# ガウス過程回帰の混合エキスパートモデルを用いた歌声 $F_0$ 軌跡の予測と生成*

◎大石康智 (NTT), 持橋大地 (統数研), 亀岡弘和, 柏野邦夫 (NTT)

## 1　はじめに

本研究の目的は，歌声の基本周波数の時間変化（$F_0$ 軌跡）に表れる歌唱表現を特徴抽出し，任意の楽譜に対して，その歌唱表現を反映した $F_0$ 軌跡を予測可能な生成過程モデルを構築することである。歌唱表現の定義はまだ十分になされていないが，歌唱表現が歌唱フォルマントや $F_0$ 動的変動成分のような信号特徴と関連付けられることが明らかにされた [1]。特に，ビブラートやポルタメントのような $F_0$ 動的変動成分は知覚的な観点からも，歌唱表現と密接に関係する [2]。ビブラートは，持続音において，$F_0$ が準周期的に変調される成分であり，速度，振幅，継続長，波形，規則性などによって特徴付けられる。一方，ポルタメントは，音符の切り替わりにおいて，$F_0$ が緩やかに変化する成分である [3]。歌唱者は，楽譜のコンテキスト（音符の高さや長さなど）が与えられた下で，これらの動的変動成分を巧みに制御して，歌声に表情を付ける。このような生成過程を計算機で実現できれば，歌唱者の個性や表現力を学習することにつながり，歌声の認識や合成 [4–6] への応用を期待できる。例えば，ある歌唱者の大量の歌声から歌唱表現を学習することによって，どんな新しい楽譜に対しても，その表現を付与して合成することが可能となるだろう。

これまで個々の $F_0$ 動的変動成分を特徴抽出して，歌声の認識や合成，歌唱力評価に利用する試みはなされてきたものの [2, 3, 7–9]，動的変動成分を楽譜のコンテキストと結びつけて，体系的にその動特性を学習する方法は十分に検討されていなかった。隠れマルコフモデル (HMM) に基づく歌声合成のように，コンテキストラベルが付与された区間毎に，例えば，5 状態の HMM を用いて $F_0$ の動特性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，短時間に細かく変動する歌声の $F_0$ を，固定された状態数と局所的な動的特徴で表現することは難しいとされる [10]。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，実際に生成される $F_0$ 軌跡は過剰に平滑化され，本来の動特性を再現できないという問題もある。

本稿では，楽譜のコンテキストから歌声の $F_0$ 軌跡が生成される入出力関係を，ガウス過程回帰の混合エキスパートモデル [11] を利用して直接的にモデル化する。混合エキスパートモデルは混合モデルの一つの拡張であり，混合係数それ自身を入力変数の関数（ゲート関数と呼ぶ）で表現する。すなわち，入力空間（楽譜のコンテキストからなる空間）の異なる領域ごとに，個々の混合要素の密度関数（エキスパートと呼ばれ，ここではガウス過程回帰に相当する）が $F_0$ 軌跡の予測を行う。ゲート関数はいずれの混合要素がどの領域において優勢であるかを判定する。これにより，歌唱者がコンテキストに依存して，歌唱表現に起因する $F_0$ 動的変動成分を多様に使い分ける動作を表現できると考える。さらに，動的変動成分を精緻に特徴づけるために，個々のガウス過程回帰におけるカーネル関数の設計を工夫する。評価実験では，新規の楽譜に対する $F_0$ 軌跡の予測の観点から提案法を評価する。また，モデルの学習データ量について検討する。

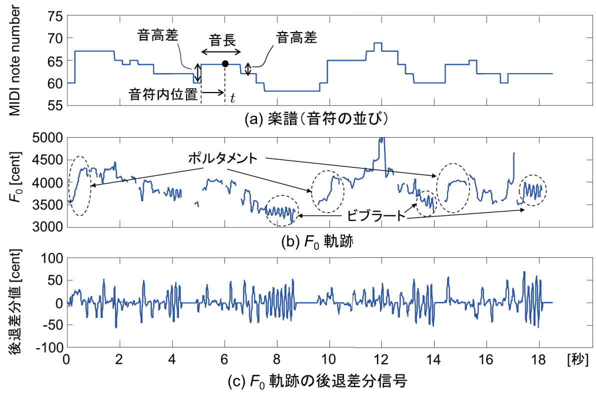


Fig. 1　メロディの楽譜とそれを歌った歌声の $F_0$ 軌跡: MIDI 信号からノートナンバと音長を算出してメロディの楽譜とした。

## 2　歌声 $F_0$ 軌跡の生成過程モデル

メロディの楽譜とその歌声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 $F_0$ 軌跡への回帰問題を考える。入力変数は楽譜に基づいて，例えば下記のように構成される。

