一般社団法人 人工知能学会
Japanese Society for Artificial Intelligence

人工知能学会資料
JSAI Technical Report
SIG-Challenge-049-2 (11/25)

アクティブ周波数レンジフィルタを用いた雑音にロバストな音源定位手法の提案

Noise-robust sound source localization method using active frequency range filter

干場功太郎 *1, 中臺一博 *1,*2, 公文誠 *3, 奥乃博 *4

Kotaro HOSHIBA[*1], **Kazuhiro NAKADAI**[*1,2], **Makoto KUMON**[*3], **Hiroshi G. OKUNO**[*4]

東京工業大学 工学院 システム制御系 *1

(株) ホンダ・リサーチ・インスティチュート・ジャパン *2

熊本大学 大学院 先端科学研究部 *3

早稲田大学 実体情報学博士プログラム *4

Department of Systems and Control Engineering, School of Engineering

Tokyo Institute of Technology[*1]

Honda Research Institute Japan Co., Ltd.[*2]

Faculty of Advanced Science and Technology, Kumamoto University[*3]

Graduate Program for Embodiment Informatics, Waseda University[*4]

{hoshiba, nakadai}@ra.sc.e.titech.ac.jp, kumon@gpo.kumamoto-u.ac.jp, okuno@nue.org

Abstract

われわれは，被災地等での要救助者の探索を目的に，UAV (Unmanned Aerial Vehicle) 搭載マイクロホンアレイを用いた音源探査の研究を行っている．これまで，屋外実環境での使用を想定し，さまざまな音源定位手法を提案してきた．しかし，これらの手法は，雑音耐性とリアルタイム性がトレードオフの関係にあった．被災地等での音源探査の場合，雑音耐性とリアルタイム性の両者が重要であり，それらを同時に満たす手法の開発が必要である．本稿では，MUSIC (MUltiple Signal Classification) 法に基づく，アクティブ周波数レンジフィルタを用いた音源定位手法を提案する．本手法では，単純な四則演算のみを用いたフィルタを使用することで，雑音耐性とリアルタイム性を確保する．計算機シミュレーションによる評価を行った結果，以前の手法に比べ，定位性能・処理遅延ともに優位性を示すことができた．

1 はじめに

屋外環境での音響処理に関する研究は，計測分野などさまざまな応用が考えられるため，さかんに行われている．われわれは，JST ImPACT タフロボティクスチャレンジの極限音響チームにおける研究開発活動の一環として，これまでに培ってきたロボット聴覚技術 [1] を用いて，被災地等での要救助者の探索を目的に，UAV (Unmanned Aerial Vehicle) 搭載マイクロホンアレイを用いた音源探査の研究を行っている．こうした手法を確立できれば，上空から広範囲かつ迅速に被災者の音声を探査する有効な手段ができると考えられる．これまで，動的に変化する UAV の自己雑音を抑制するさまざまな音源定位手法の提案や [2, 3]，屋外実環境において実時間で音源探査を行うシステムの開発を行ってきた [4, 5]．しかし，これらの手法は，雑音耐性を高めると計算コストが大きくなるためリアルタイム性が失われ，リアルタイム性を高めると雑音耐性が低くなるといった，トレードオフの関係にある．また，機械学習の一種である LSTM (Long Short Term Memory) を用いて，周波数方向のマスクを作成するという手法も提案されているが [6]，計算コストが大きく，リアルタイム性を確保できない．本研究では，音源探査システムの組み込みシステム化の検討も行っているため [7]，組み込みボードでも処理が行えるよう，組み込みボードでも処理が行えるような計算コストの小さい，リアルタイム性と雑音耐性の両者を満たす音源定位手法の開発が必要である．そこで本稿では，MUSIC (MUltiple Signal Classification) 法 [8] に基づく，単純な四則演算のみを用いたアクティブ周波数レンジフィルタを用いた音源定位手法を提案し，従来手法と比較することで音源定位性能の評価と考察を行う．

