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］音楽音響信号解析のための ガンマ過程に基づく無限重畳離散全極モデル

吉井　和佳†　　糸山　克寿†　　後藤　真孝††

† 京都大学 大学院情報学研究科　〒 606–8501 京都府京都市左京区吉田本町
†† 産業技術総合研究所 情報技術研究部門　〒 305–8568 茨城県つくば市梅園 1–1–1
E-mail: †{yoshii,itoyama}@kuis.kyoto-u.ac.jp, ††m.goto@aist.go.jp

**あらまし**　本稿では，多数の楽器音が重畳している音楽音響信号を，音の三要素である音高（基本周波数）・音色（スペクトル包絡）・音量に分解するための確率モデルについて述べる．楽器音がソース・フィルタ理論に従うとすると，そのフーリエスペクトルは音源信号に対応するスペクトル微細構造と音色を表すスペクトル包絡との積で表現できる．この仮定のもとで，短時間フーリエ変換 (STFT) で得られる混合音スペクトログラムを音の三要素に分解するには，**複合自己回帰モデル**を用いることができる．しかし，音高を持つ楽器音に対してスペクトル包絡を推定する際に，調波構造のピークのみに着目するのではなく，全周波数帯域が等しく考慮に入れられていた．また，人間の聴覚特性に則した対数周波数スペクトログラムを扱うことができなかった．これらの問題を解決するため，本研究では，**離散全極モデル**を複合自己回帰モデルの枠組みに組み入れることで，連続ウェーブレット変換で得られる混合音スペクトログラムを分解できる**無限重畳離散全極モデル**を提案する．本モデルはガンマ過程を用いたノンパラメトリックベイズモデルであり，観測データに合わせて適切な個数のスペクトル包絡と調波構造スペクトルを推定できる．
**キーワード**　音楽音響信号解析，非負値行列分解，音源分離，ガンマ過程，ノンパラメトリックベイズ

# [Poster Presentation] An Infinite Superimposed Discrete All-pole Model based on Gamma Processes for Music Signal Analysis

Kazuyoshi YOSHII†, Katsutoshi ITOYAMA†, and Masataka GOTO††

† Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan
†† National Institute of Advanced Industrial Science and Technology (AIST)
1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan
E-mail: †{yoshii,itoyama}@kuis.kyoto-u.ac.jp, ††m.goto@aist.go.jp

**Abstract**　This paper presents a probabilistic model that is used for decomposing music audio signals into the three elements of sound: pitches (fundamental frequencies: F0s), timbres (spectral envelopes), and volumes. Conventional composite autoregressive models cannot deal with log-frequency spectrograms that match the characteristics of human auditory perception. When pithed sounds are analyzed, it is inappropriate to focus on all the frequency bands not limited to harmonic partials for estimating the spectral envelopes. To solve these problems, we propose a nonparametric Bayesian model based on gamma processes called an infinite superimposed discrete all-pole model by incorporating the idea of discrete all-pole modeling into a composite autoregressive model.
**Key words**　Music signal analysis, NMF, source separation, gamma process, nonparametric Bayes

## 1. はじめに

音楽情報処理分野において，機械学習技術に基づく音楽音響信号の統計的モデリングは，ホットなトピックのひとつである．特に，多重音に対する基本周波数推定や音源分離における有用性から，非負値行列分解 (Nonnegative Matrix Factorization: NMF) は大きな注目を集めている [1–13]．標準的な NMF では，多重音の振幅あるいはパワースペクトログラム（非負値行列）を二つの非負値行列，すなわち，周波数方向の基底スペクトルの集合と各基底スペクトルに対応する時間方向の音量変化の集合とに分解することができる．このとき，観測スペクトログラムと再構成スペクトログラムとの近似誤差を表すコスト関

数を最小化するため，効率的な乗法更新アルゴリズムが利用できる [14]．このアルゴリズムは，ある特定の確率モデルの最尤推定と等価であることが分かっている [15]．

近年，音声信号の生成過程を説明する目的で考案されたソース・フィルタ理論を楽器音の統計的モデリングに援用することがしばしば行われている [2–4]．周波数領域において，楽器音の音高と音色はそれぞれ，音源信号の性質を表す微細構造（調波構造）と楽器個体の共鳴特性を表すスペクトル包絡によってよく特徴づけられる．人間の聴覚系はスペクトルのピーク（フォルマント）に対して敏感であるため，楽器音の各時間フレームにおけるスペクトル包絡を，全極型周波数伝達関数（自己回帰フィルタの周波数応答）を用いて表現することが一般的である [2]．全極型スペクトル包絡推定の古典的な方法である線形予測分析 (Linear Predictive Coding: LPC) [16] は，音源信号がガウス性白色雑音であるという強い仮定のもとで，観測音声信号のスペクトル包絡を推定する手法である．これは，ある特定の確率モデルの最尤推定に対応している．

LPC の問題点を解決する有望な統計的アプローチとして，複合自己回帰モデル (Composite Autoregressive Model: CAR) [5] と呼ばれるソース・フィルタ NMF が提案されている．本モデルでは，観測音響信号のスペクトログラムは，複数個の微細構造（ソース）と複数個のスペクトル包絡（フィルタ）との組み合わせから構成されているとみなす．このアプローチの重要な特徴は，音源信号がガウス性白色雑音であるとは仮定されておらず，全極型スペクトル包絡推定と同時に音源信号のスペクトル自体が推定される点である．本モデルは確率的な解釈が可能であるため，ノンパラメトリックベイズ理論を用いて，適切な個数のソースとフィルタを推定できる無限複合自己回帰モデルに拡張できる [6]．また，音源信号のスペクトルが調波構造をもつように制約を加えることも可能であり，音源信号の基本周波数 (F0) を最尤推定することができる [6]．

複合自己回帰モデルを含む従来のソース・フィルタ NMF の主要な問題点として，以下の二点が挙げられる．

- 調波構造のピークのみがスペクトル包絡からの信頼できるサンプルとみなせるにもかかわらず，調波構造に対してスペクトル包絡を推定する際に，すべての周波数帯域が等しく考慮に入れられていた．
- ウェーブレット変換や定 Q 変換の方が人間の聴覚特性に則しているにもかかわらず，短時間フーリエ変換 (STFT) を用いるのが一般的であった．フーリエ変換で得られる線形周波数領域では，ピアノ音やギター音に見られるような，高次倍音の非調波性（F0 の整数倍からのずれ）が，等間隔に配置された倍音成分からなる理想的な調波構造モデルと実際の観測スペクトルとの致命的な不整合を起こしていた．

これらの問題を解決するため，本研究では，無限重畳離散全極モデル (Infinite Superimposed Discrete All-pole Model: iSDAP) とよぶ多重音解析のための新しいソース・フィルタ NMF を提案する（図 1）．本モデルは，スペクトル包絡が調波構造のピークで不要に鋭いピークを持つことを防ぐため，調波構造の離散的なピークのみを用いてスペクトル包絡を推定できる離散全極モデル (Discrete All-pole Model: DAP) [17] に着想を得ている．しかし，多重音に対して DAP を適用するには，個々の調波構造を分離し，F0（スペクトル包絡推定に利用する離散的な倍音周波数）を推定する必要があった．本研究の重要な貢献の一つは，無限複合自己回帰モデルと同様に，重畳された複数の調波構造を扱えるように DAP を拡張したことである．提案モデルでは，離散的な周波数を取り扱えるという DAP の副次的な効果として,「対数周波数」スペクトログラムを適切な個数の音高（F0）・音色（スペクトル包絡）・音量に分解することができる．

