

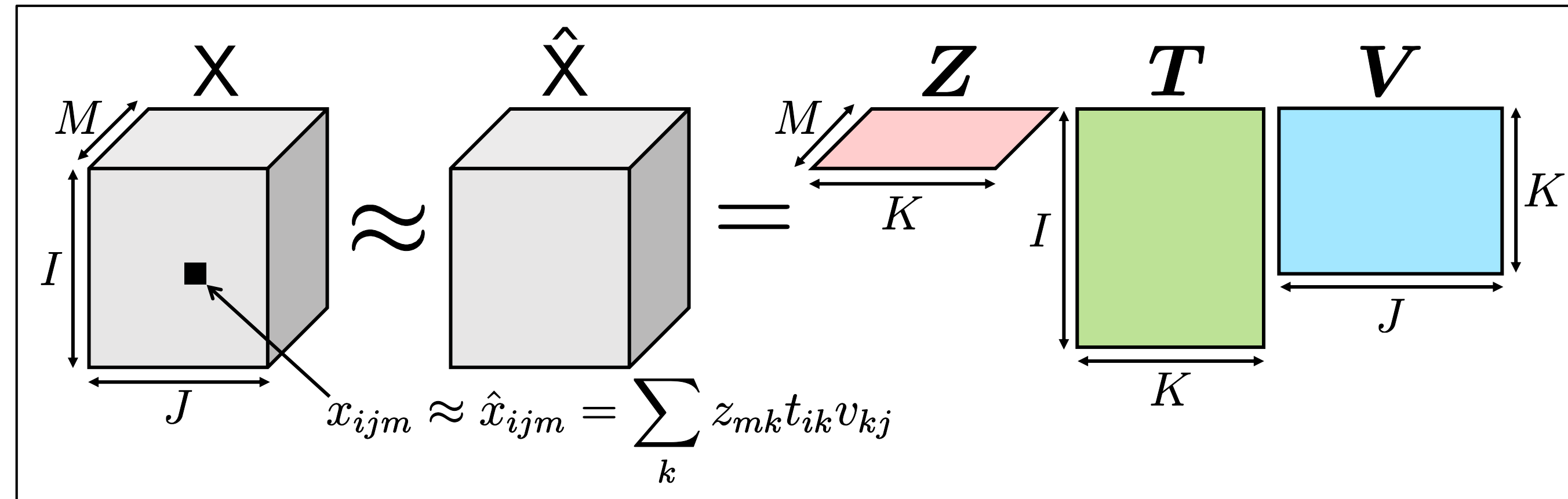
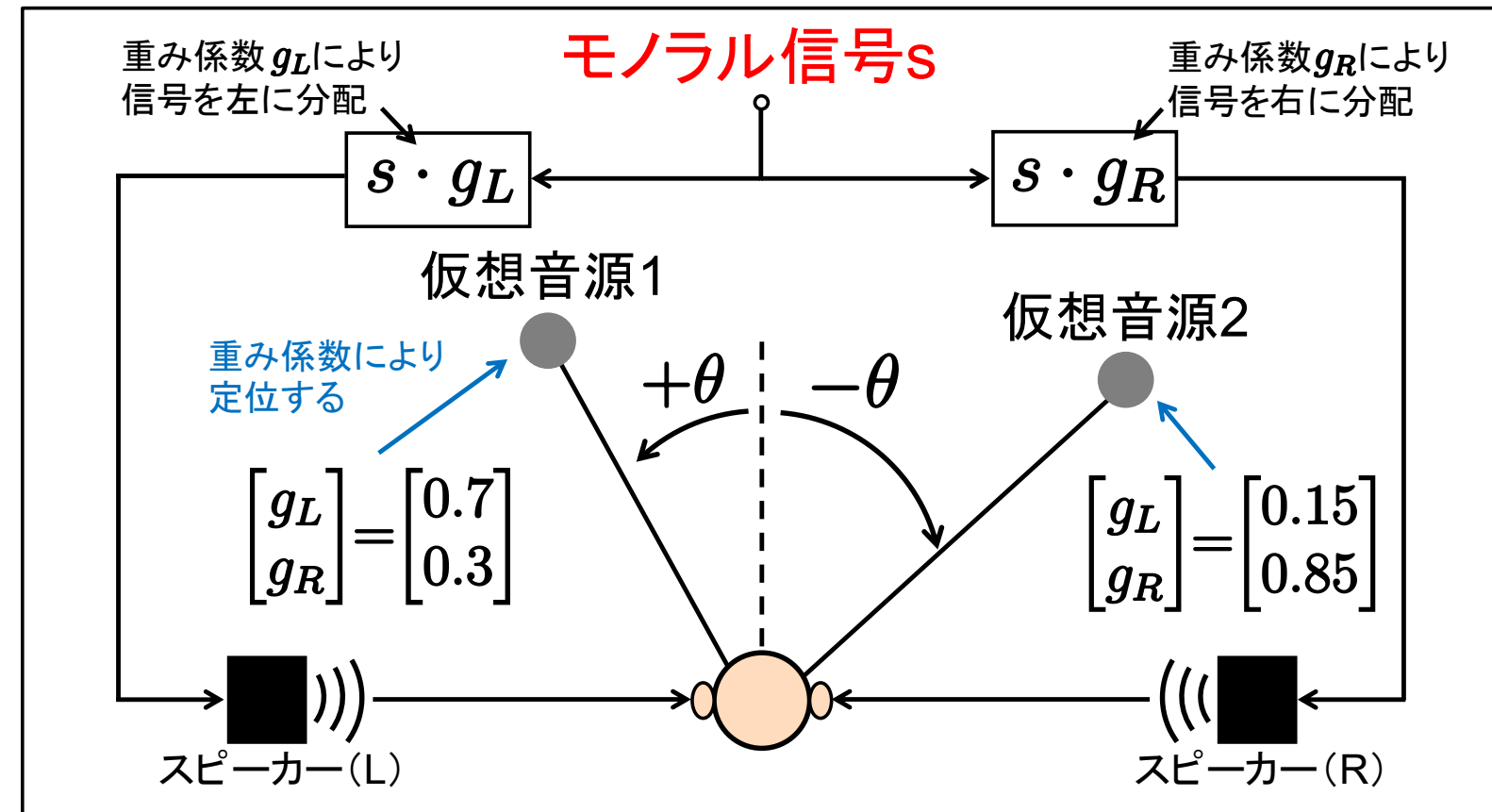
本発表の概要

