社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

［招待講演］独立低ランク行列分析に基づく音源分離とその発展

北村　大地†

† 香川高等専門学校　〒761–8058 香川県高松市勅使町 355
E-mail: †kitamura-d@t.kagawa-nct.ac.jp

あらまし　複数の音源が混合した観測信号から，混合前の音源信号を推定する技術を音源分離と呼ぶ．特に，音源やマイクロフォンの空間的配置が未知の条件の音源分離はブラインド音源分離と呼ばれ，その重要性から 20 年以上研究されてきた歴史を持つ．本講演では，ブラインド音源分離の新しいアルゴリズムの一つである独立低ランク行列分析に焦点を当て，その基本原理から近年の発展までの簡単な説明と紹介を行う．
キーワード　ブラインド音源分離，独立成分分析，非負値行列因子分解，深層学習，スペクトログラム無矛盾性

[Invited Talk] Audio Source Separation Based on Independent Low-Rank Matrix Analysis and Its Extensions

Daichi KITAMURA†

† National Institute of Technology, Kagawa College　355 Chokushi, Takamatsu, Kagawa, 761–8058 Japan
E-mail: †kitamura-d@t.kagawa-nct.ac.jp

Abstract　Audio source separation is a technique for separating individual audio sources from an observed mixture signal. In particular, blind source separation (BSS) does not require any information about locations of audio sources and microphones. Because of its importance, BSS has been investigated for more than 20 years. In this talk, one of the new BSS algorithms called independent low-rank matrix analysis (ILRMA) is explained. Also, some recent extensions of ILRMA are introduced.
Key words　Blind source separation, independent component analysis, nonnegative matrix factorization, deep learning, spectrogram consistency

1. はじめに

ブラインド音源分離（blind source separation: BSS）とは，音源位置や混合系が未知の条件で観測された信号のみから混合前の音源信号を推定する音響信号処理技術である．優決定条件（音源数 ≤ 観測チャネル数）における BSS では，1994 年に提案された独立成分分析（independent component analysis: ICA）[1] に基づく手法が歴史的に主流である．ICA は，音源間の統計的独立性と音源信号の非ガウスな生成モデルを仮定することで，混合行列の逆系である分離行列を推定する理論である．一般的な音響信号の混合系は畳み込み混合となるが，この場合は観測信号に短時間 Fourier 変換（short-time Fourier transform: STFT）を適用することで周波数毎の瞬時混合に変換できる．これを用いて，複素スペクトログラムの各周波数ビンに対して ICA を適用し周波数毎の分離行列を推定する周波数領域 ICA（frequency-domain ICA: FDICA）[2] が 1998 年に提案された．しかしながら，ICA は分離信号の順序に任意性があるため，Fig. 1 に示すように，FDICA では周波数ビン毎に推定される分離信号の順番が不揃いとなるパーミュテーション問題が生じる．

FDICA で推定される分離信号の順序を全周波数にわたって適切に並び替えるパーミュテーション解決には，これまでに様々な手法が提案されている [3]〜[7]．近年では，深層ニューラルネットワーク（deep neural network: DNN）に基づくパーミュテーション解決法も検討されている [8]．歴史的には，FDICA に基づく周波数毎の BSS の後にパーミュテーション解決法を適用する手法から，パーミュテーション問題を回避しつつ周波数毎の BSS を行う統一的な手法へと発展する．まず，2006 年に独立ベクトル分析（independent vector analysis: IVA）[9]〜[11] が提案された．IVA 各音源の周波数成分を一つにまとめた周波数ベクトルの生成モデル（多変量確率分布）を仮定し，Fig. 2 (a) に示すように「同一音源の周波数成分間は共起性（高次相関 [11]）を持つ」ことを仮定することで，パーミュテーション問題を可能な限り回避しながら周波数毎の分離行列を推定する．2010 年と 2011 年には，補助関数法 [12] に基づく高速・安定な ICA [13] 及び IVA（auxiliary-function-based IVA: AuxIVA）[14] が提案され，IVA は非常に実用的な BSS アルゴリズムとなった．

音声信号の BSS では IVA が効果的である反面，例えば複数の楽器音信号が混合した音楽信号の BSS では，IVA はしばし

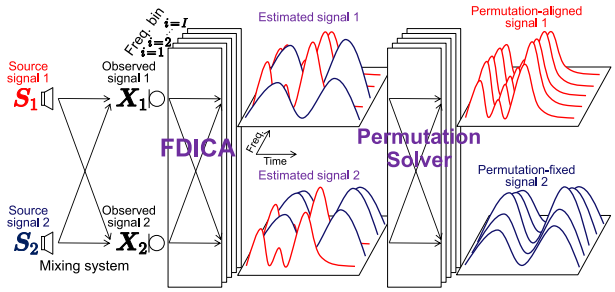

Fig. 1　Permutation problem in frequency-domain BSS based on FDICA.

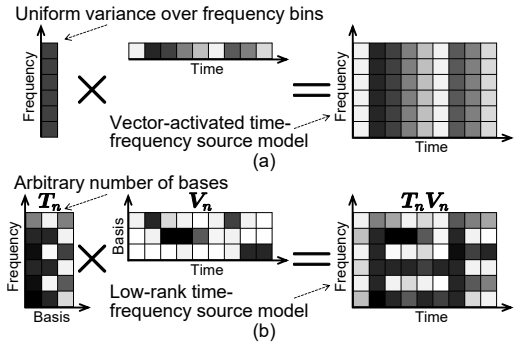

Fig. 2　Structure of source models in (a) IVA and (b) ILRMA, where grayscale depicts intensity of variance.

ば分離に失敗してしまう．これは，IVA が仮定する同一音源の周波数成分間の共起性（このような一つの音源に対する時間周波数領域での仮定を以後「音源モデル」と呼ぶ）が音楽信号に対しては不適切であることが原因として挙げられる．即ち，音楽信号では複数の異なる楽器音が同一周波数においても共起する調和成分を多分に含んでいるため，複数音源間で共起性が生じ，別の音源の成分を誤って同一音源とみなしてしまう現象が生じる．そこで，IVA の仮定する音源モデルを拡張し，多様な音源に適合できる柔軟なモデルに改良した BSS として，2016 年に独立低ランク行列分析（independent low-rank matrix analysis: ILRMA）[15]〜[18] が提案された(注1)．ILRMA では，Fig. 2 (b) に示すように「同一音源の時間周波数パワー構造は非負低ランク行列で表現できる」ことを仮定している．これは，1 本の基底ベクトルから成る IVA の音源モデルを，高々数本の基底ベクトルから成る音源モデルに一般化したものと解釈できる．また，ILRMA の音源モデルにおける基底ベクトル（Fig. 2 (b) の例では 3 本）及びその係数ベクトルは，非負値行列因子分解（nonnegative matrix factorization: NMF）[19], [20] の推定理論が用いられており，ICA とは独立に研究されてきた NMF に基づく劣決定条件（音源数 > 観測チャネル数）の音源分離技術（例えば [21]〜[26] 等）との繋がりも新たに生まれた．これらの歴史的俯瞰については，文献 [27] に詳しいので参照されたい．

