

歌声 F0 系列からの楽譜逸脱成分の抽出 - 動特性モデルに基づく楽譜との時間的対応付け - *

大石康智 (NTT), 亀岡弘和 (NTT/東大), 持橋大地 (統数研), 永野秀尚, 柏野邦夫 (NTT)

1 はじめに

楽曲のメロディを歌った歌声の音高 (F0) 軌跡には, そのメロディを構成する楽譜の音符の並びだけでなく, 楽譜に記載されない, “楽譜から逸脱した” 動的特徴成分が含まれる [1]。これらは, 発声器官の物理的制約に起因する成分 (特に, オーバーシュートやプレバレーション, 微細変動成分) と意図的表現による成分 (特に, ビブラートやポルタメント) からなると考えられ, 知覚的には, 前者は人間らしさ・自然性に関係し, 後者は巧拙感に関係することがわかってきている [2, 3]。さらに, 後者は意図して意図通りにならなかった場合と, 意図通りに制御できなかったずれとに分かれ, 習熟度に関連すると考えられる。ただし, これらの逸脱成分を F0 系列から特徴抽出することや歌唱者ごとにその多様性を学習することまではまだ十分に検討されていない。

本研究の目的は, このような物理的制約もしくは意図的表現による楽譜逸脱成分を F0 系列から楽譜と分離して抽出し, 歌唱者ごとにどのようなパターンをもちうるのか, 各パターンが文脈 (楽譜の音符列) にどう依存するかを計算機に学習させることである (Fig. 1)。ここでは, うろ覚えの状態ではなく, 楽曲のメロディを既に知っているものとして, 歌唱者なりに表情付けして歌った歌声を対象とする。本研究は, 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合成や歌声変換 [4, 5, 6] などへの応用が期待できる。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。文献 [5] と違って, 事前に歌唱者の歌い方が学習されるため, どんなメロディにもその歌い方を転写できることを特長とする。

本稿では, F0 動特性モデル [7, 8] に基づいて, 楽譜逸脱成分を特徴抽出するための, F0 系列と楽譜の音符列との時間的対応付け手法を提案する。まず歌声の F0 を

$$y(t) = y_t(t) + y_s(t) = u(t) * h(t) + y_s(t) \quad (1)$$

と, 連続時間領域 t で表現する。 $u(t)$ は楽譜の音符の並びが歌唱によって時間周波数軸上に伸縮した階段状の信号である。これに, 減衰率 ζ と固有周波数 Ω をパラメータとする 2 次系伝達関数

$$\mathcal{H}(s) = \frac{\Omega^2}{s^2 + 2\zeta\Omega s + \Omega^2} \quad (2)$$

のインパルス応答 $h(t)$ を畳み込むことによって [1], オーバーシュートやポルタメントのような, 音符が

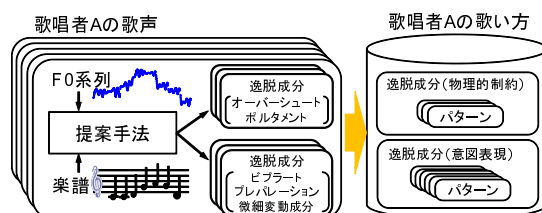


Fig. 1 F0 系列と楽譜の音符列との時間的対応付けに基づく楽譜逸脱成分の抽出と歌い方の学習

別の音符に移るときの楽譜逸脱成分を含む信号 $y_t(t)$ を得る。ここで, $u(t)$ を楽譜の音符の並びが既知の下で構成される隠れマルコフモデル (HMM) からの出力として表現する。一方, $y_s(t)$ はビブラートやプレバレーション, 微細変動成分のような音符の音高が安定するときの楽譜逸脱成分を表す。このとき, $y(t)$ である F0 と, そこに含まれる音符の並びが与えられた下で, これらを時間的に対応付けながら, 音符の遷移における逸脱成分の因子 $h(t)$ と音高が安定するときの逸脱成分 $y_s(t)$ を分離して特徴抽出することを考える。このモデルの表現方法とパラメータ推定アルゴリズムを提案し, 評価実験結果を示す。

2 F0 系列と楽譜との時間的対応付け

動特性モデル [8] では, 実測の F0 を手動で音符区間に区切り, その区間の動特性を 2 次系モデルで表現した。本手法は, 音符の並びそのものを HMM で表現し, F0 の動特性を考慮しながら音符区間 (状態系列) をも反復的に推定できることを特徴とする。

