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*

☆落合和樹 [1], 亀岡弘和 [1,2], △中野允裕 [2], 嵯峨山茂樹 [1] ([1] 東大院・情報理工, [2]NTT CS 研)

1 はじめに

音楽音響信号からの自動採譜は，楽譜作成や MIDI 変換，音楽検索など様々な応用があり，音楽音響信号処理における重要な課題のひとつとして研究が進められている．自動採譜を目的とした従来の多重音解析手法は，音高推定とオンセット検出を目指したものがほとんどである．それは，自動採譜には多重音から音高，オンセットの取得だけでなく，拍構造（リズム）認識，テンポ推定，音価認識などの問題も内在するという理由から，楽譜を作り上げることは非常に難しいと考えられてきたためである．しかし，演奏者の意図や技量によってテンポやオンセット時刻が揺らいで演奏されていた場合でも，人間が典型的なリズムに当てはめてリズムを認識することができるように，楽譜を復元するためにはリズムの情報は必要不可欠である．そこで，本稿では，音響信号から音高とオンセットだけでなく，拍構造やテンポ，音価も一挙に推定し，直接的に楽譜を作り上げる方法論について論じる．

以前我々は，多重音音響信号から音高とオンセットを推定する手法 [1] とリズム・テンポ解析手法 [2] を後段に組み合わせた多段処理によって自動採譜を行う手法を提案した [3]．しかしながら，実際には音高推定とオンセット検出，リズム・拍構造推定の間には依存関係があるため，このような多段処理による自動採譜ではその性能に限界があると考えられる．例えば，各音のオンセット時刻が予めわかっていれば，音高推定や拍構造推定性能を向上でき，その逆も然りである．このことから，音高推定とオンセット検出，リズム・拍構造推定を同時に行う枠組みが必要であると考えられる．そこで，本稿では，確率文脈自由文法をヒントにしたリズム文法モデルによって確率的に楽譜が生成され，その楽譜から人間の演奏によって音楽スペクトログラムが生成されるプロセスを階層ベイズモデルに基づき記述する．そして，その逆問題として音高，オンセット，拍構造，音価の推論を同時に行うアルゴリズムを導出し，音楽音響信号に適用してその解析性能について検討する．

2 音楽生成プロセスの階層ベイズモデル

本稿では，リズム文法に基づいて楽譜が生成され，その楽譜に描かれた曲が演奏されて観測スペクトログラムが生成される，という二段のプロセスで音楽生成プロセスをモデル化する．以下，楽譜とスペクトログラムそれぞれの生成モデルについて述べる．

2.1 音楽スペクトログラムの生成モデル

本節で述べる音楽スペクトログラム生成モデルは我々が提案した [4] に基づいている．

まずはじめに，音楽スペクトログラムを構成する各音のスペクトルやエネルギーに以下を仮定する．

1. 楽曲に登場する音高数は限られており，そのスペクトルは定常である
2. 音楽は様々な音高や音長をもつ単音（単音モデル）の組み合わせで表現される

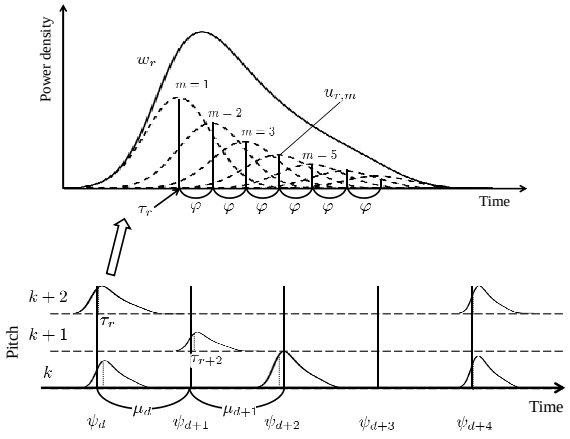


Fig. 1 正規分布の混合数を $M = 7$ としたときの単音のパワーエンベロープモデル $W_{r,t}$（上）と単音モデルが配置された拍構造モデル（下）．各単音は正規分布の配置間隔をその標準偏差に設定することで滑らかな時間変化をもつパワーを実現しており，それらが拍時刻に依存して配置される．

