

ロックボーカルレゾネータ Vocaloid 歌唱をロックボーカリスト風の歌い方に 変換するシステム

辰 巳 直 也^{†1} 馬 場 隆^{†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、Vocaloid 歌唱をロック歌手風の歌い方に変換し、Vocaloid 歌唱における作り込みを支援する「ロックボーカルレゾネータ」の提案をする。

Rock Vocal Resonator A rule-based singing design system to provide vocaloid singing with a rock singer's style

NAOYA TATSUMI,^{†1} TAKASHI BABA,^{†1}
MORISE MASANORI^{†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. In this paper, we suggest a system called “Rock Vocal Resonator” which gives Rock vocalist styles to vocaloid singing and supports the finish in Vocaloid.

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²⁾。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なっている。

Vocaloid では、コントロールパラメ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てゆくことが可能である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、Vocalistener⁴⁾ や Sinsy⁵⁾ 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ポルタメントといった歌唱技法に対応していない点が問題点として挙げられる。

本研究では、Vocaloid 歌唱をロック歌手らしい歌い方に変換するシステム「ロックボーカルレゾネータ」を提案する。ロックボーカルレゾネータでは、ビブラートやポルタメントといった歌唱技法を楽譜情報のどの位置に挿入するか、歌唱技法、音量の制御をどうするかといった問題に対し、歌唱ルールを構築することで、歌唱技法の挿入位置(条件節)、歌唱技法と音量の制御(制御節)の決定を行う。歌唱ルールの条件節は、対象とするロック歌手の楽曲を基に、ヒューリスティックなルールの設定を行い、歌唱ルールの制御節では、ロックの歌い方にみられる特徴的な歌唱技法を低次のモデルパラメタで近似し、混合ガウス分布を用いた手法でモデルパラメタを決定する。歌唱ルールにより Vocaloid に表情パラメタを付

^{†1} 関西学院大学大学院理工学研究科

Graduate 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

加することで、簡易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、分析対象とするロック歌手はポルタメントやビブラートの深さが大きく歌い方が特徴的な GACKT とし、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」を用いる。以下、第 2 章では、Vocaloid の課題とその課題に対する問題設定について述べる。第 3 章では、GACKT の歌唱音声の分析に基づき、ロック歌手らしさの特徴について述べる。第 4 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 5 章では、本システムの初期的検討を行う。

2. Vocaloid の課題と問題設定

2.1 GACKT と「がくっばいど」

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、フレーズの始めの語尾の残し方や溜め方等に独特の特徴を持つ。

「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⁸⁾。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⁹⁾。2011 年には、Vocaloid の新バージョン Vocaloid3 が開発され、その歌声はよりリアルなものとなった。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、歌手特有の歌い方を表現するには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。

2.2 問題設定

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ータの決定を行うシステムの開発を目的とする。ユーザが本システムを利用するにあたり、歌唱技法を容易に操作できるようにするために、その歌唱技法を良く表わす低次のモデルを導入する。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。

3. GACKT の歌唱の分析とルール化

3.1 GACKT の歌唱の分析

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

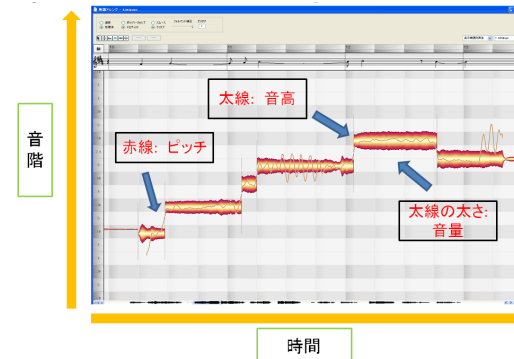


図 1 Melodyne の実行画面と説明

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、または、ヴォーカルパート以外の楽音を抑圧したデータを入力する必要がある。

本論文では、この条件を満たす楽曲として、GACKT 自身が独唱したボーカル音声「君が代」と GACKT の楽曲のオリジナルトラックとカラオケトラックの差分を取り、ヴォーカルパート以外の楽音を比較的抑圧できたデータ「12 月の Lovesong」「君に逢いたくて」「君が追いかけた夢」「月の詩」を分析対象とした。分析に用いるソフトウェアには Melodyne¹⁰⁾、ロック歌手らしさに相当するパラメータの抽出には STRAIGHT¹¹⁾ を利用した。Melodyne は入力された歌唱から、歌唱の目標音高、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、また、それぞれのパラメータを視覚的に捉えることができる。Melodyne によって、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、太線は目標音高を表し、音量は太線の太さによって表現されている。また、STRAIGHT は高品質で音声进行分析、変換、合成するツールである。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。

