

SLAMに基づく非同期分散マイクロホンアレイのキャリブレーションの評価

Evaluation of a SLAM-based Calibration Method for Asynchronous Microphone Arrays

三浦弘樹[†] 吉田尚水[†] 中村佳佑[‡] 中臺一博^{†,‡}

Hiroki MIURA, Takami YOSHIDA, Keisuke NAKAMURA, Kazuhiro NAKADAI

[†]東京工業大学大学院 情報理工学研究科

[‡](株)ホンダ・リサーチ・インスティテュート・ジャパン

Abstract

This paper evaluates an online calibration method for asynchronous microphone arrays. Conventional microphone array techniques require a lot of measurements of transfer functions to calibrate microphone locations, and a multi-channel A/D converter for inter-microphone synchronization. To solve these two problems, we proposed an online framework combining Simultaneous Localization and Mapping (SLAM) and beamforming and an implemented prototype system using an Extended Kalman Filter (EKF) showed the feasibility of the proposed framework in a simulated and a real environment. In this paper, we show the robustness of the proposed framework for different motion models, motion and observation errors to apply to real microphone array systems through numerical experiments.

1 はじめに

マイクロホンアレイ処理はロボット聴覚分野における音源定位や音源分離に有用であり、数多くの研究が報告されている [1, 2, 3, 4]。これらのマイクロホンアレイ処理には、各マイクロホンの位置もしくは音源とマイクロホンアレイ間の伝達関数が既知であること、全チャンネルを同期収録することが必要とされる。我々は、これらの問題をオンラインで解くため、拡張カルマンフィルタ (Extended Kalman Filter, EKF) に基づく Simultaneous Localization and Mapping (SLAM) と遅延和ビームフォーミングを組み合わせた手法を提案し、マイクロホンアレイの周りを人 (音源) が十数回手を叩きながら歩くだけで非同期マイクロホンアレ

イのキャリブレーションが可能であることを実験により示した [5]。

しかし、提案手法の評価は限定された環境で行われており、システムのパラメータがキャリブレーション性能にどのように影響するのかといった評価はされていなかった。

本稿では、提案手法の適用範囲を知り、実用化に向けた課題を明らかにするために、マイクロホンの初期配置、運動誤差、観測誤差に対するキャリブレーション性能の頑健性を評価する。状態遷移モデルには [5] で用いた長方形軌道に加え円運動を、各マイクロホンの初期配置には一様分布の場合と平均が真のマイク位置、標準偏差が (0.1, 0.5, 1.0) に従う乱数の場合を、運動誤差と観測誤差には実測値から求めた標準偏差、その 10 倍およびその 100 倍の場合をそれぞれ考慮し、その精度と収束速度を数値実験により評価した。

2 非同期分散マイクロホンアレイと問題の定式化

本稿では、非同期分散マイクロホンアレイを、各マイクロホンの位置が未知であり、各マイクロホンの時刻にずれがある (非同期) マイクロホンアレイとして定義する。非同期分散マイクロホンアレイを用いて、音源定位や音源分離といったマイクロホンアレイ処理が可能になれば、煩わしい伝達関数の計測作業や高価な多チャンネル同期 A/D デバイスが不要になり、より実用的な処理が実現できる。この非同期分散マイクロホンアレイを用いて、音源の位置、各マイクロホンの位置、同期時刻のずれを推定するキャリブレーション問題を Blind Alignment 問題と定義する。Blind Alignment 問題は、従来にも研究報告があり、例えば、Thrun らは、事前に各マイクロホンの位置が未知という条件の下、マイクロホン位置のオンラインキャリブレーションを実際にマイクロホンを用いて報告している [6]。しかし、彼らの手法では、音源位置は既知、マイクロホンは完全に同期されている必要があったまた、Ono

Table 1: Notation

N	マイクロホンの総数
K	発音の総数
c	音速
n	マイクロホンのインデックス
k	発音のインデックス
ω	周波数
l	EKF-SLAM における時間ステップ
$x_s[k], y_s[k], \tau_s[k]$	k 回目に発音した位置と時刻
$\xi_s[k] = [x_s[k], y_s[k], \theta_s[k]]^T$	人の位置と向き
$\xi_{mn} = [x_{mn}, y_{mn}, \tau_{mn}]^T$	マイクロホンの位置と同期時刻ずれ
$\xi_m = [\xi_{m1}, \dots, \xi_{mN}]^T$	マイクロホンの位置
$S_{[k]}(\omega)$	k 回目に発音した音
$X_n[k](\omega)$	マイク n が観測した k 回目の音
$\mathbf{X}_{[k]}(\omega)$	$[X_1[k](\omega), \dots, X_N[k](\omega)]^T$
$\mathbf{A}(\omega)$	音源とマイクロホンの間の伝達関数

