

音楽音響信号中の調波音の周波数特性および ドラムの音色の置換システム

中村 友彦^{1,a)} 吉井 和佳^{2,b)} 後藤 真孝^{3,c)} 亀岡 弘和^{1,d)}

概要：本論文では、調波楽器音の周波数特性とドラムの音色を、音楽音響信号間で楽譜を用いて置換するシステムを提案する。このシステムでは、まず置換元の音楽音響信号（インプット）と置換先の音楽音響信号（リファレンス）の振幅スペクトルをそれぞれ調波楽器音成分と打楽器音成分のスペクトルに分離し、それぞれの成分に対して独立に処理を行う。調波楽器音成分のスペクトルの周波数特性をスペクトルの山周辺と谷周辺を通る2つのスペクトル包絡によって特徴付け、インプットの調波楽器音成分の振幅スペクトルを、インプットとリファレンスの調波楽器音成分のスペクトル包絡が類似するように変形する。インプットとリファレンスの打楽器音成分のスペクトログラムは、各ドラム楽器毎のスペクトログラムに分離した後、ユーザによって指定されたインプットのドラム楽器の音色をリファレンスのドラム楽器の音色に置換する。主観評価実験により、提案するシステムが周波数特性とドラムの音色を適切に置換できることを確認した。

1. はじめに

既存楽曲をユーザ好みに合わせ自由に編集可能なシステムの実現は、音楽信号処理の重要課題の一つである。編集機能として、楽曲中の楽器の音色や周波数特性を他の楽曲のものと置換することは重要である。この置換が実現できれば、例えば楽曲のドラムを他の楽曲と差し替え、ユーザ好みに応じて雰囲気を変換したりミックス音源をユーザ自身が作成できる。楽曲制作過程では、オーディオエンジニアが音響信号の周波数特性を変えるイコライザ[1-5]などのエフェクタを用いてこのような編集を行うことがある。しかし、エフェクタを用いた変更には高度な音響編集技術が必要である。そのため、音響編集技術や音楽の専門知識がないユーザでも、音楽音響信号を直感的に編集できるシステムの開発が重要である。

このような音楽音響信号の編集システムを目指し多くの研究が進められている[6-10]。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[6]や、音色とフレーズの置換が可能なシステム[9]が提

案されている。楽譜情報の代わりに事前に得られた歌詞の音素と基本周波数を用いて、音楽音響信号中の歌声の声質を単独歌唱の声質に変換する手法も提案されている[10]。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調波楽器音と打楽器音の音量を自由に調節可能なシステム[8]や、打楽器音のリズムパターンを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[7]も提案されている。しかし、吉井らのシステム[7]では置換先の音源（以下、リファレンス）として事前にドラムの単音源を用意する必要があり、リファレンスとして既存楽曲の音響信号を用いることは対象としていなかった。

我々は、より柔軟な加工と音源の用意の簡便さを目指し、置換元の音楽音響信号（以下、インプット）とリファレンスとなる音楽音響信号を入力とし、楽譜情報を用いて調波楽器音の周波数特性とドラムの音色を置換するシステムを提案する（図1）。提案システムでは、調波楽器音と打楽器音の両方を処理したい場合だけでなく、調波楽器音または打楽器音の片方だけ処理したい場合にも対応するため、最初にインプットとリファレンスの音響信号を調波楽器音成分（歌声を含む）と打楽器音成分に分離する。この分離には、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に基づき分離を行う調波打楽器音分離[11]を用いることができる。この調波打楽器音分離により得られたインプットとリファレ

¹ 東京大学 大学院情報理工学研究科,
東京都文京区本郷 7-3-1, 113-0033.

² 京都大学 大学院情報学研究科,
京都府京都市左京区吉田本町, 606-8501.

