公開

# 研 究 終 了 報 告 書

## 「統計的時空間モデルに基づく雑踏音環境マッピング」

研究期間： 2020 年 11 月～2023 年 3 月
研 究 者： 坂東 宜昭

### 1. 研究のねらい

近年、清掃ロボットや倉庫の物流ロボットなど一部の領域で自律ロボットが実用化されている一方、日常環境で人と共存しながら活躍するロボットには未だ多くの課題が残されている。特に、人通りの多い空港やビル、繁華街といった無秩序に多くの人が集まる雑踏空間での、周辺環境の頑健な認識は大きな課題である。我々人間は、このような見通しの悪い環境で、障害物に遮蔽されず瞬時に全方位の情報を把握できる聴覚を活用しているが、同等機能を実現するロボット聴覚は未だ実用に至っていない。この理由として、複雑な雑踏音環境では個別の信号処理の組み合わせでは性能を維持できない、常に無数の音が混合される雑踏音環境のラベル付き教師データの収集は事実上困難、ロボット応用に不可欠な実時間動作できるオンライン推論の枠組みの欠如が主因に挙げられる。

本研究課題では、雑踏空間でも周辺環境を頑健に認識できるロボットの実現を目指し、コア技術として「**雑踏音環境を統一的に扱う時空間深層ベイズモデルに基づくロボット聴覚**」を開拓する。特に、最小限の事前情報で、雑踏音環境を精緻かつリアルタイムに認識する雑踏ロボット聴覚の実現をねらう。我々人間は、見通しの悪い展示会場や入り組んだ市街地で、周囲の雰囲気から自身の興味に基づいて目的地へ到達するなど、適切に周囲の音環境を認識して行動している。本研究では、同様の機能を計算機で実現する基盤技術の確立を目指す。

### 2. 研究成果

#### (1)概要

本研究課題では、大きく分けて以下の 3 つのテーマについて研究を実施した。

- 研究テーマ A「ベイズ時空間モデルに基づく音環境マッピング」
- 研究テーマ B「深層生成音源モデルによる精緻化」
- 研究テーマ C「オンライン推論と能動センシング」

テーマ A は、本研究課題の基盤となる音響マッピングの数理モデルの確立である。ここでは、非負値行列因子分解に基づく音響マッピングを開発し、従来の音源定位に基づくカスケード型の枠組みに比べより高い精度で音源位置を推定できることを確認した。また発展成果として、音源数の情報を事前に必要としないガンマ過程高速多チャネル非負値因子分解や、音源定位と分離を単一の枠組みで最適化する MUSIC-CGMM を開発した。また連携応用成果として、マーカレス位置推定法を開発した。テーマ B では、深層生成音源モデルの教師なし学習を実現した。本手法を応用し、これまで難しいとされてきたディナーパーティでの遠隔音声認識の性能向上など発展成果を得ることができた。テーマ C では、オンライン推論と能動センシングに向けて、開発した音響マッピング法の GPGPU 実装を行った結果、実時間で動作できる程度の計算時間となり、リアルタイムシステムの設計指針を得ることができた。

（2）詳細

**研究テーマ A「ベイズ時空間モデルに基づく音環境マッピング」**

**本研究課題の基盤となる雑踏音環境を統一的に理解するためのマッピング法を確立した。**移動ロボットを用いての音源の空間的な配置を推定する音環境マッピングは、知的システムが周辺環境を認識し適切な行動とるために不可欠である。従来の枠組みは初段の音源定位に強く依存した構成となっており、残響や拡散性雑音の強い雑踏環境下では性能を発揮できなかった。そこで本研究では、音源定位の代わりに非負値行列因子分解（NMF）を用いた音源分離に基づく音環境マッピングを提案した（図 1）。本手法は、まずスペクトル特徴に基づき観測信号を分解したあと、個別の音源に対して位置推定するので、雑踏環境下でも安定してマッピングできる。バブルノイズ下での音源マッピングを想定した数値シミュレーションにより提案法の有効性を評価し、ロボット聴覚ソフトウェア HARK により構築した従来の音源定位に基づくマッピングより高い精度で音源位置を推定できていることを確認した。この研究により、従来の「逐次的に空間情報を収集する枠組み」より、本研究の基盤である「雑踏音全体を統一的に扱う枠組み」が有効であることを示すことができた。本研究は人工知能学会 AI チャレンジ研究会[5]にて発表し、**研究会優秀賞を受賞**。

