

楽器音イコライザによる 音色の類似度に基づく楽曲検索システム

糸山 克寿¹ 後藤 真孝² 駒谷 和範¹
尾形 哲也¹ 奥乃 博¹

本稿では、音楽音響信号中の楽器パートごとの音量を操作することを可能とするオーディオプレーヤー、楽器音イコライザについて述べる。Query-by-Exampleに基づく音楽情報検索システムを利用する際、多様な検索結果を得るためには検索のクエリとなる楽曲を複数用意する必要があるが、楽器音イコライザを用いて楽曲中の楽器パートの音量を操作しそれをクエリとすることで、一つの楽曲から多様な検索結果を得ることができる。音量操作のためには音楽音響信号はあらかじめ楽器パートごとに分離されている必要がある。分離には、調波・非調波統合モデルと呼ぶ、調波音と非調波音の両方を表現可能な音モデルを用いる。分離におけるパラメータ推定の精度を向上させるために、多数の楽器音を用いてモデルパラメータの事前分布を学習させる。

A music information retrieval system based on timbre similarity using the instrument equalizer

KATSUTOSHI ITOYAMA,¹ MASATAKA GOTO,² KAZUNORI KOMATANI,¹
TETSUYA OGATA¹ and HIROSHI G. OKUNO¹

This paper describes a music remixing interface, called *Instrument Equalizer*, that allows users to control the volume of each instrument part within existing audio recordings in real time. Although query-by-example retrieval systems need a user to prepare favorite examples (songs) in general, our interface gives a user to generate examples from existing ones by cutting or boosting some instrument/vocal parts, resulting in a variety of retrieved results. To change the volume, all instrument parts are separated from the input sound mixture using the corresponding standard MIDI file. For the separation, we used an integrated tone (timbre) model consisting of harmonic and in-harmonic models. To improve the accuracy of parameter estimation, we train probabilistic distributions of model parameters by using various sounds.

1. はじめに

Query-by-Example (QBE) による音楽情報検索¹⁾⁴⁾とは、ユーザが指定した楽曲をクエリ (example) として与え、楽曲を相互の類似性に基づいてランキングする検索手法である。QBEは有用な検索手法であるが、ユーザは事前にクエリとなる楽曲を準備する必要がある。また、検索結果に不満がある場合、よりよい検索結果を得るためにはユーザはクエリとなる他の楽曲を探す必要がある。たとえば、検索された楽曲のボーカルやドラムスの音量が大きすぎるとユーザが感じた場合、クエリとした楽曲に雰囲気や音色などの特徴が類似しておりかつボーカルやドラムスの音量がより小さい楽

曲を探す必要がある。しかしながら、個人が所有する楽曲中からこのような条件を満たす楽曲を見つけ出すのはしばしば困難である。

我々は音楽リミックスインタフェースを用いてQBE検索におけるクエリを作成する手法を提案する。このインタフェースは、音楽音響信号 (混合音) 中の各楽器パートごとの音量を操作する機能を持つ。より好みに近い検索結果を得るため、オリジナルの楽曲とは異なるミックスバランスのもとで合成された新たなクエリをこのインタフェースを用いることで容易に生成することができる。たとえば、ボーカルやドラムスの音量を下げたクエリを生成することで、前述の問題は解決される。ステレオの音楽音響信号のリミックスに関する研究はすでに報告されている⁵⁾ものの、調波的な楽器音のみを対象としていた。これに対して我々の目標は、調波的な音と非調波的な音との両方をリミックスの対象とすることである。

本稿では、楽器音イコライザと呼ぶ音楽リミックスインタフェースについて述べる。楽器音イコライザは、

¹ 京都大学 大学院情報学研究所 知能情報学専攻
Department of Intelligence Science and Technology, Graduate
School of Informatics, Kyoto University

² 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology
(AIST)

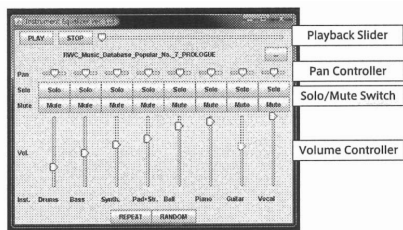


図1 楽器音イコライザのスクリーンショット. 上部に一般的なオーディオプレーヤが持つ再生スライダを, その下に各楽器パートごとのパン, ソロ・ミュート, ボリュームを操作するためのスライダ類を持つ.

