

# マイクロホンアレイを搭載した複数ロボットによる 自己位置・同期ずれ・音源位置・音源信号の推定

Sound Source Localization and Separation and Self-Localization  
Using Asynchronous Distributed Microphone Arrays

関口 航平 (京大) 坂東 宜昭 (京大) 糸山 克寿 (京大) 吉井和佳 (京大)

Kouhei Sekiguchi<sup>1</sup>, Yoshiaki bando<sup>1</sup>, Katsutoshi Itoyama<sup>1</sup>, Kazuyoshi Yoshii<sup>1</sup>

<sup>1</sup>Kyoto University

This paper presents a method of simultaneous localization and mapping (SLAM) for estimating the positions of multiple sound sources and those of stationary robots and synchronizing microphone arrays attached to those robots. Since each robot with a microphone array can solely estimate the directions of sound sources, the two-dimensional sound positions can be estimated from the source directions estimated by multiple robots. In addition, sound mixtures can be separated accurately by regarding distributed microphone arrays as one big array. To perform these tasks, the robot positions and synchronization between microphone arrays are necessary. The proposed method estimates the posterior distribution of the positions and time offsets and conducts source separation simultaneously in a Bayesian manner, given the observed signals. We conducted experiments using three robots and four sound sources. When the two of the model parameters (robot positions, sound source positions, and time offsets) were fixed to the correct value, the other one was correctly estimated and the observed signals were separated precisely. However, when all of the parameters were estimated simultaneously, they cannot be estimated correctly because of many local optimal solutions of the posterior distribution.

**Key Words:** Robot audition, Multiple robots

## 1 はじめに

実環境においてロボットが自律的に行動するうえで、周囲に存在する音源の情報を取得することは重要な課題の一つである [1]。これまでマイクロホンアレイを搭載したロボット 1 台を用いて音源の到来方向推定や音源分離を行う研究がなされてきた [2, 3]。ロボット 1 台を用いた場合には、音源の方向しか推定することができないが、複数ロボットを用いることで 2 次元平面上での音源位置を推定することが可能となる [4, 5]。また、複数台のロボットを協調させ、一つの大きなマイクロホンアレイとみなして音源分離を行うことで [6]、雑音が多く存在するような環境においても頑健な音源分離が実現できると期待される。

複数台のロボットを協調させてマイクロホンアレイ処理を行うためには、各ロボットの位置とロボット間でのマイクロホンアレイの同期が必要となるが、これらの推定には特別な機器が必要である。そのため、ロボットの周囲にある複数の音源を用いて、その音源の位置とロボットの位置、同期時刻ずれを同時に推定する研究がなされてきた [7]。この手法では、観測音から音源到来方向とロボット間の到達時間差の 2 種類の特徴量を抽出し、その特徴量を用いて FastSLAM2.0 という手法によりロボット位置、音源位置、同期時刻ずれを推定するという 2 段階で行われており、特徴量の抽出に失敗すると推定が破綻するという問題があった。

本稿では、マイクロホンで音が観測される過程をモデル化し、実際に観測音が得られた際の事後分布の期待値を推定することにより、観測音からロボット位置・音源位置・同期時刻ずれを直接推定し、さらに音源分離も同時に行う手法について述べる。事後分布の期待値を解析的に計算することが困難なため、Gibbs サンプリングを用いて事後分布からのサンプリングを行い、サンプルの集合で期待値を近似した。実験では、全ての同時推定を行うことはできていないが、ロボット位置・音源位置・同期時刻ずれのうちの 1 つと、音源分離を同時に行えることは確認できた。

## 2 マイクロホンアレイ位置・音源位置・同期時刻ずれ推定と音源分離の統合ベイズモデル

マイクロホンアレイを搭載した  $I$  台の静止したロボット、 $K$  個の静止した音源が存在する状況において、二次元平面上での各ロボットの位置・向き  $(r_i^x, r_i^y, r_i^\theta \ i = 1, \dots, I)$ 、各音源の位置  $(s_k^x, s_k^y \ n = 1, \dots, N)$ 、ロボット 1 を基準としたときのロボット  $d$  の同期時刻ずれ  $(\tau_d \ d = 2, \dots, I)$  を推定する。ただし、音源数  $K$  とロボット数  $I$  は、変数の数が制約式の数より多くなることが必要であるため、 $2KI - 4(K + I - 1) > 0$  を満たすことが必要である。以下に本章で扱う問題設定を示す。

