# Efficient multichannel nonnegative matrix factorization with rank-1 spatial model*

◎ Daichi Kitamura (SOKENDAI), Nobutaka Ono (NII/SOKENDAI), Hiroshi Sawada (NTT), Hirokazu Kameoka (The University of Tokyo/NTT), Hiroshi Saruwatari (The University of Tokyo)

## 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元信号を推定する信号処理技術である．過決定条件 (音源数 ≤ 観測チャネル数) における BSS では，独立成分分析 (independent component analysis: ICA) [1] に基づく手法が主流であり，盛んに研究されてきた [2]．一方，モノラル信号等を対象とした劣決定条件下では，非負値行列因子分解 (nonnegative matrix factorization: NMF) [3] を応用した手法が注目を集めている．BSS は一般的に，話者分離や雑音抑圧が目的であるが，音楽を対象とした音源分離の研究も増加している [4].

時間周波数領域 ICA におけるパーミュテーション問題 [5] を解決する手法の一つとして，独立ベクトル分析 (independent vector analysis: IVA) [6] が提案されている．IVA では，周波数成分をまとめたベクトルを一変数として扱うため，パーミュテーション問題を引き起こすことがなく，高い音源分離性能を達成している．ICA や IVA では各音源の統計的な独立性を仮定して分離行列を推定するが，音楽信号を対象とした場合は周波数領域での重なりや時間的な共起が頻出するため，音源間の独立性が弱まることに起因して分離性能が劣化する可能性が高い．

一方，単一チャネルにおける NMF を用いた BSS では，分解されたスペクトル基底及びアクティベーションを音源毎にクラスタリングする必要があり，これは容易ではない．そこで，従来の NMF を多チャネル信号用に拡張した多チャネル NMF (multichannel NMF: MNMF) [7] が提案された．MNMF は，音源の空間情報に相当するチャネル間相関を用いて基底をクラスタリングすることで分離信号を得る．しかし，MNMF は空間推定と音源推定を同時に行う最適化であり，モデルの自由度が高い反面，最適解を見つけることは困難であるため，反復更新回数の増加や極端な初期値依存性をまねき，分離精度が不安定となる．

本稿では，音楽信号を対象とした安定で高速な BSS アルゴリズムを目標とし，従来の MNMF の空間特徴モデルをより制限された形に近似することで，新しい効率的な多チャネル NMF を提案する．また，提案手法が従来の IVA と密接に関連している事実を解析的に明らかにし，音源の独立性が弱くなる音楽信号に対しても高精度な分離が可能であることを実験により示す．

## 2 従来手法

### 2.1 定式化

過決定条件下において，簡便のために音源数とチャネル数を同じ $M$ とし，各時間周波数の多チャネルの音源信号，観測信号，分離信号をそれぞれ，

$$\boldsymbol{s}_{ij} = (s_{ij,1} \cdots s_{ij,M})^{\mathrm{t}} \tag{1}$$

$$\boldsymbol{x}_{ij} = (x_{ij,1} \cdots x_{ij,M})^{\mathrm{t}} \tag{2}$$

$$\boldsymbol{y}_{ij} = (y_{ij,1} \cdots y_{ij,M})^{\mathrm{t}} \tag{3}$$

と表す (要素はすべて複素数)．ここで，$1 \le i \le I$ ($i \in \mathbb{N}$) は周波数インデックス，$1 \le j \le J$ ($j \in \mathbb{N}$) は時間インデックス，$1 \le m \le M$ ($m \in \mathbb{N}$) はチャネル (音源) インデックスを示し，$^{\mathrm{t}}$ は転置を示す．今，混合系が線形時不変混合で，音源からマイクロホンまでのインパルス応答長が，短時間フーリエ変換の分析窓の長さよりも短い場合には，観測信号及び分離信号は次式のように表される．

