

歌唱力転写システム制作に向けての検討

辰 巳 直 也^{†1} 森 勢 将 雅^{†2} 片 寄 晴 弘^{†1,†3}

Vocaloid「初音ミク」の発売以来、歌唱合成に対する注目が高まりつつある。Vocaloidでは、メロディーと歌詞を入力することにより歌声を生成できる。また、表情パラメータを調整することにより様々な表情を付与することができる。しかし、より人間らしい歌声にするには、表情パラメータの調整を細かく設定することが必要なため、非常に煩雑で時間がかかる。本研究では、あらかじめ、特定の歌唱者 (GACKT) の歌い方にみられるビブラートやポルタメントといった音量、音高等の特徴を表情パラメータとして抽出しておき、それらを Vocaloid の出力に付加することで、より、当該の歌唱者らしい歌い方を実現するような GACKT レゾネータの開発を目指す。

Examination for transcript of how to sing system production

NAOYA TATSUMI,^{†1} MASANORI MORISE^{†2}
and HARUHIRO KATAYOSE^{†1,†3}

Since the release of Vocaloid "Hatsune Miku," voice synthesizing applications have been known to the public people. Vocaloid generates vocal melodies, from the given lyrics and melodies, and provides users with the interfaces for adjusting parameters for expression control. However, setting parameters for elaborating natural expressiveness requires complicated operations. This paper introduces a vocal resonator that makes the vocal melodies more natural by copying the characteristics of a professional singer's (GACKT) singing.

^{†1} 関西学院大学理工学部

School of Science and Technology, Kwansai Gakuin University

^{†2} 立命館大学情報理工学部

College of Information Science and Engineering, Ritsumeikan University

^{†3} 関西学院大学大学院理工学研究科

Graduate School of Science and Technology, Kwansai Gakuin University

1. はじめに

Vocaloid「初音ミク」¹⁾の発売以来、歌唱合成に対する注目が大きく高まっている。Vocaloidとは、サンプリングされた人の声を利用した連結的合成方式に分類される歌唱合成ソフトウェアであり、メロディーと歌詞を入力することによって歌声の生成が行われる。YouTubeやニコニコ動画等の動画共有サイトの普及にも関連して、アマチュアクリエイターが昼夜作品作りに励み、その作品がインターネット上でアップロードされ、さらにその参照コンテンツが連鎖的に作られており、その盛り上がりは一種の社会現象としてとらえられるまでに至っている²⁾。

Vocaloidでは、コントロールパラメータを調整することによって、歌い方に自分なりの表情を付与したり、より人間らしい歌声に仕上げることが可能である³⁾。しかし、この作業は煩雑であり、経験の乏しいユーザにとっては極めて困難である。この問題に対し、ユーザ自身が歌を歌い、その歌い方を反映させるといったシステムの開発も行われてきた⁴⁾。ところが、このシステムを使用するにあたってはユーザ自身がある程度の歌唱力を有していることが求められる。

そこで本研究では、あらかじめ、歌唱者の歌い方にみられるビブラートやポルタメントといった音量、音高等の特徴を表情パラメータとして抽出しておき、それらを Vocaloid の出力に付加することで、より、当該の歌唱者らしい歌い方を実現するようなレゾネータの開発を目指す。今回とした対象のレゾネータ開発の対象は「GACKT」をモデルとして作られた「がくっぽいど」、レゾネータの歌手モデルとしては「GACKT」自身とした。「GACKT」レゾネータの構成に当たっては、「GACKT」自身が実際に歌った「君が代」を分析した結果を用いる。以下、第2章では、「君が代」の分析に基づいたGACKTらしさの特徴について述べる。第3章では、GACKTらしさを転写するための機構について述べる。第4章では、「ふるさと」を対象として「GACKT」レゾネータの適用例を説明する。