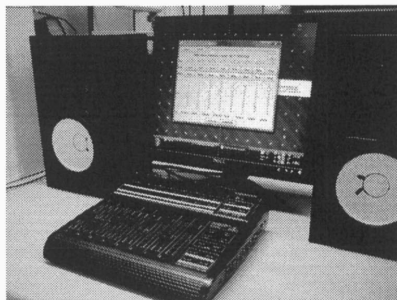


図2 フィジカルコントローラを併用した例. コントローラでの操作とディスプレイでの操作は連動しており, 一方で音量などを操作するとその変更は他方に伝達される.

各楽器パートに対応したスライダを持ち, 各楽器パートの音量をリアルタイムに操作することを可能にする. 楽器音イコライザを用いて QBE 検索を行う際は, 特定の検索システムを前提としない. 現在の楽器音イコライザの実装では, 音楽音響信号に対応する標準 MIDI ファイル (SMF) を用いて混合音の分離を行う. 多くの楽曲 (特にクラシック) に限っては Web などから SMF を入手することができ, SMF があれば MIDI 音源を用いて楽器パートの音量を操作することは容易に実現できる. しかしながら, ユーザの検索への動機付けとしては, MIDI 音源によって生成した信号をクエリとするよりは CD の信号をクエリとする方がユーザは検索へとより強く動機付けされると考えられることから, 我々は音楽音響信号 (混合音) の音量を操作し, クエリを生成することが重要だと考える.

2. 楽器音イコライザ

楽器音イコライザは, ユーザが多重奏音楽音響信号をリミックスすることを可能にするオーディオプレーヤである. 図1にインタフェースのスクリーンショットを, 図2にシステムの設置例を示す. 本システムは, 以下の2つの機能を持つ.

- (1) 音量操作. 各楽器パートごとに音量を増減させる機能. 音量操作は, 楽曲の再生中でもリアルタイムに出力される音響信号に反映される.
- (2) フィジカルコントローラとの操作連動. ディスプレイ上のインタフェースをマウスなどで操作するだけでなく, 図2のようなフィジカルコントローラを併用することを想定している. ディスプレイでの操作とコントローラでの操作とは連動している. マウスでの操作よりも, より直接的に各パートの音量を操作できると考えられる.

多重奏音楽音響信号をリミックスするためには音響信号を各楽器パートごとに分離する必要がある. このために, 我々は調波・非調波等号モデル⁶⁾を用いる. 分離処理中のモデルパラメータ推定において, 入力音響信号中の楽器音の奏法・音色個性差が推定精度に影響することを避けるため, 複数の楽器個性・奏法で演奏した楽器音を学習データとしたモデルパラメータの事前分布を用いる. 本分離手法は, 何らかの手法⁷⁻⁹⁾を

用いて音楽音響信号と同期がとられた標準 MIDI ファイル (SMF) を用いる. 分離処理は, SMF に基づいて各単音のモデルを生成し, モデルと混合音中の各単音とのスペクトログラム上での距離を最小化するようにモデルパラメータを反復推定することで実現される.

3. 実装

本章では, 楽器音イコライザ, 類似楽曲検索システム, 音源分離処理の実装について述べる.

3.1 楽器音イコライザ

本節では, 楽器音イコライザによる音量操作部分の内部構造について述べる. 楽器音イコライザによるリミックス処理は実時間で行われる. 各楽器パートの音楽音響信号は第3.3節で述べる手法で事前に分離されていることを仮定している.

$x_k(c, t)$ と $y_k(c, t)$ とをそれぞれ時刻 t , チャネル c における楽器 k の分離信号と音量とする. $y_k(c, t)$ は以下の条件を満たす.

$$\forall k, c, t: 0 \leq y_k(c, t) \leq 1,$$

また, $y_k(c, t)$ は以下のように計算する.

$$y_k(c, t) = (\text{楽器 } k \text{ の音量}) \cdot (\text{チャネル } c \text{ のパン}).$$

第2章で述べた2つの機能は, 以下のように実現する.

- (1) 音量操作. 楽器音イコライザの出力信号 $x(c, t)$ は以下のように計算する.

$$x(c, t) = \sum_k y_k(c, t) \cdot x_k(c, t).$$

$y_k(c, t)$ の値は, GUI から実時間で読み込まれる.

