基底自動分配のための正則化を用いた非負値テンソル因子分解によるスポットフォーミング*

◎綾野翔馬 (香川高専)，李莉，関翔悟 (サイバーエージェント)，北村大地 (香川高専)

1 はじめに

目的音源抽出は観測信号から目的音源の信号のみを抽出する技術であり，音声認識など様々な音声信号処理に応用できる．マイクロホンアレイを収録に用いる場合，ビームフォーミング (beamforming: BF) に基づく目的音源抽出が一般である．しかし，BF は特定の方向の音源を強調するため，目的音源と同一方向にある干渉音源は抑制できない．そこで，Fig. 1 のように複数マイクロホンアレイを用いて，空間上の特定の領域に存在する目的音源のみを抽出する「スポットフォーミング」が提案されている [1][2][3][4]．これまでに，同期された複数マイクロホンアレイで空間フィルタを作成する手法 [1] や，複数マイクロホンアレイの配置の最適化手法が提案されている [2]．

他のアプローチを行う従来法として，非負値行列因子分解 (nonnegative matrix factorization: NMF) を用いたスポットフォーミングが提案されている [3]．この手法では，各マイクロホンアレイから目的音源のある方位に向けて BF を行い，その出力信号を時間方向に結合して NMF を適用する（Fig. 2 (a) 参照）．NMF の分解結果の共通成分が目的音源に対応するため，係数行列を閾値処理することでこれを推定する．しかし，この NMF モデルは基底間や BF の出力信号間の関係性が保たれておらず解釈性に欠ける．その結果，識別的な基底学習を促す正則化の導入が難しい問題がある．また，基底数や閾値といったハイパーパラメータの設定に性能が大きく依存してしまう．

本稿では，Fig. 2 (b) に示すように，非負値テンソル因子分解 (nonnegative tensor factorization: NTF) を用いたスポットフォーミング手法を提案する [5]．提案法はより解釈性の高いモデルとなっており，識別的な基底学習を促す正則化を導入することでハイパーパラメータに対して頑健なスポットフォーミングを達成する．

2 複数のマイクロホンアレイを用いたスポットフォーミング

2.1 想定する状況と信号モデル

本研究では，Fig. 1 に示す状況を考える．ただし，各マイクロホンアレイ内では同期録音しているが，マイクロホンアレイ間での同期は必ずしも保証されていない状況を考える．Fig. 1 の場合，各マイクロホンアレイは正面方向に BF を行うことで，目的音源を強調できるが，同一方位上に存在する干渉音源も強調されてしまう．一方で，干渉音源はマイクロホンアレイ毎に異なるため，各 BF の出力信号間で共通する音源成分を抽出できれば，目的音源のみが得られると

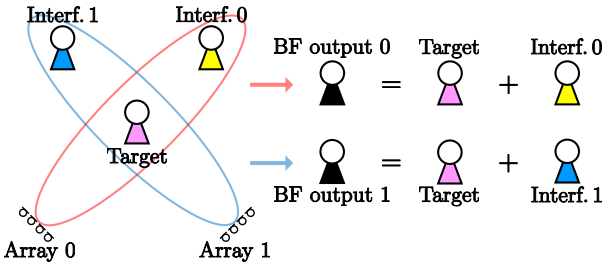
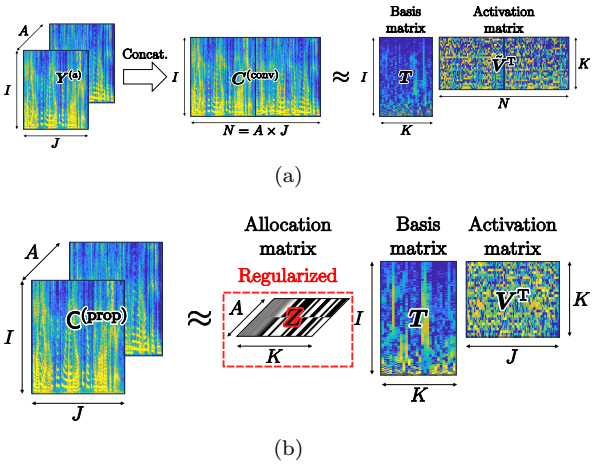


Fig. 1 Situations and signals estimated by BF.



