Vol.2013-MUS-100 No.22
2013/9/1

混合ガウス過程に基づく歌声音量軌跡の生成過程モデル

大石 康智[1,a)]　亀岡 弘和[1]　持橋 大地[2]　柏野 邦夫[1]

概要：歌声の声の大きさの変化（音量軌跡と呼ぶ）を楽譜と関連付けて特徴づけ，未知の楽譜に対して，その音量軌跡を予測できる生成過程モデルを提案する．数名の歌唱者による同一曲の歌声の音量軌跡を観察した結果，歌唱者ごとにその動特性は特有であり，楽譜や歌唱表現に起因する成分が含まれることがわかった．また，同一歌唱者による数曲の歌声の音量軌跡を観察したところ，歌唱者はいくつかの動特性パターンを所有し，楽譜が与えられた下で，パターンを使い分けて歌唱すると考えた．これらを踏まえて，楽譜における様々なコンテキスト（音符の音高や音長，音符内位置，前後の音符情報など）が与えられた下で，歌唱者が描くであろう音量軌跡を生成するモデルを構築するために，混合ガウス過程を用いる．複数のガウス過程によって音量軌跡の多様な動特性が特徴づけられ，これらの混合モデルによって歌唱者が時々刻々と動特性パターンを使い分ける動作が表現される．評価実験では，単一のガウス過程を用いるより，混合ガウス過程を用いて音量軌跡の動特性を特徴づけた方が，未知の楽譜に対する音量軌跡の予測性能が高いことを示す．また，音符のコンテキストの種類と予測性能の関係について考察する．

1. はじめに

歌声の声の大きさは声の高さや声色と同じく，歌唱者が巧妙な制御を必要とするパラメータである．本研究では，歌声の声の大きさの変化（音量軌跡）を楽譜と関連付けて特徴づけ，未知の楽譜に対して，その音量軌跡を予測できる生成過程モデルを提案する．このような生成過程モデルを記述できれば，歌唱者の歌い方や個性，癖を学習することにつながり，現在盛んに研究される歌声の認識や合成 [1–9] への工学的応用が期待できる．例えば，ある歌声を別の歌唱者の歌い方に変換して合成することが可能となるだろう．事前に歌唱者のあらゆる歌い方が学習されるため，どんな未知の楽譜が与えられても，その歌い方を転写できることを特長とする．

早速，**図 1** を見てほしい．ある楽曲の抜粋部分を，3 名の歌唱者がヘッドフォンで MIDI 伴奏を聴きながら歌った歌声の音量軌跡を示す（音量の計算方法は二乗平均平方根に基づき，次章に示す）．伴奏と同期して収録したため，メロディの MIDI ノートナンバと音長 (Interonset interval) の変化を上図に示す．各歌唱者の音量軌跡には，音量が急激に上昇する区間，ビブラートのように上下に振動する区間，緩やかに下降する区間など，様々な動特性が観測される．このような声の大きさの動特性は人間の呼吸動作に起因する．呼吸動作は，筋収縮によって肺を拡大させたり，収縮させたりする 2 つの主要な筋群によって実現される．第 1 の筋群は，通常の呼吸時に頻繁に用いられる 2 種類の肋間筋（内肋間筋と外肋間筋）であり，第 2 の筋群は，腹壁や横隔膜の筋肉によって構成され，これらの筋肉によって生み出される力は肺気量に影響を与える．肺気量の変化は声門下圧（肺の中の空気の過剰な圧力）に影響を与え，この声門下圧が声の大きさを決定すると言われている [10]．歌声は通常の話声に比べて，腹壁や横隔膜の筋肉を能動的かつ連続的に収縮させて声門下圧を迅速に変化させるため，その動作が声の大きさの動特性として表れる [11–14]．

実際，歌唱者 A と B は音大声楽科出身で発声訓練を受け

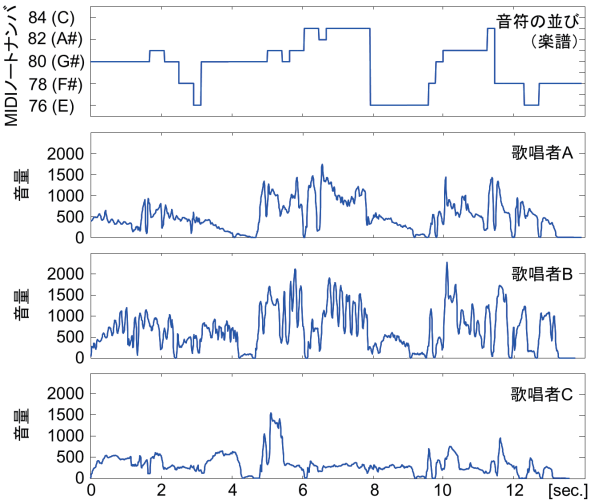

図 1　同一メロディに対する異なる歌唱者の歌声の音量軌跡

Fig. 1　Vocal volume contours of excerpts from a song sung by three singers (two professionals and one amateur)

[1] NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories, Atsugi, Kanagawa 243–0198, Japan
[2] 情報・システム研究機構 統計数理研究所
The Institute of Statistical Mathematics, Tachikawa, Tokyo 190–8562
[a)] ohishi.yasunori@lab.ntt.co.jp

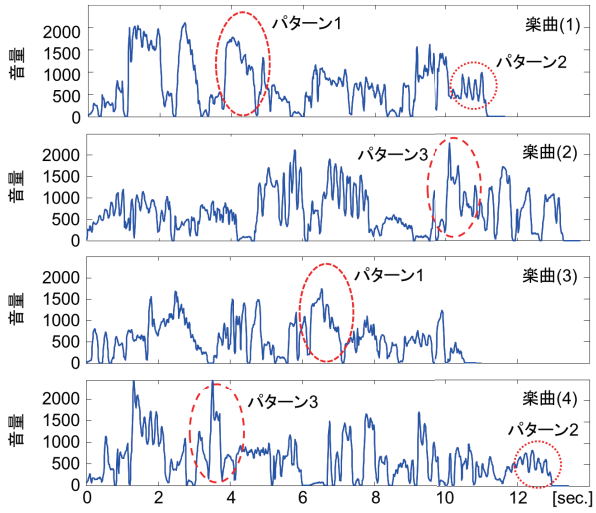

図 2 同一歌唱者が異なる楽曲のメロディを歌った歌声の音量軌跡
Fig. 2 Vocal volume contours of four songs sung by a singer

ているが，歌唱者 C は発声訓練未経験者である．その違いが音量の動特性にここまで顕著に表れることは興味深い．また，歌唱者 A と B の音量軌跡には，音符の音高が高くなるにつれて音量が大きくなったり，特定の音符で上下に振動したりする挙動が見られる．これは，声域の最も高い領域では基本周波数の上昇に伴う声門下圧の上昇が典型的に見られ，声門下圧の振動はビブラートに起因するという先行研究の調査結果 [15–17] とも整合する．**図 2** は，図 1 の歌唱者 B が 4 つの異なる楽曲のメロディを同じ収録条件で歌った歌声の音量軌跡を示す．異なるメロディにもかかわらず，類似した “パターン” が観測される．これは，歌唱者が呼吸動作において，筋肉の制御パターン（状態）をいくつか持っており，楽譜が与えられた下で，その状態を使い分けながら音量を生成することが予想される．

音量軌跡のモデリングは実は十分に検討されていない．文献 [18] では，ガウス過程を用いて，楽譜における音符のコンテキスト（音符の長さや高さ，前後の音符との相対的な差など）を入力し，ピアニストの演奏情報（音の強さ，始まりと終わりのタイミング）を出力するモデルを構築した．多くの MIDI 演奏データから，そのピアニストの演奏表情が自動的に学習される．ただし，打鍵楽器を対象とすることもあり，音符内の音量の動特性を特徴付けるまでには至っていない．文献 [19] では，状態空間モデルを用いて，音量軌跡を楽譜に起因する成分，大局的な変動成分，局所的な変動成分に分解するものの，楽譜と対応付けて動特性を学習することまでには至っていない．我々が提案する二次系に基づく基本周波数 (F0) 軌跡の生成過程モデル [20, 21] をそのまま音量軌跡に適用することも考えられるが，図 1 を見ると，もはや二次系のステップ応答から逸脱した動特性であるため，その表現が難しいと考える．