- 背景: 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
- 目的: ステレオ音楽を方位ごとに分離しその結果を**DNN学習の補助情報**として活用する
- 提案: 非負値テンソル因子分解で信号を近似し**方位ごとのスペクトルパターン学習を誘導する正則化を導入**する
- 結果: 提案法は振幅比に基づくクラスタリングより高い分離性能を示した

1. 研究背景

- パニング(panning)**
 - モノラル音源に対し, 左右の2チャンネル間の音量比で聴覚上の方位を制御する技術
 - サイン則[安藤, 2012]による導出
$$g_L = \frac{1}{2} \times (\sin(\theta) + 1)$$

$$g_R = 1 - g_L$$
- 非負値テンソル因子分解(nonnegative tensor factorization: NTF)** [Shashua+, 2005]
 - 観測した非負テンソルを3つの非負因子行列でモデル化する手法



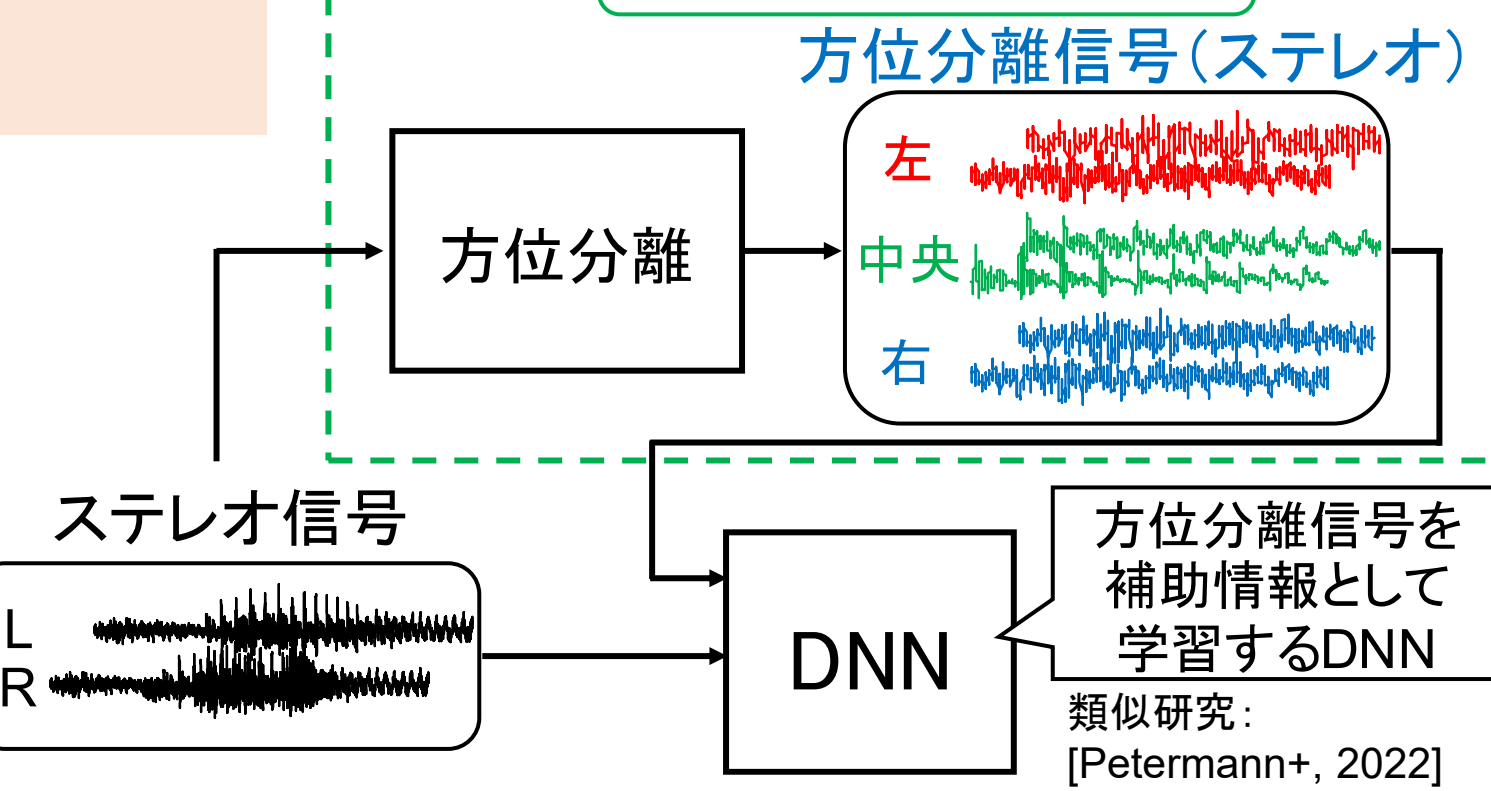
目的

ステレオ音楽信号の方位分離の実現と将来的なDNN学習支援への発展

動機

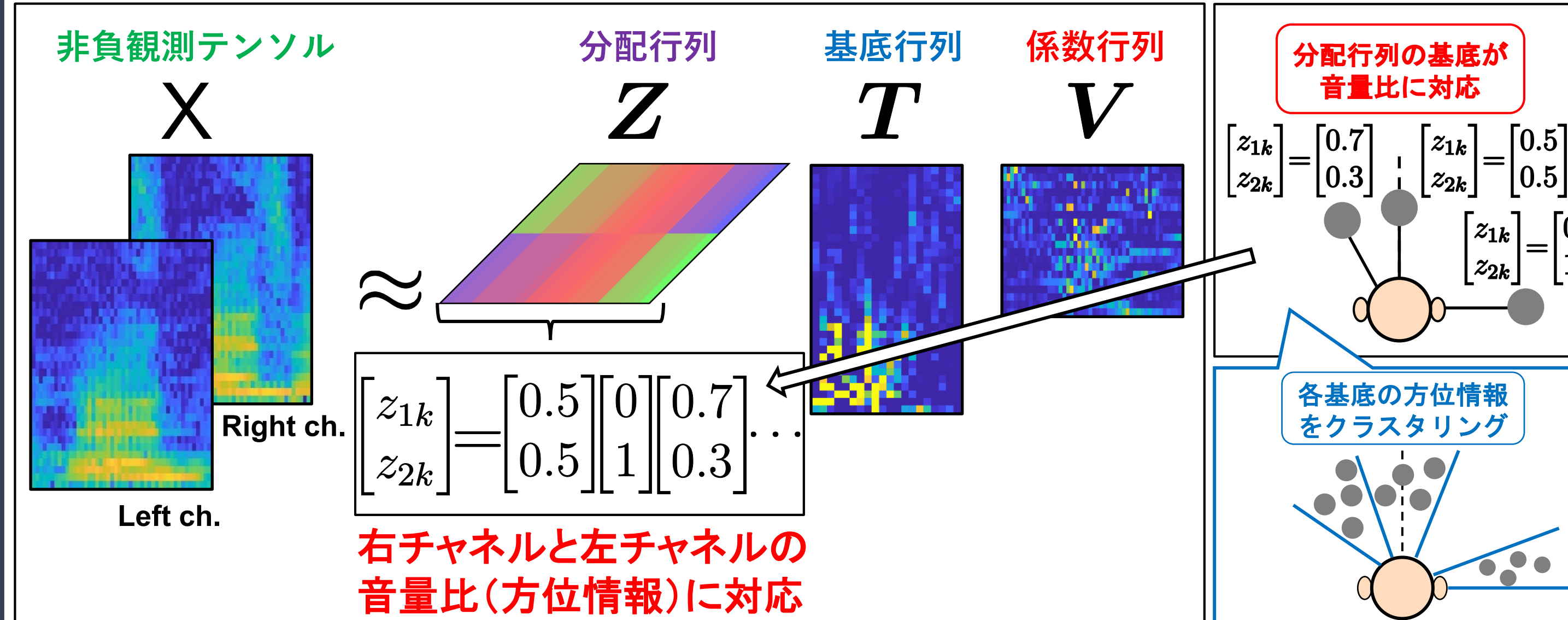
方位分離信号をDNN学習の補助として利用
大雑把な分離でもDNNの学習性能を高めるうえでの少量データ・パラメータで性能が上がるなどの有意義な役割を果たすことを期待

