情報処理学会論文誌 Vol. 52 No. 5 1910–1922 (May 2011)

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

右田尚人[†1] 森勢将雅[†1] 西浦敬信[†1]

本論文では，プロ歌手による歌唱表現（ヴィブラートやポルタメント）の差異を分析するために構築した歌唱データベースの詳細と歌唱データベースを用いて分析したヴィブラートの個人性の制御に有効な特徴量について述べる．従来，高品質な歌唱合成を実現するために，様々な楽曲が収録されたデータベースを用いてヴィブラートが分析された．基本周波数（F0）軌跡におけるヴィブラートの速さと深さに関する特徴量が用いられ，楽曲の種類による差異が確認された．我々は，プロ歌手による歌唱表現の差異を分析するために，プロ歌手 4 名がヴィブラートやポルタメントを表現した歌唱を収録し，歌唱データベースを構築した．個人性の制御に有効な特徴量を抽出することができれば，様々なプロ歌手の F0 軌跡を制御することにより，旋律に応じた柔軟な歌唱合成が可能となる．そこで，歌唱データベースよりヴィブラートに関する従来の特徴量と我々の提案する特徴量を抽出し，特徴量の有効性を検討した．結果，これらの特徴量は歌手により異なり，ヴィブラートの個人性の制御に有効であることが示された．

Study of Effective Features for Controlling the Differences of Vibratos Among Singers by Utilizing Singing Database

Naoto Migita,[†1] Masanori Morise[†1] and Takanobu Nishiura[†1]

This paper describes the details of singing database for analyzing the differences of musical expressions (vibrato and portamento) among professional singers and the effective features for controlling the differences of vibratos. Vibratos were analyzed by utilizing database composed of various types of songs for synthesizing singing voices with high-quality. The features of fundamental frequency (F0) contours about the rate and the extent of vibrato were analyzed and the result suggested that they varied according to the types of songs. We designed singing database by recording the singing voices that four professional singers expressed vibrato and portamento for analyzing the differences of musical expressions among professional singers. We can synthesize natural singing voices flexibly by controlling F0 contours of various professional singers, provided that effective features for controlling the differences of musical expressions among professional singers are extracted. Then, we studied the effectiveness of conventional features and proposed features about vibrato extracted from singing database. The results suggested that the features were different by professional singers and effective for controlling vibratos.

1. はじめに

近年，楽曲制作において歌唱合成技術が注目され，YAMAHA の VOCALOID[1)] のような歌詞と楽譜の入力により歌唱を合成する技術が利用されている．VOCALOID は，歌唱ライブラリから歌詞と楽譜を基に音素片を抽出し，つなぎ合わせることにより歌唱を合成する．様々な歌手の歌唱を収録した歌唱ライブラリが存在し，ユーザは旋律に応じて異なる歌手の歌唱を合成することができる．また，楽譜により示される声の高さ（F0）の制御により，歌唱にヴィブラートなどの歌唱表現を付与し，より自然な歌唱を合成することができる．ユーザは，テンプレートで用意されているヴィブラートの F0 軌跡のパラメータ（振幅，周期，長さ）を手動で調節するため，初心者の場合，思いどおりのヴィブラートを合成することが難しい．そこで，旋律に応じて様々な種類の自然なヴィブラートを合成するために，ユーザの歌唱入力により合成された歌唱の音高や音量を自動編集する VocaListener[2)] が提案された．この技術では，目標とする歌唱としてユーザ自身の歌唱やプロ歌手の歌唱を入力し，VOCALOID により合成される歌唱に自動でヴィブラートなどの歌唱表現を付与する．よって，VocaListener では VOCALOID に入力する歌詞を歌った歌唱データが必要である．プロ歌手のヴィブラートをモデル化することができれば，ユーザは歌詞や楽譜に依存せず，合成した歌唱にプロ歌手のヴィブラートを付与できるはずである．

従来，高品質な歌唱合成を実現するために，歌唱の F0 軌跡が分析されている[3)]．特に，ヴィブラートの F0 を制御するために，ヴィブラートの速さと深さに関する特徴量が提案された[4)]．文献 5) では，邦楽や洋楽の様々な種類の楽曲が収録されたデータベース「日本語を歌・唄・謡う」[6)] を用いて速さと深さに関する特徴量が分析され，楽曲の種類により異なることが確認された．話声を歌唱に変換する歌唱合成システムである SingBySpeaking[7)] では，

†1 立命館大学
Ritsumeikan University

速さと深さに関する特徴量を用いたヴィブラートの F0 制御モデルが提案された．文献 5) において，ヴィブラートは時間とともに変動することが報告されたが，SingBySpeaking に用いられる F0 制御モデルは定常振動であり，ヴィブラートの時間変動は考慮されていない．

本論文で目的とするプロ歌手のヴィブラート F0 制御モデルの構築には，プロ歌手がヴィブラートを表現した大量の歌唱データを用いて，プロ歌手によるヴィブラートの差異を分析する必要がある．個人性の制御に有効な特徴量を用いた F0 制御モデルを構築することができれば，歌唱データより特徴量を抽出をすることで，プロ歌手のヴィブラートを制御することが可能となる．我々は，プロ歌手による歌唱表現の差異を分析するために，プロ歌手 4 名がヴィブラートやポルタメントを表現した歌唱を収録し，歌唱データベースを構築した．歌唱データベースには，プロ歌手が普通に歌った歌唱（通常歌唱）と特定歌手を物真似した歌唱（物真似歌唱）が収録された．プロ歌手間における特徴量の差異の分析により，ヴィブラートやポルタメントがプロ歌手間でどのように異なるか，通常歌唱と物真似歌唱における特徴量の差異の分析により，プロ歌手がどのようにヴィブラートやポルタメントを制御するかについて分析することが可能である．本論文では，ヴィブラートに着目し，歌唱データベースを用いてヴィブラートの個人性の制御に有効な特徴量を検討した．歌唱データベースに収録されたヴィブラート歌唱より，従来の特徴量と提案する特徴量を抽出し，ヴィブラートにおけるプロ歌手間の差異と通常歌唱と物真似歌唱の差異を分析した．さらに，差異を確認した特徴量を用いて従来のヴィブラートの F0 制御モデルを拡張し，評価実験により提案する特徴量の有効性を検証した．

2. ヴィブラートに関する従来研究

ヴィブラートとは，ある音の高さ・強さ・音色などを感覚的には一定に保ちながら周期的に変動させる歌唱技術である．従来，ヴィブラートの声の高さ（F0）の変動を制御するために，ヴィブラートの速さと深さに関する特徴量が提案された[4]．文献 5) では，自然性の高いヴィブラート制御法を検討するために，ヴィブラートの速さを示す vibrato rate や深さを示す vibrato extent がデータベース「日本語を歌・唄・謡う」[6] に収録されている洋楽（ソプラノ・テノール・バス・バリトン）と邦楽（演歌・長唄・民謡）の歌唱データを用いて分析され，歌唱法により異なることが報告された．全データの平均では，vibrato rate が 5.6 [Hz]，vibrato extent が 87 [cent]（ヴィブラートの基準となる F0 の 5.2 [%]）であった．そして，ヴィブラートの速さと深さを制御する F0 制御モデルが提案され，話声を歌唱に変換する歌唱合成システムである SingBySpeaking [7] においてヴィブラートの F0 制御に用いられている．

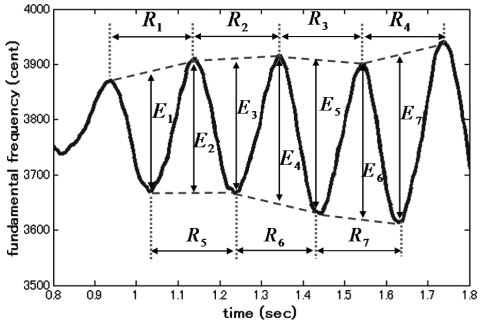

図 1　ヴィブラートの F0 軌跡
Fig. 1　F0 contour of a vibrato singing voice.

