4U-4

パーティクルフィルタを用いた動的環境下の複数音源追跡

黄 楊暘 † 大塚 琢馬 ‡ 高橋 徹 ‡ 尾形 哲也 ‡ 奥乃 博 ‡

† 京都大学 工学部情報学科 ‡ 京都大学 大学院情報学研究科 知能情報学専攻

1. はじめに

人とロボットが共生するためには,「聞き上手」なロボットの開発が不可欠である. 聞き上手ロボットを実現する上で重要な機能の一つに, いつどこで, 誰が話したのかを明らかにする speaker diarization がある. 本稿では, 移動話者も含めた speaker diarization のための複数音源追跡について報告する. 様々な環境下で音を聞くロボットは, 次のような動的環境に頑健に対応する必要がある. (1) 複数の音源が同時に存在する. (2) 音源が移動する. (3) 音源が一時的消失して再び現れる.

音源追跡手法としては, 多チャンネル音響信号の共分散モデルに基づいて, パーティクルフィルタによって環境中を動く複数音源の同時追跡が達成されている [1, 2]. しかしながら, (2), (3) という状況の下では, 音源の方向にのみ着目した音源追跡では, ある音源を別の音源と混同しうるため, 異なる音源を追跡するという問題があった. 図 1 に異なる音源を識別できないことで生じる不具合の例を示す. 図のように音源同士が接近する場合, 音源位置を連続的に繋ぐだけの追跡処理では, 個々の音源識別の曖昧性が生じる. したがって, 時々刻々の音源位置を繋ぐだけではなく, 音響的特徴も取り入れることが精度のよい音源追跡には重要である. 本稿では, パーティクルフィルタに共分散尤度と音響特徴量尤度を取り入れて, 音源追跡と識別問題を扱う.

入力: 多チャンネルの音声信号
出力: 複数移動音源の水平面上の方向の時系列情報
仮定: 追跡する音源の音響特徴を事前に入手可能

2. 提案手法

本手法は 2 段階からなる. (1) 伝達関数の測定, 音源識別用の事前学習処理, (2) オンライン音源定位, 識別処理

事前処理:伝達関数測定と音源識別パラメータ学習　事前処理は図 2 のように以下である. (1) 各方向から多チャンネル音声信号を収集するマイクロフォンアレイまでの伝達関数を測定する. (2) 追跡する音源のクリーン信号から、音響特徴量の混合ガウスモデル (GMM) のパラメータを学習する. 音響特徴量としては次のように計算する Mel Scale Log Spectrum (MSLS) を利用する. (1) 観測信号の振幅値にメル周波数窓関数を適用. (2) 得られた 13 次元のベクトルをノルム 1 に正規化する. (3) 13 次元の時間変化 Δ と, 観測信号のパワー計算し, 27 次元ベクトルを生成する. 抽出した MSLS 特徴量について, EM アルゴリズムを用いて各音源の GMM パラメータを学習する. 学習す

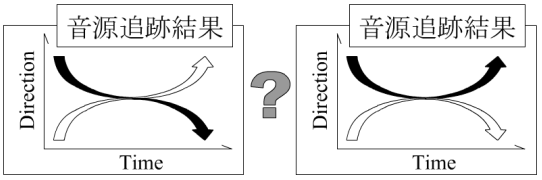

図 1: 音源の接近によって生じる音源追跡の結果の曖昧性

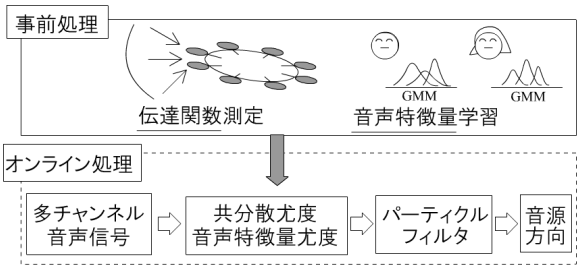

図 2: 手法の概要

るパラメータ $g^k_{c_d}$ は混合重み, $\boldsymbol{\mu}^k_{c_d}, \boldsymbol{\Sigma}^k_{c_d}, (k = 1, \cdots, K)$, はそれぞれのガウス分布の平均と共分散行列を表す. c_d, k はそれぞれ音源 ID, GMM の各混合のインデックスを表す.

オンライン処理: 音源の識別・追跡　オンライン処理ではパーティクルフィルタを利用して, 音源追跡と識別を同時に行う. パーティクルの重み評価に共分散尤度と, 学習されたパラメータを援用する.

2.1 パーティクルフィルタ

音源方向の推定は, 短時間フーリエ変換を施した多チャンネル信号を 0.25 秒間隔で区切ったブロック単位で行う. ブロック内での音源の存在方向が一定と仮定する. $\mathbf{z}_b$ を時間周波数領域での多チャンネル信号とする. 簡単のため, 周波数ビン番号とブロック内のフレーム番号を省略する. $b = 1, \cdots, B$ はブロックの添字である.

