

# STRAIGHT を用いた音声モーフィングの 事例に基づくデザイン支援への応用

Extending STRAIGHT-based Speech Morphing  
for Case-Based Design Assistance

河原 英紀\*1 豊田 健一\*2 片寄 晴弘\*2  
Hideki Kawahara Ken'ichi Toyoda Haruhiro katayose

\*1和歌山大学 Wakayama University  
\*2関西学院大学 Kwansai Gakuin University

A high-quality speech morphing based on STRAIGHT speech analysis modification and synthesis system serves as a powerful tool for assisting case-based design, in other words design reuse, in vocal performance. This article introduces a new framework for performance design reuse based on two extensions in morphing concept; one is individual morphing of constituent parameters and the other is recursive morphing that enables morphing more than two instances. These two extensions are implemented on a conceptual framework called morphing object. Demonstrations on the current state of design reuse in vocal performance will also be presented at the conference.

## 1. はじめに

モーフィング技術には、物理パラメタと心理的屬性との対応関係についての明示的な知識に依存せずに、2つの試料の中間状態を容易に制作できることや、現実には存在しない効果を生み出すことができるという利点がある。これらの特長により、モーフィングはコンテンツデザインの有用なツールとして映像メディアの領域において広く実用に供されている。しかし、聴覚メディアの領域におけるモーフィングは、適切なツールの不在などの要因により、まだ十分に利用されるには至っていない。

この状況は、著者らによって発明された高品質音声分析変換合成法 STRAIGHT [Kawahara 99b] と、それを用いたモーフィング技術 [Kawahara 03] の開発により変わりつつある。明示的な知識に依存せずに高い自然性を有する刺激連続体を作成できるという STRAIGHT に基づくモーフィングの特長は、非分節的な構造と（議論はあるものの）非範疇的な性格を有する音声のパラ言語情報、非言語情報の研究手段としてまさに待ち望まれていたものである。歌声がそれを鑑賞する者に与える印象・感動は、その多くを音声のパラ言語・非言語情報に負っている。この意味において、歌声モーフィングへの応用 [Yonezawa 05, 豊田 06] は、能動的芸術鑑賞や新しいエンタテインメントを可能にする技術 [片寄 06] を構築するための重要な第一歩であると位置づけられる。

## 2. STRAIGHT を用いた歌声モーフィング

STRAIGHT は、Channel VOCODER のアーキテクチャに基づき、音声からフィルタ情報（スペクトル包絡）と音源情報とを分離し抽出する。STRAIGHT では、音源の基本周波数に適合する相補的な時間窓と周波数領域での spline 関数の理論に基づく適応平滑化により、調波位置での値を保存することと音源の周期性により生ずるスペクトル包絡への干渉の完全な除去を両立させている [Kawahara 99b]。

音源情報は、基本周波数と、帯域毎の周期成分と非周期成分の割合を表す非周期性指標とから構成される。基本周波数の抽出には、フィルタ群の中心周波数から出力の瞬時周波数への写像における不動点を利用したアルゴリズムが用いられる。非周期性指標は、見かけ上の基本周波数が定数となるような時間軸の伸縮と楕円フィルタに基づき、シミュレーションによる補正を加えて計算される\*1 [Kawahara 99a]。

STRAIGHT における（再）合成音声は、こうして求められた音源情報とスペクトル包絡から計算される。スペクトル包絡は、最小位相のインパルス応答に変換され、群遅延を操作された混合音源（パルス + 有色雑音）と畳み込まれて overlap and add により合成音声波形が求められる。

### 2.1 STRAIGHT を用いたモーフィング

STRAIGHT を用いたモーフィングは、情報表現が独立性の高い実数値であるため、事例間の線形補間により容易に実装できる。また、スペクトル包絡と音源情報という情報表現は、聴覚における属性との類似性が高く直感的であり、これまでに蓄積されている聴覚心理学や音声知覚の知見に基づく様々な操作にも容易に対応づけられる。

聴覚メディアのモーフィングにおいても、映像メディアにおけるモーフィングと同様に、モーフィングに用いる試料間の対応関係を指定する必要がある。ここでは、スペクトル包絡の系列を時間周波数表現として表示し、特徴的な位置に参照とするための点（基準点）を設定する。基準点は、時間方向では子音と母音からなる一つの音節について 4~5 個、周波数方向では、5000Hz までに 3~5 個程度で十分であることが分かっている。モーフィングの第一段階では、それらの基準点が重なるように、一方の試料の時間周波数平面を変形する。

こうして対応づけられた時間周波数平面の上で、それぞれの点においてモーフィング率に応じてパラメタを補間し、モーフィングされたパラメタ値を求める。最後に、モーフィング率に応じて、時間周波数平面を変形する。こうして求められたパラメタを STRAIGHT の合成部に渡すことにより、モーフィングされた音声合成される。

連絡先: 河原英紀, 和歌山大学システム工学部デザイン  
情報学科聴覚メディア研究室, 〒 640-8510 和歌山市  
栄谷 930, TEL: 073-457-8461, FAX: 073-457-8112,  
kawahara@sys.wakayama-u.ac.jp

\*1 基本周波数の抽出が困難な音源については、自己相関に基づく方法を併用した信頼度の高い方法 [Kawahara 05] が開発されている。

## 2.2 歌声モーフィングへの応用

歌声を対象とした場合、時間軸の変形と周波数軸の変形およびパラメタの補間操作は歌い回しや声質に深く関わっていると考えられる。前報では、この仮定に基づいて操作を以下のように分解することにより、歌い回しと声質を独立に操作したモーフィング音声の作成を試み、知覚に対する影響を調べた。

