論文

MUSICに基づくパーティクルフィルタによる 2音源追尾の高速化

松田 尚頼, 陶山 健仁

東京電機大学 工学部 電気電子工学科, 東京都足立区千住旭町 5

Acceleration of Two Sound Source Tracking Using MUSIC-Based Particle Filters

Naoyori Matsuda and Kenji Suyama

Department of Electrical and Electronic Engineering, School of Engineering, Tokyo Denki University

5 Senju-Asahi-cho, Adachi-ku, Tokyo 120-8551, Japan

E-mail: suyama@cck.dendai.ac.jp

Abstract In this paper, MUSIC (multiple signal classification) algorithm is applied for a sound source tracking using particle filters. MUSIC algorithm is a high-resolution sound source localization method but has high computational costs. Therefore, a reduction of computational cost is attempted by using the MUSIC spectrum function as a likelihood function of particle filters. When there are two sound sources, however, a particle intensification problem arises simultaneously. Therefore, a method to overcome this problem is also proposed. Several experimental results in a reverberant environment show that the proposed method can track two sound sources in real time.

Keywords: microphone array, sound source tracking, particle filter

1. はじめに

遠隔会議システムやロボット聴覚などへの応用を目的として、マイクロホンアレー [1] による音源追尾技術が注目されている [2]–[7]。音源追尾では、通常、時間毎の定位に相関行列などの統計量を用いるが、音源の位置は常に変化しているため、短い時間区間での少数の観測サンプルのみから統計量を推定する必要がある。しかし、このようにして推定した統計量には、部屋の残響や周囲の雑音、他の音源との干渉による精度の劣化が生じやすい。文献 [8] では、ベイズ推定の枠組みで尤度関数を算出する際に、マイクロホンアレーの受信信号に加え、画像信号を用い、騒音源の存在する

環境で単一話者の追尾に成功している。文献 [2], [5], [6] では、多数の粒子（パーティクル）により時系列モデルを表現する手法であるパーティクルフィルタ [9] を用いて、安定した音源追尾に成功している。

パーティクルフィルタによる音源追尾では、観測モデルを構成するために最尤推定法 [10], MUSIC (multiple signal classification) [11], 遅延和アレー法 [12] などの音源定位手法を要する。中でも MUSIC は、代表的な音源定位法の 1 つであるものの、高い演算量を必要とするため、実時間処理困難な場合がある。文献 [13], [14] では、並列処理により実時間処理を達成している。

MUSIC が高演算負荷である要因として、信号部分

空間を推定するために相関行列の固有値分解が必要なこと、評価関数が急峻なため細かい間隔でのピークサーチが必要なことの2つがある。信号部分空間の算出に対しては、PAST (projection approximation subspace tracking) [15] を用いて演算量を低減する手法が提案されている。一方、ピークサーチにおける演算負荷についても無視できない。文献 [16] では定位を行う時間フレームを、文献 [17] では MUSIC スペクトルを算出する周波数を間引いて、演算量を低減している。

それに対して本研究では、PASTに加えて、MUSIC スペクトル関数を尤度関数として用い、ピークサーチにおける演算負荷の低減を行う。そうした場合、ピークから離れた位置にあるパーティクルはリサンプリングによって消滅するため、多数のパーティクルを用いる必要がない。そのため、パーティクル数を少なくでき、MUSIC スペクトルの計算点数の削減が可能である。

しかしながら、2つの音源が存在する状況では、次の2点が問題となる。

- (i) 音源が発話状態から無音状態へ移行した際に、他方の音源に対するパーティクルの集中化
- (ii) 2つの音源方向をパーティクルから推定する方法

(i) は、無音状態の音源が存在した場合、リサンプリング処理で発話状態の音源方向にパーティクルが集中する問題である。一旦発話状態にある音源に対してパーティクルが集中すると、無音状態の音源が発話状態に復帰しても、パーティクルがその音源の追尾を再開することは困難となる。(ii) は、音源数が2の場合、各パーティクルの帰属先音源が不明となるため、一般にパーティクルフィルタで用いられる加重平均による状態推定が困難となる問題である。問題(i)に対して、文献 [18] では遅延和アレー法を用いて大まかな音源方向を推定し、発話区間への復帰時にパラメータの再設定を行っている。本研究においても、同様にパーティクルの再設定を行う。問題(ii)に対しては、パーティクルの分布に混合正規分布を当てはめ、EM アルゴリズム [19] により各音源方向を推定する。