隠れマルコフモデル (HMM) に基づく音声合成のように，ラベル付けられた区間において，例えば，5 状態の HMM を用いて音量軌跡を学習することもできるが，状態内で出力確率分布が一定であるという HMM の制約のために，短時間に細かく変化する音量を，固定された状態数と局所的な動的特徴量で表現することは難しい．また，未知のコンテキストに対して頑健なモデルを構築するために木構造に基づくクラスタリングを行うが，その平均化処理のために，実際に生成される特徴量が過剰に平滑化され，本来の動特性を再現できないという問題もある [22]．

本稿では，混合ガウス過程 [23] に基づいて，音符のコンテキストを入力として，時間的に複雑に変動する音量の生成過程モデルを提案する．ガウス過程は，“回帰関数の確率分布” を定義し，データ $D = \{\boldsymbol{x}_t, y_t\}_{t=1}^{T}$ が与えられた時，これらを互いに独立に扱うのではなく，互いの相関関係を考慮し，新しい $\boldsymbol{x}_{t+1}$ に対する y_{t+1} を予測する．この相関関係は音量軌跡の動特性を考慮することに相当し，カーネル関数を用いて特徴づけられる．文献 [20, 21] は物理モデルを陽に定義して動特性を特徴づけたことに対し，本手法ではカーネル関数によって，潜在的に物理モデルが学習されることを期待する．混合ガウス過程はガウス過程の混合モデルであり，図 2 のように，歌唱者がいくつかの動特性パターンを使い分けながら音量を生成する動作を表現する．

評価実験では，単一のガウス過程を用いるより，混合ガウス過程を用いて音量軌跡の動特性を特徴づけた方が，未知の楽譜に対する音量軌跡の予測性能が高いことを示す．また，入力変数となるコンテキストの候補について，予測性能の観点から考察する．

2. 混合ガウス過程に基づく生成過程モデル

これまで，声質変換 [24] や音声分析 [25, 26]，音声合成 [22] にガウス過程を利用する手法が提案された．特に，文献 [22] では，テキストから得られる各フレームのコンテキストを入力変数とし，スペクトルからなる各フレームの音響特徴量を出力変数とするガウス過程を考える．動的特徴量や木構造のクラスタリングを用いずに，音素ごとに音響特徴量の動特性を直接モデル化することで，HMM に基づく従来法に比べて，高いスペクトル再現性が得られた．我々の提案法は，この枠組みに入力変数であるコンテキストのクラスタリングを含めたものとみなすこともできる．

図 1 のように，伴奏付きで歌った歌声と，そのメロディの楽譜が同期して与えられた下で，楽譜に含まれる様々なコンテキストから音量軌跡への回帰問題を考える．入力変数はメロディの楽譜から，例えば下記のように構成される．

$$\boldsymbol{x}_t = [\text{音符の発音開始時刻からの時間 (音符内位置)}, \text{音符の音高}, \text{音符の音長}]^{\mathrm{T}} \quad (1)$$

t はサンプリング周期（本稿では 10 ms とする）によって離散化された時刻を表す．もちろん，当該音符の発音停止時刻からの時間や，前後の音符の音高や音長，クレッシェ

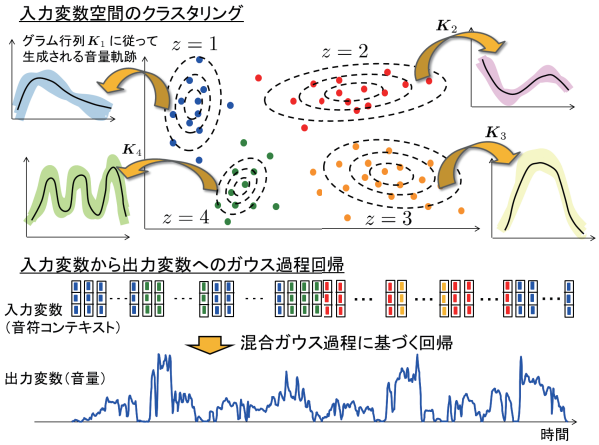

図 3 混合ガウス過程に基づく音量軌跡の生成過程

Fig. 3 Generative process of vocal volume contour based on mixture of Gaussian process experts

ンド，デクレッシェンドのような強弱記号や演奏記号の有無など，様々なコンテキストを入力変数に加えることも可能である．一方，出力変数である音量は，窓幅 N，歌声波形 $w(t)$，窓関数 $h(t)$ として，下記のように計算した [8]．

$$y_t = \sum_{\tau=-N/2}^{N/2} \left(\sqrt{(w(t+\tau) \times h(\tau))^2} \right) / N \qquad (2)$$

ここで，N は 512 点 (32 ms)，$h(t)$ はハニング窓とした．

歌唱者は，図 2 に見られる様々な動特性を生成する物理的な系（状態）をいくつか持っており，時々刻々とその状態を遷移させながら音量を生成すると想定する．このような状態遷移を考慮して，コンテキストから音量軌跡を生成するために，混合ガウス過程 [23,27] を利用する．まずは，ガウス過程回帰を説明する．入力変数 $\boldsymbol{x}$ に対する出力変数 y がガウス過程に従うとき，出力変数全体からなるベクトル $\boldsymbol{y} = [y_1, \ldots, y_T]^{\mathrm{T}}$ の確率密度関数は次の多次元ガウス分布で表される．

$$p(\boldsymbol{y}) = \mathcal{GP}(\boldsymbol{y}; \boldsymbol{0}, \boldsymbol{K} + \eta^2 \boldsymbol{I}) = \mathcal{N}(\boldsymbol{y}; \boldsymbol{0}, \boldsymbol{K} + \eta^2 \boldsymbol{I}) \qquad (3)$$

ここで，$\mathcal{GP}$ はガウス過程を表し，$\boldsymbol{K}$ は $K_{i,j} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ を要素に持つグラム行列，$k(\boldsymbol{x}_i, \boldsymbol{x}_j)$ は 2 変数間の相関を表すカーネル関数である．また，η^2 は出力変数に含まれる観測ノイズの分散パラメータ，$\boldsymbol{I}$ は単位行列を表す．

ガウス過程による回帰分析では未知の入力変数 $\boldsymbol{x}_*$ に対し，出力変数 y_* の分布を予測できる．既に与えられている入力変数集合 X と新たな入力変数 $\boldsymbol{x}_*$ とのカーネル関数の値を並べたベクトル $\boldsymbol{k}_*$ を用いると，$\boldsymbol{y}$ と y_* の同時分布は

$$p\left(\begin{bmatrix}\boldsymbol{y}\\ y_*\end{bmatrix}\right) = \mathcal{N}\left(\begin{bmatrix}\boldsymbol{y}\\ y_*\end{bmatrix}; \boldsymbol{0}, \begin{bmatrix}\boldsymbol{K}+\eta^2\boldsymbol{I} & \boldsymbol{k}_*\\ \boldsymbol{k}_*^{\mathrm{T}} & k(\boldsymbol{x}_*, \boldsymbol{x}_*)\end{bmatrix}\right)$$

となる．ゆえに未知の出力変数 y_* の予測分布は以下で与えられる．

$$p(y_*|\boldsymbol{y}, X, \boldsymbol{x}_*) = \mathcal{N}(y_*; \mu_*, \sigma_*^2) \qquad (4)$$
$$\mu_* = \boldsymbol{k}_*^{\mathrm{T}} (\boldsymbol{K} + \eta^2 \boldsymbol{I})^{-1} \boldsymbol{y}$$
$$\sigma_*^2 = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - \boldsymbol{k}_*^{\mathrm{T}} (\boldsymbol{K} + \eta^2 \boldsymbol{I})^{-1} \boldsymbol{k}_*$$

ここで，出力変数 y_* には観測ノイズを考慮しない．

ガウス過程による回帰分析のためにはカーネル関数の設計が必要である．カーネル関数に求められる条件はグラム行列が正定値対称行列となることであり，出力信号の定常性を仮定した下で，二乗指数共分散関数（ガウスカーネル）や二次有理共分散関数などを利用することが一般的である．しかしながら，図 2 より，音量軌跡は必ずしも定常な信号とは言えない．定常でない多様な動特性を，状態遷移に基づいて生成するために混合ガウス過程を利用する．

