Vol.2014-MUS-104 No.13
2014/8/26

結合動的モデルに基づく音響信号アライメント

前澤 陽[1,2,a)]　糸山 克寿[1]　吉井 和佳[1]　奥乃 博[3]　河原 達也[1]

概要：本稿では，複数の演奏者が演奏した同一楽曲の複数の音響信号の比較を支援するため，各音響信号の時刻を同一楽曲内での位置に対応づける手法（音響信号アライメント）について述べる．従来，演奏の解析において，テンポの動特性に関するモデルの有用性が指摘されていたが，一般的な音響信号アライメント手法にはテンポ推定の機構がなく，テンポ情報を活用することができなかった．本研究では，テンポの動特性を間接的にモデル化するため，楽曲の各位置で，各音響信号が演奏する，瞬時的なテンポ同士の比率をモデル化する．具体的には，瞬時的なテンポの比率が連続的であり，その変化量は音響信号間で相関があることを仮定することで，テンポ軌跡の連続性と演奏者間の類似性を同時にモデル化する．このとき，変化量を生成する背後にある共分散行列は，少数の代表的な共分散行列から構成されるマルコフ系列であるとして確率的な定式化を行う．これにより，楽曲を通して頻出する，特徴的なテンポ比率の発生箇所とその変動パターンを同時に学習することが出来るため，演奏解析に有益な情報も得られる．評価実験の結果，アライメントの精度が向上することが示され，解釈の違いの分析に対する有用性が示唆された．

1. はじめに

同じ曲に対する複数の録音には，それぞれの録音を演奏した演奏者の，解釈に対する共通性や相違が反映される．各演奏者による音楽の解釈を分析するためには，解釈の共通性と差異を切り分けることが不可欠である．特に，解釈の違いが反映されやすい演奏速度（テンポ）と音量の共通性や相違を見つけることは重要である．たとえば，テンポや音量の差異を分析・可視化することで，楽曲の演奏の仕方に対する発見を得ることができる [1–3]．また，ユーザ好みの演奏と共通したテンポや音量を持つ録音を検索することで [4]，ユーザ好みの演奏者を見つけやすくなると考えられる．このようにテンポの共通性と相違を切り分ける必要がある場合，音響信号アライメント，すなわち複数の音響信号に対して，同じ箇所を演奏する時刻を特定することが必須となる．

音楽演奏では，テンポの緩急が重要な表現の要素であるが，このような表現を音響信号アライメントで直接モデル化することは困難である．一般的に，テンポのモデル化には音符の長さ（音価）が必要であり，音響信号から音価を得ることが困難であるため，テンポの連続性といった，楽譜表現が与えられた問題設定では有効とされる性質 [5–8] などを，音響信号アライメントでは活用することができなかった [9, 10]．

そこで，我々はテンポの動特性を間接的に表現するため，楽曲上のある位置に対して，各音響信号が演奏するテンポの比率（以後**相対テンポ**と呼ぶ）の連続性と，音響信号間での相関性に着目する．楽曲中でテンポが急激に変化する箇所が限られているため，相対テンポは，滑らかに変化すると考えられる．例えば，小楽節内でのテンポは，弧を描くような軌跡を辿る傾向が一般的であるように [2]，テンポの軌跡は，演奏者を問わず，概ね滑らかである．また，テンポの変化量は幾つかのパターンに分けられ，特定の演奏者間で相関があると考えられる．例えば，前に述べた例では，フレーズ境界として曖昧な箇所では，フレーズの途中と見なした演奏者同士と，境界点と見なした演奏者同士で，相対テンポの変化量に相関が見られるだろう．このように，演奏者間の相対テンポの連続性と相関性をモデル化することで，テンポに基づく演奏解析とアライメントが同時に行える．

本研究では，音響信号間の相対テンポの関係性を明示的にモデル化する上で，複数の不確定要因を統一的に扱うことができる生成モデルを用いて音響信号アライメントに取り組む．音響信号同士の相対テンポに関する不確定さ，音響信号の背後に存在する楽曲の不確定さ，背後に存在する楽曲からどのような音響信号が生成されるかの不確定さといった，多面的な不確定要素を統一的に扱う必要があるため，生成モデルとして扱うことは必須である．具体的には，音響信号アライメントの生成モデル [10] に対して，相対テンポの連続性と関係性を統合したアライメント手法を提案

[1] 京都大学大学院情報学研究科
Kyoto University,
Yoshida Honmachi, Sakyo, Kyoto, 606–8501, Japan
[2] ヤマハ株式会社
Yamaha Corporation
[3] 早稲田大学大学院創造理工学研究科
Waseda University
[a)] akira.maezawa@gmail.com

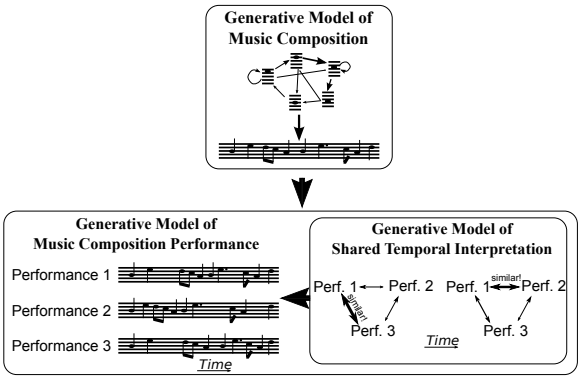

図 1 本手法の概要．同一楽曲を演奏する複数の音響信号をモデル化する際，(1) 音響信号の背後にある共通の楽曲，(2) 各々の音響信号が演奏する特徴量の系列，(3) 音響信号間における相対テンポの関連性に切り分ける．

する．つまり，**図 1** に示すように，本手法では入力された複数の音楽音響信号の背後に存在する楽曲の生成過程と，それぞれの音響信号がその楽曲を演奏する過程を同時にモデル化する．このとき，それぞれの音響信号が演奏される過程においては，それぞれの相対テンポに相関性と大まかな定常性があると考える．

2 章では楽曲の生成過程と演奏の生成過程を切り分けてモデル化できるアライメントの生成モデルについて述べる．次に，3 章で，演奏の生成過程における相対テンポのモデル化について述べたあと，4 章で提案手法を評価する．

