社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2008－MUS－76 (26)
2008／8／8

音高による音色変化を考慮した楽器音の音高・音長操作手法

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 ‡産業技術総合研究所

本稿では，ある音高・音長をもつ楽器音を音色の特徴を歪ませることなく任意の音高・音長へ操作する手法について述べる．我々は音色の聴感上の差に関する音響心理学的知見に基づき，楽器音のスペクトログラム上での音色特徴量として (i) 倍音ピーク間の相対強度，(ii) 非調波成分の分布，(iii) 時間方向エンベロープの3つを定義する．これら音色特徴量の分析には糸山らの調波・非調波統合モデルを用いる．音高操作時には，音高に対する特徴量 (i)，(ii) の分布を三次関数でモデル化し，所望の音高における特徴量の値を予測することで音高依存性を考慮する．音長操作時には，特徴量 (iii) の時間的変化がゆるやかな区間のみを伸縮させることで，楽器音の立ち上がりと立ち下がりを保存する．32種類の楽器に対して音高操作を試みたところ，音高依存性を考慮しない場合と比べて合成音と実際の楽器音とのMFCC距離が32.31％減少した．

A Method for Manipulating Pitch and Duration of Musical Instrument Sounds Dealing with Pitch-dependency of Timbre

TAKEHIRO ABE†, KATSUTOSHI ITOYAMA†, KAZUYOSHI YOSHII‡,
KAZUNORI KOMATANI†, TETSUYA OGATA† and HIROSHI G. OKUNO†

†Dept. of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University
‡ National Institute of Advanced Industrial Science and Technology (AIST)

This paper presents a manipulation method that can generate musical instrument sounds with arbitrary pitches and durations from a given musical instrument sound without distorting timbral characteristics. Based on the psychoacoustical knowledge on auditory effects of timbre, we define timbral features on the spectrogram of a musical instrument sound as (i) relative amplitudes of harmonic components, (ii) distribution of inharmonic components, and (iii) temporal envelopes of harmonic components. We use Itoyama's integrated model to analyze timbral features. For pitch manipulation, we take into account the pitch-dependency of timbre by using a cubic polynomial that approximates the distribution of features (i) and (ii) over pitches and predicting the values of each feature. To manipulate duration, we preserve feature (iii) in the attack and decay durations of a seed by expanding or shrinking only the steady duration. Experimental results showed the effectiveness of our method; the MFCC distance between synthesized sounds and real sounds of 32 instruments was reduced by 32.31 %.

1. はじめに

従来のイコライザとは音響信号全体の周波数特性を変化させるものであったが，近年，音楽音響信号に特化し，楽器単位での音量の操作や音色の置き換えが可能な楽器音イコライザと呼ばれる新技術が開発されてきている[1)~3)]．多くのオーディオプレーヤに実装されているイコライザは周波数帯域の操作によって楽曲の音響を変化させるが，楽器音イコライザが提供する楽器単位の操作によって音楽鑑賞の幅はさらに広がると期待される．吉井らのDrumix[2)]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の操作と音色の置き換えを実現している．一方，糸山らの楽器音イコライザ[3)]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が，Drumixで実現されていた音色の置き換えは扱われていない．

我々の最終目標は，任意の楽器パートをユーザーの好みの楽器音に置き換えるイコライザの開発である．これが実現できれば，例えば，ロック風の楽曲を構成するギター，ベース，キーボードなどの楽器音を，ヴァイオリン，ウッドベース，ピアノなどの楽器音で置き換えることで，ユーザーはその楽曲をクラシック風にアレンジして楽しむことができる．また，好きなギタリストが演奏した楽曲からギター音を抽出し，別の楽曲のギターパートをそのギター音で置き換えることで，ユーザーはそのギタリストにさまざまなフレーズを演

奏させることもできる．

上記イコライザを実現するための技術的課題として以下の 2 つが挙げられる．

(1) 混合音中からユーザーが置き換えに用いたい楽器音を抽出するため，**混合音から任意の楽器音を分離**する．

(2) 任意のフレーズを演奏するため，分離された楽器音をもとにして**任意の音高・音長を持つ楽器音を合成**する．

前者は，糸山らを含め多くの研究者によって継続的に取り組まれており，その成果が報告されている[4),5)]．一方，分離された音の応用についてはほとんど議論されてこなかった．そこで，我々は分離された複数の単音を入力とした楽器音の合成の課題に取り組む．

