



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 600 27 140 T2** 2007.01.11

(12) **Übersetzung der europäischen Patentschrift**

(97) **EP 1 159 738 B1**

(51) Int Cl.⁸: **G10L 19/00** (2006.01)

(21) Deutsches Aktenzeichen: **600 27 140.4**

(86) PCT-Aktenzeichen: **PCT/US00/02900**

(96) Europäisches Aktenzeichen: **00 914 511.1**

(87) PCT-Veröffentlichungs-Nr.: **WO 2000/046795**

(86) PCT-Anmeldetag: **04.02.2000**

(87) Veröffentlichungstag

der PCT-Anmeldung: **10.08.2000**

(97) Erstveröffentlichung durch das EPA: **05.12.2001**

(97) Veröffentlichungstag

der Patenterteilung beim EPA: **05.04.2006**

(47) Veröffentlichungstag im Patentblatt: **11.01.2007**

(30) Unionspriorität:

246605 08.02.1999 US

(84) Benannte Vertragsstaaten:

**AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT,
LI, LU, MC, NL, PT, SE**

(73) Patentinhaber:

Qualcomm, Inc., San Diego, Calif., US

(72) Erfinder:

CHANG, Chienchung, California 92067-3874, US

(74) Vertreter:

**WAGNER & GEYER Partnerschaft Patent- und
Rechtsanwälte, 80538 München**

(54) Bezeichnung: **SPRACHSYNTHETISIERER AUF DER BASIS VON SPRACHKODIERUNG MIT VERÄNDERLICHER
BIT-RATE**

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

Die Übersetzung ist gemäß Artikel II § 3 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patent- und Markenamt inhaltlich nicht geprüft.

Beschreibung**HINTERGRUND DER ERFINDUNG****I. Gebiet der Erfindung**

[0001] Die vorliegende Erfindung bezieht sich auf Sprachsynthese. Im Speziellen bezieht sich die vorliegende Erfindung auf die Synthese von Sprache codiert durch einen Vocoder mit variabler Rate. Die Erfindung bezieht sich weiterhin auf die Verwendung von Sprachsynthese mit Drahtloskommunikationsgeräten.

II. Beschreibung der verwandten Technik

[0002] Elektronische Sprachsynthese ist in einer Anzahl von Applikationen nützlich. Mehr und mehr sehen Computer und anderes elektronisches Zubehör die Möglichkeit von gesprochenen Benutzerhinweisen als ein Benutzerinterface vor. Sprache kann zum Beispiel zum Lesen von elektronischen Postnachrichten, zum Erzeugen von gesprochenen Benutzerhinweisen in einem Sprach-Antwort-System oder zum Vorsehen von Anweisungen für einen Fahrer in einem Fahrzeug angewendet werden.

[0003] Es gibt zwei generelle Arten von Sprachsynthesizern oder Techniken, die dazu benutzt werden, um Sprache zu erzeugen. Der erste Typ wird als ein Text-zu-Sprache-Sprachsynthesizer (TTS = text-to-speech) bezeichnet und basiert auf Grammatik. Ein TTS-basiertes System konvertiert gewöhnlichen Text in verständliche und natürlich klingende Sprache. Dies ist nützlich für Applikationen, die eine automatische Konvertierung von beliebigem Eingabetext in verständliche und natürlich klingende Sprachausgabe benötigt. Es ist speziell nützlich, wo großes Vokabular und/oder dynamisch wechselnde Daten involviert sind. Das TTS-System ist in Applikationen, wie zum Beispiel das Vorsehen von automatischen Sprachhinweisen und -eingabeaufforderungen, Korrekturlesen, Telefonzugriff zu Datenbanken und Konvertierung von elektronischer Post zu Sprachpost oder Audioausgabe nützlich. Weil TTS flexibel und leistungsfähig ist, bietet es Nutzen in vielen Applikationen. Jedoch kann die Implementierung eines TTS-Systems enormen Speicher und Verarbeitungsleistungsressourcen erfordern. Es kann ebenso einen Maschinenton enthalten, wenn der Synthesizer die menschliche Sprachintonation nicht nahe genug simuliert. Demgemäß ist TTS keine praktische Wahl für Applikationen mit begrenztem Speicher und Verarbeitungsressourcen, wie zum Beispiel in kleinen portablen Drahtlosgeräten, entfernt platzierten Kommunikationsgeräten oder Computern und Ähnlichem. Ein zweiter Typ Sprachsynthesizer basiert auf einem Sprachcodierer (Vocoder). Ein Vocoder komprimiert stimmhafte Sprache oder Audiosignale durch das Extrahieren von Parametern, die sich auf ein Modell der

menschlichen Spracherzeugung beziehen. Vocoder sind entwickelt worden, um Eingabesprache, die digital mit einer Rate von 64 Kilobits pro Sekunde (kbps) konvertiert wurde, runter auf 13 kbps, 8 kbps oder sogar niedrigere Raten zu komprimieren. Ein Vocoder, der auf einem Sprachsynthesizer basiert, erzeugt gewisse Parameter von oder für die Sprache, die synthetisiert werden soll. Die Parameter werden in einer Art von Speicher gespeichert, vorzugsweise eine Flash-Art und werden auf die Sprachsynthese codiert. Da die Parameter von allen Wörtern, die synthetisiert werden sollen, in einem Speicher gespeichert werden müssen, sind vocoderbasierte Sprachsynthesizer besser geeignet für Applikationen, die kein großes Vokabular benötigen. Sie sind speziell geeignet für Systeme mit begrenztem Speicher und Verarbeitungsressourcen.

[0004] EP-A-0 762 711 beschreibt einen Sprachspeicher in einem tragbaren Zellulartelefon. Sprache kann über ein Mikrofon eingegeben werden und komprimiert und gespeichert für nachfolgendes Wiedergeben, oder Senden über einen Kanal. Die Komprimierung kann unter der Verwendung von existierender Schaltung bewirkt werden.

[0005] Für vocoderbasierte Sprachsynthesizer gibt es einen Bedarf, die Speicherverwendung zu optimieren, während akzeptable Sprachqualität aufrecht erhalten wird. Für einige Applikationen kann es wünschenswert sein, die Größe des Vokabulars für eine gegebene Größe des Speichers zu maximieren. Weiterhin kann es ebenso wünschenswert sein, Signalverarbeitungsressourcen, die schon innerhalb eines vorhandenen Kommunikationssystemdesigns verfügbar sind, für das Ausführen von Sprachsynthese zu benutzen. Ein Sprachsynthesizer, der diese und andere synthesizer, der diese und andere Charakteristiken besitzt, wird durch die vorliegende Erfindung in der Art, wie nachstehend beschrieben, vorgesehen.