図 3 に示すように，混合ガウス過程では入力変数空間をいくつかの状態にクラスタリングする（入力コンテキストおよび音量軌跡の動特性の類似度の観点から，観測される音量軌跡が分割されクラスタリングされる）．音量軌跡の動特性は，各状態に割り当てられた入力変数から計算されるグラム行列によって特徴づけられる．最終的にこれらの状態の混合モデルとして音量軌跡を表現する．混合ガウス過程の構成方法はこれまで 2 種類提案されているが [23,27]，これらの違いは入力変数空間の密度分布を考えて完全に生成モデルで記述するか否かである．文献 [23] は完全な生成モデルを提案し，入力変数の欠損や出力変数から入力変数の予測を扱えるため，本稿で利用する．式で書くと，

$$p(\{\boldsymbol{x}_t, y_t\}_{t=1}^T | \Theta, \Omega) \qquad (5)$$
$$= \sum_{\mathcal{Z}} p(\{z_t\}_{t=1}^T | \Omega) \times \left(\prod_{r=1}^R p(\boldsymbol{y}_r | X_r, \Theta, \Omega) p(X_r | \Theta, \Omega) \right)$$

と書ける．ここで，z_t は，時刻 t の入力変数が割り当てられる状態を表し，インジケータ変数と呼ぶ．状態 r に割り当てられる入力変数の集合を $X_r \equiv \{\boldsymbol{x}_t : z_t = r\}$，これらに対応する出力変数をまとめたベクトルを $\boldsymbol{y}_r \equiv \{y_t : z_t = r\}$ と表現し，R は状態の総数とする．Θ はモデルパラメータ，Ω はハイパーパラメータを表す．式 (5) は，すべての入力変数に対するあらゆる状態割り当て $\mathcal{Z}$（T^R 個の組合せ）に関して総和を計算するが，この取り扱いが難しい．ガウス過程は個々の出力変数が互いに独立でなく，カーネル関数を用いて互いの相関関係を考慮するためである．そこで，潜在変数 z_t を周辺化することなく，直接，変数として扱う．以下にその生成過程の流れを示す．

(1) ディリクレー多項分布モデルを用いて，T 個の入力変数を R 個の状態のいずれかに割り当てる．入力変数の割り当ては集合 $\{z_t\}_{t=1}^T$ によって表現される．

(2) 状態 r におけるインジケータ変数の集合 $\{z_t : z_t = r\}$ が与えられた下で，状態 r の密度分布のパラメータ $\theta_r^{\boldsymbol{x}} = \{\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r\}$ が生成される．ここでは，密度分布として，全共分散行列をもつガウス分布を仮定する．

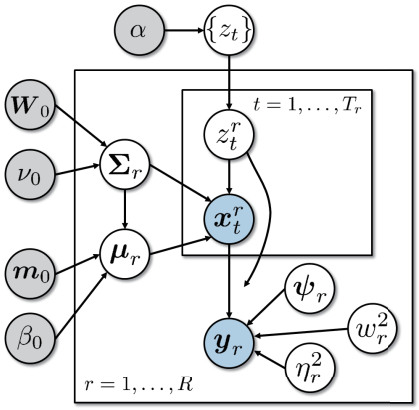

図 4 提案モデルのグラフィカル表現
Fig. 4 Graphical representation of proposed model

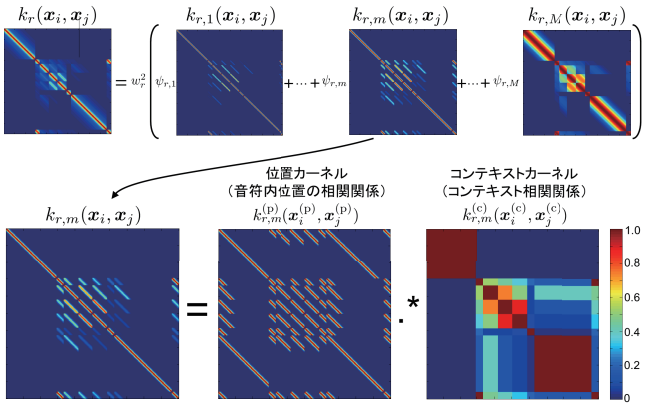

図 5 マルチカーネル学習に基づいて計算されるグラム行列
Fig. 5 Gram matrix based on multiple kernel learning

(3) 状態ごとに，カーネル関数のパラメータ θ_r^{GP} が生成される．
(4) 分布のパラメータ $\theta_r^{\boldsymbol{x}}$ が与えられた下で，状態 r に属する入力変数 X_r が生成される．
(5) 最終的に，状態ごとに，入力変数集合 X_r とカーネル関数のパラメータ θ_r^{GP} を使ってグラム行列が計算され，出力変数ベクトル $\boldsymbol{y}_r$ が生成される．

このとき，完全同時分布は

$$p(\{\boldsymbol{x}_t, y_t\}_{t=1}^T, \{z_t\}_{t=1}^T, \{\theta_r^{\mathrm{GP}}\}_{r=1}^R, \{\theta_r^{\boldsymbol{x}}\}_{r=1}^R|\Omega) \quad (6)$$
$$= \prod_{r=1}^R \left[p(\theta_r^{\boldsymbol{x}}|\Omega)p(X_r|\theta_r^{\boldsymbol{x}})p(\boldsymbol{y}_r|X_r, \theta_r^{\mathrm{GP}}, \Omega)\right] \times p(\{z_t\}_{t=1}^T|\Omega)$$

と書ける．式 (6) における各分布を次のように定義する．

$$p(\{z_t\}_{t=1}^T|\Omega) = \int p(\{z_t\}_{t=1}^T|\boldsymbol{\pi})p(\boldsymbol{\pi}|\alpha)d\boldsymbol{\pi}$$
$$= \frac{\Gamma(\alpha)}{\Gamma(\alpha+T)} \prod_{r=1}^R \frac{\Gamma(T_r + \alpha/R)}{\Gamma(\alpha/R)} \quad (7)$$
$$p(\theta_r^{\boldsymbol{x}}|\Omega) = \mathcal{N}(\boldsymbol{\mu}_r; \boldsymbol{m}_0, \boldsymbol{\Sigma}_r/\beta_0)\mathcal{W}(\boldsymbol{\Sigma}_r^{-1}; \boldsymbol{W}_0, \nu_0) \quad (8)$$
$$p(X_r|\theta_r^{\boldsymbol{x}}) = \mathcal{N}(X_r; \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r) \quad (9)$$
$$p(\boldsymbol{y}_r|X_r, \boldsymbol{\theta}_r^{\mathrm{GP}}, \Omega) = \mathcal{GP}\left(\boldsymbol{y}_r; \boldsymbol{0}, \boldsymbol{K}_r + \eta_r^2 \boldsymbol{I}_r\right) \quad (10)$$

ここで，α はディリクレー多項分布モデルのハイパーパラメータ，T_r は集合 X_r の要素数，$\boldsymbol{I}_r$ は $T_r \times T_r$ の単位行列，$\mathcal{W}$ はウィシャート分布を表す．η_r^2 は出力変数の観測ノイズを表現するための分散パラメータである．提案モデルのグラフィカル表現を**図 4** に示す．

グラム行列 $\boldsymbol{K}_r$ は集合 X_r における入力変数とカーネル関数のパラメータ θ_r^{GP} を用いて計算される．本稿では，マルチカーネル学習の考え方に基づき，**図 5** のように，複数のカーネルの線形結合によって与えられるカーネル関数

$$k_r(\boldsymbol{x}_i, \boldsymbol{x}_j) = w_r^2 \sum_{m=1}^M \psi_{r,m} k_{r,m}(\boldsymbol{x}_i, \boldsymbol{x}_j) \quad (11)$$