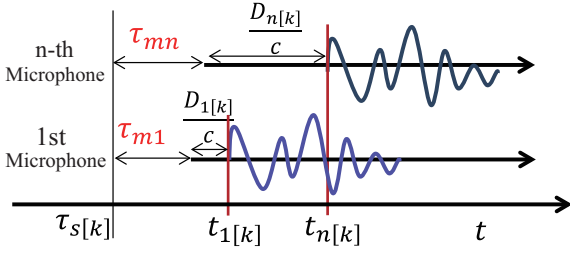


Figure 1: Observation model for each microphone

らは観測した音のみから音源位置、各マイクロホン位置、同期時刻ずれを推定する問題に取り組み、マイクロホンや発音回数など理論的に必要な条件を明らかにした [7]。しかし、彼らの手法はオフラインの手法である、計算量コスト大きい、前もってキャリブレーションの推定回数を指定する必要があるといった問題があった。

これに対して、我々が研究を行っている手法 [5] は、オンラインで Blind Alignment 問題を解決することができる。つまり、位置が未知で、かつ、完全な同期収録が保証できない、非同期分散マイクロホンアレイのオンラインキャリブレーションが可能である。具体的には、SLAM を用いて、その地図推定を各マイクロホンの位置推定、自己位置推定を音源の位置推定に当てはめ、同期時刻のずれを含む推定誤差を最小になるように推定値を更新することによって、オンラインキャリブレーションを行う。

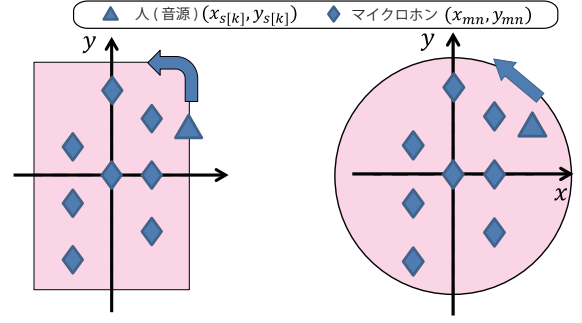
以降、手法の定式化を行う。まず、本稿では、音源はインパルスとして問題を定式化する。なお、本稿で用いる表記を Tab. 1 にまとめた。

2.1 伝達関数（ステアリングベクトル）モデル

マイクロホンで観測される信号は、マイクロホンと音源間の伝達関数を用いて、以下のように表すことができる。

$$\mathbf{X}_{[k]}(\omega) = \mathbf{A}(\omega)S_{[k]}(\omega) \quad (1)$$

この伝達関数 $\mathbf{A}(\omega)$ は、直接音のみを考慮することで、音源位置とマイク位置を用いて以下のように近似計算する



a) Rectangular motion b) Circular motion

Figure 2: Motion Models

ことができる。

$$\begin{aligned} \mathbf{A}(\omega) &\approx \mathbf{A}(\xi_s[k], \xi_m, \omega) \\ &= [\exp(-2\pi j\omega t_{1[k]}), \dots, \exp(-2\pi j\omega t_{N[k]})]^T. \quad (2) \end{aligned}$$

ここで、 $t_n[k]$ は、マイクロホン n が k 回目に発せられた音を観測した時刻である。 $t_n[k]$ は、Fig. 1 に示すように、音源が発音した時刻 $\tau_s[k]$ を用いて、以下のように求めることができる ($D_n[k]$ はマイクロホン n と音源間の距離)。

$$t_n[k] = \tau_s[k] + \frac{D_n[k]}{c} + \tau_{mn}, \quad (3)$$

$$D_n[k] = \sqrt{(x_s[k] - x_{mn})^2 + (y_s[k] - y_{mn})^2}. \quad (4)$$

$\mathbf{A}(\xi_s[k], \xi_m, \omega)$ は、音源定位で用いる際はステアリングベクトルとも呼ばれる。従来の音源定位手法 [8] ではこのステアリングベクトルを事前計測する必要があった。しかし、測定には設備が必要で、かつ時間がかかるため、簡単に計測することは難しい。提案手法では、 $\xi_s[k]$ と ξ_m が推定可能であり、 $\mathbf{A}(\xi_s[k], \xi_m, \omega)$ を事前計測なしに得ることができる。

2.2 状態遷移モデル

音源（人）の移動モデルは一般的には次の式で表される。

$$\xi_s[l+1] = g(\xi_s[l], \eta[l]) + \mathbf{w}_s[l] \quad (5)$$

ただし、 $\eta[l]$ は入力を表し、 $\mathbf{w}_s[l]$ は平均 0、分散 $[\sigma_x^2, \sigma_y^2, \sigma_\theta^2]$ の正規分布に従うモデル誤差を表す。ここで、 $g(\xi_s[l], \eta[l])$ は自由に設計できる。本稿では Fig. 2a), b) に示すように長方形軌道を描く長方形運動モデルと円軌道を描く円運動モデルの 2 つを構築する。

