

音色特徴量に基づく調波・非調波統合モデルによる楽器音モーフィング

安部 武宏 糸山 克寿 高橋 徹 駒谷 和範 尾形 哲也 奥乃 博

京都大学 大学院情報学研究科 知能情報学専攻

1. はじめに

本研究の最終目標は任意の楽器パートの音色を置換できる楽器音イコライザの開発である。吉井らの Drumix [1] では、打楽器単位での音量操作と音色置換を実現している。一方、糸山らの楽器音イコライザ [2] では、全ての楽器での音量操作が可能だが、Drumix で実現されていた音色置換は扱われていない。

本稿では、セミパラメトリックな音色特徴量の補間によって、複数の楽器音の音響特徴を含んだ音をモーフィング合成する手法を報告する*。モーフィング合成の実現により、複数の楽器音の音響特徴を含有した音色で、楽器パートの音色を置換することができる。音のモーフィングにおいて重要なことは、合成音に入力となる楽器音の音響特徴が自然性が失われることなく含有されることである。STRAIGHT によるモーフィングでは、ノンパラメトリックな音響特徴量を用いているために、分析と合成の可逆性が保証されておらず、数回の操作によって自然性が劣化すると報告されている [4]。パラメトリックな音色特徴量を用いれば、モデルが仮定されることで楽器音と音色特徴量と対応付けが一意に決定されるので、操作に対する頑健性が高まると期待される。

2. 音色特徴量に基づく楽器音のモーフィング

音響心理学の分野では、音色の聴感上の知覚の差はおもに、(1) 高周波数領域での倍音ピークの有無、(2) 発音時に発生する非調波成分、(3) 各ピークの時間方向における振幅の変動、の3つに起因する傾向があるとの報告がある [5]。我々はこれらから以下の3つの音色特徴量を定義する。

- (1) 倍音ピーク間の相対強度
- (2) 非調波成分の分布
- (3) 時間方向の振幅エンベロープ

複数個の対象音 (seed と呼ぶ) から、これら音色特徴量を分析し、各音色特徴量を補間することによってモーフィング音を合成する。本手法の概要を図1に示す。

2.1 調波・非調波統合モデルによる分析

音色特徴量を分析するために、糸山らが提案した調波・非調波統合モデル [2] に弦楽器音で観測される非調和性 (インハーモニシティ) を含む調波構造を分析できるよう拡張した音モデルを用いる。すなわち、以下のように seed のスペクトログラム $S(f, r)$ に対し、調波成分に対応するモデル $M_H(f, r)$ と非調波成分に対応するモデル $M_I(f, r)$ とを w_H 及び w_I で重み付けした混合モデルを適応させる。

$$S(f, r) = w_H M_H(f, r) + w_I M_I(f, r) \quad (1)$$

$$M_H(f, r) = \sum_n \frac{v_n E_n(r)}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(f - \mu_n(r))^2}{2\sigma^2}\right] \quad (2)$$

Musical instrument sound morphing based on psychoacoustic timbre characteristics using harmonic and inharmonic models: Takehiro Abe, Katsutoshi Itoyama, Toru Takahashi, Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno (Kyoto Univ.)

*時系列上における2音間での音響特徴の変化が音のモーフィングと指される場合 [3] もあるが、本稿では時系列とは独立な定常的な音響特徴の補間を音のモーフィングと定義している。

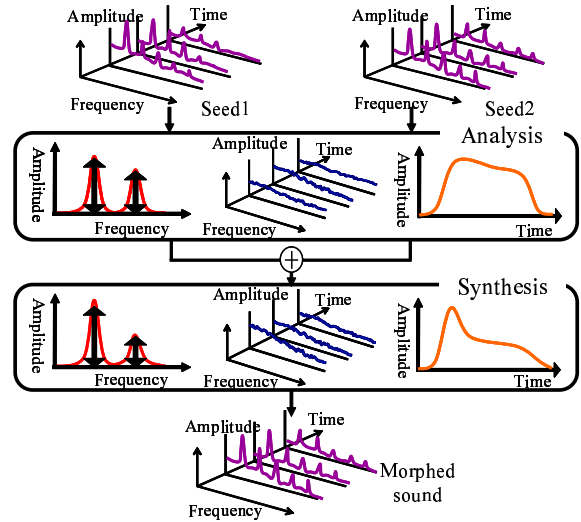


図1: 本手法の概要

ここで、 f と r は周波数とフレーム番地を表す。非調波成分は音色の繊細な差異に関係するので、ノンパラメトリックモデルで表現する。 $\mu_n(r)$ は n 次高調波成分に対応するピークの周波数である。本手法ではインハーモニシティを考慮した制約を加えることで、音高軌跡 $\mu(r)$ と非調和度 B で各ピークの周波数 $\mu_n(r)$ を以下のように表現する。