を導入し，カーネル関数全体の強度 ω_r^2 と各カーネルの優勢度 $\psi_{r,m}$ を推定すべき未知パラメータとみなす [25]．ここで，$\boldsymbol{x}_i, \boldsymbol{x}_j \in X_r$，$\sum_{m=1}^M \psi_{r,m} = 1$，$M$ は線形結合するカーネル関数の総数である．ただし，単位の異なる様々なコンテキスト（時刻や音高，音長など）を扱うため，個々のカーネル関数 $k_{r,m}(\boldsymbol{x}_i, \boldsymbol{x}_j)$ の構成方法には工夫が必要である．文献 [22] を参考に，$k_{r,m}(\boldsymbol{x}_i, \boldsymbol{x}_j)$ を，音符内位置の類似度を表す位置カーネル $k_{r,m}^{(\mathrm{p})}(\boldsymbol{x}_i^{(\mathrm{p})}, \boldsymbol{x}_j^{(\mathrm{p})})$ と，コンテキストの類似度を表すコンテキストカーネル $k_{r,m}^{(\mathrm{c})}(\boldsymbol{x}_i^{(\mathrm{c})}, \boldsymbol{x}_j^{(\mathrm{c})})$ を掛けあわせて表現する．

$$k_{r,m}(\boldsymbol{x}_i, \boldsymbol{x}_j) = k_{r,m}^{(\mathrm{p})}(\boldsymbol{x}_i^{(\mathrm{p})}, \boldsymbol{x}_j^{(\mathrm{p})})k_{r,m}^{(\mathrm{c})}(\boldsymbol{x}_i^{(\mathrm{c})}, \boldsymbol{x}_j^{(\mathrm{c})}) \quad (12)$$

すなわち，入力変数ベクトル $\boldsymbol{x}_i$ を $\boldsymbol{x}_i^{(\mathrm{p})}, \boldsymbol{x}_i^{(\mathrm{c})}$ のような二つのグループに分けてカーネル関数を計算する．位置カーネルは二乗指数共分散関数に基づいて下記のように構成する．

$$k_{r,m}^{(\mathrm{p})}(\boldsymbol{x}_i^{(\mathrm{p})}, \boldsymbol{x}_j^{(\mathrm{p})}) = \exp\left(-\frac{(\boldsymbol{x}_i^{(\mathrm{p})} - \boldsymbol{x}_j^{(\mathrm{p})})^{\mathrm{T}}(\boldsymbol{x}_i^{(\mathrm{p})} - \boldsymbol{x}_j^{(\mathrm{p})})}{2{l_m^{(\mathrm{p})}}^2}\right)$$

コンテキストカーネルは

$$k_{r,m}^{(\mathrm{c})}(\boldsymbol{x}_i^{(\mathrm{c})}, \boldsymbol{x}_j^{(\mathrm{c})}) = \exp\left(-\frac{1}{2}(\boldsymbol{x}_i^{(\mathrm{c})} - \boldsymbol{x}_j^{(\mathrm{c})})^{\mathrm{T}}\boldsymbol{\Lambda}(\boldsymbol{x}_i^{(\mathrm{c})} - \boldsymbol{x}_j^{(\mathrm{c})})\right)$$
$$\boldsymbol{\Lambda}^{-1} = \mathrm{diag}({l_{m,1}^{(\mathrm{c})}}^2, {l_{m,2}^{(\mathrm{c})}}^2, \ldots, {l_{m,D_{\mathrm{c}}}^{(\mathrm{c})}}^2)$$

とする．ここで，D_{c} は $\boldsymbol{x}_i^{(\mathrm{c})}$ の次元数である．よって，推定すべきパラメータは，$\Theta = \{\theta_1^{\boldsymbol{x}}, \ldots, \theta_R^{\boldsymbol{x}}, \theta_1^{\mathrm{GP}}, \ldots, \theta_R^{\mathrm{GP}}\}$ と整理される．ここで，$\theta_r^{\boldsymbol{x}} = \{\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r\}, \theta_r^{\mathrm{GP}} = \{w_r^2, \psi_{r,1}, \ldots, \psi_{r,M}, \eta_r^2\}$ である．一方，ハイパーパラメータは $\Omega = \{\alpha, \boldsymbol{m}_0, \boldsymbol{W}_0, \beta_0, \nu_0, l_1^{(\mathrm{p})}, \ldots, l_M^{(\mathrm{p})}, l_{1,1}^{(\mathrm{c})}, \ldots, l_{M,D_{\mathrm{c}}}^{(\mathrm{c})}\}$ となる．

混合ガウス過程における，未知の入力変数 $\boldsymbol{x}_*$ に対する出力変数 y_* の予測分布を導出する．式 (5) を参考に，

$$p(y_*|\{y_t, \boldsymbol{x}_t\}_{t=1}^T, \boldsymbol{x}_*, \Theta, \Omega) \quad (13)$$
$$= \sum_{r=1}^R p(y_*|\boldsymbol{y}_r, X_r, \boldsymbol{x}_*, z_* = r, \theta_r^{\mathrm{GP}})p(z_* = r|\boldsymbol{x}_*, \theta_r^{\boldsymbol{x}})$$

と書ける．ここで，$p(z_* = r|\boldsymbol{x}_*, \theta_r^{\boldsymbol{x}})$ は正規化して

$$p(z_*=r|\boldsymbol{x}_*,\theta_r^{\boldsymbol{x}})=\frac{p(\boldsymbol{x}_*|z_*=r,\theta_r^{\boldsymbol{x}})p(z_*=r)}{p(\boldsymbol{x}_*)} \quad (14)$$
$$p(\boldsymbol{x}_*|z_*=r,\theta_r^{\boldsymbol{x}})=\mathcal{N}(\boldsymbol{x}_*;\boldsymbol{\mu}_r,\boldsymbol{\Sigma}_r),\ p(z_*=r)=\frac{1}{T_r}$$

と計算される．一方，$p(y_*|\boldsymbol{y}_r,X_r,\boldsymbol{x}_*,z_*=r,\theta_r^{\mathrm{GP}})$ は

$$p(y_*|\boldsymbol{y}_r,X_r,\boldsymbol{x}_*,z_*=r,\theta_r^{\mathrm{GP}})=\mathcal{N}(y_*;\mu_{r,*},\sigma_{r,*}^2) \quad (15)$$
$$\mu_{r,*}=\boldsymbol{k}_{r,*}^{\mathrm{T}}(\boldsymbol{K}_r+\eta_r^2\boldsymbol{I}_r)^{-1}\boldsymbol{y}_r$$
$$\sigma_{r,*}^2=k_r(\boldsymbol{x}_*,\boldsymbol{x}_*)-\boldsymbol{k}_{r,*}^{\mathrm{T}}(\boldsymbol{K}_r+\eta_r^2\boldsymbol{I}_r)^{-1}\boldsymbol{k}_{r,*}$$

となる．式 (13) を互いに独立な確率分布から生成される確率変数の線形和とみなすと，平均と分散が求められる．

$$p(y_*|\{y_t,\boldsymbol{x}_t\}_{t=1}^T,\boldsymbol{x}_*,\Theta,\Omega)=\mathcal{N}(y_*;\mu_*,\sigma_*^2) \quad (16)$$
$$\mu_*=\sum_{r=1}^R c_r\mu_{r,*},\ \sigma_*^2=\sum_{r=1}^R c_r^2\sigma_{r,*}^2$$

ここで，$c_r=p(z_*=r|\boldsymbol{x}_*,\theta_r^{\boldsymbol{x}})$ である．

3. パラメータの推論

MCMC-EM アルゴリズム [28] を用いて，各パラメータを推論する．具体的には，インジケータ変数と入力変数空間のガウス分布のパラメータの推論にはギブスサンプリングを用い，ガウス過程のカーネル関数のパラメータの推論には EM アルゴリズムを利用する．

■ $z_1,\ldots,z_T$ の推論

z_t の事後分布は，

$$p(z_t=r|\boldsymbol{z}_{\setminus t},\{\boldsymbol{x}_t\}_{t=1}^T,\{y_t\}_{t=1}^T,\theta_r^{\boldsymbol{x}},\theta_r^{\mathrm{GP}}) \quad (17)$$
$$\propto p(y_t|\boldsymbol{y}_{r,\setminus t},X_{r,\setminus t},\boldsymbol{x}_t,\theta_r^{\mathrm{GP}})p(z_t=r|\boldsymbol{z}_{\setminus t},\{\boldsymbol{x}_t\}_{t=1}^T,\theta_r^{\boldsymbol{x}})$$

