

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6381366号
(P6381366)

(45) 発行日 平成30年8月29日 (2018. 8. 29)

(24) 登録日 平成30年8月10日 (2018. 8. 10)

(51) Int. Cl.

F I

G 1 O L 21/0224 (2013. 01)

G 1 O L 21/0224

H O 4 R 3/00 (2006. 01)

H O 4 R 3/00 3 2 O

H O 4 N 5/225 (2006. 01)

H O 4 N 5/225 1 0 O

請求項の数 14 (全 17 頁)

(21) 出願番号	特願2014-171649 (P2014-171649)	(73) 特許権者	000001007
(22) 出願日	平成26年8月26日 (2014. 8. 26)		キヤノン株式会社
(65) 公開番号	特開2016-45456 (P2016-45456A)		東京都大田区下丸子3丁目30番2号
(43) 公開日	平成28年4月4日 (2016. 4. 4)	(74) 代理人	100085006
審査請求日	平成29年8月10日 (2017. 8. 10)		弁理士 世良 和信
		(74) 代理人	100100549
			弁理士 川口 嘉之
		(74) 代理人	100106622
			弁理士 和久田 純一
		(74) 代理人	100131532
			弁理士 坂井 浩一郎
		(74) 代理人	100125357
			弁理士 中村 剛
		(74) 代理人	100131392
			弁理士 丹羽 武司

最終頁に続く

(54) 【発明の名称】 音声処理装置、音声処理方法、及び、プログラム

(57) 【特許請求の範囲】

【請求項 1】

第 1 音声信号を取得する第 1 取得手段と、

参照期間を設定する第 1 設定手段と、

前記参照期間と同じ時間幅を有する期間であって、前記参照期間とは異なる期間である、複数の比較期間を設定する第 2 設定手段と、

前記第 1 音声信号に対する処理において取得対象の周波数帯域である対象帯域以外の周波数帯域の音声信号を減衰させる減衰処理を前記第 1 音声信号に施すことにより、第 2 音声信号を取得する第 2 取得手段と、

前記参照期間における第 2 音声信号を各比較期間における第 2 音声信号と比較することにより、複数の比較期間の中から、前記参照期間における第 2 音声信号と類似した第 2 音声信号の期間である複数の類似期間を検出する検出手段と、

前記参照期間における第 1 音声信号と、前記複数の類似期間のそれぞれにおける第 1 音声信号と、に基づいて、前記参照期間における音声信号として設定すべき音声信号である置換信号を生成する生成手段と、

前記参照期間における第 1 音声信号を前記置換信号に置換する置換手段と、
を有することを特徴とする音声処理装置。

【請求項 2】

前記減衰処理は、前記対象帯域の音声信号を抽出する抽出処理であることを特徴とする請求項 1 に記載の音声処理装置。

10

20

【請求項 3】

前記減衰処理は、前記対象帯域の音声信号を通過させるフィルタを用いたフィルタ処理である

ことを特徴とする請求項 1 または 2 に記載の音声処理装置。

【請求項 4】

前記対象帯域は、500 Hz 以上且つ 1500 Hz 以下の周波数帯域を含むことを特徴とする請求項 1 ～ 3 のいずれか 1 項に記載の音声処理装置。

【請求項 5】

複数の動作モードにそれぞれ対応する複数の周波数帯域が予め定められており、

前記音声処理装置は、前記複数の周波数帯域の中から、設定されている動作モードに対応する周波数帯域を、前記対象帯域として選択する選択手段、をさらに有することを特徴とする請求項 1 ～ 4 のいずれか 1 項に記載の音声処理装置。

10

【請求項 6】

前記第 1 音声信号に基づいて前記対象帯域を決定する決定手段、をさらに有することを特徴とする請求項 1 ～ 4 のいずれか 1 項に記載の音声処理装置。

【請求項 7】

前記決定手段は、前記第 1 音声信号を用いた周波数解析の結果に基づいて前記対象帯域を決定することを特徴とする請求項 6 に記載の音声処理装置。

【請求項 8】

20

前記決定手段は、前記第 1 音声信号における第 1 フォルマントの周波数を含む周波数帯域を、前記対象帯域として決定することを特徴とする請求項 6 または 7 に記載の音声処理装置。

【請求項 9】

前記生成手段は、前記参照期間における第 1 音声信号と、前記複数の類似期間のそれぞれにおける第 1 音声信号と、を重みづけ加算することにより、前記置換信号を生成することを特徴とする請求項 1 ～ 8 のいずれか 1 項に記載の音声処理装置。

【請求項 10】

前記生成手段は、前記類似期間における第 1 音声信号の重みとして、前記参照期間における第 2 音声信号との第 2 音声信号の類似度が高いほど大きい重みを使用することを特徴とする請求項 9 に記載の音声処理装置。

30

【請求項 11】

前記検出手段は、前記参照期間における第 2 音声信号との第 2 音声信号の類似度が高い比較期間から順番に N 個（N は 2 以上の整数）の比較期間のそれぞれを、前記類似期間として検出することを特徴とする請求項 1 ～ 10 のいずれか 1 項に記載の音声処理装置。

【請求項 12】

光学レンズと、

前記光学レンズを駆動する駆動手段と、
を有し、

40

前記第 1 設定手段は、前記光学レンズの駆動期間を前記参照期間として設定することを特徴とする請求項 1 ～ 11 のいずれか 1 項に記載の音声処理装置。

【請求項 13】

第 1 音声信号を取得する第 1 取得ステップと、

参照期間を設定する第 1 設定ステップと、

前記参照期間と同じ時間幅を有する期間であって、前記参照期間とは異なる期間である、複数の比較期間を設定する第 2 設定ステップと、

前記第 1 音声信号に対する処理において取得対象の周波数帯域である対象帯域以外の周波数帯域の音声信号を減衰させる減衰処理を前記第 1 音声信号に施すことにより、第 2 音声信号を取得する第 2 取得ステップと、

50

前記参照期間における第 2 音声信号を各比較期間における第 2 音声信号と比較することにより、複数の比較期間の中から、前記参照期間における第 2 音声信号と類似した第 2 音声信号の期間である複数の類似期間を検出する検出ステップと、

前記参照期間における第 1 音声信号と、前記複数の類似期間のそれぞれにおける第 1 音声信号と、に基づいて、前記参照期間における音声信号として設定すべき音声信号である置換信号を生成する生成ステップと、

前記参照期間における第 1 音声信号を前記置換信号に置換する置換ステップと、
を有することを特徴とする音声処理方法。

【請求項 1 4】

請求項 1 3 に記載の音声処理方法の各ステップをコンピュータに実行させることを特徴とするプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音声処理装置、音声処理方法、及び、プログラムに関する。

【背景技術】

【0002】

撮像した動画とともに音声を記録することができる撮像装置では、光学系の駆動により発生する雑音を含む音声記録されてしまうことがある。

このような課題を解決するための従来技術は、例えば、特許文献 1 に開示されている。

特許文献 1 に開示の技術では、撮像装置のモータ（アイリスモータ、シャッターモータ、等）が駆動した場合に、モータの駆動する直前の音声をを用いて、雑音が発生する期間の音声補正される。

【先行技術文献】

【特許文献】

【0003】

【特許文献 1】特開 2006 - 203376 号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかしながら、特許文献 1 の技術は循環バッファを必要とするため、特許文献 1 の技術を用いて雑音を低減できる期間の長さは、物理的な制限によって制限される。