3. 各音のパワーはオンセットからオフセットまで滑らかに連続的な形状を持つ

まず，仮定 1 に関して，例えばピアノ曲なら 88 音の中からいくつかの音が選ばれて演奏されるように，楽曲に登場する音高数は数限られている．観測信号を時間周波数領域に変換したパワースペクトログラムを $Y_{\omega,t}$ とする．ω, t はそれぞれ周波数，時間のインデックスである．音高 k のスペクトル $H_{\omega,k}$ が時間に依らず一定だとし，そのエネルギー $U_{k,t}$ のみが時間変化すると考えると，モデルが生成するスペクトログラムは，$X_{\omega,t} = \sum_k H_{\omega,k} U_{k,t}$ となり，観測 $\boldsymbol{Y}$ がモデル $\boldsymbol{X}$ から Poisson 分布に従って得られるとすると，

$$Y_{\omega,t} \sim \mathrm{Poisson}\left(Y_{\omega,t}; X_{\omega,t}\right) \tag{1}$$

で与えられる．この下での $\boldsymbol{H}$ と $\boldsymbol{U}$ の最尤推定は，$\boldsymbol{Y}$ と $\boldsymbol{HU}$ の I ダイバージェンスを最小化する問題と等価である．

次に，仮定 2 から，r 番目の単音モデルが時刻 t に持つパワーを $W_{r,t}$ とすると，その時刻での音高 k のパワー $U_{k,t}$ は

$$U_{k,t} = \sum_{r=1}^{R} \delta_{\kappa_r,k} W_{r,t} \tag{2}$$

と表すことができる．ここで，κ_r は r 番目の単音モデルが対応している音高のインデックスで，$\delta_{x,k}$ は Kronecker の Delta 関数である．

さらに，仮定 3 に関して，実際の楽器音のパワーはオンセットの瞬間こそ急峻に大きくなることもあるが，時間とともに滑らかに変化していくことが多い．これを表現するモデルとして，各音のパワーの時間変化を，拘束つき混合正規分布型の関数 [1] によりモデル化する．これにより，

$$W_{r,t} = \sum_{m=1}^{M} G_{r,m,t} \tag{3}$$

---

*Hierarchical Bayesian modeling of the generating process of music signals for automatic transcription by OCHIAI Kazuki[1], KAMEOKA Hirokazu[1,2], NAKANO Masahiro[2], SAGAYAMA Shigeki[1] ([1]The University of Tokyo, [2]NTT CS Lab)

$$= \sum_{m=1}^{M} \frac{w_r u_{r,m}}{\sqrt{2\pi}\varphi} e^{-\{t-(m-1)\varphi-\tau_r\}^2/2\varphi^2} \quad (4)$$

と表現できる（Fig. 1 上側）．ここで，w_r は r 番目の単音モデルが持つ総エネルギーで，τ_r は単音モデルを構成する 1 個目の正規分布であり，推定オンセット時刻とみなせる．各正規分布の中心は，全ての単音モデルで等しい値を持つ正規分布の標準偏差 φ だけ離れているものとする．また，$u_{r,1}, \ldots, u_{r,M}$ は M 個の正規分布の重みに相当し，単音モデルの形状（パワーの時間変化）を決定しており，その大小や個数によって音長も制御できる．エネルギー項 w_r と正規分布の重み $u_{r,m}$ のスケールの任意性を避けるために，全ての r で $\sum_{m=1}^{M} u_{r,m} = 1$ が成り立つものとする．このとき，各音の音長は全て等しいわけではなく，楽曲によって長いものから短いものまで様々存在しているので，正規分布の適切な個数 M をデータから推論できるようにするのが望ましい．$u_{r,1}, \ldots, u_{r,M}$ は確率ではないが，足して 1 となる変数である．したがって，$u_{r,1}, \ldots, u_{r,M}$ の生成プロセスとして，Dirichlet 過程の構成法のひとつである Stick-Breaking 過程を形式的に採用することが可能である．すなわち，Beta 分布に従って独立に生成される M 個のパラメータ $V_{r,1}, \ldots, V_{r,M}$ を用いて，$u_{r,m}$ を $u_{r,m} = V_{r,m} \prod_{m'=1}^{m-1}(1 - V_{r,m'}), V_{r,m} \sim \mathrm{Beta}(V_{r,m}; 1, \beta_{r,m}^{V})$ と設定する．これにより，m が大きくなるにつれて $u_{r,m}$ が小さくなる傾向を持たせることができるので，楽器音が発音後徐々に小さくなる様子も表現できる．

