

リアルタイムに初音ミクを歌わせるタイプソングシステム 「HANAUTAU」とそのアジャイル型開発事例報告

竹本 拓真^{1,a)} 馬場 隆¹ 片寄 晴弘¹

概要: 初音ミクなどのバーチャル・シンガーたちが動画コミュニティサービス「ニコニコ動画」を中心に脚光を浴びており、初音ミク主演のライブコンサートが催されている。リアルタイム歌唱生成の重要性が高まり、VOCALOID キーボードを始め様々なシステムが提案されているが、歌唱表現に関する様々なパラメータの付加ができない、操作が煩雑であるなどの問題点がある。本稿では、鼻歌によるピッチ操作とPC キーボードによる歌詞入力によって、人間らしい歌唱表現の付いた歌唱を手軽に実時間生成できるシステム「HANAUTAU」を提案し、さらにニコニコ動画を活用したアジャイル型開発の概要と過程について紹介する。

a Real-Time Singing Generator Using Typing and Humming “Hanautau” Based on an Agile Software Development

TAKEMOTO TAKUMA^{1,a)} BABA TAKASHI¹ KATAYOSE HARUHIRO¹

Abstract: Virtual singer who, such as Hatsune Miku is in the spotlight at video community service “Nico Nico Douga”. Then, live concert of Hatsune Miku organizers have been held. The importance of real-time increases. So, real-time singing generation system has also been proposed multiple led by VOCALOID keyboard. But, they has some problems that we can not add various parameters on singing expression and operation of such is complicated. pitch manipulation by humming and the lyrics input by the PC keyboard. We propose a system that can generate a real-time singing with a human singing expression. And, we will introduce agile development using the Nico Nico Douga.

1. はじめに

近年、初音ミクなどの実際には存在しないバーチャル・シンガーと呼ばれるキャラクタたちが、動画コミュニティサービス「ニコニコ動画」[5]を中心に脚光を浴びている。初音ミクの歌声は、YAMAHAが開発した歌声合成ツール「VOCALOID」[6]により生成されている。PCにソフトウェアをインストールすることで誰にでも歌唱作成が可能になった。また、ニコニコ動画の特徴である創作が創作を呼び起こすN次創作?にマッチしたことで、連鎖的な創作活動が起こった。その結果、VOCALOIDを使用した曲が数万単位で制作されている。そして、初音ミク主演の

ライブコンサート [8] が催されるまでになり、リアルタイム歌唱生成が求められる場面が多くなってきている。その要求に応えるものとして、右手でピッチ・リズム指定(鍵盤演奏)しながら左手でリアルタイムに歌詞を入力するVOCALOID キーボード [1]、事前に歌詞を入力しピッチ指定をリアルタイムに行う eVY1 [9] が YAMAHA で開発されている他、VOCALOID エディタに同封されている Real-time VOCALOID VSTi を利用して、同様のシステムが開発されている [10][11]。しかし、これらのシステムには、生成された音声に対し歌唱表現が付加されていない・操作の習熟に時間がかかるなどの問題がある。本研究は、これらの問題を解決するものとして、リズム・歌詞入力をPC キーボードで指定し、鼻歌によってピッチに関する歌唱表現を付加するシステム「HANAUTAU」シリーズの開発を進めてきた [3]。開発にあたって、ニコニコ動画に本コ

¹ 関西学院大学
Kwansei Gakuin University

^{a)} takumarakan@kwansei.ac.jp

ンテンツを発表し、そこでのコメントをユーザーズ集積手段として利用し、アジャイル型のシステム開発を実施してきた。本稿では、「HANAUTAU」の概要とアジャイル型のシステム改訂過程について報告する。

2. 関連研究と本研究の位置づけ

2.1 関連研究

これまでに提案されてきた VOCALOID を使用したリアルタイム歌唱生成は、大きく以下の2種類のアプローチに分類される。

2.1.1 ピッチと歌詞をリアルタイムで同時入力するアプローチ

- 「VOCALOID キーボード」(以下 VK) は、YAMAHA が開発した VOCALOID 用のリアルタイム歌唱生成システムである [1]。ユーザは、鍵盤と文字入力用キーボードが一体化した独自デバイスを使用する。右手の鍵盤操作によりピッチ・リズムを入力し、鍵盤左方にある文字キーボードを左手で操作することで歌詞を入力する(図1)。VK には調声機能がないため、人間らしい歌唱に必要とされる表現を付加することはできない。また、VOCALOID ハードウェア音源や特殊基盤を含む独自デバイスであるにもかかわらず非売品である。さらに、両手で異なる操作を行うため、その習熟に時間がかかる。