(a)

(b)

Fig. 2 Decomposition by (a) NMF and (b) NTF.

考えられる．

a 番目のマイクロホンアレイによって得られる時間周波数成分を $\mathsf{X}^{(a)} \in \mathbb{C}^{I\times J\times M}$，その要素を $x^{(a)}_{i,j,m}$ と定義する．ここで，$i = 0, 1, \cdots, I-1$，$j = 0, 1, \cdots, J-1$，$a = 0, 1, \cdots, A-1$，及び $m = 0, 1, \cdots, M-1$ はそれぞれ周波数ビン，時間フレーム，マイクロホンアレイ，及び（各アレイ内の）マイクロホンのインデクスを示す．初めに，観測信号に対して，各マイクロホンアレイで $\boldsymbol{Y}^{(a)} = f_{\theta_a}(\mathsf{X}^{(a)})$ と表される BF を行う．ここで，$f_{\theta_a} : \mathbb{C}^{I\times J\times M} \mapsto \mathbb{C}^{I\times J}$ は方位 θ_a を強調する BF である．$\boldsymbol{Y}^{(a)} \in \mathbb{C}^{I\times J}$ は BF の出力信号であり，その要素を $y^{(a)}_{i,j}$ とする．このとき，スポットフォーミングはマイクロホンアレイ間の共通成分を BF の出力信号 $(\boldsymbol{Y}^{(a)})_{a=0}^{A-1}$ から推定する処理となる．

2.2 NMF に基づく従来法

従来法 [3] では，BF の出力 Y のマイクロホンアレイ方向と時間フレーム方向を結合して 2 次元行列を作成し，NMF の結果から時間周波数マスクを構成してスポットフォーミングする．NMF を適用する非負行列を $\boldsymbol{C}^{(\mathrm{conv})} \in \mathbb{R}^{I\times N}_{\geq 0}$ と定義する．

$$c^{(\mathrm{conv})}_{i,n} \coloneqq c^{(\mathrm{conv})}_{i,aJ+j} = \left| y^{(a)}_{i,j} \right| \tag{1}$$

*Automatic basis allocation using attractor-based regularization for NTF-based spotforming. By Shoma AYANO (NIT Kagawa), Li LI, Shogo SEKI (CyberAgent, Inc.), and Daichi KITAMURA (NIT Kagawa).

ここで，$N = AJ$ であり，$n = 0, 1, \cdots, N-1$ は $\boldsymbol{C}^{(\mathrm{conv})}$ の列インデクスである．さらに，この行列 $\boldsymbol{C}^{(\mathrm{conv})}$ に対して次式のように NMF を適用する．

$$\boldsymbol{C}^{(\mathrm{conv})} \simeq \boldsymbol{T}\tilde{\boldsymbol{V}}^{\mathrm{T}} \quad (c^{(\mathrm{conv})}_{i,n} \simeq \textstyle\sum_k t_{i,k}\tilde{v}_{n,k}) \tag{2}$$

ここで，$\boldsymbol{T} \in \{\boldsymbol{T} \in [0,1]^{I\times K} | \sum_i T_{i,k} = 1\}$ 及び $\tilde{\boldsymbol{V}} \in \mathbb{R}^{N\times K}_{\geq 0}$ は基底行列及び係数行列であり，$k = 0, 1, \cdots, K-1$ は NMF の基底ベクトルのインデクスである．NMF の目的関数は次式となる．

$$\begin{aligned} \underset{\boldsymbol{T},\tilde{\boldsymbol{V}}}{\text{minimize}} \ & \textstyle\sum_{i,n} \mathcal{D}\left(c^{(\mathrm{conv})}_{i,n} | \sum_k t_{i,k}\tilde{v}_{n,k}\right) \\ \text{s.t. } & t_{i,k}, \tilde{v}_{n,k} \geq 0 \ \forall i,k,n \end{aligned} \tag{3}$$

