ノート指令と表現指令によって駆動される歌声 F0 生成過程の統計モデル*

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

1　はじめに

歌声の音高 (F_0) 軌跡には，発声器官の物理的制約に起因する動的変動成分（オーバーシュートや微細変動成分）や歌唱者の意図的表現による動的変動成分（ビブラートやポルタメント）が含まれ，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 [1]。ただし，これらの動的変動成分を F_0 軌跡から特徴抽出し，歌唱者ごとにその特性を精緻に学習することはまだ十分に検討されていない。

本研究の目的は，このような動的変動成分を F_0 軌跡から楽譜情報と分離して特徴抽出し，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。ここでは，うろ覚えの状態で歌った歌声ではなく，楽曲のメロディまたはその楽譜を既知として，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，癖を学習することを目指しており，歌唱力評価や歌唱者識別，現在盛んに研究される歌声合成や歌声変換 [2, 3, 4] への応用が期待できる。

本研究では，物理モデルに基づいて F_0 軌跡の生成過程を記述し，そこから動的変動成分の特徴抽出に取り組む。これまで，話声の F_0 パターンを表現する藤崎モデル [5] を参考に，2 次系を利用した歌声 F_0 制御モデルが提案された [1, 6]。2 次系の伝達関数は

$$\mathcal{G}(s) = \frac{\Omega^2}{s^2 + 2\zeta\Omega s + \Omega^2} \qquad (1)$$

であり，この減衰率 ζ によって，様々な振動現象を表現できる。文献 [1] では，楽譜の音符列を表す階段状信号に式 (1) のインパルス応答を部分的に畳み込んで得られる F_0 軌跡を利用して，表情豊かな歌声合成音を実現した。制御パラメータ（減衰率 ζ と固有周波数 Ω）は手作業あるいは規則に基づいて決定された。一方，我々は観測される F_0 軌跡から制御パラメータを推定する逆問題の解法を検討してきた [7]。ただし，ビブラートや微細変動成分がすべて白色雑音としてモデル化されたため，意図的表現による動的変動成分を微細変動成分と分離して特徴付けられなかった。

本稿では，歌唱の意図的表現を特徴抽出するために，ノート指令信号と表現指令信号によって駆動される歌声 F_0 軌跡の生成過程を提案する（Fig. 1）。ここで，ノート指令信号は楽譜に記載される音符の並びを表現する。一方，表現指令信号は歌唱者の音楽的な表現意図を矩形状の細かい指令として表現する。ノート成分と表現成分は各指令信号によって駆動される制御機講の出力であり，これらの制御機講は 2 次系で表現される。ノート制御機講はオーバーシュートなどのノートの立ち上がり方を制御する。表現制御機講は矩形状の細かい指令信号を制御してビブラートやポルタメントを生成する。微細変動成分は 10 Hz 以上の不規則な振動成分を想定する [1]。最終的に，対数スケールの F_0 軌跡 $y(t)$（ここで，t は時間）は，これら 3 つの成分の重ね合わせと想定する。この F_0 生成過程を想定した理由は 2 点ある。

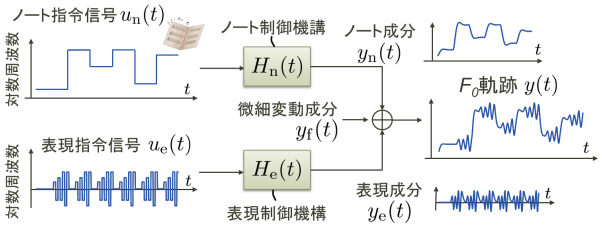

Fig. 1　提案する歌声の F_0 生成過程の概略

理由 1　藤崎モデルでは，甲状軟骨の二つの独立な運動 (平行移動と回転) に伴う声帯の長さの変化の合計が F_0 の時間的変化をもたらすと解釈され，これらの運動に関係する成分をそれぞれ，フレーズ成分とアクセント成分とした [5]。歌声にも同様の機講が存在すると仮定し，大域的な変化を表すノート成分とアクセントのような歌唱者が意図的に制御できる表現成分から歌声 F_0 軌跡が構成されると考えた。

理由 2　従来，ビブラートは正弦波で表現され，そのパラメータは Vibrato rate と Vibrato extent とされた [1]。これらのパラメータが指数的に変化するモデルも提案された [3]。しかし実際は，これらのパラメータは時々刻々と変化するものの，指数的な変化であるとは言えない。この複雑な振動現象をモデル化して歌い方を認識するために，意図を表す矩形状の細かい指令信号が 2 次系によって制御されるビブラートの生成過程を想定した。