そのほか，楽曲に登場する音符数は数限られているので，必要最低限の単音モデルだけエネルギーが大きくなることが望ましい．そこで，Gamma 過程の考え方に基づいて $w_r \sim \mathrm{Gamma}(w_r; \alpha_r^w, \beta_r^w)$ とすることでスパース化効果が得られる．

また，各音高を表す基底スペクトルが 1 音高に対応することが望ましい．そこで，$H_{\omega,k} = \mathrm{Gamma}(H_{\omega,k}; \beta_{\omega,k}^{H} \bar{H}_{\omega,k} + 1, \beta_{\omega,k}^{H})$ とすることで，スペクトルが調波構造を保たせることができる．$\bar{H}_{\omega,k}$ は $H_{\omega,k}$ がその音高を表現するために取るべき値であり，Gamma 分布の最頻値となる．

さて，楽譜として記述可能な音楽というものは，規則正しい拍構造を持っており，拍位置に基づいて各音の発音するタイミングが決定される．一般に，各音の発音されるべき時刻の間隔は拍間隔の整数倍か整数分の 1 である．しかし実際に人間により楽曲が演奏される場合，演奏者の技量や意識的な演奏表情付けに伴いテンポが揺らいだり，オンセット時刻が拍位置や拍間隔の整数分の 1 の時刻からずれたりすることがある．このことを考慮すると，楽譜に記述された楽曲が演奏されて観測スペクトログラムが生成されるプロセスに関し，以下の仮定を置くことができる．

4. 各音のオンセットは拍位置か拍位置から拍間隔の整数分の 1 倍離れた時刻に存在する
5. テンポ（拍間隔）は時々刻々緩やかに変化する

仮定 4 に関して，まず，曲が構成されている音符の中で最も短い音価単位で拍を細分化した時刻を考える．最も短い音価は曲によって異なるが，例えば 32 分音符を最も短い音価だとすると，4 分の 4 拍子の曲では拍間隔を 8 分割することに相当する．この細分化された拍の時刻を $\boldsymbol{\psi} = \{\psi_d\}_{1 \le d \le D}$ とすると，全ての音は，ψ の中のいずれかの時刻で発音を開始すると仮定していることになる．D は細分化された拍の総数である．このとき，演奏表情や演奏者の技量により，同時に複数音を鳴らした時をはじめ，必ずしも全ての音が細分化された拍時刻に正確に演奏されるわけではない．そこで，r 番目の単音のオンセット時刻を，ψ_{S_r} を中心とした正規分布

$$p(\boldsymbol{\tau}|\boldsymbol{\psi}, \boldsymbol{S}) = \prod_r \mathcal{N}\left(\tau_r; \psi_{S_r}, (\sigma^\tau)^2\right) \quad (5)$$

から生成されたと仮定する．加えて，仮定 5 を考慮すると，拍間隔の時々刻々の変化はテンポの変化によって揺らぐものであるから，テンポも拍間隔も急峻に変化することなく緩やかに変化するものと仮定する．細分化された拍時刻のテンポに対する揺らぎを正規分布で，d 番目の細分化された拍時刻での局所的なテンポ μ_d の変化を正規分布による Markov 連鎖で仮定すると，

$$p(\boldsymbol{\psi}|\boldsymbol{\mu}) = \prod_{d=2}^{D} \mathcal{N}\left(\psi_d; \psi_{d-1} + \mu_{d-1}, (\sigma^\psi)^2\right) \quad (6)$$

$$p(\boldsymbol{\mu}) = \prod_{d=2}^{D} \mathcal{N}\left(\mu_d; \mu_{d-1}, (\sigma^\mu)^2\right) \quad (7)$$

とできる．この拍構造を Fig. 1 下側に示す．

全ての非終端記号の要素に適用する生成規則：
- $\phi^T \sim \mathrm{Beta}(\phi^T; 1, \beta^T)$
  （終端記号出力か分割か）
- $\phi^N \sim \mathrm{Beta}(\phi^N; 1, \beta^N)$
  （分割が時間方向か音高方向か）
- $\phi_{l,l'}^B \sim \mathrm{Dirichlet}(\boldsymbol{\phi}_l^B; 1, \beta^B)$
  （どの拍位置で分割するか）

