

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4127792号
(P4127792)

(45) 発行日 平成20年7月30日(2008.7.30)

(24) 登録日 平成20年5月23日(2008.5.23)

(51) Int.Cl. F I
G 1 0 L 21/02 (2006.01) G 1 0 L 21/02 1 0 1 B
 G 1 0 L 21/02 1 0 2 B

請求項の数 9 (全 9 頁)

(21) 出願番号	特願2002-580312 (P2002-580312)	(73) 特許権者	507219491
(86) (22) 出願日	平成14年3月25日 (2002.3.25)		エヌエックスピー ビー ヴィ
(65) 公表番号	特表2004-519737 (P2004-519737A)		オランダ国 5656エイジー アインド
(43) 公表日	平成16年7月2日 (2004.7.2)		ーフエン ハイ テク キャンパス 60
(86) 国際出願番号	PCT/IB2002/001050	(74) 代理人	100147485
(87) 国際公開番号	W02002/082427		弁理士 杉村 憲司
(87) 国際公開日	平成14年10月17日 (2002.10.17)	(74) 代理人	100134005
審査請求日	平成17年3月24日 (2005.3.24)		弁理士 澤田 達也
(31) 優先権主張番号	01201304.1	(72) 発明者	ギギ エルカン エフ
(32) 優先日	平成13年4月9日 (2001.4.9)		オランダ国 5656 アーアー アイ
(33) 優先権主張国	欧州特許庁 (EP)		ドーフエン プロフ ホルストラーン 6
		審査官	菊池 智紀

最終頁に続く

(54) 【発明の名称】 音声強化デバイス

(57) 【特許請求の範囲】

【請求項1】

オーディオ信号の時間領域サンプルのフレームを周波数領域に変換するための時間 - 周波数変換ユニットと、前記周波数領域において雑音低減を行うためのバックグラウンド雑音低減手段と、前記雑音が低減されたオーディオ信号を前記周波数領域から前記時間領域に変換するための周波数 - 時間変換ユニットとを有する前記バックグラウンド雑音の低減のための音声強化デバイスにおいて、前記バックグラウンド雑音低減手段が、前記オーディオ信号のカレントフレームにおける各々の周波数コンポーネントに対して、前記時間 - 周波数変換ユニットからの、測定された入力強度 $S[k]$ に応じると共に、先行して計算されたバックグラウンド強度 $B_{-1}[k]$ に応じて、予測されたバックグラウンド強度 $B[k]$ を計算するためのバックグラウンドレベル更新ブロックと、各々の前記周波数コンポーネントに対して、前記予測されたバックグラウンド強度 $B[k]$ に応じると共に、前記測定された入力強度 $S[k]$ に応じて、信号対雑音比 $SNR[k]$ を計算するための信号対雑音比ブロックと、各々の前記周波数コンポーネントに対して、前記信号対雑音比 $SNR[k]$ に応じて、前記測定された入力強度 $S[k]$ に対するフィルタ強度 $F[k]$ を計算するためのフィルタ更新ブロックとを有する音声強化デバイスであって、

前記バックグラウンドレベル更新ブロックが、前記先行して計算されたバックグラウンド強度 $B_{-1}[k]$ を得るためのメモリユニットと、 $U[k]$ 及び $D[k]$ が周波数に依存するスケーリングファクタであると共に C が定数である場合、

$$B'[k] = B_{-1}[k] \cdot U[k] \text{ と、}$$

$B''[k] = (B'[k] \cdot D[k]) + (|S[k]| \cdot C \cdot (1 - D[k]))$ と
を用いる一方、最小限許容されるバックグラウンドレベル B_{min} を用いて、関係式

$$B[k] = \max\{\min\{B'[k], B''[k]\}, B_{min}\}$$

による前記先行して予測されたバックグラウンド強度を更新するための処理手段及びコン
パレータ手段とを有することを特徴とする音声強化デバイス。

【請求項 2】

$U[k] = a + k/b$ であることを特徴とする請求項 1 に記載の音声強化デバイス。

【請求項 3】

$D[k] = c - k/d$ であることを特徴とする請求項 1 に記載の音声強化デバイス。

【請求項 4】

前記信号対雑音比ブロックが、前記予測されたバックグラウンド強度 $B[k]$ に応じ
ると共に、関係式

$$SNR[k] = |S[k]| / B[k]$$

による前記測定された入力強度 $S[k]$ に応じて、信号対雑音比 $SNR[k]$ を計算する
ための手段を有することを特徴とする請求項 1 に記載の音声強化デバイス。

