

多重奏楽曲の楽器音量バランス変化による 音楽ジャンルシフト

糸山克寿^{†1} 後藤真孝^{†2} 駒谷和範^{†1}
尾形哲也^{†1} 奥乃博^{†1}

本報告では、楽曲の楽器パート音量操作によってユーザがクエリをカスタマイズすることが可能な類似楽曲検索手法を提案する。楽曲の雰囲気やジャンルは楽曲を構成する楽器およびその音量バランスと強く関係する、という仮説に基づく。楽曲の音響信号を楽譜に基づいて楽器パートへと分離し、その分離信号の音量を操作することで楽曲の音響的特徴を変化させる。楽曲の音響特徴はガウス混合分布で表現され、楽曲間の類似性を分布間の Earth Movers Distance で定義する。実験により、歌声、ギター、ドラムスパートの音量を操作した際にジャンルシフトが起こることを示す。

Musical Genre Shift of Polyphonic Musical Pieces by Changing Instrument Volume

KATSUTOSHI ITOYAMA,^{†1} MASATAKA GOTO,^{†2}
KAZUNORI KOMATANI,^{†1} TETSUYA OGATA^{†1}
and HIROSHI G. OKUNO^{†1}

This report presents a novel Query-by-Example (QBE) approach in Music Information Retrieval, which allows a user to customize query examples by directly modifying the volume of different instrument parts. The underlying hypothesis is that the musical genre *shifts* (changes) in relation to the volume balance of different instruments. Our QBE system first separates the musical audio signal into all instrument parts with the help of its musical score, and then lets a user remix those parts to change acoustic features that represent musical mood of the piece. The distribution of those features is modeled by the Gaussian Mixture Model for each musical piece, and the Earth Movers Distance between mixtures of different pieces is used as the degree of their mood similarity. Experimental results showed that the shift was actually caused by the volume change of vocal, guitar, and drums.

^{†1} 京都大学 大学院情報学研究科 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)

1. はじめに

Query-by-Example (QBE) による音楽情報検索 (類似楽曲検索)¹⁾⁻⁵⁾ 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。類似楽曲検索は有効な検索手法であるが、多様な検索結果を得るためには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。このような条件を満たす楽曲を見つけ出すのは堂々巡りであり、直接的な検索手法ではない。

我々は、既存楽曲のリミックス (楽器パートの音量操作) によって QBE 検索におけるクエリを作成する手法⁶⁾ により上記の堂々巡りを解消する。ユーザは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生成し、検索を行う。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。このようなリミックスを行うためには楽曲を楽器パートごとに分離する必要がある。我々は既存の音楽音響信号とその楽曲の楽譜を入力して楽器パートごとの音響信号を出力する音源分離手法⁷⁾ を用いる。

本リミックスに基づく検索が機能するために必要な仮説は、検索結果の楽曲のジャンルはクエリの楽曲を構成する楽器およびその音量比率に影響を受ける、すなわち、ユーザは楽曲の楽器音量バランスの操作によってクエリ楽曲のジャンルを変化させることができる、ということである^{*1}。この仮説は先行研究⁶⁾ でも示唆されている。仮説が成り立つ範囲を明らかにすることで、楽器パートの音量変化の大きさと検索結果の変化の関係を明らかにすることが本報告の目的である。

楽曲の雰囲気や類似性に基づいて検索を行う類似楽曲検索システムを実装し、上記の仮説を検証するための2つの実験を行う。この類似楽曲検索システムは、楽曲の雰囲気を音響特徴量の混合正規分布で表現し、楽曲同士の類似性を特徴量分布間の Earth Movers Distance (EMD)⁸⁾ で定量化する。様々なジャンルの楽曲から構成されるデータベースに対してリミッ

*1 本報告では、クラシック、ジャズ、ロックといった、楽器編成やその音量で分類することが可能な粒度のジャンルを対象とする。ワルツやヒップホップなど、特定のリズムパターンや歌唱スタイル、楽器演奏法によって分類可能なジャンルも存在するが、本報告ではこのようなジャンルは扱わない。

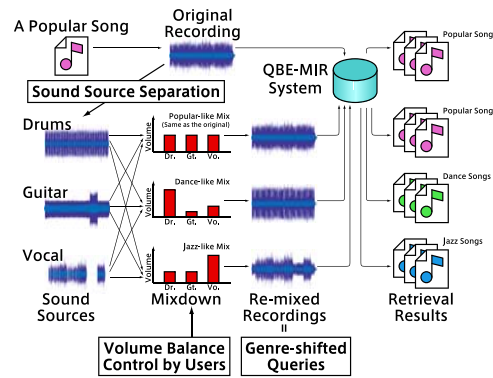


図 1 音楽ジャンルシフトを用いた類似楽曲検索の概要。
Fig. 1 Overview of QBE retrieval system based on genre shift.