実環境実験により、提案法が2音源追尾可能であり、かつ実時間処理で動作可能であることを示す。

2. 問題設定

図1のように、空間中に2つの音源が $\theta_i(n)$, $i = 1, 2$ 方向に存在し、それらの音響信号 $s_i(n)$ を間隔 d で等

図1 音源追尾モデル

Fig. 1 Sound source tracking model

間隔に配置した M 個のマイクロホンで受音する。ここで、 n は離散時間である。マイクロホン1を基準としたマイクロホン m の受音信号 $x_m(n)$ は、周波数領域において

$$X_{t,m}(k) = \sum_{i=1}^2 S_{i,t}(k) e^{-j\omega_k(m-1)\tau_{i,t}} + \Gamma_{t,m}(k) \quad (1)$$

となる。ここで、 t はフレーム番号、 k は周波数番号、 $S_{i,t}(k)$ は $s_i(n)$ の複素振幅、 $\tau_{i,t} = d \sin \theta_{i,t} / c$ は到達時間差、 $\Gamma_{t,m}(k)$ はマイクロホン m の観測雑音、 c は音速である。ただし、 $\omega_k = 2\pi f_s k / L$ 、 $\theta_{i,t} = \theta_i(Lt)$ であり、 f_s はサンプリング周波数、 L はフレーム長である。また、 $\Gamma_{t,m}(k)$ は $S_{i,t}(k)$ と無相関であり、各マイクロホンで独立であると仮定する。式(1)は、ベクトル表記を用いて次式のように書き直せる。

$$\mathbf{X}_t(k) = \sum_{i=1}^2 S_{i,t}(k) \mathbf{a}_k(\tau_{i,t}) + \mathbf{\Gamma}_t(k) \quad (2)$$

$$= [X_{t,1}(k), X_{t,2}(k), \dots, X_{t,M}(k)]^T \quad (3)$$

ここで、 $\mathbf{T}$ は転置である。また、 $\mathbf{a}_k(\tau_{i,t})$ はステアリングベクトル、 $\mathbf{\Gamma}_t(k)$ は雑音ベクトルであり、次式で表せる。

$$\mathbf{a}_k(\tau_{i,t}) = [1, e^{-j\omega_k \tau_{i,t}}, \dots, e^{-j\omega_k (M-1)\tau_{i,t}}]^T \quad (4)$$

$$\mathbf{\Gamma}_t(k) = [\Gamma_{t,1}(k), \Gamma_{t,2}(k), \dots, \Gamma_{t,M}(k)]^T \quad (5)$$

複数音源追尾問題は、 $\mathbf{X}_t(k)$ から $\theta_{i,t}$ の推定値 $\hat{\theta}_{j,t}$, $j = 1, 2$ を求める問題である。音源がすれ違う際などには、各音源に対応する定位結果が入れ替わる場合があるが、本研究ではこれを考慮しない。音源の入れ替わりに対処した手法には、例えば文献 [20] がある。

3. 音源定位手法

本手法で用いる音源定位法である MUSIC [11] と、固有値分解なしに信号部分空間を逐次的に推定する PAST [15] について簡単にまとめる。

3.1 MUSIC

MUSIC は、相関行列 $\mathbf{R}_t(k) = E[\mathbf{X}_t(k)\mathbf{X}_t^H(k)]$ の固有ベクトルによって張られる信号部分空間 $\mathbf{Q}_S(t, k)$ と雑音部分空間 $\mathbf{Q}_N(t, k)$ に基づく手法である。ここで、 $E[\cdot]$ は期待値演算、 H はエルミート転置を表す。2つの部分空間は、互いに直交補空間の関係にある。MUSIC は、 $\mathbf{a}_k(\tau)$ と $\mathbf{Q}_N(t, k)$ の直交性を次式の MUSIC スペクトル関数 $P_{\text{MU}}(k, \tau)$ によって評価し、ピークに対応する τ を用いて音源方向を推定する。

$$P_{\text{MU}}(k, \tau) = \frac{\mathbf{a}_k^H(\tau)\mathbf{a}_k(\tau)}{\mathbf{a}_k^H(\tau)\mathbf{Q}_N(t, k)\mathbf{Q}_N^H(t, k)\mathbf{a}_k(\tau)} \quad (6)$$

3.2 PAST

音源が移動する場合、それに伴い信号部分空間も変動する。そのため、観測毎に固有値分解が必要であり、高負荷な演算となる。そこで、固有値分解なしで $\mathbf{Q}_S(t, k)$ を逐次更新により求める PAST を用いる。PAST は、次式の評価関数 $J(\mathbf{Q}_S(t, k))$ を最小にする $\hat{\mathbf{Q}}_S(t, k)$ を、毎フレーム 1 ステップ更新し、求める。

$$J(\mathbf{Q}_S(t, k)) = \sum_{t'=1}^t \beta^{t-t'} \|\mathbf{X}_{t'}(k) - \mathbf{Q}_S(t, k)\mathbf{h}(t', k)\|^2 \quad (7)$$

