

(19)



Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

EP 1 288 914 A2

(12)

EUROPÄISCHE PATENTANMELDUNG

(43) Veröffentlichungstag:

05.03.2003 Patentblatt 2003/10

(51) Int Cl.7: **G10L 19/00**

(21) Anmeldenummer: **02012790.8**

(22) Anmeldetag: **10.06.2002**

(84) Benannte Vertragsstaaten:

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

Benannte Erstreckungsstaaten:

AL LT LV MK RO SI

(71) Anmelder: **Deutsche Telekom AG**

53113 Bonn (DE)

(72) Erfinder: **Berger, Jens, Dr.**

10405 Berlin (DE)

(30) Priorität: **29.08.2001 DE 10142846**

(54) **Verfahren zur Korrektur von gemessenen Sprachqualitätswerten**

(57) Instrumentelle Verfahren zur Bestimmung der Sprachqualität von Sprachsignalen werten oft nur Signalabschnitte mit Sprachaktivität aus. Hintergrundgeräusche, die in den Sprachpausen auftreten, bleiben dabei unberücksichtigt. Auch der aktuelle ITU-T Standard P.862 "PESQ" beschränkt sich für die Berechnung des Sprachqualitätswertes auf Signalabschnitte mit Sprachaktivität. Wenn Hintergrundgeräusche in Sprachpausen auftreten, aber ansonsten hochqualitative Sprachqualität vorliegt, bewerten diese Messverfahren die Sprachqualität als zu hoch.

Mit dem vorliegenden Verfahren werden die Geräusche in den Sprachpausen bei der Bestimmung der Sprachqualität berücksichtigt, in dem die gemessenen

Sprachqualitätswerte korrigiert werden. Dazu werden ein oder mehrere Intensitätswert von den Hintergrundgeräuschen in den Sprachpausen (z. B. mittlere Lautheit) bestimmt und aus diesen sowie dem Anteil der Sprachpausen im Gesamtsignal und dem Qualitätswert selbst der Korrekturwert ermittelt, um den der gemessene Sprachqualitätswert reduziert wird.

Das beschriebene Ausführungsbeispiel zeigt die Korrektur der Sprachqualitätswerte unter Benutzung der mittleren Lautheit der Hintergrundgeräusche in Sprachpausen für den aktuellen ITU-T Standard P.862 "PESQ" (Stand 2001).

EP 1 288 914 A2

Beschreibung

Vorbemerkung

[0001] Die Erfindung bezieht sich auf instrumentelle Verfahren zur Messung der Sprachqualität von aufgetragenen oder übertragenen Sprachsignalen. Dabei wird von Sprachqualitätsmesswerten ausgegangen, die z.B. mit dem ITU-T Standard P.862 ("Perceptual Evaluation of Speech Quality (PESQ), an Objective Method for end-to-end Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs", ITU-T, Genf, 2001) bestimmt werden.

[0002] Die empfundene Sprachqualität z. B. bei Telefonverbindungen oder Rundfunkübertragungen wird hauptsächlich von sprachsimultanen Störungen, also Störungen während der Sprachaktivität, bestimmt. Aber auch Geräusche in Sprachpausen gehen in das Qualitätsurteil ein, das insbesondere bei hochqualitativer Sprachwiedergabequalität.

[0003] Sprachqualitätsbestimmungen von Sprachsignalen werden in der Regel mittels auditiver ("subjektiver") Untersuchungen mit Versuchspersonen vorgenommen. Das Ziel von instrumentellen ("objektiven") Verfahren zur Sprachqualitätsbestimmung ist es dagegen, aus Eigenschaften des zu bewertenden Sprachsignals mittels geeigneter Rechenverfahren Kennwerte zu ermitteln, die die Sprachqualität des zu bewertenden Sprachsignals beschreiben ohne auf Urteile von Versuchspersonen zurückgreifen zu müssen.

