

空間正則化付き独立ベクトル抽出に対する 分離フィルタのスケール正則化の導入*

☆上田哲也 (早稲田大), 中谷智広, 池下林太郎, 木下慶介, 荒木章子 (NTT),
牧野昭二 (早稲田大)

1 はじめに

音響信号の音源分離とは、複数の音響信号が混合した観測信号から、混合前の個々の信号を推定する技術である。なかでも、ブラインド音源分離 (Blind Source Separation: BSS) は事前情報や事前収録データを用いずに音源分離を行う技術であり、音声認識率の改善 [1] など様々なアプリケーションへの応用が期待される。BSSとして、信号間の独立性を最大化するように分離フィルタを推定する独立ベクトル分析 (Independent Vector Analysis: IVA) [2,3] や独立ベクトル抽出 (Independent Vector Extraction: IVE) [4-6] などが提案されている。

しかし、BSSは分離信号の出力順序に任意性が生じてしまう。音源分離を音アプリケーションに適用するためには、分離信号が混合前の音響信号と対応づくように音源分離できる方が望ましい。分離信号の出力順序を所望の順序に誘導する (以降、出力順序誘導と呼ぶ) 方法として、BSSに空間正則化 (Spatial Regularization: SR [7]) を導入した研究が盛んに行われてきた [7-10]。空間正則化は、話者からマイクロホンまでの音響伝達関数を入力とした時の分離フィルタの出力を規定することで、伝達関数に対応する方向を強調もしくは抑圧する。なお、伝達関数は平面波仮定に基づき、各話者からの到来方向から推定可能とする。本稿では、IVAやIVEに空間正則化を導入した手法をそれぞれ空間正則化付きIVA (SRIVA) [9]、空間正則化付きIVE (SRIVE) [10] と呼ぶ。SRIVAやSRIVEは空間正則化項とIVA (IVE) による負の対数尤度関数の和を全体のコスト関数として扱うことで、音源分離と出力順序誘導を同時に行う。

しかし、従来のSRIVAやSRIVEでは次の二つが与える出力順序誘導への影響が十分には考慮されていない:

- 1) IVAやIVEのスケール任意性。
- 2) 伝達関数の推定誤差。

まず1)に関して、伝達関数を入力とした時の分離フィルタの出力は分離フィルタのスケールに依存する。一方、IVA (IVE) にはスケール任意性がある。そのため、出力を事前に規定しても、対応する話者を強調するか抑圧するかのどちらになるかが定まらないという問題があった。次に2)に関して、伝達関数に対応する方向を抑圧する空間正則化を用いるとする。このとき、SRIVA (SRIVE) が設計するコスト関数は指定した方向を抑圧する急峻な局所解を持つ可能性がある。しかし、推定伝達関数に誤差が含まれる場合、その局所解に収束した時に、想定した出力順序を得られない場合がある。実際、事前実験 (Fig. 1 (d)) では、SRIVEの出力順序正答率が高々75%までしか示せなかった。

本稿では、空間正則化に加えてさらに分離フィルタのスケール正則化を導入することで、SRIVA (SRIVE)

が上述の二つを考慮した出力順序誘導を実現することを実験的に示す。スケール正則化は分離フィルタをスケールが小さくなるように誘導し、スケール任意性を制限する。これにより、伝達関数を入力とした時の分離フィルタの出力を規定した時に、その規定が対応する話者の強調と抑圧のどちらとして機能するかを定められる。さらにスケール正則化は、SRIVA (SRIVE) が用いる空間共分散行列に diagonal loading を行うことに相当する。これにより、伝達関数に対応する方向を抑圧する局所解の急峻性を和らげ、IVEの尤度を含めたコスト関数全体として、伝達関数に対応する方向の近傍にいる話者を抑圧できるようになると期待する。実験の結果、適切なスケール正則化項を用いることで、分離性能を維持しながら出力順序誘導の性能を改善することを確認した。