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

ここで，$A_i = (\boldsymbol{a}_{i,1} \cdots \boldsymbol{a}_{i,M})$ は混合行列を表し，$\boldsymbol{a}_{i,m}$ は各音源のステアリングベクトルを表す．このとき，過決定条件下においては，分離ベクトル $\boldsymbol{w}_{i,m}$ で表現される分離行列 $W_i = (\boldsymbol{w}_{i,1} \cdots \boldsymbol{w}_{i,M})^{\mathrm{h}}$ が存在し，分離信号を次式で表現できる．

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

但し，$^{\mathrm{h}}$ はエルミート転置を表す．

### 2.2 IVA

従来の ICA を用いた BSS では，周波数ビン毎に独立な ICA を適用する．そのため，分離信号を周波数間でまとめるパーミュテーション問題 [5] を解かなければならないが，IVA では，音源毎に各周波数ビンをまとめたベクトル $\boldsymbol{y}_{j,m}$ を変数とする．

$$\boldsymbol{y}_{j,m} = (y_{1\,j,m} \cdots y_{I\,j,m})^{\mathrm{t}} \tag{6}$$

このようなベクトル変数を用いることで，周波数間の高次相関を考慮しつつ音源間は独立となるような分離行列を推定できる [6]．最小化すべきコスト関数は

$$Q_{\mathrm{IVA}}(\boldsymbol{W}) = \textstyle\sum_m \frac{1}{J} \sum_j G(\boldsymbol{y}_{j,m}) - \sum_i \log |\det \boldsymbol{W}_i| \tag{7}$$

で与えらる．ここで，$J$ は時間フレームの総数を表す．また，$G(\boldsymbol{y}_{i,m})$ はコントラスト関数と呼ばれ，$p(\boldsymbol{y}_{j,m})$ を $\boldsymbol{y}_{j,m}$ が従う確率密度分布としたとき，$G(\boldsymbol{y}_{j,m}) = -\log p(\boldsymbol{y}_{j,m})$ である．音源信号の事前分布を球状ラプラス分布と仮定する $G(\boldsymbol{y}_{i,m}) = \|\boldsymbol{y}_{i,m}\|_2$ が良く用いられている [6]．但し，$\|\cdot\|_2$ は $L_2$ ノルムを表す．式 (7) を最小化する $\boldsymbol{W}$ を求める際に，補助関数法を用いることで，効率的かつ安定的に解が求まる補助関数型 IVA が提案されている [8].

IVA は発話音声の混合などに対して高い分離性能を発揮する．しかし音楽信号のように，周波数領域での重なりや時間的な共起が頻出する信号に対しては，音源間の独立性が弱くなることに起因して，分離精度が劣化する問題がある．

### 2.3 MNMF

NMF を自然な形で多チャネル信号に拡張した MNMF では，観測信号を次のように表現する [7].

$$\mathsf{X}_{ij} = \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^{\mathrm{h}} \tag{8}$$

---

* ランク 1 空間モデルを用いた効率的な多チャネル非負値行列因子分解. by 北村大地 (総研大), 小野順貴 (NII/総研大), 澤田宏 (NTT), 亀岡弘和 (東大/NTT), 猿渡洋 (東大)

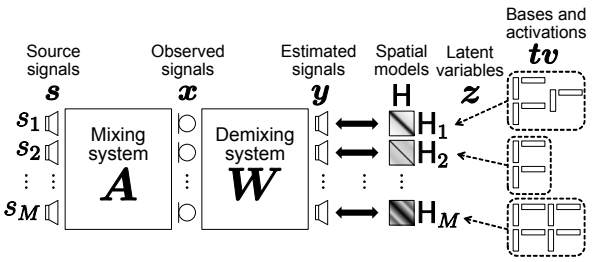


Fig. 1　Signal model of MNMF.