式 (3) を最小化する基底行列 $\boldsymbol{T}$ 及び係数行列 $\tilde{\boldsymbol{V}}$ を求めた後，$\tilde{\boldsymbol{V}}$ を用いてバイナリマスク行列 $\tilde{\boldsymbol{H}} \in \{0,1\}^{J\times K}$ を次式のように作成する．

$$\tilde{h}_{j,k} = \begin{cases} 1 \ (\text{if } \tilde{v}_{aJ+j,k} > \mu \ \forall a) \\ 0 \ (\text{o/w}) \end{cases} \tag{4}$$

ここで，$\mu \in \mathbb{R}_{\geq 0}$ は閾値である．このバイナリマスクはすべてのマイクロホンアレイで μ より大きい係数を持つ成分を共通の目的音源成分とみなし 1 とする．このバイナリマスクにより目的音源の推定振幅スペクトログラム $\hat{\boldsymbol{S}}^{(a)} \in \mathbb{C}^{I\times J}$ が次式で得られる．

$$\hat{s}^{(a)}_{i,j} = \frac{\sum_k \left(t_{i,k}\tilde{h}_{j,k}\tilde{v}_{aJ+j,k}\right)^2}{\sum_k \left(t_{i,k}\tilde{v}_{aJ+j,k}\right)^2} y^{(a)}_{i,j} \tag{5}$$

ここで，$\hat{s}^{(a)}_{i,j}$ は $\hat{\boldsymbol{S}}^{(a)}$ の (i,j) 番目の要素である．その後，$\hat{\boldsymbol{S}}^{(a)}$ に位相を与えて逆 STFT を行い，長さ L の信号 $\hat{\mathbf{s}}^{(a)} \in \mathbb{R}^L$ を得る．最後に，$(\hat{\mathbf{s}}^{(a)})^{A-1}_{a=0}$ に対して遅延和処理を計算することで，強調信号を得る．

3 提案法

3.1 NTF を用いる動機

Fig. 2 (a) に示す NMF のモデルは $\boldsymbol{T}$ 中の基底ベクトル間の関係や各 BF$(\boldsymbol{Y}^{(a)})^{A-1}_{a=0}$ 間の関係が保たれておらず，解釈性に欠ける．そのため，$\boldsymbol{T}$ や $\boldsymbol{V}$ に対して，目的音源成分と干渉音源成分の分離を誘導する正則化の導入が困難である．更に，K 及び τ のハイパーパラメータは観測信号に合わせて事前に調整する必要があり，実応用の際に難しい点となる．

本稿では，Fig. 2 (b) のように各 BF 出力に対して NTF を適用する手法を提案する．提案法は基底行列 $\boldsymbol{T}$ 中の各基底ベクトルを分配行列 $\boldsymbol{Z}$ で各マイクロホンアレイに振り分ける．さらに，$\boldsymbol{Z}$ に対して目的音源成分と干渉音源成分の分離を促進する「アトラクタ正則化」を導入し，基底ベクトルの振り分けがより識別的となるように誘導する．この正則化により目的音源に割り当てる基底ベクトル数も最適化されるため，ハイパーパラメータに対して頑健な手法となる．

3.2 NTF に基づくスポットフォーミング

Fig. 2 (b) に示すように NTF で分解する 3 階テンソル $\mathsf{C}^{(\mathrm{prop})} \in \mathbb{R}^{A\times I\times J}_{\geq 0}$ を次式で定義する．

$$c^{(\mathrm{prop})}_{a,i,j} \coloneqq \left| y^{(a)}_{i,j} \right| \tag{6}$$

ここで，$c^{(\mathrm{prop})}_{a,i,j}$ は $\mathsf{C}^{(\mathrm{prop})}$ の各要素である．3 階テンソル $\mathsf{C}^{(\mathrm{prop})}$ は，式 (1) とは異なりマイクロホンアレイ，時間，及び周波数の次元を維持している．$\mathsf{C}^{(\mathrm{prop})}$ を分配行列 $\boldsymbol{Z} = [\boldsymbol{z}_0, \boldsymbol{z}_1, \cdots, \boldsymbol{z}_{K-1}] \in \{\boldsymbol{Z} \in [0,1]^{A\times K} | \sum_a z_{a,k} = 1\}$，基底行列 $\boldsymbol{T} = [\boldsymbol{t}_0, \boldsymbol{t}_1, \cdots, \boldsymbol{t}_{K-1}]$，及び係数行列 $\boldsymbol{V} \in \mathbb{R}^{J\times K}_{\geq 0}$ の 3 つの行列に分解する．

