(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2014-21315
(P2014-21315A)
(43)公開日　平成26年2月3日(2014.2.3)

(51)Int.Cl.		F I		テーマコード（参考）
G10L 21/028	(2013.01)	G10L 21/02	201D	
G10L 21/0308	(2013.01)	G10L 21/02	203Z	
G10L 15/28	(2013.01)	G10L 15/28	400	

審査請求 未請求　請求項の数 7　OL　（全 27 頁）

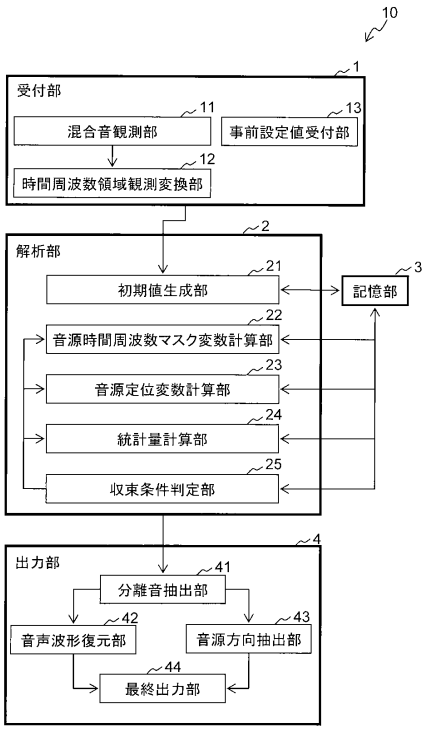
(21)出願番号　特願2012-160450 (P2012-160450)
(22)出願日　平成24年7月19日 (2012.7.19)

(71)出願人　000004226
日本電信電話株式会社
東京都千代田区大手町二丁目３番１号
(71)出願人　504132272
国立大学法人京都大学
京都府京都市左京区吉田本町３６番地１
(74)代理人　110001519
特許業務法人太陽国際特許事務所
(72)発明者　石黒　勝彦
東京都千代田区大手町二丁目３番１号　日本電信電話株式会社内
(72)発明者　澤田　宏
東京都千代田区大手町二丁目３番１号　日本電信電話株式会社内

最終頁に続く

(54)【発明の名称】音源分離定位装置、方法、及びプログラム

(57)【要約】
【課題】音源分離及び音源定位の両方の問題に対して、安定して高い性能を得る。
【解決手段】混合音観測部１１が、複数の音源の各々から発生した各音の混合音をマイクロフォンアレイにより観測した混合音信号を受け付け、時間周波数領域観測変換部１２が、混合音信号を時間周波数領域の観測信号x_{tf}に変換し、音源時間周波数マスク変数計算部２２が、統計量及び音源定位変数η_{kd}を用いたマスク変数ξ_{tfk}を計算し、音源定位変数計算部２３が、統計量及びマスク変数ξ_{tfk}を用いた音源定位変数η_{kd}を計算し、統計量計算部２４が、各種統計量を計算し、収束条件判定部２５が、音源時間周波数マスク変数計算部２２、音源定位変数計算部２３、及び統計量計算部２４の処理を、予め定めた収束条件を満たすまで反復させ、収束条件を満たした場合には、出力部４から解析結果を出力する。
【選択図】図３



【特許請求の範囲】
【請求項１】
複数の音源の各々から発せられた各音の混合音を、各々異なる位置に配置された複数の観測手段により観測した混合音信号を受け付ける受付手段と、
前記受付手段により受け付けた混合音信号を、前記複数の音源の各々に対応するように分離する音源分離と、前記観測手段を基準とした前記複数の音源の各々が存在する方向を推定する音源定位とを、前記音源分離と前記音源定位とで相互に依存させた変数を用いて反復処理する同時最適化により解析する解析手段と、
前記解析手段により解析された音源分離及び音源定位の結果を出力する出力手段と、
を含む音源分離定位装置。
【請求項２】
前記受付手段は、前記混合音信号を、時間フレームｔ及び周波数ビンｆ毎の各要素からなる時間周波数領域の観測信号x_{tf}に変換して前記解析手段に受け渡し、
前記解析手段は、
前記観測信号x_{tf}の各要素が、仮想的に設定した複数の音源の各々へ該各要素を割り当てる複数のマスクのｋ番目のマスクに対応する信号である確率を表すマスク変数ξ_{tfk}を、前記複数のマスクの各々について計算する音源時間周波数マスク変数計算手段と、
前記ｋ番目のマスクに対応した音源が、前記観測手段を基準として分割された複数の方向のｄ番目の方向に存在する確率を表す音源定位変数η_{kd}を、前記複数の方向の各々について計算する音源定位変数計算手段と、
前記マスク変数ξ_{tfk}及び前記音源定位変数η_{kd}の計算に用いられる統計量を計算する統計量計算手段と、
前記音源時間周波数マスク変数計算手段、前記音源定位変数計算手段、及び前記統計量計算手段の計算を、予め定めた収束条件を満たすまで反復させる収束条件判定手段と、を含み、
前記マスク変数ξ_{tfk}の計算に前記音源定位変数η_{kd}を用い、前記音源定位変数η_{kd}の計算に前記マスク変数ξ_{tfk}を用いる
請求項１記載の音源分離定位装置。
【請求項３】
前記解析手段は、無響環境において測定された前記複数の観測手段のステアリングベクトルを用いて、前記音源分離及び前記音源定位を解析する請求項１または請求項２記載の音源分離定位装置。
【請求項４】
受付手段と、解析手段と、出力手段とを含む音源分離定位装置における音源分離定位方法であって、
前記受付手段が、複数の音源の各々から発せられた各音の混合音を、各々異なる位置に配置された複数の観測手段により観測した混合音信号を受け付け、
前記解析手段が、前記受付手段により受け付けた混合音信号を、前記複数の音源の各々に対応するように分離する音源分離と、前記観測手段を基準とした前記複数の音源の各々が存在する方向を推定する音源定位とを、前記音源分離と前記音源定位とで相互に依存させた変数を用いて反復処理する同時最適化により解析し、
前記出力手段が、前記解析手段により解析された音源分離及び音源定位の結果を出力する
音源分離定位方法。
【請求項５】
前記解析手段が、音源時間周波数マスク変数計算手段と、音源定位変数計算手段と、統計量計算手段と、収束条件判定手段とを含む音源分離定位装置における音源分離定位方法であって、
前記受付手段が、前記混合音信号を、時間フレームｔ及び周波数ビンｆ毎の各要素からなる時間周波数領域の観測信号x_{tf}に変換して前記解析手段に受け渡し、

前記音源時間周波数マスク変数計算手段が、前記観測信号x_{tf}の各要素が、仮想的に設定した複数の音源の各々へ該各要素を割り当てる複数のマスクのｋ番目のマスクに対応する信号である確率を表すマスク変数ξ_{tfk}を、前記複数のマスクの各々について計算し、

前記音源定位変数計算手段が、前記ｋ番目のマスクに対応した音源が、前記観測手段を基準として分割された複数の方向のｄ番目の方向に存在する確率を表す音源定位変数η_{kd}を、前記複数の方向の各々について計算し、

前記統計量計算手段が、前記マスク変数ξ_{tfk}及び前記音源定位変数η_{kd}の計算に用いられる統計量を計算し、

前記収束条件判定手段が、前記音源時間周波数マスク変数計算手段、前記音源定位変数計算手段、及び前記統計量計算手段の計算を、予め定めた収束条件を満たすまで反復させ、

前記マスク変数ξ_{tfk}の計算に前記音源定位変数η_{kd}を用い、前記音源定位変数η_{kd}の計算に前記マスク変数ξ_{tfk}を用いる

請求項４記載の音源分離定位方法。

【請求項６】

前記解析手段が、無響環境において測定された前記複数の観測手段のステアリングベクトルを用いて、前記音源分離及び前記音源定位を解析する請求項４または請求項５記載の音源分離定位方法。

【請求項７】

コンピュータを、請求項１～請求項３のいずれか１項記載の音源分離定位装置を構成する各手段として機能させるための音源分離定位プログラム。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は、音源分離定位装置、方法、及びプログラムに係り、特に、複数の音源の各々から発せられた音の混合音から、個別の音源毎の音を分離すると共に、各音源の方向を定位する音源分離定位装置、方法、及びプログラムに関する。

【背景技術】

【０００２】

複数の音源の各々から発せられた音の重ね合わせである環境音（以下、混合音と呼ぶ）を個別の音源毎の音へと分離する音源分離技術は非常に古い歴史を持つ技術である。この技術は、例えば、会議の様子を録音した混合音から会議の議事録を作成するための発話者分離などに利用することができる。また、混合音を観測した複数のマイクの位置関係及び各マイクで観測された音から、各音源の相対位置及び方向を計算する音源定位技術は、例えば、環境中を自律移動するロボットや機械の自己位置同定や障害物回避などのための基礎的な技術として、非常に多くの手法が提案されている（例えば、非特許文献１～３）。

【０００３】

非特許文献１では、各時刻と各周波数とにおいては、通常高々１つの音源からの信号しか観測されない、という音源のスパース性を利用した音源分離法を提案している。非特許文献２では、ロボットでの利用を前提とした音源の分離及び定位を行うシステムを提案している。非特許文献３では、音源数よりも多いマイクを用いた音源分離手法を提案している。

【０００４】

この音源分離及び音源定位の２つの問題は、互いに深く密接に関係した相互依存の問題であることが知られている。例えば、複数の音源の位置が分かっている場合には、ビームフォーマという技術を使うことで各音源のみの分離音を精度よく復元できることが知られている。一方、各音源の音が分離できている場合に、各音源の位置を決定することも比較的容易である。

