Vol.2011-MUS-91 No.6
2011/7/27

スペクトログラムのベイジアンノンパラメトリックモデリングに基づく音楽信号の解析

中野 允裕[†1] ルルー ジョナトン[†2]
亀岡 弘和[†2]
中村 友彦[†1] 小野 順貴[†1] 嵯峨山 茂樹[†1]

本報告では, 音楽信号のような多重音を解析するための手法として, Bayesian nonparametrics に基づく音響信号スペクトログラムのモデル化方法を提案し, その構成法と推論について議論する. 近年, 非負値行列分解に代表されるようなスパース表現基づく音楽信号のモデル化が盛んに研究されている. その中で解決すべき二つの問題が注目を集めている. 一つ目は楽器音が時間変化する多様なスペクトルを持つ点であり, もう一点は観測信号中に含まれる音源の数が一般的には未知なことである. さらに, 楽器音の多様なスペクトルは音源数の推定を困難にし, また逆に音源数が未知であることによって一音一音がどの程度多様なスペクトルを持つか推定することを困難にしている. 本報告では, これら二つの課題を同時に解消するために, 信号の重畳を表す非負値行列分解型のスパース表現と時系列パターンを表現する隠れマルコフモデルを Bayesian nonparametrics 上で融合させたスペクトログラムモデルを提案する.

Bayesian nonparametric spectrogram modeling for music signal analysis

Masahiro Nakano,[†1] Jonathan Le Roux ,[†2]
Hirokazu Kameoka,[†2] Tomohiko Nakamura,[†1]
Nobutaka Ono[†1] and Shigeki Sagayama[†1]

This paper presents a Bayesian nonparametric latent source discovery method for music signal analysis. Recently, the use of latent variable decompositions, especially nonnegative matrix factorization (NMF), has been a very active area of research. These methods are facing two, mutually dependent, problems: first, instrument sounds often exhibit time-varying spectra, and grasping this time-varying nature is an important factor to characterize the diversity of each instrument; moreover, in many cases we do not know in advance the number of sources. Conventional decompositions generally fail to cope with these issues as they suffer from the difficulties of automatically determining the number of sources and automatically grouping spectra into single events. We address both these problems by developing a Bayesian nonparametric fusion of NMF and hidden Markov model (HMM).

1. はじめに

多重音の中から楽器音 1 音 1 音の音程や音色, 発音時刻を特定する技術は, 音楽信号処理分野における中心的な課題の一つであり総じて多重音解析と呼ばれる. 多重音解析は音楽信号の自動採譜や楽器音分離, 音楽加工への応用が期待される極めて重要な技術であり, その確立へ向け多くの研究が行われてきている. 従来, さまざまな観点から多重音解析の試みがなされてきたが, 特に近年ではスパース表現の考え方を利用して音楽信号をスペクトログラムの領域でモデル化しようとする研究が注目を集めている.

スパース表現に基づく音楽信号の取り扱いにおいて, 代表的な手法として非負値行列分解 (NMF: Nonnegative matrix factorization) が挙げられる[1)]. NMF に基づく音楽信号解析は一般に, 音楽信号のスペクトログラムを非負値行列に見立て, それを低ランクの非負値行列の積に分解することによって行われる. これはスペクトログラムを少数の頻出のスペクトルパターンとその音量変化によって近似しようとするものである. 音楽信号においては各楽器音は曲中で繰り返し演奏されることが多いため, 頻出のスペクトルパターンは自ずと各楽器音の平均的なスペクトルとして表出することが期待される.

NMF による音楽信号の解析は非常に活発に研究されているが, 二つの大きな問題が解決すべき課題として挙げられてきた. 一つ目は, 楽器音が本来もつ時間変化するスペクトルをどのように表現するかということである[2)–4)]. 例えばピアノの場合, 発音からアタック, ディケイ, サステイン, リリースといった音色変化を経て消音すると一般的に捉えられている. また, 歌声や弦楽器は, ヴィブラートのように基本周波数を変化させることで演奏に表情付けを行っている. このように楽器音の時間変化する音色は音楽に彩りを与える重要な要素であ

†1 東京大学情報理工学系研究科
Graduate School of Information Science and Technology, The University of Tokyo
†2 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
NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation

るが, 通常の NMF に基づく音楽信号のモデル化においては楽器音は定常な一つのスペクトルにより表されるという仮定が置かれており本来の多様な音色変化を表現出来ていない. もう一つの問題は, 入力信号中に含まれる楽器音の数が事前に分からない点にある[5),6)]. NMF は入力の行列を低ランクで近似しようとするものであるが, そのランクは分解の結果に重大な影響を及ぼす. 分解された一つ一つの要素を楽器一音ごとに対応させるためには適切なランク数の設定が必要であり, そのようなランク数を入力信号毎に見つけるのは容易ではない.

さらに, 以上の二つの問題は相互に依存しあうことで扱いをより難しくしている. 楽器音の非定常なスペクトルは適切な楽器音数の推定を困難にし, また, 未知の楽器音数が多様なスペクトルを一楽器音に割り当てることを難しくしている. そこで本研究では, この二つの問題を同時に解消するようなスペクトログラムのモデル化手法を提案する.

2. 非負値行列分解型のスパース表現に基づくスペクトログラムのモデル化

2.1 Nonnegative matrix factorization (NMF)

NMF に基づく音響信号解析は一般に, 振幅スペクトログラムもしくはパワースペクトログラム $\boldsymbol{Y} = (Y_{\omega,t})_{\Omega\times T} \in \mathbb{R}^{\geq 0,\Omega\times T}$ (ただし, $\omega = 1,\cdots,\Omega$ は周波数インデックス, $t = 1,\cdots,T$ は時間インデックスを表す) が基底 $\boldsymbol{H} = (H_{\omega,d})_{\Omega\times D} \in \mathbb{R}^{\geq 0,\Omega\times D}$ とアクティベーション $\boldsymbol{U} = (U_{d,t})_{D\times T} \in \mathbb{R}^{\geq 0,D\times T}$ の積で表現できるという仮定に基づいている. これはすなわち, $Y_{\omega,t} \approx \sum_d H_{\omega,d}U_{d,t}$, のように観測スペクトログラム $\boldsymbol{Y}$ を D 個の頻出の基底スペクトル $\boldsymbol{h}_d = [H_{1,d},\cdots,H_{\Omega,d}]$ とそれぞれの音量変化を表すアクティベーションで近似しようとしていることに相当する. 各基底スペクトル $\boldsymbol{h}_d$ とアクティベーション $U_{d,t}$ のペアをコンポーネントと呼ぶことが多い.

