

被り音混合系における独立低ランク行列分析のセミブラインド推定*

☆谷野宮蒼士, 北村大地 (香川高専), 高宗典玄, 山岡洗瑛, 猿渡洋 (東京大),
高橋祐, 近藤多伸, △山川颯人 (ヤマハ)

1 はじめに

通常の音楽ライブ演奏では, 各音源に近接させてマイクロホン配置し, 各音源の音を収録及び増幅して観客に届ける. この工程は sound reinforcement (SR) と呼ばれる. このようなマイクロホン配置は, 近接させている音源 (目的音源) の信号のみを得る目的がある. 特に, ボーカル (Vo.) 音源の明瞭な収録は SR において重要である. しかし Fig. 1 に示すように, ステージ上には楽器用アンプ, モニタースピーカ, 及び front-of-house (FoH) スピーカなど, 多数の音源が存在する. 結果, Vo. 用マイクロホンに他の音源が混入し, 被り音となって観測され, SR に悪影響を及ぼす.

多チャンネル音源分離 (multichannel audio source separation: MASS) は, 前述の被り音を抑圧する手法として期待される. 特に, 非負値行列因子分解 (non-negative matrix factorization: NMF) [1] を用いた時間チャンネル NMF (time-channel NMF: TCNMF) [2] は, 被り音抑圧手法として提案されている. TCNMF は, 振幅スペクトログラムに対する周波数毎の混合行列を推定し, 時間周波数 (time-frequency: TF) マスクを適用することで被り音を抑圧する. しかし, TF マスクは人工的な歪みを生じさせ, 出力信号の音質を劣化させてしまう問題がある.

他の手法として, ブラインド音源分離 (blind source separation: BSS) [3] の活用が考えられる. 本稿では, 独立低ランク行列分析 (independent low-rank matrix analysis: ILRMA) [4, 5] と呼ばれる BSS に着目する. ILRMA は低ランクな TF モデルを活用するため, 音楽信号の MASS 問題によく適合する. また ILRMA は, 周波数毎の複素分離行列を推定し, 線形時不変な分離処理を適用するため, 音質の劣化をある程度回避できる利点がある. しかしながら, 音楽ライブ演奏における MASS 問題に ILRMA を直接適用した場合, 分離行列の推定精度が劣化する. これは, Fig. 1 のようにマイクロホン間隔が大きい場合, 空間エイリアシングが発生することが原因である.

現実の音楽ライブ演奏環境では, Vo. 用マイクロホンに混入する被り音が最も深刻である. 一方, Vo. 用以外のマイクロホンに混入する被り音は比較的エネルギーが小さく, その影響は限定的である場合が多い. このような状況を踏まえ, 本稿では Vo. 用マイクロホンに対する被り音の抑圧に焦点を当てる. 提案手法では, Vo. 用以外のマイクロホンから得られる信号が各音源のリファレンス信号として利用可能であると仮定し, ILRMA にセミブラインドな分離モデルを導入する. このアプローチは Vo. 音源に対する線形時不変分離フィルタの推定と解釈できる. 本稿では, 実際の音楽ライブ演奏で測定したインパルス応答を用いた実験を通して, 提案手法の有効性を確認する.

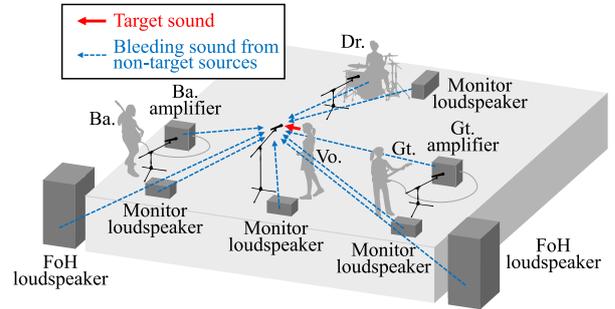


Fig. 1: Spatial arrangement of close microphones and sound sources.