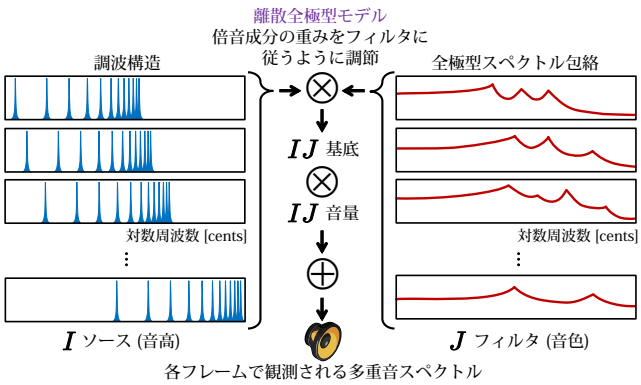


図 1　無限重畳離散全極モデル (iSDAP)：ソース数 $I$ とフィルタ数 $J$ がいずれも無限に発散した場合の極限を考える．

## 2. 関 連 研 究

本章では，音響信号に対するソース・フィルタモデルと NMF への応用について紹介する．従来のモデルは一般に，STFT で得られる線形周波数領域で定式化されている．次章で提案するモデルを定式化する際の基礎となるため，これらのモデルの確率的解釈について説明する．

### 2.1 線形予測分析（全極モデル）

線形予測分析 (LPC) は観測スペクトルに対してスペクトル包絡を推定するための音響信号モデリング手法である．LPC では，与えられたスペクトルに対応する観測音響信号 $\boldsymbol{x} = \{x_m\}_{m=1}^{M}$（$M$ は窓幅）が，$P$ 次の自己回帰過程

$$x_m = -\sum_{p=1}^{P} a_p x_{m-p} + s_m \quad i.e., \quad \sum_{p=0}^{P} a_p x_{m-p} = s_m \qquad (1)$$

に従うことを仮定している．ここで，$\boldsymbol{a} = [a_0, \cdots, a_P]^T$ は自己回帰フィルタの係数（ただし $a_0 = 1$）であり，$\boldsymbol{s} = \{s_m\}_{m=1}^{M}$ は線形予測誤差である．式 (1) はソース・フィルタモデルの観点から説明することができる．すなわち，$\boldsymbol{x}$ が音声信号であるとすると，$\boldsymbol{s}$ は声帯（ソース）から生成される音源信号に対応し，$\boldsymbol{a}$ は声道（フィルタ）の反響特性を表している．

式 (1) は，$\boldsymbol{s}$ を入力にとり，$\boldsymbol{x}$ を出力する線形系とみなすことができ，その振る舞いはパラメータ $\boldsymbol{a}$ で決定される．式 (1) は $\boldsymbol{a}$ と $\boldsymbol{x}$ との畳み込みであることから

$$A(z)X(z) = S(z) \quad i.e., \quad X(z) = S(z)F(z) \qquad (2)$$

が成立する．ここで，$X(z)$ および $S(z)$ はそれぞれ，$\boldsymbol{x}$ および

$\boldsymbol{s}$ の $z$ 変換であり，

$$X(z) = \sum_{m=1}^{M} x_m z^{-m} \qquad S(z) = \sum_{m=1}^{M} s_m z^{-m} \tag{3}$$

で与えられる．ここで，$F(z) \stackrel{\text{def}}{=} \frac{1}{A(z)}$ は全極モデルであり，次式で与えられる．

$$F(z) = \frac{1}{A(z)} = \frac{1}{\sum_{p=0}^{P} a_p z^{-p}} \tag{4}$$

いま，$2\pi\frac{m}{M} = \omega_m$ とし，$z = e^{i\omega_m}$ を式 (2) に代入すると，この線形系の伝達特性のフーリエ領域表現を得る．

$$X(e^{i\omega_m}) = S(e^{i\omega_m})F(e^{i\omega_m}) \tag{5}$$

ここで，$\{X(e^{i\omega_m})\}_{m=1}^{M}$ は観測信号 $\boldsymbol{x}$ の複素スペクトルであり，$\{S(e^{i\omega_m})\}_{m=1}^{M}$ は音源信号 $\boldsymbol{s}$ の複素スペクトル，$\{F(e^{i\omega_m})\}_{m=1}^{M}$ は全極型伝達関数の周波数応答である．

LPC では，音源信号 $\boldsymbol{s}$ がガウス性白色雑音であるという強い仮定をおくことで，フィルタ係数 $\boldsymbol{a}$ を推定する．まず，$S(e^{i\omega_m})$ が，すべての周波数帯域で同じ分散を持つ独立同分布な複素ガウス分布に従うことを仮定する．

$$S(e^{i\omega_m}) \sim \mathcal{N}_c(0, \sigma^2) \tag{6}$$

ここで，$\sigma^2$ は音源信号 $\boldsymbol{s}$ の白色スペクトルのパワーを表す．式 (5) および式 (6) を用いると，

$$X(e^{i\omega_m}) \sim \mathcal{N}_c(0, \sigma^2|F(e^{i\omega_m})|^2) \tag{7}$$

を得る．ここで，$X_m = |X(e^{i\omega_m})|^2$ および $F_m = |F(e^{i\omega_m})|^2$ と定義すると，式 (7) は簡潔に

$$X_m \sim \text{Exponential}(\sigma^2 F_m) \tag{8}$$

と書ける．ここで，図 2 で示される通り，$\{X_m\}_{m=1}^{M}$ は観測信号 $\boldsymbol{x}$ のパワースペクトル，$\{F_m\}_{m=1}^{M}$ は $\{X_m\}_{m=1}^{M}$ のスペクトル包絡である．式 (8) は LPC の確率モデルであるので，$\{F_m\}_{m=1}^{M}$（すなわち $\boldsymbol{a}$）および $\sigma^2$ は最尤推定を用いて求めることができる [16].

