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.

多チャンネル階乗隠れマルコフモデルによる音響情景分析のための統合的アプローチ

樋口　卓哉† 　亀岡　弘和†,††

† 東京大学大学院情報理工学系研究科
〒 113–8654 東京都文京区本郷 7–3–1
†† 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所
〒 243–0198 神奈川県厚木市森の里若宮 3–1
E-mail: †{higuchi,kameoka}@hil.t.u-tokyo.ac.jp

あらまし　本研究では，音源分離・音響イベント検出・残響除去・音源の到来方向推定という，音響情景分析に関する諸問題を取り扱う．これら音響情景分析に関する問題の根幹にあるのは，ブラインド音源分離の問題である．ブラインド音源分離の問題は不良設定問題であり，一般に音源に関して何らかの仮定を置かない限り，解を限定し解くことができない．本稿では，音響情景分析に関する諸問題が相互依存の関係にあることに着目し，音響情景に関する様々なパラメータによって観測信号を確率的にモデル化することで，ブラインド音源分離の問題における解を限定すると共に，パラメータ推論を通して統合的に音響情景分析を行う手法を提案する．

キーワード　非負値行列因子分解，ブラインド音源分離，残響除去，音響イベント検出，DOA

Unified approach for auditory scene analysis based on multichannel factorial hidden Markov model

Takuya HIGUCHI† and Hirokazu KAMEOKA†,††

† Graduate School of Information Science and Technology, The University of Tokyo
Hongo 7–3–1, Bunkyo-ku, Tokyo, 113–8654 Japan
†† NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation
Morinosatowakamiya 3–1, Atsugi-shi, Kanagawa, 243–0198 Japan
E-mail: †{higuchi,kameoka}@hil.t.u-tokyo.ac.jp

Abstract　This paper deals with the problems of audio source separation, audio event detection, dereverberation and DOA estimation. We focus on the fact that these problems are interdependent, and propose an unified approach for these problem. We model the generative process of observed signals with parameters each of which corresponds to a specific aspect of an audio source. Through the parameter inference, We can simultaneously salve the problems of auditory scene analysis.

Key words　Non-negative matrix factorization, blind source separation, dereverberation, audio event detection, DOA

1. はじめに

本稿では，ブラインド音源分離や残響除去，音響イベント検出，到来方向推定などの音響情景分析の諸問題を取り扱う．音響情景分析技術の応用先として，聴覚障害者へのリアルタイム音響情景提示，音を利用した動画からの特定のコンテンツ抽出，またロボットの音認識システムなどが考えられる．これらを実現するためには音響情景分析に関する諸問題を解く必要があるが，それらの問題の根底にあるのは，ブラインド音源分離の問題である．ブラインド音源分離の問題とは，音源信号や音源からマイクまでの伝達特性が未知の場合に，複数の音源信号が混合された観測信号から元の音源信号を推定する問題である．ブラインド音源分離の問題はいわゆる不良設定問題であり，一般的に，この問題を解くためには，音源信号に対して立てたなんらかの仮定を基に最適化基準を立て，最適化問題を解く必要がある．当然のように，立てた仮定が成り立たない環境下ではそ

の手法はうまく動作しないことが予想され，また立てた仮定が弱すぎる場合でも，解を十分に制限することができず，音源分離を十分に行うことができない．従って，ブラインド音源分離を解くためには，対象とする環境下において成り立つ仮定を置き，さらにその仮定が十分に解を制限することができる必要がある．

例えば，観測信号数が音源数よりも多い優決定問題では，音源信号間の独立性を仮定して分離する独立成分分析 (Independent Component Analysis; ICA) が有用であることが知られており，音源信号間の独立性を最大化するように分離フィルタを推定することが目的となる [1]．しかし，ICA では観測信号数が音源数よりも少ない劣決定問題を扱うことはできず，この場合は独立性よりもさらに強い仮定が必要である．

単チャンネルの観測信号に対するブラインド音源分離の有効なアプローチとして，非負値行列因子分解 (Non-negative Matrix Factorization; NMF) が知られている [2] [3]．この手法では観測信号のパワースペクトログラムを，2 つの非負値行列の積に分解する．分解した各行列は，いくつかの基底パワースペクトルによって構成される基底行列と，それらの基底パワースペクトルの時変な音量を表すアクティベーションによって構成されるアクティベーション行列となる．ここで重要なのは，分解された各基底パワースペクトルが，観測信号の中で主となる要素，すなわち各音源信号を表していると考えられることである．また，音源信号の空間的な情報も利用して音源分離を行うために，NMF を多チャンネルの音響信号へと拡張するアプローチがいくつか取られてきた [4] [5].

しかし NMF(あるいはその多チャンネル観測信号への拡張手法) で立てられていた仮定が，実環境において必ずしも成り立つとは限らない．例えば NMF では，観測信号を構成する限られた数の基底パワースペクトルが，それぞれ各音源信号を表していることを仮定していた．しかし実際の音源信号のパワースペクトルは時変であることが多く，1 つの基底パワースペクトルで表現するのは不十分な場合がある．そこで我々は以前，音源信号の時変なパワースペクトルを隠れマルコフモデルで表現し，多チャンネル NMF のモデルと統合することで，音響イベント検出と音源分離を統合的に行う手法 [6] を提案した．また, [4] [5] では室内インパルス応答長が時間周波数展開における時間窓長に対して十分に短いことを仮定し，瞬時混合近似を用いていたが，残響下ではこの近似は成り立たないことが多い．この問題に対して我々は，観測信号を時間周波数領域における畳み込み混合で表現することで，時間周波数展開におけるフレーム外の残響をモデル化し, [6] の手法と組み合わせることによって，残響除去・音響イベント検出・音源分離を統合的に行う手法を提案した [7].

以上の説明は，従来の NMF において立てられていた仮定のうち，実環境では成り立たなかった仮定 (強すぎた仮定) を，より実環境下の音響信号に則した形に拡張した，という側面での説明であるが，一方でこれらの手法は，音響イベント検出や残響除去の問題という，音源分離の問題と相互依存にある問題を，音源分離の問題と統合的に解くことによって，従来の NMF ベースの手法では利用されなかった，音響イベントや残響成分をブラインド音源分離の問題を解く手がかりとしているという側面も存在する．

例えば，各音源の音響イベント (いつ鳴ったか) が分かっていれば，音源分離の問題は比較的容易になると考えられるが，一方で音源分離が十分うまく行えれば，各音源信号の音響イベントを検出することは比較的容易になると考えられる．すなわち，音響イベント検出の問題と音源分離の問題は相互依存の関係にあるといえる. [6] の手法では，アクティベーションに対して音源の状態 (無音状態，音の立ち上がり，定常状態など) に依存した事前分布を仮定し，音量の大きな値のとりやすさを設定することで，音響イベント検出と音源分離を同時に実現していた．

また前述したように，従来の NMF では残響下では成り立たない瞬時混合近似を用いていたため，残響下において精度良い音源分離を行うためには，事前に残響除去を行う必要がある．すなわち，音源分離性能は残響除去の精度に依存することになる．一方で，実環境においてどのように残響がかかるかは各音源の位置によって異なるため，残響除去の精度は音源分離性能に依存する．すなわち，残響除去の問題と音源分離の問題は相互依存の関係にあるといえる. [7] の手法では，各音源ごとに残響成分を推定しながら音源分離を行うことによって，残響除去と音源分離を統合的に行う手法となっていた．

さらに，音源の到来方向推定の問題も音源分離の問題と相互依存の関係にあるといえる．点音源を仮定すると，一般的に，音源信号の直接波に対応する空間相関行列は，音源の到来方向に応じてある特定の構造を持つことが知られている．この事実に基づき，空間相関行列をモデル化する一つの手法として，音源の取りうるあらゆる到来方向を隠れ変数として，到来方向に基づいて空間相関行列をモデル化する手法が提案されている [8] [9]．また別のモデル化として，音源の取りうるあらゆる到来方向に基づいて陽に記述された空間相関行列の重みつき和として，空間相関行列をモデル化する方法 [10] も提案されている．音源の到来方向に応じて空間相関行列がある特定の構造を持つということは，音源信号の到来方向が分かっていれば，音源分離の手がかりに成りうることを意味しており，また当然ながら，音源分離が十分うまくいけば音源の到来方向推定は比較的容易であると考えられる．すなわち，音源の到来方向推定の問題と音源分離の問題もまた，相互依存の関係にあるのである．