2. GACKTらしさの分析

2.1 GACKTと「がくっぽいど」

本研究で対象とする歌手のモデルGACKTは、沖縄の男性シンガーソングライターであ

る。1990年代にヴィジュアル系バンド、MALICE MIZERのヴォーカリストとして活躍していたが、1999年に脱退し、現在はソロで活動している⁵⁾。彼の歌い方は非常に特徴的で、歌い上げるときの語尾の残し方や溜め方等が印象的である。

「がくっばいど」は、歌声ライブラリにGACKTの歌声を用いたVocaloidである⁶⁾。Vocaloidでは、歌声ライブラリに収録された歌手の声質・音高・モーラを譜面に基づいて接続することで歌唱の合成を実施する⁷⁾。しかしながら、歌手ライブラリには歌手特有の「歌い方」に相当するデータが含まれておらず、その部分を実用しようとする、使用者が手作業にてVocaloidの制御パラメータを編集する必要がある。本論文では、Vocaloidによって生成される歌唱が、よりGACKTらしいものになるよう、制御パラメータを自動修正することを目的としている。その前準備として、GACKT自身が歌唱した楽曲からGACKTらしさが表出される部分の分析を実施する。

2.2 分析の概要

人から発せられた声の特徴は、知覚的には高さ、音色と音量に分類される。歌手の声質に相当するパラメータが音色であることから、歌い方の抽出するためには、高さや音量を分析することが求められる。また、歌唱からこれらのパラメータを抽出するためには、他の楽音が存在しない、独唱のデータを入力する必要がある。

本論文では、この条件を満たすGACKT自身が独唱した歌唱として「君が代」を分析の対象とする。また、分析に用いるソフトウェアにはMelodyneを用いて抽出された音高と音量に着目し、GACKTらしさに相当するパラメータの抽出を行う。分析より得られた特徴を用いて、Vocaloidに打ち込まれた別の楽曲にGACKTらしさを転写するためのルール構築を試みる。

2.2.1 分析する歌唱の特徴

本研究における歌声の合成エンジンはVocaloidが利用される。Vocaloid外部パラメータとして制御可能な歌唱特徴は音高の変化と音量の変化であることから、本研究では、ビブラートとポルタメントと、歌唱の強弱表現に着目してGACKTらしさの分析を実施する。

ここでは、Vocaloidにおいて音高の変化として精密に制御可能であるビブラートとポルタメントと、演奏の強弱に相当する音量を対象に、GACKTらしさの抽出を行う。

ビブラートは、(楽譜上表記上)の音高を一定としながら、音高を変動させる演奏表現と

定義される^{*1}。GACKTレゾネータでは、どの歌詞をどのようなビブラートで演奏すればよいか、ルールを設定する必要がある。ポルタメントは、ある音高から別の音高に遷移する際、滑らかに変化させる演奏表現として定義される。ここでは、GACKTは歌唱において、ビブラートやポルタメントの表現にあたって、どのように音高・音量の制御を行っているのか、また、その表現は、音符の情報(音価や音高の変化量)と歌詞とどのような関係があるのかの分析を行う。

2.2.2 Melodyneを用いた分析と歌唱特徴の抽出

Melodyneは、入力された歌唱から、歌唱の音階、音高の微細変動、音量を抽出し、それぞれのパラメータを編集し再合成する機能を有している⁸⁾。解析された歌唱の音高と音量は図1のような形式で表示される。この図において、横軸は時間、縦軸は音階、図中に描かれる細い赤線は音高の詳細な情報、太線は音階を表し、音量は太線の太さによって表現されている。また、Melodyneには、これらのデータをsmf形式でexportするための機能が提供されている。

分析においては、GACKT自身が独唱した楽曲「君が代」を対象とし、第一著者がMelodyne表示画面の目視およびexportされたデータの分析に基づいて、ビブラートやポルタメントにおける制御パラメータの同定を進めた。また、この分析に基づき「がくっばいど」の制御パラメータを手修正して、実演奏と聞き比べることで、分析をさらに精緻なものへと近づけて行った。これらの制御パラメータとその制御が行われる条件を経験的に設定することで、GACKTらしさの元となるルールの集積をはかった。

