

音楽音響信号を対象とする GTTM 的アプローチによる グルーピング構造の抽出について

澤田 隼^{1,a)} 竹川 佳成^{2,b)} 平田 圭二^{2,c)}

概要：本論文では、音楽理論 GTTM に基づいて音楽音響信号からグルーピング階層構造を獲得する方法について述べる。GTTM 規則を音響信号（スペクトログラム）に直接適用することで、多数の楽曲演奏に対して高次の楽曲構造分析や楽曲類似度計算などが可能となる。これより、例えばさらに能動的な音楽鑑賞の実現が期待される。しかし、GTTM 規則の適用に関して、順序が曖昧であったり競合の発生により高い分析精度を達成するのは難しかった。本論文では、音楽音響信号のスペクトログラムのテクスチャ特徴量を使って階層的クラスタリングを行うことで各規則の変換と競合の解決をはかる。提案手法に対して RWC 研究用音楽データベース中のポピュラー音楽 100 曲を対象に評価実験を行ったところ、平均正答率は 85.5 % となった。

Grouping Structures Extraction by GTTM Approach for Musical Signal

SAWADA SHUN^{1,a)} TAKEGAWA YOSHINARI^{2,b)} HIRATA KEIJI^{2,c)}

1. はじめに

従来の研究では、音楽音響信号に含まれる各音源を分離し、Musical Instrument Digital Interface (MIDI) などの記号に変換する技術が存在し、その分離した記号に Generative Theory of Tonal Music (GTTM) を適用し分析するのが標準的である。従来の音楽音響信号の分析では、楽譜に記されている明示的な情報（表層構造）に相当する情報を抽出することに主眼が置かれている（音符、声部、演奏記号、拍節、楽曲構成、コード名など）[1]。しかし応用（編曲、演奏表情付けなど）によってはそのような表層構造に限らず、それらを生成する認知的な情報（深層構造）を理解する必要がある。音楽音響信号に GTTM を適用することができれば、フレーズ単位、小節単位の深層構造を分析することが可能になろう。

一方で、音楽音響信号全体を対象として音楽の構造上の境界を抽出することに主眼が置かれている研究は多い。Foote の Self-Similarity Matrix[2] や後藤ら [3] は音響的な類似度を用いて楽曲の構造分析と表示を可能にした。近年、この課題を解決するために音色的な特徴量と調和的な特徴量を組み合わせて k-means を用いる方法 [4] や、ニューラルネットを用いた方法も提案されている [5]。これらの研究の多くは、イントロ、サビ、A メロ、といった大きなレベルのグルーピングであり、聴取する楽曲全体の階層的な構造まで分析はしていない。

楽譜に書かれた楽曲を対象として構造や意味を分析する手法である GTTM が 1983 年に Fred Lerdahl と Ray Jackendoff によって提案された [6]。これはグルーピング階層構造と拍節構造に基づいて音イベントのゲシュタルトに基づいて生成される階層構造を抽出する分析手法である。GTTM の分析の結果得られるタイムスパン木は、旋律中の各音符の重要度を二分木で表すことができる。これは楽曲の構造の記述 [7] にとどまらず、楽曲の構造の操作を可能にするものであった。これまでに、楽曲要約システム [8]、2 つのメロディの内挿となるメロディを生成するメ

¹ 公立はこだて未来大学大学院
Graduate School of Future University Hakodate

² 公立はこだて未来大学
Future University Hakodate

a) g2116022@fun.ac.jp

b) yoshi@fun.ac.jp

c) hirata@fun.ac.jp

ロディーモーフィング [9] など広く応用されている。

GTTM を計算機上へ実装する試みとして、GTTM 分析を自動で適用するシステム、Automatic Time-span Tree Analyzer (ATTA) が浜中らによって開発された [10]。このシステムにより、楽譜に書かれている楽曲の階層構造を自動で獲得することが可能になった。GTTM の規則の適用においてしばしば競合が起こるが、ATTA ではそれらを制御するために、各ルールに重み付けを行うパラメータを導入することで解決をはかった。