そのため、特許文献 1 の技術を用いたとしても、雑音を高精度に低減することができないことがある。

【0005】

本発明は、音声に含まれている雑音を高精度に低減することができる技術を提供することを目的とする。

【課題を解決するための手段】

【0006】

本発明の第 1 の態様は、

第 1 音声信号を取得する第 1 取得手段と、

参照期間を設定する第 1 設定手段と、

前記参照期間と同じ時間幅を有する期間であって、前記参照期間とは異なる期間である、複数の比較期間を設定する第 2 設定手段と、

前記第 1 音声信号に対する処理において取得対象の周波数帯域である対象帯域以外の周波数帯域の音声信号を減衰させる減衰処理を前記第 1 音声信号に施すことにより、第 2 音声信号を取得する第 2 取得手段と、

前記参照期間における第 2 音声信号を各比較期間における第 2 音声信号と比較することにより、複数の比較期間の中から、前記参照期間における第 2 音声信号と類似した第 2 音声信号の期間である複数の類似期間を検出する検出手段と、

10

20

30

40

50

前記参照期間における第 1 音声信号と、前記複数の類似期間のそれぞれにおける第 1 音声信号と、に基づいて、前記参照期間における音声信号として設定すべき音声信号である置換信号を生成する生成手段と、

前記参照期間における第 1 音声信号を前記置換信号に置換する置換手段と、
を有することを特徴とする音声処理装置である。

【0007】

本発明の第 2 の態様は、

第 1 音声信号を取得する第 1 取得ステップと、

参照期間を設定する第 1 設定ステップと、

前記参照期間と同じ時間幅を有する期間であって、前記参照期間とは異なる期間である、複数の比較期間を設定する第 2 設定ステップと、

前記第 1 音声信号に対する処理において取得対象の周波数帯域である対象帯域以外の周波数帯域の音声信号を減衰させる減衰処理を前記第 1 音声信号に施すことにより、第 2 音声信号を取得する第 2 取得ステップと、

前記参照期間における第 2 音声信号を各比較期間における第 2 音声信号と比較することにより、複数の比較期間の中から、前記参照期間における第 2 音声信号と類似した第 2 音声信号の期間である複数の類似期間を検出する検出ステップと、

前記参照期間における第 1 音声信号と、前記複数の類似期間のそれぞれにおける第 1 音声信号と、に基づいて、前記参照期間における音声信号として設定すべき音声信号である置換信号を生成する生成ステップと、

前記参照期間における第 1 音声信号を前記置換信号に置換する置換ステップと、
を有することを特徴とする音声処理方法である。

【0008】

本発明の第 3 の態様は、上述した音声処理方法の各ステップをコンピュータに実行させることを特徴とするプログラムである。

【発明の効果】

【0009】

本発明によれば、音声に含まれている雑音を高精度に低減することができる。

【図面の簡単な説明】

【0010】

【図 1】本実施形態に係る音声処理部の機能構成の一例を示すブロック図

【図 2】本実施形態に係る撮像装置の外観及び機能構成の一例を示す図

【図 3】本実施形態に係る雑音低減処理の一例を示す図

【図 4】本実施形態に係る各音声信号と減衰処理の特性の一例を示す図

【図 5】本実施形態に係る雑音低減処理の流れの一例を示すフローチャート

【図 6】本実施形態に係る音声処理部の機能構成の一例を示すブロック図

【図 7】従来の雑音低減処理の一例を示す図

【図 8】従来の雑音低減処理において生じる課題の一例を示す図

【発明を実施するための形態】

【0011】

以下、図面を参照して本発明の実施形態に係る音声処理装置、撮像装置、及び、音声処理方法について詳細に説明する。

なお、以下の実施形態はあくまで一例であり、本発明は以下の実施形態に限定されない。

【0012】

(撮像装置の構成)

本実施形態に係る音声処理装置の一例として、以下カメラ 1 について説明する。

図 2 (a) は、カメラ 1 の外観の一例を示す斜視図である。図 2 (b) は、カメラ 1 の構成の一例を示すブロック図である。

図 2 (b) に示すように、カメラ 1 は、カメラシステム制御部 10、撮像レンズ 11、

10

20

30

40

50

マイクロフォン 12、撮像素子 13、画像処理部 14、レンズ駆動部 15、音声処理部 16、メモリ部 17、操作部 18、画像表示部 19、等を有する。

【0013】

撮像レンズ 11 を通過した光線は、撮像素子 13 の近傍で結像をし、撮像素子 13 に適正な時間だけ露光される。

撮像素子 13 は、露光された光を電気信号（アナログ信号）に光電変換する。

画像処理部 14 は、A/D変換器、ホワイトバランス回路、ガンマ補正回路、補間演算回路、等の処理部（処理回路）を有する。画像処理部 14 は、これらの処理部を用いて、撮像素子 13 で生成されたアナログ信号に様々な処理を施すことにより、デジタル信号である撮像画像データを生成する。生成された撮像画像データは、カメラシステム制御部 10 を介して、メモリ部 17 に記録される。

10

レンズ駆動部 15 は、カメラシステム制御部 10 からの指示（命令）に応じて撮像レンズ 11 を駆動することにより、撮像レンズ 11 の光学状態を調整する。具体的には、レンズ駆動部 15 は、カメラシステム制御部 10 からの指示に応じて、撮像レンズ 11 が有するフォーカスレンズ群、絞り機構、手振れ防振機構、等を駆動する。

撮像素子 13 に連続的に露光し、撮像素子 13 からアナログ信号を読み出して撮像画像データを生成する処理を一定のフレームレートで行うことにより、動画の撮像を行うことができる。

【0014】

マイクロフォン 12 には、音声が入力される。マイクロフォン 12 は、入力された音声を表す音声信号（アナログ信号またはデジタル信号）を生成する。本実施形態では、マイクロフォン 12 は、動画の撮像中に入力された音声（被写体の音声を少なくとも含む音声）を表す音声信号を生成する。

20

音声処理部 16 は、マイクロフォン 12 で生成された音声信号を取得し、取得した音声信号に種々の処理を施すことにより、デジタル信号である音声信号（出力音声信号）を生成する。音声処理部 16 が行う処理は、A/D変換処理、雑音低減処理、等を含む。雑音低減処理は、マイクロフォン 12 で生成された音声信号が表す音声に含まれている雑音を低減する処理である。雑音低減処理の詳細については後述する。生成された出力音声信号は、カメラシステム制御部 10 を介して、メモリ部 17 に記録される。出力音声信号は、例えば、動画の撮像画像データに対応付けられてメモリ部 17 に記録される。

30

【0015】

操作部 18 は、カメラ 1 に対するユーザ操作を受け付ける。本実施形態では、操作部 18 は、図 2（a）のシャッターリリース釦 18a を含む 1 つ以上の釦を有する。

カメラシステム制御部 10 は、ユーザ操作に応じて生成された操作信号（タイミング信号）に応じて、カメラ 1 の各機能部を制御する。例えば、シャッターリリース釦 18a の押下が検出されると、撮像素子 13 の駆動、画像処理部 14 の動作、音声処理部 16 の動作、メモリ部 17 に記録するデータや信号の圧縮処理、等が制御される。また、カメラシステム制御部 10 は、画像表示部 19 の画像や情報の表示を制御する。