図1 VOCALOID キーボード操作風景

- 「vocaloidpad」は、VK をモチーフに個人作成されたソフトウェアである [10]。タブレット PC 用であるため、左手の歌詞入力方法がフリック入力に変更されている。音源は、対応する日本語全ての wav ファイルを VOCALOID エディタから個別に出力する必要がある。

2.1.2 ピッチをリアルタイムに入力するアプローチ

「Real-time VOCALOID2 VSTi」は、VOCALOID 付属の音源エンジンを外部ソフトウェアから制御するための VSTi である。これを活用して、VK をモチーフにしたソフトウェアが個人制作されている [11]。VK との違いは、歌詞を事前入力しておくことで、ピッチ指定を両手での鍵

盤操作による入力にした点である。実装されていないが、MIDI のピッチベンド信号によりビブラートを制御することができるので、ピッチに関する歌唱表現を付加することも可能である。また、eVY1 は、様々なデバイスでリアルタイム VOCALOID 歌唱を生成させるために Real-time VOCALOID VSTi を外部チップ化したものである [9]。(音声合成処理量を大幅に削減した eVOCALOID を搭載している。こちらも歌詞は事前入力であり、ピッチの指定を両手での鍵盤操作にて行う。

2.2 HANAUTAU の狙い

VK は、右手によるピッチ・リズム入力と左手による歌詞入力により、リアルタイム歌唱生成を実現しており、急な歌詞変更などのアドリブにも対応可能であることが特徴である。しかし、操作が複雑である上、特殊な機材が必要となる。vocaloidpad は、より一般的なタブレット PC を使用し、VK の機能を実装している。また歌詞入力がフリック入力に変更されている。これらの2つの研究手法では、歌詞を片手で入力しなければならず、スムーズな入力や高速な曲に対応しきれない。Real-time VOCALOID2 VSTi と eVY1 では歌詞を事前に入力し、ピッチ・リズムをリアルタイムに指定している。この手法では、急な歌詞変更に対応できないが、高速な曲にも対応可能である。上記の既存研究すべてにおいて、1オクターブを12半音に分けたクロマティックな離散ピッチしか出すことができず、ビブラート・しゃくり・フォールなどの連続的に変化するピッチに関する歌唱表現を付加することができない。そこで、本研究ではピッチ指定に鼻歌を使用し、鼻歌に内包されるピッチに関する歌唱表現をそのまま合成歌唱に転写する。鼻歌は他の楽器に比べて必要とされるスキルレベルは低く、より多くのユーザが直感的に表情を付与できると期待される。ピッチ指定を鼻歌に変更したことで、両手が自由になる。歌詞入力を日本語入力において広く普及している qwerty 配列の PC キーボードによる両手入力とすることで、操作の習熟時間の削減とタイピング速度の飛躍的向上が期待できる。

3. HANAUTAU

3.1 システム概要

HANAUTAU のシステム概要(図2)を示す。HANAUTAU は Max/MSP により実装されている。音声は、VOCALOID 初音ミクあるいはフリーの音声合成ソフトウェア UTAU [12] 「重音テト」 [13] の2種類用意する。音声データには、日本語音韻 113 種について、{あ/い/う/え/お/ん} の音長は 500[ms] で、それ以外の音韻は音長 250[ms] でサンプリングしたものを使用する。ユーザは鼻歌を歌いながら、PC キーボードで歌詞を入力する。マイクから入力された鼻歌をリアルタイムに解析してピッチ F0 を推定

し、指定音源のピッチをピッチ F0 に変更し、出力する。母音入力のタイミングでメロディのリズムを制御する。

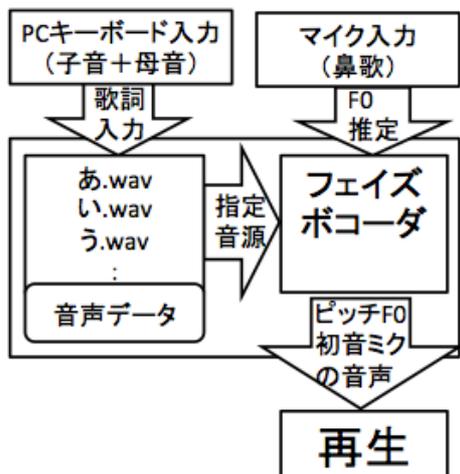


図 2 システム概要

3.2 合成音声の鼻歌ピッチによる駆動

合成音声はフェイズボコーダ [4] を用いて生成される。フェイズボコーダにおけるモジュレータ部には音声データから得られたスペクトル概形を、キャリア部には鼻歌の推定 f_0 を、それぞれ使用する。これにより初音ミクの音色で任意のピッチを生成できるだけでなく、鼻歌に含まれるしゃくり・フォル・ビブラートなどのピッチに関する音楽表現を初音ミクの音声に付加することができる。

