# 基底共有型半教師あり独立低ランク行列分析に基づく多チャネル補聴器システム*

☆宇根昌和（筑波大学）, 久保優騎（東京大学）, 高宗典玄（東京大学）,
北村大地（香川高専）, 猿渡洋（東京大学）, 牧野昭二（筑波大学）

## 1 はじめに

補聴器を利用する場合，周囲の雑音の影響で目的音声の品質が劣化するため，目的音声の抽出処理が必要となる．補聴器システムにおける音声抽出処理には，目的話者の位置や空間情報が未知であっても頑健に動作するブラインド音源分離 (blind source separation: BSS) や一部の音源に教師データがある半教師あり音源分離が有効である．中でも独立低ランク行列分析 (independent low-rank matrix analysis: ILRMA) [1] は安定で高精度な音源分離を達成している．ILRMA を半教師あり音源分離に拡張した手法として，基底共有型 ILRMA (basis-shared ILRMA: BS-ILRMA) が提案されている [2]．BS-ILRMA は災害環境用のロボットのために提案された手法であり，ロボット自身が発するエゴノイズから生存者の声を分離する．目的音声の教師信号は得られないが，エゴノイズは前もって収録できるため，半教師ありの枠組みを使うことができる．補聴器を使う状況においても，会話直前の数秒の雑音区間など，事前に雑音のサンプルを利用することで，半教師あり音源分離である BS-ILRMA を利用可能となる．しかし，雑音のサンプル数などが大きく異なるため，補聴器システムに対する BS-ILRMA の有効性は明らかでない．

ILRMA や BS-ILRMA は線形時不変フィルタであるため，雑音が全方位から到来する拡散性雑音などの目的音源方位に雑音が存在する場合，その雑音は原理的に分離が不可能である．そこで，拡散性雑音が存在する状況を対象とした音声抽出法である，ランク制約付き空間共分散行列 (spatial covariance matrix: SCM) 推定法 [3] が提案されている．この手法は非負値行列因子分解 (nonnegative matrix factorization: NMF) [4] を多チャネル化した多チャネル NMF [5] と同様に，各音源の空間伝達特性を表現する SCM を推定する．ランク制約付き SCM 推定法は次の二段階の処理からなる．前段では ILRMA を用いて空間パラメータを推定し，後段では得られた空間パラメータから目的音源方位に残留する雑音を推定し抑圧する．多チャネル NMF は推定するパラメータの数が多く計算コストが大きい一方，ランク制約付き SCM 推定法は，ILRMA で推定された高精度な空間パラメータを用いることで推定すべきパラメータの数を削減しているため，多チャネル NMF より効率的かつ初期値に頑健な分離を可能にしている．

これまで我々は，スマートフォンを用いた分散マイクロホンアレー補聴器システムを提案している [6]．スマートフォンのマイクロホンを利用することで，マイクロホンの数が増えるだけでなく両耳から離れた距離にある空間の情報が得られるため，さらに高品質な分離が可能となる．上記の分散マイクロホンアレー補聴器システムにおいて，我々は前処理として ILRMA を用いたランク制約付き SCM 推定法における有効性を確認している．ランク制約付き SCM 推定法の前段に BS-ILRMA を用いて雑音の学習を取り入れることで，さらに高品質な分離ができると期待できる．本研究では，まず分散マイクロホンアレー補聴器システムのデータにおける BS-ILRMA の有効性を明らかにする．さらに，ランク制約付き SCM 推定法と BS-ILRMA を組み合わせることでより高品質な分離を達成することを示す．

## 2 定式化及び BSS 手法

### 2.1 定式化

$N$ 個の音源信号を $M$ 個のマイクロホンで収録し，観測した信号を分離することを考える．短時間フーリエ変換 (short-time Fourier transform: STFT) によって得られる複素時間周波数成分における音源信号，観測信号，及び分離信号をそれぞれ，$\boldsymbol{s}_{ij} = (s_{ij,1},\dots,s_{ij,n},\dots,s_{ij,N})$, $\boldsymbol{x}_{ij} = (x_{ij,1},\dots,x_{ij,m},\dots,x_{ij,M})$, 及び $\boldsymbol{y}_{ij} = (y_{ij,1},\dots,y_{ij,N})$ とする．ここで，$i = 1,\dots,I$, $j = 1,\dots,J$, $n = 1,\dots,N$, 及び $m = 1,\dots,M$ はそれぞれ周波数ビン，時間フレーム，音源信号，及び観測信号のインデクスである．各音源が点音源であり，STFT の窓長が残響時間より十分短いとする瞬時混合仮定では，各周波数ビンにおいて混合行列 $\boldsymbol{A}_i = (\boldsymbol{a}_{i,1}\cdots\boldsymbol{a}_{i,N}) \in \mathbb{C}^{M\times N}$ が存在し，次のように書ける．

