複数マイクロホンアレイの同期および 3 次元位置・姿勢推定の同時最適化の検討

○杉山地塩[1] 糸山克寿[1] 西田健次[1] 中臺一博[1,2]
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・インスティチュート・ジャパン

1. はじめに

本稿では, 音源位置および複数マイクロホンアレイの位置・向き・時間同期情報を推定する問題について扱う. 従来の手法としてはマイクロホンアレイ間の同期を仮定するものや, 変数を分けて最適化計算を行うために, 局所最適解の影響を受けやすいものが提案されている. そこで, 我々はマイクロホンアレイ間の音響信号の観測時間差 (TDOA, time difference of arrival) とマイクロホンアレイから見た音源方向 (DOA, direction of arrival) の 2 種類のデータから目的関数を作り, 音源位置および複数マイクロホンアレイの位置・向き・観測時刻オフセットを同時に最適化する計算手法を提案する. この手法を用いて, マイクロホンアレイ間の同期がない場合にも局所最適解に陥らず素早い推定ができることを, 数値シミュレーションによって示した.

2. 関連研究

複数のマイクロホンで構成されるマイクロホンアレイを用いると, 観測した音響信号からその飛来方向を推定することができ, この技術は音響信号処理の分野で多く研究されている [1]. また, 位置・向きが既知であり, かつ同期された複数のマイクロホンアレイによる観測を行うことで, 三角測量から音源位置を計算できる. マイクロホンアレイ間の位置・向きの関係がわかり, 時間同期がなされることは, より大規模なマイクロホンアレイを構築することと同義である. しかし, 複数のマイクロホンアレイに対してそれらの位置・向き・時間同期を高精度に設定することはハードウェア面で難しいか高価であるため, 実際には各パラメータを校正した上で観測を行う [2, 3, 4, 5].

非同期分散マイクロホンアレイ (アドホックマイクロホンアレイ) を構築する研究は多く報告されており, TDOA に基づいて音源定位を行うもの [6], TOA (time of arrival) を利用するもの [7], TOF (time of flight) を用いるもの [8, 9] がある. TDOA を利用するものの中には, マイクロホン位置を校正する研究もある [10]. また, 非同期の状態を表すために観測時刻オフセットを導入し, それを推定する取り組みもなされている [11].

Plinge ら [12, 13] は同期された複数のマイクロホンアレイに対して, DOA と TDOA に基づく目的関数を統合し, 音源位置とマイクロホンアレイの位置・向きを最適化した. さらに, マイクロホンアレイ間の同期がない場合について, Woźniak ら [14] はマイクロホンアレイごとに観測時刻オフセットを加え, 位置および向きとともに推定した. これは初めに位置と向きを最適化し,

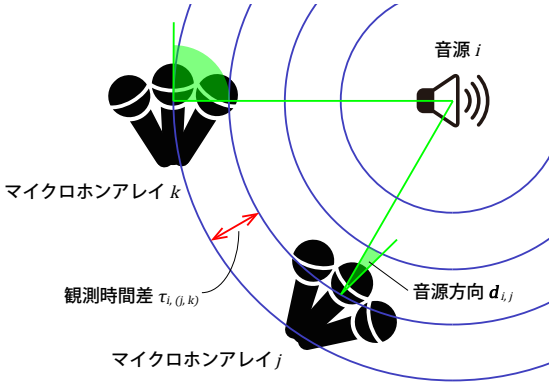


図 1: 複数マイクロホンアレイによる音響信号の観測

続いて時間オフセットと座標のスケールを計算する手法を提案している. 他にもマイクロホンアレイの大きさを利用して座標スケールを求める手法 [15] も知られている. また, 推定性能を向上するために, 最適化アルゴリズムとして RANSAC (random sample consensus) を用いた研究もある [16, 17].

本稿では, マイクロホンアレイ間の同期がない仮定で, DOA と TDOA に基づいて位置・向き・時間オフセットを同時に最適化する手法を提案し, その有効性を数値シミュレーションによって検証する. これによって局所最適解への収束を回避するとともに簡単なアルゴリズムによって最適化を行い, さらに推定に必要な音源数の削減を目指す.

