

# ばね質量系を利用した 合唱における歌声の $F_0$ ダイナミクスモデル化

川岸基成<sup>1</sup> 宮島千代美<sup>1</sup> 北岡教英<sup>1</sup> 武田一哉<sup>1</sup>

**概要:** 本報告では、合唱における歌声の基本周波数 ( $F_0$ ) に含まれる動的変動成分 (ダイナミクス) を歌声の引き込みという観点から分析し、そのモデル化手法について検討を行う。我々は、他者の歌声を聴きながら歌唱することにより生じる影響を歌声の引き込みと考え、1つの質点と2つのばねで構成されるばね質量系を利用することで、合唱における歌声の  $F_0$  ダイナミクスをモデル化する。モデルパラメータは最急降下法に基づいて推定され、合唱における歌声の  $F_0$  軌跡を 44.4 cent 以下の RMSE で生成することができる。評価実験では、歌唱力の異なる歌唱者では  $F_0$  ダイナミクスの特徴が異なると考えられるため、推定されたモデルパラメータを用いて歌唱者識別実験を行い、モデルの有効性を示した。

## 1. はじめに

合唱では、複数の歌唱者が同時に歌唱を行うため、他者の歌声から影響を受け、独唱とは異なった歌唱となる。例えば、独唱時と合唱時では、音色に関する音響特徴である歌唱ホルマント (singer's formant)[2] のパワーが異なること、合唱では他者の音圧に合わせて歌唱することが Rossingら [1] により報告されている。また、他者の歌声の音圧が非常に大きいと自身の音高を正確に制御することが困難となり、歌唱のパフォーマンスに影響を与えることが報告されている [3]。その他にも、ステージの設計や合唱団内の配置の決定に役立てることを目的として、各歌唱者の好みの音圧比 SOR(Self-to-Other Ratio) について調査した研究もある [4], [5]。このような歌唱の違いは、より自然な合唱音の生成にとって重要な要素であり、歌声情報処理にとって重要な課題である。

このような他者の行動に適応するような変化は、行動の「引き込み」 [6], [7], [8] と解釈することができ、音楽情報処理の分野においても様々な側面から研究されている。引き込みは、工学の分野において相互作用により振動体の位相や周波数が一致する現象として知られており、2つ以上の現象が同期に向かうことによる振る舞いの変化と定義することができる。文献 [9] では演奏者と聴取者の呼吸のリズムの相互作用について、文献 [10] ではボーカルとギターによる共同演奏におけるリズム及び周波数の引き込みにつ

いて分析している。文献 [11] では、より人間らしい演奏を行うアンサンブルシステムの構築を目指し、2者によるピアノ演奏のリズムの相互作用を分析し、モデル化している。合唱に関する研究では、歌声の音高の引き込みにかかる時間に着目し、歌唱技量を評価した研究 [12] がある。このように、合唱や合奏のような共同演奏における演奏者間の関係について様々な研究が行われており、合奏時における演奏者のリズムの制御を数学的にモデル化した研究もある [13]。しかしながら、合唱における歌唱の相互作用をモデル化する研究はほとんど行われていない。

本研究では合唱における歌唱を特徴づける数学的モデルを構築することを目指す。その第一段階として、歌声の基本周波数 ( $F_0$ ) に着目し、合唱における  $F_0$  の動的な特徴 ( $F_0$  ダイナミクス) をモデル化することを目的とする。歌声の  $F_0$  軌跡に含まれる動的な特徴には、ビブラート [14], [15] やオーバーシュート [16] などがある。これらの  $F_0$  の動的な特徴は、音楽表現において主要な役割を果たし、歌声の知覚に影響を及ぼすことが知られている [17], [18]。合唱における  $F_0$  ダイナミクスをモデル化することは、合唱歌唱の振る舞いを特徴付けるための重要な課題の一つである。

$F_0$  ダイナミクスに関する研究として、文献 [19] では、目標とする音高に引きつけられる  $F_0$  の挙動に着目し、 $F_0$  とその導関数で構成される相平面 ( $F_0, \dot{F}_0$ ) に  $F_0$  軌跡を描くことにより、その動的な変動を分布として表現している。ここでは、分布をパラメトリックに表現するために混合ガウス分布 (Gaussian Mixture Model: GMM) が利用されている。

<sup>1</sup> 名古屋大学大学院情報科学研究科  
Graduate School of Information Science, Nagoya University,  
Aichi, Japan

