

# 歌声波形からのビブラートパラメータ推定と性能評価

宮崎 嵩大<sup>1,a)</sup> 森勢 将雅<sup>2,3</sup>

受付日 2019年12月25日, 採録日 2020年9月10日

**概要:** 本論文では, 歌唱表現の1つであるビブラートに着目し, 波形からビブラートを構成するパラメータ(以下ではビブラートパラメータとする)を推定する手法の改良と性能評価を実施した結果について述べる. ビブラートは歌声の基本周波数( $F_0$ )を準周期的に振動させる歌唱表現であり, ビブラートパラメータは $F_0$ 軌跡における振動の周期と深さが該当する. 加えて, 本論文では, 歌唱区間においてビブラートをかけ始める時刻も推定対象とする. 歌声波形からビブラートパラメータを検出する方法はすでに検討されているため, ここでは既存の方法をベースに精度を改善する方法を提案する. 歌唱データベースを用い真値が明確となるよう人工的に付与したビブラートを用いた計算機シミュレーションを実施し, 既存の方法の性能を明らかにするとともに, 提案法によりさらに性能が改善できることを示す.

**キーワード:** 歌声情報処理, 歌声分析, 基本周波数, ビブラート

## Estimation of Vibrato Parameters from Singing Waveform and its Evaluation

TAKAHIRO MIYAZAKI<sup>1,a)</sup> MASANORI MORISE<sup>2,3</sup>

Received: December 25, 2019, Accepted: September 10, 2020

**Abstract:** In this paper, we describe a vibrato parameter estimation algorithm and show the performance evaluation. Vibrato is defined as a musical effect consisting of a quasi-periodic fluctuation of fundamental frequency ( $F_0$ ). Vibrato parameters consist of the amplitude and frequency of the  $F_0$  contour. In addition, we define the boundary in the duration of vibrato as the ratio between the duration of one note and that of vibrato section. Since there is an algorithm for estimating the vibrato parameters, we propose an algorithm for improving the performance. An evaluation was carried out to verify the effectiveness of the proposed algorithm, and the result showed that the proposed algorithm could work as expected.

**Keywords:** singing information processing, singing analysis, fundamental frequency, vibrato

### 1. はじめに

歌うことは多くの人によって楽しられている娯楽の1つであり, その代表例であるカラオケは世界各地で親しまれている. インターネットの発達にともない, ニコニコ動画<sup>\*1</sup>やYouTube<sup>\*2</sup>などの動画共有サイトでは, 歌唱音声も共有して楽しむことが文化として定着しつつある. スマホ

アプリにおいても, 歌声を公開・共有できるサービスとして nana<sup>\*3</sup>や Changba<sup>\*4</sup>がリリースされており, 特に後者は数千万単位のアクティブユーザを獲得しているとされる. このように, 誰でも歌うことや, それを共有することを楽しむことができる文化が発展し定着している状況にある.

歌声に関する研究は古くからなされているが, 歌声を用いたアプリケーションやサービスの発展にともない研究事例 [1] も増加している. 歌声は話し声とは異なり特定の音高をリズムに合わせて変化させる特色があるが, それ以外

<sup>1</sup> 山梨大学  
University of Yamanashi, Kofu, Yamanashi 400-8511, Japan

<sup>2</sup> 明治大学  
Meiji University, Nakano, Tokyo 164-8525, Japan

<sup>3</sup> JST さきがけ  
JST PRESTO, Chiyoda, Tokyo 102-0076, Japan

a) g19tk020@yamanashi.ac.jp

\*1 <https://www.nicovideo.jp/> (最終検索日: 2020年7月13日)

\*2 <https://www.youtube.com/> (最終検索日: 2020年7月13日)

\*3 <https://nana-music.com/> (最終検索日: 2020年7月13日)

\*4 <https://changba.com/> (最終検索日: 2020年7月13日)

にも歌唱ホルマント [2] という音色の変化や、 $F_0$  の動的変動成分 [3] が会話音声とは異なるという特徴を有する。歌唱ホルマントとは、Sundberg により示された、男性のオペラ歌唱の 3kHz 付近において観測される顕著なホルマントピークのことである。 $F_0$  軌跡の特徴には、オーバシュートや微細変動成分などの発声器官の物理的な制約に起因する成分 [4] のほかに、ビブラートやポルタメントなどの歌唱者による意識的な表現 [5], [6] も存在する。代表的な歌唱表現であるビブラートは、 $F_0$  軌跡を準周期的に振動させる歌唱法として定義される。ビブラートは、ポピュラー音楽で統計的性質の分析 [7]、歌唱音声のビブラート部分の類似度を測るためのビブラート特徴量の解析 [8]、ビブラートの情報を用いた歌手の識別 [9] に利用されている。加えて、ビブラートは、歌唱力や [10]、歌声の不快さ [11]、歌声の知覚に影響することが指摘されている [12]。

歌声合成において、ビブラートは VOCALOID [13] や話し声を歌声に変換する歌声合成システム [12]、ユーザ歌唱の歌い方を真似る歌声合成パラメータを自動推定するシステム [14], [15] などで用いられている。以上のことから、ビブラートは歌声分析・合成における主要なパラメータとして幅広く用いられており、重要なものの 1 つであるといえる。ビブラートパラメータの推定法はすでに提案されているものの、ビブラートの区間を検出するためのアルゴリズムである [16]。この手法は歌手ごとの統計的性質の研究にも利用されていることから、区間検出という目的に対し十分な性能が達成できているといえる [17]。

本研究では、従来法とは異なり、ビブラートパラメータを精密に推定するという目標を設定し、まず従来法における精度を低下させる要因を示す。その後、精度を向上させるための方法を提案する。ビブラートを構成するパラメータはすでいくつか提案されているが [12], [17]、本研究では組合せ数を発散させないよう 3 つの特徴に限定して性能評価を実施する。具体的には、ビブラートの速さ、深さ、およびビブラートは歌声の全体ではなくノートの途中から付与することもあるため、ビブラート判定した開始時刻を推定の対象とする。詳細なビブラートパラメータの推定が可能になれば、歌声のより精密な分析が可能になり、歌声合成におけるビブラート制御などにも利用可能になることが期待される。

## 2. ビブラートに関する関連研究

本章では、ビブラートに関する関連研究とビブラートを構成するパラメータについて説明する。その後、ビブラート区間検出手法に関する従来研究について紹介し、本研究の位置づけを述べる。

### 2.1 ビブラートパラメータの定義

ビブラートとは、音を伸ばした歌声においてその音高を

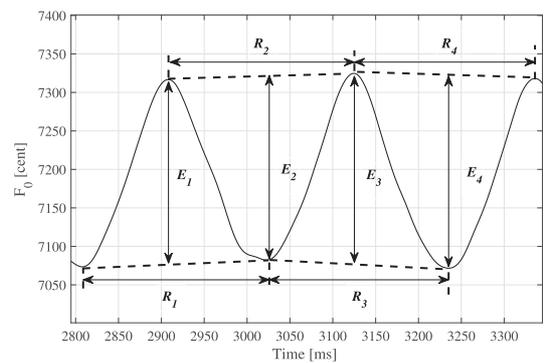


図 1 ビブラートの  $F_0$  軌跡と速さ・深さの算出法

Fig. 1 Example of an  $F_0$  contour and the definition of vibrato rate and extent.

