

スパース時系列正則化に基づく独立低ランク行列分析を用いた ブラインド音源分離*

☆三井祥幹（東大），北村大地（総研大），高道慎之介（東大），猿渡洋（東大）

1 はじめに

ブラインド音源分離とは，複数の音源からなる混合信号を観測した際に，マイクや音源の位置等の事前情報を利用することなく，混合する前の音源信号を推定する技術である．観測チャンネル数が音源数以上となる優決定条件では，音源信号間の独立性を利用して音源分離を行う独立成分分析（independent component analysis: ICA）[1]に基づく手法が盛んに研究されている．ICAに基づく音源分離では，時間周波数領域での線形時不変な瞬時混合モデルを仮定する．これは，チャンネル間空間相関行列のランクが1になることから，ランク1空間モデル[2]と呼ばれる．このような仮定の下で，周波数ビンごとに独立なICAを適用する周波数領域ICA（frequency-domain ICA: FDICA）が提案されているが，分離後にパーミュテーション問題を解く必要がある．分離とパーミュテーション問題を同時に解く手法として，周波数ビンをまとめて1つのベクトルとして取り扱う独立ベクトル分析（independent vector analysis: IVA）[3]が提案されている．

一方で，観測信号のスペクトログラムを非負行列の積に分解する非負値行列因子分解（nonnegative matrix factorization: NMF）[4]に基づく音源分離も盛んに研究されており，多チャンネルに拡張したMNMF（multichannel NMF: MNMF）[5]や，IVAとNMFを統合したブラインド音源分離手法である独立低ランク行列分析（independent low-rank matrix analysis: ILRMA）[6, 7]などが提案されている．特にILRMAは，分離信号の時間周波数構造（NMFモデル）と線形分離行列を同時に推定する手法であり，音楽信号においてIVAやMNMFよりも頑健かつ高精度な音源分離を実現している．

本稿では，音源間の独立性のみを用いるILRMAに対して，時間周波数構造に関する事前分布を新たに導入する．特に，音声信号や音楽信号を対象とし，時間周波数領域での時系列がスパースとなる正則化を新たに付与することで，従来のILRMAよりも高精度な音源分離を目指す．スパース正則化には，ベイズ型ポストフィルタの一種であるMOSIE（MMSE estimation with Optimizable Speech model and Inhomogeneous Error criterion）[8]を利用し，その統計モデルパラメータを自己推定する枠組みも併せて提案する．

2 従来の音源分離手法

2.1 表記の定義

観測チャンネル数および音源数を，それぞれ M, N とおく．また，時間周波数領域における複素多チャンネル

音源信号，観測信号および分離信号をそれぞれ

$$\mathbf{s}_{ij} = (s_{ij,1}, s_{ij,2}, \dots, s_{ij,N})^T \quad (1)$$

$$\mathbf{x}_{ij} = (x_{ij,1}, x_{ij,2}, \dots, x_{ij,M})^T \quad (2)$$

$$\mathbf{y}_{ij} = (y_{ij,1}, y_{ij,2}, \dots, y_{ij,N})^T \quad (3)$$

とおく．ここで， $i = 1, \dots, I$ は周波数ビン， $j = 1, \dots, J$ は時間フレーム， $n = 1, \dots, N$ は音源， $m = 1, \dots, M$ は観測チャンネルのインデックスをそれぞれ表し， \top は行列またはベクトルの転置を表す．