代表的な楽器音合成方式にはフェーズボコーダおよび正弦波重畳モデルを用いたものが有名である．フェーズボコーダは歴史の長い楽器音合成方式であり，多くの派生的な手法が報告されている[6),7)]．正弦波重畳モデルもまた，音声や楽器音を良く表現するモデルとして有名であり，音源分離にも応用され，様々なモデルパラメータ推定手法が報告されている[8)~10)]．これら手法では音の分析あるいは合成に焦点が当てられており，分析されたパラメータの操作については言及されていない．また，明示的に音色の特徴がパラメータとして定義されておらず，音色の特徴を考慮した操作が困難であり，音高依存性などといった音色の性質を分析するまでには至っていない．

我々は音色の聴感上の差に関する音響心理学的知見に基づいて定義した音色特徴量を分析し，音色の特徴を回避した楽器音の音高・音長操作手法を報告する．音色特徴量の分析には糸山らの調波・非調波統合モデルを用いる．音高操作時には，音高特徴量の分布を三次関数でモデル化し，所望の音高における特徴量の値を予測することで音高依存性を考慮する．また，音長操作時には，音のエネルギーの変化を表す時間エンベロープの時間的変化がゆるやかな区間のみを伸縮させることで，楽器音の立ち上がりと立ち下がりを保存する．

2. 音色の特徴を考慮した音高・音長操作

本研究の目的は，ある楽器個体の実際の音（*seed* と呼ぶ）がいくつか得られているとき，それらをもとにして同個体の任意の音高・音長をもつ音を合成することである．このとき重要な点は，音色の音響的特徴が歪まないようにすることである☆．例えば，ある音高をもつ楽器音から他の音高をもつ音を合成したとき，これら 2 つの音は同一個体から発せられる音であると感じることができなければならない．

音色の音響的特徴を歪めないで楽器音を合成するには，音色の特徴量を数学的に定義し，これを分析する必要がある．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，(i) 高周波数領域での倍音ピークの有無，(ii) 発音時に発生する非調波成分，(iii) 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[11)]．我々はこれらの要因を以下の 3 つの特徴量にそれぞれ対応付ける．

☆ 本稿における音色の音響的特徴の歪みとは，実楽器から発せられる楽器音の音色との差異と定義する．

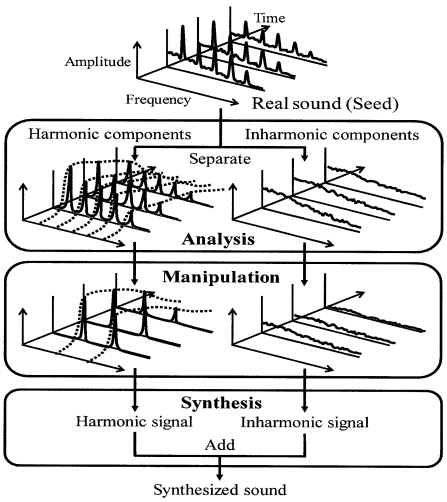


図 1 提案手法の概要

(i) **倍音ピーク間の相対強度**

(ii) **非調波成分の分布**

(iii) **時間方向の振幅エンベロープ**

図 1 に我々の提案する分析・合成手法の概要を示す．特徴量 (i) および (iii) は調波成分に関するもの，特徴量 (ii) は非調波成分に関するものであるので，まず，*seed* の調波成分と非調波成分を分離し，各特徴量を分析する．次に，音色を歪ませることなく音高・音長操作を行う．このとき，音色特徴量の値を変えずに音高・音長だけを変化させることは適切でないことに注意する．最後に，調波成分・非調波成分を別々に再合成し，足し合わせる．

