

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第4263620号
(P4263620)

(45) 発行日 平成21年5月13日 (2009. 5. 13)

(24) 登録日 平成21年2月20日 (2009. 2. 20)

(51) Int. Cl.

F I

G 1 0 L 11/00 (2006. 01)

G 1 0 L 11/00 4 0 4

G 1 0 L 19/00 (2006. 01)

G 1 0 L 11/00 1 0 1 F

G 1 0 L 19/00 3 2 0

請求項の数 4 (全 14 頁)

(21) 出願番号 特願2003-575064 (P2003-575064)
 (86) (22) 出願日 平成15年2月26日 (2003. 2. 26)
 (65) 公表番号 特表2005-519339 (P2005-519339A)
 (43) 公表日 平成17年6月30日 (2005. 6. 30)
 (86) 国際出願番号 PCT/EP2003/002058
 (87) 国際公開番号 W02003/076889
 (87) 国際公開日 平成15年9月18日 (2003. 9. 18)
 審査請求日 平成16年11月8日 (2004. 11. 8)
 (31) 優先権主張番号 02075973. 4
 (32) 優先日 平成14年3月8日 (2002. 3. 8)
 (33) 優先権主張国 欧州特許庁 (EP)
 (31) 優先権主張番号 02075997. 3
 (32) 優先日 平成14年3月11日 (2002. 3. 11)
 (33) 優先権主張国 欧州特許庁 (EP)

(73) 特許権者 504292093
 コニクリーク・ケイビーエヌ・ナムロー
 ゼ・フェンノートシャップ
 オランダ国エヌエルー2516 シーケイ
 ・ザ・ヘーグ、マーンブレイン 55
 (74) 代理人 100089705
 弁理士 社本 一夫
 (74) 代理人 100076691
 弁理士 増井 忠武
 (74) 代理人 100075270
 弁理士 小林 泰
 (74) 代理人 100080137
 弁理士 千葉 昭男
 (74) 代理人 100096013
 弁理士 富田 博行

最終頁に続く

(54) 【発明の名称】 システムの伝送品質を測定する方法及びシステム

(57) 【特許請求の範囲】

【請求項 1】

音声システムの伝送品質を測定する方法であって、入力信号 (X) が前記音声システムに入力されて前記音声システムから出力される出力信号 (Y) を生じさせ、前記入力信号 (X) 及び前記出力信号 (Y) が処理され、望ましくは比較される方法において、

前記音声システムの前記出力信号及び/又は前記入力信号が、パワー比に応じて、前記パワーの小さな偏差が補償され、より大きな偏差が部分補償される方法でスケールングされ、

前記入力のノイズの局所レベルに依存するスケールング係数により、前記原入力音声信号 (X) に存在するノイズ・レベルが低減された人工的な参照音声信号が作成され、

時間 - 周波数平面における前記入力信号 (X) 及び前記出力信号 (Y) のそれぞれに対する内部表現 $LX(f)n$ 及び $LY(f)n$ の差 $D(f)n$ が、

$LX(f)n < K$ の場合に、 $|LY(f)n - LX(f)n^b / K^{b-1}|$ と、

$LX(f)n > K$ の場合に、 $|LY(f)n - LX(f)n|$ と、

等しくなるように設定され、あるいはこの代わりに、

$LX(t) < K'$ の場合に、 $|LY(f)n - LX(f)n^b / K^{b-1}|$ と、

$LX(t) > K'$ の場合に、 $|LY(f)n - LX(f)n|$ と、

等しくなるように設定される方法。ただし、パラメータ b が 1 より大きい値に設定される第二の同調パラメータであり、パラメータ K およびパラメータ K' が所望の低レベル・ノイズ・パワー条件を表す低レベル・ノイズ・パワー条件値である。

【請求項 2】

音声システムの伝送品質を測定する方法であって、入力信号 (X) が前記音声システムに入力されて前記音声システムから出力される出力信号 (Y) を生じさせ、前記入力信号 (X) 及び前記出力信号 (Y) が処理される方法において、

前記音声システムの前記出力信号及び / 又は前記入力信号が、パワー比に応じて、前記パワーの小さな偏差が補償され、より大きな偏差が部分補償される方法でスケージングされ、

補償比 F が、前記入力信号 (X) 及び前記出力信号 (Y) のそれぞれに対するパワー表現 P_X 及び P_Y から計算されるものであって、 P_X / P_Y の比に等しく、

クリッピングされた比 C であって、 $F < m$ のとき第 1 のクリッピング値 m と等しく、 $F > M$ のとき第 2 のクリッピング値 M と等しく、それ以外のときに F と等しいクリッピングされた比 C が計算され、

ソフトスケール比 S が、 $m < m < 1$ 且つ $M > M > 1$ であるような第 1 のスケージング係数 m 及び第 2 のスケージング係数 M から計算され、 $C < m$ の場合に S が $C^a + C - C(m)^{a-1}$ と等しく、 $C > M$ の場合に S が $C^a + C - C(M)^{a-1}$ と等しく、それ以外の場合に S が C と等しい方法。ただし、パラメータ a は 0 より大きく且つ 1 より小さい値に設定される第一の同調パラメータである。

【請求項 3】

音声システムの伝送品質を測定するシステムであって、入力信号 (X) が前記音声システムに入力されて前記音声システムから出力される出力信号 (Y) を生じさせ、前記入力信号 (X) 及び前記出力信号 (Y) が互いに処理され、望ましくは比較されるシステムにおいて、

パワー比に応じて、前記パワーの小さな偏差が補償され、より大きな偏差が部分補償される方法で、前記音声システムの前記出力信号及び / 又は前記入力信号をスケージングするスケージング手段を備え、該スケージング手段が、さらに、

前記原入力音声信号 (X) に存在するノイズ・レベルが、前記入力のノイズの局所レベルに依存するスケージング係数により低減された、人工的な参照音声信号を作成する手段と、

時間 - 周波数平面における前記入力信号 (X) 及び前記出力信号 (Y) のそれぞれに対する内部表現 $L_X(f)n$ 及び $L_Y(f)n$ の差 $D(f)n$ を設定する手段であって、該差 $D(f)n$ が、

