

複数マイクロホンアレイにおける音源方向尤度に基づく三次元音源追跡

Sound Source Tracking Based on Source Direction Likelihood from Multiple Microphone Arrays

山田 泰基^{1*} 糸山 克寿¹ 西田 健次¹ 中臺 一博^{1,2}

Taiki Yamada¹, Katsutoshi Itoyama¹, Kenji Nishida¹, Kazuhiro Nakadai^{1,2}

¹ 東京工業大学

¹ Tokyo Institute of Technology

² (株) ホンダ・リサーチ・インスティテュート・ジャパン

² Honda Research Institute Japan Co.,Ltd.

Abstract: 本稿は、複数のマイクロホンアレイを用いた三次元音源位置の推定および追跡手法を提案する。マイクロホンアレイによる信号処理は、音源方向推定に用いられる技術として確立されている。しかし、マイクロホンアレイをロボットやドローンに搭載して災害救助する際など、音源方向に併せて音源との距離の取得、言い換えれば音源位置の取得が必要になる場合が存在する。そこで、本稿では各マイクロホンアレイが算出した音源方向尤度を統合することで得られる音源位置尤度を定義し、求めた音源位置尤度から音源位置の推定および追跡を行う。また、数値シミュレーションを通じて提案手法の有効性を検証し、三角測量的アプローチで音源位置を追跡する他手法との性能比較を行った。提案推定手法は 5° 刻みの伝達関数を用いたとき、20 m 離れた音源に対し、平均推定誤差 2.45 m と、三角測量を用いた音源追跡手法の平均推定誤差より 3.25 m 小さく追跡することができることが確認された。

1 はじめに

音環境理解の分野において、マイクロホンアレイを用いた音源定位は盛んに研究が行われており、例えば、災害現場で瓦礫に埋もれた要救助者探索などに応用が期待できる有用な技術である。一般に、単独マイクロホンアレイは音源方向の推定に用いられるが、単独で用いる代わりに複数マイクロホンアレイを用いることで音源位置を推定する研究が盛んに行われている。複数マイクロホンアレイを用いて音源を観測することで、音源を多角的に観測することができ、音源方向以上の情報が得られる上、移動音源をリアルタイムで追跡することができるというメリットがある。複数マイクロホンアレイで音源位置を推定する手法として、複数マイクロホンアレイから得た音源方向を元に三角測量する Brandstein らの手法 [1] や Daniel らの手法 [2] が挙げられる。三角測量的アプローチは音源が遠方にあるほど、外れ値となるような三角測量点が現れ、結果的に位置推定誤差が大きく現れてしまう問題がある。そこで、

山田らはこれまで、三角測量を用いた音源位置推定に加え、Gaussian Sum Filter を用いて外れ値となるような三角測量点を無視するような音源追跡手法を提案し、数値シミュレーションにおいて、Brandstein らの手法より低い位置推定誤差で推定することを確認した [3]。しかし、音源方向推定段階で大きな誤差が発生した際に、三角測量点が距離方向に大きくなずれて発生し、音源位置の追跡が困難になってしまうことが屋外実験を通じて分かった [4]。そこで、本稿では、音源方向推定を介さず、各マイクロホンアレイで求められる音源方向尤度をもとに直接音源位置の尤度を求め、音源位置推定を行う手法を提案する。これによって、音源位置が遠方にあたり、音源方向指定の分解能が低かったりしても、音源位置の推定誤差、特にマイクロホンアレイから音源への距離方向の推定誤差を抑えられることが期待できる。

本稿は、2 節で提案手法を説明し、3 節では数値シミュレーションによる提案手法の有効性の評価を行う。最後に、4 節にまとめ、今後の課題を記す。

*連絡先： 東京工業大学
〒 152-8552 東京都目黒区大岡山 2-12-1
E-mail: yamada@ra.sc.e.titech.ac.jp

2 音源尤度からの三次元音源追跡

本稿は、複数マイクロホンアレイから算出した音源方向の尤度を元に、音源位置の三次元位置の推定および追跡を行う。一般に単独マイクロホンアレイは音源方向の推定に用いられるが、複数マイクロホンアレイを用いれば、各マイクロホンアレイの方向推定戦略を統合することで、音源位置推定が可能となる。これまで山田らは、複数のマイクロホンアレイが推定した方向より三角測量を行うことで音源位置を推定する手法を報告しているが、三角測量を用いるアプローチは音源位置が遠方にある場合やマイクロホンアレイの伝達関数の分解能が低い場合に、著しい推定誤差を起こしたり三角測量が出来なかったりする問題がある。そこで本稿では、各マイクロホンアレイにて音源方向を推定するかわりに、音源方向推定に用いられる音源方向尤度を三次元音源位置尤度に変換する処理を施すことで、三角測量を介さない音源位置推定を実現し、遠方音源に対する位置推定誤差の抑制を試みる。