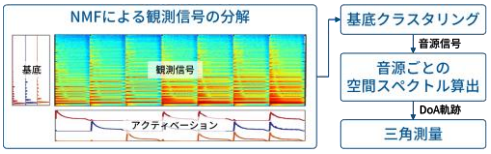


**図 2 開発したマッピング法の概要**

［発展的成果］**複雑な雑踏環境を頑健に分析するため、より情報が豊富な多チャネル信号を扱う NMF の拡張を行った。**NMF を用いた多チャネル音響信号のモデル・音源分離法として高速多チャネル NMF（FastMNMF）が高い性能を発揮することが知られている。しかし本手法は、混合音に含まれる音源数を事前に指定する必要があり、可用性に課題があった。そこで本研究では、ガンマ課程を用いて、FastMNMF が音源数を同時推定できるよう拡張した（図 2）。本拡張は、従来最尤推定として実現されていた FastMNMF を変分ベイズ EM アルゴリズムとして再定式化することで実現した。複数人の音声を混合した数値実験により、観測信号中の音源の個数が未知であっても頑健に音源分離できることを確認し、ヨーロッパ圏最大の信号処理国際会議 EUSIPCO 2021 [4]にて発表した。本研究で得られたガンマ過程の知見は、マッピングで用いている NMF の設計に活用した。

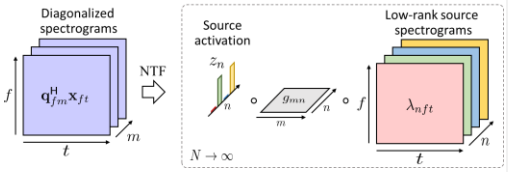


**図 3 ガンマ過程 FastMNMF の概要**

［発展的成果］**更に、音源信号の抽出だけでなく音源到来方向を同時に推定するため、音源分離と音源定位を単一の枠組みで最適化する手法を開発した。**従来の多くのロボット聴覚システムでは、音源定位と音源分離はカスケードに扱われており性能劣化の原因となっていた。また、これらを同時最適化する統計的な枠組みも提案されているが、反復推論のための計算量が膨大となる課題があった。本研究では、従来のカスケード型と同時最適化型の長所を両立するハイブリッド型システムを開発した（図 3）。本手法は、従来のカスケード型の枠組みにおける音源定位を複数のハイパーパラメータで実行し、後段の音源分離の尤度関数を用いて最も良い音源定位結果を選択する。各ハイパーパラメー

タでの試行は独立なので、近年安価になりつつある GPU を用いて容易に高速化できる。これにより、カスケード型の欠点であったハイパーパラメータに対する敏感さと、同時最適化型における計算量の問題を解決できた。本手法は、ノート型コンピュータに搭載された GPU NVIDIA GeForce GTX 1080 Max-Q を用いて実時間動作することを確認した。さらに、実雑踏音に音源信号を重畳した実験で、提案法の有効性を確認した。本研究成果は国際論文誌 IEEE Access にて誌上発表[2]した。

