

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 25 日現在

機関番号：32657

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23560468

研究課題名(和文)リアルタイム複数話者追尾システムの実現

研究課題名(英文)Realization of realtime multiple talker tracking system

研究代表者

陶山 健仁(Suyama, Kenji)

東京電機大学・工学部・教授

研究者番号：50303011

交付決定額(研究期間全体)：(直接経費) 2,700,000円、(間接経費) 810,000円

研究成果の概要(和文)：本研究では、リアルタイム複数話者追尾システムの実現を目的として、検討中の追尾アルゴリズムの検証ならびに新しい枠組みでの追尾アルゴリズムの提案を行なった。まず、検討アルゴリズムであるPAST-IPLSでは、マイクロホン間隔を拡張した際の空間エイリアシングの影響の回避方法について検討し、実環境実験によりその有効性を示した。つぎに新しい枠組みのアルゴリズムとして群知能であるPSOによる話者追尾手法について提案した。また、話者追尾問題に共通の課題である同一話者追尾問題に対する解決法も示した。その結果、5°程度の推定誤差での追尾をリアルタイムファクタ0.5以下で実現した。

研究成果の概要(英文)：In this research, an effectiveness of the realtime multiple talker tracking method developed by the representative author under the support of this grant money has been studied. Through several experimental results carried out in the actual environment, it was verified that a mis-estimation by spatial aliasing could be avoided by the proposed system even when wide width microphones were used for estimation. Moreover, a novel algorithm has been proposed under a new framework. The algorithm used the PSO (Particle Swarm Optimization) which is one of the swarm intelligence algorithms. In addition, the same talker tracking problem was resolved even if the new algorithm was used for the tracking. A realtime tracking performance was shown by the several experimental results.

研究分野：工学

科研費の分科・細目：電気電子工学

キーワード：マイクロホンアレー 音源追尾 PSO 群知能

1. 研究開始当初の背景

本研究はマイクロホンアレーによる音源追尾システムの開発を目標としている。マイクロホンアレーによる話者追尾のアプリケーションでは、精度の高い推定をリアルタイムで行なうことが求められる。高解像度音源方向推定法である Multiple Signal Classification (MUSIC)はその候補の1つである。MUSICは信号部分空間と雑音部分空間の直交性により高解像度推定が可能である。しかしながら、信号部分空間算出のための固有値分解と MUSIC スペクトルのピークサーチに多大な演算負荷を要する。研究代表者は、部分空間算出の高演算問題に対して、Projection Approximation Subspace Tracking (PAST) の一回更新により信号部分空間を逐次推定し、信号部分空間算出にかかる演算負荷を削減している(電気学会論文誌 C, Vol.129, No.1, pp.79-86, 2009)。

ピークサーチにかかる演算負荷削減に対しては、音声信号の非定常性により MUSIC スペクトルのピーク位置が急激に変化する恐れがあるため、高速な追従性と無音区間での追尾失敗を防ぐ安定性をもつ追尾法が求められる。そこで、適応信号処理の観点から Interior Point Least Square (IPLS) で MUSIC スペクトルの解を求め、MUSIC の高解像度を保ちつつリアルタイム話者追尾を実現する PAST-IPLS 法を提案し、検討してきた(電子情報通信学会論文誌 A, Vol.J93-A, No.4, pp.248-259, 2010)。しかしながら、複数音源追尾問題における共通の課題である無音区間から発話区間への復帰時における同一話者追尾問題の解決のためには高精度な音源定位法を必要とするため、高い演算負荷を要する問題が残った。そこで、本研究では、新しい枠組みでの音源追尾法について検討する。

