

平成 30 年 6 月 10 日現在

機関番号：32657

研究種目：基盤研究(C) (一般)

研究期間：2015～2017

課題番号：15K06084

研究課題名(和文) 実用性の高い話者追尾システムの開発

研究課題名(英文) A Study on Practical 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) 3,100,000円

研究成果の概要(和文)：本研究では、2マイクロホンシステムを用いた複数音源追尾について検討した。2マイクロホンシステムでは、低空間解像度のため推定精度が劣化する。そこで、高解像度スペクトル推定法として知られるMUSICを適用する。MUSICスペクトル関数は、音源方向に鋭いピークを有するが、推定精度向上を狙ってマイクロホン間隔を広げた場合、高周波数帯域で空間エイリアシングの影響で偽りのピークが発生することがある。その結果、評価関数は多峰性関数となる。本研究では、ピーク探索に確率的多点探索アルゴリズムであるPSOを適用した。さらに、ペナルティ関数の設定により、同一音源追尾問題の回避も達成した。

研究成果の概要(英文)：In this research, a method for multiple sound source tracking by using two microphones system was studied. In the two microphone system, an estimation accuracy of the DOAs (Direction-Of-Arrival) tends to degrade due to a low spatial resolution. Therefore, MUSIC (Multiple Signal Classification), which is known as a high resolution method for the spectrum estimation, is applied. Although the MUSIC spectrum function indicates sharp peaks to sound source directions, some false peaks are appeared in the function in the high frequency band when a width between two microphones is extended for improving the estimation accuracy. Thus, the evaluation function becomes to a multi-modal function. In this research, PSO (Particle Swarm Optimization) was applied for the peak search. In addition, the same source tracking problem was also resolved by setting a penalty function.

研究分野：信号処理

キーワード：音源追尾 マイクロホンアレー 確率的最適化

1. 研究開始当初の背景

音源追尾問題は、空間中を時々刻々と移動する話者の位置もしくは方向を推定する問題である。例えば、ハンズフリー通話やロボット聴覚のようなアプリケーションでは、話者は静止しているより、時刻によって位置が変動することが自然である。通常の会話環境では話者は一人に限定されず、複数存在することが容易に想定できる。そのような状況において、本来伝えたい話者音声のみを強調し、他の話者音声を抑圧するための選択的受音システムの構築を目指す場合、複数音源追尾が必要である。

音源追尾には、複数のマイクロホンと並べた收音系であるマイクロホンアレーが用いられる。代表的な音源追尾法であるパーティクルフィルタを用いる手法では、推定精度がマイクロホン数に依存するため、高精度推定には多数マイクロホンと要し、装置サイズ、コスト、演算量の点において問題が残る。さらに、一方の発話音声が無音状態の間では、パーティクルが発話状態にある他方の話者に集中し、一方の話者が発話状態に戻ってもパーティクルが復帰できない同一音源追尾問題が生じる。以上を考慮し、最小のマイクロホン数である2マイクロホンシステムを用い、かつ同一音源追尾問題を回避可能な手法が望ましい。しかし、2マイクロホンシステムでは、推定精度が劣化する傾向がある。本研究では、2マイクロホンシステムで、高い推定精度を有する音源追尾技術の開発について検討する。

2. 研究の目的

2マイクロホンシステムでは、空間解像度の低さが原因となり推定精度の劣化が懸念される。本研究では、高解像度スペクトル推定法として知られる MUSIC (Multiple Signal Classification) を適用する。MUSIC は受信信号の相関行列の固有ベクトルが張る信号部分空間の直交性に着目した手法であり、評価関数である MUSIC スペクトル関数が音源方向に対応する方向に対して、急峻なピーク特性を与えるため、2マイクロホンであっても高解像度推定が期待できる。ただし、MUSIC は狭帯域信号に対する手法であるため、音声のような広帯域信号に適用する場合は、受信信号を離散フーリエ変換 (DFT: Discrete Fourier Transform) した周波数毎に MUSIC スペクトル関数を求め、全帯域分を加算したスペクトル関数を評価関数として用いる。

MUSIC スペクトル関数は、理想的には音源方向に対応する方向のみにピークを有するが、スパース性を有する音声信号のように信号エネルギーを有しない周波数帯域のスペクトル関数や外来雑音による影響によって、帯域ごとにピーク方向がずれる場合があ

る。さらに、推定精度向上を狙ってマイクロホン間隔を広げた場合、高周波数帯域で空間エイリアシングの影響で話者方向によっては偽りの方向に鋭いピークが発生することがある。その結果、ピーク探索を行なうべく評価関数は多峰性関数となり、ピーク探索に過剰な演算負荷を要する傾向にある。

