

5M-9

パーティクルフィルタによる音源追跡の性能評価

村瀬 昌満[†]中臺 一博[‡]奥乃 博[†][†]京都大学大学院 情報学研究科 知能情報学専攻 [‡](株)ホンダ・リサーチ・インスティテュート・ジャパン

1. はじめに

ロボットが人間と社会の中で共生するためには、複数話者の同時発話、移動話者、発話中の割り込みなどの様々な形態の音声インタラクションに対応できる必要がある。

一般に、実環境では話者やロボットが同じ方向を向いて立ち止まっているわけではなく、ロボットが動くとき、相対的に話者が移動することもあるので、移動話者への対応は不可欠である。さらに、多くの音源分離手法では音源定位情報をパラメータとしているので、移動話者に対する音源定位は重要な機能である。しかし、静止音源については、音源定位、音源分離に関する多数の研究があるが、移動音源に対する音源定位、音源分離の研究は始まったばかりである。

以前、我々は履歴長の異なる複数のカルマンフィルタを用いた移動話者の実時間追跡システムを開発し [1]、これが複数の移動話者の追跡に有効であることを示した。本稿では、音源定位システムが出力する空間スペクトル値を利用し、パーティクルフィルタを用いた移動話者の追跡を行った。また、線形なフィルタリングであるカルマンフィルタを用いた場合との比較を行った。

2. システム構成

本実験では、音源定位システムによる定位結果を、履歴長が異なる複数のカルマンフィルタ、およびパーティクルフィルタを用いて追跡を行った。

2.1 音源定位システム

音源定位システムとして、MUSIC法 [2, 3] を実装し、用いた。MUSIC法では、出力として各方向の空間スペクトル値が得られる (図 1)。本実験では、この値をそれぞれの方向に音源が存在する尤度であると見なした。

2.2 履歴長が異なる複数のカルマンフィルタ

各フレームごとに、履歴長の異なる 3 つのカルマンフィルタを用いて話者位置の推定を行い、この推定結果のうち、前フレームにおける誤差が最小のものを推定位置として用いる [1]。本実験では、履歴長として 120ms, 240ms, 400ms を用いた。

2.3 パーティクルフィルタ

2.3.1 特徴

カルマンフィルタなどのフィルタリングと異なり、パーティクルフィルタの特徴として、「内部状態の遷移モデル、観測モデルを確率モデルで記述 (非線形な遷移、ガウスノイズ以外のノイズが可能)」というものがある [4]。そのため、カルマンフィルタにおける、内部状態が線形遷移するという制約をとり除くことが可能となる。

パーティクルフィルタによる内部状態の推定は以下の流れに従って行われる。

1. 状態空間の値を持つ N 個の粒子を初期化する。初期状態の確率が分かっているときは、その確率に従って初期化を行う。そうでない場合はランダムに初期化を行う。

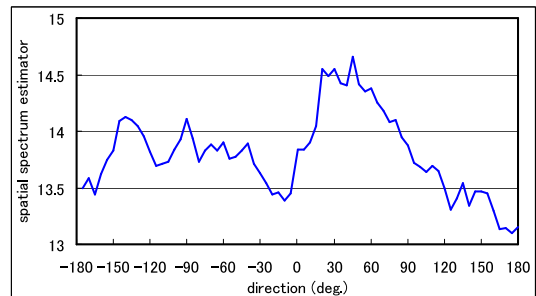


図 1: MUSIC 法による空間スペクトル値 (30度方向に音源が存在)

2. 内部状態の遷移モデル $p(x_t|x_{t-1})$ に従って、各粒子の状態を遷移させる。
3. 現在の観測値 y_t と各粒子 x_t に対し、観測モデル $p(y_t|x_t)$ から算出される尤度を各粒子の重要度として付与し、和が 1 となるように正規化する。
4. 粒子の重要度に応じて粒子を増殖、消去する。
5. 2. へ戻る。

