

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第4440937号
(P4440937)

(45) 発行日 平成22年3月24日(2010.3.24)

(24) 登録日 平成22年1月15日(2010.1.15)

(51) Int.Cl.

F I

G 1 O L 21/02 (2006.01)

G 1 O L 21/02 1 O 1 B

G 1 O L 21/02 1 O 3 Z

請求項の数 83 (全 30 頁)

(21) 出願番号	特願2006-545874 (P2006-545874)	(73) 特許権者	398012616
(86) (22) 出願日	平成16年12月29日(2004.12.29)		ノキア コーポレイション
(65) 公表番号	特表2007-517249 (P2007-517249A)		フィンランド エフイーエンーO2150
(43) 公表日	平成19年6月28日(2007.6.28)		エスプー ケイララーデンティエ 4
(86) 国際出願番号	PCT/CA2004/002203	(74) 代理人	100099759
(87) 国際公開番号	W02005/064595		弁理士 青木 篤
(87) 国際公開日	平成17年7月14日(2005.7.14)	(74) 代理人	100092624
審査請求日	平成18年8月28日(2006.8.28)		弁理士 鶴田 準一
(31) 優先権主張番号	2454296	(74) 代理人	100102819
(32) 優先日	平成15年12月29日(2003.12.29)		弁理士 島田 哲郎
(33) 優先権主張国	カナダ(CA)	(74) 代理人	100113826
			弁理士 倉地 保幸
		(74) 代理人	100108383
			弁理士 下道 晶久

最終頁に続く

(54) 【発明の名称】 暗騒音存在時の音声を改善するための方法および装置

(57) 【特許請求の範囲】

【請求項 1】

音声信号のノイズ抑制方法であって、

複数の周波数ピンを具備する音声信号のスペクトル領域表示を生成するために周波数分析を行うステップと、

前記周波数ピンを複数の周波数帯域にグループ化するステップとを有する方法において、前記音声信号の中に有声化音声活性状態を検出したときに、周波数帯域の第1の数に対しては周波数ピン毎のベースでノイズの抑制を行い、前記周波数帯域の第2の数に対しては周波数帯域毎のベースでノイズの抑制を行うことを特徴とする音声信号のノイズ抑制方法。

【請求項 2】

周波数帯域の前記第1の数が、有声化された周波数帯域の前記第1の数に従って決定される請求項1に記載の方法。

【請求項 3】

それ未満の周波数の信号が有声音であると見なされる有声化カットオフ周波数に関して、周波数帯域の前記第1の数を決定する請求項1に記載の方法。

【請求項 4】

周波数帯域の前記第1の数が、前記音声信号のすべての周波数帯域を含み、前記すべての周波数帯域が前記有声化カットオフ周波数を上回らない上位周波数を有する請求項3に記載の方法。

【請求項 5】

周波数帯域の前記第 1 の数が所定の固定数である請求項 1 に記載の方法。

【請求項 6】

前記音声信号の周波数帯域が有声化されていない場合、すべての周波数帯域に対して周波数帯域毎のベースでノイズの抑制を行う請求項 1 に記載の方法。

【請求項 7】

前記音声信号が、複数のサンプルを含む音声フレームを含み、請求項 1 に記載の方法を適用して、音声フレーム内のノイズの抑制を行う請求項 1 に記載の方法。

【請求項 8】

前記音声フレームの第 1 のサンプルに関して m 個のサンプルによってオフセットされる分析ウィンドウを利用して、請求項 1 の前記周波数分析を行うステップを有する請求項 7 に記載の方法。

10

【請求項 9】

前記音声フレームの第 1 のサンプルに関して m 個のサンプルによってオフセットされる第 1 の分析ウィンドウを利用して、第 1 の周波数分析を行うステップと、前記音声フレームの第 2 のサンプルに関して p 個のサンプルによってオフセットされる第 2 の分析ウィンドウを利用して、第 2 の周波数分析を行うステップとを有する請求項 7 に記載の方法。

【請求項 10】

$m = 24$ で、 $p = 128$ である請求項 9 に記載の方法。

【請求項 11】

20

前記第 2 の分析ウィンドウが、前記音声フレームから、後続する音声フレームの中へ延在する先読み部分を含む請求項 9 に記載の方法。

【請求項 12】

前記周波数ピンおよび周波数帯域の少なくとも一方にスケーリング利得値を印加してノイズの抑制を行うステップを有する請求項 1 に記載の方法。

【請求項 13】

周波数ピン毎のベースでノイズの抑制を行うときに、周波数ピンに対して周波数ピン固有のスケーリング利得値を決定するステップをさらに有する請求項 1 に記載の方法。

【請求項 14】

周波数帯域毎のベースでノイズの抑制を行うときに、周波数帯域に対して周波数帯域特有のスケーリング利得値を決定するステップを有する請求項 1 に記載の方法。

30

【請求項 15】

一定のスケーリング利得値を印加することによってノイズの抑制を行うステップを有する請求項 6 に記載の方法。

【請求項 16】

周波数ピンに対して決定された前記周波数ピンの信号対雑音比 (SNR) に関して、前記周波数ピンに対して周波数ピン特有のスケーリング利得値を決定するステップを有する請求項 13 に記載の方法。

【請求項 17】

周波数帯域に対して決定された前記周波数帯域の信号対雑音比 (SNR) に関して、前記周波数帯域に対して周波数帯域特有のスケーリング利得値を決定するステップを有する請求項 14 に記載の方法。

40

【請求項 18】

前記第 1 および第 2 の周波数分析の各々に対して請求項 16 に記載のステップを有する請求項 16 に記載の方法。

【請求項 19】

前記第 1 と第 2 の周波数分析の各々に対して請求項 17 に記載のステップを有する請求項 17 に記載の方法。

【請求項 20】

前記スケーリング利得値が、平滑化済みのスケーリング利得値である請求項 12 ~ 14

50

のいずれか一項に記載の方法。

【請求項 2 1】

或る特定の周波数ビン、または或る特定の周波数帯域のスケーリング利得値に対して逆比例する値を有する平滑化係数を用いて、前記特定の周波数ビン、または前記特定の周波数帯域に対して印加すべき平滑化済みのスケーリング利得値を計算するステップを有する請求項 1 2 ~ 1 4 のいずれか一項に記載の方法。

【請求項 2 2】

平滑化ステップが、小さな値のスケーリング利得に対してより強くなるように決定された値を有する平滑化係数を用いて、前記特定の周波数ビン、または前記特定の周波数帯域に対して印加すべき平滑化済みのスケーリング利得値を計算するステップを有する請求項 1 2 ~ 1 4 のいずれか一項に記載の方法。

10

【請求項 2 3】

前記スケーリング利得の値を決定するステップが、音声フレーム毎に n 回 (n は 1 よりも大きい) 発生する請求項 1 3 または 1 4 に記載の方法。

【請求項 2 4】

$n = 2$ である請求項 2 3 に記載の方法。

【請求項 2 5】

音声フレーム毎に n 回 (n は 1 よりも大きい) 前記スケーリング利得の値を決定するステップを有する方法であって、前記有声化カットオフ周波数が少なくとも部分的に前回の音声フレーム内の前記音声信号の関数である請求項 1 3 または 1 4 に記載の方法。

20

【請求項 2 6】

周波数ビン毎のベースのノイズ抑制が、1 7 の帯域に対応する 7 4 ビンの最大値で行われる請求項 1 3 に記載の方法。

【請求項 2 7】

周波数ビン毎のベースのノイズ抑制が、3 7 0 0 H z の周波数に従って最大数の周波数ビンで行われる請求項 1 3 に記載の方法。

【請求項 2 8】

第 1 の S N R 値に対しては前記スケーリング利得の値が最小値に設定され、前記第 1 の S N R 値よりも大きな第 2 の S N R 値に対しては前記スケーリング利得の値が 1 に設定される請求項 1 6 に記載の方法。

30

【請求項 2 9】

前記第 1 の S N R 値が約 1 d B に等しく、前記第 2 の S N R 値が約 4 5 d B である請求項 2 8 に記載の方法。

【請求項 3 0】

活性音声を含まない音声信号の検出セクションをさらに具備する請求項 2 0 に記載の方法。

【請求項 3 1】

活性音声を含まない音声信号のセクションの検出に応答して、前記平滑化済みのスケーリング利得を最小値にリセットするステップをさらに有する請求項 3 0 に記載の方法。

【請求項 3 2】

40

複数の周波数帯域内の最大ノイズエネルギーがしきい値未満であるときに、前記ノイズ抑制手法が活性音声フレームでは実行されない請求項 7 に記載の方法。

【請求項 3 3】

短いハングオーバー音声フレームの発生に応答して、第 1 の x 周波数帯域用の帯域毎のベースで決定された平滑化済みのスケーリング利得を印加して前記音声信号の周波数スペクトルのスケーリングを行うステップと、前記スケーリング利得の単一値を印加して前記音声信号の周波数スペクトルの残りの周波数帯域のスケーリングを行うステップとをさらに有する請求項 7 に記載の方法。

【請求項 3 4】

前記第 1 の x 周波数帯域が、1 7 0 0 H z までの周波数に対応する請求項 3 3 に記載の

50

方法。

【請求項 3 5】

狭帯域音声信号に対して、個々の周波数帯域が少なくとも 2 つの周波数ピンを具備する第 1 の x 周波数帯域であって、前記第 1 の x 周波数帯域が 3 7 0 0 H z までの周波数に対応する前記周波数帯域毎のベースで決定された平滑化済みのスケーリング利得値を印加することによってノイズの抑制を行うステップと、3 7 0 0 H z に対応する前記周波数ピンで、前記スケーリング利得の値を印可することによってノイズの抑制を行うステップと、前記音声信号の周波数スペクトルの残りの周波数帯域をゼロにするステップとをさらに有する請求項 2 0 に記載の方法。

【請求項 3 6】

前記狭帯域音声信号が、1 2 8 0 0 H z までアップサンプリングされた狭帯域音声信号である請求項 3 5 に記載の方法。

【請求項 3 7】

計算された有声音測定値を用いて前記有声化カットオフ周波数を決定する請求項 3 に記載の方法。

【請求項 3 8】

前記有声化カットオフ周波数を上回らない上位周波数を有する臨界帯域の数をさらに決定するステップを有しており、 x 周波数帯域の最小値と y 周波数帯域の最大値とに対して周波数ピン毎の処理を実行するように境界周波数を設定する請求項 3 7 に記載の方法。

【請求項 3 9】

$x = 3$ および $y = 1 7$ である請求項 3 8 に記載の方法。

【請求項 4 0】

3 2 5 H z 以上で、かつ、3 7 0 0 H z 以下となるように前記有声化カットオフ周波数の境界周波数を設ける請求項 3 7 に記載の方法。

【請求項 4 1】

音声信号内のノイズを抑制する装置であって、

複数の周波数ピンを具備する音声信号のスペクトル領域表示を生成するために周波数分析を行うように構成され、

前記周波数ピンを複数の周波数帯域にグループ化するように構成される装置において、前記音声信号の中に有声化音声活性状態を検出したときに、周波数帯域の第 1 の数に対しては周波数ピン毎のベースでノイズの抑制を行い、前記周波数帯域の第 2 の数に対しては周波数帯域毎のベースでノイズの抑制を行うことを特徴とする音声信号内のノイズを抑制する装置。