クス楽曲をクエリとした類似楽曲検索を行い、単独楽器パートの音量を操作した場合、2つの楽器の組に対して音量を操作した場合のそれぞれで検索結果結果の上位ジャンルがどのように変化するかを調べる。

2. 音楽ジャンルシフトを用いた類似楽曲検索

本節では、音楽ジャンルを用いた類似楽曲検索およびそのような検索を実現するシステムの実装について述べる。

2.1 音楽ジャンルシフト

音楽ジャンルシフトとは、楽曲の楽器音量バランスの変化によって生じた、その楽曲の音響特徴量上での音楽的なジャンルの変化である。例えば、あるポピュラー楽曲の歌声を増幅し、ギターとドラムスを減衰させると、楽曲特徴量のジャンルはジャズ寄りに変化する。図 1 に示すように、楽曲の音楽音響信号を各楽器パートへと分離することで、音量バランス操作および音楽ジャンルシフトを実現する。

2.2 音響特徴量の抽出

表 1 に示す、楽曲の雰囲気を表す音響特徴量を、音楽ムード抽出の先行研究⁹⁾を参考に設計した。これらの特徴量は、パワースペクトル $X(t, f)$ からフレーム毎に抽出される。フレーム数は毎秒 100 とした。パワースペクトルは音響信号への短時間フーリエ変換で得る。

表 1 楽曲の雰囲気を表する音響特徴量。
Table 1 Acoustic features representing musical mood.

音量に関する特徴量	
次元	概要
1	全帯域の音量
2-8	サブバンド*の音量
音色に関する特徴量	
次元	概要
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*のピーク値
20-26	サブバンド*のパレー値
27-33	サブバンド*のピーク値とパレー値の比

* バンク数 7 のオクターブフィルタバンク。

以下、 t と f でそれぞれ時刻と周波数を表す。

2.2.1 音量に関する特徴量

音量に関する特徴量として全帯域の音量 $S_1(t)$ およびサブバンドの音量 $S_2(t)$ を用いる。それぞれは以下で定義される。

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

F_N はパワースペクトルの周波数インデックスの数、 $F_L(i)$ および $F_H(i)$ は第 i サブバンドの周波数インデックスの下限と上限を表す。サブバンドの音量は、楽曲の明るさを表現するのに有効である。パワースペクトルのサブバンドへの分割には、以下で表されるバンク数 n のオクターブフィルタバンクを用いた。

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

本報告での実験では、 n は 7 とした。

2.2.2 音色に関する特徴量

音色に関する特徴量として、パワースペクトルの形状を表す特徴量とサブバンド毎のパワーの比率などを表す特徴量を用いる。パワースペクトルの形状を表す特徴量として、スペクトル重心 $S_3(t)$ 、スペクトル幅 $S_4(t)$ 、スペクトルロールオフ $S_5(t)$ 、およびスペクトルフラックス $S_6(t)$ を用いる。それぞれ、以下で定義される。

$$S_3(t) = \frac{\sum_{f=1}^{F_N} X(t, f) f}{S_1(t)} \quad S_4(t) = \frac{\sum_{f=1}^{F_N} X(t, f) (f - S_3(t))^2}{S_1(t)}$$

$$S_5(t) = \sum_{f=1}^{F_N} X(t, f) = 0.95 S_1(t) \quad S_6(t) = \sum_{f=1}^{F_N} (\log X(t, f) - \log X(t-1, f))^2$$

サブバンド毎のパワーの比率を表す特徴量を定義するため、時刻 t における第 i サブバンドのパワースペクトル

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

を、パワーで降順に並べ替えたベクトル

$$(X'(i, t, 1), \dots, X'(i, t, F_N(i))) \quad \text{s.t.} \quad X'(i, t, 1) > \dots > X'(i, t, F_N(i))$$

