Vol.2019-MUS-123 No.12
Vol.2019-SLP-127 No.12
2019/6/22

音声波形からのヴィブラートパラメータ推定の高精度化と評価

宮崎 嵩大[1,a)]　森勢 将雅[2]

概要：誰でも歌うことやそれを共有することができる文化の発展に伴い，歌声分析の需要が増加している．歌声分析に用いられるパラメータの 1 つであるヴィブラートは，歌唱力や歌声の知覚に影響することが知られており，重要なパラメータであるといえる．高精度なヴィブラート分析が実現できれば，ヴィブラートによる個人性の違いなどのより詳細な歌声の特性解析ができると考えられる．本研究では高精度なヴィブラートの深さ，速さの推定を目的としたヴィブラート区間検出手法を目指し，先行研究で提案されたヴィブラート区間検出手法を基に 3 つの改善手法を提案する．また，計算機シミュレーションによる比較実験を実施し，提案手法の有用性を確認する．

1. はじめに

歌うことは多くの人によって楽しまれており，その代表的な例として，カラオケが挙げられる．カラオケは余暇活動の一つとして挙げられ，様々な人たちに親しまれている．また，ニコニコ動画[*1]や YouTube[*2]といった動画共有サイトでは，歌唱音声を共有して楽しむ文化も存在する．他にスマホアプリでも nana[*3]があり，これも歌声や楽器演奏の投稿，共有ができる．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てきている．

これらの文化の発展に伴い歌声分析の需要が増加し，歌声を対象とした研究事例が報告されている [1]．歌声の基本周波数（F_0）は会話音声と異なる特徴を持つことが知られている [2]．歌声分析によく用いられる F_0 のパラメータには，楽譜情報から逸脱した動的変動成分が含まれる．これらにはオーバーシュートや微細変動成分等の発声器官の物理的な制約に起因する成分 [3] のほかに，ヴィブラートやポルタメント等の歌唱者が意識的に表現する成分 [4], [5] も存在する．ヴィブラートは，歌唱力に影響することや [6]，歌声の知覚に影響することが知られている [7]．また，ヴィブラートは VOCALOID [8] や，話し声を歌声に変換する歌声合成システム [7] などで用いられている．以上のことから，ヴィブラートは歌声分析に用いられるパラメータとして重要なものの 1 つであるといえる．高精度なヴィブラート分析が実現できれば，さらに詳細な歌声の特性解析ができると期待される．

本研究では，高精度なヴィブラートパラメータの分析を目的とし，ヴィブラート区間検出手法を提案する．本研究におけるヴィブラートパラメータとは，F_0 によるヴィブラートの制御に用いられるヴィブラートの速さ，深さを指す [7], [9]．本手法によるヴィブラート分析により，高精度なヴィブラートパラメータが得られる．得られたパラメータを用いて詳細な歌声合成，歌声分析ができることが期待される．

2. ヴィブラートに関する関連研究

2.1 ヴィブラートに関する定義とその構成要素

ヴィブラートとは，音を伸ばした歌唱において，その音高を保ちつつ高さなどを細かく振動させる歌唱表現である．ヴィブラートの特徴量として，ヴィブラートの速さを表す vibrato rate と深さを表す vibrato extent が挙げられる．この 2 つは，F_0 によるヴィブラートの制御に用いられている主要なパラメータである．これらは，ヴィブラート区間の F_0 軌跡より，図 1 のように速さと深さを構成するパラメータ R_n〔s〕，E_n〔cent〕を抽出し，式 1，2 によって算出される．

$$\frac{1}{\text{rate}} = \frac{1}{N}\sum_{n=1}^{N} R_n \tag{1}$$

[1] 山梨大学
University of Yamanashi
[2] 明治大学
Meiji Uniersity
[a)] g19tk020@yamanashi.ac.jp

[*1] https://www.nicovideo.jp/（最終検索日：2019 年 5 月 16 日）
[*2] https://www.youtube.com（最終検索日：2019 年 5 月 16 日）
[*3] https://nana-music.com/（最終検索日：2019 年 5 月 16 日）