保ちつつ高さなどを細かく振動させる表現であるため、ビブラートの速さを表す vibrato rate と深さを表す vibrato extent が主要なビブラートパラメータとなる。これらは、ビブラート区間の  $F_0$  軌跡より、図 1 に示す  $R_n$  [s],  $E_n$  [cent] を抽出し、式 (1), (2) によって算出される [16]。

$$\frac{1}{\text{rate}} = \frac{1}{N} \sum_{n=1}^N R_n, \quad (1)$$

$$\text{extent} = \frac{1}{2N} \sum_{n=1}^N E_n, \quad (2)$$

ここで、 $N$  はビブラート区間の  $F_0$  軌跡から抽出された極大・極小点の総数をそれぞれ示している。深さの算出では、式 (3) に示すように、中央ハ音の周波数  $f_c$  ( $= 440 \times 2^{\frac{3}{12} - 1} = 261.62... \text{ Hz}$ ) の cent 値を 4800 cent とし、周波数  $f_{\text{Hz}}$  を cent 単位の  $f_{\text{cent}}$  に変換後に求めている。

$$f_{\text{cent}} = 1200 \log_2 \left( \frac{f_{\text{Hz}}}{f_c} \right) + 4800. \quad (3)$$

本研究では、上記 2 つの他にビブラートの開始時刻も推定すべきビブラートパラメータとして扱う。人間が発するビブラートでは、速さと深さは時間変動すること [18], [19] も指摘されているが、本実験では推定対象とはしない。本実験では、時間変動をしないビブラートを人工的に付与させた歌唱音声を用いる。ビブラートの付与の詳細は 4 章で後述する。

### 2.2 ビブラート区間検出手法

先行研究として、中野らによってビブラート区間検出手法が提案されている [16]。従来法では、 $F_0$  の時系列  $F_0(t)$  [cent] の 1 次差分  $\Delta F_0(t)$  (10 ms ごと) に短時間フーリエ変換 (short-time Fourier transform: STFT) を行うことでビブラートを検出する。32 点 (320 ms) のハニング窓を用いた STFT で得られる振幅スペクトル  $X(f, t)$  をビブラート区間検出に用いる。その振幅スペクトルは、ビブラートの速さに対応する周波数において鋭いピークを有する。この推定法では、時刻  $t$  におけるビブラート速度の周波数帯

域のパワー  $\Psi_v(t)$  とピークの鋭さ  $S_v(t)$  を式 (4), (5) で定義し用いる.

$$\Psi_v(t) = \int_{F_L}^{F_H} \hat{X}(f, t) df, \quad (4)$$

$$S_v(t) = \int_{F_L}^{F_H} \left| \frac{\partial \hat{X}(f, t)}{\partial f} \right| df, \quad (5)$$

ここで,  $F_L$  と  $F_H$  はそれぞれ速さの周波数の下限, 上限を示している. 式中にある  $\hat{X}(f, t)$  は, 式 (6) に示すように, 時刻  $t$  ごとに 0 Hz からナイキスト周波数までのパワーで正規化したものとなる.

$$\hat{X}(f, t) = \frac{X(f, t)}{\int_0^{f_s/2} X(f, t) df}, \quad (6)$$

ここで,  $f_s$  は, サンプリング周波数を表す. これらを用いて, 時刻  $t$  におけるビブラートらしさ  $P_v(t)$  を式 (7) のように定義する.

$$P_v(t) = S_v(t)\Psi_v(t). \quad (7)$$

$P_v(t)$  が大きく, 速さと深さが制限内 (速さが 5–8 Hz, 深さが 30–150 cent) であり, かつ  $F_0(t)$  がその平均音高と 5 回以上交差する区間をビブラートとして検出している.

### 2.3 従来法によるビブラートパラメータの推定手順

本論文では, 文献 [16] で提案された方法を従来法として以下のように用いる. 具体的には, 2.2 節で記述した方法により, 入力された歌声の  $F_0$  軌跡からビブラート区間を検出する. ビブラート区間の検出条件に速さと深さの制限範囲があるため, 320 ms でのビブラート区間判定に速さと深さの推定が必要である. 320 ms でのビブラート区間判定で求めた速さと深さは, そこにのみ用い, その後に 1 音符の歌声から検出された全体のビブラート区間のみから求めた速さと深さを本研究では推定結果とする. 4 章で後述する従来法で本評価に用いた全テストデータを評価したところ, テストデータの約 7 割で, 判定したビブラート終端とその 1 音符の歌声に対する音符の終端が一致した. そのため, 1 つの音符に対応する歌声の場合, ビブラートの終端は音符の終端と一致する傾向があるといえる. しかしビブラートの終了時刻を音符の終端時刻としてしまうと, 1 音符の歌声でビブラート区間が途切れ途切れに複数検出された場合, 検出条件を満たせずビブラートが検出されなかった区間の極大・極小点を含み, ビブラートの速さや深さの誤差拡大の原因となることが考えられる. そのため, 本研究のビブラートの終了時刻は従来法提案法の双方において, 検出された区間の終端に従うものとする. それにともない, 1 音符の歌声でビブラート区間が途切れ途切れに複数検出された場合, 検出されたそれらのみを無条件で 1 つの区間と見なし, ビブラート区間の開始時刻を検出する.

その後, 途切れたビブラート区間すべての  $E_n, R_n$  から式 (1), (2) を用いて速さと深さを推定する. 具体的にはビブラート区間の  $F_0$  軌跡より前後 1 サンプル点の値を用いて極大・極小点となる時刻を抽出し, 図 1 に示す  $E_n, R_n$  を算出した. この処理においては, 入力するビブラートに対する柔軟性を考慮して極大・極小点の検出に速さや深さに関する制限は設けない.

### 2.4 本研究の位置付け

ビブラートのパラメータ推定では, 音声波形からの  $F_0$  推定と同様に  $F_0$  軌跡の基本周波数であるビブラートの速さを求めることに加え, ビブラートの深さや開始時刻に対応するパラメータも推定する必要がある. このように, 出力とする系列の特徴が異なるため, 音声の  $F_0$  推定精度を向上させる方法をそのまま利用することは困難である. たとえば  $F_0$  推定性能向上に用いられる低域通過フィルタを通すことで雑音を取り除くことが可能であるが, フィルタを畳み込むことによりビブラート軌跡の振幅が変化し, ビブラートの深さの推定精度に影響する. 加えて, 2.2 節で説明した平均音高と 5 回以上交差する区間が, フィルタの過渡応答により広がるため, ビブラートの開始時間の推定にも悪影響を及ぼすことが想定される. 音声の  $F_0$  の精度向上を目指した方法は提案されているものの, ビブラートパラメータ推定に関しては, 速さ以外のパラメータ推定結果に悪影響を及ぼさない独自のアルゴリズムを検討する必要がある.

