

(19) 日本国特許庁(JP)

## (12) 特許公報(B2)

(11) 特許番号

特許第6448567号  
(P6448567)

(45) 発行日 平成31年1月9日(2019.1.9)

(24) 登録日 平成30年12月14日(2018.12.14)

(51) Int.Cl.

F 1

|              |           |              |      |
|--------------|-----------|--------------|------|
| HO4R 3/00    | (2006.01) | HO4R 3/00    | 320  |
| G10L 21/028  | (2013.01) | G10L 21/028  | B    |
| G10L 21/0232 | (2013.01) | G10L 21/0232 |      |
| G10L 21/0272 | (2013.01) | G10L 21/0272 | 100A |
| G10L 21/0308 | (2013.01) | G10L 21/0308 | Z    |

請求項の数 5 (全 16 頁) 最終頁に続く

(21) 出願番号

特願2016-31801 (P2016-31801)

(22) 出願日

平成28年2月23日(2016.2.23)

(65) 公開番号

特開2017-152825 (P2017-152825A)

(43) 公開日

平成29年8月31日(2017.8.31)

審査請求日

平成29年11月13日(2017.11.13)

(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

(73) 特許権者 504137912

国立大学法人 東京大学

東京都文京区本郷七丁目3番1号

(74) 代理人 110001519

特許業務法人太陽国際特許事務所

(72) 発明者 龟岡 弘和

東京都千代田区大手町一丁目5番1号 日本電信電話株式会社内

(72) 発明者 村田 直毅

東京都文京区本郷七丁目3番1号 国立大学法人東京大学内

最終頁に続く

(54) 【発明の名称】音響信号解析装置、音響信号解析方法、及びプログラム

## (57) 【特許請求の範囲】

## 【請求項 1】

J 本の各マイクロホン  $j$  で集音した音響信号の時系列データを入力として、各時刻  $l$  における各周波数  $k$  の観測信号時間周波数成分  $y_{j,k,l}$  を出力する時間周波数解析部と、

音源  $i$  から前記マイクロホン  $j$  までの時刻  $n$  だけ遅延して集音される伝達特性を表す時変ステアリングベクトルの振幅成分  $A_{j,i,k,n}$ 、複数フレームのスペクトルを連結したスペクトル素片に対応する各時刻  $l$  における、各基底  $m$  及び各周波数  $k$  の非負値の基底スペクトル  $W_{i,k,m}$ 、前記スペクトル素片に対応する各時刻  $l$  における、各基底  $m$  及び各周波数  $k$  の非負値の基底オンセット  $H_{i,m}$  の各々に初期値を設定する初期値設定部と、

( $j, k, l$ ) の全ての組み合わせにおける、前記観測信号時間周波数成分  $y_{j,k,l}$  と、前記スペクトル素片に対応する各時刻  $l$  における、各基底  $m$  及び各周波数  $k$  の非負値の基底スペクトル  $W_{i,k,m}$ 、前記スペクトル素片に対応する各時刻  $l$  における、各基底  $m$  及び各周波数  $k$  の非負値の基底オンセット  $H_{i,m}$ 、各音源  $i$  及び各時刻  $n$  の前記振幅成分  $A_{j,i,k,n}$  に基づいて算出される前記マイクロホン  $j$  のパワースペクトログラムモデル  $X_{j,k,l}$  との距離が小さくなるように、前記基底スペクトル  $W_{i,k,m}$  と、前記基底オンセット  $H_{i,m}$  と、前記振幅成分  $A_{j,i,k,n}$  とを更新するパラメータ更新部と、

予め定められた終了条件を満たすまで、前記パラメータ更新部による更新を繰り返し行う終了判定部と、

を含む音響信号解析装置。

## 【請求項 2】

10

前記観測信号時間周波数成分  $y_{j,k,l}$  と、前記マイクロホン  $j$  のパワースペクトログラムモデル  $X_{j,k,l}$  との距離の尺度として板倉斎藤距離を用いた

請求項 1 記載の音響信号解析装置。

【請求項 3】

時間周波数解析部と、初期値設定部と、パラメータ更新部と、終了判定部とを含む音響信号解析装置における音響信号解析方法であって、

前記時間周波数解析部が、 $J$  本の各マイクロホン  $j$  で集音した音響信号の時系列データを入力として、各時刻  $l$  における各周波数  $k$  の観測信号時間周波数成分  $y_{j,k,l}$  を出力し、

前記初期値設定部が、音源  $i$  から前記マイクロホン  $j$  までの時刻  $n$  だけ遅延して集音される伝達特性を表す時変ステアリングベクトルの振幅成分  $A_{j,i,k,n}$ 、複数フレームのスペクトルを連結したスペクトル素片に対応する各時刻  $l$  における、各基底  $m$  及び各周波数  $k$  の非負値の基底スペクトル  $W_{i,k,m}$ 、前記スペクトル素片に対応する各時刻  $l$  における、各基底  $m$  及び各周波数  $k$  の非負値の基底オンセット  $H_{i,m}$  の各々に初期値を設定し、

前記パラメータ更新部が、 $(j,k,l)$  の全ての組み合わせにおける、前記観測信号時間周波数成分  $y_{j,k,l}$  と、前記スペクトル素片に対応する各時刻  $l$  における、各基底  $m$  及び各周波数  $k$  の非負値の基底スペクトル  $W_{i,k,m}$ 、前記スペクトル素片に対応する各時刻  $l$  における、各基底  $m$  及び各周波数  $k$  の非負値の基底オンセット  $H_{i,m}$ 、各音源  $i$  及び各時刻  $n$  の前記振幅成分  $A_{j,i,k,n}$  に基づいて算出される前記マイクロホン  $j$  のパワースペクトログラムモデル  $X_{j,k,l}$  との距離が小さくなるように、前記基底スペクトル  $W_{i,k,m}$  と、前記基底オンセット  $H_{i,m}$  と、前記振幅成分  $A_{j,i,k,n}$  とを更新し、

前記終了判定部が、予め定められた終了条件を満たすまで、前記パラメータ更新部による更新を繰り返し行う

音響信号解析方法。

【請求項 4】

前記観測信号時間周波数成分  $y_{j,k,l}$  と、前記マイクロホン  $j$  のパワースペクトログラムモデル  $X_{j,k,l}$  との距離の尺度として板倉斎藤距離を用いた

請求項 3 記載の音響信号解析方法。

【請求項 5】

コンピュータを、請求項 1 又は請求項 2 記載の音響信号解析装置の各部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、音響信号解析装置、音響信号解析方法、及びプログラムに係り、特に、複数のマイクロホンで取得した音響信号を用いて、残響除去と音源分離を行うための音響信号解析装置、方法、及びプログラムに関する。

【背景技術】

【0002】

複数のマイクロホンで取得した多チャネル信号を処理し、音源の空間情報を手がかりにして音源分離などを行う枠組をマイクロホンアレー信号処理という。

【0003】

近年、マイクロホンアレー信号処理に関して、ボイスレコーダ、ノートパソコン、スマートフォン、及びビデオカメラ等の身の回りにある様々な録音機器による多チャンネル録音を用いた、アドホックマイクロホンアレーの研究が盛んに行われている。アドホックマイクロホンアレーを用いた場合、特殊な装置や配線を要する従来のマイクロホンアレーを用いた場合に比べて手軽かつ安価にマイクロホンアレーシステムを構築できるため、注目を集めている。