- (2) フィジカルコントローラとの操作連動. GUI とフィジカルコントローラとは MIDI で通信を行う. ユーザが一方に操作を行うと, その情報を表す MIDI メッセージが他方に送信される. コントローラにはモータが内蔵されており, GUI からの MIDI メッセージによってフェーダを GUI のスライダに対応する位置まで移動させる.

3.2 類似楽曲検索

本節では, 類似楽曲検索の実装について述べる.

3.2.1 特徴量の抽出

音楽における雰囲気を表現する特徴量を, ムード検出に関する先行研究¹⁰⁾を参考に表1のように設計した.

これらの特徴量は、音響信号をモノラルに変換し、さらに短時間フーリエ変換することで得られるパワースペクトル $X(t, f)$ からフレーム毎 (毎秒 100 フレーム) に抽出を行う。ここでは、 t はフレーム番号、 f は周波数ピンの番号を表す。以下で、これらの特徴量の詳細について述べる。

音量に関する特徴 各フレームの音量 $S_1(t)$ 、および各フレームの i 番目のサブバンドの音量 $S_2(i, t)$ をそれぞれ以下のように定める。

$$S_1(t) = \sum_{f=1}^{F_N} X(t, f), S_2(i, t) = \sum_{f=F_L(i)}^{F_H(i)} X(t, f)$$

ここで、 F_N はパワースペクトルの周波数ピン数、 $F_L(i)$ と $F_H(i)$ はそれぞれ i 番目のサブバンドの周波数の下限と上限のピン番号である。このように全体の音量に加え、サブバンド毎の音量の特徴量を扱うことで、より詳細な音量の特徴を表現することができる。本研究では、サブバンドへの分割に以下のオクターブフィルタバンクを使用する。

$$\left[1, \frac{F_N}{2^{n-1}}\right], \left[\frac{F_N}{2^{n-1}}, \frac{F_N}{2^{n-2}}\right], \dots, \left[\frac{F_N}{2}, F_N\right]$$

ここで、 n はバンク数を表す。本稿での実験では、バンク数を 7 とした。

音色に関する特徴 音色を表現する特徴量としては、メル周波数ケプストラム係数 (Mel Frequency Cepstrum Coefficient; MFCC) を始めスペクトルの性質を表現する特徴量が使用されている^{11),12)}。本研究では、音楽のムードの検出に有効であると報告されているスペクトルの形状およびスペクトルのコントラストを表現する特徴量^{13),14)}に基づき、音色特徴量を設計する。これらのスペクトルの形状の特徴量により音楽における明るさや激しさを、スペクトルのコントラストの特徴量により音楽における調性を表現する。

まず、スペクトルの形状を表現する特徴量である、各フレームのスペクトル重心 $S_3(t)$ 、スペクトルバンド幅 $S_4(t)$ 、スペクトルロールオフ $S_5(t)$ 、スペクトルフラックス $S_6(t)$ をそれぞれ以下のように設計する。

$$S_3(t) = \frac{\sum_{f=1}^{F_N} X(t, f)f}{\sum_{f=1}^{F_N} X(t, f)},$$

$$S_4(t) = \frac{\sum_{f=1}^{F_N} X(t, f)(f - S_3(t))^2}{\sum_{f=1}^{F_N} X(t, f)},$$

$$S_5(t) = \frac{\sum_{f=1}^{F_N} X(t, f)}{\sum_{f=1}^{F_N} X(t, f)} = 0.95 \sum_{f=1}^{F_N} X(t, f),$$

$$S_6(t) = \sum_{f=1}^{F_N} (\log X(t, f) - \log X(t-1, f))^2$$

次に、スペクトルのコントラストを表現する特徴量を以下のように設計する。 t 番目の時刻のフレ

表 1 音楽の雰囲気表現する特徴量

音量に関する特徴	
1	全体の音量
2-8	サブバンド*毎の音量
音色に関する特徴	
9	スペクトル重心
10	スペクトル幅
11	スペクトルロールオフ
12	スペクトルフラックス
13-19	サブバンド*毎のピーク値
20-26	サブバンド*毎のバレー値
27-33	サブバンド*毎のピーク値とバレー値の差

