



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 695 35 709 T2** 2009.02.12

(12) **Übersetzung der europäischen Patentschrift**

(97) **EP 1 703 493 B1**

(21) Deutsches Aktenzeichen: **695 35 709.3**

(96) Europäisches Aktenzeichen: **06 013 824.5**

(96) Europäischer Anmeldetag: **01.08.1995**

(97) Erstveröffentlichung durch das EPA: **20.09.2006**

(97) Veröffentlichungstag

der Patenterteilung beim EPA: **13.02.2008**

(47) Veröffentlichungstag im Patentblatt: **12.02.2009**

(51) Int Cl.⁸: **G10L 19/14** (2006.01)

G10L 19/02 (2006.01)

G10L 11/02 (2006.01)

(30) Unionspriorität:

288413 10.08.1994 US

(73) Patentinhaber:

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

(74) Vertreter:

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

(84) Benannte Vertragsstaaten:

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

(72) Erfinder:

**Dejaco, Andrew P., 80538 München, DE; Gardner,
William R., San Diego CA 92130, US**

(54) Bezeichnung: **Verfahren und Vorrichtung zur Auswahl der Kodiertrate bei einem Vokoder mit variabler 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**I. Gebiet der Erfindung**

[0001] Die vorliegende Erfindung bezieht sich auf Vocoder. Insbesondere bezieht sich die vorliegende Erfindung auf ein neues und verbessertes Verfahren zur Bestimmung einer Sprachcodiertrate in einem Vocoder mit variabler Rate (variable rate vocoder).

II. Beschreibung des Standes der Technik

[0002] Sprachkomprimierungssysteme mit variabler Rate benutzen üblicherweise eine Art von Ratenbestimmungsalgorithmus vor dem Beginn des Codierens. Der Ratenbestimmungsalgorithmus weist ein Codierungsschema mit höherer Bitrate Segmenten des Audiosignals zu, in denen Sprache vorliegt, und weist ein Codierungsschema mit niedrigerer Rate Pausensegmenten zu. Auf diese Art und Weise wird eine niedrigere durchschnittliche Bitrate erreicht, während die Sprachqualität der rekonstruierten Sprache hoch verbleibt. Um daher effizient betrieben zu werden, benötigt ein Sprachcodierer mit variabler Rate einen robusten Raten- bzw. Geschwindigkeitsbestimmungsalgorithmus, der Sprache von Pausen (silence) in einer Vielzahl von Hintergrundrauschumgebungen unterscheiden kann.

[0003] Ein solches Sprachkomprimierungssystem mit variabler Rate bzw. Vocoder mit variabler Rate ist in der WO-A1-92/22891, eingereicht am 11. Juli 1991, betitelt „Variable rate vocoder“ und dem Rechtsnachfolger der vorliegenden Erfindung zugewiesen, offenbart. In dieser bestimmten Implementierung eines Vocoders mit variabler Rate wird Eingabesprache mittels Code Excited Linear Predictive Coding-(CELP)-Techniken mit einer von mehreren Raten codiert, und zwar bestimmt durch den Pegel der Sprachaktivität (level of speech activity). Der Pegel der Sprachaktivität wird aus der Energie in den Eingabe-Audio-Samples, die Hintergrundrauschen zusätzlich zu stimmhafter Sprache enthalten können, bestimmt. Damit der Vocoder eine hochqualitative Sprachcodierung bei verschiedenen Pegeln von Hintergrundrauschen vorsehen kann, wird eine adaptive Technik zur Schwellenwerteinstellung benötigt, um die Auswirkungen des Hintergrundrauschens auf den Ratenentscheidungsalgorithmus zu kompensieren.

[0004] Vocoder werden typischerweise in Kommunikationsvorrichtungen, wie z. B. zellularen Telefonen oder Personenkommunikationsvorrichtungen (personal communication devices) eingesetzt, um digitale Signalkomprimierung eines Analog-Audiosignals, das für die Übertragung in digitale Form konvertiert wird, vorzusehen. In einer Mobilfunkumgebung, in der ein zellulares Telefon oder eine Personenkommunikationsvorrichtung eingesetzt werden kann, gestalten es hohe Pegel von Hintergrundrauschenergie für den Ratenbestimmungsalgorithmus schwierig, unstimmhafte Töne mit niedriger Energie von Pausenhintergrundrauschen mittels eines auf Signalenergie basierenden Ratenbestimmungsalgorithmus zu unterscheiden. Die unstimmhaften Töne werden häufig mit niedrigeren Bitraten codiert, und die Sprachqualität verschlechtert sich, da Konsonanten, wie z. B. „s“, „x“, „ch“, „sh“, „t“, etc. in der rekonstruierten Sprache verloren gehen.