なお、各マイクロホンの位置は動かないので、状態遷移モデルは音源（人）のみに対して構築する。

2.2.1 長方形運動モデル

長方形運動モデルは、以下の式で表される。

$$g(\xi_s[l], \eta[l]) = \xi_s[l] + \begin{bmatrix} \sin(\theta_s[l]) & 0 \\ \cos(\theta_s[l]) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v_s[l] \\ u_s[l] \end{bmatrix}. \quad (6)$$

入力 $\eta_{[l]} = [v_{s[l]}, u_{s[l]}]^T$ は音源の移動速度と角速度を表し、角速度 $u_{s[l]}$ は長方形の四隅に音源が到達した時に 90 度回転させ、それ以外のときは直進 (0 度) である。

2.2.2 円運動モデル

円運動モデルは、半径一定の円を目標軌道としており、以下の式で表される。

$$g(\xi_{s[l]}, \eta_{[l]}) = \begin{bmatrix} \cos(\Delta\theta) & -\sin(\Delta\theta) & 0 \\ \sin(\Delta\theta) & \cos(\Delta\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \xi_{s[l]} + \begin{bmatrix} 0 \\ 0 \\ \Delta\theta \end{bmatrix}. \quad (7)$$

入力 $\eta_{[l]} = \Delta\theta$ は音源の 1 歩あたりの移動角度を表す。

2.3 観測モデル

観測は、 k 番目のインパルスの到達時刻 $t_{n[k]}$ である。音を発した時刻 $\tau_{s[k]}$ は未知であるため、基準マイクロホン (マイクロホン 1) での観測時刻との差をとると、観測モデルは、以下のように相対時刻で表すことができる。

$$\zeta_{[k]} = \begin{bmatrix} \frac{D_{2[k]} - D_{1[k]}}{c} + \tau_{m2} - \tau_{m1} \\ \vdots \\ \frac{D_{N[k]} - D_{1[k]}}{c} + \tau_{mN} - \tau_{m1} \end{bmatrix} + \delta_{[k]} \quad (8)$$

観測誤差 $\delta_{[k]}$ は平均 0 分散 σ_r^2 の正規分布に従うものとする。

3 非同期分散マイクロホンアレイのキャリブレーション

提案法は、EKF-SLAM を用い、予測、観測、更新ステップを繰り返すことでキャリブレーションを行う。

予測ステップ 音源状態の平均 $\hat{\xi}_{[l]}$ と分散 $\hat{P}_{[l]}$ は以下のように計算される。

$$\hat{\xi}_{s[l-1]} = g(\hat{\xi}_{s[l-1]}, \eta_{[l-1]}) \quad (9)$$

$$\hat{P}_{[l-1]} = \mathbf{G}_{[l]} \hat{P}_{[l-1]} \mathbf{G}_{[l]}^T + \mathbf{F}^T \mathbf{R} \mathbf{F} \quad (10)$$

$$\mathbf{F} = [\mathbf{I}^{3 \times 3}, \mathbf{O}^{3 \times 3N}] \quad (11)$$

ここで \mathbf{R} は $\mathbf{R} = \text{diag}(\sigma_x^2, \sigma_y^2, \sigma_a^2)$ で表される共分散行列であり、 $\mathbf{G}_{[l]}$ は次式で示される状態遷移モデルのヤコビアンである。

$$\mathbf{G}_{[l]} = \frac{\partial g(\xi_s, \eta)}{\partial \xi_s} \Big|_{\xi_s = \hat{\xi}_{s[l-1]}} \quad (12)$$

$$= \begin{cases} \begin{bmatrix} \mathbf{I} + \mathbf{F}^T & \begin{bmatrix} 0 & 0 & -v_{s[l]} \sin(\theta_{s[l]}) \\ 0 & 0 & v_{s[l]} \cos(\theta_{s[l]}) \\ 0 & 0 & 0 \end{bmatrix} \end{bmatrix} \mathbf{F} \quad (\text{長方形運動モデル}), \\ \begin{bmatrix} \mathbf{I} + \mathbf{F}^T & \begin{bmatrix} \cos(\Delta\theta) & -\sin(\Delta\theta) & 0 \\ \sin(\Delta\theta) & \cos(\Delta\theta) & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{bmatrix} \mathbf{F} \quad (\text{円運動モデル}) \end{cases} \quad (13)$$