本研究では、ピーク探索に確率的多点探索アルゴリズムである PSO (Particle Swarm Optimization) を適用する手法を検討する。

3. 研究の方法

図1に検討手法の流れを示す。まず、2マイクロホンの受信信号を DFT で周波数領域に変換し、周波数毎にマイクロホン間の相関行列を算出する。相関行列の要素である相関関数はマイクロホン間の音波の到達時間差に依存する。到達時間差は音源方向に依存するため、到達時間差から音源方向を推定可能である。本研究では、到達時間差 (音源方向) の推定法として MUSIC を用いる。

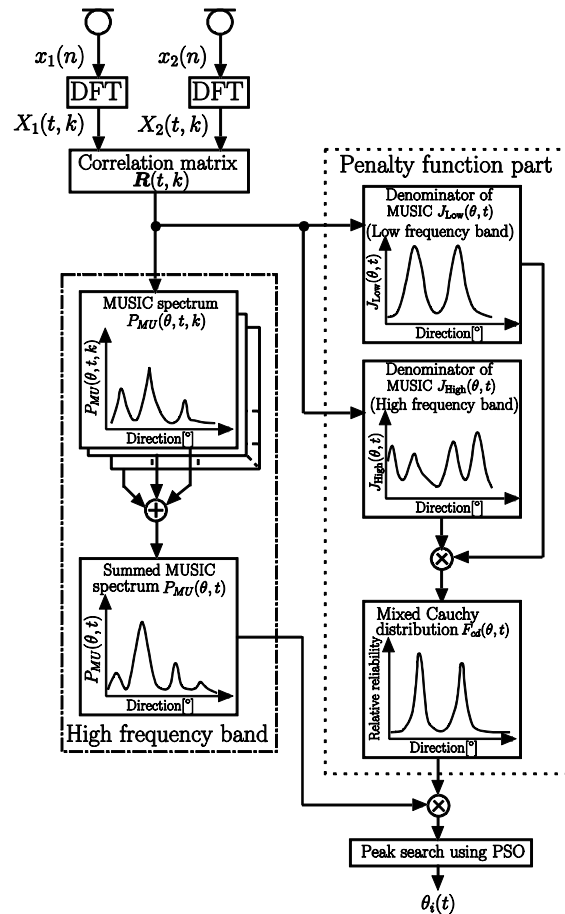


図1: 検討手法の流れ

音源信号である音声は「時間 - 周波数」平面上で信号エネルギーが疎らに存在するスパース性が仮定できる。そのため、複数音源が同時に存在しても、1音源エネルギーが支配的な「時間 - 周波数」領域が多数存在する。そのような領域での音源方向推定は成功する。問題は、そのような領域の推定結果の選

択法である。本研究では、MUSIC の評価関数である MUSIC スペクトルに対して、ペナルティ関数を乗算し、正解方向を強調する。

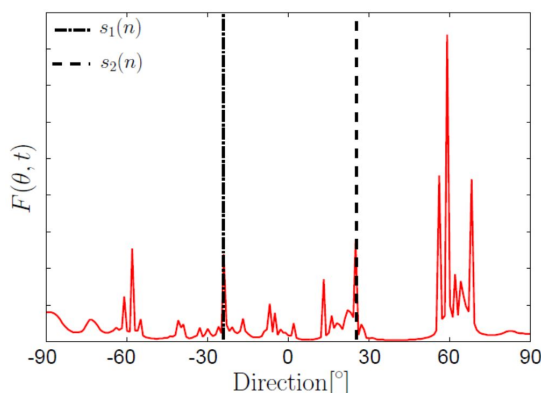


図 2：加算 MUSIC スペクトル

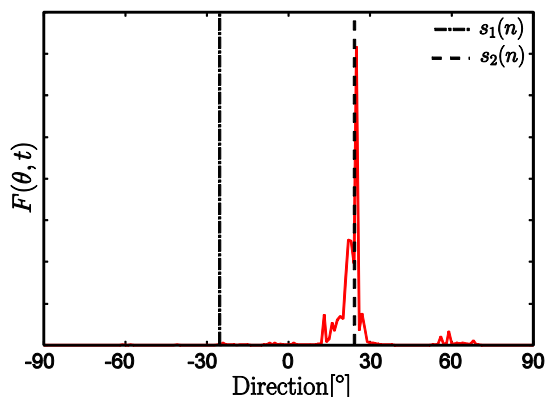


図 3：ペナルティ関数を乗じたスペクトル

MUSIC は狭帯域信号に対する手法であるため、MUSIC スペクトルは周波数ごとに得られる。音源方向推定には、全周波数の MUSIC スペクトルを加算する。その結果、正解方向が強調される。しかし、空間解像度の高い高周波数帯域では、エイリアシングの影響を受け、話者方向によっては偽りの方向に鋭いピークが発生するため、加算 MUSIC スペクトルは図 2 に示すような多峰性関数となる。音源方向に対応するピークのみを検出するため、ペナルティ関数を乗ずる。ペナルティ関数には空間解像度は低い、空間エイリアシングの生じない低周波数帯域の MUSIC スペクトルの分母項を逐次更新して算出する。図 3 に 2 音源に対して 30° 方向のみを強調するペナルティ関数を乗じた MUSIC スペクトルを示す。このような手順で各音源に対するピークを強調する。ペナルティを乗じた後のスペクトル関数は常に話者数を 1 と考えるため、複数話者が評価関数上に同時に存在する際に発生する同一話者追尾問題は考慮する必要がない。以上の結果、空間解像度の高い評価関数を用いて同一話者追尾問題を考慮せずに扱うことが可能となり、本研究で解決を目指した問題をクリアできる。

