

# スマートグラスを用いた音環境理解支援

## Assistance of Sound Scene Understanding with Smart Glasses

吉井 和佳<sup>1,2\*</sup>  
Kazuyoshi Yoshii<sup>1,2</sup>

<sup>1</sup> 京都大学 大学院情報学研究科 <sup>2</sup> 理化学研究所 革新知能統合研究センター (AIP)  
<sup>1</sup>Graduate School of Informatics, Kyoto University <sup>2</sup>AIP, RIKEN

**Abstract:** 本稿では、実環境下での音声コミュニケーションを支援するため、特定の話者の音声をリアルタイムで強調・認識・翻訳・拡張現実 (AR) 表示を行うスマートグラスの開発について紹介する。スマートグラスに搭載されたマイクアレイから得られる音響信号は、未知の環境における多数の音声・雑音・残響が重畳したものであり、音源の移動に加えて、ユーザ自身の移動や頭部の動きに影響を受けるため、研究室環境のベンチマークとは本質的に難しさが異なる。そのうえで、頑健性と即応性を備えた実用システムを開発する技術的なチャレンジについて解説する。

### 1 はじめに

近年、スマートフォンやスマートスピーカーへの指示、自動字幕・議事録作成など、日常生活中で音声認識が活用される場面が増えつつある。最新の深層学習技術に基づく音声認識手法では、接話型マイクで集音されたクリーンな発話に対する認識精度は90%を超えており、実用化の急速な進展を支えている。一方、雑音・残響のある実環境における遠隔発話認識となると、認識精度が30%以下に低下することも珍しくない。そのため、遠隔音声認識のフロントエンドとして、マイクアレイで集音された多チャンネル音響信号に対する音声強調や残響除去の研究が活発である。

我々の研究グループでは、汎用的な基盤技術として、多チャンネル音源分離・音声強調技術の改善に継続的に取り組んできた(第2章)。具体的には、高速多チャンネル非負値行列因子分解 (FastMNMF) [1,2] をコア技術として、深層学習の導入 [3-6]、残響除去の統合 [7,8]、音源数の自動推定 [9]、優ガウス性音源への対応 [10] を考案した。一連の技術は、遠隔音声認識のフロントエンドとして有用であり、アンドロイドロボットによる音声対話への応用を想定している。

将来有望な別の応用として、健常者・聴覚障害者双方に有用な、スマートグラスを用いた音声コミュニケーション支援に取り組んでいる(第3章)。具体的には、マイクアレイ・カメラを用いて視聴覚情報を取得し、特定話者の音声を強調したうえで、リアルタイムで音声認識・翻訳結果を拡張現実 (AR) 表示する [11-13]。しかし、センサー類はウェアラブルデバイスに搭載され

ているため、ユーザの移動や頭部の動きに対する頑健性に加えて、リアルタイム性(低遅延)も求められ、研究室環境における伝統的なオフライン型のベンチマークとは本質的に難しさが異なる。

本稿では、多チャンネル音源分離における我々の貢献を体系的に解説し、スマートグラスへの応用に関する先駆的取り組みを通じて今後を展望する(第4章)。

### 2 多チャンネル音源分離・残響除去

我々の研究グループは、汎用ブラインド音源分離 (BSS) の洗練化に継続的に取り組み、理論面で一定の完成をみたと言ってよい。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、多チャンネル非負値行列因子分解 (MNMF) [14] は、フルランク空間共分散行列に基づく空間モデルと、非負値行列因子分解 (NMF) に基づく音源モデルを統合した汎用 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) [15] では、空間共分散行列をランク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 BSS 手法である高速多チャンネル非負値行列因子分解 (FastMNMF) [1,2] では、空間共分散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た。さらに、FastMNMF 各部の本質的な拡張を行い、確率的な枠組みへの深層学習の導入を実現した(図1)。

\*連絡先：京都大学 大学院情報学研究科 知能情報学専攻  
〒606-8501 京都府京都市左京区吉田本町  
E-mail: yoshii@i.kyoto-u.ac.jp