本研究の目的は楽譜を対象としている GTTM 分析のアプローチにのっとり、スペクトログラムから直接グルーピング階層構造を獲得することである。スペクトログラムを時間軸方向に短冊状に分割し (bin) その bin から画像としてのテクスチャ特徴量を抽出し、階層的クラスタリングによってグルーピング階層構造を獲得する。これにより、ルールの競合の根本的な解決をはかる。

2. GTTM 分析のスペクトログラムへの適用

本システムは、特徴抽出器とグルーピング階層構造分析器からなる (図 1)。特徴抽出器ではスペクトログラムのテクスチャ特徴量を抽出する。グルーピング階層構造分析器は特徴抽出器により出力された特徴量を入力とし、拍節構造によって分割された bin に対してクラスタリングを行うことで、グルーピング階層構造を抽出する。

2.1 GTTM のグルーピング規則

GTTM のグルーピング構造分析は連続したメロディをより短いフレーズなどに階層的に分割することであり、グルーピング構造分析の規則は構成規則 (Well-formedness Rules) と選好規則 (Preference Rules) の 2 種類からなる。構成規則は、楽曲構造を生成する上で必要な条件や制約であり、可能性のある楽曲構造をすべて生成する。選好規則は、複数の構造が構成規則を満たす場合どれが好ましいかを示す規則である。

構成規則の例としては、あるグループが他のグループ (サブグループ) を含むときには、必ずサブグループの全てが含まれていなければならないという規則や、グループ同士が重なってはならないなどの規則がある。選好規則はどのようなところにグループの境界があるか (聴取した際に区切りとして聞こえるか) という規則が示されており、その規則によく当てはまる場合は、高レベルのグループの境界として分割される。例えば 4 つの連続した音を考えたときに、2 つめと 3 つめの間の音高や音量、音の長さ、アーティキュレーションが変わっていると、そこがグループの境界になる可能性があるなどの規則がある。

