

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第5629249号
(P5629249)

(45) 発行日 平成26年11月19日(2014.11.19)

(24) 登録日 平成26年10月10日(2014.10.10)

(51) Int.Cl. F I
G 1 0 L 25/51 (2013.01) G 1 0 L 25/51 4 0 0

請求項の数 5 (全 20 頁)

(21) 出願番号	特願2011-182774 (P2011-182774)	(73) 特許権者	000005326 本田技研工業株式会社 東京都港区南青山二丁目1番1号
(22) 出願日	平成23年8月24日(2011.8.24)	(74) 代理人	110000246 特許業務法人O F H特許事務所
(65) 公開番号	特開2013-44950 (P2013-44950A)	(72) 発明者	中臺 一博 埼玉県和光市本町8-1 株式会社ホンダ ・リサーチ・インスティテュート・ジャパ ン内
(43) 公開日	平成25年3月4日(2013.3.4)	(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
審査請求日	平成25年11月27日(2013.11.27)	(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 最終頁に続く

(54) 【発明の名称】 音源定位システム及び音源定位方法

(57) 【特許請求の範囲】

【請求項1】

音データを検出する音検出部と、
検出された音データに基づいて、方向及び時間ごとのMUSICスペクトルを演算する
MUSICスペクトル演算部と、

各方向の音源の有無による状態、状態の遷移を記述する状態遷移モデル、及び音源有りの
状態及び音源無しの状態に観測されるMUSICスペクトルを記述する観測モデルを定め、
MUSICスペクトルの時系列データに基づいて、前記観測モデル及び前記状態遷移
モデルのモデルパラメータの事後分布を推定する、モデルパラメータ推定部と、

推定されたモデルパラメータの事後分布に基づいて、方向及び時間ごとの音源存在事後
確率のパーティクルをサンプルすることによって、音源定位を行う音源定位部と、を備え
た音源定位システム。

【請求項2】

前記観測モデルとしてガウス混合モデルを使用する請求項1に記載の音源定位システム
。

【請求項3】

音データを検出するステップと、
検出された音データに基づいて、方向及び時間ごとのMUSICスペクトルを演算する
ステップと、

各方向の音源の有無による状態、状態の遷移を記述する状態遷移モデル、及び音源有り

10

20

の状態及び音源無しの状態に観測されるMUSICスペクトルを記述する観測モデルを定め、MUSICスペクトルの時系列データに基づいて、前記観測モデル及び前記状態遷移モデルのモデルパラメータの事後分布を推定するステップと、

推定されたモデルパラメータの事後分布に基づいて、方向及び時間ごとの音源存在事後確率のパーティクルをサンプルすることによって、音源定位を行うステップと、を含む音源定位方法。

【請求項 4】

前記観測モデルとしてガウス混合モデルを使用する請求項 3 に記載の音源定位方法。

【請求項 5】

音源定位を行うステップは、P 個のパーティクルをサンプルするステップと、各パーティクルについて重みを計算するステップと、各パーティクルの重みを正規化するステップと、各パーティクルの重みを使用してリサンプリングを行うステップと、を含む請求項 3 または 4 に記載の音源定位方法。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、音源定位システム及び音源定位方法に関する。

【背景技術】

【0002】

音源定位は、マイクロフォンアレイを使用した同時発話混合音声の分離、遠隔ロボット
のオペレータへの音源方向提示、移動ロボットによる音源検出と位置推定などに使用される重要な要素技術である。

【0003】

マイクロフォンアレイを使用した音源定位法には、ビームフォーミングに基づく方法と Multiple Signal Classification (MUSIC) に基づく方法がある。これらのうち、MUSIC に基づく MUSIC 法は、雑音に対してロバストであり、音源数がマイクロフォンの数未満という条件下では、比較的安定して複数音源の定位を行うことができる(たとえば、特許文献 1)。

【0004】

通常 MUSIC 法では、MUSIC スペクトルと呼ばれる音源到来評価関数に対して
閾値を設定して、音源がある方向に存在するかどうか判定する。適切な閾値を定めるには、環境中の音源数や残響時間などを考慮する必要がある。したがって、動的に音環境が変化する状況において音源定位を行うには、従来技術においては、閾値を手動で設定する必要があった。換言すれば、動的に音環境が変化する状況において、MUSIC スペクトルの閾値を自動的に定めることのできる音源定位システム及び音源定位方法は開発されていなかった。

【先行技術文献】

【特許文献】

【0005】

【特許文献 1】特許 4095348 号

【発明の概要】

【発明が解決しようとする課題】

【0006】

したがって、動的に音環境が変化する状況において、MUSIC スペクトルの閾値を自動的に定めることのできる音源定位システム及び音源定位方法に対するニーズがある。

【課題を解決するための手段】

【0007】

本発明の第 1 の態様による音源定位システムは、音データを検出する音検出部と、検出された音データに基づいて、方向及び時間ごとの MUSIC スペクトルを演算する MUSIC
スペクトル演算部と、を備えている。本システムは、各方向の音源の有無による状態

10

20

30

40

50

、状態の遷移を記述する状態遷移モデル、及び音源有りの状態及び音源無しの状態に観測されるMUSICスペクトルを記述する観測モデルを定め、MUSICスペクトルの時系列データに基づいて、前記観測モデル及び前記状態遷移モデルのモデルパラメータの事後分布を推定する、モデルパラメータ推定部と、推定されたモデルパラメータの事後分布に基づいて、方向及び時間ごとの音源存在事後確率のパーティクルをサンプルすることによって、音源定位を行う音源定位部と、をさらに備えている。

【0008】

本態様の音源定位システムによれば、観測モデル及び状態遷移モデルのモデルパラメータの事後分布を推定し、推定されたモデルパラメータの事後分布に基づいて音源定位を行うので、動的に音環境が変化する状況においても、手動で閾値を設定する必要なく、ロバ

10

【0009】

本発明の第1の態様の第1の実施形態による音源定位システムは、前記観測モデルとしてガウス混合モデルを使用する。

【0010】

本実施形態によれば、ガウス分布を使用することで解析的な計算が可能となる。

【0011】

本発明の第2の態様による音源定位システムは、音データを検出するステップと、検出された音データに基づいて、方向及び時間ごとのMUSICスペクトルを演算するステップと、を含む。本方法は、各方向の音源の有無による状態、状態の遷移を記述する状態遷移モデル、及び音源有りの状態及び音源無しの状態に観測されるMUSICスペクトルを記述する観測モデルを定め、MUSICスペクトルの時系列データに基づいて、前記観測モデル及び前記状態遷移モデルのモデルパラメータの事後分布を推定するステップと、推定されたモデルパラメータの事後分布に基づいて、方向及び時間ごとの音源存在事後確率のパーティクルをサンプルすることによって、音源定位を行うステップと、をさらに含む。