2 従来の BSS

2.1 定式化

今, N 及び M をそれぞれ音源数及びマイクロホン数とする. 短時間フーリエ変換 (short-time Fourier transform: STFT) で得られる各 TF スロットの音源, 観測, 及び推定信号を次式で表す.

$$\mathbf{s}_{ft} = [s_{ft1}, \dots, s_{ftn}, \dots, s_{ftN}]^T \in \mathbb{C}^N \quad (1)$$

$$\mathbf{x}_{ft} = [x_{ft1}, \dots, x_{ftm}, \dots, x_{ftM}]^T \in \mathbb{C}^M \quad (2)$$

$$\mathbf{y}_{ft} = [y_{ft1}, \dots, y_{ftn}, \dots, y_{ftN}]^T \in \mathbb{C}^N \quad (3)$$

ここで, $f \in \{1, 2, \dots, F\}$, $t \in \{1, 2, \dots, T\}$, $n \in \{1, 2, \dots, N\}$, 及び $m \in \{1, 2, \dots, M\}$ はそれぞれ周波数ビン, 時間フレーム, 音源, 及びマイクロホンのインデックスである. また, \cdot^T は転置を表す.

BSS では, 観測信号の混合系として次式を仮定する.

$$\mathbf{x}_{ft} = \mathbf{A}_f \mathbf{s}_{ft} \quad (4)$$

ここで, $\mathbf{A}_f \in \mathbb{C}^{M \times N}$ は周波数毎の時不変混合行列である. 決定条件 ($M = N$) の BSS は, 次式で定義される分離行列を推定することが目的となる.

$$\mathbf{W}_f = \begin{bmatrix} w_{f11}^* & \cdots & w_{f1M}^* \\ \vdots & \ddots & \vdots \\ w_{fN1}^* & \cdots & w_{fNM}^* \end{bmatrix} = \begin{bmatrix} \mathbf{w}_{f1}^H \\ \vdots \\ \mathbf{w}_{fN}^H \end{bmatrix} \in \mathbb{C}^{N \times M} \quad (5)$$

ここで, $*$ 及び \cdot^H はそれぞれ複素共役及びエルミート転置を表す. 以降, 本稿では決定条件である $M = N$ の場合のみを考える.

分離行列が $\mathbf{W}_f = \mathbf{A}_f^{-1}$ を満たす場合, 推定 (分離) 信号は次式によって得られる.

$$\mathbf{y}_{ft} = \mathbf{W}_f \mathbf{x}_{ft} \quad (y_{ftn} = \mathbf{w}_{fn}^H \mathbf{x}_{ft}) \quad (6)$$

すなわち, \mathbf{w}_{fn} は n 番目の音源のための線形時不変分離フィルタと解釈でき, y_{ftn} の推定は \mathbf{w}_{fn} と \mathbf{x}_{ft} の内積となる.

*Semi-blind estimation of independent low-rank matrix analysis for bleeding sound mixtures. By Soushi TANINOMIYA, Daichi KITAMURA (NIT, Kagawa), Norihiro TAKAMUNE, Kouei YAMAOKA, Hiroshi SARUWATARI (The University of Tokyo), Yu TAKAHASHI, Kazunobu KONDO, and Hayato YAMAKAWA (Yamaha Corporation).

2.2 ILRMA

ILRMA [4, 5] は、分離行列 \mathbf{W}_f を正確に推定するための強力なアプローチである。この手法は、音源間の統計的独立性の最大化と各推定信号のパワースペクトログラムの低ランク近似 (NMF) を同時に達成する。NMF に基づく低ランク TF モデリングは音楽音源によく適合するため、ILRMA は特に音楽信号の MASS に効果的である。

ILRMA のコスト関数は次式によって定義される。

$$\mathcal{J} = -T \sum_f \log |\det \mathbf{W}_f|^2 + \sum_{f,t,n} \left[\frac{|\mathbf{w}_{fn}^H \mathbf{x}_{ft}|^2}{\sum_k b_{fkn} v_{ktn}} + \log \sum_k b_{fkn} v_{ktn} \right] \quad (7)$$

ここで、 b_{fkn} 及び v_{ktn} はそれぞれ NMF における基底行列 $\mathbf{B}_n \in \mathbb{R}_+^{F \times K}$ 及びアクティベーション行列 $\mathbf{V}_n \in \mathbb{R}_+^{K \times T}$ の非負要素、 $k \in \{1, 2, \dots, K\}$ は基底ベクトルのインデクスである。ランク K 行列 $\mathbf{B}_n \mathbf{V}_n$ は n 番目の推定信号のパワースペクトログラムの低ランクモデルであり、 \mathbf{W}_f の高精度な推定を促進する。

