

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

中野倫靖[†] 後藤真孝[†]

VocaListener: A Meta Singing Synthesis Interface Based on Interactive Singing Synthesis Using Sung Examples

TOMOYASU NAKANO[†] and MASATAKA GOTO[†]

1. はじめに

歌声合成では、ユーザが歌唱の表現 (声質、歌い方) を思い通りに創出できることが望ましい。しかし、既存の歌声合成システム¹⁾ では、自分なりの表現で合成させるために細かなパラメータ調整が必要で、しかも、パラメータ調整の結果を他の音源 (合成の声質) にそのまま適用できない問題があった。そこで我々は、ユーザが歌声合成条件 (歌声合成システムやその音源データ) の違いに依存せず、「表現」の創出に全力を注げるよう支援することを目指して研究を進めている。

本稿では、歌声合成のパラメータをユーザが歌うだけで調整でき、かつ、歌声合成条件を自由自在に切り替えながら歌声合成できるメタ歌声合成インタフェース VocaListener を提案する。従来、ユーザの歌唱音声 (目標歌唱) を分析し、その結果を歌声合成パラメータとする研究はあった²⁾ が、歌声合成条件の違いには対処できなかった。VocaListener は、目標歌唱と合成歌唱の分析結果が近づくように、パラメータを反復推定することで、歌声合成条件の違いを吸収できる。

さらに、目標歌唱を真似る歌声合成では、

- (1) 目標歌唱の分析結果に誤りが含まれると、合成音の聴取時に違和感を感じてしまう
- (2) ユーザ自身が歌唱できない表現では合成できない (歌が不得手、音高が声域より高い場合など) の二つの問題点がある。そこで、違和感を感じた箇所を指摘するだけで訂正できる『ダメ出し』インタラクシオンと、ピブラートを強く・弱くという直観的操作で、自分好みの表現へ変更できる歌唱力補正機能を新たに提案する。

[†] 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST)

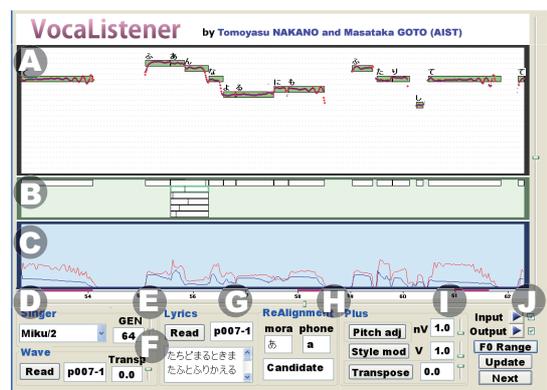


図 1 VocaListener の表示画面。『ダメ出し』インタラクシオンにより、③に音節境界候補が表示された例。

2. インタフェース概要

図 1 に VocaListener の画面例を示す。図の上部は、以下の 3 つのウィンドウで構成される。

- 楽譜ウィンドウ (図 1: ①)
歌詞と共に、自動推定された音符列 (緑の長方形) と F_0 軌跡 (赤: 目標歌唱, 青: 合成歌唱) が表示される。縦軸は音の高さで、図中右端のスライダーでその表示範囲を変更できる。
- アラインメントウィンドウ (図 1: ②)
自動推定された歌詞アラインメント (目標歌唱と歌詞の時間的対応付け) の結果が表示される。
- パワーウィンドウ (図 1: ③)
目標音量 (赤) および合成音量 (青) が表示される。歌声合成は、図の下部の各種コントロール (ボタンなど) により、インタラクティブに行う。

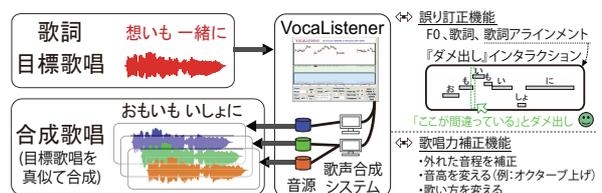


図 2 VocaListener による歌声合成

3. VocaListener による歌声合成

図 2 に、VocaListener による歌声合成の流れを示す。以降、処理を手順ごとに説明する。

3.1 ステップ 1: 歌声合成条件の選択

まず、合成に使用する歌声合成システムとその音源を選択する。現時点では、ヤマハ株式会社の開発した歌声合成技術 Vocaloid / Vocaloid 2 のうち、日本語歌詞の歌声を合成できるソフトウェア から選ぶ。

3.2 ステップ 2: 目標歌唱と歌詞の分析・誤り訂正

次に、目標歌唱とその歌詞テキスト(漢字仮名交じり文)を指定する。歌声合成のための音高(F_0)・音量推定、歌詞のテキスト処理がなされ、音高と音量がⒶⒸに、歌詞を発音系列(平仮名)へ変換した結果がⒸ下部に、それぞれ表示される。 F_0 推定結果が誤っていた場合は、ドラッグで推定する時間・音高(周波数)範囲を指定すると自動的に正しく再推定される。また、発音系列の誤りはテキスト編集で簡単に訂正できる。