3. 提案手法

この章では提案する推定手法について説明する. 初めに, 音源方向 (DOA) と観測時間差 (TDOA) それぞれに関して, 音源位置およびマイクロホンアレイの位置・向き・観測時刻オフセットを推定する目的関数 [14] を確認する. 次に, それらの目的関数を統合した新しい目的関数を提案する.

3.1 音響信号の伝播モデルと従来の計算手法

2 次元または 3 次元空間に, M 個の音源と N 個のマイクロホンアレイがあるとする. このとき, 各マイクロホンアレイで観測できる音源方向と, マイクロホンアレイ間の音響信号の観測時間差は, 図 1 に示される. 音源 i からの信号を観測すると, マイクロホンアレイ j,k について音源方向を表す単位ベクトル $\boldsymbol{d}_{i,j}$ と, 観測時間差 $\tau_{i,(j,k)}$ は次の式で計算できる [14].

$$\boldsymbol{d}_{i,j} = \mathbf{R}_j^T \frac{\boldsymbol{s}_i - \boldsymbol{a}_j}{\left\|\boldsymbol{s}_i - \boldsymbol{a}_j\right\|_2} \quad (1)$$

$$\tau_{i,(j,k)} = c\left(\left\|\boldsymbol{s}_i - \boldsymbol{a}_j\right\|_2 - \left\|\boldsymbol{s}_i - \boldsymbol{a}_k\right\|_2\right) + \delta_j - \delta_k. \quad (2)$$

ここで $\boldsymbol{s}_i, \boldsymbol{a}_j$ は, 音源 i とマイクロホンアレイ j の座標を表す. また, $\mathbf{R}_j = \mathbf{R}(\boldsymbol{\theta}_j)$ はマイクロホンアレイの向き $\boldsymbol{\theta}_j$ に対応した回転行列であり, 2 次元では $\boldsymbol{\theta}_j = \theta_{j1}$, 3 次元では ZYX オイラー角を用いて $\boldsymbol{\theta}_j = (\theta_{j1}, \theta_{j2}, \theta_{j3})$ とする. δ_j はマイクロホンアレイ j の観測時刻オフセットを表し, c は音速である.

$\boldsymbol{d}_{i,j}$ および $\tau_{i,(j,k)}$ の値が観測によって与えられたとき, 位置・向き・観測時刻オフセットを最適化する式は (3), (4) のようになる.

$$\tilde{\mathbf{S}}, \tilde{\mathbf{A}}, \hat{\boldsymbol{\theta}} = \underset{\mathbf{S},\mathbf{A},\boldsymbol{\theta}}{\operatorname{argmin}} \sum_{i=1}^{M}\sum_{j=1}^{N} D_{i,j}$$

$$D_{i,j} = \left\|\mathbf{R}_j^T(\boldsymbol{s}_i - \boldsymbol{a}_j) - \boldsymbol{d}_{i,j}\left\|\boldsymbol{s}_i - \boldsymbol{a}_j\right\|_2\right\|_2^2 \quad (3)$$

$$\hat{\mathbf{S}}, \hat{\mathbf{A}}, \hat{\boldsymbol{\delta}} = \underset{\mathbf{S},\mathbf{A},\boldsymbol{\delta}}{\operatorname{argmin}} \sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1}^{N} T_{i,j,k}$$

$$T_{i,j,k} = \left\|\boldsymbol{s}_i - \boldsymbol{a}_j\right\|_2 - \left\|\boldsymbol{s}_i - \boldsymbol{a}_k\right\|_2 + c\,\delta_j - c\,\delta_k - c\,\tau_{i,(j,k)} \quad (j < k) \quad (4)$$

ただし, $\mathbf{S} = (\boldsymbol{s}_1, ..., \boldsymbol{s}_M)$, $\mathbf{A} = (\boldsymbol{a}_1, ..., \boldsymbol{a}_N)$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_N)$, $\boldsymbol{\delta} = (\delta_1, ..., \delta_N)$ とする. 式 (3) では拡縮の自由度が残るため, 座標 $\mathbf{S}, \mathbf{A}$ は一意に定まらない. よって, 従来法 [14] (*Sequential*) は初めに式 (3) で最適化を行い, 続いて座標スケール $\alpha > 0$ を観測時間差に基づいて推定することにより, 最終的な座標 $\hat{\mathbf{S}} = \alpha\tilde{\mathbf{S}}, \hat{\mathbf{A}} = \alpha\tilde{\mathbf{A}}$ を得る. この手法は式 (3) の計算で局所最適解に陥りやすく, 以降の推定値も大きい誤差を生じる.

