

# インパルス応答のスパース性を利用したブラインド音源分離

小田 亮太<sup>†</sup> 北原 大地<sup>†</sup> 平林 晃<sup>†</sup>

<sup>†</sup> 立命館大学情報理工学部 〒525-8577 滋賀県草津市野路東1-1-1

E-mail: <sup>†</sup> d-kita@fc.ritsumei.ac.jp, akirahrb@media.ritsumei.ac.jp

あらまし ブラインド音源分離手法の独立低ランク行列分析法 (ILRMA) に混合系の先験情報を与えることにより、より高精度の分離を行う手法を提案する。ILRMA は高精度の分離手法であるが、初期値によっては局所解に陥り、十分な分離性能を得られないことがあった。この問題を解決する為に本手法では、混合系のインパルス応答のスパース性を利用する。公開されているインパルス応答を用いた計算機シミュレーションにより、提案法の有効性を示す。  
キーワード ブラインド音源分離, 独立低ランク行列分析, インパルス応答, スパース性

## Blind Source Separation Based on the Sparsity of Impulse Responses

Ryota ODA<sup>†</sup>, Daichi KITAHARA<sup>†</sup>, and Akira HIRABAYASHI<sup>†</sup>

<sup>†</sup> College of Information Science and Technology, Ritsumeikan University

1-1-1 Nojihigashi, Kusatsu, Shiga, 525-8577 Japan

**Abstract** We propose a blind source separation (BSS) algorithm using *a priori* information of the mixing process based on the state-of-the-art method for BSS, called independent low-rank matrix analysis (IRLMA). Even though IRLMA can provide the best source separation performance, it can be sometimes degraded by an inappropriate initial value for the algorithm. To solve this problem, we exploit sparsity of impulse responses of the mixing process. Simulations using publicly available impulse responses show that the proposed method outperforms IRLMA irrespective of evaluation measures, such as SDR, SIR, or SAR.

**Key words** Blind Source Separation, Independent Low-Rank Matrix Analysis, Impulse Response, Sparsity

### 1. はじめに

音源分離とは、複数の音源が混合された観測信号から、混合以前の信号を推定し復元する技術である。音源の空間的な配置が未知の場合、音源間の統計的独立性を利用する多チャンネル過決定音源分離手法として、周波数ビン毎に分離系（線形分離行列）を推定する周波数領域独立成分分析 (frequency-domain independent component analysis: FDICA) [2] に基づく手法が多数検討されている。この手法は周波数ビン毎の分離信号の順序が定まらないために、複雑な後処理が必要となる [3]。このため近年は、より高度な音源モデルを用いることで、高精度の音源分離を実現するようになっている。すなわち、独立ベクトル分析 (independent vector analysis: IVA) [4] は、周波数ビン同士に高次相関が存在することを意味する音源モデルを利用している。また、独立低ランク行列分析 (independent low-rank matrix analysis: ILRMA) [8] は非負値行列因子分解 (nonnegative matrix factorization: NMF) に基づく音源モデルを利用している。とりわけ ILRMA は、FDICA や IVA よりも高い音源分離性能を発揮することが報告されている。しかし従来の ILRMA では、まとまった帯域ブロックごとにパーミュ

テーション不整合が生じる問題や、NMF による音源モデルが適切に学習されないことにより、品質があまり良くない局所解に陥る問題などがある。これらの問題を解決する為に、本論文では混合系に関する前提知識を利用する。

この文脈での研究には、マイクロホンアレイを用いた音源分離において、音源の到来方向を利用したものがある [12]–[15]。これらの方法では、ビームフォーミングに基づいたステアリングベクトルを利用して、混合行列の推定精度を向上させている。しかし、通常のブラインド音源分離では、音源の到来方向は未知であることが多い。また、音源の到来方向が既知であったとしても、マイクが適切に並べられていなければ、上記の手法は適用できない。

そこで本研究では、音源の混合過程に対するより一般的な性質である、各音源とマイク間のインパルス応答のスパース性を利用する。実際、室内のインパルス応答を推定する研究では、インパルス応答のスパース性やブロックスパース性に基づいたコスト関数を用いることで、推定精度を改善できることが報告されている [16]–[18]。この考えを利用し、従来の最尤推定に基づくコスト関数にインパルス応答のスパース性を評価する指標を導入したコスト関数を定義し、これを最小化することによっ

て音源を分離する手法を提案する。この最適化問題は、分離行列に対して先験情報を新たに付与していることになる。したがって提案法は、従来の最尤推定に基づく手法と違い、ベイズ推定に基づく手法と解釈できる。分離行列の更新には偏微分を用い、インパルス応答の更新には重み付き  $l_0$  ノルムを用いる。これらの更新を交互に行うことで、提案法における最小化問題を解く。数値実験により、提案手法が音楽信号、音声信号のいずれにおいても、従来手法よりも高い分離性能が得られることを示す。

## 2. 過決定条件下におけるブラインド音源分離

### 2.1 ブラインド音源分離問題

音源数を  $N$  ( $\geq 2$ ) とし、マイク数を  $M$  ( $\geq 1$ ) とする。時刻  $t$  における  $n$  番目の音源の生成音を  $s_n[t]$  とし、 $m$  番目のマイクの収録音を  $x_m[t]$  とする。 $n$  番目の音源から  $m$  番目のマイクへのインパルス応答を  $h_{m,n}[\tau]$  ( $\tau = 0, 1, \dots, T-1$ ) で表せば、

$$x_m[t] = \sum_{n=1}^N (h_{m,n} * s_n)[t] = \sum_{n=1}^N \sum_{\tau=0}^{T-1} h_{m,n}[\tau] s_n[t-\tau] \quad (1)$$

が成立する。ここで、 $*$  は畳み込み演算である。 $s_n[t]$  と  $x_m[t]$  をフレーム長  $L$  で短時間フーリエ変換<sup>(注1)</sup>したものをそれぞれ  $\hat{s}_{i,j,n}$  と  $\hat{x}_{i,j,n}$  ( $i = 1, 2, \dots, L; j = 1, 2, \dots, J$ ) で表す。また、 $h_{m,n}[\tau]$  を離散時間フーリエ変換<sup>(注2)</sup>したものを  $\hat{h}_{m,n}(\omega)$  とし、 $\hat{a}_{i,m,n} := \hat{h}_{m,n}(\frac{2\pi(i-1)}{L})$  ( $i = 1, 2, \dots, L$ ) とサンプリングする。短時間フーリエ変換のフレーム長  $L$  がインパルス応答の長さ  $T$  に比べて十分大きいとき、各フレームで式 (1) の「時間領域の畳み込み演算」を「周波数領域の乗算」に変換できる [1] ため、 $i = 1, 2, \dots, \lfloor \frac{L+2}{2} \rfloor =: I$  と  $j = 1, 2, \dots, J$  に対して<sup>(注3)</sup>,

$$\hat{x}_{i,j,m} = \sum_{n=1}^N \hat{a}_{i,m,n} \hat{s}_{i,j,n} + \hat{\epsilon}_{i,j,m} \quad (2)$$

