hörbar gemacht wird. Der Puls-Rauschgenerator 16 erzeugt die durch den Verstärker 17 verstärkte Anregung x_n des Vokaltraktmodellfilters 18, und zwar im stimmlosen Falle ($p = 0$) weisses Rauschen und im stimmhaften Falle ($p \neq 0$) eine periodische Pulsfolge der durch die Pitchperiode p festgelegten Frequenz. Der Lautstärkeparameter G kontrolliert den Verstärkungsfaktor des Verstärkers 17, die Filterparameter (k_j) definieren die Uebertragungsfunktion des Klangbildungs- bzw. Vokaltraktmodellfilters 18.

Vorstehend wurde der allgemeine Aufbau und die Funktion der erfindungsgemässen Sprachverarbeitungsvorrichtung der einfacheren Verständlichkeit halber anhand diskreter Funktionsstufen erläutert. Es ist für den Fachmann jedoch selbstverständlich, dass sämtliche Funktionen bzw. Funktionsstufen zwischen dem analyseseitigen A/D-Wandler 3 und dem syntheseseitigen D/A-Wandler 19, in denen also digitale Signale verarbeitet werden, in der Praxis vorzugsweise durch einen entsprechend programmierten Computer oder einen Mikroprozessor oder dergleichen implementiert sind. Die softwarenmässige Realisierung der einzelnen Funktionssufen, wie z.B. der Parameterrechner, die diversen Digitalfilter, Autokorrelation etc. ist für den mit der

Datenverarbeitungstechnik vertrauten Fachmann Routine und in der Fachliteratur beschrieben (siehe z.B. IEEE Digital Signal Processing Committee: "Programs for Digital Signal Processing", IEEE Press Book 1980).

Für Echtzeit-Anwendungen sind insbesondere bei hohen Abtastraten und kurzen Sprachabschnitten wegen der dann grossen Anzahl von in kürzester Zeit zu bewältigenden Operationen extrem leistungsfähige Rechner erforderlich. Für solche Zwecke werden dann am besten Multi-Prozessor-Systeme mit einer geeigneten Aufgabenteilung eingesetzt. Ein Beispiel für ein solches System ist in Fig. 2 als Blockschema dargestellt.

Das dargestellte Multi-Prozessor-System umfasst im wesentlichen vier Funktionsblöcke, und zwar einen Hauptprozessor 50, zwei Nebenprozessoren 60 und 70 und eine Eingabe/Ausgabe-Einheit 80. Es implementiert sowohl Analyse als auch Synthese.

Die Eingabe/Ausgabe-Einheit 80 enthält die mit 81 bezeichneten Stufen zur analogen Signalverarbeitung, wie Verstärker, Filter und automatische Verstärkungsregelung, sowie den A/D-Wandler und den D/A-Wandler.

Der Hauptprozessor 50 führt die eigentliche Sprachanalyse bzw. -synthese durch, wozu die Bestimmung der Filterparameter und der Lautstärkeparameter (Parameterrechner 4), die Bestimmung von Energie und Nulldurchgängen des Sprachsignals (Stufen 13 und 12), die Stimmhaft-Stimmlos-Entscheidung (Stufe 11) und die Bestimmung der Pitchperiode (Stufe 9) bzw. syntheseseitig die Erzeugung des Ausgangssignals (Stufe 16), dessen Lautstärkevariation (Stufe 17) und dessen Filtrierung im Sprachmodellfilter (Filter 18) gehören.

Der Hauptprozessor 50 wird dabei vom Nebenprozessor 60 unterstützt, welcher die Zwischenspeicherung (Buffer 5), Inversfiltrierung (Stufe 6), gegebenenfalls die Tiefpassfiltrierung (Stufe 7) und die Autokorrelation (Stufe 8) durchführt.

Der Nebenprozessor 70 schliesslich befasst sich ausschliesslich mit der Codierung bzw. Decodierung der Sprachparameter sowie mit dem Datenverkehr mit z.B. einem Modem 90 oder dgl. via eine mit 71 bezeichnete Schnittstelle.

Im folgenden wird auf die Codierung der Sprachparameter eingegangen.

Die Datenrate in einem LPC-Vocoder-System wird bekanntlich bestimmt durch die sog. Frame-Rate, i.e. die Anzahl Sprachabschnitte pro Sekunde, die Anzahl der verwendeten Sprachparameter und die Anzahl Bit, die zur Codierung der Sprachparameter benötigt werden.

Bei den bisher bekannten Systemen werden gewöhnlich insgesamt etwa 10-14 Parameter verwendet, für deren Codierung pro Frame (Sprachabschnitt) in der Regel etwas über 50 bit benötigt werden. Bei einer auf 2,4 kbit/sec begrenzten Datenrate, wie sie bei Telefonnetzen üblich ist, führt dies zu einer maximalen Frame-Rate von rund 45. Wie die Praxis gezeigt hat, ist jedoch die Qualität der unter diesen Bedingungen verarbeiteten Sprache unbefriedigend.