導出木の全てのノードに適用する生成規則：
- $b_n \sim \mathrm{Categorical}(b_n; \phi^T)$
  （出力か分割かを選択）
- If $b_n = $ (EMISSION)
  - $S_r \sim \delta_{S_r,S_n}, \quad L_r \sim \delta_{L_r,L_n}$
    （終端記号を出力）
- If $b_n = $ (BINARY-PRODUCTION)
  - If $\rho_n = $ (SYNCHRONIZATION)
    - $S_{n_1} \sim \delta_{S_{n_1},S_n}, \quad S_{n_2} \sim \delta_{S_{n_2},S_n}$
    - $L_{n_1} \sim \delta_{L_{n_1},L_n}, \quad L_{n_2} \sim \delta_{L_{n_2},L_n}$
      （親ノードを複製）
  - If $\rho_n = $ (TIME-SPANNING)
    - $S_{n_1} \sim \delta_{S_{n_1},S_n}, \quad S_{n_2} \sim \delta_{S_{n_2},S_n+L_{n_1}}$
    - $L_{n_1} \sim \delta_{L_{n_1},L_n-L_{n_1}}$
    - $L_{n_2} \sim \mathrm{Categorical}(L_{n_2}; \boldsymbol{\phi}_{L_n}^B)$
      （時間分割をして子ノードを生成）

Fig. 2　提案するリズム生成文法

2.2　リズム文法に基づく楽譜生成モデル

前節で述べたような r 個のオンセット系列 $\boldsymbol{\tau} = \{\tau_r\}_{1 \le r \le R}$ と D 個の細分化された拍時刻 $\boldsymbol{\psi} = \{\psi_d\}_{1 \le d \le D}$ が得られたときに，楽譜を復元する問題は，各音がどの拍 S_r をターゲットにし，どのような音長 L_r を意図して演奏されていたかを推測する問題である．これは，どのようなリズムが楽譜に使われていたかを推測する問題と等価である．

楽譜中の音符間の構造には，時間方向に連続した構造と音高方向に連続した構造がある．このような構造が様々な形で階層的に組み合わさることで楽譜が生成されていると考えられる．そこで，このような構造をモデル化でき，かつ未知パターンにも柔軟なものとして，確率文脈自由文法が適していると考えられる．

本稿で提案するリズム文法は，確率文脈自由文法をヒントに拡張したモデルである．楽譜を生成する

リズム文法を図 2 に示す．このリズム文法に従って楽譜が生成される流れを述べる．はじめに，細分化された総拍数を D とし，曲全体を 1 拍目で発音し音長が D の音だとみなす．これをリズム文法に従って，時間方向に連続した構造や音高方向に連続した構造（非終端記号）に繰り返し分割して生成していく．そして．分割された要素から，オンセット時刻に相当する細分化された拍のインデックス S_r と音長（音価に相当）L_r （終端記号）を出力することによって演奏されている各音の情報となる終端記号列が得られる．導出木のノード n における非終端記号は，オンセットの拍のインデックス S_n と音長 L_n の情報を持っており，この非終端記号に対して毎回の分割を行う際，

1. 演奏されている各音の拍位置と音長を出力するか分割をするか
2. 分割する場合，時間方向に分割するか音高方向に複製するか
3. 時間方向に分割する場合，どの細分化された拍位置で分割するか
4. 終端記号を出力する場合，どの音高の音か

という選択をする必要がある．

まず，終端記号を出力（EMISSION）するか非終端記号に分割（BINARY-PRODUCTION）するかを Bernoulli 分布に従う生成規則パラメータ b_n で決定する．b_n の選択確率 ϕ^T は Beta 分布に従って決定される．

次に，非終端記号への分割が選ばれたときに，$n(S_n, L_n) \rightarrow n_1(S_{n_1}, L_{n_1})n_2(S_{n_2}, L_{n_2})$ というように分割するとする．時間方向に分割（TIME-SPANNING）するか音高方向に分割（SYNCHRONIZATION）するかを Bernoulli 分布に従う生成規則パラメータ ρ_n で決定する．ρ_n の選択確率 ϕ^N も Beta 分布に従って決定される．

音高方向への複製が選ばれたとき，親ノード n のオンセットの拍位置 S_n と音長 L_n を子ノード n_1, n_2 へ複製する．この複製の規則は従来の確率文脈自由文法では扱っていないものであり，複旋律の曲も解析できるためのものである [5]．前節のスペクトログラム生成モデルにより，[5] では扱えなかったテンポ変化にも本モデルでは対応可能である．

