

歌唱特徴付与システム「ロックボーカルレゾネータ」

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

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、メロディと歌詞を入力することにより、サンプリングされた人の声を元にした歌声を合成することができる。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、より人間らしい表情豊かな歌声にするには、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、ロック歌手の一人「GACKT」の歌い方に見られるピブラートやボルタメントといった音高・音量等の歌唱技法を低次のモデルパラメータで近似し、混合ガウス分布を用いた手法でモデルパラメータを決定する。それらの値をVocaloidの出力に付加することで、ロック歌手らしい歌い方を実現する「ロックボーカルレゾネータ」を提案する。

Rock Vocal Resonator: A Rule-based Singing Design System Featured a Rock Singer's Characteristics

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

Since the release of Vocaloid "Hatsune Miku", voice synthesizing applications have been known to the public people. Vocaloids generate human-like vocal melodies by giving lyrics and melodies. If the parameters for expression control are elaborated, Vocaloids yield more natural vocals. However, setting these parameters requires complicated expertise. This paper proposes a system called "Rock Vocal Resonator" that emphasizes Rock vocalist styles for Vocaloids, based on analysis of a Japanese Rock singer, GACKT's 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

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、Vocalistener citeVocalis や Sinsy citeSinsy といったシステムが提案されてきた。Vocalistenerでは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であるがユーザ自身がある程度の歌唱力を有している必要がある。Sinsyは、音声合成方式のひとつである「HMM音声合成」を歌声の合成に応用したもので与えられた歌唱データに基づいてモデルを自動学習することにより、歌唱者の声の特徴を再現する合成歌唱を得ることができるシステムである。しかし、Sinsyではロック歌手特有のボルタメントといった歌唱技法に対応していない点が挙げられる。

そこで本研究では、ロック歌手らしい歌い方をロックボーカルレゾネータの開発を提案する。ロックボーカルレゾネータではあらかじめ、ロックの歌い方にみられるピブラートやボルタメントといった特徴的な歌唱技法を低次のモデルパラメータで近似し、混合ガウス分布を用いた手法でモデルパラメータを決定する。また歌唱技法が現れる条件を経験則的に設けた。ルールの条件節と制御節によりVocaloidに表情パラメータを付加することで、簡単にロック歌手らしい歌い方を実現するような「ロックボーカルレゾネータ」の開発を目指す。今回、分析対象としたロック歌手はボルタメントやピブラートの深さが大きく歌い方が特徴的なGACKTとした。また、レゾネータ開発の対象はGACKTをモデルとして作られた「がくっばいど」を用いた。「ロックボーカルレゾネータ」の構成に当たっては、GACKT自身が実際に歌った楽曲を分析した結果を用いる。以下、第2章では、楽曲の分析に基づいたロック歌手らしさの特徴について述べる。第3章では、ロック歌手らしさを付与するための機構につ

いて述べる。第4章では、本システムの初期的検討を行う。

2. ロック歌手らしさの分析

2.1 GACKTと「がくっぽいど」

本研究で対象とする歌手のGACKTは、沖縄県出身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⁷⁾。彼の歌い方は非常に特徴的で、フレーズ始めの語尾の残し方や溜め方等において独特の特徴を持つ。

「がくっぽいど」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、使用者が手作業にてVocaloidの制御パラメ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、簡易に、よりロック歌手らしいものになるよう、歌唱技法の挿入位置と表情パラメタの決定を行うシステムの開発を目的としている。その前準備として、GACKT自身が歌唱した楽曲からロック歌手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、知覚的には高さや音色、音量に分類される。歌手の声質に相当するパラメ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、この条件を満たし、かつ、GACKTの特徴的な歌い方がよく表れている「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を利用し、ロック歌手らしさに相当するパラメタの抽出にはSTRAIGHTを利用した。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ロック歌手らしさを付与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、ピブラートとポルタメントと、歌唱の強弱表現に着目してロック歌手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ピブラートとポルタメント、演奏の強弱に相当する音量を対象に、ロック歌手らしさの抽出を行う。

ピブラートは(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ピ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、ピ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 分析方法と演奏ルール構築

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。

ルールの構築には、第一著者がMelodyne表示画面の目視およびSTRAIGHTの基本周波数推定の値に基づいて、ピブラート・ポルタメントにおけるモデルパラメタの同定を進めた。STRAIGHTでは高品質で音声进行分析、変換、合成するツールである⁹⁾。

