

# 奄美大島民謡風の歌唱デザインを支援するシステム： ゲインレゾネータ

森勢 将雅<sup>1,a)</sup> 村主 大輔<sup>2</sup> 馬場 隆<sup>2</sup> 片寄 晴弘<sup>2</sup>

受付日 2012年7月1日, 採録日 2013年1月11日

**概要：**電子的に合成される歌唱が音楽コンテンツ制作に用いられるようになって以来、パソコンを用いた音楽制作はこれまでにない盛り上がりを見せている。Vocaloid に代表的される歌唱合成ソフトウェアでは、煩雑なパラメータの調整（歌唱デザイン）が自然な歌声を生み出すために必要であり、クリエイタは、作業時間の多くを歌唱デザインに割いている。本研究では、歌唱デザインの1つの形として、歌唱素材に対して、特定の歌手の歌唱スタイルを転写する方法を取り扱う。本論文では、島唄風歌唱における歌唱技巧「ゲイン」を対象とし、入力された歌声を島唄風に変換する技術、および、歌唱デザインを支援するインタフェース「ゲインレゾネータ」を提案する。F 値により性能を評価した結果、67.8%であることが示され、主観評価では知覚的にゲインを転写できることが確認された。

キーワード：歌声分析, 歌声合成, ゲイン, STRAIGHT

## GUIN-Resonator: A System to Support Singing Voice Design for Amami Folk Songs

MASANORI MORISE<sup>1,a)</sup> DAISUKE SUGURU<sup>2</sup> TAKASHI BABA<sup>2</sup> HARUHIRO KATAYOSE<sup>2</sup>

Received: July 1, 2012, Accepted: January 11, 2013

**Abstract:** In accordance with increasing power of PCs, commercial vocal synthesizers have been used for music creation. However, to synthesize natural singing voice, the users have been forced to manipulate several parameters. Since this manipulation, named singing design, is too laborious, methods to lower the burden has been longed for by many creators. To cope with the problem, we propose a method to copy a particular singing style into an input vocal melody. The method of analysis of and copying singing style is explained, focusing on GUIN, a typical vibrato technique of *Amami folk song*. Then, this paper introduces GUIN-Resonator, as an implementation of the proposed method. An objective evaluation showed that the F-value was 67.8%, and a subjective evaluation suggested that the proposed method can transcribe the GUIN.

**Keywords:** singing analysis, singing synthesis, GUIN, STRAIGHT

### 1. はじめに

DAW (Digital Audio Workstation) と呼ばれる統合的音楽編集環境が一般のパソコン上でも動作するようになり、

かつては、職業的な専門家でしか使用できなかった音楽制作をアマチュアレベルのユーザが実施できるようになった。現在、パソコンを用いた音楽制作は、1990年代のDTM以来の盛り上がりを見せており、CGM (Consumer Generated Media) の主要コンテンツとして流通する機会も大いに高まった。音楽制作において用いられる音素材としては、制作者が録音したものほかに、電子的に合成される楽音が

<sup>1</sup> 山梨大学工学部  
Faculty of Engineering, University of Yamanashi, Kofu,  
Yamanashi 400-8511, Japan

<sup>2</sup> 関西学院大学大学院理工学研究科  
Graduate School of Science and Technology, Kwansai Gakuin  
University, Sanda, Hyogo 669-1337, Japan

a) mmorise@yamanashi.ac.jp

本論文に関連する動画コンテンツを以下で視聴できる。  
<http://www.youtube.com/watch?v=W58iaXEuLGk>  
<http://www.nicovideo.jp/watch/sm19910895>

利用される。2003年、YAMAHAからVocaloid [1]が発売されて以降、電子的に合成される歌唱が音楽コンテンツの制作に用いられるようになった。さらに、そのエンジンを利用して2007年に発売されたバーチャル・シンガ「初音ミク [2]」が、CGMの代表格としてブレイクしたのにあわせて、歌声合成技術によるボーカルパートの生成が、広く社会一般にも認知されるようになった [3]。

Vocaloidは、メロディと歌詞を入力することによって歌声が生成される画期的な市販アプリケーションである。しかし、人間的で「自然な」歌唱を作りあげるためには、煩雑なパラメータの調整作業（歌唱デザイン）が必要となる。この歌唱デザインの省力化を目指した取り組みとして、中野らは、ユーザの歌声を参照し、基本周波数と音量のエンベロープの反復推定によって、Vocaloidの制御パラメータを調整するVocaListener [4], [5]を提案した。Moriseらは、2つの歌唱を、声質、節回しの任意のモーフィグ率で合成するインタフェースv.morish [6]の提案を行った。v.morishを用いれば、音響信号のレベルでの簡易な歌唱のデザインが可能である。楽譜と歌詞の情報を入力するだけで歌唱を合成するVocaloid以外のエンジンとしては、徳田らのSinsy [7]があげられる。Sinsyでは、歌い手の声質と基本周波数パターンをHMM（隠れマルコフモデル）によりモデル化するため、統計処理に耐えうる大量の歌声データが用意できれば、モデルとなった歌い手の表情が再現される。一方、Sinsyは曲全体に対するパラメータ調整は可能なものの、特定の部分のみ作りこむような詳細な調整ができないという問題がある。

歌唱デザインの支援に向けては、そのメカニズムやモデル化に向けた取り組み、ユーザインタフェースまで、様々なレイヤでの取り組みが求められるが、本研究では、アマチュアレベルのユーザを対象に、実際に使用される応用インタフェースを主眼としたシステムの提案を目指す [8]。その1つの形として、ユーザの歌唱に、近年のポピュラーソングの歌唱ジャンルとしても注目を集めている鳥唄風歌唱における歌唱技巧「ゲイン」を付与する歌唱合成システム・ゲインレゾネータを提案する。以下、2章で、代表的な鳥唄風歌唱である奄美大島鳥唄の概要とゲイン転写に向けての課題について述べる。3章では、ゲイン音響的特徴の分析と音響的特徴の転写による歌唱デザイン支援について述べる。4章では、ゲインがどのような部位に入りやすいかに関する分析について述べる。5章では、上記の検討に基づいてデザインされたゲインレゾネータのシステムと動作例について紹介し、6章で、本方式の可能性について検討を実施する。