$$c^{(\mathrm{prop})}_{a,i,j} \approx \textstyle\sum_k z_{a,k}t_{i,k}v_{j,k} \tag{7}$$

ここで，$z_{a,k}$ 及び $v_{j,k}$ は $\boldsymbol{Z}$ 及び $\boldsymbol{V}$ の要素であり，$\boldsymbol{z}_k$ 及び $\boldsymbol{t}_k$ は $\boldsymbol{Z}$ 及び $\boldsymbol{T}$ の列ベクトルを表す．

分配行列 $\boldsymbol{Z}$ は K 本の基底ベクトルを A 個のマイクロホンアレイに割り当てる役割を持つ．目的音源は BF 出力 $(|\boldsymbol{Y}^{(a)}|)^{A-1}_{a=0}$ の共通成分であるため，目的音源成分に対応する基底ベクトルは全マイクロホンアレイに割り当てられるべきである．従って，基底ベクトル $\boldsymbol{t}_k$ が目的音源を表すならば，$\boldsymbol{z}_k$ は $\boldsymbol{z}_k \approx [1/A, 1/A, \cdots, 1/A]^{\mathrm{T}}$ となる．逆に，$\boldsymbol{t}_k$ が干渉音源を表すならば，$\boldsymbol{z}_k$ は one-hot ベクトルに近づく．このような識別的な最適化は，NTF の持つ低ランク近似である程度実現されるが，提案法ではより識別的な分解結果を誘導する正則化を導入する．

提案法の最適化問題は次式で表される．

$$\begin{aligned} \underset{\boldsymbol{Z},\boldsymbol{T},\boldsymbol{V}}{\text{minimize}} \ & \textstyle\sum_{a,i,j} \mathcal{D}\left(c^{(\mathrm{prop})}_{a,i,j} | \sum_k z_{a,k}t_{i,k}v_{j,k}\right) \\ & + \mu \textstyle\sum_k \mathcal{R}\left(\boldsymbol{p}_{b_k} | \boldsymbol{z}_k\right) \\ \text{s.t. } & z_{a,k}, t_{i,k}, v_{i,k}, \geq 0 \ \forall a,i,j,k \end{aligned} \tag{8}$$

ここで，$\mu \geq 0$ は正則化係数である．さらに，正則化項は次式で表される．

$$\mathcal{R}\left(\boldsymbol{p}_{b_k} | \boldsymbol{z}_k\right) = \textstyle\sum_a \mathcal{D}(p_{a,b_k} | z_{a,k}), \tag{9}$$

$$\mathbb{P} = \{\boldsymbol{p}_0, \boldsymbol{p}_1, \cdots, \boldsymbol{p}_{B-1}\}, \tag{10}$$

$$\begin{aligned} \boldsymbol{p}_0 &\coloneqq [1/A, 1/A, \cdots, 1/A]^{\mathrm{T}} \in \{1/A\}^A, \\ \boldsymbol{p}_1 &\coloneqq [1, 0, \cdots, 0]^{\mathrm{T}} \in \{0,1\}^A, \\ \boldsymbol{p}_2 &\coloneqq [0, 1, \cdots, 0]^{\mathrm{T}} \in \{0,1\}^A, \\ &\vdots \\ \boldsymbol{p}_{B-1} &\coloneqq [0, 0, \cdots, 1]^{\mathrm{T}} \in \{0,1\}^A, \\ b_k &\in \underset{b}{\text{argmin}} \ \textstyle\sum_a \mathcal{D}(p_{a,b} | z_{a,k}) \end{aligned} \tag{11}$$

