

# ベイジアン非負値調波因子分解と 多重基本周波数推定への応用

阪上 大地<sup>1,a)</sup> 大塚 琢馬<sup>1</sup> 糸山 克寿<sup>1</sup> 奥乃 博<sup>1</sup>

概要：本稿では、音の三要素（音量・音高・音色）を同時にモデル化し、音楽の多重音解析を行う手法、ベイジアン非負値調波領域分解について報告する。本手法では、非負値行列因子分解 (NMF) にならい、観測音のウェーブレットスペクトログラムを基底とアクティベーションの積に分解する。さらに、各基底を調波構造を模した混合ガウス分布とすることで、各単音の音高・音色を表現する。これは、NMF と調波クラスタリングを統合した非負値調波因子分解 (NHF) という手法により実現する。パラメータの推定時には、特性事前分布と呼ぶ新しい確率分布族を用いてより正確な多重音解析を行う解を探索する。従来法と提案法それぞれで、一様乱数を初期値とする多重基本周波数解析の性能評価を行った結果、F 値基準で平均 5.2% 性能が向上することを確認した。

## 1. はじめに

近年、情報技術の発達によって、音楽が様々な形で楽しめるようになった。たとえば、研究領域では、楽曲構造を可視化する音楽プレーヤ [1]、楽曲中の各パートの音量を変えられる楽器音イコライザ [2]、人間の歌い方を模倣する歌唱合成システム [3] など、以前は存在しなかった高機能なアプリケーションが開発されている。いっぽう、インターネット上ではアマチュア音楽家が既存作品の一部をアレンジして再投稿する、CGM (Consumer Generated Media) とよばれる現象が広がっている。したがって本分野は、研究・産業の両面で最も活発な領域のひとつといえる。このような急速な技術革新を支えるため、音楽の構造や演奏表現を分析するための音楽解析システムが日夜研究されている [4-9]。中でも、音楽の各時刻で鳴っている複数の音の音高を推定する多重基本周波数推定は、音楽分析における最重要タスクのひとつである。

従来、調波クラスタリングにもとづく多重音解析手法では、モデルの自由度が高すぎるため、過学習によって誤った局所解に到達する問題がある。そこで一般には、初期値・事前分布を最適化することで、よりよい推定結果を得る工夫がなされる。ここで、初期値の最適化とは、観測スペク

トログラムの値を参考にして、ヒューリスティックな規則によってパラメータの初期化を行うことを意味する。具体的には、スペクトログラム中のピークを検出したり [10]、音源モデルの倍音構造をピアノやギターに合わせて初期化する [8]。また、事前分布の最適化とは、推定の際にピアノやギターなどにみられる典型的な倍音構造を優遇し、それ以外の倍音構造を排除する手法である [4]。しかし、これらの手法は確からしい解に相対的な重みを付けているに過ぎず、不自然なパラメータをモデル上禁止しているわけではない。実際、調波クラスタリングによる多重基本周波数解析の推定エラーの多くは、不自然な倍音構造を誤って推定することで発生する。

本稿では、以上の問題を解決するため、音の三要素（音量・音高・音色）を明示的にモデル化し、かつ各パラメータの探索範囲を限定して推論を行う手法を報告する。音の三要素を同時に推論するため、本手法では、非負値行列因子分解 (NMF) と調波クラスタリングを統合する。さらに、音高・音色のパラメータの探索範囲を限定するため、特性事前分布と呼ぶ新しい確率分布族を導入する。音色の探索範囲の決定では、我々が先に提案した倍音コーパスモデル [8] を踏襲した。これは、楽器音として適切な倍音構造の範囲がある凸包の内部に制限されるという考え方である。以下で提案法について詳細に説明する。

## 2. 非負値調波因子分解法

本節では、Cemgil らの Bayesian NMF [11] を起点として、各基底に適応的に更新可能な混合ガウス分布を組み

<sup>1</sup> 京都大学 大学院情報学研究所  
606-8501 京都市左京区吉田本町 京都大学 工学部 10 号館  
<sup>a)</sup> dsakaue@kuis.kyoto-u.ac.jp

Bayesian Nonnegative Harmonic Factorization and Its Application to Multipitch Estimation: Daichi Sakaue, Takuma Otsuka, Katsutoshi Itoyama, and Hiroshi G. Okuno (Kyoto Univ.)