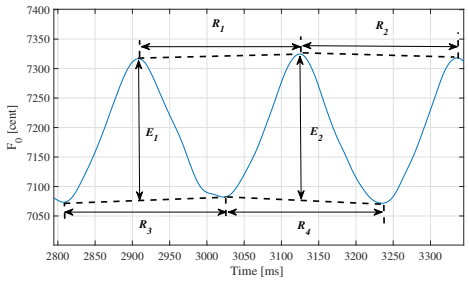

図 1　ヴィブラートの F0 軌跡

$$\text{extent} = \frac{1}{2N}\sum_{n=1}^{N} E_n \tag{2}$$

N はヴィブラート区間の F_0 軌跡から抽出された各パラメータの総数を示している．また，今回の cent 単位への変換では，式 3 に示すように，中央ハ音の周波数 f_c $(= 440 * 2^{\frac{3}{12}-1} = 261.62...\ \mathrm{Hz})$ の cent 値を 4800 cent として，周波数 f_Hz を，cent 単位の f_cent に変換している．

$$f_\mathrm{cent} = 1200 \log_2 \left(\frac{f_\mathrm{Hz}}{f_\mathrm{c}} + 4800 \right) \tag{3}$$

本研究では，目的に対してヴィブラート区間検出法が改善出来たか精度比較の計算シミュレーションを行う必要がある．それに伴い，ヴィブラートの速さと深さの真値が存在するヴィブラートが必要となる．しかし，ヴィブラートの速さ，深さが時間変動しているものを用いる場合，ヴィブラートを作成したパラメータから速さと深さの真値を求めることは困難である．そこで本研究は，ヴィブラートを作成したパラメータから，速さと深さの真値を求めることが可能な速さと深さの時間変動なしのヴィブラートを扱う．

2.2　ヴィブラート区間検出手法

先行研究として，中野らによってヴィブラート区間検出手法が提案されている [10]．この検出手法は，楽譜情報を用いずに歌唱力を自動で評価することを目的として作られた．従来手法では，F_0 の時間変化 $F_0(t)$〔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) を行うことでヴィ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 $X(f,t)$ をヴィブラート区間判定に用いる．その振幅スペクトルのヴィブラートの速さに対応する周波数成分が鋭いピークになる事を利用している．時刻 t におけるヴィブラート速さの周波数帯域のパワー $\Psi_v(t)$ とピークの鋭さ $S_v(t)$ が式 4，5 のように定義されている．

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

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

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 にあるように，各時刻 t ごとに全周波数帯域のパワーで正規化したものとなる．

$$\hat{X}(f,t) = \frac{X(f,t)}{\int X(f,t) df} \tag{6}$$

これらを用いて，時刻 t におけるヴィブラートらしさ $P_v(t)$ が式 7 のように定義されている．

$$P_v(t) = S_v(t)\Psi_v(t) \tag{7}$$

そして $P_v(t)$ が大きく，速さと深さが制限内で，$F_0(t)$ がその平均音高と 5 回以上交差する区間をヴィブラートとして判定している．また，速さと深さのそれぞれの制限範囲は，5 – 8 Hz と 30 – 150 cent にしていた．

2.3　本研究の位置付け

2.2 節では，歌唱力評価を目的としたヴィブラート区間検出手法を紹介した．この検出手法では，速さと深さの制限範囲を，5 – 8 Hz と 30 – 150 cent にしていたが，その制限範囲を超えるヴィブラートも存在している [9]．また，歌声合成で作成した音声を扱う場合もその制限範囲を超える可能性がある．

本研究では，先行研究の制限範囲をある程度広げても，高精度なヴィブラートの速さ，深さの推定が可能であるようにヴィブラート区間検出法の改善を図る．その際，文献 [10] に具体的に書かれていないパラメータの値の調整方法と，ヴィブラート区間検出手法の改善手法を提案する．また，提案した改善手法によってヴィブラート音声を分析し，分析精度を確認する．

3.　ヴィブラート区間検出手法の提案

従来手法ではヴィブラートの有無を判定することを目的としており，高精度なパラメータ推定を目的とした場合，パラメータ推定誤差の原因の対処を行う必要がある．そのため，本ヴィブラート区間検出手法では主に下記の三つの改善手法を中野らが提案した従来手法に適用させる．

3.1　$\mathrm{F_0}$ の分析シフト幅の変更

従来手法では F_0 の分析シフト幅を 10 ms としている．この分析シフト幅をより細かくすることで，F_0 波形の各極大点，極小点を正確に求められるようになり，パラメータ推定の精度が高くなることが見込まれる．本研究の分析に用いる WORLD では 1 ms 以下の分析シフト幅を扱う場合，1 ms で分析した点と点を補間している．本研究では，各極大点と極小点を求めることを考慮して，分析シフト幅を 1 ms にした．

