様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　４ 年　 ６ 月 ２２ 日現在

機関番号：１４３０１
研究種目：挑戦的研究（萌芽）
研究期間：2020～2021
課題番号：２０Ｋ２１８１３

研究課題名（和文）あらゆる音の定位・分離・分類のためのユニバーサル音響理解モデル

研究課題名（英文）A Universal Audio Understanding Model for Localization, Separation, and Classification of Various Sounds

研究代表者
吉井　和佳（Yoshii, Kazuyoshi）

京都大学・情報学研究科・准教授

研究者番号：２０５１０００１

交付決定額（研究期間全体）：（直接経費）　　4,900,000円

研究成果の概要（和文）：本研究の目的は、音声・音楽・環境音など多岐にわたるあらゆる種類の音を、適応的かつ頑健に分析できるユニバーサル音響理解モデルを確立することである。具体的には、最近我々が提案した、高速かつ高精度な最新の汎用ブラインド音源分離（BSS）手法である高速多チャネル非負値行列因子分解（FastMNMF）に関して、音源モデル・空間モデル・尤度関数の改良を行い、分離モデルや残響モデルとの同時学習を実現した。また、音声認識との統合についても取り組んだ。

研究成果の学術的意義や社会的意義
本研究を通じて、人間が持つ音理解能力の創発的な側面、すなわち、正解の教示を受けなくても、多様な音が重畳する実環境とのインタラクションを通じて、音を個別に理解する能力に対し、一定の構成論的説明と統計的エビデンスを与えることができた。技術的には、ペアデータを前提とした深層学習モデルの教師あり学習から脱却し、尤度最大化の枠組みに基づく教師なし学習を主軸とすることで、大規模な音響信号データ利用への道筋を開いた。

研究成果の概要（英文）：Our goal is to formulate a universal audio understanding model for various kinds of sounds including speech, music, and environmental sounds. More specifically, we have improved the source and spatial models and the likelihood function of the state-of-the-art blind source separation (BSS) method called FastMNMF and achieved joint optimization of FastMNMF with separation and reverberation models. We also tackled integration of speech enhancement and recognition.

研究分野：音響信号処理

キーワード：音響信号処理　音源分離　残響除去　深層学習　最尤推定　音声強調　音声認識

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

音響信号処理分野では、近年の深層学習の台頭で目覚ましい発展に伴って、音声以外の様々な音が解析対象とされているが、本質的な限界は以下の二点であった。

(1) 解析したい音の特性に応じて数理モデルを人手で定式化する点

音源分離、音声強調、雑音抑圧、残響除去などの目的ごとにモデルが多数考案され、それらをアドホックに組み合わせることも多く、分野全体の見通しが悪くなっている。近年の深層学習ブームは状況を一層悪化させており、個別タスクの精度向上を追い求めた DNN のアーキテクチャ改善競争が過熱していた。

(2) 数理モデルを学習するために大量の教師データを必要とする点

大量のペアデータを用いて DNN を教師あり学習するアプローチは、実用面では有益であるが、人間の理解能力とは様式が異なる。人間は、初めて聴く音でも、過去の経験と照らし合わせながら、適切な類型化・体系化を行うことができる。しかし、計算機上での教師なし音響理解の試みとして、一般の音を対象とした取り組みが必要であった。

２．研究の目的

本研究の目的は、屋内外の多様な音響条件のもとで、音声・音楽・環境音など多岐にわたる種類の音を、**適応的かつ頑健に分析できるユニバーサル音響理解モデル**を確立することである。これにより、人間が持つ音響理解能力の創発的側面、すなわち、正解の教示を受けなくても、多様な音が重畳する実環境とのインタラクションを通じて、個別の音源を類型化する能力に対し、構成論的説明と統計的エビデンスを与えることを主眼とした。

本研究を通じて、確率的な枠組みに基づく音響表現学習の体系化を行うとともに、音声強調や音楽分離など、各種タスク特化型の End-to-End アプローチ、すなわち、ペアデータを両エンドとした**教師あり学習一辺倒なアプローチからのパラダイムシフト**を目標とした。

