(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) **EP 1 386 307 B2**

(12) **NEUE EUROPÄISCHE PATENTSCHRIFT**
Nach dem Einspruchsverfahren

(45) Veröffentlichungstag und Bekanntmachung des Hinweises auf die Entscheidung über den Einspruch:
17.04.2013 Patentblatt 2013/16

(45) Hinweis auf die Patenterteilung:
09.02.2005 Patentblatt 2005/06

(21) Anmeldenummer: **02703438.8**

(22) Anmeldetag: **19.03.2002**

(51) Int Cl.:
G10L 19/00 *(2013.01)*

(86) Internationale Anmeldenummer:
PCT/CH2002/000164

(87) Internationale Veröffentlichungsnummer:
WO 2002/075725 (26.09.2002 Gazette 2002/39)

(54) VERFAHREN UND VORRICHTUNG ZUR BESTIMMUNG EINES QUALITÄTSMASSES EINES AUDIOSIGNALS

METHOD AND DEVICE FOR DETERMINING A QUALITY MEASURE FOR AN AUDIO SIGNAL

PROCEDE ET DISPOSITIF POUR DETERMINER UN NIVEAU DE QUALITE D'UN SIGNAL AUDIO

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

(30) Priorität: **20.03.2001 EP 01810285**

(43) Veröffentlichungstag der Anmeldung:
04.02.2004 Patentblatt 2004/06

(73) Patentinhaber: **SwissQual License AG**
4528 Zuchwil (CH)

(72) Erfinder:
- **JURIC, Pero**
  CH-4513 Langendorf (CH)
- **THOMET, Bendicht**
  CH-3020 Riedbach (CH)

(74) Vertreter: **Schalch, Rainer et al**
c/o E. Blum & Co. Patentanwälte
Vorderberg 11
8044 Zürich (CH)

(56) Entgegenhaltungen:
EP-A- 0 644 526 **WO-A-00/72453**
US-A- 5 583 968

- **SEOK JONG WON ET AL: "Speech enhancement with reduction of noise components in the wavelet domain" PROCEEDINGS OF THE 1997 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, ICASSP. PART 2 (OF 5);MUNICH, GER APR 21-24 1997, Bd. 2, 1997, Seiten 1323-1326, XP002170620 ICASSP IEEE Int Conf Acoust Speech Signal Process Proc;ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings; Speech Processing 1997 IEEE, Piscataway, NJ, USA**
- **LIANG J ET AL: "OUTPUT-BASED OBJECTIVE SPEECH QUALITY" PROCEEDINGS OF THE VEHICULAR TECHNOLOGY CONFERENCE,US, NEW YORK, IEEE, Bd. CONF. 44, 8. Juni 1994 (1994-06-08), Seiten 1719-1723, XP000497716 ISBN: 0-7803-1928-1**
- **HAUENSTEIN M ET AL: "INSTRUMENTELLE SPRACHGUETEBEURTEILUNG" FUNKSCHAU, DE,FRANZIS-VERLAG K.G. MUNCHEN, Bd. 71, Nr. 3, 23. Januar 1998 (1998-01-23), Seiten 61-64, XP000765678 ISSN: 0016-2841**

EP 1 386 307 B2

Beschreibung

Technisches Gebiet

[0001] Die Erfindung betrifft ein Verfahren zur Bestimmung eines Qualitätsmasses eines Audiosignals. Weiter betrifft die Erfindung eine Vorrichtung zur Durchführung dieses Verfahrens sowie ein Rauschunterdrückungsmodul und ein Unterbruchdetektions- und interpolationsmodul zur Verwendung in einer derartigen Vorrichtung.

Stand der Technik

[0002] Die Beurteilung der Qualität eines Telekommunikationsnetzes ist ein wichtiges Instrument zur Erreichung bzw. Erhaltung einer gewünschten Service-Qualität. Eine Möglichkeit, die Service-Qualität eines Telekommunikationsnetzes zu beurteilen besteht darin, die Qualität eines über das Telekommunikationsnetz übertragenen Signals zu bestimmen. Bei Audiosignalen, insbesondere bei Sprachsignalen sind hierfür verschiedene intrusive Verfahren bekannt. Bei derartigen Verfahren wird, wie der Name schon sagt, in das zu testende System eingegriffen, indem ein Übertragungskanal belegt und darin ein Referenzsignal übermittelt wird. Die Qualitätsbeurteilung erfolgt anschliessend durch einen Vergleich des bekannten Referenzsignals mit dem empfangenen Signal beispielsweise subjektiv durch eine oder eine Mehrzahl von Testpersonen. Dies ist jedoch aufwändig und damit teuer.

[0003] In der EP 0 980 064 ist ein weiteres intrusives Verfahren zur maschinengestützten Qualitätsbeurteilung eines Audiosignals beschrieben, wobei zur Beurteilung der Übertragungsqualität ein spektraler Ähnlichkeitswert des bekannten Quellsignals und des Empfangssignals bestimmt wird. Dieser Ähnlichkeitswert beruht auf einer Berechnung der Kovarianz der Spektren des Quellsignals und des Empfangssignal und einer Division der Kovarianz durch die Standardabweichungen der beiden genannten Spektren.

[0004] Intrusive Methoden haben generell jedoch den Nachteil, dass wie bereits erwähnt in das zu testende System eingegriffen werden muss. Zur Bestimmung der Signalqualität muss nämlich mindestens ein Übertragungskanal belegt und darin ein Referenzsignal übermittelt werden. Dieser Übertragungskanal kann während dieser Zeit nicht für eine Datenübermittlung verwendet werden. Zudem ist es bei einem Broadcastingsystem wie beispielsweise einem Rundfunkdienst prinzipiell zwar möglich, die Signalquelle zur Übermittlung von Testsignalen zu belegen, da damit aber sämtliche Kanäle besetzt und das Testsignal zu allen Empfängern übermittelt würde, ist dieses Vorgehen äusserst unpraktisch. Intrusive Verfahren sind ebenso ungeeignet, um gleichzeitig die Qualität einer Vielzahl von Übertragungskanälen zu überwachen.

[0005] EP-A-644 526 offenbart ein nicht-intrusives Verfahren zur Geräuschreduktion, welches zur Berechnung der gewünschten Signalinformation eine Schätzung der Rauschenergie verwendet. US-A-5 848 384 zeigt ein Verfahren und eine Vorrichtung zur Bestimmung der Qualität eines Audiosignals.

Darstellung der Erfindung

[0006] Aufgabe der Erfindung ist es, ein Verfahren der oben genannten Art anzugeben, welches die Nachteile des Standes der Technik vermeidet und insbesondere eine Möglichkeit bietet zur Beurteilung der Signalqualität eines über ein Telekommunikationsnetz übertragenen Signals ohne Kenntnis des ursprünglich gesendeten Signals.

[0007] Die Lösung der Aufgabe ist durch die Merkmale des Verfahrensanspruchs 1 und des Vorrichtungsanspruchs definiert. Bei dem erfindungsgemässen Verfahren zur maschinengestützten Bestimmung eines Qualitätsmasses eines Audiosignals wird aus dem Audiosignal zunächst ein Referenzsignal ermittelt. Mittels Vergleichen des ermittelten Referenzsignals mit dem Audiosignal wird ein Qualitätswert bestimmt, der zur Bestimmung des Qualitätsmasses verwendet wird.

[0008] Das erfindungsgemässe Verfahren erlaubt somit eine Beurteilung der Qualität eines Audiosignals an einem beliebigen Anschluss des Telekommunikationsnetzwerkes. D. h. es erlaubt damit auch die Qualitätsbeurteilung von vielen Übertragungskanälen gleichzeitig, wobei sogar eine gleichzeitige Beurteilung sämtlicher Kanäle möglich wäre. Die Qualitätsbeurteilung erfolgt hierbei allein aufgrund der Eigenschaften des empfangenen Signals, d. h. ohne Kenntnis des Quellsignals oder der Signalquelle.

[0009] Die Erfindung ermöglicht somit nicht nur eine Überwachung der Übertragungsqualität des Telekommunikationsnetzwerkes, sondern beispielsweise auch eine qualitätsbasierte Kostenverrechnung, ein qualitätsbasiertes Routing im Netz, ein Test des Deckungsgrades beispielsweise bei Mobilfunknetzen, eine QOS (Quality of Service) Steuerung der Netzknoten oder ein Qualitätsvergleich innerhalb eines Netzes oder auch netzübergreifend.