³ 産業技術総合研究所, 茨城県つくば市梅園 1-1-1, 305-8568.

a) nakamura@hil.t.u-tokyo.ac.jp

b) yoshii@kuis.kyoto-u.ac.jp

c) m.goto@aist.go.jp

d) kameoka@hil.t.u-tokyo.ac.jp

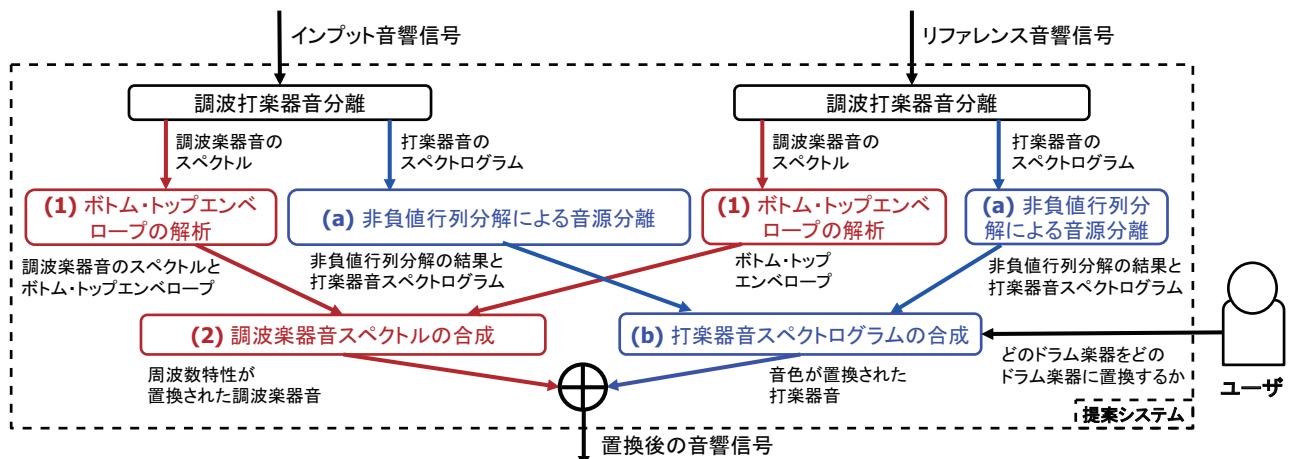


図 1 提案システムの処理フロー。赤のモジュールが調波楽器音、青のモジュールが打楽器音を処理する。

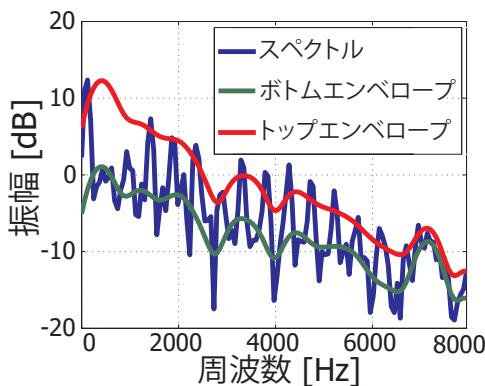


図 2 振幅スペクトル（青線）に対するボトムエンベロープ（緑線）とトップエンベロープ（赤線）の例。

ンスの歌声を含む調波楽器音成分と打楽器音成分のスペクトログラムに対し処理を行う。(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) 音高を変化させずにインプットの調波楽器音成分の周波数特性をリファレンスの調波楽器音成分の周波数特性に置換する。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) インプットのドラムの音色をリファレンスのドラムの音色に置換する。

2. 調波楽器音の周波数特性の置換

周波数特性を解析し置換するために、亀岡らの手法 [12] が利用できる。この手法では、図 2 のようにスペクトルの谷周辺と山周辺を通るようなエンベロープ（以下、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、近似的に調波楽器音の周波数特性を変換できる。