本発表のスコープ



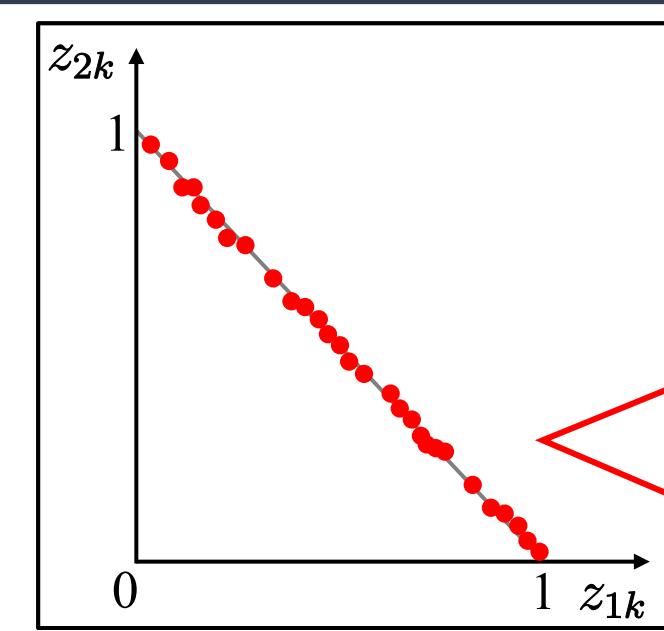
2. 提案手法

- NTFによるパニングされたステレオ信号の方位分離**
 - ステレオ信号を観測テンソルXとすると分配行列Zの基底が方位情報となる



- 得られた分配行列Zを基底ごとに**クラスタリング**することで方位分離を実現

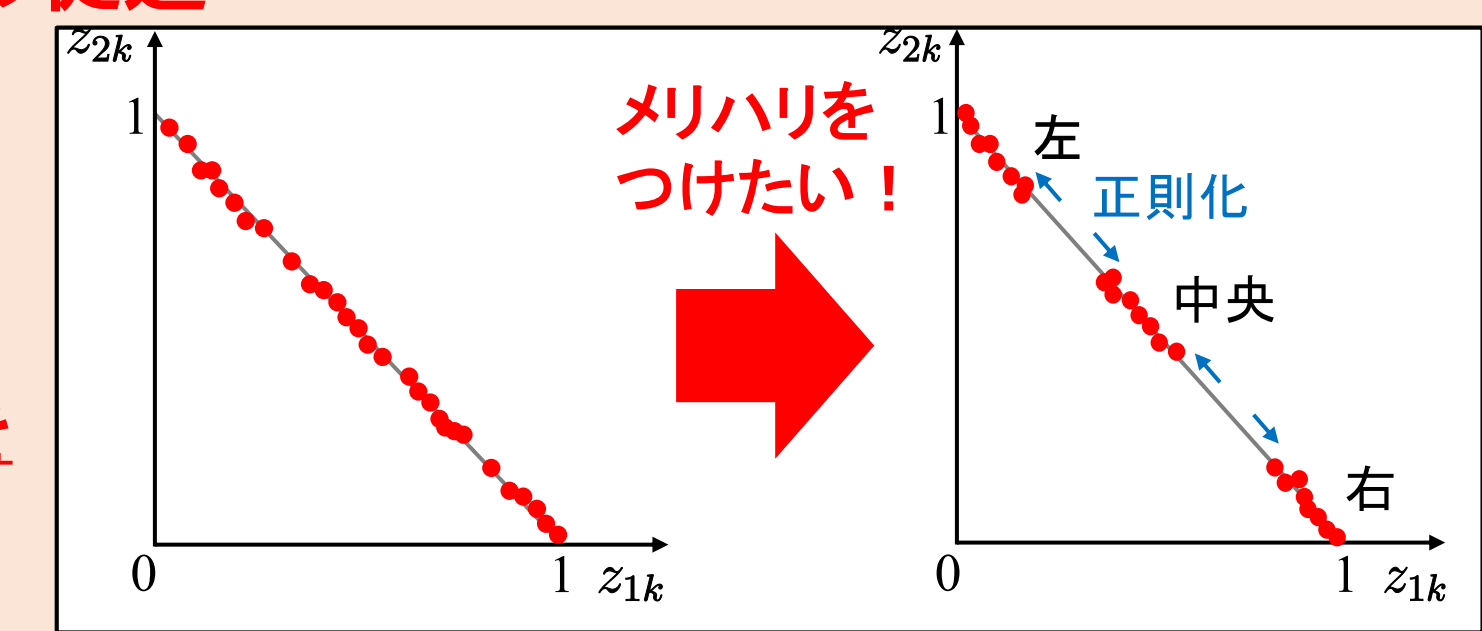
- 実際NTFで方位分離は可能だが, 分配行列Zの**基底ベクトルの値はばらつきが大きい**
- ばらつきにより**クラスタリング誤差**が生じ, **分離精度が低下**する



