

## 編曲を考慮した旋律検索方法の提案

向後 憲治<sup>†</sup> 黄 宏軒<sup>†</sup> 川越 恭二<sup>†</sup>

<sup>†</sup>立命館大学 情報理工学部

### 1 はじめに

手軽に持ち運び可能な音楽再生機器の普及に伴い、iTunes やハイレゾ音源提供サイト [1] など音楽ファイルを提供するサービスが充実している。これらの音楽サービスの中では、歌詞から特定の楽曲を検索する手法 [2] や、アーティスト名やアルバム名から検索する手法 [3] などが使用されている。

また、クラシック曲を基調としたポピュラー音楽や、ジャズ曲をアレンジしたダンスミュージック等、原曲が編曲された楽曲も数多く存在する。しかし、視聴中の楽曲と同一の原曲から編曲された、異なるジャンルやアーティストの楽曲を得ることが可能な既存楽曲検索システムは存在しない。

そこで本研究では、原曲からの編曲で得られた曲が原曲と同一あるいは非常に類似した曲として検索可能な編曲を考慮した類似フレーズ抽出方法を有する旋律検索手法を提案する。提案手法は、音の高さ（音高）と音の長さ（音価）をそれぞれ時系列データとみなし、時系列データ間における類似度計算を行う。曲間類似度の算出には、時系列データ間における類似度を測る方法として頻用される Dynamic Time Wrapping (DTW) を用いる。しかし、従来の DTW では、編曲による原曲とのキーの違いやテンポの違いを考慮できない。そこで、本手法では時系列データの正規化変換処理を行うことでこれらの問題を解決する。

### 2 旋律検索の関連研究と要件

旋律検索手法の考案は様々行われており、矢澤ら [4] は旋律において、連続する 3 つの音を 1 つのシンボルとし、そのシンボルに名前を与えた。これらを旋律順に並べ、名前が並んだ文字列とみなすことで、旋律間類似度を文字列間類似度と定義し直すことが出来る。これによって矢澤らは N-gram 法を用いて楽曲間類似度を測定することを可能とした。連続した 3 音の上がり下がりに基づいた類似度になるため、曲の雰囲気似ているかどうかを判断する際に有効といえる。しかし、本研究では音高に着目しており、音価に関しては考慮されていない。そのため、異なる音価を持つ楽曲であっても、音高が類似していれば、類似した楽曲であると判断されてしまう。

本研究で着目している原曲と編曲の類似性を測るためには、音高と音価どちらも考慮することが必要であ

る。さらに、編曲された旋律において原曲より音高が高くなっている場合や、音価が大きくなりテンポが変化している場合、また部分的に原曲が編曲内において使用されている場合においても類似しているとみなすような手法が必要である。本稿で提案する手法では、これらの場合に対応した手法を提案している。

### 3 曲同士の距離計算

本研究の最終的な目標は、原曲旋律と編曲された旋律間、及び編曲された旋律同士が類似しているという計算結果を出力することである。そのために旋律情報を取得し正規化処理を行った後、DTW を適用する。

#### 3.1 数値正規化変換処理

本研究では、基底として使用する手法は時系列データの類似度算出手法として有名な DTW である。音高と音価を時系列データとみなして DTW を適用する。

ただし、DTW を旋律情報に対して使用するにあたって、本研究の目的を達成するには問題点がある。DTW は旋律間での距離を測るため、例えば、原曲に比べて編曲された曲が全体的に 1 音ずれている状況にあった場合、音の流れとしては全く同じであっても、距離を算出した場合はずれた分だけ数値が大きくなる。これが 1 音ではなく 1 オクターブずれとなった場合、算出された数値は顕著に大きくなる。また、これは音価に対しても同様のことが言える。例えば、音価が原曲に比べて全体的に 1.15 倍に伸びた編曲旋律が合った場合、リズムパターンは完全に一致しているにもかかわらず、距離として大きな数値が算出され、類似していないと判定されてしまう。

これらの問題を解決するために、本研究では旋律情報に対して前処理を行う。まず、時系列旋律データ  $f(t)$  の旋律全体の平均値  $average[f(t)]$  を、旋律全体の音数を  $n$  とする。

$$average[f(t)] = \frac{1}{n} \sum_{i=1}^n f(t) \quad (1)$$

次に、個々の  $f(t)$  に対して  $average[f(t)]$  の差を取る。これを  $degree[f(t)]$  とする。

$$degree[f(t)] = f(t) - average[f(t)] \quad (2)$$

最後に、旋律全体の最大値  $max[f(t)]$  を取得し、先ほど計算した差分  $degree[f(t)]$  を  $max[f(t)]$  で割る。この値を  $F(t)$  すると式 (3) のように表すことができる。

$$F(t) = \frac{degree[f(t)]}{max[f(t)]} \times K \quad (3)$$

A Music Scores Search Method using Arrangement Detection

Kenji KOGO<sup>†</sup>, Hung-Hsuan HUANG<sup>†</sup>, Kyoji KAWAGOE<sup>†</sup>

<sup>†</sup>Ritsumeikan University

この式を音高と音価にそれぞれ適用する。ここで、最後に値を  $K(=1)$  倍にするのは、DTW を適用した結果が、極めて小さい値となるので、視覚的に確認を容易にするためである。この式により旋律情報を時間軸で正規化することが可能になり、音高のずれや音価の伸び縮みを考慮した類似度を算出することが可能である。