$$\boldsymbol{x}_{ij} = \boldsymbol{A}_i \boldsymbol{s}_{ij} \tag{1}$$

ただし，$\boldsymbol{a}_{i,n}$ は周波数 $i$ における音源 $n$ のステアリングベクトルである．$M = N$ かつ $\boldsymbol{A}_i$ が正則である場合，$\boldsymbol{A}_i$ の逆行列 $\boldsymbol{W}_i = (\boldsymbol{w}_{i,1}\cdots\boldsymbol{w}_{i,N})^{\mathrm{H}} \in \mathbb{C}^{N\times M}$ を推定することで，次のように分離信号が得られる．

$$\boldsymbol{y}_{ij} = \boldsymbol{W}_i \boldsymbol{x}_{ij} \tag{2}$$

### 2.2 ILRMA [1]

ILRMA では，音源 $n$ の各時間周波数成分が

$$s_{ij,n} \sim \mathcal{N}_c\left(0, r_{ij,n}\right) \tag{3}$$

$$r_{ij,n} = \sum_l t_{il,n} v_{lj,n} \tag{4}$$

なる単変量複素ガウス分布に従い生起する確率生成モデルを仮定する．ここで，$t_{il,n} \geq 0$, $v_{lj,n} \geq 0$ は NMF における基底行列 $\boldsymbol{T}_n \in \mathbb{R}^{I\times L}$ とアクティベーション行列 $\boldsymbol{V}_n \in \mathbb{R}^{L\times J}$ の成分であり，$l = 1,\dots,L$ は基底のインデクス，$L$ は基底数である．また，$r_{ij,n}$ は音源 $n$ の音源モデルに相当し，NMF により低ランクの仮定が導入されている．式 (1) と多変量複素ガウ

*Multichannel hearing-aid system based on basis-shared semi-supervised independent low-rank matrix analysis by Masakzu Une (The University of Tsukuba), Yuki Kubo (The University of Tokyo), Norihiro Takamune (University of Tokyo), Daichi Kitamura (National Institute of Technology, Kagawa College), Hiroshi Saruwatari (The University of Tokyo), Shoji Makino (The University of Tsukuba).

ス分布の再生性より，$\boldsymbol{x}_{ij}$ も多変量複素ガウス分布

$$\boldsymbol{x}_{ij} \sim \mathcal{N}_c\left(\boldsymbol{0}, \sum_n r_{ij,n}\boldsymbol{a}_{i,n}\boldsymbol{a}_{i,n}^{\mathrm{H}}\right) \tag{5}$$

に従う．ILRMA のコスト関数 $\mathcal{J}_{\mathrm{ILRMA}}$ は観測信号の負対数尤度関数であり，次のように定義される．

$$\begin{aligned}\mathcal{J}_{\mathrm{ILRMA}} =& \sum_n^N \sum_{i,j} \left[\frac{|y_{ij,n}|^2}{r_{ij,n}} + \log r_{ij,n}\right] \\ & - 2J\sum_i \log|\det \boldsymbol{W}_i| + \mathrm{const.}\end{aligned} \tag{6}$$

NMF 変数 $t_{il,n}$，$v_{lj,n}$ 及び分離行列 $\boldsymbol{W}_i = \boldsymbol{A}_i^{-1}$ は尤度最大化により推定される．