2.2 節で紹介したビブラート区間検出手法では速さと深さの範囲を制限していたが, その制限範囲を超えるビブラートも存在している [17]. また, 歌声合成で作成した歌唱音声を扱う場合では, 人間の発声可能な範囲を超えたビブラートを用いることも可能である. 本研究では人間の歌声の分析を目指すため, 文献 [17] で示される例に基づき, ビブラートの速さと深さが 3.5–8 Hz と 20–400 cent 以内をビブラートの検出範囲とする. 本論文では, 2.3 節で述べた方法を従来法と定義する. 以下では, 従来法の精度が不十分となる要因について示し, それを解決するための方法を提案する. その後, 計算機シミュレーションにより, 推定法のパラメータ調整を行う.

### 3. ビブラートパラメータ推定精度の向上手法

本研究の目的は, ビブラートパラメータの推定精度の検証と, 性能向上を目指した改善法の提案にある. 本章では, 提案する下記の 2 つの改善法を従来法に適用させることでビブラートパラメータの推定精度の向上を図る. ビブラートの速さと深さの誤差は誤差割合  $\epsilon_1$  として式 (8), 開始時刻の誤差  $\epsilon_2$  は式 (9) を用いて算出する.

$$\epsilon_1 = \frac{|x - \hat{x}|}{x}, \quad (8)$$

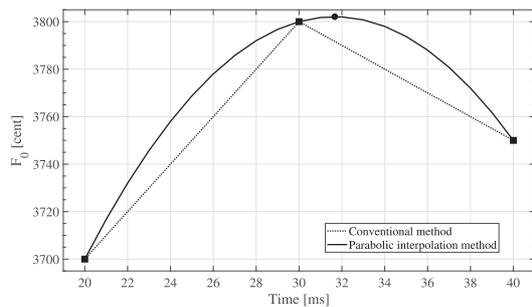


図 2 放物線補間による極大点の推定例

Fig. 2 Example of maximum point estimation by parabolic interpolation.

$$\epsilon_2 = x - \hat{x}, \tag{9}$$

ここで  $x$  は、真値、 $\hat{x}$  は推定値を表す。

### 3.1 放物線補間による推定精度の改善

従来法は、 $F_0$  を 10 ms のシフト幅で推定し用いていた。ビブラートの有無が検出できればよい場合、あるいはビブラートの速度が遅い場合は単純に前後 1 サンプル点の値から判断した極大・極小点を取り出すだけでよいが、速度が速い場合では誤差が拡大する原因となる。改善法としてフレームシフト幅を狭める方法が考えられるが、 $F_0$  推定の段階でフレームシフト幅を狭めた場合、人間の歌声が持つ揺らぎの影響により局所的な凹凸が生じ誤差が拡大する原因となる可能性がある。低域通過フィルタなどでこの影響を除去することは可能であるが、2.4 節で説明したとおり、ビブラート開始時刻の推定誤差の拡大につながる。これは速さ、深さの誤差拡大も意味する。

本研究では、シフト幅は変えず、極大・極小点を中心とし、その前後 1 サンプルを利用した 3 点を対象に二次関数を用いた放物線補間を実施し、微分係数が 0 となる時刻を極大・極小点の時刻として検出することで、開始時刻の推定に影響しないよう、速さと深さの精度改善を狙う。3 点を利用した放物線補間による  $F_0$  軌跡の極大点の推定例を、図 2 に示す。この図は、縦軸が  $F_0$ 、横軸が時間を示しており、実線は図中における 3 つの四角マークから放物線補間したもの、破線は放物線補間前を示し、丸マークは放物線補間により推定された極大点を示す。放物線補間前の極大点である 30 ms に位置する四角マークと放物線補間後の極大点である丸マークを比較すると、 $F_0$  と時間双方において異なる。

### 3.2 ビブラート開始時刻の補正

ビブラート区間検出手法では、320 ms の区間を用いてビブラートの有無を判定するため、切り出された 320 ms の区間に対し局所的にビブラートが含まれることで、ビブラート区間として検出されることが想定できる。窓関数の振幅が 1 となる時刻を中心と見なした本実装では、本来のビブ

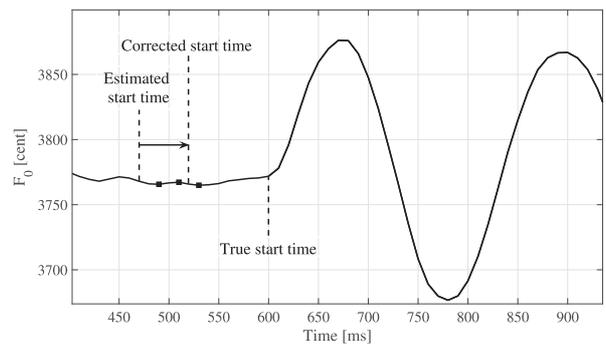


図 3 補正前後のビブラート開始時刻を示した  $F_0$  軌跡の例（四角マークは、補正前の開始時刻から真値の開始時刻までのビブラート区間ではないところに存在する極大・極小点を示す）

Fig. 3 Example of  $F_0$  trajectory showing the vibrato start time before and after correction (The square marks indicate the maximum and minimum points that exist in the non-vibrato section from the start time before correction to the start time of the true value).

ラート開始時刻はビブラートが判定区間の半分の区間で含まれる。これによりビブラートが  $\frac{1}{3}$  程度含む場合でも周波数の検出ができ、ビブラートと判定される可能性がある。主にこの問題は、ビブラートの開始時刻に対し、本来の開始時刻よりも手前を推定結果とする誤差が入ることを意味する。

開始時刻を本来の開始時刻よりも手前に誤推定することは、通常発声で生じる揺らぎ成分の極大・極小点が余分に混入することによる誤差の拡大につながる。その結果、検出において制限範囲を外れることや、制限を外れないまでも誤差の拡大の原因となる。そのためビブラート区間から通常歌唱の領域が取り除かれることで、各ビブラートパラメータの推定精度の向上が期待できる。

本論文では、人工的にビブラートを付与した多数の歌声に対し開始時刻の計算機シミュレーションを行い、中央値として算出したずれを補正することでこの問題に対応する。このビブラート付与には、高品質音声分析合成システム WORLD [20] (D4C edition [21]) を用い、ビブラートがかかっていない歌声を分析し、人工的に生成した  $F_0$  軌跡を付与した後に波形を生成することで実施する。 $F_0$  軌跡は cent の単位で扱っているため、 $F_0$  軌跡の付与は加算することで実現させた。ビブラートの深さと速さの条件の詳細は、4.2 節で説明する。ビブラートの開始時刻では、大幅にずれたデータも観測されることから、平均値ではなく中央値を用いて推定結果を補正することとする。開始時刻の補正を行った  $F_0$  軌跡の例を、図 3 に示す。この図は、縦軸が  $F_0$ 、横軸が時間を示しており、図中の Estimated start time は補正前に推定された開始時刻、Corrected start time は補正した後の開始時刻、True start time は開始時刻の真値を示す。このように従来法どおりに開始時刻を推定し、その時刻から補正する。図中にある四角マークは、補正前

