

# 合唱における歌声の引き込みを利用した歌声 $F_0$ 制御の検討

川岸基成<sup>1</sup> 川渕将太<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:** 合唱歌唱時に生じる歌声の“引き込み”を利用して歌唱の基本周波数 ( $F_0$ ) を目標とする音高に誘導制御することを試みる。我々はこれまで、他者の歌声を受聴しながら歌唱したときの歌声への影響を歌声の引き込みという観点から分析し、1つの質点と2つのばねで構成されるばね質量系を用いて、歌声の  $F_0$  動特性をモデル化した。本稿では、合唱歌唱の  $F_0$  動特性を表現するばね質量モデルを利用して歌唱の  $F_0$  を制御する手法を提案する。本手法では、合唱歌唱のばね質量モデルに基づき各歌唱者の引き込みの特徴を分析し、引き込みを考慮した誘導音を合成、受聴させることにより、目標音高に近づくように  $F_0$  を制御する。評価実験では、目標音高を受聴しながら歌唱した歌声と誘導音高を受聴しながら歌唱した歌声を目標音高との RMSE で評価し、誘導音高を受聴することで被験者 8 人中 5 人の RMSE が減少するという結果を得た。

## 1. はじめに

歌声は声帯が振動し、音響管に対応する声道を通して生成される音響信号であるため、歌声を生成する発声器官は音を出す道具である楽器とみなすことができる。そのため、歌声は最も簡単に誰もが楽しむことのできる楽器である。その反面、ピアノやギターのように出る音の高さや音色が決まっている楽器に対し、歌声は歌唱者自身で調音するため歌唱力の乏しい人にとって楽譜通りの音程での歌唱や表情豊かな歌唱をすることは困難である。

歌唱能力向上のためにこれまで様々な研究が行われてきた。歌唱能力を向上させるには、持続的なトレーニングにより能力を向上させるアプローチと技術的な支援により能力を向上させるアプローチが考えられる。後者のアプローチとして、視覚フィードバックによる支援システム [1, 2], 触覚フィードバックを利用した盲ろう者のための歌唱訓練 [3] などが挙げられる。このように様々な側面から歌唱力向上のための様々なアプローチや歌唱支援システムが検討されており、歌唱能力向上のための支援技術は重要な研究課題である。

そこで本稿では、歌唱力向上を目的として、歌声の“引き込み”という観点からアプローチする。引き込み [4-6] は、他者の行動に適応するような変化と解釈でき、工学の分野では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、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。歌声においても、合唱のように複数の歌唱者で歌唱を行う場合には、他者の歌声から影響を受け、引き込み現象が生じると考えられる [7]。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 [8-10]。

我々はこれまで、合唱における歌唱を特徴付ける数理モデルの構築を目指し、その第一段階として歌声の基本周波数に焦点を当て、合唱における  $F_0$  の動的な特徴 ( $F_0$  ダイナミクス) をばね質量系を用いてモデル化した [11]。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 [12]。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 [13-16]。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。

$$\alpha \frac{d^2 y(t)}{dt^2} + \beta \frac{dy(t)}{dt} + \gamma y(t) = u(t) \quad (1)$$

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、 $u(t)$  は目標とする旋律である楽譜の旋律、 $y(t)$  は歌声の  $F_0$  軌跡を表し、 $\alpha, \beta$  及び  $\gamma$  は減衰振動や指数減衰など様々な変動を表現するパラメータである。我々が提案したモデルは、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り、引き込みの影響を受けた歌声の  $F_0$  ダイナミクスを特徴付ける。そこで、合唱歌唱の OMTS モデルを利用して各歌唱者の引き込みの特徴を分析し、引き込みを考慮した誘導音を合成、受聴させることにより、目標