2.1 歌声の F0 の離散時間表現

実測の F0 は離散時間信号であるので, 式 (2) を後退差分変換 ($s \simeq (1 - z^{-1})/t_0$ で近似する変換であり, t_0 は離散時間表現におけるサンプリング周期) によって, 離散時間領域で表現する。この変換により, $\mathcal{H}(s)$ の逆システムの伝達関数を z 領域で

$$\mathcal{H}^{-1}(z) = a_2 z^{-2} + a_1 z^{-1} + a_0 \quad (3)$$

と書ける。ただし, $a_2 = \varphi^2$, $a_1 = -2\varphi(\psi + \varphi)$, $a_0 = 1 + 2\varphi\psi + \varphi^2$ であり, φ と ψ は $\varphi = 1/(\Omega t_0)$, $\psi = \zeta$ と表現される。ここで, k を離散時刻インデックスとすると, 階段状信号 $u[k]$ はパラメータ φ と ψ によって特性が決まる拘束つき全極モデルからの出力

$$u[k] = a_0 y_t[k] + a_1 y_t[k-1] + a_2 y_t[k-2] \quad (4)$$

とみなすことができる。一方, $y_s(t)$ も離散時間表現すると, 歌声の F0 系列はこれらの成分の和 $y[k] =$

* Extraction of Note Deviation Manner in F0 contour of Singing Voice. by OHISHI, Yasunori (NTT), KAMEOKA, Hirokazu (NTT / University of Tokyo), MOCHIHASHI, Daichi (The Institute of Statistical Mathematics), NAGANO Hidehisa, KASHINO Kunio (NTT)

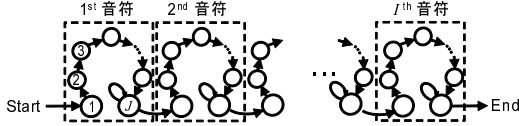


Fig. 2 楽譜の音符の並びが既知の下で構成される隠れマルコフモデル

$y_t[k] + y_s[k]$ と表現される。

2.2 歌声の F0 動特性モデルの統計モデル化

Left-to-Right 型 HMM を利用して、階段状に変化する信号 $u[k]$ を表現する (Fig. 2)。 $u[k]$ を HMM の出力値系列として、 $u[k] \sim \mathcal{N}(\mu[k] + d, \sigma_u^2)$ のように正規分布する確率変数と見なし、平均 $\mu[k] + d$ が HMM の状態遷移に伴って階段状に変化するモデルを考える。 Fig. 2 の点線枠で囲った部分は、メロディの楽譜の i 番目の音符を形成する状態集合 $\{S_{i,1}, S_{i,2}, \dots, S_{i,J-1}, S_{i,J}\}$ であり、これらの状態の出力分布はすべて同じものとする。状態 $S_{i,j}$ ($j = 1, \dots, J-1$) は、 $S_{i,j+1}$ だけに遷移し (遷移確率 = 1)、状態 $S_{i,J}$ は自己遷移、もしくは次の音符の初期状態 $S_{i+1,1}$ に遷移する。このような HMM を用いる理由は、短い音符区間が形成されることを抑制するためである。 I は楽譜に含まれる音符の総数とする。 HMM の構成は以下のとおりである。

出力値系列: $\{u[k]\}_{k=1}^K$

状態集合: $\mathcal{S} := \{S_{1,1}, S_{1,2}, \dots, S_{1,J}, \dots, S_{I,J}\}$

状態系列: $\{s_k\}_{k=1}^K$

状態出力分布: $P(u[k]|s_k = S_{i,j}) = \mathcal{N}(c_{S_{i,j}}[k] + d, \sigma_u^2)$

$$c_{S_{i,j}}[k] = A^{(i)}$$

状態遷移確率: $\phi_{S_{i,j}, S_{i',j'}}$

$$:= \log P(s_k = S_{i',j'} | s_{k-1} = S_{i,j})$$

ここで、状態出力分布の平均値の $A^{(i)}$ は楽譜から与えられる i 番目の音符の、対数周波数軸上の平均律の音高値とする。一方、 d はその音高からのずれを表現し、キーシフトパラメータと呼ぶ。例えば、歌唱者が楽譜よりも少し高めに歌ったり、低めに歌ったりすることに対応でき、このパラメータも楽譜逸脱成分を表す。これは全体にわたる定数パラメータ d としても、音符ごと (Fig. 2 の点線枠) に変化するパラメータ $\{d_1, d_2, \dots, d_I\}$ としてもよい。以降は簡単のため、パラメータ d としてモデルを組み立てる。 $\phi_{S_{i,j}, S_{i',j'}}$ を定数とすると、推定すべきパラメータは $\theta_u := \{\{s_k\}_{k=1}^K, d, \sigma_u^2\}$ となる。また、 $\{\mu[k]\}_{k=1}^K$ は、 $\{s_k\}_{k=1}^K$ が与えられたもとで $\mu[k] = c_{s_k}[k]$ となる。