【0016】

（音声処理部 16 の構成）

40

図 2（b）の音声処理部 16 の構成について説明する。

動画撮像中にマイクロフォン 12 で生成された音声信号が表す音声には、被写体の音声のみが含まれていることが好ましい。しかしながら、動画撮像中にマイクロフォン 12 で生成された音声信号が表す音声には、撮像レンズ 11 の駆動に伴って生じたレンズ駆動雑音、マイクロフォン 12 の性能に起因して生じた白色雑音である暗雑音、等の雑音が重畳されていることがある。このように、マイクロフォン 12 で生成された音声信号が表す音声には、雑音が含まれていることがある。

音声処理部 16 は、雑音低減処理を行うことにより、上述した雑音を低減する。

【0017】

図 1 は、音声処理部 16 の機能構成の一例を示すブロック図である。

50

図 1 では、データ（信号）と機能部とを区別しやすくするために、駆動部は四隅がとがった四角、データ（信号）は四隅が丸まった四角で図示されている。

図 1 に示すように、音声処理部 16 は、音声信号減衰部 31、類似期間検出部 32、置換信号生成部 33、参照期間設定部 34、音声信号置換部 35、等を有する。

図 1 において、入力音声信号 21 は、音声信号であり、マイクロフォン 12 で生成されたデジタル信号、マイクロフォン 12 で生成されたアナログ信号に A/D 変換処理を施したデジタル信号、等である。

【0018】

音声信号減衰部 31 は、入力音声信号 21（第 1 音声信号）を取得する（第 1 取得処理）。

10

また、音声信号減衰部 31 は、注目帯域以外の周波数帯域である非注目帯域の音声信号を減衰させる減衰処理を入力音声信号 21 に施すことにより、減衰音声信号（第 2 音声信号）を取得（生成）する（第 2 取得処理）。注目帯域は、入力音声信号 21 に対する雑音低減処理において注目すべき周波数帯域である。減衰処理は、注目帯域の音声信号を抽出する抽出処理と言うこともできる。

音声信号減衰部 31 は、減衰音声信号を類似期間検出部 32 に出力する。

【0019】

なお、減衰処理（抽出処理）の方法は特に限定されない。例えば、減衰処理は、注目帯域の音声信号を通過させるフィルタ（バンドパスフィルタ；BPF）を用いたフィルタ処理であってもよい。

20

なお、注目帯域や非注目帯域は特に限定されない。注目帯域と非注目帯域は、予め定められた周波数帯域であってもよいし、ユーザによって設定可能なものであってもよい。例えば、注目帯域と非注目帯域の少なくとも一方が、撮像対象、カメラ 1 の動作モード、ユーザ操作、等に応じて決定されてもよい。

【0020】

なお、人間が発する声がマイクロフォン 12 に入力される音声として想定されている場合には、注目帯域は、人間が発する声の周波数帯域を含むことが好ましい。具体的には、注目帯域は、人間が発する声の第 1 フォルマントに相当する周波数帯域を含むことが好ましい。一般的に、成人が発する声の第 1 フォルマントの周波数帯域は、500 Hz 以上且つ 1500 Hz 以下の周波数帯域と言われている。そのため、注目帯域は、500 Hz 以上且つ 1500 Hz 以下の周波数帯域を含むことが好ましい。また、成人が発する声の第 2 フォルマントの周波数帯域は、1500 Hz 以上且つ 3000 Hz 以下の周波数帯域と言われている。そして、成人が発する声には、第 2 フォルマントの周波数よりも高い周波数帯域に、第 3 フォルマントの周波数および第 4 フォルマントの周波数が存在していると言われている。

30

本実施形態では、注目帯域が、第 1 フォルマントの周波数帯域と第 2 フォルマントの周波数帯域とを含む周波数帯域である場合の例を説明する。具体的には、注目帯域が、500 Hz 以上且つ 3000 Hz 以下の周波数帯域である場合の例を説明する。

なお、第 1 取得処理は、音声信号減衰部 31 とは異なる機能部によって実行されてもよい。

40

【0021】

参照期間設定部 34 は、音声信号減衰部 31 から出力された減衰音声信号に対して、参照期間を設定する（第 1 設定処理）。参照期間は、雑音低減処理の対象の期間である。本実施形態では、参照期間として、所定の時間幅を有する期間が設定される。参照期間設定部 34 は、参照期間を類似期間検出部 32 に通知する。

なお、参照期間の時間幅は予め定められていなくてもよい。例えば、参照期間の時間幅は、撮像対象、カメラ 1 の動作モード、ユーザ操作、等に応じて決定されてもよい。

【0022】

類似期間検出部 32 は、音声信号減衰部 31 から出力された減衰音声信号に対して、複数の比較期間を設定する（第 2 設定処理）。比較期間は、参照期間と同じ時間幅を有する

50

期間であり、且つ、参照期間とは異なる期間である。

また、類似期間検出部 3 2 は、参照期間における減衰音声信号を各比較期間における減衰音声信号と比較する。そして、類似期間検出部 3 2 は、その比較結果に基づいて、複数の比較期間の中から、参照期間における減衰音声信号と類似した減衰音声信号の期間である複数の類似期間を検出する（検出処理）。例えば、参照期間における減衰音声信号との減衰音声信号の類似度が高い比較期間から順番に N 個（N は 2 以上の整数）の比較期間のそれぞれが、類似期間として検出される。

そして、類似期間検出部 3 2 は、各類似期間を少なくとも表す類似期間信号 2 2 を出力する。本実施形態では、類似期間信号 2 2 として、複数の類似期間にそれぞれ対応する複数の信号類似度の大小関係をさらに表す信号が、生成され、出力される。信号類似度は、参照期間における減衰音声信号と類似期間における減衰音声信号との間の類似度である。

【 0 0 2 3 】

なお、第 2 設定処理は、類似期間検出部 3 2 とは異なる機能部によって実行されてもよい。

なお、類似期間の検出方法は上記方法に限らない。例えば、信号類似度が閾値以上である複数の比較期間のうち、参照期間に時間的に近い比較期間から順番に N 個の比較期間のそれぞれが、類似期間として検出されてもよい。

なお、N の値は、予め定められた固定値であってもよいし、ユーザによって設定されてもよい。例えば、N の値は、撮像対象、カメラ 1 の動作モード、ユーザ操作、等に応じて決定されてもよい。

【 0 0 2 4 】

置換信号生成部 3 3 は、参照期間における入力音声信号と、複数の類似期間のそれぞれにおける入力音声信号と、に基づいて、置換信号 2 3 を生成する。置換信号 2 3 は、参照期間における出力音声信号として設定すべき音声信号である。参照期間における入力音声信号は、参照期間における減衰音声信号に対応する入力音声信号であり、類似期間における入力音声信号は、類似期間における減衰音声信号に対応する入力音声信号である。

