

メディア操作

Directability

歌唱

体験デモ

## 歌声を混ぜるインタフェース

森勢将雅 河原英紀<sup>†</sup> 片寄晴弘

(片寄グループ：関西学院大学理工学部片寄研究室)

<sup>†</sup>和歌山大学システム工学部

**概要** 音声分析変換合成方式STRAIGHTを用いた歌唱モーフィングを、実時間で動作するインタフェースとして実装した。歌唱のモーフィングでは、歌声を「声質」「歌い回し」へと分離し、2歌手の各特徴を混ぜることができる。感情のモーフィングでは、1名の歌手が「喜」「怒」「哀」について歌った歌唱をモーフィングすることにより、3感情を自由な比率で混合できる。これらのモーフィング率を再生中でもリアルタイムで変更することができるインタフェースを紹介する。シンポジウム会場では、素人、プロ歌手、合成システムで作られた合成歌唱を混ぜることで、どのような歌声が生み出されるか体験して頂きたい。

**キーワード**：STRAIGHT、モーフィング、歌声変換インタフェース。

### 1. はじめに

高品質音声分析変換合成方式 STRAIGHT[1]、および歌唱モーフィング [2] の提案により、高い品質を有したまま歌唱の特徴を操作することが可能となった。また、いくつもの感情で歌った歌声を混ぜることにより、感情の細かい操作が可能となりつつある。歌唱モーフィングに基づいて合成された歌声を使った主観評価では、歌唱と区別する手がかりは、歌い回しよりも声質であることを明らかにした [3]。このように、STRAIGHTによるモーフィング技術は、高品質領域における歌唱の特徴分析や、プロ歌唱の声質・歌い回しを自分の歌声に転写するような、ユニークなツールが提供できる可能性を秘めている。

ここでは、STRAIGHT を用いた歌唱モーフィングを実時間で行うための仕組み、実時間でモーフィング率を操作するインタフェースを紹介する。また、インタラクティブセッションで展示するリアルタイム歌唱モーフィングインタフェースで実装した機能について説明する。

## 2. STRAIGHT と歌唱モーフィング

### 2.1 STRAIGHT

STRAIGHT とは、音声を基本周波数・スペクトル包絡・非周期性指標に分離し、再合成する方式である。音声を音響パラメータに分離し再合成する方式は Vocoder と呼ばれ、品質は低いものとされていた。STRAIGHT は Vocoder 方式でありながら、元音に匹敵する品質の音声で合成できる特長を有する。基本周波数は声の高さ、スペクトル包絡は音色に関する情報であり、これらを高い品質を保ったまま操作できることが STRAIGHT の大きな長所となる。

### 2.2 歌唱モーフィング

音声のモーフィングは、基準となる音声、目標となる音声を STRAIGHT により分析し、各要素につ

いて中間的な音響パラメータを求めることで実現された。歌唱モーフィングでは、基本周波数と発話タイミングに関する情報を「歌い回し」、スペクトル包絡、非周期性指標に関する情報を「声質」とし、歌い回し・声質それぞれについて独立したモーフィング率を与えることができる。

この歌唱モーフィングをインタフェースとして実装したものが、歌唱モーフィングインタフェース (図 1) である。このインタフェースを用いることにより、ある歌手の歌声の特徴を、別の歌手の歌声へと転写できることを示した。

### 2.3 歌唱デザイン支援を目指した歌唱モーフィングインタフェース

歌唱モーフィングインタフェースは、マウスでクリックした座標のモーフィング率で合成した歌唱が再生される静的なものである。モーフィング技術を用いた歌唱加工ツールを実現するためには、モーフィング率を時刻と共に変えることができる時系列操作機能を有することが望ましい。また、再生中にリアルタイムでモーフィング率を変化させることができる機能を実装することで、使用者は即座にフィードバックを得ることができる。この機能により生産性の向上が期待できる。

## 3. リアルタイム歌唱モーフィング

### 3.1 実装

C++ を用いて歌唱モーフィング、感情モーフィングを実装した。GUI 部分は Direct3D により Windows 上で実装した。以下では実装した機能について説明する。図 2 は歌唱モーフィングインタフェース、図 3 は感情モーフィングインタフェースの実行画面である。

### 3.2 モーフィング歌唱の実時間合成

現状の STRAIGHT は、モーフィングをリアルタイムで行うことができない。そのため、事前に様々なモーフィング率の歌唱を合成しコーパスとして

蓄える。合成は、システムが要求するモーフィング率の波形を Hanning 窓により切り出し、振幅変調が生じないように接続することで実現した。

### 3.3 再生中のモーフィング率操作

歌唱・感情どちらのインタフェースも、再生中のマウス操作でリアルタイムに音色の変化をフィードバックする機能が備わっている。図 2 の左図、あるいは図 3 の右に示す三角形中の任意の座標にマウスカーソルを移動することで、その座標に基づくモーフィング率の歌唱が再生される。