ZUSAMMENFASSUNG DER ERFINDUNG

[0006] Die vorliegende Erfindung ist eine Vorrichtung und ein Verfahren zur Sprachsynthese, wie dargestellt in den Ansprüchen 1 und 11, entsprechend, und basiert auf einem Sprachcodieren mit variabler Rate. Die Sprache, die synthetisiert werden soll, wird durch einen Vocoder mit variabler Rate codiert. Ein Vocoder mit variabler Rate codiert einen Rahmen von Sprache bei einer aus einem Satz von vorbestimmten Raten basierend auf die Sprachaktivität, die innerhalb des Sprachrahmens stattfindet. In einem Ausführungsbeispiel ist der Vocoder mit variabler Rate ein Code angeregter Linearprädiktionscodierer bzw. Code-Excited-Linear-Prediction-(CELP)-Codierer mit vier Bitraten. Somit wird das Eingabesprachsignal in Sprachparameter mit einer der vier Raten unter Verwendung eines CELP-Codierungsschemas für die

gewählte Rate codiert. Die Sprachparameter werden generell einem Decodierer geliefert, der ein Decodierungsschema mit variabler Rate entsprechend der variablen Rate des angewendeten Codierungsschemas durchführt. Der Decoder erzeugt Sprachsamples, die zu einem Codierer/Decodierer oder Codec zur Digital-zu-Analog-Konvertierung geliefert werden. Das resultierende analoge Signal, das von dem Codec erzeugt wurde, wird anschließend über einen Lautsprecher oder ein anderes bekanntes Audioausgabegerät als synthetisierte Sprache abgestrahlt.

[0007] Der Sprachsynthesizer der vorliegenden Erfindung ist speziell geeignet für die Verwendung in drahtlosen Kommunikationssystemen, in denen Sprachcodieren mit variabler Rate bereits implementiert ist. In diesen Systemen können die existierenden Sprachcodierungsressourcen für die Sprachsynthese verwendet werden. Als Alternative können DSP-Elemente, die bereits vorliegen oder leicht eingearbeitet werden können, in Verbindung mit einer kleinen Menge Speicher benutzt werden, um die Sprachsynthesizerfunktion vorzusehen. Zusätzlich ist ein Sprachsynthesizer basierend auf das Sprachcodieren mit variabler Rate dazu in der Lage, gute Sprachqualität ohne eine große Menge an Speicher zu benötigen, vorzusehen. Der Grad der Komprimierung, der von einem Vocoder mit variabler Rate vorgesehen wird, macht ihn geeignet für Applikationen mit begrenztem Speicher.

KURZE BESCHREIBUNG DER ZEICHNUNGEN

[0008] Die Merkmale, Ziele und Vorteile der vorliegenden Erfindung werden ausgehend von der detaillierten Beschreibung, die nachstehend dargelegt ist, noch deutlicher werden, wenn sie in Verbindung mit den Zeichnungen gebracht werden, in denen gleiche Bezugszeichen durchgehend Entsprechendes identifizieren, und wobei:

[0009] [Fig. 1](#) ein Blockdiagramm eines Vocoders mit variabler Rate ist; und

[0010] [Fig. 2](#) ein Blockdiagramm des Sprachsynthesizers der vorliegenden Erfindung ist.

DETAILLIERTE BESCHREIBUNG DER BEVORZUGTEN AUSFÜHRUNGSBEISPIELE

[0011] Die vorliegende Erfindung sieht eine Vorrichtung und ein Verfahren zum Synthetisieren von Sprache vor, die sehr nützlich ist, wenn sie mit drahtlosem Kommunikationszubehör benutzt werden. Die Erfindung kann Vorteil aus existierenden Signalverarbeitungsressourcen in Drahtloskommunikationsanlagen bzw. -zubehör oder einem Minimum von zusätzlicher Hardware ziehen, um Sprache auf eine Art und Weise zu synthetisieren, die hoch qualitative Sprache vorsieht und eine kleine Speichergröße benötigt.

[0012] Die vorliegende Erfindung ist sehr nützlich, wenn sie in Verbindung mit einer Vielfalt von bekannten Kommunikationsgeräten oder Systemen verwendet wird, und es ist nachstehend beschrieben, in Bezug auf ein CDMA-Drahtloskommunikationssystem. Zusätzlich wird in Erwägung gezogen, dass sie besonders gut für spezifische Applikationen, wie z.B. Freisprechautoanlagen, die dazu verwendet werden, um Drahtlosgeräte in Fahrzeugen aufzunehmen und zu operieren bzw. betreiben geeignet ist. Jedoch wird der Fachmann gut verstehen, dass dies keine Begrenzung der vorliegenden Erfindung ist, und dass sie mit anderen Typen von Kommunikationsgeräten einschließlich von Systemen, des drahtgebunden, Drahtleitungs- oder des optischen Kabeltyps kommunizieren und, solchen, die andere Signalmodulationstechniken benutzen, benutzt werden kann.

[0013] Ein beispielhaftes Drahtloskommunikationssystem macht Gebrauch von Codemultiplex-Vielfachzugriff-Modulationstechniken (CDMA = code division multiple access). Obwohl andere Techniken, wie zum Beispiel Zeitmultiplex-Vielfachzugriff (TDMA = time division multiple access), Frequenzmultiplex-Vielfachzugriff (FDMA = frequency division multiple access) und Amplitudenmodulationsschemata (AM = amplitude modulation), wie zum Beispiel Amplitude Companded Single Sideband (ACSSB) bekannt sind, hat CDMA signifikante Vorteile über diese anderen Techniken. Die Verwendung von CDMA-Techniken in einem Vielfachzugriffskommunikationssystem ist offenbart im U.S.-Patent Nr. 4,901,307 mit dem Titel "Spread Spectrum Multiple Access Communication System Using Satellite Or Terrestrial Repeaters", dem Rechtsnachfolger der vorliegenden Erfindung zugeordnet.

[0014] Ein Sprachsynthesizer kann in Drahtloskommunikationsgeräten oder -zubehör bzw. -anlagen aus verschiedenen von Gründen implementiert werden. Sprachsynthese kann zum Beispiel Teil eines Spracherkennungssystems in einem Drahtlostelefon oder einem "Freisprech"-Autokit bzw. Autofreisprechanlage sein, das dazu verwendet wird, um den Betrieb in einem Fahrzeug zu unterstützen. Ein Sprachsynthesizer kann Informationen in hörbarer Form vorsehen, wenn ein Gerätebenutzer oder -betreiber eine Ausgabeanzeige oder -anzeigen auf dem Gerät visuell nicht betrachten kann. Informationen können zum Beispiel vorgesehen werden, um den Gerätebetrieb oder die Geräteausgabe zu ermöglichen, wenn ein Fahrzeugfahrer oder Maschinenbetreiber nicht nahe genug sicher auf das Kommunikationsgerät sehen kann. Der Sprachsynthesizer würde ebenso die Freisprechoperation von Geräten durch Vorsehen von Spracheingabeaufforderungen für Operationen, die durchgeführt werden sollen, erlauben. Der Sprachsynthesizer kann zum Beispiel nach dem Namen der Person, die angerufen werden soll, fragen, was dem Gerät erlaubt, automatisch eine Telefonnummer zu

wählen, oder nach einem Befehl, der implementiert werden soll, zu fragen, wie zum Beispiel das Wählen, Speichern, das Öffnen von Post, das Beenden eines Anrufversuchs oder das Ausschalten.