2.1 楽器音の分析

音色の特徴量を分析するためには，調波成分と非調波成分とを明示的に分けて取り扱い，それぞれにおける特徴量を定義する必要がある．この問題を解決するため，我々は糸山らが提案した調波・非調波統合モデル[3)]を利用して楽器音を表現することを試みる．すなわち，*seed* のスペクトログラム $M(f,r)$ に対し，調波成分に対応するパラメトリックモデル $M_H(f,r)$ と非調波成分に対応するノンパラメトリックモデル $M_I(f,r)$ を w_H および w_I で重み付けした混合モデルをフィッティングさせる．

$$M(f,r) = w_H M_H(f,r) + w_I M_I(f,r) \qquad (1)$$

ここで，f と r はそれぞれ周波数と時間を表す．また，$\sum_{f,r} M_I(f,r) = 1$ という制約が与えられているので，重み w_I は非調波成分のエネルギーと考えることができ，$w_I M_I(f,r)$ は非調波成分のスペクトログラムそのものを表す．一方，$M_H(f,r)$ は，各倍音 n に対するパラメトリックモデルの重み付き混合モデルとして表現される．

$$M_H(f,r) = \sum_n F_n(f,r) E_n(r) \qquad (2)$$

ここで，$F_n(f,r)$ および $E_n(r)$ は，**図 2** と**図 3** に示すような周波数エンベロープおよび時間エンベロープの

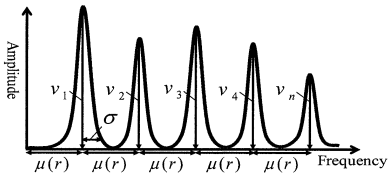


図 2 周波数エンベロープ

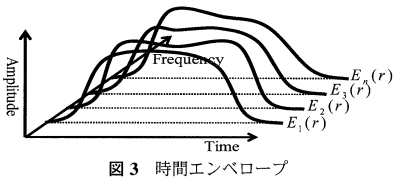


図 3 時間エンベロープ

モデルとなっている．

$F_n(f,r)$ は混合正規分布として表現される．

$$F_n(f,r) = v_n \mathcal{N}(f - n\mu(r), \sigma^2) \tag{3}$$

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，σ は倍音ピークの周波数方向への広がりを表す．$\mu(r)$ は *seed* の音高の時間方向の軌跡である．また，v_n は $\sum_n v_n = 1$ を満たす重みである．

一方，$E_n(r)$ は $\sum_r E_n(r) = 1$ を満たすノンパラメトリックな関数である．糸山らは $E_n(r)$ に対しても $F_n(f,r)$ と同様のパラメトリックモデルを構成していたが，より詳細な分析を可能とするため本稿ではこのような方法をとった．

この統合モデルにおいて，音色の特徴量 (i)，(ii) および (iii) は，それぞれ v_n，$w_I M_I(f,r)$ および $E_n(r)$ に対応する．これらの求め方は 3.1 節にて述べる．

2.2 音高操作

音高を操作するには，音高軌跡 $\mu(r)$ に所望の倍率を乗算すればよいが，このとき音色特徴量の値を変化させずにそのまま利用することはできない．なぜなら，音色は音高依存性をもつことが知られており[12]，音高の操作が大きくなるにつれて音色は歪みは増加するからである．**図 6** に示すように，音高を $\mu(r)$ から $\mu'(r)$ に変化させる場合には，相対強度を v_n から v'_n へと適切に変化させる必要がある．

この問題を解決するため，我々は北原らの提案した音高依存性を考慮した楽器音識別手法[13] に着目する．彼らは音高に対する音響的特徴量の分布を三次関数を用いて近似し，音高依存性を除去したあとの特徴量分布を学習することで，楽器音識別率が向上したと報告している．本研究では，音高よりも奏法に依存すると考えられる特徴量 (iii) を除き，音高に対する特徴量 (i), (ii) の分布を三次関数（**音高依存特徴関数**と呼ぶ）で近似する．具体的には，以下の 2 つのパラメータに着目する．

(1) 各倍音の倍音ピーク間の相対強度 v_n
(2) 調波成分のエネルギーに対する非調波成分のエネルギーの比 w_H/w_I

異なった音高をもつ複数の *seed* が与えられれば，それらの音色特徴量を分析し，最小二乗法によって音高依存特徴関数を求めることができる．得られた音高依存特徴関数を用いれば，所望の音高における音色特徴量を予測することができる．例として，**図 4** にトランペットの第 1 次倍音，第 4 次倍音，第 10 次倍音の相対強度，および調波成分と非調波成分のエネルギー比の音高特徴依存関数を示す．