NMF は一般的に観測データとモデルの間の何らかの距離尺度を用いて目的関数を設計し, それを最小化する制約付きの最適化問題として定式化される. この距離尺度の選び方は重要であり従来さまざまな研究が行われてきている. よく用いられる尺度としては, Euclidean distance, 一般化 Kullback-Leibler divergence や Itakura-Saito divergence などが挙げられるが, 最近ではこれらを含むより広いクラスの β-divergence が用いられることも多い[7)]. 本稿では距離尺度の選び方は中心的な話題ではないため, 音源分離において性能が良いと報告されている振幅スペクトログラムに対する一般化 Kullback-Leibler (KL) divergence を用いた状況に限定して議論する[7)]. ただし, 提案手法は距離尺度の選び方に依存したものではなく軽微な修正によって他の尺度を用いることが可能である.

さらに, 一般化 Kullback-Leibler (KL) divergence 規準の NMF は次のようなモデルの最尤推定問題と等価であることが知られている[8)]:

$$Y_{\omega,t} \sim \delta\Big(Y_{\omega,t} - \sum_d C_{\omega,t,d}\Big) \ , \quad C_{\omega,t,d} \sim \mathrm{Poisson}(H_{\omega,d}U_{d,t}). \tag{1}$$

さらに基底とアクティベーションに対する事前分布を導入し, ベイズ推定によりモデルを推論する枠組みも広く研究されてきている. 例えば, Poisson 分布の共役事前分布である Gamma 分布を用いて, $H_{\omega,d} \sim \mathrm{Gamma}(a_H, b_H)$, $U_{d,t} \sim \mathrm{Gamma}(a_U, b_U)$ とするのが最も標準的な Bayesian NMF である[9)]. Euclidean distance 規準や Itakura-Saito divergence 規準の Bayesian NMF に対する研究も報告されている[6),10)].

従来, NMF を用いて観測データを分解する際, そのコンポーネントの数 D は事前に与える必要があった. しかし, 適切なコンポーネント数 D を事前に知っている状況は稀であり, 人手によるチューニングやモデル選択手法に頼る場合が多かった. このように様々な入力信号に対してそれぞれに適切なモデルの複雑度 D をチューニングしなければいけない問題に対し, 次に述べる Bayesian nonparametrics を用いた NMF の拡張が提案されている[6),10)].

2.2 Bayesian nonparametric NMF

NMF におけるコンポーネント数 D の決定はモデルの複雑度を決める上で重要な役割を果たしている. 例えば, ピアノの C, E, G の 3 音高から構成された楽曲に NMF を適用する場合を考えてみる. NMF に基づくスパース表現は多重音のスペクトログラムを頻出の基底スペクトルパターンによって表現しようとするモデルであるから, この場合 $D = 3$ として NMF を適用することで, 3 つの基底スペクトルはそれぞれ C, E, G 各音の平均的な楽器音スペクトルに対応することが期待される. しかし, この場合の $D = 3$ のように, 楽曲の構成要素の数が事前に分かっている状況は稀であり, 観測信号から適切なコンポーネント数 D を自動的に推定出来るような枠組みが NMF に求められてきた.

では, 適切なコンポーネント数とは何であるのかを具体例を通してもう少し考えてみたい. 先ほどのピアノ 3 音高の楽曲の例において, もし録音中にわずかにノイズが混入してしまっていた場合はどうなるであろうか. 例えば定常なスペクトルを持つ白色ノイズが楽曲の一部に 0dB で混入してしまっていたとする. この場合は, 楽曲スペクトログラムを構成する頻出スペクトルパターンはピアノ 3 音高の平均的なスペクトルに加え, ノイズな定常なスペクトルとなるはずであり, $D = 4$ を選択するのが妥当と思われる. では, このノイズが-100dB だった場合はどうであろうか. このノイズはすでに無視できる程度であり, $D = 3$ とするのが適切かもしれない. では, -10dB であった場合はどうなるのか. このときは $D = 3$ とすべきか $D = 4$ とすべきか実際に適用してみないと分からないかもしれない.

この例のように, さまざまな多重音の混在した複雑な音楽信号において, コンポーネント数を “10 個” や “20 個” のように決め打ちにするモデル化には限界があることが分かる. そこで Bayesian nonparametrics に基づく NMF の拡張が注目されている. ここでは, その一つである Gamma process NMF (GaP-NMF)[6] を中心に議論する.

ピアノとノイズの例のように各コンポーネント全体のゲイン (音量に相当) はモデルの複雑度を決める上で重要な役割を果たしていた. GaP-NMF は, 従来の Bayesian NMF に対して各コンポーネントに対するゲインの概念を導入し, その一連のゲインの生成が Gamma process により行われると考えたものである. ここでは直観的な理解を与えたい. 楽曲全体のゲイン Ξ は仮想的な時間 τ に伴って増えていくと考えることが出来る. (τ は楽曲に流れる時間 t とは別なものであることに注意して頂きたい. 例えば楽曲製作過程において流れる時間だと解釈して頂き, 作曲の過程において音数やパート数が増えていき楽曲全体のゲインが徐々に増えていく様子を想像すると分かりやすい.) 時間 τ に伴って単調に増加していく確率変数 Ξ_τ のモデル化として次のような Gamma process[11] を考えることが出来る:

$$\Xi_0 = 0 \ , \quad \Xi_{\tau'} - \Xi_\tau \sim \mathrm{Gamma}(\eta(\tau' - \tau),\ \eta\lambda)\ (\forall 0 \leq \tau < \tau') \ . \tag{2}$$

ただし, η, λ は非負の実数である. ここで, 時刻 $(d-1)/D$ から d/D のゲインの増分 $\theta_d = \Xi_{d/D} - \Xi_{(d-1)/D}$ を d 番目のコンポーネントのゲインとして割り当てることにする.

GaP-NMF は上述のような各コンポーネントの音量に相当するゲイン $\boldsymbol{\theta} = \{\theta_1, \theta_d, \dots\}$ を導入して次のように表現される:

$$Y_{\omega,t} = \delta\Big(Y_{\omega,t} - \sum_d C_{\omega,t,d}\Big) \ , \ C_{\omega,t,d} \sim \mathrm{Poisson}(\theta_d H_{\omega,d} U_{d,t}) \ , \ \theta_d \sim \mathrm{Gamma}(\eta/D, \eta\lambda) \ .$$

D を無限に増大させていくと, $\boldsymbol{\theta}$ が shape パラメータ η, inverse-scale パラメータ $\eta\lambda$ の Gamma process に従って生成されたと見なすことが出来る[6),11)].

