社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

［ポスター講演］マルチチャネル音源分離のための低ランク音源モデルとスパース重畳過程に基づくネスト型ベイズ混合・因子モデル

板倉　光佑†　坂東　宜昭†　中村　栄太†　糸山　克寿†　吉井　和佳†
河原　達也†

† 京都大学 大学院情報学研究科
E-mail: †{itakura,bando,enakamura,itoyama,yoshii,kawahara}@sap.ist.i.kyoto-u.ac.jp

あらまし　本稿では，低ランク音源モデルとスパース重畳過程に基づくネスト型ベイズ混合・因子モデルを用いたマルチチャネル音源分離法について述べる．従来の音源分離では，音源モデルに対して低ランク性を仮定し因子モデルを用いたモデル化を行う非負値行列因子分解や重畳過程において音源のスパース性を仮定し混合モデルに基づいた時間周波数クラスタリングによる音源分離法などがあった．提案法ではこの音源モデルと重畳過程を統合したモデル化により音源分離を行う．また，因子モデルと混合モデルの関係性に着目し，音源モデルと重畳過程のそれぞれに対し因子モデルと混合モデルによるモデル化を行うことで複数の分離法を提案する．

キーワード　マルチチャネル音源分離，時間周波数クラスタリング，潜在的ディリクレ配分法, 非負値行列因子分解

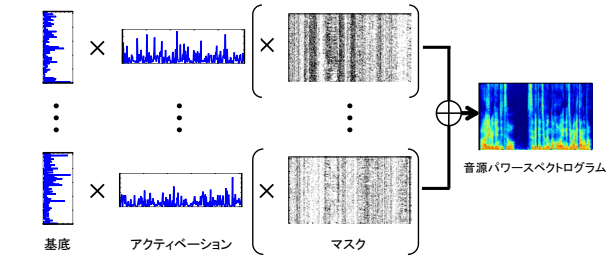
1. はじめに

マイクロホンアレイを用いたマルチチャネル音源分離において，これまでに多くの手法が提案されてきた．広く用いられている手法のうちの一つに，独立成分分析 (ICA) [1] がある．ICA は各音源の統計的な独立性を仮定することにより混合行列の逆行列である分離行列を推定し，分離を行う．この ICA をもとに独立ベクトル分析 (IVA) [2] や FastICA [3] などのさまざまな手法が提案されているが，これらの手法は共通してマイク数が音源数より少ない劣決定条件では分離できないという問題点がある．

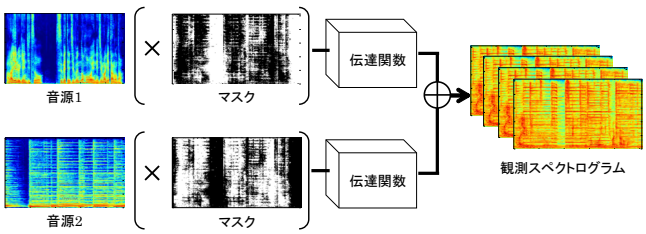
これに対し，劣決定条件でも分離が可能な手法として時間・周波数クラスタリングに基づく音源分離法が着目されている [4–9]．このアプローチでは，各音源スペクトログラムが時間・周波数領域でスパースであると仮定することで，混合音スペクトログラムの各時間・周波数ビンにおける観測はそれぞれいずれか一つの音源成分が直接観測されたものであるとみなす．これにより，音の重畳過程が混合モデルにより表される．この混合モデルを推定するため，マイク間の音の位相差とパワー差を特徴量として用いた混合ワトソン分布のクラスタリング [5–7] や各マイクでの音の位相とパワーを特徴量として用いた混合ガウス分布のクラスタリング [9,10] による分離法が提案されている．これらのクラスタリングはそれぞれの時間周波数ビンでそれぞれ独立に行われるためパーミュテーション問題が発生するが，大塚ら [9] は潜在的ディリクレ配分法 (LDA) を用いて各時間周波数ビンを各音源に割り当てたあと，それらの音源をさらにいずれかの方向に割り当てることでこのパーミュテーション問題を解決した．

また，単チャネル音源分離のための手法として非負値行列因子分解 (NMF) [11] が広く使われている．単チャネル音源分離ではマイク間の位相差などの空間的な情報を用いることができないため，NMF では音源の低ランク性を仮定することで音源を構成する因子を推定し，音源分離を行う．具体的には，音源が基底とアクティベーションの二つの因子から成るとし，観測のパワースペクトログラムを基底行列とアクティベーション行列の積により近似する．このような因子分解によるモデルを因子モデルと呼ぶ．また，この NMF をマルチチャネル音源分離のために拡張したマルチチャネル NMF (MNMF) [12] も提案されている．MNMF では観測スペクトログラムを基底行列とアクティベーション行列と空間特徴量に分解する．MNMF は空間特徴量も因子分解により推定するため，音源モデル・重畳過程がともに因子モデルによりモデル化された手法である．

本稿では，音の重畳過程と音源モデルを統合したモデルにより音源分離を行う．また，従来法では重畳過程が混合モデルでも因子モデルでもモデル化されていたことから，その二つのモデル化の関係性に着目し，音源モデルにおいても混合モデルと因子モデルの二つのモデル化を行った．つまり，因子モデルでは音源パワースペクトログラムの各時間周波数ビンが複数の基底とアクティベーションの積の和で表されたが，混合モデルでは各時間周波数ビンをいずれか一つの基底とアクティベーションの積により表す．このように，本稿では重畳過程・音源モデルに対して，混合モデル・因子モデルのそれぞれのモデル化を行うことで，それらの組み合わせにより複数の音源分離法を実現した．具体的には，図 1 に示すように，因子モデルでは音源のパワースペクトログラムが基底とアクティベーションの積の和から成り，観測スペクトログラムは音源スペクトログラムと



(a) 音源モデル



(b) 重畳過程

図 1　提案法の生成モデル．音源モデルでは各音源のパワースペクトログラムが基底・アクティベーション (・マスク) により構成される．重畳過程では混合音スペクトログラムが音源スペクトログラム・伝達関数 (・マスク) により構成される．音源モデル・重畳過程ともに混合モデルではマスクによりどれか一つの基底，もしくは音源が選択されるが，因子モデルではマスクはなくすべての基底，もしくは音源が使用される．

伝達関数から生成されるとする．混合モデルでは音源のパワースペクトログラムがマスクによって選択された基底とアクティベーションから成り，観測スペクトログラムはマスクによって選択された音源スペクトログラムと伝達関数から生成されるとする．提案法では混合モデルには LDA，因子モデルには NMF の枠組みを用いてギブスサンプリングにより音源分離を行う．

2. モデル化

提案法では，音の重畳過程と各音源の音源モデルを考慮することで音源分離を行う．モデル化には NMF のような因子モデルと LDA のような混合モデルの二通りの方法がある．本章では重畳過程と音源モデルのそれぞれに対してこれらの二通りのモデル化を行う方法について述べる．それらの組み合わせにより，表 1 に示すように音源モデル-重畳過程を因子-混合モデル，混合-混合モデル，混合-因子モデル，因子-因子モデルでそれぞれモデル化した NMF-LDA，LDA-LDA，LDA-NMF，NMF-NMF について述べる．