2.1 問題設定

本稿では、複数のマイクロホンアレイを用いて単音源を追跡することについて考える。マイクロホンアレイは N 個存在し、それぞれ

$$MA_1, \dots, MA_N$$

と定義する。各マイクロホンアレイは三次元空間上を移動・回転をすることができ、あるマイクロホンアレイ MA_n の時刻 t における状態を

$$\mathbf{m}_n(t) = [\mathbf{m}_{n,xyz}^T, \mathbf{m}_{n,\phi\theta\psi}^T]^T \quad (1)$$

$$\mathbf{m}_{n,xyz} = [x_n(t), y_n(t), z_n(t)]^T \quad (2)$$

$$\mathbf{m}_{n,\phi\theta\psi} = [\phi_n(t), \theta_n(t), \psi_n(t)]^T \quad (3)$$

とおき、既知であるとする。 $x_n(t), y_n(t), z_n(t)$ は MA_n の中心の3次元位置座標を指し、 $\phi_n(t), \theta_n(t), \psi_n(t)$ はそれぞれ MA_n のロール、ピッチ、ヨー角を指す。また、各マイクロホンアレイは M 個のマイクロホンから構成されており、 MA_n に収録される音響信号は時間領域で $\mathbf{s}_n(t) \in \mathbb{R}^M$ と記述する。音源は一つのみ存在し、点音源であると仮定する。この音源の三次元座標は、

$$\mathbf{e}(t) = [x_e(t), y_e(t), z_e(t)]^T \quad (4)$$

とする。本稿で取り組む問題は、各マイクロホンアレイ状態 $\mathbf{m}_n(t)$ と収録信号 $\mathbf{s}_n(t)$ から音源位置 $\mathbf{e}(t)$ を一定時間おきに推定することで、音源軌跡を推定することである。

2.2 追跡手法説明

これまで、マイクロホンアレイを用いて音源方向を推定する際は、各方向において音源が存在する尤もらしさ $P(\phi, \theta)$ (ϕ は方位角, θ は仰角) を算出し、最も $P(\phi, \theta)$ が大きくなる方向が音源方向であるとしている。この各マイクロホンアレイで得られる方向尤度 $P(\phi, \theta)$ を三次元位置に対する尤度に変換することで音源位置を推定することが提案手法のアプローチである。本手法は T_k おきに音源位置を推定する。つまり、 T_k 秒おきに推定音源軌跡を更新する。三次元空間上にグリッド点を設け、各グリッド点にて各マイクロホンアレイが算出する方向尤度 $P(\phi, \theta)$ を考慮に入れた評価関数を求めることで、音源位置に対する尤度を表現する。各行程の詳細は以下に記述する。

2.2.1 音源方向尤度の算出

音源方向に対する尤度と見なせる指標は多数報告されている。2つのマイクロホンでTDOA (Time Difference Of Arrival) を推定する手法の一つであるCSP法 [5] で用いられるCSP係数や、Delay-and-Sumビームフォーマから求める空間スペクトルは、マイクロホンアレイから見た方向をパラメータに持つスカラー量であり、一般に音源が存在する方向にピークが立つ性質を持つ [6]。本稿では、3つ以上のマイクロホンで構成されるマイクロホンアレイを想定し、Delay-and-Sumビームフォーマによる空間スペクトルより鋭いピークを音源方向に出すMUSICスペクトルを音源方向尤度として用いることを考える。MUSIC法 [7] とは、空間相関行列が張る固有空間を解析手法であり、目的音源の部分空間と雑音部分空間の直交性を用いて音源の方位・仰角を推定する手法である。角周波数 ω , 方位角 ϕ , 仰角 θ の音源からマイクロホンアレイへの伝達関数を $\mathbf{a}(\omega, \phi, \theta)$ とすると、 (ϕ, θ) における空間スペクトル $P(\phi, \theta)$ は

$$P(\phi, \theta) = \frac{1}{\omega_H - \omega_L + 1} \sum_{\omega=\omega_L}^{\omega_H} \frac{\mathbf{a}(\phi, \theta)^H \mathbf{a}(\phi, \theta)}{\mathbf{a}(\phi, \theta)^H \mathbf{E}(\omega) \mathbf{E}(\omega)^H \mathbf{a}(\phi, \theta)} \quad (5)$$