$u[k]$ の統計モデルに基づき、F0 系列の確率密度関数を導く。まず、 $\mathbf{y} := (y[1], \dots, y[K])^T$, $\mathbf{u} := (u[1], \dots, u[K])^T$, $\boldsymbol{\mu} := (\mu[1], \dots, \mu[K])^T$, $\mathbf{d} := d\mathbf{1}_K$ とおくと、

$$\mathbf{u} | \theta_u \sim \mathcal{N}(\boldsymbol{\mu} + \mathbf{d}, \boldsymbol{\Sigma}_u), \quad \boldsymbol{\Sigma}_u = \sigma_u^2 \mathbf{I}_K \quad (5)$$

と書ける。ここで、 $\mathbf{1}_K$ は K 個の 1 からなるベクトル、 \mathbf{I}_K は $K \times K$ の単位行列である。前節で述べたように、音符遷移における楽譜逸脱成分を含む信号 $\mathbf{y}_t := (y_t[1], \dots, y_t[K])^T$ と \mathbf{u} の関係は、

$$\boldsymbol{\Psi} = \begin{bmatrix} a_0 & & & & \mathbf{O} \\ a_1 & a_0 & & & \\ a_2 & a_1 & a_0 & & \\ & \ddots & \ddots & \ddots & \\ \mathbf{O} & & a_2 & a_1 & a_0 \end{bmatrix} \quad (6)$$

と置くと、 $\mathbf{u} = \boldsymbol{\Psi} \mathbf{y}_t$ のように表現でき、式 (5) より、

$$\mathbf{y}_t | \theta_u, \varphi, \psi \sim \mathcal{N}(\boldsymbol{\Psi}^{-1}(\boldsymbol{\mu} + \mathbf{d}), \sigma_u^2 \boldsymbol{\Psi}^{-1}(\boldsymbol{\Psi}^{-1})^T) \quad (7)$$

が導かれる。ここで、音符ごとに 2 次系の制御パラメータ φ, ψ が異なるとしてもよく、その場合は、 $\boldsymbol{\Psi} = (\boldsymbol{\Psi}_1^T, \boldsymbol{\Psi}_2^T, \dots, \boldsymbol{\Psi}_I^T)^T$ とし、各 $\boldsymbol{\Psi}_i$ は φ_i, ψ_i で表現される。以降は簡単のため、 φ, ψ は全体で共通とする。

楽譜逸脱成分 $\mathbf{y}_s := (y_s[1], \dots, y_s[K])^T$ は Gauss 性白色雑音を仮定し、 $\mathbf{y}_s \sim \mathcal{N}(\mathbf{0}, \sigma_s^2 \mathbf{I}_K)$ と表現する。これに含まれるビブラート、ブレパレーション、微細変動成分の特徴抽出については今後の課題とする。 $\mathbf{y}_t, \mathbf{y}_s$ は互いに独立と仮定し、 $\Theta := \{\theta_u, \varphi, \psi, \sigma_s^2\}$ の下での F0 系列 $\mathbf{y} = \mathbf{y}_t + \mathbf{y}_s$ の確率密度関数は、

$$\mathbf{y} | \Theta \sim \mathcal{N}(\boldsymbol{\Psi}^{-1}(\boldsymbol{\mu} + \mathbf{d}), \sigma_u^2 \boldsymbol{\Psi}^{-1}(\boldsymbol{\Psi}^{-1})^T + \sigma_s^2 \mathbf{I}_K)$$

で与えられる。これが、パラメータ Θ に関する F0 系列 \mathbf{y} の尤度関数となる。 Θ の事前確率は、各要素は独立で、状態系列 $\{s[k]\}_{k=1}^K, \varphi, \psi, d$ 以外のパラメータは一様に分布すると仮定し、 $P(\Theta) \propto P(s_1) \prod_{k=2}^K P(s_k | s_{k-1}) P(\varphi) P(\psi) P(d)$ とする。