が成立する。ここで、 $\hat{\epsilon}_{i,j,m}$  は短時間フーリエ変換で得られる周波数成分と真の周波数成分の差に起因するモデル誤差である。ベクトル  $\mathbf{s}_{i,j}$ ,  $\mathbf{x}_{i,j}$ ,  $\boldsymbol{\epsilon}_{i,j}$ ,  $\mathbf{a}_{i,j}$  と行列  $A_i$  を

$$\begin{cases} \mathbf{s}_{i,j} := (\hat{s}_{i,j,1}, \hat{s}_{i,j,2}, \dots, \hat{s}_{i,j,N})^T \in \mathbb{C}^N \\ \mathbf{x}_{i,j} := (\hat{x}_{i,j,1}, \hat{x}_{i,j,2}, \dots, \hat{x}_{i,j,M})^T \in \mathbb{C}^M \\ \boldsymbol{\epsilon}_{i,j} := (\hat{\epsilon}_{i,j,1}, \hat{\epsilon}_{i,j,2}, \dots, \hat{\epsilon}_{i,j,M})^T \in \mathbb{C}^M \\ \mathbf{a}_{i,n} := (\hat{a}_{i,1,n}, \hat{a}_{i,2,n}, \dots, \hat{a}_{i,M,n})^T \in \mathbb{C}^M \\ A_i := (\mathbf{a}_{i,1}, \mathbf{a}_{i,2}, \dots, \mathbf{a}_{i,N}) \in \mathbb{C}^{M \times N} \end{cases}$$

(注1) : 虚数単位を  $i \in \mathbb{C}$  とする。離散時間信号  $s_n[t]$  の短時間フーリエ変換は

$$\hat{s}_{i,j,n} := \sum_{\tau=0}^{L-1} \psi[\tau] s_n[(j-1)\eta + \tau] \exp\left(-i \frac{2\pi(i-1)\tau}{L}\right)$$

と定義される。 $L$  はフレーム長、 $\psi$  は窓関数、 $\eta$  はフレームシフト量である。

(注2) : インパルス応答  $h_{m,n}[\tau]$  の離散時間フーリエ変換は

$$\hat{h}_{m,n}(\omega) := \sum_{\tau=0}^{T-1} h_{m,n}[\tau] \exp(-i\omega\tau)$$

と定義される。

(注3) :  $i = 2, 3, \dots, \lfloor \frac{L+1}{2} \rfloor$  に対して、 $\hat{s}_{i,j,n}$  と  $\hat{s}_{L-i+2,j,n}$  は複素共役な関係にあるので、 $i = 1, 2, \dots, \lfloor \frac{L+2}{2} \rfloor$  のときのみを考えれば十分である。

のように定義すれば、式 (2) から

$$\mathbf{x}_{i,j} = A_i \mathbf{s}_{i,j} + \boldsymbol{\epsilon}_{i,j} \quad (i = 1, 2, \dots, I; j = 1, 2, \dots, J) \quad (3)$$

が成立する。ブラインド音源分離とは、式 (3) における混合行列  $A_i$  が未知の状態、 $\mathbf{x}_{i,j}$  から  $\mathbf{s}_{i,j}$  を推定し、最終的に各音源信号  $s_n[t]$  ( $n = 1, 2, \dots, N$ ) を推定する問題<sup>(注4)</sup>である [1]。

過決定条件 ( $M \geq N$ )<sup>(注5)</sup>下においては、 $A_i$  の一般逆行列

$$W_i := (\mathbf{w}_{i,1}, \mathbf{w}_{i,2}, \dots, \mathbf{w}_{i,N})^H := (A_i^H A_i)^{-1} A_i^H \in \mathbb{C}^{N \times M}$$

を求めることで、 $\mathbf{s}_{i,j}$  を

$$\begin{aligned} \mathbf{y}_{i,j} &:= (\hat{y}_{i,j,1}, \hat{y}_{i,j,2}, \dots, \hat{y}_{i,j,N})^T \in \mathbb{C}^N \\ &:= (\mathbf{w}_{i,1}^H \mathbf{x}_{i,j}, \mathbf{w}_{i,2}^H \mathbf{x}_{i,j}, \dots, \mathbf{w}_{i,N}^H \mathbf{x}_{i,j})^T = W_i \mathbf{x}_{i,j} \end{aligned}$$

のように推定する手法が主流である [1]–[5], [8]–[15]。ここで、 $W_i$  は分離行列と呼ばれ、特に  $\boldsymbol{\epsilon}_{i,j} = \mathbf{0}$  の場合には、真の分離行列が求めれば  $\mathbf{y}_{i,j} = \mathbf{s}_{i,j}$  となる。また、 $\boldsymbol{\epsilon}_{i,j} \neq \mathbf{0}$  の場合でも、二乗誤差  $\|\mathbf{x}_{i,j} - A_i \mathbf{y}_{i,j}\|_2^2$  を最小にする  $\mathbf{y}_{i,j}$  が得られる。

### 2.2 音源信号の独立性と優ガウス性を利用した分離手法

音源信号  $s_n[t]$  の複素スペクトログラム  $\bar{S}_n$  は

$$\begin{cases} \bar{s}_{j,n} := (\hat{s}_{1,j,n}, \hat{s}_{2,j,n}, \dots, \hat{s}_{I,j,n})^T \in \mathbb{C}^I \\ \bar{S}_n := (\bar{s}_{1,n}, \bar{s}_{2,n}, \dots, \bar{s}_{J,n}) \in \mathbb{C}^{I \times J} \end{cases}$$

のように定義される。代表的な従来法である、独立成分分析 (ICA) [2]、独立ベクトル分析 (IVA) [5]、独立低ランク行列分析 (ILRMA) [8] はいずれも、「各複素スペクトログラム  $\bar{S}_n$  が優ガウスのな分布<sup>(注6)</sup>から独立に生成される」と仮定して分離行列  $W_i$  を求める手法である。以下では、式 (3) において  $\boldsymbol{\epsilon}_{i,j} = \mathbf{0}$  が成立しているものとして議論を進める。

各成分の値が  $\hat{s}_{i,j,n}$ ,  $\hat{x}_{i,j,m}$  であるテンソルを  $\boldsymbol{S} \in \mathbb{C}^{I \times J \times N}$ ,  $\boldsymbol{X} \in \mathbb{C}^{I \times J \times M}$  とし、それぞれの確率密度関数を  $p_s, p_x$  とする。 $\boldsymbol{\epsilon}_{i,j} = \mathbf{0}$  のとき、 $A_i$  ( $i = 1, 2, \dots, I$ ) は  $\boldsymbol{S}$  から  $\boldsymbol{X}$  への 1 対 1 の線形写像となり、 $p_x$  は  $p_s$  に比例する。更に  $\bar{S}_n$  の独立性から、

$$p_x(\boldsymbol{X}) = \frac{p_s(\boldsymbol{S})}{\prod_{i=1}^I (\det(A_i^H A_i))^J} = \prod_{i=1}^I (\det(W_i W_i^H))^J \prod_{n=1}^N p_n(\bar{S}_n)$$