そこで本稿では，多チャンネル NMF [5] における観測信号の生成モデルを拡張し，音響イベント・残響・音源の到来方向を表すパラメータによって観測信号の生成プロセスを確率的に記述することによって，パラメータ推論を通して音響イベント検出・残響除去・音源の到来方向推定・音源分離を統合的に行う手法を提案する．本稿ではこのモデルを多チャンネル階乗隠れマルコフモデルと呼ぶ．

2. 音響イベント検出と音源分離の統合的アプローチ

2.1 多チャンネル階乗隠れマルコフモデル

2.1.1 瞬時混合近似を用いた混合モデル

まず，観測信号の生成プロセスについて述べる．I 個の音源信号が M 個のマイクロフォンで観測される場合を考える．$y_m(\omega_k, t_l) \in \mathbb{C}$ を m 番目のマイクで観測された観測信号の周波数 ω_k，時刻 t_l における時間周波数成分，$s_i(\omega_k, t_l) \in \mathbb{C}$ を i 番目の音源信号の周波数 ω_k，時刻 t_l における時間周波数成分とし，$1 \leqq k \leqq K$ と $1 \leqq l \leqq L$ をそれぞれ時間周波数領域にお

ける周波数と時間のインデックスとする．ここで，室内インパルス応答長が時間周波数展開における時間窓長よりも十分に短い場合を仮定すると，瞬時混合近似を用いて観測信号は以下のように時間周波数領域において記述できる．

$$\boldsymbol{y}(\omega_k, t_l) = \sum_{i=1}^{I} \boldsymbol{a}_i(\omega_k) s_i(\omega_k, t_l), \tag{1}$$

ただし $\boldsymbol{y}(\omega_k, t_l) = (y_1(\omega_k, t_l), \ldots, y_M(\omega_k, t_l))^{\mathsf{T}} \in \mathbb{C}^M$ である．$\boldsymbol{a}_i(\omega_k)$ は i 番目の音源信号に対する周波数 ω_k における伝達周波数特性を表す．表記の都合上，以下では ω_k，t_l を k，l の添え字でそれぞれ表す．

2.1.2 観測信号の生成プロセス

次に，式 (1) に基づいて観測信号の生成プロセスを確率的に記述する．まず，音源信号が区分的に定常であることを仮定し，各時間周波数点で $s_{i,k,l}$ が平均 0，分散 $\sigma^2_{i,k,l}$ の複素正規分布に従うとすると，音源信号の生成プロセスは，

$$s_{i,k,l}|\sigma_{i,k,l} \sim \mathcal{N}_{\mathbb{C}}(s_{i,k,l}; 0, \sigma^2_{i,k,l}), \tag{2}$$

と書き表せる．ここで $\sigma^2_{i,k,l}$ は周波数 k，時刻 l における i 番目の音源のパワースペクトル密度を表す．式 (1) と式 (2) から，$\boldsymbol{a}_{1:I,k}$ と $\sigma_{1:I,k,l}$ が既知の条件下で観測信号 $\boldsymbol{y}_{k,l}$ は同じく複素正規分布に従う．

$$\boldsymbol{y}_{k,l}|\boldsymbol{a}_{1:I,k}, \sigma_{1:I,k,l} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{y}_{k,l}; 0, \sum_i \boldsymbol{C}_{i,k}\sigma^2_{i,k,l}), \tag{3}$$

ただし $\boldsymbol{C}_{i,k} = \boldsymbol{a}_{i,k}\boldsymbol{a}^{\mathsf{H}}_{i,k}$ は i 番目の音源に対する周波数 k における空間相関行列と呼ばれる行列であり，また，$\mathcal{N}_{\mathbb{C}}(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma) \propto \exp(-(\boldsymbol{x}-\boldsymbol{\mu})^{\mathsf{H}}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}))$ である．

2.1.3 多チャンネル NMF [4] [5] における生成モデル

通常の NMF のモデルにおいては ([11] 参照)，音源信号のパワースペクトルはスケールを除いて時不変であることが仮定されていた． 上記のモデルにこの仮定を組み込むと，

$$\sigma^2_{i,k,l} = w_{i,k}h_{i,l}, \tag{4}$$

となる．ここで $\sigma^2_{i,k,l}$ は時不変の基底パワースペクトル $w_{i,k}$ と時変な音量を表す $h_{i,l}$ の積の形で表現されている．これにより $s_{i,k,l}$ の生成モデルは以下のように書き直せる．

$$\boldsymbol{s}_{i,k,l}|w_{i,k}, h_{i,l} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{s}_{i,k,l}; 0, w_{i,k}h_{i,l}). \tag{5}$$

式 (5) による $\boldsymbol{Y} = \{\boldsymbol{y}_{k,l}\}_{k,l}$ の生成モデルは，NMF の多チャンネル観測信号への自然な拡張とみなすことができ，このモデルに基づく BSS のアプローチは多チャンネル NMF [4] [5] と呼ばれている．

2.1.4 音響イベントに基づく音源信号の生成モデル [6]

上記のように，多チャンネル NMF ではそれぞれの音源信号のパワースペクトルはスケールを除いて時不変であることが仮定されていた．この仮定により，解を限定しブラインド音源分離の問題を解く手がかりとしていたのである．しかし多くの場合，音源信号は時変なパワースペクトルを持つため，NMF で仮定されている時不変なパワースペクトルでは，音源信号を十分に表現することができない場合がある．そこで，音響イベントに基づいて音源信号の生成プロセスを記述することで，時変なパワースペクトルを表現すると共に，音響イベントを音源分離の手がかりとすることのできるモデル化を行う．

まず，多くの音源のパワースペクトルは，無音状態，音の立ち上がり，定常状態などその音源の状態 (音響イベント) に応じて異なると考えられるので，時間変化する音源の状態に応じて各音源信号が異なるパワースペクトルを持つと仮定する．時刻 l における i 番目の音源の状態を表す隠れ変数 $z_{i,l} \in \{1, \ldots, Q\}$ を導入し，状態の時系列 $z_{i,1}, \ldots, z_{i,L}$ がマルコフ連鎖に従うと仮定すると，

$$z_{i,l}|z_{i,l-1} \sim \mathrm{Categorical}(z_{i,l}; \boldsymbol{\rho}_{z_{i,l-1}}), \tag{6}$$

と書ける．ここで $\mathrm{Categorical}(x; \boldsymbol{y}) = y_x$ であり，$\boldsymbol{\rho}_q = (\rho_{q,1}, \ldots, \rho_{q,Q})$ は状態 q から各状態 $1, \ldots, Q$ への遷移確率を表し， $\boldsymbol{\rho} = (\rho_{q,q'})_{Q\times Q}$ は遷移行列である．状態 q である i 番目の音源の基底パワースペクトルを $w_{i,k,q}$ と表すとすると，時刻 l における i 番目の音源信号のパワースペクトルは $z_{i,l}$ に依存し，$s_{i,k,l}$ の生成モデルは以下のように書き直せる．

$$s_{i,k,l}|w_{i,k,1:Q}, h_{i,l}, z_{i,l} \sim \mathcal{N}_{\mathbb{C}}(s_{i,k,l}; 0, w_{i,k,z_{i,l}}h_{i,l}). \tag{7}$$