[0015] In einem Ausführungsbeispiel macht der Sprachsynthesizer der vorliegenden Erfindung Gebrauch von der Vocoderschaltung, die bereits in einer Anzahl von Drahtlosgeräten vorliegt, wie zum Beispiel Drähtlostelefonen und anderen Produkten, die von Kommunikationsdienstteilnehmern benutzt werden, um stimmhafte Sprache zu erzeugen. Der Sprachsynthesizer basiert speziell auf einem Vocoder mit variabler Rate. Ein Vocoder mit variabler Rate benutzt Sprachaktivität, um seine momentane Datenrate zu variieren. Während aktiver Sprache benutzt der Vocodercodierer eine große Anzahl von Bits, um die Sprachsamples zu codieren. Während Perioden mit Stille, benutzt der Vocodercodierer wenig oder weniger Bits, um das Hintergrundgeräusch zu codieren. Beispielhafte Ausführungsbeispiele von Vocodern mit variabler Rate sind im US-Patent Nr. 5,414,796 und 5,657,420 beschrieben, mit dem Titel "Variable Rate Vocoder", und sind dem Rechtsnachfolger der vorliegenden Erfindung zugeordnet.

[0016] Vocoder mit variabler Rate werden gewöhnlich in Kommunikationssystemen des CDMA-Typs benutzt, um Systemkapazität durch Vermindern der Anzahl der Bits, die gewöhnlich von jedem Kommunikationssignal benutzt werden, zu erhöhen. Ein Vocoder mit variabler Rate kann zum Beispiel in dem CDMA-Kommunikationssystem der Patent-Nr. 4,901,307, wie oben diskutiert, implementiert werden. In einem CDMA-Kommunikationssystem kommunizieren verschiedene Benutzer unter Verwendung der gleichen Bandbreite, benutzen aber verschiedene Codekanäle. Ein Vocoder mit variabler Rate zieht in einem CDMA-Kommunikationssystem Vorteil aus der Tatsache, dass ein Benutzer nur ungefähr 40% der Zeit auf jedem gegebenen Kanal aktiv spricht. Durch das Senden von weniger Bits, wenn ein Benutzer still ist, erlaubt es der Vocoder mit variabler Rate mehr Benutzern, sich die gleiche Bandbreite zu teilen.

[0017] Ein schematisches Blockdiagramm eines typischen Vocoders mit variabler Rate ist in [Fig. 1](#) gezeigt und ist generell durch **100** angezeigt. Der Vocoder, der in [Fig. 1](#) gezeigt ist, benutzt vier verschiedene Datenraten, obwohl es angemerkt ist, dass eine unterschiedliche Anzahl von Datenraten stattdessen angewendet werden kann, so wie es auf dem Fachgebiet bekannt ist. In dem Satz von vier Raten, wenn die Spitzenrate 13,2 kbps ist, dann entspricht die Vollrate 13,2 kbps, die Halbrate entspricht näherungsweise 6,2 kbps, die Viertelrate entspricht annäherungsweise 2,7 kbps und die Achtelrate entspricht näherungsweise 1,0 kbps. Es sei angemerkt, dass die tatsächliche Bitrate für Raten anders als die Voll-

rate wegen der Verwendung von Overhead-Bits angenähert ist, so wie es auf dem Fachgebiet bekannt ist.

[0018] Immer noch bezogen auf die [Fig. 1](#) kann erkannt werden, dass der Vocoder mit variabler Rate **100** einen Codierer **102** und einen Decodierer **104** beinhaltet. Der Codierer **102** empfängt Sprachsamples für Rahmen mit Sprachdaten als eine Eingabe, zum Beispiel als 8-Bit-PCM-Samples mit einer 64-kbps-Datenrate, in entweder einem mu-law- oder a-law-Format. Codierer **102** codiert diese Sprachsamples in Sprachparameter mit einer von vier Datenraten, abhängig von der Sprachaktivität. Diese Eingabesprachsamples werden auch zum Ratenbestimmungselement **106** geliefert.

[0019] Ratenbestimmungselement **106** kann jegliche von einer Anzahl von Ratenentscheidungsalgorithmen implementieren. In einem Ausführungsbeispiel werden die Energieschwellen relativ zum Hintergrundgeräuschenergiepegel benutzt, um die Sprachaktivität zu bestimmen und dadurch die Rate mit der die Eingabesamples codiert werden sollen. Wenn die Energie des aktuellen Sprachsamplers weit über der Hintergrundgeräuschenergie liegt, dann wird das Ratenbestimmungselement **106** bestimmen, dass der Rahmen mit der Vollrate codiert werden soll. Wenn die Energie des aktuellen Rahmens nahe der Hintergrundgeräuschenergie liegt, dann wird das Ratenbestimmungselement **106** bestimmen, dass der Rahmen mit Achtelrate codiert werden soll, und so weiter, wie bekannt ist.

[0020] Eine andere Ratenbestimmungstechnik ist offenbart in EP-A-1,339,044 mit dem Titel "Method and Apparatus For Performing Reduced Rate Variable Rate Vocoding", dem Rechtsnachfolger der vorliegenden Erfindung zugeordnet. Diese Technik sieht einen Satz von Ratenentscheidungskriterien vor, die als Modusmessungen bezeichnet werden. Eine erste Modusmessung ist der Zielabgleichssignal-Rauschabstand (TMSNR = target matching signal to noise ratio) von dem vorhergehenden Codierrahmen, der Informationen vorsieht, wie gut das Codierungsmodell arbeitet, und zwar durch das Vergleichen eines synthetisierten Sprachsignals mit dem Eingabesprachsignal. Eine zweite Modusmessung ist die normalisierte Autokorrelationsfunktion (NACF = normalized autocorrelation function), die die Periodizität des Sprachrahmens misst. Eine dritte Modusmessung ist der Nulldurchgangparameter (ZC = zero crossing), der den Hochfrequenzinhalt in einem Eingabesprachrahmen misst. Eine vierte Messung, der Vorhersageverstärkungsunterschied (PGD = prediction gain differential) bestimmt, ob der Codierer seine Vorhersageeffizienz aufrecht erhält. Eine fünfte Messung ist der Energieunterschied (ED = energy differential), der die Energie in dem aktuellen Rahmen mit einer Durchschnittsrahmenenergie vergleicht.

