メディア操作　Directability　歌唱

ユーザ歌唱を真似て歌声合成する

中野倫靖　　後藤真孝
（後藤グループ：産業技術総合研究所）

概要　本研究では，歌声合成を使用する音楽制作を支援するために，ユーザの歌唱音声から歌声合成パラメータを自動推定するシステム VocaListener を実現した．従来，ユーザの歌唱音声から，音高や音量等を推定して歌声合成パラメータとする研究はあったが，歌声合成の条件（歌声合成システムやその音源データ）の違いに対してロバストでなかった．そこで VocaListener では，合成された歌唱が入力歌唱と近くなるように，合成パラメータを反復更新することで，上記の条件の変化へ対処した．さらに，入力歌唱を真似るだけではユーザの歌唱力やその表現能力を超えることができないが，音高のずれやビブラートなどの歌唱要素を修正できる支援機能も提供することで，そのような問題にも対処した．
キーワード：歌声合成，ユーザ歌唱，合成パラメータの自動推定，歌詞アラインメント，歌唱力補正．

1. はじめに

歌声合成システムは，個人が歌唱付き楽曲を制作するのを容易にし，歌唱の表現を自在にコントロールできる重要なツールである．現在では，高品質な歌声を合成できる市販ソフトウェアが普及し始めており，インターネットを介した音楽の共同制作等，新しいコミュニケーションを生み出している．さらに，高品質な歌声合成を目指すことは，人間の歌声知覚・生成機構のさらなる解明に繋がる取り組みでもある．

歌声合成システムを普及させるためにはいくつかの課題があるが，まず自然な歌声を合成できる必要がある．また，歌声合成システムを利用する多くのユーザが，魅力的な歌声を自由自在に作れるようになるために，歌唱の表情付けのためのインタフェースを充実させる必要がある．さらに，利用できる歌声合成システムやその音源データ（声質）は，今後増えていくと考えられるが，そのような歌声合成の条件の違いに依存せずに歌声合成を行える必要がある．

従来，より自然な歌声を合成するために，歌声の表情パラメータを細かく調整できる方式 [1] があったが，ユーザによっては自分の望む歌声を作るのを困難にしていた．一方で，そのような表情付けを容易にするために，歌唱音声から音高や音長などを抽出して表情パラメータとする研究があった [2]．しかし，音高や音量などの合成結果は，歌声合成システムやその音源データに依存しているため，その条件が変わると，同じパラメータを与えても合成結果が異なってしまう問題があった．

そこで我々は，入力としてユーザが歌唱音声を与え，それを真似るために歌声合成パラメータを入力と比較しながら反復推定する VocaListener を実現した [3]．これによって歌声合成の条件に依存せず，ユーザは歌うだけでそれを基にした表情豊かで自然な歌声を合成できる．さらに，ユーザ歌唱の分析結果を編集することで，ユーザ自身が歌唱できない表現（音高が声域より高い場合など）に対しても歌声合成を行えることも可能とした．

2. VocaListener: ユーザ歌唱を真似る歌声合成パラメータ推定システム

本研究では，合成歌唱を目標歌唱（入力）へ近づけるコア技術を VocaListener-core，目標歌唱自体を編集する技術を VocaListener-plus と呼ぶ．また，それぞれで必要な要素技術を VocaListener-front-end と呼ぶ．これ以降，ユーザによって与えられた歌唱を目標歌唱，歌声合成システムによって合成された歌唱を合成歌唱と呼ぶ．

図 1 にシステム全体の流れを示す。ユーザは，歌唱音声とその歌詞を入力として与える（A）．システムは，それらの入力に対して分析を行うが，漢字かな混じり文をかな文字列に変換する際の誤りや，歌詞の割り当てでフレーズをまたがるような大きな誤りがあった場合は，ユーザが手作業で訂正する（B，C）．次に，VocaListener-plus によって，声域を変更したり，ビブラートの深さ等を調節したりできる（D）．最後に，VocaListener-core によって，入力歌唱を真似る合成パラメータを推定する（E）．この際，歌詞アラインメントの音節境界に誤りが生じていたら，ユーザはその箇所を指摘して訂正する（F）．最後に，ユーザは推定されたパラメータによって合成された歌唱を得る（G）．

2.1 VocaListener-front-end

VocaListener-front-end は，歌声分析及び歌声合成に関する要素技術群である．これらの要素技術は，状況に応じて任意の手法を利用できる．

まず，歌声分析としては「音高」，「音量」，「発音開始時刻」，「音長」及び「ビブラート区間」を自動推定する技術が必要である．音高は歌唱音声の基本周波数（F0）を既存の手法で抽出し，音量は音声波形の振幅の実効値を計算した．発音開始時刻及び音長は，音声認識で利用される Viterbi アラインメントによって自動的に推定して利用した．

また歌声合成システムとその音源データとしては，ヤマハの開発した Vocaloid2 [1] の応用商品である，クリプトン・フューチャー・メディアの初音ミク及び鏡音リン [4] を利用した．