2.3 音長操作

音長を操作するには，時間エンベロープ $E_n(r)$ を所望の音長になるように伸縮させる方法は適切ではない．なぜなら，同一楽器個体では音長にかかわらず，発音の立ち上がりと立ち下がり，および音高の変動周期は類似することが知られており，音長の操作が大きくなるにつれて歪みは増加するからである．特に楽器音の立ち上がりや立ち下がりはエネルギーが大きく変化する部分で音色の印象への関わりが深い．

この問題を解決するため，我々は時間エンベロープにおける立ち上がりと立ち下がり部分を保存する手法および音高軌跡の時間的変動を再現する手法を提案する．まず，特徴量 (iii) において，エネルギーの急峻な立ち上がり終了時をオンセット r_{on}，エネルギーの急峻な立ち下がり開始時をオフセット r_{off} として定義する．音長を操作するには，**図 5** に示すようにオンセット・オフセット間のみを伸縮させればよい．オンセット以前及びオフセット以降の音高軌跡は操作前のものを用いる．

3. 処理系の実装

本章では，2 章で述べた手法の具体的な実装について説明する．

3.1 楽器音の分析

ここで問題となるのは，2.1 節で示した統合モデルにおける未知パラメータ $w_H, w_I, F_n(f,r), E_n(r), v_n, \mu(r), \sigma, M_I(f,r)$ を推定することである．そのため，糸山らは統合モデルと *seed* のスペクトログラムとの Kullback-Leibler Divergence (KLD) を減少させるようにパラメータを反復更新する手法を提案している．この反復過程は Expectation-Maximization (EM) アルゴリズムと解釈でき，効率的にパラメータを推定することができる．

3.2 音高操作

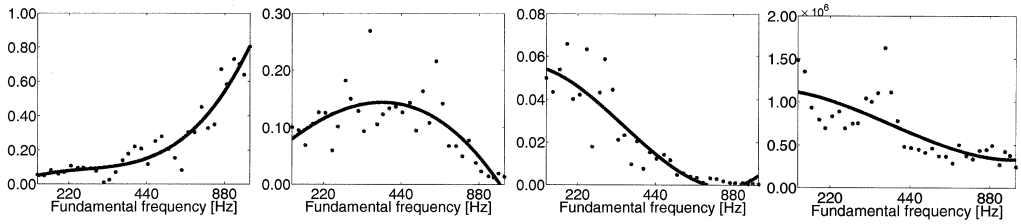
音高操作を行うには，周波数エンベロープを構成する音高軌跡 $\mu(r)$ に対して，実数 α （音高を低くする場合: $0 \leq \alpha < 1$, 音高を高くする場合: $1 < \alpha$) を乗算する．ここで，$\mu'(r)$ は所望する操作後の音高とすると以下が成り立つ．

$$\mu'(r) = \alpha\mu(r) \tag{4}$$

例えば，α を 2 とすれば，*seed* の 1 オクターブ上の音高の楽器音が合成できる．操作後の楽器音の倍音ピーク間の相対強度 v_n は，音高依存特徴関数から予測される各倍音ごとの倍音ピーク間の相対強度を制約条件 $\sum_n v_n = 1$ より正規化することで得られる．また，非調波成分のエネルギー w_I は，調波成分のエネルギー w_H を音高特徴依存関数から予測される調波成分に対する非調波成分のエネルギーの比 w_H/w_I で割ることで得られる．