さて，時間方向への分割が選ばれたとき，どの細分化された拍数 L_1 で分割する生成規則が使われるかは，離散分布に従って決定する．離散分布は，$\mathrm{Categorical}(k_i; \boldsymbol{p}) = p_i$ である．生成規則の適用確率 $\boldsymbol{\phi}^B$ は Dirichlet 分布に従って決定される．このとき，時間方向に分割されたノード n_1, n_2 は，n_1 が時間的に早く鳴り始める音に対応しているとすると，n_1 のオンセット位置 S_{n_1} は n のオンセット位置 S_n に等しく，n_1 と n_2 の音長の和 $L_{n_1} + L_{n_1}$ は n の音長 L_n に等しくなるように n_1, n_2 を生成する．

最後に，終端記号の出力が選ばれた場合，拍のインデックス S_n と音長 L_n を，曲全体で r 番目の音符として出力し，さらに，その音高 κ_r を離散分布 $p(\kappa_r) = \mathrm{Categorical}(\kappa_r; \phi_r^K)$ に従って選択する．$\boldsymbol{\phi}^K$ は Dirichlet 分布 $p(\boldsymbol{\phi}^K) = \prod_r \mathrm{Dirichlet}(\boldsymbol{\phi}_r^K; \boldsymbol{\alpha}_r^K)$ に従って与えることができる．音高を決定する $\boldsymbol{\kappa}$ も非終端記号 n の要素に加えるモデルも考えられるが，その場合は本稿で扱っていない調や和声の情報も考える必要があるので，本稿における $\boldsymbol{\kappa}$ は，リズム生成モデルとは別個に扱うことにする．

このようにして楽譜の導出木を生成することができるが，本稿で求めたいものは，導出木の形そのものではなく，終端記号としてどのような音符系列がどのような確率で出力されたか，であるので，推論すべき分布は，$\boldsymbol{b}$ と $\boldsymbol{\rho}$ に関して導出木の生成確率 $p(\boldsymbol{S}, \boldsymbol{L}|\boldsymbol{b}, \boldsymbol{\rho}, \boldsymbol{\Phi})$ を周辺化した，

$$p(\boldsymbol{S}, \boldsymbol{L}|\boldsymbol{\Phi}) = \sum_{\boldsymbol{b}, \boldsymbol{\rho}} p(\boldsymbol{S}, \boldsymbol{L}|\boldsymbol{b}, \boldsymbol{\rho}, \boldsymbol{\Phi}) p(\boldsymbol{b}|\phi^T) p(\boldsymbol{\rho}|\phi^N) \tag{8}$$

である．ここで，$\boldsymbol{\Phi} = \{\boldsymbol{\phi}^B, \phi^T, \phi^N\}$ とした．

3 パラメータ推論アルゴリズム

観測スペクトログラム $\boldsymbol{Y}$ が得られたときの各パラメータの集合 θ に関する事後分布 $p(\theta|\boldsymbol{Y})$ を求めるのが今回の問題である．一般に，事後分布を求めるための積分計算は処理が困難なので，近似した事後分布 $q(\theta)$ を推論する変分ベイズ法を用いる．

パラメータ推論の前段階として，後の変分事後分布計算を簡略化するため，観測スペクトログラムを構成する潜在変数 $C_{r,m,\omega,t}$ を用いて，

$$Y_{\omega,t} \sim \delta\left(Y_{\omega,t} - \sum_{r,m} C_{r,m,\omega,t}\right) \tag{9}$$

$$C_{r,m,\omega,t} \sim \mathrm{Poisson}\left(C_{r,m,\omega,t}; H_{\omega,\kappa_r} G_{r,m,\omega,t}\right) \tag{10}$$

