

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6334895号
(P6334895)

(45) 発行日 平成30年5月30日 (2018.5.30)

(24) 登録日 平成30年5月11日 (2018.5.11)

(51) Int.Cl.

H04R 3/00 (2006.01)

F I

H04R 3/00 320

請求項の数 18 (全 20 頁)

(21) 出願番号	特願2013-237350 (P2013-237350)	(73) 特許権者	000001007
(22) 出願日	平成25年11月15日 (2013.11.15)		キヤノン株式会社
(65) 公開番号	特開2015-97355 (P2015-97355A)		東京都大田区下丸子3丁目30番2号
(43) 公開日	平成27年5月21日 (2015.5.21)	(74) 代理人	100076428
審査請求日	平成28年11月14日 (2016.11.14)		弁理士 大塚 康德
		(74) 代理人	100112508
			弁理士 高柳 司郎
		(74) 代理人	100115071
			弁理士 大塚 康弘
		(74) 代理人	100116894
			弁理士 木村 秀二
		(74) 代理人	100130409
			弁理士 下山 治
		(74) 代理人	100134175
			弁理士 永川 行光

最終頁に続く

(54) 【発明の名称】 信号処理装置及びその制御方法、プログラム

(57) 【特許請求の範囲】

【請求項 1】

收音手段により收音される收音信号を取得する取得手段と、
前記取得手段により取得される第1收音信号に含まれる雑音を抑制する抑制手段と、
前記取得手段により前記第1收音信号よりも前に取得された第2收音信号を用いた学習の結果に基づいて、前記第1收音信号に対応する目的音信号を生成する生成手段と、
前記生成手段により生成される前記第1收音信号に対応する目的音信号を出力する第1の出力形態と、前記抑制手段により前記第1收音信号から雑音が抑制された雑音抑制後信号を出力する第2の出力形態とを含む複数の出力形態から、適用すべき出力形態を決定する決定手段と、
前記決定手段により決定される出力形態に応じた信号を出力する出力手段と、
を備えることを特徴とする信号処理装置。

【請求項 2】

前記決定手段は、前記目的音信号と前記雑音抑制後信号とを混合することにより得られる混合信号を出力する第3の出力形態と、前記第1の出力形態と、前記第2の出力形態とを含む複数の出力形態から、前記適用すべき出力形態を決定することを特徴とする請求項1に記載の信号処理装置。

【請求項 3】

前記取得手段により取得される收音信号に含まれる雑音が所定の大きさより小さいかを検出する検出手段と、

前記第 2 收音信号に含まれる雑音が前記所定の大きさより小さいことが前記検出手段により検出された場合に、前記第 2 收音信号を用いて学習を行う学習手段と、
を更に備え、

前記生成手段は、前記学習手段による学習の結果に基づいて前記第 1 收音信号に対応する目的音信号を生成する
ことを特徴とする請求項 1 又は 2 に記載の信号処理装置。

【請求項 4】

前記取得手段により取得される收音信号から雑音信号を推定する推定手段を更に備え、
前記検出手段は、前記取得手段により取得される收音信号に含まれる雑音が前記所定の大きさより小さいかを、前記推定手段により推定される雑音信号に基づいて検出し、
前記抑制手段は、前記取得手段により取得される收音信号に含まれる雑音を、前記推定手段により推定される雑音信号に基づいて抑制する
ことを特徴とする請求項 3 に記載の信号処理装置。

10

【請求項 5】

前記学習手段は、前記第 2 收音信号を解析することで得られる特性を学習してモデル化することで目的音モデルを生成し、
前記生成手段は、前記学習手段により生成される目的音モデルによって前記第 1 收音信号をモデリングすることで前記第 1 收音信号に対応する目的音信号を生成する
ことを特徴とする請求項 3 に記載の信号処理装置。

20

【請求項 6】

前記決定手段は、前記目的音モデルの活性度に応じて、前記適用すべき出力形態を決定する
ことを特徴とする請求項 5 に記載の信号処理装置。

【請求項 7】

前記検出手段は、前記推定手段で推定した雑音信号の処理単位フレームにおける時間振幅絶対値の平均値が、予め定められた閾値以下である場合に、前記取得手段により取得される收音信号に含まれる雑音が前記所定の大きさより小さいことを検出する
ことを特徴とする請求項 4 に記載の信号処理装置。

【請求項 8】

前記検出手段は、処理単位フレームにおける複数の收音手段それぞれにより收音された收音信号間の相関度が、予め定められた閾値よりも大きい場合に、当該收音信号に含まれる雑音が前記所定の大きさより小さいことを検出する
ことを特徴とする請求項 4 に記載の信号処理装置。

30

【請求項 9】

前記取得手段により取得される收音信号に含まれる雑音が所定の大きさより小さいかを検出する検出手段と、
前記第 2 收音信号に含まれる雑音が前記所定の大きさより小さいことが前記検出手段により検出された場合に、前記第 2 收音信号を記憶する記憶手段と、
前記記憶手段に記憶されている收音信号を用いて、非負値行列因子分解を繰り返し行うことにより、基底スペクトルを学習する学習手段と、
を更に備え、

40

前記生成手段は、前記学習手段で学習した基底スペクトルを用いて、前記第 1 收音信号の非負値行列因子分解を行って基底アクティベートを計算し、当該計算の結果に基づいて目的音を生成し、

前記決定手段は、前記生成手段が出力する基底アクティベートの係数の大きさに応じて、前記適用すべき出力形態を決定する
ことを特徴とする請求項 1 に記載の信号処理装置。

【請求項 10】

前記取得手段により取得される收音信号から雑音信号を推定する推定手段と、
前記推定手段により推定される雑音信号のスペクトル分布に応じて、收音信号を低域と

50

高域に分割する分割周波数を決定する第 2 決定手段と、
を更に備え、

前記生成手段は、前記第 2 決定手段で決定した分割周波数より高域の基底スペクトルに基づいて、前記取得手段により取得される收音信号の非負値行列因子分解を行って基底アクティベートを計算する

ことを特徴とする請求項 9 に記載の信号処理装置。

【請求項 11】

前記出力手段は、前記決定手段により第 3 の出力形態が前記適用すべき出力形態として決定された場合に、前記生成手段が出力する基底アクティベートの係数の大きさに応じて、前記雑音抑制後信号の前記分割周波数より低域の低域成分を、前記目的音信号の低域成分に置換して出力する

ことを特徴とする請求項 10 に記載の信号処理装置。

【請求項 12】

前記出力手段は、前記決定手段により第 3 の出力形態が前記適用すべき出力形態として決定された場合に、前記生成手段が出力する基底アクティベートの係数の大きさに応じて、前記雑音抑制後信号の前記分割周波数より低域の低域成分に、前記目的音信号の低域成分を混合して出力する

ことを特徴とする請求項 10 に記載の信号処理装置。

【請求項 13】

前記抑制手段は、スペクトル減算、ハイパスフィルタ及びウィナーフィルタの少なくともいづれかを用いて、前記取得手段により取得される收音信号に含まれる雑音を抑制する

ことを特徴とする請求項 1 乃至 12 のいずれか 1 項に記載の信号処理装置。

【請求項 14】

複数の收音手段を備え、

前記推定手段は、ビームフォーマー及び独立成分分析の少なくともいづれかを用いて、前記取得手段により取得される收音信号から雑音信号を推定する

ことを特徴とする請求項 4 又は 10 に記載の信号処理装置。

【請求項 15】

信号処理装置の制御方法であって、

收音手段により收音される收音信号を取得する取得工程と、

前記取得工程において取得される第 1 收音信号に含まれる雑音を抑制する抑制工程と、

前記取得工程において前記第 1 收音信号よりも前に取得された第 2 收音信号を用いた学習の結果に基づいて、前記第 1 收音信号に対応する目的音信号を生成する生成工程と、

前記生成工程において生成される前記第 1 收音信号に対応する目的音信号を出力する第 1 の出力形態と、前記抑制工程において前記第 1 收音信号から雑音が抑制された雑音抑制後信号を出力する第 2 の出力形態とを含む複数の出力形態から、適用すべき出力形態を決定する決定工程と、