[0010] Ein über ein Telekommunikationsnetz übertragenes Audiosignal weist neben der gewünschten Signalinformation typischerweise auch unerwünschte Komponenten wie beispielsweise verschiedene Rauschanteile auf, welche im ursprünglichen Quellsignal nicht vorhanden waren.

[0011] Um eine möglichst gute Qualitätsbeurteilung durchführen zu können, ist eine möglichst gute Schätzung des ursprünglich gesendeten Signals notwendig. Um dieses Referenzsignal zu rekonstruieren, gibt es verschiedene Methoden. Eine Möglichkeit besteht darin, eine Schätzung der Charakteristika des Übertragungskanals zu bestimmen und ausgehend vom empfangenen Signal quasi rückwärts zu rechnen. Eine weitere Möglichkeit besteht in einer direkten Schätzung des Referenzsignals anhand der bekannten Informationen über

das Empfangssignal und den Übertragungskanal.

[0012] Bei der vorliegend angewandten Methode wird das Referenzsignal ermittelt, indem die im empfangenen Signal vorhandenen Störsignalanteile geschätzt und anschliessend aus dem empfangenen Signal entfernt werden. Indem die Rauschanteile aus dem Audiosignal entfernt werden, wird zunächst ein entrauschtes Audiosignal bestimmt, welches bevorzugt als Referenzsignal zur Beurteilung der Übertragungsqualität verwendet wird.

[0013] Es gibt verschiedene Methoden, Rauschanteile aus dem empfangenen Audiosignal zu entfernen. Das Audiosignal könnte beispielsweise über entsprechende Filter geführt werden. Bei einer bevorzugten Methode, die Rauschanteile aus dem Audiosignal zu entfemen, wird hierfür jedoch ein neuronales Netzwerk verwendet.

[0014] Das Audiosignal wird jedoch nicht direkt als Eingangssignal verwendet. Zunächst wird auf das Audiosignal eine diskrete Wavelet Transformation (DWT) angewendet. Diese Transformation liefert eine Mehrzahl von DWT-Koeffizienten des Audiosignals, welche dem neuronalen Netzwerk als Eingangssignal zugeführt werden. Das neuronale Netzwerk liefert am Ausgang eine Mehrzahl von korrigierten DWT-Koeffizienten, aus welchen mit der inversen DWT das Referenzsignal gewonnen wird. Dieses entspricht der entrauschten Version des Audiosignals.

[0015] Um dies zu erreichen, müssen die Koeffizienten des neuronalen Netzwerkes derart eingestellt sein, dass dieses zu den DWT-Koeffizienten eines rauschbehafteten Eingangssignals die DWT-Koeffizienten des entsprechenden entrauschten Eingangssignals liefert. Damit das neuronale Netzwerk die gewünschten Koeffizienten liefert, muss es zuvor mit einem Set von korrespondierenden rauschbehafteten bzw. entrauschten Signalpaaren trainiert werden.

[0016] Auf diese Weise lässt sich sowohl stationäres Rauschen wie beispielsweise weisses, thermisches sowie Fahrzeug- oder Strassenrauschen, als auch Impulsrauschen unterdrücken. Auch Echostörungen und Interferenzen lassen sich mit dem neuronalen Netzwerk unterdrücken bzw. beseitigen.

[0017] Bei der Bestimmung des Qualitätsmasses können neben dem Qualitätswert, der durch den Vergleich des empfangenen Audiosignals mit dem daraus ermittelten Referenzsignal ermittelt wird, auch beliebige andere Informationen berücksichtigt werden. Dies können sowohl im Audiosignal enthaltene Informationen, als auch Informationen über den Übertragungskanal oder das Telekommunikationsnetz selber sein.

[0018] Es ist von Vorteil, bei der Bestimmung des Qualitätsmasses Informationen zu verwenden, welche sich mit geeigneten Mitteln aus dem empfangenen Audiosignal selber gewinnen lassen. So wird die Qualität des empfangenen Audiosignal beispielsweise durch die bei der Übermittlung durchlaufenen Codec's (Coder - Decoder) beeinflusst. Es ist schwierig, derartige Signal-Degradationen festzustellen, denn beispielsweise bei zu kleinen Codec-Bitraten geht ein Teil der ursprünglichen Signalinformation verloren. Allerdings haben zu kleine Codec-Bitraten eine Veränderung der Grundfrequenz (Pitch) des Audiosignals zur Folge, weshalb mit Vorteil der Verlauf und die Dynamik der Grundfrequenz im Audiosignal untersucht wird. Da sich solche Änderungen am einfachsten anhand von Audiosignalabschnitten mit Vokalen untersuchen lassen, werden zunächst vorzugsweise Signalanteile im Audiosignal mit Vokalen detektiert und danach auf Pitch-Variationen hin untersucht.

[0019] Zurück zur Ermittlung des Referenzsignals aus dem empfangenen Audiosignal. Dieses kann nämlich nicht nur unerwünschte Signalanteile aufweisen, es können unterwegs auch teilweise gewünschte Informationen verloren gegangen sein. So kann das empfangene Audiosignal beispielsweise mehr oder weniger lange Signalunterbrüche aufweisen.

[0020] Je näher nun aber das aus dem Audiosignal generierte Referenzsignal beim ursprünglichen Quellsignal liegt, desto präziser ist die Beurteilung der Übertragungsqualität. Dies ist der Grund dafür, Signalunterbrüche durch geeignete Signale zu ersetzen. Hierfür könnten beispielsweise geeignete Rauschsignale oder auch bereits übermittelte Signalabschnitte verwendet werden.

[0021] Um jedoch eine möglichst genaue Schätzung des Referenzsignals zu erhalten, ist es von Vorteil, derartige Signalunterbrüche im Audiosignal zunächst zu detektieren und danach die fehlenden Signalabschnitte durch möglichst genaue, durch Interpolation erreichte Schätzungen zu ersetzen. Die Art der Interpolation der verlorengegangenen Signalabschnitte hängt hierbei ab von der Länge des Signalunterbruches. Bei kurzen Unterbrüchen, d. h. bei Unterbrüchen bis zu einigen wenigen Abtastwerten im Audiosignal wird bevorzugt eine polynomische und bei mittellangen Unterbrüchen, d. h. von einigen wenigen bis einigen Dutzend Abtastwerten wird bevorzugt eine modellbasierte Interpolation verwendet.

[0022] Längere Signalunterbrüche, d. h. Unterbrüche ab einigen Dutzend Abtastwerten, können jedoch kaum sinnvoll rekonstruiert werden. Anstatt diese Informationen als überflüssig zu betrachten und zu verwerfen, werden sie und teilweise auch die Informationen über die kurzen und mittellangen Signalunterbrüche vorzugsweise bei der Beurteilung der Übertragungsqualität berücksichtigt. Sie fliessen bei der Bestimmung des Qualitätsmasses mit in die Berechnungen ein.

[0023] Das empfangene Audiosignal kann verschiedene Arten von Audiosignalen umfassen. So kann es beispielsweise Sprach-, Musik-, Rausch- oder auch Ruhesignalanteile beinhalten. Die Qualitätsbeurteilung kann natürlich anhand der gesamten oder anhand eines Teils dieser Signalanteile erfolgen. Bei einer bevorzugten Variante der Erfindung wird die Beurteilung der Signalqualität hingegen beschränkt auf die Sprachsignalanteile. Mit einem Audio-Diskriminator werden aus dem Audiosignal daher zunächst die Sprachsignalanteile extrahiert und nur diese Sprachsignalanteile zur Bestimmung des Qualitätsmasses, d. h. zur Ermittlung des Referenzsignals verwendet. Um den Qualitätswert zu bestimmen

wird in diesem Fall das ermittelte Referenzsignal natürlich auch nicht mit dem empfangenen Audiosignal, sondern nur mit dem daraus extrahierten Sprachsignalanteil verglichen.

[0024] Die erfindungsgemässe Vorrichtung zur maschinengestützten Bestimmung eines Qualitätsmasses eines Audiosignals umfasst erste Mittel zur Bestimmung eines Referenzsignals aus dem Audiosignal, zweite Mittel zur Bestimmung eines Qualitätswertes mittels Vergleichen des ermittelten Referenzsignals mit dem Audiosignal sowie dritte Mittel zur Bestimmung des Qualitätsmasses unter Berücksichtigung des Qualitätswertes.

[0025] Die ersten Mittel zur Bestimmung eines Referenzsignals aus dem Audiosignal können mehrere Module umfassen. So ist vorzugsweise ein Rauschunterdrückungsmodul und/oder ein Unterbruchdetektions- und interpolationsmodul vorgesehen.

