

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4301514号
(P4301514)

(45) 発行日 平成21年7月22日(2009.7.22)

(24) 登録日 平成21年5月1日(2009.5.1)

(51) Int.Cl.

F I

G 1 0 L 11/00 (2006.01)
H 0 4 M 1/24 (2006.01)G 1 0 L 11/00 4 0 4
H 0 4 M 1/24 F

請求項の数 13 (全 9 頁)

(21) 出願番号 特願2004-517987 (P2004-517987)
 (86) (22) 出願日 平成15年6月27日(2003.6.27)
 (65) 公表番号 特表2005-531990 (P2005-531990A)
 (43) 公表日 平成17年10月20日(2005.10.20)
 (86) 国際出願番号 PCT/US2003/020354
 (87) 国際公開番号 W02004/003499
 (87) 国際公開日 平成16年1月8日(2004.1.8)
 審査請求日 平成18年3月30日(2006.3.30)
 (31) 優先権主張番号 10/186,862
 (32) 優先日 平成14年7月1日(2002.7.1)
 (33) 優先権主張国 米国 (US)

前置審査

(73) 特許権者 596092698
 アルカテルルーセント ユーエスエー
 インコーポレーテッド
 アメリカ合衆国 07974 ニュージャ
 ーシー, マレイ ヒル, マウンテン アヴ
 ェニュー 600-700
 (73) 特許権者 305029944
 キム, ドーースク
 アメリカ合衆国 07920 ニュージャ
 ーシー, バスキングリッジ, ハンティング
 トン ロード 42
 (74) 代理人 100064447
 弁理士 岡部 正夫
 (74) 代理人 100085176
 弁理士 加藤 伸晃

最終頁に続く

(54) 【発明の名称】 音声品質を評価する方法

(57) 【特許請求の範囲】

【請求項 1】

音声品質を評価する方法であって、

第 1 音声信号および第 2 音声信号についての第 1 音声品質評価および第 2 音声品質評価を、前記第 1 音声および第 2 音声信号について明瞭発音電力と非明瞭発音電力とを比較することにより決定する工程であって、前記第 2 音声信号が処理された音声信号であり、および前記第 1 音声信号が前記第 2 音声信号のひずんだバージョンであり、前記明瞭発音電力および非明瞭発音電力が、前記第 1 音声信号および第 2 音声信号の明瞭発音周波数および非明瞭発音周波数に関連付けられている工程と、

前記第 1 音声品質評価と前記第 2 音声品質評価とを比較して、補償された音声品質評価を得る工程であって、この補償された音声品質評価が前記第 1 音声品質評価と前記第 2 音声品質評価の間の差に相当するものである工程とを含み、

明瞭発音電力と非明瞭発音電力とを比較する前記処理は、

(i) 複数の臨界帯域信号を得るために、前記第 1 音声信号および第 2 音声信号をろ波する処理、(i i) 複数の変調スペクトルを得るために、前記複数の臨界帯域信号について包絡線分析を実施する処理、および(i i i) 前記複数の変調スペクトルの各々に対してフーリエ変換を実施する処理を含む方法。

【請求項 2】

前記第 1 音声品質評価および前記第 2 音声品質評価を決定する前に、前記第 2 音声信号をひずませて、前記第 1 音声信号を生成する工程をさらに含む、請求項 1 に記載の方法。