<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

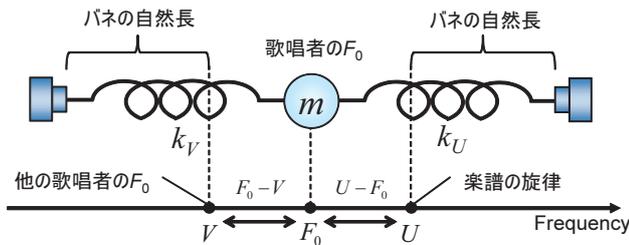


図 1 合唱歌唱の OTMS モデル

Fig. 1 OTMS model of chorus singing

音高に近づくように  $F_0$  を誘導すること手法を提案する。

本稿は 5 つの節から構成され、次節では合唱における  $F_0$  ダイナミクスを表現する OTMS モデルについて述べる。3 節では、合唱歌唱の OTMS モデルを用いて、歌唱者が受聴する音高を変化させたときの影響を分析し引き込みの影響を明確化する。4 節では、歌声の音高の引き込みを利用して  $F_0$  を制御する手法と評価実験について述べる。最後に、5 節で本稿のまとめと今後の課題を述べる。

## 2. OTMS モデルによる合唱歌唱の $F_0$ ダイナミクスの表現

合唱歌唱における歌声の  $F_0$  ダイナミクスを表現する OTMS (One-Mass Two-Spring) モデル [11] は 1 つの質点と 2 つのばねから構成される簡単なシステム (図 1) である。このシステムの運動方程式は以下で与えられる。

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

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。OTMS 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ず、楽譜に忠実な歌唱となる。また、質点の質量  $m$  と動摩擦係数  $\lambda$  は物理的性質として外力に対する質点の変動の大きさを表すため、 $F_0$  ダイナミクスに適用する場合  $F_0$  の変動の大きさを表現するパラメータとなる。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  $F_0$  ダイナミクスを特徴づけるパラメータとなる。

## 3. 引き込み現象の基礎的分析

他者と同時に歌唱を行う合唱では、他者の音高から影響を受け引き込み現象が生じると考えられる。特に、歌唱経験の少ない素人歌唱者は、他者の音高の影響を受け、強く引き込まれる。一方、歌唱訓練を積んだ経験者は、楽譜の

表 1 経験者の歌唱訓練歴

Table 1 The number of vocal training of experienced subjects.

| 歌唱者   | 歌唱訓練歴 |
|-------|-------|
| 経験者 1 | 4 年   |
| 経験者 2 | 2 年   |
| 経験者 3 | 3 年   |
| 経験者 4 | 1 年   |

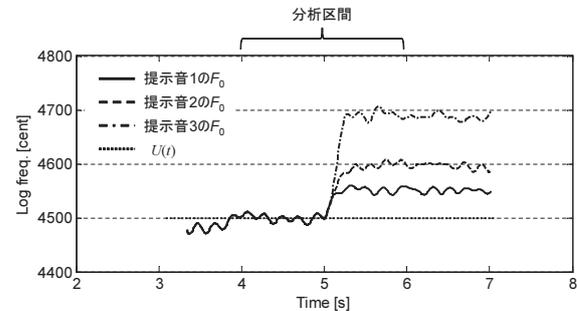


図 2 提示音の  $F_0$  軌跡

Fig. 2  $F_0$  contours of presented sounds

旋律に相当する安定した目標を持つため、素人と比較して他者の音高の影響を受けにくいという仮説が立てられる。そこで本節では、歌唱者が受聴する音高を変化させたときの影響を OTMS モデルに基づいて分析し、引き込みの影響について議論する。

### 3.1 収録データ

歌唱力の異なる歌声を収録するため、大学の混声合唱団に所属する男性歌唱者 (経験者) 4 名と歌唱訓練を受けたことのない男性学生 (素人) 4 名に対して収録実験を行った。各経験者の歌唱訓練歴は表 1 に示す。歌唱者は提示された音 (以後、提示音と呼ぶ) を受聴しながら、220 Hz (4500 cent) の一定の旋律で「ラ」の音を 4 秒間発声する。このロングトーンを各提示音に対し 2 回ずつ収録した。歌唱者に受聴させる提示音は、開始 2 秒は 4500 cent、その後異なる音高 (4550, 4600, 4700 cent) に変化する歌声を生成し使用した。それぞれ提示音 1, 提示音 2, 提示音 3 とする。なお、本稿では  $F_0$  の周波数 [Hz] は、式 (3) を用いて変換した対数スケールの周波数 [cent] を用いる。対数スケール上では、半音が 100 cent に対応する。