Dieses durch die Begrenzung der Datenrate auf 2,4 kbit/sec bedingte Dilemma wird nun durch die vorliegende Erfindung durch eine bessere Ausnützung der Redundanzeigenschaften der menschlichen Sprache gelöst. Das grundlegende Prinzip der Erfindung besteht in der Ueberlegung, dass, wenn das Sprachsignal öfter analysiert wird, also die Frame-Rate erhöht wird, eine bessere Verfolgung der Instationaritäten des Sprachsignals möglich ist. Damit wird bei stationären Sprachabschnitten eine grössere Korrelation zwischen den Parametern der aufeinanderfolgenden Sprachabschnitte erreicht, welche wiederum zu einer effizienteren, d.h. bitsparenden Codierung ausgenützt werden kann, sodass die Gesamtdatenrate trotz erhöhter Frame-Rate nicht erhöht, die Sprachqualität hingegen erheblich verbessert wird. Diese spezielle, erfindungsgemässe Codierung der Sprachparameter ist nachstehend näher erläutert.

Der Grundgedanke der erfindungsgemässen Parameter-Codierung ist das sog. Blockcodierungsprinzip, d.h., die Sprachparameter werden nicht für jeden einzelnen Sprachabschnitt unabhängig voneinander codiert, sondern jeweils zwei oder drei Sprachabschnitte werden zu einem Block zusammengefasst und innerhalb dieses Blocks erfolgt die Codierung der Parameter aller zwei oder drei Sprachabschnitte nach einheitlichen Regeln und zwar derart, dass jeweils nur die Parameter des ersten Abschnitts in vollständiger Form codiert werden, während die Parameter des bzw. der übrigen Sprachabschnitte in differentieller Form codiert oder eventuell gänzlich weggelassen bzw. substituiert werden. Die Codierung innerhalb jedes Blocks wird ferner in Berücksichtigung der typischen Eigenschaften der menschlichen Sprache unterschiedlich vorgenommen je nach dem, ob es sich um einen stimmhaften oder einen stimmlosen Block handelt, wobei für den Stimmhaftigkeitscharakter des Blocks jeweils der erste Sprachabschnitt darin bestimmend ist.

Unter Codierung in vollständiger Form wird die übliche Codierung der Parameter verstanden, bei der z.B. für den Pitch-Parameter 6 bit, für den Lautstärkeparameter 5 bit und (bei einem z.B. zehnpoligen Filter) für die ersten vier Filterkoeffizienten je 5 bit, für die nächsten vier je 4 bit und für die beiden letzten 3 bzw. 2 bit reserviert werden. (Die abnehmende Bitanzahl für die höheren Filterkoeffizienten erklärt sich daraus, dass die gewöhnlich verwendeten Reflexionskoeffizienten im Betrag mit steigender Ordnungszahl abnehmen und im wesentlichen nur die Feinstruktur des Kurzzeitsprachspektrums mitbestimmen.)

Die erfindungsgemässe Codierung ist für die einzelnen Parameter-Typen (Filterkoeffizienten, Lautstärke, Pitch) unterschiedlich. Sie wird im folgenden am Beispiel von aus jeweils drei Sprachabschnitten bestehenden Blöcken erläutert.

Filterparameter (-koeffizienten):

Wenn der Block, d.h. der erste Sprachabschnitt darin stimmhaft

($p \neq 0$) ist, werden die Filterparameter des ersten Abschnitts in vollständiger Form codiert, die Filterparameter des zweiten und des dritten Abschnitts hingegen in differentieller Form, d.h., nur in Form ihrer Differenz gegenüber den entsprechenden Parametern des ersten bzw. gegebenenfalls auch des zweiten Abschnitts. Für die jeweilige Differenz wird z.B. um ein Bit weniger veranschlagt als für die vollständige Form, die Differenz eines 5-bit-Parameters wird also z.B. durch ein 4-bit-Wort dargestellt, u.s.f. Im Prinzip könnte so auch der letzte, nur 2 bit umfassende Parameter codiert werden, allerdings wäre dies bei nur 2 bit wenig sinnvoll. Der letzte Filterparameter des zweiten und des dritten Abschnitts wird daher entweder durch den des ersten Abschnitts ersetzt oder gleich Null gesetzt, was in beiden Fällen die Uebertragung erspart.

Gemäss einer ebenfalls bewährten Variante können die Filterkoeffizienten des zweiten Sprachabschnitts auch gleich mit denen des ersten Abschnitts angenommen werden und brauchen demzufolge überhaupt nicht codiert bzw. übertragen zu werden. Die dabei freiwerdenden Bits können dazu verwendet werden, die Differenz der Filterparameter des dritten Abschnitts zu denen des ersten Abschnitts mit grösserer Auflösung zu codieren.

Im stimmlosen Fall, d.h. also wenn der erste Sprachabschnitt des Blocks stimmlos ist ($p = 0$), erfolgt die Codierung in anderer Weise. Zwar werden die Filterparameter des ersten Abschnitts wieder voll, d.h. in vollständiger Form bzw. voller Bitlänge codiert, die Filterparameter der beiden übrigen Abschnitte werden jedoch nicht differentiell, sondern ebenso in vollständiger Form codiert. Damit dennoch eine Bitreduktion möglich ist, wird von der Tatsache Gebrauch gemacht, dass im stimmlosen Fall die höheren Filterkoeffizienten wenig zum Klangbild beitragen, und dementsprechend werden die höheren Filterkoeffizienten, z.B. ab dem siebenten, überhaupt nicht codiert bzw. übertragen. Syntheseseitig werden sie dann als Null interpretiert.

