

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第5791081号
(P5791081)

(45) 発行日 平成27年10月7日(2015.10.7)

(24) 登録日 平成27年8月14日(2015.8.14)

(51) Int.Cl. F I
G 1 O L 21/028 (2013.01) G 1 O L 21/028 B
G 1 O L 21/0308 (2013.01) G 1 O L 21/0308 Z

請求項の数 5 (全 26 頁)

(21) 出願番号	特願2012-160450 (P2012-160450)	(73) 特許権者	000004226 日本電信電話株式会社 東京都千代田区大手町一丁目5番1号
(22) 出願日	平成24年7月19日(2012.7.19)	(73) 特許権者	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
(65) 公開番号	特開2014-21315 (P2014-21315A)	(74) 代理人	110001519 特許業務法人太陽国際特許事務所
(43) 公開日	平成26年2月3日(2014.2.3)	(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
審査請求日	平成26年7月31日(2014.7.31)	(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) 【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
含み、

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 $x_{t,f}$ 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 $x_{t,f}$ 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 $t_{f,k}$ 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 k_d 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 $t_{f,k}$ 及び前記音源定位変数 k_d 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 $t_{f,k}$ の計算に前記音源定位変数 k_d を用い、前記音源定位変数 k_d の計算に前記マスク変数 $t_{f,k}$ を用いる
音源分離定位装置。

【請求項 2】

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 記載の音源分離定位装置。

10

【請求項 3】

受付手段と、音源時間周波数マスク変数計算手段、音源定位変数計算手段、統計量計算手段、及び収束条件判定手段を含む解析手段と、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、

20

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において、

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 $x_{t,f}$ 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 $x_{t,f}$ 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 $t_{f,k}$ 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 k_d を、前記複数の方向の各々について計算し、

30

前記統計量計算手段が、前記マスク変数 $t_{f,k}$ 及び前記音源定位変数 k_d 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 $t_{f,k}$ の計算に前記音源定位変数 k_d を用い、前記音源定位変数 k_d の計算に前記マスク変数 $t_{f,k}$ を用いる
音源分離定位方法。

40

【請求項 4】

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 3 記載の音源分離定位方法。

【請求項 5】

コンピュータを、請求項 1 または請求項 2 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】

50

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

10

【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。

20

【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 Convolutional 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.

30

【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 Convolutional 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。

40

【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

50

、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。

10

【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。

20

【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 $x_{t f}$ 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 $x_{t f}$ 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 $t_{f k}$ 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 k_d 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 $t_{f k}$ 及び前記音源定位変数 k_d 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 $t_{f k}$ の計算に前記音源定位変数 k_d を用い、前記音源定位変数 k_d の計算に前記マスク変数 $t_{f k}$ 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。

30

40

【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】

50

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 $x_{t f}$ 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 $x_{t f}$ 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 $t f k$ 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 $k d$ 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 $t f k$ 及び前記音源定位変数 $k d$ 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 $t f k$ の計算に前記音源定位変数 $k d$ を用い、前記音源定位変数 $k d$ の計算に前記マスク変数 $t f k$ 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。

10

【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 などによって利用されており、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。

20

【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。

30

【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。

40

【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】

50

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。

10

【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】

20

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。

30

【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。

40

【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】

50

観測信号記憶部 3 2 には、時間周波数領域観測変換部 1 2 で変換された観測信号が記憶される。

【 0 0 4 1 】

音源時間周波数マスク変数記憶部 3 3 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。

【 0 0 4 2 】

音源定位変数記憶部 3 4 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。

【 0 0 4 3 】

統計量記憶部 3 5 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。

10

【 0 0 4 4 】

統計量初期値記憶部 3 6 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。

【 0 0 4 5 】

収束判定閾値記憶部 3 7 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。

【 0 0 4 6 】

出力部 4 は、さらに、分離音抽出部 4 1 と、音声波形復元部 4 2 と、音源方向抽出部 4 3 と、最終出力部 4 4 とを含んだ構成で表すことができる。

【 0 0 4 7 】

20

分離音抽出部 4 1 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 4 2 へと渡す。

【 0 0 4 8 】

音声波形復元部 4 2 は、記憶部 3 に保存された情報、及び分離音抽出部 4 1 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。

【 0 0 4 9 】

音源方向抽出部 4 3 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。

【 0 0 5 0 】

30

最終出力部 4 4 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。

【 0 0 5 1 】

< 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 1 0 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 1 0 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。