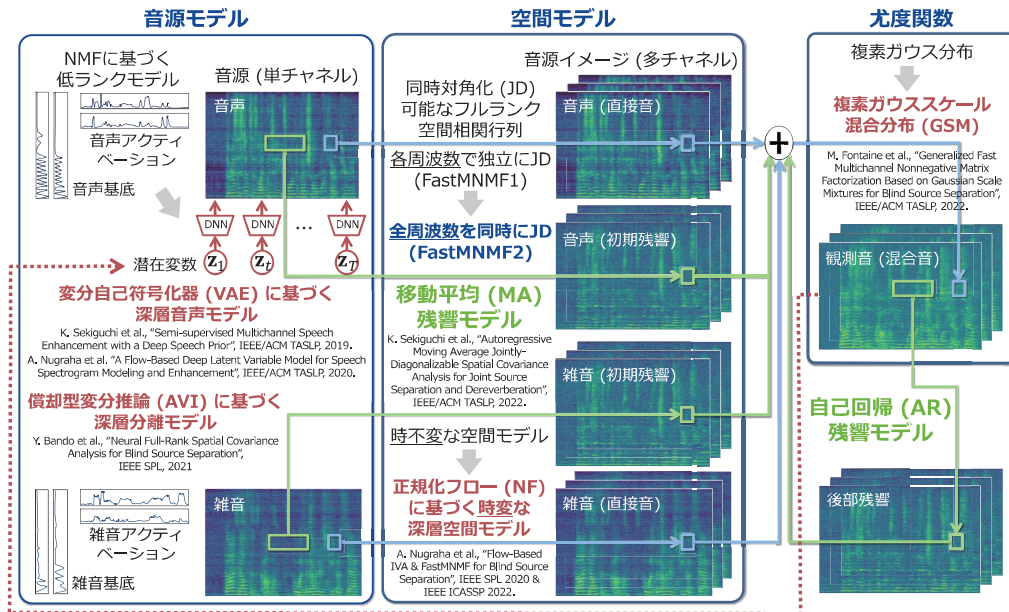


図 1: FastMNMF を中心とした多チャンネル音源分離技術体系

## 2.1 フルランク空間共分散分析の系譜

まず、フルランク空間共分散行列に基づく BSS 手法の歴史を整理しておく。一連の研究の発端となったのは、Duong ら [16] が考案した、音源の時間周波数平面のパワースペクトル密度を表現する音源モデルと、マイク間の相関（位相差）を表現する空間モデルからなるフルランク空間共分散分析 (FCA) である。FCA は統一的な確率モデルの最尤推定に対応しており、数学的に見通しが良く、汎用性が高い一方で、音源・空間モデルの自由度が高すぎるゆえに、物理的に妥当なパラメータを推定することは容易ではなかった。

澤田ら [14] は、低ランク音源モデルに基づく多チャンネル非負値行列因子分解 (MNMF) を考案した。MNMF では、音源のパワースペクトル密度は非負値行列因子分解 (NMF) で表現されており、パラメータの最尤推定に収束保証付きの Majorization-Minimization (MM) アルゴリズムを用いた点で画期的である。しかし、コスト関数（負の対数尤度）の上限関数を導出する過程に誤りがあり、フルランク空間共分散行列の更新則はリッカ方程式の解として与えられていた。我々はその後、NMF の本質的なテンソル拡張である相関テンソル分解 (CTF) [17] およびその高速版 (FastCTF) [18] の研究において、上限関数の導出を正しく行ったうえで、上記リッカ方程式の解が、その係数である二つの正定値行列の幾何平均で与えられることを報告した。この閉経式での更新則は、後に続く研究で標準的に用いられるようになっていく。MNMF は優れた分離性能を達成しようが、空間モデルの自由度は依然として高く、局所解へ陥りやすい問題が残されていた。

近年、奇しくも同時期に、我々の研究グループ [1,2] と伊藤らの研究グループ [19,20] から独立に、音源の空間共分散行列を同時対角化可能なものに制限した高速多チャンネル非負値行列因子分解 (FastMNMF) が提案されている（名称・確率モデルともに同一）。FastMNMF は最先端の BSS 手法であり、空間モデルの自由度を制限しつつもフルランク性を維持することにより、独立低ランク行列分析 (ILRMA) [15] に匹敵する計算効率と、MNMF より優れた分離性能を兼ね備えている。伊藤らは当時すでに、同時対角化制約を導入した FCA である FastFCA を考案していた [21,22]。同時期に、我々は、CTF の計算量を削減するための同時対角化制約の導入を考案していた [17,18]。そのため、同時対角化制約の MNMF への導入は自然な流れでもあった。