2 空間正則化付き独立ベクトル抽出

2.1 音源分離問題

M 個のマイクロホンを用いて N 個の音源信号を取り出す場合を考える。観測信号、音源信号、雑音信号の時間周波数成分をそれぞれ

$$\mathbf{x}(f, t) = [x_1(f, t), \dots, x_M(f, t)]^T \in \mathbb{C}^M, \quad (1)$$

$$\mathbf{s}(f, t) = [s_1(f, t), \dots, s_N(f, t)]^T \in \mathbb{C}^N, \quad (2)$$

$$\mathbf{z}(f, t) = [z_{N+1}(f, t), \dots, z_M(f, t)]^T \in \mathbb{C}^{M-N} \quad (3)$$

とする。ただし、 f, t, n, m はそれぞれ周波数ビン、時刻、音源、マイクロホンのインデックスであり、 F, T は f, t の総数である。 $(\cdot)^T$ は行列の転置である。観測信号 $\mathbf{x}(f, t)$ と分離信号 $\hat{\mathbf{s}}(f, t), \hat{\mathbf{z}}(f, t)$ の関係が次式で表せると仮定する:

$$\begin{pmatrix} \hat{\mathbf{s}}(f, t) \\ \hat{\mathbf{z}}(f, t) \end{pmatrix} = \mathbf{W}^H(f) \mathbf{x}(f, t). \quad (4)$$

ここで、 $\mathbf{W}(f) = [\mathbf{w}_1(f), \dots, \mathbf{w}_M(f)] \in \mathbb{C}^{M \times M}$ を分離行列、 $\mathbf{w}_n(f) \in \mathbb{C}^M$ を分離フィルタと呼ぶ。 $(\cdot)^H$ はエルミート転置である。また、 n 番目の音源信号 $s_n(f, t)$ から m 番目のマイクロホンに到達するまでの伝達関数 $a_{n,m}(f)$ が平面波仮定に基づき、到来方向から推定可能とする。本稿では、入力された観測信号 $\mathcal{X} = \{x_m(f, t)\}_{f,t,m}$ と伝達関数 $\mathcal{A} = \{a_{n,m}(f)\}_{n,m,f}$ から $\hat{\mathbf{s}}(f, t) \approx \mathbf{s}(f, t)$ となるように、すなわち、出力の順序も含めて音源信号を取り出す分離フィルタ $\mathcal{W} = \{\mathbf{w}_n(f)\}_{n,f}$ を推定する問題を考える。

2.2 独立ベクトル抽出

分離信号 $\hat{\mathbf{s}}_n(t) = [\hat{s}_n(1, t), \dots, \hat{s}_n(F, t)]^T$ と雑音信号 $\hat{\mathbf{z}}(f, t)$ が統計的に独立であり、 $\hat{\mathbf{s}}_n(t)$ は優ガウス分布、 $\hat{\mathbf{z}}(f, t)$ は定常ガウス分布に従うと仮定する:

*Scale Regularization for Spatially Regularized Independent Vector Extraction. by Tetsuya UEDA (Waseda University), Tomohiro NAKATANI, Rintaro IKESHITA, Keisuke KINOSHITA, Shoko ARAKI (NTT), Shoji MAKINO (Waseda University)

$$p(\{\hat{\mathbf{s}}_n(t), \hat{\mathbf{z}}(f, t)\}_{n, f, t}) = \prod_{n, t} p(\hat{\mathbf{s}}_n(t)) \prod_{f, t} p(\hat{\mathbf{z}}(f, t)) \quad (5)$$

$$p(\hat{\mathbf{s}}_n(t)) = C \exp(-G(r_n(t))), \quad r_n(t) = \|\hat{\mathbf{s}}_n(t)\|, \quad (6)$$

$$p(\hat{\mathbf{z}}(f, t)) = \frac{1}{\pi^{M-N} \det \mathbf{\Omega}(f)} e^{-\hat{\mathbf{z}}^H(f, t) \mathbf{\Omega}^{-1}(f) \hat{\mathbf{z}}(f, t)}. \quad (7)$$

ただし, C は正規化係数である. $\|\cdot\|$ は L2 ノルムである. 関数 $G: \mathbb{R} \rightarrow \mathbb{R}$ は微分可能であり, $\varphi(r) = G'(r)/r$ が $r > 0$ について単調減少するような関数である [11].