ここで、 $\mathbf{h}(t', k) = \hat{\mathbf{Q}}_S^H(t' - 1, k)\mathbf{X}_{t'}(k)$ 、 β は忘却係数である。PAST により推定した $\hat{\mathbf{Q}}_S(t, k)$ から $\hat{\mathbf{Q}}_N(t, k)\hat{\mathbf{Q}}_N^H(t, k) = \mathbf{I} - \hat{\mathbf{Q}}_S(t, k)\hat{\mathbf{Q}}_S^H(t, k)$ を算出し、 $P_{\text{MU}}(k, \tau)$ に用いる。ここで、 $\mathbf{I}$ は単位行列である。

4. パーティクルフィルタによる音源追尾

本章では、MUSIC に基づくパーティクルフィルタについて説明する。また、尤度関数として MUSIC スペクトル関数を用いた場合、そのままでは複数音源追尾が困難であることを示す。

4.1 パーティクルフィルタ

パーティクルフィルタは複数のパーティクルから成り、各パーティクルが個別に対象の追跡を行う。フ

レーム t において、 u 番目のパーティクル $\boldsymbol{\alpha}_t^{[u]}$ 、 $u = 1, 2, \dots, U$ は、次式のように位置 $\theta_t^{[u]}$ をもつ。

$$\boldsymbol{\alpha}_t^{[u]} = [\theta_t^{[u]}]^T \quad (8)$$

また、 $\theta_t^{[u]}$ の評価値を重み $w_t^{[u]}$ で表す。

4.1.1 重みの評価

パーティクルの位置 $\theta_t^{[u]}$ を尤度関数により評価し、重み $w_t^{[u]}$ を与える。本研究では、尤度関数として次式の広帯域 MUSIC スペクトル関数を用いる。

$$q_t^{[u]} = \sum_k P_{\text{MU}}(k, \tau(\boldsymbol{\alpha}_t^{[u]})) \quad (9)$$

ここで、 $\tau(\boldsymbol{\alpha}_t^{[u]}) = d \sin \theta_t^{[u]} / c$ である。この関数は真の確率密度関数ではないが、本研究では擬似的に、この値を $\boldsymbol{\alpha}_t^{[u]}$ の尤度とみなした [21]。 $w_t^{[u]}$ は次式の通り、 $q_t^{[u]}$ を正規化したものを用いる。

$$w_t^{[u]} = \frac{q_t^{[u]}}{\sum_{\ell=1}^U q_t^{[\ell]}} \quad (10)$$

4.1.2 リサンプリング

均一な重みのパーティクルを生成するために、リサンプリングを行う。 $w_t^{[u]}$ に基づいて、重みの大きいパーティクルは数を増やし、重みの小さいパーティクルは数を減らす。

4.1.3 状態遷移

システムモデルに従って、現在のパーティクル $\boldsymbol{\alpha}_t^{[u]}$ から次フレームの $\boldsymbol{\alpha}_{t+1}^{[u]}$ を予測する。本研究では次式のように、状態遷移分布 $p(\theta_{t+1}^{[u]} | \theta_t^{[u]})$ として平均 $\theta_t^{[u]}$ 、分散 σ_p^2 の正規分布を用いる。

$$p(\theta_{t+1}^{[u]} | \theta_t^{[u]}) = \mathcal{N}(\theta_{t+1}^{[u]}; \theta_t^{[u]}, \sigma_p^2) \quad (11)$$

4.2 複数音源追尾における問題点

図 2(I) に示すように、2つの音源がいずれも発話状態にある場合には、それぞれの音源方向に対して MUSIC スペクトルのピークが生じ、パーティクルも各音源方向付近に存在する。しかしながら、一方の音源が無音状態へと移行した場合、リサンプリング処理によってパーティクルが他方の音源に対して集中する(図 2(II))。その後は、無音状態の音源が発話状態に復帰したとしても、その音源の存在する方向付近にパーティクルが移動しない限り、それぞれの音源を追尾することは困難となる(図 2(III))。

図2 無音区間前後におけるパーティクルの状態
Fig. 2 State transition of particles when a silent period is included