10

20

【請求項 3】

前記第 1 音声品質評価および前記第 2 音声品質評価が、客観的音声品質評価のための同一技法を使用して決定される、請求項 1 に記載の方法。

【請求項 4】

前記補償された音声品質評価が、前記第 1 音声品質評価と前記第 2 音声品質評価との比に相当する、請求項 1 に記載の方法。

【請求項 5】

前記第 1 音声品質評価および前記第 2 音声品質評価が、音声の聴覚明瞭度分析を使用して決定される、請求項 1 に記載の方法。

【請求項 6】

前記明瞭発音周波数が、約 $2 \sim 12.5 \text{ Hz}$ である、請求項 1 に記載の方法。

【請求項 7】

前記明瞭発音周波数が、人間の明瞭発音速度にほぼ対応する、請求項 1 に記載の方法。

【請求項 8】

前記非明瞭発音周波数が、明瞭発音周波数よりおよそ高い、請求項 1 に記載の方法。

【請求項 9】

前記明瞭発音電力と前記非明瞭発音電力との前記比較が、前記明瞭発音電力と前記非明瞭発音電力との比を求めることである、請求項 1 に記載の方法。

【請求項 10】

前記比が、前記明瞭発音電力および小さい定数を含む分子と、前記非明瞭発音電力と前記小さい定数との和を含む分母とを含む、請求項 9 に記載の方法。

【請求項 11】

前記明瞭発音電力と前記非明瞭発音電力との比較が、前記明瞭発音電力と前記非明瞭発音電力との差である、請求項 1 に記載の方法。

【請求項 12】

前記第 1 音声品質評価及び前記第 2 音声品質評価を決定する工程が、
前記明瞭発音電力と前記非明瞭発音電力の間の前記比較を行うことによりおよび DC 成分電力に基づく重み付けファクタを使用することにより 局部音声品質を決定する工程を含む、請求項 1 に記載の方法。

【請求項 13】

前記第 1 音声品質評価または前記第 2 音声品質評価が、前記局部音声品質を使用して決定される、請求項 12 に記載の方法。

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、一般に、通信システムに関し、具体的には、音声品質評価に関する。

【背景技術】**【0002】**

無線通信システムの性能は、とりわけ、音声品質の観点から測定することができる。当技術分野では、音声品質を評価する 2 つの技法が存在する。第 1 の技法は、主観的技法である（これ以後「主観的音声品質評価」と呼ぶ）。主観的音声品質評価では、聴取者を使用して、処理された音声の音声品質を評価する。この場合、処理された音声は、受信器において処理された伝送後の音声信号である。この技法は、個人の知覚に基づくので主観的であり、人間による音声品質評価は、通常、表音内容、話し方、または個々の話者の相違を考慮に入れる。主観的音声品質評価は、コストがかかり、時間がかかることがある。

【0003】

第 2 の技法は、客観的技法である（これ以後「客観的音声品質評価」と呼ぶ）。客観的音声品質評価は、個人の知覚に基づいていない。ほとんどの客観的音声品質評価の技法は、既知の元の音声、または処理された音声から推定された、再構築された元の音声に基づく。しかし、これらの客観的技法は、表音内容、話し方、または個々の話者の相違を考慮

10

20

30

40

50

に入れない。

【発明の開示】

【発明が解決しようとする課題】

【0004】

したがって、表音内容、話し方、または個々の話者の相違を考慮に入れて客観的に音声品質を評価することが求められている。

【課題を解決するための手段】

【0005】

本発明は、音声品質評価される音声信号をひずませることによって、表音内容、話し方、または個々の話者の相違を明らかにする客観的音声品質評価の方法である。音声信号のひずんだバージョンを使用することによって、音声品質を評価する際に、異なる表音内容、異なる個々の話者、および異なる話し方を補償することが可能である。音声信号をひずませることによって客観的音声品質評価が低下する量は、特に音声信号のひずんだバージョンのひずみ量が厳密であるとき、異なる音声信号について同様に維持される。発声に依存する明瞭度について補償された音声品質評価を得るために、ひずんだ音声信号についての客観的音声品質評価と元のひずんでいない音声信号についての客観的音声品質評価とを比較する。一実施形態では、比較は、ひずんだ音声信号とひずんでいない音声信号との客観的音声品質評価の相違に相当する。

本発明の特徴、態様、および利点は、以下の記述、添付の請求項、および付随する図面に関してより良く理解されるであろう。

【発明を実施するための最良の形態】

【0006】

本発明は、処理された音声信号をひずませることによって、表音内容、話し方、または個々の話者の相違を明らかにする客観的音声品質評価方法である。客観的音声品質評価は、同じ主観的音声品質スコアを有する異なる音声信号について異なる値をもたらす傾向がある。これらの値が異なる理由は、変調スペクトル領域におけるスペクトル内容の分布が異なるためである。処理された音声信号のひずんだバージョンを使用することによって、異なる表音内容、異なる個々の話者、および異なる話し方を補償することが可能である。音声信号をひずませることによって客観的音声品質評価が低下する量は、特にひずみが厳密であるとき、異なる音声信号について同様に維持される。発声に依存明瞭度について補償された音声品質評価を得るために、ひずんだ音声信号と元のひずんでいない音声信号との客観的音声信号評価を比較する。