[0026] Mit dem Rauschunterdrückungsmodul lassen sich Rauschsignalanteile im empfangenen Audiosignal unterdrücken. Es beinhaltet die Mittel zur Durchführung der bereits beschriebenen Wavelet-Transformationen sowie das neuronale Netz zur Bestimmung der neuen DWT-Koeffizienten. Das Unterbruchdetektions- und interpolationsmodul weist diejenigen Mittel auf, welche einerseits zur Detektion von Signalunterbrüchen im Audiosignal und andererseits zur polynomischen Interpolation von kurzen sowie zur modellbasierten Interpolation von mittellangen Signalunterbrüchen benötigt werden. Das dermassen ermittelte Referenzsignal entspricht somit einer entrauschten Version des empfangenen Audiosignals und weist typischerweise nur noch grössere Signalunterbrüche auf.

[0027] Die Informationen über die Signalunterbrüche des Audiosignals werden jedoch nicht nur zur Ermittlung eines besseren Referenzsignals verwendet, sie können auch zur Bestimmung eines besseren Qualitätsmasses verwendet werden. Die dritten Mittel zur Bestimmung des Qualitätsmasses sind deshalb bevorzugt derart ausgebildet, dass Informationen über Signalunterbrüche im Audiosignal berücksichtigt werden können.

[0028] Je mehr Informationen über das Audiosignal bei der Bestimmung des Qualitätsmasses einbezogen werden, umso genauer kann die Qualitätsbeurteilung erfolgen. Die Vorrichtung weist daher mit Vorteil vierte Mittel zur Bestimmung von Informationen über Codec-bedingte Signalverzerrungen auf. Diese umfassen beispielsweise ein Vokaldetektionsmodul, mit welchem sich im Audiosignal Signalanteile mit Vokalen detektieren lassen. Diese Vokal-Signalanteile werden an ein Bewertungsmodul weitergegeben, welches anhand dieser Signalanteile Informationen über Codec-bedingte Signalverzerrungen bestimmt, welche ebenfalls zur Beurteilung der Signalqualität verwendet werden. Die dritten Mittel sind entsprechend derart ausgebildet, dass diese Informationen über die Codec-bedingten Signalverzerrungen bei der Bestimmung des Qualitätsmasses berücksichtigt werden können.

[0029] Mit Vorteil wird jedoch nicht das gesamte Audiosignal, sondern nur dessen Sprachsignalanteile zur Qualitätsbeurteilung verwendet. Entsprechend dem bereits geschilderten Verfahren weist die Vorrichtung daher insbesondere fünfte Mittel zur Extraktion der Sprachsignalanteile aus dem Audiosignal auf. Dementsprechend wird zur Ermittlung des Referenzsignals nicht das Audiosignal selber, sondern nur dessen Sprachsignalanteil entrauscht und auf Unterbrüche hin untersucht. Ebenso wird natürlich nicht das Audiosignal, sondern nur dessen Sprachsignalanteil mit diesem Referenzsignal verglichen. Damit erfolgt die Bestimmung des Qualitätsmasses lediglich anhand der Informationen im Sprachsignalanteil, wobei die Informationen aus den restlichen Signalanteilen nicht berücksichtigt werden.

[0030] Aus der nachfolgenden Detailbeschreibung und der Gesamtheit der Patentansprüche ergeben sich weitere vorteilhafte Ausführungsformen und Merkmalskombinationen der Erfindung.

Kurze Beschreibung der Zeichnungen

[0031] Die zur Erläuterung des Ausführungsbeispiels verwendeten Zeichnungen zeigen:

- Fig. 1 ein schematisch dargestelltes Blockdiagramm des erfindungsgemässen Verfahrens;
- Fig. 2 das Rauschunterdrückungsmodul im Betriebszustand;
- Fig. 3 das Rauschunterdrückungsmodul im Trainingszustand;
- Fig. 4 das neuronale Netzwerk des Rauschunterdrükkungsmoduls und
- Fig. 5 ein Beispiel für ein Audiosignal mit einem Unterbruch.

[0032] Grundsätzlich sind in den Figuren gleiche Teile mit gleichen Bezugszeichen versehen.

Wege zur Ausführung der Erfindung

[0033] Figur 1 zeigt ein Blockdiagramm des erfindungsgemässen Verfahrens. Hierbei wird für ein Audiosignal 1 ein Qualitätsmass 2 bestimmt, welches beispielsweise auch zur Bewertung des benutzten (nicht dargestellten) Telekommunikationsnetzes verwendet werden kann. Unter dem Audiosignal 1 wird hier dasjenige Signal verstanden, welches ein Empfänger nach der Übertragung über das Telekommunikationsnetz empfängt. Dieses Audiosignal 1 stimmt nämlich typischerweise nicht mit dem vom (nicht dargestellten) Sender gesendeten Signal überein, denn auf dem Weg vom Sender zum Empfänger wird das Sendesignal auf vielfältige Art und Weise verändert. So durchläuft es beispielsweise verschiedene Module wie Sprachcoder und -decoder,

Multiplexer und Demultiplexer oder auch Sprachverbesserer und Echokompensatoren. Aber auch der Übertragungskanal selber kann einen grossen Einfluss auf das Signal haben, welche sich beispielsweise in Form von Interferenzen, Fading, Übertragungsab- oder unterbrüchen, Echogenerierung etc. äussern.

[0034] Des Audiosignal 1 enthält somit nicht nur gewünschte Signalanteile, d. h. das ursprüngliche Sendesignal, sondern auch unerwünschte Störsignalanteile. Es kann auch sein, dass Signalanteile des Sendesignals fehlen, d. h. während der Übertragung verloren gegangen sind.

[0035] Bei dem dargestellten Beispiel erfolgt die Beurteilung der Signalqualität jedoch nicht anhand des gesamten Audiosignals 1, sondern lediglich anhand des darin enthaltenen Sprachanteils. Das Audiosignal 1 wird zunächst mit einem Audio-Diskriminator 3 auf Sprachsignalanteile 4 hin untersucht. Gefundene Sprachsignalanteile 4 werden zur weiteren Verarbeitung weitergeleitet, wohingegen andere Signalanteile wie beispielsweise Musik 5.1, Pausen 5.2 oder starke Signalstörungen 5.3 aussortiert und anderweitig weiterverarbeitet oder verworfen werden können. Um diese Unterscheidung durchführen zu können, wird das Audiosignal 1 stückweise, d. h. zu Stückchen a jeweils etwa 100 ms bis 500 ms, an den Audio-Diskriminator 3 übergeben. Dieser zerlegt diese Stückchen weiter in einzelne Buffer von etwa 20 ms Länge, verarbeitet diese Buffer und ordnet sie dann jeweils einer der zu unterscheidenden Signalgruppen Sprachsignal, Musik, Pause oder starke Störung zu.

[0036] Der Audio-Diskriminator 3 verwendet zur Beurteilung der Signalstückchen beispielsweise eine LPC (linear predictive coding) Transformation, mit welcher die Koeffizienten eines dem menschlichen Sprachtrakt entsprechenden, adaptiven Filters berechnet werden. Die Zuordnung der Signalstückchen zu den verschiedenen Signalgruppen erfolgt anhand der Form der Übertragungs-Charakteristika dieses Filters.

[0037] Um die Qualität der Übertragung beurteilen zu können, wird aus diesem Sprachsignalanteil 4 nun ein Referenzsignal 6, d. h. eine möglichst gute Schätzung des vom Sender ursprünglich übermittelten Sendesignals, ermittelt. Diese Referenzsignal-Schätzung erfolgt mehrstufig.

[0038] In einer ersten Stufe, einem Rauschunterdrükkungsmodul 7, werden zunächst unerwünschte Signalanteile wie stationäres Rauschen oder Impulsstörungen aus dem Sprachsignalanteil 4 entfernt bzw. unterdrückt. Dies geschieht mit Hilfe eines neuronalen Netzwerkes, welches zuvor mittels einer Vielzahl von verrauschten Signalen als Eingang und jeweils der entsprechenden rauschfreien Version des Eingangssignals als Zielsignal trainiert worden ist. Das auf diese Weise erhaltene, entrauschte Sprachignal 11 wird an die zweite Stufe weitergeleitet.

[0039] In der zweiten Stufe, dem Unterbruchdetektions- und interpolationsmodul 8 werden Unterbrüche im Audiosignal 1 bzw. in dessen Sprachsignalanteil 4 detektiert und wenn möglich interpoliert, d. h. die fehlenden Samples werden durch geeignet geschätzte Werte ersetzt.