が成立する。ここで、 $p_n$  は  $\bar{S}_n$  の確率密度関数である。 $W_i$  を真の分離行列とすると、 $\hat{s}_{i,j,n} = \hat{y}_{i,j,n} = \mathbf{w}_{i,n}^H \mathbf{x}_{i,j}$  であるため、 $p_x$  は全ての分離行列の組からなるテンソル  $\boldsymbol{W} \in \mathbb{C}^{N \times M \times I}$  をパラメータとする関数  $p_x(\boldsymbol{X} | \boldsymbol{W})$  になる。したがって、最尤推定の考えに基づき、尤度  $p_x(\boldsymbol{X} | \boldsymbol{W})$  の最大化問題、すなわち負対数尤度  $-\log(p_x(\boldsymbol{X} | \boldsymbol{W}))$  の最小化問題

$$\underset{\boldsymbol{W}}{\text{minimize}} \quad - \sum_{n=1}^N \log(p_n(\bar{S}_n)) - J \sum_{i=1}^I \log(\det(W_i W_i^H)) \quad (4)$$

を解くことで分離行列  $W_i$  を決定することができる。

(注4) : 実際には、 $(h_{m,n} * s_n)[t]$  ( $m = 1, 2, \dots, M; n = 1, 2, \dots, N$ ) を推定する場合はほとんどである。

(注5) : より厳密には、 $\text{rank}(A_i) = N$  ( $\forall i = 1, 2, \dots, I$ ) の場合であり、更に  $A_i$  の条件数は大きくないことが望まれる。

(注6) :  $X$  を実数値確率変数とする。 $X$  の平均を  $\mu := E[X]$ 、分散を  $\sigma^2 := E[(X - \mu)^2]$  とすると、尖度は  $v := E[(X - \mu)^4] / \sigma^4 - 3$  と定義される。 $X$  は  $v > 0$  のとき優ガウスのであるといい、 $v < 0$  のとき劣ガウスのであるという。

$p_n$  には優ガウスの確率密度関数、つまり、ガウス関数に比べてピークが鋭く、かつ大きな裾を持つ確率密度関数を仮定する。例えば、ICA では複素ラプラス分布<sup>(注7)</sup>、IVA では球対称複素ラプラス分布<sup>(注8)</sup>、ILRMA では時変複素ガウス分布<sup>(注9)</sup>を仮定した上で、式(4)の最適化問題を解いている。いずれの場合においても、必要であれば補助関数法を用いることで、分離行列  $W_i$  ( $i = 1, 2, \dots, I$ ) ごとの最適化問題

$$\underset{W_i}{\text{minimize}} \sum_{n=1}^N \mathbf{w}_{i,n}^H U_{i,n}^{(l+1)} \mathbf{w}_{i,n} - \log(\det(W_i W_i^H)) \quad (5)$$

に帰着させて  $\mathcal{W}$  の推定値を更新する。式(5)は  $(l+1)$  回目の更新に用いられる最適化問題であり、行列  $U_{i,n}^{(l+1)} \in \mathbb{C}^{M \times M}$  は

$$U_{i,n}^{(l+1)} := \frac{1}{J} \sum_{j=1}^J \frac{X_{i,j}}{r_{i,j,n}^{(l+1)}} := \frac{1}{J} \sum_{j=1}^J \frac{\mathbf{x}_{i,j} \mathbf{x}_{i,j}^H}{r_{i,j,n}^{(l+1)}}$$

のように定義される。ここで、 $r_{i,j,n}^{(l+1)}$  の値はアルゴリズムごとに

$$r_{i,j,n}^{(l+1)} := \begin{cases} 2\sigma_n |\hat{y}_{i,j,n}^{(l)}| & (\text{ICA のとき}) \\ 2\sigma_n \|\hat{\mathbf{y}}_{j,n}^{(l)}\|_2 & (\text{IVA のとき}) \\ \sum_{k=1}^{K_n} b_{i,k,n}^{(l+1)} c_{k,j,n}^{(l+1)} & (\text{ILRMA のとき}) \end{cases}$$

と与えられる<sup>(注10)</sup>。分離行列  $W_i$  の推定値は、式(5)の最適化問題の近似解として、

$$\begin{cases} \mathbf{v}_{i,n}^{(l+1)} = (U_{i,n}^{(l+1)})^{-1} \mathbf{a}_{i,n}^{(l)} \\ \mathbf{w}_{i,n}^{(l+1)} = \frac{\mathbf{v}_{i,n}^{(l+1)}}{\sqrt{\mathbf{v}_{i,n}^{(l+1)H} U_{i,n}^{(l+1)} \mathbf{v}_{i,n}^{(l+1)}}} \end{cases} \quad (6)$$

のように更新される。ここで、 $\mathbf{a}_{i,n}^{(l)}$  は  $W_i^{(l)}$  の一般逆行列  $A_i^{(l)} = W_i^{(l)H} (W_i^{(l)} W_i^{(l)H})^{-1}$  の第  $n$  列ベクトルである。

(注7)：複素ラプラス分布

$$p_n(\bar{S}_n) := \prod_{i=1}^I \prod_{j=1}^J p_{i,j,n}(\hat{s}_{i,j,n}) := \prod_{i=1}^I \prod_{j=1}^J \frac{1}{2\pi\sigma_n^2} \exp\left(-\frac{|\hat{s}_{i,j,n}|}{\sigma_n}\right).$$

(注8)：球対称複素ラプラス分布

$$p_n(\bar{S}_n) := \prod_{j=1}^J p_{j,n}(\bar{\mathbf{s}}_{j,n}) := \prod_{j=1}^J \frac{2^{I-1} (I-1)!}{(2\pi\sigma_n^2)^J (2I-1)!} \exp\left(-\frac{\|\bar{\mathbf{s}}_{j,n}\|_2}{\sigma_n}\right).$$

(注9)：時変複素ガウス分布

$$p_n(\bar{S}_n) := \prod_{i=1}^I \prod_{j=1}^J p_{i,j,n}(\hat{s}_{i,j,n}) := \prod_{i=1}^I \prod_{j=1}^J \frac{1}{\pi r_{i,j,n}} \exp\left(-\frac{|\hat{s}_{i,j,n}|^2}{r_{i,j,n}}\right).$$