3.2 GACKT の歌い方の特徴

図 2 に GACKT の典型的な歌い方の事例（ビブラート、ポルタメント、音量の変化）を示

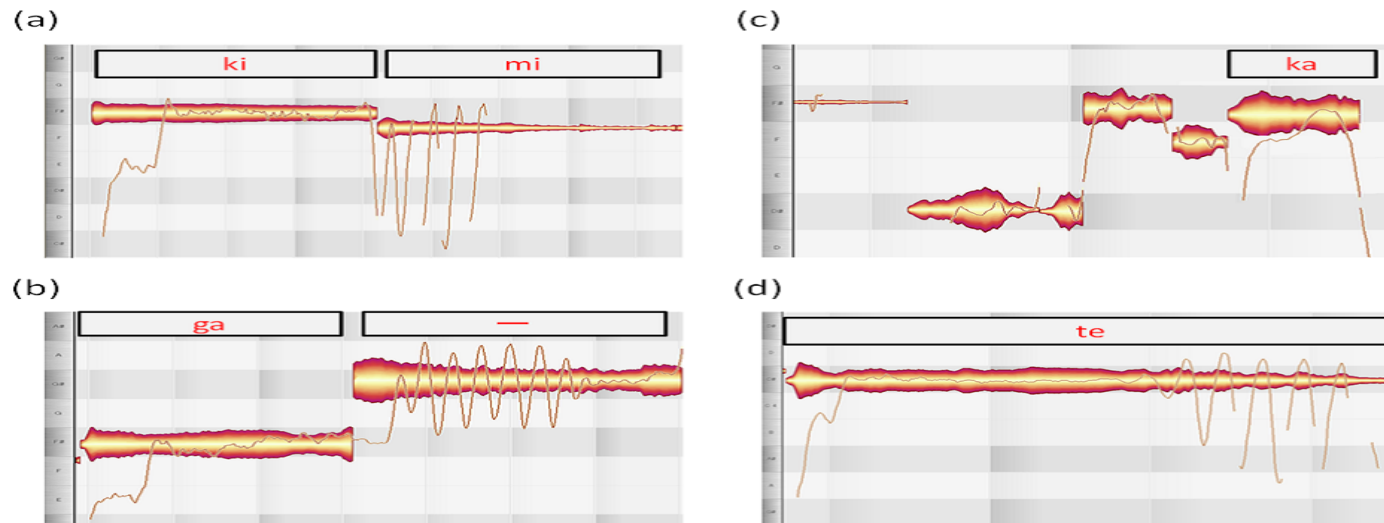


図 2 GACKTらしい歌い方の特徴

す。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^{*2}。また、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。図2の(a)は、君が代における歌詞「きみがーよーは」の「きみ」/kimi/の分析結果を示す。/ki/の歌唱では、譜面よりも低い音高で歌唱してから、ポルタメントにより目標音高へ遷移したことが分かる。/mi/の歌唱では、すぐにピブラートを表現していることが分かる。図2の(b)は、君が代における歌詞「きみがーよーは」の「がー」/ga-/の分析結果を示す。(a)と同様に、/ga-/の歌唱では、譜面よりも低い音高で歌唱してから、ポルタメントにより目標音高へ遷移したことを示している。/-/の歌唱では、/ga-/の音高から/-/の目標音高へ遷移し、その後ピブラートを表現している。また、目標音高へ遷移する前に音量を上げてから、目標音高へ遷移している。図2の(c)は、12月のLovesongにおける歌詞「ひとだかーら」の「かー」/ka-/の分析結果を示す。(c)では、当該音符の一つ前の音符の音高で/ka-/を歌唱してからポルタメントで目標音高へ遷移していることがわかる。図2の(d)は、君に逢いたくてにおける歌詞「きみにあいたくて」の

*2 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。

「て」/te/の分析結果を示す。(d)では、目標音高で歌唱した後に、途中からピブラートをかけていることが分かる。

(a)(b)(c)(d)以外の箇所でも、このような歌唱特徴がみられる。そこで、本研究では、ピブラート、ポルタメント、歌唱の強弱表現に着目して、ロック歌手らしさの抽出を行う。すなわち、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、音符の情報とどのような関係があるのかといったルールの構築を行う。

3.3 歌唱ルールの構築

本研究では、ピブラート、ポルタメント、音量の変化を決定する歌唱ルールの構築を行う。歌唱ルールは条件節と制御節からなり、条件節では、歌唱技法がどういった箇所で見られるかを決定し、制御節では、その歌唱技法のパラメタ値を決定する。この歌唱ルールを適用することにより、Vocaloidに表情パラメタを付与し、ロック歌手風の歌声に変換する。条件節の構築にあたって、用いる説明変数を決定する必要がある。そこで、人手による条件節ルールの抽出を行い、ルールの適合率・再現率・F値の調査を行った。その調査結果より、F値の向上に必要な説明変数の考察を行っていく。