2.1 問題設定

まず，単一の音源を観測した時の観測モデルについて述べる．提案法では時間領域の信号に対して短時間フーリエ変換 (STFT) を行うことにより得られる時間周波数領域の信号に対してモデル化を行う．まず K 個の音源を M 個のマイクを用いて録音するとし，時刻 t，周波数 f での混合音の観測 $\boldsymbol{x}_{tf}$ と音源信号 $\boldsymbol{y}_{tf}$ を以下のように定義する．

$$\boldsymbol{x}_{tf} = [x_{tf1}, \cdots, x_{tfM}]^T \in \mathbb{C}^M \tag{1}$$

$$\boldsymbol{y}_{tf} = [y_{tf1}, \cdots, y_{tfK}]^T \in \mathbb{C}^K \tag{2}$$

表 1　モデル化方法の組み合わせ

音源モデル ＼ 重畳過程	混合モデル	因子モデル
混合モデル	LDA-LDA	LDA-NMF
因子モデル	NMF-LDA	NMF-NMF

このとき周波数領域での瞬時混合を仮定すると，音源 k のみを観測したときの観測スペクトル $\boldsymbol{x}_{tfk}$ は以下のように表される．

$$\boldsymbol{x}_{tfk} = \boldsymbol{a}_{fk} \cdot y_{tfk} \tag{3}$$

ただし，$\boldsymbol{a}_{fk}$ は周波数 f での音源 k の伝達関数である．ここで，y_{tfk} が次のような複素ガウス分布に従うとする．

$$y_{tfk} \sim \mathcal{N}_{\mathbb{C}}(0, \lambda_{tfk}) \tag{4}$$

$\lambda_{tfk} = \mathbb{E}[y_{tfk}^2]$ は時刻 t，周波数 f での音源 k のパワーを表す．このとき音源 k のみを観測したときの観測 $\boldsymbol{x}_{tfk}$ は次のような複素ガウス分布に従う．

$$\boldsymbol{x}_{tfk} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \lambda_{tfk}\boldsymbol{G}_{fk}^{-1}) \tag{5}$$

ここで，$\boldsymbol{G}_{fk}^{-1}$ は周波数 f での音源 k の空間相関行列であり，$\boldsymbol{G}_{fk}^{-1} = \boldsymbol{a}_{fk}\boldsymbol{a}_{fk}^H$ である．ただし $*^H$ はエルミート共役を示す．

2.2 重畳過程

本節では音の重畳過程に対して因子モデルと混合モデルを用いてモデル化を行う．

2.2.1 因子モデル

混合音の観測が各音源ごとの観測の和に等しいとすると式 (3) より，混合音の観測は次式で与えられる．

$$\boldsymbol{x}_{tf} = \sum_{k=1}^{K} \boldsymbol{x}_{tfk} = \sum_{k=1}^{K} \boldsymbol{a}_{fk} \cdot y_{tfk} \tag{6}$$

このとき，式 (5)，(6) とガウスの加法性により混合音の観測は次のような複素ガウス分布に従うとされる．

$$\boldsymbol{x}_{tf} \mid \boldsymbol{\lambda}, \boldsymbol{S}, \boldsymbol{G} \sim \mathcal{N}_{\mathbb{C}}\left(\boldsymbol{0}, \sum_{k=1}^{K} \lambda_{tfk}\boldsymbol{G}_{fk}^{-1}\right) \tag{7}$$

ここで，空間相関行列は音源の種類ではなく音源の方向に依存するため，音源ごとの空間相関行列 $\boldsymbol{G}_{fk}$ を方向ごとの空間相関行列 $\boldsymbol{G}_{fd}$ の重み付き和を用いて表す．このとき，方向 d の音源 k への重みを r_{kd} とすると式 (7) は次のように表される．

$$\boldsymbol{x}_{tf} \mid \boldsymbol{\lambda}, \boldsymbol{S}, \boldsymbol{G} \sim \mathcal{N}_{\mathbb{C}}\left(\boldsymbol{0}, \sum_{k=1}^{K}\sum_{d=1}^{D} \lambda_{tfk} r_{kd} \boldsymbol{G}_{fd}^{-1}\right), \tag{8}$$

2.2.2 混合モデル

音源スペクトログラムがスパースである，すなわち，各時間周波数ビンにおいて観測される音は高々一つであるとし，そのときの混合音の観測モデルについて考える．まず各時間周波数ビンにおいて観測される音源を示すための変数を $\boldsymbol{z}_{tf} = [z_{tf1}, \cdots, z_{tfK}]^T$ とする．ただし，$\boldsymbol{z}_{tf}$ は 1 of K 表現のベクトルであり，音源 k が観測されるときは $z_{tfk} = 1$ となり，それ以外のときは 0 となる．このとき，観測 $\boldsymbol{x}_{tf}$ は次式で与えられる．

$$\boldsymbol{x}_{tf} = \prod_{k=1}^{K} \boldsymbol{x}_{tfk}^{z_{tfk}} = \prod_{k=1}^{K} (\boldsymbol{a}_{fk} \cdot y_{tfk})^{z_{tfk}} \tag{9}$$

式 (5)，(9) より，観測は次の分布にしたがって生成される．

$$\boldsymbol{x}_{tf} \sim \prod_{k=1}^{K} \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \lambda_{tfk} \boldsymbol{G}_{fk}^{-1})^{z_{tfk}} \tag{10}$$

ここで，空間相関行列は音源の種類ではなく音源の方向に依存するため，空間相関行列 $\boldsymbol{G}_{fk}$ を音源ごとに独立な変数ではなく，方向ごとに独立な変数 $\boldsymbol{G}_{fd}$ として考える．このとき，式 (10) は音源 k の方向を示すベクトル $\boldsymbol{s}_k = [s_{k1}, \cdots, s_{kD}]^T$ を用いると次のように表される．

$$\boldsymbol{x}_{tf} \sim \prod_{k=1}^{K} \prod_{d=1}^{D} \mathcal{N}_{\mathbb{C}}(\boldsymbol{0}, \lambda_{tfk} \boldsymbol{G}_{fd}^{-1})^{z_{tfk} s_{kd}} \tag{11}$$

ただし，$\boldsymbol{s}_k$ は 1 of D 表現のベクトルであり，音源 k が方向 d にあるときは $s_{kd} = 1$，それ以外のときは 0 となる．