【先行技術文献】

【非特許文献】
【０００５】
【非特許文献１】Sawada, H., Araki, S. and Makino, S. “Underdetermined Convolutive Blind Source Separation via Frequency Bin-Wise Clustering and Permutation Alignment”, IEEE Transactions on Audio, Speech and Language Processing, Vol. 19, No. 3, pp. 516-527, 2011.
【非特許文献２】Nakadai, K. Lourens, T., Okuno, H. G. and Kitano, H. “Active Audition for Humanoid”, in Proc. AAAI, 2000.
【非特許文献３】Lee, I., Kim, T. and Lee, T.-W., “Fast Fixed-point Independent Vector Analysis Algorithms for Convolutive Blind Source Separation”, Signal Processing, Vol. 87, No. 8, pp.1859-1871, 2007.
【発明の概要】
【発明が解決しようとする課題】
【０００６】
上述の音源分離及び音源定位の２つの問題を同時に解決することができれば、例えば、自律ロボットが障害物回避を行いながら、騒音環境下で特定のユーザの指令コマンドを音声で受け取って行動することなど、非常に高度な知能システムを実現することができる。
【０００７】
しかしながら、非特許文献１に代表される既存手法では、音源分離及び音源定位という相互依存する問題を個別に解決している。例えば、非特許文献１の手法は音源分離を主目的としており、音源分離完了後に各音源の定位を行うことを前提にしている。また、非特許文献２の手法は逆に各音源の定位を完了した後に、各音源が発する音声信号を分離している。これらの従来手法のように、まず、音源分離及び音源定位の一方の問題を何らかの事前情報や強い仮定を伴う方法で解決した後に、他方の問題を解決する場合には、最初に解決した一方の問題の精度が悪かった場合に、他方の問題の精度も大きく劣化してしまう、という問題がある。
【０００８】
本発明は上記問題点を解決するために成されたものであり、音源分離及び音源定位の両方の問題に対して、安定して高い性能を得ることができる音源分離定位装置、方法、及びプログラム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】
【０００９】
上記目的を達成するために、本発明の音源分離定位装置は、複数の音源の各々から発せられた各音の混合音を、各々異なる位置に配置された複数の観測手段により観測した混合音信号を受け付ける受付手段と、前記受付手段により受け付けた混合音信号を、前記複数の音源の各々に対応するように分離する音源分離と、前記観測手段を基準とした前記複数の音源の各々が存在する方向を推定する音源定位とを、前記音源分離と前記音源定位とで相互に依存させた変数を用いて反復処理する同時最適化により解析する解析手段と、前記解析手段により解析された音源分離及び音源定位の結果を出力する出力手段と、を含んで構成されている。
【００１０】
本発明の音源分離定位装置によれば、受付手段が、複数の音源の各々から発せられた各音の混合音を、各々異なる位置に配置された複数の観測手段により観測した混合音信号を受け付ける。そして、解析手段が、受付手段により受け付けた混合音信号を、複数の音源の各々に対応するように分離する音源分離と、観測手段を基準とした複数の音源の各々が存在する方向を推定する音源定位とを、音源分離と音源定位とで相互に依存させた変数を用いて反復処理する同時最適化により解析する。相互に依存させた変数を用いるとは、音源分離及び音源定位の一方で求めた変数を、他方の変数を求める際に用いることである。最後に、出力手段が、解析手段により解析された音源分離及び音源定位の結果を出力する。

【００１１】
このように、音源分離と音源定位とを相互に依存させて同時最適化により解析することにより、音源分離及び音源定位の両方の問題に対して、安定して高い性能を得ることができる。
【００１２】
また、前記受付手段は、前記混合音信号を、時間フレームｔ及び周波数ビンｆ毎の各要素からなる時間周波数領域の観測信号x_{tf}に変換して前記解析手段に受け渡すことができる。また、前記解析手段は、前記観測信号x_{tf}の各要素が、仮想的に設定した複数の音源の各々へ該各要素を割り当てる複数のマスクのｋ番目のマスクに対応する信号である確率を表すマスク変数ξ_{tfk}を、前記複数のマスクの各々について計算する音源時間周波数マスク変数計算手段と、前記ｋ番目のマスクに対応した音源が、前記観測手段を基準として分割された複数の方向のｄ番目の方向に存在する確率を表す音源定位変数η_{kd}を、前記複数の方向の各々について計算する音源定位変数計算手段と、前記マスク変数ξ_{tfk}及び前記音源定位変数η_{kd}の計算に用いられる統計量を計算する統計量計算手段と、前記音源時間周波数マスク変数計算手段、前記音源定位変数計算手段、及び前記統計量計算手段の計算を、予め定めた収束条件を満たすまで反復させる収束条件判定手段と、を含んで構成することができ、前記マスク変数ξ_{tfk}の計算に前記音源定位変数η_{kd}を用い、前記音源定位変数η_{kd}の計算に前記マスク変数ξ_{tfk}を用いることができる。これにより、音源分離と音源定位とを相互に依存させて、効率よく同時最適化を行うことができる。
【００１３】
また、前記解析手段は、無響環境において測定された前記複数の観測手段のステアリングベクトルを用いて、前記音源分離及び前記音源定位を解析することができる。これにより、様々な残響環境にも適用することができる。
【００１４】
また、本発明の音源分離定位方法は、受付手段と、解析手段と、出力手段とを含む音源分離定位装置における音源分離定位方法であって、前記受付手段が、複数の音源の各々から発せられた各音の混合音を、各々異なる位置に配置された複数の観測手段により観測した混合音信号を受け付け、前記解析手段が、前記受付手段により受け付けた混合音信号を、前記複数の音源の各々に対応するように分離する音源分離と、前記観測手段を基準とした前記複数の音源の各々が存在する方向を推定する音源定位とを、前記音源分離と前記音源定位とで相互に依存させた変数を用いて反復処理する同時最適化により解析し、前記出力手段が、前記解析手段により解析された音源分離及び音源定位の結果を出力する方法である。
【００１５】
また、前記解析手段が、音源時間周波数マスク変数計算手段と、音源定位変数計算手段と、統計量計算手段と、収束条件判定手段とを含む音源分離定位装置における音源分離定位方法であって、前記受付手段が、前記混合音信号を、時間フレームｔ及び周波数ビンｆ毎の各要素からなる時間周波数領域の観測信号x_{tf}に変換して前記解析手段に受け渡し、前記音源時間周波数マスク変数計算手段が、前記観測信号x_{tf}の各要素が、仮想的に設定した複数の音源の各々へ該各要素を割り当てる複数のマスクのｋ番目のマスクに対応する信号である確率を表すマスク変数ξ_{tfk}を、前記複数のマスクの各々について計算し、前記音源定位変数計算手段が、前記ｋ番目のマスクに対応した音源が、前記観測手段を基準として分割された複数の方向のｄ番目の方向に存在する確率を表す音源定位変数η_{kd}を、前記複数の方向の各々について計算し、前記統計量計算手段が、前記マスク変数ξ_{tfk}及び前記音源定位変数η_{kd}の計算に用いられる統計量を計算し、前記収束条件判定手段が、前記音源時間周波数マスク変数計算手段、前記音源定位変数計算手段、及び前記統計量計算手段の計算を、予め定めた収束条件を満たすまで反復させ、前記マスク変数ξ_{tfk}の計算に前記音源定位変数η_{kd}を用い、前記音源定位変数η_{kd}の計算に前記マスク変数ξ_{tfk}を用いることができる。

【００１６】
　また、本発明の音源分離定位方法において、前記解析手段が、無響環境において測定された前記複数の観測手段のステアリングベクトルを用いて、前記音源分離及び前記音源定位を解析することができる。
【００１７】
　また、本発明の音源分離定位プログラムは、コンピュータを、上記の音源分離定位装置を構成する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】
【００１８】
　以上説明したように、本発明の音源分離定位装置、方法、及びプログラムによれば、音源分離と音源定位とを相互に依存させて同時最適化により解析することにより、音源分離及び音源定位の両方の問題に対して、安定して高い性能を得ることができる、という効果が得られる。
【図面の簡単な説明】
【００１９】
【図１】本実施の形態の概要（音源分離）を示すイメージ図である。
【図２】本実施の形態の概要（音源定位）を示すイメージ図である。
【図３】本実施の形態に係る音源分離定位装置の機能的構成を示すブロック図である。
【図４】記憶部の構成を示す図である。
【図５】本実施の形態における音源分離定位処理ルーチンの内容を示すフローチャートである。
【図６】初期値生成処理ルーチンの内容を示すフローチャートである。
【図７】音源時間周波数マスク変数計算処理ルーチンの内容を示すフローチャートである。
【図８】音源定位変数計算処理ルーチンの内容を示すフローチャートである。
【図９】統計量計算処理ルーチンの内容を示すフローチャートである。
【図１０】実験例のセットアップを示す概略図である。
【図１１】実験例における音源定位の性能を示すグラフである。
【図１２】実験例における音源分離の性能を示すグラフである。
【発明を実施するための形態】
【００２０】
　以下、図面を参照して本発明の実施の形態を詳細に説明する。
【００２１】
＜概要＞
　まず、本実施の形態の概要について説明する。図１及び図２は、本実施の形態の概要を示すイメージ図である。
【００２２】
　図１に示すように、音源分離は、観測した混合音をフーリエ変換によって時間周波数領域の信号に変換した観測信号中の各（ｔ，ｆ）要素を、Ｋ種類の音源に割り振ることで実現する。なお、ｔは時間フレームを表すインデックス、ｆは周波数ビンを表すインデックスである。この方法は、時間周波数領域の観測信号のスパース性を利用した音源分離手法として、非特許文献１などで利用されており、良い音源分離性能を示すことが知られている。時間周波数領域の観測信号の各音源への割り当ては「音源時間周波数マスク」と呼ばれる。このマスクによって混合音を音源毎に分離し、各音源から発せられた音を復元することができる。
【００２３】
　図２に示すように、音源定位は、Ｋ種類の音源の方向を、マイクロフォンアレイを中心とした３６０度方向のいずれかに決定することで実現する。数学的には、マイクロフォンの方向解像度などの制約に従って、方向をＤ種類へ離散化（分割）する。そして、各音源をＤ種類の方向中のいずれかの方向１つへ割り当てる、すなわちＤ種類の方向へクラスタ

リングすることによって定位する。

【００２４】

これら個々の手法自体は新しいものではないが、本実施の形態では、音源分離及び音源定位を同時に、かつ相互依存する形で解決する枠組みを特徴とする。すなわち、本実施の形態では、音源時間周波数マスクを計算することで音源分離を可能とする。また、音源クラスタ毎にその方向を計算することで音源定位を可能とする。さらに、これらの音源分離及び音源定位を交互反復して同時最適化し、繰り返し計算手法により収束させることで、複数音源の同時分離及び定位を可能とすることを特徴とする。