ここで，$p_{a,b}$ は $\boldsymbol{p}_b$ の要素，$b = 0, 1, \cdots, B-1$ はアトラクタとなるベクトル $(\boldsymbol{p}_0, \boldsymbol{p}_1, \cdots, \boldsymbol{p}_{B-1})$ のインデクスである（$B = A+1$ を満たす）．$\boldsymbol{p}_0$ は目的音源，$\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_{B-1}$ は各マイクロホンアレイごとに固有の干渉音源にそれぞれ対応するアトラクタベクトルである．b_k は式 (11) によって計算され，現在の $\boldsymbol{z}_k$ に最も近いアトラクタベクトルのインデクスとなる．つまり，式 (9) の正則化項は $\boldsymbol{z}_k$ を最も近いアトラクタベクトル $\boldsymbol{p}_{b_k}$ に近づける働きを持ち，基底 $(\boldsymbol{t}_k)^{K-1}_{k=0}$ のクラスタリングを強調する．さらにこの正則化は K 本の基底ベクトルを自動的に目的音源と干渉音源の成分にクラスタリングするため，目的音源に対応する基底ベクトル数は最適化の中で自動的に決まる．

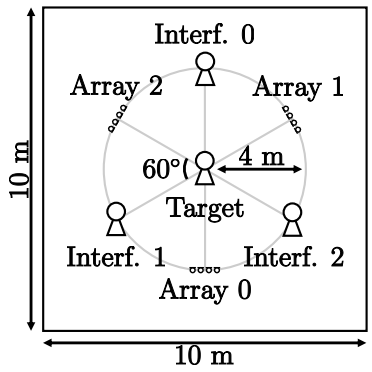


Fig. 3　Recording environment.

$\boldsymbol{Z}$, $\boldsymbol{T}$, 及び $\boldsymbol{V}$ を推定した後，バイナリマスク $\boldsymbol{h} = [h_0, h_1, \cdots, h_{K-1}]^{\mathrm{T}} \in \{0,1\}^K$ を次式で計算する．

$$h_k = \begin{cases} 1 & (\text{if } b_k = 0) \\ 0 & (\text{otherwise}) \end{cases} \tag{12}$$

$h_k = 1$ は $\boldsymbol{t}_k$ が目的音源の基底ベクトルであることを表す．式 (4) と比較して，式 (12) は時間フレームに非依存なバイナリマスクである．

式 (5) と同様に，目的音源のスペクトログラムは次式で得られる．

$$\hat{s}_{i,j}^{(a)} = \frac{\sum_k (h_k z_{a,k} t_{i,k} v_{j,k})^2}{\sum_k (z_{a,k} t_{i,k} v_{j,k})^2} y_{i,j}^{(a)} \tag{13}$$

その後の処理は従来法と同様である．

3.3　NTF における各変数の反復更新式の導出

式 (8) の最適化問題には上界最小化アルゴリズム [6] が適用できる．本研究では，式 (3)，(8)，及び (11) のコスト関数に，次式の一般化 Kullback–Leibler 擬距離を用いる．

$$\mathcal{D}(\mathrm{b}|\mathrm{a}) = \mathrm{b}\log\frac{\mathrm{b}}{\mathrm{a}} + \mathrm{a} - \mathrm{b} \tag{14}$$

式 (8) 中のコスト関数を $\mathcal{J}$ とおき，Jensen の不等式を適用することでコスト関数の上界 $\mathcal{J}^+ \geq \mathcal{J}$ を得る．

$$\begin{aligned} \mathcal{J}^+ \overset{c}{=} & \sum_{a,i,j} \Big(-c_{a,i,j}^{(\mathrm{prop})} \sum_k \alpha_{a,i,j,k} \log \tfrac{z_{a,k} t_{i,k} v_{j,k}}{\alpha_{a,i,j,k}} \\ & + \textstyle\sum_k z_{a,k} t_{i,k} v_{j,k} \Big) \\ & + \mu \textstyle\sum_{a,k} \left(-p_{a,b_k} \log z_{a,k} + z_{a,k} \right) \end{aligned} \tag{15}$$