【 0 0 2 5 】

音声信号置換部 3 5 は、参照期間における入力音声信号を置換信号 2 3 に置換することにより、出力音声信号 2 4 を生成する。

【 0 0 2 6 】

（従来の雑音低減処理）

従来の雑音低減処理の一例について説明する。詳細は以下で述べるが、従来の雑音低減処理では、減衰音声信号は生成されない。

図 7 (a) ~ 7 (e) は、従来の雑音低減処理の一例を示す模式図である。図 7 (a) の上側には、被写体の音声に白色雑音（暗雑音）が重畳された入力音声信号の一例が示されている。図 7 (a) の下側には、参照期間における入力音声信号と各類似期間における入力音声信号とが、他の期間における入力音声信号から切り離されて図示されている。図 7 (b) は、置換信号の一例を示す。図 7 (c) は、参照期間における入力音声信号を置換信号に置き換えて得られる出力音声信号の一例を示す。図 7 (d) は、出力音声信号の他の例を示す。図 7 (e) は、被写体の音声に対して一時的にレンズ駆動雑音が重畳された入力音声信号の一例を示す。図 7 (a) ~ 7 (e) において、横軸は時間位置を示し、縦軸は音声信号レベル（音声信号の信号レベル）を示す。図 7 (a) の上側，7 (c) ，7 (d) ，7 (e) は、入力音声信号 2 1 や出力音声信号 2 4 の一部を拡大した拡大図である。図 7 (a) の上側，7 (c) ，7 (d) ，7 (e) に示す音声信号は、0 . 2 秒程度の音声信号である。図 7 (a) の上側の音声信号を局所的に観察すると、音声信号の繰り返し性が非常に高いことが分かる。以下で説明する従来の雑音低減処理は、音声信号が有する短時間での繰り返し性の高さに着目した処理である。短時間での繰り返し性の高さは、本実施形態でも着目される。

【 0 0 2 7 】

まず、図 7 (a) に示すように、入力音声信号に対して、参照期間 1 0 0 が設定される

。参照期間における入力音声信号に第 1 フォルマントの周波数の 1 周期分の音声信号が含まれるように、参照期間の長さ（時間）が設定されていることが好ましい。即ち、参照期間の長さは、第 1 フォルマントの周波数の 1 周期以上であることが好ましい。例えば、成人が発する声の第 1 フォルマントの周波数は 5 0 0 H z 以上且つ 1 0 0 0 H z 以下の周波数と言われているため、参照期間の長さは、2 m s e c (= 0 . 0 0 2 s e c = 1 ÷ 5 0 0 H z) 以上であることが好ましい。

【 0 0 2 8 】

次に、複数の比較期間が設定される。例えば、参照期間に対して時間的に前の期間と、参照期間に対して時間的に後の期間と、の少なくとも一方を含む複数の期間が、複数の比較期間として設定される。上述したように、比較期間の時間幅は、参照期間の時間幅と等しい。

10

なお、比較期間と、当該比較期間に隣接する隣接期間（参照期間または比較期間）と、の間の時間差は、特に限定されない。上記時間差は、例えば、処理負荷、想定される音声の周波数、等を考慮して決定される。上記時間差は、音声信号レベルのサンプリングレートの 1 ビット分であることが好ましい。比較期間の一部が隣接期間の一部に重畳されていてもよいし、比較期間は隣接期間から離れていてもよい。

【 0 0 2 9 】

そして、参照期間 1 0 0 における入力音声信号を各比較期間における入力音声信号と比較することにより、複数の比較期間の中から複数の類似期間が検出される。図 7 (a) の例では、3 つの類似期間 1 0 1 a , 1 0 1 b , 1 0 1 c が検出されている。

20

【 0 0 3 0 】

類似期間の検出方法の一例を以下に説明する。

なお、類似期間の検出方法は、以下の方法に限らない。

【 0 0 3 1 】

まず、比較期間毎に、参照期間における入力音声信号と比較期間における入力音声信号との間の類似度が算出される。類似度は、例えば、以下の式 1 を用いて算出される。

【数 1】

$$D = \sum_{i=1}^M |S_C(i) - S_R(i)| \quad \cdots (式 1)$$

30

参照期間及び比較期間は、M 個（M は 2 以上の整数）の離散時間位置を含む。M の値は、参照期間（または比較期間）の長さを音声信号レベルのサンプリングレートで除算することにより、算出することができる。式 1 において、 $S_C(i)$ は比較期間の i 番目（ i は 1 以上且つ M 以下の整数）の離散時間位置における入力信号レベル（入力音声信号の信号レベル）であり、 $S_R(i)$ は参照期間の i 番目の離散時間位置における入力信号レベルである。D は、非類似度である。類似度は、例えば、非類似度 D の逆数である。

【 0 0 3 2 】

式 1 では、各離散時間位置におけるレベル差（参照期間における入力信号レベルと比較期間における入力信号レベルとの間の差の絶対値）の総和が、非類似度 D として算出される。そのため、比較期間における入力音声信号が参照期間における入力音声信号に近いほど小さい値が、非類似度 D として算出される。そして、比較期間における入力音声信号が参照期間における入力音声信号と完全に一致する場合に、非類似度 D として 0 が算出される。

40

【 0 0 3 3 】

次に、類似度が高い比較期間から順番に N 個（N は 3 以上の整数）の比較期間のそれぞれが、類似期間として検出される。具体的には、非類似度 D が小さい比較期間から順番に N 個の比較期間のそれぞれが、類似期間として検出される。

【 0 0 3 4 】

50

類似期間が検出された後、参照期間における入力音声信号と、各類似期間における入力音声信号と、を用いて、置換信号が生成される。置換信号は、例えば、以下の式 2 を用いて算出される。

【数 2】

$$S_0(i) = w_R S_R(i) + \sum_{K=1}^N \sum_{i=1}^M w_K S_{CK}(i) \quad \cdots \text{(式 2)}$$

式 2 において、 i と M は式 1 と同じである。 N は類似期間の総数であり、 K は類似期間の番号である。 K は、1 以上且つ N 以下の整数である。 $S_0(i)$ は i 番目の離散時間位置における置換信号レベル（置換信号の信号レベル）であり、 $S_R(i)$ は参照期間の i 番目の離散時間位置における入力信号レベルである。 $S_{CK}(i)$ は、番号 K の類似期間の i 番目の離散時間位置における入力信号レベルである。 w_R は参照期間における入力音声信号の重みであり、 w_K は番号 K の類似期間における入力音声信号の重みである。式 2 では、参照期間における入力音声信号と各類似期間における入力音声信号とを重みづけ加算することにより、置換信号が生成される。類似期間における音声信号の重み w_K としては、例えば、参照期間における音声信号との音声信号の類似度が高いほど大きい重みを使用される。即ち、重み w_K としては、非類似度 D が小さいほど大きい重みを使用される。

【0035】

図 7 (b) の音声信号 102 は、参照期間 100 における入力音声信号と、類似期間 101a, 101b, 101c における入力音声信号と、を用いて生成された置換信号である。図 7 (b) から、雑音が低減された音声信号が置換信号 102 として生成されていることがわかる。

なお、置換信号の生成方法は上記方法に限らない。例えば、重み w_R , w_K として 1 を使用し、参照期間における入力音声信号と各類似期間における入力音声信号との平均の音声信号が、置換信号として生成されてもよい。また、重み w_K として、参照期間と類似期間の間の時間差が小さいほど大きい重みを使用されてもよい。

【0036】

次に、参照期間 100 における入力音声信号が、置換信号 102 に置換される。それにより、図 7 (c) の出力音声信号が生成される。図 7 (c) の出力音声信号では、参照期間 100 における暗雑音が低減されている。