先の尤度関数は全区間で F0 が観測されることが暗に想定されているが、実際には F0 は有声音が発せられるときのみ観測可能であり、無声音の区間では F0 は通常観測できない。従って、実測 F0 系列からのパラメータ推定を行うためには一般に欠損データの問題を扱う必要がある。統計的枠組では、この問題は不完全データ問題に他ならず、EM アルゴリズムにより扱うことができる。詳細は文献 [9] を参照されたい。

2.3 パラメータ推定アルゴリズム

実測 F0 系列 \mathbf{y} が与えられたとき、事後確率 $P(\Theta | \mathbf{y}) \propto P(\mathbf{y} | \Theta) P(\Theta)$ を最大化するパラメータ Θ を推定する問題を考える。 $P(\Theta | \mathbf{y})$ を最大化する問題は解析的に解くことはできないが、 $\mathbf{x} := [\mathbf{y}_t^T, \mathbf{y}_s^T]^T$ を完全データと見なすことで EM アルゴリズムによる不完全データ問題に帰着できる。不完全データと完全データの関係は、 $\mathbf{y} = \mathbf{H} \mathbf{x}$ となり、ここで、 $\mathbf{H} := [\mathbf{I}_K \ \mathbf{I}_K]$ とする。このとき、 Q 関数は、

$$Q(\Theta, \Theta') \stackrel{c}{=} \frac{1}{2} \left[\log |\boldsymbol{\Lambda}^{-1}| - \text{tr}(\boldsymbol{\Lambda}^{-1} \mathbb{E}[\mathbf{x} \mathbf{x}^T | \mathbf{y}; \Theta']) \right] \\ + 2 \mathbf{m}^T \boldsymbol{\Lambda}^{-1} \mathbb{E}[\mathbf{x} | \mathbf{y}; \Theta'] - \mathbf{m}^T \boldsymbol{\Lambda}^{-1} \mathbf{m} + \log P(\Theta) \\ \mathbf{m} := \begin{bmatrix} \boldsymbol{\Psi}^{-1}(\boldsymbol{\mu} + \mathbf{d}) \\ \mathbf{0} \end{bmatrix}, \quad \boldsymbol{\Lambda}^{-1} := \begin{bmatrix} \boldsymbol{\Psi}^T \boldsymbol{\Psi} / \sigma_u^2 & \mathbf{O} \\ \mathbf{O} & \mathbf{I}_K / \sigma_s^2 \end{bmatrix}$$

となる。ここで、 $\text{tr}(\cdot)$ は行列のトレースを表し、 $\mathbb{E}[x|y; \Theta']$ と $\mathbb{E}[xx^T|y; \Theta']$ は、条件付きガウス分布の性質より、

$$\begin{aligned}\mathbb{E}[x|y; \Theta'] &= m + \Lambda H^T (H \Lambda H^T)^{-1} (y - Hm) \\ \mathbb{E}[xx^T|y; \Theta'] &= \Lambda - \Lambda H^T (H \Lambda H^T)^{-1} H \Lambda \\ &\quad + \mathbb{E}[x|y; \Theta'] \mathbb{E}[x|y; \Theta']^T \quad (8)\end{aligned}$$

と書ける。E ステップでは、直前のステップで更新されたパラメータを Θ' に代入し、 $\mathbb{E}[x|y; \Theta']$ と $\mathbb{E}[xx^T|y; \Theta']$ を算出する。M ステップでは、各パラメータの更新を行う。便宜上、 y_t 、 y_s に対応するように $\mathbb{E}[x|y; \Theta']$ および $\mathbb{E}[xx^T|y; \Theta']$ を

$$\mathbb{E}[x|y; \Theta'] = \begin{bmatrix} \bar{x}_t \\ \bar{x}_s \end{bmatrix}, \quad \mathbb{E}[xx^T|y; \Theta'] = \begin{bmatrix} R_t & * \\ * & R_s \end{bmatrix}$$

のように区分表現する。各パラメータの更新式を以下に示す。

1) 状態系列: Q 関数で $s := \{s_k\}_{k=1}^K$ に関する項は

$$\begin{aligned}\mathcal{I}_1(s) &:= -\frac{1}{2\sigma_u^2} \sum_{k=1}^K (o[k] - c_{s_k}[k])^2 + \log P(s_1) \\ &\quad + \sum_{k=2}^K \log P(s_k | s_{k-1}) \quad (9)\end{aligned}$$