【請求項 4 2】

周波数帯域の前記第 1 の数が、有声化された周波数帯域の第 1 の数に従って決定される請求項 4 1 に記載の装置。

【請求項 4 3】

それ未満の周波数の信号は有声音であると見なされる有声化カットオフ周波数に関して、周波数帯域の前記第 1 の数を決定するように構成される請求項 4 1 に記載の装置。

【請求項 4 4】

周波数帯域の前記第 1 の数が、前記音声信号のすべての周波数帯域を含み、前記すべての周波数帯域が前記有声化カットオフ周波数を上回らない上位周波数を有する請求項 4 3 に記載の装置。

【請求項 4 5】

周波数帯域の前記第 1 の数が所定の固定数である請求項 4 1 に記載の装置。

【請求項 4 6】

前記音声信号の周波数帯域が有声化されていない場合、すべての周波数帯域に対して周波数帯域毎のベースでノイズの抑制を行う請求項 4 1 に記載の装置。

【請求項 4 7】

前記音声信号が複数のサンプルを含む音声フレームを含み、音声フレーム内のノイズの

10

20

30

40

50

抑制を行うように構成される請求項 4 1 に記載の装置。

【請求項 4 8】

前記音声フレームの第 1 のサンプルに関して m 個のサンプルによってオフセットされる分析ウィンドウを利用して、前記周波数分析を行うように構成される請求項 4 7 に記載の装置。

【請求項 4 9】

前記音声フレームの第 1 のサンプルに関して m 個のサンプルによってオフセットされる第 1 の分析ウィンドウを利用して、第 1 の周波数分析を行い、前記音声フレームの第 2 のサンプルに関して p 個のサンプルによってオフセットされる第 2 の分析ウィンドウを利用して、第 2 の周波数分析を行うように構成される請求項 4 7 に記載の装置。

10

【請求項 5 0】

$m = 24$ で、 $p = 128$ である請求項 4 9 に記載の装置。

【請求項 5 1】

前記第 2 の分析ウィンドウが、前記音声フレームから、後続する音声フレームの中へ延在する先読み部分を含む請求項 4 9 に記載の装置。

【請求項 5 2】

前記周波数ビンおよび周波数帯域の少なくとも一方にスケーリング利得値を印加してノイズの抑制を行うように構成される請求項 4 1 に記載の装置。

【請求項 5 3】

周波数ビン毎のベースでノイズの抑制を行うときに、周波数ビンに対して周波数ビン特有のスケーリング利得値を決定するようにさらに構成される請求項 4 1 に記載の装置。

20

【請求項 5 4】

周波数帯域毎のベースでノイズの抑制を行うときに、周波数帯域に対して周波数帯域固有のスケーリング利得値を決定するようにさらに構成される請求項 4 1 に記載の装置。

【請求項 5 5】

すべての周波数帯域に対して一定のスケーリング利得値を印加することによってノイズの抑制を行うように構成される請求項 4 6 に記載の装置。

【請求項 5 6】

前記周波数ビンに対して決定された周波数ビンの信号対雑音比 (SNR) に関して、前記周波数ビンに対して周波数ビン固有のスケーリング利得値を決定するように構成される請求項 5 3 に記載の装置。

30

【請求項 5 7】

前記周波数帯域に対して決定された前記周波数帯域の信号対雑音比 (SNR) に関して、前記周波数帯域に対して周波数帯域固有のスケーリング利得値を決定するように構成される請求項 5 4 に記載の装置。

【請求項 5 8】

前記第 1 と第 2 の周波数分析の各々に対して請求項 5 6 に記載のステップを実行するように構成される請求項 5 6 に記載の装置。

【請求項 5 9】

前記第 1 と第 2 の周波数分析の各々に対して請求項 5 7 に記載のステップを実行するように構成される請求項 5 7 に記載の装置。

40

【請求項 6 0】

前記スケーリング利得値が平滑化済みのスケーリング利得値である請求項 5 2 ~ 5 4 のいずれか一項に記載の装置。

【請求項 6 1】

或る特定の周波数ビン、または或る特定の周波数帯域のスケーリング利得値に対して逆比例する値を有する平滑化係数を用いて、前記特定の周波数ビン、または前記特定の周波数帯域に対して印加すべき平滑化済みのスケーリング利得値を計算するように構成される請求項 5 2 ~ 5 4 のいずれか一項に記載の装置。

【請求項 6 2】

50

平滑化ステップが、小さな値のスケーリング利得値に対してより強くなるように決定された値を有する平滑化係数を用いて、前記特定の周波数ピン、または前記特定の周波数帯域に対して印加すべき平滑化済みのスケーリング利得値を計算するように構成される請求項 5 2 ~ 5 4 のいずれか一項に記載の装置。

【請求項 6 3】

音声フレーム毎に n 回 (n は 1 よりも大) 前記スケーリング利得値を決定するように構成される請求項 5 3 または 5 4 に記載の装置。

【請求項 6 4】

$n = 2$ である請求項 6 3 に記載の装置。

【請求項 6 5】

音声フレーム毎に n 回 (n は 1 よりも大きい) 前記スケーリング利得値を決定するように構成される装置であって、前記有声化カットオフ周波数が少なくとも部分的に前回の音声フレーム内の前記音声信号の関数である請求項 5 3 または 5 4 に記載の装置。

【請求項 6 6】

1 7 の帯域に対応する 7 4 ピンの最大値で周波数ピン毎のベースのノイズ抑制を行うように構成される請求項 5 3 に記載の装置。

【請求項 6 7】

3 7 0 0 H z の周波数に対応する最大数の周波数ピンで周波数ピン毎のベースのノイズ抑制を行うように構成される請求項 5 3 に記載の装置。

【請求項 6 8】

第 1 の S N R 値に対しては前記スケーリング利得値を最小値に設定し、前記第 1 の S N R 値よりも大きな第 2 の S N R 値に対しては前記スケーリング利得値を 1 に設定する請求項 5 6 に記載の装置。

【請求項 6 9】

前記第 1 の S N R 値が約 1 d B に等しく、前記第 2 の S N R 値が約 4 5 d B である請求項 6 8 に記載の装置。

【請求項 7 0】

活性音声を含まない音声信号のセクションを検出するように構成される請求項 6 0 に記載の装置。

【請求項 7 1】

活性音声を含まない音声信号のセクションの検出に応答して、前記平滑化済みのスケーリング利得値を最小値にリセットするように構成される請求項 7 0 に記載の装置。

【請求項 7 2】

複数の周波数帯域内の最大ノイズエネルギーがしきい値未満であるときに、活性音声フレームでノイズ抑制を行わないように構成される請求項 4 7 に記載の装置。

【請求項 7 3】

短いハングオーバー音声フレームの発生に応答して、第 1 の x 周波数帯域に対して周波数帯域毎のベースで決定されたスケーリング利得値を平滑化することによってノイズ抑制を行い、残りの周波数帯域に対してスケーリング利得の単一値を印加することによってノイズ抑制を行うように構成される請求項 4 7 に記載の装置。

【請求項 7 4】

前記第 1 の x 周波数帯域が、1 7 0 0 H z までの周波数に対応する請求項 7 3 に記載の装置。

【請求項 7 5】

狭帯域音声信号に対して、3 7 0 0 H z までの周波数に対応する第 1 の x 周波数帯域に対して、前記周波数帯域毎のベースで決定された平滑化済みのスケーリング利得値を印加することによってノイズの抑制を行い、3 7 0 0 H z に対応する前記周波数ピンのスケーリング利得値を 3 7 0 0 H z と 4 0 0 0 H z との間の前記周波数ピンに印加することによってノイズの抑制を行い、前記第 1 の x 周波数帯域が前記音声信号の周波数スペクトルの残りの周波数帯域をゼロにするように構成される請求項 6 0 に記載の装置。

10

20

30

40

50

【請求項 7 6】

前記狭帯域音声信号が、12800 Hz までアップサンプリングされた狭帯域音声信号である請求項 7 5 に記載の装置。

【請求項 7 7】

計算された有声音測定値を用いて前記有声化カットオフ周波数を決定する請求項 4 3 に記載の装置。

【請求項 7 8】

前記有声化カットオフ周波数を上回らない上位周波数を有する臨界帯域の数をさらに決定するステップを有しており、 x 周波数帯域の最小値と y 周波数帯域の最大値とで周波数ビン毎の処理を実行するように構成される境界周波数を設定する請求項 7 7 に記載の装置。

10

【請求項 7 9】

$x = 3$ および $y = 17$ である請求項 7 8 に記載の装置。

【請求項 8 0】

325 Hz 以上で、かつ、3700 Hz 以下となるように前記有声化カットオフ周波数の境界周波数を設ける請求項 7 7 に記載の装置。

【請求項 8 1】

ノイズ抑制装置を具備する音声エンコーダであって、前記ノイズ抑制装置が、複数の周波数ピンを具備する音声信号のスペクトル領域表示を生成するために周波数分析を行うように構成され、

20

前記周波数ピンを複数の周波数帯域にグループ化するように構成される装置において、前記音声信号の中に有声化音声活性状態を検出したときに、周波数帯域の第 1 の数に対しては周波数ビン毎のベースでノイズの抑制を行い、前記周波数帯域の第 2 の数に対しては周波数帯域毎のベースでノイズの抑制を行うことを特徴とする音声エンコーダ。

【請求項 8 2】

ノイズ抑制装置を具備する自動音声認識システムであって、前記ノイズ抑制装置が、複数の周波数ピンを具備する音声信号のスペクトル領域表示を生成するために周波数分析を行うように構成され、

前記周波数ピンを複数の周波数帯域にグループ化するように構成される装置において、前記音声信号の中に有声化音声活性状態を検出したときに、周波数帯域の第 1 の数に対しては周波数ビン毎のベースでノイズの抑制を行い、前記周波数帯域の第 2 の数に対しては周波数帯域毎のベースでノイズの抑制を行うことを特徴とする自動音声認識システム。

30

【請求項 8 3】

ノイズ抑制装置を具備する移動電話であって、前記ノイズ抑制装置が、複数の周波数ピンを具備する音声信号のスペクトル領域表示を生成するために周波数分析を行うように構成され、

前記周波数ピンを複数の周波数帯域にグループ化するように構成される装置において、前記音声信号の中に有声化音声活性状態を検出したときに、周波数帯域の第 1 の数に対しては周波数ビン毎のベースでノイズの抑制を行い、前記周波数帯域の第 2 の数に対しては周波数帯域毎のベースでノイズの抑制を行うことを特徴とする移動電話。

40

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、音声信号を強めて、暗騒音存在時の通信を改善する手法に関する。特に、本発明は、音声信号内の暗騒音のレベルを低減させるためのノイズ低減回路の設計に関するものであるが、これ以外の事項を排除するものではない。