3.2　LPF による $\mathrm{F_0}$ 軌跡の平滑化

ヴィブラートが付与された F_0 区間には，準周期的な振

動だけでなく細かい振動が混入している．この細かい振動が原因で，ヴィブラートの速さと深さの誤差が大きくなることや，それに伴って速さと深さが制限範囲から外れることが，正しくヴィブラート区間を判定できない理由として考えられる．そこで本手法は，一定以上の周波数を減衰させる LPF を F_0 に適用することで細かい振動の影響を抑制し，この問題の解決を図る．

LPF を適用するにあたり，通過域リップルの影響で，ヴィブラートの周波数もある程度減衰することが予測される．その結果，ヴィブラートの深さの誤差が大きくなることや，それに伴い深さが制限範囲を外れ，正しくヴィブラート区間を判定できないことが予想される．そこで，本手法では LPF が周波数によってどのくらい減衰させるかを求め，ヴィブラートの深さに適した補正を適用した．

3.3 ヴィブラート判定区間の補正

ヴィブラート判定した区間の両端付近に細かい振動が誤って混入していることが確認された．この誤判定が原因で，ヴィブラートの速さと深さの誤差が大きくなることや，それに伴って速さと深さが制限範囲から外れることが，正しくヴィブラート区間を判定できない理由として考えられる．そこで全データを分析し算出されたヴィブラート判定の開始時間，終了時間の平均誤差〔s〕を用いて，それぞれの開始時間，終了時間を短縮する．

4. 精度比較を目的とした計算シミュレーションの計画

4.1 計算シミュレーション内容

本計算シミュレーションの目的は，ヴィブラートの深さと速さの制限範囲を広げた際，提案手法の有用性を検証することである．実験手順としては，ヴィブラートのかかっていない音声に真値ありのヴィブラートをかけ，ヴィブラート音声を作成する．それらのデータを各手法ごとに分析し，分析結果から精度を比較する．また，実験で必要となる F_0 推定は音声分析合成システム WORLD [11]（D4C edition [12]）の Harvest [13] で行う．

4.2 評価の指標

評価は以下の 6 つの指標をそれぞれ分析した結果を用いる．全ての指標は，0 に近いほど精度が高いことを意味する．

- Gross error [%]
  Gross error は，全体に対するヴィブラートの深さと速さが許容誤差割合範囲内に無いデータ数の率を示す．許容誤差割合の詳細については 5.3 節で説明する．
- Fine error [%]
  Fine error は，ヴィブラートの深さと速さが許容誤差割合範囲内にあるデータの平均二乗誤差率を表す．誤差率とは真値に対してどのくらいの割合の誤差かを計算したものとなる．速さについては Fine rate error，深さについては Fine extent error とする．
- Boundry error [s]
  ヴィブラート判定した区間の両端の平均二乗誤差を表す．区間の開始地点については Start boundry error，区間の終了地点については End boundry error とする．
- Absence error [%]
  Absence error は，ヴィブラートがかかっていない全データに対してヴィブラートと判定した率を表す．また他の指標の分析に用いたデータでなく．ヴィブラートのかかっていないデータを用いる．

表 1　実験に用いる音声加工の条件

パラメータ	値
ヴィブラートの速さ〔Hz〕	1 – 12 (11 パターン)
ヴィブラートの深さ〔cent〕	20 – 300 (11 パターン)
ヴィブラートの割合〔%〕	55 – 85 (11 パターン)

4.3 計算シミュレーションに使用した音声

本計算シミュレーションでは，右田らが作成した歌声データベース [9] を利用する．男性 2 人がヴィブラートをかけないように発した音声の中から，音高が高い，普通，低いの 3 通りを使用した．そして歌唱内容が/a/の計 6 音声を，以下の表 1 に示す条件によって加工した音声をテストデータとした．また 4.2 節で説明した Absence error の分析では，前述した男性 2 人がヴィ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 130 音声を用いた．