【0007】

図1は、本発明による発声依存明瞭度を補償する客観的音声品質評価構成10を示す。客観的音声品質評価構成10は、複数の客観的音声品質評価モジュール12、14と、ひずみモジュール16と、発声固有バイアス補償モジュール18とを備える。音声信号 $s(t)$ が、入力としてひずみモジュール16および客観的音声品質評価モジュール12に提供される。ひずみモジュール16において、音声信号 $s(t)$ をひずませて、変調雑音基準単位(modulated Noise Reference Unit MNRU)音声信号 $s'(t)$ を生成する。すなわち、ひずみモジュール16は、入力信号 $s(t)$ の雑音バージョンを生成する。次いで、MNRU音声信号 $s'(t)$ は、入力として客観的音声品質評価モジュール14に提供される。

【0008】

客観的音声品質評価モジュール12、14において、音声信号 $s(t)$ およびMNRU音声信号 $s'(t)$ を処理して、客観的音声品質評価 $SQ(s(t))$ および $SQ(s'(t))$ を得る。客観的音声品質評価モジュール12、14は、あらゆる入力音声信号に対して実施される処理のタイプについて本質的に同一である。すなわち、両方の客観的音声品質評価モジュール12、14が、同じ入力音声信号を受信する場合、両方のモジュール12、14の出力信号は、ほぼ同一である。他の実施形態では、客観的音声品質評価モジュール12、14は、互いに異なる方式で音声信号 $s(t)$ および $s'(t)$ を処理す

ることが可能であることに留意されたい。客観的音声信号評価モジュールは、当技術分野では周知である。そのようなモジュールの例について、以下で記述する。

【0009】

次いで、客観的音声品質評価 $SQ(s(t))$ と $SQ(s'(t))$ とを比較して、発声依存明瞭度を補償する音声品質評価 $SQ_{compensated}$ を獲得する。一実施形態では、音声品質評価 $SQ_{compensated}$ は、客観的音声品質評価 $SQ(s(t))$ と $SQ(s'(t))$ との相違を使用して決定される。たとえば、 $SQ_{compensated}$ は、 $SQ(s(t))$ から $SQ(s'(t))$ を減算したもの、または $SQ(s'(t))$ から $SQ(s(t))$ を減算したものに等しい。他の実施形態では、音声品質評価 $SQ_{compensated}$ は、客観的音声品質評価 $SQ(s(t))$ と $SQ(s'(t))$ との比に基づいて決定される。たとえば、

10

【数1】

$$SQ_{compensated} = \frac{SQ(s(t)) + \mu}{SQ(s'(t)) + \mu} \quad \text{or} \quad SQ_{compensated} = \frac{SQ(s'(t)) + \mu}{SQ(s(t)) + \mu}$$

であり、上式で、 μ は、小さい一定値である。

20

【0010】

前述したように、客観的音声品質評価モジュール12、14は、当技術分野では周知である。図2は、本発明による音声の聴覚明瞭度分析モジュールを使用する客観的音声品質評価モジュール12、14の実施形態20を示す。図2に示すように、客観的品質評価モジュール20は、蝸牛フィルタバンク22と、包絡線分析モジュール24と、明瞭度分析モジュール26とを備える。客観的品質評価モジュール20において、音声信号 $s(t)$ が、入力として蝸牛フィルタバンク22に提供される。蝸牛フィルタバンク22は、周辺聴覚システムの第1段階に従って音声信号 $s(t)$ を処理するための複数の蝸牛フィルタ $h_i(t)$ を備える。ここで $i = 1, 2, \dots, N_c$ は、特定の蝸牛フィルタ・チャンネルを表し、 N_c は、蝸牛フィルタ・チャンネルの全数を表す。具体的には、蝸牛フィルタバンク22は、複数の臨界帯域信号 $s_i(t)$ を生成するように音声信号 $s(t)$ をろ波する。臨界帯域信号 $s_i(t)$ は、 $s(t)$ と $h_i(t)$ との積に等しい。

