Vol.2021-MUS-131 No.37
Vol.2021-SLP-137 No.37
2021/6/19

非負値行列因子分解を導入した product of experts型音源モデルに基づく 独立深層学習行列分析による多チャネル音源分離

蓮実 拓也[1]　中村 友彦[1]　高宗 典玄[1]　猿渡 洋[1]　北村 大地[2]　高橋 祐[3]　近藤 多伸[3]

概要：独立深層学習行列分析（independent deeply learned matrix analysis: IDLMA）は，事前に学習した深層ニューラルネットワーク（deep neural network: DNN）による音源のパワースペクトログラム推定と音源間の統計的独立性に基づく多チャネル音源分離手法である．適切に学習データを用意できれば従来の IDLMA は高い分離性能を示しているものの，学習データとテストデータの間で対象音源の音色に乖離がある場合に IDLMA の分離性能は低下してしまう．一方，パワースペクトログラム推定に教師なし非負値行列因子分解（nonnegative matrix factorization: NMF）を用いる独立低ランク行列分析（independent low-rank matrix analysis: ILRMA）では適切に学習された IDLMA に比べると分離性能は低いものの，教師なし手法であるためテストデータにしかない音源成分も表現できる．そこで，本研究では IDLMA の学習済み DNN による音源生成モデルと ILRMA の教師なし NMF の音源生成モデルを product of experts により統合した音源モデルを提案する．また，補助関数法と呼ばれる最適化原理を用いて，提案モデルに対する効率的なパラメータ更新アルゴリズムも導出する．楽音分離実験により，IDLMA に提案音源モデルを導入することで音源分離性能が向上することを示す．

1. はじめに

多チャネル音源分離は，マイクロホンアレイで観測された多チャネル信号から混合前の音源を推定する技術である [1]．ブラインド多チャネル音源分離の分野では，周波数領域独立成分分析（frequency domain independent component analysis: FDICA）[2]，独立ベクトル分析（independent vector analysis: IVA）[3], [4]，独立低ランク行列分析（independent low-rank matrix analysis: ILRMA）[5] といった音源間の独立性を最大化する手法が提案されている．この中でも，ILRMA は音源モデルが低ランクであることを仮定して，パワースペクトログラム推定に非負値行列因子分解（nonnegative matrix factorization: NMF）[6] を適用する手法であり，高い性能を示している．さらに，各音源の学習データが事前に十分入手できる教師あり音源分離の場合には，ILRMA の音源モデルを深層ニューラルネットワーク（deep neural network: DNN）に置き換えた独立深層学習行列分析（independent deeply learned matrix analysis: IDLMA）[7] がより高い分離性能を示している．

IDLMA では，パワースペクトログラムを推定するための DNN を音源毎に学習する．そのため，学習データに近い音源に対しては高い分離性能を示している．しかし，テストデータの性質が学習データの性質と同じであるとは必ずしも限らない．例えば，4 章で扱う楽音のデータセットである DSD100 データセット [8] 中のベース音は主にエレキベースであるため，DNN はエレキベースに基づいた音源モデルを学習することになる．推論時にもエレキベースの音であれば DNN が精緻な音源モデルを表現できるはずである．しかし，シンセベースのような音色の異なるベースが含まれる曲の場合には，4.2 節で示すように IDLMA の DNN 推論によって分離性能が低下する例が見られる．これは，DNN がテストデータ特有の成分を表現することができず，エレキベースの場合よりも音源モデルの推定が困難になるためだと考えられる．一方，ILRMA は音源モデルとして教師なし NMF を用いるため，エレキベースとシンセベースのどちらが含まれる場合でも，安定して音源モデルを推定することができる．

そこで，IDLMA の音源モデルに教師なし NMF を導入することによって，学習データとテストデータの不一致か

[1] 東京大学
The University of Tokyo, Tokyo, Japan
[2] 香川高等専門学校
National Institute of Technology, Kagawa College, Kagawa, Japan
[3] ヤマハ株式会社
Yamaha Corporation, Shizuoka, Japan

ⓒ 2021 Information Processing Society of Japan

ら生じる DNN 音源モデルの分離性能低下を軽減する方策について考える．本研究では，product of experts（PoE）[9] と呼ばれる確率モデル統合手法を用いて，DNN に基づく確率モデルと NMF に基づく確率モデルの両者の性質を兼ね備える新たな音源モデルを提案し，IDLMA に導入する．また，この提案モデルに対し，補助関数法 [10] と呼ばれる最適化原理に基づき効率的な音源分離アルゴリズムを導出する．提案法では，推論時に事前学習済み DNN のみの音源モデルで表現できない成分を教師なし NMF が表現しうる．実験により，学習データにほとんど含まれない性質を有するテストデータに対しても，分離性能がさらに向上することを確認する．

2. 従来法

2.1 定式化

音源数, チャネル数をそれぞれ N, M とし, 音源信号, 観測信号, 分離信号の短時間 Fourier 変換（short time Fourier transform: STFT）をそれぞれ以下のように定義する．

$$\boldsymbol{s}_{ij} = (s_{ij1}, \ldots, s_{ijN})^{\mathsf{T}} \in \mathbb{C}^N \tag{1}$$

$$\boldsymbol{x}_{ij} = (x_{ij1}, \ldots, x_{ijM})^{\mathsf{T}} \in \mathbb{C}^M \tag{2}$$

$$\boldsymbol{y}_{ij} = (y_{ij1}, \ldots, y_{ijN})^{\mathsf{T}} \in \mathbb{C}^N \tag{3}$$

ただし，$i = 1, \ldots, I$, $j = 1, \ldots, J$, $n = 1, \ldots, N$, $m = 1, \ldots, M$ はそれぞれ周波数，時間フレーム，音源，チャネルのインデクスであり，$^{\mathsf{T}}$ は転置を表す．また，(i, j) 成分が x_{ijm}，y_{ijn} である行列をそれぞれ $\boldsymbol{X}_m \in \mathbb{C}^{I\times J}$，$\boldsymbol{Y}_n \in \mathbb{C}^{I\times J}$ と定義する．混合系が時不変であり STFT の窓長が残響時間よりも長い場合，混合行列を $\boldsymbol{A}_i \in \mathbb{C}^{M\times N}$ として，観測信号 $\boldsymbol{x}_{ij}$ は