4.3.1 音声加工手法

計 6 音声の F_0 に，WORLD を用いて表 1 に示す条件の正弦波をそれぞれ付与し，音声を合成することでヴィブラートの付与を実現した．VOCALOID ソフトである HATSUNE MIKU V4X[*4]にある速さのパラメータの値を最大にして WORLD で分析した F_0 から目視で 9.5 Hz 辺りであると算出した．ヴィブラートの速さの上限は，それを参考に一部その値を超えるように設定した．ヴィブラートの深さの上限とヴィブラート区間の割合では，右田らが作成した歌声データベースの分析結果 [9] を参考に設定した．また，速さと深さの下限は聴いた際に，一部ヴィブラートと判断しづらいようなものを含むように設定した．これらの上限下限から，ヴィブラートの速さと深さ，割合は 10 等分した各 11 パターンを使用した．本研究は，これらの計 7986 データを用いた． また，検出手法が 320 ms の区間を用いて判定していることから，周期が 320 ms 以上でなければ， ヴィブラートの速さと深さの判定ができない．したがって，3.125 Hz 未満の速さが判定ができないことが予測される．その確認のため，一部該当周波数以下の

[*4] https://ec.crypton.co.jp/pages/prod/vocaloid/mikuv4x （最終検索日：2019 年 5 月 16 日）

表 2　提案手法の実験のヴィブラート判定条件

パラメータ	値
$P_v(t)$	0.63 以上
ヴィブラートの速さ〔Hz〕	1 – 12
ヴィブラートの深さ〔cent〕	20 – 300
平均音高との交差数	5 回以上

表 3　実験に用いるヴィブラート区間検出手法

ラベル	実装した改善手法
A	なし
B	F_0 の分析シフト幅の変更
C	F_0 に適用させる LPF と補正の実装
D	ヴィブラート判定区間の補正
E	F_0 の分析シフト幅の変更
	F_0 に適用させる LPF と補正の実装
F	F_0 の分析シフト幅の変更
	ヴィブラート判定区間の補正
G	F_0 に適用させる LPF と補正の実装
	ヴィブラート判定区間の補正
H	F_0 の分析シフト幅の変更
	F_0 に適用させる LPF と補正の実装
	ヴィブラート判定区間の補正

条件も加えた．

4.4　比較に用いるヴィブラート区間検出手法

ヴィブラート区間判定条件は表 2 のように統一し，従来手法を含んだ表 3 のように 3 章で説明した改善手法を用いた計 8 通りの手法を用いて比較した．$P_v(t)$ の基準の設定については，5.2 節で詳しく説明する．比較する従来手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が分析できず，提案手法に対して不利になると考えられる．よってヴィブラートの速さと深さを 1 – 12 Hz，20 – 300 cent とテストデータの範囲を全て含むように設定した．それに伴い，$P_v(t)$ を求める際に扱う F_L と F_H も同様に変更した．

5.　従来手法のパラメータ設定

5.1　STFT の FFT ポイント数の設定

F_0 の分析シフト幅の変更を行うにあたり，STFT の FFT（fast Fourier transform）ポイント数を，後述する提案手法の周波数成分の分析シフト幅に近い値になるように設定する必要がある．これによって従来手法と提案手法の精度を比較する際に，周波数成分の分析シフト幅の違いによる影響を抑制でき，F_0 の分析シフト幅の変更による精度の変化を確認できる．提案手法の FFT ポイント数を 1024 としたことから，FFT ポイント数を 103 と設定した．

5.2　ヴィブラート判定に対する $P_v(t)$ の基準の設定

ヴィブラートらしさ $P_v(t)$ の基準を決定する際，速さと深さが従来手法の制限範囲内にあるヴィブラートの $P_v(t)$ を本来の従来手法で分析し，累積分布関数を用いた．使用

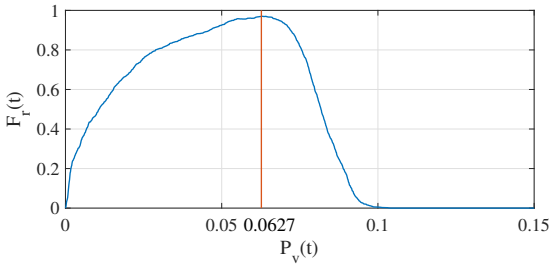

図 2　$F_r(t)$(青実線) と $P_v(t)$ の参考値 (赤実線)