### 2.3 ランク制約付き SCM 推定法 [3]

ランク制約付き SCM 推定法は 1 個の方向性目的音源と拡散性雑音が混合している状況を対象とした手法である．ランク制約付き SCM 推定法は二段階の処理からなる．前段では ILRMA などを用いて線形時不変フィルタを推定し，後段では得られた空間パラメータから目的音源方位に残留する雑音を推定し抑圧する．ILRMA を適用した $M$ 個の分離音のうち，方向性目的音が含まれる分離音には雑音が混入する一方，それ以外の $M-1$ 個の分離音はほぼ雑音のみで目的音の混入が非常に少ない [7]．この $M-1$ 個の雑音のみの信号から推定された SCM はランクが $M-1$ となる [3]．フルランク（即ちランク $M$）であることが期待される拡散性雑音の SCM の推定値は，ランクが 1 つ落ちている．目的音源方位の雑音を除去するためにはフルランクの SCM が必要なため，不足したランク 1 空間基底を加算するモデル化を行い，パラメータを推定する．最後に多チャネルウィーナフィルタを構成し，目的音源方位の拡散性雑音を低減する．

ランク制約付き SCM 推定法のモデルは観測信号 $\boldsymbol{x}_{ij}$ を目的音源のソースイメージ $\boldsymbol{h}_{ij} = (h_{ij,1},\ldots,h_{ij,M})^\top$ と拡散性音源のソースイメージ $\boldsymbol{u}_{ij} = (u_{ij,1},\ldots,u_{ij,M})^\top$ の和として次のように表す．

$$\boldsymbol{x}_{ij} = \boldsymbol{h}_{ij} + \boldsymbol{u}_{ij} \tag{7}$$

目的音源のソースイメージ $\boldsymbol{h}_{ij}$ は，ILRMA によって得られたステアリングベクトル $\boldsymbol{a}_{i,1},\ldots,\boldsymbol{a}_{i,N}$ のうち目的音源に対応するベクトル $\boldsymbol{a}_i^{(h)} =: \boldsymbol{a}_{i,n_h}$ と，目的音源のドライソース $s_{ij}^{(h)}$ を用いて次のように表す．

$$\boldsymbol{h}_{ij} = \boldsymbol{a}_i^{(h)} s_{ij}^{(h)} \tag{8}$$

$$s_{ij}^{(h)} \sim \mathcal{N}_c\left(0, r_{ij}^{(h)}\right) \tag{9}$$

ここで，$n_h$ は目的音源に対応する音源インデクス，$r_{ij}^{(h)}$ は目的音源の分散である．目的音源として音声を想定しているため，目的音源の分散 $r_{ij}^{(h)}$ はスパース性を有するとし，事前分布として逆ガンマ分布を仮定する．

$$p(r_{ij}^{(h)};\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\left(r_{ij}^{(h)}\right)^{-\alpha-1}\exp\left(-\frac{\beta}{r_{ij}^{(h)}}\right) \tag{10}$$

ここで，$\alpha > 0$ は形状母数，$\beta > 0$ は尺度母数，$\Gamma(\cdot)$ はガンマ関数を表す．

一方，拡散性音源のソースイメージ $\boldsymbol{u}_{ij}$ は目的音源のソースイメージ $\boldsymbol{h}_{ij}$ とは独立な多変量複素ガウス分布に従うと仮定する．

$$\boldsymbol{u}_{ij} \sim \mathcal{N}_c\left(\boldsymbol{0}, r_{ij}^{(u)}\boldsymbol{R}_i^{(u)}\right) \tag{11}$$

ここで，$r_{ij}^{(u)}$ と $\boldsymbol{R}_i^{(u)}$ はそれぞれ拡散性音源の分散と SCM である．ILRMA によって推定された分離フィルタ $\boldsymbol{w}_{i,n}$ を用いて，拡散性音源の SCM $\boldsymbol{R}_i^{(u)}$ は次のようにモデル化される．

$$\boldsymbol{R}_i^{(u)} = \boldsymbol{R'}_i^{(u)} + \lambda_i \boldsymbol{b}_i \boldsymbol{b}_i^{\mathrm{H}} \tag{12}$$

$$\begin{aligned}\boldsymbol{R'}_i^{(u)} =& \frac{1}{J}\sum_j \boldsymbol{W}_i^{-1}(|\boldsymbol{w}_{i,1}^{\mathrm{H}}\boldsymbol{x}_{ij}|^2,\ldots,|\boldsymbol{w}_{i,n_h-1}^{\mathrm{H}}\boldsymbol{x}_{ij}|^2, 0, \\ & |\boldsymbol{w}_{i,n_h+1}^{\mathrm{H}}\boldsymbol{x}_{ij}|^2,\ldots,|\boldsymbol{w}_{i,N}^{\mathrm{H}}\boldsymbol{x}_{ij}|^2)(\boldsymbol{W}_i^{-1})^{\mathrm{H}}\end{aligned} \tag{13}$$