2　音源定位手法

本章では，本稿で扱う音源定位手法について述べる．

2.1　SEVD-MUSIC 法

音源定位において多く用いられる，一般的な MUSIC 法である SEVD-MUSIC (MUSIC based on Standard Eigen Value Decomposition) 法のアルゴリズムを示す．M チャンネル入力音響信号の f フレーム目をフーリエ変換して得られる $Z(\omega, f)$ から，以下のように相関行列 $R(\omega, f)$ を定義する．

$$R(\omega, f) = \frac{1}{T_R} \sum_{\tau=f}^{f+T_R-1} Z(\omega, \tau) Z^*(\omega, \tau) \tag{1}$$

ここで，ω は周波数ビン番号，T_R は相関行列の計算に用いるフレーム数，Z^* は Z の共役転置である．次に，f 番目のフレームに対して，f_s 前のフレームから，T_N フレーム分の信号は雑音区間であると仮定して，雑音の相関行列 $K(\omega, f)$ を求める．SEVD-MUSIC 法では，こうして得られた $R(\omega, f)$ を固有値展開して固有ベクトルを計算する．

$$R(\omega, f) = E(\omega, f)\Lambda(\omega, f)E^*(\omega, f) \tag{2}$$

ただし，$\Lambda(\omega, f)$ は降順に並んだ固有値を対角成分に持つ行列，$E(\omega, f)$ は $\Lambda(\omega, f)$ に対応する固有ベクトルを並べた行列である．これと，UAV 座標系での音源方向 ψ に対応した伝達関数 $G(\omega, \psi)$ を用いて MUSIC 空間スペクトル $P(\omega, \psi, f)$ を計算する．

$$P(\omega, \psi, f) = \frac{|G^*(\omega, \psi)G(\omega, \psi)|}{\sum_{m=L+1}^{M} |G^*(\omega, \psi)e_m(\omega, \psi)|}. \tag{3}$$

ただし，L は目的音源数，e_m は，E に含まれる m 番目 $(1 \leq m \leq M)$ の固有値ベクトルを表す．また，ψ は UAV に対する方位角 θ，仰角 ϕ から $\psi = (\theta, \phi)$ と定義する．このようにして得られた $P(\omega, \psi, f)$ を，音源方向を推定するために以下のように ω 方向に平均する．

$$\bar{P}(\psi, f) = \frac{1}{\omega_H - \omega_L + 1} \sum_{\omega=\omega_L}^{\omega_H} P(\omega, \psi, f) \tag{4}$$

なお，ω_H，ω_L は使用する周波数ビンの上限と下限に対応したインデックスである．$\bar{P}(\psi, f)$ に対して閾値処理，ピーク検出を行い，得られたピークに対する ψ を音源方向として検出する．

SEVD-MUSIC 法では，雑音に対する耐性が低いが，計算コストが小さいため，処理遅延が少ないという特徴が挙げられる．また，ω_H，ω_L を目的音源に応じ狭帯域に設定することで雑音耐性を高めることもできるが，異なる周波数特性を持った目的音源が複数あった場合，対応することができない．

2.2　iGSVD-MUSIC 法

われわれは，これまで屋外音環境特有の問題を解決するため，MUSIC 法を拡張した音源定位手法を提案してきた．その一つとして，MUSIC 法に一般化特異値展開を導入した iGSVD-MUSIC ((MUSIC based on incremental Generalized Singular Value Decomposition) 法 [3] が挙げられる．iGSVD-MUSIC 法では，時間変化する雑音に対応するため，逐次的に雑音モデルを推定し，雑音モデルを用いた白色化により，音源定位性能の向上を図る．