データはテストデータから，ヴィブラートあり，無しのフレームを 1000 フレームずつ抽出し，それぞれ分析を行った．フレームとは，判定に用いる 320 ms の区間のことを指す．ヴィブラート無しの累積分布関数を, $F_n(p)$ とする．これは計算結果の $P_v(t)$ の値が p 以下だった際，ヴィブラート無しを正しく判定する確率を表す．ヴィブラート有りの累積分布関数を, $F_v(p)$ とする．これは計算結果の $P_v(t)$ の値が p 以下だった際，ヴィブラートが存在する確率を表す．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，ヴィブラートが存在しない確率を表し，ヴィブラートありをヴィブラート無しとは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，ヴィブラート有りを正しく判定している確率を表している．これらを用いて，式 8 で算出された $F_r(p)$ はヴィブラートの有無を正しく判定する確率となる．

$$F_r(p) = (\ 1 - F_v(p)\ )F_n(p) \tag{8}$$

この $F_r(p)$ が, 最大値となる p を $P_v(t)$ の基準として決定した．結果は図 2 のようになった．この図は，縦軸が $F_r(p)$，横軸が p の $F_r(p)$ を示しており，最大値が 0.97 で p が 0.0627 であったことから，$P_v(t)$ の基準を 0.063 とした．

5.3　従来手法からの許容誤差割合範囲の決定

本研究は，制限範囲をある程度広げても制度に支障のないことを目的としていたため，従来手法とどの程度同じ精度か確認する目安が必要となる．そこで，文献 [10] の深さと速さの制限範囲に設定した従来手法から，目安となる速さ，深さの誤差割合の許容値を求め，本研究の評価に用いる許容誤差割合範囲を決定する．本研究では許容誤差割合範囲内の分析結果は，ある程度の外れ値を省けていることを期待し，評価の一部とする．

使用するデータはテストデータから従来手法の制限範囲内にある計 792 データを用いた．このデータを用いて，それぞれのヴィブラートの速さ，深さの誤差割合を求め，累積分布関数でそれぞれ約 95%を含む値を許容誤差割合範囲とした．速さと深さの累積分布関数は，図 3 のようになった．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，青実線はヴィブラートの速さ，赤実線はヴィブラートの深さを示す．黄色実線は 95%を示

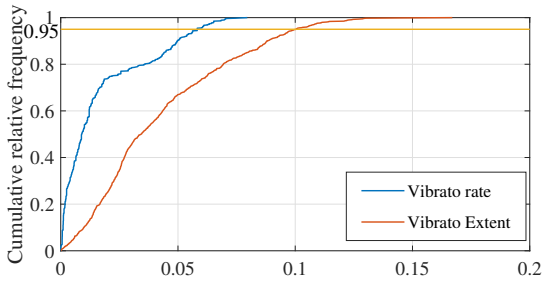

図 3 ヴィブラートの深さ，速さの誤差割合の累積分布関数

表 4 手法 C の LPF 次数の検討

LPF 次数	無し	2	4	8	16	32	64	128
Gross error [%]	79.1	**78.2**	79.7	82.0	83.8	92.5	94.4	96.9
Fine rate error [%]	**1.92**	2.57	2.47	2.71	2.50	2.67	2.70	2.95
Fine extent error [%]	6.92	6.95	6.95	6.95	6.88	6.48	6.50	**6.37**
Start boundry error [s]	**0.174**	0.244	0.246	0.283	0.313	0.380	0.469	0.600
End boundry error [s]	0.141	0.144	**0.140**	0.168	0.163	0.265	0.341	0.352
Absence error [%]	15.4	15.4	17.7	**8.46**	13.8	35.4	40.0	66.9

表 5 手法 E の LPF 次数の検討

LPF 次数	無し	2	4	8	16	32	64	128
Gross error [%]	**74.8**	76.6	76.5	76.5	77.0	79.1	83.1	85.4
Fine rate error [%]	**2.10**	2.23	2.23	2.25	2.25	2.37	2.38	2.59
Fine extent error [%]	**6.07**	6.12	6.14	6.19	6.35	6.54	6.45	6.53
Start boundry error [s]	**0.186**	0.189	0.188	0.188	0.189	0.242	0.279	0.291
End boundry error [s]	0.111	0.101	0.101	0.0999	0.100	**0.0972**	0.118	0.107
Absence error [%]	18.5	20.0	17.7	16.2	19.2	16.2	**12.3**	20.0