2.3 音源モデル

本節では音源 k のパワー λ_{tfk} に対して因子モデルと混合モデルを用いてモデル化を行う．

2.3.1 因子モデル

音源 k のパワースペクトログラム λ_{tfk} が低ランク性を持つと仮定すると，パワースペクトログラムは基底スペクトル w_{klf} とアクティベーション h_{klt} の積の和で表現される．この基底スペクトルとアクティベーションはそれぞれ各基底の周波数的特徴とその基底の各時刻ごとの音の大きさを示す．基底の数を L とするとパワースペクトログラムは次式で表される．

$$\lambda_{tfk} = \sum_{l} w_{klf} h_{klt} \tag{12}$$

2.3.2 混合モデル

混合モデルでも因子モデルのように基底スペクトル w_{klf} とアクティベーション h_{klt} を用いてパワースペクトログラム λ_{tfk} を表現する．ただし，混合モデルでは全ての基底の和でパワーを表現するのに対し，提案法では各時間周波数ビンごとに一つの基底 l' を選択することによりパワー λ_{tfk} を

$$\lambda_{tfk} = w_{kl'f} h_{kl't} \tag{13}$$

と表す．その基底を選択するための変数を $\boldsymbol{u}_{tfk} = [u_{tfk1}, \cdots, u_{tfkL}]^T$ とする．ただし，$\boldsymbol{u}_{tfk}$ は 1 of L 表現のベクトルであり，基底 l が用いられるときは $u_{tfkl} = 1$，それ以外のときは 0 となる．

2.4 各モデルの尤度関数

本章では，音源モデル-重畳過程を因子-混合モデル，混合-混合モデル，因子-因子モデル，混合-因子モデルでそれぞれモデル化した NMF-LDA，LDA-LDA，NMF-NMF，LDA-NMF の観測モデルについて述べる．

2.4.1 NMF-LDA

NMF-LDA は重畳過程が混合モデル，音源モデルが因子モデルなので式 (8)，(13) より尤度関数は次のようになる．

$$\boldsymbol{x}_{tf} \mid \boldsymbol{W}, \boldsymbol{H}, \boldsymbol{Z}, \boldsymbol{S}, \boldsymbol{G} \sim \prod_{k,d} \mathcal{N}_{\mathbb{C}} \left(\boldsymbol{0}, \sum_{l} w_{klf} h_{klt} \boldsymbol{G}_{fd}^{-1} \right)^{z_{tfk} s_{kd}} \tag{14}$$

2.4.2 LDA-LDA

LDA-LDA は重畳過程が混合モデル，音源モデルが混合モデルなので式 (11)，(13) より尤度関数は次のようになる．

$$\boldsymbol{x}_{tf} \mid \boldsymbol{W}, \boldsymbol{H}, \boldsymbol{Z}, \boldsymbol{S}, \boldsymbol{U}, \boldsymbol{G} \sim \prod_{k,d,l} \mathcal{N}_{\mathbb{C}} \left(\boldsymbol{0}, w_{klf} h_{klt} \boldsymbol{G}_{fd}^{-1} \right)^{z_{tfk} s_{kd} u_{tfkl}} \tag{15}$$

2.4.3 NMF-NMF

NMF-NMF は重畳過程が因子モデル，音源モデルが因子モデルなので式 (8)，(12) より尤度関数は次のようになる．

$$\boldsymbol{x}_{tf} \mid \boldsymbol{W}, \boldsymbol{H}, \boldsymbol{R}, \boldsymbol{G} \sim \mathcal{N}_{\mathbb{C}} \left(\boldsymbol{0}, \sum_{k,d,l} w_{klf} h_{klt} r_{kd} \boldsymbol{G}_{fd}^{-1} \right) \tag{16}$$

2.4.4 LDA-NMF

LDA-NMF は重畳過程が混合モデル，音源モデルが因子モデルなので式 (11)，(12) より尤度関数は次のようになる．

$$\boldsymbol{x}_{tf} \mid \boldsymbol{W}, \boldsymbol{H}, \boldsymbol{R}, \boldsymbol{G} \sim \prod_{l} \mathcal{N}_{\mathbb{C}} \left(\boldsymbol{0}, \sum_{k,d} w_{klf} h_{klt} r_{kd} \boldsymbol{G}_{fd}^{-1} \right)^{u_{tfkl}} \tag{17}$$

しかし，このモデルでは音源 k に関するパラメータはガウス分布の分散パラメータとして因子分解により推定されるべきであるが，基底の選択を行う u_{tfkl} が混合ガウス分布における分布を選択するパラメータとして存在している．これにより，u_{tfkl} は因子分解では推定することができないためこのモデル化方法では LDA-NMF を実現することができない．

2.5 事前分布の設計

提案法では式 (14)–(16) のパラメータに対し，それぞれ適切な事前分布を与えることで推論を行う．まず，$\boldsymbol{z}_{tf}$，$\boldsymbol{s}_k$，$\boldsymbol{u}_{tfk}$ はクラスタリングによる推論を行うためにカテゴリカル分布から生成されるとする．

$$\boldsymbol{z}_{tf} \mid \boldsymbol{\pi}_t \sim \text{Categorical}(\boldsymbol{\pi}_t) \tag{18}$$

$$\boldsymbol{s}_k \mid \boldsymbol{\phi} \sim \text{Categorical}(\boldsymbol{\phi}) \tag{19}$$

$$\boldsymbol{u}_{tfk} \mid \boldsymbol{\psi}_{tk} \sim \text{Categorical}(\boldsymbol{\psi}_{tk}) \tag{20}$$

ここで，ハイパーパラメータ $\boldsymbol{\pi}_t$，$\boldsymbol{\phi}$，$\boldsymbol{\psi}_{tk}$ は観測に依存して変動するため推論を必要とする．したがって $\boldsymbol{\pi}_t$，$\boldsymbol{\phi}$，$\boldsymbol{\psi}_{tk}$ はカテゴリカル分布と共役なディリクレ分布から生成されるとする．

$$\boldsymbol{\pi}_t \sim \text{Dirichlet}(a_0^{\pi} \boldsymbol{1}_K) \tag{21}$$

$$\boldsymbol{\phi} \sim \text{Dirichlet}(a_0^{\phi} \boldsymbol{1}_D) \tag{22}$$

$$\boldsymbol{\psi}_{tk} \sim \text{Dirichlet}(a_0^{\psi} \boldsymbol{1}_L) \tag{23}$$

ここで，$\boldsymbol{1}_N$ は要素が全て 1 の N 次元のベクトルとし，a_0^* はハイパーパラメータとする．また，空間相関行列 $\boldsymbol{G}_{fd}$，基底 w_{klf}，アクティベーション h_{klt}，重み r_{kd} は式 (14)–(16) と共

役になるように事前分布として次のような分布を与える．

$$\boldsymbol{G}_{fd} \sim \mathcal{W}_{\mathbb{C}}(\nu, \boldsymbol{G}^0_{fd}) \tag{24}$$

$$w_{klf} \sim \mathrm{Gamma}(a^w_0, b^w_0) \tag{25}$$

$$h_{klt} \sim \mathrm{Gamma}(a^h_0, b^h_0) \tag{26}$$

$$r_{kd} \sim \mathrm{Gamma}(a^r_0, b^r_0) \tag{27}$$

ここで，ν，a^*_0，b^*_0 はハイパーパラメータであり，$\mathcal{W}_{\mathbb{C}}$ は複素ウィシャート分布 (付録参照) とする．

