

SingBySpeaking: ユーザの話声を歌声に変換する歌声合成インタフェース

齋藤 毅[†] 後藤 真孝[†]
鵜木 祐史^{††} 赤木 正人^{††}

SingBySpeaking: Singing Voice Synthesis Interface by Converting User's Speaking Voices into Singing Voices

TAKESHI SAITOU,[†] MASATAKA GOTO,[†] MASASHI UNOKI^{††}
and MASATO AKAGI^{††}

1. はじめに

歌声の自動合成技術は、計算機による新たな音楽の楽しみ方を創造するだけでなく、人間の音声コミュニケーションを理解する上でも重要である。従来の歌声合成では、歌詞テキストから歌声を合成するアプローチが主流^{1),2)}である。これらは、波形接続合成やHMM合成といったコーパスベースの手法に基づいたものが多く、誰でも手軽に歌声を合成することが可能である。一方で、予め用意されたコーパスの声質の歌声しか合成できず、ユーザの声質の歌声を合成するにはユーザ自身のコーパスが必要となるため非常に困難であった。

それに対して、本研究では、話声から歌声を合成するという新しいアプローチに基づいた歌声合成インタフェース SingBySpeaking を提案する。ユーザは、合成したい歌の歌詞を朗読し、更にはその朗読音声に対して歌声特有の音響特徴を選択的に付与することで、自分の声質の歌声を得ることができる。これにより、「歌詞を朗読さえすれば自分自身の歌声を合成できる」という、従来は実現されていなかった新たな歌声合成の楽しみ方を提供できる。更には、話声と歌声の音響的違いを直感的に学ぶことができ、人間の歌声知覚・生成機構の理解の一助となることが期待できる。

2. SingBySpeaking の機能

SingBySpeaking の概要を図1に示す。図の中央にインタフェース画面を、その周りに合成処理の概要を示

す。ユーザは、合成する歌の選択、歌詞の朗読を行った後、朗読音声の時間長、音高、音色の操作を順に行うことで、ユーザ自身の歌声を合成できる。以下、各操作について述べる。

1: 歌の選択

リストボックス (図中の①) から、事前に譜面情報 (MIDI データ) と歌詞情報 (テキスト) データが登録された曲を選択する。選択すると、ウィンドウ (図中の②) に朗読する歌詞が表示される。

2: 歌詞の朗読

録音ボタン (図中の③) を押し、表示された歌詞をマイクに向かって朗読する。朗読音声の音色は歌声合成音でも保持されるため、同じ曲でも音色を変えて朗読することで、異なる表情の歌声を合成できる。その後、分析ボタン (図中の④) によって、朗読音声のアラインメント (音声信号と歌詞の時間対応付け)、及び音高 (F0) と音色 (スペクトル包絡の時間系列) の音響パラメータを抽出する。朗読音声の波形と抽出した各種音響パラメータは、ウィンドウ (図中の⑤) に表示される。朗読音声の再生は、ボタン (図中の⑥) で行う。

3: 時間長 (Duration) 操作

時間長操作 (図中の⑦) では、朗読音声中の各音韻を、割り当てられた音符の種類と曲のテンポに従って時間伸長する (図中の時間長操作の概要を参照)。これにより、朗読音声中の各音韻の長さのみを歌声に変換する。

4: 音高 (F0) 操作

音高操作 (図中の⑧) では、選択曲のメロディ変化 (譜面に対応する階段状の変化) に対して歌声特有の変動を付与し、歌声の音高変化を生成する (図中の音高操作の概要を参照)。この変動を F0 動的変動と呼

[†] 産業技術総合研究所

National Institute of Advanced Industrial Science and Technology (AIST)

^{††} 北陸先端科学技術大学院大学 情報科学研究科

School of Information Science, Japan Advanced Institute of Science and Technology