$$\mu_n(r) = n\mu(r)\sqrt{1 + Bn^2} \quad (3)$$

$E_n(r)$ と $M_I(f, r)$ にはそれぞれ制約条件として $\int E_n(r) dr = 1$, $\iint M_I(f, r) df dr = 1$ を設けており、EM アルゴリズムを用いた MAP 推定より音モデルのパラメータを導出することができる。M ステップにおけるパラメータ更新式の一部を示す。

$$\hat{\mu}(r) = \frac{\sum_n n \int S_n^{(H)}(f, r) f \sqrt{1 + Bn^2} df}{\sum_n n^2 \int S_n^{(H)}(f, r) (1 + Bn^2) df} \quad (4)$$

$$\hat{B} = \frac{2 \sum_n \iint S_n^{(H)}(f, r) n^3 \mu(r) (f - n\mu(r)) df dr}{\sum_n \iint S_n^{(H)}(f, r) (n^3 \mu(r))^2 df dr} \quad (5)$$

ここで、 $S_n^{(H)}(f, r)$ は分離された n 次高調波成分である。その他のパラメータ更新式は紙面の制約により省略する。我々が定義した音色特徴量である、(1) 倍音ピーク間の相対強度、(2) 非調波成分の分布及び (3) 時間方向の振幅エンベロープは、それぞれ v_n , $w_I M_I(f, r)$ 及び $E_n(r)$ に対応する。

2.2 音色特徴量の補間

音色特徴量は楽器音のオンセットとオフセットに基づいてアラインメントがとられた後、一定の混合率で補間される。各 seed の混合率 α_k は制約条件 $\sum_k \alpha_k = 1$ を満たし、 $0 < \alpha_k < 1$ では内挿、 $1 < \alpha_k$ あるいは $\alpha_k < 0$ では外挿となる。補間された各音色特徴量は次式によって得られる。

$$Feature^{(P)} = \exp\left[\sum_k \alpha_k \log(Feature^{(k)})\right] \quad (6)$$

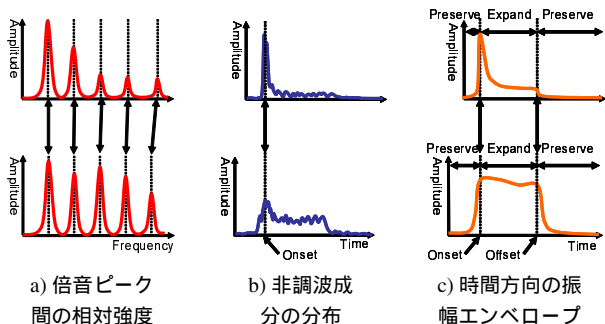


図 2: 音色特徴量のアラインメントの取り方

表 1: 本稿の実験で用いた楽器音の内訳

カテゴリ	楽器名	基本周波数 [Hz]
打弦楽器	ピアノ (PF)	55, 110, 220, 440, 880
撥弦楽器	アコースティックギター (AG)	110, 220, 440
	エレキベース (EB)	55, 110
擦弦楽器	バイオリン (VN)	220, 440, 880
	コントラバス (CB)	55, 110, 220
金管楽器	トランペット (TR)	220, 440, 880
	チューバ (TU)	55, 110, 220
木管楽器	アルトサクソ (AS)	220, 440, 880
	バリトンサクソ (BS)	110, 220

Feature には $v_n, M_I(f, r), E_n(r)$ といった音色特徴量が当てはまる。また、 k, P は、それぞれ各 seed へのインデックスと、補間された特徴量へのインデックスである。音色特徴量のアラインメントの取り方を図 2 に示す。倍音ピーク間の相対強度 v_n ではアラインメントは不要である。非調波成分の分布 $M_I(f, r)$ ではオンセットのみでアラインメントがとられる。一方、時間方向の振幅エンベロープ $E_n(r)$ ではオンセットとオフセットが揃うように音長操作された後アラインメントがとられる。

3. 評価実験

音のモーフィングにおいて音色特徴量をセミパラメトリックに表現することの有効性を評価するために、合成音と seed との距離を計測した。

3.1 実験条件

モーフィングの過程で生じた調波成分と非調波成分の歪みを評価するために、合成音と seed との距離を線形スペクトル距離と対数スペクトル距離との 2 つの評価尺度を用いた。線形スペクトル距離はエネルギーの大きい調波成分の歪み、対数スペクトル距離はエネルギーの小さい非調波成分の歪みを評価することができる。評価実験に用いた実楽器音には、RWC 研究用音楽データベースの楽器音データベース RWC-MDB-I-2001 に登録されている楽器音を利用した [6]。楽器音の内訳を表 1 に示す。評価に用いる合成音は内挿と外挿後、seed と同じ音響特徴を持つよう合成される。まず、seed の 2 音を混合率 $(\alpha_1, \alpha_2) = (1/3, 2/3), (2/3, 1/3)$ で内挿して 2 音のモーフィング音を得る。この 2 音のモーフィング音をさらに各 seed へ戻す混合率で外挿することで評価に用いる合成音が得られる。評価に用いた合成音の操作過程を図 3 に示す。モーフィングの組み合わせは表 1 に示した楽器の内、同じ基本周波数を持つ楽器から選択した。