前記決定工程において決定される出力形態に応じた信号を出力する出力工程と、を含むことを特徴とする制御方法。

【請求項 16】

前記決定工程においては、前記目的音信号と前記雑音抑制後信号とを混合することにより得られる混合信号を出力する第 3 の出力形態と、前記第 1 の出力形態と、前記第 2 の出力形態とを含む複数の出力形態から、前記適用すべき出力形態が決定される

ことを特徴とする請求項 15 に記載の制御方法。

【請求項 17】

前記取得工程において取得される收音信号に含まれる雑音が所定の大きさより小さいかを検出する検出工程と、

前記第 2 收音信号に含まれる雑音が前記所定の大きさより小さいことが前記検出工程において検出された場合に、前記第 2 收音信号を用いて学習を行う学習工程と、

10

20

30

40

50

を更に含み、

前記生成工程においては、前記学習工程における学習の結果に基づいて前記第1收音信号に対応する目的音信号が生成される

ことを特徴とする請求項15又は16に記載の制御方法。

【請求項18】

コンピュータを請求項1乃至14の何れか1項に記載の信号処理装置の各手段として動作させるためのプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

10

本発明は、風雑音を抑制しつつ周囲の音を収録する收音技術に関する。

【背景技術】

【0002】

近年、カムコーダやカメラ、スマートフォン等の撮像装置の普及により気軽に画像が撮影できるようになってきている。また、高音質録音が可能なポータブルオーディオレコーダも多く普及しており、画像が付随する・しないに関わらず、屋外で周囲、もしくは目的物の音を録音する機会が増えている。

【0003】

このような屋外で收音する場合において、風が收音用マイクロフォンに作用することによって生じる雑音（以下、風雑音と呼称）が收音信号に混じると、目的音が聞き取りにくくなり、また、不快な音になる。そこで、風雑音を除去、または抑制することが、従来から重要な課題になっている。

20

【0004】

風雑音の周波数特性を分析すると、そのエネルギーの多くは500Hz以下の低周波数域に偏るという特徴を持っている。そこで、風雑音を抑制する従来技術の一つとして、高周波数域通過フィルタ（以下、ハイパスフィルタと呼称）を用いて風雑音を抑制する手法がある。

【0005】

ところが、ハイパスフィルタを用いた風雑音抑制手法では、風雑音のレベルが大きい場合、ハイパスフィルタもそれに応じて抑制量を大きくする必要がある。そのため、目的音成分の低周波数域が丸ごと抑制され、目的音の音色が変化してしまうという問題がある。

30

【0006】

また、風雑音を抑制する従来技術の一つとして、風雑音信号を推定して、收音信号からスペクトル減算を行うことにより抑制する技術がある。

【0007】

しかしながら、スペクトル減算を用いた抑制方法においても、風雑音のレベルが大きくなりすぎると目的音成分自体がかき消されてしまい、風雑音を減算すると目的音成分までなくなってしまうという問題がある。

【0008】

そこで、風雑音抑制処理によって失われる目的音成分を、風雑音抑制後に復元してその目的音成分を補完するという従来技術が存在する。

40

【0009】

例えば、特許文献1では、入力信号を低・中・高の三帯域に分離し、中帯域から低帯域の復元信号を生成し、風雑音の影響度合いを推定して入力信号の低帯域信号と混合している。また、中帯域の信号レベルを低減して混合している。このような構成により、歪の発生を抑制しつつ風雑音を低減するという技術が開示されている。

【先行技術文献】

【特許文献】

【0010】

【特許文献1】特開2009-55583号公報

50

【発明の概要】

【発明が解決しようとする課題】

【0011】

しかしながら、特許文献1の技術では、調波性のある中帯域、高帯域信号を利用して基本波や低次高調波を復元するものであり、調波性のある信号しか復元できないという課題がある。また、基本波を特定する情報は持っておらず、低次高調波のレベルバランスも考慮しないため、不正確な低帯域成分を付加してしまい、かえって音質が劣化する、あるいは、音色が変化してしまう恐れがあった。

【0012】

本発明は上記の課題を解決するためになされたものであり、雑音を抑制しつつ、音色変化や目的音成分の欠落を防止して、精密な目的音の復元を行うことができる收音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】

【0013】

上記の目的を達成するための本発明による信号処理装置は以下の構成を備える。即ち、信号処理装置は、

收音手段により收音される收音信号を取得する取得手段と、

前記取得手段により取得される第1收音信号に含まれる雑音を抑制する抑制手段と、

前記取得手段により前記第1收音信号よりも前に取得された第2收音信号を用いた学習の結果に基づいて、前記第1收音信号に対応する目的音信号を生成する生成手段と、

前記生成手段により生成される前記第1收音信号に対応する目的音信号を出力する第1の出力形態と、前記抑制手段により前記第1收音信号から雑音が抑制された雑音抑制後信号を出力する第2の出力形態とを含む複数の出力形態から、適用すべき出力形態を決定する決定手段と、

前記決定手段により決定される出力形態に応じた信号を出力する出力手段と、
を備える。

【発明の効果】

【0014】

本発明によれば、雑音を抑制しつつ、音色変化や目的音成分の欠落を防止して、精密な目的音の復元を行うことができる。

【図面の簡単な説明】

【0015】

【図1】実施形態1の收音装置の構成を示すブロック図である。

【図2】実施形態1の收音装置の收音処理を示すフローチャートである。

【図3】実施形態2の收音装置の構成を示すブロック図である。

【図4】実施形態2の收音装置の收音処理を示すフローチャートである。

【図5】実施形態3の收音装置の構成を示すブロック図である。

【図6】実施形態3の收音装置の收音処理を示すフローチャートである。

【発明を実施するための形態】

【0016】

以下、本発明の実施の形態について図面を用いて詳細に説明する。尚、以下の実施形態において示す構成は一例に過ぎず、本発明は図示された構成に限定されるものではない。

<実施形態1>

図1は、実施形態1の收音装置の構成を示すブロック図である。

【0017】

図1において、1は音入力部としてのマイクロフォンユニットであり、目的音を含む周囲の音を收音し、電気信号に変換する。2はマイクロフォンアンプであり、マイクロフォンユニット1が出力する微弱なアナログ音響信号を増幅して出力する。3はアナログデジタル変換器(ADC)であり、入力されたアナログ音響信号をデジタル音響信号に変換し、收音信号として出力する。

【 0 0 1 8 】

1 0 1 は雑音推定器であり、入力された收音信号に含まれる非定常雑音を推定して、推定雑音信号を出力する。1 0 2 は無雑音状態推定器であり、雑音推定器 1 0 1 が出力する推定雑音信号が無雑音状態（雑音が弱い、もしくは、雑音が発生していない状態）であるか否かを検出し、無雑音状態である場合にのみスイッチ ON 信号をスイッチ 1 0 8 に出力する。尚、無雑音状態をより定量的に表現すれば、無雑音状態とは、雑音の強度を示す雑音レベルが、雑音として知覚されない所定レベル以下である状態を意味する。

【 0 0 1 9 】

1 0 3 は目的音学習器であり、入力されたデジタル音響信号を目的音信号として解析し、そのスペクトル包絡や調波構造等の特性を学習し、これらの特性を複数のパターンに類型化して、目的音モデル 1 0 4 に出力する。

10

【 0 0 2 0 】

1 0 4 は目的音モデルであり、目的音学習器 1 0 3 が出力した目的音信号のパターン情報を格納し、目的音復元器 1 0 6 に適宜供給する。1 0 5 は雑音抑制器であり、雑音推定器 1 0 1 が出力する推定雑音信号に従って、收音信号から推定雑音を抑制した信号（雑音抑制後信号）を出力する。1 0 6 は目的音復元器であり、收音信号と目的音モデル 1 0 4 に格納されているパターン情報とのパターンマッチングを行うことにより、目的音信号を復元し、目的音復元信号として出力する。また、この時の目的音パターンの活性度を出力する。

