



(10) **DE 695 34 285 T3** 2010.09.09

(12) **Übersetzung der geänderten europäischen Patentschrift**

(97) **EP 1 239 465 B2**

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

(21) Deutsches Aktenzeichen: **695 34 285.1**

(96) Europäisches Aktenzeichen: **02 009 467.8**

(96) Europäischer Anmeldetag: **01.08.1995**

(97) Erstveröffentlichung durch das EPA: **11.09.2002**

(97) Veröffentlichungstag

der Patenterteilung beim EPA: **15.06.2005**

(97) Veröffentlichungstag

des geänderten Patents beim EPA: **17.02.2010**

(47) Veröffentlichungstag im Patentblatt: **09.09.2010**

Patentschrift wurde im Einspruchsverfahren geändert

(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., San Diego, CA 92126, US;
Gardner, William R., San Diego, CA 92130, US**

(54) Bezeichnung: **Verfahren und Vorrichtung zur Auswahl der Kodierate in einem Vocoder mit variabler Rate**

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 zum Hinzufügen von Hangover bzw. Überhang-Rahmen.

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 dem ebenfalls anhängigen U.S. Patent 5,414,796, eingereicht am 11. Juli 1991, betitelt „Variable rate vocoder“ und dem Rechtsnachfolger der vorliegenden Erfindung zugewiesen, offenbart. In dieser bestimmten Implementierung eines Vocoder 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 unstimmhafte Töne werden häufig mit niedrigeren Bitraten codiert, und die Sprachqualität verschlechtert sich, da Konsonanten, wie z. B. „s“, „x“, „ch“, „sh“, „f“, 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] Weiterhin wird auch auf das Dokument betitelt „Voice Activity Detection For Cellular Networks“ von K. Srinivasan und A. Gersho, Proceedings: IEEE Workshop an Speech Coding for Telecommunications, 13–15. Oktober 1993, Seiten 85–86, XP002204645, University of California aufmerksam gemacht. Das Dokument dis-

kutiert Algorithmen zur Sprachaktivitätsdetektierung wenn Fahrzeugrauschen und sprachähnliches Rauschen vorhanden ist. Insbesondere offenbart das Dokument einen Sprachaktivitätsdetektierungsalgorithmus, in dem eine adaptive Hangover- bzw. Überhangperiode, die zwischen 40 ms und 180 ms variiert, eingeführt wird. Die tatsächliche Überhangperiode basiert auf dem Verhältnis r der Rauschunterdrückungsfilterausgangsleistung zu der entsprechenden adaptiven Schwelle.

[0008] Weiterhin wird auf das Dokument Paksoy E et al: "Variable rate speech coding for multiple access wireless networks", Electrotechnical Conference, 1994, Proceedings, 7th Mediterranean Antalya, Türkei 12.–14. April 1994, New York, NY, USA, IEEE, 12, April 1994, Seiten 47–50, XP10130866 ISBN: 0-7803-1772-6 aufmerksam gemacht, das Sprachkodierung mit variabler Rate für Mehrfachzugriffsdrahtlosnetzwerke diskutiert und das insbesondere eine Sprachaktivitätsdetektierung mit einer Adaption der Überhangperiode auf die detektierten Signalpegel hin erwähnt.

[0009] Weiterhin wird verwiesen auf die WO-A1-93/13516, die eine Berechnung einer CAD Überhangzeit und der Verwendung eines SNR und auf die Recommendation GSM 06,32, "VOICE ACTIVITY DETECTION"; Februar 1992, die ein VAD Überhang-Hinzufügen zu Sprach-Bursts, die eine Gewisse Länge überschreiten, offenbart.

Zusammenfassung der Erfindung

[0010] Gemäß der vorliegenden Erfindung wird ein Verfahren und eine Vorrichtung zum Hinzufügen von Überhangrahmen zu einer Vielzahl von Rahmen, die von einem Sprachdekodierer bzw. Vocoder kodiert werden, gemäß Ansprüchen 1 und 3 vorgesehen.

[0011] Die vorliegende Beschreibung beschreibt ein neues und verbessertes Verfahren und eine Vorrichtung zum Bestimmen einer Kodierrate in einem Vocoder mit variabler Rate. Es ist ein erstes Ziel ein Verfahren vorzusehen, bei dem die Wahrscheinlichkeit des Kodierens von unstimmhafter Sprache mit niedriger Energie als Hintergrundrauschen reduziert wird. Das Eingangssignal wird 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.

[0012] Ein zweites Ziel ist es, ein Mittel vorzusehen, mit dem die Schwellenwertpegel eingestellt werden, und das Signalenergie sowie Hintergrundrauschenergie berücksichtigt. Das Einstellen der Sprachdetektierschwellenwerte basiert auf einer Schätzung des Signal-zu-Rausch-Verhältnisses (signal to noise ratio (SNR)) des Eingangssignals. In dem Beispiel wird die Signalenergie als die Maximalsignalenergie während Zeiten von aktiver Sprache geschätzt, und die Hintergrundrauschenergie wird als die minimale Signalenergie während Pausenzeiten geschätzt.