以下にそのアルゴリズムを示す．iGSVD-MUSIC 法では，f 番目のフレームに対して，f_s 前のフレームから，T_N フレーム分の信号は雑音区間であると仮定して，雑音の相関行列 $K(\omega, f)$ を求める．

$$K(\omega, f) = \frac{1}{T_N} \sum_{\tau=f-f_s-T_N}^{f+f_s} Z(\omega, \tau) Z^*(\omega, \tau) \tag{5}$$

iGSVD-MUSIC 法では，フレームごとに雑音が推定できるため，動的な雑音変化に対応できることができる．R の左から K^{-1} を掛けることで，雑音成分を白色化する．こうして得られた $K^{-1}(\omega, f)R(\omega, f)$ を一般化特異値展開して特異値ベクトルを計算する．

$$K^{-1}(\omega, f)R(\omega, f) = Y_l(\omega, f)\Sigma(\omega, f)Y_r^*(\omega, f) \tag{6}$$

ただし，$\Sigma(\omega, f)$ は降順に並んだ特異値を対角成分に持つ行列，$Y_l(\omega, f)$，$Y_r(\omega, f)$ は $\Sigma(\omega, f)$ に対応する特異値ベクトルを並べた行列である．ここから，SEVD-MUSIC 法と同様に，MUSIC 空間スペクトル $P(\omega, \psi, f)$ を計算する．

$$P(\omega, \psi, f) = \frac{|G^*(\omega, \psi)G(\omega, \psi)|}{\sum_{m=L+1}^{M} |G^*(\omega, \psi)y_m(\omega, \psi)|} \tag{7}$$

ただし，y_m は，Y_l に含まれる m 番目の特異値ベクトルを表す．このようにして得られた $P(\omega, \psi, f)$ を，Eq. 4 と同様に，ω 方向に平均する．その後，閾値処理，ピーク検出を行う．

iGSVD-MUSIC 法では，UAV のローター音など時間変化する雑音を抑制することができる一方，計算コストが大きく，2 s 以上の遅延が発生することがわかった [5]

2.3　アクティブ周波数フィルタを用いた音源定位手法の提案

上記の 2 つの手法は，雑音耐性とリアルタイム性がトレードオフの関係にあり，実環境で音源探査を行う場合，両者を満たす手法の開発が必要である．そこで，MUSIC 法における，使用する周波数レンジをアクティブに変化させる手法を提案する．SEVD-MUSIC 法をベースに，Eq. 4 における ω_H，ω_L を状況に応じて変化させることで，雑音に対する耐性とリアルタイム性を確保する．

アルゴリズムを以下に示す．f 番目のフレームに対して，f_s 前のフレームから，T_N フレーム分の信号は雑音

区間であると仮定して，雑音の周波数スペクトル Z_n をフレーム間，チャンネル間の平均から求める．

$$Z_n(\omega, f) = \frac{1}{T_N \cdot M} \sum_{m} \sum_{\tau=f-f_s+1}^{f-f_s+T_N} Z(\omega, \tau) \tag{8}$$

得られた Z_n と現在の周波数スペクトル Z の差分から，評価関数 $J(\omega, f)$ を算出する．

$$J(\omega, f) = Z(\omega, f) - Z_n(\omega, f) \tag{9}$$

J は，雑音に対する各周波数のパワーの変化量を意味する．つまり，J の値が大きい周波数は目的音源によるものと考えられる．そこで，J が最も大きい周波数 $\omega_J(f)$ を求める．

$$\omega_J(f) = \underset{\omega}{\operatorname{argmax}}\, J(\omega, f) \tag{10}$$

得られた ω_J から，使用する周波数レンジを決定する．

$$\omega_H = \omega_J + \frac{f_w}{2} \tag{11}$$

$$\omega_L = \omega_J - \frac{f_w}{2} \tag{12}$$

ここで，f_w は周波数レンジの大きさである．f_w は目的音源の周波数特性を考慮し，可能な限り小さく設定する．そうすることで，雑音の影響を抑制する．このように算出された ω_H，ω_L を用いて，SEVD-MUSIC 法による処理を行う．以降，本手法を AFRF-MUSIC (MUSIC using Active Frequency Range Filter) 法と呼ぶ．