$$\boldsymbol{x}_t = [\ \text{音符内位置（発音開始時刻からの時間）, 音長, 先行音符との音高差, 後続音符との音高差}\ ]^{\mathrm{T}}$$

もちろん，歌詞の音素情報やクレッシェンドのような演奏情報の有無などを入力変数に加えることも可能である。一方で，出力変数 $\{y_t\}_{t=1}^{T}$ には，$F_0$ 軌跡から音符情報を取り除いた後退差分信号を利用する。差分信号を扱うことで，Fig. 1 に示すように，動的変動成分を顕著に観測することができる。$F_0$ は YIN[12]

*Mixture of Gaussian process experts for predicting and generating sung melodic contour. by OHISHI, Yasunori (NTT), MOCHIHASHI, Daichi (ISM), KAMEOKA, Hirokazu, KASHINO, Kunio (NTT)

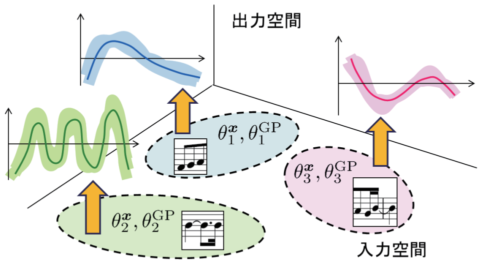


Fig. 2　MoGPEs における入力空間の割り当て: 各楕円領域は全共分散型ガウス分布を表し，この領域では同じ GPR によって出力が予測される。

を利用して 10 ms ごとに推定され，下記のように Hz で表される周波数単位を cent に変換する。

$$y_{\mathtt{cent}} = 1200 \log_2 \frac{y_{\mathtt{Hz}}}{440 \times 2^{\frac{3}{12}-5}} \tag{1}$$

まず，ガウス過程回帰 (Gaussian process regression, GPR) を説明する。学習データ $\mathcal{D} = \{\boldsymbol{x}_t, y_t\}_{t=1}^T$ が与えられ，出力変数を並べたベクトル $\boldsymbol{y} = [y_1, \ldots, y_T]^{\mathrm{T}}$ を GPR によって回帰する場合，その確率密度関数は

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

のような多次元ガウス分布で表現される。ここで，$\boldsymbol{K}$ 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 変数間の類似性を定義するカーネル関数である。また，$\eta^2$ と $\boldsymbol{I}$ はそれぞれ，観測雑音の分散値と単位行列を表す。GPR は新しい入力変数 $\boldsymbol{x}_*$ が与えられたときの出力変数 $y_*$ の予測分布を推論できる。予測分布は，

$$\begin{aligned} p(y_*|\mathcal{D}, \boldsymbol{x}_*) &= \mathcal{N}(y_*; \mu_*, \sigma_*^2), \\ \mu_* &= \boldsymbol{k}_*^{\mathrm{T}} (\boldsymbol{K} + \eta^2 \boldsymbol{I})^{-1} \boldsymbol{y}, \\ \sigma_*^2 &= k(\boldsymbol{x}_*, \boldsymbol{x}_*) - \boldsymbol{k}_*^{\mathrm{T}} (\boldsymbol{K} + \eta^2 \boldsymbol{I})^{-1} \boldsymbol{k}_* \end{aligned} \tag{3}$$

となり，$\boldsymbol{k}_*$ は $k(\boldsymbol{x}_t, \boldsymbol{x}_*)$ $(t = 1, \ldots, T)$ からなるベクトルを表す。学習データが “事例” となって，新たな入力変数に対する予測に貢献することがわかる。