2. 潜在共通構造モデルに基づく音楽音響信号アライメントの生成モデル

本章では，提案手法の基礎となる潜在共通構造モデル [10] について述べる．この手法では，同一楽曲を演奏する複数の音響信号の生成モデルを考え，音響信号間アライメントを生成過程の事後分布推定問題として定式化する．特徴的なのは，与えられた音響信号が演奏する「楽曲」の生成過程と，生成された楽曲が，それぞれの音響信号で「演奏」される過程を，切り分けてモデル化することである．これにより，3 章で述べる，音響信号間の相対テンポに関するモデルを，演奏の生成過程として，統一的にモデル化できる．本章で説明する生成モデルに対するグラフィカルモデルを**図 2** に図示する．

2.1 共通潜在構造と共通楽曲

楽曲とは，長さ N，状態数 S の状態系列 $\boldsymbol{Z} = \{\boldsymbol{z}_n\}_{n=1}^{N}$ として表現する．ただし，$\boldsymbol{z}_n$ は S 次元の二値変数とし，状態が s であるとき $z_{ns} = 1$ でありそれ以外の成分が 0 となるように表現する．それぞれの状態 s には，音響信号の生成過程を表すパラメータ $\boldsymbol{\theta}_s$ が割り当てられており，$\boldsymbol{\theta}_s$ の遷移により楽曲が出力する音響特徴量の順序を表現する．つまり，$\boldsymbol{z}$ は，入力音響信号の背後にある，演奏者が共通に演奏する楽曲を表現する．そこで，$\boldsymbol{z}$ を**共通楽曲**と呼ぶ．

共通楽曲 $\boldsymbol{z}$ は，初期状態確率を π，状態遷移確率を τ とする ergodic なマルコフ連鎖として，次のように表せると仮定する:

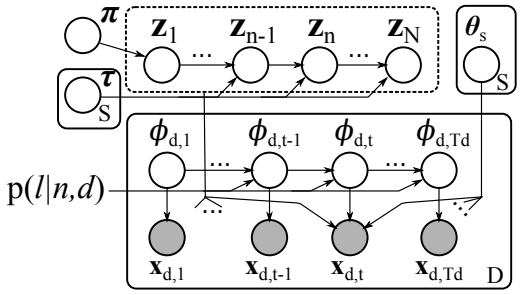

図 2 潜在共通構造モデル．破線の枠を始点とした矢印は，破線で囲った変数に，矢印の先の変数が依存することを意味する．$p(l|n,d)$ には，録音 d における潜在楽曲の位置 n に対する状態継続長音長 l の尤度である．

$$p(\boldsymbol{z}|\boldsymbol{\pi},\boldsymbol{\tau}) = \prod_{s=1}^{S} \pi_s^{z_{1,s}} \prod_{n=2}^{N} \prod_{s'=1}^{S} \prod_{s=1}^{S} \tau_{s,s'}^{z_{n-1,s'} z_{ns}} \qquad (1)$$

また，推論の簡単のため，$\boldsymbol{\tau}_s$ は共役事前分布である Dirichlet 分布から生成されたと仮定する．すなわち，$\boldsymbol{\tau}_s \sim \mathrm{Dir}(\boldsymbol{\tau}_{0,s})$ とする．初期状態確率 $\boldsymbol{\pi}$ も同様に $\boldsymbol{\pi} \sim \mathrm{Dir}(\boldsymbol{\pi}_0)$ とする．超パラメータ $\boldsymbol{\tau}_{0,s}$ と $\boldsymbol{\pi}_0$ は，$\boldsymbol{\tau}_s$ と $\boldsymbol{\pi}$ がコンパクトな構造を持つよう設定することが望ましいため，これらを 1 未満の正の値に設定することで，事後分布がスパースになるよう誘導する．

2.2 演奏系列

音響信号アライメントの生成モデルでは，共通系列の生成モデルに加え，共通楽曲を，各々の音源がどのように演奏するのかをモデル化する必要がある．そこで，同一楽曲を演奏した複数の音響信号は，二つの条件を満たすことに着目する．第一に，全ての音響信号は同一楽曲を演奏する以上，全ての音響信号は，共通楽曲に現れる状態系列と同じ**順序**で，パラメータ $\boldsymbol{\theta}_s$ から生成される音響特徴量を出力する．第二に，それぞれの音響信号は共通楽曲を独自のテンポで演奏するため，共通楽曲上のある位置 n に留まる時間は，音響信号ごとに異なる．これらの要件を満たす系列を**演奏系列**と呼ぶ．

これらの条件を満たすよう，音響信号 d に割り当てられた演奏系列を，長さ T_d，状態数 N の Left-to-right マルコフ連鎖 $\boldsymbol{\Phi}_d = \{\boldsymbol{\phi}_{d,t}\}_{t=1}^{T_d}$ として表現する．このとき，$\boldsymbol{\phi}_{d,1}$ は共通楽曲の先頭，$\boldsymbol{\phi}_{d,T_d}$ は共通楽曲の末尾に割り当てる．つまり，長さ N の共通系列 $\boldsymbol{Z}$ で定められたそれぞれの状態に留まりながら，最初から最後まで遷移することで，長さ $T_d > N$ の演奏系列を生成する．

本稿で対象とするようなテンポの相互関係をモデル化する場合，演奏系列がある状態に停留する時間を自由にモデル化できることが望ましい．そこで，$\boldsymbol{\phi}_{dt}$ の状態空間を $[1 \dots N] \times [1 \dots L]$ という積空間とし，状態 (n,l) を「状態 n にあと l フレーム留まる」とみなす．このとき，音響信号 d の演奏系列 n が停留するフレーム数 l を $p(l|n,d)$ とす

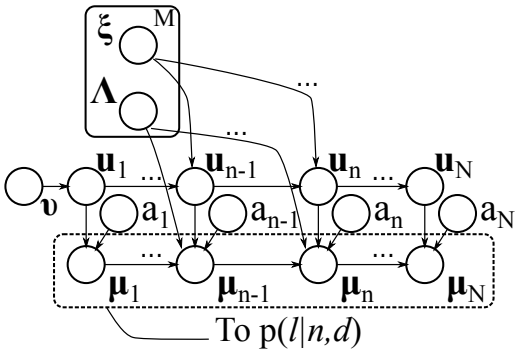

図 3 結合動的モデル．階層的なマルコフ過程から生成された μ を図 2 の $p(l|n,d)$ として用いる．

れば，このようなモデルにおける演奏系列の尤度は次のように与えられる:

$$p(\boldsymbol{\phi}_{d,t=\{1\cdots T_d\}}) = \prod_{l=1}^{L} p(l|n,d)\delta(n,1)^{\phi_{d,1,n,1}} \times \delta(n,S)^{\phi_{d,T_d,n,1}} \times \prod_{t=1}^{T_d}\prod_{n=2}^{N}\Big[\prod_{l=1}^{L} p(l|n,d)^{\phi_{dt-1,n-1,1}\phi_{dt,nl}} \prod_{l=2}^{L} 1^{\phi_{dt-1,nl}\phi_{dt,nl-1}}\Big] \quad (2)$$