込む方法を説明する．本稿では，これを非負値調波因子分解法と呼ぶ．以下，Normal, Wishart, Poisson, Multinomial, Gamma, Dirichlet 分布の確率密度関数をそれぞれ  $\mathcal{N}, \mathcal{W}, \mathcal{P}, \mathcal{M}, \mathcal{G}, \mathcal{D}$  であらわす．また， $t, f, k, m$  はそれぞれ時間フレーム，対数周波数ビン，基底，倍音のインデックスである．角括弧  $\llbracket \rrbracket$  は括弧中の添字に関する集合またはベクトルをあらわす．

## 2.1 Bayesian NMF

Bayesian NMF では，音響信号の観測スペクトログラム  $Y_{tf}$  の各時刻でのスペクトルが高々  $K$  個の基底によって生成されると仮定する． $k$  番目の基底のスペクトル構造を  $h_f^k$  とし，時刻  $t$  での基底の音量を  $u_t^k$  とすると， $Y_{tf} \approx \sum_k u_t^k h_f^k$  となる．実際には，観測スペクトログラムの値と音源モデル間のコスト  $\sum_{tf} D(Y_{tf} || \sum_k u_t^k h_f^k)$  を最小化するようにモデルパラメータを学習する．Bayesian NMF ではコスト関数として KL ダイバージェンスを用いる．このモデルの確率密度関数は以下ようになる．

$$p(Y_{tf} | X_{tf}^{[k]}) = \delta \left( Y_{tf} - \sum_k X_{tf}^k \right) \quad (1)$$

$$p(X_{tf}^k) = \mathcal{P}(X_{tf}^k | h_f^k u_t^k) \quad (2)$$

ここで， $X_{tf}^k$  は  $k$  番目の基底が生成したスペクトログラムである．Bayesian NMF では，潜在変数に共役事前分布

$$p(u_t^k) = \mathcal{G}(u_t^k | a_0, b_0) \quad (3)$$

$$p(h_f^k) = \mathcal{G}(h_f^k | a_0, b_0) \quad (4)$$

を与え，ベイズ推論を行う．

一般的に，NMF では  $h_f^k$  は各周波数で独立にモデル化される．しかし，ピアノやギターなどの調波音のスペクトルは調波構造と呼ばれる特徴的な構造を持つため，これを混合ガウス分布 (GMM) としてモデル化したい．さらに，GMM のパラメータは観測信号に応じて適応的に変化することが望ましい．これを式で表現すると，

$$h_f^k \propto \sum_m \tau_{km} \mathcal{N}(x_f | \mu_k + o_m, \lambda_k^{-1}) \quad (5)$$

となる．ここで， $x_f$  は  $f$  番目の周波数ビンの対数周波数， $o_m$  は第  $m$  倍音のオフセット， $\tau_{km}$  は各倍音コンポーネントの相対振幅比をあらわす．

## 2.2 提案法のアイデア

ここで，非負値調波因子分解法のアイデアを説明する．本手法ではまず，時刻  $t$  で基底  $k$  から発生する音粒子の数  $S_t^k$  を，ポアソン分布  $\mathcal{P}(S_t^k | u_t^k)$  にしたがって決定する．次に， $S_t^k$  個の音粒子を基底に対応する混合ガウス分布を用いて対数周波数軸上にばらまく．GMM のパラメータを  $\tau_{km}, \mu_k, \lambda_k$  とすると，粒子ひとつの観測尤度は次のようになる．

$$p(x | \tau_k, \mu_k, \lambda_k) = \sum_m \tau_{km} \mathcal{N}(x | \mu_k, \lambda_k^{-1}) \quad (6)$$

次に，実際の観測スペクトログラムは，多数の音粒子からなるヒストグラムであると考えられる．ただし便宜上， $f$  番目の周波数ビンは  $x_f - \epsilon/2 \leq x \leq x_f + \epsilon/2$  の周波数幅に対応させる．この時， $k$  番目の基底， $m$  番目の倍音の，時刻  $t$ ，周波数  $f$  での観測粒子数  $s_{tf}^{km}$  は以下の分布にしたがう．

$$\begin{aligned} p(s_{tf}^{km} | u_t^k, \tau_k, \mu_k, \lambda_k) &= \sum_{S_t^k} p(s_{tf}^{km} | S_t^k, \tau_k, \mu_k, \lambda_k) p(S_t^k | u_t^k) \\ &= \mathcal{P}(s_{tf}^{km} | u_t^k \tau_{km} \int_{x_f - \epsilon/2}^{x_f + \epsilon/2} \mathcal{N}(x | \mu_k, \lambda_k^{-1}) dx) \\ &\approx \mathcal{P}(s_{tf}^{km} | \epsilon u_t^k \tau_{km} \mathcal{N}(x_f | \mu_k, \lambda_k^{-1})) \end{aligned} \quad (7)$$

