

# ドローン聴覚におけるヒストグラム情報を用いた 音源定位手法の提案 -周波数抽出とスケージングの導入による性能向上-

## Proposal of sound source localization method for drone audition using histogram -Improvement of performance by introducing frequency extraction and scaling-

小松崎和泉<sup>1\*</sup> 干場功太郎<sup>1</sup> 岩附信行<sup>1</sup>  
Izumi Komatsuzaki<sup>1</sup> Kotaro Hoshiba<sup>1</sup> Nobuyuki Iwatsuki<sup>1</sup>

<sup>1</sup> 東京工業大学

<sup>1</sup> Tokyo Institute of Technology

**Abstract:** ドローンをを用いた被災者捜索のための音源探査技術において、これまで、著しい時刻変化を伴うドローン自身のエゴノイズに対する耐性、広い搜索範囲、低い計算コスト、汎用性をすべて満たす音源定位手法の開発を目的に、ヒストグラム情報と周波数情報を用いて空間スペクトルにおけるエゴノイズの除去を動的に行う音源定位手法の提案を行った。本稿では、これまでの提案手法におけるリアルタイム性および定位精度をより向上するために、目標音成分の存在する周波数を抽出し、周波数毎に適した基準値を設けてエゴノイズを除去することにより、エゴノイズ近傍を含む広い範囲に存在する目標音に対して、より正確な定位が可能となるよう提案手法の改良を行った。実環境での屋外実験とシミュレーションにより、提案手法の性能を評価した結果、本稿で紹介する改良した提案手法を用いることで、高いリアルタイム性を持ちながら、より高いノイズ耐性と広い搜索可能範囲を獲得することができ、本手法の有用性が確認された。

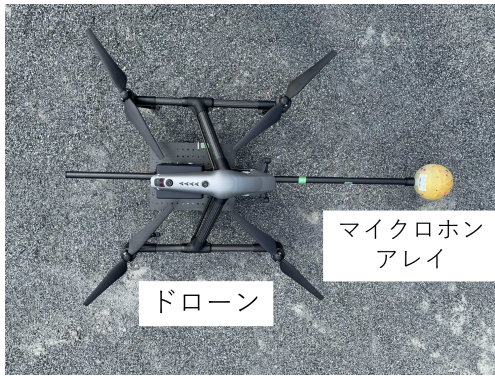
## 1 はじめに

近年、災害地において、人が侵入できない場所にも容易に侵入できること、迅速な活動ができることから、ドローンをを用いた要救助者の搜索手法が注目されている。ドローンをを用いた搜索では、カメラによる方法が一般的であるが [1]、暗い時間帯の搜索活動、および瓦礫等に埋もれた被災者といったカメラに映らない対象の搜索は困難である。そこで、音情報による搜索手法の確立を目的に、ドローン搭載マイクロホンアレイを用いた音源探査技術の開発が進められている。このような技術はドローン聴覚と呼ばれ、様々な研究が行われている [2][3]。ドローンをを用いた音源探査の実用化にあたり、課題の一つがドローンのエゴノイズである。風や飛行状態の影響により著しい時刻変化を伴うエゴノイズに対する耐性、広い搜索範囲、ドローン搭載の小型コンピュータを用いて実時間で搜索を行うためのリアルタイム性、どのような機体・状況でも搜索可能な高い汎用性を持った音源探査手法が求められる。

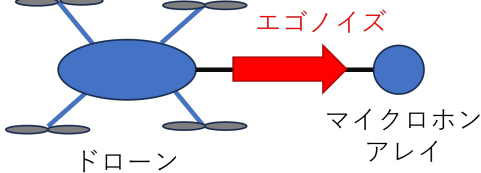
これまで、音源探査手法として、その分解能の高さから MUSIC (MUltiple SIgnal Classification) 法 [4] に

基づく音源定位手法が提案されてきている。一般的な MUSIC 法である SEVD-MUSIC (MUSIC based on Standard Eigen Value Decomposition) 法は、計算コストが低くリアルタイム性が高い反面、ノイズ耐性が低い。そこで、事前収録したノイズの相関行列を用いてノイズ成分を除去する GEVD-MUSIC (MUSIC based on Generalized Eigen Value Decomposition)[5] や GSVD-MUSIC (MUSIC based on Generalized Singular Value Decomposition)[6] が提案された。これらは、SNR (Signal-to-Noise Ratio) の低い状況でも高い音源定位性能を持つが、計算コストが高く、事前に収録したノイズ情報を用いているため時刻変化するノイズへの耐性および汎用性がない。時刻変化するノイズへの耐性の強化および汎用性を補うことを目的に、直前の時刻の収録音をノイズと仮定してノイズ除去を行う iGEVD-MUSIC (incremental GEVD-MUSIC)[7] や iGSVD-MUSIC (incremental GSVD-MUSIC)[8] が提案されている。しかし、これらは、計算コストが高いことに加え、著しく時刻変化するノイズに対する耐性は不十分である。AFRF-MUSIC (MUSIC using Active Frequency ange Filter)[9] は、直前の時刻の収録音を用いて解析周波数を制限することで、小さい計算コストでありながら、エゴノイズを抑制することに成功し

\*連絡先：東京工業大学 工学院 機械系  
〒152-8552 東京都目黒区大岡山 2-12-1 11-27  
E-mail: komatsuzaki.i.aa@m.titech.ac.jp



(a) マイクロホンアレイ搭載ドローンの一例.



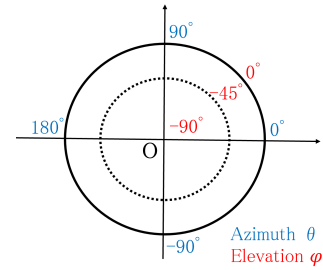
(b) マイクロホンアレイに対するエゴノイズの到来方向.

図 1: 一定の方向からエゴノイズが到来する配置のマイクロホンアレイ搭載ドローン.

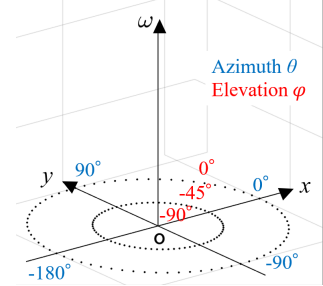
た. しかし, こちらも著しく時刻変化するノイズへの耐性は不十分である. また, 角度制限 SEVD-MUSIC[10] は, 一定の方向からエゴノイズが到来するようなマイクロホンアレイを設計し, SEVD-MUSIC にて得られた空間スペクトルに対し, 目標音の搜索範囲からエゴノイズの到来範囲を事前に除外することでノイズ耐性を向上させる. しかし, 除外範囲をあらかじめ設定する必要があり, 狭い範囲を除外すると時刻変化を伴うノイズに対する十分な耐性が見込めず, 広い範囲を除外すると搜索範囲が狭まる.