- 入力 マイクロホンアレイを搭載した  $I$  台のロボットでの観測音  $\mathbf{x} \in \mathbb{C}^{T \times F \times M}$
- 出力 (1) ロボット  $i$  の位置、向き  $(r_i^x, r_i^y, r_i^\theta)$   
(2) 音源  $k$  の位置  $(s_k^x, s_k^y)$   
(3) ロボット 1 と  $d$  の間の同期時刻ずれ  $\tau_d$   
(4) 音源  $k$  の分離音  $\mathbf{y} \in \mathbb{C}^{T \times F \times N}$
- 仮定 (1) 各ロボット、音源は静止

ただし、 $T, F, M$  はそれぞれ時間フレーム数、周波数ビン数、全てのロボットに搭載されたマイクロホン数の合計を表す。

### 2.1 モデル化

$M$  チャンネルの観測音  $\mathbf{x}$  と  $K$  個の音源信号  $\mathbf{z} \in \mathbb{C}^{T \times F \times K}$  の関係は、周波数領域での瞬時混合を仮定すると、以下の式で表される。

$$\mathbf{x}_{tf} = \sum_k \mathbf{a}_{fk} z_{tfk} \quad (1)$$

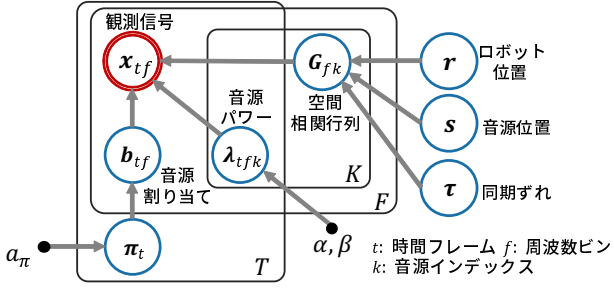


Fig.1 Graphical model of the proposed method

ここで、 $\mathbf{a}_{fk} \in \mathbb{C}^{M \times 1}$  は周波数  $f$  での音源  $k$  とマイクロホン間の伝達系を表すステアリングベクトルであり、次式で計算される。

$$a_{fkm} = \frac{1}{l_{km}} \exp(-2\pi f j \{ \frac{l_{km}}{C} - \tau_{1i} \}) \quad (2)$$

ここで、 $C$  は音速、 $j$  は複素数、 $l_{km}$  は音源  $k$  とマイクロホン  $m$  の間の距離、 $\tau_{1i}$  はマイクロホン  $m$  が搭載されたロボットとロボット 1 の間の同期時刻ずれを表す。音源信号  $z_{tfk}$  のパワーが  $\lambda_{tfk}$  であるとすると、 $z_{tfk}$  は平均 0、分散  $\lambda_{tfk}$  の複素 Gauss 分布に従うと仮定することができ、観測音  $x_{tf}$  は以下の複素 Gauss 分布に従う。

$$\mathbf{x}_{tf} \sim \mathcal{N}_C(0, \lambda_{tfk} \mathbf{G}_{fk}^{-1}) \quad (3)$$

ただし、 $G_{fk}$  は周波数  $f$  での音源  $k$  の空間相関行列を表し、ステアリングベクトル  $\mathbf{a}_{fk}$  を用いて  $G_{fk} = \mathbf{a}_{fk} \mathbf{a}_{fk}^H$  で計算される。更に、観測音  $x$  の 1 つの時間周波数ビンが 1 つの音源のみに依存することを仮定することにより、提案法の観測モデルを次式で定める。

$$p(\mathbf{x} | \lambda, G, \mathbf{b}) = \prod_{t,f,k} \mathcal{N}_C(\mathbf{x}_{tf} | 0, \lambda_{tfk} \mathbf{G}_{fk}^{-1})^{b_{tfk}} \quad (4)$$