【請求項 5】

前記フィルタ更新ブロックが、内部フィルタ値 $F'[k]$ を計算するための第一の手段
と、前記内部フィルタ値から前記測定された入力強度に対する前記フィルタ強度を求める
第二の手段とを有し、前記第一の手段が、先行して計算された内部フィルタ強度 F'_{-1}
 $[k]$ を得るためのメモリユニットと、前記先行して計算された内部フィルタ強度を更新
するための処理手段とを有することを特徴とする請求項 1 に記載の音声強化デバイス。

【請求項 6】

前記第二の手段は、 H が定数であり、 F_{min} が最小のフィルタ値であり、 $F'[k]$
が前記内部フィルタ値である場合、関係式

$$F[k] = \max\{\min\{H \cdot F'[k], 1\}, F_{min}\}$$

による前記フィルタ強度をスケールリングすると共にクリッピングするためのコンパレータ
手段を有することを特徴とする請求項 5 に記載の音声強化デバイス。

【請求項 7】

請求項 1 乃至 6 の何れか一項に記載の音声強化デバイスを備える音声符号化システム。

【請求項 8】

請求項 1 乃至 6 の何れか一項に記載の音声強化デバイスを有する音声エンコーダを備
える音声符号化システム。

【請求項 9】

前記音声エンコーダが、スペクトル振幅ワーピング手段と、ADPCMエンコーダとを
含むプリプロセッサを有する P^2CM エンコーダであり、前記音声強化デバイスの前記バ
ックグラウンド雑音低減手段が、前記プリプロセッサの前記スペクトル振幅ワーピング手
段に統合されることを特徴とする請求項 8 に記載の音声符号化システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、オーディオ信号の時間領域サンプルのフレームを周波数領域に変換する時間 -
周波数変換ユニットと、前記周波数領域における雑音低減を行うバックグラウンド雑音低
減手段と、雑音が低減された前記オーディオ信号を周波数領域から時間領域に変換する周
波数 - 時間変換ユニットとを有する、バックグラウンド雑音の低減のための音声強化デバ
イス (speech enhancement device) に関する。

【0002】

当該音声強化デバイスは、音声符号化システムにおいて、例えば、デジタル電話応答装
置及びボイスメール用途における記憶用途と、“車内 (in-car)” ナビゲーション
システムにおけるボイス応答システムと、インターネット電話のような通信用途とに対して
適用されてもよい。

10

20

30

40

50

【 0 0 0 3 】

【従来の技術】

雑音の高い音声記録の音質を高めるため、雑音レベルが知られていなければならない。単体のマイクロフォンの記録のために、雑音の高い音声のみが利用可能となる。前記雑音レベルは当該信号のみから見積られなければならない。前記雑音を測定する方法は、音声作用 (speech activity) のない記録帯 (region) を使用し、音声作用のある間のサンプルのフレームのスペクトラムを、音声作用のない間に得られるサンプルのフレームのスペクトラムと比較すると共に更新することである。例えば米国特許第 US - A - 6 0 7 0 1 3 7 号が参照される。当該方法が有する問題は、音声作用検出器が使用されなければならないことにある。信号対雑音比が相対的に高いときにさえ、うまく動作するロバストな音声検出器を製作することは困難である。他の問題は、非音声作用帯が非常に狭くてもよく、又はむしろなくてもよいことにある。前記雑音が非定常状態にあるとき、当該雑音特性は、音声作用のある間、変化し得るので、当該アプローチは更により困難となる。