AFRF-MUSIC 法では，f_w を狭帯域に設定することで雑音への耐性が期待できる．また，アクティブに周波数レンジを変化させるため，異なる周波数特性を持った複数の音源が存在した場合にも対応することができる．さらに，本手法では，SEVD-MUSIC 法に加えて単純な四則演算のみを使用しているため，計算コストが小さく，処理遅延が少なくなり，リアルタイム性を確保できる．

3 検証実験

3.1 実験方法

計算機シミュレーションにより，音響信号を作成し，各音源定位手法を行うことで，定位性能評価を行った．探査対象の音源として，周波数特性の異なる，笛の音と人の声の 2 種類を用いた．これらの音源を用いて，方位角 θ を $-180°$〜$180°$，仰角 ϕ を $-90°$〜$0°$ の範囲で，5° 刻みで各方向から到達した場合の信号をシミュレーションにより作成する．また，雑音として実際のアレイで収録した UAV のローター音を加え，信号の SNR (Signal-to-Noise Ratio) はを 20〜-20 dB の間で変化させ，各音源定位手法により信号処理を行う．マイクロホンアレイには，Fig. 1(a) に示される，球形のマイクロホンアレイを用いる．本マイクロホンアレイはカスケード接続された 12 ch の MEMS マイクロホンで構成され，Fig. 1(b) のように UAV のアームの先に接続し，使用する．各マイクロホンのアレイは，Fig. 1(c) のように球形の筐体の下半球に配置されている．探査対象音源，アレイで収録したローター音のスペクトログラムを Fig. 2 に示す．笛の音は 2.5〜3 kHz にパワーが集中していることがわかる．また，声は 850 Hz 付近が最もパワーが大きく，約 4 kHz まで倍音が存在している．ローター音の周波数帯域は広域に渡っているが，特に 2 kHz 以下の成分が大きい．作成した音響信号を，前述の各音源定位法にて処理を行う．アルゴリズムの実装には，ロボット聴覚オープンソースソフトウェア HARK (Honda Research Institute Japan Audition for Robots

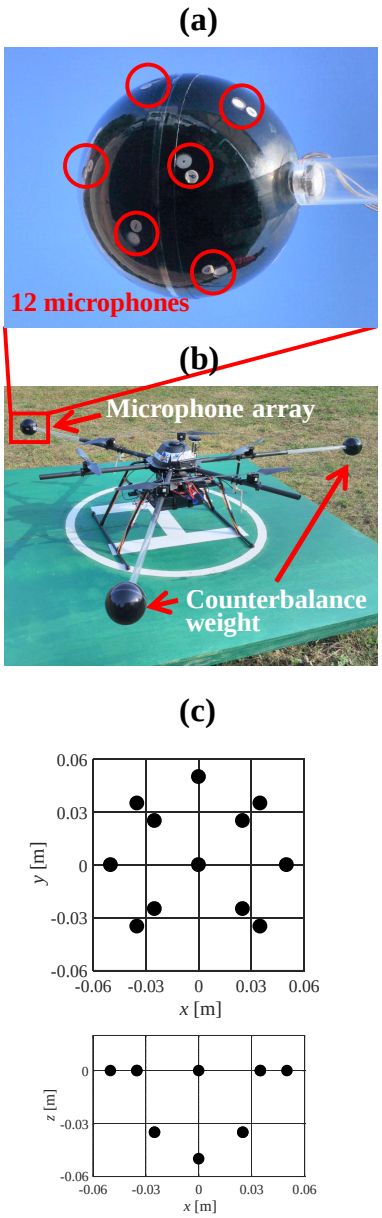

Fig. 1: (a) Microphone array, (b) UAV with microphone array, (c) coordinates of microphone positions in the microphone array.