ここで，$\boldsymbol{R'}_i^{(u)}$ は ILRMA によって推定された雑音のランク $M-1$ SCM であり，$\boldsymbol{b}_i$ は $\boldsymbol{R'}_i^{(u)}$ の零固有値に対応する単位固有ベクトル，$\lambda_i$ は補完される成分の大きさを表す．ここで，$\boldsymbol{R}_i^{(u)}$ において推定すべき変数は $\lambda_i$ だけであり，ILRMA によって推定された $\boldsymbol{R'}_i^{(u)}$ と $\boldsymbol{b}_i$ を固定して最適化を行う．以上より，式 (10) の目的音源の分散の事前分布を考慮したランク制約付き SCM モデルの負対数事後確率 $\mathcal{L}$ は次のように表される．

$$\begin{aligned}\mathcal{L}(r_{ij}^{(h)}, r_{ij}^{(u)}, \lambda_i) =& \sum_{i,j}\Bigg[\boldsymbol{x}_{ij}^{\mathrm{H}}(\boldsymbol{R}_{ij}^{(x)})^{-1}\boldsymbol{x}_{ij} + \log\det\boldsymbol{R}_{ij}^{(x)} \\ & + (\alpha+1)\log r_{ij}^{(h)} + \frac{\beta}{r_{ij}^{(h)}}\Bigg] + \mathrm{const.}\end{aligned} \tag{14}$$

$$\boldsymbol{R}_{ij}^{(x)} = r_{ij}^{(h)}\boldsymbol{a}_i^{(h)}\boldsymbol{a}_i^{(h)\mathrm{H}} + r_{ij}^{(u)}\boldsymbol{R}_i^{(u)} \tag{15}$$

変数 $r_{ij}^{(h)}$，$r_{ij}^{(u)}$，$\lambda_i$ は、$\mathcal{L}$ を最小化するように expectation-maximization アルゴリズムによって最適化される [3]．

## 3 BS-ILRMA [2] の分散マイクロホンアレー補聴器システムへの適用

BS-ILRMA は人の進入が困難な災害環境で生存者の声を検知するロボットのための手法として提案された．BS-ILRMA は前もって収録されたロボットのエゴノイズを利用した半教師あり音源分離である．一方，補聴器を使う状況においても，会話直前の数秒の雑音区間など事前に雑音のサンプルを利用することで半教師あり音源分離を適用することが可能である．

BS-ILRMA は $M = N$ の条件のうち，$N' = M' = N$–1 個の雑音源と 1 個の目的音源が存在する状況を仮定している．事前に得られる $M'$ チャネルの雑音のサンプルを $\boldsymbol{x}_{ij'}^{(\mathrm{noise})} = (x_{ij',1}^{(\mathrm{noise})},\ldots,x_{ij',m'}^{(\mathrm{noise})},\ldots,x_{ij',M'}^{(\mathrm{noise})})$，分離対象である $M$ チャネルの観測信号を $\boldsymbol{x}_{ij}^{(\mathrm{mix})} = (x_{ij,1}^{(\mathrm{mix})},\ldots,x_{ij,M}^{(\mathrm{mix})})^\top$ と表す．ここで，$j' = 1,\ldots,J'$ 及び $m' = 1,\ldots,M'$ は雑音サンプルの観測信号のフレーム及びインデクスである．

単純に ILRMA を半教師ありにする場合，半教師あり NMF [8] と ILRMA を組み合わる手法が考えられる．即ち，雑音のサンプル $\boldsymbol{x}_{ij'}^{(\mathrm{noise})}$ を ILRMA に適用し，$N'$ 個の雑音の基底行列 $\boldsymbol{T}_{n'}^{(\mathrm{noise})} \in \mathbb{R}_{\geq 0}^{I\times L}$ を学