## 2. 奄美大島鳥唄と転写型デザイン支援のフレームワーク

歌唱は、音楽コンテンツの主要要素であるが、最近の音

楽制作現場では、計算機を用いた加工、デザインを実施するのが通例となってきた。歌唱デザインの支援としては、Sinsyのように、パラメータを設定することで歌唱全体の生成を目指す自動方式のものと、DAWを利用し、ユーザが、収録歌唱の特定の区間を切り出して、ピッチ制御を行ったり、プラグインエフェクタを適用していくといったヒューマンインザループタイプのものに大別される。自動方式には、効率性とまったくの初心者を利用できるというメリットがある一方で、実際のプロダクション現場では、人間の介入を前提とした作業がとり行われている。その前提で、デザイン全体の効率をいかにして向上させるかという取り組みが必要不可欠である [8]。

市販DAWやプラグインの機能を利用することで、歌唱に、「ビブラート」や「コブシ」などの歌唱技巧を付与することもできる時代にはなっているが、それら機能を使いこなすことは、アマチュアクリエイターにとっては必ずしも容易なことではない。この問題に対し、本研究では、プラグインスタイルで、歌唱技巧を簡単に転写する方式を提案する。歌唱の特定の区間を選び、その部分の編集を行うというDAWの基本的な操作方式を踏襲し、そのうえで、歌唱技巧転写処理の自動化と、歌唱技巧を挿入する際に入りやすさをガイドする機能によって歌唱デザインの支援を実現する。

歌唱素材に対して歌唱技巧を転写するという方式の取扱いにあたっては、1) その歌唱技巧が転写されたかどうか明らかに分かること、2) その歌唱技巧が、実際のコンテンツの視点からも一定のニーズがあること、を満たすことが望まれる。これらの条件を満たすものとして、本研究では、近年、J-POP領域において、複数のボーカリストが出現し、その様式が確立した鳥唄風の歌唱を取り上げる。本章では、鳥唄風の歌唱法の代表として、奄美大島鳥唄の特徴的な歌唱法について概説し、続いて、表情転写のデザインの支援を行うにあたっての考慮事項のポイントを整理する。

### 2.1 奄美大島鳥唄の特徴

J-POP領域において、鳥唄風の歌唱を行っているとされる歌手の多くは、奄美大島鳥唄の歌い回しを実践している。奄美大島鳥唄の歌い方の特徴として、ファルセット（裏声）の多用と「ゲイン」と呼ばれる独特の歌唱技巧があげられる。ファルセットは近年のJ-POPでもごく普通に使われるのに対し、ゲインは他の民謡および奄美大島出身の歌唱者以外のJ-POPでは聴くことのできない歌唱技巧とされる。ゲインの特徴的な性質として、地声-裏声-地声と瞬間的に変化させることが鳥唄歌唱者によって指摘されている [14]。裏声は高さだけではなく声質の変化もともなうため、ゲインの転写を実現するためには声質（スペクトル包絡。以下では単にスペクトルとする）を変化させる必要が

ある。

日本には民謡や長唄といった様々な伝統音楽があり、これまでにも、音楽学における分類学的な興味に根ざした調査・検討がなされてきた。それらに比べて、定量的な分析を実施した研究例は限られている [9], [10], [11], [12], [13]. 提案システムの実現においては、合成系処理の適用をふまえた奄美大島鳥唄特徴の定量的分析が最初の課題となる。

## 2.2 歌唱技巧の転写による歌唱デザイン支援

現在、表情付け支援システムは実用化段階に至っており、自動処理の実現に加えて、ユーザのデザインタスクを支援する実用的な方式が求められる時代に入ってきている [8]. 歌唱を含む表現技巧の転写方式はその代表的なものの1つとして位置づけられ、以下の機能のサポートが求められる。

### 1) 歌唱技巧転写のための信号処理

歌唱技巧を転写した結果がユーザの想定した歌唱技巧として知覚され、かつ、F0、スペクトルを元々の歌唱と自然につながるように変換する歌唱技巧転写処理。

### 2) 歌唱技巧の出現位置に関するガイダンス

システムがユーザをサポートするため、歌唱技巧が入る可能性のある場所をガイダンスとして表示し、ユーザがそのガイダンスに基づいてデザイン転写適用の可否を判断できるようなインタフェース。

以下の3章では、上記の条件を考慮したゲインの音響的特徴の分析と転写処理、出現位置の分析手法について述べる。

## 3. ゲインの音響的特徴の分析と歌唱への転写方式

2章では、ゲインは地声から瞬間的に裏声となり、また地声に戻る歌唱技巧であることを述べた。この変化は、高さだけではなく音色の変化も生ずるため、F0とスペクトル構造が短時間で変化することとなる。したがって、歌唱へゲインを転写するためには、F0の変動に同期してスペクトルも制御する必要がある。本章では、プロ歌手の奄美大島鳥唄を分析し、ゲインのF0・スペクトルの音響的特徴を明らかにする。ついで、明らかとされた音響的特徴を歌唱へ転写する方法を提案する。

### 3.1 ゲインの音響的特徴の分析

我々は、奄美大島出身の歌手として、中孝介 [15]、元ちとせ [16] の歌唱、また、鳥唄風の歌唱ができるプロ歌手の歌唱の分析を進めてきた [17]。その結果、ゲインの音響的特徴は、歌唱者によって若干の差異はあるものの、共通の特徴が存在することが確認された。ここでは、典型的なゲインの分析対象事例として、中孝介の「花」の「いいわけさえも」というフレーズを分析事例として紹介する。なお、このスペクトル分析には、音声の分析・再合成におい

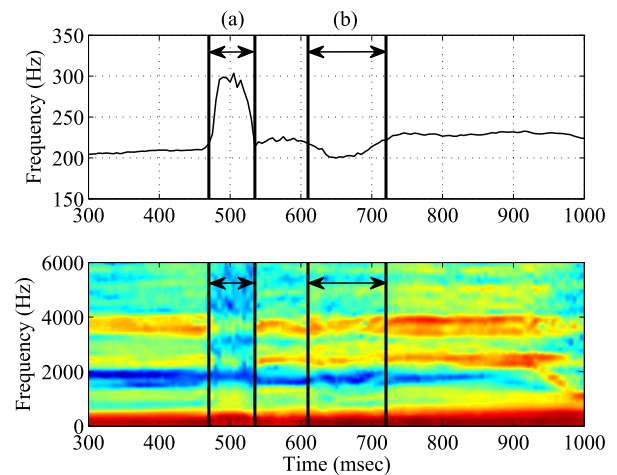


図1 ゲインにおけるF0(上図)とスペクトログラム(下図)の特徴的な変動

Fig. 1 F0 (upper) and spectrogram (bottom) of the singing voice including GUIN.

