

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6381367号
(P6381367)

(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 15 (全 22 頁)

(21) 出願番号	特願2014-171660 (P2014-171660)	(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2016-45457 (P2016-45457A)		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24日 (2017. 8. 24)		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】

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、
複数の比較期間の中から、前記参照期間における音声信号と類似した音声信号の期間である複数の類似期間を検出する第1検出手段と、
 前記参照期間における音声信号に含まれる基本周波数を検出する第2検出手段と、
 前記第2検出手段で検出された前記基本周波数に基づいて、前記類似期間毎に、前記参照期間における前記基本周波数の音声信号と、前記類似期間における前記基本周波数の音声信号と、の間の位相差を取得する第2取得手段と、
 前記第2取得手段で取得された前記類似期間毎の位相差に基づいて、前記複数の類似期間の中から、2つ以上の類似期間を選択する選択手段と、
 前記参照期間における音声信号と、前記選択手段で選択された前記2つ以上の類似期間のそれぞれにおける音声信号と、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】

10

20

前記選択手段は、前記2つ以上の類似期間の位相差のばらつきが増すように、前記2つ以上の類似期間を選択する

ことを特徴とする請求項1に記載の音声処理装置。

【請求項3】

前記選択手段は、前記位相差が互いに等しい複数の類似期間のうち、前記参照期間における音声信号との音声信号の類似度が最も高い類似期間を、前記2つ以上の類似期間の1つとして優先的に選択する

ことを特徴とする請求項2に記載の音声処理装置。

【請求項4】

前記第2取得手段は、前記参照期間と前記類似期間の間の時間差を前記基本周波数の周期で除算することによって得られる整数値と余りのうち、前記余りを前記参照期間と前記類似期間の間の位相差として取得する

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】

前記生成手段は、前記参照期間における音声信号と前記2つ以上の類似期間のそれぞれにおける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1～4のいずれか1項に記載の音声処理装置。

【請求項6】

前記生成手段は、前記類似期間における音声信号の重みとして、前記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する

ことを特徴とする請求項5に記載の音声処理装置。

【請求項7】

前記第2検出手段は、前記参照期間における音声信号をフーリエ変換することにより、前記基本周波数を検出する

ことを特徴とする請求項1～6のいずれか1項に記載の音声処理装置。

【請求項8】

前記第2検出手段は、前記参照期間における音声信号をフーリエ変換することにより得られる複数の周波数のうち、大きさが所定値より大きい周波数を、前記基本周波数として検出する

ことを特徴とする請求項7に記載の音声処理装置。

【請求項9】

前記第2検出手段は、前記参照期間における音声信号をフーリエ変換することにより大きさが所定値より大きい周波数が複数得られた場合に、大きさが所定値より大きい複数の周波数のうち、最も低い周波数を、前記基本周波数として検出する

ことを特徴とする請求項7または8に記載の音声処理装置。

【請求項10】

前記第1検出手段は、前記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1～9のいずれか1項に記載の音声処理装置。

【請求項11】

前記第1取得手段は、第1音声信号を取得する第3取得手段と、前記第1音声信号に対して遅延した第2音声信号を取得する第4取得手段と、を有し、

前記音声処理装置は、前記第1音声信号に対する前記第2音声信号の遅延時間を検出する第3検出手段、をさらに有し、

前記生成手段は、前記参照期間における第1音声信号として設定する第1置換信号を生成する際に、前記2つ以上の類似期間を前記遅延時間だけ遅延させた2つ以上の期間のそれぞれにおける第2音声信号をさらに使用する

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】

前記第 1 設定手段は、第 1 参照期間と、前記第 1 参照期間を前記遅延時間だけ遅延させた第 2 参照期間と、を設定し、

前記第 1 置換信号は、前記第 1 参照期間における第 1 音声信号として設定する音声信号であり、

前記生成手段は、前記第 2 参照期間における第 2 音声信号として設定する音声信号である第 2 置換信号を生成する際に、前記 2 つ以上の類似期間として、前記第 1 置換信号を生成する際に用いた 2 つ以上の類似期間を前記遅延時間だけ遅延させた 2 つ以上の期間を使用する

ことを特徴とする請求項 1 1 に記載の音声処理装置。

【請求項 1 3】

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、

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 ～ 1 2 のいずれか 1 項に記載の音声処理装置。

【請求項 1 4】

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、

複数の比較期間の中から、前記参照期間における音声信号と類似した音声信号の期間である複数の類似期間を検出する第 1 検出ステップと、

前記参照期間における音声信号に含まれる基本周波数を検出する第 2 検出ステップと、

前記第 2 検出ステップで検出された前記基本周波数に基づいて、前記類似期間毎に、前記参照期間における前記基本周波数の音声信号と、前記類似期間における前記基本周波数の音声信号と、の間の位相差を取得する第 2 取得ステップと、

前記第 2 取得ステップで取得された前記類似期間毎の位相差に基づいて、前記複数の類似期間の中から、2 つ以上の類似期間を選択する選択ステップと、

前記参照期間における音声信号と、前記選択ステップで選択された前記 2 つ以上の類似期間のそれぞれにおける音声信号と、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 5】

コンピュータを請求項 1 ～ 1 3 のいずれか 1 項に記載の音声処理装置の各手段として機能させるためのプログラム。

【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、モータが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】

10

20

30

40

50

【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】

10

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、

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、

複数の比較期間の中から、前記参照期間における音声信号と類似した音声信号の期間である複数の類似期間を検出する第1検出手段と、

20

前記参照期間における音声信号に含まれる基本周波数を検出する第2検出手段と、

前記第2検出手段で検出された前記基本周波数に基づいて、前記類似期間毎に、前記参照期間における前記基本周波数の音声信号と、前記類似期間における前記基本周波数の音声信号と、の間の位相差を取得する第2取得手段と、

前記第2取得手段で取得された前記類似期間毎の位相差に基づいて、前記複数の類似期間の中から、2つ以上の類似期間を選択する選択手段と、

前記参照期間における音声信号と、前記選択手段で選択された前記2つ以上の類似期間のそれぞれにおける音声信号と、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る置換信号を生成する生成手段と、

30

前記参照期間における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、

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、

複数の比較期間の中から、前記参照期間における音声信号と類似した音声信号の期間である複数の類似期間を検出する第1検出ステップと、

40

前記参照期間における音声信号に含まれる基本周波数を検出する第2検出ステップと、

前記第2検出ステップで検出された前記基本周波数に基づいて、前記類似期間毎に、前記参照期間における前記基本周波数の音声信号と、前記類似期間における前記基本周波数の音声信号と、の間の位相差を取得する第2取得ステップと、

