

5R-5

A Music Retrieval Approach from Alternative Genres of Query by Adjusting Instrument Volume

Wang Kaiping[†] Katsutoshi Itoyama[†] Masataka Goto[‡]
 Kazunori Komatani[†] Tetsuya Ogata[†] Hiroshi G. Okuno[†]

[†] Graduate School of Informatics, Kyoto University [‡] 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 [1] 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 で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 [2] 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 [3] を用いる。

本リミックスに基づく検索が機能するための仮説は、楽曲のジャンルはその楽曲を構成する楽器およびその音量比率によって決まる、ということである。この仮説は先行研究 [3] 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が明らかにする。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クス楽曲をクエリとした類似楽曲検索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. 類似楽曲検索

本章では、類似楽曲検索の実装について述べる。

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 [4] 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からフレーム毎 (毎秒 100 フレーム) に抽出を行う。

2.2 類似度の計算

楽曲ごとの全体的な雰囲気を表現するために、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

表 1: 楽曲の雰囲気を表現する特徴量。

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

*バンク数 7 のオクターブフィルタバンク。

は、混合数を 8 とした。各楽曲内での雰囲気の時間的な変動は扱わない。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) [5] を用いる。

3. 音源分離

本稿では、音源分離問題を入力混合音のパワースペクトル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。このような分解を行うために、各単音のパワースペ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 [3] を用いる。このモデルは、調波的な音のパワースペクトルを表現する調波構造モデルと非調波的な音のパワースペ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。分離処理の詳細については、紙面の制約上省略する。

4. 実験

類似楽曲検索において、楽曲のリミックスによってどのように検索結果が変化するかを調査する実験を行った。

4.1 実験条件

クエリとなる楽曲には RWC 音楽データベース: ポピュラー音楽, No. 1-10 の 10 楽曲を用いた。これらの楽曲と、同データベースに付属の標準 MIDI ファイル (楽譜に相当) を用いて前述の音源分離手法で楽器パートごとの音源分離を行った。また、同データベース: 音楽ジャンルからジャンルの大分類がポップス, ロック, ダンス, ジャズ, クラシックである合計 50 楽曲を選択し、検索対象となるデータベースを構築した。

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。楽器パートの音量操作で楽曲のジャンルを変化させるためには、操作対象の楽器パートは楽曲中での演奏時間が十分に長くなければならない*。そ

*演奏時間 5 分の楽曲中で 10 秒程度しか演奏されない楽曲の音量を操作したとしても、楽曲のジャンルが変化することはないだろう。

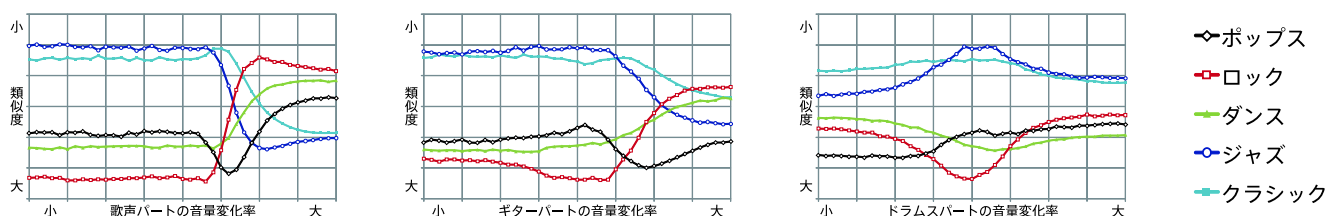


図 1: 楽器パートの音量を操作して検索を行い, ジャンルごとの平均 EMD の全ジャンルの平均 EMD に対する比率を求めたもの. 左から順に, 歌声, ギター, ドラムスパートの音量を操作した場合. EMD の比率が小さい (グラフの下方にプロットされている) ほど類似度は大きい.