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

と表される．チャネル数が音源数に等しく（$M = N$），混合行列 $\boldsymbol{A}_i$ が正則であれば，分離行列を $\boldsymbol{W}_i = (\boldsymbol{w}_{i1}, \ldots, \boldsymbol{w}_{iN})^{\mathsf{H}} \in \mathbb{C}^{N\times M}$ として，分離信号 $\boldsymbol{y}_{ij}$ は

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

と表される．ただし，$^{\mathsf{H}}$ は Hermite 転置を表す．分離信号が次の平均 0，分散 r_{ijn} の円対称な複素 Gauss 分布に従うと仮定する [5], [7].

$$p(y_{ijn}; r_{ijn}) = \frac{1}{\pi r_{ijn}} \exp\left(-\frac{|y_{ijn}|^2}{r_{ijn}}\right) \tag{6}$$

式 (6) で表される生成モデルのもとで，r_{ijn}，$\boldsymbol{W}_i$ に関する最尤推定問題として定式化すると，観測信号の負対数尤度がコスト関数に対応する．当該コスト関数は，観測信号，分離信号のスペクトログラムの集合をそれぞれ $\mathcal{X} = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M\}$ と $\mathcal{Y} = \{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N\}$ として

$$\begin{aligned}\mathcal{L} &= -\log p(\mathcal{X}) \\ &= -\log p(\mathcal{Y}) - 2J\sum_i \log|\det \boldsymbol{W}_i| \\ &\overset{c}{=} \sum_{i,j,n}\left(\log r_{ijn} + \frac{|\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}|^2}{r_{ijn}}\right) - 2J\sum_i \log|\det \boldsymbol{W}_i|\end{aligned} \tag{7}$$

と表される [1]．ただし，$\overset{c}{=}$ は定数項を除いて等しいことを示す．r_{ijn} を (i, j) 成分に持つ行列を $\boldsymbol{R}_n \in \mathbb{R}_{\geq 0}^{I\times J}$ とする．$\boldsymbol{R}_n$ を，ILRMA では NMF で表現し，IDLMA では DNN で表現する．$\boldsymbol{R}_n$ の NMF による表現を $\boldsymbol{R}_n^{(\mathrm{NMF})} \in \mathbb{R}_{\geq 0}^{I\times J}$，DNN による表現を $\boldsymbol{R}_n^{(\mathrm{DNN})} \in \mathbb{R}_{\geq 0}^{I\times J}$ とし，それぞれの (i, j) 成分を $r_{ijn}^{(\mathrm{NMF})}$，$r_{ijn}^{(\mathrm{DNN})}$ とする．

2.2 独立低ランク行列分析 [5]

ILRMA では，$r_{ijn}^{(\mathrm{NMF})}$ を次のような低ランクな行列でモデル化する．

$$r_{ijn}^{(\mathrm{NMF})} = \sum_k t_{ik,n} v_{kj,n} \tag{8}$$

ただし，$t_{ik,n}$，$v_{kj,n}$ はそれぞれ音源 n に関する基底行列 $\boldsymbol{T}_n \in \mathbb{R}_{\geq 0}^{I\times K}$ の (i, k) 成分，アクティベーション行列 $\boldsymbol{V}_n \in \mathbb{R}_{\geq 0}^{K\times J}$ の (k, j) 成分である．$K \ll \min(I, J)$ は NMF の基底数である．式 (8) を式 (7) に代入すると，ILRMA のコスト関数 $\mathcal{L}_{\mathrm{ILRMA}}$ は次のように表される．

$$\begin{aligned}\mathcal{L}_{\mathrm{ILRMA}} \overset{c}{=} &\sum_{i,j,n}\left(\log\sum_k t_{ik,n}v_{kj,n} + \frac{|\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}|^2}{\sum_k t_{ik,n}v_{kj,n}}\right) \\ &- 2J\sum_i \log|\det \boldsymbol{W}_i|\end{aligned} \tag{9}$$

式 (9) のコスト関数は，$t_{ik,n}$，$v_{kj,n}$ に関して，板倉斎藤ダイバージェンス基準 NMF [11] のコスト関数と定数項を除いて等価である．板倉斎藤ダイバージェンスに基づく NMF に対して，補助関数法 [10] と呼ばれる最適化による反復型のパラメータ更新アルゴリズムが提案されている [11].

補助関数法 [10] は，コスト関数を最小化するための反復法の一種であり，反復によるコスト関数の単調非増加性が保証されている．目的変数 $\boldsymbol{\theta}$ を用いて表されるコスト関数 $f(\boldsymbol{\theta})$ について，以下の条件を満たす関数 $f^+(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}})$ を $f(\boldsymbol{\theta})$ の補助関数と呼び，$\bar{\boldsymbol{\theta}}$ を補助変数と呼ぶ.

$$f(\boldsymbol{\theta}) \leq f^+(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}}) \tag{10}$$

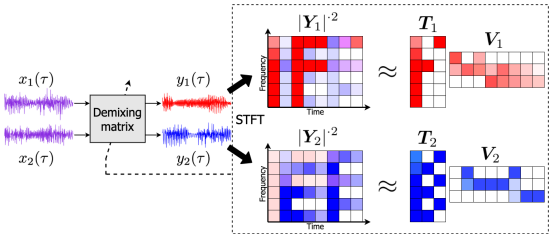
補助関数法では，補助変数に関する更新と目的変数に関する更新を以下のように繰り返す.

$$\bar{\boldsymbol{\theta}} \leftarrow \arg\min_{\bar{\boldsymbol{\theta}}} f^+(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}}) \tag{11}$$

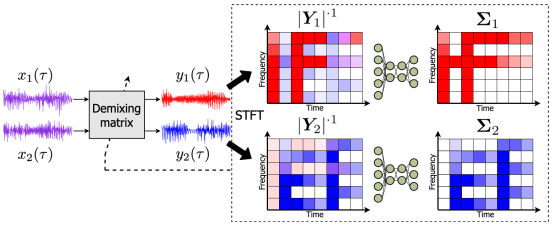
$$\boldsymbol{\theta} \leftarrow \arg\min_{\boldsymbol{\theta}} f^+(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}}) \tag{12}$$

