

## 複数楽器混合モデルのパラメータ推定と 楽器名同定への応用

糸山克寿<sup>†1</sup> 後藤真孝<sup>†2</sup> 駒谷和範<sup>†1</sup>  
尾形哲也<sup>†1</sup> 奥乃博<sup>†1</sup>

本報告では、音源分離に適用可能な音モデルである、調波・非調波統合モデルを用いた複数楽器混合モデルのパラメータ推定とその楽器名同定への応用について述べる。観測パワースペクトルに適合する複数楽器混合モデルのパラメータは変分法に基づくベイズ推論を用いて推定する。調波・非調波統合モデルのパラメータ分布は楽器によって異なるため、楽器間相対重みが最大となる楽器を選択することで楽器名が同定される。楽器音データベース中から選択した10楽器に対する評価実験で、平均認識率81.6%を得た。

### Parameter Estimation of Mixture Model of Multiple Instruments and Application to Musical Instrument Identification

KATSUTOSHI ITOYAMA,<sup>†1</sup> MASATAKA GOTO,<sup>†2</sup>  
KAZUNORI KOMATANI,<sup>†1</sup> TETSUYA OGATA<sup>†1</sup>  
and HIROSHI G. OKUNO<sup>†1</sup>

This report presents parameter estimation of mixture model of multiple instruments based on the integrated harmonic and inharmonic model, and its application to musical instrument identification. Parameters of the integrated model, which fit an observed power spectrogram, are estimated by a bayesian inference method based on calculus of variations. Since parameter distributions of the integrated model depend on each instrument, the instrument name is identified by selecting an instrument which has maximum relative instrument weight. Experimental results showed 81.6% of accuracy for 10 instruments.

<sup>†1</sup> 京都大学 大学院情報学研究科 知能情報学専攻  
Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University  
<sup>†2</sup> 産業技術総合研究所  
National Institute of Advanced Industrial Science and Technology (AIST)

### 1. はじめに

複雑な音楽音響信号中の楽器音を認識し、さらに信号からその構成要素である楽器音や歌声を分離することは、近年の音楽情報処理における重要な要素技術である。楽器音の認識に関しては混合音に対する基本周波数の推定<sup>1)–3)</sup>、テンポ推定やビートトラッキング<sup>4),5)</sup>などを要素技術として、これらを応用した楽器音認識<sup>6),7)</sup>が取り組まれてきた。これらの技術の開発は、要素間で独立に進められてきており、複数の要素を包括的に推定する研究は少なかった。これに対して、混合音に含まれる調波音の音響特徴量を包括的に推定する手法<sup>8),9)</sup>が開発された。これらは、時間周波数平面上のエネルギー分布を制約付きの混合モデルでフィッティングする、調波時間構造化クラスタリングに基づく、複雑な音楽音響信号に対する包括的な音響特徴量の抽出を可能としたが、対象は調波構造を持つ楽器音に限られていた。一方、音源分離に関しては、調波的な楽器音の分離<sup>10),11)</sup>や複雑な音楽音響信号からの打楽器音の分離<sup>7),12)</sup>が取り組まれてきた。しかしこれらの技術は、調波構造を持つ音やドラム音など、混合音中の一部の楽器音のみを対象としたものであった。これに対して、楽曲の楽譜を既知として調波音、非調波音の両方を分離可能な音源分離手法<sup>13)</sup>が開発された。この手法は、楽器名、音高、音長といった情報が楽譜から得られることを前提としており、一般の音楽音響信号に適用するためには汎用性に欠けるという問題点があった。

本報告では、楽器音認識と音源分離との統合をめざし、音源分離に適用可能なフレームワークに基づく楽器名同定手法について述べる。このフレームワークでは混合音の音楽音響信号および楽曲中の各単音の音高と発音区間を入力として、各単音の分離音響信号と楽器名同定結果を出力する問題を想定しているが、本報告では楽器単独音の信号とその音高、発音区間を入力し、楽器名同定結果を出力する問題を考える。複数楽器の調波・非調波統合モデルの重ね合わせとして定義される複数楽器混合モデルに対して、観測パワースペクトルに最も適合するパラメータの分布を推定することで、楽器名同定問題を最適関数形の選択として解く。パラメータの分布は楽器ごとに異なるため、楽器ごとに学習したパラメータの事前分布を用いることで重みパラメータが最大となる楽器を楽器名同定結果とみなすことができる。

### 2. 調波・非調波統合モデル

楽器音のパワースペクトル  $X(t, f)$  を、時刻  $t$ 、周波数  $f$  の2次元平面で定義される確率密度分布であるとみなす。本節での目的は、観測パワースペクトル  $X(t, f)$  を楽器音パワー

表 1 調波・非調波統合モデルパラメータの制約条件と音響的な意味  
Table 1 Constraints for the parameters of the integrated tone model and their acoustical meanings.