周波数特性の置換では、インプットとリファレンスの各スペクトルにボトムエンベロープとトップエンベロープの推定をそれぞれ行う。その後、インプットとリファレンスのエンベロープが類似するようにインプットのエンベロープを変形し、インプットの調波打楽器音成分の振幅スペクトルを変形されたエンベロープを持つように修正する。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 [12] を利用できるため、以下ではインプットのエンベロープをリファレンスのエンベロープに類似させる手法を提案する。

2.1 ボトムエンベロープとトップエンベロープの置換法

エンベロープの統計量（時間平均や分散）は、調波楽器音成分全体の周波数特性を反映していると考えられる。そのため、インプ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ゲインを用いて、インプットの調波楽器音成分の周波数特性をリファレンスの調波楽器音成分の周波数特性に置換する。

インプットのボトムエンベロープに周波数インデックス ω 每にゲイン 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、 g_ω を導出できる。周波数 ω 毎のインプ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 $(\mu_\omega^{(in)}, V_\omega^{(in)})$, $(\mu_\omega^{(ref)}, V_\omega^{(ref)})$ 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、

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

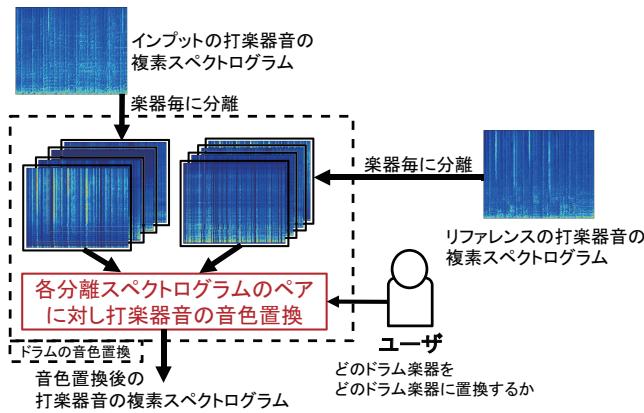


図 3 ドラムの音色置換の処理フロー .

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、得られたゲインを適用したボトムエンベロープとトップエンベロープを持つように、インプットの振幅スペクトルを周波数成分毎に変形する。これらの操作によって、リファレンスのボトムエンベロープとトップエンベロープの時間平均と分散をもつ、インプットの振幅スペクトルが得られる。

2.2 ステレオ音響信号への簡易な拡張

上述の処理はモノラルの音楽音響信号に対する処理であるが、インプットとリファレンスがステレオ音響信号の場合にも簡易に拡張できる。全チャネルにわたって平均した信号(以下、代表信号)を入力された信号とみなしてボトムエンベロープとトップエンベロープおよびゲインを推定する。その後、インプットの代表信号の調波楽器音成分の振幅スペクトルを周波数成分毎に変形するゲインを使って、インプットの各チャネルの振幅スペクトルを変形する。

3. ドラムの音色置換

ドラムの音色置換では、まず打楽器音成分のスペクトログラムを、非負値行列因子分解 [13] 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(以下、基底スペクトログラム)に分解する。非負値行列因子分解は、振幅スペクトログラムを低次元の 2 つの非負値の行列の積、すなわち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。ここでは簡単のため、標準的な非負値行列因子分解を用いたが、non-negative matrix factor deconvolution [14] などの様々な拡張手法も用いることができる。この分解後、図 3 に示すようにインプ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置換するかを、ユーザが決定する。これは、非負値行列因子分解によって得られたインプットとリファレンスの基底インデックスを選択することで実現できる。この選択に従って、基底のペア毎に置換を行い、リファレンスのドラ

