

## 楽曲の音響信号からの高揚度変化の推定\*

福谷 和貴 酒向 慎司

名古屋工業大学 工学部

### 1 はじめに

現在、定額制音楽配信サービスを用いて音楽を聴取することが広まってきており、好きなCDを買ってきて好きな楽曲を聴取するというスタイルから、膨大な量の楽曲を聴取するというスタイルに聴取の方法が変わりつつある。しかしその一方で、膨大な量の楽曲をすべて把握することはできず、自分が聴きたい楽曲を見つけることは困難になっている。定額制音楽配信サービスではプレイリストを用意することでユーザーが好きな楽曲を見つける補助を行っているが、1曲全てを聴く必要があるため、時間がかかるという欠点がある。他の解決策の一つとして、様々な楽曲からメドレー曲を作成し、短時間で様々な楽曲を聴取することを容易にする方法も有用である。

自動メドレー生成に必要な主要要素技術として、楽曲のどの部分を切り出し、メドレー曲に使用するかを決める区間検出と、様々な楽曲区間をどのように接続するか二つが考えられる。メドレー曲が備えておくべき要件としては、個々の楽曲が違和感なく接続されていることが重要であり、本研究では楽曲中の展開が一致しているかどうかに着目する。曲の展開の不一致とは、メドレー曲において、曲が変化する前は盛り上がりつつ進行しているのに、曲が変化すると、落ち着いて進行してしまうなどの状況である。メドレー生成においてこのような問題が生じないようにするためには、楽曲から人間が感じる高揚度の変化を適切に推定することが重要であると考えられる。

本研究では、ポピュラー音楽を対象に、楽曲の音響信号から高揚度変化の推定を行うことを目標として、主に高揚度変化の指標化とその推定に有効な音響特徴量について検討する。

### 2 高揚度変化の要因

小山らは演奏動画における盛り上がり検出に用いる特徴量について検討し [1]、音量RMS、オンセット数が高揚度と相関があることが確認さ

れた。これを参考に以下の3つの点を高揚度変化に関係が深い要因として挙げる。

- 音量の変化  
音量を大きくしていく箇所は、盛り上がる箇所の直前に位置し、高揚度を高まらせることが多い。
- 音数の変化  
音の鳴る密度が高くなる、例えばドラムが連打される状況では、高揚度が変化すると考えられる。
- 音色の変化  
暗い音色から明るい音色に変化するなど、音色の変化に伴い高揚度も変化すると考えられる。

### 3 高揚度変化推定に用いる音響特徴量

2章で挙げた高揚度変化の要因を反映した以下の音響特徴量を高揚度変化推定に用いる。それぞれの音響特徴量の抽出には音楽信号処理ライブラリ LibROSA [2]を用いる。

- 音量RMS (I)  
ある時間窓幅における音量の振幅の二乗平均平方根を特徴量として用いることで音量変化の検出が期待される。
- オンセット数  
オンセット数は音の立ち上がり数であるので、音数の変化を表現することが出来る。調波音と打楽器音では高揚度の変化に与える影響の大きさが異なるので、調波音(II)、打楽器音(III)に音源分離し、ある時間窓幅におけるそれぞれのオンセット数をそれぞれ特徴量として用いる。
- スペクトル重心(IV)  
スペクトル重心はパワースペクトルの重心であり、音色の明るさと関係がある。スペクトル重心を特徴量として用いることで音色の変化を検出することが期待される。

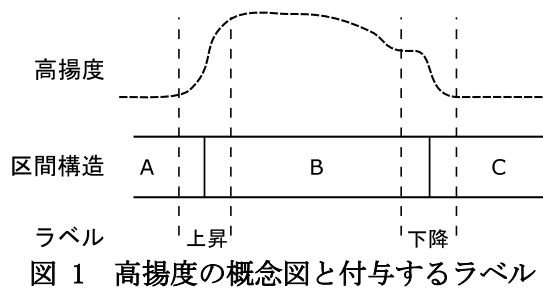
音響特徴量を以下の4通りの条件で抽出しそれぞれの音響特徴量で機械学習を試行する。

1. 窓幅：0.5秒    ホップ幅：0.25秒
2. 窓幅：0.5秒    ホップ幅：0.375秒
3. 窓幅：0.25秒    ホップ幅：0.125秒
4. 窓幅：0.25秒    ホップ幅：0.1875秒

