

ディリクレ分布に基づく非負値テンソル因子分解の ステレオ音楽分離への応用*

©小川遼, 北村大地 (香川高専)

1 はじめに

現在, 流通している一般的な音楽は, 2 チャンネルのステレオ信号で構成されることが多い. ステレオ信号では, ボーカルや各楽器音といった個々の音源に対し, ステレオエフェクト等を通して左右のチャンネル間の波形の違いを持たせ, 特定の方位から聞こえる感覚や空間的な広がりリスナーに提供している. 特に, Fig. 1 のように左右の音量比のみを用いて音源の方位を操作する技術はパニングと呼ばれ最も単純かつ基本的なステレオエフェクトである.

ステレオ音源を対象とした音源分離を考える際, パニングによって付与された左右チャンネル間の音量比は, 各音源を区別するための重要な特徴量となる. 近年, 音源分離では深層学習 (deep neural network: DNN) を用いた手法 (例えば [1, 2] 等) が主流となっている. 例えば, モノラル信号を対象とした音楽音源分離として Demucs [1] や Band-Split RNN [2] が有名なモデルである. また, ステレオ信号の情報を積極的に音楽音源分離に利用するモデルとして Hybrid Transformer Demucs [3] や Spain-Net [4] が提案されている. これらの DNN の学習には大規模かつ高品質なデータセットが必要となる. この課題に対するアプローチの1つとして, あらかじめ教師無し手法等で大まかに分離した仮の分離信号を補助情報として DNN に入力する方法が考えられる. DNN 内部で補助情報をうまく活用できれば, 要求されるデータセットの品質や規模の緩和が期待できる. 以上から, 教師無し手法での分離技術も依然として重要であり, その性能向上は DNN への応用可能性を高めることにつながる.

本稿では, ステレオ信号の最も単純な分離として, パニングで作成されたステレオ音楽信号の方位分離について検討する. 具体的には, ステレオ音楽信号の振幅スペクトログラムを非負値テンソル因子分解 (nonnegative tensor factorization: NTF) [6, 7] で近似するアプローチを考え, 各方位の音源のスペクトルパターンの識別的な学習を誘導する正則化を提案する. また, 提案手法が振幅比に基づく方位クラスタリングよりも高い分離性能を発揮することを示す.

2 要素技術とディリクレ NMF

2.1 パニング

パニングとは, Fig. 1 に示すように, モノラル音源に対し, 左右の 2 チャンネル間で音量比を調整することで, 聴覚上の方位を制御する技術である. 本稿で用いるサイン則 [5] によるパニングでは, モノラル信号を $s(n)$ としたとき, 左右のチャンネル信号 $y_L(n)$ 及び $y_R(n)$ を $y_L(n) = g_L \times s(n)$ 及び $y_R(n) = g_R \times s(n)$ と定義する. ここで, 係数 g_L 及び g_R は, 音源の聴感上の方位角 θ (Fig. 1 のように正面 0° からの角度)

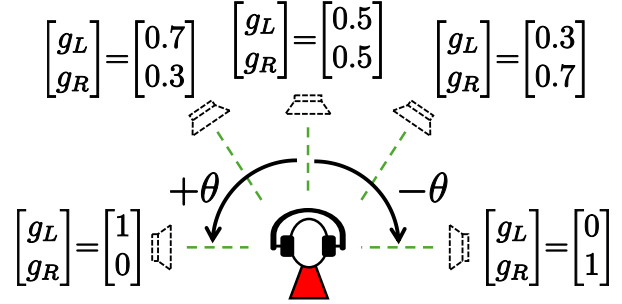


Fig. 1 Stereo panning based on sine law.

から次式で計算される.

$$g_L = \frac{1}{2} \times \left(\frac{\sin(\theta)}{\sin(\tilde{\theta})} + 1 \right), \quad g_R = 1 - g_L \quad (1)$$

ここで, $\tilde{\theta}$ はスピーカの配置角度 (角度の定義は θ と同様) である.