す．この結果から深さの許容誤差割合を 0.1 に，速さの許容誤差割合を 0.06 に決定した．使用したデータの許容誤差割合範囲内にある割合を分析したところ，約 92.1%となった．これは速さ，深さで別々に約 95%内の値を含むようにしたことが原因となる．

6. 提案手法のパラメータ設定

6.1 STFT の FFT ポイント数の設定

従来手法と同様，提案手法では STFT により求められる正規化した振幅スペクトルの中の設定したヴィブラートの速さに対応する周波数成分を利用し，ヴィブラート区間判定条件に使われる $P_v(t)$ を算出する．よって周波数成分の分析シフト幅が粗すぎると，正確な判定ができない恐れがある．そこで，1 Hz 近くごとの判定ができる 2 のべき乗の FFT ポイント数を考慮し，本研究では FFT ポイント数を 1024 とした．周波数成分の分析シフト幅をそろえる際，提案手法の FFT ポイント数のほうが従来手法より大きくなるため，FFT の計算時間を考慮し，提案手法の FFT ポイントが 2 のべき乗となるように設定した．これにより，分析シフト幅が 1 ms なので，周波数成分は約 0.977 Hz ごとに判定ができることが見込まれる．

6.2 深さの補正を適用した LPF の検討

LPF を実装するにあたり，適切な LPF 次数とカットオフ周波数を設定する必要がある．LPF はカットオフ周波数より高い周波数成分を低減させる．カットオフ周波数は今回実験に用いる速さの上限である 12 Hz とした．深さの補正の詳細については，6.3 節で後述する．深さの補正を適用した LPF 次数の検討では，**表 3** の提案手法 C と E を用いて，LPF をかけない F0 軌跡と，7 種類の次数の LPF をそれぞれ適用した F0 軌跡の，ヴィブラートの速さ，深さの分析を行った．3.125 Hz 未満の速さである条件で実験した場合，ヴィブラート区間外をヴィブラートと誤判定するため，実験後のデータから取り除いた．その結果を**表 4**，**表 5** に示す．最も性能が良い数値をそれぞれ太字で表している．

6.2.1 手法 C の LPF 次数の検討

表 4 で，LPF をかけていないものより一部性能が高いものは，LPF 次数 2，4，8，128 である．本研究ではパラメータ推定の精度の向上を目的としているため，Gross error で性能が高い LPF 次数 2 を採用する．また手法 C と同じ F_0 の分析シフト幅で LPF を適用する手法 G も，同じ LPF 次数を用いる．

6.2.2 提案手法 E の LPF 次数の検討

表 5 で，LPF をかけていないものより一部性能が高いものは，LPF 次数 32，64 である．ヴィブラートは本来音声のある地点から最後までかかることが一般的であり，音声の中間のみでかけるということは見られないと考え，実際の分析の際には影響が少ないと考えられる．そのため End boundry error の性能の優先度は低いと考えられ，Absence error で性能が高い LPF 次数 64 を採用する．また手法 E と同じ F_0 の分析シフト幅で LPF を適用する手法 E と H も，同じ LPF 次数を用いる．

6.3 LPF に対する深さの補正の検討

LPF に対する深さの補正値の検討として LPF 次数ごとの LPF を，ヴィブラートの速さの異なる正弦波に適用し，それぞれの減少割合を分析した．これを F_0 の分析シフト幅の 2 種類と LPF 次数ごとに行った．分析には速さを 1 – 12 Hz で 200 等分した 201 通りのデータを用いた．ヴィブラートの深さは全て 1 cent で統一されている．分析結果の一部を図 4 に示す．この図は，縦軸がヴィブラートの深さの LPF による減少割合，横軸がヴィブラートの速さを示している．このデータを用いて，速さの制限範囲内である 1 – 12 Hz 内にある波形の深さに減少割合に対する補正をかける．補正は式 9 に示すように行い，検出されたヴィブラートの深さ extent を，補正済みの深さ $\mathrm{extent}_{\mathrm{correct}}$ に変換する．$d(r)$ は，ヴィブラートの速さ r における深さの減少割合 d を示す．

$$\mathrm{extent}_{\mathrm{correct}} = \mathrm{extent} \ / \ (1 - d(r)) \tag{9}$$