音価において式 (3) を適用し、変換された値を DTW に適用した結果を *DistanceLength*、音高において適用した結果を *DistanceHeight* とする。

### 3.2 最終距離

3.1 節で述べた *DistanceLength* と *DistanceHeight* を統合したものを最終距離 *DistanceResult* とする。このとき、距離数値の高い方を重視した距離を *DistanceResult* と定義する。*DistanceLength* の重みを  $W_L$ 、*DistanceHeight* の重みを  $W_H$  ( $W_L + W_H = 1.0$ ) とすると、

$$DistanceResult = DistanceLength \times W_L + DistanceHeight \times W_H \quad (4)$$

と表現できる。

## 4 評価実験

### 4.1 実験内容

今回は式 (3) の正当性を確認するための実験として、単音からなる平均音数 46 の No.1 から No.30 という midi 音源 30 種を用意し、この音源を原曲として、No.1 から No.10 の音高を全体的に  $-8 \sim +8$  音シフトさせた音源を 10 個 (No.31 ~ No.40)、No.11 から No.20 の音価値を全体に  $0.75 \sim 1.09$  倍変化させた音源を 10 個 (No.41 ~ No.50)、また、No.21 から No.30 音高を  $-8 \sim +8$  音シフト、音価値を全体に  $0.75 \sim 1.09$  倍全体的に変化させたものを 10 個 (No.51 ~ No.60)、合計 60 個の音源を対象に式 (3) を適用後、DTW を適用し交差検証を行った。

手法比較として、編集距離を用いた距離計算を行った。ただし、DTW と同じく、編集距離をそのまま旋律に対して適用すると音高のずれを考慮できないため、今回は隣り合う 2 つの音の音高差分をとった数値を用いた距離計算を行っている。この手法では音価の増減にはできないため、音高のみの評価とする。

また、今回式 (4) において、今回の実験では  $W_H = 0.7$ 、 $W_L = 0.3$  で実験を行う。

### 4.2 実験結果

式 (3) 適用後の DTW 距離で、原曲と編曲を行った楽曲の組み合わせについて着目すると、すべての組み合わせで極めて小さな距離が算出された。

表 1 に結果の一例を示す。今回の例では No.1 と No.15 を挙げている。表 2 では今回交差検証を行った 3481 組における音高と音価それぞれの平均距離と、原曲と編曲間に限定した全 30 組の音高と音価それぞれの平均距離を示す。また表 3 で手法比較の結果を示す。最後に表 4 で式 (4) の評価結果を示す。表 1, 2, 3 から原曲と編曲間距離が他の曲間より小さくなり、表 4 より式 (4) の *DistanceResult* が、音高のみの評価である編集距離の値より小さくなるという結果を得ることができた。

表 1: 提案手法による結果 ( $K = 100$ )

比較対象	No.1(音高比較)	No.15(音価比較)
全楽曲との平均距離 (A)	5.86	109.9
編曲された旋律との距離 (B)	0.0297	0.000119
(A) / (B)	197.3	924000

表 2: 交差検証結果

評価項目	音高	音価
全組み合わせの平均距離 (C)	8.11	168.5
原曲と編曲間の平均距離 (D)	0.0846	0.000790
(C) / (D)	95.8	213000

表 3: No.1 における手法比較結果

比較対象	提案手法	編集距離
全楽曲との平均距離 (E)	5.86	49.9
No.1 と No.31(F)	0.0297	52.0
(E) / (F)	197.3	0.960

表 4: 式 (4) の評価実験 ( $W_H = 0.7$ ,  $W_L = 0.3$ )

比較対象	<i>Distance Height</i>	<i>Distance Length</i>	最終距離	編集距離 (音高のみ)
No.1 と No.31	0.0297	0.00	0.021	52.0
No.15 と No.45	0.00	0.000119	0.0000357	0.00
No.30 と No.60	0.100	0.000330	0.070	30.0

## 5 おわりに

本研究では原曲と編曲された旋律間の類似性を考慮した類似度算出手法を提案した。本手法では、ある原曲と編曲間の距離を著しく小さくすることが可能であるため、カバー曲の検索や、曲の一部に原曲が使用された編曲を検索する際に有効な手法である。今後式 (4) における  $W_H$  及び  $W_L$  の適切な値を定める評価実験を行う予定である。

## 参考文献

- [1] ONKYO ENTERTAINMENT TECHNOLOGY CORPORATION: ハイレゾ音源配信サイト【e-onkyo music】入手先 <<http://www.e-onkyo.com/music/>> (2015/01/04 参照)
- [2] 歌詞検索 J-Lyric.net 入手先 <<http://j-lyric.net/>> (2015/01/04 参照)
- [3] 【レコチョク】50音 ABC 検索:歌詞検索 入手先 <[http://recochoku.jp/search\\_50abc/](http://recochoku.jp/search_50abc/)> (2015/01/04 参照)
- [4] 矢澤 櫻子, 浜中 雅俊 (2014), 「主観的類似度を反映した暗意実現モデルの拡張」, 研究報告音楽情報科学 (MUS) 2014-MUS-102 PP.1 - 5