2.2 NTF

NTF は, 非負値行列因子分解 (nonnegative matrix factorization: NMF) [9] をテンソルに拡張した手法である. NTF では, 全要素が非負の 3 階テンソル $\mathbf{X} \in \mathbb{R}_{\geq 0}^{I \times J \times M}$ を, 3 つの非負因子行列 $\mathbf{Z} = [\mathbf{z}_1 \ \mathbf{z}_2 \ \cdots \ \mathbf{z}_K] \in \mathbb{R}_{\geq 0}^{M \times K}$, $\mathbf{T} = [\mathbf{t}_1 \ \mathbf{t}_2 \ \cdots \ \mathbf{t}_K] \in \mathbb{R}_{\geq 0}^{I \times K}$, 及び $\mathbf{V} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_K]^T \in \mathbb{R}_{\geq 0}^{K \times J}$ を用いて以下のようにモデル化する.

$$x_{ijm} \approx \hat{x}_{ijm} = \sum_k z_{mk} t_{ik} v_{kj} \quad \forall i, j, m \quad (2)$$

ここで, x_{ijm} , z_{mk} , t_{ik} , 及び v_{kj} はそれぞれ $\mathbf{X}$, $\mathbf{Z}$, $\mathbf{T}$, 及び $\mathbf{V}$ の要素であり, $\hat{x}_{ijm}$ をまとめたテンソルを $\hat{\mathbf{X}} \in \mathbb{R}_{\geq 0}^{I \times J \times M}$ と定義する. また, $i \in \{1, 2, \dots, I\}$, $j \in \{1, 2, \dots, J\}$, $m \in \{1, 2, \dots, M\}$, 及び $k \in \{1, 2, \dots, K\}$ は各次元のインデックスであり, K は基底数である. さらに, $\cdot^T$ は転置を表す. これらの因子行列の推定は観測テンソル $\mathbf{X}$ と近似テンソル $\hat{\mathbf{X}}$ 間の乖離度を測るコスト関数の最小化問題となる.

$$\underset{\mathbf{Z}, \mathbf{T}, \mathbf{V}}{\text{Minimize}} \ \mathcal{D}(\mathbf{X} | \hat{\mathbf{X}}) \text{ s.t. } z_{mk}, t_{ik}, v_{kj} \geq 0 \ \forall i, j, m, k \quad (3)$$

本稿では, 乖離度関数 $\mathcal{D}(\mathbf{X} | \hat{\mathbf{X}})$ に次式の一般化 Kullback-Leibler (KL) 擬距離を用いる.

$$\mathcal{D}(\mathbf{X} | \hat{\mathbf{X}}) = \sum_{i,j,m} \left(x_{ijm} \log \frac{x_{ijm}}{\hat{x}_{ijm}} - x_{ijm} + \hat{x}_{ijm} \right) \quad (4)$$

2.3 ディリクレ分布と正則化への応用

ディリクレ分布は, 非負制約とノルム制約を満たす確率変数ベクトルに対する確率密度関数である. い

* Application of Dirichlet-distribution-based nonnegative tensor factorization to stereo music source separation. By Haru OGAWA and Daichi KITAMURA (NIT, Kagawa).

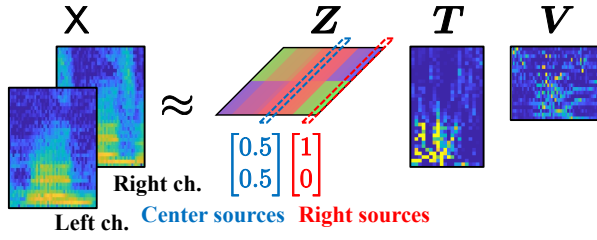


Fig. 2 NTF decomposition of a stereo spectrogram into channel-partitioning weights $\mathbf{Z}$, spectral bases $\mathbf{T}$, and temporal activations $\mathbf{V}$.