[0040] Im vorliegenden Beispiel erfolgt die Detektion von Signalunterbrüchen mittels einer Untersuchung von Diskontinuitäten der Signalgrundfrequenz (pitch-tracing). Die Interpolation wird in Abhängigkeit der Länge des detektierten Unterbruches vorgenommen. Bei kurzen Unterbrüchen, d. h. Unterbrüchen von wenigen Samples Länge wird eine polynomische Interpolation wie beispielsweise ein Lagrange-, Newton-, Hermite-, oder Cubic Spline-Interpolation angewendet. Bei mittellangen Unterbrüchen (einige wenige bis einige Dutzend Samples) werden modellbasierte Interpolationen wie beispielsweise eine Maximum a posteriori-, eine autoregressive- oder eine frequency-time-Interpolation angewendet. Bei längeren Signalunterbrüchen ist eine Interpolation oder eine andere Signalrekonstruktion in der Regel nicht mehr auf sinnvolle Art und Weise möglich.

[0041] Das Ganze wird erschwert durch die Tatsache, dass es sowohl unterschiedliche Arten von Unterbrüchen - es ist zu unterscheiden zwischen Silben- bzw. Wortpausen und richtigen Signalunterbrüchen - als auch unterschiedliche Arten von Techniken zur Bearbeitung solcher Unterbrüche im Übertragungskanal gibt. So kann von einem Endgerät, beispielsweise in Abhängigkeit von Informationen über das Übertragungsnetz, unterschiedlich auf fehlende Frames reagiert werden. Bei einer ersten Methode werden verlorene Frames beispielsweise einfach durch Nullen ersetzt. Bei einer zweiten Methode werden anstelle der verlorenen Frames andere, richtig empfangene Frames eingesetzt und bei einer dritten Methode werden anstelle der verlorenen Frames lokal generierte Rauschsignale, sogenannter "comfort noise" eingesetzt.

[0042] Nach dem Ermitteln des Referenzsignals 6 mit dem Rauschunterdrückungsmodul 7 und dem Unterbruchdetektions- und interpolationsmodul 8 wird es mit Hilfe des Vergleichsmoduls 9 mit dem Sprachsignalanteil 4 verglichen. Für diesen Vergleich kann ein Algorithmus verwendet werden, wie er beispielsweise bei intrusiven Verfahren für den Vergleich des bekannten Quellsignals mit dem empfangenen Signal verwendet wird. Geeignet sind beispielsweise psychoakustische Modelle, die Signale perzeptiv, d. h. wahrnehmbar vergleichen. Das Resultat dieses Vergleichs ist ein intrusiver Qualitätswert 10. Zur Bestimmung dieses intrusiven Qualitätswertes 10 werden die Eingangssignale, also der Sprachsignalanteil 4 und das Referenzsignal 6, in Signalstücke von etwa 20 bis 30 ms Länge zerlegt und für jedes Signalstück ein Teilqualitätswert berechnet. Nach etwa 20 bis 30 Signalstücken, was etwa einer Signaldauer von 0.5 Sekunden entspricht, wird der intrusive Qualitätswert 10 als arithmethisches Mittel dieser Teilqualitätswerte ermittelt. Der intrusive Qualitätswert 10 bildet das Ausgangssignal des Vegleichsmoduls 9.

[0043] Bei der Bestimmung des Qualitätsmasses 2 können jedoch neben der Information über Störsignal-

anteile bzw. Signalunterbrüche auch noch andere Informationen über das Audiosignal 1 berücksichtigt werden. So kann beispielsweise ein Sprachcoder bzw. Sprachdecoder, den das gesendete Signal auf seinem Weg vom Sender zum Empfänger durchlaufen hat, einen Einfluss auf das Audiosignal 1 haben. Diese Einflüsse bestehen beispielsweise darin, dass sowohl die Grundfrequenz als auch die Frequenzen der höheren Harmonischen des Signals variieren. Je kleiner die Bitrate der verwendeten Sprachcodecs, desto grösser die Frequenzverschiebungen und damit die Signalverzerrungen.

[0044] Derartige Einflüsse lassen sich am einfachsten bei Vokalen untersuchen, weshalb das entrauschte Sprachsignal 11 zunächst einem Vokaldetektor 12 zugeführt wird. Dieser umfasst beispielsweise ein neuronales Netz, das vorher für die Erkennung von bestimmten (einzelne oder alle) Vokalen trainiert worden ist. Vokalsignale 13, d. h. Signalanteile welche das neuronale Netz als Vokale erkennt, werden an ein Bewertungsmodul 14 weitergeleitet, andere Signalanteile werden verworfen.

[0045] Das Bewertungsmodul 14 teilt das Vokalsignal 13 in Signalstücke von etwa 30 ms auf und berechnet daran jeweils eine DFT (diskrete Fourier Transformation) mit einer Frequenzauflösung von ungefähr 2 Hz bei einer Abtastfrequenz von etwa 8 kHz. Damit lassen sich dann die Grundfrequenz sowie die Frequenzen der höheren Harmonischen bestimmen und auf Variationen hin untersuchen. Ein weiteres Merkmal zur Bewertung der Codec-bedingten Verzerrungen bildet die Dynamik des Signalspektrums, wobei eine kleinere Dynamik eine schlechtere Signalqualität bedeutet. Die Referenzwerte für die Dynamikbewertung werden für die einzelnen Vokale aus Beispielsignalen gewonnen. Aus den Informationen über den Einfluss von Codecs auf die Frequenzverschiebungen und die Spektrumdynamik des Audiosignals 1 bzw. des entrauschten Sprachsignals 11 wird ein Codec-Qualitätswert 15 abgeleitet.

[0046] Bei der Bestimmung des Qualitätsmasses 2 durch das Auswertemodul 16 wird zusätzlich zum intrusiven Qualitätswert 10 und zum Codec-Qualitätswert 15 auch ein Unterbruchs-Qualitätswert 17 berücksichtigt. Dieser Wert beinhaltet Informationen über die Länge und die Anzahl der vom Unterbruchdetektions- und interpolationsmodul 8 festgestellten Unterbrüche, wobei bei einem bevorzugten Ausführungsbeispiel der Erfindung nur die Informationen über die langen Unterbrüche berücksichtigt werden. Zusätzlich können natürlich auch weitere Qualitäts-Informationen 18 über das empfangene Audiosignal 1 bzw. das entrauschte Sprachsignal 11, welche mit anderen Modulen oder Untersuchungen ermittelt werden, in die Berechnungen des Qualitätsmasses 2 einfliessen.

[0047] Die einzelnen Qualitätswerte werden nun derart skaliert, dass sie im Zahlenbereich zwischen 0 und 1 liegen, wobei ein Qualitätswert von 1 eine unverminderte Qualität und Werte unter 1 eine entsprechend verminderte Qualität bezeichnen. Das Qualitätsmass 2 wird schliesslich als Linearkombination der einzelnen Qualitätswerte berechnet, wobei die einzelnen Gewichtungskoeffizienten experimentell bestimmt und derart festgelegt werden, dass ihre Summe 1 ergibt.

[0048] Stehen weitere qualitätsrelevante informationen über das Telekommunikationsnetz zur Verfügung oder treten neue Effekte in den Übertragungskanälen auf, ist es auf einfache Art und Weise möglich, weitere Module zur Berechnung von weiteren Qualitätswerten hinzuzufügen und bei der Bestimmung des Qualitätsmasses 2 in der beschriebenen Art und Weise zu berücksichtigen.

[0049] Im Folgenden werden anhand der Figuren 2 bis 5 einige der Module näher erläutert. Figur 2 zeigt das Rauschunterdrückungsmodul 7. Der Sprachsignalanteil 4 des Audiosignals 1 wird zunächst einer an sich bekannten DWT 19 (diskrete Wavelet Transformation) unterworfen. DWT's werden ähnlich wie DFT's zur Signalanalyse eingesetzt. Ein wesentlicher Unterschied ist jedoch, im Gegensatz zu den bei einer DFT verwendeten, zeitlich unbegrenzten und damit zeitlich nicht lokalisierten Sinus- bzw. Kosinus-wellenformen, der Einsatz von sogenannten Wavelets, d. h. zeitlich begrenzten und damit zeitlich lokalisierten Wellenformen mit Mittelwert 0.