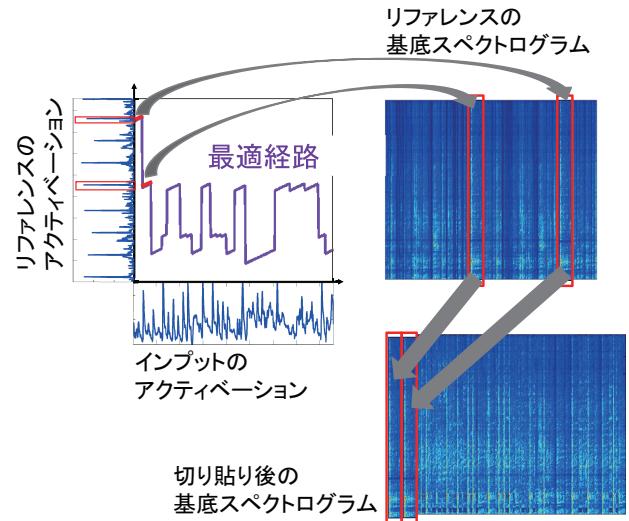


図 4 切り貼り法による基底スペクトログラムペア毎のドラムの音色置換の処理概要 .

ムの音色でインプットのリズムの複素スペクトログラムを得る。以下では、基底のペアに対するドラムの音色置換を議論する。

3.1 イコライジング法

簡易な手法の一つは、イコライザのように選択されたインプ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、インプ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)。周波数インデックスとインプットの基底スペクトログラムの時間インデックスをそれぞれ $\omega \in [0, W - 1]$, $t \in [0, T^{(\text{in})} - 1]$ とする。インプットの基底とその複素スペクトログラムを $H_{\omega}^{(\text{in})}$, $Y_{\omega,t}^{(\text{in})}$ とし、対応するリファレンスの基底を $H_{\omega}^{(\text{ref})}$ とすると、インプットの基底に対応する置換後の複素スペクトログラムは

$$Y_{\omega,t}^{(\text{synth})} = Y_{\omega,t}^{(\text{in})} \frac{H_{\omega}^{(\text{ref})}}{H_{\omega}^{(\text{in})}} \quad (2)$$

と計算できる。

この方法は、インプットの基底スペクトログラムに時不変なゲインを適用するだけでよく高速に計算できる。しかし、基底スペクトログラムのペア間でドラムの音色が大きく異なる場合、例えば片方が高周波帯域、もう片方が低周波帯域にエネルギーを持つ場合には、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、置換後の音質が劣化しドラムの音色が適切に置換されないことがある。

3.2 切り貼り法

3.2.1 最適経路問題としての切り貼り法

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(以下、基底スペクトル)を、インプ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(以下, 切り貼り法)を提案する(図4). 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, インプットのドラムの音量の時間発展に類似させる必要がある. 音量の時間発展を表す特徴量として, 非負値行列因子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. さらに, 以下の3つの要請を課すことにより置換時に生じうる雑音を低減する. 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, 雜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, 楽譜情報なしで各楽器音への音源分離は未だ難しい問題であるため, 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,

$$\mathcal{I}_t(\tau) := \begin{cases} O_{t,\tau} & (t = 1) \\ O_{t,\tau} + \min_{\tau'} \{C_{\tau',\tau} + \mathcal{I}_{t-1}(\tau')\} & (t > 1) \end{cases} \quad (3)$$

$$O_{t,\tau} := \alpha D_{\text{Idiv}}(\tilde{U}_t^{(\text{in})} || \tilde{U}_\tau^{(\text{ref})}) + \beta P_\tau \quad (4)$$

と定義できる. ここで, t, τ はそれぞれインプ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り, 正規化されたアクティベーション $\tilde{U}_t^{(\text{in})}, \tilde{U}_\tau^{(\text{ref})}$ と一般化 I ダイバージェンスは

$$\tilde{U}_t^{(\text{in})} := \frac{U_t^{(\text{in})}}{\max_{t'} \{U_{t'}^{(\text{in})}\}}, \quad \tilde{U}_\tau^{(\text{ref})} := \frac{U_\tau^{(\text{ref})}}{\max_{\tau'} \{U_{\tau'}^{(\text{ref})}\}} \quad (5)$$

$$D_{\text{Idiv}}(x; y) = x \ln \frac{x}{y} - (x - y) \quad (6)$$

と定義される. (4) の第1項は, 2つの正規化されたアクティベーション間の一般化 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,

$$C_{\tau',\tau} := \begin{cases} 1 & (\tau = \tau' + 1) \\ c + \gamma(\tilde{U}_{\tau'}^{(\text{ref})} + \tilde{U}_\tau^{(\text{ref})}) & (\tau \neq \tau' + 1) \end{cases} \quad (7)$$