このとき, \mathcal{X} が与えられた時の負の対数尤度関数 $\mathcal{L}(\mathcal{W}) = -\frac{1}{T} \log p(\mathcal{X} | \mathcal{W})$ は以下のように記述できる:

$$\begin{aligned} \mathcal{L}(\mathcal{W}) \stackrel{c}{=} & \sum_{f=1}^F \log \det \mathbf{\Omega}(f) - 2 \log |\det \mathbf{W}(f)| \\ & + \frac{1}{T} \sum_{t=1}^T \left[\sum_{n=1}^N G(r_n(t)) + \sum_{f=1}^F \hat{\mathbf{z}}^H(f, t) \mathbf{\Omega}^{-1}(f) \hat{\mathbf{z}}(f, t) \right], \end{aligned} \quad (8)$$

ただし, $\stackrel{c}{=}$ は \mathcal{W} に依存しない項を除いた等号を表す.

2.3 空間正則化

空間正則化付き音源分離 [7–10] では, 負の対数尤度関数と空間情報を用いて設計される空間正則化項を足したコスト関数を最小にするような分離フィルタを推定する:

$$\underset{\mathcal{W}}{\text{minimize}} \mathcal{L}(\mathcal{W}) + \mathcal{J}(\mathcal{W}). \quad (9)$$

$\mathcal{J}(\mathcal{W})$ は, 各分離フィルタ $\mathbf{w}_n(f)$ が所望の音源信号 $s_n(f, t)$ を強調ように誘導する正則化項であり, これまでに様々な正則化項が提案されている [7–10]. 本稿では次の正則化項を扱う:

$$\mathcal{J}(\mathcal{W}) = \mathcal{J}_{\text{null}}(\mathcal{W}) + \mathcal{J}_{\text{unit}}(\mathcal{W}) \quad (10)$$

$$\mathcal{J}_{\text{null}}(\mathcal{W}) = \lambda^{\text{null}} \sum_{f=1}^F \sum_{n=1}^N \sum_{i \neq n} |\mathbf{w}_n^H(f) \mathbf{a}_i(f)|^2 \quad (11)$$

$$\mathcal{J}_{\text{unit}}(\mathcal{W}) = \lambda^{\text{unit}} \sum_{f=1}^F \sum_{n=1}^N |\mathbf{w}_n^H(f) \mathbf{a}_n(f) - 1|^2 \quad (12)$$

ここで, λ^{null} と λ^{unit} は各正則化項の重みである. また, $\mathbf{a}_n(f) = [a_{n,1}(f), \dots, a_{n,M}(f)]^T \in \mathbb{C}^M$ は伝達関数 $a_{n,m}$ を要素とするステアリングベクトルである. $\mathcal{J}_{\text{null}}(\mathcal{W})$ はヌル正則化項 [7] であり, $\mathbf{w}_n(f)$ が $\mathbf{a}_{i \neq n}(f)$ に対応する音源信号 $s_{i \neq n}(f, t)$ を除去するように誘導する. $\mathcal{J}_{\text{unit}}(\mathcal{W})$ は単位応答正則化項 [7,9] であり, $\mathbf{w}_n^H(f) \mathbf{a}_n(f) = 1$ を誘導する.

2.4 空間正則化付き IVE の最適化アルゴリズム

SRIVE では, 最適化問題 (9) の局所最適解を効率的に求めるために, 上界最小化アルゴリズムに基づいて次式に示す補助関数 $\mathcal{L}_{\text{Aux}}(\mathcal{W})$ の設計とその最小