【００２５】

さらに、本実施の形態のもう一つの特徴として、マイクロフォンアレイの無響ステアリングベクトルを利用する点がある。無響ステアリングベクトルとは、各マイクロフォンアレイの音響的な固有の性質である、無響室のインパルス応答である。この情報は実際の有響環境下における観測状況でのインパルス応答は異なるが、そのインパルス応答を予測する上では非常に有効であることが多い。本実施の形態では、この無響ステアリングベクトルを事前に計測、入力しておき、実際の混合音に適した音響特性の推定を音源分離及び定位と同時に行う。これにより、残響環境によらず、良い分離及び定位性能を得ることができる。

【００２６】

＜システム構成＞

本実施の形態に係る音源分離定位装置１０は、ＣＰＵ（Central Processing Unit）と、ＲＡＭ（Random Access Memory）と、後述する音源分離定位処理ルーチンを実行するためのプログラムを記憶したＲＯＭ（Read Only Memory）とを備えたコンピュータで構成されており、ＣＰＵが音源分離定位処理ルーチンを実行するためのプログラムを、内部記憶装置であるＲＯＭから読み込んで実行することにより形成される。

【００２７】

このコンピュータは、機能的には、図３に示すように、解析したい混合音及びマイクロフォンアレイの音響特性を示すデータの入力を受け付ける受付部１と、音源分離及び音源定位の解析に必要な変数を計算及び更新する解析部２と、受け付けたデータ及び計算された情報を記憶する記憶部３と、解析結果を出力する出力部４とを含んだ構成で表すことができる。

【００２８】

受付部１は、さらに、混合音観測部１１と、時間周波数領域観測変換部１２と、事前設定値受付部１３とを含んだ構成で表すことができる。

【００２９】

混合音観測部１１は、記憶装置などの入力器または本装置に付随するマイクロフォンアレイから、観測された混合音が電子データに変換された混合音信号を受け付ける。例えば、既にマイクロフォンアレイによって観測され、電子データに変換された上で、一旦記憶装置に記憶された混合音信号を記憶装置から読み込むことにより、入力データとして受け付けることができる。また、本装置に付随するマイクロフォンアレイで観測された混合音を、直接電子データに変換して受け付けることもできる。

【００３０】

時間周波数領域観測変換部１２では、混合音観測部１１で受け付けた混合音信号を、フーリエ変換を利用して時間周波数領域の信号へと変換する。以下、混合音信号を時間周波数領域に変換した信号を観測信号と呼ぶ。

【００３１】

事前設定値受付部１３は、キーボードや記憶装置などの入力器から、後述する本装置の実装したモデルに必要な定数、入力された混合音を観測したマイクロフォンアレイの無響ステアリングベクトル情報を含む統計量初期値の一部、及び収束判定閾値の値を受け付ける。

【００３２】

解析部２は、さらに、初期値生成部２１と、音源時間周波数マスク変数計算部２２と、音源定位変数計算部２３と、統計量計算部２４と、収束条件判定部２５とを含んだ構成で表すことができる。

【００３３】

初期値生成部２１は、受付部１で受け付けた情報を記憶部３の各部へ記憶すると共に、記憶部３の各部に記憶された値の初期化を行う。

【００３４】

音源時間周波数マスク変数計算部２２は、記憶部３に保存された情報を利用して、音源時間周波数マスク変数を計算し、保存及び更新する。

【００３５】

音源定位変数計算部２３は、記憶部３に保存された情報を利用して、音源定位変数を計算し、保存及び更新する。

【００３６】

統計量計算部２４は、記憶部３に保存された情報を利用して、統計量を計算し、保存及び更新する。

【００３７】

収束条件判定部２５は、記憶部３に保存された情報を利用して、解析部２の計算処理を継続するか、終了するかを判定する。終了する場合は、解析結果を出力部４へ渡す。

【００３８】

記憶部３には、図４に示すように、定数記憶部３１、観測信号記憶部３２、音源時間周波数マスク変数記憶部３３、音源定位変数記憶部３４、統計量記憶部３５、統計量初期値記憶部３６、及び収束判定閾値記憶部３７の各記憶部が設けられている。

【００３９】

定数記憶部３１には、本装置の実装したモデルに必要な定数が記憶される。

【００４０】

観測信号記憶部３２には、時間周波数領域観測変換部１２で変換された観測信号が記憶される。

【００４１】

音源時間周波数マスク変数記憶部３３には、主に音源分離の解析結果を表現する情報を表す音源時間周波数マスク変数が記憶される。

【００４２】

音源定位変数記憶部３４には、主に音源定位の解析結果を表現する情報を表す音源定位変数が記憶される。

【００４３】

統計量記憶部３５には、音源分離及び音源定位に必要となる各種統計量が記憶される。

【００４４】

統計量初期値記憶部３６には、統計量の計算に必要となる初期値である統計量初期値が記憶される。

【００４５】

収束判定閾値記憶部３７には、解析結果の収束を判定するために用いる閾値が記憶される。

【００４６】

出力部４は、さらに、分離音抽出部４１と、音声波形復元部４２と、音源方向抽出部４３と、最終出力部４４とを含んだ構成で表すことができる。

【００４７】

分離音抽出部４１は、記憶部３に保存された情報を利用して、時間周波数領域での分離音信号を計算して音声波形復元部４２へと渡す。

【００４８】

音声波形復元部４２は、記憶部３に保存された情報、及び分離音抽出部４１から渡された各分離音の時間周波数領域信号を利用して、各分離音の時間周波数領域信号を逆フーリ

エ変換によって分離音の音声信号へと復元する。

【００４９】

音源方向抽出部４３は、記憶部３に保存された情報を利用して、各音源の方向を計算及び定位する。

【００５０】

最終出力部４４は、ディスプレイ、プリンタ、スピーカー、磁気ディスクなどで実装された出力装置に、ユーザの所望の形式で音源分離及び音源定位の解析結果を出力する。

【００５１】

＜本実施の形態の作用＞

次に、本実施の形態に係る音源分離定位装置１０の作用について説明する。まず、複数のマイクロフォンを任意の配置で設置したマイクロフォンアレイを利用して観測された混合音が記憶装置に混合音信号として記憶された状態、または本装置に付随するマイクロフォンアレイにより混合音が観測されている状態で、音源分離定位装置１０において、図５に示す音源分離定位処理ルーチンが実行される。

【００５２】

ステップ１００で、混合音観測部１１が、記憶装置に記憶された混合音信号を読み込むことにより受け付けるか、または、マイクロフォンアレイにより観測された混合音を電子データである混合音信号に変換して直接受け付ける。

【００５３】

次に、ステップ１０２で、時間周波数領域観測変換部１２が、上記ステップ１００で受け付けた混合音信号を時間周波数領域の信号である観測信号へ変換する。変換には短時間フーリエ変換（ＳＴＦＴ）あるいは高速フーリエ変換（ＦＦＴ）を利用することができる。変換した観測信号をx_{tf}で表す。各x_{tf}は時間フレームｔ（ｔ＝１，...，Ｔ）、フーリエ変換によるｆ（ｆ＝１，...，Ｆ）番目の周波数ビン（周波数帯）における音声信号の変換表現である。各x_{tf}はマイク数に相当する要素数のベクトルであり、各要素は複素数である。以後、このx_{tf}を本装置にとっての観測量として用いる。この観測量は、本実施の形態の中では複素正規分布から生成されるものと仮定する。入力された混合音信号全てを時間周波数領域へ変換し、変換した各x_{tf}を初期値生成部２１に渡す。

【００５４】

次に、ステップ１０４で、事前設定値受付部１３が、音源分離定位装置１０における解析処理に必要な定数を受け付ける。定数には、観測信号の総時間フレーム数Ｔ、観測信号の総周波数ビン数Ｆ、仮想的に設定する音源の最大数であるマスク数Ｋ、音源方向をクラスタリングするための方向クラス数Ｄ、混合音を観測したマイクロフォンアレイのマイク数Ｍ、及び統計量の初期値を計算する際に利用される正実数である正則化定数εが含まれる。マスク数Ｋは、例えば１２とすることができるが、さらに多数の音源が予想される場合には、音源数を十分上回る数を設定する。また、正則化定数εは、例えば０．０００１と設定することができる。

【００５５】

また、事前設定値受付部１３は、解析に必要な統計量の初期値の一部（β_k^0、κ_d^0、a_{tf}^0、v_{fd}^0）も受け付ける。β_k^0は音源時間周波数マスク変数の数学モデルに利用されるディリクレ分布の初期パラメータであり、ｋ＝１，...，Ｋに対し０より大きい値を設定する。例えば、$\beta_1^0=\beta_2^0=...=\beta_K^0$とすることができる。$\kappa_d^0$は音源定位変数の数学モデルに利用されるディリクレ分布の初期パラメータであり、ｄ＝１，...，Ｄに対し０より大きい値を設定する。例えば、$\kappa_1^0=\kappa_2^0=...=\kappa_D^0$とすることができる。$a_{tf}^0$は時間周波数領域観測信号の数学モデルで利用されるガンマ分布の初期パラメータである。v_{fd}^0は時間周波数領域観測信号の数学モデルで利用される複素ウィシャート分布の初期パラメータである。例えば、ｔ＝１，...，Ｔ，ｆ＝１，...，Ｆ，ｄ＝１，...，Ｄに対し$a_{tf}^0=1$、$v_{fd}^0=M$と設定することができる。

【００５６】

さらに、事前設定値受付部１３は、マイクロフォンアレイの音響的特性を表す、無響ステアリングベクトルｑも受け付ける。無響ステアリングベクトルは周波数ビンｆ及び方向ｄ毎に式（１）に示すように、Ｍ本のマイク毎に事前に無響室で測定したものである。

【００５７】

【数１】

$$q_{fd} = \left[q_{fd}^{1}, q_{fd}^{2}, \cdots, q_{fd}^{M} \right] \qquad (1)$$

【００５８】

例えば非特許文献１や非特許文献３等の従来手法では、利用するマイクの配置や無響室でのインパルス応答といった、システム固有の音響的特性を利用していない。これは利用するシステムの設定に寄らない一般性を持つものの、音響特性の事前情報が利用できないことで様々な残響環境において高精度な音源分離及び定位性能を得る可能性が低くなってしまう。

【００５９】

そこで、本実施の形態では、システム固有の音響特性である無響ステアリングベクトルを事前に測定及び入力しておくことで、様々な残響環境にも適応できる高精度な解析を実現することができる。

【００６０】

さらに、事前設定値受付部１３は、解析処理の収束判定に利用する収束判定閾値θも受け付ける。収束判定閾値θの値は、ユーザの設定した収束判定基準によって変わるが、本実施の形態では、音源分離解析の変化幅を利用するため、正の実数となる。