ただし、 $b_{tfk}$  は  $x_{tf}$  が音源  $k$  に依存するかどうかを表す。音源分離を行う際には  $\mathbf{b}$  の期待値からソフトマスクを作成し、それを用いて観測音を各音源に割り当てる。

観測モデルに式 (4) を用いる場合、観測音  $\mathbf{x}$  は空間相関行列  $G$ 、音源パワー  $\lambda$ 、音源の割り当て  $\mathbf{b}$  に依存し、音源の割り当て  $\mathbf{b}$  は各音源の出やすさを表す  $\pi$  に依存する。また、空間相関行列  $G$  はロボット位置  $r$ 、音源位置  $s$ 、同期時刻ずれ  $\tau$  に依存する。従って、提案法のグラフィカルモデルは図 1 のようになる。 $G$  は理論上  $r, s, \tau$  から計算することができるにもかかわらず、提案法が  $G$  を潜在変数として持つのは、残響や障害物などの影響により幾何的に計算される空間相関行列と実際の空間相関行列が異なり、その違いを補完するためである。

観測モデル (式 (4)) から、 $G, \lambda, \mathbf{b}$  の事前分布を以下のように定める。

$$p(G_{fk} | r, s, \tau) = \mathcal{W}_C(G_{fk} | \nu, \frac{1}{\nu} G_{fk}^0) \quad (5)$$

$$p(\lambda_{tfk} | \alpha, \beta) = \text{GIG}(\lambda_{tfk} | \alpha, \beta, 0) \quad (6)$$

$$p(\mathbf{b}_{tf} | \pi_t) = \text{Categorical}(\mathbf{b}_{tf} | \pi_t) \quad (7)$$

ただし、 $\mathcal{W}_C$  は複素 Wishart 分布であり以下の式で表される。

$$\mathcal{W}_C(G | \nu, G_0) = \frac{|G|^{\nu-M} \exp(-\text{tr}(G G_0^{-1}))}{|G_0|^\nu \pi^{M(M-1)/2} \prod_{m=0}^{M-1} \Gamma(\nu-m)} \quad (8)$$

ただし、 $M$  は  $G$  の次元数を表す。 $\nu (> M)$  は自由度パラメータであり、値が大きいくほど  $G$  の事後分布に対する事前分布の影響が大きくなる。 $G_{fk}^0$  はロボット位置  $r$ 、音源位置  $s$ 、同期時刻ずれ  $\tau$  から幾何的に計算される空間相関行列であり、式 (2) で計算したステアリングベクトル  $\mathbf{a}_{fk}$  を用いて  $\mathbf{a}_{fk} \mathbf{a}_{fk}^H + \epsilon I$  で計算され

る。 $\epsilon I$  を加えたのは、 $G_{fk}^0$  を正則行列にするためである。GIG は Generalized Inverse Gaussian 分布であり、以下の式で表される。

$$\text{GIG}(x | \gamma, \rho, \tau) = \frac{\exp((\gamma-1) \log x - \rho x - \frac{\tau}{x})}{2\tau^{\gamma/2} K_\gamma(2\sqrt{\rho\tau})} \quad (9)$$

ここで、 $K_\gamma$  は第 2 種変形ベッセル関数である。

Categorical 分布の共役事前分布は Dirichlet 分布であるため、 $\mathbf{b}$  のハイパーパラメータである  $\pi$  の事前分布は Dirichlet 分布とする。

$$p(\pi_t | a_\pi) = \text{Dirichlet}(\pi_t | a_\pi \mathbf{1}_K) \quad (10)$$

ここで、 $\mathbf{1}_K$  は全ての要素が 1 である  $K$  次元ベクトル、 $a_\pi$  はハイパーパラメータである。

ロボット位置  $r$ 、音源位置  $s$ 、同期時刻ずれ  $\tau$  の事前分布は一様分布とする。通常、一様分布はどの範囲で一様であるかを定める必要があるが、本手法では定める必要がない。なぜなら、一様分布は事後分布を計算するためだけに用いられ、このとき一様分布は  $r, s, \tau$  に対して定数であり、事後分布に影響を与えないためである。

