

Multicopter UAV を用いた音源定位のための雑音相関行列推定

古川孝太郎 大塚琢馬 (京都大学)

糸山克寿 (京都大学) 中臺一博 ((株) HRI-JP) 奥乃博 (京都大学)

1. 序論

Multicopter UAV (Unmanned Aerial Vehicle) は、機敏な動作および自在性の高い運用が可能であることから、環境センシングのためのユニバーサルプラットフォームとして有効である。Multicopter UAV を用いたセンシングについての近年の研究 [1-3] は主として画像情報を取り扱うが、障害物による対象の遮蔽に頑健ではないという問題がある。

無人航空機を用いて音源を検出できれば、図 1 に示されるような、障害物により遮蔽された場所にいる人や動物の発見が可能となる。このため音を用いたセンシングは捜索救難、環境や災害の監視また警備などの応用に有効と考えられる [4, 5]。そこで本研究は機体に搭載したマイクロフォンアレイにより音を収集し、周囲の環境中の音源位置特定のために音源定位、すなわち音の到来方向を推定する手法を開発する。

Multicopter UAV を用いる上での主な課題は、飛行時に発生する非定常でパワーの強い自己雑音による定位性能の劣化である。自己雑音は主としてロータの回転によって起こるもので、機体上のマイクロフォンを雑音源から一定以上遠ざけることは困難である。また飛行中の機体の制御に伴って各ロータの回転速度は変化するため、雑音も非定常なものとなる。

GEVD-MUSIC (Generalized Eigenvalue Decomposition-based Multiple Signal Classification) [6] は、雑音を含む観測からの音源方向推定に広く用いられる音源定位アルゴリズムである。GEVD-MUSIC は雑音についての空間相関行列 (以下、雑音相関行列) を用いて観測信号中の雑音成分を抑圧し、音源方向を表す空間スペクトルを生成する。ここで雑音相関行列は時々刻々と変化するから、非定常な雑音環境下での GEVD-MUSIC の適用のためには各時刻における雑音相関行列の推定が最大の課題となる。

GEVD-MUSIC による音源定位は、雑音相関行列を予め推定する必要があるため、動的な雑音への対応に問題があった。この問題を解消するため、次のようなアルゴリズム拡張が報告されている。奥谷らによって提案された iGEVD-MUSIC [7] は直前の観測を雑音とみなして雑音相関行列を生成することによって、動的雑音に対処する手法であるが、目的音が自己雑音より大きく変化するという仮定に基づいているため、定常な目的音が誤って雑音とみなされるおそれがある。また Ince らの提案する手法 [8] は、ヒューマノイドロボットの自己動作音を雑音テンプレートを組み合わせることで動的に推定し、音源定位における動的な雑音抑圧を実現している。しかし、動作雑音の推定に最近傍法を用いているため、雑音テンプレートに存在しない動作雑音の推定精度に問題がある。

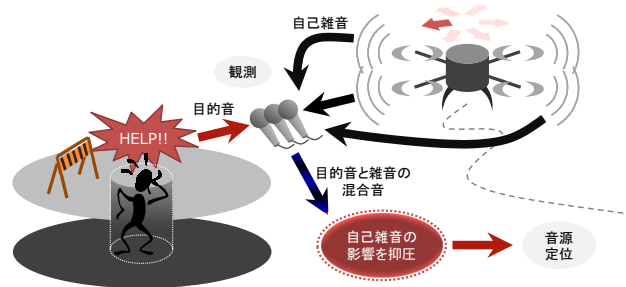


図 1: Multicopter UAV 飛行時の自己雑音による定位性能劣化の回避

本研究では Gauss 過程 [9] を用いて UAV に搭載された機体センサのモニタリング値から雑音相関行列を回帰する手法を開発する。本手法では雑音相関行列の推定に観測音を用いないので iGEVD-MUSIC のように定常的な目的音を誤って雑音とみなすことがない。また学習データ中に現れないモニタリング値に対しても適切な雑音相関行列を内挿によって推定できる。

本手法は図 2 に示されるように、雑音相関行列の回帰および得られた雑音相関行列を用いた GEVD-MUSIC による音源定位の二段階の処理から構成される。以下では、これらの処理について説明する。

