

科学研究費助成事業（学術研究助成基金助成金）研究成果報告書

平成25年6月4日現在

機関番号：12102
 研究種目：若手研究（B）
 研究期間：2011～2012
 課題番号：23760325
 研究課題名（和文） 二次カーネルに基づく非線形アレー信号処理による
 音源分離・音源定位の研究
 研究課題名（英文） Research about source separation and source localization by
 nonlinear array signal processing based on quadratic kernel
 研究代表者
 宮部 滋樹（MIYABE SHIGEKI）
 筑波大学・システム情報系・助教
 研究者番号：50598745

研究成果の概要（和文）：カーネルを用いた高次元写像による非線形マイクロホンアレー信号処理を用いた音源分離と音源定位に取り組み、特に二次および更に高次の写像を用いた高解像度の音源定位手法の開発に成功した。また、音源定位の際に分析される非線形写像の共分散が、写像の2倍の次数のクロスモーメント分析に相当するという統計的な性質を明らかにした。従来研究されていた音源定位手法よりも計算量・解像度の両面で有意であるということが確認された。

研究成果の概要（英文）：We studied source separation and source localization by nonlinear microphone array signal processing using higher dimensional map with kernels, and successfully developed high-resolution source localization using quadratic or higher-degree mapping. We showed that the covariance analysis of the mapping corresponds to the moment analysis with the order twice of the degree of the map, and the statistical property of the analysis in the source localization is clarified. It is ascertained that the proposed source localization has higher resolution and is more computationally efficient than the conventional methods.

交付決定額

(金額単位：円)

| | 直接経費 | 間接経費 | 合計 |
|-------|-----------|-----------|-----------|
| 交付決定額 | 3,400,000 | 1,020,000 | 4,420,000 |

研究分野：通信・ネットワーク工学

科研費の分科・細目：知覚情報処理

キーワード：パターン認識

1. 研究開始当初の背景

音源分離と音源定位は、音声理解、環境理解、音楽分析など、音の情報処理の全ての分野における情報抽出のための根本課題である。そのアプローチとして最も成功を収めているのがマイクロホンアレーによる多チャネル処理である。音の観測は画像などと違って線形に重ね合わさった波の観測しか得られず、分析対象の個々の音源を別々に収録できる特別な環境でないかぎり、目的の音を取り出す処理を施さなければあらゆる音の情報分析は破綻してしまう。このうち音源定位

は、マイクロホンアレーから音の方位を取り出す処理であり、話者の方位が目的となる情報処理、あるいは音源分離・音声強調の事前情報を与える処理として非常に重要な問題である。しかし古典的な線形マイクロホンアレー信号処理は、音源の数がマイクロホンアレーの素子数以下または未満であることが前提となった信号の部分空間分析に基づいているため、音源のスパース性を利用するなどの近似的な仮定を入れる必要があった。スパース性を導入したアプローチは現在でも盛んに研究されているが、近似的にしか成立

しないスパース性の仮定は、線形アレー信号処理に対する性能劣化は不可避である。

2. 研究の目的

すでに一般的となっていたスパース性を導入するアプローチに対して我々は、二次関数を用いた観測信号の高次元写像によって分析する部分空間を拡大し、写像先で線形アレー信号処理を実施することにより、スパース性の仮定を排除するアプローチを見出していた。本研究課題ではこの高次元写像に基づく非線形アレー信号処理の音源分離と音源定位に対する応用可能性を探った。

3. 研究の方法

本手法の仕組は、非線形写像を用いて観測信号空間の次元を拡大したのち、拡大された空間において線形アレー信号処理を施すという、線形アレー信号処理の非線形な拡張である。この枠組みを、音源数が素子数未満となる線形アレー信号処理でも扱える問題、および音源数が素子数よりも多く線形アレー信号処理が扱えない問題に対して適用し、従来よりも性能の高い音源定位を試みる。当初の目的は、音源定位だけでなく音源分離と両方の問題の解決を目指したが、信号復元を目的とする音源分離問題では、非線形写像による音のひずみの解決が難しく、研究の方向性は音源定位のみに的を絞ることとなった。しかし音源定位においては非線形写像による歪は問題とならず、二次写像よりも高次の写像を問題なく利用することができるため、その性質解析と性能評価を行い、当初の見込みよりも大幅に良い結果が得られた。