2. 研究の目的

近年、制御や最適化の分野において、群による情報の共有と集中化により解を探索する Particle Swarm Optimization(PSO)[5]による動的システムのモデル同定が盛んに研究されている。音源追尾問題は確率的に状態が変動する動的モデルと見なしうるため、PSO に合致した問題であるといえる。PSO の高速な解探索は MUSIC スペクトルのピーク位置の変動にすばやく対応できることが見込め、パラメータ調整により実時間処理も容易に実現可能と考えられる。そこで、本研究では話者追尾問題を動的システムのモデル同定とみなし、PSO を用いた話者追尾法を提案する。PSO の集中化は高速な探索に貢献するが、探索の多様性を失うため、評価関数の変化に対応できなくなることが指摘されている。特に、話者追尾問題では、無音区間において話者方向にピークが存在し

ない可能性があり、その際に PSO が探索の多様性を失うため、追尾の安定性が損なわれる。

この問題に対し、提案法ではスペクトルの値に応じた群最良解更新と個体再配置を行う。これにより追尾失敗時に個体が集中化することを避ける。さらに、発話者が複数存在する際には、話者数と同数の群を用いることで複数の話者を追跡する。この際、複数の群が同一のピークを探索することを防ぐため、群毎に探索禁止範囲を設定する。実環境実験により、提案法が従来法よりも高精度な追尾を同じ演算時間で実現できることを目的とする。

3. 研究の方法

提案法の流れを図1に示す。まず、Mチャンネルのマイクロホンの受信音を Discrete Fourier Transform(DFT)により周波数領域に変換する。周波数領域の各帯域において相関行列を算出する。

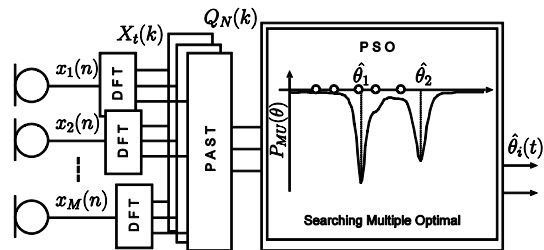


図1: 提案法の流れ

相関行列に対し、PAST を用いて信号部分空間を張る固有ベクトルを1回更新で算出する。PASTはRLSアルゴリズムに基づいたアルゴリズムであるため、高い追従性能を有する。また、本来3乗オーダーを要する演算を2乗オーダーで実現可能である。信号部分空間を張る固有ベクトルから MUSIC スペクトル算出に必要な雑音部分空間の計量行列を算出する。

MUSIC スペクトルが算出できれば、そのピークを与える方向から音源方向を推定可能である。そのために、音源追尾に利用する周波数帯域の MUSIC スペクトルを平均化する。平均化 MUSIC スペクトルは多峰性関数となるが、音源方向に対する方向では鋭いピークを有する。そのピーク検出に PSO を用いる。

PSO は生物の群行動をモデル化した多点探索アルゴリズムである。PSO は複数の個体と全個体からなる群により構成される。各個体は位置と速度をもち、良解に関する情報を共有しながら大域的最適解を探索する。PSO は評価関数の勾配情報を用いないため、様々な問題に対しフレキシブルに適用可能な手法である。また、良解に関する情報の共有により局所解に対して探索の強い指向性を有するため、MUSIC スペクトルに対して適用した場合はすばやいピーク検出が可能である。さらに、その更新則が四則演算のみで記述され

るため、演算負荷を抑えることが可能である。

しかしながら、複数音源追尾の場合は、推定すべき解も複数存在するため、通常の PSO で推定可能な音源は、鋭いスペクトル値を与える 1 つの音源方向に限定される。提案法ではこれを防ぐために複数群を用意し、各群が同一音源を追尾することなく別々の音源を追尾するようなメカニズムの構築が必要である。このような状況は、1 人の話者が無音状態になった場合、他の話者方向のピークをすばやくとらえるため、無音状態にあった話者が再び発話状態になった場合にも、その話者に対応するピークを捕らえられないという問題が生じる。従来法では、この状態を防ぐために高負荷な演算を要する音源定位法を適用していたが、本研究では、PSO の枠組みの範囲内で実行可能な手法について 2 通り検討した。