【００６１】

次に、ステップ２００～６００で、解析部２が、記憶部３に定義される変数を最適化するための計算を実施する。記憶部３に定義される変数の最適化には様々な最適化法（例えば、非特許文献４「Ｃ．Ｍ．ビショップ、“パターン認識と機械学習　上・下”、シュプリンガー・ジャパン、２００７．」）を利用できるが、本実施の形態では、変分ベイズ法に基づく音源分離及び音源定位の同時最適化を行う。

【００６２】

まず、ステップ２００で、初期値生成部２１が図６に示す初期値生成処理ルーチンを実行して初期値を設定する。そして、ステップ３００で、音源時間周波数マスク変数計算部２２が図７に示す音源時間周波数マスク変数計算処理ルーチンを実行し、ステップ４００で、音限定位変数計算部２３が図８に示す音限定位変数計算処理ルーチンを実行し、ステップ５００で、統計量計算部２４が図９に示す統計量計算処理ルーチンを実行して順番に各値を計算し、収束条件を満足するまで繰り返し反復計算を行うことで最適化する。変分ベイズ法に基づく計算では、必ず計算結果が収束することが保証されている。

【００６３】

以下、各処理について詳述する。なお、音源時間周波数マスク変数計算処理ルーチン、音限定位変数計算処理ルーチン、及び統計量計算処理ルーチンの実行の順番は任意でよい。

【００６４】

まず、初期値生成処理ルーチン（図６）では、ステップ２０２で、上記ステップ１０４において事前設定値受付部１３が受け付けた総時間フレーム数Ｔ、総周波数ビン数Ｆ、マスク数Ｋ、方向クラス数Ｄ、マイク数Ｍ、及び正則化定数εを、定数記憶部３１に保存する。

【００６５】

次に、ステップ２０４で、上記ステップ１０２において時間周波数領域観測変換部１２の計算の結果得られた観測信号x_{tf}を、観測信号記憶部３２に保存する。

【００６６】
次に、ステップ２０６で、上記ステップ１０４において事前設定値受付部１３が受け付けた収束判定閾値θを、収束判定閾値記憶部３７に保存する。
【００６７】
次に、ステップ２０８で、統計量の初期値を設定する。まず、上記ステップ１０４において事前設定値受付部１３が受け付けた統計量初期値の一部（β_k^0、κ_d^0、a_{tf}^0、v_{fd}^0）を統計量初期値記憶部３６に保存する。さらに、上記ステップ１０４において事前設定値受付部１３が受け付けた無響ステアリングベクトルq_{fd}、及び上記ステップ１０２で計算された観測信号x_{tf}を利用して、式（２）及び（３）に示すように、統計量初期値の一部であるG_{fd}^0及びb_{tf}^0を計算する。なお、Ｈはエルミート転置を示し、I_MはＭ次元の単位行列を表す。
【００６８】
【数２】

$$G_{fd}^{\ 0} = \frac{1}{M}\left(q_{fd}q_{fd}^H + \varepsilon I_M\right)^{-1} \qquad (2)$$

$$b_{tf}^{\ 0} = x_{tf}^H x_{tf} \qquad (3)$$

【００６９】
式（２）及び（３）により計算された統計量初期値の一部（G_{fd}^0、b_{tf}^0）を統計量初期値記憶部３６に保存する。
【００７０】
次に、ステップ２１０で、式（４）により、音源時間周波数マスク変数の初期値ξ_{tfk}^0を計算し、計算したξ_{tfk}^0をξ_{tfk}として音源時間周波数マスク変数記憶部３３に保存する。なお、Ｚはｋに関する和を１にするための正規化項である。
【００７１】
【数３】

$$\xi_{tfk}^0 = \frac{1}{Z}\exp\left(-x_{tf}^H \sum_{d=1}^{D}\left(\eta_{kd}^0 G_{fd}\right)x_{tf}\right)^{-1} \qquad (4)$$

【００７２】
次に、ステップ２１２で、式（５）及び（６）により、音源定位変数の初期値η_{kd}^0を計算し、計算したη_{kd}^0をη_{kd}として音源定位変数記憶部３４に保存する。
【００７３】
【数４】

$$\eta_{kd}^0 = \begin{cases} \dfrac{K}{D} & (k-1)\dfrac{D}{K} \le d \le k\dfrac{D}{K} \qquad (5) \\ 0 & \text{otherwise.} \qquad (6) \end{cases}$$

【００７４】

以上、初期値の設定が終了すると、初期値生成処理ルーチンを終了して、音源分離定位処理ルーチンへリターンする。

【００７５】

次に、音源時間周波数マスク変数計算処理ルーチン（図７）では、ステップ３０２で、記憶部３から必要な情報をロードし、次に、ステップ３０４で、時間フレームに対応する変数ｔを１にセットし、次に、ステップ３０６で、周波数ビンに対応する変数ｆを１にセットする。

【００７６】

次に、ステップ３０８で、式（７）により、ｋ＝１，...，Ｋについて音源時間周波数マスク変数ξ_{tfk}を計算する。なお、Ψはディガンマ関数である。

【００７７】

【数５】

$$\xi_{tfk} = \exp\left[\begin{array}{l}\Psi(\beta_{tk})-\Psi\left(\sum_{k=1}^{K}\beta_{tk}\right)+M\left\{\Psi(a_{tfk})-\log b_{tfk}\right\} \\ +\sum_{d=1}^{D}\eta_{kd}\left\{\sum_{m=0}^{M-1}\Psi(v_{fd}-m)+\log\left|G_{fd}\right|-\frac{a_{tfk}}{b_{tfk}}x_{td}^{H}v_{fd}G_{fd}x_{tf}\right\}\end{array}\right] \quad (7)$$

【００７８】

音源時間周波数マスク変数ξ_{tfk}は、観測信号を分離するために計算する変数である。ｔ＝１，...，Ｔ、ｆ＝１，...，Ｆ、ｋ＝１，...，Ｋとする。ξ_{tfk}は時間フレームｔ、周波数ビンｆにおける観測信号がｋ番目の音源（マスク）による信号である確率を表す。この音源時間周波数マスク変数ξ_{tfk}に従って、観測信号をＫ音源に分離することで、音源毎の分離音を復元することができる。この音源時間周波数マスク変数ξ_{tfk}は、本実施の形態では多項分布から生成されると仮定しており、その多項分布のパラメータは統計量βでパラメタライズされたディリクレ分布によって決定されるものとする。

【００７９】

式（７）のポイントは、右辺第４項にあるように、音源定位変数η_{kd}が必要であるという点である。これは、音源定位の情報を使って音源分離が改善されることを表している。

【００８０】

次に、ステップ３１０で、上記ステップ３０８でｋ＝１，...，Ｋについて計算された音源時間周波数マスク変数ξ_{tfk}を、式（８）により正規化する。ξ_{tfk}は確率であるので、各ｔ及びｆ毎に、全てのｋに対する和が常に１となるように正規化する。

【００８１】

【数６】

$$\xi_{tfk} = \frac{\xi_{tfk}}{\sum_{k'=1}^{K} \xi_{tfk'}} \qquad (8)$$

【００８２】

次に、ステップ３１２で、ｆを１インクリメントして、次のステップ３１４で、ｆが総周波数ビン数Ｆを超えたか否かを判定し、ｆが未だＦに到達していない場合には、ステップ３０８へ戻って、ステップ３０８～３１２の処理を繰り返す。

【００８３】

一方、ｆがＦを超えた場合には、ステップ３１６へ移行し、ｔを１インクリメントして、次のステップ３１８で、ｔが総時間フレーム数Ｔを超えたか否かを判定し、ｔが未だＴに到達していない場合には、ステップ３０６へ戻って、ステップ３０６～３１６の処理を繰り返す。

【００８４】

一方、ｔがＴを超えた場合には、ステップ３２０へ移行し、計算された音源時間周波数マスク変数ξ_{tfk}を音源時間周波数マスク変数記憶部３３に保存して更新し、音源時間周波数マスク変数計算処理ルーチンを終了して、音源分離定位処理ルーチンへリターンする。

【００８５】

次に、音源時定位変数計算処理ルーチン（図８）では、ステップ４０２で、記憶部３から必要な情報をロードし、次に、ステップ４０４で、各マスクに対応する変数ｋを１にセットする。

【００８６】

次に、ステップ４０６で、式（９）により、ｄ＝１，...，Ｄについて音源定位変数η_{kd}を計算する。

【００８７】

【数７】

$$\eta_{kd} = \exp\left[\Psi(\kappa_d) - \Psi\left(\sum_{d=1}^{D}\kappa_d\right) + \sum_{t=1}^{T}\sum_{f=1}^{F}\xi_{tfk}\left\{M\left(\Psi(a_{tfk}) - \log b_{tfk}\right) + \sum_{m=0}^{M-1}\Psi(v_{fd} - m) + \log|G_{fd}| - \frac{a_{tfk}}{b_{tfk}}x_{td}^{H}v_{fd}G_{fd}x_{tf}\right\}\right] \qquad (9)$$

【００８８】

音源定位変数η_{kd}は、複数音源の定位、すなわち各音源のマイクロフォンアレイに対する方向を推定するために計算する変数である。ｋ＝１，...，Ｋ、ｄ＝１，...，Ｄとする。η_{kd}は音源ｋの方向がｄ番目の離散化された方向にある確率を表す。この変数に従って各音源の方向を推定することができる。この変数は、本実施の形態では多項分布から生成されると仮定しており、その多項分布のパラメータは統計量κでパラメタライズされたディリクレ分布によって決定されるものとする。

【００８９】

式（９）のポイントは、右辺第３項にあるように、音源時間周波数マスク変数ξ_{tfk}が必要であるという点である。これは、音源分離の情報を使って音源定位が改善されることを表している。

【００９０】

次に、ステップ４０８で、上記ステップ４０６でｄ＝１，...，Ｄについて計算された音源定位変数η_{kd}を、式（１０）により正規化する。η_{kd}は確率であるので、ｋ毎に、全てのｄに対する和が常に１となるように正規化される。

【００９１】

【数８】

$$\eta_{kd} = \frac{\eta_{kd}}{\sum_{d'=1}^{D} \eta_{kd'}} \qquad (10)$$

【００９２】

次に、ステップ４１０で、ｋを１インクリメントして、次のステップ４１２で、ｋが設定された最大マスク数Ｋを超えたか否かを判定し、ｋが未だＫに到達していない場合には、ステップ４０６へ戻って、ステップ４０６～４１０の処理を繰り返す。

