3-P-31

複数振動基底に基づく歌声のF0動特性の統計的モデリング*

◎大石康智, 亀岡弘和, 持橋大地, 永野秀尚, 柏野邦夫 (NTT CS 研)

1 はじめに

歌声の声の高さを表す基本周波数（F0）系列には，階段状に変化する旋律成分と，ビブラートやオーバーシュートのような歌唱者の歌唱動作を表す動的な変動成分が複雑に重ね合わされている（図 1）．この F0 動的変動成分は，歌声知覚および個人性知覚に寄与する特徴であることが知られており [1, 2]，歌唱スタイルの記述やそれを利用した歌声検索，歌声評価 [4, 3, 8] のための有用な尺度となる．また，表情豊かかつ多様な歌声合成 [5, 6, 7] を実現するためにも必要不可欠な成分と言える．

これまで多くの研究では，2 次系モデルを利用して，歌唱の F0 動的変動成分を表現した [1, 9, 10]．これらの研究では，日本語の話声の F0 パターンを表現する藤崎モデル [11] を参考にする．藤崎モデルは，臨界制動 2 次系のインパルス応答とステップ応答を利用して，日本語の句頭から句末に向けて緩やかに下降するフレーズ成分と，語句に対応して上昇下降するアクセント成分を表現し，これらを重畳することで，F0 系列を記述する．ただし，歌声の旋律に伴った急激な F0 の上昇・下降の制御及び，ビブラートのような周期的な振動は，臨界制動系では表現できないため，歌声の F0 制御モデルでは 2 次系の伝達関数

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

の減衰率 ζ によって表現される，指数減衰 $(\zeta > 1)$，減衰振動 $(0 < \zeta < 1)$，臨界制動 $(\zeta = 1)$，定常振動 $(\zeta = 0)$ を利用する [1]．齋藤らは，減衰率 ζ と固有周波数 Ω を聴取実験に基づいて手動で調整し，それらによって得られる式 (1) のインパルス応答を，階段状に変化する信号に畳み込んで生成される F0 系列を利用して，表情豊かな歌声合成音を実現した [1]．

これに対し，我々は旋律成分および 2 次系の制御パラメータ ζ, Ω がいずれも未知の下で，観測される F0 系列だけから，これらを同時に推定する統計的手法の枠組を提唱した [12]．これは，旋律成分を表す隠れマルコフモデル（HMM）と，差分近似に基づく式 (1) のパラメトリックな表現によって，最尤のモデルパラメータを反復推定するアルゴリズムであった．しかし，このアルゴリズムでは，モデルパラメータの推定性能が悪かった．この理由は，図 1 の上部に示すように F0 系列には各ノート（音符）の切り替わりやオーバーシュート，ビブラートのような動的変動成分が所々で混在するのに対し，自由度が高すぎるモデルで，これらの動的変動成分を学習しようとしていたためである．本報告は，このようなオーバーフィッティングの問題を解消するために，2 つの方策を導入した歌声の F0 生成過程の新しい確率モデルを提案する．

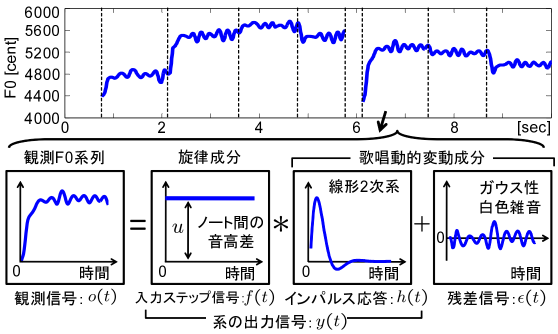

Fig. 1 想定する歌声の F0 生成過程の概略図

1. モデルの自由度を効果的に下げる目的で，2 次系のインパルス応答 $h(t)$ を，あらかじめ用意したいくつかの振動基底の疎（スパース）な線形和によって構成する．

2. ノートが切り替わる時点を始点終点と考え，F0 系列をいくつかのセグメントに分割し，セグメントごとに，図 1 に示す信号の生成過程を仮定する．$h(t)$ と $\epsilon(t)$ はどちらも動的変動成分を表す信号であるが，$h(t)$ はノートの切り替わりやオーバーシュートを表現する信号の大局的動特性を，$\epsilon(t)$ はビブラートのような音高が安定するときの局所的動特性を表現する．