$M \times M$ のエルミート半正定値行列となる $\mathsf{X}_{ij}$ は，その対角要素が各マイクロホンで観測した $i, j$ 成分のパワー (実数) を示し，非対角要素がマイクロホン間の相関 (位相差) を示す複素数となる．この $\mathsf{X}_{ij}$ を，すべての $i$ と $j$ に対して近似する分解モデル $\hat{\mathsf{X}}_{ij}$ は以下で定義される．

$$\mathsf{X}_{ij} \approx \hat{\mathsf{X}}_{ij} = \sum_k \left(\sum_m \mathsf{H}_{i,m} z_{mk}\right) t_{ik} v_{kj} \tag{9}$$

ここで，$1 \le k \le K$ ($k \in \mathbb{N}$) は NMF における基底 (スペクトルパターン) のインデックスを示し，$\mathsf{H}_{i,m}$ は周波数 $i$ における音源 $m$ の空間相関行列を表す $M \times M$ のエルミート半正定値行列である．また，$z_{mk} \in \mathbb{R}_{\ge 0}$ は $k$ 番目の基底を $m$ 番目の音源に対応付ける潜在変数に相当し，$\sum_m z_{mk} = 1$ であり，$z_{mk} = 1$ のとき，$k$ 番目の基底は $m$ 番目の音源のみに寄与する．さらに，$t_{ik} \in \mathbb{R}_{\ge 0}$ 及び $v_{kj} \in \mathbb{R}_{\ge 0}$ はそれぞれ単一チャネル NMF の基底行列 $\boldsymbol{T}$ 及びアクティベーション行列 $\boldsymbol{V}$ の要素と等価である．MNMF のモデルの概念を Fig. 1 に示す．BSS においては Fig. 1 に示す混合系や分離系は未知である．MNMF では，観測信号を $\boldsymbol{TV}$ で近似分解すると同時に，各音源に一意に対応する空間相関行列 $\mathsf{H}$ を最適化し，潜在変数 $\boldsymbol{z}$ を用いて空間相関行列と基底及びアクティベーションを対応付けることで，分離信号 $\boldsymbol{y}$ を得る．$\mathsf{X}_{ij}$ と $\hat{\mathsf{X}}_{ij}$ 間の板倉斎藤擬距離は，定数項を省略すると

$$Q_{\mathrm{MNMF}} = \sum_{i,j} \left[\mathrm{tr}(\mathsf{X}_{ij} \hat{\mathsf{X}}_{ij}^{-1}) + \log\det \hat{\mathsf{X}}_{ij}\right] \tag{10}$$

で表される．MNMF においても補助関数法に基づく最適化が適用されており，単一チャネル NMF と同様に乗法型の反復更新式が導出されている [7].

MNMF は，各音源の空間相関行列 $\mathsf{H}$ に基づいて，潜在変数 $\boldsymbol{z}$ が基底を音源毎にまとめ上げることで，音源分離が達成される．しかし，この自由度の高いモデルでは，最適化すべき変数の増大に伴い局所解も増えるため，最適化が極めて困難になる．そのため，MNMF は分離精度が初期値に強く依存し，非常に不安定となる問題がある．

## 3　提案手法

### 3.1　空間相関行列のランク 1 モデルによる近似

Fig. 1 に示す混合系が，式 (4) のように混合行列 $\boldsymbol{A}_i = (\boldsymbol{a}_{i,1} \cdots \boldsymbol{a}_{i,M})$ で表現できる場合を考える．このとき，各音源の伝達系はステアリングベクトル $\boldsymbol{a}_{i,m}$ で与えられ，その外積となるランク 1 の半正定値エルミート行列 $\boldsymbol{a}_{i,m}\boldsymbol{a}_{i,m}^{\mathrm{h}}$ は MNMF における空間相関行列 $\mathsf{H}_{im}$ に相当する．

$$\mathsf{H}_{i,m} = \boldsymbol{a}_{i,m}\boldsymbol{a}_{i,m}^{\mathrm{h}} \tag{11}$$

### 3.2　分離行列と分離信号への変数変換

まず，式 (11) を式 (9) に代入すると

$$\begin{aligned}\hat{\mathsf{X}}_{ij} &= \sum_k \left(\sum_m \boldsymbol{a}_{i,m}\boldsymbol{a}_{i,m}^{\mathrm{h}} z_{mk}\right) t_{ik} v_{kj} \\ &= \sum_m \boldsymbol{a}_{i,m}\boldsymbol{a}_{i,m}^{\mathrm{h}} \sum_k z_{mk} t_{ik} v_{kj}\end{aligned} \tag{12}$$