記号	制約条件	音響的な意味
$\alpha(l)$	$\sum_l \alpha(l) = 1$	楽器 $l$ の相対重み
$\beta(H), \beta(I)$	$\beta(H) + \beta(I) = 1$	調波構成成分と非調波成分とのパワー比率
$\gamma(m, n)$	$\sum_{m,n} \gamma(m, n) = 1$	調波構成成分の時間周波数平面上でのパワー分布
$\delta(m, n)$	$\sum_{m,n} \delta(m, n) = 1$	非調波成分の時間周波数平面上でのパワー分布
$\tau$		発音時刻
$\rho_H$	$\rho_H > 0$	$m\rho_H$ で調波構成成分の音長
$\rho_I$	$\rho_I > 0$	$m\rho_I$ で非調波成分の音長
$\omega$	$\omega > 0$	基本周波数
$\sigma$	$\sigma > 0$	調波構造の周波数方向へのパワーばらつき
$\kappa, \lambda$	$\kappa > 0, \lambda > 1$	非調波構造モデルの周波数方向への基底分布の配置を定める係数

スペクトルを表すモデルである調波・非調波統合モデルで表現し、適切なモデルパラメータを推定することである。

### 2.1 モデル構造

調波・非調波統合モデル  $J(t, f)$  は、時刻と周波数の 2 次元平面で定義される確率密度関数であり、調波構造モデル  $H(t, f)$  と非調波構造モデル  $I(t, f)$  との混合モデルで表現される。

$$J(t, f) = \beta(H)H(t, f) + \beta(I)I(t, f)$$

さらに、調波構造モデルと非調波構造モデルはそれぞれ時間周波数上の配置に拘束条件を持つ基底分布の混合分布として表現される。

$$H(t, f) = \sum_{m=0}^{M_H-1} \sum_{n=1}^{N_H} \gamma(m, n)H(t, f; m, n) \quad I(t, f) = \sum_{m=0}^{M_I-1} \sum_{n=1}^{N_I} \delta(m, n)I(t, f; m, n)$$

$$H(t, f; m, n) = \mathcal{N}(t; \mathcal{T}_H(m), \rho_H^2) \mathcal{N}(f; \mathcal{F}_H(n), \sigma_H^2)$$

$$I(t, f; m, n) = \mathcal{N}(t; \mathcal{T}_I(m), \rho_I^2) \mathcal{M}(f; n)$$

$$\mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$\mathcal{M}(f; \mu) = \frac{1}{\sqrt{2\pi}(f+\kappa)\log\lambda} \exp\left(-\frac{(\mathcal{F}_I(f)-\mu)^2}{2}\right)$$

$$\mathcal{T}_H(m) = \tau + m\rho_H \quad \mathcal{T}_I(m) = \tau + m\rho_I \quad \mathcal{F}_H(n) = n\omega \quad \mathcal{F}_I(f) = \log_\lambda\left(\frac{f}{\kappa} + 1\right)$$

各パラメータの制約条件と音響的な意味を表 1 に示す。

### 2.1.1 潜在変数の導入

本モデルは、ガウス分布などを基底分布とする混合分布で定義されているため、本モデルから直接設計した目的関数の解析的な最適化は困難である。そこで、EM アルゴリズム、もしくはそれを拡張したパラメータ推定手法を適用するため、潜在変数

$$z = \left\{ \{z_\beta(H), z_\beta(I)\}, \{z_\gamma(m, n)\}_{\substack{m=0, \dots, M_H-1 \\ n=1, \dots, N_H}}, \{z_\delta(m, n)\}_{\substack{m=0, \dots, M_I-1 \\ n=1, \dots, N_I}} \right\}$$

を導入する。これらの潜在変数は、いずれも 1-of- $K$  表現（どれか 1 つの要素だけが 1 で、残りが 0 であるような  $K$  次元ベクトル）をとる。

これらの潜在変数と、時刻、周波数の組  $(t, f)$  との同時分布は以下のように書き下される。

$$p(t, f, z | \theta) = p(z_\beta(H) = 1 | \theta) p(t, f, \{z_\gamma(m, n)\} | z_\beta(H) = 1, \theta) \\ + p(z_\beta(I) = 1 | \theta) p(t, f, \{z_\delta(m, n)\} | z_\beta(I) = 1, \theta)$$

$$p(z_\beta(H) = 1 | \theta) = \beta(H) \quad p(z_\beta(I) = 1 | \theta) = \beta(I)$$