[0005] Vocoder, die Ratenentscheidungen nur auf der Energie des Hintergrundrauschens basieren, berücksichtigen nicht die Signalstärke relativ zu dem Hintergrundrauschen beim Setzen von Schwellenwerten. Ein Vocoder, der seine Schwellenwertpegel nur auf Hintergrundrauschen basiert, tendiert dazu, die Schwellenwertpegel zusammenzukomprimieren, wenn das Hintergrundrauschen ansteigt. Verbliebe der Signalpegel festgelegt, wäre dies der korrekte Ansatz, die Schwellenwertpegel einzustellen. Wenn der Signalpegel jedoch mit dem Hintergrundrauschpegel ansteigt, dann ist die Kompression bzw. Komprimierung der Schwellenwertpegel nicht eine optimale Lösung. Ein alternatives Verfahren zum Einstellen der Schwellenwertpegel, das die Signalstärke berücksichtigt, wird für Vocoder mit variabler Rate benötigt.

[0006] Ein Schlussproblem, das noch verbleibt, ergibt sich während des Abspielens von Musik bei Vocodern, deren Ratenentscheidung auf der Hintergrundrauschenergie basiert. Wenn Personen sprechen, müssen sie zwischendurch anhalten, um zu atmen, was es den Schwellenwertpegeln ermöglicht, auf den richtigen Hintergrundrauschpegel zurückgesetzt zu werden. Bei der Übertragung von Musik durch einen Vocoder, wie es z. B. bei Musikwarteschleifen-Zuständen auftritt, treten keine Pausen auf, und die Schwellenwertpegel werden kontinuierlich ansteigen bis damit begonnen wird, die Musik mit einer Rate, die geringer ist als die Vollrate zu codieren. In einem solchen Zustand hat der Codierer mit variabler Rate Musik mit Hintergrundrauschen verwechselt.

[0007] Das Dokument „QCELP: The North American CDMA Digital Cellular Variable Rate Speech Coding Standard“, Proc. IEEE Workshop on Speech Processing for Telecommunications, 1993, Seiten 85–86, by De

Jaco et al., offenbart einen Variabelratenselector, der auf der Verwendung von drei variablen Schwellenwerten basiert, die über der Hintergrundrauschschätzung gehalten sind.

[0008] Die vorliegende Erfindung ist in den unabhängigen Ansprüchen 1, 6 und 11 dargelegt.

Zusammenfassung der Erfindung

[0009] Die vorliegende Erfindung ist ein neues und verbessertes Verfahren und eine Vorrichtung zum Bestimmen einer Kodierate in einem Vocoder mit variabler Rate. Es ist ein erstes Ziel der vorliegenden Erfindung ein Verfahren vorzusehen, bei dem die Wahrscheinlichkeit des Kodierens von unstimmhafter Sprache mit niedriger Energie als Hintergrundrauschen reduziert wird. In der vorliegenden Erfindung wird das Eingangssignal in eine Hochfrequenzkomponente und eine Niedrigfrequenzkomponente gefiltert. Die gefilterten Komponenten des Eingangssignals werden dann individuell analysiert, um das Vorliegen von Sprache zu detektieren. Da unstimmhafte Sprache (unvoiced speech) eine Hochfrequenzkomponente hat, ist deren Stärke relativ zu einem Hochfrequenzband gegenüber dem Hintergrundrauschen in dem Frequenzband ausgeprägter als deren Stärke im Vergleich zu dem Hintergrundrauschen über das gesamte Frequenzband.

[0010] Ein zweites Ziel der vorliegenden Erfindung ist es, ein Mittel vorzusehen, mit dem die Schwellenwertpegel eingestellt werden, und das Signalenergie sowie Hintergrundrauschenergie berücksichtigt. In der vorliegenden Erfindung ist das Einstellen der Sprachdetektierschwellenwerte auf einer Schätzung des Signal-zu-Rausch-Verhältnisses (signal to noise ratio (SNR)) des Eingangssignals basiert. In dem Ausführungsbeispiel wird die Signalenergie als die Maximalsignalenergie während Zeiten von aktiver Sprache geschätzt, und die Hintergrundrauschenergie wird als die minimale Signalenergie während Pausenzeiten geschätzt.

[0011] Ein drittes Ziel der vorliegenden Erfindung ist es, ein Verfahren zum Kodieren von Musik, die durch einen Vocoder mit variabler Rate tritt, vorzusehen. In dem Ausführungsbeispiel detektiert die Ratenauswahlvorrichtung eine Anzahl von aufeinanderfolgenden Rahmen, über denen die Schwellenwertpegel angestiegen sind und überprüft hinsichtlich einer Periodizität über die Anzahl der Rahmen hinweg. Wenn das Eingangssignal periodisch ist, würde dies das Vorliegen von Musik anzeigen. Wenn das Vorliegen von Musik detektiert wird, werden die Schwellenwerte auf solche Pegel gesetzt, dass das Signal mit voller Rate kodiert wird.

Kurze Beschreibung der Zeichnungen

[0012] Die Merkmale, Ziele und Vorteile der vorliegenden Erfindung werden aus der unten folgenden detaillierten Beschreibung noch offensichtlicher, wenn diese in Zusammenhang mit den Zeichnungen gesehen wird, wobei in den Zeichnungen gleiche Bezugszeichen Entsprechendes durchgängig identifizieren, und wobei:

[0013] [Fig. 1](#) ein Blockdiagramm der vorliegenden Erfindung ist.