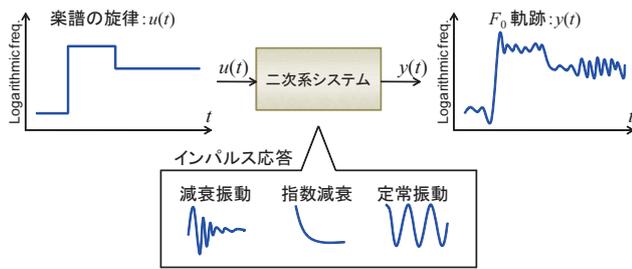


図 1 歌声の  $F_0$  生成過程の概略図

Fig. 1 Schematic of  $F_0$  creation process of singing voice.

$$(F_0, \dot{F}_0) \sim \sum_i \omega_i \mathcal{N}(F_0, \dot{F}_0; \bar{\mu}_i, \bar{\sigma}_i^2) \quad (1)$$

加古らは、歌声の  $F_0$  の動的特徴は相平面上に描かれる  $F_0$  軌跡によって特徴付けることができるという考え方を利用して、独唱時と合唱時では  $F_0$  ダイナミクスが異なることを示した [20]。また、ばね質量系を用いて合唱における歌声の  $F_0$  ダイナミクスを表現するモデルも提案したが、そのモデルについては深く議論されていない。そこで本稿では、加古らの提案したモデルに基づいて、他者の歌声を受聴しながら歌唱した歌声の  $F_0$  ダイナミクスをモデル化する手法について検討を行い、モデルパラメータの推定手法について議論する。

ばね質量系は、二次系のダイナミクスを分析する簡単なモデルであり、音声の  $F_0$  軌跡をモデル化するために用いられてきた [21]。このモデルは、音声だけでなく、独唱における歌声の  $F_0$  軌跡を表現するモデルとしても用いられてきた [17], [22], [23], [24]。歌声の  $F_0$  ダイナミクスを表現するばね質量系は次式のような二階の微分方程式で表される。

$$\alpha \frac{d^2 y(t)}{dt^2} + \beta \frac{dy(t)}{dt} + \gamma y(t) = u(t) \quad (2)$$

ここで、 $u(t)$  は外力、 $y(t)$  は質点の位置を表す。この方程式を  $F_0$  ダイナミクスに適用する場合、 $u(t)$  は目標とする音高軌跡である楽譜の旋律、 $y(t)$  は歌声の  $F_0$  軌跡を表し、入出力の関係は図 1 のようになる。  $\alpha, \beta$  及び  $\gamma$  はこのシステムを特徴づけるパラメータであり、減衰振動や指数減衰など様々な変動を表現する。加古らが提案したモデルは、引き込みの影響を受けた歌声の  $F_0$  を表現することを目的として、この基本的なばね質量系を 1 つの質点と 2 つのばねで構成される結合システム (One-Mass-Two-Spring (OMTS) system) に拡張したモデルである。

本稿は 5 つの節から構成され、次節では歌声データの収録と  $F_0$  軌跡の分析について述べる。ここでは、独唱時と合唱時における歌声の  $F_0$  の違いを比較し、合唱時の  $F_0$  が他者の  $F_0$  に近づくことを示す。3 節では、合唱における  $F_0$  ダイナミクスを表現する OMTS モデルについての述べ、モデルのパラメータ推定方法について議論する。4 節では、モデルの評価実験を行う。最後に、5 節で本稿のまとめと今後の展望を述べる。

表 1  $F_0$  推定の実験条件

Table 1 Signal analysis conditions for  $F_0$  estimation.

サンプリング周波数	16 kHz
量子化ビット数	16 bits
窓関数	ハニング窓
窓幅	64 ms
窓シフト長	10 ms
$F_0$ 平滑化 (移動平均窓長)	30 ms

## 2. 合唱における歌声の $F_0$ 軌跡

### 2.1 収録データ

本稿では、合唱における歌声の  $F_0$  の動的変動成分を歌声の引き込みという観点から議論する。  $F_0$  の引き込みを分析するため、次のような条件の下、斉唱における 6 名の歌声を収録した。歌唱者は、大学の混成合唱団に所属する歌唱訓練歴 3-7 年の男性 (経験者) 3 名と、歌唱訓練を積んでいない男性 (素人) 3 名である。歌唱曲は「ふるさと」とし、その 1 番のバスパートを用い、低残響室において収録した。各歌唱者はそれぞれ、事前に収録した同じパートの歌声 (随伴歌唱) と伴奏をヘッドホンで聴きながら 3 回ずつ歌唱した。伴奏音は、合唱パート練習用の Chorus ONTA Vol.10 の CD に収録された音源を用いた。随伴歌唱は、3 名とは別の経験者が伴奏と同 CD に収録されたプロの音楽家の歌声をヘッドホンで聴きながら歌唱した歌声である。また、伴奏をヘッドホンで聴きながら歌唱した独唱の歌声も収録した。つまり合計で、6 名の各歌唱者の独唱 3 回と随伴歌唱受聴時の歌唱 (合唱) 3 回を収録した。