1) 探索禁止範囲の設定による手法

第 1 の手法として図 2 に示すような元々の評価関数に対し、探索禁止範囲を設定し、同一音源追尾を回避する手法について検討した。この手法は、図 3 に示すように最初の群が検出した音源方向を含む領域をつぎの群が探索する領域から外し、探索を禁止することを意味する。この手続きは、その領域の評価関数値を 0 に設定すれば、通常の PSO の枠組みを崩すことなく実行可能である。

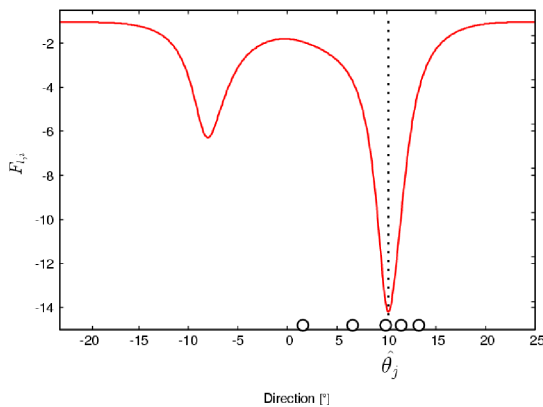


図 2 探索禁止範囲設定前

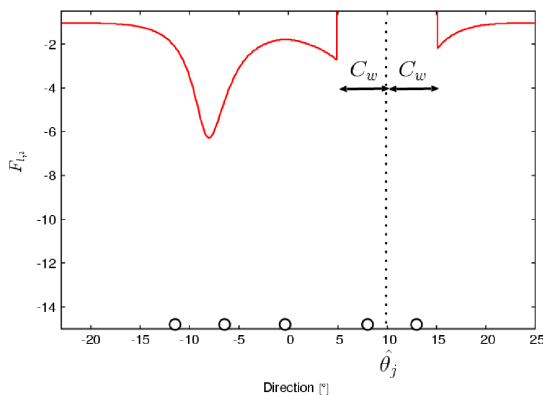


図 3 探索禁止範囲設定後

2) 部分空間減算による手法

第 2 の手法では、PAST で求めた信号部分空間に対し、最初の群で推定した音源方向に対応するステアリングベクトルを減算し、つぎの群が探索する対象から外す手法について検討した。この手法も PSO の評価関数を変更するのみであるため、PSO の枠組みを崩すことなく低演算という PSO の利点を活かせる手法である。

4. 研究成果

提案法の追尾精度、演算時間を検証するために実環境で録音した音声を用いて実験を行った。音声は標準的な会議室で収録した。収録時の騒音レベルは 34.0 [dB]、残響時間は 0.51 [s] であった。人がスピーカを持ち、スピーカをマイクロホンアレーの方向に向けながら歩いて移動したものを音源とした。音源の移動パターンは 3 種類、音声は 4 種類、いずれも無音区間を含み、計 12 パターンに対して実験を行なった。演算時間は Intel CoreTMi3 - 2130 3.4[GHz]、メモリ 4[GByte] を搭載した PC を使用し計測した。試行毎の追尾誤差は、平均絶対値誤差で評価した。評価のための真値としてカメラアレーによる画像計測により計測した位置を用いた。表 1 に実験条件を示す。演算量については 20[s] の音声に対する平均演算時間で評価した。

表 1 実験条件

マイクロホン数 M	16
マイクロホン間隔 d	40[mm]
サンプリング周波数 f_s	8[kHz]
使用周波数帯域	2[kHz] - 4[kHz]
フレーム長 K	256
オーバーラップ	50[%]
音源数 N_s	2