【 0 0 5 2 】

40

ステップ 1 0 0 で、混合音観測部 1 1 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。

【 0 0 5 3 】

次に、ステップ 1 0 2 で、時間周波数領域観測変換部 1 2 が、上記ステップ 1 0 0 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 (S T F T) あるいは高速フーリエ変換 (F F T) を利用することができる。変換した観測信号を $x_{t f}$ で表す。各 $x_{t f}$ は時間フレーム t ($t = 1, \dots, T$)、フーリエ変換による f ($f = 1, \dots, F$) 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 $x_{t f}$ はマイク数に相当する要素数のベクトルであり、各要素

50

は複素数である。以後、この $x_{t f}$ 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 $x_{t f}$ 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規化定数が含まれる。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、音源数を十分に上回る数を設定する。また、正規化定数は、例えば0.0001と設定することができる。

10

【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(k^0 、 d^0 、 $a_{t f}^0$ 、 $v_{f d}^0$)も受け付ける。 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、 $k = 1, \dots, K$ に対し0より大きい値を設定する。例えば、 $k_1^0 = k_2^0 = \dots = k_K^0$ とすることができる。 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、 $d = 1, \dots, D$ に対し0より大きい値を設定する。例えば、 $d_1^0 = d_2^0 = \dots = d_D^0$ とすることができる。 $a_{t f}^0$ 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。 $v_{f d}^0$ 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、 $t = 1, \dots, T$ 、 $f = 1, \dots, F$ 、 $d = 1, \dots, D$ に対し $a_{t f}^0 = 1$ 、 $v_{f d}^0 = M$ と設定することができる。

20

【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】

$$q_{fd} = [q_{fd}^1, q_{fd}^2, \dots, q_{fd}^M] \quad (1)$$

30

【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ふ。

40

【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も受け付ける。収束判定閾値の値は、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】

50

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」、シュブリンガー・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。

10

【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 、及び正規化定数 を、定数記憶部 31 に保存する。

20

【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 $x_{t f}$ 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 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（ k^0 、 d^0 、 $a_{t f}^0$ 、 $v_{f d}^0$ ）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 $q_{f d}$ 、及び上記ステップ 102 で計算された観測信号 $x_{t f}$ 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 $G_{f d}^0$ 及び $b_{t f}^0$ 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。

30

【0068】

【数 2】

$$G_{fd}^0 = \frac{1}{M} (q_{fd} q_{fd}^H + \epsilon I_M)^{-1} \quad (2)$$

40

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

【0069】

式（2）及び（3）により計算された統計量初期値の一部（ $G_{f d}^0$ 、 $b_{t f}^0$ ）を統計量初期値記憶部 36 に保存する。

【0070】

50

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】

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

10

【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】

$$\eta_{kd}^0 = \begin{cases} \frac{K}{D} & (k-1)\frac{D}{K} \leq d \leq k\frac{D}{K} \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

(6)

20

【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。

30

【0076】

次に、ステップ308で、式(7)により、 $k = 1, \dots, K$ 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、 Γ はディガンマ関数である。

【0077】

【数5】

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

【0078】

音源時間周波数マスク変数 $\xi_{t f k}$ は、観測信号を分離するために計算する変数である。 $t = 1, \dots, T$ 、 $f = 1, \dots, F$ 、 $k = 1, \dots, K$ とする。 $\xi_{t f k}$ 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 $\xi_{t f k}$ 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 $\xi_{t f k}$ 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 $k = 1, \dots, K$ について計算された音源時間周波数マスク変数 $\xi_{t f k}$ を、式(8)により正規化する。 $\xi_{t f k}$ 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】

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

【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。

10

20

30

40

50

【 0 0 8 4 】

一方、 t が T を超えた場合には、ステップ 3 2 0 へ移行し、計算された音源時間周波数マスク変数 $t_{f k}$ を音源時間周波数マスク変数記憶部 3 3 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。

【 0 0 8 5 】

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 4 0 2 で、記憶部 3 から必要な情報をロードし、次に、ステップ 4 0 4 で、各マスクに対応する変数 k を 1 にセットする。

【 0 0 8 6 】

次に、ステップ 4 0 6 で、式（9）により、 $d = 1, \dots, D$ について音源定位変数 η_{kd} を計算する。

【 0 0 8 7 】

【数 7】

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

【 0 0 8 8 】

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。 $k = 1, \dots, K$ 、 $d = 1, \dots, D$ 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。