[0050] Der Sprachsignalanteil 4 wird in Signalstücke von etwa 20 ms bis 30 ms unterteilt, welche jeweils der DWT 19 unterworfen werden. Das Resultat der DWT 19 ist ein Satz von DWT-Koeffizienten 20.1, welche als Eingangsvektor einem neuronalen Netz 20 eingespiesen werden. Dessen Koeffizienten wurden vorgängig so trainiert, dass sie zu einem gegebenen Satz von DWT-Koeffizienten 20.1 eines verrauschten Signals einen neuen Satz von DWT-Koeffizienten 20.2 der unverrauschten Version dieses Signals liefern. Dieser neue Satz von DWT-Koeffizienten 20.2 wird nun der IDWT 21, d. h. der zur DWT 19 inversen DWT unterworfen. Diese IDWT 21 liefert auf diese Weise eine mehrheitlich unverrauschte Version der Sprachsignalanteile 4, eben das gewünschte, entrauschte Sprachsignal 11.

[0051] Die Trainingskonfiguration des neuronalen Netzes 20 ist in Figur 3 dargestellt. Es wird mit Paaren von verrauschten und unverrauschten Versionen von Beispielsignalen trainiert. Ein unverrauschtes Beispielsignal 22.1 wird der DWT 19 unterworfen und es wird ein erster Satz 20.3 von DWT-Koeffizienten erhalten. Auch das verrauschte Beispielsignal 22.2 wird der gleichen DWT 19 unterworfen und ein zweiter Satz 20.4 von DWT-Koeffizienten generiert, der in das neuronale Netz 20 eingespiesen wird. Der Ausgangsvektor des neuronalen Netzes 20, die neuen DWT-Koeffizienten 20.5, wird in einem Komparator 23 mit dem ersten Satz 20.3 von DWT-Koeffizienten verglichen. Aufgrund der Unterschiede zwischen diesen beiden Sätzen von DWT-Koeffizienten erfolgt eine Korrektur 24 der Koeffizienten des neuronalen Netzes 20. Dieser Vorgang wird mit einer Vielzahl von Beispielsignal-Paaren wiederholt, sodass die Koeffizienten des neuronalen Netzes 20 die gewünschte Funktion immer präziser durchführen. Vorteilhafterweise werden für das Training des neuronalen Netzes 20 Beispielsi-

gnale 22.1, 22.2 verwendet, welche menschliche Laute aus verschiedenen Sprachen darstellen. Ebenso ist es von Vorteil, hierfür sowohl Frauen- als auch Männer- und Kinderstimmen zu verwenden. Die erwähnte Grösse der einzeln zu verarbeitenden Signalstücke von 20 ms bis 30 ms Dauer ist so gewählt, dass die Verarbeitung des Sprachsignalanteils 4 unabhängig von der Sprache und des Sprechers durchgeführt werden kann. Auch Sprechpausen und sehr ruhige Signalabschnitte werden trainiert, damit auch diese korrekt erkannt werden.

[0052] Bei dem vorliegenden Ausführungsbeispiel wurde als neuronales Netzwerk 20 ein Mehrschicht-Perceptron mit einer Eingangsschicht 25, einer verborgenen Schicht 26 und einer Ausgangsschicht 27 verwendet. Trainiert wurde das Perceptron mit einem Backpropagation-Algorithmus. Die Eingangsschicht 25 weist eine Mehrzahl von Eingangs-Neuronen 25.1, die verborgene Schicht 26 eine Mehrzahl von verborgenen Neuronen 26.1 und die Ausgangsschicht 27 eine Mehrzahl von Ausgangs-Neuronen 27.1 auf. Jedem Eingangs-Neuron 25.1 wird jeweils einer der DWT-Koeffizienten 20.1 der vorangegangenen DWT 19 zugeführt. Nachdem die Eingangssignale das neuronale Netzwerk durchlaufen haben, wobei die jeweiligen Werte mit den eingestellten Koeffizienten der jeweiligen Neuronen bestimmt und die Wertekombinationen in den einzelnen Neuronen berechnet werden, liefert jedes Ausgangs-Neuron 27.1 einen der neuen DWT-Koeffizienten 20.2. Wie bereits erwähnt, zerlegt der Audio-Diskriminator 3 die Signalstückchen in einzelne Buffer der Länge 20 ms. Bei einer Abtastrate von 8 kHz entspricht dies 160 Abtastwerten. Für diesen Fall kann beispielsweise ein neuronales Netz 20 mit je 160 Eingangs- und Ausgangs-Neuronen 25.1, 27.1 sowie etwa 50 bis 60 verborgenen Neuronen 26.1 verwendet werden.

[0053] Anhand der Figur 5 soll die Interpolation eines Signalunterbruches kurz beschrieben werden. Für die Signalrekonstruktion wird beispielsweise eine Zeit-Frequenz Interpolation angewendet. Hierzu wird zunächst ein Kurzzeitspektrum für Signatframes mit 64 Samples Länge (8 ms) berechnet. Dies geschieht, indem die Signalframes mit Hamming-Fenstem bei einer Überschneidung von 50% multipliziert werden.

[0054] Das Ziel der Interpolation ist die Behandlung dieser Lücke. Zunächst wird eine Frequenz-Zeit Transformation durchgeführt. Dies führt zu einer dreidimensionalen Signaldarstellung, welche für jeden Punkt in der Zeit-Frequenz Ebene (x-y Ebene) das Leistungsspektrum in Richtung der z-Achse liefert. Ein Unterbruch zu einem gegebenen Zeitpunkt t ist einfach zu erkennen als Nullpunkte entlang der Linie x = t in der Zeit-Frequenz Ebene.

[0055] Figur 5 zeigt ein derartiges Signal 28 von etwa 200 Samples Länge. Um die Periodizität einfacher erkennen zu können, zeigt Figur 5 das Signal 28 in der zeitlichen Domäne. Auf der Abszissenachse 32 sind die Anzahl Samples und auf der Ordinatenachse 33 die Magnituden aufgetragen. Die Interpolation erfolgt jedoch in der Frequenz-Zeit Domäne. In Figur 5 ist der Unterbruch 29 unschwer zu erkennen als Lücke von knapp 10 Samples Länge.

[0056] Für jeden Frequenzanteil erfolgt nun eine polynomische Interpolation sowohl für die Phase, als auch die Magnitude, wobei diese mit minimaler Phasen- und Magnitudendiskontinuität erfolgt. Hierfür wird zunächst wiederum die Pitch-Periode 30 des Signals 28 bestimmt. Für die Interpolation werden Information aus den Samples vor und nach der Lücke innerhalb dieser Pitch-Periode 30 berücksichtigt. Die Signalbereiche 31.1, 31.2 zeigen diejenigen Bereiche des Signals 28 je eine Pitch-Periode vor bzw. hinter dem Unterbruch 29. Diese Signalbereiche 31.1, 31.2 sind zwar nicht identisch mit dem ursprünglichen Signalstück beim Unterbruch 29, zeigen aber dennoch ein hohes Mass an Ähnlichkeit dazu. Für kleine Lücken bis etwa 10 Samples wird angenommen, dass noch genügend Signalinformation vorhanden ist, um eine korrekte Interpolation durchführen zu können. Bei längeren Lücken können zusätzliche Informationen aus Samples der Umgebung verwendet werden.

[0057] Zusammenfassend ist festzustellen, dass es die Erfindung erlaubt, die Signalqualität eines empfangenen Audiosignals zu beurteilen, ohne das ursprüngliche Sendesignal zu kennen. Aus der Signalqualität kann natürlich auch auf die Qualität der benutzten Übertragungskanäle und somit auf die Service-Qualität des gesamten Telekommunikationsnetzes geschlossen werden. Die schnellen Antwortzeiten des erfindungsgemässen Verfahrens, welche in der Grössenordnung von etwa 100 ms bis 500 ms liegen, ermöglichen somit verschiedenen Anwendungen wie beispielsweise generelle Vergleiche der Servicequalität verschiedener Netze oder Teilnetze, eine qualitätsbasierte Kostenverrechnung oder ein qualitätsbasiertes Routing in einem Netz oder über mehrere Netze hinweg mittels entsprechender Steuerung der Netzknoten (Gateways, Router etc.).

Patentansprüche

1. Verfahren zur maschinengestützten Bestimmung eines Qualitätsmasses eines Audiosignals, wobei aus dem Audiosignal ein Referenzsignal ermittelt wird, welches eine Schätzung eines ursprünglich gesendeten Audiosignals darstellt, und dass mittels Vergleichen des Referenzsignals mit dem Audiosignal ein Qualitätswert bestimmt wird, der zur Bestimmung des Qualitätsmasses verwendet wird, **dadurch gekennzeichnet, dass** mittels Entfernen von Rauschsignalanteilen aus dem Audiosignal ein entrauschtes Audiosignal ermittelt und dieses als Referenzsignal verwendet wird, und dass im entrauschten Audiosignal Signalanteile mit Vokalen detektiert, daraus Informationen über Codec-bedingte Signalverzerrungen ermittelt und diese bei der Bestimmung des Qualitätsmasses berücksichtigt werden.