さらに，音源の音量に着目すると，無音状態と有音状態では当然音量の大きな値の取りやすさが異なると考えられるので，音量もまた音源の状態に依存して異なる振る舞いをするといえる．そこで，$h_{i,l}$ が，$z_{i,l}$ によって異なるハイパーパラメータを持つガンマ分布に従うと仮定すると，

$$h_{i,l}|z_{i,l} \sim \mathrm{Gamma}(h_{i,l}; \alpha_{z_{i,l}}, \beta_{z_{i,l}}), \tag{8}$$

となる．ここで $\alpha_{1:Q}$ と $\beta_{1:Q}$ はそれぞれガンマ分布の形状パラメータとスケールパラメータであり，$\mathrm{Gamma}(x; \alpha, \beta) = \frac{x^{\alpha-1}e^{-x/\beta}}{\Gamma(\alpha)\beta^{\alpha}}$ である．$z_{i,l}$ が無音状態に対応するときは $h_{i,l}$ は小さな値をとってほしいので，小さな値をとる確率が高くなるようにガンマ分布のハイパーパラメータを設定し，$z_{i,l}$ が有音状態に対応するときは一様分布に近くなるようにガンマ分布のハイパーパラメータを設定すればよい．具体的には，無音状態に対応する状態では α と β をそれぞれ 1，10^{-2} などと設定し，有音状態のときはそれぞれ 1，10^{20} などと設定すればよい．このような設定により，音響イベント検出と音源分離を協調的に解くことが可能となる．

観測信号の最終的な生成モデルは $\boldsymbol{a}_{1:I,k}$，$w_{1:I,k,1:Q}$， $h_{1:I,l}$，$z_{1:I,l}$ が既知の条件下で，式 (6)，式 (8) と合わせて以下のように書ける．

$$\boldsymbol{y}_{k,l}|\boldsymbol{a}_{1:I,k}, w_{1:I,k,1:Q}, h_{1:I,l}, z_{1:I,l} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{y}_{k,l}; 0, \sum_i \boldsymbol{C}_{i,k}w_{i,k,z_{i,l}}h_{i,l}). \tag{9}$$

観測信号の生成モデルは，各音源信号に関するパラメータ $w_{1:I,1:K,1:Q}$， $h_{1:I,1:L}$ や，各音源信号の音響イベントに関するパラメータ $z_{1:I,1:L}$ などに基づいて記述されているため，この生成モデルに基づいて最適なパラメータを求めることは，音源分離の問題と音響イベント検出の問題を統合的に解くことに相当している．

2.2 パラメータ推論アルゴリズム

2.2.1 目 的 関 数

ここでは，上記の生成モデルに対する，補助関数法に基づくパ

ラメータ推論アルゴリズムについて述べる．モデルにおける推定したい変数は $\boldsymbol{W} = w_{1:I,1:K,1:Q}$, $\boldsymbol{H} = h_{1:I,1:L}$, $\boldsymbol{C} = \boldsymbol{C}_{1:I,1:K}$, $\boldsymbol{Z} = z_{1:I,1:L}$ である．上記の変数の集合を Θ で表す．以下では $\boldsymbol{\rho}$ は実験的に定められた定数とする．我々の目的は以下の式を満たす $\hat{\Theta}$ を求めることである．

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmax}} \log p(\Theta|\boldsymbol{Y}). \tag{10}$$

ここで $\boldsymbol{Y} = \boldsymbol{y}_{1:K,1:L}$ は多チャンネル観測信号の時間周波数成分の集合である．ベイズの定理から，

$$p(\Theta|\boldsymbol{Y}) = \frac{p(\boldsymbol{Y},\Theta)}{p(\boldsymbol{Y})}, \tag{11}$$

であり，また 2.1 節で定義された条件つき確率を用いて，同時確率 $p(\boldsymbol{Y},\Theta)$ は以下のように記述できる．

$$p(\boldsymbol{Y},\Theta) \propto p(\boldsymbol{Y}|\Theta)p(\boldsymbol{H}|\boldsymbol{Z})p(\boldsymbol{Z}). \tag{12}$$

従って，式 (10)，(11)，(12) から，この最適化問題は以下のように書き直せる．

$$\begin{aligned}\hat{\Theta} = \underset{\Theta}{\operatorname{argmax}}\big(&\log p(\boldsymbol{Y}|\Theta) \\ &+ \log p(\boldsymbol{H}|\boldsymbol{Z}) + \log p(\boldsymbol{Z})\big).\end{aligned} \tag{13}$$

ここで，2.1 節で記述した生成モデルから，$\log p(\boldsymbol{Y}|\Theta)$ は以下のように書ける．

$$\begin{aligned}&\log p(\boldsymbol{Y}|\Theta) \\ =&\sum_{k,l}(-\frac{M}{2}\log 2\pi - \frac{1}{2}\log|\hat{\boldsymbol{X}}_{k,l}| - \frac{1}{2}{\boldsymbol{y}_{k,l}}^{\mathsf{H}}\hat{\boldsymbol{X}}_{k,l}^{-1}\boldsymbol{y}_{k,l}).\end{aligned} \tag{14}$$

ただし $\hat{\boldsymbol{X}}_{k,l} = \sum_i \boldsymbol{C}_{i,k} w_{i,k,z_{i,l}} h_{i,l}$ である．

2.2.2 補助関数法に基づく最適化アルゴリズム

式 (13)，(14) をみると，今回の目的関数は各パラメータ同士がお互いに関係しあう複雑な形をしており，一般的に最適化が困難であるが，補助関数法に基づく反復計算によって局所最適となる Θ を求めることができる．補助関数法による，目的関数 $F(\Theta)$ の最大化問題の最適化アルゴリズムでは，まず補助変数 Λ を導入し，$F(\Theta) = \max_\Lambda F^+(\Theta,\Lambda)$ を満たす補助関数 $F^+(\Theta,\Lambda)$ を設計する．そして，$\Theta \leftarrow \operatorname{argmax}_\Theta F^+(\Theta,\Lambda)$ と $\Lambda \leftarrow \operatorname{argmax}_\Lambda F^+(\Theta,\Lambda)$ を交互に繰り返し，補助関数 $F^+(\Theta,\Lambda)$ の最大化を行うことで，間接的に元の目的関数 $F(\Theta)$ を最大化する．ここで重要なのは，Θ について最大化しやすい $F^+(\Theta,\Lambda)$ を設計することである．

今回の最適化問題に補助関数法を適用するため，まず $L(\Theta) = \max_\Lambda L^+(\Theta,\Lambda)$ を満たす補助関数 $L^+(\Theta,\Lambda)$ を設計する．今回の場合は，以下のように補助関数 $L^+(\Theta,\Lambda)$ を設計できる．

$$\begin{aligned}&L(\Theta) \\ \geq&L^+(\Theta,\Lambda) \\ =&-\frac{1}{2}\sum_{k,l}\Bigg(\sum_i\Bigg(\frac{\operatorname{tr}(\boldsymbol{y}_{k,l}{\boldsymbol{y}_{k,l}}^{\mathsf{H}}\boldsymbol{R}_{i,k,l}\boldsymbol{C}_{i,k}^{-1}\boldsymbol{R}_{i,k,l})}{w_{i,k,z_{i,l}}h_{i,l}} \\ &+\operatorname{tr}(\boldsymbol{U}_{k,l}^{-1}\boldsymbol{C}_{i,k})w_{i,k,z_{i,l}}h_{i,l}\Bigg) + \log|\boldsymbol{U}_{k,l}| - M\Bigg) \\ &+\sum_{i,l}\Big((\alpha_{z_{i,l}}-1)\log h_{i,l} - h_{i,l}/\beta_{z_{i,l}} - \alpha_{z_{i,l}}\log\beta_{z_{i,l}}\Big) \\ &+\log p(\boldsymbol{Z}).\end{aligned} \tag{15}$$