このような信号表現方法に基づき，パラメータ最適化アルゴリズムを導く．そして，パラメータの推定性能を評価し，推定されたパラメータの歌唱者ごとの違いから歌唱スタイルの違いについて分析する．

2 複数の振動基底を利用した線形 2 次系のインパルス応答表現

式 (1) のラプラス逆変換によって得られるインパルス応答は，ζ の値によって場合分けされる．

$$h(t) = \begin{cases} \frac{\Omega e^{-\zeta\Omega t}}{2\sqrt{\zeta^2-1}}\left(e^{\sqrt{\zeta^2-1}\Omega t} - e^{-\sqrt{\zeta^2-1}\Omega t}\right), & (\zeta > 1) \\ \frac{\Omega e^{-\zeta\Omega t}}{\sqrt{1-\zeta^2}}\left(\sin(\sqrt{1-\zeta^2}\Omega t\right), & (0 < \zeta < 1) \\ \Omega^2 t e^{-\Omega t}, & (\zeta = 1) \\ \Omega \sin(\Omega t), & (\zeta = 0) \end{cases}$$

これらのインパルス応答をサンプリング周期 Δ で離散化すると，図 1 の系の入出力関係は $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{f}$ と記述できる．ここで，$\boldsymbol{y} = [y_1, y_2, \ldots, y_N]^{\mathrm{T}}$，$\boldsymbol{f} = [f_1, f_2, \ldots, f_N]^{\mathrm{T}}$ は，出力信号 $y(t)$ と入力ステップ信

*Statistical Modeling of F0 Dynamics in Singing Voices Based on Multiple Oscillation Bases. by OHISHI, Yasunori, KAMEOKA, Hirokazu, MOCHIHASHI, Daichi, NAGANO Hidehisa, KASHINO Kunio (NTT Communication Science Laboratories)

号 $f(t)$ をサンプリング周期 Δ で離散化した時系列信号のベクトルを表す（N は信号の長さ）．この $\mathbf{\Phi}$ が，系のインパルス応答を表し，例えば，$\zeta = 1$ の場合，$\mathbf{\Phi}$ は下三角行列

$$\mathbf{\Phi} = \begin{bmatrix} \Omega^2\Delta e^{-\Omega\Delta} & & & \mathbf{0} \\ 2\Omega^2\Delta e^{-2\Omega\Delta} & \Omega^2\Delta e^{-\Omega\Delta} & & \\ \vdots & \ddots & \ddots & \\ N\Omega^2\Delta e^{-N\Omega\Delta} & \dots & 2\Omega^2\Delta e^{-2\Omega\Delta} & \Omega^2\Delta e^{-\Omega\Delta} \end{bmatrix}$$

となる．しかし，$h(t)$ は，上式のように複数の場合からなるので，行列 $\mathbf{\Phi}$ を以下のように構成する．

$$\mathbf{\Phi}^{-1} \simeq w_1\mathbf{\Upsilon}^{(1)} + w_2\mathbf{\Upsilon}^{(2)} + \ldots + w_I\mathbf{\Upsilon}^{(I)} \quad (2)$$

ここでは，予め手動で ζ, Ω を決定し，I 個の振動現象を表すインパルス応答 $\{\mathbf{\Phi}^{(1)}, \mathbf{\Phi}^{(2)}, \ldots, \mathbf{\Phi}^{(I)}\}$ を計算する．そして，これらの逆行列 $\mathbf{\Upsilon}^{(i)} := (\mathbf{\Phi}^{(i)})^{-1}$（逆フィルタのインパルス応答を表す．以後，振動基底と呼ぶ）の重み付き和で $\mathbf{\Phi}^{-1}$ を近似する．ただし，この重みパラメータ $\boldsymbol{w} := \{w_1, w_2, \ldots, w_I\}$ は疎（スパース）となるように正則化する．これは，$\mathbf{\Phi}^{-1}$ が，ある限られた種類の振動基底のみによって表現されることを意味し，モデルの自由度を効果的に下げる目的として，このような操作を行う．後に説明するが，これは $\boldsymbol{w}$ の事前確率を課すことで実現される．したがって，系の入出力関係は以下のように表現される．

$$(w_1\mathbf{\Upsilon}^{(1)} + w_2\mathbf{\Upsilon}^{(2)} + \ldots + w_I\mathbf{\Upsilon}^{(I)})\boldsymbol{y} = \boldsymbol{f} \quad (3)$$

