(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第6371516号
(P6371516)

(45)発行日　平成30年8月8日(2018.8.8)　(24)登録日　平成30年7月20日(2018.7.20)

(51)Int.Cl.　F I
G10L 21/0308 (2013.01)　G10L 21/0308 Z
G10L 25/18 (2013.01)　G10L 25/18

請求項の数 15　(全 17 頁)

(21)出願番号　特願2013-237353(P2013-237353)
(22)出願日　平成25年11月15日(2013.11.15)
(65)公開番号　特開2015-96921(P2015-96921A)
(43)公開日　平成27年5月21日(2015.5.21)
審査請求日　平成28年11月15日(2016.11.15)

(73)特許権者　000001007
キヤノン株式会社
東京都大田区下丸子３丁目３０番２号
(74)代理人　100076428
弁理士　大塚　康徳
(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】音響信号処理装置および方法

(57)【特許請求の範囲】
【請求項１】
音響信号に非負値行列因子分解を適用することで、複数の基底スペクトルから構成される基底行列を取得する取得手段と、
前記取得手段により取得される前記基底行列を構成する基底スペクトルに対応する基底ケプストラムの、所定範囲のケフレンシーにおける最大値に基づいて、該基底スペクトルを第一の基底スペクトル群と第二の基底スペクトル群の何れかに分類する分類手段と、を備えることを特徴とする音響信号処理装置。
【請求項２】
前記取得手段により取得された前記複数の基底スペクトルを前記分類手段により分類した結果に基づいて、前記音響信号に含まれる雑音成分が抑制された目的音信号を生成する生成手段を有することを特徴とする請求項１に記載の音響信号処理装置。
【請求項３】
前記基底行列を構成する複数の基底スペクトルのうち、前記分類手段により前記第一の基底スペクトル群に分類される基底スペクトルの重心周波数は、前記分類手段により前記第二の基底スペクトル群に分類される基底スペクトルの重心周波数よりも高いことを特徴とする請求項１又は２に記載の音響信号処理装置。
【請求項４】
前記分類手段は、ケフレンシーが所定値以下である低ケフレンシー部分における該基底ケプストラムの最大値に基づいて、該基底スペクトルを前記第一の基底スペクトル群と前

記第二の基底スペクトル群の何れかに分類することを特徴とする請求項１又は２に記載の音響信号処理装置。
【請求項５】
前記分類手段は、前記低ケフレンシー部分における該基底ケプストラムの最大値と閾値との比較により、該基底スペクトルを前記第一の基底スペクトル群と前記第二の基底スペクトル群の何れかに分類することを特徴とする請求項４に記載の音響信号処理装置。
【請求項６】
音響信号に非負値行列因子分解を適用することで、複数の基底スペクトルから構成される基底行列を取得する取得手段と、
前記取得手段により取得される前記基底行列を構成する基底スペクトルに対応する基底ケプストラムの、所定の基本周波数範囲に対応する部分の値に基づいて、該基底スペクトルを、第一の基底スペクトル群と、前記第一の基底スペクトル群に分類される基底スペクトルよりも調波成分が大きい基底スペクトルが分類される第二の基底スペクトル群との何れかに分類する分類手段と、を備えることを特徴とする音響信号処理装置。
【請求項７】
前記分類手段は、該基底ケプストラムの前記所定の基本周波数範囲に対応する部分の最大値と閾値との比較により、該基底スペクトルを前記第一の基底スペクトル群と前記第二の基底スペクトル群の何れかに分類することを特徴とする請求項６に記載の音響信号処理装置。
【請求項８】
前記第一の基底スペクトル群へ分類される基底スペクトルの数である第一の基底数と、前記第二の基底スペクトル群へ分類される基底スペクトルの数である第二の基底数の少なくとも一方を調整する調整手段をさらに備える請求項１乃至７の何れか１項に記載の音響信号処理装置。
【請求項９】
前記分類手段により前記第一の基底スペクトル群に分類される基底スペクトルと、前記非負値行列因子分解の適用により取得されるアクティビティ行列を構成する複数のアクティビティベクトルのうちの前記第一の基底スペクトル群に対応するアクティビティベクトルとを用いて、第一の音響復元信号を生成する生成手段を更に備えることを特徴とする請求項１乃至８の何れか１項に記載の音響信号処理装置。
【請求項１０】
前記分類手段により前記第二の基底スペクトル群に分類される基底スペクトルと、前記非負値行列因子分解の適用により取得されるアクティビティ行列を構成する複数のアクティビティベクトルのうち前記第二の基底スペクトル群に対応するアクティビティベクトルとを用いて、第二の音響復元信号を生成する生成手段を更に備えることを特徴とする請求項１乃至８の何れか１項に記載の音響信号処理装置。
【請求項１１】
前記分類手段により前記第一の基底スペクトル群に分類される基底スペクトルと、前記非負値行列因子分解の適用により取得されるアクティビティ行列を構成する複数のアクティビティベクトルのうちの前記第一の基底スペクトル群に対応するアクティビティベクトルとを用いて、第一の音響復元信号を生成する第一生成手段と、
前記分類手段により前記第二の基底スペクトル群に分類される基底スペクトルと、前記アクティビティ行列を構成する複数のアクティビティベクトルのうち前記第二の基底スペクトル群に対応するアクティビティベクトルとを用いて、第二の音響復元信号を生成する第二生成手段をさらに備え、
前記第一生成手段により生成される前記第一の音響復元信号と前記第二生成手段により生成される前記第二の音響復元信号の少なくとも一方を用いて、前記音響信号から非目的音を除去することを特徴とする請求項１乃至８のいずれか１項に記載の音響信号処理装置。
【請求項１２】

前記取得手段は、前記音響信号に対して時間周波数変換を行うことで得られる行列を、前記非負値行列因子分解により基底行列とアクティビティ行列に分解することで、前記基底行列を取得することを特徴とする請求項１乃至１１のいずれか１項に記載の音響信号処理装置。

【請求項１３】

音響信号処理装置の制御方法であって、

音響信号に非負値行列因子分解を適用することで、複数の基底スペクトルから構成される基底行列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前記基底行列を構成する基底スペクトルに対応する基底ケプストラムの、所定範囲のケフレンシーにおける最大値に基づいて、該基底スペクトルを第一の基底スペクトル群と第二の基底スペクトル群の何れかに分類する分類工程と、

を有することを特徴とする音響信号処理装置の制御方法。

【請求項１４】

音響信号処理装置の制御方法であって、

音響信号に非負値行列因子分解を適用することで、複数の基底スペクトルから構成される基底行列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前記基底行列を構成する基底スペクトルに対応する基底ケプストラムの、所定の基本周波数範囲に対応する部分の値に基づいて、該基底スペクトルを、第一の基底スペクトル群と、前記第一の基底スペクトル群に分類される基底スペクトルよりも調波成分が大きい基底スペクトルが分類される第二の基底スペクトル群との何れかに分類する分類工程と、を有することを特徴とする音響信号処理装置の制御方法。

【請求項１５】

コンピュータを、請求項１乃至１２のいずれか１項に記載の音響信号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、音響信号をたとえば目的音と雑音といったように複数の音響信号に分離する音響信号処理装置および方法に関する。