LPC の主な問題のひとつは，音源信号の周期性（例：バイオリンの弦振動）に起因する音高をもつ音響信号を解析すると，観測スペクトル $\{X_m\}_{m=1}^{M}$ に含まれる調波構造に影響を受け，推定されるスペクトル包絡 $\{F_m\}_{m=1}^{M}$ は倍音周波数において不要に急峻なピークをもつことである．この理由は，実際には調波構造を構成する倍音のみがスペクトル包絡からの信頼できるサンプルとみなせるにもかかわらず，$M$ 個すべての周波数帯域が全極モデルの推定に用いられるからである．

### 2.2 離散全極モデル

離散全極モデル (DAP) は，LPC のもつ問題を解決するために提案されたスペクトル包絡推定手法のひとつである．DAP は LPC の拡張であるとみなすことができるため，その確率モデルは式 (8) と同じ形をとるが，一部の周波数ビン $\Omega$ 上にのみ定義されている点が異なる．

$$X_m \sim \text{Exponential}(\sigma^2 F_m) \quad m \in \Omega \tag{9}$$

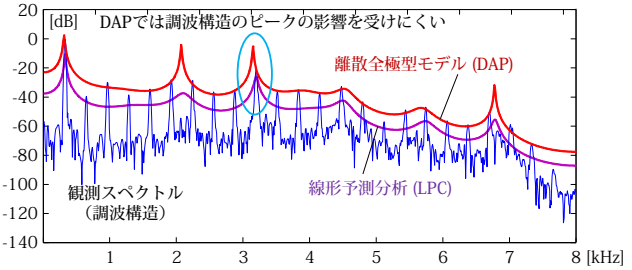


図 2　調波構造をもつ観測スペクトルに対して線形予測分析 (LPC) および離散全極モデル (DAP) で推定されたスペクトル包絡

ここで，$\Omega = \{1, \cdots, M\}$ とすると，DAP は LPC と等価になる．調波構造に対してスペクトル包絡を推定するためには，離散的な倍音周波数の集合 $\Omega$ を定義すればよい．こうすることで，図 2 に示すように，調波構造のピークの近くを正確に通り，必ずしも倍音周波数ではない位置に極を持つスペクトル包絡を推定することができる．式 (9) で示される尤度を最大化するため，$\boldsymbol{a}$ および $\sigma^2$ を交互に反復最適化する手法が提案されている [17]．この手法は，乗法更新アルゴリズムの一種とみなすことができる [2, 9].

最尤推定における式 (9) の最大化は，次式で与えられる IS ダイバージェンスの最小化と等価である．

$$\mathcal{D}_{\text{IS}}(\boldsymbol{X}|\sigma^2\boldsymbol{F}) = \sum_{m\in\Omega}\left(\frac{X_m}{\sigma^2 F_m} - \log\frac{X_m}{\sigma^2 F_m} - 1\right) \tag{10}$$

ここで，$F_m$ は次式で与えられる．

$$F_m = \frac{1}{\left|\sum_{p=0}^{P} a_p e^{-i\omega_m p}\right|^2} = \frac{1}{\boldsymbol{a}^T\boldsymbol{U}_m\boldsymbol{a}} \tag{11}$$

$\boldsymbol{U}_m$ は $(P+1)\times(P+1)$ のテプリッツ行列であり，各要素は $[\boldsymbol{U}_m]_{pq} = \cos(\omega_m(p-q))$ で与えられる．式 (10) を $\sigma^2$ に関して偏微分してゼロとおくことで，更新則

$$\sigma^2 = \frac{1}{|\Omega|}\sum_{m\in\Omega}\frac{X_m}{F_m} \tag{12}$$

を得る．ここで，$|\Omega|$ は $\Omega$ に含まれる周波数の個数を表す．一方，式 (10) を $\boldsymbol{a}$ に関して偏微分すると，二項の差の形

$$\frac{\partial\mathcal{D}_{\text{IS}}(\boldsymbol{X}|\sigma^2\boldsymbol{F})}{\partial\boldsymbol{a}} = 2(\boldsymbol{R} - \boldsymbol{R}')\boldsymbol{a} \tag{13}$$

を得る．ここで，$\boldsymbol{R}$ および $\boldsymbol{R}'$ は正定値行列であり，

$$\boldsymbol{R} = \frac{1}{\sigma^2}\sum_{m\in\Omega} X_m\boldsymbol{U}_m \qquad \boldsymbol{R}' = \sum_{m\in\Omega} F_m\boldsymbol{U}_m \tag{14}$$

で与えられる．ベクトル $\boldsymbol{a}$ に関する乗法更新則は

$$\boldsymbol{a} \leftarrow \boldsymbol{R}^{-1}\boldsymbol{R}'\boldsymbol{a} \tag{15}$$

で与えられる．$\boldsymbol{R}$ は正定値行列であることから常に逆行列をもち，式 (15) は安定的に動作する．式 (10) が収束するまで式 (12) および式 (15) を交互に反復することでパラメータの最適化を行う．ただし，$\sigma^2$ を調節することで，$a_0 = 1$ を満たすようスペクトル包絡 $\{F_m\}_{m\in\Omega}$ をそのつど正規化しておく．

調波構造に対する DAP の主な問題のひとつは，$\Omega$ を定義するうえで，観測スペクトルの F0 をあらかじめ与える必要があることである．したがって，複数の調波構造が重畳した多重音を解析する場合には，個々の調波構造を分離し，F0 を推定しておく必要がある．一方，調波構造を分離するうえで，スペクトル包絡は倍音の相対強度比に関して重要な手掛かりを与える．このような鶏と卵の問題は容易に解くことはできない．

### 2.3 複合自己回帰モデル

複合自己回帰モデルは，図 3 に示す通り，多重音のスペクトログラムを $I$ 個の微細構造（ソース）と $J$ 個のスペクトル包絡（フィルタ）とに分解することができるソース・フィルタ NMF の一種である [5]．いま，観測パワースペクトログラムを $\boldsymbol{X} \in \mathbb{R}^{M\times N}$ する．ここで，$M$ は周波数ビンの数，$N$ はフレーム数である．非負値行列 $\boldsymbol{X}$ を，三つの因子 $\boldsymbol{S}$，$\boldsymbol{F}$，$\boldsymbol{H}$ に分解することを考える．

$$X_{mn} \approx \sum_{i=1}^{I}\sum_{j=1}^{J} S_{im}F_{jm}H_{nij} \stackrel{\text{def}}{=} Y_{mn} \tag{16}$$

ここで，$\{S_{im}\}_{m=1}^{M}$ はソース $i$ のパワースペクトル，$\{F_{jm}\}_{m=1}^{M}$ はフィルタ $j$ のパワースペクトル，$H_{nij}$ はフレーム $n$ におけるソース $i$・フィルタ $j$ の組み合わせの音量を表す．これらすべての変数は，$\boldsymbol{X}$ から推定することができる．

#### 2.3.1 基本的な定式化

CAR の確率モデルを定式化するためには，音源信号のスペクトルが従う確率分布を仮定する必要がある．LPC では式 (6) で示すように音源信号がガウス性白色雑音であると仮定していたのに対し，CAR では各周波数ビンごとに異なるパラメータを持つ独立な確率分布を許容する．

$$S_i(e^{i\omega_m}) \sim \mathcal{N}_c\left(0, S_{im}\right) \tag{17}$$

ここで，$\{S_i(e^{i\omega_m})\}_{m=1}^{M}$ はソース $i$ の複素スペクトルである．式 (5) および式 (17) を用いると，

$$X_{ijmn}(e^{i\omega_m}) \sim \mathcal{N}_c\left(0, S_{im}F_{jm}H_{nij}\right) \tag{18}$$