ここで $\boldsymbol{R}_{i,k,l}$ と $\boldsymbol{U}_{k,l}$ は $\sum_i \boldsymbol{R}_{i,k,l} = \boldsymbol{I}$ を満たすエルミート正定値行列であり，$\boldsymbol{R}$ と $\boldsymbol{U}$ の集合を Λ で表す．$\operatorname{tr}(\cdot)$ は行列のトレースを表す．式 (15) の等号成立条件は

$$\boldsymbol{R}_{i,k,l} = \boldsymbol{C}_{i,k} w_{i,k,z_{i,l}} h_{i,l} \hat{\boldsymbol{X}}_{k,l}^{-1}, \tag{16}$$

$$\boldsymbol{U}_{k,l} = \hat{\boldsymbol{X}}_{k,l}, \tag{17}$$

となる．

以上から，L は次の 2 つのステップを繰り返すことによって局所最大化できる．

（1） $\boldsymbol{R}$ と $\boldsymbol{U}$ について L^+ を最大化．

（2） $\boldsymbol{W}$，$\boldsymbol{H}$，$\boldsymbol{C}$，$\boldsymbol{Z}$ について L^+ を最大化．

元の目的関数 $L(\Theta)$ では，各パラメータがお互いに関係しあう複雑な形をしていたが，補助関数 $L^+(\Theta,\Lambda)$ は各時間，各周波数，各音源のパラメータごとの和の形で書かれているために，ステップ 2 において並列計算で各パラメータごとに更新を行うことができる．

ステップ 1 における $\boldsymbol{R}$ と $\boldsymbol{U}$ の更新では，式 (16) と式 (17) を使えばよい．ステップ 2 では，L^+ をそれぞれの変数で変微分して 0 となるものを求めることで，更新則が導ける．$\boldsymbol{W}$ と $\boldsymbol{H}$ に関する L^+ の変微分はそれぞれ以下のようになる．

$$\begin{aligned}\frac{\partial L^+}{\partial w_{i,k,z_{i,l}}} = \frac{1}{2}\sum_l\Bigg(&\frac{\operatorname{tr}(\boldsymbol{y}_{k,l}{\boldsymbol{y}_{k,l}}^{\mathsf{H}}\boldsymbol{R}_{i,k,l}\boldsymbol{C}_{i,k}^{-1}\boldsymbol{R}_{i,k,l})}{w_{i,k,z_{i,l}}^2 h_{i,l}} \\ &-\operatorname{tr}(\boldsymbol{U}_{k,l}^{-1}\boldsymbol{C}_{i,k})h_{i,l}\Bigg),\end{aligned} \tag{18}$$

$$\begin{aligned}\frac{\partial L^+}{\partial h_{i,l}} = \frac{1}{2}\sum_k\Bigg(&\frac{\operatorname{tr}(\boldsymbol{y}_{k,l}{\boldsymbol{y}_{k,l}}^{\mathsf{H}}\boldsymbol{R}_{i,k,l}\boldsymbol{C}_{i,k}^{-1}\boldsymbol{R}_{i,k,l})}{w_{i,k,z_{i,l}} h_{i,l}^2} \\ &-\operatorname{tr}(\boldsymbol{U}_{k,l}^{-1}\boldsymbol{C}_{i,k})w_{i,k,z_{i,l}}\Bigg) \\ &+(\alpha_{z_{i,l}}-1)/h_{i,l} - 1/\beta_{z_{i,l}}.\end{aligned} \tag{19}$$

これらを 0 と置くことで，以下の更新式が導ける．

$$w_{i,k,z_{i,l}} \leftarrow \sqrt{\frac{\sum_l \frac{\operatorname{tr}(\boldsymbol{y}_{k,l}{\boldsymbol{y}_{k,l}}^{\mathsf{H}}\boldsymbol{R}_{i,k,l}\boldsymbol{C}_{i,k}^{-1}\boldsymbol{R}_{i,k,l})}{h_{i,l}}}{\sum_l \operatorname{tr}(\boldsymbol{U}_{k,l}^{-1}\boldsymbol{C}_{i,k})h_{i,l}}}, \tag{20}$$

$$h_{i,l} \leftarrow \frac{(\alpha_{z_{i,l}}-1) + \sqrt{(\alpha_{z_{i,l}}-1)^2 + \mu_{i,l}\nu_{i,l}}}{\nu_{i,l}}, \tag{21}$$

ただし

$$\mu_{i,l} = \sum_k \frac{\operatorname{tr}(\boldsymbol{y}_{k,l}{\boldsymbol{y}_{k,l}}^{\mathsf{H}}\boldsymbol{R}_{i,k,l}\boldsymbol{C}_{i,k}^{-1}\boldsymbol{R}_{i,k,l})}{w_{i,k,z_{i,l}}}, \tag{22}$$

$$\nu_{i,l} = \sum_k \operatorname{tr}(\boldsymbol{U}_{k,l}^{-1}\boldsymbol{C}_{i,k})w_{i,k,z_{i,l}} + 2/\beta_{z_{i,l}}, \tag{23}$$

である．$\boldsymbol{C}$ についての L^+ の変微分は以下のようになる．

$$\begin{aligned}\frac{\partial L^+}{\partial \boldsymbol{C}_{i,k}} = \sum_l\Bigg(&\frac{\boldsymbol{C}_{i,k}^{-1}\boldsymbol{R}_{i,k,l}\boldsymbol{y}_{k,l}{\boldsymbol{y}_{k,l}}^{\mathsf{H}}\boldsymbol{R}_{i,k,l}\boldsymbol{C}_{i,k}^{-1}}{w_{i,k,z_{i,l}}h_{i,l}} \\ &-\boldsymbol{U}_{k,l}^{-1}w_{i,k,z_{i,l}}h_{i,l}\Bigg).\end{aligned} \tag{24}$$

これを 0 と置くと，以下の Riccati 方程式が得られる．

$$\boldsymbol{C}_{i,k}\boldsymbol{A}_{i,k}\boldsymbol{C}_{i,k} = \boldsymbol{B}_{i,k}, \tag{25}$$

ただし

$$\boldsymbol{A}_{i,k} = \sum_l w_{i,k,z_{i,l}} h_{i,l} \hat{\boldsymbol{X}}_{k,l}^{-1}, \tag{26}$$

$$\boldsymbol{B}_{i,k} = \boldsymbol{C}_{i,k}\Big(\sum_l w_{i,k,z_{i,l}} h_{i,l} \hat{\boldsymbol{X}}_{k,l}^{-1} \boldsymbol{y}_{k,l}\boldsymbol{y}_{k,l}^{\mathsf{H}} \hat{\boldsymbol{X}}_{k,l}^{-1}\Big)\boldsymbol{C}_{i,k}, \tag{27}$$

である．以下の方法でこの Riccati 方程式を解くことで $\boldsymbol{C}$ の更新則が得られる [5]．まず，以下の $2M \times 2M$ の行列に対して固有値分解を行う．

$$\begin{bmatrix} 0 & -\boldsymbol{A}_{i,k} \\ -\boldsymbol{B}_{i,k} & 0 \end{bmatrix}. \tag{28}$$

ここで $\boldsymbol{e}_{1,i,k} \ldots \boldsymbol{e}_{M,i,k}$ を負の固有値に対応する固有ベクトルだとし，$2M$ 次元の固有ベクトルを $m = 1 \ldots M$ において以下のように分解する．

$$\boldsymbol{e}_{m,i,k} = \begin{bmatrix} \boldsymbol{f}_{m,i,k} \\ \boldsymbol{g}_{m,i,k\cdot} \end{bmatrix} \tag{29}$$

ここで $\boldsymbol{f}_{m,i,k}$ と $\boldsymbol{g}_{m,i,k}$ は M 次元のベクトルである．$\boldsymbol{C}_{i,k}$ の更新則は以下のように得られる．

$$\boldsymbol{C}_{i,k} \leftarrow \boldsymbol{G}_{i,k}\boldsymbol{F}_{i,k}^{-1}. \tag{30}$$

ただし $\boldsymbol{F}_{i,k} = [\boldsymbol{f}_{1,i,k}, \ldots, \boldsymbol{f}_{M,i,k}]$，$\boldsymbol{G}_{i,k} = [\boldsymbol{g}_{1,i,k}, \ldots, \boldsymbol{g}_{M,i,k}]$ である．