短時間フーリエ変換の窓長が各音源から各マイクまでのインパルス応答長より十分長い場合，ランク1空間モデルが成立し，観測信号を次式で表せる．

$$\mathbf{x}_{ij} = \mathbf{A}_i \mathbf{s}_{ij} \quad (4)$$

ここで， $\mathbf{A}_i = (\mathbf{a}_{i,1}, \dots, \mathbf{a}_{i,N})$ は混合行列であり， $\mathbf{a}_{i,n}$ は各音源のステアリングベクトルである． $M = N$ となる場合は，分離行列 $\mathbf{W}_i = (\mathbf{w}_{i,1}, \dots, \mathbf{w}_{i,M})^H$ が存在し，分離信号は

$$\mathbf{y}_{ij} = \mathbf{W}_i \mathbf{x}_{ij} \quad (5)$$

と書ける．ここで H は行列またはベクトルの共役転置を表し， $\mathbf{W}_i = \mathbf{A}_i^{-1}$ の関係にある．

2.2 NMF

NMFは，非負値の要素のみで構成される行列を，非負の2つの行列の積として分解する手法であり，行列の低ランク構造を抽出できる．NMFでは，非負行列 \mathbf{D} を

$$\mathbf{D} \approx \mathbf{T}\mathbf{V} \quad (6)$$

として近似的に分解する．ここで， \mathbf{T} および \mathbf{V} は非負行列である．音響信号に対しては， \mathbf{D} をパワースペクトログラムとして分解する．このとき， \mathbf{T} の各列にはパワースペクトログラム中の頻出スペクトルパターンを表す基底ベクトルが得られ， \mathbf{V} の各行には各基底ベクトルの時間的な強度変化を表すアクティベーションが得られる．

NMFの目的関数は，観測 \mathbf{D} とモデル $\mathbf{T}\mathbf{V}$ のダイバージェンスで定義される．

$$Q_{\text{NMF}} = \mathcal{D}(\mathbf{D} | \mathbf{T}\mathbf{V}) \quad (7)$$

ここで， $\mathcal{D}(\cdot | \cdot)$ はダイバージェンスであり，行列の要素ごとの二乗ユークリッド距離，一般化カルバック・ライブラー情報量，板倉斎藤擬距離などが用いられ，その最適化手法として乗算更新型の反復アルゴリズムが提案されている[4, 9]．

2.3 ILRMA

IVAはFDICAの独立なスカラー変数をベクトル変数に拡張したモデルであるが，ILRMAはIVAのベクトル変数をさらに低ランク行列変数（NMF変数）に拡張したアルゴリズムである．また，MNMF[5]に

*Blind source separation using independent low-rank matrix analysis with sparse regularization for time-series activations, by Yoshiki Mitsui (The University of Tokyo), Daichi Kitamura (SOKENDAI), Shinnosuke Takamichi (The University of Tokyo), Hiroshi Saruwatari (The University of Tokyo).

対してランク 1 空間モデルを仮定した理論と等価なため、Rank-1 MNMF と呼ばれる [7].

ILRMA の目的関数は次式で表される [7]. ただし、 $M = N$ およびランク 1 空間モデルを仮定している.

$$Q_{\text{ILRMA}} = \sum_{i,j} \left[\sum_n \frac{|y_{ij,n}|^2}{\sum_k z_{nk} t_{ik} v_{kj}} - 2 \log |\det \mathbf{W}_i| + \sum_n \log \sum_k z_{nk} t_{ik} v_{kj} \right] \quad (8)$$

ここで、 z_{nk} は n 番目の音源信号に対する k 番目の基底の寄与を表す隠れ変数であり、 $0 \leq z_{nk} \leq 1$ かつ $\sum_n z_{nk} = 1$ が成り立つ。 t_{ik} および v_{kj} は、NMF の基底行列 \mathbf{T} およびアクティベーション行列 \mathbf{V} の非負要素である。式 (8) の第一項と第二項は IVA の目的関数に一致し、第一項と第三項は板倉斎藤擬距離に基づく NMF (Itakura-Saito NMF: ISNMF) の目的関数に一致する。文献 [7] では式 (8) の効率的な最適化手法が提案されている。