変数 \mathbf{W}_f , \mathbf{B}_n , 及び \mathbf{V}_n は、式 (7) のコスト関数の最小化で推定される。ILRMA [4] では、 \mathbf{W}_f の高速・安定な更新則である反復射影法 (iterative projection: IP) [6] と、 \mathbf{B}_n 及び \mathbf{V}_n の乗法更新則 [7] を用いており、次式の最適化アルゴリズムとなる。

$$\mathbf{U}_{fn} = \frac{1}{T} \sum_t \frac{1}{\sum_k b_{fkn} v_{ktn}} \mathbf{x}_{ft} \mathbf{x}_{ft}^H \quad (8)$$

$$\mathbf{w}_{fn} \leftarrow (\mathbf{W}_f \mathbf{U}_{fn})^{-1} \mathbf{e}_n \quad (9)$$

$$\mathbf{w}_{fn} \leftarrow \frac{\mathbf{w}_{fn}}{\sqrt{\mathbf{w}_{fn}^H \mathbf{U}_{fn} \mathbf{w}_{fn}}} \quad (10)$$

$$b_{fkn} \leftarrow b_{fkn} \sqrt{\frac{\sum_t |\mathbf{w}_{fn}^H \mathbf{x}_{ft}|^2 v_{ktn} (\sum_{k'} b_{fk'n} v_{k'tn})^{-2}}{\sum_t v_{ktn} (\sum_{k'} b_{fk'n} v_{k'tn})^{-1}}} \quad (11)$$

$$v_{ktn} \leftarrow v_{ktn} \sqrt{\frac{\sum_f |\mathbf{w}_{fn}^H \mathbf{x}_{ft}|^2 b_{fkn} (\sum_{k'} b_{fk'n} v_{k'tn})^{-2}}{\sum_f b_{fkn} (\sum_{k'} b_{fk'n} v_{k'tn})^{-1}}} \quad (12)$$

ここで、 $\mathbf{e}_n \in \{0, 1\}^N$ は n 番目の要素が 1 の one-hot ベクトルである。最適化後は推定信号を式 (6) で得る。

3 提案手法

3.1 手法提案の動機

実際の音楽ライブ演奏環境では、Vo. 用マイクロホンに混入する被り音は他のマイクロホンに混入する被り音よりも深刻な問題となる。特に、Vo. 用マイクロホンに混入するドラム (Dr.) 音源の被り音は大きな課題となる。この原因として、Fig. 2 に示すように、Dr. 音源本来の音量が大きい点、Vo. 用マイクロホンの空間的配置が Dr. 音源を收音しやすい点、及び Vo. 用マイクロホンのヘッドアンプ (head amplifier: HA) ゲインが高く設定される点が挙げられる。一方、ギター (Gt.) アンプ、ベース (Ba.) アンプ、及び Dr. などの他のマイクロホンに混入する被り音はエネルギーが比較的小さく、その影響は限定的な場合が多い。

前述の条件に基づき、本稿では Vo. 用マイクロホンで観測される被り音の抑圧に特化したセミブライン

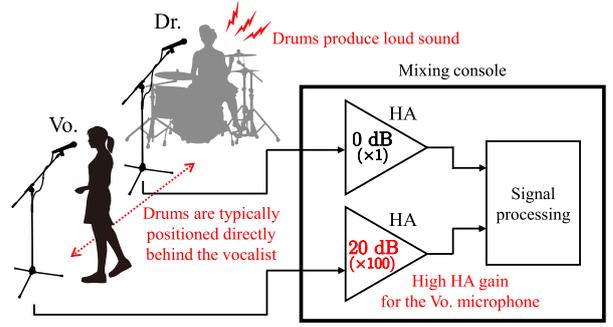


Fig. 2: Typical stage setup in live music performances.