これらの問題を解決するため, これまでに, 過去の情報を用いず, 得られた現在時刻の空間スペクトルから, ヒストグラム情報と周波数情報に基づきエゴノイズ成分の判定を行い, 動的に搜索範囲の制限による目標音成分の抽出を行う手法である HIST-MUSIC (MUSIC with HISTogram information)[13] および HIST-MUSIC-3D (three-dimensional HIST-MUSIC)[14] を提案した. これにより, 高いノイズ耐性, 広い搜索範囲, リアルタイム性, 汎用性をすべて満たすことができた. しかし, 被災者搜索にて活躍するためには, リアルタイム性およびノイズと近傍の方向にある目標音に対する目標音成分の抽出精度が不十分であり, 更なる向上が必要と考えた.

本稿では, 著しい時刻変化を伴うノイズに対する耐性, 広い搜索範囲, 高いリアルタイム性, 高い汎用性をすべて満たす音源定位手法の開発を目的に, HIST-MUSIC-3D の改良を行い, より精度の高い空間スペクトルからの目標音成分抽出手法を提案する. 本手法では, 得られた現在時刻の空間スペクトルから, ヒストグラム情報と周波数情報に基づき, 目標音成分の存在



(a) 二次元空間スペクトルの座標設定.



(b) 三次元空間スペクトルの座標設定.

図 2: 座標設定.

する周波数の抽出と, 各周波数で最適なノイズ判定の基準値の設定を行うことで, 目標音成分の抽出を行う. 周波数抽出により, 計算コストの大きいノイズ判定の処理が行われる周波数が制限されること, また, 周波数毎のノイズ判定の基準値の設定により, 各周波数に存在する目標音成分を適切に抽出できることから, リアルタイム性を向上させながら, 目標音成分の抽出精度も向上させることができると期待される. 本稿では, 提案手法のノイズ耐性, 搜索範囲, リアルタイム性および汎用性について, シミュレーションおよび屋外実験により評価する.

## 2 HIST-MUSIC-3D

これまでに, 現在時刻の空間スペクトルのヒストグラム情報と周波数情報のみからノイズの判定を行い, 動的に探索範囲の制限を行う HIST-MUSIC-3D を提案した [14]. 空間スペクトルは, 計算コストが低い SEVD-MUSIC[4] により得られる MUSIC スペクトルとする. また, これまでに図 1 のようなマイクロホンアレイ搭載ドローンの構造が開発されており, 本論文では, 一定の方向からエゴノイズが到来するマイクロホンアレイの使用を前提に考える. 以下に HIST-MUSIC-3D のアルゴリズムを示す.

SEVD-MUSIC 法では一般的に, 周波数毎に算出した空間スペクトル  $P(\psi, \omega)$  を周波数方向に加算した二次元空間スペクトル  $\bar{P}(\psi)$  を解析に用いる.

$$\bar{P}(\psi) = \sum_{\omega} P(\psi, \omega) \quad (1)$$

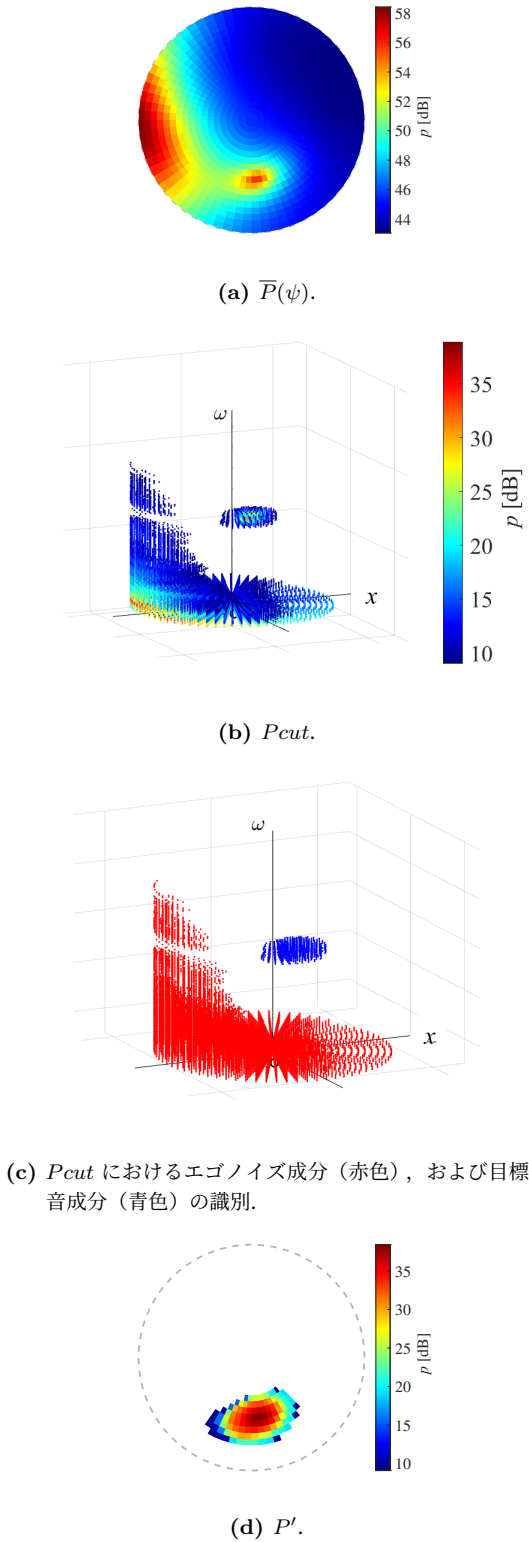


図 3: HIST-MUSIC-3D における目標音成分抽出の過程.

$\omega$  とは周波数ビンを表し,  $\psi$  はマイクロホンアレイに対する方位角  $\theta$ , 仰角  $\phi$  から,  $\psi = (\theta, \phi)$  と定義する. 二次元空間スペクトルの一例を図 3a に示す. ここでは, 図 2a の設定軸に従ってプロットされており, 各方向から到来した音のパワーをカラーマップで示している. 図 3a の目標音方向は, 方位角  $-90^\circ$ , 仰角  $-45^\circ$  である. 本手法では, 周波数方向に加算する前の周波数情報を

含む三次元空間スペクトル  $P(\psi, \omega)$  に着目する.