である。ただし、 $o[k] := [\Psi \bar{x}_t]_k$ であり、 $[\cdot]_k$ はベクトルの k 番目の要素を表す。これを最大化する $\{s_k\}_{k=1}^K$ は Viterbi 探索により効率的に解くことができる。

2) 2 次系制御パラメータ: φ の事前分布を $\varphi \sim \mathcal{N}(\mu_\varphi, \sigma_\varphi^2)$ 、 ψ の事前分布を $\psi \sim \mathcal{N}(\mu_\psi, \sigma_\psi^2)$ とする。 Q 関数より、 φ, ψ に関する項は

$$\begin{aligned}\mathcal{I}_2(\varphi, \psi) &:= -\frac{1}{2\sigma_u^2} \text{tr}(\Psi^T \Psi R_t) + \frac{1}{\sigma_u^2} (\mu + d)^T \Psi \bar{x}_t \\ &\quad + \log |\Psi| - \frac{1}{2\sigma_\varphi^2} (\varphi - \mu_\varphi)^2 - \frac{1}{2\sigma_\psi^2} (\psi - \mu_\psi)^2 \quad (10)\end{aligned}$$

となる。ここで、式 (3) より、 Ψ は定数行列 U_0, U_1, U_2 を用いて $\Psi = U_2 \varphi^2 + U_1 \varphi \psi + U_0$ と表せるので、式 (10) の φ と ψ に関する偏導関数を求めて、ニュートン・ラフソン法を利用すれば、数値的に φ と ψ が求められる。

$\Psi = (\Psi_1^T, \Psi_2^T, \dots, \Psi_I^T)^T$ と表現した場合も同様に、各 i に関して事前分布 $\varphi_i \sim \mathcal{N}(\mu_{\varphi_i}, \sigma_{\varphi_i}^2)$ 、 $\psi_i \sim \mathcal{N}(\mu_{\psi_i}, \sigma_{\psi_i}^2)$ を導入する。そして、 Q 関数の φ_i と ψ_i に関する偏導関数から、ニュートン・ラフソン法により数値的に φ_i と ψ_i が求められる。

3) キーシフトパラメータ: d の事前分布を $d \sim \mathcal{N}(0, \sigma_d^2)$ として、 Q 関数を d に関して最大化すると

$$d = (1^T \Psi \bar{x}_t - 1^T \mu) / (K + \sigma_u^2 / \sigma_d^2) \quad (11)$$

の更新式を得る。

音符ごとに変化するキーシフトパラメータを導入した場合、 $d = M(d_1, d_2, \dots, d_I)^T$ と表現できる。こ

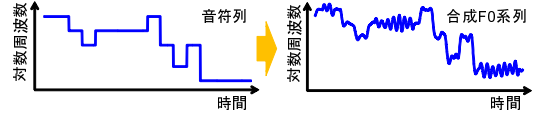


Fig. 3 文献 [1] に基づく合成 F0 系列の作成

Table 1 合成 F0 系列を利用した評価実験結果

	音符区間の正解率 [%]	二乗平均平方根 (RMS)		
		ζ	Ω	d
提案法	99.0	0.0398	0.00209	2.28
比較法	94.8	0.0972	0.0161	20.66

ここで、 M は各行に 1 が一個あり残りはすべて 0 であるような $K \times I$ のバイナリ行列であり、階段状に変化する μ の各要素に基づいて構成される。各 i に関して事前分布 $d_i \sim \mathcal{N}(0, \sigma_{d_i}^2)$ を導入し、 Q 関数を d_i に関して最大化すると更新式が求められる。

4) 分散パラメータ: σ_u^2 、 σ_s^2 の更新式は、

$$\sigma_u^2 = (\text{tr}(\Psi^T \Psi R_t) - 2r^T \Psi \bar{x}_t + r^T r) / K \quad (12)$$