カーネル関数の必要条件はグラム行列が半正定値対称行列となることである。出力信号が定常であることを仮定し，Squared exponential (SE) 共分散関数が一般的に利用されるが，Fig. 1 に示すように，差分信号は必ずしも定常ではない。歌唱者は音符の並びに依存して，様々な動的変動成分を使い分けながら歌唱するためである。このような非定常な動特性をモデル化するために，ガウス過程回帰の混合エキスパートモデル (Mixture of Gaussian process experts, MoGPEs) を利用する。

MoGPEs では，入力空間がゲート関数によって確率的に分割される (Fig. 2)。それぞれの領域では，異なる GPR によって，動的変動成分が特徴付けられて，出力が予測される。最終的に，出力は個々のゲート関数と GPR との混合モデルとして表現される。MoGPEs はこれまで，2 つの構成方法 [11, 13] が提案されているが，文献 [11] は MoGPEs の完全な生成モデルを定義し，入力変数の欠損や出力変数から入力変数への逆予測が可能であるため，この構成方法を利用する。生成過程の流れを以下に示す。

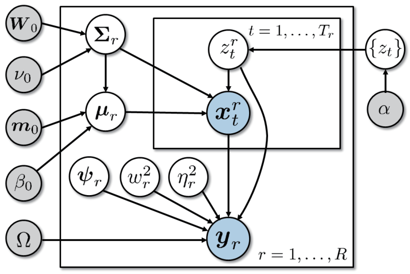


Fig. 3　提案法のグラフィカル表現

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 $R$ 個の領域のいずれかに割り当てる。これは指示変数集合 $\{z_t\}_{t=1}^T$ によって表現される。
2. 指示変数の集合 $\{z_t : z_t = r\}$ が与えられた下で，領域 $r$ の入力変数の密度分布パラメータ $\theta_r^{\boldsymbol{x}} \equiv \{\boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r\}$ が生成される。 ここで，密度分布は全共分散型ガウス分布を仮定する。
3. $\theta_r^{\boldsymbol{x}}$ が与えられた下で，領域 $r$ に属する入力変数集合 $X_r \equiv \{\boldsymbol{x}_t : z_t = r\}$ が生成される。
4. カーネル関数のパラメータ $\theta_r^{\mathrm{GP}} \equiv \{\boldsymbol{\psi}_r, w_r^2, \eta_r^2\}$ が領域毎に生成される。
5. 領域毎に $X_r$ と $\theta_r^{\mathrm{GP}}$ を使ってグラム行列が計算され，出力変数を並べたベクトル $\boldsymbol{y}_r \equiv \{y_t : z_t = r\}$ が生成される。

Fig. 3 はこのグラフィカル表現であり，同時分布は，

$$\begin{aligned} &p(\{\boldsymbol{x}_t, y_t\}_{t=1}^T, \{z_t\}_{t=1}^T, \{\theta_r^{\mathrm{GP}}\}_{r=1}^R, \{\theta_r^{\boldsymbol{x}}\}_{r=1}^R | \Omega) \\ &= \prod_{r=1}^R \left[ p(\theta_r^{\boldsymbol{x}}|\Omega) p(X_r|\theta_r^{\boldsymbol{x}}) p(\boldsymbol{y}_r|X_r, \theta_r^{\mathrm{GP}}, \Omega) \right] \times p(\{z_t\}_{t=1}^T|\Omega) \end{aligned} \tag{4}$$

と書ける。$R$ と $\Omega$ はそれぞれ，エキスパートの総数と超パラメータ集合を表す。本稿では，指示変数集合 $\{z_t\}_{t=1}^T$ を積分消去せず直接表現して，これらの依存関係を考慮する。式 (4) における個々の分布を以下に示す。

$$\begin{aligned} p(\{z_t\}_{t=1}^T|\Omega) &= \frac{\Gamma(\alpha)}{\Gamma(\alpha + T)} \prod_{r=1}^R \frac{\Gamma(T_r + \alpha/R)}{\Gamma(\alpha/R)} \\ p(\theta_r^{\boldsymbol{x}}|\Omega) &= \mathcal{N}(\boldsymbol{\mu}_r; \boldsymbol{m}_0, \boldsymbol{\Sigma}_r/\beta_0) \mathcal{W}(\boldsymbol{\Sigma}_r^{-1}; \boldsymbol{W}_0, \nu_0) \\ p(X_r|\theta_r^{\boldsymbol{x}}) &= \mathcal{N}(X_r; \boldsymbol{\mu}_r, \boldsymbol{\Sigma}_r) \\ p(\boldsymbol{y}_r|X_r, \boldsymbol{\theta}_r^{\mathrm{GP}}, \Omega) &= \mathcal{GP}\left(\boldsymbol{y}_r; \boldsymbol{0}, \boldsymbol{K}_r + \eta_r^2 \boldsymbol{I}_r\right) \end{aligned} \tag{5}$$