【背景技術】

【０００２】

音響信号から非目的音である雑音を除去する技術は、音響信号に含まれる目的音に対する聴感を改善し、また、音声認識においては認識率を高めるために重要な技術である。

【０００３】

音響信号から雑音を除去する技術として、非負値行列因子分解を用いたものがある。これは、音響信号を時間周波数変換した行列を非負値行列因子分解によって基底行列とアクティビティ行列に分解するもので、これらの行列が目的音に係る部分行列と雑音に係る部分行列に分けられるという仮定に基づく。そして、目的音に係る部分基底行列である目的音基底行列と、目的音に係る部分アクティビティ行列である目的音アクティビティ行列を用いて、雑音が除去された目的音復元信号を生成する。

【０００４】

特許文献１では、雑音除去対象である音響信号とは別に目的音と雑音をそれぞれ用意し、それらを事前学習することで目的音と雑音それぞれの教師基底行列および教師アクティビティ行列を得る。そして教師基底行列および、教師アクティビティ行列の統計量情報を用い、音響信号を時間周波数変換した行列を分解して目的音復元信号を得る。

【０００５】

特許文献２では、２ｃｈの音響信号それぞれを時間周波数変換した２つの行列を非負値行列因子分解する。そして、各ｃｈの基底行列の各列を構成する基底スペクトルについて、ｃｈ間の相関が高いものを雑音基底スペクトル、それ以外のものを目的音基底スペクトルとする。そして、目的音基底スペクトルで構成される目的音基底行列と、それに対応す

る目的音アクティビティ行列を用いて目的音復元信号を生成する。
【先行技術文献】
【特許文献】
【０００６】
【特許文献１】特開２００９－１２８９０６号公報
【特許文献２】特開２０１２－０２２１２０号公報
【発明の概要】
【発明が解決しようとする課題】
【０００７】
特許文献１の方法では、別に用意した音から基底行列を事前学習し、これを用いて復元信号を生成することになる。これは、調波楽器の音階のように、基底スペクトルの形状（調波構造）が大体決まっている場合の楽器音分離（例えば自動採譜に用いる）などには好適であると考えられる。しかし、それ以外の場合は、分離対象の音響信号に含まれる音とは異なる基底スペクトルを用いて復元信号を生成する可能性があるため、音質劣化につながり得る。
【０００８】
特許文献２の方法は、雑音除去対象の音響信号から基底行列を得るため、目的音基底行列と雑音基底行列にうまく分けられれば、実際の目的音に対応する基底スペクトルを用いて目的音復元信号を生成できると期待される。しかしながら、目的音基底スペクトルと雑音基底スペクトルへの分類はｃｈ間の相関に基づいて行うため、複数ｃｈの音響信号を必要としている。
【０００９】
また、相関は２つの基底スペクトルの組み合わせに対して算出される量であるが、基底スペクトル間のユークリッド距離や内積を用いるとしている。しかし、このような単純な相関指標は物理的な意味が不明瞭であり、基底スペクトルの分類において必ずしも好適ではない。
【００１０】
本発明は上述した問題を解決するためになされたものであり、音響信号の各々の基底スペクトルを高精度に分類することが可能な音響信号処理装置および制御方法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】
【００１１】
上記目的を達成するための本発明の一態様による音響信号処理装置は以下の構成を備える。すなわち、
音響信号に非負値行列因子分解を適用することで、複数の基底スペクトルから構成される基底行列を取得する取得手段と、
前記取得手段により取得される前記基底行列を構成する基底スペクトルに対応する基底ケプストラムの、所定範囲のケフレンシーにおける最大値に基づいて、該基底スペクトルを第一の基底スペクトル群と第二の基底スペクトル群の何れかに分類する分類手段と、を備える。
【発明の効果】
【００１２】
本発明によれば、音響信号の各々の基底スペクトルから算出された指標を用いて基底スペクトルを分類するので、基底スペクトルを高精度に分類することができる。
【図面の簡単な説明】
【００１３】
【図１】実施形態に係る音源分離装置のブロック図。
【図２】音響信号および振幅スペクトログラムを説明するための図。
【図３】実施形態に係る音響分離処理のフローチャート。
【図４】第一実施形態に係る基底スペクトルの評価指標を説明する図。

【図５】第一実施形態に係る基底番号のソートおよび目的音復元信号のＳＮＲを説明する図。
【図６】実施形態に係る目的音復元信号を説明する図。
【図７】実施形態に係る基底スペクトルの分類を説明する図。
【図８】第二実施形態に係る基底スペクトルの評価指標を説明する図。
【図９】第二実施形態に係る基底番号のソートおよび目的音復元信号のＳＮＲを説明する図。
【図１０】第三実施形態に係る基底スペクトルの評価指標を説明する図。
【図１１】第三実施形態に係る基底番号のソートおよび目的音復元信号のＳＮＲを説明する図。

【発明を実施するための形態】

【００１４】
以下、添付の図面を参照して、本発明をその好適な実施形態に基づいて詳細に説明する。なお、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、以下に説明される構成に限定されるものではない。

【００１５】
＜第一実施形態＞
図１は、第一実施形態による音響信号処理装置としての音源分離装置の構成例を示すブロック図である。図１に示す音源分離装置は、主たるシステムコントローラ１００の中に、全構成要素の制御を行うシステム制御部１０１、各種データを記憶しておく記憶部１０２、信号の解析処理を行う信号解析処理部１０３を備える。また、システムコントローラ１００は、音響信号を入出力するための音響信号入出力部１０４を備える。

【００１６】
図１に示す音源分離装置は、音響信号入出力部１０４を介して、例えば外部の記憶媒体やネットワークから音響信号が入力され、記憶部１０２へ記録される。ここで、音響信号とは、目的音に除去対象となる雑音が混合した混合音を指すものとする。なお、たとえば、不図示のマイクによって収音されたマイク信号に、増幅およびＡＤ変換が施されたものが音響信号として入力され、記憶部１０２へ逐次記録されるようにしてもよい。

【００１７】
第一実施形態においては、目的音を図２（ａ）に示すような音声、雑音を図２（ｂ）に示すような風雑音とし、これらが混合した図２（ｃ）に示すような混合音を雑音除去の対象である音響信号とする。以下、信号解析処理部１０３が中心となり、図３のフローチャートに沿って雑音除去処理が行われる。なお、図３のフローチャートは、音響信号の所定の時間ブロック長ごとの処理を表すものとし、第一実施形態では時間ブロック長を３秒としている。

【００１８】
Ｓ３０１において、信号解析処理部１０３は、音響信号入出力部１０４より入力され記憶部１０２に記憶されている音響信号から時間ブロック長の音響信号を得て、これを時間周波数変換して音響信号の複素スペクトログラムＹを取得する。具体的には、時間ブロック長より短い所定の時間フレーム長ずつ音響信号を切り出して行き、フーリエ変換することで複素フーリエ係数を得る。このとき、切り出される時間フレームは時間フレーム長の半分ずつシフトして行くものとし、時間フレーム長によって時間周波数変換における時間解像度および周波数解像度が決まる。時間ブロックに含まれる時間フレーム数をＴ、ナイキスト周波数までの周波数分割数をＦとすると、上記の処理によってＦ×Ｔ個の複素フーリエ係数が得られる。複素スペクトログラムＹは、これらのフーリエ係数を要素とするサイズ｛Ｆ×Ｔ｝の複素行列である。なお、フーリエ変換の前に時間信号に対して窓掛けを行うのが好適であり、窓掛けは逆フーリエ変換によって再び時間信号に戻した後にも行う。このため、５０％ずつオーバーラップする時間フレームに対し、２回の窓掛けにおける再構成条件を考慮して、窓関数にはサイン窓などを用いる。