と定義される. ここで, 定数 c を $c > 1$ とすることにより, $\tau' \neq \tau + 1$ のとき

$$C_{\tau,\tau'} > C_{\tau,\tau+1} \quad (8)$$

となるため, $\tau + 1$ 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なる(要請(i)). $\tau \neq \tau' + 1$ の(7)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, 音源分離が完全でないために, インプ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, 置換後の音が薄くなりやすい. インプ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. 切り貼り法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, 全周波数ビンの振幅の和が定数 ϵ よりも大きければインプットの打楽器音スペクトルに置換する操作をフレーム毎に行う.

3.2.2 切り貼り法の高速化

切り貼り法で最適経路を求める際に, (3) に従って式通りに計算すると, インプットの打楽器音スペクトログラムのフレーム数を $T^{(\text{in})}$, リファレンスの打楽器音スペクトログラムのフレーム数 $T^{(\text{ref})}$ として計算量は $O(T^{(\text{in})}(T^{(\text{ref})})^2)$ である. リファレンスの音響信号が長くなれば $T^{(\text{ref})}$ に関して2乗オーダーで計算時間が増大してしまい, ユーザがシステムを使用する際のストレス要因となりうるため, 計算量の削減が必要である.

計算量の削減には, 遠隔のフレームへの遷移コストが一様であることを利用した高速 Viterbi 法 [15] を用いることができる. $t > 1$ の場合に $C_{\tau,\tau'}$ の定義(7)を(3)に代入すると,

$$\mathcal{I}_t(\tau) = O_{t,\tau} + \min \{1 + \mathcal{I}_{t-1}(\tau - 1), c + \gamma \tilde{U}_\tau^{(\text{ref})} + \min_{\tau' \neq \tau-1} \{\gamma \tilde{U}_{\tau'}^{(\text{ref})} + \mathcal{I}_{t-1}(\tau')\}\} \quad (9)$$

が得られる. 計算量の増大する原因是(9)の右辺の τ' に関する最小値演算が τ に依存するためである. そこで, この最小値演算を τ に依存せずに計算することを考える.

ここで, (8) から

$$c + \gamma(\tilde{U}_\tau^{(\text{ref})} + \tilde{U}_{\tau-1}^{(\text{ref})}) + \mathcal{I}_{t-1}(\tau - 1) > 1 + \mathcal{I}_{t-1}(\tau - 1) \quad (10)$$

が常に成り立つ. そのため, $\mathcal{I}_t(\tau)$ は

$$\begin{aligned}
 \mathcal{I}_t(\tau) &= O_{t,\tau} + \min \{1 + \mathcal{I}_{t-1}(\tau - 1), \\
 &\quad c + \gamma(\tilde{U}_{\tau}^{(\text{ref})} + \tilde{U}_{\tau-1}^{(\text{ref})}) + \mathcal{I}_{t-1}(\tau - 1), \\
 &\quad c + \gamma\tilde{U}_{\tau}^{(\text{ref})} + \min_{\tau' \neq \tau-1} \left\{ \gamma\tilde{U}_{\tau'}^{(\text{ref})} + \mathcal{I}_{t-1}(\tau') \right\} \} \\
 &= O_{t,\tau} + \min \{1 + \mathcal{I}_{t-1}(\tau - 1), \\
 &\quad c + \gamma\tilde{U}_{\tau}^{(\text{ref})} + \min_{\tau'} \left\{ \gamma\tilde{U}_{\tau'}^{(\text{ref})} + \mathcal{I}_{t-1}(\tau') \right\} \} \quad (11)
 \end{aligned}$$