パラメータの最尤推定には共通して MM アルゴリズムが利用されているが、その実装は異なっていた。空間共分散行列を対角化する行列を更新するため、我々は、独立ベクトル分析 (IVA) [23] を始め、ILRMA [15] や FastCTF [18] と同様に、反復射影法 (IP) が適用できることを見出したのに対して [1]、伊藤らは不動点反復法を利用して [19]。収束性が保証され、収束も速いことから、現在では IP の利用が標準的である。

我々は最終的に、FastMNMF1 [2,19,20] より性能に優れた FastMNMF2 [2] を考案した。FastMNMF1 では、周波数ごとに空間共分散行列を同時対角化可能なものに制限するのに対して、FastMNMF2 では、すべての周波数・音源に関する空間共分散行列を一挙に制限する点で異なる。これにより、各音源の方向情報がすべての周波数間で共有され、周波数間のパーミュテーション問題の解決に寄与することが期待できる。

## 2.2 深層音源モデルの導入 [3]

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 [24] 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 FastMNMF-DP を提案した。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も可能な枠組みとなっている。

## 2.3 深層空間モデルの導入 [4, 6]

独立成分分析 (ICA) [25] や独立ベクトル分析 (IVA) [26] 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 [4]。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 である NF-FastMNMF の導出に成功した。

## 2.4 残響モデルの導入 [7, 8]

実環境下において、音声強調と残響除去はいずれも重要であり、相互依存関係のある両タスクを一举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 ARMA-FastMNMF 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。

## 2.5 裾の重い尤度関数の導入 [10]

通常、混合音スペクトルに対する尤度関数として、時変な空間共分散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) を用いた GSM-FastMNMF 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 [27], 裾が重い一般化ガウス分布 (GG) [28],  $\alpha$ -対称安定分布 [29] 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。

## 2.6 深層分離モデルの導入 [5]

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、実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かった。そこで、FastMNMF の基盤となる、音源信号から多チャン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。

## 2.7 ノンパラメトリックベイズの導入 [9]

理論的には、MNMF は音源数がマイク数よりも多い劣決定条件でも適用できる一方で、FastMNMF は高々マイク数と同じ音源数まで扱うことができる。実用上は、決定条件を前提とした BSS 手法である IVA や ILRMA などと同様に、音源数をマイク数と同一に設定することも多いが、適切に音源数を指定することで、一つの音源が複数の音源として分割されることを防ぐことができる。そこで、理論上は無数個の音源を考えるガンマ過程に基づくノンパラメトリックベイズモデルを定式化し、観測データに合わせて必要な個数の音源を自動的に推定する手法を考案した。

## 3 スマートグラスを用いたリアルタイム音声強調・認識

我々のグループでは、音響信号処理技術のキラーアプリケーションとして、リアルタイム音声コミュニケーション支援のため、スマートグラスを用いたオンライン音声強調・認識に取り組んでいる (図 2)。この応用においては、事前学習が困難な未知の実環境にどのように適応するのかが重要な課題となる。また、ユーザの身体や頭部の動作に伴うマイクアレイの位置変化 (相対的音源方向の変化) も問題となる。これらは、研究室環境下での従来のオフラインベンチマークでは存在しなかった極めて困難な課題である。