で表せる。ただし、 \mathbf{E} は空間相関行列において雑音部分空間が張る固有ベクトル行列であり、 ω_L, ω_H はそれぞれ空間スペクトルの評価に用いる角周波数の下限と上限である。空間スペクトル $P(\phi, \theta)$ はMUSICスペクトルとも呼ばれ、一般に方向推定をする際は、MUSICスペクトルがピークを取る方向を推定方向とする。本稿では MA_n で求めたMUSICスペクトルを $P_n(\phi, \theta)$ と記述し、 MA_n における音源方向に対する尤度であると見なす。

2.2.2 音源位置尤度への変換

各マイクロホンアレイで算出した音源方向尤度 $P_n(\phi, \theta)$ を用いて、音源位置尤度を表現することで、音源位置を推定する。具体的には、想定される三次元空間をグリッド状に分割し、各グリッド点ごとに音源位置尤度を求め、最も音源位置尤度が大きいグリッド点を推定音源位置とする。あるグリッド点 i の座標を

$$\mathbf{g}_i = [x_i, y_i, z_i]^T \quad (6)$$

とし、あるマイクロホンアレイ MA_n の中心からグリッド点 i までの方位角、仰角をそれぞれ ϕ_{ni} 、 θ_{ni} とおく。方向 (ϕ_{ni}, θ_{ni}) は、ベクトル $\mathbf{m}_{n,xyz} - \mathbf{g}_i$ を極座標変換することで得られる。あるグリッド点 i における音源位置尤度 $L(\mathbf{g}_i)$ は以下の式のように定義する。

$$L(\mathbf{g}_i) = \sum_n P(\tilde{\phi}_{ni}^{\text{round}}, \tilde{\theta}_{ni}^{\text{round}}) \quad (7)$$

$$\tilde{\phi}_{ni}^{\text{round}} = \text{round}(\tilde{\phi}_{ni}), \quad \tilde{\theta}_{ni}^{\text{round}} = \text{round}(\tilde{\theta}_{ni}) \quad (8)$$

$$\begin{bmatrix} \cos \tilde{\phi}_{ni} \cos \tilde{\theta}_{ni} \\ \sin \tilde{\phi}_{ni} \cos \tilde{\theta}_{ni} \\ \sin \tilde{\theta}_{ni} \end{bmatrix} = R_n^{-1} \begin{bmatrix} \cos \phi_{ni} \cos \theta_{ni} \\ \sin \phi_{ni} \cos \theta_{ni} \\ \sin \theta_{ni} \end{bmatrix} \quad (9)$$

ここで、 $\text{round}(\cdot)$ は伝達関数 $\mathbf{a}(\omega, \phi, \theta)$ の方位角・仰角の分解能に合わせて方向 (ϕ_{ni}, θ_{ni}) を丸める関数であり、 R_n は MA_n の姿勢を表す回転行列である。つまり、各マイクロホンアレイから見たグリッド点 i への方向を算出し、その各方向に対応する音源方向尤度を足し合わせた値を、グリッド点 i の音源位置尤度としている。

2.2.3 音源位置推定と追跡

全てのグリッド点において音源位置尤度を求めたあと、最大値を取るグリッド点を推定音源位置とする。このように推定音源位置を求めると、推定音源位置は必ずどれかのグリッド点上に求められる。よって、グリッド点間隔が大きい場合、最終的に推定される音源軌跡は大きく振動するような軌跡になる。そこで、移動平均フィルタやカルマンフィルタを用いてスムージングすることで、スムージングしない場合と比べてより良く音源軌跡を推定することができる。この効果は、次節『数値シミュレーション』で確認できる。