$$p(t, f, \{z_\gamma(m, n)\} | z_\beta(H) = 1, \theta) =$$

$$\prod_{m=0}^{M_H-1} \prod_{n=1}^{N_H} p(z_\gamma(m, n) = 1 | z_\beta(H) = 1, \theta) p(t, f | z_\beta(H) = 1, z_\gamma(m, n) = 1, \theta)$$

$$p(t, f, \{z_\delta(m, n)\} | z_\beta(I) = 1, \theta) =$$

$$\prod_{m=0}^{M_I-1} \prod_{n=1}^{N_I} p(z_\delta(m, n) = 1 | z_\beta(I) = 1, \theta) p(t, f | z_\beta(I) = 1, z_\delta(m, n) = 1, \theta)$$

$$p(z_\gamma(m, n) = 1 | z_\beta(H) = 1, \theta) = \gamma(m, n) \quad p(z_\delta(m, n) = 1 | z_\beta(I) = 1, \theta) = \delta(m, n)$$

$$p(t, f | z_\beta(H) = 1, z_\gamma(m, n) = 1, \theta) = H(t, f; m, n)$$

$$p(t, f | z_\beta(I) = 1, z_\delta(m, n) = 1, \theta) = I(t, f; m, n)$$

$\theta$  はパラメータの集合である。

### 2.2 事前分布

後述するパラメータ推定において、パラメータが「常識的」な範囲から逸脱することを防ぐために事前分布を設定する。表 1 に示したパラメータは、制約条件の範囲内で自由な値を取り得るものの、調波・非調波統合モデルが常識的な楽器音を表現していればその値の範囲、すなわち分布、はある程度集中する。例えば多くの楽器では、非調波成分のパワー分布は発音時刻直後に集中し、発音から相当の時刻が経過した後に突然強いパワーが集中することはない。また、パラメータの分布は楽器ごとにある程度異なるものになるため、観測パワースペクトルがどの楽器に起因するかが与えられれば、その楽器として自然なパラメータ

が容易に推定できる．発音時刻  $\tau$  や基本周波数  $\omega$  の分布を楽器ごとに定めることは困難だが，楽譜などから単音の発音時刻や音高が得られればパワー比率などと同様に事前分布を設定することができる．

ここでは，各パラメータに対する以下の事前分布を用いる．

$$\begin{aligned} p(\beta(H), \beta(I)) &= \mathcal{D}(\beta(H), \beta(I); \hat{\phi}_\beta(H), \hat{\phi}_\beta(I)) \\ p(\{\gamma(m, n)\}) &= \mathcal{D}(\{\gamma(m, n)\}; \{\hat{\phi}_\gamma(m, n)\}) \quad p(\{\delta(m, n)\}) = \mathcal{D}(\{\delta(m, n)\}; \{\hat{\phi}_\delta(m, n)\}) \\ p(\tau) &= \mathcal{N}\left(\tau; \hat{\nu}_\tau, \left(\frac{\hat{\xi}_H}{\rho_H^2} + \frac{\hat{\xi}_I}{\rho_I^2}\right)^{-1}\right) \quad p(\omega, \sigma) = \mathcal{N}\left(\omega; \hat{\nu}_\omega, \left(\frac{\hat{\xi}_\omega}{\sigma^2}\right)^{-1}\right) \mathcal{G}(\sigma^{-2}; \hat{\eta}, \hat{\zeta}) \end{aligned}$$

ただし， $\mathcal{D}(\cdot)$  および  $\mathcal{G}(\cdot)$  はそれぞれディリクレ分布，ガンマ分布である．これらの確率密度関数は正規化係数を除いて以下で与えられる．

$$\mathcal{D}(x_1, \dots, x_N; \phi_1, \dots, \phi_N) \propto \prod_{n=1}^N x_n^{\phi_n - 1} \quad \mathcal{G}(x; \eta, \zeta) \propto x^{\eta-1} e^{-\zeta x}$$

各パラメータの事前分布は，それぞれのパラメータに対して共役となるように選んだ．事前分布を設定していないパラメータ  $\rho_H, \rho_I, \kappa$  および  $\lambda$  は定数として扱う．

### 2.3 パラメータ推定

観測パワースペクトル  $X(t, f)$  から，前述の事前分布を用いて調波・非調波統合モデルのパラメータを推定する．ここでは，パラメータの値を推定するのではなく， $X(t, f)$  が与えられた基でのパラメータの分布を推定するベイズ推定を行う．

推定すべき分布は以下で与えられる．

$$p(\theta | X) = \sum_z p(z, \theta | X) = \sum_z \frac{\iint X(t, f) p(t, f, z, \theta) dt df}{\sum_z \int p(X, z, \theta) d\theta}$$

しかしながら，この積分は計算量が膨大であるため，これを直接計算することは実用的ではない．そこで，パラメータの事後分布を直接求めるのではなく，変分法に基づく手法で事後分布を近似的に求める．