【００１９】

次にＳ３０２において、信号解析処理部１０３は、音響信号の振幅スペクトログラム｜Ｙ｜を、基底行列Ｈとアクティビティ行列Ｕに非負値行列因子分解する。ここで、振幅スペクトログラム｜Ｙ｜とは、Ｓ３０１で取得した複素スペクトログラムＹについて、要素ごとに複素数の絶対値を取った非負値行列である。

【００２０】

図２（ｄ）は、図２（ｃ）の音響信号の振幅スペクトログラムを表している。ただし、表示においては振幅値を二値化しており、白が大きい方を、黒が小さい方を表す。図２（ｄ）より、低域では風雑音が卓越しており、また中高域では音声の調波成分による縞模様が見えていることが分かる。

【００２１】

非負値行列因子分解に際して指定する基底数をＫとすると、サイズ｛Ｆ×Ｔ｝の振幅スペクトログラム｜Ｙ｜を、式（１）のようにサイズ｛Ｆ×Ｋ｝の基底行列Ｈとサイズ｛Ｋ×Ｔ｝のアクティビティ行列Ｕに分解できる。ここで、「＊」は行列（ベクトル、スカラ含む）の積を表すものとする。

｜Ｙ｜　Ｈ＊Ｕ　…　（１）

なお、式（１）の収束計算におけるＨとＵの更新式は、（Ｈ＊Ｕ）の｜Ｙ｜からの乖離度を表す規準に応じたものを用いればよい。行列の乖離度を表す規準としては、ユークリッド距離（二乗誤差）、一般化Kullback-Leiblerダイバージェンス、板倉斎藤距離などが挙げられる。

【００２２】

図２（ｄ）の振幅スペクトログラム｜Ｙ｜を例えば基底数Ｋ=２０で非負値行列因子分解すると、サイズ｛Ｆ×２０｝の基底行列Ｈとサイズ｛２０×Ｔ｝のアクティビティ行列Ｕが得られる。図４（ｂ）は、基底行列Ｈの各列を構成する、サイズ｛Ｆ×１｝の２０個の基底スペクトルを正規化して表示したものであり、図の縦軸の番号は基底番号を表している。なお、人の聴感に合わせて周波数軸は対数で、振幅はデシベルで表示している。また、図４（ｃ）は、アクティビティ行列Ｕの各行を構成する、サイズ｛１×Ｔ｝の２０個のアクティビティベクトルを正規化して表示したものである。

【００２３】

ここで、同じ基底番号の基底スペクトルとアクティビティベクトルを掛け合わせることで、サイズ｛Ｆ×Ｔ｝の基底別振幅スペクトログラム｜Ｙｉ｜が式（２）のように得られる。ただし、（：，ｉ）は行列の第ｉ列を取り出す操作を、（ｉ，：）は行列の第ｉ行を取り出す操作を表すものとする。

｜Ｙｉ｜＝Ｈ（：，ｉ）＊Ｕ（ｉ，：）　［ｉ＝１～Ｋ］　…　（２）

【００２４】

さらに、基底別振幅スペクトログラム｜Ｙｉ｜に複素スペクトログラムＹの位相成分を掛け合わせることで、サイズ｛Ｆ×Ｔ｝の基底別複素スペクトログラムＹｉが式（３）のように得られる。ここで、「．＊」は行列の要素ごとの積を、ｊは虚数単位を表すものとする。また、ａｒｇ（Ｙ）はＹの要素ごとに複素数の偏角を取った行列を表すものとする。

Ｙｉ＝｜Ｙｉ｜．＊ｅｘｐ（ｊ＊ａｒｇ（Ｙ））　［ｉ＝１～Ｋ］　…　（３）

【００２５】

そして、基底別複素スペクトログラムＹｉを逆時間周波数変換することで、基底別復元信号ｙｉ［ｉ=１～Ｋ］を生成することができる。具体的には、複素スペクトログラムをサンプリング周波数まで対称復元した後、列ごとに逆フーリエ変換することで各時間フレームの復元信号が得られるため、これを窓掛けしてオーバーラップ加算すればよい。

【００２６】

図４（ｄ）は、２０個の基底別復元信号を正規化して表示したものである。図４（ｄ）の一番上に示す目的音の波形と見比べてみると、例えば基底番号４番や１６番は、目的音と雑音のうち目的音に係るものであると予想できる。このように、例えば基底別復元信号を人が見たり聴いたりすれば、各基底を目的音と雑音のものに高精度に分類できそうだが

、本実施形態では以下のように、各々の基底スペクトルから物理的意味が明確な指標を算出することで自動的に行う。

【００２７】

Ｓ３０３において、信号解析処理部１０３は、各々の基底スペクトルから算出する評価指標に基づいて、基底番号のソートを行う。具体的には、風雑音の基底スペクトルなら低域が卓越しており、音声の基底スペクトルなら中高域含む広い周波数範囲に分布しているであろうという考え方に基づく。そこで、このような基底スペクトルの周波数軸上での分布状態を数値化するため、第一実施形態では各々の基底スペクトルの評価指標として、各基底スペクトルの周波数分布における重心周波数を算出する。

【００２８】

まず、図４（ｂ）に示すような基底スペクトルの表現を得るため、基底スペクトルをデシベル表現する。ただし、微小な値がデシベル表現によって負の大きな値となると不都合であるため、基底行列の最大値－６０ｄＢ未満の値は、最大値－６０ｄＢに丸めるなどする。そして、例えば０から１の範囲内に正規化した後、対数周波数軸上で等間隔に振幅値を得るためオクターブスムージングを行う。

【００２９】

このような表現を行った基底スペクトルをｈ、対数周波数軸上の対象周波数範囲（例えば５０～３ｋＨｚ）における等間隔なサンプル点の番号をｓ（＝０～）とすると、重心周波数に対応するサンプル点番号ｓｇを式（４）のように算出することができる。ただし、ｈ（ｓ）は基底スペクトルｈのサンプル点番号ｓにおける値を表し、Σはｓに関して和を取る操作を表すものとする。

ｓｇ＝Σ（ｓ＊ｈ（ｓ））／Σ（ｈ（ｓ））　…　（４）

式（４）で算出されるｓｇは一般に小数値であり、これを対数周波数軸上に対応付けた値が重心周波数となる。

【００３０】

図４（ｂ）の各々の基底スペクトルについて、上記のようにして求めた重心周波数を模式的に示したのが図４（ｂ）の黒丸であり、また重心周波数の値を棒グラフで示したのが図４（ａ）である。これらの図より、目的音に係ると予想した基底番号４番や１６番の基底スペクトルは広い周波数範囲に分布しており、それゆえ重心周波数も他と比べて高くなっていることが分かる。

【００３１】

図５（ａ）は、図４（ａ）の重心周波数に従って基底スペクトルを重心周波数の昇順にソートしたものであり、横軸がソートされた基底番号を表している。ここで、左の基底番号ほど、基底スペクトルの重心周波数が低いため、低域が卓越した風雑音である可能性が高い。また、右の基底番号ほど重心周波数が高いため、基底スペクトルが広い周波数範囲に分布した音声である可能性が高いと考えられる。

【００３２】

Ｓ３０４において、信号解析処理部１０３は、基底スペクトルを第一の基底スペクトルとしての目的音基底スペクトルと、第二の基底スペクトルとしての非目的音基底スペクトル（雑音基底スペクトルともいう）とに分類する。そして、分類された基底スペクトルを用いて音響復元信号を生成する。はじめに、信号解析処理部１０３は、Ｓ３０３でソートされた基底番号に基づいて、基底行列Ｈの各列（基底スペクトル）を並べ替える。すなわち、図５（ａ）に示されるソート結果に従って元の基底行列の第１５列を並べ替え後の基底行列の第１列とし、元の基底行列の第１２列を並べ替え後の基底行列の第２列とする、といった具合に２０個の基底スペクトルを並べ替える。また、アクティビティ行列Ｕの各行（アクティビティベクトル）も同様に並べ替える。