10

【 0 0 0 4 】

音声又は非音声の二者選択を使用せずに、前記信号内の各スペクトルコンポーネントの変化を測定する統計モデルを使用することは更に知られている (IEEE、ASSPに関する通信分野、第32巻、第6号 (1984年12月) (IEEE Trans. on ASSP, vol. 32, No. 6, Dec. 1984) に記載されている、エフライム、マラー (Ephraim, Malah) による "MMSE 短期間スペクトル振幅見積り器を使用する音声強化 (Speech Enhancement Using MMSE Short-Time Spectral Amplitude Estimator)" 参照)。当該方法が有する問題は、前記バックグラウンド雑音が非定常状態にあるとき、前記見積りは最も隣接した時点のフレームに基づいていなければならないことにある。音声発生長 (length speech utterance) において、いくつかの音声スペクトラム帯は、常に、実際の雑音レベル上にあってもよい。これにより、当該スペクトル帯に対する雑音レベルの誤見積りがもたらされる。

20

【 0 0 0 5 】

【発明が解決しようとする課題】

本発明の目的は、音声作用検出器を使用することなく、且つ著しく低減された、雑音レベルの誤見積りを伴って、単体のマイクロフォン音声記録においてバックグラウンド雑音のレベルを予測することにある。

30

【 0 0 0 6 】

【課題を解決するための手段】

従って、本発明によれば、冒頭の段落において記載されているように、前記音声強化デバイスは、前記バックグラウンド雑音低減手段が、前記オーディオ信号のカレントフレームにおける各々の周波数コンポーネントに対して、前記時間 - 周波数変換ユニットからの、測定された入力強度 (magnitude) $S[k]$ に応答すると共に、先行して計算されたバックグラウンド強度 $B_{-1}[k]$ に応答して、予測されたバックグラウンド強度 $B[k]$ を計算するためのバックグラウンドレベル更新ブロックと、各々の前記周波数コンポーネントに対して、前記予測されたバックグラウンド強度 $B[k]$ に応答すると共に、前記測定された入力強度 $S[k]$ に応答して、信号対雑音比 $SNR[k]$ を計算するための信号対雑音比ブロックと、各々の前記周波数コンポーネントに対して、前記信号対雑音比 $SNR[k]$ に応答して、前記測定された入力強度 $S[k]$ に対する前記フィルタ強度 $F[k]$ を計算するためのフィルタ更新ブロックとを有することを特徴とする。

40

【 0 0 0 7 】

本発明は、更に、音声符号化システム、及び、本発明による音声強化デバイスを備える当該音声符号化システム、特に P^2CM オーディオ符号化システムのための音声エンコーダにも関する。特に、適応型差動パルスコード変調 (adaptive differential pulse code modulation (ADPCM)) コーダ、及び

50

上記音声強化システムを具備するプリプロセッサユニットが、前記P²CMオーディオ符号化システムのエンコーダに設けられる。

【0008】

本発明のこれら及び他の態様は、以下に記載された実施例から明らかであり、これらの実施例を参照して説明される。

【0009】

本発明の目的は、上記問題点を解消する方法を提供することにある。

【0010】

【発明の実施の形態】

例として、音声強化デバイスにおいて、前記オーディオ入力信号は、例えば10ミリ秒(millisecond)のフレームに分けられる。例えば8kHzのサンプリング周波数の場合、フレームは80個のサンプルから構成される。各々のサンプルは、例えば16ビットによって表される。

【0011】

BNSは、基本的には、周波数領域適応型フィルタ(frequency domain adaptive filter)である。実際のフィルタリングに先行して、音声強化デバイスの入力フレームは、前記周波数領域に変換されなければならない。フィルタリングの後、前記周波数領域情報は、時間領域に逆変換される。BNSのフィルタの特性は期間に渡って変化するので、フレームの境界における不連続を防止する特別の配慮がなされなければならない。