これは，Poisson 分布の thinning として知られている [12]．

さらに，以下の分布との同時分布を考えることで，完全同時分布  $p(Y, S, u, \tau, \mu, \lambda)$  のベイズ推論が可能となる．

$$p(Y_{tf} | s_{tf}^{[km]}) = \delta \left( Y_{tf} - \sum_{km} s_{tf}^{km} \right) \quad (8)$$

$$p(u_t^k) = \mathcal{G}(u_t^k | a_0, b_0) \quad (9)$$

$$p(\tau_k) = \mathcal{D}(\tau_k | \alpha_0) \quad (10)$$

$$p(\mu_k, \lambda_k) = \mathcal{N}(\mu_k | m_k, (\beta_0 \lambda_k)^{-1}) \mathcal{W}(\lambda_k | w_0, \nu_0) \quad (11)$$

## 2.3 更新式の導出

本手法は， $\epsilon \rightarrow 0$  の極限で各変数の事後分布が共役形となり，変分ベイズ法や Gibbs サンプリングを用いて容易にパラメータ更新ができる．変分ベイズ法を用いる場合， $u_t^k$  の最適な事後分布は以下の形になる．

$$\begin{aligned} \ln q^*(u_t^k) &= \mathbb{E}[\ln p(Y, S, u, \tau, \mu, \lambda)] \\ &= \sum_{fm} \mathbb{E}[s_{tf}^{km}] \ln u_t^k + (a_0 - 1) \ln u_t^k \\ &\quad - \sum_{fm} \epsilon u_t^k \tau_{km} \mathcal{N}(x_f | \mu_k + o_m, \lambda_k^{-1}) \\ &\approx \left( a_0 + \sum_{fm} \mathbb{E}[s_{tf}^{km}] - 1 \right) \ln u_t^k \end{aligned} \quad (12)$$

上式では，推論に影響しない定数項を省略した．両辺の指数をとることで，以下の形となる．

$$q^*(u_t^k) = \mathcal{G}(u_t^k | a_t^k, b_0) \quad (13)$$

$$a_t^k = a_0 + \sum_{fm} \mathbb{E}[s_{tf}^{km}] \quad (14)$$

$\tau_k, \mu_k, \lambda_k$  についても，同様の近似を行うことで推論が可能である．なお， $s_{tf}^{km}$  の更新の際には近似は不要である．本手法の，変分ベイズ法を用いた場合の性能は LHA の性能と厳密に一致する [13]．

この手法は更新式の形が LHA と同じになるため，多重基本周波数推定の性能も同等である．しかし，本手法は以下の二点において LHA よりも有用であると考えられる．

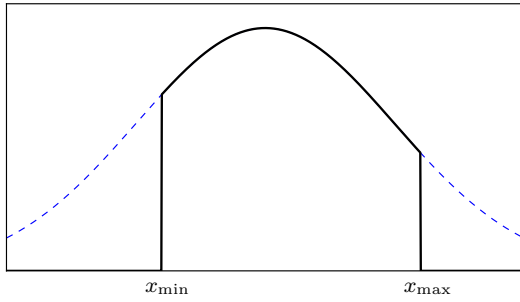


図 1 特性事前分布の図．青の破線が通常の正規分布，黒の実線が特性正規分布をあらわす．この図では，潜在変数の存在範囲が  $x_{\min} \leq x \leq x_{\max}$  に制限されている．

- (1) 本手法では基底の各時刻での音量を  $u_t^k$  の形で陽にモデル化するため，楽器音の音量変化に関する事前知識をモデルに組み込むことができる．既に我々は，HTC [10] にならって音量の時間構造に GMM を導入した場合，推定性能が向上することを確認した [13]．
- (2) LHA の潜在変数の数は音粒子数に比例して爆発的に大きくなるが，本手法では  $O(TFKM)$  である．したがって，EM アルゴリズムによる確定的なパラメータ更新の難しい，非共役な事前分布を用いたモデルに柔軟に応用することができる．本稿では，このアプローチにもとづき推定性能を向上する．

### 3. 特性事前分布

#### 3.1 特性事前分布の定義

LHA をはじめとする，調波クラスタリングに基づく多重基本周波数推定では，音量  $u_t^k$ ，音高  $\mu_k$ ，ピーク形状  $\lambda_k$ ，音色  $\tau_k$  などのパラメータの同時推論が課題となる．多くの場合，観測スペクトログラムと音源モデル間のモデル化誤差を最小化するようにパラメータを学習するが，だからといって，任意のパラメータが許容されるわけではない．というのも，これらのパラメータによって張られる潜在変数の空間の中には，音楽的に不適切な領域が多数含まれているからである．たとえば，調波モデルの目標は調波構造のもつ鋭いピークを捕捉することにあるため，ガウス分布の精度  $\lambda_k$  は一定以上に大きい必要がある．