【００９３】

一方、ｋがＫを超えた場合には、ステップ４１４へ移行し、計算された音源定位変数η_{kd}を音源定位変数記憶部３４に保存して更新し、音源定位変数計算処理ルーチンを終了して、音源分離定位処理ルーチンへリターンする。

【００９４】

次に、統計量計算処理ルーチン（図９）では、音源時間周波数マスク変数計算処理ルーチン及び音源定位変数計算処理ルーチンで用いる各統計量を計算する。まず、ステップ５０２で、記憶部３から必要な情報をロードする。

【００９５】

次に、ステップ５０４で、音源時間周波数マスク変数の数学モデルに利用されるディリクレ分布のパラメータであるβ_{tk}を、ｔ＝１，...，Ｔ及びｋ＝１，...，Ｋについて、式（１１）により計算する。β_{tk}は、直感的には時間フレームｔにおいて、各周波数ビンｆ上の観測信号が音源ｋからの信号で説明される可能性の強さを表すパラメータであり、０より大きい値となる。。

【００９６】

【数９】

$$\beta_{tk} = \beta_k^0 + \sum_{f=1}^{F} \xi_{tfk} \qquad (11)$$

【００９７】

次に、ステップ５０６で、音源定位変数の数学モデルに利用されるディリクレ分布のパラメータであるκ_dを、ｄ＝１，...，Ｄについて、式（１２）により計算する。κ_dは、直感的には音源ｋが方向ｄに存在する可能性の強さを表すパラメータであり、０より大きい値となる。

【００９８】

【数１０】

$$\kappa_d = \kappa^0 + \sum_{k=1}^{K} \eta_{kd} \qquad (12)$$

【００９９】

次に、ステップ５０８で、時間周波数領域観測信号の数学モデルで利用されるガンマ分布のパラメータであるa_{tfk}を、ｔ＝１，...，Ｔ、ｆ＝１，...，Ｆ、及びｋ＝１，...，Ｋについて、式（１３）により計算する。

【０１００】

【数１１】

$$a_{tfk} = a_{tf}^0 + M\xi_{tfk} \qquad (13)$$

【０１０１】

次に、ステップ５１０で、時間周波数領域観測信号の数学モデルで利用されるガンマ分布のパラメータであるb_{tfk}を、ｔ＝１，...，Ｔ、ｆ＝１，...，Ｆ、及びｋ＝１，...，Ｋについて、式（１４）により計算する。

【０１０２】

【数１２】

$$b_{tfk} = b_{tf}^0 + \xi_{tfk} \sum_{d=1}^{D} \eta_{kd} v_{fd} x_{tf}^H G_{fd} x_{tf} \qquad (14)$$

【０１０３】

次に、ステップ５１２で、時間周波数領域観測信号の数学モデルで利用される複素ウィシャート分布のパラメータであるv_{fd}を、ｆ＝１，...，Ｆ及びｄ＝１，...，Ｄについて、式（１５）により計算する。

【０１０４】

【数１３】

$$v_{fd} = v_{fd}^0 + \sum_{t=1}^{T} \sum_{k=1}^{K} \xi_{tfk} \eta_{kd} \qquad (15)$$

【０１０５】

次に、ステップ５１４で、時間周波数領域観測信号の数学モデルで利用される複素ウィシャート分布のパラメータであるG_{fd}を、ｆ＝１，...，Ｆ及びｄ＝１，...，Ｄについて、式（１６）により計算する。G_{fd}は、無響ステアリングベクトルの情報を取り込んだ、実際の有響環境の音響情報を含む行列である。

【０１０６】

【数１４】

$$G_{fd}=\left(G_{fd}^{-1}+\sum_{t=1}^{T}\sum_{k=1}^{K}\xi_{tfk}\eta_{kd}\frac{a_{tfk}}{b_{tfk}}x_{tf}x_{tf}^{H}\right)^{-1} \qquad (16)$$

【０１０７】
次に、ステップ５１６で、上記ステップ５０４～５１４で計算した各統計量を、統計量記憶部３５に保存して更新し、統計量計算処理ルーチンを終了して、音源分離定位処理ルーチンにリターンする。

【０１０８】
音源分離定位処理ルーチンでは、次に、ステップ６００へ移行し、収束条件判定部２５が、記憶部３に保存された各値を監視して、計算の収束条件が満たされたか否かを判定する。収束条件は反復計算の繰り返し回数など任意に設定してよいが、変分ベイズ法に基づく解析計算を行う本実施の形態では、例えば、式（１７）に示すような収束条件を用いることができる。ただし、ξ’は更新前のξの値を表す。この収束条件を用いた場合には、必ず各値の計算が収束することが知られている。

【０１０９】
【数１５】

$$\frac{1}{TF}\sum_{t=1}^{T}\sum_{f=1}^{F}\left\|\xi_{tfk}-\xi'_{tfk}\right\| \qquad (17)$$

【０１１０】
なお、式（１７）では、音源時間周波数マスク変数ξ_{thk}の変化幅を収束条件として用いているが、音源定位変数η_{kd}の変化幅を収束条件として用いてもよい。

【０１１１】
収束条件を満たしていない場合には、ステップ３００へ戻り、各値の計算を繰り返す。一方、収束条件を満たした場合には、ステップ７００へ移行する。

【０１１２】
ステップ７００では、分離音抽出部４１が、観測信号x_{tf}及び音源時間周波数マスク変数ξ_{tfk}を利用することで、各音源に対応した時間周波数領域での分離音信号を計算する。まず、Ｋ個の音源マスク数、すなわち仮想的な最大の音源数に対して、Ｎ＜＝Ｋとなる抽出音源数Ｎを決定する。これは事前に指定しておいてもよいし、解析終了後に記憶部３の情報を利用して何らかの決定則に基づいて自動的または人手で決定してもよい。

【０１１３】
同時に、音源時間周波数マスク変数ξ_{tfk}のインデックスの順番を入れ替える。具体的に入れ替えるインデックスはξ_{tfk}の最後のインデックスであるｋである。インデックスの順番を入れ替える方法は、下記に示すように、音源インデックスｋ毎に全てのマスク変数の総和を計算し、この総和が大きい順番に入れ替える。

【０１１４】

【数１６】

$$\sum_{t=1}^{T}\sum_{f=1}^{F}\xi_{tf1} > \sum_{t=1}^{T}\sum_{f=1}^{F}\xi_{tf2} > \cdots > \sum_{t=1}^{T}\sum_{f=1}^{F}\xi_{tfk} > \cdots > \sum_{t=1}^{T}\sum_{f=1}^{F}\xi_{tfK}$$

【０１１５】

上記のインデックスの並べ替えは、直観的には、音源ｋを予想される音量が大きい順番に入れ替えることに相当する。入れ替えた音源のインデックスをｎで表す。

【０１１６】

そして、入れ替えたインデックスのもとで、式（１８）を利用して、ｎ番目に音量の大きい音源の時間フレームｔ、周波数ビンｆでの時間周波数領域の分離音信号y_{tf}^{n}を計算する。

【０１１７】

【数１７】

$$y_{tf}^{n} = \frac{\xi_{tfn}}{\sum_{k=1}^{N}\xi_{tfk}} x_{tf} \qquad (18)$$

【０１１８】

式（１８）の意味は、右辺第１項の時間周波数マスク変数ξ_{tfk}の分数によって、Ｎ個の音源の中で各（ｔ，ｆ）における音源ｎの音が占める割合を計算し、この割合で混合音の時間周波数領域表現であるx_{tf}を分配するというものである。全てのｔ＝１，...，Ｔ、ｆ＝１，...，Ｆ、及びｎ＝１，...，Ｎに対して式（１８）の計算が終了したら、音源毎の時間周波数領域の分離音信号の計算結果を音声波形復元部４２へ渡すと共に、入れ替えた音源インデックス情報ｎを音源方向抽出部４３へ渡す。

【０１１９】

次に、ステップ７０２で、音声波形復元部４２が、上記ステップ７００において分離音抽出部４１より受け取った分離音信号y_{tf}^{n}を変換して、通常の音声波形を復元する。具体的には、時間周波数領域観測変換部１２の逆変換である逆フーリエ変換を音源ｎ毎に実施する。

【０１２０】

次に、ステップ７０４で、音源方向抽出部４３が、上記ステップ７００において分離音抽出部４１より受け取った入れ替えた音源インデックスｎ、及び音源定位変数η_{kd}を利用して、Ｎ個の音源の方向を計算する。具体的には、各ｎに対して式（１９）の計算を行えばよい。これによって音源ｎの存在する方向のインデックスd_nを求めることができる。

【０１２１】

【数１８】

$$d_n = \arg\max_{d} \eta_{nd} \qquad (19)$$

【０１２２】

次に、ステップ７０６で、最終出力部４４が、記憶部３、分離音抽出部４１、音声波形復元部４２、及び音源方向抽出部４３の情報を用いて、ユーザの所望の形で解析結果を出力して、音源分離定位処理ルーチンを終了する。

【０１２３】

なお、上記ステップ７０２及び７０４の処理はいずれを先に行ってもよい。

【０１２４】

以上説明したように、本実施の形態に係る音源分離定位装置によれば、複数音源の各々から発せられた音の混合音を観測した際に、各音源への音源分離と音源の方向定位とを、同時に一つの統計的枠組みによって解決することにより、既存手法のように「一方の問題で失敗した結果、他方の問題まで失敗する」という状況を回避し、両方の問題に対して、安定して高い性能を得ることができる。

【０１２５】

また、音源分離と音源定位とは相互依存の問題であるため、両問題を同時に解決することにより、各問題に対して個別に解決するよりも高い精度を得ることができる。

【０１２６】

さらに、複数音源の同時分離及び定位に、混合音を観測するマイクロフォンアレイについて事前に計測した無響環境のステアリングベクトルを利用することにより、現実の未知有響環境下でステアリングベクトルを再計測することなく、様々な環境に適合して、音源分離及び定位を実施することができる。

【０１２７】

＜実験例＞

次に、本実施の形態に係る音源分離定位装置における実験の結果について説明する。

【０１２８】

図１０に実験のセットアップを示す。本実験では、音源数Ｎ＝２または３を既知として、マイク数Ｍ＝２，４，８のパターンで各音源の分離及び定位の性能評価を行う。実験では、離散化した方向の数はＤ＝７２、すなわち５度おきに方向を区分けする。