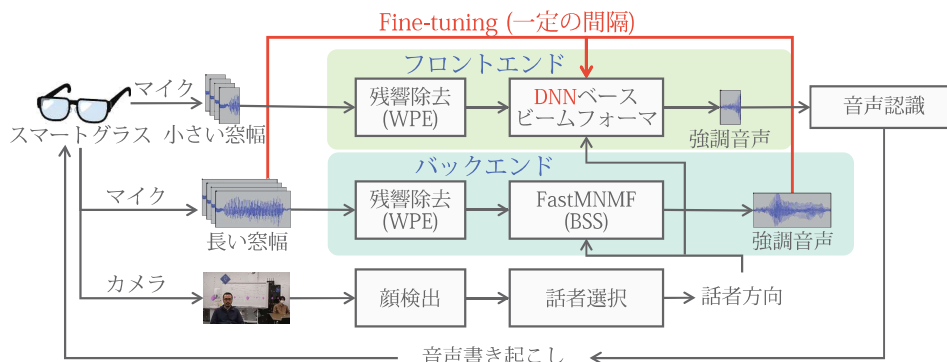


図2: スマートグラスを用いたリアルタイム音声強調・認識とオンライン適応

スマートグラスを用いた音声強調において、FastMNMFは、適応的なパラメータ推定（教師なし学習）によって未知の環境でも頑健に動作する点で有望であるが、そ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。一方、多チャンネル音声強調の主流である最小分散 (MVDR) ビームフォーミングは、極めて高速に動作する一方で、いかに適切な音声と雑音の空間共分散行列を与えるが性能を左右する。その典型的な解決法の一つは、深層ニューラルネットワーク (DNN) を用いて、観測音（混合音）のスペクトログラムに対して音声マスクを推定するものである。この方式では、音声と雑音の空間共分散行列は適応的に計算されるが、DNNは事前に教師あり学習する必要があるため、学習時と運用時の環境のミスマッチによって性能が大幅に低下する問題は避けられない。

このような性質に鑑み、我々は、FastMNMFをバックエンド、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ている。具体的には、バックエンドでは、十分な反復推定が行える一定の間隔で、大きなサイズの時間窓に対してFastMNMFを行うことで、音声と雑音の空間共分散行列を推定することができる。これらに基づく最小分散 (MVDR) ビームフォーミングを行えば音声強調を行うことができる [11]。DNNベースのビームフォーミングでは、観測音（混合音）とFastMNMFで分離された音声とのペアデータを用いて音声マスク推定用のDNNをFine-tuningすることで環境へのオンライン適応が実現できる [12]。

我々はさらに、マスク推定用のDNNと音声認識用のDNNとを結合し、運用時に全体をFine-tuningすることにも取り組んでいる [13]。この方式は、音声強調と音声認識との同時最適化に関する研究に着想を得ているが、運用時には音声認識の正解データが存在しない点が本質的に異なる。そのため、認識結果から、出力の長さが十分にあり、音声認識モデルおよび大規模言語モデルが与えるスコアがいずれも大きいものを、擬似的な正解データとして選択する方法を考案している。

## 4 おわりに

本稿では、実環境下での音声コミュニケーションを支援するため、特定の話者の音声をリアルタイムで強調・認識・翻訳・拡張現実 (AR) 表示を行うスマートグラスの開発について紹介した。音響信号処理分野では、深層学習の導入により、音源分離や音声強調などの基盤技術で目覚ましい進展があったが、キラーアプリケーションが存在しなかった。近い将来、スマートグラスはスマートコンタクトレンズまで小型化されることは間違いなく、我々は、スマートフォンに変わる、人の視聴覚と密接にリンクしたウェアラブルデバイスへの応用こそがキラーアプリケーションになると確信している。実際、MetaやGoogleなどから同様の目的をもったスマートグラスの開発プロジェクトがつい最近相次いで発表されており、この研究課題はにわかに注目を浴びつつある。我々は昨年からいち早く取り組み、すでにいくつかの成果発表を行ってきているが、さらに研究を加速させていきたい。

## 謝辞

本稿で紹介した一連の研究は、京都大学 大学院情報科学研究科 知能情報学専攻 音声メディア分野および理化学研究所 革新知能統合研究センター (AIP) 音響情景理解チームにて実施したものである。中でも、關口航平氏 (当時理研 AIP/京大大学院生, 現在 DENSO), Aditya Arie Nugraha 氏 (理研 AIP), Mathieu Fontaine 氏 (当時理研 AIP, 現在 Télécom Paris), 坂東宜昭氏 (産総研), Yicheng Du 氏 (当時京大大学院生) らの貢献は特に大きく、彼らの協力なくしては成し得なかった。その他、多くの研究者・学生との議論を通じて、数多くの有益な助言を得た。この場を借りて最大限の謝意を表明したい。

また、本研究の一部は、JSPS 科研費 No. 20H00602, 20K21813, 19H04137 および JST さきがけ No. JP-MJPR20CB の支援を受けた。