本稿では，このような形の事前知識を，特性事前分布と呼ぶ新しい確率分布族を用いて表現する．これは，共役事前分布と特性関数の積として表される．特性関数とは，関数の定義域  $A$  の一部  $B$  の内部でのみ 1 となるような関数である．

$$\chi(x) = \begin{cases} 1 & (x \in B) \\ 0 & (\text{otherwise}) \end{cases} \quad (15)$$

したがって，たとえば，特性 Dirichlet 分布  $\mathcal{D}^*$  は以下の形になる．

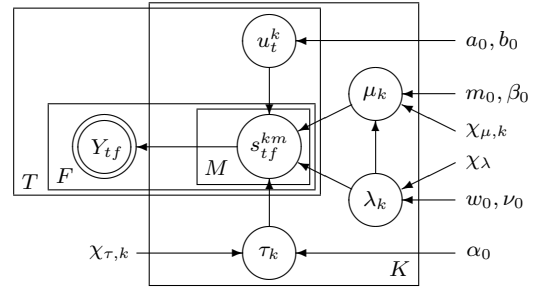


図 2 提案法のグラフィカルモデル．丸印は潜在変数，二重丸は観測変数をあらわす．

$$\mathcal{D}^*(x|\alpha, \chi) = \frac{\mathcal{D}(x|\alpha)}{\int_{\chi} \mathcal{D}(x'|\alpha) dx'} \quad (16)$$

同様に，特性 Normal 分布  $\mathcal{N}^*$ ，特性 Wishart 分布  $\mathcal{W}^*$  を定義する．図 1 に特性 Normal 分布の一例を示す．これらの詳細な定義については付録を参照されたい．

#### 3.2 音高・ピーク形状・音色の事前分布

はじめに，音高  $\mu_k$  の特性関数を決定する．音楽音響信号中に含まれる複数の音の音高は一般的に，平均律にしたがって対数周波数軸上で等間隔にあらわれることが知られている．したがって，各基底を半音間隔で配置できるように特性関数を設定する．

$$\chi_{\mu,k}(x) = \begin{cases} 1 & (\hat{\mu}_k - 50 \leq x \leq \hat{\mu}_k + 50) \\ 0 & (\text{otherwise}) \end{cases} \quad (17)$$

本稿では，基底数  $K$  を 73 とし，6 オクターブの音域をカバーできるようにした．式中の定数  $\hat{\mu}_k$  は  $k$  番目のノートナンバーに対応する対数周波数である．また， $\lambda_k$  について，各倍音成分の標準偏差が 100 [cents] を超えないように設定する．

$$\chi_{\lambda}(x) = \begin{cases} 1 & (x \geq 1/10000[\text{cents}^{-2}]) \\ 0 & (\text{otherwise}) \end{cases} \quad (18)$$

最後に，倍音構造に対する特性関数を設計する．本稿では，我々が既に倍音コーパスモデルとして発表した手法 [8] を踏襲し，適切な倍音構造の範囲が凸包になると仮定する．凸包の頂点座標は，MIDI 音源を使用して録音した様々な楽器音の倍音成分の相対振幅比とする．この倍音凸包は音高ごとに個別に作成する．以下，この凸包の特性関数を  $\chi_{\tau,k}$  であらわす．

#### 4. 非負値調波領域分解法

本節で，非負値調波領域分解法と呼ぶ，音楽音響信号のための新しい確率モデルを定式化する．本手法では，非負値調波因子分解法の事前分布を以下の通り修正する．

$$p(u_t^k) = \mathcal{G}(u_t^k|a_0, b_0) \quad (19)$$

$$p(\tau_{km}) = \mathcal{D}^*(\tau_{km}|\alpha_0, \chi_{\tau,k}) \quad (20)$$

$$p(\mu_k|\lambda_k) = \mathcal{N}^*(\mu_k|m_0, (\beta_0\lambda_k)^{-1}, \chi_{\mu,k}) \quad (21)$$

$$p(\lambda_k) = \mathcal{W}^*(\lambda_k|w_0, \nu_0, \chi_{\lambda}) \quad (22)$$

音量  $u_t^k$  についても特性事前分布を与えることができるが、本稿では省略する。

#### 4.1 推論アルゴリズム

提案法では非共役な事前分布を使用するため、変分ベイズ法のような確定的アルゴリズムを使用することは困難である。そこで、潜在変数の更新に Gibbs サンプルングを使用する。以下で、各変数の更新手順を説明する。