図 7 (a) の例では、入力音声信号の全期間にわたって暗雑音が重畳されている。参照期間の時間位置を少しずつずらしながら上述した処理を繰り返し行うことにより、図 7 (d) の出力音声信号を生成することができる。図 7 (d) の出力音声信号では、入力音声信号の全期間にわたって暗雑音が低減されている。

なお、暗雑音以外の雑音についても、上述した処理により低減することができる。例えば、図 7 (e) の入力音声信号に重畳されている雑音（一部の期間 103 に重畳されているレンズ駆動雑音）も、上述した処理により低減することができる。具体的には、参照期間 104a と参照期間 104b を含む複数の参照期間を順番に設定して上述した処理を行うことにより、図 7 (e) の入力音声信号に重畳されている全てのレンズ駆動雑音を低減することができる。

【0037】

しかしながら、上述した従来の雑音低減処理では、高精度に雑音を低減することができないことがある。以下、図 8 (a) ~ 8 (d) を用いて、従来の雑音低減処理において生じる課題について説明する。

【0038】

図 8 (a) は、被写体の音声を表す音声信号（被写体音声信号；雑音が重畳されていない音声信号）の一例を示す図である。図 8 (b) は、雑音を表す音声信号（雑音信号）の一例を示す図である。図 8 (c) は、図 8 (a) の被写体音声信号に図 8 (b) の雑音信

10

20

30

40

50

号が重畳された音声信号を示す図である。図 8 (d) は、図 8 (a) の被写体信号に風雑音と暗雑音とが重畳された音声信号の一例を示す図である。以下では、簡略化のために被写体の音声を表す音声信号の周波数が $F_b [Hz]$ であるものとする。

【 0 0 3 9 】

図 8 (b) に示す雑音の周波数 $[Hz]$ 及びパワー (大きさ) $[dB]$ は、図 7 (a) の入力音声信号に重畳されている暗雑音に比べて大きい。そのため、図 8 (c) に示す音声信号が入力音声信号である場合、入力音声信号に対する雑音の影響が大きいため、類似期間として検出されるべき比較期間の信号類似度が低下し、類似期間の検出精度が低下してしまう。具体的には、被写体音声信号の繰り返し単位の比較期間が類似期間として検出され難くなる。このように、入力音声信号に重畳されている雑音の周波数及びパワーが大

10

きい場合、類似期間の検出精度が低下してしまう。その結果、雑音低減処理の処理精度が低下してしまう。

周波数及びパワーが大きい雑音は、例えば、手振れ防振機構の駆動雑音である。

【 0 0 4 0 】

図 8 (d) に示す音声信号 (風雑音が重畳されている音声信号) が入力音声信号である場合にも、類似期間の検出精度が低下し、雑音低減処理の処理精度が低下してしまう。風雑音は、低周波成分を多く含む。一般的には、風雑音は、 $400Hz$ 以下の周波数帯域に強いパワーを有すると言われている。低周波成分を多く含む雑音が入力音声信号に重畳されている場合にも、類似期間の高精度な検出が困難となる。その結果、類似期間の検出精度が低下し、雑音低減処理の処理精度が低下してしまう。

20

【 0 0 4 1 】

(本実施形態に係る雑音低減処理)

そこで、本実施形態では、入力音声信号に減衰処理を施すことにより、低周波成分を多く含む雑音、周波数及びパワーが大きい雑音、等が低減された減衰音声信号を取得 (生成) する。そして、入力音声信号の代わりに減衰音声信号を用いて、類似期間を検出する。その後、上述した従来の雑音低減処理と同様に、入力音声信号を用いて置換信号及び出力音声信号を生成する。減衰音声信号を用いることにより、類似期間を高精度に検出することができる。その結果、音声に含まれている雑音を高精度に低減することができる。

本実施形態に係る雑音低減処理の一例について説明する。

【 0 0 4 2 】

図 3 (a) は、被写体音声信号の一例を示す図であり、図 3 (b) は、周波数及びパワーが大きい雑音を表す雑音信号の一例を示す図である。図 3 (c) は、入力音声信号の一例を示す図であり、図 3 (d) は、減衰音声信号の一例を示す図である。図 3 (c) の下側には、図 3 (a) の被写体音声信号に図 3 (b) の雑音信号が重畳された入力音声信号の一例が示されている。図 3 (c) の上側には、参照期間における入力音声信号と各類似期間における入力音声信号とが、他の期間における入力音声信号から切り離されて図示されている。図 3 (d) の減衰音声信号は、図 3 (c) の入力音声信号に減衰処理を施すことにより得られた音声信号である。

30

【 0 0 4 3 】

図 4 (a) , 4 (b) は、各音声信号の周波数特性及び減衰処理の処理特性 (フィルタ特性) の一例を示す図である。

40

図 4 (a) , 4 (b) において、横軸は周波数を示し、縦軸はパワーを示す。

図 4 (a) において、実線 6 1 は、図 3 (a) の被写体音声信号の周波数特性を表し、破線 6 2 は、図 3 (b) の雑音信号の周波数特性を表す。図 4 (a) , 4 (b) において、太実線 6 3 は、図 3 (c) の入力音声信号の周波数特性を表す。図 4 (b) において、一点鎖線 6 4 は、減衰処理のフィルタ特性を表し、実線 6 5 は、図 3 (d) の減衰音声信号の周波数特性を表す。

【 0 0 4 4 】

被写体音声信号の周波数特性 6 1 は、周波数帯域 F_1 , F_2 , F_3 , F_4 にピークを有する。周波数帯域 F_1 が第 1 フォルマントの周波数帯域であり、周波数帯域 F_2 が第 2 フ

50

ォルマンの周波数帯域であり、周波数帯域 F 3 が第 3 フォルマンの周波数帯域であり、周波数帯域 F 4 が第 4 フォルマンの周波数帯域である。

雑音信号の周波数特性 6 2 は、高周波数の側にある周波数帯域 F 4 に、他の周波数帯域よりも強めの成分が存在している。このような成分は、類似期間の検出精度を低下させる。

本実施形態では、フィルタ特性 6 4 を有するフィルタを用いた減衰処理（フィルタ処理）を行うことにより、入力音声信号から、第 1 フォルマンの周波数帯域 F 1 と第 2 フォルマンの周波数帯域 F 2 とを含む周波数帯域の音声信号が抽出される。

そのため、減衰音声信号の周波数特性 6 5 では、周波数帯域 F 2 よりも高い周波数の成分が入力音声信号の周波数特性 6 3 から低減されている。

このように、本実施形態では、減衰処理を行うことにより、類似期間の検出精度を低下させる成分が低減された減衰音声信号が得られる。

なお、図 3 (a) ~ 3 (b) と図 4 (a) , 4 (b) とを用いて、周波数及びパワーが大きい雑音を低減する減衰処理を説明したが、上記減衰処理と同様の方法で他の雑音（低周波成分を多く含む雑音、等）を低減することもできる。

【 0 0 4 5 】

図 5 は、本実施形態に係る雑音低減処理の流れの一例を示すフローチャートである。

以下、本実施形態に係る雑音低減処理の流れの一例について説明する。

【 0 0 4 6 】

まず、音声処理部 1 6 が、マイクロフォン 1 2 から入力音声信号を取得し、メモリ部 1 7 に記録する（ S 1 1 0 ）。例えば、図 3 (c) の入力音声信号が取得される。