本稿では，2016 年に登場した BSS アルゴリズムである ILRMA に焦点を当て，原理の簡単な説明を述べる．また，その後の研究成果として，ILRMA を起源とした発展手法である DNN に基づく教師有り音源モデル拡張及びスペクトログラム無矛盾性を担保する制約付き ILRMA の二つを新たに紹介する．いずれも，Fig. 1 に示すパーミュテーション問題を高精度に回避しながら BSS を達成するための発展であり，そのためにより良い音源モデルの探求した手法である．

(注1)：ILRMA は文献 [15], [16] で rank-1 MNMF と呼ばれているが，後の文献で改名された．

2.　ILRMA に基づく BSS の原理

2.1　BSS で解くべき問題の定式化

音源数と観測チャネル数をそれぞれ N 及び M とし，各時間周波数における音源信号，観測信号，及び分離信号をそれぞれ

$$\boldsymbol{s}_{ij} = [s_{ij1}, s_{ij2}, \cdots, s_{ijn}, \cdots, s_{ijN}]^{\mathrm{T}} \in \mathbb{C}^N \tag{1}$$

$$\boldsymbol{x}_{ij} = [x_{ij1}, x_{ij2}, \cdots, x_{ijm}, \cdots, x_{ijM}]^{\mathrm{T}} \in \mathbb{C}^M \tag{2}$$

$$\boldsymbol{y}_{ij} = [y_{ij1}, y_{ij2}, \cdots, y_{ijn}, \cdots, y_{ijN}]^{\mathrm{T}} \in \mathbb{C}^N \tag{3}$$

と表す．ここで，$i = 1, 2, \cdots, I$，$j = 1, 2, \cdots, J$，$n = 1, 2, \cdots, N$，及び $m = 1, 2, \cdots, M$ はそれぞれ周波数ビン，時間フレーム，音源，及び観測チャネルのインデクスを示し，$\cdot^{\mathrm{T}}$ は転置を表す．また，各信号の複素スペクトログラム行列を $\boldsymbol{S}_n \in \mathbb{C}^{I\times J}$，$\boldsymbol{X}_m \in \mathbb{C}^{I\times J}$，及び $\boldsymbol{Y}_n \in \mathbb{C}^{I\times J}$ で表す．これらの行列の要素はそれぞれ s_{ijn}, x_{ijm}, 及び y_{ijn} に一致する．混合系が線形時不変であり，時間周波数領域での複素瞬時混合で表現できる場合，周波数毎の時不変な複素瞬時混合行列 $\boldsymbol{A}_i = [\boldsymbol{a}_{i1}\, \boldsymbol{a}_{i2}\, \cdots\, \boldsymbol{a}_{iN}] \in \mathbb{C}^{M\times N}$（$\boldsymbol{a}_{in} = [a_{in1}, a_{in2}, \cdots, a_{inM}]^{\mathrm{T}}$ は各音源のステアリングベクトル）が定義でき，観測信号を次式で表現できる．

$$\boldsymbol{x}_{ij} = \boldsymbol{A}_i \boldsymbol{s}_{ij} \tag{4}$$

この混合モデルは，時不変混合系の残響時間が STFT の窓長よりも十分短い場合に成立する．このとき，$M = N$ かつ $\boldsymbol{A}_i$ が正則であれば，分離ベクトル $\boldsymbol{w}_{in} = [w_{in1}, w_{in2}, \cdots, w_{inM}]^{\mathrm{T}}$ で構成される分離行列 $\boldsymbol{A}_i^{-1} \approx \boldsymbol{W}_i = [\boldsymbol{w}_{i1}\, \boldsymbol{w}_{i2}\, \cdots\, \boldsymbol{w}_{iN}]^{\mathrm{H}} \in \mathbb{C}^{N\times M}$ が存在し，分離信号は次式で与えられる．

$$\boldsymbol{y}_{ij} = \boldsymbol{W}_i \boldsymbol{x}_{ij} \tag{5}$$

ここで，$\cdot^{\mathrm{H}}$ はエルミート転置を示す．優決定条件 BSS では，式 (5) 中の分離行列 $\boldsymbol{W}_i$ を全周波数 $i = 1, 2, \cdots, I$ において推定することが最終的な目標となる．本稿では，以後常に決定的な系（$M = N$）を考える．

2.2　動　　機

前章で述べたように，IVA はパーミュテーション問題を回避する為に，Fig. 2 (a) に示すように，同一音源の周波数成分間が一様な強度の共起性を持つことを音源モデルとして仮定している．しかしながら，実際の音響信号では，基本周波数とその倍音は強い共起性を持つが，それ以外の成分や周波数差の大きい成分間の共起性は弱くなる等，共起性の度合いは時間周波数のグリッド (i, j) に依存して大きく変動する．このような理由から，IVA は特に楽器音信号等のように明確な時間周波数構造を持つ音響信号の分離には適していない．各音源の時間周波数領域での共変性をより詳細にモデル化できれば，高精度なパーミュテーション問題の回避や高精度な分離行列の推定につながると考えられる．

時間周波数領域での共変性は，板倉斎藤擬距離に基づく NMF（Itakura–Saito NMF: ISNMF）[28] として提案された統計的生成モデルによって効率的かつ詳細に表現可能である．この生成モ

デルは局所 Gauss モデル（local Gauss model: LGM）と呼ばれ，文献 [29], [30] に詳しい．ILRMA では，この LGM 及び ISNMF に基づく音源モデルを IVA に導入し，高精度な BSS を達成している．次節にてその詳細を述べる．

2.3 ILRMA で仮定する音源の生成モデル

ILRMA では，音源の複素スペクトログラムに対して次のような統計的仮定をおく．今，複素スペクトログラム $\boldsymbol{S}_n$ の各時間周波数要素 s_{ijn} が L 個の複素要素 $c_{ij1n}, c_{ij2n}, \cdots, c_{ijLn}$ の合成 $s_{ijn} = \sum_l c_{ijln}$ で与えられると考える．ここで，$l = 1, 2, \cdots, L$ は複素要素のインデクスである．さらに，複素要素 c_{ijln} の生成モデルを，次式に示す原点対称な複素ガウス分布と仮定する．

$$p(c_{ijln}; \rho_{ijln}) = \frac{1}{\pi \rho_{ijln}} \exp\left(-\frac{|c_{ijln}|^2}{\rho_{ijln}}\right) \tag{6}$$