[0013] Ein drittes Ziel ist es, ein Verfahren zum Kodieren von Musik, die durch einen Vocoder mit variabler Rate tritt, vorzusehen. In dem Beispiel 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

[0014] 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:

[0015] [Fig. 1](#) ein Blockdiagramm der vorliegenden Erfindung ist.

Detaillierte Beschreibung der bevorzugten Ausführungsbeispiele

[0016] Bezugnehmend auf [Fig. 1](#) wird das Eingangssignal bzw. Eingabesignal $S(n)$ an ein Teilbandenergieberechnungselement **4** und ein Teilbandenergieberechnungselement **6** geliefert. Das Eingangssignal $S(n)$ be-

RATE_L = Achtelrate
 RATE_L = Halbrate
 RATE_L = Vollrate

$$R_L(0) \leq T_{L1/2} \quad (4)$$

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

$$R_L(0) > T_{Lfull} \quad (6)$$

[0024] Teilbandratenentscheidungselement **14** operiert auf ähnliche Art und Weise und wählt eine vorgeschlagene Kodierrate RATE_N aus, und zwar gemäß dem Hochfrequenzenergiewert R_H(0) und basierend auf einem unterschiedlichen Satz von Schwellenwerten T_{H1/2} und T_{Hfull}. Das Teilbandratenentscheidungselement **12** liefert seine vorgeschlagene Kodierrate, RATE, an Kodierratenauswahlelement **16**, und Teilbandratenentscheidungselement **14** liefert seine vorgeschlagene Kodierrate, RATE_N, an Kodierratenauswahlelement **16**. In dem Ausführungsbeispiel wählt Kodierratenauswahlelement **16** die höhere der zwei vorgeschlagenen Raten und liefert die höhere Rate als die ENCODING RATE bzw. Kodierrate.

[0025] Teilbandenergieberechnungselement **4** liefert außerdem den Tieffrequenzenergiewert 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 Hochfrequenzenergiewert R_H(0) an Schwellenwerteinstellelement **10**, wo die Schwellenwerte T_{L1/2} und T_{Lfull} für den nächsten Eingaberahmen berechnet werden.

[0026] 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

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

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

[0027] Die Konstruktion eines LPC-Filters, sowie das Filtern eines Signals durch einen LPC-Filter ist auf dem Fachgebiet bekannt und ist in dem zuvor erwähnten U.S. Patent 5,414,796 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 aufweist und fast immer niedrige Werte für NACF aufweist.

[0028] 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.

[0029] 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 α₁BGN_L gesetzt, wobei α₁ eine Zahl größer als 1 ist. In dem Ausführungsbeispiel ist α₁ 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.

[0030] 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.

[0031] Das Schwellenwertelement **8** berechnet dann eine Signal-zu-Rausch-Verhältnisschätzung gemäß der folgenden Gleichung 8:

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

[0032] Das Schwellenwertelement **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.

[0033] Schwellenwertelement **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

[0034] Diese zwei Werte werden eingesetzt, um die Schwellenwerte für die Ratenwahl 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.

[0035] Das Schwellenwertelement **8** bzw. das Schwellenwertanpasselement **8** liefert die angepassten Schwellenwerte $T_{L1/2}$ und T_{Lfull} an das Ratenentscheidungselement **12**. Das Schwellenwertelement **10** operiert auf ähnliche Art und Weise und liefert die Schwellenwerte $T_{H1/2}$ und T_{Hfull} an das Teilbandratenentscheidungselement **14**.

[0036] Der Anfangswert der Audiosignalenergieschätzung S , wobei S für S_L oder S_H steht, wird, wie folgt, ein-

gestellt: 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.

[0037] 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.

[0038] 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 Hintergrundrauschschätzung auf den Wert des empfangenden Teilbandenergiepegels gesetzt, und die Generierung der Hintergrundrausch-BGN_L-Schätzung fährt, wie zuvor beschrieben, fort.

[0039] Gemäß der vorliegenden Erfindung 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 $27,5$ dBas, gemäß der Definition in Gleichung 8.

[0040] 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:

#Hangover-Rahmen = 1	$22,5 < \text{SNR} < 27,5$ (13)
#Hangover-Rahmen = 2	$\text{SNR} \leq 22,5$ (14)
#Hangover-Rahmen = 0	$\text{SNR} \geq 27,5$ (15)

[0041] Die vorliegende Beschreibung 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 eine anfängliche Hintergrundrauschenergie, BGN_{init} genau zu schätzen. Da Musik im Gegensatz zu Hintergrundrauschen eine periodische Charakteristik hat, wird der Wert der NACF untersucht, um Musik von Hintergrundrauschen zu unterscheiden. Das Musikdetektierverfahren 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.

[0042] 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.

[0043] 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 durch die beigefügten Ansprüche definiert ist, zuzuordnen.