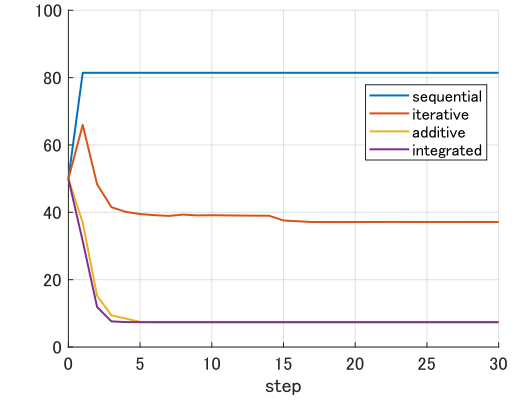
3.2 目的関数の提案

この課題を解決するため, 式 (3), (4) を交互に計算することで推定値を更新する手法 (*Iterative*) を提案する. また, 音源方向と観測時間差についての目的関数を組み合わせる方法を考える. 式 (3), (4) による最適化計算はどちらも目的関数 D, T を最小化することから, それらの和を最小化することで位置・向き・観測時刻オフセットの変数を一度に評価できる (*Additive*).

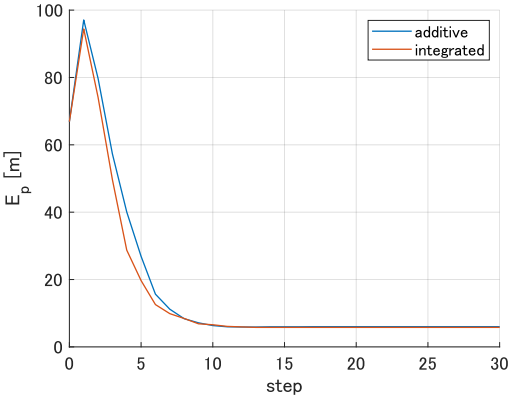
$$\hat{\mathbf{S}}, \hat{\mathbf{A}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\delta}} = \underset{\mathbf{S},\mathbf{A},\boldsymbol{\theta},\boldsymbol{\delta}}{\operatorname{argmin}} \sum_{i=1}^{M}\sum_{j=1}^{N}\left(D_{i,j} + \textstyle\sum_{k=1}^{N} T_{i,j,k}\right)$$

$$(j < k) \quad (5)$$

式 (5) において, これは i, j, k の組み合わせに応じた $D = 0, T = 0$ の連立方程式に近似できる. そこで計算ステップを削減するために, 共通項である $\left\|\boldsymbol{s}_i - \boldsymbol{a}_j\right\|_2$ を消去し, 新たな目的関数を作る.



(a) 2D



(b) 3D

図 2: 音源・マイクロホンアレイ定位の平均誤差

$$\hat{\mathbf{S}}, \hat{\mathbf{A}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\delta}} = \underset{\mathbf{S},\mathbf{A},\boldsymbol{\theta},\boldsymbol{\delta}}{\operatorname{argmin}} \sum_{i=1}^{M}\sum_{j=1}^{N}\sum_{k=1}^{N} I_{i,j,k}$$

$$I_{i,j,k} = \left\|\mathbf{R}_j^T(\boldsymbol{s}_i - \boldsymbol{a}_j) - \boldsymbol{d}_{i,j}\left(T_{i,j,k} - \left\|\boldsymbol{s}_i - \boldsymbol{a}_j\right\|_2\right)\right\|_2^2$$

$$(j \le k) \quad (6)$$

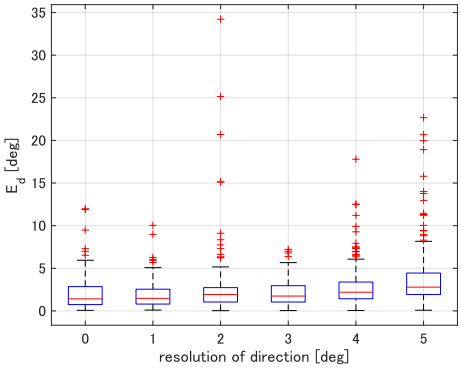
この計算を *Integrated* とし, 本稿の提案手法とする.