実際に、2音源に対して追尾を行ったときのパーティクルの分布を、図3(a)に示す。正規乱数の分散は $\sigma_p^2 = 1 [\text{deg}^2]$ 、パーティクル数は $U = 20$ とした。MUSICスペクトルは急峻であり、また、ピーク値が音源により大きく異なるため、図3(a)のように、リサンプリングによってパーティクルが1音源に対して集中化する。加えて、集中化した音源が無音状態となるたび、他方の音源に対して一斉に移動することもわかる。次に、集中化を回避するために分散を $\sigma_p^2 = 30 [\text{deg}^2]$ と大きくとった場合の分布を図3(b)に示す。この場合、パーティクルは両方の音源に対して分布している。しかしながら、適当なクラスタリング法により各クラスタの中心を抽出したとしても、図4(a)のように、 σ_p^2 が大きいため追尾結果が大きく乱れてしまう。すなわち、通常の状態遷移では、複数音源追尾は困難である。一方、 $\sigma_p^2 = 1 [\text{deg}^2]$ としたまま一部のパーティクルを毎フレーム初期化し、大域的な探索能力を補う方法も考えられる。初期化用パーティクルを10個追加した場合の追尾結果を、図4(b)に示す。大域的探索が可能となり、2音源を追尾できているが、0[s]、8[s]付近において誤推定が生じている。これは、広帯域MUSICスペクトルが一般に多数のピークをもつためである。

5. 提案法

図5に提案法の流れを示す。まず、マイクロホンで受信した信号 $x_m(n)$ を L 点DFTし、周波数領域の信号 $X_{t,m}(k)$ へ変換する。そして、PASTを用いて雑音部分空間 $\hat{\mathbf{Q}}_N(t, k)$ を算出し、パーティクルフィルタ

図3 パーティクルの分布
Fig. 3 Distribution of particles

図4 追尾結果

Fig. 4 Tracking results

に用いる。パーティクルフィルタでは、パーティクルの重み $w_t^{[u]}$ を評価し、遅延和アレー法により推定した音源数[18]に従い、推定音源数が前フレーム以下の場合リサンプリングを、増加した場合パーティクルの再設定を行う。そして、EMアルゴリズムを用いてパーティクルに対して混合正規分布を当てはめ、音源方向 $\hat{\theta}_{j,t}$ を推定する。最後に、状態遷移モデルに基づきパーティクルの状態遷移を行う。

5.1 パーティクルの再設定

図6に遅延和アレー法によるパーティクルの再設定の流れを示す。本手法では空間分解能を向上するために、用いるマイクロホンを間引き、マイクロホン間隔を広げている。まず、 $\mathbf{X}_t(k)$ から遅延和アレー法[12]を用いて帯域毎のパワースペクトル $P_{DS,k}(\tau)$ を算出し、全帯域分加算する。加算したスペクトル $P'_{DS}(\tau)$ を粗くピークサーチし、音源方向の候補 $\tilde{\theta}_{j,t}$ を算出する。しかしながら、マイクロホン間隔を広げた場合、空間エイリアシングによる偽りのピークが生じる場合がある。そこで、音源に対応するピークを判別するために音源選択を行う。音源選択では、 $\hat{\mathbf{Q}}_S(t, k)$ の列ベクトルと、 $\tilde{\theta}_{j,t}$ に対応するステアリングベクトル $\mathbf{a}_k(\tilde{\tau}_{j,t})$ の内積の2乗和 $\psi(\tilde{\tau}_{j,t}) = \mathbf{a}_k^H(\tilde{\tau}_{j,t}) \hat{\mathbf{Q}}_S(t, k) \hat{\mathbf{Q}}_S^H(t, k) \mathbf{a}_k(\tilde{\tau}_{j,t})$ を算出し、全帯域分平均した $\psi(\tilde{\tau}_{j,t})$ が閾値 C_{th} を超えた場合、音源に対応するピークと判定する。音源選

図5 提案法の流れ

Fig. 5 Block diagram of the proposed method

図6 パーティクルの再設定

Fig. 6 Resetting procedure of particles

択した後、 $\tilde{\theta}_{j,t}$ をパーティクルの再設定に用いる。また、フレーム t における $\tilde{\theta}_{j,t}$ の個数を推定音源数 $\hat{N}_t$ とする。パーティクルの再設定は $\hat{N}_{t-1} < \hat{N}_t$ のとき、各音源方向に対してパーティクルを $U/2$ 個ずつ、平均 $\tilde{\theta}_{j,t}$ 、分散 σ_r^2 の次式の正規分布に基づき再設定する。

$$p(\theta_t^{[u]} | \tilde{\theta}_{j,t}) = \mathcal{N}(\theta_t^{[u]}; \tilde{\theta}_{j,t}, \sigma_r^2) \quad (12)$$

$\hat{N}_t > 2$ のときは、 $\psi(\tilde{\tau}_{j,t})$ が大きい 2 個の $\tilde{\theta}_{j,t}$ を音源方向とみなし、 $\hat{N}_t = 2$ とする。その他の場合はパーティクルの再設定を行わず、リサンプリング処理を行う。