Stand der Technik und Nachteile bekannter Verfahren

[0004] Bekannte Verfahren zur instrumentellen Bestimmung der Sprachqualität (z. B. ITU-T Rec. P.862 "PESQ") bestimmen die Sprachqualität anhand eines Vergleichs zwischen ungestörtem Referenzsprachsignal (Quellsprachsignal) und dem zu bewertenden und möglicherweise gestörten Signal.

[0005] Solche Messverfahren werden in sogenannten Probeverbindungssystemen eingesetzt, bei denen ein bekanntes Referenzsprachsignal (Quellsprachsignal) an der Quelle eingespeist und z. B. über eine Telefonverbindung übertragen und an der Senke aufgetraget wird. Nach der Aufzeichnung des Sprachsignals erfolgt die Berechnung eines Sprachqualitätswertes.

[0006] Instrumentelle Verfahren zur Sprachqualitätsbestimmung beschränken sich meist auf die Auswertung von Abschnitten mit Sprachaktivität. Der aktuelle ITU-T Standard P.862 beschränkt sich bei der Bestimmung der Sprachqualität ebenfalls nur auf Abschnitte mit aktiver Sprache. Insbesondere bei hochqualitativer Sprachwiedergabe und Geräuschen, die ausschließlich in Sprachpausen auftreten, liefern diese Verfahren (z. B. Messverfahren nach ITU-T Rec. P.862) unzuverlässige Qualitätswerte. Die Sprachqualität wird hier zu op-

timistisch bewertet, da die von einem Zuhörer empfundene Sprachqualität sich auf das gesamte Signal einschließlich möglicher Geräusche in den Sprachpausen bezieht.

Aufgabe

[0007] Einige instrumentelle Verfahren zur Sprachqualitätsbestimmung, wie z. B. das Verfahren nach ITU-T Rec. P.862, berücksichtigen bei der Berechnung der Sprachqualitätswerte die Geräusche in den Sprachpausen nicht. Die sich ergebenden Messwerte sind dadurch, insbesondere bei hoher Wiedergabequalität bei Sprachaktivität aber auftretenden Geräuschen in Sprachpausen, unzuverlässig. Mit dem vorliegenden Verfahren sollen die Hintergeräusche in den Sprachpausen bei der Bestimmung der Sprachqualitätswerte berücksichtigt werden.

Lösungsprinzip

[0008] Die Lösung der Aufgabe geht davon aus, dass die Hintergrundgeräusche in den Sprachpausen hinsichtlich ihres störenden Einflusses auf die empfundene Sprachqualität bewertet werden. Dazu werden Intensitätskennwerte des Hintergrundgeräusches ermittelt und mit diesen Werten die Sprachqualitätsmesswerte, die durch ein instrumentales Verfahren, z. B. nach ITU-T Rec. P.862, gewonnenen wurden, korrigiert.

[0009] Zunächst wird der Sprachqualitätswert unter Benutzung des Quellsprachsignals und des zu bewertenden gestörten Sprachsignals, z. B. mit dem Verfahren nach ITU-T Rec. P.962, berechnet. Diese Sprachsignale stehen dem anschließenden Korrekturverfahren ebenfalls als Eingangsparameter zur Verfügung. Zusätzlich benötigt das hier beschriebene Korrekturverfahren noch den berechneten Sprachqualitätswert, der z. B. mit dem Verfahren nach ITU-T Rec. P.862 berechnet wurde.

[0010] Ausgehend von den beiden Sprachsignalen werden ein oder mehrere Intensitätskennwerte des Geräusches in Sprachpausen berechnet. Dies kann z. B. die mittlere Lautheit nach ISO 532 der Hintergrundgeräusche in Sprachpausen sein. Ein sehr effizientes Verfahren dafür ist in der Patentanmeldung DE 101 20 168 beschrieben. Auch andere Intensitätskennwerte (z.B. Schärfe, Impulshaftigkeit, Schwankungsstärke) können in den Korrekturwert einfließen. Es wird davon ausgegangen, dass zunehmende Intensitätskennwerte von einer ebenfalls zunehmenden Störung durch das Geräusch in Sprachpausen verursacht werden und damit zu einer stärkeren Verminderung der empfundenen Sprachqualität führen.