以降の節では，上記の歌声 F_0 軌跡の生成過程を離散時間表現して確率モデル化し，統計的手法に基づいてモデルパラメータの推定アルゴリズムを導出する。そして，提案モデルの有効性を実測の F_0 軌跡を用いて評価する。文献 [8] では，同じ枠組みで藤崎モデルを確率過程に基づいて統計モデル化する試みも行っているのでこちらも参照されたい。

---

* A Statistical Model of Singing Voice F_0 Contours Driven by Musical Note and Expression Command Functions by OHISHI, Yasunori (NTT), KAMEOKA, Hirokazu (NTT / The University of Tokyo), MOCHIHASHI, Daichi (The Institute of Statistical Mathematics), KASHINO Kunio (NTT)

2　歌声 F_0 生成過程の確率モデル

歌声 F_0 軌跡の生成過程を離散時間表現し，確率過程に基づいて統計モデル化する。連続時間領域で表現される 2 次系制御機講の伝達関数の離散時間表現を得るために，後退差分変換を利用する [7]。t_0 をサンプリング周期とすると，この変換によりノート制御機構の逆システムの伝達関数 $\mathcal{H}_{\rm n}^{-1}(s)$ は z 領域で，

$$\mathcal{H}_{\rm n}^{-1}(z) = a_2 z^{-2} + a_1 z^{-1} + a_0 \qquad (2)$$

と書くことができる。ただし，$a_2 = \varphi^2, a_1 = -2\varphi(\psi+\varphi), a_0 = 1+2\varphi\psi+\varphi^2$，および，$\varphi = 1/(\Omega t_0)$, $\psi = \zeta$ と表現される。ここで，k を離散時刻インデックスとし，ノート指令信号およびノート成分の離散時間表現をそれぞれ $u_{\rm n}[k]$，$y_{\rm n}[k]$ とすると，$y_{\rm n}[k]$ は，ノート制御パラメータ φ, 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

$$u_{\rm n}[k] = a_0 y_{\rm n}[k] + a_1 y_{\rm n}[k-1] + a_2 y_{\rm n}[k-2] \qquad (3)$$

と見なすことができる。同様に，表現指令信号 $u_{\rm e}[k]$ と表現成分 $y_{\rm e}[k]$ の関係も $u_{\rm e}[k] = b_0 y_{\rm e}[k] + b_1 y_{\rm e}[k-1] + b_2 y_{\rm e}[k-2]$ と書ける。ただし，表現制御機講は臨界制動系 ($\zeta = 1$ の場合) を想定し，$b_2 = \xi^2$, $b_1 = -2\xi(1+\xi)$, $b_0 = 1+2\xi+\xi^2$，表現制御パラメータを $\xi = 1/(\Omega t_0)$ とする。微細変動成分 $y_{\rm f}(t)$ の離散時間表現を $y_{\rm f}[k]$ として，歌声 F_0 軌跡の離散時間表現は，$y[k] = y_{\rm n}[k] + y_{\rm e}[k] + y_{\rm f}[k]$ と想定する。

次に，F_0 生成過程を確率モデル化する。ノート指令信号と表現指令信号はそれぞれ，楽譜に記載されるメロディの音符の並びと歌唱者の音楽的な表現意図を表す (Fig. 1)。これらの指令信号を表現するために，HMM を利用して，$u_{\rm n}[k]$ と $u_{\rm e}[k]$ を確率モデル化する（Fig. 2)。$\boldsymbol{o}[k] := (u_{\rm n}[k], u_{\rm e}[k])^{\rm T}$ を

$$\boldsymbol{o}[k] \sim \mathcal{N}(\boldsymbol{\nu}[k], \boldsymbol{\Upsilon}), \boldsymbol{\nu}[k] := \begin{bmatrix}\mu_{\rm n}[k] \\ \mu_{\rm e}[k]\end{bmatrix}, \boldsymbol{\Upsilon} := \begin{bmatrix}\sigma_{\rm n}^2 & 0 \\ 0 & \sigma_{\rm e}^2\end{bmatrix}$$