前記第2取得ステップで取得された前記類似期間毎の位相差に基づいて、前記複数の類似期間の中から、2つ以上の類似期間を選択する選択ステップと、

前記参照期間における音声信号と、前記選択ステップで選択された前記2つ以上の類似期間のそれぞれにおける音声信号と、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音声信号を前記置換信号に置換する置換ステップと、

50

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、コンピュータを上述した音声処理装置の各手段として機能させるためのプログラムである。

【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】実施例1に係る音声処理部の機能構成の一例を示すブロック図

10

【図2】実施例1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】従来の雑音低減処理の一例を示す図

【図4】従来の入力音声信号と出力音声信号の一例を示す図

【図5】従来の雑音低減処理の一例を示す図

【図6】従来の入力音声信号と出力音声信号の周波数特性の一例を示す図

【図7】実施例1に係る入力音声信号の一例を示す図

【図8】実施例2に係る撮像装置の外観及び機能構成の一例を示す図

【図9】実施例2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10】実施例2に係る入力音声信号の一例を示す図

【発明を実施するための形態】

20

【0011】

以下、図面を参照して本発明の実施例について詳細に説明する。

なお、以下の実施例はあくまで一例であり、本発明は以下の実施例に限定されない。

【0012】

<実施例1>

図1～7を参照して、本発明の実施例1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説明する。

【0013】

(撮像装置の構成)

本実施例に係る音声処理装置の一例として、以下カメラ1 (撮像装置) について説明する。

30

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、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。

【0014】

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。

40

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。

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

50

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。

【0015】

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。本実施例では、マイクロフォン12は、動画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。

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。

【0016】

操作部18は、カメラ1に対するユーザ操作を受け付ける。本実施例では、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。

【0017】

（音声処理部16の構成）

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。

【0018】

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。

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、参照期間設定部31、類似期間検出部32、置換信号生成部33、基本周波数検出部34、位相差検出部35、音声信号置換部36、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。

【0019】

参照期間設定部31は、入力音声信号21に対し、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。本実施例では、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31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。

【0020】

類似期間検出部 3 2 は、入力音声信号 2 1 を取得する（第 1 取得処理）。

また、類似期間検出部 3 2 は、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。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、参照期間における入力音声信号と類似した入力音声信号の期間である複数の類似期間を検出する（第 1 検出処理）。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。

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。本実施例では、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、参照期間における入力音声信号と類似期間における入力音声信号との間の類似度である。

【0021】

なお、第 1 取得処理と第 2 設定処理の少なくとも一方は、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。第 1 取得処理と第 2 設定処理は、互いに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るものであってもよい。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。

【0022】

基本周波数検出部 3 4 は、参照期間における入力音声信号に含まれる基本周波数を検出する（第 2 検出処理）。本実施例では、人間が発する声の第 1 フォルマントに相当する周波数が、基本周波数として検出される。一般的に、成人が発する声の第 1 フォルマントの周波数は、500 Hz 以上且つ 1000 Hz 以下の周波数と言われている。

【0023】

位相差検出部 3 5 は、基本周波数検出部 3 4 によって検出された基本周波数に基づいて、類似期間毎に信号位相差を検出（取得）する（第 2 取得処理）。信号位相差は、参照期間における基本周波数の音声信号と、類似期間における基本周波数の音声信号と、の間の位相差である。

【0024】

置換信号生成部 3 3 は、各類似期間の信号位相差に基づいて、複数の類似期間の中から 2 つ以上の類似期間を選択する（選択処理）。

そして、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、選択した 2 つ以上の類似期間（選択類似期間）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（生成処理）。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。

なお、選択処理は、置換信号生成部 3 3 とは異なる機能部によって実行されてもよい。

【0025】

音声信号置換部 3 6 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。

【0026】

（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。

図 3（a）～3（e）は、従来の雑音低減処理の一例を示す模式図である。図 3（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 3（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

10

20

30

40

50

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 (b) は、置換信号の一例を示す。図 3 (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 3 (d) は、出力音声信号の他の例を示す。図 3 (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 3 (a) ~ 3 (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 3 (a) の上側、3 (c)、3 (d)、3 (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 3 (a) の上側、3 (c)、3 (d)、3 (e) に示す音声信号は、0.2 秒程度の音声信号である。図 3 (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、本実施例でも着目される。

10

【0027】

まず、図 3 (a) に示すように、入力音声信号に対して、参照期間 100 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 500 Hz 以上且つ 1000 Hz 以下の周波数と言われているため、参照期間の長さは、2 msec (= 0.002 sec = 1 ÷ 500 Hz) 以上であることが好ましい。また、参照期間の長さとして、入力音声信号の基本周波数の数倍程度までの長さが設定されていることが好ましい。

20

【0028】

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。

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。

30

【0029】

そして、参照期間 100 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 3 (a) の例では、3 つの類似期間 101 a、101 b、101 c が検出されている。

【0030】

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。

【0031】

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。

40

【数 1】

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

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

50

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。

【0032】

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。

【0033】

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。

【0034】

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】

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

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。 $S_O(i)$ 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3(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3(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3(c)の出力音声信号が生成される。図3(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3(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3(d)の出力音声信号を生成することができる。図3(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

10

20

30

40

50

ば、図3(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3(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4(a)、4(b)、5、6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4(a)は、被写体の音声に暗雑音が重畳された入力音声信号の一例を示す図である。図4(b)は、従来の雑音低減処理を図4(a)の入力音声信号に施すことによって生成された出力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。

【0039】

図4(a)の例では、参照期間110が設定されており、類似期間111-1、111-2、111-3、111-pを含む複数の類似期間が検出されている。 $T_r(K)$ は、番号Kの類似期間と参照期間110との時間差である。図4(a)では、時間的に最も早い類似期間の番号Kが1となり、時間が進むにつれて番号Kが1ずつ増加するように各類似期間の番号Kが定められている。類似期間111-1は、時間的に最も早い類似期間である。そのため、類似期間111-1と参照期間110の時間差は、 $T_r(1)$ である。類似期間111-2は、2番目に早い類似期間である。そのため、類似期間111-2と参照期間110の時間差は、 $T_r(2)$ である。類似期間111-3は3番目に早い類似期間である。そのため、類似期間111-3と参照期間110の時間差は、 $T_r(3)$ である。類似期間111-pは、p番目に早い類似期間である。そのため、類似期間111-pと参照期間110の時間差は $T_r(p)$ である。

【0040】

参照期間110における入力音声信号の基本周波数は F_b [Hz]であり、基本周波数 F_b [Hz]の周期 T_b [sec]は基本周波数 F_b [Hz]の逆数である。

上述したように、音声信号を局所的に観察すると、音声信号の繰り返し性は非常に高い。そのため、参照期間110に時間的に近い複数の類似期間は、周期 T_b の時間間隔で存在している可能性が高い。つまり、時間差 $T_r(K)$ は、周期 T_b の整数倍である可能性が高い。そのため、時間差 $T_r(K)$ が周期 T_b の整数倍であるN個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を生成する処理(式2の演算)は、図5に示す信号処理と同等である。