式 (11)，(12) の計算が容易であれば，$f(\boldsymbol{\theta})$ を直接最小化

ⓒ 2021 Information Processing Society of Japan



(a) ILRMA の分離プロセス



(b) IDLMA の分離プロセス

図 1 (a) ILRMA および (b) IDLMA の分離プロセス．$|\cdot|^{\cdot q}$ は行列の要素ごとの絶対値の q 乗を表す．

するかわりに $f^{+}(\boldsymbol{\theta}, \bar{\boldsymbol{\theta}})$ の最小化を反復的に解くことで，明示的に最小化することが困難なコスト関数 $f(\boldsymbol{\theta})$ の局所最小解を求めることができる．

補助関数法を式 (9) の $t_{ik,n}$，$v_{kj,n}$ に関する最小化に適用すると，以下の更新式を得る [5]．

$$t_{ik,n} \leftarrow t_{ik,n} \left[\frac{\sum_j \frac{v_{kj,n}}{\left(\sum_{k'} t_{ik',n} v_{k'j,n}\right)^2} |y_{ij}|^2}{\sum_j \frac{v_{kj,n}}{\sum_{k'} t_{ik',n} v_{k'j,n}}} \right]^{\frac{1}{2}} \tag{13}$$

$$v_{kj,n} \leftarrow v_{kj,n} \left[\frac{\sum_i \frac{t_{ik,n}}{\left(\sum_{k'} t_{ik',n} v_{k'j,n}\right)^2} |y_{ij}|^2}{\sum_i \frac{t_{ik,n}}{\sum_{k'} t_{ik',n} v_{k'j,n}}} \right]^{\frac{1}{2}} \tag{14}$$

式 (13)，(14) より得られる t_{ik}，$v_{kj,n}$ を用いて，$r_{ijn}^{(\mathrm{NMF})}$ は式 (8) で計算される．

$r_{ijn}^{(\mathrm{NMF})}$ が与えられているもとでは，式 (9) は $\boldsymbol{w}_{in}$ の 2 次形式と $-\log\det$ 項の和の形式となる．この形式のコスト関数の最小化する手法の一つとして反復射影法（iterative projection: IP）[12] と呼ばれるコスト関数が単調非増加となる更新アルゴリズムが知られている．IP では周波数，音源ごとに以下の更新をおこなう．

$$\boldsymbol{w}_{in} \leftarrow (\boldsymbol{W}_i \boldsymbol{U}_{in})^{-1} \boldsymbol{e}_n \tag{15}$$

$$\boldsymbol{w}_{in} \leftarrow \frac{\boldsymbol{w}_{in}}{\sqrt{\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{U}_{in} \boldsymbol{w}_{in}}} \tag{16}$$

$$\boldsymbol{U}_{in} := \frac{1}{J} \sum_j \frac{1}{r_{ijn}^{(\mathrm{NMF})}} \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^{\mathsf{H}} \tag{17}$$

ただし，$\boldsymbol{e}_n \in \mathbb{R}^N$ は n 番目の要素が 1 でその他の要素が 0 の単位ベクトルである．

分離フィルタ $\boldsymbol{w}_{in}$ と $r_{ijn}^{(\mathrm{NMF})}$ の間にはスケールの不定性が存在するため，projection back（PB）[13] を適用することで，この問題を解決する．PB は分離信号の総和が基準マイクロホンの観測に一致するようにスケールを揃える．基準マイクロホンのインデクスを m_{ref} とすると

$$x_{ijm_{\mathrm{ref}}} = \sum_n y_{ijn} = \sum_n \boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij} \tag{18}$$

を満たすように，$\boldsymbol{y}_{ij}$ に以下の補正をおこなう．

$$\boldsymbol{y}_{ij} \leftarrow \mathrm{diag}(\boldsymbol{d}_i) \boldsymbol{y}_{ij} \tag{19}$$

ただし，$\mathrm{diag}(\boldsymbol{d}_i) \in \mathbb{C}^{N \times N}$ は $\boldsymbol{d}_i \in \mathbb{C}^N$ の成分を対角成分に持ち，非対角成分が 0 の対角行列である．$\boldsymbol{d}_i$ は次の式で与えられる．

$$\boldsymbol{d}_i = (\boldsymbol{W}_i^{\mathsf{T}})^{-1} \boldsymbol{e}_{m_{\mathrm{ref}}} \tag{20}$$

ILRMA の分離プロセスの概要を図 1(a) に示す．

2.3 独立深層学習行列分析 [7]

IDLMA では，DNN で $r_{ijn}^{(\mathrm{DNN})}$ をモデル化する．DNN_n は n 番目の音源についての $\sigma_{ijn} = \left(r_{ijn}^{(\mathrm{DNN})}\right)^{1/2}$ を出力する．つまり，IDLMA のコスト関数 $\mathcal{L}_{\mathrm{IDLMA}}$ は

$$\mathcal{L}_{\mathrm{IDLMA}} \overset{c}{=} \sum_{i,j,n} \left(\log \sigma_{ijn}^2 + \frac{|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|^2}{\sigma_{ijn}^2} \right) - 2J \sum_i \log |\det \boldsymbol{W}_i| \tag{21}$$

で定義される．ここで，(i,j) 成分が σ_{ijn} である行列を $\boldsymbol{\Sigma}_n \in \mathbb{R}_{\geq 0}^{I \times J}$ と定義する．

学習時には，単チャネルの瞬時混合信号を音源ごとに作成し，n 番目の音源に対する DNN_n を用意する．n 番目の音源信号のスペクトログラムの (i,j) 成分を $\tilde{s}_{ijn} \in \mathbb{C}$ とし，DNN_n の出力の (i,j) 成分を $\hat{\sigma}_{ijn} \in \mathbb{R}_{\geq 0}$ とすると，板倉斎藤ダイバージェンス基準で振幅スペクトルの誤差を最小化するように DNN を学習させる．

$$\mathcal{L}_{\mathrm{IDLMA}}^{(\mathrm{DNN})} = \sum_{i,j} \left(\frac{|\tilde{s}_{ijn}|^2 + \delta}{\hat{\sigma}_{ijn}^2 + \delta} - \log \frac{|\tilde{s}_{ijn}|^2 + \delta}{\hat{\sigma}_{ijn}^2 + \delta} \right) \tag{22}$$