2.2 スペクトログラムのテクスチャ解析

音楽音響信号の音高や音量、アーティキュレーション、

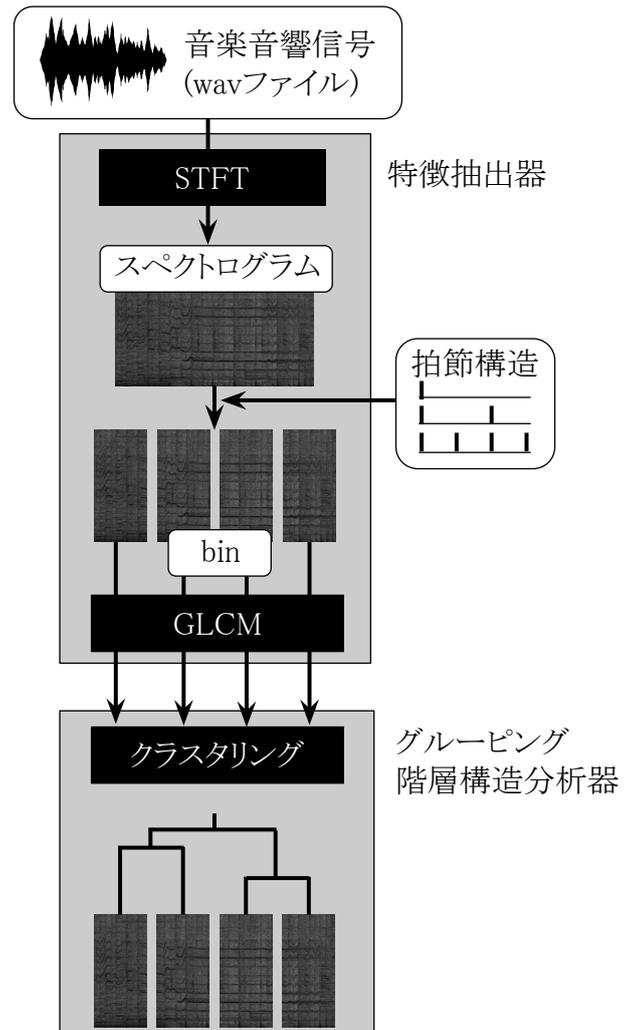


図 1 システム構成図

音の長さの違いはスペクトログラム上でテクスチャの違いとして現れる (図 2)。代表的なテクスチャ特徴量に R. M. Haralick らによって提案されたグレーレベル同時生起行列 (GLCM) がある [11]。これは、2 つの画素の対が画像の中で発生する頻度をあらわした行列であり、この行列から統計情報を抽出することによってテクスチャの特徴を記述することができる。この特徴量を用いて階層的クラスタリングをすることによって、テクスチャが異なる場所でグループの境界になり、テクスチャが大きく異なる場合は高レベルのグループの境界として分割されるため、GTTM の選好規則を満たすグルーピング構造が獲得できる。

2.3 特徴抽出器

特徴抽出器では入力した音楽音響信号から周波数スペクトルを抽出し、それをもとに出力されたスペクトログラムからテクスチャ特徴量であるグレーレベル同時生起行列 (GLCM) を抽出する。

まず、A/D 変換された音響信号に対してハニング窓を

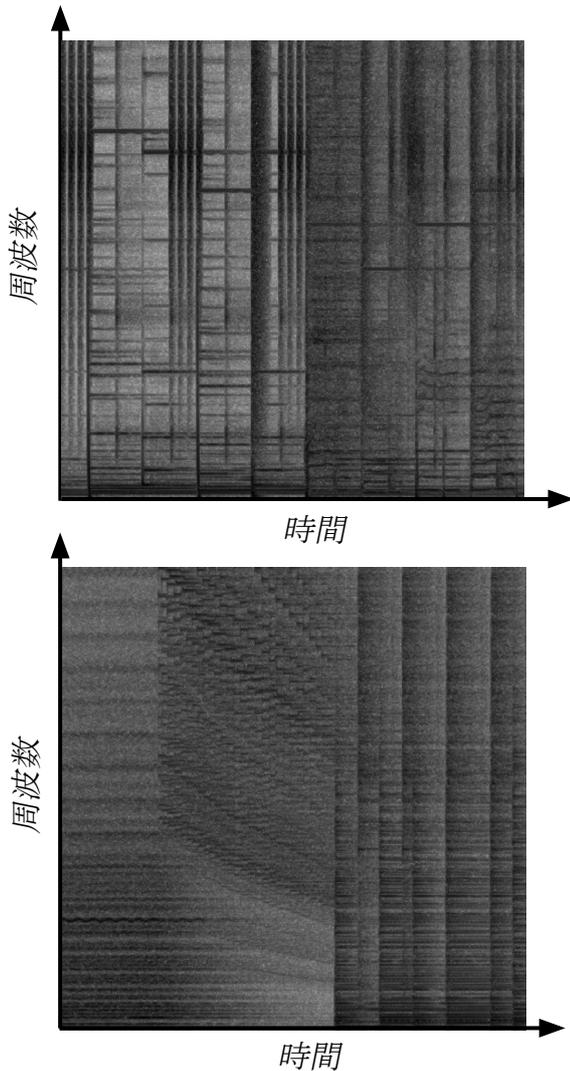


図 2 GLCM を用いるためにグレースケール化したスペクトログラム上のテクスチャの違い

用いた短時間フーリエ変換 (STFT) をし、周波数スペクトルを得る。STFT は観測区間を時間軸方向に単位時間ずつずらしながら適用する。本システムのウィンドウ幅は 1024、シフト幅は 441 点である。次にスペクトログラムを 256 階調のグレースケールで描画する。スペクトログラムは STFT の 1 サンプルを 1 ピクセルとして描画し、縦 512 ピクセル、横が曲の長さとなっている。このスペクトログラムから GLCM を求める。