このとき，状態継続長のモデルである $p(l|n,d)$ の形により，演奏系列のモデルが大きく変わることが想定されるが，このモデルについては 3 章で述べる．

2.3 音響特徴量の生成

上記の議論を踏まえると，音響信号 d の時刻 t では，まず，演奏系列 $\boldsymbol{\phi}_{dt}$ が示す共通楽曲の位置 n に割り当てられた状態 $\boldsymbol{z}_n$ を求め，その状態 $\boldsymbol{z}_n$ に関連付けられているパラメータ $\boldsymbol{\theta}_{\boldsymbol{z}_n}$ を用いて音響特徴量を生成すればよい．つまり，次のような形で観測尤度は表現される:

$$p(\boldsymbol{x}_{dt}|\boldsymbol{z},\boldsymbol{\phi},\boldsymbol{\theta}) = \prod_{s=1}^{S}\prod_{n=1}^{N} p(\boldsymbol{x}_{dt}|\boldsymbol{\theta}_s)^{z_{ns}\phi_{dtn}} \quad (3)$$

ただし，$p(\boldsymbol{x}|\boldsymbol{\theta}_s)$ は特徴量 $\boldsymbol{x}$ を状態 s で観測する尤度である．また，パラメータ $\boldsymbol{\theta}_s$ は適当な事前分布 $p(\boldsymbol{\theta}_s;\boldsymbol{\theta}_0)$ から生成されたと仮定する．

本稿では，簡単のため，$p(\boldsymbol{x}_{dt}|\boldsymbol{\theta}_s)$ を $\dim(\boldsymbol{x}_{dt})$ 次元の正規分布とし，$\boldsymbol{\theta}_s$ を共役事前分布である Normal-Gamma 分布から生成されると仮定する．すなわち，$\boldsymbol{\theta}_s = \{\tilde{\boldsymbol{\mu}}_s, \tilde{\boldsymbol{\lambda}}_s\}$, $\boldsymbol{\theta}_0 = \{\boldsymbol{m}_0, \nu_0, u_0, \boldsymbol{k}_0\}$ とした上で，$\boldsymbol{x}_{dt}|\tilde{\boldsymbol{\mu}}_s, \tilde{\boldsymbol{\lambda}}_s \sim \mathcal{N}(\tilde{\boldsymbol{\mu}}_s, \tilde{\boldsymbol{\lambda}}_s^{-1})$ と，$\tilde{\mu}_{si}, \tilde{\lambda}_{si} \sim \mathcal{NG}(m_{0,i}, \nu_{0,i}, u_{0,i}, k_{0,i})$ を仮定する．

3. 結合動的モデルに基づく演奏系列に対する状態継続長のモデル化

本節では 2.2 節で導入した演奏系列の状態継続長をモデル化することを考える．状態継続長は，楽曲に内在する音価，相対テンポの動特性，そして演奏系列間の相対テンポに対する相関性に影響されるため，これらを統一的にモデル化する必要がある．音価は，それぞれの共通楽曲位置に対して，相対テンポに対する状態継続長の比率に影響を与える．また，相対テンポの動特性は隣接する状態継続長の関係性に影響を与え，相対テンポにおける演奏系列間の相関性は，同一状態内における演奏系列間の状態継続長に影響を与える．

相対テンポは連続的であり，その変化量は楽曲の箇所に依存すると考えられる．例えば，テンポの緩急で音楽性を表現しない箇所では，相対テンポの変化量は少なく，テンポの緩急で音楽性を表現できる箇所では，相対テンポは大きく変化する可能性がある．また，相対テンポの変化量は，一部の演奏系列間で相関があると考えられる．例えば，音長に緩急を付けることが許される箇所では，緩急を付ける演奏系列同士と，緩急を付けない演奏系列同士で相関がある．また，緩急の付け方はテンポにも依存するため，速めに弾いている演奏系列同士や，遅めに弾いている演奏系列同士で，相関があると予想される．

これらを踏まえると，状態継続長は，相対テンポを音価に比例しながら伸縮したものであり，相対テンポは，局所的な連続性と演奏系列間の相関性を持っていると考えられる．そこで，このような要件を表現する相対テンポの結合動的モデルを提案する．本章で説明する確率モデルに対するグラフィカルモデルを**図 3** に示す．

3.1 状態継続長の生成モデル

まず，相対テンポの連続性と，相対テンポから共通楽曲に内在する絶対状態継続長への変換を同時にモデル化することを考える．ここで，音響信号 d に対する演奏系列において，状態 n であるときの継続長が l_{nd} フレームである確率がある平均値を中心に分布していると仮定する．この時，継続長の平均値は，状態 n が持つ音価と，その時点での演奏系列 d のテンポに依存する．絶対的な音価やテンポは音響信号から得られないため，音価とテンポに相当する，代替するパラメータを考える．そこで，状態 n の代表的な継続長 a_n を考え，それぞれの演奏系列に対する状態継続長は，a_n に，相対テンポ値 $\boldsymbol{\mu}_n = [\mu_{n,1}\cdots\mu_{n,d}]$ を掛けたものであるとする [*1]．つまり，a_n は音価に相当するパラメータ（**平均音価**と呼ぶ）であり，μ は相対テンポを表現する．これらを踏まえ，$\boldsymbol{l}_n$ は，次のように，相対テンポに平均音価を掛けた値を中心に分布すると仮定する:

$$p(\boldsymbol{l}_n|a_n,\boldsymbol{\mu}_n,\lambda_0) = \mathcal{N}\big(\boldsymbol{l}_n|a_n\boldsymbol{\mu}_n,\lambda_0^{-1}\big) \quad (4)$$

また，平均音価は，s とは独立に分布していると仮定する:

$$p(a_n|\kappa,\iota) = \mathcal{N}\big(a_n|\kappa,\iota^{-1}\big) \quad (5)$$

このとき，$\kappa = 0$ にすることで平均音価をスケール不変にでき，また ι を小さな正の値にすることにより，平均音価の分布に関する事前知識を弱めることができる．

相対テンポ $\boldsymbol{\mu}$ は滑らかに遷移するだけでなく，$\boldsymbol{\mu}$ の期待値が，ある平均値 $\boldsymbol{m}$（例えば 1）を中心に有限の分散を持つことが望ましい．なぜならば，$\boldsymbol{\mu}$ は相対値を表すた