【背景技術】**【0002】**

多くの通信システムにおいて暗騒音レベルを低減させることは非常に重要である。例えば、移動電話は、高い暗騒音レベルが存在する多くの環境で利用される。そのような利用

50

環境として、(ますます手を使わなくなっている)自動車内や街路での利用状況があり、これによって、車両による高レベルの騒音や街路の騒音が存在する中で通信システムの操作を行う必要が生じることになる。テレビ会議およびハンドフリー型のインターネットアプリケーションのようなオフィスでの利用時には、システムは、オフィスの騒音に効率良く対処する必要がある。実際には別のタイプの環境騒音も経験する場合がある。低い信号対雑音比(SNR)での操作を必要とする場合が多い上記のアプリケーションにとって、ノイズ抑制としてやはり知られているノイズの低減すなわち音声強化が重要となる。ノイズ低減は、種々の現実の環境でますます採用されている自動音声認識システムにおいても重要である。ノイズの低減は、上述のアプリケーションで通常用いられる音声符号化アルゴリズムや音声認識アルゴリズムのパフォーマンスを改善することになる。

10

【0003】

ノイズ低減のために最も利用される手法の1つとしてスペクトル減算手法がある(エス・エフ・ボル(S.F.Boll)の「スペクトル減算を利用する音声内の音響ノイズの抑制」(IEEE会報、音響、音声、信号処理、巻ASSP-27、第113頁~第120頁、1979年4月)を参照のこと)。スペクトル減算は、ノイズの多い音声からノイズ推定値の減算を行うことによって音声の短時間スペクトル振幅値の推定を試みるものである。位相歪みが人間の聴覚によって知覚されないという仮説に基づいてノイズの多い音声の位相は処理されない。実際には、スペクトル減算は、ノイズスペクトルとノイズの多い音声スペクトルとの推定値からSNRベースの利得関数を形成することによって実行される。この利得関数に入力スペクトルを乗じて、低いSNRを用いて周波数成分が抑制される。従来方式のスペクトル減算アルゴリズムを利用する上での主要な問題点として、結果として生じる「楽音(musical tones)」からなる音楽的残留ノイズがあり、この残留ノイズはリスナにとってのみならず(音声符号化のような)後続信号処理用アルゴリズムにとっても妨害となる。これらの楽音は主として、スペクトル推定時の分散に起因して生じるものである。この問題に対処するために、スペクトル平滑化ステップが提案され、この結果として低減された分散と分解能とが得られた。楽音を低減させるための別の公知の方法として、スペクトルフロアと組み合わせた過減算係数を利用する方法がある(エム・ベルーチ(M.Berouti)、アール・シュワルツ(R.Schwartz)およびジェイ・マックホール(J.Makhoul)の「音響ノイズによって損なわれた音声の改善」(IEEE ICASSP会報、ワシントンDC、1979年4月、第208頁~第211頁)を参照のこと)。この方法には、楽音が十分に低減されたとき、音声が悪化するという欠点がある。別のアプローチとして、ソフト・デシジョン・ノイズ抑制フィルタリング(アール・ジェイ・マックオーリー(R.J.McAulay)およびエム・エル・マルパス(M.L.Malpasse)の「ソフトノイズ抑制フィルタを利用する音声の改善」(IEEE会報、音響、音声、信号処理、巻ASSP-28、第137頁~第145頁、1980年4月)を参照のこと)、および、非線形スペクトル減算(ピー・ロックウッド(P.Lockwood)およびジェイ・パウディ(J.Boudy)の「非線形スペクトル減算回路(NSS)と、隠れマルコフモデルと、投影とを用いた車内でのロバスト(robust)認識のための実験」(音声通信、巻11、第215頁~第228頁、1992年6月)を参照のこと)を利用する方法もある。

20

30

40

【発明の開示】

【課題を解決するための手段】

【0004】

上記問題点を解決するために、本発明の1つの態様では、音声信号のノイズ抑制方法であって、

複数の周波数ピンを具備する音声信号のスペクトル領域表示を生成するために周波数分析を行うステップと、

上記周波数ピンを複数の周波数帯域にグループ化するステップとを有する方法において、上記音声信号の中に有声化音声活性状態を検出したときに、周波数帯域の第1の数に対しては周波数ピン毎のベースでノイズの抑制を行い、上記周波数帯域第2の数に対しては

50

周波数帯域毎のベースでノイズの抑制を行うことを特徴とする方法が提供される。

【 0 0 0 5 】

本発明の別の態様では、音声信号内のノイズを抑制する装置であって、

複数の周波数ピンを具備する音声信号のスペクトル領域表示を生成するために周波数分析を行うように構成され、

上記周波数ピンを複数の周波数帯域にグループ化するように構成される装置において、上記音声信号の中に有声化音声活性状態を検出したときに、周波数帯域の第 1 の数に対しては周波数ピン毎のベースでノイズの抑制を行い、上記周波数帯域第 2 の数に対しては周波数帯域毎のベースでノイズの抑制を行うことを特徴とする装置が提供される。

【 0 0 0 6 】

本発明のさらなる態様では、ノイズ抑制装置を具備する音声エンコーダであって、上記ノイズ抑制装置が、

複数の周波数ピンを具備する音声信号のスペクトル領域表示を生成するために周波数分析を行うように構成され、

上記周波数ピンを複数の周波数帯域にグループ化するように構成される装置において、前記音声信号の中に有声化音声活性状態を検出したときに、周波数帯域の第 1 の数に対しては周波数ピン毎のベースでノイズの抑制を行い、上記周波数帯域の第 2 の数に対しては周波数帯域毎のベースでノイズの抑制を行うことを特徴とする音声エンコーダが提供される。

【 0 0 0 7 】

本発明のさらに別の態様では、ノイズ抑制装置を具備する自動音声認識システムであって、上記ノイズ抑制装置が、

複数の周波数ピンを具備する音声信号のスペクトル領域表示を生成するために周波数分析を行うように構成され、

上記周波数ピンを複数の周波数帯域にグループ化するように構成される装置において、上記音声信号の中に有声化音声活性状態を検出したときに、周波数帯域の第 1 の数に対しては周波数ピン毎のベースでノイズの抑制を行い、上記周波数帯域の第 2 の数に対しては周波数帯域毎のベースでノイズの抑制を行うことを特徴とする自動音声認識システムが提供される。

【 0 0 0 8 】

本発明のさらに別の態様では、ノイズ抑制装置を具備する移動電話であって、上記ノイズ抑制装置が、

複数の周波数ピンを具備する音声信号のスペクトル領域表示を生成するために周波数分析を行うように構成され、

上記周波数ピンを複数の周波数帯域にグループ化するように構成される装置において、上記音声信号の中に有声化音声活性状態を検出したときに、周波数帯域の第 1 の数に対しては周波数ピン毎のベースでノイズの抑制を行い、上記周波数帯域の第 2 の数に対しては周波数帯域毎のベースでノイズの抑制を行うことを特徴とする移動電話が提供される。

【 0 0 1 1 】

添付図面と共に添付図面で単に例示として示される本発明の実施例についての以下の本発明を限定するものではない説明を読むときに、本発明の上述の目的およびその他の目的と、利点ならびに特徴はさらに明らかになるであろう。

【 発明を実施するための最良の形態 】

【 0 0 1 2 】

本明細書では、ノイズ低減のための効率の良い手法が開示される。これらの手法は、少なくとも部分的な臨界帯域における振幅スペクトルの分割と、臨界帯域毎の S N R に基づく利得関数の計算とに基づく手法であり、E V R C 音声コーデックで用いられるアプローチと類似した手法である (3 G P P 2 C . S 0 0 1 4 - 0 の「広帯域スペクトラム拡散通信システム用の拡張機能を備えた可変レートコーデック (E V R C) サービスオプション」 (3 G P P 2 技術仕様、1999 年 1 2 月) を参照のこと) 。例えば、処理対象の音声

10

20

30

40

50

フレームの性質に基づいて各種の処理手法を利用する特徴が開示される。無声音フレームでは、帯域毎の処理がスペクトル全体の中で利用される。或る一定の周波数まで有声音が検出されたフレームでは、有声音が検出されたスペクトルの低い部分で周波数ビン毎の処理が利用され、残りの帯域では帯域毎の処理が利用される。暗騒音フレームの場合には、スペクトル全体で同じスケール値を利用することによって一定のノイズフロアが除去される。さらに、個々の帯域内でのスケール値の平滑化ステップの手法が開示され、実際のスケール値に逆比例する平滑化係数を利用して周波数ビンが実行される（平滑化は小さな利得ほど強くなる）。例えば、有声音での立ち上りの場合のように、このアプローチによって低S N Rフレームに先行する高いS N R音の部分内の歪みが防止される。

10

【 0 0 1 3 】

本発明を限定するものではない本発明の1つの態様として、スペクトル減算に基づくノイズ低減用の新規の方法を提供する態様がある。この態様では、ノイズ低減方法は処理対象の音声フレームの性質に依存することになる。例えば、有声音フレームでは、処理は或る一定の周波数未満の周波数ビン毎のベースで実行することも可能である。

【 0 0 1 4 】

ある実施例では、ノイズの低減は音声符号化システムの範囲内で実行され、符号化前に音声信号の暗騒音レベルが低減される。8000サンプル/秒でサンプリングされた狭帯域音声信号もしくは16000サンプル/秒でサンプリングされた広帯域音声信号、または他の任意のサンプリング周波数に関して上記の開示された手法を利用することができる。この実施例で用いられるエンコーダは、AMR-WBコーデック（エス・エフ・ボル（S・F・B o l l））の「スペクトル減算を利用する音声内の音響ノイズの抑制」（IEEE会報、音響、音声、信号処理、巻ASSP-27、第113頁～第120頁、1979年4月）を参照のこ）をベースとするものであり、内部サンプリング変換を利用して、12800サンプル/秒（6.4kHzの帯域で動作）への信号サンプリング周波数変換が行われる。

20

【 0 0 1 5 】

したがって、本実施例の上記の開示されたノイズ低減手法は、12.8kHzへのサンプリング変換後、狭帯域信号かまたは広帯域信号かのいずれかの信号で動作することになる。

30

【 0 0 1 6 】

広帯域入力の場合、入力信号は16kHzから12.8kHzまでデシメーション（decimated）を行う必要がある。デシメーションは4だけの第1のアップサンプリングを行い、次いで、6.4kHzで周波数をカットオフするローパスFIRフィルタを介して出力のフィルタリングを行うことによって実行される。次いで、信号は5だけダウンサンプリングされる。フィルタリング遅延時間は16kHzのサンプリング周波数で15サンプルとなる。

【 0 0 1 7 】

狭帯域入力の場合、8kHzから12.8kHzまで信号のアップサンプリングを行う必要がある。このアップサンプリングは8だけの第1のアップサンプリングによって、次いで、6.4kHzのカットオフ周波数を有するローパスFIRフィルタによって出力のフィルタリングを行うことにより実行される。次いで、信号は5だけダウンサンプリングされる。フィルタリング遅延時間は8kHzのサンプリング周波数で8サンプルとなる。

40

【 0 0 1 8 】

サンプリング変換を行った後、符号化処理プロセス（ハイパスフィルタ処理とプレエンファシス処理）を行う前に2つの前処理関数が信号に適用される。

【 0 0 1 9 】

ハイパスフィルタは、望ましくない低周波成分に対する予防策として機能するものである。本実施例では50Hzのカットオフ周波数のフィルタが使用され、下記の式によって示されている。