【 0 0 2 1 】

20

1 0 7 は信号選択・混合器であり、雑音抑制器 1 0 5 から出力される雑音抑制後信号と、目的音復元器 1 0 6 が出力する目的音復元信号とを、学習モデルである目的音モデルの活性度に従って、適宜置換、もしくは混合を行って出力する。

【 0 0 2 2 】

尚、收音装置は、上記の構成以外に、汎用コンピュータに搭載される標準的な構成要素（例えば、CPU、RAM、ROM、ハードディスク、外部記憶装置、ネットワークインタフェース、ディスプレイ、キーボード、マウス等）を有することができる。そして、例えば、ハードディスク等に記憶されているプログラムを CPU が読み出し実行することによって、以下で説明する各種フローチャートの処理を実行することもできる。

【 0 0 2 3 】

30

以下、図 1 の構成において、目的音の成分欠落や音質劣化を防止しつつ、收音信号に含まれる非定常雑音を抑制する一連の動作をフローに従って説明する。

【 0 0 2 4 】

図 2 は、実施形態 1 の收音装置が実行する收音処理を示すフローチャートである。

【 0 0 2 5 】

まず、ステップ S 1 で、マイクロフォンユニット 1 によって目的音を含む周囲の音を電気信号に変換し、マイクロフォンアンプ 2 によって増幅し、ADC 3 において、デジタル信号に変換し、所定サンプル長の処理単位フレームに切り出して出力する。

【 0 0 2 6 】

ステップ S 2 で、雑音推定器 1 0 1 において、ステップ S 1 で切り出した收音信号の処理フレームに含まれる雑音信号を推定する。実施形態 1 において、モノラル音響信号から非定常雑音を推定する方法としては、線形予測を用いて予測できなかった成分を非定常雑音とする方法や、予め学習した音源（音声）信号モデルに合致しない成分を非定常雑音とする方法等を用いる。尚、これらの雑音推定処理は公知であり、一般的に利用されているものであるため、詳細な説明は行わない。

40

【 0 0 2 7 】

ステップ S 3 で、無雑音状態検出器 1 0 2 において、ステップ S 2 で得られた推定雑音信号の当該処理フレームにおける時間振幅絶対値の平均（雑音レベル）を計算する。これは、以下の式（1）によって計算できる。

【 0 0 2 8 】

50

【数 1】

$$\frac{\sum_{t=1}^T |a_t|}{T} \quad (1)$$

【0029】

但し、 T はフレームサンプル数、 a_t はフレーム内の時間 t における推定雑音信号の時間振幅である。

【0030】

ステップ $S4$ で、無雑音状態検出器 102 において、ステップ $S3$ で計算した時間振幅絶対値の平均が、予め定められた閾値以下であるか否かを判定する。時間振幅絶対値の平均が閾値より大きい場合（ステップ $S4$ で NO ）、無雑音状態検出器 102 は、当該処理フレームの時間区間を雑音状態と判定して、ステップ $S7$ へ進む。この場合、無雑音状態検出器 102 は、信号を出力しない。

10

【0031】

一方、時間振幅絶対値の平均が閾値以下である場合（ステップ $S4$ で YES ）、無雑音状態検出器 102 は、当該処理フレームの時間区間を無雑音状態であると判定し、ステップ $S5$ へ進む。この場合、無雑音状態検出器 102 は、スイッチ ON 信号をスイッチ 108 に出力する。これにより、スイッチ 108 が接続されるため、目的音学習器 103 に收音信号が入力される。

20

【0032】

ステップ $S5$ で、目的音学習器 103 において、当該処理フレームの收音信号を目的音として、その特性を解析する。この解析によって、收音信号のスペクトル包絡や調波構造、時間波形包絡等が解析結果として得られる。

【0033】

ステップ $S6$ で、目的音学習器 103 において、ステップ $S5$ で得られた收音信号の特性を目的音モデル変数として目的音モデル 104 に追加することにより、目的音モデル 104 の再構築を行う。

【0034】

以上の処理により、ステップ $S4$ で無雑音状態と判定した処理フレームの收音信号を目的音信号としてステップ $S5$ で解析し、ステップ $S6$ でその特性を目的音モデル変数として追加することにより目的音モデル 104 を再構築する。これにより、非定常雑音の影響を避けつつ、より正確な目的音モデル変数を收音信号から学習することができる。

30

【0035】

ステップ $S7$ で、雑音抑制器 105 において、ステップ $S2$ で得られた推定雑音信号に基づいて、当該処理フレームの收音信号に対して雑音抑制を行う。実施形態 1 において、この処理は、收音信号のスペクトル振幅から推定雑音信号のスペクトル振幅を減算することによって行われる。

【0036】

尚、実施形態 1 において、スペクトル減算を用いるのはあくまでも一例である。例えば、推定雑音信号のスペクトルエネルギー分布に基づいてカットオフ周波数を定めたハイパスフィルタ処理を行うようにしても、同様な処理が可能である。あるいは、処理単位フレームの周波数成分毎に、推定雑音が占めるエネルギーの割合を計算することで、ウィナーフィルタを設計して收音信号から推定雑音成分を除去する処理を行ってもよく、本発明の範囲を限定するものではない。

40

【0037】

ステップ $S8$ で、目的音復元器 106 において、收音信号の特性を解析して、目的音モデル 104 に格納されている目的音モデル変数を用いてモデリングを行うことにより、目的音を復元する。具体的には、收音信号を解析して得られるスペクトル包絡や調波構造等の特性と、目的音モデル 104 に格納されている目的音モデル変数とのパターンマッチン

50

グを行う。次に、マッチングしたパターンを組み合わせることでにより収音信号をモデル化することによって、目的音信号を復元し、出力する。

【 0 0 3 8 】

例えば、実施形態 1 では、スペクトル包絡のモデル変数として、当分野で一般的に用いられている L P C (L i n e a r P r e d i c t i o n C o d i n g : 線形予測符号) スペクトル包絡を用いる。処理対象フレームの収音信号を線形予測分析して得られる L P C スペクトル包絡を $g (\quad)$ 、目的音モデル 1 0 4 に格納されている i 番目の L P C スペクトル包絡を $f_i (\quad)$ とする。実施形態 1 では、この 2 つのマッチングを $c o s h$ 尺度によって計算する。 $c o s h$ 尺度は、以下の式 (2) で計算する。

【 0 0 3 9 】

【 数 2 】

$$COSH_{fi} = \frac{1}{2\pi} \int_{-\pi}^{\pi} 2 \cdot \{ \cosh (\log f_i(\lambda) - \log g(\lambda)) - 1 \} d\lambda \quad (2)$$

【 0 0 4 0 】

但し、 λ は角周波数 ($-\pi < \lambda < \pi$) である。

【 0 0 4 1 】

ここで、 $f_i (\quad)$ と $g (\quad)$ の対数スペクトル差分を $V (\quad)$ とする。

【 0 0 4 2 】

【 数 3 】

$$V(\lambda) = \log f_i(\lambda) - \log g(\lambda) \quad (3)$$

【 0 0 4 3 】

式 (2) より、 $C O S H_{fi}$ の値は、 $V (\quad)$ を用いて、以下の式 (4) で記述できる。

【 0 0 4 4 】

【 数 4 】

$$COSH_{fi} = \frac{1}{2\pi} \int_{-\pi}^{\pi} (e^{V(\lambda)} + e^{-V(\lambda)} - 2) d\lambda \quad (4)$$

【 0 0 4 5 】

式 (4) の積分項を $V (\quad) = 0$ のまわりでテーラー展開すると、以下の式 (5) になる。

【 0 0 4 6 】

【 数 5 】

$$e^{V(\lambda)} + e^{-V(\lambda)} - 2 = \sum_{i=1}^{\infty} \frac{2}{(2i)!} V(\lambda)^{2i} = V(\lambda)^2 + \frac{1}{12} V(\lambda)^4 + \frac{1}{360} V(\lambda)^6 + \dots$$

(5)

【 0 0 4 7 】

よって、 $| V (\quad) |$ が小さい場合、すなわち、マッチング度合いが高い場合は、 $C O S H_{fi}$ の値はその値の二乗に極めて近い重みになる。一方、 $| V (\quad) |$ が大きい場合、すなわち、マッチング度合いが低い場合は、 $C O S H_{fi}$ の値は指数関数 $e^{| V (\quad) |}$ の重みになる。