3.3 音長操作

音長操作を行うには，オンセット・オフセット間の



a: 第 1 次倍音の相対強度　b: 第 4 次倍音の相対強度　c: 第 10 次倍音の相対強度　d: 調波と非調波のエネルギー比

図 4　トランペットの音高依存特徴関数．点と実線はそれぞれ，音高ごとに分析された音色の特徴量と，導出された音高依存特徴関数である．

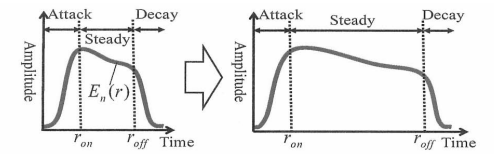


図 5　時間エンベロープの操作

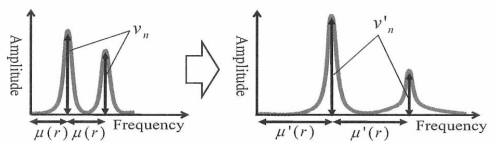


図 6　周波数エンベロープの操作

時間エンベロープ $E_n(r)$ と音高軌跡 $\mu(r)$ を操作する．

本稿におけるオンセットとは，楽器音の時間方向の振幅が十分に大きくなってから，振幅の変動が一定になる瞬間である．オフセットとは，時間方向の振幅が十分な大きさを持っており，振幅の変動が一定の状態が得られなくなる瞬間である．具体的には，オンセット r_{on} とオフセット r_{off} を以下の 2 条件を満たす r の区間の始点と終点とする．

$$\left|\frac{dE_n(r)}{dr}\right| \leq \epsilon,\ E_n(r) \geq Th \tag{5}$$

ここで Th は楽器音の時間方向の振幅の十分な大きさを示す閾値である．持続系の楽器はこれでよいが，打弦楽器や撥弦楽器といった減衰楽器のオンセットとオフセットは，ほぼ同時時刻となり，オンセット・オフセット間を伸縮させることができない．よって，減衰楽器音の場合には，時間エンベロープの終端を減衰楽器音のオフセットとみなし，オンセット以降の時間エンベロープを伸縮の対象とする．

3.4 楽器音の合成

調波モデルから調波信号 $s_H(t)$ を，非調波モデル $s_I(t)$ から非調波信号を合成し，以下のように重ね合わせることで最終的な楽器音 $s(t)$ を合成する．

$$s(t) = s_H(t) + s_I(t) \tag{6}$$

3.4.1 調波信号の合成

調波信号 $s_H(t)$ を合成するには，正弦波重畳モデルを用いる．

$$s_H(t) = \sum_n A_n(t)\exp[j\phi_n(t)] \tag{7}$$

$$\phi_n(t) = \phi_n(0) + \int_0^t \omega_n(\tau)d\tau \tag{8}$$

ここで，$A_n(t)$，$\phi_n(t)$ と $\omega_n(t)$ はそれぞれ n 番目の正弦波の振幅，瞬時位相と瞬時周波数である．瞬時周波数は音高軌跡 $\mu(r)$ のスプライン補間によって得られる．振幅は，次式のように調波・非調波統合モデルのパラメータから導出できる．

$$A_n(t) = \frac{w_H \hat{E}_n(t) v'_n}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} w(\tau)d\tau \tag{9}$$

ここで $w(t)$ は *seed* のスペクトログラムを計算する際に使用された分析窓である．また，$E_n(r)$ のスプライン補間によって得られる，サンプル単位で表現された時間エンベロープである．

3.4.2 非調波信号の合成

非調波信号 $s_I(t)$ を合成するには，オーバーラップ加算法を用いる．このとき，非調波成分のエネルギーを乗算した非調波モデルをスペクトログラムとみなして信号に変換する．位相は *seed* のものをそのまま利用する．

4. 評価実験

本手法の有効性を評価するために行った評価実験について報告する．

4.1 評価条件

合成した楽器音の品質を評価するため，合成音と実楽器音との距離を以下に示すメル周波数ケプストラム係数 (MFCC) 距離尺度を用いて算出した．

$$D_M = \sum_{d,r}(M_{syn} - M_{real})^2/R \tag{10}$$

ここで，M_i は MFCC を表し，添え字 *syn* と *real* は合成音と実楽器音に対応する．R は楽器音の時間長である．この距離が小さいほど，合成音が実楽器音に近いことを示す．MFCC 距離尺度は聴感上の尺度としてしばしば用いられる．周波数軸は対数スケールで表わされるため，調波成分に含まれる各倍音ピークの差だけでなく，それらよりずっと小さなエネルギーしかない非調波成分の差も含めて評価できる．MFCC の次元は 12 次元とした．

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[14]．このデータベースに含まれる楽器音は，単独発音を半音ごとに収録（サンプリング周波数: 44.1 kHz，16 ビット量子化，モノラル）されている．このデータベースから 32 種類の楽器ごとに 3 個体を選択し，フォルテで通常の奏法で