[*1] 厳密には，μ はテンポの逆数に比例するパラメータである．

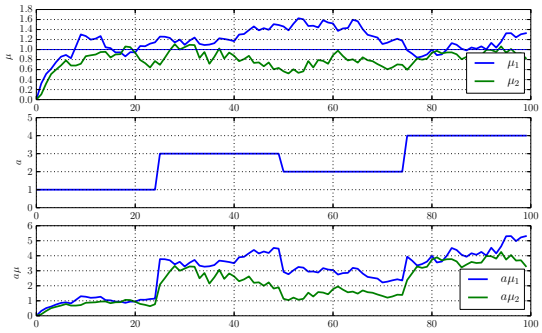

図 4 状態継続長の期待値が生成される過程．平均値 $\boldsymbol{m}=1$ を持つ相対テンポ μ を生成し（上図），平均音価 a を生成し（中図），それらの積 $a\mu$ を音長の平均値とする（下図）．

め，相対テンポは一定であることと，相対テンポは身体的な制約などから，相対テンポが極端に増えることはないためである．そこで，相対テンポ $\boldsymbol{\mu}$ は，任意の値へ戻る性質（Mean-reverting）がある線形動的システム（AR(1)）であると仮定する．すなわち，$\boldsymbol{\mu}_n$ と $\boldsymbol{\mu}_{n-1}$ の差分が，以下のような自己回帰過程に従うと仮定する:

$$\boldsymbol{\mu}_n - \boldsymbol{\mu}_{n-1} = \alpha(\boldsymbol{m} - \boldsymbol{\mu}_n) + \boldsymbol{\epsilon}_n \tag{6}$$

ここで，$\boldsymbol{\epsilon}_n$ は精度行列 Λ_n に従う正規分布から生成される雑音である．なお，Λ_n の特性については後述する．このとき，$\boldsymbol{\mu}_n$ の分散は $(1-(1+\alpha)^{-2})^{-1}$ に比例するため，$0<\alpha<1$ とすることで，分散を有限にすることができる．つまり，$\boldsymbol{\mu}_n$ が m から逸脱すればするほど，m に向けた力が α に比例して働くため，μ が発散しない．

ここまでの議論を踏まえて，Λ_n が単一の対角行列であった場合の生成プロセスを，**図 4** に図示する．この図から，μ が平均値 $m=1$ を中心とした，連続的な経路であることが分かる．また，出力値が，平均音価 a と相対テンポに比例するようになっていることが分かる．このような生成過程を推定する場合，a は μ の連続性を保つような任意の値が選ばれ，μ は連続的かつ平均値が m に近づくような経路が選ばれると考えられる．

3.2 共分散行列系列の生成モデル

式 6 で用いられる $\boldsymbol{\Lambda}_n$ は，相対テンポの連続性と，演奏系列間に対する相対テンポの相関を制御する．相対テンポが大きく変わる地点は，楽曲中の一部のみであり，それ以外では相対テンポの動特性は大きく変化しないと考えられる．また，テンポが変化する地点では，似た演奏では似た変化の仕方をすると考えられる．

そこで，$\boldsymbol{\Lambda}_n$ は，テンポの動特性を表す M 状態のマルコフ連鎖 $\boldsymbol{u}$ から選択された，M 個のうちのどれか一つの全共分散行列であると仮定する．このようにテンポの動特性のパターンに対する系列 $\boldsymbol{u}$ を**テンポ共分散系列**と呼ぶ．テンポ共分散系列 $\boldsymbol{u}$ は，状態遷移確率 $\boldsymbol{\xi}_m$，初期状態確率 $\boldsymbol{\upsilon}$ とすると，次のように表される：

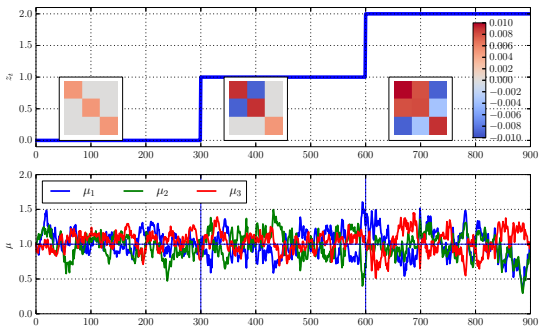

図 5 相対テンポの生成過程．演奏系列間における相対テンポ変化に対する相関の系列を，テンポ共分散系列から生成することで，音響信号間の相対テンポ変動に相関を持たせることができる（上図）．横軸 300 から 600 では，μ_1 と μ_2 の間に負の相関があり，横軸 600 から 900 では，μ_1 と μ_2 の間に正の相関，μ_2 と μ_3 の間に負の相関がある系列が生成される．

$$p(\boldsymbol{u}|\xi,\upsilon) = \prod_{m=1}^{M} \upsilon_m^{u_{1,m}} \prod_{n=1}^{N}\prod_{m=1}^{M}\prod_{m'=1}^{M} \xi_{m,m'}^{u_{n-1,m}u_{n,m'}} \tag{7}$$

$\boldsymbol{u}$ の各状態に対応する状態には，相対テンポの音響信号間における関係性を示すような精度行列が保持される．この精度行列は，共通楽曲の位置 n が定常的であるとか，演奏系列 $d=1$ と 3 の相対テンポ変化に正の相関があるといった情報を表している．$\boldsymbol{u}_n$ は，そのうち，共通楽曲の位置 n を記述するのに適切な精度行列を，精度行列の前後関係を加味しながら選択する．すると $\boldsymbol{\mu}$ は，共分散のみが時変な Switching-state Kalman filter として表現できる:

$$p(\boldsymbol{\mu}_n|\boldsymbol{\mu}_{n-1},\boldsymbol{\Lambda},\boldsymbol{u}_n) = \mathcal{N}\left(\boldsymbol{\mu}_n \middle| \frac{\boldsymbol{\mu}_{n-1}+\alpha\boldsymbol{m}}{1+\alpha}, \Lambda_m^{-1}\right)^{u_{nm}} \tag{8}$$

テンポ共分散系列 $\boldsymbol{u}$ の状態遷移確率 $\boldsymbol{\xi}_m$ と初期状態確率 $\boldsymbol{\upsilon}$ には，共役事前分布である Dirichlet 分布を仮定する．また，$\boldsymbol{\Lambda}_m$ には，共役事前分布である Wishart 分布 $\mathcal{W}(n_m,\boldsymbol{W}_m)$ を仮定する．このような過程から得られる相対テンポの系列を**図 5** に示す．このように，各状態に対して異なる共分散行列が割り当てられており，平均テンポの動特性が，各時刻に割り当てられた共分散行列に従い遷移する．

