

# 楽器固有の音響的特徴を考慮した楽器音の音高操作手法

安部 武宏 糸山 克寿 吉井 和佳 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

音楽鑑賞を支援する技術として楽器単位でのイコライザの研究がいくつか報告されている。吉井らの Drumix [1] ではスネアドラムやバスドラムといった打楽器単位での音量操作と音色の置き換えを実現している。それに対して糸山らの楽器音イコライザ [2] では打楽器だけではなく、全ての楽器の音量を操作させることができる。しかし、Drumix で実現されていた音色の置き換えまでは及んでいない。

我々の最終目標は楽器音イコライザで実現されていなかった全ての楽器における音色の置き換えを実現することである。音色の置き換えを実現するためには、混合音から目的の楽器音を分離することと、分離された音から置き換える楽曲に含まれない音高・音長を持つ楽器音を合成することが要求される。そこで我々は後者に注目し、分離された楽器音から音色の特徴を保持したまま任意の音高・音長を持つ楽器音を合成する課題に取り組む。

本稿では特に楽器固有の音響的特徴が歪まないように音高を操作する手法を報告する。Grey によると音色の聴感上知覚の差をスペクトル上では 3 つの要因: 1. 高周波数領域での倍音ピークの有無, 2. 各ピークの時間方向における振幅の変動, 3. 励振時に発生する非調波成分の分布によって対応づけられる傾向があるという [3]。この知見に基づいて我々はスペクトル上に表れる音色の特徴量として 1. 倍音ピーク間の相対強度, 2. 時間方向の振幅変動, 3. 非調波成分を定義する。これら 3 つの特徴量を調波・非調波統合モデル [2] に基づいて分析することで、合成時の音高操作による楽器固有の音響的特徴の歪みを回避する。

## 2. 楽器音の分析

本章では、調波・非調波統合モデルによって楽器固有の音響的特徴を分析する方法について述べる。

### 2.1 調波・非調波統合モデルによる分析

音高を操作する楽器音 (本稿では *seed* と呼ぶ) に含まれる楽器固有の音響的特徴を分析するために、図 1 に示すような調波・非調波統合モデル [2] を用いる。このモデルは調波モデル  $M_H$  と非調波モデル  $M_I$  から成り立ち、以下のように定義される。

$$M(f, t) = w(M_H(f, t) + M_I(f, t)) \quad (1)$$

ここで、 $f$  と  $t$  はそれぞれ周波数と時間、 $w$  は *seed* の音量である。 $M_I$  はパワースペクトルによって直接表現される。対して  $M_H$  は以下のように 2 つの独立した関数  $E_f$  (図 2) と  $E_t$  (図 3) の積によって表現される。

$$M_H(f, t) = E_f(f)E_t(t) \quad (2)$$

$E_f$  と  $E_t$  はそれぞれ周波数方向と時間方向のエンベロープであり、以下のように定義される。

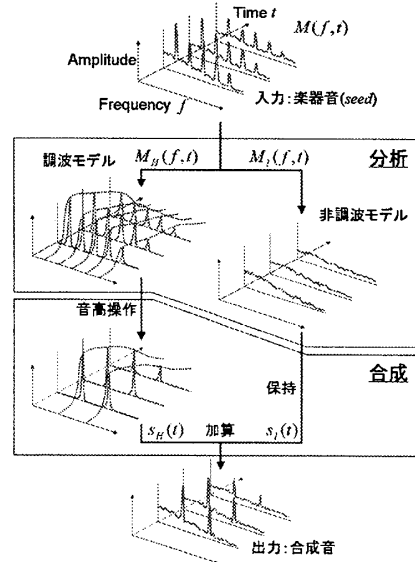


図 1: 音高の操作の概要

$$E_f(f) = \sum_n w_n \mathcal{N}(f - n\mu, \sigma^2) \quad (3)$$

$$E_t(t) = \sum_m w_m \mathcal{N}(t - \tau - m\phi, \phi^2) \quad (4)$$

ここで、 $\mathcal{N}(x, y^2)$  は平均  $x$ 、分散  $y^2$  のガウス関数であり、*seed* の音高に関するパラメータは  $E_f(f)$  の中に存在する  $\mu$  である。本稿では、 $w_n$ 、 $E_t$ 、 $M_I$  をそれぞれ、倍音ピーク間の相対強度、時間方向の振幅変動、非調波成分といった音色の特徴量として扱っている。推定の対象となるのは  $w_n$ 、 $\mu$ 、 $\sigma$ 、 $w_m$ 、 $\tau$ 、 $\phi$ 、 $w$ 、 $M_I$  であり、EM アルゴリズムに基づく反復計算によって統合モデルを *seed* のスペクトログラムへ適応させることで推定する。