Detaillierte Beschreibung der bevorzugten Ausführungsbeispiele

[0014] Bezugnehmend auf [Fig. 1](#) wird das Eingangssignal bzw. Eingabesignal $S(n)$ an ein Teilbandenergieberechnungselement **4** und ein Teilbandenergieberechnungselement **6** geliefert. Das Eingangssignal $S(n)$ besteht aus einem Audiosignal und Hintergrundrauschen. Das Audiosignal ist typischerweise Sprache, kann jedoch auch Musik sein. In dem Ausführungsbeispiel wird $S(n)$ in 20 Millisekundenrahmen von jeweils 160 Samples bzw. Abtastwerten vorgesehen. In dem Ausführungsbeispiel hat das Eingangssignal $S(n)$ Frequenzkomponenten von 0 kHz bis 4 kHz, was ungefähr die Bandbreite eines menschlichen Sprachsignals ist.

[0015] In dem Ausführungsbeispiel wird das 4 kHz Eingangssignal $S(n)$ auf zwei separate Teilbänder (subbands) gefiltert. Die zwei separaten Teilbänder liegen zwischen 0 und 2 kHz bzw. 2 kHz und 4 kHz. Bei einem Ausführungsbeispiel kann das Eingangssignal durch Teilbandfilter, deren Konstruktion im Fachgebiet bekannt ist, und z. B. detailliert in dem U.S. Patent 5,644,596, dem Rechtsnachfolger der vorliegenden Erfindung zugewiesen ist, dargestellt, auf Teilbänder aufgeteilt werden.

[0016] Die Impulsantworten der Teilbandfilter werden mit $h_L(n)$ für den Tiefpassfilter und $h_H(n)$ für den Hochpassfilter bezeichnet. Die Energie der resultierenden Teilbandkomponenten des Signals kann auch durch einfaches Summieren der Quadrate der Teilbandfilterausgabesamples berechnet werden, um die Werte $R_L(0)$ und $R_H(0)$ zu liefern, wie es auf dem Fachgebiet bekannt ist.

[0017] In dem bevorzugten Ausführungsbeispiel wird, wenn das Eingangssignal $S(n)$ an das Teilbandenergie-

berechnungselement 4 geliefert wird, der Energiewert der Niedrig- bzw. Tieffrequenzkomponente des Eingaberahmens $R_L(0)$, wie folgt berechnet:

$$R_L(0) = R_s(0) \cdot R_{h_L}(0) + 2 \cdot \sum_{i=1}^{L-1} R_s(i) \cdot R_{h_L}(i), \quad (1)$$

wobei L die Anzahl der Taps bzw. Abgriffe in dem Tiefpassfilter mit der Impulsantwort $h_L(n)$ ist, wobei $R_s(i)$ die Autokorrelationsfunktion des Eingangssignals $S(n)$ ist, und zwar gegeben durch die Gleichung:

$$R_s(i) = \sum_{n=1}^N S(n) \cdot S(n-i), \quad \text{für } i \in [0, L-1] \quad (2)$$

wobei N die Anzahl der Samples in dem Rahmen ist, und wobei R_{h_L} die Autokorrelationsfunktion des Tiefpassfilters $h_L(n)$ ist, und zwar gegeben durch:

$$R_{h_L}(i) = \sum_{n=0}^{L-1} h_L(n) \cdot h_L(n-i), \quad \text{für } i \in [0, L-1] \quad (3)$$

$$= 0, \quad \text{anderenfalls.}$$

[0018] Die Hochfrequenzenergie, $R_H(0)$ wird auf ähnliche Art und Weise in dem Teilbandenergieberechnungselement 6 berechnet.

[0019] Die Werte der Autokorrelationsfunktion der Teilbandfilter können zeitlich voraus berechnet werden, um die Rechenlast zu reduzieren. Weiterhin werden einige der berechneten Werte von $R_s(i)$ in anderen Berechnungen bei der Kodierung des Eingangssignals $S(n)$ verwendet, was weiterhin die effektive Berechnungslast des Kodierungsratenauswahlverfahrens der vorliegenden Erfindung reduziert. So benötigt z. B. die Herleitung der LPC-Filter-Tap-Werte die Berechnung eines Satzes von Eingangssignalauskorrelationskoeffizienten.

[0020] Die Berechnung von LPC-Filter-Tap-Werten ist auf dem Fachgebiet bekannt und wird in der oben zitierten WO-A1-92/22891 detailliert dargestellt. Würde man die Sprache mit einem Verfahren kodieren, das einen LPC-Filter mit zehn Taps benötigt, müssten nur die Werte von $R_s(i)$ für Werte für i zwischen 11 und $L-1$ berechnet werden, und zwar zusätzlich zu denen, die in der Kodierung des Signals eingesetzt werden, da die $R_s(i)$ für Werte von i zwischen 0 und 10 in der Berechnung der LPC-Filter-Tap-Werte verwendet werden. In dem Ausführungsbeispiel haben die Teilbandfilter 17 Taps, $L = 17$.

