音楽音響信号に含まれる調波音の周波数特性とドラムの音色の転写システム*

☆中村友彦[1]，吉井 和佳[2]，後藤 真孝[2]，亀岡 弘和[1]
[1] 東大院・情報理工，[2] 産総研

1　はじめに

既存楽曲を自由に加工するシステムの実現は，音楽信号処理の重要課題の一つであり，音楽的な専門知識のないユーザでも直感的に利用できるシステムの構築を目指し研究が進められている [1–3]．楽譜情報が利用できる場合には，音楽音響信号中の各楽器の音量を変えられるシステムや，音色とフレーズの置換が可能なシステムが提案された [1,2]．また，楽譜情報が利用できない場合にも，音楽音響信号に含まれるバスドラムとスネアドラムの音量や音色，リズムパターンを置換可能なシステムも提案されている [3]．しかし，このシステムではドラムのみを対象としており，変換後のドラムの単音源を用意する必要があった．

我々は，より柔軟な加工と音源の用意の簡便さを目指し，ある音楽音響信号（ターゲット）と異なる音楽音響信号（リファレンス）を入力とし，楽譜情報を用いずに調波楽器音の周波数特性とドラムの音色を転写するシステムを提案する（図 1）．提案システムでは，最初にスペクトログラム上で調波楽器音が時間方向に滑らか，打楽器音が周波数方向に滑らかという性質から分離を行う調波打楽器音分離 [4] を用いて，ターゲットとリファレンスのスペクトログラムを調波楽器音成分（歌声を含む）と打楽器音成分に分離する．その後，(1) 調波楽器音成分のスペクトルの周波数特性を解析し，(2) リファレンスからターゲットに周波数特性を転写する．一方打楽器音成分に対しては，(a) さらに各ドラム楽器に分離し，ユーザがどのドラム楽器をどのドラム楽器に変換するかを選択した後，(b) ドラムの音色をリファレンスからターゲットに転写する．

2　調波楽器音の周波数特性転写モジュール

周波数特性を解析し転写するために，亀岡らによって提案された手法が利用できる [5]．この手法では，スペクトルのディップとピークを通るようなエンベロープ（ボトムエンベロープとトップエンベロープ）を介して振幅スペクトルを変形することにより，音高を変化させることなくスペクトルの周波数特性を変える．ここで，ボトムエンベロープは歌声の子音や楽器音のアタック時の音に含まれる傾向のある平坦なスペクトル成分，トップエンベロープは調波楽器音のもつ調波構造に相当するラインスペクトル状の成分に対応する．そのため，これらのエンベロープを変形することにより，近似的に調波楽器の周波数特性を変換できる．

当該モジュールでは，各スペクトルにボトム・トップエンベロープの推定を行ったのち，ターゲットとリファレンスのエンベロープを近づけるように，ターゲットの調波打楽器音成分の振幅スペクトルを変形する．エンベロープの推定法と与えられたエンベロープに対する振幅スペクトルの変形法は亀岡らの方法 [5] を利用できるため，以下ではターゲットとリファレンスのエンベロープを近づける手法を提案する．

2.1　ボトム・トップエンベロープの転写法

エンベロープの統計量（時間平均や分散）は，調波楽器音成分の全体的な周波数特性を表すと考えられる．そのため，ターゲットのエンベロープの統計量をリファレンスのエンベロープの統計量に一致させるような転写を考え，リファレンスからターゲットに周波数特性を転写する．

ターゲットのボトムエンベロープに周波数インデックス ω 毎にゲイン g_ω を加えた時間平均と分散を，リファレンスのボトムエンベロープの時間平均と分散に近づけたい．ここで，ボトムエンベロープが ω に関して独立な正規分布に従うとすれば，この正規分布同士の距離を最小化することによって，そのような g_ω を導出できる．周波数 ω 毎のターゲットとリファレンスのボトムエンベロープの時間平均と分散をそれぞれ $(\mu_\omega^{(\mathrm{tar})}, V_\omega^{(\mathrm{tar})}), (\mu_\omega^{(\mathrm{ref})}, V_\omega^{(\mathrm{ref})})$ とし，分布同士の距離基準として Kullback-Leibler ダイバージェンスを用いると，