観測ステップ 各マイクロホンでの観測時刻から、式 (8) に従って、以下を得る。

$$h(\hat{\xi}_{[k|k-1]}) = \begin{bmatrix} \frac{\hat{D}_{2[k]} - \hat{D}_{1[k]}}{c} + \hat{\tau}_{m2} - \hat{\tau}_{m1} \\ \vdots \\ \frac{\hat{D}_{N[k]} - \hat{D}_{1[k]}}{c} + \hat{\tau}_{mN} - \hat{\tau}_{m1} \end{bmatrix} \quad (14)$$

更新ステップ 予測ステップと観測ステップを元に、音源の位置・向きとマイクロホンの位置・同期時刻のずれの推定値を更新する。まず、 $h(\hat{\xi}_{[k|k-1]})$ と $\zeta_{[k]}$ の差を最小にするようにカルマンゲインを導出する。

$$\mathbf{K}_{[k]} = \mathbf{P}_{[k|k-1]} \mathbf{H}_{[k]}^T \left(\mathbf{H}_{[k]} \mathbf{P}_{[k|k-1]} \mathbf{H}_{[k]}^T + \mathbf{Q}_{[k]} \right)^{-1} \quad (15)$$

ここで、 $\mathbf{H}_{[k]} = \frac{\partial h(\xi)}{\partial \xi} \Big|_{\xi = \hat{\xi}_{[k|k-1]}}$ は観測モデルのヤコビアンであり、 $\mathbf{Q}_{[k]}$ は $\mathbf{Q}_{[k]} = \text{diag}(\sigma_r^2, \dots, \sigma_r^2)$ で定義される共分散行列である。

求めたカルマンゲインを用いて、推定値を以下のように更新する。

$$\hat{\xi}_{[k]} = \hat{\xi}_{[k|k-1]} + \mathbf{K}_{[k]} \left(\zeta_{[k]} - h(\hat{\xi}_{[k|k-1]}) \right), \quad (16)$$

$$\hat{P}_{[k]} = (\mathbf{I} - \mathbf{K}_{[k]} \mathbf{H}_{[k]}) \hat{P}_{[k|k-1]}. \quad (17)$$

4 キャリブレーション性能の評価

ここでは、以下の条件においてキャリブレーションの収束速度と収束後の音源位置・マイク位置・同期時刻のずれの推定精度を評価した。

状態遷移モデル 長方形運動モデル・円運動モデル

各マイクの初期値 一様分布・真値を平均とした正規分布

運動誤差 実測値・その 10 倍・その 100 倍 (σ_x, σ_y)

実測値・その 25 倍・その 100 倍 (σ_θ)

観測誤差 実測値・その 10 倍・その 100 倍

なお、更新ステップによるマイクロホンの推定位置の変化量 $\hat{\xi}_m[k+1] - \hat{\xi}_m[k]$ が平均で 0.01 [m] 以下になったら収束したとみなし、それまでの発音回数を収束速度とした。

各パラメータは、77 回の拍手を収録した実測データ [5] から算出した値を基準にした。

実測データの収録条件 ハードウェアには、(株) システムインフロンティア社製の多チャンネル録音機器 RASP24 と MEMS マイクロホンを用い、8 ch, 24 bits, 16 kHz サンプリングで収録した。観測誤差には A/D コンバータの影響、配線長による影響、マイクと音源位置の計測誤差、そして観測された音の波形からの到達時刻の抽出精度が含まれる。しかし、A/D コンバータと配線長による影響は到達時刻の抽出による誤差に比べ小さいので無視するものとする。

実験時には 1.2 m × 2.4 m の机の上に 8 チャネルマイクロホンアレイを配置し、音源 (拍手) を一定間隔で動かすことによって 77 回録音した。得られた音の波形から到達時刻を各チャンネルごとに抽出した。あらかじめ計測しておいた音源位置、各マイクロホン位置から到達時間差を計算し、マイクロホン 1 を基準とした観測モデルの観測誤差を計算した。Fig.4 はマイクロホンの観測誤差を分布を示すヒストグラムであり、平均 -2.75×10^{-4} [s]、標準偏差