ただし，δ はゼロ除算を防ぐための微小な値である．式 (22) の最小化は，学習データにおける σ_{ijn} についての式 (21) の最小化に相当する．

推論時には，事前学習済み DNN によって現在の分離信号のスペクトログラム $\boldsymbol{Y}_n$ から，

$$\sigma_{ijn} \leftarrow \max\left(\left[\mathrm{DNN}_n(|\boldsymbol{Y}_n|^{\cdot 1})\right]_{ij}, \varepsilon\right) \tag{23}$$

$$r_{ijn}^{(\mathrm{DNN})} \leftarrow \sigma_{ijn}^2 \tag{24}$$

と $r_{ijn}^{(\mathrm{DNN})}$ を更新する．ただし，$|\cdot|^{\cdot q}$ は行列の要素ごとの

絶対値の q 乗を表し，$[\cdot]_{ij}$ は行列の (i,j) 成分を表す．ε は数値的不安定性回避のための微小な値である．DNN$_n$ は式 (22) によって学習されているため，DNN の学習は式 (21) を σ_{ijn}^2 に関して最小化するシミュレーションに対応する．そのため，式 (24) による DNN 推定値を用いることはコスト関数の面で一貫性がある．

IDLMA のコスト関数 (21) は ILRMA のコスト関数 (9) と同様に式 (7) に由来している．そのため，$\boldsymbol{w}_{in}$ の更新に関しては，式 (17) の $r_{ijn}^{(\mathrm{NMF})}$ を $r_{ijn}^{(\mathrm{DNN})}$ に置き換えることで式 (15)，(16) による IP を適用できる．

IDLMA では DNN 出力のスケールが統一されているため，ILRMA のような $\boldsymbol{w}_{in}$ と $r_{ijn}^{(\mathrm{DNN})}$ の間のスケールの不定性の問題は生じないが，線形歪み低減のために式 (19) による PB を適用する．IDLMA の分離プロセスの概要を図 1(b) に示す．

3. 提案法

3.1 Product of experts 型音源モデル導入の動機と定式化

教師あり音源分離手法である IDLMA は，ILRMA の音源モデルに事前学習した DNN を導入することで，NMF が要請する対象音源のスペクトログラムの低ランク性を仮定できない場合でも高性能な分離を実現した [7]．しかし，DNN は事前学習が必要なため，学習時とテスト時のデータの性質が大きく異なる場合，性能が低下しうる．一方，ブラインド音源分離手法としての ILRMA では音源モデルとして教師なし NMF が用いられるため，このような場合でも頑健に動作する．したがって，DNN が表現可能な学習データに基づく音源の確率モデルと，テストデータに固有な成分を表現できる教師なし NMF による音源の確率モデルの両者に適合するような音源モデルを構築できれば，DNN の高い分離性能を活かしつつデータの不一致に起因する性能低下を軽減できるはずである．

そこで，本研究では分離信号の音源モデルを NMF と DNN の PoE [9] で表現する方法を提案する．PoE は所望の確率分布を expert と呼ばれる複数の確率分布の積で表現する手法であり，experts の論理積に対応する確率モデルを実現する．$r_{ijn}^{(\mathrm{NMF})}$ を用いた分離信号の確率モデルを $p(y_{ijn};r_{ijn}^{(\mathrm{NMF})})$，$r_{ijn}^{(\mathrm{DNN})}$ を用いた分離信号の確率モデルを $p(y_{ijn};r_{ijn}^{(\mathrm{DNN})})$ とすると，PoE を用いた音源モデルは以下のように表現できる．

$$p(y_{ijn};r_{ijn}^{(\mathrm{NMF})},r_{ijn}^{(\mathrm{DNN})}) \propto \left[p(y_{ijn};r_{ijn}^{(\mathrm{NMF})})\right]^{\alpha}\left[p(y_{ijn};r_{ijn}^{(\mathrm{DNN})})\right]^{\beta} \quad (25)$$

ただし，α，$\beta>0$ は各確率分布の PoE に対する寄与を調節するハイパーパラメータである．例えば，$\alpha<\beta$ であれば $p(y_{ijn};r_{ijn}^{(\mathrm{NMF})})$ よりも $p(y_{ijn};r_{ijn}^{(\mathrm{DNN})})$ を重視した PoE となる．$p(y_{ijn};r_{ijn}^{(\mathrm{NMF})})$ と $p(y_{ijn};r_{ijn}^{(\mathrm{DNN})})$ はどちらも複素ガウス分布であるためそれらの積も複素ガウス分布となり，式 (25) を整理すると以下の生成モデルを得られる．

$$p(y_{ijn};r_{ijn}^{(\mathrm{NMF})},r_{ijn}^{(\mathrm{DNN})}) = \frac{1}{\pi\tilde{r}_{ijn}}\exp\left(-\frac{|y_{ijn}|^2}{\tilde{r}_{ijn}}\right) \quad (26)$$

ただし，

$$\frac{1}{\tilde{r}_{ijn}} = \frac{\alpha}{r_{ijn}^{(\mathrm{NMF})}} + \frac{\beta}{r_{ijn}^{(\mathrm{DNN})}} \quad (27)$$

である．IDLMA と同様に式 (25) で表される生成モデルに対する最尤推定問題として定式化すると，提案法のコスト関数 $\mathcal{L}_{\mathrm{prop}}$ は次のように表される．

$$\begin{aligned}\mathcal{L}_{\mathrm{prop}} \overset{c}{=} &-\sum_{i,j,n}\log\left(\frac{\alpha}{\sum_k t_{ik,n}v_{kj,n}}+\frac{\beta}{\sigma_{ijn}^2}\right)\\ &+\sum_{i,j,n}\left(\frac{\alpha}{\sum_k t_{ik,n}v_{kj,n}}+\frac{\beta}{\sigma_{ijn}^2}\right)|\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}|^2\\ &-2J\sum_i\log|\det\boldsymbol{W}_i| \end{aligned} \quad (28)$$

提案法のコスト関数 (28) において，$\alpha=1$，$\beta=0$ とすると ILRMA のコスト関数 (9) と一致する．また，$\alpha=0$，$\beta=1$ とすると IDLMA のコスト関数 (21) と一致する．$\alpha+\beta=1$ であれば，式 (27) の $\tilde{r}_{ijn}$ は，$r_{ijn}^{(\mathrm{NMF})}$ と $r_{ijn}^{(\mathrm{DNN})}$ の重みつき調和平均であると解釈することができ，提案法は ILRMA と IDLMA の両者の利点を併せ持つ手法である．