便宜上 $\mathbf{\Psi} := w_1\mathbf{\Upsilon}^{(1)} + w_2\mathbf{\Upsilon}^{(2)} + \ldots + w_I\mathbf{\Upsilon}^{(I)}$ とおく．

3　2 次系 F0 生成過程の統計的モデリング

2 次系の入出力関係を統計的にモデル化する．

3.1　入力信号と出力信号の確率モデル

入力信号 $\boldsymbol{f}$ はノート間の音高差を表すよう，ステップ信号を想定する．そのために，常に同じ値をもつベクトル $\boldsymbol{u} = [u_1, \ldots, u_N]^{\mathrm{T}} = u[1, 1, \ldots, 1]^{\mathrm{T}} = u\mathbf{1}$ を用意する．ここで，スカラー値 u は音高差を表すパラメータ，$\mathbf{1}$ は N 個の 1 の値が並ぶベクトルとする．このベクトル $\boldsymbol{u}$ を平均とする多次元ガウス分布 $\mathcal{N}(\boldsymbol{u}, \alpha\boldsymbol{I}_N)$ から生成される確率変数として，入力信号 $\boldsymbol{f}$ を表現する．α は分散を表す超パラメータであり，あらかじめ手動で値を設定する．$\boldsymbol{I}_N$ は $N \times N$ の単位行列を示す．

出力信号 $\boldsymbol{y}$ は，ガウス分布に従う変数集合 $\boldsymbol{f}$ の線形結合（$\boldsymbol{y} = \mathbf{\Psi}^{-1}\boldsymbol{f}$）であるから，$\boldsymbol{y}$ 自身もガウス分布に従い，$\boldsymbol{y} \sim \mathcal{N}\left(\mathbf{\Psi}^{-1}\boldsymbol{u},\ \alpha\mathbf{\Psi}^{-1}(\mathbf{\Psi}^{-1})^{\mathrm{T}}\right)$ と書ける．

3.2　尤度関数と事前確率

残差信号 $\boldsymbol{\epsilon} = [\epsilon_1, \epsilon_2, \ldots, \epsilon_N]^{\mathrm{T}}$ は，ガウス性白色雑音 $\mathcal{N}(\mathbf{0}, \beta\boldsymbol{I}_N)$ に従うとし，観測 F0 系列 $\boldsymbol{o} = [o_1, o_2, \ldots, o_N]^{\mathrm{T}}$ は，系の出力信号 $\boldsymbol{y}$ に残差信号 $\boldsymbol{\epsilon}$ が加わった信号 $\boldsymbol{o} = \boldsymbol{y} + \boldsymbol{\epsilon}$ と仮定する．ここで，β は残差信号の分散を表す超パラメータである．$\boldsymbol{y}$ と $\boldsymbol{\epsilon}$ は互いに独立であると仮定すると，観測信号 $\boldsymbol{o}$ が与えられたときのモデルパラメータ $\Theta := \{\boldsymbol{w}, u, \beta\}$ の対数尤度関数は，以下のように書ける．

$$\log P(\boldsymbol{o}|\Theta) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{\Sigma}| - \frac{1}{2}(\boldsymbol{o}-\boldsymbol{\mu})^{\mathrm{T}}\mathbf{\Sigma}^{-1}(\boldsymbol{o}-\boldsymbol{\mu}) \quad (4)$$

$$\left(\boldsymbol{\mu} = \mathbf{\Psi}^{-1}\boldsymbol{u},\ \mathbf{\Sigma} = \alpha\mathbf{\Psi}^{-1}(\mathbf{\Psi}^{-1})^{\mathrm{T}} + \beta\boldsymbol{I}_N\right)$$