【0012】

図1は、BNSを具備する音声強化デバイスのブロック図を示している。前記音声強化デバイスは、入力窓形成ユニット(input window forming unit)1、FFTユニット2、バックグラウンド雑音減算器(background noise subtractor(BNS))3、逆FFT(IFFT)ユニット4、出力窓形成ユニット(output window forming unit)5、及びオーバーラップ加算ユニット(overlap-add unit)6を有している。当該例において、入力窓形成ユニット1の80個のサンプル入力フレームは、入力窓 $s[n]$ を形成するために、前記フレームのサイズの2倍、すなわち160個のサンプルのバッファにシフトされる。前記入力窓は、正弦窓 $w[n]$ で加重される。当該例において、スペクトラム $S[k]$ が、256ポイントのFFT2を使用して計算される。BNSブロック3は、当該スペクトラムにおいて周波数領域のフィルタリングをもたらす。その結果 $S^b[k]$ は、IFFT4を使用して時間領域に逆変換される。当該結果は、時間領域表示 $S^b[n]$ をもたらす。ユニット5において、時間領域出力は、前記入力のために使用される正弦窓と同じ正弦窓で加重される。正弦窓による2倍の加重の正味の結果は、ハニングの窓(Hanning window)による加重をもたらす。ユニット5の出力は、 $S^b_w[n]$ によって表される。ハニングの窓は、後続する処理ブロック6、すなわちオーバーラップ加算に対して使用される、好ましい窓の型である。オーバーラップ加算は、二つの連続する出力フレームの間のスムーズな遷移を達成するために使用される。フレーム“i”に対するオーバーラップ加算ユニット6の出力は、

$$S^{*b}_{w,i}[n] = S^b_{w,i}[n] + S^b_{w,i-1}[n+80] \quad (0 \leq n < 80)$$

によって表される。

【0013】

図2は、使用されるウインドウイング及びフレーミングを示している。音声強化デバイスの出力は、一つのフレームの全ディレイ、すなわち当該例において10ミリ秒で、処理された場合(version)の入力信号である。

【0014】

図3は、強度(magnitude)ブロック7、バックグラウンドレベル更新ブロック8、信号対雑音比ブロック9、フィルタ更新ブロック10、及び処理手段11を有する、

周波数領域における適応型フィルタリングのブロック図を示している。後続する演算は、スペクトラム $S[k]$ の各周波数コンポーネント k についてなされる。まず、強度ブロック 7 において、絶対値 $|S[k]|$ は、関係式

$$|S[k]| = [(R\{S[k]\})^2 + (I\{S[k]\})^2]^{1/2}$$

を使用して計算される。ここで、 $R\{S[k]\}$ 及び $I\{S[k]\}$ はそれぞれ、当該例において $0 < k < 129$ の場合のスペクトラムの実数部及び虚数部である。それから、バックグラウンドレベル更新ブロックは、カレントフレームに対して、予期されたバックグラウンド強度 $B[k]$ を計算するために、入力強度 $|S[k]|$ を使用する。

【0015】

信号対雑音比 (SNR) は、関係式

$$SNR[k] = |S[k]| / B[k]$$

を使用して計算され、フィルタ更新ブロック 10 によってフィルタ強度 $F[k]$ を計算するために使用される。

【0016】

最後に、式

$$R^b\{S^b[k]\} = R\{S[k]\} \cdot F[k] \text{ と、}$$

$$I^b\{S^b[k]\} = I\{S[k]\} \cdot F[k] \text{ と}$$

を使用してフィルタリングがなされる。

【0017】

バックグラウンド雑音への全体的な位相の寄与 (overall phase contribution of background noise) は、前記スペクトラムの実数部及び虚数部に渡って一様に分散しているため、周波数領域における振幅の局所的な低減は、加算された位相情報も低減させることが想定される。しかしながら、振幅スペクトラムのみを変化させ、バックグラウンド信号への位相の寄与を変化させなくても十分であるかどうかは議論の余地がある。前記バックグラウンド雑音が周期的な信号のみから構成されている場合には、当該信号の振幅及び位相コンポーネントを測定し、同じ周期と振幅とを備え、180度回転した位相を備える合成信号を加算することは容易であろう。解析期間中の雑音信号への位相の寄与は一定ではなく、さらに、信号対雑音比のみが測定されるので、各周波数領域に対して個別のファクタで入力信号のエネルギーを抑制することし