ド音源分離手法を提案する。提案手法では、ILRMA に対してセミブラインド分離モデルを導入する。推定パラメータの数を大幅に削減できるため、観測信号に空間エイリアシングが発生する場合においても、頑健な被り音抑圧が期待できる。

3.2 Semi-blind ILRMA

今、 s_{ft1} ($n = 1$) 及び x_{ft1} ($m = 1$) をそれぞれ Vo. 音源及び Vo. 用マイクロホンの TF 成分とする。 x_{ft1} には他の音源 s_{ft2}, \dots, s_{ftN} からの被り音が混入しているが、他のマイクロホンの観測信号 x_{ft2}, \dots, x_{ftM} には被り音が混入しておらず、リファレンスチャネルとして利用できることを仮定する。従って、 $x_{ftm} = s_{ftm} \forall m = n \in \{2, \dots, N\}$ が成り立つ。この仮定の妥当性は、各マイクロホンの信号対雑音比 (signal-to-noise ratio: SNR) に依存するが、4.1 章で実際に測定し検証する。

この仮定に基づくと、分離行列のセミブラインドモデル $\tilde{\mathbf{W}}_f \in \mathbb{C}^{N \times M}$ は次式のように与えられる。

$$\tilde{\mathbf{W}}_f = \begin{bmatrix} \tilde{w}_{f1}^* & \tilde{w}_{f2}^* & \tilde{w}_{f3}^* & \dots & \tilde{w}_{fM}^* \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{w}}_f^H \\ \mathbf{e}_2^H \\ \mathbf{e}_3^H \\ \vdots \\ \mathbf{e}_N^H \end{bmatrix} \forall f. \quad (13)$$

式 (13) を ILRMA のコスト関数 (7) に代入し、 $\tilde{\mathbf{W}}_f$ の構造を用いて簡略化することで、Semi-blind ILRMA のコスト関数が次式のように得られる。

$$\mathcal{J} = -T \sum_f \log |\tilde{\mathbf{w}}_f^H \mathbf{e}_1|^2 + \sum_{f,t} \left[\frac{|\tilde{\mathbf{w}}_f^H \mathbf{x}_{ft}|^2}{\sum_k \tilde{b}_{fk} \tilde{v}_{kt}} + \log \sum_k \tilde{b}_{fk} \tilde{v}_{kt} \right] \quad (14)$$

ここで、 \tilde{b}_{fk} 及び \tilde{v}_{kt} はそれぞれ基底行列 $\tilde{\mathbf{B}} \in \mathbb{R}_+^{F \times K}$ 及びアクティベーション行列 $\tilde{\mathbf{V}} \in \mathbb{R}_+^{K \times T}$ の非負要素である。

提案手法では、 $\tilde{\mathbf{W}}_f$ の最初の行 ($\tilde{\mathbf{w}}_f$) のみが推定すべき変数となる。このベクトルは、Vo. 用マイクロホンの観測信号中の被り音を抑圧する空間分離フィルタに対応する。また、Vo. 音源のパワースペクトログラムは $\tilde{\mathbf{B}} \tilde{\mathbf{V}}$ でモデル化されるが、他の音源の TF モデルは推定されない。その結果、ILRMA の最適化変数は $\mathbf{W}_f, \mathbf{B}_n, \mathbf{V}_n \forall f, n$ から $\tilde{\mathbf{w}}_f, \tilde{\mathbf{B}}, \tilde{\mathbf{V}} \forall f$ に削減され、推定すべきパラメータ数は $1/N$ 倍となる。この削

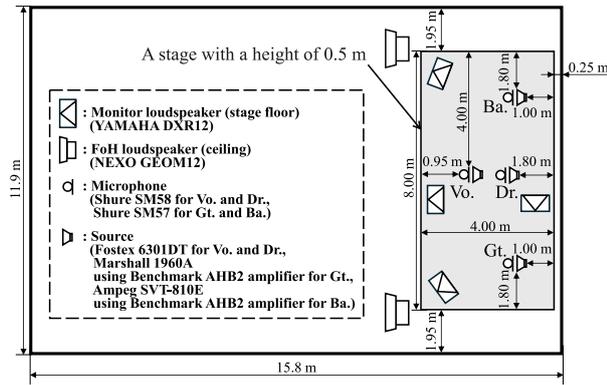


Fig. 3: Top-view schematic of the microphone and loudspeaker arrangement in a live music venue. Distances between each source and its corresponding microphone are set to 0.01 m.

