

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 3D Sound Source Tracking for Drones Using Direction Likelihood Integration

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、ドローンに搭載された複数のマイクロホンアレイを用いた3次元音源追跡手法を提案する。一般的にマイクロホンアレイは、信号処理を通じて音源方向を推定するツールとして用いられている。マイクロホンアレイを搭載することで、ドローンは危険な災害現場で助けを呼ぶ人の捜索がすることが可能になる。多くの音源が存在する屋外環境では、音源方向の情報だけでは音源を特定するには不十分であり、代わりに音源位置を取得する必要がある。そこで、単一のマイクロホンアレイでは三次元位置を取得することは困難であるため、三角測量を用いた三次元音源追跡手法が多数報告されている。しかし、実際の屋外環境ではドローン自体からのノイズや未知の外部ノイズが存在するため、方向推定が悪化しやすく、音源方向推定に含まれる離散性も相まって、三角測量による音源位置推定もより悪化しやすい。本研究では、三角測量に代えて、音源位置の尤度分布を推定することで音源位置を追跡することを提案する。従来、各マイクロホンアレイが音源の方向を定位する際には、音源の方向の尤度分布を計算し、最も尤度の高い方向を取るのが一般的であるが、本研究では、最大尤度に着目するのではなく、全てのマイクロホンアレイの方向尤度を統合し、音源位置尤度の分布を推定しながら音源追跡を行う手法を提案する。これにより、三角測量では困難な音源位置分布を非ガウス形で表現することが可能となる。提案手法は、実在するドローンのノイズを用いて数値的に推定し、三角測量に基づく他のトラッキング手法と比較して評価した。シミュレーションの結果、提案手法の有効性が実証され、40m先の音源を4m以下のRMSEで追跡できることがわかった。

1 はじめに

音響シーン解析の分野では、マイクロホンアレイを用いた音源の定位が盛んに研究されており、有用な技術として期待されている。例えば、ドローンと組み合わせることで、災害現場の瓦礫の下に埋もれている人をドローンが捜索することが可能となる [1,2]。一般に、音源の方向を推定するには、単独のマイクロホンアレイを用いるが、単独マイクロホンアレイを用いるのではなく、複数マイクロホンアレイを用いることで、推定された方向を統合して音源の位置を推定することが

できる。複数のマイクロホンアレイを用いて音源位置を推定する方法の一つとして、音源方向に基づく三角測量が報告されている [3-5]。しかし、ドローン聴覚の分野では、定位すべき音源がドローンから遠く離れている場合が多く、加えてドローンのノイズによって方向推定が悪化するため、三角測量が悪化する可能性がある。そこで、本稿では、方向推定ではなく、位置尤度の分布を推定することで音源位置を追跡する手法を提案する (図 1)。これにより、音源位置の分布を一般的な形で表現することができ、マイクロホンアレイのペアごとの情報ではなく、すべてのマイクロホンアレイの情報を用いた音源位置の推定を行うことができる。本稿の残りは以下のように構成されている。第2節では提案手法を説明し、第3節では数値シミュレーションによる評価を行う。最後に結論と今後の課題を述べる。

*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

本稿はシンポジウム Quiet Drones 2020 で発表した内容を和訳したものである。

<https://www.quietdrones.org/conferences/1-quiet-drones-2020/>

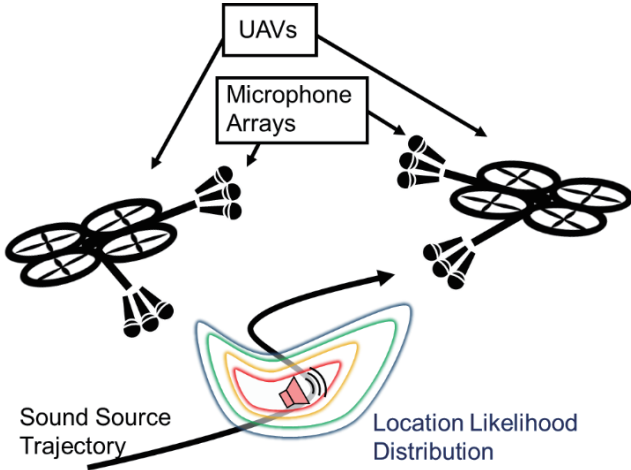


図 1: 複数マイクロホンアレイ搭載ドローンを用いた音源追跡

2 三次元音源追跡

本節では、位置尤度推定に基づく音源追跡手法を提案する。音源位置尤度分布を、各マイクロホンアレイから得られた音源方向尤度分布を統合することで求め、得られた音源位置尤度を元に Particle Filter を通じて音源位置を追跡する。