3 提案手法

3.1 動機

ILRMA は分離信号の音源モデル (時間周波数モデル) を \mathbf{T} と \mathbf{V} で捉えながら分離行列 \mathbf{W}_i を推定する。しかし、より複雑な時間周波数構造を持つ音源信号に対しては、NMF の音源モデル推定に失敗し、分離精度が上がらないという問題がある。そこで「音源信号の時間周波数構造の時系列 (スペクトログラム行列上の任意周波数ビンにおける行ベクトル) はスパースになる」という事前知識を踏まえ、ILRMA の音源モデル $\mathbf{T}\mathbf{V}$ の時系列がスパースになる正則化を付与することで、音声や音楽等の複雑な構造を持つ音源信号の分離を促進する。ここで、一般の音響信号に関しては、各周波数ビンにおける時系列のスパース度合いはその周波数位置ごとに大きく異なる。よって、アクティベーション行列 \mathbf{V} を直接スパースに誘導するのではなく、各周波数ビンのスパース度合いを MOSIE の事前分布の形状母数として推定する手法も提案する。

3.2 定式化

提案手法は、以下に示す二段階最適化として定式化できる。ここでは、正則化を表す制約式 (10) も最小化問題を含むのが特徴である。

$$\min_{\mathbf{W}_i, \mathbf{T}, \mathbf{V}, \mathbf{Z}} Q_{\text{ILRMA}} \quad (9)$$

$$\text{s.t. } \min_{\mathbf{V}} \sum_n \mathcal{D}_{\text{IS}} \left(|\tilde{\mathbf{Y}}_n|^2 \mid \mathbf{T}_n \mathbf{V}_n \right) \quad (10)$$

ここで、 $\mathcal{D}_{\text{IS}}(\cdot \mid \cdot)$ は板倉斎藤擬距離であり、行列 \mathbf{X} に対する演算 $|\mathbf{X}|^2$ は行列の要素ごとの絶対値の 2 乗を表す。また、 $\tilde{\mathbf{Y}}_n$ は、時系列のスパース性が強調された n 番目の分離信号のスペクトログラムであり、3.3.3 節で詳しく述べる。 \mathbf{T}_n および \mathbf{V}_n は、それぞれ n 番目の音源信号に対する寄与が大きい基底とアクティベーションのみで構成される行列であり、3.3.4 節で詳しく述べる。この二段階最適化を直接解くのは困難であるが、本稿では文献 [10] に倣い、目的関数の最適化更新を数回行い、制約式の最適化更新を数回行う、といった流れを繰り返すことにより、全体

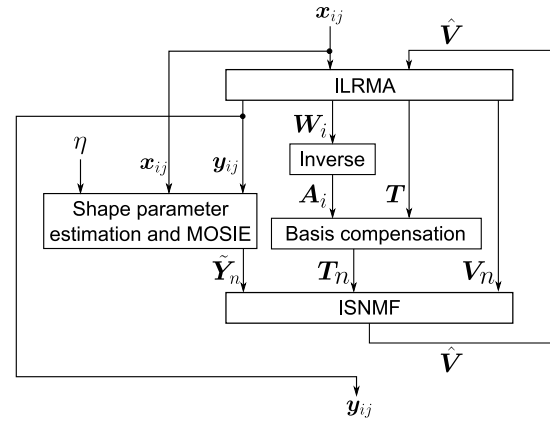


Fig. 1 Process flow of proposed method.

の最適化を図る。

3.3 処理の流れ

Figure 1 は提案手法の処理の流れを示している。まず、従来の ILRMA を数回反復更新し、分離行列 \mathbf{W}_i 、基底行列 \mathbf{T} およびアクティベーション行列 \mathbf{V} を得る。この基底行列 \mathbf{T} のスペクトル形状を補償した上で、 n 番目の音源に寄与する基底ベクトルのみで構成される部分基底行列 \mathbf{T}_n を構成する。次に、観測信号 x_{ij} と ILRMA の暫定的な分離信号 y_{ij} を用いて周波数ビンごとの適切な形状母数推定および MOSIE を適用し、スパース性を強調した n 番目の分離信号 $\tilde{\mathbf{Y}}_n$ を得る。このとき、反復初期においては、推定される形状母数値よりもさらに強い度合いで強制的にスパース性を誘導するためのハイパーパラメータ η を用いることで、音源分離を促進する。ここで得た観測信号 $\tilde{\mathbf{Y}}_n$ に対して、基底行列 \mathbf{T}_n を固定した ISNMF を適用し、スパース正則化された分離信号に対応するアクティベーション行列 $\hat{\mathbf{V}}_n$ を得る。これを ILRMA の内部モデル \mathbf{V} に戻し、再び ILRMA を数回反復更新する。この手順を繰り返すことで、ILRMA の音源モデル $\mathbf{T}\mathbf{V}$ の時系列がスパースになるように誘導し、音源分離を促進する。以降では、各ステップにおける処理の詳細を説明する。