を得る．ここで，$\{X_{ijmn}(e^{i\omega_m})\}_{m=1}^{M}$ はフレーム $n$ におけるソース $i$・フィルタ $j$ の組み合わせに起因する複素スペクトルである．ガウス分布の再生成を考慮すると，

$$X_{mn}(e^{i\omega_m}) \sim \mathcal{N}_c\left(0, Y_{mn}\right) \tag{19}$$

を得る．$\{X_{mn}(e^{i\omega_m})\}_{m=1}^{M}$ はフレーム $n$ で観測される複素スペクトルである．式 (19) は次式と等価である．

$$X_{mn} \sim \text{Exponential}\left(Y_{mn}\right) \tag{20}$$

ここで，$\mathbb{E}[X_{mn}] = Y_{mn}$ が成立しており，$\{X_{mn}\}_{m=1}^{M}$ および $\{Y_{mn}\}_{m=1}^{M}$ はフレーム $n$ における観測および再構成パワースペクトルである．これは，式 (16) において $X_{mn}$ と $Y_{mn}$ の近似誤差を評価するコスト関数として，理論的には板倉-斎藤 (IS) ダイバージェンスが適切であることを示している [5]．ただし，IS ダイバージェンスは $Y_{mn}$ に対して凸関数ではないため，$Y_{mn}$ の最適化は局所解に陥りやすい傾向がある．

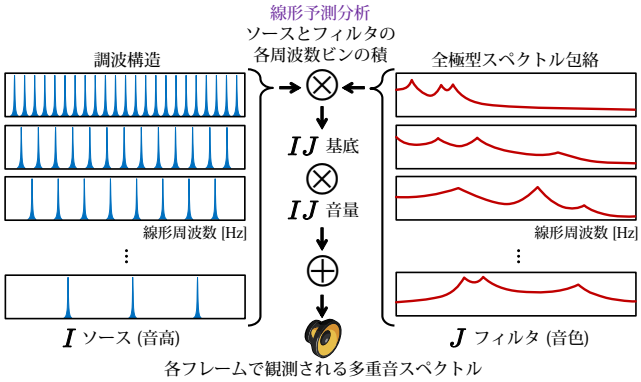


図 3　線形周波数領域における複合自己回帰モデル (CAR)

#### 2.3.2 確率モデルの拡張

$X_{mn}$ と $Y_{mn}$ との近似誤差を評価するコスト関数として，IS ダイバージェンスの代わりにカルバック・ライブラ (KL) ダイバージェンスを用いると異なる確率モデルが得られる [6]．すなわち，式 (20) の代わりに次式を仮定する．

$$X_{mn} \sim \text{Poisson}\left(Y_{mn}\right) \tag{21}$$

ここで，$\mathbb{E}[X_{mn}] = Y_{mn}$ が成立し，$\{X_{mn}\}_{m=1}^{M}$ および $\{Y_{mn}\}_{m=1}^{M}$ はフレーム $n$ における観測および再構成振幅スペクトルである．ただし，KL ダイバージェンスを用いる場合には，「パワー」ではなく「振幅」スペクトログラムに対して定式化することが一般的である [14, 18]．

CAR の拡張として，ノンパラメトリックベイズ理論を用いることで，観測データ $\boldsymbol{X}$ に合わせて適切な個数のソースとフィルタを推定することができる無限複合自己回帰モデル (iCAR) が提案されている [6]．

$$X_{mn} \approx \sum_{i=1}^{I\to\infty}\sum_{j=1}^{J\to\infty} \theta_i\phi_j S_{im}F_{jm}H_{nij} \tag{22}$$

ここで，$\theta_i$ および $\phi_j$ はそれぞれソース $i$ およびフィルタ $j$ の大域的な重みを表す．ガンマ過程 NMF [10] と同様に，$\boldsymbol{\theta}$ および $\boldsymbol{\phi}$ に対してガンマ過程事前分布を仮定することにより，理論的には無限個存在するソース・フィルタのうちで，観測データに合わせて高々有限個だけが実質的にアクティベートされる機構を実現できる．具体的には，$I$ および $J$ を十分に大きい値に設定し（大きければ大きいほど有限打ち切りによる近似が正確になる），そのほとんどの要素がゼロとなるようなスパースが学習を行う．

CAR の別の拡張として，ソーススペクトル $\{S_{im}\}_{m=1}^{M}$ が調波構造をもつような制約を導入する手法も提案されている [6, 19]．もし，音源信号が周期的インパルスの系列であるとすると（声道の理想的なモデル），$\{S_{im}\}_{m=1}^{M}$ は同じ強度で線形周波数軸に等間隔に並んだ倍音成分からなる調波構造をもつことになる．F0 は，式 (21) で定義される尤度を最大化するように最尤推定で求めることができる．このモデルにおいては，スペクトル包絡推定と同時に F0 推定を行うことができるため，DAP の問題点を解決することができる可能性を秘めている．

## 3. 無限重畳離散全極モデル

本章では，連続ウェーブレット変換で得られる対数周波数スペクトログラムに対してソース・フィルタ分解を行うための無限重畳離散全極モデル (iSDAP) を提案する．本モデルでは，与えられた音楽音響信号に対し，各フレームに含まれる複数の F0（調波構造）を推定すると同時に，複数のスペクトル包絡（楽器の音色）を発見することができる．これを実現するため，スペクトル包絡推定のための離散全極モデル (DAP) [17] を，F0 とスペクトル包絡の同時推定のための複合自己回帰モデル (CAR) [5, 6] の枠組みに確率的に統合する．これによる本研究の主な貢献は以下の通りである．

（1） **重畳離散全極モデル**：多重音に含まれる複数の調波構造それぞれに対してスペクトル包絡を推定することができる．通常の DAP では，単独音のスペクトルに対するスペクトル包絡推定を目的としていた．

（2） **基本周波数モデリング**：各フレームごとに異なる F0（ソーススペクトル）の集合が含まれることを許容することにより，ビブラートやポルタメントなどの基本周波数の微細変動を精緻にモデル化することができる．通常の CAR においては，ソーススペクトルの集合がすべてのフレームで共有されているため，各ソーススペクトルは半音レベルの F0 に対応していることが期待されていた．

（3） **対数周波数領域での定式化**：本モデルは，人間の聴覚特性にあった対数周波数スペクトログラムを分解することができる初めてのソース・フィルタ NMF である．これを実現するには，スペクトル包絡推定の際に離散的な周波数のみを考慮することができる DAP が不可欠である．これまで，対数周波数領域における単独発話スペクトルの解析に DAP が利用された例 [20] はあったが，混合音への適用は初めてである．

本研究では，観測スペクトログラムに合わせてソースとフィルタの個数を自動調節するため，ノンパラメトリックベイズモデルを定式化する．本モデルは，観測データが無限にあれば，理論的には無限個のソースとフィルタが存在することを仮定している．一方，現実的には有限の観測データが与えられると，そこに含まれる高々有限個のソースとフィルタを推定する必要がある．ノンパラメトリックベイズ理論を用いると無限次元の空間内でスパースな学習が可能になる．