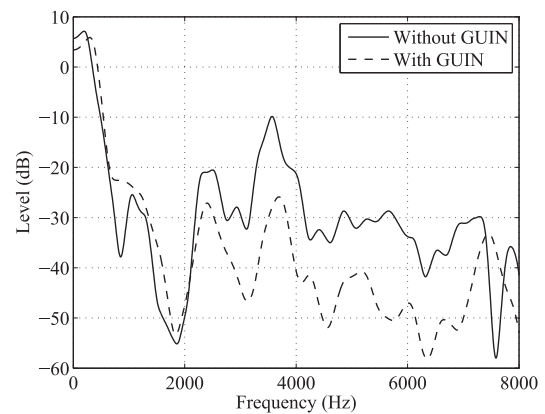


図2 ゲイン部と非ゲイン部のスペクトル

Fig. 2 Spectra in GUIN and non-GUIN.

てF0・スペクトルを高精度に推定可能な音声分析変換合成システムTANDEM-STRAIGHT [18]を用いた。分析の音声はTANDEM-STRAIGHTではF0に同期した窓を用いることでF0に依存したスペクトルの変動成分を除去することができる。また、分析音声のサンプリング周波数は44.1kHzであり、F0、スペクトル包絡は5msecごとに推定することとした。

分析されたF0軌跡とスペクトログラムを図1に示す。特に特徴的な箇所については図中に(a)、(b)として示している。また、ゲイン部分のスペクトルとゲイン直前のスペクトルを図2に示す。この結果から次の2つの事項が推察される。

- F0が65msecの区間で85Hz程度変化した(図の(a)部分。本論文では「隆起」と定義する)後、さらに110msecの区間でF0が20Hz程度下がって元のF0に戻っている(図の(b)部分。本論文では「沈降」と定義する)。これは、長唄でも用いられている「あたり」[10], [11]に似た特徴である。ただし、あたりでは

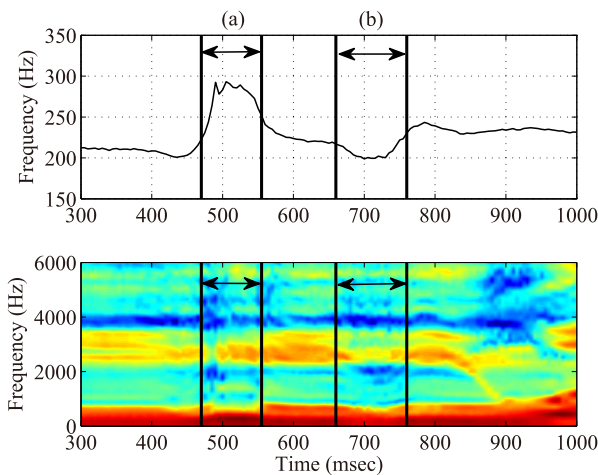


図 3 奄美大島民謡の歌唱が可能なプロ歌手によるグインにおける F0 (上図) とスペクトログラム (下図) との特徴的な変動

Fig. 3 F0 (upper) and spectrogram (bottom) of the singing voice including GUIN sung by another professional singer.

沈降は見られないため、この沈降がグインの大きな特徴を表している可能性がある。

- 同一音素を発声中にもかかわらずスペクトルが急激に変化している。特に、F0 の隆起時に高次倍音 (2~7kHz 付近) が約 14.5 dB 減少している。裏声はスペクトルの変化として観測できるため、このスペクトルの変化は裏声を用いられた可能性を示している。

次に、図 1 で取り上げたフレーズの中孝介の歌い方を、奄美大島民謡の歌唱が可能なプロ歌手 (奄美大島出身ではないスタジオミュージシャン) が模倣したものの分析結果を図 3 に示す。

奄美大島民謡の技巧を使える歌手と中孝介の F0・スペクトログラムでは同様の特徴が図 3 では見られる。その他複数の楽曲から取り出したグインを分析した結果、グインを表現するには

- (1) F0 における 50~80 msec と短時間で生じる 85 Hz 程度の隆起と 100 msec 程度の区間で生じる 20 Hz 程度の沈降
- (2) 裏声のスペクトルにおける倍音成分の減少
- (3) F0 の隆起に合わせたスペクトル変動 (F0 変動とスペクトル変動の同期)

が必要といえる。

### 3.2 グインの音響的特徴の転写

本節では、前節で示した 3 つの条件を満足するように音響的特徴を転写する手法について述べる。具体的には F0 の操作により (1) の隆起と沈降を再現し、スペクトルの操作により (2) の倍音成分の減少を実現する。スペクトルの操作は F0 の隆起と同期させることで (3) を満たすようにする。なお、グインの音響的特徴は多くの場合同様の傾向

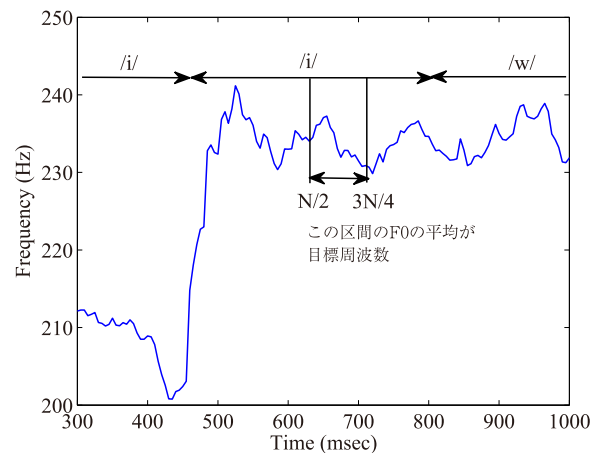


図 4 目標周波数を決定する範囲

Fig. 4 Length used to determine the target frequency.

であったため、本論文ではモデルを構築するのではなく、実際に歌われたグインを事例のテンプレートとしてそのまま入力された歌唱に転写するアプローチを採用する。

#### 3.2.1 F0 の操作

まず、2 章で分析に用いた「中孝介」の「花」のフレーズ「いいわけさえも」からグイン部分の F0 変動をテンプレートとして取得する。このテンプレートを歌唱音声のグインを挿入する位置で入力音声の F0 と置き換えることによりグインの F0 変動を表現できる。