50

【数 1】

$$H_{hl}(z) = \frac{0.982910156 - 1.965820313z^{-1} + 0.982910156z^{-2}}{1 - 1.965820313z^{-1} + 0.966308593z^{-2}}$$

【0020】

10

プレエンファシス回路では、1次ハイパスフィルタを利用してより高い周波数が強調され、この1次ハイパスフィルタは $H_{pre-emph}(z) = 1 - 0.68z^{-1}$ によって示される。

【0021】

プレエンファシス回路は、高周波でコーデックのパフォーマンスを改善し、エンコーダで使用されるエラー最小化処理時の聴覚重み付けを改善するために AMR-WB コーデックの中で用いられる。

【0022】

本実施例の残り部分では、ノイズ低減アルゴリズムの入力時の信号は 12.8 kHz サンプル周波数に変換され、前述のような前処理が行われる。しかし、上記の開示された手法は、前処理を用いて、および、前処理を用いることなく 8 kHz や 16 kHz のような別のサンプル周波数の信号に等しく適用することができる。

20

【0023】

以下、ノイズ低減アルゴリズムについて詳細に説明する。ノイズ低減アルゴリズムを利用する音声エンコーダは、12.8 kHz サンプル周波数で 256 サンプルを含む 20 ms フレームの処理を行う。さらに、コーデックはその解析時に将来のフレームから 13 ms の先読みを利用する。ノイズ低減は同じフレームの同期構造に従う。しかし、エンコーダフレーミングとノイズ低減フレーミングとの間に何らかのシフトを導入して、先読みの利用の最大化を図るようにすることが可能である。本説明では、サンプルのインデックスはノイズ低減フレーミングを反映するものである。

30

【0024】

図1は、ノイズ低減を含む音声通信システムを概略的に示すブロック図である。ブロック101では、前述の実施例のように前処理が実行される。

【0025】

ブロック102では、スペクトル分析と音声活性化検出 (VAD: voice activity detection) とが実行される。50%のオーバーラップ部分を有する 20 ms のウィンドウを利用して、2つのスペクトル分析が個々のフレームで実行される。ブロック103では、ノイズ低減をスペクトルパラメータに適用し、次いで、離散的逆フーリエ変換を利用して強化された信号を変換して、元の時間領域へこの信号を戻す。次いで、オーバーラップ追加処理を利用して信号の再構成が行われる。

40

【0026】

ブロック104では、線形予測 (LP) 分析とオープンループピッチ分析とが (通常音声符号化アルゴリズムの一部として) 実行される。本実施例では、ブロック104の結果として生じるパラメータを決定時に使用して、臨界帯域でノイズ推定値の更新を行う (ブロック105)。VAD決定はノイズ更新決定として使用することができる。ブロック105で更新されたノイズエネルギー推定値はノイズ低減ブロック103内の次のフレームで利用され、スケール値が計算される。ブロック106は、強化された音声信号に対して音声符号化を実行する。別のアプリケーションでは、ブロック106を自動音声認識システムとすることができる。ブロック104内の諸機能は、音声符号化アルゴリズムの一体部分にしてもよいことに留意されたい。

50

【 0 0 2 7 】

スペクトル分析

離散フーリエ変換を利用してスペクトル分析並びにスペクトルエネルギーの推定が行われる。50%のオーバーラップを用いて256ポイント高速フーリエ変換（FFT）を利用してフレーム毎に2回周波数分析が行われる（図2に例示されているように）。分析ウィンドウが配置され、それによってすべての先読みが利用されることになる。音声エンコーダの現フレームの開始点の24サンプル後に第1のウィンドウの開始点が配置される。第2のウィンドウはさらに128サンプル後に配置される。周波数分析用の入力信号の重み付けを行うために、ハニング（Hanning）ウィンドウ（サインウィンドウに相当する）の平方根が利用されてきた。このウィンドウは、オーバーラップ追加方法用として特に好適である（したがって、この特別のスペクトル分析は、スペクトル減算およびオーバーラップ追加分析／合成に基づいてノイズ抑制アルゴリズムで利用される）。平方根ハニング・ウィンドウは下記の式によって与えられる。

【 数 2 】

$$w_{FFT}(n) = \sqrt{0.5 - 0.5 \cos\left(\frac{2\pi n}{L_{FFT}}\right)} = \sin\left(\frac{\pi n}{L_{FFT}}\right) \quad (n = 0, \dots, L_{FFT} - 1) \quad (1)$$

但し、 $L_{FFT} = 256$ はFFT分析のサイズである。ウィンドウの1/2のみが計算され、格納されることに留意されたい。というのは、このウィンドウが左右対称（0から $L_{FFT}/2$ まで）であるからである。

【 0 0 2 8 】

(n) は、ノイズ低減フレーム内の第1のサンプルに対応するインデックス0付きの信号を示すものとする（本実施例では、第1のサンプルは、音声エンコーダフレームの開始点から24サンプル多いサンプルである）。双方のスペクトル分析用のウィンドウ化された信号が下記の式の形で得られる。

【 数 3 】

$$\begin{aligned} x_w^{(1)}(n) &= w_{FFT}(n) s'(n) & (n = 0, \dots, L_{FFT} - 1) \\ x_w^{(2)}(n) &= w_{FFT}(n) s'(n + L_{FFT}/2) & (n = 0, \dots, L_{FFT} - 1) \end{aligned}$$

但し、 $s'(0)$ は、現在のノイズ低減フレーム内の第1のサンプルである。

【 0 0 2 9 】

FFTは双方のウィンドウ化された信号に対して実行され、下記のようなフレーム当たり2組のスペクトルパラメータが得られる。

【数 4】

$$X^{(1)}(k) = \sum_{n=0}^{N-1} x_w^{(1)}(n) e^{-j2\pi \frac{kn}{N}} \quad (k=0, \dots, L_{FFT}-1)$$

$$X^{(2)}(k) = \sum_{n=0}^{N-1} x_w^{(2)}(n) e^{-j2\pi \frac{kn}{N}} \quad (k=0, \dots, L_{FFT}-1)$$

10

【0030】

FFTの出力は、 $X_R(k)$ ($k=0 \dots 128$) および $X_I(k)$ ($k=0 \dots 127$) によって示されるスペクトルの実数部および虚数部を与える。 $X_R(0)$ は、0 Hz (DC) におけるスペクトルに対応し、 $X_R(128)$ は、6400 Hz におけるスペクトルに対応することに留意されたい。これらのポイントにおけるスペクトルは実数値化され、その後の分析時には通常単に無視される。

【0031】

FFT分析を行った後、結果として生じるスペクトルは以下の上限値を有するインタバル (周波数帯域範囲 0 ~ 6400 Hz 内の 20 帯域) を利用して臨界帯域に分割される。
 臨界帯域 = { 100.0, 20, 0.0, 300.0, 400.0, 510.0, 630.0, 770.0, 920.0, 1080.0, 1270.0, 1480.0, 1720.0, 2000.0, 2320.0, 2700.0, 3150.0, 3700.0, 4400.0, 5300.0, 6350.0 } Hz

20

(ディー・ジョンストン (D. Johnston) の「知覚ノイズ基準を利用するオーディオ信号の変換符号化」(IEEE J セレクション、通信分野、巻 6、第 314 頁 ~ 第 323 頁、1988 年 2 月) を参照のこと)

【0032】

256 ポイント FFT は 50 Hz の周波数分解能 (6400 / 128) をもたらすことになる。したがって、スペクトルの直流成分を無視した後、臨界帯域毎の周波数ビンの数は、それぞれ、 $M_{CB} = \{ 2, 2, 2, 2, 2, 2, 3, 3, 3, 4, 4, 5, 6, 6, 8, 9, 11, 14, 18, 21 \}$ となる。

30

【0033】

臨界帯域内の平均エネルギーは、下記の式に示すように、

【数 5】

$$E_{CB}(i) = \frac{1}{(L_{FFT}/2)^2 M_{CB}(i)} \sum_{k=0}^{M_{CB}(i)-1} (X_R^2(k+j_i) + X_I^2(k+j_i)) \quad (i=0, \dots, 19) \quad (2)$$

40

として計算される。但し、 $X_R(k)$ および $X_I(k)$ は、それぞれ k 番目の周波数ビンの実数部および虚数部であり、 j_i は、 $j_i = \{ 1, 3, 5, 7, 9, 11, 13, 16, 19, 22, 26, 30, 35, 41, 47, 55, 64, 75, 89, 107 \}$ によって与えられる i 番目の臨界帯域における第 1 の周波数ビンのインデックスである。

【0034】

第 1 の 17 個の臨界帯域 (直流成分を除く 74 ビン) 用スペクトル分析モジュールも周波数ビン当たりのエネルギー $E_{BIN}(k)$ を計算するモジュールである。このエネルギー E_{BIN}

50

(k) は下記の式により得られる。

【数 6】

$$E_{BIN}(k) = X_R^2(k) + X_I^2(k) \quad (k = 0, \dots, 73) \quad (3)$$

【0035】

10

最後に、双方のFFT分析用スペクトル分析モジュールは、平均臨界帯域 E_{CB} によって 20ms フレーム内の平均総エネルギーを計算する。すなわち、或る一定のスペクトル分析用のスペクトルエネルギーは下記の式として計算される。

【数 7】

$$E_{frame} = \sum_{i=0}^{19} E_{CB}(i) \quad (4)$$

20

さらに、総フレームエネルギーはフレーム内の双方のスペクトル分析のスペクトルエネルギーの平均値として計算される。すなわち、下記の式により表される。

$$E_t = 10 * \log(0.5(E_{frame}(0) + E_{frame}(1))), \text{ dB} \quad (5)$$

【0036】

スペクトル分析モジュールの出力パラメータ、すなわち、臨界帯域毎の平均エネルギーと、周波数ビン当たりのエネルギーと、総エネルギーとは、VADモジュール、ノイズ低減モジュールおよびレート選択モジュールの中で利用される。

【0037】

8000 サンプル/秒でサンプリングされた狭帯域入力用として、12800 サンプル/秒でサンプリング変換を行った後、スペクトルの両端部には内容が存在しなくなり、したがって、第1の低周波数臨界帯域並びに最後の3つの高周波数帯域は出力パラメータの計算時に考慮されなくなる ($i = 1$ から 16 までの帯域のみが考慮される) ことに留意されたい。

30

【0038】

音声活性状態検出

前述のスペクトル分析はフレーム当たり2回実行される。前述の式(2)により計算されているように、 $E_{CB}^{(1)}(i)$ と $E_{CB}^{(2)}(i)$ とは、それぞれ第1と第2のスペクトル分析用の臨界帯域情報当たりのエネルギーを示すものとする。フレーム全体に対する臨界帯域毎の平均エネルギーおよび前回のフレームの一部は下記の式として計算される。

40

【数 8】

$$E_{av}(i) = 0.2E_{CB}^{(0)}(i) + 0.4E_{CB}^{(1)}(i) + 0.4E_{CB}^{(2)}(i) \quad (6)$$

但し、 $E_{CB}^{(0)}(i)$ は、前回のフレームの第2の分析から得られる臨界帯域情報当たりのエネルギーを示す。次いで、臨界帯域毎の信号対雑音比 (SNR) は下記の式として計算さ