2. Verfahren nach Anspruch 1, **dadurch gekennzeichnet, dass** das entrauschte Audiosignal ermittelt wird, indem das Audiosignal einer diskreten Wavelet Transformation unterworfen wird, deren Koeffizienten in ein zuvor trainiertes neuronales Netz eingespiesen und dessen Ausgangssignale der inversen, diskreten Wavelet Transformation unterworfen werden.

3. Verfahren nach einem der Ansprüche 1 oder 2, **dadurch gekennzeichnet, dass** Signalunterbrüche im Audiosignal detektiert und das Referenzsignal ermittelt wird, indem es bei den Signalunterbrüchen zumindest teilweise rekonstruiert wird, wobei das Referenzsignal bei kurzen Signalunterbrüchen vorzugsweise mit einer polynomischen und bei mittellangen Signalunterbrüchen vorzugsweise mit einer modellbasierten Interpolation rekonstruiert wird.

4. Verfahren nach Anspruch 3, **dadurch gekennzeichnet, dass** bei der Bestimmung des Qualitätsmasses Informationen über die Signalunterbrüche berücksichtigt werden.

5. Verfahren nach einem der Ansprüche 1 bis 4, **dadurch gekennzeichnet, dass** vor dem Ermitteln des Referenzsignals aus dem Audiosignal ein Sprachsignalanteil extrahiert und die Bestimmung des Qualitätsmasses auf den Sprachsignalanteil beschränkt wird.

6. Vorrichtung zur maschinengestützten Bestimmung eines Qualitätsmasses eines Audiosignals, welche erste Mittel zur Bestimmung eines Referenzsignals aus dem Audiosignal, zweite Mittel zur Bestimmung eines Qualitätswertes mittels Vergleichen des Referenzsignals mit dem Audiosignal sowie dritte Mittel zur Bestimmung des Qualitätsmasses unter Berücksichtigung des Qualitätswertes aufweist, wobei das Referenzsignal eine Schätzung eines ursprünglich gesendeten Audiosignals darstellt, **dadurch gekennzeichnet, dass** sie Mittel zum Entfernen von Rauschsignalanteilen aus dem Audiosignal und Mittel zur Bestimmung von Codec-bedingten Signalverzerrungen aufweist, wobei diese ein Vokaldetektionsmodul zur Detektion von Vokal-Signalanteilen im entrauschten Audiosignal sowie ein Bewertungsmodul zur Bestimmung der Codec-bedingten Signalverzerrungen umfassen, wobei die dritten Mittel derart ausgebildet sind, dass die Codec-bedingten Signalverzerrungen bei der Bestimmung des Qualitätsmasses berücksichtigbar sind.

7. Vorrichtung nach Anspruch 6, **dadurch gekennzeichnet, dass** die ersten Mittel ein Rauschunterdrückungsmodul zur Unterdrückung von Rauschsignalanteilen und/oder ein Unterbruchdetektions- und interpolationsmodul zur Detektion und Interpolation von Signalunterbrüchen im Audiosignal aufweisen, und die dritten Mittel derart ausgebildet sind, dass Signalunterbrüche bei der Bestimmung des Qualitätsmasses berücksichtigt werden können.

8. Vorrichtung nach einem der Ansprüche 6 oder 7, **dadurch gekennzeichnet, dass** sie Mittel zur Extraktion eines Sprachsignalanteils aus dem Audiosignal aufweist und zur Bestimmung des Qualitätsmasses des Sprachsignalanteils ausgebildet ist.

9. Vorrichtung nach Anspruch 7, wobei die ersten Mittel das Rauschunterdrückungsmodul aufweisen, **dadurch gekennzeichnet, dass** das Rauschunterdrückungsmodul Mittel zur Durchführung einer diskreten Wavelet-Transformation zur Berechnung von Signalkoeffizienten eines Audiosignals, ein neuronales Netz zur Berechnung von korrigierten Signalkoeffizienten sowie Mittel zur Durchführung einer inversen Wavelet-Transformation der korrigierten Signalkoeffizienten zur Bestimmung des Audiosignals ohne Rauschsignalanteile aufweist.

10. Vorrichtung nach Anspruch 7, wobei die ersten Mittel das Unterbruchdetektions- und Interpolationsmodul aufweisen, **dadurch gekennzeichnet, dass** das Unterbruchdetektions- und Interpolationsmodul Mittel zur Detektion von Signalunterbrüchen in einem Audiosignal sowie Mittel zur interpolation von Signalunterbrüchen des Audiosignals aufweist, wobei diese vorzugsweise zur polynomischen Interpolation von kurzen bzw. zur modellbasierten Interpolation von mittellangen Signalunterbrüchen ausgebildet sind.

Claims

1. A method for the machine-assisted determination of a measure of quality of an audio signal, in which a reference signal that represents an estimate of an audio signal originally transmitted is determined from the audio signal, and a quality value, which is used for determining the measure of quality, is determined by means of comparing the reference signal with the audio signal, **characterised in that** a de-noised audio signal is determined by removing noise signal components from the audio signal and is used as the reference signal, and that signal components with vowels are detected in the de-noised audio signal information on codec-related signal distortions is determined therefrom and is taken into consideration in determining the measure of quality.

2. A method according to claim 1, **characterised in that** the de-noised audio signal is determined by subjecting the audio signal to discrete wavelet transformation, feeding the coefficients of the latter into a

previously trained neural network and subjecting the output signals of the latter to inverse discrete wavelet transformation.

3. A method according to either of claims 1 and 2, **characterised in that** signal interruptions in the audio signal are detected and the reference signal is determined by at least partially reconstructing it in the case of the signal interruptions, the reference signal being reconstructed preferably by polynomial interpolation in the case of short signal interruptions and preferably by model-based interpolation in the case of medium-length signal interruptions.

4. A method according to claim 3, **characterised in that** information on the signal interruptions is taken into consideration in determining the measure of quality.

5. A method according to any one of claims 1 to 4, **characterised in that**, before the reference signal is determined, a speech signal component is extracted from the audio signal and the determination of the measure of quality is restricted to the speech signal component.

6. A device for the machine-assisted determination of a measure of quality of an audio signal, which has first means for determining a reference signal from the audio signal, second means for determining a quality value by comparing the reference signal with the audio signal, and third means for determining the measure of quality while taking the quality value into consideration, the reference signal representing an estimate of an audio signal originally transmitted, **characterised in that** it has means for removing noise signal components from the audio signal and means for determining codec-related signal distortions, the latter means including a vowel detection module for detecting vowel signal components in the de-noised audio signal and an evaluation module for determining the codec-related signal distortions, the third means being so designed that the codec-related signal distortions can be taken into consideration in determining the measure of quality.

7. A device according to claim 6, **characterised in that** the first means have a noise suppression module for suppressing noise signal components and/or an interruption detection and interpolation module for detection and interpolation of signal interruptions in the audio signal, and the third means are so designed that signal interruptions can be taken into consideration in determining the measure of quality.

8. A device according to either of claims 6 and 7, **characterised in that** it has means for extracting a speech signal component from the audio signal and is designed for the purpose of determining the measure of quality of the speech signal component.

9. A device according to claim 7, wherein the first means have the noise suppression module, **characterised in that** the noise suppression module has means for performing discrete wavelet transformation for calculating signal coefficients of an audio signal, a neural network for calculating corrected signal coefficients and means for performing inverse discrete wavelet transformation of the corrected signal coefficients for determining the audio signal without noise signal components.

10. A device according to claim 7, wherein the first means have the interruption detection and interpolation module, **characterised in that** the interruption detection and interpolation module has means for detecting signal interruptions in an audio signal and means for interpolating signal interruptions of the audio signal, the latter means preferably being designed for the purpose of polynomial interpolation of short signal interruptions and model-based interpolation of medium-length signal interruptions.

Revendications

1. Procédé pour la détermination, assistée par ordinateur, d'une mesure de qualité d'un signal audio, par lequel on détermine à partir du signal audio, un signal de référence qui représente une estimation d'un signal audio initialement émis, et par lequel l'on détermine, au moyen d'une comparaison du signal de référence au signal audio, une valeur de qualité qui est utilisée pour la détermination de la mesure de qualité, **caractérisé en ce que** l'on détermine un signal audio non bruité en éliminant des composantes de signal bruitées du signal audio et on l'utilise comme signal de référence, et **en ce que** l'on détecte dans le signal audio non bruité, des composantes de signal avec éléments vocaux, et l'on détermine des informations sur des distorsions de signal dues au codeur-décodeur et on prend en compte celles-ci lors de la détermination de la mesure de qualité.