2. 雑音相関行列推定

本研究で取り扱う雑音相関行列は機体の状態や制御入力と複雑に関係しておりパラメトリックなモデルを適用することが困難であるため、Gauss 過程による回帰によって推定する。Gauss 過程は平均関数 m とカーネル k によって特徴付けられる連続空間上の確率過程であり、特徴量と目的変数を事前に収集したデータから、値が Gauss 分布に従うという仮定に基づいて、新たに観測された特徴量についての適切な目的変数の値を与える。以下では特徴量である期待センサのモニタリング値を \mathbf{u}_t 、目的変数であるベクトル化された雑音相関行列を $\mathbf{v}_{t,f}$ と表すことにする。

$$\mathbf{u}_t = [u_{t,1}, \dots, u_{t,10}]^T, \quad (1)$$

$$\mathbf{v}_{t,f} = [v_{t,f,1}, \dots, v_{t,f,d_v}]^T. \quad (2)$$

ここで t, f はそれぞれ時間と周波数のインデックスであり、 M はマイクロフォンの数、 $d_v = M(M+1)/2$ である。各特徴量の値は

- $u_{t,1}, \dots, u_{t,3}$: Pitch, Roll, Yaw
- $u_{t,4}, \dots, u_{t,6}$: xyz 各軸の速度
- $u_{t,7}, \dots, u_{t,10}$: パルス幅変調信号

Pitch, Roll, Yaw とはそれぞれ機首の上下方向の角度、左右の傾きの角度、機首の方向である。パルス幅変調信号とはモータを制御する信号であり、パルス長のパルス周

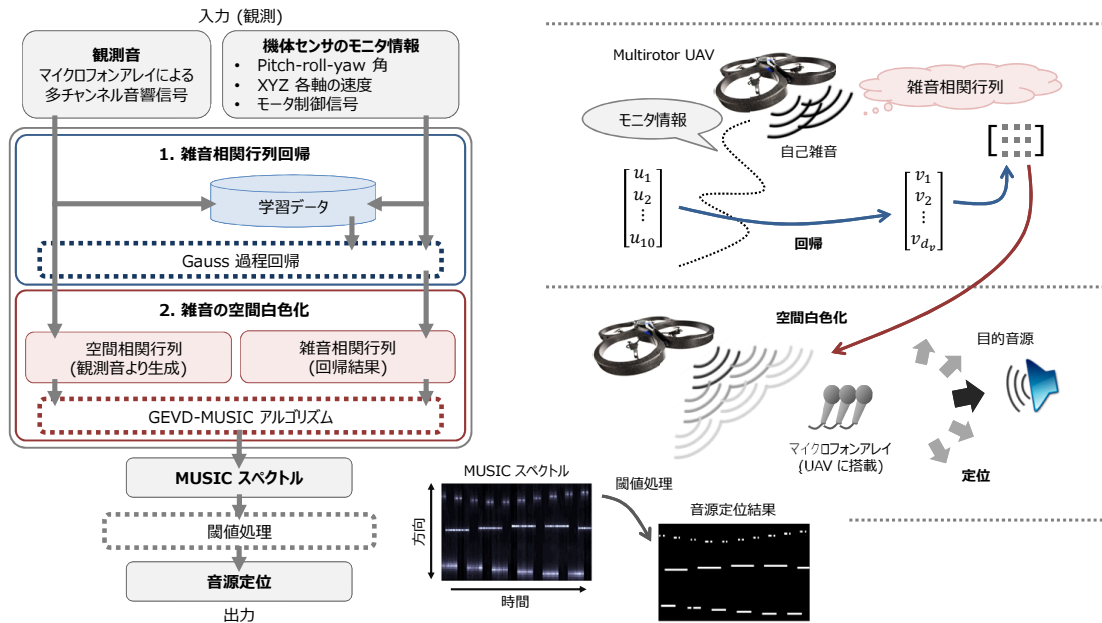


図 2: Gauss 過程による雑音相関行列回帰と GEVD-MUSIC による音源定位の概要

期に対する割合として得られるためモータに供給される電力を表している。

雑音相関行列 $\mathbf{Q}_{t,f}$ を Gauss 過程によって扱うことためベクトル化する。雑音相関行列は Hermite 行列であり、冗長性を減らすため Cholesky 分解により上三角行列 $\mathbf{L}_{t,f}$ を求める。得られた上三角行列の要素を連結してベクトルとする。ただし方向性雑音のない環境での雑音相関行列が単位行列であることから、平均が単位行列となるようにする。具体的には、以下のようにベクトル $\mathbf{v}_{t,f}$ の要素を求める。

$$\mathbf{L}_{t,f} = \begin{bmatrix} v_{t,f,1}^2 & v_{t,f,2} & \cdots & v_{t,f,d_v-(M-1)} \\ & v_{t,f,3}^2 & \cdots & v_{t,f,d_v-(M-2)} \\ & & \ddots & \vdots \\ 0 & & & v_{t,f,d_v}^2 \end{bmatrix} - \mathbf{I}_M. \quad (3)$$