一般的に空間スペクトルのピーク検出により音源定位を行うが, 目標音成分の最大パワーより大きいエゴノイズ成分がある場合, 正確な目標音源の定位ができない. そこで, 目標音成分の最大パワーより小さくかつ近傍な基準値を設定し, 基準値より大きいエゴノイズ成分をピーク検出範囲から除外することで, 搜索範囲が最大かつ正確な定位が可能になる [13]. 基準値として, 空間スペクトル  $P(\psi)$  の全要素のヒストグラム  $H$  を求める.

$$H(p) = \text{histogram}(P(\psi)) \quad (2)$$

ここで,  $p$  は空間スペクトルのパワーに対する階級である. 得られた  $H$  のピーク以降の変曲点を基準値  $p_t$  とする [13].

$$p_t = p |_{H''(p)=0} \quad (3)$$

各周波数ビンの空間スペクトル  $P(\psi, \omega)$  に対して,  $p_t$  以上のパワーを持つ範囲  $\Psi_{cut}(\omega)$  を抽出する.

$$\Psi_{cut}(\omega) = \{\psi | P(\psi, \omega) > p_t\} \quad (4)$$

$\Psi_{cut}(\omega)$  に含まれる三次元の空間スペクトルの一例を図 3b に示す. ここでは, 図 2b の設定軸に従ってプロットされている. エゴノイズは一定方向より到来することから, 基準方向  $\psi_0 = (\theta_0, \phi_0)$  を設定し,  $\Psi_{cut}(\omega)$  に対して,  $\psi_0$  を含まない連続している部分を目標音成分  $\Psi_{target}(\omega)$  として抽出する.

$$\Psi_{target}(\omega) = \Psi_{cut}(\omega) \not\ni \psi_0 \quad (5)$$

それぞれの成分を図示したものが図 3c であり, 赤色がエゴノイズ成分, 青色が目標音成分を表す. そして, 得られた  $\Psi_{target}$  に対応する空間スペクトルを周波数方向に足し合わせ, 二次元の空間スペクトル  $\bar{P}'(\psi)$  を得る (3d).

$$\bar{P}'(\psi) = \sum_{\omega} P(\Psi_{target}, \omega) \quad (6)$$

$\bar{P}'(\psi)$  に対して, 最大値をとる方向を目標音方向  $\psi_{target}$  として検出する.

$$\psi_{target} = \text{argmax}_{\psi} \bar{P}'(\psi) \quad (7)$$

### 3 提案手法の改良

HIST-MUSIC-3D を用いた場合, すべての周波数ビンに対してノイズ除外操作を行っていたため, 目標音成分が存在していない周波数に対してもノイズ除外操作を行うといった余分な計算コストがかかっている問題点があった. また, すべての周波数ビンに対して等しい基準値でノイズ判定を行っていたため, 各周波数ごとの空間スペクトルの値に合ったノイズ判定の基準値でノイズの判定が行えていない問題点があった. その一例を図 4c に示す. 図 4a はある 1 つの周波数ビンの空間スペクトルである. 図 4b は, 図 4a に対して, すべての周波数で等しく設定された基準値以上の成分をプロットしたものであるが, 定位に影響のあるノイ

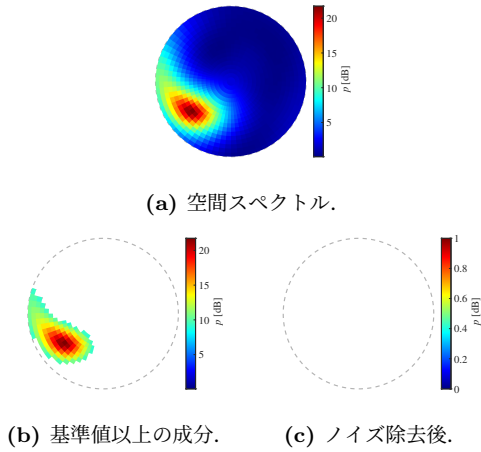


図 4: HIST-MUSIC-3D の失敗例.

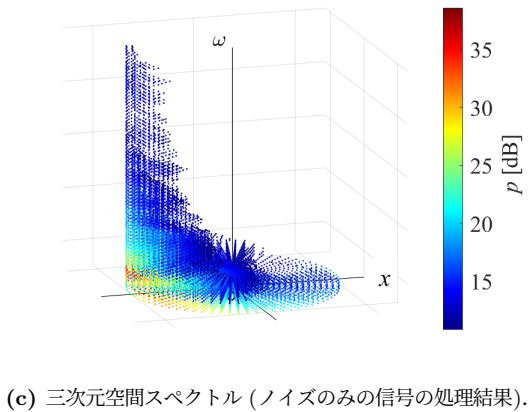
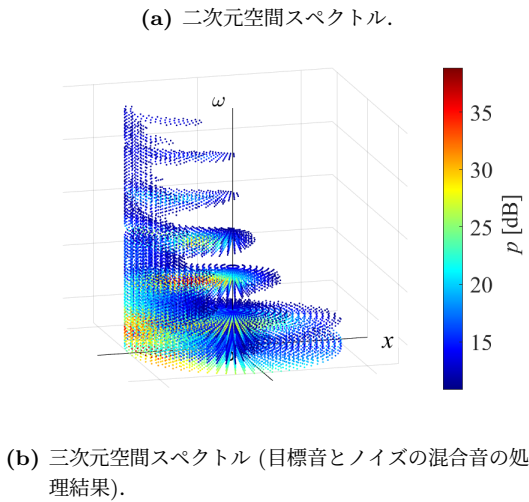
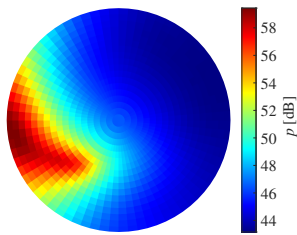


図 5: 空間スペクトルの一例.

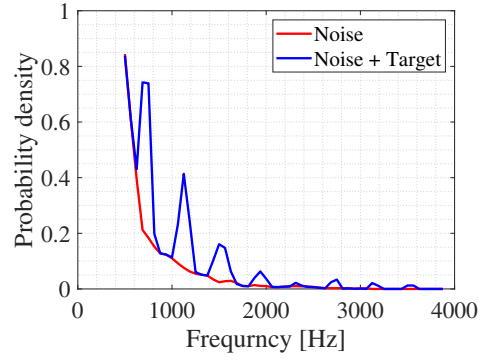


図 6: 全方向に対する  $\Psi_{cut}(\omega)$  が占める確立密度.

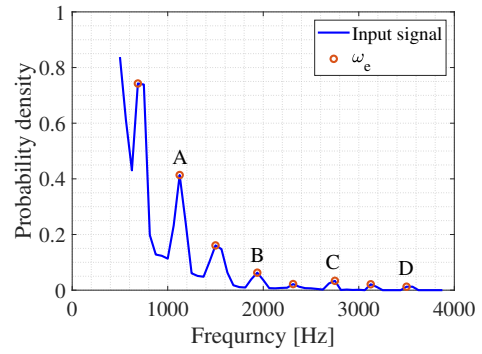


図 7:  $N$  と  $\omega_e$  の一例.

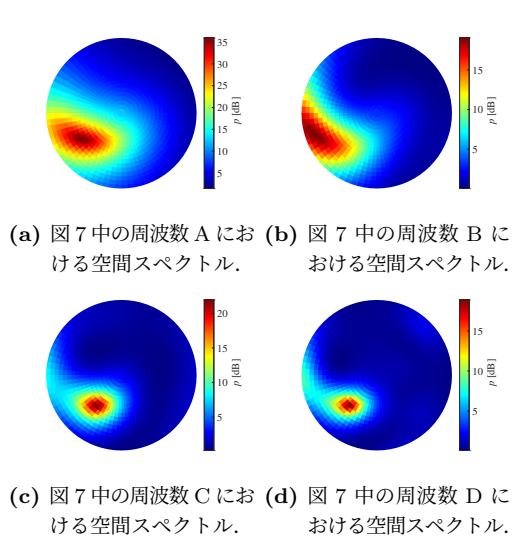


図 8: 抽出した周波数における空間スペクトルの一例.