【 0 0 8 9 】

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 $t_{f k}$ 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。

【 0 0 9 0 】

次に、ステップ 4 0 8 で、上記ステップ 4 0 6 で $d = 1, \dots, D$ 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。

【 0 0 9 1 】

【数 8】

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

【 0 0 9 2 】

次に、ステップ 4 1 0 で、 k を 1 インクリメントして、次のステップ 4 1 2 で、 k が設

10

20

30

40

50

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 κ_d 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。

10

【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、 $t = 1, \dots, T$ 及び $k = 1, \dots, K$ 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】

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

20

【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、 $d = 1, \dots, D$ 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】

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

30

【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、 $t = 1, \dots, T$ 、 $f = 1, \dots, F$ 、及び $k = 1, \dots, K$ について、式(13)により計算する。

40

【0100】

【数11】

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

【0101】

50

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、 $t = 1, \dots, T$ 、 $f = 1, \dots, F$ 、及び $k = 1, \dots, K$ について、式(14)により計算する。

【0102】

【数12】

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

10

【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、 $f = 1, \dots, F$ 及び $d = 1, \dots, D$ について、式(15)により計算する。

【0104】

【数13】

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

20

【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、 $f = 1, \dots, F$ 及び $d = 1, \dots, D$ 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】

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

30

【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、 \cdot は更新前の \cdot 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。

40

50

【 0 1 0 9 】

【 数 1 5 】

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

【 0 1 1 0 】

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 ξ_{kd} の変化幅を収束条件として用いてもよい。 10

【 0 1 1 1 】

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。

【 0 1 1 2 】

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、 $N \leq K$ 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。 20

【 0 1 1 3 】

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。

【 0 1 1 4 】

【 数 1 6 】

$$\sum_{t=1}^T \sum_{f=1}^F \xi_{tf1} > \sum_{t=1}^T \sum_{f=1}^F \xi_{tf2} > \dots > \sum_{t=1}^T \sum_{f=1}^F \xi_{tfk} > \dots > \sum_{t=1}^T \sum_{f=1}^F \xi_{tfK} \quad 30$$

【 0 1 1 5 】

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。

【 0 1 1 6 】

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。 40

【 0 1 1 7 】

【 数 1 7 】

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

【 0 1 1 8 】

式(18)の意味は、右辺第1項の時間周波数マスク変数 t_{fk} 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 $t = 1, \dots, T$ 、 $f = 1, \dots, F$ 、及び $n = 1, \dots, N$ 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。

【 0 1 1 9 】

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。

10

【 0 1 2 0 】

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 k_d 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。

【 0 1 2 1 】

【 数 1 8 】

20

$$d_n = \arg \max_d \eta_{nd} \quad (19)$$

【 0 1 2 2 】

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。

【 0 1 2 3 】

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。

30

【 0 1 2 4 】

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。

【 0 1 2 5 】

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。

40

【 0 1 2 6 】

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。

【 0 1 2 7 】

< 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。

【 0 1 2 8 】

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

50

、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、 $M \geq N$ 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 I V A 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 (T F - p e r m) 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 $RT_{60} = 20 \text{ msec}$ 、(c) 及び (d) は $RT_{60} = 400 \text{ msec}$ 、(e) 及び (f) は $RT_{60} = 600 \text{ msec}$ 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