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

提示音は絶対音感のある歌唱者に依頼し、4500, 4600, 4700 cent のロングトーンを収録し、音高が変化する音源を作成した。4550 cent については、Griffin-Lim の反復 STFT 法 [17] を用いて音高変換し合成した音源を利用した。提示音の  $F_0$  軌跡を図 2 に示す。

表 2  $F_0$  推定の実験条件

Table 2 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 bit |
| 窓シフト長           | 10 ms  |
| $F_0$ 平滑化移動平均窓長 | 50 ms  |

表 3 モデルパラメータ推定結果と RMSE

Table 3 Estimated model parameters and RMSE.

|     | 提示音 | $\alpha$ | $\beta_U$ | $\beta_V$ | $\frac{k_V}{k_U + k_V}$ | RMSE<br>[cent] |
|-----|-----|----------|-----------|-----------|-------------------------|----------------|
|     |     |          |           |           |                         |                |
| 経験者 | 1   | 0.47     | 0.18      | 0.01      | -0.048                  | 9.6            |
|     | 2   | 0.41     | 0.34      | 0.06      | 0.012                   | 6.5            |
|     | 3   | 0.59     | 2.47      | 0.1       | -0.005                  | 17.3           |
| 素人  | 1   | 0.42     | 0.12      | 0.1       | 0.346                   | 15.8           |
|     | 2   | 0.43     | 1.16      | 0.79      | 0.152                   | 11.4           |
|     | 3   | 0.52     | 1.16      | -0.21     | 0.137                   | 14.3           |

**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。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。

### 3.2 OMTS モデルによる分析

収録した歌声から OMTS モデルパラメータを推定し、受聴音の音高変化による影響を分析する。歌声の  $F_0$  は、河原らの提案した TANDEM-STRAIGHT[18] を利用して推定した。  $F_0$  推定の条件を表 2 に示す。その他のパラメータはデフォルトの値で行った。 OMTS モデルを 3.1 節の収録データに適用する場合、提示音の  $F_0$  が随伴歌唱音高  $V$ , 4500 cent の一定音高が楽譜の旋律  $U$  にそれぞれ対応する。推定するモデルパラメータは  $\{\lambda, k_U, k_V\}$  を質量  $m$  で割り、正規化した値  $\{\alpha, \beta_U, \beta_V\}$  となる。 OMTS モデルパラメータの推定は最急降下法を用いて行った [11]。最急降下法におけるパラメータの更新回数は 100 回、各モデルパラメータの初期値は  $\alpha^{(0)} = 0.4, \beta_U^{(0)} = 0.05, \beta_V^{(0)} = 0.05$ 、学習率パラメータ  $\eta$  は  $\eta_\alpha = 1.0 \times 10^{-5}, \eta_{\beta_U} = 1.0 \times 10^{-6}, \eta_{\beta_V} = 1.0 \times 10^{-6}$  として、モデルパラメータを推定した。モデルパラメータの初期値は先行研究 [16] に基づいて、学習率パラメータの値は予備実験に基づいて決定した。分析には音高変化の前後 1 秒間の歌声を使用した。

経験者と素人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  $k_V/(k_U + k_V)$  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 4 名の歌唱者の発声データ 8 回分から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。パラメータ推定結果におけるばね定数比を比較すると、素人より経験者の方が 0 に近い値をとっていることがわかる。つまり、提示音に対応する振動子に接続するばねのばね定数が非常に小さいということを意味する。この結果から、経験者は提示音の音高に影響されず、楽譜に従って歌唱するということが推

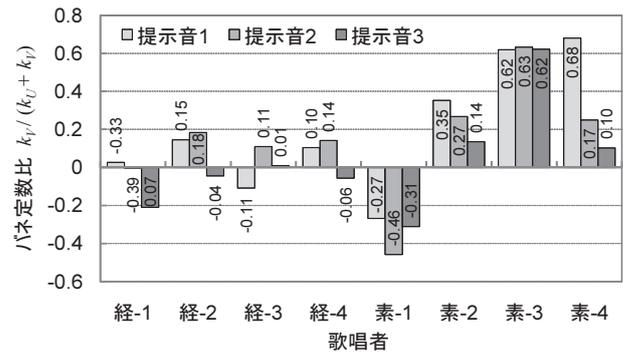


図 3 各歌唱者のばね定数比

Fig. 3 Calculated spring constant ratio using estimated model parameters  $\beta_U, \beta_V$ .