- 課題①: **ばらつきが大きく方位分離の精度が悪い**
- 課題②: 通常のNTFの場合**基底ベクトルが持ちうる値は総和が1ではない**

提案手法: 正則化による方位分離の促進

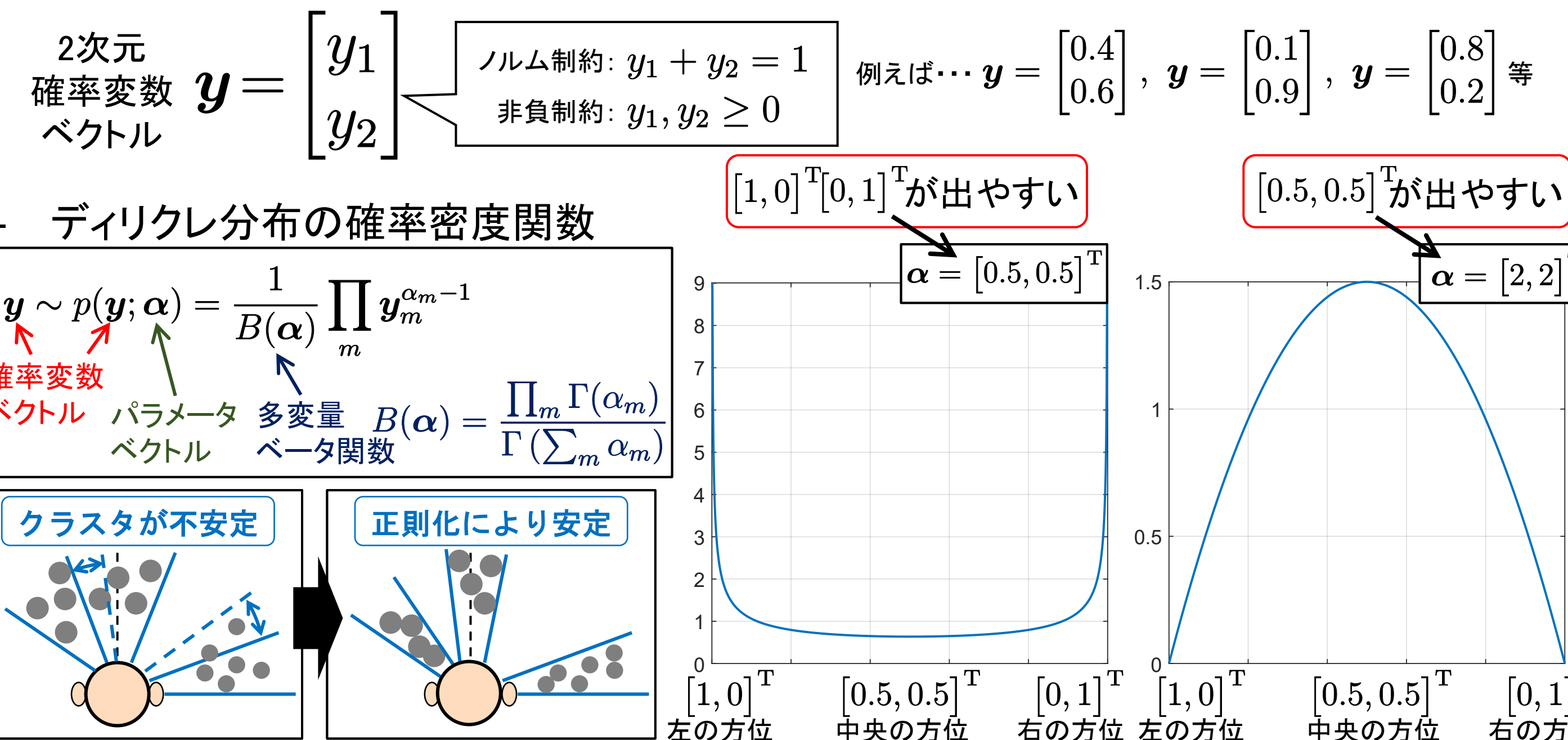
- 基底ベクトルの値に**メリハリ**をつけ方位分離の向上を狙う
- ディリクレ分布をNTFの事前分布に導入**
- 母数によりメリハリのある方位情報を得られるように誘導**