[0021] Unter der Verwendung der Modusmessungen, die oben diskutiert wurden, wählt die Ratenbestimmungslogik eine Codierungsrate für jeden Rahmen der Eingabesprachdaten aus. Die Werte für die verschiedenen Modi wählen einen von vier oder mehr Modi, in denen operiert wird. D.h., die Werte, die für jede Modusmessung relativ zu einer Schwelle oder anderen Kriterien detektiert wurden, bestimmen, welche Codierate gewählt wird, und zwar basierend auf einem vorausgewählten Muster oder Hierarchie. Wenn zum Beispiel der Wert für NACF niedriger ist als eine vorausgewählte Schwelle und ZC größer ist als eine zweite vorausgewählte Schwelle, könnte eine Rate gewählt werden. Wenn jedoch diese Bedingungen nicht erfüllt werden, aber ED niedriger ist als eine dritte Schwelle, dann könnte eine Viertelrate gewählt werden. Wenn der Wert für TSNR größer ist, PGD kleiner ist und NACF größer ist als die vierte, fünfte und sechste Schwelle, entsprechend, dann könnte eine Halbrate gewählt werden. Verschiedene solcher Kombinationen und Schwellen können vom Fachmann auf gewählte Codieraten angewendet werden.

[0022] Es sei angemerkt, dass noch andere Ratenbestimmungstechniken von dem Ratenbestimmungselement **106** angenommen werden können.

[0023] Noch immer Bezug nehmend auf die [Fig. 1](#) wird ein Signal als Anzeige für die Datenrate, die von dem Ratenbestimmungselement **106** bestimmt wurde, zu einem Switch **108** geliefert. Switch **108** wählt ein Element zum Codieren eines Rahmens der Eingabesprachsamples unter einem Vollratencodierelement **110**, einem Halbratencodierelement **112**, einem Viertelratencodierelement **114** und einem Achtelratencodierelement **116** aus, wie vorgesehen von dem Datenratensignal. Das gewählte Codierelement codiert die Sprachsamples, um ein Signal eines codierten Datenpakets zu erzeugen bzw. zu produzieren. Das Ratenbestimmungselement **106** liefert auch ein Signal als Anzeige für die Datenrate zu einem Switch **118**, der das gleiche Codierelement wie Switch **108** auswählt, so dass das Signal des codierten Datenpakets, das vom gewählten Codierelement erzeugt wurde, zu einem Ausgang des Vocoders mit variabler Rate geliefert werden kann.

[0024] Jedes der Codierelemente **110**, **112**, **114** und **116** ist konfiguriert, um Sprache unter Verwendung eines vorbestimmten Codierschemas zu codieren. Ein Linear-Vorhersage-basiertes Codierschema, wie zum Beispiel der Code-Excited-Linear-Predictive-(CELP)-Codierer wird in einem bevorzugten Ausführungsbeispiel benutzt. Der CELP-Codierer wird in dem Paper "A 4.8 Kbps Code Excited Linear Predictive Coder", von Thomas E. Tremain et al., Proceedings of the Mobile Satellite Conference, 1988, beschrieben. Linear-Vorhersage-basierte Codierer komprimieren Sprache durch Entfernen der natürli-

chen Redundanzen, die der Sprache anhaftend bzw. inhärent sind. Sprache besitzt typischerweise kurzzeitige Redundanzen, die von der mechanischen Aktion der Lippen und der Zunge resultieren, und Langzeitredundanzen, die von der Vibration der Stimmbänder resultieren. Linear-Vorhersage-Schemata modellieren diese Operationen als Filter, entfernen die Redundanzen und modulieren anschließend das resultierende Restsignal als weißes Gauss'sches Rauschen. Linear-Vorhersage-Codierer erreichen deswegen eine verminderte Bitrate durch Senden von Filterkoeffizienten und quantisiertes Rauschen im Vergleich zu einem Vollbandbreitensprachsignal.

[0025] Ein Linear-Vorhersage-Codierungsschema, das variable Raten anwendet, bietet weitere Reduktionen in der Bitrate ohne die Sprachqualität zu beeinträchtigen. In [Fig. 1](#) codiert das Vollratencodierelement **110** die Parameter des Eingabesprachsignals unter Verwendung mehrerer Bits, um die Charakteristiken der Eingabe besser zu erhalten. Für Perioden, wo keine Sprache detektiert wird, codiert das Achtelratencodierelement **116** die Parameter unter Verwendung weniger Bits, da es typischerweise geringes Detail oder nützliche Informationen gibt, die es zu erfassen gilt. Übergänge zwischen Perioden von aktiver Sprache und Perioden mit nicht detektierter Sprache werden mit dem Halbratencodierelement **112** und dem Viertelratencodierelement **114** codiert.

[0026] Nun Bezug nehmend auf das Decodierelement des Vocoders mit variabler Rate empfängt der Decodierer **104** ein Signal der codierten Sprachparameter, wie auch ein Signal als Anzeige für die Rate, die verwendet wird, um die Sprache zu codieren. Ein Ratenextraktionselement **128** empfängt dieses Eingangssignal und bestimmt die Datenrate der Sprache. Ein Signal der Datenrate wird auch zu einem Switch **130** geliefert, der das Decodierelement von einem Satz von Decodierelementen auswählt, um die Eingabeparameter richtig zu decodieren. In [Fig. 1](#) sind vier Decodierelemente, Vollratendecodierelement **120**, Halbratendecodierelement **122**, Viertelratendecodierelement **124** und Achtelratendecodierelement **126** vorgesehen, und zwar zum Decodieren der Sprachparameter mit den vier möglichen Raten. Das gewählte Decodierelement decodiert die Eingabeparameter basierend auf der Datenrate, um ein Signal der decodierten Samples zu produzieren, die typischerweise 64-kbps-Puls-Code-modulierte (PCM = pulse code modulated) Samples sind. Ein Signal der Datenrate, die von dem Ratenextraktionselement **128** bestimmt wurde, wird auch zu einem Switch **132** geliefert. Der Switch **132** wählt das gleiche Decodierelement wie Switch **130** aus, so dass ein Signal der decodierten Samples zu einem Ausgang des Vocoders gesendet wird.

[0027] Nun Bezug nehmend auf die [Fig. 2](#) ist ein Blockdiagramm eines Sprachsynthesystems ge-

zeigt, das gemäß den Prinzipien der vorliegenden Erfindung operiert und wobei das System einen Vocoder mit variabler Rate inkorporiert bzw. beinhaltet. Das Sprachsynthesesystem weist einen Codierer mit variabler Rate **202** und einen Sprachsynthesizer **204** auf. Ein Beispiel des Codierers mit variabler Rate **202** ist Codierer **102** von [Fig. 1](#). Codierer mit variabler Rate **202** empfängt ein Sprachsignal als Eingabe und codiert die Sprache mit einer von einem Satz von, vorbestimmten Raten. In einem bevorzugten Ausführungsbeispiel ist der Codierer mit variabler Rate **202** ein CELP-Codierer, der Sprachparameter bei einer von den Raten basierend auf der Sprachaktivität in dem Eingabesegment der Sprache erzeugt.

