



(11) **EP 1 279 168 B1**

(12) **EUROPÄISCHE PATENTSCHRIFT**

(45) Veröffentlichungstag und Bekanntmachung des Hinweises auf die Patenterteilung:
25.07.2007 Patentblatt 2007/30

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

(21) Anmeldenummer: **01911752.2**

(86) Internationale Anmeldenummer:
PCT/EP2001/002603

(22) Anmeldetag: **08.03.2001**

(87) Internationale Veröffentlichungsnummer:
WO 2001/084541 (08.11.2001 Gazette 2001/45)

(54) **VERFAHREN ZUR VERBESSERUNG DER SPRACHQUALITÄT BEI SPRACHÜBERTRAGUNGSAUFGABEN**

METHOD FOR IMPROVING SPEECH QUALITY IN SPEECH TRANSMISSION TASKS

PROCEDE POUR AMELIORER LA QUALITE SONORE DE TRANSMISSION DE LA PAROLE

(84) Benannte Vertragsstaaten:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR

• **ERDMANN, Christoph**
52062 Aachen (DE)

(30) Priorität: **28.04.2000 DE 10020863**

(56) Entgegenhaltungen:
WO-A-00/13174 US-A- 4 133 976

(43) Veröffentlichungstag der Anmeldung:
29.01.2003 Patentblatt 2003/05

• **HAGEN R ET AL: "AN 8 KBIT/S ACELP CODER WITH IMPROVED BACKGROUND NOISE PERFORMANCE" PHOENIX, AZ, MARCH 15 - 19, 1999, NEW YORK, NY: IEEE, US, 15. März 1999 (1999-03-15), Seiten 25-28, XP000898256 ISBN: 0-7803-5042-1**

(73) Patentinhaber: **Deutsche Telekom AG**
53113 Bonn (DE)

(72) Erfinder:
• **FISCHER, Kyrill, Alexander**
64347 Griesheim (DE)

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).

EP 1 279 168 B1

Beschreibung

[0001] Die vorliegende Erfindung betrifft ein Verfahren gemäß des Oberbegriffs des Anspruchs 1.

[0002] Im Bereich der Sprachübertragung und im Bereich der digitalen Signal- und Sprachspeicherung ist die Anwendung spezieller digitaler Codierungsverfahren zu Datenkompressionszwecken weit verbreitet und aufgrund der hohen Datenaufkommen sowie der begrenzten Übertragungskapazitäten zwingend notwendig. Ein für die Übertragung von Sprache besonders geeignetes Verfahren ist das aus der US 4133976 bekannte Code Excited Linear Prediction (CELP)-Verfahren. Bei diesem Verfahren wird das Sprachsignal in kleinen zeitlichen Abschnitten ("Sprachrahmen", "Rahmen", "zeitlicher Ausschnitt", "zeitlicher Abschnitt") von jeweils ca. 5 ms bis 50 ms Länge codiert und übertragen. Jeder dieser zeitlichen Abschnitte wird nicht exakt, sondern nur durch eine Annäherung an die tatsächliche Signalfom dargestellt. Die den Signalabschnitt beschreibende Approximation wird dabei im wesentlichen aus drei Komponenten gewonnen, die Decoder-Seitig zur Rekonstruktion des Signals verwendet werden: Erstens einem Filter, das die spektrale Struktur des jeweiligen Signalausschnittes annähernd beschreibt, zweitens einem sog. Anregungssignal, das durch dieses Filter gefiltert wird, und drittens einem Verstärkungsfaktor ("gain"), mit dem das Anregungssignal vor der Filterung multipliziert wird. Der Verstärkungsfaktor ist für die Lautstärke des jeweiligen Abschnitts des rekonstruierten Signals verantwortlich.

[0003] Das Ergebnis dieser Filterung, stellt dann die Approximation des zu übertragenden Signalstückes dar. Für jeden Abschnitt muß die Information über die Filtereinstellungen und die Information über das zu verwendende Anregungssignal und dessen Skalierung ("gain"), die die Lautstärke beschreibt, übertragen werden. Im allgemeinen werden diese Parameter aus verschiedenen, dem Encoder und Decoder in identischen Kopien vorliegenden Codebüchern gewonnen, so daß zur Rekonstruktion nur die Nummer der am besten geeigneten Codebucheinträge übertragen werden muß. Bei der Codierung eines Sprachsignals sind also für jeden Abschnitt diese am besten geeigneten Codebucheinträge zu bestimmen, wobei alle relevanten Codebucheinträge in allen relevanten Kombinationen durchsucht werden, und diejenigen Einträge ausgewählt werden, die die im Sinne eines sinnvollen Abstandsmaßes kleinste Abweichung zum Originalsignal liefern.

[0004] Es existieren verschiedene Verfahren zur Optimierung der Struktur der Codebücher (z.B. Mehrstufigkeit, Lineare Prädiktion basierend auf den vergangenen Werten, spezifische Abstandsmaße, optimierte Suchverfahren, etc.). Außerdem gibt es verschiedene Verfahren, die den Aufbau und das Durchsuchungsverfahren für die Bestimmung der Anregungsvektoren beschreiben.