ズが除外しきれておらず、図 4c のように、目標音成分の抽出に失敗している。これらの問題を解決するため、目標音の存在する周波数の抽出およびヒストグラムのスケールリングを用いた各周波数で異なる基準値の設定を行う。以下にそのアルゴリズムを示す。

### 3.1 周波数抽出

三次元空間スペクトルにおいて、基準値  $p_t$  以上の要素は、ノイズのみの場合は図 5c、目標音とノイズを含

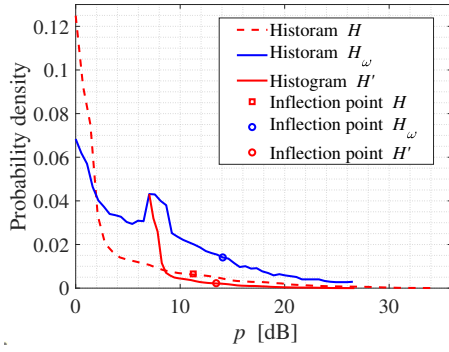


図 9: ヒストグラムのスケールリング.

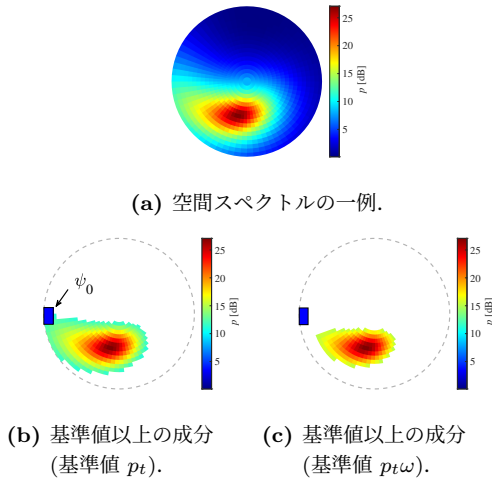


図 10: 基準値の違いによる空間スペクトルの除去される範囲の変化.

む場合は図 5b のようになる。このとき、それぞれのプロットされている周波数毎の基準値  $p_t$  以上の要素数、つまり全要素数に対する基準値  $p_t$  以上の要素数の確立密度を求める (図 6)。ノイズのみの場合は、周波数が高くなるにつれ確立密度が滑らかに減少するものの、目標音とノイズを含む場合は、目標音成分の存在する周波数で極大値をとる。よって、全要素数に対する基準値  $p_t$  以上の要素数の確立密度が極大値となる周波数を抽出することで、目標音成分の存在する周波数を抽出することが可能となる。これより、計算コストが大きいノイズ除外操作を目標音成分の存在する周波数対してのみ行うことができるようになり、計算コストが削減されると期待される。

三次元空間スペクトルの全方向に対する  $\Psi_{cut}(\omega)$  が占める確立密度を  $N$  としたとき、抽出する周波数  $\omega_e$  は次のアルゴリズムで求められる。

$$N(\omega) = P_{sicut}(\omega) / \Psi \quad (8)$$

$$\omega_e = \text{arglocalmaximum}_{\omega}(N(\omega)) \quad (9)$$

図 7 は  $N$  と  $\omega_e$  の一例である。抽出した周波数のうち、A~D の周波数における空間スペクトル  $P(\psi, \omega)$  を図 8 に示す。このとき、目標音方向は、方位角  $-135^\circ$ 、仰角  $-45^\circ$  である。周波数が低い A, B では、ノイズの

パワーも大きいいため、目標音成分がノイズに引っ張られてしまっているが、全ての抽出した周波数で目標音成分の存在が確認できる。特に、高周波数である C, D では、目標音成分が正確な方向にはっきりと存在していることがわかる。

### 3.2 周波数毎のノイズ判定基準値の設定

HIST-MUSIC-3D では、すべての周波数ビンに対して等しい基準値でノイズ判定を行っていたため、各周波数ごとの空間スペクトルの値に合ったノイズ判定の基準値を周波数毎に設定することで、目標音抽出の精度が向上できると期待される。

HIST-MUSIC-3D にて、ノイズ判定の基準値  $p_t$  である変曲点を求める際、空間スペクトルのヒストグラムを滑らかな関数にするために、近似関数を適応している。しかし、近似計算は計算コストが大きいため、すべての周波数に対してヒストグラムに近似関数を適応し変曲点を求めることは、リアルタイム性を考慮すると現実的ではない。そこで、小さい計算コストで周波数毎のノイズ判定基準値の設定するため、空間スペクトルの全要素のヒストグラムを、空間スペクトルの各周波数の要素のヒストグラムにスケールリングすることで、簡易的に各周波数のヒストグラムにおける変曲点を求める。アルゴリズムは以下の通りである。

スケールリングを行う空間スペクトルの全要素のヒストグラム  $H$  に対して、 $H$  の極大値のうち階級  $p$  の大きい点における極大値  $pd_{lm}$  と階級  $p_{lm}$ 、階級の最大値  $p_{max}$  を求めておく。

$$pd_{lm}^M = \text{localmax}(H(p)) \quad (10)$$

$$p_{lm}^M = p|_{H(p)=pd_{lm}^M} \quad (11)$$

$$p_{lm} = \max(p_{lm}^M) \quad (12)$$

$$pd_{lm} = H(p_{lm}) \quad (13)$$

$$p_{max} = \text{argmax}(H(p)) \quad (14)$$

各周波数の空間スペクトルの要素  $P(\psi, \omega)$  のヒストグラム  $H_{\omega}$  を求める。

$$H_{\omega} = \text{histogram}(P(\psi, \omega)) \quad (15)$$

式 (10)-(14) と同様に、 $H_{\omega}$  の極大値のうち階級の大きい点における極大値  $pd_{lm}^{\omega}$  と階級  $p_{lm}^{\omega}$ 、階級の最大値  $p_{max}^{\omega}$  を求める。そして、 $pd_{lm}$  を  $pd_{lm}^{\omega}$  に、 $p_{lm}$  を  $p_{lm}^{\omega}$  に、 $p_{max}$  を  $p_{max}^{\omega}$  に合わせるよう、 $H$  をスケールリングしたヒストグラム  $H'(\omega)$  を求める。

$$H'(p) = aH(bp + c)|_{pd_{lm}=pd_{lm}^{\omega}, p_{lm}=p_{lm}^{\omega}, p_{max}=p_{max}^{\omega}} \quad (16)$$