10

20

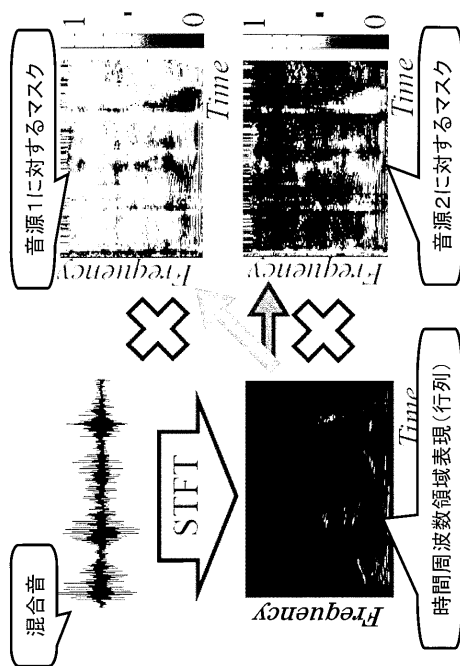
30

40

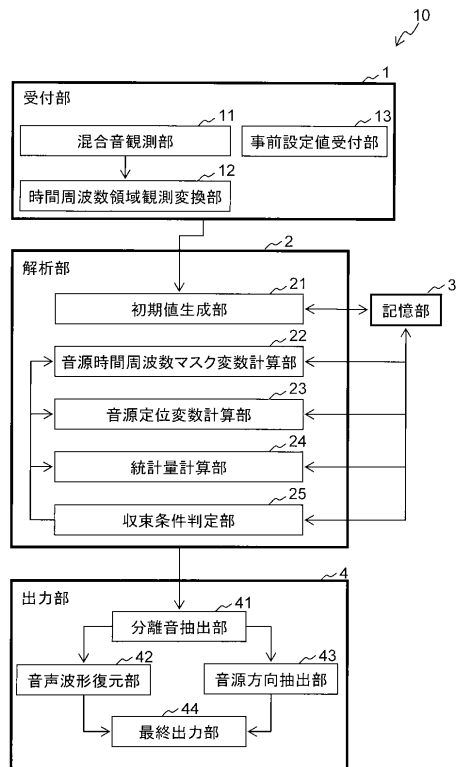
50

- 2 3 音源定位変数計算部
- 2 3 音源分離変数計算部
- 2 4 統計量計算部
- 2 5 収束条件判定部
- 4 1 分離音抽出部
- 4 2 音声波形復元部
- 4 3 音源方向抽出部
- 4 4 最終出力部