50

れる。

$$SNR_{CB}(i) = E_{av}(i) / N_{CB}(i) \quad (SNR_{CB} \text{ 1により規定される}) \quad (7)$$

但し、 $N_{CB}(i)$ は、次のセクションで説明することになるような臨界帯域毎の推定されるノイズエネルギーである。次いで、フレーム当たりの平均SNRは、下記の式として計算される。

【数9】

$$SNR_{av} = 10 \log \left(\sum_{i=b_{min}}^{b_{max}} SNR_{CB}(i) \right) \quad (8) \quad 10$$

但し、広帯域信号の場合、 $b_{min} = 0$ かつ $b_{max} = 19$ で、狭帯域信号の場合、 $b_{min} = 1$ かつ $b_{max} = 16$ である。

【0039】

音声活性状態は、長期SNRの関数である或る一定のしきい値とフレーム当たりの平均SNRとを比較することによって検出される。長期SNRは下記の式によって与えられる。 20

【数10】

$$SNR_{LT} = \overline{E}_f - \overline{N}_f \quad (9)$$

但し、 $\overline{E}_f$ と $\overline{N}_f$ （本来なら上記の式(9)のように、EやN等の上部に横線を引いて表すべきであるが、現行の電子出願形式では不可能なため、以後、EやN等の左側の部分に $\overline{}$ を添付して表すこととする）は、それぞれ、後程説明する式(12)と式(13)とを利用して計算される。 $\overline{N}_f$ の初期値は45dBである。 30

【0040】

しきい値は長期SNRの区分一次関数である。2つの関数が利用され、1つの関数はノイズのない音声用の関数であり、1つの関数はノイズの多い音声用の関数である。

【0041】

広帯域信号の場合、 $SNR_{LT} < 35$ （ノイズの多い音声）ならば、

$$th_{VAD} = 0.4346 SNR_{LT} + 13.9575$$

そうでない（ノイズのない音声の）場合には、 40

$$th_{VAD} = 1.0333 SNR_{LT} - 7$$

狭帯域信号の場合、 $SNR_{LT} < 29.6$ （ノイズの多い音声）ならば、

$$th_{VAD} = 0.313 SNR_{LT} + 14.6$$

そうでない（ノイズのない音声の）場合には、

$$th_{VAD} = 1.0333 SNR_{LT} - 7$$

【0042】

さらに、VAD決定時のヒステリシスを活性音声の終了時に追加して、頻繁な切替えを防止する。フレームがソフトハングオーバー期間にある場合に、あるいは、最後のフレームが活性音声フレームである場合に、上記ヒステリシスが印加される。ソフトハングオーバー期間は、2個の連続フレームよりも長い個々の活性音声バースト後の第1の10個のフレ 50

ームから構成される。ノイズの多い音声 ($SNR_{LT} < 3.5$) の場合、上記ヒステリシスは VAD 決定しきい値を下げることになる。

$$th_{VAD} = 0.95 \cdot th_{VAD}$$

【0043】

ノイズのない音声の場合、ヒステリシスは下記の式分だけ VAD 決定しきい値を下げることになる。

$$th_{VAD} = th_{VAD} - 1.1$$

【0044】

フレーム当たりの平均 SNR が VAD 決定しきい値よりも大きければ、すなわち、 $SNR_{av} > th_{VAD}$ ならば、フレームは活性音声フレームと宣言され、VAD フラグとローカルの VAD フラグとは 1 に設定される。フレーム当たりの平均 SNR が VAD 決定しきい値よりも大きくなければ、VAD フラグとローカルの VAD フラグとは 0 に設定される。しかし、ノイズの多い音声の場合、フレーム内でのハードハングオーバー時に、VAD フラグは 1 まで変化する。すなわち、1 つまたは 2 つの非活性フレームが 2 つの連続フレームよりも長い音声時間の後続くことになる（この場合に、ローカルの VAD フラグは 0 に等しくなるが、VAD フラグは 1 まで変化する）。

【0045】

第 1 のレベルのノイズ推定と更新

本セクションでは、総ノイズエネルギー、相対フレームエネルギー、長期平均ノイズエネルギーと長期平均フレームエネルギーとの更新値、および、臨界帯域毎の平均エネルギー並びに騒音補正係数が計算される。さらに、ノイズエネルギーの初期化と下方への更新とが行われる。

【0046】

フレーム当たりの総ノイズエネルギーは下記の式によって与えられる。

【数 1 1】

$$N_{tot} = 10 \log \left(\sum_{i=0}^{19} N_{CB}(i) \right) \quad (10)$$

但し、 $N_{CB}(i)$ は臨界帯域毎の推定ノイズエネルギーである。

【0047】

フレームの相対エネルギーは、dB の形でのフレームエネルギーと長期平均エネルギーとの差分によって与えられる。相対フレームエネルギーは下記の式によって与えられる。

【数 1 2】

$$E_{rel} = E_t - \bar{E}_f \quad (11)$$

但し、 E_t は前述の式 (5) で与えられる。

【0048】

長期平均ノイズエネルギーまたは長期平均フレームエネルギーは、すべてのフレームで更新される。活性音声フレームの場合 (VAD フラグ = 1)、長期平均フレームエネルギーは下記の関係式を利用して更新される。

10

20

30

40

【数 1 3】

$$\overline{E}_f = 0.99\overline{E}_f + 0.01E_t \quad (12)$$

この場合、初期値 / $E_f = 45 \text{ dB}$ を用いる。

【0049】

非活性音声フレーム (VADフラグ = 0) の場合、長期平均ノイズエネルギーは下記の式 10
によって更新される。

【数 1 4】

$$\overline{N}_f = 0.99\overline{N}_f + 0.01N_{tot} \quad (13)$$

【0050】

/ N_f の初期値は、第 1 の 4 つのフレームに対して N_{tot} に等しく設定される。さらに、 20
第 1 の 4 つのフレームでは、/ E_f の値には / E_f / $N_{tot} + 10$ という制限がある。

【0051】

臨界帯域毎のフレームエネルギー、ノイズ初期化、および、下方へのノイズ更新

フレーム全体の臨界帯域毎のフレームエネルギーはフレーム内の双方のスペクトル分析からエネルギーを平均化することによって計算される。すなわち、下記の式により計算される。

【数 1 5】

$$\overline{E}_{CB}(i) = 0.5E_{CB}^{(1)}(i) + 0.5E_{CB}^{(2)}(i) \quad (14) \quad 30$$

【0052】

臨界帯域毎のノイズエネルギー $N_{CB}(i)$ は最初 0.03 に初期化される。しかし、第 1
の 5 個のサブフレームで、信号エネルギーが高すぎないか、信号が強い高周波成分を持たなければ、臨界帯域毎のエネルギーを利用して、ノイズエネルギーの初期化が行われ、これによ
って、処理の当初の開始点からノイズ低減アルゴリズムを効率の良いものにすることが可能となる。2 つの高い周波数比が計算される。 $r_{15,16}$ は、臨界帯域 15 と 16 の平均エ 40
ネルギーと、第 1 の 10 個の帯域での平均エネルギーとの間の比率 (双方のスペクトル分析の
平均値) であり、帯域 18 と 19 とがなければ $r_{18,19}$ は同じである。

【0053】

第 1 の 5 個のフレームにおいて、 $E_t < 49$ かつ $r_{15,16} < 2$ かつ $r_{18,19} < 1.5$ ならば、最初の 3 個のフレームに対して、下記の式が得られる。

【数 1 6】

$$N_{CB}(i) = \overline{E}_{CB}(i) \quad (i = 0, \dots, 19) \quad (15)$$

さらに、次の 2 つのフレームに対して、 $N_{CB}(i)$ は下記の式によって更新される。

【数 1 7】

10

$$N_{CB}(i) = 0.33N_{CB}(i) + 0.66\overline{E}_{CB}(i) \quad (i = 0, \dots, 19) \quad (16)$$

【0 0 5 4】

以下のフレームに対して、この段階で、臨界帯域用の下方へのノイズエネルギー更新が実行され、これによって、エネルギーは暗騒音エネルギー未満となる。第 1 に、一時的に更新されたノイズエネルギーは下記の式として計算される。

20

【数 1 8】

$$N_{imp}(i) = 0.9N_{CB}(i) + 0.1(0.25E_{CB}^{(0)}(i) + 0.75\overline{E}_{CB}(i)) \quad (17)$$

但し、 $E_{CB}^{(0)}(i)$ は前回のフレームからの第 2 のスペクトル分析に対応する。

【0 0 5 5】

次いで、 $i = 0$ から 19 について、 $N_{tmp}(i) < N_{CB}(i)$ ならば、 $N_{CB}(i) = N_{tmp}(i)$ となる。

30

【0 0 5 6】

その後、フレームを非活性フレームとして宣言した場合、 $N_{CB}(i) = N_{tmp}(i)$ と設定することによって第 2 のノイズ更新レベルが実行される。ノイズエネルギー更新値を 2 つの部分に分割する (fragment) 理由として、ノイズ更新が非活性音声フレーム中にのみ実行可能であり、したがって、音声活性状態の決定に必要なすべてのパラメータを必要とするという点が挙げられる。しかし、これらのパラメータは LP 予測分析とオープンループピッチ分析とに依存し、脱ノイズ化 (denoised) 音声信号に対して実行される。可能な限り正確なノイズ推定値を取得するノイズ低減アルゴリズム用として、ノイズ低減の実行前にノイズ推定更新値を下方へ更新し、次いで、フレームが不活性であれば、ノイズ推定更新値は後程上方へ更新する。下方へのノイズ更新は安全であり、この更新は音声活性状態に左右されずに行うことが可能である。

40

【0 0 5 7】

ノイズ低減

信号領域に対してノイズ低減処理を適用し、オーバーラップと追加とを利用して脱ノイズ化信号の再構成が行われる。低減処理は、 g_{min} と 1 との間に制限値を設けたスケーリング利得値であって、当該臨界帯域における信号対雑音比 (SNR) から導き出したスケーリング利得値を用いて、個々の臨界帯域においてスペクトルのスケーリングを行うことにより実行される。ノイズ抑制における新たな特徴として、信号の有声化に関係する或る一定の周波数よりも低い周波数に対して、周波数ビン毎のベースで処理を実行し、臨界帯域

50

ベースでは実行しないという特徴が挙げられる。このようにして、当該周波数ビン内の S N R から導き出されるすべての周波数ビンに対してスケーリング利得値が印加されることになる (S N R は、当該ビンを含む臨界帯域のノイズエネルギーによって分割されたビンエネルギーを利用して計算される)。この新たな特徴によって、歪みを防止するために高調波間でノイズを強く低減させながら、高調波近くの周波数でエネルギーを保持することが可能となる。有声信号に対してのみ、そして、利用される周波数分析の周波数分解能が与えられている場合に、相対的に短いピッチ周期を有する信号に対して上記特徴を利用することが可能となる。しかし、正確に言えば、これらの信号は高調波間のノイズを最も知覚できる信号である。