３．研究の方法

本研究では、上記問題を一挙に解決するため、物理拘束に立脚した音響信号の統一的な深層生成モデルの定式化と、その逆問題としての教師なし学習について取り組んだ。

(1) 任意の空間/音特性を表現可能な「ユニバーサル音響生成モデル」の定式化

音の伝達特性を示す空間モデルと、時間周波数構造を示す音源モデルを内包する多チャネル音響信号の統一的な深層生成モデルを定式化した。いずれのモデルも、内部に環境/音源特性を精緻に表現可能な深層構造をもち、環境変化への追従を試みた。

(2) 混合音に対する「ユニバーサル音響理解モデル」への拡張

上記生成モデルの逆問題ソルバとして、観測音響信号の潜在状態を推定する深層分離モデルの利用を考案した。具体的には、変分自己符号化器 (VAE) を構成し、音響生成モデルと理解モデルを同時最適化することにより、教師なし/半教師あり学習を試みた。

(3) 音声認識タスクと「ユニバーサル音響理解モデル」の同時最適化

ユニバーサル音響理解モデルは、空間/音源情報を手掛かりとする聴覚抹消系に対応し、音声認識は、記号接地を担う聴覚中枢系に対応している。そこで、両者のモジュラビリティを保持したまま、尤度基準での確率的統合および同時最適化を試みた。

４．研究成果

本研究を通して得られた主要な研究成果について報告する。まず、**研究方法(1), (2)に関して、汎用ブラインド音源分離 (BSS) 技術の本質的な進展**を得た。これまで、BSS 問題においては、観測音の生成モデルに基づく統計的アプローチが広く用いられてきた。中でも、多チャネル非負値行列因子分解 (MNMF) [Sawada+ 2013] は、フルランクな空間相関行列に基づく空間モデルと、NMF に基づく音源モデルを統合した汎用的 BSS 手法として注目されている。しかし、計算負荷が過大であり、局所解に陥りやすい欠点があった。独立低ランク行列分析 (ILRMA) [Kitamura+ 2016] では、空間相関行列をランク 1 に制限することで、計算量を削減し、初期値依存性を軽減しているが、残響に対する頑健性が低下する代償があった。最近我々が提案した、高速かつ高精度な最新の BSS 手法である高速多チャネル非負値行列因子分解 (FastMNMF) [Sekiguchi+ 2020] では、空間相関行列のフルランク性を保ちつつも、同時対角化制約を導入することで、計算量の削減と分離精度の改善を両立することに成功している。本研究では、FastMNMF 各部の本質的な拡張を行い、確率的な枠組みと深層学習の融合に成功した（図 1）。

(1) 深層学習に基づく音源モデルの表現力向上

FastMNMF では、各音源スペクトログラムのパワースペクトル密度が低ランク構造を持つと仮定しているため、ある種の音源（例 : 音声スペクトログラム）に対してこの仮定は成り立たなかった。そこで、雑音環境下での音声強調を目的として、音声に対しては変分自己符号化器 (VAE) に基づく深層生成モデルを用い、雑音に対しては NMF に基づく低ランクモデルを用いた音声強調法を提案し、継続的に改良を行った。VAE はあらかじめ事前学習しておくことが想定されるが、理論上は教師なし学習が可能な枠組みを確立した。