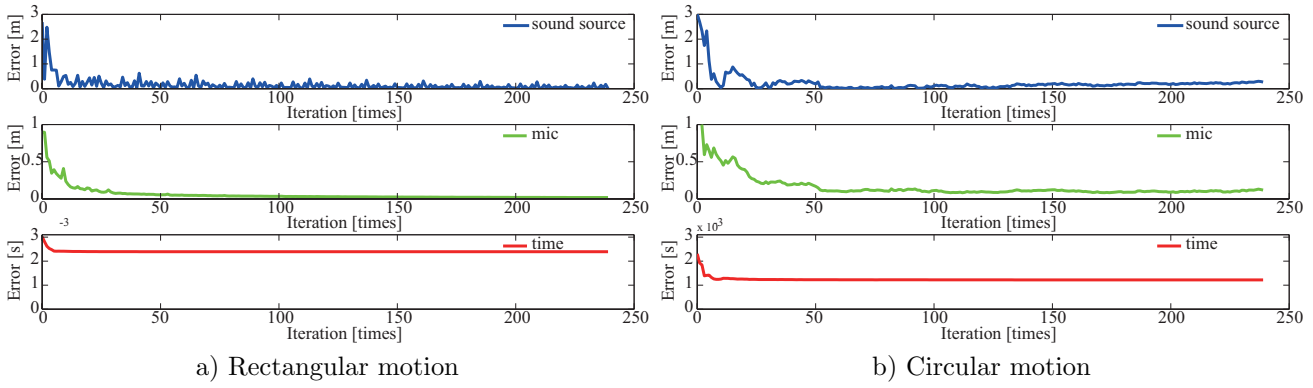


Figure 3: Calibration results

2.1×10^{-3} [s] であった．サンプリング周波数が 16 kHz であるので，観測誤差の標準偏差は 16 サンプル程度である．

マイクロホンをランダムに部屋に配置し，長方形運動モデルと円運動モデルをそれぞれ用い，数値実験により評価する．さらに，実際の観測誤差を計測し，移動モデル誤差に対する頑健性を検証し，キャリブレーション性能を評価する．数値実験では， 1.2 [m] \times 2.4 [m] の部屋（長方形運動モデル），と半径 1.2 [m] の部屋（円運動モデル）を想定して検証を行った．マイクロホン数は 8 であり，図 2 に従い配置した．基準となるマイクロホン 1 の位置を原点とし，回転方向の曖昧性を解消するため，マイクロホン 2 の位置は y 座標を 0， x 座標を正とした．音源はインパルスを想定し，部屋の隅に沿って反時計回りに移動する．初期位置は実際の初期位置である部屋の左下隅座標に対して，平均 0 [m]，標準偏差 0.5 [m] に従うガウス雑音を与えた．音源の移動は，1 歩あたり 0.3 [m] とし，5 歩進むごとに 1 回音を発するものとした．状態遷移モデルの位置と角度の標準偏差はそれぞれ 0.1 [m]， 1 [度]，観測誤差の標準偏差は 0.5×10^{-3} [s] (0.17 [m] に相当) とした．各マイクロホンの時刻のずれは固定であり，初期状態では，ずれは 0 [s]，標準偏差を 0.1 [s] とした．

4.1 状態遷移モデルのキャリブレーション評価

図 3a), b) はそれぞれ長方形運動モデル，円運動モデルを用いたマイクロホンアレイのキャリブレーションの結果を示す．マイクロホン位置の平均誤差と同期時刻のずれの誤差には明確な差は無いが，長方形運動モデルを用いた場合の人の位置の誤差は振動的になっている．これは人が部屋の壁にたどり着いたら直角に曲がるという非線形性の強いモデルを使用していることが原因と考えられる．一方円運動モデルでは長方形運動モデルほど振動的でないことがわかる．

4.2 分散パラメータに対するキャリブレーションの性能
 実測した観測誤差の分散 $\sigma_r^2 = 4.41 \times 10^{-6}$ [s²] を用いマイクロホンアレイのキャリブレーションの性能を評価する．状態遷移モデルは長方形運動モデルとし，誤差は期待値 0 [m]，

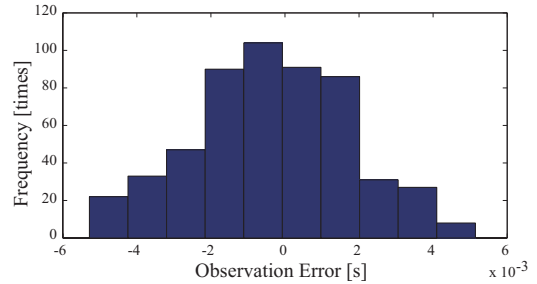


Figure 4: Histogram of Observation Errors

分散 $[\sigma_x^2, \sigma_y^2, \sigma_\theta^2] = [1.0 \times 10^{-2} [m^2], 1.0 \times 10^{-2} [m^2], 1.75 \times 10^{-2} [rad^2]]$ に従うガウス分布とする．マイクロホンの初期位置は一様分布に従い発生させ，数値シミュレーションにより性能検証した．シミュレーションは 100 回を行い，その平均を求めた．

図 5-7 の左図はマイクロホン位置推定が収束したときの手を叩いた回数（総計 100 回）のヒストグラムを表す．横軸が収束するまでにかかったインパルス回数，縦軸が度数（総計 100 回）である．ここで，収束とは変化率が 1.0×10^{-2} [m] を下回った時とする．

また，右図は，マイクロホンの位置推定誤差の平均を折れ線グラフで，最大値，最小値をエラーバーで示したものである．縦軸は各マイクロホン位置推定の平均誤差，横軸はインパルス回数であり 50 回ごとに平均計算を行った．