論できる。一方、素人は経験者より大きい値をとっているため、提示音の音高に影響されやすいということが推論できる。

図 3 に各歌唱者について計算されたばね定数比を示す。図 3 から経験者の方が 0 に近い値をとり、素人の方が大きい値をとっていることが読み取れる。次に、推定されたモデルパラメータと実際の  $F_0$  軌跡との関係について議論するため、各提示音における典型的な例を図 4 に示す。図 4(a) は、経験者 3 が提示音 3 を受聴しながら歌唱した歌声の  $F_0$  と推定されたモデルパラメータから生成した信号である。この  $F_0$  軌跡から、経験者 3 は提示音高が変化しても安定した旋律で歌唱していることがわかる。そのため、ばね定数比も 0 に近い値が推定されている。図 4(b) は、素人 4 が提示音 2 を受聴しながら歌唱した歌声の  $F_0$  と推定されたモデルパラメータから生成した信号である。素人 4 の場合、提示音高が変化すると、その影響を受け提示音高に近づく  $F_0$  が観測される。このとき提示音高に対応するばねが硬くなるため、ばね定数比も大きくなる。図 4(c) は、素人 1 が提示音 1 を受聴しながら歌唱した歌声の  $F_0$  と推定されたモデルパラメータから生成した信号である。素人 1 の  $F_0$  は、提示音高が変化すると、提示音の音高変化とは逆方向に変化している。素人 1 は提示音の音高変化に対して影響されないように意識して発声した結果、このような  $F_0$  軌跡になり、負の値のばね定数比が推定されたと考えられる。この結果から受聴音高が、歌唱者が目標とする音高である楽譜情報と大きく異なる場合、受聴音高変化とは逆方向に引き込みの影響が生じる可能性がある。

## 4. 音高の引き込みを利用した $F_0$ 制御

### 4.1 $F_0$ 制御の枠組み

3 節では、OMTS モデルを用いて他者の歌声を受聴することにより生じる引き込みについて分析し、他者の歌声からの影響の大きさを表現するモデルパラメータが歌唱者によって異なることがわかった。そこで、合唱歌唱の OMTS