Lautstärkeparameter (Verstärkungsfaktor):

Bei diesem Parameter erfolgt die Codierung im stimmhaften und im stimmlosen Falle weitestgehend oder in einer Variante sogar vollständig gleich. Der Parameter des ersten und des dritten Abschnitts wird jeweils voll codiert, der des mittleren Abschnitts in Form seiner Differenz zu dem des ersten Abschnitts. Im stimmhaften Falle kann der Lautstärkeparameter des mittleren Sprachabschnitts auch gleich wie der des ersten Abschnitts angenommen werden und braucht demzufolge überhaupt nicht codiert bzw. übertragen zu werden. Der syntheseseseitige Decoder erzeugt dann diesen Parameter selbsttätig aus dem Parameter des ersten Sprachabschnitts.

Pitch-Parameter:

Die Codierung des Pitch-parameters erfolgt für stimmhafte und für stimmlose Blöcke gleich, und zwar so wie die der Filterkoeffizienten im stimmhaften Falle, d.h. für den ersten Sprachabschnitt (z.B. 7 bit) voll und für die beiden übrigen Abschnitte differentiell. Die Differenzen werden dabei vorzugsweise mit 3 bit dargestellt.

Eine Schwierigkeit ergibt sich jedoch, wenn innerhalb eines Blocks nicht alle Sprachabschnitte stimmlos oder stimmhaft sind, der Stimmhaftigkeitscharakter also wechselt. Zur Behebung dieser Schwierigkeit wird gemäss einem weiteren Gedanken der Erfindung ein solcher Wechsel durch ein spezielles Codewort angezeigt, indem die anstatt der dann den darstellbaren Differenzbereich in der Regel ohnehin übersteigende Differenz zum Pitch-Parameter des ersten Sprachabschnitts durch dieses Codewort ersetzt wird. Das Codewort hat dabei natürlich dasselbe Format wie die Pitch-Parameter-Differenzen.

Im Falle eines Wechsels von stimmhaft zu stimmlos, also $p \neq 0$ zu $p = 0$, ist klar, wie das Codewort syntheseseseitig decodiert werden muss - es braucht dann lediglich der betreffende Pitch-Parameter gleich Null gesetzt zu werden. Im umgekehrten Falle weiss man jedoch lediglich, dass ein Wechsel stattgefunden hat, aber nicht, wie

gross der betreffende Pitch-Parameter ist. Aus diesem Grunde wird syntheseseitig in diesem Falle als betreffender Pitch-Parameter ein laufender Mittelwert aus den Pitch-Parametern einer Anzahl, z.B. 2 bis 7 vorangegangener Sprachabschnitte verwendet.

Als weitere Sicherung gegen Fehlcodierungen und Fehlübertragungen und auch gegen Fehlberechnungen der Pitch-Parameter wird syntheseseitig vorzugsweise der decodierte Pitch-Parameter mit einem laufenden Mittelwert aus den Pitch-Parametern einer Anzahl, z.B. 2 bis 7 vorangegangener Sprachabschnitte verglichen und beim Ueber-schreiten einer vorgegebenen Maximalabweichung, beispielsweise etwa $\pm 30\%$ bis $\pm 60\%$, durch den laufenden Mittelwert ersetzt. Der "Ausreisser" geht dann natürlich auch nicht in die weitere Mittelwertbildung ein.

Bei Blöcken mit nur zwei Sprachabschnitten erfolgt die Codierung im Prinzip gleich wie bei den Blöcken mit drei Abschnitten. Sämtliche Parameter des ersten Abschnitts werden in vollständiger Form codiert. Die Filterparameter des zweiten Sprachabschnitts werden bei stimmhaften Blöcken entweder in differentieller Form codiert oder als gleich wie beim ersten Abschnitt angenommen und dementsprechend überhaupt nicht codiert. Bei stimmlosen Blöcken werden wiederum auch die Filterkoeffizienten des zweiten Sprachabschnitts in vollständiger Form codiert, dafür werden aber die höheren Koeffizienten weggelassen.

Der Pitch-Parameter des zweiten Sprachabschnitts wird im stimmhaften und im stimmlosen Fall wieder gleich codiert, und zwar in Form seiner Differenz zum Pitch-Parameter des ersten Abschnitts. Für den Fall eines Stimmhaft-Stimmlos-Wechsels innerhalb eines Blocks wird wiederum ein Code-Wort verwendet.

Der Lautstärkeparameter des zweiten Sprachabschnitts wird gleich codiert wie im Falle von Blöcken mit drei Abschnitten, also in differentieller Form oder gar nicht.

Vorstehend wurde bis auf einige Ausnahmen lediglich von der Codierung der Sprachparameter auf der Analyseseite des kompletten Sprachverarbeitungssystems gesprochen. Es versteht sich jedoch von selbst, dass auf der Syntheseseite eine entsprechende Decodierung der Parameter erfolgen muss, welche Decodierung auch die Erzeugung (vorvereinbarter Werte) der nicht codierten Parameter mit einschliesst.