ここで，$\overset{c}{=}$ は定数を除いて一致を示す演算子であり，$\alpha_{a,i,j,k} > 0$ は $\sum_k \alpha_{a,i,j,k} = 1$ を満たす補助変数である．等式 $\mathcal{J}^+ = \mathcal{J}$ の必要十分条件は次式である．

$$\alpha_{a,i,j,k} = \frac{z_{a,k} t_{i,k} v_{j,k}}{\sum_{k'} z_{a,k'} t_{i,k'} z_{j,k'}} \tag{16}$$

$\partial \mathcal{J}^+ / \partial z_{a,k} = 0$ より，次式を得る．

$$\sum_{i,j} \left(-c_{a,i,j}^{(\mathrm{prop})} \tfrac{\alpha_{a,i,j,k}}{z_{a,k}} + t_{i,k} v_{j,k} \right) + \mu \left(-\tfrac{p_{a,b_k}}{z_{a,k}} + 1 \right) = 0$$

これを $z_{a,k}$ について解けば，次式となる．

$$z_{a,k} = \frac{\sum_{i,j} c_{a,i,j}^{(\mathrm{prop})} \alpha_{a,i,j,k} + \mu p_{a,b_k}}{\sum_{i,j} t_{i,k} v_{j,k} + \mu} \tag{17}$$

Table 1　Dry sources

File name	Source	Duration [s]
84_121123_000008_000002.wav	Target	3.38
652_130737_000012_000000.wav	Interf. 0	4.05
3000_15664_000020_000005.wav	Interf. 1	4.07
1272_141231_000024_000005.wav	Interf. 2	3.47

Table 2　Experimental conditions

Sampling frequency	Down sampled to 16 kHz
Window function used in STFT	Hann window
Window length in STFT	32 ms
Window shift length in STFT	16 ms
Number of iterations in NMF/NTF	100 times
Initial values of $\boldsymbol{T}$, $\tilde{\boldsymbol{V}}$, and $\boldsymbol{V}$	Uniform random values in the range $(0, 1)$
Initial values of $\boldsymbol{Z}$	All the elements are set to $1/A$
Weight coefficient μ	$\mu = 0$ for first 50 iterations, and $\mu > 0$ for the rest of iterations

式 (16) を式 (17) に代入して，$\boldsymbol{Z}$ の反復更新式を得る．

$$z_{a,k} \leftarrow \frac{z_{a,k} \sum_{i,j} c_{a,i,j}^{(\mathrm{prop})} \frac{t_{i,k} v_{j,k}}{\sum_{k'} z_{a,k'} t_{i,k'} v_{j,k'}} + \mu p_{a,b_k}}{\sum_{i,j} t_{i,k} v_{j,k} + \mu} \tag{18}$$

ただし，b_k は $\boldsymbol{Z}$ の更新前に式 (11) で更新する必要がある．$\boldsymbol{T}$ 及び $\boldsymbol{V}$ の反復更新式は次式として得られる．

$$t_{i,k} \leftarrow t_{i,k} \frac{\sum_{a,j} c_{a,i,j}^{(\mathrm{prop})} \frac{z_{a,k} v_{j,k}}{\sum_{k'} z_{a,k'} t_{i,k'} v_{j,k'}}}{\sum_{a,j} z_{a,k} v_{j,k}} \tag{19}$$

$$v_{j,k} \leftarrow v_{j,k} \frac{\sum_{a,i} c_{a,i,j}^{(\mathrm{prop})} \frac{z_{a,k} t_{i,k}}{\sum_{k'} z_{a,k'} t_{i,k'} v_{j,k'}}}{\sum_{a,i} z_{a,k} t_{i,k}} \tag{20}$$

$\boldsymbol{Z}$ や $\boldsymbol{T}$ の更新後は，$\sum_a z_{a,k} = 1$ や $\sum_i t_{i,k} = 1$ を満たすように $\boldsymbol{z}_k$ と $\boldsymbol{t}_k$ のスケールをそれぞれ調整する．ただし，その逆数を適宜 $\boldsymbol{V}$ に乗じて，コスト関数の値がスケール調整によって変化しないようにする．