のように正規分布する確率変数と見なし，平均 $\boldsymbol{\nu}[k]$ が状態遷移に伴って変化するモデルを考える。

具体的には，この HMM は $I \times J$ 個の状態集合 $\mathcal{S} := \{{\rm S}_{i,j}\}_{i=1,j=1}^{I,J}$ からなる。状態 ${\rm S}_{i,j}$ では，$\mu_{\rm n}[k]$ は $A_{\rm n}^{(i)} + d_i$ の値をとる。ここで，$A_{\rm n}^{(i)}$ は楽譜に記載されるメロディの i 番目の音符の音高を表し（この値は楽譜から与えられるものとする)，d_i はその絶対音高からのずれ（音高シフトパラメータと呼ぶ)，I は音符の総数を表す。Fig. 2 に示す状態遷移により，$\mu_{\rm n}[k]$ は I 個の音符区間からなる階段状信号を表現する。一方，$\mu_{\rm e}[k]$ は $B_{\rm e}^{(i,j)}$ の値をとり，これは歌唱者の表現意図を表すための，矩形の指令の大きさを表す。ここで，i 番目の音符では，${\rm S}_{i,1}$ を通らずして，状態 ${\rm S}_{i,j}$ から別の状態 ${\rm S}_{i,j'}$ $(j \neq j',\ 2 \leq j \leq J,\ 2 \leq j' \leq J)$ へ直接に遷移できない制約を設けることによって，$\mu_{\rm e}[k]$ は Fig. 2 に示すよ

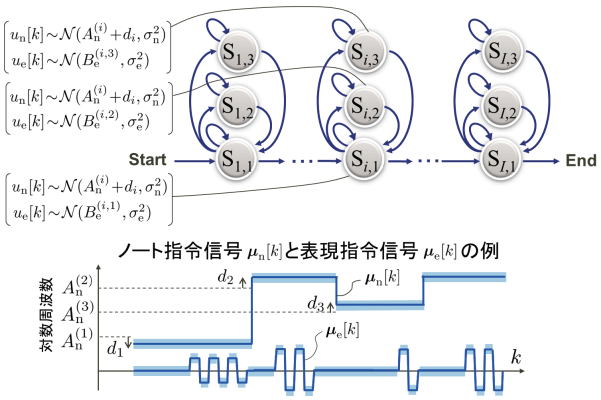

Fig. 2　HMM による指令信号の統計モデル化

うな矩形信号を表現する。J は表現指令を構成するための状態数を表す（Fig. 2 は $J = 3$ である)。指令信号を生成する HMM（指令信号モデル）の構成をまとめると，出力系列: $\{\boldsymbol{o}[k]\}_{k=1}^{K}$，状態集合: $\{{\rm S}_{i,j}\}_{i=1,j=1}^{I,J}$，状態系列: $\{s_k\}_{k=1}^{K}$，出力確率分布: $P(\boldsymbol{o}[k]|s_k = n) = \mathcal{N}(\boldsymbol{c}_n[k], \boldsymbol{\Upsilon})$, $\boldsymbol{c}_n[k] = (A_{\rm n}^{(i)} + d_i,\ B_{\rm e}^{(i,j)})^{\rm T}$, $(n = {\rm S}_{i,j})$，状態遷移確率 $\phi_{n',n} := \log P(s_k = n | s_{k-1} = n')$ となる。簡単のため，状態遷移確率 $\phi_{n',n}$ は定数とすると，指令信号モデルにおいて推定すべきパラメータは，HMM の経路を表す状態系列 $\{s_k\}_{k=1}^{K}$，音高シフトパラメータ $\{d_i\}_{i=1}^{I}$，表現指令の大きさパラメータ $\{B_{\rm e}^{(i,j)}\}_{i=1,j=1}^{I,J}$，指令信号の分散パラメータ $\sigma_{\rm n}^2$，$\sigma_{\rm e}^2$ であり，これらをまとめて $\theta_{\rm u}$ と記述する。また，平均値系列 $\{\mu_{\rm n}[k]\}_{k=1}^{K}$ および $\{\mu_{\rm e}[k]\}_{k=1}^{K}$ は，状態系列 $\{s_k\}_{k=1}^{K}$ が与えられたもとで，$(\mu_{\rm n}[k],\ \mu_{\rm e}[k])^{\rm T} \leftarrow \boldsymbol{c}_{s_k}[k]$ で与えられる。

$\boldsymbol{y} = (y[1], \ldots, y[K])^{\rm T}$ の確率密度関数を導く。$\boldsymbol{u}_{\rm n} := (u_{\rm n}[1], \ldots, u_{\rm n}[K])^{\rm T}$，$\boldsymbol{u}_{\rm e} := (u_{\rm e}[1], \ldots, u_{\rm e}[K])^{\rm T}$，$\boldsymbol{\mu}_{\rm n} := (\mu_{\rm n}[1], \ldots, \mu_{\rm n}[K])^{\rm T}$，$\boldsymbol{\mu}_{\rm e} := (\mu_{\rm e}[1], \ldots, \mu_{\rm e}[K])^{\rm T}$ とすると

$$\boldsymbol{u}_{\rm n}|\theta_u \sim \mathcal{N}(\boldsymbol{\mu}_{\rm n}, \boldsymbol{\Sigma}_{\rm n}), \quad \boldsymbol{\Sigma}_{\rm n} = \sigma_{\rm n}^2 \boldsymbol{I}_K \qquad (4)$$

$$\boldsymbol{u}_{\rm e}|\theta_u \sim \mathcal{N}(\boldsymbol{\mu}_{\rm e}, \boldsymbol{\Sigma}_{\rm e}), \quad \boldsymbol{\Sigma}_{\rm e} = \sigma_{\rm e}^2 \boldsymbol{I}_K \qquad (5)$$