2.6 分離音の生成

各音源信号は，重畳過程が因子モデルのときはマルチチャネル Wiener フィルタ，混合モデルのときは時間周波数マスキングにより復元することができる．マルチチャネル Wiener フィルタでは音源 k の信号 $\boldsymbol{x}^k_{tf}$ は以下の式により復元される．

$$\boldsymbol{x}^k_{tf} = \boldsymbol{Y}_{tfk}\boldsymbol{Y}_{tf}^{-1}\boldsymbol{x}_{tf} \tag{28}$$

ここで，モデルの推論の結果を $\boldsymbol{Y}_{tfk} = \sum_{d,l} w_{klf}h_{klt}r_{kd}\boldsymbol{G}^{-1}_{fd}$，$\boldsymbol{Y}_{tf} = \sum_k \boldsymbol{Y}_{tfk}$ と定義した．

時間周波数マスキングでは，時間周波数領域で音源方向ごとにマスクを推定することにより分離音を生成する．ギブスサンプリングで i 回目の試行により得られるサンプルを $\boldsymbol{W}^{(i)}, \boldsymbol{H}^{(i)}, \boldsymbol{G}^{(i)}, \boldsymbol{Z}^{(i)}, \boldsymbol{S}^{(i)}, (\boldsymbol{U}^{(i)})$ とすると，時刻 t 周波数 f での音源方向 d に対するマスク M^d_{tf} は次のようになる．

$$M^d_{tf} = \frac{1}{I}\sum_{i=1}^{I}\sum_{k=1}^{K} z^{(i)}_{tfk}s^{(i)}_{kd} \tag{29}$$

このマスクを用いると方向 d の音源の信号 $\boldsymbol{x}^d_{tf}$ は

$$\boldsymbol{x}^d_{tf} = M^d_{tf}\boldsymbol{x}_{tf} \tag{30}$$

により復元される．

3. 推　　論

提案法では，観測データ集合 $\boldsymbol{X}$ に対するすべてのパラメータの事後分布 $p(\boldsymbol{G}, \boldsymbol{W}, \boldsymbol{H}, (\boldsymbol{Z}, \boldsymbol{S}, \boldsymbol{R}, \boldsymbol{U}, \boldsymbol{\pi}, \boldsymbol{\phi}, \boldsymbol{\psi})|\boldsymbol{X})$ を最大とするパラメータを求めることを目標とする．ただし，これらのパラメータを解析的に求めることは困難なので，提案法ではこれらのパラメータをギブスサンプリングにより求めることとする．ただし，提案法では $\boldsymbol{\pi}$，$\boldsymbol{\phi}$，$\boldsymbol{\psi}$ は積分消去を行い，残りのパラメータ $\boldsymbol{\Theta} = \{\boldsymbol{G}, \boldsymbol{W}, \boldsymbol{H}, (\boldsymbol{Z}, \boldsymbol{S}, \boldsymbol{R}, \boldsymbol{U})\}$ を求めることとする．ギブスサンプリングではそれぞれのパラメータの事後分布を求め，それらの事後分布からサンプリングを繰り返すことにより推定を行う．

3.1 NMF-LDA の更新

NMF-LDA では $\boldsymbol{G}_{fd}$，$\boldsymbol{z}_{tf}$，$\boldsymbol{s}_k$ は尤度と事前分布の積により得られる以下の事後分布からサンプルすることにより更新される．

$$\boldsymbol{G}_{fd} \mid \boldsymbol{X}, \boldsymbol{\Theta}_{\neg \boldsymbol{G}_{fd}} \sim \mathcal{W}_{\mathbb{C}}(\nu'_{fd}, \boldsymbol{G}'_{fd}) \tag{31}$$

$$\boldsymbol{z}_{tf} \mid \boldsymbol{X}, \boldsymbol{\Theta}_{\neg \boldsymbol{z}_{tf}} \sim \mathrm{Categorical}(\boldsymbol{\pi}'_{tf}) \tag{32}$$

$$\boldsymbol{s}_k \mid \boldsymbol{X}, \boldsymbol{\Theta}_{\neg \boldsymbol{s}_k} \sim \mathrm{Categorical}(\boldsymbol{\phi}'_k) \tag{33}$$

ここで，$\boldsymbol{\Theta}_{\neg *}$ は $\boldsymbol{\Theta}$ から $*$ の要素のみを除いた集合とする．また，ハイパーパラメータ ν'_{fd}，$\boldsymbol{G}'_{fd}$，$\boldsymbol{\pi}'_{tf}$，$\boldsymbol{\phi}'_k$ は次のようになる．

$$\nu'_{fd} = \nu + \sum_{t,k} z_{tfk}s_{kd} \tag{34}$$

$$\boldsymbol{G}'^{-1}_{fd} = (\boldsymbol{G}^0_{fd})^{-1} + \sum_{t,k}\frac{\boldsymbol{x}_{tf}\boldsymbol{x}_{tf}{}^H}{\sum_l w_{klf}h_{klt}} z_{tfk}s_{kd} \tag{35}$$

$$\pi'_{tfk} = \prod_d \left\{ \left| \frac{\boldsymbol{G}_{fd}}{\sum_l w_{klf}h_{klt}} \right| \exp\left(-\frac{\boldsymbol{x}^H_{tf}\boldsymbol{G}_{fd}\boldsymbol{x}_{tf}}{\sum_l w_{klf}h_{klt}} \right) \right\}^{s_{kd}} \times (a^\pi_0 + n^{\neg tf}_{tk}) \tag{36}$$

$$\phi'_{kd} = \prod_{t,f} \left\{ \left| \frac{\boldsymbol{G}_{fd}}{\sum_l w_{klf}h_{klt}} \right| \exp\left(-\frac{\boldsymbol{x}^H_{tf}\boldsymbol{G}_{fd}\boldsymbol{x}_{tf}}{\sum_l w_{klf}h_{klt}} \right) \right\}^{z_{tfk}} \times (a^\phi_0 + c^{\neg k}_d) \tag{37}$$

ここで，$n^{\neg tf}_{tk}$ は時刻 t 周波数 f でのサンプルを除いて，時刻 t において音源 k に割り当てられた時間周波数ビンの数を表し，$c^{\neg k}_d$ は音源 k を除いて方向 d に割り当てられた音源の数を表す．また，n_{tkl} は時刻 t において音源 k，基底 l に割り当てられた時間周波数ビンの数を示し，$n^{\neg tfk}_{tkl}$ は n_{tkl} から時刻 t 周波数 f で音源 k に割り当てられた要素を除いたものである．

w_{klf} と h_{klt} は事後分布から直接サンプルを得ることは難しいため，補助関数法を用いて得られる下限からサンプリングを行う．ただし，この下限は補助変数について最大化したとき，もとの対数尤度と一致する．まず，対数尤度は次式で与えられる．

$$\log p(\boldsymbol{X} \mid \boldsymbol{W}, \boldsymbol{H}, \boldsymbol{G}, \boldsymbol{Z}, \boldsymbol{S}) \overset{\mathrm{c}}{=} \sum_{t,f,k,d} z_{tfk}s_{kd}\left(-M\log\left|\sum_l w_{klf}h_{klt}\right| - \frac{\boldsymbol{x}^H_{tf}\boldsymbol{G}_{fd}\boldsymbol{x}_{tf}}{\sum_l w_{klf}h_{klt}} \right) \tag{38}$$