*サブバンドにはバンク数 7 のオクターブフィルタバンクを使用。

ム、 i 番目のサブバンド内のスペクトログラムの、各周波数ピンの係数のベクトルを

$$(X(i, t, 1), X(i, t, 2), \dots, X(i, t, F_N(i)))$$

とおき、このベクトルを降順に整列したものを改めて

$$(X'(i, t, 1), X'(i, t, 2), \dots, X'(i, t, F_N(i)))$$

とする。ただし、 $F_N(i)$ は i 番目のサブバンドの周波数ピン数で、

$$F_N(i) = F_H(i) - F_L(i)$$

である。ここで、

$$X'(i, t, 1) > X'(i, t, 2) > \dots > X'(i, t, F_N(i))$$

である。このとき、第 t フレームにおける i 番目のサブバンドのスペクトルピーク $S_7(i, t)$ 、スペクトルバレー $S_8(i, t)$ 、スペクトルコントラスト $S_9(i, t)$ をそれぞれ以下のように設計する。

$$S_7(i, t) = \log \left(\frac{1}{\beta F_N(i)} \sum_{f=1}^{\beta F_N(i)} X'(i, t, f) \right),$$

$$S_8(i, t) = \log \left(\frac{1}{\beta F_N(i)} \sum_{f=(1-\beta)F_N(i)}^{F_N(i)} X'(i, t, f) \right),$$

$$S_9(i, t) = S_7(i, t) - S_8(i, t)$$

ここで β は安定したピークとバレーの値を抽出するためのパラメタであり、本稿での実験では 0.2 とした。

これらの特徴を抽出した後、主成分分析によって特徴量の次元を圧縮する。本稿での実験では、累積寄与率が 0.95 になるように次元圧縮をしたところ、特徴量は 9 次元に圧縮された。

3.2.2 類似度の計算

楽曲ごとの全体的な雰囲気表現のために、各楽曲内では雰囲気は一定であると仮定し、楽曲毎に特徴量を混合正規分布でモデル化する。本稿の実験では、混合数を 4 としてモデル化した。こうして得られた各楽曲に対応する混合正規分布同士の距離を、楽曲間の類似度として用いる。距離には、Earth Movers Distance (EMD)¹⁵⁾を用いる。

3.3 音源分離

本研究では、音源分離問題を入力混合音のパワースペクトル $X(c, t, f)$ を楽曲中の各単音に対応するパワースペクトルへ分解することと定義する。 c, t, f はそれぞれ

チャンネル, 時刻, 周波数を表す. このような分解を行うために, 各単音のワースベクトルを近似的に表現するモデルとして, 調波・非調波統合モデル⁶⁾ $J(k, l, c, t, f)$ を用いる. ただし, k, l はそれぞれ楽器の番号, 単音の番号である. このモデルは, 調波的な音のワースベクトルを表現する調波構造モデルと非調波的な音のワースベクトルを表現する非調波構造モデルとの和で定義される. このモデルが含むパラメータを, 表 2 に示す.

調波・非調波統合モデルを用いて音源分離を行うため, 入力ワースベクトルを (k, l) 番目の単音へと分配するための分配関数 $\Delta_J(k, l, c, t, f)$ を導入する. この関数は以下の条件を満たす.

$$\forall c, t, f: \sum_{k, l} \Delta_J(k, l, c, t, f) = 1$$

この関数を用いると, 分離された (k, l) 番目の単音のワースベクトルは以下のように表される.

$$X_J(k, l, c, t, f) = \Delta_J(k, l, c, t, f) X(c, t, f)$$

ここで, どのように $\Delta_J(k, l, c, t, f)$ を定めると最もよい分離が行えるかを考える. よい分離が行われる状態を, $J(k, l, c, t, f)$ が入力ワースベクトルを正確に表現している状態と捉え, 分離の善し悪しをこれらの間の Kullback-Leibler Divergence (KLD) で表す.

$$\sum_c \iint X(c, t, f) \log \frac{X(c, t, f)}{\sum_{k, l} J(k, l, c, t, f)} dt df$$

さらに, Jensen の不等式より, 以下が導出できる.

$$\begin{aligned} & \sum_c \iint X(c, t, f) \log \frac{X(c, t, f)}{\sum_{k, l} J(k, l, c, t, f)} dt df \\ & \leq \sum_{k, l, c} \iint X_J(k, l, c, t, f) \log \frac{X_J(k, l, c, t, f)}{J(k, l, c, t, f)} dt df \end{aligned}$$

このことから, 上記の不等式の右辺を最小化するような $\Delta_J(k, l, c, t, f)$ および $J(k, l, c, t, f)$ を求めることで, 音源分離処理を実現できる. この分離処理は, 6) にて詳細が述べられている.