2.3 GACKTの歌い回しの特徴

図2にGACKTの典型的な歌い回し事例を示す。図2のa)は、君が代における歌詞「君が代は」の/kimi/の分析結果を示す。/k/から/i/へ音階が変化していることは、譜面よりも意図的に低い音高で/k/を歌唱し、ポルタメントにより/i/の高さへと遷移したことを示している。/m/から/i/についても同様に、/m/では、譜面上の音階よりも低い音高から、目的とする高さへとポルタメントにより遷移させている。また、どちらの/i/も数秒の持続時間を有するが、ビブラートを表現していないことも分かる。

(b)は、君が代における歌詞「君が代は」の/ga/の分析結果を示す。(a)の分析結果と同

*1 実際の演奏表現を精緻に分析すると、音高だけではなく音量にも変動が入っていることが確認される

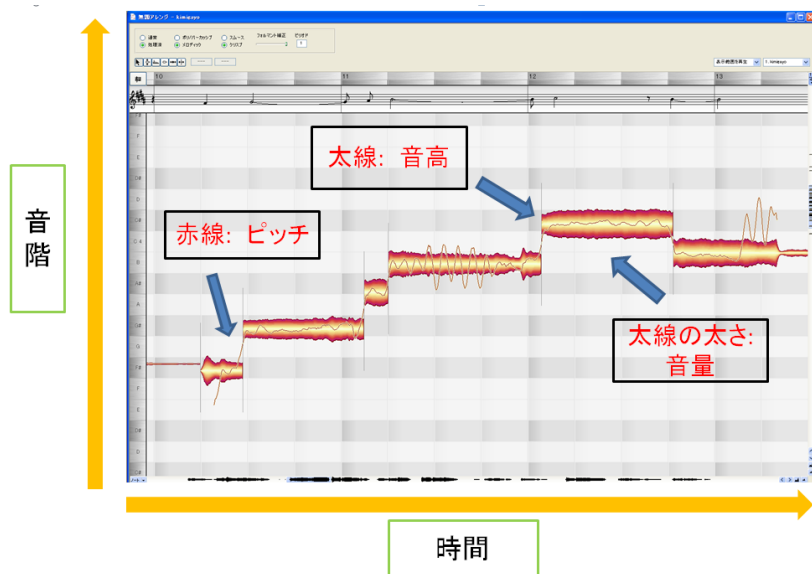


図1 Melodyne の説明

Fig. 1 Explanation of Melodyne

様に、/g/から/a/へポルタメント気味に遷移させていることが分かる。また、譜面では/a/の前半と後半で音階が遷移するが、遷移の前に音量を下げていること、遷移後にビブラートを表現していることが分かる。

(c)は、君が代における歌詞「君が代は」の/yo/の分析結果を示す。/yo/では、(b)における/a/の音階のまま/y/を発声し、ポルタメントにより高さを遷移させていることが分かる。/o/については、前半と後半で音階が異なる。この音高遷移については、遷移前の音量の減少は認められないが、遷移後のビブラートは認められる。

「君が代」の分析から得られた歌い回しに関するルールを以下にまとめる。

ビブラートとポルタメントに関するルール

- (1) フレーズの初めの音が低音域（以後低音域は B3 以下、高音域は C4 以上とする）のモーラに対し、（ビブラートの）振幅の大きいビブラートをかける。
- (2) フレーズ内で、対象となるモーラの音階が、その1つ前の音階以下である場合、ビブ

ラートを付与する。また、ビブラートは音の途中からかける。

- (3) 低音域で歌い始める場合、そのモーラは譜面の音階よりも低い音高から遷移するポルタメントをかける。
- (4) フレーズの終りのモーラには、ビブラートを付与する。また、ビブラートは音の始めからかける。
- (5) 低音域で長音符号 (-) がある場合、伸ばす後半のモーラへの遷移においてポルタメントをかけ、遷移後にビブラートを付与する。
- (6) フレーズの最後で高音域、さらにビブラートがかけられている場合、ビブラートの後にピッチの大きな減少が起こる、それと同時に音量も減少する。