習させ，学習済みの基底行列を用いて，もう一つの ILRMA で $M$ チャネルの観測信号 $\boldsymbol{x}_{M}^{(\mathrm{mix})}$ を分離する方法 (semi-supervised ILRMA: SS-ILRMA) である．ここで，$n'=1,\ldots,N'$ は雑音サンプルの音源のインデクスである．しかし，単純な半教師ありアプローチでは $\boldsymbol{W}_i$ と基底行列の間のスケールの不定性が雑音の教師基底行列 $\boldsymbol{T}_{n'}^{(\mathrm{noise})}$ のスペクトル構造を崩壊させる可能性がある [2]．

この問題に対処するため BS-ILRMA が提案された．BS-ILRMA の概要を Fig. 1 に示す．ここで，$\boldsymbol{W}_i^{(\mathrm{noise})} \in \mathbb{C}^{N'\times M'}$ 及び $\boldsymbol{W}_i^{(\mathrm{mix})} \in \mathbb{C}^{N\times M}$ は，雑音サンプル $\boldsymbol{x}_{ij'}^{(\mathrm{noise})}$ 及び観測信号 $\boldsymbol{x}_{ij}^{(\mathrm{mix})}$ に対する分離行列である．$\boldsymbol{X}_{m'}^{(\mathrm{noise})} \in \mathbb{C}^{I\times J'}$ 及び $\boldsymbol{Y}_{n'}^{(\mathrm{noise})} \in \mathbb{C}^{I\times J'}$ はそれぞれ $\boldsymbol{x}_{ij'}^{(\mathrm{noise})}$ 及び $\boldsymbol{y}_{ij'}^{(\mathrm{noise})} = (y_{ij',1}^{(\mathrm{noise})},\ldots,y_{ij',N'}^{(\mathrm{noise})})^\top$ の $m'$ 及び $n'$ 番目のスペクトログラムである．$\boldsymbol{X}_{m}^{(\mathrm{mix})} \in \mathbb{C}^{I\times J}$ 及び $\boldsymbol{Y}_{n}^{(\mathrm{mix})} \in \mathbb{C}^{I\times J}$ はそれぞれ $\boldsymbol{x}_{ij}^{(\mathrm{mix})}$ 及び $\boldsymbol{y}_{ij}^{(\mathrm{mix})} = (y_{ij,1}^{(\mathrm{mix})},\ldots,y_{ij,N}^{(\mathrm{mix})})^\top$ の $m$ 及び $n$ 番目のスペクトログラムである．また，$|\cdot|^{\cdot 2}$ は要素毎の 2 乗を表す．$\boldsymbol{T}_{n'} \in \mathbb{R}_{\geq 0}^{I\times L}$ は雑音サンプルの音源に対する共有基底行列，$\boldsymbol{T}_N \in \mathbb{R}_{\geq 0}^{I\times L}$ は目的音源に対する非共有基底行列，$\boldsymbol{V}_{n'}^{(\mathrm{noise})} \in \mathbb{R}_{\geq 0}^{L\times J'}$ 及び $\boldsymbol{V}_{n}^{(\mathrm{mix})} \in \mathbb{R}_{\geq 0}^{L\times J}$ は $\boldsymbol{Y}_{n'}^{(\mathrm{noise})}$ 及び $\boldsymbol{Y}_{n}^{(\mathrm{mix})}$ に対応するアクティベーション行列である．BS-ILRMA は二つの ILRMA を使用する．一方では $\boldsymbol{W}_i^{(\mathrm{noise})}$ 及び $\boldsymbol{y}_{ij'}^{(\mathrm{noise})}$ を推定するため，ILRMA を雑音サンプル $\boldsymbol{x}_{ij'}^{(\mathrm{noise})}$ に適用する．もう一方では $\boldsymbol{W}_i^{(\mathrm{mix})}$ 及び $\boldsymbol{y}_{ij}^{(\mathrm{mix})}$ を推定するため，ILRMA を観測信号 $\boldsymbol{x}_{ij}^{(\mathrm{mix})}$ に適用する．最も重要な点は，雑音サンプルの音源に対する基底行列 $\boldsymbol{T}_{n'}$ は二つの ILRMA 間で共有されており，これらのモデルにおける全ての変数は同時に最適化されていることである．共有基底行列 $\boldsymbol{T}_{n'}$ は $\boldsymbol{x}_{ij'}^{(\mathrm{noise})}$ 及び $\boldsymbol{x}_{ij}^{(\mathrm{mix})}$ の両方で類似したスペクトルを表さなければならないため，雑音サンプルのスペクトルパターンは $\boldsymbol{T}_{n'}$ によって捉えられ，基底行列 $\boldsymbol{T}_N$ は結果的に残った目的音源のスペクトルパターンを表現する．