事後分布を近似する分布である，テスト分布  $q(z, \theta)$  を導入する．さらに，最適なテスト分布を計算可能にするため，テスト分布が以下のように分解できると仮定する．

$$\begin{aligned} q(z, \theta) &= q(z) q(\theta) \\ q(z) &= q(z_\beta(H), z_\beta(I)) q(\{\gamma(m, n)\}) q(\{z_\delta(m, n)\}) \end{aligned}$$

$q(\theta) = q(\beta(H), \beta(I)) q(\{\gamma(m, n)\}) q(\{\delta(m, n)\}) q(\tau) q(\omega, \sigma)$   
このテスト分布に関する目的関数を以下で定義する．

$$\mathcal{F}[q] = \iiint \sum_z X(t, f) q(z, \theta) \log \frac{p(t, f, z, \theta)}{q(z, \theta)} dt df d\theta$$

この目的関数は関数  $q$  に依存する関数，すなわち汎関数となっている．さらに，目的関数を最大化するテスト分布は，真の事後分布を最もよく近似する分布となる（証明は省略）．

目的関数を最大化するテスト分布を求めるためには，オイラー＝ラグランジュ方程式を解くことになる．パラメータに関するテスト分布  $q(\theta)$  が得られたとき，潜在変数に関する最適テスト分布  $q(z)$  は以下で得られる（詳細な導出は省略）．

$$\begin{aligned} q(\{z_\gamma(m, n)\} | z_\beta(H) = 1) &= C_\gamma \exp \langle \log p(t, f, \{z_\gamma(m, n)\} | z_\beta(H) = 1, \theta) \rangle_{X(t, f), q(\theta)} \\ q(\{z_\delta(m, n)\} | z_\beta(I) = 1) &= C_\delta \exp \langle \log p(t, f, \{z_\delta(m, n)\} | z_\beta(I) = 1, \theta) \rangle_{X(t, f), q(\theta)} \\ q(z_\beta(H), z_\beta(I)) &= C_\beta \exp \langle \log p(t, f, z_\beta(H), z_\beta(I) | \theta) \rangle_{X(t, f), q(\theta)} \end{aligned}$$

$C_\beta, C_\gamma, C_\delta$  は  $\sum_z q(z) = 1$  となるための規格化定数， $\langle x \rangle_{f(x)}$  は変数  $x$  の分布  $f(x)$  に対する期待値である．

また，潜在変数に関するテスト分布が得られたときのパラメータに関する最適テスト分布は以下で得られる．

$$\begin{aligned} q(\beta(H), \beta(I)) &= C'_\beta p(\beta(H), \beta(I)) \exp \langle \log p(t, f, z | \theta) \rangle_{X(t, f), q(z), q(\theta/\{\beta(H), \beta(I)\})} \\ &= \mathcal{D}(\beta(H), \beta(I); \hat{\phi}_\beta(H), \hat{\phi}_\beta(I)) \\ q(\{\gamma(m, n)\}) &= C'_\gamma p(\{\gamma(m, n)\}) \exp \langle \log p(t, f, z | \theta) \rangle_{X(t, f), q(z), q(\theta/\{\gamma(m, n)\})} \\ &= \mathcal{D}(\{\gamma(m, n)\}; \{\hat{\phi}_\gamma(m, n)\}) \\ q(\{\delta(m, n)\}) &= C'_\delta p(\{\delta(m, n)\}) \exp \langle \log p(t, f, z | \theta) \rangle_{X(t, f), q(z), q(\theta/\{\delta(m, n)\})} \\ &= \mathcal{D}(\{\delta(m, n)\}; \{\hat{\phi}_\delta(m, n)\}) \\ q(\tau) &= C'_\tau p(\tau) \exp \langle \log p(t, f, z | \theta) \rangle_{X(t, f), q(z), q(\theta/\{\tau\})} = \mathcal{N}\left(\tau; \hat{\nu}_\tau, \left(\frac{\hat{\xi}_H}{\rho_H^2} + \frac{\hat{\xi}_I}{\rho_I^2}\right)^{-1}\right) \\ q(\omega, \sigma) &= C'_{\omega, \sigma} p(\omega, \sigma) \exp \langle \log p(t, f, z | \theta) \rangle_{X(t, f), q(z), q(\theta/\{\omega, \sigma\})} \\ &= \mathcal{N}\left(\omega; \hat{\nu}_\omega, \left(\frac{\hat{\xi}_\omega}{\sigma^2}\right)^{-1}\right) \mathcal{G}(\sigma^{-2}; \hat{\eta}, \hat{\zeta}) \end{aligned}$$