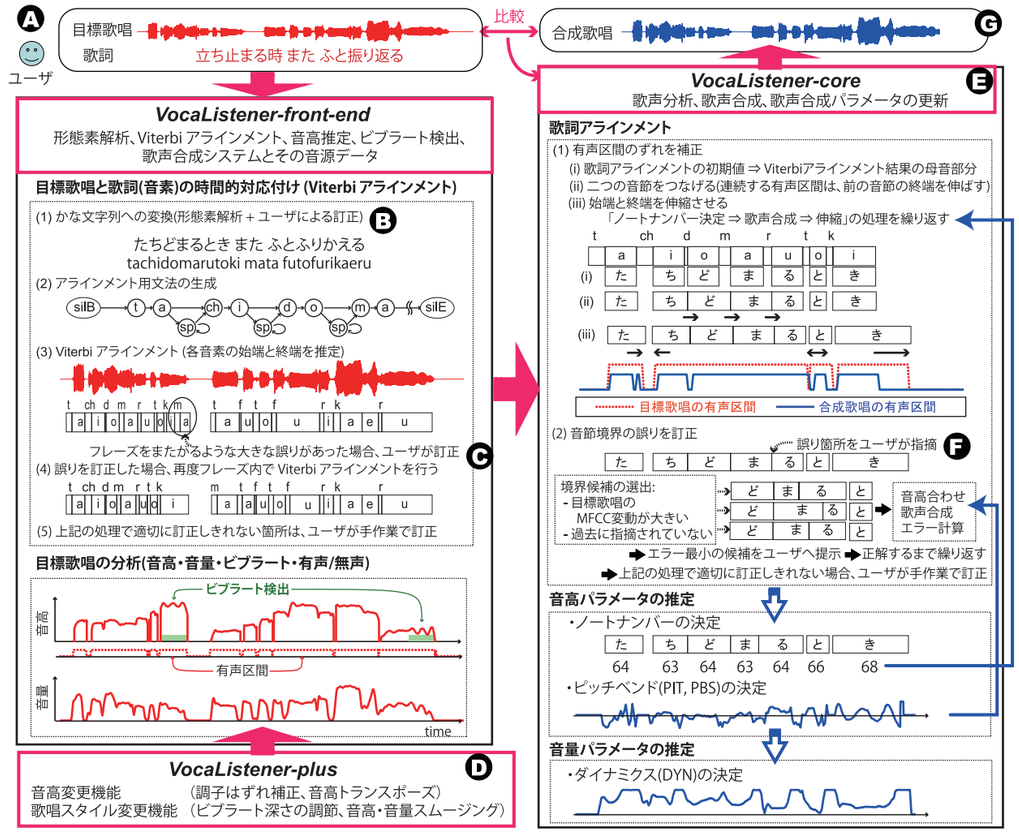

図 1　VocaListener の全体像（VocaListener-front-end, VocaListener-plus, 及び VocaListener-core）

2.2 VocaListener-core

VocaListener-core では，歌詞と目標歌唱の時間的対応付けを行う**歌詞アラインメント**と，音高と音量に関する**合成パラメータの反復推定**を行う．歌詞アラインメントでは，Viterbi アラインメント性能や歌声合成システムの特性が原因で合成結果が目標とずれるため，「有声区間を合わせる」「音節境界の誤り箇所をユーザが指摘する」ことで訂正する．続いて，音高・音量パラメータを実際に合成しながら反復更新していくことで，歌声合成の条件の違いを吸収してロバストに推定する．

2.3 VocaListener-plus

VocaListener-plus は，歌唱入力の表現を広げるために目標歌唱自体を編集する機能であり，現在，以下の二種類がある．これらは，状況に応じて利用すればよく，使わないという選択も可能である．

音高変更機能　音高の変化が半音単位となるように音高をずらす「調子はずれ（off-pitch）補正」と，部分的もしくは歌唱全体の音高を上下させる「音高トランスポーズ」がある．

歌唱スタイル変更機能　音高と音量を変更する機能であり，ビブラート区間とそれ以外を分けて，個別に音高・音量を強調したり抑制したりできる．

3. おわりに

本稿では，ユーザ歌唱を真似る機能と歌唱力補正機能を持つ歌声合成パラメータの自動推定システム VocaListener を紹介した．VocaListener の有効性は実際の歌唱（4 人分）を用いて評価を行い，歌声合成パラメータの反復推定と，音節境界誤りの指摘について，その有効性を確認した [3]．

参 考 文 献

[1] 剣持秀紀，大下隼人：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, pp.51–58, Vol.2008, No.12, pp.51–58, (2008).
[2] J. Janer, J. Bonada, M. Blaauw: Performance-Driven Control for Sample-Based Singing Voice Synthesis, In Proc. of the 9th Int. Conference on Digital Audio Effects (DAFx-06), pp.42–44 (2006).
[3] 中野倫靖，後藤真孝：VocaListener: ユーザ歌唱を真似る歌声合成パラメータを自動推定するシステムの提案，情報処理学会研究報告 2008-MUS-75, Vol.2008, No.50, pp.49–56, (2008).
[4] クリプトン，VOCALOID2 特集，<http://www.crypton.co.jp/mp/pages/prod/vocaloid/>