【０１２９】

図１１は音源定位の性能を示すグラフである。同図（ａ）は音源数が２、（ｂ）は音源数が３の場合である。図中RT_{60}は、混合音の観測環境の残響時間を表している。これによれば、残響が長くなるとその分定位性能が落ちることがわかる。しかし、マイク数が多い、例えばＭ＝８の場合には、定位誤差はほとんどゼロで済むという結果である。すなわち、特にマイク数が多い場合、本実施の形態に係る音源分離定位装置によれば、高精度に音源の定位を実現することができる。

【０１３０】

図１２は音源分離の性能を示すグラフである。ここでは従来手法との比較を行った。比較対象は、マイク数と音源数との組み合わせによって変更する。まず、Ｍ＞＝Ｎの場合、すなわちマイク数の方が音源よりも多い場合は非特許文献３のＩＶＡ法を利用する。一方、Ｍ＜Ｎ、すなわちマイク数の方が音源よりも少ない場合は非特許文献１の方法（ＴＦ－ｐｅｒｍ）を利用する。

【０１３１】

図１２では、（ａ）、（ｃ）、及び（ｅ）は音源数が２の場合、（ｂ）、（ｄ）、及び（ｆ）は音源数が３の場合である。また、（ａ）及び（ｂ）はRT_{60}＝２０ｍｓｅｃ、（ｃ）及び（ｄ）はRT_{60}＝４００ｍｓｅｃ、（ｅ）及び（ｆ）はRT_{60}＝６００ｍｓｅｃである。すなわち、上から下の行に移る毎に残響時間が長い環境での実験結果を示している。（ｂ）、（ｄ）、及び（ｆ）ではマイク数が２のときに非特許文献１の手法の結果を掲載している。

【０１３２】

同図より明らかなように、ほぼ全ての実験環境及びマイク数で、本実施の形態に係る音源分離定位装置は、従来手法よりも良い分離精度を達成することができている。これは、音源分離を音源定位と同時に問題を解決することで、より良い音源分離が達成できること

を意味しており、本発明の有効性を示すものである。

【０１３３】

なお、本発明は、上述した実施形態に限定されるものではなく、この発明の要旨を逸脱しない範囲内で様々な変形や応用が可能である。

【０１３４】

例えば、上述の音源分離定位装置は、内部にコンピュータシステムを有しているが、「コンピュータシステム」は、ＷＷＷシステムを利用している場合であれば、ホームページ提供環境（あるいは表示環境）も含むものとする。

【０１３５】

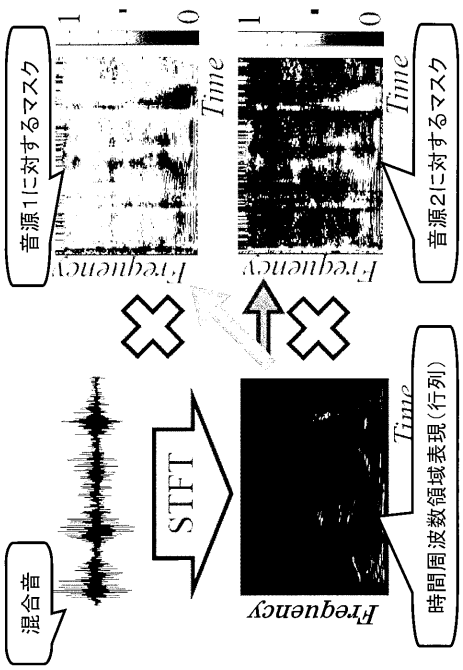
また、本願明細書中において、プログラムが予めインストールされている実施形態として説明したが、当該プログラムを、コンピュータ読み取り可能な記録媒体に格納して提供することも可能である。

【符号の説明】

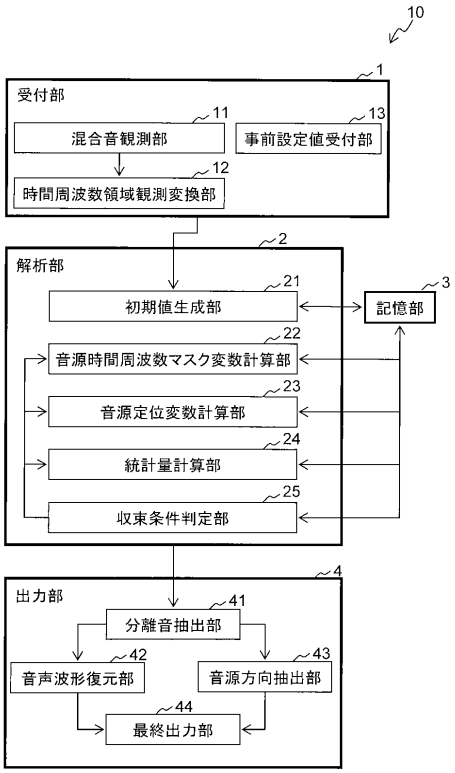
【０１３６】

１　受付部
２　解析部
３　記憶部
４　出力部
１０　音源分離定位装置
１１　混合音観測部
１２　時間周波数領域観測変換部
１３　事前設定値受付部
２１　初期値生成部
２２　音源時間周波数マスク変数計算部
２３　音源定位変数計算部
２３　音源分離変数計算部
２４　統計量計算部
２５　収束条件判定部
４１　分離音抽出部
４２　音声波形復元部
４３　音源方向抽出部
４４　最終出力部

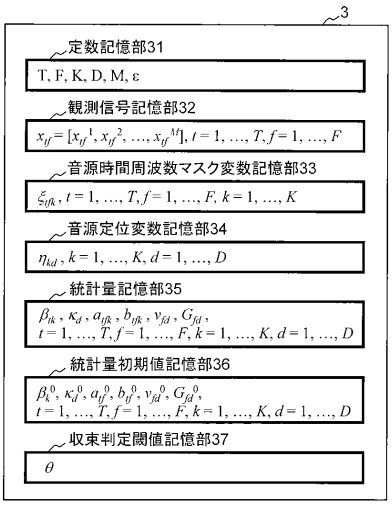
【図１】


音源1に対するマスク
音源2に対するマスク
Time
Frequency
STFT
混合音
時間周波数領域表現（行列）

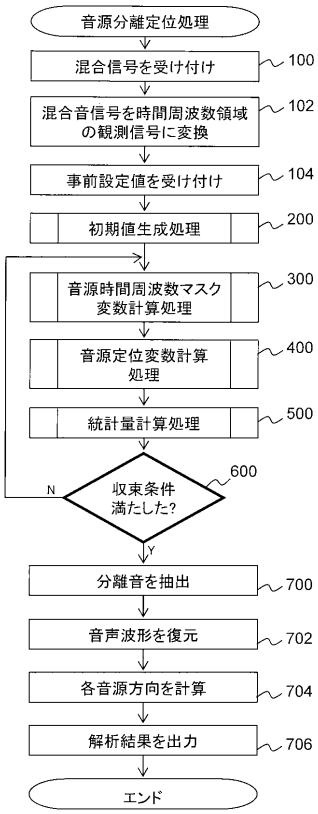
【図３】


10
1
受付部
11
混合音観測部
13
事前設定値受付部
12
時間周波数領域観測変換部
2
解析部
21
初期値生成部
3
記憶部
22
音源時間周波数マスク変数計算部
23
音源定位変数計算部
24
統計量計算部
25
収束条件判定部
4
出力部
41
分離音抽出部
42
音声波形復元部
43
音源方向抽出部
44
最終出力部

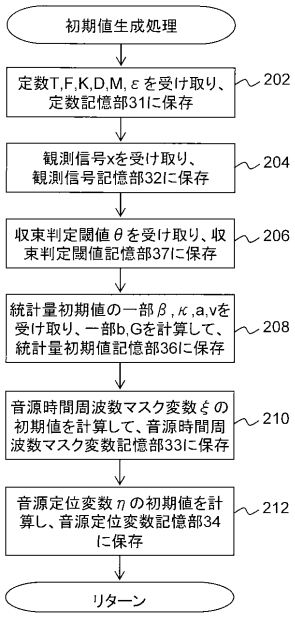
【図４】


3
定数記憶部31
T, F, K, D, M, ε
観測信号記憶部32
$x_{tf}=[x_{tf}^1, x_{tf}^2, \ldots, x_{tf}^M], t=1,\ldots,T, f=1,\ldots,F$
音源時間周波数マスク変数記憶部33
$\xi_{tfk}, t=1,\ldots,T, f=1,\ldots,F, k=1,\ldots,K$
音源定位変数記憶部34
$\eta_{kd}, k=1,\ldots,K, d=1,\ldots,D$
統計量記憶部35
$\beta_{tk}, \kappa_d, a_{tfk}, b_{tfk}, v_{fd}, G_{fd}$,
$t=1,\ldots,T, f=1,\ldots,F, k=1,\ldots,K, d=1,\ldots,D$
統計量初期値記憶部36
$\beta_k^0, \kappa_d^0, a_{tf}^0, b_{tf}^0, v_{fd}^0, G_{fd}^0$,
$t=1,\ldots,T, f=1,\ldots,F, k=1,\ldots,K, d=1,\ldots,D$
収束判定閾値記憶部37
θ

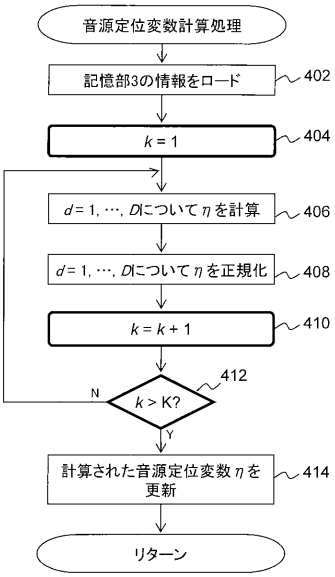
【図５】


音源分離定位処理
混合信号を受け付け
100
混合音信号を時間周波数領域の観測信号に変換
102
事前設定値を受け付け
104
初期値生成処理
200
音源時間周波数マスク変数計算処理
300
音源定位変数計算処理
400
統計量計算処理
500
収束条件満たした?
600
N
Y
分離音を抽出
700
音声波形を復元
702
各音源方向を計算
704
解析結果を出力
706
エンド