ここで $M \times M$ 単位行列を \mathbf{I}_M として表す

特徴量の新たな観測 \mathbf{u}_{T+1} についての目的変数 $\mathbf{v}_{T+1,f}$ は、 \mathbf{u}_{T+1} と学習データ \mathcal{D} による条件付き分布 $p(\mathbf{v}_{T+1}|\mathcal{D}, \mathbf{u}_{T+1})$ の平均として求められる。学習データ \mathcal{D} は目的音の存在しない状況で、事前に収集した特徴量のモニタリング値と目的変数の雑音相関行列の対の集合である。ゆえに、平均 $m(\mathbf{v}_{T+1,f})$ は以下のように求められる。

$$m(\mathbf{v}_{T+1,f}) = \mathbf{k}_T^T (\mathbf{K}_T + \rho^2 \mathbf{I}_T)^{-1} \mathbf{V}_{T,f}^T, \quad (4)$$

$$\mathbf{V}_{T,f} = (\mathbf{v}_{1,f}, \dots, \mathbf{v}_{T,f}). \quad (5)$$

ρ^2 は特徴量に付加していると仮定する加法的雑音の分散でありグリッド探索などにより最適な値を求める。また \mathbf{k}_T , \mathbf{K}_T はそれぞれ以下で表される。

$$\mathbf{k}_T = \begin{bmatrix} k_{1,T+1} \\ \vdots \\ k_{T,T+1} \end{bmatrix}, \quad \mathbf{K}_T = \begin{bmatrix} k_{1,1} & \cdots & k_{1,T} \\ \vdots & \ddots & \vdots \\ k_{T,1} & \cdots & k_{T,T} \end{bmatrix}. \quad (6)$$

ただし $k(\mathbf{u}_i, \mathbf{u}_j)$ を $k_{i,j}$ と表している。ここでカーネルとして、RBF (Radial Basis Function) カーネルをもとにした Mahalanobis カーネル [10] を用いている。

$$k(\mathbf{u}_i, \mathbf{u}_j) = \exp\left(-\frac{\gamma(\mathbf{u}_i - \mathbf{u}_j)^T \boldsymbol{\Sigma}^{-1}(\mathbf{u}_i - \mathbf{u}_j)}{\dim(\mathbf{u}_i)}\right). \quad (7)$$

ただし $\boldsymbol{\Sigma}^{-1}$ は特徴量の各次元の分散を対角成分に持つ行列である。

3. 音源定位

本研究で用いる音源定位アルゴリズム GEVD-MUSIC は、方向性雑音環境下においても雑音についての追加の情報である雑音相関行列を与えることによってその影響を抑圧し音源方向を推定する。出力は目的音の到来方向にピーク値を持つようなスペクトルであり、真の音源方向を θ_i ($i = 1, \dots, N$) として以下の $\mathbf{p}_{t,f}$ で表される。

$$\mathbf{p}_{t,f} = [p_{t,f,\theta_1}, \dots, p_{t,f,\theta_A}]^T, \quad (8)$$

$$p_{t,f,\theta_i} = \frac{\|\mathbf{a}_{f,\theta_i}^H \mathbf{a}_{f,\theta_i}\|}{\sum_{i=N+1}^M |\mathbf{a}_{f,\theta_i}^H \mathbf{e}_{\mathbf{R},i}|^2}. \quad (9)$$

ここで変数右肩の H は共役転置を表す。 \mathbf{a}_{f,θ_i} はマイクロフォンアレイ配置についての事前知識であるステアリングベクトル、 A はその数で定位可能な方向の数を表す任意の値である。この定義により、音源方向とステアリングベクトルの方向が近い場合には分母が 0 へと近づき、スペクトル中のその方向にピークが立つ。ただし目的音は一般に広帯域に広がることが考えられるため、このスペクトルの加重和をとって出力とする。

$$\mathbf{p}_t = \sum_f w_{t,f} \mathbf{p}_{t,f}. \quad (10)$$

重み $w_{t,f}$ には $\mathbf{R}_{t,f}$ の跡を用いた。

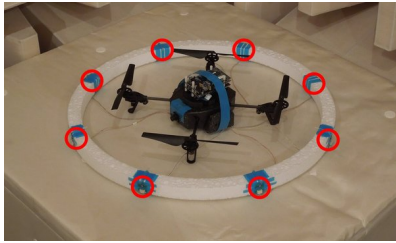


図 3: マイクロフォンアレイを搭載した録音システム (丸印にマイク)