を考える。ここで、 $F_N(i) = F_H(i) - F_L(i)$ 。このベクトルを用いて、第 i サブバンドのピーク $S_7(i, t)$ 、パレー $S_8(i, t)$ 、およびそれらの比 $S_9(i, t)$ を以下で定義する。

$$S_7(i, t) = \log \left(\frac{\sum_{f=1}^{\beta F_N(i)} X'(i, t, f)}{\beta F_N(i)} \right) \quad S_8(i, t) = \log \left(\frac{\sum_{f=(1-\beta)F_N(i)}^{F_N(i)} X'(i, t, f)}{\beta F_N(i)} \right)$$

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

ただし、 β は安定したピークとバレーを抽出するためのハイパーパラメータで、本報告の実験ではその値を 0.2 とした。

2.3 楽曲間類似度の計算

本報告の実験で用いる類似楽曲検索システムは、クエリ楽曲の音響特徴を抽出し、クエリ楽曲とデータベース中の楽曲のそれぞれとの間の音響的な類似度を計算し、データベース中の楽曲を類似度に基づいてランキングして出力する。本節では、楽曲間類似度をどのように定量化するかを述べる。

第 1 で述べたように、本報告では楽器編成やその音量比で分類される程度のジャンルを扱うため、楽曲間類似性の定量化では、楽曲の詳細な構造などよりも大まかな雰囲気をつかむ必要がある。ここでは、楽曲の雰囲気を表現するため、音響特徴量を混合ガウス分布で表す。分布の混合数は 8 とした。混合分布のパラメータを推定する際、前処理として主成分分析で特徴量の次元を圧縮する。累積寄与率が 0.95 となるように主成分を選ぶと、次元は 33 から 9 へと圧縮された。

分布間の Earth Movers Distance (EMD)⁸⁾ で楽曲間の類似度を定量化する。EMD は、一方の分布を他方の分布に変換する際の最小輸送コストに基づいて計算する。

3. 音源分離

第 1 節で述べたように、楽器音量バランスを操作するために事前に楽曲の音響信号を楽器パートへ分離する。本節では、その際の音源分離手法を述べる。

音源分離の入出力は以下のように定義される。

入力 楽曲のパワースペクトルと楽譜の組^{*1}。パワースペクトルと楽譜とは、事前に何らかの手法で時間的な同期がとられていると仮定する。

出力 各単音に分解されたパワースペクトル。

音源分離を行うため、パワースペクトルの加法性が近似的に成り立つことを仮定する。分解されたパワースペクトルと、入力スペクトログラムの位相に対して逆短時間フーリエ変換を

表 2 調波・非調波統合モデルのパラメータ。

Table 2 Parameters of integrated tone model.

記号	概要
$w_{kl}^{(J)}$	全体の強度 (音量)
$w_{kl}^{(H)}, w_{kl}^{(I)}$	調波構造モデルと非調波構造モデルの相対強度
$u_{kl}^{(H)}$	調波構造モデルの時間方向のパワー変動
$v_{kl}^{(H)}$	第 n 次倍音の相対強度
$u_{kl}^{(I)}$	非調波構造モデルの時間方向のパワー変動
$v_{kl}^{(I)}$	非調波構造モデルの周波数方向の第 n 次基底関数の相対強度
τ_{kl}	発音時刻
$\rho_{kl}^{(H)}$	調波構造モデル基底関数の時間方向への広がり
$\rho_{kl}^{(I)}$	非調波調波構造モデル基底関数の時間方向への広がり
$\omega_{kl}^{(H)}$	調波構造モデルの基本周波数
$\sigma_{kl}^{(H)}$	調波構造モデル基底関数の周波数方向への広がり
β, κ	非調波構造モデル基底関数の周波数方向への配置を定める定数

行うことで、音響信号の再合成を行う。多くの楽譜には楽器情報が含まれており、これを音源分離の反復計算アルゴリズム中で活用するため、後述する調波・非調波統合モデルの楽器ごとのモデルパラメータの分布を楽器音データベース等を用いて事前に学習する。

3.1 調波・非調波統合モデル

時刻 t 、周波数 f の 2 次元平面上で定義されたパワースペクトル $X(t, f)$ に対して、各単音へと分解する問題として音源分離をとらえる。ここで、パワースペクトルには K の楽器が含まれ、第 k 番目の楽器は L_k 個の単音を演奏しているものとする。