GLCM は 2 つの画素の対 (図 3 上) が画像の中で発生する頻度をあらわした行列であり (図 3 下)、今回は距離 1 ピクセルとなる画素の対で、水平方向 (0°)、斜め方向 (45° , 135°)、垂直方向 (90°) の各 4 方向毎に作成した。それらをひとつの行列に足し合わせて 0~1 に正規化したものを使用した。この行列を使えば、例えばこの行列の対角成分は隣接する画素が同じであることを意味するが、対角成分から離れば離れるほど画素の差が大きいことを意

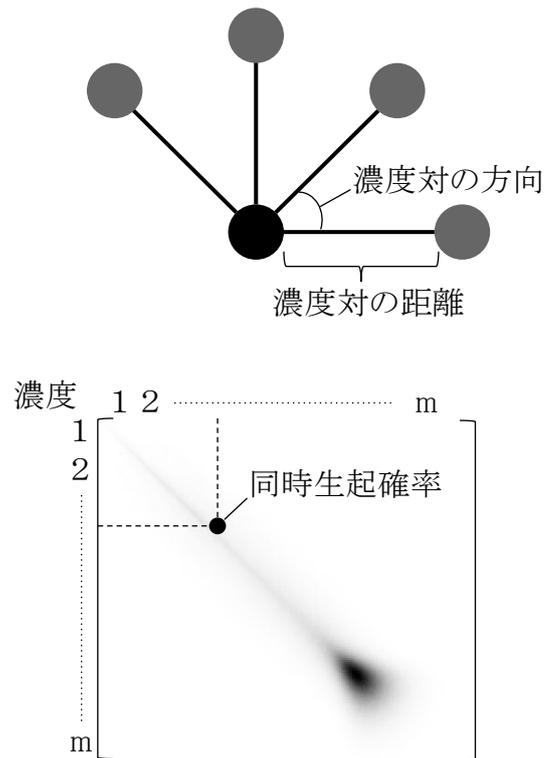


図 3 濃度対と GLCM

味する。つまり対角成分から離れた位置に多く分布していると、その画像のコントラストが高いことがわかる。

GLCM が求められたら、そこから角 2 次モーメント (均一性)、コントラスト、平均、逆差分モーメント、標準偏差、エントロピーの 6 つの特徴量を算出し、それらを 6 次元の特徴ベクトルとして用いる。これらの特徴量のみを用いてクラスタリングを行うと、時系列に関係なくテクスチャの類似した bin 同士がクラスタリングされる。これは似たようなフレーズを抽出する際には有効であるが、今回は隣り合う bin 以外にクラスタを形成させないために、bin の時刻も特徴量として持たせた。

2.4 グルーピング階層構造分析器

グルーピング階層構造分析器では、特徴抽出器により抽出された 7 次元の特徴ベクトル (テクスチャ特徴量 6 次元 + bin 時刻) を用いてクラスタリングを行う。

GTTM のグルーピング構成規則によると、全てのピッチイベントに対してグルーピングが行われる必要がある。これはグルーピングの最小単位がピッチイベントであることを意味する。音楽音響信号を全体を対象とする本手法では、ピッチイベントが存在しないため拍節構造によって分割されたスペクトログラムの 1 拍分の bin をグルーピングの最小単位とした。さらに、GTTM のグルーピング構造は階層的であるため、グルーピング階層構造の獲得を階層的クラスタリング問題として解き、デンドログラムを出力

した (図 4)。

階層的クラスタリングはあらかじめクラスタ数を決める必要がない。各 bin がどのように結合していくかの過程を知ることができ、分割する階層によってクラスタ数を自由に変えることができる。さらに高さがクラスタ間の距離、つまり類似度をあらわしているため、楽曲の構造を知るための重要な手がかりになると考えられる。今回は bin 間の特徴量ベクトルのユークリッド距離を算出し、完全連結法によるクラスタリングを用いた。