BS-ILRMA のコスト関数は，二つの ILRMA のコストの和として次のように定義される．

$$
\begin{aligned}
\mathcal{J} = & \frac{1}{N'}\left\{\sum_{n'=1}^{N'}\sum_{i,j'}\left[\frac{|y_{i,j',n'}^{(\mathrm{noise})}|^2}{\sum_l t_{il,n'}v_{lj',n'}^{(\mathrm{noise})}} + \log\sum_l t_{il,n'}v_{lj',n'}^{(\mathrm{noise})}\right]\right. \\
& \left. - 2J'\sum_i \log|\det \boldsymbol{W}_i^{(\mathrm{noise})}|\right\} \\
& + \frac{1}{N}\left\{\sum_{n=1}^{N}\sum_{i,j}\left[\frac{|y_{i,j,n}^{(\mathrm{mix})}|^2}{\sum_l t_{il,n}v_{lj,n}^{(\mathrm{mix})}} + \log\sum_l t_{il,n}v_{lj,n}^{(\mathrm{mix})}\right]\right. \\
& + \sum_{i,j}\left[\frac{|y_{i,j,N}^{(\mathrm{mix})}|^2}{\sum_l t_{il,N}v_{lj,N}^{(\mathrm{mix})}} + \log\sum_l t_{il,N}v_{lj,N}^{(\mathrm{mix})}\right] \\
& \left. - 2J\sum_i \log|\det \boldsymbol{W}_i^{(\mathrm{mix})}|\right\} \qquad (16)
\end{aligned}
$$

ここで，$t_{il,n'}$ 及び $t_{il,N}$ はそれぞれ $\boldsymbol{T}_{n'}$ 及び $\boldsymbol{T}_N$ の要素，$\boldsymbol{v}_{lj,n'}^{(\mathrm{noise})}$ 及び $\boldsymbol{v}_{lj,n}^{(\mathrm{mix})}$ はそれぞれ $\boldsymbol{V}_{n'}^{(\mathrm{noise})}$ 及び $\boldsymbol{V}_{n}^{(\mathrm{mix})}$ の要素を表す．共有されないパラメータの更新式は [1] と同じである．一方，共有基底 $t_{il,n'}$ に関して式 (16) を直接最小化することは困難であるため，

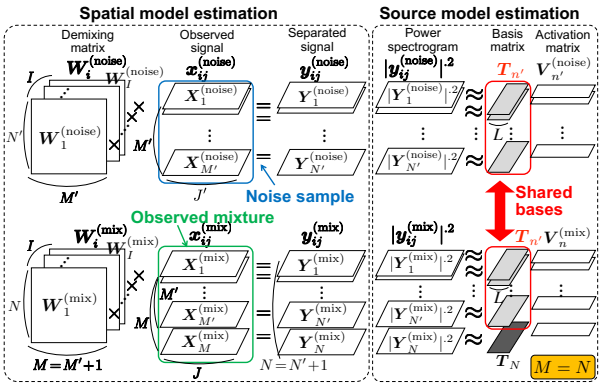


Fig. 1　Overview of BS-ILRMA, where upper and lower models are *simultaneously* optimized.

補助関数法に基づき補助関数を設計し，これを最小化することで局所最適解を得る [2]．

BS-ILRMA はロボットのエゴノイズと目的音源を分離するタスクにおいて，有効な手法であることが明らかにされている [2]．ランク制約付き SCM 推定法は目的音方位の雑音を抑圧するため，処理の前段として ILRMA によって空間パラメータを推定する．本研究では，ランク制約付き SCM 推定法の前段部分に BS-ILRMA を適用した手法を提案する．BS-ILRMA は ILRMA と同様に線形フィルタであるが，ILRMA に比べて高い性能が期待できるため，ランク制約付き SCM 推定法の前処理として BS-ILRMA を用いて空間パラメータを推定することで，さらに高品質な音声抽出処理が行えると考えられる．