$H'$  において、 $H$  の変曲点に対応する点  $p_t\omega$  を求める。

$$p_t\omega = p_t|_{H'} \quad (17)$$

この値を周波数毎に算出し、それぞれの周波数におけるノイズ判定の基準値として用いる。

スケールリングを行ったヒストグラムの一例を図 9 に示す。赤色の点線で表された  $H$  を青の実線で表された

$h$ に合わせてスケーリングすることで、赤色の実線で表された  $H'$  を得た。ヒストグラムのスケーリングに伴い、赤四角の位置にある  $H$  の変曲点は、赤丸の位置に移動する。これは、近似を用いて算出した、青丸の位置にある  $h$  の変曲点と近い階級値を取る。よって、ヒストグラムのスケーリングを行うことで、近似を用いずに簡易的に各周波数のヒストグラムにおける変曲点を求めることができることがわかった。また、図 10 に改良前と改良後のノイズ判定の基準値を用いた際の、目標音抽出の精度を示す。図 10a に対して、改良前の基準値  $p_t$  を用いたとき、多くのノイズが除外しきれず、青四角で示されたドローンの方向である基準方向  $\psi_0$  を含む連続した部分に目標音成分も含まれてしまっている (図 10b)。一方で、改良後の基準値  $p_{t\omega}$  を用いたときは、目標音成分が  $\psi_0$  と離れて抽出できており、目標音がノイズと誤判定されずに抽出できていることがわかる (図 10c)。

## 4 評価実験

提案手法のノイズ耐性、検索範囲、計算コストおよび汎用性を検証するため、評価実験を行った。屋外環境にて収録したエゴノイズと、シミュレーションにより作成した任意の方向から到来した目標音を加算することにより評価用信号を作成し、解析を行った。

エゴノイズは、DJI 社製 Inspire 2 (図 11a) と ACSL 社製 MS-06LA (図 11b) の 2 種類のドローンに搭載したマイクロホンアレイにて収録した。プロペラには normal と high の異なる 2 種類のプロペラを用いた [11]。normal は一般的なプロペラであり、低周波数に強いエゴノイズが発生する。high altitude はエゴノイズの低周波数成分を抑制する代わりに高周波数のパワーが強い特徴を持つ。DJI 社製 Inspire 2 を用いた際のマイクロホンアレイには、下半球に 12ch、上半球に 4ch の MEMS マイクロホンが設置されている 16ch 球形マイクロホンアレイ (図 11c) [12] を用いた。また、ACSL 社製 MS-06LA を用いた際のマイクロホンアレイには、上半球の 4ch は使用せず、下半球のみの 12ch 球形マイクロホンアレイを用いた。本マイクロホンアレイでは、サンプリング周波数 16 kHz、量子化ビット数 24 bit で音響信号が収録される。マイクロホンアレイはドローンの中心から 600 mm の位置に設置した (図 c)。高度 10 m でホバリング中、および速度 1 m/s, 2 m/s, 3 m/s で飛行時のエゴノイズを収録した。目標音のサンプルには声およびホイッスルを用い、方位角が  $-180^\circ \leq \theta < 180^\circ$ 、仰角が  $-90^\circ \leq \phi \leq 0^\circ$  の範囲で  $5^\circ$  刻みで到来方向を設定し、幾何計算により得た伝達関数から各方向から到来した目標音を作成した。そして、収録したノイズと作成した目標音を、SNR が 4 dB 刻みで  $-20 \sim 0$  dB となるよう加算し、評価用信号を作成した。

まず、評価用信号を SEVD-MUSIC により解析し、MUSIC スペクトルを求めた。SEVD-MUSIC のパラメータは、目標音源数を 2、相関行列に用いる平均化フレーム数を 50、最小解析周波数を 500 Hz、最大解

析周波数を 4000 Hz とした。得られた MUSIC スペクトルに対し、HIST-MUSIC-3D [14]、本稿で改良した提案手法において、周波数抽出のみ導入した手法 (以後、HIST-MUSIC-3D+FE (Frequency Extraction) と呼ぶ)、周波数毎のノイズ判定の基準値のみ導入した手法 (以後、HIST-MUSIC-3D+SC (SCaling) と呼ぶ)、周波数抽出を行った後、抽出された周波数のみに対して周波数毎のノイズ判定の基準値の設定と以降のノイズ除外操作を行う手法 (以後、HIST-MUSIC-3D+FE,SC と呼ぶ) により処理を行い比較した。表 1 に比較条件、図 12 にそれぞれの比較条件の流れを示す。基準方向は、 $\psi_0 = (-180^\circ, 0^\circ)$  とした。

ノイズ耐性および検索範囲は、normal のプロペラを取り付けた DJI 社製 Inspire 2 のエゴノイズを用いた評価用信号に対して処理を行った。

リアルタイム性は、仮想マシンを用い、評価用信号に対して処理を行い評価した。仮想マシンは Jetson Nano を想定し、メモリ数 4GB、CPU4 コアとした。評価指標にはリアルタイム性能を示す指標である RTF (Real Time Factor) を用いた。RTF は、(演算時間)/(処理信号の時間) のように算出する。つまり、 $RTF < 1$  の場合、処理信号の時間以内に演算が終わるため、遅延時間が蓄積せずにリアルタイムに処理が行えることとなる。0.5 秒の信号を用いて 200 回試行を行い、平均値を評価値とした。

汎用性は、high のプロペラを取り付けた DJI 社製 Inspire 2 と ACSL 社製 MS-06LA の 2 種類のエゴノイズを用いた評価用信号に対し処理を行い、音源定位性能を比較した。

表 1: 比較に用いた手法

1. SEVD-MUSIC
2. HIST-MUSIC-3D
3. HIST-MUSIC-3D+FE
4. HIST-MUSIC-3D+SC
5. HIST-MUSIC-3D+FE,SC

## 5 結果

SNR と目標音源、ドローンの機種、プロペラ、飛行状況の異なる様々な条件で作成した評価用信号を、表 1 で示した各手法で処理した結果を図 13 に示す。

空間スペクトルは図 2a の設定軸に従ってプロットされており、各方向から到来した音のパワーをカラーマップで示している。SNR はすべての結果で  $-12$  dB である。エゴノイズは、a, b はともにエゴノイズに normal のプロペラを取り付けた DJI 社製 Inspire 2、c は high のプロペラを取り付けた DJI 社製 Inspire 2、d は ACSL 社製 MS-06LA のエゴノイズを用いた。目標音サンプルは、a は声、b~d はホイッスルとした。目標音方向は、a, c, d が方位角  $\theta = -135^\circ$ 、仰角  $\phi = -45^\circ$ 、b が方位角  $\theta = -160^\circ$ 、仰角  $\phi = -13^\circ$  の結果である。い