の開始時刻から真値の開始時刻までのビブラート区間ではないところに存在する極大・極小点を示す。この図では、補正を行うことでそれらの極大・極小点が3個中2個省かれ、極大・極小点を用いて算出されるビブラートの速さと深さの誤差が減少することが見込まれる。開始時刻に対して終了時刻は2.3節で説明したように音符の終端である傾向があることから通常歌唱の領域を含むことは少なく、補正によるビブラート区間から通常歌唱の領域が取り除かれる効果が低いと考え、補正を行わない。

#### 4. 計算機シミュレーション

人間の歌声を対象に推定精度が検証できる評価が理想的であるが、人間の歌声におけるビブラートパラメータの真値は定義することができない。よって、3.2節と同様に、ビブラートパラメータに基づく人工的な  $F_0$  軌跡を付与した合成音声を用いる。人工的なビブラート軌跡における真値は明確なため、精密な誤差評価が可能となる。本章では、提案法の有用性を検証することを目的に実施した計算機シミュレーションの計画について述べる。それにともない、使用した評価指標やデータ、各方法の閾値設定についても説明する。

##### 4.1 計算機シミュレーションの目的

本シミュレーションの目的は、ビブラートパラメータの推定精度について、従来法と提案法を比較し提案法の有用性を検証することである。実験で必要となる歌唱データベースには右田らが作成した歌唱データベース [17] を、 $F_0$  推定には Harvest [22] を用いた。2.1節で説明したように、時間変化パラメータを含まないビブラートのみを用いて実験する。

##### 4.2 計算機シミュレーションに用いる歌唱音声の生成方法

本実験では、歌唱データベースに収録されている男性2名と女性2名がビブラートをかけないように発した歌声の中から、音高が高い、普通、低いの3通りを使用した。男性2名を男性1, 2, 女性2名を女性1, 2とする。男性1の音高は D3, F#3, D4, 男性2は E3, A#3, E4, 女性1は B3, F4, B4, 女性2は C3, F#3, C4である。今回の評価では、 $F_0$  が推定できれば音色は無関係のため、歌唱内容が/a/の歌声のみを対象とした。音高3種類、歌手4名からなる計12個の歌唱音声を、以下の表1に示す条

表1 実験に用いるビブラートパラメータの条件

Table 1 Vibrato parameter conditions used in the evaluation.

パラメータ	値
ビブラートの速さ [Hz]	3.5, 4.4, 5.3, 6.2, 7.1, 8.0
ビブラートの深さ [cent]	20, 96, 172, 248, 324, 400
ビブラートの割合 [%]	40, 50, 60, 70, 80, 90

件によって、ビブラートとして正弦波を  $F_0$  に付与したそれぞれの歌唱音声を評価に用いた。ビブラートの付与は、3.2節と同様に WORLD を用いて実施した。ビブラートが付与された  $F_0$  軌跡の例を図4に示す。この図の縦軸は  $F_0$ 、横軸は時間を表す。ビブラートの割合は、図4中の Singing section 内に占める Vibrato section の割合で算出される。ビブラートの速さ、深さ、割合の上限下限は、右田らが作成した歌唱データベースの分析結果 [17] を参考に設定した。これらの上限下限から、ビブラートの速さと深さ、割合は5等分した各6パターンを使用した。歌手ごとのオープンテストを実施するため、これらの計2,592データから3名分の1,944データを訓練データとして用い、1名分の648データをテストデータとした。評価に用いる4つのデータグループが作成されるため、男性1をテストデータとしてその他3名を訓練データとしたデータグループを M1 とする。同様に男性2を M2, 女性1を F1, 女性2を F2 とする。

本計算機シミュレーションでは、ビブラートを含まない歌声をビブラートと誤推定するエラーも評価する必要がある。ここでは、4名がビブラートをかけないように発した、13音階で歌唱内容が/a/, /i/, /u/, /e/, /o/の5つである計260データの歌声からデータグループのテストデータにおける歌手と一致する計65データをテストデータとして用い、残り3名分の195データを訓練データとして用いた。

##### 4.3 評価の指標

音声の評価指標には、 $F_0$  推定に用いられる Gross Pitch Error (GPE) [23], Gross Error [24] や Voicing Decision Error (VDE) [25] がある。たとえば GPE は真値に対し誤差が1ms以内であるか否かで評価する指標であるが、このような指標は、音声と比較して周波数が低いビブラートに対する許容値が狭すぎるため、適切であるとはいえない。Gross Error も真値に対し  $\pm 20\%$  以内であることを条件している。特に、音声の  $F_0$  推定ではハーフピッチエラーのように大きな誤差をとまうことがあるため、局所的にハー

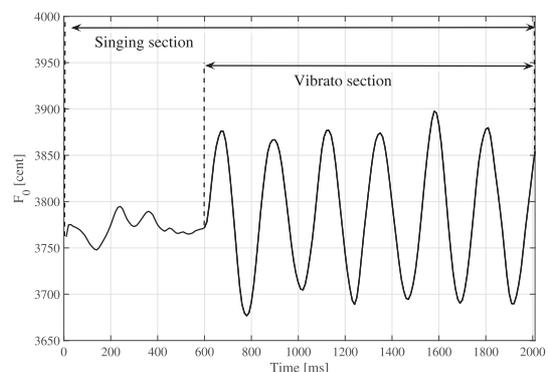


図4 ビブラートが付与された  $F_0$  軌跡の例

Fig. 4 Example of an  $F_0$  contour including a vibrato.

フピッチエラーを含むが他のフレームで誤差が生じない手法と、ハーフピッチエラーがない一方全フレームに小さい誤差が蓄積する方法を単純な誤差の平均で議論することは困難である。従来の指標は、このような音声の  $F_0$  推定固有の問題への対処法としては適切であるが、ビブラートパラメータ推定では誤差の生じ方も異なる点でそのまま利用すべきではない。

加えて、VDEは有声無声の推定ミスを計測する指標であるが、本論文ではビブラートの開始時刻のみ検出できればよいため、これらの指標での評価は適切とはいえない。したがって、本計算機シミュレーションでは、以下に示す8つの評価指標を用いる。

● Allowable number

Allowable number は、全評価データに対し、各ビブラートパラメータについて設定された許容範囲に含まれるデータの数である。許容範囲は、真値に対して誤差が後述する許容値により決定する範囲のことを指す。ビブラートの速さについては Allowable rate number, 深さは Allowable extent number, ビブラートの開始時刻は Allowable boundary number とする。

● Fine error

Fine error は、各推定結果の許容範囲内にあるデータのみから算出した誤差を表す。速さと深さについては、真値からのズレをパーセンテージで表すこととし、速さについては Fine rate error, 深さは Fine extent error とする。また、ビブラートの開始時刻は Fine boundary error とする。ここでのビブラートの開始時刻の誤差は、式 (9) で算出された誤差の絶対値を用いる。

● False error

False error は、全データ数とビブラート判定誤りが生じたデータ数との比率として定義される。ビブラートを非ビブラートと誤判定したものについて、False positive error とする。それに対して非ビブラートをビブラートと誤判定したものについて、False negative error とする。False negative error では、他の指標の分析に用いたビブラートデータでなく、歌唱データベースに含まれるビブラートを含まない実歌唱データを対象に評価する。

Allowable number では数値が高いほど、それ以外の指標は 0 に近いほど精度が高いことを意味する。Allowable number において許容範囲に含まれなかったデータ数については、推定結果が大幅にずれていたものだけでなく、ビブラートを非ビブラートと誤判定する False positive error の影響も含まれる。

4.4 比較に用いるビブラート区間検出手法と閾値設定

ビブラート区間判定条件は表 2 のように統一し、表 3

表 2 ビブラート判定に用いる条件

Table 2 Conditions used for the vibrato identification.