化を繰り返し行う [9–11]:

$$\begin{aligned} \mathcal{L}_{\text{Aux}}(\mathcal{W}) \stackrel{c}{=} & \log \det \mathbf{\Omega}(f) - 2 \log |\det \mathbf{W}(f)| \\ & + \sum_{n=1}^N \left\{ \mathbf{w}_n^H(f) \mathbf{\Sigma}'_n(f) \mathbf{w}_n(f) - \lambda^{\text{unit}} (\mathbf{w}_n^H(f) \mathbf{a}_n(f) \right. \\ & \left. + \mathbf{a}_n^H(f) \mathbf{w}_n(f)) \right\} + \text{tr}(\mathbf{W}_Z(f) \mathbf{\Sigma}_Z(f) \mathbf{W}_Z(f)), \end{aligned} \quad (13)$$

ただし, $\mathbf{W}_Z(f) = [\mathbf{w}_{N+1}(f), \dots, \mathbf{w}_M(f)]$ である. $\mathbf{\Sigma}'_n(f)$ は空間正則化項を考慮した重み付き共分散行列であり, 次のように求められる:

$$\mathbf{\Sigma}'_n(f) = \mathbf{\Sigma}_n(f) + \sum_{i=1}^N (\lambda_{n,i} \mathbf{a}_i(f) \mathbf{a}_i^H(f)), \quad (14)$$

$$\lambda_{n,i} = \begin{cases} \lambda^{\text{unit}} & (\text{if } i = n), \\ \lambda^{\text{null}} & (\text{otherwise}). \end{cases} \quad (15)$$

重み付き共分散行列 $\mathbf{\Sigma}_n(f), \mathbf{\Sigma}_Z(f)$ は次のように求まる:

$$\mathbf{\Sigma}_n(f) = \frac{1}{T} \sum_t \varphi(r_n(t)) \mathbf{x}(f, t) \mathbf{x}^H(f, t) \quad (16)$$

$$\mathbf{\Sigma}_Z(f) = \frac{1}{T} \sum_t \mathbf{x}(f, t) \mathbf{x}^H(f, t) \quad (17)$$

この補助関数 $\mathcal{L}_{\text{Aux}}(\mathcal{W})$ を \mathcal{W} について最小化する部分問題に対して, $\mathbf{w}_1(f) \rightarrow \mathbf{w}_2(f) \rightarrow \dots \rightarrow \mathbf{w}_N(f) \rightarrow \mathbf{W}_Z(f)$ の順に更新するブロック座標降下法が提案されている [10]. $\mathbf{w}_n(f)$ の更新は

$$\mathbf{w}_n(f) \leftarrow \underset{\mathbf{w}_n(f)}{\text{argmin}} \mathcal{L}_{\text{Aux}}(\mathcal{W}) \quad (18)$$

のように行う. この問題は Vectorwise Coordinate Descent [8] を適用することで大域的に解ける:

$$\mathbf{w}_n(f) = (\mathbf{W}^H(f) \mathbf{\Sigma}'_n(f))^{-1} \mathbf{e}_n, \quad (19)$$

$$\hat{\mathbf{w}}_n(f) = \lambda^{\text{unit}} \mathbf{\Sigma}'_n^{-1}(f) \mathbf{a}_n(f), \quad (20)$$

$$\hat{h}_n(f) = \mathbf{w}_n^H(f) \mathbf{\Sigma}'_n(f) \mathbf{w}_n(f), \quad (21)$$

$$\hat{h}_n(f) = \mathbf{w}_n^H(f) \mathbf{\Sigma}'_n(f) \hat{\mathbf{w}}_n(f), \quad (22)$$

$$\mathbf{w}_n(f) = \begin{cases} \frac{1}{\sqrt{\hat{h}_n(f)}} \mathbf{w}_n(f) + \hat{\mathbf{w}}_n(f) & (\text{if } \hat{h}_n(f) = 0), \\ \frac{\hat{h}_n(f)}{2\hat{h}_n(f)} \left[-1 + \sqrt{1 + \frac{4\hat{h}_n(f)}{|\hat{h}_n(f)|^2}} \right] \mathbf{w}_n(f) + \hat{\mathbf{w}}_n(f) & (\text{otherwise}). \end{cases} \quad (23)$$

$\mathbf{W}_Z(f)$ の更新は

$$\mathbf{W}_Z(f) \leftarrow \underset{\mathbf{W}_Z(f)}{\text{argmin}} \mathcal{L}_{\text{Aux}}(\mathcal{W}) \quad (24)$$