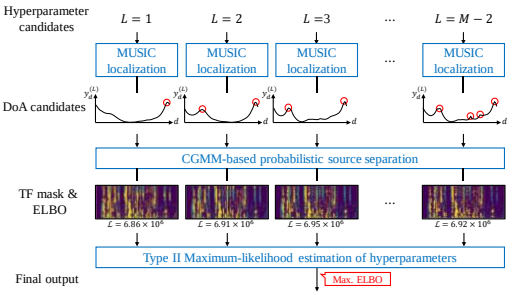


**図 4 MUSIC-CGMM の概要**

［連携成果］**本研究課題の成果応用・連携として、産総研 小木曽 里樹 研究員と連携し、音源地図の応用であるビーコンレス位置推定法を開拓した。**ロボットに限らず人間や物体の位置を計測する技術は、行動計画や生産性向上など様々な産業・学術応用の基盤である。既存の枠組みでは、Bluetooth ビーコンやカメラ、Wi-Fi フィンガープリントを用いる方法などが主流であるが、いずれもビーコン設置のコストやプライバシーの観点から課題があった。そこで本研究では、環境音を計測することで、場所ごとの聞こえ方の違いから音源位置を推定する枠組みを確立した（図 4）。事前に音源地図を準備しておくことで、環境音をフィンガープリントとしてセンサー位置を推定する。提案法では、NMF を活用することで、単純な回帰より雑音耐性を改善した。実際にオフィス環境で収録した環境音を用いて、提案法の有効性を確認し、国際会議 IEEE/SICE SII 2023 にて発表予定である。

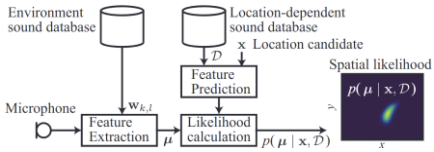


**図 5 位置推定法の概要**

Satoki Ogiso, Yoshiaki Bando, Takeshi Kurata, Takashi Okuma. Infrastructure-less Localization from Indoor Environmental Sounds Based on Spectral Decomposition and Spatial Likelihood Model. IEEE/SICE International Symposium on System Integration. 2023, in print.

**研究テーマ B「深層生成音源モデルによる精緻化」**

**当初計画では事前学習した深層生成音源モデルを用いる予定であったが、より容易に構築可能な教師なしで学習できる枠組みの確立に成功した。**深層生成音源モデルは、従来の NMF を用いた音源モデルと比較して、より精緻な音源表現を実現できるが、その学習には事前に収集した音源信号のデータセットを要する課題があった。本手法は、多チャネル混合音を、各音源信号の時不変空間相関行列と時変パワースペクトル密度に分解する、深層フルランク空間相関分析（Neural FCA）と呼ばれる。本枠組みは、図 5 のように変分オートエンコーダ型のアーキテクチャを取り、デコーダ側に音の物理的伝播過程を埋め込むことにより、教師なしで音源信号の深層生

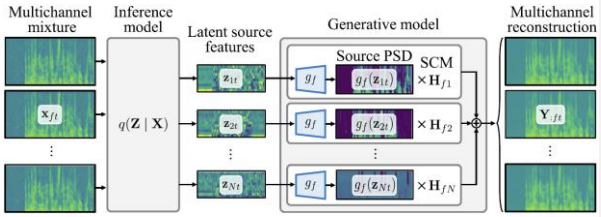


**図 5 深層フルランク相関分析の概要**

成モデルを学習できる。Neural FCA により、これまで教師あり事前学習が必要だった深層生成モデルを、多チャネル混合音のみから教師なし学習できるようになったことを示した。音声混合音の分離タスク（Spatialized WSJ0-2mix）での定量評価を行ったところ、従来の教師なし深層分離法やブラインド音源分離法を凌駕し、最新の教師あり多チャネル音源分離法の一つである多チャネル変分オートエンコーダ（MVAE）法と同程度の性能を達成した。本研究成果は、信号処理分野のトップ論文誌の一つである IEEE SP Letters にて誌上発表［1］した。