7. 計算シミュレーション結果

3.125 Hz 未満の速さである条件で実験した場合，ヴィブ

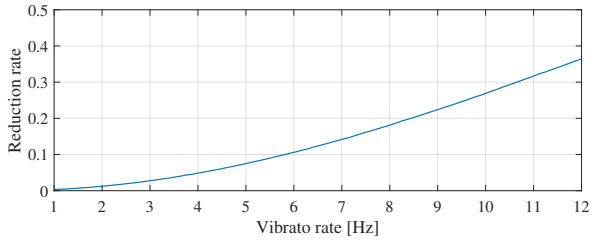

図 4　LPF によるヴィブラートの速さに対する深さの減少割合

表 6　計算シミュレーション結果

手法名	A	B	C	D	E	F	G	H
Gross error [%]	79.1	74.8	78.2	27.9	83.1	**19.0**	33.5	27.1
Fine rate error [%]	1.92	2.1	2.57	0.922	2.38	0.893	0.852	**0.682**
Fine extent error [%]	6.92	6.07	6.95	4.46	6.45	3.48	3.26	**3.08**
Start boundry error [s]	0.174	0.186	0.244	**0.127**	0.279	0.149	0.181	0.172
End boundry error [s]	0.141	**0.111**	0.144	0.177	0.118	0.167	0.194	0.244
Absence error [%]	15.4	18.5	15.4	3.85	12.3	5.38	3.08	**0**

ラート区間外をヴィブラートと誤判定するため，実験後のデータから取り除いた．その結果を**表 6** に示す．最も性能が良い数値をそれぞれ太字で表している．全ての手法で従来手法 A と比べて一部の評価指標で精度が向上していることを確認でき，3 つの本手法は精度向上に有効な手段であると考えられる．

7.1　Gross error の分析結果

Gross error では従来手法 A と比べ，手法 E 以外は精度が向上している．最も性能が高い手法は F であり，上位 3 手法に入る手法 D と H ともに，ヴィブラート判定区間の補正が行われている共通点がある．このことからヴィブラート判定区間の補正がヴィブラート判定の精度に特に有効であることが考えられる．

7.2　Fine rate error の分析結果

Fine rate error では従来手法 A と比べ，手法 B，C，E は精度が低下している．このことから速さの推定精度の向上では F_0 の分析シフト幅の変更と，F_0 に適用させる LPF と補正の実装の単体での実装と組合せる実装は効果がないと考えられる．最も性能が高い手法は H であり，上位 3 手法に入る手法 F と G ともに，ヴィブラート判定区間の補正が行われている共通点がある．このことからヴィブラート判定区間の補正が特に有効であることが考えられる．

7.3　Fine extent error の分析結果

Fine extent error では従来手法 A と比べ，手法 C 以外は精度が向上している．7.2 節と同様，最も性能が高い手法は H であり，上位 3 手法に入る手法 F と G ともに，ヴィブラート判定区間の補正が行われている共通点がある．このことからヴィブラート判定区間の補正が特に有効であることが考えられる．

7.4　Start boundry error の分析結果

Start boundry error では従来手法 A と比べ，手法 B，C，E と G は精度が低下している．このことから判定区間の開始地点の推定精度の向上では F_0 の分析シフト幅の変更と，F_0 に適用させる LPF と補正の実装の単体での実装と組合せる実装は効果がないと考えられる．最も性能が高い手法は D であり，ヴィブラート判定区間の補正が特に有効であることが考えられる．

7.5　End boundry error の分析結果

End boundry error では従来手法 A と比べ，手法 B と E は精度が向上している．最も性能が高い手法は B であり，判定区間の終了地点の推定精度の向上では F_0 の分析シフト幅の変更が有効であることが考えられる．

7.6　Absence error の分析結果

Absence error では従来手法 A と比べ，手法 B は精度が低下している．このことからヴィブラートが無いことを判定する精度の向上では F_0 の分析シフト幅の変更のみの実装は効果がないと考えられる．最も性能が高い手法は H であり，上位 3 手法に入る手法 D と G ともに，ヴィブラート判定区間の補正が行われている共通点がある．このことからヴィブラート判定区間の補正が特に有効であることが考えられる．