5.2 EM アルゴリズム

各パーティクルは、平均 μ_j 、分散 σ_j^2 の g 要素をもつ混合正規分布から独立に発生したと仮定する。ここで、 g は分割幅 B で作成した $\theta_t^{[u]}$ のヒストグラムのピーク数である。状態推定後もなお不完全データであるパーティクルから各音源方向を推定するために、EM

表1 実験条件

Condition	Value
Sampling frequency f_s	8 kHz
Number of microphones M	16
Microphone spacing d	40 mm
Frame length L	256
Frame shift	128
Frequency band	2 kHz–4 kHz

アルゴリズムにより最尤推定を行う。EM アルゴリズムは、次式の混合正規分布の μ_j 、 σ_j^2 、混合比 ρ_j を推定する。

$$F(\theta) = \sum_{j=1}^g \rho_j \left\{ \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(\theta - \mu_j)^2}{2\sigma_j^2}\right) \right\} \quad (13)$$

推定した μ_j を推定音源方向 $\hat{\theta}_{j,t}$ とする。

6. 実環境実験

提案法の性能を検証するため、実環境実験を行った。音源にはスピーカを用い、音声を再生しながら実際に人が持って移動した。移動パターンによらない追尾性能の検証のために、音源間隔が狭い場合、広い場合、音源が接近し交差する場合について実験を行った。実験環境は、大きさが 26.0 [m] × 9.0 [m] × 2.5 [m] の通常の会議室である。騒音レベルは 38.4 [dB]、残響時間は 0.51 [s] であった。実験条件を表 1 に示す。

音源追尾性能の評価には、次式の RMSE (root mean square error) ε を用いた。

$$\varepsilon = \sqrt{\frac{1}{2} \sum_{i=1}^2 |\hat{\theta}_{\Pi(i),t} - \theta_{i,t}|^2} \quad (14)$$

ここで、 $\Pi(i)$ は真値の音源番号 i を推定方向番号に対応させる関数、 $\theta_{i,t}$ は真値であり、カメラアレーを用いた動画像計測により求めた。カメラアレーは、4 つのカメラからなり、フレームレートは 30 [Hz] である。 ε は音声パワーが比較的大きいフレームのみで評価を行った。演算時間の測定には、Intel® Core™ i3-2130 3.4 [GHz]、メモリ 4 [GByte] を搭載した PC を使用した。

比較手法として、4.2 節で用いた一部のパーティクルを毎フレーム初期化する手法（比較手法 1）と、文献 [17] の MUSIC スペクトルのピークサーチによる手法（比較手法 2）を用いた。提案法と 2 つの比較手法は、いずれも MUSIC に基づく音源追尾法である。比較手法 1 は、演算負荷と集中化回避の有効性を、提案

表2 提案法1と比較手法1のパラメータ
Table 2 Parameters used in the proposed method and the comparison method 1

Condition	Value	
	Prop.	Comp.1
Forgetting factor (for PAST) β	0.9	0.9
Particle filter		
Number of particles U	20	20
Number of initialized particles	-	10
Variance for state transition σ_p^2	1 deg ²	1 deg ²

表3 比較手法2 [17]のパラメータ
Table 3 Parameters used in the comparison method 2

Condition	Value
Forgetting factor (for correlation matrices)	0.96
Number of frequency components	20
Peak search step size	1°

法と比較するために用いた。比較手法2は、MUSICの定位精度のリファレンスとして用いた。

6.1 パラメータの設定

表2に実験に用いた提案法(Prop.)と比較手法1(Comp.1)のパラメータを示す。比較手法1では、 $U = 20$ に加えて10個のパーティクルを用いているが、これは5.1節の再設定法に要する演算負荷をパーティクル数の増加に利用するという考えによるものである。提案法、比較手法1ともに、パーティクルのクラスタリング法としてEMアルゴリズムを用いた。本実験で用いる音源の移動パターンは、交差する場合を除き、最も接近した場合でも音源同士の間隔が約20°であるため、要素分布数 g の推定に用いるヒストグラムの分割幅は $B = 10^\circ$ とした。提案法における再配置の分散は、 $B = 10^\circ$ 幅に99.73%の確率($\tilde{\theta}_{j,t} \pm 3\sigma_r$)で収まるよう $\sigma_r^2 = 2.8 [\text{deg}^2]$ とした。音源選択の閾値は、表2と同じ条件で、 C_{th} に対するRMSEを評価し、RMSEが最小となった $C_{th} = 0.46$ とした。