20

【0012】

本態様の音源定位方法によれば、観測モデル及び状態遷移モデルのモデルパラメータの事後分布を推定し、推定されたモデルパラメータの事後分布に基づいて音源定位を行うので、動的に音環境が変化する状況においても、手動で閾値を設定する必要なく、ロバ

30

【0013】

本発明の第2の態様の第1の実施形態による音源定位方法は、前記観測モデルとしてガウス混合モデルを使用する。

【0014】

本実施形態によれば、ガウス分布を使用することで解析的な計算が可能となる。

【0015】

本発明の第2の態様の第2の実施形態による音源定位方法において、音源定位を行うステップは、P個のパーティクルをサンプルするステップと、各パーティクルについて重みを計算するステップと、各パーティクルの重みを正規化するステップと、各パーティクルの重みを使用してリサンプリングを行うステップと、を含む。

40

【0016】

本実施形態によれば、推定されたモデルパラメータの分布に基づいてパーティクルをサンプルすることによって、方向及び時間ごとの音源事後確率のパーティクルを簡単な手順で求めることができる。

【図面の簡単な説明】

【0017】

【図1】本発明の一実施形態による音源定位システムの構成を示す図である。

【図2】M個のマイクロフォンからなるマイクロフォンアレイの構成を示す図である。

【図3】対数スケールのMUSICスペクトルの分布を示す図である。

50

【図4】VB-HMMの確率変数間の条件付き独立性を示すグラフィカルモデルを示す図である。

【図5】モデルパラメータ推定部による、モデルパラメータの分布を推定する手順を示す流れ図である。

【図6】音源定位部が、各方向ビンにおける音源存在事後確率を表す、P個のパーティクルを求める手順を示す流れ図である。

【図7】オンラインの音源定位実験に使用した音源の配置を示す図である。

【図8】従来の音源定位システムによるオンライン音源定位の結果を示す図である。

【図9】本実施形態の音源定位システムによるオンライン音源定位の結果を示す図である。

10

【発明を実施するための形態】

【0018】

図1は、本発明の一実施形態による音源定位システム100の構成を示す図である。音源定位システム100は、音検出部101と、MUSICスペクトル演算部103と、モデルパラメータ推定部105と、音源定位部107とを含む。

【0019】

音検出部101は、M個のマイクロフォンからなるマイクロフォンアレイであってもよい。

【0020】

図2は、M個のマイクロフォン1011からなるマイクロフォンアレイ101の構成を示す図である。図2において、M=8である。一例として、8個のマイクロフォン1011は水平面上に配置されており、音源定位システム100は、音源が、水平面上のどの方向に存在するかを定める。一例として、方向の解像度は5°であり、360°/5°=72のいずれかの方向を定める。

20

【0021】

たとえば、音検出部であるマイクロフォンアレイによって、Mチャンネルの音響信号が得られる。また、周波数ビンごとに水平面上のD方向(D=72)からの伝達関数を与えられているとする。音源定位システム100は、N個の音源方向を定める。ただし、同時に定位可能な最大音源数 N_{max} は、マイクロフォンの数未満である。

$$N \quad N_{max} < M$$

30

【0022】

MUSICスペクトル演算部103におけるMUSIC(Multiple Signal Classification)スペクトルの算出方法について説明する。より詳細は、文献[R. O. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," IEEE Trans. on Antennas and Propagation, vol. 34, no. 3, pp. 276-280, 1986; P. Dan`es and J. Bonnal, "Information-Theoretic Detection of Broadband Sources in a Coherent Beamspace MUSIC Scheme," in Proc. of IROS-2010, 2011, pp. 1976-1981.]に記載されている。MUSIC法は、時間周波数領域において適用される。具体的には、サンプリング周波数16000[Hz]で、窓長12[pt]、シフト幅160[pt]の短時間フーリエ変換を行っている。

40

【数1】

$$x_{r,\omega} \in \mathbb{C}^M$$

をMチャンネル音響信号の時間フレーム、周波数ビンにおける入力Mチャンネル音響信号の複素振幅ベクトルとする。各周波数ビン、T[sec]間隔の時刻tに対して、

(1)入力信号の自己相関行列

【数2】

$$R_{t,\omega}$$

の計算、

50

【数 3】

$$(2) \mathbf{R}_{t,\omega}$$

の固有値分解、

(3) 固有ベクトルと伝達関数を用いた MUSIC スペクトルの計算を行う。

【0023】

上記の(1)乃至(3)について以下に説明する。

(1) 入力信号の自己相関行列の計算

入力信号の自己相関行列は、時間 T で観測したサンプル値の相関として計算する。

【数 4】

$$\mathbf{R}_{t,\omega} = \frac{1}{\hat{\tau}(t) - \hat{\tau}(t - \Delta T)} \sum_{\tau=\hat{\tau}(t-\Delta T)}^{\hat{\tau}(t)} \mathbf{x}_{\tau,\omega} \mathbf{x}_{\tau,\omega}^H, \quad (1)$$

10

ただし、

【数 5】

$$(\cdot)^H$$

はエルミート転置、

【数 6】

$$\hat{\tau}(t)$$

20

は時刻 t における時間フレームを表す。入力ベクトル

【数 7】

$$\mathbf{x}_{\tau,\omega}$$

の M 個の要素は各チャネルに対応する。

(2) 固有値分解

【数 8】

$$\mathbf{R}_{t,\omega}$$

30

を次のように固有値分解する。

【数 9】

$$\mathbf{R}_{t,\omega} = \mathbf{E}_{t,\omega}^H \mathbf{Q}_{t,\omega} \mathbf{E}_{t,\omega}, \quad (2)$$

ここで、

【数 10】

$$\mathbf{E}_{t,\omega}$$

は固有ベクトル、

【数 11】

$$\mathbf{Q}_{t,\omega}$$

40

は固有値からなる対角行列である。

【数 12】

$$\mathbf{Q}_{t,\omega}$$

は、

【数 1 3】

$$\mathbf{E}_{t,\omega} = [\mathbf{e}_{t,\omega}^1 \dots \mathbf{e}_{t,\omega}^M]$$