(注10)：ILRMA では、各音源信号  $s_n[t]$  のパワースペクトログラム  $|\bar{S}_n|^2$  が2つの非負値行列  $B_n \in \mathbb{R}_+^{I \times K_n}$  (基底行列) と  $C_n \in \mathbb{R}_+^{K_n \times J}$  (係数行列) の積で低ランク近似できると仮定する。ここで、 $K_n$  は基底数を表し、低ランク近似のため  $K_n \ll \min(I, J)$  を満たす値に設定される。 $b_{i,k,n}$  と  $c_{k,j,n}$  は  $B_n$  と  $C_n$  の各成分であり、時変複素ガウス分布の分散は  $r_{i,j,n} := \sum_{k=1}^{K_n} b_{i,k,n} c_{k,j,n}$  と定義される。 $b_{i,k,n}$  と  $c_{k,j,n}$  は、補助関数法を用いて

$$\begin{cases} b_{i,k,n}^{(l+1)} = b_{i,k,n}^{(l)} \sqrt{\frac{\sum_{j=1}^J |\hat{y}_{i,j,n}^{(l)}|^2 c_{k,j,n}^{(l)} \left(\sum_{k'=1}^{K_n} b_{i,k',n}^{(l)} c_{k',j,n}^{(l)}\right)^{-2}}{\sum_{j=1}^J c_{k,j,n}^{(l)} \left(\sum_{k'=1}^{K_n} b_{i,k',n}^{(l)} c_{k',j,n}^{(l)}\right)^{-1}}} \\ c_{k,j,n}^{(l+1)} = c_{k,j,n}^{(l)} \sqrt{\frac{\sum_{i=1}^I |\hat{y}_{i,j,n}^{(l)}|^2 b_{i,k,n}^{(l+1)} \left(\sum_{k'=1}^{K_n} b_{i,k',n}^{(l+1)} c_{k',j,n}^{(l)}\right)^{-2}}{\sum_{i=1}^I b_{i,k,n}^{(l+1)} \left(\sum_{k'=1}^{K_n} b_{i,k',n}^{(l+1)} c_{k',j,n}^{(l)}\right)^{-1}}} \end{cases}$$

のように更新される。

### 3. 混合系の情報を考慮したブラインド音源分離

#### 3.1 混合過程に関する先験情報の導入

第2節で紹介した従来法では、「音源信号の振幅値  $|s_n[e]|$  は高い頻度で0になる」という先験情報に基づき、複素スペクトログラム  $S_n$  の確率密度関数  $p_n$  に優ガウスの分布を仮定した。結果として、これらの手法は、音源信号の先験情報に基づいて設計された最適化問題を解くことにより、分離行列  $W_i$  を決定していることになる。一方で、マイクロホンアレイを用いて音の収録を行う場合には、音源信号の到来方向が既知であれば、混合行列  $A_i$  を推定できる。この推定値を利用して、より分離性能の高い分離行列を求める手法も提案されている[12]–[15]。しかしながら、ブラインド音源分離では、通常、音源信号の到来方向は未知である。また、到来方向が既知であったとしても、マイク配置が適切でない場合には、上記の手法は適用できない。

混合過程において一般的に成立する性質として、インパルス応答のスパース性が挙げられる。実際に、室内のインパルス応答推定に対して、スパース性やブロックスパース性を考慮することで、推定精度が向上することが報告されている[16]–[18]。本論文では、音源信号の先験情報のみならず、「インパルス応答  $h_{m,n}[\tau]$  はスパースである」という混合過程の先験情報も考慮した最適化問題を解くブラインド音源分離手法を提案する。

#### 3.2 インパルス応答のスパース性を利用した分離手法

インパルス応答  $h_{m,n}[\tau]$  ( $\tau = 0, 1, \dots, T-1$ ) をベクトル

$$\mathbf{h}_{m,n} := (h_{m,n}[0], h_{m,n}[1], \dots, h_{m,n}[T-1])^T \in \mathbb{R}^T$$

で表現し、全てのインパルス応答  $\mathbf{h}_{m,n}$  の組からなるテンソルを  $\mathcal{H} \in \mathbb{R}^{T \times M \times N}$  と定義する。 $i = 2, 3, \dots, \lfloor \frac{L+1}{2} \rfloor$  に対して、 $A_i = W_i^H (W_i W_i^H)^{-1}$  の各成分  $\hat{a}_{i,m,n}$  の複素共役をとることで  $\hat{a}_{L-i+2,m,n}$  を定義し、更に  $\hat{a}_{i,m,n}$  ( $i = 1, 2, \dots, L$ ) を並べて

$$\bar{\mathbf{a}}_{m,n} := (\hat{a}_{1,m,n}, \hat{a}_{2,m,n}, \dots, \hat{a}_{L,m,n})^T \in \mathbb{C}^L$$

を定義する。本論文では、負対数尤度  $-\log(p_x(\mathcal{X} | \mathcal{W}))$  に加えインパルス応答のスパース性も考慮した、制約付き最適化問題

$$\underset{\mathcal{W}, \mathcal{H}, (B), (C)}{\text{minimize}} -\log(p_x(\mathcal{X} | \mathcal{W})) + \lambda J [\mathcal{F}(\mathcal{W}, \mathcal{H}) + \nu \mathcal{G}(\mathcal{H})]$$

$$\text{subject to } \forall n \sum_{m=1}^M \|\bar{\mathbf{a}}_{m,n}\|_2^2 = L \quad \text{and} \quad \forall n \sum_{m=1}^M \|\mathbf{h}_{m,n}\|_2^2 = 1 \quad (7)$$

を解く手法を提案する。ここで、 $\mathcal{F} : \mathbb{C}^{N \times M \times I} \times \mathbb{R}^{T \times M \times N} \rightarrow \mathbb{R}_+$  は  $\mathcal{W}$  と  $\mathcal{H}$  の間の整合性を評価する関数、 $\mathcal{G} : \mathbb{R}^{T \times M \times N} \rightarrow \mathbb{R}_+$  はインパルス応答  $\mathbf{h}_{m,n}$  のスパース性を評価する関数である。制約条件  $\sum_{m=1}^M \|\bar{\mathbf{a}}_{m,n}\|_2^2 = L$  と  $\sum_{m=1}^M \|\mathbf{h}_{m,n}\|_2^2 = 1$  は音源信号と混合系の間に存在するスケールの任意性を取り除くために追加されている。非負値行列  $B := (B_1, B_2, \dots, B_N) \in \mathbb{R}_+^{I \times K}$  と  $C := (C_1^T, C_2^T, \dots, C_N^T)^T \in \mathbb{R}_+^{K \times J}$  ( $K := \sum_{n=1}^N K_n$ ) は、 $\bar{S}_n$  の確率密度関数  $p_n$  に IRLMA と同様の時変複素ガウス分布を仮定した際に必要となる変数である。また、 $\lambda > 0$  と  $\nu > 0$  は各項の重みを調整するパラメータである。提案手法では、 $\mathcal{W}$  と  $\mathcal{H}$  の更新を交互に行うことで、式(7)の最適化問題を解く。