次に、音声信号減衰部 3 1 が、 S 1 1 0 で取得された入力音声信号に減衰処理を施すことにより、減衰音声信号を生成する（ S 1 1 1 ）。例えば、図 3 (d) の減衰音声信号が生成される。

【 0 0 4 7 】

そして、参照期間設定部 3 4 が、 S 1 1 1 で生成された減衰音声信号に対して、参照期間を設定する（ S 1 1 2 ）。参照期間の情報は、類似期間検出部 3 2 と置換信号生成部 3 3 とに出力される。例えば、図 3 (d) の参照期間 5 1 が設定される。

次に、類似期間検出部 3 2 が、 S 1 1 1 で生成された減衰音声信号を用いて、複数の類似期間を検出する（ S 1 1 3 ）。具体的には、入力音声信号の代わりに減衰音声信号を用いて従来の処理と同様の処理を行うことにより、複数の類似期間が検出される。例えば、図 3 (d) の 3 つの類似期間 5 2 a , 5 2 b , 5 2 c が検出される。類似期間検出部 3 2 は、検出した各類似期間を表す類似期間信号を置換信号生成部 3 3 に出力する。例えば、図 3 (d) の時刻 t 1 , t 2 , t 3 を表す情報が、類似期間信号として出力される。

【 0 0 4 8 】

そして、置換信号生成部 3 3 が、 S 1 1 0 で取得された入力音声信号から、 S 1 1 2 で設定された参照期間における入力音声信号と、 S 1 1 3 で検出された複数の類似期間における入力音声信号と、を抽出する（ S 1 1 4 ）。例えば、図 3 (c) の上側に示すように、参照期間 5 1 における入力音声信号 4 1、類似期間 5 2 a における入力音声信号 4 2 a、類似期間 5 2 b における入力音声信号 4 2 b、及び、類似期間 5 2 c における入力音声信号 4 2 c、が抽出される。

【 0 0 4 9 】

次に、置換信号生成部 3 3 が、 S 1 1 4 で抽出された入力音声信号を用いて、置換信号を生成する（ S 1 1 5 ）。置換信号は、従来の処理と同様の処理により生成される。置換信号生成部 3 3 は、生成した置換信号を音声信号置換部 3 5 に出力する。

そして、音声信号置換部 3 5 が、 S 1 1 2 で設定された参照期間における入力音声信号を S 1 1 5 で生成された置換信号に置換することにより、出力音声信号を生成または更新する（ S 1 1 6 ）。1 回目の処理では、 S 1 1 0 で取得された入力音声信号の一部が S 1 1 5 で生成された置換信号に置換される。それにより、出力音声信号が生成される。2 回目以降の処理では、前回の S 1 1 6 で生成された出力音声信号の一部が S 1 1 5 で生成さ

10

20

30

40

50

れた置換信号に置換される。それにより、出力音声信号が更新される。

次に、置換信号生成部33が、S116で得られた出力音声信号を、メモリ部17に記録する(S117)。1回目の処理では、S116で得られた出力音声信号がメモリ部17に新規保存され、2回目以降の処理では、メモリ部17に記録されている出力音声信号がS116で得られた出力音声信号に更新される。

【0050】

そして、参照期間設定部34が、雑音を低減すべき期間であり、且つ、参照期間として設定されていない期間である、未処理期間が存在するかを判断する(S118)。未処理期間が存在する場合には、S112に処理が戻される。そして、S112において、未処理期間の少なくとも一部を含む参照期間が設定される。その後、S113～S118の処理が行われる。そして、未処理期間が存在しなくなるまで、S112～S118の処理が繰り返し行われる。未処理期間が存在しなくなると、本フローが終了される。

10

なお、複数の参照期間の設定方法は特に限定されない。複数の参照期間は、例えば、時間位置を少しずつずらしながら順番に設定される。参照期間の一部が隣接参照期間の一部に重畳されていてもよいし、参照期間が隣接参照期間から離れていてもよい。参照期間の終了時間位置と隣接参照期間の開始時間位置とが一致するように、複数の参照期間が設定されてもよい。隣接参照期間は、参照期間に隣接する参照期間である。

【0051】

S111では、類似期間の検出精度を低下させる成分が低減された減衰音声信号が得られる。減衰処理のフィルタ特性が図4(b)のフィルタ特性64である場合には、第1フォルマントの周波数帯域と第2フォルマントの周波数帯域とを含む注目帯域における音声信号(被写体音声信号及び雑音信号)を表す減衰音声信号が得られる。換言すれば、低周波数の側及び低周波数の側の音声信号(被写体音声信号及び雑音信号)を減衰させた減衰音声信号が得られる。そして、S113では、このような減衰音声信号を用いて複数の類似期間が検出される。それにより、複数の類似期間を高精度に検出することができる。具体的には、注目帯域における音声信号に着目して類似期間が検出されるため、高精度に類似期間を検出することができる。

20

【0052】

ここで、減衰音声信号では、図4(b)の減衰音声信号の周波数特性65に示すように、非注目帯域(注目帯域以外の周波数帯域)における雑音信号だけでなく、非注目帯域における被写体音声信号も減衰している。そのため、図3(a)の被写体音声信号のうち、低周波数の側及び低周波数の側の音声信号は、図3(d)の減衰音声信号には含まれていない。そのため、図3(d)の減衰音声信号(参照期間51における減衰音声信号、及び、3つの類似期間52a, 52b, 52cのそれぞれにおける減衰音声信号)を用いて置換信号を生成すると、被写体の音声が悪化した置換信号が生成されてしまう。具体的には、低周波数の側及び低周波数の側の被写体音声信号を含まない置換信号が生成されてしまう。その結果、被写体の音声が悪化した出力音声信号が生成されてしまう。

30

【0053】

本実施形態では、S115において、減衰されていない入力音声信号(全周波数帯域における音声信号)を用いて、置換信号が生成される。それにより、被写体の音声が悪化しておらず、且つ、雑音が高精度に低減された置換信号を生成することができる。その結果、雑音が高精度に低減された出力音声信号を生成することができる。

40

具体的には、ランダム性の高い雑音は、S115の処理(例えば、参照期間における入力音声信号と、各類似期間における入力音声信号と、を重みづけ合成する処理)によって低減することができる。例えば、風雑音はランダム性が非常に高いため、S115の処理によって低減することができる。そして、類似期間が高精度に検出されているため、S115の処理によって雑音を高精度に低減することができる。

また、繰り返し性の高い被写体音声信号は、S115の処理によって、低減されず、強調される。そして、入力音声信号では、全周波数帯域において音声信号が減衰されていないため、上述した被写体の音声の劣化を抑制することができる。

50

【0054】

以上述べたように、本実施形態によれば、減衰音声信号を用いて複数の類似期間が検出される。それにより、複数の類似期間を高精度に検出することができる。そして、本実施形態によれば、入力音声信号（参照期間における入力音声信号、及び、複数の類似期間のそれぞれにおける入力音声信号）を用いて置換信号が生成される。それにより、雑音が高精度に低減され、且つ、被写体の音声をよく表す置換信号及び出力音声信号を生成することができる。