3.3.1 事前分布の利用

同じ種類の楽器であっても, 楽器の個体が異なればそれを演奏することで得られる楽器音の音色には差が現れる. また, 人間が楽器を演奏する場合, 全く同じ楽器演奏動作を繰り返すことは不可能であるので, 演奏動作の違いによって楽器音の音色も異なるものになる. しかしながら, 我々が以前に開発した音源分離手法⁶⁾ は MIDI 音源で生成したテンプレート音を分離に用いていたため, 音調や音高, 音量といった情報だけから機械的に生成されたテンプレート音では上記の単音毎の音色の違いを表現することはできなかった.

このような楽器毎の音色のぶれを適切に扱い, かつパラメータ推定処理のテンプレート音への依存性をなくすため, あらかじめ統合モデルパラメータの事前分布を楽器音データベースを用いて学習しておき, この

表 2 統合モデルのパラメータ.

記号	意味
$w_J(k, l)$	モデル全体の音量
$r(k, l, c)$	各チャンネルの音量比
$w_H(k, l), w_I(k, l)$	調波構造モデルと非調波構造モデルとの音量比
$v_H(k, l, m, n)$	第 n 次倍音の第 m 番目の時刻での相対強度
$\tau(k, l)$	発音時刻
$\phi_H(k, l)$	調波構造を構成するガウス分布関数の時間方向への広がり
$\omega_H(k, l, t)$	F0 時系列
$\sigma_H(k, l)$	調波構造を構成するガウス分布関数の周波数方向への広がり
$v_I(k, l, m, n)$	非調波構造モデルの第 m 番目の時刻, 第 n 番目の周波数成分の相対強度
ϕ_I	非調波構造モデルを構成するガウス分布関数の時間方向への広がり
$\omega_I(n)$	非調波構造モデルの第 n 番目の周波数成分の中心周波数
$\sigma_I(n, f)$	非調波構造モデルの第 n 番目の周波数成分の周波数方向への広がり

事前分布をテンプレート音の代わりに用いてパラメータ推定を行う手法を提案する.

事前分布としてモデル化する対象のパラメータとして, $(w_H(k, l), w_I(k, l), v_H(k, l, m, n), v_I(k, l, m, n))$ を用いる. これらを Dirichlet 分布として学習する. すなわち, 以下の 3 つの事前分布を学習する.

- (1) $p(w_H(k, l), w_I(k, l))$,
- (2) $p(v_H(k, l, 0, 1), \dots, v_H(k, l, M_H - 1, N_H))$,
- (3) $p(v_I(k, l, 0, 1), \dots, v_I(k, l, M_I - 1, N_I))$.

本稿の実験では, RWC 音楽データベース: 楽器音¹⁶⁾ を学習データとした.

パラメータ推定処理において, 最小化する目的関数は以下のようになる.

$$\begin{aligned} & \sum_c \iint X_J(k, l, c, t, f) \log \frac{X_J(k, l, c, t, f)}{J(k, l, c, t, f)} dt df \\ & - (\alpha_{w_H}(k) - 1) \log w_H(k, l) \\ & - (\alpha_{w_I}(k) - 1) \log w_I(k, l) \\ & - \sum_{m, n} (\alpha_{v_H}(k, m, n) - 1) \log v_H(k, l, m, n) \\ & - \sum_{m, n} (\alpha_{v_I}(k, m, n) - 1) \log v_I(k, l, m, n). \end{aligned}$$

4. 評価実験

音源分離処理において事前分布を利用したことの分離性能への影響, および類似楽曲検索において楽器音イコライザを用いることでどのように検索結果が変化するかを調査するため, 評価実験を行った.