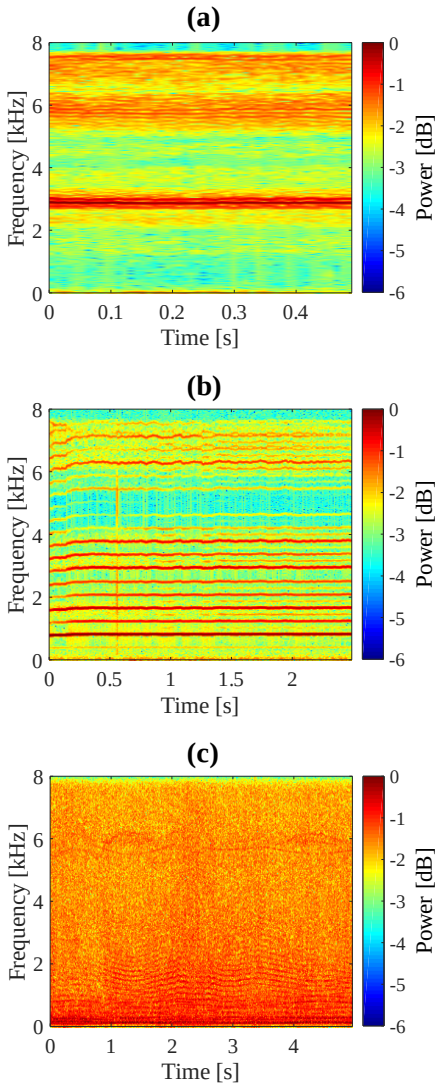

Fig. 2: Spectrograms. (a) Whistle, (b) voice, (c) noise of UAV recorded by microphone aray.

with Kyoto University)[1][9] を用いた．MUSIC 法で用いる伝達関数 (Eq. 3 における G) については，幾何計算で算出した．各方向・各音源につき，周波数レンジを 500–3000 Hz に設定した SEVD-MUSIC・iGSVD-MUSIC，レンジを 2500–3000 Hz に設定した SEVD-MUSIC，AFRF-MUSIC の 5 つの手法を用いて MUSIC スペクトル P を各 250 フレーム算出し，評価を行った．AFRF-MUSIC における f_w は 500 Hz に設定した．本稿では，Fig. 3 に示すように方位角 θ，仰角 ϕ を設定し，MUSIC スペクトルの評価を行う．

3.2 実験結果

実験結果について述べる．SNR が 0 dB の場合に算出された MUSIC スペクトルの一例を Fig. 4 に示す．(a) が SEVD-MUSIC (500–3000 Hz)，(b) が iGSVD-MUSIC (500–3000 Hz)，(c)，(d) が SEVD-MUSIC (2500–3000 Hz)，(e) が AFRF-MUSIC による結果である．また，音源として，(a)，(b)，(c)，(e) は笛の音，(d) は声を用いている．Fig. 3 に従ってプロットされており，カラーバーで各方向からの音のパワーを示している．また，音源の方向は $\theta = -30°$，$\phi = -45°$ に設定している．Fig. 4(a) に示されるように，周波数レンジを大きく設定した場合の SEVD-MUSIC では，音源方向にピークが確認できるものの，MUSIC スペクトルに UAV のローター雑音が大きく現れているため，定位精度に大きく影響が出ると考えられる．iGSVD-MUSIC を用いた場合，Fig. 4(b) のように雑音が大きく抑制され，音源方向に鋭いピークを確認することができる．また，SEVD-MUSIC における周波数レンジを，笛の周波数特性を考慮し，2500–3000 Hz と設定した場合，Fig. 4(c) のように，ピークの鋭さはないが，ローター雑音を抑制することができている．このことから，周波数レンジを狭帯域に設定することで雑音耐性を確保することが可能であるとわかる．しかし，このように固定の周波数レンジを用いると，周波数特性の異なる声が音源の場合，Fig. 4(d) のようにピークがほぼ確認できなくなることから，2 種類以上の異なる音源が存在した場合に定位性能が低下してしまう．AFRF-MUSIC を用いた場合，使用する狭帯域周波数レンジが適切に設定されるため，Fig. 4(e) に示されるように，レンジを 2500–3000 Hz に設定した SEVD-MUSIC と同様の結果になる．また，AFRF-MUSIC では音源の周波数特性に応じてレンジが変化するため，異なる音源が存在する場合であっても定位性能が低下しにくいと考えられる．