[0005] Auch die Bestimmung des Verstärkungsfaktors (gain-Wertes) läßt sich auf verschiedene Weise sinnvoll realisieren. Der Verstärkungsfaktor kann im Prinzip mittels zweier nachfolgend beschriebener Methoden angenähert werden:

Methode 1: "waveform matching"

[0006] Bei dieser Methode wird der Verstärkungsfaktor unter Berücksichtigung der Wellenform des Anregungssignals aus dem Codebuch berechnet. Zur Berechnung wird die Abweichung E_1 zwischen ursprünglichem, d.h. zu übertragendem Signal \underline{x} (in der Darstellung als Vektor und dem rekonstruierten Signal $\underline{g H c}$ minimiert. Dabei ist \underline{g} der zu bestimmende Verstärkungsfaktor, \underline{H} die die Filteroperation beschreibende Matrix und \underline{c} der ebenfalls zu bestimmende bestgeeignete Anregungscodebuchvektor, der die gleiche Dimension hat wie der Zielvektor \underline{x} .

$$E_1 = || \underline{x} - \underline{g H c} ||^2$$

[0007] Zur Berechnung wird im allgemeinen zunächst der optimale Codebuchvektor \underline{c} -opt bestimmt. Danach wird der hierfür optimale Verstärkungsfaktor \underline{g} zunächst berechnet und dann der hierzu passende Codebuchvektor \underline{g} -opt bestimmt. Diese Berechnung liefert gute Werte immer dann, wenn die Wellenform des mit \underline{H} gefilterten Anregungscodebuchvektors aus dem Codebuch möglichst gut mit der vorgegebenen Wellenform übereinstimmt. Dies ist z. B. bei klarer Sprache ohne Hintergrundgeräusche im allgemeinen häufiger der Fall als bei Sprachsignalen mit Hintergrundgeräuschen. Bei starken Hintergrundgeräuschen kann eine Verstärkungsfaktor-Berechnung nach Methode 1 daher zu störenden Effekten führen, die sich z. B. in Form von Lautstärkeschwankungen äußern können.

[0008] Eine Verstärkungsfaktor-Berechnung nach Methode 1 wird im Dokument 'HAGEN ET AL: "AN 8 KBIT/S ACELP CODER WITH IMPROVED BACKGROUND NOISE PERFORMANCE" PHOENIX, AZ, MARCH 15 - 19, 1999, NEW YORK, 15. März 1999, Seiten 25-28, XP000898256 beschrieben.

Methode 2: "energy matching"

[0009] Bei dieser Methode wird der Verstärkungsfaktor g ohne Berücksichtigung der Wellenform des Sprachsignals berechnet. Bei der Berechnung wird die Abweichung E_2 minimiert:

$$E_2 = (|| \text{exc}(g) || - || \text{res} ||)^2$$

[0010] Dabei ist exc der skalierte Codebuchvektor, der von dem Verstärkungsfaktor g abhängt, res bezeichnet das "ideale" Anregungssignal. Außerdem können noch andere, vorher bestimmte konstante Codebucheinträge d hinzukommen:

$$\text{exc}(g) = c_{\text{opt}} * g + d$$

[0011] Dieses Verfahren liefert gute Werte z. B. bei wenig periodischen Signalen, zu denen z. B. Sprachsignale gehören können, die einen hohen Hintergrund-Geräuschpegel aufweisen. Die nach Methode 2 berechneten Verstärkungswerte liefern andererseits bei geringen Hintergrundgeräuschen im allgemeinen schlechtere Werte als Methode 1.

[0012] Bei dem heute eingesetzten Verfahren wird zunächst der aus Methode 1 folgende optimale Codebucheintrag g_{opt} bestimmt, und dann der tatsächlich zu verwendenden quantisierte, d.h. im Codebuch gefundene Verstärkungsfaktor g_{opt2} durch Minimierung der Größe E_3 bestimmt:

$$\begin{aligned} \text{G1. (1)} \quad E_3(g_{\text{opt2}}) = & (1-a) * ||c_{\text{opt}}||^2 * (g_{\text{opt2}} - g_{\text{opt}})^2 \\ & + a * (||\text{exc}(g_{\text{opt2}})|| - ||\text{res}||)^2 \end{aligned}$$

[0013] Der Gewichtungsfaktor a kann dabei Werte zwischen 0 und 1 annehmen, und ist mittels geeigneter Algorithmen vorzugeben. Für den Extremfall $a = 0$ wird in dieser Gleichung nur der erste Summand beachtet. Die Minimierung von E_3 führt in diesem Fall stets auf $g_{\text{opt2}} = g_{\text{opt}}$, so daß der zuvor nach Methode 1 berechnete Wert g_{opt} als Ergebnis der endgültigen Verstärkungsfaktor-Berechnung übernommen wird (reines "waveform matching"). Für den anderen Extremfall $a = 1$ wird dagegen nur der zweite Summand betrachtet. In diesem Falle wird dann für g_{opt2} stets die gleiche Lösung resultieren wie bei Anwendung der Methode 2 (reines "energy matching,"). Im allgemeinen wird der Wert von a zwischen 0 und 1 liegen, und somit zu einem Ergebniswert für g_{opt2} führen, der beide Methoden 1 "waveform matching" und 2 "energy matching" berücksichtigt.