## 参考文献

- [1] Kouhei Sekiguchi, Aditya A. Nugraha, Yoshiaki Bando, and Kazuyoshi Yoshii. Fast multichannel source separation based on jointly diagonalizable spatial covariance matrices. In *European Signal Processing Conf. (EUSIPCO)*, pp. 1–5, 2019.
- [2] Kouhei Sekiguchi, Yoshiaki Bando, Aditya A. Nugraha, Kazuyoshi Yoshii, and Tatsuya Kawahara. Fast multichannel nonnegative matrix factorization with directivity-aware jointly-diagonalizable spatial covariance matrices for blind source separation. *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, Vol. 28, pp. 2610–2625, 2020.
- [3] Kouhei Sekiguchi, Yoshiaki Bando, Aditya A. Nugraha, Kazuyoshi Yoshii, and Tatsuya Kawahara. Semi-supervised multichannel speech enhancement with a deep speech prior. *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, Vol. 27, No. 12, pp. 2197–2212, 2019.
- [4] Aditya A. Nugraha, Kouhei Sekiguchi, Mathieu Fontaine, Yoshiaki Bando, and Kazuyoshi Yoshii. Flow-based independent vector analysis for blind source separation. *IEEE Signal Processing Letters*, Vol. 27, pp. 2173–2177, 2020.
- [5] Yoshiaki Bando, Kouhei Sekiguchi, Yoshiki Masuyama, Aditya A. Nugraha, Mathieu Fontaine, and Kazuyoshi Yoshii. Neural full-rank spatial covariance analysis for blind source separation. *IEEE Signal Processing Letters*, Vol. 28, pp. 1670–1674, 2021.
- [6] Aditya A. Nugraha, Kouhei Sekiguchi, Mathieu Fontaine, Yoshiaki Bando, and Kazuyoshi Yoshii. Flow-based fast multichannel nonnegative matrix factorization for blind source separation. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 501–505, 2022.
- [7] Kouhei Sekiguchi, Yoshiaki Bando, Aditya A. Nugraha, Mathieu Fontaine, and Kazuyoshi Yoshii. Autoregressive fast multichannel nonnegative matrix factorization for joint blind source separation and dereverberation. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 511–515, 2021.
- [8] Kouhei Sekiguchi, Yoshiaki Bando, Aditya A. Nugraha, Mathieu Fontaine, Kazuyoshi Yoshii, and Tatsuya Kawahara. Autoregressive moving average jointly-diagonalizable spatial covariance analysis for joint source separation and dereverberation. *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, Vol. 30, pp. 2368–2382, 2022.
- [9] Yoshiaki Bando, Kouhei Sekiguchi, and Kazuyoshi Yoshii. Gamma process FastMNMF for separating an unknown number of sound sources. In *European Signal Processing Conf. (EUSIPCO)*, pp. 291–295, 2021.
- [10] Mathieu Fontaine, Kouhei Sekiguchi, Aditya A. Nugraha, Yoshiaki Bando, and Kazuyoshi Yoshii. Generalized fast multichannel nonnegative matrix factorization based on gaussian scale mixtures for blind source separation. *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, Vol. 30, pp. 1734–1748, 2022.
- [11] Aditya A. Nugraha, Kouhei Sekiguchi, Mathieu Fontaine, Yoshiaki Bando, and Kazuyoshi Yoshii. DNN-free low-latency adaptive speech enhancement based on frame-online beamforming powered by block-online FastMNMF. In *IEEE Int. Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 1–5, 2022.
- [12] Kouhei Sekiguchi, Aditya A. Nugraha, Yicheng Du, Yoshiaki Bando, Mathieu Fontaine, and Kazuyoshi Yoshii. Direction-aware adaptive online neural speech enhancement with an augmented reality headset in real noisy conversational environments. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pp. 1–8, 2022.
- [13] Yicheng Du, Aditya A. Nugraha, Kouhei Sekiguchi, Yoshiaki Bando, Mathieu Fontaine, and Kazuyoshi Yoshii. Direction-aware joint adaptation of neural speech enhancement and recognition in real multiparty conversational environments. In *Annual Conf. of Int. Speech Communication Association (Interspeech)*, pp. 2918–2922, 2022.
- [14] Hiroshi Sawada, Hirokazu Kameoka, Shoko Araki, and Naonori Ueda. Multichannel extensions of non-negative matrix factorization with complex-valued data. *IEEE Trans. on Audio,*