図 11: 実験にて使用したドローンおよびマイクロホンアレイ。

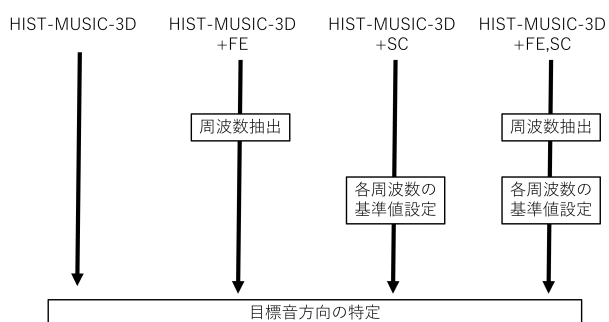


図 12: 比較条件の流れ。

ずれの結果も、エゴノイズは空間スペクトルの左側に、目標音は左下側、ノイズに対して右下側にノイズと重なる位置に現れる。

### 5.1 ノイズ耐性と搜索範囲の評価結果

ノイズ耐性と搜索範囲の評価結果を図 13a,b に示す。SEVD-MUSIC の結果は、エゴノイズの右下に目標音が飛び出すように現れているが、エゴノイズの方がパワーが大きい。HIST-MUSIC-3D は、a の結果は目標音の抽出が行えているが、より目標音がエゴノイズと近い方向にある b の結果では目標音がノイズと誤認識され抽出ができていない。HIST-MUSIC-3D+FE も、同様の理由で、b の結果では目標音がノイズと誤認識され抽出ができていない。しかし、HIST-MUSIC-3D の結果 b にて左上側に僅かであるが残っているパワーの大きいエゴノイズが、HIST-MUSIC-3D+FE の b の結果では除外できている。これより、目標音の存在する周波数の抽出することは、リアルタイム性を向上させるだけでなく、目標音の定位に影響のあるパワーの大きいエゴノイズのみ存在する周波数を除外する効果もあるがわかった。HIST-MUSIC-3D+SC は、b の結果で目標音方向とエゴノイズ側にずれた位置に最大ピークが表れており、定位に影響のあるパワーの大きいエゴノイズが除外しきれていないため、正確な定位が行えていない。しかし、HIST-MUSIC-3D では除外されてしまった目標音成分の存在が確認できるため、周波数毎にノイズ判定の基準値を設定することで、目標音

成分抽出の精度が向上していることがわかる。一方で、HIST-MUSIC-3D+FE,SC は、どちらの結果も目標音の抽出が行えており、目標音が最大ピークとなった。これは、目標音成分の存在する周波数の抽出を行うことで、パワーの大きいエゴノイズのみ存在する周波数では除去しきれなかったノイズが周波数抽出の時点で除去できるため、HIST-MUSIC-3D+SC は、b の結果では定位が失敗していたものの、HIST-MUSIC-3D+FE,SC では定位が可能になったと考えられる。よって、目標音の存在する周波数の抽出を行い、周波数毎に最適なノイズ判定の基準値を設けることで、エゴノイズと近い方向にある目標音を含む広い搜索範囲において、定位が可能であると期待される。

さらに、評価用音響信号 50 フレーム (25 秒分)、飛行状況 4 種類、目標音 2 種類、音源方向 1,297 通り、SNR6 通りの全 3,112,800 回の試行を行い、各手法における、搜索可能率 (図 14b) を求めた。

音源定位の成功率は、全試行のうち、定位が成功した試行数の割合を示す。つまり、ノイズ耐性と搜索可能範囲を統合した結果である。ここでは、表 1 で示した各手法に加えて、従来手法として、GEVD-MUSIC, iGEVD-MUSIC, AFRF-MUSIC, 角度制限 SEVD-MUSIC についても評価を行った。GEVD-MUSIC で使用するノイズ相関行列には、別日の実験にて収録した DJI 社製 Inspire 2 のエゴノイズを用いた。iGEVD-MUSIC と AFRF-MUSIC のパラメータは、ノイズと定位音源のフレーム差を 50、ノイズの相関行列のフレーム数は iGEVD-MUSIC は 100、AFRF-MUSIC は 50 とした。また、AFRF-MUSIC の解析周波数帯域は 500~4000 Hz の範囲内の 500Hz とした。角度制限 SEVD-MUSIC は、 $-90 \sim 90^\circ$  (角度制限 SEVD(1)) と  $-135 \sim 135^\circ$  (角度制限 SEVD(2)) の 2 つの搜索範囲を設定した。

HIST-MUSIC-3D は、ほとんどの SNR で従来手法に比べて高い成功率を獲得したが、SNR が  $-4$  dB のとき、AFRF-MUSIC よりも成功率が下回った。一方で、HIST-MUSIC-3D+FE, HIST-MUSIC-3D+SC, HIST-MUSIC-3D+FE,SC は、すべての SNR で従来手法に比べて高い成功率を獲得した。しかし、HIST-MUSIC-3D+FE は、SNR が  $-20$  dB のとき、HIST-MUSIC-3D より低い成功率となった。これは、HIST-MUSIC-3D は、エゴノイズと近い方向にある目標音に