と変形できる。(11) の右辺の τ' に関する最小値演算は τ によらず計算できるため, t 毎に計算すれば十分である。したがって, 計算量は $O(T^{(\text{in})}T^{(\text{ref})})$ に削減され, 高速に最適経路を計算できる。

3.3 ステレオ音響信号への簡易な拡張

上述の処理はモノラルの音楽音響信号に対する処理であるが, インプットとリファレンスがステレオ音響信号の場合にも簡易に拡張できる。ステレオ音響信号の場合は, 調波打楽器音分離を左右チャネルの音響信号に対してそれぞれ行い, 得られた 2 つの打楽器音スペクトログラムに対して, チャネル間で基底のみを共有した非負値行列因子分解を行う。その後, イコライジング法では, 共有した基底を用いて左右チャネルのインプットの打楽器音スペクトログラムを変形する。一方, 切り貼り法では, 時刻 t の左右チャネルのアクティベーションを 2 次元のベクトルとみなし, モノラルの音響信号の場合と同様に計算を行う。すなわち, チャネルインデックス $l = 0, 1$ とインプットとリファレンスの l 番目のチャネルのアクティベーション $U_{l,t}^{(\text{in})}, U_{l,t}^{(\text{ref})}$ を用いて, (4) と (7) を

$$O_{t,\tau} := \alpha \sum_{l=1}^2 D_{\text{Idiv}} \left(\tilde{U}_{l,t}^{(\text{in})}; \tilde{U}_{l,\tau}^{(\text{ref})} \right) + \beta P_{\tau} \quad (12)$$

$$C_{\tau',\tau} := \begin{cases} 1 & (\tau = \tau' + 1) \\ c + \gamma \sum_l (\tilde{U}_{l,\tau'}^{(\text{ref})} + \tilde{U}_{l,\tau}^{(\text{ref})}) & (\tau \neq \tau' + 1) \end{cases} \quad (13)$$

として計算する。ここで, $\tilde{U}_{l,t}^{(\text{in})}, \tilde{U}_{l,t}^{(\text{ref})}$ は

$$\tilde{U}_{l,t}^{(\text{in})} := \frac{U_{l,t}^{(\text{in})}}{\max_{t',l'} \{U_{l',t'}^{(\text{in})}\}}, \quad \tilde{U}_{l,t}^{(\text{ref})} := \frac{U_{l,t}^{(\text{ref})}}{\max_{t',l'} \{U_{l',t'}^{(\text{ref})}\}} \quad (14)$$

である。

4. 主観評価実験

4.1 実験条件

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。インプットとリファレンスとして, RWC ポピュラー音楽データベースとジャンルデータベース [16] 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置換を行った。ドラムの音色置換に関しては, 著者の一人がどのドラ

表 1 各置換例に対するイコライジング法と切り貼り法による評価項目 (1), および周波数特性の置換による評価項目 (2) の平均 MOS 値と標準誤差。

置換例	評価項目 (1)		評価項目 (2)
	イコライジング法	切り貼り法	
例 1	2.4 ± 0.3	3.5 ± 0.3	2.5 ± 0.2
例 2	3.0 ± 0.3	2.8 ± 0.2	2.4 ± 0.2
例 3	2.7 ± 0.3	2.2 ± 0.4	2.3 ± 0.2
例 4	2.5 ± 0.3	2.8 ± 0.3	2.5 ± 0.4
例 5	2.3 ± 0.4	2.1 ± 0.3	2.5 ± 0.3
例 6	1.9 ± 0.3	3.2 ± 0.3	2.7 ± 0.3