Ferner versteht es sich, dass die Codierung und die Decodierung vorzugsweise per Software mittels des für die übrige Sprachverarbeitung ohnehin vorhandene Computersystems durchgeführt wird. Die Erstellung eines geeigneten Programms liegt im Bereich des Könnens des durchschnittlichen Fachmanns. Ein Beispiel für ein Flussschema eines solchen Programms, und zwar für den Fall von Blöcken mit je drei Sprachabschnitten, ist in den Fig. 3 und 4 dargestellt. Die Fluss-schemen sind aus sich heraus verständlich, es sei lediglich erwähnt, dass der Index i laufend die einzelnen Sprachabschnitte numeriert und zählt, während der Index $N = i \bmod 3$ die Nummer der Abschnitte innerhalb jedes einzelnen Blocks angibt. Die in Fig. 3 enthaltenen Codierungsvorschriften A_1 , A_2 und A_3 sowie B_1 , B_2 und B_3 sind in Fig. 4 detaillierter dargestellt und geben jeweils das Format (Bitzuteilungen) der zu codierenden Parameter an.

Die Programme für die Decodierung sind natürlich analog.

Patentansprüche

1. Redundanzverminderndes Sprachverarbeitungsverfahren nach der Methode der linearen Prädiktion, wobei analyseseitig das durch Abtastung des gegebenenfalls bandbegrenzten Analogsprachsignals gewonnene digitale Sprachsignal in Abschnitte eingeteilt wird und für jeden Sprachabschnitt die Parameter eines Sprachmodellfilters, ein Lautstärkeparameter und der Pitchparameter (Periode der Stimmbandgrundfrequenz) bestimmt und in codierter Form zur Uebertragung bereit gestellt bzw. übertragen werden, und wobei syntheseseitig die Filterparameter, der Lautstärkeparameter und der Pitchparameter decodiert werden und mit diesen Parametern eine im wesentlichen aus einem Anregungsgenerator und einem Sprachmodellfilter bestehende Synthesestufe zur Wiedergewinnung des Sprachsignals gesteuert wird, dadurch gekennzeichnet, dass die Codierung der Parameter blockweise über jeweils zwei oder drei aufeinanderfolgende Sprachabschnitte erfolgt, wobei die Parameter des ersten Sprachabschnitts jeweils in vollständiger Form und wenigstens ein Teil der Parameter der übrigen Abschnitte in differentieller Form codiert oder weggelassen werden.
2. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass die Codierung der Parameter in unterschiedlicher Weise vorgenommen wird in Abhängigkeit davon, ob der erste Sprachabschnitt eines Sprachabschnittblocks stimmhaft oder stimmlos ist.
3. Verfahren nach Anspruch 2, dadurch gekennzeichnet, dass im Falle von je drei Sprachabschnitte umfassenden Blöcken bei stimmhaftem ersten Sprachabschnitt die Filter- und der Pitchparameter des ersten Abschnitts in vollständiger Form und die Filter- und der Pitchparameter der beiden übrigen Abschnitte in Form ihrer Differenzen zu den entsprechenden Parametern des ersten Abschnitts bzw. zweiten Abschnitts codiert werden, und dass bei stimmlosem ersten Sprachabschnitt Filterparameter höherer Ordnungen weggelassen und die verbleibenden Filterparameter aller drei Sprachabschnitte in vollständiger Form und die

Pitchparameter gleich wie im stimmhaften Fall codiert werden.

4. Verfahren nach Anspruch 2, dadurch gekennzeichnet, dass im Falle von je drei Sprachabschnitte umfassenden Blöcken bei stimmhaftem ersten Sprachabschnitt die Filter- und die Pitchparameter des ersten Abschnitts in vollständiger Form, die Filterparameter des mittleren Sprachabschnitts überhaupt nicht und der Pitchparameter dieses Abschnitts in Form seiner Differenz zum Pitchparameter des ersten Abschnitts, und die Filter- und der Pitchparameter des letzten Abschnitts in Form ihrer Differenzen zu den entsprechenden Parametern des ersten Abschnitts codiert werden, und dass bei stimmlosem ersten Sprachabschnitt Filterparameter höherer Ordnungen weggelassen und die verbleibenden Filterparameter aller drei Sprachabschnitte in vollständiger Form und die Pitchparameter gleich wie im stimmhaften Fall codiert werden.

5. Verfahren nach Anspruch 2, dadurch gekennzeichnet, dass im Falle von je zwei Sprachabschnitte umfassenden Blöcken bei stimmhaftem ersten Sprachabschnitt die Filter- und der Pitchparameter des ersten Sprachabschnitts in vollständiger Form und die Filterparameter des zweiten Abschnitts überhaupt nicht oder in Form ihrer Differenzen zu den jeweiligen Parametern des ersten Abschnitts und der Pitchparameter des zweiten Abschnitts in Form seiner Differenz zum Pitchparameter des ersten Abschnitts codiert werden, und dass bei stimmlosem ersten Sprachabschnitt Filterparameter höherer Ordnungen weggelassen und die verbleibenden Filterparameter beider Sprachabschnitte in vollständiger Form und die Pitchparameter gleich wie im stimmhaften Fall codiert werden.

6. Verfahren nach Anspruch 3 oder 4, dadurch gekennzeichnet, dass bei stimmhaftem ersten Sprachabschnitt die Lautstärkeparameter des ersten und des letzten Sprachabschnitts in vollständiger Form und der des mittleren Abschnitts überhaupt nicht oder in Form seiner Differenz zum Lautstärkeparameter des ersten Abschnitts codiert werden, und dass bei stimmlosem ersten Sprachabschnitt die Lautstärkeparameter des ersten und des letzten Sprachabschnitts in vollständiger Form und der

des mittleren Abschnitts in Form seiner Differenz zum Lautstärkeparameter des ersten Abschnitts codiert werden.