と書ける．ここで，$\boldsymbol{z}_{\setminus t}=\{z_1,\ldots,z_{t-1},z_{t+1},\ldots,z_T\}$，$\boldsymbol{y}_{r,\setminus t}=\{y_i:i\neq t,z_i=r\}$，$X_{r,\setminus t}=\{\boldsymbol{x}_i:i\neq t,z_i=r\}$ とすると，第一項の条件付き確率は，

$$p(y_t|\boldsymbol{y}_{r,\setminus t},X_{r,\setminus t},\boldsymbol{x}_t,\theta_r^{\mathrm{GP}})=\mathcal{N}(y_t;\mu_t,\sigma_t^2) \quad (18)$$
$$\mu_t=\boldsymbol{k}_{r,t}^{\mathrm{T}}[\boldsymbol{K}_{r,\setminus t}+\eta_r^2\boldsymbol{I}_r]^{-1}\boldsymbol{y}_{r,\setminus t}$$
$$\sigma_t^2=k_r(\boldsymbol{x}_t,\boldsymbol{x}_t)-\boldsymbol{k}_{r,t}^{\mathrm{T}}[\boldsymbol{K}_{r,\setminus t}+\eta_r^2\boldsymbol{I}_r]^{-1}\boldsymbol{k}_{r,t}$$

となる．ここで，$\boldsymbol{K}_{r,\setminus t}$ は集合 $X_{r,\setminus t}$ における入力変数を使って計算されるグラム行列である．$\boldsymbol{k}_{r,t}$ は $X_{r,\setminus t}$ における入力変数と $\boldsymbol{x}_t$ とのカーネル関数の値を並べたベクトルである．第二項は，

$$p(z_t=r|\boldsymbol{z}_{\setminus t},\{\boldsymbol{x}_t\}_{t=1}^T,\theta_r^{\boldsymbol{x}}) \quad (19)$$
$$\propto\frac{T_{r,\setminus t}+\alpha/R}{T-1+\alpha}\int p(\boldsymbol{x}_t|\theta_r^{\boldsymbol{x}})p(\theta_r^{\boldsymbol{x}}|\boldsymbol{x}_{\setminus t})d\theta_r^{\boldsymbol{x}}$$
$$=\frac{T_{r,\setminus t}+\alpha/R}{T-1+\alpha}\frac{\int p(\boldsymbol{x}_t|\theta_r^{\boldsymbol{x}})\prod_{i:z_i=r,i\neq t}p(\boldsymbol{x}_i|\theta_r^{\boldsymbol{x}})p(\theta_r^{\boldsymbol{x}})d\theta_r^{\boldsymbol{x}}}{\int\prod_{i:z_i=r,i\neq t}p(\boldsymbol{x}_i|\theta_r^{\boldsymbol{x}})p(\theta_r^{\boldsymbol{x}})d\theta_r^{\boldsymbol{x}}}$$

となる．$T_{r,\setminus t}$ は集合 $X_{r,\setminus t}$ の要素数を表す．ここで，積分を含む後半部分の詳細な計算方法は付録 A に示す．式 (19) を用いて計算されるすべての状態の事後確率に基づいて，z_t の割り当てをサンプリングする．

■ $\boldsymbol{\psi}_r$，w_r^2，η_r^2 の推論

EM アルゴリズムを用いることで目的関数の増加を保証する更新式を導くことができる [25]．まず，状態 r の出力変数ベクトル $\boldsymbol{y}_r$ を $M+1$ 個の独立な確率変数

$$\boldsymbol{u}_{r,m}\sim\mathcal{N}(\boldsymbol{u}_{r,m};\boldsymbol{0},w_r^2\psi_{r,m}\boldsymbol{K}_{r,m}),\ m=1,\ldots,M \quad (20)$$
$$\boldsymbol{u}_{r,M+1}\sim\mathcal{N}(\boldsymbol{u}_{r,M+1};\boldsymbol{0},\eta_r^2\boldsymbol{I}_r) \quad (21)$$

の和に分解し，これらを完全データと扱う．よって，完全データ $\boldsymbol{u}_r=(\boldsymbol{u}_{r,1}^{\mathrm{T}},\ldots,\boldsymbol{u}_{r,M+1}^{\mathrm{T}})^{\mathrm{T}}$ に対する対数尤度関数は，

$$\log p(\boldsymbol{u}_r;\theta_r^{\mathrm{GP}})\overset{\mathrm{c}}{=}-\frac{1}{2}(\log|\boldsymbol{S}_r|+\boldsymbol{u}_r^{\mathrm{T}}\boldsymbol{S}_r^{-1}\boldsymbol{u}_r) \quad (22)$$
$$\boldsymbol{S}_r=\begin{bmatrix}w_r^2\psi_{r,1}\boldsymbol{K}_{r,1} & & & O\\ & \ddots & & \\ & & w_r^2\psi_{r,M}\boldsymbol{K}_{r,M} & \\ O & & & \eta_r^2\boldsymbol{I}_r\end{bmatrix}$$

で与えられる．ただし，$\overset{\mathrm{c}}{=}$ は定数項以外の等号を表す．上式に対し，$\boldsymbol{y}_r$，$\theta_r^{\mathrm{GP}}=\theta_r^{\mathrm{GP}\prime}$ が与えられたときの条件付き期待値をとると，Q 関数は

$$Q(\theta_r^{\mathrm{GP}},\theta_r^{\mathrm{GP}\prime})=-\frac{1}{2}\left(\log|\boldsymbol{S}_r|+\mathrm{tr}\left(\boldsymbol{S}_r^{-1}\mathbb{E}[\boldsymbol{u}_r\boldsymbol{u}_r^{\mathrm{T}}|\boldsymbol{y}_r;\theta_r^{\mathrm{GP}}]\right)\right)$$

となる．ここで，$\boldsymbol{H}_r\equiv[\boldsymbol{I}_r,\ldots,\boldsymbol{I}_r]$ とおくと，不完全データ $\boldsymbol{y}_r$ と完全データ $\boldsymbol{u}_r$ との間には $\boldsymbol{y}_r=\boldsymbol{H}_r\boldsymbol{u}_r$ なる関係式が成り立つことから，$\mathbb{E}[\boldsymbol{u}_r\boldsymbol{u}_r^{\mathrm{T}}|\boldsymbol{y}_r;\theta_r^{\mathrm{GP}}]$ は

$$\mathbb{E}[\boldsymbol{u}_r\boldsymbol{u}_r^{\mathrm{T}}|\boldsymbol{y}_r;\theta_r^{\mathrm{GP}}]=\boldsymbol{S}_r-\boldsymbol{S}_r\boldsymbol{H}_r^{\mathrm{T}}(\boldsymbol{H}_r\boldsymbol{S}_r\boldsymbol{H}_r^{\mathrm{T}})^{-1}\boldsymbol{H}_r\boldsymbol{S}_r$$
$$+\boldsymbol{S}_r\boldsymbol{H}_r^{\mathrm{T}}(\boldsymbol{H}_r\boldsymbol{S}_r\boldsymbol{H}_r)^{-1}\boldsymbol{y}_r\boldsymbol{y}_r^{\mathrm{T}}(\boldsymbol{H}_r\boldsymbol{S}_r\boldsymbol{H}_r^{\mathrm{T}})^{-\mathrm{T}}\boldsymbol{H}_r\boldsymbol{S}_r^{\mathrm{T}}$$

と具体的に計算される．この各対角ブロックを $\boldsymbol{R}_{r,1},\ldots,\boldsymbol{R}_{r,M+1}$ と置くと，Q 関数は

$$Q(\theta_r^{\mathrm{GP}},\theta_r^{\mathrm{GP}\prime})=-\frac{T_r}{2}\log(w_r^{2L}\eta_r^2\psi_{r,1}\ldots\psi_{r,M})$$
$$-\frac{1}{2w_r^2}\sum_{m=1}^M\frac{1}{\psi_{r,m}}\mathrm{tr}(\boldsymbol{K}_{r,m}^{-1}\boldsymbol{R}_{r,m})-\frac{1}{2\eta_r^2}\mathrm{tr}(\boldsymbol{R}_{r,M+1}) \quad (23)$$