【0055】

なお、注目帯域は、500Hz以上且つ3000Hz以下の周波数帯域に限らない。類似期間の検出精度に影響を与える雑音が小さい場合には、注目帯域が広いほど高精度に類似範囲を検出することができる。そのため、類似期間の検出精度に影響を与える雑音として想定される雑音（想定雑音）の周波数に基づいて、注目帯域を決定することが好ましい。例えば、想定雑音が撮像レンズ11の駆動に伴うレンズ駆動雑音であり、且つ、当該レンズ駆動雑音が8000Hzの近傍に強い成分を有する場合には、500Hz以上且つ7000Hz以下の周波数帯域が注目帯域として設定されてもよい。7000Hz以下の周波数帯域が注目帯域として設定されてもよい。想定雑音が風雑音である場合には、500Hz以上の周波数帯域が注目帯域として設定されてもよい。

【0056】

なお、注目帯域は固定値でなくてもよい。

例えば、音声処理装置や撮像装置が複数の動作モードを有しており、複数の動作モードにそれぞれ対応する複数の周波数帯域が予め定められていてもよい。そして、音声処理装置は、複数の周波数帯域の中から、設定されている動作モードに対応する周波数帯域を、注目帯域として選択する選択部を有していてもよい。

具体的には、複数の動作モードは、屋内での撮像時に設定すべき屋内撮像モード、屋外での撮像時に設定すべき屋外撮像モード、等を含む。そして、屋内撮像モードが設定されている場合には、風雑音が重畳されていないと判断され、3000Hz以下の周波数帯域が注目帯域として設定される。屋外撮像モードが設定されている場合には、風雑音が重畳されていると判断され、500Hz以上の周波数帯域が注目帯域として設定される。

【0057】

また、撮像装置が有する光学レンズの複数の駆動状態にそれぞれ対応する複数の周波数帯域が予め定められていてもよい。そして、音声処理装置は、複数の周波数帯域の中から、光学レンズの駆動状態に対応する周波数帯域を、注目帯域として選択する選択部を有していてもよい。

なお、入力音声信号にレンズ駆動雑音が重畳されている場合には、撮像装置が有する光学レンズの駆動期間を、参照期間として設定すればよい。具体的には、カメラシステム制御部10からの駆動命令に応じてレンズ駆動部15が撮像レンズ11を駆動している期間を、参照期間として設定すればよい。

【0058】

また、音声処理装置は、入力音声信号に基づいて注目帯域（または非注目帯域）を決定する決定部を有していてもよい。例えば、決定部は、入力音声信号における第1フォルマントの周波数を検出し、検出した周波数を含む周波数帯域を、注目帯域として決定する。

入力音声信号に基づく注目帯域の決定方法は特に限定されない。入力音声信号に基づく注目帯域は、例えば、入力音声信号を用いた周波数解析の結果に基づいて決定することができる。

【0059】

具体的には、図6に示すように、音声処理部16が、周波数解析部37と注目帯域決定部36とをさらに有していてもよい。図6は、音声処理部16の機能構成の一例を示すブロック図である。図6において、図1と同じ機能部には図1と同じ符号を付し、その説明は省略する。

周波数解析部37は、入力音声信号21を周波数解析することにより、入力音声信号2

10

20

30

40

50

1（入力音声信号21を含む被写体音声信号）における第1フォルマン트의周波数を検出する。例えば、周波数解析部37は、入力音声信号21をフーリエ変換し、フーリエ変換の結果に基づいて第1フォルマン트의周波数を検出する。

なお、入力音声信号21（入力音声信号21を含む被写体音声信号）における特徴的な他の周波数をさらに含む複数の周波数が検出されてもよい。

注目帯域決定部36は、周波数解析部37で検出された1つ以上の周波数（検出周波数）を含む周波数帯域を注目帯域として決定する。1つ以上の検出周波数は、第1フォルマン트의周波数を含む。

一般的には、成人が発する声の第1フォルマン트의周波数帯域は、500Hz以上且つ1500Hz以下の周波数帯域と言われている。図6の構成によれば、被写体音声信号の第1フォルマン트의周波数が500Hz以上且つ1500Hz以下の周波数帯域の外側の周波数である場合にも、適切な検出帯域を設定することができ、適切な雑音低減処理を行うことができる。

【0060】

なお、本実施形態では、音声処理装置としてカメラ1のような撮像装置を例示し、撮像装置が上述した雑音低減処理を実行する例を説明したが、これに限らない。撮像装置と異なる他の電子機器が上述した雑音低減処理を実行してもよい。

【0061】

<その他の実施形態>

記憶装置に記録されたプログラムを読み込み実行することで前述した実施形態の機能を実現するシステムや装置のコンピュータ（又はCPU、MPU等のデバイス）によっても、本発明を実施することができる。また、例えば、記憶装置に記録されたプログラムを読み込み実行することで前述した実施形態の機能を実現するシステムや装置のコンピュータによって実行されるステップからなる方法によっても、本発明を実施することができる。この目的のために、上記プログラムは、例えば、ネットワークを通じて、又は、上記記憶装置となり得る様々なタイプの記録媒体（つまり、非一時的にデータを保持するコンピュータ読取可能な記録媒体）から、上記コンピュータに提供される。したがって、上記コンピュータ（CPU、MPU等のデバイスを含む）、上記方法、上記プログラム（プログラムコード、プログラムプロダクトを含む）、上記プログラムを非一時的に保持するコンピュータ読取可能な記録媒体は、いずれも本発明の範疇に含まれる。

【符号の説明】

【0062】

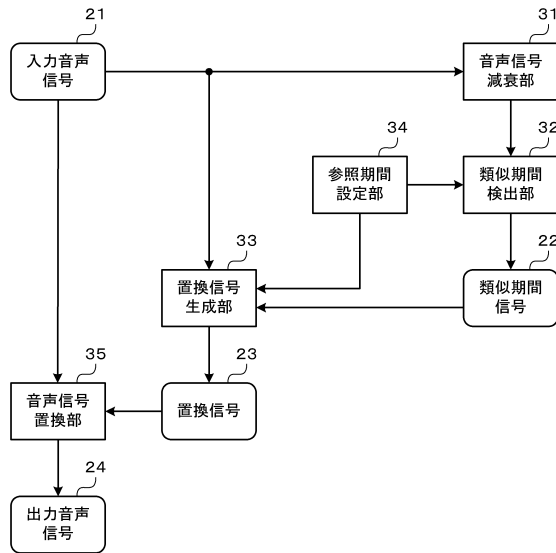
1：カメラ 11：撮像レンズ 15：レンズ駆動部 16：音声処理部
31：音声信号減衰部 32：類似期間検出部 33：置換信号生成部
34：参照期間設定部 35：音声信号置換部