図 4 に提案法 1 による追尾結果、図 5 に提案法 2 による追尾結果を示す。

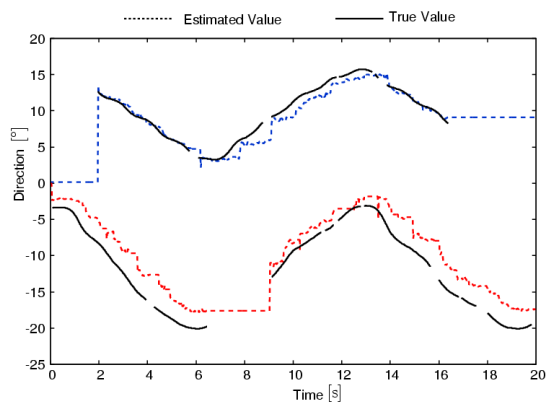


図 4 提案法 1 による追尾結果

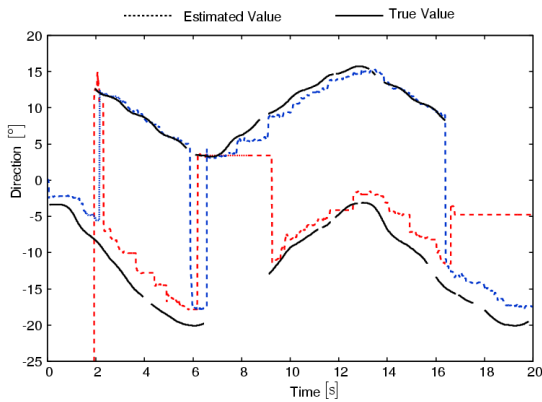


図4 提案法2による追尾結果

図4, 図5の結果よりいずれの提案法でも各音源を区別・追尾していることが確認できる。特に無音状態から発話状態に復帰した場合、もしくは追尾開始時にすばやく音源の移動に追従していることがわかる。追従性能については図6, 図7に示す個体配置からも確認できる。表2に各使用音声, 各移動パターンに対する追尾結果の平均誤差, 分散, 平均演算時間を示す。この結果からも, 提案法が最大7°程度の追尾誤差をリアルタイムファクタ0.5未満で実現していることが確認できる。以上の結果から, 本研究の目的である実時間複数音源追尾が実現できたことが確認できた。

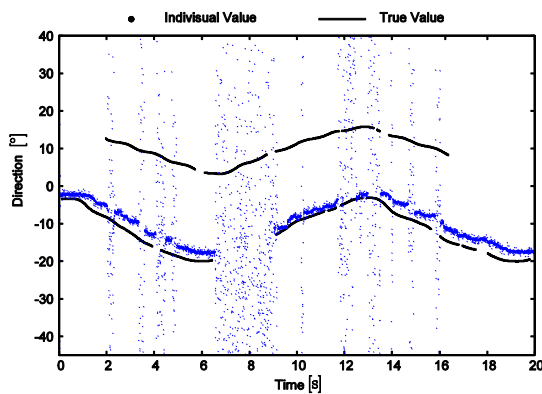


図5 提案法1における個体配置

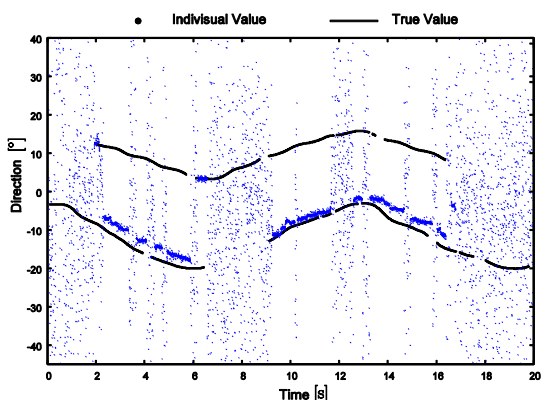


図6 提案法2における個体配置

表2 実験結果

使用音声	移動パターン	提案法1			提案法2		
		\bar{e} [°]	σ^2	\bar{T} [s]	\bar{e} [°]	σ^2	\bar{T} [s]
1	1	2.13	0.021	6.91	2.09	0.001	7.21
	2	2.27	0.008	6.85	2.47	0.046	7.20
	3	4.50	0.063	6.85	5.42	1.796	7.20
2	1	1.95	0.026	6.93	1.92	0.004	7.28
	2	2.45	0.022	6.89	2.60	0.036	7.25
	3	4.17	0.003	6.86	5.87	1.629	7.20
3	1	1.48	0.002	6.90	1.51	0.007	7.22
	2	3.85	0.246	6.86	2.77	0.251	7.20
	3	5.76	0.014	6.86	6.66	16.335	7.20
4	1	2.28	0.015	6.85	2.30	0.007	7.20
	2	2.91	0.063	6.86	4.97	0.113	7.20
	3	4.14	0.018	6.89	4.26	0.033	7.20