ここで, 式 (5) または (6) による最適化計算は, 式 (3) と (4) を独立して計算するよりも, 推定する変数の個数に対して最小化する i, j, k の組み合わせが多い. よって, 少ない音源数 (観測回数) でも劣決定系にならず安定した推定が可能となる.

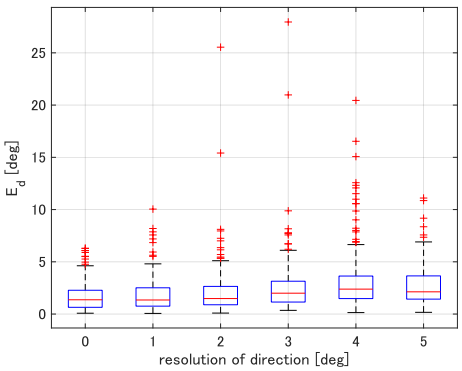
本稿では, 2 次元空間に対して上に挙げた 4 手法すべてを, また 3 次元空間に対して *Additive* および *Integrated* を実装し, その推定精度を検証する. 解を一意に定めるため, あるマイクロホンアレイを基準として座標 $\boldsymbol{s}_1$ を原点 O に, 回転 $\boldsymbol{\theta}_1$ を $\mathbf{0}$[deg] に, 観測時刻オフセット δ_1 を 0[s] に固定する. また, 最適化計算には内点法を用いる.

4. 評価実験

提案手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. 音源 7 個とマイクロホンアレイ 3 個を

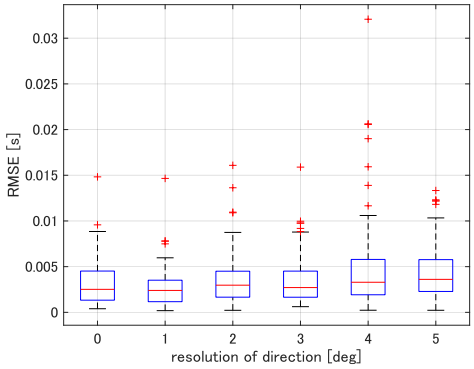


(a) *Additive*

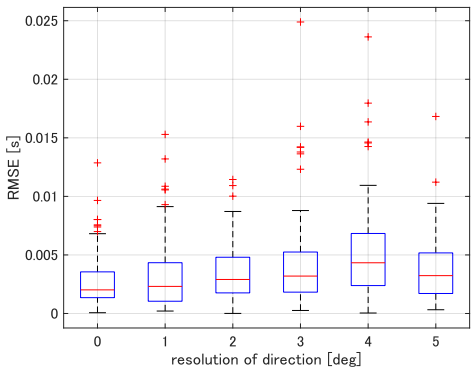


(b) *Integrated*

図 3: マイクロホンアレイの向きの誤差 E_d (3D)



(a) *Additive*



(b) *Integrated*

図 4: 配置ごとの観測時刻オフセットの RMSE (3D)

$120 \times 120[\mathrm{m}^2]$ (2D), $120 \times 120 \times 120[\mathrm{m}^3]$ (3D) の空間に一様分布の乱数で配置し, これを 100 パターン用意した. また, マイクロホンアレイの向きは一様分布, 観測時刻オフセットは $\mathcal{N}(0, 10^{-4})[\mathrm{s}]$ に従う乱数で与え, これらを真値とする.