ま、ディリクレ分布に従う M 次元確率変数ベクトルを $\mathbf{y} = [y_1, \dots, y_M]^T \in \mathbb{R}_{\geq 0}^M$ と定義する。但し、 $\mathbf{y}$ は $\sum_m y_m = 1$ を満たす。

$$\mathbf{y} \sim p(\mathbf{y}; \boldsymbol{\alpha}) = \frac{1}{B(\boldsymbol{\alpha})} \prod_m y_m^{\alpha_m - 1} \quad (5)$$

ここで、 $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_M]^T \in \mathbb{R}_{> 0}^M$ は母数ベクトル、 $B(\boldsymbol{\alpha})$ は多変量ベータ関数である。

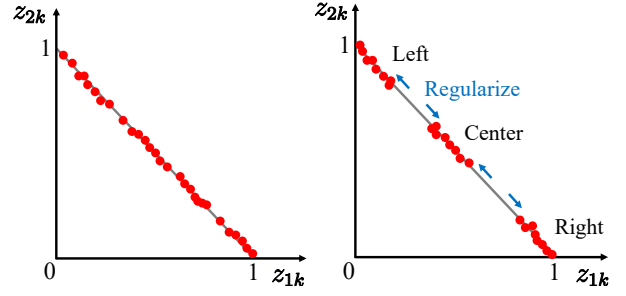
母数 $\boldsymbol{\alpha}$ の全要素が同じ値 α となる対称ディリクレ分布を考えた場合、 $\alpha < 1$ 及び $\alpha > 1$ において、それぞれ $\mathbf{y}$ は one-hot ベクトル及び dense ベクトル（全要素が同じ値を均等に含むベクトル）となる確率が高くなる。この性質を用いた NMF の変数の正則化としてディリクレ NMF [11] が提案されている。本稿では、ディリクレ NMF のさらなる拡張として、NTF の変数 $\mathbf{z}_k$ にディリクレ分布に基づく正則化を導入し、ステレオ音楽分離に活用する手法を提案する。

3 提案手法

3.1 提案手法の原理

提案手法では、複数の音源が異なる方位にパニングで配置されたステレオ音楽信号の分離を目的とする。まず、Fig. 2 に示すようにステレオの観測信号の振幅スペクトログラムをテンソル $\mathbf{X} \in \mathbb{R}_{\geq 0}^{I \times J \times 2}$ と定義する（すなわち、 $M = 2$ となる）。このテンソル $\mathbf{X}$ を NTF で式 (2) のように近似分解するとき、分配行列 $\mathbf{Z} \in \mathbb{R}_{\geq 0}^{2 \times K}$ の列ベクトル $\mathbf{z}_k = [z_{1k}, z_{2k}]^T$ は、 k 番目の基底（ランク 1 スペクトログラム $\mathbf{t}_k \mathbf{v}_k^T$ ）が分配される方位の情報を 2 次元ベクトルとして含む。例えば、dense ベクトル $\mathbf{z}_k = [0.5, 0.5]^T$ に対応するランク 1 スペクトログラム成分 $\mathbf{t}_k \mathbf{v}_k^T$ は左チャネルと右チャネルに等しい振幅 0.5 で分配されるため中央に定位する音源の成分となり、one-hot ベクトル $\mathbf{z}_k = [1, 0]^T$ 又は $\mathbf{z}_k = [0, 1]^T$ に対応する $\mathbf{t}_k \mathbf{v}_k^T$ は、それぞれ左チャネル又は右チャネルに定位する音源の成分となる。実際には $\mathbf{z}_k$ は前述の具体例のように明確な one-hot 又は dense ベクトルとはならず適当な値を含む形で推定されるが、それでもステレオ観測信号 $\mathbf{X}$ に含まれる音源がパニングによって明確な方位の違いを持つ場合、NTF 変数の推定後に $\mathbf{z}_1, \dots, \mathbf{z}_K$ を音源数分でクラスタリングすれば、各ランク 1 スペクトログラム成分 $\mathbf{t}_k \mathbf{v}_k^T$ がどの音源に属するかを分類することができ、教師無し音源分離（方位分離）が可能となる。