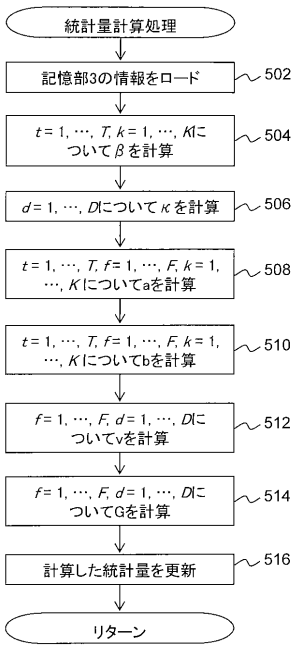
【図６】


初期値生成処理
定数T,F,K,D,M,εを受け取り、定数記憶部31に保存
202
観測信号xを受け取り、観測信号記憶部32に保存
204
収束判定閾値θを受け取り、収束判定閾値記憶部37に保存
206
統計量初期値の一部β,κ,a,vを受け取り、一部b,Gを計算して、統計量初期値記憶部36に保存
208
音源時間周波数マスク変数ξの初期値を計算して、音源時間周波数マスク変数記憶部33に保存
210
音源定位変数ηの初期値を計算し、音源定位変数記憶部34に保存
212
リターン

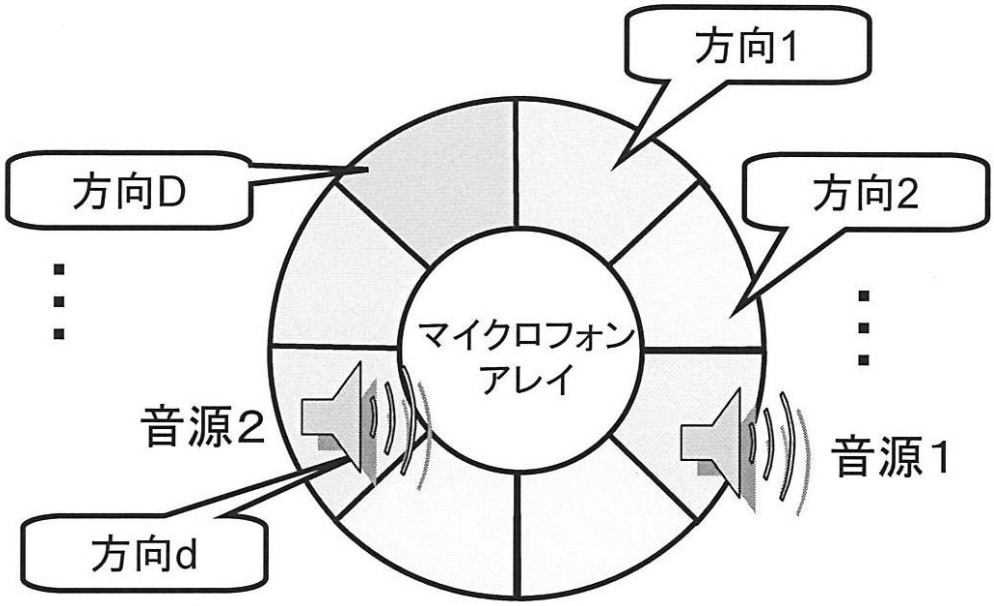
【図８】


音源定位変数計算処理
記憶部3の情報をロード
402
k = 1
404
d = 1, ⋯, Dについてηを計算
406
d = 1, ⋯, Dについてηを正規化
408
k = k + 1
410
k > K?
412
N
Y
計算された音源定位変数ηを更新
414
リターン

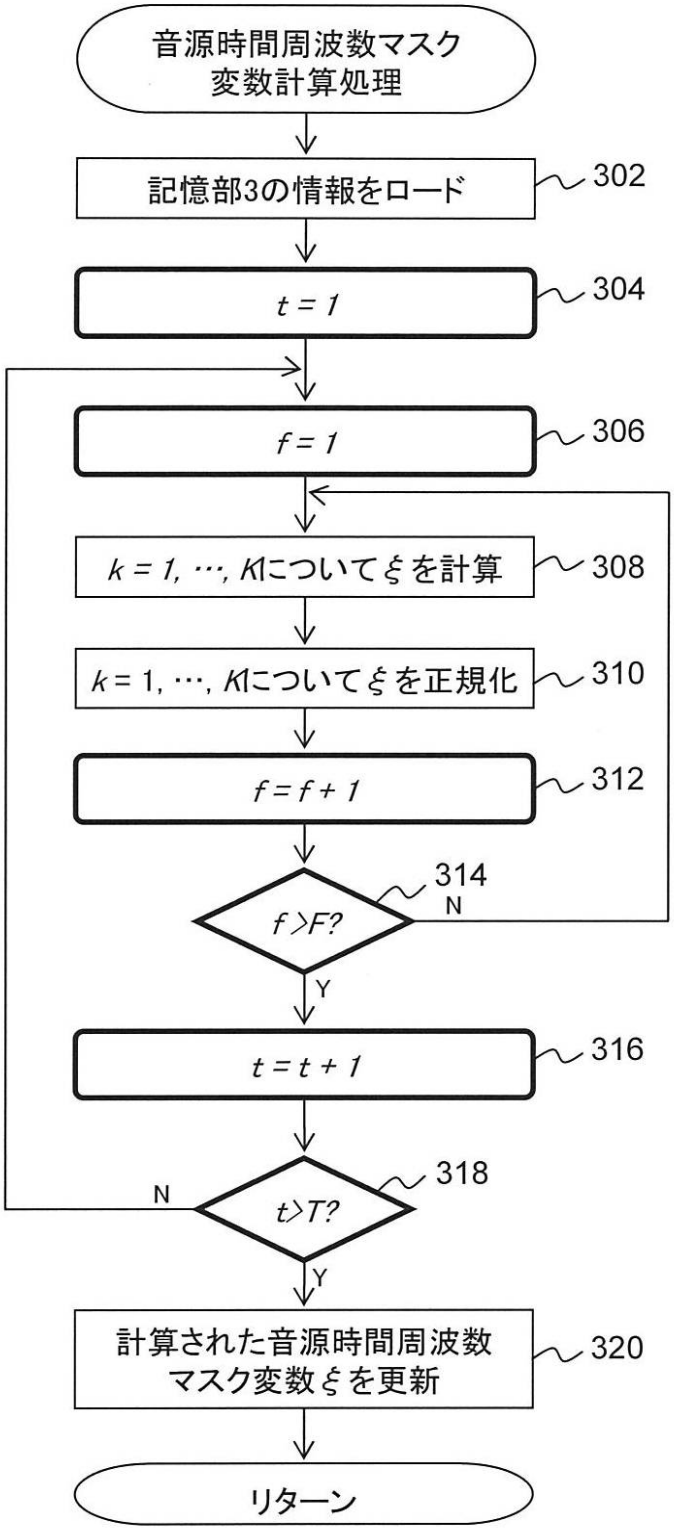
【図９】


統計量計算処理
記憶部3の情報をロード
502
t = 1, ⋯, T, k = 1, ⋯, Kについてβを計算
504
d = 1, ⋯, Dについてκを計算
506
t = 1, ⋯, T, f = 1, ⋯, F, k = 1, ⋯, Kについてaを計算
508
t = 1, ⋯, T, f = 1, ⋯, F, k = 1, ⋯, Kについてbを計算
510
f = 1, ⋯, F, d = 1, ⋯, Dについてvを計算
512
f = 1, ⋯, F, d = 1, ⋯, Dについて Gを計算
514
計算した統計量を更新
516
リターン