【 0 0 5 8 】

10

図 3 は、開示された処理手順の概観を示す図である。ブロック 3 0 1 でスペクトル分析が実行される。ブロック 3 0 2 は、有声臨界帯域の数が 0 よりも大きいかどうかを検証するブロックである。この場合、第 1 の有声の K 個の帯域でピン毎の処理を実行するブロック 3 0 4 でノイズ低減処理手順が実行され、残りの帯域で帯域毎の処理が実行される。K = 0 ならば、帯域毎の処理がすべての臨界帯域に適用される。後程説明するように、スペクトルに対するノイズ低減処理を行った後、ブロック 3 0 5 は逆 D F T 分析を実行し、オーバーラップ追加処理を利用して強化された音声信号を再構成する。

【 0 0 5 9 】

最小のスケーリング利得値 $g_{\min}$ が、最大の許されるノイズ低減値 ($N R_{\max}$) から d B で導き出される。この最大の許されるノイズ低減値は、1 4 d B のデフォルト値を有する。したがって、最小のスケーリング利得値は下記の式によって与えられることになる。

20

【 数 1 9 】

$$g_{\min} = 10^{-NR_{\max}/20} \quad (18)$$

そして、この最小のスケーリング利得値は 1 4 d B のデフォルト値として 0 . 1 9 9 5 3 に等しくなる。

30

【 0 0 6 0 】

V A D = 0 の非活性フレームの場合、同じスケーリングがスペクトル全体にわたって適用され、ノイズ抑制が活性状態になった場合 ($g_{\min}$ が 1 よりも低い場合)、このスケーリングは $g_s = 0 . 9 g_{\min}$ によって示される。すなわち、スペクトルのスケーリングがなされた実数成分と虚数成分とは下記の式によって与えられる。

【 数 2 0 】

$$X'_R(k) = g_s X_R(k) \quad (k = 1, \dots, 128) \text{、および、}$$

40

$$X'_I(k) = g_s X_I(k) \quad (k = 1, \dots, 127) \quad (19)$$

狭帯域入力用として前述の式 (1 9) での上限値が 7 9 に設定される (3 9 5 0 H z まで) ことに留意されたい。

【 0 0 6 1 】

活性フレームの場合、臨界帯域毎の S N R と関係して、あるいは、第 1 の有声帯域用として周波数ビン毎にスケーリング利得値が計算される。K_{voic} > 0 の場合、ビン毎のノイ

50

ズ抑制は第1の K_{VOIC} 帯域で実行される。帯域毎のノイズ抑制が残りの帯域に対して利用される。 $K_{VOIC} = 0$ の場合、帯域毎のノイズ抑制がスペクトル全体に対して利用される。後程説明するように、 K_{VOIC} の値が更新される。 K_{VOIC} の最大値は17であり、したがって、3700 Hzの最大周波数に対応する第1の17個の臨界帯域に対してのみ周波数ビン毎の処理を適用することが可能となる。周波数ビン毎の処理を利用できる対象となる周波数ビンの最大数は74（第1の17の帯域内のビンの数）である。例外として、本セクションで後程説明するハードハングオーバーフレームの場合がある。

【0062】

代替実施例では、 K_{VOIC} の値を固定値にすることも可能である。この場合、すべてのタイプの音声フレームで、或る一定の帯域まで周波数ビン毎の処理が実行され、帯域毎の処理が別の帯域に適用される。

10

【0063】

或る一定の臨界帯域内の、あるいは、或る一定の周波数ビンに対するスケーリング利得値がSNRの関数として計算され、下記の式によって与えられる。

$$(g_s)^2 = k_s \text{SNR} + c_s \quad (g_{\min} \leq g_s \leq 1) \quad (20)$$

【0064】

SNR = 1 に対して $g_s = g_{\min}$ 、SNR = 45 に対して $g_s = 1$ のようにして、 k_s と c_s の値が決定される。すなわち、1 dB 以下のSNRに対してはスケーリングを g_s に限定し、45 dB 以上のSNRに対しては所定の臨界帯域内ではノイズ抑制は実行されない ($g_s = 1$)。したがって、これら2つのエンドポイントが与えられれば、前述の式(20)内の k_s と c_s の値は下記の式によって与えられる。

20

【数21】

$$k_s = (1 - g_{\min}^2) / 44, \text{ および } c_s = (45g_{\min}^2 - 1) / 44 \quad (21)$$

【0065】

前述の式(20)内の変数SNRは、処理のタイプに応じて、臨界帯域毎のSNR ($\text{SNR}_{CB}(i)$) か、周波数ビン毎のSNR ($\text{SNR}_{BIN}(k)$) かのいずれかになる。

30

【0066】

フレーム内の第1のスペクトル分析の場合、臨界帯域毎のSNRは下記の式として計算される。

【数22】

$$\text{SNR}_{CB}(i) = \frac{0.2E_{CB}^{(0)}(i) + 0.6E_{CB}^{(1)}(i) + 0.2E_{CB}^{(2)}(i)}{N_{CB}(i)} \quad (i = 0, \dots, 19) \quad (22)$$

40

そして、第2のスペクトル分析の場合、SNRは下記の式として計算される。

【数 2 3】

$$SNR_{CB}(i) = \frac{0.4E_{CB}^{(1)}(i) + 0.6E_{CB}^{(2)}(i)}{N_{CB}(i)} \quad (i = 0, \dots, 19) \quad (23)$$

但し、 $E_{CB}^{(1)}(i)$ と $E_{CB}^{(2)}(i)$ とは、それぞれ、(前述の式(2)で計算されるような)第1および第2のスペクトル分析に対する臨界帯域情報毎のエネルギーを示し、 $E_{CB}^{(0)}(i)$ は、前回のフレームの第2の分析から得られる臨界帯域情報毎のエネルギーを示し、 $N_{CB}(i)$ は臨界帯域毎のノイズエネルギー推定値を示す。

【0067】

第1のスペクトル分析の場合、フレーム内の或る一定の臨界帯域*i*での臨界ピン毎のSNRは下記の式として計算される。

【数 2 4】

$$SNR_{BIN}(k) = \frac{0.2E_{BIN}^{(0)}(k) + 0.6E_{BIN}^{(1)}(k) + 0.2E_{BIN}^{(2)}(k)}{N_{CB}(i)} \quad (k = j_i, \dots, j_i + M_{CB}(i) - 1) \quad (24)$$

20

そして、第2のスペクトル分析の場合、SNRは下記の式として計算される。

【数 2 5】

$$SNR_{BIN}(k) = \frac{0.4E_{BIN}^{(1)}(k) + 0.6E_{BIN}^{(2)}(k)}{N_{CB}(i)} \quad (k = j_i, \dots, j_i + M_{CB}(i) - 1) \quad (25)$$

30

但し、 $E_{BIN}^{(1)}(k)$ と $E_{BIN}^{(2)}(k)$ とは、それぞれ、(前述の式(3)で計算されているような)第1および第2のスペクトル分析用の周波数毎のエネルギーを示し、 $E_{BIN}^{(0)}(k)$ は前回のフレームの第2の分析から得られる周波数毎のエネルギーを示し、 $N_{CB}(i)$ は臨界帯域毎のノイズエネルギー推定値を示す。 j_i は*i*番目の臨界帯域における第1の周波数ピンのインデックスであり、 $M_{CB}(i)$ は、上記で定義された臨界帯域*i*での周波数ピンの数である。

【0068】

インデックス*i*を有する帯域用の臨界帯域毎の処理の場合、前述の式(22)の場合のようにスケーリング利得値を決定した後、そして、前述の式(24)または式(25)に定義されるようにSNRを利用して、周波数毎の分析時に更新される平滑化済みのスケーリング利得値を利用して、実際のスケーリングが実行される。

40

$$g_{BIN,LP}(i) = \alpha_{gs} g_{BIN,LP}(i) + (1 - \alpha_{gs}) g_s \quad (26)$$

【0069】

本発明では、平滑化係数が適応型の係数となり、上記利得自身に逆比例する関係になる新規な特徴が開示される。本実施例では、平滑化係数は $\alpha_{gs} = 1 - g_s$ によって示される。すなわち、平滑化ステップは小さな g_s に対するほど強くなる。このアプローチによって、有声音での立ち上りの場合のように低いSNRフレームに先行する高いSNR音の部分での歪みが防止されることになる。例えば、無声の音声フレームでは、SNRが低くな

50

るため、強いスケーリング利得値を利用して、スペクトル内のノイズの低減が図られることになる。有声音での立ち上りが無声音フレームに後続する場合、S N R はより高い比率になり、利得平滑化ステップがスケーリング利得値の迅速な更新を防止する場合、強いスケーリングが、良好でないパフォーマンスを結果としてもたらしことになる有声音での立ち上り時に利用されることが予想される。提案されたアプローチでは、平滑化处理手順は、立ち上り時に低いスケーリング利得値を迅速に適合させ、このスケーリング利得値を利用することが可能となる。

【 0 0 7 0 】

臨界帯域でのスケーリングは下記の式として実行される。

【 数 2 6 】

10

$$\begin{aligned} X'_R(k+j_i) &= g_{CB,LP}(i) X_R(k+j_i) \text{、および、} \\ X'_I(k+j_i) &= g_{CB,LP}(i) X_I(k+j_i) \quad (k=0,\dots,M_{CB}(i)-1) \end{aligned} \quad (27)$$

但し、 j_i は、臨界帯域 i における第 1 の周波数ピンのインデックスであり、 $M_{CB}(i)$ は当該臨界帯域内の周波数ピンの数である。

20

【 0 0 7 1 】

前述の式 (2 0) に記載のように、インデックス i を持つ帯域内の周波数ピン毎の処理の場合、スケーリング利得値を決定し、前述の式 (2 4) または (2 5) に定義されているように S N R を利用した後、周波数分析毎に更新される平滑化済みのスケーリング利得値を利用して、実際のスケーリングが下記の式として実行される。

$$g_{BIN,LP}(k) = \alpha_{gs} g_{BIN,LP}(k) + (1 - \alpha_{gs}) g_s \quad (28)$$

但し、前述の式 (2 6) の場合と同様、 $\alpha_{gs} = 1 - g_s$

【 0 0 7 2 】

利得の時間的平滑化ステップによって、可聴エネルギーの発振が防止され、一方、例えば、有声音での立ち上りの場合のように、 α_{gs} を利用する平滑化制御ステップによって、低 S N R フレームに先行する高い S N R 音の部分内の歪みが防止される。

30

【 0 0 7 3 】

臨界帯域 i におけるスケーリングは下記の式として実行される。

【 数 2 7 】

$$\begin{aligned} X'_R(k+j_i) &= g_{BIN,LP}(k+j_i) X_R(k+j_i) \text{、および、} \\ X'_I(k+j_i) &= g_{BIN,LP}(k+j_i) X_I(k+j_i) \quad (k=0,\dots,M_{CB}(i)-1) \end{aligned} \quad (29)$$

40

但し、 j_i は、臨界帯域 i 内の第 1 のピンのインデックスであり、 $M_{CB}(i)$ は当該臨界帯域内のピンの数である。

【 0 0 7 4 】

平滑化済みのスケーリング利得値 $g_{BIN,LP}(k)$ と $g_{BIN,LP}(i)$ とは最初 1 に設定される。非活性フレームが処理される度に (V A D = 0)、平滑化された利得値が前述の式 (1 8) で定義された g_{min} にリセットされる。