と書ける。ここで，$\boldsymbol{I}_K$ は $K \times K$ の単位行列を表す。ノート成分 $\boldsymbol{y}_{\rm n} := (y_{\rm n}[1], \ldots, y_{\rm n}[K])^{\rm T}$ とノート指令信号 $\boldsymbol{u}_{\rm n}$ の関係，および表現成分 $\boldsymbol{y}_{\rm e} := (y_{\rm e}[1], \ldots, y_{\rm e}[K])^{\rm T}$ と表現指令信号 $\boldsymbol{u}_{\rm e}$ の関係は，

$$\boldsymbol{A} := \begin{bmatrix}\boldsymbol{A}^{(1)} \\ \vdots \\ \boldsymbol{A}^{(i)} \\ \vdots \\ \boldsymbol{A}^{(I)}\end{bmatrix} = \begin{bmatrix} a_0^{(1)} & & & & O \\ a_1^{(1)} & a_0^{(1)} & & & \\ \ddots & \ddots & \ddots & & \\ & a_2^{(i)} & a_1^{(i)} & a_0^{(i)} & \\ & & \ddots & \ddots & \ddots \\ O & & & a_2^{(I)} & a_1^{(I)} \quad a_0^{(I)} \end{bmatrix}$$

$$\boldsymbol{B} := \begin{bmatrix} b_0 & & & O \\ b_1 & b_0 & & \\ b_2 & b_1 & b_0 & \\ \ddots & \ddots & \ddots & \ddots \\ O & & b_2 & b_1 \quad b_0 \end{bmatrix}$$

と置くと，それぞれ，$\boldsymbol{u}_{\rm n}=\boldsymbol{A}\boldsymbol{y}_{\rm n}, \boldsymbol{u}_{\rm e}=\boldsymbol{B}\boldsymbol{y}_{\rm e}$ と表現できる。ここでは，先行研究で得られた知見 [1] に基づいて，ノート制御パラメータがノートごとに異なるものと想定し，$a_0^{(i)}$, $a_1^{(i)}$, $a_2^{(i)}$ は，$a_2^{(i)}=(\varphi^{(i)})^2$, $a_1^{(i)}=-2\varphi^{(i)}(\psi^{(i)}+\varphi^{(i)})$, $a_0^{(i)}=1+2\varphi^{(i)}\psi^{(i)}+(\varphi^{(i)})^2$ と表現され，$\{\varphi^{(i)},\ \psi^{(i)}\}_{i=1}^{I}$ を制御パラメータとする。そこで，ノート成分と表現成分の確率密度関数は，

$$\boldsymbol{y}_{\rm n}\sim\mathcal{N}(\boldsymbol{A}^{-1}\boldsymbol{\mu}_{\rm n},\boldsymbol{A}^{-1}\boldsymbol{\Sigma}_{\rm n}(\boldsymbol{A}^{-1})^{\rm T}) \tag{6}$$

$$\boldsymbol{y}_{\rm e}\sim\mathcal{N}(\boldsymbol{B}^{-1}\boldsymbol{\mu}_{\rm e},\boldsymbol{B}^{-1}\boldsymbol{\Sigma}_{\rm e}(\boldsymbol{B}^{-1})^{\rm T}) \tag{7}$$

が導出される。微細変動成分 $\boldsymbol{y}_{\rm f}:=(y_{\rm f}[1],\ldots,y_{\rm f}[K])^{\rm T}$ はガウス性白色雑音を想定する。

$$\boldsymbol{y}_{\rm f}|\sigma_{\rm f}^2\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{\Sigma}_{\rm f}),\quad \boldsymbol{\Sigma}_{\rm f}=\sigma_{\rm f}^2\boldsymbol{I}_K \tag{8}$$