### 2.2 $F_0$ 軌跡の分析

収録した歌唱データを用いて独唱時と合唱時における  $F_0$  軌跡の違いについて分析を行う。歌声の  $F_0$  軌跡は、河原ら [25] の提案した TANDEM-STRAIGHT を利用して推定した。  $F_0$  推定の条件を表 1 に示す。なお、本稿では  $F_0$  の周波数 [Hz] は、式 (3) を用いて対数スケールの周波数 [cent] に変換する。

$$f [\text{cent}] = 1200 \log_2 \frac{f [\text{Hz}]}{440 \times 2^{\frac{3}{12} - 5}} \quad (3)$$

このように対数スケールに変換することで、半音が 100 cent に対応する。

図 2 は、独唱時及び合唱時における  $F_0$  軌跡と、その相平面表現の例である [19], [26]。図 2 には、楽譜の旋律情報及び随伴歌唱の  $F_0$  軌跡も示してある。  $F_0 - \dot{F}_0$  構成の相平面の図より、独唱時と比較し、合唱時の渦軌跡の中心が随伴歌唱の中心に近づいたことがわかる。これは、合唱歌唱において  $F_0$  の引き込みが生じた典型的な結果であると考えられる。そこで随伴歌唱を受聴することによる  $F_0$  の変化を議論するために、独唱及び合唱における歌声の音高 ( $F_0$ )

表 2  $d$  の平均  $\mu$  と標準偏差  $\sigma$

Table 2 Average  $\mu$  and standard deviation  $\sigma$  of  $d$

	$\mu$ [cent]		$\sigma$ [cent]	
	独唱	合唱	独唱	合唱
素人	25.18	22.81	38.13	37.27
経験者	12.72	7.93	34.93	33.08

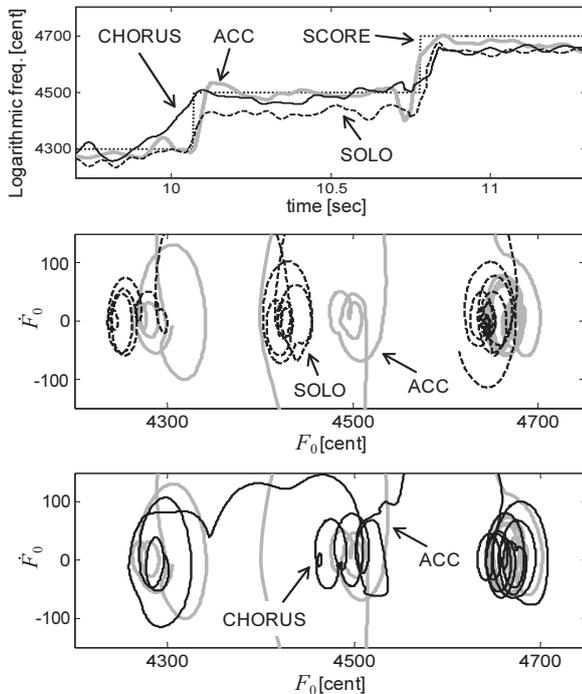


図 2 独唱時及び合唱時における  $F_0$  軌跡, 上段:  $F_0 - \dot{F}_0$  構成の相平面上に描かれる独唱時の  $F_0$  軌跡. 点線: 楽譜の旋律情報 (SCORE), 実線 (グレー): 随伴歌唱音高 (ACC), 破線: 独唱時の  $F_0$  (SOLO), 実線 (黒): 合唱時の  $F_0$  (CHORUS).

Fig. 2 Melodic contours (top) and corresponding phase plane for  $F_0 - \dot{F}_0$  in solo and chorus singing (middle and bottom). Dotted line: original musical score (SCORE); solid gray line: accompanying vocal (ACC); dashed line: solo singing (SOLO); solid black line: chorus singing (CHORUS).

と随伴歌唱音高 ( $V$ ) との差を計算した.

$$d(t) = V(t) - F_0(t) \quad (4)$$

素人と経験者のカテゴリ別に計算した  $d(t)$  の平均と標準偏差を表 2 に示す. 平均と分散について t 検定及び F 検定を行った結果, 独唱時と合唱時の間に 1% の有意水準で有意差があることが確認された. この結果より, 合唱歌唱では随伴歌唱を受聴することによる影響が生じることが明らかとなった. また, 合唱における  $d(t)$  の標準偏差が減少したことから, 統計的に歌唱者の歌声の音高は随伴歌唱に近づくといえる.