2.1 ヴィブラート特徴量

ヴィブラートの速さを示す vibrato rate と深さを示す vibrato extent は，ヴィブラート区間の F0 軌跡より算出される．文献 8) において，vibrato rate と vibrato extent は楽譜情報を用いずに歌唱力を自動で評価するために用いられ，式 (1)，(2) により算出された．R_n[sec]，E_n[cent] は，**図 1** に示すパラメータであり，これらのパラメータはヴィブラート区間の F0 軌跡より抽出される．ヴィブラート区間は F0 軌跡の 1 次差分の短時間フーリエ変換により得られるスペクトルを用いて抽出される．N は，ヴィブラート区間の F0 軌跡から抽出された各パラメータの総数を示しており，vibrato rate と vibrato extent はヴィブラート区間における平均値である．図 1 における F0 は，式 (3) により周波数 f_{Hz} を対数化した値 f_{cent} を示す．

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

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

$$f_{cent} = 1200\log_2\left(\frac{f_{Hz}}{261}\right) + 4800. \tag{3}$$

文献 9) において，プロのテノール歌手の場合，vibrato rate は一定ではなくヴィブラートの終端に向けて上昇する傾向が報告された．そこで，ヴィブラートの速さの時間変動に着

目し，ヴィブラートの始端と終端の速さの変化量が分析された[5]．データベース「日本語を歌・唄・謡う」[6] を分析した結果，洋楽では，ヴィブラートの速さは終端に向かって上昇する傾向が強く，邦楽では上昇する場合だけでなく下降する場合も多数確認された．上昇・下降それぞれの平均は，上昇率が 14 [%]（0.8 [Hz]），下降率が 8 [%]（0.5 [Hz]）であった．文献 10) では，10 名のソプラノ歌手のヴィブラートの深さの変化量も分析されており，17〜80 [cent] の範囲で変化していることが確認された．

2.2　ヴィブラート F0 制御モデル

話声を歌唱に変換する歌唱合成システムである SingBySpeaking [7] では，定常振動の F0 制御モデルが用いられ，そのモデルは式 (4) のようにパラメータ ω，k により表される．これらのパラメータは，データベース「日本語を歌・唄・謡う」[6] の歌唱データから抽出される F0 と F0 制御モデルにより合成される F0 の誤差が最小となるように非線形最小自乗法により決定される．vibrato extent は k/ω であり，vibrato rate を示す ω に応じて変化する．ヴィブラート F0 制御モデルと歌唱知覚の関係の分析では，自然なヴィブラートを合成するための vibrato rate は 6.3 [Hz]，vibrato extent は 68〜84 [cent]（ヴィブラートの基準となる F0 の 4〜5 [%]）と報告された[5]．

$$v_1(t) = \frac{k}{\omega}\sin(\omega t). \tag{4}$$

また，ヴィブラートの速さの時間変動を制御するモデルも提案されており，式 (5) の m（ヴィブラート区間長と速さの変化量を用いて算出される係数）により速さの時間変動を制御する．定常振動モデル同様に自然性を分析した結果，14 [%] 程度の上昇がヴィブラートの自然性を向上させることが示された[5]．

$$v_2(t) = \frac{k}{\omega}\sin(\omega t + \exp(mt)). \tag{5}$$

文献 5) では，様々な種類の楽曲の歌唱が収録されたデータベースを用いてヴィブラート特徴量が分析され，分析結果を基に F0 制御モデルが検討された．よって，旋律の変化が特徴量に影響を与えるため，プロ歌手によるヴィブラートの差異を分析することは困難であり，様々なプロ歌手のヴィブラートを高精度に制御することは不可能である．

3.　歌唱データベースの構築

旋律を歌った歌唱が収録されたデータベースを用いてヴィブラートを分析する場合，抽出するヴィブラート区間により音高，音量，音長などの条件が異なるため，旋律によるヴィブラートの変化が分析結果に影響を与えると考えられる．プロ歌手によるヴィブラートの差異を分析するには，複数名のプロ歌手がヴィブラートを表現した大量の歌唱データが必要である．文献 11) では，複数名のプロ歌手がヴィブラートを表現した歌唱が収録されている RWC 研究用音楽データベース[12] が用いられたが，歌唱の長さが様々で，周期的な変動を表現できていないヴィブラートが存在していることが分かった．

表 1　歌唱データベースの収録条件
Table 1　Recording conditions of singing database.

歌唱内容	単母音（/a/，/i/，/u/，/e/，/o/）
歌唱の長さ	2 [sec]
サンプリング周波数	96 [kHz]
量子化ビット数	24 [bit]
チャンネル数	モノラル
マイクロホン	NEUMANN U87Ai
場所	レコーディングスタジオ（NC-15）

我々はプロ歌手による歌唱表現の差異を分析するために，プロ歌手 4 名（女性 2 名，男性 2 名）が，旋律ではなくヴィブラートやポルタメントのみを表現した歌唱を収録し，歌唱データベースを構築した[13]．このデータベースには，プロ歌手が普通に歌った歌唱（通常歌唱）だけでなく特定のプロ歌手を物真似した歌唱（物真似歌唱）も収録されており，歌唱制御によるヴィブラートやポルタメントの変化を分析することが可能である．物真似対象歌手には，ポップス系と演歌系の代表的な歌手を 1 名ずつ選定した．ヴィブラートとは，ある音高を基準として周期的に変動させる歌唱表現であるため，基準となる音高として，各プロ歌手が得意とする声域（1 オクターブ）を収録した．一方，ポルタメントとは，ある音高から別の音高に移す際に，滑らかに音高を変動させる歌唱表現であるため，各プロ歌手が変動前の基準となる音高を選択し，その音高から別の音高（±1 オクターブ）まで，上昇する場合と下降する場合を収録した．さらに全条件に対してヴィブラートやポルタメントを表現していない歌唱も収録した．よって，歌唱データベースに収録されたヴィブラートは，歌手 4 名，5 母音，13 音階，物真似の有無，ヴィブラートの有無の計 1,040 データ，ポルタメントは，歌手 4 名，5 母音，24 音階（上昇：12 音階，下降：12 音階），物真似の有無，ポルタメントの有無の計 1,920 データである．この歌唱データベースの詳細を**表 1** と**表 2** 示す．収録は，NC 値が NC-15 のレコーディングスタジオにおいて行われた．

歌唱データベースには，歌唱内容として単母音を収録しており，旋律に依存しないプロ歌手自身の歌唱表現を分析することができる．また，通常歌唱だけでなく物真似歌唱も収録し

表 2　歌唱データベースの構成
Table 2　Composition of singing database.

	プロ歌手名	物真似対象歌手名	ヴィブラート範囲	ポルタメント範囲（基準）
女性 1	荒牧陽子	宇多田ヒカル	B3. . .B4	B2. . .B4（B3）
女性 2	千田かおり	美空ひばり	C3. . .C4	C2. . .C4（C3）
男性 1	風雅なおと	GACKT	D3. . .D4	G2. . .G4（G3）
男性 2	西一男	五木ひろし	E3. . .E4	E2. . .E4（E3）

ており，プロ歌手がどのようにヴィブラートやポルタメントを制御するのか分析することができる．

3.1　データベースの有効性の検証

歌唱データベースには，ポップス系と演歌系の代表的な歌手を物真似した歌唱が収録されている．本論文ではヴィブラートに着目し，プロ歌手のヴィブラートの F0 を制御するモデルの構築を目的とした．そこで，収録したヴィブラートの物真似歌唱が，物真似対象歌手の F0 を制御するモデルの構築に有効かどうかを検証するために主観評価実験を行った．評価法として，複数の歌唱に対して 1 から 5 の 5 段階で評価し，それらの平均を結果とする MOS（Mean Opinion Score）を用いた．正常な聴力を有する成人 10 名（女性 5 名，男性 5 名）の被験者に収録したヴィブラート歌唱を呈示し，5 段階評定尺度（5：似ている，4：少し似ている，3：どちらともいえない，2：あまり似ていない，1：似ていない）を用いて，どれぐらい物真似対象歌手の歌唱に似ているかを評価させた．今回，母音の差異には着目しないため，評価用のヴィブラート歌唱として母音/a/のみを用い，音階は各歌手異なる 3 音階（収録した最も低い音階から 1 度，3 度，5 度）とした．騒音レベルが 20.1 [dBA] の防音室で評価実験を行い，被験者にはヘッドホン（SONY MDR-CD900ST）を用いてヴィブラート歌唱 24 データ（プロ歌手 4 名，物真似の有無，3 音階，1 母音/a/）を呈示した．また，呈示順による影響を考慮して各歌手の 6 データ（物真似の有無，3 音階）をランダムに呈示した．これらのヴィブラート歌唱を評価する前に，被験者に対して各物真似対象歌手の代表曲（宇多田ヒカル：First Love，美空ひばり：川の流れのように，GACKT：Vanilla，五木ひろし：契り）のサビ部分を呈示し，その歌唱を基準にヴィブラート歌唱を評価するように指示した．