かできない。これは、通常、バックグラウンドエネルギーだけでなく、音声信号のエネルギーも抑制し得る。しかしながら、知覚のために重要となる音声信号の成分は、通常、他の領域よりも高い信号対雑音比を有しているため、実際上は、本方法は十分満足できる方法である。

【0018】

図 4 は、バックグラウンドレベル更新ブロック 8 を、より詳細に示している。ブロック 8 は、処理手段 12 乃至 16 と、コンパレータ 18 及び 19 を具備するコンパレータ手段 17 と、メモリユニット 20 とを有している。

【0019】

前記バックグラウンドレベルは、後続するステップで更新される。

- まず、メモリユニット 20 及び処理手段 14 を介して、バックグラウンドレベルの先行する値 $B_{-1}[k]$ が、 $B'[k]$ をもたらすファクタ $U[k]$ によって増幅される。

- それから、前記出力は、処理手段 12、13、15、及び 16 を介して得られる現時点の絶対値入力レベル $|S[k]|$ と、増幅されたバックグラウンドレベル $B'[k]$ とのスケールされた合成である値 $B''[k]$ と比較される。コンパレータ 18 によって、より小さな方が、バックグラウンドレベル $B'''[k]$ に対する候補として選択される。

- 最後に、コンパレータ 19 によって、バックグラウンドレベル $B'''[k]$ は、最小限許容されるバックグラウンドレベル B_{min} によって制限され、新たなバックグラウンドレベルをもたらす。これは、バックグラウンドレベル更新ブロック 8 の出力ともなる。

【0020】

10

20

30

40

50

従って、計算されたバックグラウンドの強度は、 $U[k]$ 及び $D[k]$ が周波数に依存するスケールングファクタであり、 C が定数である場合、

$$B'[k] = B_{-1}[k] \cdot U[k] \text{ と、}$$

$$B''[k] = (B'[k] \cdot D[k]) + (|S[k]| \cdot C \cdot (1 - D[k]))$$

とを用いる一方、最小限許容されるバックグラウンドレベル B_{min} を用いて、関係式 $B[k] = \max\{\min\{B'[k], B''[k]\}, B_{min}\}$

で表され得る。

【0021】

本実施例において、入力スケールファクタ C は 4 にセットされる。 B_{min} は 64 にセットされる。スケールング関数 $U[k]$ 及び $D[k]$ は、各フレームに対して一定であり、周波数インデックス k のみに依存している。当該関数は

$$U[k] = a + k/b、\text{ 及び } D[k] = c - k/d$$

として規定される。ここで、 a は 1.002、 b は 16384、 c は 0.97、及び d は 1024 にセットされてもよい。

【0022】

図5は、フィルタ更新ブロック10をより詳細に示している。ブロック10は、処理手段21乃至27と、コンパレータ29及び30を具備するコンパレータ手段28と、メモリユニット31とを有している。

【0023】

ブロック10は、2段、すなわち、内部フィルタ値 $F'[k]$ の適応のための1段と、出力フィルタ値のスケールング及びクリッピング (clipping) のための他の1段とを有している。内部フィルタ値 $F'[k]$ の適応は、関係式 $F''[k] = F'_{-1}[k] \cdot E$ と、

$$F''[k] = (1 - F''[k]) \cdot SNR[k] \text{ と、}$$

$F'[k] = F''[k] (F''[k] < 1 \text{ の場合})$ 、又は $F'[k] = F''[k] + G \cdot F''[k] (F''[k] \geq 1 \text{ 以外の場合})$ と

によれば、入力及びフィルタリングレベルが依存するステップ値によって、先行するフレームの、スケールダウンされた内部フィルタ値を増加させることによってなされる。ここで、 E は 0.9375 にセットされてもよく、 G は 0.0416 にセットされてもよい。