ムの音色をどのドラムの音色に置換するかを選択し, 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 $(\alpha, \beta, \gamma, c, \epsilon) = (0.5, 3, 10, 3, 100)$ に設定した。ポピュラー音楽の打楽器音の構成は, バスドラム, スネアドラム, シンバルとその他の計 4 楽器であることが多いため, 非負値行列因子分解の基底数を 4 とし, 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} には, L2 正則化付き L1 損失サポートベクターマシン [17] 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 [16] に収録されている打楽器音とそれ以外の楽器音の 2 クラスで学習した。

主観評価の項目を

- (1) 「ドラムの音色がインプットのドラムの音色に変わったか」
- (2) 「調波成分の全体的な音色がインプットの調波成分の全体的な音色に変わったか」

として, 11 人の被験者により 5 段階 MOS 評価 (1 点を全く置換されていない, 5 点を完全に置換されていたとする) を行った。被験者は, インプットとリファレンス, 置換後の音楽音響信号とそれらの調波楽器音と打楽器音をそれぞれ何度も聞き直せた。

4.2 結果と考察

項目 (1) に対し, 標準誤差付きの平均スコアはイコライジング法が 2.5 ± 0.1 , 切り貼り法が 2.8 ± 0.1 であった。イコライジング法と切り貼り法共に MOS 値が 1 に比べて有意に高く (t 検定, 有意水準 1 %), ドラムの音色置換が期待通り動作していることを確認した。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, 置換例 6 例中 2 例に対する MOS 値が有意に高かった (t 検定, 有意水準 5 %)。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。この項目に対しても, MOS 値が 1 に比べて有意に高く (t 検定, 有意水準 1 %), 調波音楽器に対する周波数特性の置換が期待通り動作していることを確認した。

置換例を聴取すると、音源分離精度が高い場合には、ドラムの音色置換について切り貼り法の方がイコライジング法よりもリファレンスの音源に近い印象を受けた。これは、イコライジング法では時不变なゲインを用いるため残響やスペクトルの時間変動を扱えないが、切り貼り法はリファレンスの打楽器音スペクトルを使用するためこれらの時間的な変動も扱えるためであると考えられる。一方で、調波打楽器音分離や非負値行列因子分解による分離精度が低い場合には、切り貼り法よりもイコライジング法の方がリファレンスの音源に近い印象を受けた。分離精度が低い場合には、打楽器音スペクトログラムに歌声が含まれてしまう場合が多く、切り貼り法で得られた打楽器音スペクトログラムにリファレンスの歌声の一部が混入することがあった。この混入は、より精度の高い音源分離手法を用いることで抑制できると考えられる。

置換例は <http://hil.t.u-tokyo.ac.jp/~nakamura/demo/TimbreReplacement.html> で試聴可能である。

5. 結論

本論文では、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置換するシステムを提案した。ボトムエンベロープとトップエンベロープを用いたスペクトルの変形法 [12] を利用し、調波打楽器音成分の周波数特性の置換手法を提案した。また、ドラムの音色置換については、インプットの打楽器音成分のスペクトルを時不变なゲインを用いて変形するイコライジング法と、リファレンスの各ドラム楽器のスペクトルを切り貼りする切り貼り法 2 種類の方法を提案した。主観評価実験により、調波打楽器の周波数特性とドラムの音色の置換について提案システムが適切に動作していることを確認した。

今後は、提案システムの定量評価を通してどの程度の精度で置換が可能なのかや、どの打楽器により適しているのかを調べる必要がある。また、ユーザが自由に置換度合いを調節可能なユーザインターフェースの開発や、声質や音高の変換も課題である。