を得る．ここで，$d_{ij,m} = \sum_k z_{mk} t_{ik} v_{kj}$ とおき，

$$\mathsf{D}_{ij} = \begin{pmatrix} d_{ij,1} & 0 & \cdots & 0 \\ 0 & d_{ij,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & d_{ij,M} \end{pmatrix} \tag{13}$$

なる対角行列を定義すると，$\hat{\mathsf{X}}_{ij}$ は混合行列 $\boldsymbol{A}_i$ を含む形で表すことができる．

$$\hat{\mathsf{X}}_{ij} = \boldsymbol{A}_i \mathsf{D}_{ij} \boldsymbol{A}_i^{\mathrm{h}} \tag{14}$$

次に，式 (14) を MNMF のコスト関数である式 (10) に代入する．

$$Q = \sum_{i,j} \left[\mathrm{tr}\left(\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathrm{h}} \left(\boldsymbol{A}_i^{\mathrm{h}}\right)^{-1} \mathsf{D}_{ij}^{-1} \boldsymbol{A}_i^{-1}\right) + \log\det \boldsymbol{A}_i \mathsf{D}_{ij} \boldsymbol{A}_i^{\mathrm{h}}\right] \tag{15}$$

ここで，過決定条件下では分離行列 $\boldsymbol{W}_i$ が存在するため，$\boldsymbol{W}_i = \boldsymbol{A}_i^{-1}$ 及び $\boldsymbol{y}_{ij} = \boldsymbol{W}_i \boldsymbol{x}_{ij}$ を用いて，混合行列から分離行列へ，観測信号から分離信号へそれぞれ変数変換を行うと，最終的に下記のコスト関数が得られる．

$$\begin{aligned} Q &= \sum_{i,j} \Big[\mathrm{tr}\left(\boldsymbol{W}_i^{-1}\boldsymbol{y}_{ij}\boldsymbol{y}_{ij}^{\mathrm{h}} \left(\boldsymbol{W}_i^{\mathrm{h}}\right)^{-1} \boldsymbol{W}_i^{\mathrm{h}} \mathsf{D}_{ij}^{-1} \boldsymbol{W}_i\right) \\ &\qquad + \log\left(\det \boldsymbol{A}_i\right)\left(\det \mathsf{D}_{ij}\right)\left(\det \boldsymbol{A}_i\right)^{\mathrm{h}}\Big] \\ &= \sum_{i,j} \Big[\mathrm{tr}\left(\boldsymbol{W}_i \boldsymbol{W}_i^{-1}\boldsymbol{y}_{ij}\boldsymbol{y}_{ij}^{\mathrm{h}} \left(\boldsymbol{W}_i^{\mathrm{h}}\right)^{-1} \boldsymbol{W}_i^{\mathrm{h}} \mathsf{D}_{ij}^{-1}\right) \\ &\qquad + 2\log\left|\det \boldsymbol{A}_i\right| + \log\det \mathsf{D}_{ij}\Big] \\ &= \sum_{i,j} \left[\mathrm{tr}\left(\boldsymbol{y}_{ij}\boldsymbol{y}_{ij}^{\mathrm{h}} \mathsf{D}_{ij}^{-1}\right) - 2\log\left|\det \boldsymbol{W}_i\right| + \sum_m \log d_{ij,m}\right] \\ &= \sum_{i,j} \left[\sum_m \frac{|\boldsymbol{y}_{ij}|^2}{\sum_k z_{mk} t_{ik} v_{kj}} - 2\log\left|\det \boldsymbol{W}_i\right| \right. \\ &\qquad \left. + \sum_m \log \sum_k z_{mk} t_{ik} v_{kj}\right] \end{aligned} \tag{16}$$