判定条件	条件
$P_v(t)$	データグループごと
ビブラートの速さ [Hz]	3.5–8
ビブラートの深さ [cent]	20–400
平均音高との交差数	5 回以上

表 3 実験に用いるビブラートパラメータ推定法

Table 3 Vibrato parameter estimators used for the evaluation.

ラベル	実装した改善法
A	なし (従来法)
B	ビブラートの放物線補間
C	ビブラート開始時刻の補正
D	ビブラートの放物線補間 ビブラート開始時刻の補正

のように従来法と 2 つの改善法、および両改善法を組み合わせた計 4 通りの手法を用いて比較した。従来法の判定条件をそのまま使用した場合、テストデータの制限範囲外のもの分析できないため、フェアな評価のため従来法においてもビブラートの速さと深さを 3.5–8 Hz, 20–400 cent に設定した。それにともない、 $P_v(t)$  を求める際に扱う  $F_L$  と  $F_H$  も同様に 3.5, 8 Hz に統一して変更した。

4.4.1 ビブラート判定における閾値の策定

ここでは、文献 [16] の判定条件に設定した手法から、ビブラート判定や評価に用いる閾値の策定について述べる。ここで用いる手法における判定条件の速さと深さの制限範囲は、4.4 節で説明した評価に用いる 4 手法で統一されている独自に設定した制限範囲とは異なる。

2.2 節で説明したとおり、ビブラートらしさ  $P_v(t)$  が大きいという条件がビブラート判定条件の 1 つにあるため、 $P_v(t)$  が大きいと判断する閾値を設定する必要がある。 $P_v(t)$  の閾値を決定する際、速さと深さが制限範囲内にあるビブラートの  $P_v(t)$  を文献 [16] の方法で分析し、 $P_v(t)$  の累積分布関数を用いた。使用データは、訓練データからビブラート、非ビブラートのフレームを無作為に 1,000 フレームずつ抽出し、それぞれ分析を行った。非ビブラートのフレームから抽出された  $P_v(t)$  を用いた累積分布関数  $F_n(p)$  は、計算結果の  $P_v(t)$  の値が  $p$  以下だった際、非ビブラートを正しく判定する確率に対応する。ビブラートのフレームから抽出された  $P_v(t)$  を用いた累積分布関数  $F_v(p)$  は、計算結果の  $P_v(t)$  の値が  $p$  以下だった際、ビブラートが存在する確率に対応する。よって  $1 - F_v(p)$  は、計算結果の  $P_v(t)$  の値が  $p$  以下だった際ビブラートが存在しない確率を表し、ビブラートを非ビブラートと判定する確率となる。このことから  $1 - F_v(p)$  は、計算結果の  $P_v(t)$  の値が  $p$  以上だった際、ビブラートを正しく判定している確率を表す。これらを用いて、式 (10) で算出された  $F_r(p)$  は、

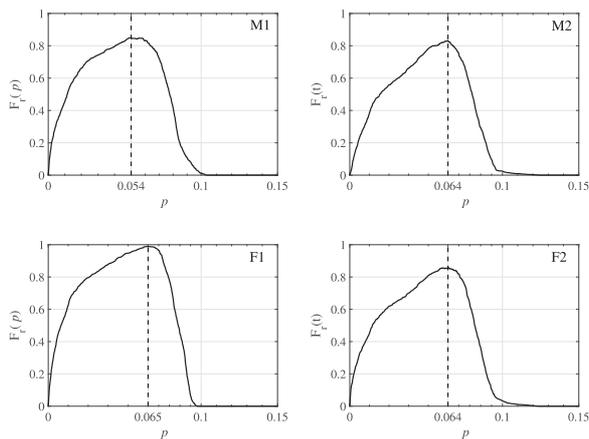


図 5 データグループごとの  $F_r(p)$  (実線) と  $p$

Fig. 5 The functions of  $F_r(p)$  (solid line) and the values of  $p$  in each data group.

ビブラートの有無を正しく判定できる確率となる。

$$F_r(p) = (1 - F_v(p))F_n(p), \quad (10)$$

この式に使われる各パラメータは、ビブラートパラメータを推定する前に算出されるものであることから、算出された  $F_r(p)$  は計算シミュレーションに用いる 4つの手法ごとに影響の差はない。ここで、 $F_r(p)$  が最大値となる  $p$  を、 $P_v(t)$  のビブラート判定の閾値として決定した。図 5 に各データグループの結果を示す。この図は、縦軸が  $F_r(p)$ 、横軸が  $p$  を示しており、 $F_r(p)$  を実線、最大値となる  $p$  を破線で示す。それぞれデータグループにより  $F_r(p)$  は異なるため、本実験においては、 $P_v(t)$  の閾値を M1 は男性 1 以外の 3 名からなる訓練データを用いて 0.054、同様に各訓練データから M2 は 0.064、F1 は 0.065、F2 は 0.064 と異なる値に設定した。

#### 4.4.2 誤差における許容範囲の策定

文献 [16] の手法から、目安となる速さ、深さの誤差割合とビブラート判定した開始時刻の誤差の許容値を求め、本研究の評価に用いる誤差における許容範囲を決定する。ここでのビブラート判定した開始時刻の誤差は、式 (9) で算出された誤差の絶対値を用いる。使用するデータには、訓練データから従来法の制限範囲内にある計 216 データを用いた。このデータを用いて、それぞれのビブラートの速さ、深さの誤差割合とビブラート判定した開始時刻の誤差を求め、累積分布関数でそれぞれ約 95% を含む値を許容範囲とした。データグループごとの速さと深さの累積分布関数は、図 6 に示す。この図は、縦軸が累積相対度数、横軸が誤差割合の累積分布関数を示しており、実線はビブラートの速さ、破線はビブラートの深さを示し、点線は 95% となる値を示す。また、図 7 は、データグループごとのビブラート判定した開始時刻の誤差の累積分布関数を示す。この結果から、深さと速さの誤差割合とビブラート判定した開始時刻の誤差の許容値を表 4 に示す値とした。

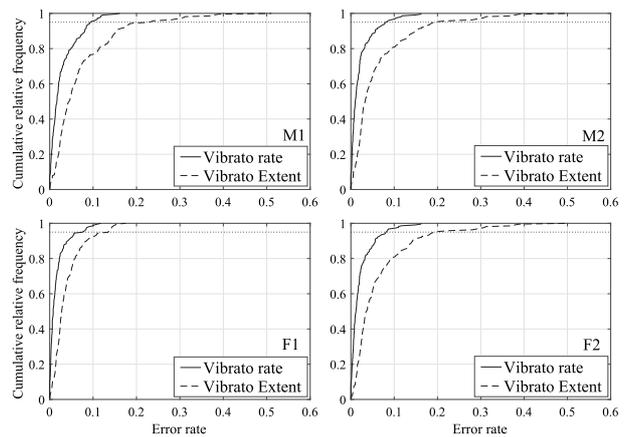


図 6 ビブラートの深さ、速さの誤差に関する累積分布関数

Fig. 6 Cumulative distribution functions in the errors of vibrato rate and extent.