Θ の事前確率 $P(\Theta)$ は，各要素の独立性 $P(\Theta) = P(\boldsymbol{w})P(u)P(\beta)$ を仮定し，u と β はそれぞれ一様分布に従うとする．パラメータ $\boldsymbol{w}$ の要素には疎な制約をもたせるため，その事前確率は一般化正規分布

$$P(\boldsymbol{w}) = \prod_{i=1}^{I}\frac{\lambda p}{2\Gamma(1/p)}\exp^{-\lambda^p|w_i|^p} \quad (5)$$

に従うとする．p, λ は一般化正規分布の形状を規定する定数であり，$0 < p < 2$ のとき $p(\boldsymbol{w})$ は優ガウス的となり，スパースネスを測るための尺度となる．

4　パラメータ最適化アルゴリズム

観測 F0 系列 $\boldsymbol{o}$ が与えられたとき，事後確率 $P(\Theta|\boldsymbol{o}) \propto P(\boldsymbol{o}|\Theta)P(\Theta)$ を最大化するパラメータ Θ を決定したいが，① $\boldsymbol{o}$ が出力信号 $\boldsymbol{y}$ と残差信号 $\boldsymbol{\epsilon}$ の和で構成される，② 尤度関数が $\boldsymbol{w}$ に関して非線形となるため解析的に求めることは難しい．そこで，① EM 法 [13] を適用して，その E-step で，$\boldsymbol{o}$ を $\boldsymbol{y}$ と $\boldsymbol{\epsilon}$ に分離する，② M-step に補助関数法 [14] を適用して，Q 関数の補助関数を設計する，からなる 2 つの方策に基づいて，パラメータ最適化アルゴリズムを導出する．

4.1　完全データの定義

この事後推定問題に EM 法を適用する際の最初のステップは完全データを定義することである．ここでは，$\boldsymbol{y}$ と $\boldsymbol{\epsilon}$ を完全データ $\boldsymbol{x} = [\boldsymbol{y}^{\mathrm{T}}\ \boldsymbol{\epsilon}^{\mathrm{T}}]^{\mathrm{T}}$ と見なして，EM 法を適用する．不完全データと完全データの関係は，$\boldsymbol{o} = \boldsymbol{H}\boldsymbol{x}$ となり，ここで，$\boldsymbol{H} := [\boldsymbol{I}_N\ \boldsymbol{I}_N]$ とする．$\boldsymbol{x}$ と現在のパラメータ Θ' が与えられたときの，対数尤度関数の条件付き期待値を計算し，さらに $\log P(\Theta)$ を加算すると，以下のような Q 関数を得る．

$$Q(\Theta, \Theta') \stackrel{c}{=} \frac{1}{2}\Big[\log|\mathbf{\Lambda}^{-1}| - \mathrm{tr}\left(\mathbf{\Lambda}^{-1}\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}|\boldsymbol{o};\Theta']\right) + 2\boldsymbol{m}^{\mathrm{T}}\mathbf{\Lambda}^{-1}\mathbb{E}[\boldsymbol{x}|\boldsymbol{o};\Theta'] - \boldsymbol{m}^{\mathrm{T}}\mathbf{\Lambda}^{-1}\boldsymbol{m}\Big] + \log P(\Theta)$$

$$\left(\boldsymbol{m} := \begin{bmatrix}\mathbf{\Psi}^{-1}\boldsymbol{u} \\ \mathbf{0}\end{bmatrix},\ \mathbf{\Lambda}^{-1} := \begin{bmatrix}\frac{1}{\alpha}\mathbf{\Psi}^{\mathrm{T}}\mathbf{\Psi} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\beta}\boldsymbol{I}_N\end{bmatrix}\right) \quad (6)$$

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

$$\mathbb{E}[\boldsymbol{x}|\boldsymbol{o};\Theta'] = \boldsymbol{m} + \mathbf{\Lambda}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\mathbf{\Lambda}\boldsymbol{H}^{\mathrm{T}})^{-1}(\boldsymbol{o} - \boldsymbol{H}\boldsymbol{m})$$

$$\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}|\boldsymbol{o};\Theta'] = \mathbf{\Lambda} - \mathbf{\Lambda}\boldsymbol{H}^{\mathrm{T}}(\boldsymbol{H}\mathbf{\Lambda}\boldsymbol{H}^{\mathrm{T}})^{-1}\boldsymbol{H}\mathbf{\Lambda} + \mathbb{E}[\boldsymbol{x}|\boldsymbol{o};\Theta']\mathbb{E}[\boldsymbol{x}|\boldsymbol{o};\Theta']^{\mathrm{T}} \quad (7)$$