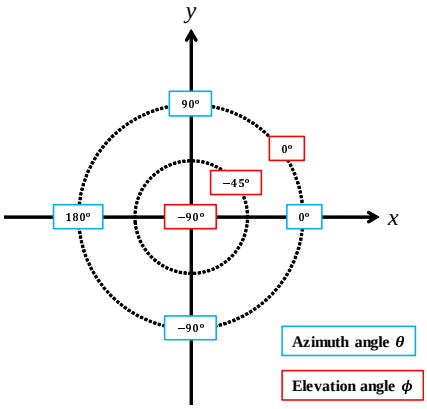

Fig. 3: Setting of azimuth angle θ and elevation angle ϕ.

3.3 考察

各手法を用いた場合の音源定位の正解率による定位性能の評価を行った．MUSIC スペクトルの最大ピークの方向が，音源の設定方向と一致した場合に正解とし，笛の音・声を用いたすべての方向からのシミュレーション音源にて処理・ピーク探査を行い，正解率を算出する．Fig. 5 が算出

[1] http://www.hark.jp/

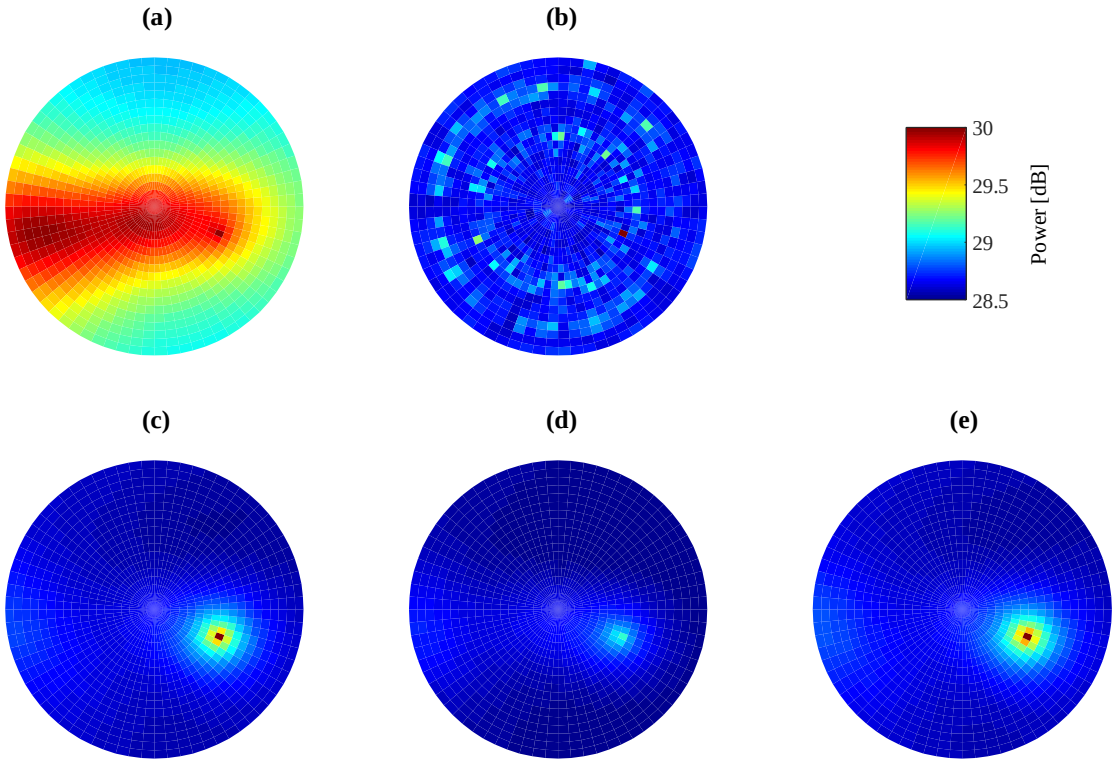

Fig. 4: MUSIC spectra. (a) whistle sound using SEVD-MUSIC (500–3000 Hz), (b) whistle sound using iGSVD-MUSIC (500–3000 Hz), (c) whistle sound using SEVD-MUSIC (2500–3000 Hz), (d) voice using SEVD-MUSIC (2500–3000 Hz), (e) whistle sound using AFRF-MUSIC.

した各手法の正解率である．横軸が SNR，縦軸が正解率である．周波数レンジを広帯域に設定した SEVD-MIUSIC では，SNR が低下するとともに，正答率も低下していることがわかる．iGSVD-MUSIC では，SNR が-10 dB までは正解率が 100 %に近く，-20 dB の場合でも 80 %を超えている．このことから，雑音耐性が非常に高いことがわかる．また，レンジを狭帯域に設定した SEVD-MUSIC では，SNR が 0 dB 以下の場合には広帯域の SEVD-MUSIC と比べ正解率が高いが，SNR が 0 dB 以上の場合には 70 %程度となっている．