【0004】

10

20

30

40

50

従来のマイクロホンアレーで商用化されているものの多くは、各マイクロホンが小規模に集中配置されているため、録音チャンネル間の音声信号の時間差が音源分離のための手がかりとなるのに対し、アドホックマイクロホンアレーを用いた場合、従来のマイクロホンアレーに比べてマイクロホンを広範囲に分散して配置することが容易となるため、録音チャンネル間の時間差に加えて、更に音声信号の強度比も音源分離のための手がかりとなる。

#### 【0005】

一般に、音声信号に残響及び雑音が重畠され、観測信号が得られるプロセスを順問題と捉えると、アドホックマイクロホンアレーで集音した観測信号から目的音声のみを分離抽出する問題は逆問題とみなすことができる。雑音、或いは室内伝達系の情報が未知の場合で、且つ、マイクロホン数より音源数が多いという劣決定条件の場合、この逆問題には解が無数に存在しうるため、解を絞り込むための何らかの仮定が必要となる。

10

#### 【0006】

一方、近年、劣決定条件における音源分離手法の一例として、非負値行列因子分解(Non-negative Matrix Factorization: NMF)の多チャンネル拡張を用いたアプローチが注目されている(非特許文献1、2)。

#### 【0007】

NMFとは、非負値行列を二つの非負値行列(基底行列と係数行列)の積に分解することをいい、スペクトログラムを非負値行列と見なしてNMFを適用することはスペクトログラムを低ランクの非負値行列で近似することに相当し、各時刻のスペクトルを基底行列の列数分のスペクトルテンプレートの非負結合で説明しようとしていることを意味する。そして、NMFの多チャンネル拡張は、各音源のパワースペクトログラムにこの構造を仮定した多チャンネル音源分離手法の一例である。なお、音源数よりマイクロホン数が多いという過剰決定条件でのNMFの多チャンネル拡張の例も提案されている(非特許文献3)。

20

#### 【先行技術文献】

#### 【非特許文献】

#### 【0008】

【非特許文献1】A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," IEEE Transactions on Audio, Speech and Language Processing, vol. 18, no. 3, pp. 550-563, 2010.

30

【非特許文献2】A. Ozerov, C. Févotte, R. Blouet and J.L. Durrieu, "Multichannel nonnegative tensor factorization with structured constraints for user-guided audio source separation," in Proc. 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2011), pp. 257-260, 2011.

【非特許文献3】Hiroyuki Kameoka, Takuya Yoshioka, Mariko Hamamura, Jonathan Le Roux, Kunio Kashino, "Statistical model of speech signals based on composite autoregressive system with application to blind source separation," in Proc. 9th International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA 2010), LNCS 6365, pp. 245-253, Sep. 2010.

【非特許文献4】T. Yoshioka, Tomohiro Nakatani, Masato Miyoshi, and Hiroshi G. Okuno, "Blind separation and dereverberation of speech mixtures by joint optimization," IEEE Transactions on Audio, Speech and Language Processing, vol. 19, no. 1, pp. 69-84, 2011.

40

#### 【発明の概要】

#### 【発明が解決しようとする課題】

#### 【0009】

上述した従来のアプローチでは、室内伝達系に時不变性などの制約が置かれ、その条件の下で逆問題が定式化されるが、アドホックマイクロホンアレーの構組では手軽にアレーシステムが構築できる利点がある一方で、各マイクロホンの位置は固定されないため音源及びマイクロホンの相対位置関係が録音中に変化しやすいという脆弱性を有している。こ

50

のように録音中に音源及びマイクロホンの相対位置関係が変化する場合、室内伝達系に対する上述の仮定が成立しなくなり、当該仮定の下で設計されたアルゴリズムは、音源分離に関して高い性能を発揮できなくなるという問題があった。

【0010】

本発明は、上記の事情を鑑みてなされたものであり、音源及びマイクロホンの相対位置関係が変化する時変残響環境下であっても、複数の音源の音声が重畠した観測信号から、各音源信号を精度よく分離することができる音響信号解析装置、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】

【0011】

上記の目的を達成するために本発明に係る音響信号解析装置は、J本の各マイクロホン $j$ で集音した音響信号の時系列データを入力として、各時刻 $l$ における各周波数 $k$ の観測信号時間周波数成分 $y_{j,k,l}$ を出力する時間周波数解析部と、音源 $i$ から前記マイクロホン $j$ までの時刻 $n$ だけ遅延して集音される伝達特性を表す時変ステアリングベクトルの振幅成分 $A_{j,i,k,n}$ 、複数フレームのスペクトルを連結したスペクトル素片に対応する各時刻 $l$ における、各基底 $m$ 及び各周波数 $k$ の非負値の基底スペクトル $W_{i,k,m}$ 、前記スペクトル素片に対応する各時刻 $l$ における、各基底 $m$ 及び各周波数 $k$ の非負値の基底オンセット $H_{i,m}$ 、の各々に初期値を設定する初期値設定部と、 $(j,k,l)$ の全ての組み合わせにおける、前記観測信号時間周波数成分 $y_{j,k,l}$ と、前記スペクトル素片に対応する各時刻 $l$ における、各基底 $m$ 及び各周波数 $k$ の非負値の基底スペクトル $W_{i,k,m}$ 、前記スペクトル素片に対応する各時刻 $l$ における、各基底 $m$ 及び各周波数 $k$ の非負値の基底オンセット $H_{i,m}$ 、各音源 $i$ 及び各時刻 $n$ の前記振幅成分 $A_{j,i,k,n}$ に基づいて算出される前記マイクロホン $j$ のパワースペクトログラムモデル $X_{j,k,l}$ との距離が小さくなるように、前記基底スペクトル $W_{i,k,m}$ と、前記基底オンセット $H_{i,m}$ と、前記振幅成分 $A_{j,i,k,n}$ とを更新するパラメータ更新部と、予め定められた終了条件を満たすまで、前記パラメータ更新部による更新を繰り返し行う終了判定部と、を含んで構成されている。