と書ける．後の計算のため，$\boldsymbol{y}$, $\boldsymbol{\epsilon}$ に対応するように $\mathbb{E}[\boldsymbol{x}|\boldsymbol{o};\Theta']$ と $\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}|\boldsymbol{o};\Theta']$ を区分表現する．

$$\mathbb{E}[\boldsymbol{x}|\boldsymbol{o};\Theta'] = \begin{bmatrix} \bar{\boldsymbol{x}}_y \\ \bar{\boldsymbol{x}}_\epsilon \end{bmatrix},\ \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}|\boldsymbol{o};\Theta'] = \begin{bmatrix} \boldsymbol{R}_y & * \\ * & \boldsymbol{R}_\epsilon \end{bmatrix}$$

ここで，$\bar{\boldsymbol{x}}_y$ と $\bar{\boldsymbol{x}}_\epsilon$ はどちらも長さ N のベクトルであり，$\boldsymbol{R}_y$ と $\boldsymbol{R}_\epsilon$ はどちらも $N \times N$ の正方行列を表す．

4.2 M-step 更新式

式 (6) からパラメータ集合 Θ に関連する項を取り出し，最大化する目的関数を，

$$\begin{aligned} f(\Theta) := &-\frac{N}{2}\log\alpha\beta + \sum_{n=1}^{N}\log\Big(\sum_{i=1}^{I} w_i \Upsilon_{n,n}^{(i)}\Big) \\ &+ \frac{1}{\alpha}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Psi}\bar{\boldsymbol{x}}_y - \frac{1}{2\alpha}\mathrm{tr}(\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi}\boldsymbol{R}_y) \\ &- \frac{1}{2\beta}\mathrm{tr}(\boldsymbol{R}_\epsilon) - \frac{1}{2\alpha}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} - \lambda^p \sum_{i=1}^{I}|w_i|^p \end{aligned} \tag{8}$$

と改めて定義する．ここで，$\Upsilon_{n,n}^{(i)}$ は $\boldsymbol{\Upsilon}^{(i)}$ の n 行 n 列目の対角要素を表す．この最大化問題を解く更新式を補助関数法 [14] により導く．式 (8) で与えられる目的関数の補助関数を以下の 2 つの不等式（証明略）を用いて導く．

$$\begin{aligned} &\sum_{n=1}^{N}\log\Big(\sum_{i}^{I} w_i \Upsilon_{n,n}^{(i)}\Big) \geq \sum_{n=1}^{N}\sum_{i=1}^{I}\gamma_{i,n}\log\frac{w_i\Upsilon_{n,n}^{(i)}}{\gamma_{i,n}} \\ &|w_i|^p \leq p|\bar{w}_i|^{p-1}w_i + |\bar{w}_i|^p - p|\bar{w}_i|^p,\ (0 < p \leq 1) \end{aligned} \tag{9}$$

ここで，補助変数 $\bar{\boldsymbol{w}} := \{\bar{w}_1, \bar{w}_2, \ldots, \bar{w}_I\}$，$\boldsymbol{\gamma} := \{\gamma_{1,1}, \ldots, \gamma_{I,N}\}$ を定義する．式 (9) の二つの不等式を式 (8) に代入すると，

$$\begin{aligned} f^{+}(\Theta, \bar{\boldsymbol{w}}, \boldsymbol{\gamma}) := &-\frac{N}{2}\log\alpha\beta + \sum_{n=1}^{N}\sum_{i=1}^{I}\gamma_{i,n}\log\frac{w_i\Upsilon_{n,n}^{(i)}}{\gamma_{i,n}} \\ &+ \frac{1}{\alpha}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Psi}\bar{\boldsymbol{x}}_y - \frac{1}{2\alpha}\mathrm{tr}(\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi}\boldsymbol{R}_y) - \frac{1}{2\beta}\mathrm{tr}(\boldsymbol{R}_\epsilon) - \frac{1}{2\alpha}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{u} \\ &- \lambda^p\sum_{i=1}^{I}\Big(p|\bar{w}_i|^{p-1}w_i + |\bar{w}_i|^p - p|\bar{w}_i|^p\Big) \end{aligned} \tag{10}$$