【００３３】

このように、ソートされた基底番号に基づいて基底行列を並べ替えれば、後は目的音基底数または雑音基底数を定めることで、基底スペクトルを目的音基底スペクトルと雑音基底スペクトルに分類することができる。すなわち、雑音基底数をＫｎとすれば、並べ替え

られた基底行列Ｈの第１列から第Ｋｎ列までの基底スペクトルが雑音基底スペクトルとして、第Ｋｎ＋１列から第Ｋ列までの基底スペクトルが目的音基底スペクトルとして分類される。そして、雑音基底スペクトルで構成される雑音基底行列Ｈｎと、目的音基底スペクトルで構成される目的音基底行列Ｈｓが、それぞれ式（５）と式（６）のように得られる。ただし、（：，１：Ｋｎ）は行列の第１列から第Ｋｎ列を取り出す操作を、（：，Ｋｎ＋１：Ｋ）は行列の第Ｋｎ＋１列から第Ｋ列を取り出す操作を表すものとする。

Ｈｎ＝Ｈ（：，１：Ｋｎ）　　…　（５）

Ｈｓ＝Ｈ（：，Ｋｎ＋１：Ｋ）　　…　（６）

【００３４】

また、基底スペクトルと同様に、アクティビティベクトルも第一のアクティビティベクトルとしての目的音アクティビティベクトルと第二のアクティビティベクトルとしての雑音アクティビティベクトル（非目的音アクティビティベクトル）に分類できる。雑音アクティビティベクトルで構成される雑音アクティビティ行列Ｕｎと、目的音アクティビティベクトルで構成される目的音アクティビティ行列Ｕｓが、それぞれ式（７）と式（８）のように得られる。ただし、（１：Ｋｎ，：）は行列の第１行から第Ｋｎ行を取り出す操作を、（Ｋｎ＋１：Ｋ，：）は行列の第Ｋｎ＋１行から第Ｋ行を取り出す操作を表すものとする。

Ｕｎ＝Ｕ（１：Ｋｎ，：）　　…　（７）

Ｕｓ＝Ｕ（Ｋｎ＋１：Ｋ，：）　…　（８）

【００３５】

目的音基底数をＫｓ（＝Ｋ－Ｋｎ）とすると、サイズ｛Ｆ×Ｋｓ｝の目的音基底行列Ｈｓと、サイズ｛Ｋｓ×Ｔ｝の目的音アクティビティ行列Ｕｓを掛け合わせることで、サイズ｛Ｆ×Ｔ｝の目的音振幅スペクトログラム｜Ｙｓ｜が式（９）のように得られる。

｜Ｙｓ｜＝Ｈｓ＊Ｕｓ　…　（９）

【００３６】

さらに、複素スペクトログラムＹの位相成分を掛け合わせることで、サイズ｛Ｆ×Ｔ｝の目的音複素スペクトログラムＹｓが式（１０）のように得られる。

Ｙｓ＝｜Ｙｓ｜．＊ｅｘｐ（ｊ＊ａｒｇ（Ｙ））　…　（１０）

【００３７】

そして、目的音複素スペクトログラムＹｓを逆時間周波数変換することで、音響復元信号として目的音復元信号ｙｓを生成することができる。なお、サイズ｛Ｆ×Ｋｎ｝の雑音基底行列Ｈｎと、サイズ｛Ｋｎ×Ｔ｝の雑音アクティビティ行列Ｕｎを用いて、音響復元信号としての雑音復元信号ｙｎも同様に生成することができる。

【００３８】

図６（ａ）は、雑音基底数を０から２０まで増やして行ったときの、それぞれの目的音復元信号を示したものである。図６（ａ）より、雑音基底数Ｋｎを多くするにつれて、言い換えれば目的音基底数Ｋｓを絞って行くにつれて、風雑音が除去されて音声が復元されて行く様子が分かる。また図５（ｂ）は、雑音基底数と目的音復元信号のＳＮＲの関係を示したグラフであり、雑音基底数Ｋｎ＝１７（目的音基底数Ｋｓ＝３）のとき、ＳＮＲが最大の２．２１ｄＢとなる。このとき図５（ａ）より、基底番号１６番、４番、７番が用いられていることが分かる。しかしながら図５（ｂ）より、さらに雑音基底数を増やして目的音基底数を絞って行くと、ＳＮＲが低下してしまう。このため、目的音基底数または雑音基底数を適切に定めることが大切と考えられる。

【００３９】

目的音基底数の決め方としては、重心周波数は［Ｈｚ］の単位を持つ物理的意味が明らかな指標であるため、例えば閾値を２００Ｈｚとして、重心周波数が閾値以上の基底スペクトルの数を目的音基底数としてもよい。図７（ｃ）は、図４（ａ）をヒストグラム化して重心周波数の分布を表したものであり、閾値を２００Ｈｚとすると図７（ｃ）の実線で分けることになるため、目的音基底数は３となる。また、図７（ｃ）のヒストグラムを混合正規分布と見なして、ＥＭアルゴリズムにより２群に分類することで目的音基底数を決

定してもよい。

【００４０】

他にも、音響信号とは別に音声と風雑音を用意し、それぞれから求めた重心周波数のヒストグラムを利用してもよい。例えば、図７（ｂ）に示す風雑音のみの音響信号から得られたヒストグラムの範囲を考慮して、図７の実線で図７（ｃ）を分けると目的音基底数は３となる。または、図７（ａ）に示す音声のみの音響信号から得られたヒストグラムの範囲を考慮して、図７の点線で図７（ｃ）を分けると目的音基底数は４となる。この方法は、特許文献１の事前学習に似ているようにも見えるが、別に用意する目的音または雑音は、目的音基底数または雑音基底数の決定に用いるだけであり、基底行列は音響信号から得るため特許文献１とは異なる。

【００４１】

なお、システム制御部１０１と相互に結ばれた不図示の入出力ＧＵＩ部（例えばタッチパネルで構成される）を介して、ユーザが目的音基底数または雑音基底数を調整できるようにしてもよい。

【００４２】

以上のようにして、目的音基底数または雑音基底数を適切に定めて生成された目的音復元信号は、記憶部１０２へ記録される。記録された目的音復元信号は、音響信号入出力部１０４を介して外部に出力されたり、ＤＡ変換および増幅を行ったのち、不図示のイヤホン、ヘッドホン、スピーカ等によって再生されたりする。

【００４３】

＜第二実施形態＞

第一実施形態では、Ｓ３０３において、各々の基底スペクトルから重心周波数という評価指標を算出したが、評価指標の算出はこれに限られるものではない。第二実施形態においては、基底スペクトルを変換することで得られるケプストラム（以下、基底ケプストラムと呼ぶ）から評価指標を算出する例を説明する。

【００４４】

基底ケプストラムは、サンプリング周波数まで対称復元した基底スペクトルに対し、対数を取って逆フーリエ変換した結果の実部として得られる。図８（ｃ）は、図８（ｂ）の各々の基底スペクトルから求めた基底ケプストラムを正規化して表示したものであり、横軸は時間の次元を持つケフレンシーとなる。なお、図８（ｂ）、（ｄ）は、それぞれ図４（ｂ）、（ｄ）と同じものである。