L^+ は各音源のパラメータごとの和の形で書かれているために，L^+ を $\boldsymbol{Z}$ の関数と見ると，各音源に対応する隠れマルコフモデルの対数事後確率の和とみなすことができる．従って最適な状態の時系列 $z_{i,1}, \ldots, z_{i,L}$ を動的計画法によって効率的に，各音源ごとに個別に求めることができる．

以上の更新則から $\boldsymbol{W}$，$\boldsymbol{H}$，$\boldsymbol{C}$，$\boldsymbol{Z}$ を反復計算により求めることは，ブラインド音源分離の問題と音響イベント検出の問題を画一的最適化規準に基づき協調的に解いていることに相当している．

2.3 動作実験例

以上の提案法により劣決定音源分離実験を行った結果，多チャンネル NMF [5] と比較して，高い音源分離性能を示した．また，音響イベント検出もある程度の精度で行うことができた．実験の詳細は [6] を参照してほしい．

3. 残響除去との統合的アプローチ

3.1 畳み込み混合による観測信号の混合モデル

式 (1) による混合モデルでは，瞬時混合近似を仮定していた．しかし残響がある場合には一般に，室内インパルス応答長は時間窓長に対して十分に短いとはいえず，瞬時混合近似は成り立たない．そこで，時間周波数領域における畳み込み混合の形で観測信号を近似する．

$$\boldsymbol{y}(\omega_k, t_l) \approx \sum_{i=1}^{I}\sum_{n=0}^{N} \boldsymbol{a}_i(\omega_k, t_n) s_i(\omega_k, t_l - t_n). \tag{31}$$

ここで $\boldsymbol{a}_i(\omega_k, n)$ は i 番目の音源信号に対する周波数 ω_k における伝達周波数特性の時刻 t_n の成分であり，$0 \leqq n \leqq N$ は伝達周波数特性の時間周波数領域における時間インデックスである．ここで $\boldsymbol{a}_i(\omega_k, t_n)$ は i 番目の音源信号が時間周波数領域において t_n だけ先の時刻にどれだけ影響を与えるか，すなわち時間周波数領域においてどれだけ残響がかかるかを表している．観測信号の最終的な生成モデルは $\boldsymbol{a}_{1:I,k,0:N}$，$w_{1:I,k,1:Q}$，$h_{1:I,l-N:l}$，$z_{1:I,l-N:l}$ が既知の条件下で，式 (6)，式 (8) と合わせて以下のように書き直せる．

$$\boldsymbol{y}_{k,l} | \boldsymbol{a}_{1:I,k,0:N}, w_{1:I,k,1:Q}, h_{1:I,l-N:l}, z_{1:I,l-N:l} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{y}_{k,l}; 0, \sum_{i,n} \boldsymbol{C}_{i,k,n} w_{i,k,z_{i,l-n}} h_{i,l-n}). \tag{32}$$

この生成モデルに基づいて最適なパラメータを求めることは，残響除去・音響イベント検出・音源分離の問題を統合的に解くことに相当している [7]．

3.2 パラメータ推論アルゴリズム

3.2.1 目的関数

式 (32) の生成モデルにより，2.2.1 節の目的関数における式 (14) の $\hat{\boldsymbol{X}}_{k,l}$ を，$\hat{\boldsymbol{X}}_{k,l} = \sum_{i,n} \boldsymbol{C}_{i,k,n} w_{i,k,z_{i,l-n}} h_{i,l-n}$ と置き換えたものが，今回のモデルにおける目的関数に相当する．これにより，$L(\Theta)$ は，各時刻のパラメータ同士が互いに関係しあう形となってしまうが，この場合も補助関数法を適用し，各時刻ごとのパラメータの和の形で表された補助関数を設計することで，効率的な最適化アルゴリズムを導出できる．

3.2.2 補助関数法に基づく最適化アルゴリズム

この場合の補助関数は，以下のように設計できる．

$$\begin{aligned}
&L(\Theta) \\
\geq& L^+(\Theta, \Lambda) \\
=& -\frac{1}{2}\sum_{k,l}\Bigg(\sum_{i,n}\Bigg(\frac{\mathrm{tr}(\boldsymbol{y}_{k,l}\boldsymbol{y}_{k,l}^{\mathsf{H}}\boldsymbol{R}_{i,k,l,n}\boldsymbol{C}_{i,k,n}^{-1}\boldsymbol{R}_{i,k,l,n})}{w_{i,k,z_{i,l-n}} h_{i,l-n}} \\
&+ \mathrm{tr}(\boldsymbol{U}_{k,l}^{-1}\boldsymbol{C}_{i,k,n}) w_{i,k,z_{i,l-n}} h_{i,l-n}\Bigg) + \log|\boldsymbol{U}_{k,l}| - M\Bigg) \\
&+ \sum_{i,l}\Big((\alpha_{z_{i,l}} - 1)\log h_{i,l} - h_{i,l}/\beta_{z_{i,l}} - \alpha_{z_{i,l}}\log\beta_{z_{i,l}}\Big) \\
&+ \log p(\boldsymbol{Z}).
\end{aligned} \tag{33}$$

ここで $\boldsymbol{R}_{i,k,l,n}$ と $\boldsymbol{U}_{k,l}$ は $\sum_{i,n}\boldsymbol{R}_{i,k,l,n} = \boldsymbol{I}$ を満たすエルミート正定値行列であり，$\boldsymbol{R}$ と $\boldsymbol{U}$ の集合を Λ で表す．式 (33) の等号成立条件は

$$\boldsymbol{R}_{i,k,l,n} = \boldsymbol{C}_{i,k,n} w_{i,k,z_{i,l-n}} h_{i,l-n} \hat{\boldsymbol{X}}_{k,l}^{-1}, \tag{34}$$

$$\boldsymbol{U}_{k,l} = \hat{\boldsymbol{X}}_{k,l}, \tag{35}$$

となる．観測信号の生成プロセスを畳み込み混合の形で表現していたがゆえに，元の目的関数は各時刻のパラメータが他の時刻のパラメータと関係しあう複雑な形をしていたが，設計した補助関数は，各時刻ごとのパラメータの和の形で表現されていることが分かる．これにより，各時刻ごとのパラメータを並列計算によって求めることができると共に，音響イベントを表す隠れ変数 $z_{i,1:L}$ の時系列を，動的計画法により効率的に求めることが可能となる．

紙面の都合上詳細は省略するが，2.2.2 節の最適化アルゴリズムと同様，各パラメータの更新則は以下の形で求まる．

$$w_{i,k,z_{i,l}} \leftarrow \sqrt{\frac{\sum_{l,n} \frac{\mathrm{tr}(\boldsymbol{y}_{k,l+n}\boldsymbol{y}_{k,l+n}{}^{\mathsf{H}}\boldsymbol{R}_{i,k,l+n,n}\boldsymbol{C}_{i,k,n}^{-1}\boldsymbol{R}_{i,k,l+n,n})}{h_{i,l}}}{\sum_{l,n}\mathrm{tr}(\boldsymbol{U}_{k,l+n}^{-1}\boldsymbol{C}_{i,k,n})h_{i,l}}}, \quad (36)$$

$$h_{i,l} \leftarrow \frac{(\alpha_{z_{i,l}}-1)+\sqrt{(\alpha_{z_{i,l}}-1)^2+\mu_{i,l}\nu_{i,l}}}{\nu_{i,l}}, \quad (37)$$

ただし

$$\mu_{i,l} = \sum_{k,n} \frac{\mathrm{tr}(\boldsymbol{y}_{k,l+n}\boldsymbol{y}_{k,l+n}{}^{\mathsf{H}}\boldsymbol{R}_{i,k,l+n,n}\boldsymbol{C}_{i,k,n}^{-1}\boldsymbol{R}_{i,k,l+n,n})}{w_{i,k,z_{i,l}}}, \quad (38)$$

$$\nu_{i,l} = \sum_{k,n} \mathrm{tr}(\boldsymbol{U}_{k,l+n}^{-1}\boldsymbol{C}_{i,k,n})w_{i,k,z_{i,l}} + 2/\beta_{z_{i,l}}, \quad (39)$$