図 2 は評価実験結果を示し，横軸は評価対象の歌唱，縦軸は MOS による評価結果，エラーバーは標準偏差を示す．t 検定[14] による有意差検定（有意水準：0.05）を行った結果，女性 1，男性 1，男性 2 の場合，通常歌唱と物真似歌唱の間に有意な差が存在した．この結果は，通常の歌唱法と物真似の歌唱法が異なることを意味する．また，物真似歌唱の MOS 値は 4 前後であり，通常歌唱よりも物真似対象に似ていることが分かった．一方，女性 2 の MOS 値において，通常歌唱と物真似歌唱の差は小さく，t 検定の結果においても有意な差は存在しなかった．女性 2 の通常歌唱は物真似対象歌手の歌唱に似ており，物真似による歌唱法の変化が小さいことが分かった．以上より，収録した 4 種類の物真似歌唱は物真似対象歌手の歌唱と似ており，この歌唱データベースを用いることで物真似対象歌手のヴィブラートを分析することが可能となる．つまり，歌唱データベースは物真似対象歌手の F0 を制御するモデルの構築に有効であると考えられる．

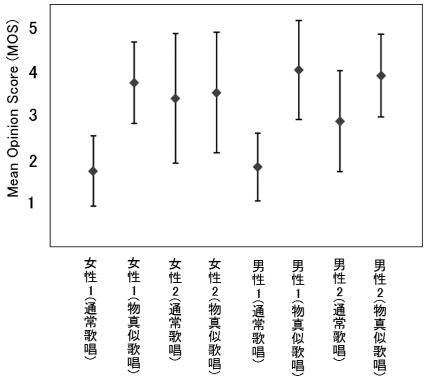

図 2　歌唱データベースの評価実験結果（MOS）
Fig. 2　Results of evaluation experiment about singing database (MOS).

4. ヴィブラート特徴量の提案

SingBySpeaking[7] では，ヴィブラート区間の平均である vibrato rate と vibrato extent を用いたヴィブラート F0 制御モデルが提案された．また，ヴィブラートの速さは時間とともに変動することが報告され，速さの時間変動を制御するモデル[5] も提案された．文献 10) では，ヴィブラートの速さだけでなく深さも時間とともに変動することが報告された．よって，歌手によるヴィブラートの差異を高精度に制御するには，ヴィブラートの時間変動を分析する必要がある．我々は，ヴィブラートの時間変動を分析するために，ヴィブラートの速さと深さに関する新たな特徴量を提案する．また，ヴィブラートの速さや深さだけではなく，ヴィブラート区間と歌唱区間の関係に着目し，ヴィブラートの長さに関する特徴量を提

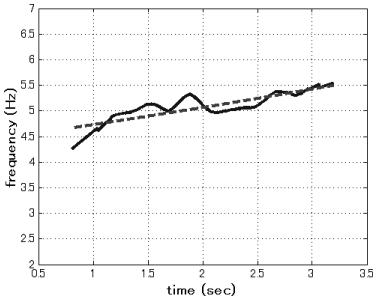

図 3　速さの軌跡（実線）と近似曲線（点線）
Fig. 3　Time fluctuation of vibrato rate (solid line) and approximated curve (dotted line).

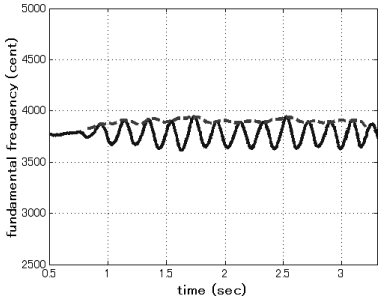

図 4　F0 軌跡（実線）と深さの軌跡（点線）
Fig. 4　F0 contour (solid line) and time fluctuation of vibrato extent (dotted line).

案する．

4.1　速さの時間変動に関する特徴量（vibrato rate's fluctuation）

我々はヴィブラート区間長に対する速さの変化量に着目し，速さの時間変動に関する特徴量（以下 vibrato rate's fluctuation）を提案する．**図 3** において，実線はヴィブラートの F0 軌跡から従来法[15] を用いて抽出したヴィブラートの速さの軌跡を示し，点線は速さの軌跡から最小自乗法により近似した曲線を示す．従来の F0 制御モデル[5] で，速さの時間変動を制御するために指数関数が用いられた．本論文でも，ヴィブラート区間長に対する速さの変化量を分析するために，近似曲線として式 (6) に示す指数関数を用い，係数 β を vibrato rate's fluctuation と定義する．α は，変動前のヴィブラートの速さを示す．

$$r(t) = \alpha \exp(\beta t). \tag{6}$$

4.2　深さの時間変動に関する特徴量（vibrato extent's fluctuation）

図 4 において実線は，あるヴィブラートの F0 軌跡を示し，点線は F0 軌跡の瞬時振幅を示す．瞬時振幅とは，F0 軌跡である $f(t)$ をヒルベルト変換した $f_a(t)$ の絶対値で定義され，式 (7)，(8)，(9) により算出される．j は虚数単位であり，$f_h(t)$ は $f_a(t)$ の虚部を示す．また，$IDFT$ は逆離散フーリエ変換を示しており，ω は角周波数，$F(\omega)$ は $f(t)$ のスペクトルである．点線の瞬時振幅軌跡より，ヴィブラートの深さは時間とともに変動することが分かる．そこで，本論文では瞬時振幅軌跡の標準偏差をヴィブラートの深さの時間変動に関する特徴量（以下 vibrato extent's fluctuation）と定義する．

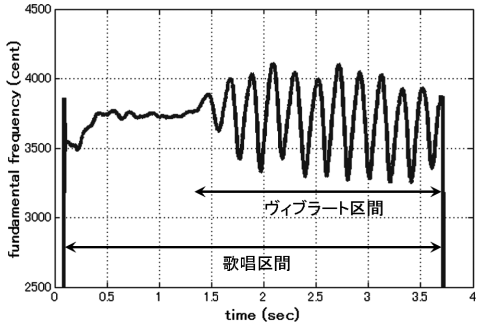

図 5　歌唱区間とヴィブラート区間
Fig. 5　Singing section and vibrato section.

$$f_a(t) = f(t) + j f_h(t), \tag{7}$$

$$f_h = IDFT(F_h(\omega)), \tag{8}$$

$$F_h(\omega) = \begin{cases} -jF(\omega), \omega > 0, \\ jF(\omega), \omega < 0. \end{cases} \tag{9}$$

4.3　長さに関する特徴量（vibrato duration）

ヴィブラートは，主に声を伸ばす際に用いられ，旋律に応じてヴィブラート区間の長さは制御される．**図 5** は，あるプロ歌手が単母音/a/を歌ったヴィブラートの F0 軌跡であり，ヴィブラート区間と歌唱区間の時間長が大きく異なる．つまり，この歌手は旋律を歌っていない場合でも，ヴィブラートの開始時刻を制御していると考えられる．そこで，歌唱区間内に占めるヴィブラート区間の割合を vibrato duration と定義する．

5.　ヴィブラート特徴量の分析

我々は，ヴィブラートの個人性の制御に有効な特徴量を検討するために，歌唱データベースに収録されたヴィブラート歌唱より，従来の特徴量と提案する特徴量を抽出し，ヒストグラムを用いてプロ歌手による差異と物真似による差異を分析した．データベースの有効性の検証により，収録された女性 1，男性 1，男性 2 の物真似した際の歌唱法は，通常の歌唱法と異なることが分かった．そこで，物真似による差異，つまり歌唱法の変化の分析により，プロ歌手がどのようにヴィブラートを制御するかについて検討する．STRAIGHT[16] を用いて，歌唱データベースに収録されたヴィブラート歌唱 520 データ（4 名の歌手，物真似

表 3　各特徴量の平均
Table 3　Averages of each feature.