【００４５】

図８（ｃ）において、基底番号４番や１６番の基底ケプストラムは、図８（ｂ）の点線丸で示すように低ケフレンシー部分が一際大きくなっている。ここで、一般にケプストラムの低ケフレンシー部分はスペクトルの包絡成分に対応するが、スペクトルの包絡成分の大きさは、スペクトルの広がり状態を表していると考えることもできる。実際に、基底ケプストラムにおいて低ケフレンシー部分に大きい値を持つ基底番号４番や１６番では、図８（ｂ）の点線丸で示すように、基底スペクトルが広い周波数範囲に分布していることが確認できる。なお、低ケフレンシー部分とは、ケフレンシーが所定値以下の部分であり、本実施形態では、たとえば２ｍｓ以下の部分とする。

【００４６】

そこで、第二実施形態では、Ｓ３０３において基底スペクトルの周波数軸上での分布状態を数値化するための評価指標を、各々の基底ケプストラムから算出する。例えば、基底スペクトルの広がり状態を表す包絡成分の大きさを数値化するため、基底ケプストラムの所定のケフレンシー以下の部分について、その最大値を評価指標とする。より簡単には、図８（ｃ）より基底ケプストラムの低ケフレンシー部分の最大値は、基底ケプストラム全体の最大値であるとして差し支えなさそうであるため、これを評価指標として用いてもよい。

【００４７】

図８（ｃ）の各々の基底ケプストラムについて、基底ケプストラム全体の最大値を棒グ

ラフで示したのが図８（ａ）であり、これを昇順にソートしたものが図９（ａ）である。図８（ａ）において、棒グラフが右へ延びる基底番号ほど基底スペクトルの包絡成分が大きいため、基底スペクトルが広い周波数範囲に分布した音声である可能性が高い。逆に、棒グラフが左側にとどまる基底番号ほど、基底スペクトルの包絡成分が小さいため、基底スペクトルが狭い周波数範囲（低域）に集中した風雑音である可能性が高いと考えられる。

【００４８】

ここで、第一実施形態と第二実施形態はともに、基底スペクトルの周波数軸上での分布状態を数値化するという考えに基づいている。このため、図５（ａ）と図９（ａ）の棒グラフは全体的に傾向が似ており、ソートされた基底番号の並び順も類似している。特に、上位４つ（１６番、４番、７番、２番）は同じである。そのため、第二実施形態における雑音基底数と目的音復元信号のＳＮＲの関係を示した図９（ｂ）は、図５（ｂ）と同じく雑音基底数が１７（目的音基底数が３）のときにＳＮＲが最大の２．２１ｄＢとなる。

【００４９】

なお、Ｓ３０４の目的音基底数または雑音基底数の決定において、図７の実線や点線のように評価指標の値で分ける場合、音響信号の大きさに評価指標の値が依存しないことが望ましい。第一実施形態の重心周波数はその性質上、音響信号の大きさには依存しないが、一般に基底スペクトルや基底ケプストラムの大きさは音響信号の大きさに依存する。そこで、式（１）の非負値行列因子分解において基底行列Ｈを正規化しておけば、基底スペクトルや基底ケプストラムが音響信号の大きさに依存しなくなり、基底ケプストラムから算出する評価指標も音響信号の大きさに依存しなくなるため好適である。

【００５０】

なお、図３のフローチャートにおいて上述したＳ３０３の処理以外は、第一実施形態と同様である。

【００５１】

＜第三実施形態＞

上述の第一実施形態、第二実施形態では、Ｓ３０３において、基底スペクトルの周波数軸上での分布状態を数値化した評価指標を用いた。第三実施形態においては、音声の基底スペクトルなら風雑音の基底スペクトルより調波成分が大きいであろうという考え方に基づき、このような調波成分の大きさを数値化するため、各々の基底ケプストラムから評価指標を算出する。

【００５２】

一般にケプストラムのピークは、スペクトルの調波成分の大きさと、その基本周波数を示している。例えば、ケプストラムがケフレンシー５ｍｓの位置にピークを持てば、スペクトルはその逆数の２００Ｈｚを基本周波数とする調波成分を持つ。

【００５３】

基底スペクトルの調波成分の大きさを数値化するためには、基底ケプストラムのピークの大きさを調べればよいため、簡単には基底ケプストラムの最大値を算出することが考えられる。しかしながら第二実施形態で述べたように、基底ケプストラム全体の最大値は、実際上は低ケフレンシー部分の最大値であるため、結局は基底スペクトルの周波数軸上での分布状態を見ていることになる。そこで第三実施形態では、基底ケプストラムの音声の基本周波数範囲に対応する部分においてその最大値を取ることで調波成分の大きさを数値化した評価指標とする。

【００５４】

図１０（ｃ）は、図１０（ｂ）の各々の基底スペクトルから求めた基底ケプストラムについて、音声の基本周波数範囲である１００～４００Ｈｚに対応する部分、すなわちケフレンシーで２．５～１０ｍｓの部分を拡大して表示したものである。なお、図１０（ｂ）、（ｄ）は、それぞれ図４（ｂ）、（ｄ）と同じものである。

【００５５】

図１０（ｃ）の点線丸で示すように、目的音に係ると予想される基底番号４番や１６番

の基底ケプストラムは、音声の基本周波数範囲に対応する部分にピークを持っており、このようなピークは基底番号７番や１１番の基底ケプストラムにも見られる。

【００５６】

音声の基本周波数範囲に対応する部分である、図１０（ｃ）の各々の基底ケプストラムについて最大値を棒グラフで示したのが図１０（ａ）であり、これを昇順にソートしたものが図１１（ａ）である。図１１（ａ）において、右の基底番号ほど、基底スペクトルの調波成分が大きいため音声である可能性が高く、逆に左の基底番号ほど、基底スペクトルの調波成分が小さいため風雑音である可能性が高いと考えられる。なお、上位４つ（７番、４番、１１番、１６番）の組み合わせを見ると、上記第一、第二実施形態における２番の代わりに、１１番が含まれていることが分かる。

【００５７】

図１１（ｂ）は、本実施形態における雑音基底数と目的音復元信号のＳＮＲの関係を示したグラフである。第一、第二実施形態と少し異なり、雑音基底数が１６（目的音基底数が４）のときＳＮＲが最大の２．９８ｄＢとなっており、上記第一、第二実施形態における２．２１ｄＢよりも高い。これは、先に述べた基底番号１１番が用いられたためと考えられる。特に、図１０（ｄ）の基底別復元信号で示される点線丸の部分が加わることで、第三実施形態の目的音復元信号を示した図６（ｂ）において、図６（ａ）と比較して点線丸の部分の音声が復元されていることが分かる。

【００５８】

なお、特許文献２には、雑音復元信号から調波成分を抽出して目的音復元信号に合成する処理があるが、このような処理は特に、目的音基底スペクトルが雑音基底スペクトルとして分類されてしまった場合に必要になると考えられる。第三実施形態は、基底スペクトルの分類における評価指標を調波成分の大きさとすることで、上記のような分類ミスを防止しようとするものであり、特許文献２とは異なる。

【００５９】

なお、時間ブロック長は３秒として説明を行ってきたが、第三実施形態において音声の音素（好ましくは母音）ごとに基底スペクトルを得るために、例えば０．３秒といった短い時間ブロック長を用いてもよい。この場合、非負値行列因子分解における行列サイズが縮小されるため、計算時間の短縮にもつながる。