### 3.1 確率モデルの定式化

本節では iSDAP の確率モデルの定式化を行う．いま，対数周波数領域における振幅スペクトログラムを $\boldsymbol{X} \in \mathbb{R}^{M \times N}$ とする．ここで，$M$ は周波数ビンの個数であり，$N$ はフレームの個数である．非負値行列 $\boldsymbol{X}$ を，二つの因子 $\boldsymbol{W}$ および $\boldsymbol{H}$ に分解することを考える．

$$X_{mn} \sim \text{Poisson}\left(\sum_{i=1}^{I\to\infty}\sum_{j=1}^{J\to\infty} \theta_{ni}\phi_j W_{nijm} H_{nij}\right) \tag{23}$$

ここで，$\theta_{ni}$ はフレーム $n$ におけるソース $i$ の局所的な重み，$\phi_j$ は全フレームにおけるフィルタ $j$ の大域的な重みを表す．$H_{nij}$ はフレーム $n$ におけるソース $i$・フィルタ $j$ の組み合わせの音量

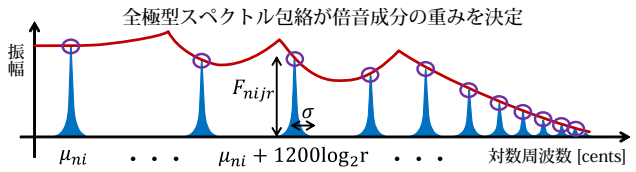


図 4　対数周波数領域におけるフレーム $n$ におけるソース $i$ とフィルタ $j$ との組み合わせによる楽器音スペクトルの生成

を表す．$\{W_{nijm}\}_{m=1}^{M}$ は，フレーム $n$ におけるソース $i$・フィルタ $j$ から生成されたある楽器音の振幅スペクトルである．もし，$\theta_{ni}$ および $W_{nijm}$ が時間変化しないとすると，すなわち，ソーススペクトルの集合がすべてのフレームで共有されているとすると，式 (16) で示される無限複合自己回帰モデルと類似した形式をとる．ただし，周波数軸は異なることに注意する．

#### 3.1.1 ソース・フィルタの組み合わせ

図 4 で示す通り，振幅スペクトル $\{W_{nijm}\}_{m=1}^{M}$ は対数周波数領域において調波構造を持つことを仮定する．

$$W_{nijm} = \sum_{r=1}^{R} F_{nijr} S_{mnir} \tag{24}$$

ここで，$R$ は倍音の個数であり，$\{S_{mnir}\}_{m=1}^{M}$ はフレーム $n$ におけるソース $i$ の第 $r$ 次倍音成分の単峰的なスペクトルであり，次式で表すものとする．

$$S_{mnir} = \exp\left(-\frac{1}{2\sigma^2}\left(f_m - (\mu_{ni} + 1200\log_2 r)\right)^2\right) \tag{25}$$

ここで，$\mu_{ni}$ はフレーム $n$ におけるソース $i$ の対数基本周波数 F0 [cents]，$f_m$ はスペクトログラムにおける $m$ 番目の周波数ビンに対応する対数周波数，$\sigma^2$ は各倍音を中心としたメインローブの広がりである．

ここで，倍音成分の重み $\{F_{nijr}\}_{r=1}^{R}$ は，対数周波数領域において全極型伝達関数を用いて表現する．

$$F_{nijr} = \frac{1}{\left|\sum_{p=0}^{P} a_{jp} e^{-\omega_{nir} p i}\right|} = \left(\boldsymbol{a}_j^T \boldsymbol{U}(\omega_{nir}) \boldsymbol{a}_j\right)^{-\frac{1}{2}} \tag{26}$$

ここで，$\boldsymbol{a}_j \equiv [a_{j0}, \cdots, a_{jP}]^T$ であり，$\omega_{nir}$ はフレーム $n$ におけるソース $i$ の第 $r$ 次倍音に対応する正規化角周波数 [rad] であり，$\boldsymbol{U}(\omega)$ は $(P+1)\times(P+1)$ の行列であり，各要素は $[\boldsymbol{U}(\omega)]_{pq} = \cos(\omega(p-q))$ で与えられる．$F_{nijr}$ はパワーではなく，振幅を表すことに注意する．

重畳離散全極モデル (SDAP) の重要な特徴は，ソースとフィルタを組み合わせる際に，ソースの倍音成分の重みのみが全極型スペクトル包絡によって制御される点である．一方，複合自己回帰モデル (CAR) においては，式 (16) で示す通り，全周波数帯域にわたって周波数ビンごとにソースとフィルタとの要素積を計算する．したがって，スペクトル包絡を推定する際に，SDAP では倍音成分（調波構造のピーク）のみが参照されるのに対して，CAR では調波構造スペクトル全体が参照される．これは，SDAP が離散全極モデル (DAP) の多重音拡張であるとみなせるのに対して，CAR は線形予測分析 (LPC) の多重音拡張とみなせることを意味する．

### 3.1.2 事前分布の設計

無限次元のベクトル $\boldsymbol{\theta}_n = [\theta_{n1}, \cdots, \theta_{nI}]^T$ および $\boldsymbol{\phi} = [\phi_1, \cdots, \phi_J]^T$ に対してスパースな学習を行うため，ガンマ過程事前分布を仮定する [6,10]．これを近似的に実現するひとつの方法として，$\boldsymbol{\theta}_n$ および $\boldsymbol{\phi}$ の各要素に対して独立なガンマ事前分布を仮定する．

$$\theta_{ni} \sim \mathrm{Gamma}\left(\frac{\alpha_\theta}{I}, \alpha_\theta\right) \tag{27}$$

$$\phi_j \sim \mathrm{Gamma}\left(\frac{\alpha_\phi}{J}, \alpha_\phi\right) \tag{28}$$

ここで，$\alpha_\theta$ および $\alpha_\phi$ は超パラメータであり，ガンマ過程の集中度と呼ばれる．打ち切りレベル $I$ を無限に大きくしていけば，ベクトル $\boldsymbol{\theta}_n$ は集中度 $\alpha_\theta$ をもつガンマ過程からのランダムサンプルとみなすことができる．ここで，任意の正の実数 $\epsilon$ に対して $\theta_{ni} > \epsilon$ を満たす実効的な要素数 $I^+$ はほとんど確実に有限であることが証明されている．現実的には，$I$ を $\alpha_\theta$ に比べて十分大きく設定すれば，$\boldsymbol{\theta}_n$ の $I$ 個の要素のいくつかだけがゼロよりある程度大きな値をとることが期待できる．一方，音量 $H_{nij}$ に関しては事後分布計算の容易さから，ガンマ事前分布を仮定する．

$$H_{nij} \sim \mathrm{Gamma}(a_H, b_H) \tag{29}$$

ここで，$a_H$ および $b_H$ は超パラメータである．

## 3.2 確率モデルの変分推論

我々の目標は，ベイズの定理に従って確率変数の事後分布 $p(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{H}|\boldsymbol{X}; \boldsymbol{\mu}, \boldsymbol{a}) = \frac{p(\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{H}; \boldsymbol{\mu}, \boldsymbol{a}) p(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{H})}{p(\boldsymbol{X}; \boldsymbol{\mu}, \boldsymbol{a})}$ を求め，パラメータ $\boldsymbol{\mu}$ および $\boldsymbol{a}$ については周辺尤度 $p(\boldsymbol{X}; \boldsymbol{\mu}, \boldsymbol{a})$ を最大化するよう最尤推定（経験ベイズ法による最適化）を行うことである．しかし，周辺尤度 $p(\boldsymbol{X}; \boldsymbol{\mu}, \boldsymbol{a})$ を解析的に計算することはできないため，変分ベイズ法を用いて，事後分布を因子分解できる形

$$q(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{H}) = \prod_{ni} q(\theta_{ni}) \prod_j q(\phi_j) \prod_{nij} q(H_{nij}) \tag{30}$$