歌唱	vibrato rate [Hz]	vibrato extent [cent]	vibrato rate's fluctuation β	vibrato extent's fluctuation [cent]	vibrato duration [%]
女性 1 通常歌唱	5.33	54	0.003	16	79
女性 1 物真似歌唱	5.70	48	−0.006	17	81
女性 2 通常歌唱	3.94	133	−0.022	38	62
女性 2 物真似歌唱	3.90	145	0.021	41	55
男性 1 通常歌唱	4.85	98	0.022	27	80
男性 1 物真似歌唱	4.82	291	0.039	70	62
男性 2 通常歌唱	5.36	68	0.030	29	71
男性 2 物真似歌唱	5.39	109	0.014	33	85

の有無，5 母音，13 音階）の F0 軌跡を推定し，以下に示す方法で各特徴量を自動で抽出した．抽出した 520 データの各特徴量の平均を**表 3**，各特徴量のヒストグラムを**図 6**，**図 7**，**図 8**，**図 9**，**図 10** に示す．

(1)　STRAIGHT[16] により推定した F0 軌跡において，F0 が 3,000〜6,500 [cent] の区間を抽出し，その区間を歌唱区間とする．この F0 範囲は，収録されたヴィブラート歌唱の全音階を含む範囲である．本論文では，1 [msec] 間隔で F0 を推定する．

(2)　歌唱区間の F0 軌跡より，従来のヴィブラート区間を抽出する手法[8] を用いてヴィブラート区間を抽出する．

(3)　歌唱区間とヴィブラート区間の長さより vibrato duration を算出する．
カットオフ周波数が 10 [Hz] の LPF（Low Pass Filter）を畳み込んだヴィブラート区間の F0 軌跡より，1 次微分が 0 となる時刻を抽出し，図 1 の R_n と E_n を用いて vibrato rate と vibrato extent を算出する．
LPF 処理前のヴィブラート区間の F0 軌跡より，4 章で述べた手法を用いて vibrato rate's fluctuation と vibrato extent's fluctuation を算出する．

また，t 検定[14] を用いて各特徴量分布における通常歌唱と物真似歌唱の間の有意差を検定し，多重検定法である Tukey の方法[14] を用いてプロ歌手間の有意差を検定した．Tukey の方法により，4 名の歌手間の組合せ ${}_4C_2 = 6$ パターン（女性 1 と女性 2 の間，女性 1 と男性 1 の間，女性 1 と男性 2 の間，女性 2 と男性 1 の間，女性 2 と男性 2 の間，男性 1 と男性 2 の間）の有意差を検定する．t 検定と Tukey の方法による多重検定の有意水準は 0.05 とした．t 検定の結果を**表 4** に示し，Tukey の方法を用いて通常歌唱のプロ歌手間の有意差を検定した結果を**表 5**，物真似歌唱のプロ歌手間の有意差を検定した結果を**表 6** に示す．

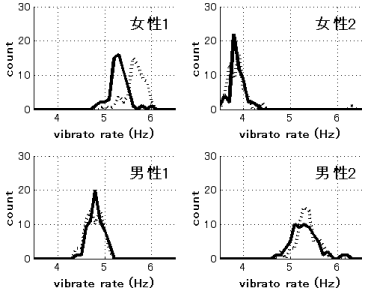

図 6　vibrato rate のヒストグラム（実線：通常歌唱，点線：物真似歌唱）
Fig. 6　Histograms of vibrato rates (Solid lines: normal voices, dotted lines: imitation voices).

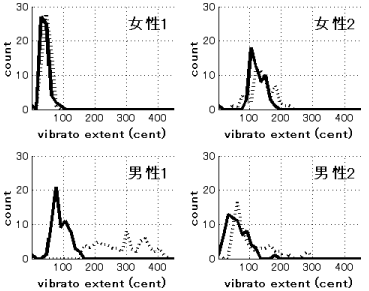

図 7　vibrato extent のヒストグラム（実線：通常歌唱，点線：物真似歌唱）
Fig. 7　Histograms of vibrato extents (Solid lines: normal voices, dotted lines: imitation voices).

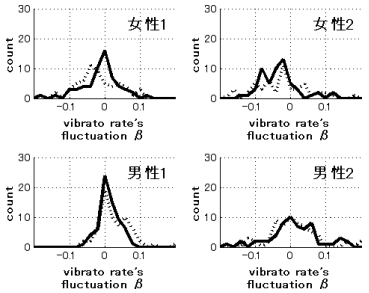

図 8　vibrato rate's fluctuation のヒストグラム（実線：通常歌唱，点線：物真似歌唱）
Fig. 8　Histograms of vibrato rate's fluctuations (Solid lines: normal voices, dotted lines: imitation voices).

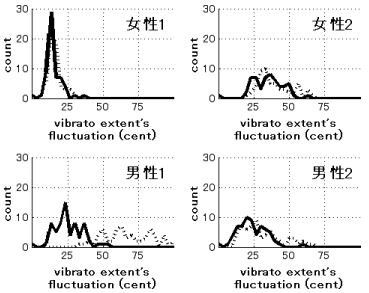

図 9　vibrato extent's fluctuation のヒストグラム（実線：通常歌唱，点線：物真似歌唱）
Fig. 9　Histograms of vibrato extent's fluctuations (Solid lines: normal voices, dotted lines: imitation voices).

5.1　Vibrato rate の分析結果

通常歌唱の vibrato rate の平均は，女性 1 が 5.33 [Hz]，女性 2 が 3.94 [Hz]，男性 1 が 4.85 [Hz]，男性 2 が 5.36 [Hz] であり，多重検定の結果である表 5 より，プロ歌手間の有意差を確認できる．図 6 より，6.0 [Hz] を超える vibrato rate は少なく，ヴィブラートの自然性の研究[5] で示された 6.3 [Hz] に比べ低いことが分かった．通常歌唱の分布と物真似歌唱の

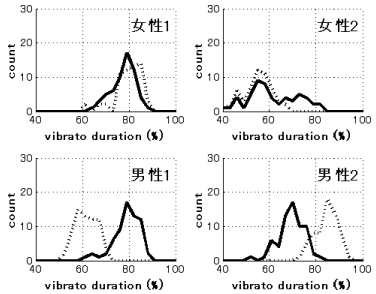

図 10　vibrato duration のヒストグラム（実線：通常歌唱，点線：物真似歌唱）
Fig. 10　Histograms of vibrato durations (Solid lines: normal voices, dotted lines: imitation voices).

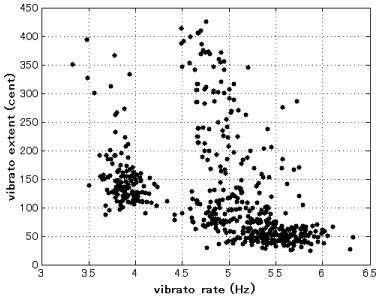

図 11　vibrato rate と vibrato extent の関係
Fig. 11　Relationship between vibrato rates and vibrato extents.

分布を比較すると，女性 1 の場合，通常歌唱に比べ物真似歌唱の vibrato rate は高い．表 4 に示す t 検定の結果においても，通常歌唱と物真似歌唱の間に有意差が存在した．よって，女性 1 は vibrato rate の制御により宇多田ヒカルに似たヴィブラートを表現していたことが分かる．また，表 6 より，vibrato rate は物真似歌唱のプロ歌手間でも異なり，すべての組合せにおいて有意差が存在した．

5.2　Vibrato extent の分析結果

通常歌唱の vibrato extent の平均は，女性 1 が 54 [cent]，女性 2 が 133 [cent]，男性 1 が 98 [cent]，男性 2 が 68 [cent] であり，多重検定の結果より，プロ歌手間の有意差を確認できる．4 名の vibrato extent の平均は 88 [cent] であり，様々な歌唱法のデータベースの分析[5] で示された 87 [cent] に近い値であった．ただし，図 7 に示すように自然性の研究[5] で示された 68〜84 [cent] の範囲外の vibrato extent が多数存在していた．通常歌唱の分布と物真似歌唱の分布を比較すると，男性 1 の場合，通常歌唱に比べ物真似歌唱の vibrato extent は大きいが，ばらつきが大きく不安定であった．t 検定の結果では，男性 1 以外に男性 2 の場合でも有意差が存在した．また，物真似歌唱におけるプロ歌手間の多重検定を行った結果，すべての組合せで有意差が存在した．

5.3　Vibrato rate's fluctuation の分析結果

図 8 に示すように，すべての歌手において，vibrato rate's fluctuation が正の値と負の値，つまり時間とともに上昇する場合と下降する場合が存在した．ただし，通常歌唱において女性 2 の平均値は −0.022，男性 1 の平均値は 0.022，男性 2 の平均値は 0.030 であり，上昇または下降の傾向を確認した．プロ歌手間の有意差検定を行った結果，女性 2 と男性 1 の間，女性 2 と男性 2 の間に有意差が存在した．しかし，通常歌唱と物真似歌唱を比較すると，vibrato rate's fluctuation において vibrato rate のような顕著な差異は存在せず．t 検定でもすべての組合せにおいて有意差は存在しなかった．

表 4　通常歌唱と物真似歌唱の t 検定結果
Table 4　Results of t-test between normal voices and imitation voices.

	vibrato rate	vibrato extent	vibrato rate's fluctuation	vibrato extent's fluctuation	vibrato duration
女性 1 通常歌唱，女性 1 物真似歌唱	＊	−	−	−	−
女性 2 通常歌唱，女性 2 物真似歌唱	−	−	−	−	＊
男性 1 通常歌唱，男性 1 物真似歌唱	−	＊	−	＊	＊
男性 2 通常歌唱，男性 2 物真似歌唱	−	＊	−	−	＊

＊：有意差あり，−：有意差なし

表 5　通常歌唱の多重比較検定結果（vibrato rate，vibrato extent，vibrato rate's fluctuation，vibrato extent's fluctuation and vibrato duration）
Table 5　Results of multiple comparisons of normal voices (vibrato rate，vibrato extent，vibrato rate's fluctuation，vibrato extent's fluctuation and vibrato duration).