この際、歌唱者の音声の F0 の遷移に合うようにグインの F0 の隆起と沈降を表現することが課題となる。ここでは、グインを挿入するノートの音価<sup>\*1</sup>に相当する入力音声の区間で、基準となる F0 の値 (目標周波数と呼ぶ) を入力音声と譜面から決定することとする。目標周波数は、図 4 に示されるように、目標周波数を求める音声区間内 (ここでは /i/) で区間長の中央から 3/4 までの F0 の平均値とする。音声区間内の後半以降の F0 を使用する理由には、楽譜音高から大きく F0 が逸脱するポルタメント、音高変化の際目標とする F0 を瞬時的に超えるオーバシュート、音高変化前に変化と逆方向に振れるプレパレーション [19] が歌唱において発生しにくいことがあげられる。区間の長さを音響信号のフレーム数  $N$  とし、 $i$  番目のフレームの F0 を  $F_i$  とすると目標周波数  $F_t$  は

$$F_t = \frac{4}{N} \left( \sum_{i=N/2+1}^{3N/4} F_i \right), \quad (1)$$

と表される。 $N$  は TANDEM-STRAIGHT により分析されたフレーム数を示している。本分析ではシフト幅を 5 msec としたため、 $N$  は  $5N$  msec の時刻の分析結果と一致する。

図 1 より、F0 の隆起と沈降の間の上下動が少ない部分は沈降後の F0 の安定した箇所とほぼ同じことが確認できるため、歌唱の際に目標とする音の高さに相当すると考え

<sup>\*1</sup> 音価とは、譜面に記載された音符の長さで、実際の継続長に対応する音長とは異なる。

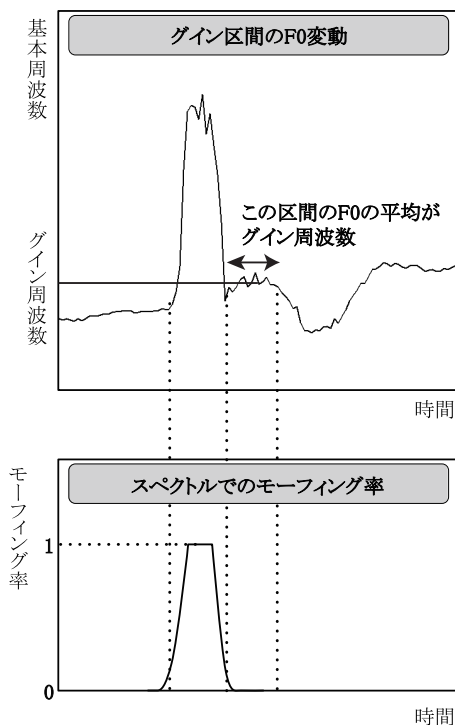


図 5 グインの F0 テンプレート (上図) とスペクトルのモーフィング率 (下図)  
**Fig. 5** F0 template of GUIN (upper) and the morphing rate of the spectrum (bottom).

ることができる。この隆起と沈降の間の F0 の平均値をテンプレートの基準の F0 (グイン周波数と呼ぶ) とする。また、我々は、中孝介の「花」、元ちとせの「ワダツミの木」「幻の月」「語り継ぐこと」「千の夜と千の昼」「ミヨリの森」の楽曲を受聴し、グインは楽曲中で音高遷移が起こった直後に表現される傾向を確認した。そのため、音韻のアライメント情報から発声開始時刻に合わせて、グイン周波数と入力音声から決定される目標周波数が一致するように置き換えることとする。本論文では図 1 のグインをテンプレートとして採用した。具体的には、隆起は 65 msec の区間で 85 Hz 変動させ、沈降は 110 msec の区間で 20 Hz 変動させている。また、グイン周波数は図 1 より 221.7 Hz とした。

グインを転写するためには、転写元の歌声の発声長が、テンプレートとなるグインよりも長いことが求められる。たとえば転写元の歌声がテンプレートの歌声よりも短い場合、その時刻にはグインは挿入されないと見なすことで対応する。

### 3.2.2 スペクトルの操作

スペクトルの操作では、グインの挿入を行う場所の歌詞から事前に取得した母音の音韻情報が利用される。具体的には、前項で説明した F0 の操作で決定した入力音声におけるグインの挿入位置と合わせ、地声から裏声に変換するフィルタ (以下では単に変換フィルタとする) を用いてスペクトルを制御する方法を採用する。本論文では、一般歌唱者 3 名の 5 母音の地声と裏声の音声を収録し、裏声の

パワースペクトルと地声のパワースペクトルの比を各周波数について求めることにより変換フィルタを構築した。なお、TANDEM-STRAIGHT ではスペクトル包絡から最小位相を計算し波形を得るため、位相の制御は行わない。

実際の発声においては、地声から裏声に切り替わる際に過渡的な状態が存在する。その部分を省略し、地声と裏声の部分の直接つないでしまうと、自然な聞こえは得られない。ここでは、地声と、変換フィルタによって得られた裏声間のスペクトル変換に歌唱モーフィング [20], [21] を利用し、過渡的な部分を作り出す。具体的には、隆起中の音響信号が、隆起開始から隆起のピークまでは徐々にスペクトルに近づき、ピーク後から隆起の終了までは地声に戻るよう入力音声のスペクトルと変換フィルタで処理後のスペクトルを用い、図 5 のように F0 の遷移の仕方にあわせてモーフィング率を変化させる。同一母音による歌唱モーフィングではフォルマントの対応付けが不要なため、単純な重み付き加算によりモーフィング可能である [6]。

## 4. グインの出現部位の自動探索

鳥唄歌唱者の歌唱において、音の並び (メロディ) における特定の音符においてグイン唱法が使われている一方で、まったくグインが使われない音符もある。グイン唱法には、入りやすい部位とそうではない部位が存在すると思われるのが妥当である。本章では、事例探索に基づいてグイン唱法が入る可能性のある部位を自動探索するアプローチについて述べる。

### 4.1 事例探索で用いる情報

グイン唱法が入りやすい音符の探索にあたり、網羅的な探索を実施するのは効率的とはいえない。我々は、3.2.1 項で実施した検討から、主観的な判断として以下の傾向を確認している。1) プレス間で区切られた音列 (以下、フレーズと呼ぶ) 中で、出現するグインの数はたかだか数個に限られる、2) フレーズの後半でグイン唱法が出現しやすい。本論文では、これらの傾向に着目し、まず、プレス位置による楽曲のフレーズ分割を行う。そのうえで、分割したフレーズごとに、ピッチコントロールの基準や目標となる音高に関する情報と、継続時間に関する情報を事例探索のキー (以下では探索キーとする) として探索を実施する。具体的には、以下の 7 つをキーとしている。

- (1) 相対音高: 楽曲内の平均音高からの差
- (2) 音高差: 1 つ前の音符との音高差
- (3) フレーズ内での相対的な音高: フレーズ内での最小音高からの差
- (4) フレーズ内での音符の位置: フレーズ内でおおよその位置にいるのか
- (5) フレーズ内での相対的な音価: フレーズ内での最小音価からの比