ここで，$\rho_{ijln} > 0$ は分散である．式 (6) は原点対称であるため，確率値は複素要素 c_{ijln} の位相によらず振幅 $|c_{ijln}|$ あるいはパワー $|c_{ijln}|^2$ にのみ依存する．従って，分散はパワーの期待値 $\rho_{ijln} = \mathrm{E}[|c_{ijln}|^2]$ に対応する．合成成分 s_{ijn} の生成モデルは次式となる．

$$p(s_{ijn}; r_{ijn}) = \frac{1}{\pi r_{ijn}} \exp\left(-\frac{|s_{ijn}|^2}{r_{ijn}}\right) \tag{7}$$

ここで，$r_{ijn} > 0$ は分散であり，複素 Gauss 分布の再生性より $r_{ijn} = \sum_l \rho_{ijln}$ が成立する．即ち $p(s_{ijn})$ は，個々の複素要素 c_{ijln} の分散 ρ_{ijln} の合成 $\sum_l \rho_{ijln}$ を新たな分散とした原点対称な複素ガウス分布となる．

式 (7) が時間周波数の各グリッド (i, j) で互いに独立と仮定すると，複素スペクトログラム $\boldsymbol{S}_n$ の生成モデルは次式となる．

$$p(\boldsymbol{S}_n; \boldsymbol{R}_n) = \prod_{i,j} \frac{1}{\pi r_{ijn}} \exp\left(-\frac{|s_{ijn}|^2}{r_{ijn}}\right) \tag{8}$$

ここで，$\boldsymbol{R}_n \in \mathbb{R}_{>0}^{I \times J}$ は分散 r_{ijn} を要素に持つ分散行列であり，$\boldsymbol{R}_n$ のランクは高々 L である．複素スペクトログラムに対する生成モデル (8) は LGM と呼ばれる．複素要素 c_{ijln} の分散を $\rho_{ijln} = t_{iln} v_{ljn}$ とおけば，合成成分 s_{ijn} の分散 r_{ijn} は

$$r_{ijn} = \sum_l t_{iln} v_{ljn} \tag{9}$$

と表現でき，$\boldsymbol{S}_n$ のパワースペクトログラム $|\boldsymbol{S}_n|^{\cdot 2}$ を ISNMF で低ランク近似することが，LGM を仮定した分散の最尤推定と等価になる [28]〜[30]．但し，行列に対する絶対値 $|\cdot|$ 及びドット付き指数 $\cdot^{\cdot p}$ はそれぞれ要素毎の絶対値及び要素毎の p 乗を施した行列を表す．また，$t_{iln} \geq 0$ 及び $v_{ljn} \geq 0$ は非負変数であり，行列表記 $\boldsymbol{T}_n \in \mathbb{R}_{\geq 0}^{I \times L}$ 及び $\boldsymbol{V}_n \in \mathbb{R}_{\geq 0}^{L \times J}$ を導入すれば，式 (9) は次式の行列積で表される．

$$\boldsymbol{R}_n = \boldsymbol{T}_n \boldsymbol{V}_n \tag{10}$$

2.4 音源間の独立性の導入と目的関数

理想的には，式 (5) で得られる分離信号 $\boldsymbol{Y}_n$ は音源信号 $\boldsymbol{S}_n$ と一致することから，$p(\boldsymbol{Y}_n) = p(\boldsymbol{S}_n)$ と考えたうえで，分離行列 $\boldsymbol{W}_i$ の最尤推定問題を考える．音源間の独立性の仮定より，$p(\boldsymbol{y}_{ij}) = p(y_{ij1}, y_{ij,2}, \cdots, y_{ijN}) = \prod_n p(y_{ijn})$ を用いると，観測信号の尤度関数は次式となる [18].

$$\begin{aligned}
\mathcal{L} &= \prod_{i,j} p(\boldsymbol{x}_{ij} | \boldsymbol{W}_i) \\
&= \prod_{i,j} \left[|\det \boldsymbol{W}_i|^2 \cdot p(\boldsymbol{y}_{ij}) \right] \\
&= \prod_{i,j} \left[|\det \boldsymbol{W}_i|^2 \cdot \prod_n p(y_{ijn}) \right]
\end{aligned} \tag{11}$$

従って，負対数尤度関数は $p(y_{ijn}) = p(s_{ijn})$ より次式となる．

$$\begin{aligned}
-\log \mathcal{L} &= -\sum_{i,j} \log |\det \boldsymbol{W}_i|^2 - \sum_{i,j,n} \log p(y_{ijn}) \\
&= -2J \sum_i \log |\det \boldsymbol{W}_i| + \sum_{i,j,n} \left[\frac{|y_{ijn}|^2}{r_{ijn}} + \log r_{ijn} \right] + \text{const.}
\end{aligned}$$

分散 r_{ijn} を式 (9) とおけば，分散行列 $\boldsymbol{R}_n$（即ち分離信号 $\boldsymbol{Y}_n$ のパワースペクトログラムの期待値）に式 (10) なる非負低ランク制約を NMF 音源モデルとして与えることができ，結果として ILRMA の目的関数が次式として得られる．

$$\mathcal{J} = -2J \sum_i \log |\det \boldsymbol{W}_i| + \sum_{i,j,n} \left[\frac{|y_{ijn}|^2}{\sum_l t_{iln} v_{ljn}} + \log \sum_l t_{iln} v_{ljn} \right] \tag{12}$$

上式を分離行列 $\{\boldsymbol{W}_i\}_{i=1}^I$ と NMF の変数行列（基底行列 $\{\boldsymbol{T}_n\}_{n=1}^N$ 及びアクティベーション行列 $\{\boldsymbol{V}_n\}_{n=1}^N$）に関して最小化することで，各分離信号の時間周波数の共変性の強度（分散 $\boldsymbol{R}_n$）を低ランクに制約しながら，分離行列を最尤推定できる．最適化の過程で，混合されている各音源の時間周波数構造（パワースペクトログラム）を $\boldsymbol{R}_n = \boldsymbol{T}_n \boldsymbol{V}_n$ として緻密にモデル化することに成功すれば，パーミュテーション問題も高精度に回避できる．このような ILRMA の音源分離の原理を Fig. 3 に示す．$\boldsymbol{W}_i$ 及び NMF 音源モデル最適化の過程では，分離信号のパワースペクトログラム $|\boldsymbol{Y}_n|^{\cdot 2}$ を低ランク行列としてモデル化しながら，その時間周波数構造を共変性として加味した分離行列 $\boldsymbol{W}_i$ を推定する．混合前の各音源のパワースペクトログラム $|\boldsymbol{S}_n|^{\cdot 2}$ が低ランクであれば，混合信号のパワースペクトログラム $|\boldsymbol{X}_m|^{\cdot 2}$ のランクはいずれの m においても基本的に増加することから，ILRMA は分離信号の時間周波数構造を低ランクに誘導することで，パーミュテーション問題を避けつつ．互いに独立となる分離信号を推定していると解釈できる．