であると仮定し，対数周辺尤度 $\log p(\boldsymbol{X}; \boldsymbol{\mu}, \boldsymbol{a})$ の変分下限 $\mathcal{L}$ を最大化するように最適化を行う．まず，対数周辺尤度 $\log p(\boldsymbol{X}; \boldsymbol{\mu}, \boldsymbol{a})$ に対する第一の変分下限 $\mathcal{L}_0$ は

$$\begin{aligned}
&\log p(\boldsymbol{X}; \boldsymbol{\mu}, \boldsymbol{a}) \geqq \mathbb{E}[\log p(\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{H}; \boldsymbol{\mu}, \boldsymbol{a})] \\
&+ \mathbb{E}[\log p(\boldsymbol{\theta})] + \mathbb{E}[\log p(\boldsymbol{\phi})] + \mathbb{E}[\log p(\boldsymbol{H})] \\
&- \mathbb{E}[\log q(\boldsymbol{\theta})] - \mathbb{E}[\log q(\boldsymbol{\phi})] - \mathbb{E}[\log q(\boldsymbol{H})] \equiv \mathcal{L}_0
\end{aligned} \tag{31}$$

で与えられる．右辺第一項は依然として解析的に計算できないが，対数関数に対する Jensen の不等式を利用することで，さらなる変分下限を求めることができる．

$$\begin{aligned}
&\mathbb{E}[\log p(\boldsymbol{X}|\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{H}; \boldsymbol{\mu}, \boldsymbol{a})] \\
&= \sum_{mn} X_{mn} \mathbb{E}\left[\log \sum_{ijr} \theta_{ni} \phi_j S_{mnir} F_{nijr} H_{nij}\right] \\
&\quad - \sum_{mnijr} \mathbb{E}[\theta_{ni} \phi_j S_{mnir} F_{nijr} H_{nij}] + \mathrm{const.} \\
&\geqq \sum_{mnijr} \lambda_{mnijr} X_{mn} \mathbb{E}\left[\log \frac{\theta_{ni} \phi_j S_{mnir} F_{nijr} H_{nij}}{\lambda_{mnijr}}\right] \\
&\quad - \sum_{mnijr} \mathbb{E}[\theta_{ni} \phi_j S_{mnir} F_{nijr} H_{nij}] + \mathrm{const.}
\end{aligned} \tag{32}$$

ここで，$\lambda_{mnijr}$ は補助変数であり，$\sum_{ijr} \lambda_{mnijr} = 1$ を満たす．等号が成立する，すなわち $\mathcal{L}_0$ が最大化される条件は，

$$\lambda_{mnijr} \propto \exp(\mathbb{E}[\log(\theta_{ni} \phi_j S_{mnir} F_{nijr} H_{nij})]) \tag{33}$$

で与えられる．式 (32) を用いることで，最大化すべき目的関数 $\mathcal{L}$ は，$\mathcal{L}_0$ の下限として求めることができる．

ここで，以降で説明する数学的導出のために，$X_{mnijr}$ および $Y_{mnijr}$ を次式で定義しておく．

$$X_{mnijr} = \lambda_{mnijr} X_{mn} \tag{34}$$

$$Y_{mnijr} = \mathbb{E}[\theta_{ni} \phi_j S_{mnir} F_{nijr} H_{nij}] \tag{35}$$

## 3.3 確率変数 $\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{H}$ に対する変分ベイズ推論

一般に，式 (30) で分解されたそれぞれの確率変数に関する変分事後分布は次式で求められる．

$$q(\boldsymbol{\theta}) \propto \exp(\mathbb{E}_{q(\boldsymbol{\phi}, \boldsymbol{H})}[\log p(\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{H}; \boldsymbol{\mu}, \boldsymbol{a})]) \tag{36}$$

$$q(\boldsymbol{\phi}) \propto \exp(\mathbb{E}_{q(\boldsymbol{\theta}, \boldsymbol{H})}[\log p(\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{H}; \boldsymbol{\mu}, \boldsymbol{a})]) \tag{37}$$

$$q(\boldsymbol{H}) \propto \exp(\mathbb{E}_{q(\boldsymbol{\theta}, \boldsymbol{\phi})}[\log p(\boldsymbol{X}, \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{H}; \boldsymbol{\mu}, \boldsymbol{a})]) \tag{38}$$

対数周辺尤度の変分下限 $\mathcal{L}$ は各確率変数に関する線形項および対数項を含んでいることから，各確率変数に対応する変分事後分布は，事前分布と同じ形であるガンマ分布をとることが分かる．したがって，

$$q(\theta_{ni}) = \mathrm{Gamma}(a^\theta_{ni}, b^\theta_{ni}) \tag{39}$$

$$q(\phi_j) = \mathrm{Gamma}(a^\phi_j, b^\phi_j) \tag{40}$$

$$q(H_{nij}) = \mathrm{Gamma}(a^H_{nij}, b^H_{nij}) \tag{41}$$

とすると，変分パラメータは次式で求められる．

$$a^\theta_{ni} = \frac{\alpha_\theta}{I} + \sum_{mjr} X_{mnijr} \tag{42}$$

$$b^\theta_{ni} = \alpha_\theta + \sum_{mjr} \mathbb{E}[\phi_j H_{nij}] W_{nijm} \tag{43}$$

$$a^\phi_j = \frac{\alpha_\phi}{J} + \sum_{mnir} X_{mnijr} \tag{44}$$

$$b^\phi_j = \alpha_\phi + \sum_{mnir} \mathbb{E}[\theta_{ni} H_{nij}] W_{nijm} \tag{45}$$

$$a^H_{nij} = a_H + \sum_{mr} X_{mnijr} \tag{46}$$

$$b^H_{nij} = b_H + \sum_{mr} \mathbb{E}[\theta_{ni} \phi_j] W_{nijm} \tag{47}$$

## 3.4 パラメータ $\boldsymbol{\mu}, \boldsymbol{a}$ に対する乗法更新

パラメータ $\boldsymbol{\mu}$ および $\boldsymbol{a}$ を推定するためには，乗法更新アルゴリズム [2,9] を用いることができる．一般に，あるスカラパラメータ $x$ に依存するコスト関数 $\mathcal{C}$ が与えられ，$\mathcal{C}$ が最小となるような $x$ を求める問題を考える．$\mathcal{C}$ の $x$ に関する偏微分が $\frac{\partial \mathcal{C}}{\partial x} = R - R'$ という二つの正の項の差で表現可能であるとき，$x$ の乗法更新則は $x \leftarrow \frac{R'}{R} x$ で与えられる．この更新によって，$x$ が正であるとき，正値性は自動的に保持される．本研究では，