$$g_\omega = \frac{\mu_\omega^{(\mathrm{tar})}\mu_\omega^{(\mathrm{ref})} + \sqrt{(\mu_\omega^{(\mathrm{tar})}\mu_\omega^{(\mathrm{ref})})^2 - 4\{V_\omega^{(\mathrm{tar})} + (\mu_\omega^{(\mathrm{tar})})^2\}V_\omega^{(\mathrm{ref})}}}{2\{V_\omega^{(\mathrm{tar})} + (\mu_\omega^{(\mathrm{tar})})^2\}} \tag{1}$$

とゲインが計算できる．トップエンベロープにも同一の処理を行った後，得られたゲインをそれぞれ掛けあわせたボトム・トップエンベロープを持つように，ターゲットの振幅スペクトルを変形すれば，リファレンスの周波数特性が転写できる．

3　ドラムの音色転写モジュール

当該モジュールでは，まず打楽器音成分のスペクトログラムを，非負値行列分解 [6] と Wiener フィルタを用いて近似的に各ドラム楽器のスペクトログラム（基底スペクトログラム）に分解する．非負値行列分解は，振幅スペクトログラムを低次元の 2 つの非負値の行列（楽器音スペクトルのテンプレートを表す基底行列と，各楽器の音量の時間発展を表すアクティベーション行列）の積によって近似する．この分解後，ターゲットのどのドラムの音色をリファレンスのどのドラムの音色に変換するかを，ユーザが決定する．これは，非負値行列分解によって得られた基底を選択することで実現できる．この選択に従って，基底のペア毎に変換を行うことによりリファレンスからターゲットにドラムの音色を転写する．以下では，各基底のペア毎にドラムの音色転写を議論する．

3.1　イコライジング法

簡易な手法の一つは，イコライザーのように，選択されたターゲットの基底と対応するリファレンスの基底の各周波数でのパワー比をゲインとしてターゲットの基底スペクトログラムを変形する方法である（以下，イコライジング法）．この方法では，基底のペア間でドラムの音色が大きく異なる場合（片方が高周波帯域に，もう片方が低周波帯域にエネルギーを持つ場合等）には，エネルギーの低い周波数帯域が過剰に増幅されてしまい，転写後の音質が劣化しやすく，ドラムの音色転写がユーザに知覚されにくい．

3.2　切り貼り法

上述の問題を避けるために，我々はリファレンスの基底スペクトログラムの各フレームでのスペクトル（基底スペクトル）を，ターゲットのドラムの音量の時間発展に合わせ切り貼りする方法（切り貼り法）を提案する（図 2）．異なる楽曲同士のドラムの音量の時間発展は一般に異なるため，切り貼り法を実現するためには，リファレンスのドラムの音量を参照しつつ基底スペクトルを適切に切り貼りし，ターゲットのドラムの音量の時間発展に近づける必要がある．音量の時間発展を表す特徴量として，非負値行列分解によって得られたアクティベーションが利用できる．

---

* System for Replacement of Drum Timbres and Frequency Characteristics of Harmonic Sound in Music Acoustic Signals by Tomohiko Nakamura (The University of Tokyo), Kazuyoshi Yoshii, Masataka Goto (National Institute of Advanced Industrial Science and Technology) and Hirokazu Kameoka (The University of Tokyo)

