

## 歌詞の構造に基づく楽曲の繰り返し構造推定

## A Method of Music Structural Segmentation Based on the Lyrics

大野 将樹<sup>†</sup>

小川 卓人

Masaki Oono

Takuto Ogawa

## 1. はじめに

本論文では、楽曲の繰り返し構造を推定する手法を提案する。本論文が取り扱う楽曲の繰り返し構造とは、Aメロ、Bメロ、サビ(主題)などの、人が容易に知覚可能なリフレイン構造である。楽曲の繰り返し構造を抽出すれば、音楽の要約や楽曲検索、ジャンル分類の精度向上に有用である。従来の楽曲の繰り返し構造推定法は、音楽音響信号を対象とし、主にサビ位置あるいはサビ区間を検出することを目的としていた[1]。提案手法は、歌詞テキストの記述形式に着目し、音楽音響信号を用いずにサビを含む全ての繰り返し構造を推定するものである。近年、歌詞は、インターネット上の歌詞配布サイトから容易に入手可能であり、その記述形式には、一行あたりのモーラ数や改行など、楽曲の繰り返し構造を表す特徴が数多く含まれている。提案手法は、歌詞記述者の繰り返し構造の解釈を利用したものであり、より人間の認知に近い構造推定を行うことを目的とする。

## 2. 提案手法

## 2.1 歌詞に着目した構造推定

歌詞には、歌詞記述者が楽曲を聴き、その構造を解釈した痕跡が残されている。本研究では、楽曲構造解釈の痕跡として、歌詞中の改行、空白行、一行の音数を利用し、楽曲の繰り返し構造を検出する。提案システムの流れを図1に示す。まず、入力として音楽音響信号(wav)あるいはMIDI信号(smfi)が歌詞検索部に渡される。歌詞検索部では、入力楽曲のタイトル・歌手名などのメタデータから歌詞データベースを参照し、歌詞テキストを得る。次に、特徴抽出部において、歌詞の音数に基づく特徴量の抽出、空白行に基づくブロック分割を行う。分類部では、先に得られたブロック群に対するクラスタリングを行う。構造推定部では、各クラスタについて、繰り返し構造の種類(Aメロ、Bメロ、サビ等)を推定し、歌詞に繰り返し区間を付与したxmlファイルを出力する。

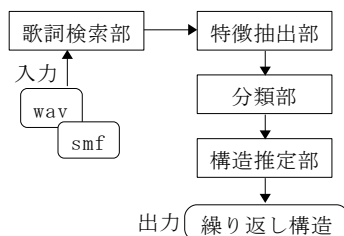


図1: 提案システムの概要

## 2.2 特徴抽出部

特徴抽出部では歌詞テキストから特徴量を算出する。ここで、歌詞記述者の楽曲構造認知とその痕跡に関する2つの仮定を設ける。

**(仮定1)** 一曲中における同一の繰り返し区間(1番のサビと2番のサビなど)において、その各行(1番のサビの1行目と2番のサビの1行目など)で発音される音の数は、ほぼ等しい。

**(仮定2)** 歌詞中に挿入された空白行は、繰り返し区間の区切りに対応する。

歌詞記述者が楽曲を聴き、繰り返し構造を解釈した上で改行を挿入したと仮定するものが仮定1であり、空白行を挿入したと仮定するものが仮定2である。本研究では、仮定1に基づき、歌詞テキストの特徴量として歌詞1行ごとのモーラ数を用いる。仮定2に基づき、空白行を繰り返し区間の区切りとして、空白行と空白行に挟まれた区間*i*(ブロックと呼ぶ)の特徴ベクトル  $X_i$  を作成する。

$$X_i = (x_0, x_1, \dots, x_j, \dots, x_n) \quad (1)$$

$x_j$  はブロック  $i$  の  $j$  行目のモーラ数、 $n$  はブロック  $i$  の行数である。

## 2.3 分類部

分類部では、前節で得られた全てのブロックの特徴ベクトル  $X$  をクラスタリングすることによって、同一の繰り返し区間を検出する。本研究では、クラスタリング手法として階層的手法を採用し、距離関数にはWard法を用いた。各クラスタに属するブロックには、同一の繰り返し区間であることを示すクラスタID(0, 1, 2, ...,  $m$ ;  $m$  はクラスタ数)を付与した。例えば、楽曲の先頭ブロックと2番目のブロックが同一クラスタであり、3番目と4番目のブロックが同一クラスタならば、クラスタID系列は0, 0, 1, 1となり、全て異なるクラスタならば0, 1, 2, 3となる。

## 2.4 構造推定部

構造推定部では、分類部で得られた各クラスタに対して、繰り返し構造のラベル(Aメロ、Bメロ、サビなど)を推定する。クラスタID系列に適合する最適なラベル付けを行うために、考え得る全てのラベル付けパターンから、最適なラベル系列を得る。本研究では、ラベル系列の確率モデルとしてbi-gramモデルを仮定した。ラベル系列  $w_1, w_2, \dots, w_m$  の生成確率は式(2)で示される。

$$P(w_1^m) = \prod_{i=1}^m P(w_i | w_{i-1}) \quad (2)$$

<sup>†</sup> 電気通信大学大学院 情報理工学研究所

式(2)が最大となるラベル系列を求め、歌詞テキストに繰り返し区間を表すXMLタグを記述する。

### 3. 実験と評価

実験データとして、RWC研究用音楽データベース[2] (ポピュラー音楽)のうち、RWC-MDB-P-2001 No. 1 ~ No. 50 の50曲の歌詞を使用した。この50曲すべてについて、人手で正解ラベルを付与した。主観による評価の揺れを避けるため、AIST Annotation for RWC Music Database[3]に記載されているアノテーションを基準として正解ラベルを付与した。本実験で用いたラベルの種類を表1に、実験システムの出力例を図2に示す。