GaP-NMF の重要な性質は, 適切なコンポーネント数 D を見つける必要はなく, 十分に大きな D を与えておけばよい点にある. GaP-NMF を観測データに適用すると, データを説明するのに重大な影響を持つコンポーネントには大きな θ_d が割り当てられ, 影響の小さいコンポーネントには小さな θ_d が割り当てられることが期待される. 上述のピアノ 3 音高とノイズの例であれば, $D = 50$ に設定した GaP-NMF でも $D = 100$ に設定した GaP-NMF であっても, 主にアクティブになるコンポーネントは 3 つであり, 他のコンポーネントは小さな θ_d によって抑圧されるはずである. これは $D = 3$ や $D = 4$ のように期待した分解のために適切なコンポーネント数を見つけなければならない通常の NMF にはない重要な性質である.

3. スペクトルの状態遷移に基づく楽器音スペクトログラムのモデル化

3.1 Hidden Markov model (HMM)

GaP-NMF は観測データを説明する上で必要な分の適切なコンポーネントを与えてくれる. しかし, 我々の期待は一つのコンポーネントが楽器音 1 音に対応してくれることであり, その意味においては GaP-NMF は必ずしも期待通りの振る舞いをしてくれるとは限らない. その主たる要因となるのが楽器音の持つ時間変化するスペクトルである. 例えば, ヴァイオリンのビブラートに GaP-NMF を適用した場合を考えてみる. ビブラートの中には多様な音色が繰り返し登場し, それを説明するためにはいくつかのアクティブなコンポーネントを割り当てる必要があると予想出来る. 実用的には, そのように分断されてしまったいくつかのコンポーネントも, 一つのヴァイオリンの 1 音であるとして扱いたい場合の方が多いが, 一度分断されてしまったものを一つにまとめるのは容易なことではない.

そこでこのような問題を解消するための方法として, 基底スペクトルに状態遷移モデルを導入した NMF の拡張が提案されてきている[3]. 状態の概念を持った基底スペクトルとして $\boldsymbol{H} = \{(H^{(k)}_{\omega,1})_{\Omega\times K}, \dots, (H^{(k)}_{\omega,D})_{\Omega\times K}\}$ を導入し, $\boldsymbol{Z} = (Z_{d,t})_{D\times T} \in \mathbb{N}$ を用いて時刻 t に d 番目のコンポーネントにてアクティブな状態のインデックスを $Z_{d,t}$ で表すことにすると, GaP-NMF に状態遷移スペクトルを取り入れたモデルは

$$\begin{aligned}
&H^{(k)}_{\omega,d} \sim \mathrm{Gamma}(a_H, b_H) && U_{d,t} \mid W_{d,t} \sim \mathrm{Gamma}(a_U, a_U W_{d,t}) \\
&W_{d,t} \mid U_{d,t-1} \sim \mathrm{Gamma}(a_U, a_U U_{d,t-1}) && \theta_d \sim \mathrm{Gamma}(\eta/D, \eta\lambda) \\
&Z_{d,t} \mid Z_{d,t-1}, (\pi_{d,k})_{k=1}^{\infty} \sim \boldsymbol{\pi}_{d,Z_{d,t-1}} && Y_{\omega,t} \sim \delta\left(Y_{\omega,t} - \textstyle\sum_d C_{\omega,t,d}\right) \\
&\multicolumn{2}{c}{C_{\omega,t,d} \mid (H^{(k)}_{\omega,d})_{k=1}^{K}, U_{d,t}, Z_{d,t} \sim \mathrm{Poisson}\left(\theta_d H^{(Z_{d,t})}_{\omega,d} U_{d,t}\right)}
\end{aligned} \tag{3}$$

のように表現することが出来る. 補助変数 $\boldsymbol{W}$ は音量 $\boldsymbol{U}$ の変化が時間的になめらかになるような意図で導入している[4),12)]. $\mathbb{E}[U_{d,t}] = 1/W_{d,t}$, $\mathbb{E}[W_{d,t}] = 1/U_{d,t-1}$ となる事前分布として機能していることに注意して頂きたい. この時間連続性の扱いは他にもさまざまな候補が考えられる.

ここで状態数 K の扱いについて考えたい. ピアノの場合の音色の変化は少数のスペクトルであると考えられる. 一方, ヴァイオリンのビブラートのように多様な音色変化を持つ楽器音の場合は多数のスペクトルが出現すると考えられる. したがって, 状態数に関するモデルの複雑度においても Bayesian nonparametrics の導入が有効であると考えられる.

3.2 Bayesian nonparametric HMM

前述の状態遷移スペクトルを導入した GaP-NMF において, 一つのコンポーネントに着目したい. モデル選択の考え方においては, K を色々な値で固定した上で状態遷移確率 $(\boldsymbol{\pi}_{d,k})_{k=1}^{K}$ やスペクトルの状態 $(\boldsymbol{h}_d^k)_{k=1}^{K}$ を推定し, その中で何らかの規準で一つを選ぶことが標準的である. それに対しここでは, K や $(\boldsymbol{\pi}_{d,k})_{k=1}^{K}$, $(\boldsymbol{h}_d^k)_{k=1}^{K}$ のあらゆる可能性に対し, 背後に確率分布があると考え, それらを同時に推定することを考える. つまり $K = 1, 2, \ldots$ ありとあらゆる可能性に対する事前分布が必要になる. HMM に対して Hierarchical Dirichlet process がこのようなことを可能にすることが知られている[13]. まず, 準備として Dirichlet process[14] について簡単な説明を与える.

3.2.1 Dirichlet process

可測空間 $(\Theta, \mathcal{B})$ 上の確率測度 G_0 と正の実数 α_0 に対する Dirichlet process を $\mathrm{DP}(\alpha_0, G_0)$ と表すことにする. Dirichlet process はこの可測空間上の確率測度 G に対する確率分布であり, 標本空間 Θ の任意の有限の分割 $(A_1, A_2 \ldots A_r)$ に対して,

$$(G(A_1), \ldots G(A_r)) \sim \mathrm{Dirichlet}(\alpha_0 G_0(A_1), \ldots \alpha_0 G_0(A_r)) \ . \tag{4}$$

が成り立つとして定義される[14]. 以下では, Dirichlet process $\mathrm{DP}(\alpha_0, G_0)$ が与えられた時, この Dirichlet process から生成された確率測度 G を, $G \sim \mathrm{DP}(\alpha_0, G_0)$ と表すことにする. Dirichlet process が提案されて以降, さまざまな構成法が与えられてきた. どのような構成法を用いるかは, モデル化や推論における戦略に大きく関わっている. 本稿では, 実用的によく用いられる 2 つの構成法を紹介する.