ここで，$\alpha, T_r, \boldsymbol{I}_r, \eta_r^2, \mathcal{W}$ はそれぞれ，ディリクレー多項分布の超パラメータ，$X_r$ の要素数，$T_r \times T_r$ の単位行列，観測雑音の分散値，ウィシャート分布を示す。

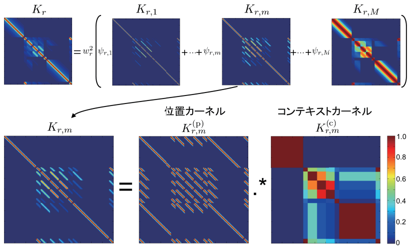


Fig. 4　マルチカーネル学習に基づくグラム行列表現

本稿では，マルチカーネル学習に基づき，Fig. 4 のような複数のカーネルの線形結合

$$k_r(\boldsymbol{x}_i, \boldsymbol{x}_j) = w_r^2 \sum_{m=1}^{M} \psi_{r,m} k_{r,m}(\boldsymbol{x}_i, \boldsymbol{x}_j) \qquad (6)$$

でグラム行列 $\boldsymbol{K}_r$ を表現し，強度 $w_r^2$ 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 [14]。ここで，$\boldsymbol{x}_i, \boldsymbol{x}_j \in X_r$，$\sum_{m=1}^{M} \psi_{r,m} = 1$ とする。$M$ は線形結合するカーネル関数の総数である。単位の異なる様々なコンテキスト（音符内位置や音高差，音長など）を扱うため，個々のカーネル関数 $k_{r,m}(\boldsymbol{x}_i, \boldsymbol{x}_j)$ を 2 つのカーネル関数の積で表現する [15]。

$$k_{r,m}(\boldsymbol{x}_i, \boldsymbol{x}_j) = k_{r,m}^{(\mathrm{p})}(x_i^{(\mathrm{p})}, x_j^{(\mathrm{p})}) k_{r,m}^{(\mathrm{c})}(\boldsymbol{x}_i^{(\mathrm{c})}, \boldsymbol{x}_j^{(\mathrm{c})}) \qquad (7)$$

ここで，$k_{r,m}^{(\mathrm{p})}(x_i^{(\mathrm{p})}, x_j^{(\mathrm{p})})$ と $k_{r,m}^{(\mathrm{c})}(\boldsymbol{x}_i^{(\mathrm{c})}, \boldsymbol{x}_j^{(\mathrm{c})})$ はそれぞれ，音符内位置の類似を表す位置カーネルと音符の類似度を表すコンテキストカーネルと呼ぶ。入力変数ベクトル $\boldsymbol{x}_i$ を $x_i^{(\mathrm{p})}, \boldsymbol{x}_i^{(\mathrm{c})}$ のように位置に関する変数と音符のコンテキストに関する変数に分け，カーネル関数を計算する。位置カーネルは動的変動成分の時間的連続性や周期性を捉えるために，SE 共分散関数と周期共分散関数を利用する。

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

$$k_{r,m}^{(\mathrm{p})}(x_i^{(\mathrm{p})}, x_j^{(\mathrm{p})}) = \exp\left(-2\sin^2\left(\frac{l_m^{(\mathrm{p})}}{2\pi}(x_i^{(\mathrm{p})} - x_j^{(\mathrm{p})})\right)\right)$$

コンテキストカーネルにはSE共分散関数を利用する。

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

$$\boldsymbol{\Lambda}^{-1} = \mathrm{diag}({l_{m,1}^{(\mathrm{c})}}^2, {l_{m,2}^{(\mathrm{c})}}^2, \ldots, {l_{m,D_\mathrm{c}}^{(\mathrm{c})}}^2)$$