音源方向 $\boldsymbol{d}$ および観測時間差 τ は生成した位置・向き・観測時刻オフセットから計算される. このとき音源方向には $\mathcal{N}(0, 2^2)[\mathrm{deg}]$ の誤差を加算し, 分解能を 1[deg] とする. 観測時間差には $\mathcal{N}(0, 0.005^2)[\mathrm{s}]$ の誤差を与え, 16[kHz] 単位で時間差を測定できると仮定した.

$$E_p = \frac{1}{M+N}\left(\sum_{i=1}^{M}\left\|\hat{\boldsymbol{s}}_i - \check{\boldsymbol{s}}_i\right\|_2 + \sum_{j=1}^{N}\left\|\hat{\boldsymbol{a}}_j - \check{\boldsymbol{a}}_j\right\|_2\right) \tag{7}$$

手法ごとの定位の誤差 E_p の推移を図 2 に示す. これは各音源・マイクロホンアレイについて真値と推定値のユークリッド距離を計算し, その平均を表した値である. ただし $\hat{\cdot}$ は推定値, $\check{\cdot}$ は真値に対応する. 最適化計算における一定の評価回数を 1 ステップとして, 100 通りの配置に対する平均の誤差を記録した.

Sequential は最初のステップで局所最適解に収束し, *Iterative* は *Sequential* よりも小さい誤差を示すが, これも局所最適解に陥っている. 対照的に, *Additive* と *Integrated* の誤差は約 8[m] であり, 音源方向と観測時間差に誤差が含まれることを考慮して, これらは正しく推定できていると考える.

このシミュレーションは音源方向と観測時間差を高い精度で測定できる条件のもと行われている. よって, 反響などによる劣化の考慮を必要とする, 実際に観測した音響信号を用いた評価が今後の課題である.

4.1 向きと観測時刻オフセットの推定精度

次に, 3 次元におけるマイクロホンアレイの向きと観測時刻オフセットの推定精度を検証する. 音源方向データの観測分解能を 0–5[deg] の範囲で 1[deg] ごとに変え, 100 通りの配置を生成した. このとき分解能 0[deg] とは, 音源方向 $\boldsymbol{d}$ に丸め込みを行わないことを意味する. その他の条件は先のシミュレーションと同一である. 基準とするマイクロホンアレイについての誤差は 0 になるため, それを除いた 2 個のマイクロホンアレイの値を使用する. 結果は図 3 および 4 の通りである.

向きの誤差 E_d は次の式で定義される.

$$E_d = \arccos\left(\frac{\hat{\mathbf{R}}_j\boldsymbol{v} \cdot \check{\mathbf{R}}_j\boldsymbol{v}}{\|\boldsymbol{v}\|_2^2}\right) \ (j = 2, \ldots, N). \tag{8}$$

ベクトル $\boldsymbol{v} = [1, 1, 1]^T$ を回転行列 $\mathbf{R}$ で回転したものに対して, コサイン類似度から方向のずれを計算する. 観測時刻オフセットは配置パターンごとの RMSE (root mean square error) で評価する.

2 手法に精度の差はなく, 向きは 10[deg] 未満, 時間オフセットは 10[ms] 未満の誤差を示す. また, 音源方

表 1: 1 回の推定に要する時間 [s]

手法	2D	3D
Sequential	1.71	–
Iterative	4.36	–
Additive	1.97	9.42
Integrated	1.54	8.32

向の分解能が低下するほど, いずれの誤差も大きくなっている.

4.2 計算時間

数値シミュレーションについて, 1 配置の推定に要した時間の平均を表 1 に示す.

2 次元では *Iterative* が最長であり, 局所最適解に素早く収束した *Sequential* よりも, *Integrated* の計算時間は短い. 推定精度の高い *Additive* と *Integrated* を比較すると, 2 次元・3 次元ともに *Integrated* の方が短時間で処理できる. しかし, 3 次元での計算に 8–9[s] 要することを鑑みると, オンラインシステムとして実装するには計算時間のさらなる短縮が必要である.