図 6 は，図 5 で用いた σ_r^2 を 10 倍，100 倍に変化させた場合の結果であり，図 7 は，図 5 の $[\sigma_x^2, \sigma_y^2, \sigma_\theta^2]$ を $[\times 25, \times 25, \times 1]$ (a), b), $[\times 100, \times 100, \times 1]$ (c), d), $[\times 1, \times 1, \times 10]$ (e), f), $[\times 1, \times 1, \times 100]$ (g), h) と変化させた場合の結果である．

図 6, 7 の左図から，収束までに必要なインパルス回数は，観測，状態遷移モデルに関わらず，実際の分散の 10 倍程度までなら，ほとんど変化がないことがわかる．100 倍程度になると，ヒストグラムの形が崩れ，収束までの時間が大きくなることがわかる．つまり，これらの分散は，実際の値の 10 倍程度までの値を設定する必要があると言える．一方，右図からは，インパルス回数に対するマイク

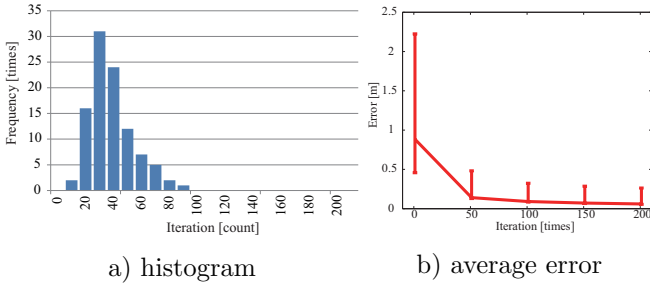
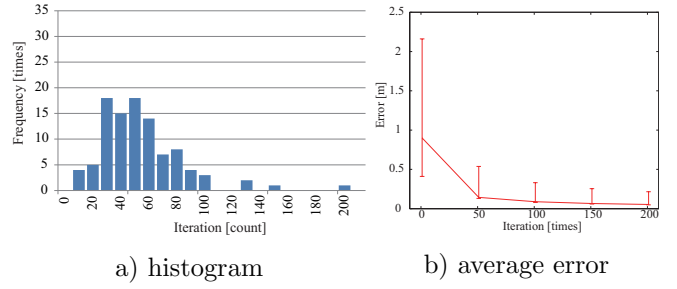
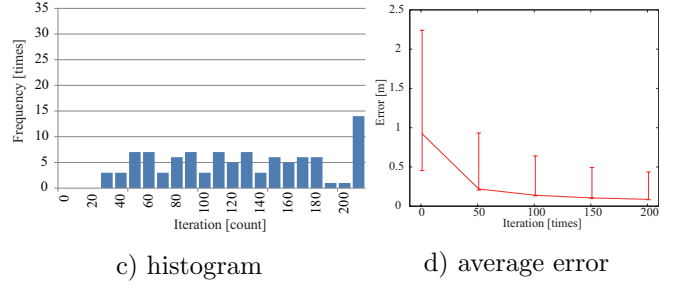
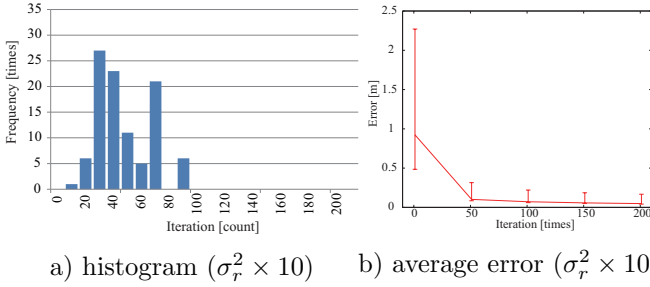


Figure 5: Reference Result



$(\sigma_x^2 \times 25, \sigma_y^2 \times 25, \sigma_\theta^2 \times 1)$ $(\sigma_x^2 \times 25, \sigma_y^2 \times 25, \sigma_\theta^2 \times 1)$



$(\sigma_x^2 \times 100, \sigma_y^2 \times 100, \sigma_\theta^2 \times 1)$ $(\sigma_x^2 \times 100, \sigma_y^2 \times 100, \sigma_\theta^2 \times 1)$