3. 実験システムの評価

3.1 データセット

音楽情報処理の分野で共通楽曲データベースとして使われている、Real World Computing Music Database (RWCMDDB) 内のポピュラーソング 100 曲を対象に、グルーピング階層構造の獲得を試みた。正解データとして、後藤によって提供されている AIST Annotation for RWCMDDB を用いる。正解データにはグループの境界の時刻と、それに対応したラベル (intro, chorus, bridge, verse, ending) がついている。これらのラベルに対応したグループを正解クラスタとする。また、我々のグルーピング階層構造分析器は、初めに拍節構造を入力として必要とする。今回は AIST Annotation for RWCMDDB の拍節構造を入力として与えた。

3.2 評価

グルーピング階層構造分析器が出力した特定の階層でのクラスタが、正解データと一致するかどうかを評価する。正解データのクラスタと同じクラスタに含まれる拍の数がシステムの出力に含まれている割合を正答率とする。クラスタ数は曲毎に変え、最も正答率が高いクラスタ数で分割した。各曲毎の正答率の結果を以下に示す (表 1)。

表 1 RWC 研究用音楽データベース各曲の正答率 (No.1~10 のみ)

No.	正答率 (%)
1	97.4
2	95.6
3	97.6
4	96.2
5	1.8
6	95.1
7	1.3
8	96.2
9	93.7
10	96.5

グルーピング構造が適切に獲得できている楽曲の大半は正答率が 95 % を超えている。反対に、うまく獲得できていない楽曲の正答率は 5 % を下回っていた。全体の平均正答率は 85.5 % となった。

4. おわりに

本研究では、音楽音響信号を対象にスペクトログラムのテクスチャ特徴量を用いてグルーピング階層構造の獲得を行った。グルーピング階層構造の獲得を階層的クラスタリング問題として解き、デンドログラムを出力した。このデンドログラムによって、各 bin がどのように結合していくかの過程を知ることができ、さらに高さがクラスタ間の距離をあらわしているため、楽曲の構造を知るための重要な手がかりになると考えられる。RWCMDDB 中のポピュラーソング 100 曲を対象に評価を行ったところ、平均正答率は 85.5 % となった。したがって、提案手法は特定の階層で正解データのグルーピングと一致することが確認できた。

未検討の課題を 3 つあげる。第一に、クラスタリングに関しての課題がある。今回はユークリッド距離を算出し、完全連結法によるクラスタリングを用いたが、その他のクラスタリング手法と比較する必要がある。さらに、クラスタ間の距離をどのように定義するか、何を用いるかも検証しなければならない。第二に、評価方法の課題がある。今回は楽曲毎に正答率が最大となるクラスタ数に分割したが、応用によっては正解データが無い楽曲に対してもクラスタをいくつに分けるべきかを自動で判別することができる必要がある。さらに、正解データよりも下位・上位のグルーピング構造が正しいかどうかの証明にはなっていない。第三に、特徴量に関しての課題がある。今回はスペクトログラムのテクスチャ特徴量のみを用いたが、本来は音楽音響信号としての音楽的、音響的な特徴量 (クロマベクトル、MFCC, Self-similarity matrix など) と組み合わせるべきであり、そうすることによってより音楽構造を反映したグルーピング階層構造が獲得できると考えられる。テクスチャ特徴量として用いた GLCM も、濃度対の距離などのパラメータの調節や、GLCM から抽出される特徴量の選択も課題である。提案手法ではスペクトログラムを画像として扱っているため、そのテクスチャ特徴量はスペクトログラムの描画に左右される。例えば、周波数軸を対数尺度やメル尺度、バーク尺度に変換することで、より人間の聴覚特性を反映した結果が得られる可能性があると考えている。

謝辞

本研究は JSPS 科研費 26280089, 16H01744 の助成を受けたものです。

参考文献

- [1] 尾島優太, 中村栄太, 糸山克寿, 吉井和佳: コード進行と多重音スペクトルの階層ベイズモデルに基づく音楽音響信号の音高推定, 情報処理学会全国大会講演論文集, 3Q-06, 2016.
- [2] J. Foote: Visualizing Music and Audio using Self Simi-