HANAUTAU は、鼻歌に合わせてクロマティックな 12 音以外の連続的なピッチで VOCALOID を歌わせられるため、従来のエディタでは表現が難しかった話口調の生成も可能となっている。VOCALOID 音声による会話の他、ラップなど話口調の歌唱の生成も可能である。

3.3 PC キーボードによる歌詞入力及び旋律各音のオンセットオフセット指定

歌詞はローマ字入力とし、qwerty 配列の PC キーボードを用いて入力する。各音符のアタック (発音)、サスティン (保持)、リリース (消音) はそれぞれ母音キー (“a”, “i”, “u”, “e”, “o”) または “n” キーを押す/押し続ける/離すタイミングとし、子音は母音キーを押す前に押すようにする。これにより特に母音キーを押すタイミングでメロディのリズムを制御する。

歌唱の各発音単位によって発音時間は異なるため、音声データは音韻別に短くサンプリングしたものを扱い、アタックからリリースまでの間ループ再生させる。“n” 以外の子音に母音を付けた 107 音韻は音長 250[ms] とし、アタック時に 1 回だけ再生させる。“n” 及び母音の 6 音韻は音長 500[ms] とし、歌詞が母音のみの場合は、アタック

からリリースまでダブルバッファによるクロスフェード再生処理を施した上でループ再生とし、歌詞が子音+母音の場合は、該当する子音音韻を 1 回再生した後同様の処理を施す。

4. ニコニコ動画を活用したアジャイル型開発

ニコニコ動画は 3000 万人以上が登録している動画コミュニティサービスである。アカウントを作れば誰でも動画を投稿することができ、開発したコンテンツの紹介動画を多くのユーザに迅速に公開することができる。また、ニコニコ動画の特徴である時刻同期コメント [2] が得られ、その瞬間ごとの感想や提案をユーザから得ることができる。ニコニコ動画は、ソフトウェア工学において迅速かつ適応的に開発を行う開発手法であるアジャイル型開発との親和性が高く、ニコニコ動画を活用したアジャイル型開発は、効率の良いコンテンツ開発を可能にする。そこで本研究では、HANAUTAU の動画を投稿し、得られたユーザ意見をシステムに反映させて再度投稿を行うアジャイル型開発手法を実施した。

なお、VOCALOID および初音ミクについては規約により一般配布が困難であるため、より自由に使用可能な UTAU の重音テトの音源を音声データとして使用した一般配布版アプリケーションを別途作成し、(音声データ及びキャラクタ画像以外は MIKUTYPE と同様)。初音ミク版を「MIKUTYPE」重音テト版を「TETOTYPE」と称し、併せて「HANAUTAU」シリーズとする。



図 3 ニコニコ動画での MIKUTYPE 紹介動画

4.1 動画共有サイトへの投稿

本コンテンツの有用性の調査とシステムの改訂に役立てるため、2013 年 9 月 29 日にニコニコ動画へ MIKUTYPE の紹介動画を投稿した (図 3)。11 月 25 日現在で再生数 186、308 コメント数 5、866 毎リスト数 7、379 となっており、デイリーカテゴリ総合ランキング最高 1 位を獲得し、投稿から 29 日連続 10 位以上をキープするほどの反響があっ

た。また、ニコニコ動画経由での twitter のつぶやきが 2304 回あった。多くの紹介サイト [14][15] や週刊アスキー [16] などに記事が掲載された。一般配布した TETOTYPE の紹介動画も投稿し、11 月 21 日現在 37, 586 再生、ダウンロード件数 3384 件である。さらに、TETOTYPE を使用した二次創作の動画が 3 件確認されている。

4.2 ユーザコメントの収集とシステム改訂

ユーザコメントを整理し、機能に関する主な要望には、**1) 歌詞の事前入力、2) 録音機能、3) MIDI データでの出力、4) 音声認識**が挙げられており、第一回目の改訂では、特にコメント数が多かった 1) と 2) の機能について実装した。2) については、TETOTYPE を配布する際の標準機能として実装した。1) の機能改訂については以下で詳しく述べる。