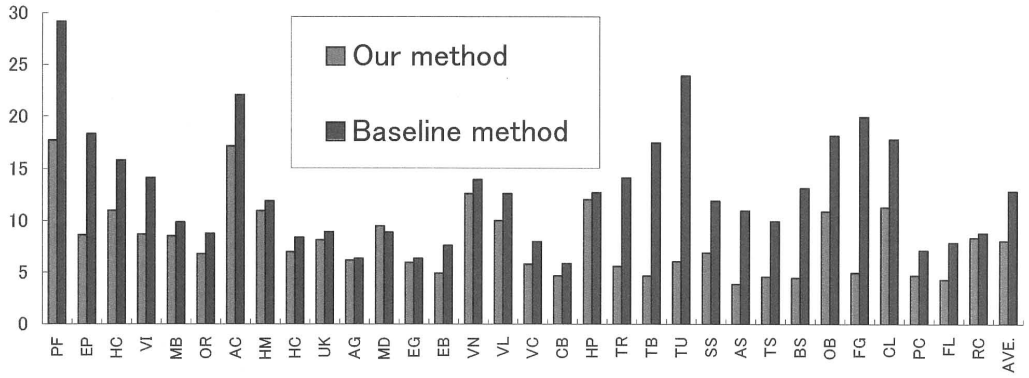


図 7 本手法とベースライン手法における MFCC 距離.

表 1 本実験で用いた楽器音の内訳

楽器名 (楽器記号)	ピアノ (PF), エレクトリックピアノ (EP), ハープシコード (HC), ビブラフォン (VI), マリンバ (MB), オルガン (OR), アコーディオン (AC), ハーモニカ (HM), クラシックギター (HC), ウクレレ (UK), アコースティックギター (AG), マンドリン (MD), エレキギター (EG), エレキベース (EB), バイオリン (VN), ビオラ (VL), チェロ (VC), コントラバス (CB), ハープ (HP), トランペット (TR), トロンボーン (TB), チューバ (TU), ソプラノサックス (SS), アルトサックス (AS), テナーサックス (TS), バリトンサックス (BS), オーボエ (OB), ファゴット (FG), クラリネット (CL), ピッコロ (PC), フルート (FL), リコーダー (RC)
楽器個体数	3 個体
強さ	フォルテのみ.
奏法	通常の奏法のみ
データ数	PF: 264, EP: 206, HC: 178, VI: 104, MB: 145, OR: 178, AC: 141, HM: 101, HC: 111, UK: 71, AG: 111, MD: 123, EG: 111, EB: 88, VN: 138, VL: 126, VC: 134, CB: 111, HP: 241, TR: 103, TB: 96, TU: 90, SS: 99, AS: 99, TS: 98, BS: 98, OB: 96, FG: 120, CL: 120, PC: 98, FL: 111, RC: 75

演奏されたものを実験に用いた⋆. 実験に用いた楽器音の内訳を **表 1** に示す.

各楽器の個体ごとに 10-fold cross validation を行い, 合成音と実楽器音との距離の平均を求めた. まず, ある個体の全楽器音データを, 90%を学習用, 10%を評価用としてランダムに分割した. 次に, 学習用データを用いて音高依存特徴関数を学習した. 学習用データの各楽器音を *seed* として, 評価用データの各楽器音と同じ音高を持つ楽器音を合成する. 最後に, 合成音と実際の楽器音との距離を計算した.

ところで, 合成品質を距離に適切に反映するには, 実楽器音の時間エンベロープ $E_n(r)$ や音高軌跡 $\mu(r)$ における演奏のゆらぎの影響を排除しなければならない. そのため, 本実験において合成音を生成する際には, $E_n(r)$ および $\mu(r)$ は評価用データのものを用い, その他のパラメータを学習用データから推定することにした. すなわち, 現状では音高操作に関してのみ距離尺度を用いて評価できることを示す. 本実験では, 音高依存特徴関数を用いない場合をベースライン手法とし, 音高依存性を考慮した提案手法と比較した.