[0014] Über den Gewichtungsfaktor a wird also gesteuert, in welchem Ausmaß das Ergebnis der Methode 1 bzw. das Ergebnis der Methode 2 verwendet werden soll. Der nach G1. (1) durch Minimierung von E_3 berechnete, quantisierte Wert gain-eff2 wird dann übertragen und auf der Decoderseite verwendet.

[0015] Das zugrunde liegende Problem besteht nun darin, für jeden zu codierenden Signalabschnitt den Gewichtungsfaktor a so zu bestimmen, daß die Berechnung nach G1. (1) oder einer anderen Minimierungsfunktion, bei der eine Gewichtung zwischen zwei Methoden Verwendung findet, möglichst sinnvolle Werte gefunden werden. "Sinnvolle Werte" sind im Sinne der Sprachqualität der Übertragung solche Werte, die der im aktuellen Signalabschnitt gegebenen Signalsituation möglichst gut angepaßt sind. Für z. B. störgeräuschfreie Sprache wäre z. B. a nahe 0 zu wählen. bei starken Hintergrundgeräuschen wäre a nahe 1 zu wählen.

[0016] Bei den heute verwendeten Verfahren wird der Wert des Gewichtungsfaktors a über ein Periodizitätsmaß gesteuert, indem der Prädiktionsgewinn als Grundlage für die Bestimmung der Periodizität des vorliegenden Signals genommen wird. Aus der den aktuellen Signalzustand beschreibenden Angabe des Periodizitätsmaßes, das mit p bezeichnet sei, wird der zu verwendende Wert von a über eine feste Kennlinie $f(p)$ ermittelt. Diese Kennlinie ist so gestaltet, daß sie für stark periodische Signale einen niedrigeren Wert für a liefert. Das heißt, für stark periodische Signale wird die Methode 1 des "waveform matching," bevorzugt. Für weniger stark periodische Signale wird dagegen über $f(p)$ ein höherer Wert für a , d. h. näher bei 1, vorgegeben.

[0017] Es zeigt sich jedoch in der Praxis, daß diese Methode bei bestimmten Signalen immer noch zu Artefakten führt. Hierzu zählen z. B. der Beginn stimmhafter Signalstücke, sog. Onsets, oder aber auch Rauschsignale ohne periodische

Anteile.

[0018] Aufgabe der vorliegenden Erfindung ist es daher, ein Verfahren bereit zu stellen, mittels dem ein optimaler Gewichtungsfaktor a für die Berechnung eines möglichst optimalen Verstärkungsfaktors für nahezu alle Signale ermittelbar ist.

[0019] Diese Aufgabe wird erfindungsgemäß mit einem Verfahren mit den Merkmalen des Anspruchs 1 gelöst. Weitere vorteilhafte Ausgestaltungen des Verfahrens ergeben sich durch die Merkmale der Unteransprüche.

[0020] Das erfindungsgemäße Verfahren sieht vor, nicht nur die Periodizität S_1 des Signals für die Bestimmung des Gewichtungsfaktors, sondern zusätzlich die Stationarität S_2 des Signals zu verwenden. In Abhängigkeit von der Güte des zu ermittelnden Gewichtungsfaktors a können weitere Parameter, welche charakteristisch für die vorliegenden Signale sind, wie es z.B. die kontinuierliche Schätzung des Störpegels, bei der Bestimmung des Gewichtungsfaktors Berücksichtigung finden. Der Gewichtungsfaktor a wird demnach vorteilhaft nicht nur anhand der Periodizität S_1 , sondern von einer Mehrzahl von Parametern ermittelt. Die Anzahl der zum Einsatz kommenden Parameter bzw. Maße sei mit N bezeichnet. Aus der Kombination der Ergebnisse der einzelnen Maße kann eine verbesserte, robustere Bestimmung von a vorgenommen werden. Damit wird der zu verwendende Wert von a nicht mehr von nur einem Maß abhängig gemacht, sondern er hängt über eine Vorschrift h von den den aktuellen Signalzustand beschreibenden Angaben aller N Maße S_1, S_2, \dots, S_N ab. Es ergibt sich der in Gl. (2) gezeigte Zusammenhang:

$$(Gl. 2) \quad a = h(S_1, S_2, \dots, S_N)$$

[0021] Eine beispielhafte erfindungsmässige Umsetzung wäre demnach in einem System zu sehen, daß einerseits ein Periodizitätsmaß S_1 und zusätzlich auch ein Stationaritätsmaß S_2 verwendet. Durch die zusätzliche Berücksichtigung der Stationarität S_2 des Signals können z. B. die oben genannten Problemfälle (Onsets, Rauschen) besser behandelt werden. In einem Sprachcodiersystem, das das erfindungsgemäße Verfahren verwendet, werden dabei zunächst die Ergebnisse des Periodizitätsmaßes S_1 und des Stationaritätsmaßes S_2 berechnet. Dann werden gemäß Gl. (2) aus den beiden Maßen der passende Wert für den Gewichtungsfaktor a berechnet. Dieser Wert wird dann in Gl. (1) zur Bestimmung des besten Wertes für den Verstärkungsfaktor verwendet.

[0022] Eine konkrete Möglichkeit, die Zuordnungsvorschrift $h(S_1)$ zu realisieren, besteht z. B. darin, eine Anzahl von K verschiedenen Kennlinienverläufen $h_1(S_1) \dots h_K(S_1)$ zu verwenden und den im vorliegenden Signalfall zu verwendenden Kennlinienverlauf $h_i(S_1)$ über einen Parameter S_2 zu steuern:

[0023] Für $K = 3$ könnten dabei folgende Unterscheidungen vorgenommen werden:

verwende $a = h_1(S_1)$, wenn $S_{2a} < S_2 \leq S_{2b}$,

verwende $a = h_2(S_1)$, wenn $S_{2b} < S_2 \leq S_{2c}$,

verwende $a = h_3(S_1)$, wenn $S_{2c} < S_2 \leq S_{2d}$,

wobei $S_{2a} < S_2 < S_{2d}$

[0024] Nachfolgend wird das erfinderische Verfahren am Beispiel für $K=2$ näher erläutert. Die verwendete Zuordnungsvorschrift $h(\cdot)$ sieht in diesem Fall zwei unterschiedliche Kennlinienverläufe $h_1(S_1)$ und $h_2(S_1)$ vor. Die Auswahl der jeweiligen Kennlinie erfolgt in Abhängigkeit von einem weiteren Parameter S_2 der entweder 0 oder 1 ist.

[0025] Der Parameter S_1 beschreibt die Stimmhaftigkeit (Periodizität) des Signals. Die Information über die Stimmhaftigkeit ergibt sich aus der Kenntnis des Eingangssignals $s(n)$ ($n=0 \dots L$, L : Länge des betrachteten Signalabschnitts) sowie des Schätzwerts τ der Pitch (Dauer der Grundperiode des momentanen Sprachsegments). Zunächst ist ein stimmhaft/stimmlos Kriterium gemäss

$$\chi = \frac{\sum_{i=0}^{L-1} s(i) \cdot s(i - \tau)}{\sqrt{\sum_{i=0}^{L-1} s^2(i) \cdot \sum_{i=0}^{L-1} s^2(i - \tau)}}$$

zu berechnen. Den verwendeten Parameter S_1 erhält man nun durch Bildung des kurzzeitigen Mittelwertes von χ über den letzten 10 Signalabschnitten (m_{cur} : Index des momentanen Signalabschnitts):

$$S_1 = \frac{1}{10} \sum_{i=m_{cur}-10}^{m_{cur}} \chi_i .$$

[0026] Die Figur 1 zeigt schematisch die Abhängigkeit des Gewichtungsfaktors a von S_1 .

[0027] Die Form der Kennlinie hängt demnach von der Wahl der Schwellwerte a_l und a_h sowie s_{1l} und s_{1h} ab.

[0028] Die angegebene Auswahl der Kennlinie h_1 bzw. h_2 in Abhängigkeit von s_2 bedeutet, daß für unterschiedliche Werte von S_2 verschiedene Schwellwertkombinationen (a_l, a_h, s_{1l}, s_{1h}) gewählt werden.

[0029] Der Parameter S_2 enthält eine Aussage über die Stationarität des vorliegenden Signalabschnitts. Konkret handelt es sich um eine Zustandsinformation die angibt, ob im momentan betrachteten Signalabschnitt Sprachaktivität ($s_2 = 1$) oder eine Sprachpause vorliegt ($S_2 = 0$).

[0030] Diese Information muß von einem Algorithmus zur Detektierung von Sprachpausen (engl. VAD = Voice Activity Detection) geliefert werden.

[0031] Da die Erkennung von Sprachpausen und stationären Signalabschnitten vom Prinzip her ähnlich sind, ist die VAD nicht auf eine exakte Bemessung der Sprachpausen (wie sonst üblich) sondern auf eine Klassifikation solcher Signalabschnitte hin optimiert, die hinsichtlich der Bestimmung des Verstärkungsfaktors als stationär gelten.

[0032] Da die Stationarität S_2 eines Signals keine eindeutig festgelegte Meßgröße ist, wird sie nachfolgend genauer definiert.