仮定より，$\boldsymbol{y}_{\rm n}$, $\boldsymbol{y}_{\rm e}$, $\boldsymbol{y}_{\rm f}$ は独立なので，パラメータ $\Theta:=\{\theta_{\rm u},\ \{\varphi^{(i)},\ \psi^{(i)}\}_{i=1}^{I},\ \xi,\ \sigma_{\rm f}^2\}$ が与えられた下での歌声 F_0 軌跡 $\boldsymbol{y}=\boldsymbol{y}_{\rm n}+\boldsymbol{y}_{\rm e}+\boldsymbol{y}_{\rm f}$ の確率密度関数（尤度関数）は式 (6)，(7) と式 (8) より，

$$\begin{aligned}\boldsymbol{y}|\Theta\sim\mathcal{N}\big(&\boldsymbol{A}^{-1}\boldsymbol{\mu}_{\rm n}+\boldsymbol{B}^{-1}\boldsymbol{\mu}_{\rm e},\\ &\boldsymbol{A}^{-1}\boldsymbol{\Sigma}_{\rm n}(\boldsymbol{A}^{-1})^{\rm T}+\boldsymbol{B}^{-1}\boldsymbol{\Sigma}_{\rm e}(\boldsymbol{B}^{-1})^{\rm T}+\boldsymbol{\Sigma}_{\rm f}\big)\end{aligned} \tag{9}$$

となる。Θ の事前確率は，各パラメータは互いに独立で，パラメータ $\sigma_{\rm n}^2$, $\sigma_{\rm e}^2$, $\sigma_{\rm f}^2$ は一様に分布すると仮定し，$P(\Theta)\propto P(\varphi^{(1)})\ldots P(\varphi^{(I)})P(\psi^{(1)})\ldots P(\psi^{(I)})P(\xi)P(d_1)\ldots P(d_I)P(B_{\rm e}^{(1,1)})\ldots P(B_{\rm e}^{(I,J)})P(s_1)\prod_{k=2}^{K}P(s_k|s_{k-1})$ と記述される。

3　パラメータ推定アルゴリズム

Θ の事後確率 $P(\Theta|\boldsymbol{y})$ を最大化する問題を解析的に解くことはできないが，$\boldsymbol{x}:=(\boldsymbol{y}_{\rm n}^{\rm T},\ \boldsymbol{y}_{\rm e}^{\rm T},\ \boldsymbol{y}_{\rm f}^{\rm T})^{\rm T}$ 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 $\boldsymbol{y}=\boldsymbol{H}\boldsymbol{x}$ となり，ここで $\boldsymbol{H}:=[\boldsymbol{I}_K\ \boldsymbol{I}_K\ \boldsymbol{I}_K]$ とする。このとき，Q 関数は，

$$\begin{aligned}Q(\Theta,\Theta')\overset{\rm c}{=}\frac{1}{2}\big[&\log|\boldsymbol{\Lambda}^{-1}|-{\rm tr}(\boldsymbol{\Lambda}^{-1}\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\rm T}|\boldsymbol{y};\Theta'])\\ &+2\boldsymbol{m}^{\rm T}\boldsymbol{\Lambda}^{-1}\mathbb{E}[\boldsymbol{x}|\boldsymbol{y};\Theta']-\boldsymbol{m}^{\rm T}\boldsymbol{\Lambda}^{-1}\boldsymbol{m}\big]+\log P(\Theta)\end{aligned}$$

$$\boldsymbol{m}:=\begin{bmatrix}\boldsymbol{A}^{-1}\boldsymbol{\mu}_{\rm n}\\ \boldsymbol{B}^{-1}\boldsymbol{\mu}_{\rm e}\\ \boldsymbol{0}\end{bmatrix},\boldsymbol{\Lambda}^{-1}:=\begin{bmatrix}\boldsymbol{A}^{\rm T}\boldsymbol{\Sigma}_{\rm n}^{-1}\boldsymbol{A} & \boldsymbol{O} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{B}^{\rm T}\boldsymbol{\Sigma}_{\rm e}^{-1}\boldsymbol{B} & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{\Sigma}_{\rm f}^{-1}\end{bmatrix}$$

となる。ここで，${\rm tr}(\cdot)$ は行列のトレースを表す。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，条件付きガウス分布の性質から $\mathbb{E}[\boldsymbol{x}|\boldsymbol{y};\Theta']$ と $\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\rm T}|\boldsymbol{y};\Theta']$ を計算する [7]。$\boldsymbol{y}_{\rm n}$, $\boldsymbol{y}_{\rm e}$, $\boldsymbol{y}_{\rm f}$ に対応するように，$\mathbb{E}[\boldsymbol{x}|\boldsymbol{y};\Theta']$ および $\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\rm T}|\boldsymbol{y};\Theta']$ を以下のように区分表現する。

$$\mathbb{E}[\boldsymbol{x}|\boldsymbol{y};\Theta']=\begin{bmatrix}\bar{\boldsymbol{x}}_{\rm n}\\ \bar{\boldsymbol{x}}_{\rm e}\\ \bar{\boldsymbol{x}}_{\rm f}\end{bmatrix},\ \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\rm T}|\boldsymbol{y};\Theta']=\begin{bmatrix}\boldsymbol{R}_{\rm n} & * & *\\ * & \boldsymbol{R}_{\rm e} & *\\ * & * & \boldsymbol{R}_{\rm f}\end{bmatrix}$$