4.1 音源分離における実験

本実験の目的は, 音源分離処理において事前分布を用いることが分離性能へどのように影響するかを調査することである. 実験内容は, SMF に基づいて楽器音データベース¹⁶⁾ の単音を混合して生成した音楽音響信号を分離するというものである. SMF は, RWC 音楽

データベース：クラシックと同データベース：ジャズ¹⁷⁾から10曲ずつ、合計20曲を使用した。表3に用いた楽曲を、表4に用いた楽器を示す。それぞれの楽器に対して、3つの楽器個体、2つの奏法を用いたため、1曲あたり6つの音響信号を生成した。

事前分布の学習は、音響信号の生成に用いなかった楽器音を用いて行った。データがスパースになることを避けるため、モデルパラメータが以下のように分解できると仮定して分解されたそれぞれのパラメータに対して事前分布を学習した。

$$v_H(k, l, m, n) = v_H(k, l, m)v_H(k, l, n),$$

$$v_I(k, l, m, n) = v_I(k, l, m)v_I(k, l, n).$$

本実験においては、以下の2条件を比較した。

- (1) 事前分布を用いる場合 (提案手法)
- (2) テンプレート音を用いる場合 (以前の手法¹⁸⁾)

各条件で分離されたパワースペクトルに対して、ミックス前の楽器音のパワースペクトルとのSNRをもって分離性能を評価した。テンプレート音と入力混合音との間には必ず音響的な差ができるようにテンプレート音はMIDI音源で生成した。

4.1.1 実験結果

楽曲毎に平均したSNRを図3に示す。事前分布を用いてパラメータ推定を行うことで、SNRは平均で4.89から8.48dB向上した。この結果から、テンプレート音と入力混合音との間に音響的な差がある場合でも事前分布を使うことでパラメータ推定がよりロバストに行われたと考えられる。また、楽器と奏法の組み合わせ毎に平均したSNRを図4に示す。多くの楽器と奏法に関しては事前分布を用いることでSNRが向上したが、EBPN, CBNO, CBNV, TRNVの楽器と奏法の組み合わせに関してはSNRが低下した。これは、以下のような原因と改善策が考えられる。

- (1) 事前分布が楽器音の音色を正確に表現できていなかった。本研究では事前分布に単峰的なDirichlet分布を用いたが、パラメータの分布にピークが複数存在する場合などは単独のDirichlet分布では十分にその分布を近似できない。これは、事前分布を混合分布化するなどによって改善される。
- (2) 低周波数帯域での周波数解像度が不十分だった。これは、基本周波数が低い音では調波構造のピーク同士が近づくため、調波構造モデルで調波構造を正確に近似できなくなってしまう。これに対しては、短時間フーリエ変換の窓長を増やしたり、可変な周波数解像度を扱えるウェーブレット変換などを用いることで改善される。

4.2 類似楽曲検索における実験

本実験の目的は、楽曲の楽器ミックスバランスを操作し合成した信号をクエリとすることで、検索結果がその操作に応じて適切に変化するかどうかを確認することである。検索のクエリとする楽曲はRWC音楽データベース：ポピュラー No. 7。検索対象とする楽曲の集合はRWC音楽データベース：音楽ジャンルの100曲を、それぞれ用いた。

表3 実験に用いた楽曲。

Data Symbol	Instruments
Classical No. 2	VN, VL, VC, CB, TR, OB, FG, FL
Classical No. 3	VN, VL, VC, CB, TR, OB, FG, CL, FL
Classical No. 12	VN, VL, VC, CB, FL
Classical No. 16	VN, VL, VC, CL
Classical No. 17	VN, VL, VC, CL
Classical No. 22	PF
Classical No. 30	PF
Classical No. 34	PF
Classical No. 39	PF, VN
Classical No. 40	PF, VN
Jazz No. 1	PF
Jazz No. 5	PF
Jazz No. 8	EG
Jazz No. 9	EG
Jazz No. 16	PF, EB
Jazz No. 17	PF, EB
Jazz No. 23	PF, EB, TS
Jazz No. 24	PF, EB, TS
Jazz No. 27	PF, AG, EB, AS, TS, BS
Jazz No. 28	PF, AG, EB, AS, TS, BS