2.1 問題設定

複数マイクロホンアレイを用いた音源位置追跡を考える。マイクロホンアレイは全部で N 個存在すると仮定し、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。追跡する音源は点音源であると仮定し、音源の三次元座標は、

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

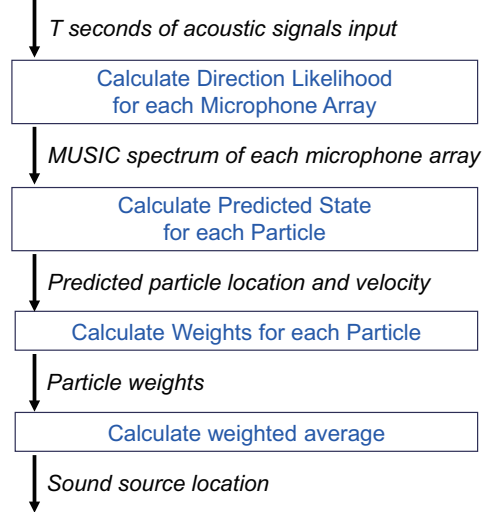


図 2: 追跡手法のフロー (1 サイクル分)

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

本手法は方向推定手法によって得られる方向尤度分布の統合に基づいており、図 2 に提案追跡手法のフローを示している。一般的に、音源方向推定を行う際は、音源方向尤度 $P(\phi, \theta)$ を計算し、 $P(\phi, \theta)$ の最大値をとるような方位角 ϕ 、仰角 θ を推定方向としている。本稿では各マイクロホンアレイで得られた音源方向尤度 $P(\phi, \theta)$ を統合し、三次元音源位置尤度として変換することで、三次元音源位置の追跡を図る。推定された三次元音源位置尤度は Particle filter に適用し、各粒子に位置尤度を重みとして与えることで、音源軌跡を推定する。

2.2.1 方向尤度分布の推定

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 [6] 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 [7] 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。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 [10] 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, 音源位置尤度分布を表現することで, 音源位置を追跡する。任意の三次元位置 \mathbf{x} において, $\mathbf{m}_{n,xyz}$ から x の方向を (ϕ_n, θ_n) とおく。このとき, 三次元位置 \mathbf{x} における音源位置尤度は以下のように定義する。

$$L(\mathbf{x}) = \sum_n P(\tilde{\phi}_{ni}^{\text{round}}, \tilde{\theta}_{ni}^{\text{round}}) \quad (6)$$

$$\tilde{\phi}_{ni}^{\text{round}} = \text{round}(\tilde{\phi}_{ni}), \quad \tilde{\theta}_{ni}^{\text{round}} = \text{round}(\tilde{\theta}_{ni}) \quad (7)$$

$$\begin{bmatrix} \cos \tilde{\phi}_{ni} & \cos \tilde{\theta}_{ni} \\ \sin \tilde{\phi}_{ni} & \cos \tilde{\theta}_{ni} \\ \sin \tilde{\theta}_{ni} \end{bmatrix} = \mathbf{R}_n^{-1} \begin{bmatrix} \cos \phi_{ni} & \cos \theta_{ni} \\ \sin \phi_{ni} & \cos \theta_{ni} \\ \sin \theta_{ni} \end{bmatrix} \quad (8)$$

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, \mathbf{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点 \mathbf{x} 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, 点 x の音源位置尤度としている。

2.2.3 音源位置尤度分布に基づく追跡

本手法では, 前小節で求めた音源位置尤度 $L(\mathbf{x})$ を Particle filter に適用し, 音源位置を追跡する。パーティクルフィルタに用いるパーティクルの個数を I とし, 時