[0021] Das Teilbandenergieberechnungselement 4 liefert die berechneten Werte von $R_L(0)$ an das Teilbandratenentscheidungselement 12, und das Teilbandenergieberechnungselement 6 liefert die berechneten Werte von $R_H(0)$ an Teilbandratenentscheidungselement 14. Das Ratenentscheidungselement 12 vergleicht die Werte von $R_L(0)$ mit zwei vorbestimmten Schwellenwerten $T_{L1/2}$ und T_{Lfull} und weist eine vorgeschlagene Kodierate, $RATE_s$ zu, und zwar gemäß dem Vergleich. Die Ratenzuweisung wird wie folgt ausgeführt:

$$RATE_L = \text{Achtelrate} \quad R_L(0) \leq T_{L1/2} \quad (4)$$

$$RATE_L = \text{Halbrate} \quad T_{L1/2} < R_L(0) \leq T_{Lfull} \quad (5)$$

$$RATE_L = \text{Vollrate} \quad R_L(0) > T_{Lfull} \quad (6)$$

[0022] Teilbandratenentscheidungselement 14 operiert auf ähnliche Art und Weise und wählt eine vorgeschlagene Kodierate $RATE_N$ aus, und zwar gemäß dem Hochfrequenzenergiwert $R_H(0)$ und basierend auf einem unterschiedlichen Satz von Schwellenwerten $T_{H1/2}$ und T_{Hfull} . Das Teilbandratenentscheidungselement 12 liefert seine vorgeschlagene Kodierate, $RATE$, an Kodieratenauswahlelement 16, und Teilbandratenentscheidungselement 14 liefert seine vorgeschlagene Kodierate, $RATE_H$, an Kodieratenauswahlelement 16. In dem Ausführungsbeispiel wählt Kodieratenauswahlelement 16 die höhere der zwei vorgeschlagenen Raten und liefert die höhere Rate als die ENCODING RATE bzw. Kodierate.

[0023] Teilbandenergieberechnungselement 4 liefert außerdem den Tieffrequenzenergiwert $R_L(0)$ an das Schwellenwerteinstellelement 8, wo die Schwellenwerte $T_{L1/2}$ und T_{Lfull} für den nächsten Eingaberahmen berechnet werden. Ähnlich liefert Teilbandenergieberechnungselement 6 den Hochfrequenzenergiwert $R_H(0)$ an Schwellenwerteinstellelement 10, wo die Schwellenwerte $T_{H1/2}$ und T_{Hfull} für den nächsten Eingaberahmen be-

rechnet werden.

[0024] Schwellenwerteinstellelement **8** empfängt den Tieffrequenzenergiewert, $R_L(0)$, und berechnet ob $S(n)$ Hintergrundrauschen oder ein Audiosignal enthält. In einer beispielhaften Implementierung geschieht das Verfahren, durch welches das Schwellenwerteinstellelement **8** bestimmt, ob ein Audiosignal vorliegt, durch Untersuchen der normalisierten Autokorrelationsfunktion (normalized autocorrelation function NACF), die durch die folgende Gleichung gegeben ist

$$\text{NACF} = \max_T \frac{\sum_{n=0}^{N-1} e(n) \cdot e(n-T)}{\frac{1}{2} \cdot \left[\sum_{n=0}^{N-1} e^2(n) + \sum_{n=0}^{N-1} e^2(n-T) \right]}, \quad (7)$$

wobei $e(n)$ das Formant-Restsignal ist, das vom Filtern des Eingangssignals $S(n)$, durch einen LPC-Filter resultiert.

[0025] Die Konstruktion eines LPC-Filters, sowie das Filtern eines Signals durch einen LPC-Filter ist auf dem Fachgebiet bekannt und ist in der zuvor erwähnten WO-A1-92/22891 detailliert dargestellt. Das Eingangssignal, $S(n)$ wird durch den LPC-Filter gefiltert, um Wechselwirkungen der Formanten zu entfernen. Die NACF wird mit einem Schwellenwert verglichen, um zu bestimmen, ob ein Audiosignal vorliegt. Wenn die NACF größer ist als ein vorbestimmter Schwellenwert, zeigt dies an, dass der Eingaberahmen eine periodische Charakteristik hat, die anzeigend ist für das Vorliegen eines Audiosignals, wie z. B. Sprache oder Musik. Es ist anzumerken, dass während Teile von Sprache und Musik nicht periodisch sind und niedrige Werte für NACF aufzeigen, Hintergrundrauschen typischerweise niemals eine Periodizität aufzeigt und fast immer niedrige Werte für NACF aufweist.

[0026] Wenn bestimmt ist, dass $S(n)$ Hintergrundrauschen enthält, ist der Wert von NACF geringer als ein Schwellenwert TH1, wobei dann der Wert $R_L(0)$ eingesetzt wird, den Wert der momentanen Hintergrundrauschschätzung BGN_L zu aktualisieren. In dem Ausführungsbeispiel ist TH1 0,35. $R_L(0)$ wird mit dem momentanen Wert der Hintergrundrauschschätzung BGN_L verglichen. Wenn $R_L(0)$ kleiner ist als BGN_L , dann wird die Hintergrundrauschschätzung BGN_L gleich $R_L(0)$ gesetzt, unabhängig von dem Wert von NACF.