3.3.1 ILRMA の更新

ILRMA の乗算更新式 [7] に基づき、分離行列の \mathbf{W}_i 、基底行列 \mathbf{T} およびアクティベーション行列 \mathbf{V} を数回更新する。

3.3.2 基底ベクトルの補償

ILRMA は ICA と同様に分離信号のスケールを決められないため、分離行列 \mathbf{W}_i および基底 t_{ik} は周波数ごとにスケールがばらばらになっている。そこで、観測信号の m 番目のチャンネルのスケールに補償した基底 $t_{ik}^{(m)}$ を次式のように計算する。ただし、本稿では簡略化のため、 $n = m$ の場合のみを考える。

$$t_{ik}^{(n)} = |a_{i,nn}|^2 z_{nk} t_{ik} \quad (11)$$

ここで、 $a_{i,nn}$ は ILRMA で推定した分離行列 \mathbf{W}_i の逆行列から得られる。

3.3.3 MOSIE による時系列スパース性の強調

MOSIE はベイズ型ポストフィルタの一種であり、目的音源信号および妨害音信号に対してそれぞれ事

前分布を仮定し、妨害音の分散の推定値を利用して目的音源信号の強調および妨害音の抑圧を行うフィルタである [8]。MOSIE では、各周波数ビンにおいて、目的音の振幅スペクトルの事前分布としてカイ分布を仮定する。ここで、 i 番目の周波数ビンのある時系列に対し、 $\chi \in \mathbb{R}_{\geq 0}$ を目的音の振幅スペクトルを表す確率変数とすると、カイ分布の確率密度関数は次式で与えられる。

$$p(\chi) = \frac{2}{\Gamma(\rho_i)} \left(\frac{\rho_i}{\mathbb{E}[\chi^2]} \right)^{\rho_i} \chi^{2\rho_i-1} \exp\left(-\frac{\rho_i}{\mathbb{E}[\chi^2]} \chi^2\right) \quad (12)$$

ここで、 ρ_i は i 番目の周波数ビンにおけるカイ分布の形状母数、 $\mathbb{E}[\cdot]$ は期待値を表す。

まず、式 (12) で表される時系列統計モデルに関して、そのスパース性を制御する周波数別形状母数 ρ_i の決定方法について述べる。3.3.2 節と同様に $m = n$ の場合を考えると、 n 番目のチャネルにおける観測信号 $x_{ij,n}$ のうち、ILRMA の出力から推定される妨害音成分を $q_{ij,n}$ とおく。本研究においては、 ρ_i の推定値を以下の式で算出する。

$$\begin{aligned} \rho_i = & \eta \cdot \{\mu_2^2(|x_{ij,n}|) + \mu_2^2(|q_{ij,n}|) \\ & - 2\mu_2(|x_{ij,n}|)\mu_2(|q_{ij,n}|)\} \\ & \cdot \{\mu_4(|x_{ij,n}|) - \mu_4(|q_{ij,n}|) - \mu_2^2(|x_{ij,n}|) \\ & - 2\mu_2(|x_{ij,n}|)\mu_2(|q_{ij,n}|) + 3\mu_2^2(|q_{ij,n}|)\}^{-1} \quad (13) \end{aligned}$$