## 4　評価実験

### 4.1　BS-ILRMA の性能評価

本研究の目的は分散マイクロホンアレー補聴器システムで収録したデータに対して，ランク制約付き SCM 推定法の初期化方法に BS-ILRMA を用いた場合の分離性能を評価することである．そこでまず，分散マイクロホンアレー補聴器システムに対する BS-ILRMA の有効性を確認する．分散マイクロホンアレー補聴器システムで収録したデータに対して，通常の ILRMA，雑音サンプルを用いて事前に基底を学習して分離を行う SS-ILRMA，及び基底共有により雑音学習用と分離用のモデルを同時に最適化する BS-ILRMA を比較した．残響時間 300 ms の室内に，片耳に 3 つ，スマートフォンに 2 つ，計 8 つのマイクロホンを装備したダミーヘッドを置き，収録したインパンルス応答と拡散性雑音を用いて評価した [6]．ダミーヘッドから目的音源までの距離を 75 cm，100 cm，150 cm，また正面方位を 0°，左側をマイナスとして，角度を −20°，0°，20° に変更して収録した．音声データベース JNAS [9] の女声データ 1 文に収録したインパルス応答を 16 kHz にダウンサンプリングし畳み込んだものを目的信号とした．目的音声が発話される直前の 2 秒間に雑音区間を設け，この 2 秒の雑音区間を学習に用いた．STFT において窓長 64 ms のハミング窓を 32 ms シフトで用いた．実験するにあたり，入力 SNR は −10 dB, −5 dB, 0 dB の 3 通りを用い，ILRMA，SS-ILRMA，BS-ILRMA の基底数は 10，更新回数はそれぞれ 50 回とした．主成分分析を用いて観測信号

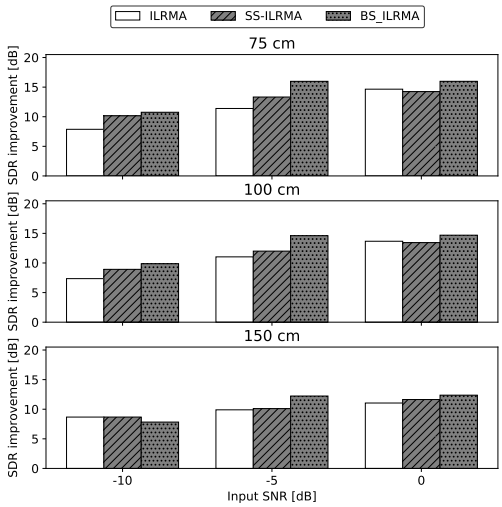


Fig. 2 Average SDR improvements under each input SNR condition.

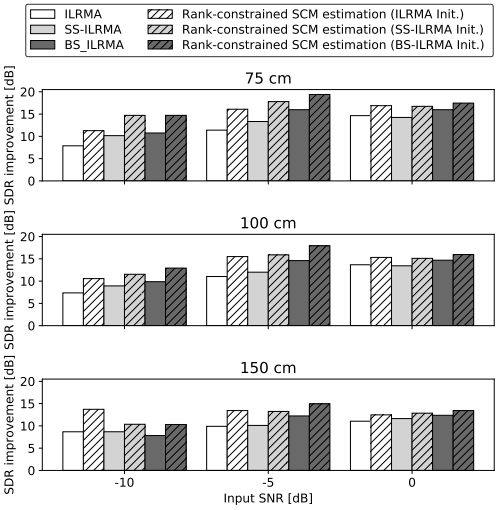


Fig. 3 Average SDR improvements of rank-constrained SCM estimation initialized by ILRMA, SS-ILRMA and BS-ILRMA, where number of iterations of rank-constrained SCM estimation was two.