のように行う. この問題は, [5] で提案された更新式を用いることで解くことができる [6, 命題 4]:

$$\begin{aligned} \mathbf{W}_Z(f) \leftarrow & \left(\begin{array}{c} -(\mathbf{W}_S^H(f) \mathbf{\Sigma}_Z(f) \mathbf{E}_N)^{-1} (\mathbf{W}_S^H(f) \mathbf{\Sigma}_Z(f) \mathbf{E}_{M-N}) \\ \mathbf{I}_{M-N} \end{array} \right), \end{aligned} \quad (25)$$

ただし, $\mathbf{W}_S(f) = [\mathbf{w}_1(f), \dots, \mathbf{w}_N(f)]$, $\mathbf{E}_N = [\mathbf{e}_1, \dots, \mathbf{e}_N]$, $\mathbf{E}_{M-N} = [\mathbf{e}_{N+1}, \dots, \mathbf{e}_M]$ である.

3 提案法: 空間正則化に対するスケール正則化の導入

前節で述べた空間正則化付き独立ベクトル抽出 (SRIVE) は以下の問題を抱えている:

- 1) $\mathcal{J}_{\text{unit}}(\mathcal{W})$ は $\mathbf{a}_n(f)$ の方向に対して $\mathbf{w}_n^H(f)\mathbf{a}_n(f) = 1$ を促す正則化項だが, その他の方向については何も規定されていない. 分離フィルタのスケール任意性に基づき, $\|\mathbf{w}_n(f)\| \gg 1$ となるように全体コスト関数が収束した時, $\mathcal{J}_{\text{unit}}(\mathcal{W})$ は $\mathbf{w}_n(f)$ が $\mathbf{a}_n(f)$ の方向を抑圧するように機能する可能性がある.
- 2) $\mathcal{J}_{\text{null}}(\mathcal{W})$ に関して, コスト関数 $\mathcal{L}(\mathcal{W}) + \mathcal{J}_{\text{null}}(\mathcal{W})$ は $\mathbf{w}_n^H(f)\mathbf{a}_i(f) \rightarrow 0$ となる急峻な局所解を持つ可能性がある. もし, 推定した $\mathbf{a}_n(f)$ に誤差がある場合, その局所解に収束しても, 所望と異なる方向の音を抑圧し, 想定した出力順序を得られない場合がある.

この問題に対し, 我々は次の最適化問題を設定することで上記の問題の解決を試みる.

$$\underset{\mathcal{W}}{\text{minimize}} \mathcal{L}(\mathcal{W}) + \mathcal{J}_{\text{unit}}(\mathcal{W}) + \mathcal{J}_{\text{null}}(\mathcal{W}) + \mathcal{J}_{\text{scale}}(\mathcal{W}), \quad (26)$$

$$\text{where } \mathcal{J}_{\text{scale}}(\mathcal{W}) = \lambda^{\text{scale}} \sum_n \sum_f \mathbf{w}_n^H(f)\mathbf{w}_n(f). \quad (27)$$

式 (27) は, 分離フィルタ $\mathbf{w}_n(f)$ のスケールを小さくする正則化項であり, スケール正則化項と呼ぶことにする. λ^{scale} は重み係数である. スケール正則化を導入することで一つ目の課題に対し, $\|\mathbf{w}_n(f)\| \gg 1$ となる場合を回避することができ, $\mathcal{J}_{\text{unit}}(\mathcal{W})$ が $\mathbf{a}_n(f)$ の方向を強調するようになると期待する.

最適化問題 (26) に対するアルゴリズムとして, 空間正則化付き IVE のアルゴリズムにおいて, 更新式 (14) を次の (28) に変更したものを導出できる:

$$\Sigma'_n(f) = \Sigma_n(f) + \sum_i (\lambda_{n,i} \mathbf{a}_i(f) \mathbf{a}_i^H(f)) + \lambda^{\text{scale}} \mathbf{I}_M. \quad (28)$$

ただし, \mathbf{I}_M は M 行 M 列の単位行列である. つまり, スケール正則化は共分散行列 $\Sigma'_n(f)$ に diagonal loading を行うことに相当する. これにより二つ目の課題に対し, $\mathbf{w}_n^H(f)\mathbf{a}_i(f) \rightarrow 0$ となる局所解の急峻性を和らげることで, 最適化問題 (26) は $\mathbf{a}_n(f)$ に対応する方向の近傍にいる話者を抑圧し, 所望の出力順序が得られる局所解に収束すると期待できる.