10

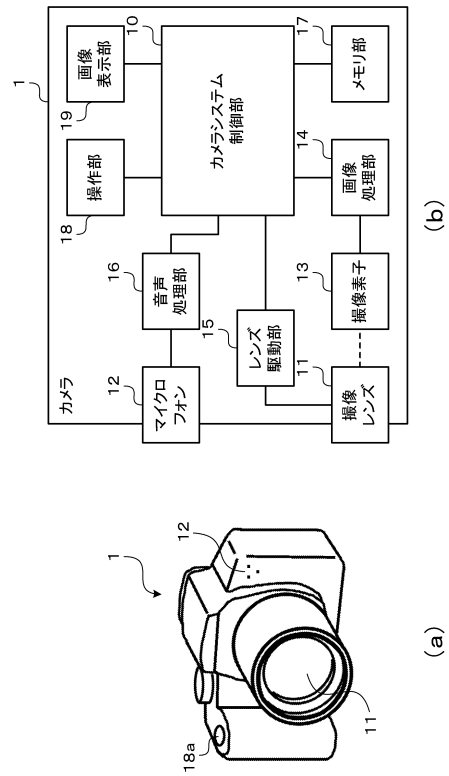
20

30

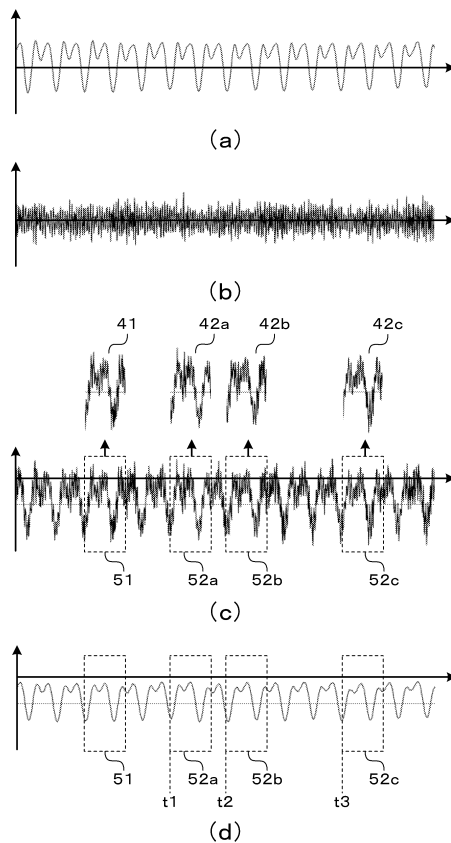
【図 1】



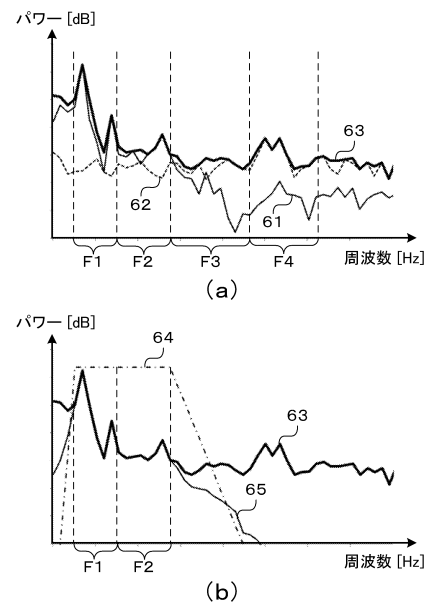
【図 2】



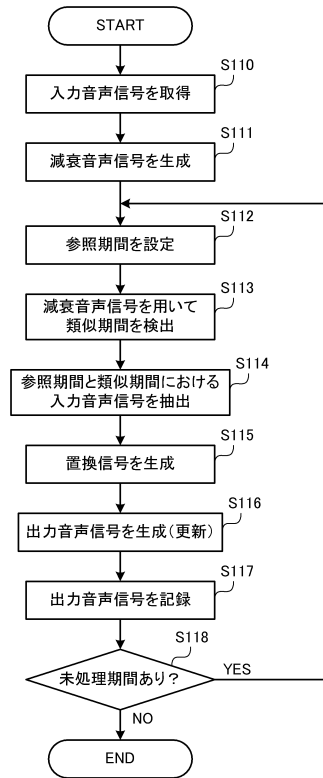
【図 3】



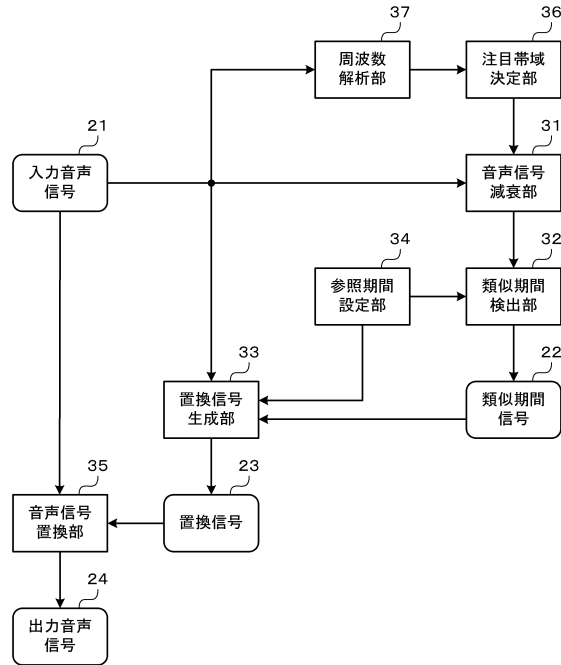
【図 4】



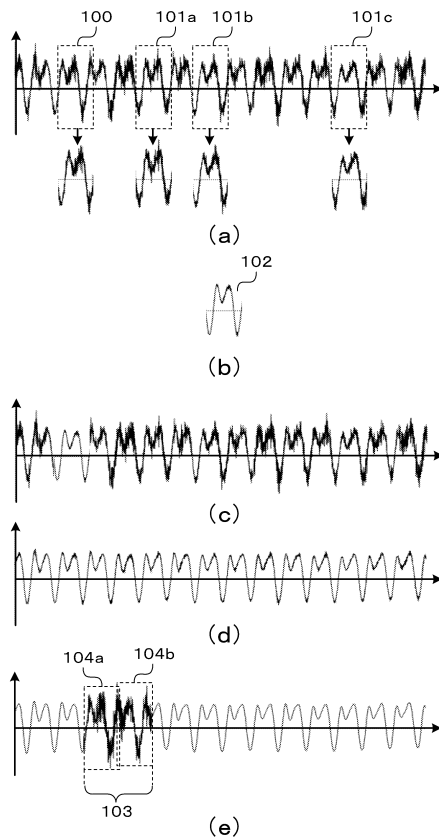
【図 5】



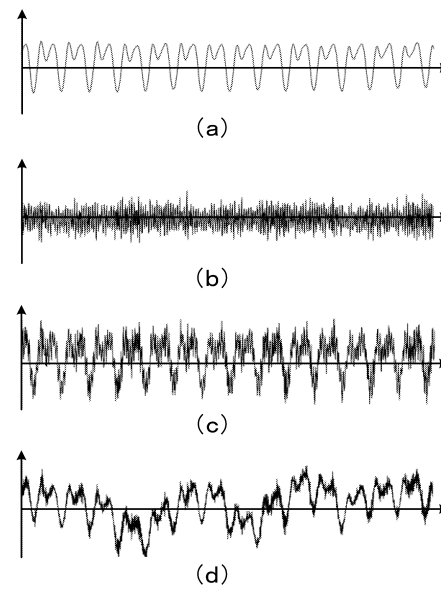
【図 6】



【図 7】



【図 8】



フロントページの続き

(72)発明者 梶村 文裕

東京都大田区下丸子3丁目30番2号 キヤノン株式会社 内

審査官 富澤 直樹

(56)参考文献 特開2003-150191(JP,A)

特開2013-117585(JP,A)

特開2010-117501(JP,A)

特開2011-095378(JP,A)

再公表特許第2011/129421(JP,A1)

(58)調査した分野(Int.Cl., DB名)

G10L 13/00 - 99/00

H04N 5/225

H04R 3/00