$L_X(f)n < K$ の場合に、 $|L_Y(f)n - L_X(f)n^b / K^{b-1}|$ と、

$L_X(f)n > K$ の場合に、 $|L_Y(f)n - L_X(f)n|$ と、

等しくなるように、あるいはこの代わりに、

$L_X(t) < K'$ の場合に、 $|L_Y(f)n - L_X(f)n^b / K^{b-1}|$ と、

$L_X(t) > K'$ の場合に、 $|L_Y(f)n - L_X(f)n|$ と、

等しくなるよう設定される手段とを備えることを特徴とするシステム。ただし、パラメータ b が 1 より大きい値に設定される第二の同調パラメータであり、パラメータ K およびパラメータ K' が所望の低レベル・ノイズ・パワー条件を表す低レベル・ノイズ・パワー条件値である。

【請求項 4】

音声システムの伝送品質を測定するシステムであって、入力信号 (X) が前記音声システムに入力されて前記音声システムから出力される出力信号 (Y) を生じさせ、前記入力信号及び前記出力信号が互いに処理され、望ましくは比較されるシステムにおいて、

パワー比に応じて、前記パワーの小さな偏差が補償され、より大きな偏差が部分補償される方法で、前記音声システムの前記出力信号及び / 又は前記入力信号をスケージングするスケージング手段を備え、該スケージング手段が、さらに、

前記原入力音声信号 (X) に存在するノイズ・レベルが、該入力のノイズの局所レベルに依存するスケージング係数により低減された、人工的な参照音声信号を作成する手段と

補償比 F であって、前記入力信号 (X) 及び前記出力信号 (Y) のそれぞれに対するパワー表現 PX 及び PY から、 PX / PY の比に等しい補償比 F を計算する手段と、

クリッピングされた比 C であって、 $F < mm$ の場合に第一のクリッピング値 mm と等しく、 $F > MM$ の場合に第二のクリッピング値 MM と等しく、それ以外の場合には F と等しいクリッピングされた比 C を計算する手段と、

$mm < m < 1$ 且つ $MM > M > 1$ であるような第一のスケーリング係数 m 及び第二のスケーリング係数 M からソフトスケール比 S を計算する手段であって、 $C < m$ の場合に S が $C^a + C - C(m)^{a-1}$ と等しく、 $C > M$ の場合に S が $C^a + C - C(M)^{a-1}$ と等しく、それ以外の場合に S が C と等しく計算する手段と、

を備えることを特徴とするシステム。ただし、パラメータ a が 0 より大きく且つ 1 より小さい値に設定される第一の同調パラメータである。

10

【発明の詳細な説明】

【技術分野】

【0001】

発明の分野

本発明は、被試験システムの伝送品質を測定する方法及びシステムであって、被試験システムに入力された入力信号と、被試験システムから生じた出力信号とを処理して、互いに比較する方法及びシステムに関する。

【0002】

発明の背景

20

2001年2月に出されたITU-TのP.862勧告草案「電話伝送品質、電話設置、ローカル・ライン・ネットワーク - 客観的及び主観的品質評価方法 - 音声品質の知覚的評価方法 (PESQ)、ナローバンド電話網及び音声コーデックのエンド・ツー・エンドの音声品質を評価する客観的方法」は、従来技術であるPESQの方法及びシステムを開示する。

【0003】

音声処理システム又は音声伝送システムにおいて劣化した音声信号の品質測定は、入力信号における非常に弱い又は無音の部分に対しては劣悪な結果を生じ得る。P.862勧告により知られる方法及びシステムは、フレームごとのパワー・レベルの差を正しく補償しないという欠点を有する。これらの差は、入力信号の利得の変化又はノイズにより引き起こされる。不正確な補償は、特に原参照入力音声信号のノイズ・レベルが低い場合に、主観的スコアと客観的スコアとの間の相関を低くする。

30

【0004】

出願人による欧州特許出願第01200945号に開示された従来技術の方法及びシステムによれば、前処理段階における第1のスケーリング段階に、調整値を加えられた出力信号のパワーの逆数の関数である第1のスケーリング係数を適用することにより、改良がなされる。第2のスケーリング段階には第2のスケーリング係数が適用されるが、該係数は、0から1までの調整値を有する指数で累乗された第1のスケーリング係数と実質的に等しい。第2のスケーリング段階は装置の様々な場所で実行され得、調整値は、試験信号を用いて、よく定義された主観的品質スコアと調整される。P.862勧告及び欧州特許出願01200945号の方法及びシステムにおいては、パワー・ドメインにおいて参照入力信号と適合するよう、劣化出力信号が局所的にスケーリングされる。

40

【0005】

(知覚的) 品質測定処理の結果は、方法及びシステムのそれぞれの少なくとも1つの段階における「ソフト・スケーリング」の適用により改良されることがわかっている。(「ハード」スケーリング閾値を用いる)「ハード・スケーリング」に代わる「ソフト・スケーリング」の導入は、人間の音声知覚メカニズムは「ハードな閾値」よりむしろ「ソフトな閾値」を用いるという観察及び理解に基づいており、本発明の分野は、人間であるユーザーにより経験されるのと同様の音声品質の評価に関する。この観察、及びそれらの人間の音声スケーリング・メカニズムがどのように働くかについてのより良い理解に基づい

50

て、本発明は、従来技術による方法又はシステムのそれぞれに追加又は挿入されるべき「ソフト・スケーリング」メカニズムを提示する。

【 0 0 0 6 】

発明の概要

本発明の一つの態様によれば、システムの出力信号及び／又は入力信号は、パワー比に応じて、パワーの小さなひずみが補償され、より大きなひずみが部分補償される方法でスケーリングされる。

【 0 0 0 7 】

本発明の更なる工夫によれば、原入力音声信号に存在するノイズ・レベルが、この入力に含まれる局所ノイズ・レベルに依存するスケーリング係数により低減された人工的な参照音声信号が作成され得る。

10

【 0 0 0 8 】

本発明に係る方法は、結果として、特に、穏やかな音声部分及び静寂が低レベルのノイズにより劣化される場合に、局所スケーリングの変動を含む音声信号について主観的に知覚されるエンド・ツー・エンドの音声品質をより正確に予測することができる。