3.3 ステップ 3: 歌詞アラインメント結果の誤り訂正(『ダメ出し』インタラクシオン)

歌詞アラインメント結果に含まれる誤りを訂正する。まず、まれに比較的長い無音をまたいだ大きな対応付け誤りが発生するので、そのおおよそ正しい時刻を指定すると、自動的に正しく再推定される。次に、より細かい音節境界誤りを訂正する(図 1)。誤っている箇所を指摘するだけで、可能性の高い新たな境界候補が自動生成されて表示されるので、選択するだけで簡単に訂正できる。最後に、上記で半自動的に訂正できなかった誤りを、音符位置や長さを微調整して訂正する。

3.4 ステップ 4: 歌唱力補正

補正したい時間範囲をドラッグで指定し、3 種類の補正ができる。まず、「調子はずれ補正」では、音程がずれている箇所を自動補正して合成できる。次に、「歌唱スタイルの変更」では、ビブラート区間とそれ以外の区間の音高と音量を、別々に強調・抑制して補正でき、自分好みの表現で合成できる。最後に、「音高トランスポーズ」では、合成の声域を変え、ユーザが歌えない声域での歌唱も合成できる。

4. VocaListener の実現方法

まず、単独歌唱からの F_0 推定技術、漢字仮名交じり文から発音系列を得る技術、歌詞アラインメント技術が必要である。それぞれ、入力信号中で最も優勢な

高調波構造を求める手法⁶⁾、形態素解析器 MeCab⁷⁾、歌詞同期で使われる Viterbi アラインメントを用いた。

そして、歌声合成条件に依存しない歌声合成パラメータの推定は、目標歌唱だけでなく合成歌唱も分析し、パラメータを反復推定することで実現した。

さらに、誤った音節境界を訂正するために、新しい境界候補を自動選出する技術が必要がある。本稿では、目標歌唱の MFCC 変動が大きく、目標歌唱と合成歌唱との MFCC 距離が小さい箇所を候補として選出した。また、ビブラート区間の検出には、 F_0 軌跡の 1 次差分の短時間フーリエ変換によってビブラートらしさを得る手法⁸⁾を用い、調子はずれ補正では、有声区間毎に、半音単位に大きな重みを与える関数が F_0 軌跡と最も適合する値を補正值とした。

5. おわりに

歌声合成パラメータをユーザが歌って調整でき、インタラクティブに誤り訂正や歌唱力補正が行えるメタ歌声合成インタフェース VocaListener を提案した。VocaListener を実装して運用した結果、音源を替えても常に同じ歌い方で合成できることを確認した。また、推定結果が誤っていた場合、ユーザが専門的知識を持っていなくても、誤り箇所を訂正できた。

今後は、誤りを全て『ダメ出し』のみで訂正できたり、歌唱力の自動評価⁸⁾を組み合わせて「歌を作っていたら自分も歌が上手く」なれる、新しいインタラクシオンの実現を目指す。

謝辞 本研究の一部は、科学技術振興機構 CrestMuse プロジェクトによる支援を受けた。

参考文献

- [1] 剣持 他: 歌声合成システム VOCALOID - 現状と課題, 情報研報 (MUS), Vol. 2008, No. 12, pp. 51-58 (2008).
- [2] Janer, J.: Performance-Driven Control for Sample-Based Singing Voice Synthesis, *Proc. of DAFX-06*, pp. 42-44 (2006).
- [3] クリプトン: ソフト音源 (VOCALOID ENGINE), <http://www.crypton.co.jp/mp/do/prod/vi/engine?id=10>.
- [4] クリプトン: VOCALOID2 特集, <http://www.crypton.co.jp/mp/pages/prod/vocaloid/>.
- [5] 株式会社インターネット: がくっほいど, <http://www.ssw.co.jp/products/vocal/gackpoid/>.
- [6] 後藤 他: 自然発話中の有声休止箇所のリアルタイム検出システム, 信学論, Vol. J83-D-II, No. 11, pp. 2330-2340 (2000).
- [7] 工藤: MeCab: Yet Another Part-of-Speech and Morphological Analyzer, <http://mecab.sourceforge.net/>.
- [8] 中野 他: 楽譜情報を用いない歌唱力自動評価手法, 情報学論, Vol. 48, No. 1, pp. 227-236 (2007).

ただし、合成の過程で好きな時に変更することが可能である。Vocaloid 「KAITO」, 「MEIKO」³⁾、Vocaloid 2 「初音ミク」, 「鏡音リン」, 「鏡音レン」⁴⁾、「がくっほいど」⁵⁾。

合成結果の具体例は、<http://staff.aist.go.jp/t.nakano/VocaListener/index-j.html> で聞くことができる。