The stick-breaking construction

Dirichlet process から生成された確率測度の構成法の一つとして, 棒を次々に折っていく様子に準えた stick-breaking construction が知られている[15]. $G \sim \mathrm{DP}(\alpha_0, G_0)$ によって生成された確率測度 G は, 独立な確率変数 $(\beta'_k)_{k=1}^{\infty}$, $(\phi_k)_{k=1}^{\infty}$:

$$\beta'_k \mid \alpha_0, G_0 \sim \mathrm{Beta}(1, \alpha_0) \ , \quad \phi_k \mid \alpha_0, G_0 \sim G_0 \tag{5}$$

を用いて,

$$G = \sum_{k=1}^{\infty} \beta_k \delta_{\phi_k} \ , \quad \beta_k = \beta'_k \prod_{l=1}^{k-1} (1 - \beta'_l) \tag{6}$$

と表すことが出来ることが知られている. δ_ϕ は Dirac 測度を表している. 重要なのは $\boldsymbol{\beta} = (\beta'_k)_{k=1}^{\infty}$ についてほとんど確実に $\sum_{k=1}^{\infty} \beta_k = 1$ が成り立つ点である. つまり $\boldsymbol{\beta}$ は正の整数 $k = 1, 2, \ldots$ 上の無限次元多項分布とみなせ, したがって, Dirichlet process は無限混合モデルの構成に用いることが出来ることを意味している. 簡単のために $\boldsymbol{\beta}$ の生成過程を $\boldsymbol{\beta} \sim \mathrm{GEM}(\alpha_0)$ と表されることがよくある.

The Chinese restaurant process

Dirichlet process は Polya の壺と呼ばれる観点から捉えることも出来る[16]. 先ほどの stick-breaking construction では Dirichlet process から生成された確率測度 G を陽に扱ったが, ここでは G を参照せず（つまり G は周辺化によって消去されている）G から生成された確率変数 $\vartheta_1, \vartheta_2, \ldots$ に注目する. G が与えられた時, 変数 $\vartheta_1, \vartheta_2, \ldots$ は互いに独立なため交換可能であることに注意して頂きたい. つまり, 以降で扱うデータの生成過程において, その生成の順番は重要ではないということである. $\vartheta_1, \ldots, \vartheta_{i-1}$ が与えられた時, G を周辺化して消去することによって, ϑ_i は

$$\vartheta_i \mid \vartheta_1, \ldots, \vartheta_{i-1}, \alpha_0, G_0 \sim \sum_{l=1}^{i-1} \frac{1}{i - 1 + \alpha_0} \delta_{\vartheta_l} + \frac{1}{i - 1 + \alpha_0} G_0 \ , \tag{7}$$

と与えられることが知られている. ここで, 各変数 ϑ_k を色のついた玉だと見なすと, 式 (7) は以下のような Polya の壺からのデータの生成過程を表していることが簡単に確認できる. 今, ある玉の入った壺から一つずつ玉を取り出し記録しながら一つずつ玉を追加している状況を想像して頂きたい. ここではより分かりやすくするため, 初めに壺の中には α_0 個の黒玉が入っていたと考えることにする. (α_0 は正の実数であればよいので, “α_0 個” という表現は適切ではないことに注意して頂きたい. しかし簡単のため, ここでは仮に “1 個” などのように考えると煩雑さがなく分かりやすい) i 回目の試行では, 次のようなルールで玉の記録と追加を行っていく: 壺の中から一つの玉を取り出し, 黒玉以外だった場合, それを ϑ_i として記録し, 取り出した玉を元に戻して同色の玉を一つ追加する. もし取り出した玉が黒玉だった場合, 取り出した黒玉を元に戻し, 壺の中に入っている玉とは別の新しい色の玉を一つ追加し, それを ϑ_i として記録する. Polya の壺による解釈から分かる重要な性質は, 今までの取り出された回数の多い色玉ほど取り出されやすくなり, 逆にあまり取り出されたことのない色玉は取り出されにくくなっていくことである.

ここで, G から生成された $\vartheta_1, \ldots, \vartheta_{i-1}$ の中に K 種類の色 $\phi_1, \ldots, \phi_K$ が含まれていたとすると式 (7) は次のように書き直すことが出来る:

$$\vartheta_i \mid \vartheta_1, \ldots, \vartheta_{i-1}, \alpha_0, G_0 \sim \sum_{k=1}^{K} \frac{m_k}{i - 1 + \alpha_0} \delta_{\phi_k} + \frac{1}{i - 1 + \alpha_0} G_0 \ . \tag{8}$$

ただし, m_k は $\vartheta_1, \ldots, \vartheta_{i-1}$ のうち ϕ_k に一致するものの個数を表している. 式 (8) は Chinese restaurant process と呼ばれる隠喩で説明される[13]. 今, 限りない数のテーブルを持った中国料理店を考えているとする. それぞれの ϑ_i はこの中国料理店を訪れた客を表し, ϕ_k は

テーブルを表している. 客は一つのテーブルを選んで座る. 式 (8) は客 ϑ_i がテーブル ϕ_k に座る確率はテーブル ϕ_k に座っている客の人数 m_k に比例し, まだ誰も座っていない新しいテーブル ϕ_{K+1} に座る確率が α_0 に比例することを表している.

3.2.2 Hierarchical Dirichlet process HMM

ここでは無限状態を持つ隠れマルコフモデル infinite HMM について議論する. 最も広く用いられているものとして Hierarchical Dirichlet process を用いた Bayesian nonparametric HMM が知られている[13]. 通常の HMM と同様に考えると, 隠れ状態の系列 $\vartheta_1, \ldots, \vartheta_T$ に関して, 各 $\vartheta_t (t = 1, \ldots T)$ は無限個のアトム $\phi_1, \phi_2, \ldots$ の中の一つを選んでいると捉えればよい. では, どのようにして隠れ状態の系列を生成していくのかを見ていきたい. 今, ϑ_t に k 番目のアトム ϕ_k が割り当てられたとしよう. このとき ϑ_{t+1} は, k 番目のアトム ϕ_k から次の状態への遷移を司る確率分布 G_k によって,

$$\vartheta_{t+1} \sim G_k = \sum_{j=1}^{\infty} \beta_j \phi_j \tag{9}$$

として生成されると考えればよい. 同様に $\vartheta_{t'}$ に k' 番目のアトム $\phi_{k'}$ が割り当てられた際には, $\vartheta_{t'+1}$ を