7. Verfahren nach Anspruch 5, dadurch gekennzeichnet, dass bei stimmhaftem ersten Sprachabschnitt der Lautstärkeparameter des ersten Sprachabschnitts in vollständiger Form und der des zweiten Sprachabschnitts überhaupt nicht oder in Form seiner Differenz zum Lautstärkeparameter des ersten Sprachabschnitts codiert werden, und dass bei stimmlosem ersten Sprachabschnitt der Lautstärkeparameter des ersten Abschnitts in vollständiger Form und der des zweiten Abschnitts in Form seiner Differenz zum Lautstärkeparameter des ersten Sprachabschnitts codiert werden.

8. Verfahren nach einem der Ansprüche 3 bis 7, dadurch gekennzeichnet, dass bei einem Wechsel von stimmhaft zu stimmlos oder umgekehrt innerhalb eines Sprachabschnittblocks der Pitchparameter des betreffenden Abschnitts durch ein spezielles Codewort ersetzt wird.

9. Verfahren nach Anspruch 8, dadurch gekennzeichnet, dass syntheseseitig bei Auftreten des Codeworts und wenn der vorhergehende Sprachabschnitt stimmlos war, als entsprechender Pitchparameter ein laufender Mittelwert aus den Pitchparametern einer Anzahl vorangegangener Sprachabschnitte verwendet wird.

10. Verfahren nach einem der vorangehenden Ansprüche, dadurch gekennzeichnet, dass syntheseseitig der decodierte Pitchparameter mit einem laufenden Mittelwert aus den Pitchparametern einer Anzahl vorangegangener Sprachabschnitte verglichen und bei Ueberschreiten einer vorgegebenen Maximalabweichung durch den laufenden Mittelwert ersetzt wird.

11. Verfahren nach einem der vorangehenden Ansprüche, dadurch gekennzeichnet, dass die Länge der einzelnen Sprachabschnitte, für welche jeweils die Sprachparameter ermittelt werden, höchstens etwa 30 msec, vorzugsweise etwa 20 msec beträgt.
12. Verfahren nach einem der vorangehenden Ansprüche, dadurch gekennzeichnet, dass die Anzahl Sprachabschnitte pro Sekunde mindestens etwa 55, vorzugsweise mindestens 60 beträgt.
13. Vorrichtung zur Durchführung des Verfahrens gemäss einem der vorangehenden Ansprüche, mit einem Signalaufbereitungsteil, der das analoge Sprachsignal taktweise abtastet und die dabei erhaltenen Abtastwerte digitalisiert, mit einem Analyseteil, welcher das digitalisierte Sprachsignal abschnittsweise analysiert und einen Parameterrechner, eine Pitchentscheidungsstufe und eine Pitchberechnungsstufe enthält, und mit einer Codierstufe, welche die vom Analyseteil ermittelten Sprachparameter codiert, dadurch gekennzeichnet, dass der Analyseteil und die Codierstufe durch einen Rechner oder ein Rechnersystem gebildet sind, welches zur Durchführung der in einem oder mehreren der vorangehenden Ansprüche beschriebenen Verfahrensschritte programmiert ist.
14. Vorrichtung nach Anspruch 13, dadurch gekennzeichnet, dass der Analyseteil ein Multiprozessorsystem mit einem Hauptprozessor (50) und zwei Nebenprozessoren (60, 70) ist, wobei ein Nebenprozessor (60) das Sprachsignal zwischenspeichert, aus dem zwischengespeicherten Sprachsignal durch eine Inversfiltrierung das Prädikationsfehlersignal erzeugt und aus diesem, gegebenenfalls nach einer Tiefpassfiltrierung, die normierte Autokorrelationsfunktion bildet, wobei der Hauptprozessor (50) die eigentliche Analyse des Sprachsignals durchführt, und wobei der andere Nebenprozessor (70) für die Codierung der vom Hauptprozessor in Verbindung mit dem ersten Nebenprozessor ermittelten Sprachparameter verantwortlich ist.