3 数値シミュレーション

提案手法の有効性と性能を検証するために、MATLAB[®] を用いて数値シミュレーションを行った。また、他の三次元追跡手法も MATLAB[®] で実装し、シミュレーション結果を比較することで提案手法の性能を評価する。

3.1 シミュレーション内容

図1のような三次元空間上を点音源が $z = 0$ で等速円運動をしているシナリオを考える。点音源は以下のダイナミクスに従って等速円運動を行い、ホワイトノイズを出力している。

$$\mathbf{e}(0) = [5, 0, 0]^T \quad (10)$$

$$\dot{\mathbf{e}}(t) = \left[-\frac{\pi}{5} y_e(t), \frac{\pi}{5} x_e(t), 0 \right]^T \quad (11)$$

音源位置は $T_k = 0.1$ 秒おきに推定し、計 10 秒間シミュレーションを行った。マイクロホンアレイは 4 つ用意し、各マイクロホンアレイの状態は表 3 に記されている。マイクロホンアレイは 16 ch の球形マイクロホンアレイであり、44.1 kHz、24 bit で収録を行う。また、三次元空間上の、

$$-6\text{m} \leq x \leq 6\text{m}$$

$$-6\text{m} \leq y \leq 6\text{m}$$

$$-5\text{m} \leq z \leq 5\text{m}$$

の範囲に、0.1 m おきにグリッド点を定義し、各グリッド点で音源位置尤度 $L(\mathbf{g}_i)$ を求めることで、各時刻の音源位置を推定する。本稿では、マイクロホンアレイ-音源間の距離、伝達関数の刻みが各音源追跡手法に与える影響を考察するために、音源との距離が 10 m の場合と 20 m の場合、伝達関数が 1° 刻みの場合と 5° 刻みの場合と異なるシナリオを用意し、計 4 種のシナリオに対してシミュレーションを行う。

また、提案手法の性能を他の既存手法と比較するため、複数マイクロホンアレイを用いて三角測量的に音源位置推定を行う LI (Linear Intersection) 法 [1] と、LI 法にガウスフィルタを加えることで外れ値となるような三角測量点を無視する MT-GSFT (Multiple Triangulation and Gaussian Sum Filter Tracking) 法 [3] を実装し、シミュレーションを行った。また、音源位置推定を行ったあと、スムージングをすることによる性能の変化を考察するため、移動平均フィルタを用いて音源軌跡をスムージングする場合としない場合の両方についてシミュレーションを行った。

3.2 シミュレーション結果・考察

図 1,2 はそれぞれマイクロホンアレイ-音源間距離が 10 m かつ伝達関数が 1° 刻みの場合と、距離が 20 m かつ伝達関数が 5° 刻みの場合シミュレーション結果である。両図より、提案手法は真の音源軌跡周辺を推定できていることが分かる。また、推定音源位置と真値間のユークリッド距離を推定誤差として、各シナリオにおける各手法の平均推定誤差と最大推定誤差を求め、表 1,2