［発展的成果］**開発した Neural FCA を、複数人の自由に会話を書き起こす遠隔音声認識のフロントエンドとして応用した。**混合音から個別の音声を抽出する音声強調や音源分離は、雑音や他の話者の音声が混入する遠隔音声認識のフロントエンドとして不可欠である。スマートスピーカに代表されるように、単一話者の遠隔音声認識は高い性能を達成しているが、複数人が参加する会話の音声認識は未だ多くの課題が残されている。従来、複数話者の混合音から個別の音声を抽出するには、音源やマイクの空間的配置情報を殆ど必要としないブラインド音源分離(BSS)が活用されてきた。しかし従来の BSS の多くは、線形の生成モデルに基づくため性能に限界があった。そこで本研究では、非線形生成モデルに基づく Neural FCA を用いた遠隔音声認識のフロントエンドを構築した（図 6）。多くの BSS と同じく Neural FCA は、混合音に含まれる音源数が既知かつ固定と仮定しており、遠隔音声認識には適さない。本研究では、実用性を重視し、他の手法で得た発話区間情報を生成モデルに導入し、音源数の変動に対応した弱教師あり学習へ拡張した。提案法は、ホームパーティでの会話を収録した CHiME-6 データセットを用いて評価し、CHiME-6 の公式ベースラインであるガイド付き音源分離法（GSS）を上回る性能を達成した。本研究成果は、主導的に指導したインターン生により情報処理学会 全国大会で発表［7］し、**学生奨励賞を受賞**した。またより性能改善した内容を音声処理のトップ国際会議の 1 つである INTERSPEECH 2022 にて発表[3]した。

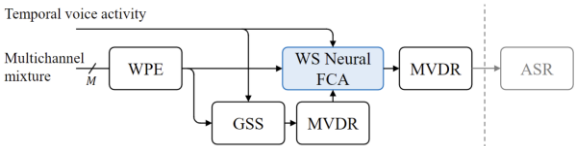


**図 6　深層フルランク空間相関分析法を用いた遠隔音声認識システムの概要**

**研究テーマ C「オンライン推論と能動センシング」**

**本研究課題で確立した音環境マッピングの枠組みを GPGPU 処理として実装し、オンライン推論アルゴリズム化の検討を行った。**AI 橋渡しクラウド（ABCI）上の GPU NVIDIA V100 を用いた実験で、混合音のデータ長より短い時間で推論できることを確認しており、本手法はリアルタイムに動作できる程度の計算量であった。今後は、クラウド計算機と連携したエッジ-クラウド型のリアルタイム地図構築システム・能動センシングを目指す。

## 3.　今後の展開

本研究課題の成果により、雑踏音環境を分析するロボット聴覚の実現が近づいた一方、未だ日常生活のあらゆる環境で動作する枠組みを実現するには課題も残る。具体的には、より頑健な音響マッピングを目指した NMF とクラスタリングの同時最適化と、Neural FCA のより難しい実データへの適用・拡張を進める。本研究課題の成果である音響マッピング法は、観測信号を NMF で

一挙に分析する点で従来の枠組みより一貫した処理を行えているが、いまだ NMF 部分とクラスタリング部分が独立しており、性能劣化の要因と考えられる。今後は、これらを統一しつつ局所解の少ない安定した枠組みの確立を目指す。また、Neural FCA は一定の環境下では既存手法を凌駕する性能を達成しているので、その適用可能範囲を広げる研究を進める。特に、音源数の同時推定・移動音源の定位および分離を同時に学習できる枠組みの確立を目指す。これらの研究は今後 5 年以内で実現することを目指しながら、発展成果の遠隔音声認識のように、要素技術ごとに逐次応用問題での有効性を実証していく。

## 4. 自己評価

**研究目的の達成状況**

当初計画の NMF を用いた雑踏音環境を統一的に扱うマッピングの枠組みは確立することができた。また、音響マッピングでの成果に比べ、音源定位や分離における理論的・実装的な発展成果が多くなった。これらは、音響マッピング自身との統合は期間内に行えなかったが、今後の雑踏音環境理解における基盤になると考えている。特に Neural FCA は、高い性能を得られる上に数値的にも安定しており、新しい高性能かつ頑健な枠組みを確立できたことは大きな成果であった。以上より、研究目標をある程度達成できたと言えるだろう。