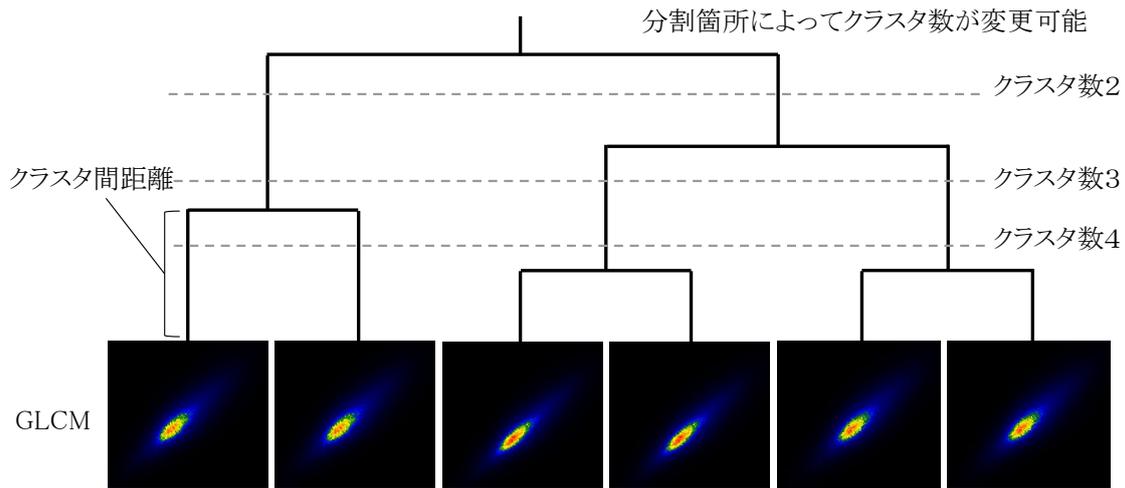


図 4 デンドログラムと GLCM

- larity, ACM Multimedia, pp. 77-80, 1999.
- [3] R. B. Dannenberg, M. Goto : Music Structure Analysis from Acoustic Signals, In D. Havelock, S. Kuwano, M. Vorländer, editors, Handbook of Signal Processing in Acoustics, pp. 305-331, 2008.
 - [4] R. Chen, M. Li : Music Structure Segmentation By Combining Harmonic and Timbral Information, In Proc. ISMIR, pp. 477-482, 2011.
 - [5] K. Ullrich, J. Schlüter and T. Grill : Boundary Detection in Music Structure Analysis using Convolutional Neural Networks, In Proc. ISMIR, Taipei, Taiwan, 2014.
 - [6] F. Lerdahl, R. Jackendoff : A Generative Theory of Tonal Music, The MIT Press, 1983.
 - [7] K. Hirata, T. Aoyanagi : Computational Music Representation Based on the Generative Theory of Tonal Music and the Deductive Object-Oriented Database, Computer Music Journal, Vol. 27, No. 3, pp. 73-89, 2003.
 - [8] K. Hirata, S. Matsuda : Interactive Music Summarization based on Generative Theory of Tonal Music, Journal of New Music Research, Vol. 35, No. 2, pp. 165-177, 2003.
 - [9] M. Hamanaka, K. Hirata, K. Tojo : Melody Morphing Method Based on GTTM, In Proc. International Computer Music Conference, pp. 155-158, 2008.
 - [10] M. Hamanaka, K. Hirata, S. Tojo : ATTA : Automatic Time-Span Analyzer Based On Extended GTTM, In Proc. ISMIR, pp. 358-365, 2005.
 - [11] R. M. Haralick : Statistical and structural approaches to texture, In Proc. IEEE, vol. 67, pp. 786-804, 1979.