【 0 0 0 9 】

ソフト・スケーリング・アルゴリズムにおいて、主観的に知覚される品質と客観的に測定される品質との相関を高めるために、2つの異なる種類の信号処理が利用される。

第1のサブ・アルゴリズムに制御される第1のソフトスケール処理において、出力信号における局所利得変動を補正するために $P \cdot 862$ 勧告で用いられる補償は、パワー比に応じて、パワーの小さな偏差が（好ましくは時間フレーム又は時間期間ごとに）補償され、より大きな偏差が部分補償される方法で出力（又は入力）をスケーリングすることにより改良される。

20

【 0 0 1 0 】

好ましい単純且つ効果的な実施の形態は、局所パワー、即ち（例えば 30 ms の）フレームごとのパワーを取得し、局所補償比 F を計算する。即ち、

$$F = (P_X + \Delta) / (P_Y + \Delta)$$

である。ただし、 Δ は P_Y の小さな値に対して C の値を最適化するために利用される。 F は、クリッピングされた比 C を得るためにレベル m m 及び M M でクリッピングされた振幅である。即ち、

30

$$F < m \text{ 且つ } 0 < m \leq 1 \text{ のとき } C = m$$

$$F > M \text{ 且つ } M \leq 1 \text{ のとき } C = M$$

となり、それ以外の場合には、

$$C = F$$

となる。クリッピングされた比 C は、 $m < m \leq 1$ 且つ $M > M \leq 1$ である係数 m 及び M を用いてソフトスケール比 S を計算するために利用される。即ち、

$$C < m \text{ 且つ } 0 < a < 1 \text{ のとき } S = C^a + C - C(m)^{a-1}$$

$$C > M \text{ 且つ } 0 < a < 1 \text{ のとき } S = C^a + C - C(M)^{a-1}$$

であり、それ以外の場合には、

$$S = C$$

40

である。ただし、 a は（第1の）同調パラメータとして利用され得る。このように、本発明に係る局所スケーリングは、 $m \leq F \leq M$ である限り、先行技術文献である $P \cdot 862$ 勧告及び欧州特許出願 01200945 号において与えられたスケーリングと等しい。しかし、 $F < m$ 又は $F > M$ である値については、スケーリングは徐々に 1.0 より、次いで従来技術におけるスケーリングより低くなる。ソフトスケール係数 S は、各フレームの出力パワーを局所的に補償するように従来技術の方法及びシステムにおいて F が利用されるのと同じ方法で利用される。

【 0 0 1 1 】

第2のサブ・アルゴリズムに制御される第2のソフトスケール処理において、利用される補償は、入力信号の低レベルの部分に焦点を当てる。

50

入力信号（参照信号）が低レベルのノイズを含む場合、トランスペアレントな音声伝達システムは、やはり低レベルのノイズを含む出力音声信号を生じさせる。音声伝達システムの出力は、伝達システムにより導入されたノイズのために期待以下の品質を有するとの判定される。入力音声信号を聞いて比較をすることができる場合に、ノイズは伝達システムにより引き起こされたものではないという事実に気が付くに過ぎない。しかし、たいていの主観的音声品質試験では、入力参照信号は被験者に提示されず、従って、被験者は、入力信号における低レベルのノイズの差は音声伝達システムの品質の差であると判定する。客観的試験システムにおいてそのような主観的試験との相関を高めるため、進歩した客観的音声品質評価アルゴリズムにおいて、この効果がエミュレートされる。

【 0 0 1 2 】

10

本発明のこの好ましいオプションは、入力信号の局所ノイズ・レベルに依存するスケーリング係数によりノイズのパワー・レベルが低減された、パワー表現ドメインにおける新しい架空の人工的な参照音声信号を効果的に作成することにより、これをエミュレートする。従って、新しく作成された人工的な参照信号は、この入力信号の低レベル部分について原入力信号より速くゼロに収束する。参照入力信号に存在するような低レベル信号の期間に劣化出力信号における外乱が計算される場合、ゼロに近づく際に入力信号のラウドネスより速くゼロに到達するレベルへの入力ラウドネス信号のスケーリングの後に、内部表現のラウドネス・ドメインにおける差の計算が実行される。

【 0 0 1 3 】

20

欧州特許出願第 0 1 2 0 0 9 4 5 号に開示された従来技術の方法によれば、処理は、人間の聴覚システムの心理物理的知覚モデルに従って（劣化）出力信号（ $Y(t)$ ）と参照信号（ $X(t)$ ）とを代表信号 LY 及び LX にマッピングすることを含む。差分信号又は外乱信号（ D ）はこれらの代表信号から「識別手段」により決定され、該外乱信号は、次いで、品質信号 Q を得るために、被験者の特定の特性がモデル化される認識モデルに従うモデル化手段により処理される。

【 0 0 1 4 】

上記の通り、内部表現ラウドネス・ドメインにおける差の計算は、本発明の範囲内において、ゼロに近づく際に入力信号のラウドネスより速くゼロに達するレベルへの入力ラウドネス信号のスケーリングの後に実行されることが望ましい。

【 0 0 1 5 】

30

これの効果的な実現形態は、 $LX(f)_n$ 及び $LY(f)_n$ から計算された時間 - 周波数平面における内部表現の差を利用することにより達成される（欧州特許出願第 0 1 2 0 0 9 4 5 号参照）。即ち、

$$D(f)_n = |LY(f)_n - LX(f)_n|$$

及び、これを置き換えた

$$D(f)_n = |LY(f)_n - H(t, f)|$$

である。ただし、 $LX(f)_n < K$ のとき

$$H(t, f) = LX(f)_n^b / K^{b-1}$$

であり、 $LX(f)_n \geq K$ のとき

$$H(t, f) = LX(f)_n$$

40

である。これらの公式において、 K は、特定の實現形態に応じて時間周波数セルごとの低レベルのノイズ・パワー条件を表し、 $b > 1$ である。

【 0 0 1 6 】

この第 2 のソフトスケール処理サブ・アルゴリズムは、 $LX(f)_n < K$ という条件を単一の時間フレームにおけるパワー条件で置き換えることによっても實現され得る。即ち、