(6) 音長：音の長さ．単位は msec

(7) 音価：音符の長さ

(1) から (3) は音高に関する情報，(4) から (7) が継続時間に関する情報を示す．

本論文では，音高差・フレーズ内での音符の位置・音長の3つの条件に加え，ビブラートなどの歌唱技巧を使用する際にピッチコントロールの基準や目標となる相対音高とその技巧の継続時間に関係する音価を条件として楽譜情報から取り出すこととした．また，コブシの一種であるゲインを挿入する条件として，論文 [23] を参考にフレーズ内での相対音高と相対音価も条件としてあげることとした．これは，ゲインがフレーズの内では先頭と終端のそれぞれの付近で音高遷移の起こった場合に楽曲内で相対的に音長が大きい音符に対して挿入されるケースが多いためである．

#### 4.2 事例データベースの構築

事例データベースでは，前節で述べた探索キーに基づきゲインが挿入される頻度を保存する．本研究では，探索キーのもとにある音符にゲインが挿入される頻度をゲインの入りやすさとして扱う．楽譜情報である探索キーの各値は確率変数  $Y$  で表し，MIDI より入力される楽曲データは，各音符に対しゲインが入る場合を正事例 ( $X = 1$ )，入らない場合を負事例 ( $X = 0$ ) とし，それぞれの出現回数からすべての確率変数の条件  $Y$  における条件付き確率  $p(X|Y)$  を次のように設ける．

$$p(X|Y) = \begin{cases} \frac{C_{\text{positive}}}{(C_{\text{positive}} + C_{\text{negative}})} & (X = 1) \\ 0 & (X = 0) \end{cases} \quad (2)$$

ただし， $C_{\text{positive}}$  は条件  $Y$  のときにある音符で正事例が起きた回数， $C_{\text{negative}}$  は負事例が起きた回数を表す．この確率をゲインの入りやすさとして事前に奄美歌唱者の楽曲から求め，事例データベースを構築する．

#### 4.3 事例探索アルゴリズム

事例探索は，入力する一般歌唱の楽曲（対象楽曲と呼ぶ）に対して図 6 に示される手順で行われる．なお， $X$ ， $Y$  は事例データベースの構築で述べた確率変数を示す．

以上の処理を対象楽曲のすべての音符に対して行い，各音符に格納されている確率をゲインの入りやすさを表す信頼度とする．この信頼度がある一定の閾値  $T$  を超えた音符に対してのみゲインの挿入を可能とする．探索範囲を拡張するためのガウス分布の標準偏差の  $\sigma$  の最大更新回数は，ガウス分布の最大値が信頼度の閾値  $T$  以下になるまでとする．上記のステップで事例探索を行った際にも事例が見つからない場合は，設定する楽譜情報の各条件の数を 1 つずつ減少をさせ次元を落として再度事例を探索する．この条件の次元落としは，ゲインの入る位置の制約が緩くなるため，次元を落とすごとにある，一定の割合の重みを掛けて

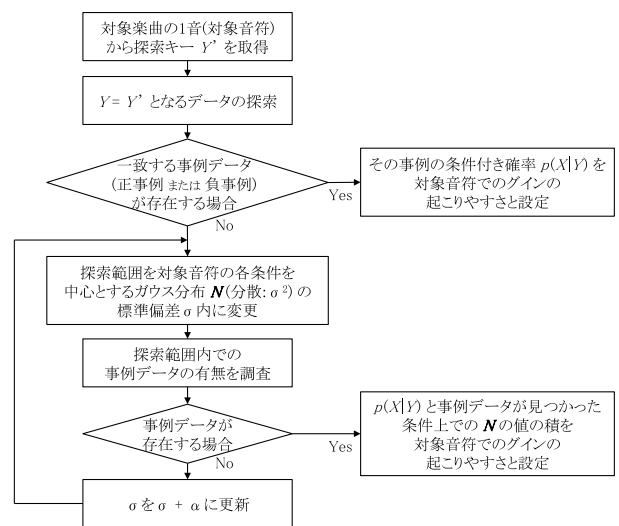


図 6 事例探索のフローチャート

Fig. 6 Flowchart for the case search.

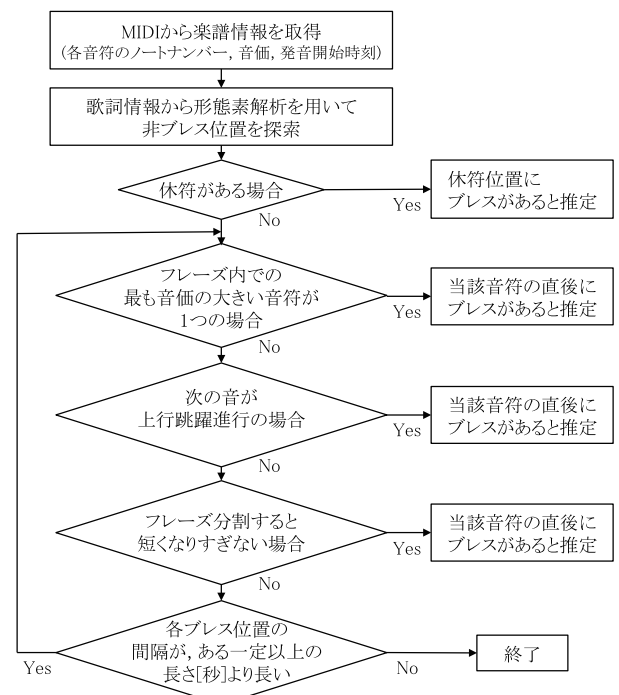


図 7 フレーズ分割のフローチャート

Fig. 7 Flowchart for the phrase segmentation.