の白色化を行い，分離行列の初期化方法は単位行列とした．基底行列とアクティベーション行列は一様乱数により初期化した．SS-ILRMA については雑音の事前学習として，基底を 50 回更新したものを使用した．異なる乱数初期値で 10 回試行した．以上の条件で，評価尺度として source-to-distortion ratio (SDR) 改善量 [10] を用いて，右耳外耳道付近のマイクロホンでの各角度及び異なる乱数初期値での結果を平均して比較した．

結果を Fig. 2 に示す．ほぼ全ての場合において，BS-ILRMA が通常の ILRMA や SS-ILRMA に比べて SDR 改善量が高い．このことから，ランク制約付き SCM 推定法の初期化方法に BS-ILRMA を用いることで従来に比べ高品質な分離ができると考えられる．

### 4.2 ランク制約付き SCM 推定法へ適用した場合の性能評価

4.1 節の結果を踏まえ，次に ILRMA，SS-ILRMA，BS-ILRMA のそれぞれを初期化方法としてランク制約付き SCM 推定法を適用した場合の分離性能について調査する．実験データや ILRMA，SS-ILRMA，BS-ILRMA の条件は 4.1 節と同様である．ランク制約付き SCM 推定法における形状母数パラメータ $\alpha$ は 20 とし，尺度母数パラメータ $\beta$ は $10^{-16}$ とした．また少ない反復回数で高い分離性能を達成することが分かっているため，ランク制約付き SCM 推定法の反復回数が 2 回目の結果を用いて比較した [6]．

ILRMA, SS-ILRMA, BS-ILRMA と各手法を初期値とした場合のランク制約付き SCM 推定法の結果を Fig. 3 に示す．結果から全ての場合においてランク制約付き SCM 推定法を適用した場合に SDR 改善量が向上している．中でも，BS-ILRMA を初期値とした場合の SDR 改善量が他の場合と比べて高いため，ランク制約付き SCM 推定法の初期化方法に BS-ILRMA を用いることでより高い分離性能を達成できることが分かる．

## 5 おわりに

本研究では分散マイクロホンアレー補聴器システムにおける，BS-ILRMA の有効性および，ランク制約付き SCM 推定法の初期化方法として利用した場合の有効性について調査した．実験結果から，補聴器システムに対しても ILRMA や SS-ILRMA と比較して BS-ILRMA は分離性能の点で有効であり，ランク制約付き SCM 推定法の初期化方法として用いることで高い分離性能を達成することが分かった．今後は，ランク制約付き SCM 推定法の半教師ありへの拡張を目指す．

**謝辞** 本研究の一部は，セコム科学技術振興財団，JSPS 科研費 19H01116 及び 19K20306 の助成を受けたものである．

## 参考文献

[1] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization," *IEEE/ACM Trans. on ASLP*, vol. 24, no. 9, pp. 1626–1641, 2016.

[2] M. Takakusaki, D. Kitamura, N. Ono, T. Yamada, S. Makino, and H. Saruwatari, "Ego-noise reduction for a hose-shaped rescue robot using basis shared semi-supervised independent low-rank matrix analysis," in *Proc. NCSP*, 2018, pp. 351–354.

[3] Y. Kubo, N. Takamune, D. Kitamura, and H. Saruwatari, "Efficient full-rank spatial covariance estimation using independent low-rank matrix analysis for blind source separation," in *Proc. EUSIPCO*, 2019, pp. 1814–1818.

[4] 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.

[5] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, "Multichannel extensions of non-negative matrix factorization with complex valued data," *IEEE Trans. ASLP*, vol. 21, no. 5, pp. 971–982, 2013.

[6] M. Une, Y. Kubo, N. Takamune, D. Kitamura, H. Saruwatari, and S. Makino, "Evaluation of multichannel hearing aid system using rank-constrained spatial covariance matrix estimation," in *Proc. APSIPA*, 2019, pp. 1874–1879.

[7] Y. Takahashi, T. Takatani, K. Osako, H. Saruwatari, and K. Shikano, "Blind spatial subtraction array for speech enhancement in noisy environment," *IEEE Trans. ASLP*, vol. 17, no. 4, pp. 650–664, 2009.

[8] P. Smaragdis, B. Raj, and M. Shashanka, "Supervised and semi-supervised separation of sounds from single-channel mixtures," in *Proc. ICA*, 2007, pp. 414–421.

[9] K. Itou, M. Yamamoto, K. Takeda, T. Takezawa, T. Matsuoka, T. Kobayashi, K. Shikano, and S. Itahashi, "JNAS: Japanese speech corpus for large vocabulary continuous speech recognition research," *The Journal of Acoustical Society of Japan (E)*, vol. 20, no. 3, pp. 199–206, 1999.

[10] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Trans. ASLP*, vol. 14, no. 4, pp. 1462–1469, 2006.