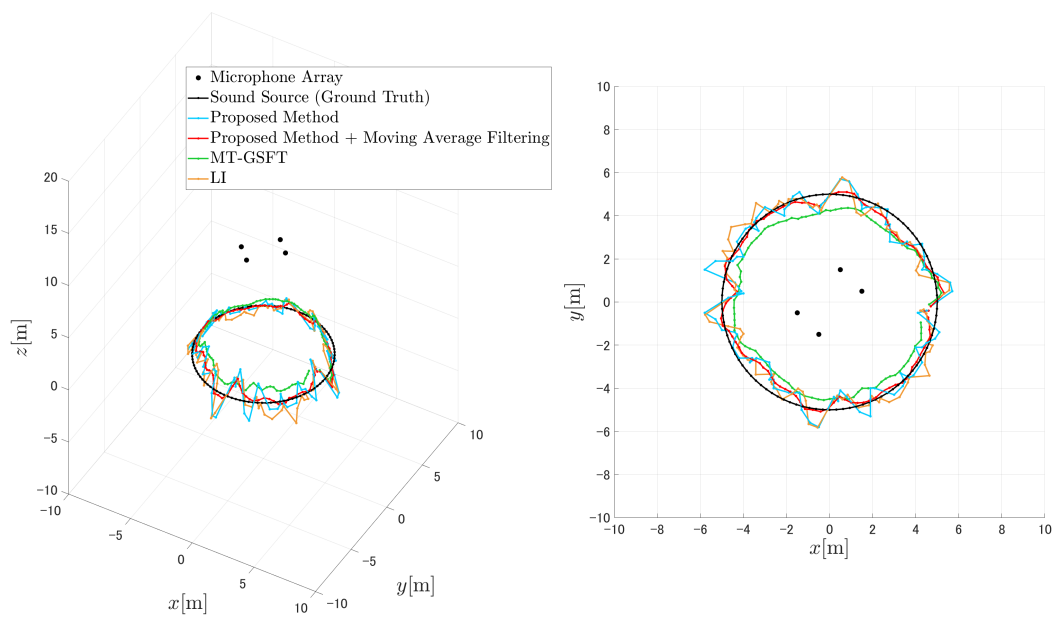


図 1: マイクロホンアレイの高さが 10 m, 伝達関数が 1° 刻みの場合の音源位置追跡結果 (左図は俯瞰図, 右図は上から見た図)

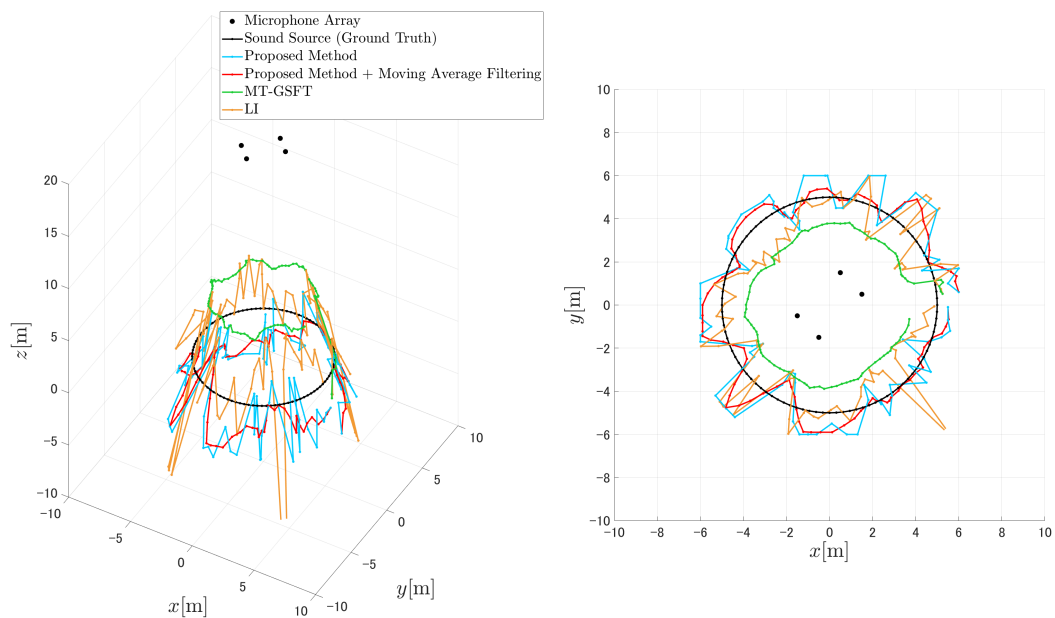


図 2: マイクロホンアレイの高さが 20 m, 伝達関数が 5° 刻みの場合の音源位置追跡結果 (左図は俯瞰図, 右図は上から見た図)

表 1: 音源位置推定誤差 (伝達関数が 1° 刻みの場合)

	高さ 10 m			高さ 20 m		
	平均誤差 [m]	最大誤差 [m]	分散 [m ²]	平均誤差 [m]	最大誤差 [m]	分散 [m ²]
提案手法	0.99	2.41	0.32	2.12	5.16	1.37
提案手法 + 移動平均フィルタ	0.53	1.57	0.13	1.25	4.16	0.61
MT-GSFT	1.45	2.03	0.11	1.88	3.66	0.23
LI	0.94	2.16	0.35	3.11	15.31	5.45

表 2: 音源位置推定誤差 (伝達関数が 5° 刻みの場合)