【0041】

図5は、置換信号を生成する信号処理を実現する機能部(置換信号生成部)の機能構成の一例を示すブロック図である。

遅延器121は、入力音声信号 $x(n)$ を遅延量 $D_s(K)$ の離散時間だけ遅延させて出力する。入力音声信号 $x(n)$ は、時間位置nにおける入力音声信号の信号レベルであり、参照期間における入力音声信号の信号レベルに相当する。遅延量 $D_s(K)$ は、番号Kの類似期間に対応する遅延量である。

ゲイン器122は、遅延器121から出力された音声信号にゲイン値 $\alpha(K)$ を乗算し、乗算結果の音声信号を出力する。ゲイン値 $\alpha(K)$ は、番号Kの類似期間に対応するゲイン値であり、上述した重み w_K である。ここでは、説明を簡単にするために、全てのゲイン値 $\alpha(K)$ が1であるものとする。

加算器123は、入力音声信号 $x(n)$ と、N個のゲイン器122から出力されたN個の音声信号と、の総和を算出し、算出結果である音声信号(総和)を出力する。

ゲイン器124は、加算器123から出力された音声信号にゲイン値 β を乗算することにより、出力音声信号 $y(n)$ を算出する。ゲイン値 β は $1/N$ である。出力音声信号 y

10

20

30

40

50

(n) は、時間位置 n における入力音声信号の信号レベルとして設定すべき音声信号レベルである。

【0042】

遅延量 $D_s(K)$ は、以下の式 3 で算出された値である。

$$D_s(K) = K \times T_b \quad \dots (式 3)$$

そのため、ゲイン器 122 からは、入力音声信号 $x(n + K \times T_b)$ が出力される。入力音声信号 $x(n + K \times T_b)$ は、時間位置 “ $n + K \times T_b$ ” における入力音声信号の信号レベルであり、K 番目の類似期間における入力音声信号の信号レベルに相当する。そして、入力音声信号 $x(n)$ と N 個の入力音声信号 $x(n + K \times T_b)$ の平均値が、出力音声信号 $y(n)$ として算出される。入力音声信号 $x(n)$ と出力音声信号 $y(n)$ のこのような関係（伝達特性）は、いわゆるコンフィルタのフィルタ特性と同様であり、以下の式 4 で表すことができる。N 個の入力音声信号 $x(n + K \times T_b)$ は、“時間位置 n から周期 T_b の整数倍だけずらした N 個の時間位置における N 個の入力信号レベル” と言うこともできる。

【数 3】

$$y(n) = \frac{1}{N+1} \left\{ x(n) + \sum_{K=1}^N x(n + K \times T_b) \right\} \quad \dots (式 4)$$

【0043】

図 4 (a)、4 (b) に示す音声信号の周波数特性（周波数とパワーとの対応関係）を、図 6 に示す。図 6 において、横軸は周波数 [Hz] を示し、縦軸はパワー [dB] を示す。周波数特性 131 は図 4 (a) の入力音声信号の周波数特性であり、周波数特性 132 は図 4 (b) の出力音声信号の周波数特性である。

図 4 (a) の入力音声信号には、周波数 F_b [Hz] のサイン波と、白色雑音と、が含まれている。周波数 F_b [Hz] のサイン波は、被写体の音性に相当する。そして、周波数特性 131 において、周波数 F_b [Hz] のパワーは約 G_2 [dB] であり、周波数 F_b [Hz] 以外の周波数帯域のパワーは約 G_1 [dB] である。

図 6 から、図 5 に示す伝達特性で図 4 (a) の入力音声信号を図 4 (b) の出力音声信号に変換することにより、白色雑音のパワーが約 G_1 [dB] から約 G_3 [dB] へ低下することがわかる。しかしながら、周波数特性 132 において、周波数 F_b [Hz] の整数倍の周波数のパワーは、周波数特性 131 のパワーから変化していない。つまり、図 4 (b) の出力音声信号は、雑音が低減されているように見えるが、周波数 F_b の整数倍の周波数の雑音を低減できていない。遅延量 $D_s(K)$ を時間差 $T_r(K)$ に置き換えても、同様のことが言える。

【0044】

（本実施例に係る雑音低減処理）

そこで、本実施例では、参照期間と各類似期間との時間差に注目し、周波数 F_b の整数倍の周波数の雑音も低減されるように、置換信号の生成に用いる類似期間を選択する。

本実施例に係る雑音低減処理の一例について説明する。

図 7 は、被写体の音声に暗雑音が重畳された入力音声信号の一例を示す図である。図 7 の入力音声信号は、図 4 (a) の入力音声信号と等しい。符号 40 は参照期間を示し、符号 41-1、41-2、41-q、41-r は類似期間を示す。類似期間 41-1 は、参照期間 40 から基本周波数の 1 周期だけずれた時間位置の類似期間であり、類似期間 41-2 は、参照期間 40 から基本周波数の 2 周期だけずれた時間位置の類似期間である。類似期間 41-q、41-r は、参照期間 40 からある程度離れた時間位置の類似期間である。図 7 でも、図 4 (a) と同様に、時間的に最も早い類似期間の番号 K が 1 となり、時間が進むにつれて番号 K が 1 ずつ増加するように各類似期間の番号 K が定められている。類似期間 41-1 は、時間的に最も早い類似期間である。そのため、類似期間 41-1 の番号 K は、1 である。類似期間 41-2 は、2 番目に早い類似期間である。そのため、類

10

20

30

40

50

似期間 4 1 - 2 の番号 K は、2 である。類似期間 4 1 - q は、q 番目に早い類似期間である。そのため、類似期間 4 1 - q の番号 K は、q である。類似期間 4 1 - r は、r 番目に早い類似期間である。そのため、類似期間 4 1 - r の番号 K は、r である。

【0045】

本実施例では、基本周波数検出部 3 4 により、参照期間における入力音声信号に含まれる基本周波数が検出される。基本周波数検出部 3 4 では、例えば、参照期間における入力音声信号をフーリエ変換することにより、基本周波数が検出される。

なお、参照期間における入力音声信号をフーリエ変換することにより周波数が複数得られた場合には、当該複数の周波数のうち、大きさ（パワー）が所定値より大きい周波数を、基本周波数として検出すればよい。

10

そして、参照期間における入力音声信号をフーリエ変換することにより大きさが所定値より大きい周波数が複数得られた場合には、当該複数の周波数のうち、最も低い周波数を、基本周波数として検出すればよい。