各パラメータの M ステップ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

$$\begin{aligned}\mathcal{I}_1(s):=-\frac{1}{2}\sum_{k=1}^{K}&(\boldsymbol{o}[k]-\boldsymbol{c}_{s_k}[k])^{\rm T}\boldsymbol{\Upsilon}^{-1}(\boldsymbol{o}[k]-\boldsymbol{c}_{s_k}[k])\\ &+\log P(s_1)+\sum_{k=2}^{K}\log P(s_k|s_{k-1})\end{aligned} \tag{10}$$

である。ただし，$\boldsymbol{o}[k]:=([\boldsymbol{A}\bar{\boldsymbol{x}}_{\rm n}]_k,[\boldsymbol{B}\bar{\boldsymbol{x}}_{\rm e}]_k)^{\rm T}$ 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 探索により導出できる。

2) ノート制御パラメータ: $\varphi^{(i)}$ と $\psi^{(i)}$ に関する事前分布を $\varphi^{(i)}\sim\mathcal{N}(\mu_\varphi,\sigma_\varphi^2)$, $\psi^{(i)}\sim\mathcal{N}(\mu_\psi,\sigma_\psi^2)$ とする。Q 関数の中で $\varphi^{(i)}$ と $\psi^{(i)}$ に関する項は，

$$\begin{aligned}\mathcal{I}_2(\varphi^{(i)},\psi^{(i)})=&|\mathcal{T}_{{\rm S}_{i,\cdot}}|\log(1+2\varphi^{(i)}\psi^{(i)}+(\varphi^{(i)})^2)\\ &-\frac{1}{2\sigma_{\rm n}^2}{\rm tr}((\boldsymbol{A}^{(i)})^{\rm T}\boldsymbol{A}^{(i)}\boldsymbol{R}_{\rm n})+\frac{1}{\sigma_{\rm n}^2}([\boldsymbol{\mu}_{\rm n}]_{\mathcal{T}_{{\rm S}_{i,\cdot}}})^{\rm T}\boldsymbol{A}^{(i)}\bar{\boldsymbol{x}}_{\rm n}\\ &-\frac{1}{2\sigma_\varphi^2}(\varphi^{(i)}-\mu_\varphi)^2-\frac{1}{2\sigma_\psi^2}(\psi^{(i)}-\mu_\psi)^2\end{aligned}$$

$$\mathcal{T}_{{\rm S}_{i,\cdot}}=\{k|s_k\in\{{\rm S}_{i,1},\ldots,{\rm S}_{i,J}\}\} \tag{11}$$

となる。ここで，$|\mathcal{T}|$ は集合 $\mathcal{T}$ の要素数を表す。また，$[\boldsymbol{\mu}]_{\mathcal{T}}$ は，集合 $\mathcal{T}$ の要素を添え字として，その添え字に相当する $\boldsymbol{\mu}$ の要素を取り出した部分ベクトルを表す。今，定数行列 $\boldsymbol{U}_0,\boldsymbol{U}_1,\boldsymbol{U}_2$ を用いて $\boldsymbol{A}^{(i)}$ は，

$$\boldsymbol{A}^{(i)}=(\varphi^{(i)})^2[\boldsymbol{U}_2]_{\mathcal{T}_{{\rm S}_{i,\cdot}}}+\varphi^{(i)}\psi^{(i)}[\boldsymbol{U}_1]_{\mathcal{T}_{{\rm S}_{i,\cdot}}}+[\boldsymbol{U}_0]_{\mathcal{T}_{{\rm S}_{i,\cdot}}}$$

と表現される。ここで，$[\boldsymbol{U}]_{\mathcal{T}}$ は集合 $\mathcal{T}$ の要素を添え字として，行列 $\boldsymbol{U}$ からその添え字に相当する行ベクトルを取り出して構成される部分行列を意味する。ニュートン・ラフソン法を利用して，$\mathcal{I}_2(\varphi^{(i)},\psi^{(i)})$ を最大化する $\varphi^{(i)}$ と $\psi^{(i)}$ が数値的に導出される。