【 0 0 4 8 】

以上のように、式 (2) の計算を目的音モデル 1 0 4 に格納されている全ての L P C ス

10

20

30

40

50

ペクトル包絡に対して行い、C O S H 値が最も小さい値となる L P C スペクトル包絡 f を目的音復元に用いるモデル変数として使用する。

【 0 0 4 9 】

このとき、選択した L P C スペクトル包絡 f の活性度 α_{spctr} を、以下の式 (6) で計算する。

【 0 0 5 0 】

【 数 6 】

$$\alpha_{\text{spctr}} = \frac{1}{1 + \text{COSH}_f} \quad (6)$$

10

【 0 0 5 1 】

モデル変数として参照される L P C スペクトル包絡と收音信号の L P C スペクトル包絡との差が少ないほど、C O S H 値の値は小さくなり限りなく 0 に近づくため、モデル変数とのマッチング度合いが高いほど α_{spctr} の値は 1 に近づく。また、マッチング度合いが小さいほど C O S H 値は大きくなるため、 α_{spctr} の値は 0 に近づく。

【 0 0 5 2 】

次に、目的音復元器 1 0 6 は、目的音モデル 1 0 4 に格納されている全ての調波構造と、收音信号の調波構造とのマッチングを取り、最もマッチングする調波構造を目的音復元に用いるモデル変数として選択する。さらに、その活性度 α_{harm} を α_{spctr} と同様な値域

20

【 0 0 5 3 】

次に、目的音復元器 1 0 6 は、最も活性度が大きいスペクトル包絡と調波構造を周波数領域で畳み込み、逆 F F T を行うことにより、時間領域の目的音復元信号を復元する。

【 0 0 5 4 】

このとき、目的音モデル 1 0 4 全体の活性度 α を、以下の式 (7) で計算する。

【 0 0 5 5 】

【 数 7 】

$$\alpha = \frac{\alpha_{\text{spctr}} + \alpha_{\text{harm}}}{2} \quad (7)$$

30

【 0 0 5 6 】

目的音復元器 1 0 6 は、活性度 α を目的音復元信号と同時に信号選択・混合器 1 0 7 に出力する。

【 0 0 5 7 】

ステップ S 9 で、信号選択・混合器 1 0 7 において、ステップ S 8 で計算した目的音モデル 1 0 4 の活性度 α の値を確認し、予め定められた閾値、A、B と比較する。尚、 $A > B$ である。

【 0 0 5 8 】

ここで、A、B の実際の値は、例えば、様々な α の条件で復元した目的音復元信号と実際の目的音信号との聴感上の比較実験を行い、その結果において、5 % の有意水準で有意性が認められた α 値とする。つまり、目的音復元信号と目的音信号がほぼ等しいことが 5 % の有意水準で有意性が認められた場合の α 値の内、最小値を A とする。また、目的音復元信号と目的音信号が全く異なっていることが 5 % の有意水準で有意性が認められた場合の α 値の内、最大値を B とする。

40

【 0 0 5 9 】

ステップ S 9 における比較の結果、 $\alpha > A$ となる場合は、信号選択・混合器 1 0 7 において、ステップ S 8 で得られた目的音復元信号が実際の目的音とほぼ等しいと判定する。そして、ステップ S 1 0 で、信号選択・混合器 1 0 7 において、目的音復元器 1 0 6 から

50

入力した目的音復元信号をそのまま出力する（第１の出力形態）。

【００６０】

ステップＳ９における比較の結果、 $B < A$ となる場合は、信号選択・混合器１０７において、ステップＳ８で得られた目的音復元信号には実際の目的音がある程度含まれていると判定する。そして、ステップＳ１１で、信号選択・混合器１０７において、雑音抑制信号と目的音復元信号の混合率を計算する。これは、例えば、目的音モデル１０４の活性度に基づいて、以下の式（８）で計算する。

【００６１】

【数８】

$$\beta = (A - \alpha) / (A - B) \quad (8)$$

10

【００６２】

ステップＳ１２で、ステップＳ１１で計算した混合率に基づいて、雑音抑制信号と目的音復元信号を混合して出力する（第２の出力形態）。ある時間 t に対する雑音抑制信号の時間振幅を z_t 、目的音復元信号の時間振幅を s_t とすると、時間 t に対する混合信号 m_t は、以下の式（９）で計算する。

【００６３】

【数９】

$$m_t = \beta \cdot z_t + (1 - \beta) \cdot s_t \quad (9)$$

20

【００６４】

式（８）より、活性度が大きいほど、混合率は小さくなるので、式（９）より混合信号における目的音復元信号の割合が大きくなることになる。

【００６５】

尚、実施形態１では、時間領域信号において混合しているが、周波数領域で混合してもよい。

30

【００６６】

ステップＳ９における比較の結果、 $A < B$ となる場合は、信号選択・混合器１０７において、ステップＳ８で得られた目的音復元信号には実際の目的音はほぼ含まれていないと判定する。そして、ステップＳ１３で、信号選択・混合器１０７において、ステップＳ７で生成した雑音抑制信号を出力する（第３の出力形態）。このようにすることによって、学習モデルが活性化されない場合に、誤って復元された信号が最終的な出力に反映されることを防止することができる。

【００６７】

ステップＳ９からステップＳ１３までの処理を実行することによって、学習した目的音モデルの活性度に応じて、目的音復元信号の確からしさを判定し、それによって目的音復元信号と雑音抑制信号の置換・混合の出力形態を決定することができる。このようにすることで、雑音によって失われる目的音成分を補完しつつ、不完全な学習モデルによる不完全な目的音復元信号が混入することを避けることが可能になるため、より正確な目的音信号を取り出すことができる。

40

【００６８】

ステップＳ１４で、收音処理を終了する制御部（不図示）による指示があるか否かを判定する。指示がない場合（ステップＳ１４でＮＯ）、ステップＳ１へ戻る。一方、指示がある場合（ステップＳ１４でＹＥＳ）、收音処理を終了する。

【００６９】

以上説明したように、実施形態１によれば、無雑音区間における入力信号から目的音の

50

特性を学習し、雑音抑制で失われる目的音成分を学習モデルによって復元する。また、学習モデルと入力信号による学習モデルの活性度に応じて雑音抑制信号を補正する。これによって、風雑音を抑制しつつ、音色変化や目的音成分の欠落を防止することができる。

【0070】

より具体的には、雑音の非定常性を利用することにより、雑音が弱い、もしくは、雑音が発生していない区間（無雑音区間）において、目的音の特性を学習し、学習モデルと入力信号のマッチング状態に応じて雑音抑制後の信号補正を制御する。これにより、たとえ、調波性を持たない目的音信号であっても、雑音抑制処理によって欠落する目的音信号を学習したモデルにより復元し、風雑音抑制後の信号をより精密に補正することができる。

【0071】

<実施形態2>

実施形態2では、入力信号が複数で、かつ、目的音の学習方法として非負値行列因子分解（NMF：Nonnegative Matrix Factorization）を用いる構成について説明する。

【0072】

図3は、実施形態2の收音装置の構成を示すブロック図である。

【0073】

図中のマイクロフォンユニット1、マイクロフォンアンプ2、ADC3は、図1の構成と同様であるので説明を省略する。実施形態2の構成では、マイクロフォンユニット1、マイクロフォンアンプ2、ADC3の各々が、1chからLchまでのL個（Lチャンネル：Lは自然数）分用意され、Lchの收音信号を收音する。L個のマイクロフォンユニット1は、同一球面上の上下左右前後の様々な方向に向けられていてもよいし、同一の平面上、もしくは線上において、全て同じ方向に並行して向けられていてもよい。

【0074】

201は風雑音推定器であり、Lchの收音信号から各チャンネルの風雑音信号を推定して、推定雑音信号を出力する。202は無雑音状態検出器であり、Lchの推定雑音信号各々に対して、無雑音状態であるか否かを判定し、無雑音状態であると判定したチャンネルに対するスイッチON信号をスイッチ109各々に出力する。203は無雑音信号DB（データベース）であり、当該フレームの無雑音状態であると判定された各チャンネルの入力信号を記憶、保存する。