$$\vartheta_{t'+1} \sim G_{k'} = \sum_{j=1}^{\infty} \beta_j \phi_j \tag{10}$$

から生成すると考えればよい. 以上から, infinite HMM を表現するためには, それぞれが同じアトム $\phi_1, \phi_2, \ldots$ を共有するような無限次元多項分布 $G_1, G_2, \ldots$ を用意すればよいことが分かる. このような $G_1, G_2, \ldots$ を表現するのに Hierarchical Dirichlet process を用いることが出来る. まず $G_0 \sim DP(\gamma, F)$ によって, 無限個のアトム $\phi_1, \phi_2, \ldots$ を用意する. そして, この G_0 を base measure にし, 各 $G_k (k = 1, 2, \ldots)$ を $G_k \sim DP(\alpha, G_0)$ から生成する. G_0 はアトム $\phi_1, \phi_2, \ldots$ にしかアクティブになっていないため, これを base measure とした Dirichlet process から生成される確率測度も $\phi_1, \phi_2, \ldots$ でしかアクティブにならない. このように Dirichlet process の階層化によって無限次元多項分布 $G_1, G_2, \ldots$ のアトムを共有させることが出来, 無限状態をもつ HMM を表現することが出来る.

Hierarchical Dirichlet process の構成法もさまざまな方法が提案されている. Dirichlet process に対する Chinese restaurant process と同様に, Hierarchical Dirichlet process の表現にも Chinese restaurant frachise と呼ばれる隠喩を用いた方法がある[13]. フランチャイズとなっている中国料理店たちがあり, そこでは共通のメニューが共有されている. 各料理店の各テーブルには最初に座った客によって注文された一つの料理が割り当てられる. 料理店、テーブルをまたいで同一の料理が別のテーブルにのっていても構わないとする.

まず, $G_0 \sim DP(\gamma, F)$ によって, 無限個のアトム $\phi_1, \phi_2, \ldots$ が用意されていく部分を考えていく. アトム $\phi_1, \phi_2, \ldots$ は全料理店の共有のメニューだと捉えればよい. ここで, それぞれの料理店のそれぞれのテーブルに料理が割り当てていく様子を考える. 上述の通り, テーブルの上の料理を決めるのは最初にそのテーブルに座った客である. つまり, 客が座って初めてテーブルの上の料理が決まっていくことに注意して頂きたい. j 番目の料理店の g 番目のテーブルの上にのっている料理を $\psi_{j,g}$ と表すことにする. 今, ちょうど j 番目の料理店の g 番目のテーブルに客が座り料理が決まるところだとしよう. すでに料理のおかれたテーブルには $\phi_1, \ldots, \phi_K$ が割り当てられているとする. G_0 を周辺化して消去すると, j 番目の料理店の g 番目のテーブルの料理 $\psi_{j,g}$ の生成の様子は Chinese restaurant process と同様に表すことが出来る. つまり, $\psi_{j,g}$ としては その料理が今までに使われた回数に比例して選ばれやすくなり, γ に比例して新しい料理が選ばれる:

$$\psi_{j,g} \mid \psi_{1,1}, \psi_{1,2}, \ldots, \psi_{2,1}, \ldots, \psi_{j,g-1} \sim \sum_{k=1}^{K} \frac{m_{\cdot,k}}{\sum_k m_{\cdot,k} + \gamma} \delta_{\psi_k} + \frac{\gamma}{\sum_k m_{\cdot,k} + \gamma} F \ . \tag{11}$$

ただし, $m_{j,k}$ は j 番目の料理店で料理 ϕ_k が使われている回数とし, $m_{\cdot,k} = \sum_j m_{j,k}$ とする. 料理を選ぶにあたって, その料理が今までに何回選ばれてきたのかを参照する際には料理店の垣根を越えてフランチャイズ全体を見ていることに注意して頂きたい.

次に $G_k (k = 1, 2, \ldots)$ に基づいて料理店において客がテーブルに座っていく様子を考えていく. k 番目の料理店に訪れた i 番目の客が選ぶ料理を $\vartheta_{k,i}$ と表すことにする. では, j 番目の料理店に i 番目の客が来た際にどのように料理を選ぶかを考えていこう. G_j を周辺化して消去すると, Chinese restaurant process に基づいて, 新しく来た客がつくテーブルは, そこに座っている人数に比例して選ばれやすくなり, α に比例して新しいテーブルが選ばれる. 客がテーブルを選ぶ際に, そのテーブルにのった料理を気にしていないことに注意して頂きたい. したがって, $\vartheta_{k,i}$ の生成は次のように表現することが出来る:

$$\vartheta_{j,i} \mid \vartheta_{j,1}, \ldots, \vartheta_{j,i-1} \sim \sum_{g=1}^{m_{j,\cdot}} \frac{n_{j,g}}{i - 1 + \alpha} \delta_{\psi_{j,g}} + \frac{\alpha}{i - 1 + \alpha} G_0 \tag{12}$$

ただし, $m_{j,\cdot} = \sum_k m_{j,k}$ とする.

Chinese restaurant franchise の重要な点は, テーブルを介して 2 つの階層の Dirichlet process が結びついていることにある. テーブルの上の料理を決める際には, 各料理がどの程度テーブルの上におかれているかに注目し, 客がどのようにテーブルに座っているかを気

にしていない．また, 客が料理を選ぶ際には, 各テーブルにどれくらいの人が座っているかに注目し, テーブルにどんな料理をのっているのかを気にしていない．この性質が後述のモデルの構成法の際に重要になってくる．

4. Infinite factorial hidden Markov model

4.1 提案モデルの構成法

提案手法は, 前述の状態遷移スペクトルを持つ GaP-NMF に対し, さらに HDP-HMM に基づく状態遷移スペクトルへの生成モデルを導入したものである．我々はこのモデルを Infinite factorial infinite hidden Markov model (iFiHMM) と呼んでいる．ここでは推論における戦略も考慮した二つの iFiHMM の構成法について議論する．