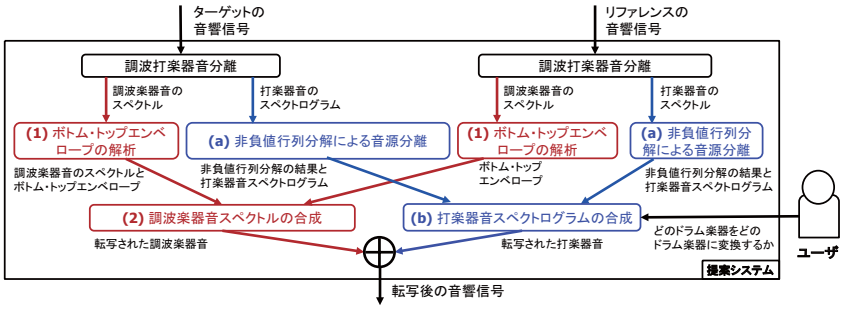

Fig. 1　提案システムの処理フロー．赤のモジュールが調波楽器音，青のモジュールが打楽器音を処理する．

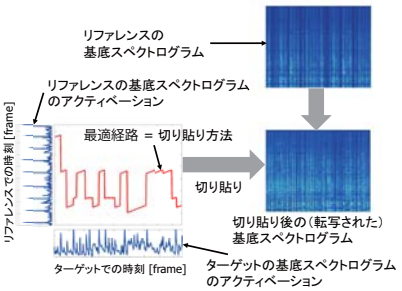

Fig. 2　切り貼り法による基底ペア毎のドラムの音色転写の処理概要．

さらに，合成時に生じうる雑音を低減するため以下の 3 つの要請を課す．離れた時刻の高いエネルギーを持つスペクトルを隣接して並べると，雑音が発生しやすいため，(i) 可能な限り時間的に連続したセグメントを使用し，(ii) セグメント境界は低いエネルギーの時刻に位置させる．また，楽譜情報なしの楽器音への分離は完全に行うことは難しいため，基底スペクトログラムは打楽器音以外の音を含むことがある．そのため，切り貼りする際に (iii) 打楽器音以外の音を含むスペクトルの使用を避ける．

これらの要請を満たす切り貼り方法は，要請 (i), (ii), (iii) をコスト関数に含めた動的計画法により得られる．累積コスト $\mathcal{I}_t(\tau)$ は，

$$\mathcal{I}_t(\tau) := \begin{cases} O_{t,\tau} & (t=1) \\ O_{t,\tau} + \max_{\tau'}\{C_{\tau',\tau} + \mathcal{I}_{t-1}(\tau')\} & (t>1) \end{cases}, \quad (2)$$

$$O_{t,\tau} := \alpha D(\tilde{U}_t^{(\mathrm{tar})} \| \tilde{U}_\tau^{(\mathrm{ref})}) + \beta P_\tau, \quad (3)$$

と定義できる．ここで，t, τ はそれぞれターゲットとリファレンスの時刻インデックス，$\alpha, \beta > 0$ は各項の累積コストに対する寄与を調節するパラメータである．(3) の第 1 項は，2 つの正規化されたアクティベーション $\tilde{U}_t^{(l)} := U_t^{(l)} / \max_t\{U_t^{(l)}\}, (l = \mathrm{tar}, \mathrm{ref})$ 間の一般化 I ダイバージェンスである．P_τ は，τ 番目のリファレンスの基底スペクトルが打楽器音以外の音をどの程度含むかを表し，打楽器音以外の音を含むほど大きくなる（要請 (iii)）．$C_{\tau',\tau}$ はリファレンスでの τ' 番目から τ 番目のフレームへの遷移コストを表し，

$$C_{\tau',\tau} := \begin{cases} 1 & (\tau = \tau' + 1) \\ c + \gamma(\tilde{U}_{\tau'}^{(\mathrm{ref})} + \tilde{U}_\tau^{(\mathrm{ref})}) & (\tau \neq \tau' + 1) \end{cases}. \quad (4)$$

と定義される．ここで，定数 c を $c > 1$ とすることにより，$\tau + 1$ 番目のフレームへの遷移を他の遷移よりも起こりやすくできる（要請 (i)）．$\tau \neq \tau' + 1$ での (4) の第 2 項は，アクティベーションが低い時に離れた時刻への遷移が起こりやすいことを示しており（要請 (ii)），$\gamma > 0$ は累積コストへの寄与を調節するパラメータである．このように定義された累積コストを最小にする最適経路として，切り貼り方法が得られる．