この対数尤度に対しテイラー展開と Jensen の不等式を用いて下限を計算する [13]．まず一次のテイラー展開により，凸関数 $f(y) = -\log y$ の下限は以下のように計算される．

$$-\log y \geqq -\log\alpha + 1 - \frac{y}{\alpha} \tag{39}$$

α は補助変数であり，等号は $\alpha = y$ のとき成立する．次に Jensen の不等式により，凹関数 $g(y) = -\frac{1}{y}$ に対し次の不等式が成立する．

$$-\frac{1}{\sum_k y_k} \geqq -\sum_k \beta^2_k \frac{1}{y_k} \tag{40}$$

β_k は $\beta_k \geqq 0$ かつ $\sum_k \beta_k = 1$ を満たす補助変数であり，等号は $\beta_k = y_k(\sum_k y_k)^{-1}$ のとき成立する．

これらの二つの不等式により対数尤度の下限は次式により求められる．

$$\log p(\boldsymbol{X} \mid \boldsymbol{W}, \boldsymbol{H}, \boldsymbol{G}, \boldsymbol{Z}, \boldsymbol{S}) \overset{\mathrm{c}}{\geqq} \sum_{t,f,k,d} z_{tfk}s_{kd}\left(-M\left(\log|\alpha_{tf}| - 1 + \frac{\sum_l w_{klf}h_{klt}}{\alpha_{tf}} \right) - \boldsymbol{x}^H_{tf}\boldsymbol{G}_{fd}\boldsymbol{x}_{tf}\sum_l \frac{\beta_{tfl}}{w_{klf}h_{klt}} \right) \tag{41}$$

ここで，α_{tf} と β_{tfl} は補助変数であり，次の式を満たすとき等号が成立する．

$$\alpha_{tf} = \sum_l w_{klf} h_{klt} \tag{42}$$

$$\beta_{tfl} = w_{klf} h_{klt} \left(\sum_l w_{klf} h_{klt} \right)^{-1} \tag{43}$$

w_{klf} と h_{klt} は式 (25)，(26)，(41) の積により得られる次の分布からサンプルされる．

$$w_{kfl} \mid \boldsymbol{X}, \boldsymbol{\Theta}_{\neg w_{klf}} \sim \mathrm{GIG}(\gamma^w_{klf}, \rho^w_{kfl}, \tau^w_{kfl}), \tag{44}$$

$$h_{klt} \mid \boldsymbol{X}, \boldsymbol{\Theta}_{\neg h_{klt}} \sim \mathrm{GIG}(\gamma^h_{klt}, \rho^h_{klt}, \tau^h_{klt}), \tag{45}$$

ここで，GIG は一般化逆ガウス分布 (付録参照) を示す．また，ハイパーパラメータ γ^*_*，ρ^*_*，τ^*_* は次のようになる．

$$\gamma^w_{klf} = a^w_0, \tag{46}$$

$$\rho^w_{klf} = b^w_0 + \sum_{t,d} \frac{M h_{klt}}{\alpha_{tfk}}, \tag{47}$$

$$\tau^w_{klf} = \sum_{t,d} \mathrm{tr}(\boldsymbol{X}_{tf} \boldsymbol{G}_{fd}) \frac{\beta^2_{tfkl}}{h_{klt}}, \tag{48}$$

$$\gamma^h_{klt} = a^h_0, \tag{49}$$

$$\rho^h_{klt} = b^h_0 + \sum_{f,d} \frac{M w_{klf}}{\alpha_{tfk}}, \tag{50}$$

$$\tau^h_{klt} = \sum_{f,d} \mathrm{tr}(\boldsymbol{X}_{tf} \boldsymbol{G}_{fd}) \frac{\beta^2_{tfkl}}{w_{klf}}. \tag{51}$$

3.2 LDA-LDA の更新

LDA-LDA ではすべてのパラメータが，尤度関数と事前分布の積により得られる事後分布からサンプルすることが可能である．事後分布は次のように定められる．

$$\boldsymbol{G}_{fd} \mid \boldsymbol{X}, \boldsymbol{\Theta}_{\neg \boldsymbol{G}_{fd}} \sim \mathcal{W}_{\mathbb{C}}(\nu'_{fd}, \boldsymbol{G}'_{fd}) \tag{52}$$

$$\boldsymbol{z}_{tf} \mid \boldsymbol{X}, \boldsymbol{\Theta}_{\neg \boldsymbol{z}_{tf}} \sim \mathrm{Categorical}(\boldsymbol{\pi}'_{tf}) \tag{53}$$

$$\boldsymbol{s}_k \mid \boldsymbol{X}, \boldsymbol{\Theta}_{\neg \boldsymbol{s}_k} \sim \mathrm{Categorical}(\boldsymbol{\phi}'_k) \tag{54}$$

$$\boldsymbol{u}_{tfkl} \mid \boldsymbol{X}, \boldsymbol{\Theta}_{\neg \boldsymbol{u}_{tfkl}} \sim \mathrm{Categorical}(\boldsymbol{\psi}'_{tfk}) \tag{55}$$

$$w_{klf} \mid \boldsymbol{X}, \boldsymbol{\Theta}_{\neg w_{klf}} \sim \mathrm{GIG}(\gamma^w_{klf}, \rho^w_{klf}, \tau^w_{klf}) \tag{56}$$

$$h_{klt} \mid \boldsymbol{X}, \boldsymbol{\Theta}_{\neg h_{klt}} \sim \mathrm{GIG}(\gamma^h_{klt}, \rho^h_{klt}, \tau^h_{klt}) \tag{57}$$