2. Procédé selon la revendication 1, **caractérisé en ce que** l'on détermine le signal audio non bruité en soumettant le signal audio à une transformation d'ondelettes discrète dont les coefficients sont introduits dans un réseau neuronal ayant subi auparavant un apprentissage et dont les signaux de sortie sont soumis à la transformation d'ondelettes discrète inverse.

3. Procédé selon l'une des revendications 1 et 2, **caractérisé en ce que** l'on détecte des interruptions de signal dans le signal audio et **en ce que** l'on détermine le signal de référence en le reconstruisant

au moins partiellement au niveau des interruptions de signal, le signal de référence étant reconstruit de préférence avec une interpolation polynomiale en cas d'interruptions de signal courtes et de préférence avec une interpolation basée sur un modèle en cas d'interruptions de signal moyennement longues.

4. Procédé selon la revendication 3, **caractérisé en ce que**, lors de la détermination de la mesure de qualité, on prend en compte des informations sur les interruptions de signal.

5. Procédé selon l'une des revendications 1 à 4, **caractérisé en ce que**, avant la détermination du signal de référence, on extrait une composante de signal vocale du signal audio, et **en ce qu'**on limite la détermination de la mesure de qualité à la composante de signal vocale.

6. Dispositif pour la détermination d'une mesure de qualité, lequel présente des premiers moyens pour déterminer un signal de référence à partir du signal audio, des deuxièmes moyens pour déterminer une valeur de qualité au moyen d'une comparaison du signal de référence au signal audio et des troisièmes moyens pour déterminer la mesure de qualité en tenant compte de la valeur de qualité, grâce à quoi le signal de référence représente une estimation d'un signal audio émis à l'origine, **caractérisé en ce qu'**il présente des moyens pour éliminer du signal audio des composantes de signal bruitées et des moyens pour déterminer des distorsions de signal dues au codeur-décodeur, celles-ci comportant un module de détection d'éléments vocaux pour détecter des composantes de signal avec éléments vocaux dans le signal audio non bruité, ainsi qu'un module d'évaluation pour déterminer les distorsions de signal provoquées par le codeur-décodeur, les troisièmes moyens étant conçus de sorte que les distorsions de signal causées par le codeur-décodeur puissent être prises en considération lors de la détermination de la mesure de qualité.

7. Dispositif selon la revendication 6, **caractérisé en ce que** les premiers moyens comportent un module de suppression de bruit pour supprimer des composantes de signal bruitées et/ou un module de détection d'interruptions et d'interpolation pour détecter et interpoler des interruptions de signal dans le signal audio, et **en ce que** les troisièmes moyens sont conçus de telle sorte que des interruptions de signal peuvent être prises en compte lors de la détermination de la mesure de qualité.

8. Dispositif selon l'une des revendications 6 et 7, **caractérisé en ce qu'**il comporte des moyens pour extraire du signal audio une composante de signal vocale, et **en ce qu'**il est conçu pour la détermination de la mesure de qualité de la composante de signal vocale.

9. Dispositif selon la revendication 7, les premiers moyens comportant le module de suppression de bruit, **caractérisé en ce que** le module de suppression de bruit comporte des moyens pour la mise en oeuvre d'une transformation d'ondelettes discrète en vue du calcul de coefficients de signal d'un signal audio, un réseau neuronal en vue du calcul de coefficients de signal corrigés, ainsi que des moyens pour la mise en oeuvre d'une transformation d'ondelettes inverse des coefficients de signal corrigés en vue de la détermination du signal audio sans composantes de signal bruitées.

10. Dispositif selon la revendication 7, les premiers moyens comportant le module de détection d'interruptions et d'interpolation, **caractérisé en ce que** le module de détection d'interruptions et d'interpolation comporte des moyens pour détecter des interruptions de signal dans un signal audio ainsi que des moyens pour interpoler des interruptions du signal audio, ces derniers étant conçus de préférence pour une interpolation polynomiale d'interruptions de signal courtes et pour une interpolation, basée sur un modèle, d'interruptions de signal moyennement longues.