【0012】

本発明に係る音響信号解析方法は、時間周波数解析部と、初期値設定部と、パラメータ更新部と、終了判定部とを含む音響信号解析装置における音響信号解析方法であって、前記時間周波数解析部が、J本の各マイクロホン $j$ で集音した音響信号の時系列データを入力として、各時刻 $l$ における各周波数 $k$ の観測信号時間周波数成分 $y_{j,k,l}$ を出力し、前記初期値設定部が、音源 $i$ から前記マイクロホン $j$ までの時刻 $n$ だけ遅延して集音される伝達特性を表す時変ステアリングベクトルの振幅成分 $A_{j,i,k,n}$ 、複数フレームのスペクトルを連結したスペクトル素片に対応する各時刻 $l$ における、各基底 $m$ 及び各周波数 $k$ の非負値の基底スペクトル $W_{i,k,m}$ 、前記スペクトル素片に対応する各時刻 $l$ における、各基底 $m$ 及び各周波数 $k$ の非負値の基底オンセット $H_{i,m}$ 、の各々に初期値を設定し、前記パラメータ更新部が、 $(j,k,l)$ の全ての組み合わせにおける、前記観測信号時間周波数成分 $y_{j,k,l}$ と、前記スペクトル素片に対応する各時刻 $l$ における、各基底 $m$ 及び各周波数 $k$ の非負値の基底スペクトル $W_{i,k,m}$ 、前記スペクトル素片に対応する各時刻 $l$ における、各基底 $m$ 及び各周波数 $k$ の非負値の基底オンセット $H_{i,m}$ 、各音源 $i$ 及び各時刻 $n$ の前記振幅成分 $A_{j,i,k,n}$ に基づいて算出される前記マイクロホン $j$ のパワースペクトログラムモデル $X_{j,k,l}$ との距離が小さくなるように、前記基底スペクトル $W_{i,k,m}$ と、前記基底オンセット $H_{i,m}$ と、前記振幅成分 $A_{j,i,k,n}$ とを更新し、前記終了判定部が、予め定められた終了条件を満たすまで、前記パラメータ更新部による更新を繰り返し行う。

【0013】

本発明に係るプログラムは、上記の音響信号解析装置の各部としてコンピュータを機能させるためのプログラムである。

【発明の効果】

【0014】

以上説明したように、本発明の音響信号解析装置、方法、及びプログラムによれば、観測信号時間周波数成分  $y_{j,k,l}$  と、スペクトル素片に対応する各時刻 における、各基底  $m$  及び各周波数  $k$  の非負値の基底スペクトル  $W_{i,k,m}$  、スペクトル素片に対応する各時刻 における、各基底  $m$  及び各周波数  $k$  の非負値の基底オンセット  $H_{i,m}$  、各音源  $i$  及び各時刻  $n$  における時変ステアリングベクトルの振幅成分  $A_{j,i,k,n}$  に基づいて算出されるマイクロホン  $j$  のパワースペクトログラムモデル  $X_{j,k,l}$  との距離が小さくなるように、基底スペクトル  $W_{i,k,m}$  と、基底オンセット  $H_{i,m}$  と、振幅成分  $A_{j,i,k,n}$  とを更新することを繰り返すことにより、複数の音源の音声が重畠した観測信号から、各音源信号を精度よく分離することができる、という効果が得られる。

【図面の簡単な説明】

10

【0015】

【図1】本発明の実施の形態に係る音響信号解析装置の構成を示す概略図である。

【図2】本発明の実施の形態に係る音響信号解析装置における音響信号解析処理ルーチンの内容を示すフローチャートである。

【図3】音響信号解析装置の評価実験を実施する部屋の環境を示す図である。

【図4】音響信号解析装置の評価実験における、壁の反射係数に対するSource to Distortion Ratio(SDR)の変化の一例を示すグラフである。

【図5】音響信号解析装置の評価実験における、伝達系への外乱に対するSDRの変化の一例を示すグラフである。

【発明を実施するための形態】

20

【0016】

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。

【0017】

(発明の原理)

【0018】

まず、本発明の提案モデルについて説明する。

【0019】

(時間周波数領域における畳み込み混合モデル)

【0020】

音源からマイクロホンアレーへの伝達系が線形時不变であり、また、残響成分が時間周波数解析の窓長内に収まっていると仮定できる場合、マイクロホンアレーで得られる信号は音源の瞬時混合で記述することができる。

30

【0021】

一方、窓長を超える残響成分が無視できない状況下では、マイクロホンで観測される観測信号は、例えば、非特許文献4に示されているように、時間周波数領域の畳み込み混合モデルを用いて(1)式のように表される。

【0022】

【数1】

$$\hat{y}_{k,l} = \sum_i \sum_n \hat{a}_{i,k,n} \hat{s}_{i,k,l-n} \quad \dots \quad (1)$$

40

【0023】

ここで、変数  $i$ 、 $k$  は、それぞれ音源と周波数のインデックスを表し、 $l$ 、 $n$  は時間フレームのインデックスを表す。 $\hat{y}_{k,l}$ 、 $\hat{s}_{i,k,l}$  は、マイクロホンアレーで観測される観測信号ベクトルであり、 $j$  はマイクロホンのインデックスを表し、 $J$  はマイクロホン数を表す。 $\hat{a}_{i,k,n}$  は、各音源からマイクロホンへのステアリングベクトルであり、時間が  $n$  フレーム遅れて到来する反響成分に対応するものとする。また、 $\hat{s}_{i,k,l}$  は、各音源の時間周波数領域における複素スペクトログラムである。以降、ベクトル、行列又は確率変数を表す変数には “ $\hat{ }$ ” を付すものとする。

【0024】

今、音源或いはマイクロホンの位置が時間変化すると仮定した場合、ステアリングベク

50

トル  $a_{i,k,n}^h$  は時刻 1 に依存し、(1)式の混合過程は、(2)式のような時変ステアリングベクトル  $a_{i,k,n,1}^h$  を用いて表される。

【0025】

【数2】

$$\hat{y}_{k,1} = \sum_i \sum_n \hat{a}_{i,k,n,1}^h s_{i,k,n,1} \quad \dots \quad (2)$$

【0026】

ここで、各音源  $i$  の複素スペクトログラムが複素ガウス分布に従う、すなわち、 $s_{i,k,n,1} \sim N_c(0, P_{i,k,n,1})$  と仮定すれば、マイクロホンでの観測信号ベクトル  $\hat{y}_{k,1}$  は、(3)式に示す分布に従う。なお、 $P_{i,k,n,1}$  は、音源  $i$  の時刻 1 におけるパワースペクトログラムを表し、 $a_{i,k,n,1}^H$  は、時変ステアリングベクトル  $a_{i,k,n,1}^h$  のエルミート転置を表す。

【0027】

【数3】

$$\hat{y}_{k,1} \sim N_c \left( 0, \sum_i \sum_n P_{i,k,n,1} \hat{a}_{i,k,n,1}^h \hat{a}_{i,k,n,1}^{H*} \right) \quad \dots \quad (3)$$

【0028】

ここで、時変ステアリングベクトル  $a_{i,k,n,1}^h$  を絶対値と偏角の要素に分解して表した式を(4)式に示す。

【0029】

【数4】

$$\hat{a}_{i,k,n,1}^h = \begin{bmatrix} |\hat{a}_{1,i,k,n,1}^h| & 0 & \begin{bmatrix} e^{j\phi_{1,i,k,n,1}} \\ \vdots \\ e^{j\phi_{J,i,k,n,1}} \end{bmatrix} \\ \ddots & & \\ 0 & |\hat{a}_{J,i,k,n,1}^h| & \end{bmatrix} \quad \dots \quad (4)$$