- 声質のモーフィング
  - 周波数軸の伸縮
  - スペクトル包絡および非周期性指標の補間
- 歌い回しのモーフィング
  - 時間軸の伸縮
  - 基本周波数の補間
  - 音量（ここではエネルギー）の補間

なお、心理属性との類似度を高くすることを狙い、基本周波数の補間では、対数周波数、スペクトル強度と音量の補間では、それぞれの dB 表示を用いた。対応点の付与においては、スケールスペースフィルタを用いた半自動の時間方向の選択機能を組み込んだ GUI[豊田 06] を利用した。

## 2.3 歌声モーフィングの知覚的検討

予備実験は、RWC 研究用音楽データベース：ポピュラー音楽に収録された楽曲 “So Long” (RWC-MDB-P-2001 No.64) のうち、サビ部分にあたる約 10 秒のフレーズを用いて行われた。モーフィングは、このスタジオミュージシャンによる歌唱と、同一曲のプロではない 20 代の男性の歌唱とを試料とした。

知覚実験では、声質のモーフィング率の評価、歌い回しのモーフィング率の評価、歌声の自然性評価、歌唱者の判別が被験者に求められた。詳細は前報 [豊田 06] に譲るが、補間の領域のモーフィング率に対しては、声質および歌い回しとの対応の単調性が示され、歌唱者の判別においては、声質が支配的であることが示された。物理パラメタと被験者の受容した属性値との対応は、非線形であった。直感的な操作の実現のためには、関数近似等による線形化が必要であることを意味する。

## 3. デザイン支援のための拡張

予備実験により、パラメタおよび時間軸、周波数軸の独立な変形操作が有用であることが確認された。重要なことは、デザイン支援への応用では、STRAIGHT により直接求められるパラメタのみではなくそれから導出される様々な（今回の例では音量）パラメタの操作が必要であることと、時間周波数写像という単一の関数ではなく時間写像と周波数写像という 2 つの独立な写像としての操作が必要であることの発見である。また、事例間の関係を二項関係に制限する必然性はない。言い替えば、任意の個数の事例間のモーフィングを実現する必要がある。これらの要請に柔軟に対応するために、モーフィングオブジェクトを導入し、モーフィングを、2 つのモーフィングオブジェクトから一つのモーフィングオブジェクトを生成する演算として定義することで再帰的なモーフィングの適用を可能とした。また、モーフィング率を、STRAIGHT により直接求められるパラメタだけではなく導出されたパラメタや、時間軸、周波数軸それぞれについて独立の値を持つベクトル量として定義する。図 1 に、Matlab に実装されたモーフィングオブジェクトの構造をそれぞれのメンバーの簡単な説明とともに示す。

```

date: % Date of creation
pwd: % Directory of the original data
waveform: % Original speech sample
samplingFrequency:
F0: % Fundamental frequency (Hz)
vuv: % Voiced or unvoiced indicator
spectrogram:
aperiodicityIndex:
frameUpdateInterval: % In ms
anchorTimeLocation:
maximumFrequencyPoints:
anchorFrequency:
F0extractionConditions: % Structured variable
SpectrumExtractionConditions: % ditto
creatorInformation: % Text string
  
```

図 1: Members of a morphing object (structured variable). Brief explanations start from % symbols.

## 4. まとめ

STRAIGHT を用いた歌唱モーフィングにおける事例ベースのデザイン支援に向けた拡張について紹介した。ここで紹介した拡張は、複数の演奏や歌手間のモーフィングを可能にし、また、歌い回しと声質のモーフィングを個別に操作することを可能にする。この拡張により、モーフィングによる歌唱制御の自由度が大きく広がるとともに、デザイン支援の検討における操作パラメタの追加などにも柔軟に対応することのできる基盤が整備された。今後は、多数の試料を用いた心理実験により、演奏や声質の様々な心理的属性に対する操作パラメタの貢献度を明らかにすることと、事例間のモーフィングを、事例間の『関係』の転写にまで拡張することが課題となる。

謝辞：本研究は、科学技術振興機構の CrestMuse プロジェクトによる支援を受けた。

## 参考文献

- [片寄 06] 片寄, 後藤: 音楽のデザイン転写技術の開発に向けて - CrestMuse プロジェクトの「価値」創出視点からの紹介 -, 本予稿集 (2006).
- [Kawahara 99a] Kawahara, H., Katayose, H., Cheveigné, de A., and Patterson, R. D.: Fixed Point Analysis of Frequency to Instantaneous Frequency Mapping for Accurate Estimation of F0 and Periodicity, *Eurospeech'99*, Vol. 6, pp. 2781-2784 (1999).
- [Kawahara 99b] Kawahara, H., Masuda-Katsuse, I., and Cheveigné, de A.: Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction, *Speech Communication*, Vol. 27, No. 3-4, pp. 187-207 (1999).
- [Kawahara 03] Kawahara, H. and Matsui, H.: Auditory morphing based on an elastic perceptual distance metric in an interference-free time-frequency representation, *ICASSP 2003*, Vol. I, pp. 256-259, Hong Kong (2003).
- [Kawahara 05] Kawahara, H., Cheveigné, de A., Banno, H., Takahashi, T., and Irino, T.: Nearly Defect-free F0 Trajectory Extraction for Expressive Speech Modifications based on STRAIGHT, *Interspeech2005*, pp. 537-540, Lisboa (2005).
- [豊田 06] 豊田, 片寄, 河原: STRAIGHT による歌声モーフィングの初期的検討, 情報処理学会研究報告, 第 2006-MUS-64 巻 (2006).
- [Yonezawa 05] Yonezawa, T., Suzuki, N., Mase, K., and Kogure, K.: Gradually Changing Expression of Singing Voice based on Morphing, *Interspeech2005*, pp. 541-544, Lisboa (2005).