である．$\boldsymbol{C}_{i,k,n}$ の更新については，以下の Riccati 方程式を章と同様の手順で解けばよい．

$$\boldsymbol{C}_{i,k,n}\boldsymbol{A}_{i,k,n}\boldsymbol{C}_{i,k,n} = \boldsymbol{B}_{i,k,n}, \quad (40)$$

ただし

$$\boldsymbol{A}_{i,k,n} = \sum_{l} w_{i,k,z_{i,l-n}} h_{i,l-n} \hat{\boldsymbol{X}}_{k,l}^{-1}, \quad (41)$$

$$\boldsymbol{B}_{i,k,n} = \boldsymbol{C}_{i,k,n}\Big(\sum_{l} w_{i,k,z_{i,l-n}} h_{i,l-n} \hat{\boldsymbol{X}}_{k,l}^{-1}\boldsymbol{y}_{k,l}\boldsymbol{y}_{k,l}{}^{\mathsf{H}}\hat{\boldsymbol{X}}_{k,l}^{-1}\Big)\boldsymbol{C}_{i,k,n}, \quad (42)$$

である．$\boldsymbol{Z}$ の更新については，前述したように，L^+ を各音源，各時刻ごとのパラメータの和の形になるように設計したことによって，最適な音源の状態 $z_{i,1},\ldots,z_{i,L}$ を動的計画法によって効率的に，各音源ごとに個別に求めることができる．

3.3 動作実験例

以上の提案法により残響下で録音された観測信号に対して，教師あり音源分離・残響除去・音響イベント検出を試みた結果，2 章で説明した [6] の手法と比較して，高い音源分離性能を示した．図 1 に無響下で録音された音源信号のスペクトログラム (上)，2 章の手法 [6] によって得られた分離音のスペクトログラム (中上)，本章の提案法によって得られた残響除去済み分離音のスペクトログラム (中下)，提案法による音響イベント検出結果 (下) を示す．実験の詳細は [7] を参照してほしい．

4. 音源の到来方向推定との統合的アプローチ

4.1 音源の到来方向に基づく空間相関行列のモデル化

上記のモデルでは，空間相関行列に対して何の仮定も置かれておらず，自由度の高いモデルとなっていたがゆえに，推定した空間相関行列が望ましくない局所解に陥ってしまうことがあった．しかし一般的に，音源信号の直接波に対応する実際の空間相関行列は，点音源を仮定すると，音源の到来方向に応じてある特定の構造を持つことが知られている．従って，音源の到来方向に基づいて空間相関行列をモデル化することが可能である．マイクロフォンの数 $M=2$ の場合では，方向 $\theta(0 \leqq \theta \leqq \pi)$ にある音源の空間相関行列は，以下のように陽に記述可能である．

$$\boldsymbol{J}(\theta,\omega) = \begin{bmatrix} 1 \\ e^{\mathrm{J}\omega B\cos\theta/C} \end{bmatrix}\begin{bmatrix} 1 & e^{\mathrm{J}\omega B\cos\theta/C} \end{bmatrix}^* \quad (43)$$

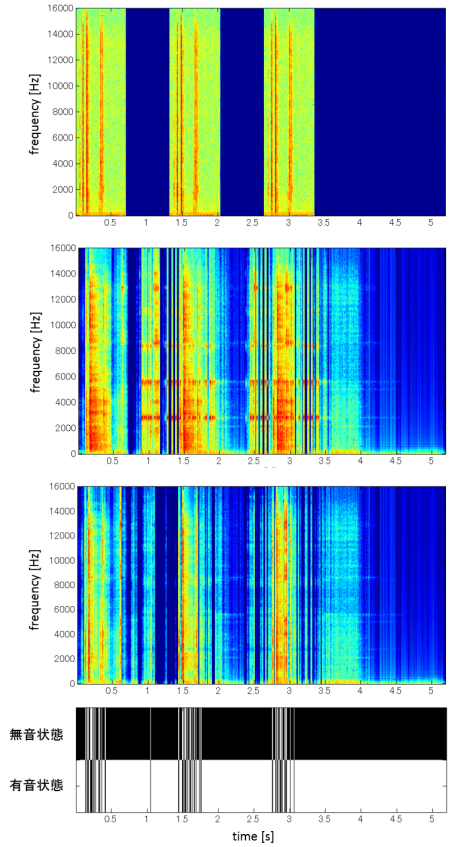

図 1　無響下で録音された音源信号のスペクトログラム (上)，2 章の手法によって得られた分離音のスペクトログラム (中上)，本章の提案法によって得られた残響除去済み分離音のスペクトログラム (中下)，提案法による音響イベント検出結果 (下)．音響イベント検出結果は，黒がその時刻に推定された状態を表す．

ここで J は虚数単位，B [m] はマイクロフォン間の距離，C [m/s] は音速である．i 番目の音源の到来方向 θ_i が既知の場合では，直接波に対応する空間相関行列は $\boldsymbol{J}(\theta_i,\omega_k)$ と等しくなることが期待される．しかし実際には，時間周波数展開におけるフレーム内に含まれる残響成分によって，$\boldsymbol{C}_{i,k,0}$ は $\boldsymbol{J}(\theta_i,\omega_k)$ と等しくならないことが予想されるばかりでなく，音響信号から直接音源の到来方向を観測することはできない．そこで，[10] の手法と同様に，まず到来方向を O 個の到来方向 $\vartheta_1,\ldots,\vartheta_O$ に離散化し，さらに重み定数 $d_{i,1}\ldots d_{i,O}$ を導入することで，空間相関行列を以下のようにモデル化する．

$$\boldsymbol{C}_{i,k,0} = \sum_{o} d_{i,o}\boldsymbol{J}(\vartheta_o,\omega_k). \quad (44)$$

ここで $d_{i,1}\ldots d_{i,O}$ は非負値の値であり，$\sum_o d_{i,o}=1$ を満たす．このモデル化により，実際の伝達特性にある程度即した形で，空間相関行列の自由度を制限することができる．さらに，$d_{i,1}\ldots d_{i,O}$ を推定することによって，音源の到来方向推定を同時に行うことが可能となる．なぜなら，比較的大きな値として推定された重み定数に対応した到来方向が，音源の到来方向であると期待されるからである．

観測信号の最終的な生成モデルは $\boldsymbol{a}_{1:I,k,0:N}$，$w_{1:I,k,1:Q}$，$h_{1:I,l-N:l}$，$z_{1:I,l-N:l}$，$d_{1:I,1:O}$ が既知の条件下で，式 (6)，式 (8) と合わせて以下のように書き直せる．

$$\boldsymbol{y}_{k,l}|\boldsymbol{a}_{1:I,k,0:N}, w_{1:I,k,1:Q}, h_{1:I,l-N:l}, z_{1:I,l-N:l}, d_{1:I,1:O} \sim \mathcal{N}_{\mathbb{C}}(\boldsymbol{y}_{k,l}; 0, \sum_{i,o} d_{i,o}\boldsymbol{J}(\vartheta_o, \omega_k) w_{i,k,z_{i,l}} h_{i,l} + \sum_{i,n\neq 0} \boldsymbol{C}_{i,k,n} w_{i,k,z_{i,l-n}} h_{i,l-n}). \quad (45)$$