##### 4.1.1 更新式

モデル中のパラメータは各潜在変数の条件付き事後確率からサンプルする。サンプルングは、 $s_{tf}^{km}, u_t^k, \tau_k, \lambda_k, \mu_k$  の順番に行う。 $s_{tf}^{km}$  の事後確率は次のようになる。

$$p(s_{tf}^{[km]} | Y, S_{-tf}^{[km]}, u, \pi, \mu, \lambda) = \mathcal{M}(s_{tf}^{[km]} | Y_{tf}, \rho_{tf}^{[km]}) \quad (23)$$

$$\tilde{\rho}_{tf}^{km} = u_t^k \tau_{km} \mathcal{N}(x_f | \mu_k + o_m, \lambda_k^{-1}) \quad (24)$$

$$\rho_{tf}^{km} = \frac{\tilde{\rho}_{tf}^{km}}{\sum_{k'm'} \tilde{\rho}_{tf}^{k'm'}} \quad (25)$$

ここで、 $\rho_{tf}^{km}$  は  $s_{tf}^{km}$  のサンプルングに必要なパラメータである。

$u_t^k, \tau_k, \mu_k, \lambda_k$  の事後確率は次の通りである。はじめに、事後分布のハイパーパラメータ  $a_t^k, \alpha_{km}, m_k, \beta_k, w_k, \nu_k$  を計算する。

$$a_t^k = a_0 + \sum_{fm} s_{tf}^{km} \quad \alpha_{km} = \alpha_0 + \sum_{tf} s_{tf}^{km} \quad (26)$$

$$\beta_k = \beta_0 + \sum_{tfm} s_{tf}^{km} \quad \nu_k = \nu_0 + \sum_{tfm} s_{tf}^{km} \quad (27)$$

$$m_k = \frac{m_0 \beta_0 + \sum_{tfm} s_{tf}^{km} (x_f - o_m)}{\beta_0 + \sum_{tfm} s_{tf}^{km}} \quad (28)$$

$$w_k^{-1} = w_0^{-1} + \beta_0 m_0^2 + \sum_{tfm} s_{tf}^{km} (x_f - o_m)^2 - \beta_k m_k^2 \quad (29)$$

各変数の事後確率は、非負値調波因子分解法と同様に、 $\epsilon \rightarrow 0$  の極限をとることで計算が可能となる。

$$p(u_t^k | Y, S, u_{-kt}, \tau, \mu, \lambda) \approx \mathcal{G}(u_t^k | a_t^k, b_0) \quad (30)$$

$$p(\tau_k | Y, S, u, \tau_{-k}, \mu, \lambda) \approx \mathcal{D}^*(\tau_k | \alpha_{k[m]}, \chi_{\tau,k}) \quad (31)$$

$$p(\mu_k | Y, S, u, \tau, \mu_{-k}, \lambda) \approx \mathcal{N}^*(\mu_k | m_k, (\beta_k \lambda_k)^{-1}, \chi_{\mu,k}) \quad (32)$$

$$p(\lambda_k | Y, S, u, \tau, \mu, \lambda_{-k}) \approx \mathcal{W}^*(\lambda_k | w_k, \nu_k, \chi_\lambda) \quad (33)$$

提案法のグラフィカルモデルを図 2 に示す。

##### 4.1.2 実装上の注意

上式に基づいてパラメータを更新する際、特性 Dirichlet 分布、特性 Normal 分布、特性 Wishart 分布からサンプルングを行う必要がある。これは、通常の Dirichlet 分布などからサンプルングを行い、特性関数の値が 1 となるサンプルが得られるまで棄却を繰り返すことで実現できる。しかし実際には、条件付き事後分布の確率質量のほとんどが

特性領域外に分布し、棄却ベースのサンプルングに長大な時間がかかることがある。そこで実装上は以下の近似的なサンプルングを使用した。

特性 Normal 分布、特性 Wishart 分布 はじめに、先述の厳密なサンプルングを 100 回試し、特性領域内のサンプルが得られた場合はそこでサンプルングを完了する。そうでない場合かつ、最後に得られたサンプル値が下限を下回った場合、下限値を、上限を上回った場合は上限値をサンプルング結果とする。これは、反復条件付きモード (ICM) [14] として知られる方法に類似している。

特性 Dirichlet 分布 この場合は、多次元のサンプルングとなるので、問題が複雑である。正規分布、Cauchy 分布などを提案分布とする Metropolis-Hastings アルゴリズムを使用することもできるが、倍音領域の構造によっては独立なサンプルが得られるまでに多くの繰り返しが必要のため、不適切である。代わりに、凸包上の一様分布を提案分布とする Metropolis アルゴリズムを 200 回繰り返し、最後に得られた値をサンプルング結果とする。これは、Delaunay 単体分割によって凸包を単体の和に分割し、単体の超体積の比率によってサンプルング対象の単体を選択し、 $\mathcal{D}(x|1)$  からサンプルングした値を単体上に線形変換することで実現できる。単体分割には Qhull ライブラリ [15] を使用した。