6. 謝辞

本研究は JST CREST 「OngaCREST プロジェクト」と科研費 26700020 の支援を受けた。

参考文献

- [1] Swamy, M. N. S. and Thyagarajan, K. S.: Digital bandpass and bandstop filters with variable center frequency and bandwidth, *Proc. IEEE*, Vol. 64, No. 11, pp. 1632–1634 (1976).
- [2] Erfani, S. and Peikari, B.: Variable cut-off digital ladder filters, *Int. J. Electron.*, Vol. 45, No. 5, pp. 535–549 (1978).
- [3] Tan, E. C.: Variable lowpass wave-digital filters, *Electron. Lett.*, Vol. 18, pp. 324–326 (1982).
- [4] Regalia, P. A. and Mitra, S. K.: Tunable digital frequency response equalization filters, *IEEE Trans. Acoust., Speech, and Language Process.*, Vol. 35, No. 1, pp. 118–120 (1987).
- [5] Orfanidis, S. J.: Digital parametric equalizer design with prescribed nyquist-frequency gain, *J. Audio Eng. Soc.*, Vol. 45, No. 6, pp. 444–455 (1997).
- [6] Itoyama, K., Goto, M., Komatani, K., Ogata, T. and Okuno, H. G.: Integration and Adaptation of Harmonic and Inharmonic Models for Separating Polyphonic Musical Signals, *Proc. Int. Conf. Acoust. Speech Signal Process.*, Vol. 1, pp. I-57–I-60 (2007).
- [7] Yoshii, K., Goto, M., Komatani, K., Ogata, T. and Okuno, H. G.: Drumix: An Audio player with real-time drum-part rearrangement functions for active music listening, *Trans. Info. Process. Soc. Japan*, Vol. 48, No. 3, pp. 1229–1239 (2007).
- [8] Ono, N., Miyamoto, K., Kameoka, H. and Sagayama, S.: A real-time equalizer of harmonic and percussive components in music signals, pp. 139–144 (2008).
- [9] Yasuraoka, N., Abe, T., Itoyama, K., Takahashi, T., Ogata, T. and Okuno, H. G.: Changing timbre and phrase in existing musical performances as you like: manipulations of single part using harmonic and inharmonic models, *Proc. ACM Multimedia*, pp. 203–212 (2009).
- [10] 藤原 弘将, 後藤真孝 : 混合音中の歌声スペクトル包絡推定に基づく歌声の声質変換手法, 情処研報, Vol. 2010-MUS-86, No. 7, pp. 1–10 (2010).
- [11] Tachibana, H., Kameoka, H., Ono, N. and Sagayama, S.: Comparative Evaluation of Multiple Harmonic/Percussive Sound Separation Techniques based on Anisotropic Smoothness of Spectrogram, *Proc. Int. Conf. Acoust. Speech Signal Process.*, pp. 465–468 (2012).
- [12] 亀岡 弘和, 後藤 真孝, 嵐嶽山茂樹 : スペクトル制御エンベロープによる混合音中の周期および非周期成分の選択的イコライザ, 情処研報, Vol. 2006-MUS-66, No. 13, pp. 77–84 (2006).
- [13] Seung, D. and Lee, L.: Algorithms for non-negative matrix factorization, *Proc. Adv. Neural Inf. Process. Syst.*, Vol. 13, pp. 556–562 (2001).
- [14] Smaragdis, P.: Non-negative matrix factor deconvolution; extraction of multiple sound sources from monophonic inputs, *Independent Component Analysis and Blind Signal Separation*, Springer, pp. 494–499 (2004).
- [15] Nakamura, T., Nakamura, E. and Sagayama, S.: Acoustic Score Following to Musical Performances with Errors and Arbitrary Repeats and Skips for Automatic Accompaniment, *Proc. Sound and Music Computing*, No. 166 (2013).
- [16] Goto, M.: Development of the RWC Music Database, *Proc. Int. Congress Acoust.*, pp. 1–553–556 (2004).
- [17] Fan, R., Chang, K., Hsieh, C., Wang, X. and Lin, C.: LIBLINEAR: A Library for Large Linear Classification, *J. Mach. Learn. Res.*, Vol. 9, pp. 1871–1874 (2008).