[0028] Die vorliegende Erfindung benutzt einen Vocoder mit variabler Rate, wie beschrieben im US-Patent Nr. 5,414,796, wie oben beschrieben, welches kommerziell verfügbar ist, zum Beispiel als ein 13-kbps-Vocoder-Produkt von Qualcomm Incorporated. In einem bevorzugten Ausführungsbeispiel ist der Decodierer mit variabler Rate ein erweiterter Decodierer mit variabler Rate, wie zum Beispiel beschrieben mit Bezug auf den IS127-Standard.

[0029] In einem Ausführungsbeispiel der vorliegenden Erfindung basieren die Codierratenentscheidungen auf "Modusmessungen", wie oben diskutiert. Die verschiedenen Kombinationen der Kriterien, die verwendet werden, um Ratenauswahl zu treffen, werden dazu verwendet, um etwas zu erschaffen, das als "reduzierter Ratenmodus" oder "-modi" bezeichnet wird und einfach als Modus 0, Modus 1, Modus 2 und so weiter bezeichnet wird, so wie es für den Fachmann klar sein würde. Die vorliegende Erfindung kann aus solchen Modi für die Zwecke der Sprachsynthese einen Vorteil ziehen.

[0030] Die Sprache, die vom Codierer mit variabler Rate **202** empfangen wird, kann ein Wort oder eine Phrase bzw. Satz aus einem vorausgewählten Vokabular sein, für das ein Kommunikationsgerät, wie zum Beispiel ein Drahtlostelefon, Autofreisprechanlage oder ein anderes Kommunikationsgerät entwickelt wurde, zu synthetisieren. Das Vokabular würde Eingabeaufforderungen und Warnsignale beinhalten, die einem Gerätebenutzer gegeben werden sollen. Durch Extrahieren und Synthetisieren von fünf Individualvokabularwörtern: "Call" bzw. "Anrufen", "Redial" bzw. "Wiederwählen", "Program" bzw. "Programmieren", "Or" bzw. "Oder" und "Exit" bzw. "Verlassen", kann der Sprachsynthesizer zum Beispiel entwickelt worden sein, um die Eingabeaufforderungen "Call, Redial, Program, Or, Exit" in der Ansuchung einer Antwort von dem Benutzer vorsehen. Als Alternative kann der Sprachsynthesizer dafür entwickelt werden, um vorangegangene gespeicherte Information, wie zum Beispiel ein Telefonbuch, Nachschlagetabellen oder Datenbanken für einen Gerätebenutzer ansprechend auf verschiedene Geräteeingaben, einschließ-

lich Audio, vorzusehen. Die Sprache, die vom Codierer mit variabler Rate **202** empfangen wird, wird codiert und die codierten Parameter werden zu einem Speicherelement oder Schaltung **206** des Sprachsynthesizers **204** für die Aufbewahrung geliefert.

[0031] Der Speicher **206** ist dazu gedacht, die Parameter über einige Zeit zu halten oder zu speichern, und zwar für die Operation des gewünschten Geräts. Jedoch ist es ebenso generell wünschenswert, die Parameter auf eine Art und Weise gespeichert zu haben, die sie aktualisierbar oder ersetzbar macht, wie zum Beispiel, wenn das Vokabular für sich ändernde Bedingungen oder Aktualisierungen bzw. Upgrades für die Gerätemerkmale geändert werden soll. Deswegen ist der Speicher **206** in der Form des nicht-flüchtigen, aber wieder beschreibbaren Speichers konfiguriert, was unter der Verwendung von Flash-Typ-Speicherelementen, wie auf dem Fachgebiet bekannt ist, erreicht werden kann.

[0032] Wie erkannt werden würde, kann die Operation des Ladens der Parameter während der Herstellung eines Kommunikationsgeräts für das die Erfindung benutzt werden soll durchgeführt werden. Da die Eingabeaufforderungen und Alarmsignale, die synthetisiert werden sollen, vorbestimmt sein können, können diese während der Herstellung codiert werden und im Flash-Speicher **206** vor der Verwendung gespeichert werden. Die Parameter können geändert oder ersetzt werden und zwar während des Services des Geräts oder über neu entwickelte Über-die-Luft-Programmierungstechniken für Drahtlosgeräte.

[0033] Als Alternative kann der Codierer mit variabler Rate **202** eine Sprachsignaleingabe während der Operation bzw. des Betriebs des Kommunikationsgeräts empfangen. Ansprechend auf eine Eingabeaufforderung von dem Sprachsynthesizer kann zum Beispiel der Benutzer eine gesprochene Antwort liefern. Der Codierer mit variabler Rate **202** wird anschließend die Sprache des Benutzers codieren und die codierten Parameter können zum Flash-Speicher **206** zur Speicherung und/oder zu einem Spracherkenner (nicht gezeigt) für Spracherkennungszwecke geliefert werden. Auf diese Weise werden die Parameter nach der Produktion, wie zum Beispiel sofort, wenn das Gerät einen nützlichen Dienst aufnimmt oder über die Zeit, wie zum Beispiel durch Aufbauen einer persönlichen Vokabularsammlung für jeden Geräte-(Vocoder)-Benutzer bezogen auf diese Anforderungen des Benutzers eingegeben.

[0034] Der Flash-Speicher **206** sollte von einer Größe sein, die groß genug ist, um die Parameter des vorausgewählten Vokabulars wie auch die Parameter der Sprache, die vom Benutzer erwartet wird, zu speichern. Somit kann die Größe des Flash-Speichers **206** basierend auf den Anforderungen der spe-

zifischen Applikation variieren. Postproduktionsspeicherung kann einen Vorteil haben, und zwar das Reduzieren der Speicheranforderungen, wo jeder Gerätebenutzer kein so extensives Vokabular benötigt, verglichen zu was ein Hersteller installieren müsste, um einen gesamten größeren Gerätemarkt abzudecken. Der Sprachsynthesizer kann Namen oder andere Wörter aufnehmen, wie zum Beispiel "Fred Smith" durch Detektieren der Endpunkte des Ziels oder der gewünschten Phrase oder Sprache, durch Entfernen der Stille oder Redundanzen und dessen Codierung. Deswegen kann Sprache "on-line" aufgenommen werden und später dazu verwendet werden, um die Sprachausgabe zu synthetisieren.