本問題を, 各音源の方向と, 学習された音響特徴との対応付けである識別 ID を隠れ変数とする状態空間モデルとして定式化する. パーティクルフィルタは状態空間の確率分布を多数のパーティクルと呼ばれるサンプル値とその重みによって近似し, ブロックごとにパーティクルの

表 1: パーティクルフィルタのアルゴリズム

初期化:	$\{\mathbf{X}_0^{(i)}, w_0^{(i)}\}, i = 1, \cdots, N_p$
反復: **For** $b = 1, 2, \cdots$	
サンプリング:	$\mathbf{X}_b^{(i)} = \mathbf{f}(\mathbf{X}_{b-1}^{(i)})$
重みの更新:	$w_b^{(i)} = p(\mathbf{z}_b \mid \mathbf{X}_b^{(i)})$
	$\tilde{w}_b^{(i)} = w_b^{(i)} / \sum_{j=1}^{N_p} w_b^{(j)}$
リサンプリング:	$\{\mathbf{X}_b^{(i)}, \tilde{w}_b^{(i)}\} \rightarrow \{\tilde{\mathbf{X}}_b^{(j)}, N_p^{-1}\}$
事後確率密度推定:	$\hat{p}(\mathbf{X}_b \mid \mathbf{Z}_{1,\cdots,b}) = \frac{1}{N_p} \sum_{j=1}^{N_p} \delta(\mathbf{X}_b - \tilde{\mathbf{X}}_b^{(j)})$
状態推定:	$\hat{\mathbf{X}}_b = \frac{1}{N_p} \sum_{j=1}^{N_p} \tilde{\mathbf{X}}_b^{(j)}$
置き換え:	$\mathbf{X}_b^{(i)} = \tilde{\mathbf{X}}_b^{(i)}$
End	

Sound Sources Tracking using Particle Filter in Dynamic Environments: *Yangyang Huang, Takuma Otsuka, Toru Takahashi, Tetsuya Ogata, and Hiroshi G. Okuno (Kyoto Univ.)*

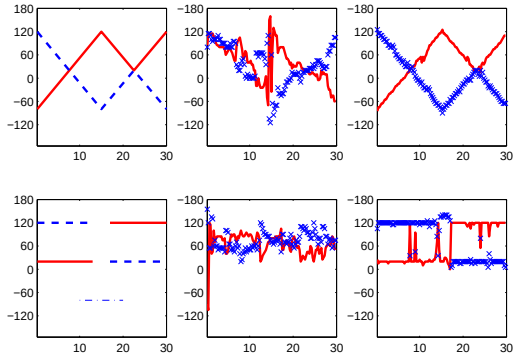

図 3: 正解音源軌跡と推定結果. 横軸は時間 (sec), 縦軸は方向 (deg) を表す. 左から順に評価用軌跡 (正解軌跡). 共分散尤度による追跡結果, 右が共分散尤度＋音響特徴量尤度による追跡結果. 上段: 交差軌跡, 下段: 消失軌跡. 実線, 破線, 点線は異なる音源を表す.

持つ値を更新する. アルゴリズムを表 1 にまとめる. N_p はパーティクルの数である. $\mathbf{X}_b^{(i)}$ はブロック b での i 番目のパーティクルの状態を表す. $w_b^{(i)}$ はそのパーティクルの重みを表す.

$$\mathbf{X}_b^{(i)} = (s_1, \cdots, s_D, c_1, \cdots, c_D) \tag{1}$$

$s_1, \cdots, s_D$は追跡する D 個の音源方向を表す. $c_1, \cdots, c_D$は追跡する D 個の音源の ID を表す. $\mathbf{f}(\mathbf{X}_{b-1}^{(i)})$ は音源の遷移に対応する提案分布であり, 等確率で前時刻の方向からその場に留まる, あるいは隣接方向に動くモデルを利用する. つまり, $s_d^{(b+1)}$は $[s_d^{(b)}+1, s_d^{(b)}, s_d^{(b)}-1]$ から等しい確率で選択する. 提案分布を表す, 簡単のため, s_{d+1} は $[s_d+1, s_d, s_d-1]$ の中からランダムに出す. また, 音源 ID は 20 パーセントの確率で入れ替える. 同一音源 ID の方向は状態変数 s_d の重み付き平均で推定する. パーティクルの重み計算に用いる尤度関数は次の通り.

$$p(\mathbf{z}_b|\mathbf{X}_b^{(i)}) = L_m(\mathbf{z}_b|\mathbf{X}_b^{(i)}) L_f(\mathbf{z}_b|\mathbf{X}_b^{(i)})$$

共分散尤度:$L_m(\mathbf{z}_b|\mathbf{X}_b^{(i)})$　共分散モデルによる音源追跡は関連文献 [1] に詳しい. ここで, 数式をまとめる. 一つのブロックでの観測値を $\mathbf{z}$ とする. まずは, 共分散尤度の計算式はこうなる. 各周波数ビンに対してさらに尤度を統合する.