2.5 反復最適化更新式

式 (12) を最小化する $\boldsymbol{W}_i$，$\boldsymbol{T}_n$，及び $\boldsymbol{V}_n$ は，AuxIVA で提案された反復射影法（iterative projection: IP）[14] 及び ISNMF の更新式を交互に反復計算することで最適化できる．ILRMA の最適化アルゴリズムを Algorithm 1 に示す．但し，$\boldsymbol{e}_n \in \{0, 1\}^N$ は n 番目の要素のみが 1 で他の要素が 0 のベクトルである．また，行列間の演算子 $\odot$ 及び分数はそれぞれ要素毎の積及び商，$[\cdot]_{r,c}$ は行列の (r, c) 番目の要素を表す．これらの反復更新式は，1 回の更新で目的関数 $\mathcal{J}$ の値が増加しないことが理論的に保証されている．反復最適化の後は，分離行列の周波数毎のスケール任意性を解消するために，プロジェクションバック（projection back: PB）法 [31] を適用して分離信号を再計算し，その後 $\boldsymbol{Y}_n$ に逆 STFT を適用して時間領域の分離信号を得る．ILRMA の具

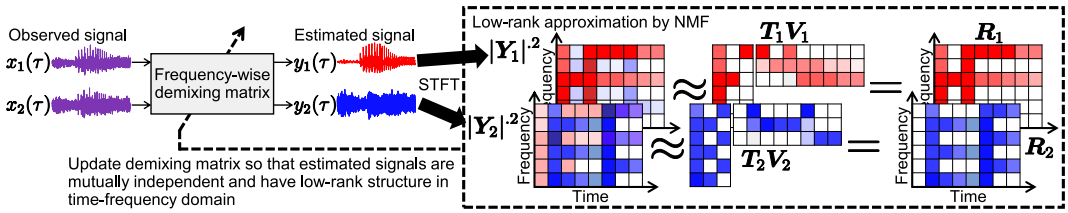

Fig. 3 Principle of source separation based on ILRMA, where $N = 2$ and τ is index of discrete time.

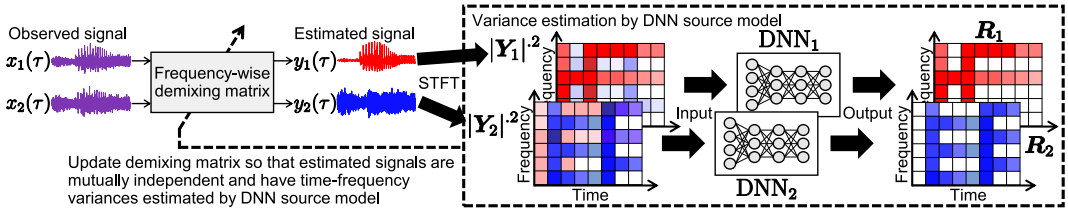

Fig. 4 Principle of source separation based on IDLMA, where $N = 2$ and τ is index of discrete time.

Algorithm 1 ILRMA

Input: $\{x_{ij}\}_{i=1,j=1}^{I,J}$, maxIter

Output: $\{y_{ij}\}_{i=1,j=1}^{I,J}$

1: Initialize $\{\boldsymbol{T}_n\}_{n=1}^{N}, \{\boldsymbol{V}_n\}_{n=1}^{N}, \{\boldsymbol{W}_i\}_{i=1}^{I}$

2: **for** iter $= 1, 2, \cdots,$ maxIter **do**

3: $\boldsymbol{T}_n \leftarrow \boldsymbol{T}_n \odot \left\{ \frac{[|\boldsymbol{Y}_n|^{\cdot 2} \odot (\boldsymbol{T}_n \boldsymbol{V}_n)^{\cdot -2}] \boldsymbol{V}_n^{\mathrm{T}}}{(\boldsymbol{T}_n \boldsymbol{V}_n)^{\cdot -1} \boldsymbol{V}_n^{\mathrm{T}}} \right\}^{\cdot \frac{1}{2}} \quad \forall n$

4: $\boldsymbol{V}_n \leftarrow \boldsymbol{V}_n \odot \left\{ \frac{\boldsymbol{T}_n^{\mathrm{T}} [|\boldsymbol{Y}_n|^{\cdot 2} \odot (\boldsymbol{T}_n \boldsymbol{V}_n)^{\cdot -2}]}{\boldsymbol{T}_n^{\mathrm{T}} (\boldsymbol{T}_n \boldsymbol{V}_n)^{\cdot -1}} \right\}^{\cdot \frac{1}{2}} \quad \forall n$

5: $\boldsymbol{U}_{in} \leftarrow \frac{1}{J} \sum_j \frac{1}{[\boldsymbol{T}_n \boldsymbol{V}_n]_{i,j}} \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^{\mathrm{H}} \ \forall i, n$

6: $\boldsymbol{w}_{in} \leftarrow (\boldsymbol{W}_i \boldsymbol{U}_{in})^{-1} \boldsymbol{e}_n \ \forall i, n$

7: $\boldsymbol{w}_{in} \leftarrow \boldsymbol{w}_{in} \left(\boldsymbol{w}_{in}^{\mathrm{H}} \boldsymbol{U}_{in} \boldsymbol{w}_{in} \right)^{-\frac{1}{2}} \quad \forall i, n$

8: **end for**

体的な実装は MATLAB(注2) と Python(注3) がそれぞれ公開されている．また，ILRMA による BSS のデモンストレーション(注4) も公開されているため，参照されたい．

3. DNN 音源モデルと ILRMA の融合

3.1 教師有り音源モデル導入の意図

ILRMA の BSS 実験 [16] で示されるように，混合前の各音源の時間周波数間の共変性に関する仮定（音源モデル）が適切であれば，パーミュテーション問題をほとんど起こすことなく分離行列を推定でき，高精度な音源分離が達成される．確かに，音楽信号中の各楽器音は NMF に基づく低ランク音源モデルが良く適合し，ILRMA は高い性能を発揮する．しかしながら，時間的にスペクトルが大きく変動する音声信号やボーカル信号等は低ランク音源モデルはあまり適切ではない．不適切な音源モデルを仮定した場合には，やはりパーミュテーション問題が発生し，分離性能が劣化する．音源の性質を陽に仮定せずとも適切な音源モデル $\boldsymbol{R}_n$ を推定できれば理想的であるが，それは BSS の枠組みでは非常に困難である．