ここで，ハイパーパラメータ ν'_{fd}，$\boldsymbol{G}'_{fd}$，$\boldsymbol{\pi}'_{tf}$，$\boldsymbol{\phi}'_k$，$\boldsymbol{\psi}$，$\boldsymbol{\gamma}^*$，$\boldsymbol{\rho}^*$，$\boldsymbol{\tau}^*$ は次のようになる．

$$\nu'_{fd} = \nu + \sum_{t,k} z_{tfk} s_{kd} \tag{58}$$

$$\boldsymbol{G}'^{-1}_{fd} = (\boldsymbol{G}^0_{fd})^{-1} + \sum_{t,k,l} \frac{\boldsymbol{x}_{tf} \boldsymbol{x}_{tf}{}^H}{w_{klf} h_{klt}} z_{tfk} s_{kd} u_{tfkl} \tag{59}$$

$$\pi'_{tfk} = \prod_{d,l} \left\{ \left| \frac{\boldsymbol{G}_{fd}}{w_{klf} h_{klt}} \right| \exp\left(-\frac{\boldsymbol{x}^H_{tf} \boldsymbol{G}_{fd} \boldsymbol{x}_{tf}}{w_{klf} h_{klt}} \right) \right\}^{s_{kd} u_{tfkl}} \times (a^\pi_0 + n^{\neg tf}_{tk}) \tag{60}$$

$$\phi'_{kd} = \prod_{t,f,l} \left\{ \left| \frac{\boldsymbol{G}_{fd}}{w_{klf} h_{klt}} \right| \exp\left(-\frac{\boldsymbol{x}^H_{tf} \boldsymbol{G}_{fd} \boldsymbol{x}_{tf}}{w_{klf} h_{klt}} \right) \right\}^{z_{tfk} u_{tfkl}} \times (a^\phi_0 + c^{\neg k}_d) \tag{61}$$

$$\psi'_{tfkl} = \prod_d \left\{ \left| \frac{\boldsymbol{G}_{fd}}{w_{klf} h_{klt}} \right| \exp\left(-\frac{\boldsymbol{x}^H_{tf} \boldsymbol{G}_{fd} \boldsymbol{x}_{tf}}{w_{klf} h_{klt}} \right) \right\}^{z_{tfk} s_{kd}} \times (a^\psi_0 + n^{\neg tfk}_{tkl}) \tag{62}$$

$$\gamma^w_{klf} = a^w_0 - M n_{fkl} \tag{63}$$

$$\rho^w_{klf} = b^w_0 \tag{64}$$

$$\tau^w_{klf} = \sum_{t,d} \frac{\boldsymbol{x}_{tf} \boldsymbol{G}_{fd} \boldsymbol{x}_{tf}}{h_{klt}} z_{tfk} s_{kd} u_{tfkl} \tag{65}$$

$$\gamma^h_{klt} = a^h_0 - M n_{tkl} \tag{66}$$

$$\rho^h_{klt} = b^h_0 \tag{67}$$

$$\tau^h_{klt} = \sum_{f,d} \frac{\boldsymbol{x}_{tf} \boldsymbol{G}_{fd} \boldsymbol{x}_{tf}}{w_{klf}} z_{tfk} s_{kd} u_{tfkl} \tag{68}$$

ここで，n_{fkl} は周波数 f において音源 k，基底 l に割り当てられた時間周波数ビンの数を示す．

3.3 NMF-NMF の更新

NMF-NMF ではどのパラメータも事後分布から直接サンプルを得ることは難しいため，補助関数法を用いて得られる下限からサンプリングを行う．ただし，この補助関数は補助変数について最大化したとき，もとの対数尤度と一致する．まず，対数尤度は次式で与えられる．

$$\log p(\boldsymbol{X} | \boldsymbol{W}, \boldsymbol{H}, \boldsymbol{R}, \boldsymbol{G}) \overset{\mathrm{c}}{=} \sum_{tf} \left(-\log |\boldsymbol{Y}_{tf}| - \mathrm{tr}(\boldsymbol{X}_{tf} \boldsymbol{Y}^{-1}_{tf}) \right) \tag{69}$$

ここで，$\boldsymbol{X}_{tf} = \boldsymbol{x}^H_{tf} \boldsymbol{x}_{tf}$，$\boldsymbol{Y}_{tfkld} = w_{klf} h_{klt} r_{kd} \boldsymbol{G}^{-1}_{fd}$，$\boldsymbol{Y}_{tf} = \sum_{kld} \boldsymbol{Y}_{tfkld}$ とした．式 (69) の下限を求めるために二つの不等式を用いる [14]．まず，多変量の凸関数 $f(\boldsymbol{Y}) = -\log |\boldsymbol{Y}|$ $(\boldsymbol{Y} \succeq \boldsymbol{0} \in \mathbb{C}^{M \times M})$ に対して一次の多変量のテイラー展開を用いると次の不等式が得られる．

$$-\log |\boldsymbol{Y}| \geqq -\log |\boldsymbol{\Omega}| - \mathrm{tr}(\boldsymbol{\Omega}^{-1} \boldsymbol{Y}) + M, \tag{70}$$

等号は $\boldsymbol{\Omega} = \boldsymbol{Z}$ のとき成立する．次に，凹関数 $g(\boldsymbol{Y}) = -\mathrm{tr}(\boldsymbol{Y}^{-1} \boldsymbol{A})$ に対して次の不等式が成り立つ．

$$-\mathrm{tr}\left(\left(\sum_{k=1}^K \boldsymbol{Y}_k \right)^{-1} \boldsymbol{A} \right) \geqq -\sum_{k=1}^K \mathrm{tr}\left(\boldsymbol{Y}_k^{-1} \boldsymbol{\Phi}_k \boldsymbol{A} \boldsymbol{\Phi}_k^H \right) \tag{71}$$

ここで $\{\boldsymbol{\Phi}_k\}_{k=1}^K$ は補助変数であり，$\sum_k \boldsymbol{\Phi}_k = \boldsymbol{I}$ を満たす．また，等号は $\boldsymbol{\Phi}_k = \boldsymbol{Z}_k (\sum_{k'} \boldsymbol{Z}_{k'})^{-1}$ のとき成立する．

不等式 (70)，(71) を用いると対数尤度 (式 (69)) の下限は次のように求められる．

$$\begin{aligned} &\log p(\boldsymbol{X} | \boldsymbol{W}, \boldsymbol{H}, \boldsymbol{S}, \boldsymbol{G}) \\ &\overset{\mathrm{c}}{\geqq} \sum_{tf} \left(-\mathrm{tr}\left(\boldsymbol{Y}_{tf} \boldsymbol{\Omega}^{-1}_{tf} \right) - \log |\boldsymbol{\Omega}_{tf}| + M \right) \\ &\quad - \sum_{tfkld} \mathrm{tr}\left(\boldsymbol{Y}^{-1}_{tfkld} \boldsymbol{\Phi}_{tfkld} \boldsymbol{X}_{tf} \boldsymbol{\Phi}_{tfkld} \right) \end{aligned} \tag{72}$$