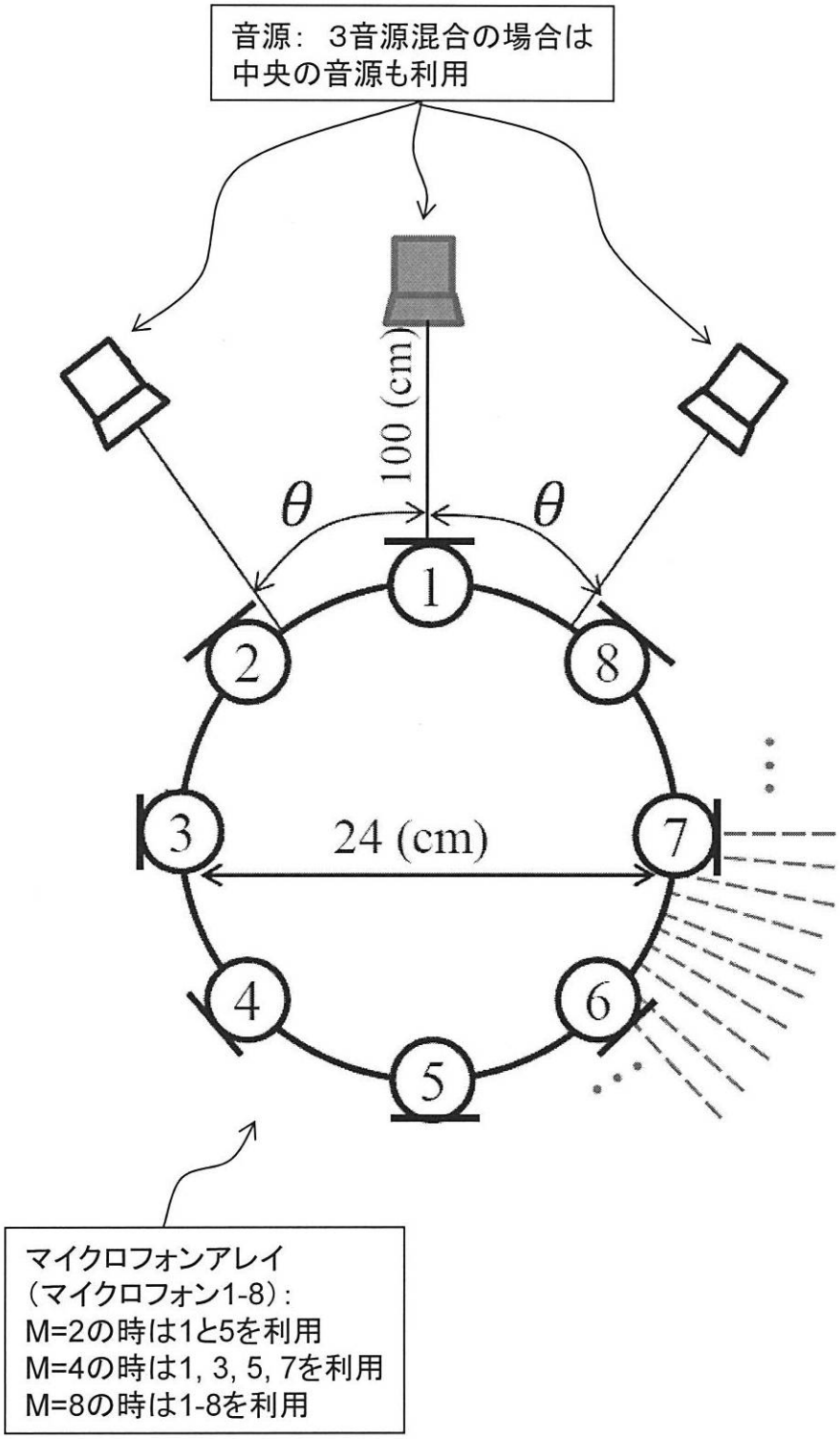
【図２】



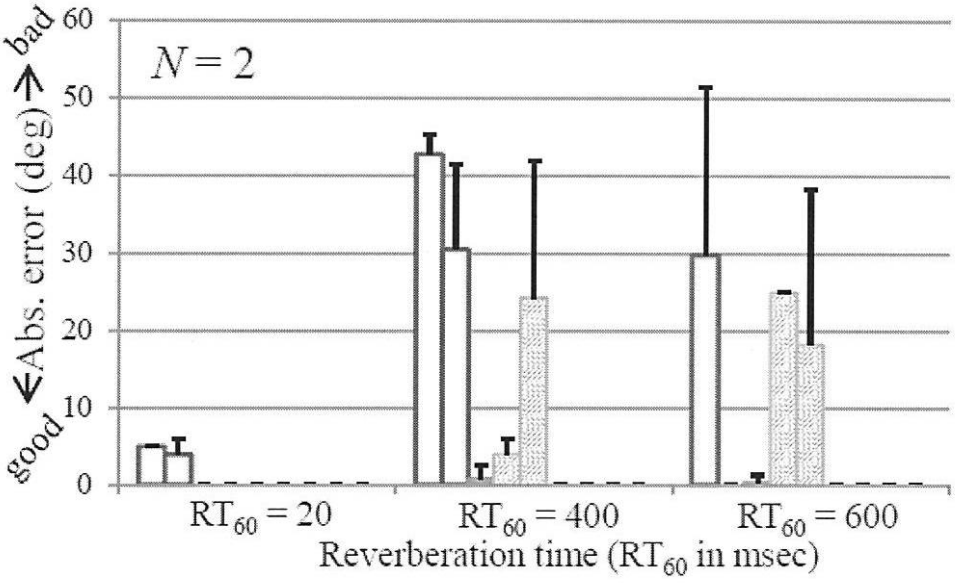
【図７】


音源時間周波数マスク
変数計算処理
記憶部3の情報をロード
302
t = 1
304
f = 1
306
k = 1, ⋯, Kについてξを計算
308
k = 1, ⋯, Kについてξを正規化
310
f = f + 1
312
f>F?
314
N
Y
t = t + 1
316
t>T?
318
N
Y
計算された音源時間周波数
マスク変数ξを更新
320
リターン

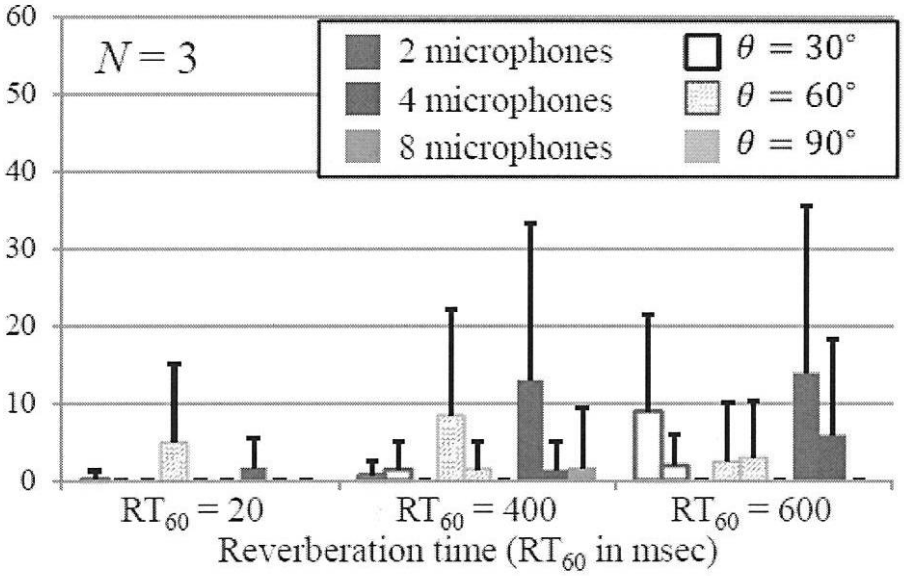
【図１０】


音源：３音源混合の場合は
中央の音源も利用
100 (cm)
θ
θ
1
2
8
3
24 (cm)
7
4
6
5
マイクロフォンアレイ
（マイクロフォン1-8）：
M=2の時は1と5を利用
M=4の時は1, 3, 5, 7を利用
M=8の時は1-8を利用

【図１１】

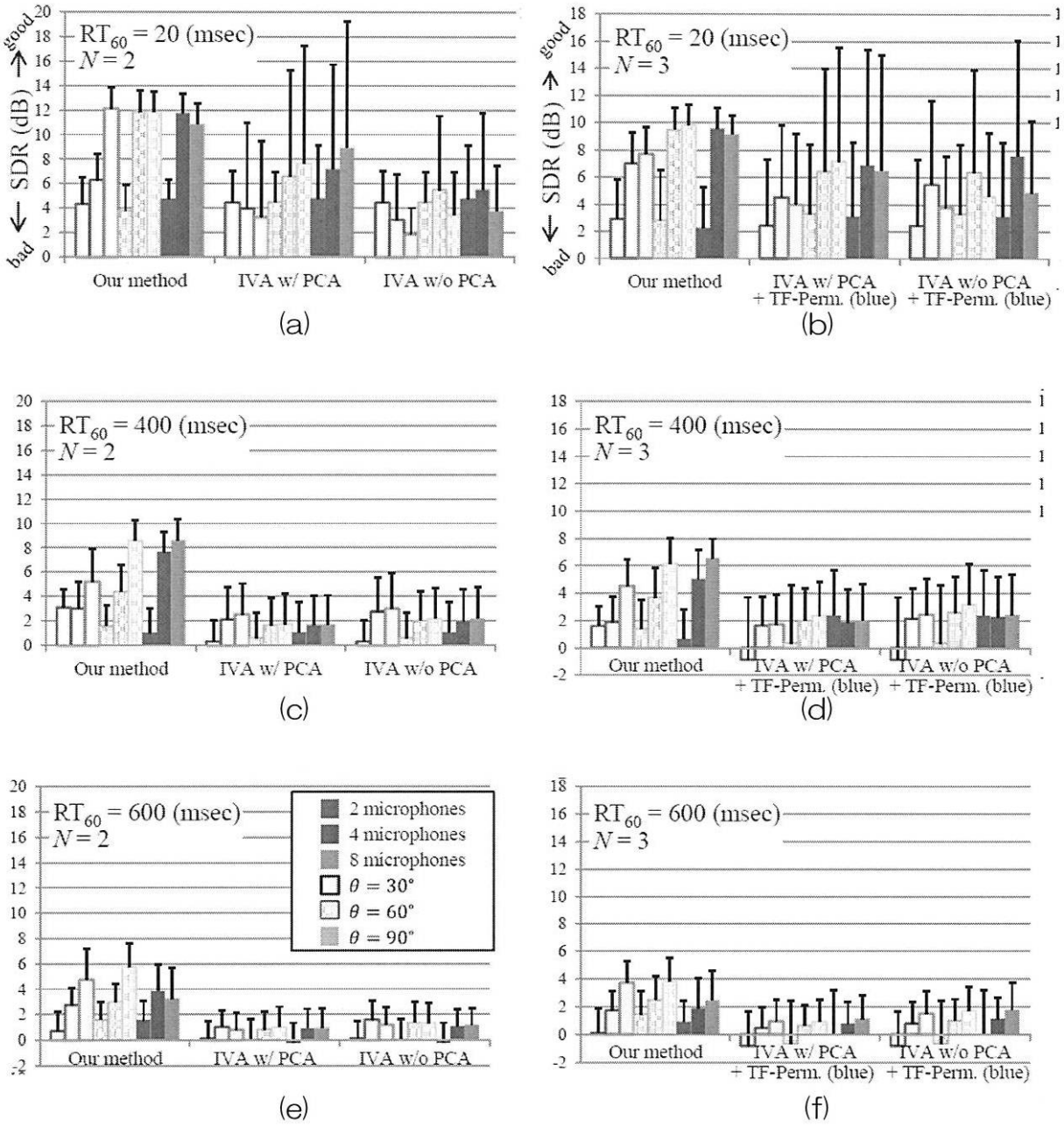

N = 2
Abs. error (deg)
good
bad
60
50
40
30
20
10
0
RT60 = 20
RT60 = 400
RT60 = 600
Reverberation time (RT60 in msec)

(a)


N = 3
2 microphones
4 microphones
8 microphones
θ = 30°
θ = 60°
θ = 90°
60
50
40
30
20
10
0
RT60 = 20
RT60 = 400
RT60 = 600
Reverberation time (RT60 in msec)

(b)

【図１２】


RT60 = 20 (msec)
N = 2
good
bad
SDR (dB)
Our method
IVA w/ PCA
IVA w/o PCA
(a)
RT60 = 20 (msec)
N = 3
IVA w/ PCA + TF-Perm. (blue)
IVA w/o PCA + TF-Perm. (blue)
(b)
RT60 = 400 (msec)
N = 2
(c)
RT60 = 400 (msec)
N = 3
(d)
RT60 = 600 (msec)
N = 2
2 microphones
4 microphones
8 microphones
θ = 30°
θ = 60°
θ = 90°
(e)
RT60 = 600 (msec)
N = 3
(f)

フロントページの続き

(72)発明者　大塚　琢馬
京都府京都市左京区吉田本町　国立大学法人京都大学大学院情報学研究科内
(72)発明者　奥乃　博
京都府京都市左京区吉田本町　国立大学法人京都大学大学院情報学研究科内