### 3.3 分離行列の更新

インパルス応答の推定値  $\mathbf{h}_{m,n}$  を離散時間フーリエ変換して得られる混合行列を  $\tilde{A}_i$  ( $i = 1, 2, \dots, L$ )<sup>(注11)</sup> と定義し、対応する分離行列を  $\tilde{W}_i := (\tilde{A}_i^H \tilde{A}_i)^{-1} \tilde{A}_i^H$  ( $i = 1, 2, \dots, L$ ) とする。まず、 $\mathcal{H} = \mathcal{H}^{(l)}$  と固定し  $\mathcal{W}$  を更新する。 $\mathcal{W}$  の更新時には、 $\mathcal{W}$  と  $\mathcal{H}$  の間の整合性を評価する関数  $\mathcal{F}$  を

$$\mathcal{F}(\mathcal{W}, \mathcal{H}) := \sum_{i=1}^L \|\mathbf{W}_i - \tilde{W}_i\|_F^2 = \sum_{i=1}^L \sum_{n=1}^N \|\mathbf{w}_{i,n} - \tilde{\mathbf{w}}_{i,n}\|_2^2 \quad (8)$$

と定義する。このとき、制約条件  $\sum_{m=1}^M \|\bar{\mathbf{a}}_{m,n}\|_2^2 = L$  を無視すれば、式 (5) から、 $\tilde{S}_n$  の確率密度関数に複素ラプラス分布、球対称複素ラプラス分布、時変複素ガウス分布のいずれを採用した場合においても、分離行列  $W_i$  ごとに最適化問題

$$\underset{W_i}{\text{minimize}} \sum_{n=1}^N \left( \mathbf{w}_{i,n}^H U_{i,n}^{(l+1)} \mathbf{w}_{i,n} + \lambda \|\mathbf{w}_{i,n} - \tilde{\mathbf{w}}_{i,n}^{(l)}\|_2^2 \right) - \log(\det(W_i W_i^H)) \quad (9)$$

を解けばよいことが分かる。式 (9) のコスト関数を  $Q_i^{(l+1)}$  とすると、 $Q_i^{(l+1)}$  の値を局所的に最小にする  $W_i$  において、勾配  $\nabla Q_i^{(l+1)} := \frac{\partial Q_i^{(l+1)}}{\partial \Re[\mathbf{w}_i]} + j \frac{\partial Q_i^{(l+1)}}{\partial \Im[\mathbf{w}_i]}$  は零行列となるため、

$$\nabla Q_i^{(l+1)}(W_i) = 2 \begin{pmatrix} \mathbf{w}_{i,1}^H U_{i,1}^{(l+1)} \\ \mathbf{w}_{i,2}^H U_{i,2}^{(l+1)} \\ \vdots \\ \mathbf{w}_{i,N}^H U_{i,N}^{(l+1)} \end{pmatrix} + 2\lambda \begin{pmatrix} \mathbf{w}_{i,1}^H - \tilde{\mathbf{w}}_{i,1}^{(l)H} \\ \mathbf{w}_{i,2}^H - \tilde{\mathbf{w}}_{i,2}^{(l)H} \\ \vdots \\ \mathbf{w}_{i,N}^H - \tilde{\mathbf{w}}_{i,N}^{(l)H} \end{pmatrix} - 2(W_i W_i^H)^{-1} W_i = O_{N \times M}$$

が成立する。両辺のエルミート転置をとり、更に 2 で割れば、

$$\begin{pmatrix} \tilde{U}_{i,2}^{(l+1)} \mathbf{w}_{i,1} & \tilde{U}_{i,2}^{(l+1)} \mathbf{w}_{i,2} & \cdots & \tilde{U}_{i,N}^{(l+1)} \mathbf{w}_{i,N} \end{pmatrix} = W_i^H (W_i W_i^H)^{-1} + \lambda \tilde{W}_i^{(l)H}$$

が成立する。ここで、行列  $\tilde{U}_{i,n}^{(l+1)} \in \mathbb{C}^{M \times M}$  は  $M$  次の単位行列  $E_M$  を用いて、 $\tilde{U}_{i,n}^{(l+1)} := U_{i,n}^{(l+1)} + \lambda E_M$  のように定義される。左から  $W_i$  を掛ければ、

$$W_i \begin{pmatrix} \tilde{U}_{i,1}^{(l+1)} \mathbf{w}_{i,1} & \tilde{U}_{i,2}^{(l+1)} \mathbf{w}_{i,2} & \cdots & \tilde{U}_{i,N}^{(l+1)} \mathbf{w}_{i,N} \end{pmatrix} = E_N + \lambda W_i \tilde{W}_i^{(l)H}$$

が成立するため、結果として、 $p, q = 1, 2, \dots, N$  に対して

$$\mathbf{w}_{i,p}^H \tilde{U}_{i,q}^{(l+1)} \mathbf{w}_{i,q} = \delta_{p,q} + \lambda \mathbf{w}_{i,p}^H \tilde{\mathbf{w}}_{i,q}^{(l)} \quad (10)$$

を満たす  $W_i$  を求めればよい。ここで、 $\delta_{p,q}$  はクロネッカーのデルタである。式 (10) を近似し、 $n = 1, 2, \dots, N$  に対して

$$\begin{cases} \mathbf{w}_{i,p}^{(l)H} \tilde{U}_{i,n}^{(l+1)} \mathbf{w}_{i,n}^{(l+1)} = \lambda \mathbf{w}_{i,p}^{(l)H} \tilde{\mathbf{w}}_{i,n}^{(l)} & (\forall p \neq n) \\ \mathbf{w}_{i,n}^{(l+1)H} \tilde{U}_{i,n}^{(l+1)} \mathbf{w}_{i,n}^{(l+1)} = 1 + \lambda \mathbf{w}_{i,n}^{(l+1)H} \tilde{\mathbf{w}}_{i,n}^{(l)} \end{cases} \quad (11)$$

を満たすような  $W_i^{(l+1)}$  を求める。式 (11) を満たす  $W_i^{(l+1)}$  は、特に  $M > N$  のとき無数に存在するが、例えば、

(注11) :  $i = 1, 2, \dots, L$  ではなく、 $i = 1, 2, \dots, L$  である点に注意されたい。

$$\begin{cases} \mathbf{v}_{i,n}^{(l+1)} = (\tilde{U}_{i,n}^{(l+1)})^{-1} \mathbf{a}_{i,n}^{(l)} \\ \tilde{\mathbf{v}}_{i,n}^{(l+1)} = \lambda (\tilde{U}_{i,n}^{(l+1)})^{-1} \tilde{\mathbf{w}}_{i,n}^{(l)} \\ d_{i,n}^{(l+1)} = \mathbf{v}_{i,n}^{(l+1)H} \tilde{U}_{i,n}^{(l+1)} \mathbf{v}_{i,n}^{(l+1)} \\ \tilde{d}_{i,n}^{(l+1)} = \mathbf{v}_{i,n}^{(l+1)H} \tilde{U}_{i,n}^{(l+1)} \tilde{\mathbf{v}}_{i,n}^{(l+1)} \\ \mathbf{w}_{i,n}^{(l+1)} = \begin{cases} \frac{\mathbf{v}_{i,n}^{(l+1)}}{\sqrt{d_{i,n}^{(l+1)}}} + \tilde{\mathbf{v}}_{i,n}^{(l+1)} & \text{if } \tilde{d}_{i,n}^{(l+1)} = 0, \\ \frac{\tilde{d}_{i,n}^{(l+1)}}{2d_{i,n}^{(l+1)}} \left[ -1 + \sqrt{1 + \frac{4d_{i,n}^{(l+1)}}{|\tilde{d}_{i,n}^{(l+1)}|^2}} \right] \mathbf{v}_{i,n}^{(l+1)} + \tilde{\mathbf{v}}_{i,n}^{(l+1)} & \text{otherwise,} \end{cases} \end{cases} \quad (12)$$

