

## 複数マイクロホンアレイを用いた尤度分布統合による移動音源追跡

山田 泰基<sup>1</sup>, 糸山 克寿<sup>1</sup>, 西田 健次<sup>1</sup>, 中臺一博<sup>1,2</sup>

1 東京工業大学 工学院 システム制御系 2 (株) ホンダ・リサーチ・インスティテュート・ジャパン

## 1 はじめに

音環境理解の分野において、音源位置は音源分離や音源識別に有用な情報である [1]. 特に、複数のマイクロホンからなるマイクロホンアレイを用いることによる、音源の位置推定および追跡は盛んに研究されている [2, 3]. 例えば、複数マイクロホンアレイを用いて音源方向を推定し、推定方向より三角測量することで音源追跡を行う研究が報告されている. しかし、三角測量ベースの音源追跡は遠方音源を追跡する際、微小な方向推定誤差が、位置推定誤差として大きく現れる問題がある. そこで、本稿では音源位置ではなく、三次元空間上の音源位置尤度の分布を推定することで、位置推定誤差が現れやすい方向に対する誤差の抑制が可能な音源追跡を目指す.

## 2 提案手法

本稿では、複数マイクロホンアレイを用いて、単独移動音源の追跡を行うことについて考える. マイクロホンアレイは全てで  $M (\geq 2)$  個存在し、それぞれの位置と姿勢は既知であるとする. 移動音源は 3 次元空間上を制約なく移動しているとし、 $T$  秒ごとにその位置を推定することを音源追跡とする.

提案手法のフローを Algorithm 1 に示す. 本手法は複数マイクロホンアレイが収録した音響信号から、音源位置尤度を求め、求めた尤度をパーティクルフィルタの各パーティクルフィルタに与えることで、音源位置分布を反映した音源追跡を可能にする.

ある位置  $\mathbf{x}$  における音源位置尤度を [4] で定義された音源位置尤度を参考に定義する. Rieken らは各マイクロホンアレイから得られる MUSIC スペクトルの総和を音源位置尤度と定義している. MUSIC スペクトルとは、MUltiple SIgnal Classification (MUSIC) 法 [5] による音源方向推定に用いられる空間スペクトルである. MUSIC スペクトルは音源方向  $(\phi, \theta)$  を変数に持つ関数  $P(\phi, \theta)$  と表すことができ、理論上は音源方向対

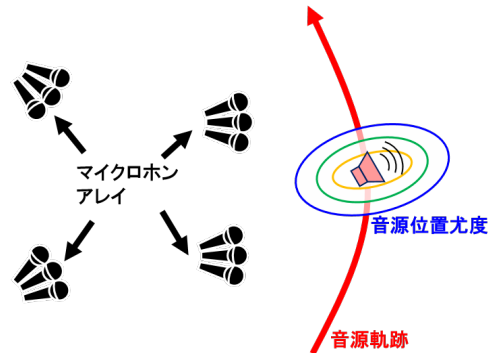


Fig. 1 Image of sound source tracking with microphone arrays by spatial likelihood estimation

して局所的最大値を取る. よって、音源位置尤度  $L(\mathbf{x})$  は以下のように書き表せる.

$$L(\mathbf{x}) = \sum_{m=1}^M P_m(\phi_m, \theta_m) \quad (1)$$

$$\text{ただし, } \begin{bmatrix} \cos \phi_m \cos \theta_m \\ \sin \phi_m \cos \theta_m \\ \sin \theta_m \end{bmatrix} = \mathbf{R}_{\text{rot},m}(\mathbf{x} - \mathbf{x}_m) \quad (2)$$

ここで、 $P_m$  はマイクロホンアレイ  $m$  で求めた MUSIC スペクトル、 $\mathbf{x}_m \in \mathbb{R}^3$  はマイクロホンアレイ  $m$  の三次元位置、 $\mathbf{R}_{\text{rot},m} \in \mathbb{R}^{3 \times 3}$  はマイクロホンアレイ  $m$  の姿勢を表す回転行列である. つまり、 $(\phi_m, \theta_m)$  はマイクロホンアレイから見た位置  $\mathbf{x}$  の方向である. 一般的に、MUSIC スペクトルは用意された伝達関数上で計算されることが多く、その伝達関数の分解能によって、MUSIC スペクトルを導出できる方向が決定される. よって、伝達関数が数度刻みに定義されている場合、 $(\phi_m, \theta_m)$  もその刻み値に丸める必要がある.

式 (1) より、任意の位置の音源位置尤度  $L(\mathbf{x})$  を得ることができる. 従来の音源位置推定では、 $L(\mathbf{x})$  を局所最大化するような位置  $\mathbf{x}$  を勾配法など、反復的な計算によって求める. しかし、音源追跡をする場合、短時間の入力信号を元に、音源位置尤度を求める必要があるため、雑音によって局所最大値を取るような位置が真の音源位置でない場所にも現れる可能性がある. そこで、本手法ではパーティクルフィルタを用いることで、不安定な音源位置尤度にロバストな音源追跡を行う. 従来のパーティクルフィルタを用いた音源追跡では、予測後のパーティクルの尤度は、観測モデルを用いて推定された

3D sound source tracking by spatial likelihood integration using microphone arrays

Taiki Yamada<sup>1</sup>, Katsutoshi Itoyama<sup>1</sup>, Kenji Nishida<sup>1</sup>, Kazuhiro Nakadai<sup>1,2</sup>