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

(雑誌論文)(計 2 件)

1. 梅津浩史, 陶山健仁: 局在音声エネルギーを追従する複数音源定位法, 電子情報通信学会論文誌, 査読有, Vol. J95-A, No. 10, pp.756-759, 2012
2. 前田隼一郎, 梅津浩史, 陶山健仁: マイクロホン間隔の拡張による複数音源定位の性能改善, 電子情報通信学会論文誌, 査読有, Vol. J95-A, No. 6, pp.523-526, 2012

(学会発表)(計 10 件)

1. 平川真里, 陶山健仁: MUSIC に基づくペナルティ付加による複数音源追尾, 電子情報通信学会技術報告, CAS2013-98, pp.47-52, 2014年3月6日, 大阪
2. 平川真里, 陶山健仁: PSO を用いた音声のスパース性に基づく複数音源追尾, 電子情報通信学会技術報告, CAS2013-76, pp.19-24, 2014年2月6日, 神奈川
3. 鈴木毅, 陶山健仁: 0-1 信頼性分布を用いた複数音源追尾の一提案, 電子情報通信学会技術報告, SIP2013-106, pp.123-128, 2014年1月23日, 福岡
4. Masato Hirakawa and Kenji Suyama: Multiple Sound Source Tracking by Two Microphones using PSO, Proceedings of IEEE ISPACS2013, 査読有, 2013年11月13日, 沖縄
5. 大森征一, 陶山健仁: PSO による複数話者追尾における同一方向推定回避, 第26回回路とシステムワークショップ予稿集, 査読有, pp.132-137, 2013年7月29日, 兵庫
6. 松田尚頼, 陶山健仁: 複数音源追尾におけるパーティクルの集中化回避, 電子情報通信学会技術報告, SIP2013-34,

- pp.19-24, 2013年7月11日, 熊本
7. 鈴木毅, 陶山健仁: 逐次更新ヒストグラムに基づく音源追尾の複数音源への拡張, 電子情報通信学会技術報告, SIP2012-141, pp.81-86, 2013年3月14日, 山形
 8. 大森征一, 前田線一朗, 松本拓也, 陶山健仁: PSOを用いた複数解探索による音源追尾法の提案, 第25回回路とシステムワークショップ予稿集, 査読有, pp.337-342, 2012年7月31日, 兵庫
 9. Hiroshi Umezu and Kenji Suyama: Multiple Sound Source Localization based on Local Existence Property of Speech Signal, Proceedings of 2011 WASET International Conference on Communications and Signal Processing, ICCSP2011, pp.676-679, 査読有, 2011年12月22日, タイ
 10. Kenji Suyama, Masahiro Oshida, and Noboru Owada: Multiple Moving Talker Tracking by Integration of Two Successive Algorithms, Proceedings of 2011 WASET International Conference on Communications and Signal Processing, ICCSP2011, pp.738-741, 査読有, 2011年12月22日, タイ

(2)研究分担者 ()

研究者番号:

(3)連携研究者 ()

研究者番号:

〔図書〕(計 0 件)

〔産業財産権〕
出願状況(計 0 件)

名称:
発明者:
権利者:
種類:
番号:
出願年月日:
国内外の別:

取得状況(計 0 件)

名称:
発明者:
権利者:
種類:
番号:
取得年月日:
国内外の別:

〔その他〕
ホームページ等

6. 研究組織

(1)研究代表者

陶山 健仁 (SUYAMA, Kenji)
東京電機大学・工学部・教授
研究者番号: 50303011