ディリクレ分布

- 生成されるベクトルの総和が1かつすべての要素が0以上という制約を持つ
- パラメータベクトル $\alpha = [\alpha_1, \alpha_2]^T$ は**確率密度の各頂点への集中度**に対応

→ One-hotやデンスなベクトルの生成確率密度を α で制御可能



母数 α を調整することでメリハリがつき中央や左右の方位情報を精度よく捉えることが可能

ディリクレNMFの目的関数と最適化の方針

- ディリクレ分布から分配行列の各列(基底ベクトル)が独立に生成されると仮定

$$\text{Minimize } \mathcal{D}(X|\hat{X}) + \sum_{m,k} (\alpha_{mk} - 1) \log z_{mk}^{-1} \text{ s.t. } z_{mk}, t_{ik}, v_{kj} \geq 0 \forall i, j, m, k, \\ \sum_m z_{mk} = 1 \forall k, \sum_i t_{ik} = 1 \forall k$$

- NMFでは常套手段の**補助関数法**を用いるが, ディリクレ分布由来の**ノルム制約**を陽に扱って解く必要がある → **補助関数法の上限関数最小化に制約条件を付与**

ディリクレNTFの反復更新式

- 補助関数法を用いて最小化問題を解くとそれぞれの更新式が得られる

$$z_{mk} = \begin{cases} \hat{z}_{mk} & (\text{if } \hat{z}_{mk} \geq 0) \\ 0 & (\text{otherwise}) \end{cases} \quad \hat{z}_{mk} = \frac{z_{mk} \sum_{i,j} \frac{x_{ijm}}{\hat{x}_{ijm}} t_{ik} v_{kj} + \alpha_{mk} - 1}{\sum_{m \in \mathcal{M}_k} (z_{mk} \sum_{i,j} \frac{x_{ijm}}{\hat{x}_{ijm}} t_{ik} v_{kj} + \alpha_{mk} - 1)}$$

補助関数の最適化問題に対するラグランジアン $\mathcal{L}$ を考えたときに, $\partial \mathcal{L} / \partial z_{mk} = 0$ を満たす z_{mk} が非負となるインデックス m の集合(k 毎にこの集合を定義している)

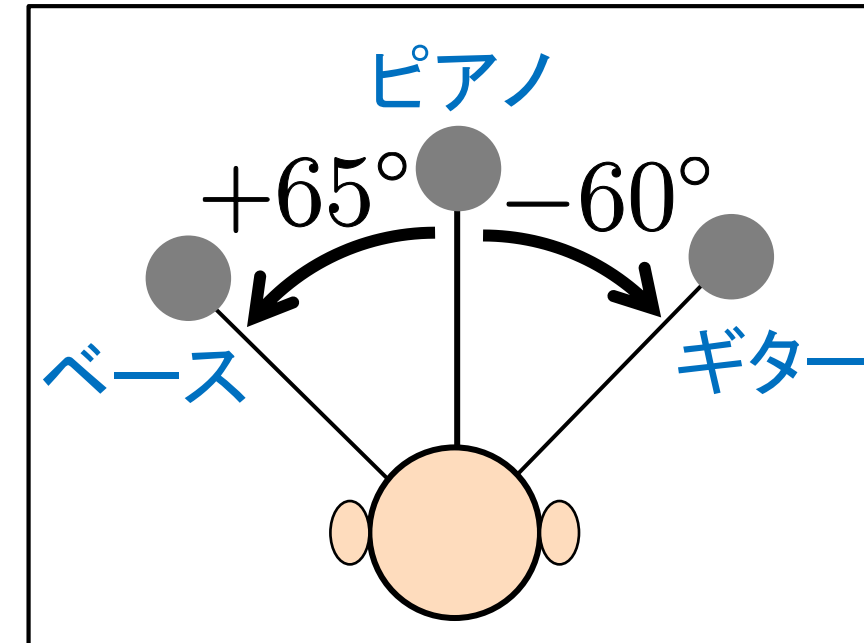
- スケール任意性の回避のため基底行列の基底にもノルム制約 $\sum_i t_{ik} = 1 \forall k$ を付与