表3に提案法2(Comp.2)のパラメータを示す。忘却係数は、事前実験により求めた値である。

6.2 追尾性能の評価

音源と移動のパターン8種類に対し、追尾実験を各100回試行した際の平均RMSE $\bar{\varepsilon}$ により追尾性能を評価した。表4に実験結果を示す。表4において、 $\varepsilon_{\min}$ 、 $\varepsilon_{\max}$ は、それぞれ100試行中の最小RMSEと最大RMSE、 σ は100試行におけるRMSEの標準偏

差である。表4より、どのパターンにおいても提案法の平均RMSEが最も小さかった。さらに、全てのパターンにおいて、提案法の最大RMSEは比較手法1の最大RMSE、比較手法2のRMSEより小さかった。提案法の標準偏差は1.0°未満に抑えられ、試行中のばらつきは比較手法1と同程度もしくはそれ以下に抑えることができた。これは、パーティクルの再設定による集中化回避の効果である。信号長1[s]当たりの平均演算時間は、提案法が0.42[s]、比較手法1が0.50[s]、比較手法2が0.86[s]であった。ゆえに、いずれの手法も実時間処理可能である。通常のMUSICでは、1°間隔でのピークサーチにおけるMUSICスペクトルの計算点数は181点なのに対し、MUSICスペクトル関数を尤度関数として用いた場合は $U = 20$ 点であり、1/9以下の演算量で追尾できるため、提案法と比較手法1の演算時間は小さかった。標準固有値分解を行う比較手法2にPASTを適用した場合も、演算時間を0.37[s]まで高速化可能であるが、比較手法2は S/N の高い一部の帯域のみを用いるため、広帯域雑音が重畳した場合の雑音耐性は提案法の方が高いと考えられる。また、比較手法1と比べることによって、提案法におけるパーティクルの再設定の演算量は、MUSICスペクトル10点分に満たないことも示された。

試行中RMSEが最も小さかったときの追尾結果を図7-14に示す。提案法と比較手法1の追尾結果に一定値となっている区間があるのは、発話音源数が2未満と推定された場合、以前の結果をホールドしているためである。また、比較手法2のみプロットの仕方が異なるが、比較手法2は定位結果のばらつきが激しく、線で表示すると見辛いためであり、手法による結果の出力方法に違いはない。比較手法1がフレーム毎にパーティクルの再設定をしているのに対し、提案法は遅延とアレーによる音源数推定により効果的に再設定ができているため、いずれのパターンに対しても良好な追尾を行えている。比較手法1と比べると追尾の軌跡が荒れているが、真値から大きく外れているフレームは無音区間であるため、表4の結果からわかるように、発話状態の音源を正しく追尾できている。特に、図13-14のパターンのように、音源が近接し、交差するパターンにおいても、提案法が最も良好に追尾できている。1音源と推定された区間に限定すれば、比較手法1は、非常に滑らかな軌跡で追尾を行えているが、2音源区間におけるパーティクルの集中化が完全には回避できず、追尾結果が激しく乱れている。また、初期化を行った場合でも発話状態に復帰した音源を検出できず、2音源区間においても1音源しか追え

(a) Proposed method

(b) Comparison method 1

(c) Comparison method 2

図7 追尾結果 (パターン 1)

Fig. 7 Tracking results in the source pattern 1

(a) Proposed method

(b) Comparison method 1

(c) Comparison method 2

図8 追尾結果 (パターン 2)

Fig. 8 Tracking results in the source pattern 2

(a) Proposed method

(b) Comparison method 1

(c) Comparison method 2

図9 追尾結果 (パターン 3)

Fig. 9 Tracking results in the source pattern 3

(a) Proposed method

(b) Comparison method 1

(c) Comparison method 2

図10 追尾結果 (パターン 4)

Fig. 10 Tracking results in the source pattern 4

図 11 追尾結果 (パターン 5)

Fig. 11 Tracking results in the source pattern 5

図 12 追尾結果 (パターン 6)

Fig. 12 Tracking results in the source pattern 6

図 13 追尾結果 (パターン 7)

Fig. 13 Tracking results in the source pattern 7

図 14 追尾結果 (パターン 8)

Fig. 14 Tracking results in the source pattern 8

表4 実験結果

Table 4 Experimental results

Pattern	Prop.				Comp.1				Comp.2 $\bar{\varepsilon}$ [°]
	$\bar{\varepsilon}$ [°]	$\varepsilon_{\min}$ [°]	$\varepsilon_{\max}$ [°]	σ [°]	$\bar{\varepsilon}$ [°]	$\varepsilon_{\min}$ [°]	$\varepsilon_{\max}$ [°]	σ [°]	
1	3.82	3.33	4.69	0.24	7.97	5.57	19.82	1.90	6.24
2	3.04	2.40	3.85	0.32	5.33	3.15	9.01	1.40	4.54
3	4.49	3.48	5.80	0.52	8.80	6.88	10.77	0.93	7.05
4	5.16	4.27	6.04	0.33	8.13	7.04	10.28	0.65	7.75
5	8.48	6.67	10.36	0.99	16.00	13.77	17.63	0.60	13.54
6	7.29	6.13	9.01	0.58	15.28	12.87	16.58	0.73	9.03
7	5.65	4.57	6.49	0.47	9.38	8.48	13.24	0.55	12.98
8	4.83	4.23	5.64	0.31	6.58	5.08	11.94	1.24	9.90