と

【数 1 4】

$$\mathbf{R}_{t,\omega}$$

の M 個の固有ベクトルで表せ、

【数 1 5】

$$\mathbf{Q}_{t,\omega} = \text{diag}(q_{t,\omega}^1 \dots q_{t,\omega}^M)$$

10

となる。固有値

【数 1 6】

$$q_{t,\omega}^m$$

は降順に並べられているものとする。

【0 0 2 4】

入力信号に N 個の音源が含まれる場合、固有値

【数 1 7】

$$q_{t,\omega}^1$$

20

から

【数 1 8】

$$q_{t,\omega}^N$$

までが、音源のエネルギーに対応する大きな値を有する。それに対し、残りの固有値

【数 1 9】

$$q_{t,\omega}^{N+1}$$

から

30

【数 2 0】

$$q_{t,\omega}^M$$

まではマイクロフォンの観測ノイズなどに対応する小さな値を有する。ここで、重要な点は、

【数 2 1】

$$\mathbf{e}_{t,\omega}^{N+1}$$

から

【数 2 2】

$$\mathbf{e}_{t,\omega}^M$$

40

のノイズに対応する固有ベクトルは、音源方向に対応する伝達関数ベクトルと直交するという点である[R. O. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," IEEE Trans. on Antennas and Propagation, vol. 34, no. 3, pp. 276-280, 1986.]。

(3) 固有ベクトルと伝達関数を用いた MUSIC スペクトルの計算

MUSIC スペクトルは以下の式にしたがって計算する。

【数 2 3】

$$P_{t,d,\omega} = \frac{\|\mathbf{a}_{d,\omega}^H \mathbf{a}_{d,\omega}\|}{\sum_{m=N_{\max}+1}^M \|\mathbf{a}_{d,\omega}^H \mathbf{e}_{t,\omega}^m\|} \quad (3)$$

ただし、

【数 2 4】

$$\mathbf{a}_{d,\omega}$$

は方向 d 、周波数ビン に対応する M 次元の伝達関数ベクトルである。これらの伝達関数は、マイクロフォンアレイを使用して事前に測定したものである。ここで、観測される最大の音源数は N_{\max} である。そのため、

【数 2 5】

$$\mathbf{e}_{t,\omega}^{N_{\max}+1}$$

から

【数 2 6】

$$\mathbf{e}_{t,\omega}^M$$

までの固有ベクトルは、音源方向 d に対応する伝達関数

【数 2 7】

$$\mathbf{a}_{d,\omega}$$

と直交する。したがって、式 (3) の分母は音源方向の d に対して 0 となる。すなわち、式 (3) の MUSIC スペクトル

【数 2 8】

$$P_{t,d,\omega}$$

は発散する。ただし、実際には、壁からの反射音などの影響で MUSIC スペクトルは発散せず鋭いピークとして観測されることが多い。

【0025】

つぎに、周波数ビンごとの MUSIC スペクトルを以下の式にしたがって合算する。

【数 2 9】

$$P'_{t,d} = \sum_{\omega=\omega_{\min}}^{\omega_{\max}} \sqrt{q_{t,\omega}^1} P_{t,d,\omega}, \quad (4)$$

ここで、

【数 3 0】

$$q_{t,\omega}^1$$

は、周波数ビン における最大固有値である。本実施形態では、音声信号を対象とするため、

【数 3 1】

$$\omega_{\min} = 500 [\text{Hz}]$$

$$\omega_{\max} = 2800 [\text{Hz}]$$

とした。

【0026】

つぎに、モデルパラメータ推定部 105 の機能について説明する。モデルパラメータ推定部 105 は、変分ベイズ隠れマルコフモデル (VB-HMM) を使用する。

【 0 0 2 7 】

状態ベクトルとしてD次元の2値ベクトルを使用する。各次元のベクトル値は、その方向に音源が存在するか否かを示す。

【 0 0 2 8 】

また、MUSICスペクトルがガウス分布にしたがう観測値であるとみなし、音源有りの場合及び音源無しの場合に対応するガウス分布からなるガウス混合分布を観測モデルとする。ガウス分布を使用する理由は、複数の周波数ピンの値を加算して対数とったMUSICスペクトルが近似的にガウス分布とみなせることと、ガウス分布を使用することで解析的な計算が可能となることである。

【 0 0 2 9 】

図3は、対数スケールのMUSICスペクトルの分布を示す図である。図3の横軸は、対数スケールのMUSICスペクトルを示す。対数スケールのMUSICスペクトルは、以下の式によって求める。

【 数 3 2 】

$$x_{t,d} = 10 \log_{10} P'_{t,d} \quad (5)$$

図3の縦軸は、観測数を表す。図3に点線で示す音源が存在しない(オフ)のときのガウス分布は、狭くかつ値の小さなMUSICスペクトルの領域に形成され、図3に実線で示す音源が存在する(オン)のときのガウス分布は、広くかつ値の大きなMUSICスペクトルの領域を覆っている。

【 0 0 3 0 】

モデルパラメータ推定部105で使用される観測モデルは以下の式で表せる。

【 数 3 3 】

$$p(\mathbf{x}_t | \mathbf{s}_t, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \prod_{d=1}^D \prod_{j=0}^1 N(x_{t,d} | \mu_j, \lambda_j^{-1})^{\delta_j(s_{t,d})} \quad (6)$$

ただし、

【 数 3 4 】

$$\delta_y(x)$$

は $x = y$ のとき

【 数 3 5 】

$$\delta_y(x) = 1$$

そうでなければ、

【 数 3 6 】

$$\delta_y(x) = 0$$

を表す。また、

【 数 3 7 】

$$N(\cdot | \mu, \lambda^{-1})$$

は、平均 μ 、精度 λ の正規分布の確率密度関数を表す。パラメータ μ と λ には、正規・ガンマ分布を使用する。

【 数 3 8 】

$$p(\boldsymbol{\mu}, \boldsymbol{\lambda} | \beta_0, m_0, a_0, b_0) = \prod_{j=0}^1 N(\mu_j | m_0, (\beta_0 \lambda_j)^{-1}) G(\boldsymbol{\lambda}_j | a_0, b_0) \quad (7)$$