5. おわりに

本稿では, 複数音源からの音響信号の観測によって, 音源位置およびマイクロホンアレイの位置・向き・観測時刻オフセットを推定する際の最適化手法について述べた. この問題について簡単なアルゴリズムで最適化計算を行う場合, マイクロホンアレイから見た音源方向とマイクロホンアレイ間の観測時間差を同時に評価する手法は, それらを独立して評価する手法よりも高い精度で推定できることを示した. また, 同時に最適化する手法について *Additive* および *Integrated* の推定精度はほぼ等しいことから, 計算時間で優れている *Integrated* を, 本稿では最良の手法として提案する.

ただし, 本稿では入力データに加わる誤差が小さい仮定でシミュレーションを行っている. よって, 実際に観測を行い, 反響などを考慮した入力で性能を比較することが今後の課題となる. さらに, 3 次元空間を動きながら観測するロボットやドローンなどへの実装を想定し, 計算時間の短縮も求められる.

謝 辞 本研究は JSPS 科研費 JP19K12017, JP19KK0260 および JP20H00475 の助成を受けた.

参 考 文 献

[1] 浅野 太: “音のアレイ信号処理”, 音響テクノロジーシリーズ/日本音響学会編, コロナ社, 2011.

[2] T. Yamada, et al.: “マイクロホンアレイ搭載ドローンによる音源方向尤度統合に基づく音源追跡”, 第 *57* 回人工知能学会 *AI* チャレンジ研究会予稿集, 2020.

[3] D. Gabriel, et al.: “2D sound source position estimation using microphone arrays and its application to a VR-based bird song analysis system”, *Advanced Robotics*, vol. 33, No. 7-8, pp. 403–414, 2019.

[4] S. D. Valente, et al.: “Geometric calibration of distributed microphone arrays from acoustic source correspondences”, *2010 IEEE International Workshop on Multimedia Signal Processing*, pp. 13–18, 2010.

[5] A. Plinge and G. A. Fink: “Geometry calibration of distributed microphone arrays exploiting audio-visual correspondences”, *2014 22nd European Signal Processing Conference (EUSIPCO)*, pp. 116–120, 2014.

[6] A. Canclini, et al.: “Acoustic Source Localization With Distributed Asynchronous Microphone Networks”, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 21, No. 2, pp. 439–443, 2013.

[7] S. T. Birchfield: “Geometric microphone array calibration by multidimensional scaling”, *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03).*, vol. 5, pp. V–157, 2003.

[8] M. Crocco, et al.: “A Bilinear Approach to the Position Self-Calibration of Multiple Sensors”, *IEEE Transactions on Signal Processing*, Vol. 60, No. 2, pp. 660–673, 2012.

[9] M. Crocco, et al.: “A Closed Form Solution to the Microphone Position Self-Calibration Problem”, *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2597–2600, 2012.

[10] Y. Kuang and K. Åström: “Stratified sensor network self-calibration from TDOA measurements”, *21st European Signal Processing Conference (EUSIPCO 2013)*, pp. 1–5, 2013.

[11] P. Pertilä, et al.: “Passive Temporal Offset Estimation of Multichannel Recordings of an Ad-Hoc Microphone Array”, *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, No. 11, pp. 2393–2402, 2013.

[12] A. Plinge and G. A. Fink: “Geometry calibration of multiple microphone arrays in highly reverberant environments”, *2014 14th IWAENC*, pp. 243–247, 2014.

[13] A. Plinge, et al.: “Passive online geometry calibration of acoustic sensor networks”, *IEEE Signal Processing Letters*, vol. 24, No. 3, pp. 324–328, 2017.

[14] S. Woźniak and K. Kowalczyk: “Passive joint localization and synchronization of distributed microphone arrays”, *IEEE Signal Processing Letters*, vol. 26, No. 2, pp. 292–296, 2019.

[15] F. Jacob, et al.: “DOA-based microphone array postion self-calibration using circular statistics”, *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, pp. 116–120, 2013.

[16] J. Schmalenstroeer, et al.: “Unsupervised geometry calibration of acoustic sensor networks using source correspondences”, *Twelfth Annual Conference of the International Speech Communication Association*, 2011.

[17] F. Jacob, et al.: “Microphone array position self-calibration from reverberant speech input”, *IWAENC 2012; International Workshop on Acoustic Signal Enhancement*, VDE, pp. 1–4, 2012.