【 0 0 7 5 】

前述のように、 $K_{VOIC} > 0$ ならば、周波数ピン毎のノイズ抑制が第 1 の K_{VOIC} の帯域に対して実行され、前述の処理手順を利用して帯域毎のノイズ抑制が残りの帯域に対して実

50

行される。すべてのスペクトル分析では、平滑化済みのスケーリング利得値 $g_{\text{BIN}, \text{LP}}(i)$ がすべての臨界帯域に対して更新されることに留意されたい（周波数ビン毎の処理によって処理される有声化帯域の場合であっても、このケースでは、 $g_{\text{BIN}, \text{LP}}(i)$ が帯域 i に属する $g_{\text{BIN}, \text{LP}}(k)$ の平均値を用いて更新される）。同様に、スケーリング利得値 $g_{\text{BIN}, \text{LP}}(k)$ は第 1 の 17 の帯域ですべての周波数ビンに対して更新される。帯域毎の処理を用いる処理帯域の場合、これら 17 の特定の帯域内でスケーリング利得値 $g_{\text{BIN}, \text{LP}}(k)$ を $g_{\text{BIN}, \text{LP}}(i)$ に等しく設定することによって更新される。

【0076】

ノイズのない音声の場合、ノイズ抑制が活性音声フレームでは実行されないことに留意されたい（ $VAD = 1$ ）。これはすべての臨界帯域で最大ノイズエネルギーを発見することによって検出される。 $\max(N_{\text{CB}}(i))$ 、 $i = 0, \dots, 19$ 、そして、この値が 15 以下の場合、ノイズ抑制は実行されない。

【0077】

上述のように、非活性フレーム（ $VAD = 0$ ）に対して、 $0.9 g_{\text{min}}$ のスケーリングがスペクトル全体に適用され、これは、一定のノイズフロアの除去に相当する。 VAD の短いハングオーバーフレームに対して（ $VAD = 1$ かつローカルの $VAD = 0$ ）、帯域毎の処理が前述のように第 1 の 10 の帯域に対して適用され（1700 Hz に対応）、次いで、スペクトルの残り部分に対して定数値 g_{min} によって残りのスペクトルのスケーリングを行うことによって一定のノイズフロアが減算される。この測定ステップによって、高周波数ノイズエネルギー発振が大幅に低減される。10 番目の帯域以上のこれらの帯域に対しては、平滑化済みのスケーリング利得値 $g_{\text{BIN}, \text{LP}}(i)$ はリセットされず、 $g_s = g_{\text{min}}$ で前述の式（26）を利用して更新が行われ、周波数ビン毎の平滑化済みのスケーリング利得値 $g_{\text{BIN}, \text{LP}}(k)$ が、これらのスケーリング利得値を $g_{\text{BIN}, \text{LP}}(i)$ に等しく設定することによって更新される。

【0078】

前述の処理手順は、低減アルゴリズムが処理対象の音声フレームの性質に依存するクラス固有のノイズ低減処理と理解することができる。これは図 4 に例示されている。ブロック 401 は、 VAD フラグが 0（不活性音声）であるかどうかを検証するブロックである。 VAD フラグが 0 であれば、スペクトル全体に同じスケーリング利得値を印加することによって、スペクトルから一定のノイズフロアが除去される（ブロック 402）。 VAD フラグが 0 でなければ、ブロック 403 は、フレームが VAD ハングオーバーフレームであるかどうかの検証を行う。フレームが VAD ハングオーバーフレームであれば、第 1 の 10 個の帯域内で帯域毎の処理が利用され、残りの帯域では同じスケーリング利得値が利用される（ブロック 406）。フレームが VAD ハングオーバーフレームでなければ、ブロック 405 はスペクトルの形で第 1 の帯域内で有声化が検出されたかどうかを検証する。有声化が検出された場合、第 1 の K 個の有声化帯域でビン毎の処理が実行され、帯域毎の処理が残りの帯域で実行される（ブロック 406）。有声帯域が検出されなかった場合、すべての臨界帯域で帯域毎の処理が実行されることになる（ブロック 407）。

【0079】

（12800 Hz までアップサンプリングされる）狭帯域信号の処理の場合、ノイズ化された抑制が第 1 の 17 個の帯域に対して実行される（3700 Hz まで）。3700 Hz と 4000 Hz 間の残りの 5 個の周波数ビンの場合、スペクトルは 3700 Hz でビンにおける最後のスケーリング利得値 g_s を利用してスケールされる。スペクトルの残り部分（4000 Hz から 6400 Hz まで）に対しては、スペクトルはゼロになる。

【0080】

脱ノイズ化信号の再構成

スケーリングされたスペクトル成分 $X_R(k)$ と $X'_I(k)$ の決定後、逆 FFT がスケーリングされたスペクトルに適用され、下記の式に示すように、時間領域においてウィンドウ化された脱ノイズ化信号が取得される。

10

20

30

40

【数 2 8】

$$x_{w,d}(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j2\pi \frac{kn}{N}} \quad (n=0, \dots, L_{FFT}-1)$$

【0081】

10

上記処理がフレーム内の双方のスペクトル分析用として反復されて、脱ノイズ化されたウィンドウ化信号 $x_{w,d}^{(1)}(n)$ と $x_{w,d}^{(2)}(n)$ とが取得される。1/2 フレーム毎に、オーバーラップ追加処理を利用して、分析のオーバーラップ部分に対する信号の再構成が行われる。平方ハニング・ウィンドウがスペクトル分析より前に原信号に対して用いられるため、同じウィンドウがオーバーラップ追加処理より前に逆FFTの出力側で適用される。したがって、2倍のウィンドウ化された脱ノイズ化信号が下記の式によって与えられることになる。

【数 2 9】

20

$$\begin{aligned} x_{ww,d}^{(1)}(n) &= w_{FFT}(n) x_{w,d}^{(1)}(n) \quad (n=0, \dots, L_{FFT}-1) \\ x_{ww,d}^{(2)}(n) &= w_{FFT}(n) x_{w,d}^{(2)}(n) \quad (n=0, \dots, L_{FFT}-1) \end{aligned} \quad (30)$$

【0082】

分析ウィンドウの第1の半分に対して、脱ノイズ化信号を構成するためのオーバーラップ追加処理が下記の式として実行される。

30

【数 3 0】

$$s(n) = x_{ww,d}^{(0)}(n + L_{FFT}/2) + x_{ww,d}^{(1)}(n) \quad (n=0, \dots, L_{FFT}/2-1)$$

次いで、分析ウィンドウの第2の半分に対して、脱ノイズ化信号を構成するオーバーラップ追加処理が下記の式として実行される。

40

【数 3 1】

$$s(n + L_{FFT}/2) = x_{ww,d}^{(1)}(n + L_{FFT}/2) + x_{ww,d}^{(2)}(n) \quad (n=0, \dots, L_{FFT}/2-1)$$

但し、 $x_{w,d}^{(0)}(n)$ は、前回のフレームでの第2の分析から得られる2倍のウィンドウ化された脱ノイズ化信号である。

50

【 0 0 8 3 】

オーバーラップ追加処理と共に、音声エンコーダフレームとノイズ低減フレームとの間に 2 4 サンプルのシフトが存在するため、本フレームに加えて先読みから 2 4 サンプルまで脱ノイズ化信号の再構成が可能であることに留意されたい。しかし、線形予測 (L P) 分析とオープンループピッチ分析のための音声エンコーダによる先読みの終了のためには、別の 1 2 8 サンプルがまだ必要となる。この部分は、オーバーラップ追加処理を実行することなく脱ノイズ化されたウィンドウ化信号 $x_{w,d}^{(2)}(n)$ の第 2 の半分の逆ウィンドウ操作を行うことによって一時的に取得される。すなわち、下記の式により表される。

【 数 3 2 】

$$s(n+L_{FFT}) = x_{wv,d}^{(2)}(n+L_{FFT}/2) / w_{FFT}^2(n+L_{FFT}/2) \quad (n=0, \dots, L_{FFT}/2-1)$$

10

信号の当該部分は、オーバーラップ追加処理の利用時に次のフレームで適正に再計算されることに留意されたい。

【 0 0 8 4 】

ノイズエネルギー推定値更新

このモジュールは、ノイズ抑制用として臨界帯域毎のノイズエネルギー推定値を更新するものである。更新は不活性発話時間中に実行される。しかし、臨界帯域毎の S N R に基づいて上記で実行される V A D 決定は、ノイズエネルギー推定値を更新するかどうかの決定には用いられない。別の決定が、臨界帯域毎の S N R に依存しない別のパラメータに基づいて実行される。ノイズ更新決定用として使用されるパラメータには、以下の、ピッチ安定性、信号の非静的状態、有声化、および、2 次 L P 線形予測残差エネルギーと 1 6 次 L P 線形予測残差エネルギーとの間の比率パラメータがあり、一般に低感度のノイズレベルの変動を有する。

20

【 0 0 8 5 】

ノイズ更新用のエンコーダ V A D の決定を利用しない理由として、ノイズレベルを敏速に変更するためにノイズ推定をロバストなものにするという理由がある。エンコーダの V A D の決定がノイズ更新に利用された場合、ノイズレベルの突然の上昇が、非活性音声フレームに対してさえも S N R の上昇を引き起こす原因になり、ノイズ推定装置の更新を妨げることになり、これが、今度は、次のフレームで S N R を高く維持することになる、等々。したがって、ノイズ更新が阻止されることになり、別の何らかの論理回路がノイズの適合化を再開する必要が生じることになる。

30

【 0 0 8 6 】

本実施例では、オープンループピッチ分析がエンコーダで実行され、3 つのオープンループピッチ推定値が、第 1 のハーフフレーム、第 2 のハーフフレームおよび先読みにそれぞれ対応してフレーム: d_0 、 d_1 、 d_2 毎に計算される。ピッチ安定性カウンタは下記の式として計算される。

40

$$p c = |d_0 - d_{-1}| + |d_1 - d_0| + |d_2 - d_1| \quad (31)$$

但し、 d_{-1} は、前回のフレームの第 2 のハーフフレームの通信のタイムラグである。本実施例では、1 2 2 よりもさらに大きなピッチラグ用として、オープンループピッチ探索モジュールが $d_2 = d_1$ を設定する。したがって、このような通信のタイムラグに対して、前述の式 (3 1) の $p c$ の値に 3 / 2 を乗じて、数式の欠落している第 3 項の補正が行われる。ピッチ安定性は $p c$ の値が 1 2 未満であれば、真となる。さらに、低い有声化用フレームに対して、 $p c$ を 1 2 に設定して、ピッチの不安定性を示す。

すなわち、下記の式により表される。

$$(c_{norm}(d_0) + c_{norm}(d_1) + c_{norm}(d_2)) / 3 + r_e < 0.7 \text{ ならば、 } p c = 12$$

50

3 2)

但し、 $c_{norm}(d)$ は生の正規化相関値であり、 r_e は、低減値を補正するために、暗騒音の存在時の正規化相関値に追加されるオプションの補正值である。本実施例では、デシメーションを受けた重み付き音声信号 $s_{wd}(n)$ に基づいて正規化相関値が計算され、下記の式によって与えられる。