要請 (iii) で述べた通り，分離が完全でないために，ターゲットの基底スペクトルも打楽器音以外の音を含むことがある．切り貼りして得られた基底スペクトログラムではこの楽器音成分が失われているため，転写後の音が薄くなりやすい．ターゲットの基底スペクトルに含まれる打楽器音以外の楽器音成分を復元するために，この楽器音成分が打楽器音成分に比べ低いエネルギーを持つ傾向があることを利用する．切り貼りによって得られた基底スペクトログラムを全て加算した打楽器音スペクトログラムに対して，フレーム毎に全周波数ビンの振幅の和が定数 ϵ よりも大きければターゲットの打楽器音スペクトルと置換する．

4　主観評価実験

提案システムの性能を評価するため主観評価実験を行った．ターゲットとリファレンスとして，RWC ポピュラー音楽・ジャンルデータベース [7] から 3 曲の音響信号をそれぞれ 10 s 切り出し，22.05 kHz にダウンサンプルして用いた．3 楽曲の全組み合わせ計 6 ペアに関して，調波楽器音の周波数特性とドラムの音色の転写を行った．ドラムの音色転写に関しては，著者の一人がどのドラムの音色をどのドラムの音色に変換するかを選択し，イコライジング法と切り貼り法を比較した．音響信号とスペクトログラム間の変換は，512 点のハン窓，256 点のシフト長の短時間フーリエ変換と短時間逆フーリエ変換を用いた．パラメータは $(\alpha, \beta, \gamma, c, \epsilon) = (0.5, 3, 10, 3, 100)$ に設定し，非負値行列分解の基底数は 4 として一般化 I ダイバージェンスを距離基準として使用した．P_τ として，L2 正則化付き L1 損失サポートベクターマシン [8] で得られた負の対数事後確率を用いた．サポートベクターマシンは，RWC の楽器音データベース [7] を打楽器音とそれ以外の楽器音の 2 クラスで学習した．

主観評価の項目として，(1)「ドラムの音色がリファレンスからターゲットに適切に転写されているか」，(2)「調波楽器音の音色がリファレンスからターゲットに適切に転写されているか」を用いて，11 人の被験者により 5 段階評価（1 点を全く転写されていない，5 点を完全に転写されていた）を行った．被験者は，ターゲットとリファレンス，転写後の音楽音響信号とそれらの調波楽器音と打楽器音を何度も聞き直せた．

項目 (1) に対し，標準誤差付きの平均スコアは 2.5 ± 0.1（イコライジング法），2.8 ± 0.1（切り貼り法）であった．切り貼り法のスコアがイコライジング法に比べ平均値は高く，特に 3 節で述べたようにドラムの音色が大きく異なる場合に切り貼り法のスコアが高かった．また，項目 (2) に対する標準誤差付きの平均スコアは 2.5 ± 0.1 であった．特に，高周波帯域が変化する際にスコアが上昇する傾向があった．これらの結果は，被験者がドラムの音色と周波数特性の転写を知覚したことを示しており，システムが期待通り動作していることを確認した．

5　結論

本研究では，楽譜情報なしで異なる音楽音響信号間での周波数特性とドラムの音色を転写するシステムを提案した．主観評価実験により，異なる音楽音響信号間での転写が周波数特性の転写とドラムの音色の両者について有効であることが確認された．今後は，ユーザが自由に転写度合いを調節可能なユーザインターフェースの開発や，声質の転写も課題である．

参考文献

[1] K. Itoyama *et al.*, In *Proc. of ICASSP*, 1, pp. I–57–I–60, 2007.

[2] N. Yasuraoka *et al.*, In *Proc. of ACM-MM*, pp. 203–212, 2009.

[3] K. Yoshii *et al.*, *Trans. IPSJ*, 48(3), pp. 1229–1239, 2007.

[4] H. Tachibana *et al.*, In *Proc. of ICASSP*, pp. 465–468, 2012.

[5] 亀岡 弘和 他, 情処研報, 49, pp. 139–144, 8, 2006.

[6] D. Seung *et al.*, *Adv. Neural Inf. Process. Syst.*, 13, pp. 556–562, 2001.

[7] M. Goto, In *Proc. of ICA*, pp. I–553–556, 2004.

[8] R. Fan *et al.*, *JMLR*, vol. 9, pp. 1871–1874, 2008.