と具体的に書ける．パラメータの更新式は解析的に得ることができ，それぞれ，

$$\psi_{r,m}=\frac{1}{T_rw_r^2}\mathrm{tr}(\boldsymbol{K}_{r,m}^{-1}\boldsymbol{R}_{r,m}) \quad (24)$$
$$\omega_r^2=\frac{1}{T_rM}\sum_{m=1}^M\frac{1}{\psi_{r,m}}\mathrm{tr}(\boldsymbol{K}_{r,m}^{-1}\boldsymbol{R}_{r,m}) \quad (25)$$
$$\eta_r^2=\frac{1}{T_r}\mathrm{tr}(\boldsymbol{R}_{r,M+1}) \quad (26)$$

となる．ただし，$\sum_{m=1}^M\psi_{r,m}=1$ となるように更新後に

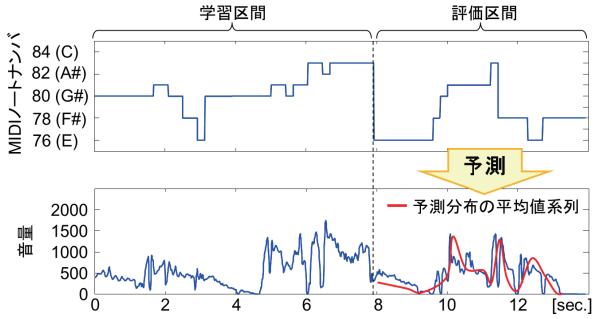

図 6 学習区間と評価区間の作成：収録された歌声ごとに，歌唱した音符数の割合が 6:4 となるように学習区間と評価区間に分離した．評価区間の入力変数から音量軌跡を予測する．

Fig. 6 Training and test segments for evaluation

$\psi_{r,m}$ は正規化される．パラメータの値が収束するまで，式 (23)，(24)，(25)，(26) の計算を順番に繰り返す．

■ $\boldsymbol{\Sigma}_r$ の推論

次の事後分布にしたがって，$\boldsymbol{\Sigma}_r$ をサンプリングする．

$$p(\boldsymbol{\Sigma}_r^{-1}|X_r,\{z_t : z_t = r\}) = \mathcal{W}(\boldsymbol{\Sigma}_r^{-1};\boldsymbol{W}_r,\nu_r) \qquad (27)$$

$$\beta_r = \beta_0 + T_r,\ \nu_r = \nu_0 + T_r$$

$$\bar{\boldsymbol{x}}_r = \frac{1}{T_r}\sum_{i=1}^{T_r}\boldsymbol{x}_{r,i},\ (\boldsymbol{x}_{r,i} \in X_r)$$

$$\boldsymbol{W}_r^{-1} = \boldsymbol{W}_0^{-1} + \sum_{i=1}^{T_r}\boldsymbol{x}_{r,i}\boldsymbol{x}_{r,i}{}^{\mathrm{T}} + \frac{\beta_0 T_r}{\beta_r}(\bar{\boldsymbol{x}}_r - \boldsymbol{\mu}_0)(\bar{\boldsymbol{x}}_r - \boldsymbol{\mu}_0)^{\mathrm{T}}$$

■ $\boldsymbol{\mu}_r$ の推論

次の事後分布にしたがって，$\boldsymbol{\mu}_r$ をサンプリングする．

$$p(\boldsymbol{\mu}_r|X_r,\{z_t : z_t = r\}) = \mathcal{N}(\boldsymbol{\mu}_r;\boldsymbol{m}_r,\boldsymbol{\Sigma}_r/\beta_r) \qquad (28)$$

$$\boldsymbol{m}_r = \frac{1}{\beta_r}(\beta_0\boldsymbol{m}_0 + T_r\bar{\boldsymbol{x}}_r)$$

4. 評価実験

提案法の基本動作を，未知の楽譜に対する音量軌跡の予測性能の観点から評価する．音大声楽科出身の歌唱者 1 名が，J-pop に分類される合計 10 曲のサビの 4 小節を MIDI 伴奏を聴きながら歌った歌声（総時間 63.8 秒）を実験データとして利用する．歌唱者には，事前に原曲を 4 回聴きながら自由に練習してもらったため，収録された歌声はうろ覚えの状態ではない．サンプリング周波数は 16 kHz，量子化ビット数は 16 ビットで歌声を収録した．

まず，収録された歌声ごとに，式 (1) に基づいて，音符内位置を表す発音開始時刻からの時間 (sec.)，音符の音高 (MIDI ノートナンバ)，音符の音長 (sec.) からなる 3 次元ベクトルの入力変数 $\boldsymbol{x}_t$ を作成した（$x_{t,1}^{(\mathrm{p})}$ = 音符内位置，$x_{t,1}^{(\mathrm{c})}$ = 音高，$x_{t,2}^{(\mathrm{c})}$ = 音長 とする）．次に，式 (2) に基づいて出力変数となる音量を計算した．そして，**図 6** に示すように，収録された歌声ごとに，歌唱した音符数の割合が 6:4 となるように歌声を学習区間と評価区間に分けた．すべての歌

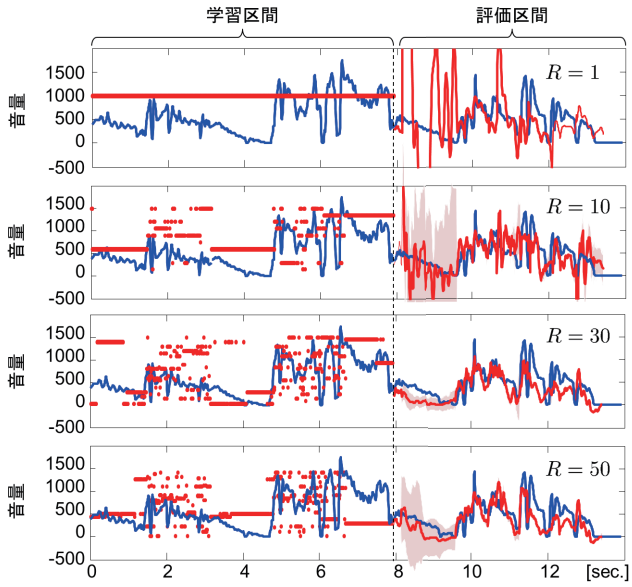

図 7 学習区間における音量軌跡の状態割り当て結果と評価区間における音量軌跡の予測結果：学習区間における同じ高さの赤点は同じ状態を表す．評価区間では予測分布を示す（赤線が予測分布の平均値系列，桃色は平均値から $\pm 2\sigma_*$ の範囲を表す）

Fig. 7 State assignments in training segments and predictive distributions in test segments

声の学習区間を用いてモデルパラメータを推定し，評価区間の入力変数から音量軌跡を予測する．

インジケータ変数 $z_1,\ldots,z_T$ の初期値は，すべての学習区間の入力変数を k-means クラスタリングし，R 個の状態に割り当てた結果を利用する．各状態に割り当てられた入力変数を用いて計算される平均と共分散行列を $\theta_r^{\boldsymbol{x}}$ の初期値とする．また，マルチカーネル学習におけるカーネル関数の個数は $M = 30$ と固定し，$w_r^2 = 100$，$\psi_{r,1} = 1/M,\ldots,\psi_{r,M} = 1/M$，$\eta_r^2 = 10$ $(r = 1,\ldots,R)$ を初期値とする．ここで，観測ノイズの分散値 η_r^2 は推論せず固定する．ハイパーパラメータは $\alpha = 1$，$\beta_0 = 0.1$，$\nu_0 = D+1$ とする．D は入力変数の次元数である．$\boldsymbol{m}_0$ は学習区間の入力変数全体の平均とする．$\boldsymbol{W}_0$ は入力変数全体から計算される共分散行列の逆行列を ν_0 で割った行列に設定する．位置カーネルのハイパーパラメータは，$l_{1:10}^{(\mathrm{p})} = l_{11:20}^{(\mathrm{p})} = l_{21:30}^{(\mathrm{p})}$ として，0.005 から 0.5 までを対数スケールで 10 等分した値を設定した．またコンテキストカーネルのハイパーパラメータは，$l_{1:10,1}^{(\mathrm{c})} = 1$，$l_{11:20,1}^{(\mathrm{c})} = 2$，$l_{21:30,1}^{(\mathrm{c})} = 3$，$l_{1:10,2}^{(\mathrm{c})} = 0.1$，$l_{11:20,2}^{(\mathrm{c})} = 0.2$，$l_{21:30,2}^{(\mathrm{c})} = 0.3$ と設定した．これらは，予備実験を通して決定した値である．パラメータのサンプリング回数は 100 回とし，$\boldsymbol{\psi}_r, w_r^2, \eta_r^2$ を推論するための EM アルゴリズムも 100 回繰り返した．