図15 パターン1におけるパーティクルの分布

Fig. 15 Distribution of particles in the source pattern 1

ていない区間（ホールドされている区間）が多く存在する。比較手法2は、発話区間においては各時刻において高精度に定位できている。しかし、 $\pm 60^\circ$ 付近など図の範囲外にも多く定位結果が存在しており、無音区間においては3手法中もっとも定位結果のばらつきが大きい。

図7-14(a), (b)のように、提案法と比較手法1では2つの定位結果が互いに近接する場合があるが、これはパーティクルの集中化が回避できなかったためではなく、混合正規分布の最尤推定に原因があると考えられる。図15に示す4[s]付近におけるパーティクルの分布を見ると、パーティクルが存在しない位置に正規分布の平均をたびたび推定していることがわかる。この現象は、正規分布よりも尖度の高い分布を用いれば解決できると考えられる。

6.3 パラメータの検証

提案法の状態遷移における分散 σ_p^2 とパーティクル数 U について、検証を行った。図16に σ_p^2 の検証結果を示す。図16のように、 σ_p^2 が増加するにつれ平均

図16 分散 σ_p^2 の影響Fig. 16 Effect of variance σ_p^2

RMSE $\bar{\varepsilon}$ は増大した。一方、RMSEの標準偏差 σ は、 σ_p^2 によらず約 0.4° で大域的にほぼ変化しなかった。 $\sigma_p^2 \approx 0$ においても $\bar{\varepsilon}$ が増大しない理由は、提案法では遅延とアレー法による再配置を併用しているため、音源が高速に移動した場合でもパーティクルが追従できるためと考えられる。以上より、 $\sigma_p^2 \in [0, 1]$ 程度にしておけば、多くの音源パターンについて良好な結果が得られると言える。

図17は、 U の検証結果である。図17より、 U を増加するほど平均RMSE $\bar{\varepsilon}$ を減少させることができる。一方、パーティクル数が $U = 6$ 程度でも、それ以上用いた場合と追尾精度に大きな差がない。これは、パーティクルフィルタをピークサーチであるかのように用いるため、ピークサーチにおける計算点数の削減が可能となる。本実験の条件では、 $U = 70$ でも実時間処理可能であった。しかしながら、20以上のどの U でも $\bar{\varepsilon}$ が同等であり、信号長当たりの演算時間は U に比例して増加するため、パーティクルは $U = 20$ 個あれば十分であると言える。

図 17 パーティクル数 U の影響

Fig. 17 Effect of the number of particles U

7. まとめ

本研究では、パーティクルフィルタによる音源追尾への MUSIC の適用を検討した。パーティクルフィルタの尤度関数に高分解能音源定位が可能な MUSIC スペクトル関数を用い、MUSIC で必要な固有ベクトルの算出には逐次推定により演算量の低減が可能な PAST を用いた。音源が複数存在する場合の問題点である無音区間におけるパーティクルの集中化回避のために、遅延和アレー法による再設定を行った。不完全データであるパーティクルの分布に対し混合正規分布を当てはめ、各音源方向を最尤推定した。実環境実験により、提案法が 2 音源追尾を実時間で示した。