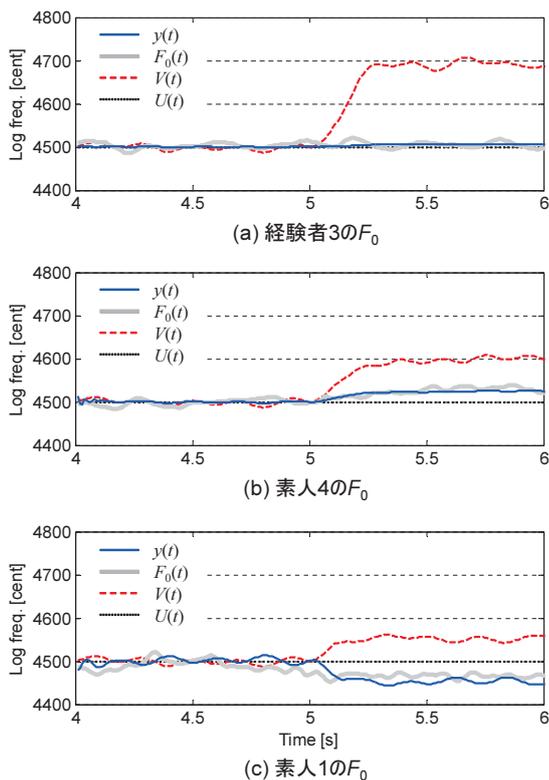


図 4 実際の歌声の  $F_0$  軌跡の信号  $F_0(t)$  と推定されたモデルパラメータから生成した信号  $y(t)$

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

モデルの応用として、合唱における引き込みを利用して歌唱の  $F_0$  を誘導制御する手法を提案する。本手法では、合唱歌唱の OMTS モデルに基づき各歌唱者の引き込みの特徴を分析し、引き込みを考慮した誘導音を合成、受聴させることにより、目標音高に近づくように  $F_0$  を制御することを試みる。引き込みを利用した  $F_0$  制御の枠組みを図 5 に示す。まず、随伴歌唱  $V(t)$  を受聴しながら歌唱した歌声  $F_0$  からモデルパラメータ  $\theta = \{\alpha, \beta_U, \beta_V\}$  を推定する。その後、モデルパラメータ  $\theta$  と、歌唱者に歌唱させたい音高系列  $T(t)$ 、楽譜情報  $U(t)$  から歌唱者に受聴させる随伴歌唱である誘導音高系列  $V_g(t)$  を生成する。生成した  $V_g(t)$  に従って歌声を合成し、歌唱者に受聴させることで目標音高  $T(t)$  に近づくように  $F_0$  を制御する。

#### 4.2 誘導音の生成方法

各歌唱者に対して推定されたモデルパラメータ  $\theta = \{\alpha, \beta_U, \beta_V\}$  と、歌唱者が目標とする音高  $T(t)$ 、楽譜情報  $U(t)$  から随伴歌唱音高  $V_g(t)$  を推定する方法について述べる。ここでは、後退差分変換を用いて  $V_g(t)$  を求める。後退差分変換は、時間微分演算子  $s$  を  $z$  領域における後退差分演算子  $s \simeq (1 - z^{-1})/\tau_s$  に置き換える変換である ( $\tau_s$  は離散時間表現におけるサンプリング周期とする)。まず、式 (2) の各係数を質点の質量  $m$  で割って書き換え、

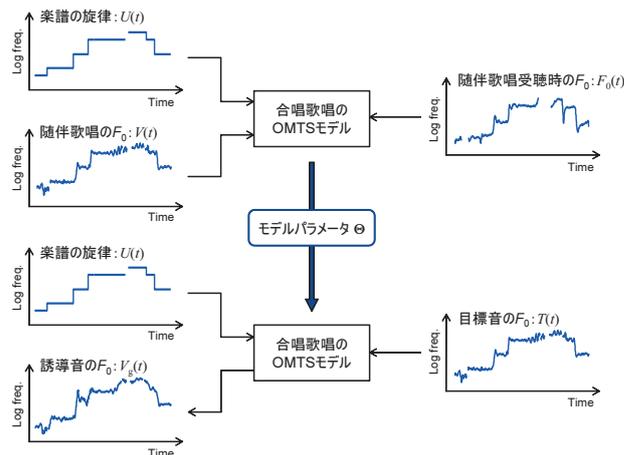


図 5  $F_0$  制御の枠組み

Fig. 5 Framework of  $F_0$  control system.

$$\ddot{y}(t) + \alpha \dot{y}(t) + (\beta_U + \beta_V) y(t) = \beta_U U(t) + \beta_V V(t) \quad (4)$$

後退差分変換を適用すると、

$$a_2 y(t_{n-2}) + a_1 y(t_{n-1}) + a_0 y(t_n) = \beta_U U(t_n) + \beta_V V(t_n) \quad (5)$$

を得る。ただし、

$$a_2 = \varphi^2, a_1 = -\varphi(2\varphi + \alpha), a_0 = \varphi^2 + \alpha\varphi + \beta_U + \beta_V \quad (6)$$

である。ここで、 $\varphi = 1/\tau_s$  である。歌唱者に受聴させる誘導音高  $V_g(t)$  を求めるために、 $y(t)$  を  $T(t)$ 、 $V(t)$  を  $V_g(t)$  に置き換え、 $V_g(t)$  について解くと、

$$V_g(t_n) = \frac{1}{\beta_V} \{a_0 T(t_n) + a_1 T(t_{n-1}) + a_2 T(t_{n-2}) - \beta_U U(t_n)\} \quad (7)$$

を得る。よって、楽譜情報  $U(t)$ 、目標音高  $T(t)$  と各歌唱者について推定した OMTS モデルパラメータ  $\theta$  から誘導音高  $V_g(t)$  を生成できる。