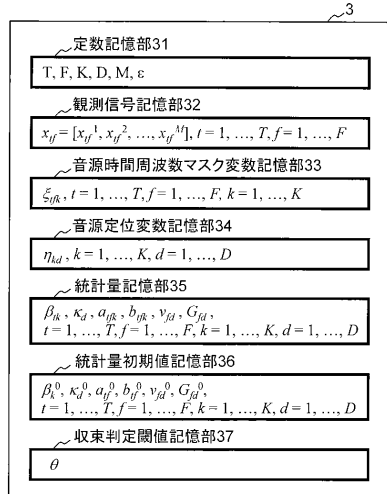
【図1】



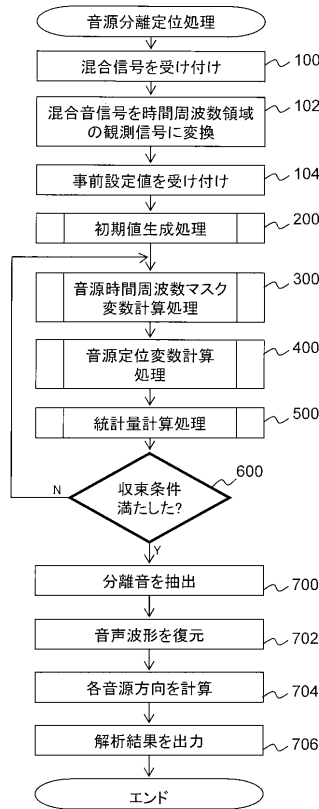
【図3】



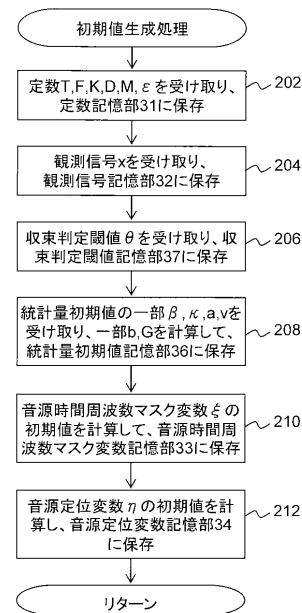
【図4】



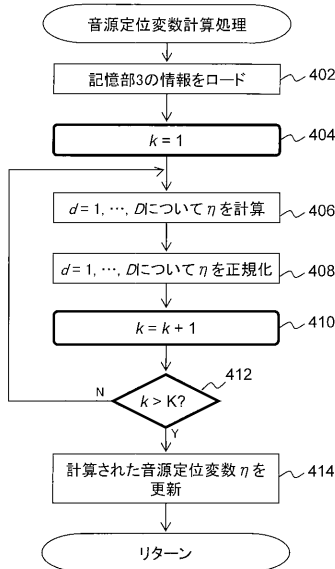
