

# スペクトログラムの階層的クラスタリングを用いた グルーピング構造分析について

澤田 隼<sup>1,a)</sup> 竹川 佳成<sup>2,b)</sup> 平田 圭二<sup>2,c)</sup>

概要：本論文では、音楽音響信号を対象とした GTTM に基づくグルーピング構造分析について述べる。楽譜を対象とした GTTM 規則を音響信号に直接適用することができれば、楽譜のない楽曲や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。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用いて階層的クラスタリングを行うことで、グルーピング構造の獲得を試みたが、GTTM のグルーピング構造分析と十分に比較、考察がなされていなかった。本論文では、スペクトログラムの階層的クラスタリングを用いたグルーピング構造分析の結果を GTTM グルーピング構造分析結果と比較し、その意味を考察する。

## 1. はじめに

近年、大量の音楽がデジタル化され、音楽音響信号を対象とした音楽検索や楽曲推薦などのシステムの需要が高まっている。それに伴い、編曲や演奏表情付けなど楽曲を深く理解して楽しめるようなシステムの需要も増加している。しかし、これらのシステムを実現するためには楽曲の構造や意味を分析する必要がある。本研究では音楽音響信号を対象とした楽曲の構造を分析することを目指す。

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 [1]。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 [2] 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 [3] や 2 つのメロディの内挿となるメロディを生成するメロディーモーフィング [4] など広く応用されている。GTTM を計算機上に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 [5]。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。

音楽音響信号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 [6] や後藤ら [7] 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 k-means を用いる方法 [8] や、ニューラルネットを用いた方法も提案されている [9]。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、人間が音楽を聴取する際の認知的な階層構造まで分析していない。

最近では、音楽音響信号をスペクトログラムに変換し画像処理の分野で培われてきた技術を適用する研究が注目されている。中鹿ら [10] や Yandre ら [11] はスペクトログラムの画像特徴量を用いて楽曲のジャンル分類を行っている。スペクトログラム上に現れるジャンル毎のテクスチャに着目し、音響特徴量である Mel-Frequency Cepstrum Coefficients (MFCC) を用いるよりも高精度でジャンルの分類を実現した。

我々はこれまで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用いた階層的クラスタリングにより、楽譜を対象としている GTTM 分析のアプローチに則ったグルーピング構造の獲得を試みた [12]。しかし、提案手法の分析結果と GTTM のグルーピング構造の分析結果を十分に比較、考察がなされていなかった。本論文では、スペクトログラムのテクスチャ特徴量を用いたグルーピング構造分析の結果を GTTM の結果と比較し、その意味を考察

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

<sup>2</sup> 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

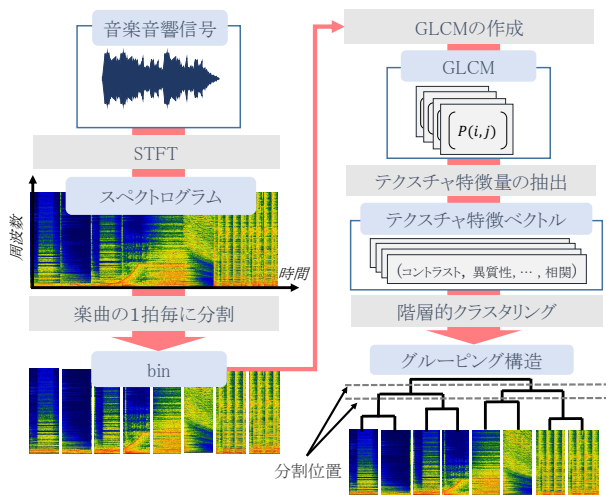


図 1 システム構成図

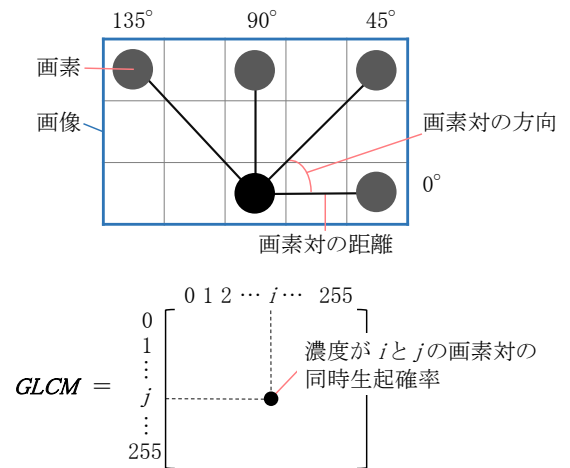


図 2 画像中の画素対の距離と方向 (上) と GLCM (下)