30

【0011】

複数の臨界帯域信号 $s_i(t)$ は、入力として包絡線分析モジュール24に提供される。包絡線分析モジュール24において、複数の臨界帯域信号 $s_i(t)$ を処理して、複数の包絡線 $a_i(t)$ を得る。ここで、

【数2】

$$a_i(t) = \sqrt{s_i^2(t) + \hat{s}_i^2(t)}$$

40

であり、

【数3】

$$\hat{s}_i(t)$$

50

は、 $s_i(t)$ のヒルベルト変換である。

【0012】

次いで、複数の包絡線 $a_i(t)$ は、入力として明瞭度分析モジュール 26 に提供される。明瞭度分析モジュール 26 において、複数の包絡線 $a_i(t)$ を処理して、音声信号 $s(t)$ の音声品質評価を得る。具体的には、明瞭度分析モジュール 26 は、人間の明瞭発音システムから生成された信号に関連付けられた電力（これ以後「明瞭発音電力 $P_A(m, i)$ 」と呼ぶ）を、人間の明瞭発音システムから生成されない信号に関連付けられた電力（これ以後「非明瞭発音電力 $P_{NA}(m, i)$ 」と呼ぶ）と比較する。次いで、そのような比較を使用して、音声品質評価を実施する。

【0013】

図 3 は、明瞭度分析モジュール 26 において、本発明の一実施形態による複数の包絡線 $a_i(t)$ を処理するためのフローチャート 300 を示す。工程 310 において、複数の包絡線 $a_i(t)$ のそれぞれのフレーム m に対してフーリエ変換を実施して、変調スペクトル $A_i(m, f)$ を生成する。 f は周波数である。

【0014】

図 4 は、電力対周波数の観点から変調スペクトル $A_i(m, f)$ を示す例 40 である。例 40 では、明瞭発音電力 $P_A(m, i)$ は、周波数 $2 \sim 12.5 \text{ Hz}$ に関連付けられた電力であり、非明瞭発音電力 $P_{NA}(m, i)$ は、 12.5 Hz より高い周波数に関連付けられた電力である。 2 Hz 未満の周波数に関連する電力 $P_{N0}(m, i)$ は、臨界帯域信号 $a_i(t)$ のフレーム m の DC 成分である。この例では、明瞭発音電力 $P_A(m, i)$ は、人間の明瞭発音速度が $2 \sim 12.5 \text{ Hz}$ であるということに基づいて、周波数 $2 \sim 12.5 \text{ Hz}$ に関連付けられた電力として選択される。明瞭発音電力 $P_A(m, i)$ に関連付けられた周波数範囲と非明瞭発音電力 $P_{NA}(m, i)$ に関連付けられた周波数範囲（これ以後それぞれ「明瞭発音周波数範囲」および「非明瞭発音周波数範囲」と呼ぶ）とは、隣接する重複しない周波数範囲である。本願の目的のため、「明瞭発音電力 $P_A(m, i)$ 」という用語は、人間が明瞭な発音の周波数範囲または上述した周波数範囲 $2 \sim 12.5 \text{ Hz}$ に限定すべきではないことを理解されたい。同様に、「非明瞭発音電力 $P_{NA}(m, i)$ 」という用語は、明瞭発音電力 $P_A(m, i)$ に関連付けられた周波数範囲より高い周波数範囲に限定すべきではない。非明瞭発音周波数範囲は、明瞭発音周波数範囲と重複するまたはしない可能性があり、もしくは明瞭発音周波数範囲と隣接するまたはしない可能性がある。非明瞭発音周波数の範囲は、臨界帯域信号 $a_i(t)$ のフレーム m の DC 成分に関連付けられた周波数など、明瞭発音周波数範囲の最低周波数より小さい周波数を含む可能性もある。