[0033] Betrachtet man zunächst das Frequenzspektrum eines Signalabschnitts, so weist es für den betrachteten Zeitraum eine charakteristische Form auf. Ist die Änderung der Frequenzspektren zeitlich aufeinanderfolgender Signalabschnitte hinreichend gering, d.h. die charakteristische Form der jeweiligen Spektren bleibt mehr oder weniger erhalten, so kann man von spektraler Stationarität sprechen.

[0034] Betrachtet man einen Signalabschnitt im Zeitbereich, so weist es einen für den betrachteten Zeitraum charakteristischen Amplituden- bzw. Energieverlauf auf. Bleibt die Energie zeitlich aufeinanderfolgender Signalabschnitte konstant, bzw. die Abweichung der Energie ist auf ein hinreichend kleines Toleranzintervall begrenzt, so kann man von zeitlicher Stationarität sprechen.

[0035] Sind zeitlich aufeinanderfolgende Signalabschnitte sowohl spektral als auch zeitlich stationär so werden sie allgemein als stationär bezeichnet. Die Bemessung spektraler und zeitlicher Stationarität erfolgt in zwei separaten Stufen. Zunächst wird die spektrale Stationarität untersucht:

Spektrale Stationarität (1. Stufe)

[0036] Zur Feststellung, ob spektrale Stationarität vorliegt, wird zunächst ein spektrales Abstandsmaß, die sog. *spektrale Verzerrung* SD (engl.: spectral distortion) aufeinanderfolgender Signalabschnitte betrachtet. Die Berechnung ergibt sich gemäß:

$$SD = \sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} \left(10 \log \left[\frac{1}{|A(e^{j\omega})|^2} \right] - 10 \log \left[\frac{1}{|A'(e^{j\omega})|^2} \right] \right)^2 d\omega} .$$

[0037] Dabei bezeichnet

$$10 \log \left[\frac{1}{|A(e^{j\omega})|^2} \right]$$

den logarithmierten Einhüllendenfrequenzgang des aktuellen Signalabschnitts und

5

$$10 \log \left[\frac{1}{|A'(e^{j\omega})|^2} \right]$$

10 den logarithmierten Einhüllendenfrequenzgang des vorangegangenen Signalabschnitts. Zur Entscheidung wird sowohl SD selbst als auch sein kurzzeitlicher Mittelwert über den letzten 10 Signalabschnitten \overline{SD} betrachtet. Liegen beide Maße SD und \overline{SD} unterhalb eines für sie spezifischen Schwellwertes SD_g , bzw. \overline{SD}_g , so wird spektrale Stationarität angenommen.

[0038] Konkret gilt:

15

$$\begin{aligned} SD_g &= 2.6 \text{ dB} \\ \overline{SD}_g &= 2.6 \text{ dB} \end{aligned}$$

20 **[0039]** Problematisch ist, daß auch extrem periodische (stimmhafte) Signalabschnitte diese spektrale Stationarität aufweisen. Sie werden über das Periodizitätsmaß s_1 ausgeschlossen. Es gilt:

$$\begin{aligned} \text{Falls } s_1 &\geq 0.7 \\ \text{oder } s_1 &< 0.3 \end{aligned}$$

25 ist, wird der betrachteten Signalabschnitt als nicht spektral stationär angenommen.

Zeitliche Stationarität (2. Stufe):

30 **[0040]** Die Feststellung der zeitlichen Stationarität erfolgt in einer zweiten Stufe, deren Entscheidungsschwellen von der Detektion spektral stationärer Signalabschnitte der ersten Stufe abhängen. Ist der vorliegende Signalabschnitt von der ersten Stufe als spektral stationär klassifiziert worden, so wird sein Einhüllendenfrequenzgang

35

$$\frac{1}{|A(e^{j\omega})|^2}$$

40 gespeichert. Ebenfalls gespeichert wird die Referenz-Energie $E_{\text{reference}}$ des Restsignals $d_{\text{reference}}$, welches sich aus der Filterung des vorliegenden Signalabschnittes mit einem Filter ergibt, das den zu diesem Signalabschnitt inversen Frequenzgang $|A(e^{j\omega})|^2$ besitzt. $E_{\text{reference}}$ ergibt sich durch

45

$$E_{\text{reference}} = \sum_{n=0}^{L-1} d_{\text{reference}}^2(n),$$

wobei L der Länge des betrachteten Signalabschnitts entspricht.