	女性 2 通常歌唱	男性 1 通常歌唱	男性 2 通常歌唱
女性 1 通常歌唱	(＊，＊，−，＊，＊)	(＊，＊，−，＊，−)	(−，＊，−，＊，＊)
女性 2 通常歌唱		(＊，＊，＊，＊，＊)	(＊，＊，＊，＊，＊)
男性 1 通常歌唱			(＊，＊，−，−，＊)

＊：有意差あり，−：有意差なし

表 6　物真似歌唱の多重比較検定結果（vibrato rate，vibrato extent，vibrato rate's fluctuation，vibrato extent's fluctuation and vibrato duration）
Table 6　Results of multiple comparisons of imitation voices (vibrato rate，vibrato extent，vibrato rate's fluctuation，vibrato extent's fluctuation and vibrato duration).

	女性 2 物真似歌唱	男性 1 物真似歌唱	男性 2 物真似歌唱
女性 1 物真似歌唱	(＊，＊，−，＊，＊)	(＊，＊，−，＊，＊)	(＊，＊，−，＊，＊)
女性 2 物真似歌唱		(＊，＊，−，＊，＊)	(＊，＊，−，＊，＊)
男性 1 物真似歌唱			(＊，＊，−，＊，＊)

＊：有意差あり，−：有意差なし

5.4　Vibrato extent's fluctuation の分析結果

図 7 と図 9 を比較すると，通常歌唱においてプロ歌手間の関係が類似しており，さらに通

常歌唱と物真似歌唱の関係も類似していた．男性 1 の物真似歌唱の平均は，vibrato extent 同様に通常歌唱の平均に比べ大きい．これは，t 検定の結果からも有意差を確認することができた．よって，vibrato extent's fluctuation は vibrato extent に依存し，ヴィブラートの揺れ幅が大きいほど時間変動も大きい．vibrato extent と vibrato extent's fluctuation の結果より，男性 1 は，GACKT に似たヴィブラートを表現するために，ヴィブラートの深さを意識的に制御していたと考えられる．また，物真似歌唱におけるプロ歌手間の多重検定を行った結果，すべての組合せで有意差が存在し，物真似間で異なることが分かった．

5.5 Vibrato duration の分析結果

通常歌唱の vibrato duration の平均は，女性 1 が 79 [%]，女性 2 が 62 [%]，男性 1 が 80 [%]，男性 2 が 71 [%] であり，他の特徴量同様に歌手による差異を確認できる．多重検定の結果では，女性 1 と男性 1 以外の組合せで有意差が存在した．また，図 10 に示すように，男性 1 と男性 2 の場合，通常歌唱の分布と物真似歌唱の分布は明確に異なる．男性 1 は，ヴィブラート開始時刻を遅らせることにより GACKT に似たヴィブラートを表現し，男性 2 は，早めることにより五木ひろしに似たヴィブラートを表現していたと考えられる．t 検定の結果では，男性 1 や男性 2 だけでなく女性 2 の場合も有意な差が存在した．また，物真似歌唱におけるプロ歌手間の有意差検定を行った結果，すべての組合せで有意差が存在した．

5.6 考　　察

ヴィブラート特徴量を分析した結果，全特徴量においてプロ歌手による差異を確認した．また，vibrato rate，vibrato extent，vibrato extent's fluctuation，vibrato duration では通常歌唱の分布と物真似歌唱の分布が異なる場合が存在し，プロ歌手がこれらの特徴量を制御していることが分かった．女性 1 の場合は vibrato rate を，女性 2 の場合は vibrato duration を，男性 1 の場合は vibrato extent，vibrato extent's fluctuation や vibrato duration を，男性 2 の場合は vibrato extent，vibrato duration を意識的に制御することにより，特定のプロ歌手のヴィブラートを表現していた．一方，vibrato rate's fluctuation において，歌手による差異は存在したが，物真似による差異は確認されず，歌唱データベースに収録した 4 名は，物真似の際，vibrato rate's fluctuation を制御しなかったと考えられる．

図 11 は 520 データの vibrato rate と vibrato extent の関係を示しており，従来の F0 制御モデルのように vibrato rate が高くなるほど，vibrato extent が小さくなる傾向がある．しかし，vibrato rate が 4.5〜5.0 [Hz] 付近では，4.0 [Hz] に比べ vibrato extent が大きい場合も多く，歌唱データベースに収録された全プロ歌手のヴィブラートを制御するには従来の F0 制御モデルを拡張する必要がある．

6. F0 制御モデルの拡張と評価

話声を歌唱に変換する SingBySpeaking [7] のヴィブラート F0 制御モデルでは，速さと深さを考慮した定常振動モデルが用いられた．しかし，実際のヴィブラートでは，速さと深さは時間とともに変動し，非定常である．そこで，式 (5) に示すヴィブラートの速さの時間変動を考慮したモデルが提案され，自然性の高いヴィブラートの制御法が検討された．また，ヴィブラートの速さと深さの時間変動を考慮した F0 制御モデル[17] が提案されたが，全極モデルを用いたフレーム処理により F0 を制御するため，パラメータの数が多い．モデルの評価では，フレーム長が 250 [msec]，フレームシフトが 100 [msec]，伝達関数の次数が 3 であった．本論文では速さと深さの時間変動や長さに関する特徴量を提案し，歌唱データベースを用いた分析の結果，プロ歌手による特徴量の差異と物真似による特徴量の差異を確認した．そこで，提案する特徴量を用いて式 (5) の従来モデルを拡張し，6 種類のパラメータでヴィブラートの F0 を制御するモデルを提案する．提案するモデルは，歌唱データベースに収録されたヴィブラート歌唱の F0 軌跡を制御するためのモデルであり，収録されたヴィブラートなしの歌唱にヴィブラートを付与する目的で構築された．以下に，提案モデルの詳細と提案モデルの有効性を検証するために行った評価実験の結果を示す．

6.1 深さの時間変動と長さを考慮した F0 制御モデル

自然性の高いヴィブラートを合成するためにヴィブラートの速さの時間変動を考慮した F0 制御モデルが提案された[5]．我々は，vibrato extent's fluctuation と vibrato duration を用いて式 (5) の従来モデルを拡張し，深さの時間変動と長さの制御を可能にする．深さの時間変動は，正弦波を用いて表現され，長さはヴィブラート軌跡に遅延を加えることにより表現される．提案するモデルは式 (10)〜式 (14) により示され，歌唱データベースのヴィブラート歌唱から抽出される 6 種類のパラメータ（$v_1 \dots v_6$）を用いてヴィブラートの F0 軌跡 $v(t)$ を合成する．v_r は，vibrato rate である v_1 と vibrato rate's fluctuation である v_2 を用いて算出される速さの時間変動の軌跡であり，v_e は，vibrato extent である v_3 と vibrato extent's fluctuation である v_4 を用いて算出される深さの時間変動の軌跡である．式 (11) の v_5 は，ヴィブラート区間全体における F0 軌跡の瞬時振幅のフーリエ変換により得られる振幅スペクトルの最低次ピークに対応する周波数を示す．v_d は，vibrato duration である v_6 と歌唱区間の長さ T を用いて算出される遅延時間を示す．式 (13) に v_r，v_e，v_d を代入し，ヴィブラートの F0 軌跡 $v(t)$ を算出する．そして，歌唱データベースに収録さ

れたヴィブラートなしの歌唱の F0 軌跡 $f(t)$ に，ヴィブラートの F0 軌跡 $v(t)$ を加算（単位：cent）し，歌唱にヴィブラートを付与する．式 (14) において $f'(t)$ は，ヴィブラートを付与された F0 軌跡を示す．本論文では，STRAIGHT [16] を用いて歌唱の時間波形より F0 軌跡 $f(t)$ を推定する．5 章の特徴量抽出と同様に，1 [msec] 間隔で F0 軌跡を推定する．

$$v_r(t) = v_1 t + \exp(v_2 t) - 1, \tag{10}$$

$$v_e(t) = v_3 + v_4 \sin(2\pi v_5 t), \tag{11}$$

$$v_d = \frac{T(100 - v_6)}{100}, \tag{12}$$

$$v(t) = v_e(t - v_d)\sin(2\pi v_r(t - v_d)), \tag{13}$$

$$f'(t) = \begin{cases} f(t), t < v_d, \\ f(t) + v(t), t \geq v_d. \end{cases} \tag{14}$$