4.2 実験結果・考察

図 7 に本手法とベースライン手法によって合成した楽器音の実楽器音に対する距離を示す. これらの値は楽器個体ごとに得られた距離を平均して示してある. 32 種類の楽器のうち, マンドリンを除いての全ての楽器において MFCC 距離が減少している. 全楽器で平均すると, 本手法を用いたことによる MFCC 距離の減少率 (改善率) は 32.31 %となり, 音色の音高依存性を考慮した本手法の有効性が示されている.

距離の改善率が高かった例として, ファゴット (MFCC 距離: 75.17 %) における音高操作の変化量に対応する距離を**図 8** a) に示す. ースライン手法では, 音高操作の変化量が増えるにつれて, 距離が増加するのを確認できる. 一方, 本手法では音高操作の変化量に対しての距離の増加がみられない. 以下に述べる距離の改善の大きかった楽器においても, 音高操作の変化量にかかわらず, 距離の増加が抑えられることを確認した.

一方, アコーディオン, マリンバ, マンドリンを例に挙げるように, いくつかの楽器では距離の改善がさほどされなかった. これらの楽器の改善について以下で考察する.

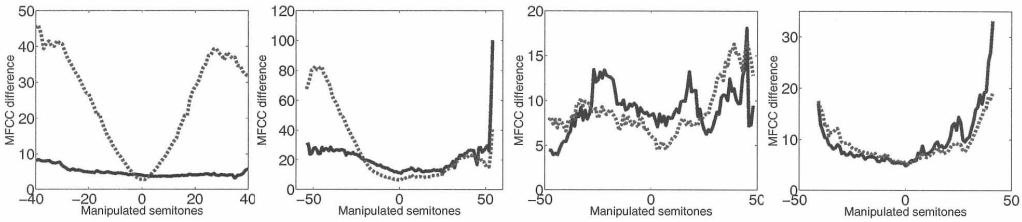
(1) 音色の音高依存性が小さい場合の改善

アコーディオン (MFCC 距離: 22.08 %) における音高操作の変化量に対応する距離を図 8 b) に示す. 高を下げる操作では, 本手法により距離が改善されている. 一方, 音高を上げる操作では, 本手法による改善がみられない. ベースライン手法においても音高をあげる操作では距離の増加がないところから, アコーディオンのような楽器では高い音高では, 音色の音高依存性をさほど持たないためと考えられる.

(2) 音色の音高依存性の高次元への対応

マリンバ (MFCC 距離: 13.99 %) における音高操作の

⋆ ビブラート, スタッカートなどを除く一般的な奏法を指す. ただし, RWC 研究用音楽データベースのバイオリン音においては, ビブラート奏法が通常奏法とタグづけされているため,「ビブラート無」とタグづけされている音を用いる



a) ファゴット　b) アコーディオン　c) マリンバ　d) マンドリン

図 8　本手法とベースライン手法における半音操作ごとの距離．実線と破線はそれぞれ本手法での距離とベースライン手法での距離を示す．

変化量に対応する距離を図 8 c) に示す．手法において音高操作の変化量に対しての距離が複雑に変化しており，マリンバの音色の音高依存性を学習できていないことがわかる．マリンバの音は打楽器音の要素を含み，さらに発音機構がピアノのように各音高で独立である．本稿では，音色の特徴量を 3 次関数で近似することで，音高依存特徴関数を導出したが，マリンバのように音高ごとに複雑に音色が変わる楽器では，3 次関数で近似するのは不十分であると考えられる．

(3)　非調波成分における強い音高依存性への対応

マリンバ (MFCC 距離: -6.64 %) における音高操作の変化量に対応する距離を図 8 d) に示す．の MFCC 距離の改悪は，合成音の非調波成分の分布が実楽器音のものと異なることが原因と考えられる．本手法では，音高特徴量関数によって非調波成分のエネルギーの音高依存性を扱ってはいるが，分布自体の音高依存性の考慮までには至っていない．他の撥弦楽器においても，MFCC 距離の改善はさほどみられなかった．撥弦楽器は発音時に高周波数領域に多くの非調波成分を含むことが知られており[11]，非調波成分の音高依存性の影響が他の楽器より大きい．