刻 k におけるパーティクル i の状態と重みをそれぞれ \mathbf{x}_k^i, w_k^i とする。状態 \mathbf{x}_k^i はパーティクルの三次元位置と速度を含み, 以下のように記述する。

$$\mathbf{x}_k^i = \left[x_k^i, y_k^i, z_k^i, \dot{x}_k^i, \dot{y}_k^i, \dot{z}_k^i \right]^T \quad (9)$$

また, パーティクルは excitation-damping モデル [9] に従うと仮定し, 以下のような挙動を示すと仮定する。

$$\mathbf{x}_k^i = F \mathbf{x}_{k-1}^i + H \mathbf{v} \quad (10)$$

$$F = \begin{bmatrix} \mathbf{I} & T\mathbf{I} \\ \mathbf{O} & a\mathbf{I} \end{bmatrix}, \quad H = \begin{bmatrix} \mathbf{O} \\ b\mathbf{I} \end{bmatrix}, \quad (11)$$

$$\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (12)$$

ここで, $\mathbf{I} \in \mathbb{R}^3$ は単位行列, \mathbf{O} はゼロ行列を示す。各パーティクルは $L(\mathbf{x}_{k, \text{pos}}^i)$ に比例した重みが与えられるため, 重みは下式で表される。

$$w_k^i = w_{k-1}^i \frac{L(\mathbf{x}_{k, \text{pos}}^i)}{\sum_i L(\mathbf{x}_{k, \text{pos}}^i)} \quad (13)$$

ここで, $\mathbf{x}_{k, \text{pos}}^i = [x_k^i, y_k^i, z_k^i]^T$ である。リサンプリングは有効パーティクル数が閾値 N_{thr} を下回ったときに行う。つまり, 下式の条件が満たされる時, 各パーティクルの重みは $1/I$ にリセットされる。

$$\frac{1}{\sum_i (w_k^i)^2} \leq N_{thr} \quad (14)$$

パーティクルの初期化は, 以下の分布からパーティクルをサンプルすることで得られる。

$$\mathbf{x}_0^i \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \quad (15)$$

$$\boldsymbol{\mu}_0 = [\boldsymbol{\mu}_{0, \text{pos}}^T, 0, 0, 0]^T \quad (16)$$

$$\boldsymbol{\Sigma}_0 = \begin{bmatrix} \sigma_{\text{pos}}^2 \mathbf{I} & \mathbf{O} \\ \mathbf{O} & \sigma_{\text{vel}}^2 \mathbf{I} \end{bmatrix} \quad (17)$$

ここで, $\boldsymbol{\mu}_0$ は三角測量を用いることで得る三次元位置である [4, 5]。具体的には, 複数の三角測量点を計算し, 全ての三角測量点の平均点を $\boldsymbol{\mu}_0$ とおく。

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 シミュレーション設定

図3のように, 2台のドローン(黒点)に2個ずつマイクロホンアレイ(黒丸)を取り付け, 高さ30mでホバ

表 1: 音源追跡に用いた変数値

Variables	Value
a	0.5
b	3
I	500
N_{thr}	350
σ_{pos}^2	25
σ_{vel}^2	25

リングして停止しているシナリオを考える。追跡する音源（赤線）は半径5mの円を描くように等速で移動する。音源の速さは π m/sであり、10秒で1周する。円軌道の中心と両ドローンとの水平距離は $l = 10, 20, 30, 40, 50$ mの5種類のパターンに対してシミュレーションを行った。各マイクロホンアレイは図4のような球型マイクロホンであり、1つのマイクロホンアレイは $M = 16$ 個のマイクロホンから成っている。音響信号は16 kHz, 24 bitで収録される。音源位置は0.2秒ごとに10秒間推定されることで、音源軌跡を推定する。MUSICスペクトルは水平角、仰角共に5度刻みで推定される。提案手法に用いられる各種パラメータは表1に示す。音源は日本音響学会 新聞記事読み上げ音声コーパス (JNAS) 内の男声コーパス・女声コーパス各10種、ホワイトノイズ10種の計30種を出力させ、各種音源に対してシミュレーションを行った。また、他手法との比較のため、三角測量点の平均点に対してカルマンフィルタを適用した手法 [8] と推定音源方向を元にパーティクルフィルタで推定する手法 [9] を実装し、同じ条件の元でシミュレーションを行った。