ここで、 $\mu_d(\cdot)$ は時間フレーム j に関する d 次のモーメントを表し、 $\mu_d(\alpha_{ij}) = (\sum_j \alpha_{ij}^d)/J$ である。また、 η は後述するハイパーパラメータを表す。式 (13) の右辺から η を除いた因子は、現状の分離音から得られる形状母数の empirical 推定値であり、モーメントマッチングによって各周波数別に最尤推定されたものである [11]。ただし、ILRMA の反復初期においては音源分離が不十分な状態であるため、中心極限定理により、この推定値は過大推定されていることに留意する（これは、音源のスパース性を過小評価していることと等価である）。このため、ハイパーパラメータ $0 < \eta \leq 1$ を乗ずることによって、上述の empirical 推定値を補正し、スパース性を強調する。この η は、反復処理が進行するにつれて音源分離の精度が向上することを考慮し、次第に 1 へと収束する簡単な数列を適当に定めればよい。本稿ではこれを $\eta = l/L$ と与える。ここで、 l は提案手法の現在の反復回数を、 L は総反復回数を表す。

前述の ρ_i が与えられるとき、目的音複素スペクトル $\tilde{y}_{ij,n}$ の最適な推定値は以下のように書ける [8]。

$$\begin{aligned} \tilde{y}_{ij,n} = & x_{ij,n} \cdot \sqrt{\frac{\mu_2(|q_{ij,n}|)}{|x_{ij,n}|^2}} \cdot \frac{\xi_{ij}}{\rho_i + \xi_{ij}} \\ & \cdot \left[\frac{\Phi\left(1 - \frac{\beta}{2} - \rho_i, 1; -\frac{|x_{ij,n}|^2}{\mu_2(|q_{ij,n}|)} \frac{\xi_{ij}}{\rho_i + \xi_{ij}}\right)}{\Phi\left(1 - \rho_i, 1; -\frac{|x_{ij,n}|^2}{\mu_2(|q_{ij,n}|)} \frac{\xi_{ij}}{\rho_i + \xi_{ij}}\right)} \right] \\ & \cdot \frac{\Gamma\left(\frac{\beta}{2} + \rho_i\right)}{\Gamma(\rho_i)} \Bigg]^{1/\beta} \quad (14) \end{aligned}$$

ここで、 β はドメインパラメータ、 $\Gamma(\cdot)$ はガンマ関数、 $\Phi(e, f; g) = {}_1F_1(e, f; g)$ は第 1 種合流型超幾何関数である。 ξ_{ij} は事前 SN 比であり、以下のように定義さ

れる。

$$\begin{aligned} \xi_{ij} = & \kappa \frac{|\tilde{y}_{i(j-1),n}|^2}{\mu_2(|q_{ij,n}|)} \\ & + (1 - \kappa) \max\left[\frac{|x_{ij,n}|^2}{\mu_2(|q_{ij,n}|)} - 1, 0\right] \quad (15) \end{aligned}$$

また、 κ は忘却係数である。

以上より、全ての周波数 i に関して MOSIE を通すことで、時系列のスパース性が強調された分離信号の振幅スペクトログラム $|\tilde{y}_{ij,n}|$ が得られる。これを要素に持つ n 番目の分離信号の振幅スペクトログラム行列を $\tilde{\mathbf{Y}}_n$ と定義する。

3.3.4 アクティベーション行列の再推定

基底インデクスの集合 \mathcal{K}_n を次のように定義する。

$$\mathcal{K}_n = \left\{ k \mid z_{nk} > \frac{1}{N} \right\} \quad (16)$$

\mathcal{K}_n は n 番目の分離信号に寄与する基底インデクスを表す集合である。 $\mathbf{T}^{(n)}$ (要素に $t_{ik}^{(n)}$ を持つ行列) の $k \in \mathcal{K}_n$ 列目の基底ベクトルから構成される部分行列を \mathbf{T}_n とすると、これは n 番目の音源のみを表現する部分基底行列となる。同様に、 n 番目の音源のみを表現する部分アクティベーション行列 \mathbf{V}_n も得られる。時系列のスパース性が強調された分離信号 $\tilde{\mathbf{Y}}_n$ に対して式 (10) のように \mathbf{T}_n を固定した ISNMF を適用することで、 $\tilde{\mathbf{Y}}_n$ のアクティベーション行列 $\hat{\mathbf{V}}_n$ を再推定できる。この NMF の更新式は、文献 [9] より次式で与えられる。