### 2.2 平均化処理による分析誤差の抑制

我々の最終目標では分析される楽器音 (*seed*) は混合音から分離された音である。つまり、完全に分離できていない音を分析することが想定され、これにより分析された特徴量に誤差を含む可能性がある。そこで、分析誤差を抑制する手段として同一楽器の複数の *seed* からこれらを推定し、楽器固有の音響的特徴として扱っている  $w_n$  と  $M_I$  に対して平均化処理を施す。平均化処理によって正しく分離された楽器音の特徴量を強めることができ、分析誤差が抑制されることが期待される。以下では、平均化処理を施した  $w_n$  と  $M_I$  をそれぞれ  $\bar{w}_n$  と  $\bar{M}_I$  とする。

## 3. 楽器音の合成

本章では、分析された楽器固有の音響的特徴を保存しながら音高を操作する方法について述べる。図 1 のように調波成分と非調波成分を分離した後、音高を操作した調波成分と非調波成分からそれぞれ調波信号  $s_H(t)$  と非調波信号  $s_I(t)$  を別々に合成する。これらを以下のように合算することによって操作された音高  $\mu'$  をもつ合成

Synthesis approach for manipulating pitch of a musical instrument sound with considering timbral characteristics: Takehiro Abe, Katsutoshi Itoyama, Kazuyoshi Yoshii, Kazunori Komatani, Tetsuya Ogata, and Hiroshi G. Okuno (Kyoto Univ.)

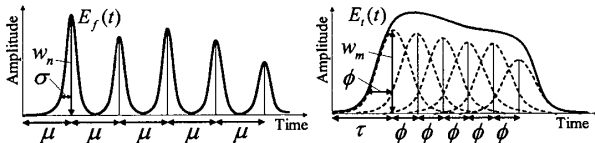


図2: 周波数エンベロープ 図3: 時間エンベロープ

音源  $s(t)$  を得る.

$$s(t) = s_H(t) + s_I(t) \quad (5)$$

調波信号  $s_H(t)$  は正弦波重畳モデルに基づいて合成される。我々は  $n$  番目の時間  $t$  における音高  $n\mu'$  を持つ正弦波の振幅  $A_n(t)$  を以下のように定義する。

$$A_n(t) = w\bar{w}_n E_t(t) \quad (6)$$

よって、調波信号は以下の式で表現することができる。

$$s_H(t) = \sum_n A_n(t) e^{jn\mu't} \quad (7)$$

非調波信号  $s_I(t)$  は、 $\bar{M}_I$  と  $seed$  の信号から得られる位相を用いたオーバーラップ加算を行うことで合成される。

#### 4. 評価実験

本手法の有効性と限界を評価するために様々な音高で本手法にて楽器音を合成し、同じ音高を持つ実際の楽器音とのスペクトル距離を計測することで定量的な評価を行った。

##### 4.1 実験条件

スペクトル距離は  $\sum_{f,t} (S_{synth}(f,t) - S_{seed}(f,t))^2 / T$  のように、合成音のスペクトログラム  $S_{synth}$  と合成音と同じ音高をもつ実際の楽器音のスペクトログラム  $S_{seed}$  の差分のフレーム間平均として定義される。ここで  $T$  はフレーム数である。スペクトル距離が小さいほうが実際の楽器音に近いといえる。距離を測る前の処理として比較する音源の同期をとり、持続時間が短いほうの音源のフレーム数にて距離の計算を行った。また、同一楽器の楽器音間での音量の違いによって発生するスペクトル距離の差を解消するために、実際の楽器音と合成音とのエネルギー(スペクトログラムの合計値)が等しくなるよう調節した。

$seed$  には7種類の楽器(ピアノ、ギター、ヴァイオリン、オーボエ、クラリネット、トランペット、アルトサククス)ごとに1オクターブ内で半音づつ異なる12種類の楽器音(C, C#, D, D#, ..., B)を用意した。これらの  $seed$  を全て分析し、それらを平均化して得られた特徴量を用いて、各  $seed$  に対して  $k$  半音 ( $k = -12, \dots, 12$ ) ほどの音高を操作し、合成した。なお、 $seed$  や比較に使う実際の楽器音には RWC 楽器音データベース (RWC-MDB-I-2001) [4] で提供されている楽器音 (44.1 kHz, モノラル) を用いた。