【0075】

204は目的音基底スペクトル学習器であり、NMFを用いて無雑音信号DB203に記憶されている入力信号の学習を行う。205は目的音モデルであり、目的音基底スペクトル学習器204における目的音学習結果として出力される基底スペクトルを格納し、必要に応じて出力する。206は風雑音抑制器であり、Lchの收音信号に対して、風雑音推定器201によって出力されるLchの推定雑音信号に基づいて風雑音の抑制処理を行い、雑音抑制後信号を出力する。

【0076】

207は目的音復元器であり、Lchの收音信号に対して、目的音モデル205に格納された基底スペクトルによる制限付NMFを行い、Lch分の基底アクティベートを計算し、それによって收音信号に含まれるLch分の目的音信号を復元し、目的音復元信号として出力する。208は信号選択・混合器であり、風雑音抑制器206から出力されるLch分の雑音抑制後信号と、目的音復元器207から出力されるLch分の目的音復元信号を、各チャンネル毎に選択・混合して出力する。尚、選択・混合の判断は、目的音復元器207から出力されるLch分の基底アクティベートの係数の大きさに基づいて行う。

【0077】

以下、図3の構成において、收音信号に含まれる非定常雑音（風雑音）を抑制しつつ、NMFにより学習したモデルに基づいて雑音抑制によって欠落する目的音の補正を行う一連の動作をフローに従って説明する。

【0078】

10

20

30

40

50

図4は、実施形態2の收音装置が実行する收音処理を示すフローチャートである。

【0079】

まず、ステップS101で、マイクロフォンユニット1で周囲の音を收音して電気信号に変換し、マイクロフォンアンプ2によって増幅し、ADC3において、デジタル信号に変換し、所定サンプル長の処理単位のフレームに切り出して出力する。ステップS101では、この処理をLch分並行して行う。

【0080】

ステップS102で、風雑音推定器201において、ステップS1で切り出したLch分の收音信号を分析し、それらに含まれる風雑音を推定する。多チャンネル收音信号から風雑音のような拡散性のある雑音を推定する方法としては、次のようなものがある。ビームフォーマーを用いて、指向性を持つ成分、つまり、目的音の到来する方向にヌルを向けるようにすることで、無指向性の雑音を取り出す方法がある。また、ICA（独立成分分析）を用いて拡散性を持つ信号だけを取り出す方法がある。風雑音と目的音では、空間における拡散性や指向性が全く異なるため、このような方法を用いることで有効に風雑音を推定することができる。

【0081】

尚、これらの方法で推定した推定雑音信号は、手法によってはLch分全てがモノラル信号に統合されて出力される場合もあるが、推定する際の多チャンネル処理の逆変換を推定雑音信号に対して行うことにより、Lch分の信号に変換することができる。実施形態2では、ステップS102によって收音信号の各チャンネルに対応するLch分の推定雑音信号が得られるものとする。これらの方法は、音源分離技術として一般に用いられており、公知であるため、詳細な説明は行わない。

【0082】

ステップS103で、無雑音状態検出器202において、ステップS102で推定したLch分の推定雑音信号各々に対して、時間振幅絶対値の平均を計算する。この計算は、図2のステップS3と同様に、式(1)で計算する。

【0083】

ステップS104で、無雑音状態検出器202において、ステップS103で計算した各チャンネルの時間振幅絶対値の平均が、予め定められた閾値以下であるか否かを判定し、閾値以下のチャンネルのスイッチON信号をスイッチ209それぞれに出力する。この処理によって、スイッチON信号が出力されたチャンネルの收音信号と無雑音信号DB203を接続するスイッチ209がONになる。

【0084】

ステップS105で、無雑音信号DB203において、ステップS104によってスイッチON信号が出力されたチャンネルの收音信号を、それぞれ無雑音信号として保存する。

【0085】

ステップS106で、目的音基底スペクトル学習器204において、ステップS105によって更新した無雑音信号DB203に基づいて、NMFによる学習を行う。具体的には、この学習は、以下のように行う。

【0086】

まず、無雑音信号DB203に新たに格納された收音信号の各々に対して、短時間フーリエ変換を行って、スペクトログラムを作成し、これまでのフレーム処理で作成したスペクトログラムの最後尾に追加する。このスペクトログラムを $M \times N$ の大きさの二次元行列 V で表現する。ここで、 M はスペクトルの分解能、 N はスペクトログラムの時間サンプルである。次に、これを、 K 個の基底スペクトルとその各々の活性度に分解する。つまり、 $M \times K$ の非負値の基底スペクトル行列 H と $K \times N$ の非負値の基底アクティベート U の積に分解する。

【0087】

10

20

30

40

【数 1 0】

$$\mathbf{V} \cong \mathbf{H}\mathbf{U} \quad (10)$$

【0088】

ここで、コスト関数は、以下の式(11)のようになる。

【0089】

【数 1 1】

$$\|\mathbf{V} - \mathbf{H}\mathbf{U}\|_F^2 = \sum_{m,n} |V_{m,n} - \sum_k H_{m,k} U_{k,n}|^2 \quad (11)$$

10

【0090】

式(11)は、Frobeniusノルム規準と呼ばれる。

【0091】

実施形態2では、式(11)の値が最小となるように基底スペクトルと基底アクティベートを最適化することにより学習を行う。Frobeniusノルム規準の一般的な解法として、Jensenの不等式を用いて補助関数を作成し、それを最適化する式を代入することによって、次の最適化式が得られる。

【0092】

20

【数 1 2】

$$H_{m,k} \leftarrow H_{m,k} \frac{\sum_n V_{m,n} U_{k,n}}{\sum_n U_{k,n} \sum_{k'} H_{m,k'} U_{k',n}} \quad (12)$$

【0093】

【数 1 3】

$$U_{k,n} \leftarrow U_{k,n} \frac{\sum_m V_{m,n} H_{m,k}}{\sum_m H_{m,k} \sum_{k'} H_{m,k'} U_{k',n}} \quad (13)$$

30

【0094】

式(12)と式(13)による基底スペクトルと基底アクティベートの更新を、値が収束するまで繰り返すことにより、最適化、つまり、目的音モデル変数の学習を行う。

【0095】

この処理の結果、上記のように更新された目的音基底スペクトル行列Hが目的音モデル205に出力される。また、作成したスペクトログラムと基底スペクトル行列H、基底アクティベート行列Uは次フレームにおけるNMF処理の初期値として用いるために、無雑音信号DB203に格納される。このようにすることで、無雑音信号DB203に保存される無雑音信号が増えるほど、基底スペクトル行列Hをより目的音信号に忠実に学習させることができる。

40

【0096】

ステップS107で、風雑音抑制器206において、チャンネル毎に収音信号に対する風雑音抑制を行う。これは、図2のステップS7と同様な手法を用いて、チャンネル毎に行う。

【0097】

ステップS108で、目的音復元器207において、目的音モデル205に格納された基底スペクトルを変化させずに最適化を行う。まず、各チャンネルの収音信号を、 $M \times T$ のスペクトログラム行列 V_{ch} に変換する。ここで、 T は収音信号の当該処理フレームの時間サンプル数である。次に、式(13)の V を V_{ch} 、 n を t に各々置き換えた計算式を用

50

いて、基底アクティベートのみを値が収束するまで繰り返し計算する。

【 0 0 9 8 】

このようにして、各チャンネルの受信信号に対する $K \times T$ の大きさの基底アクティベート行列 U_{ch} を計算する。また、同時に、計算した基底アクティベートと基底スペクトルを用いて、各チャンネルの目的音復元信号 S_{ch} を生成する。これは、以下の式 (1 4) によって計算する。