6.2　客観評価実験

歌唱データベースには，8 種類の声色（プロ歌手 4 名，物真似の有無）につき，65 パターン（5 母音，13 音階）のヴィブラートを表現した歌唱が収録された．この 520 データを用いて，従来モデルと提案モデルにより合成される F0 軌跡を比較し，提案モデルの有効性を検証した．従来モデルとして式 (5) に示す 3 種類のパラメータに基づくモデルを用い，各パラメータ（速さ ω，速さの時間変動 m，深さ k/ω）は，提案モデルの v_1，v_2，v_3 とした．各データから特徴量を抽出した後に，従来モデル，提案モデルにより F0 軌跡を合成し，歌唱データベースに収録されたヴィブラートなしの歌唱の F0 軌跡に付与した．客観評価実験では，**図 12** に示す，収録されたヴィブラート歌唱から抽出された特徴量 $a_{x,y,z}$，従来モデルにより合成された F0 軌跡から抽出された特徴量 $b_{x,y,z}$，提案モデルにより合成された F0 軌跡から抽出された特徴量 $c_{x,y,z}$ を，以下の式に代入した D_b と D_c を用いた．各特徴量は，5 章に示した特徴量抽出と同様の流れで，自動的に抽出された．

$$D_{b_{x,z}} = \frac{1}{65}\sum_{y=1}^{65} |a_{x,y,z} - b_{x,y,z}|, \tag{15}$$

$$D_{c_{x,z}} = \frac{1}{65}\sum_{y=1}^{65} |a_{x,y,z} - c_{x,y,z}|. \tag{16}$$

x は 8 種類の声色（プロ歌手 4 名，物真似の有無），y は 65 種類の発声パターン（5 母音，13 音階），z は 5 種類の特徴量（vibrato rate，vibrato extent，vibrato rate's fluctuation，vibrato extent's fluctuation，vibrato duration）を示す．

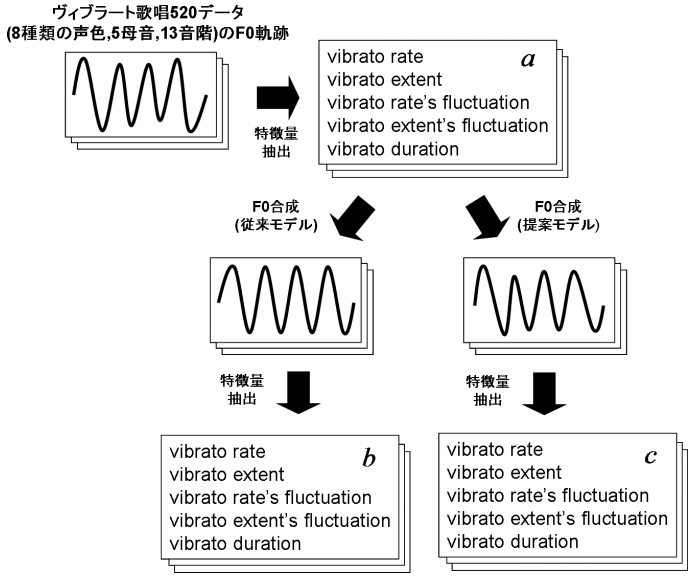

図 12　F0 制御モデルの評価実験
Fig. 12　Evaluation experiments of F0 models.

6.2.1　客観評価実験結果

表 7，**表 8**，**表 9**，**表 10** は，各プロ歌手の通常歌唱と物真似歌唱の評価実験結果 D_b，D_c を示す．D_b は従来モデルにより合成された F0 軌跡と収録されたヴィブラート歌唱の F0 軌跡の距離，D_c は提案モデルにより合成された F0 軌跡と収録されたヴィブラート歌唱の F0 軌跡の距離を示し，距離が小さいほど高精度にヴィブラートを制御できていることを意味する．提案する特徴量である vibrato extent's fluctuation と vibrato duration の実験結果に着目すると，すべての歌唱法（プロ歌手 4 名の通常歌唱と物真似歌唱）において，D_c は D_b に比べ小さい値であり，従来モデルより提案モデルの方が高精度にヴィブラートの深さの時間変動と長さを制御できている．5 章の分析結果において，vibrato extent's fluctuation の値が大きい男性 1 の物真似歌唱の場合，D_b は 65.8 [cent]，D_c は 19.6 [cent] であり，提案モデルの距離は従来モデルの 30 [%] 以下である．また，5 章の分析結果において，vibrato duration が低くなる傾向が強い女性 2 の場合，通常歌唱では従来モデルに比べ提案モデルは 32.8 小さく，物真似の歌唱においても 36.9 小さい値となり，vibrato duration の有効性を

表 7　男性 1 の実験結果（上段：通常歌唱，下段：物真似歌唱）
Table 7　Experimental results of male 1 (upper: normal voices, lower: imitation voices).

	vibrato rate [Hz]	vibrato extent [cent]	vibrato rate's fluctuation β	vibrato extent's fluctuation [cent]	vibrato duration [%]
D_b	0.029	**0.01**	0.028	26.2	18.3
D_c	**0.027**	1.78	**0.027**	**11.5**	**2.0**
	vibrato rate [Hz]	vibrato extent [cent]	vibrato rate's fluctuation β	vibrato extent's fluctuation [cent]	vibrato duration [%]
D_b	0.048	**0.04**	0.043	65.8	35.5
D_c	**0.045**	3.59	**0.041**	**19.6**	**2.7**

表 8　男性 2 の実験結果（上段：通常歌唱，下段：物真似歌唱）
Table 8　Experimental results of male 2 (upper: normal voices, lower: imitation voices).

	vibrato rate [Hz]	vibrato extent [cent]	vibrato rate's fluctuation β	vibrato extent's fluctuation [cent]	vibrato duration [%]
D_b	0.098	**0.01**	**0.081**	27.7	26.9
D_c	**0.051**	2.46	0.088	**11.0**	**6.7**
	vibrato rate [Hz]	vibrato extent [cent]	vibrato rate's fluctuation β	vibrato extent's fluctuation [cent]	vibrato duration [%]
D_b	**0.06**	**0.02**	**0.063**	31.0	12.5
D_c	0.11	2.13	0.079	**10.9**	**2.6**

表 9　女性 1 の実験結果（上段：通常歌唱，下段：物真似歌唱）
Table 9　Experimental results of female 1 (upper: normal voices, lower: imitation voices).

	vibrato rate [Hz]	vibrato extent [cent]	vibrato rate's fluctuation β	vibrato extent's fluctuation [cent]	vibrato duration [%]
D_b	**0.04**	**0.44**	**0.044**	15.9	18.8
D_c	0.24	1.42	0.059	**6.0**	**7.5**
	vibrato rate [Hz]	vibrato extent [cent]	vibrato rate's fluctuation β	vibrato extent's fluctuation [cent]	vibrato duration [%]
D_b	**0.05**	**0.01**	**0.053**	16.8	16.9
D_c	0.06	1.38	0.072	**6.0**	**5.9**

確認できる．ただし，vibrato rate，vibrato extent，vibrato rate's fluctuation の評価実験結果では，従来モデルよりも提案モデルの方が距離が大きい場合が多数存在した．提案モデルは時間変動を考慮した F0 制御モデルであり，ヴィブラート区間の平均値である vibrato rate や vibrato extent では，定常振動である従来モデルに比べ，ヴィブラート歌唱との距離は大きくなると考えられる．

6.3　主観評価実験

提案したヴィブラート特徴量が個人性の制御に有効であるかを検証するために，2 種類の主観評価実験を行った．1 つ目の実験では，式 (5) の従来モデルと式 (10)〜式 (14) の提案モデルにより合成されたヴィブラート歌唱を比較し，提案するモデルが従来モデルよりも高精度なヴィブラート歌唱を制御可能であるかを検証した．従来モデルに用いるパラメータ（速さ ω，速さの時間変動 m，深さ k/ω）には，提案モデルの v_1，v_2，v_3 を用いた．2 つ目の実験では，提案モデルにより合成された 4 名のヴィブラート歌唱を比較し，歌手間の差異を知覚可能であるかを検証した．これらの実験には，歌唱データベースに収録されたヴィブラートありの歌唱とヴィブラートなしの歌唱を用いた．合成されたヴィブラート歌唱とは，ヴィブラートありの歌唱から抽出されるパラメータ（$v_1 \ldots v_6$）を用いてモデルにより合成されるヴィブラートの F0 軌跡を，STRAIGHT [16] により推定したヴィブラートなしの歌唱の F0 に付与し，再合成した歌唱である．