これは，周波数レンジを狭くすることで雑音を抑制するため，SNR が低い場合には有効的であるが，笛の音に応じてレンジを設定したため，音源が声の場合の正解率が低下し，SNR が高い場合でも 100 %にはならない．AFRF-MUSIC では，雑音を抑制し，さらに複数の種類の音源がある場合でも，それぞれに応じた周波数レンジが設定されるため，iGSVD-MUSIC よりは正解率は低いものの，近い値となっていることがわかる．

また，処理遅延を計測することで，リアルタイム性の評価を行った．各手法の処理遅延を Table 1 に示す．iGSVD が最も遅延が大きく，3 s 以上の遅延が発生していた．また，広帯域に設定した場合と狭帯域に設定した場合の SEVD-MUSIC について，遅延は使用する周波数レンジの大きさに比例して大きくなることがわかっている [5]．今回の実験の場合でも，広帯域に設定した場合と比べ，狭帯域に設定した場合は遅延が少なく，リアルタイム性が高くなっている．AFRF の場合，f_w を 500 Hz と設定しているため，計算コストは狭帯域に設定した SEVD-MUSIC とほぼ変わらず，遅延もほぼ同程度である．

これらの結果から，AFRF-MUSIC の定位性能は雑音耐性の高い iGSVD-MUSIC と同程度，遅延は計算量の少ない，狭帯域に設定した SEVD-MUSIC と同程度であることから，雑音耐性・リアルタイム性の両者を満たす音源定位手法として有用性が確認できた．本実験では単一のレンジを用いたが，複数のレンジを同時に使用するといった拡張もできるため，同時に発話されている複数音源への対応も可能であり，今後検討を行っていく予定である．

4 おわりに

本稿では，リアルタイム性と雑音耐性の両者を満たす音源定位手法の開発を目的に，MUSIC 法に基づく，アクティブ周波数レンジフィルタを用いた音源定位手法を提案した．最も計算コストの小さい SEVD-MUSIC をベースに，単純な四則演算のみで構成されるアクティブ周波数レンジフィルタを適用した AFRF-MUSIC を開発し，リアル