減により、空間エイリアシングを含む観測信号に対しても頑健に $\tilde{\mathbf{w}}_f$ を推定できる。

$\tilde{\mathbf{w}}_f$, $\tilde{\mathbf{B}}$, 及び $\tilde{\mathbf{V}}$ の更新則は、ILRMA と同様の方法で次式のように導出される。

$$\tilde{\mathbf{U}}_f = \frac{1}{T} \sum_t \frac{1}{\sum_k \tilde{b}_{fk} \tilde{v}_{kt}} \mathbf{x}_{ft} \mathbf{x}_{ft}^H \quad (15)$$

$$\tilde{\mathbf{w}}_f \leftarrow \tilde{\mathbf{U}}_f^{-1} \mathbf{e}_1 \quad (16)$$

$$\tilde{\mathbf{w}}_f \leftarrow \frac{\tilde{\mathbf{w}}_f}{\sqrt{\mathbf{e}_1^H \tilde{\mathbf{w}}_f}} \quad (17)$$

$$\tilde{b}_{fk} \leftarrow \tilde{b}_{fk} \sqrt{\frac{\sum_t |\tilde{\mathbf{w}}_f^H \mathbf{x}_{ft}|^2 \tilde{v}_{kt} \left(\sum_{k'} \tilde{b}_{fk'} \tilde{v}_{k't} \right)^{-2}}{\sum_t \tilde{v}_{kt} \left(\sum_{k'} \tilde{b}_{fk'} \tilde{v}_{k't} \right)^{-1}}} \quad (18)$$

$$\tilde{v}_{kt} \leftarrow \tilde{v}_{kt} \sqrt{\frac{\sum_f |\tilde{\mathbf{w}}_f^H \mathbf{x}_{ft}|^2 \tilde{b}_{fk} \left(\sum_{k'} \tilde{b}_{fk'} \tilde{v}_{k't} \right)^{-2}}{\sum_f \tilde{b}_{fk} \left(\sum_{k'} \tilde{b}_{fk'} \tilde{v}_{k't} \right)^{-1}}} \quad (19)$$

4 被り音抑圧実験

4.1 音楽ライブ演奏環境のインパルス応答測定

現実のライブステージ環境を模擬するため、実際の音楽ライブ演奏環境でインパルス応答の測定を行った。Fig. 3 に収録環境の上面図を示す。4 音源 4 マイクホンの構成を仮定し、各音源から各マイクロホンまでのインパルス応答を測定した。各音源の HA ゲインは、プロの SR エンジニアが実際の楽器を用いて調整し、Vo., Gt., Ba., 及び Dr. に対してそれぞれ 26 dB, 20 dB, 18 dB, 及び 0 dB に設定された。

また、Fig. 4 に示す 2 条件 (Case A 及び Case B) で収録を行った。Case A では、モニタースピーカ及び FoH スピーカを使用せず、ステージ上の純粋なインパルス応答を測定した。一方、Case B ではより現実的な状況を模擬し、Vo. 用、Gt. 用、及び Ba. 用の 3 個のマイクロホンの観測音の混合信号をモニタースピーカから等しい音量でリアルタイムに出力し、FoH スピーカからは全マイクロホンの観測音の混合信号をリアルタイムに出力した。ただし、各スピーカの音量は SR エンジニアによって調整された。これらスピーカの出力音も、Fig. 1 に示すように各マイクロホンに被り音として観測される。

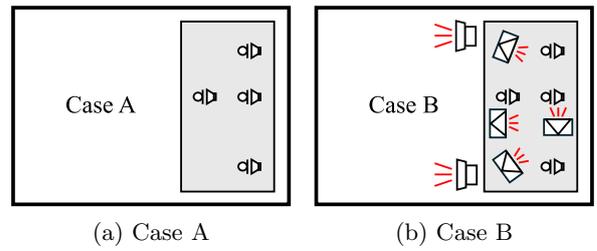


Fig. 4: Two recording conditions for the impulse response measurement: (a) all loudspeakers are muted and (b) all loudspeakers are active.