[0011] Die Sprachqualität wird bei guter Sprachwiedergabequalität und gleichzeitigen Geräuschen in den Sprachpausen von den beschriebenen Sprachqualitätsmessverfahren zu hoch bewertet. Ein oder mehrere Intensitätswerte des Hintergrundgeräusches in Sprach-

pausen werden dazu benutzt, den gemessenen Sprachqualitätswert zu korrigieren. Ausgehend davon, dass der gemessene Sprachqualitätswert auf einer Skala von 1 (niedrige Qualität) bis 5 (sehr gute Qualität) entsprechend der ITU-T Recommendation P.800 ("Methods for objective and subjective assessment of quality", ITU-T, Genf 1996), abgebildet wird, werden alle Werte oberhalb eines bestimmten Sprachqualitätsschwellwertes (z. B. bei dem Verfahren nach ITU-T Rec. P.862 oberhalb von 3,0) bei Auftreten von Hintergrundgeräuschen in Sprachpausen reduziert. Diese Reduktion ist von den Intensitätskennwerten des Hintergrundgeräusches, dem Anteil der Sprachpausen im Gesamtsignal PA und dem Sprachqualitätswert Y abhängig. Eine allgemeine Beschreibung dieser Korrektur ist für alle Sprachqualitätswerte Y oberhalb eines Sprachqualitätsschwellwertes YS , bei der nur ein Intensitätskennwert (hier N) berücksichtigt wird. Dies kann z. B. die mittlere Lautheit des Hintergrundgeräusches in Sprachpausen sein:

$$Y(\text{korrigiert}) = Y - ((Y - YS) * a(N) * PA).$$

[0012] Die Funktion $a(N)$ stellt eine eindeutige und steigende Wichtungsfunktion des Intensitätskennwertes N dar. Der korrigierte Sprachqualitätswert ist stets kleiner oder gleich dem unkorrigierten Wert. Die Korrektur fällt gering aus, wenn die Intensität des Geräusches gering ist (N ist klein), nur wenig Sprachpausen auftreten (PA klein) oder sich der Sprachqualitätswert in der Nähe des Sprachqualitätsschwellwertes befindet ($Y - YS$ klein). Stärker wird korrigiert, wenn starke Pausengeräusche bei ansonsten hoher Sprachqualität Y auftreten. In Analogie dazu, kann die Korrektur der Sprachqualitätswerte auch mit Berücksichtigung mehrerer Intensitätskennwerte (hier als Beispiel N , M und O) erfolgen:

$$Y(\text{korrigiert}) = Y - ((Y - YS) * a(N) * b(M) * c(O) * PA)$$

[0013] Die Intensitätskennwerte werden mit den Wichtungsfunktionen $a(N)$, $b(M)$ bzw. $c(O)$ gewichtet. Da von einer zunehmenden Störung bei steigenden Werten ausgegangen wird, sind hier auch eindeutige und steigende Wichtungsfunktionen anzusetzen.

Ausführungsbeispiel für die Korrektur von Sprachqualitätswerten, die mit dem Verfahren "PESQ" nach ITU-T P.862 bestimmt wurden

[0014] Das hier vorgestellte Ausführungsbeispiel zeigt beispielhaft eine Korrektur der mit dem Verfahren nach ITU-T Rec. P.862 "PESQ" (Stand 2001) ermittelten Sprachqualitätswerte. Dieses Verfahren liefert durch den Vergleich eines ungestörten Quellsprachsignals mit dem zu bewertenden gestörten Sprachsignal einen

Sprachqualitätswert. Diese beiden Sprachsignale dienen zur Bestimmung der mittleren Lautheit des Hintergrundgeräusches entsprechend Patentanmeldung DE 101 20 168. Der so berechnete Wert N in $sones$ wird zur Korrektur des mit dem Verfahren nach ITU-T Rec. P.862 "PESQ" (Stand 2001) berechneten Qualitätswertes benutzt. Die Korrektur des Qualitätswertes setzt erst ein, wenn N einen Wert von 1,3 $sones$ überschritten hat (unterer Schwellwert):

$$N1 = N - 1,3,$$

wenn $N1 < 0$, dann $N1 = 0$

wenn $N1 > 4$, dann $N1 = 4$.