【0024】

前記出力フィルタ値のスケールング及びクリッピングは、

$$F[k] = \max\{\min\{H \cdot F'[k], 1\}, F_{min}\}$$

を用いてなされる。ここで、 H は 1.5 にセットされてもよく、 F_{min} は 0.2 にセットされてもよい。

【0025】

前記出力フィルタリングをさらにスケールング及びクリッピングする理由は、前記バックグラウンド雑音よりもずっと高いエネルギーを有するスペクトル領域に対する帯域通過特性を備えるフィルタを得ることにある。

【0026】

図6は、バックグラウンド雑音が混入している有音声セグメントのフレームに対するフィルタ更新ブロック及び前記バックグラウンドレベルの出力を示している。

【0027】

上記のような独立のバックグラウンド雑音減算器 (BNS) を具備する音声強化デバイスは、音声符号化システム、特に P^2CM 符号化システムのエンコーダにおいて適用されてもよい。前記 P^2CM 符号化システムのエンコーダは、プリプロセッサ及び ADPCM エンコーダを有している。前記プリプロセッサは、特に、例えば、R. リフェブレ及び C. ラフラメによる ICASSP 第1巻、335乃至338頁 (1997年) (R. Lefebvre, C. Laflamme, ICASSP, vol. 1, p. 335 - 338, 1997) に記載の "オーディオ符号化における雑音スペクトラム整形のためのスペクトル振幅ワーピング (Spectral Amplitude Warping

10

20

30

40

50

(SAW) for Noise Spectrum Shaping in Audio Coding)”に記載されているような振幅ワーピングを適用することによって、符号化に先行してオーディオ入力信号の信号スペクトラムを整形(modify)する。当該振幅ワーピングは、周波数領域においてなされるので、前記バックグラウンド雑音の低減が、前記プリプロセッサにおいて実現される。時間-周波数変換後、バックグラウンド雑音の低減及び振幅ワーピングは連続的に達成され、その後、周波数-時間変換がなされる。この場合、前記音声強化デバイスの入力信号は、前記プリプロセッサの入力信号によって形成される。前記プリプロセッサにおいて、当該入力信号は、もたらされる信号における雑音低減が達成される態様で変化するので、ワーピングは、雑音が低減された信号についてなされる。前記入力信号にตอบสนองして得られる、前記プリプロセッサの出力は、遅延型の場合の前記入力フレームを形成し、ADPCMエンコーダに供給される。当該遅延(当該例においては10ミリ秒)は、BNSの内部処理にほぼ起因している。ADPCMエンコーダに対する他の入力信号は、ADPCMエンコーダのビットストリーム出力における符号語のためのビット割り当てを決定するコーデックモード信号(codec mode signal)によって形成される。ADPCMエンコーダは、前段処理された信号フレーム(pre-processed signal frame)における各サンプルに対して符号語を生成する。符号語は、それから、当該例においては、80個の符号のフレームにおさめられる。選択されたコーデックモードに依存して、もたらされたビットストリームは、例えば11.2、12.8、16、21.6、24、又は32kbit/sのビットレートを有する。

10

20

【0028】

上記の実施例は、P²CMオーディオエンコーダにおける信号処理手段において実行可能なコンピュータプログラムの形態であってもよいアルゴリズムによって実現される。図の一部が、あるプログラミング可能な機能を実行するためのユニットを示している場合、当該ユニットは前記コンピュータプログラムのサブパーツとみなされなければならない。

【0029】

記載されている本発明は、記載されている実施例に限定されるものではない。それらに関する変形例は可能である。特に、a、b、c、d、E、G及びHの値が単なる例として与えられており、他の値が可能であることは注意されてもよい。