Patentansprüche

1. Verfahren zur Addieren von Überhangrahmen zu einer Vielzahl von Rahmen codiert durch einen Vocoder, wobei das Verfahren folgendes vorsieht:

Detektieren, dass eine vorbestimmte Anzahl von aufeinander folgenden Rahmen mit einer Vollrate codiert ist; Bestimmen, dass ein nächst folgender Rahmen mit einer Rate aus einer Vielzahl von Raten, die kleiner sind als die Vollrate, codiert werden soll; und

Auswahl einer Anzahl von aufeinander folgenden Überhangrahmen beginnend mit dem erwähnten nächst folgenden Rahmen, der codiert werden soll mit der Rate aus der Vielzahl von Raten, die kleiner sind als die Vollrate, wobei die Zahl eine Funktion eines Signal-zu-Rausch-Verhältnisses ist, und zwar bestimmt aus dem Eingangssignal $S(n)$, das zu codieren ist.

2. Verfahren nach Anspruch 1, wobei das Detektieren folgendes aufweist:

Detektieren, dass eine vordefinierte Anzahl von aufeinander folgenden Rahmen mit der Vollrate codiert ist, die für Sprachrahmen gedacht ist.

3. Eine Vorrichtung zum Hinzuaddieren von Überhangrahmen zu einer Vielzahl von Rahmen codiert durch einen Vocoder, wobei die Vorrichtung folgendes aufweist:

Mittel zum Detektieren, dass eine vorbestimmte Anzahl von aufeinander folgenden Rahmen, die mit der Vollrate codiert ist;

Mittel zur Bestimmung, dass ein nächst darauf folgender Rahmen mit einer Rate aus einer Vielzahl von Raten, die kleiner sind als die Vollrate, codiert werden soll;

Mittel zur Auswahl einer Anzahl von aufeinander folgenden Überhangrahmen beginnend mit dem erwähnten nächst folgenden Rahmen, der mit der erwähnten zweiten Rate aus der Vielzahl von Raten, die kleiner sind als die Vollrate, codiert werden soll, wobei die Anzahl eine Funktion eines Signal-zu-Rausch-Verhältnisses ist, und zwar bestimmt aus dem Eingangssignal $S(n)$, das codiert werden soll.

4. Vorrichtung nach Anspruch 3, wobei die Mittel zum Detektieren folgendes aufweisen:

Mittel zum Detektieren, dass eine vorgeschriebene bzw. vordefinierte Anzahl von aufeinander folgenden Rahmen mit der Vollrate codiert wurde, vorgesehen für die Codierung von Sprachrahmen.

Es folgt ein Blatt Zeichnungen

Anhängende Zeichnungen

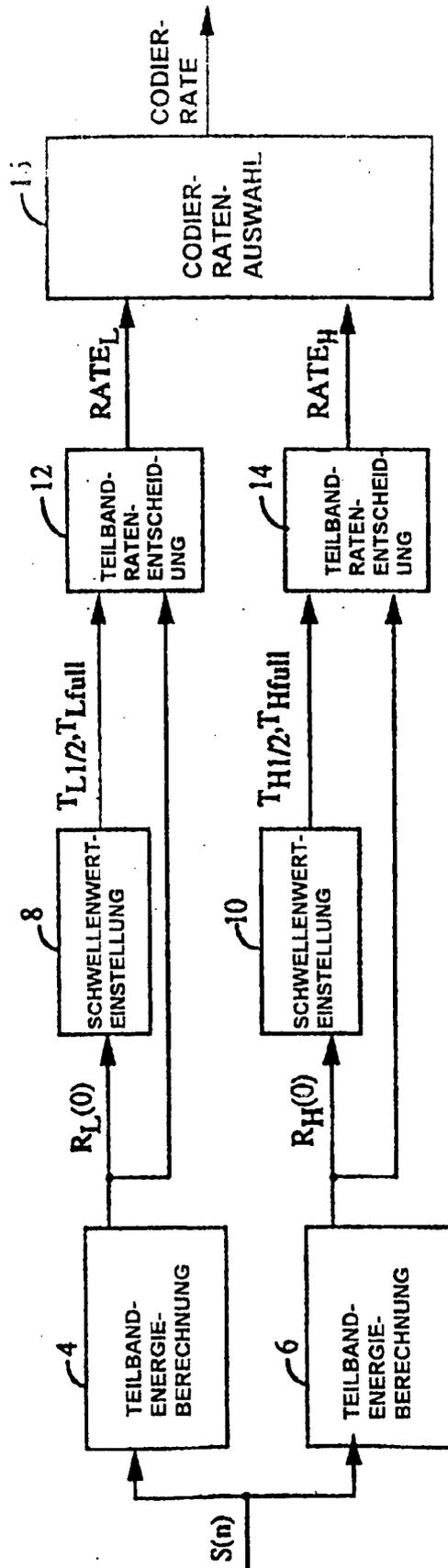


FIG. 1