実環境での音源分離において, diagonal loading を用いることでより安定した分離が行えることはよく知られている. また, SRIVE において $\mathcal{J}_{\text{scale}}(\mathcal{W})$ と $\mathcal{J}_{\text{null}}(\mathcal{W})$ を組み合わせることについては先行研究がある [10]. 一方, 本研究の主成果は, スケール正則化を導入することで「出力順序の正答率を大きく改善できること」を実験的に示したことにある.

4 実験評価

本実験では, スケール正則化項の空間正則化付き音源分離手法への影響を調査する. 適切なスケール正則化項の重みを用いることで, 音源分離性能を維持しつつ出力順序誘導の性能を改善できることを示す.

4.1 実験条件

観測信号の生成に用いるインパルス応答として, RWCP 実環境音声・音響データベース (RWCP Sourced Source Dataset in Real Acoustic Environment; RWCP-SSD) [12] の残響室 (E2A) を使用する. 分離対象の音声信号として, ATR デジタル音声データベースのセット B の収録データを使用する [13]. そして, 次の操作に従って観測信号を作成する: 1) 話者データセットの中から 3 人の異なる話者の発話をランダムに選択し, 信号長が 20 秒になるように発話を繰り返す. 2) 各発話者からマイクロホンまでの実測インパルス応答と発話信号を畳み込む. 3) 畳み込まれた信号を各マイクロホンで足し合わせた信号を観測信号とする. 以上の操作を行い, 観測信号 (入力テストデータ) を 100 データ作成した. なお, 各発話者から第一マイクロホンへの音像を入力テストデータに対応する正解テストデータとした. また, 実測インパルス応答との畳み込みはサンプリング周波数 48 kHz で行った後に 16 kHz にリサンプリングした. マイクロホンは RWCP データセット内の 19, 20, ..., 23 番目のマイクロホンを選択し, データセット内に記されている $\{70^\circ, 150^\circ, 110^\circ\}$ の方向から到来した固定インパルス応答を使用した.

音源分離には SRIVA [9] や SRIVE [10] を用いた. また, 統計モデルは時変ガウス分布 ($\varphi(r_n(t)) = Fr_n^{-2}(t)$) を使用し, 上界最小化アルゴリズムを 50 回反復して行った. 短時間フーリエ変換の窓長と移動長は 2048, 1024 samples に設定した. 空間正則化項を設計するために, 音響伝達関数 $a_{n,m}(f)$ は平面波仮定に基づいて次のように求めた:

$$a_{n,m}(f) = \exp(-j2\pi f\tau_{n,m})/\sqrt{M}, \quad (29)$$

$$\tau_{n,m} = d(m-1)\cos\theta_n/c. \quad (30)$$

ただし, $j = \sqrt{-1}$ であり, $\tau_{n,m}$ は n 番目の音源から m 番目のマイクロホンに平面波が到来した時の相対的な遅延時間である. $d = 0.0281$ m は各マイク間の距離で, $c = 343$ m/s は音速である. $\{\theta_1, \theta_2, \theta_3\} = \{70^\circ, 150^\circ, 110^\circ\}$ とした¹.

分離性能の評価には, Source-to-Distortion Ratio (SDR) と出力順序正答率を用いる. なお, SDR の計算には BSSEval v3 [14] で実装されている `bss_eval_sources` 関数を使用した. 出力順序正答率は分離信号の平均 Source-to-Interference Ratio が最大になる順序を BSSEval v3 を用いて選択し, その順序が正解データの順序と合致するテストデータを数えた値である.