HANAUTAU では、リアルタイム歌詞入力を重視したため歌詞の事前入力の機能は、初期の機能では採用しなかった。しかし、ユーザからの要望が多かったため、歌詞データを事前入力とし、ボタンを押すことで歌詞進行する機能を付加した。歌詞データはテキスト形式とし、ユーザが自由に作成できる。歌詞はローマ字表記、日本語 1 文字ごとに改行を設ける。ボタンが一だと、前音と次音の間に空白ができるため、とぎれとぎれの歌唱になってしまう。そこでボタンを二つ(方向キーの下・右)用意し、交互に連打することでスムーズかつ高速な歌詞進行を実現した。歌詞の事前入力機能により片方の手が自由になるため、新たな機能の追加を実施し、今回は、歌唱表現に大きく影響を与える音量の操作と、リバーブエフェクトの制御を片手でできる機能を試験的に追加する。操作の煩雑化を回避するため本機能はオプション機能とした。ユーザインターフェースには PC のタッチパッドを使用し、Y 座標を音量・X 座標をリバーブにそれぞれマッピングした。11 月末にユーザからの意見を基に改訂した TETOTYPE を一般公開し、それに関する動画をニコニコ動画に投稿した。得られたコメントを解析し、さらなるフィードバックを行う予定である。

5. まとめと今後の展開

初音ミク主演のライブコンサートが催されるなど、リアルタイム歌唱の重要性が高まってきており、VK のようなリアルタイム歌唱生成システムが提案されていた。しかし、生成された歌唱には歌唱表現が付加されていない・操作が煩雑であるなどの問題があった。そこで本研究では、鼻歌によるピッチ指定を採用することにより、滑らかなピッチ変化やビブラートなどの歌唱表現をリアルタイムに付加することで、より人間らしいリアルタイム歌唱生成システムを実現した。歌詞入力について、PC キーボードを両手で操作することにより入力的高速化を図った。

ニコニコ動画を活用したアジャイル型開発を行うこと

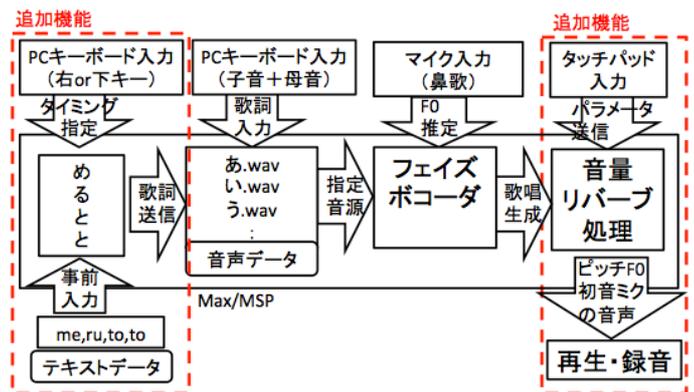


図 4 ユーザコメントに基づいての追加された機能

で、使用者側のニーズを迅速にとらえることができる。また、ニコニコ動画は、3000 万人以上のユーザに加え SMS との連携が充実し、全世界へのコンテンツ情報の拡散に効果的である。本コンテンツにおいても英語・スペイン語・韓国語によるコメントが寄せられている。

ユーザから得られたコメントを基に歌詞の事前入力・録音・音量制御・エフェクト制御機能の追加を実施した。今後の展開として、フェイズボコーダによる音質劣化を最小限に抑えるためのシステム改訂や、音色の追加・切り替えと、さらなる歌唱表現の付加を実施する。

参考文献

- [1] 加々見 翔太:リアルタイム日本語歌唱鍵盤楽器“VOCALOID キーボード”の開発, インタラクシオン 2012(2012)
- [2] 特集「CGM の現在と未来: 初音ミク, ニコニコ動画, ピアプロの切り開いた世界」(解説 5 件), 情報処理(情報処理学会誌), Vol. 53, No5, pp.464-494 (May 2012).
- [3] 竹本 拓真, 馬場 隆, 片寄 晴弘: MIKUTYPE: リアルタイムで初音ミクに歌わせるタイプソングシステム, エンタテインメントコンピューティングシンポジウム 2013 論文集, pp.330-331 (2013)
- [4] Flanagan, J. L, Golden, R. M: Phase vocoder, Bell System Technical Journal 45, pp.1493-1509 (1966)
- [5] <http://www.nicovideo.jp/>
- [6] <http://www.VOCALOID.com/>.
- [7] <http://www.crypton.co.jp/mp/pages/prod/vocaloid/cv01.jsp>
- [8] <http://mikufs.jp/>
- [9] <http://jp.yamaha.com/newsrelease/2013/pdf/1310230101.pdf>
- [10] <http://www.nicovideo.jp/watch/sm17357529>
- [11] <http://www.nicovideo.jp/watch/sm17499508>
- [12] <http://www10.atwiki.jp/utau2008/>
- [13] <http://kasaneteto.jp/>
- [14] <http://nlab.itmedia.co.jp/>
- [15] <http://getnews.jp/>
- [16] <http://weekly.ascii.jp/>