ピークの検出には、PSO を用いる。PSO は

強い集中化能力をもつ探索手法であり、一般に多数のパーティクルを要するが、ペナルティ関数を乗じた MUSIC スペクトルは鋭いピークを有するため、高々数個のパーティクルを用いた 1 回の更新のみでピークを検出可能である。

4. 研究成果

検討手法の有効性を実環境実験で確認した。音源にはスピーカを用い、人が手でもって移動した。音源信号には音声をを用い、2 音源が同時に移動した。マイクロホン間隔は 20[cm]とし、サンプリング周波数は 8000[Hz]とした。図 3 に実験結果を示す。図 3 に示されるように、無音区間から発話区間に復帰した際に、一方の話者に追尾結果が偏ることなく、複数話者追尾に成功していることが確認できる。全部で 15 通りの音声の組み合わせに対して実験を行なった結果、平均 RMSE が 5.338°, 平均 RTF が 0.108 であり、高い追尾性能を実時間処理で実現したことが確認できた。

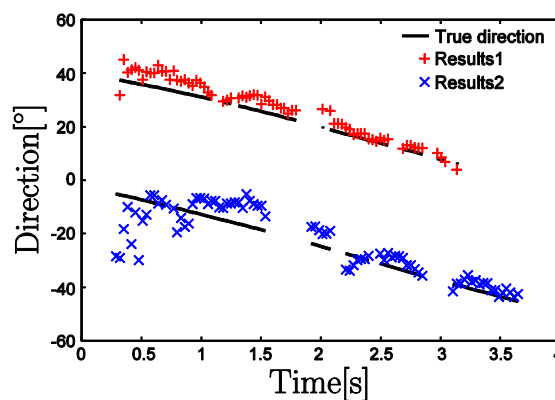


図 3：音源追尾結果

5. 主な発表論文等

〔雑誌論文〕(計 1 件)

平川 真里, 陶山 健仁, Kinect 搭載の不等間隔マイクロホンアレイによる 2 音源追尾, 査読有, 電子情報通信学会論文誌, Vol. J98-A, No. 11, 2015, pp. 600-609.

〔学会発表〕(計 10 件)

大宮 健太, 陶山 健仁, 瞬時位相差に基づく逐次更新ヒストグラムを用いた複数音源追尾, 電子情報通信学会技術報告, SIP2017-62, 2017, pp. 79-84
Kenta Omiya and Kenji Suyama, Multiple Sound Source Tracking Using Low Complexity Directional Estimation, Proceedings of IEEE ISCIT2017, 2017, pp. 355-360

大宮 健太, 陶山 健仁, 三角関数の近似による複数音源追尾, 第 30 回回路とシステムワークショップ予稿集, 2017, pp.271-276

Kenta Omiya and Kenji Suyama, Two Microphone Approach for Multiple Sound Source Tracking, Proceedings of IEEE ISESD2016, 2016, PID-93

Kenta Omiya and Kenji Suyama, Sound Source Tracking via Two Microphones Based on MUSIC Using PSO, Proceedings of IEEE ISPACS2016, 2016, pp.50-55

大宮 健太, 陶山 健仁, MUSIC に基づく 2 マイクホン音源追尾の一検討, 電子情報通信学会技術報告, SIP2016-51, 2016, pp.91-96

陶山 健仁, マイクホンアレーによる音源追尾技術, 第 29 回回路とシステムワークショップ予稿集, 2016, pp.181-183

大宮 健太, 平川 真里, 陶山 健仁, 可変探索パラメータをもつ PSO による 2 マイクホン音源追尾, 電子情報通信学会技術報告, SIP2015-137, 2016, pp.129-133

Yusuke Shiiki and Kenji Suyama, Omnidirectional Sound Source Tracking Based on Sequential Updating Histogram, Proceedings of APSIPA ASC 2015, 2015, pp.1249-1256

Yusuke Shiiki and Kenji Suyama, Multiple Omnidirectional Sound Source Tracking using Sequential Updating Histogram, Proceedings of ISPACS2015, 2015, pp.112-117

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

取得状況(計 0 件)

〔その他〕

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