【 0 0 9 9 】

【 数 1 4 】

$$S_{ch} = H U_{ch} \quad (14)$$

10

【 0 1 0 0 】

基底アクティベートと目的音復元信号は、信号選択・混合器 2 0 8 に出力される。

【 0 1 0 1 】

ステップ S 1 0 9 からステップ S 1 1 6 までの処理は、受信信号の全てのチャンネルに対して、個別の処理を繰り返して行う。

【 0 1 0 2 】

ステップ S 1 0 9 で、信号選択・混合器 2 0 8 において、処理対象となる次のチャンネルを選択する。処理対象のチャンネルは、受信信号の 1_{ch} から L_{ch} まで順に選択する。

20

【 0 1 0 3 】

ステップ S 1 1 0 で、処理対象のチャンネルに対応する受信信号に対して、ステップ S 1 0 8 で計算した基底アクティベートの処理フレーム全体の基底アクティベート平均値 (係数の大きさ) を計算する。

【 0 1 0 4 】

基底スペクトル k の t 番目の時間サンプルにおける基底アクティベートの振幅を $A_{k,t}$ 、スペクトル基底の数を K 、フレームの時間サンプル数を T とすると、基底アクティベート平均値 は以下の式 (1 5) で計算する。

【 0 1 0 5 】

【 数 1 5 】

$$\alpha = \frac{\sum_{k=1}^K \sum_{t=1}^T |A_{k,t}|}{K \cdot T} \quad (15)$$

30

【 0 1 0 6 】

ステップ S 1 1 1 で、信号選択・混合器 2 0 8 において、ステップ S 1 1 0 で計算した目的音モデル変数の基底アクティベート平均値 の値を確認し、予め定められた閾値、 A 、 B と比較する。尚、 $A > B$ である。

【 0 1 0 7 】

ステップ S 1 1 1 における比較の結果、 A となる場合は、信号選択・混合器 2 0 8 において、ステップ S 1 0 8 で得られた目的音復元信号が実際の目的音とほぼ等しいと判定し、ステップ S 1 1 2 へ進む。

40

【 0 1 0 8 】

また、ステップ S 1 1 1 における比較の結果、 $B < A$ となる場合は、信号選択・混合器 2 0 8 において、ステップ S 1 0 8 で得られた目的音復元信号には実際の目的音がある程度含まれていると判定し、ステップ S 1 1 3 へ進む。

【 0 1 0 9 】

また、ステップ S 1 1 1 における比較の結果、 $< B$ となる場合は、信号選択・混合器 2 0 8 において、ステップ S 1 0 8 で得られた目的音復元信号には実際の目的音はほぼ含

50

まれていないと判定し、ステップS 1 1 5へ進む。

【0 1 1 0】

ステップS 1 1 2からステップS 1 1 5までの処理は、実施形態1における図2のステップS 1 0からステップS 1 3までの処理と同様であるので、説明を省略する。これらの処理を終えると、ステップS 1 1 6へ進む。

【0 1 1 1】

ステップS 1 1 6で、全てのチャンネルに対して、信号選択・混合処理が終了したか否かを判定する。全てのチャンネルに対する処理が終了していない場合（ステップS 1 1 6でNO）、ステップS 1 0 9へ戻る。一方、全てのチャンネルに対する処理が終了した場合（ステップS 1 1 6でYES）、ステップS 1 1 7へ進む。

10

【0 1 1 2】

ステップS 1 0 9からステップS 1 1 6の処理を実行することによって、收音信号の各チャンネル毎に、基底スペクトルの活性度に応じて、目的音復元信号の確からしさを判定し、それによって目的音復元信号と雑音抑制信号の選択、混合を決定することができる。このようにすることで、雑音によって失われる目的音成分を補完しつつ、不完全な学習モデルによる不完全な目的音復元信号が混入することを避けることが可能になるため、より正確な目的音信号を取り出すことができる。

【0 1 1 3】

ステップS 1 1 7で、收音処理を終了する制御部（不図示）による指示があるか否かを判定する。指示がない場合（ステップS 1 1 7でNO）、ステップS 1 0 1へ戻る。一方、指示がある場合（ステップS 1 1 7でYES）、收音処理を終了する。

20

【0 1 1 4】

以上説明したように、実施形態2によれば、無雑音区間における入力信号から目的音の特性を学習し、雑音抑制で失われる目的音成分を学習した目的音モデルによって復元する。また、目的音モデルと入力信号による目的音モデルの活性度に応じて雑音抑制信号を補正する。これによって、風雑音を抑制しつつ、音色変化や目的音成分の欠落を防止することができる。

【0 1 1 5】

尚、実施形態2では、図4のステップS 1 0 4において、各チャンネルの時間振幅絶対値の平均が予め定められた閾値以下のチャンネルの推定雑音信号を、それぞれ無雑音信号としているが、その他の雑音の性質に基づいて判定することもできる。例えば、風雑音はマイクユニット毎に独立して生じる現象によって生じるため、チャンネル間の相関性を持たない。この性質を利用して、各チャンネル間の相関を調べ、他のチャンネルとの相関度が一つでも予め定められた閾値より大きい場合、無雑音信号として判定することができる。

30

【0 1 1 6】

<実施形態3>

実施形態3では、NMFによって目的音を復元する場合に、基底スペクトルの高域をキーにしてマッチングを行うことによって、処理量を抑えつつマッチング時の風雑音の影響を抑える構成について説明する。また、実施形態3では、風雑音の影響を受ける低域のみを補正することによって、より正確な目的音を得る場合について説明する。

40

【0 1 1 7】

図5（a）は、実施形態3の收音装置の構成を示すブロック図である。

【0 1 1 8】

図5（a）において、1から3と、201から206までの構成は、実施形態2における図3と同一であるため、説明を省略する。

【0 1 1 9】

301は風雑音スペクトル分布計算器であり、風雑音推定器201によって出力されたLch分の推定雑音信号に対して、チャンネル毎に周波数成分に変換する。そして、風雑音スペクトル分布計算器301は、各周波数成分のチャンネル平均を取ることにによって、

50

L c h 分の推定雑音信号全体のスペクトル分布を計算して出力する。

【 0 1 2 0 】

3 0 2 は分割周波数決定器であり、風雑音スペクトル分布計算器 3 0 1 によって出力されたスペクトル分布に基づいて、收音信号を低域と高域に分割する周波数を決定する。ここで、風雑音のスペクトルエネルギーは低域に偏っている。そのため、分割周波数決定器 3 0 2 は、低域から高域にかけて急激にスペクトルエネルギーが減衰し、かつ、それより高域には大きなエネルギーが存在しない周波数を探索し、それを分割周波数として出力する。

【 0 1 2 1 】

3 0 3 は目的音復元器であり、L c h の收音信号の各チャンネル信号に対して、分割周波数より高域のスペクトル基底を用いて N M F 処理を行い、各チャンネルに対する基底アクティベートを計算する。また、目的音復元器 3 0 3 は、計算した基底アクティベートと低域の基底スペクトルを用いて、目的音低域復元信号を生成して出力する。尚、3 0 3 の詳細構成は図 5 (b) を用いて後述する。

10

【 0 1 2 2 】

3 0 4 は信号選択・混合器であり、風雑音抑制器 2 0 6 から出力される L c h 分の雑音抑制後信号の低域成分と、目的音復元器 3 0 3 から出力される L c h 分の目的音低域復元信号（低域成分の目的音復元信号）を、チャンネル毎に選択・混合して出力する。尚、選択・混合の判断は、分割周波数決定器 3 0 2 から出力される分割周波数に基づいて行う。

【 0 1 2 3 】

20

図 5 (b) は、目的音復元器 3 0 3 の詳細構成を示すブロック図である。

【 0 1 2 4 】

図 5 (b) において、3 1 1 は基底スペクトル分割器であり、分割周波数決定器 3 0 2 が出力する分割周波数に従って、目的音モデル 2 0 5 に格納されている基底スペクトルを低域、高域に分割して出力する。

【 0 1 2 5 】

3 1 2 は高域スペクトログラム生成器であり、L c h 分の收音信号の各チャンネル信号に対して、短時間フーリエ変換を行い、時間周波数情報であるスペクトログラムを生成する。さらに、分割周波数決定器 3 0 2 が出力する分割周波数に基づき、收音信号において雑音の影響を受けていない分割周波数以上の高周波成分を抜き出して出力する。