提案法の更新式の単調非増加性は次定理で示される．

Theorem 1. 更新式 (11)，(18)–(20) による更新で式 (8) は単調非増加となる．

Proof. 上界最小化アルゴリズムに基づく更新式 (18)–(20) は式 (8) の単調非増加性を保証する．従って，更新式 (11)，(18)–(20) の単調非増加性は式 (11) が単調非増加性を有するか否かに依存する．$b_k^{(\mathrm{old})}$ 及び $b_k^{(\mathrm{new})}$ をそれぞれ更新前後のアトラクタ基底ベクトルのインデクスとおく．b_k は式 (11) の通り $\sum_a \mathcal{D}(p_{a,b_k}|z_{a,k})$ が最小となるように更新されるため，次が成り立つ．

$$\begin{aligned} \mathcal{R}(\boldsymbol{p}_{b_k^{(\mathrm{old})}}|\boldsymbol{z}_k) &= \sum_a \mathcal{D}(p_{a,b_k^{(\mathrm{old})}}|z_{a,k}) \\ &\geq \sum_a \mathcal{D}(p_{a,b_k^{(\mathrm{new})}}|z_{a,k}) \\ &= \mathcal{R}(\boldsymbol{p}_{b_k^{(\mathrm{new})}}|\boldsymbol{z}_k) \quad \forall k \end{aligned} \tag{21}$$

従って，更新式 (11) は式 (8) の値を増加させない．□

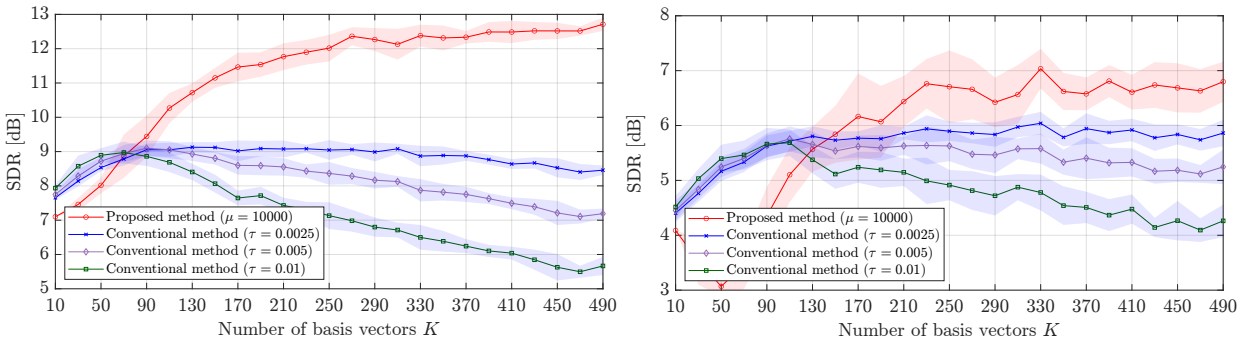


Fig. 4　SDR scores with various K: (a) $T_{60} = 0$ ms and (b) $T_{60} = 256$ ms.

4　実験

4.1　実験条件

従来法及び提案法のスポットフォーミングの性能を比較した．Pyroomacoustics [7] を用いて Fig. 3 に示される部屋に 2 次元鏡像法を用いて $T_{60} = 0$ 及び $T_{60} = 256$ ms となるように調整後，観測信号を生成した．ドライソースには LibriTTS [8] からランダムに抽出した Table 1 の音源を正規化して用いた．

各マイクロホンアレイの観測信号 $\mathsf{X}^{(a)}$ に対して最小分散無歪 (minimum-variance distortionless response: MVDR) BF を適用し，BF 出力 $(\boldsymbol{Y}^{(a)})_{a=0}^{A-1}$ を得た．MVDR BF で用いる目的音源のステアリングベクトル及びノイズの分散共分散行列には，各インパルス応答から計算できる理想的な値を用いた．また，全マイクロホンは完全同期しているものとした．