ただし、 $N(\cdot | m, L^{-1})$ は平均 m 、精度 L (分散 $1/L$) の正規分布であり、以下の式で表せる。

。

10

20

30

40

50

【数 3 9】

$$N(x|m, L^{-1}) = \sqrt{\frac{L}{2\pi}} \exp\left(-\frac{L(x-m)^2}{2}\right),$$

$G(\cdot|a, b)$ は形状 a 、尺度 b のガンマ分布であり、以下の式で表せる。

【数 4 0】

$$G(x|a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx)$$

正規分布の、ガンマ分布の a は事前分布の影響の大きさ(事前に観測したとみなす観測数)を表す。本実施形態では、事前情報よりも学習時に観測されるデータを重視するため

10

【数 4 1】

$$\beta_0=1, a_0=1$$

に設定する。

【0 0 3 1】

m_0 は平均パラメータ μ の事前情報で与えられた平均値を示す。本実施形態では、およそ25程度が適している。あるいは、VB-HMMの学習に用いる観測値のサンプル平均でも問題ない。

20

【0 0 3 2】

b_0 は精度パラメータの事前情報で与えられた「ばらけ具合」を示す。実験的に500に設定している。あるいは、VB-HMMの学習に用いる観測値のサンプル分散程度でも問題ない。

【0 0 3 3】

図4は、VB-HMMの確率変数間の条件付き独立性を示すグラフィカルモデルを示す図である。VB-HMMは、状態遷移確率のパラメータ κ や、観測確率のパラメータ μ 、 σ が数値ではなく、確率変数として扱われる点が通常のHMMと異なる。モデルパラメータ推定部105は、これらのパラメータの確率分布を学習する。

【0 0 3 4】

モデルパラメータ推定部105で使用される状態遷移モデルは基本的に、各方向ビン d について、前状態で音源がない場合

30

【数 4 2】

$$s_{t,d} = 0$$

と音源がある場合

【数 4 3】

$$s_{t,d} = 1$$

から、次状態で音源が出現する、継続する、消滅するといった遷移を考える。本実施形態においては、移動する音源についても考慮するために、以下の表1に示すように前状態の組み合わせから成る四つの場合を考える。すなわち、前時刻の同方向ビン

40

【数 4 4】

$$s_{t-1,d}$$

に音源が存在するかどうかと、前時刻の隣接方向ビン

【数 4 5】

$$s_{t-1,d\pm 1}$$

のいずれかに音源が存在するかによって分類する。たとえば、 $s_{t-1,d}$ は前時刻に当該方向 d 及び隣接ビン $d\pm 1$ に音源が存在しない状態から音源が出現する確率、 $s_{t-1,d}$ は、前時刻に方向 d に音源が存在しないが、隣接ビン $d\pm 1$ には音源が存在したため、その音源が方向 d に移動してきて

50

【数 4 6】

$$s_{t,d} = 1$$

となる確率を表す。

【表 1】

前状態 $s_{t-1,d}$	隣接前状態 $1 - s_{t-1,d-1} s_{t-1,d+1}$	音源存在確率 $p(s_{t,d} = 1 s_{t-1,d-1:d+1})$
0 (off)	0	θ_1
0 (off)	1	θ_2
1 (on)	0	θ_3
1 (on)	1	θ_4

10

【0 0 3 5】

状態遷移確率は、以下の式によって表せる。

【数 4 7】

$$p(s_t | s_{t-1}, \theta) = \prod_{d=1}^D \prod_{k=1}^4 \prod_{s_{t,d}=0}^1 (\theta_k^{s_{t,d}} (1-\theta_k)^{1-s_{t,d}})^{f_k(\theta_{t-1,d})} \quad (8)$$

ここで、

【数 4 8】

$$f_k(s_{t-1,d})$$

20

は表1 に従って、方向ビンd の周りの前状態の値

【数 4 9】

$$s_{t-1,d-1}, s_{t-1,d}, s_{t-1,d+1}$$

によって条件k に合致するときに

【数 5 0】

$$f_k(\cdot, d) = 1$$

30

その他の場合は0 を返す条件識別関数である。初期状態としては、音源は存在しない、すなわちすべてのd に対して

【数 5 1】

$$s_{0,d} = 0$$

とする。

【0 0 3 6】

状態遷移パラメータである

【数 5 2】

$$\theta = [\theta_1, \dots, \theta_4]$$

40

には、式(8)の共役事前分布としてベータ分布を用いる。

【数 5 3】

$$p(\theta | \alpha_0) = \prod_{k=1}^4 B(\theta_k | \alpha_{0,0}, \alpha_{0,1}) \quad (9)$$

ただし

【数 5 4】

$$B(\cdot | c, d)$$

はパラメータ c、d を持つ 分布の確率密度関数である。

50

【 0 0 3 7 】

モデルパラメータ推定部 1 0 5 における V B - H M M の学習は、事後分布

【数 5 5】

$$p(\mathbf{s}_{1:T}, \boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\lambda} | \mathbf{x}_{1:T})$$

を以下のように因数分解可能な分布に近似して推定する。

【数 5 6】

$$\begin{aligned} p(\mathbf{s}_{1:T}, \boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\lambda} | \mathbf{x}_{1:T}) &\approx q(\mathbf{s}_{1:T}, \boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\lambda}), \\ &= q(\mathbf{s}_{1:T})q(\boldsymbol{\theta})q(\boldsymbol{\mu}, \boldsymbol{\lambda}), \end{aligned} \quad (10) \quad 10$$

$(\cdot)_{1:T}$ は、時刻 1 から T までの確率変数の集合を表す。一般的な V B - H M M の推論については、文献[M. J. Beal, "Variational Algorithms for Approximate Bayesian Inference," Ph.D. dissertation, Gatsby Computational Neuroscience Unit, University College London, 2003.]に記載されている。

【数 5 7】

$$q(\boldsymbol{\theta}) = \prod_k q(\theta_k)$$

はそれぞれの k に対し、式(11)に示すパラメータ

【数 5 8】

$$\hat{\alpha}_{k,0}, \hat{\alpha}_{k,1} \quad 20$$

を持つベータ分布となり、

【数 5 9】

$$q(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \prod_j q(\mu_j, \lambda_j)$$