3.2 結果

シミュレーション結果を図5と図6に示す。いずれの図からも、提案手法が音源軌跡の概形を追跡できていることがわかる。推定序盤では z 軸方向に推定誤差が大きく発生しているが、これはパーティクル初期化のために行った三角測量がドローンノイズの影響で不安定であり、初期位置が音源に近くないためである。しかし、 z 軸の初期位置が外れているにもかかわらず、提案手法によって真値に収束している様子が図6で見られ、提案手法が移動音源に対して有効であることが分かる。追従誤差は表2-4に示しており、表は以下のように定義される Root Mean Squared Error (RMSE) を比較したものである。

$$\text{RMSE} = \sqrt{\frac{1}{K} \sum \text{error}^2} \quad (18)$$

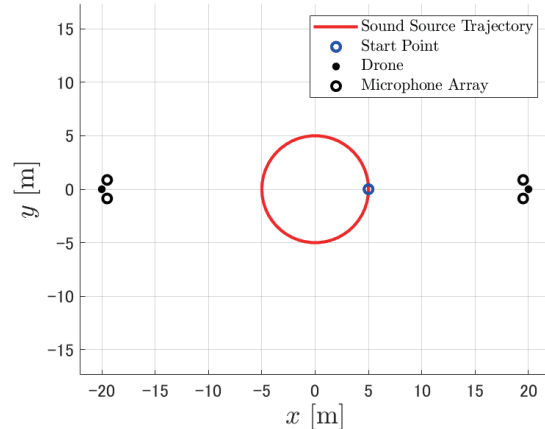


図 3: 上から見たシミュレーションシナリオ ($l = 20$ m)

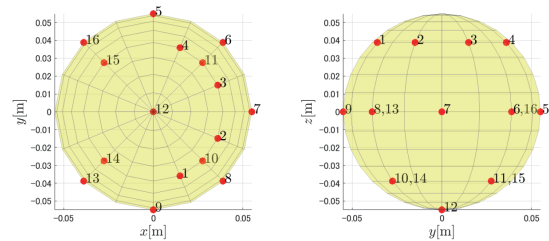


図 4: マイクロホンアレイ中のマイクロホン配置

ここで、 error は推定位置と真値のユークリッド誤差を指しており、 K は一シミュレーション中のタイムステップ数を示している。また、表2-4中の太字で示されている数値は、同じ水平距離において最も小さいRMSEを指している。表2-4より、提案手法は他手法と比べて最も小さい誤差で音源追跡を行ったことが分かる。これは提案手法は三角測量点を求めるのではなく、位置尤度分布を推定しようとするため、三角測量の離散性より生じる外れ値の影響を受けにくくなっているからだと考えられる。また、他手法と同じく、音源距離が長くなるにつれてRMSEが増加しているが、これは方向尤度に分解能があるため、位置尤度分布にも離散性が生じるからである。

4 終わりに

本稿では、音源の方向尤度分布を位置尤度として、統合することで音源追跡を行う手法を提案した。音源位置尤度分布は各マイクロホンアレイから得られる方向尤度分布を足し合わせることで得られる。音源位置尤度分布を元に音源追跡を行うことで、音源方向推定誤差や三角測量誤差が発生しやすいケースに対してロバストな追跡が行えることが期待される。提案手法を数