従来の MNMF では，音源毎の空間相関行列と基底及びアクティベーションを，潜在変数が結びつけることで分離信号を得ていたが，提案手法では，分離行列 $\boldsymbol{W}_i$ を求めることで音源分離が達成される．このとき，最適化の過程で暫定的に求まる $\boldsymbol{W}_i$ から仮の分離信号 $\boldsymbol{y}_{ij}$ が計算され，より良い $\boldsymbol{W}_i$ を得るために，$\boldsymbol{y}_{ij}$ を近似分解表現する $z_{mk}$，$t_{ik}$，及び $v_{kj}$ を求める必要がある．

### 3.3　各変数の反復更新式の導出

式 (16) を見ると，分離行列を含む第一項及び第二項は，式 (7) に示す IVA のコスト関数と本質的に等価であることが確認できる．この事実は，IVA と MNMF の関連性を明らかにする．即ち，時間周波数領域での線形混合仮説を導入した MNMF は，従来の IVA に NMF の基底分解を導入したモデルと本質的に等価である．但し，IVA では式 (7) において $G(\boldsymbol{y}_{i,m}) = \|\boldsymbol{y}_{i,m}\|_2$ として，音源の事前分布に球状ラプラス分布 (各周波数ビンで分散が一定) を仮定することが一般的である

Table 1　Music sources

| ID | Song (Artist / Name / Snip) | Part (Source 1 / Source2) |
|---|---|---|
| 1 | Bearlin / Roads / 85-99 | Bass / Piano |
| 2 | Tamy / Que pena tanto faz / 6-19 | Guitar / Vocal |
| 3 | Another dreamer / The ones we love / 69-94 | Guitar / Vocal |
| 4 | Fort minor / Remember the name / 54-78 | Violins_synth / Vocal |

が，式 (16) では板倉斎藤擬距離に基づいていることから，音源分布として，分散が時間周波数成分毎に定義された分散変動型ガウス分布を仮定したことに相当する．以上より，式 (16) を最小化する分離行列 $\boldsymbol{W}_i$ は，IVA における更新式を用いることで最適化が可能である．補助関数型 IVA [8] と同様に，$\boldsymbol{W}_i$ の更新式は次のように導出できる．

$$r_{ij,m} = \sum_k z_{mk} t_{ik} v_{kj} \tag{17}$$

$$V_{i,m} = \frac{1}{J} \sum_j \frac{1}{r_{ij,m}} \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^{\mathrm{h}} \tag{18}$$

$$\boldsymbol{w}_{i,m} \leftarrow \left(\boldsymbol{W}_i V_{i,m}\right)^{-1} \boldsymbol{e}_m \tag{19}$$

ここで，$\boldsymbol{e}_m$ は $m$ 番目の要素のみが 1 の単位ベクトルを示す．

次に，$z_{mk}$，$t_{ik}$，及び $v_{kj}$ について考える．式 (16) の第一項及び第三項は，潜在変数 $z_{mk}$ の存在を除けば，下記に示す板倉斎藤擬距離を用いた単一チャネル NMF のコスト関数と等価であることが確認できる．

$$Q_{\mathrm{ISNMF}} = \sum_{i,j} \left[ \frac{|y_{ij}|^2}{\sum_l t_{il} v_{lj}} + \log \sum_l t_{il} v_{lj} \right] \tag{20}$$

但し，$1 \leq l \leq L$ ($l \in \mathbb{N}$) は基底インデックスを示す．従って，潜在変数が $z_{mk} \in \{0, 1\}$ かつ $M$ 個ある音源の一つ一つが等しい数 $L$ 個の基底で表される場合 (即ち $L \times M = K$)，$z_{mk}$ を消すことができ，次に示す従来の単一チャネル NMF の更新式をチャネル $m$ 毎に適用することで，$t_{il,m}$ 及び $v_{lj,m}$ として更新できる．

$$t_{il,m} \leftarrow t_{il,m} \sqrt{\frac{\sum_j |y_{ij,m}|^2 v_{lj,m} \left(\sum_{l'} t_{il',m} v_{l'j,m}\right)^{-2}}{\sum_j v_{lj,m} \left(\sum_{l'} t_{il',m} v_{l'j,m}\right)^{-1}}} \tag{21}$$

$$v_{lj,m} \leftarrow v_{lj,m} \sqrt{\frac{\sum_i |y_{ij,m}|^2 t_{il,m} \left(\sum_{l'} t_{il',m} v_{l'j,m}\right)^{-2}}{\sum_i t_{il,m} \left(\sum_{l'} t_{il',m} v_{l'j,m}\right)^{-1}}} \tag{22}$$