は、式(12), (13) で表されるパラメータ

【数 6 0】

$$\hat{\beta}_j, \hat{m}_j, \hat{a}_j, \hat{b}_j$$

を持つ正規ガウス分布となる。

【数 6 1】

$$\hat{\alpha}_{k,j} = \alpha_{0,j} + \sum_{t,d} \langle s_{t,d,j} f_k(s_{t-1}, d) \rangle, \quad (11)$$

$$\hat{\beta}_j = \beta_0 + w_j, \hat{m}_j = (\beta_0 m_0 + w_j \bar{x}_j) / (\beta_0 + w_j), \quad (12)$$

$$\hat{a}_j = a_0 + \frac{w_j}{2}, \hat{b}_j = b_0 + \frac{w_j S_j^2}{2} + \frac{\beta_0 w_j (\bar{x}_j - m_0)^2}{2(\beta_0 + w_j)}, \quad (13) \quad 40$$

ただし、変数 $s_{t,d,j}$ は、 $s_{t,d} = 0$ のとき、 $s_{t,d,0} = 1$ 、また、 $s_{t,d} = 1$ のとき、 $s_{t,d,1} = 1$ となる変数である。式(12)、(13)に用いられる正規分布の十分統計量は

【数 6 2】

$$w_j = \sum_{t,d} \langle s_{t,d,j} \rangle, \bar{x}_j = \frac{\sum_{t,d} \langle s_{t,d,j} \rangle x_{t,d}}{w_j}, S_j^2 = \frac{\sum_{t,d} \langle s_{t,d,j} \rangle (x_{t,d} - \bar{x}_j)^2}{w_j}.$$

と定義する。また、

【数 6 3】

$$\langle \cdot \rangle$$

は式(10)の分布による期待値演算子である。各時刻の状態変数と状態遷移の期待値

50

【数 6 4】

$$\langle s_{t,d,j} \rangle, \langle s_{t,d,j} f_k(s_{t-1}, d) \rangle$$

は次のように計算する。

【数 6 5】

$$\langle s_{t,d,j} \rangle \propto \alpha(s_{t,d,j}) \beta(s_{t,d,j}), \quad (14)$$

$$\langle s_{t,d,j} f_k(s_{t-1}, d) \rangle \propto \tilde{\alpha}(s_{t-1,d,k}) \tilde{p}(s_{t,d}|s_{t-1}) \tilde{p}(x_{t,d}|s_{t,d}) \beta(s_{t,d,j}), \quad (15)$$

ただし、 $(s_{t,d,j})$ と $(s_{t-1,d,k})$ はそれぞれ前向き、後ろ向き再帰式により計算される。

10

【数 6 6】

$$\alpha(s_{t,d,j}) \propto \sum_{k=1}^4 \tilde{\alpha}(s_{t-1,d,k}) \tilde{p}(s_{t,d}|s_{t-1}) \tilde{p}(x_{t,d}|s_{t,d}), \quad (16)$$

$$\beta(s_{t,d,j}) = \sum_{j'=0}^1 \beta(s_{t+1,d,j'}) \tilde{p}(s_{t+1,d,j'}|s_{t,d,j}) \tilde{p}(x_{t,d}|s_{t,d}). \quad (17)$$

ここで、

20

【数 6 7】

$$\begin{aligned} \tilde{p}(s_{t,d}|s_{t-1}) &\equiv C \exp(E_{q(\theta)}[\log p(s_{t,d}|s_{t-1}, \theta)]) \\ &= C \exp\left(\int \log p(s_{t,d}|s_{t-1}, \theta) q(\theta) d\theta\right) \end{aligned}$$

$$\begin{aligned} \tilde{p}(x_{t,d}|s_{t,d}) &= C \exp(E_{q(\mu, \lambda)}[\log p(x_{t,d}|s_{t,d}, \mu, \lambda)]) \\ &= C \exp\left(\iint \log p(x_{t,d}|s_{t,d}, \mu, \lambda) q(\mu, \lambda) d\mu d\lambda\right) \end{aligned}$$

である。遷移、観測確率の幾何平均は以下の式で表せる。

【数 6 8】

30

$$\tilde{p}(s_{t,d} = j | s_{t-1}) \propto \exp\{\psi(\hat{\alpha}_{k,j}) - \psi(\hat{\alpha}_{k,0} + \hat{\alpha}_{k,1})\} \quad (18)$$

$$\tilde{p}(x_{t,d}|s_{t,d}) \propto \prod_j \exp\left\{ \frac{\psi(\hat{a}_j) - \log \hat{b}_j - 1/\hat{\beta}_j}{2} - \frac{a_j(x_{t,d} - \hat{m}_j)^2}{2\hat{b}_j} \right\}^{s_{t,d,j}} \quad (19)$$

ここで、

【数 6 9】

$$\Psi(\cdot)$$

40

は、以下の式で定義されるディガンマ関数である。

【数 7 0】

$$\Psi(x) \equiv \frac{d}{dx} \log \Gamma(x)$$

【0 0 3 8】

式(14)及び(15)は、ともに添え字 j 、 k を動かしたとき総和が1になるように正規化されている。

【数 7 1】

$$\tilde{\alpha}(s_{t-1,d,k})$$

50

は、状態遷移の条件 k に関する前向き確率である。

【 0 0 3 9 】

図 5 は、モデルパラメータ推定部 1 0 5 による、モデルパラメータの分布を推定する手順を示す流れ図である。

【 0 0 4 0 】

図 5 のステップ S 1 0 1 0 において、モデルパラメータ推定部 1 0 5 は、初期値を定める。初期値は、式(14)及び式(15)の値に対して、たとえば、以下手順によって定めることができる。

【 0 0 4 1 】

式(14) の左辺 $\langle s_{t,d,j} \rangle$ は、時刻 t 、方向ビン d において音源が存在しない場合は $s_{t,d,0} = 1$ かつ $s_{t,d,1} = 0$ 、音源が存在する場合は $s_{t,d,0} = 0$ かつ $s_{t,d,1} = 1$ となる2値変数の期待値である。観測値 $x_{t,d}$ が適当な閾値 (たとえば、 m_0 の値) を超える場合は、たとえば

$$\langle s_{t,d,1} \rangle = 0.8, \quad \langle s_{t,d,0} \rangle = 1 - 0.8 = 0.2$$