表 10　女性 2 の実験結果（上段：通常歌唱，下段：物真似歌唱）
Table 10　Experimental results of female 2 (upper: normal voices, lower: imitation voices).

	vibrato rate [Hz]	vibrato extent [cent]	vibrato rate's fluctuation β	vibrato extent's fluctuation [cent]	vibrato duration [%]
D_b	**0.04**	**0.01**	0.050	35.3	35.3
D_c	0.05	2.55	**0.049**	**12.7**	**2.5**
	vibrato rate [Hz]	vibrato extent [cent]	vibrato rate's fluctuation β	vibrato extent's fluctuation [cent]	vibrato duration [%]
D_b	**0.30**	**0.23**	**0.116**	37.5	40.8
D_c	0.31	2.79	0.117	**15.9**	**3.9**

また，F0 のみを制御しており，スペクトルや時間波形は未制御である．騒音レベルが 20.1 [dBA] の防音室で本実験を行い，正常な聴力を有する被験者 10 名（女性 5 名，男性 5 名）にヘッドホン（SONY MDR-CD900ST）を介してヴィブラート歌唱を呈示した．

6.3.1　提案モデルの有効性の検証（ABX テスト 1）

従来モデルにより合成したヴィブラート歌唱と提案モデルにより合成したヴィブラート歌唱のうち，どちらが収録したヴィブラート歌唱に近いかを被験者に選択させる ABX テストを行った．被験者に対して，AB（従来モデルにより合成されたヴィブラート歌唱，提案モデルにより合成されたヴィブラート歌唱）と X（収録されたヴィブラート歌唱）を順番に呈示し，3 秒以内に X に近いヴィブラート歌唱（A あるいは B）を選択させた．合成に用いるヴィブラートなしの歌唱とヴィブラートありの歌唱には，同じ条件（音階，物真似の有無）の歌唱を用いた．3 章で述べたヴィブラート歌唱の有効性の検証と同じ条件である単母音/a/の 3 音階（収録した最も低い音階から 1 度，3 度，5 度）を用い，48 パターン（プロ歌手 4 名，物真似の有無，3 音階，単母音/a/，AB の入れ替え）の ABX テストを行った．

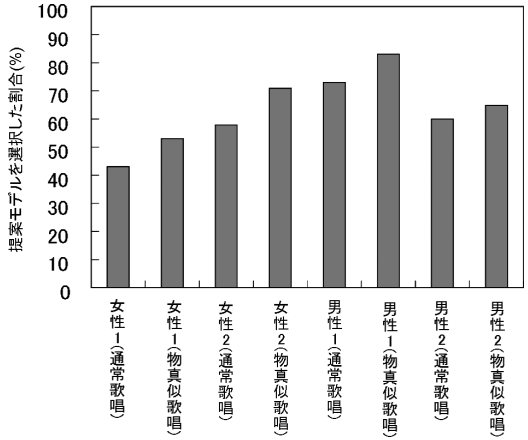

図 13　ABX テスト 1 の結果
Fig. 13　Results of ABX test 1.

以下に示す女性 1 の通常歌唱の場合のように，4 名のプロ歌手の通常歌唱と物真似歌唱を別々に評価した．

- **A**，**B**：以下の 2 種類
  （1）　女性 1 の通常のヴィブラートなしの歌唱に，女性 1 の通常のヴィブラートありの歌唱を基に従来モデルを用いて合成した F0 軌跡を付与した歌唱
  （2）　女性 1 の通常のヴィブラートなしの歌唱に，女性 1 の通常のヴィブラートありの歌唱を基に提案モデルを用いて合成した F0 軌跡を付与した歌唱
- **X**：女性 1 の通常のヴィブラートありの歌唱

従来モデルは vibrato rate，vibrato extent，vibrato rate's fluctuation を制御するモデルであり，提案モデルは，従来モデルに加え vibrato extent's fluctuation，vibrato duration の制御を可能にしたモデルである．vibrato extent's fluctuation とは深さの時間変動の大きさであり，大きな値であるほど従来モデルとの差が大きくなり，歌唱区間とヴィブラート区間の比である vibrato duration では，小さい値であるほど従来モデルとの差が明確になる．**図 13** に示す ABX テスト 1 の結果において，5 章の分析の結果，vibrato extent's fluctuation が大きく，vibrato duration が小さい値であった女性 2 の物真似歌唱や男性 1 の物真似歌唱の場合，提案モデルを選択した割合は 70 [%] 以上であり，提案モデルは収録されたヴィブラート歌唱の F0 を高精度に制御していることが分かった．一方，5 章の分析の結果で vibrato extent's fluctuation が小さく，vibrato duration が大きい値であった女性 1 の通常歌唱や物真似歌唱の場合，提案モデルを選択した割合は 50 [%] 前後であり，従来モデルとの差が小さいことが分かった．

6.3.2　歌手間の差異の知覚に関する検証（ABX テスト 2）

歌手間の差異を知覚可能かどうかを検証するために，収録されたヴィブラート歌唱と提案モデルを用いて合成されたヴィブラート歌唱を用いて ABX テストを行った．この検証では，歌唱データベースに収録された宇多田ヒカルの物真似歌唱と美空ひばりの物真似歌唱の差異，GACKT の物真似歌唱と五木ひろしの物真似歌唱の差異に着目した．歌唱データベースには，各プロ歌手が得意とする 13 音階（1 オクターブ）のヴィブラート歌唱が収録されており，4 名の音階の範囲は異なる．そこで，音階の差異の影響を考慮して，女性の場合の音階は C4，男性の場合の音階は G3 とし，女性と男性に分けて評価した．よって，16 パターン（性別：2（女性，男性），ヴィブラートの種類：2（宇多田ヒカルの物真似，美空ひばりの物真似あるいは GACKT の物真似，五木ひろしの物真似），声色の種類：2（女性 1，女性 2 あるいは男性 1，男性 2），1 音階，単母音/a/，AB の入れ替え）の ABX テストを行った．以下に女性の場合の ABX を示す．

- **A**，**B**：以下の 2 種類
  （1）　女性 1 が宇多田ヒカルの物真似をしたヴィブラート歌唱
  （2）　女性 2 が美空ひばりの物真似をしたヴィブラート歌唱
- **X**：以下の 4 種類
  （1）　女性 1 の通常のヴィブラートなしの歌唱に，女性 1 が宇多田ヒカルの物真似をしたヴィブラート歌唱を基に合成した F0 を付与した歌唱
  （2）　女性 1 の通常のヴィブラートなしの歌唱に，女性 2 が美空ひばりの物真似をしたヴィブラート歌唱を基に合成した F0 を付与した歌唱
  （3）　女性 2 の通常のヴィブラートなしの歌唱に，女性 1 が宇多田ヒカルの物真似をしたヴィブラート歌唱を基に合成した F0 を付与した歌唱
  （4）　女性 2 の通常のヴィブラートなしの歌唱に，女性 2 が美空ひばりの物真似をしたヴィブラート歌唱を基に合成した F0 を付与した歌唱

図 14 に示す ABX テスト 2 の結果において，歌手間の差異を知覚した割合とは，被験者の選択した物真似対象歌手と，X の F0 合成の基となった物真似対象歌手が同じであった割合を意味する．すべてのパターンにおいて 70 [%] 以上であり，女性の場合と男性の場合，ともに歌手間の差異を知覚可能であった．特に，女性 1 の通常のヴィブラートなしの歌唱

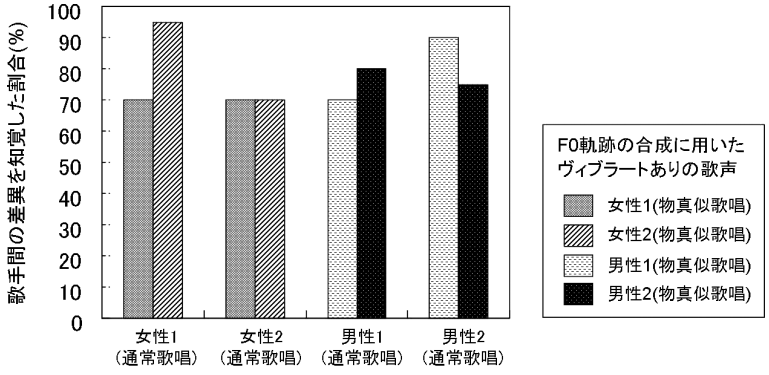

図 14　ABX テスト 2 の結果
Fig. 14　Results of ABX test 2.