【0030】

今、音源或いはマイクロホンの微小移動等、音響信号解析環境の軽微な変化については、時変ステアリングベクトル  $a_{i,k,n,1}^h$  の振幅成分を時不变、位相成分を時変と仮定する特殊な時変系を設定し、このような混合過程を「半時変形」と呼ぶことにする。すなわち、 $|\hat{a}_{j,i,k,n,1}^h|$  は時刻 1 に依存しないことになる。したがって、 $|\hat{a}_{j,i,k,n,1}^h| = A_{j,i,k,n}$  とすれば、(4)式は(5)式のように表すことができる。

【0031】

【数5】

$$\hat{a}_{i,k,n,1}^h = A_{i,k,n}^h \hat{\psi}_{i,k,n,1}^h \quad \dots \quad (5)$$

ただし、

$$A_{i,k,n}^h = \begin{bmatrix} A_{1,i,k,n} & 0 \\ \ddots & \\ 0 & A_{J,i,k,n} \end{bmatrix}$$

$$\hat{\psi}_{i,k,n,1}^h = \begin{bmatrix} e^{j\phi_{1,i,k,n,1}} \\ \vdots \\ e^{j\phi_{J,i,k,n,1}} \end{bmatrix}$$

【0032】

(5)式を(3)式に代入すれば、(6)式が得られる。

【0033】

10

20

30

40

## 【数6】

$$\hat{y}_{k,l} \sim N_c \left( 0, \sum_i \sum_n P_{i,k,l-n} A_{i,k,n} \hat{\psi}_{i,k,n,l} \hat{\psi}_{i,k,n,l}^H A_{i,k,n}^H \right) \quad \dots \quad (6)$$

## 【0034】

(非負値テンソル二重畳み込みモデル)

## 【0035】

アドホックマイクロホンアレーでは、アレー素子が非同期であることによる音声信号のサンプリング周波数の僅かなずれ、並びに音源或いはマイクロホンの僅かな位置の変化が、通常のマイクロホンアレーに比べて起こりやすいため、半時変系の混合過程として取り扱う必要がある。 10

## 【0036】

半時変系の混合過程を取り扱う際の一つの解決策は、時変ステアリングベクトル  $a_{i,k,n,l}^i$  の時間変化量をオンラインで推定して補償した後に、公知のアレー信号処理を適用することである。

## 【0037】

一方、時変ステアリングベクトル  $a_{i,k,n,l}^i$  の位相成分を確率的に変動する確率変数として扱う方法も考えられるが、ここでは、後者の考え方について、時変ステアリングベクトル  $a_{i,k,n,l}^i$  の位相成分に、次の2つの条件を設定する。

## 【0038】

(条件1)  $a_{j,i,k,n,l}^i$  と  $a_{j',i,k,n,l}^i$  ( $j = j'$  または  $l = l'$ ) は互いに独立である 20

## 【0039】

(条件2)  $a_{j,i,k,n,l}^i$  は区間  $[0, 1]$  で一様分布に従う。

## 【0040】

(条件1) 及び (条件2) の下で位相成分  $a_{j,i,k,n,l}^i$  について周辺化を行うと、 $E[a_{j,i,k,n,l}^i a_{j',i,k,n,l}^i]$  は単位行列となるため、(6)式は(7)式のように表される。

## 【0041】

## 【数7】

$$\hat{y}_{j,k,l} \sim N_c \left( 0, \sum_i \sum_n P_{i,k,l-n} A_{j,i,k,n}^2 \right) \quad \dots \quad (7)$$

## 【0042】

次に、音源  $i$  のパワースペクトログラム  $P_{i,k,l}$  について検討する。

## 【0043】

周波数のインデックス  $k$  の最大値を  $K$ 、及び時間フレームのインデックス  $l$  の最大値を  $L$  とすれば、NMFの多チャンネル拡張では、 $K$  行  $L$  列で表されるパワースペクトログラムベクトル  $P_i^i = (P_{i,k,l})_{K \times L}$  を2つの非負値行列の積で表される。これは、音源  $i$  の時刻  $l$  におけるパワースペクトログラム  $P_{i,k,l}$  が、限られた数のスペクトルテンプレートの非負結合で表されるという仮定に基づくものである。

## 【0044】

しかし、音声には単一の時間フレーム  $l$  におけるスペクトルのみならず、そのダイナミクス、すなわち、局所的な時間変化パターンに大きな特徴があるものと考えられる。したがって、各時間フレーム  $l$  におけるスペクトルを音声を構成する要素と考えるよりも、複数の時間フレーム  $l$  に亘るスペクトルを連結したものを、音声を構成する要素単位とみなした方が、音声をより特徴付けた好みの表現を考えることができる。 40

## 【0045】

そこで、本実施の形態では、複数の時間フレーム  $l$  に亘るスペクトルを連結したスペクトログラム素片のテンプレートとアクティベーション系列とを畳み込む混合モデルによって、パワースペクトログラム  $P_{i,k,l}$  のモデル化を行う。具体的には、公知の非負値行列因子逆畳み込み(Nonnegative Matrix Factor Deconvolution:NMFD)の考え方を、パワース

ペクトログラム  $P_{i,k,l}$  のモデル化に適用する。この場合、音源  $i$  の時刻  $l$  におけるパワースペクトログラム  $P_{i,k,l}$  は (8) 式で表される。

【0046】

【数8】

$$P_{i,k,l} = \sum_m \sum_r W_{i,k,m,r} H_{i,m,l-r} \quad \dots \quad (8)$$

【0047】

ここで、 $W_{i,k,m,r}$  は音源  $i$  のパワースペクトログラム  $P_{i,k,l}$  の基底スペクトルを表し、 $H_{i,m,l-r}$  は基底オンセットを表す。また、変数  $m$  は基底のインデックスを表し、変数  $r$  は時間フレームのインデックスを表す。なお、 $= \{0\}$  ( $\{x\}$  は、“ $x$ ”が集合の要素に含まれていることを示す) の場合は、NMFを用いた混合モデルに一致する。 10

【0048】

(パラメータ推定)

【0049】

次に、最尤推定を用いたパラメータ推定について説明する。まず、(7)式の対数尤度関数の負値を  $C_{ML}$  として (9) 式のように定義する。なお、 $Y^{\wedge}$  は、観測信号  $y_{j,k,l}$  に対応する確率変数、 $P^{\wedge}$  は、パワースペクトログラム  $P_{i,k,l}$  に対応する確率変数、及び  $A^{\wedge}$  は、時変ステアリングベクトル  $a_{i,k,n,l}^{\wedge}$  の絶対値に対応する確率変数である。また、観測信号  $y_{j,k,l}$  は、観測信号時間周波数成分の一例である。 20

【0050】

【数9】

$$C_{ML} \stackrel{\text{def}}{=} -\log p(Y^{\wedge} | P^{\wedge}, A^{\wedge}) \quad \dots \quad (9)$$