$C'_\beta, C'_\gamma, C'_\delta, C'_\tau, C'_{\omega, \sigma}$  は  $\int q(\theta) d\theta = 1$  となるための規格化定数である．パラメータに対して共役となる事前分布を選んでいるため，事後分布は事前分布のハイパーパラメータを変化させたものとして表現できる．

潜在変数とパラメータのテスト分布の導出は相互に依存しているため、これらを閉じた形式で解くことはできない。最適な事後分布を推定するためには、交互に更新を繰り返す必要がある。事後分布の具体的な更新式を付録に示す。

### 3. 楽器名同定

パラメータ推定においてベイズ推定を用いることの利点の一つは、観測データに対する最適な関数形の選択が可能にある。ここでは、調波・非調波統合モデルと楽器ごとに学習したパラメータの事前分布を用いた、観測したパワースペクトルに最もよく適合する楽器の選択について述べる。

既知の楽器の総数を  $L$  として、 $L$  個の調波・非調波統合モデルを重み付きで重ね合わせた新たなモデル  $J'(t, f)$  を考える。以降、第2節で述べたパラメータ等には、次式のように「どの楽器に帰属するか」を表す変数  $l$  が追加されたものとする。

$$J'(t, f) = \sum_{l=1}^L \alpha(l) J(l, t, f)$$

$\alpha(l)$  は楽器  $l$  の重みである。さらに、 $\{\alpha(l)\}$  に対する事前分布  $p(\{\alpha(l)\})$

$$p(\{\alpha(l)\}) = \mathcal{D}(\{\alpha(l)\}; \{\tilde{\phi}_\alpha(l)\})$$

を導入し、テスト分布が以下のように分解されると仮定する。

$$q(z, \theta) = q(z) q(\theta)$$

$$q(z) = q(\{z_\alpha(l)\}) \prod_{l=1}^L q(z_\beta(l, H), z_\beta(l, I)) q(\{z_\gamma(m, n)\}) q(\{z_\delta(m, n)\})$$

$$q(\theta) = q(\{\alpha(l)\}) \prod_{l=1}^L q(\beta(l, H), \beta(l, I)) q(\{\gamma(l, m, n)\}) q(\{\delta(l, m, n)\}) q(\tau(l)) q(\omega(l), \sigma(l))$$

ここで、目的関数  $\mathcal{F}[q]$  を最大化するテスト分布  $q(\{z_\alpha(l)\})$  は、真の事後分布  $p(\{z_\alpha(l)\} | X)$  の最良の近似となる。潜在変数  $z_\alpha(l)$  は、観測されたパワースペクトルがどの楽器に帰属するかを表す変数であるため、 $\operatorname{argmax}_l q(z_\alpha(l) = 1)$  を満たす  $l$  は、 $X(t, f)$  に最もよく適合する楽器を表す。

表2 本実験で用いた楽器音。

Table 2 Instrument sounds in this experiment.

Instrument name (Abbreviation)	Number of tones
Piano (PF)	792
Acoustic Guitar (AG)	702
Electric Guitar (EG)	702
Violin (VN)	576
Cello (VC)	565
Trumpet, Cornet (TR)	302
Trombone (TB)	285
Alto Sax (AS)	297
Tenor Sax (TS)	294
Clarinet (CL)	360

表3 調波・非調波統合モデル中の定数

Table 3 Constants of the integrated model.

記号	値
$M_H$	30
$N_H$	100
$N_H$	30
$N_H$	100
$\rho_H$	0.05 sec.
$\rho_I$	0.05 sec.
$\kappa$	440.0 Hz
$\lambda$	1.134

### 4. 評価実験

以上で述べた手法は、楽器音の混合音に対して各単音の楽器名と音色を表すパラメータを同時推定するものであるが、今回は評価実験の第一段階として、RWC 研究用音楽データベース：楽器音<sup>14)</sup>より表2に示す10楽器を選び、それぞれの楽器音から音モデルのパラメータを推定して楽器名を同定する実験を行った。楽器音を楽器ごとに3つのグループに分け、2グループを学習データ、残り1グループを認識データとした。各楽器の事前分布は、学習データから推定した音モデルパラメータの平均値を基に作成した。調波・非調波統合モデル中の定数は表3のように設定した。

#### 4.1 結果と考察

実験によって得られた楽器ごとの認識率を表4に、混同行列を表5に示す。参考として、楽器音に対する音源同定の先行研究<sup>15)</sup>でのF0依存正規分布を用いない場合の認識率を挙げた。本実験とは楽器セットが異なるため認識率差に基づいて性能差を論じることはできないが、同程度の認識率が得られている。

楽器ごとに認識率を見ると、PF, EG, VN, VC, CLが平均より高く、AG, TR, TB, AS, TSが50–70%と平均より低い。これらの認識率が低い楽器では、VN, VC, CLへの誤認識が多いことが混同行列より分かる。これらの楽器では、図1に示すように、非調波構造モデルの周波数方向への相対重みパラメータの楽器内での分散が全ての楽器の平均よりも極めて大きな値をとっている。本来調波構造モデルで表現すべき調波構成成分が、非調波構造モデルで表現されてしまうなどの原因でモデルパラメータに外れ値が生じ、そのパラメータの事前

表 4 楽器ごとの音源同定正解率 .  
Table 4 Accuracy for each instrument.