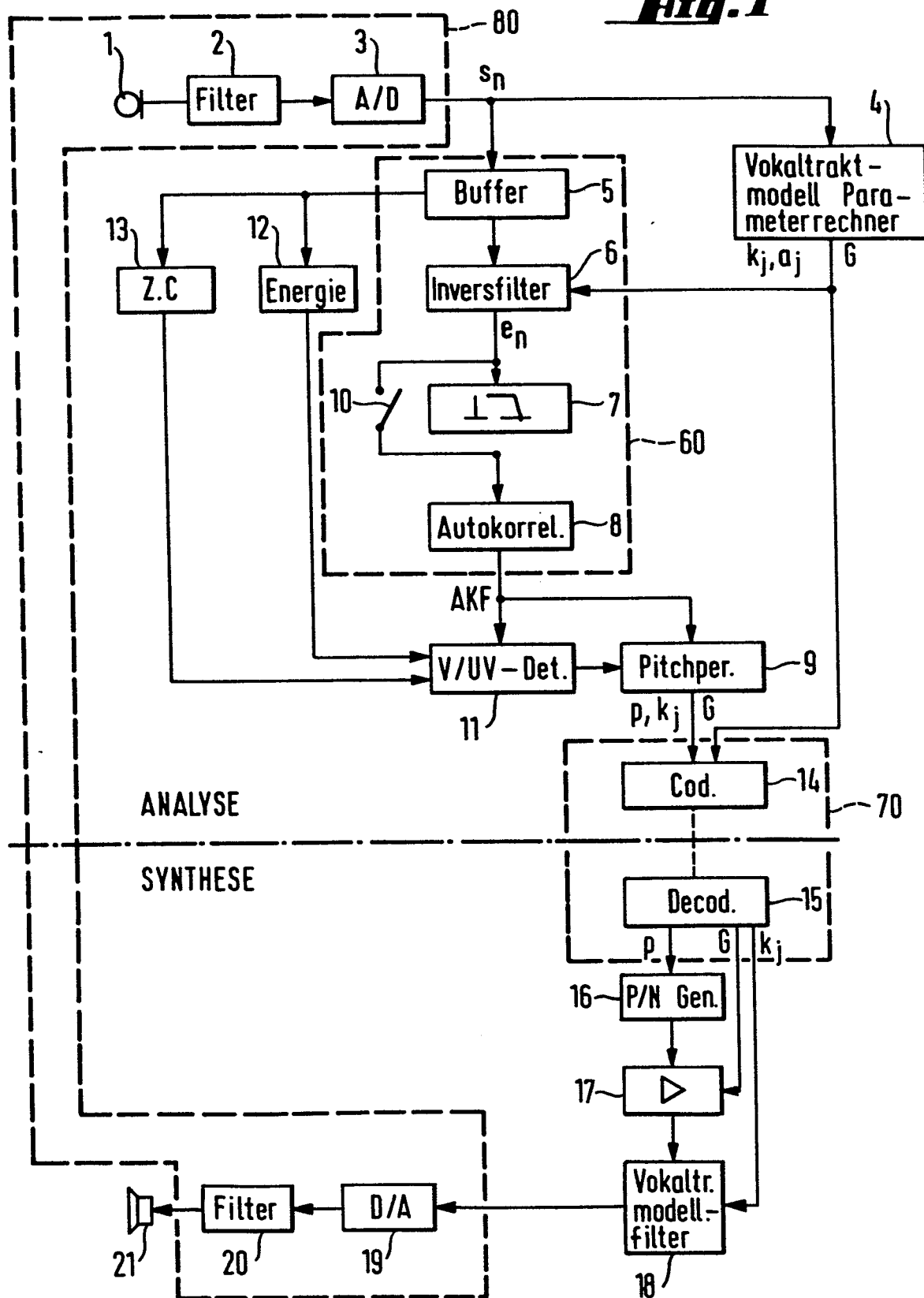
Fig. 1

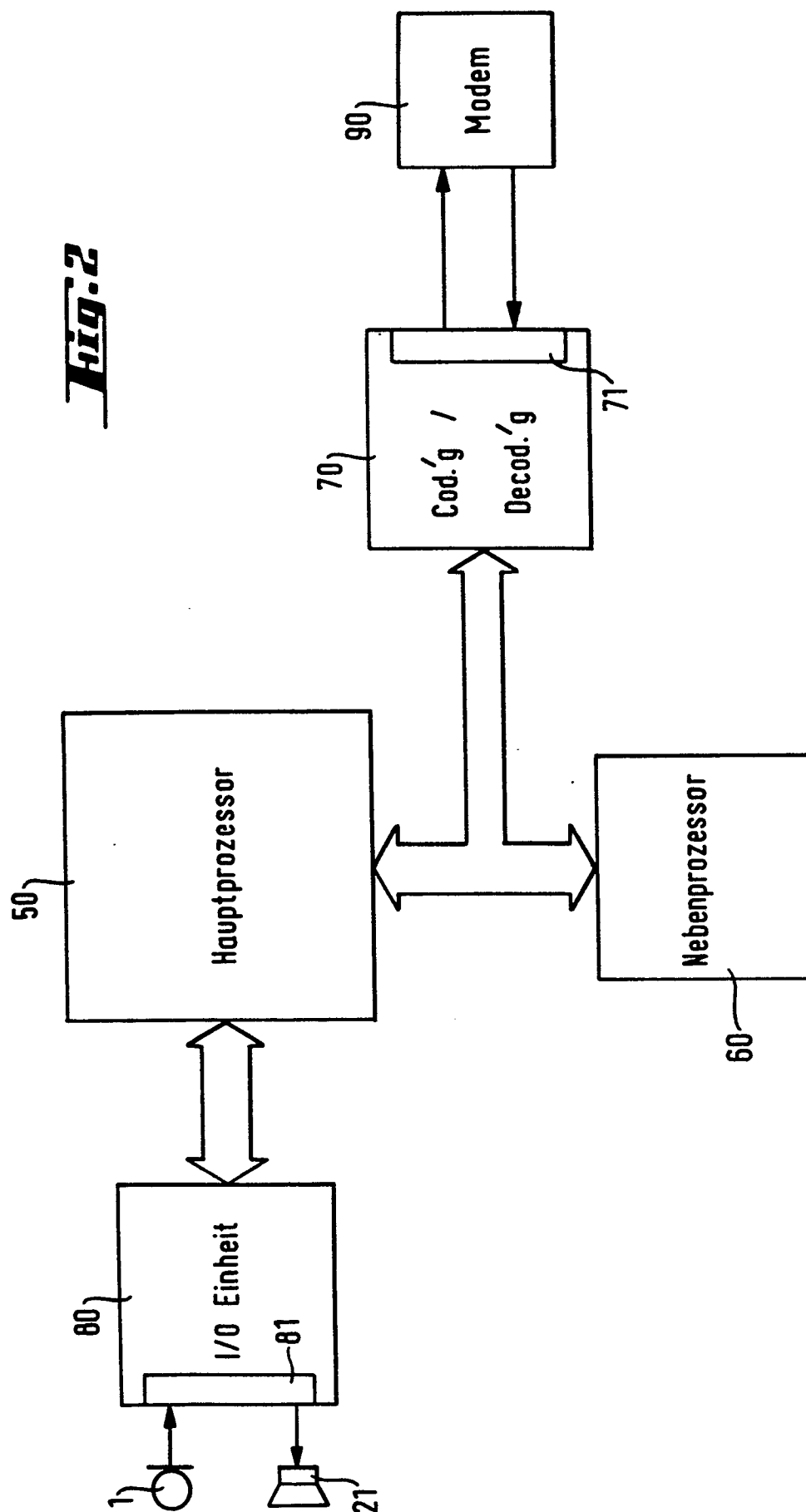
Fig. 2

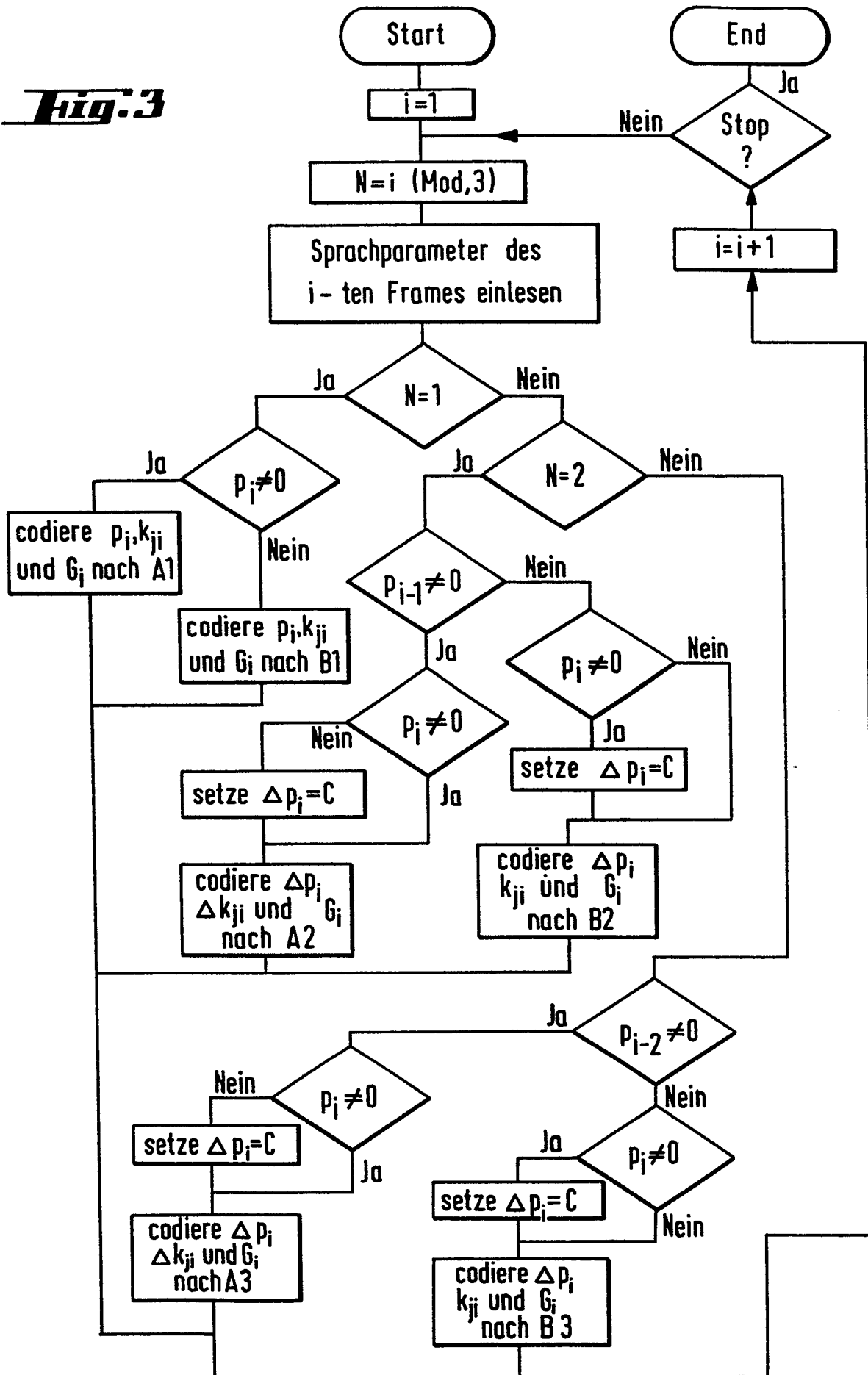
Fig. 3

Fig. 4

Bitzuteilungs- vorschrift	p_i	k_{1i}	k_{2i}	k_{3i}	k_{4i}	k_{5i}	k_{6i}	k_{7i}	k_{8i}	k_{9i}	k_{10i}	g_i
A 1	7	7	6	5	5	5	4	4	4	3	2	5
B 1	7	6	5	5	4	4	3	-	-	-	-	5