【図面の簡単な説明】

30

【図1】 本発明による、単独のバックグラウンド雑音減算器(BNS)を具備する音声強化デバイスの基本的なブロック図を示している。

【図2】 BNSにおけるフレーミング及びウィンドウイングを示している。

【図3】 BNSにおける周波数領域適用型フィルタのブロック図を示している。

【図4】 BNSにおけるバックグラウンドレベルの更新のブロック図を示している。

【図5】 BNSにおけるフィルタの更新のブロック図を示している。

【図6】 測定されたバックグラウンドレベル及びもたらされた周波数領域フィルタリングを伴うバックグラウンド雑音が混入しているボイス音声セグメントを示している。

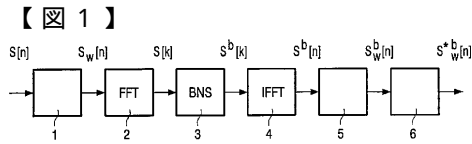


FIG. 1

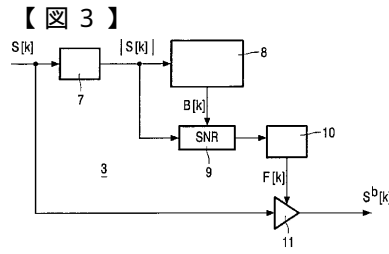
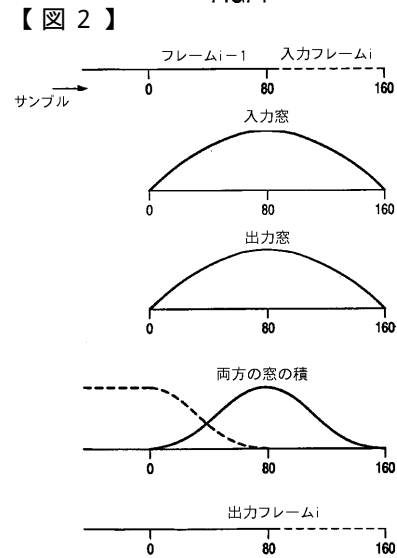


FIG. 3

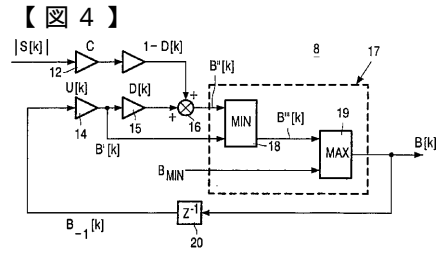


FIG. 4

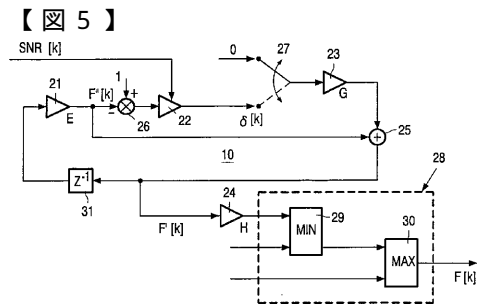
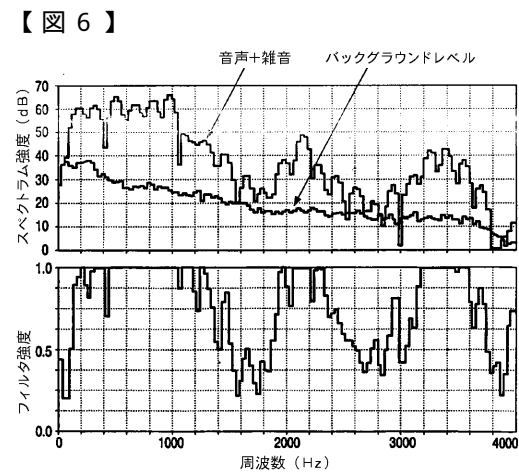


FIG. 5



フロントページの続き

(56)参考文献 米国特許第05706395 (US, A)

加藤正徳 他, "重みつき雑音推定とMMSE STSA法に基づく高音質雑音抑圧", 電子情報通信学会2001年総合大会講演論文集 基礎・境界, 2001年 3月 7日, A-4-21, p.121

(58)調査した分野(Int.Cl., DB名)

G10L 11/00-21/06

JSTPlus(JDreamII)