## 2.2 推論

事後分布  $p(G, \lambda, \mathbf{b}, r, s, \tau | \mathbf{x})$  の期待値をロボット位置、音源位置、同期時刻ずれの推定結果とし、音源の割り当て  $\mathbf{b}$  からマスクを生成することで音源分離を行うため、事後分布の期待値を求めることが必要である。期待値を計算するためには積分計算を行うことが必要であるが、提案法の事後分布の積分計算は解析的に解くことができない。そのため、事後分布からのサンプリングを行い、モンテカルロ積分を用いてサンプルから期待値を近似する。事後分布  $p(G, \lambda, \mathbf{b}, r, s, \tau | \mathbf{x})$  から直接サンプリングを行うことは困難であるため、Gibbs サンプリングという手法を用いる。Gibbs サンプリングでは事後分布から直接サンプリングする代わりに、ある潜在変数  $z$  をその潜在変数以外の潜在変数  $\Theta_{-z}$  で条件付けられた事後分布  $p(z | \mathbf{x}, \Theta_{-z})$  からサンプリングする操作を全ての潜在変数に対して繰り返し行うことで、事後分布に従うサンプル列を得る手法である。 $p(z | \mathbf{x}, \Theta_{-z})$  は  $z$  に共役事前分布が存在する場合、直接サンプリングすることができるが、共役事前分布が存在しない場合には直接サンプリングすることが困難な場合が多い。その場合には Metropolis Hastings 法 (MH) を用いる。MH では、まずサンプリングしたい分布  $p(z | \mathbf{x}, \Theta_{-z})$  とは別のサンプリングが容易な提案分布  $q(z^* | z)$  からサンプリングを行う。次に、新しいサンプル  $z^*$  を受理するかどうかを決定する受理率  $\alpha(z^*, z)$  を以下の式で計算する。

$$\alpha(z^*, z) = \min \left( \frac{p(z^* | \mathbf{x}, \Theta_{-z}) q(z | z^*)}{p(z | \mathbf{x}, \Theta_{-z}) q(z^* | z)}, 1 \right) \quad (11)$$

0 から 1 の範囲の一様分布から値を一つサンプリングし、その値が受理率より小さい場合にはそのサンプルを受理して、受理されなかった場合には 1 つ前のサンプルを新たなサンプルとする。

### 2.2.1 空間相関行列 $G$ ・音源パワー $\lambda$

空間相関行列  $G$ 、音源パワー  $\lambda$  については式 (5, 6) の共役事前分布を置くことができるため、条件付き事後分布からのサンプリングを行うことができる。 $G, \lambda$  の事後分布は尤度 (式 (4)) と事前分布 (式 (5, 6)) より次式で表される。

$$p(G_{fk} | \mathbf{x}, \Theta_{-G_{fk}}) = \mathcal{W}_C(G_{fk} | \nu', G'_{fk}) \quad (12)$$

$$p(\lambda_{tfk} | \mathbf{x}, \Theta_{-\lambda_{tfk}}) = \text{GIG}(\lambda_{tfk} | \alpha_{tfk}, \beta_{tfk}, \gamma_{tfk}) \quad (13)$$

パラメータ  $\nu', G'_{fk}, \alpha_{tfk}, \beta_{tfk}, \gamma_{tfk}$  は以下の式で表される。

$$\nu' = \nu + \sum_t b_{tfk} \quad (14)$$

$$G'_{fk} = \nu(G_{fk}^0)^{-1} + \sum_t \frac{\mathbf{x}_{tf} \mathbf{x}_{tf}^H}{\lambda_{tfk}} b_{tfk} \quad (15)$$

$$\alpha_{tfk} = \alpha - M b_{tfk} \quad (16)$$

$$\beta_{tfk} = \beta \quad (17)$$

$$\gamma_{tfk} = \mathbf{x}_{tf}^H G_{fk} \mathbf{x}_{tf} b_{tfk} \quad (18)$$

$G_{f,k}^{\prime-1}$  の第 2 項の  $\mathbf{x}_{t,f} \mathbf{x}_{t,f}^H / \lambda_{t,f,k}$  は観測音から計算された空間相関行列である。従って、 $G_{f,k}$  の事前分布は  $\nu$  個の  $(G_{f,k}^0)^{-1}$  を観測したと等価であり、尤度は  $\sum_t b_{t,f,k}$  個の空間相関行列を観測したとみなすことができる。そのため、 $\nu$  は事後分布に与える事前分布の影響の大きさを表す。