付 録

A.1 付録

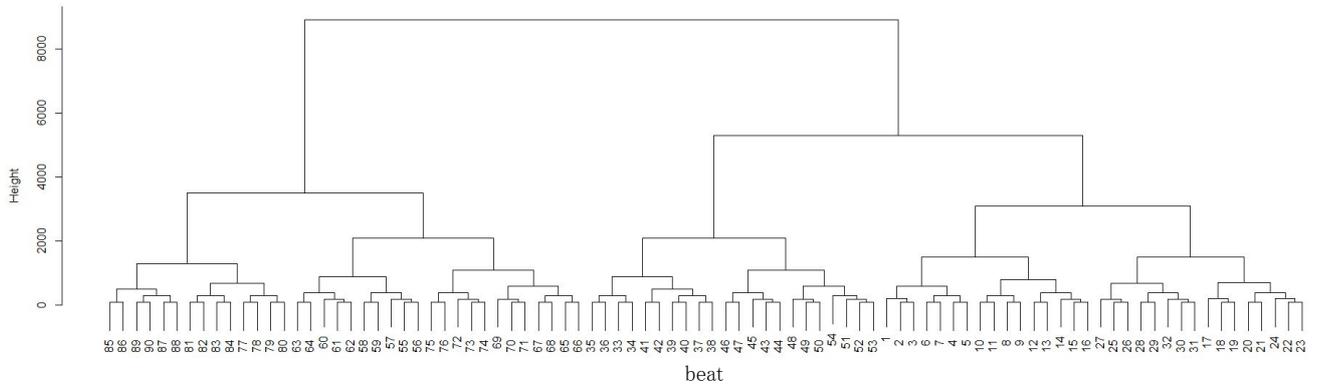


図 A.1 RWCMDDB 中の楽曲 No.1 : 永遠のレプリカのデンドログラム

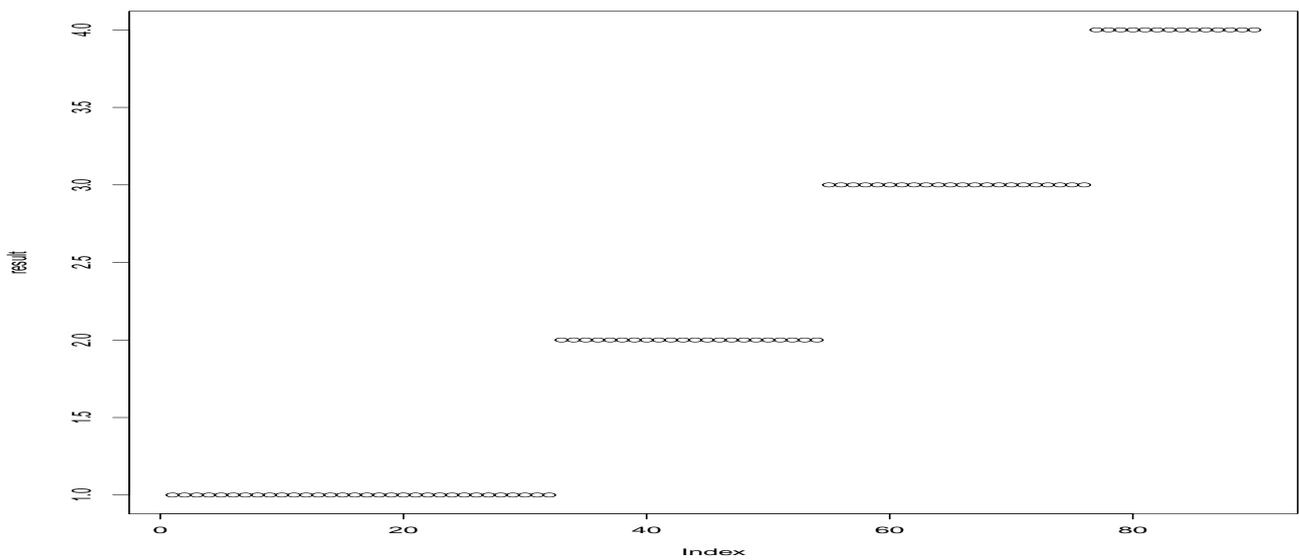


図 A.2 RWCMDDB 中の楽曲 No.1 : 永遠のレプリカのクラスタリング