この生成モデルに基づいて最適なパラメータを求めることは，音源の到来方向推定・残響除去・音響イベント検出・音源分離の問題を統合的に解くことに相当している．

4.2 パラメータ推論アルゴリズム

4.2.1 目的関数

3.2.1 節の目的関数において，$\boldsymbol{C}_{i,k,0}$ を $\sum_o d_{i,o}\boldsymbol{J}(\vartheta_o, \omega_k)$ と置き換え，また音源の状態数 Q を増やした場合に，i 番目の音源の音響イベントを表す $z_{i,1:L}$ が望ましくない局所解に落ちてしまうことを防ぐため，目的関数を $\boldsymbol{Z}$ について周辺化することを考える．今回の目的関数は，

$$\begin{aligned} L(\Theta) &= \log \sum_{\boldsymbol{Z}} p(\boldsymbol{Y}|\boldsymbol{W}, \boldsymbol{H}, \boldsymbol{C}, \boldsymbol{D}, \boldsymbol{Z}) \\ &\quad + \log \sum_{\boldsymbol{Z}} p(\boldsymbol{H}|\boldsymbol{Z}) + \log \sum_{\boldsymbol{Z}} p(\boldsymbol{Z}) \\ &= \sum_{k,l} (-\frac{1}{2} \log \sum_{\boldsymbol{Z}} |\hat{\boldsymbol{X}}_{k,l}| \\ &\quad - \frac{1}{2} \log \sum_{\boldsymbol{Z}} \exp \boldsymbol{y}_{k,l}{}^{\mathsf{H}} \hat{\boldsymbol{X}}_{k,l}^{-1} \boldsymbol{y}_{k,l}) \\ &\quad + \log \sum_{\boldsymbol{Z}} p(\boldsymbol{H}|\boldsymbol{Z}) + \log \sum_{\boldsymbol{Z}} p(\boldsymbol{Z}), \end{aligned} \quad (46)$$

となる．ここで $\hat{\boldsymbol{X}}_{k,l} = \sum_i \left(\sum_o d_{i,o}\boldsymbol{J}(\vartheta_o, \omega_k) w_{i,k,z_{i,l}} h_{i,l} + \sum_{n\neq 0} \boldsymbol{C}_{i,k,n} w_{i,k,z_{i,l-n}} h_{i,l-n} \right)$ であり，今回は $\boldsymbol{C}_{1:I,1:K,0}$ の代わりに $d_{1:I,1:O}$ を推定する必要がある．目的関数をみると，ここでもまた，各到来方向に対応する重み定数 $d_{1:I,1:O}$ が互いに関係しあう形となっているが，これも補助関数法を用いることにより，各音源ごと，各到来方向ごとの和の形になるように補助関数を設計し，効率的な最適化アルゴリズムを導出することができる．

4.2.2 補助関数法に基づく最適化アルゴリズム

2, 3 章の最適化アルゴリズムでは，i 番目の音源の音響イベントを表す $z_{i,1:L}$ の最適な時系列をビタビアルゴリズムにより求めていたが，本章では各時刻でどの状態にあるらしいかを求めるアルゴリズムを導出する．この場合の補助関数は，以下のように設計できる．

$$\begin{aligned} & L(\boldsymbol{\Theta}) \\ \geqq & L^+(\boldsymbol{\Theta}, \Lambda) \\ = & -\frac{1}{2} \sum_{k,l} \Bigg(\sum_{i,q,o} \lambda_{q,i,l} \Bigg(\frac{\mathrm{tr}(\boldsymbol{y}_{k,l}\boldsymbol{y}_{k,l}{}^{\mathsf{H}} \boldsymbol{R}_{i,k,l,0,q} \boldsymbol{J}_{i,k,o}^{-1} \boldsymbol{R}_{i,k,l,0,q})}{d_{i,o} w_{i,k,q_{i,l}} h_{i,l}} \\ & + \mathrm{tr}(\boldsymbol{U}_{k,l}^{-1} \boldsymbol{J}_{i,o}) d_{i,o} w_{i,k,q_{i,l}} h_{i,l} \Bigg) \\ & + \sum_{i,q,n\neq 0} \lambda_{q,i,l} \Bigg(\frac{\mathrm{tr}(\boldsymbol{y}_{k,l}\boldsymbol{y}_{k,l}{}^{\mathsf{H}} \boldsymbol{R}_{i,k,l,n,q} \boldsymbol{C}_{i,k,n}^{-1} \boldsymbol{R}_{i,k,l,n,q})}{w_{i,k,q_{i,l-n}} h_{i,l-n}} \\ & + \mathrm{tr}(\boldsymbol{U}_{k,l}^{-1} \boldsymbol{C}_{i,k,n}) w_{i,k,q_{i,l-n}} h_{i,l-n} \Bigg) + \log |\boldsymbol{U}_{k,l}| - M \Bigg) \\ & + \sum_{i,l,q} \lambda_{q,i,l} \Big((\alpha_{q_{i,l}} - 1) \log h_{i,l} - h_{i,l}/\beta_{q_{i,l}} - \alpha_{q_{i,l}} \log \beta_{q_{i,l}} \Big) \\ & + \sum_q \lambda_{q,i,l} \log p(\boldsymbol{Z}), \end{aligned} \quad (47)$$

ここで $\boldsymbol{R}_{i,k,l,n,q}$，$\boldsymbol{U}_{k,l}$ は $\sum_{i,n} \boldsymbol{R}_{i,k,l,n} = \boldsymbol{I}$ を満たすエルミート正定値行列であり，$\lambda_{q,i,l}$ は $\sum_q \lambda_{q,i,l} = 1$ を満たす非負値のスカラー値である．式 (47) の等号成立条件は，$\boldsymbol{R}_{i,k,l,n,q}$，$\boldsymbol{U}_{k,l}$ については

$$\boldsymbol{R}_{i,k,l,n,q} = \boldsymbol{C}_{i,k,n} w_{i,k,z_{i,l-n}=q} h_{i,l-n} \hat{\boldsymbol{X}}_{k,l}^{-1}, \quad (48)$$

$$\boldsymbol{U}_{k,l} = \hat{\boldsymbol{X}}_{k,l}. \quad (49)$$

である．さらに $\lambda_{q,i,l}$ に関しては Forward-Backward アルゴリズムを用いて，

$$\lambda_{q,i,l} = F_{q,i,l} B_{q,i,l} / \sum_q F_{q,i,l} B_{q,i,l}, \quad (50)$$

$$F_{q,i,l} = p(\Theta|z_{i,l} = q) \sum_{q'} F_{q,i,l-1} \rho_{z_{i,l-1}=q', z_{i,l}=q}, \quad (51)$$

$$B_{q,i,l} = \sum_{q'} B_{q',i,l+1} p(\Theta|z_{i,l+1} = q') \rho_{z_{i,l}=q, z_{i,l+1}=q'}, \quad (52)$$

ただし

$$\begin{aligned} & p(\Theta|z_{i,l} = q) \\ & \propto \exp\Bigg(-\frac{1}{2} \sum_{k,l} \Bigg(\sum_{i,o} \Bigg(\frac{\mathrm{tr}(\boldsymbol{y}_{k,l}\boldsymbol{y}_{k,l}{}^{\mathsf{H}} \boldsymbol{R}_{i,k,l,0,q} \boldsymbol{J}_{i,k,o}^{-1} \boldsymbol{R}_{i,k,l,0,q})}{d_{i,o} w_{i,k,z_{i,l}=q} h_{i,l}} \\ & + \mathrm{tr}(\boldsymbol{U}_{k,l}^{-1} \boldsymbol{J}_{i,k,o}) d_{i,o} w_{i,k,z_{i,l}=q} h_{i,l} \Bigg) \\ & + \sum_{i,n\neq 0} \Bigg(\frac{\mathrm{tr}(\boldsymbol{y}_{k,l+n}\boldsymbol{y}_{k,l+n}{}^{\mathsf{H}} \boldsymbol{R}_{i,k,l+n,n,q} \boldsymbol{C}_{i,k,n}^{-1} \boldsymbol{R}_{i,k,l+n,n,q})}{w_{i,k,z_{i,l}=q} h_{i,l}} \\ & + \mathrm{tr}(\boldsymbol{U}_{k,l+n}^{-1} \boldsymbol{C}_{i,k,n}) w_{i,k,z_{i,l}=q} h_{i,l} \Bigg) \Bigg) \\ & + \sum_{i,l} \Big((\alpha_{z_{i,l}=q} - 1) \log h_{i,l} - h_{i,l}/\beta_{z_{i,l}=q} - \alpha_{z_{i,l}=q} \log \beta_{z_{i,l}=q} \Big) \Bigg). \end{aligned} \quad (53)$$