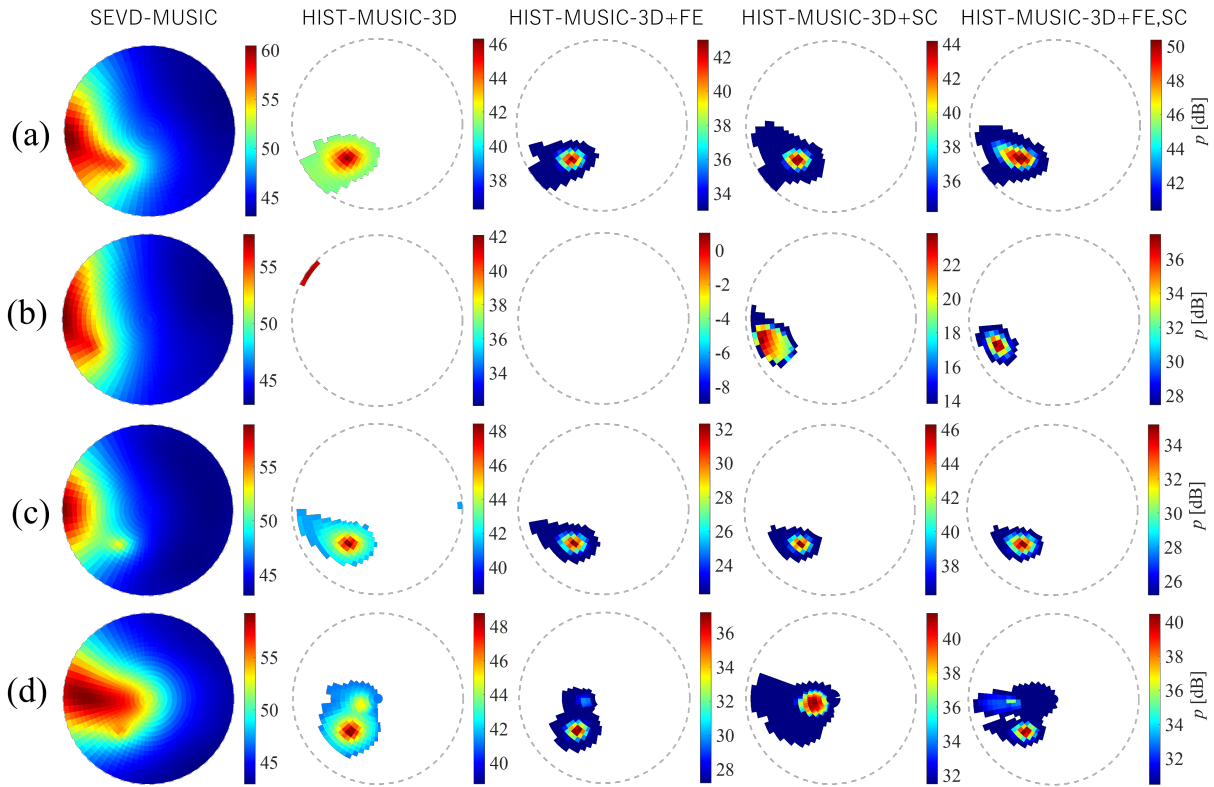


図 13: 各手法を用いて得られた空間スペクトル. SNR = -12dB.

(a) エゴノイズ: Inspire 2 (normal), 目標音源: 声,  $\psi_{target} = (-135^\circ, -45^\circ)$ , (b) エゴノイズ: Inspire 2 (normal), 目標音源: ホイッスル,  $\psi_{target} = (-160^\circ, -25^\circ)$ , (c) エゴノイズ: Inspire 2 (high altitude), 目標音源: ホイッスル,  $\psi_{target} = (-135^\circ, -45^\circ)$ . (d) エゴノイズ: MS-06LA, 目標音源: ホイッスル,  $\psi_{target} = (-135^\circ, -45^\circ)$ .

表 2: 各手法での RTF の結果.

処理条件	RTF
SEVD-MUSIC	0.92
HIST-MUSIC-3D	0.97
HIST-MUSIC-3D+FE	0.95
HIST-MUSIC-3D+SC	1.09
HIST-MUSIC-3D+FE,SC	0.96

対して、目標音が除外されているにも関わらず、目標音付近（目標音方向と誤差  $\pm 5^\circ$  以内）のエゴノイズが除外しきれず最大ピークとなり定位成功となる場合があること、また、目標音の存在する周波数の抽出に失敗し、定位が失敗していることも原因として考えられる。HIST-MUSIC-3D+SC はすべての SNR で最も高い検索可能性を獲得した。これは、HIST-MUSIC-3D+FE,SC では、目標音の存在する周波数の抽出ができず定位が失敗している場合があることが原因として挙げられる。一方で、SNR が -20dB と低いとき、HIST-MUSIC-3D+FE,SC の成功率は HIST-MUSIC-3D+SC とほとんど等しく、高性能であることがわかった。

## 5.2 リアルタイム性の評価結果

リアルタイム性の評価結果を表 2 に示す。

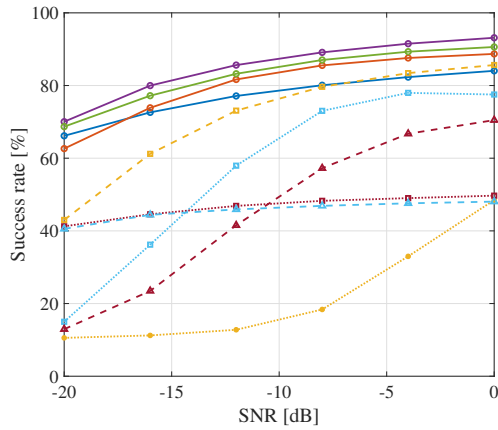
SEVD-MUSIC は各方向の MUSIC スペクトルを固有値展開により算出しているため、計算コストが小さい。HIST-MUSIC-3D は SEVD-MUSIC に単純な計算を加える処理を行うため、SEVD-MUSIC より僅かに大きくなったもののリアルタイム性を獲得できている。HIST-MUSIC-3D+FE は、最も RTF が小さくなった。これは、抽出された周波数のみに対してノイズ除外操作を行うため、計算コストが比較的大きいノイズ除外操作の回数が減ることで、周波数抽出の計算を加えても処理全体としては計算コストが小さくなったためである。HIST-MUSIC-3D+SC は、RTF > 1 となり、リアルタイム性が獲得できなかった。これは、すべての周波数に対してヒストグラム算出とノイズ除外操作を行うため、計算コストが高くなったことが原因である。HIST-MUSIC-3D+FE,SC は、RTF > 1 となり、また、これまでの提案手法に比べても小さい RTF となった。よって、周波数の抽出を行うことで、ヒストグラムの算出とノイズ除外操作を行う周波数を制限することで、リアルタイム性を獲得できることがわかった。

## 5.3 汎用性

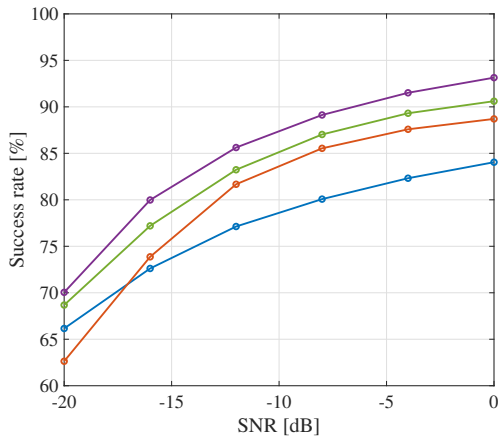
汎用性の評価結果を図 13c,d に示す。

SEVD-MUSIC の結果は、左側にエゴノイズ、その右下側に目標音が確認できるが、エゴノイズの方がパワーが大きい。HIST-MUSIC-3D は、どちらの結果も





(a) 各手法の成功率の比較.



(b) (a) の拡大図 (提案手法のみ).

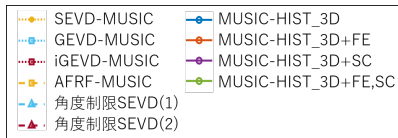


図 14: 音源定位の成功率.

目標音の定位が行えている. HIST-MUSIC-3D+FE も, 両方の結果で目標音が最大ピークとなったが, dの結果において, これまでの提案手法の結果にも表れているノイズが, 目標音の上側に残っている. よって, SEVD-MUSIC の dの結果のようにエゴノイズが仰角方向に大きく広がっているとき, HIST-MUSIC-3D+FE では, エゴノイズの先端部分, つまり仰角  $-90^\circ$  付近のエゴノイズ成分が目標音成分のようにふるまうことが原因で, ノイズのみの周波数をすべて除外できないことがわかった. また, この除外しきれないノイズは, HIST-MUSIC-3D+SC で大きく現れており, 最大ピークとなってしまっている. こちらも, エゴノイズの先端部分の成分が目標音成分のようにふるまうことで, エゴノイズの先端成分を目標音として誤って抽出するようなノイズ判定の基準値が設定されてしまっていることがわかる. HIST-MUSIC-3D+FE,SC についても, 同様に目標音の上側にエゴノイズが除外しきれず残っているが, 目標音が最大ピークとなり, 周波数の抽出と周波数毎のノイズ判定の基準値の設定をどちらも取り

入れる手法は, 汎用性を獲得できることがわかった.