歌声合成は、Griffin-Lim の反復 STFT 法 [17] を用いて音高変換し合成した [19]。Griffin-Lim の反復 STFT 法は振幅スペクトルに位相を付加することによって音響信号を復元する手法である。元音源の  $F_0$  軌跡と変換したい  $F_0$  軌跡の比に応じて駆動音源スペクトルを線形伸縮し、スペクトル包絡を掛け合わせ、任意の基本周波数に変換した振幅スペクトルに Griffin-Lim の反復 STFT 法を用いて位相を付加することで目的の  $F_0$  軌跡を持つ音響信号を合成した。

#### 4.3 収録データ

歌唱訓練を積んでいない男性学生 8 名の歌声を以下の条件の下で収録した。収録場所は低残響室、歌唱曲は「ふるさと」を選曲し、その 1 番の歌詞のバスパートを収録した。歌唱者が目標とする歌唱には、合唱パート練習用の Chorus ONTA Vol.10 の CD に収録されているプロの声楽家の歌声 (バスパート) を用いた。また、伴奏音として同 CD に収録されている音源を用いた。以下の手順に従って各歌唱

者が各条件 3 回ずつ歌唱した歌声を収録した。

1. 伴奏と目標歌唱を受聴しながら歌唱 (パラメータ推定用)
2. 伴奏と誘導歌唱を受聴しながら歌唱 (評価用)
3. 伴奏と目標歌唱を受聴しながら歌唱 (評価用)

なお、歌唱者に受聴させる歌声と伴奏はヘッドホンを通じて流した。

#### 4.4 評価実験

##### 4.4.1 評価方法

提案手法の有効性を評価するため、目標音高を受聴しながら歌唱した歌声の  $F_0$  と誘導音高を受聴しながら歌唱した歌声の  $F_0$  を目標音高 ( $T$ ) との RMSE (式 (8)) で評価する。

$$E(F_0, T) = \sqrt{\frac{1}{N} \sum_{n=1}^N (F_0(t_n) - T(t_n))^2} \quad (8)$$

この値が小さいほど、目標音高との二乗誤差が小さい。つまり、目標音高に近いことを意味する。

##### 4.4.2 実験結果

$F_0$  の推定及び OMTS モデルパラメータの推定に関する実験条件は 3 節と同様である。実験には 4.3 節において示した収録データの歌い出しから 17 秒 (8 小節分) を使用した。

各歌唱者に対して計算した RMSE を図 6 に示す。図 6 から 8 名中 5 名は誘導音を受聴した方が小さい RMSE となっていることがわかる。しかし、残り 3 名では誘導音を受聴することにより目標音受聴時より大きな値となってしまった。目標音受聴と比較して誘導音受聴時の RMSE が減少した歌唱者、増加した歌唱者、それぞれの  $F_0$  軌跡の例を図 7 に示す。図 7 (a)(b) は RMSE が減少した歌唱者 2 の  $F_0$  軌跡、図 7 (c)(d) は RMSE が増加した歌唱者 1 の  $F_0$  軌跡である。歌唱者 2 の場合、目標音受聴時の 11~15 秒区間の音高変化において目標音高から大きく外れた  $F_0$  となっているのに対し、誘導音受聴時の同区間では目標音高に近づいた軌跡が観測される。歌唱者 1 の場合、目標音受聴時の 15~18 秒区間において目標音高に近い  $F_0$  が観測されるが、誘導音受聴時の同区間では目標音高から低く外れた  $F_0$  が観測される。歌唱者 1 の 15~18 秒区間では、誘導音の  $F_0$  軌跡が目標音高から大きく外れる。このため、歌唱者が誘導音に影響されないように低く歌唱することを意識することで、目標音高より低い  $F_0$  となったと考えられる。歌唱者 1 は 3 節の素人 1 と同一人物であり、3 節の分析において受聴音高の変化とは逆の方向に影響を受けるという結果が得られている。よって、この歌唱者は目標音高から大きく外れる  $F_0$  の歌声を受聴した場合、随伴歌唱音高に対応するばねのばね定数が負となると推測できる。このため、誘導音受聴時における RMSE が大きくなったと考えられる。