図 2 に提案法の分離プロセスの概要を示す．図 1(a) や (b) と比較すると，ILRMA や IDLMA と提案法の違いは，音源モデルの部分にあると言える．提案法では DNN 音源モデルを IDLMA と同様に，NMF 音源モデルを ILRMA と同様に更新し，それらの重みつき調和平均で統合された音源モデルを得る．得られた音源モデルに IP を適用することで分離フィルタを更新し，PB をおこなうという流れで処理が行われる．

3.2 DNN の事前学習と分離アルゴリズム

3.2.1 DNN の学習と推論

提案法では，IDLMA と同様に DNN の事前学習をおこなう．DNN 学習のコスト関数について考えると，NMF がテストデータに由来する成分を表現するため，学習時には $t_{ik,n}$，$v_{kj,n}$ に関する項を含むべきではない．したがって，提案法の DNN 学習時のコスト関数は，式 (28) で $\alpha=0$，$\beta=1$ として設計する必要がある．3.1 節で述べたように，$\alpha=0$，$\beta=1$ の場合，提案法のコスト関数 (28) は IDLMA のコスト関数 (21) と一致するため，提案法の DNN 学習のコスト関数も IDLMA の DNN 学習のコスト関数 (22) に一致する．推論時には事前学習済み DNN は，IDLMA と同様に，式 (23) によって σ_{ijn} を更新し，式 (24) によって $r_{ijn}^{(\mathrm{DNN})}$ を得る．

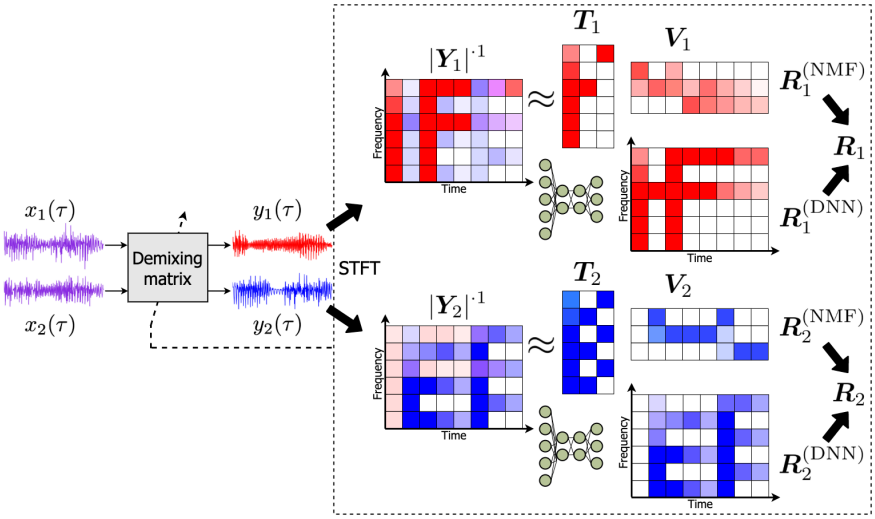


図 2 提案法の分離プロセス.

3.2.2 NMF に関する更新則

$t_{ik,n}$，$v_{kj,n}$ は $r_{ijn}^{(\mathrm{DNN})}$ を用いて，コスト関数 (28) が単調非増加となるように更新する．式 (9) と式 (28) の第 1 項に着目すると，ILRMA の場合と異なり対数関数の中に σ_{ijn} に関する項も含まれており，式 (13)，(14) の $t_{ik,n}$,$v_{kj,n}$ の更新アルゴリズムを用いてもコスト関数の単調非増加性は保証されない．また，式 (28) の第 1 項では対数関数の中に，第 2 項では分数関数の分母に k に関する和が含まれており，$t_{ik,n}$，$v_{kj,n}$ に関する最小化問題を解析的に解くことは難しい．そこで，文献 [5] と同様に，補助関数法 [10] を用いてコスト関数 (28) の補助関数を求め，コスト関数の単調非増加性が保証された $t_{ik,n}$，$v_{kj,n}$ に関する更新アルゴリズムを導出する．

まず，式 (28) の第 1 項について，α，β，$z>0$ の場合，

$$\frac{\partial^2}{\partial z^2}\left[-\log\left(\frac{\alpha}{z}+\frac{\beta}{\sigma^2}\right)\right]=-\frac{\alpha\sigma^2(\alpha\sigma^2+2\beta z)}{z^2(\alpha\sigma^2+\beta z)^2}<0 \tag{29}$$

より，$-\log\left(\alpha/z+\beta/\sigma^2\right)$ が上に凸となるので，当該項に接線不等式を適用すると以下の上界が得られる．

$$\begin{aligned}&-\log\left(\frac{\alpha}{\sum_k t_{ik,n}v_{kj,n}}+\frac{\beta}{\sigma_{ijn}^2}\right)\\&\le\frac{\alpha\sigma_{ijn}^2}{\alpha\sigma_{ijn}^2+\beta c_{ijn}}\frac{1}{c_{ijn}}\left(\sum_k t_{ik,n}v_{kj,n}-c_{ijn}\right)\\&\quad-\log\left(\frac{\alpha}{c_{ijn}}+\frac{\beta}{\sigma_{ijn}^2}\right)\end{aligned} \tag{30}$$

ただし，$c_{ijn}>0$ は補助変数である．不等式 (30) の等号成立条件は，

$$c_{ijn}=\sum_k t_{ik,n}v_{kj,n} \tag{31}$$

である．次に，式 (28) の第 2 項について，$1/z$ が $z>0$ で下に凸であるから Jensen の不等式より，

$$\frac{1}{\sum_k t_{ik,n}v_{kj,n}}\le\sum_k\frac{\lambda_{ijk,n}^2}{t_{ik,n}v_{kj,n}} \tag{32}$$