これらを踏まえ，式 2 に出現した $p(l|n,d)$ を，結合動的モデルを用い次のように表す:

$$\begin{aligned} p(\boldsymbol{l}) = \prod_{n=1}^{N} \Bigg[& \mathcal{N}\left(l_{n,d}|a_n\mu_{n,d},\lambda_d^{-1}\right)\mathcal{NG}(a_n,\lambda|a_{n,m},a_{n,l}^{-1}) \\ & \prod_{m=1}^{M}\left[\mathcal{N}\left(\boldsymbol{\mu}_n|\boldsymbol{\mu}_{n-1},\Lambda_m^{-1}\right)^{u_{n,m}} \upsilon_m^{u_{1,m}} \prod_{m'=1}^{M} \xi_{m,m'}^{u_{n-1,m}u_{n,m'}}\right]\Bigg] \\ & \times \mathrm{Dir}(\boldsymbol{\upsilon};\upsilon_0) \prod_{m=1}^{M}\left[\mathrm{Dir}(\boldsymbol{\xi}_m;\xi_0)\mathcal{W}(\boldsymbol{\Lambda}_m;n_m,\boldsymbol{W}_m)\right] \end{aligned} \tag{9}$$

3.3 事後分布の推論

本章では，2 章で提示したモデルの事後分布推定方法について述べる．このモデルでは，事後分布をベイズ則を用

いて求めることが困難であるため，変分ベイズ法を用いて，次のように因子分解できる分布 q の中から，事後分布に KL 距離の意味で最も近くなるような分布を推定する:

$$q(\boldsymbol{\phi},\boldsymbol{z},\boldsymbol{\theta},\boldsymbol{\pi},\boldsymbol{\tau},\boldsymbol{\mu},a,\boldsymbol{u},\boldsymbol{\upsilon},\boldsymbol{\xi},\boldsymbol{\Lambda}) = \prod_d q(\boldsymbol{\phi}_{d,\cdot})q(\boldsymbol{z})q(\boldsymbol{\pi})\prod_s(q(\boldsymbol{\theta}_s)q(\boldsymbol{\tau}_s)) \quad q(\boldsymbol{\mu})q(\boldsymbol{u})q(\boldsymbol{\upsilon})\prod_n(q(a_n)q(\boldsymbol{\xi}_n))\prod_m q(\boldsymbol{\Lambda}) \quad (10)$$

このとき，各因子から事後分布への KL 距離を交互に最小化することによって，局所解への収束が保証される．

まず，潜在共通構造モデルの事後分布更新方法について述べる．$q(\boldsymbol{\phi})$ と $q(\boldsymbol{z})$ は HMM と同様に，前向き後ろ向きアルゴリズムで更新することができる．つまり，データ $\boldsymbol{x}$ を観測し，時刻 t に，状態系列 s の状態 n の出力確率を $O_{t,n}$，状態 n から n' までの遷移確率を $T_{n,n'}$ としたとき，$p(s_{tn}|x_{1\cdots t}) \propto \sum_{n'} O_{t,n}T_{n',n}p(s_{t-1,n'}|x_{1\cdots t-1})$ という漸化式と $p(x_{t+1\cdots T}|s_{tn}) = \sum_{n'} p(x_{t\cdots T}|s_{t-1,n'})O_{t+1,n}T_{n',n}$ という漸化式を逐時的に求める．すると，$p(s_{tn}) \propto p(x_{t+1\cdots T}|s_{tn})p(s_{tn}|x_{1\cdots t})$ となる．$q(\boldsymbol{z})$ の推定においては，共通楽曲の位置 n における出力確率を以下に示す $\boldsymbol{g}_n$ とし，状態 s からの遷移確率を $\boldsymbol{\upsilon}_s$ とし，前向き後ろ向きアルゴリズムを実行すればよい:

$$\log g_{n,s} = \sum_{d,t}\langle\phi_{d,t,n}\rangle\langle\log p(\boldsymbol{x}_{d,t}|\boldsymbol{\theta}_s)\rangle \quad (11)$$

$$\log \upsilon_{s,s'} = \langle\log\tau_{s,s'}\rangle \quad (12)$$

ここで $\langle f(x)\rangle$ は q に対する $f(x)$ の期待値である．また，$\boldsymbol{\pi}$ は $q(\boldsymbol{\pi}) = \mathrm{Dir}(\boldsymbol{\pi}_0 + \langle\boldsymbol{z}_1\rangle)$ と更新ができ，$\boldsymbol{\tau}$ は，$q(\boldsymbol{\tau}_s) = \mathrm{Dir}(\boldsymbol{\tau}_{0,s} + \sum_{n>1}^{N}\langle z_{n-1,s}\boldsymbol{z}_n\rangle)$ と更新ができる．

$q(\boldsymbol{\phi}_{d,t})$ も $q(\boldsymbol{z})$ と同様に，出力確率を以下に示す $\boldsymbol{h}_{d,n}$，状態 (n',l') が (n,l) に遷移する確率を以下に示す $\boldsymbol{w}_{d,(n,l),(n',l')}$ として，前向き後ろ向きアルゴリズムを実行すればよい:

$$\log h_{d,t,n} = \sum_s\langle z_{n,s}\rangle\langle\log p(\boldsymbol{x}_{d,t}|\boldsymbol{\theta}_s)\rangle \quad (13)$$

$$\log w_{d,(n',l'),(n,l)} = \begin{cases} 1 & l'>1, n'=n \\ E_{l,n} & l'=1, n=n'+1 \\ 0 & \text{otherwise}\end{cases} \quad (14)$$

ただし，$E_{l,n}$ は次のように定義される:

$$E_{l,n} = \langle -\frac{1}{2}\lambda_0(\boldsymbol{l}_{nd} - a_n\mu_{nd})^2 + \frac{D}{2}\log\lambda_0\rangle \quad (15)$$

つまり，状態継続長の確率は，$a_n\boldsymbol{\mu}_n$ を中心とした正規分布となり，$l=1$ の時に，状態 $n-1$ から n へ遷移する．