制御節の構築にあたっては、特徴的な歌唱技法を低次のモデルを利用する。歌唱技法のモデルを用いるのは、ユーザが本システムを利用する際に、簡易に歌唱技法を操作することができるためであり、またパラメタの補間も容易となるからである。モデルパラメタを最小二乗法により決定し、いくつかの説明変数の基で、 K 個の混合ガウス分布フィッティングを行うことで、制御節の構築を行う。

3.3.1 条件節の構築

GACKT の歌唱において、条件節のルール構築をするため、人手による条件節のルール抽出を行った。以下に抽出したルールのうちを示す。

- ルール1:フレーズの終わりにはビブラートをかける
- ルール2:フレーズの始めにはポルタメントをかける
- ルール3:当該音符の音高より一つ前の音符の音高が低く、かつ後続の音符が長音符の場合ポルタメントをかける
- ルール4:ルール2とルール3のOR結合

ここで述べるフレーズとは、歌唱者のプレスを入れる位置であり、旋律のひとまとまりを表わす。

3.3.2 制御節の構築

制御節では、いくつかの説明変数の基づき、どのような表情付けを行うを求める。それぞれの歌唱技法を低次のパラメタで表れすことができるモデルを利用し、そのモデルと歌唱データの最小二乗誤差により、モデルパラメタを決定する。目標音高より低い音高から目標音高へ遷移するポルタメントモデルは以下のモデルを導入する。

$$Por(t) = \begin{cases} -\frac{A}{2}, & (\theta(t-d) < 0) \\ -\frac{A}{2}(\cos(\theta(t-d)) - 1), & (0 \leq \theta(t-d) \leq 2\pi) \\ 0, & (2\pi < \theta(t-d)) \end{cases} \quad (1)$$

ここで、 A は目標音高よりどれだけ低い音高で歌唱するか、 θ は目標音高へ遷移する速さを表わし、また d は低い音高の歌唱持続時間を表わす。

次に、(楽譜上表記上)の音高を一定としながら、音高を変動させるビブラートには、ビブラートの周期変動を考慮した精度のよい以下のモデルを利用した¹²⁾。

$$V(t) = A \sin(\theta + \exp(Mt)) \quad (2)$$

ここで A はビブラートの揺れの大きさ、 θ はビブラートの揺れの速さ、 M はビブラート終盤での揺れの速さの急激な上昇・下降を表わす。

表 1 3曲のテストデータに対する再現率・適合率・F値

	再現率	適合率	F値
ルール1	81.5	77.9	79.6
ルール2	48.7	37.5	48.7
ルール3	43.3	71.4	54.0
ルール4	92.0	49.7	64.5

本研究では、それぞれの歌唱技法には複数のパターンがあると仮定する。(例えば、歌唱技法の一つであるビブラートを取っても、振幅が遅いビブラートや、周期が遅いビブラート等)そこで、説明変数上に k 個のクラスがあると考え、混合ガウス分布のパラメタフィッティングを行った。また、説明変数には当該音符と一つ前の音符の音高差、音価、音高、フレーズ情報、テンポを用いた。

3.3.3 抽出したルールの評価実験

本実験では、人手による条件節に関するルールの精度を確かめるため、いくつかの楽曲に対して適合率・再現率・F値を調査した。再現率 (Recall) は人手で準備した正解例の中の何割を提案したルールで抽出できたかを表し、適合率 (Precision) はルールで抽出したもので人手で準備した正解例に含まれるものが何割だったかを表わす。また、F値 (F-measure) は適合率と再現率の調和平均で表わされ、以下の式で計算される。

$$F = \left(\frac{2}{1/Precision + 1/Recall} \right) \quad (3)$$

実験では、「12月の Lovesong」「君に逢いたくて」「君が追いかけた夢」の3曲を用いた。

3.3.4 実験結果

評価実験で得られた再現率、適合率、F値を表1に示す。ルール1の再現率、適合率を見ると、どちらも高い値を示していた。この結果より、ビブラートの有無はある程度ルール1で表せることができるが、ビブラートをかけるのはフレーズの終わりだけではなく、フレーズ内の音価の大きい音符や長音符のある箇所にも表れていた。また、当該音符の次の音符が休符の場合はビブラートを付与しないこともあった。そのため音価の大きさ、後続音符の有無を加えたルールを用いることにより、再現率、適合率の向上が望めると考えられる。