## 5.4 結果のまとめ

ノイズ耐性, 搜索範囲の結果から, HIST-MUSIC-3D+SC が最も搜索可能率が高く, 少し性能が下回るものの, HIST-MUSIC-3D+FE,SC も特に低い SNR において高い搜索可能率を獲得していることがわかった. 一方で, HIST-MUSIC-3D+FE は, 搜索可能率がこれまでの提案手法と同程度であり, 他の改良した提案手法に比べて劣っていた. 計算コストの結果から, HIST-MUSIC-3D+FE は最も計算コストが小さく, また HIST-MUSIC-3D+FE,SC もリアルタイム性を獲得していることがわかった. 一方で, HIST-MUSIC-3D+SC は, リアルタイム性を獲得できなかった. 汎用性の評価結果から, HIST-MUSIC-3D+FE と HIST-MUSIC-3D+FE,SC は汎用性を得られたが, HIST-MUSIC-3D+SC は汎用性を得られなかった. 以上のことを踏まえて, HIST-MUSIC-3D+FE,SC が最もすべての求められる音源定位手法の性能を満たしており, 最も有用性があると期待できる.

## 6 結論

本稿では, 著しい時刻変化を伴うノイズに対する頑健性, 広い搜索範囲, 低い計算コスト, 汎用性をすべて満たす音源定位手法の開発を目的に, 過去の情報を用いず, 得られた現在の空間スペクトルから, ヒストグラム情報と周波数情報に基づき, 目標音成分の存在する周波数の抽出と各周波数で最適なノイズ判定の基準値を設定することで, ノイズの判定と目標音成分の抽出を行う手法を提案した. これまでの提案手法では, パワーの大きいエゴノイズが除外しきれない場合があることや, エゴノイズと近傍の方向に存在する目標音がエゴノイズと誤判定され定位が不可能であるといった問題点があったが, ノイズ除外操作を行う周波数を目標音の存在する周波数のみ制限すること, 各周波数に存在するエゴノイズ, 目標音成分のパワーに合ったノイズ判定の基準値を設けるよう提案手法を改良することにより, これまでの提案手法における問題点を解決することができた. 評価実験の結果, 高いノイズ耐性, 広い搜索範囲, リアルタイム性, 汎用性をすべて考えたとき, 周波数の抽出と周波数毎のノイズ判定の基準値を設定する処理をどちらも導入する手法が最も有用性があると確認できた. 今後は, 屋外実環境にて評価を行い, 改良した提案手法の実用性の評価を行っていく.

## 謝辞

本研究は, JSPS 科研費 22K14218 の助成を受けた.

## 参考文献

- [1] 加藤, 寺島, 高見: 要救助者の複数ドローンによる協調探索のためのエッジサーバ集約型自動スケジューリング手法とシミュレーション評価マルチメディア, 分散協調とモバイルシンポジウム 2019 論文集, pp.291–296 (2019)
- [2] L. wang, A. Cavallaro: Deep-Learning-Assisted Sound Source Localization From a Flying Drone *IEEE Sensors Journal* VOL. 22, NO. 21, pp. 20828-20838 (2022)
- [3] L. wang, A. Cavallaro: Drone Ego-Noise Cancellation for Improved Speech Capture using Deep Convolution Autoencoder Assisted Multi-stage Beamforming *2022 25th International Conference on Information Fusion (FUSION)* (2022)
- [4] R. O. Schmidt: Multiple Emitter Location and Signal Parameter Estimation *IEEE Trans. Antennas and Propagation*, VOL. 34, NO. 3, pp. 276-280 (1986)
- [5] K. Nakamura, K. Nakadai, F. Asano, Y. Hasegawa, H. Tsujino: Intelligent Sound Source Localization for Dynamic Environment, *IEEE/RSJ International Conference on Intelligent Robots and Systemd* (2009)
- [6] K. Nakamura, K. Nakadai, G. Ince: Real-time Super-resolution Sound Source Localization for Robots *Proc. of IEEE/RSJ International Conference on Robots and Intelligent Systems (IROS)* pp. 694-699 (2012)
- [7] K. Okutani, T. Yoshida, K. Nakamura, K. Nakadai: Intelligent Sound Source Localization for Dynamic Environment, *Outdoor Auditory Scene Analysis Using a Moving Microphone Array Embedded in a Quadcopter* (2012)
- [8] T. Ohata, K. Nakamura, A. Nagamine, T. Mizumoto, T. Ishizaki, R. Kojima, O. Sugiyama, K. Nakadai: Outdoor Sound Source Detection Using a Quadcopter with Microphone Array *J. of Robotics and Mechatronics* VOL. 29, NO. 1, pp. 177-187, (2017)
- [9] K. Hoshihara, K. Nakadai, M. Kumon, H. G. Okuno: Assessment of MUSIC-Based Noise-Robust Sound Source Localization with Active Frequency Range Filtering, *Journal of Robotics and Mechatronics*, VOL.30, NO. 3, pp. 426-435, (2018)
- [10] K. Hoshihara, K. Washizaki, M. Wakabayashi, T. Ishiki, M. Kumon, Y. Bando, D. Gabriel, K. Nakadai, H. G. Okuno: Design of UAV-Embedded Microphone Array System for Sound Source Localization in Outdoor Environments *Sensors* VOL. 17, NO. 11, pp. 1-16, (2017)
- [11] R. Noda, T. Nakata, K. Senda, H. Liu : Multi-scale morphological effect on noise level and frequency characteristics of drone propellers *QUIET DRONS International e-Symposium on UAV/UAS Noise* pp. 77–142, (2020)
- [12] K. Nonami, K. Hoshihara, K. Nakadai, M. Kumon, H.G. Okuno, Y. Tanabe, K. Yonezawa, H. Tokutake, S. Suzuki, K. Yamaguchi, S. Sunada, T. Takaki, T. Nakata, R. Noda, H. Liu, S. Tadokoro: Recent R&D Technologies and Future Prospective of Flying Robot in Tough Robotics Challenge *Disaster Robotics - Results from the ImPACT Tough Robotics Challenge, Satoshi Tadokoro Ed., Springer International Publishing* pp. 77–142, (2019)
- [13] 小松崎, 干場, 武田, 菅原: ”ヒストグラム情報を用いた時刻変化の著しい雑音に対する体制の高い音源定位手法の提案”, 第 40 回日本ロボット学会 学術講演会, RSJ2022AC4J3-06, (2022)
- [14] 小松崎, 干場, 岩附:”ドローン聴覚におけるヒストグラム情報と周波数情報を用いた音源定位性能向上の検討”, 第 61 回人工知能学会 AI チャレンジ研究会, pp. 26-32, (2022)