音量に関するルール

- (1) 無声破裂音や破擦音、両唇鼻音/m/でかつ低音域の場合、その音素の音量を増加させる。
- (2) フレーズの終わりのモーラは、緩やかに音量を下げる。
- (3) 対象となるモーラの音階が、その1つ前の音階よりも低い場合、遷移時に音量を減少させる。
- (4) 対象となるモーラの音階が、その1つ前の音階よりも高い場合、遷移時に音量を増加させる。

3. GACKT レゾネータの構成

3.1 GACKT レゾネータ概要

この章では、2章にて述べた GACKT らしさを表す 10 のルールに基づいて、「がくっぼいど」における Vocaloid パラメータを修正し、より GACKT らしいものとする「GACKT レゾネータ」について述べる。

GACKT レゾネータでは、対象となる楽曲の歌詞ならびにメロディの情報を、Vocaloid の外部出力データ形式である VSQ ファイルから読み取る。続いて、2.3 節にて示されたルールと VSQ ファイルに含まれる音符推移と各モーラとの照合を行い、条件を満たす部分をユーザに提示する。特定の音符に複数のルールが当てはまった際、排他的な判断が求められる可能性がある。ユーザが排他判断をしなくてもシステムが動作するよう、各ルールには優先度が設定されているが、必要に応じてユーザはマニュアルでルールの適用条件を設定する

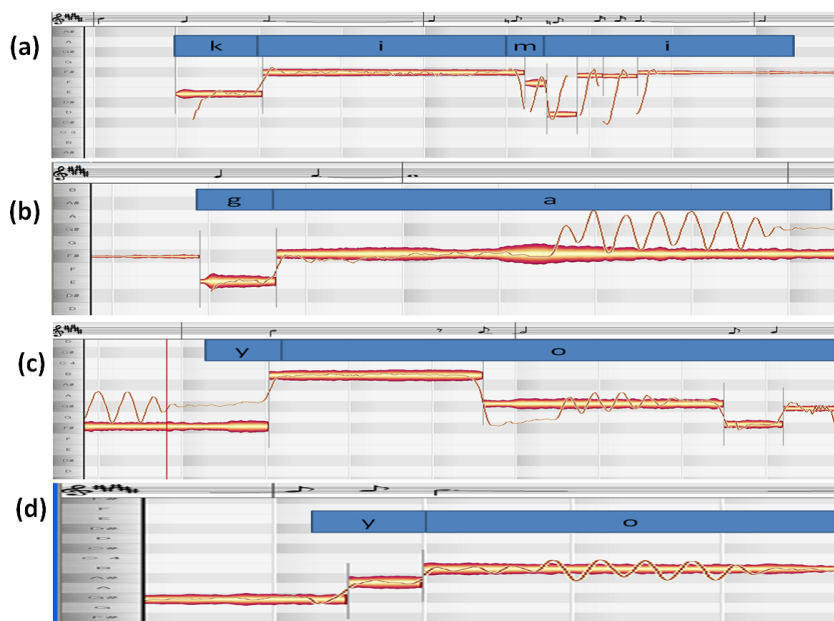


図2 GACKTらしい歌い方の特徴箇所の一部(1)

Fig. 2 Example of feature of how to sing GACKT(1)

ことができる。以上により、適用されるルールが決定し、対応する部分のVSQ中の音高および音量パラメータを書き換えていくことによって、表情付けのためのパラメータが生成される。書き換えられたVSQファイルを再度Vocaloidで読み込み、歌唱合成を行うことで、より、GACKTっぽさが強化された歌唱が得られる。

3.2 適用するルールの決定

GACKTレゾネータでは、全てのモーラについて、10のルールの1つを適用することで歌唱の制御を行う。ここでは、ルールの重複時における適用ルールの選択について述べる。