なお、基本周波数の検出方法は特に限定されない。基本周波数を検出する処理として、従来技術を適宜利用することができる。

【0046】

そして、本実施例では、位相差検出部 3 5 により、類似期間毎に、参照期間における基本周波数の音声信号と、類似期間における基本周波数の音声信号と、の間の位相差（信号位相差）が検出される。信号位相差 $P_b(K)$ [sec] は、例えば、式 5 を用いて算出することができる。信号位相差 $P_b(K)$ [sec] は、K 番目の類似期間に対応する信号位相差である。

20

$$P_b(K) = \text{mod}[T_r(K), T_b] \cdots (\text{式} 5)$$

$\text{mod}[A, B]$ は、A を B で除算することによって得られる整数値と余りのうち、余りを返す関数である。そのため、式 5 を用いた場合には、時間差 $T_r(K)$ を基本周波数の周期 T_b で除算することによって得られる整数値と余りのうち、余りが、信号位相差 $P_b(K)$ として算出（取得）される。

なお、信号位相差の検出方法は、上記方法に限らない。

【0047】

上述したように、時間差 $T_r(K)$ が T_b の整数倍である数多くの類似期間を使用して置換信号を生成すると、基本周波数の整数倍の周波数の雑音を高精度に低減できない置換信号が生成されてしまう。具体的には、信号位相差 $P_b(K)$ が略等しい数多くの類似期間を使用して置換信号を生成すると、基本周波数の整数倍の周波数の雑音を高精度に低減できない置換信号が生成されてしまう。

30

【0048】

そこで、本実施例では、置換信号生成部 3 3 が、各類似期間の信号位相差に基づいて、複数の類似期間の中から 2 つ以上の類似期間のそれぞれを、置換信号の生成に用いる類似期間（選択類似期間）として選択する。具体的には、置換信号生成部 3 3 は、複数の選択類似期間の信号位相差のばらつきが増すように、複数の選択類似期間を選択する。

選択類似期間が選択された後、従来の雑音低減処理と同様の処理によって、置換信号が生成され、出力音声信号が生成される。

40

複数の選択類似期間の信号位相差のばらつきが増すように、複数の選択類似期間を選択することにより、従来よりも高精度に雑音を低減することができる置換信号を生成することができる。具体的には、基本周波数の整数倍の周波数の雑音をも十分に低減することができる。

【0049】

以下、選択類似期間を選択する選択処理の具体例について説明する。

以下では、“類似期間 4 1 - 1 の信号類似度” > “類似期間 4 1 - 2 の信号類似度” > “類似期間 4 1 - q の信号類似度” > “類似期間 4 1 - r の信号類似度” であるものとする。類似期間 4 1 - 1 の信号類似度は、参照期間 4 0 における入力音声信号と類似期間 4 1 - 1 における入力音声信号との間の類似度である。類似期間 4 1 - 2 の信号類似度は、

50

参照期間 4 0 における入力音声信号と類似期間 4 1 - 2 における入力音声信号との間の類似度である。類似期間 4 1 - q の信号類似度は、参照期間 4 0 における入力音声信号と類似期間 4 1 - q における入力音声信号との間の類似度である。類似期間 4 1 - r の信号類似度は、参照期間 4 0 における入力音声信号と類似期間 4 1 - r における入力音声信号との間の類似度である。

また、以下では、類似期間 4 1 - 1 の信号位相差 $P_b(1)$ が 0 であり、類似期間 4 1 - 2 の信号位相差 $P_b(2)$ が 0 であるものとする。そして、類似期間 4 1 - q の信号位相差 $P_b(q)$ が $0.5 T_b$ であり、類似期間 4 1 - r の信号位相差 $P_b(r)$ が $0.1 T_b$ であるものとする。

【0050】

10

本実施例では、置換信号生成部 3 3 は、信号類似度が高い類似期間から順番に N 個の類似期間のそれぞれを選択類似期間として選択する。但し、置換信号生成部 3 3 は、信号位相差が選択類似期間の信号位相差と等しい（略等しい）類似期間を、選択類似期間として選択しない。

【0051】

そのため、最初に、信号類似度が最も高い類似期間 4 1 - 1 が、選択類似期間として選択される。

信号類似度が 2 番目に高い類似期間 4 1 - 1 は、信号位相差 $P_b(2)$ が類似期間 4 1 - 1 の信号位相差 $P_b(1)$ と等しいため、選択類似期間として選択されない。

信号類似度が 3 番目に高い類似期間 4 1 - q の信号位相差 $P_b(q)$ は、類似期間 4 1 - 1 の信号位相差 $P_b(1)$ と大きく異なる。そのため、類似期間 4 1 - 1 の次に、類似期間 4 1 - q が選択類似期間として選択される。

20

信号類似度が 4 番目に高い類似期間 4 1 - r の信号位相差 $P_b(q)$ は、類似期間 4 1 - 1 の信号位相差 $P_b(1)$ 及び類似期間 4 1 - q の信号位相差 $P_b(q)$ と大きく異なる。そのため、類似期間 4 1 - q の次に、類似期間 4 1 - r が選択類似期間として選択される。

【0052】

なお、選択処理の方法は上記方法に限らない。信号位相差を考慮して選択類似期間が選択されれば、どのような方法で選択類似期間が選択されてもよい。例えば、信号位相差が等しい（略等しい）類似期間が複数の選択類似期間にわずかに含まれていてもよい。但し、複数の選択類似期間の信号位相差のばらつきが大きい方が、ばらつきが小さい場合に比べ、基本周波数の整数倍の周波数の雑音がより高精度に低減された置換信号を生成することができる。そのため、複数の選択類似期間の信号位相差のばらつきが大きいことが好ましい。また、信号類似度が高い類似期間を使用した方が、信号類似度が低い類似期間を使用する場合に比べて、雑音より高精度に低減された置換信号を生成することができる。そのため、信号位相差が互いに等しい複数の類似期間が存在する場合には、当該複数の類似期間のうち、信号類似度が最も高い類似期間が、選択類似期間として優先的に選択されることが好ましい。

30

【0053】

以上述べたように、本実施例によれば、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。具体的には、本実施例では、循環バッファ等を必要としないため、雑音を低減できる期間の長さが物理的な制限によって制限されることがない。また、本実施例では、各類似期間の信号位相差を考慮して、複数の類似期間の中から複数の選択類似期間（置換信号の生成に使用する類似期間）が選択される。それにより、信号位相差が略等しい類似期間を数多く使用して置換信号が生成されることを抑制することができ、基本周波数の整数倍の周波数の雑音をも十分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得ることができる。

40

【0054】

なお、本実施例では、雑音が暗雑音である場合の例について詳しく説明したが、上述した処理を行うことにより、他の雑音も低減することができる。例えば、上述したレンズ駆

50

動雑音も低減することができる。低減対象の雑音がレンズ駆動雑音である場合には、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。