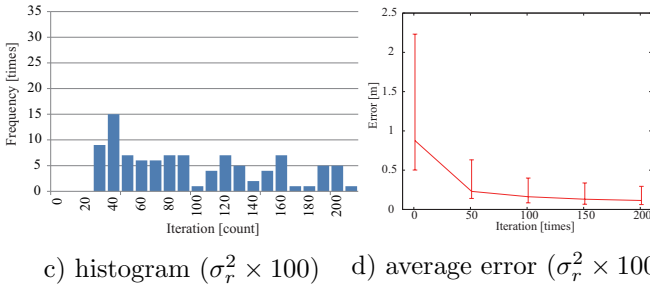


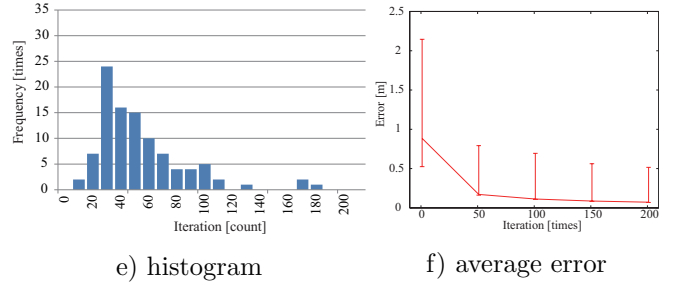
Figure 6: Changes of observation variances

口ホンの位置推定誤差の傾向は、どの場合も同様であるということが言える。50 回程度までは、推定性能が向上するものの、それを越えると推定性能の向上は緩やかになる。例えば、必要な精度を 0.2 [m] とすると、分散の値が 100 倍以上ずれていたとしても、50 回程度のインパルスで収束するということが言える。実際に、文献 [5] では、音源定位で用いるビームフォーミングの解像度が 0.2 [m] を用いていた。従って、これらの図から、観測、状態遷移モデルの誤差分散は、実際の値の 10 倍程度までなら、収束にほとんど影響しないこと、また、100 倍程度であっても、必要な解像度によっては、十分実用に耐えうることが分かった。

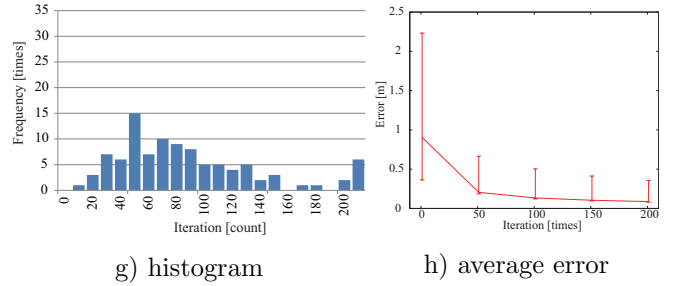
4.3 マイク初期位置に対するキャリブレーションの性能

図 8a)–f) はマイクロホンの初期配置を変化させた時のキャリブレーション性能を示している。左図、右図は、4.2 節と同様に、収束時のインパルス回数のヒストグラムとインパルス回数に対するマイクロホン位置推定の平均誤差を示している。

図 8a),b) は、実際のマイクロホンの位置に対する誤差



$(\sigma_x^2 \times 1, \sigma_y^2 \times 1, \sigma_\theta^2 \times 10)$ $(\sigma_x^2 \times 1, \sigma_y^2 \times 1, \sigma_\theta^2 \times 10)$



$(\sigma_x^2 \times 1, \sigma_y^2 \times 1, \sigma_\theta^2 \times 100)$ $(\sigma_x^2 \times 1, \sigma_y^2 \times 1, \sigma_\theta^2 \times 100)$

Figure 7: Changes of motion variances

の標準偏差 σ_m が 0.1 [m] になるようにマイクロホンの初期位置を設定した場合の結果であり、図 8c),d) は、 $\sigma_m = 0.5$ [m]、図 8e),f) は、 $\sigma_m = 1.0$ [m] の結果である。

σ_m が 0.1 [m] と、初期位置が正解位置に比較的近い場合は、20 回以下のインパルスでほぼ収束することから、提案手法の正当性が示されている。 σ_m が 0.5 [m] の場合であっても、収束までのインパルス回数は増加するものの 30 回程度で、大半が収束していることがわかる。一方、 σ_m が 1.0 [m] と大きくなってしまると、収束にかかるインパルス回数の度数分布はなだらかになり、一概に何回インパ