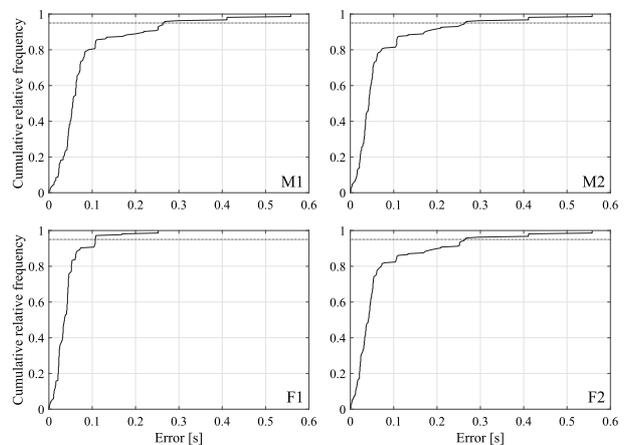


図 7 ビブラート判定した開始時刻の誤差の累積分布関数

Fig. 7 Cumulative distribution function in the error of vibrato's boundary.

表 4 ビブラートの深さ、速さ、開始時刻の誤差の許容値

Table 4 Allowable errors in the vibrato rate and extent and vibrato's boundary.

データグループ	速さの誤差	深さの誤差	開始時刻の誤差 [s]
M1	0.094	0.23	0.26
M2	0.080	0.19	0.26
F1	0.074	0.13	0.11
F2	0.081	0.19	0.26

#### 4.5 提案法の閾値設定

提案法では、ビブラート開始時刻を補正する必要がある。補正值算出のため、4つのデータグループごとに訓練データを分析し、ビブラート判定した開始時刻の誤差の中央値を補正值とした。分析に用いる手法は手法 A、B でそれぞれ用いて求められた補正值を、テストデータのそれぞれ手法 C、手法 D での分析に使用した。補正值を算出した結果は表 5 に示す。 $F_0$  推定のフレームシフト幅が 10ms のため、得られた値を 10ms 単位で四捨五入することと丸めることとした。

表 5 実験に用いた開始時刻補正值

Table 5 Boundary offset used for the evaluation.

データグループ	手法 C の補正值 [ms]	手法 D の補正值 [ms]
M1	40.0	40.0
M2	40.0	40.0
F1	20.0	20.0
F2	40.0	40.0

## 5. 計算機シミュレーション結果

データグループごとと全データグループをまとめた計算機シミュレーション結果を表 6 に示す。この表では、各評価項目で従来法 A より性能が高いものを太字で表し、最も性能が高いのものには太字に加え下線で示す。

### 5.1 従来法と提案法の Fine error における 2 標本 t 検定結果

本研究が目的とする提案法である手法 D と従来法である手法 A の Fine error の有意差検定の結果を表 7 に示す。この表ではデータグループごとの各 Fine error に対する  $p$  値の範囲を表している。データグループ F1 の Fine boundary error 以外は少なくとも  $p < 0.01$  で有意であり、手法 D が高い性能を示している。それに加え、その Fine boundary error では手法 D の結果が手法 A を下回っている状態で有意差が認められないことから、同程度の性能を示している。そのため、手法 A と比べ提案法である手法 D の性能は向上したことを示せたといえる。

### 5.2 Allowable number の分析結果

Allowable number は、データグループ F2 の Boundary number を除き提案法による精度の向上が認められるため、提案法はおおむね期待どおり動作しているといえる。Rate と Extent については、データグループ F1 の Rate number 以外において提案した 2 種を組み合わせた手法 D の結果が最良である。ビブラートの特性には歌手への依存性があるため全データにおいて手法 D が最良の結果ではないが、提案した工夫は平均的に良好な結果をもたらすといえる。

### 5.3 Fine error の分析結果

Fine error については、特に Rate と Extent error についてビブラート開始時刻補正による効果が顕著であり、手法 C か手法 D のどちらかがつねに最良の結果を示している。Boundary error については、傾向が分かれており、従来法よりも低い性能を示す結果も得られている。ただし、これらは、Allowable boundary number の向上した場合、許容範囲の境界付近かつ外側のデータが内側に補正されることにより増加した可能性もある。Allowable boundary number が向上しつつ Fine boundary error も向上してい

表 6 計算機シミュレーション結果。太字は A (従来法) よりも高い性能を示したもので、4 手法の中で最も優れていたものは太字に下線を加えている

Table 6 Results of the evaluation. Bold font represents that its performance was better than the conventional method (A), and we added the underline in the best value of all methods.

データグループ M1				
	A	B	C	D
Allowable rate number	494	<b>503</b>	<b>542</b>	<b>553</b>
Allowable extent number	531	<b>545</b>	<b>547</b>	<b>560</b>
Allowable boundary number	555	<b>558</b>	547	538
Fine rate error [%]	2.74	<b>2.58</b>	<b>1.49</b>	1.54
Fine extent error [%]	7.70	<b>7.17</b>	<b>4.79</b>	<b>4.70</b>
Fine boundary error [ms]	71.9	75.7	<b>53.4</b>	<b>53.9</b>
False positive error [%]	11.1	<b>8.64</b>	11.1	<b>8.64</b>
False negative error [%]	<b>12.3</b>	<b>12.3</b>	<b>12.3</b>	<b>12.3</b>

データグループ M2				
	A	B	C	D
Allowable rate number	438	<b>443</b>	<b>506</b>	<b>519</b>
Allowable extent number	508	<b>516</b>	<b>532</b>	<b>551</b>
Allowable boundary number	515	<b>525</b>	<b>542</b>	<b>555</b>
Fine rate error [%]	2.45	<b>2.39</b>	<b>1.20</b>	1.33
Fine extent error [%]	7.52	<b>7.26</b>	<b>4.16</b>	4.28
Fine boundary error [ms]	68.9	<b>63.6</b>	<b>62.9</b>	<b>58.1</b>
False positive error [%]	12.3	<b>10.2</b>	12.5	<b>10.2</b>
False negative error [%]	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>

データグループ F1				
	A	B	C	D
Allowable rate number	326	322	<b>356</b>	<b>350</b>
Allowable extent number	232	<b>240</b>	<b>263</b>	<b>270</b>
Allowable boundary number	179	170	<b>219</b>	<b>209</b>
Fine rate error [%]	2.91	<b>2.74</b>	<b>2.62</b>	<b>2.47</b>
Fine extent error [%]	6.09	<b>5.58</b>	<b>5.56</b>	<b>5.04</b>
Fine boundary error [ms]	54.8	60.3	<b>53.7</b>	56.2
False positive error [%]	8.80	<b>8.18</b>	8.80	<b>8.18</b>
False negative error [%]	<b>43.1</b>	<b>43.1</b>	<b>43.1</b>	<b>43.1</b>