Inst.	Baseline <sup>15)</sup>	Proposed
PF	74.2%	86.7%
AG	81.2%	66.1%
EG	N/A	99.4%
VN	69.7%	96.0%
VC	73.5%	97.7%
TR	73.5%	51.7%
TB	76.7%	54.7%
AS	41.5%	61.6%
TS	64.7%	63.6%
CL	90.7%	94.7%
Avg.	73.1%	81.6%

表 5 混同行列 .  
Table 5 Confusion matrix.

True inst.	Classified as											Total
	PF	AG	EG	VN	VC	TR	TB	AS	TS	CL		
PF	687	4	0	81	20	0	0	0	0	0	0	792
AG	2	464	4	141	11	0	0	0	0	80	0	702
EG	0	0	698	0	3	0	0	0	0	1	0	702
VN	2	0	0	553	10	0	0	0	0	11	0	576
VC	3	0	0	10	552	0	0	0	0	0	0	565
TR	1	0	2	79	7	156	0	0	0	57	0	302
TB	3	1	1	64	53	0	156	0	0	7	0	285
AS	0	0	0	95	16	0	0	183	0	3	0	297
TS	4	1	0	69	13	0	0	0	187	20	0	294
CL	1	0	0	15	3	0	0	0	0	341	0	360
Total	703	470	705	1107	688	156	156	183	187	520	0	4875

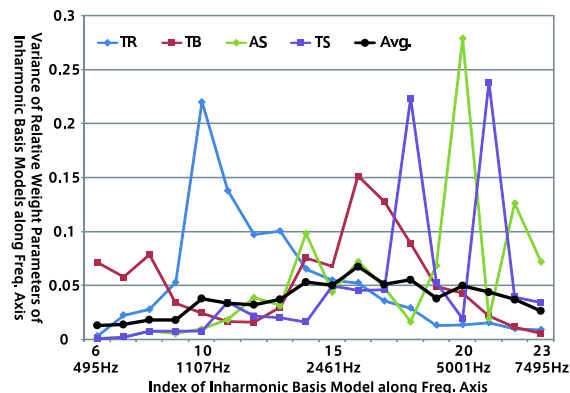


図 1 非調波構造モデルの周波数方向への相対重みパラメータの楽器内での分散 .  
Fig. 1 Variance of relative weight parameters of inharmonic model along frequency axis.

分布が正しく学習できなかったと考えられる .

## 5. おわりに

本報告では、音源分離に適用可能な音モデルを用いた複数楽器混合モデルのパラメータ推定とその楽器音同定への応用について述べた . 変分ベイズ法に基づいてモデルパラメータを推定し、楽器間相対重みが最大となる楽器を選択することで楽器同定を行う . 今後は、単音

に対する楽器同定だけでなく、混合音に対する楽器同定と音源分離の並行処理に取り組む予定である .

謝辞 本研究の一部は、科研費、CrestMuse、およびグローバル COE プログラムに支援を受けた .