【0055】

なお、本実施例では、撮像装置に設けられている音声処理部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。他の電子機器に設けられている音声処理部が、上述した雑音低減処理を実行してもよい。撮像装置などの電子機器とは独立した音声処理装置が、上述した雑音低減処理を実行してもよい。撮像装置では、動画撮像時に入力音声信号が記憶部に記録されてもよい。そして、動画撮像後に、撮像装置以外の電子機器が上述した雑音低減処理を実行することにより、入力音声信号から出力音声信号が生成されてもよい。

10

【0056】

<実施例2>

図8(a)、8(b)、9、10を参照して、本発明の実施例3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説明する。

実施例1では、1つのマイクロフォン12で生成された1つの入力音声信号（モノラル音声信号）に雑音低減処理を施す例を説明した。

本実施例では、2つのマイクロフォンで生成された2つの入力音声信号（ステレオ音声信号；マルチチャンネル音声信号）に雑音低減処理を施す例を説明する。

20

なお、本実施例で説明する雑音低減処理は、3つ以上のマイクロフォンで生成された3つ以上の入力音声信号（マルチチャンネル音声信号）に施す雑音低減処理として実行することもできる。

【0057】

（撮像装置の構成）

図8(a)は、本実施例に係る音声処理装置の一例であるカメラ81の外観の一例を示す斜視図である。図8(b)は、カメラ81の機能構成の一例を示すブロック図である。

なお、図8(a)、8(b)において実施例1（図2(a)、2(b)）と同じ機能部には実施例1と同じ符号を付し、その説明は省略する。

図8(a)、8(b)に示すように、本実施例では、カメラ81に2つのマイクロフォン（左側のLchマイクロフォン82aと右側のRchマイクロフォン82b）が設けられている。

30

Lchマイクロフォン82aとRchマイクロフォン82bは、実施例1のマイクロフォン12と同様の機能を有する。以後、Lchマイクロフォン82aで生成された音声信号を“Lch入力音声信号”と記載し、Rchマイクロフォン82bで生成された音声信号を“Rch入力音声信号”と記載する。

音声処理部83は、Lchマイクロフォン82aからLch入力音声信号を取得し、Rchマイクロフォン82bからRch入力音声信号を取得する。そして、音声処理部83は、Lch入力音声信号とRchマイクロフォン82bのそれぞれに種々の処理を施すことにより、2つの出力音声信号を生成する。以後、Lch入力音声信号に種々の処理を施すことによって生成された出力音声信号を“Lch出力音声信号”と記載し、Rch入力音声信号に種々の処理を施すことによって生成された出力音声信号を“Rch出力音声信号”と記載する。音声処理部83が行う処理は、雑音低減処理を含む。

40

【0058】

（音声処理部83の構成）

図9は、音声処理部83の機能構成の一例を示すブロック図である。

なお、図9において実施例1（図1）と同じ機能部には実施例1と同じ符号を付し、その説明は省略する。

【0059】

本実施例では、Rch入力音声信号（第2音声信号）が、Lch入力音声信号（第1音

50

声信号)が表す音声に対して遅延した音声を表す場合の例を説明する。

遅延検出部91は、Lch入力音声信号とRch入力音声信号のそれぞれを入力音声信号21として取得し、Lch入力音声に対するRch入力音声の遅延時間を検出する(第3検出処理)。Lch入力音声は、Lch入力音声信号が表す音声であり、Rch入力音声は、Rch入力音声信号が表す音声である。

なお、遅延時間の検出方法は特に限定されない。例えば、Lch入力音声信号とRch入力音声信号の間の相互相関値に基づいて、遅延時間を検出することができる(従来技術であるため、詳細な説明は省略する)。

【0060】

参照期間設定部92は、Lch入力音声信号に対して第1参照期間を設定し、Rch入力音声信号に対して第2参照期間を設定する。第2参照期間は、第1参照期間と、遅延検出部91で検出された遅延時間と、に基づいて設定される。具体的には、第1参照期間を遅延時間だけ遅延させた期間が、第2参照期間として設定される。

【0061】

類似期間検出部93は、Lch入力音声信号を取得する処理(第3取得処理)と、Rch入力音声信号を取得する処理(第4取得処理)と、を行う。

なお、第3取得処理、第4設定処理、及び、実施例1で述べた第2設定処理のうちの少なくとも1つは、類似期間検出部32とは異なる機能部によって実行されてもよい。第3取得処理、第4設定処理、及び、第2設定処理は、互いに異なる機能部によって実行されてもよい。

【0062】

そして、類似期間検出部93は、Lch入力音声信号とRch入力音声信号のそれぞれに対して複数の類似期間を設定する。Lch入力音声信号に対しては、実施例1の類似期間検出部32と同様の処理を行うことによって、複数の第1類似期間が設定(検出)される。第1類似期間は、第1参照期間におけるLch入力音声信号と類似したLch入力音声信号の期間である。Rch入力音声信号に対する複数の類似期間は、Lch入力音声信号の複数の類似期間と、遅延検出部91で検出された遅延時間と、に基づいて設定される。具体的には、複数の第1類似期間を遅延時間だけ遅延させた複数の期間が、Rch入力音声信号の複数の類似期間(第2類似期間)として設定される。

【0063】

本実施例では、第2類似期間は、Lch入力音声信号に対する類似期間としても使用される。

そのため、本実施例では、第1置換信号を生成する際に、2つ以上の第1類似期間だけでなく、当該2つ以上の第1類似期間を遅延時間だけ遅延させた2つ以上の第2類似期間も使用される。具体的には、2つ以上の第2類似期間のそれぞれにおけるRch入力音声信号がさらに使用される。第1置換信号は、第1参照期間におけるLch入力音声信号として設定すべき置換信号である。

そして、本実施例では、第2置換信号を生成する際に、2つ以上の第2類似期間として、第1置換信号を生成する際に用いた2つ以上の第1類似期間を遅延時間だけ遅延させた2つ以上の第2類似期間が使用される。第2置換信号は、第2参照期間におけるRch入力音声信号として設定すべき音声信号である。

なお、第1類似期間は、Rch入力音声信号に対する類似期間としてさらに使用されてもよい。即ち、第1置換信号を生成する際に、2つ以上の第2類似期間を遅延時間だけ遅延させた2つ以上の第1類似期間のそれぞれにおけるLch入力音声信号がさらに使用されてもよい。

なお、類似期間検出部93で複数の第1類似期間のみが設定されてもよい。その場合には、置換信号生成部33で、2つ以上の第1類似期間を遅延時間だけ遅延させることによって、2つ以上の第2類似期間が設定されてもよい。