50 **[0041]** Diese Energie dient als Referenzwert bis zur Detektion des nächsten spektral stationären Abschnitts. Alle folgenden Signalabschnitte werden nun mit demselben gespeicherten Filter gefiltert. Gemessen wird nun die Energie E_{rest} des nach Filterung entstandenen Restsignals d_{rest} . Sie ergibt sich entsprechend zu

55

$$E_{\text{rest}} = \sum_{n=0}^{L-1} d_{\text{rest}}^2(n).$$

[0042] Die endgültige Entscheidung, ob der betrachtete Signalabschnitt stationär ist folgt folgender Vorschrift:

Falls: $E_{\text{rest}} < E_{\text{reference}} + \text{Toleranz}$
 $s_2 = 1$, Signal stationär,
 sonst: $s_2 = 0$, Signal instationär

[0043] Es gilt dabei beispielhaft die in Fig. 2 dargestellte Zuordnung, wobei für

$s_2 = 1$ ($h_1(s_1)$, instationär): und

$s_2 = 0$ ($h_2(s_1)$, stationär/pause) $\rightarrow a = 1.0$ für alle s_1

[0044] Das bedeutet die Kennlinie ist flach und a hat unabhängig von s_1 den Wert 1.

[0045] Natürlich ist auch eine Abhängigkeit denkbar, in der ein kontinuierlicher Parameter S_2 ($0 \leq s_2 \leq 1$) eine Aussage über die Stationarität S_2 enthält. In diesem Fall tritt anstelle der unterschiedlichen Kennlinien h_1 und h_2 eine dreidimensionale Fläche $h(s_1, s_2)$ durch die a bestimmt wird.

[0046] Es versteht sich von selbst, daß die Algorithmen zur Bestimmung der Stationarität und der Periodizität den jeweils gegebenen Umständen entsprechend angepaßt werden müssen bzw. können. Die einzelnen o.a. Schwellwerte und Funktionen sind lediglich exemplarisch und müssen in der Regel durch eigene Versuche herausgefunden werden.

Patentansprüche

1. Verfahren zur Berechnung des die Lautstärke mitbestimmenden Verstärkungsfaktors für ein codiert übertragenes Sprachsignal, wobei das Sprachsignal in kurze zeitliche Signalabschnitte unterteilt und die einzelnen Signalabschnitte getrennt voneinander codiert und übertragen werden, und zu jedem Signalabschnitt der Verstärkungsfaktor berechnet, übertragen und vom Decoder zur Rekonstruktion des Signals verwendet wird, wobei der Verstärkungsfaktor durch Minimierung der Größe $E(g_{\text{opt}2}) = (1-a) \cdot f_1(g_{\text{opt}2}) + a \cdot f_2(g_{\text{opt}2})$ bestimmt wird, wobei $g_{\text{opt}2}$ einen zu verwendeten quantisierten Verstärkungsfaktor darstellt, und f_1 und f_2 entsprechende Funktionen darstellen, **dadurch gekennzeichnet, daß** die Bestimmung des Gewichtungsfaktors a unter Berücksichtigung sowohl eines Periodizitätsmaßes S_1 als auch eines Stationaritätsmaßes S_2 des codierten Sprachsignals erfolgt.

2. Verfahren nach Anspruch 1, **dadurch gekennzeichnet, daß** die Minimierung der Größe $E(g_{\text{opt}2})$ erfolgt mittels der Formel:

$$E(g_{\text{opt}2}) = (1-a) \cdot ||c_{\text{opt}}||^2 \cdot (g_{\text{opt}2} - g_{\text{opt}})^2 + a \cdot (||\text{exc}(g_{\text{opt}2})|| - ||\text{res}||)^2,$$

wobei c_{opt} , g_{opt} , exc und res einen optimalen Codebuchvektor, einen optimalen Verstärkungsfaktor, einen skalierten Codebuchvektor und ein ideales Anregungssignal darstellen.

3. Verfahren nach Anspruch 1 oder 2, **dadurch gekennzeichnet, daß** in Abhängigkeit des ermittelten Wertes für das Stationaritätsmaß S_2 des Sprachsignals eine bestimmte Funktion $h_1(S_1)$ zur Bestimmung des Gewichtungsfaktors a ausgewählt wird, wobei S_1 ein Maß für die Periodizität des Sprachsignals ist.

4. Verfahren nach Anspruch 3, **dadurch gekennzeichnet, daß** das Stationaritätsmaß S_2 ein Maß für die Sprachaktivität eines Signalabschnittes ist.

5. Verfahren nach einem der Ansprüche 3 oder 4, **dadurch gekennzeichnet, daß** das Stationaritätsmaß S_2 ein Maß für das Verhältnis von Sprachpegel zu Hintergrundgeräuschpegel des zu betrachtenden Sprachsignalabschnittes ist.

6. Verfahren nach einem der vorhergehenden Ansprüche, **dadurch gekennzeichnet, daß** das Stationaritätsmaß S_2 in Abhängigkeit von der spektralen Veränderung sowie der Energieveränderung, d.h. zeitliche Stationarität berechnet wird.

7. Verfahren nach Anspruch 6, **dadurch gekennzeichnet, daß** zur Berechnung der spektralen Stationarität sowie der Energieveränderung d.h. zeitliche Stationarität, mindestens ein zeitlich vorangegangener Signalabschnitt berücksichtigt wird.
8. Verfahren nach Anspruch 7, **dadurch gekennzeichnet, daß** die ermittelten Werte der spektralen Veränderung die Bewertung der Energieveränderung, d. h. zeitlichen Stationarität, beeinflußt.