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

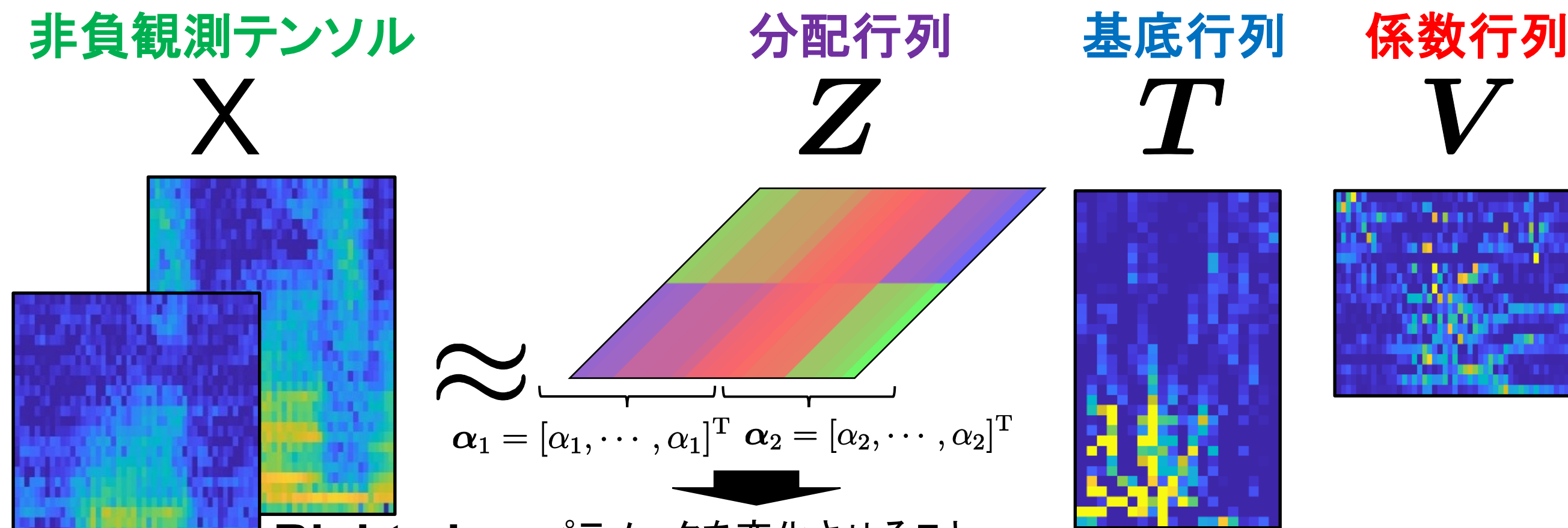
3. 実験

パニング音の分離実験

条件	値
比較対象	振幅比のみを用いる単純な方位クラスタリング(DC)
信号	信号長 17.9s のベース, ピアノ, ギターの合成音
音源の定位	65°, 0°, -60°
窓長	92.9 ms
シフト長	46.4 ms
窓関数	Blackman窓
試行設定	30個の異なる乱数シード



- ディリクレNTFのパラメータ設定**
 - 基底数は $K = 150$ とし, 基底を $K_1 = \{1, 2, \dots, N\}$, $K_2 = \{N + 1, \dots, 150\}$ の2つのグループに分けた
 - それぞれのグループに対して $\alpha_1 = [\alpha_1, \dots, \alpha_1]^T$, $\alpha_2 = [\alpha_2, \dots, \alpha_2]^T$ を与えた