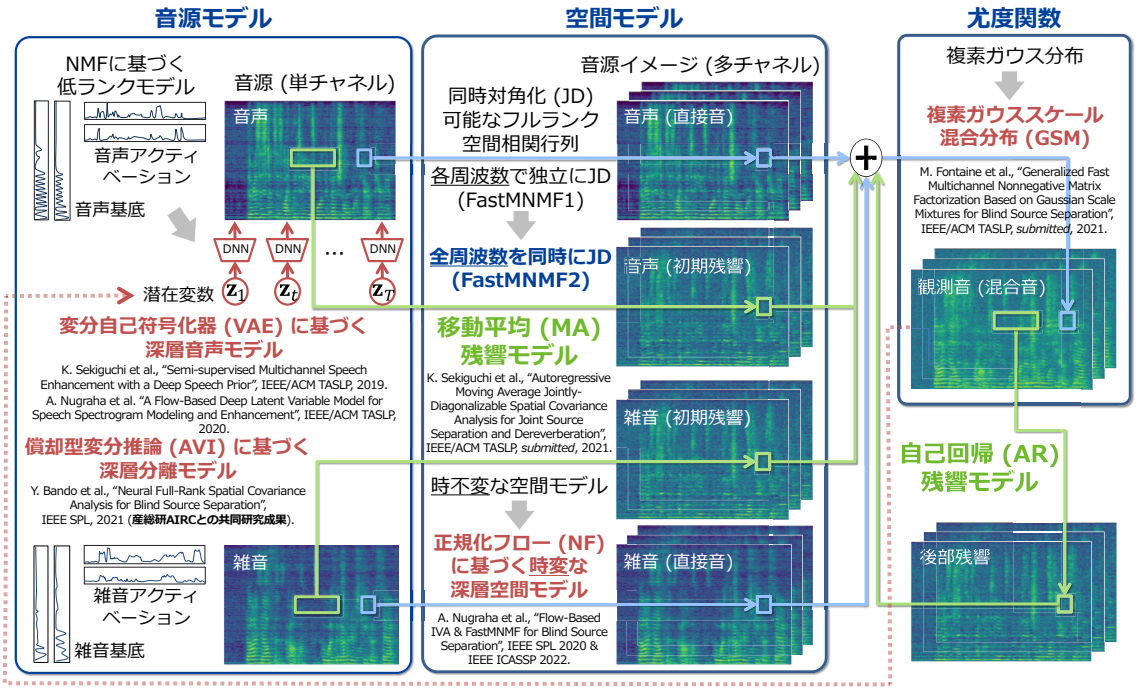

図 1：GSM-FastMNMF と深層学習の融合に基づく汎用的ブラインド音源分離・残響除去手法

(2) 深層学習に基づく空間モデルの表現力向上

独立成分分析 (ICA) [Smaragdis 1998] や独立ベクトル分析 (IVA) [Kim+ 2006] などの線形時不変型決定系 BSS 手法は、マイク数と音源数が等しい決定条件のもとで、周波数ごとに、混合音スペクトルを統計的に独立な要素からなる音源スペクトルに変換するための時不変な分離行列を推定する。この種の BSS では、分離行列と混合行列は逆行列の関係にあり、混合音スペクトルと音源スペクトルは可逆な確率変数である。ここで、このような確率変数の可逆変換は、正規化フロー (NF) の一種であるとみなせる。そこで、FastMNMF の空間モデルに含まれる対角行列（空間変換行列）を NF で表現することで、音環境の変化（音源移動）にも対応可能な線形時変非決定系 BSS 手法の導出に成功した。

(3) 同時的音源分離・残響除去のための残響モデルの統合

実環境下音声認識において、音声強調と残響除去はいずれも重要な役割を果たしており、相互依存関係のある両タスクを一挙に実行することが望ましい。そこで、自己回帰移動平均 (ARMA) 過程に基づく残響モデルを FastMNMF に統合することで、音源分離と残響除去を高速かつ一挙に行う手法を提案した。具体的には、NMF に基づく音源モデルに従う各音源信号に対し、移動平均 (MA) 過程に従って初期残響（音源に依存）が付加され、それらが重畳した混合信号に対し、AR 過程に従って後部残響（音源に非依存）が付加される生成モデルを考案し、混合音からの教師なし学習を実現した。

(4) 裾の重い確率分布に基づく頑健性向上

通常、混合音スペクトルに対する尤度関数として、時変な空間相関行列をもつ複素ガウス分布を用いるのが一般的であり、突発性音源に対しては性能が劣化しやすい問題があった。そこで、裾の重い分布であるガウススケール混合分布 (GSM) に基づく FastMNMF の拡張を考案し、EM アルゴリズムに基づく統一的なパラメータ推定法を導出した。これまで独立に提案されてきた様々な裾の重い分布（t 分布、一般化ガウス分布 (GG)、α-対称安定分布など）に基づく FastMNMF を統一的に説明することに成功し、一般化双曲型分布 (GH) やその特殊系である正規逆ガウス分布 (NIG) が高い性能を示すことを発見した。