のように与えられ、式 (12) は式 (6) の一般化となっている [11]。しかし、式 (9) の問題を解く際に条件  $\sum_{m=1}^M \|\bar{\mathbf{a}}_{m,n}\|_2^2 = L$  を無視していたため、これを満たすように  $\mathcal{W}^{(l+1)}$  の修正を行う。

まず、混合行列  $A_i^{(l+1)} = W_i^{(l+1)H} (W_i^{(l+1)} W_i^{(l+1)H})^{-1}$  ( $i = 1, 2, \dots, L$ ) を計算する。次に、 $i = 2, 3, \dots, \lfloor \frac{L+1}{2} \rfloor$  に対して、 $A_i^{(l+1)}$  の各成分の複素共役を取ることで  $A_{L-i+2}^{(l+1)}$  を計算する。 $\gamma_n^{(l+1)} := \sqrt{\sum_{m=1}^M \|\bar{\mathbf{a}}_{m,n}^{(l+1)}\|_2^2 / L}$  ( $n = 1, 2, \dots, N$ ) とし、 $\mathbf{a}_{i,n}^{(l+1)}$  ( $i = 1, 2, \dots, L$ ) を  $\gamma_n^{(l+1)}$  で割る、すなわち  $\mathbf{w}_{i,n}^{(l+1)}$  を  $\gamma_n^{(l+1)}$  倍することで、 $\sum_{m=1}^M \|\bar{\mathbf{a}}_{m,n}\|_2^2 = L$  を満たす  $\mathcal{W}^{(l+1)}$  が求まる。

### 3.4 インパルス応答の更新

$\mathcal{W}$  の更新後は、 $\mathcal{W} = \mathcal{W}^{(l+1)}$  と固定し  $\mathcal{H}$  を更新する。 $\mathcal{H}$  の更新時には、関数  $\mathcal{F}$  を

$$\mathcal{F}(\mathcal{W}, \mathcal{H}) := \sum_{i=1}^L \|\mathbf{A}_i - \tilde{A}_i\|_F^2 = \sum_{m=1}^M \sum_{n=1}^N \|\bar{\mathbf{a}}_{m,n} - \Phi \mathbf{h}_{m,n}\|_2^2 \quad (13)$$

と定義<sup>(注12)</sup>する。ここで、行列  $\Phi$  は  $\phi_{i,\tau} := \exp(j \frac{2\pi(i-1)\tau}{L})$  ( $i = 1, 2, \dots, L; \tau = 0, 1, \dots, T-1$ ) を用いて、

$$\begin{cases} \phi_\tau := (\phi_{1,\tau}, \phi_{2,\tau}, \dots, \phi_{L,\tau})^T \in \mathbb{C}^L \\ \Phi := (\phi_0, \phi_1, \dots, \phi_{T-1}) \in \mathbb{C}^{L \times T} \end{cases}$$

のように定義される。インパルス応答のスパース性を評価する関数  $\mathcal{G}$  には、重み付き  $\ell_0$  ノルムを採用し、

$$\mathcal{G}(\mathcal{H}) := \sum_{m=1}^M \sum_{n=1}^N \|\mathbf{h}_{m,n}\|_0^\kappa := \sum_{m=1}^M \sum_{n=1}^N \sum_{\tau=0}^{T-1} \kappa[\tau] \Gamma(\mathbf{h}_{m,n}[\tau]) \quad (14)$$

と定義する。ここで、重み  $\kappa[\tau] > 0$  ( $\tau = 0, 1, \dots, T-1$ ) の値は、インパルス応答の性質を考慮し、 $\tau$  に関して単調非減少になるように設定するとよい。また 2 値関数  $\Gamma: \mathbb{R} \rightarrow \{0, 1\}$  は

$$\Gamma(h) := \begin{cases} 0 & \text{if } h = 0, \\ 1 & \text{if } h \neq 0, \end{cases}$$

と定義される。このとき、制約条件  $\sum_{m=1}^M \|\mathbf{h}_{m,n}\|_2^2 = 1$  を無視

(注12) : 同じサイズの行列  $X$  と  $Y$ 、及びそれぞれの一般逆行列  $X^\dagger$  と  $Y^\dagger$  に対して、通常  $\|X - Y\|_F^2 \neq \|X^\dagger - Y^\dagger\|_F^2$  であるため、式 (8) と式 (13) で、実際には別々の関数を定義していることになる。しかしながら、 $X = Y$  のときには  $\|X - Y\|_F^2 = \|X^\dagger - Y^\dagger\|_F^2 = 0$  が成立することから、本論文では式 (8) と式 (13) を  $\mathcal{W}$  と  $\mathcal{H}$  の間の整合性を評価する関数として用いる。

すれば、インパルス応答  $\mathbf{h}_{m,n}$  ごとに最適化問題

$$\underset{\mathbf{h}_{m,n}}{\text{minimize}} \|\bar{\mathbf{a}}_{m,n}^{(l+1)} - \Phi \mathbf{h}_{m,n}\|_2^2 + \nu \|\mathbf{h}_{m,n}\|_0^\kappa \quad (15)$$

を解けばよいことが分かる．ここで、 $\phi_p^H \phi_q = L\delta_{p,q}$  ( $p, q = 0, 1, \dots, T-1$ ) を用いれば、式 (15) の問題の解  $\mathbf{h}_{m,n}^{(l+1)}$  に関して

$$\begin{aligned} \mathbf{h}_{m,n}^{(l+1)} &= \underset{\mathbf{h}_{m,n}}{\text{argmin}} \|\bar{\mathbf{a}}_{m,n}^{(l+1)} - \Phi \mathbf{h}_{m,n}\|_2^2 + \nu \|\mathbf{h}_{m,n}\|_0^\kappa \\ &= \underset{\mathbf{h}_{m,n}}{\text{argmin}} \left\| \frac{\Phi^H \bar{\mathbf{a}}_{m,n}^{(l+1)}}{\sqrt{L}} - \sqrt{L} \mathbf{h}_{m,n} \right\|_2^2 + \nu \|\mathbf{h}_{m,n}\|_0^\kappa \\ &= \underset{\mathbf{h}_{m,n}}{\text{argmin}} L \left\| \frac{\Phi^H \bar{\mathbf{a}}_{m,n}^{(l+1)}}{L} - \mathbf{h}_{m,n} \right\|_2^2 + \nu \|\mathbf{h}_{m,n}\|_0^\kappa \\ &= \underset{\mathbf{h}_{m,n}}{\text{argmin}} \left\| \frac{\Phi^H \bar{\mathbf{a}}_{m,n}^{(l+1)}}{L} - \mathbf{h}_{m,n} \right\|_2^2 + \frac{\nu}{L} \|\mathbf{h}_{m,n}\|_0^\kappa \end{aligned}$$