図 7 は状態数 R を変化させたときの，学習区間における音量軌跡の状態割り当て結果と評価区間における音量軌跡の予測結果を示す．学習区間における同じ高さの赤点は同じ状態に割り当てられたことを表す．評価区間には，式 (16) で計算される予測分布を示す．赤線は予測分布の平均

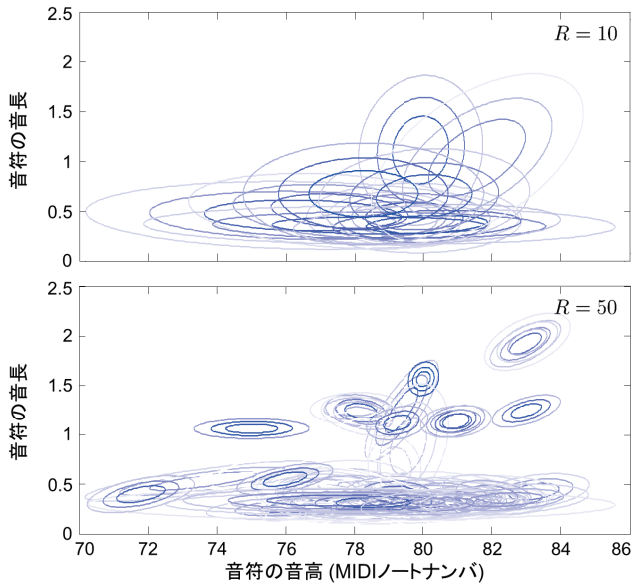

図 8 入力変数空間の一部（音符の音高と音長からなる平面）において学習されたガウス分布

Fig. 8 Gaussian distributions trained in input variable space

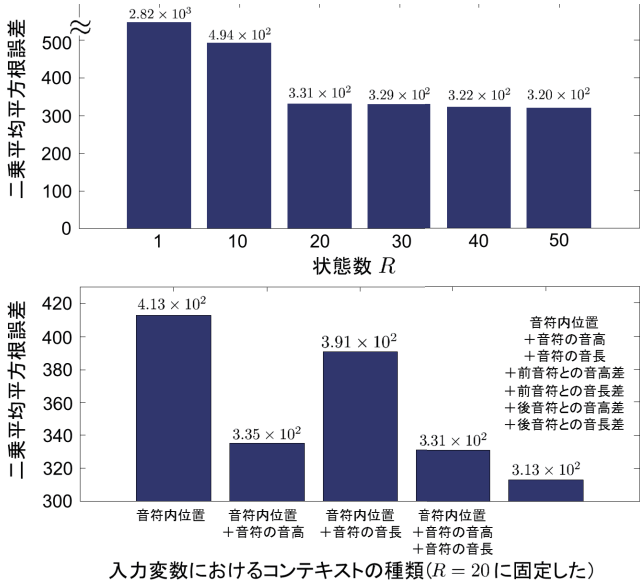

図 9 評価区間の音量軌跡と予測分布の平均値系列との二乗平均平方根誤差

Fig. 9 Root mean square errors between volume contours and predictive means

値 μ_* 系列，桃色は平均値から $\pm 2\sigma_*$ の範囲を表す．学習区間をみると，類似する動特性を含む部分が同じ状態に割り当てられることがわかる．ただし，急激に音量が変化する部分で状態が激しく切り替わるため，例えば，インジケータ変数の系列にマルコフ性を導入する（この変数の系列を HMM で表現する）ことなど検討の余地が残る．評価区間の観測される音量軌跡（青線）を正解と考えると，状態数 R を増やすにつれて，青線と赤線が近い結果となり，予測性能が向上すると言える．つまり，音量軌跡全体を単一のガウス過程で表現するよりも，動特性の違いを考慮していくつかの状態を用意し，各々のガウス過程の混合モデルで音量軌跡を表現することの有用性を定性的に評価できる．

図 8 は入力変数の一部である，音符の音高と音長からなる平面において，各状態から入力変数を生成するガウス分布の学習結果を，$R = 10$，$R = 50$ の場合について比較する．式 (14) を用いて分布を計算し，等高線で図示した．$R = 50$ の場合，各状態は個々の音符を表現するように散らばっており，そこから生成されうる音量軌跡の動特性は状態ごとに特徴づけられる．そのため，音量軌跡の予測性能も向上する．一方，状態数が少ない場合，入力変数空間における各状態のガウス分布が重なってしまい，動特性を特徴づけることが不十分であるため，予測性能が低下したと考えられる．また，計算量の観点からも，この状態数 R の設定は重要である．パラメータの推論のときに，$T_r \times T_r$ 行列の逆行列をはじめとする様々な代数演算を必要とするためである．式 (7) にディリクレ過程を導入する（$R \to \infty$ とする）ことで，モデル選択パラメータである状態数を，学習データから自動的に推論する枠組みに拡張できるが，これは今後の課題とする．

提案法を定量的に評価するために，評価区間における式 (16) の予測分布の平均値系列と，観測された音量との二乗平均平方根誤差 (RMSE) を計算した結果を**図 9** に示す．状態数を増やすにつれて RMSE が低下し，今回の学習データに対しては $R = 50$ のときに最も値が小さい．この状態数のときに音量軌跡の予測性能が最も高いと考える．図 9 の下図では，入力変数のコンテキストの種類を変化させたときの RMSE を示す．音符内位置と当該音符の音高や音長を単独や組合せで扱うよりも，すべてを入力変数に利用した場合に RMSE が小さくなり，予測性能が向上した．特に，音符の音長よりも音高の方が音量の予測に効果的である結果が得られたことは興味深い．さらに，前後の音符との音高差や音長差を入力変数に導入することの有効性も確認できた．ここで，$x^{(c)}_{t,3}$ = 前の音符との音高差，$x^{(c)}_{t,4}$ = 前の音符との音長差，$x^{(c)}_{t,5}$ = 後の音符との音高差，$x^{(c)}_{t,6}$ = 後の音符との音長差 とした．より広範囲にわたる音符の並びや強弱記号，様々な演奏記号を 2 値で入力変数に導入して評価することが今後の課題である．

5. おわりに

本稿では，混合ガウス過程を利用して，歌声音量軌跡の動特性を特徴付け，未知の楽譜に対して，その音量軌跡を予測できる生成過程モデルを提案した．また，そのモデルパラメータの推論アルゴリズムと新たな入力変数に対する予測分布を導出した．評価実験では，予測性能の観点から，多様な動特性を表現するために混合ガウス過程を導入することの有用性を確認した．今後の課題は，学習データを増やしながら，より広範囲にわたる音符の並びや強弱記号，

演奏記号などを入力変数に導入することである．また，インジケータ変数の系列を HMM で表現することや，ディリクレ過程に基づいて状態数を学習データから推論すること，カーネル関数の優勢度 $\psi_{r,m}$ から動特性に関する歌唱者の個性を見出すことも課題として挙げられる．最終的には，提案法を歌声の $F0$ 軌跡やスペクトログラムにも適用し，歌声の認識や合成に応用することを目指している．