30

【 0 1 2 6 】

3 1 3 は制限付 N M F であり、基底スペクトル分割器 3 1 1 が出力する高域基底スペクトルを変化させずに、L c h 分の收音信号の高域成分を N M F によって分解することで、L c h 分の基底アクティベートを計算する。

【 0 1 2 7 】

3 1 4 は目的音復元信号生成器であり、基底スペクトル分割器 3 1 1 が出力する低域基底スペクトルと、制限付 N M F 3 1 3 が出力する L c h 分の基底アクティベートの行列積を取ることににより、L c h 分の目的音低域復元信号を生成して出力する。

【 0 1 2 8 】

以下、図 5 の構成において、N M F による目的音復元処理時に、雑音の影響を受けていない高域において基底アクティベートを計算することで正確に目的音信号を復元し、かつ、雑音の影響を受けている目的音信号の低域を基底アクティベートによって復元して補正することにより、風雑音抑制後の信号をより正確に補正する一連の動作をフローに従って説明する。

40

【 0 1 2 9 】

図 6 は、実施形態 3 の收音装置が実行する收音処理を示すフローチャートである。

【 0 1 3 0 】

ステップ S 2 0 1 からステップ S 2 0 7 までの処理は、実施形態 2 の図 4 におけるステップ S 1 0 1 からステップ S 1 0 7 までの処理と同一であるため説明を省略する。

【 0 1 3 1 】

50

ステップS 2 0 8で、風雑音スペクトル分布計算器3 0 1において、風雑音推定器2 0 1によって出力したL c h分の推定雑音信号に対して、チャンネル毎に時間周波数変換処理（FFT等）を行って周波数成分に変換する。次に、風雑音スペクトル分布計算器3 0 1において、各周波数成分の振幅絶対値のチャンネル平均を取ることによって、L c h分の推定雑音信号全体のスペクトル分布を計算して出力する。このような処理は当分野において公知であるので詳細説明はしない。

【0 1 3 2】

ステップS 2 0 9で、分割周波数決定器3 0 2において、ステップS 2 0 8で計算した風雑音スペクトル分布を解析し、風雑音成分の大部分が集中する低周波数域と、風雑音成分があまり存在しない高周波数域とを分割する分割周波数を決定する。これは、例えば、風雑音スペクトル分布において、振幅が急激に減衰する変化点となる周波数を探索し、変化点から高域の全ての周波数振幅の平均が、ピーク振幅を基準として、予め定められた閾値以下のdB差となる最低周波数を分割周波数とする。

10

【0 1 3 3】

ステップS 2 1 0で、基底スペクトル分割器3 1 1において、目的音モデル2 0 5に格納されている基底スペクトルをステップS 2 0 9で決定した分割周波数に基づいて低域と高域に分割する。実施形態3における基底スペクトルは行列で表現されている。この行列において、各行は特定の周波数成分を示し、周波数順にソートされている。また、各列が個別の基底スペクトルを表現している。よって、この分割は、分割周波数前後の行となる部分で、行列を上下に分割することによってなされる。

20

【0 1 3 4】

ステップS 2 1 1で、高域スペクトログラム生成器3 1 2において、L c h分の收音信号の高域スペクトログラムを生成する。この処理の詳細は、高域スペクトログラム生成器3 1 2の説明において前述しているので省略する。

【0 1 3 5】

ステップS 2 1 2で、制限付NMF3 1 3において、ステップS 2 1 1で生成したL c h分の高域スペクトログラムを、ステップS 2 1 0で分割した高域基底スペクトルでNMFによる分解を行うことにより、L c h分の基底アクティベートを計算する。

【0 1 3 6】

ステップS 2 1 3で、目的音復元信号生成器3 1 4において、ステップS 2 1 0で分割した低域基底スペクトルと、ステップS 2 1 2で算出されたL c h分の基底アクティベートの行列積を計算することにより、L c h分の目的音低域復元信号を生成する。

30

【0 1 3 7】

ステップS 2 1 4からステップS 2 2 3までの処理は、実施形態2の図4と同様に、L c hの收音信号の全てのチャンネルに対して、個別の処理を繰り返して行う。

【0 1 3 8】

ステップS 2 1 4からステップS 2 1 6までの処理は、実施形態2の図4におけるステップS 1 0 9からステップS 1 1 1までの処理と同様であるため、説明を省略する。

【0 1 3 9】

ステップS 2 1 7で、信号選択・混合器3 0 4において、分割周波数決定器3 0 2が出力する分割周波数に基づき、ステップS 2 0 7で生成したL c h分の雑音抑制信号の低域成分を、ステップS 2 1 3で生成した対応するチャンネルの目的音低域復元信号に置換する。

40

【0 1 4 0】

ステップS 2 1 8の処理は、実施形態2の図4におけるステップS 1 1 3と同様であるため説明を省略する。

【0 1 4 1】

ステップS 2 1 9で、信号選択・混合器3 0 4において、ステップS 2 0 7で生成したL c h分の雑音抑制信号の各チャンネルに対して、分割周波数以下の低域成分を取り出す。

50

【 0 1 4 2 】

ステップS 2 2 0で、信号選択・混合器3 0 4において、ステップS 2 1 9で取り出した雑音抑制信号の低域成分と、ステップS 2 1 3で生成した目的音低域復元信号を、ステップS 2 1 8で算出した混合率で混合する。

【 0 1 4 3 】

ステップS 2 2 1で、信号選択・混合器3 0 4において、雑音抑制信号の低域成分を、ステップS 2 2 0で生成した混合信号に置換する。このようにすることで、基底アクティベートに応じて目的音低域復元信号を雑音抑制信号に反映させることができるため、より正確な補正が可能になる。

【 0 1 4 4 】

ステップS 2 2 2からステップS 2 2 4までの処理は、実施形態2の図4におけるステップS 1 1 5からステップS 1 1 7までの処理と同様であるため、説明を省略する。

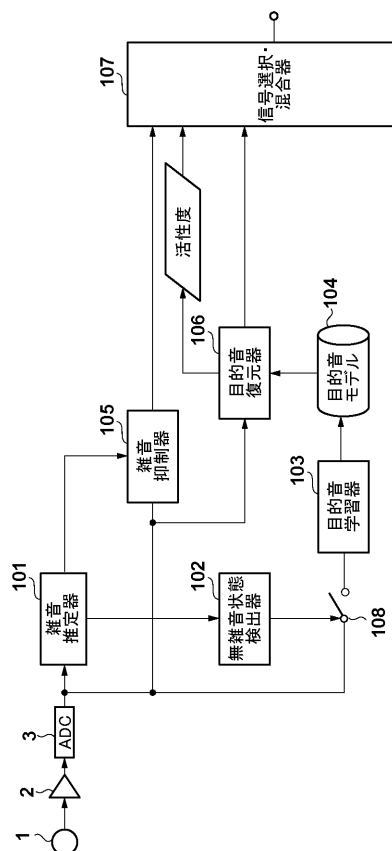
【 0 1 4 5 】

以上説明したように、実施形態3によれば、NMFによる目的音復元処理時に、雑音の影響を受けていない高域收音信号を分解することによって基底アクティベートを正確に計算する。また、低域基底スペクトルによって目的音信号の低域を復元する。これにより、風雑音抑制後の信号をより正確に復元することができる。