### 2.2.2 音源割り当て $b$

音源割り当て  $b$  については式 (7) の共役事前分布を置くことができるため、条件付き事後分布からのサンプリングを行うことができる。 $\pi$  は値を直接求める必要がないため、積分消去してサンプリングは行わない。また、 $G$  を積分消去することにより全ての  $G$  を考慮することができ、現在の  $G$  に引きずられて正しく  $b$  が更新できず、 $b$  に引きずられて  $G$  もうまく更新できないということが起こりにくくなることが期待される。 $\pi$  と  $G$  を積分消去した事後確率は以下の式で表される。

$$p(b_{t,f} | \mathbf{x}, \lambda, b_{-t,f}) \propto (a_\pi + \sum_{f' \neq f} b_{t,f',k'}) \prod_{m=0}^M (\nu + \sum_{t' \neq t} b_{t',f',k'} - m) \prod_k \left| \left( \nu (G_{f,k}^0)^{-1} + \sum_{t'} \frac{\mathbf{x}_{t',f} \mathbf{x}_{t',f}^H}{\lambda_{t',f,k}} b_{t',f,k} \right)^{-1} \right|^{\nu + \sum_{t'} b_{t',f,k}} \quad (19)$$

ただし、 $k'$  は  $b_{t,f,k'} = 1$  を満たす。

### 2.2.3 ロボット位置 $r$ ・音源位置 $s$ ・同期時刻ずれ $\tau$

ロボット位置  $r$ 、音源位置  $s$ 、同期時刻ずれ  $\tau$  は、事前分布として一様分布を用いているため、事後分布は尤度に一致する。 $r, s, \tau$  は尤度関数  $p(G | r, s, \tau)$  の中で  $G_{f,k}^0$  の中でのみ現れ、 $G_{f,k}^0$  は  $r, s, \tau$  の非線形な変換となっているため、事後分布から直接サンプリングを行うことが出来ない。そこで、 $w \in \{r, s, \tau\}$  は現在の値を平均、 $\Sigma_w \in \{\Sigma_r, \Sigma_s, \Sigma_\tau\}$  を分散共分散行列に持つ Gauss 分布を提案分布としてサンプリングを行い、Metropolis 法により受率率を計算し受率率を決定する。

$r, s, \tau$  の尤度として  $p(G_{f,k} | r, s, \tau) = \mathcal{W}_C(G_{f,k} | \nu, \frac{1}{\nu} G_{f,k}^0)$  を用いることが考えられるが、その場合には受率率が非常に小さくなってしまい、全く受率されないという問題が起きてしまう。この問題を解決するための方法として、 $G$  を積分消去した尤度を用いる方法がある。この尤度は以下のように計算される。

$$p(\mathbf{x} | r, s, \tau) = \prod_{f,k} \frac{|(G_{f,k}^0)^{-1}|^\nu}{|(G_{f,k}^0)^{-1} + \sum_t \frac{\mathbf{x}_{t,f} \mathbf{x}_{t,f}^H}{\lambda_{t,f,k}} b_{t,f,k}|^{\nu + \sum_t b_{t,f,k}}} \quad (20)$$

同期時刻ずれについてはグリッドサーチを行うことも有効である。ハードウェア的にある程度の精度でロボット間での同期ができると考えられるため、同期時刻ずれは有限の区間内を調べればよく、さらに各ロボットの同期時刻ずれは 1 次元の値であるためグリッドサーチしても少ない計算量で済むためである。図 3 に、ロボットと音源を正解位置に固定し、同期時刻ずれを変化させた時の各値での  $G$  を積分消去した事後確率 (式 20) を示す。 $x$  軸は同期時刻ずれの正解との誤差を表しており、原点が正解の同期時刻ずれを表している。正解で鋭く尖っており、正解の周辺では少し谷になっており、局所解が存在することが分かる。MH では山登り法のようにサンプリングしていくため、正解周辺の谷を超えることが難しく、局所解に陥りやすい。グリッドサーチで正解周辺を調べることが出来れば、周り比べて事後確率が非常に高いため、正解に辿り着くことができると考えられる。