ここで、各時点での内部状態の確率分布は、step 4. 直後における粒子の分布として得られる。得られた内部状態の分布から、平均の近傍の粒子のみの平均を内部状態の推定値として出力する。また、本実験では 500 個の粒子を用いた。

2.3.2 モデル化

我々は、内部状態、観測状態、状態遷移モデル、観測モデルを以下のように定めた。

内部状態 θ_p

θ_p はマイクロフォンアレイから見たときの音源方向

観測状態

音源定位システムの出力である、各方向の空間スペクトル値

状態遷移モデル

random walk モデルを用いた。

観測モデル

内部状態が θ_p である粒子に対する尤度として、音源定位システムの方向 θ_p の空間スペクトル値を用いた。

3. 移動話者の追跡実験

移動話者の定位結果を履歴長の異なる複数のカルマンフィルタ、パーティクルフィルタでフィルタリングを行い、その結果を比較した。

3.1 実験・解析条件

8ch のマイクロフォンアレイをロボット頭部に円形にほぼ等間隔になるように取り付け、ターンテーブルに設置した。録音は、1面がガラスの壁、残りの3面が吸音材に覆われた、7m×4mの部屋で行った [5]。ターンテーブルをこの部屋のほぼ中央(3m, 2m)に配置し、正面 1.5m で発話を行い、録音した。録音時に、マイクロフォンア

Sound Source Tracking based on Particle Filter and Accuracy Evaluation Masamitsu Murase (Kyoto Univ.), Kazuhiro Nakadai (HRI Japan), and Hiroshi G. Okuno (Kyoto Univ.)

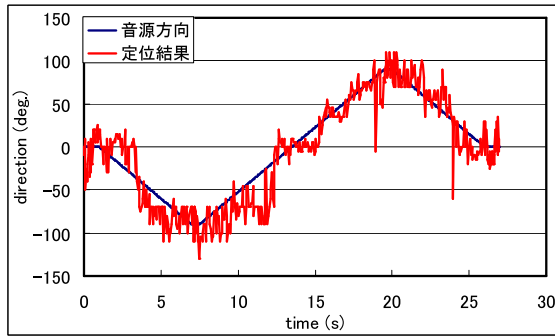


図 2: 音源定位結果

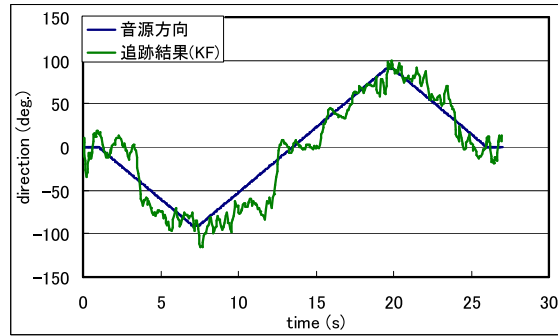


図 3: カルマンフィルタ (KF) による追跡結果

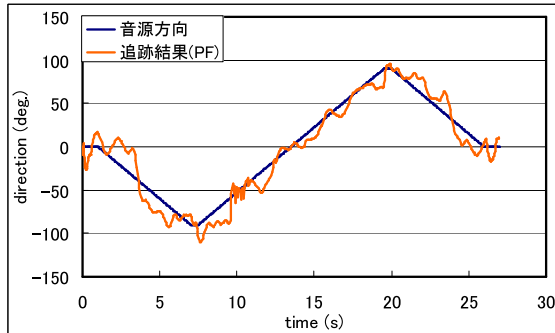


図 4: パーティクルフィルタ (PF) による追跡結果

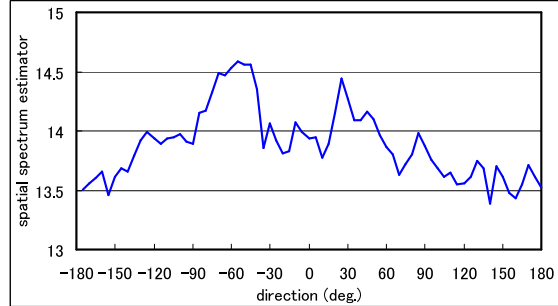


図 5: MUSIC 法による空間スペクトル値
(30 度方向に音源が存在)