を得る．このとき，$f(\Theta) \geq f^{+}(\Theta, \bar{\boldsymbol{w}}, \boldsymbol{\gamma})$ が成り立ち，等号成立は，$\bar{w}_i = w_i$，$\gamma_{i,n} = \frac{\bar{w}_i\Upsilon_{n,n}^{(i)}}{\sum_{i'=1}^{I}\bar{w}_{i'}\Upsilon_{n,n}^{(i')}}$，$(i = 1, 2, \ldots, I,\ n = 1, 2, \ldots, N)$ のときであるため，式 (10) は補助関数の定義を満たす．

式 (10) を $w_{i'}$ に関して微分して 0 とおくと，

$$\begin{aligned} &\frac{1}{\alpha}\sum_{i=1}^{I}\mathrm{tr}\big({\boldsymbol{R}_y}^{\mathrm{T}}{\boldsymbol{\Upsilon}^{(i)}}^{\mathrm{T}}\boldsymbol{\Upsilon}^{(i')}\big)w_i - \frac{1}{\alpha}\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Upsilon}^{(i')}\bar{\boldsymbol{x}}_y \\ &\qquad + \lambda^p p|\bar{w}_{i'}|^{p-1} - \sum_{n=1}^{N}\frac{\gamma_{i',n}}{w_{i'}} = 0 \end{aligned} \tag{11}$$

を得る．ただし，式 (11) は，$i' = 1, 2, \ldots, I$ として，$w_1, w_2, \ldots, w_I$ に関する非線形な連立方程式となるため，Coordinate descent 法 [15] を利用して解く．

一方，$f^{+}(\Theta, \bar{\boldsymbol{w}}, \boldsymbol{\gamma})$ を u, β それぞれに関して微分して 0 とおくと，

$$u = \frac{1}{N}\boldsymbol{1}^{\mathrm{T}}\boldsymbol{\Psi}\bar{\boldsymbol{x}}_y, \qquad \beta = \frac{1}{N}\mathrm{tr}\big(\boldsymbol{R}_\epsilon\big) \tag{12}$$

が得られ，u と β の更新式となる．パラメータ最適化アルゴリズムの流れを以下にまとめる．

初期化: パラメータ Θ に初期値を与える．
E-step: 条件付き期待値 $\mathbb{E}[\boldsymbol{x}|\boldsymbol{o};\Theta']$，$\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}|\boldsymbol{o};\Theta']$，補助変数 $\bar{\boldsymbol{w}}$, $\boldsymbol{\gamma}$ の更新．
M-step: 式 (11)，(12) から，パラメータ Θ の更新．
収束判定: 式 (10) の値が収束していなければ，$\Theta' = \Theta$ として **E-step** に戻る．

5 評価実験

提案するパラメータ最適化アルゴリズムの有効性を評価する．まず，オーバーフィッティングの問題を改善できるかどうか確認するために，提案法によって F0 系列から分解されたステップ信号とインパルス応答信号を評価する．そのために，ランダムに決定された ζ, Ω, u（値の範囲は，$0 < \zeta < 2$，$0.05 < \Omega < 0.3$，$0 < u < 200$ とする）を使って，人工的に 100 個の擬似 F0 系列を作成した．ここで，N, α, β はそれぞれ 300，2，100 とした．そして，F0 系列ごとにアルゴリズムを適用して，パラメータを推定する．前処理として，まず $\boldsymbol{\Upsilon}$ を作成するために，ζ は 0 から 2 までの間を 0.02 刻みで，Ω は 0.05 から 0.3 までの間を 0.005 刻みで変化させた．その結果，$I = 3100$ となる．したがって，$w_i = 1/I$ を初期値とした．また，$\beta = 100$ を初期値とした．u の初期値は，F0 系列の中央値とした．$\alpha = 2$，$\lambda = 10000$，$p = 0.8$ に固定する．これらの初期値は実験的に決定した．