$$D(f)_n = |LY(f)_n - H(t, f)|$$

であって、 $LX(t) < K'$ のとき

$$H(t, f) = LX(f)_n^b / K'^{b-1}$$

であり、 $LX(t) \geq K'$ のとき

50

$$H(t, f) = L X(f) n$$

である。これらの公式において、 K' は、特定の実現形態に依存する時間フレームごとの低レベルのノイズのパワー条件を表し、 $b > 1$ である。

【0017】

図面の詳細な説明

図1に示されるPESQシステムは、原信号（入力信号） $X(t)$ と、例えば通信システムに $X(t)$ を通過させた結果である劣化信号（出力信号） $Y(t)$ とを比較する。PESQシステムの出力は、主観的聴音試験において被験者により $Y(t)$ に与えられるであろう知覚品質の予測である。PESQシステムにより実行される第1の段階において、遅延が以前の時間区間と大きく異なる時間区間ごとに、原入力と劣化出力との間の一連の遅延が計算される。これらの各区間について、対応する始点及び終点が計算される。整列アルゴリズムは、或る時間区間内に2つの遅延があるという確度と、当該区間に単一の遅延があるという確度との比較の原理に基づく。該アルゴリズムは、無音部分及びアクティブな音声部分の両期間において、遅延の変動を処理することができる。

【0018】

検出された一組の遅延に基づいて、PESQシステムは、原（入力）信号と、被試験装置の整列された劣化出力とを、知覚モデルを用いて比較する。このプロセスの要点は、原信号と劣化信号とを、知覚周波数（Bark）及びラウドネス（Sone）を考慮に入れながら、人間の聴覚システムにおける音声信号の精神物理学的表現と類似する内部表現（ LX 、 LY ）に変換することである。これは、時間合わせ、校正された聴覚レベルへのレベル合わせ、時間 - 周波数マッピング、周波数ワーピング、及び圧縮型ラウドネス・スケールリングなどの幾つかの段階により実現される。

【0019】

内部表現は、過度に激しくない場合には、知覚的にほとんど重要性を有さない局所利得変動及び線形フィルタリングなどの影響を考慮に入れて処理される。これは、補償の量を制限し、補償に影響よりも遅れさせることにより達成される。こうして、原信号と劣化信号との間の小さな安定状態の差が補償される。より重大な影響又は急激な変動は部分的にしか補償されないため、未解決の影響が残り、全体としての知覚の妨げとなる。このため、全ての主観的な影響をモデル化するために、少数の品質指標のみしか利用することができない。PESQシステムでは、認識モデルにおいて2つの誤差パラメータが計算される。これらは組み合わされて、客観的聴音品質MOS（平均オピニオン・スコア）を提供する。PESQシステムで用いられる基本的な概念は、参考文献一覧の(1)から(5)に記載されている。

【0020】

従来技術のPESQシステムにおける知覚モデル

図1に示されるPESQシステムの知覚モデルは、原音声信号と劣化音声信号との間の距離（「PESQスコア」）を計算するために利用される。これは、所与の主観的試験における主観的MOSの予測を得るために単調関数に通され得る。PESQスコアは、-0.5 ~ 4.5の範囲内の単一の数値であるMOSと同様のスケールにマッピングされるが、大抵の場合、出力範囲はACR聴音品質実験におけるMOS値の正常範囲である1.0 ~ 4.5である。

【0021】

定数設定の事前計算

或る種の定数値及び関数は事前計算される。これらはサンプル周波数に依存するため、8kHzと16kHzのサンプル周波数のバージョンがプログラムに格納されている。

【0022】

FFTウィンドウ・サイズ及びサンプル周波数

PESQシステムにおいて、時刻信号は、32msサイズのハニング窓による短期間FFT（高速フーリエ変換）を用いて時間周波数ドメインにマッピングされる。8kHzの場合、これは1ウィンドウ当たり256個のサンプルに相当し、16kHzの場合、ウィン

10

20

30

40

50

ドウは隣接フレームが 50 % だけ重なり合うが、512 個のサンプルを計数する。

【0023】

絶対聴覚閾値

絶対聴覚閾値 $P_0(f)$ は、利用される Bark 帯域の中央における値を得るために補間される。これらの値は配列に記憶され、ツビッカーのラウドネス公式において利用される。

【0024】

パワー・スケーリング係数

時間 - 周波数分析のための FFT に付随する任意の利得定数が存在する。この定数は、32 ms 以上のウィンドウを用いた FFT により周波数ドメインに変換された、29.5 4 (40 dB SPL) の振幅を有する周波数 1000 Hz の正弦波から計算される。次いで、(離散)周波数軸が、FFT 帯域のビンニングにより修正 Bark スケールに変換される。Bark 周波数スケールにビンニングされたスペクトルのピーク振幅(「ピッチ・パワー密度」と呼ばれる)は、10000 (40 dB SPL) になるはずである。後者は、パワー・スケーリング係数 S_p である定数を右から乗ずることにより実行される。

【0025】

ラウドネス・スケーリング係数

同じ 40 dB の参照音が、心理音響的な (Sone) ラウドネス・スケールを校正するために利用される。修正 Bark スケールへのビンニングの後、強度軸は、絶対聴覚閾値に基づき、ツビッカーの法則を用いてラウドネス・スケールにワープされる。1000 Hz 及び 40 dB SPL の校正音を用いて、Bark 周波数スケールに対してラウドネス密度を積分すると、1 Sone の値が得られる。後者は、ラウドネス・スケーリング係数 S_1 である定数を右から乗ずることにより実行される。

【0026】

IRS 受信フィルタリング

セクション 10.1.2 で述べられているように、聴音試験は、ハンドセットにおける IRS 受信特性又は修正された IRS 受信特性を用いて実行されると想定される。音声信号に対する必要なフィルタリングは、既に前処理において適用されている。

【0027】

アクティブな音声時間区間の計算

原音声ファイル及び劣化音声ファイルが大きな無音区間で開始又は終了している場合、これは、ファイルにおける特定の平均ひずみ値の計算に影響し得る。そのため、これらのファイルの最初又は最後に、無音部分が見積もられる。或る位置がアクティブな区間の始点又は終点であるとみなされる為には、連続した 5 つの絶対サンプル値の和が、原音声ファイルの最初及び最後から 500 を超えなければならない。この始点と終点との間の区間が、アクティブな音声時間区間と定義される。計算サイクル及び / 又は記憶サイズを節約するため、幾つかの計算はアクティブな区間に限られる。