まとめると，パラメータと超パラメータはそれぞれ，$\Theta = \{z_1, \ldots, z_T, \theta_1^{\boldsymbol{x}}, \ldots, \theta_R^{\boldsymbol{x}}, \theta_1^{\mathrm{GP}}, \ldots, \theta_R^{\mathrm{GP}}\}$，$\Omega = \{\alpha, \boldsymbol{m}_0, \boldsymbol{W}_0, \beta_0, \nu_0, l_1^{(\mathrm{p})}, \ldots, l_M^{(\mathrm{p})}, l_{1,1}^{(\mathrm{c})}, \ldots, l_{M,D_\mathrm{c}}^{(\mathrm{c})}\}$ となる。$D_\mathrm{c}$ は $\boldsymbol{x}_i^{(\mathrm{c})}$ の次元数を表す。

新たな入力変数 $\boldsymbol{x}_*$ に対する出力変数 $y_*$ の予測分布を考える。式 (3) を参考にすると，MoGPEs では，

$$\begin{aligned} &p(y_*|\mathcal{D}, \boldsymbol{x}_*, \Theta, \Omega) \qquad (8) \\ &= \sum_{r=1}^{R} p(y_*|\boldsymbol{y}_r, X_r, \boldsymbol{x}_*, z_* = r, \theta_r^{\mathrm{GP}}) p(z_* = r|\boldsymbol{x}_*, \theta_r^{\boldsymbol{x}}) \\ &= \mathcal{N}(y_*; \mu_*, \sigma_*^2) \end{aligned}$$

となり，右辺の第 1 式が GPR とゲート関数の混合モデルとなることがわかる。ここで，

$$\mu_* = \sum_{r=1}^{R} c_r \boldsymbol{k}_{r,*}^{\mathrm{T}} (\boldsymbol{K}_r + \eta_r^2 \boldsymbol{I}_r)^{-1} \boldsymbol{y}_r$$

$$\sigma_*^2 = \sum_{r=1}^{R} c_r^2 (k_r(\boldsymbol{x}_*, \boldsymbol{x}_*) - \boldsymbol{k}_{r,*}^{\mathrm{T}} (\boldsymbol{K}_r + \eta_r^2 \boldsymbol{I}_r)^{-1} \boldsymbol{k}_{r,*})$$

$$c_r = p(z_* = r|\boldsymbol{x}_*, \theta_r^{\boldsymbol{x}})$$

であり，ゲート関数 $c_r$ はベイズの定理を利用して容易に計算できる。この予測分布に基づいて，$F_0$ 軌跡 $\{f_t\}_{t=1}^{T_m}$ を以下のように音符毎に再合成する。

$$f_t = o_t + c - \bar{o},\ o_t = \sum_{n=1}^{t} \mu_{*,n},\ \bar{o} = \sum_{t=1}^{T_m} o_t / T_m \qquad (9)$$

ここで，$c$ と $T_m$ はそれぞれ，楽譜に記載される音符の音高と音長を表す。

GPR はこれまで音声合成 [15] や $F_0$ 生成 [16]，声質変換 [17] などに適用されており，合成音の観点から性能の向上が確認されている。提案法はこれらの研究と関連するが，動的変動成分の動特性を精緻に特徴づけるために，MoGPEs と多様なカーネル関数を利用することが特徴的である。また，MoGPEs を利用すれば，通常の GPR よりも計算コストを低減できることも特筆すべき点である [11]。

## 3　パラメータの推論

MCMC-EM アルゴリズムを用いて，すべてのパラメータを推論する。具体的には，$\{z_t\}_{t=1}^{T}$ と $\{\theta_r^{\boldsymbol{x}}\}_{r=1}^{R}$ の推論のためにギブスサンプリングを利用する [11]。また，式 (5) の多次元ガウス分布を独立なガウス分布の和とみなして，$\{\theta_r^{\mathrm{GP}}\}_{r=1}^{R}$ の推論のために EM アルゴリズムを利用する [14]。