$$\sigma_s^2 = \text{tr}(R_s) / K \quad (13)$$

となる。ここで、 $r := \mu + d$ とした。

3 評価実験

提案法による F0 系列と楽譜の音符列との時間的対応付け、および楽譜逸脱成分を表現するモデルパラメータの推定性能を評価する。実測の F0 系列を用いた場合、これらの正解となる値はわからないため、まずは人工的に作成した合成 F0 系列で評価する。事前に童謡や歌謡曲から抜粋したメロディの楽譜（合計 16 種類、抜粋部分の長さは平均 11.4 秒）に基づいて、対数周波数（単位は cent、半音は 100 cent に相当する）軸上で階段状に変化する音符列を作成する（サンプリング周期 5 ms、休符区間はその前の音符の音高を休符区間も伸ばす）。これに、文献 [1] で調査されたプロ歌唱者とアマチュア歌唱者の F0 系列を合成するためのパラメータを用いて、楽譜逸脱成分を付加する (Fig. 3)。具体的には、オーバーシュート（プロ： $\zeta = 0.570$, $\Omega = 0.0363$ [rad/ms]、アマ： $\zeta = 0.518$, $\Omega = 0.0312$ [rad/ms]）、プレパレーション（プロ： $\zeta = 0.675$, $\Omega = 0.0363$ [rad/ms]、アマ： $\zeta = 0.660$, $\Omega = 0.0287$ [rad/ms]）、ビブラート（プロ：Vibrato extent = 83.0 cent, Vibrato rate = 6.25 Hz、アマ：Vibrato extent = 33.5 cent, Vibrato rate = 4.70 Hz）、微細変動成分（白色雑音をカットオフ周波数 10 Hz のローパスフィルタに通し、振幅の平均値を 20 cent に調整）、キーシフト（+25 cent, 0 cent, -25 cent の 3 パターン）とした。

作成された全 96 パターンの合成 F0 系列にパラメータ推定アルゴリズムを適用する。固定パラメータは、文献 [1] を参考に、 $J = 20$ 、 $\mu_\varphi = 5.74$ 、 $\sigma_\varphi^2 = 0.1$ 、 $\mu_\psi = 0.57$ 、 $\sigma_\psi^2 = 0.02$ 、 $\sigma_d^2 = 400$ 、 $\phi_{S_i, J, S_i, J} =$

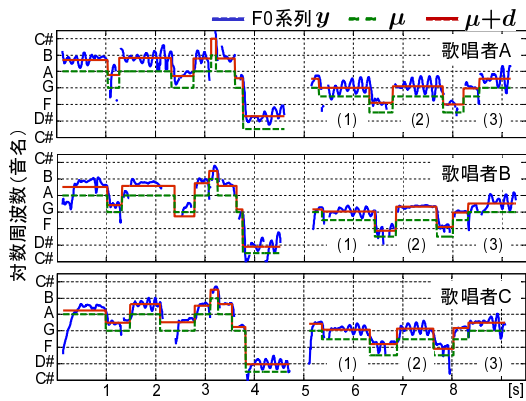


Fig. 4 実測のF0系列と楽譜の音符列との時間的対応付け結果

0.99999, $\phi_{S_{i+1,1}, S_{i,J}} = 0.00001$ とした。評価尺度は、(1) 正解の音符区間と、推定された状態系列に基づく音符区間の一致する割合（正解率）、(2) オーバーシュート作成の ζ, Ω と、推定されたパラメータ ψ, φ から計算される ζ, Ω との二乗平均平方根（RMS）、(3) キーシフトの値と、推定された d との RMS, とする。

Tab. 1 に評価結果を示す。 σ_s^2 が 10^{-5} のオーダーで変化しなくなるまで反復推定を繰り返した。比較法とは、2.2 節の HMM を用いた Viterbi 探索により、F0 系列を前処理で音符区間に分割し、文献 [8] の動特性モデルによってその区間ごとにパラメータを推定して評価した結果である。提案法は F0 の音符遷移における動特性を考慮しながら Viterbi 探索を反復して行うため、前処理のみの比較法に比べて、音符区間（状態系列）の推定性能が高くなる。それによって、 ζ, Ω, d の RMS が小さくなり、比較法に比べて、楽譜逸脱成分を抽出できていると言える。提案法の ζ, Ω の RMS が許容できる誤差であるか確認するため、音符列に、正解の ζ, Ω を代入した 2 次系インパルス応答を畳み込んだ信号と、推定された ζ, Ω を代入した 2 次系インパルス応答を畳み込んだ信号との RMS を計算したところ、平均 6.57 cent であった。これは半音 100 cent の 6.57% の誤差であるので、提案法は楽譜逸脱成分を抽出するための妥当な方法であると考えている。