とする．この $\delta(\cdot)$ は Dirac の Delta 関数である．これは，Poisson 分布の再生性より，式 (1) の関係性は失われていない．このとき，パラメータの事後分布は，

$$\begin{aligned}
&p(\boldsymbol{H}, \boldsymbol{w}, \boldsymbol{V}, \boldsymbol{\tau}, \boldsymbol{\kappa}, \boldsymbol{\psi}, \boldsymbol{\mu}, \boldsymbol{S}, \boldsymbol{L}, \boldsymbol{\phi}^B, \phi^T, \phi^N, \boldsymbol{\phi}^K|\boldsymbol{Y}) \\
&\propto p(\boldsymbol{Y}|\boldsymbol{C}) p(\boldsymbol{C}|\boldsymbol{H}, \boldsymbol{w}, \boldsymbol{V}, \boldsymbol{\tau}, \boldsymbol{\kappa}) p(\boldsymbol{H}) p(\boldsymbol{V}) p(\boldsymbol{w}) \\
&\quad \cdot p(\boldsymbol{\tau}|\boldsymbol{\psi}, \boldsymbol{S}) p(\boldsymbol{\psi}|\boldsymbol{\mu}) p(\boldsymbol{\mu}) p(\boldsymbol{\kappa}|\boldsymbol{\phi}^{\boldsymbol{K}}) p(\boldsymbol{\phi}^{\boldsymbol{K}}) \\
&\quad \cdot p(\boldsymbol{S}, \boldsymbol{L}|\boldsymbol{\phi}^B, \phi^T, \phi^N) p(\boldsymbol{\phi}^B) p(\phi^T) p(\phi^N) \qquad (11) \\
&= p(\boldsymbol{Y}, \theta) \qquad (12)
\end{aligned}$$

のようになる．

3.1 変分事後分布

変分ベイズ法では，事後分布 $p(\theta|\boldsymbol{Y})$ を近似する $q(\theta)$ 求めることが目的である．$p(\theta|Y)$ と $q(\theta)$ の KL ダイバージェンスを最小化する問題は，Jensen の不等式に基づき立てられる対数周辺尤度の下界

$$\mathcal{B}[q] \equiv \sum_{\boldsymbol{\kappa}, \boldsymbol{S}, \boldsymbol{L}} \int q(\theta) \log \frac{p(\boldsymbol{Y}, \theta)}{q(\theta)} \mathrm{d}\boldsymbol{C} \mathrm{d}\boldsymbol{H} \mathrm{d}\boldsymbol{U} \tag{13}$$

を最大化するような $q(\theta)$ を求めることと同じになる．ここで，各パラメータが独立であるとし，

$$\begin{aligned}
q(\theta) = &q(\boldsymbol{C}) q(\boldsymbol{H}) q(\boldsymbol{w}) q(\boldsymbol{V}) q(\boldsymbol{\tau}, \boldsymbol{\psi}, \boldsymbol{\mu}) q(\boldsymbol{\kappa}) q(\boldsymbol{\phi}^K) \\
&\cdot q(\boldsymbol{S}, \boldsymbol{L}) q(\boldsymbol{\phi}^{\boldsymbol{B}}) q(\phi^T) q(\phi^N)
\end{aligned} \tag{14}$$

として各 $q(\cdot)$ の更新する．これは平均場近似と呼ばれ，変分ベイズ法では良く用いられる仮定である．一般に，$B[q]$ を最大にする q は，

$$q(Z_i) \propto \exp \left\langle \log p(\boldsymbol{Y}, \boldsymbol{Z}) \right\rangle_{q(/Z_i)} \tag{15}$$

で与えられる．ここで，$\langle \cdot \rangle_{q(/Z)}$ は Z 以外の変数で期待値を取るという意味とする．

本稿で提案するモデルの各パラメータの変分事後分布は，自然共役事前分布を設定したことにより，

$$q(\boldsymbol{C}_{\omega,t}) = \mathrm{Multinomial}(\boldsymbol{C}_{\omega,t}; Y_{\omega,t}, \boldsymbol{f}_{\omega,t}^C) \tag{16}$$

$$q(H_{\omega,k}) = \mathrm{Gamma}(H_{\omega,k}; \xi_{\omega,k}^H, \zeta_{\omega,k}^H) \tag{17}$$