3.2 評価結果

図 4 に、本手法によって合成した音の seed に対する線形スペクトル距離と対数スペクトル距離を示す[†]。9 種類の楽器のうち、線形スペクトル距離においてはチューバ、対数スペクトル距離においてはピアノにおいて距離

[†]組み合わせごとではなく、混合に用いた seed ごとに平均して距離を算出している。すなわち、本稿では 2 つの seed からの合成音について評価しているので、平均前の元データには同じ値が 2 つ含まれる。

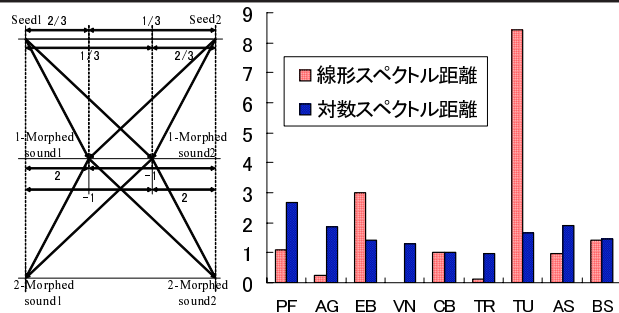


図 3: 評価に用いた合成音と seed 間のスペクトル成音の操作過程

図 4: 合成音と seed 間のスペクトル距離 (コントラバスで正規化)

が大きくなっていることがわかる。これら結果の考察を以下に示す。

1. チューバ (TU) における線形スペクトル距離の増加
チューバ音を調波モデルで十分に分析できていないのが原因である。分析ができない原因には主に以下の 2 つが挙げられる。

- (a) 低音の各ピークは周波数領域で密集する。
- (b) 発音体の慣性より各ピークが急峻に変動する [7]。

2. ピアノ (PF) における対数スペクトル距離の増加
ピアノ音の高周波数領域のピークが非調波モデルで過剰に適應されていることが原因である。調波モデルではピアノ音でみられるインハーモニシティを考慮しているが、非調波モデルへの過剰適應が起こると高周波領域のピークに依存する非調和度の分析精度が低下する。

合成音と seed を聴取にて比較してみたところ、パラメトリックな音色特徴量である倍音ピーク間の相対強度と時間方向の振幅エンベロープから合成された調波成分からは歪みがほとんど感じられなかった。2 回の操作で自然性が大きく劣化したという STRAIGHT でのモーフィングの報告 [4] を踏まえるとパラメトリックな音色特徴量によるモーフィングは頑健性が高いといえる。一方、ノンパラメトリックな音色特徴量である非調波成分の分布から合成された非調波成分からは歪みが大きく感じられた。本手法では非調波成分の繊細な差異を表現するために、ノンパラメトリックモデルを用いたが、非調波成分の微細構造を損なわずに、一部をパラメトリックに表現できれば、歪みが抑えられると期待される。

4. おわりに

本稿では、セミパラメトリックな音色特徴量を用いた頑健性の高い楽器音のモーフィング手法について報告し、本手法の楽器ごとでの有効性を定量的な評価より確認した。今後は音のモーフィングに使われる STRAIGHT などの従来手法と比較実験していく。

謝辞 本研究の一部は、科研費、GCOE、CREST-Muse の支援を受けた。

5. 参考文献

- [1] K. Yoshii et al. Drumix: An audio player with real-time drum-part rearrangement functions for active music listening. Vol. 48, No. 3, pp. 1229–1239, 2007.
- [2] 糸山他. 楽譜情報を援用した多重音楽音響信号の音源分離と調波・非調波統合モデルの制約付パラメータ推定の同時実現. Vol. 49, No. 3, pp. 1465–1479, 2008.
- [3] M. Slaney, M. Covell, and B. Lassiter. Automatic audio morphing. In Proc. ICASSP, pp. 1001–1004, 1996.
- [4] 高橋他. Straight を用いた反復分析再合成音声の評価. 日本音響学会秋季, pp. 289–290, 2007.
- [5] J. M. Grey. Multidimensional perceptual scaling of musical timbres. J. Acoust Soc. Am., Vol. 61, No. 5, pp. 1270–1277, 1977.
- [6] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka. RWC music database: Music genre database and musical instrument sound database. In Proc. ISMIR, pp. 229–230, 2003.
- [7] 早坂寿雄. 楽器の科学. 電子情報通信学会, 1992.