【0028】

短期間 FFT

人間の耳は、時間 - 周波数変換を実行する。PESQ システムにおいて、これは、ウィンドウ・サイズが 32 ms の短期間 FFT により実現される。連続する時間ウィンドウ(フレーム)どうしの重なりは 50 % である。パワースペクトル、即ち複素数 FFT 要素の実部の二乗と虚部の二乗との和は、原信号及び劣化信号について別々の実数値配列に格納される。単一のハニング窓内の位相情報は PESQ システムにおいて廃棄され、全ての計算はパワー表現 $PX_{WIRSS}(f)_n$ 及び $PY_{WIRSS}(f)_n$ のみに基づく。劣化信号におけるウィンドウの始点は、遅延のために移動される。原音声信号の時間軸はそのままに留められる。遅延が増加すると、劣化信号の一部が処理から省略され、遅延が減少すると、部分が繰り返される、

ピッチ・パワー密度の計算

Bark スケールは、人間の聴覚システムが低周波数において高周波数におけるよりも

10

20

30

40

50

優れた周波数分解能を有することを反映する。これは、F F T 帯域をビンニングして、F F T 帯域の対応するパワーを加算し、加算された部分を正規化することにより実現される。H e r t z の周波数スケールをB a r k のピッチ・スケールにマッピングするワーピング関数は、与えられた通りの値に正確に従うのではない。結果としての信号は、ピッチ・パワー密度 $PPX_{WIRSS}(f)_n$ 及び $PPY_{WIRSS}(f)_n$ として知られる。

【 0 0 2 9 】

原ピッチ・パワー密度の部分補償

被試験システムのフィルタリングを扱うために、原ピッチ・パワー密度及び劣化ピッチ・パワー密度のパワー・スペクトルは、時間平均される。この平均は、パワーが絶対聴音閾値の 1 0 0 0 倍より大きい時間 - 周波数セルを用いて、アクティブな音声フレームについてのみ計算される。修正されたB a r k ビンごとに、劣化スペクトルの原スペクトルに対する比率から部分補償係数が計算される。補償は最大でも 2 0 d B を超えることは無い。個々のフレーム n の原ピッチ・パワー密度 $PPX_{WIRSS}(f)_n$ は、原信号を劣化信号と等しくするよう、この部分補償係数を乗じられる。この結果、逆フィルタ処理された原ピッチ・パワー密度 $PPX'_{WIRSS}(f)_n$ が得られる。厳格なフィルタ処理は聴取者の妨げとなるため、この部分補償が利用される。A C R 実験において劣化信号は被験者により判断されるものであるため、補償は原信号に実行される。

【 0 0 3 0 】

ひずみのあるピッチ・パワー密度の部分補償

短期間利得変動は、ピッチ・パワー密度をフレームごとに処理することにより部分補償される。原ピッチ・パワー密度及び劣化ピッチ・パワー密度について、個々のフレーム n における絶対聴音閾値を超える全ての値の合計が計算される。原ファイル及び劣化ファイルにおけるパワーの比率が計算され、範囲 $[3 \cdot 10^{-4}, 5]$ に結び付けられる。(時間軸に沿った) 第一次のローパス・フィルタは、この比率に適用される。各フレーム n におけるひずみのあるピッチ・パワー密度は、この比率を乗じられて、部分的に利得補償されたひずみのあるピッチ・パワー密度 $PPY'_{WIRSS}(f)_n$ を生ずる。

【 0 0 3 1 】

ラウドネス密度の計算

フィルタ処理及び短期間利得変動に対する部分補償の後、原ピッチ・パワー密度及び劣化ピッチ・パワー密度が、ツビッカーの法則 [7] を用いてS o n e ラウドネス・スケールに変換される。即ち、

【 0 0 3 2 】

【数 1】

$$LX(f)_n = S_1 \cdot \left(\frac{P_0(f)}{0.5} \right)^r \cdot \left[\left(0.5 + 0.5 \cdot \frac{PPX'_{WIRSS}(f)_n}{P_0(f)} \right)^r - 1 \right]$$

【 0 0 3 3 】

である。ただし、 $P_0(f)$ は絶対閾値であり、 S_1 はラウドネス・スケーリング係数である。4 B a r k より大きい場合、ツビッカー・パワー y は、文字通り与えられる値である 0 . 2 3 である。4 B a r k より小さい場合、ツビッカー・パワーはいわゆる漸増効果を補償するようわずかに増加する。結果として生じる 2 次元配列 $LX(f)_n$ 及び $LY(f)_n$ はラウドネス密度と言われる。

【 0 0 3 4 】

外乱密度の計算

ひずみのあるラウドネス密度と原ラウドネス密度との符号付きの差が計算される。この差が正である場合、ノイズなどの要素が追加されている。この差が負である場合、原信号から要素が省略されている。この差の配列は、未処理外乱密度と呼ばれる。

【 0 0 3 5 】

原ラウドネス密度と劣化ラウドネス密度との最小値が、時間周波数セルごとに計算され

10

20

30

40

50

る。これらの最小値は、0.25 を乗じられる。対応する二次元配列は、マスク配列と呼ばれる。以下の法則が、それぞれの時間 - 周波数セルに適用される。

- ・未処理外乱密度が正でありマスク値より大きい場合、マスク値が未処理外乱から減じられる。

- ・未処理外乱密度がマスク値の大きさの正の値と負の値との間にある場合、外乱密度はゼロに設定される。

- ・未処理外乱密度がマスク値の負の値より更に小さい場合、マスク値が未処理外乱密度に加算される。

【0036】

最終的な効果は、未処理外乱密度がゼロに近づくことである。これは、実際の時間周波数セルが歪みとして知覚される前の不感帯を表している。これは、それぞれの時間 - 周波数セルにおける、大きな信号（マスク）が存在する場合には聞こえないような小さな差の処理をモデル化する。結果は、時間（ウィンドウ番号 n ）及び周波数の関数としての外乱密度 $D(f)_n$ である。