【図5】



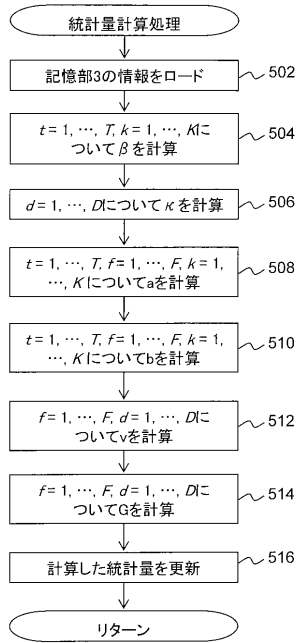
【図6】



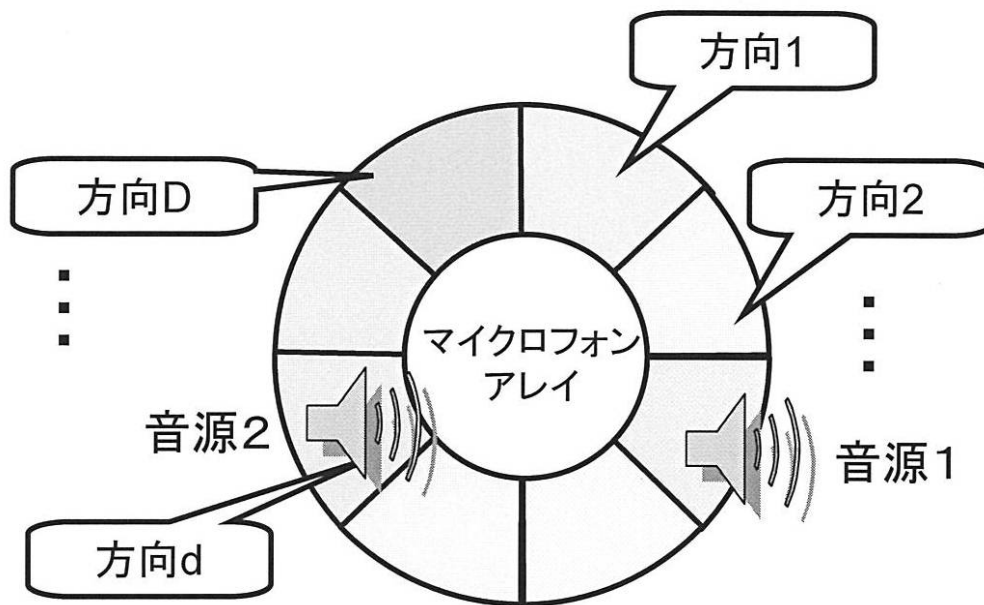
【図8】



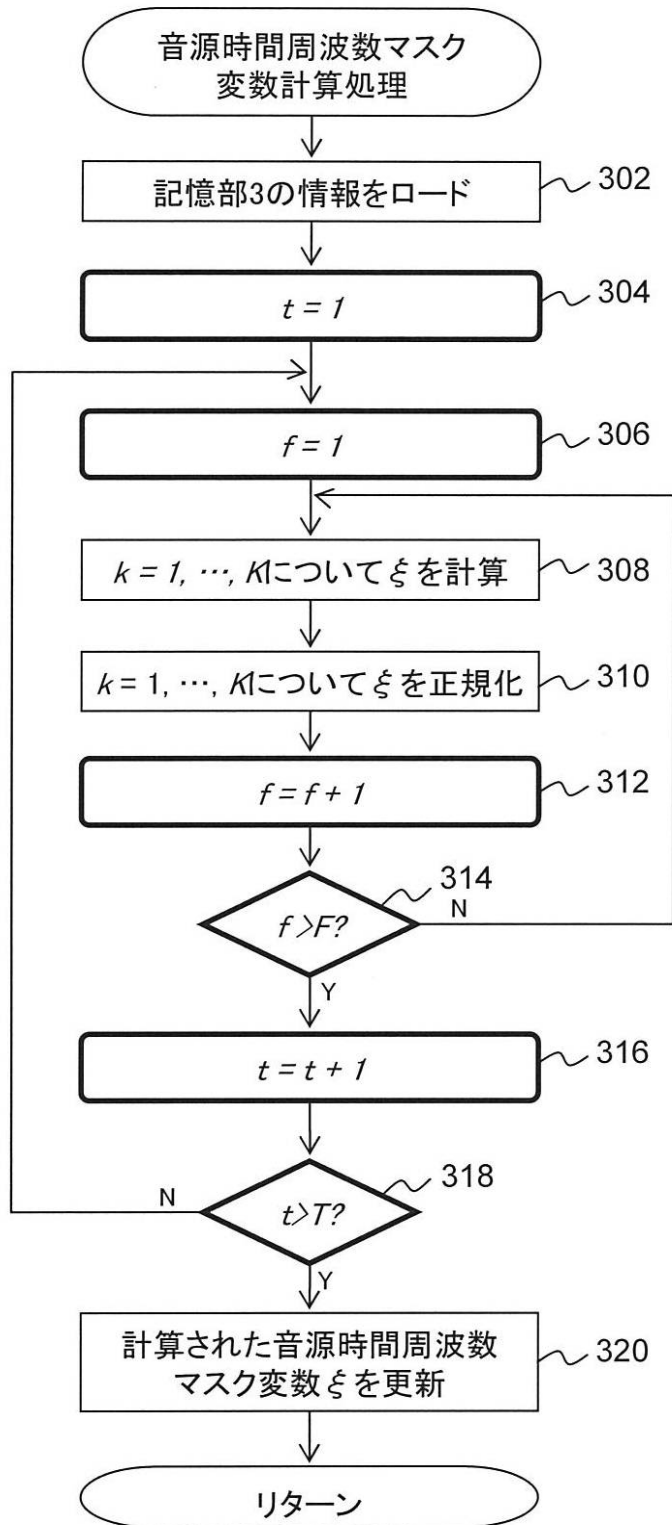
【図9】



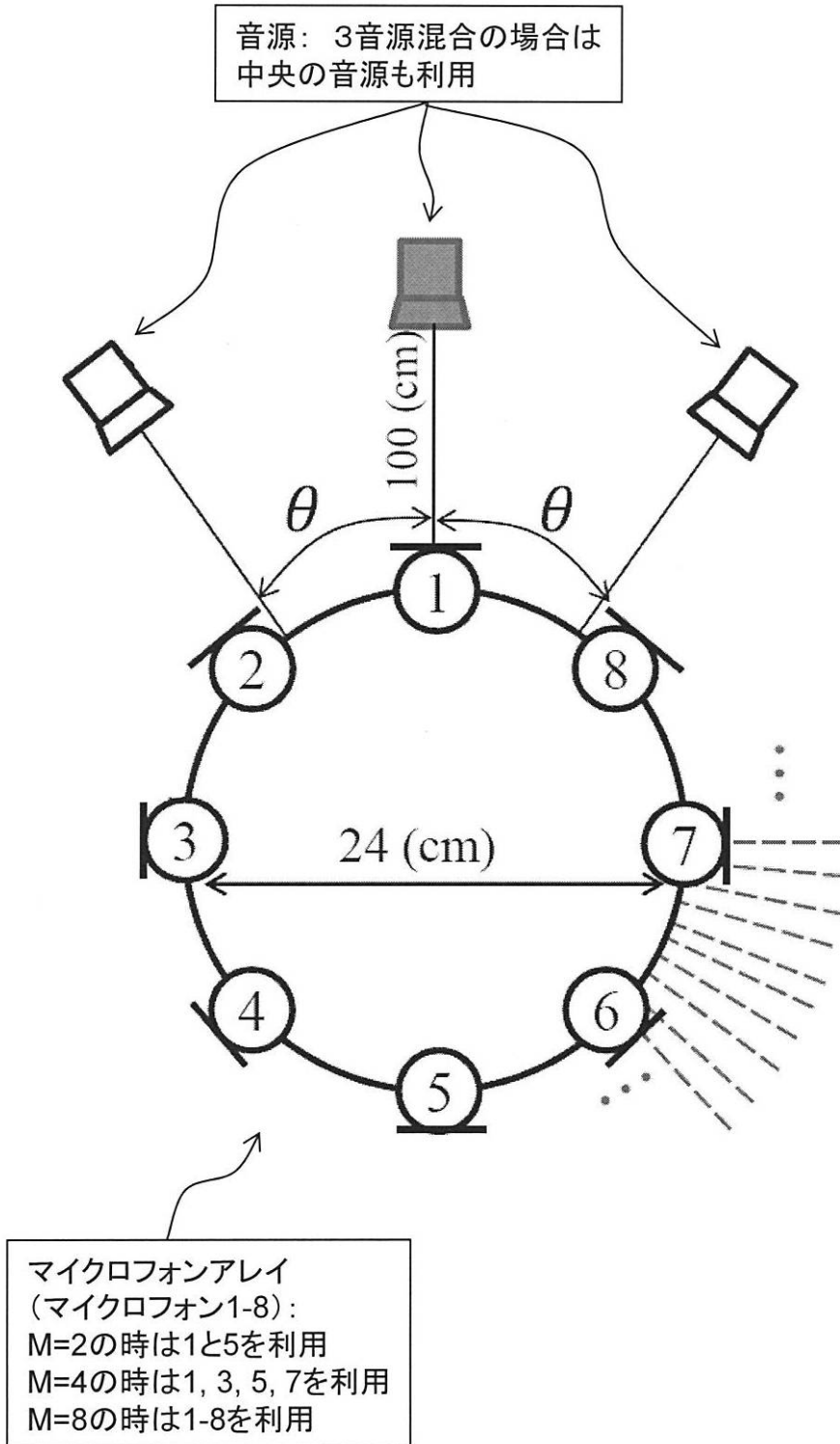