あるいは MNMF と同様に，各分離音源に寄与する基底が $z_{mk}$ によって自動的に割り当てられるような柔軟なモデルを考える場合は，式 (16) を補助関数法で最小化することで，次の更新式を得ることができる．

$$z_{mk} \leftarrow z_{mk} \sqrt{\frac{\sum_{i,j} |y_{ij,m}|^2 t_{ik} v_{kj} \left(\sum_{k'} z_{mk'} t_{ik'} v_{k'j}\right)^{-2}}{\sum_{i,j} t_{ik} v_{kj} \left(\sum_{k'} z_{mk'} t_{ik'} v_{k'j}\right)^{-1}}} \tag{23}$$

$$t_{ik} \leftarrow t_{ik} \sqrt{\frac{\sum_{j,m} |y_{ij,m}|^2 z_{mk} v_{kj} \left(\sum_{k'} z_{mk'} t_{ik'} v_{k'j}\right)^{-2}}{\sum_{j,m} z_{mk} v_{kj} \left(\sum_{k'} z_{mk'} t_{ik'} v_{k'j}\right)^{-1}}} \tag{24}$$

$$v_{kj} \leftarrow v_{kj} \sqrt{\frac{\sum_{i,m} |y_{ij,m}|^2 z_{mk} t_{ik} \left(\sum_{k'} z_{mk'} t_{ik'} v_{k'j}\right)^{-2}}{\sum_{i,m} z_{mk} t_{ik} \left(\sum_{k'} z_{mk'} t_{ik'} v_{k'j}\right)^{-1}}} \tag{25}$$

但し，$\sum_m z_{mk} = 1$ を満たすために，更新毎に $z_{mk} \leftarrow z_{mk} / \sum_{m'} z_{m'k}$ とする．

以上より，式 (16) を最小化する変数は，式 (17)–(19) と，式 (21)–(22) あるいは式 (23)–(25) を交互に反復することで求まる．しかしながら，本手法では分散と分離行列がともに変数となっているため，スケールが決まらず，更新を重ねると推定分散 $r_{ij,m}$ が発散する可能性がある．そこで，更新毎に $\boldsymbol{W}_i$ と $\sum_k z_{mk} t_{ik} v_{kj}$ に同じ正規化をかけることで，これを防ぐ．最終的に projection back [9] をかけることで，信号を正しいスケールに戻すことができる．

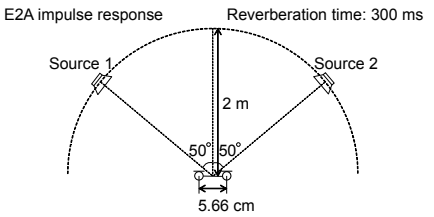

Fig. 2　Recording condition of room impulse response.

Table 2　Experimental conditions

| | |
|---|---|
| Sampling frequency | Down sampled from 44.1 kHz to 16 kHz |
| FFT length | 512 ms |
| Window shift | 128 ms |
| Initialization | $\boldsymbol{W}_i$ : identity matrix<br>$z_{mk}, t_{ik}, v_{kj}$ : nonnegative random values |
| Number of bases $K$ | Proposed method 1: $L = 30$ ($K = 60$)<br>Proposed method 2: $K = 60$ |
| Number of iterations | 200 |