Vo. 音源から Vo. 用マイクロホンへのインパルス応答を用いて残響時間を算出した結果、Case A は $T_{60} = 760$ ms, Case B は $T_{60} = 775$ ms であった。Tables 1 及び 2 は、各マイクロホンで観測された音源毎の相対的な被り音のエネルギーを示す。表中の値は目的音源のエネルギーで正規化されているため、対角成分は常に 0 dB となっている。この結果から、Vo. 用マイクロホンは被り音が深刻であり、特に Dr. 音源成分は 0 dB を超えるエネルギーの被り音が観測されている。すなわち、Vo. 用マイクロホンには、Vo. 音源よりも Dr. 音源の方が大きなエネルギーで観測されることを意味している。一方、Gt., Ba., 及び Dr. 用マイクロホンは被り音のエネルギーが小さく、各目的音源は Case A 及び Case B でそれぞれ 25 dB 及び 16 dB 以上の SNR マージンを確保しており、3.1 節で述べたセミブラインド仮定の妥当性を示している。

4.2 4 音源混合における性能比較

提案手法の有効性を検証するため、TCNMF [2], 通常の ILRMA [4], 及び Semi-blind ILRMA を 4 音源 4 チャンルの観測信号に適用し、被り音抑圧性能を比較した。被り音抑圧の評価指標として、Vo. 音源の信号対歪み比 (source-to-distortion ratio: SDR) [8] の改善量を用いた。観測信号は 4.1 節で測定したインパルス応答を用いて作成した。ドライソース信号には DSD100 [9] から選んだ 3 曲 (nos. 022, 023, 及び 040) を使用した。DSD100 は Vo., Ba., Dr., 及びその他 (選択した 3 曲では Gt.) の音源信号が含まれており、本実験では各音源信号を 20 s の長さでカットした信号をドライソース信号とした。

本実験では、手法毎に STFT における最適な窓長が異なるため、各手法で最適な性能が得られる値 (Table 3) の結果について示す。TCNMF におけるハイパーパラメタ (κ, θ, α) (κ 及び θ はガンマ分布の形状及び尺度パラメタ, α は正規化パラメタ [2]) も事前に実験的に探索し、最良の結果をもたらす条件として Case A では (1.1, 21.5, 0.00077), Case B では (2, 1, 0.0001) を用いた。その他の実験条件を Table 4 に示す。なお、いずれの手法も乱数初期値に依存して性能が変化するため、各手法で 30 種類の異なる乱数シードで実行した。

Fig. 5 は SDR 改善量のバイオリン図を示している。各バイオリンは計 90 種の結果 (3 曲・30 種の乱数シード) を含んでいる。Fig. 5 (a) に示す TCNMF は、最適化変数の初期値にほぼ依存しない結果となっており、バイオリン図に現れる 3 つの層がそれぞれ 3 曲に対応している。この現象については、文献 [2] で解析されている通りである。しかしながら、この TCNMF

Table 1: Observed relative energy [dB] in Case A

	Vo. src.	Gt. src.	Ba. src.	Dr. src.
Vo. mic.	0.0	-17.3	-12.9	1.2
Gt. mic.	-56.3	0.0	-25.3	-29.4
Ba. mic.	-59.7	-34.8	0.0	-31.2
Dr. mic.	-72.7	-44.6	-40.6	0.0

Table 2: Observed relative energy [dB] in Case B

	Vo. src.	Gt. src.	Ba. src.	Dr. src.
Vo. mic.	0.0	-9.8	-6.9	1.3
Gt. mic.	-31.5	0.0	-16.9	-19.8
Ba. mic.	-32.5	-21.5	0.0	-23.1
Dr. mic.	-43.9	-34.1	-31.4	0.0

Table 3: Best window lengths [ms] for each method

Case	Gamma-TCNMF	ILRMA	Semi-blind ILRMA
Case A	186	186	1115
Case B	1858	743	2229

の初期値頑健性は Fig. 5 (b) において失われている。これは、Case B ではモニータスピーカ及び FoH スピーカが稼働しているため、これらの仮想的な音源の存在が TCNMF の初期値頑健性を失わせたと推測される。一方、ILRMA 及び Semi-blind ILRMA の結果では、Case A 及び Case B の両条件において、提案手法である Semi-blind ILRMA が高い被り音抑圧性能を発揮しており、Semi-blind ILRMA が空間エイリアシングが生じている観測信号に対しても、頑健に分離フィルタを推定できることを示唆している。しかし、被り音抑圧の性能はドライソースに依存する傾向が確認でき、さらなる改良が求められる。