[0035] Es sei angemerkt, dass der Codierer mit variabler Rate **202** basierend auf dem verfügbaren Speicher und der benötigten Sprachqualität konfiguriert werden kann. In dem System mit vier Raten, wobei die Vollrate bei 13 kbps liegt, wird die Durchschnittsrate generell bei 5,88 kbps basierend auf 40% Sprachaktivität sein. Die Verwendung der variablen Raten wird Hochsprachqualität vorsehen. Wenn jedoch die Speichergröße begrenzt ist, kann der Codierer mit variabler Rate **202** so konfiguriert werden, um bei, sozusagen, einer festen Halbrate von näherungsweise 800 Bytes pro Sekunden zu operieren. Anderenfalls kann die Rate aus einem Sub-Satz des vorbestimmten Satzes von Raten statt aus dem ganzen Satz von Raten gewählt werden. Die reduzierten Ratenmodi, die oben diskutiert wurden, können zum Beispiel benutzt werden, um verschiedene Raten auszuwählen. In einem Ausführungsbeispiel der Erfindung sind die Raten in einem Satz von vier Modi, gekennzeichnet als Modus 0, 1, 2 und 3, aufgeteilt. Unter Verwendung von festen Raten gemäß dem Modus, Raten der Ordnung von 1800 Bytes pro Sekunde, 1540 Bytes pro Sekunde, 1400 Bytes pro Sekunde und 1100 Bytes pro Sekunde können entsprechend benutzt werden. Die Verwendung von solchen festen reduzierten Raten erlaubt die Lieferung von sehr hoch qualitativer Sprache angesichts einer vordefinierten Datenrate, eine Landleitungsqualität annähernd. Diese vier Modi sehen die beste Abwägung zwischen synthetisierter Sprachqualität und Speicheranforderung vor.

[0036] Des Weiteren kann der Codierer mit variabler Rate **202** zwischen verschiedenen Betriebsmodi bzw. Operationsmodi (variable Rate, immer Halbrate, einen Sub-Satz der variablen Raten, etc.) basierend auf den momentanen Anforderungen der Applikation wechseln. Da es eine Abwägung zwischen Sprachqualität und Speichergröße geben kann, wird die Konfiguration, die angenommen werden soll, von der Applikation, die implementiert worden ist, abhängen.

[0037] Die Sprachparameter, die im Flash-Speicher **206** gespeichert sind, werden zu einem Decodierer mit variabler Rate **208** geliefert, wenn Sprachsynthe-

se gewünscht ist. Der Decodierer mit variabler Rate **208** ist konfiguriert, um die Parameter, die vom entsprechenden Codierer mit variabler Rate **202** erzeugt wurden, zu decodieren. Ein Beispiel für einen Decodierer mit variabler Rate **208** ist Decodierer **104** in [Fig. 1](#).

[0038] Generell wird der Decodierer mit variabler Rate **208** als Teil eines digitalen Signalprozessors (DSP), der innerhalb des Kommunikationsgeräts benutzt wird, implementiert werden. Solche DSPs werden benutzt, als oder, um Verarbeitungselemente für Signalcodierung/Decodierung, Kombination, CD-MA-Codierung, Leistungsanpassung und so weiter zu bilden. Da solche Elemente typischerweise in Drahtlosgeräten benutzt werden und in vielen anderen Geräten, in denen die Erfindung Nutzen finden kann, kann ein Vorteil aus deren Präsenz gezogen werden, um sehr kosteneffektiv die vorliegende Erfindung zu implementieren.

[0039] Um die Decodierungsfunktionalität der vorliegenden Erfindung zu implementieren wird nur eine kleine Menge an Speicher in oder gekoppelt an einen DSP benötigt. Ein Stand-Alone-Decodierer innerhalb oder unter Verwendung eines DSPs benötigt eine sehr kleine Menge an Speicher (beides, Programm und Daten), um eine Sprachsynthesefähigkeit zu erlangen. Der Sprachsynthesizer kann unter Verwendung von bekannten DSP-Schaltungen und Geräten, wie zum Beispiel kommerziell verfügbaren von Analog Devices und Qualcomm Incorporated, implementiert werden.

[0040] Die decodierten Parameter werden anschließend typischerweise in der Form von Puls-Code-modulierten (PCM) Samples zu einem Codec **210** geliefert. Codec **210** konvertiert die PCM-Samples von einem digitalen Format in ein analoges Signal. Das analoge Signal wird zu einem Lautsprecher oder einem anderem bekannten Audioausgabegerät **212** geliefert, das die synthetisierte Sprache in die umgebende Geräteumgebung projiziert oder ausstrahlt, wo es gehört werden kann.

[0041] Deswegen ist ein Sprachsynthesizer basierend auf dem Sprachcodieren mit variabler Rate von der vorliegenden Erfindung vorgesehen. Der Sprachsynthesizer ist speziell geeignet für die Verwendung in drahtlosen Kommunikationsgeräten, die bereits einen Vocoder mit variabler Rate aufweisen. In anderen Worten kann ein existierender Vocoder mit variabler Rate vom Sprachsynthesizer angewendet werden, und zwar durch die Verwendung von passenden bzw. geeigneten Änderungen im Programm oder in den Operationsinstruktionen oder unter der Verwendung von Steuerhardware. Zusätzlich kann über die Verwendung des Sprachcodierens mit variabler Rate die Komprimierung, die erreicht wird, es erlauben, ein vorbestimmtes Vokabular, das in einem Speicher von

begrenzter Größe assoziiert mit dem Drahtlosgerät oder anderem Zubehör, mit dem es verbunden ist, gespeichert werden soll, zu ermöglichen. Weiterhin kann die Abwägung von Sprachqualität und Speichergröße bei der Konfiguration des Vocoder mit variabler Rate berücksichtigt werden, um einen Sprachsynthesizer mit der gewünschten Sprachqualität und Speichergröße vorzusehen.

[0042] Die vorliegende Erfindung kann Anwendung finden in einer Vielfalt von Kommunikationsgeräten und Schnittstellenzubehör. Die obigen Beispielausführungsbeispiele wurden mit Bezug auf Drahtloskommunikationsgeräte diskutiert, wie zum Beispiel, aber nicht begrenzt auf, zellulare und Satellitentelefone, oft bezeichnet als Benutzerterminals, Teilnehmerseinheiten, Mobilstationen oder einfach "Benutzer", "Handys" bzw. "Mobiles" oder "Teilnehmer". Zusätzlich sind auch andere Geräte in Erwägung gezogen, wie zum Beispiel Nachrichtenempfänger und Datentransfergeräte (zum Beispiel tragbare Computer, persönliche Datenassistenten, Modems, Maschinensteuerungsgeräte), oder Schnittstellen bzw. Interface für öffentliche Telefonnetzwerke oder dedizierte Kommunikationskanäle.