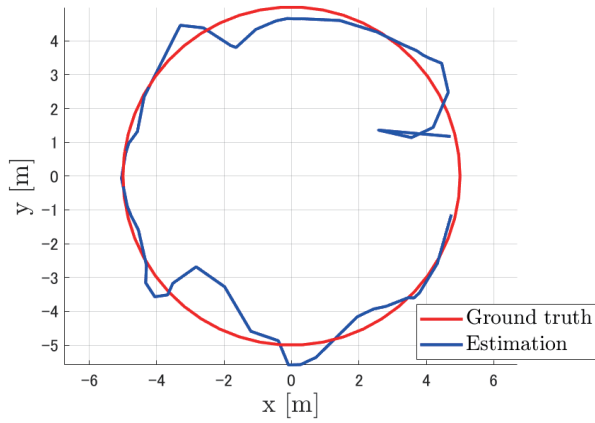


図 5: 上から見た推定軌跡

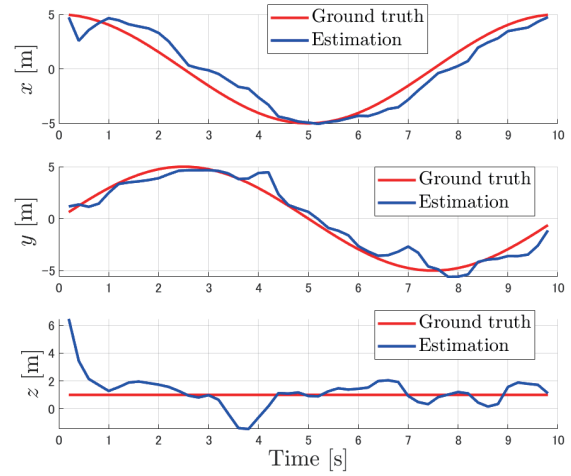


図 6: 各軸における推定結果

表 2: 各音源水平距離の追跡誤差 (RMSE) (女声)

	10 m	20 m	30 m	40 m	50 m
Proposed method	1.66	2.43	1.75	2.89	5.61
Method [8]	12.47	13.76	14.15	14.68	16.43
Method [9]	4.97	4.01	3.85	4.30	5.69

表 3: 各音源水平距離の追跡誤差 (RMSE) (男声)

	10 m	20 m	30 m	40 m	50 m
Proposed method	4.80	4.10	4.18	3.92	4.32
Method [8]	12.47	13.76	14.34	15.17	16.27
Method [9]	4.77	4.20	3.98	4.26	5.32

表 4: 各音源水平距離の追跡誤差 (RMSE) (ホワイトノイズ)

	10 m	20 m	30 m	40 m	50 m
Proposed method	1.81	2.39	1.93	2.26	2.95
Method [8]	8.73	10.57	10.87	11.16	11.32
Method [9]	2.14	1.81	2.14	2.55	3.66

値シミュレーションにより評価した結果、提案手法は40m離れた音源に対して4m以下の誤差で追跡できることがわかった。実環境での評価や、音源分離・音源認識を加えた音環境理解システムへの拡張は今後の課題である。

謝辞

科研費 JP19K12017, JP19KK0260 および JP20H00475 の助成を受けた。

参考文献

- [1] Nakadai, K. et al. (2017). Development of microphone-array-embedded UAV for search and rescue task, 2017 IEEE International Conference on Intelligent Robots and Systems (IROS), pp. 5985-5990
- [2] Washizaki, K., Wakabayashi, M., and Kumon, M. (2016). Position estimation of sound source on ground by multirotor helicopter with microphone array, 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1980-1985
- [3] Brandstein, M. S., and Silverman, H. F. (1997). A practical methodology for speech source localization with microphone arrays, *Computer speech & language*, 11(2), 91-126.
- [4] Gabriel, D., Kojima, R., Hoshiya, K., Itoyama, K., Nishida, K., and Nakadai, K. (2019). 2D sound source position estimation using microphone arrays and its application to a VR-based bird song analysis system, *Advanced Robotics*, 33(7-8), pp. 403-414.
- [5] Yamada, T., Itoyama, K., Nishida, K., and Nakadai, K. (2020). Sound Source Tracking by Drones with Microphone Arrays, *IEEE/SICE International Symposium on System Integration (SII)*, pp. 796-801
- [6] Knapp, C., and Carter, G. (1976). The generalized correlation method for estimation of time delay, *IEEE transactions on acoustics, speech, and signal processing*, 24(4), pp. 320-327.
- [7] Valin, J. M., Michaud, F., and Rouat, J. (2006). Robust 3D localization and tracking of sound sources using beamforming and particle filtering, 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings (Vol. 4, pp. IV-IV).
- [8] Potamitis, I., Chen, H., and Tremoulis, G. (2004). Tracking of multiple moving speakers with multiple microphone arrays, *IEEE Transactions on Speech and Audio Processing*, 12(5), pp. 520-529.
- [9] Lauzon, J. S., Grondin, F., Létourneau, D., Desbiens, A. L., and Michaud, F. (2017). Localization of RW-UAVs using particle filtering over distributed microphone arrays, 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 2479-2484)
- [10] Schmidt, R. (1986). Multiple emitter location and signal parameter estimation, *IEEE transactions on antennas and propagation*, 34(3), pp. 276-280.