Claims

1. Method for calculating the gain, said gain codetermining the loudness, for a codedly transmitted speech signal, wherein the speech signal is divided into short time signal portions and the individual signal portions are separately coded and transmitted, wherein, for each signal portion, the gain is calculated, transmitted and used by the decoder to reconstruct the signal, the gain factor being determined by minimization of the quantity $E(g_opt2) = (1-a)*f_1(g_opt2) + a*f_2(g_opt2)$, where g_opt2 represents a quantized gain to be used, and f_1 and f_2 represent corresponding functions, **characterized in that** the weighting factor a is determined in consideration both of a periodicity factor S_1 and also of a stationarity factor S_2 of the coded speech signal.

2. Method according to claim 1, **characterized in that** the quantity $E(g_opt2)$ is minimized by means of the formula:

$$E(g_opt2) = (1-a) * \|c_opt\|^2 * (g_opt2 - g_opt)^2 + a * (\|exc(g_opt2)\| - \|res\|)^2,$$

where c_opt , g_opt , exc and res represent an optimum code book vector, an optimum gain, a scaled code book vector and an ideal excitation signal.

3. Method according to claim 1 or 2, **characterized in that**, depending on the determined value for the stationarity factor S_2 of the speech signal, a determined function $h_i(S_1)$ is selected for determination of the weighting factor a , where S_1 is a measure of the periodicity of the speech signal.
4. Method according to claim 3, **characterized in that** the stationarity factor S_2 is a measure of the speech activity of a signal portion.
5. Method according to any one of claims 3 or 4, **characterized in that** the stationarity factor S_2 is a measure of the ratio of speech level to background noise level of the speech signal portion under consideration.
6. Method according to any one of the preceding claims, **characterized in that** the stationarity factor S_2 is calculated as a function of the spectral change as well as the energy change, i.e. time stationarity.
7. Method according to claim 6, **characterized in that**, for calculation of the spectral stationarity as well as the energy change, i.e. time stationarity, at least one chronologically preceding signal portion is taken into consideration.
8. Method according to claim 7, **characterized in that** the determined values of the spectral change influence the evaluation of the energy change, i.e. time stationarity.

Revendications

1. Procédé de calcul du facteur d'amplification influant sur le volume sonore d'un signal vocal transmis sous forme codée, le signal vocal étant divisé en sections de signal courtes dans le temps qui sont codées et transmises séparément les unes des autres, et pour chacune de ces sections de signal, le facteur d'amplification étant calculé, transmis et utilisé par le décodeur pour reconstruire le signal, ce facteur d'amplification étant déterminé par la minimisation de la grandeur $E(g_opt2) = (1 - a) * f_1(g_opt2) + a * f_2(g_opt2)$, où g_opt2 représente un facteur d'amplification quantifié à utiliser et où f_1 et f_2 représentent des fonctions correspondantes, **caractérisé en ce que** le facteur de pondération a est déterminé en tenant compte à la fois d'un degré de périodicité S_1 et d'un degré de

stationnarité S_2 du signal vocal codé.

2. Procédé selon la revendication 1, **caractérisé en ce que** la minimisation de la grandeur $E(g_{opt2})$ s'effectue moyennant la formule

$$E(g_{opt2}) = (1 - a) * |c_{opt}|^2 * (g_{opt2} - g_{opt})^2 + a * (|exc(g_{opt2})| - |res|)^2$$

où c_{opt} , g_{opt} , exc et res représentent un vecteur de dictionnaire de codes optimal, un facteur d'amplification optimal, un vecteur de dictionnaire de codes normé et un signal d'excitation idéal.

3. Procédé selon la revendication 1 ou 2, **caractérisé en ce que**, en fonction de la valeur obtenue pour le degré de stationnarité S_2 du signal vocal, une fonction donnée $h_1(S_1)$ est sélectionnée pour déterminer le facteur de pondération a , S_1 étant une mesure de la périodicité du signal vocal.

4. Procédé selon la revendication 3, **caractérisé en ce que** le degré de stationnarité S_2 est une mesure de l'activité vocale d'une section de signal.

5. Procédé selon l'une des revendications 3 ou 4, **caractérisé en ce que** le degré de stationnarité S_2 est une mesure du rapport entre le niveau vocal et le niveau de bruit de fond de la section de signal vocal à observer.

6. Procédé selon l'une des revendications précédentes, **caractérisé en ce que** le degré de stationnarité S_2 est calculé en fonction de la modification spectrale ainsi que de la variation d'énergie, à savoir la stationnarité temporelle.

7. Procédé selon la revendication 6, **caractérisé en ce que** pour calculer la stationnarité spectrale ainsi que la variation d'énergie, à savoir la stationnarité temporelle, on tient compte d'au moins une section de signal qui a précédé dans le temps.