【数 3 3】

$$C_{norm}(d) = \frac{\sum_{n=0}^{L_{sec}} s_{wd}(n) s_{wd}(n-d)}{\sqrt{\sum_{n=0}^{L_{sec}} s_{wd}^2(n) \sum_{n=0}^{L_{sec}} s_{wd}^2(n-d)}} \quad 10$$

但し、加重制限値は自身の遅延時間に依存する。本実施例では、オープンループピッチ分析で使用する重み付き信号に対して 2 だけデシメーションが行われ、加重制限値は、下記の式に従って与えられる。

$d = 10, \dots, 16$ の場合 $L_{sec} = 40$

$d = 17, \dots, 31$ の場合 $L_{sec} = 40$

$d = 32, \dots, 61$ の場合 $L_{sec} = 62$

$d = 62, \dots, 115$ の場合 $L_{sec} = 115$

【0087】

信号の非静的状態推定値は、臨界帯域毎のエネルギーと臨界帯域毎の平均長期エネルギーとの間の比率の積に基づいて実行される。

【0088】

臨界帯域毎の平均長期エネルギーは下記の式によって更新される。

【数 3 4】

$$E_{CB,LT}(i) = \alpha_e E_{CB,LT}(i) + (1 - \alpha_e) \bar{E}_{CB}(i) \quad (i = b_{min}, \dots, b_{max}) \quad (33)$$

但し、広帯域信号の場合、 $b_{min} = 0$ かつ $b_{max} = 19$ であり、狭帯域信号の場合、 $b_{min} = 1$ かつ $b_{max} = 16$ であり、 $E_{CB}(i)$ は、前述の式 (14) で定義される臨界帯域毎のフレームエネルギーである。更新係数 α_e は総フレームエネルギーの一次関数であり、前述の式 (5) で定義され以下のように与えられる。

広帯域信号の場合、 $\alpha_e = 0.0245 E_{tot} - 0.235 \quad (0.5 \leq \alpha_e \leq 0.99)$

狭帯域信号の場合、 $\alpha_e = 0.00091 E_{tot} + 0.3185 \quad (0.5 \leq \alpha_e \leq 0.999)$

)

【0089】

フレームの非静的状態は、フレームエネルギーと臨界帯域毎の平均長期エネルギーとの間の比率の積によって与えられる。すなわち、下記の式により表される。

10

20

30

40

【数 3 5】

$$nonstat = \prod_{i=b_{\min}}^{b_{\max}} \frac{\max(\bar{E}_{CB}(i), E_{CB,LT}(i))}{\min(\bar{E}_{CB}(i), E_{CB,LT}(i))} \quad (34)$$

【0090】

10

ノイズ更新の有声化係数 (voicing) は下記の式によって与えられる。

$$voicing = (c_{norm}(d_0) + c_{norm}(d_1)) / 2 + r_e \quad (35)$$

【0091】

最後に、2次および16次の分析後のLP残留エネルギーとの間の比率は、下記の式によって与えられる。

$$resid_ratio = E(2) / E(16) \quad (36)$$

但し、E(2)およびE(16)は、2次および16次の分析後のLP残留エネルギーであり、当業者にとっては周知のレヴィンソン-ダービン漸化式 (Levinson-Durbin recursion) で計算される。この比率は、信号スペクトル包絡を表すためには、一般に、より高次のLPの方がノイズ用の音声信号にとって必要であるという事実を反映する比率である。言い換えれば、E(2)とE(16)との間の差分はノイズ用の活性音声用の場合よりも低くなるように仮定されている。

20

【0092】

更新値の決定は、当初6に設定される可変ノイズ更新値に基づいて行われ、非活性フレームが検出された場合、1だけ減少し、活性フレームが検出された場合、2だけ増分する。さらに、ノイズ更新値は0と6によって制限される。ノイズエネルギーはnoise_update = 0のときにのみ更新される。

【0093】

可変ノイズの更新値は、個々のフレームにおいて以下のように更新される。

(nonstat > th_{stat}) または (pc < 12) (有声化 > 0.85) または (resid_ratio > th_{resid}) の場合、

30

$$noise_update = noise_update + 2$$

そうでない場合には、

$$noise_update = noise_update - 10$$

但し、広帯域信号に対して、th_{stat} = 350000かつth_{resid} = 1.9、および、狭帯域信号に対して、th_{stat} = 500000かつth_{resid} = 11

【0094】

言い換えれば、(nonstat > th_{stat}) かつ (pc < 12) (有声化 > 0.85) かつ (resid_ratio > th_{resid})、かつ、ノイズ更新が行われる前に6個のフレームのハングオーバーが使用されるとき、フレームはノイズ更新に対して不活性と宣言される。

40

【0095】

したがって、noise_update = 0ならば、

i = 0 . . . 19 に対して、N_{CB}(i) = N_{tmp}(i) になる。但し、N_{tmp}(i) は、前述の式 (17) ですでに計算された一時的に更新されたノイズエネルギーである。

【0096】

有声化カットオフ周波数の更新値

それ未満の周波数の信号が有声音であると見なされるカットオフ周波数が更新される。この周波数を利用して臨界帯域の数が決定され、この臨界帯域に対して周波数ビン毎の処理を利用してノイズ抑制が実行される。

50

【 0 0 9 7 】

第 1 に有声音測定値が下記の式として計算される。

$$v_g = 0.4 c_{norm}(d_1) + 0.6 c_{norm}(d_2) + r_e \quad (37)$$

そして、有声化カットオフ周波数が下記の式によって与えられる。

【 数 3 6 】

$$f_c = 0.00017118 e^{17.9772 v_g} \quad (325 \leq f_c \leq 3700 \text{ により規定される}) \quad (38)$$

10

【 0 0 9 8 】

次いで、 f_c を上回らない上位周波数を有する臨界帯域の数、 K_{voic} が決定される。325 f_c 3700の境界周波数が、最小値3の帯域と最大値17の帯域とで周波数ビン毎の処理が実行されるように設定される（上記の定義された臨界帯域上限値を意味する）。有声音測定値計算時に、先読みの正規化相関に対してさらなる重み付けが与えられることに留意されたい。というのは、所定数の有声帯域が次のフレームで使用されることになるからである。

【 0 0 9 9 】

したがって、上記に記載のように、次のフレームで、第1の K_{voic} 臨界帯域に対して、ノイズ抑制は周波数ビン毎の処理を利用することになる。

20

【 0 1 0 0 】

低い有声化を用いたフレームの場合、および、大きなピッチ遅延時間の場合、臨界帯域毎の処理のみが利用され、この結果 K_{voic} が0に設定されることに留意されたい。以下の条件は次の場合に用いられる。

$(0.4 c_{norm}(d_1) + 0.6 c_{norm}(d_2) < 0.72)$ または $(d_1 > 116)$ または $(d_2 > 116)$ ならば、 $k = 0$

【 0 1 0 1 】

いうまでもなく、他に多くの修正および変更例が可能である。以上述べたような本発明の実施形態についての詳細な説明例および関連図面に鑑みて、上記のような他の修正および変更例は当業者にとっては明らかにである。また一方で、自明なことであるが、本発明の精神と範囲から逸脱することなく、上記のような他の変更例を実施することも可能である。

30

【 図面の簡単な説明 】

【 0 1 0 2 】

【 図 1 】 ノイズ低減を含む音声通信システムを概略的に示すブロック図である。

【 図 2 】 スペクトル分析時のウィンドウ操作を例示する図である。

【 図 3 】 ノイズ低減アルゴリズムの実施例の概観図である。

【 図 4 】 クラスを特定するノイズ低減の実施例を概略的に示すブロック図であって、ノイズ低減アルゴリズムが処理対象の音声フレームの性質に依存するブロック図である。

40

【図 1】

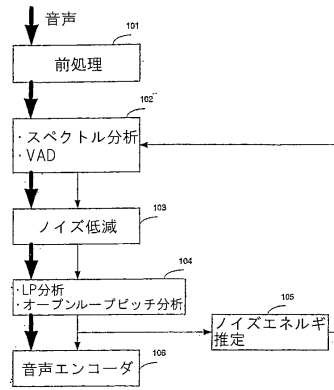


Figure 1 ノイズ低減を含む音声通信システムを概略的に示すブロック図

【図 2】

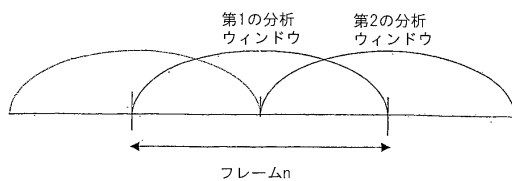


Figure 2 スペクトル分析時のウィンドウ操作を例示する図

【図 3】

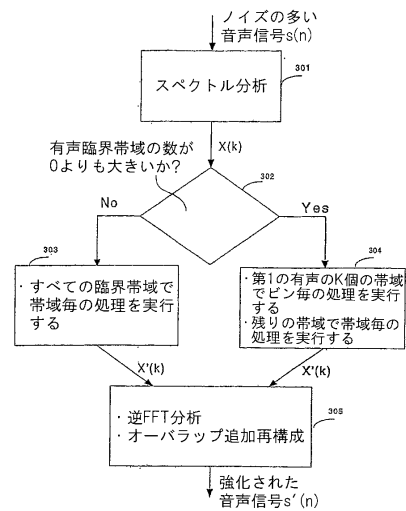


Figure 3 開示されたノイズ低減アルゴリズムの概観図

【図 4】

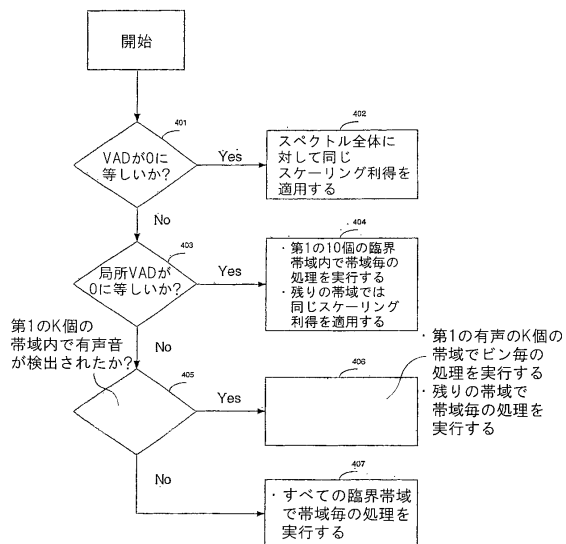


Figure 4 クラスを特定するノイズ低減の概略ブロック図であり、ノイズ低減アルゴリズムは処理対象の音声フレームの性質に依存する

フロントページの続き

(72)発明者 ジェリネック, ミラン

カナダ国, ケベック ジ1 アッシュ 1 ジ2, シェルブルック, モンカルム 8 4 1 # 1

審査官 山下 剛史

(56)参考文献 特開昭57-161800(JP, A)

特開2001-44844(JP, A)

特開2002-149200(JP, A)

特表2002-519719(JP, A)

(58)調査した分野(Int.Cl., DB名)

G10L 15/20, 21/02