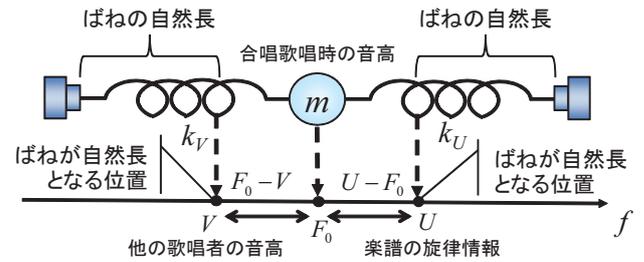


図 3 合唱歌唱の OMTS モデル

Fig. 3 OMTS model of chorus singing

### 3. 合唱における歌声の $F_0$ ダイナミクスのモデル化

#### 3.1 合唱歌唱の OMTS モデル

前節において, 合唱歌唱時の音高が随伴歌唱音高に近づくことが統計的に確認された. そこで本節では,  $F_0$  の動的な特徴について更なる分析を行うため, 加古ら [20] によって提案された 1 つの質点と 2 つのばねから構成されるシステム (One-Mass-Two-Spring (OMTS) モデル) を利用する. OMTS モデルは図 3 に示すような簡単なシステムである. このシステムの運動方程式は以下で与えられる.

$$m \frac{d^2 F_0}{dt^2} = -\lambda \frac{dF_0}{dt} + k_U (U - F_0) + k_V (V - F_0) \quad (5)$$

ここで  $m$  は質点の質量,  $\lambda$  は動摩擦係数, 楽譜の旋律情報  $U$  と随伴歌唱音高  $V$  はそれぞれのばねの自然長の位置と定義される.  $k_U$  と  $k_V$  はそれぞれ楽譜の旋律情報と随伴歌唱音高に対応する振動子に接続するばねのばね定数である. このモデルは, 従来の独唱における歌声の  $F_0$  ダイナミクスを表現するモデル (式 (2)) に, 他者の音高からの影響を表す項として  $k_V (V - F_0)$  を新たに加えたモデルである. OMTS モデルにおける 2 つのばねのばね定数の値は, 楽譜情報と随伴歌唱からの影響の度合いを表す. 例えば,  $k_V$  が大きければ合唱時の歌唱は随伴歌唱からの影響をより受けた歌唱となる. 逆に,  $k_U$  が大きければ随伴歌唱にあまり影響されない歌唱となる. よって, モデルパラメータ  $\Theta = \{m, \lambda, k_U, k_V\}$  は, 合唱における歌声の引き込みを特徴づけるパラメータとなる.

#### 3.2 モデルパラメータ推定方法

本節では, 観測信号からモデルパラメータを推定する方法について述べる. 我々は, モデルパラメータを推定する方法として, 最急降下法を用いた. 最急降下法とは目的関数  $J(\Theta)$  を最小にする重みベクトル (モデルパラメータ  $\Theta$ ) を目的関数の勾配情報を利用して反復的に求める方法である. ここでは, 最小化する目的関数  $J(\Theta)$  として, 式 (6) で与えられる二乗誤差を用いた.

$$J(\Theta) = \frac{1}{N} \sum_{n=1}^N (F_0(t_n) - y(t_n, \Theta))^2 \quad (6)$$

ここで、 $F_0(t)$  は観測した歌声の  $F_0$ 、 $y(t)$  は式 (5) を解くことによって得られる関数である。目的関数を式 (6) とすると、モデルパラメータ  $\Theta$  の更新式は以下ようになる。

$$\Theta^{(\tau+1)} = \Theta^{(\tau)} - \eta \frac{\partial J(\Theta^{(\tau)})}{\partial \Theta} \quad (7)$$

$$\frac{\partial J(\Theta)}{\partial \Theta} = -\frac{2}{N} \sum_{n=1}^N (F_0(t_n) - y(t_n, \Theta)) \frac{\partial y(t_n, \Theta)}{\partial \Theta} \quad (8)$$

ここで、 $\eta$  は各モデルパラメータに対する学習率パラメータである。誤差関数  $J(\Theta)$  の  $\Theta$  に関する導関数は、モデルパラメータセット  $\Theta$  が与えられた下で式 (5) を解くことによって計算することができる。二階微分方程式である式 (5) は、フーリエ級数展開とラプラス変換を用いて以下のように解くことが可能である。