$$q(w_r) = \mathrm{Gamma}(w_r; \xi^w_r, \zeta^w_r) \tag{18}$$
$$q(V_{r,m}) = \mathrm{Beta}(V_{r,m}; \xi^V_{r,m}, \zeta^V_{r,m}) \tag{19}$$
$$q(\boldsymbol{\tau}, \boldsymbol{\psi}, \boldsymbol{\mu}) = \mathcal{N}(\boldsymbol{\chi}; \boldsymbol{\xi}^\chi, \boldsymbol{\zeta}^\chi) \tag{20}$$
$$q(\kappa_r) = \mathrm{Categorical}(\kappa_r; \boldsymbol{f}^\kappa_r) \tag{21}$$
$$q(\boldsymbol{\phi}^K_r) = \mathrm{Dirichlet}(\boldsymbol{\phi}^K_r; \boldsymbol{\xi}^K_r) \tag{22}$$
$$q(S_r, L_r) = \mathrm{Categorical}(S_r, L_r; \boldsymbol{f}^{SL}_r) \tag{23}$$
$$q(\boldsymbol{\phi}^B_l) = \mathrm{Dirichlet}(\boldsymbol{\phi}^B_l; \boldsymbol{\xi}^B_l) \tag{24}$$
$$q(\phi^T) = \mathrm{Beta}(\phi^T; \xi^T, \zeta^T) \tag{25}$$
$$q(\phi^N) = \mathrm{Beta}(\phi^N; \xi^N, \zeta^N) \tag{26}$$

の形となる．ここで，$\boldsymbol{C}_{\omega,t}, \boldsymbol{\phi}^K_r, \boldsymbol{\phi}^B_l$ はそれぞれ添え字に書かれていない各要素を並べたベクトルで，

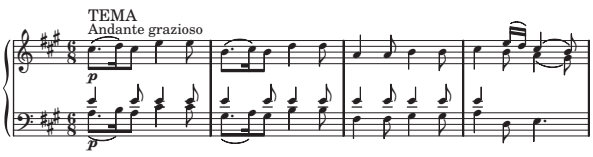
$$\boldsymbol{\chi} = \begin{bmatrix} \boldsymbol{\tau} \\ \boldsymbol{\psi} \\ \boldsymbol{\mu} \end{bmatrix}, \boldsymbol{\xi}^\chi = \begin{bmatrix} \boldsymbol{\eta}^\tau \\ \boldsymbol{\eta}^\psi \\ \boldsymbol{\eta}^\mu \end{bmatrix}, \boldsymbol{\zeta}^\chi = \begin{bmatrix} \boldsymbol{\nu}^\tau & \boldsymbol{\nu}^{\tau\psi} & \boldsymbol{\nu}^{\tau\mu} \\ \boldsymbol{\nu}^{\tau\psi} & \boldsymbol{\nu}^\psi & \boldsymbol{\nu}^{\psi\mu} \\ \boldsymbol{\nu}^{\tau\mu} & \boldsymbol{\nu}^{\psi\mu} & \boldsymbol{\nu}^\mu \end{bmatrix} \tag{27}$$

である．式 (24) から (26) は，各生成規則を更新する場合のみ用い，式 (23) から (26) は Inside-Outside アルゴリズムによって導くことができる．各変分事後分布を求めることは，各変分事後分布のパラメータを求めることと等価である．

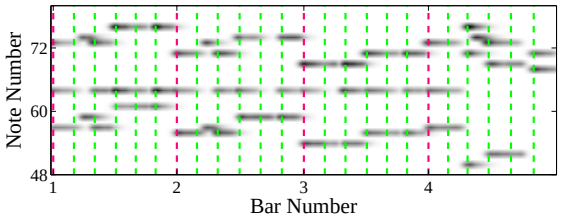
詳しい導出や更新式は紙面の都合上省略するが，式 (15) に従って各パラメータを求めることで得られる．