## 5. 実験

### 5.1 倍音領域の決定

実験に先立ち、MIDI 音源 (Roland SD-80) を使用し、各音高で許容される倍音構造の範囲を決定した。73 種類の音高それぞれに対し、MIDI プログラム 1 番から 80 番までの楽器音を 1 秒間再生し、これをウェーブレット変換したものを学習データとした。ただし、このデータのうち、非調波成分の割合が 50% を超えるもの、ノートナンバーと異なるピッチを持つものはあらかじめ除去した。ピッチの決定には Subharmonic Summation (SHS) [16] を使用した。有効な楽器音のスペクトログラムを各倍音コンポーネントの周波数帯域ごとに積分し、倍音ベクトルの値とした。

$$\tilde{f}_m = \left(m - \frac{1}{2}\right) \times f_0, \quad \tau_{jm}^0 \propto \sum_{t=1}^T \sum_{\tilde{f}_m^{(\log)} \leq x_f < \tilde{f}_{m+1}^{(\log)}} Y_{tf}^{(j)}, \quad (34)$$

ここで、 $Y_{tf}^{(j)}$  は  $j$  番目の楽器音のスペクトログラム、 $f_0$  は基本周波数、 $\tilde{f}_m^{(\log)}$  は  $\tilde{f}_m$  に対応する対数周波数である。得られた倍音ベクトルの一部を倍音コーパスモデル [8] と同じ方法で省略し、残ったベクトルによって張られる凸包を倍音領域とした。

### 5.2 実験データ

実験には、RWC 音楽データベース [17] から、ピアノソロ

表 1 多重基本周波数推定結果の F 値．太字は最高性能を示す．

Music Type	LHA	OC-LHA	OR-NHF
Piano Solo	0.339	0.563	<b>0.610</b>
Guitar Solo	0.137	0.659	<b>0.678</b>
Jazz (Duo)	0.228	0.484	<b>0.547</b>
Jazz (Trio～)	0.258	0.474	<b>0.520</b>
Chamber	0.247	0.464	<b>0.529</b>

5 曲 (Jazz, No. 1-5), ギターソロ 5 曲 (Jazz, No. 6-10), ジャズデュオ 10 曲 (Jazz, No. 11-20), ジャズバンド 10 曲 (Jazz, No. 21-30), 室内楽 10 曲 (Classic, No. 12-21) を使用した．各楽曲の MIDI ファイルを正解データとし, 音響信号は MIDI 音源 (YAMAHA MOTIF-XS) を用いて作成した．録音時, ドラムトラックはすべてミュートし, 実験から除外した．ジャズデュオ, ジャズバンドはそれぞれ 2 人または 3 人以上の奏者による演奏に対応しており, この人数はドラム奏者を含めずに数えた．実験には各楽曲の先頭 32 秒をウェーブレット変換したものを使用した．

### 5.3 実験条件

提案法の楽曲分析能力を調べるため, 提案法を含む 3 種類の手法で多重基本周波数推定を行い, 推定性能を比較した．推定には LHA, 倍音コーパス LHA, 提案法を用いた．LHA は本分野の一般的な手法のひとつであり, 倍音コーパス LHA は初期値に対し頑健な推論が行える手法として, 我々が以前に提案したものである．実験に際し, ハイパーパラメータはすべて無情報とした．

実験では, はじめに各音源モデルの負担率を一樣乱数を使用して初期化した．これは, 初期値による最適化を一切行わず, 純粋に確率モデルの性能を比較するためである．手法ごとのアルゴリズム繰り返し回数はそれぞれ, 1000 回, 1000 回, 100 回とした．この回数は, F 値の変化をもとに実験的に決定した．

実験終了後, 時刻  $t$  で基底  $k$  が放出したと考えられる音粒子の数の推定値  $N_{tk} = \sum_{f_m} s_{tf}^{km}$  を計算し, この値が一定の閾値を超えるかどうかで, 音の発音・消音を判定した．本システムの出力は  $T \times 128$  の二値行列とした．性能は MIDI ファイルから直接生成した正解データとの比較によって行った．比較基準には F 値を使用した．各手法の潜在的な最高性能を比較するため, 閾値は手法・楽曲ごとに個別に最適化した．

### 5.4 実験結果

実験結果を表 1 に示す．提案法は五種類のデータセットすべてで従来法を上回る性能で, 性能向上率は 1 曲あたり 5.2% であった．図 3 は各楽曲に対する性能向上率を表している．横軸, すなわち従来法の性能が低い場合に性能向上が顕著であるため, 提案法は従来推定の難しかった楽曲に対しても適切に動作することが分かる．唯一, 大幅な悪