時刻 $t=T1, \dots, T2$ においてピブラート区間とすると、ピブラートは次式でモデル化を行った。

$$(A, \theta) = A * \sin(2\pi\theta(t - T1)); \quad (1)$$

Aはピブラートの振幅、 θ はピブラートの周期を表わす。
また、ポルタメントに関しては、次式でモデル化を行った。

*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

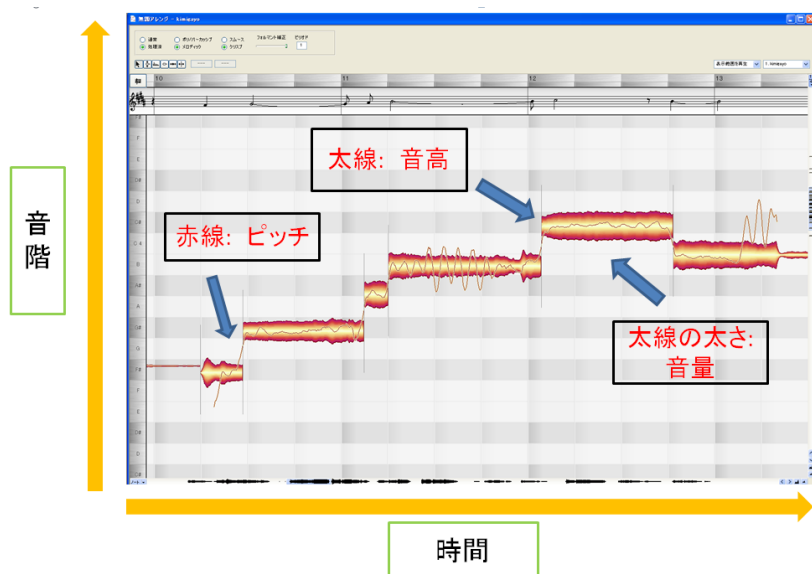


図1 Melodyne の説明

Fig.1 Explanation of Melodyne

$$p(a, c) = 0.5 * a^{\exp(-ct)}; \quad (2)$$

a はポルタメントの深さ, c は立ち上がりの時間を表す. これらのモデルパラメタとその音符の音価, 音長の下で, クラスタ数 k 個の混合ガウス分布のフィッティングを EM アルゴリズムを用いて行い, ルールの制御節を構築した. またルールの条件節を経験的に設定することで, ロック歌手らしさの元となるルールの集積をはかった.

2.3 GACKT の歌い方の特徴

図2にGACKTの典型的な歌い直し事例を示す. 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

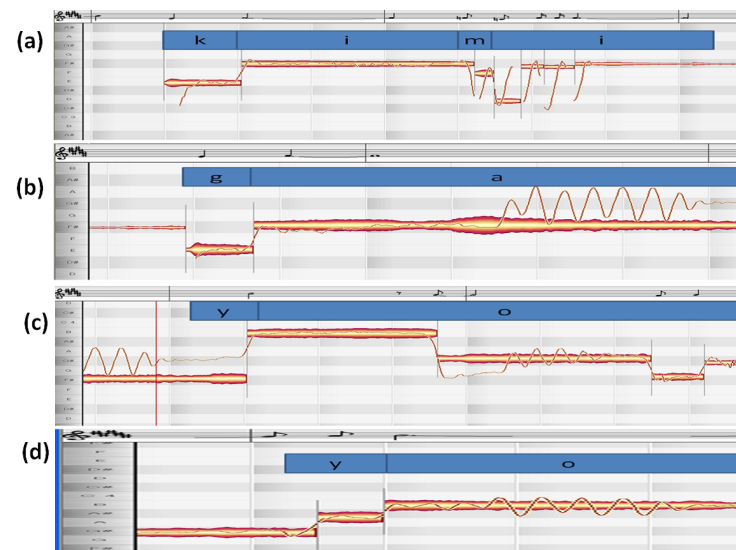


図2 GACKTらしい歌い方の特徴箇所

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

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へと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直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のモーラに対し, (ビブラートの) 振幅の大きいビブラー



図3 楽曲のフレーズングの一例

Fig. 3 A exaple of phrasing

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、ピブラートを付与する。また、ピ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、ピブラートを付与する。また、ピブラートは音の始めからかける。
- (5) 低音域で長音符(-)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、遷移後にピブラートを付与する。
- (6) フレーズの最後で高音域、さらにピブラートがかけられている場合、ピ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。