まず、式 (5) を次式のように書き換える。

$$\ddot{y}(t) + \alpha \dot{y}(t) + (\beta_U + \beta_V) y(t) = \beta_U U(t) + \beta_V V(t) \quad (9)$$

ここで  $\alpha, \beta_U, \beta_V$  は、ばね定数を質量  $m$  で割り、正規化した値である。次に、 $U(t)$  と  $V(t)$  をフーリエ級数展開を利用し正弦関数と余弦関数に展開する。

$$U(t) = u_0 + \sum_{k=1}^{\infty} (u_{1k} \cos k\omega_0 t + u_{2k} \sin k\omega_0 t) \quad (10)$$

$$V(t) = v_0 + \sum_{k=1}^{\infty} (v_{1k} \cos k\omega_0 t + v_{2k} \sin k\omega_0 t) \quad (11)$$

これらを利用すると、OMTS モデルの運動方程式 (式 (5)) を変形して得られた式 (9) は、更に異なる式に変形することができる。

$$\begin{aligned} \ddot{y}(t) + \alpha \dot{y}(t) + (\beta_U + \beta_V) y(t) \\ = c_0 + \sum_{k=1}^{\infty} (c_{1k} \cos k\omega_0 t + c_{2k} \sin k\omega_0 t) \end{aligned} \quad (12)$$

ここで  $c_k$  は、フーリエ係数の重み付け和であり、 $c_k = \beta_U u_{.k} + \beta_V v_{.k}$  と定義する。式 (12) の解はラプラス演算子  $s$  を用いて以下で与えられる。

$$y(t) = \mathcal{L}^{-1} \{[(s + \alpha)y(0) + \dot{y}(0) + R(s)] Q(s)\} \quad (13)$$

ここで、 $\mathcal{L}^{-1}$  はラプラス逆変換、 $R(s)$  は式 (12) における右辺をラプラス変換したものを表す。また、 $Q(s)$  は次式で与えられる。

$$Q(s) = \frac{1}{s^2 + \alpha s + (\beta_V + \beta_U)} \quad (14)$$

式 (13) を解くと、最終的に式 (12) の解は以下の関数で与えられる。

$$\begin{aligned} y(t) = & \frac{1}{\xi_1 - \xi_2} \{ (y(0)\xi_1 + \dot{y}(0) - \alpha y(0)) e^{\xi_1 t} \\ & - (y(0)\xi_2 + \dot{y}(0) - \alpha y(0)) e^{\xi_2 t} \} \\ & + \frac{c_0}{\xi_1 \xi_2} + \frac{c_0 e^{\xi_1 t}}{(\xi_1 - \xi_2)\xi_1} + \frac{c_0 e^{\xi_2 t}}{(\xi_2 - \xi_1)\xi_2} \\ & + \sum_{k=1}^{\infty} \left\{ \frac{(\xi_1 c_{1k} + k\omega_0 c_{2k}) e^{\xi_1 t}}{(k^2 \omega_0^2 + \xi_1^2)(\xi_1 - \xi_2)} + \frac{(\xi_2 c_{1k} + k\omega_0 c_{2k}) e^{\xi_2 t}}{(k^2 \omega_0^2 + \xi_2^2)(\xi_2 - \xi_1)} \right. \\ & + \frac{(\xi_1 \xi_2 - k^2 \omega_0^2) c_{1k} + k\omega_0 (\xi_1 + \xi_2) c_{2k}}{(k^2 \omega_0^2 + \xi_1^2)(k^2 \omega_0^2 + \xi_2^2)} \cos k\omega_0 t \\ & \left. - \frac{(k^2 \omega_0^2 - \xi_1 \xi_2) c_{2k} + k\omega_0 (\xi_1 + \xi_2) c_{1k}}{(k^2 \omega_0^2 + \xi_1^2)(k^2 \omega_0^2 + \xi_2^2)} \sin k\omega_0 t \right\} \end{aligned} \quad (15)$$

但し、

$$\xi_{1,2} = \frac{\alpha \pm \sqrt{\alpha^2 - 4(\beta_U + \beta_V)}}{2}$$

このようにフーリエ級数展開とラプラス変換を利用して  $y(t)$  の関数を得ることができるため、 $\Theta$  を式 (7) に従って反復更新することで、与えられた信号  $F_0, U, V$  からモデルパラメータを推定することができる。