レイを載せたターンテーブルを、時計回りに 90 度、反時計回りに 180 度、時計回りに 90 度回転させた。そのため、マイクロフォンアレイからは、話者が反時計回りに 90 度、時計回りに 180 度、反時計回りに 90 度移動したように観測される。また、あらかじめ 5 度おきにインパルス応答を測定しておき、音源定位システムで用いた。録音はすべてサンプリング周波数 48kHz で行った。

3.2 結果

図 2 に話者の方向と定位結果を、図 3 に話者の方向とカルマンフィルタ (KF) による追跡結果を、図 4 に話者の方向とパーティクルフィルタ (PF) による追跡結果を示す。

また、表 1 にそれぞれの追跡結果の平均誤差を示す。

表 1: 追跡結果の平均誤差	
	平均誤差 (度)
MUSIC 法による定位結果	20.84
KF による追跡結果	17.38
PF による追跡結果	13.34

3.3 考察

3 つのカルマンフィルタの各履歴長をこの問題で最適となるように調整したが、表 1 から分かるように、そのようなパラメータの調整が不要であるパーティクルフィルタによる追跡では、より精度の高い結果を得ることができた。これより、パーティクルフィルタによる追跡手法は、より頑健であるといえる。

図 2 において、24 秒付近で定位誤差が大きくなっている。このときの各方向の空間スペクトル値 (各方向に音源が存在する尤度) は図 5 のようになっている。最尤な方向は -60 度となる。しかし、正しい方向である 30 度付近の尤度も大きくなっている。粒子が適切にサンプリングされ、正しく追跡が行われたと考えられる。音

源定位では、このように大きく定位結果が分散することが起こりやすい。そのため、音源定位システムによる定位結果のみを用いるのではなく、空間スペクトル値を各方向の音源存在尤度として用いるという本手法が有効であると考えられる。

また、パーティクルフィルタでは観測モデルに任意のモデルを用いることができるので、画像情報を用いることが可能となるような状況では、画像による定位結果と音声による定位結果を観測モデルにおいて統合することが可能となると考えられる。

4. おわりに

本稿では、カルマンフィルタおよびパーティクルフィルタによる話者の追跡精度の比較を行った。その結果、調整すべきパラメータが少ないパーティクルフィルタで、より高い追跡結果が得られ、パーティクルフィルタを用いた本手法の頑健性が示された。また、MUSIC 法の出力である空間スペクトル値を音源存在の尤度として用いることにより、追跡精度が向上することを確認した。

謝辞 本研究の一部は、科研費、21 世紀 COE、SCAT 研究助成の支援を受けた。

参考文献

- [1] 村瀬昌満, 他: “カルマンフィルタによる音声の時系列特徴を用いた複数移動話者の追跡,” 第 67 回情処全大, vol.2, 385–386, 2005.
- [2] Ralph O. Schmidt: “Multiple Emitter Location and Signal Parameter Estimation,” *IEEE Trans. Antennas Propagation*, vol.34, 276–280, 1986.
- [3] Futoshi Asano, et al.: “Fusion of Audio and Video Information for Detecting Speech Events,” *Proc. of Fusion 2003*, 386–393, 2003.
- [4] Rudolph van der Merwe, et al.: “The Unscented Particle Filter,” *Advances in Neural Information Processing System 13*, 2001.
- [5] Kazuhiro Nakadai, et al.: “Sound Source Tracking with Directivity Pattern Estimation Using a 64ch Microphone Array,” *IEEE/RSJ IROS 2005*, 196–202, 2005.