なお、第1類似期間の信号類似度は実施例1と同様の処理により決定されるが、第2類似期間の信号類似度は、どのように決定されてもよい。例えば、実施例1と同様の処理に

10

20

30

40

50

よって第2類似期間の信号類似度が決定されてもよい。また、第2類似期間の元となった第1類似期間の信号類似度が、第2類似期間の信号類似度として設定されてもよい。

【0064】

(雑音低減処理)

本実施例に係る雑音低減処理の具体例について説明する。

なお、実施例1と同じ処理については説明を省略する。

図10は、被写体の音声に暗雑音が重畳された入力音声信号の一例を示す図である。図10の上側にはLch入力音声信号が示されており、図10の下側にはRch入力音声信号が示されている。暗雑音は、Lch入力音声信号とRch入力音声信号の両方に重畳されている。図10において、横軸は時間を示し、縦軸は信号レベルを示す。

10

【0065】

音声を発する被写体（音発生源）がカメラ1の正面からずれている場合、Lchマイクロフォン82aから音発生源までの距離は、Rchマイクロフォン82bから音発生源までの距離と異なる。そのため、Lch入力音声信号とRch入力音声信号の一方が、Lch入力音声信号とRch入力音声信号の他方が表す音声に対して遅延した音声を表すことがある。図10では、Rch入力音声信号がLch入力音声信号に対して遅延時間Dtだけ遅延している。

【0066】

遅延時間Dtが0である場合には、Lch入力音声信号とRch入力音声信号の一方は、Lch入力音声信号とRch入力音声信号の他方と同等みなすことができる。そのため、Lch入力音声信号に対する第1置換信号を生成する際にRch入力音声信号を使用したり、Rch入力音声信号に対する第2置換信号を生成する際にLch入力音声信号を使用したりすることができる。また、置換信号を生成する際にLch入力音声信号とRch入力音声信号の両方を使用すれば、使用される類似期間の数が増すため、雑音がより低減された置換信号を生成することができる。暗雑音のようなランダム性の高い雑音に関しては、Lch入力音声信号とRch入力音声信号との間の雑音の相関関係は無い。そのため、ランダム性の高い雑音が重畳されている場合には、Lch入力音声信号とRch入力音声信号の両方を使用して置換信号を生成することは、極めて有効である。

20

【0067】

図10において、符号50は、Lch入力音声信号の第1参照期間を示し、符号52は、Rch入力音声信号の第2参照期間を示す。第2参照期間52は、第1参照期間50におけるLch入力音声信号に対応するRch入力音声信号の期間であり、第1参照期間50を遅延時間Dtだけ遅延させた期間である。

30

このように、本実施例では、第1参照期間を遅延時間Dtだけ遅延させた期間が、第2参照期間として設定される。

【0068】

符号51-1、51-qは、Lch入力音声信号の第1類似期間を示す。第1類似期間51-1は、第1参照期間50から基本周波数の1周期だけずれた時間位置の第1類似期間であり、第1類似期間51-qは、第1参照期間50からある程度離れた時間位置の第1類似期間である。

40

符号51-1'、51-q'は、Rch入力音声信号の第2類似期間を示す。第2類似期間51-1'は、第2参照期間52から基本周波数の1周期だけずれた時間位置の第2類似期間であり、第2類似期間51-q'は、第2参照期間52からある程度離れた時間位置の第2類似期間である。第2類似期間51-1'は、第1類似期間51-1を遅延時間Dtだけ遅延させた期間であり、第2類似期間51-q'は、第1類似期間51-qを遅延時間Dtだけ遅延させた期間である。

このように、本実施例では、第1類似期間を遅延時間Dtだけ遅延させた期間が、第2類似期間として設定される。

なお、基本周波数として、第1類似期間におけるLch入力音声信号の基本周波数と、第2類似期間におけるRch入力音声信号の基本周波数と、の両方が検出されてもよいし

50

、いずれか一方が検出されてもよい。

【0069】

図10において、第1参照期間50と第1類似期間51-1の間の時間差は $T_r(1)$ であり、第1参照期間50と第1類似期間51-qの間の時間差は $T_r(q)$ である。上述したように、遅延時間 D_t が0である場合には、Lch入力音声信号とRch入力音声信号の一方は、Lch入力音声信号とRch入力音声信号の他方と同等みなすことができる。そのため、第2参照期間52と第2類似期間51-1'の間の時間差は、第1参照期間50と第1類似期間51-1の間の時間差と同じ時間差 $T_r(1)$ となる。そして、第2参照期間52と第2類似期間51-q'の間の時間差は、第1参照期間50と第1類似期間51-qの間の時間差と同じ時間差 $T_r(q)$ となる。従って、第1参照期間50と第2類似期間51-1'の間の時間差 $T_r(1')$ は、式6-1の関係を満たし、第1参照期間50と第2類似期間51-q'の間の時間差 $T_r(q')$ は、式6-2の関係を満たす。

$$T_r(1') = T_r(1) + D_t \quad \cdots \text{(式6-1)}$$

$$T_r(q') = T_r(q) + D_t \quad \cdots \text{(式6-2)}$$

【0070】

そして、図10に示すように、第1参照期間50におけるLch入力音声信号に含まれる基本周波数と、第1類似期間51-1におけるLch入力音声信号に含まれる基本周波数と、の間の信号位相差 $P_b(1)$ は0となる。そして、第1参照期間50におけるLch入力音声信号に含まれる基本周波数と、第2類似期間51-1'におけるRch入力音声信号に含まれる基本周波数と、の間の信号位相差 $P_b(1')$ は D_t (=遅延時間)となる。そのため、信号位相差 $P_b(1)$ と信号位相差 $P_b(1')$ は D_t となる。

同様に、信号位相差 $P_b(q)$ と信号位相差 $P_b(q')$ の差も D_t となる。信号位相差 $P_b(q)$ は、第1参照期間50におけるLch入力音声信号に含まれる基本周波数と、第1類似期間51-qにおけるLch入力音声信号に含まれる基本周波数と、の間の信号位相差である。信号位相差 $P_b(q')$ は、第1参照期間50におけるLch入力音声信号に含まれる基本周波数と、第2類似期間51-q'におけるRch入力音声信号に含まれる基本周波数と、の間の信号位相差である。

このように、遅延時間だけずれた第1類似期間と第2類似期間の間の信号位相差の差は、置換時間 D_t と一致する。

【0071】

そのため、遅延時間 D_t が閾値よりも大きければ、第1置換信号を生成する際に、第1類似期間だけでなく、当該第1類似期間を遅延時間 D_t だけ遅延させた第2類似期間も選択されることとなる。例えば、第1類似期間51-qが選択された場合に、第2類似期間51-q'も選択されることとなる。これにより、第1置換信号を生成する際に、信号位相差のばらつきの低下を抑制して、より多くの類似期間(第1参照期間におけるLch入力音声信号との類似度が高い音声信号の期間)を使用することができる。その結果、雑音により低減された第1置換信号を生成することができる。