【図2】



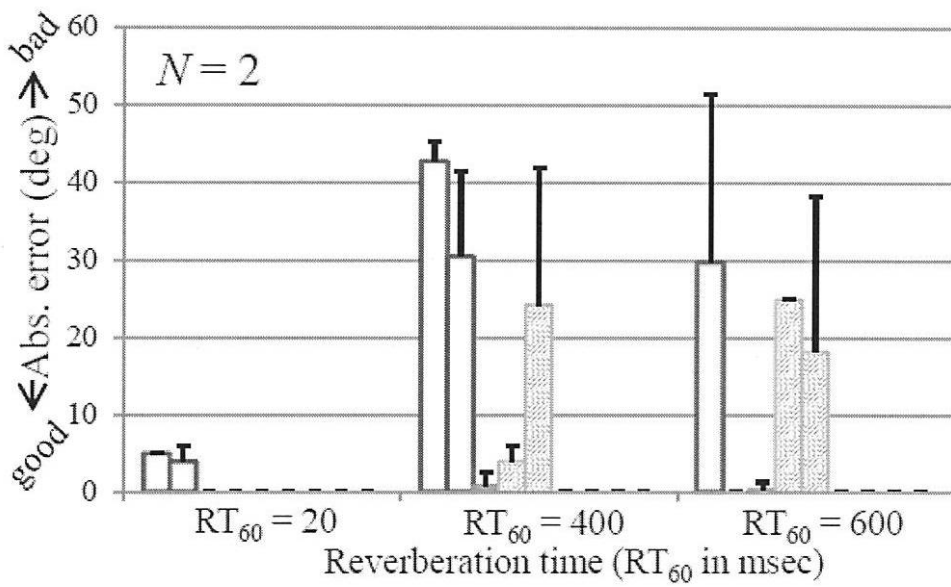
【図7】



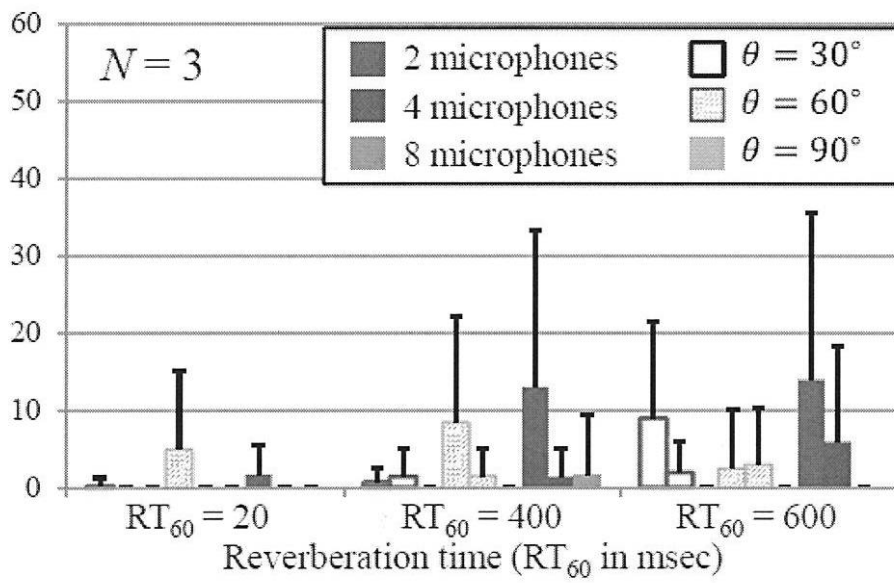
【図10】



【 図 1 1 】

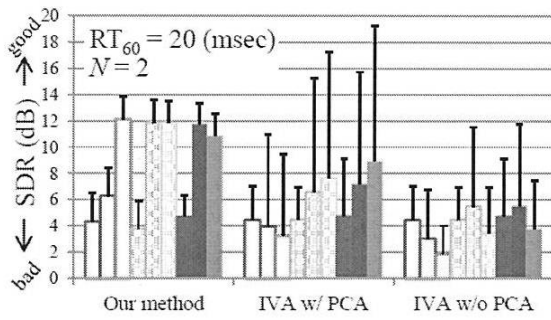


(a)