評価方法として，疑似 F0 系列のステップ信号の作成に用いた u と推定された u の二乗平均平方根誤差 (RMSE) を計算する．同様に，疑似 F0 系列の作成に用いた ζ, Ω から計算されるインパルス応答と，推定された $\boldsymbol{w}$ から計算されるインパルス応答の RMSE を計算する．ステップ信号とインパルス応答への分解性能が高ければ，オーバーフィッティングを解消できていると考える．表 1 は，それぞれの RMSE のすべての疑似 F0 系列に対する平均値を従来法 [12] と比較する．従来法に比べて値が小さくなり，モデルの自由度を下げた提案法の有効性を確認できた．

次に，実際の F0 系列を利用してパラメータの推定性能を定性的に評価する．声楽家，プロのポップス歌手，音楽的な訓練を受けていない素人 (それぞれ男女 1 名ずつの計 6 名) の歌声からなるデータベースを利用する．歌唱者は，きらきら星と喜びの歌を譜面を見ながら伴奏なしで歌唱した．2 パターンの日本語歌詞とハミングによる歌声を収録した．歌声信号は計 36 サンプルとなる．F0 は，YIN[17] を利用して，5ms ごとに推定される ($\Delta = 5$ms)．そして，半音が 100cent となるように，Hz を cent の単位に変換する．また，300ms 以内の F0 が推定されない無声音の区間は近傍の F0 値で線形補間を行う．本実験ではモデル

Table 1　二乗平均平方根誤差 (RMSE) による評価

	ステップ信号に関する RMSE	インパルス応答に関する RMSE
従来法	21.7	5.29×10^{-2}
提案法	0.67	2.98×10^{-4}

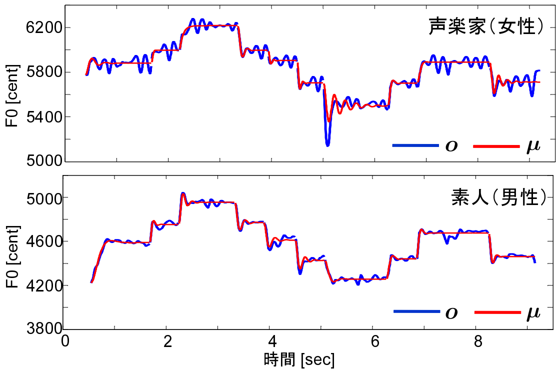

Fig. 2　観測 F0 系列 $\boldsymbol{o}$ と生成 F0 系列 $\boldsymbol{\mu}$ の比較

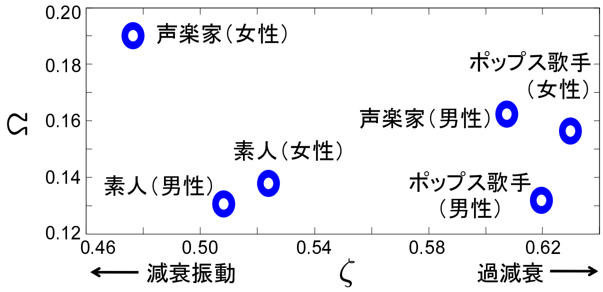

Fig. 3　歌唱者ごとに推定された ζ と Ω の平均値

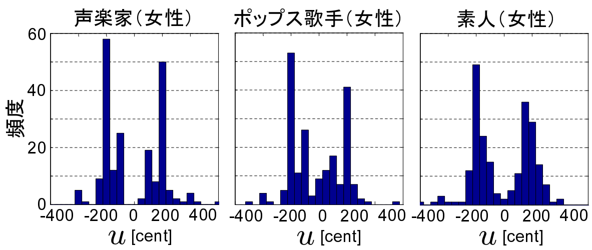

Fig. 4　歌唱者ごとに推定された u の頻度分布

の性能を評価するために，図 1 に示すように，F0 系列に対してセグメントを手作業でラベル付けし，各セグメントごとにパラメータを推定する．歌声データベース全体で，1323 個のセグメントをラベル付けした．前処理として，セグメントの最初の F0 値を，そのセグメントにおけるすべての F0 値から減算した．