参考文献

- [1] 大賀寿郎, 山崎芳男, 金田豊: マイクロホン系における信号処理, 音響システムとデジタル処理, pp. 173–218, 電子情報通信学会, 1995.
- [2] A. Quinlan, M. Kawamoto, Y. Matsusaka, H. Asoh and F. Asano: Tracking intermittently speaking multiple speakers using a particle filter, EURASIP J. Audio, Speech, Music Process., Vol. 2009, No. 8, pp. 1–11, Jan. 2009.
- [3] 大和田昇, 陶山健仁: 多峰性評価関数に対する逐次解法に基づく複数移動音源追尾, 信学論 (A), Vol. J93-A, No. 4, pp. 248–259, Apr. 2010.
- [4] 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, Apr. 2010.
- [5] M. F. Fallon and S. J. Godsill: Acoustic source localization and tracking of a time-varying number of speakers, IEEE Trans. Audio, Speech, Lang. Process., Vol. 20, No. 4, pp. 1409–1415, May 2012.
- [6] A. Kijima, Y. Hioka and N. Hamada: Tracking of multiple moving sound sources using particle filter for arbitrary microphone array configurations, Proc. IEEE Int. Symp. Intell. Signal Process. Commun. Syst. (ISPACS 2012), pp. 108–113, Nov. 2012.
- [7] D. Pavlidis, A. Griffin, M. Puigt and A. Mouchtaris: Real-time multiple sound source localization and counting using a circular microphone array, IEEE Trans. Audio, Speech, Lang. Process., Vol. 21, No. 10, pp. 2193–2206, Oct. 2013.
- [8] H. Asoh, I. Hara, F. Asano and K. Yamamoto: Tracking human speech events using a particle filter, Proc. ICASSP2005, Vol. II, pp. 1153–1156, Mar. 2005.
- [9] 北川源四郎: モンテカルロ・フィルタおよび平滑化について, 統計数理, Vol. 44, No. 1, pp. 31–48, 1996.
- [10] M. I. Miller and D. R. Fuhrmann: Maximum-likelihood narrow-band direction finding and the EM algorithm, IEEE Trans. Acoust., Speech, Signal Process., Vol. 38, No. 9, pp. 1560–1577, Sept. 1990.
- [11] R. O. Schmidt: Multiple emitter location and signal parameter estimation, IEEE Trans. Antennas Propag., Vol. AP-34, No. 3, pp. 276–280, Mar. 1986.
- [12] J. C. Chen, K. Yao and R. E. Hudson: Source localization and beamforming, IEEE Signal Process. Mag., Vol. 19, No. 2, pp. 30–39, Mar. 2002.
- [13] F. Asano, M. Goto, K. Itou and H. Asoh: Real-time sound source localization and separation system and its application to automatic speech recognition, Proc. Eurospeech 2001, pp. 1013–1016, Sept. 2001.
- [14] K. Nakadai, H. Nakajima, M. Murase, H. G. Okuno, Y. Hasegawa and H. Tsujino: Real-time tracking of multiple sound sources by integration of in-room and robot-embedded microphone arrays, Proc. IEEE/RSJ Int. Conf. Intell. Robots and Systems, pp. 852–859, Oct. 2006.
- [15] B. Yang: Projection approximation subspace tracking, IEEE Trans. Signal Process., Vol. 43, No. 1, pp. 95–107, Jan. 1995.
- [16] 奥乃博, 中臺一博, 高橋徹, 武田龍, 中村圭佑, 水本志志, 吉田尚水, 大塚琢馬, 柳楽浩平, 糸原達彦: HARK 1.2.0 ドキュメント, <http://winnie.kuis.kyoto-u.ac.jp/HARK/hark-document-ja-ver1.2.pdf>, Mar. 2013.
- [17] T. Pham and M. Fong: Real-time implementation of MUSIC for wideband acoustic detection and tracking, Proc. SPIE, Automatic Target Recognition VII, Vol. 3069, pp. 250–256, Jun. 1997.
- [18] 押田賢浩, 大和田昇, 陶山健仁: 遅延和アレー法を用いた複数移動音源追尾の性能改善, 第 9 回情報科学技術フォーラム, Vol. 2, pp. 25–30, Sept. 2010.
- [19] A. P. Dempster, N. M. Laird and D. B. Rubin: Maximum likelihood from incomplete data via the EM algorithm, J. Royal Stat. Soc. B, Methodological, Vol. 39, No. 1, pp. 1–38, Dec. 1977.
- [20] C. R. Sastry, E. W. Kamen and M. Simaan: An efficient algorithm for tracking the angles of arrival of moving target, IEEE Trans. Signal Process., Vol. 39, No. 1,

- pp. 242–246, Mar. 1991.
- [21] D. B. Ward and R. C. Williamson: Particle filter beam-forming for acoustic source localization in a reverberant environment, Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., Vol. 2, pp. II-1777–II-1780, May 2002.

松田 尚頼 2012 年 3 月東京電機大・工・電気電子卒。2014 年 3 月，同大大学院修士課程修了。在学中，マイクロホンアレーの研究に従事。電子情報通信学会会員。

陶山 健仁 1998 年 3 月電気通信大学大学院博士課程修了。1998 年 4 月同助手。1999 年 4 月東京理科大学工学部助手。2002 年 9 月東京電機大学工学部第一部電気工学科講師。現在，同大学電気電子工学科教授。博士（工学）。マイクロホンアレー，音源定位，デジタルフィルタ設計に関する研究に従事。電気学会，電子情報通信学会，信号処理学会会員。

(20014 年 3 月 11 日受付；2014 年 9 月 7 日再受付)