【0051】

(9) 式を用いて  $C_{ML}$  を具体的に計算すれば (10) 式が得られる。

【0052】

【数10】

$$\begin{aligned} C_{ML} &= \sum_{j,k,l} \log N_c(y_{j,k,l} | 0, \sum_{i,n} P_{i,k,l-n} A_{i,j,k,n}^2) \\ &= JKL \log \pi + \sum_{j,k,l} \log \left( \sum_{i,n} P_{i,k,l-n} A_{i,j,k,n}^2 \right) + \frac{|y_{j,k,l}|^2}{\sum_{i,n} P_{i,k,l-n} A_{i,j,k,n}^2} \\ &\stackrel{c}{=} \sum_{j,k,l} d_{IS} \left( |y_{j,k,l}|^2 \left| \sum_{i,n} P_{i,k,l-n} A_{i,j,k,n}^2 \right. \right) \quad \dots \quad (10) \end{aligned} \quad 30$$

【0053】

ここで、記号  $\stackrel{c}{=}$  は、定数項以外の他の項は等しいことを表す。

【0054】

すなわち、最尤推定を用いたパラメータ推定は、マイクロホン  $j$  の観測パワースペクトログラム  $Y_{j,k,l}$  と、音源  $i$  のパワースペクトログラム  $P_{i,k,l}$  として (8) 式を適用した場合における、マイクロホン  $j$  のパワースペクトログラムモデル  $X_{j,k,l}$  との板倉斎藤距離の最小化問題に帰着し、目的関数である (11) 式を  $A_{i,j,k,n} = 0$ 、 $W_{i,k,m,r} = 0$ 、及び  $H_{i,m,l-r} = 0$  の制約の下で最適化する問題になる。 40

【0055】

【数11】

$$\min \sum_{j,k,l} d_{IS} (Y_{j,k,l} | X_{j,k,l}) \quad \dots \quad (11)$$

$$\text{ただし、} X_{j,k,l} = \sum_{i,n} P_{i,k,l-n} A_{i,j,k,n}^2 \quad \dots \quad (12)$$

【0056】

そこで、(11)式を公知の補助関数法の原理に基づいて最適化する。なお、ここでは、板倉斎藤距離を内包した、より一般的な乖離度規準である ダイバージェンスを規準として最適化アルゴリズムを導出する。

【0057】

最適化アルゴリズムの詳細な導出過程の説明は省略するが、(11)式で表される目的関数に対して、Jensenの不等式と接線不等式とを用いて補助関数を設計することで、(13)～(15)式に示す乗法更新式を得ることができる。

【0058】

【数12】

$$A_{j,i,k,n} \leftarrow A_{j,i,k,n} \left( \frac{\sum_{l} Y_{j,k,l} X_{j,k,l}^{\beta-2} P_{i,k,l-n}}{\sum_{l} X_{j,k,l}^{\beta-1} P_{i,k,l-n}} \right)^{\rho(\beta)} \quad \dots \quad (13)$$

$$W_{i,k,m,\tau} \leftarrow W_{i,k,m,\tau} \left( \frac{\sum_{j,l,n} Y_{j,k,l} X_{j,k,l}^{\beta-2} A_{j,i,k,l-n} H_{i,m,n-\tau}}{\sum_{j,l,n} X_{j,k,l}^{\beta-1} A_{j,i,k,l-n} H_{i,m,n-\tau}} \right)^{\rho(\beta)} \quad \dots \quad (14)$$

$$H_{i,m,\tau} \leftarrow H_{i,m,\tau} \left( \frac{\sum_{j,k,l,n} Y_{j,k,l} X_{j,k,l}^{\beta-2} A_{j,i,k,l-n} W_{i,k,m,n-\tau}}{\sum_{j,k,l,n} X_{j,k,l}^{\beta-1} A_{j,i,k,l-n} W_{i,k,m,n-\tau}} \right)^{\rho(\beta)} \quad \dots \quad (15)$$

10

20

【0059】

ただし、( )は、の値に応じて設定される値であり、(16)式で定義される。

【0060】

【数13】

$$\rho(\beta) = \begin{cases} \frac{1}{2-\beta} & (\beta < 1) \\ 1 & (1 \leq \beta \leq 2) \\ \frac{1}{\beta-1} & (2 < \beta) \end{cases} \quad \dots \quad (16)$$

30

【0061】

(システム構成)

【0062】

次に、アドホックマイクロホンアレーで取得した、伝達系への外乱及び残響成分が重畠された複数の音源の観測信号を解析して、複数の音源の各々の音源信号に分離する音響信号解析装置に本発明を適用した場合を例にして、本発明の実施の形態を説明する。

【0063】

図1に示すように、本発明の実施の形態に係る音響信号解析装置は、CPUと、RAMと、後述する音響信号解析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とを備えたコンピュータで構成され、機能的には次に示すように構成されている。

40

【0064】

音響信号解析装置100は、入力部10と、演算部20と、記憶部30と、出力部40とを備えている。

【0065】

入力部10により、伝達系への外乱及び残響成分が重畠された複数の音源iを含む観測信号 $y_j[1]$ の時系列データが入力される。記憶部30は、入力部10により入力された観測信号 $y_j[1]$ の時系列データを記憶する。また、記憶部30は、後述する各処理での結果を記憶すると共に、本処理ルーチンで用いる各パラメータの初期値を記憶している。

【0066】

50

演算部 20 は、時間周波数解析部 21 と、初期設定部 22 と、パラメータ更新部 23 と、終了判定部 24 と、信号変換部 25 とを備えている。

#### 【0067】

時間周波数解析部 21 は、例えばマイクロホン  $j$  の時系列信号としての観測された観測信号  $y_j[1]$  を入力として、マイクロホン  $j$  の観測パワースペクトログラム  $Y_{j,k,1}$  を計算する。また、計算したマイクロホン  $j$  の観測パワースペクトログラム  $Y_{j,k,1}$  を、記憶部 30 に記憶しておく。より詳細には、時間周波数解析部 21 は、例えばマイクロホン  $j$  で観測された観測信号の時系列データを入力として、短時間フーリエ変換 (Short-Time Fourier Transform:STFT) を用いて時間周波数解析を行うことにより、マイクロホン  $j$  の観測パワースペクトログラム  $Y_{j,k,1}$  を計算する。

10

#### 【0068】

初期設定部 22 は、後述する処理で用いる各パラメータ  $A_{j,i,k,n}$ 、 $W_{i,k,m}$ 、及び  $H_{i,m}$  の各初期値を設定する。なお、各パラメータの初期値は、例えば乱数を用いて適当な値に設定すればよい。この場合、 $A_{j,i,k,n}$ 、 $W_{i,k,m}$ 、及び  $H_{i,m}$  の各パラメータの初期値は非負値となるように設定する。

#### 【0069】

パラメータ更新部 23 は、( $i$ 、 $k$ 、 $l$ ) の全ての組み合わせの各々について、記憶部 30 に記憶されている  $W_{i,k,m}$  及び  $H_{i,m}$  に基づいて、上記(8)式に従って、音源  $i$  のパワースペクトログラム  $P_{i,k,1}$  を計算し、記憶部 30 に格納する。