また，$\boldsymbol{\theta}_s$ の事後分布は，次のように更新できる:

$$q(\tilde{\boldsymbol{\mu}}_s,\tilde{\boldsymbol{\lambda}}_s) = \mathcal{NG}\Big(\nu_0+\bar{N}_s, \frac{\nu_0\boldsymbol{m}_0+\bar{N}_s\bar{\tilde{\boldsymbol{\mu}}}_s}{\nu_0+\bar{N}_s}, u_0+\frac{\bar{N}_s}{2}, \boldsymbol{k}_0+\frac{1}{2}\Big(\bar{N}_s\bar{\boldsymbol{\Sigma}}_s+\frac{\nu_0\bar{N}_s}{\nu_0+\bar{N}_s}\left(\bar{\tilde{\boldsymbol{\mu}}}_s-\boldsymbol{m}_0\right)^2\Big)\Big) \quad (16)$$

ただし，$\bar{N}_s$，$\bar{\boldsymbol{\mu}}_s$，$\bar{\boldsymbol{\Sigma}}_s$ は次のように定義される:

$$\bar{N}_s = \sum_{d,n,t}\langle z_{n,s}\rangle\langle\phi_{d,t,n}\rangle \quad (17)$$

$$\bar{\tilde{\boldsymbol{\mu}}}_s = \frac{1}{\bar{N}_s}\sum_{d,n,t}\langle z_{n,s}\rangle\langle\phi_{d,t,n}\rangle\boldsymbol{x}_{d,t} \quad (18)$$

$$\bar{\boldsymbol{\Sigma}}_s = \frac{1}{\bar{N}_s}\sum_{d,n,t}\langle z_{n,s}\rangle\langle\phi_{d,t,n}\rangle(\boldsymbol{x}_{d,t}-\bar{\tilde{\boldsymbol{\mu}}}_s)^2 \quad (19)$$

続いて，結合動的モデルの事後分布を更新することを考える．本モデルは Switching-state Kalman filter [11] と形が近い．そこで，Switching-state Kalman filter の平均場近似に基づく推論法を変分ベイズに応用すると，事後分布の推定には，$\boldsymbol{\mu}$ の推定に Kalman smoother を用い，$\boldsymbol{u}$ の推定に HMM の前向き後ろ向きアルゴリズムを用いればよいことが分かる．なお，以下では次の量を導入する:

$$\mathcal{X}_{ndl} = \sum_{t=1}^{T_d}\langle\phi_{d,t,n-1,1}(t-1)\phi_{d,t,n,l}(t)\rangle \quad (20)$$

$$\mathcal{C}_{nd} = \sum_{l=1}^{L}\mathcal{X}_{ndl} \quad (21)$$

$$\mathcal{M}_{nd} = \frac{1}{\mathcal{C}_{nd}}\sum_{l=1}^{L}l\mathcal{X}_{ndl} \quad (22)$$

テンポ共分散系列 $\boldsymbol{u}$ は，HMM と同様に前向き後ろ向きアルゴリズムで求まる．ただし，観測尤度 O_{nm} と状態遷移確率 $T_{m,m'}$ は次のように与えられる:

$$\log O_{nm} = -\frac{1}{2}\mathrm{tr}(\Gamma_n\boldsymbol{\Lambda}_m) + \frac{1}{2}\langle\log\det\boldsymbol{\Lambda}_m\rangle \quad (23)$$

$$\log T_{m,m'} = \langle\log\xi_{m,m'}\rangle \quad (24)$$

また，初期状態 $\boldsymbol{\upsilon}$ は $q(\boldsymbol{\upsilon}) = \mathrm{Dir}(\boldsymbol{\upsilon}_0 + \langle\boldsymbol{u}_1\rangle)$ と更新ができ，状態遷移 $\boldsymbol{\xi}$ は $q(\boldsymbol{\xi}_{m,:}) = \mathrm{Dir}(\boldsymbol{\xi}_0 + \sum_{n>1}^{N}\langle u_{n-1,m}\boldsymbol{z}_{n,:}\rangle)$ と更新ができる．

次に，相対テンポ $\boldsymbol{\mu}$ の事後分布を求める．$p(\mu_n|\mathcal{X}_{1\ldots n},\cdots)$ を $\mathcal{N}(\mu_n|\boldsymbol{g}_n,\boldsymbol{V}_n)$ とすると，Kalman smoother の前向きアルゴリズムを用いて $\boldsymbol{g}$，$\boldsymbol{V}$ を求めることができる．ここで，表記の簡単のため次のような変数を導入する:

$$\boldsymbol{\gamma} = \frac{\alpha}{1+\alpha}\boldsymbol{m} \quad (25)$$

$$\beta = \frac{1}{1+\alpha} \quad (26)$$

$$\Gamma_n = \langle(\boldsymbol{\mu}_n-\beta\boldsymbol{\mu}_{n-1}-\boldsymbol{\gamma})(\boldsymbol{\mu}_n-\beta\boldsymbol{\mu}_{n-1}-\boldsymbol{\gamma})^T\rangle \quad (27)$$

$$\boldsymbol{S}_n = \sum_{m=1}^{M}\langle u_{nm}\rangle\langle\boldsymbol{\Lambda}_m\rangle \quad (28)$$

$$\boldsymbol{A}_n = \mathrm{diag}\langle\boldsymbol{a}_n\rangle \quad (29)$$

すると，前向きアルゴリズムは次のように求まる:

$$\boldsymbol{P}_n^{-1} = \boldsymbol{V}_{n-1} + \beta^2 \boldsymbol{S}_n \tag{30}$$

$$\boldsymbol{V}_n = \boldsymbol{S}_n + \boldsymbol{A}_n \boldsymbol{\Lambda}_n \boldsymbol{A}_n - \beta^2 \boldsymbol{S}_n \boldsymbol{P}_n \boldsymbol{S}_n \tag{31}$$

$$\boldsymbol{g}_n = \boldsymbol{V}_n^{-1} \Big(\beta \boldsymbol{S}_n \boldsymbol{P}_n (\boldsymbol{V}_{n-1} \boldsymbol{g}_{n-1} - \beta \boldsymbol{S}_n \boldsymbol{\gamma}) + \boldsymbol{S}_n \boldsymbol{\gamma} + \boldsymbol{A}_n \boldsymbol{\Lambda}_n \mathcal{M}_n \Big) \tag{32}$$