表1: 提案手法で用いた繰り返し構造のラベル

ラベルの種類	ラベルの意味
A1, A2	Aメロの1種類目、2種類目
B1, B2	Bメロの1種類目、2種類目
C1, C2	Cメロの1種類目、2種類目
S1, S2, S3	サビの1種類目 ~ 3種類目

<p>&lt;A1&gt; いつでも近すぎて ごまかせないため息 さり気ない仕草に めくもりを感じていた&lt;/A1&gt; &lt;B1&gt; 君に届くなら 不思議なチカラ 勇気に変えて 今 星の向こうへ飛んでいこう&lt;/B1&gt; &lt;S1&gt; I just fly away ふたりなら あざやかに輝いて I just fly away ためらわず 終わらない 光の中へ&lt;/S1&gt; &lt;A1&gt; 特別な気持ちを 感じあえる瞳は 憎らしいくらいに 思わせぶりの視線&lt;/A1&gt;</p>
----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

図2: 実験システムの出力例

繰り返し構造推定の正解率を求めるには、正解ラベル系列  $V = \{v_1, \dots, v_s, \dots, v_S\}$  とシステムが出力したラベル系列  $W = \{w_1, \dots, w_t, \dots, w_T\}$  とのDPマッチングにより、累積距離  $g(S, T)$  を計算する。ここで、 $g(s, t)$  は以下で与えられる。

$$g(s, t) = \min \left\{ \begin{array}{l} g(s-1, t) + 0.5 \\ g(s-1, t-1) + d(s, t) \\ g(s, t-1) + 0.5 \end{array} \right\} \quad (3)$$

$d(s, t)$  は  $v_s = w_t$  のとき 0,  $v_s \neq w_t$  のとき 1 で与えられる。以上より、繰り返し構造の正解率は次により求められる。

$$\text{accuracy} = \frac{S - g(S, T)}{S} \quad (4)$$

式(4)の正解率が 0.75 以上となった曲は、繰り返し構造推定が正しく行われたとして、実験に使用した50曲について評価を行ったところ、26曲で正しく推定できた。

不正解となる原因を分析するため、分類部のクラスタリング精度について考察する。クラスタリングにより、各ブロックの特徴ベクトル  $X$  にはクラスタ ID が割り当てられている。このクラスタ ID を曲の開始から並べた順列と、あらかじめ人手でクラスタリングを行い付与したクラスタ ID の順列の間で、ハミング距離  $D_{hamm}$  を算出する。上記2つのクラスタ ID 系列の長さを  $Len$  とするとき、クラスタリングの評価値  $C$  は次の式(5)で与えられる。

$$C = 1 - \frac{D_{hamm}}{Len} \quad (5)$$

ここで、正解曲および不正解曲に対する、正解率  $accuracy$  の平均とクラスタリングの評価値  $C$  の比較結果を表2に示す。

表2: システムの正解率とクラスタリング評価値の比較

	正解の曲	不正解の曲
正解率 $accuracy$ の平均	0.84	0.53
クラスタリングの評価値 $C$ の平均	0.88	0.5

提案システムの出力が正解となった曲と不正解となった曲との間には、正解率の平均だけでなく、クラスタリングの評価値にも大きな差が存在し、正の相関が伺える。このことから、システム全体の正解率を下げる原因は、クラスタリングの誤りによるものと考えられる。

ここで、音楽要約や音楽検索において有用となるサビの検出精度について考えてみると、後藤[1]の手法では、80% (100曲中80曲で正解) の精度でサビ区間を正しく検出できたのに対して、提案手法では74% (50曲中37曲で正解) と同等の精度を得ていることが分かる。

### 4. 今後の課題

今後の課題として、分類部において階層的クラスタリングを行う際に、クラスタ数が適切となるようにクラスタの集約を一定数で停止し、クラスタリングの精度を高める必要がある。また、特徴抽出部では、特徴量としてモーラ数のみを用いたが、文節数や形態素数など、より詳細な特徴を採用することが考えられる。さらに、空白行の存在しない歌詞については、頻出系列パターンマイニング手法を導入することで対処できると考える。

### 参考文献

- [1] 後藤真孝: SmartMusicIOSK: サビ出し機能付き音楽試聴機, 情報処理学会論文誌, vol. 44, No. 11, pp. 2737-2747 (2003).
- [2] 後藤真孝, 橋口博樹, 西村拓一, 岡隆一: RWC研究用音楽データベース: ポピュラー音楽データベースと著作権切れ音楽データベース, 情報処理学会音楽情報科学研究会, 2001-MUS-42-6, pp. 35-42 (2001).
- [3] M. Goto: AIST Annotation for the RWC Music Database, Proc. of ISMIR, Victoria, Canada (2006).