**研究の進め方（研究実施体制及び研究費執行状況）**

代表者主導で研究を集中して遂行することができた。また領域会議を通して知り合った研究者と情報交換を進めることで、他分野における深層ベイズ学習の動向を把握するなど、効率的に研究を進めることができた。また、本研究課題をきっかけとして、マーカレス位置推定手法への応用など、他の研究者との連携にも繋がった。研究費執行については、新型コロナウィルス感染症の蔓延により旅費は最小限にとどめ、計算資源やロボット部品の調達に効果的に利用できた。

**研究成果の科学技術及び社会・経済への波及効果**

得られた研究成果の多くは、信号処理分野の主要国際会議や国際誌で発表できたほか、国内発表でも受賞するなど、分野内で高い評価を得ていると考えている。今後は、得られた成果の実証実験を通じて、産業応用など橋渡しを進めていきたい。

## 5. 主な研究成果リスト

### （1）代表的な論文（原著論文）発表

研究期間累積件数：2件

| |
|---|
| 1. Yoshiaki Bando, Kouhei Sekiguchi, Yoshiki Masuyama, Aditya Arie Nugraha, Mathieu Fontaine, Kazuyoshi Yoshii. **Neural Full-Rank Spatial Covariance Analysis for Blind Source Separation.** IEEE Signal Processing Letters. 2021, 28, 1670–1674 |
| 概要 多チャネル混合音のみから深層生成音源モデルと音源分離を教師なし学習できる、新しいブラインド音源分離の枠組みを確立した。本枠組みにより、教師データの収集が困難な雑踏環境音を分析するための深層生成モデルの基盤を整備できた。音声信号を混合した数値実験により、提案法は従来のブラインド音源分離を上回る性能を有することを確認し、教師あり学習した深層生成モデルに基づく音源分離法と同程度の性能を有することを示した。 |
| 2. Yoshiaki Bando, Yoshiki Masuyama, Yoko Sasaki, Masaki Onishi. **Robust Auditory** |

| |
|---|
| **Functions Based on Probabilistic Integration of MUSIC and CGMM**. IEEE Access. 2021, 9, 38718-38730. |
| 概要　音源定位と分離を単一の枠組みで一挙に推論する新しいロボット聴覚の枠組みを確立した。従来のカスケード型の欠点であったハイパーパラメータに対する敏感さと、同時最適化型における計算量の問題を解決できた。ノート型コンピュータに搭載された GPU でも実時間で動作することを確認したうえ、実雑踏雑音が重畳された音声信号の混合音において提案法の有効性を示した。 |

**（2）特許出願**

**該当なし**

**（3）その他の成果（主要な学会発表、受賞、著作物、プレスリリース等）**

3. Yoshiaki Bando, Takahiro Aizawa, Katsutoshi Itoyama, Kazuhiro Nakadai. **Weakly-Supervised Neural Full-Rank Spatial Covariance Analysis for a Front-End System of Distant Speech Recognition**. INTERSPEECH. 2022, 3824-3828
4. Yoshiaki Bando, Kohei Sekiguchi, Kazuyoshi Yoshii. **Gamma Process FastMNMF for Separating an Unknown Number of Sound Sources**. European Signal Processing Conference (EUSIPCO). 2021, 291-295
5. 坂東宜昭, 升山義紀, 佐々木洋子, 大西正輝. **雑踏環境における音源地図の生成. 人工知能学会 AI チャレンジ研究会**. 2021, 43–46. **研究会優秀賞受賞**
6. 坂東宜昭, 関口航平, Aditya Arie Nugraha, Mathieu Fontaine, 吉井和佳. **深層フルランク空間相関分析に基づくブラインド音源分離**. 日本音響学会研究発表会. 2021, 1–2.
7. 合澤隆拓, 坂東宜昭, 糸山克寿, 西田健次, 中臺一博. **深層フルランク空間相関分析に基づく遠隔音声認識のフロントエンド**. 情報処理学会 全国大会. 2022, 1R-02. **学生奨励賞受賞**