のように設定する。0.8の代わりに1にしてもほぼ同様に動作する。

【 0 0 4 2 】

式(15) の左辺 $\langle s_{t,d,j} f_k(s_t, d) \rangle$ も、 $x_{t,d}$ が閾値を超えるか否かによって算出する。この値は、 $s_{t,d,1} = 0, 1$ の2通りと、 $k=1 \sim 4$ のいずれかで $f_k(s_t, d) = 1$ となる4通りの組み合わせで、計8通りの組み合わせが存在する。 $x_{t,d}$ の閾値処理と、前時刻の $x_{t-1,d}$ の閾値処理、前時刻・近傍方向の $x_{t-1,d \pm 1}$ の閾値処理結果から、表1を参照して、該当する k の値を求める。たとえば、前時刻 $x_{t-1,d}$ は閾値以下だが、隣接ビンの $x_{t-1,d+1}$ では閾値を超えた場合は $k = 2$ にあたる。 $x_{t,d}$ が閾値を超えた場合、 $\langle s_{t,d,1} f_2(s_t, d) \rangle = 0.8$ 、その他の7通りの組み合わせでは、 $\langle s_{t,d,j} f_k(s_t, d) \rangle = (1 - 0.8) / 7$ のように設定する。

【 0 0 4 3 】

図 5 のステップ S 1 0 2 0 において、モデルパラメータ推定部 1 0 5 は、式(18)及び式(19)を使用して、遷移、観測確率の幾何平均を求める。

【 0 0 4 4 】

図 5 のステップ S 1 0 3 0 において、モデルパラメータ推定部 1 0 5 は、ステップ S 1 0 2 0 で求めた遷移、観測確率の幾何平均、式(16)及び式(17)を使用して、 $(s_{t,d,j})$ と $(s_{t,d,j})$ を計算する。

【 0 0 4 5 】

図 5 のステップ S 1 0 4 0 において、モデルパラメータ推定部 1 0 5 は、ステップ S 1 0 3 0 で求めた $(s_{t,d,j})$ と $(s_{t,d,j})$ 、式(14)及び式(15)を使用して、各時刻の状態変数と状態遷移の期待値を求める。

【 0 0 4 6 】

図 5 のステップ S 1 0 5 0 において、モデルパラメータ推定部 1 0 5 は、ステップ S 1 0 4 0 で求めた各時刻の状態変数と状態遷移の期待値、式(11)から式(13)を使用して、モデルパラメータの事後分布を推定する。

【 0 0 4 7 】

図 5 のステップ S 1 0 6 0 において、モデルパラメータ推定部 1 0 5 は、収束判定を行う。具体的に、式(12)、(13) で算出されるパラメータ m, a, b の値の変動がなくなることによって収束判定を行う。収束していないと判断されれば、ステップ S 1 0 2 0 に戻る。収束していると判断されれば、処理を終了する。

【 0 0 4 8 】

つぎに、音源定位部 1 0 7 の機能について説明する。音源定位部 1 0 7 は、モデルパラメータ推定部 1 0 5 が推定した、モデルパラメータの事後分布に基づいて、パーティクルフィルタを使用して、複数音源の存在事後確率計算を行う。パーティクルフィルタの推定対象は、MUSICスペクトルの時系列データが与えられたときの、各方向ビンにおける音源存在事後確率である。この分布を、 P 個のパーティクルを用いて以下のように近似計算する。

10

20

30

40

50

【数 7 2】

$$p(\mathbf{s}_t | \mathbf{x}_{1:t}) \approx w_p \mathbf{s}_t^p, \quad (20)$$

ただし、

【数 7 3】

$$w_p$$

はパーティクル p の重み、

【数 7 4】

$$\mathbf{s}_t^p$$

10

は状態ベクトルの値である。

【0049】

図6は、音源定位部107が、各方向ビンにおける音源存在事後確率を表す、 P 個のパーティクルを求める手順を示す流れ図である。

【0050】

図6のステップS2010において、音源定位部107は、 P 個のパーティクルをサンプリングする。

【0051】

P は以下のように定める。 P が大きいほど式(20)の近似は正確になるが、 P の大きさに比例した計算時間がかかる。そこで、 P を決める一般的な手続きとしては、実用的な近似を得るほど十分大きな P を与え、このような P の処理時間が大きすぎる場合は、 P の大きさを適宜縮小する。本実施形態においては、近似結果が収束し、理速度も十分高速な値として、 $P = 500$ とした。

20

【0052】

P 個のパーティクルのサンプルは、以下の式で表せる分布を使用して行う。

【数 7 5】

$$\mathbf{s}_t^p \sim q(\mathbf{s}_t | \mathbf{x}_t, m, a, b), \quad (21)$$

$$q(\mathbf{s}_t^p | \mathbf{x}_t, \hat{m}, \hat{a}, \hat{b}) \propto \prod_{d=1}^D \prod_{j=0}^1 C(x_{t,d})^{s_{t,d,j}^p} \exp(-\Delta_{d,j}^2 / 2)^{s_{t,d,j}^p} \quad (22)$$

30

ただし、 $x_{t,d}$ が極大値をとる d のとき、 $C(x_{t,d})=1$ でその他の場合は、 $C(x_{t,d})=0$ となる。上記の分布の重みには、以下の式で表せるマハラノビス距離を使用する。

【数 7 6】

$$\Delta_{d,j}^2 = (x_{t,d} - \hat{m}_j)^2 \hat{a}_j / \hat{b}_j$$

【0053】

時刻 t において、式(22)で計算される分布 q は、計 D 個のビンに対してオン

40

【数 7 7】

$$(s_{t,d,1}^p = 1)$$

であるか、またはオフ

【数 7 8】

$$(s_{t,d,0}^p = 1)$$

であるかの確率を与える。サンプル方法としては、各 d について

a) $C(x_{t,d})=0$ の場合、 $j=0$ とする。つまり

【数 7 9】

$$s_{t,d,0}^p = 1$$

b) $C(x_{t,d})=1$ の場合、 $j=0,1$ それぞれの場合の分布 q の確率を参照する。たとえば、

【数 8 0】

$$\exp(-\Delta_{d,0}^2) : \exp(-\Delta_{d,1}^2) = 8 : 2$$

の場合、0~1の区間から一様乱数を生成し、その値が0.8以下の場合、

【数 8 1】

$$s_{t,d,0}^p = 1$$

10

、0.8を超えた場合は

【数 8 2】

$$s_{t,d,1}^p = 1$$

のように値を決定する。

【0 0 5 4】

図6のステップS2020において、音源定位部107は、以下の式にしたがって、各パーティクルについて重み w_p を算出する。

【数 8 3】

20

$$w_p \propto \frac{\bar{p}(x_t | s_t^p) \bar{p}(s_t^p | s_{t-1}^p)}{q(s_t^p | x_t, \hat{m}, \hat{a}, \hat{b})}, \quad (23)$$

$$\bar{p}(x_t | s_t^p) = \int p(x_t | s_t^p, \mu, \lambda) q(\mu, \lambda) d\mu d\lambda, \quad (24)$$

$$\bar{p}(s_t^p | s_{t-1}^p) = \int p(s_t^p | s_{t-1}^p, \theta) q(\theta). \quad (25) \quad 30$$