する。

## 2. スペクトログラムのグルーピング構造分析

本システムは、音楽音響信号を入力とし、グルーピング構造を出力する (図 1)。まず、音響信号を短時間フーリエ変換 (Short-time Fourier transform : STFT) し、スペクトログラムを描画する。その後、拍でスペクトログラムを分割し、その分割された短冊 (bin) 毎にテクスチャ特徴量を抽出する。各 bin 毎に濃度共起行列を求め、それらの行列から統計的な二次特徴量を抽出する。このテクスチャ特徴量を特徴ベクトルとし、特徴空間で階層的クラスタリングを行い、画像空間 (スペクトログラム) に戻すことによってグルーピング階層構造を獲得する。

### 2.1 スペクトログラムのテクスチャ特徴量の抽出

#### 2.1.1 スペクトログラムの描画

音楽音響信号をスペクトログラムへ変換する方法について述べる。まず分析対象の音響信号が入力されると、STFT をしてパワースペクトルを求める。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 STFT のウィンドウ幅は 1024、シフト幅は 256 とした。横軸に時間軸、縦軸に周波数軸をとったスペクトログラムを 1 サンプル 1 ピクセルとなる様に 256 階調のグレースケールで描画する。周波数軸が線形なスペクトログラムは、音高が平行移動した場合にピクセル間の位置関係が異なってしまう。移調した際にも同じパターンが抽出できることが望ましいため、スペクトログラムの周波数軸は対数をとって描画する。こうして得られたスペクトログラムを拍で分割し、各 bin 毎にテクスチャ特徴量を抽出する。

#### 2.1.2 濃度共起行列

人間が音楽を聴取した際の音楽の構造的なグループの境界は、スペクトログラム上ではテクスチャの違いとして

現れる。代表的なテクスチャ特徴量に Haralick らによって提案されたグレーレベル同時生起行列 (Gray-Level Co-Occurrence Matrix : GLCM) がある [13]。これは、隣接する 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 (図 2 下)、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。GLCM は画素対の相対的な位置関係を記述するものであり、その画素対がどこに出現しているかという情報は捨象されている。そのため、時間軸方向や周波数軸方向への平行移動には頑健な特徴量である。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 GLCM を作成した (図 2 上)。GLCM の要素は生起確率となるように全ての要素の和が 1 になるよう 0~1 の値に正規化した。

#### 2.1.3 テクスチャ特徴量

Haralick らは GLCM からコントラストや異質性などの二次特徴量を計算しテクスチャを分類する方法を提案した。本システムでは、GLCM からコントラスト、異質性、均質性、角二次モーメント、相関の 5 つの二次特徴量を算出し、それぞれ平均 0、標準偏差を 1 に標準化したものを 5 次元の特徴ベクトルとした。それぞれの算出方法を式 (1)~式 (5) に示す。

$$\text{コントラスト (contrast)} : \sum_{i,j=0}^{255} (i-j)^2 P(i,j) \quad (1)$$

$$\text{異質性 (dissimilarity)} : \sum_{i,j=0}^{255} |i-j| P(i,j) \quad (2)$$

$$\text{均質性 (homogeneity)} : \sum_{i,j=0}^{255} \frac{P(i,j)}{1+(i-j)^2} \quad (3)$$

$$\text{角二次モーメント (ASM)} : \sum_{i,j=0}^{255} P(i,j)^2 \quad (4)$$

$$\text{相関 (correlation)} : \sum_{i,j=0}^{255} P(i,j) \frac{(i - \mu_i)(j - \mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \quad (5)$$

ここで  $P(i, j)$  は bin 中の画素の濃度が  $i$  と  $j$  の画素対の生起確率を表し、 $\mu_i$  は平均を表し、 $\sigma_i^2$  は分散を表す。

式 (1) と式 (2) は濃度対の画素の濃度差 (コントラスト) に関する特徴量であり、濃度差が増加すると式 (1) の値は指数関数的に増加し、式 (2) の値は線形に増加する。そのため、濃度差の高い画素対が多いほどこれらの特徴量の値は高くなる。式 (3) は GLCM の要素の分布が GLCM の対角線に対してどの程度接近しているのかを表す特徴量である。コントラストは濃度差が増加すると値が指数関数的に増加したのに対して、均質性は対角線から離れると指数関数的に値が減少する。式 (4) はテクスチャが整然としているときに最も高い値を示す。全ての画素が同じ値の時に最大値 1 をとる。式 (5) は画素対の相関を表す特徴量で、濃度の値が比例している程大きくなり、方向性の周期的なパターンを示している。