2.3.1 フレーズの定義

本稿では、フレーズは歌唱者がブレスを入れる位置と定義する。楽曲のブレス位置での分割(以後、フレーズング)は、与えられた Voaloid ファイルの楽譜情報から、休符と音価・音高、形態素を利用して求める。形態素の解析には日本語形態素解析「Sen」を利用した。楽曲のフレーズング方法を以下に示す。

- (1) 休符の位置をフレーズの区切りとする。

- (2) (1)で決定したフレーズ区間がある一定以上空く場合、フレーズ内で最も音価の大きい音符の後をフレーズの区切りとする。一意に決まらない場合、音価と現在の音符から次の音符への音高遷移度数の大きい箇所をフレーズの区切りとする。

(但し、形態素を分割するようなフレーズングはしないとする。)

(2)の方法では、息の持続時間を制約条件とし、エネルギーを多く使う(例えば、音価の長い)音符の後ろ等にフレーズングを行う。上記の方法で、GACKTの楽曲「12月のLoveSong」の数小節にフレーズングした例を示す。赤い線は休符の位置でフレーズを決定したものである。次に、緑の線は(2)の方法によってフレーズングを行った。最後に青の線は、ある一定以上のフレーズ区間がまだ残されているため、さらに(2)の方法でフレーズングを行ったものである。

3. ロックボーカルレゾネータの構成

この章では、2章にて述べたルールに基づいて、「がくっばいど」における Vocaloid パラメタを修正し、簡易によりロック歌手らしいものとする「ロックボーカルレゾネータ」について述べる。

3.1 ロックボーカルレゾネータ概要

ロックボーカルレゾネータでは、対象となる楽曲の歌詞ならびにメロディの情報を Vocaloid の外部出力形式である VSQ ファイルから読み取る。続いて、2.3.1節で定義したフレーズング手法を用いて、フレーズ箇所を推定する。フレーズの挿入位置は人によって異なる場合があるため、ユーザによってはフレーズングの結果に不満がでる場合がある。そのため、本システムではフレーズの位置をユーザが変更できるようにしている。

決定されたフレーズ情報と2.3節にて示されたルールより、旋律の音高推移と各モーラとの照合を行い、特定の歌唱技法の入る箇所をユーザに提示する。ルールの条件節により、特定の歌唱技法が一つの音符に複数当てはまる場合、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ことができる。ルールの制御節により、2.2.2節で求めた混合ガウス分布のパラメタより現在の音符の音高・音長のもとで最尤パラメタを決定し確率の高いパラメタから順にユーザに提示する。ユーザは提示されたパラメタのいずれかを選択・適用することで、表情パラメタを付与する。また、

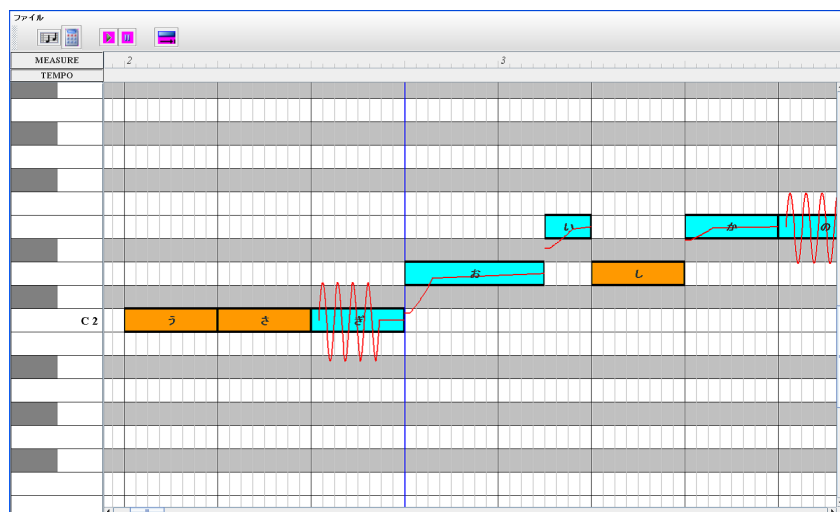


図 4 ロックボーカルレゾネータ システム画面

Fig. 4 Rock Vocal Resonator: System screen

ユーザが歌唱技法の表現を誇張・抑制を微調整できるように、適用されたパラメタの深さを調整することができる。

生成されたパラメタを適用した Vocaloid ファイルより歌唱合成を行うことで、簡易によりロック歌手らしい歌唱が得られる。

この GUI により、使用者は適用された結果を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。

3.2 システム画面

ロックボーカルレゾネータのシステム画面の図 5 は、ロックボーカルレゾネータのシステム画面になる。縦軸は音階情報、横軸は時間情報を表わす。ノート上の赤い細線は音高の揺らぎを表わす。またオレンジ色のノートには表情パラメタの適用が行われず、シアン色のノートには表情パラメタの適用が行われている。シアン色のノートを選択することで、表情パラメタを変更することができ、また赤い細線をドラックして上下に動かすことで、表情パラメタの伸縮を行う事ができる。