【００６０】

なお、図３のフローチャートにおいて上述したＳ３０３の処理以外は、第一実施形態と同様である。

【００６１】

ここで、風雑音に音声が少し混ざったような基底スペクトルがある場合、各実施形態でどのように分類され得るかを考える。基底スペクトルの概形としては、低域が卓越した風雑音が支配的となるため、第一実施形態では、重心周波数が低くなって雑音基底スペクトルに分類される可能性が高い。また、基底スペクトルが狭い周波数範囲（低域）に集中しており、基底スペクトルの包絡成分が小さくなるため、第二実施形態においても雑音基底スペクトルに分類されると考えられる。一方、音声の調波成分は含まれているため、第三実施形態では目的音基底スペクトルに分類される可能性がある。以上より、第一実施形態や第二実施形態のように、基底スペクトルの周波数軸上での分布状態を見る方法は、雑音の除去を重視した方法であると解釈することができる。一方で、第三実施形態のように基底スペクトルに含まれる調波成分を見る方法は、目的音の保存を重視した方法であると解釈でき、このような重視する点の違いが、目的音復元信号のＳＮＲの値にも表れたと考えられる。

【００６２】

以上の各実施形態においては、目的音を音声、雑音を風雑音として説明を行ってきたが、本発明は他の音の組み合わせに対しても適用できることは言うまでもない。第一実施形態や第二実施形態の方法は、たとえ目的音と雑音で調波成分の強さが同程度であっても、基底スペクトルの周波数軸上での分布状態が異なっていればよいため、例えば目的音がせ

せらぎの音で、雑音が車のロードノイズであるような場合にも適用できる。また、第三実施形態の方法は、たとえ目的音と雑音で周波数帯が重なっていても、基底スペクトルの調波成分の大きさが異なっていればよい。よって、例えば音声や鳥の鳴き声のように調波成分を持つ目的音と、ざわめきのような雑音の組み合わせに対しても適用可能であるし、調波楽器と打楽器のような音の組み合わせにも用いることができる。このように、基底スペクトルの周波数軸上での分布状態を見る方法と、基底スペクトルに含まれる調波成分を見る方法によって、本発明は様々な音の組み合わせに対応することができる。

【００６３】

なお、目的音復元信号と雑音復元信号の少なくとも一方を用いて、元の音響信号から雑音除去を行うようにしてもよい。例えば、雑音復元信号を音響信号からスペクトル減算することで雑音除去を行ってもよいし、目的音復元信号と雑音復元信号から生成したウィナーフィルタを音響信号に適用してもよい。

【００６４】

なお、雑音をもう一つの目的音と考えれば、本発明を雑音除去ではなく音源分離に用いることもできる。さらに、ソートされた基底番号を２つではなく３つ以上に分けることで、音響信号を３つ以上の音に分離することも可能である。また、複数ｃｈの音響信号に対しても、ｃｈごとに本発明の処理を適用できることは言うまでもない。

【００６５】

なお、以上の実施形態においては、音響信号の振幅スペクトログラム｜Ｙ｜を非負値行列因子分解することで、基底行列とアクティビティ行列を得ていたがこれに限られるものではない。たとえば、複素ＮＭＦを用いることで、音響信号の複素スペクトログラムＹを、基底行列とアクティビティ行列および、サイズ｛Ｆ×Ｔ｝のＫ個の位相スペクトログラムＰｉ［ｉ＝１～Ｋ］に分解することができる。このとき、例えば目的音複素スペクトログラムＹｓは、式（１１）のように算出することになる。ただし、Σは目的音に対応するＫｓ個のｉに関して、和を取る操作を表すものとする。

$$Ys = \Sigma(H(:, i) * U(i, :) . * Pi) \quad \cdots \quad (11)$$

【００６６】

なお、振幅スペクトログラム｜Ｙ｜は、複素スペクトログラムＹの要素ごとに複素数の絶対値を取ったものとしていたが、代わりに絶対値の指数乗（例えば０．５乗や２乗）を取ったものとしてもよい。また、時間周波数変換において、フーリエ変換の他にウェーブレット変換などを用いてもよく、その場合はスカログラムが振幅スペクトログラムの代わりとなる。

【００６７】

以上説明したように、上記各実施形態によれば、音響信号の各々の基底スペクトルから物理的意味が明確な指標を算出し、算出された指標により基底スペクトルを目的音と雑音に分類するため、音響信号から高精度に雑音を除去することができる。また、単一の音響信号から教師基底なしで高精度に雑音除去することができる。

【００６８】

＜その他の実施形態＞

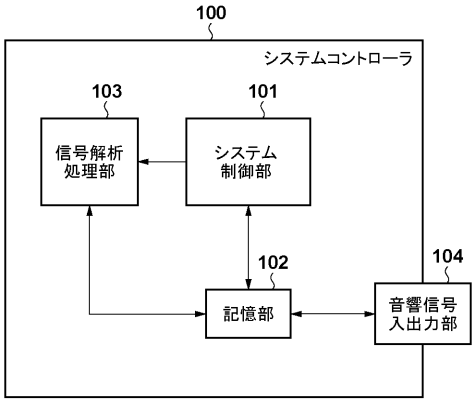
以上、実施形態例を詳述したが、本発明は例えば、システム、装置、方法、プログラム若しくは記録媒体(記憶媒体)等としての実施態様をとることが可能である。具体的には、複数の機器（例えば、ホストコンピュータ、インタフェース機器、撮像装置、Ｗｅｂアプリケーション等）から構成されるシステムに適用しても良いし、また、一つの機器からなる装置に適用しても良い。

【００６９】

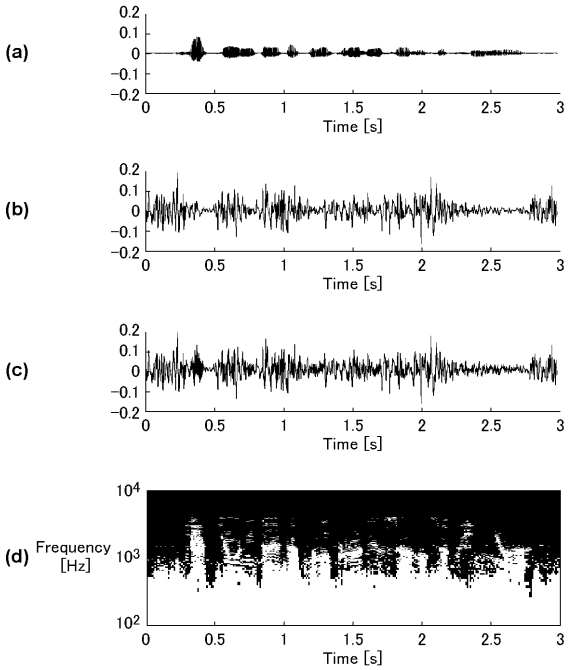
また、本発明の目的は、以下のようにすることによって達成されることはいうまでもない。即ち、前述した実施形態の機能を実現するソフトウェアのプログラムコード（コンピュータプログラム）を記録した記録媒体（または記憶媒体）を、システムあるいは装置に供給する。係る記憶媒体は言うまでもなく、コンピュータ読み取り可能な記憶媒体である。そして、そのシステムあるいは装置のコンピュータ（またはＣＰＵやＭＰＵ）が記録媒

体に格納されたプログラムコードを読み出し実行する。この場合、記録媒体から読み出されたプログラムコード自体が前述した実施形態の機能を実現することになり、そのプログラムコードを記録した記録媒体は本発明を構成することになる。