周辺尤度の変分下限 $\mathcal{L}$ の符号を反転したものをコスト関数とみなせば，乗法更新アルゴリズムを適用できる．

まず，$-\mathcal{L}$ を基本周波数 $\mu_{ni}$ に関して偏微分すると

$$\frac{-\partial\mathcal{L}}{\partial\mu_{ni}} = R_{ni} - R'_{ni} \tag{48}$$

を得る．ここで，$R_{ni}$ および $R'_{ni}$ は正の項であり，

$$R_{ni} = \sum_{mjr}(\mu_{ni} + 1200\log_2 r)X_{mnijr} + f_m Y_{mnijr} \tag{49}$$

$$R'_{ni} = \sum_{mjr} f_m X_{mnijr} + (\mu_{ni} + 1200\log_2 r)Y_{mnijr} \tag{50}$$

で求まる．したがって，スカラパラメータである $\mu_{ni}$ の乗法更新則は次式で与えられる．

$$\mu_{ni} \leftarrow R_{ni}^{-1} R'_{ni}\mu_{ni} \tag{51}$$

この更新によって $\mu_{ni}$ の正値性は保証されるが，$\mu_{ni}$ は対数基本周波数 [cents] であるから理論的には負の値をとってもよい．しかし，本研究は過去の研究を参考に，対数周波数軸の基準となる 0 [cents] をピアノの最低音よりもさらに低い $440 \times 2^{3/12-5} = 16.3516$ [Hz] に設定しているため，$\mu_{ni}$ は正の値であると考えて実用上は問題ない．

次に，関連研究 [2,9] で行われているのと同様に，$-\mathcal{L}$ を $\boldsymbol{a}_j$ に関して微分すると

$$\frac{-\partial\mathcal{L}}{\partial\boldsymbol{a}_j} = (\boldsymbol{R}_j - \boldsymbol{R}'_j)\boldsymbol{a}_j \tag{52}$$

を得る．ここで，$\boldsymbol{R}_j$ および $\boldsymbol{R}'_j$ は正定値行列であり，

$$\boldsymbol{R}_j = \sum_{mnir} X_{mnijr}F_{nijr}^2\boldsymbol{U}(\omega_{nir}) \tag{53}$$

$$\boldsymbol{R}'_j = \sum_{mnir} Y_{mnijr}F_{nijr}^2\boldsymbol{U}(\omega_{nir}) \tag{54}$$

で求まる．したがって，ベクトルパラメータである $\boldsymbol{a}_j$ に関するベクトル型の乗法更新則は次式で与えられる．

$$\boldsymbol{a}_j \leftarrow \boldsymbol{R}_j^{-1}\boldsymbol{R}'_j\boldsymbol{a}_j \tag{55}$$

これは，DAP を用いたスペクトル包絡推定における乗法更新則式 (15) と同様の形式をとる．ここで，$\boldsymbol{a}_j$ に関しては何の制約もおいておらず，更新を繰り返すと限りなく 0 に近づいたり，発散したりする可能性がある．これを防ぐには，$a_0 = 1$ となるように，強制的にスペクトル包絡全体をスケーリングするステップを加えることが有効である．これにより学習アルゴリズムの収束性は保証されなくなるが，実用上はそれほど問題にはならないことが多い．この解決については今後の課題とする．

## 4. 評価実験

本章では，無限重畳離散全極モデル (iSDAP) を多重音解析に適用した結果について報告する．

### 4.1 実験条件

実験には，MAP ピアノデータベース [21] に含まれるピアノ演奏 “MUS-mz_570_1_ENSTDkCl” から，文献 [8] と同様に冒頭 30 秒を切り出した 44.1kHz・モノラルの音響信号を用いた．

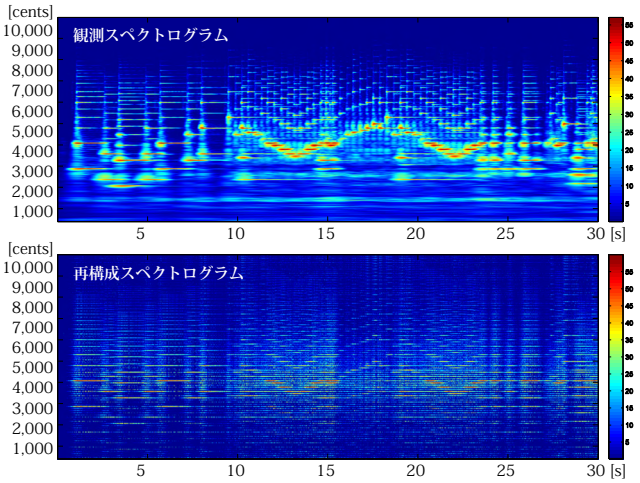


図 5 “MUS-mz_570_1_ENSTDkCl” の観測ウェーブレットスペクトログラムと無限重畳離散全極モデルの推定結果から生成した再構成スペクトログラム

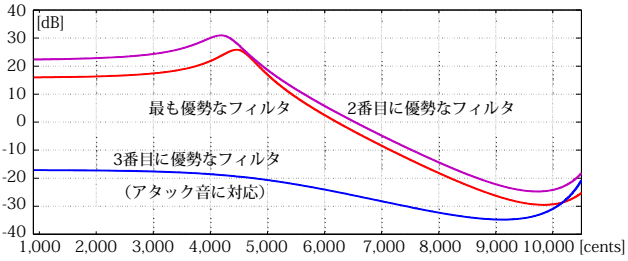


図 6 “MUS-mz_570_1_ENSTDkCl” から推定された全極型スペクトル包絡（フィルタ）

対数周波数領域の振幅スペクトログラムを得るため，ガボール関数を用いた連続ウェーブレット変換を行った．その際，時間方向に 10 [ms] ずつシフトさせながら，900 [cents] から 10400 [cents] まで 10 [cents] 間隔で解析を行った（$M = 960$ および $N = 3000$）．超パラメータは $I = 88$, $J = 3$, $\alpha_\theta = \alpha_\phi = 1$, $R = 20$, $P = 12$, $\sigma^2 = 100$, $a_H = 1$, $b_H = \mathbb{E}_{\mathrm{emp}}[X_{mn}]^{-1}$ と設定した．観測スペクトログラム $\boldsymbol{X}$ にはピアノ音しか含まれていないため，フィルタ数は $J = 3$ とした．各フレームにおける対数基本周波数 $\{\mu_{ni}\}_{i=1}^{I}$ は，ピアノの 88 鍵の音高に対応するように初期化し，他のパラメータはランダムに初期化した．