[0043] Die Erfindung kann implementiert werden unter der Verwendung von getrennten Schaltungen in der Form von dedizierten Komponenten oder applikationsspezifischen integrierten Schaltungen (ASIC = application specific integrated circuit), um einen Sprachsynthesizer zu bilden, der innerhalb eines gewünschten Geräts installiert wird. Als Alternative kann sie innerhalb anderer ASICs und Geräte inkorporiert werden, und zwar durch Benutzen einer kleinen Menge von zusätzlichem Speicher, um mit existierenden Digitalsignalverarbeitungselementen zusammenzuarbeiten.

[0044] Die vorangegangene Beschreibung der bevorzugten Ausführungsbeispiele ist vorgesehen, um jedem Fachmann zu ermöglichen, die vorliegende Erfindung zu produzieren oder zu benutzen. Die verschiedenen Modifikationen zu diesen Ausführungsbeispielen werden dem Fachmann leicht ersichtlich sein und die ursprünglichen Prinzipien, die hierin definiert wurden, können ohne die Verwendung der erfinderischen Fähigkeit auf andere Ausführungsbeispiele angewandt werden. Somit ist es nicht beabsichtigt, die vorliegende Erfindung auf die Ausführungsbeispiele, die hierin gezeigt sind, zu begrenzen, sondern ihr soll der breiteste Schutzbereich, wie definiert durch die Ansprüche, eingeräumt werden.

Patentansprüche

1. Eine Vorrichtung (**204**) zum Synthetisieren von vorselektiertem Vokabular in einem drahtlosen Kommunikationssystem, wobei das vorselektierte Vokabular von einem Codierer mit variabler Rate (**102**,

202) mit einem Satz variabler Raten codiert wurde, wobei die Vorrichtung (**204**) folgendes aufweist: Einen Speicher (**206**) zum Speichern von einem Satz von Sprachparametern; einen Prozessor, der konfiguriert ist, eine sprachliche Eingabe von einem Anwender zu akzeptieren; einen Decodierer mit variabler Rate (**104**, **208**) zum Decodieren der Sprachparameter, um decodierte Sprachsamples zu generieren bzw. zu erzeugen; und einen Digital-zu-Analog-Konverter (**210**) zum Konvertieren der Sprachsamples in ein analoges Signal zum Ausstrahlen als synthetisierte Sprache, **dadurch gekennzeichnet**, dass: der Satz von Sprachparametern das Codierte, vorselektierte Vokabular repräsentiert; der Prozessor ist weiterhin konfiguriert, einen Sub-Satz von Sprachparametern dieser Sprachparameter gemäß der sprachlichen Eingabe des Anwenders auszuwählen; und der Decodierer mit variabler Rate (**104**, **208**) ist konfiguriert, um den Sub-Satz von Sprachparametern zu decodieren.

2. Vorrichtung (**204**) nach Anspruch 1, wobei der Codierer mit variabler Rate (**102**, **202**) auf linearer Prädiktion bzw. Linear-Prediction basiert.

3. Vorrichtung (**204**) nach Anspruch 1, wobei der Decodierer mit variabler Rate (**104**, **208**) auf linearer Prädiktion basiert.

4. Vorrichtung (**204**) nach Anspruch 1, wobei der Satz von Sprachparametern mit einem Satz von variablen Raten codiert wurde, welcher eine Vollrate, eine Halbrate, eine Viertelrate und eine Achtelrate aufweist.

5. Vorrichtung (**204**) nach Anspruch 4, wobei die Vollrate 13,2 kbps beträgt, die Halbrate ungefähr 6,2 kbps und die Viertelrate ungefähr 2,7 kbps und die Achtelrate ungefähr 1,0 kbps.

6. Vorrichtung (**204**) nach Anspruch 4, wobei der Satz von Sprachparametern mit einer fixen Rate, ansprechend auf eine oder mehrere gemessene Moduskriterien codiert wurde.

7. Vorrichtung (**204**) nach Anspruch 4, wobei der Satz von Sprachparametern mit einer Rate festgelegt auf die Halbrate codiert wird.

8. Vorrichtung (**204**) nach Anspruch 4, wobei die Codierrate gemäß den Anforderungen der Sprachqualität und der Größe des Speichers (**206**) selektiert bzw. ausgewählt wird.

9. Vorrichtung (**204**) nach Anspruch 1, wobei das drahtlose Kommunikationssystem ein CDMA-System ist.

10. Vorrichtung (**204**) nach Anspruch 1 wobei der Codierer mit variabler Rate (**102, 202**) einen erweiterten bzw. verbesserten Codierer mit variabler Rate aufweist.

11. Ein Verfahren zur Synthetisierung eines vorselektierten Vokabulars in einem drahtlosen Kommunikationssystem, wobei das vorselektierte Vokabular von einem Codierer mit variabler Rate (**102, 202**) mit einem Satz variabler Raten codiert wurde, wobei das Verfahren Folgendes aufweist:

Empfangen einer sprachlichen Anwendereingabe;
Abrufen eines Satzes von Sprachparametern, gespeichert in einem Speicher (**206**);
Decodieren (**104, 208**) des Satzes von Sprachparametern unter Verwendung von einem Codierungsschema mit variabler Rate, um die decodierten Sprachsamples zu erzeugen; und
Konvertieren (**210**) der Sprachsamples in ein analoges Signal für das Ausstrahlen als synthetisierte Sprache, gekennzeichnet dadurch, dass:
der Satz von Sprachparametern das codierte vorselektierte Vokabular repräsentiert;
ein Sub-Satz von Sprachparametern wird aus dem Satz von Sprachparametern gemäß der sprachlichen Anwendereingabe ausgewählt; und
der Sub-Satz von Sprachparametern wird decodiert unter Verwendung des Decodierungsschemas mit variabler Rate (**104, 208**).

12. Verfahren nach Anspruch 11, wobei der Codierer mit variabler Rate (**102, 202**) ein Codierungsschema mit variabler Rate basierend auf linearer Prädiktion durchführt.

13. Verfahren nach Anspruch 11, wobei das Codierungsschema mit variabler Rate auf linearer Prädiktion basiert.

14. Verfahren nach Anspruch 11, wobei der Satz von Sprachparametern mit einem Satz von variablen Raten codiert wurde, welcher eine Vollrate, eine Halbrate, eine Viertelrate und eine Achtelrate aufweist.

15. Verfahren nach Anspruch 14, wobei die Vollrate 13,2 kbps beträgt, die Halbrate ungefähr 6,2 kbps, die Viertelrate ungefähr 2,7 kbps und die Achtelrate ungefähr 1,0 kbps.

16. Verfahren nach Anspruch 14, wobei der Satz von Sprachparametern mit einer fixen Rate ansprechend auf eine oder mehrere gemessenen Moduskriterien codiert wurde.