楽器単音のパワースペクトルを表現するモデルである、調波・非調波統合モデルを考える。第 k 番目の楽器、第 l 番目の単音 ((k, l) 番目の単音) のパワースペクトルを表現するモデルを $J_{kl}(t, f)$ で表す。本モデルは調波構造モデル $H_{kl}(t, f)$ と非調波構造モデル $I_{kl}(t, f)$ にそれぞれの相対強度 $w_{kl}^{(H)}$ と $w_{kl}^{(I)}$ を乗じて和を取り、さらに全体の強度 $w_{kl}^{(J)}$ を乗じたものとして定義される。

$$J_{kl}(t, f) = w_{kl}^{(J)} (w_{kl}^{(H)} H_{kl}(t, f) + w_{kl}^{(I)} I_{kl}(t, f))$$

$w_{kl}^{(J)}$ および $(w_{kl}^{(H)}, w_{kl}^{(I)})$ は以下の制約条件を満たす。

$$\sum_{k,l} w_{kl}^{(J)} = \iint X(t, f) dt df \quad \forall k, l : w_{kl}^{(H)} + w_{kl}^{(I)} = 1$$

調波構造モデル $H_{kl}(t, f)$ は時間周波数上での配置が拘束された 2 次元混合正規分布とし

*1 カラオケソフトなどで、有名な楽曲については楽譜 (標準 MIDI ファイルなど) を比較的容易に入手できる。

て定義される．さらに，この混合正規分布は， $\sum u_{klm}^{(H)} E_{klm}^{(H)}(t)$ および $\sum v_{kln}^{(H)} F_{kln}^{(H)}(f)$ で表される 2 つの 1 次元混合正規分布の積で表現される．このモデルは，調波時間構造化クラスタリング¹⁰⁾ で用いられた音モデルを基に構成される．これと同様に，非調波構造モデル $I_{kl}(t, f)$ は配置が拘束された 2 次元混合分布として定義され， $\sum u_{klm}^{(I)} E_{klm}^{(I)}(t)$ および $\sum v_{kln}^{(I)} F_{kln}^{(I)}(f)$ で表される 1 次元混合分布の積で表現される．これらのモデルの時間方向の関数は数式上では同一の構造をなしているが，周波数方向の関数は異なる構造を取る．調波構造モデルおよび非調波構造モデルは以下で定義する．

$$\begin{aligned}
 H_{kl}(t, f) &= \sum_{m=0}^{M_H-1} \sum_{n=1}^{N_H} u_{klm}^{(H)} E_{klm}^{(H)}(t) v_{kln}^{(H)} F_{kln}^{(H)}(f) \\
 I_{kl}(t, f) &= \sum_{m=0}^{M_I-1} \sum_{n=1}^{N_I} u_{klm}^{(I)} E_{klm}^{(I)}(t) v_{kln}^{(I)} F_{kln}^{(I)}(f) \\
 E_{klm}^{(H)}(t) &= \mathcal{N}(t; \tau_{klm}^{(H)}, (\rho_{kl}^{(H)})^2) & F_{kln}^{(H)}(f) &= \mathcal{N}(f; \omega_{kln}^{(H)}, (\sigma_{kl}^{(H)})^2) \\
 E_{klm}^{(I)}(t) &= \mathcal{N}(t; \tau_{klm}^{(I)}, (\rho_{kl}^{(I)})^2) & F_{kln}^{(I)}(f) &= \frac{1}{\sqrt{2\pi}(f+\kappa)\log\beta} \exp\left(-\frac{(\mathcal{F}(f)-n)^2}{2}\right) \\
 \tau_{klm}^{(H)} &= \tau_{kl} + m\rho_{kl}^{(H)} & \omega_{kln}^{(H)} &= n\omega_{kl}^{(H)} & \tau_{klm}^{(I)} &= \tau_{kl} + m\rho_{kl}^{(I)} & \mathcal{F}(f) &= \log_{\beta}\left(\frac{f}{\kappa} + 1\right) \\
 \mathcal{N}(x; \mu, \sigma) &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)
 \end{aligned}$$