5. おわりに

本稿では，ある音高・音長をもつ楽器音，これを扱った楽器音の音高・音長操作手法を報告した．ここで定義した音色の特徴量は (i) 倍音ピーク間の相対強度，(iii) 時間方向の振幅エンベロープ，(ii) 非調波成分の 3 つであり，Grey が報告した音色の聴感上の知覚の差に対応するスペクトルの要因の知見を参考にしたものである．音高に対する音色の変化と，立ち上がりと立ち下がりの振幅，及び音高の変動の周期の同一楽器内での類似といった音色の性質を音色特徴量によって扱った．本手法と音色の音高依存性を扱っていないベースライン手法との比較実験を行ったの結果，合成音と実楽器音との MFCC 距離が，32.31 %減少し，本手法の有効性を確認した．

今後は，本研究の最終目標に向けて，実際に楽曲から分離された楽器音に対して本手法を適用した場合の問題について対処していく．このとき，分離音には様々なノイズが含まれているので，出来る限りノイズの含まれない *seed* を選択することが重要になる．加えて，本手法の音長操作の評価と非調波成分の分布の音高依存性を扱えるよう本手法の拡張を行う．

謝辞　本研究の一部は，科学研究費補助金（基盤研究 (S)），グローバル COE プログラム「知識社会基盤構築のための情報学拠点形成」，科学技術振興機構 CrestMuse プロジェクトによる支援を受けた．

参考文献

1) Gillet, O. and Richard, G.: Extraction and Remixing of Drum Tracks from Polyphonic Music Signals, *WASPAA*, pp.315–318 (2005).
2) Yoshii, K., Goto, M. and Okuno, H.G.: Drumix: An Audio Player with Real-time Drum-part Rearrangement Functions for Active Music Listening, *IPSJ Journal*, Vol.48, No.3, pp. 1229–1239 (2007).
3) 糸山克寿，後藤真孝，駒谷和範，尾形哲也，奥乃　博：楽譜情報を援用した多重奏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，情報処理学会論文誌，Vol.49, No.3, pp.1465–1479 (2008).
4) Helen, M. and Virtanen, T.: Separation of Drums from Polyphonic Music Using Non-negative Matrix Factorization and Support Vector Machine, *Proc. EUSIPCO* (2005).
5) Fitzgerald, D., Cranitch, M. and Coyle, E.: Sound Source Separation using Shifted Non-negative Tensor Factorization, *ICASSP*, Vol.V, pp.653–656 (2006).
6) Laroche, J. and Dolson, M.: Improved phase vocoder timescale modification of audio, *IEEE Trans. Speech and Audio Processing*, Vol.7, No.3, pp.323–332 (1999).
7) Robel, A.: A New Approach to Transient Processing in the Phase Vocoder, *DAFx*, pp.1–6 (2003).
8) McAulay, R. and Quatieri, T.: Speech Analysis/Synthesis based on a Sinusoidal Representation, *IEEE Trans. Acoust., Speech, and Signal Processing*, pp.744–754 (1986).
9) Jinachitra, P.: Constrained EM Estimates for Harmonic Source Separation, *ICASSP*, Vol.VI, pp.609–612 (2003).
10) 亀岡弘和，小野順貴，嵯峨山茂樹：正弦波重畳モデルのパラメータ最適化アルゴリズムの導出，電子情報通信学会技術研究報告，Vol.106, No.432, pp.49–54 (2006).
11) Grey, J.M.: Multidimensional perceptual scaling of musical timbres, *J. Acoust Soc. Am.*, Vol.61, No.5, pp.1270–1277 (1977).
12) Marozeau, J., Cheveigne, A., McAdams, S. and Winsberg, S.: The dependency of timbre on fundamental frequency, *J. Acoust Soc. Am.*, Vol.114, No.5, pp.2946–2957 (2003).
13) 北原鉄朗，後藤真孝，奥乃　博：音高による音色変化に着目した楽器音の音源同定：F0 依存多次元正規分布に基づく識別手法，情報処理学会論文誌，Vol.44, No.10, pp.2448–2458 (2003).
14) Goto, M., Hashiguchi, H., Nishimura, T. and Oka, R.: RWC Music Database: Music Genre Database and Musical Instrument Sound Database, *ISMIR*, pp.229–230 (2003).