また，$p(\mathcal{X}_{n+1\ldots N}|\mu_n, \cdots)$ を $\mathcal{N}(\mu_n|\boldsymbol{h}_n, \boldsymbol{W}_n)$ とすると，後向きアルゴリズムを用いて $\boldsymbol{h}$ と $\boldsymbol{W}$ を次のように求めることができる:

$$\boldsymbol{Q}_n^{-1} = \boldsymbol{W}_n + \boldsymbol{S}_n + \boldsymbol{A}_n \boldsymbol{\Lambda}_n \boldsymbol{A}_n \tag{33}$$

$$\boldsymbol{W}_{n-1} = \beta^2 \boldsymbol{S}_n (\boldsymbol{I} - \boldsymbol{Q}_n \boldsymbol{S}_n) \tag{34}$$

$$\boldsymbol{h}_{n-1} = \beta \boldsymbol{W}_n^{-1} \boldsymbol{S}_n (\boldsymbol{Q}_n (\boldsymbol{W}_n \boldsymbol{h}_n + \boldsymbol{S}_n \boldsymbol{\gamma} + \boldsymbol{A}_n \boldsymbol{\Lambda}_n \mathcal{M}_n) - \boldsymbol{\gamma}) \tag{35}$$

これらを用いると，$\boldsymbol{\mu}_n$ の事後分布は次のように与えられる:

$$q(\boldsymbol{\mu}_n) = \mathcal{N}(\boldsymbol{U}_n (\boldsymbol{V}_n \boldsymbol{g}_n + \boldsymbol{W}_n \boldsymbol{h}_n), \boldsymbol{U}_n) \tag{36}$$

ただし，$\boldsymbol{U}_n = (\boldsymbol{V}_n + \boldsymbol{W}_n)^{-1}$ である．また，Γ_n を求める際に必要である $\langle \mu_n \mu_n \rangle$ と $\langle \mu_{n-1} \mu_n \rangle$ は，以下に与えられる:

$$\langle \mu_n \mu_n \rangle = \boldsymbol{U}_n \tag{37}$$

$$\langle \mu_{n-1} \mu_n \rangle = \beta \boldsymbol{P}_n^{-1} \boldsymbol{S}_n (\boldsymbol{Q}_n^{-1} + \beta^2 \boldsymbol{S}_n^T \boldsymbol{P}_n^{-1} \boldsymbol{S}_n)^{-1} \tag{38}$$

すると，平均音価 a は次のように更新される:

$$a_n \sim \mathcal{N}\left(\frac{\iota\kappa + \lambda_0 \sum_{d=1}^{D} \mathcal{C}_{nd} \langle \mu_{nd} \rangle \mathcal{M}_{nd}}{\iota + \lambda_0 \sum_{d=1}^{D} \mathcal{C}_{nd} \langle \mu_{nd}^2 \rangle}, \left(\iota + \lambda_0 \sum_{d=1}^{D} \mathcal{C}_{nd} \langle \mu_{nd}^2 \rangle \right)^{-1} \right) \tag{39}$$

また，共分散 Λ_m は次のように更新される:

$$\Lambda_m \sim \mathcal{W}\left(n_m + \sum_{n=1}^{N} \langle u_{nm} \rangle, \left(\boldsymbol{W}_m^{-1} + \sum_{n=1}^{N} \langle u_n m \rangle \Gamma_n \right)^{-1} \right) \tag{40}$$

なお，確率変数の期待値は，共役系かつ指数型なので，HMM や Kalman smoother など特筆したもの以外は，シンプルな形で解析的に求まる．これらについては，紙面の制約上割愛する．

4. 評価実験

本手法を評価するために，まず，結合動的モデルの有効性を，アライメントの精度面で検証する．次に，演奏の解釈と結合動的モデルの事後分布が一致するかを検証する．

4.1 アライメント精度の評価

Chopin の Mazurka9 曲に対してそれぞれ 2 曲から 5 曲までの音楽音響信号を用意した．これらに対し，手動で得られた拍点のアノテーションデータ [1] から算出されるアライメントと，音響信号アライメント手法で得られたものの絶対誤差パーセンタイルを評価した．特徴量には，Chroma vector [12] と Δ-chroma を半波整流したものの 24 次元を，サンプリング周波数 44.1kHz の音源に対して，フレーム長 8192 サンプルとホップサイズ 1764 サンプルで抽出した．

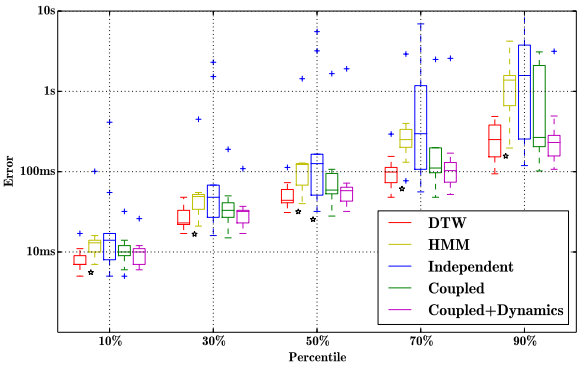

図 6 アライメントの誤差パーセンタイル．「*」は，Kruskal-Wallis 検定を行った結果，DTW と優位差があったもの ($p = 0.05$).

ベースライン手法として二乗距離最小化規準に基づく DTW を用いた．このとき，DTW の経路制約は [13] を用い，二乗誤差を最小化するような経路を求めた．また，提案法において，(1) 結合動的モデルを用いず，演奏系列を HSMM ではなく HMM として扱ったもの（「HMM」），(2) 演奏系列を HSMM として扱い，演奏系列の各位置に対する状態継続長を，演奏系列間で独立の正規分布とするもの（「Independent」），(3) 演奏系列の各位置に対する状態継続長を，演奏系列間で共通の単一の正規分布とするもの（「Coupled」），(4) 提案法（「Coupled+Dynamics」）の 4 種類を用意した．これにより，セミマルコフモデルの有効性，演奏系列間で状態継続長を共有することの有効性，相対テンポの連続性を仮定することの有効性の 3 つを検証することができる．なお，超パラメータは $\alpha = 0.1$, $\boldsymbol{m} = 1$, $\boldsymbol{W}_m = 100\boldsymbol{I}_d$, $n_m = D$, $\lambda_0 = 30$, $\iota = 0.01$, $\kappa = 0$, $\xi_0 = 0.1$, $\upsilon_0 = 0.1$ とした．

図 6 に示す通り，提案法は DTW と同程度の精度が出ることが分かる．また，相対テンポの連続性を明示的にモデル化することで，大きなアライメントの誤差が減ることが確認された．また，Independent と Coupled の結果に大きな違いがあるから，継続長に関するパラメータを複数の演奏系列で共有することが，性能向上において支配的な貢献を果たすことを示唆する．