#### 【0070】

また、パラメータ更新部 23 は、( $j$ 、 $k$ 、 $l$ ) の全ての組み合わせの各々について、記憶部 30 に記憶されている  $P_{i,k,1}$  及び  $A_{j,i,k,n}$  に基づいて、上記(12)式に従って、マイクロホン  $j$  のパワースペクトログラムモデル  $X_{j,k,1}$  を計算し、記憶部 30 に格納する。

20

#### 【0071】

また、パラメータ更新部 23 は、( $j$ 、 $i$ 、 $k$ 、 $n$ ) の全ての組み合わせの各々について、記憶部 30 に記憶されている  $Y_{j,k,1}$ 、 $X_{j,k,1}$ 、 $A_{j,i,k,n}$  及び  $P_{i,k,1}$  に基づいて、上記(11)式の目的関数を小さくするように、(13)式に従って、時変ステアリングベクトルの振幅成分  $A_{j,i,k,n}$  を更新し、記憶部 30 に格納する。この際、パラメータ更新部 23 は、更新した  $A_{j,i,k,n}$  を用いて、上記(12)式に従って、マイクロホン  $j$  のパワースペクトログラムモデル  $X_{j,k,1}$  を更新し、記憶部 30 に格納する。

30

#### 【0072】

また、パラメータ更新部 23 は、( $i$ 、 $k$ 、 $m$ 、 $l$ ) の全ての組み合わせの各々について、記憶部 30 に記憶されている  $Y_{j,k,1}$ 、 $X_{j,k,1}$ 、 $A_{j,i,k,n}$ 、 $W_{i,k,m}$ 、及び  $H_{i,m}$  に基づいて、上記(11)の目的関数を小さくするように、(14)式に従って、音源  $i$  のパワースペクトログラム  $P_{i,k,1}$  の基底スペクトル  $W_{i,k,m}$  を更新し、記憶部 30 に格納する。この際、パラメータ更新部 23 は、更新した  $W_{i,k,m}$  を用いて、上記(8)式及び(12)式に従って、音源  $i$  のパワースペクトログラム  $P_{i,k,1}$  及びマイクロホン  $j$  のパワースペクトログラムモデル  $X_{j,k,1}$  を更新し、記憶部 30 に格納する。

#### 【0073】

更に、パラメータ更新部 23 は、( $i$ 、 $m$ 、 $l$ ) の全ての組み合わせの各々について、記憶部 30 に記憶されている  $Y_{j,k,1}$ 、 $X_{j,k,1}$ 、 $A_{j,i,k,n}$ 、 $H_{i,m}$ 、及び  $W_{i,k,m}$  に基づいて、上記(11)の目的関数を小さくするように、(15)式に従って、音源  $i$  のパワースペクトログラム  $P_{i,k,1}$  の基底オンセット  $H_{i,m}$  を更新し、記憶部 30 に格納する。この際、パラメータ更新部 23 は、更新した  $H_{i,m}$  を用いて、上記(8)式及び(12)式に従って、音源  $i$  のパワースペクトログラム  $P_{i,k,1}$  及びマイクロホン  $j$  のパワースペクトログラムモデル  $X_{j,k,1}$  を更新し、記憶部 30 に格納する。

40

#### 【0074】

終了判定部 24 は、予め定められた終了条件を満足するか否かを判定し、終了条件を満足していない場合には、パラメータ更新部 23 の各処理を繰り返す。終了判定部 24 は、

50

終了条件を満足したと判定した場合には、信号変換部 25 による処理に移行する。

【0075】

信号変換部 25 は、記憶部 30 に記憶されている音源  $i$  のパワースペクトログラム  $P_{i,k,l}$  の基底スペクトル  $W_{i,k,m}$ 、及び音源  $i$  のパワースペクトログラム  $P_{i,k,l}$  の基底オクターブセット  $H_{i,m}$  に基づいて、複数の音源  $i$  の各々について、当該音源  $i$  の音源信号を生成して出力部 40 に出力する。出力部 40 は、複数の音源  $i$  の各々の音源信号を出力する。

【0076】

なお、終了条件としては、繰り返し回数が  $L-1$  回目の目的関数 (11) 式の値と、繰り返し回数が  $L$  回目の目的関数 (11) 式の値との差が、予め定めた閾値よりも小さくなつたことを用いればよい。あるいは、終了条件として、繰り返し回数が、予め定められた上限回数に到達したことを用いてもよい。

【0077】

(音響信号解析装置の作用)

【0078】

次に、本実施の形態に係る音響信号解析装置 100 の作用について説明する。J 本のマイクロホン  $j$  からなるアドホックマイクロホンアレーで取得された、伝達系への外乱及び残響成分が重畠された複数の音源  $i$  を含む観測信号  $y_j[1]$  の時系列データが、音響信号解析装置 100 に入力され、記憶部 30 に格納される。そして、音響信号解析装置 100 において、図 2 に示す音響信号解析処理ルーチンが実行される。

【0079】

まず、ステップ S100において、記憶部 30 から、観測信号  $y_j[1]$  を読み込み、当該観測信号  $y_j[1]$  に対して、短時間フーリエ変換を用いた時間周波数分析を行い、マイクロホン  $j$  の観測パワースペクトログラム  $Y_{j,k,l}$  を算出すると共に、得られたマイクロホン  $j$  の観測パワースペクトログラム  $Y_{j,k,l}$  を記憶部 30 に記憶する。

【0080】

そして、ステップ S102において、乱数を用いて、 $A_{j,i,k,n}$ 、 $W_{i,k,m}$ 、及び  $H_{i,m}$  の各初期値を設定して、記憶部 30 に記憶する。

【0081】

次に、ステップ S104では、ステップ S102で設定された  $W_{i,k,m}$ 、及び  $H_{i,m}$  に基づいて、上記 (8) 式に従って、音源  $i$  のパワースペクトログラム  $P_{i,k,l}$  を各 ( $i$ 、 $k$ 、 $l$ ) の組み合わせについて算出して、記憶部 30 に格納する。更に、ステップ S104 では、ステップ S102 で設定された  $A_{j,i,k,n}$  と、本ステップで算出された  $P_{i,k,l}$  に基づいて、上記 (12) 式に従って、マイクロホン  $j$  のパワースペクトログラムモデル  $X_{j,k,l}$  を各 ( $j$ 、 $k$ 、 $l$ ) の組み合わせについて算出して、記憶部 30 に格納する。

【0082】

ステップ S106 では、ステップ S100 で算出された  $Y_{j,k,l}$  と、ステップ S102 で設定された  $A_{j,i,k,n}$  と、ステップ S104 で算出された  $P_{i,k,l}$  及び  $X_{j,k,l}$  に基づいて、上記 (12) 式に従って、時変ステアリングベクトルの振幅成分  $A_{j,i,k,n}$  を各 ( $j$ 、 $i$ 、 $k$ 、 $n$ ) の組み合わせについて更新して、記憶部 30 に格納する。

【0083】

ステップ S108 では、ステップ S106 で更新された  $A_{j,i,k,n}$  と、ステップ S104 で算出した  $P_{i,k,l}$  に基づいて、(12) 式に従って、マイクロホン  $j$  のパワースペクトログラムモデル  $X_{j,k,l}$  を各 ( $j$ 、 $k$ 、 $l$ ) の組み合わせについて更新して、記憶部 30 に格納する。