本モデルに含まれるパラメータを表 2 に示す．ここで， M_H および N_H は調波構造モデルの時間方向及び周波数方向のガウス基底関数の数を，同様に M_I および N_I は非調波構造モデルの時間方向，周波数方向の基底関数の数をそれぞれ表す． β および κ は非調波構造モデルの周波数方向への基底関数の配置を定める定数である^{*1}． $u_{klm}^{(H)}$ ， $v_{kln}^{(H)}$ ， $u_{klm}^{(I)}$ および $v_{kln}^{(I)}$ は以下の制約条件を満たす．

$$\forall k, l: \left[\sum_m u_{klm}^{(H)} = 1 \quad \sum_n v_{kln}^{(H)} = 1 \quad \sum_m u_{klm}^{(I)} = 1 \quad \sum_n v_{kln}^{(I)} = 1 \right]$$

$F_{kln}^{(I)}(f)$ は，確率密度関数 $\mathcal{N}(g; n, 1)$ の変数変換によって得られる．

*1 $1/(\log\beta)$ と κ をそれぞれ 1127, 700 とすると， $\mathcal{F}(f)$ は fHz のメル周波数と一致する．

$$F_{kln}^{(I)}(f) = \frac{dg}{df} \mathcal{N}(\mathcal{F}(f); n, 1) = \frac{1}{(f+\kappa)\log\beta} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\mathcal{F}(f)-n)^2}{2}\right)$$

3.2 反復分離アルゴリズム

パワースペクトルの分解を行うため，以下の条件を満たす分配関数 $\Delta^{(J)}(k, l; t, f)$ を導入する．

$$\forall k, l, t, f: 0 \leq \Delta^{(J)}(k, l; t, f) \leq 1 \quad \forall t, f: \sum_{k,l} \Delta^{(J)}(k, l; t, f) = 1$$

(k, l) 番目の単音のパワースペクトル $X_{kl}^{(J)}(t, f)$ は，観測パワースペクトル $X(t, f)$ と分配関数 $\Delta^{(J)}(k, l; t, f)$ との積として得られる．さらに，分配関数 $\Delta^{(H)}(m, n; k, l, t, f)$ および $\Delta^{(I)}(m, n; k, l, t, f)$ を導入する．これらは， $X_{kl}^{(J)}(t, f)$ を調波構造モデルおよび非調波構造モデルの基底関数へと分配する関数で，以下の条件を満たす．

$$\begin{aligned}
 \forall k, l, m, n, t, f: & [0 \leq \Delta^{(H)}(m, n; k, l, t, f) \leq 1 \quad 0 \leq \Delta^{(I)}(m, n; k, l, t, f) \leq 1] \\
 \forall k, l, t, f: & \sum_{m,n} \Delta^{(H)}(m, n; k, l, t, f) + \sum_{m,n} \Delta^{(I)}(m, n; k, l, t, f) = 1
 \end{aligned}$$

それぞれの基底関数へと分配されたパワースペクトル， $X_{klmn}^{(H)}(t, f)$ および $X_{klmn}^{(I)}(t, f)$ は，それぞれ $\Delta^{(H)}(m, n; k, l, t, f)$ と $X_{kl}^{(J)}(t, f)$ との積， $\Delta^{(I)}(m, n; k, l, t, f)$ と $X_{kl}^{(J)}(t, f)$ との積として得られる．

観測パワースペクトルに対する最適な分配関数と調波・非調波統合モデルのパラメータを求めるため，分配されたパワースペクトル $X_{klmn}^{(H)}(t, f)$ ， $X_{klmn}^{(I)}(t, f)$ と，これらに対応するそれぞれの基底関数との Kullback-Leibler ダイバージェンスで目的関数を定義し，分配関数とパラメータのそれぞれに関して $Q^{(\Delta)}$ を最小化する．それぞれを最適化する式を導出すると閉じた形式として解くことができないため，一方を固定した状態でもう一方を最適化する，すなわち，反復的な最適化を行う．