しかしながら、通常の NTF では、各変数間でスケール任意性が存在するため、 $\mathbf{z}_k$ の $z_{1k} + z_{2k} = 1$ なるノルム制約が保証されない。そこで本稿では、基底ベクトル $\mathbf{z}_k$ の事前分布にディリクレ分布を仮定する。2.3 節で述べた通り、ディリクレ分布に従う確率変数



(a) $\mathbf{z}_k$ without regularization (b) $\mathbf{z}_k$ with regularization

Fig. 3 Estimated channel-partitioning weights $\mathbf{z}_1, \dots, \mathbf{z}_K$ (a) without regularization and (b) with Dirichlet-distribution-based regularization.

ベクトルは非負性とノルム制約が保証される。さらに、ディリクレ事前分布の母数 $\boldsymbol{\alpha}$ をハイパーパラメタとして与えることで、NTF における $\mathbf{z}_k$ の推定を one-hot ベクトルや dense ベクトルに誘導することができる。その結果、Fig. 3 のように $\mathbf{z}_1, \dots, \mathbf{z}_K$ が明確なクラスタ構造を持つように正則化され、方位分離の性能向上が期待される。以後、この提案手法をディリクレ NTF と呼ぶ。

3.2 ディリクレ NTF の導出

K 個のベクトル $\mathbf{z}_1, \dots, \mathbf{z}_K$ が独立にディリクレ分布 $p(\mathbf{z}; \boldsymbol{\alpha}_k)$ に従うことを仮定する。

$$\mathbf{z}_k \sim p(\mathbf{z}; \boldsymbol{\alpha}_k) \quad \forall k \quad (6)$$

$$\mathbf{Z} \sim p(\mathbf{Z}) = \prod_k p(\mathbf{z}_k; \boldsymbol{\alpha}_k) |_{\mathbf{z}=\mathbf{z}_k} \quad (7)$$

ディリクレ事前分布の負対数尤度関数から得られる正則化項 $\mathcal{R}(\mathbf{Z})$ は次式となる。

$$\begin{aligned} \mathcal{R}(\mathbf{Z}) &= -\log p(\mathbf{Z}) \\ &\stackrel{c}{=} \sum_{m,k} (\alpha_{mk} - 1) \log z_{mk}^{-1} \end{aligned} \quad (8)$$

ここで、 α_{mk} は $\boldsymbol{\alpha}_k$ の要素であり、 $\stackrel{c}{=}$ は定数項の違いを除いて等しいことを表す。従って、ディリクレ NTF は次のように定式化される。

$$\begin{aligned} \underset{\mathbf{Z}, \mathbf{T}, \mathbf{V}}{\text{Minimize}} \quad & \mathcal{D}(\mathbf{X} | \hat{\mathbf{X}}) + \sum_{m,k} (\alpha_{mk} - 1) \log z_{mk}^{-1} \\ \text{s.t.} \quad & z_{mk}, t_{ik}, v_{kj} \geq 0 \quad \forall i, j, m, k, \\ & \sum_m z_{mk} = 1 \quad \forall k, \quad \sum_i t_{ik} = 1 \quad \forall k \end{aligned} \quad (9)$$

制約条件として、各変数の非負制約に加えて、 $\mathbf{z}_k$ のノルム制約及びスケール任意性の問題を回避するために、 $\mathbf{t}_k$ のノルム制約も追加している。以後、ディリクレ NMF の目的関数を $\mathcal{J} = \mathcal{D}(\mathbf{X} | \hat{\mathbf{X}}) + \mathcal{R}(\mathbf{Z})$ と表記する。

NMF や NTF の文脈では補助関数法に基づく乗算型反復更新式 [6] がよく導出される。この手法は、直接最適化が難しい目的関数に対して、(a) 現在の変数の地点で一定の要件を満たす補助関数を設計、(b) 補助関数の停留点を閉形式で求めて次の地点の変数として更新、の 2 ステップを反復することで、本来の目的関数を間接的に最小化する。本稿においても、最適化問題 (9) に対して同様のアプローチを適用する。