## 4　評価実験

新たな入力（楽譜）に対する出力（$F_0$ 軌跡の差分信号）の予測性能の観点から提案法を評価する。「RWC 研究用音楽データベース: ポピュラー音楽」(RWC-MDB-P-2001) における 10 楽曲 (No. 38, 39, 42, 44, 45, 46, 64, 72, 74, 76) の歌声の $F_0$ 軌跡を手作業でラベル付けした結果 [18] を利用する。また，その MIDI 信号を楽譜として利用する。これらの楽曲はすべて同一歌手によって歌われたものである。本来ならばこれらの楽曲の音響信号から $F_0$ を推定すべきであるが，今回は提案法の性能の上限を調べるためにこのようなデータを用いた。$\boldsymbol{x}_t$ と $F_0$ の後退差分値 $y_t$ は 10 ms ごとに計算され，無声音のため $F_0$ が観測されない区間は除去した。楽曲 No. 38, 39, 42, 44, 45 はモデルパラメータの学習に，楽曲 No. 46, 64, 72, 74, 76 は予測性能の評価に利用した。学習データと評価データ

Table 1　RMSE に基づく予測性能の比較

| $R$ | 10 | 20 | 30 | 40 | 50 |
|---|---|---|---|---|---|
| MoPRs | 71.8 | 59.7 | 73.1 | 79.6 | 56.5 |
| MoGPEs | **25.0** | **24.0** | **23.9** | **23.1** | **22.3** |

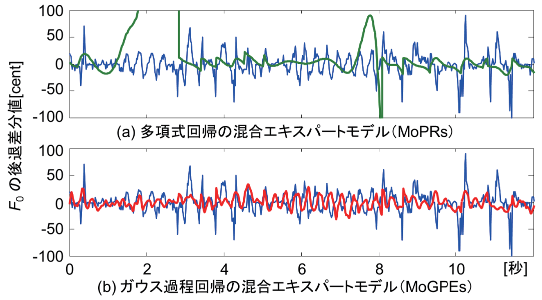


Fig. 5　予測結果の比較

の量はそれぞれ，649.3 秒，625.6 秒である。 指示変数の初期値は，学習データの入力ベクトル集合 $\{\boldsymbol{x}_t\}$ を k-means クラスタリングすることによって得られた割当て結果を利用した。各領域に割当てられた入力ベクトルを利用して，$\theta_r^{\boldsymbol{x}}$ を初期化した。$M = 216$ とし，$\theta_r^{\mathrm{GP}}$ の初期値はそれぞれ，$w_r^2 = 100$, $\psi_{r,1} = 1/M, \ldots, \psi_{r,M} = 1/M$, $\eta_r^2 = 10$ $(r = 1, \ldots, R)$ とした。超パラメータの設定を脚注に示す[1]。パラメータの推論回数は 50 回とした。

評価尺度として，実際の出力値系列と式 (8) の予測平均値系列間の二乗平均平方根誤差 (Root mean square error, RMSE) を利用した。

$$\mathrm{RMSE} = \sqrt{\textstyle\sum_{t=1}^{T_t}(y_t - \mu_{*,t})^2/T_t}, \qquad (10)$$

ここで，$T_t$ は評価データの長さを表す。比較として，多項式回帰の混合エキスパートモデル (Mixture of polynomial regression experts, MoPRs) を定義した。式 (5) の GPR が $n$ 次の多項式回帰に置き換わる。

Tab. 1 はエキスパートの数 $R$ を変化させたときの RMSE の平均値を示す。この結果は楽曲 No. 35 だけを利用してパラメータを学習し，評価した結果である。MoPRs の多項式の次数は $n = 8$ とした。これより高次の多項式を利用した場合，多項式の重み係数の推定が安定しなかったためである。RMSE の観点から，提案法は多項式回帰からなる単純な手法と比較して，予測性能が上回ることを確認した。また $R$ を増やすにつれて予測性能は向上した。Fig. 5 は 2 つの手法の予測結果を比較する。動的変動成分の動特性を表現するために，周期カーネルを利用する提案法の有効性を確認できた。Tab. 2 は学習データの量に対する予測性能を検証する。$R = 50$ に固定した。実験結果より，学習データを増やすに連れて RMSE が小さくなることを確認した。Fig. 6 は式 (9) によって再合成される $F_0$ 軌跡を示す。正解となる実際の $F_0$ 軌跡とほぼ同じ軌跡が得られるが，細かい動特性に関してはまだ予測が十分でない。予測性能を高めるために，エキスパートの数や学習データ量，入力コンテキストの種類を今後さらに検討する予定である。