[0027] Die Hintergrundrauschschätzung BGN_L wird nur erhöht, wenn NACF kleiner ist der Schwellenwert TH1. Wenn $R_L(0)$ größer ist als BGN_L und NACF kleiner ist als TH1, dann wird die Hintergrundrauschschätzung BGN_L auf $\alpha_1 BGN_L$ gesetzt, wobei α_1 eine Zahl größer als 1 ist. In dem Ausführungsbeispiel ist α_1 gleich 1,03. Die BGN_L wird solange fortfahren sich zu erhöhen, solange NACF kleiner ist als der Schwellenwert TH1 und $R_L(0)$ größer ist als der momentane Wert von BGN_L , und zwar bis BGN_L einen vorbestimmten Maximalwert BGN_{\max} erreicht, wobei bei diesem Punkt die Hintergrundrauschschätzung BGN_L auf BGN_{\max} gesetzt wird.

[0028] Wenn ein Audiosignal detektiert wird, was dadurch gekennzeichnet ist, dass der Wert von NACF einen zweiten Schwellenwert TH2 überschreitet, dann wird die Signalenergieschätzung, S_L , aktualisiert. In dem Ausführungsbeispiel ist TH2 auf 0,5 gesetzt. Der Wert von $R_L(0)$ wird mit einer momentanen Tiefpasssignalenergieschätzung S_L verglichen. Wenn $R_L(0)$ größer ist als der momentane Wert von S_L , dann wird S_L gleich $R_L(0)$ gesetzt. Wenn $R_L(0)$ kleiner ist als der momentane Wert von S_L , dann wird S_L gleich $\alpha_2 \cdot S_L$ gesetzt, und zwar nur dann, wenn NACF größer ist als TH2. In dem beispielhaften Ausführungsbeispiel ist α_2 auf 0,96 gesetzt.

[0029] Das Schwellenwerteinstellelement **8** berechnet dann eine Signal-zu-Rausch-Verhältnisschätzung gemäß der folgenden Gleichung 8:

$$\text{SNR}_L = 10 \cdot \log \left[\frac{S_L}{BGN_L} \right], \quad (8)$$

[0030] Das Schwellenwerteinstellelement **8** bestimmt dann einen Index des quantisierten Signal-zu-Rausch-Verhältnisses I_{SNRL} gemäß der folgenden Gleichung 9 bis 12:

$$I_{\text{SNRL}} = \text{nint} \left[\frac{\text{SNR}_L - 20}{5} \right], \quad \text{für } 20 < \text{SNR}_L < 55, \quad (9)$$

$$= 0, \quad \text{für } \text{SNR}_L \leq 20 \quad (10)$$

$$= 7 \quad \text{für } \text{SNR}_L \geq 55$$

wobei nint eine Funktion ist, die den Bruchwert auf den nächsten integer rundet.

[0031] Schwellenwerteinstellelement **8** wählt dann zwei Skalierungsfaktoren, $k_{L1/2}$ und k_{Lfull} aus oder berechnet diese, und zwar gemäß dem Signal-zu-Rausch-Index, I_{SNRL} . Eine Beispielnachschlagetabelle für Skalierungswerte wird in der folgenden Tabelle 1 angegeben.

Tabelle 1

I_{SNRL}	$K_{L1/2}$	K_{Lfull}
0	7,0	9,0
1	7,0	12,6
2	8,0	17,0
3	8,6	18,5
4	8,9	19,4
5	9,4	20,9
6	11,0	25,5
7	15,8	39,8

[0032] Diese zwei Werte werden eingesetzt, um die Schwellenwerte für die Ratenauswahl gemäß der folgenden Gleichungen zu verwenden.

$$T_{L1/2} = K_{L1/2} \cdot \text{BGN}_L, \quad \text{und} \quad (11)$$

$$T_{Lfull} = K_{Lfull} \cdot \text{BGN}_L, \quad (12)$$

wobei $T_{L1/2}$ der Tieffrequenz-Halbraten-Schwellenwert ist und T_{Lfull} der Tieffrequenz-Vollraten-Schwellenwert ist.

[0033] Das Schwellenwerteinstellelement **8** bzw. das Schwellenwertanpasselement **8** liefert die angepassten Schwellenwerte $T_{L1/2}$ und T_{Lfull} an das Ratenentscheidungselement **12**. Das Schwellenwerteinstellelement **10** operiert auf ähnliche Art und Weise und liefert die Schwellenwerte $T_{H1/2}$ und T_{Lfull} an das Teilbandratenentscheidungselement **14**.

[0034] Der Anfangswert der Audiosignalenergieschätzung S , wobei S für S_L oder S_H steht, wird, wie folgt, eingestellt: Die anfängliche Signalenergieschätzung S_{INIT} wird auf $-18,0$ dBmO gesetzt, wobei $3,17$ dBmO die Signalstärke einer ganzen Sinuswelle bezeichnet, wobei die Sinuswelle in dem Ausführungsbeispiel eine digitale Sinuswelle mit einem Amplitudenbereich von -8031 bis 8031 ist. S_{INIT} wird verwendet, bis bestimmt wird, dass ein akustisches Signal vorliegt.