4.2 実験結果

図 1 は出力順序正答率と SDR の二つの評価指標を, スケール正則化の有無で比較した図である. 各図では, 使用するマイクロホンの数と空間正則化項を変化させた時の結果を示している.

まず, 図 1 中の出力順序正答率に着目する. Fig. 1 (a) において, スケール正則化を用いない ($\lambda^{\text{scale}} = 0$) 場合の出力順序正答率は λ^{unit} の値に関わらず 0 に近い値を示した. これは, $\|\mathbf{w}_n(f)\|$ が 1 より大きい値になるように学習が収束し, $\mathcal{J}_{\text{unit}}(\mathcal{W})$ が目的音を

¹式 (29) and (30) は残響などの影響を考慮しない平面波仮定モデルであるため, データベース上で公表された角度を使っても実収録の伝達関数に誤差を含んだものになることに注意してほしい.

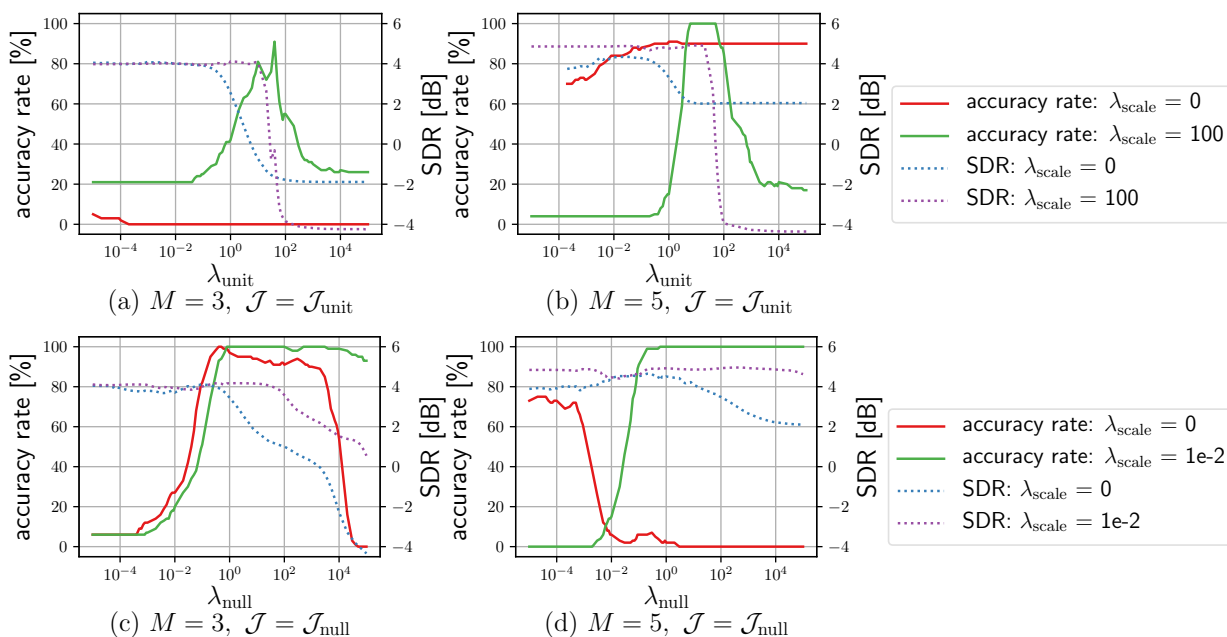


Fig. 1: 空間正則化の重みに対する出力順序正答率 (accuracy rate [%]) と SDR. 各副見出しでは、使用したマイクロホン数 M と 空間正則化を記述している.

抑圧するように機能したことが原因である. 一方で, $\mathcal{J}_{\text{scale}}(W)$ を導入して $\|w_n(f)\|$ を小さくすることで出力順序正答率が改善した. これにより, $\mathcal{J}_{\text{unit}}(W)$ を適切に機能させるためにスケール正則化が有効であると考えられる.