データグループ F2				
	A	B	C	D
Allowable rate number	466	<b>488</b>	<b>540</b>	<b>551</b>
Allowable extent number	513	<b>534</b>	<b>543</b>	<b>557</b>
Allowable boundary number	<b>558</b>	552	555	550
Fine rate error [%]	2.22	<b>2.14</b>	<b>1.30</b>	<b>1.05</b>
Fine extent error [%]	6.51	<b>6.04</b>	<b>3.72</b>	<b>3.25</b>
Fine boundary error [ms]	75.3	<b>72.4</b>	<b>59.4</b>	<b>55.1</b>
False positive error [%]	10.5	<b>7.56</b>	10.5	<b>7.56</b>
False negative error [%]	<b>1.54</b>	<b>1.54</b>	<b>1.54</b>	<b>1.54</b>

総合結果				
	A	B	C	D
Allowable rate number	1724	<b>1756</b>	<b>1944</b>	<b>1973</b>
Allowable extent number	1784	<b>1835</b>	<b>1885</b>	<b>1938</b>
Allowable boundary number	1807	1805	<b>1863</b>	<b>1852</b>
Fine rate error [%]	2.56	<b>2.44</b>	<b>1.57</b>	<b>1.51</b>
Fine extent error [%]	7.10	<b>6.66</b>	<b>4.41</b>	<b>4.21</b>
Fine boundary error [ms]	70.4	<b>69.4</b>	<b>58.0</b>	<b>55.8</b>
False positive error [%]	10.7	<b>8.60</b>	<b>10.7</b>	<b>8.60</b>
False negative error [%]	<b>14.2</b>	<b>14.2</b>	<b>14.2</b>	<b>14.2</b>

るデータグループ F1 は、提案法による改善が有効に機能した結果だといえる。

表 7 従来法 (手法 A) と提案法 (手法 D) の Fine error に対する t 検定結果. \*:  $p < 0.01$ , \*\*:  $p < 0.001$ , \*\*\*:  $p < 0.0001$  としている

Table 7 Results of t-test in the fine error between the conventional method (method A) and proposed method (method D). Define as \*:  $p < 0.01$ , \*\*:  $p < 0.001$ , and \*\*\*:  $p < 0.0001$ .

	Fine rate error	Fine extent error	Fine extent error
M1	***	***	***
M2	***	***	**
F1	*	**	
F2	***	***	***
総合結果	***	***	***

### 5.4 False error の分析結果

False negative error については、全手法において数値の差は見られず、提案法による性能改善効果はないといえる。これは、提案法が精度の改善が目的であり、ビブラート検出においては影響が出なかったことを意味する。歌手による差が顕著である理由については、次章で説明する。

False positive error では、全データグループでビブラートの放物線補間を用いた手法 B と手法 D が従来法 A と比べて向上している。これは、許容範囲の境界付近の外側にあった検出用パラメータが、精度の改善により許容範囲に含まれたことが原因である。

## 6. 考察

本章では、計算機シミュレーション結果から、提案法の有効性について考察する。また、さらに精度を改善するための課題についても論じる。

### 6.1 ビブラートの速さと深さの推定精度の向上

ビブラートの放物線補間と開始時刻の補正を従来法に組み込むことにより、False error の性能を保持または向上させつつ、Allowable rate number と Allowable extent number の性能が向上した。全データグループのテストデータにおける手法 A でビブラート判定した開始時刻の誤差の累積分布関数を図 8 に示す。この図は、縦軸が累積相対度数、横軸が誤差の累積分布関数を示しており、実線はデータグループ M1、破線は M2、点線は F1、鎖線は F2 を示す。3 章で説明したとおり、開始時刻の誤差は真値と推定値との差で計算している。どのデータグループでも誤差が正の数である割合が負の数である割合より高いことから、ビブラート判定が本来の開始時刻よりも手前にずれているという仮説が正しいと考えられる。ビブラートパラメータ推定の極大・極小点の検出が深さと速さにおける制限がないため、開始時刻よりやや手前にある  $F_0$  軌跡の微細な変動が、誤差の原因となる、ビブラート開始時刻の補正が特に有効であった点については、開始時刻補正がこの影響を除去す

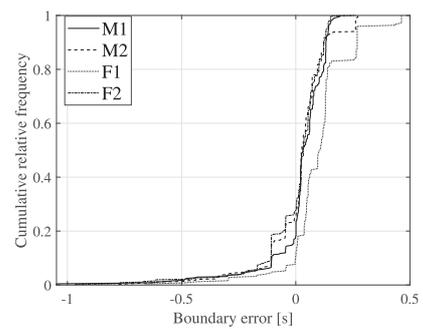


図 8 ビブラート判定した開始時刻の誤差の累積分布関数  
Fig. 8 Cumulative distribution function in the error of the vibrato's boundary.

ることができていたと考えられる。3.2 節で説明した図 3 は本計算シミュレーションのビブラート開始時刻の補正結果の 1 つを示す。3.2 節で説明したように、補正を行うことで  $F_0$  軌跡の微細な変動の極大・極小点が省かれ、極大・極小点を用いて算出されるビブラートの速さと深さの誤差が減少したと考えられる。

ビブラートの放物線補間が精度向上に有効であったことは、Fine rate error と Fine extent error の精度向上から裏付けられる。この原因としては、放物線補間によって極大・極小点の推定が高精度化し、特に速度の速いビブラートの検出精度が向上していたためであると考えられる。以上のことから、ビブラートの速さ、深さの推定精度向上に 2 つの改善法は有効であるといえる。

### 6.2 ビブラートの開始時刻の推定精度の向上

ビブラートの放物線補間とビブラート開始時刻の補正を従来法に組み込むことにより、False error の性能を保持または向上したが、Allowable boundary number と Fine boundary error の性能が一部で低下することも確認できた。ビブラート開始時刻の補正に関する性能低下の原因として、女性 1 の歌声による影響があげられる。図 8 より、データグループ F1 は他の 3 グループと比べ、正方向にずれていることから、女性 1 のデータが訓練、テストデータに悪影響を及ぼしたと考えられる。

### 6.3 提案法の有効性

本評価では、ビブラートの有無の検出、検出できた場合許容範囲に含まれるか、許容範囲内の推定精度という複数の項目に分かれて評価している。たとえば許容範囲内に含まれるデータ数の増加は、許容範囲の境界付近のデータが境界内部に含まれた場合、推定精度に悪影響を及ぼす。手法の優劣については、何を目的にするかにより結論が異なることとなる。

提案法については、特定の項目で従来法が最良の結果となる場合もあるが、全体的には両改善法を組み合わせ手法 D が良好な結果を与えている。特に Allowable number

についてはデータグループ M1, F2 の Boundary number 以外において従来法よりも良好であり, 8 項目について最良の結果である. このことや総合結果では全手法が従来法と比べ全項目で良好な結果になっていることから, 提案法は従来法よりも優れていると考えられる.

## 7. おわりに

本研究では, 高精度なビブラートの速さと深さ, 開始時刻の推定を目的としたビブラートパラメータ推定精度の向上手法を提案した. 提案法では, 2 つの改善法を組み合わせることでビブラートパラメータ推定の改善を図った. また提案手法の有用性を確認するため, 2 つの改善法の組合せに対するオープンテストの計算機シミュレーションを行った. 計算機シミュレーション結果から, 従来法と比べすべての手法が一部の精度に対して向上していることが確認できた.