[0035] Das Verfahren, mit dem ein akustisches Signal anfänglich detektiert wird, geschieht durch Vergleichen des NACF-Wertes mit einem Schwellenwert, wobei dann ein akustisches Signal als vorliegend bestimmt wird, wenn die NACF den Schwellenwert für eine vorbestimmte Anzahl von aufeinanderfolgenden Rahmen überschreitet. In dem Ausführungsbeispiel muss die NACF den Schwellenwert zehn aufeinanderfolgende Rahmen lang überschreiten. Nachdem diese Bedingung erfüllt ist, wird die Signalenergieschätzung, S , auf die maximale Signalenergie in den vorhergehenden zehn Rahmen gesetzt.

[0036] Der anfängliche Wert der Hintergrundrauschschätzung BGN_L wird anfänglich auf BGN_{max} gesetzt. Sobald eine Teilbandrahmenenergie empfangen wird, die geringer ist als BGN_{max} , wird die Hintergrundrausch-

schätzung auf den Wert des empfangenden Teilbandenergiepegels gesetzt, und die Generierung der Hintergrundrausch-BGN_{init}-Schätzung fährt, wie zuvor beschrieben, fort.

[0037] In einem bevorzugten Ausführungsbeispiel wird ein Hangover- bzw. Überhangzustand betätigt, wenn nach einer Serie von Sprachrahmen mit voller Rate ein Rahmen mit niedrigerer Rate detektiert wird. In dem Ausführungsbeispiel wird, wenn vier aufeinanderfolgende Sprachrahmen mit Vollrate kodiert werden, gefolgt von einem Rahmen, bei dem die Rate auf eine niedrigere als die Vollrate gesetzt ist, und die berechneten Signal-zu-Rausch-Verhältnisse geringer sind als ein vorbestimmtes Minimum SNR, die ENCODING RATE für diesen Rahmen auf Vollrate gesetzt. In dem Ausführungsbeispiel ist das vordefinierte Minimal-SNR 27,5 dB gemäß der Definition in Gleichung 8.

[0038] In dem bevorzugten Ausführungsbeispiel ist die Zahl der Überhangrahmen eine Funktion des Signal-zu-Rausch-Verhältnisses. In dem Ausführungsbeispiel wird die Zahl der Überhangrahmen wie folgt bestimmt:

$$\# \text{Hangover-Rahmen} = 1 \quad 22,5 < \text{SNR} < 27,5 \quad (13)$$

$$\# \text{Hangover-Rahmen} = 2 \quad \text{SNR} \leq 22,5 \quad (14)$$

$$\# \text{Hangover-Rahmen} = 0 \quad \text{SNR} \geq 27,5 \quad (15)$$

[0039] Die vorliegende Erfindung liefert außerdem ein Verfahren, mit dem das Vorliegen von Musik detektiert werden kann, bei der, wie zuvor beschrieben, Pausen fehlen, die es ermöglichen, die Hintergrundrauschmessungen zurückzusetzen. Das Verfahren zum Detektieren des Vorliegens von Musik geht davon aus, dass Musik zu Beginn des Anrufes nicht vorliegt. Dies erlaubt es der Kodieratenauswahlvorrichtung der vorliegenden Erfindung eine anfängliche Hintergrundrauschenergie, BGN_{init} genau zu schätzen. Da Musik im Gegensatz zu Hintergrundrauschen eine periodische Charakteristik hat, untersucht die vorliegende Erfindung den Wert der NACF, um Musik von Hintergrundrauschen zu unterscheiden. Das Musikdetektierverfahren der vorliegenden Erfindung berechnet eine durchschnittliche NACF gemäß der folgenden Gleichung:

$$\text{NACF}_{\text{AVE}} = \frac{1}{T} \sum_{i=1}^T \text{NACF}(i), \quad (16)$$

wobei NACF in Gleichung 7 definiert ist, und

wobei T die Anzahl von aufeinanderfolgenden Rahmen ist, in denen sich der geschätzte Wert des Hintergrundrauschens, ausgehend von einer anfänglichen Hintergrundrauschschätzung BGN_{init} erhöht hat.

[0040] Wenn sich das Hintergrundrauschen BGN eine vorbestimmte Anzahl von Rahmen von T lang erhöht hat und NACF_{AVE} einen vorbestimmten Schwellenwert überschreitet, wird Musik detektiert und das Hintergrundrauschen BGN wird auf BGN_{init} zurückgesetzt. Es ist anzumerken, dass für eine gewisse Effektivität der Wert T niedrig genug gesetzt werden muss, so dass die Kodierungsrate nicht unter die Vollrate fällt. Daher sollte der Wert von T als eine Funktion des akustischen Signals und BGN_{init} gesetzt werden.