ただし、従来の NMF や NTF における補助関数法では、変数の非負制約を陽に扱うことなく、「乗算型の更新式を導出すること」及び「変数の初期値を正值とすること」の2点で暗に非負制約を満たすような変数更新を実現している。ディリクレ NTF では、非負制約に加えて z_k 及び t_k のノルム制約も課す必要があるため、前述の (a) ステップで設計した補助関数に対して、(b) ステップを等式・不等式制約条件付き最小化問題と捉え、ラグランジュの未定乗数法とカルーシュ・クーン・タッカー (Karush-Kuhn-Tucker: KKT) 条件を用いて解く。

まず、一般化 KL 擬距離 (4) にイエンゼンの不等式を適用し次式を得る。

$$\begin{aligned} \mathcal{D}(\mathbf{X}|\hat{\mathbf{X}}) &= \sum_{i,j,m} (-x_{ijm} \log \hat{x}_{ijm} + \hat{x}_{ijm}) \\ &\leq \sum_{i,j,m} \left(-x_{ijm} \sum_k \delta_{ijmk} \log \frac{z_{mk} t_{ik} v_{kj}}{\delta_{ijmk}} + \hat{x}_{ijm} \right) \end{aligned}$$

ここで、 $\delta_{ijmk} > 0$ はイエンゼンの不等式の補助変数であり $\sum_k \delta_{ijmk} = 1 \forall i, j, m$ を満たす。上式を用いて、目的関数全体の補助関数 $\mathcal{J}^+ \geq \mathcal{J}$ が次のように設計できる。

$$\begin{aligned} \mathcal{J}^+ &\stackrel{c}{=} \sum_{i,j,m} \left(-x_{ijm} \sum_k \delta_{ijmk} \log \frac{z_{mk} t_{ik} v_{kj}}{\delta_{ijmk}} + \hat{x}_{ijm} \right) \\ &\quad + \sum_{m,k} (\alpha_{mk} - 1) \log z_{mk}^{-1} \end{aligned}$$

ここで、 $\mathcal{J}^+ = \mathcal{J}$ を満たす補助変数は次式となる。

$$\delta_{ijmk} = \frac{z_{mk} t_{ik} v_{kj}}{\sum_{k'} z_{mk'} t_{ik'} v_{kj'}} \quad (10)$$

次に、補助関数 $\mathcal{J}^+$ に対して、次の等式・不等式制約条件付き最小化問題を考える。

$$\begin{aligned} \text{Minimize } \mathcal{J}^+ \text{ s.t. } z_{mk}, t_{ik}, v_{kj} &\geq 0 \forall i, j, m, k, \\ \sum_m z_{mk} &= 1 \forall k, \sum_i t_{ik} = 1 \forall k \end{aligned} \quad (11)$$

ここで、 Δ は δ_{ijmk} を全インデックスに関してまとめた集合である。ラグランジアンは次式となる。

$$\mathcal{L} = \mathcal{J}^+ + \sum_k \lambda_k \left(\sum_m z_{mk} - 1 \right) + \sum_k \mu_k \left(\sum_i t_{ik} - 1 \right)$$