$$\hat{v}_{kj,n} \leftarrow \hat{v}_{kj,n} \sqrt{\frac{\sum_i t_{ik,n} |\tilde{y}_{ij,n}|^2 (\sum_{k'} t_{ik',n} \hat{v}_{k',j,n})^{-2}}{\sum_i t_{ik,n} (\sum_{k'} t_{ik',n} \hat{v}_{k',j,n})^{-2}}} \quad (17)$$

ここで $t_{ik,n}$ および $\hat{v}_{kj,n}$ は、それぞれ行列 \mathbf{T}_n および $\hat{\mathbf{V}}_n$ の各要素である。

3.3.5 スパース正則化の反映

全音源に対する $\hat{\mathbf{V}}_n$ を統合して得られるアクティベーション行列 $\hat{\mathbf{V}}$ を、ILRMA の音源モデル中の \mathbf{V} に代入することで、MOSIE によるスパース正則化を ILRMA に反映できる。ただし、基底インデクスの順番が \mathbf{V} と一致するように $\hat{\mathbf{V}}_n$ を統合する必要がある。

4 評価実験

4.1 実験条件

提案手法の有効性を確認するため、音声と音楽の混合信号の音源分離実験を行った。比較対象法は、IVA および従来の ILRMA である。観測信号として、SiSEC [12] に収録されている音声および音楽信号を、RWCP database [13] に収録されている E2A インパルス応答（残響時間 300 ms）と畳み込むことで、擬似的な 2 チャネル 2 音源の信号を作成した。音源信号の到来方向として、正面を 90° として、 70° と 110° 、 50° と 130° 、 50° と 70° の三種の組み合わせについて検証した。分離性能の客観評価尺度には signal-to-distortion ratio (SDR) [14] の改善量を用いた。提案手法は、ILRMA を 20 回反復すること一度 MOSIE による正則化を行うように設定した。また、式 (17) の ISNMF の更新回数は 100 回とした。その他の条件

Table 1 Experimental conditions

Sampling frequency	16000 Hz
FFT length	256 ms (4096 points)
Window shift	64 ms (1024 points)
Number of bases K	20
Number of ILRMA iterations L	200
Domain parameter β	1.0
Forgetting factor κ	0.98

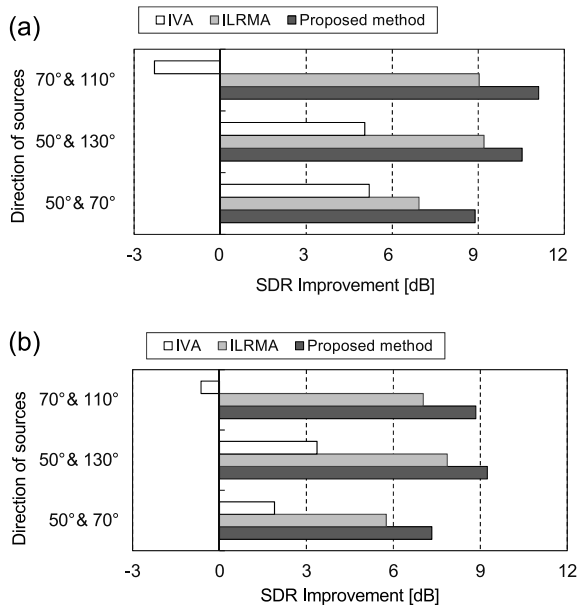


Fig. 2 Comparison of signal separation performance among IVA, conventional ILRMA and proposed method, extracting (a) instrumental signal and (b) speech signal.