が成立する．ただし，$\lambda_{ijk,n}$ は補助変数で $\lambda_{ijk,n}\ge 0$，$\sum_k\lambda_{ijk,n}=1$ を満たす．不等式 (32) の等号成立条件は

$$\lambda_{ijk,n}=\frac{t_{ik}v_{kj}}{\sum_{k'}t_{ik',n}v_{k'j,n}} \tag{33}$$

である．

式 (30)，(32) より，$\mathcal{L}_{\mathrm{prop}}$ の補助関数 $\mathcal{L}_{\mathrm{prop}}^{+}$ は

$$\begin{aligned}&\mathcal{L}_{\mathrm{prop}}\\&\le\mathcal{L}_{\mathrm{prop}}^{+}\\&\overset{c}{=}\sum_{i,j,n}\frac{\alpha\sigma_{ijn}^2}{\alpha\sigma_{ijn}^2+\beta c_{ijn}}\frac{1}{c_{ijn}}\sum_k t_{ik,n}v_{kj,n}\\&\quad+\sum_{i,j,n}\left(\alpha\sum_k\frac{\lambda_{ijk,n}^2}{t_{ik,n}v_{kj,n}}+\frac{\beta}{\sigma_{ijn}^2}\right)|y_{ijn}|^2\\&\quad-\sum_{i,j,n}\left[\frac{\alpha\sigma_{ijn}^2}{\alpha\sigma_{ijn}^2+\beta c_{ijn}}+\log\left(\frac{\alpha}{c_{ijn}}+\frac{\beta}{\sigma_{ijn}^2}\right)\right]\end{aligned} \tag{34}$$

で与えられる．よって，$\mathcal{L}_{\mathrm{prop}}^{+}$ の $t_{ik,n}$，$v_{kj,n}$ に関する停留点を求めると

$$t_{ik,n}=\left[\frac{\displaystyle\sum_j\frac{\lambda_{ijk,n}^2}{v_{kj,n}}|y_{ijn}|^2}{\displaystyle\sum_j\frac{\sigma_{ij}^2}{\alpha\sigma_{ijn}^2+\beta c_{ijn}}\frac{v_{kj,n}}{c_{ijn}}}\right]^{\frac{1}{2}} \tag{35}$$

$$v_{kj,n}=\left[\frac{\displaystyle\sum_i\frac{\lambda_{ijk,n}^2}{t_{ik,n}}|y_{ijn}|^2}{\displaystyle\sum_i\frac{\sigma_{ijn}^2}{\alpha\sigma_{ijn}^2+\beta c_{ijn}}\frac{t_{ik,n}}{c_{ijn}}}\right]^{\frac{1}{2}} \tag{36}$$

ⓒ 2021 Information Processing Society of Japan

を得る．不等式の等号成立条件 (31)，(33) を式 (35)，(36) に代入すると，以下の更新式を得られる．

$$t_{ik,n} \leftarrow t_{ik,n} \left[\frac{\sum_j \frac{v_{kj,n}}{\left(\sum_{k'} t_{ik',n} v_{k'j,n}\right)^2} |y_{ijn}|^2}{\sum_j \frac{v_{kj}}{\left(\sum_{k'} t_{ik',n} v_{k'j,n}\right)^2} \tilde{r}_{ijn}} \right]^{\frac{1}{2}} \tag{37}$$

$$v_{kj,n} \leftarrow v_{kj,n} \left[\frac{\sum_i \frac{t_{ik,n}}{\left(\sum_{k'} t_{ik',n} v_{k'j,n}\right)^2} |y_{ijn}|^2}{\sum_i \frac{t_{ik,n}}{\left(\sum_{k'} t_{ik',n} v_{k'j,n}\right)^2} \tilde{r}_{ijn}} \right]^{\frac{1}{2}} \tag{38}$$

3.2.3 分離行列の推定

提案法では，$\tilde{r}_{ijn}$ は $r_{ijn}^{(\mathrm{NMF})}$ と $r_{ijn}^{(\mathrm{DNN})}$ を用いて式 (27) により得られる．$\tilde{r}_{ijn}$ を固定した場合，提案法のコスト関数は ILRMA や IDLMA と同様に式 (7) の形で書けるため，式 (17) の $r_{ijn}^{(\mathrm{NMF})}$ を $\tilde{r}_{ijn}$ に置き換えることで式 (15)，(16) による IP を適用できる．その後，式 (19) の PB をおこなう．ただし，提案法では，σ_{ijn} の更新に比べて $t_{ik,n}$，$v_{kj,n}$ や $\boldsymbol{w}_{in}$ の更新は収束が遅いため，1 回の σ_{ijn} の更新につき，$t_{ik,n}$，$v_{kj,n}$ および $\boldsymbol{w}_{in}$ を複数回更新する．以上をまとめて，提案法の更新アルゴリズムを Algorithm 1 に示す．ここで，$\tilde{\boldsymbol{R}}_n \in \mathbb{R}_{\geq 0}^{I \times J}$ は (i,j) 成分が $\tilde{r}_{ijn}$ の行列である．また，Algorithm 1 中の L は DNN 音源モデルの更新回数を，L' は DNN 音源モデル更新 1 回ごとの NMF 音源モデルと空間フィルタの更新回数を表す．全体としての合計の反復回数は LL' となる．

Algorithm 1 Iterative algorithm of proposed method

Input: $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_M$, $\mathrm{DNN}_1, \ldots, \mathrm{DNN}_N$
Output: $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N$

for l of number of DNN source model updates L **do**
  for all source index n **do**
    Update DNN source model $\boldsymbol{R}_n^{(\mathrm{DNN})}$ by Eq. (23) and (24)
    Update source model $\tilde{\boldsymbol{R}}_n$ by Eq. (27)
  end for
  for l' of number of NMF source model and demixing filter updates (per DNN source model update) L' **do**
    for all source index n **do**
      Update NMF source model $\boldsymbol{R}_n^{(\mathrm{NMF})}$ by Eq. (37) and (38)
      Update source model $\tilde{\boldsymbol{R}}_n$ by Eq. (27)
    end for
    for all frequency bin i and source index n **do**
      Update $\boldsymbol{w}_{in}$ by Eq. (15) and Eq. (16) with Eq. (17)
    end for
    for all frequency bin i and time frame j **do**
      Update $\boldsymbol{y}_{ij}$ by Eq. (5)
    end for
    for all frequency bin i and time frame j **do**
      Apply back-projection to $\boldsymbol{y}_{ij}$ by Eq. (19) with Eq. (20)
    end for
  end for
end for