次に, Figs. 1 (b), (d) に着目する. スケール正則化を用いない ($\lambda^{\text{scale}} = 0$) 場合, どのような空間正則化項の重みを与えても出力順序正答率を 100 にすることができなかった. 一方, λ^{scale} と空間正則化項の重みを適切に与えることで, 出力順序正答率を 100 にすることができた. また, Fig. 1 (c) に着目すると, スケール正則化を用いない場合でも λ^{null} を適切に設定することで, 出力順序正答率を 100 にすることができた. しかし, その λ^{null} の範囲は非常に狭い. 一方で, λ^{scale} を適切に与えることで, 出力順序正答率が 100 を示す λ^{null} の範囲を広げることができた. 以上より, スケール正則化は出力順序誘導の性能改善に貢献することを確認した.

最後に, 出力順序正答率と分離性能を統合して評価する. 図 1 中に記されている SDR は BSSEval v3 によって出力順序が解決された場合での SDR を意味している. $\mathcal{J}_{\text{null}}(W)$ を用いた場合に関して, Figs. 1 (c), (d) より, $\mathcal{J}_{\text{scale}}(W)$ を用いることで SDR を維持しながら出力順序正答率を向上させることができた. $\mathcal{J}_{\text{unit}}(W)$ を用いた場合に関しては, Figs. 1 (a), (b) より, SDR が λ^{unit} に依存する結果となった. しかし, 適切な λ^{unit} を用いることで, $\lambda^{\text{scale}} = 0$ を用いた時の SDR を維持しながらより高い出力順序正答率を示す λ^{unit} が存在した. 以上より, 実収録データでの空間正則化付き音源分離において, スケール正則化が分離性能を維持しつつ, 出力順序誘導の性能改善に貢献することを確認した.

5 結論

本稿では, 空間正則化付き IVE (SRIVE) の出力順序誘導の性能を改善するために, 分離フィルタに

スケール正則化を導入することを提案した. 従来の SRIVE はスケール任意性や伝達関数の推定誤差による出力順序誘導への影響が十分には考慮されていなかった. それに対しスケール正則化項を導入することで, 伝達関数に対する分離フィルタの応答を規定した時に, 対応する話者の強調と抑圧のどちらとして機能するかを定められることを実験的に確かめた. また, 平面波仮定に基づいた音響伝達関数を用いてもスケール正則化が分離性能を維持しつつ出力順序誘導の性能改善に貢献することを確認した.

謝辞 本研究は科研費 19H04131 の助成を受けた.

参考文献

- [1] T. Higuchi *et al.*, *IEEE/ACM Trans. ASLP*, vol. 25, no. 4, pp. 780–793, 2017.
- [2] T. Kim *et al.*, in *Proc. ICA*, 2006, pp. 165–172.
- [3] A. Hiroe, in *Proc. ICA*, 2006, pp. 601–608.
- [4] Z. Koldovsky *et al.*, *IEEE Trans. SR*, vol. 67, no. 4, pp. 1050–1064, 2018.
- [5] R. Scheibler *et al.*, in *Proc. WASPAA*, 2019, pp. 185–189.
- [6] R. Ikeshita *et al.*, *IEEE Trans. SR*, vol. 69, pp. 3252–3267, 2021.
- [7] L. C. Parra *et al.*, *IEEE Trans. SAP*, vol. 10, no. 6, pp. 352–362, 2002.
- [8] Y. Mitsui *et al.*, in *Proc. ICASSP*, 2018, pp. 746–750.
- [9] L. Li *et al.*, in *Proc. ICASSP*, 2020, pp. 846–850.
- [10] A. Brendel *et al.*, *IEEE Trans. SR*, vol. 68, pp. 3545–3558, 2020.
- [11] N. Ono *et al.*, in *Proc. LVA/ICA*, 2010, pp. 165–172.
- [12] S. Nakamura *et al.*, in *Proc. LREC*, 2000, pp. 965–968.
- [13] A. Kurematsu *et al.*, *Speech communication*, vol. 9, no. 4, pp. 357–363, 1990.
- [14] E. Vincent *et al.*, *IEEE/ACM Trans. ASLP*, vol. 14, no. 4, pp. 1462–1469, 2006.