この研究の土台としては、線形アレー信号処理の中では最も高精度な音源定位を達成する multiple signal classification (MUSIC) を用いる。MUSIC は、信号部分空間が音源の伝達関数で張られるという性質を利用し、雑音部分空間と呼ばれる信号部分空間の直交補空間の分析に基づく手法で、雑音部分空間を用いる手法よりも解像度が高いのが利点である。しかし雑音部分空間を得るためには音源数がマイクロホン素子数未満である必要があり、また雑音部分空間の次元数が大きいほど精度が上がるため、提案する高次元写像によって次元を拡大する効果が非常に大きいと期待できる。

高次元写像を用いたアレー信号処理を構築するために、以下のサブテーマを設定する。

(1) 写像の幾何学的・統計的解釈

これまでに我々が提案した手法は、パターン認識の分野で利用されるカーネル関数という表現法を用いていたが、具体的にどのような写像が行われるのかを明に書き出す写像ではないため、その幾何学的構造がわかりづらい。また、計算に使用するグラム行列が

観測サンプル数次元に一致して非常に大きなものになってしまい、計算コストの面でも非効率である。そのためカーネルを用いずに写像関数を書き出した写像の構造分析が必要である。またこれにより、音源定位の際にどのような統計的分析が行われているのかも明らかになる。

(2) 写像関数の最適化と音源定位

これまでに我々が研究してきたのは二次の写像であるが、逆写像が必要となる音源分離と異なり、取り出す情報が元の観測信号と違う領域で、かつ方位情報として縮退したものを取り出す音源定位問題では、写像関数による信号の変形を気にすることなく高次のものを用いることができる。そこで、高次の情報を取り出すための写像関数の設計について取り組む。そして設計した写像を基にした MUSIC に基づく音源定位を評価する。

4. 研究成果

当初の目的としては二次の写像を用いた音源分離と音源定位の確立であった。しかし音源分離においては写像による出力信号の歪の解決に至らず、主に音源定位に重点を置いた研究をすることになった。しかし、音源定位への写像の効果は予想をはるかに上回って大きく、また高次写像関数の分析も予定よりもはるかに進んだ。そのため研究課題全体としては当初の予定よりもはるかに大きな成果が得られたといえる。

(1) 写像の幾何学的・統計的解釈

従来取り組んでいた手法ではカーネル関数による内積の置き換えを用いた写像を行っていたが、内積がカーネル関数と一致する解析的な写像を求めることができればカーネル関数よりも効率的な計算と性質解析を行うことができる。内積の性質を分析するうちに、2次カーネルによる写像は直積により得られる行列を縦ベクトルとして並べたおしたものと一致するということがわかった。また、この写像の共分散行列は4次のクロスモーメントに一致し、提案手法が高次統計量を用いた枠組みであるということがわかった。

高次クロスモーメント量を利用する分析であるという提案する写像 MUSIC の正体が明らかになったため、従来の高次クロスモーメントを用いた 2 q -MUSIC との類似性がわかり、比較を行った。その結果、モーメントはキュムラントよりも統計的バイアスに対して頑健であるため、提案手法の方が高い精度の推定が行えるということがわかった。また、キュムラントの計算には複数次数のモーメントの分析が必要なために計算量が非常に大きい。それに対して提案手法は、写像の共分散行列という単純な形式で単一のモーメントのみを利用するため、演算量において従