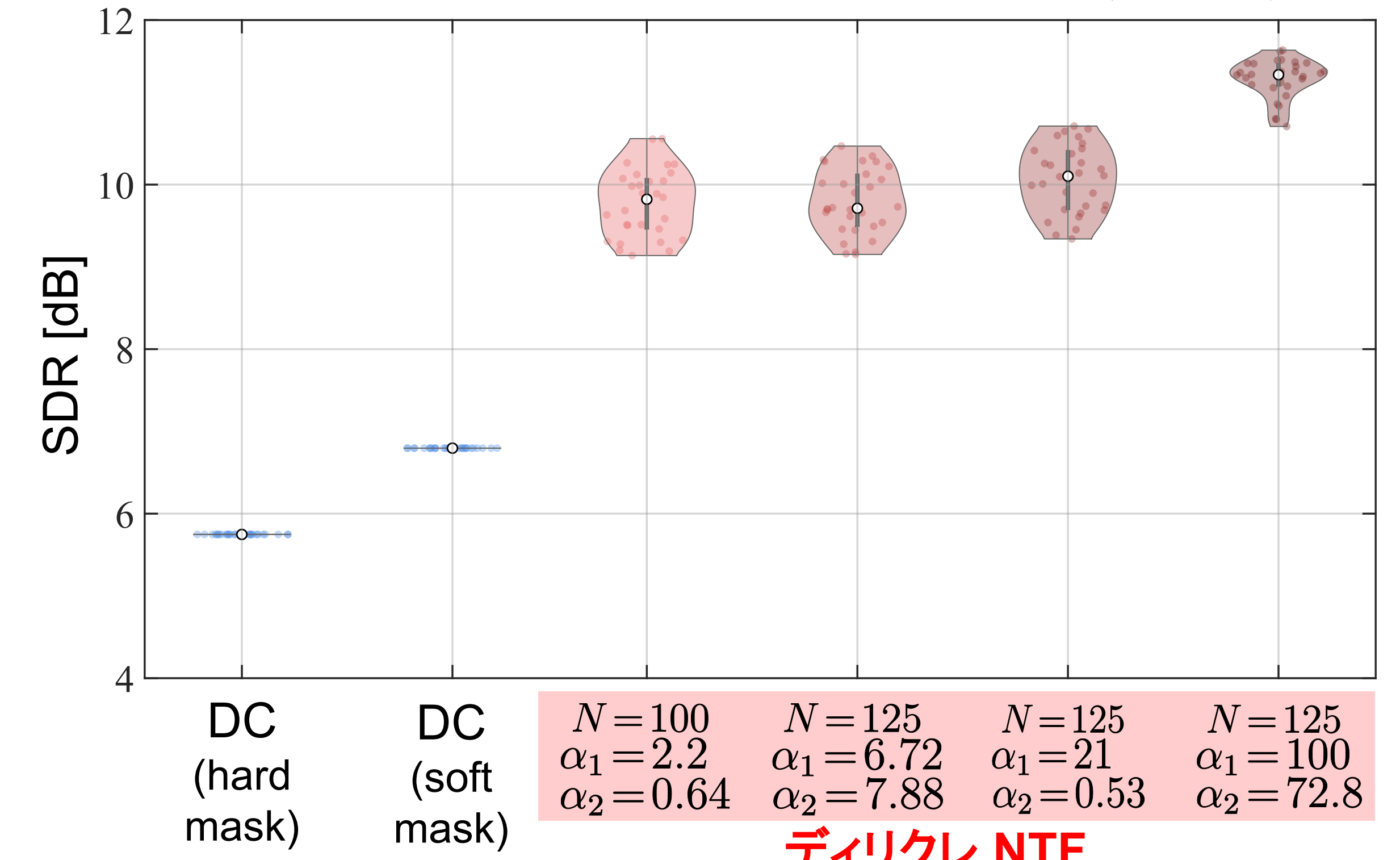
- One-hot** $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ or $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ or **デンス** $\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$
- K-means**でクラスタリングを行い, チャンネルごとにWienerフィルタを作成し分離

$$C_{ijmn} = \frac{(\sum_{k \in K_d} z_{mk} t_{ik} v_{kj})^2}{(\sum_k z_{mk} t_{ik} v_{kj})^2} x_{ijm}$$

d 番目のクラスタの方位を抽出するウィーナフィルタ

実験結果

- 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す
- SDRは**bss_eval_images**を使用(分離された方位の正解度合いも加味されたSDR)
- 提案手法ではどちらもデンスに誘導するパラメータを用いた場合が最良となった



4. まとめ

- 補助関数法において上限関数に制約条件付き最適化を適用した場合でも正しく正則化を行えることを確認した
- ディリクレNTFと単純な方位クラスタリングを比較し, パラメータを適切に設定した場合ディリクレNTFの性能が上回っていることを確認した