第2置換信号を生成する際にも、第1置換信号を生成する場合と同様に、信号位相差のばらつきの低下を抑制して、より多くの類似期間(第2参照期間におけるRch入力音声信号との類似度が高い音声信号の期間)を使用することができる。その結果、雑音により低減された第2置換信号を生成することができる。

【0072】

また、本実施例では、第2類似期間は、第1類似期間を遅延時間だけ遅延させることによって設定される。そのため、第2類似期間を設定する際には、複数の比較期間を設定する処理、音声信号を比較する処理、信号類似度を算出する処理、等は省略される。それにより、類似期間検出部93の処理負荷を大幅に低減することができる。

【0073】

以上述べたように、本実施例によれば、第1置換信号を生成する際に、2つ以上の第1類似期間を遅延時間だけ遅延させた2つ以上の期間のそれぞれにおけるRch音声信号が

さらに使用する。それにより、第1置換信号を生成する際に、信号位相差のばらつきの低下を抑制して、より多くの類似期間を使用することができる。その結果、雑音がより低減された第1置換信号を生成することができ、雑音がより低減されたR c h出力音声信号を生成することができる。

また、本実施例によれば、第1置換信号を生成する際に用いた2つ以上の第1類似期間を遅延時間だけ遅延させることにより、第2置換信号を生成する際に使用する2つ以上の第2類似期間が設定される。それにより、第2類似期間を設定する処理の処理負荷を大幅に低減することができる。

【0074】

なお、本実施例では、2つの入力音声信号に雑音低減処理を施す例を説明したが、3つ以上の入力音声信号に雑音低減処理を施す場合にも上述した処理と同様の処理を行うことによって、処理負荷を低減したり、雑音低減処理の精度を高めたりすることができる。具体的には、3つ以上の入力音声信号に雑音低減処理を施す場合には、3つ以上の入力音声信号の間の遅延時間を検出し、遅延時間を用いて類似期間を設定すればよい。

なお、本実施例では、第1音声信号がL c h入力音声信号であり、第2音声信号がR c h入力音声信号である場合の例を説明したが、これに限らない。例えば、第2音声信号がL c h入力音声信号であり、第1音声信号がR c h入力音声信号であってもよい。即ち、上述した説明において、“L c h入力音声信号”を“R c h入力音声信号”と読み替え、“R c h入力音声信号”を“L c h入力音声信号”と読み替えてもよい。

【0075】

＜その他の実施例＞

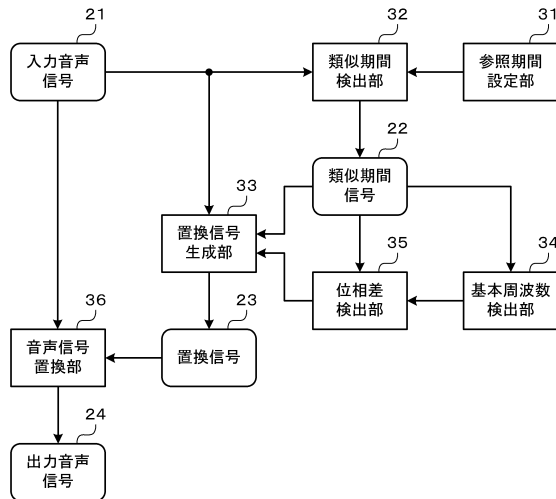
記憶装置に記録されたプログラムを読み込み実行することで前述した実施例の機能を実現するシステムや装置のコンピュータ（又はCPU、MPU等のデバイス）によっても、本発明を実施することができる。また、例えば、記憶装置に記録されたプログラムを読み込み実行することで前述した実施例の機能を実現するシステムや装置のコンピュータによって実行されるステップからなる方法によっても、本発明を実施することができる。この目的のために、上記プログラムは、例えば、ネットワークを通じて、又は、上記記憶装置となり得る様々なタイプの記録媒体（つまり、非一時的にデータを保持するコンピュータ読取可能な記録媒体）から、上記コンピュータ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

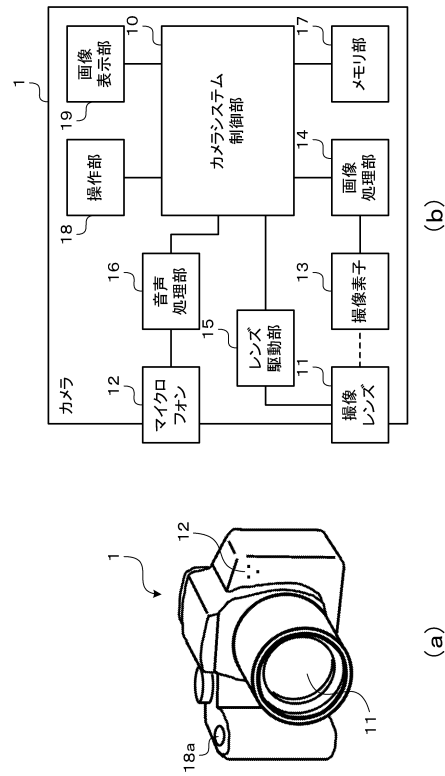
【0076】

- 1, 81 : カメラ 11 : 撮像レンズ 15 : レンズ駆動部
- 16, 83 : 音声処理部 31, 92 : 参照期間設定部
- 32, 93 : 類似期間検出部 33 : 置換信号生成部
- 34 : 基本周波数検出部 35 : 位相差検出部 36 : 音声信号置換部
- 91 : 遅延検出部