4.2 演奏解釈の相違

本節では，本手法における共分散行列とその状態系列に，音楽的に意味をなす情報が含まれるのかを評価する．音楽的に意味があるためには，演奏者が明確に異なる解釈を持って楽曲を演奏した箇所で，共分散の構造に大きな変化が現れることが望ましい．そこで，同一の演奏者が複数の解釈で同一楽曲を演奏した場合に出現する共分散構造を分析した．まず，ヴァイオリン奏者一名が，J.S.Bach の無伴奏パルティータから 2 小節の旋律（15 秒程度）を演奏した音響信号を用意した．このとき，同一フレーズに対して，

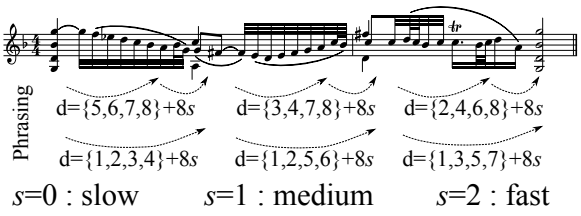

図 7 評価に用いた小楽節．破線の矢印はフレージングを表している．

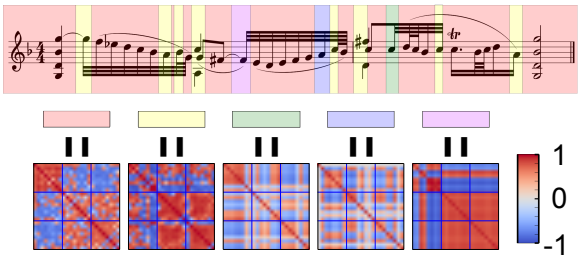

図 8 テンポ共分散系列 $\langle \boldsymbol{u} \rangle$ と，各状態に割り当てられた精度行列 $\boldsymbol{\Lambda}$ を共分散行列に変換し，要素毎の符号あり対数を取ったもの（$\mathrm{sgn}((\langle\Lambda\rangle)^{-1}) \log \mathrm{abs}((\langle\Lambda\rangle)^{-1})$）．

図 7 に示すように，３箇所で２通りのフレージングを用意し，それらの組み合わせ 2^3 通りを，３種類のテンポ（「速め」「普通」「遅め」）で演奏した．音響信号のインデックス d は図 7 に示すような割り当て方を行った．

図 8 には，得られたテンポ共分散系列を楽譜に重ねて表示し，テンポ共分散系列のそれぞれの状態に割り当てられた，相対テンポ変動の共分散行列を図示している．まず，最も頻繁に表れる状態（赤で塗りつぶされた領域）の共分散行列では，遅い演奏と，それ以外で大きなブロックを形成していることが分かる．つまり，大局的なテンポ変動は，頻出する状態の共分散構造に反映されることを示唆している．また，黄色のブロックには，$d = 1 \ldots 4$ と $d = 5 \ldots 8$ に弱い負の相関が見られ，その影響が $s = 1$, $s = 2$ にも反映されていることが示唆された．共分散行列が，期待されたようなブロック構造を持っていないものもあるが，フレーズの区切りとフレーズの違いが明記されている地点で赤色以外の共分散構造を持つことから，テンポの緩急で解釈の違いを表す余地がある箇所を検出できていることが示唆された．

5. まとめ

本稿では，生成モデルに基づく音響信号アライメントにおけるある音が演奏される音長のモデルを提案した．本モデルでは，相対テンポの連続性と演奏系列間の相関性に着目し，似た演奏同士でテンポ変化の傾向を共有することができる．

本稿で提案した結合動的モデルは，音響信号アライメント以外の枠組みでも有効であると考えられる．例えば，手動で入力されたテンポ軌跡を結合動的モデルでモデル化することにより，テンポ軌跡のパラメトリックなモデルに頼ることなく [2]，解釈の幅が広い箇所や，類似した演奏者のグルーピングに応用できる可能性がある．

本手法の共分散行列では，必要な分散行列の数が，解釈の相違を起こす数に対して指数的に増加する．今後は，因子モデルなどを活用し，解釈の違いをなす精度行列の構成要素を同時に推定することで，より表現力の高い類似性のモデルを提案したい．また，楽曲の生成モデルを同時に推定するため，楽節単位など，音楽の構成要素としてのテンポ緩急を解析するための手法へと応用したい．

参考文献

[1] Sapp, C. S.: Comparative Analysis of Multiple Musical Performances, *ISMIR*, pp. 2–5 (2007).
[2] Stowell, D. and Chew, E.: Maximum a Posteriori Estimation of Piecewise Arcs in Tempo Time-Series, *From Sounds to Music and Emotions*, LNCS(7900), Springer, pp. 387–399 (2013).
[3] Konz, V.: Automated Methods for Audio-Based Music Analysis with Applications to Musicology, PhD Thesis, Saarland University (2012).
[4] Miki, S., Baba, T. and Katayose, H.: PEVI: Interface for retrieving and analyzing expressive musical performances with scape plots, *SMC*, pp. 748–753 (2013).
[5] Raphael, C.: A Hybrid Graphical Model for Aligning Polyphonic Audio with Musical Scores, *ISMIR*, pp. 387–394 (2004).
[6] Cont, A.: A Coupled Duration-Focused Architecture for Real-Time Music-to-Score Alignment, *PAMI*, Vol. 32, No. 6, pp. 974–987 (2010).
[7] Otsuka, T., Nakadai, K., Ogata, T. and Okuno, H. G.: Incremental Bayesian Audio-to-Score Alignment with Flexible Harmonic Structure Models, *ISMIR*, pp. 525–530 (2011).
[8] Sako, S., Yamamoto, R. and Kitamura, T.: Ryry: A Real-Time Score-Following Automatic Accompaniment Playback System Capable of Real Performances with Errors, Repeats and Jumps, *AMT*, pp. 134–145 (2014).
[9] Miotto, R., Montecchio, N. and Orio, N.: Statistical Music Modeling Aimed at Identification and Alignment, *AdMIRe*, pp. 187–212 (2010).
[10] 前澤　陽，糸山克寿，吉井和佳，奥乃　博：潜在共通構造モデルに基づくオーディオ信号間のアライメント，情報処理学会音楽情報処理研究会, 2014-MUS-103 (2014).
[11] Ghahramani, Z. and Hinton, G. E.: Variational learning for switching state-space models, *Neural Computation*, Vol. 12, pp. 963–996 (1998).
[12] Fujishima, T.: Realtime Chord Recognition of Musical Sound: A System Using Common Lisp Music, *ICMC*, pp. 464–467 (1999).
[13] Hu, N., Dannenberg, R. B. and Tzanetakis, G.: Polyphonic Audio Matching and Alignment for Music Retrieval, *WASPAA*, pp. 185–188 (2003).