分離対象音源の十分な学習データが用意できる場合は，その音源に対して適切な音源モデルを構築することは比較的容易である．特に，教師有り学習において大きな成果を上げている DNN に基づく手法は，音源分離問題に対してもその有効性が多くの文献で示されている（例えば [32]〜[36] 等）．一方，空間的な伝達系は，音源位置やマイクロホン位置，部屋の形状，残響時間等膨大な物理要因に依存することから，それらを網羅する学習データを用意することは非現実的である．従って，音源モデルには学習済の DNN を用い，分離行列は従来通りブラインドに推定する手法が合理的である．

上記の理由より，音源間の統計的独立性に基づくブラインドな分離行列推定と DNN に基づく教師有り音源モデルを組み合わせた手法として，2018 年に独立深層学習行列分析（independent deeply learned matrix analysis: IDLMA）[37]〜[39] が提案された．IDLMA は，ILRMA と同じような分離行列推定の過程で，DNN 音源モデルで推定される時間周波数分散行列 $\boldsymbol{R}_n$ を活用する BSS であり，ILRMA における教師無し NMF 音源モデルが教師有り DNN 音源モデルへと入れ替えられたものである．

3.2 IDLMA の処理の概要

Fig. 4 に IDLMA による音源分離の原理図を示す．IDLMA は ILRMA と同様に，式 (8) の生成モデルに基づいて音源モデル $\boldsymbol{R}_n$ 及び分離行列 $\boldsymbol{W}_i$ を推定する．このとき，混合信号から n 番目の音源の分散行列 $\boldsymbol{R}_n$ を推定する DNN を事前に学習しておき，これを DNN_n とする．例えば，ボーカル信号とその他の雑多な音源が混合した信号を入力とし，ボーカル信号のみの分散行列を出力するように DNN を学習することで，ボーカル信号を強調する DNN 音源モデルが得られる．このような特定の音源を強調する DNN を全音源（$\mathrm{DNN}_1, \mathrm{DNN}_2, \cdots, \mathrm{DNN}_N$）に対して学習することで，低ランク性やスパース性等の陽なモデルではなく学習データから得た適切な音源モデルを構築でき，より高精度な分散行列 $\boldsymbol{R}_n$ 及び分離行列 $\boldsymbol{W}_i$ の推定に活用できる．

IDLMA の処理の流れを Fig. 5 に示す．分離行列 $\boldsymbol{W}_i$ は観測信号 $\boldsymbol{X}_m$ と推定分散行列 $\boldsymbol{R}_n$ を用いて IP によって更新され，暫定的な分離信号 $\boldsymbol{Y}_n$ が得られる．空間モデルの推定には周波数

(注2)：https://github.com/d-kitamura/ILRMA/blob/master/ILRMA.m

(注3)：https://pyroomacoustics.readthedocs.io/en/pypi-release/pyroomacoustics.bss.ilrma.html

(注4)：http://d-kitamura.net/demo-ILRMA.html

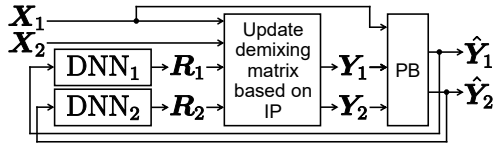

Fig. 5 Process flow of IDLMA in two sources case, where first channel is used as reference for PB.

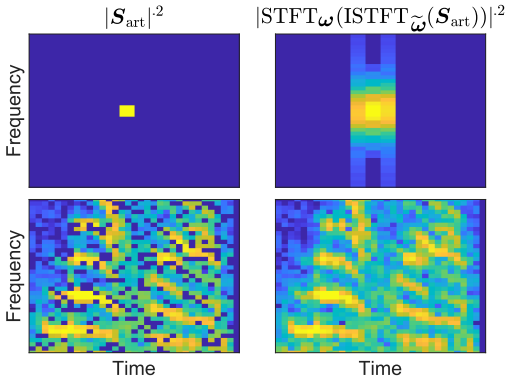

Fig. 6 Inconsistent power spectrograms $|S_{\mathrm{art}}|^{\cdot 2}$ (left column) and their consistent version (right column) obtained via inverse STFT and STFT.

毎のスケールの任意性があるため，$\boldsymbol{Y}_n$ に対してリファレンスチャネルを用いた PB 法を適用し，スケール補正後の分離信号 $\hat{\boldsymbol{Y}}_n \in \mathbb{C}^{I\times J}$ を得る．反復最適化の初期は分離が不十分であるため，$\hat{\boldsymbol{Y}}_n$ は他の音源成分が多く残留している．これを混合信号とみなし，音源モデル DNN$_n$ に入力することで，より分離が進んだ推定分散行列 $\boldsymbol{R}_n$ が得られ，これを再び分離行列 $\boldsymbol{W}_i$ の更新に用いる．このプロセスを繰り返すことで，より高精度な $\boldsymbol{W}_i$ が推定される．IDLMA の最終的な出力は $\hat{\boldsymbol{Y}}_n$ であり，これは時不変線形空間フィルタ $\boldsymbol{w}_{i,n}$ の出力であるため，IDLMA は ILRMA と同様に歪みの少ない分離信号が得られる．これは，例えば分離信号をさらに音声認識に適用する場合などにおいて大きな利点となる．

より具体的な処理内容や実験による性能評価については，文献 [37]〜[39] を参照されたい．また，ILRMA による BSS と IDLMA による教師有り音源分離を比較したデモンストレーション(注5)も公開されている．さらに近年では，IDLMA においてより安定かつ高速な分離行列の更新を行う手法 [40]，周波数ビン間相関をモデル化する独立深層学習テンソル分析 [41]，経験 Bayes を用いて DNN 音源モデルの信頼度を考慮する IDLMA [42] 等が提案されている．

4. スペクトログラム無矛盾性制約付き ILRMA

4.1 スペクトログラム無矛盾性とパーミュテーション問題

時間波形に STFT を施して得られるスペクトログラムは，STFT 中の窓関数とそのオーバーラップの影響を受け，近傍時間周波数グリッドに一貫した共起性を持つ．このような状態をスペクトログラム無矛盾 [43], [44] と呼ぶ．スペクトログラムに対して BSS 等の時間周波数領域での信号処理を施した場合，この一貫した共起性は通常崩され，矛盾したスペクトログラムとなる．矛盾スペクトログラムに逆 STFT を施した場合，無矛盾