式(24)及び式(25)の状態遷移、観測確率は、モデルパラメータ推定部105で使用された式(6)及び式(8)の事後分布で積分消去することで計算することができる。分布の共役性を用いると、この積分計算は、次のように解析的に求まる。

【数 8 4】

$$\bar{p}(x_t | s_t^p) = \prod_d St(x_{t,d} | \hat{m}_j, \frac{\hat{\beta}_j \hat{a}_j}{(1 + \hat{\beta}_j) \hat{b}_j}, 2\hat{a}_j)^{s_{t,d,j}^p} \quad (26)$$

$$\bar{p}(s_t^p | s_{t-1}^p) = \prod_d \prod_k \left(\hat{\alpha}_{k,s_{t,d}} / (\hat{\alpha}_{k,0} + \hat{\alpha}_{k,1}) \right)^{f_k(s_{t-1}^p, d)} \quad (27) \quad 40$$

ただし、

【数 8 5】

$$St(\cdot | m, \lambda, \nu)$$

は、平均 m 、精度 λ 、自由度 ν のStudent t-分布である。さらに、最大の音源数を N_{max} に抑えるため、状態ベクトル

【数 8 6】

$$s_t^p$$

に存在する音源数が N_{\max} を超える場合には観測確率は

【数 8 7】

$$\tilde{p}(x_t | s_t^p) = 0$$

とする。

【0 0 5 5】

図 6 のステップ S 2 0 3 0 において、音源定位部 1 0 7 は、各パーティクルの重み w_p が

【数 8 8】

$$\sum_{p=1}^P w_p = 1$$

となるよう正規化する。

【0 0 5 6】

図 6 のステップ S 2 0 4 0 において、処理を終了するかどうか判断する。たとえば、スイッチの状態によって処理を終了するかどうか判断してもよい。処理を終了しないと判断した場合には、ステップ S 2 0 5 0 に進む。そうでなければ処理を終了する。

【0 0 5 7】

図 6 のステップ S 2 0 5 0 において、音源定位部 1 0 7 は、リサンプリングを行う。リサンプリングは、パーティクル p が持つ値

【数 8 9】

$$s_t^p$$

を、そのパーティクルの重み

【数 9 0】

$$w_p$$

に比例した確率で複製することによって行う。方法の一例としては、以下の処理を $p' = 1 \sim P$ まで繰り返す。

a) $0 \sim 1$ 区間から一様乱数を生成 $r_{p'}$ とする。

b) $p = 1 \sim P$ まで

$$i \cdot r_{p'} - r_{p'} \cdot w_p$$

$$i \cdot i \cdot r_{p'} < 0 \text{ になったら、}$$

【数 9 1】

$$s_t^{p'} \leftarrow s_t^p$$

として、 p のループを抜ける。

$$i \cdot i \cdot i \cdot w_p = 1/P \text{ (リサンプル後の重みは全てのパーティクルで等しい)}$$

c) a) に戻る。

【0 0 5 8】

つぎに、評価実験について説明する。評価実験は、本実施形態による音源定位システム及び固定閾値を使用する従来の音源定位システムの性能を比較した。モデルパラメータ推定部 1 0 5 による、オフラインでの V B - H M M での学習は、一人の話者がマイクロフォンの周囲を発話しながら動く音響信号で行った。

【0 0 5 9】

図 7 は、オンラインの音源定位実験に使用した音源の配置を示す図である。二人の話者 3 0 1 及び 3 0 3 が、マイクロフォンアレイ 1 0 1 の周囲を移動しながら発話し、固定されたスピーカ 2 0 1 から音楽が再生されている。オフライン及びオンラインで用いられ

10

20

30

40

50

た信号の長さはともに20 [sec] である。

【 0 0 6 0 】

パラメータの設定は以下のとおりである。

$$N_{\max} = 3, \quad \theta_0 = [1, 1], \quad \sigma_0 = 1, \quad a_0 = 1, \quad b_0 = 500$$

また、パーティクル数は $P = 500$ とした。実験で使用した室内の残響時間は $RT_{20} = 840$ [msec] であった。

【 0 0 6 1 】

図8は、従来の音源定位システムによるオンライン音源定位の結果を示す図である。図の横軸は秒単位の時間を示し、図の縦軸は度単位の方向を示す。従来の音源定位システムの閾値は $P_{\text{thres}} = 23, 25, 27$ に設定されている。図8(a)、(b)、(c)は、それぞれ、閾値が23, 25, 27の場合の音源定位の結果を示す。図8(a)、(b)、(c)においては、閾値を超えるピンを、音源が存在するとして黒く表示している。図8(a)、(b)、(c)において、固定されたスピーカ及び移動する話者が黒で表示されている。しかし、図8(a)、(b)の実線の囲みで示すように、閾値を低く設定した場合は音源の誤検出が頻発する。

10

【 0 0 6 2 】

図9は、本実施形態の音源定位システムによるオンライン音源定位の結果を示す図である。図の横軸は秒単位の時間を示し、図の縦軸は度単位の方向を示す。本実施形態の音源定位システムの初期値は $m_0 = 23, 25, 27$ に設定されている。図9(a)、(b)、(c)は、それぞれ、初期値が23, 25, 27の場合の音源定位の結果を示す。図9(a)、(b)、(c)においては、事後分布の音源存在確率が0.95以上のピンを、音源が存在するとして黒く表示している。図9(a)、(b)、(c)において、固定されたスピーカ及び移動する話者が黒で表示されている。図8(a)、(b)の実線の囲みの位置に対応する、図9(a)、(b)の点線の囲みの位置において音源の誤検出は少ない。このことは、本実施形態の音源定位システムにおいて、学習の初期値にかかわらず音源の誤検出が少ないことを意味する。また、本実施形態の音源定位システムにおいて音源存在確率の閾値を0.95から1.00まで動かして結果を検証し、これらの閾値に対してもロバストに同様の結果を示すことを確認した。この結果から、本実施形態の音源定位システムのモデルパラメータ推定部105によるオフライン学習及び音源定位部107によるオンライン定位の枠組みが、自動的に音源定位に適したパラメータに収束することが確認できた。さらに、今回の実験条件から、本実施形態の音源定位方法は、1音源のみで学習した場合であっても、複数音源に対して安定した音源定位結果をもたらすことが実証された。