評価指標には目的音源の信号対歪み比 (source-to-distortion ratio: SDR) [9] を用いた．NMF 及び NTF は初期値に依存して性能が変化するため，10 種類の乱数シードを用いた．その他の条件を Table 2 に示す．

4.2　実験結果

実験結果を Fig. 4 に示す．折れ線グラフは SDR 平均値を，影は SDR の標準偏差を表している．BF 出力 $\boldsymbol{Y}^{(0)}$，$\boldsymbol{Y}^{(1)}$，及び $\boldsymbol{Y}^{(2)}$ の SDR 平均値は $T_{60} = 0$ ms では 7.1 dB，$T_{60} = 256$ ms では 4.1 dB であった．従来法のハイパーパラメータ τ は 12 種類とし，その中で 3 つの良い性能を示した結果を表している．これらの結果より，提案法は従来法に比べて残響やマイクロホンアレイ数によらずに高い性能を示した．さらに，従来法はいくつかの τ において基底数 K が増加するにつれて性能が低下したが，提案法はより高い性能を示した．

Fig. 4 では $\mu = 10000$ の結果を示したが，提案法は μ を十分大きく（例えば $\mu \geq 100$ 等）すると，一貫して安定した良い性能を示すことを確認している．μ を大きくすることで式 (9) はより強力な正則化となり，各 $\boldsymbol{z}_k$ はアトラクタベクトル $(\boldsymbol{p}_b)_{b=0}^{B-1}$ のいずれかと一致することとなる．これは，基底ベクトル $(\boldsymbol{t}_k)_{k=0}^{K-1}$ をハードクラスタリングすることに相当するが，そのような割り当てが効果的であることが分かる．以上より，提案法のハイパーパラメータ μ 及び K は大きな値にしておくことで頑健なスポットフォーミングが達成できる．

5　おわりに

本研究では，基底ベクトルを自動的に分配する正則化を用いた NTF によるスポットフォーミングを提案した．分配行列をより識別的な結果に誘導する正則化を導入し，NTF の基底ベクトルを目的音源成分及び干渉音源成分に自動的に振り分けるように設計した．実験結果より，提案法はハイパーパラメータの設定に頑健かつ高性能なスポットフォーミングを達成した．

参考文献

[1] M. Taseska and E. A. P. Habets, “Spotforming: spatial filtering with distributed arrays for position-selective sound acquisition,” *IEEE/ACM Trans. ASLP*, vol. 24, no. 7, pp. 1291–1304, 2016.

[2] K. Sekiguchi et al., “Layout optimization of cooperative distributed microphone arrays based on estimation of source separation performance,” *J. Robotics and Mechatronics*, vol. 29, no. 1, pp.83–93, 2017.

[3] Y. Kagimoto et al., “Spotforming by NMF using multiple microphone arrays,” *Proc. IROS*, pp. 9253–9258, 2022.

[4] 綾野翔馬ら, “非負値テンソル因子分解に基づく分散マイクロホンアレイを用いたスポットフォーミング,” **日本音響学会** *2024* **年春季研究発表会講演論文集**, pp. 137–140, 2024.

[5] S. Ayano et al., “Audio spotforming using nonnegative tensor factorization with attractor-based regularization,” *Proc. EUSIPCO*, 2024 (in press).

[6] D. R. Hunter and K. Lange, “Quantile regression via an MM algorithm,” *J. Comput. Graph. Stat.*, vol. 9, no. 1, pp. 60–77, 2000.

[7] R. Scheibler et al., “Pyroomacoustics: a Python package for audio room simulation and array processing algorithms,” *Proc. ICASSP*, pp. 351–355, 2018.

[8] H. Zen et al., “LibriTTS: a corpus derived from librispeech for text-to-speech,” *Proc. INTERSPEECH*, pp. 1526–1530, 2019.

[9] E. Vincent et al., “Performance measurement in blind audio source separation,” *IEEE Trans. ASLP*, vol. 14, no. 4, pp. 1462–1469, 2006.