4. 実験的評価

4.1 実験条件

提案法の有効性を評価するために, SiSEC2016 の DSD100 データセット [8] を用いて，楽音の多チャネル音源分離実験をおこなった．データセットは 8 kHz にダウンサンプリングし，STFT は 512 ms のハミング窓，256 ms のシフト長とした．テストデータとして，DSD100 データセットの `test` セットのうちのアルファベット順の上位 25 曲の 30 秒から 60 秒の区間をドライソースとして用いた．楽器の組み合わせはボーカルとベース（Vo./Ba.），ベースとドラム（Ba./Dr.），ボーカルとドラム（Vo./Dr.）の 3 種類であり，それぞれの組み合わせに RWCP データベース [14] の E2A インパルス応答（残響時間 300 ms）を畳み込むことで多チャネル信号を作成した．マイクロホンと音源の配置は図 3 に示す 2 通りであり，各楽器の組み合わせに対してテストデータの数はそれぞれ 50 曲である．

比較手法として，ILRMA [5]，IDLMA [7] および提案法を用いた．ILRMA では NMF 音源モデルの更新 1 回につき空間モデルを 1 回更新し，IDLMA では DNN 音源モデルの更新 1 回につき空間モデルを 10 回更新した．提案法では，DNN 音源モデルの更新 1 回につき NMF 音源モデル

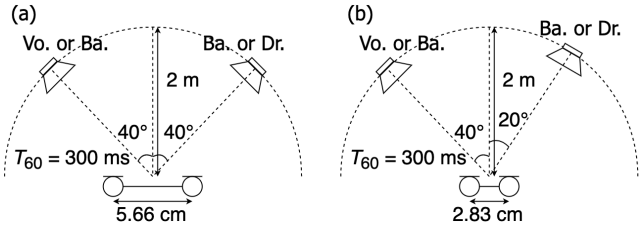
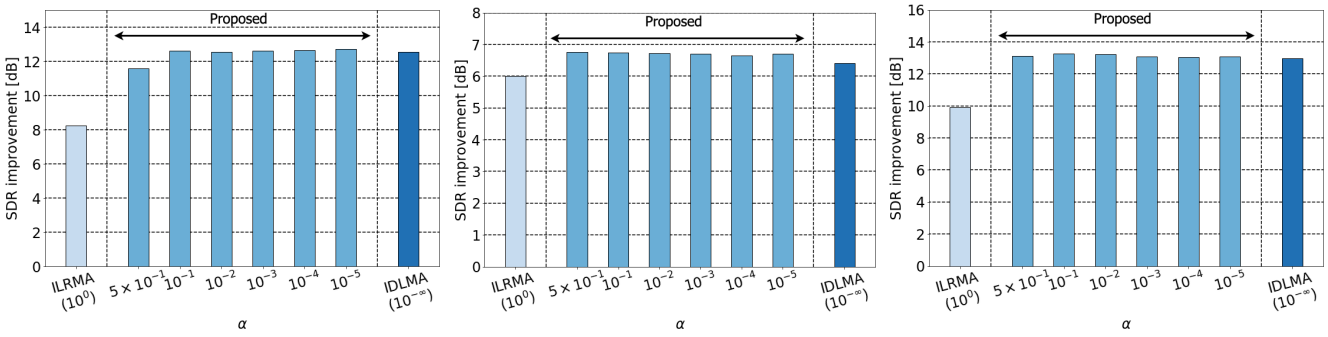


図 3 インパルス応答の録音条件.

と空間モデルを 10 回更新した．いずれの手法でも全体の更新回数は，空間モデルの更新回数が 100 回となるように揃えた．また，文献 [7] と同様に，ILRMA および提案法の NMF の基底数は $K = 20$ とし，IDLMA および提案法の DNN は 5 層の多層パーセプトロンとした．さらに，3.2.1 節で述べたように，提案法の DNN は IDLMA と同じ枠組みで学習させるため，提案法と IDLMA で同一の学習済み DNN を用いた．提案法では，$\alpha + \beta = 1$ の条件下で α の値を 5×10^{-1}，1×10^{-1}，1×10^{-2}，1×10^{-3}，1×10^{-4}，1×10^{-5} の 6 通りとして実験した．DNN を学習する際には，学習データとして DSD100 データセットの `dev` セット中の全 50 曲を用い，検証用データとして `test` セット中ののアルファベット順の下位 25 曲を用いた．DNN 学習では 2.3 項で述べたように単チャネルの混合信号を作成する．本実験では入力する混合音の各音源のゲインを除



(a) Vo./Ba. (b) Ba./Dr. (c) Vo./Dr.

図 4 $\alpha+\beta=1$ の制約下で α の値を変化させた場合の SDR 改善量.

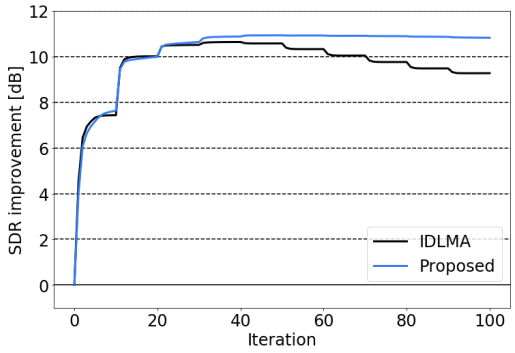


図 5 提案法と IDLMA の SDR 改善量の変化例（Ba./Dr.）.