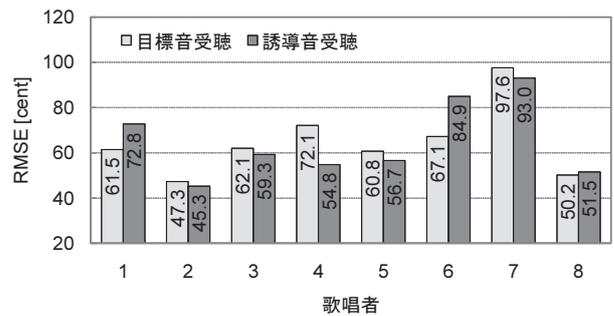


図 6 RMSE の計算結果

Fig. 6 Calculation results of RMSE

## 5. まとめ

本稿では、合唱歌唱の OMTS モデルに基づき歌唱者の引き込みの特徴を分析し、歌唱音高の引き込みを利用して歌声の  $F_0$  を目標の音高に近づけるように制御することを試みた。我々は最初に、歌唱者が受聴する音高が変化したときの影響について“引き込み”という観点から分析し、歌唱訓練を受けたことのある歌唱力の高い歌唱者は受聴音の音高が変化してもあまり影響を受けず、歌唱力の低い歌唱者は影響を受けやすいことがわかった。そこで、各歌唱者の引き込みの特徴を分析し、他者の歌声を受聴することによる影響を考慮した誘導音を合成、受聴させることにより目標音高に近づくように  $F_0$  を制御する手法を提案した。評価実験では、目標音高を受聴しながら歌唱した歌声の  $F_0$  と誘導音高を受聴しながら歌唱した歌声の  $F_0$  を目標音の  $F_0$  との RMSE で評価し、誘導音高を受聴することで被験者 8 名中 5 名の RMSE が減少するという結果を得た。

しかし、歌唱者によっては誘導音高を受聴することで目標音高から大きく外れ、RMSE が増加してしまった。その原因として、各歌唱者の引き込みの特徴を考慮して生成した誘導音高が目標音高から大きく外れた場合、妨害音として働くことが挙げられた。この問題を解決するためには、リアルタイムで  $F_0$  を分析し、誘導音高を更新することが必要であると考えられる。

**謝辞** 本研究は、科学技術振興事業団の戦略的基礎研究推進事業 CREST により行われた。

## 参考文献

- [1] T. Nakano, M. Goto and Y. Hiraga.: MiruSinger: A Singing Skill Visualization Interface Using Real-Time Feedback and Music CD Recordings as Referential Data, *ISM workshop 2007*, pp. 75–76, (2007).
- [2] D. M. Howard and G. F. Welch.: Visual displays for the assesment of vocal pitch matching development, *Applied Acoustics*, Vol. 39, pp. 235–252, (1993).
- [3] 坂尻正次, 三好茂樹, 中邑賢龍, 福島智, 伊福部達.: 触覚フィードバックを用いた音声ピッチ制御方式による盲ろう者の歌唱訓練, ライフサポート, Vol. 22, pp. 138–145, (2010).