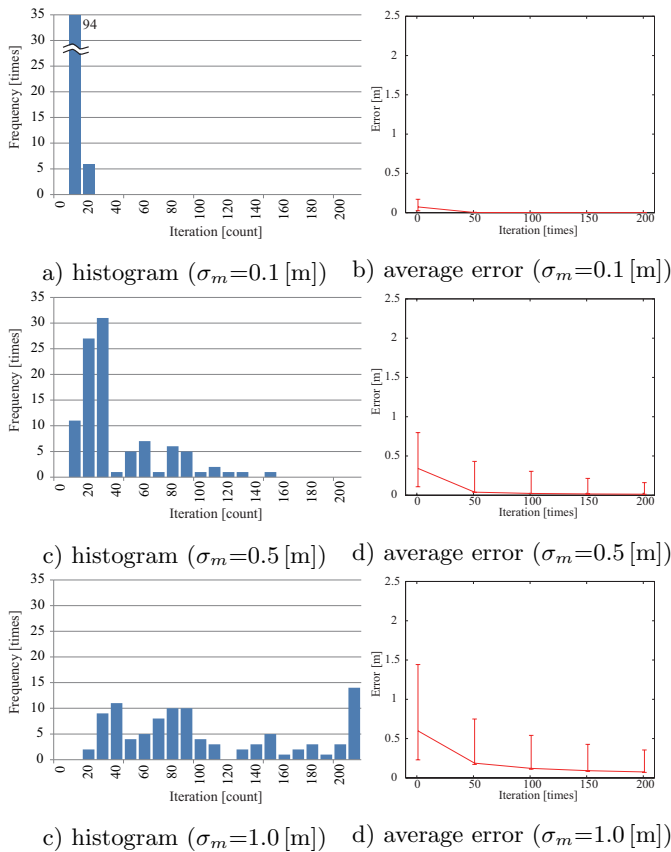


Figure 8: Performance for initial microphone location errors

ルスを出力する必要があるか明確には言えない。

マイクロホンの位置推定誤差については、インパルス回数が増えるに応じて、収束していくこと、また 50 回程度で推定精度の向上が徐々に頭打ちになることがわかる。また、 σ_m が 0.5 [m] 以下の場合には、マイクロホンの位置推定が 50 回程度で正確に行えるのに対して、 σ_m が 1.0 [m] になると、なかなか初期位置の誤差が吸収しきれないことがわかる。

以上のことから、マイクロホンアレイの初期位置はキャリブレーションを行う際に重要なパラメータであり、本稿のケースでは、 σ_m を 0.5 [m] 以下に設定することが望ましいと言える。

5 おわりに

本稿では、非同期分散マイクロホンアレイのオンラインキャリブレーション問題を解決するために提案している EKF-SLAM ベースの手法の評価を行った。状態遷移モデルに長方形運動モデルと円運動モデルを使用し、数値実験で両者を比較した。また、状態遷移モデルの誤差、観測誤差、そしてマイクロホンの初期位置を変化させ、提案手法のロバスト性、適用範囲の評価を行った。結果として、本手法を利用する際には、本稿のマイクロホンアレイ設定条件では、観測モデル、状態遷移モデルの誤差を実際の

誤差の 10 倍程度に抑えるべきであるが、100 倍程度でも場合によっては十分実用に耐えうること、マイクロホンの初期位置は、実際の位置に対して標準偏差が 0.5 [m] 以下であれば、高精度なキャリブレーションが可能であることが示された。今後はカルマンフィルタの理論的解析をし、実環境で評価する必要がある。さらに、非線形な状態遷移モデルに対し、よりロバストな UKF やパーティクルフィルタの適用を試みる予定である。

謝辞

本研究の一部は科研費若手研究 (B)(22700165)、科研費 (S)(19100003)、新学術領域研究 (22118502)、特別研究員奨励費の補助を受けた。

参考文献

- [1] J.-M. Valin, J. Rouat, and F. Michaud, “Enhanced robot audition based on microphone array source separation with post-filter,” in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004)*. IEEE, 2004, pp. 2123–2128.
- [2] F. Asano, H. Asoh, and T. Matsui, “Sound source localization and signal separation for office robot “Jijo-2”,” in *Proc. of IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI-99)*, 1999, pp. 243–248.
- [3] S. Yamamoto, J.-M. Valin, K. Nakadai, T. Ogata, and H. G. Okuno, “Enhanced robot speech recognition based on microphone array source separation and missing feature theory,” in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005)*. IEEE, 2005, pp. 1489–1494.
- [4] H. Saruwatari, Y. Mori, T. Takatani, S. Ukai, K. Shikano, T. Hiekata, and T. Morita, “Two-stage blind source separation based on ICA and binary masking for real-time robot audition system,” in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2005)*. IEEE, 2005, pp. 209–214.
- [5] H. Miura, T. Yoshida, K. Nakamura, and K. Nakadai, “SLAM-based online calibration of asynchronous microphone array for robot audition,” in *Proceedings of 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2011)*, IEEE, 2011, pp. 524–529.
- [6] S. Thrun, “Affine structure from sound,” *Advances in Neural Information Processing Systems*, vol. 18, pp. 1353–1360, 2006.
- [7] N. Ono, H. Kohno, N. Ito, S. Sagayama, “Blind alignment of asynchronously recorded signals for distributed microphone array,” in *2009 IEEE Workshop Applications of Signal Processing to Audio and Acoustics*, IEEE, 2009, pp. 161–164.
- [8] K. Nakamura, K. Nakadai, F. Asano, Y. Hasegawa, and H. Tsujino, “Intelligent sound source localization for dynamic environments,” in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2009)*, IEEE, 2009, pp. 664–669.