が成立する．結果として、 $\tilde{\mathbf{h}}_{m,n}^{(l+1)} := \Phi^H \bar{\mathbf{a}}_{m,n}^{(l+1)} / L$  の第  $\tau$  成分を  $\tilde{h}_{m,n}^{(l+1)}[\tau - 1]$  とおけば、インパルス応答の推定値  $\mathbf{h}_{m,n}^{(l+1)}$  は

$$h_{m,n}^{(l+1)}[\tau] = \begin{cases} \tilde{h}_{m,n}^{(l+1)}[\tau] & \text{if } |\tilde{h}_{m,n}^{(l+1)}[\tau]| \geq \sqrt{\frac{\nu\kappa[\tau]}{L}}, \\ 0 & \text{if } |\tilde{h}_{m,n}^{(l+1)}[\tau]| < \sqrt{\frac{\nu\kappa[\tau]}{L}}, \end{cases} \quad (16)$$

のように与えられる．

しかし、式 (16) で得られる  $\mathcal{H}^{(l+1)}$  は条件  $\sum_{m=1}^M \|\mathbf{h}_{m,n}\|_2^2 = 1$  を満たさないため、これを満たすように  $\mathcal{H}^{(l+1)}$  の修正を行う． $\xi_n^{(l+1)} := \sqrt{\sum_{m=1}^M \|\mathbf{h}_{m,n}^{(l+1)}\|_2^2}$  ( $n = 1, 2, \dots, N$ ) として、 $\mathbf{h}_{m,n}^{(l+1)}$  ( $m = 1, 2, \dots, M$ ) を  $\xi_n^{(l+1)}$  で割ることで、 $\sum_{m=1}^M \|\mathbf{h}_{m,n}\|_2^2 = 1$  を満たす  $\mathcal{H}^{(l+1)}$  が求まる．

$\mathcal{W}$  と  $\mathcal{H}$  の更新を既定の回数繰り返すことで、最終的な分離行列  $W_i$  ( $i = 1, 2, \dots, I$ ) の推定値を決定する．離散時間信号  $(h_{m,n} * s_n)[t]$  ( $m = 1, 2, \dots, M; n = 1, 2, \dots, N$ ) の推定値は、

$$\hat{\mathbf{y}} := (\hat{y}_{i,j,1,n}, \hat{y}_{i,j,2,n}, \dots, \hat{y}_{i,j,M,n})^T := \hat{y}_{i,j,n} \mathbf{a}_{i,n} \in \mathbb{C}^M$$

を計算し、 $\hat{y}_{i,j,m,n}$  ( $i = 1, 2, \dots, I; j = 1, 2, \dots, J$ ) を逆短時間フーリエ変換 [19] することにより得られる．

## 4. 数値実験

### 4.1 実験概要

提案法の有効性を示すために、音源数  $N = 2$ 、マイク数  $M = 2$  の場合でシミュレーションを行った．サンプリング周波数 16,000 [Hz] の音源信号として、無償配布されている音楽信号 (ギター [22] とピアノ [23]) を用いた実験と、音源分離コンペティションで利用した音声信号 (二人の男性の声) [22] を用いた実験を行った．図 1 のような音源とマイクの位置関係を想定し、Aachen Impulse Response Database [20] 内にあるインパルス応答を参考に、 $\mathbf{h}_{1,1} = \mathbf{h}_{2,2}$  を図 2 のように、 $\mathbf{h}_{1,2}$  を図 3 のように、 $\mathbf{h}_{2,1}$  を図 4 のように定義した．短時間フーリエ変換を行う際は、フレーム長を  $L = 8192$ 、フレームシフト量を  $\eta = 2048$ 、窓関数をハミング窓  $\psi[\tau] := 0.54 - 0.46 \cos(2\pi\tau/L)$  とした．

比較手法を ILRMA [7] とし、提案法における  $\bar{S}_n$  の確率密度関数  $p_n$  にも ILRMA と同様の時変ガウス分布を採用した．基底行列  $B_n$  の基底数は、音楽信号に対する実験では  $K_1 = K_2 = 30$

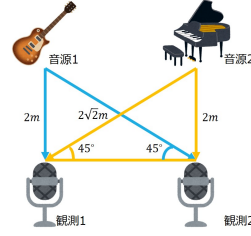


図 1 音源とマイクの位置関係

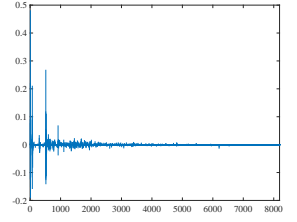


図 2  $\mathbf{h}_{1,1} = \mathbf{h}_{2,2}$

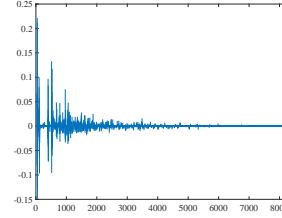


図 3  $\mathbf{h}_{1,2}$

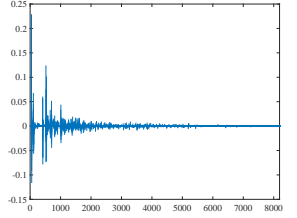


図 4  $\mathbf{h}_{2,1}$

とし、音声信号に対する実験では  $K_1 = K_2 = 5$  とした．提案法のパラメータは、 $T = 4096$ 、 $\lambda = 0.075$ 、 $\nu = 8192$  とし、式 (14) において重みは、 $\kappa[\tau] := -\log_{10}(1 - \exp(-432/(\tau + 1)))$  とした．提案法と ILRMA で共通の初期値から 100 回更新を行い、分離行列の初期値は  $W_i^{(0)} = E_2$  ( $i = 1, 2, \dots, I$ ) とし、基底行列と係数行列の初期値  $B_n^{(0)}$  と  $C_n^{(0)}$  の各非負成分は、 $[0, 1]$  の一様分布からランダムに生成した．音源分離の評価指標には、文献 [20] で定義されている source-to-distortion ratio (SDR), source-to-interferences ratio (SIR), sources-to-artifacts ratio (SAR) の改善量を用い、 $B_n$  と  $C_n$  の初期値を変えた 10 回の平均値で比較を行った．

### 4.2 実験結果

まず音楽信号に対する実験結果を述べる．図 5 に提案法と従来法 (ILRMA) の音源分離性能を示す．図 5 から、提案法による分離結果は、総合的な分離性能を表す SDR 値で平均約 0.64 [dB]、非目的音の除去性能を表す SIR 値で約 1.18 [dB]、音源分離時に生じた歪みの少なさを表す SAR 値で約 0.39 [dB]、従来法による結果を上回っていた．また、どちらの手法も 40 回程度の更新回数で SDR 値が収束していた．分離結果を実際に聞いてみると、従来法よりも提案法を用いた場合の方が、音源どおしの干渉度合いが少ない、クリアな音を得ることができた．