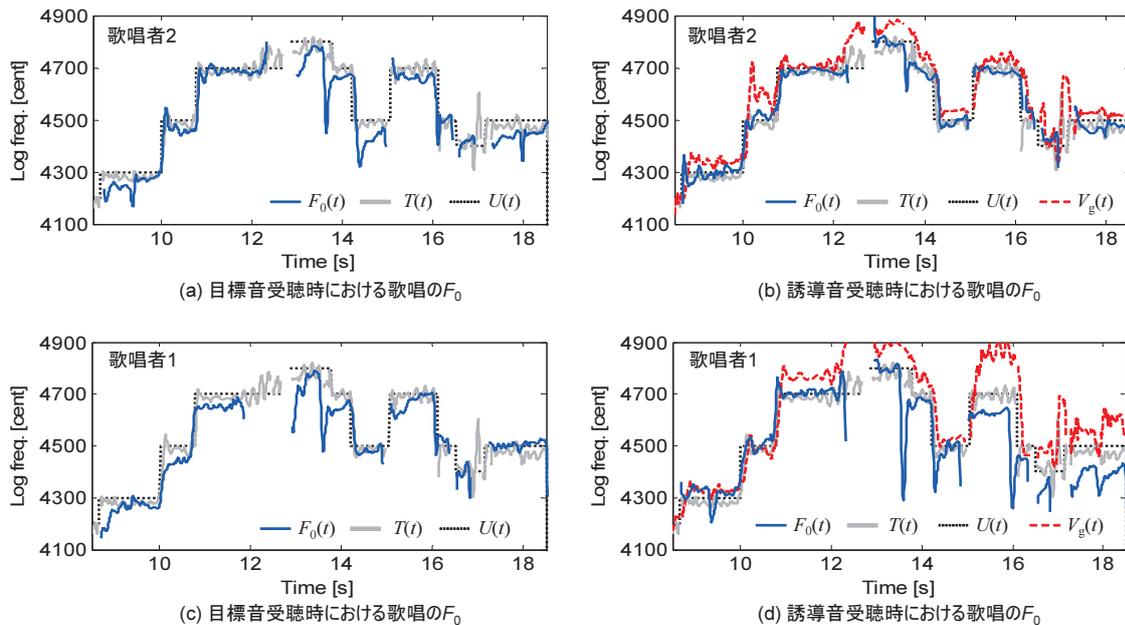


図 7 目標音及び誘導音受聴時の  $F_0$  軌跡. 点線: 楽譜の旋律 ( $U(t)$ ), 実線 (グレー): 目標音の  $F_0(T(t))$ , 実線 (青): 歌声の  $F_0(F_0(t))$ , 破線: 誘導音の  $F_0(V_g(t))$ .

Fig. 7 An example of  $F_0$  contours of the actually sung signals when the subjects sang while listening to target sound and guiding sound, respectively. Dotted line: original musical score  $U(t)$ ; solid gray line:  $F_0$  contour of target sound  $T(t)$ ; solid blue line:  $F_0$  contour of actually sung signal in each conditions  $F_0(t)$ ; dashed line:  $F_0$  contour of guiding sound  $V_g(t)$ .

- [4] J. Buck and E. Buck.: Mechanism of Rhythmic Synchronous Flashing of Fireflies, *Science*, Vol. 159, pp. 1319–1327, (1968.03).
- [5] T. J. Walker.: Acoustic Synchrony: Two Mechanisms in the Snowy Tree Cricket, *Science*, Vol. 166, pp. 891–894, (1969.11).
- [6] 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).
- [7] 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).
- [8] T. Yamamoto and Y. Miyake.: Analysis of interaction in musical communication and its modeling, *Proc. SMC 2000*, Vol. 2, pp. 763–768, (2000).
- [9] 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).
- [10] Y. Kobayashi and Y. Miyake.: New ensemble system based on mutual entrainment, *Proc. RO-MAN 2003*, pp. 235–240, (2003).
- [11] M. Kawagishi, S. Kawabuchi, C. Miyajima, N. Kitaoka, and K. Takeda.: Analysis and modeling of entrainment in chorus singing, *Proc. ICASSP 2013*, pp. 7795–7799, (2013.05).
- [12] 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).
- [13] 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, *Proc. WASSPA*, pp. 215–218, (2007).
- [14] N. Minematsu, B. Matsuoka, and K. Hirose.: Prosodic Modeling of Nagauta Singing and Its Evaluation, *Proc. SpeechProsody 2004*, pp. 487–490, (2004.09).
- [15] H. Mori, W. Odagiri, and H. Hirose.: F0 dynamics in singing: Evidence from the data of a baritone singer, *IEICE Trans. Inf. and Syst.*, Vol. E87-D, No. 5, pp. 1086–1092, (2004).
- [16] Y. Ohishi, H. Kameoka, D. Mochihashi, and K. Kashino.: A Stochastic Model of Singing Voice F0 Contours for Characterizing Expressive Dynamic Components, *Proc. INTERSPEECH 2012*, (2012.09).
- [17] Daniel W. Griffin and Jae S. Limarticle.: Signal estimation from modified short-time Fourier Transform, *IEEE Trans. Acoustics, Speech, and Signal*, Vol. 32, No. 2, pp. 236–243, (1984).
- [18] 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, *Proc. ICASSP 2008*, pp. 3933–3936, (2008.04).
- [19] 大石康智, 亀岡弘和, 持橋大地, 永野秀尚, 柏野邦夫.: Vocal Dynamics Controller: 歌声の F0 動特性をノート単位で編集し, 合成できるインタフェース, 情報処理学会研究報告, Vol. 2010-MUS-86, No. 9, (2010.07).