## 4　評価実験

### 4.1　実験条件

提案手法の有効性を確認するために，音楽信号を対象とした分離評価実験を行った．実験では，IVA，提案手法 1 (式 (17)–(19) 及び (21)–(22) で更新)，及び提案手法 2 (式 (17)–(19) 及び (23)–(25) で更新) の 3 手法の比較を行った．提案手法 1 は各音源に同じ数 $L$ 個の基底を仮定して分離するが，提案手法 2 は基底の総数 $K$ のみを設定し $z_{mk}$ を同時に最適化することで，各音源に適応的に基底が割り当てられる．信号は Table 1 に示すように，SiSEC [10] の 4 種の音楽データ，各 2 楽器を選択した．さらに，RWCP database [11] に収録されている 2 つのインパルス応答 (E2A，Fig. 2 参照) を各楽器信号に畳み込み，ステレオ混合信号 ($M = 2$) を作成した．その他の実験条件は Table 2 に示す通りである．分離精度を示す客観評価値には，文献 [12] で定義されている signal-to-distortion ratio (SDR)，source-to-interference ratio (SIR)，及び sources-to-artificial ratio (SAR) を用いた．SDR は総合的な分離性能，SIR は非目的音の除去性能，SAR は人工的歪みの少なさの良い指標となる．

### 4.2　実験結果

Figs. 3–6 は，各手法において NMF の変数の初期値を変えて 10 回試行した際の平均と標準偏差を楽曲毎に示している．いずれの楽曲においても，提案手法 1 及び 2 が IVA よりも高い分離結果を示している．これは，音源間の統計的独立性のみを用いて分離する IVA よりも，厳密なスペクトル特徴を捉える基底分解を導入した提案手法が有効であったためと考えられる，また，Fig. 4 は提案手法 1 と 2 の違いが顕著に現れている．この楽曲は Source 1 の Guitar が同じフレーズを繰り返しており，Source 2 の Vocal よりも，遥かに少ない基底数で表現できると推測できる．しかし，提案手法 1 はいずれの音源にも同一数 $L$ の基底が用いられる為，適切な基底が学習されない可能性が高い．一方，提案手法 2 では $K$ 個の基底が各音源に適応的に割り当てられるため，より良いモデル化が可能となり，分離精度が大きく改善したと推測できる．

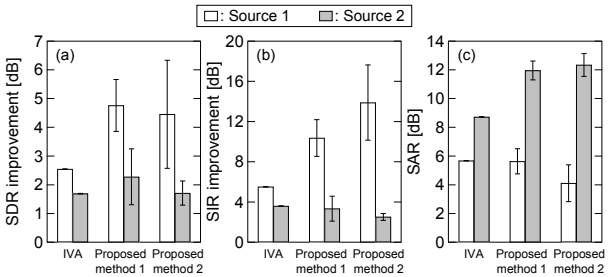


Fig. 3 Average scores for ID1 data: (a) SDR improvement, (b) SIR improvement, and (c) SAR.

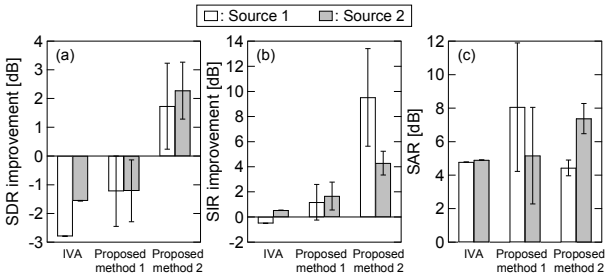


Fig. 4 Average scores for ID2 data: (a) SDR improvement, (b) SIR improvement, and (c) SAR.

Fig. 7 は ID4 の楽曲のスペクトログラム例を示している．この楽曲では，1 秒付近から長音の Violins_synth が生じており，Vocal 成分との重なりが顕著に現れる．Fig. 7 (c) の IVA では，1 秒以降の Vocal の推定精度が劣化しているが，Fig. 7 (d) の提案手法 2 では良く推定できている．IVA は全ての周波数成分を均一に扱うモデルのため，Violins_synth のように調波構造がはっきりしている音源などに対しては，分離性能が劣化してしまう．一方提案法では，これを基底分解により精度よくモデル化することができ，副次的に Vocal の分離精度が向上したと考えられる．