【0084】

ステップ S110 では、ステップ S100 で算出された  $Y_{j,k,l}$  と、ステップ S108 で更新された  $X_{j,k,l}$  と、ステップ S106 で更新された  $A_{j,i,k,n}$  と、ステップ S102 で設定された  $H_{i,m}$ 、及び  $W_{i,k,m}$  に基づいて、(14) 式に従って、音源  $i$  のパワースペクトログラム  $P_{i,k,l}$  の基底スペクトル  $W_{i,k,m}$  を各 ( $i$ 、 $k$ 、 $m$ ) の組み合

10

20

30

40

50

わせについて更新して、記憶部30に格納する。

【0085】

ステップS112では、ステップS110で更新された $W_{i,k,m}$ と、ステップS102で設定された $H_{i,m}$ に基づいて、(8)式に従って、音源iのパワースペクトログラム $P_{i,k,l}$ を各(i, k, l)の組み合わせについて更新して、記憶部30に格納する。また、本ステップで更新された $P_{i,k,l}$ と、ステップS106で更新された $A_{j,i,k,n}$ に基づいて、(12)式に従って、マイクロホンjのパワースペクトログラムモデル $X_{j,k,l}$ を各(j, k, l)の組み合わせについて更新して、記憶部30に格納する。

【0086】

ステップS114では、ステップS100で算出された $Y_{j,k,l}$ と、ステップS102で設定された $H_{i,m}$ と、ステップS112で更新された $X_{j,k,l}$ と、ステップS106で更新された $A_{j,i,k,n}$ と、ステップS110で更新された $W_{i,k,m}$ に基づいて、(15)式に従って、音源iのパワースペクトログラム $P_{i,k,l}$ の基底オンセット $H_{i,m}$ を各(i, m, l)の組み合わせについて更新して、記憶部30に格納する。

10

【0087】

ステップS116では、ステップS110で更新された $W_{i,k,m}$ と、ステップS114で更新された $H_{i,m}$ に基づいて、(8)式に従って、音源iのパワースペクトログラム $P_{i,k,l}$ を各(i, k, l)の組み合わせについて更新して、記憶部30に格納する。また、本ステップで更新された $P_{i,k,l}$ と、ステップS106で更新された $A_{j,i,k,n}$ に基づいて、(12)式に従って、マイクロホンjのパワースペクトログラムモデル $X_{j,k,l}$ を各(j, k, l)の組み合わせについて更新して、記憶部30に格納する。

20

【0088】

次のステップS118では、ステップS100で算出した $Y_{j,k,l}$ と、ステップS116で更新された $X_{j,k,l}$ に基づいて、(11)式に従って、目的関数の値を算出して、記憶部30に記憶する。そして、前回のステップS118で算出した目的関数の値を記憶部30から読み込み、今回のステップS118で算出した目的関数の値と、前回のステップS108で算出した目的関数の値との差分が、予め記憶部30に記憶されている予め定められた閾値よりも小さいか否かを判定し、差分が予め定められた閾値以上の場合には、終了条件を満足していないと判断して、上記ステップS106へ戻り、上記ステップS106～ステップS118の処理を繰り返す。

30

【0089】

一方、差分が予め定められた閾値未満の場合には、終了条件を満足したと判断して、ステップS120で、記憶部30に記憶されている音源iのパワースペクトログラム $P_{i,k,l}$ の基底スペクトル $W_{i,k,m}$ 、及び音源iのパワースペクトログラム $P_{i,k,l}$ の基底オンセット $H_{i,m}$ に基づいて、複数の音源iの各々について、当該音源iの音源信号を生成して出力部40から出力して、音響信号解析処理ルーチンを終了する。

【0090】

(実施結果)

【0091】

次に、本実施の形態に係る手法の有効性を示す目的で、2つの残響環境下での劣決定条件における音源分離実験を行った。

40

【0092】

まず、残響環境下での頑健さの度合いを確認するため、異なる残響の強さを持つ環境での音源分離実験を行った。次に、伝達系への外乱に対する頑健さの度合いを確認するため、マイクロホンの位置が観測信号取得中に変化した場合における音源分離実験を行った。

【0093】

なお、劣決定条件として、部屋の中に音源数を3つ、マイクロホン数を2つ設置し、部屋の形状と、音源の位置と、マイクロホンの位置から鏡像法を用いてインパルス応答を生成した。

【0094】

50

図3に、音源分離実験を行う部屋（以降、単に「部屋」と称す）の形状と、各音源の位置及び各マイクロホンの位置を示す。なお、図3に示すように、部屋は、 $5\text{ m} \times 10\text{ m}$ の長方形形状の大きさを有し、S1～S3で示される“ $\times$ ”は音源の位置を、M1及びM2で示される“ $\square$ ”はマイクロホンの位置をそれぞれ示している。

#### 【0095】

音源分離実験で使用する音源としては、ATR音声対話データベースの3話者15発話を用いた。3話者の音源1～3のうち、音源1、2の話者は女性であり、音源3の話者は男性である。

#### 【0096】

また、残響の強さは、部屋の壁の反射係数を調整することで変化させた。具体的には、  
部屋の壁の反射係数を0.5にした場合、残響時間は60msとなり、部屋の壁の反射係数を0.8にした場合、残響時間は210msとなる。

#### 【0097】

音源分離実験では、公知の多チャンネルNMF (Multichannel extensions of Non-negative Matrix Factorization: MNMF) を用いた手法と、本実施の形態に係る提案手法を用いた場合について比較した。MNMFの場合、瞬時混合モデルが仮定されているため、STFTのフレーム外に残響成分が存在する場合は、性能が低下することが予想される。

#### 【0098】

また、用意した各話者の15発話をうち、1つの発話を分離用の信号とし、残りの14個の発話を教師データとして事前の学習に用いた。なお、本実施の形態に係る提案手法の事前学習にはNMFD、MNMFにはNMFを用いて、基底スペクトルの学習を行い、各音源1～3に対して、それぞれ40個と20個の基底を学習した。また、距離尺度としては、一般化KLダイバージェンスを用い、評価指標としては、Source-to-distortion ratio (SDR) を用いた。この際、STFTのフレーム長を32msとし、シフト長は16msとした。

#### 【0099】

図4に、部屋の壁の反射係数の変化に対する各比較手法の性能の変化の一例を示すグラフを示す。図4に示すグラフの横軸は部屋の壁の反射係数を表すと共に、縦軸はSDRを表し、グラフ50がMNMFを用いた場合、グラフ51が本実施の形態に係る提案手法を用いた場合のグラフを示す。なお、各マイクロホンにおける各音源のSDRを平均した値を、対応する部屋の壁の反射係数に対するSDRとして表している。

#### 【0100】

図4に示すように、反射係数の低い、瞬時混合モデルが成立しているとみなせる範囲においては、MNMFは本実施の形態に係る提案手法に比べて良い性能を示している。しかし、反射係数が高くなるにしたがって、本実施の形態に係る提案手法では、スペクトログラム素片のテンプレートとアクティベーション系列とを畳み込む混合モデルを採用していることから、残響成分を精度よく推定し、MNMFより良い性能を示す傾向が見られる。