## 3 評価実験

提案法を評価するためにロボット 3 台、音源 4 つを用いたシミュレーションでの実験を行った。

### 3.1 実験設定

3ch マイクロホンアレイを搭載した 3 台のロボットと 4 つの音源 (図 3) を用いて、以下の 6 パターンについて実験を行った。

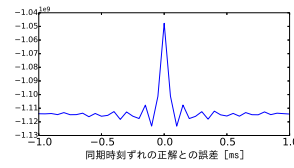


Fig.2 The posterior distribution when only time offset is changed

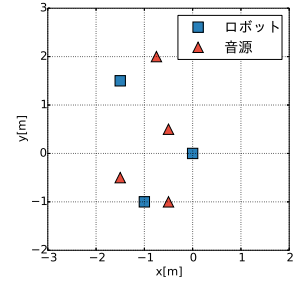


Fig.3 The layout of robots and sound sources

1. パターン 1-3: ロボット位置・音源位置・同期時刻ずれのうち 2 つを正解に固定し、残り 1 つのみを推定する。初期値は正解に  $[-l, l]$  ( $l \in \{0.4, 0.8\}$ ) の一様ノイズを加えた値とした。更新回数は 1 0 0 0 回とした。
2. パターン 4-6: ロボット位置・音源位置・同期時刻ずれのうち 1 つを正解に固定し、残り 2 つを推定する。初期値は正解に  $[-l, l]$  ( $l \in \{0.2, 0.4\}$ ) の一様ノイズを加えた値とした。更新回数は 2 0 0 0 回とした。

ロボットの方向はロボットに搭載された地磁気センサなどを用いて推定できると仮定し、正解を与え推定しないこととした。音源信号は ATR 音素バランス文 [8] からランダムに 4 つを選んだ。観測音はマイクロホンの位置と音源位置から幾何的に計算されたインパルス応答を音源信号に畳み込んで作成した。同期時刻ずれについては、更新を 1 0 0 回行うごとに、 $[-10\text{ms}, 10\text{ms}]$  の範囲を 0.1ms で区切ってグリッドサーチを行った。空間相関行列を  $r, s, \tau$  から幾何的に計算する際のステアリングベクトルの計算において、式 (2) の距離減衰の項  $1/l_{kn}$  を除いて計算した。これは、距離減衰の項を入れると音源位置が発散してしまう問題が起きてしまうためである。 $G$  の初期値は  $r, s, \tau$  の初期値を用いて幾何的に計算した空間相関行列の値をそのまま用いた。 $\lambda_{t,f,k}$  の初期値は観測音  $\mathbf{x}_{t,f}$  のパワーの平均を用いた。 $b$  の初期値は  $G$  の初期値を用いて事後確率からサンプリングした値を用いた。パラメータは  $\alpha = \beta = a_\pi = 1$ ,  $\nu = M + 1$ ,  $\Sigma_r = \text{diag}[1e-1, 1e-1]$ ,  $\Sigma_s = \text{diag}[1e-1, 1e-1]$ ,  $\Sigma_\tau = \text{diag}[1e-6, 1e-6]$  とした。ロボット位置・音源位置・同期時刻ずれの正解との誤差の平均と、分離音の Signal-to-distortion ratio (SDR) [9, 10] を評価した。