ここで、 λ_k 及び μ_k はラグランジュ乗数を表す。

まず、 z_{mk} について考える。非負制約条件付き最適化問題の KKT 条件より $\mathcal{L}$ の z_{mk} に関する最小解は $z_{mk} (\partial \mathcal{L} / \partial z_{mk}) = 0 \forall m, k$ を満たすため、 $z_{mk} = 0$ 又は $\partial \mathcal{L} / \partial z_{mk} = 0$ のいずれかを満たす。今、 $\partial \mathcal{L} / \partial z_{mk} = 0$ を満たす z_{mk} を $\hat{z}_{mk}$ とおくと、 $z_{mk} \geq 0$ 及び $\partial \mathcal{L} / \partial z_{mk} \geq 0 \forall m, k$ の2条件より、 $\hat{z}_{mk} \geq 0$ ならば $\hat{z}_{mk}$ を最小解とし、 $\hat{z}_{mk} < 0$ ならば $z_{mk} = 0$ を最小解とすれば、最適化問題 (11) の z_{mk} に関する最小化ができる。即ち、 t_{ik} , v_{kj} , 及び δ_{ijmk} を固定した際の z_{mk} の制約条件付き最小解は次式と

なる。

$$z_{mk} = \begin{cases} \hat{z}_{mk} & (\text{if } \hat{z}_{mk} \geq 0) \\ 0 & (\text{otherwise}) \end{cases} \quad (12)$$

$\hat{z}_{mk}$ は $\partial \mathcal{L} / \partial z_{mk} = 0$ を整理して、次式のように得られる。

$$\hat{z}_{mk} = \frac{\sum_{i,j} x_{ijm} \delta_{ijmk} + \alpha_{mk} - 1}{\sum_{i,j} t_{ik} v_{kj} + \lambda_k} \quad (13)$$

式 (12) を考慮すると、式 (13) で求まる $\hat{z}_{mk}$ が非負になる要素のみ着目すればよい。ため、 $\hat{z}_{mk} \geq 0$ を満たすインデックス m の集合を k 毎に $\mathcal{M}_k$ とおく。また、 $\partial \mathcal{L} / \partial \lambda_k = 0 \forall k$ より $\sum_m z_{mk} - 1 = 0$ を得るが、これはノルム制約と等価であり、次式に書き換えられる。

$$\sum_m z_{mk} = \sum_{m \in \mathcal{M}_k} \hat{z}_{mk} = 1 \forall k \quad (14)$$

式 (13) を式 (14) に代入して整理すると

$$\sum_{i,j} t_{ik} v_{kj} + \lambda_k = \sum_{m \in \mathcal{M}_k} \left(\sum_{i,j} x_{ijm} \delta_{ijmk} + \alpha_{mk} - 1 \right) \quad (15)$$

が得られる。式 (15) を式 (13) の分母に代入してラグランジュ乗数 λ_k を消去すると、次式となる。

$$\hat{z}_{mk} = \frac{\sum_{i,j} x_{ijm} \delta_{ijmk} + \alpha_{mk} - 1}{\sum_{m \in \mathcal{M}_k} \left(\sum_{i,j} x_{ijm} \delta_{ijmk} + \alpha_{mk} - 1 \right)} \quad \forall m \in \mathcal{M}_k, \forall k \quad (16)$$

以上より、補助関数 $\mathcal{J}^+$ の制約条件付き最小解は式 (12) 及び (16) である。なお、補助変数の等号成立条件 (10) を式 (16) に代入すると、 $\hat{z}_{mk}$ は次式となる。

$$\hat{z}_{mk} = \frac{z_{mk} \sum_{i,j} \frac{x_{ijm}}{\hat{x}_{ijm}} t_{ik} v_{kj} + \alpha_{mk} - 1}{\sum_{m \in \mathcal{M}_k} \left(z_{mk} \sum_{i,j} \frac{x_{ijm}}{\hat{x}_{ijm}} t_{ik} v_{kj} + \alpha_{mk} - 1 \right)} \quad (17)$$

同様の手順で T 及び V について解くと

$$t_{ik} \leftarrow \frac{t_{ik} \sum_{j,m} \frac{x_{ijm}}{\hat{x}_{ijm}} z_{mk} v_{kj}}{\sum_i \left(t_{ik} \sum_{j,m} \frac{x_{ijm}}{\hat{x}_{ijm}} z_{mk} v_{kj} \right)} \quad (18)$$

$$v_{kj} \leftarrow \frac{v_{kj} \sum_{i,m} \frac{x_{ijm}}{\hat{x}_{ijm}} z_{mk} t_{ik}}{\sum_{i,m} z_{mk} t_{ik}} \quad (19)$$