#### 【0101】

次に、伝達系への外乱に対する頑健さの度合いを確認するため、マイクロホンの位置が観測信号取得中に変化した場合における音源分離実験について説明する。

#### 【0102】

本実験では、図3に示すマイクロホンM1、M2で取得した観測信号と、マイクロホンM1、M2の位置からそれぞれ $\times\text{m}$ ずらした位置で取得した観測信号とを繋げることにより、本音源分離実験で用いる観測信号を生成した。この場合、伝達系の振幅成分に比べて、位相成分には大きな外乱が生じることになる。なお、本音源分離実験における部屋の壁の反射係数は0.8に設定した。

#### 【0103】

図5に、観測信号取得位置のずれ量 $\times$ に対する各比較手法の性能の変化の一例を示すグラフを示す。図5に示すグラフの横軸は観測信号取得位置のずれ量 $\times$ を表すと共に、縦軸はSDRを表し、グラフ52がMNMFを用いた場合、グラフ53が本実施の形態に係る提案手法を用いた場合のグラフを示す。

10

20

30

40

50

**【 0 1 0 4 】**

図5に示すように、MNMFにおける性能を示すグラフ5-2は、 $x$ が大きくなるにしたがって、すなわち、外乱が大きくなるにしたがって、SDRの値が大きく落ち込んでいることがわかる。具体的には、 $x = 0$ の場合のSDRに対して、 $x = 0.2$ の場合のSDRは約86%の値まで減少している。

**【 0 1 0 5 】**

一方、提案手法の場合、 $x = 0$ の場合のSDRに対して、 $x = 0.2$ の場合のSDRは約93%の値までしか減少していない。したがって、本実施の形態に係る提案手法は、MNMFを用いる場合に比べて、伝達系への外乱に対しての音源分離性能の低下が少なく、伝達系への外乱に対する頑健さの度合いが高いことができる。

10

**【 0 1 0 6 】**

このように、本発明に係る提案手法では、音源或いはマイクロホンの微小移動等、音響信号解析環境の軽微な変化に対して半時変形モデルを設定し、複数の時間フレームに亘るスペクトルを連結したスペクトログラム素片のテンプレートとアクティベーション系列とを畳み込む混合モデルを適用することで、音源及びマイクロホンの相対位置関係が変化する時变残響環境下であっても、複数の音源の音声が重畳した観測信号から、各音源の音源信号を精度よく分離することができる。

**【 0 1 0 7 】**

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。

20

**【 0 1 0 8 】**

例えば、上述の音響信号解析装置は、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。

**【 0 1 0 9 】**

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 】****【 0 1 1 0 】**

30

10 入力部

20 演算部

21 時間周波数解析部

22 初期設定部

23 パラメータ更新部

24 終了判定部

25 信号変換部

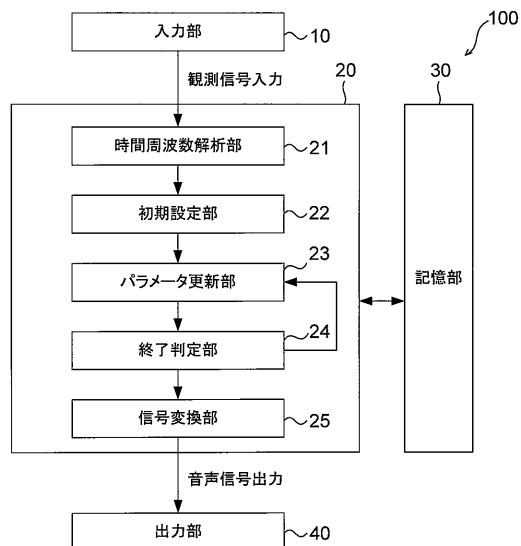
30 記憶部

40 出力部

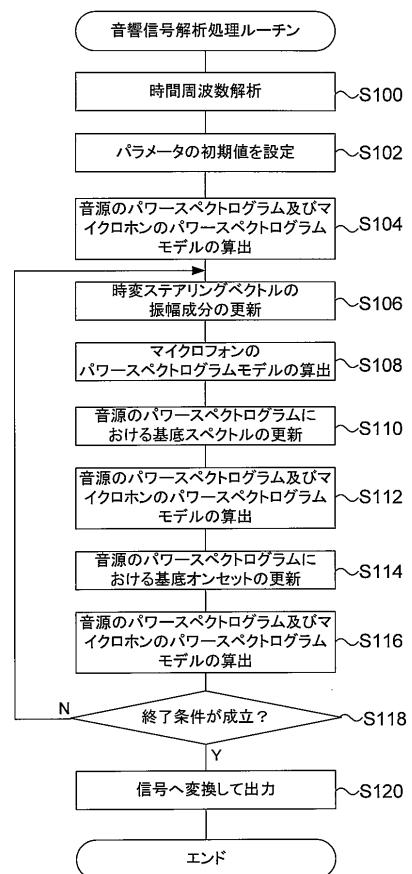
100 音響信号解析装置

40

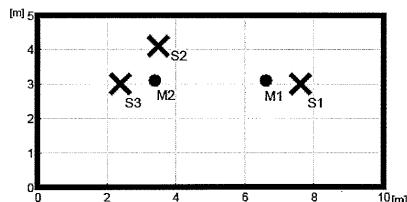
【図1】



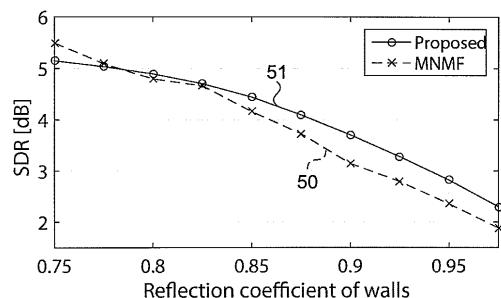
【図2】



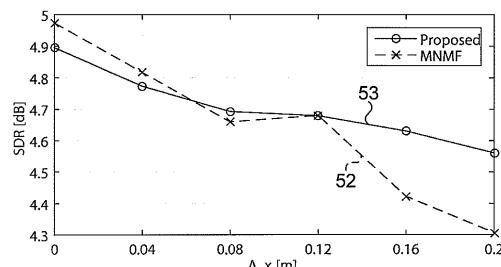
【図3】



【図4】



【図5】



---

フロントページの続き

(51)Int.Cl. F I  
H 04 R 1/40 (2006.01) H 04 R 1/40 320 A

審査官 大石 剛

(56)参考文献 特開2011-133780 (JP, A)  
特開2015-049406 (JP, A)  
特開2014-048398 (JP, A)  
特開2015-135437 (JP, A)  
特開2012-027196 (JP, A)

(58)調査した分野(Int.Cl., DB名)

G 10 L 21/0232  
G 10 L 21/0272  
G 10 L 21/028  
G 10 L 21/0308  
H 04 R 1/40  
H 04 R 3/00