信頼度を低減させる．

#### 4.4 プレスによるフレーズ分割

プレスが入る位置には，音列が切れる休符や持続する1音の長さなどの楽譜情報が関係している．本論文では単語内でのプレスはほぼ起こらないと仮定し，プレス位置からのフレーズ分割を試みる．本論文では，ユーザが入力するMIDI（楽譜）情報と歌詞情報を用いて，図 7 の手順によりプレスの位置を推定してフレーズ分割を行う．

## 5. 奄美大島民謡風歌声合成システム「グインレゾネータ」

本章では、前章、前々章で述べたグインの挿入部位の分析とグインの音響的特徴の転写処理を利用した、一般歌唱にグインを挿入する奄美大島民謡風歌唱デザイン支援システム「グインレゾネータ」について述べる。

グインレゾネータの処理概要を図 8 に示す。グインレゾネータを用いてユーザが実施することは、1) 楽譜情報 (ボーカルパートのメロディと歌詞情報) の入力, 2) レコーディング (歌唱の入力), 3) エディティング (ガイダンスに沿ったグイン転写音符の選択) である。システムは、グインの転写処理に加えて、これらの処理を支援する GUI を備えている。

### 5.1 ユーザインタフェース

グインレゾネータを起動し、楽譜情報をインプットすると図 9 のようにピアノロール形式で楽曲情報が表される。ピアノロールのノートに配された色は、グイン唱法の入りやすさの違いを表現しており、水色はグイン唱法の可能性がない音、赤はグイン唱法が入りやすい音、緑はグイン唱法が入る可能性があるが、赤ほどその可能性が高くないことを表している。本実装では、上位から順番にソートし、おおむね下位 6 割を水色、上位 2 割を赤色、それ以外のも

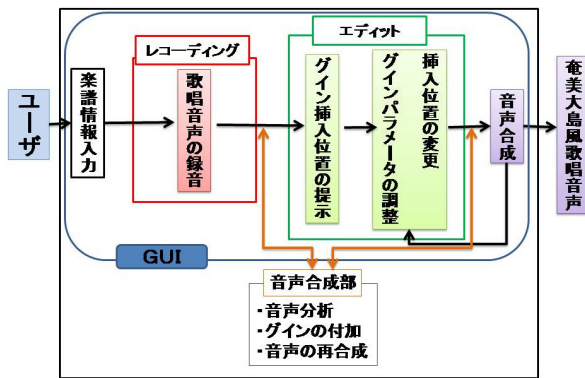


図 8 提案する歌声合成システムの概要

Fig. 8 Overview of the proposed singing voice synthesis system.

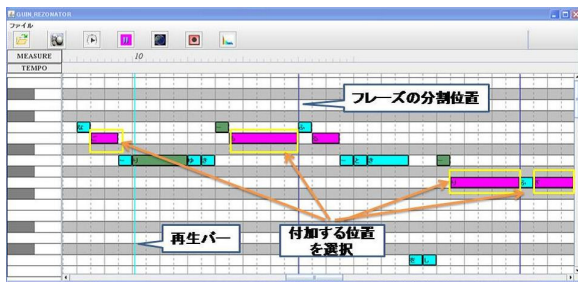


図 9 グインレゾネータによるグイン挿入位置の提示

Fig. 9 Suggestion of the insert position of GIN by GIN-resonator.

のを緑で表示するようにした。また、色の明るさによってユーザの選択した場所を分けており、図 9 の場合は、黄色の四角で囲った赤色のノートでグイン唱法の挿入を選択している状態を示している。

### 5.2 レコーディングモード

レコーディングモードでは、ユーザが入力した MIDI 音源にあわせて歌唱を行い、音声合成・編集を行うための歌声を録音する。この際、ユーザはカラオケのようにシステムから流れてくるガイドメロディを聞きながらインタフェース上の画面に表示される歌詞とノートの流れにあわせて歌唱する。

### 5.3 エディットモード

エディットモードでは、ユーザは、グイン唱法の候補とし

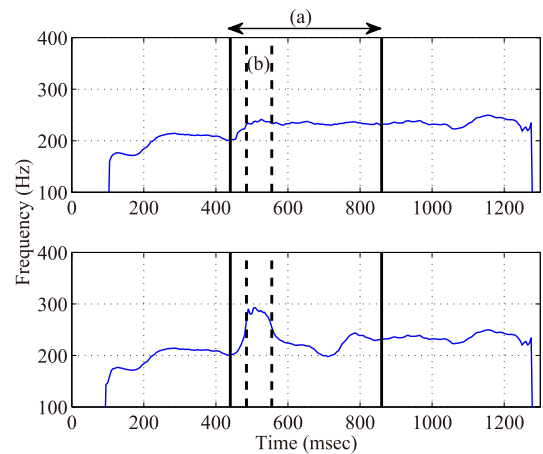


図 10 グインを挿入する前後の F0 の遷移 (上図は入力音声, 下図は出力音声)

Fig. 10 F0-transition between before and after processing to insert GIN (upper: before processing, bottom: after processing).

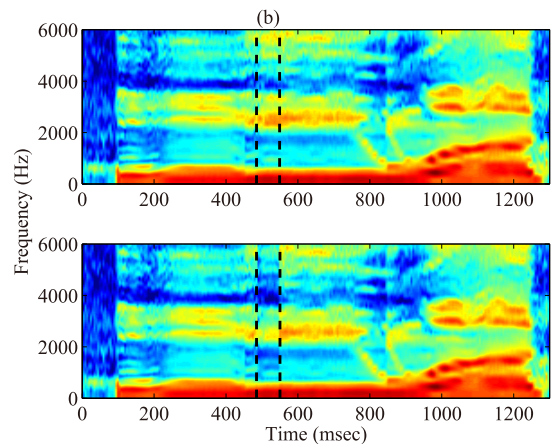


図 11 グインを挿入する前後のスペクトル (上図は入力音声, 下図は出力音声)

Fig. 11 Spectrum between before and after processing to insert GIN (upper: before processing, bottom: after processing).

てあげられたノートの中から音符を指定し、指定された音符に対して、グイン唱法の転写を行う。ユーザは、聴取と編集を繰り返しながら、必要に応じて、TANDEM-STRAIGHTが提供しているF0とスペクトルの情報(図10, 図11)を確認しつつ、効率的に所望の奄美大島民謡風の歌唱音声を得ることができる。

図10は転写前後のF0を示しており、(a)の区間で転写がなされている。(b)は、スペクトルのモーフィング率を制御する隆起の区間を示す。図11は、転写前後のスペクトログラムを示している。(b)の区間で高域が減衰していることが確認できる。

## 6. 提案法の有効性に関する実験

本章では、上述のグインの音響的特徴を挿入する方法、およびグインを挿入する位置の自動探索方法の基本性能に関する実験について述べる。

### 6.1 グインを転写した歌唱の品質評価

提案法が奄美大島歌唱者のグインを適切に転写できているか確認するため、奄美大島歌唱者の歌唱と提案法によりグインが転写された歌唱の類似度を、非計量多次元尺度法により主観的に評価した。

ここでは中孝介の「花」の「いいわけさえも」というフレーズを対象とし、奄美大島民謡技巧を使用できるプロ歌唱者の歌唱を用意した(刺激 $S_o$ )。その際、「いいわけさえも」の2つ目の「い」においてグインをいれるよう指示した。同様の歌声に関して、グインを入れないよう指示して収録した歌唱を刺激の1つとした(刺激 $S_n$ )。刺激 $S_n$ に対し、本論文で提案するグインの音響特徴を転写した音声を3つ目の刺激 $S_r$ とした。