- Speech, and Language Processing*, Vol. 21, No. 5, pp. 971–982, 2013.
- [15] Daichi Kitamura, Nobutaka Ono, Hiroshi Sawada, Hirokazu Kameoka, and Hiroshi Saruwatari. Determined blind source separation unifying independent vector analysis and non-negative matrix factorization. *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, Vol. 24, No. 9, pp. 1626–1641, 2016.
- [16] Ngoc Q. K. Duong, Emmanuel Vincent, and Remi Gribonval. Under-determined reverberant audio source separation using a full-rank spatial covariance model. *IEEE Trans. on Audio, Speech, and Language Processing*, Vol. 18, No. 7, pp. 1830–1840, 2010.
- [17] Kazuyoshi Yoshii. Correlated tensor factorization for audio source separation. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 731–735, 2018.
- [18] Kazuyoshi Yoshii, Koichi Kitamura, Yoshiaki Bando, Eita Nakamura, and Tatsuya Kawahara. Independent low-rank tensor analysis for audio source separation. In *European Signal Processing Conf. (EUSIPCO)*, pp. 1657–1661, 2018.
- [19] Nobutaka Ito and Tomohiro Nakatani. FastM-NMF: Joint diagonalization based accelerated algorithms for multichannel nonnegative matrix factorization. In *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 371–375, 2019.
- [20] Nobutaka Ito, Rintaro Ikeshita, Hiroshi Sawada, and Tomohiro Nakatani. A joint diagonalization based efficient approach to underdetermined blind audio source separation using the multichannel wiener filter. *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, Vol. 29, pp. 1950–1965, 2021.
- [21] Nobutaka Ito and Tomohiro Nakatani. FastFCA-AS: Joint diagonalization based acceleration of full-rank spatial covariance analysis for separating any number of sources. In *IEEE Int. Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 151–155, 2018.
- [22] Nobutaka Ito and Tomohiro Nakatani. Multiplicative updates and joint diagonalization based acceleration for under-determined bss using a full-rank spatial covariance model. In *IEEE Global Conf. on Signal and Information Processing (GlobalSIP)*, pp. 231–235, 2018.
- [23] Nobutaka Ono. Stable and fast update rules for independent vector analysis based on auxiliary function technique. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 189–192, 2011.
- [24] Aditya A. Nugraha, Kouhei Sekiguchi, and Kazuyoshi Yoshii. A flow-based deep latent variable model for speech spectrogram modeling and enhancement. *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, Vol. 28, pp. 1104–1117, 2020.
- [25] Paris Smaragdis. Blind separation of convolved mixtures in the frequency domain. *Neurocomputing*, Vol. 22, No. 1, pp. 21–34, 1998.
- [26] Taesu Kim, Torbjørn Eltoft, and Te-Won Lee. Independent vector analysis: An extension of ICA to multivariate components. In *Int. Conf. on Independent Component Analysis and Signal Separation (ICA)*, pp. 165–172, 2006.
- [27] Keigo Kamo, Yuki Kubo, Norihiro Takamune, Daichi Kitamura, Hiroshi Saruwatari, Yu Takahashi, and Kazunobu Kondo. Joint-diagonalizability-constrained multichannel nonnegative matrix factorization based on multivariate complex Student’s  $t$ -distribution. In *Asia-Pacific Signal and Information Processing Association Annual Summit and Conf. (APSIPA ASC)*, pp. 869–874, 2020.
- [28] Keigo Kamo, Yuki Kubo, Norihiro Takamune, Daichi Kitamura, Hiroshi Saruwatari, Yu Takahashi, and Kazunobu Kondo. Joint-diagonalizability-constrained multichannel nonnegative matrix factorization based on multivariate complex sub-Gaussian distribution. In *European Signal Processing Conf. (EUSIPCO)*, pp. 890–894, 2021.
- [29] Mathieu Fontaine, Kouhei Sekiguchi, Aditya A. Nugraha, and Kazuyoshi Yoshii. Unsupervised robust speech enhancement based on alpha-stable fast multichannel nonnegative matrix factorization. In *Annual Conf. of Int. Speech Communication Association (Interspeech)*, pp. 4541–4545, 2020.