まず一つ目として, GaP-NMF と Hierarchical Dirichlet process の最も標準的な融合に基づく構成法を与える．各コンポーネントについて, G_0 を Stick-breaking construction により生成する: $\boldsymbol{\beta}_d \sim \mathrm{GEM}(\gamma)$, $H^{(k)}_{\omega,d} \sim \mathrm{Gamma}(a_H, b_H)$．各 G_k は G_0 を base measure とした Dirichlet process から生成すればよい．ここで, 各 G_k のアトムに対する重み $\boldsymbol{\pi}_{d,k}$ のみに着目すると, $\boldsymbol{\pi}_{d,k} \sim \mathrm{DP}(\alpha, \boldsymbol{\beta}_d)$ とすることで表現できる．$\pi_{d,k,k'}$ は d 番目のコンポーネントにおいて, インデックス k の状態からインデックス k' への状態への遷移確率に相当している．各コンポーネントを以上のように構成し, GaP-NMF と同様の重ね合わせで観測スペクトログラムが生成されていると考えると, 提案モデルは以下のように書き表せる:

$$
\begin{aligned}
&H^{(k)}_{\omega,d} \sim \mathrm{Gamma}(a_H, b_H) && U_{d,t} \mid W_{d,t} \sim \mathrm{Gamma}(a_U, a_U W_{d,t}) \\
&W_{d,t} \mid U_{d,t-1} \sim \mathrm{Gamma}(a_U, a_U U_{d,t-1}) && \theta_d \sim \mathrm{Gamma}(\eta/D, \eta\lambda) \\
&Y_{\omega,t} \sim \delta\left(Y_{\omega,t} - \textstyle\sum_d C_{\omega,t,d}\right) && \boldsymbol{\beta}_d \sim \mathrm{GEM}(\gamma) \\
&\boldsymbol{\pi}_{d,k} \mid \boldsymbol{\beta}_d \sim \mathrm{DP}(\alpha, \boldsymbol{\beta}_d) && Z_{d,t} \mid Z_{d,t-1}, (\pi_{d,k})_{k=1}^{\infty} \sim \boldsymbol{\pi}_{d,Z_{d,t-1}} \\
&\multicolumn{2}{c}{C_{\omega,t,d} \mid (H^{(k)}_{\omega,d})_{k=1}^{\infty}, U_{d,t}, Z_{d,t} \sim \mathrm{Poisson}\left(\theta_d H^{(Z_{d,t})}_{\omega,d} U_{d,t}\right)} \quad .
\end{aligned}
$$

この構成法を BnNMF-HDPHMM と呼ぶ．

次にもう一つの構成法として, Hierarchical Dirichlet process の表現に Chinese restaurant franchise を用いた方法を提案する． これは次節で説明するように推論の上で扱いやすい構成法となっている[22]． では d 番目のフランチャイズ（コンポーネント）に注目しよう． $t-1$ 番目の客が $Z_{d,t-1}$ 番目の料理（スペクトルの状態）を選んだとする．このとき, t 番目の客は $Z_{d,t-1}$ 番目の料理店に行くことを強いられる． t 番目の客は $\eta_{d,Z_{d,t-1},t} \sim \boldsymbol{\pi}_{d,Z_{d,t-1}}(\boldsymbol{\pi}_{d,Z_{d,t-1}} \sim \mathrm{GEM}(\alpha))$ に従って, $\eta_{d,Z_{d,t-1},t}$ 番目のテーブルを選択する．ここで, $\varsigma_{d,j,i}$ は j 番目の料理店の i 番目のテーブルの料理のインデックスを表しているとしよう．各 $\varsigma_{d,j,i}$ は $\varsigma_{d,j,i} \sim \boldsymbol{\beta}_d(\boldsymbol{\beta}_d \sim \mathrm{GEM}(\gamma))$ によって生成されたと考えることが出来る．以上をまとめると

$$
\begin{aligned}
&H^{(k)}_{\omega,d} \sim \mathrm{Gamma}(a_H, b_H) && U_{d,t} \mid W_{d,t} \sim \mathrm{Gamma}(a_U, a_U W_{d,t}) \\
&W_{d,t} \mid U_{d,t-1} \sim \mathrm{Gamma}(a_U, a_U U_{d,t-1}) && \theta_d \sim \mathrm{Gamma}(\eta/D, \eta\lambda) \\
&Y_{\omega,t} \sim \delta\left(Y_{\omega,t} - \textstyle\sum_d C_{\omega,t,d}\right) && \boldsymbol{\beta}_d \sim \mathrm{GEM}(\gamma) \\
&\boldsymbol{\pi}_{d,j} \sim \mathrm{GEM}(\alpha) && \varsigma_{d,j,i} \sim \boldsymbol{\beta}_d \\
&\eta_{d,j,t} \sim \boldsymbol{\pi}_{d,j} && Z_{d,t} = \varsigma_{d,Z_{d,t-1},\eta_{d,Z_{t-1},t}} \\
&\multicolumn{2}{c}{C_{\omega,t,d} \mid (H^{(k)}_{\omega,d})_{k=1}^{\infty}, U_{d,t}, Z_{d,t} \sim \mathrm{Poisson}\left(\theta_d H^{(Z_{d,t})}_{\omega,d} U_{d,t}\right)}
\end{aligned}
$$

と表すことが出来る．この構成法を BnNMF-CRFHMM と呼ぶ.

4.2 提案手法のパラメータ推定

HDP-HMM に関連したモデルのベイズ推定法として, 従来マルコフ連鎖モンテカルロ法を用いた手法が多く提案されている[13),17),18)]．一方で, ベイズ推定の強力な手法としては変分推論があり, 特に大規模なデータを扱う推論においては変分推論が好まれる傾向にある．提案手法では観測データとして音響信号のスペクトログラムを扱うため, 入力信号が短時間であったとしても非常に多くのパラメータを持ち得る．そこで, 本稿では変分推論に基づく 2 つの提案モデルの構成法についてそれぞれのパラメータ推定法について述べる．

4.2.1 BnNMF-HDPHMM

変分推論の基本的な考えは, 真の事後分布を計算上扱いやすい別の分布で近似することに基づいている．近似に用いる分布としては, 推定したいパラメータをそれぞれ独立と仮定し, パラメータごとに分解された分布を設定するのが一般的である．しかし, 推定したいパラメータの中には非常に強力な結びつきを持ったものもあり, 時に推定の精度を低下させる原因にもなりえる．BnNMF-HDPHMM において, $\boldsymbol{\pi}_d$ は $\boldsymbol{Z}_d$ に非常に大きな影響を与えており, 独立だと仮定するのが躊躇われる．そこで近年, 変分推論において, 他のパラメータに強い影響を与えているものを周辺化して消去してしまう手法が提案されている．特に, HDP に関連したモデルにおいてもこの周辺化変分推論は有効に働くことが報告されている[20)]．そこで, BnNMF-HDPHMM に対しては, この周辺化変分推論を適用する．