### 3.2 実験結果

図 4 にパターン 1-3 の実験結果を示す。図 4(a) に同期時刻ずれのみを推定した場合の推定誤差、図 4(d) に各試行での SDR を示す。同期時刻ずれはすべての場合で高い精度で推定できている。このとき SDR は全て約 8dB となっているため、提案法での分離精度の限界は 8dB 程度だと考えられる。図 4(b) に音源位置のみを推定した場合の推定誤差、図 4(e) に各試行での SDR を示す。8 回中 6 回は高い精度で推定され、このときの SDR はどれも約 8dB となった。同期時刻ずれ推定時と同程度の精度で分離できているため、音源位置は十分な精度で推定されていることが分かる。残りの 2 回では 2 つの音源が同じ場所に収束してしまっている。図 5 に、3 音源を正解に固定して 1 音源を動かした場合の各位置での事後確率を示す。最も事後確率が高くなっている位置が正解音源位置であるが、他の音源と重なる位置でも事後確率は周り比べて非常に高いため、この局所解から抜けることは困難である。この問題を解決するためには音源間の距離に制約を入れ、音源同士が近すぎる場合は片方の音源をランダムに移動させる方法などが考えられる。図 4(c) にロボット位置のみを推定した場合の推定誤差、図 4(f) に各試行での SDR を示す。8 回中 5 回は正しくロボット位置が推定され、高い精度で分離できていることが分かる。音源位置推定と比較してロボット位置推定の方が更新される回数が少なく局所解に陥りやすいのは、音源  $k$  が移動すると  $FK$  個の空間相関行列の内、 $G_{f=0:F,k}$  だけが変化するのに対し、ロボットが移動すると全ての空間相関行列が変化してしまい、より多数の変数について最適な位置でなければ移動するこ

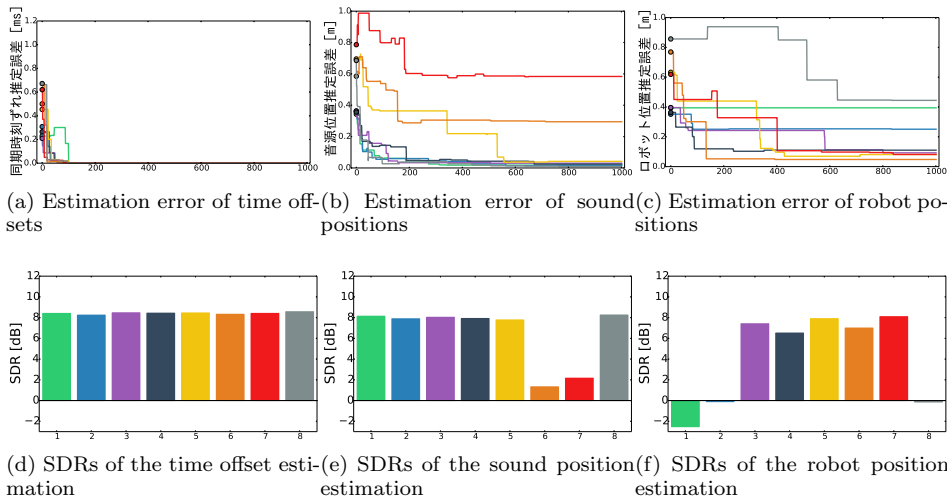


Fig.4 Experimental result

とができないためであると考えられる。

図 6-8 にパターン 4-6 (ロボット位置・音源位置・同期時刻ずれのうち 1 つを正解に固定して、残りの 2 つを推定する場合) の実験結果を示す。1 つのみを推定するとき音源位置と同期時刻ずれは比較的推定精度が高かったため、図 6 の音源位置と同期時刻ずれを同時推定する場合は、他の場合と比較するとある程度正解に近づいており、分離精度も高くなっている。ただし、1 つのみを推定する場合と比較すると分離精度は下がっているため、十分な精度で推定できていないとは言えない。図 7 と 8 を見ると、ロボット位置はあまり更新されておらず局所解に陥っていることが分かる。同期時刻ずれについても数回更新した後では値は変化する可能性があるが、それ以降はほとんど変化しておらず、局所解に陥っていると考えられる。このように複数の同時推定では局所解が増えるため、アニーリング法などを用いて局所解に陥らずにサンプリングすることが必要となる。

#### 4 おわりに

本稿では、複数音源存在下でマイクロホンアレイを搭載した各ロボットの位置、音源位置、ロボット間でのマイクロホンアレイの同期時刻ずれを推定する手法を開発した。従来法では観測音から到達時間差と音源到来方向を抽出し、そこから位置を推定するという 2 段階になっていたが、提案法では階層ベイズモデルを用いてモデル化し、観測音を得られた後の事後分布の期待値を求めることにより、観測音から直接ロボット位置と音源位置、同期時刻ずれを推定し、さらに音源分離も同時に行う。実験では、全ての同時推定はできていないが、ロボット位置・音源位置・同期時刻ずれのうちの 1 つと音源分離を同時に行えることを確かめた。