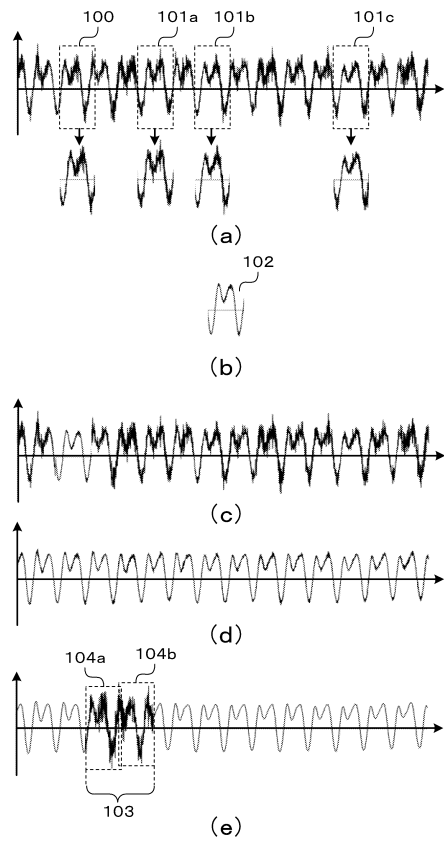
【図 1】



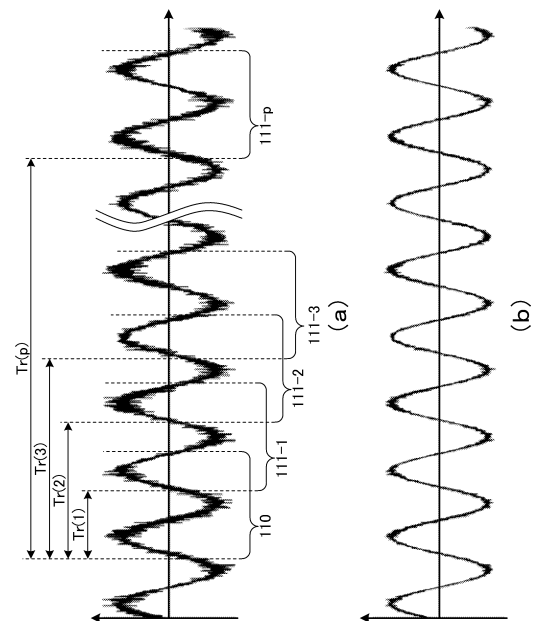
【図 2】



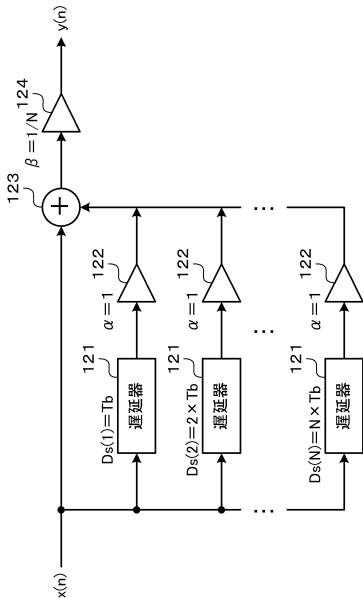
【図 3】



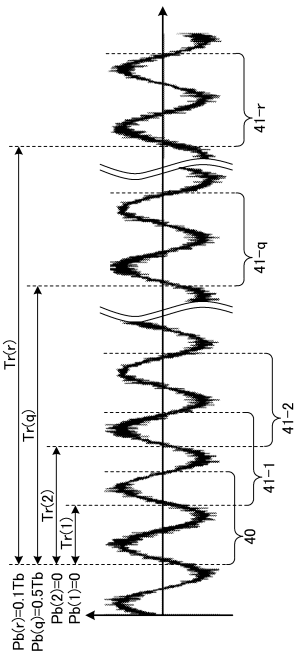
【図 4】



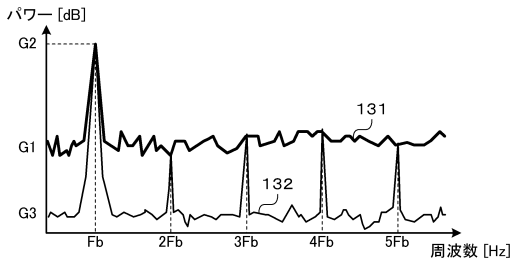
【図 5】



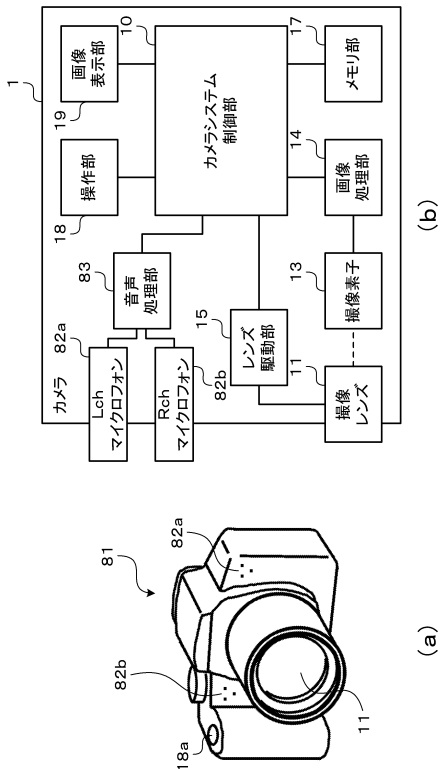
【図 7】



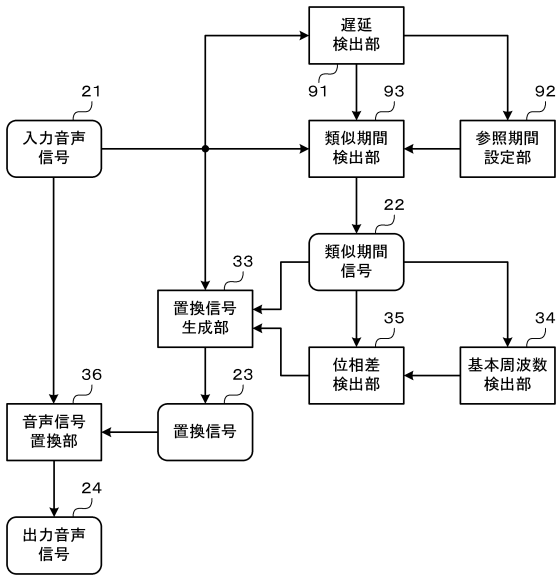
【図 6】



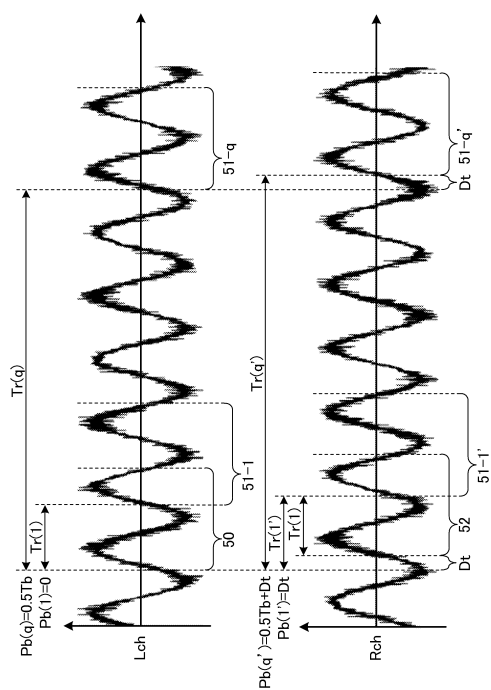
【図 8】



【図 9】



【図 10】



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2011-095378 (JP, A)
再公表特許第2007/086365 (JP, A1)

(58)調査した分野(Int.Cl., DB名)
G10L 13/00-99/00
H04N 5/225
H04R 3/00