3) 表現制御パラメータ: ノート制御パラメータと同様の手順で更新式を導出できる。ただし，推定すべきパラメータ数が多いため，今回は $\xi=3$ に固定する。

4) その他のパラメータ: d_i と $B_{\rm e}^{(i,j)}$ の事前分布をそれぞれ $d_i\sim\mathcal{N}(0,\sigma_{\rm d}^2)$ と $B_{\rm e}^{(i,j)}\sim\mathcal{N}(\mu_{B^{(i,j)}},\sigma_B^2)$ とする。残されたパラメータの更新式は

$$d_i=\frac{1}{|\mathcal{T}_{{\rm S}_{i,\cdot}}|+\sigma_{\rm n}^2/\sigma_{\rm d}^2}\sum_{k\in\mathcal{T}_{{\rm S}_{i,\cdot}}}([\boldsymbol{A}\bar{\boldsymbol{x}}_{\rm n}]_k-A_{\rm n}^{(i)}) \tag{12}$$

$$B_{\rm e}^{(i,j)}=\frac{1}{|\mathcal{T}_{{\rm S}_{i,j}}|/\sigma_{\rm e}^2+1/\sigma_B^2}\Big(\sum_{k\in\mathcal{T}_{{\rm S}_{i,j}}}\frac{[\boldsymbol{B}\bar{\boldsymbol{x}}_{\rm e}]_k}{\sigma_{\rm e}^2}+\frac{\mu_{B^{(i,j)}}}{\sigma_B^2}\Big)$$

$$\mathcal{T}_{{\rm S}_{i,j}}=\{k|s_k={\rm S}_{i,j}\} \tag{13}$$

$$\sigma_{\rm n}^2=\Big({\rm tr}(\boldsymbol{A}^{\rm T}\boldsymbol{A}\boldsymbol{R}_{\rm n})-2\boldsymbol{\mu}_{\rm n}^{\rm T}\boldsymbol{A}\bar{\boldsymbol{x}}_{\rm n}+\boldsymbol{\mu}_{\rm n}^{\rm T}\boldsymbol{\mu}_{\rm n}\Big)/K \tag{14}$$

$$\sigma_{\rm e}^2=\Big({\rm tr}(\boldsymbol{B}^{\rm T}\boldsymbol{B}\boldsymbol{R}_{\rm e})-2\boldsymbol{\mu}_{\rm e}^{\rm T}\boldsymbol{B}\bar{\boldsymbol{x}}_{\rm e}+\boldsymbol{\mu}_{\rm e}^{\rm T}\boldsymbol{\mu}_{\rm e}\Big)/K \tag{15}$$

$$\sigma_{\rm f}^2={\rm tr}(\boldsymbol{R}_{\rm f})/K \tag{16}$$

と導出される。無声音のため F_0 が推定されない区間の取り扱いや各パラメータの実際の更新方法については，文献 [9] を参照されたい。

4　評価実験

実測の F_0 軌跡を用いて，提案手法の性能を評価する。「RWC 研究用音楽データベース：ポピュラー音楽」(RWC-MDB-P-2001)[10] における，歌手名：緒方智美，曲番号：No.07，曲名：PROLOGUE のメロディの F_0 を手作業でラベル付した結果 [11] を実測 F_0 軌跡として利用した。本来ならば音響信号から F_0 を推定すべきであるが，今回は提案手法の性能の上限を調べるためにこのようなデータを利用した。F_0 は 10ms ごとにラベル付されているので，アップサンプリングによって，5ms ごとの実測 F_0 軌跡を得た。同楽曲の楽譜情報はその MIDI データを利用した。

パラメータ推定アルゴリズムの設定値を以下に示す。I と $\{A_{\mathrm{n}}^{(i)}\}_{i=1}^{I}$ は MIDI データから与えられる。表現指令信号を構成するための状態数 J は 5 とした。HMM の状態遷移確率は，$\phi_{\mathrm{S}_{i,1},\mathrm{S}_{i,1}} = \log(0.9999 \times (J-1)/J)$，$\phi_{\mathrm{S}_{i,1},\mathrm{S}_{i,j}} = \log(0.9999/J)$，$\phi_{\mathrm{S}_{i,1},\mathrm{S}_{i+1,1}} = \log(0.0001)$，$\phi_{\mathrm{S}_{i,j},\mathrm{S}_{i,j}} = \log(0.9999)$，$\phi_{\mathrm{S}_{i,j},\mathrm{S}_{i,1}} = \log(0.0001)$，$(1 \le i \le I,\ 2 \le j \le J)$ とした。パラメータの事前分布における固定パラメータは，$\mu_\varphi = 6$，$\sigma_\varphi^2 = 0.1$，$\mu_\psi = 0.6$，$\sigma_\psi^2 = 0.02$，$\sigma_{\mathrm{d}}^2 = 2500$，$\sigma_B^2 = 100$，$\mu_{B^{(i,1)}} = 0$，$\mu_{B^{(i,2)}} = 30$，$\mu_{B^{(i,3)}} = -30$，$\mu_{B^{(i,4)}} = 60$，$\mu_{B^{(i,5)}} = -60$，$(1 \le i \le I)$ とした。これらは文献 [1] と予備実験に基づいて決定した。