8. Procédé selon la revendication 7, **caractérisé en ce que** les valeurs obtenues pour la modification spectrale influent sur l'évaluation de la variation d'énergie, à savoir la stationnarité temporelle.

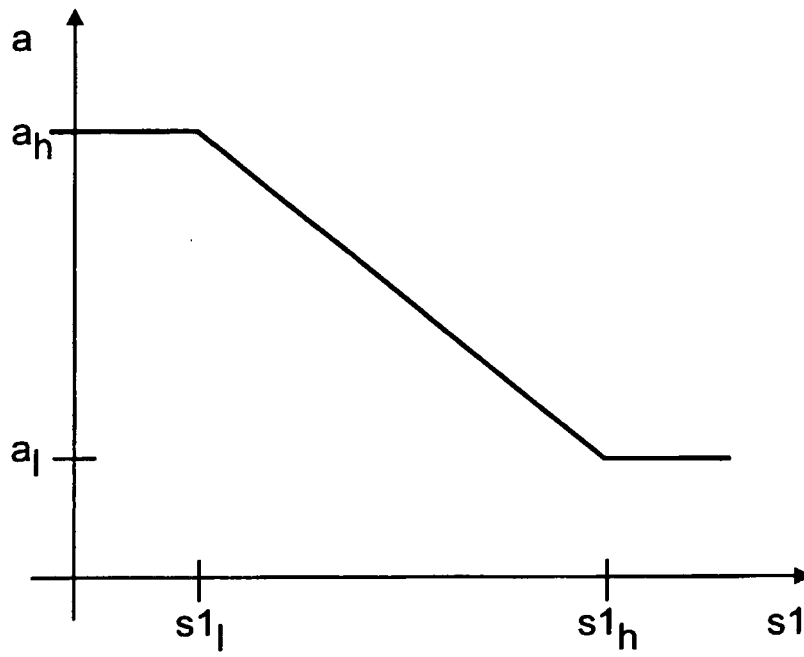
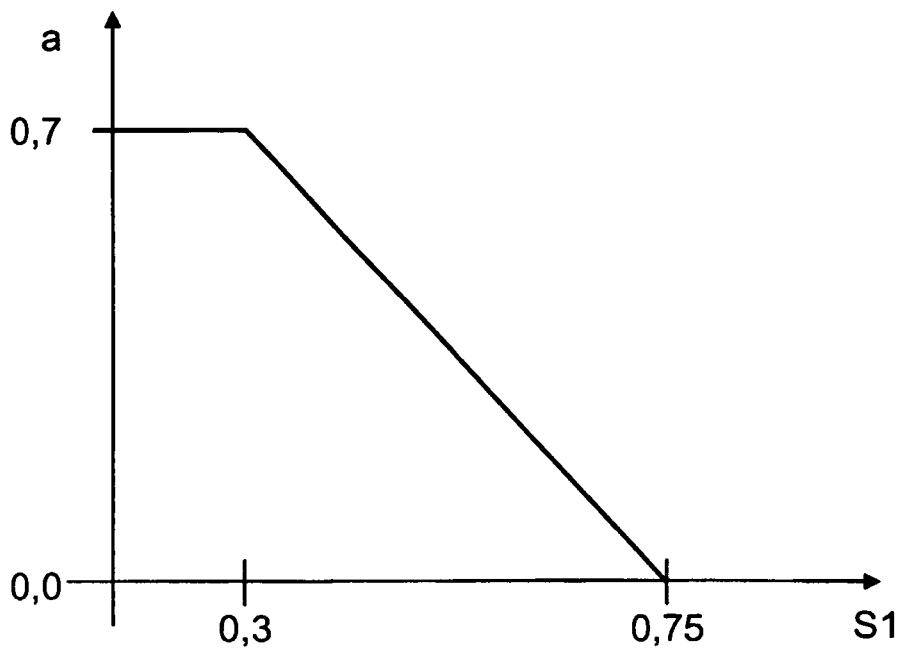


Fig. 1



$$\begin{aligned} a_l &= 0,0 \\ a_h &= 0,7 \\ s_{1l} &= 0,3 \\ s_{1h} &= 0,75 \end{aligned}$$

Fig. 2

IN DER BESCHREIBUNG AUFGEFÜHRTE DOKUMENTE

Diese Liste der vom Anmelder aufgeführten Dokumente wurde ausschließlich zur Information des Lesers aufgenommen und ist nicht Bestandteil des europäischen Patentdokumentes. Sie wurde mit größter Sorgfalt zusammengestellt; das EPA übernimmt jedoch keinerlei Haftung für etwaige Fehler oder Auslassungen.

In der Beschreibung aufgeführte Patentdokumente

- US 4133976 A [0002]
- XP 000898256 [0008]

In der Beschreibung aufgeführte Nicht-Patentliteratur

- **HAGEN et al.** AN 8 KBIT/S ACELP CODER WITH IMPROVED BACKGROUND NOISE PERFORMANCE. *PHOENIX, AZ*, 15. März 1999, 25-28 [0008]