[0015] Der Wert der Schwellenüberschreitung $N1$ wird für die weiteren Schritte benutzt. Negative $N1$ werden nicht zugelassen, sondern auf $N1 = 0$ gesetzt. Ebenso wird $N1$ bei $N1 = 4$ begrenzt.

[0016] In einem zweiten Schritt wird der Wert $N1$ mit einer Konstanten und dem Anteil der Sprachpausen im Gesamtsignal PA ($PA = 0 \dots 1$) gewichtet:

$$N2 = 0,45 * N1 * PA$$

[0017] Im nächsten Schritt wird der Wert der Schwellenüberschreitung YD ermittelt, d.h. um welchen Anteil die gemessene Sprachqualität Y den Sprachqualitätsschwellwert YS überschreitet. Der dafür erforderliche Sprachqualitätsschwellwert YS ist gleich 3,0:

$$YD = Y - YS,$$

wenn $YD < 0$, dann $YD = 0$

[0018] Der Korrekturwert YK , der die Reduktion des Sprachqualitätswertes durch das Hintergrundgeräusch in Sprachpausen beschreibt, ist der mit der Schwellenwertüberschreitung gewichtete Intensitätswert $N2$:

$$YK = YD * N2$$

[0019] Dieser Korrekturwert wird vom ursprünglichen Y subtrahiert und damit ergibt sich der korrigierte Wert für die Sprachqualität $Y(\text{korrigiert})$:

$$Y(\text{korrigiert}) = Y - YK.$$

Patentansprüche

1. Verfahren zur Korrektur von gemessenen Sprachqualitätswerten, wobei die Sprachqualitätsmesswerte mittels instrumenteller Verfahren ermittelt werden, die ausschließlich die Signalabschnitte mit

Sprachaktivität auswerten,

dadurch gekennzeichnet,

dass die Geräusche in den Sprachpausen bei der Bestimmung der Sprachqualitätswerte berücksichtigt werden,

5

dass dafür die Intensität der Hintergrundgeräusche in den Sprachpausen bestimmt wird und diese Kennwerte für die Korrektur der gemessenen Sprachqualitätswerte verwendet werden, und

dass die Korrektur in der Weise erfolgt, dass der gemessene Sprachqualitätswert um einen Betrag reduziert wird, der aus einem oder mehreren gewichteten Intensitätskennwerten des Hintergrundgeräusches, dem Anteil an Sprachpausen im zu bewertenden Sprachsignal und dem gemessenen Sprachqualitätswert berechnet wird.

10

15

2. Verfahren nach Anspruch 1, **dadurch gekennzeichnet, dass** die Korrektur für nach dem ITU-T Standard P.862 gemessene Sprachqualitätswerte in der Weise erfolgt, dass als Intensitätskennwert die mittlere Lautheit in Sprachpausen bestimmt wird und für die Korrekturberechnung der Betrag berücksichtigt wird, der den Schwellwert 1,3 sone überschreitet und auf maximal 4,0 sone begrenzt wird und dieser Wert mit einem konstanten Faktor (0,45), dem Anteil der Sprachpausen im Gesamtsignal und mit dem Anteil des gemessenen Sprachqualitätswertes multipliziert wird, der den Sprachqualitätsschwellwert von 3,0 übersteigt und einen Korrekturwert ergibt, der zur Ermittlung des korrigierten Sprachqualitätswertes vom gemessenen Sprachqualitätswert subtrahiert wird.

20

25

30

35

40

45

50

55