本論文にて提案した10のルールは、特にビブラートに関して重複が生じる。例えば、現時刻の音高が1つ前の音高以下の場合にビブラートを付与するルールは、フレーズの終わ

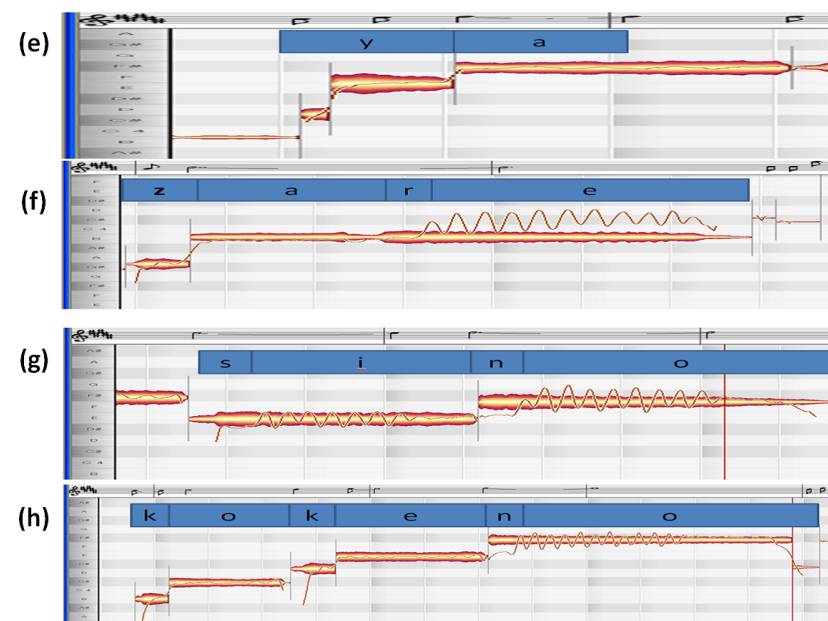


図3 GACKTらしい歌い方の特徴箇所の一部(2)

Fig. 3 Example of feature of how to sing GACKT(2)

りにビブラートを付与するルール、さらに、フレーズ最後の音の音高が高い場合にかけられる特殊なビブラートを付与するルールと重複する可能性が高い。本論文では、ルール抽出時に得られたGACKTらしさに関する知見に基づき、フレーズの始まり・終わりに関するルールは、音高遷移に関するルールよりも高く優先度を設定した。

ただし、君が代におけるルール抽出の検討より、条件を満たした場合においてもルールが適用されないモーラがあることを確認している。現在のGACKTレゾネータでは、条件を満たす音節全てに10のルールを適用するよう実装されているが、使用者が、ルールを適用するか否か、あるいは意図して別のルールを適用するための機能を実装する必要がある。これらの機能については、ユーザインタフェースの節にて述べる。

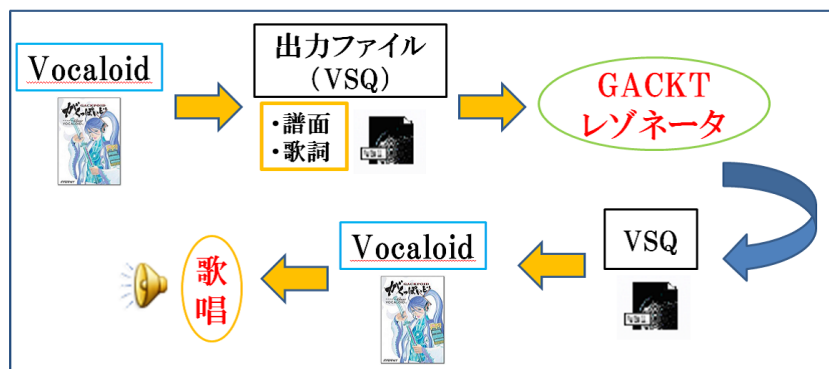


図 4 システムの概要図

Fig. 4 Outline chart of the system

3.3 ユーザインタフェース

GACKT レゾネータは、図 4 に示される手順で実行される。3 章で述べた通り、GACKT レゾネータでは、適用するルールが重複する場合、優先度を用いて適用されるルールを一意に定める。しかしながら、使用者が適用するルールを変更する、あるいは意図的にルールを適用しないという選択を GUI により視覚的に行う。この GUI により、使用者は転写された結果が容易に操作できるため、Vocaloid における歌唱作りこみ作業の効率化が期待される。