	高さ 10 m			高さ 20 m		
	平均誤差 [m]	最大誤差 [m]	分散 [m ²]	平均誤差 [m]	最大誤差 [m]	分散 [m ²]
提案手法	2.04	5.64	1.92	3.21	5.33	2.39
提案手法 + 移動平均フィルタ	1.69	3.45	0.65	2.45	5.31	1.72
MT-GSFT	1.47	2.32	0.17	5.70	6.56	1.19
LI	1.36	4.78	0.91	3.67	22.83	8.76

表 3: 各マイクロホンアレイの状態. 高さ z_n はシナリオによって異なる. (10 m or 20 m)

マイクロホンアレイ	状態 [m, m, m, rad, rad, rad]
MA ₁	[0.5, 1.5, *, 0, 0, π]
MA ₂	[1.5, 0.5, *, 0, 0, $3\pi/2$]
MA ₃	[-0.5, -1.5, *, 0, 0, 0]
MA ₄	[-1.5, -0.5, *, 0, 0, $\pi/2$]

にまとめた. 表 1, 2 より, 提案手法に移動平均フィルタを組み合わせることで, 特に距離が 20 m の場合に推定誤差を他手法と比べて小さく抑えられることが分かる. また, 移動平均フィルタの導入より, 音源軌跡のバラつきが抑えられたことが, 図 1, 2 の軌跡の形と, 表 3 の分散の減少から確認できる.

マイクロホンアレイ-音源間の距離が大きくなるにつれ, どの手法も特に距離方向の推定誤差が大きくなっていることが確認された. また, 伝達関数の刻みが大きくなると, 同様に距離方向の推定誤差が大きくなることが確認された. これはいずれの手法も, 音源が遠方にあるとき, 方向推定の微小な誤差が位置推定に大きく影響を与え, 距離方向にずれてしまう特性を持つからであると考えられる.

マイクロホンアレイ-音源間の距離が 20 m かつ伝達関数の刻みが 5° のとき, MT-GSFT は著しい推定誤差を示している. これは, MT-GSFT は前ステップで求めた三角測量点と近い位置にある三角測量点付近を推定する傾向にあることと, 三角測量は伝達関数の分解能が低いほど観測が不得意な領域が大きく現れることに起因すると考えられる. 提案手法は, グリッド点が十分に三次元空間を網羅していれば, このような三角測量的なアプローチに見られる観測が難しい領域が現れることがないため, 伝達関数の刻みが大きくなっても, 伝達

関数の刻みが小さい場合と比べた推定誤差の増分は小さく抑えることができたと思われる.

4 終わりに

本稿は, 複数マイクロホンアレイにおける音源方向尤度を元に, 三次元音源位置の推定および追跡を行う手法の提案をした. 提案手法の有効性を数値シミュレーションによって検証したところ, 三角測量的に音源位置を推定する他手法に比べて, マイクロホンアレイ-音源間の距離の増大や, 伝達関数の刻みの増大による推定誤差の増加を抑えられる効果があることが確認できた. 提案手法の複数音源に対する拡張や, 音源位置尤度の分布を応用した音源ダイナミクスの推定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 の助成をうけた.

参考文献

- [1] Brandstein, M. S., and Silverman, H. F., A practical methodology for speech source localization with microphone arrays, *Computer Speech & Language*, 11(2), pp. 91–126, (1997)
- [2] Gabriel, D., *et al*, 2D sound source position estimation using microphone arrays and its application to a VR-based bird song analysis system, *Advanced Robotics*, 30(7–8), pp. 403–414, (2019)

- [3] Yamada, T., *et al*, Sound Source Tracking by Drones with Microphone Arrays, *SII2020* (Submitted)
- [4] 山田他, 複数マイクロホンアレイを搭載した複数の UAV による移動音源の三次元追跡手法の実収録音評価, 第 37 回日本ロボット学会学術講演会 (*RSJ2019*), (2019)
- [5] Knapp, C. and Carter, G., The generalized correlation method for estimation of time delay, *IEEE transactions on acoustics, speech, and signal processing*, 24(4), pp. 320–327, (1976)
- [6] 浅野 太, 音のアレイ信号処理, コロナ社, (2009)
- [7] Schmidt, R., Multiple emitter location and signal parameter estimation, *IEEE Trans. on Antennas and Propagation*, 34(3), pp. 276—280, 1986.