17. Verfahren nach Anspruch 14, wobei der Satz von Sprachparametern mit einer Rate festgelegt auf die Halbrate codiert wird.

18. Verfahren nach Anspruch 14, wobei die Co-

dierrate gemäß den Anforderungen der Sprachqualität und der Größe des Speichers (**206**) selektiert wird.

19. Verfahren nach Anspruch 11, wobei das drahtlose Kommunikationssystem ein CDMA-System aufweist.

20. Verfahren nach Anspruch 11, welches weiterhin die folgenden Schritte aufweist:

Codieren (**102, 202**) der sprachlichen Anwendereingabe; und
Hinzufügen der codierten sprachlichen Anwendereingabe zu dem Speicher (**206**) als ein Teil des Satzes von Sprachparametern.

Es folgen 2 Blatt Zeichnungen

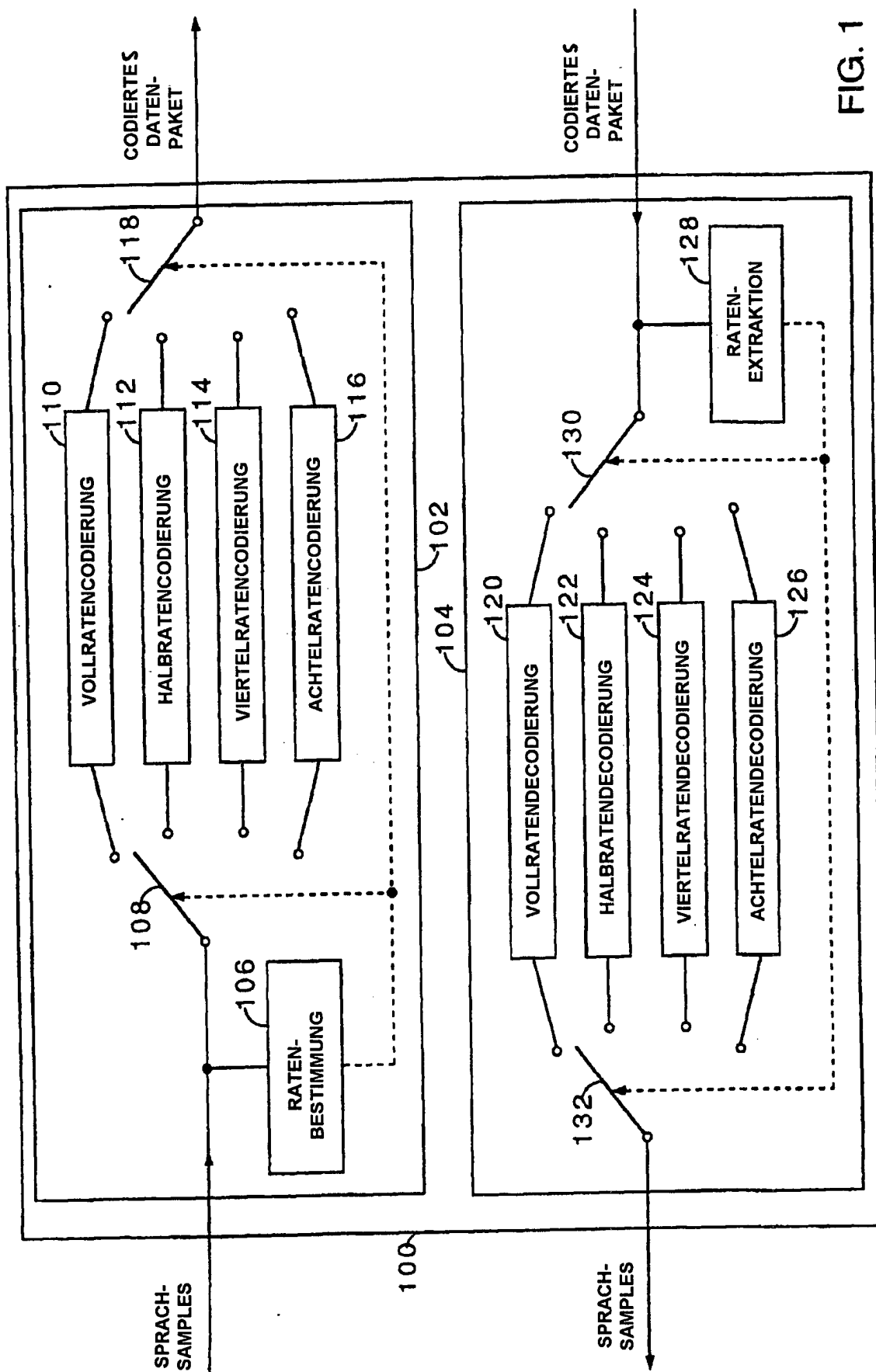


FIG. 1

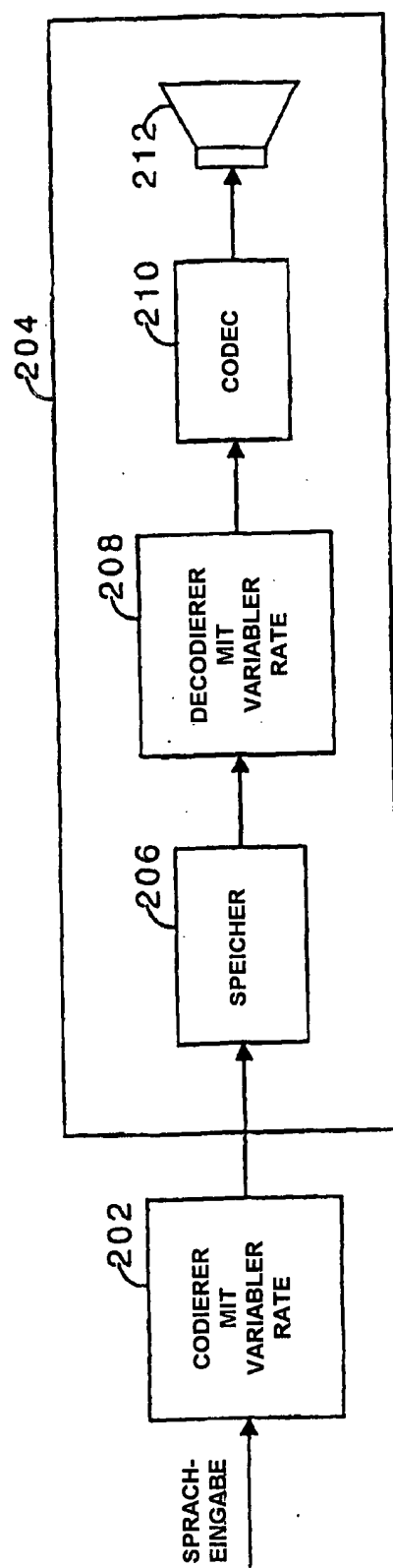


FIG. 2