4. GACKT レゾネータの適用

本章では、文部省唱歌「ふるさと」一番「兎追ひし かの山 小鮒（こぶな）釣りしかの川 夢は今も めぐりて 忘れがたき 故郷（ふるさと）」を題材とし、GACKT レゾネータによる歌唱生成状況について説明する。

図 5, 6 は、ふるさとの楽譜情報と歌詞を GACKT レゾネータに入力し、出力された VSQ ファイルをがくっばいどにて読み込ませたエディタを示す。がくっばいどのエディット画面は上下の 2 つから構成される。上には音階、持続時間、歌詞情報が表示され、下には、音高や音量の微細変動に関する情報が表示される。どちらも横軸は時間を表す。GACKT レゾ

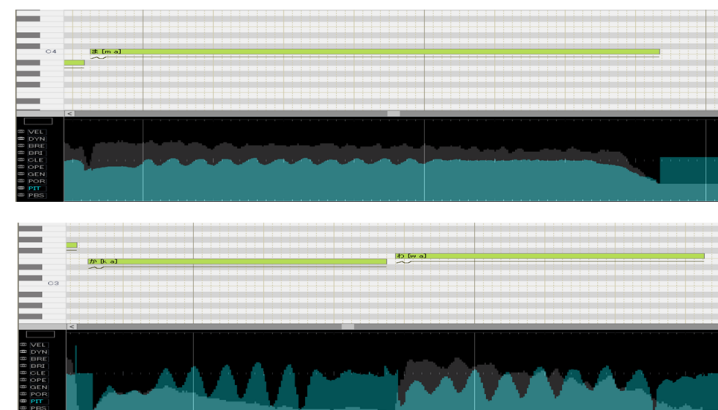


図 5 GACKT レゾネータの適用後 (上図) かの山 (下図) かの川

Fig. 5 After applying GACKT Resonator (figure above) kanoyama (figure below) kanokawa

ネータでは、下の微細変動にルールを適用することで、歌い方の転写を行う。

図 5 に、GACKT らしい表情付けが行われていると判断された例を示す。図 5 の上図では、高音域でフレーズの最後のため「ま」にビブラートの付与とその後での、ピッチと音量の減少に関するルールが転写され、知覚的にも GACKT に近づいていることが確認できた。また下図では、低音域で対象となるモーラの音階が、その 1 つ前の音階以下であるため「か」に遅いポルタメントやビブラートが付与されている。また、「わ」はフレーズの最後のためビブラートが付与され、音量の減少が確認できる。

しかしながら、ルール通りに Vocaloid パラメタが転写されているにも関わらず、合成歌唱が GACKT らしく知覚されない例も存在した。図 6 は、転写後の歌唱が GACKT らしくないと知覚された一例のエディット画面を示す。まず、図 6 の上図では、低音域で長音記号があるため、ポルタメントの後にビブラートが付与されるというルールが適用される。図より、転写は適切に行われていることが確認できるが、知覚的には GACKT らしい歌唱とは感じられなかった。この原因として、このルールには後ろの音高や音韻に依存している可能性が考えられる。また、下図では、同じ音階のノートが 3 つ並んでおり、2 つ目の「さ」と 3 つ目の「ぎ」に適用するルールは全て同一となる。個々の音を切り出して聴取した場合は GACKT らしい表情パラメタが付与されたように知覚されるが、3 モーラを連続して聴