パラメータの推定をロバストに行うため， $(w_{kl}^{(H)}, w_{kl}^{(I)})$ ， $(u_{klm}^{(H)}, v_{kln}^{(H)})$ ， $(u_{klm}^{(I)}, v_{kln}^{(I)})$ および $(\alpha_{w_k}^{(H)}, \alpha_{w_k}^{(I)})$ ， $(\alpha_{u_{klm}}^{(H)}, \alpha_{u_{klm}}^{(I)})$ ， $(\alpha_{v_{kln}}^{(H)}, \alpha_{v_{kln}}^{(I)})$ ， $(\alpha_{v_{kln}}^{(I)}, \alpha_{v_{kln}}^{(I)})$ に関する事前分布を導入する．事前分布はベータ分布およびディリクレ分布を用いる．以下で定義される新たな目的関数 $Q^{(\theta)}$ のもとでパラメータ最適化を行う．

$$\begin{aligned}
 Q^{(\theta)} &= Q^{(\Delta)} - \log \mathcal{B}(w_{kl}^{(H)}, w_{kl}^{(I)}; \alpha_{w_k}^{(H)}, \alpha_{w_k}^{(I)}) - \log \mathcal{D}(\{u_{klm}^{(H)}\}; \{\alpha_{u_{klm}}^{(H)}\}) \\
 &\quad - \log \mathcal{D}(\{v_{kln}^{(H)}\}; \{\alpha_{v_{kln}}^{(H)}\}) - \log \mathcal{D}(\{u_{klm}^{(I)}\}; \{\alpha_{u_{klm}}^{(I)}\}) - \log \mathcal{D}(\{v_{kln}^{(I)}\}; \{\alpha_{v_{kln}}^{(I)}\})
 \end{aligned}$$

$\mathcal{B}(\cdot)$ および $\mathcal{D}(\cdot)$ は，ベータ分布およびディリクレ分布の確率密度関数， $\alpha_{w_k}^{(H)}$ ， $\alpha_{w_k}^{(I)}$ ， $\{\alpha_{u_{klm}}^{(H)}\}$ ， $\{\alpha_{v_{kln}}^{(H)}\}$ ， $\{\alpha_{u_{klm}}^{(I)}\}$ ，および $\{\alpha_{v_{kln}}^{(I)}\}$ は楽器ごとに設定する，事前分布のパラメータである．

4. 実験

楽器音量バランスとジャンルとの関係を調査するため、実験を行った。楽器音量バランスを操作したクエリ楽曲を用いて類似楽曲検索を行い、検索結果の楽曲のジャンルからクエリ楽曲のジャンルシフトを調査した。

RWC 研究用音楽データベース:ポピュラー音楽(RWC-MDB-P-2001 No. 1-10)¹¹⁾ より、クエリ楽曲として 10 楽曲を利用した。楽曲の音響信号は、AIST アノテーション¹²⁾ として提供されている時刻同期した標準 MIDI ファイルを用いて各楽器パートへと分離した。検索対象のデータベースには、同データベース:音楽ジャンル(RWC-MDB-G-2001)¹³⁾ より、ジャンルの大分類がポップス、ロック、ダンス、ジャズ、クラシックである 50 楽曲を抜粋した。

本実験では、歌声、ギター、ドラムスの 3 楽器パートの音量を操作した。音量操作によってジャンルをシフトさせるためには、その楽器パートが十分な演奏時間と音量を持つ必要がある*1。そこで、以下の 2 条件を満たす楽器パートとして、上記の 3 パートを選択した。

- (1) クエリとなる 10 楽曲の全てで演奏されている。
- (2) 各楽曲中の 60%以上の区間で演奏されている。

楽器パートの音量を -20dB から +20dB の間で変化させながら、各クエリ楽曲とデータベース中の各楽曲との間の音響特徴量分布間の EMD を計算した。図 2 に単独楽器の音量を、残り 2 楽器の音量を固定して、操作した場合において、類似楽曲検索結果から各ジャンルの平均 EMD を計算したグラフを示す。横軸に楽器パートの音量操作量を、縦軸にジャンル毎の平均 EMD の比率を示す。平均 EMD が小さいほどグラフの下方に位置し、類似度が大きいことを表す。EMD の比率は以下で求める。

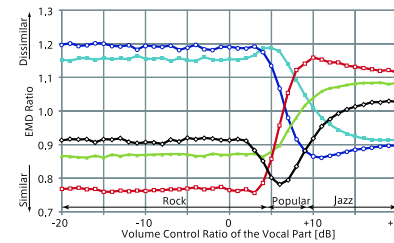
$$\text{EMD 比率} = (\text{各ジャンルの平均 EMD}) / (\text{全ジャンルの平均 EMD})$$