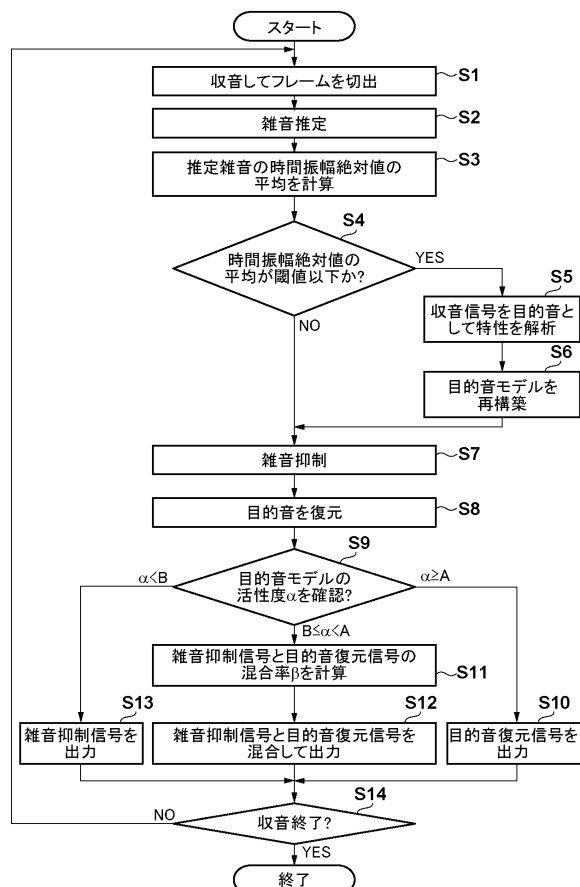
【 0 1 4 6 】

尚、本発明は、以下の処理を実行することによっても実現される。即ち、上述した実施形態の機能を実現するソフトウェア（プログラム）を、ネットワーク又は各種記憶媒体を介してシステムまたは装置に供給し、そのシステムまたは装置のコンピュータ（またはCPUやMPU等）がプログラムを読み出して実行する処理である。

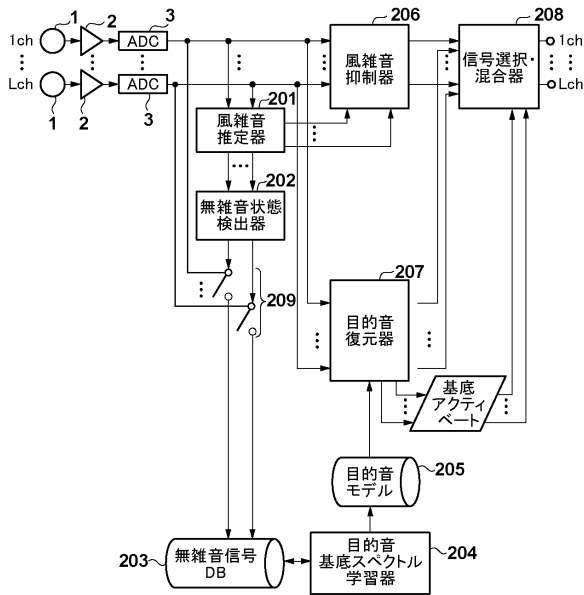
【 図 1 】



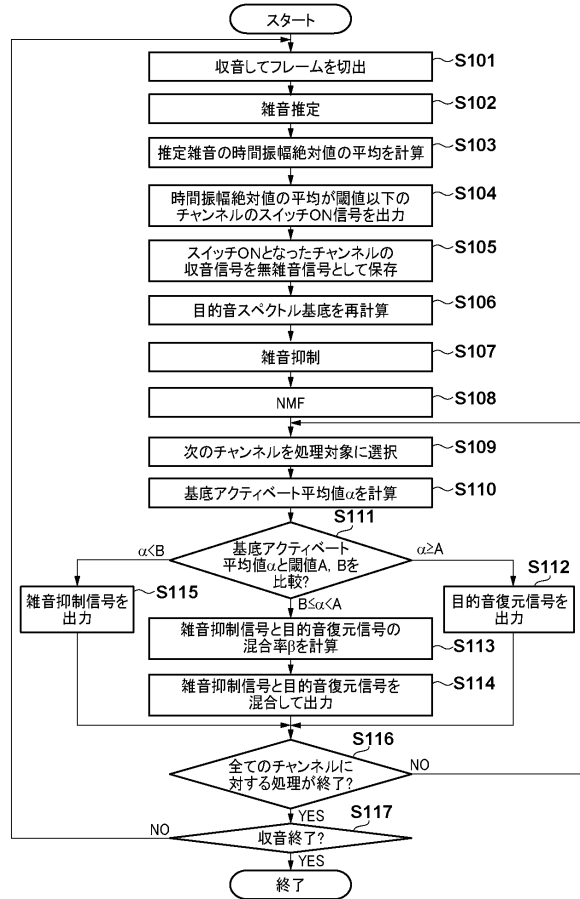
【 図 2 】



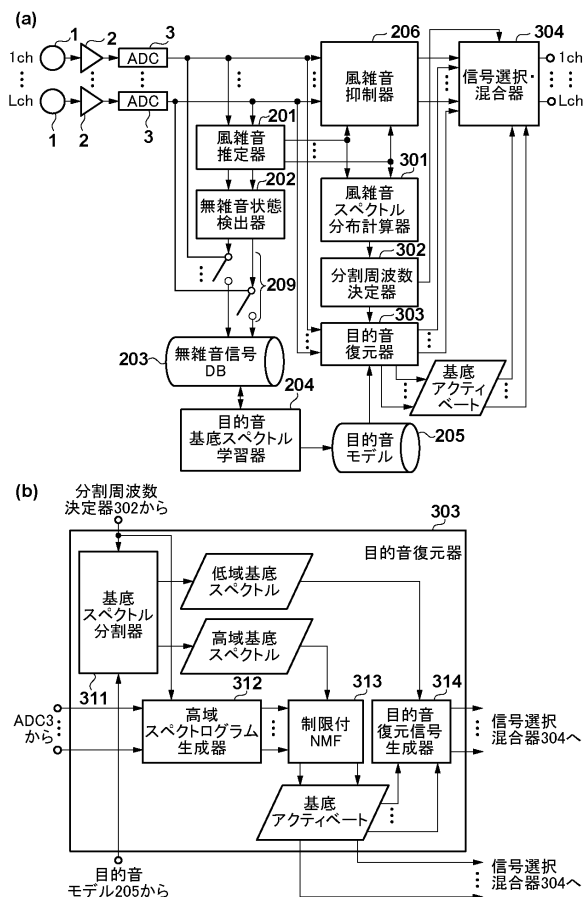
【図 3】



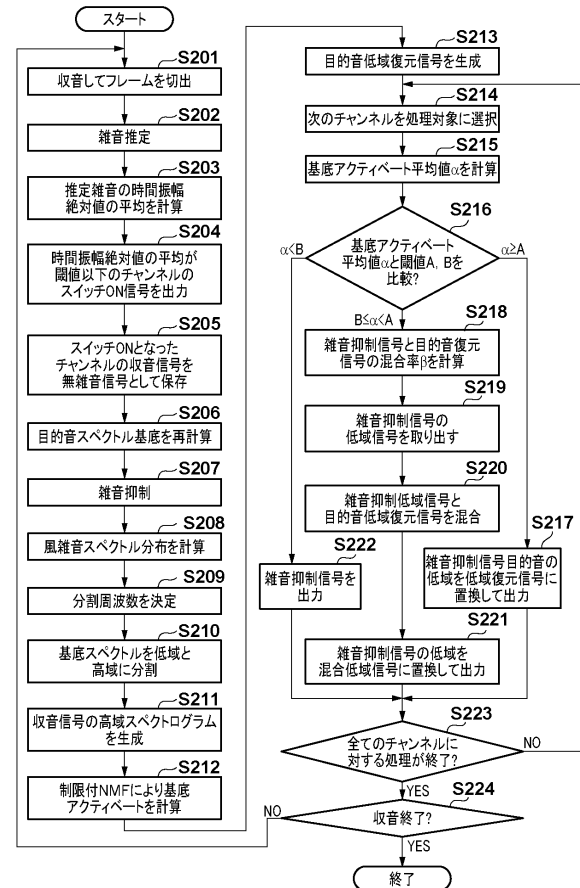
【図 4】



【図 5】



【図 6】



フロントページの続き

(72)発明者 船越 正伸
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 下林 義明

(56)参考文献 特開2003-241787(JP,A)
特開2009-128906(JP,A)

(58)調査した分野(Int.Cl., DB名)

H04R	3/00	-	3/14
H04R	1/20	-	1/40
G10L	13/00	-	13/10
G10L	19/00	-	99/00
G10L	15/00	-	17/26