およびパラメータ設定を Table 1 に示す。

文献 [7] では、 z_{nk} を 0 または 1 に固定する基底数固定型 ILRMA と、 z_{nk} を同時に推定する基底数適応法型 ILRMA の二手法が提案されている。本稿の実験では、前者の基底数固定型 ILRMA を用いている。

4.2 実験結果

Figure 2 は、異なる初期値に対して 20 回の試行を行った際の平均 SDR 改善量を示している。いずれの音源配置に関しても、提案法は従来 IVA と比較して 3 dB から 13 dB 程度、従来 ILRMA と比較して 1 dB から 2 dB 程度の SDR の改善を達成しており、より高精度な音源分離が実現されている。以上より、提案するスパース時系列正則化が音源分離に関して有効に機能していることが分かる。

5 おわりに

本稿では、より高精度なブラインド音源分離手法を目的とし、音源の独立性だけを用いる ILRMA に対して、分離信号の時間周波数の時系列がスパースになるような正則化を施す手法を提案した。正則化には、ベイズ型ポストフィルタの MOSIE を用いており、周波数ビンごとのスパース性を同時に推定する。音声と音楽の混合信号を用いた実験では、提案手法の分離精度の向上を確認した。

謝辞 本研究は、総合科学技術・イノベーション会議による革新的研究開発推進プログラム (ImPACT) およびセコム科学技術振興財団の助成を受けた。

参考文献

- [1] P. Comon, “Independent component analysis, A new concept?” *Signal processing*, vol. 36, no. 3, pp. 287–314, 1994.
- [2] N. Q. K. Duong, E. Vincent and R. Gribonval, “Under-determined reverberant audio source separation using a full-rank spatial covariance model,” *IEEE Trans. ASLP*, vol. 18, no. 7, pp. 1830–1840, 2010.
- [3] T. Kim, H. T. Attias, S.-Y. Lee and T.-W. Lee, “Blind source separation exploiting higher-order frequency dependencies,” *IEEE Trans. ASLP*, vol. 15, no. 1, pp. 70–79, 2007.
- [4] D. D. Lee and H. S. Seung, “Algorithms for non-negative matrix factorization,” *Proc. NIPS 2000*, pp. 556–562, 2001.
- [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] H. Kameoka, T. Yoshioka, M. Hamamura, J. Le Roux and K. Kashino, “Statistical model of speech signals based on composite autoregressive system with application to blind source separation,” *Proc. LVA/ICA 2010*, pp. 245–253, 2010.
- [7] 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. ASLP*, vol. 24, no. 9, pp. 1622–1637, 2016.
- [8] C. Breithaupt and R. Martin, “Analysis of the decision-directed SNR estimator for speech enhancement with respect to low-SNR and transient conditions,” *IEEE Trans. ASLP*, vol. 19, no. 2, pp. 277–289, 2011.
- [9] 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,” *Proc. IEEE MLSP 2010*, pp. 283–288, 2010.
- [10] H. Nakajima, D. Kitamura, N. Takamune, S. Koyama, H. Saruwatari, N. Ono, Y. Takahashi and K. Kondo, “Music signal separation using supervised NMF with all-pole-model-based discriminative basis deformation,” *Proc. EUSIPCO*, 2016.
- [11] Y. Murota, D. Kitamura, S. Nakai, H. Saruwatari, S. Nakamura, Y. Takahashi and K. Kondo, “Music signal separation based on bayesian spectral amplitude estimator with automatic target prior adaptation,” *Proc. ICASSP 2014*, pp. 7490–7494, 2014.
- [12] S. Araki, F. Nesta, E. Vincent, Z. Koldovský, G. Nolte, A. Ziehe and A. Benichoux, “The 2011 signal separation evaluation campaign (SiSEC2011): – Audio Source Separation –,” *Proc. LVA/ICA 2012*, pp. 414–422, 2012.
- [13] S. Nakamura, K. Hiyane, F. Asano, T. Nishiura and T. Yamada, “Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition,” *Proc. LREC 2000*, pp. 965–968, 2000.
- [14] 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.