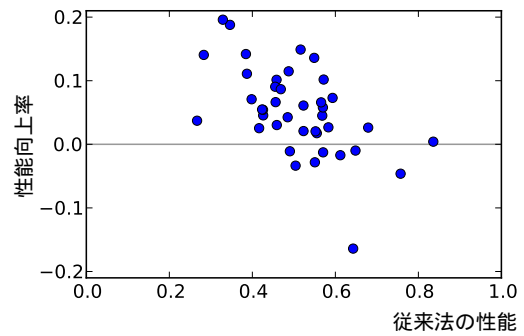


図 3 横軸を倍音コーパス LHA の F 値, 縦軸を (性能向上率) = (本手法の性能 - 倍音コーパス LHA の F 値) としてプロットしたもの．

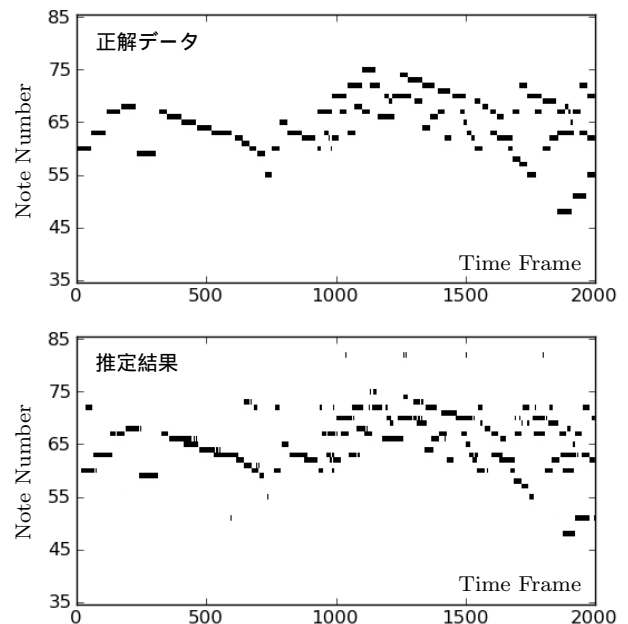


図 4 RM-C012 に対する提案法の推定結果．上から順に, 正解データ, 推定結果に対応している．

化を示した楽曲は Jazz の No. 20 である．しかし, この楽曲では開始直後からピアノのペダルが押下され続けており, MIDI ファイルから自動的に生成した正解データと音響信号の間に大きな食い違いがあった．今後は, 正解データのあり方を含め, 議論をする必要があると考えられる．図 4 に推定結果の一例を示す．

## 6. 関連研究

多重基本周波数解析の本研究以外の手法として, ノンパラメトリックベイズ [4, 7] と音楽構造解析 [5, 7, 9] が挙げられる．本節では, これらについて簡単に説明する．

### 6.1 ノンパラメトリックベイズ

本稿では対処しなかったが, 音楽音響信号のベイズ推論ではモデル選択問題と呼ばれる課題が存在している．たとえば, 本手法では基底数  $K = 73$ , 倍音数  $M = 6$  に限定し

て実装・評価を行ったが、これが適切な基底数・倍音数かどうかは議論の余地がある。ノンパラメトリックベイズでは、モデル次数と呼ばれるこれらの決定困難なパラメータを確率的に決定する。多重基本周波数解析に対する応用例としては、infinite LHA [4] などが挙げられる。

## 6.2 音楽構造解析

本稿では、音楽音響信号をモデル化する際、各音の音量・音高・倍音構造が相互に独立であるという仮定を置いて定式化を行った。しかし実際には、音楽理論の示すとおり、楽曲中にあらわれる音高のパターンには強い相互依存性が存在している [7]。また、楽器音の音量変化を単音単位でテンプレート化する、などの方法も考えられる [5]。こうした高次の依存性を適切にモデル化することは、本研究の今後の課題である。

## 7. おわりに

本稿では、音の三要素（音量・音高・音色）を精密にモデル化し、音楽の多重音解析を行う手法、ベイジアン非負値調波領域分解について報告した。本手法では NMF にならない、観測音のウェーブレットスペクトログラムを基底とアクティベーションの積に分解することで、音楽信号中の各単音をモデル化した。さらに、各基底を調波構造を模した混合ガウス分布を用いて表現した。我々は独自に開発した非負値調波因子分解という手法により NMF と調波クラスタリングを統合し、さらに特性事前分布と呼ぶ新しい確率分布族を用いて潜在変数の探索範囲を望ましい範囲内に制限した。一様乱数を初期値とする多重基本周波数解析の性能評価を行った結果、F 値基準で平均 5.2% の性能向上を確認した。今後は、音色の変化や和音進行などをモデルに組み込み、さらに性能の高い多重音解析手法を開発したい。なお、本研究の一部は、科研費 (S) No. 24220006 の支援を受けた。