3.3 適用するルールの決定

ロックボーカルレゾネータでは、歌詞中の全てのモーラについて、2.3 節のルールの条件に該当する箇所に、最尤パラメタ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。

2.3 節にて提案したルールは、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。現在のロックボーカルレゾネータでは、条件を満たす音節全てに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。

4. システムの初期的検討

本章では、システムの初期検討を行う。システムを通して、ロック歌手らしい歌い方になるようにパラメタを変更した歌唱がロック歌手に近付いているかの確認・また GUI の主観的評価を行った。

入力とした楽曲は文部省唱歌「ふるさと」である。ロック歌手らしい歌い方には図 5 でのボルタメントが付与されている箇所は第一著者の知覚的にロック歌手に近付いていることが確認できたが、ビブラートの付与されている箇所は、制御パラメタの深さがほぼ同一で、連続してモーラを聴取すると、ロック歌手らしくないと知覚された。この表情パラメタを GUI によりビブラートの深さを変更することで、ある程度、ロック歌手らしい歌い方に変更することができた。しかし、ビブラートの深さの変更だけでなく、速さ、立ち上がり時間もある程度、調整できるようにすることで、より簡易にロック歌手らしい歌い方を作れると考えらる。また、ビブラートモデルの深さ・速さが動的に変化していないため、ロック歌手らしい歌い方にはならない箇所がある。この点を踏まえ、ビブラートモデルの検討を行う必要があると考えられる。

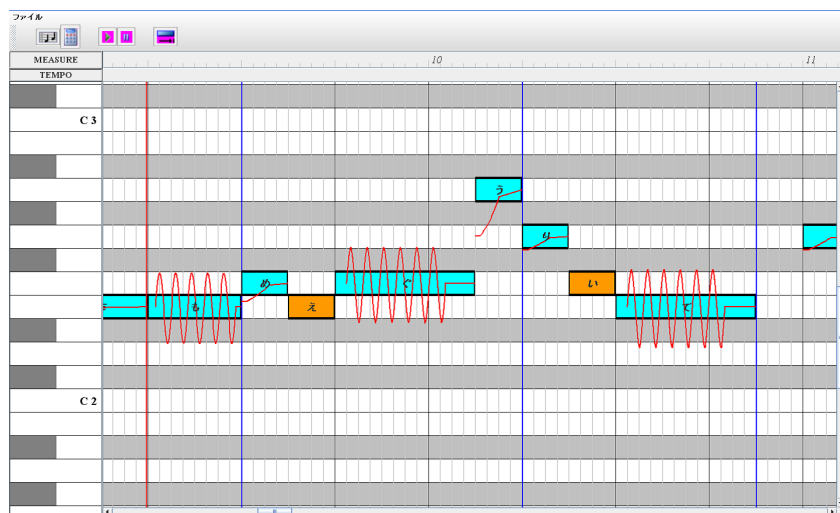


図 5 ロック歌手らしい表情パラメタの適用後「夢はいまもめぐりて」

Fig. 5 After applying an expression parameter

5. おわりに

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, ロック歌手らしい歌い方を実現する「ロックボーカルレゾネータ」を開発した. 本検討は, Vocaloid の歌手ライブラリに音声を提供し, なおかつ本人もロック歌手である「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 STRAIGHT で抽出し, 別の楽曲にロック歌手らしさを付与するための表現が, 譜面や歌詞とどのような関係にあるのか分析した. 得られたルールの条件節と制御節に基づき, Vocaloid の制御パラメタとして付与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. ピブラート表現に関するモデルの構築を静的にすることで, ピブラートの多様性を抑えていたため, 動的に変化するピブラートの速さ・深さをモデル化する必要がある. また, パラメタの深さだ

けでなく, 立ち上がり速度などのパラメタも調整できるようなインターフェースも用意していく必要がある.

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に 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.

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) Sinsy: 「あの人に歌ってほしい」をかなえる HMM 歌声合成システム情報処理学会研究報告 2010 年 7/28
- 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>
- 9) 河原英紀, Vocoder のもう一つの可能性を探る音声分析変換合成システム, ” 日本音響学会誌, vol.63, no.8, pp.442-449, 2007.