4. 実験と評価

本手法の評価のためにいくつかの実験を行った。この実験では入力として観測音の多チャンネル音響信号と機体センサのモニタリング値を与え、出力として音源の推定方向を表す空間スペクトルを得た。

4.1 音源定位システムの構築

Multicopter UAV 上にマイクロフォンアレイを搭載して図 3 のような音源定位システムを構築した。Multicopter UAV としては、6 自由度の慣性計測装置や超音波距離計、対地速度センサなどを備えた AR.Drone¹ を用いた。この AR.Drone 上に 8 チャンネルマイクロフォンアレイと多チャンネル同期音響処理装置である RASP-24² を搭載して録音モジュールを構築した。各マイクロフォンは円形のフレーム上で等間隔に、外向きに配置した。またこれらの機器搭載のために、機体本体の重量を削減してペイロードを増加させた。

録音用のソフトウェア環境として、HARK³ および ROS⁴ を用いた。HARK はロボット聴覚のためのオープンソースソフトウェア群であり、多チャンネル音響信号の同期録音やその ROS トピックとしての発行などに用いた。ROS はロボット制御のためのプラットフォームであり、AR.Drone の操縦や各モジュールで観測した情報の同期とデータ形式の統一などに用いた。

4.2 実験条件

実験データとして AR.Drone のホバリング中約 200 秒間および移動中 400 秒間について、無響室で自己雑音を収録した。そのうちおよそ五分の一をテストのために使い、残りを学習データとした。テストデータは上記 AR.Drone のホバリング中と移動中のそれぞれの場合について、この自己雑音データと音声、純音、白色雑音による目的音のシミュレーション混合によって生成した。マイクロフォンアレイの配置についての事前知識として、TSP (Time-Stretched Pulse) 応答を収録することにより周囲 5 度おきに 72 方向のステアリングベクトルを求めた。

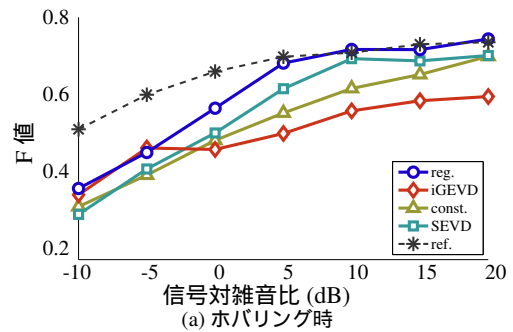
音響信号の標本化は 16kHz、モニタリング値の取得頻度はおよそ 60 ミリ秒に 1 度である。時間領域の音響信号に対しフレーム長 512、シフト長 256、Hann 窓を用いた短時間 Fourier 変換を適用して、時間周波数表現を得た。