を得る。以上より、NTF 変数の最適化は式 (12) 及び (17)–(19) となる。これらは補助関数法に基づく反復更新式であるため、反復更新におけるコスト関数値の単調非増加性を満たす。

3.3 Z のクラスタリングと時間周波数マスク

推定された $z_1, \dots, z_K$ に対して、クラスタ数 (方位数) を D する k 平均法を適用する。その後、 d 番目のクラスタに属するインデックス k の集合を $\mathcal{K}_d$ と定義すれば、 $\mathbf{X}$ から d 番目の方位を抽出するウィーナフィルタの適用は次式となる。

$$c_{ijmn} = \frac{\left(\sum_{k \in \mathcal{K}_d} z_{mk} t_{ik} v_{kj} \right)^2}{\left(\sum_k z_{mk} t_{ik} v_{kj} \right)^2} x_{ijm} \quad (20)$$

ここで、 c_{ijmn} 及び x_{ijm} はそれぞれ推定方位信号及び観測信号の複素スペクトログラムの要素である。

4 パニング音の分離実験

4.1 実験条件

本稿では、左右チャンネルの振幅比のみを用いる単純な方位クラスタリング (directional clustering: DC) [12] と提案手法の方位分離性能を比較した。分離対象の音源には信号長 17.9 s のベース、ピアノ、及びギターの合成音を用いた。これらのモノラル音源に対し、式 (1) によるパニング処理を施したステレオ観測信号を作成した。各音源の定位は、それぞれ $\theta = 65^\circ$, 0° , 及び -60° に設定した。またスピーカの配置は $\bar{\theta} = 90^\circ$ とした。従って、各方位に 1 つの音源が存在するため、本実験条件では方位を分離することが音源分離となる。また、観測信号に適用する短時間フーリエ変換は、窓長 92.9 ms, シフト長 46.4 ms, 及び窓関数 Blackman 窓とした。

提案手法の基底数は $K = 150$ としこの 150 個の基底を 2 グループに分け、 $K_1 = \{1, 2, \dots, N\}$ 及び $K_2 = \{N + 1, \dots, 150\}$ とした。前半の K_1 個の基底に対応する z_k には母数 $\alpha_1 = [\alpha_1, \dots, \alpha_1]^T$ を与え、後半の K_2 個の基底に対応する z_k には母数 $\alpha_2 = [\alpha_2, \dots, \alpha_2]^T$ を与えた。事前に行ったパラメタ探索に基づき、本稿では代表的なパラメタ設定の組み合わせについての結果を示す。

音源分離の評価指標には、`bss_eval_images` [13] で計算される多チャンネルの音源対歪み比 (source-to-distortion ratio: SDR) を採用した。この SDR 値は、分離された信号に含まれる目的音源の品質・非目的音源の抑圧度合いの 2 点に加え、観測信号中の目的音源の方位 (音量比) が分離信号で維持されたかについても加味される評価値である。

4.2 実験結果

比較手法の DC は、時間周波数マスクをハード (バイナリ) 及びソフトとする 2 条件で実施した。DC と提案手法のいずれも初期値を必要とするため、本実験では異なる 30 個の乱数シードで乱数初期値を与えて試行した。

Fig. 4 は、3 つの音源で平均した SDR 値の 30 回の試行結果のバイオリン図を示している。各手法の最適条件下での平均 SDR の平均値は、ハードマスクの DC が 5.75 dB, ソフトマスクの DC が 6.76 dB, ディリクレ NTF が 11.88 dB となり、方位分離における提案手法の有効性が確認された。また提案手法において、最良のパラメタは α_1 及び α_2 どちらも dense ベクトルに誘導するパラメタ ($\alpha_1, \alpha_2 > 1$) であった。これは、本実験で用いた観測信号の各音源の定位は、比較的 z_k が dense ベクトルとなるように配置されていることに起因すると推測される。しかしながら、one-hot ベクトル及び dense ベクトルの両方にそれぞれ誘導するパラメタ ($\alpha_1 > 1$ かつ $\alpha_2 < 1$) を与えた際も、平均 SDR の平均値は 10.75 dB となり良い性能を見せた。最適なパラメタの決定方法については今後の課題となる。