5 おわりに

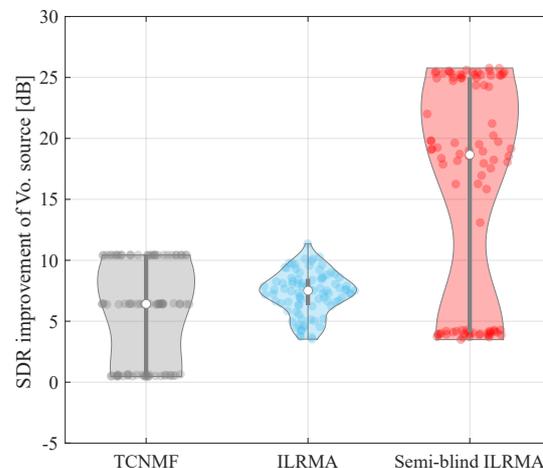
本稿では、Vo. 用マイクロホンの被り音抑圧に対して効果的な手法を提案した。提案手法は、Vo. 用以外のマイクロホンで得られた信号がそれらの音源のリファレンス信号として利用できるという仮定に基づいている。このセミブラインド分離モデルを ILRMA に導入することで、空間エイリアシングが生じる観測信号に対しても頑健に分離フィルタを推定できることが示された。

参考文献

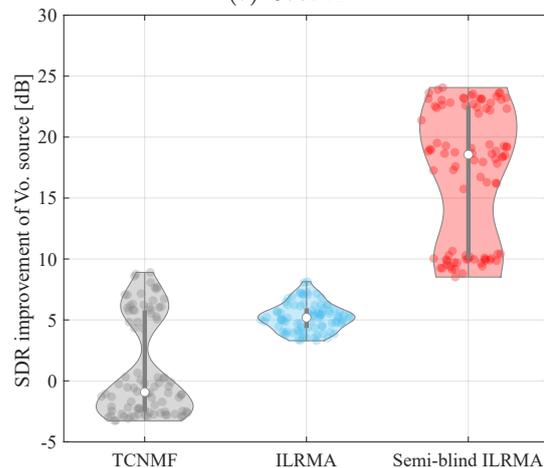
- [1] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
- [2] Y. Mizobuchi, D. Kitamura, T. Nakamura, N. Takamune, H. Saruwatari, Y. Takahashi, and K. Kondo, "Music bleeding-sound reduction based on time-channel nonnegative matrix factorization," *APSIPA Trans. Signal Inf. Process.*, 2025 (in press).
- [3] H. Sawada, N. Ono, H. Kameoka, D. Kitamura, and H. Saruwatari, "A review of blind source separation methods: Two converging routes to ILRMA originating from ICA and NMF," *APSIPA Trans. Signal Inf. Process.*, vol. 8, no. e12, pp. 1–14, 2019.
- [4] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source separation unifying independent vector analysis and non-negative matrix factorization," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 24, no. 9, pp. 1626–1641, 2016.
- [5] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source sep-

Table 4: Experimental conditions

Parameter	Condition
Window function in STFT	Blackman window
Window shift length in STFT	1/8 of window length
Number of iterations	100
Number of basis vectors for ILRMA and semi-blind ILRMA	$K = 10$



(a) Case A



(b) Case B

Fig. 5: Violin plot of SDR improvements of the Vo. source in four-source mixture case for (a) Case A and (b) Case B.

- aration with independent low-rank matrix analysis," in *Audio Source Separation*, S. Makino, Ed., pp. 125–155. Springer, Cham, 2018.
- [6] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, pp. 189–192, 2011.
- [7] M. Nakano, H. Kameoka, J. Le Roux, Y. Kitano, N. Ono and S. Sagayama, "Convergence-guaranteed multiplicative algorithms for nonnegative matrix factorization with β -divergence," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process.*, pp. 283–288, 2010.
- [8] 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.
- [9] A. Liutkus, F. R. Stöter, Z. Rafii, D. Kitamura, B. Rivet, N. Ito, N. Ono, and J. Fontcave, "The 2016 signal separation evaluation campaign," in *Proc. Int. Conf. Latent Var. Anal. Signal Separ.*, pp. 323–332, 2017.