ルール2に関してだが、再現率、適合率が共に低い値を示していた。この理由として、フレーズの始めであっても高い音を出す箇所ではポルタメントをかけることが少ないこと、また、別のルールによってフレーズの始め以外にも頻出することがあげられる。ルール3もポルタメントに関するルールだが、適合率に関しては良い値を示している。しかし、ルール3以外の箇所以外にもポルタメントが存在するため、再現率は高くない。そこで、ポルタメントに関するルール2とルール3のOR結合を取ったルール4を用いてみると、再現率は非常に高い値を示すことがわかる。しかし、適合率はあまり高くないため、不必要なデータも混じっている。ルール4のF値を向上するには、ルール2に関する高い音を出す箇所を除くルールを構築することが考えられる。

4. ロックボーカルレゾネータの構成

この章では、3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機構とルールを適用した歌唱データに対して、歌唱技法の挿入位置、歌唱技法や音量といったパラメタ制御の変更を簡易に行える機構を合わせた「ロックボーカルレゾネータ」について述べる。

4.1 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、Vocaloid歌唱を簡易にロック歌手風の歌い方に変換し、ユーザのVocaloid歌唱の作り込み支援を目的とする。本システムでは、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、フレーズ箇所を推定する¹³⁾。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ジ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条件節のルールより、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し、再現率を基に入れるかどうかの決定を行うルールの制御では、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。付与された表情に納得がいけない場合には、歌唱技法の位置を変更、歌唱技法の表現を誇張・抑制を簡易に調整できるように、適用されたパラメタを調整することができるGUIを構築した。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

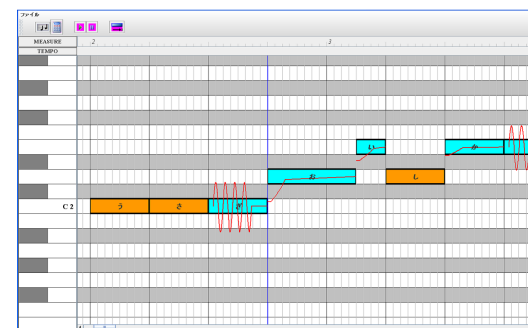


図3 ロックボーカルレゾネータ システム画面

る歌唱作りこみ作業の効率化が期待される。

4.2 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を図3に示す。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グ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。

5. 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い歌い方になっているかの確認・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。提案システムによりピブラート、ポルタメントが付与された結果を図4に示す。第一著者の主観ではあるが、ポルタメントに関しては、知覚的にロック歌手風の歌い方に近づくことが確認できた。一方、ピ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同一で、自然な歌声とはならない印象を受けた。この表情パラメタをGUIによりピ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、ピ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

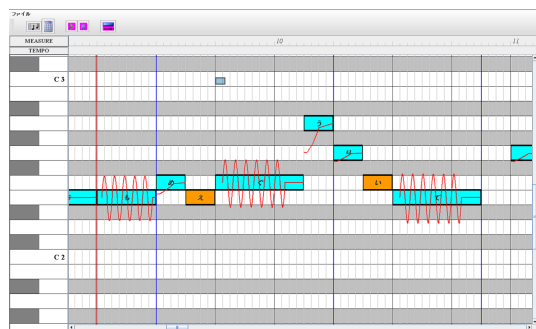


図 4 ロック歌手らしい表情パラメタの適用後

ク歌手らしい歌い方を作れると考えられる。また、ビブラートモデルの深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。

6. おわりに

本研究では、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 (GACKT) の歌い方を、Vocaloid パラメタとして付与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。また、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。ビ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、ビブラートの多様性を抑えていたため、動的に変化するビブラートの深さをモデル化する必要がある。今後は、条件節のルールにおいては、F 値の向上を目指すため、問題点を改善したルールの構築と与えられた説明変数の基での自動的なルール構築を行っていく。また、制御節のルールではビブラートの深さを加えたモデルの構築を行っていく。

参考文献

- 1) <http://www.crypton.co.jp/mp/pages/prod/vocaloid/>
- 2) 剣持秀紀 他: 歌声合成システム VOCALOID-現状と課題, 情報処理学会研究報告 2008-MUS-74(9), 2008
- 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) 大浦 圭一郎 他: 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 Vol2010-MUS-86 No.1 2010/7/28
- 6) 中里南子: J・POP 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」の分析を通じて-
- 7) <http://gackt.com/jp-h/profile/index.html>
- 8) <http://www.ssw.co.jp/products/vocal/gackpoid/gaiyo/index.html>
- 9) 「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
- 10) <http://www.celemony.com/cms/index.php?id=home>
- 11) 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.
- 12) 齋藤毅, 辻直也, 鶴木祐史, 赤木正人, “ 歌声らしさの知覚モデルに基づいた歌声特有の音響特徴量の分析, ” 日本音響学会誌, vol.64, no.5, 2008.
- 13) 辰巳直也 他: 歌唱特徴付与システム「ロックボーカルレゾネータ」, 情報処理学会研究報告 2010- MUS-87(7), 2010