## 4. 評価実験

### 4.1 パラメータ推定

OMTS モデルの評価実験として、合唱における歌声の引き込みの違いを特徴付ける能力について評価する。そこでまず、パラメータ推定を行い、素人と経験者のパラメータの違いについて分析する。実験に使用する歌声データ長は、2.1 節において示した収録データの歌い出しから 17 sec. 最急降下法に基づくパラメータの更新回数は 100 回として、モデルパラメータの推定を行った。なお、各モデルパラメータの初期値は  $\alpha^{(0)} = 1, \beta_U^{(0)} = 0.5, \beta_V^{(0)} = 0.5$  とし、学習率パラメータ  $\eta$  は  $\eta_\alpha = 1.0 \times 10^{-5}, \eta_{\beta_U} = 1.0 \times 10^{-6}, \eta_{\beta_V} = 1.0 \times 10^{-6}$  とした。パラメータ推定結果の一例として、推定されたモデルパラメータから生成した  $F_0$  軌跡と実際の歌声の  $F_0$  軌跡を図 4 に示す。

素人と経験者について推定されたモデルパラメータの平均と、ばね定数  $k_U, k_V$  の比  $k_V/(k_U + k_V)$  及び減衰率  $\zeta$  を表 3 に示す。この平均は、各歌唱者カテゴリに分類される 3 名の歌唱者の歌声データ 9 回分によって計算された値である。ばね定数比  $k_V/(k_U + k_V)$  は、楽譜の旋律情報と随伴歌唱音高のどちらからより影響を受けるかの指標となる。以下の式に基づき計算される減衰率  $\zeta$  は、変動の大きさを表し、その値が小さいほどシステムへの入力の変化に対して大きな変動となる。

$$\zeta = \frac{\alpha}{2\sqrt{\beta_V + \beta_U}} \quad (16)$$

パラメータ推定の結果より、経験者より素人の方が随伴歌唱音高に対応する振動子に接続するばねのばね定数が大き

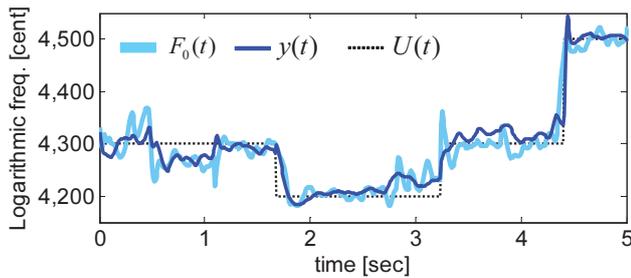


図 4 実際の歌声の  $F_0$  軌跡の信号  $F_0(t)$  と推定されたモデルパラメータから生成した信号  $y(t)$

Fig. 4 An example of the actual sang signal  $F_0(t)$  and  $F_0$  contour  $y(t)$  derived from the estimated model parameters.

表 3 モデルパラメータ推定結果と RMSE

Table 3 Estimated model parameters and RMSE

	$\alpha$	$\beta_V$	$\beta_U$	$\zeta$	$\frac{k_V}{k_U + k_V}$	RMSE [cent]
素人	1.53	1.27	1.21	0.49	0.51	44.4
経験者	1.41	1.29	1.37	0.43	0.48	39.4

Note: RMSE の値は、観測信号と推定されたモデルパラメータから生成した信号から計算された値である。減衰率  $\zeta$  は変動の大きさを表し、ばね定数比  $k_V/(k_U + k_V)$  は  $\beta_U, \beta_V$  から計算され、随伴歌唱を重視する割合を表す。

いことがわかる。この素人と経験者に対して推定されたパラメータの違いの有意差を確認したわけではないが、経験者より素人の方が随伴歌唱に影響されやすいということが示唆される。また、経験者の減衰率  $\zeta$  の方が素人より小さい値となっている。この減衰率  $\zeta$  の違いから経験者の  $F_0$  ダイナミクスの方が変動が大きく、素人より表情豊かな歌唱であると推論できよう。

#### 4.2 歌唱者識別実験

合唱における引き込みの影響は各歌唱者によって異なると考えられるため、歌唱者毎に異なるモデルパラメータが推定されると仮定できる。そこで、観測した歌声の  $F_0$  系列と OMTS モデルから生成した  $F_0$  系列の二乗平均平方根誤差 (RMSE) を評価尺度として歌唱者識別実験を行うことにより、歌唱の個人性を特徴付ける能力を評価する。本実験では、6名の歌声データを収録したため、6分類の識別実験となる。各歌唱者に対して3回分の歌声データがあるため、2回分を学習データとしてモデルパラメータ  $\Theta$  を推定する。残り1回分を評価データとして RMSE の値を計算し、式 (17) により歌唱者識別を行う。

$$\hat{s} = \arg \min_s E(F_0, y_s) \quad (17)$$

$$E(F_0, y_s) = \sqrt{\frac{1}{N} \sum_{n=1}^N (F_0(t_n) - y_s(t_n, \Theta_s))^2} \quad (18)$$