【0015】

工程 320 において、各変調スペクトル $A_i(m, f)$ について、明瞭度分析モジュール 26 は、明瞭発音電力 $P_A(m, i)$ と非明瞭発音電力 $P_{NA}(m, i)$ との比較を実施する。明瞭度分析モジュール 26 のこの実施形態では、明瞭発音電力 $P_A(m, i)$ と非明瞭発音電力 $P_{NA}(m, i)$ との比較は、明瞭発音対非明瞭発音の比 $ANR(m, i)$ である。 ANR は、以下の式によって定義される。

【数 4】

$$ANR(m, i) = \frac{P_A(m, i) + \epsilon}{P_{NA}(m, i) + \epsilon} \quad \text{式(1)}$$

上式で、 ϵ は、ある程度小さい一定値である。明瞭発音電力 $P_A(m, i)$ と非明瞭発音電力 $P_{NA}(m, i)$ との他の比較が可能である。たとえば、比較は、式(1)の逆数とすることが可能であり、または、比較は、明瞭発音電力 $P_A(m, i)$ と非明瞭発音電力 $P_{NA}(m, i)$ との差とすることが可能である。議論を簡単にするために、フローチャート 300 によって示した明瞭度分析モジュール 26 の実施形態について、式(1)の A

10

20

30

40

50

$NR(m, i)$ を使用する比較に関して議論する。しかし、これは決して、本発明を限定すると解釈すべきではない。

【 0 0 1 6 】

工程 3 3 0 において、 $ANR(m, i)$ を使用して、フレーム m について局部音声品質 $LSQ(m)$ を決定する。局部音声品質 $LSQ(m)$ は、 DC 成分電力 $P_{No}(m, i)$ に基く重み付けファクタ $R(m, i)$ と、すべてのチャネル i にわたる明瞭発音対非明瞭発音の比 $ANR(m, i)$ を使用して決定される。具体的には、局部音声品質 $LSQ(m)$ は、以下の式を使用して決定される。

【数 5】

$$LSQ(m) = \log \left[\sum_{i=1}^{N_c} ANR(m, i) R(m, i) \right] \quad \text{式(2)}$$

10

上式で

【数 6】

$$R(m, i) = \frac{\log(1 + P_{No}(m, i))}{\sum_{k=1}^{N_c} \log(1 + P_{No}(m, k))} \quad \text{(式3)}$$

20

であり、 k は、周波数インデックスである。

【 0 0 1 7 】

工程 3 4 0 において、音声信号 $s(t)$ の全音声品質 SQ が、フレーム m についての局部音声品質 $LSQ(m)$ および対数電力 $P_s(m)$ を使用して決定される。具体的には、音声品質 SQ は、以下の式を使用して決定される。

【数 7】

$$SQ = L \left\{ P_s(m) LSQ(m) \right\}_{m=1}^T = \left[\sum_{\substack{m=1 \\ P_s > P_{th}}}^T P_s^{\lambda}(m) LSQ^{\lambda}(m) \right]^{\frac{1}{\lambda}} \quad \text{(式4)}$$

30

上式で、

【数 8】

$$P_s(m) = \log \left[\sum_{t=m} s^2(t) \right]$$

40

であり、 L は L_p ノルム (norm)、 T は音声信号 $s(t)$ におけるフレームの全数、 λ は任意の値、 P_{th} は、可聴信号と沈黙とを識別する閾値である。一実施形態では、 λ は奇数値であることが好ましい。

【 0 0 1 8 】

明瞭度分析モジュール 2 6 の出力は、すべてのフレーム m にわたる音声品質 SQ の評価である。すなわち、音声品質 SQ は、音声信号 $s(t)$ に対する音声品質評価である。