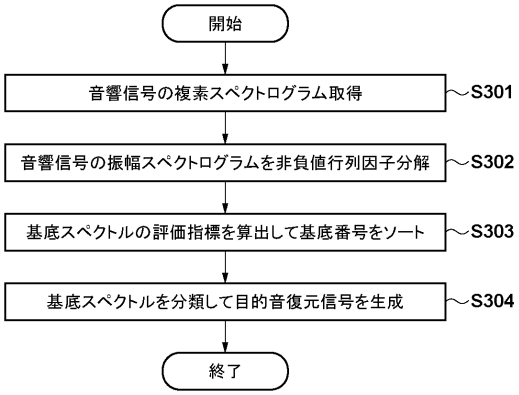
【図１】



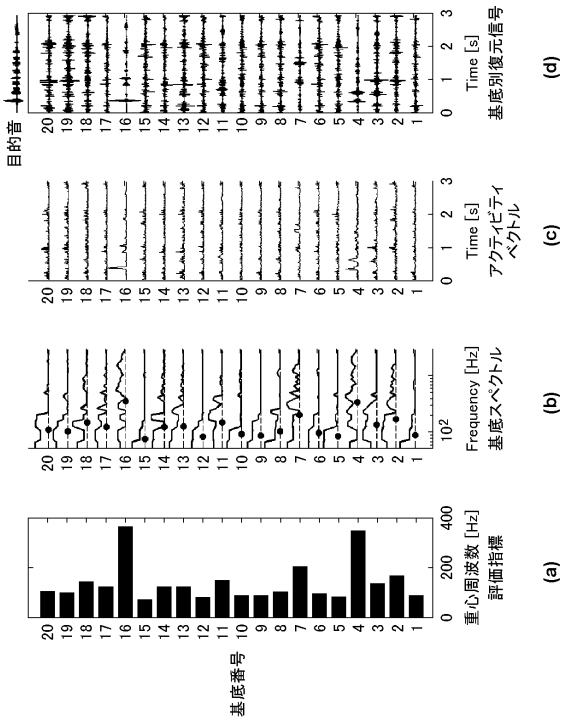
【図２】



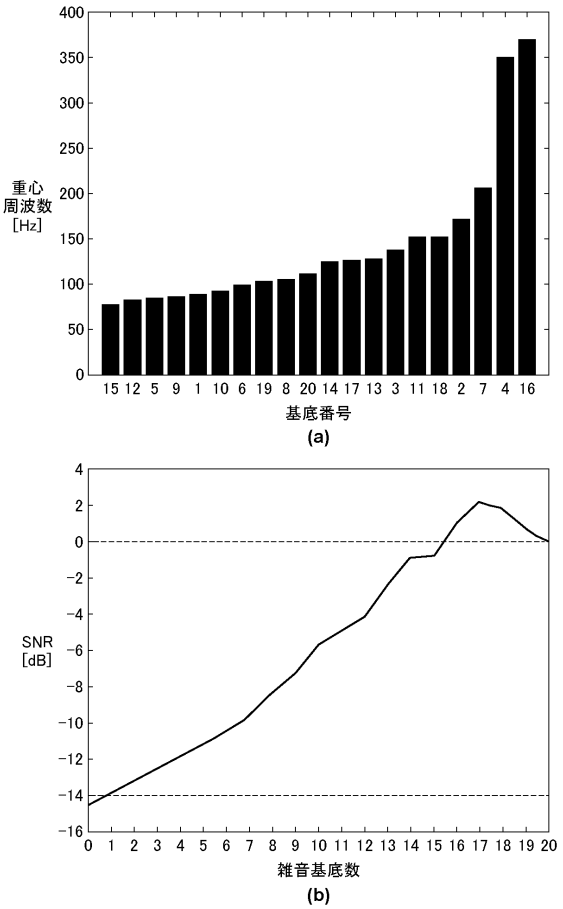
【図３】


開始
音響信号の複素スペクトログラム取得
S301
音響信号の振幅スペクトログラムを非負値行列因子分解
S302
基底スペクトルの評価指標を算出して基底番号をソート
S303
基底スペクトルを分類して目的音復元信号を生成
S304
終了

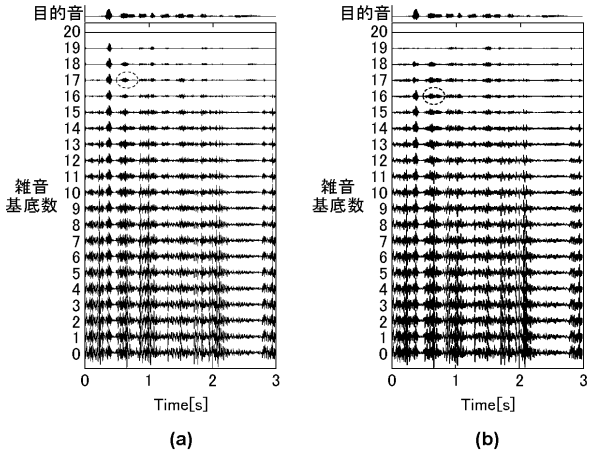
【図４】


目的音
Time [s]
基底別復元信号
(d)
アクティビティベクトル
(c)
Frequency [Hz]
基底スペクトル
(b)
重心周波数 [Hz]
評価指標
(a)
基底番号

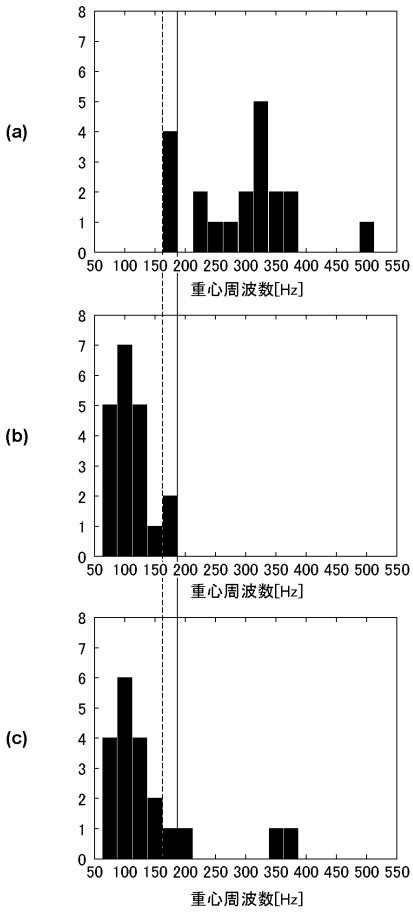
【図５】


重心周波数 [Hz]
基底番号
(a)
SNR [dB]
雑音基底数
(b)

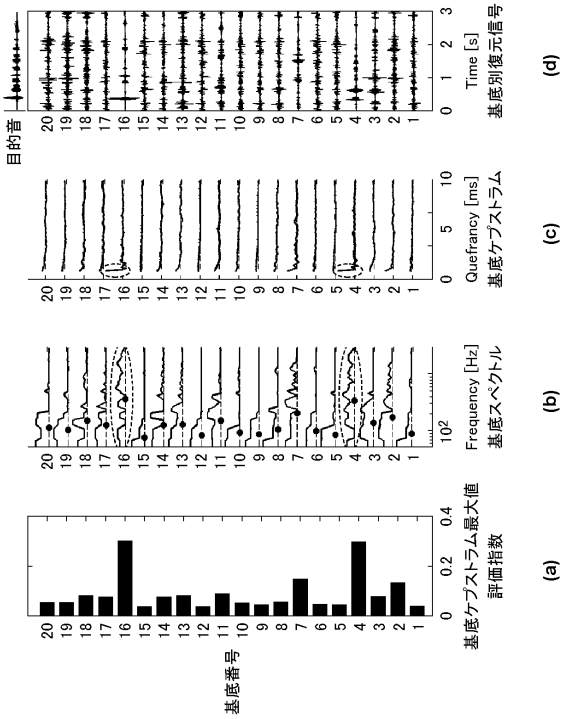
【図６】


目的音
雑音基底数
Time[s]
(a)
(b)