[0041] Die vorliegende Beschreibung der bevorzugten Ausführungsbeispiele wurde vorgesehen, um es einem Fachmann zu ermöglichen, die vorliegende Erfindung herzustellen oder zu verwenden. Die verschiedenen Modifikationen dieses Ausführungsbeispiels werden einem Fachmann leicht offensichtlich werden, und die Grundprinzipien, die in den Ausführungsbeispielen definiert sind, können auf andere Ausführungsbeispiele, ohne den Einsatz einer erfinderischen Tätigkeit, angewendet werden. Daher ist die vorliegende Erfindung nicht als durch die Ausführungsbeispiele eingeschränkt anzusehen, sondern ihr ist ein Schutzbereich, wie er nur durch die beigefügten Ansprüche definiert ist, zuzuordnen.

Patentansprüche

1. Ein Verfahren zum Detektieren, ob ein Rahmen eines Eingabesignals ein Audiosignal oder Stille besitzt bzw. wiedergibt, wobei das Verfahren Folgendes aufweist:

Setzen von Detektierungsschwellen basierend auf einer Schätzung eines Signal-zu-Rausch-Verhältnisses (SNR = signal to noise ratio) des Eingabesignals, wobei die Signalenergie des SNRs als eine maximale Signalenergie während einer Zeit von aktiver Sprache geschätzt wird; und

Verwenden der Detektierungsschwellen, um zu detektieren, ob der Rahmen des Eingabesignals ein Audiosignal oder Stille aufweist.

2. Verfahren nach Anspruch 1, wobei die Hintergrundrauschenergie des SNRs geschätzt wird als die minimale Signalenergie während einer Zeit der Stille.
3. Verfahren nach Anspruch 1, wobei der Schritt des Einstellens von Detektierungsschwellen Folgendes aufweist:
Bestimmen eines Index des SNRs des Eingabesignals;
Verwenden des Index des SNRs, um einen ersten Skalierungsfaktor und einen zweiten Faktor auszuwählen oder zu berechnen;
Verwenden des ersten Skalierungsfaktors und des zweiten Skalierungsfaktors, um einen Niedrigfrequenzschwellenwert und einen Hochfrequenzschwellenwert zu berechnen.
4. Verfahren nach Anspruch 1, wobei der Schritt des Verwendens der Detektierungsschwelle zum Detektieren, ob der Rahmen des Eingabesignals ein Audiosignal oder Stille aufweist, Folgendes aufweist:
Filtern des Eingabesignals mit einem linear prädiktiv Codierungs- bzw. LPC-Filter (LPC = linear predictive coding), um ein Formant-Restsignal zu erhalten; und
Vergleichen einer normalisierten Autokorrelationsfunktion des Formant-Restsignals mit den Detektierungsschwellen.
5. Verfahren nach Ansprüchen 2 und 4, wobei das Vergleichen der normalisierten Autokorrelationsfunktion des Formant-Restsignals mit den Detektierungsschwellen Folgendes aufweist:
Vergleichen der normalisierten Autokorrelationsfunktion des Formant-Restsignals mit einer ersten Schwelle;
Aktualisieren der Hintergrundrauschenergie-Schätzung, wenn die normalisierte Autokorrelationsfunktion des Formant-Restsignals geringer ist als die erste Schwelle;
Vergleichen der normalisierten Autokorrelationsfunktion des Formant-Restsignals mit einer zweiten Schwelle, wobei die zweite Schwelle höher ist als die erste Schwelle;
Aktualisieren der Signalenergie-Schätzung, wenn die normalisierte Autokorrelationsfunktion des Formant-Restsignals größer ist als die zweite Schwelle; und
Verwenden der aktualisierten Hintergrundrauschenergie-Schätzung und der aktualisierten Signalenergieschätzung, um zu bestimmen, ob das Eingabesignal ein Audiosignal oder Stille aufweist.
6. Vorrichtung zum Detektieren, ob ein Rahmen eines Eingabesignals ein Audiosignal oder Stille besitzt bzw. aufweist, wobei die Vorrichtung Folgendes aufweist:
Mittel zum Einstellen von Detektierungsschwellen basierend auf einer Schätzung eines Signal-zu-Rausch-Verhältnisses bzw. SNRs ($SNR = \text{signal-to-noise-ratio}$) des Eingabesignals, wobei die Signalenergie des SNRs geschätzt wird als eine maximale Signalenergie während einer Zeit von aktiver Sprache; und
Mittel zum Verwenden der Detektierungsschwellen, um zu detektieren, ob der Rahmen des Eingabesignals ein Audiosignal oder Stille aufweist.
7. Vorrichtung nach Anspruch 6, wobei die Hintergrundrauschenergie des SNRs geschätzt wird als die minimale Signalenergie während einer Zeit der Stille.
8. Vorrichtung nach Anspruch 6, wobei die Mittel zum Einstellen der Detektierungsschwellen weiterhin konfiguriert sind zum:
Bestimmen eines Index des SNRs des Eingabesignals;
Verwenden des Index des SNRs, um einen ersten Skalierungsfaktor und einen zweiten Faktor zu berechnen oder auszuwählen;
Verwenden des ersten Skalierungsfaktors und des zweiten Skalierungsfaktors, um einen Niedrigfrequenzschwellenwert und einen Hochfrequenzschwellenwert zu berechnen.
9. Vorrichtung nach Anspruch 6, wobei die Mittel zum Verwenden der Detektierungsschwellen zum Detektieren, ob der Rahmen des Eingabesignals ein Audiosignal oder Stille aufweist weiterhin konfiguriert sind zum Filtern des Eingabesignals mit einem linear prädiktiven Codierungs- bzw. LPC-Filter, um ein Formant-Restsignal zu erhalten; und Vergleichen einer normalisierten Autokorrelationsfunktion des Formant-Restsignals mit den Detektierungsschwellen.
10. Vorrichtung nach Anspruch 9, wobei die Mittel zum Vergleichen der normalisierten Autokorrelationfunktion des Formant-Restsignals mit den Detektierungsschwellen weiterhin konfiguriert sind zum:
Vergleichen der normalisierten Autokorrelationfunktion des Formant-Restsignals mit einer ersten Schwelle;
Aktualisieren der Hintergrundrauschenergie-Schätzung, wenn die normalisierte Autokorrelationfunktion des Formant-Restsignals geringer ist als die erste Schwelle;