ここで，補助変数 $\boldsymbol{\Omega}_{tf}$ と $\boldsymbol{\Phi}_{tfkld}$ が次式を満たすとき等号が成立する．

$$\boldsymbol{\Omega}_{tf} = \boldsymbol{Y}_{tf} \tag{73}$$

$$\boldsymbol{\Phi}_{tfkld} = \boldsymbol{Y}_{tfkld}\boldsymbol{Y}_{tf}^{-1} \tag{74}$$

w_{klf}，h_{klt}，s_{kd}，$\boldsymbol{G}_{fd}$ は式 (24)–(27)，(72) の積により得られる次の分布からサンプルされる．

$$w_{klf} \mid \boldsymbol{X}, \boldsymbol{\Theta}_{\neg w_{klf}} \sim \mathrm{GIG}(\gamma^w_{klf}, \rho^w_{klf}, \tau^w_{klf}) \tag{75}$$

$$h_{klt} \mid \boldsymbol{X}, \boldsymbol{\Theta}_{\neg h_{klt}} \sim \mathrm{GIG}(\gamma^h_{klt}, \rho^h_{klt}, \tau^h_{klt}) \tag{76}$$

$$r_{kd} \mid \boldsymbol{X}, \boldsymbol{\Theta}_{\neg r_{kd}} \sim \mathrm{GIG}(\gamma^r_{kd}, \rho^r_{kd}, \tau^r_{kd}) \tag{77}$$

$$\boldsymbol{G}_{fd} \mid \boldsymbol{X}, \boldsymbol{\Theta}_{\neg \boldsymbol{G}_{fd}} \sim \mathrm{MGIG}_{\mathbb{C}}(\nu_0, \boldsymbol{R}_{fd}, \boldsymbol{U}_{fd}) \tag{78}$$

ここで，$\mathrm{MGIG}_{\mathbb{C}}$ は complex matrix generalized inverse Gaussian (付録参照) を示す．ここで，ハイパーパラメータ γ^*_*，ρ^*_*，τ^*_*，$\boldsymbol{R}_{fd}$，$\boldsymbol{U}_{fd}$ は次のように定められる．

$$\gamma^w_{klf} = a^w_0 \tag{79}$$

$$\rho^w_{klf} = b^w_0 + \sum_{td} h_{klt} r_{kd} \mathrm{tr}(\boldsymbol{G}_{fd}^{-1}\boldsymbol{\Omega}_{tf}^{-1}) \tag{80}$$

$$\tau^w_{klf} = \sum_{td} h^{-1}_{klt} r^{-1}_{kd} \mathrm{tr}\left(\boldsymbol{G}_{fd}\boldsymbol{\Phi}_{tfkld}\boldsymbol{X}_{tf}\boldsymbol{\Phi}_{tfkld}\right) \tag{81}$$

$$\gamma^h_{klt} = a^h_0 \tag{82}$$

$$\rho^h_{klt} = b^h_0 + \sum_{fd} w_{klf} r_{kd} \mathrm{tr}(\boldsymbol{G}_{fd}^{-1}\boldsymbol{\Omega}_{tf}^{-1}) \tag{83}$$

$$\tau^h_{klt} = \sum_{fd} w^{-1}_{klf} r^{-1}_{kd} \mathrm{tr}\left(\boldsymbol{G}_{fd}\boldsymbol{\Phi}_{tfkld}\boldsymbol{X}_{tf}\boldsymbol{\Phi}_{tfkld}\right) \tag{84}$$

$$\gamma^r_{klf} = a^r_0 \tag{85}$$

$$\rho^r_{kd} = b^r_0 + \sum_{tfl} w_{klf} h_{klt} \mathrm{tr}(\boldsymbol{G}_{fd}^{-1}\boldsymbol{\Omega}_{tf}^{-1}) \tag{86}$$

$$\tau^r_{kd} = \sum_{tfl} w^{-1}_{klf} h^{-1}_{klt} \mathrm{tr}\left(\boldsymbol{G}_{fd}\boldsymbol{\Phi}_{tfkld}\boldsymbol{X}_{tf}\boldsymbol{\Phi}_{tfkld}\right) \tag{87}$$

$$\boldsymbol{R}_{fd} = (\boldsymbol{G}^0_{fd})^{-1} + \sum_{tkl} w^{-1}_{klf} h^{-1}_{klt} r^{-1}_{kd} \boldsymbol{\Phi}_{tfkld}\boldsymbol{X}_{tf}\boldsymbol{\Phi}_{tfkld} \tag{88}$$

$$\boldsymbol{U}_{fd} = \sum_{tkl} w_{klf} h_{klt} r_{kd} \boldsymbol{\Omega}^{-1}_{tf} \tag{89}$$

4. 評価実験

提案法の分離性能を評価するため，シミュレーションにより混合した音を用いた実験を行った．比較手法として，IVA [15]，マルチチャネル NMF (MNMF) [12]，音源モデルをスパースとし空間モデルに LDA を用いた分離法 (LDA) [9] を用いた．また，LDA では音源数の同時推定も行うが，条件を対等にするため音源数は既知とした．

4.1 実験条件

図 2 に音源とマイクの配置を示す．残響時間 400 ms のインパルス応答を用いて 3 音源を混合した音声を用いた．マイク数は 4 とした．混合音には，音声のみの混合音と音楽のみの混合音，音声と音楽の混合音をそれぞれ 10 個ずつ使用した．用いる音楽と音声は SISEC [16] と JNAS の音素バランス文 [17] から選択した．サンプリング周波数は 16 kHz とし，STFT では窓幅 512 のハミング窓をシフト幅 256 で使用した．基底数 $L = 20$ とし，ハイパーパラメータは，$\nu = M + 1$，$a^\pi_0 = a^\phi_0 = 10$，

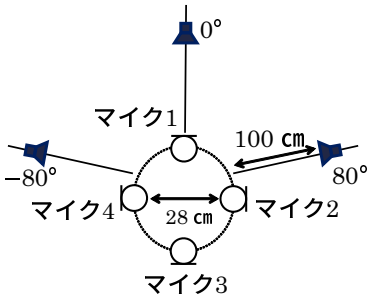


図 2　マイク配置

表 2　音楽による評価

	SDR	SIR	SAR
IVA	0.3 dB	4.9 dB	5.7 dB
MNMF	1.0 dB	6.2 dB	6.7 dB
LDA	0.7 dB	7.4 dB	4.1 dB
NMF-LDA	0.5 dB	**8.7** dB	3.2 dB
LDA-LDA	0.7 dB	**8.7** dB	3.3 dB
NMF-NMF	**3.2** dB	8.1 dB	**7.5** dB

表 3　音声による評価

	SDR	SIR	SAR
IVA	3.4 dB	7.5 dB	7.1 dB
MNMF	4.8 dB	10.0 dB	**7.7** dB
LDA	5.5 dB	15.1 dB	6.3 dB
NMF-LDA	4.2 dB	14.0 dB	5.2 dB
LDA-LDA	5.8 dB	**17.0** dB	6.3 dB
NMF-NMF	**6.0** dB	12.6 dB	7.5 dB

表 4　音楽 + 音声による評価

	SDR	SIR	SAR
IVA	0.1 dB	5.3 dB	5.3 dB
MNMF	1.8 dB	8.6 dB	6.1 dB
LDA	2.4 dB	11.5 dB	4.5 dB
NMF-LDA	1.1 dB	9.8 dB	4.1 dB
LDA-LDA	2.8 dB	**14.2** dB	3.9 dB
NMF-NMF	**4.9** dB	13.0 dB	**6.6** dB