5 おわりに

本稿では、ディリクレ分布に基づく正則化を導入した NTF を提案し、その応用としてステレオ音楽の方位分離の実験を行った。ディリクレ分布による正則化は one-hot ベクトルや dense ベクトルに誘導することができるため、基底の方位に対応するベクトルに導

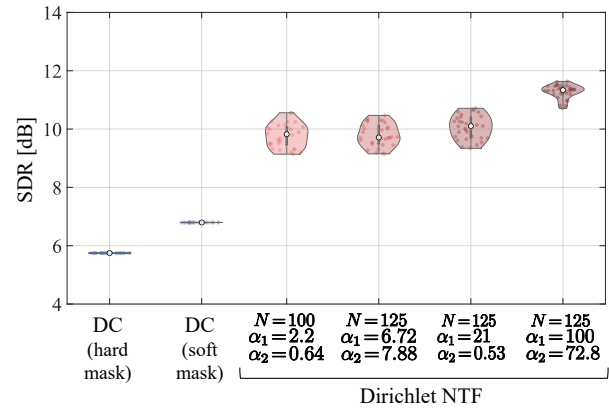


Fig. 4 Violin plots of SDR values. The white circle indicates median value, the gray vertical line shows the 25–75 percentile range, and the violin curve is an estimated distribution.

入し、性能向上を確認した。

謝辞 本研究の一部は科研費 23K24908 の助成を受けた。

参考文献

- [1] A. Défossez, N. Usunier, L. Bottou, and F. Bach, “Music source separation in the waveform domain,” *arXiv preprint*, 1911.13254, 2019.
- [2] Y. Luo and J. Yu, “Music source separation with band-split RNN,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 31, pp. 1893–1901, 2023.
- [3] S. Rouard, F. Massa, and A. Défossez, “Hybrid transformers for music source separation,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. 1–5, 2023.
- [4] D. Petermann and M. Kim, “Spain-Net: spatially-informed stereophonic music source separation,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. 106–110, 2022.
- [5] 安藤彰男, “音響の高臨場感技術,” *映像情報メディア学会誌*, vol. 66, no. 8, pp. 671–678, 2012.
- [6] A. Shashua and T. Hazan, “Non-negative tensor factorization with applications to statistics and computer vision,” in *Proc. Int. Conf. Mach. Learn.* pp. 792–799, 2005.
- [7] A. Cichocki, R. Zdunek, S. Choi, R. Plemmons, and S.-I. Amari, “Non-negative tensor factorization using alpha and beta divergences,” in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. III-1393–III-1396, 2007.
- [8] S. Araki, H. Sawada, R. Mukai, and S. Makino, “Under-determined blind sparse source separation for arbitrarily arranged multiple sensors,” *Signal Process.*, vol. 87, no. 8, pp. 1833–1847, 2007.
- [9] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” in *Proc. Adv. Neural Info. Process. Syst.*, 2000.
- [10] T. G. Kolda and B. W. Bader, “Tensor decompositions and applications,” *SIAM Rev.*, vol. 51, no. 3, pp. 455–500, 2009.
- [11] 小川遼, 北村大地, 綾野翔馬, “ディリクレ分布に基づく正則化付き非負値行列因子分解,” *日本音響学会 2025 年春季研究発表会講演論文集*, pp. 225–228, 2025.
- [12] 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. Audio Speech Lang. Process.*, vol. 23, no. 4, pp. 654–669, 2015.
- [13] E. Vincent, R. Gribonval, and C. Févotte, “Performance measurement in blind audio source separation,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 14, no. 4, pp. 1462–1469, 2006.