Algorithm 2 Consistent ILRMA

Input: $\{x_{ij}\}_{i=1,j=1}^{I,J}$, maxIter

Output: $\{y_{ij}\}_{i=1,j=1}^{I,J}$

1: Initialize $\{\boldsymbol{T}_n\}_{n=1}^{N}, \{\boldsymbol{V}_n\}_{n=1}^{N}, \{\boldsymbol{W}_i\}_{i=1}^{I}$
2: **for** iter = $1, 2, \cdots,$ maxIter **do**
3: $\quad \boldsymbol{Y}_n \leftarrow \mathrm{STFT}_\omega(\mathrm{ISTFT}_{\tilde{\omega}}(\boldsymbol{Y}_n))\ \forall n$
4: $\quad \boldsymbol{T}_n \leftarrow \boldsymbol{T}_n \odot \left\{ \frac{[|\boldsymbol{Y}_n|^{\cdot 2} \odot (\boldsymbol{T}_n\boldsymbol{V}_n)^{\cdot -2}]\boldsymbol{V}_n^{\mathrm{T}}}{(\boldsymbol{T}_n\boldsymbol{V}_n)^{\cdot -1}\boldsymbol{V}_n^{\mathrm{T}}} \right\}^{\cdot \frac{1}{2}}\ \forall n$
5: $\quad \boldsymbol{V}_n \leftarrow \boldsymbol{V}_n \odot \left\{ \frac{\boldsymbol{T}_n^{\mathrm{T}}[|\boldsymbol{Y}_n|^{\cdot 2} \odot (\boldsymbol{T}_n\boldsymbol{V}_n)^{\cdot -2}]}{\boldsymbol{T}_n^{\mathrm{T}}(\boldsymbol{T}_n\boldsymbol{V}_n)^{\cdot -1}} \right\}^{\cdot \frac{1}{2}}\ \forall n$
6: $\quad \boldsymbol{U}_{in} \leftarrow \frac{1}{J}\sum_j \frac{1}{[\boldsymbol{T}_n\boldsymbol{V}_n]_{i,j}} \boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathrm{H}}\ \forall i, n$
7: $\quad \boldsymbol{w}_{in} \leftarrow (\boldsymbol{W}_i\boldsymbol{U}_{in})^{-1}\boldsymbol{e}_n\ \forall i, n$
8: $\quad \boldsymbol{w}_{in} \leftarrow \boldsymbol{w}_{in}\left(\boldsymbol{w}_{in}^{\mathrm{H}}\boldsymbol{U}_{in}\boldsymbol{w}_{in}\right)^{-\frac{1}{2}}\ \forall i, n$
9: $\quad \lambda_{in} \leftarrow \left[\boldsymbol{W}_i^{-1}\right]_{m_{\mathrm{ref}},n}\ \forall i, n$
10: $\quad \boldsymbol{w}_{in} \leftarrow \lambda_{in}\boldsymbol{w}_{in}\ \forall i, n$
11: $\quad y_{ijn} \leftarrow \boldsymbol{w}_{in}^{\mathrm{H}}\boldsymbol{x}_{ij}\ \forall i, j, n$
12: $\quad [\boldsymbol{T}_n]_{i,k} \leftarrow |\lambda_{in}|^2\,[\boldsymbol{T}_n]_{i,k}\ \forall i, k, n$
13: **end for**

なスペクトログラムへの射影と時間領域への変換が起こる．この例を Fig. 6 に示す．図の左列は人工的に作成した矛盾スペクトログラム $\boldsymbol{S}_{\mathrm{art}} \in \mathbb{C}^{I\times J}$ のパワーであり，右列は $\boldsymbol{S}_{\mathrm{art}}$ を逆 STFT し STFT した無矛盾なスペクトログラムのパワーである．矛盾スペクトログラムには無かった近傍時間周波数の共起成分が，逆 STFT 及び STFT によって復元されていることが分かる．

2020 年に，スペクトログラムの無矛盾性を最適化の過程で担保することでパーミュテーション問題を緩和する BSS [45] が提案された．また，同様の制約を ILRMA に導入したスペクトログラム無矛盾 ILRMA（consistent ILRMA）[46], [47] が新たに提案された．Fig. 1 に示すように，パーミュテーション問題が生じた分離信号のスペクトログラムは隣接周波数成分が不連続となるため，無矛盾性が大きく損なわれる [45], [47]．Consistent ILRMA では，NMF 音源モデルに加えて，分離信号 $\boldsymbol{Y}_n$ のスペクトログラム無矛盾性を最適化の過程で常に担保する．即ち，最適化の毎反復において，分離信号 $\boldsymbol{Y}_n$ を $\mathrm{STFT}_\omega(\mathrm{ISTFT}_{\tilde{\omega}}(\boldsymbol{Y}_n))$ に更新する．ここで，ω と $\tilde{\omega}$ はそれぞれ完全再構成条件を満たす解析窓及び合成窓である．この結果，周波数毎の音源成分が隣接周波数間で正しく整列された状態に誘導され，パーミュテーション問題を回避する能力が向上する．

4.2 Consistent ILRMA のアルゴリズム

Consistent ILRMA の最適化アルゴリズムを Algorithm 2 に示す．ここで，m_{ref} は PB 法のためのリファレンスチャネルインデクスを表す．Algorithm 1 と比較して，Algorithm 2 中の 3 行目（$\boldsymbol{Y}_n$ の無矛盾性を担保する更新）及び 9–12 行目（m_{ref} チャネルへの PB 法の適用）が新たに挿入されている．本手法についても，詳細は文献 [47] を参照されたい．また，MATLAB による具体的な実装が公開されている(注6)．

(注5)：http://d-kitamura.net/demo-IDLMA.html

(注6)：https://github.com/d-kitamura/ILRMA/blob/master/consistentILRMA.m

5. ま と め

本稿では，ILRMA に基づく BSS の原理について説明し，その後の発展手法である IDLMA 及び consistent ILRMA の二手法を紹介した．これら以外にも，ILRMA を起源とした拡張は多く提案されている．例えば，理論的な拡張においては，音源の生成モデルである LGM を優 Gauss 分布に拡張した ILRMA [48]，劣 Gauss 分布に拡張した ILRMA [49]，及び複素 Poisson 分布を用いた ILRMA [50]，IVA や ILRMA の最適化に主双対近接分離法を適用したアルゴリズム [51]，周波数ビン間を無相関化した領域での ILRMA [52]，音源モデルの概念をさらに一般化した時間周波数マスク BSS [53] とその応用である調波ベクトル分析 [54] 及び調波・打撃音 BSS [55]，分離行列の IP に基づく最適化を高速化した ILRMA [56] 等が挙げられる．また，応用的な拡張においては，災害救助ロボットにおける音声強調 [57]，拡散性雑音下での音声強調の前段処理 [58]，工業機器の異常音検知の前段処理 [59]，ユーザからの補助情報を用いるインタラクティブ ILRMA [60] 等が挙げられる．

謝辞　JSPS 19K20306 及び 19H01116 の助成を受けた．

文　献

[1] P. Comon, “Independent component analysis, a new concept?,” *Signal processing*, vol.36, no.3, pp.287–314, 1994.

[2] P. Smaragdis, “Blind separation of convolved mixtures in the frequency domain,” *Neurocomputing*, vol.22, pp.21–34, 1998.