ここで  $s$  は歌唱者、 $\Theta_s$  は歌唱者  $s$  のモデルパラメータ、 $N$

表 4 RMSE の計算結果

Table 4 Calculation result of RMSE

	素-A	素-B	素-C	経-A	経-B	経-C
素-A	<b>69.26</b>	71.01	69.97	70.90	69.69	69.47
素-B	69.88	<b>68.53</b>	68.58	74.15	72.07	71.46
素-C	51.55	<b>51.08</b>	51.62	56.05	53.61	52.78
経-A	50.46	56.22	54.05	<b>48.55</b>	48.94	49.15
経-B	53.20	57.96	55.64	53.55	52.30	<b>52.27</b>
経-C	46.99	51.66	49.04	48.05	46.39	<b>46.36</b>

Note: “素”と“経”はそれぞれ素人と経験者を表す。横のラベルはモデルの名称、縦のラベルは評価データの各歌唱者の名称を表す。識別結果の正解は、対角の値が最も小さくなったときである

は評価データ長、 $F_0$  は評価データ、 $y_s$  はモデルパラメータ  $\Theta$  から生成した信号を表す。

結果を表 4 に示す。最小の RMSE の値を示すモデルをボールド体で強調した。横のラベルはモデルの名称、縦のラベルは評価データの各歌唱者の名称を表し、対角の値が最も小さくなったときが正解となる。表 4 より、6名中4名の歌唱者が自身のモデルにおいて最も小さい RMSE の値をとっていることが分かる。つまり、6名のうち4名の歌唱者を識別することができた。素-C と経-B は、歌唱者自身のモデルにおいて最も小さい値とならなかったが、自身のモデルにおいて比較的小さい RMSE の値となっている。また、RMSE の値から素人と経験者を正しく識別できていることが確認される。これらの結果から、合唱における  $F_0$  の引き込みの個人性を OMTS モデルにより特徴付けることができることが示唆された。

#### 5. まとめと今後の展開

本稿では、合唱における歌声の  $F_0$  ダイナミクスを引き込みという観点から分析し、ばね質量系に基づいてモデル化する手法について議論した。我々は観測される信号からモデルパラメータを推定する手法として、最急降下法を用いた。OMTS モデルの評価として、歌唱者識別実験を行い、6名のうち4名を識別することができた。また歌唱者識別実験から、素人と経験者を正確に識別することが確認された。これらのことから、 $F_0$  の引き込みをパラメトリックに表現することのできる手法として、OMTS モデルの有効性が示唆された。

しかし、モデルをより広いアプリケーションに適用するためには、今後更なる研究が必要である。まず第一に、歌唱者同士が互いの歌声を聴きながら歌唱するような相互作用の生じる状況下においてモデルの有効性を確かめる必要がある。なぜなら、複数の歌唱者が同時に歌唱する場合、互いの歌声が影響を及ぼすため相互引き込みが生じるが、本稿では事前に収録した随伴歌唱を聴きながら歌唱するという限定的な条件下でのみ  $F_0$  の引き込みに関して議論しているためである。また、モデルの一般的な効果について