被験者は正常な聴力を有する男性5名とし、用意した3つの刺激のうちランダムに抽出された2つの刺激に対し、その2つの刺激の類似度を6段階で評価した。聴取順序による判断への影響を除去するため、比較刺激の順序はランダムとした。同じ刺激を用いた施行を10回行い、そのうち5回の比較刺激の順序を逆にして提示した。

実験では、ヘッドフォンを介して刺激音を両耳に呈示し、被験者は、PCディスプレイ上に表示されたスライドを操作することで回答した。実験環境は、楽曲制作と同様の条件とするため、防音室ではなく通常の室内とした。また、回答時間に制限は設定しないが、刺激音の聴き直しを許可しないこととした。

非計量多次元尺度法による各音声刺激の類似度を図12に示す。図12の表内の各刺激の数値は人間の印象による距離に対応し、刺激間の距離が近いほど刺激どうしが似ていることを示している。この結果から、提案法により転写されたグインは、プロ歌唱者による奄美大島民謡技巧と類似しているといえる。また、今回の実験に使用した刺激の

刺激	$S_o$	$S_r$	$S_n$
座標	-1.29	-1.24	2.53

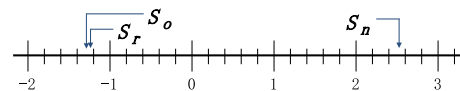


図12 非計量多次元尺度法による各刺激の類似度を示した距離  
Fig. 12 Distances in the similarity among all stimuli by non-metric multi dimensional scaling.

条件数	7	6	5	4
F値(%)	20.8	32.8	41.2	40.0

図13 F値と条件の数との関係

Fig. 13 Relationship between F-values and the number of conditions.

条件	相対音高	前後の音高差	音長	音価	フレーズ内での音符の位置	F値(%)	F値(%) ( $\sigma=0$ )
$\sigma$ (標準偏差)	0.08	0.14	0.05	0.09	0.17	67.8	41.2

図14 各条件下における標準偏差と性能評価(F値)

Fig. 14 Standard deviations and F-values in each condition.

組合せと非計量多次元尺度法の類似度の評価方法としての精度は、ストレス値が $1.1627e-017$ ときわめて小さいことから、非常に良いといえる。

### 6.2 グインの挿入位置探索法の基本性能

次いで、事例データベースとして奄美大島出身の歌手である元ちとせの楽曲情報をMIDIで与えて学習し、未知の楽曲を与えた際のアルゴリズムの性能を交差検定(cross-validation)により評価する。本論文では、性能評価のため、奄美大島民謡に該当する6つの楽曲を用いた。選定する条件のうち、グインが挿入される条件としてふさわしいかを調べるため、使用する条件の数と組合せを変更しF値を比較して検証する。F値は、各楽曲を受聴し主観的にグインの有無を判定した結果と、提案法により得られたグイン挿入判定結果から求めた。また、情報該当事例がない場合に近傍の事例を取得する方法を採用しているため、近傍事例を取得するためのパラメータであるガウス分布の標準偏差 $\sigma$ の値を変化させ、 $\sigma$ の値による性能評価の差も調査する。

これらの条件を用いる数と組合せを変更し、探索範囲を拡大せず完全一致のみで事例を探索した場合の性能評価の結果を図13に示す。図13中の表では、条件数が7未満の場合、最もF値が高かった結果を示している。F値が最大となる条件数5では相対音高、前後の音高差、音長、音価、フレーズ内での音符の位置の組合せが最大であった。

最も性能評価が高くなった条件のF値と、その際の $\sigma$ を図14に示す。図14では比較のため、提案したアルゴリ



ズムでは事例探索を行わなかった場合において、探索範囲を拡大する部分を排除した標準偏差  $\sigma = 0$  の場合の F 値も記載している。なお、本実験では  $\alpha$  を 0.01 とし、閾値  $T$  は、相対的にゲインが挿入されにくい 6 割を棄却するため、テストする楽曲に依存して変化させた。各  $\sigma$  の値については、各条件における最大の値が 1 になるようにスケールを統一しており、値が大きいほど探索範囲が広いことを意味している。図 14 を見ると、探索範囲を決める各条件の  $\sigma$  を図中の値に設定した際の F 値は、完全一致 ( $\sigma = 0$ ) の事例だけを探索する場合と比べると性能が向上しているといえる。これは、本実験で使用した楽曲データと 5 条件においては標準偏差  $\sigma$  を設けることで性能評価を向上させることができることを示している。

## 7. 展望

歌唱の品質評価の実験では、歌唱技巧転写のための信号処理が実現されていることが確認された。一方、歌唱技巧の出現位置に関するガイダンスのユーザ評価に関して、6 名の一般男性被験者を実験協力者として実施した初期的なアンケートからは、「ゲイン表現が最も入りやすいとシステムが示唆した音符のうち、実際に挿入してみて、妥当だと思えたのは半分程度であった」「候補の中から適宜選択していくことで、適切なゲイン歌唱が得られた」と総括されるコメントが得られた。これらコメントは、提案システムが、ヒューマンインザループ型のデザイン支援システムとして、使用されうることを示している。一方、第 1 候補での推定率が半分程度にとどまった理由としては、1) 判定の根拠とすべき情報（データベースを構築するため対象楽曲から抽出した楽譜データ）に不足があったこと、2) ゲインを転写した結果の判断基準が被験者により異なること、の 2 つが考えられる。これらの定量的な把握は、大きな困難が予想されるが、音楽の本質の理解につながる魅力的な研究課題でもある。本研究の発展研究の 1 つとして取り組んでいきたい。

本システムの利用にあたっては、歌唱者の歌唱と歌詞（楽譜データ）とのアライメントが必要となる。現時点での実装では、ビートにあわせて歌唱者が歌うというインタフェースを採用しているが、システム実利用の観点からは、より制約の少ない、自動アライメントの導入が望まれる。ショートレンジの課題としては、自動アライメントの導入と、それにあわせて、インタフェースデザインの改良と評価を実施する必要がある。

本論文では、奄美大島風鳥唄の唱法ゲインを対象として、音響信号として入力された歌唱に対して、特定の歌唱法を転写する手法について提案した。本方式は、ゲインに対して実施したものであるが、提案の、1) 歌唱技巧分析・転写のための信号処理、2) 歌唱技巧の出現位置に関するガイダンス、による歌唱デザイン支援のフレームワークは、他の