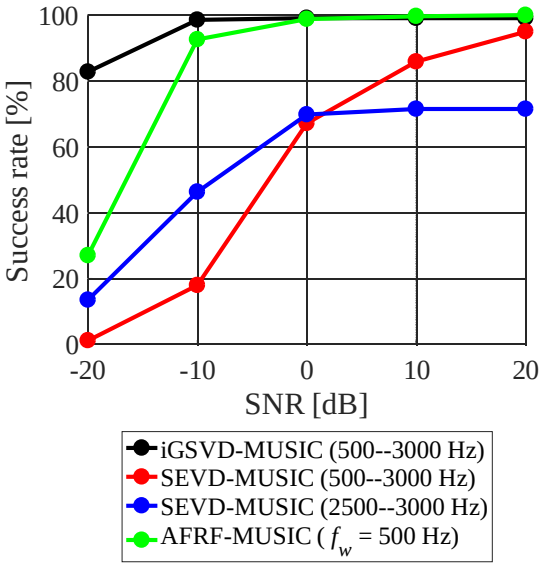

Fig. 5: Success rate of localization.

Table 1: Success rate of localization.

Algorithm	Delay [s]
iGSVD-MUSIC	3.219
SEVD-MUSIC (500–3000 Hz)	0.533
SEVD-MUSIC (2500–3000 Hz)	0.194
AFRF-MUSIC	0.204

タイム性と雑音耐性を確保した．評価実験により，その有用性が確認できた．しかし，本手法では f_s フレーム前の信号を雑音として扱うため，目的音源が f_s フレームを超えた場合，差分として目的音源が検出されない場合がある．また，UAV の飛行中のような環境が変わりやすい場面では，環境やノイズによるパラメータチューニングが難しく，定位性能が低下する可能性がある．今後は，パラメータチューニングの自動化について検討を行っていく予定である．

謝辞

本研究は，JSPS 科研費 16H02884，16K00294,，17K00365 および，JST ImPACT タフロボティクスチャレンジの助成をうけた．

参考文献

[1] K. Nakadai, T. Lourens, H. G. Okuno and H. Kitano, “Active audition for humanoid”, Proceedings of 17th National Conference on Artificial Intelligence (AAAI-2000), pp. 832-839, 2000.

[2] K. Okutani, T. Yoshida, K. Nakamura and K. Nakadai, “Outdoor auditory scene analysis using a moving microphone array embedded in a quadrocopter”, Proceedings of the IEEE/RSJ International Conference on Robots and Intelligent Systems (IROS), pp. 3288-3293, 2012.

[3] T. Ohata, K. Nakamura, T. Mizumoto, T. Tezuka and K. Nakadai, “Improvement in outdoor sound source detection using a quadrotor-embedded microphone array”, Proceedings of the IEEE/RSJ International Conference on Robots and Intelligent Systems (IROS), pp. 1902-1907, 2014.

[4] K. Hoshiba, O. Sugiyama, A. Nagamine, R. Kojima, M. Kumon, K. Nakadai, “Design and assessment of sound source localization system with a UAV-embedded microphone array”, Journal of Robotics and Mechatronics, vol. 29, No. 1, pp. 154-167, 2017.

[5] K. Hoshiba, 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.

[6] C. Xu, X. Xiao, S. Sun, W. Rao, E. S. Chng, H. Li, “Weighted Spatial Covariance Matrix Estimation for MUSIC based TDOA Estimation of Speech Source”, Proceedings of the INTERSPEECH 2017, pp. 1894-1898, 2017.

[7] 干場功太郎, 若林瑞保, 鷲崎海, 石木隆洋, 公文誠, Daniel Gabriel, 中臺一博, 奥乃博, “UAV 搭載マイクロホンアレイを用いた組み込みシステムによる音源探査性能の評価”, 第 35 回日本ロボット学会学術講演会, pp. 1-4, 2017.

[8] R. O. Schmidt, “Multiple emitter location and signal parameter estimation”, IEEE Transactions on Antennas and Propagation, Vol. 34, No. 3, pp. 276-280, 1986.

[9] K. Nakadai, T. Takahashi, H. G. Okuno, H. Nakajima, Y. Hasegawa and H. Tsujino, “Design and Implementation of Robot Audition System ’HARK’ – Open Source Software for Listening to Three Simultaneous Speakers”, Advanced Robotics, Vol. 24, No. 5-6, pp. 739-761, 2010.