表4 実験に用いた楽器。

Inst. name (Abbr.)	Perf. style	
	set A (Abbr.)	set B (Abbr.)
Pianoforte (PF)	Normal (NO)	Staccato (ST)
Electric Guitar (EG)	Legato/Pick (LP)	Vibrato/Pick (VP)
Electric Bass (EB)	Normal/Pick (PN)	Normal/Two-finger (TN)
Violin (VN)	Normal (NO)	Non-vibrato (NV)
Viola (VL)	Normal (NO)	Non-vibrato (NV)
Cello (VC)	Normal (NO)	Non-vibrato (NV)
Contrabass (CB)	Normal (NO)	Non-vibrato (NV)
Trumpet (TR)	Normal (NO)	Vibrato (VI)
Alto Sax (AS)	Normal (NO)	Vibrato (VI)
Tenor Sax (TS)	Normal (NO)	Vibrato (VI)
Baritone Sax (BS)	Normal (NO)	Vibrato (VI)
Oboe (OB)	Normal (NO)	Vibrato (VI)
Fagotto (FG)	Normal (NO)	Vibrato (VI)
Clarinet (CL)	Normal (NO)	Vibrato (VI)
Flute (FL)	Normal (NO)	Vibrato (VI)

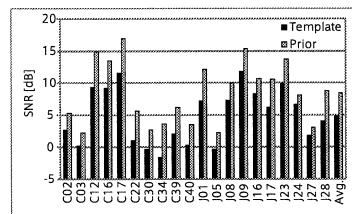


図3 楽曲ごとに平均したSNR。

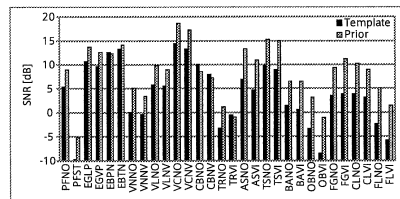


図4 楽器・奏法の組み合わせごとに平均したSNR。

表5 (a) 音量操作を行わなかった場合の検索結果

Data Symbol	Genre	Distance
Genre No. 27	ソウル/R & B	0.0157
Genre No. 24	ファンク	0.0158
Genre No. 13	ラップ/ヒップホップ	0.0161
Genre No. 11	ヘビーメタル	0.0166
Genre No. 5	バラード	0.0167
Genre No. 1	ポップス	0.0171

表6 (b) ドラムスパートの音量を増幅させた場合の検索結果

Data Symbol	Genre	Distance
Genre No. 13	ラップ/ヒップホップ	0.0143
Genre No. 18	ハウス	0.0175
Genre No. 21	テクノ	0.0179
Genre No. 16	ハウス	0.0201
Genre No. 24	ファンク	0.0201
Genre No. 20	テクノ	0.0206

表7 (c) ピアノ・ボーカルを増幅、ドラムス・ギターを減衰させた場合の検索結果

Data Symbol	Genre	Distance
Genre No. 34	フュージョン	0.0195
Genre No. 43	レゲエ	0.0199
Genre No. 5	バラード	0.0201
Genre No. 4	バラード	0.0204
Genre No. 69	フォーク	0.0205
Genre No. 66	ブルース	0.0206

4.2.1 実験結果

表5, 6, 7にそれぞれ(a)楽器音イコライザによる音量操作を行わなかった場合, (b)ドラムスパートの音量を1.3倍に増幅させた場合, (c)ボーカルとピアノパートの音量を1.3倍に増幅, ドラムスとギターパートの音量を0.6倍に減衰させた場合の距離(EMD)が小さかった6楽曲を示す。

(a)では, 歌声が重要な役割を果たす音楽(ソウルミュージックやバラード), ギターが重要な役割を果たす音楽(ヘビーメタル)など, 様々なジャンルの楽曲が検索されている。これに対して(b)ではドラムパートの音量を増幅させた結果, ハウスやテクノといったリズムパートが重要となる音楽が多く検索されている。また(c)ではボーカルとピアノを増幅した結果, バラード, フォーク, ブルースといった比較的「静かな」楽曲が多く検索されている。これらの結果より, 楽器パートの音量操作により検索結果もそれに依って変化し, かつ妥当性の高いジャンルの楽曲が検索されたといえる。

5. おわりに

本稿では, 楽器音イコライザを用いてQBEによる音楽情報検索におけるクエリを生成する手法と, 楽器音イコライザに必要な音源分離のための調波・非調波統合モデルのパラメータ推定におけるパラメータ事前分布を用いた推定手法について述べた。楽器音イコライザを用いた検索に関しては, 被験者実験を含むさらなる実験と検証を行う予定である。音源分離で用いた事前分布は分離だけでなく, 楽器音認識などに応用することもできるため, 音源分離と楽器音認識を同時に行うフレームワークを検討している。また, 楽器音イコ