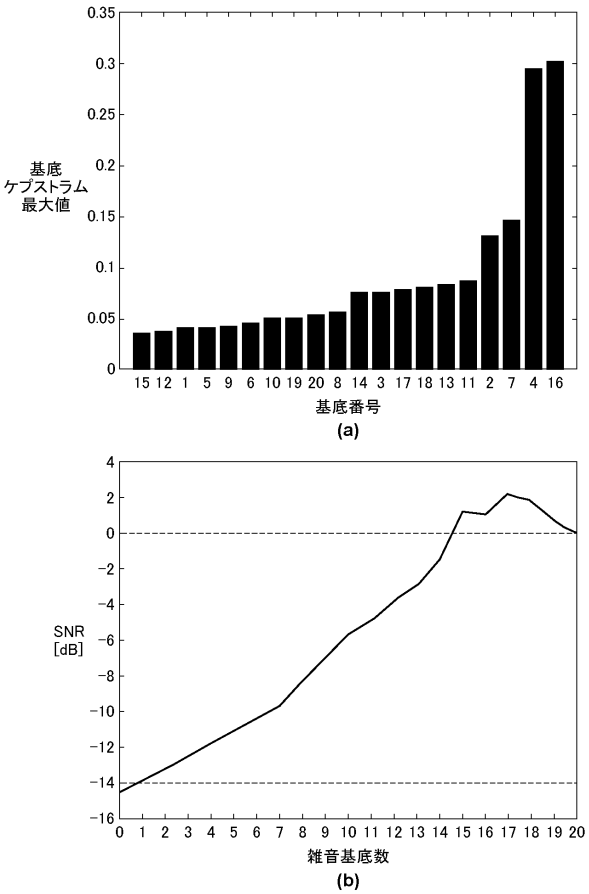
【図７】


(a)
(b)
(c)
重心周波数[Hz]

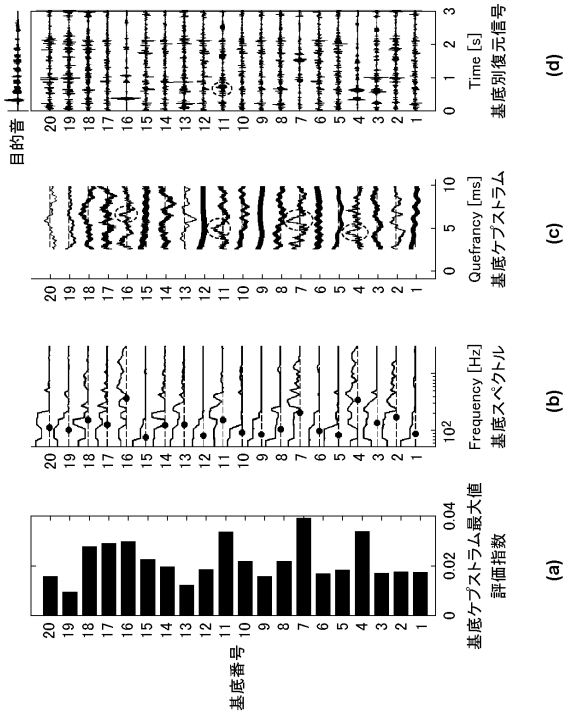
【図８】


目的音
Time [s]
基底別復元信号
(d)
Quefrancy [ms]
基底ケプストラム
(c)
Frequency [Hz]
基底スペクトル
(b)
基底ケプストラム最大値
評価指数
(a)
基底番号

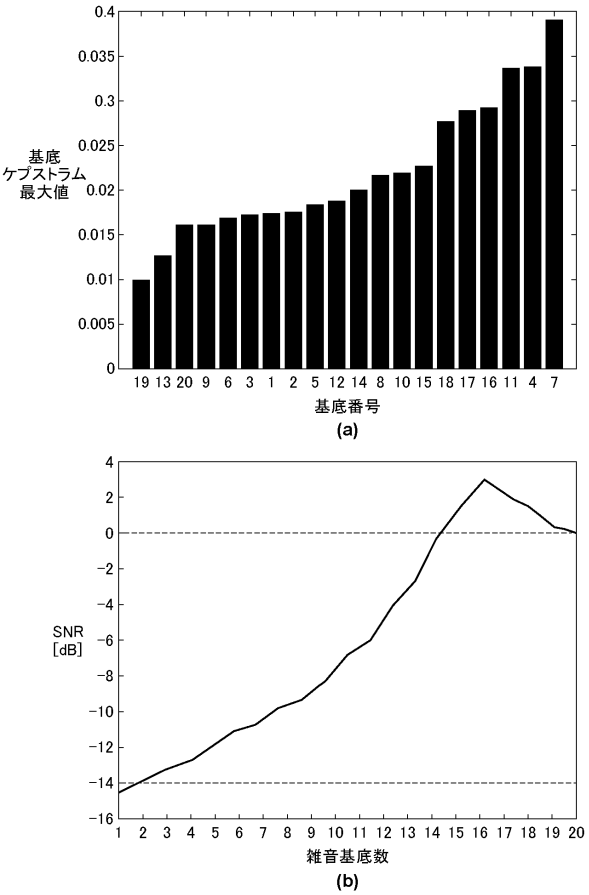
【図９】


基底
ケプストラム
最大値
基底番号
(a)
SNR
[dB]
雑音基底数
(b)

【図１０】


目的音
Time [s]
基底別復元信号
(d)
Quefrancy [ms]
基底ケプストラム
(c)
Frequency [Hz]
基底スペクトル
(b)
基底ケプストラム最大値
評価指数
(a)
基底番号

【図１１】


基底
ケプストラム
最大値
0.4
0.035
0.3
0.025
0.02
0.015
0.01
0.005
0
19 13 20 9 6 3 1 2 5 12 14 8 10 15 18 17 16 11 4 7
基底番号
(a)
SNR
[dB]
4
2
0
−2
−4
−6
−8
−10
−12
−14
−16
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
雑音基底数
(b)

フロントページの続き

(72)発明者　多和田　典朗
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　上田　雄

(56)参考文献　特開２０１３－０３７１５２（ＪＰ，Ａ）
特開２０１２－１６３９１８（ＪＰ，Ａ）
特開２０１３－０３３１９６（ＪＰ，Ａ）
特開２０１２－１３３３４６（ＪＰ，Ａ）
特開２００６－２４３１７８（ＪＰ，Ａ）
Marko Helen and Tuomas Virtanen, SEPARATION OF DRUMS FROM POLYPHONIC MUSIC USING NON-NEGATIVE MATRIX FACTORIZATION AND SUPPORT VECTOR MACHINE, Proc. 13th European Signal Processing Conference, IEEE, ２００５年　9月

(58)調査した分野(Int.Cl., ＤＢ名)
Ｇ１０Ｌ　２１／００－２１／１８
Ｇ１０Ｌ　２５／００－２５／９３