本発明について、ある実施形態を参照してかなり詳細に記述してきたが、他の形態も可

50

能である。したがって、本発明の精神および範囲は、本明細書に包含される実施形態の記述に限定すべきではない。

【図面の簡単な説明】

【0019】

【図1】本発明による発声依存明瞭度を補償する客観的音声品質評価構成を示す図である。

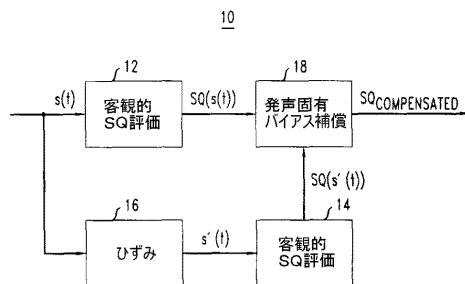
【図2】本発明による音声の聴覚明瞭度分析モジュールを使用する客観的音声品質評価の実施形態を示す図である。

【図3】明瞭度分析モジュールにおいて、本発明の一実施形態による複数の包絡線 $a_i(t)$ を処理するためのフローチャートである。

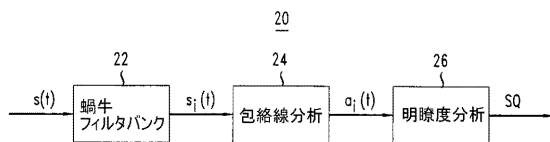
【図4】電力対周波数の観点から、変調スペクトル $A_i(m, f)$ を例示する図である。

10

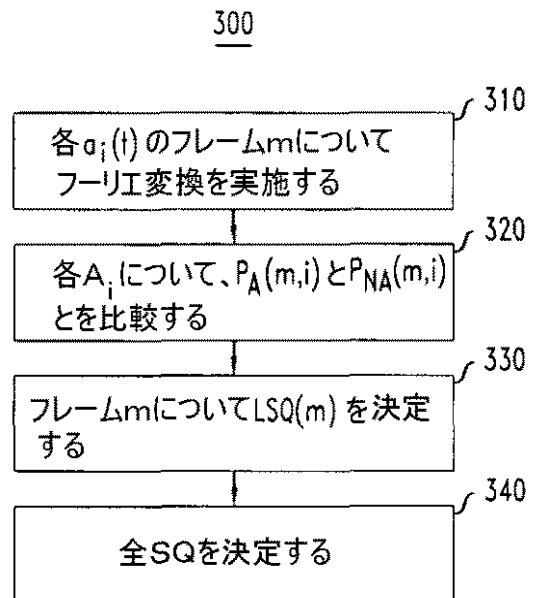
【図1】



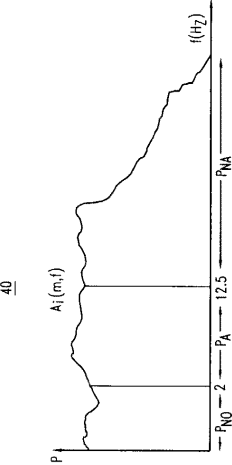
【図2】



【図3】



【 図 4 】



フロントページの続き

(74)代理人 100096943

弁理士 臼井 伸一

(74)代理人 100101498

弁理士 越智 隆夫

(74)代理人 100104352

弁理士 朝日 伸光

(72)発明者 キム, ドー - スク

アメリカ合衆国 07920 ニュージャージー, バスキング リッジ, ハンティングトン ロード 42

審査官 菊池 智紀

(56)参考文献 国際公開第02/039707(WO, A1)

国際公開第01/052600(WO, A1)

特表平08-501910(JP, A)

特表平09-505701(JP, A)

特開2000-206979(JP, A)

特開2000-349893(JP, A)

特開平08-275297(JP, A)

特開平03-034700(JP, A)

"VoIPアナライザー QUALITY STAR VoIPの音声品質評価", [online], 興和株式会社, 2002年4月, インターネット<URL:<http://www.kowa.co.jp/interstream/pr5.htm>>

入井寛 他, "電話網の総合的な通話品質を推定するプログラムを開発パソコン上の簡易な操作でMOSを求める", NTT技術ジャーナル, 1990年11月1日, Vol.2, No.11, p.74-77

(58)調査した分野(Int.Cl., DB名)

G10L 11/00-21/06

H04M 1/24

JSTPlus(JDreamII)