$$L_m(\mathbf{z}_b|\mathbf{X}_b^{(i)}) = |\mathbf{K}_\mathbf{z}|^{-1} \exp(-\mathrm{tr}(\mathbf{K}_z^{-1}\mathbf{R}_z)) \tag{2}$$

$$\mathbf{R}_z = \frac{1}{25}\sum_{i=1}^{25} \mathbf{z}(i)\mathbf{z}^H(i) \tag{3}$$

$$\mathbf{K}_z = [\mathbf{H}(s_d)]\mathbf{K}_s[\mathbf{H}(s_d)]^H + \sigma^2\mathbf{I} \tag{4}$$

ただし,$\mathbf{R}_z$ は観測信号のサンプル共分散, K_z がモデルから導出される共分散である. K_s は各音源のパワーを表す. 本来, 方向との同時推定が必要であるが, 式 (6) で近似する.

$$\mathbf{K}_s = \mathrm{diag}(\lambda_1, \cdots, \lambda_D) \tag{5}$$

$$\lambda_d = \frac{\mathbf{H}^H(s_d)\mathbf{R}_z\mathbf{H}(s_d)}{(\mathbf{H}^H(s_d)\mathbf{H}(s_d))^2}, (\text{for } d = 1, \cdots, D) \tag{6}$$

表 2: 追跡結果の平均誤差 (deg)

	共分散尤度のみ		共分散尤度＋音声特徴量尤度	
	交差軌跡	消失軌跡	交差軌跡	消失軌跡
合成信号	50.25	42.25	5.85	0.35
音声発話	75.95	64.95	6.50	13 .80

$\mathbf{H}(s_d)$ は s_d 方向の伝達関数の列ベクトルを表す. $[\mathbf{H}(s_d)] \in \mathbb{C}^{(\text{マイク数}\times\text{音源数})}$ は $\mathbf{H}(s_d)$ からなる行列である. $[\mathbf{H}]^H$ は行列 $[\mathbf{H}]$ のエルミート転置を表す.

音響特徴量尤度:$L_f(\mathbf{z}_b|\mathbf{X}_b^{(i)})$　周波数領域のビームフォーミングによる観測 $\mathbf{z}_b$ の s_d 方向の分離音声は $\mathbf{H}(s_d)^{-1}\mathbf{z}_b$ で計算される. $\mathbf{M}(\mathbf{H}(s_d)^{-1}\mathbf{z}_b)$ は分離信号の MSLS 特徴量を表す. 観測値の音響特徴量尤度は次の式で計算される.

$$L_f(\mathbf{z}_b|\mathbf{X}_b^{(i)}) = \prod_{d=1}^{D}\{\sum_{k=1}^{K} g_{c_d}^k N(\mathbf{M}(\mathbf{H}(s_d)^{-1}\mathbf{z}_b)|\boldsymbol{\mu}_{c_d}^k, \boldsymbol{\Sigma}_{c_d}^k)\}$$

3. 評価実験

評価実験では, 8 チャンネルの 45° おきに配置された円形マイクロフォンアレイを用いる. 定位する方向は, マイクロホンアレイの周囲を 5° おきの解像度で 72 方向である. テストデータはサンプリングレートが 16000Hz であり, 窓長 512point, シフト長 160point の STFT で変換して、分析する. テストデータは調波構造を持つ合成信号と, 男性及び女性発話を用いた. 合成信号は, 基本周波数 440(Hz) と 349(Hz) で, 30 倍音まで加算されている. 10 秒の信号を学習に用い, 30 秒を評価に用いた. 音声信号は, 男女の JNAS 音素バランス文のうち, 40 文を学習, 10 文を評価に用いた. 評価用データは伝達関数を用いた合成データである. 図 3, 表 2 が示すように, 従来の共分散尤度のみを用いる定位よりも, 音響特徴を組み込んだ本手法が精度の高い音源定位を実現している. この理由は, 共分散尤度のみを用いる手法では, 音源 ID が特に音源同士が接近した場合に入れ替わりやすいため, 各パーティクルの持つ方向の値を平均した結果が安定しないためである. また, 図 3 右に示すように, 音源が一時的に消失した後でも, 音源を区別しながら追跡する要すが分かる. これらの結果から, 本手法は音源同士の接近にや一時的消失に頑健な音源追跡に優れていると言える.

4. まとめ

本稿では複数音源追跡に対する音響特徴尤度を加味したパーティクルフィルタによる解法を示した. 評価実験では, 音源同士の接近と一時的消失状況では有用であることが示された. 今後の課題は (1) 残響時間が長い環境での頑健な音源追跡, (2) 音源識別パラメータのオンライン学習化などが挙げられる.

謝辞 本研究は科研費 (S), GCOE の援助を受けた.

参考文献

[1] 浅野他: パーティクルフィルタを用いた移動音源の追跡技術, 日本音響学会誌 61 巻 12 号, pp. 720-727, 2005

[2] 村瀬他:パーティクルフィルタによる音源追跡の性能評価, 情報処理学会第 68 回全国大会, 5M-9, Mar. 2006.

[3] 浅野:「音のアレイ信号処理」, コロナ社, 2011