いて，文献 [7] と同様に混合信号を作成した．本実験では DNN_n に入力する混合信号中の n 番目の目的音源のゲインのみ $[0.05, 1]$ の一様分布から生成し，それ以外の妨害音に対するゲインを形状母数が 0.1 と 1 のベータ分布から生成した．妨害音に対するゲインをこのように選ぶことで，目的音源のゲインが干渉音源よりも大きい混合音が多く生成され，IDLMA によって分離が進行した際の DNN の入力を模倣できる．

DNN は，バッチサイズを 128，学習率を 1.0，l^2 正則化の係数を 10^{-5} とした Adadelta [15] で 2000 エポック学習させた．また，勾配は l^2 ノルムが最大 10 となるようにクリッピングを適用した．他のハイパーパラメータに関しては，文献 [7] に倣い，$\delta=10^{-5}$ および $\varepsilon=10^{-1/2}$ とした．

4.2 結果

BSSEval toolbox [16] を用いて評価した，楽器の組み合わせごとのテストデータ 50 曲における SDR 改善量の平均を図 4 に示す．ILRMA は提案法の $\alpha=1$ に，IDLMA は $\alpha=0$ にそれぞれ対応する．従来法である ILRMA と IDLMA を比較すると，Vo./Ba. および Vo./Dr. の組み合わせでは，IDLMA の SDR 改善量は ILRMA よりも 3 dB 以上高い．しかし，Ba./Dr. の組み合わせでは 0.4 dB だけ高く，他の楽器の組み合わせでの SDR 改善量の向上よりも低かった．これは，Vo. のスペクトログラムと比較した場合，Ba. と Dr. のスペクトログラムが低ランクであるために NMF で音源モデルを表しやすいためであると考えられる．一方，提案法はどの楽器の組み合わせについても適切な α，β の値を選ぶことで IDLMA を上回る SDR 改善量となり，学習時とテスト時のデータの不一致を軽減することで，IDLMA の分離性能のさらなる向上が見込める．

図 5 にテストデータ中の `Girls Under Glass - We Feel Alright` の Ba./Dr. 分離における，提案法と IDLMA の反復ごとの 2 楽器の平均 SDR 改善量を示す．IDLMA では，40 反復目以降，$r_{ijn}^{(\mathrm{DNN})}$ の更新を行った反復（$41, 51, \cdots, 91$ 反復目）で SDR 改善量が低下している．DNN 学習のデータ中の Ba. の種類は主にエレキベースであるが，`Girls Under Glass - We Feel Alright` の Ba. はシンセベースである．このデータの性質の不一致が $r_{ijn}^{(\mathrm{DNN})}$ の更新による SDR 改善量を低下させていると考えられる．一方，提案法では 40 反復目以降も SDR 改善量が安定しており，学習データとテストデータの不一致による性能低下を，PoE を用いて NMF 音源モデルを導入し軽減できることを示している．

5. まとめ

本研究では，多チャネル音源分離において，従来の IDLMA の DNN 音源モデルでは表現するのが困難なテストデータ特有の成分を NMF 音源モデルで表現するために，NMF に基づく音源の確率モデルと DNN に基づく音源の確率モデルを PoE により統合した音源モデルを提案した．この音源モデルを用いたコスト関数に，NMF の基底とアクティベーションに関して補助関数法を適用することで，単調非増加となる更新を導出した．評価実験から，提案法では DNN 学習のデータ中にほとんど含まれない楽器種別だった場合でも安定して分離ができ，提案法の有効性が示された．

謝辞 本研究の一部は JSPS-CAS 二国間交流事業 JPJSBP120197203 と JSPS 科研費 19K20306，19H01116 及び 17H06101 の助成を受けた．

ⓒ 2021 Information Processing Society of Japan

参考文献

[1] Sawada, H., Ono, N., Kameoka, H., Kitamura, D. and Saruwatari, H.: A review of blind source separation methods: two converging routes to ILRMA originating from ICA and NMF, *APSIPA Transactions on Signal and Information Processing*, Vol. 8, No. e12, 14 pages (2019).

[2] Ikeda, S. and Murata, N.: A method of ICA in time-frequency domain, *Proceedings of International Conference on Independent Component Analysis and Signal Separation*, pp. 365–371 (1999).

[3] Kim, T., Attias, H. T., Lee, S.-Y. and Lee, T.-W.: Blind source separation exploiting higher-order frequency dependencies, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 15, No. 1, pp. 70–79 (2006).

[4] Hiroe, A.: Solution of permutation problem in frequency domain ICA, using multivariate probability density functions, *Proceedings of International Conference on Independent Component Analysis and Signal Separation*, pp. 601–608 (2006).

[5] Kitamura, D., Ono, N., Sawada, H., Kameoka, H. and Saruwatari, H.: Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 24, No. 9, pp. 1626–1641 (2016).

[6] Lee, D. D. and Seung, H. S.: Learning the parts of objects by non-negative matrix factorization, *Nature*, Vol. 401, No. 6755, pp. 788–791 (1999).

[7] Makishima, N., Mogami, S., Takamune, N., Kitamura, D., Sumino, H., Takamichi, S., Saruwatari, H. and Ono, N.: Independent deeply learned matrix analysis for determined audio source separation, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 27, No. 10, pp. 1601–1615 (2019).

[8] Liutkus, A., Stöter, F.-R., Rafii, Z., Kitamura, D., Rivet, B., Ito, N., Ono, N. and Fontecave, J.: The 2016 signal separation evaluation campaign, *Proceedings of International Conference on Latent Variable Analysis and Signal Separation*, pp. 323–332 (2017).

[9] Hinton, G. E.: Training products of experts by minimizing contrastive divergence, *Neural Computation*, Vol. 14, No. 8, pp. 1771–1800 (2002).

[10] Hunter, D. R. and Lange, K.: Quantile regression via an MM algorithm, *Journal of Computational and Graphical Statistics*, Vol. 9, No. 1, pp. 60–77 (2000).

[11] Févotte, C., Bertin, N. and Durrieu, J.-L.: Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis, *Neural Computation*, Vol. 21, No. 3, pp. 793–830 (2009).

[12] Ono, N.: Stable and fast update rules for independent vector analysis based on auxiliary function technique, *Proceedings of 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 189–192 (2011).

[13] Murata, N., Ikeda, S. and Ziehe, A.: An approach to blind source separation based on temporal structure of speech signals, *Neurocomputing*, Vol. 41, No. 1-4, pp. 1–24 (2001).

[14] Nakamura, S., Hiyane, K., Asano, F., Nishiura, T. and Yamada, T.: Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition, *Proceedings of the Second International Conference on Language Resources and Evaluation*, pp. 965–968 (2000).

[15] Zeiler, M. D.: Adadelta: an adaptive learning rate method, *arXiv preprint arXiv:1212.5701* (2012).

[16] Vincent, E., Gribonval, R. and Févotte, C.: Performance measurement in blind audio source separation, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 14, No. 4, pp. 1462–1469 (2006).

ⓒ 2021 Information Processing Society of Japan