また、図 3 に 2 楽器の音量を操作した場合において EMD 比率が最小となった、すなわち最大類似度を持つジャンルをプロットしたグラフを示す。

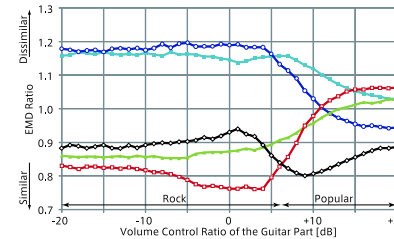
4.1 単一楽器の音量操作

図 2 より、楽器音量を操作することでジャンルシフトが起こったことが分かる。各楽器パートの音量操作において、操作量が 0dB の場合はオリジナルのクエリ楽曲を用いた場合に相当するため、(a), (b), (c) のいずれも同一の結果となることに留意する。クエリに用いた 10 楽曲はポピュラー楽曲 DB から抜粋したが、いずれもギターとドラムスの音量が比較

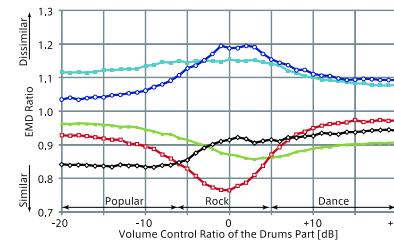
*1 5 分間の楽曲中で 10 秒程度しか演奏されない楽曲の音量を操作しても、楽曲のジャンルは変化しないだろう。



(a) 歌声の音量操作によるジャンルシフト。類似度最大のジャンルはロックからポップス、ジャズへとシフトした。

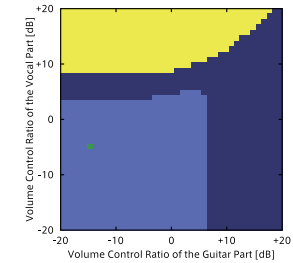


(b) ギターの音量操作によるジャンルシフト。類似度最大のジャンルはロックからポップスへとシフトした。

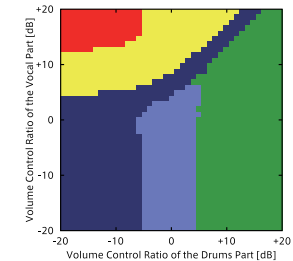


(c) ドラムスの音量操作によるジャンルシフト。類似度最大のジャンルはポップス、ロック、ダンスとシフトした。

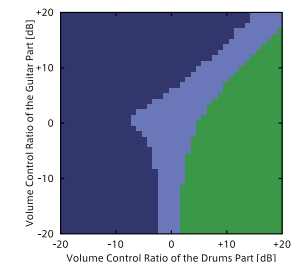
◀ Popular ◀ Rock ◀ Dance
 ▶ Jazz ▶ Classical



(a) 歌声とギターの音量操作によるジャンルシフト。



(b) 歌声とドラムスの音量操作によるジャンルシフト。



(c) ギターとドラムスの音量操作によるジャンルシフト。

■ Popular ■ Rock ■ Dance
 ■ Jazz ■ Classical

図 2 単一楽器の音量を減衰もしくは増幅した場合のジャンル毎の EMD 比率の変化。 図 3 2 楽器の音量を操作した場合の最小 EMD (最大類似度) のジャンル。

Fig. 2 Ratio of average EMD per genre to average EMD of all genres while reducing or boosting the volume of single instrument part. Fig. 3 Genres that have the smallest EMD (the highest similarity) while reducing or boosting the volume of two instrument parts.

的大きい楽曲であるため、音量操作を行わない場合にロックの類似度が最大となったと考えられる。

図 2 (a) より、歌声の音量を -20dB から増幅するにつれて、類似度が最大のジャンルはロック (-20 から 4dB)、ポップス (5 から 9dB)、ジャズ (10 から 20dB) と変化したことが分かる。同様に、図 2 (b) より、ギターの音量を増幅するにつれて類似度最大のジャンルはロック (-20 から 7dB) からポップス (8 から 20dB) へと変化したことが分かる。ジャンルがロックからポップスへとシフトした点は歌声とギターの音量操作で共通しているが、ジャズへとシフトしたのは歌声を操作した場合のみであった。このことから、歌声とギターとはジャズ音楽において異なる役割を果たしていることが示唆される。また、図 2 (c) より、ドラムスの音量を増幅するにつれて類似度最大のジャンルはポップス (-20 から -7dB)、ロック (-6 から 4dB)、ダンス (5 から 20dB) へと変化したことが分かる。これらの結果より、楽器音量バランスとジャンルシフトとの間には合理的な関係があり、音楽ジャンルの典型的なイメージと整合していることが示される。