各フレームに含まれる複数の F0 は閾値処理によって求めた．具体的には，ソース $i$ の基本周波数成分 $\sum_j \theta_{ni}\phi_j F_{nij1}S_{mni1}H_{nij}$ がある閾値より大きい場合，フレーム $n$ には $\mu_{ni}$ で示される F0 が含まれているとした．閾値は，フレームレベルの適合率と再現率のバランスをとって F 値が最大化するように設定した．

### 4.2 実験結果

図 5 に示す通り，変分ベイズ法と乗法更新アルゴリズムを用いることにより，観測スペクトログラムをよく近似するパラメータを学習することができた．しかし，主に発音時刻付近に見られる周波数方向に広く分布しているノイズ成分（例：ハンマーがピアノ弦をたたく際の打撃音）に対しては，必要以上に多数の調波構造の重ね合わせによってできるだけ正確に近似し

ようとする振る舞いが見られた．したがって，現状では最新の手法 [7, 8] に比べて F0 の推定精度は遠く及ばないが，音響理論に裏付けられた統計的モデリングアプローチは有望であると考えている．図 6 に示すように，推定されたスペクトル包絡は，調波構造のピークに影響を受けず，なだらかな形状であった．このことは，観測データのみから，倍音成分の重みは指数的に減衰する傾向があることが学習できたことを示す．

iSDAP の問題点として，ソースやフィルタの実効的な個数は自動調節されるものの，F0 推定のためには最終的に音量に対する閾値処理が必要になることが挙げられる．この問題を解決するには，音符の存在を表現するバイナリ潜在変数を導入する方法が考えられる．また，F0 推定精度を改善するには，文献 [6, 11] と同様に，音高をもつ楽器音と音高を持たない打楽器音に対してそれぞれ異なるモデルを定式化し，それらの重ね合わせで音楽音響信号を表現する方法が考えられる．ただし，ベイズモデルにおいては，事前分布の影響で異なる性質を持つモデルを適切な重みで安定的に組み合わせることは簡単ではない．予備実験では，打楽器音に対するモデルのみで混合音スペクトログラム全体を表現してしまう例が頻繁に見られた．この問題を解決するには，さらなる研究が必要である．

## 5. おわりに

本稿では，無限重畳離散全極モデル (iSDAP) と呼ぶノンパラメトリックベイズソース・フィルタ NMF について述べた．本モデルは，ウェーブレット変換で得られる対数周波数スペクトログラムを，音源の調波構造スペクトル（音高），フィルタの全極型スペクトル包絡（音色），およびそれらの組み合わせの音量に分解することができる．この結果，線形周波数領域で定式化されている無限複合自己回帰モデル (iCAR) のもつ問題点を克服することができた．未知変数の事後分布の推定およびパラメータの最適化のために変分ベイズ法と乗法更新アルゴリズムを統合する手法を提案し，実験で動作を確認した．今後は，多重音基本周波数推定と自動採譜の間にあるギャップを埋めるため，音符配置に関する事前分布をベイズ的に統合することを検討している．

**謝辞:** 本研究の一部は，JSPS 科研費 26700020，24220006，24700168 および JST CREST「OngaCREST プロジェクト」の支援を受けた．

## 文　献

[1] P. Smaragdis, C. Févotte, G. Mysore, N. Mohammadiha, and M. Hoffman. Dynamic source separation using nonnegative factorizations: A unified view. *IEEE Signal Processing Magazine*, 31(3):66–75, 2014.

[2] R. Hennequin, R. Badeau, and B. David. NMF with time-frequency activations to model nonstationary audio events. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(4):744–753, 2011.

[3] J.-L. Durrieu, G. Richard, B. David, and C. Févotte. Source/filter model for unsupervised main melody extraction from polyphonic audio signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3):564–575, 2010.

[4] T. Virtanen and A. Klapuri. Analysis of polyphonic audio using source-filter model and non-negative matrix factorization. In *NIPS Workshop on Advances in Models for Acoustic Processing*, 2009.

[5] H. Kameoka and K. Kashino. Composite autoregressive system for sparse source-filter representation of speech. In *IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 2477–2480, 2009.

[6] K. Yoshii and M. Goto. Infinite composite autoregressive models for music signal analysis. In *International Conference on Music Information Retrieval (ISMIR)*, pages 79–84, 2012.

[7] J. J. Carabias-Orti, T. Virtanen, P. Vera-Candeas, N. Ruiz-Reyes, and F. J. Cañadas-Quesada. Musical instrument sound multi-excitation model for non-negative spectrogram factorization. *IEEE Journal of Selected Topics in Signal Processing*, 5(6):1144–1158, 2011.

[8] N. Bertin, R. Badeau, and E. Vincent. Enforcing harmonicity and smoothness in Bayesian non-negative matrix factorization applied to polyphonic music transcription. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3):538–549, 2010.

[9] R. Badeau and A. Ozerov. Multiplicative updates for modeling mixtures of non-stationary signals in the time-frequency domain. In *European Signal Processing Conference (EUSIPCO)*, 2013.

[10] M. Hoffman, D. Blei, and P. Cook. Bayesian nonparametric matrix factorization for recorded music. In *International Conference on Machine Learning (ICML)*, pages 439–446, 2010.

[11] E. Benetos, S. Ewert, and T. Weyde. Automatic transcription of pitched and unpitched sounds from polyphonic music. In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 3131–3135, 2014.

[12] E. Vincent, N. Bertin, and R. Badeau. Adaptive harmonic spectral decomposition for multiple pitch estimation. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(3):528–537, 2010.

[13] N. J. Bryan, G. Mysore, and G. Wang. Source separation of polyphonic music with interactive user-feedback on a piano roll display. In *International Conference on Music Information Retrieval (ISMIR)*, pages 119–124, 2013.

[14] D. Lee and H. Seung. Algorithms for non-negative matrix factorization. In *Neural Information Processing Systems (NIPS)*, pages 556–562, 2000.

[15] A. T. Cemgil. Bayesian inference for nonnegative matrix factorisation models. *Computational Intelligence and Neuroscience*, 2009:Article ID 785152, 2009.

[16] F. Itakura and S. Saito. Analysis synthesis telephony based on the maximum likelihood method. In *International Congress on Acoustics (ICA)*, pages C17–C20, 1968.

[17] A. El-Jaroudi and J. Makhoul. Discrete all-pole modeling. *IEEE Transactions on Signal Processing*, 39(2):411–423, 1991.

[18] C. Févotte, N. Bertin, and J.-L. Durrieu. Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis. *Neural Computation*, 21(3):793–830, 2009.

[19] 安良岡直希, 奥乃博: 調波・非調波・音色構造因子分解による音響信号分析と音源分離インターフェースへの応用, **情報処理学会研究報告**, Vol. 2012-MUS-94, pp. 1–8, 2012.

[20] 亀岡弘和: 全極型声道モデルと F0 パターン生成過程モデルを内部にもつ統一的音声生成モデル, **日本音響学会秋季研究発表会講演論文集**, pp. 211—214, 2010.

[21] V. Emiya, R. Badeau, and B. David. Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle. *IEEE Transactions on Audio, Speech and Language Processing*, 18(6):1643–1654, 2010.