1 Dept. of Systems and Control Engineering, School of Engineering, Tokyo Institute of Technology

2 Honda Research Institute Japan Co., Ltd.

**Algorithm 1** Particle Filter (For time  $t$ )

**Input:** 入力信号  $z_t$ , 前時刻のパーティクル集合  $\mathcal{X}_{t-1}$   
 1:  $\bar{\mathcal{X}}_t \leftarrow$  音源モデルから予測したパーティクル集合  
 2: 各パーティクルの位置と入力信号  $z_t$  から  $L(\mathbf{x})$  を求め, 重みを付与 (式 1 参照)  
 3:  $\mathcal{X}_t \leftarrow$  パーティクル集合  $\bar{\mathcal{X}}_t$  において復元抽出を行って得られる集合  
 4:  $\hat{\mathbf{x}}_{e,t} \leftarrow$  パーティクルの重み付き平均  
**Output:** 現時刻のパーティクル集合  $\mathcal{X}_t$ , 現時刻の推定音源状態  $\hat{\mathbf{x}}_{e,t}$

音源方向や到達時間差から導出するが, 本手法ではパーティクルの位置を式 (1) に代入することでパーティクルの尤度を得る. これによって, パーティクルフィルタは音源位置尤度の分布を反映しながら, 前時刻の推定を加味した音源追跡を行うことができる. パーティクルは音源位置推定の誤差が生じやすい方向に分散するため, それらパーティクルの重み付き平均を取ることによって, その方向の誤差を抑えられることが期待できる.

**3 実験・考察**

提案法の有効性を検証するため, 数値シミュレーションを行った. Fig. 2 の赤線のように, 半径 40 m の円運動を行う点音源の追跡について考える. 点音源の高さは  $z = 1$  m である. マイクロホンアレイは表 1 の位置に静止している. マイクロホンアレイは 24 bit, 44.1 kHz で 8 チャンネルであり, マイクロホンの配置は文献 [6] の通りである. 点音源は日本音響学会 新聞記事読み上げ音声コーパス (JNAS) 内のニュース文読み上げ音声を 10 秒間出力させた. また, 三角測量を用いる手法との比較のため, 文献 [3] の手法も実装した. 両手法とも, 方位角・仰角が 5 度刻みの伝達関数を元に MUSIC スペクトルを計算し, 0.1 秒周期で音源位置を追跡した.

Fig. 2 は提案手法と比較手法の音源追跡結果である. Fig. 2 より, 両手法共に真の音源軌跡周辺を推定できていることが分かる. 提案手法の RMSE は 1.67 m, 文献 [3] の手法は 1.87 m であった. また, 同様の実験を, 音源が半径 50 m の円運動を行った場合について行ったところ, 提案手法の RMSE は 2.15 m, 文献 [3] の手法は 2.09 m であり, 両手法の推定誤差に有意な差は見られなかった. これより, 遠方にある音源に対しては, 三角測量を用いた追跡と位置尤度を用いた追跡の性能は同程度であることが考えられる. しかし, 今回の実験では, マイク入力にノイズを加えていないため, 音源方向推定の誤差が大きく現れたり, MUSIC スペクトルが実際の音源方向でない方向でピークを取ったりすることがなかった. 両手法の対ノイズにおける性能差の検証は今後の課題であると言える.

Table 1 State of microphone arrays in absolute coordinates.

	$x$ [m]	$y$ [m]	$z$ [m]	方位角 [rad]
アレイ 1	5	5	5	0
アレイ 2	-5	5	5	$7\pi/4$
アレイ 3	-5	-5	5	$\pi/4$
アレイ 4	5	-5	5	0

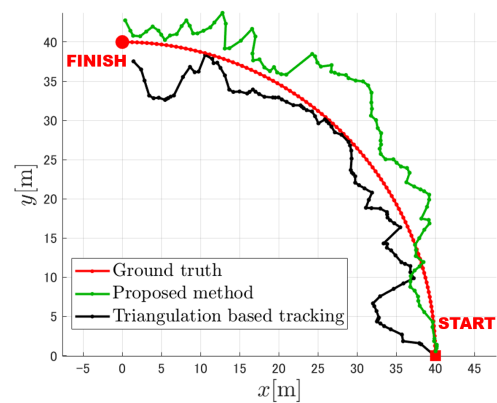


Fig. 2 Topview of estimation results (Radius of sound source trajectory is 40 m)

**4 おわりに**

本稿は, 複数マイクロホンアレイを用いて, 音源位置尤度分布を統合することで, 音源追跡を行う手法を提案した. 数値シミュレーションにより, 40 m 遠い音源に対して, RMSE 1.67 m と, 以前提案した三角測量ベースの音源追跡手法と同等の十分な追跡性能を発揮した. 本稿では, ノイズを考慮していない状況下での性能評価を行ったため, ノイズによる性能劣化の評価が今後の課題である.

謝辞 本研究は JSPS 科研費 16H02884, 17K00365 および 19K12017 の助成を受けた.

**参考文献**

[1] O. Sugiyama, *et al.*, “Outdoor acoustic event identification with DNN using a quadrotor-embedded microphone array.”, *JRM*, 29(1), pp.188–197, 2017.  
 [2] M. S. Brandstein and H. F. Silverman, “A practical methodology for speech source localization with microphone arrays.”, *Computer Speech & Language*, 11(2), pp.91–126, 1997.  
 [3] 山田他, “複数の移動マイクロホンアレイによる移動音源の 3 次元定位”, *IPSSJ2019*  
 [4] D. W. Rieken and D. R. Fuhrmann, “Generalizing MUSIC and MVDR for multiple noncoherent arrays.”, *IEEE Trans. on signal processing*, 52(9), pp.2396–2406, 2004  
 [5] R. Schmidt, “Multiple emitter location and signal parameter estimation.”, *IEEE Trans. on Antennas and Propagation*, 34(3), pp.276–280. 1986.  
 [6] M. Strauss, *et al.*, “DREGON: Dataset and Methods for UAV-Embedded Sound Source Localization.”, *IROS2018*,