歌手・ジャンルにおけるピッチと音色の制御により表現される歌唱技巧に関して適用可能である。その先の応用としては、故人の歌手が新しく作られた楽曲を歌ってみたいかどうか、というような使い方も想定される。この目標は、Sinsy でも対応可能であるが、ユーザ意図の反映の手段とガイダンス機能を提供しているという点において、提案システムに優位性がある。歌唱デザインのほかには、技法や歌い回しの習得は困難である [22] とされる民謡指導や学習におけるシステムの利用が想定される。

## 8. おわりに

本論文では、奄美大島出身の歌唱者の歌声を分析し、奄美大島民謡独特の歌唱法であるゲインの音響的特徴を明らかにした。抽出された 1) F0 の隆起と沈降、2) 歌声による倍音成分の欠落、3) F0 の隆起に合わせたスペクトル変動を一般歌唱に転写することで、ゲインを転写できることが示された。また、ゲインを適切な位置に挿入するための手法について、主観評価により有効性を検証した。今後は、インタフェースデザインに関するデザインの改良とその評価、歌声合成におけるパラメータ調整の省力化について検討してゆきたい。

謝辞 本研究は、科学技術振興機構 CREST の支援を受け、CrestMuse プロジェクトの一部として実施された。

## 参考文献

- [1] 剣持秀紀, 大下隼人: 歌声合成システム VOCALOID, 情報処理学会音楽情報科学研究会研究報告, 2007-MUS-72, pp.25–28 (2007).
- [2] 剣持秀紀: 歌声合成技術の動向: 「初音ミク」を支える技術, 日本音響学会誌, Vol.67, No.1, pp.46–50 (2011).
- [3] Kenmochi, H.: VOCALOID and Hatsune Miku phenomenon in Japan, *Proc. InterSinging 2010*, pp.1–4 (2010).
- [4] 中野倫靖, 後藤真孝: VocaListener: 事例を与えてインタラクティブに歌声合成できるメタ歌声合成インタフェース, 情報処理学会インタラクティブ 2009 論文集, pp.65–66 (2009).
- [5] 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, No.50, pp.49–56 (2008).
- [6] Morise, M., Onishi, M., Kawahara, H. and Katayose, H.: v.morish'09: A morphing-based singing design interface for vocal melodies, *Lecture Note in Computer Sciences, LNCS5709: The 10th International Conference on Entertainment Computing (ICEC 2009)*, pp.185–190 (2009).
- [7] 徳田恵一, 大浦圭一郎: 自動学習により人間のよう歌う音声合成システム—Sinsy, 情報処理学会音楽情報科学研究会研究報告, 2012-MUS-94, No.1, pp.1–6 (2012).
- [8] 片寄晴弘: 音楽における自動処理と Directability, オペレーションズ・リサーチ, 音楽と OR 特集, Vol.54, No.9, pp.540–545 (2009).
- [9] 桂 博章: 秋田民謡についての覚書—「小節 (こぶし)」について, 秋田大学教育学部研究紀要 (1998).
- [10] 月溪恒子: 邦楽と洋楽の歌唱表現法—音響的特徴の比較,

- 大阪芸術大学芸術研究所研究調査報告書 (2001).
- [11] 松岡文吾：長唄における韻律の特徴のモデル化とそれに基づく五線譜からの長唄の韻律遷移の生成，東京大学修士論文 (2003).
- [12] 小林範子，東倉洋一，天白成一，新美成二：日本音楽の歌唱における生成的特徴，音講論集 (秋)，pp.245–246 (1989).
- [13] 中山一郎，小林範子：歌の声—声質の魅力と問題点，音響誌，Vol.52, No.5, pp.383–388 (1996).
- [14] Uta-Net INTERVIEW, available from <http://blog.uta-net.com/kizuki/> (accessed 2012-06).
- [15] 中孝介オフィシャルホームページ，入手先 <http://www.atarikousuke.jp/> (参照 2012-06).
- [16] Hajime Chitose Official Website, available from <http://www.office-augusta.com/hajime/> (accessed 2012-06).
- [17] 村主大輔，馬場 隆，森勢将雅，片寄晴弘：奄美大島民謡風歌声合成システム：グインレゾネータ，情報処理学会音楽情報科学研究会，Vol.2012-MUS-094, No.3, p.8 (2012).
- [18] Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T. and Banno, H.: 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).
- [19] Saito, 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. 2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp.215–218 (2007).
- [20] 河原英紀，生駒太一，森勢将雅，高橋 徹，豊田健一，片寄晴弘：モーフィングに基づく歌唱デザインインタフェースの提案と初期的検討，情報処理学会論文誌，Vol.48, No.12, pp.3637–3648 (2007).
- [21] Kawahara, H. and Matsui, H.: Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation, *Proc. ICASSP 2003*, Vol.I, pp.256–259 (2003).
- [22] 虫眞真砂子：日本の伝統的な歌唱授業の試みに関する考察，岡山大学教育実践総合センター紀要 (2006).
- [23] 中里南子：J・ポップにみられる装飾的旋律の歌い方—平井堅・桑田佳祐・ケミストリー・ドリカムの「コブシ」の分析を通して，音楽教育実践ジャーナル，Vol.5, No.1, pp.32–39 (2007).



森勢 将雅 (正会員)

2008年和歌山大学大学院システム工学研究科博士後期課程修了。博士(工学)。関西学院大学，立命館大学を経て，現在，山梨大学工学部テニユアトラック特任助教。聴覚情報処理，音声信号処理，インタフェース設計の研究に従事。電子情報通信学会，日本音響学会各会員。



村主 大輔

2010年関西学院大学理工学部卒業。2012年関西学院大学大学院理工学研究科博士前期課程修了。研究テーマ「奄美大島民謡風の歌唱デザインを支援するシステム「グインレゾネータ」の開発」。



馬場 隆 (学生会員)

2009年くらしき作陽大学音楽学部音楽学科情報音楽専修首席卒業。2011年関西学院大学大学院理工学研究科情報科学専攻博士前期課程修了。工学修士。現在，関西学院大学理工学研究科情報科学専攻博士後期課程。指揮システム“VirtualPhilharmony”の開発をはじめ，音楽とdirectabilityに関する研究に従事。



片寄 晴弘 (正会員)

1991年大阪大学大学院基礎工学研究科博士課程修了。工学博士。イメージ情報科学研究所，和歌山大学を経て，現在，関西学院大学理工学部教授。音楽情報処理，感性情報処理，HCIの研究に従事。科学技術振興機構さきがけ研究 21「協調と制御」領域研究者。科学技術振興機構CREST「デジタルメディア (略称)」領域CrestMuseプロジェクト研究代表者。電子情報通信学会，人工知能学会各会員。