評価するためには、より大人数の歌声を収録し、実験を行う必要があるだろう。

**謝辞** 本研究は、挑戦的萌芽研究 (23650088) の支援により行われた。

## 参考文献

- [1] T. D. Rossing, J. Sundberg, and S. Ternström.: Acoustic comparison of voice use in solo and choir singing, *Acoustical Society of America*, Vol. 79, pp. 1975–1981, (1986).
- [2] Johan Sundberg.: *The Science of the Singing Voice*. the Northern Illinois University Press, (1987).
- [3] S. Ternström and J. Sundberg.: Intonation precision of choir singing, *Acoustical Society of America*, Vol. 84, pp. 59–69, (1988).
- [4] S. Ternström.: Preferred self-to-other ratios in choir singing, *Acoustical Society of America*, Vol. 105 pp. 3563–3574, (1999).
- [5] S. Ternström and J. Sundberg.: Self-to-other ratios measured in an opera chorus in performance, *Acoustical Society of America*, Vol. 118, pp. 3903–3911, (2005).
- [6] J. Buck and E. Buck.: Mechanism of Rhythmic Synchronous Flashing of Fireflies, *Science*, Vol. 159, pp. 1319–1327, (1968.03).
- [7] T. J. Walker.: Acoustic Synchrony: Two Mechanisms in the Snowy Tree Cricket, *Science*, Vol. 166, pp. 891–894, (1969.11).
- [8] Z. Neda, E. Ravasz, Y. Brechet, T. Vicsek, and A.L. Barabasi.: The sound of many hands clapping, *Nature*, Vol.403, pp. 849–850, (2000).
- [9] T. Yamamoto and Y. Miyake.: Analysis of interaction in musical communication and its modeling, *In Proc. IEEE International Conference on Systems, Man, and Cybernetics (SMC 2000)*, Vol. 2, pp. 763–768, (2000).
- [10] 小野寺進, 徳田功, 山本知仁: 共同演奏におけるリズム・周波数引き込みの解析, 日本音響学会講演論文集, 2-1-5, pp. 1017–1018, (2011.03).
- [11] Y. Kobayashi and Y. Miyake.: New ensemble system based on mutual entrainment, *In Proc. IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN 2003)*, pp. 235–240, (2003).
- [12] 野田雄也, 徳田功, 榊原健一: 合唱における基本周波数の同期現象に関する基礎研究, 日本音響学会講演論文集, 2-5-13, pp. 913–916, (2008.03).
- [13] 堀内靖雄, 坂本圭司, 市川薫: 合奏時の人間の演奏制御の分析・推定, 情報処理学会論文誌, Vol. 45, No. 3, pp. 690–697, (2004).
- [14] C. E. Seashore.: A musical ornament, the vibrato, *Psychology of Music*, McGraw-Hill Book Company, pp. 33–52, (1938).
- [15] H. B. Rothman and A. A. Arroyo.: Acoustic variability in vibrato and its perceptual significance, *In Proc. Psychology of Music*, Vol. 1, No. 2, pp. 123–141, (1987).
- [16] G. de Krom and G. Bloothoof.: Timing and accuracy of fundamental frequency changes in singing, *ICPhS*, pp. 206–209, (1995).
- [17] T. Saitou, M. Goto, M. Unoki, and M. Akagi.: Speech-To-Singing Synthesis: Converting Speaking Voices to Singing Voices by Controlling Acoustic Features Unique to Singing Voices, *WASSPA*, pp. 215–218, (2007).
- [18] T. Saitou and M. Goto.: Acoustic and Perceptual Effects of Vocal training in Amateur Male Singing, *International Conference on Spoken Language Processing (INTER-SPEECH 2009)*, pp. 832–835, (2009).
- [19] Y. Ohishi, M. Goto, K. Itou, and K. Takeda.: A Stochastic Representation of the Dynamics of Sung Melody, *In Proc. International Conference on Music Information Retrieval (ISMIR 2007)*, pp. 371–372, (2007.09).
- [20] 加古達也, 大石康智, 亀岡弘和, 長野秀尚, 柏野邦夫, 武田一哉: 合唱における歌声の基本周波数軌跡の分析, 日本音響学会講演論文集, 1-Q-44, pp. 403–404, (2011.03).
- [21] H. Fujisaki.: A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour, *Vocal Physiology: Voice Production, Mechanisms and Functions*, (O.Fujimura, ed.), Raven Press, pp. 347–355, (1988).
- [22] N. Minematsu, B. Matsuoka, and K. Hirose.: Prosodic Modeling of Nagauta Singing and Its Evaluation, *Speech-Prosody 2004*, pp. 487–490, (2004.09).
- [23] H. Mori, W. Odagiri, and H. Hirose.: F0 dynamics in singing: Evidence from the data of a baritone singer, *IE-ICE Trans. Inf. and Syst.*, Vol.E87-D, No.5, pp. 1086–1092, (2004).
- [24] Y. Ohishi, H. Kameoka, D. Mochihashi, and K. Kashino.: A Stochastic Model of Singing Voice F0 Contours for Characterizing Expressive Dynamic Components, *International Conference on Spoken Language Processing (INTER-SPEECH 2012)*, (2012.09).
- [25] H. Kawahara, M. Morise, T. Takahashi, R. Nishimura, T. Irino, and H. Banno.: Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation, *ICASSP 2008*, pp. 3933–3936, (2008.04).
- [26] T. Kako, Y. Ohishi, H. Kameoka, K. Kashino, and K. Takeda.: Automatic Identification for Singing Style Based on Sung Melodic Contour Characterized in Phase Plane, *International Conference on Music Information Retrieval (ISMIR 2009)*, pp. 393–397, (2009.10).