付録 A

式 (19) における，積分を含む後半部分の具体的な計算方法を示す．対数をとると，

$$\log\frac{\int p(\boldsymbol{x}_t|\theta_r^{\boldsymbol{x}})\prod_{i:z_i=r,i\neq t}p(\boldsymbol{x}_i|\theta_r^{\boldsymbol{x}})p(\theta_r^{\boldsymbol{x}})d\theta_r^{\boldsymbol{x}}}{\int\prod_{i:z_i=r,i\neq t}p(\boldsymbol{x}_i|\theta_r^{\boldsymbol{x}})p(\theta_r^{\boldsymbol{x}})d\theta_r^{\boldsymbol{x}}} \tag{29}$$
$$\stackrel{c}{=} -\log\frac{B(\boldsymbol{W}_{r,t},\nu_{r,t})}{B(\boldsymbol{W}_{r,\backslash t},\nu_{r,\backslash t})} - \frac{D}{2}\log\frac{\beta_0+T_{r,\backslash t}+1}{\beta_0+T_{r,\backslash t}}$$

と書ける．$B(\boldsymbol{W},\nu)$ はウィシャート分布の正規化項である．

$$B(\boldsymbol{W},\nu)\equiv|\boldsymbol{W}|^{-\frac{\nu}{2}}\Big(2^{\frac{\nu D}{2}}\pi^{\frac{D(D-1)}{4}}\prod_{i=1}^{D}\Gamma\big(\frac{\nu+1-i}{2}\big)\Big)^{-1}$$

式 (29) を計算するために，以下の値を算出する．

$$\beta_{r,t}=\beta_0+T_{r,\backslash t}+1,\ \beta_{r,\backslash t}=\beta_0+T_{r,\backslash t}$$
$$\nu_{r,t}=\nu_0+T_{r,\backslash t}+1,\ \nu_{r,\backslash t}=\nu_0+T_{r,\backslash t}$$
$$\bar{\boldsymbol{x}}_{r,t}=\frac{1}{T_{r,\backslash t}+1}\Big(\boldsymbol{x}_t+\sum_{i=1}^{T_{r,\backslash t}}\boldsymbol{x}_{r,i}\Big),\ \bar{\boldsymbol{x}}_{r,\backslash t}=\frac{1}{T_{r,\backslash t}}\sum_{i=1}^{T_{r,\backslash t}}\boldsymbol{x}_{r,i}$$
$$\boldsymbol{W}_{r,t}^{-1}=\boldsymbol{W}_0^{-1}+\sum_{i=1}^{T_{r,\backslash t}}\boldsymbol{x}_{r,i}\boldsymbol{x}_{r,i}{}^{\mathrm{T}}+\boldsymbol{x}_t\boldsymbol{x}_t^{\mathrm{T}}+\frac{\beta_0(T_{r,\backslash t}+1)}{\beta_{r,t}}(\bar{\boldsymbol{x}}_{r,t}-\boldsymbol{\mu}_0)(\bar{\boldsymbol{x}}_{r,t}-\boldsymbol{\mu}_0)^{\mathrm{T}}$$
$$\boldsymbol{W}_{r,\backslash t}^{-1}=\boldsymbol{W}_0^{-1}+\sum_{i=1}^{T_{r,\backslash t}}\boldsymbol{x}_{r,i}\boldsymbol{x}_{r,i}{}^{\mathrm{T}}+\frac{\beta_0T_{r,\backslash t}}{\beta_{r,\backslash t}}(\bar{\boldsymbol{x}}_{r,\backslash t}-\boldsymbol{\mu}_0)(\bar{\boldsymbol{x}}_{r,\backslash t}-\boldsymbol{\mu}_0)^{\mathrm{T}}$$

ここで，$\boldsymbol{x}_{r,i}\in X_{r,\backslash t}$ とし，D は入力変数の次元数を表す．

参考文献

[1] 河原英紀ほか：高品質音声分析変換合成システム STRAIGHT を用いたスキャット生成研究の提案，情報処理学会論文誌，Vol. 43, No. 2, pp. 208–218 (2002).
[2] Nakano, T. et al.: An Automatic Singing Skill Evaluation Method for Unknown Melodies Using Pitch Interval Accuracy and Vibrato Features, *in Proc. ICSLP 2006*, pp. 1706–1709 (2006).
[3] Mayor, O. et al.: The Singing Tutor: Expression Categorization and Segmentation of the Singing Voice, *in Proc. AES 121st Convention* (2006).
[4] 中野倫靖ほか：VocaListener：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).
[5] Bonada, J. et al.: Synthesis of the Singing Voice by Performance Sampling and Spectral Models, *IEEE Signal Processing Magazine*, Vol. 24, pp. 67–79 (2007).
[6] Kako, T. et al.: Automatic Identification for Singing Style based on Sung Melodic Contour Characterized in Phase Plane, *in Proc. ISMIR 2009*, pp. 393–397 (2009).
[7] Fukayama, S. et al.: Orpheus: Automatic Composition System Considering Prosody of Japanese Lyrics, *in Proc. ICEC 2009*, pp. 309–310 (2009).
[8] Nakano, T. et al.: VocaListener2: A Singing Synthesis System Able to Mimic a User's Singing in terms of Voice Timbre Chages as well as Pitch and Dynamics, *in Proc. ICASSP 2011*, pp. 453–456 (2011).
[9] Mase, A. et al.: HMM-based Singing Voice Synthesis System Using Pitch-shifted Pseudo Training Data, *in Proc. INTERSPEECH 2010*, pp. 845–848 (2010).
[10] Sundberg, J.: *The Science of the Singing Voice*, Northern Illinois University Press, Illinois (1987).
[11] Proctor, D. F.: *Breathing, Speech and Song*, Springer-Verlag, New York (1980).
[12] Bouhuys, A. et al.: Kinetic Aspects of Singing, *Journal of Applied Physiology*, Vol. 21, pp. 483–496 (1966).
[13] Sundberg, J.: Activation of the Diaphragm in Singing, *in Proc. SMAC 1983*, pp. 279–290 (1983).
[14] Tsui, W. H. et al.: Method and System on Detecting Abdominals for Singing, *in Proc. IEEE EMBC 2013* (2013).
[15] Rubin, H. J. et al.: Vocal Intensity, Subglottic Pressure and Air Flow Relationships in Singers, *Folia Phoniatr*, Vol. 19, No. 6, pp. 393–413 (1967).
[16] Cleveland, T. et al.: Acoustic Analysis of Three Male Voices of Different Quality, *in Proc. SMAC 1983*, pp. 143–156 (1983).
[17] Scully, C. et al.: Simulation of Singing with a Composite Model of Speech Production, *in Proc. SMAC 1983*, pp. 247–260 (1983).
[18] Teramura, K. et al.: Gaussian Process Regression for Rendering Music Performance, *in Proc. ICMPC 2008* (2008).
[19] 小泉悠馬ほか：演奏音の音量時系列からの奏者の意図表現成分の推定，情報処理学会第 75 回全国大会, 3R-7，pp. 257–258 (2013).
[20] Kameoka, H. et al.: A Statistical Model of Speech F0 Contours, *in Proc. SAPA 2010*, pp. 43–48 (2010).
[21] Ohishi, Y. et al.: A Stochastic Model of Singing Voice F0 Contours for Characterizing Expressive Dynamic Components, *in Proc. INTERSPEECH 2012* (2012).
[22] Koriyama, T. et al.: Frame-level Acoustic Modeling based on Gaussian Process Regression for Statistical Nonparametric Speech Synthesis, *in Proc. ICASSP 2013*, pp. 8007–8011 (2013).
[23] Meeds, E. et al.: An Alternative Infinite Mixture of Gaussian Process Experts, *in Proc. NIPS 2006* (2006).
[24] Pilkington, N. C. V. et al.: Gaussian Process Experts for Voice Conversion, *in Proc. INTERSPEECH 2011*, pp. 2761–2764 (2011).
[25] 亀岡弘和ほか：マルチカーネル線形予測モデルによる音声分析，音講論集, 2-Q-24，pp. 499–502 (2010).
[26] Henter, G. E. et al.: Gaussian Process Dynamical Models for Nonparametric Speech Representation and Synthesis, *in Proc. ICASSP 2012*, pp. 4505–4508 (2012).
[27] Rasmussen, C. E. et al.: Infinite Mixtures of Gaussian Process Experts, *in Proc. NIPS 2002* (2002).
[28] Andrieu, C. et al.: An Introduction to MCMC for Machine Learning, *Machine Learning*, Vol. 50, No. 1-2, pp. 5–43 (2003).