$a^\psi_0 = a^w_0 = a^h_0 = a^r_0 = b^w_0 = b^h_0 = b^r_0 = 1$ とした．また，$G^0_{fd} = (\boldsymbol{a}_{fd}\boldsymbol{a}^H_{fd} + 0.01 \times \boldsymbol{I})^{-1}$ とし，$\boldsymbol{a}_{fd}$ には無響室で 5° 間隔で録音したインパルス応答を用いた．ギブスサンプリングの試行回数は 200 回とし，はじめの 180 回は burn-in として棄却した．評価尺度として，Signal-to-distortion ratio (SDR)，Signal-to-inference ratio (SIR)，Signal-to-artificial ratio (SAR) を用いた．SDR は総合的な分離性能，SIR は目的音以外の音の除去性能，SAR は分離音の歪みの少なさを表す尺度である．

4.2 実験結果

表 2，3，4 に実験結果を示す．それぞれの条件において最も数値が大きくなったものを太字で示した．SIR は LDA-LDA が最も高い性能を示したが，SDR と SAR は NMF-NMF が最も高い性能を示した．また，提案法では事前分布のパラメータとして使用するマイクロホンアレイのインパルス応答を必要とするが，無響室で録音したインパルス応答を用いても残響時間 400ms の環境下での混合音の分離ができたことから，未知の環境下においても使用するマイクロホンアレイが同じであれば提案法により音源分離が可能であると考えられる．

5. おわりに

本稿では，音源モデル・重畳過程のそれぞれに対して，因子モデルと混合モデルを用いたモデル化を行うことで複数の音源分離法を提案した．それぞれのモデル化による音源分離法を比較すると，SDR，SAR の観点では音源モデルと重畳過程が共に因子モデルによりモデル化された手法が最も性能が高く，SIR の観点では音源モデルと重畳過程が共に混合モデルでモデル化された手法が最も性能が高いことを確認した．ただし，音源モデルが混合モデル，重畳過程が因子モデルでモデル化された手法はまだ実現できておらず，今後はそのモデル化を行う方法について考える必要がある．

付　　録

複素ウィシャート分布と一般化逆ガウス分布の確率密度関数は以下のとおりである．

$$\mathcal{W}_{\mathbb{C}}(\boldsymbol{G}|\nu,\boldsymbol{G}^0)=\frac{|\boldsymbol{G}|^{\nu-M}\exp(-\mathrm{tr}(\boldsymbol{G}(\boldsymbol{G}^0)^{-1}))}{|\boldsymbol{G}^0|^{\nu}\pi^{M(M-1)/2}\prod_{m=0}^{M-1}\Gamma(\nu-m)} \quad \text{(A·1)}$$

$$\mathrm{GIG}(y|\gamma,\rho,\tau)=\frac{\exp\{(\gamma-1)\log y-\rho y-\tau/y\}\rho^{\gamma/2}}{2\tau^{\gamma/2}\mathcal{K}_{\gamma}(2\sqrt{\rho\tau})} \quad \text{(A·2)}$$

ただし $\mathcal{K}_{\gamma}$ は第 2 種変形ベッセル関数である．また，$\mathrm{MGIG}_{\mathbb{C}}$ の確率密度関数は以下の比例式で表される．

$$\mathrm{MGIG}_{\mathbb{C}}(\boldsymbol{X}|\gamma,\boldsymbol{R},\boldsymbol{U})\propto|\boldsymbol{X}|^{\gamma-M}\exp\{-\mathrm{tr}(\boldsymbol{R}\boldsymbol{X}+\boldsymbol{U}\boldsymbol{X}^{-1})\} \quad \text{(A·3)}$$

謝辞 本研究の一部は、JSPS 科研費 24220006, 15K12063 の支援を受けた.

文　　献

[1] A. Hyvärinen, J. Karhunen, and E. Oja, Independent component analysis, John Wiley & Sons, 2004.

[2] I. Lee, T. Kim, and T. Lee, “Fast fixed-point independent vector analysis algorithms for convolutive blind source separation,” Signal Processing, vol.87, no.8, pp.1859–1871, 2007.

[3] A. Hyvarinen, “Fast and robust fixed-point algorithms for independent component analysis,” IEEE TNN, vol.10, no.3, pp.626–634, 1999.

[4] O. Yilmaz and S. Rickard, “Blind separation of speech mixtures via time-frequency masking,” IEEE TSP, vol.52, no.7, pp.1830–1847, 2004.

[5] N. Ito, S. Araki, and T. Nakatani, “Permutation-free convolutive blind source separation via full-band clustering based on frequency-independent source presence priors,” IEEE ICASSP, pp.3238–3242, 2013.

[6] H. Sawada, S. Araki, and S. Makino, “Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment,” IEEE TASLP, vol.19, no.3, pp.516–527, 2011.

[7] L. Drude, C. Boeddeker, and R. Haeb-Umbach, “Blind speech separation based on complex spherical k-mode clustering,” IEEE ICASSP, pp.141–145, 2016.

[8] M.I. Mandel, R.J. Weiss, and D.P.W. Ellis, “Model-based expectation-maximization source separation and localization,” IEEE TASLP, vol.18, no.2, pp.382–394, 2010.

[9] T. Otsuka, K. Ishiguro, H. Sawada, and H.G. Okuno, “Bayesian nonparametrics for microphone array processing,” IEEE TASLP, vol.22, pp.493–504, 2014.

[10] K. Itakura, Y. Bando, E. Nakamura, K. Itoyama, and K. Yoshii, “A unified Bayesian model of time-frequency clustering and low-rank approximation for multi-channel source separation,” EUSIPCO, pp.2280–2284, 2016.

[11] P. Smaragdis and J.C. Brown, “Non-negative matrix factorization for polyphonic music transcription,” IEEE WASPAA, pp.177–180, 2003.

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

[13] D.M. Blei, P.R. Cook, and M.D. Hoffman, “Bayesian nonparametric matrix factorization for recorded music,” ICML, pp.439–446, 2010.

[14] K. Yoshii, K. Itoyama, and M. Goto, “Student’s t nonnegative matrix factorization and positive semidefinite tensor factorization for single-channel audio source separation,” IEEE ICASSP, pp.51–55, 2016.

[15] N. Ono, “Stable and fast update rules for independent vector analysis based on auxiliary function technique,” IEEE WASPAA, pp.189–192, 2011.

[16] S. Araki, F. Nesta, E. Vincent, Z. Koldovskỳ, G. Nolte, A. Ziehe, and A. Benichoux, “The 2011 signal separation evaluation campaign (SiSEC2011):-audio source separation,” Latent Variable Analysis and Signal Separation, pp.414–422, Springer, 2012.

[17] K. Itou, M. Yamamoto, K. Takeda, T. Takezawa, T. Matsuoka, T. Kobayashi, K. Shikano, and S. Itahashi, “The design of the newspaper-based Japanese large vocabulary continuous speech recognition corpus,” ICSLP, pp.3261–3264, 1998.