8.　考察

判定区間の補正によるヴィブラート判定，速さと深さの推定精度の向上では，今回のヴィブラートが定常波であることが挙げられる．判定区間を縮めることで，正しいパラメータの割合がより大きくなるので精度が向上したと考えられる．また，判定区間の開始地点の精度の向上では，ヴィブラート開始地点での細かい振動を誤判定していた区間を短縮することで，開始地点の真値に近づいたためと考えられる．Absence error ではヴィブラート無しをヴィブラートと判定したものは，音声の開始終了地点に現れる F_0 軌跡の不規則な変化をヴィブラートと判定していることが多かった．そのため区間の補正でその軌跡の不規則な変化の部分を判定区間から一部除去できたため，精度が向上したと考えられる．判定区間の終了地点の推定精度の低下では，全データで音声の終端が終了時間の真値だったため，縮めることで真値からさらに遠ざかったことが原因である．F_0 の分析シフト幅の変更によるヴィブラート判定精度の向上は，F_0 軌跡の極大点と極小点の推定誤差による影響を抑制できたことで，ヴィブラート判定のミスが減少したことが原因と考えられる．速さの推定精度と判定区間の開始地点の推定精度の低下では，F_0 の分析シフト幅を細かくすることにより，細かい振動が混入しやすくなったためと考えられる．

F_0 に適用させる LPF と補正の実装によるヴィブラート判定の低下は，LPF に対する深さの補正が原因と考えられる．深さの補正ではヴィブラート判定をして検出した深さに，補正をかけていた．しかし実際の判定した F_0 区間の全てがその深さに統一されているわけではないので，そこの差が精度の低下につながった．また，判定区間の開始終了地点の推定精度の低下では，LPF に対する時間の遅延の補正の誤差が原因と考えられる．他の 2 つの手法と組み合わせると精度が一部向上した理由については，それらの誤差の原因を 2 つの手法が緩和させたと考えられる．

9. おわりに

本研究では，高精度なヴィブラートの速さ, 深さの推定を目的としたヴィブラート区間検出手法の開発を行った．提案手法では，3 つの手法を組み合わせて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，3 つの改善手法の組み合わせに対する計算シミュレーションを行った．その際に，比較する提案手法と従来手法のパラメータの設定についても行った．計算シミュレーション結果から，従来手法と比べ全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．以上の結果より提案手法では，3 つの改善手法によってパラメータ推定の精度が向上する可能性が示唆された．

今後の課題としては，提案手法によって時間変動を持つヴィブラートの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挙げられる．本研究では, 定常波のヴィブラートを用いたが，人間が発するヴィブラートでは，速さと深さは時間振動することが先行研究から知られている [14]，[15]．またヴィブラート区間の補正が，補正値の算出に使用していないデータに対しても有効であるか検証を行う必要がある．

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] 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).
[3] Akagi, M. and Kitakaze, H.: Perception of Synthesized Singing Voices with Fine Fluctuations in Their Fundamental Frequency Contours, *in Proc. ICSLP 2000*, pp. 458–461 (2000).
[4] Kojima, K., Yanagida, M. and Nakayama, I.: Variability of Vibrato —A Comparative Study Between Japanese Traditional Singing and Bel Canto—, *in Proc. Speech Prosody 2004*, pp. 151–154 (2004).
[5] Nakayama，I.：Comparative Studies on Vocal Expressions in Japanese Traditional and Western Classical-Style Singing，Using a Common Verse，in Proc．ICA 2004，pp. 1295–1296 (2014).
[6] 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).
[7] 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, *in Proc. WASSPA 2007*, pp. 215–218 (2007).
[8] Kenmochi, H. and Ohshita, H.: VOCALOID - Commercial Singing Synthesizer Based on Sample Concatenation, *in Proc. INTERSPEECH 2007*, pp. 4010–4011 (2007).
[9] 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).
[10] 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).
[11] Morise, M., Yokomori, F. and Ozawa, K.: WORLD: A Vocoder-Based High-Quality Speech Synthesis System for Real-Time Applications, *IEICE Transactions on Information and Systems*, Vol. E99-D, No. 7, pp. 1877–1884 (2016).
[12] Morise, M.: D4C, a band-aperiodicity estimator for high-quality speech synthesis, *Speech Communication*, Vol. 84, pp. 57–65 (2016).
[13] Morise, M.: Harvest: A high-performance fundamental frequency estimator from speech signals, *in Proc. INTERSPEECH 2017*, pp. 2321–2325 (2017).
[14] Prame, E.: Measurements of The Vibrato Rate of Ten Singers, *STL–QPSR*, Vol. 33, No. 4, pp. 73–86 (1992).
[15] 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).