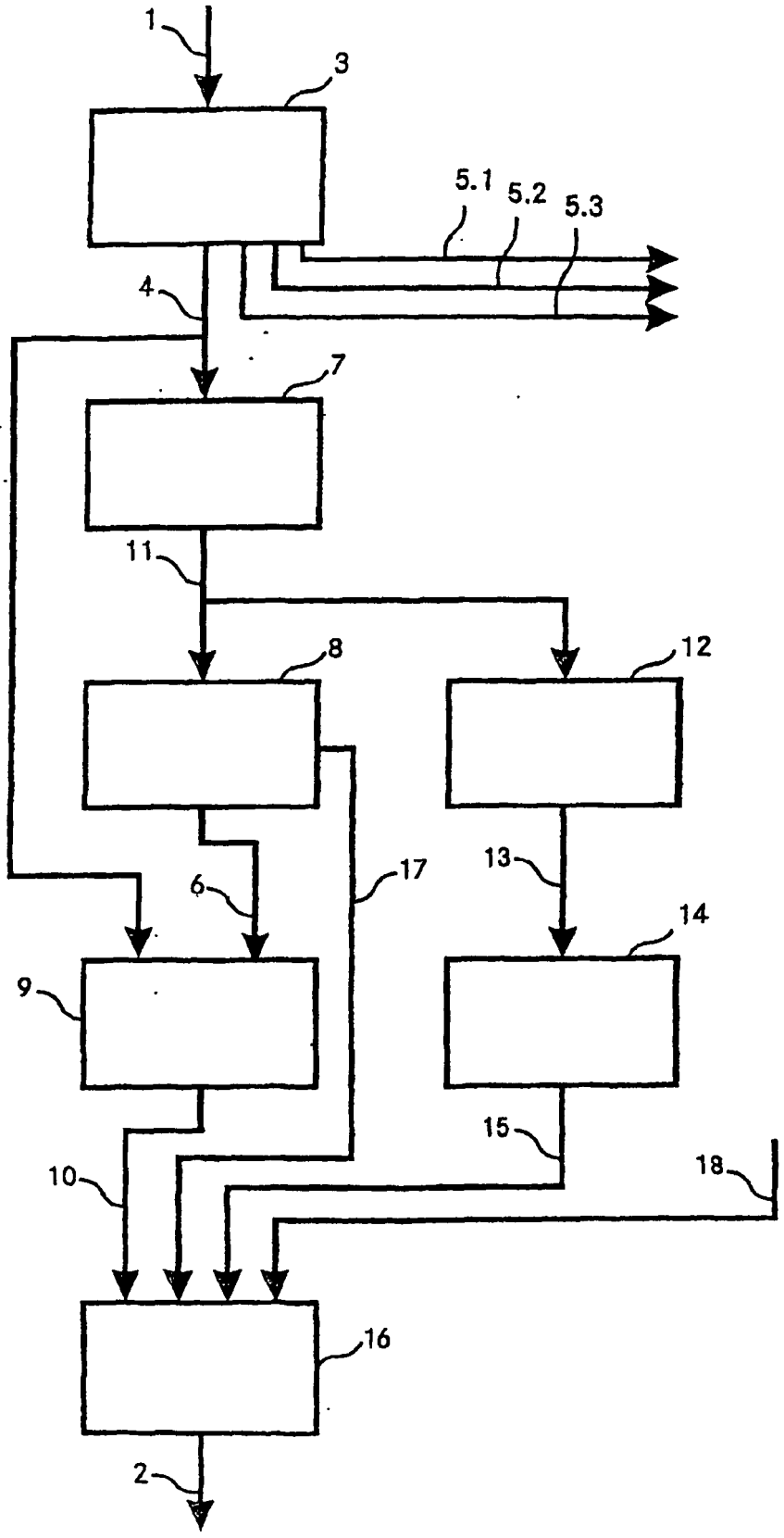

Fig. 1

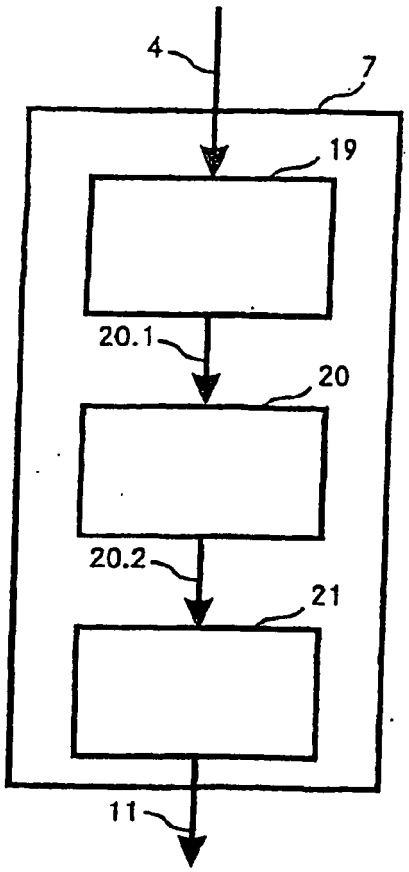

Fig. 2

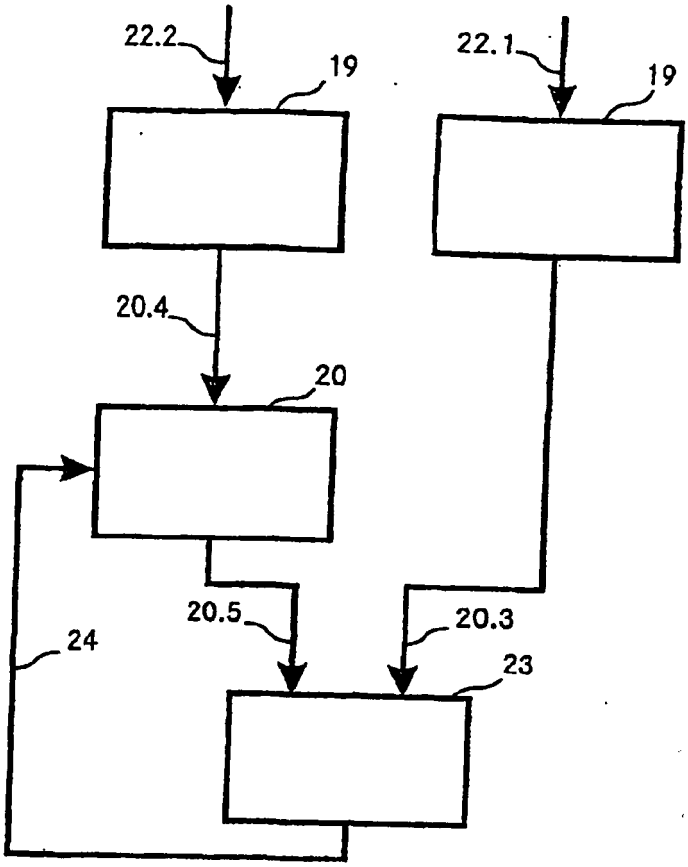

Fig. 3

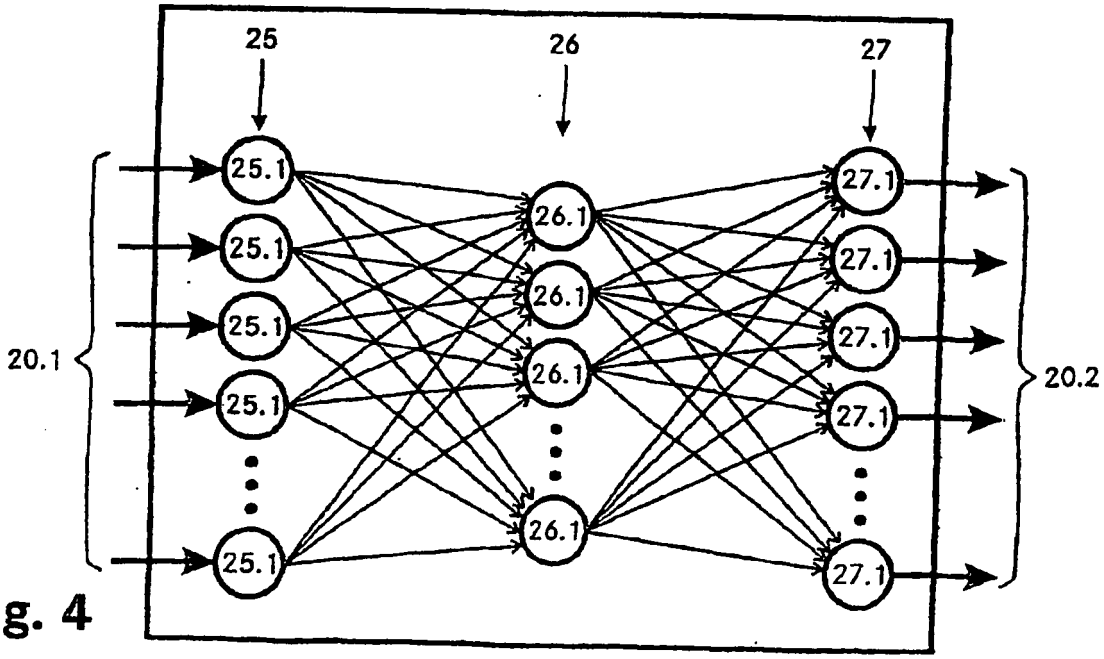

Fig. 4

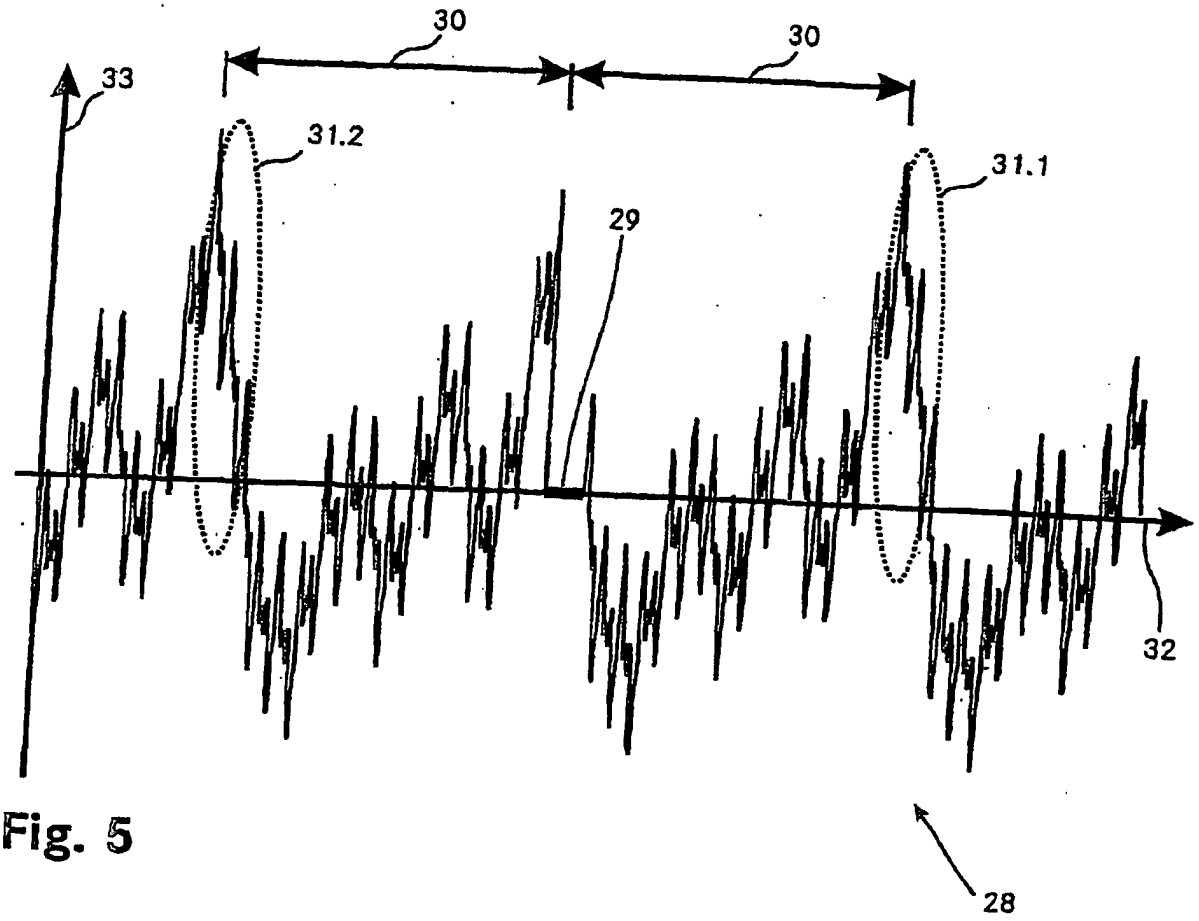

Fig. 5

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

- EP 0980064 A **[0003]**
- EP 644526 A **[0005]**
- US 5848384 A **[0005]**