ライザを検索以外の新たなアプリケーションへの応用を模索することも今後の課題である。

謝辞 本研究の一部は, グローバル COE プログラム, 科学研究費補助金(基盤研究(S), 特定領域「情報爆発IT基盤」), 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) Rauber, A., Pampalk, E. and Merkl, D.: Using Psycho-acoustic Models and Self-organizing Maps to Create a Hierarchical Structuring of Music by Sound Similarity, *Proc. ISMIR*, pp.71–80 (2002).
- 2) Yang, C.: The MACSIS Acoustic Indexing Framework for Music Retrieval: An Experimental Study, *Proc. ISMIR*, pp. 53–62 (2002).
- 3) Thoshkanna, B. and Ramakrishnan, K.R.: Projekt Quebex: A Query by Example System for Audio Retrieval, *Proc. ICME*, pp.265–268 (2005).
- 4) Kitahara, T., Goto, M., Komatani, K., Ogata, T. and Okuno, H.G.: Musical Instrument Recognizer “Instrogram” and Its Application to Music Retrieval Based on Instrumentation Similarity, *ISM*, pp.265–274 (2006).
- 5) Woodruff, J., Pardo, B. and Dannenberg, R.: Remixing Stereo Music with Score-informed Source Separation, *Proc. ISMIR*, pp.314–319 (2006).
- 6) 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, *情報論*, Vol.49, No.3, pp.1465–1479 (2008).
- 7) Cano, P., Loscos, A. and Bonada, J.: Score-performance Matching using HMMs, *ICMC*, pp.441–444 (1999).
- 8) Adams, N., Marquez, D., and Wakefield, G.: Iterative Deepening for Melody Alignment and Retrieval, *Proc. ISMIR*, pp.199–206 (2005).
- 9) Cont, A.: Realtime Audio to Score Alignment for Polyphonic Music Instruments using Sparse Non-negative Constraints and Hierarchical HMMs, *Proc. ICASSP*, pp.641–644 (2006).
- 10) 西山正紘, 北原鉄朗, 駒谷和範, 尾形哲也, 奥乃 博: マルチメディアコンテンツにおける音楽と映像の調和度計算モデル, *情報研報* 69-6 (2007).
- 11) Tzanetakis, G. and Cook, P.: Musical Genre Classification of Audio Signals, *IEEE Trans. Speech and Audio Process.*, Vol.10, No.5, pp.293–302 (2002).
- 12) Essid, S., Richard, G. and David, B.: Instrument Recognition in Polyphonic Music Based on Automatic Taxonomies, *IEEE Trans. Audio, Speech and Lang. Process.*, Vol.14, No.1, pp.68–80 (2006).
- 13) Lu, L., Liu, D. and Zhang, H.J.: Some Issues on Detecting Emotions in Music, *IEEE Trans. Audio, Speech and Lang. Process.*, Vol.14, No.1, pp.5–18 (2006).
- 14) Jiang, D.N., Lu, L., Zhang, H.J., Tao, J.H. and Cai, L.H.: Music Type Classification by Spectral Contrast Features, *Proc. ICME*, pp.113–116 (2002).
- 15) Rubner, Y., Tomasi, C. and Guibas, L.J.: A Metric for Distributions with Applications to Image Databases, *Proc. ICCV*, pp.59–66 (1998).
- 16) Goto, M., Hashiguchi, H., Nishimura, T. and Oka, R.: RWC Music Database: Music Genre Database and Musical Instrument Sound Database, *Proc. ISMIR*, pp.229–230 (2003).
- 17) Goto, M., Hashiguchi, H., Nishimura, T. and Oka, R.: RWC Music Database: Popular, Classical, and Jazz Music Databases, *Proc. ISMIR*, pp.287–288 (2002).
- 18) Itoyama, K., Goto, M., Komatani, K., Ogata, T. and Okuno, H.: Integration and Adaptation of Harmonic and In-harmonic Models for Separating Polyphonic Musical Signals, *Proc. ICASSP*, pp.57–60 (2006).