【0037】

非対称係数とのセルに関する乗算

コーデックが入力信号をひずませる際に、入力信号と統合する新しい時間 - 周波数要素を導入することは一般に非常に難しいため、結果としての出力信号は入力信号とひずみとの2つの異なる知覚対象に分解され、明らかに聞き取れるひずみ [2] をもたらしという事実により、非対称効果が引き起こされる。コーデックが時間 - 周波数要素を無視する場合、結果としての出力信号は同様には分解され得ず、ひずみの不快さは小さくなる。この効果は、外乱密度 $D(f)_n$ に非対称係数を乗じてフレームごとの非対称外乱密度 $DA(f)_n$ を計算することによりモデル化される。この非対称係数は、ひずみピッチ・パワー密度と1.2乗された原ピッチ・パワー密度との比率に等しい。非対称係数が3より小さい場合、ゼロに設定される。非対称係数が1.2を超える場合、その値でクリッピングされる。従って、劣化ピッチ・パワー密度が原ピッチ・パワー密度を超える時間周波数セルのみが非ゼロの値として留まり得る。

【0038】

外乱密度の集合

外乱密度 $D(f)_n$ と非対称外乱密度 $DA(f)_n$ とは、2つの異なる Lp 標準及びソフト・フレーム上の（ラウドネスの低い）重み付けを用いて、周波数軸に沿って統合（加算）される。即ち、

【0039】

【数2】

$$D_n = M_n \sqrt[3]{\sum_{f=1, \dots, \text{バース域の数}} (|D(f)_n| W_f)^3}$$

$$DA_n = M_n \sum_{f=1, \dots, \text{バース域の数}} (|DA(f)_n| W_f)$$

【0040】

である。ただし、 M_n は $1 / (\text{原フレームのパワーと定数との和})^{0.04}$ である乗算係数であり、原音声フラグメントの静寂時における外乱の強調をもたらす。 W_f は修正された Bark ピンの幅に比例する一連の定数である。この乗算の後、フレーム外乱値は最大 4.5 に制限される。これらの集合された値 D_n 及び DA_n は、フレーム外乱と呼ばれる。

【0041】

フレーム外乱のゼロ化

ひずみ信号が 16 ms（ウィンドウの半分）より大きな遅延の減少を含む場合、10.2.4 に述べられた繰返し戦略が修正される。客観音声品質の計算において、そのような

10

20

30

40

50

イベント期間のフレーム外乱は無視したほうがよいことがわかっている。その結果、これが起こる場合、フレーム外乱がゼロにされる。結果としてのフレーム外乱を、 D'_{n1} 及び DA'_{n1} と呼ぶ。

【0042】

不良区間の再整列

閾値を超えるフレーム外乱を有する連続したフレームは、不良区間と呼ばれる。少数のケースにおいて、客観的測定は、前処理により観察される不正確な時間遅延に起因する最小数の不良フレームにおける大きな歪みを予測する。いわゆる不良区間に対して、絶対原信号と、前処理により観察された遅延に応じて調整された絶対劣化信号との相関を最大化することにより、新しい遅延値が予測される。最大化された相関が閾値以下である場合、該区間ではノイズとノイズとを釣り合わせており、該区間はもはや不良と呼ばれず、該区間に対する処理が中止されることが結論付けられる。そうではない場合、不良区間のフレームに対するフレーム外乱が再計算され、より小さい場合には、原フレーム外乱と置き換えられる。結果として、知覚品質を計算するために利用される最終的なフレーム外乱 D'_{n1} 及び DA'_{n1} が得られる。

10

【0043】

瞬時区間における外乱の集計

次に、フレーム外乱値及び非対称フレーム外乱値が、 L_6 標準と、音声ファイル長の集合におけるのと同様のより高いp値とを用いて、20フレームからなる瞬時区間（約320msであるフレームの重なりを考慮する）を通して集計される。これらの区間もまた、50%ずつ重なりあい、窓関数は利用されない。

20

【0044】

信号の持続時間における外乱の集計

瞬時外乱値及び非対称瞬時外乱値は、 L_2 標準を用いて、音声ファイルのアクティブな区間（対応するフレーム）を通して集計される。音声ファイルの第1のセンテンスがひずんだ場合でも他のセンテンスの品質はそのままであるのに対し、瞬時区間の一部がひずむとその瞬時が意味を失うという事実のために、音声ファイルの集合における低いp値よりも瞬時区間内の集合に対するp値は高くなる。

【0045】

PESQスコアの計算

最終的なPESQスコアは、平均外乱値と平均非対称外乱値との線形結合である。PESQスコアの範囲は-0.5~4.5であるが、たいていの場合、出力範囲は聴音品質を示すMOSと同様の1.0~4.5のスコアであり、これはACR（絶対範囲評価）実験におけるMOS値の正常範囲である。

30

【0046】

図2は、局所スケール係数を計算するための従来技術によるモジュールを置き換える第1の新しいモジュールと、知覚的減算のための従来技術によるモジュールを置き換える新しい第2のモジュールとを除き、図1と等しい。

【0047】

第1の新しいモジュールは、本発明に係る方法の実行に適しており、パワー比に応じてパワーの小さな偏差を補償し、より大きな偏差を部分的に補償する新しい「ソフト・スケールリング」アルゴリズムの制御の下に、被試験システムの出力信号及び/又は入力信号をスケールリングする手段を備える。第1のモジュールは図3に示される。

40

【0048】

第2の新しいモジュールは、本発明の更なる精巧さの実行に適しており、原入力音声信号に存在するノイズ・レベルが、この入力におけるノイズの局所レベルに依存するスケールリング係数により低減される人工的な参照音声信号を作成する手段を備える。

【0049】

両方の新しいモジュールの動作は、各モジュールの動作を表すフロー図の形で示される。両モジュールは、ハードウェア又はソフトウェアにより実現され得る。

50

図3は、図2に示された第1の新しいモジュールの動作を示す。図3におけるモジュールの動作は、パワー比に応じて、パワーの小さな偏差が望ましくは時間フレーム又は時間期間ごとに補償され、より大きな偏差が部分補償される方法で入力に対して出力をスケールリングすることにより、出力信号における局所利得変動を補正するための補償機能を改良する、フロー図に示された第1のサブ・アルゴリズムにより制御される。本発明の好ましい単純且つ効果的な実現形態は、局所パワー、即ち（例えば、30msの）フレームごとのパワーをとり、局所補償比Fを計算する。即ち、 $F = (P_X + \Delta) / (P_Y + \Delta)$ であるが、 P_X 及び P_Y はそれぞれ、図1、2及び3で使われている $P P X_{WIRSS}(f)_n$ 及び $P P Y_{WIRSS}(f)_n$ の短縮形であり、Fはクリッピングされた比Cを得るためにレベル m 及び MM でクリッピングされた振幅である。即ち、 $F < m$ ≤ 1.0 のとき $C = m$ であり、 $F > MM$ ≥ 1.0 のとき $C = MM$ であり、又は $C = F$ である。ただし、 Δ は P_X 及び/又は P_Y の小さな値についてCを最適化する。