Vergleichen der normalisierten Autokorrelationsfunktion des Formant-Restsignals mit einer zweiten Schwelle, wobei die zweite Schwelle höher ist als die erste Schwelle;
Aktualisieren der Signalenergieschätzung, wenn die normalisierte Autokorrelationsfunktion des Formant-Restsignals größer ist als die zweite Schwelle; und
Verwenden der aktualisierten Hintergrundrauschenergie-Schätzung und der aktualisierten Signalenergieschätzung, um zu bestimmen, ob das Eingabesignal ein Audiosignal oder Stille aufweist.

11. Eine Vorrichtung zum Bestimmen einer Codierrate für einen Vocoder mit variabler Rate, wobei die Vorrichtung Folgendes aufweist:

Unterbandenergie-Berechnungsmittel zum Empfangen eines Eingabesignals und zum Bestimmen einer Vielzahl von Unterbandenergiewerten gemäß einem vorbestimmten Unterbandenergie-Berechnungsformat;

Ratenbestimmungsmittel zum Empfangen der Vielzahl von Unterbandenergiewerten und zum Bestimmen der Codierrate gemäß der Vielzahl von Unterbandenergiewerten;

Schwellenberechnungsmittel angeordnet zwischen den Unterbandenergieberechnungsmitteln und den Ratenbestimmungsmitteln zum Empfangen der Unterbandenergiewerte und zum Bestimmen eines Satzes von Codierratenschwellenwerten gemäß einer Vielzahl von Unterbandenergiewerten, wobei die Schwellenberechnungsmittel ein Signal-zu-Rausch-Verhältniswert bestimmen gemäß der Vielzahl von Unterbandenergiewerten.

Es folgt ein Blatt Zeichnungen

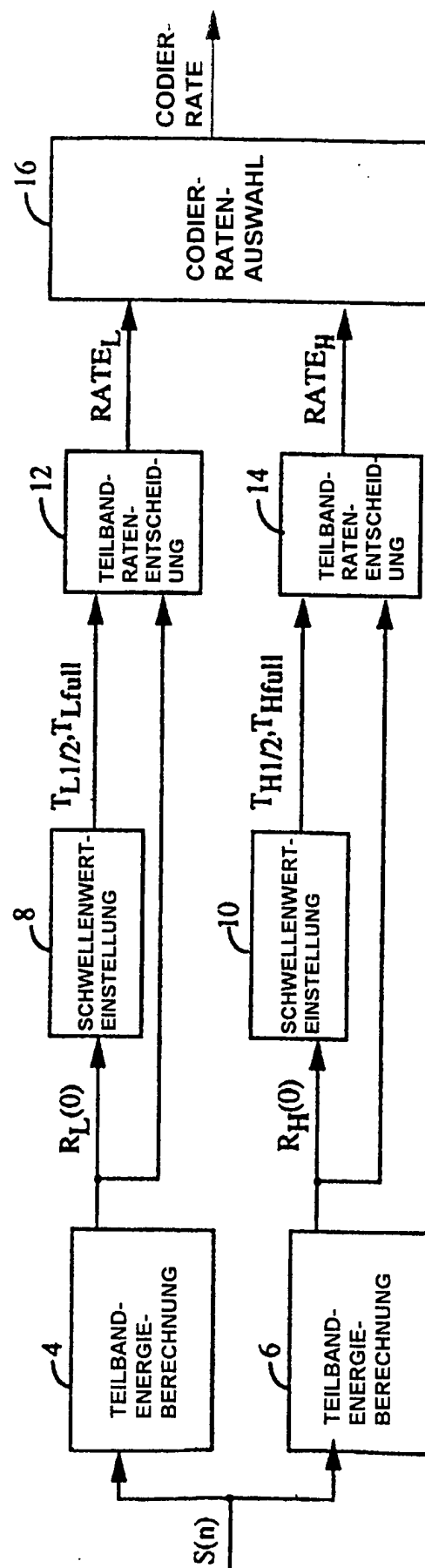


FIG. 1