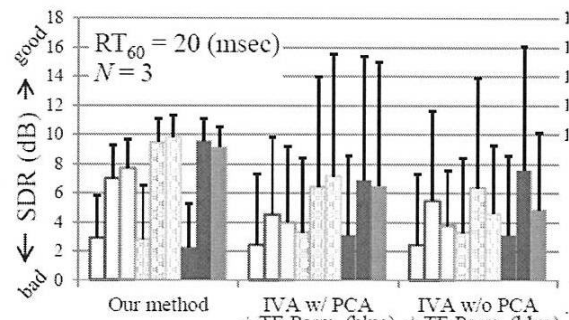


(b)

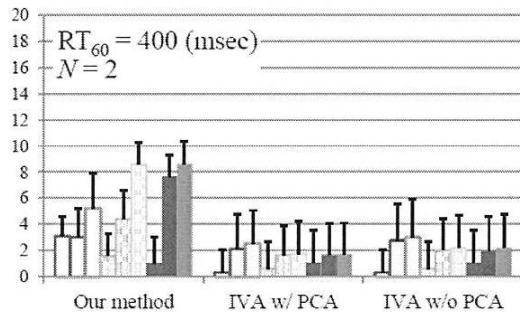
【 図 1 2 】



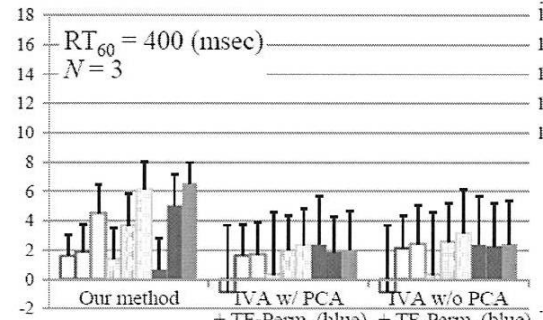
(a)



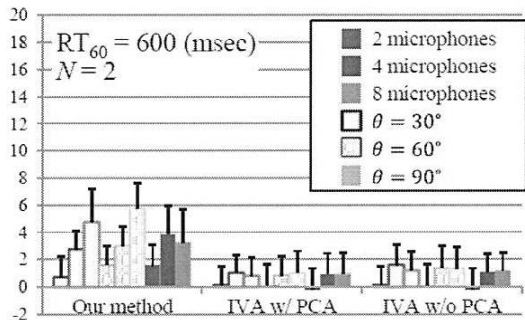
(b)



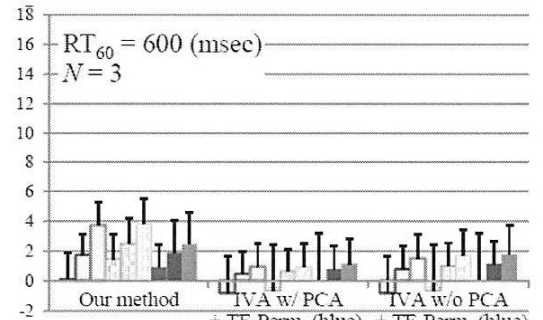
(c)



(d)



(e)



(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

審査官 山下 剛史

- (56)参考文献 特開2008-145610(JP,A)
特開2011-107603(JP,A)
中谷智広他, "音源スペクトルHMMと音源方向モデルの教師無し同時学習に基づく多チャンネル音源分離", 日本音響学会2011年春季研究発表会講演論文集[CD-ROM], 2011年 3月, pp.805-808
高橋徹他, "実環境音声認識のためのロボット聴覚システム開発とパラメータチューニング", 情報処理学会創立50周年記念(第72回)全国大会講演論文集, 2010年 3月, pp.5-29
~5-30

- (58)調査した分野(Int.Cl., DB名)
G10L 21/0272 - 21/0308