である．$d_{i,o}$ の更新則は，2.2.2 節の最適化アルゴリズムと同様の手順により，以下の形で求まる．

$$d_{i,o} \leftarrow \sqrt{\frac{\sum_{k,l,q} \lambda_{q,i,l} \frac{\mathrm{tr}(\boldsymbol{y}_{k,l}\boldsymbol{y}_{k,l}{}^{\mathsf{H}} \boldsymbol{R}_{i,k,l,0} \boldsymbol{J}_{i,k,o}^{-1} \boldsymbol{R}_{i,k,l,0})}{w_{i,k,z_{i,l}=q} h_{i,l}}}{\sum_{k,l,q} \lambda_{q,i,l} \mathrm{tr}(\boldsymbol{U}_{k,l}^{-1} \boldsymbol{J}_{i,k,o}) w_{i,k,z_{i,l}=q} h_{i,l}}}. \quad (54)$$

4.3 評価実験

提案法の音源分離・到来方向推定性能の評価のために，実験を行った．ATR 音声データベース [12] の中の音声 (男性 1 人，女性 2 人) に RWCP データベース [13] のインパルス応答 (残響時間 380 ms，マイク間距離 11.48 cm，マイクの数 $M = 2$) を畳み込み，人工的に残響下での多チャンネルの混合信号を作成した．音源の到来方向はそれぞれ $\pi/6$，$\pi/2$，$13\pi/18$ [rad] である．音声を代えて 10 個の混合信号を作成し，実験に用いた．サンプリング周波数は 16 kHz とした．フレーム長 64 ms，フレームシフト長 16 ms で STFT を行い，時間周波数展開を行った．HMM の状態数 Q は 5 とした．α_1 と β_1 を 1，10^{-1} とそれぞれ設定し，$\alpha_{2:10}$ と $\beta_{2:10}$ を 1 と 10^{10} と設定することで，$d = 1$ を無音状態とみなした．$\boldsymbol{D}$ の初期値は図 2(上) のよ

表 1　分離処理前の混合音の SDR/SIR の平均値と，提案法と [7] の手法によって得られた分離音の SDR/SIR の平均値.

	SDR [dB]	SIR [dB]
提案法	-4.17	5.90
[7] の手法	-6.49	1.94
分離処理前	-40.33	-4.22

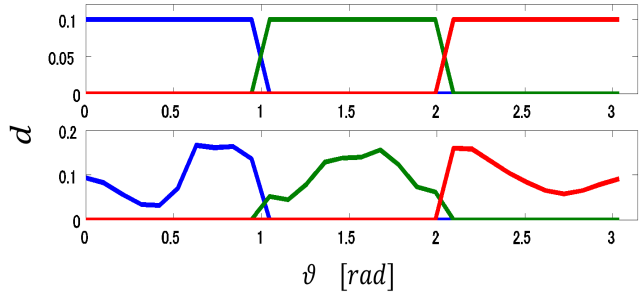

図 2　音源ごとに色分けされた，$\boldsymbol{D}$ の初期値 (上) と提案法による $\boldsymbol{D}$ の推定結果例 (下).

うに設定した．$\boldsymbol{C}_{1:I,1:K,1:N}$ の初期値については，対角成分を $10^{-1}/\sqrt{M}$，それ以外の成分を 0 とした．$\boldsymbol{W}$ の初期値は乱数で与え，$\boldsymbol{H}$ の初期値は一様とし，$\lambda_{q,i,l}$ の初期値は $1/Q$ とした．パラメータ推定アルゴリズムはまず $N=0$ としてその後徐々に N を $N=3$ となるまで増やしながら全体として 30 回反復した．比較対象には [7] の手法を用いた．分離音 $\hat{\boldsymbol{y}}_{i,k,l}$ はウィーナーフィルタ

$$\hat{\boldsymbol{y}}_{i,k,l} = w_{i,k,z_{i,l}} h_{i,l} \boldsymbol{C}_{i,k,0} \hat{\boldsymbol{X}}_{k,l}^{-1} \boldsymbol{y}_{k,l}. \qquad (55)$$

によって得た．客観評価基準として，signal-to-distortion /interference ratio (SDR/SIR) [14] を用いた．高い SDR/SIR は高い音源分離性能を表す.

表 1 に提案法と従来法で得られた SDR/SIR の平均値を示す．提案法によって得られた分離音の SDR と SIR は従来法を上回った．図 2(下) に $\boldsymbol{D}$ の推定結果例を示す．音源の到来方向がおおむね推定できていることが分かる.

5. おわりに

本稿では，音響情景分析に関する様々な諸問題を取り扱った．それらの諸問題が相互依存の関係にあることに着目し，音響情景を表す様々なパラメータによって観測信号の生成プロセスを確率的に記述することで，パラメータ推論を通して統合的な音響情景分析を行うアプローチを提案した．さらに，本提案モデルにおける対数事後確率は，様々なパラメータが互いに関係しあう複雑な形をしているが，補助関数法を用いることによって，目的関数を最適化する効率的な反復アルゴリズムを導出した．本稿は [6] と [7] の内容をまとめた上で，到来方向推定を同時に行うために，[7] の手法を拡張したものである.

6. 謝　　辞

本研究は JSPS 科研費 26730100 の助成を受けたものです.

文　　献

[1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.

[2] D. D. Lee, and H. S. Seung, “Learning the parts of objects with nonnegative matrix factorization,” *Nature*, vol. 401, pp.788–791, 1999.

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

[4] A. Ozerov, and C. Févotte, “Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation,” *IEEE Trans. Audio, Speech and Language Processing*, vol. 18, no. 3, pp. 550-563, Mar.2010.

[5] H. Sawada, H. Kameoka, S. Araki and N. Ueda, “Efficient algorithms for multichannel extensions of Itakura-Saito nonnegative matrix factorization,” *ICASSP 2012*, pp. 261–264, 2012.

[6] T. Higuchi, H. Takeda, T. Nakamura and H. Kameoka, “A unified approach for underdetermined blind signal separation and source activity detection by multichannel factorial hidden Markov models,” *Interspeech 2014*, pp. 850–854, 2014.

[7] T. Higuchi and H. Kameoka, “Joint audio source separation and dereverberation based on multichannel factorial hidden Markov model,” *MLSP 2014*.

[8] H. Kameoka, M. Sato, T. Ono, N. Ono, S. Sagayama, “Blind separation of infinitely many sparse sources,” *IWAENC 2012*, H-09, Sep. 2012.

[9] T. Higuchi, N. Takamune, T. Nakamura, H. Kameoka, ”Underdetermined blind separation and tracking of moving sources based on DOA-HMM,” *ICASSP 2014*, pp. 3215-3219, May 2014.

[10] J. Nikunen and T. Virtanen, “Multichannel audio separation by direction of arrival based spatial covariance model and non-negative matrix factorization,” *ICASSP 2014*, pp. 6727–6731, May 2014.

[11] C. Févotte, N. Bertin, and J. -L. Durrieu, “Nonnegative matrix factorization with the itakura-saito divergence. with application to music analysis,” *Neural Computation*, vol. 21, no. 3, 2009.

[12] A. Kurematsu, K. Takeda, Y. Sagisaka, S. Katagiri, H. Kuwabara, and K. Shikano,, K. Takeda, Y. Sagisaka, S. Katagiri, H. Kuwabara, and K. Shikano, “ATR Japanese speech database as a tool of speech recognition and synthesis,” *Speech Communication*, pp. 357–363, 1990.

[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,” *LREC 2000*, pp. 965–968, 2000.

[14] E. Vincent, R. Gribonval, and C. Févotte, “Performance measurement in blind audio source separation,” *IEEE Transactions on Audio, Speech, and Language Processing*, pp. 1462–1469, 2006.