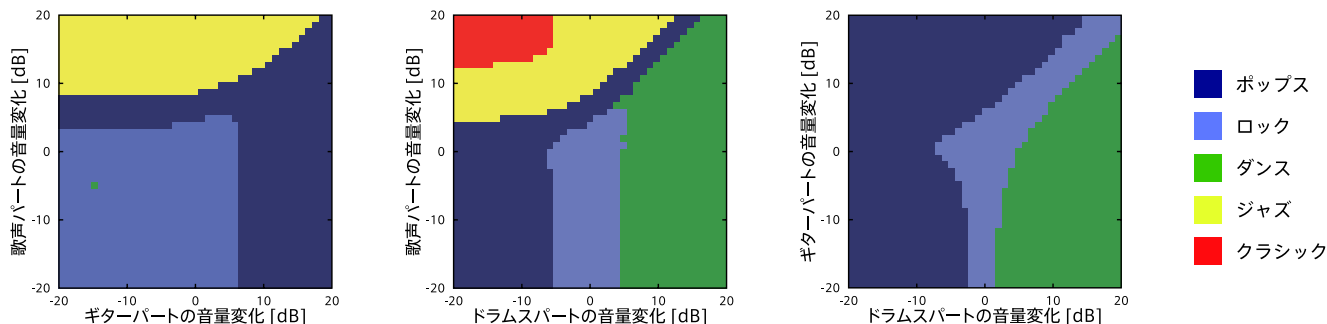


図 2: 2 つの楽器パートの音量を操作して検索を行い, 平均 EMD が最も小さくなった (類似度が最も大きい) ジャンルをプロットしたもの.

ここで, 以下の 2 条件:

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

を満たす楽器パートとして, 上記の 3 パートを選択した.

これらの楽器パートの音量を -20 – 20 dB の範囲で操作し, リミックスされた音響信号をクエリとしてデータベース中の楽曲の特徴量分布との EMD を計算した. 単独の楽器パートを操作した場合の結果を図 1 に, 2 つを操作した場合を図 2 に, それぞれ示す. クエリのジャンル変化を調べるためには, EMD そのものよりも全楽曲の平均 EMD に対する相対 EMD の変化がよりよい指標となるため, ジャンルごとの平均 EMD の全ジャンルでの平均 EMD に対する比率を示している.

4.2 考察

歌声パートの音量を -20 dB から 20 dB へと徐々に大きくすることで, もっとも類似度が大きいジャンルはロック, ポップス, クラシックと変化した. 同様に, ギターパートではロック, ポップスと, ドラムスパートではポップス, ロック, ダンスと変化した. ドラムスを増幅することで 5 ジャンル中では最もリズムが重視され, ドラムスが多用されるダンスの類似度が大きくなったこと, 歌声やギターを増幅することでロックに比べて調波的な楽器が用いられるポップスやクラシックの類似度が大きくなった. すなわち, 楽曲のリミックス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たと言える. ギターを増幅することでギターが多用されるロックの類似度が小さくなっている. これは相対的にドラムスなどの音量が小さくなったことがジャンルにより大きな影響を与えたことによると考えられる.

2 つの楽器パートを操作した場合, おおむね単独パートをした場合と同様の結果が得られた. しかし歌声パートを増幅し, ドラムスパートを減衰させた場合に限ってはクラシックの類似度が最も大きくなっており, 楽器の関係が独立でないことを表している.

5. おわりに

本稿では, 楽曲中の楽器パートの音量操作によって楽曲のジャンルがどのように変化するかを調べ, 複数の楽器の間の依存関係を確認した. 今後は, 音源分離によって楽器パートのリミックスを行うだけでなく, 楽曲制作時のクリーンな楽器パートの音響信号を用いてミックスダウンを行うことで, 本手法に必要な音源分離性能などについて検討する予定である.

謝辞 本研究の一部は, 科研費, グローバル COE, Crest-Muse の支援を受けた.

参考文献

- [1] A. Rauber *et al.*: Using Psycho-acoustic Models and Self-organizing Maps to Create a Hierarchical Structuring of Music by Sound Similarity, ISMIR, 2002.
- [2] K. Itoyama *et al.*: Instrument Equalizer for Query-by-Example Retrieval: Improving Sound Source Separation based on Integrated Harmonic and Inharmonic Models, ISMIR, 2008.
- [3] 糸山他: 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, 情処論, Vol. 49, No. 3, 2008.
- [4] 西山他: 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, 情処研報 2007-MUS-069, 2007.
- [5] Y. Rubner *et al.*: A Metric for Distributions with Applications to Image Databases, ICCV, 1998.
- [6] 後藤他: RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース, 情処論, Vol. 45, No. 3, 2004.