## 参 考 文 献

- 1) 柏野邦夫：音楽音響信号を対象とする聴覚的情景分析に関する研究，博士論文，東京大学 (1994).
- 2) Goto, M.: A Real-time Music-scene-analysis System: Predominant-F0 Estimation for Detecting Melody and Bass Lines in Real-world Audio Signals, *Speech Communication*, Vol.43, No.4, pp.311–329 (2004).
- 3) Klapuri, A.: Multipitch Analysis of Polyphonic Music and Speech Signals Using an Auditory Model, *IEEE Trans. Audio, Speech and Lang. Process.*, Vol. 16, No. 2, pp.255–266 (2008).
- 4) Goto, M.: An Audio-based Real-time Beat Tracking System for Music with or without Drum-sounds, *J. New Music Res.*, Vol.30, No.2, pp.159–171 (2001).
- 5) Dixon, S.: Automatic Extraction of Tempo and Beat from Expressive Performances, *J. New Music Res.*, Vol.30, No.1, pp.39–58 (2001).
- 6) Kitahara, T.: Computational Musical Instrument Recognition and Its Application to Content-based Music Information Retrieval, PhD Thesis, Kyoto University (2007).
- 7) Yoshii, K., Goto, M. and Okuno, H.G.: Drum Sound Recognition for Polyphonic Audio Signals by Adaptation and Matching of Spectrogram Templates with Harmonic Structure Suppression, *IEEE Trans. Audio, Speech and Lang. Process.*, Vol.15, No.1, pp.333–345 (2007).
- 8) Kameoka, H., Nishimoto, T. and Sagayama, S.: A Multipitch Analyzer Based on Harmonic Temporal Structured Clustering, *IEEE Trans. Audio, Speech and Lang. Process.*, Vol. 15, No.3, pp.982–994 (2007).
- 9) 宮本賢一，亀岡弘和，西本卓也，小野順貴，嵯峨山茂樹：調波構造・時間包絡・音色の統合的クラスタリングによる楽音分析，*情処研報*，Vol.2007, No.71, pp.155–160 (2007).
- 10) Woodruff, J., Pardo, B. and Dannenberg, R.: Remixing Stereo Music with Score-informed Source Separation, *Proc. ISMIR*, pp.314–319 (2006).
- 11) Viste, H. and Evangelista, G.: A Method for Separation of Overlapping Partial Based on Similarity of Temporal Envelopes in Multichannel Mixtures, *IEEE Trans. Audio, Speech and Lang. Process.*, Vol.14, No.3, pp.1051–1061 (2006).
- 12) Casey, M. and Westner, A.: Separation of Mixed Audio Sources by Independent Subspace Analysis, *Proc. ICMC*, pp.154–161 (2000).
- 13) Itoyama, K., Goto, M., Komatani, K., Ogata, T. and Okuno, H.G.: Parameter Estimation for Harmonic and Inharmonic Models by Using Timbre Feature Distributions, *IPJS Journal*,

Vol.50, No.7 (2009).

- 14) Goto, M., Hashiguchi, H., Nishimura, T. and Oka, R.: RWC Music Database: Music Genre Database and Musical Instrument Sound Database, *Proc. ISMIR*, pp.229–230 (2003).  
15) 北原鉄朗, 後藤真孝, 奥乃 博: 音高による音色変化に着目した楽器音の音源同定: F0 依存多次元正規分布に基づく識別手法, 情報処理学会論文誌, Vol.44, No.10, pp. 2448–2458 (2003).

## 付録 テスト分布の更新式

ここでは, 潜在変数およびパラメータの最適テスト分布を求めるための具体的な更新式について述べる.

まず, 潜在変数の最適テスト分布の期待値部分を計算する.

$$s(l, H, m, n, t, f) = \Psi(\hat{\phi}_\alpha(l)) + \Psi(\hat{\phi}_\beta(l, H)) + \Psi(\hat{\phi}_\gamma(l, m, n)) \\ - \Psi(\hat{\phi}_\beta(l, H) + \hat{\phi}_\beta(l, I)) - \Psi\left(\sum_{m=0}^{M_H-1} \sum_{n=1}^{N_H} \hat{\phi}_\gamma(l, m, n)\right) \\ - \frac{\hat{\xi}_\tau(l)^{-1} + (\hat{\nu}_\tau(l) - t + m\rho_H)^2}{2\rho_H^2} - \frac{\log \hat{\zeta}(l) - \Psi(\hat{\eta}(l))}{2} - \frac{n^2}{2\hat{\xi}_\omega(l)} - \frac{(n\hat{\nu}_\omega(l) - f)^2 \hat{\eta}(l)}{2\hat{\zeta}(l)}$$

$$s(l, I, m, n, t, f) = \Psi(\hat{\phi}_\alpha(l)) + \Psi(\hat{\phi}_\beta(l, I)) + \Psi(\hat{\phi}_\delta(l, m, n)) \\ - \Psi(\hat{\phi}_\beta(l, H) + \hat{\phi}_\beta(l, I)) - \Psi\left(\sum_{m=0}^{M_I-1} \sum_{n=1}^{N_I} \hat{\phi}_\delta(l, m, n)\right) \\ - \frac{\hat{\xi}_\tau(l)^{-1} + (\hat{\nu}_\tau(l) - t + m\rho_I)^2}{2\rho_I^2} - \log(f + \kappa) - \log \log \lambda - \frac{(\mathcal{F}_I(f) - n)^2}{2}$$

$$\Psi(x) = \frac{d}{dx} \log \Gamma(x) \quad \hat{\xi}_\tau(l) = \frac{\hat{\xi}_H(l)}{\rho_H^2} + \frac{\hat{\xi}_I(l)}{\rho_I^2}$$

これらの指数関数を求め, 正規化することで最適テスト分布を求める.