[1] 入力ベクトルの次元数 $D$ は 4 であり，$\alpha = 1$, $\beta_0 = 0.1$, $\nu_0 = D + 1$ に固定する。$\boldsymbol{m}_0$ と $\boldsymbol{W}_0$ はそれぞれ，学習データのすべての入力ベクトルの平均ベクトルおよび共分散行列の逆行列を $\nu_0$ で割算した行列に設定する。位置カーネルの超パラメータとして，SE カーネルは $\{l_m^{(\mathrm{p})}, m = 1, \ldots, 108 | 0.05, 0.11, 0.23, 0.5\}$，周期カーネルは $\{l_m^{(\mathrm{p})}, m = 109, \ldots, 216 | 0.13, 0.15, 0.17, 0.2\}$ から選択した。コンテキストカーネルの超パラメータは $\{l_{m,1}^{(\mathrm{c})} | 1, 2.2, 5\}$, $\{l_{m,2}^{(\mathrm{c})} | 1, 2.2, 5\}$，$\{l_{m,3}^{(\mathrm{c})} | 0.1, 0.55, 3\}$ の組合せで設定した。

Table 2　学習データの量に対する RMSE

| 学習データの曲数 | 1 | 3 | 5 |
|---|---|---|---|
| MoGPEs | 22.3 | 20.7 | **20.5** |

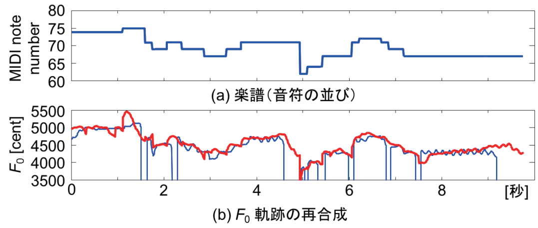


Fig. 6　メロディの楽譜とそれを歌った歌声の $F_0$ 軌跡

## 5　おわりに

歌声の $F_0$ 軌跡に表れる様々な動的変動成分を特徴抽出して歌唱者の歌唱表現を学習し，任意の楽譜に対して，その歌唱表現を反映した $F_0$ 軌跡を予測する生成過程モデルを提案した。実験結果より，単純な多項式回帰を用いるよりも GPR の混合エキスパートモデルを利用することの有効性を確認した。予測性能を高めるために，学習データを増やすこと，エキスパートの数や超パラメータを調整すること，エキスパートの割当てにディリクレ過程を導入することが挙げられる。この枠組を歌唱者認識や歌唱表現認識，歌声合成に応用することも検討中である。

## 参考文献


[1] Sundberg, “The Science of the Singing Voice” Northern Illinois University Press, 1987.
[2] Saitou *et al.*, in *Proc. WASPAA 2007.*
[3] Regnier, “Localization, Characterization and Recognition of Singing Voices” IRCAM / UPMC in Paris, France, 2013.
[4] Oura *et al.*, in *Proc. ICASSP 2012.*
[5] Nose *et al.*, in *Proc. INTERSPEECH 2013.*
[6] Doi *et al.*, in *Proc. INTERSPEECH 2013.*
[7] Nakano *et al.*, in *Proc. INTERSPEECH 2006.*
[8] Migita *et al.*, in *Proc. ICA 2010.*
[9] Ohishi *et al.*, in *Proc. INTERSPEECH 2012.*
[10] Zen *et al.*, Speech Communication, Vol. 51, No. 11, pp. 1039–1064, 2009.
[11] Meeds *et al.*, in *Proc. NIPS 2006.*
[12] de Cheveigné *et al.*, JASA, Vol. 111, No. 4, pp. 1917–1930, 2002.
[13] Rasmussen *et al.*, in *Proc. NIPS 2002.*
[14] Yoshii *et al.*, in *Proc. ICASSP 2013.*
[15] Koriyama *et al.*, in *Proc. INTERSPEECH 2013.*
[16] Fernandez *et al.*, in *Proc. ICASSP 2013.*
[17] Pilkington *et al.*, in *Proc. INTERSPEECH 2011.*
[18] Goto, in *Proc. ISMIR 2006.*