2.1 節で述べたように人間が発したビブラートの速さと深さは時間変動が存在する [18], [19]. 本研究では, 定常波のビブラートのみを用いたため, 今後の課題としては, 提案法によって時間変動を持つビブラートパラメータ推定の高精度化も期待できるかの調査を行うことがあげられる. それに加えて, 将来的にビブラートの速さと深さにおける時間変動に対応するためにフレーム単位で速さと深さを高精度に推定する方法も構築する必要がある. また, 改善法の 1 つであるビブラート開始時刻の補正では, 歌手によるばらつきを調べ, パラメータを最適化をすることがあげられる.

謝辞 本研究は, JST さきがけ JPMJPR18J8 の支援を受けた.

## 参考文献

- [1] Sundberg, J.: Research on The Singing Voice in Retrospect, *TMH-QPSR*, Vol.45, No.1, pp.11–22 (2013).
- [2] Sundberg, J.: An articulatory interpretation of the ‘singing formant’, *STL-QPSR*, Vol.13, No.1, pp.45–53 (1972).
- [3] 矢永龍一郎, 河原英紀: 会話音声と歌唱音声の基本周波数制御の動特性について, 情報処理学会研究報告音楽情報科学, pp.71–76 (2003).
- [4] Akagi, M. and Kitakaze, H.: Perception of Synthesized Singing Voices with Fine Fluctuations in Their Fundamental Frequency Contours, *Proc. ICSLP 2000*, pp.458–461 (2000).
- [5] Kojima, K., Yanagida, M. and Nakayama, I.: Variability of Vibrato – A Comparative Study Between Japanese Traditional Singing and Bel Canto, *Proc. Speech Prosody 2004*, pp.151–154 (2004).
- [6] Nakayama, I.: Comparative Studies on Vocal Expressions in Japanese Traditional and Western Classical-Style Singing, Using a Common Verse, *Proc. ICA 2004*, pp.1295–1296 (2004).
- [7] 森勢将雅, 平地由美, 坂野秀樹, 入野俊夫, 河原英紀: STRAIGHT を用いたビブラート歌唱音声の統計的性質, 日本音響学会研究発表会講演論文集, Vol.2005, No.1, pp.269–270 (2005).
- [8] 鈴木千文, 坂野秀樹, 旭 健作, 森勢将雅: 歌唱音声の類似度評価を目的とした基本周波数の変動量を反映するビブラート特徴量の提案, 電気学会論文誌 C (電子・情報・システム部門誌), Vol.137, No.12, pp.1607–1614 (2017).
- [9] Nwe, T.L. and Li, H.: Exploring Vibrato-Motivated Acoustic Features for Singer Identification, *IEEE Trans. Audio, Speech, and Language Processing*, Vol.15, No.2, pp.519–530 (2007).
- [10] 齋藤 毅, 後藤真孝: 歌唱指導による歌声中の音響特徴の変化: 歌唱力評価に寄与する音響特徴の検討, 日本音響学会講演論文集, No.2-Q-16, pp.583–586 (2009).
- [11] Sundberg, J.: The KTH synthesis of singing, *Adv. Cognit. Psychol. (Special issue on Music Performance)*, Vol.2, pp.131–143 (2006).
- [12] Saitou, T., Goto, M., Unoki, M. and Akagi, M.: Speech-To-Singing Synthesis: Converting Speaking Voices to Singing Voices by Controlling Acoustic Features Unique to Singing Voices, *Proc. WASSPA 2007*, pp.215–218 (2007).
- [13] Kenmochi, H. and Ohshita, H.: VOCALOID - Commercial Singing Synthesizer Based on Sample Concatenation, *Proc. INTERSPEECH 2007*, pp.4010–4011 (2007).
- [14] 中野倫靖, 後藤真孝: VocaListener: ユーザ歌唱を真似る歌声合成パラメータを自動推定するシステムの提案, 情報処理学会音楽情報科学研究会研究報告 2008-MUS-75, Vol.2008, No.50, pp.49–56 (2008).
- [15] Nakano, T. and Goto, M.: VocaListener: A singing-to-singing synthesis system based on iterative parameter estimation, *Proc. SMC*, pp.343–348 (2009).
- [16] 中野倫靖, 後藤真孝, 平賀 譲: 楽譜情報を用いない歌唱力自動評価手法, 情報処理学会論文誌, Vol.48, No.1, pp.227–236 (2007).
- [17] 右田尚人, 森勢将雅, 西浦敬信: 歌唱データベースを用いたヴィブラートの個人性の制御に有効な特徴量の検討, 情報処理学会論文誌, Vol.52, No.5, pp.1910–1922 (2011).
- [18] Prame, E.: Measurements of The Vibrato Rate of Ten Singers, *STL-QPSR*, Vol.33, No.4, pp.73–86 (1992).
- [19] Sundberg, J. and Bretos, J.: Measurements of Vibrato Parameters in Long Sustained Crescendo Notes as Sung by Ten Sopranos, *TMH-QPSR*, Vol.43, No.1, pp.37–44 (2002).
- [20] Morise, M., Yokomori, F. and Ozawa, K.: WORLD: A Vocoder-Based High-Quality Speech Synthesis System for Real-Time Applications, *IEICE Trans. Information and Systems*, Vol.E99-D, No.7, pp.1877–1884 (2016).
- [21] Morise, M.: D4C, a band-a-periodicity estimator for high-quality speech synthesis, *Speech Communication*, Vol.84, pp.57–65 (2016).
- [22] Morise, M.: Harvest: A high-performance fundamental frequency estimator from speech signals, *Proc. INTERSPEECH 2017*, pp.2321–2325 (2017).
- [23] Rabiner, L., Cheng, M., Rosenberg, A. and McGonegal, C.: A comparative performance study of several pitch detection algorithms, *IEEE Trans. Acoustics, Speech, and Signal Processing*, Vol.24, No.5, pp.399–418 (1976).
- [24] de Cheveigné, A. and Kawahara, H.: YIN, a fundamental frequency estimator for speech and music, *The Journal of the Acoustical Society of America*, Vol.111, No.4, pp.1917–1930 (2002).
- [25] Nakatani, T., Amano, S., Irino, T., Ishizuka, K. and Kondo, T.: A method for fundamental frequency estimation and voicing decision: Application to infant utterances recorded in real acoustical environments, *Speech Communication*, Vol.50, No.3, pp.203–214 (2008).



宮崎 嵩大

1997年生。2019年山梨大学工学部コンピュータ理工学科卒業。現在、山梨大学大学院医工農学総合教育部工学専攻コンピュータ理工学コースに在籍中。音声分析に関する研究に従事。



森勢 将雅 (正会員)

1981年生。2004年和歌山大学卒業。2008年和歌山大学大学院博士課程修了，博士（工学）取得。関西学院大学，立命館大学，山梨大学を経て，2019年より明治大学総合数理学部専任准教授。人間の知覚情報を活用した音声分析・合成・デザインの研究に従事。