に，女性 2 が美空ひばりの物真似をした歌唱を基に合成した F0 を付与した場合の割合は 90 [%] 以上であったが，女性 2 の通常のヴィブラートなしの歌唱に同じ F0 を付与した場合は 70 [%] であった．これは，3 章に示したように，女性 2 の通常歌唱と物真似歌唱が似ており，AB に用いた歌唱と X の用いた歌唱の声色が似ていたため，歌手間の差異を知覚した割合が低くなったと考えられる．

6.4 考　　察

従来モデルに vibrato extent's fluctuation と vibrato duration を組み込むことにより，vibrato rate，vibrato extent，vibrato rate's fluctuation の制御精度が低下した．しかし，ABX テスト 1 において，従来モデルと提案モデルを比較した結果，提案モデルはより高精度にヴィブラートの F0 を制御可能であることが分かった．また，ABX テスト 2 の結果では，声色が異なる歌唱において，提案モデルにより合成された F0 軌跡の差異だけでも歌手の識別が可能であることが分かった．これは，5 章で示した vibrato rate，vibrato extent，vibrato extent's fluctuation，vibrato duraion における宇多田ヒカルの物真似歌唱と美空ひばりの物真似歌唱の間，GACKT の物真似歌唱と五木ひろしの物真似歌唱の間の有意差を知覚したことを意味する．よって，提案する特徴量は個人性の制御に有効であると考えられる．

7. おわりに

我々は，プロ歌手の歌唱表現の差異を分析するために，プロ歌手 4 名がヴィブラートやポルタメントを表現した歌唱を収録し，歌唱データベースを構築した．ヴィブラートやポルタメントの制御法を分析するために，プロ歌手が普通に歌った歌唱（通常歌唱）だけでなく，特定のプロ歌手を物真似した歌唱（物真似歌唱）も収録した．本論文では，ヴィブラートに着目し，様々なプロ歌手のヴィブラートを制御する F0 モデルを構築するために，歌唱データベースを用いてヴィブラートの個人性の制御に有効な特徴量を検討した．歌唱データベースに収録したヴィブラート歌唱より，従来の特徴量とヴィブラートの時間変動や長さに関する特徴量を自動的に抽出し，ヒストグラム，t 検定，多重検定を用いて特徴量を分析した．全特徴量においてプロ歌手による差異が存在し，さらに vibrato rate，vibrato extent，vibrato extent's fluctuation や vibrato duration の場合，通常歌唱と物真似歌唱は異なり，プロ歌手が意識的に制御していることが分かった．そして，我々は，プロ歌手による差異と物真似による差異を確認した vibrato extent's fluctuation と vibrato duration を用いて，従来の F0 制御モデルを拡張した．歌唱データベースを用いた F0 制御モデルの評価実験の結果，提案モデルは従来モデルに比べ高精度にヴィブラートを制御可能であり，さらに提案する特徴量がヴィブラートの個人性の制御に有効であることを確認した．よって，特定のプロ歌手の歌唱から，本論文で有効性を確認した特徴量を抽出することにより，VOCALOID などで合成した歌唱に特定のプロ歌手のヴィブラートを付与することが可能となる．プロ歌手はこれらの特徴量を意識的に制御することが可能であるため，今後，旋律に応じて特徴量がどのように変化するかについて分析する必要がある．

謝辞　本研究の一部は，文部科学省のデジタル・ミュージアム開発プロジェクト，科学研究費補助金，および科学技術振興機構の CrestMuse プロジェクトの支援を受けて行われた．

参 考 文 献

1) 剣持秀紀，大下隼人：歌声合成システム VOCALOID，情報処理学会研究報告，2007-MUS-72, pp.25–28 (2007).
2) 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，情報処理学会研究報告，2008-MUS-75, pp.49–56 (2008).
3) 齋藤　毅，鵜木祐史，赤木正人：歌声における F0 動的変動成分の抽出と F0 制御モデル，日本音響学会聴覚研究会，H-2001-92, pp.683–690 (2001).
4) 小田切わか菜，粕谷英樹：歌声のビブラートの分析・合成・知覚に関する検討，日本

音響学会 1999 年秋季講演論文集，pp.545–546 (1999).
5) 齋藤　毅，鵜木祐史，赤木正人：自然性の高い歌声合成のためのヴィブラート変調周波数の制御法の検討，電子情報通信学会技術報告，TL2005-10, pp.13–18 (2005).
6) 中山一郎：日本語を歌・唄・謡う—共通の歌詞をうたい分けた音声試料の紹介，電子情報通信学会技術報告，SP2000-130, pp.1–4 (2001).
7) 齋藤　毅，後藤真孝，鵜木祐史，赤木正人：SingBySpeaking：歌声知覚に重要な音響特徴を制御して話声を歌声に変換するシステム，情報処理学会研究報告，2008-MUS-74, pp.25–32 (2008).
8) 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).
9) Prame, E.: Measurement of the vibrato rate of ten singers, STL-QPSR, *KTH*, Vol.33, No.4, pp.73–86 (1992).
10) Bretos, J. and Sundberg, J.: Measurements of vibrato parameters in long sustained crescendo notes as sung by ten sopranos, TMH-QPSR, *KTH*, Vol.43, No.1, pp.37–44 (2002).
11) 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 2005 年春季講演論文集，pp.269–270 (2005).
12) 後藤真孝，橋口博樹，西村拓一，岡　隆一：RWC 研究用音楽データベース：研究目的で利用可能な著作権処理済み楽曲・楽器音データベース，情報処理学会論文誌，Vol.45, No.3, pp.728–738 (2004).
13) 右田尚人，森勢将雅，西浦敬信：歌唱データベースの構築と歌手識別に有効な特徴量に関する基礎的検討，日本音響学会 2010 年春季講演論文集，pp.509–510 (2010).
14) 永田　靖，吉田道弘：統計的多重比較法の基礎，サイエンティスト社 (1997).
15) 森勢将雅，河原英紀，西浦敬信：基本波検出に基づく高 SNR の音声を対象とした高速な F0 推定法，電子情報通信学会論文誌，Vol.J93-D, No.2, pp.109–117 (2010).
16) Kawahara, H.: STRAIGHT, Exploration of the other as-pect of VOCODER: Perceptually isomorphic decomposi-tion of speech sounds, Acoustic Science and Technology, Vol.27, pp.349–353 (2006).
17) 大石康智，亀岡弘和，柏野邦夫，武田一哉：畳み込み HMM に基づく歌声の基本周波数制御モデルの提案とそのパラメータ学習方法，情報処理学会研究報告，2008-MUS-76, pp.89–96 (2008).

(平成 22 年 8 月 12 日受付)
(平成 23 年 2 月 4 日採録)

右田　尚人

昭和 61 年生．平成 21 年立命館大学情報理工学部メディア情報学科卒業．同年同大学大学院理工学研究科博士前期課程入学，現在に至る．音響信号処理の研究に従事．日本音響学会会員．

森勢　将雅（正会員）

昭和 56 年生．平成 16 年和歌山大学システム工学部デザイン情報学科卒業．平成 18 年同大学大学院システム研究科博士前期課程修了．同年 4 月より日本学術振興会特別研究員（DC1）．平成 20 年和歌山大学大学院博士後期課程修了．同年 4 月より関西学院大学理工学研究科ヒューマンメディア研究センター博士研究員．平成 21 年立命館大学情報理工学部助教，現在に至る．博士（工学）．音声・音響信号処理，インタフェース設計および聴覚情報処理の研究に従事．平成 18 年電気通信普及財団賞．日本音響学会，電子情報通信学会，日本バーチャルリアリティ学会各会員．

西浦　敬信（正会員）

昭和 49 年生．平成 9 年奈良工業高等専門学校専攻科電子情報工学専攻修了．平成 11 年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了．平成 13 年同大学院博士後期課程修了．同年和歌山大学システム工学部助手．平成 16 年立命館大学情報理工学部助教授．平成 19 年同准教授，現在に至る．博士（工学）．音響信号処理，主として音環境の解析・理解・再現・生成に関する研究に従事．平成 13 年電気通信普及財団賞，平成 13 年 ATR 発明・論文表彰．平成 21 年日本バーチャルリアリティ学会論文賞．日本音響学会，電子情報通信学会，日本騒音制御工学会，日本バーチャルリアリティ学会各会員．