(5) 深層学習に基づく分離モデルの教師なし学習

通常、深層学習を用いた音源分離を行うには、混合音と音源信号とのペアデータから分離モデルを教師あり学習する必要があったが、テスト環境を網羅的にカバーする学習データを集めることは現実的ではないため、性能に限界があった。そこで、FastMNMF の基盤となる、音源信号から多チャネル混合音が生成される生成モデルに対して、混合音から音源信号を推定する深層分離モデルを導入することで、償却型変分推論 (AVI) の枠組みを用いて、両者を一挙に同時学習する方法を考案した。

一方で、**研究方法(3)に関して、汎用 BSS 技術のオンライン音声認識との統合**についても取り組んだ。一連の基礎研究を通じて開発した FastMNMF 及びその拡張は、汎用 BSS 手法として性能や頑健性に加えて、教師あり学習が不要である点でも優れているが、推論時の計算量は依然として大きく、リアルタイム動作は容易ではなかった。そこで、深層学習を用いた音声マスク推定に基づくビームフォーミングは推論時は高速に動作することに着目し、FastMNMF をバックエンドに、ビームフォーミングをフロントエンドとした音声強調部を構築し、音声認識部と一体的に運用することを試みた。両者の同時最適化については予備的なテスト段階であるが、挑戦的研究（萌芽）としては、当初の想定以上の成果を上げることができた。

５．主な発表論文等

〔雑誌論文〕　計7件（うち査読付論文　7件／うち国際共著　2件／うちオープンアクセス　2件）

１．著者名 Yoshiaki Bando, Kouhei Sekiguchi, Yoshiki Masuyama, Aditya Arie Nugraha, Mathieu Fontaine, Kazuyoshi Yoshii	４．巻 28
２．論文標題 Neural Full-Rank Spatial Covariance Analysis for Blind Source Separation	５．発行年 2021年
３．雑誌名 IEEE Signal Processing Letters	６．最初と最後の頁 1670-1674
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/lsp.2021.3101699	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Takayuki Nakatsuka, Kazuyoshi Yoshii, Yuki Koyama, Satoru Fukayama, Masataka Goto, Shigeo Morishima	４．巻 29
２．論文標題 MirrorNet: A Deep Reflective Approach to 2D Pose Estimation for Single-Person Images	５．発行年 2021年
３．雑誌名 Journal of Information Processing	６．最初と最後の頁 406-423
掲載論文のDOI（デジタルオブジェクト識別子） 10.2197/ipsjjip.29.406	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Yicheng Du, Robin Scheibler, Masahito Togami, Kazuyoshi Yoshii, Tatsuya Kawahara	４．巻 29
２．論文標題 Computationally-Efficient Overdetermined Blind Source Separation Based on Iterative Source Steering	５．発行年 2021年
３．雑誌名 IEEE Signal Processing Letters	６．最初と最後の頁 927-931
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/lsp.2021.3134939	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Kouhei Sekiguchi, Yoshiaki Bando, Aditya Arie Nugraha, Kazuyoshi Yoshii, Tatsuya Kawahara	４．巻 28
２．論文標題 Fast Multichannel Nonnegative Matrix Factorization With Directivity-Aware Jointly-Diagonalizable Spatial Covariance Matrices for Blind Source Separation	５．発行年 2020年
３．雑誌名 IEEE/ACM Transactions on Audio, Speech, and Language Processing	６．最初と最後の頁 2610～2625
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TASLP.2020.3019181	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

1．著者名 Aditya Arie Nugraha, Kouhei Sekiguchi, Mathieu Fontaine, Yoshiaki Bando, Kazuyoshi Yoshii	4．巻 27
2．論文標題 Flow-Based Independent Vector Analysis for Blind Source Separation	5．発行年 2020年
3．雑誌名 IEEE Signal Processing Letters	6．最初と最後の頁 2173～2177
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/LSP.2020.3039944	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