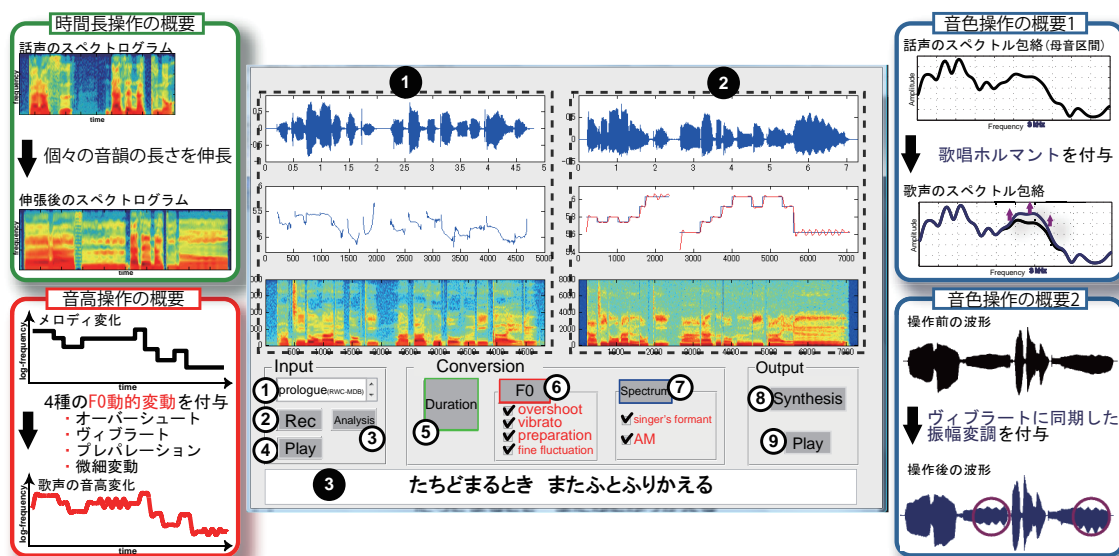


図 1 SingBySpeaking の概要

び、オーバーシュート (overshoot)、ヴィブラート (vibrato)、プレパレーション (preparation)、微細変動 (fine fluctuation) の 4 種から成る。ここでは、付与したい変動を選択し、様々な音高変化を生成できる。

5:音色 (Spectrum) 操作

音色操作 (図中の⑦) では、朗読音声のスペクトル構造に対して歌声特有の特徴を付与し、話声から歌声らしい音色に変換する (図中の音色操作の概要を参照)。付与する特徴は、スペクトルの 3 kHz 付近に存在する強いピーク成分である歌唱ホルマント (singer's formant) と、音高操作で付与したヴィブラートに同期した音声の振幅変調成分 (AM) である。ここでは、これら 2 種の特徴を選択し、音色を操作できる。

上記 3~5 の操作による各種音響パラメータの変化は、ウィンドウ (図中の②) に表示され、更に変形した音声はボタン (図中の⑧と⑨) で随時合成・再生できる。これにより、歌声特有の各種特徴が歌声変換に及ぼす影響を視覚・聴覚を通じて確認でき、最終的には「自然で人間らしい歌声の音響的条件」を理解できる。

3. 実装方法

SingBySpeaking は、音声分析合成システム STRAIGHT³⁾ の分析・合成処理過程に、時間長、音高、音色に関する音響パラメータに対して歌声特有の音響特徴を制御可能なモデルを組み込むことで実装されている。しかし、これまでに歌声合成に必要な音響特徴は十分に調査されておらず、それら特徴を自由に制御できるモデルも提案されていなかった。本研究では、筆

者らによって明らかにした歌声知覚に重要な特徴^{4),5)} に着目し、それらの特徴を自由に制御可能なモデルを構築^{5),6)} することで、従来の歌声合成技術とは全く異なる処理体系を実現した。

4. おわりに

SingBySpeaking では、ユーザによる歌詞の朗読音声に対して、ユーザ自身が様々な歌声特有の音響特徴を付与していくことで歌声に変換することを可能にした。このシステムにより、自分自身の歌声を合成でき、更には自分の話声が歌声に変換される過程を理解できる新しい歌声合成の枠組みを実現した。今後は、より多様な歌声を表現可能な歌声合成インタフェースへと発展させる予定である。

謝辞 本研究の一部は、科学技術振興機構 Crest-Muse 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，情処研報，2007-MUS-072，pp. 25-28 (2007).
- 2) 酒向他：隠れマルコフモデルに基づいた歌声合成システム，情処学論，Vol.45，No.3，pp. 719-727 (2004).
- 3) 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 STRAIGHT の背景と展開，音響誌，Vol.63，No.8，pp. 442-449 (2007).
- 4) 齋藤他：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，音響誌，Vol.64，No.5，pp. 267-277 (2008).
- 5) T. Saitou, *et al*: Development of an F0 control Model Based on F0 Dynamic Characteristics for Singing-Voice Synthesis, Speech Commun., Vol.46, pp. 405-417 (2005).
- 6) T. Saitou, *et al*: Analysis of Acoustic Features Affecting "singing-ness" and Its Application to Singing-Voice Synthesis from Speaking-Voice, Proc. Interspeech, pp. 1925-1928 (2004).

<http://staff.aist.go.jp/saitou-t/singbyspeaking.html> で全特徴を付与した歌声合成音のサンプルを聴くことができる。