Bitzuteilungs- vorschrift	Δp_i	Δk_{1i}	Δk_{2i}	Δk_{3i}	Δk_{4i}	Δk_{5i}	Δk_{6i}	Δk_{7i}	Δk_{8i}	Δk_{9i}	Δk_{10i}	g_i
A 2	3	-	-	-	-	-	-	-	-	-	-	4
A 3	3	6	5	4	4	4	3	3	3	2	2	5

Mit $\Delta p_i = p_i - p_{i-1}$ und $\Delta k_{ji} = k_{ji} - k_{ji-2}$ und C = Codewort = -4

Bitzuteilungs- vorschrift	Δp_i	k_{1i}	k_{2i}	k_{3i}	k_{4i}	k_{5i}	k_{6i}	k_{7i}	k_{8i}	k_{9i}	k_{10i}	g_i
B 2	3	6	5	5	4	4	3	-	-	-	-	4
B 3	3	6	5	5	4	4	3	-	-	-	-	5



Europäisches
Patentamt

EUROPÄISCHER RECHERCHENBERICHT

0076234
Nummer der Anmeldung

EP 82 81 0391

EINSCHLÄGIGE DOKUMENTE			
Kategorie	Kennzeichnung des Dokuments mit Angabe, soweit erforderlich, der maßgeblichen Teile	Betrifft Anspruch	KLASSIFIKATION DER ANMELDUNG (Int. Cl. ³)
A	--- IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, Band ASSP-25, Nr. 4, August 1977, Seiten 322-330, New York, USA S. CHANDRA et al.: "Linear prediction with a variable analysis frame size" * Zusammenfassung *	1	G 10 L 1/08
A	--- IEEE TRANSACTIONS ON COMMUNICATIONS, Band COM-23, Nr. 12, Dezember 1975, Seiten 1466-1474, New York, USA C.K. UN et al.: "The residual-excited linear prediction vocoder with transmission rate below 9.6 kbits/s" * Zusammenfassung *	1	
A	--- ICASSP 80 PROCEEDINGS-IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH & SIGNAL PROCESSING, Band 1, 9.-11. April 1980, Denver IEEE, New York, USA S. MAITRA: "Improvements in the classical model for better speech quality" * Abbildung 3 *	1, 2	RECHERCHIERTE SACHGEBIETE (Int. Cl. ³) G 10 L 1/08
A	--- US-A-3 439 753 (F.W. MOUNTS) * Spalte 2, Zeilen 38-62 * --- -/-	1	
Der vorliegende Recherchenbericht wurde für alle Patentansprüche erstellt.			
Recherchenort DEN HAAG		Abschlußdatum der Recherche 17-12-1982	Prüfer ARMSPACH J.F.A.M.
KATEGORIE DER GENANNTEN DOKUMENTEN X : von besonderer Bedeutung allein betrachtet Y : von besonderer Bedeutung in Verbindung mit einer anderen Veröffentlichung derselben Kategorie A : technologischer Hintergrund O : nichtschriftliche Offenbarung P : Zwischenliteratur T : der Erfindung zugrunde liegende Theorien oder Grundsätze E : älteres Patentedokument, das jedoch erst am oder nach dem Anmeldedatum veröffentlicht worden ist D : in der Anmeldung angeführtes Dokument L : aus andern Gründen angeführtes Dokument & : Mitglied der gleichen Patentfamilie, übereinstimmendes Dokument			



Europäisches
Patentamt

EUROPÄISCHER RECHERCHENBERICHT

0076234
Nummer der Anmeldung

EP 82 81 0391

EINSCHLÄGIGE DOKUMENTE			Seite 2
Kategorie	Kennzeichnung des Dokuments mit Angabe, soweit erforderlich, der maßgeblichen Teile	Betrifft Anspruch	KLASSIFIKATION DER ANMELDUNG (Int. Cl. ³)
A	IEEE TRANSACTIONS ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, Band ASSP-25, Nr. 5, Oktober 1977, Seiten 379-387, New York, USA E.M. HOFSTETTER et al.: "Micro-processor realization of a linear predictive vocoder" * Abbildung 1 * -----	13, 14	
			RECHERCHIERTE SACHGEBIETE (Int. Cl. ³)
Der vorliegende Recherchenbericht wurde für alle Patentansprüche erstellt.			
Recherchenort DEN HAAG		Abschlußdatum der Recherche 17-12-1982	Prüfer ARMSPACH J.F.A.M.
KATEGORIE DER GENANNTEN DOKUMENTEN			
X : von besonderer Bedeutung allein betrachtet		E : älteres Patentdokument, das jedoch erst am oder nach dem Anmeldedatum veröffentlicht worden ist	
Y : von besonderer Bedeutung in Verbindung mit einer anderen Veröffentlichung derselben Kategorie		D : in der Anmeldung angeführtes Dokument	
A : technologischer Hintergrund		L : aus andern Gründen angeführtes Dokument	
O : mündliche Offenbarung			
P : Zwischenliteratur		& : Mitglied der gleichen Patentfamilie, übereinstimmendes Dokument	
T : der Erfindung zugrunde liegende Theorien oder Grundsätze			