[3] S. Kurita, H. Saruwatari, S. Kajita K. Takeda, and F. Itakura, “Evaluation of blind signal separation method using directivity pattern under reverberant conditions,” *Proc. ICASSP*, vol.5, pp.3140–3143, 2000.

[4] N. Murata, S. Ikeda, and A. Ziehe, “An approach to blind source separation based on temporal structure of speech signals,” *Neurocomputing*, vol.41, no.1–4, pp.1–24, 2001.

[5] H. Sawada, R. Mukai, S. Araki, and S. Makino, “A robust and precise method for solving the permutation problem of frequency-domain blind source separation,” *IEEE Trans. ASLP*, vol.12, no.5, pp.530–538, 2004.

[6] H. Sawada, R. Mukai, S. Araki, and S. Makino, “Convolutive blind source separation for more than two sources in the frequency domain,” *Proc. ICASSP*, pp.Ⅲ-885–Ⅲ-888, 2004.

[7] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, “Blind source separation based on a fast-convergence algorithm combining ICA and beamforming,” *IEEE Trans. ASLP*, vol.14, no.2, pp.666–678, 2006.

[8] S. Yamaji and D. Kitamura, “DNN-based permutation solver for frequency-domain independent component analysis in two-source mixture case,” *Proc. APSIPA ASC*, pp. 781–787, 2020.

[9] A. Hiroe, “Solution of permutation problem in frequency domain ICA using multivariate probability density functions,” *Proc. ICA*, pp.601–608, 2006.

[10] T. Kim, T. Eltoft, and T.-W. Lee, “Independent vector analysis: an extension of ICA to multivariate components,” *Proc. ICA*, pp. 165–172, 2006.

[11] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, “Blind source separation exploiting higher-order frequency dependencies,” *IEEE Trans. ASLP*, vol.15, no.1, pp.70–79, 2007.

[12] D. R. Hunter and K. Lange, “A tutorial on MM algorithms.” *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

[13] N. Ono and S. Miyabe, “Auxiliary-function-based independent component analysis for super-Gaussian sources,” *Proc. LVA/ICA*, pp.165–172, 2010.

[14] N. Ono, “Stable and fast update rules for independent vector analysis based on auxiliary function technique,” *Proc. WASPAA*, pp.189–192, 2011.

[15] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, “Efficient multichannel nonnegative matrix factorization exploiting rank-1 spatial model,” *Proc. ICASSP*, pp.276–280, 2015.

[16] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, “Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization,” *IEEE/ACM Trans. ASLP*, vol. 24, no. 9, pp. 1626–1641, 2016.

[17] 北村大地, 小野順貴, 澤田宏, 亀岡弘和, 猿渡洋, “独立低ランク行列分析に基づくブラインド音源分離,” **電子情報通信学会技術研究報告**, EA2017-56, vol. 117, no. 255, pp. 73–80, 2017.

[18] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, “Determined blind source separation with independent low-rank matrix analysis,” *Audio Source Separation. Signals and Communication Technology.*, S. Makino, Ed. Springer, Cham, pp. 125–155, 2018.

[19] D. D. Lee and H. S. Seung, “Learning the parts of objects by non-negative matrix factorization,” *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[20] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization” *Proc. NIPS*, pp. 556–562, 2000.

[21] H. Kameoka, T. Yoshioka, M. Hamamura, J. Le Roux, and K. Kashino, “Statistical model of speech signals based on composite autoregressive system with application to blind source separation,” *Proc. LVA/ICA*, pp. 245–253, 2010.

[22] A. Ozerov and C. Févotte, “Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation,” *IEEE Trans. ASLP*, vol. 18, no. 3, pp. 550–563, 2010.

[23] H. Kameoka, M. Nakano, K. Ochiai, Y. Imoto, K. Kashino, and S. Sagayama, “Constrained and regularized variants of non-negative matrix factorization incorporating music-specific constraints,” *Proc. ICASSP*, pp.5365–5368, 2012.

[24] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, “Multichannel extensions of non-negative matrix factorization with complex-valued data,” *IEEE Trans. ASLP*, vol. 21, no. 5, pp. 971–982, 2013.

[25] D. Kitamura, H. Saruwatari, K. Yagi, K. Shikano, Y. Takahashi, and K. Kondo, “Music signal separation based on supervised nonnegative matrix factorization with orthogonality and maximum-divergence penalties,” *IEICE Trans. Fundamentals*, vol. E97-A, no. 5, pp. 1113–1118, 2014.

[26] D. Kitamura, H. Saruwatari, H. Kameoka, Y. Takahashi, K. Kondo, and S. Nakamura, “Multichannel signal separation combining directional clustering and nonnegative matrix factorization with spectrogram restoration,” *IEEE/ACM Trans. ASLP*, vol. 23, no. 4, pp. 654–669, 2015.

[27] H. Sawada, N. Ono, H. Kameoka, D. Kitamura, and H. Saruwatari, “A review of blind source separation methods: Two converging routes to ILRMA originating from ICA and NMF,” *APSIPA Trans. SIP*, vol. 8, no. e12, pp. 1–14, 2019.

[28] C. Févotte, N. Bertin, and J.-L. Durrieu, “Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis,” *Neural Computation*, vol.21, no.3, pp.793–830, 2009.

[29] 北村大地, “複素生成モデルに基づく非負値行列因子分解と音源分離への応用,” **日本音響学会誌**, vol. 75, no. 3, pp. 130–138, 2019.

[30] D. Kitamura, “Nonnegative matrix factorization based on complex generative model,” *Acoustical Science and Technology*, vol. 40, no. 3, pp. 155–161, 2019.

[31] K. Matsuoka and S. Nakashima, “Minimal distortion principle for blind source separation,” *Proc. ICA*, pp. 722–727, 2001.

[32] P.-S. Huang, M. Kim, M. H.-Johnson, and P. Smaragdis, “Joint optimization of masks and deep recurrent neural networks for monaural source separation,” *IEEE/ACM Trans. ASLP*, vol. 23, no. 12, pp. 2136–2147, 2015.

[33] S. Uhlich, F. Giron, and Y. Mitsufuji, “Deep neural network based instrument extraction from music,” *Proc. ICASSP*, pp. 2135–2139, 2015.

[34] S. Araki, T. Hayashi, M. Delcroix, M. Fujimoto, K. Takeda, and T. Nakatani, “Exploring multi-channel features for denoising-autoencoder-based speech enhancement,” *Proc. ICASSP*, pp. 116–120, 2015.

[35] T. Nakatani, N. Ito, T. Higuchi, S. Araki, and K. Kinoshita, “Integrating DNN-based and spatial clustering-based mask estimation for robust MVDR beamforming,” *Proc. ICASSP*, pp. 286–290, 2017.