## 参考文献

- [1] Goto, M.: A Chorus-Section Detection Method for Musical Audio Signals and Its Application to a Music Listening Station, *IEEE Trans. on ASLP*, Vol. 14, No. 5, pp. 1783–1794 (2006).
- [2] Itoyama, K., Goto, M., Komatani, K., Ogata, T. and Okuno, H. G.: Query-by-Example Music Information Retrieval by Score-Informed Source Separation and Remixing Technologies, *EURASIP J. Adv. in Signal Process.*, Vol. 2010, pp. 1–14 (2010).
- [3] 中野倫靖ほか：VocalListener：ユーザ歌唱の音高および音量を真似る歌声合成システム，情報処理学会論文誌，Vol. 52, No. 12, pp. 3853–3867 (2011).
- [4] Yoshii, K. and Goto, M.: A nonparametric Bayesian multipitch analyzer based on infinite latent harmonic allocation, *IEEE Trans. on ASLP*, Vol. 20, No. 3, pp. 717–730 (2012).
- [5] 落合和樹ほか：音楽生成プロセスの階層ベイズモデリン

グによる音響信号の自動採譜，日本音響学会 2012 年春季研究発表会，pp. – (2012).

- [6] Yoshii, K. and Goto, M.: Unsupervised Music Understanding Based on Nonparametric Bayesian Models, *Proc. ICASSP*.
- [7] Nakano, M. et al.: Nonparametric Bayesian music parser, *Proc. ICASSP*, pp. 461–464 (2012).
- [8] Sakaue, D. et al.: Initialization-Robust Multipitch Estimation based on Latent Harmonic Allocation using Overtone Corpus, *Proc. ICASSP*, pp. 425–428 (2012).
- [9] 大石康智ほか：ノート指令と表現指令によって駆動される歌声 F0 生成過程の統計モデル，日本音響学会 2012 年春季研究発表会，pp. 345–348 (2012).
- [10] Kameoka, H. et al.: A Multipitch Analyzer Based on Harmonic Temporal Structured Clustering, *IEEE Trans. on ASLP*, Vol. 15, No. 3, pp. 982–994 (2007).
- [11] Cemgil, A. T.: Bayesian inference for nonnegative matrix factorization models, *Technical Report CUED/F-INFENG/TR.609* (2008).
- [12] Durrett, R.: *Essentials of Stochastic Processes*, Springer-Verlag, 2nd edition (2012).
- [13] Sakaue, D. et al.: Bayesian Nonnegative Harmonic-Temporal Factorization and Its Application to Multipitch Analysis, *Proc. ISMIR*, to appear (2012).
- [14] Kittler, J. and Föglein, J.: Contextual Classification of Multispectral Pixel Data, *Image and Vision Computing*, Vol. 2, No. 1, pp. 13–29 (1984).
- [15] Barber, C. B. et al.: The Quickhull Algorithm for Convex Hulls, *ACM Trans. on Mathematical Software*, Vol. 22, No. 4, pp. 469–483 (1996).
- [16] Hermes, D. J.: Measurement of pitch by subharmonic summation, *J. Acoust. Soc. Am.*, Vol. 83, No. 1, pp. 257–264 (1988).
- [17] Goto, M. et al.: RWC Music Database: Popular, Classical, and Jazz Music Databases, *Proc. ISMIR*, pp. 287–288 (2002).

## 付 録

### A.1 特性関数

定義  $B$  の特性関数とは、集合  $A$  の任意の点  $x$  と、 $A$  の部分集合  $B$  に対し、以下の性質を満たす関数  $\chi$  のことである。

$$\chi(x) = \begin{cases} 1 & (x \in B) \\ 0 & (\text{otherwise}) \end{cases} \quad (\text{A.1})$$

本文中で、 $B$  に相当する集合が自明な場合、これを単に特性関数と呼ぶ。

### A.2 特性分布

特性 Normal 分布，特性 Wishart 分布，特性 Dirichlet 分布の確率密度関数は以下の形になる。

$$\mathcal{N}^*(x|\mu, \sigma^2, \chi) = \frac{\mathcal{N}(x|\mu, \sigma^2)}{\int_{\chi} \mathcal{N}(x'|\mu, \sigma^2) dx'} \quad (\text{A.2})$$

$$\mathcal{W}^*(x|a, b, \chi) = \frac{\mathcal{W}(x|a, b)}{\int_{\chi} \mathcal{W}(x'|a, b) dx'} \quad (\text{A.3})$$

$$\mathcal{D}^*(x|\alpha, \chi) = \frac{\mathcal{D}(x|\alpha)}{\int_{\chi} \mathcal{D}(x'|\alpha) dx'} \quad (\text{A.4})$$