¹<http://ardrone.parrot.com>

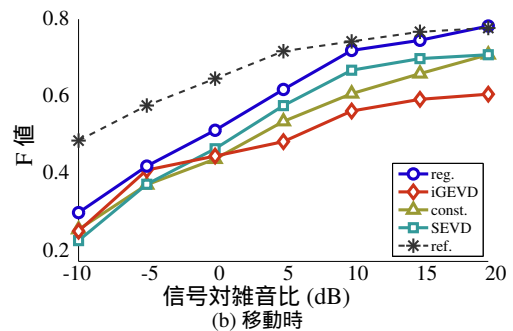
²<http://www.sifi.co.jp>

³<http://winnie.kuis.kyoto-u.ac.jp/HARK/>

⁴<http://www.ros.org>



(a) ホバリング時



(b) 移動時

図 4: 適合率・再現率の等しい点における F 値

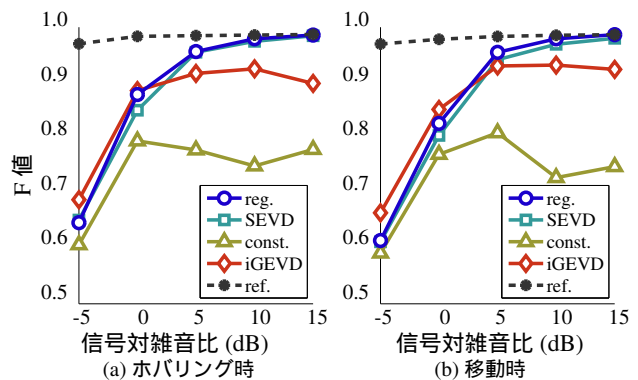


図 5: VBHMM に基づく閾値決定法を用いた場合の F 値

4.3 シミュレーションデータに対する適用

上記のデータを用いて本手法と既存手法による定位性能を比較した。比較は本手法 (図中 reg.) と 3 種類の既存手法、および真の雑音を用いた場合 (ref.) について行った。既存手法は、標準固有値分解を用いた MUSIC 法 (SEVD)、時間平均による一定の雑音を仮定した GEVD-MUSIC 法 (const.), iGEVD-MUSIC 法 (iGEVD) である。

評価は 2 つの基準によって行った。まず各時刻における空間スペクトル \mathbf{p}_t を列ベクトルに持つ \mathbf{P} に対して、閾値 ξ を設定して以下の式で適合率 Pre と再現率 Rec を求め、その等しい点における適合率と再現率の調和平均である F 値によって比較する方法である。

$$\text{Pre}(\mathbf{P}) = \frac{\#\{(t, \theta) \mid p_{t,\theta} \geq \xi \text{ and } p'_{t,\theta} = 1\}}{\#\{(t, \theta) \mid p_{t,\theta} \geq \xi\}}, \quad (11)$$

$$\text{Rec}(\mathbf{P}) = \frac{\#\{(t, \theta) \mid p_{t,\theta} \geq \xi \text{ and } p'_{t,\theta} = 1\}}{\#\{(t, \theta) \mid p'_{t,\theta} = 1\}}. \quad (12)$$

ここで # は集合の要素数を表し、 $p'_{t,\theta}$ は正解データで

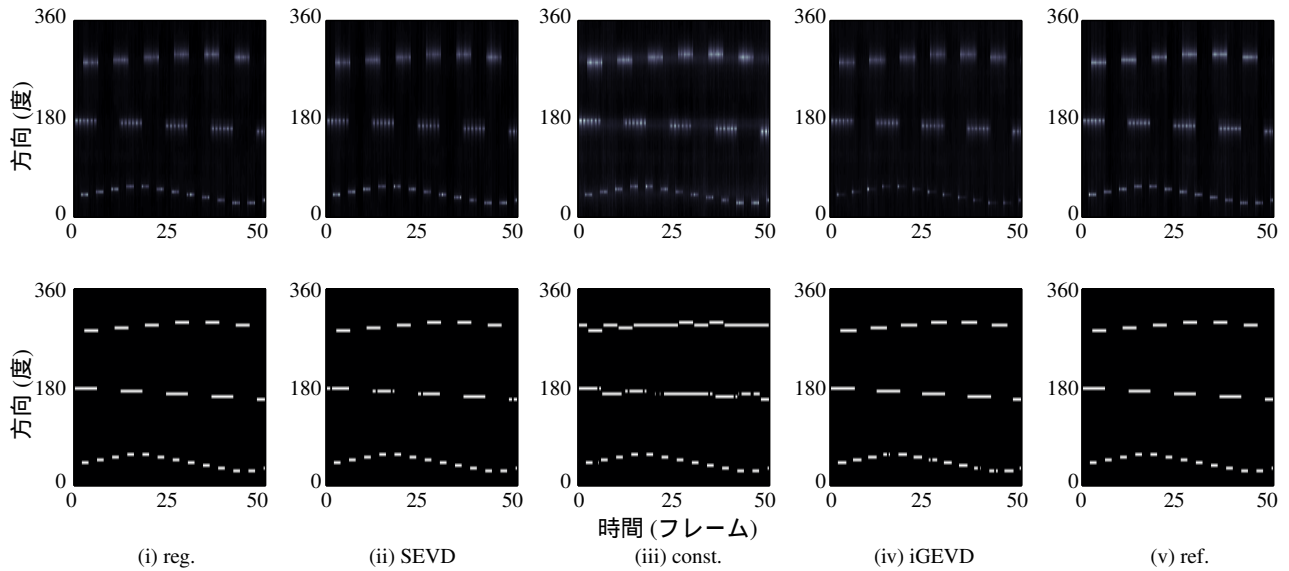


図 6: (上段) 空間スペクトル (下段) 閾値処理後の音源定位結果

あり, 真の音源方向に値 1 を持つスペクトルである. もう一方は VBHMM (Variational Bayesian Hidden Markov Model) に基づいた閾値決定法 [11] を用い, F 値によって比較する手法である. 評価結果を図 4, 5 に示す.