図 2 は，各セグメントで推定されたパラメータ $\boldsymbol{w}, \beta, u$ から計算される式 (4) の $\boldsymbol{\mu}$（生成 F0 系列）と観測 F0 系列 $\boldsymbol{o}$ を比較する．声楽家と素人のどちらも，生成 F0 系列は，オーバーシュートやポルタメントのようなノートが切り替わるときの大局的な動特性を表現できている．素人に比べて，声楽家の観測 F0 系列には所々にビブラートが観測される．提案モデルでは，このような変動成分はすべて，ガウス性白色雑音に従うとし，分散パラメータ β によって表現される．この残差信号 $\boldsymbol{\epsilon}$ をより精緻にモデル化して，ビブラートのような周期的な変動成分をも特徴抽出することは今後の課題である．例えば，近年機械学習の分野で注目を集める Multiple Kernel Learning を利用したガウス過程 [16] に基づく信号表現との類推から，$\boldsymbol{\epsilon}$ が従う確率分布の共分散行列を周期カーネルによって表現する方法が考えられる．

図 3 は，各セグメントで推定される $\boldsymbol{w}$ の要素の最も大きい値 w_i に着目し，それに対応する ζ と Ω の，すべてのセグメントにわたる平均値を歌唱者ごとに計算した結果である．ζ が小さな値であることは，その振動現象がオーバーシュートのような減衰振動であることを意味する．一方で，Ω が小さな値であることは，ノートの立ち上がり時間が長いことを意味する．したがって、素人歌唱者は歌唱技術が乏しいため，ζ と Ω が他の歌唱者に比べて，小さい値になったと考えられる．図 4 は各セグメントで推定された u の歌唱者ごとの頻度分布を示す．声楽家やポップス歌手は，半音（100cent）の整数倍の位置にピークが観測されるが，素人は，そのピークが不鮮明となった．したがって，素人は正確な半音単位の音階で歌唱することが難しいと見ることができる．将来は，大規模歌声データベースを利用して，推定される ζ, Ω, u から歌唱スタイルについて詳細に分析する予定である．

6　おわりに

歌声の F0 動特性のモデリングとそのモデルパラメータ推定に関する新しい解法を検討した．これを応用した歌声合成インタフェース [18]，さらに話声の F0 生成過程の確率モデル [19] について現在検討中である．今後は，提案法を多変量化して MFCC 信号に適用し，声質の動特性の制御についても検討する予定である．

参考文献

[1] Saitou, T. et al., *Proc. WASSPA 2007*, pp. 215–218, 2007.
[2] Saitou, T. et al., *Proc. EUROSPEECH 2009*, pp. 832–835, 2009.
[3] Kako, T. et al, *Proc. ISMIR 2009*, pp. 393–397, 2009.
[4] Nakano, T. et al., *Proc. ICSLP 2006*, pp. 1706–1709, 2006.
[5] Sundberg, J., *Advances in Cognitive Psychology. Special issue on Music Performance*, Vol. 2, No. 2-3, pp. 131–143, 2006.
[6] Bonada, J. et al., *Proc. SMAC 2003*, 2003.
[7] Nakano, T. et al., *Proc. SMC 2009*, pp. 343–348, 2009.
[8] Proutskova, P. et al., *Proc. ISMIR 2009*, pp. 759–764, 2009.
[9] 相野邦夫ほか, 音講論集, 2-9-1, pp. 625–626 1998.
[10] Minematsu, N. et al., *Proc. SpeechProsody 2004*, pp. 487–490, 2004.
[11] Fujisaki, H., *Vocal Physiology: Voice Production, Mechanisms and Functions, (O. Fujimura, ed.)*, Raven Press, pp. 347–355, 1988.
[12] Ohishi, Y. et al., *Proc. ICSLP 2008*, pp. 139–142, 2008.
[13] Feder, M. et al., *IEEE Trans. ASSP*, Vol. 36, No. 4, pp. 477–489, 1988.
[14] Kameoka, H. et al., *Proc. ICASSP 2009*, pp. 3437–3440, 2009.
[15] Meng, X. L. et al., *Biometrika*, Vol. 80, pp. 267–278, 1993.
[16] Rasmussen, C. E. et al., *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, Mass, USA, 2006.
[17] de Cheveigné, A. et al., *JASA*, Vol. 111, No. 4, pp. 1917–1930, 2002.
[18] 大石康智ほか, 情処研報音楽情報科学, 2010–MUS–86–9, 2010.
[19] 亀岡弘和ほか, 音講論集, 1-1-3, 2010.