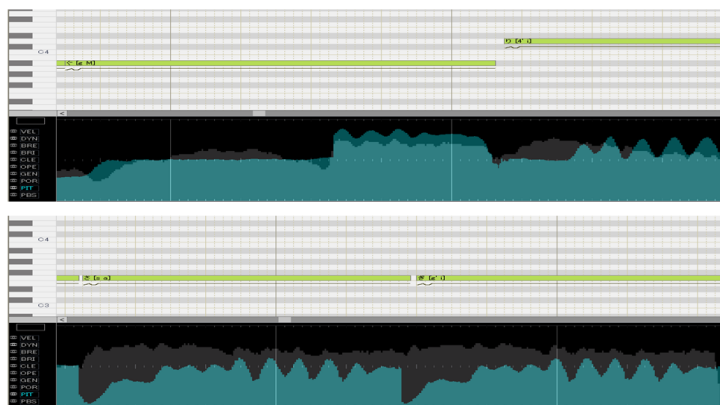


図 6 GACKT レゾネータの適用後 (上図) めぐりて (下図) 兎追ひし

Fig.6 After applying GACKT Resonator (figure above) kanoyama (figure below) kanokawa

取すると人間の歌唱として不自然に知覚された。

本実験結果をまとめると、フレーズの最後にビブラートを適用するルールにより GACKT らしさを向上できるといえる。しかしながら、音高遷移後のビブラートに関しては、モーラ単独での聴取では GACKT らしさを感じられるが、前後のモーラを含めた聴取では GACKT らしさよりも歌唱として不自然に知覚された。この結果は、音高遷移後にビブラートを付与するルールについては、より多くの事例を分析し多くのルールを構築する必要性を示している。

5. まとめと今後

本稿では、Vocaloid において煩雑な作業となる歌唱表現の作り込み支援を目的とし、プロ歌手の歌い方を、Vocaloid パラメタとして転写するための方法について述べた。本検討は、Vocaloid の歌手ライブラリに音声を提供し、なおかつ本人もプロの歌手である「GACKT」を対象に行われた。GACKT 自身が独唱した歌唱である「君が代」の音高と音量を Melodyne にて抽出し、別の楽曲に GACKT らしさを転写するための表現が、譜面や歌詞とどのような関係にあるのか分析した。得られた定性的な特徴に基づき、Vocaloid の制御パラメタと

して転写するためのルールを構築した。

君が代以外の楽曲に提案したルールを適用し、得られた楽曲を聴取したところ、GACKT らしさが向上する例と GACKT らしさが向上しない例があることを確認した。君が代では特にビブラート表現に関する事例数が不足していたため、不適切なルールを適用したことが原因といえるだろう。

今後は、様々な楽曲を対象として、どのような楽曲においても適用できるよう GACKT らしさのルールを増加させることが必要となる。また、これらのルール適用を容易に行うユーザインタフェースの実装と評価も併せて行ってゆきたい。

謝辞 本研究の一部は、科学技術振興機構による戦略的創造研究推進事業のデジタルメディア領域 CrestMuse プロジェクトの支援を受けて行われた。

参考文献

- 1) <http://www.crypton.co.jp/mp/pages/prod/vocaloid/>
- 2) <http://www.yamaha.co.jp/news/2009/09042701.html>
- 3) <http://www.ipsj.or.jp/sigmus/PAPERS/DTMM200808nakano.pdf>
- 4) 中野倫靖 他: VocaListener: ユーザ歌唱とその歌詞を用いた歌声合成パラメータの自動推定システム. 日本音響学会 2008 年 秋季研究発表会 講演論文集.
- 5) <http://gackt.com/jp-h/biography/1999.html>
- 6) <http://www.ssw.co.jp/products/vocal/gackpoid/gaiyo/index.html>
『情報処理学会研究報告』Vol.2008, No.12, 情報処理学会, 51 頁
- 7) 「剣持秀紀 interview」、『DTM magazine』第 15 巻 2 号 (通号 166) 「1 月号増刊 The VOCALOID CV01 初音ミク」、寺島情報企画、2008 年 1 月、36 頁
- 8) <http://www.celemony.com/cms/index.php?id=home>