1．著者名 Mathieu Fontaine, Kouhei Sekiguchi, Aditya Arie Nugraha, Yoshiaki Bando, Kazuyoshi Yoshii	4．巻 30
2．論文標題 Generalized Fast Multichannel Nonnegative Matrix Factorization Based on Gaussian Scale Mixtures for Blind Source Separation	5．発行年 2022年
3．雑誌名 IEEE/ACM Transactions on Audio, Speech, and Language Processing	6．最初と最後の頁 1734～1748
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TASLP.2022.3172631	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1．著者名 Kouhei Sekiguchi, Yoshiaki Bando, Aditya Arie Nugraha, Mathieu Fontaine, Kazuyoshi Yoshii, Tatsuya Kawahara	4．巻 -
2．論文標題 Autoregressive Moving Average Jointly-Diagonalizable Spatial Covariance Analysis for Joint Source Separation and Dereverberation	5．発行年 2022年
3．雑誌名 IEEE/ACM Transactions on Audio, Speech, and Language Processing	6．最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

〔学会発表〕　計8件（うち招待講演　0件／うち国際学会　5件）

1．発表者名 Mathieu Fontaine, Kouhei Sekiguchi, Aditya Arie Nugraha, Yoshiaki Bando, Kazuyoshi Yoshii
2．発表標題 Alpha-Stable Autoregressive Fast Multichannel Nonnegative Matrix Factorization for Joint Speech Enhancement and Dereverberation
3．学会等名 Annual Conference of the International Speech Communication Association (Interspeech)（国際学会）
4．発表年 2021年

１．発表者名
Yoshiaki Bando, Kouhei Sekiguchi, Kazuyoshi Yoshii

２．発表標題
Gamma Process FastMNMF for Separating an Unknown Number of Sound Sources

３．学会等名
European Signal Processing Conference (EUSIPCO)（国際学会）

４．発表年
2021年

１．発表者名
田中啓太郎，錦見亮，坂東宜昭，吉井和佳，森島繁生

２．発表標題
変分自己符号化器を用いた距離学習による楽器音の音高・音色分離表現

３．学会等名
情報処理学会 第131回音楽情報科学研究会

４．発表年
2021年

１．発表者名
Kouhei Sekiguchi, Yoshiaki Bando, Aditya Arie Nugraha, Mathieu Fontaine, Kazuyoshi Yoshii

２．発表標題
Autoregressive Fast Multichannel Nonnegative Matrix Factorization for Joint Blind Source Separation and Dereverberation

３．学会等名
IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)（国際学会）

４．発表年
2021年

１．発表者名
Keitaro Tanaka, Ryo Nishikimi, Yoshiaki Bando, Kazuyoshi Yoshii, Shigeo Morishima

２．発表標題
Pitch-Timbre Disentanglement of Musical Instrument Sounds Based on VEA-Based Metric Learning

３．学会等名
IEEE Internationzal Conference on Acoustics, Speech and Signal Processing (ICASSP)（国際学会）

４．発表年
2021年

１．発表者名
関口 航平, 坂東 宜昭, Aditya Arie Nugraha, Mathieu Fontaine, 吉井 和佳

２．発表標題
ARMA-FastMNMFに基づく同時的ブラインド音源分離・残響除去

３．学会等名
日本音響学会 2021年春季研究発表会

４．発表年
2021年

１．発表者名
Aditya Arie Nugraha, 関口 航平, Mathieu Fontaine, 坂東 宜昭, 吉井 和佳

２．発表標題
NF-IVAに基づく線形時変型決定系ブラインド音源分離

３．学会等名
日本音響学会 2021年春季研究発表会

４．発表年
2021年

１．発表者名
Aditya Arie Nugraha, Kouhei Sekiguchi, Mathieu Fontaine, Yoshiaki Bando, Kazuyoshi Yoshii

２．発表標題
Flow-Based Fast Multichannel Nonnegative Matrix Factorization for Blind Source Separation

３．学会等名
IEEE Internationzal Conference on Acoustics, Speech and Signal Processing (ICASSP) (国際学会)

４．発表年
2022年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関