来手法を大幅に凌駕していることがわかった。

(2) 写像関数の最適化

2次写像の分析から天下りの、直積演算を繰り返すことによって任意次数の高次元写像を求めることができることがわかった。また自明な統計的解釈として、高次写像を用いた音源定位は写像次数の2倍の次数のクロスモーメント分析であるということがわかった。これの統計的解釈により、長いデータに対しては高次写像が非常に有効であるが、短いデータに対しては2次程度の低い写像に留める方が性能が高いという、写像次数と統計的バイアスのトレードオフ関係が明らかになった。

また、単一次数の高次写像を用いるよりも、複数次数の写像の直和によって更に解像度が上昇するということがわかった。これは複数の次数のクロスモーメントの同時分析に相当する。

図1に、マイク数4、音源5の条件における、 $2q$ -MUSIC、写像 MUSIC と、更に改良した複数次数の写像 MUSIC の性能比較を示す。写像 MUSIC は $2q$ -MUSIC よりも誤差が小さく、また複数次数を評価する提案手法は最も解像度が高い。この提案手法は、素子数が3以上、音源数が2~5程度の条件では、これまでに提案されたどの音源方位推定手法よりも推定解像度が高いものであると思われる。

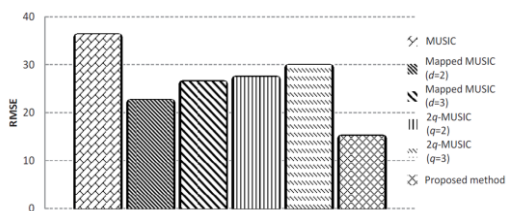


図1. 音源定位精度

最後に関連する成果について述べる。素子数がある程度大きい場合は MUSIC に基づく音源定位が有効であるが、特に素子数が少ない2素子のような場合には、雑音部分空間よりも信号部分空間を用いた方が有効であり、現在その方法について取り組んでいる。また、カーネルで表される写像では歪が大きくなるため、統計的性質よりも物理的性質を考慮して次元数を拡大する方が有効であることがわかり、スパース性の仮定を入れた波面モデルに基づいて観測空間のバーチャル観測を加える写像を提案した。

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6件)

- (1) Hiroki Katahira, Nobutaka Ono, Shigeki Miyabe, Takeshi Yamada, Shoji Makino, "Virtually Increasing Microphone Array Elements by Interpolation in Complex-Logarithmic Domain," Proc. European Signal Processing Conference (EUSIPCO), September 2013. (accepted) (査読有)
- (2) Shigeki Miyabe, Nobutaka Ono, Shoji Makino, "Blind compensation of inter-channel sampling frequency mismatch with maximum likelihood estimation in STFT domain," Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 674.678, May 2013. (査読有)
- (3) Ryutaro Sakanashi, Shigeki Miyabe, Takeshi Yamada, Shoji Makino, "Comparison of superimposition and sparse models in blind source separation by multichannel Wiener filter," Proc. APSIPA ASC, December 2012. (査読有)
- (4) Shigeki Miyabe, Takaya Hayashi, Takeshi Yamada, Shoji Makino, "Distance attenuation control of spherical loudspeaker array," Proc. APSIPA ASC, December 2012. (査読有)
- (5) Yuya Sugimoto, Shigeki Miyabe, Takeshi Yamada, Shoji Makino, Bing-Hwang (Fred) Juang, "Underdetermined DOA estimation by the non-linear MUSIC exploiting higher-order moments," Proc. International Workshop on Acoustic Signal Enhancement (IWAENC), 2012, September 2012. (査読有)
- (6) Shin Miura, Hirofumi Nakajima, Shigeki Miyabe, Shoji Makino, Takeshi Yamada, Kazuhiro Nakadai, "Restoration of clipped audio signal using recursive vector projection," Proc. TENCON, pp. 394-397, November 2011. (査読有)

[学会発表] (計12件)

- ① 宮部滋樹, 小野順貴, 牧野昭二, "非同期録音ブラインド同期のための線形位相補償の効率的最尤解探索," 日本音響学会講演論文集 2-10-4, pp. 733. 736, 2013年3月13日~15日, 東京工科大学 (東京都).
- ② 片平拓希, 小野順貴, 宮部滋樹, 山田武志, 牧野昭二, "複素対数補間によるヴァーチャル観測に基づく劣決定条件での音声強調," 日本音響学会講演論文集 2-10-6, pp. 741. 744, 2013年3月13日~15日, 東京工科大学 (東京都).