【0050】

クリッピングされた比Cは、係数 m 及び M を用いてソフトスケール比Sを計算するために利用される。ただし、 $m < m \leq 1.0$ 且つ $MM > M \geq 1.0$ である。ソフトスケール比Sは、 $C < m$ ($0.5 < a < 1.0$) のとき $S = C^a + C - C(m)^{a-1}$ であり、 $C > M$ のとき $S = C^a + C - C(M)^{a-1}$ であり、又は $S = C$ である。このように、本発明における局所スケールリングは、 $m \leq F \leq M$ である限り、従来技術文献である勧告P.862及び欧州特許出願第01200945号で与えられたスケールリングと等しい。しかし、 $F < m$ 又は $F > M$ である値に対しては、スケールリングは徐々に、従来技術におけるスケールリングより1.0から小さく外れる。ソフト・スケールリング係数Sは、各フレームにおいて局所的に出力パワーを補償するために従来技術による方法及びシステムにおいてFが利用されたのと同じ方法で利用される。

【0051】

第2のサブ・アルゴリズムに制御される第2のソフトスケール処理において、入力信号の低レベルの部分に対し、進歩したスケールリングが適用される。入力信号（参照信号）が低レベルのノイズを含む場合、トランスペアレントな音声伝達システムは、やはり低レベルのノイズを含む出力音声信号を与える。すると、音声伝達システムの出力は、伝達システムにより導入されたノイズのために期待以下の品質を有すると判定される。入力音声信号を聴いて比較することができるシステムである場合に、ノイズが伝達により生じたものではないと言う事実気付くだけである。しかし、たいていの主観的音声品質試験では、入力参照信号は被験者に提示されず、従って、被験者は入力信号における低ノイズ・レベルの違いを、音声伝達システムの品質の違いであると判断する。客観的試験システムでは、そのような主観的試験との高い相関を得るため、進歩した客観的音声品質評価アルゴリズムにおいて、この効果がエミュレートされる。図4に示される本発明の好ましいオプションとしての実施の形態は、入力信号のノイズの局所レベルに依存するスケールリング係数によりノイズ・パワー・レベルが低減されるパワー表現ドメインにおいて人工的な参照音声信号を作成することにより、これをエミュレートする。従って、人工的な参照信号は、この入力信号の低レベルについて原入力信号より速くゼロに収束する。参照入力信号に存在する低レベル信号期間において、劣化出力信号の外乱が計算される場合、内部表現のラウドネス・ドメインにおける減算は、ゼロに近づく際に入力信号のラウドネスより速くゼロに達するレベルへの入力ラウドネス信号のスケールリングの後に実行される。

【0052】

時間 - 周波数平面における内部表現の差は、
 $LX(f)_n < K$ のとき $D(f)_n = |LY(f)_n - LX(f)_n|^b / K^{b-1}$ であり、
 $LX(f)_n \geq K$ のとき $D(f)_n = |LY(f)_n - LX(f)_n|$
 に設定される。これらの公式において、 $b > 1$ であり、Kは時間周波数セルごとの低レベルのノイズ・パワー条件を表す。

【0053】

代わりに、第2のソフトスケール処理サブ・アルゴリズムは、 $LX(f)_n < K$ の条件

を単一の時間フレームにおけるパワー条件で置き換えることにより実現され得る。この代わりのオプションにおいて、時間 - 周波数平面における内部表現の差は、

$LX(t) < K'$ のとき $D(f)n = |LY(f)n - LX(f)n|^b / K^{b-1}$ であり、
 $LX(t) \geq K'$ のとき $D(f)n = |LY(f)n - LX(f)n|$

に設定される。これらの代わりの公式において、 $b > 1$ であり、 K' は時間フレームごとの低レベルのノイズ・パワー基準を表す。

【 0 0 5 4 】

参照により本明細書に援用される参考文献

- [1] BEERENDS (J.G.), STEMERDINK (J.A.): A Perceptual Speech-Quality Measure Based on a Psychoacoustic Sound Representation, J. Audio Eng. Soc., Vol. 42, No. 3, pp. 115-123, March 1994. 10
- [2] BEERENDS (J.G.): Modeling Cognitive Effects that Play a Role in the Perception of Speech Quality, Speech Quality Assessment, Workshop papers, Bochum, pp. 1-9, November 1994.
- [3] BEERENDS (J.G.): Measuring the quality of speech and music codecs, an integrated psychoacoustic approach, 98th AES Convention, pre-print No. 3945, 1995.
- [4] HOLLIER (M.P.), HAWKSFORD (M.O.), GUARD (D.R.): Error activity and error entropy as a measure of psychoacoustic significance in the perceptual domain, IEE Proceedings-Vision, Image and Signal Processing, 141 (3), 203-208, June 1994.
- [5] RIX (A.W.), REYNOLDS (R.), HOLLIER (M.P.): Perceptual measurement of end-to-end speech quality over audio and packet-based networks, 106th AES Convention, pre-print No. 4873, May 1999. 20
- [6] HOLLIER (M.P.), HAWKSFORD (M.O.), GUARD (D.R.), Characterization of communications systems using a speech-like test stimulus, Journal of the AES, 41 (12), 1008-1021, December 1993.
- [7] ZWICKER (Feldtkeller): Das Ohr als Nachrichtenempfänger, S. Hirzel Verlag, Stuttgart, 1967.
- [8] ITU - T P. 862 勧告草案、"Telephone transmission quality, telephone installations, local line networks-Methods for objective and subjective assessment of quality-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 02.2001 30
- [9] 欧州特許出願第01200945号, 出願人Koninklijke KPN n. v.