## 5 おわりに

本稿では，音楽信号に適した過決定条件における BSS として，従来の MNMF の空間特徴モデルがランク 1 となる近似を導入することで，より最適化が容易で効率的な MNMF モデルを提案した．また，提案手法は従来の IVA に NMF の基底分解を導入したモデルと本質的に等価であることを，解析的に明らかにした．客観評価実験の結果，提案手法は従来手法と比して，高精度な分離が可能であることが確認された．

**謝辞** 本研究は JSPS 特別研究員奨励費 26 · 10796 の助成を受けたものである．

## References

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

[2] 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.

[3] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” *Proc. Advances in Neural Information Processing Systems*, vol.13, pp.556–562, 2001.

[4] 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.

[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.

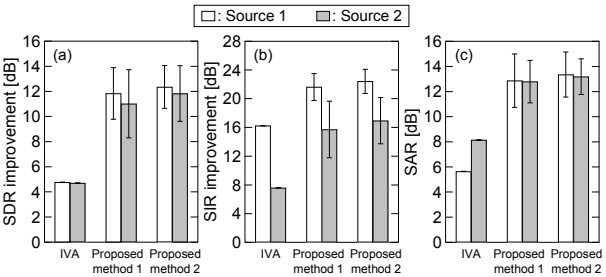


Fig. 5 Average scores for ID3 data: (a) SDR improvement, (b) SIR improvement, and (c) SAR.

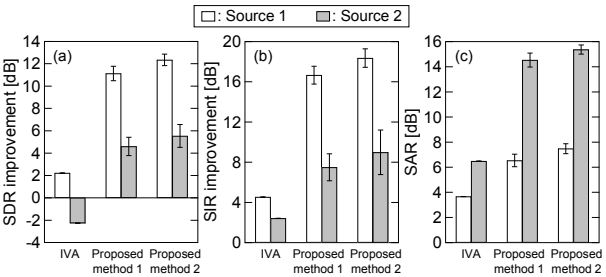


Fig. 6 Average scores for ID4 data: (a) SDR improvement, (b) SIR improvement, and (c) SAR.

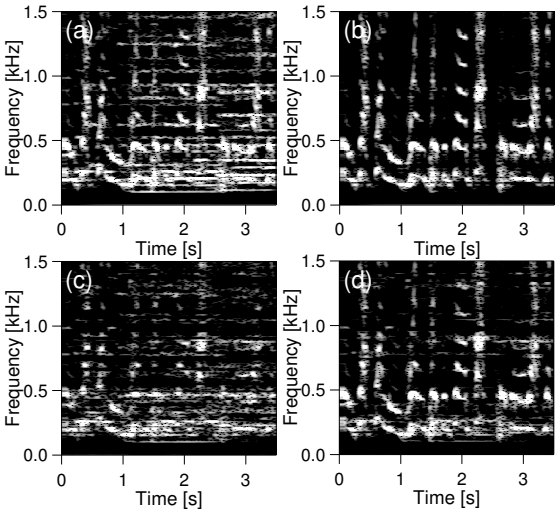


Fig. 7 Spectrogram of (a) mixed signal, (b) oracle vocal signal, (c) vocal signal estimated by IVA, and (d) vocal signal estimated by proposed method 2.

[6] 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.

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

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

[9] 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, pp.1–24, 2001.

[10] S. Araki, F. Nesta, E. Vincent, Z. Koldovsky, G. Nolte, A. Ziehe and A. Benichoux, “The 2011 signal separation evaluation campaign (SiSEC2011):-audio source separation,” *Proc. Latent Variable Analysis and Signal Separation*, pp.414–422, 2012.

[11] S. Nakamura, K. Hiyane, F. Asano, T. Nishiura and T. Yamada, “Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition,” *Proc. LREC*, pp.965–968, 2000.

[12] E. Vincent, R. Gribonval and C. Fevotte, “Performance measurement in blind audio source separation,” *IEEE Trans. ASLP*, vol.14, no.4, pp.1462–1469, 2006.