4 拍位置検出動作確認実験

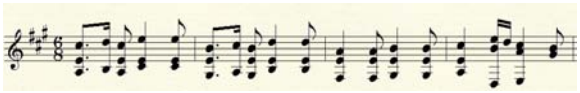
提案手法の拍構造解析性能を評価するために，音符数と各音のオンセット時刻が既知という条件下で音楽音響信号を用いて実験を行った．これは，オンセット時刻や音符数が未知の状態で非常に多くの単音モデルを配置をし，パラメータ推論にあたりエネルギーの小さい単音モデルを無音とみなしつつ更新していき，最終的に必要最低限の単音モデルのみが残ったという状況を想定している．実験には RWC クラシック音楽データベース・ジャズ音楽データベース [6] からピアノ曲（RWC-MDB-C-2001 No. 26, 27, 30）を選び，空間計算量の都合上，各曲最初の数小節分（約 10 s 程度）をモノラル信号にミックスダウンして 16 kHz にしたものを用いた．解析には Wavelet 変換されたスペクトログラムを用い，その条件は，時間分解能 16 ms，最低周波数 30 Hz，周波数分解能 12 cent とした．モデルの各事前分布のパラメータや，求めるパラメータの初期値は，$K = 74, M = 40, \varphi = 3, \alpha^H_{\omega,k} = \beta^H_{\omega,k} \bar{H}_{\omega,k} + 1, \beta^H_{\omega,k} = 500, \alpha^w_r = \beta^w_r = 0, \beta^V_{r,m} = 10e^{-m/8} / \sum_{m'} e^{-m'/8}, \sigma^\tau = 2, \sigma^\psi = 1, \sigma^\mu = 0.5, \alpha_{r,k} = 2, \beta^T = 1, \beta^N = 2$ とした．$\bar{\boldsymbol{H}}$ は基底スペクトル $\boldsymbol{H}$ の初期値であり，RWC 楽器音データベース [7] のピアノの単音データに対して，非負値行列分解を適用して求めたものとした．拍子やテンポ（総拍数）は採譜アプリケーションを想定し，ユーザが作りたい楽譜の様式に合わせて選択できるときの利便性を考慮し，拍子と小節数，最も短い音価を入力できるようにした．本実験では拍子と小節数は正解と同じものを入力し，最も短い音価は 16 分音符とした．拍子決定後に N 分の M 拍子の N に対応する拍の長さのみ $\alpha^B_{l,l'} = 200$ とし，それ以外の細分化された拍に関しては $\alpha^B_{l,l'} = 2$ とした．拍時刻や拍の時間間隔の初期値に関しては，動的計画法によるビートトラッキング手法 [8] を利用して推定したものに対しテンポを考慮して細分化した値を用いた．パラメータ更新の反復回数は 10 回とした．解析後の各分布からピアノロールや楽譜を作成する際には，連続分布は期待値を，離散分布は確率最大のインデックスを選択した．



(a) 正解楽譜



(b) 基本周波数成分のみを抽出したスペクトログラムのピアノロール表示



(c) 提案手法により推定した楽譜

Fig. 3 Morzart: Piano Sonata No. 11 in A major, K. 331/300i の正解楽譜 (a) と基本周波数成分のみを抽出したスペクトログラムをピアノロール表示したもの (b) および提案手法により推定した楽譜 (c)．(b) の赤い破線は小節線であり，緑の破線は拍時刻である．

Mozart の Sonata (No. 26) を解析して，ピアノロール表示したものと楽譜化したものを Fig. 3 に示す．いずれの曲においても，一部の音価こそ正解とは異なる結果も得られたが，装飾音を除きオンセットの拍位置は正しく推定できた．また，どのデータにおいても，いくつかのオンセット時刻が推定された拍位置と異なる拍の時刻に近いという推定結果が得られていた．これは，量子化による採譜では誤った拍位置にオンセットがあると推定されてしまうところを，リズム文法を取り入れることによって正しい拍位置に推定できたと考えられる．

5 おわりに

本稿では，音楽スペクトログラム生成モデルと楽譜生成モデルを統合した音楽生成プロセスを階層ベイズモデルに基づいて記述することにより，音楽音響信号から拍節構造を踏まえて自動採譜をする手法を提案した．オンセット時刻既知という条件下での動作実験により，各オンセットが所属する拍位置がリズム文法を導入したことにより正しく推定できることを確認した．音価を正しく推定するために，生成文法適用確率のハイパーパラメータの学習や，単音モデルを構成する正規分布の重みやエネルギーと音長パラメータの対応関係も考慮したモデル化が今後の課題である．

参考文献

[1] H. Kameoka *et al.*, *IEEE Trans. on ASLP*, 15, 982–994, 2007.
[2] H. Takeda *et al.*, *Proc. ICASSP*, 4, 1317–1320, 2007.
[3] K. Miyamoto *et al.*, *Proc. ICASSP*, 2, 697–700, 2007.
[4] K. Ochiai *et al.*, *Proc. ICASSP*, 2012. to appear.
[5] M. Nakano *et al.*, *Proc. ICASSP*, 2012. to appear.
[6] M. Goto *et al.*, *Proc. ISMIR*, 287–288, 2002.
[7] M. Goto, *Proc. ICA*, I–553–556, 2004.
[8] D. P. W. Ellis, *Journal of New Music Research*, 36, 51–60, 2007.