【図面の簡単な説明】

【 0 0 5 5 】

【図 1】図 1 は、ITU - T の P. 862 勧告に開示された従来技術による PESQ システムを概略的に示す。

【図 2】図 2 は、同一の PESQ システムであるが、第一、及び望ましくは第 2 の新しいモジュールを用いて上記の方法の実行に適するよう修正された PESQ システムを示す。

【図 3】図 3 は、PESQ システムの第 1 の新しいモジュールを示す。 40

【図 4】図 4 は、PESQ システムの第 2 の新しいモジュールを示す。

【図 1】

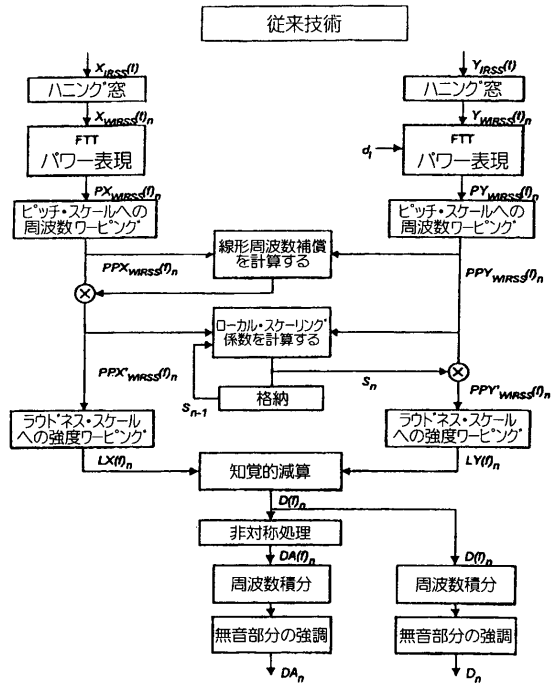


FIG. 1

【図 2】

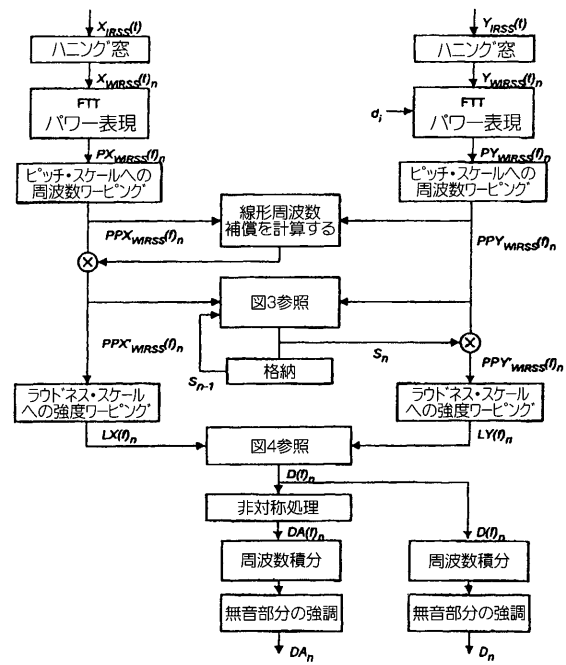


FIG. 2

【図 3】

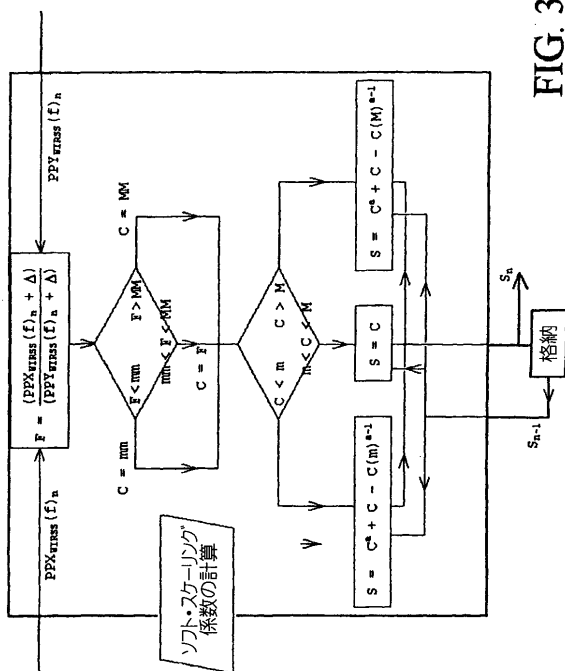


FIG. 3

【図 4】

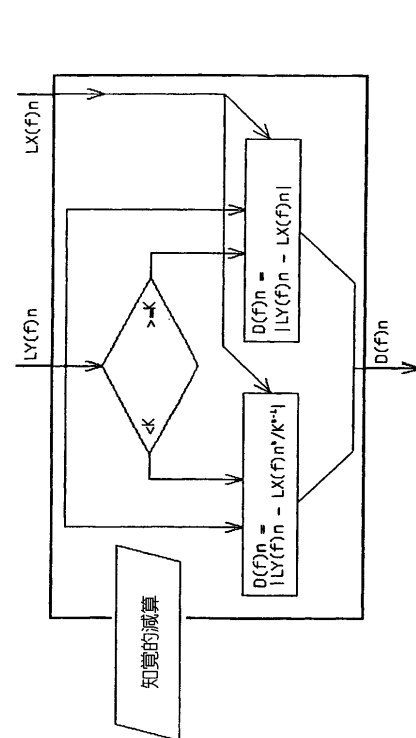


FIG. 4

フロントページの続き

(74)代理人 100091063

弁理士 田中 英夫

(72)発明者 ベーレンズ, ジョン・ジェラード

オランダ国エヌエル - 4 5 8 5 ピージー・ヘングストディーク, ポルダーウェグ 2 6

審査官 山下 剛史

(56)参考文献 特表平 1 1 - 5 0 2 0 7 1 (J P , A)

特開平 0 6 - 2 3 6 1 9 8 (J P , A)

J.G.BEERENDS, J.A.STEMERDINK, "A perceptual audio quality measure based on a psychoacoustic sound representation", Journal of the Audio Engineering Society, Vol.40, No12 (1992-12), p.963-978

(58)調査した分野(Int.Cl., D B 名)

G10L 11/00-21/06

H04R 3/00- 3/04, 29/00