### 4 実験

本研究では実験に用いるデータセットにRWC研

\*An estimation of degree of excitement by song audio signals  
Kazuki Fukutani, Shinji Sako, Faculty of Engineering, Nagoya Institute of Technology



究用音楽データベース [3]のポピュラー音楽の内、楽曲番号が奇数である 50 曲を用いる。まず、被験者実験を行い高揚度変化ラベルの収集を行った後、機械学習アルゴリズムを用いて高揚度変化ラベル推定の評価実験を行う。

#### 4.1 高揚度変化ラベルの収集

AIST Annotation [4]で与えられるパート構造の境界前後 5 秒間を被験者に聴取させ、その高揚度の変化を“上昇”，“維持”，“下降”の 3 種類に分類させる。図 1 に付与するラベルのイメージを示す。5 秒に満たないパートでは高揚度は変化しないと考える。本研究では 5 秒未満のパートは例外として前のパートと結合する処理を行う。ただし曲の最初のパートが 5 秒未満だった場合には後ろのパートと結合する。1 サンプルにつき 3 名のラベルを収集する。

#### 4.2 機械学習による実験

3 章で挙げた 4 つの音響特徴量、前節で収集した高揚度変化ラベルを用いて、機械学習アルゴリズムによる音響特徴量から高揚度変化ラベルの推定を行う。収集したラベルは全てで 736 サンプルあり、データセットとして①多数決を行い 3 人の回答が異なるものは最も多数決の結果と一致する人物のラベルを採用したもの(736 サンプル)、②多数決を行い 3 人の回答が異なるものは排除したもの(691 サンプル)、③3 名が一致したもののみ(268 サンプル)の 3 種類に分け、それぞれのデータセットで学習、評価を行う。機械学習のアルゴリズムは Linear SVC と K 近傍法を用いる。

表 1 にそれぞれのデータセット、アルゴリズムで最も良い結果が出た音響特徴量の組み合わせ、その結果を示す。すべてのデータセットにおいて、機械学習のアルゴリズムに k 近傍法を用いた方が、正解率が高くなっている。

データセットの違いに着目すると、データセット③が最も正しく分類されており、ほか 2 つのデータセットには大きな違いは見られない。多数決が決まらなかったサンプルがあることから、人によって高揚度が変化を判断する要因が異な

表 1 高揚度変化の推定結果

データセット	アルゴリズム	音響特徴量	窓幅 (s)	ホップ幅 (s)	正解率 (%)
①	SVC	I, IV	0.5	0.375	56
	k 近傍法	I, III	0.25	0.1875	57
②	SVC	I, IV	0.5	0.375	57
	k 近傍法	I, III	0.25	0.125	58
③	SVC	全て	0.25	0.125	65
	k 近傍法	I, II, IV	0.5	0.5	70

ると考えられる。データセット①、②の結果からそのようなサンプルは機械学習で推定することが困難である。しかし、全員が一致したサンプルもあることから、着目している要因が異なっても、同じように高揚度が変化するサンプルが存在する。そのようなサンプルは機械学習での推定ができることがデータセット③の結果から推測される。

#### 5 まとめ

本研究では、高揚度変化のラベル付けを行った。さらに音量 RMS、調波音・打楽器音のオンセット数、スペクトル重心を音響特徴量に用いて音楽音響信号に対する高揚度変化の推定を行った。3 種類のデータセットを用いて高揚度変化の推定を行った結果、Linear SVC と比べ k 近傍法の方が良い結果を示した。また、3 人一致のみデータセットを用いた推定結果が最も良いものとなった。今後は、人物による解釈の差異を検証するために学習に用いたラベルデータに対して未知の人物のラベルを評価対象とする実験や、音響特徴量を時系列に並べたもの以外にも全体の統計量を付与することなど、さらなる検証を進める。

#### 参考文献

- [1] 小山健一, 他, “E-041 演奏動画の盛り上がり検出に用いる特徴量の検討,” *情報科学技術フォーラム講演論文集*, vol. 10, no. 2, pp. 305-306, 2011.
- [2] B. McFee, et al. “librosa: Audio and music signal analysis in python,” *Proceedings of the 14th python in science conference*, pp. 18-25, 2015.
- [3] M. Goto, et al. “RWC Music Database: Popular, Classical and Jazz Music Databases.,” *ISMIR*, vol. 2, pp. 287-288, 2002.
- [4] M. Goto, “AIST Annotation for the RWC Music Database.,” *ISMIR*, pp. 359-360, 2006.