[36] A. A. Nugraha, A. Liutkus, and E. Vincent, “Multichannel audio source separation with deep neural networks ,” *IEEE/ACM Trans. ASLP*, vol. 24, no. 9, pp. 1652–1664, 2016.

[37] 北村大地, 角野隼斗, 高宗典玄, 高道慎之介, 猿渡洋, 猿渡洋, 小野順貴, “独立深層学習行列分析に基づく多チャネル音源分離の実験的評価,” **電子情報通信学会技術研究報告**, EA2017-104, vol. 117, no. 515, pp. 13–20, 2018.

[38] S. Mogami, H. Sumino, D. Kitamura, N. Takamune, S. Takamichi, H. Saruwatari, and N. Ono, “Independent deeply learned matrix analysis for multichannel audio source separation,” *Proc. EUSIPCO*, pp. 1571–1575, 2018.

[39] N. Makishima, S. Mogami, N. Takamune, D. Kitamura, H. Sumino, S. Takamichi, H. Saruwatari, and N. Ono, “Independent deeply learned matrix analysis for determined audio source separation,” *IEEE/ACM Trans. ASLP*, vol. 27, no. 10, pp. 1601–1615, 2019.

[40] N. Makishima, Y. Mitsui, N. Takamune, D. Kitamura, H. Saruwatari, Y. Takahashi, and K. Kondo, “Independent deeply learned matrix analysis with automatic selection of stable microphone-wise update and fast sourcewise update of demixing matrix,” *Signal Processing*, vol. 178, 107753, 2021.

[41] N. Narisawa, R. Ikeshita, N. Takamune, D. Kitamura, T. Nakamura, H. Saruwatari, and T. Nakatani, “Independent deeply learned tensor analysis for determined audio source separation,” *Proc. EUSIPCO*, 2021 (in press).

[42] T. Hasumi, T. Nakamura, N. Takamune, H. Saruwatari, D. Kitamura, Y. Takahashi, and K. Kondo, “Empirical Bayesian independent deeply learned matrix analysis for multichannel audio source separation,” *Proc. EUSIPCO*, 2021 (in press).

[43] J. Le Roux, H. Kameoka, N. Ono, and S. Sagayama, “Fast signal reconstruction from magnitude STFT spectrogram based on spectrogram consistency,” *Proc. DAFx*, 2010.

[44] J. Le Roux and E. Vincent, “Consistent Wiener filtering for audio source separation,” *IEEE SPL*, vol. 20, no. 3, pp. 217–220, 2013.

[45] K. Yatabe, “Consistent ICA: Determined BSS meets spectrogram consistency,” *IEEE SPL*, vol. 27, pp. 870–874, 2020.

[46] 北村大地, 矢田部浩平, “スペクトログラム無矛盾性を用いた独立低ランク行列分析の実験的評価,” **日本音響学会** *2021* **年春季研究発表会講演論文集**, pp. 121–124, 2021.

[47] D. Kitamura and K. Yatabe, “Consistent independent low-rank matrix analysis for determined blind source separation,” *EURASIP J. ASP*, vol. 2020, no. 46, 2020.

[48] D. Kitamura, S. Mogami, Y. Mitsui, N. Takamune, H. Saruwatari, N. Ono, Y. Takahashi, and K. Kondo, “Generalized independent low-rank matrix analysis using heavy-tailed distributions for blind source separation,” *EURASIP J. ASP*, vol. 2018, no. 28, 2018.

[49] S. Mogami, N. Takamune, D. Kitamura, H. Saruwatari, Y. Takahashi, K. Kondo, and N. Ono, “Independent low-rank matrix analysis based on time-variant sub-Gaussian source model for determined blind source separation,” *IEEE/ACM Trans. ASLP*, vol. 28, pp. 503–518, 2019.

[50] S. Mogami, Y. Mitsui, N. Takamune, D. Kitamura, H. Saruwatari, Y. Takahashi, K. Kondo, H. Nakajima, and H. Kameoka, “Independent low-rank matrix analysis based on generalized Kullback-Leibler divergence,” *IEICE Trans. Fundamentals*, vol. E102-A, no. 2, pp. 458–463, 2019.

[51] K. Yatabe and D. Kitamura, “Determined blind source separation via proximal splitting algorithm,” *Proc. ICASSP*, pp. 776–780, 2018.

[52] R. Ikeshita, N. Ito, T. Nakatani, and H. Sawada, “Independent low-rank matrix analysis with decorrelation learning,” *Proc. WASPAA*, pp. 288–292, 2019.

[53] K. Yatabe and D. Kitamura, “Time-frequency-masking-based determined BSS with application to sparse IVA,” *Proc. ICASSP*, pp. 715–719, 2019.

[54] K. Yatabe and D. Kitamura, “Determined BSS based on time-frequency masking and its application to harmonic vector analysis,” *IEEE/ACM Trans. ASLP*, vol. 29, pp. 1609–1625, 2021.

[55] S. Oyabu, D. Kitamura, and K. Yatabe, “Linear multichannel blind source separation based on time-frequency mask obtained by harmonic/percussive sound separation,” *Proc. ICASSP*, pp. 201–205 2021.

[56] T. Nakashima, R. Scheibler, Y. Wakabayashi, and N. Ono, “Faster independent low-rank matrix analysis with pairwise updates of demixing vectors,” *Proc. EUSIPCO*, pp. 301–305, 2020.

[57] Y. Bando, H. Saruwatari, N. Ono, S. Makino, K. Itoyama, D. Kitamura, M. Ishimura, M. Takakusaki, N. Mae, K. Yamaoka, Y. Matsui, Y. Ambe, M. Konyo, S. Tadokoro, K. Yoshii, and H. G. Okuno, “Low-latency and high-quality two-stage human-voice-enhancement system for a hose-shaped rescue robot,” *J. Robotics and Mechatronics*, vol. 29, no. 1, pp.198–212, 2017.

[58] Y. Kubo, N. Takamune, D. Kitamura, and H. Saruwatari, “Blind speech extraction based on rank-constrained spatial covariance matrix estimation with multivariate generalized Gaussian distribution,” *IEEE/ACM Trans. ASLP*, vol. 28, pp. 1948–1963, 2020.

[59] A. Aiba, M. Yoshida, D. Kitamura, S. Takamichi, and H. Saruwatari, “Noise robust acoustic anomaly detection system with nonnegative matrix factorization based on generalized Gaussian distribution,” *IEICE Trans. Inf. & Syst.*, vol. E104-D, no. 3, pp. 441–449, 2021.

[60] F. Oshima, M. Nakano, and D. Kitamura, “Interactive speech source separation based on independent low-rank matrix analysis,” *Acoustical Science and Technology*, vol. 42, no. 4, pp. 222–225, 2021.