##### 4.2 評価結果

図4に本実験の結果を示す。ほぼ全ての楽器において音高を操作する半音数の絶対値が大きくなるほどスペクトル距離が大きくなる傾向を確認できる。中でもピアノ、クラリネット、トランペットにおいて、音高が高くなるように操作したときのスペクトル距離の大きさが顕著に表れている。つまりこれらの楽器には音高依存性が確認できる。一方ヴァイオリン、オーボエ、アルトサククスにおいては音高依存性がさほど確認されなかった。また、ギターにおいては音高操作の半音数とスペクトル距離の差に相関がみられなかった。この結果はギターにおいては音高依存性よりもギターフレットの押弦位置の違いに

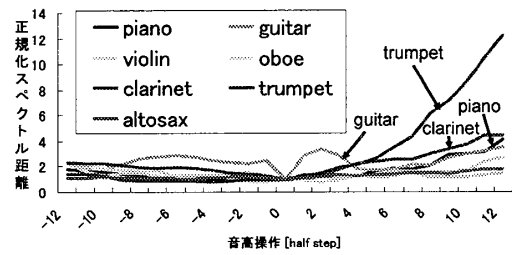


図4: 実際の楽器音と合成音との正規化スペクトル距離。横軸は半音単位での音高操作の値、縦軸は正規化スペクトル距離で音高操作されていないときのスペクトル距離で正規化している。

よる音色の変化の影響が大きいためだと考えられる。なお、楽器音の分類の研究 [5] では音高依存性を考慮することによって誤認識が抑制されたという報告もあり、改善された誤認識率\*と本実験でのスペクトル距離にも相関がみられる。

本稿では音色の特徴量を音高に依存しないものとして扱った。しかし、音高依存性が確認される楽器音の合成においてはこの仮定は好ましくない。音高依存性のある音響的特徴を考慮することによって、さらに実際の楽器音に近い音を合成できることが期待される。

#### 5. 音長操作の課題

本稿では特に音高操作手法について報告したが、楽器固有の音響的特徴を歪めないで音長を操作するにはいくつかの課題がある。単純には時間方向のエンベロープである  $E_t$  に含まれるガウス関数の分散  $\phi^2$  を調節することによって本手法でも音長操作することが可能である。しかし、単純な時間方向のエンベロープの均一な引き延ばしではスペクトル上での音色の特徴として定義している時間方向の振幅変動が歪められる。特に、楽器音の立ち上がりや楽器音の知覚に重要であることは古くから知られ、これの歪曲は音響的特徴を大きく変化させる。この問題を考慮した音長操作手法の提案を今後の課題とする。

#### 6. おわりに

本稿では調波・非調波統合モデルによって楽器固有の音響的特徴を分析することで、これが歪まないように音高を操作し、合成する手法を報告した。今後は、音色の特徴が歪まないような音長操作方法を検討し、音高依存性のある特徴も考慮した手法へと拡張する。

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

#### 7. 参考文献

- [1] K. Yoshii et al. Drumix: An audio player with real-time drum-part rearrangement functions for active music listening. *IPSP Journal*, Vol. 48, No. 3, pp. 1229-1239, 2007.
- [2] 糸山他. 多重奏音楽音響信号の音源分離のための調波・非調波モデルの制約付きパラメータ推定. *音情研*, 第71巻, pp. 57-60, 2007.
- [3] J. M. Grey. Multidimensional perceptual scaling of musical timbres. *J. Acoust. Soc. Am.*, Vol. 61, No. 5, pp. 1270-127, 1977.
- [4] 後藤他. RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース. *情処論*, Vol. 45, No. 3, pp. 728-738, 2004.
- [5] 北原他. 音高による音色変化に着目した楽器音の音源同定:  $f_0$  依存多次元正規分布に基づく識別手法. *情処論*, pp. 2448-2458, 2003.

\*改善された誤認識率はそれぞれピアノ 35.13%, ギター 14.38%, ヴァイオリン 13.10%, オーボエ 2.35%, クラリネット 14.82%, トランペット 32.50%, アルトサククス 10.90%, であった。