20

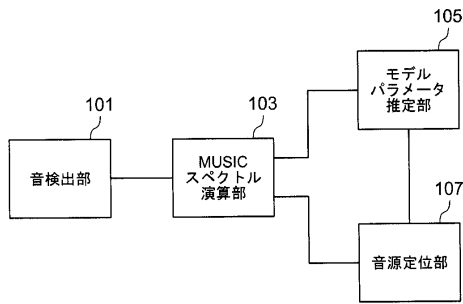
30

【符号の説明】

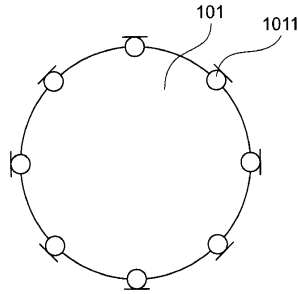
【 0 0 6 3 】

101...音検出部、103...MUSICスペクトル演算部、105...モデルパラメータ推定部、107...音源定位部

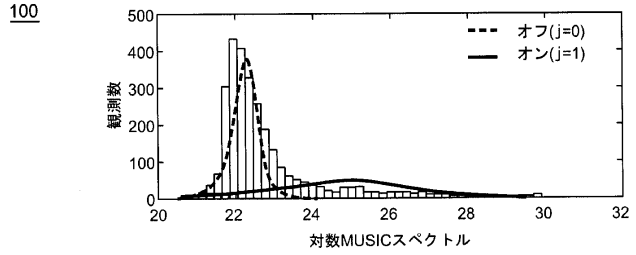
【図1】



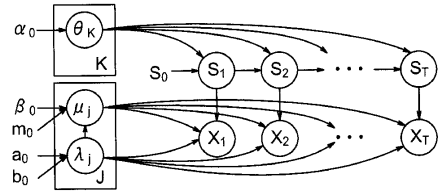
【図2】



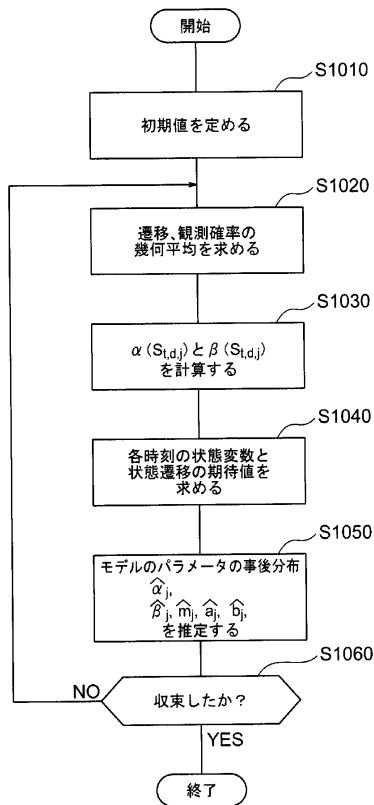
【図3】



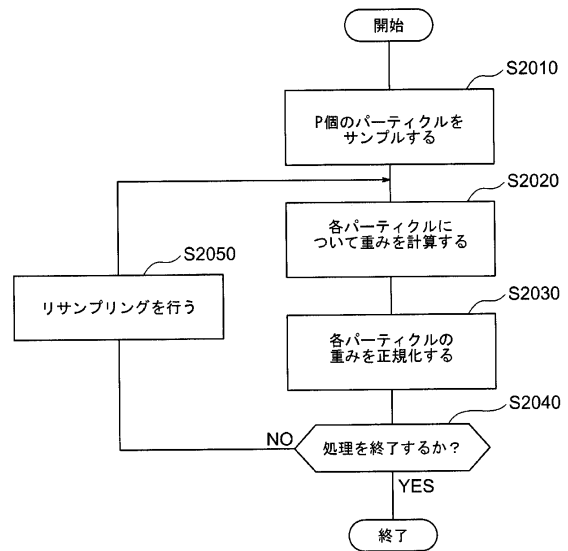
【図4】



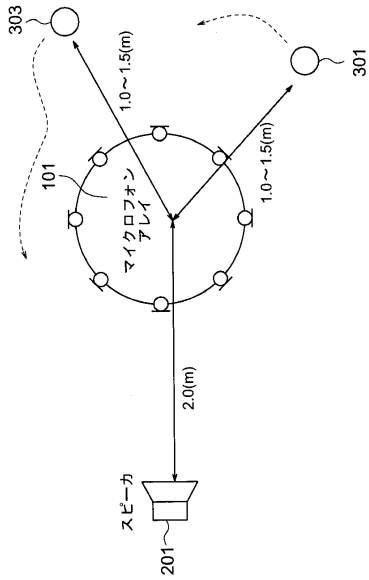
【図5】



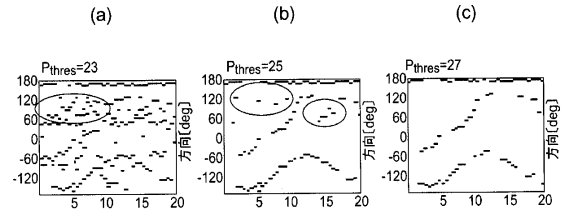
【図6】



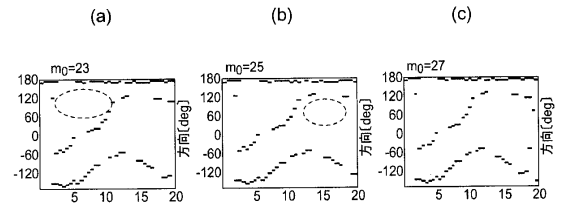
【 図 7 】



【 図 8 】



【 図 9 】



フロントページの続き

審査官 上田 雄

(56)参考文献 特開2010-121975(JP,A)
米国特許出願公開第2006/0245601(US,A1)

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