#### 参考文献

- [1] H. G. Okuno et al. Robot audition: Missing feature theory approach and active audition. In *Robotics Research*, volume 70, pages 227–244. Springer, 2011.
- [2] I. Lee et al. Fast fixed-point independent vector analysis algorithms for convolutive blind source separation. *J. Signal Processing*, 87(8):1859–1871, 2007.
- [3] R. Schmidt et al. Multiple emitter location and signal parameter estimation. *IEEE Trans. Antennas and Propagation*, 34(3):276–280, 1986.
- [4] T. Nakashima et al. *Natural Interaction with Robots, Knowbots and Smartphones*, chapter Integration of Multiple Sound Source Localization Results for Speaker Identification in Multiparty Dialogue System, pages 153–165. Springer, 2014.
- [5] E. Martinson et al. Optimizing a reconfigurable robotic microphone array. In *IEEE/RSJ IROS*, pages 125–130, 2011.
- [6] K. Sekiguchi et al. Optimizing the layout of multiple mobile robots for cooperative sound source separation. In *IEEE/RSJ IROS*, pages 5548–5554, 2015.

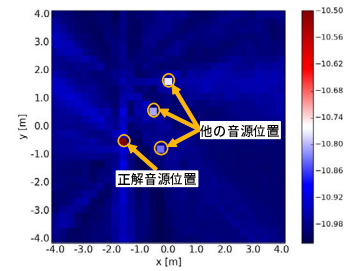


Fig.5 The posterior probability when we moved only one sound source

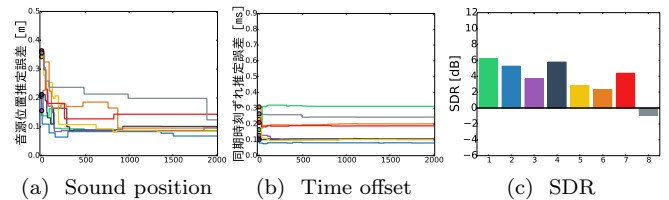


Fig.6 Estimation of the sound positions and time offsets

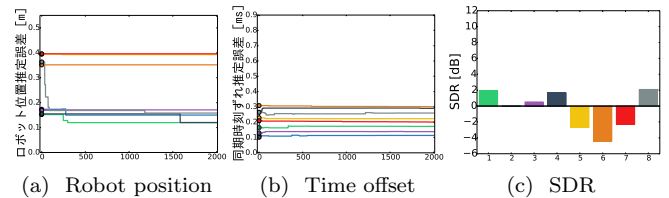


Fig.7 Estimation of the robot positions and time offsets

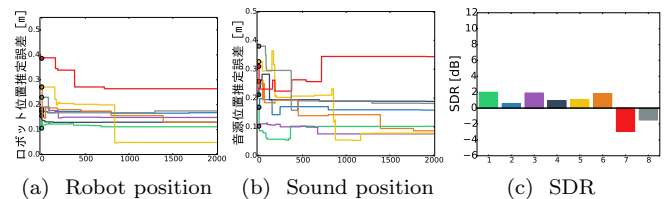


Fig.8 Estimation of the robot and sound positions

- [7] K. Sekiguchi et al. Online simultaneous localization and mapping of multiple sound sources and asynchronous microphone arrays. In *IEEE/RSJ IROS*, pages 1973–1979, 2016.
- [8] Y. Sagisaka and N. Uratani. ATR spoken language database. *J. The Acoustic Society of Japan*, 48(12):878–882, 1992.
- [9] C. Raffel et al. mir-eval: A transparent implementation of common MIR metrics. In *ISMIR*, pages 367–372, 2014.
- [10] E. Vincent et al. Performance measurement in blind audio source separation. *IEEE Trans. Audio, Speech and Language Processing*, 14(4):1462–1469, 2006.