4.2 2 楽器の音量操作

図 3 に 2 楽器パートの音量操作によるジャンルシフトを示す。一方の楽器の音量を操作しない (0dB の軸上) の場合、結果は図 2 のものと同一となる。

単一楽器操作の場合と基本的なジャンルシフトの傾向は同じだが、図 2 ではクラシックの類似度はどのような操作を施しても最大とはならなかったが、歌声を増幅、ドラムスを減衰させた場合である図 3 (b) では最大類似度となっている。ギターとドラムを個別に増幅した場合はロックの類似度が減少したが、図 3 (c) に示すようにこれらを同時に増幅した場合にはロックが最大類似度を維持していることは興味深い。この結果はロックの典型的なイメージ (ギターとドラムスの両方が用いられる) と一致しており、類似楽曲検索におけるクエリのカスタマイズが有効であることを示唆している。

5. おわりに

本報告では、楽器音量バランスを操作することによる楽曲のジャンルシフトと、ジャンルシフトを用いた類似楽曲検索手法について述べた。従来はユーザがクエリをカスタマイズすることは困難であり、類似楽曲検索において異なる検索結果を得るためには新たなクエリを用意する必要があったが、本手法は単一のクエリ楽曲から多様な検索結果を引き出す。我々が開発した音源分離手法を用いたジャンルシフトは、楽器音量バランスを操作するという単純で直感的なクエリのカスタマイズを可能にした。実験によって、ジャンルシフトと楽器音

量バランス操作との関係性を示した。

本報告では、音量バランス操作によるジャンルシフトのみを対象としたが、リズムパターン、エフェクト、コード進行など、音楽ジャンルと関係のある他の音楽的要素に関しても、これらを操作することによってジャンルシフトは起こると考えられる。今後は、ユーザが意図や興味をより簡単に反映できるような類似楽曲検索システムの実現を目指す。

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) Feng, Y., Zhuang, Y. and Pan, Y.: Music Information Retrieval by Detecting Mood via Computational Media Aesthetics, *Proc. WI*, pp.235–241 (2003).
- 4) Thoshkanna, B. and Ramakrishnan, K.R.: Projekt Quebex: A Query by Example System for Audio Retrieval, *Proc. ICME*, pp.265–268 (2005).
- 5) Vignoli, F. and Pauws, S.: A Music Retrieval System Based on User-driven Similarity and Its Evaluation, *Proc. ISMIR*, pp.272–279 (2005).
- 6) Itoyama, K., Goto, M., Komatani, K., Ogata, T. and Okuno, H.: Instrument Equalizer for Query-by-Example Retrieval: Improving Sound Source Separation based on Integrated Harmonic and Inharmonic Models, *Proc. ISMIR*, pp.133–138 (2008).
- 7) Itoyama, K., Goto, M., Komatani, K., Ogata, T. and Okuno, H.G.: Parameter Estimation for Harmonic and Inharmonic Models by Using Timbre Feature Distributions, *IPSJ Journal*, Vol.50, No.7 (2009).
- 8) Rubner, Y., Tomasi, C. and Guibas, L.J.: A Metric for Distributions with Applications to Image Databases, *Proc. ICCV*, pp.59–66 (1998).
- 9) Lu, L., Liu, D. and Zhang, H.J.: Automatic Mood Detection and Tracking of Music Audio Signals, *IEEE Trans. Audio, Speech and Lang. Process.*, Vol.14, No.1, pp.5–18 (2006).
- 10) Kameoka, H., Nishimoto, T. and Sagayama, S.: Harmonic-temporal Structured Clustering via Deterministic Annealing EM Algorithm for Audio Feature Extraction, *Proc. ISMIR*, pp. 115–122 (2005).
- 11) Goto, M., Hashiguchi, H., Nishimura, T. and Oka, R.: RWC Music Database: Popular, Classical, and Jazz Music Databases, *Proc. ISMIR*, pp.287–288 (2002).
- 12) Goto, M.: AIST Annotation for the RWC Music Database, *Proc. ISMIR*, pp.359–360 (2006).
- 13) 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).