図 4 から本手法が既存手法よりも高い値をとるため, 空間スペクトルに鮮明なピークが立っていると確認できる. 図 5 中の VBHMM に基づく閾値決定によって得られた定位結果の F 値については, 既存手法よりも本手法がやや優れていることを示しているものの, その差は僅かとなっている. また, 図 6 は各手法, 条件における実際の空間スペクトルの出力と閾値処理後の音源定位結果である.

4.4 考察

まず上記の結果の信号対雑音比が高い条件下において, 本手法によるスペクトルが既存手法よりも鮮明である理由については以下のような推測が可能である. すなわち観測音を用いて雑音相関行列を推定する既存手法は, 目的音が顕著であるほどそれを雑音として誤抑圧した場合の影響が強く現れるのに対し, 本手法では観測音を雑音相関行列推定に用いないためこの悪影響が現れなかったと考えられる. 一方で信号対雑音比の低い, すなわち目的音のパワーの小さい条件下においては, いずれの評価基準についても, 真の雑音を用いた場合のスペクトルの鮮明さに対して大きく下回る結果となった. 真の雑音を用いた結果は, GEVD-MUSIC 法に基づく音源定位手法の性能の上限であり, この値へ近いほど雑音相関行列の推定精度が高いとみなせる. すなわち本手法による雑音相関行列の推定精度は依然として高いとは言えない. この推定精度の改善のために, 学習データとしてより多様なモニタリング値の雑音相関行列を収集することや, 特徴選択の最適化などが考えられる.

5. 結論

本稿では Multirotor UAV に搭載したマイクロフォンアレイを用いた音源定位性能の向上について述べた.

Multirotor UAV を用いる問題点は飛行中に非定常でパワーの強い自己雑音が発生することである. この問題に対し Gauss 過程を用いて機体のモニタリング値から雑音相関行列を回帰し, その回帰結果を GEVD-MUSIC に適用することで自己雑音の影響を緩和する手法を提案した.

実験結果は本手法により特に信号対雑音比の高い場合に定位性能が向上したことを示した. 今後の課題として, 特徴選択や学習データの増加によって雑音相関行列の回帰精度を向上するほか, 実環境下での収録データによる評価が必要と考える.

謝辞

本研究の一部は科研費 24220006 の補助を受けた.

参考文献

- [1] L. Meier *et al.*, PIXHAWK: A system for autonomous flight using onboard computer vision, *IEEE ICRA*, 2011, pp. 2992–2997.
- [2] M. W. Achtelik *et al.*, Visual-Inertial SLAM for a Small Helicopter in Large Outdoor Environments, *IEEE/RSJ IROS*, 2012, pp. 2651–2652.
- [3] A. Natraj *et al.*, A Geometrical Approach for Vision Based Attitude and Altitude Estimation for UAVs in Dark Environments, *IEEE/RSJ IROS*, 2012, pp. 4565–4570.
- [4] M. Basiri *et al.*, Robust Acoustic Source Localization of Emergency Signals from Micro Air Vehicle, *IEEE/RSJ IROS*, 2012, pp. 4737–4742.
- [5] H. Yoshinaga *et al.*, A sound source localization technique to support search and rescue in loud noise environments, *Journal of Japan Society of Civil Engineers* 67 (2012) 11–16.
- [6] K. Nakamura *et al.*, Intelligent sound source localization for dynamic environments, *IEEE/RSJ IROS*, 2009, pp. 664–669.
- [7] K. Okutani *et al.*, Outdoor Auditory Scene Analysis Using a Moving Microphone Array Embedded in a Quadcopter, *IEEE/RSJ IROS*, 2012, pp. 3288–3293.
- [8] G. Ince *et al.*, Assessment of general applicability of ego noise estimation, *IEEE ICRA*, 2011, pp. 3517–3522.
- [9] C. E. Rasmussen, C. K. I. Williams, *Gaussian Processes for Machine Learning*, Adaptive Computation and Machine Learning, MIT Press, Cambridge, MA, 2006.
- [10] S. Abe, Training of support vector machines with Mahalanobis kernels, *Artificial Neural Networks: Formal Models and Their Applications (ICANN 2005)*, Vol. 3697, 2005, pp. 571–576.
- [11] T. Otsuka *et al.*, Bayesian Extension of MUSIC for Sound Source Localization and Tracking, *Interspeech*, 2011, pp. 3109–3112.