次に実測の F0 系列から楽譜逸脱成分の特徴抽出を試みる。音大出身の歌唱者 3 名（男性 1 名、女性 2 名）がレミオロメンの「粉雪」の出だしの部分をヘッドフォンで伴奏を聞きながら歌詞付きで歌った歌声を収録した。F0 は YIN[10] を利用して、5 ms ごとに推定した ($t_0 = 5$ ms)。ここでは、音符ごとに 2 次系パラメータおよびキーシフトパラメータが変化するものとして、アルゴリズムを実行した。従って、事前分布のパラメータを $\mu_{\varphi_i} = 5.74, \sigma_{\varphi_i}^2 = 0.1, \mu_{\psi_i} = 0.57, \sigma_{\psi_i}^2 = 0.02, \sigma_{d_i}^2 = 400$ ($i = 1, \dots, I$) とした。

Fig. 4 の点線は楽譜の音符列を時間軸上で伸縮させた結果であり、さらに各音符にキーシフトパラメータ d_i を周波数軸上に足し合わせた結果が赤実線（式 (5) の平均値）となる。定性的にみると階段状信号がうまく推定され、どの歌唱者も楽譜の音符の音高よ

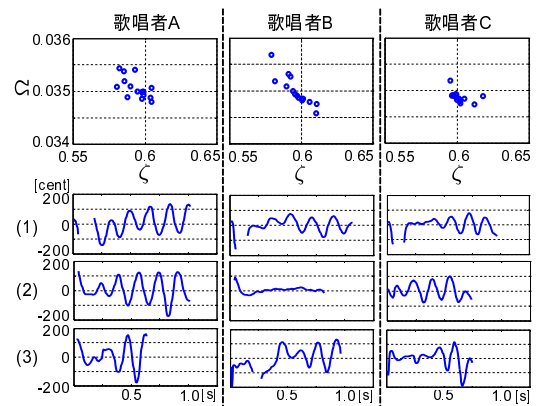


Fig. 5 各歌唱者の F0 系列における楽譜逸脱成分の抽出結果：図中の (1) ~ (3) は Fig. 4 の音符区間 (1) ~ (3) の逸脱成分に対応する。

りも半音程度高めに歌う楽譜逸脱成分を観測できる。

Fig. 5 の上部は、各音符の遷移における楽譜逸脱成分を構成する 2 次系パラメータ ζ, Ω の分布を示す。これは歌唱者ごとに大きな違いはなく、先行研究 [1] の結果と整合する。一方、Fig. 5 の下部には Fig. 4 の (1) ~ (3) の音符区間における音高が安定するときの楽譜逸脱成分を示す。ここでは、各音符の \bar{x}_s を図示した。同じ音符なのに、歌唱者ごとにその挙動が異なり、例えば、歌唱者 A はオンセットから周期的に振動するが、歌唱者 C はオンセットからしばらくして振動する傾向にある。この楽譜逸脱成分を周期的なビブラート、プレパレーション、微細変動成分に分離すること、そしてそれらが歌唱者ごとにどのようなパターンをもつのか、音符の並びにどう影響を受けるか学習することを今後の課題とする。

4 おわりに

歌唱者の歌い方や個性、癖の学習を目指して、F0 系列に含まれる楽譜逸脱成分を抽出するための、F0 系列と楽譜の音符列との時間的対応付け手法を提案し、その評価実験を行った。今後は、抽出された楽譜逸脱成分のパターンを学習すること、音高とともに音色、音量に関する楽譜逸脱成分の特徴抽出に取り組むことも検討中である。

参考文献

- [1] Saitou, T. et al., *Proc. EUROSPEECH 2009*, pp. 832–835, 2009.
- [2] 齋藤ほか, 音講論集, 2-Q-26, 2007.
- [3] 右田ほか, 音講論集, 3-P-28, 2010.
- [4] Bonada, J. et al., *Proc. SMAC 2003*, 2003.
- [5] Nakano, T. et al., *Proc. SMC 2009*, pp. 343–348, 2009.
- [6] 大浦ほか, 情処研報音楽情報科学, 2010–MUS–86–1, 2010.
- [7] 柏野ほか, 音講論集, 2-9-1, pp. 625–626 1998.
- [8] Ohishi, Y. et al., *Proc. INTERSPEECH 2010*, pp. 2598–2601, 2010.
- [9] 亀岡ほか, 音講論集, 1-1-3, 2010.
- [10] de Cheveigné, A. et al., *JASA*, Vol. 111, No. 4, pp. 1917–1930, 2002.