### 3.4 トレース再生

歌唱モーフィングでは、モーフィング率の時系列を編集する機能を備えている。図 2 の右上図が声質に対応し、右下図が歌い回しに対応する。横軸は時系列となる。使用者はマウスを用いてフリーハンドで時系列を描くことができる。トレース再生機能は、こうして編集したモーフィング率の時系列に基づいた再生方式である。再生中の動的操作とは異なり、同一条件で何度でも再生でき、モーフィング率も微調整可能な特徴を有する。

感情モーフィングでは、歌唱のように独立した操作軸を持たないので、各要素の時系列を操作する機能は持たない。代わりにマウス操作でリアルタイムに変化させたモーフィング率の時系列を再現した再生が可能な「リピート」ボタンを持つ。リピートボタンを押すことで、1つ前の再生で行ったモーフィング率の軌跡で再生することが可能となる。

### 3.5 伴奏のオン・オフ

どちらのインタフェースも、再生中にカラオケ音源の有無を切り替えることができる。図 2 では「VoiceOnly」図 3 では「声のみ」のボタンを押すことで、再生中・停止中どちらでも切り替えることができる。この機能を用いることで、同一条件でモーフィングした場合の品質が、カラオケ版の有無によりどのように変わるか容易に比較できる。

## 4. おわりに

STRAIGHT のモーフィングを擬似的に実時間で行うインタフェースを紹介した。今後は、STRAIGHT 分析・合成・モーフィング全てを実時間で動作できるアプリケーションの構築を行う。また、歌唱モーフィングだけでなく、ビブラートやクレッシェンドといった一般的な演奏表現も転写できる、歌唱デザイン支援ツールの開発を行う。

## 参考文献

- [1] H. Kawahara, M. Morise, T. Takahashi, R. Nisimura, T. Irino, H. Banno.: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, f0, and aperiodicity estimation, In ICASSP2008, pp.3933-3936(2008).
- [2] 河原英紀, 生駒太一, 森勢将雅, 高橋徹, 豊田健一, 片寄晴弘: モーフィングに基づく歌唱デザインイ

ンタフェースの提案と初期的検討, 情報処理学会論文誌, vol.48, no.12, pp.3637-3648, (2007).

- [3] 森勢将雅, 河原英紀, 片寄晴弘: STRAIGHT によるリアルタイム歌唱モーフィングシステムの実装, 情報処理学会研究報告, vol.2008, no.50, pp.117-122, (2003)

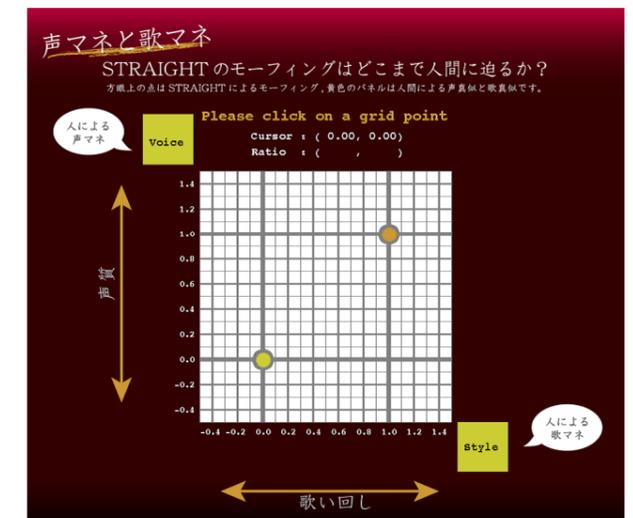


図 1 歌唱モーフィングインタフェース

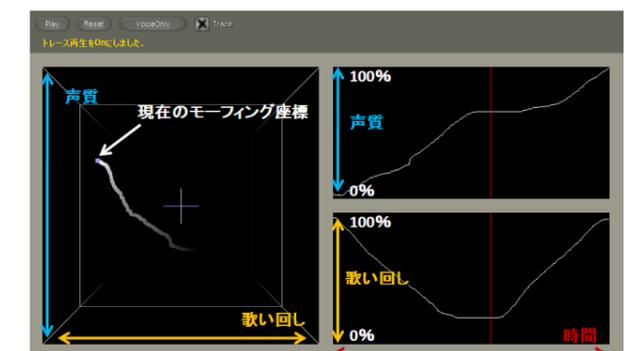


図 2 リアルタイム歌唱モーフィング



図 3 リアルタイム感情モーフィング