まず提案モデルから $\boldsymbol{\pi}$ を消去すると, $\boldsymbol{Z}, \alpha, \boldsymbol{\beta}$ に関して

$$
p(\boldsymbol{Z} \mid \alpha, \boldsymbol{\beta}) = \prod_{d,j} \left\{ \frac{\Gamma(\alpha)}{\Gamma(\alpha + n_{d,j,\cdot})} \prod_k \frac{\Gamma(\alpha\beta_{d,k} + n_{d,j,k})}{\Gamma(\alpha\beta_{d,k})} \right\} \tag{13}
$$

が得られる．しかし，このままでは式 (4.2.1) に含まれる Gamma 関数がパラメータ推定の上で取り扱いが難しくなってしまう．そこで，新たに補助変数を追加して Gamma 関数を解消し，かつ補助変数を周辺化して消去した時に式 (4.2.1) と等価になるようにパラメータを拡張する技巧が提案されている[19]．新たに自然数 $(s_{d,j,k})_{D\times K\times K}(1 \leq s_{d,j,k} \leq n_{d,j,k})$ を追加すると，式 (4.2.1) は

$$p(\boldsymbol{Z}, \boldsymbol{s} \mid \alpha, \boldsymbol{\beta}) = \prod_{d,j} \left\{ \frac{\Gamma(\alpha)}{\Gamma(\alpha + n_{d,j,\cdot})} \prod_{k} \begin{pmatrix} n_{d,j,k} \\ s_{d,j,k} \end{pmatrix} (\alpha\beta_{d,k})^{s_{d,j,k}} \right\}$$

のように拡張出来る．ただし，$\begin{pmatrix} \cdot \\ \cdot \end{pmatrix}$ は第一種スターリング数とする．ここで，$(s_{d,j,k})_{D\times K\times K}$ について周辺化すると，式 (4.2.1) と等価になり元のモデルに戻ることに注意して頂きたい．

以上から変分推論は，入力信号のスペクトログラム $\boldsymbol{Y}$ が与えられた時に事後分布 $p(\boldsymbol{C}, \boldsymbol{\theta}, \boldsymbol{H}, \boldsymbol{U}, \boldsymbol{Z}, \boldsymbol{\beta}, \boldsymbol{s} | \alpha, \eta, \lambda, a, b)$ を変分事後分布 $q(\boldsymbol{C})q(\boldsymbol{\theta})q(\boldsymbol{H})q(\boldsymbol{U})q(\boldsymbol{Z})q(\boldsymbol{\beta})q(\boldsymbol{s}|\boldsymbol{Z})$ によって近似することによって実行される．ここでは，他を固定し一つの $q(\cdot)$ を繰り返し最適化していくことによって一つの局所解を得ることにした．スペースの都合上，各反復における更新則は省略するが，従来の HDP に対する周辺化変分推論と同様にいくつかの期待値に関しては直接計算することが困難になるため，テーラー展開に基づく近似を行った[19]．

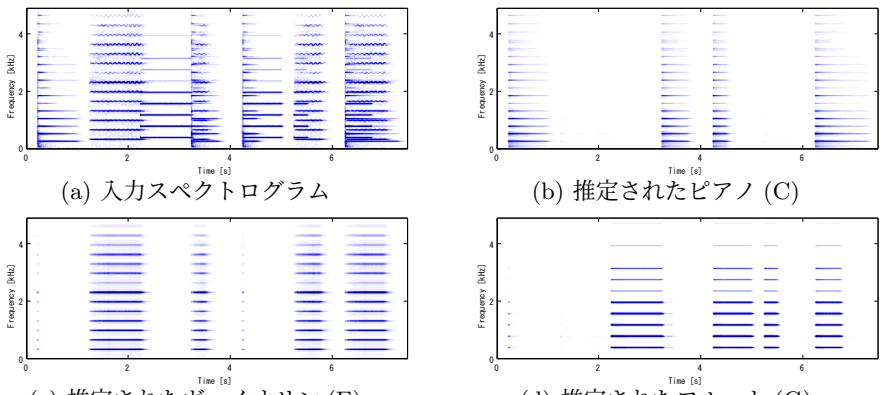
4.2.2 BnNMF-CRFHMM

最後に提案モデルのもう一つの構成法であった BnNMF-CRFHMM の変分推論について述べる．この構成法の最大の特徴は，各パラメータについて尤度関数と事前分布の間に共役性が成り立っていることである[22]．BnNMF-HDPHMM においては，$\boldsymbol{\beta}_d$ においてこのような共役性が成り立っていないため，$\boldsymbol{\beta}_d$ の事後分布の推定が容易ではない．このような問題に対し，点推定を盛り込むことで推論を行う手法もある[21]．本稿の周辺化変分推論においてはパラメータの周辺化と補助変数の追加によってこの問題を回避している．

BnNMF-CRFHMM はこのようなパラメータの共役性を保持したことによって，変分推論の更新則が非常に簡単に導出できる．スペースの都合上詳細は省略するが，期待値の計算も含め閉形式での更新則が導き出せる．

5. 実　　験

提案手法によるスペクトログラムのモデル化は音楽信号の音源分離や音楽加工への応用が可能である．スペースの都合上，本稿では提案手法の効果が分かりやすい音源分離の一例を示す．ピアノ (C)，ヴァイオリン (E)，フルート (G) の異なる 4 つの音長の MIDI 信号を用意し，はじめに 1 音ごとに演奏し，次に 2 音ずつの組み合わせを演奏し，最後に 3 音を同時に発音することで入力信号とした．振幅スペクトログラムは短時間フーリエ変換 (サンプリング周波数 16kHz，フレーム長 64ms，フレームシフト 32ms，Hanning 窓) により計算した．簡単な比較のため，標準的な NMF($D = 3$) を適用した場合の結果を図 1 に示した．モデルとしてはピアノのアタック音やヴァイオリンのビブラートが表現出来ていないことが確認出来る．無限モデルの打ち切りを $D = 10, K = 30$ とし $\alpha = \gamma = 1$, $a_H = b_H = 0.001$, $a_U = 1$, $\eta = 0.1$, $\lambda = \Omega T / \sum_{\omega,t} Y_{\omega,t}$ として提案手法 (BnNMF-HDPHMM) を適用した結果を図 2 に示す．提案手法では，楽器音数やどのような楽器が演奏されているかを事前に与えることなく，観測データを適切なコンポーネントで説明し，主要なコンポーネントが楽器 1 音と対応していることが確認出来る．