実測 F_0 軌跡からノート指令信号と表現指令信号，制御機構の出力であるノート成分と表現成分の推定結果を Fig. 3 に示す。これらを足し合わせたものを合成 F_0 軌跡とし，定性的には実測 F_0 軌跡に近い軌跡が得られた。また，微細変動成分パラメータである σ_{f}^2 の値は 187.7 であるため，振幅が 13 cent 程度の変動成分が推定され，先行研究の知見 [1] と整合する。無声音の区間を欠損データとして扱うため，合成 F_0 軌跡ではその部分が周囲の F_0 から補間される。

Fig. 4 は具体的に，ポルタメントとビブラートが生成される様子を示す。楽譜（MIDI）では一つの音を伸ばし続けるよう記述されているものの，歌手の表現意図によって，半音だけ滑らかに変化させるポルタメントが観測される。このとき，ノート指令信号は同一音符区間として推定され，表現指令信号は音高を段階的に変化させる指令が推定される。一方，実測 F_0 軌跡には必ずしも正弦波とは言えないビブラートが観測される。提案法では，部分的に矩形変化する表現指令が推定されてビブラートを生成する。ポルタメントとビブラートの生成のどちらにおいても合成 F_0 軌跡は実測 F_0 軌跡に近い軌跡が得られた。

5　おわりに

歌声 F_0 軌跡に含まれる動的変動成分を楽譜情報と分離して抽出することを目的として，2 つの指令信号によって駆動される F_0 軌跡の生成過程を記述し，モデルパラメータの推定アルゴリズムを提案した。評価実験では，実測の F_0 軌跡から推定される歌唱者の表現意図を考察した。提案モデルの有効性を主観的に評価する聴取実験も行なっているので，文献 [9] も参照されたい。今後の課題は，抽出された動的変動成分のパターンを歌唱者ごとに学習することである。

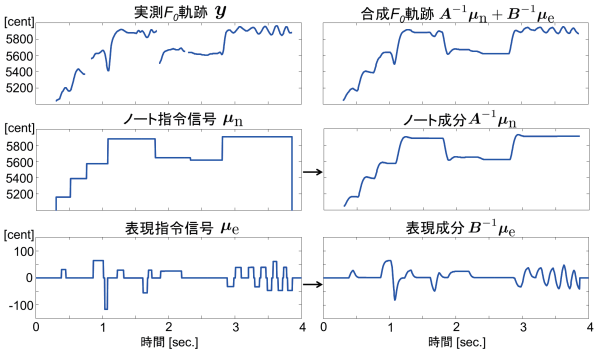

Fig. 3　実測 F_0 軌跡に対するパラメータ推定結果

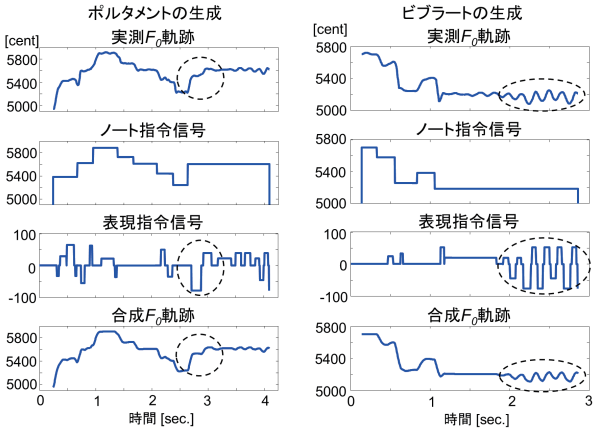

Fig. 4　ポルタメントとビブラートの生成

参考文献

[1] Saitou *et al.*, *Proc. EUROSPEECH'09*, pp. 832–835, 2009.

[2] Mase *et al.*, *Proc. INTERSPEECH'10*, pp. 845–848, 2010.

[3] 右田ほか, 情処論文誌, Vol.52, No.5, pp. 1910–1922, 2011.

[4] 中野ほか, 情処論文誌, Vol.52, No.12, pp. 3853–3867, 2011.

[5] Fujisaki, In *Vocal Physiology: Voice Production, Mechanisms and Functions*, (O. Fujimura, ed.) Raven Press, pp. 347–355, 1988.

[6] 柏野ほか, 音講論集, 2-9-1, pp. 625–626 1998.

[7] 大石ほか, 音講論集, 1-8-19 pp. 279–282, 2011.

[8] Kameoka *et al.*, *Proc. SAPA'10*, pp. 43–48, 2010.

[9] 大石ほか, 情処研報, 2012–MUS94–12, 2012.

[10] 後藤ほか, 情処論文誌, Vol.45, No.3, pp. 728–738, 2004.

[11] Goto, *Proc. ISMIR'06*, 2006.