- ③ 栗花悠輔, 宮部滋樹, Tomasz Rutkowski, 松本佳泰, 山田武志, 牧野昭二, “多変量確率モデルによる脳波の信号分離,” 電子情報通信学会技術研究報告 MBE2012-119, pp. 161. 166, 2013 年 3 月 15 日, 玉川大学 (東京都) .
- ④ 宮部滋樹, 小野順貴, 牧野昭二, “非同期マイクロホンアレーのためのサンプリング周波数ミスマッチのブラインド補償,” 電子情報通信学会技術研究報告 EA2012-111, pp. 11. 16, 2012 年 12 月 13 日, 国立情報学研究所.
- ⑤ 坂梨龍太郎, 小野順貴, 宮部滋樹, 山田武志, 牧野昭二, “単一音源区間情報を用いた非同期マイクロホンアレーによる音声強調,” 電子情報通信学会技術研究報告 EA2012-112, pp. 17. 22, 2012 年 12 月 13 日, 国立情報学研究所.
- ⑥ 坂梨龍太郎, 宮部滋樹, 山田武志, 牧野昭二, “多チャンネルウィーナーフィルタを用いた音源分離における観測モデルの調査,” 日本音響学会講演論文集 1-P-14, pp. 757. 760, 2012 年 9 月 19 日, 信州大学.
- ⑦ 宮部滋樹, 小野順貴, 牧野昭二, “非同期録音信号の線形位相補償によるブラインド同期と音源分離への応用,” 日本音響学会講演論文集 3-9-8, pp. 689. 690, 2012 年 9 月 21 日, 信州大学.
- ⑧ 杉本侑哉, 宮部滋樹, 山田武志, 牧野昭二, “高次モーメント分析に基づく非線形 MUSIC による劣決定方向推定,” 電子情報通信学会技術研究報告 EA2012-41, pp. 49. 54, 2012 年 6 月 8 日, ヤマハ株式会社.
- ⑨ 坂梨龍太郎, 宮部滋樹, 山田武志, 牧野昭二, “多チャンネルウィーナーフィルタを用いた音源分離におけるスパースモデルと最小二乗誤差モデルの性能比較,” 電子情報通信学会技術研究報告 EA2012-43, pp. 61. 66, 2012 年 6 月 8 日, ヤマハ株式会社.
- ⑩ 林貴哉, 宮部滋樹, 山田武志, 牧野昭二, “球状スピーカアレーを用いた放射特性制御のシミュレーション,” 電子情報通信学会技術研究報告 EA2012-36, pp. 19. 24, 2012 年 6 月 8 日, ヤマハ株式会社.
- ⑪ 三浦晋, 宮部滋樹, 山田武志, 牧野昭二, 中島弘史, 中臺一博, “時間周波数領域におけるグリッド間の整合性に基づくクリッピングの除去,” 日本音響学会講演論文集 1-Q-10, pp. 843. 846, 2012 年 3 月 13 日, 神奈川大学.
- ⑫ 杉本侑哉, 宮部滋樹, 山田武志, 牧野昭二, “高次相関を用いた非線形 MUSIC による高分解能方位推定,” 日本音響学会

講演論文集 3-1-6, pp. 763. 766, 2012 年 3 月 15 日, 神奈川大学.

6. 研究組織

(1) 研究代表者

宮部 滋樹 (MIYABE SHIGEKI)

筑波大学・システム情報系・助教

研究者番号：50598745