次に、音声信号に対する実験結果を述べる．図 6 に提案法と従来法の音源分離性能を示す．案法による分離結果は、SDR 値で約 0.92 [dB]、SIR 値で約 0.95 [dB]、SAR 値で約 0.84 [dB]、従来法による結果を上回っていた．また、どちらの手法も 30 回程度の更新回数で SDR 値が収束していた．これらの結果から、音声信号に対しても提案法により高品質な分離音源を得られることが確認された．

## 5. おわりに

本研究では、近年提案されたブラインド音源分離手法の独立低ランク行列分析に加えて、インパルス応答のスパース性を用いて音源分離を行う手法を提案した．従来の音源分離からイン

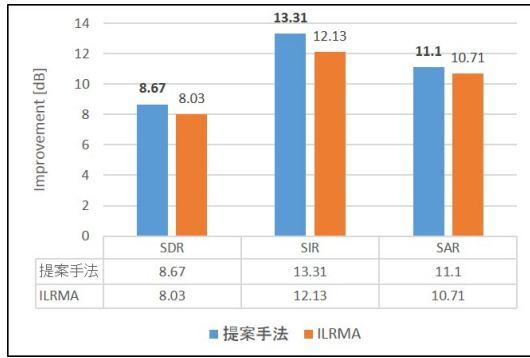


図5 音楽信号に対する提案法と ILRMA の分離結果

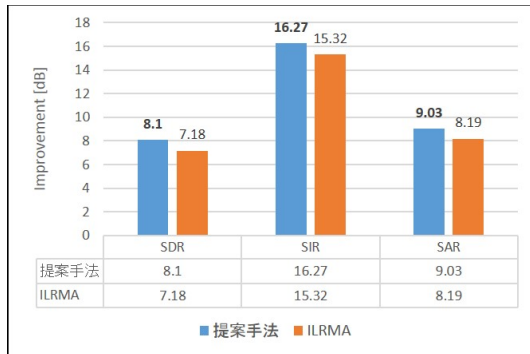


図6 音声信号に対する提案法と ILRMA の分離結果

パルス応答を推定し、インパルス応答のスパース性を加えた最適化問題を解き、分離音源を推定した。その際、この手法は分離行列を音源ごとに正規化を行うと数値計算が不安定になるため、提案手法では分離行列を推定する際、インパルス応答ごとに正規化を行った。この結果、分離性能を表す SDR 値は平均約 0.7dB、非目的音の除去性能を表す SIR 値は平均約 1.48dB、提案手法が従来法を上回った。

なお、本論文で用いた混合系インパルス応答のスパース性という考え方は、ICA や IVA にも適用できるものである。そこで今後は、これらの手法においてもインパルス応答のスパース性を利用することで、推定精度を向上できるかを確認していく必要がある。

## 文献

[1] P. Comon, "Independent component analysis, a new concept?," *Signal Process.*, vol. 36, no. 3, pp. 287–314, 1994.

[2] P. Smaragdakis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, pp. 21–34, 1998.

[3] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Trans. Audio, Speech, and Lang. Process.*, vol. 12, no. 5, pp. 530–538, 2004.

[4] T. Kim, H. T. Attias, S. Y. Lee, and T. W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 1, pp. 70–79, 2007.

[5] N. Ono and S. Miyabe, "Auxiliary-function-based independent component analysis for super-Gaussian sources," in *Proc. LVA/ICA*, vol. 14, no. 2, pp. 165–172, 2010.

[6] D. D. Lee and H. S. Seung "Algorithms for non-negative matrix factorization," in *Proc. NIPS*, 535–541, 2001.

[7] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, "Multichannel extensions of non-negative matrix factorization with complexvalued data," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 5, pp. 971–982, 2013.

[8] 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. Audio, Speech, Lang. Process.*, vol. 24, no. 9, pp. 1626–1641, 2016.

[9] D. Kitamura, "Algorithms for independent low-rank matrix analysis." [Online] <http://d-kitamura.net/pdf/misc/AlgorithmsForIndependentLowRankMatrixAnalysis.pdf>

[10] Y. Mitsui, N. Takamune, D. Kitamura, H. Saruwatari, Y. Takahashi, and K. Kondo, "Vectorwise coordinate descent algorithm for spatially regularized independent low-rank matrix analysis," in *Proc. IEEE ICASSP*, 2018 (to appear).

[11] Y. Mitsui, D. Kitamura, S. Takamichi, N. Ono, and H. Saruwatari, "Blind source separation based on independent low-rank matrix analysis with sparse regularization for time-series activity," in *Proc. IEEE ICASSP*, 2017, pp. 21–25.

[12] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, "Blind source separation based on a fast-convergence algorithm combining ICA and beamforming," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 2, pp. 666–678, 2006.

[13] L. C. Parra and C. V. Alvino, "Geometric source separation: Merging convolutive source separation with geometric beamforming," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 10, no. 6, pp. 352–362, 2002.

[14] K. Osako, Y. Mori, Y. Takahashi, H. Saruwatari, and K. Shikano, "Fast convergence blind source separation using frequency subband interpolation by null beamforming," *IEICE Trans. Fundam.*, vol. E91-A, no. 6, pp. 1357–1361, 2008.

[15] Y. Zheng, K. Reindl, and W. Kellermann, "BSS for improved interference estimation for blind speech signal extraction with two microphones," in *Proc. CAMSAP*, 2009, pp. 253–256.

[16] B. L. Sturm and G. Defrance, "Detection and estimation of arrivals in room impulse responses by greedy sparse approximation," in *Proc. EUSIPCO 2010*, pp. 1934–1938, 2010.

[17] M. Crocco and A. D. Bue, "Room impulse response estimation by iterative weighted  $L_1$ -norm," in *Proc. EUSIPCO 2015*, pp. 1895–1899, 2015.

[18] G. Vairetti, E. D. Sena, and M. Catrysse, "A scalable algorithm for physically motivated and sparse approximation of room impulse responses with orthonormal basis functions," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, pp. 1547–1561, vol. 25, no. 7, 2017.

[19] D. W. Griffin and J. S. Lim, "Signal estimation from modified short-time Fourier transform," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 32, no. 2, pp. 236–243, 1984.

[20] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1462–1469, 2006.

[21] *Aachen Impulse Response Database*, available online at: <https://www.iks.rwth-aachen.de/en/research/tools-downloads/databases/aachen-impulse-response-database/>

[22] S. Araki, A. Ozerov, V. Gowreesunker, H. Sawada, F. Theis, G. Nolte, D. Lutter, and N. Q. K. Duong, "The 2010 signal separation evaluation campaign (SiSEC2010): Audio source separation," in *Proc. LVA/ICA*, 2010, pp. 114–122.

[23] Musicpalette, BGM\_012「せつないギターアルペジオ」, available online at: <https://www.youtube.com/watch?v=GyYUKuJ3JL0&t=3s>

[24] Musicpalette, BGM\_285「ゴージャス」, available online at: <https://www.youtube.com/watch?v=FFVv3v6CH218>