$$q(z_\alpha(l) = 1, z_\beta(l, H) = 1, z_\gamma(l, m, n) = 1 | t, f) = \frac{\exp(s(l, H, m, n, t, f))}{\sum_l (\sum_{m,n} \exp(s(l, H, m, n, t, f)) + \sum_{m,n} \exp(s(l, I, m, n, t, f)))}$$

$$q(z_\alpha(l) = 1, z_\beta(l, I) = 1, z_\delta(l, m, n) = 1 | t, f) = \frac{\exp(s(l, I, m, n, t, f))}{\sum_l (\sum_{m,n} \exp(s(l, H, m, n, t, f)) + \sum_{m,n} \exp(s(l, I, m, n, t, f)))}$$

パラメータの事前分布を共役事前分布としているため, パラメータの最適テスト分布を求めることは, 目的関数を最大化するハイパーパラメータを求めることと等価である. 観測パワースペクトル  $X(t, f)$  と潜在変数の最適テスト分布との積を

$\bar{X}(l, H, m, n, t, f) = X(t, f) q(z_\alpha(l) = 1, z_\beta(l, H) = 1, z_\gamma(l, m, n) = 1 | t, f)$  などと表し, さらにこの積の適当な変数を積分して消去したものを

$$\bar{X}(l, H, m) = \sum_{n=1}^{N_H} \iint \bar{X}(l, H, m, n, t, f) dt df$$

などと表すことにして, 以下でハイパーパラメータの更新式を述べる.

$$\hat{\phi}_\alpha(l) = \tilde{\phi}_\alpha(l) + \bar{X}(l) \\ \hat{\phi}_\beta(l, H) = \tilde{\phi}_\beta(l, H) + \bar{X}(l, H) \quad \hat{\phi}_\beta(l, I) = \tilde{\phi}_\beta(l, I) + \bar{X}(l, I) \\ \hat{\phi}_\gamma(l, m, n) = \tilde{\phi}_\gamma(l, m, n) + \bar{X}(l, H, m, n) \quad \hat{\phi}_\delta(l, m, n) = \tilde{\phi}_\delta(l, m, n) + \bar{X}(l, I, m, n) \\ \hat{\nu}_\tau(l) = \frac{(\frac{\tilde{\xi}_H(l)}{\rho_H^2} + \frac{\tilde{\xi}_I(l)}{\rho_I^2}) \tilde{\nu}_\tau + \frac{\tilde{T}_H(l)}{\rho_H^2} + \frac{\tilde{T}_I(l)}{\rho_I^2}}{\frac{\tilde{\xi}_H(l)}{\rho_H^2} + \frac{\tilde{\xi}_I(l)}{\rho_I^2} + \frac{\bar{X}(l, H)}{\rho_H^2} + \frac{\bar{X}(l, I)}{\rho_I^2}} \\ \hat{\xi}_H(l) = \tilde{\xi}_H(l) + \bar{X}(l, H) \quad \hat{\xi}_I(l) = \tilde{\xi}_I(l) + \bar{X}(l, I) \\ \hat{\nu}_\omega(l) = \frac{\tilde{\xi}_\omega(l) \tilde{\nu}_\omega(l) + \tilde{F}_1(l)}{\tilde{\xi}_\omega(l) + \tilde{F}_0(l)} \quad \hat{\xi}_\omega(l) = \tilde{\xi}_\omega(l) + \tilde{F}_0(l) \\ \hat{\eta}_\sigma(l) = \tilde{\eta}_\sigma(l) + \frac{\bar{X}(l, H)}{2} \\ \hat{\zeta}_\sigma(l) = \tilde{\zeta}_\sigma(l) + \frac{\tilde{F}_2(l) - \frac{\tilde{F}_1(l)^2}{\tilde{F}_0(l)}}{2} + \frac{\tilde{\xi}_\omega(l) \tilde{F}_0(l) (\frac{\tilde{F}_1(l)}{\tilde{F}_0(l)} - \tilde{\nu}_\omega(l))^2}{2(\tilde{\xi}_\omega(l) + \tilde{F}_0(l))}$$

ただし, 補助関数を以下のように定める.

$$\tilde{T}_H(l) = \sum_{m=0}^{M_H-1} \int \bar{X}(l, H, m, t) (t - m\rho_H) dt \quad \tilde{T}_I(l) = \sum_{m=0}^{M_I-1} \int \bar{X}(l, I, m, t) (t - m\rho_I) dt \\ \tilde{F}_2(l) = \int \bar{X}(l, H, f) f^2 df \quad \tilde{F}_1(l) = \sum_{n=1}^{N_H} \int \bar{X}(l, H, n, f) n f df \\ \tilde{F}_0(l) = \sum_{n=1}^{N_H} \bar{X}(l, H, n) n^2$$