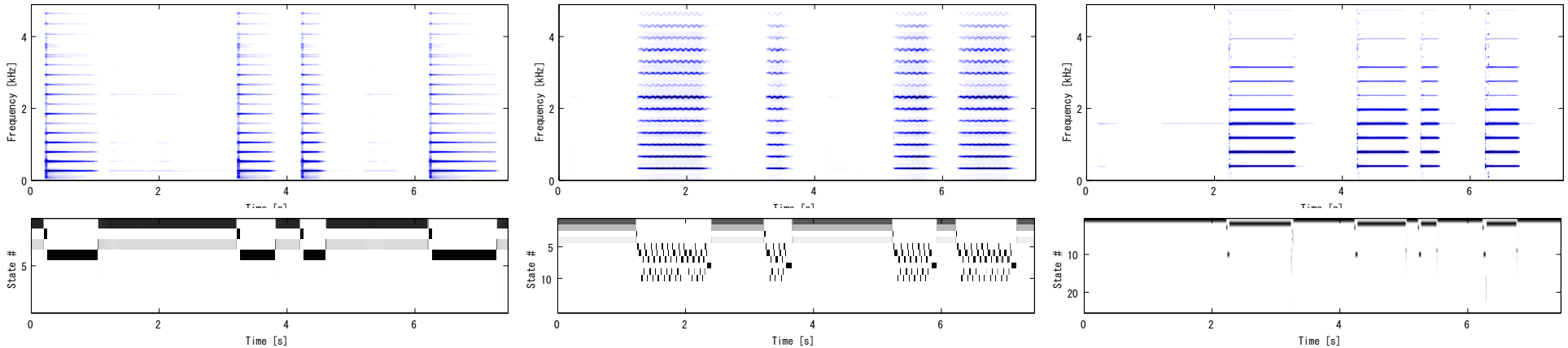
(a) 入力スペクトログラム　(b) 推定されたピアノ (C)

(c) 推定されたヴァイオリン (E)　(d) 推定されたフルート (G)

図 1　入力スペクトログラム (a) と通常の NMF により推定されたモデル $H_{\omega,d}U_{d,t}$(b), (c), (d).

6. おわりに

本報告では，音楽音響信号のモデル化として Infinite factorial infinite hidden Markov model を提案した．簡単な実験で，提案手法は入力信号にどんな楽器が含まれているか，楽器音数がいくつ含まれているかを事前に与えることなく信号をアクティブないくつかのコンポーネントに分解し，その一つ一つを楽器 1 音に対応させることが出来ることを確認した．音楽信号のモデル化の次の展開として，アクティベーションの背後に潜む楽譜を陽に扱うことが考えられる．今後は提案手法を元に音楽らしさを導入した音楽信号のモデル化への拡張を進めていきたい．



(a) 推定されたピアノ (C)　(b) 推定されたヴァイオリン (E)　(c) 推定されたフルート (G)

図 2　推定されたモデル: $\mathbb{E}[\theta_d H_{\omega,d}^{(Z_{d,t})} U_{d,t}]$ (上), $q(Z_{d,t}=k)$ (下). この 3 つが主にアクティブになり, 残りのコンポーネントは合わせて 109.2dB であった.

参 考 文 献

1) D.D. Lee and H.S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, pp. 788–791, Oct. 1999.
2) R.Hennequin, R.Badeau and B.David, "NMF with time-frequency activations to model non stationary audio events," *IEEE Trans. on Audio, Speech, and Language Process.*, 2010.
3) A. Ozerov, C. Févotte, and M. Charbit, "Factorial scaled hidden Markov model for polyphonic audio representation and source separation," in *Proc. WASPAA*, 2009.
4) M. Nakano, J. Le Roux, H. Kameoka, N. Ono and S. Sagayama, "Infinite-state spectrum model for music signal analysis," in *Proc. ICASSP*, May 2011.
5) V. Y. F. Tan and C. Févotte, "Automatic relevance determination in nonnegative matrix factorization," in *Proc. SPARS*, 2009.
6) M. Hoffman, D. Blei, and P. Cook, "Bayesian nonparametric matrix factorization for recorded music," in *Proc. ICML*, 2010.
7) D.FitzGerald, M.Cranitch and E.Coyle, "On the use of the Beta Divergence for Musical Source Separation," in *Proc. ISSC*, 2009.
8) T.Virtanen, A.T.Cemgil, and S.J.Godsill, "Bayesian Extensions to Non-negative Matrix Factorisation for Audio Signal Modelling," in *Proc. ICASSP*, 2008.
9) A.T.Cemgil, "Bayesian inference in non-negative matrix factorisation models," in *University of Cambridge*, 2008.
10) M. N. Schmidt and M. Mørup, "Infinite non-negative matrix factorization," in *Proc. EUSIPCO*, 2010.
11) J. F. C. Kingman, "Poisson processes," *Oxford University Press*, 1993.
12) A.T.Cemgil and O.Dikmen, "Conjugate gamma markov random fields for modelling nonstationary sources," in *Proc. ICA*, 2007.
13) Y. Teh, M. Jordan, M. Beal and D. Blei, "Hierarchical Dirichlet processes," in *Proc. NIPS*, 2004.
14) T.Ferguson, "A Bayesian Analysis of Some Nonparametric Problems," *Annals of Statistics*, 1973.
15) J.Sethuraman, "A Constructive Definition of Dirichlet Priors," *Statistica Sinica*, 1994.
16) D.Blackwell and J.MacQueen, "Ferguson Distributions via Pólya Urn Schemes," *Annals of Statistics*, 1973.
17) J. V. Gael, Y. Saatci, Y. W. Teh and Z. Ghahramani, "Beam sampling for the infinite hidden Markov model," in *Proc. ICML*, 2008.
18) E. B. Fox, E. B. Sudderth, M. I. Jordan and A. S. Willsky, "An HDP-HMM for systems with state persistence," in *Proc. ICML*, 2008.
19) Y. W. Teh, K. Kurihara and M. Welling, "Collapsed Variational Inference for HDP," in *Proc. NIPS*, 2008.
20) K.Yoshii and M.Goto, "Infinite Latent Harmonic Allocation: A Nonparametric Bayesian Approach to Multipitch Analysis," in *Proc. ISMIR*, 2010.
21) P.Liang, S.Petrov, M.I.Jordan and D.Klein, "The infinite PCFG using hierarchical Dirichlet processes," in *Proc. EMNLP/CoNLL*, 2007.
22) C. Wang, J. Paisley and D. M. Blei, "Online Variational Inference for the Hierarchical Dirichlet Process," in *Proc. AISTATS*, 2011.