## 2.2 階層的クラスタリング

グルーピング階層構造を獲得するために、各 bin 毎に得られた特徴ベクトルを用いて階層的クラスタリングを行う。テクスチャ特徴量のみの 5 次元特徴ベクトル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。これは類似した構造やフレーズを抽出する際には有効であるが、今回は時系列で隣り合う bin 同士でクラスタを形成させるために、bin の時刻を表すパラメータを特徴ベクトルに加え、合計 6 次元特徴ベクトルを用いてクラスタリングを行う。

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 3)。階層的クラスタリングは、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定めることができる。さらに、デンドログラムの高さがクラスタ間の距離、つまりクラスタ同士の類似度を表しているため、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。完全連結法は併合されて新たにできたクラスタは、以後併合されにくくなる性質 (空間拡散) を持つ。これにより、クラスタ間のクラスタの大きさの分散が小さくなり、各クラスタの大きさが均等になる。

## 3. 提案手法と GTTM の比較

### 3.1 GTTM のグルーピング規則

GTTM のグルーピング規則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境界がどこに検出され

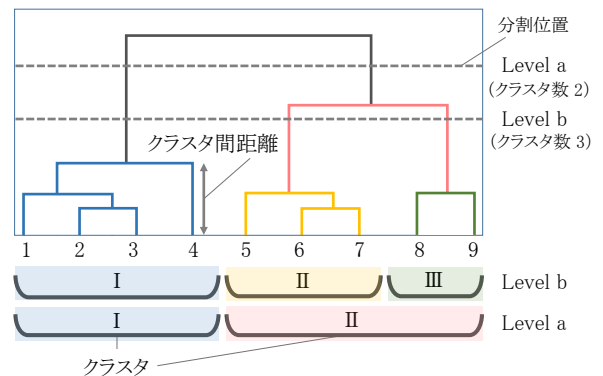


図 3 デンドログラムの分割位置と形成されるクラスタ

表 1 GTTM の選好規則の概要

規則	概要
GPR1	単音がグループとすることを避ける
GPR2	音符間の時間間隔が他より長いときに境界になりやすい
GPR3	音符間の変化が他より大きいときに境界になりやすい
GPR4	GPR2, 3 が良く当てはまる場合の境界は深くなる
GPR5	グループの分割は長さが等しくなることが望ましい
GPR6	類似箇所は同じようにグルーピングされるのが望ましい

るかを表した規則である。グルーピング選好規則 (Grouping Preference Rules : GPR) には局所的な規則 (GPR1, 2, 3) と大域的な規則 (GPR4, 5, 6) がある。GPR の概要を表 2 に示す。GPR2 は音符間の時間間隔に関する規則であり、時間間隔が他よりも長い場合にそこがグループの境界になりやすいという規則である。GPR3 は音符間の音高差や音量差、アーティキュレーションパターンの変化や音の長さの変化が他よりも大きい場合にそこがグループの境界になりやすいという規則である。

### 3.2 予備実験方法

GTTM グルーピング構造分析の結果とスペクトログラムのテクスチャ特徴量を用いた階層的クラスタリングによるグルーピング構造分析 (提案手法) の結果を比較し、その意味を考察するために 3 つの予備実験を行う。

#### 実験 1 : 提案手法と GTTM の局所的な GPR の比較

GTTM の局所的な GPR が適用される譜面を用意し、その規則が適用される位置に提案手法でも境界が検出されるかを確かめる。また、各 GPR とテクスチャ特徴量の対応付けや関係を検証する。

#### 実験 2 : 提案手法と GTTM のグルーピング構造の比較

GTTM のグルーピング構造が既知な楽曲に対して提案手法を適用し、GTTM のグルーピング構造と境界の位置や境界の深さ (順位) を比較する。また、提案手法の Monophony 楽曲と Polyphony 楽曲の分析結果

表 2 譜面に適用される規則とその適用理由

譜面	適用規則	適用理由
(a)	GPR2	3 番目の音符のみ 8 分音符で、その他は 4 分音符である
(b)	GPR2	3 番目の音符のみ 2 分音符で、その他は 4 分音符である
(c)	GPR3	3 番目の音符と 4 つ目の音符の間の音高差がその他の音符の音高差より大きい
(d)	GPR3	3 番目の音符までは弱く演奏され、4 番目からは強く演奏されている
(e)	GPR3	3 番目の音符までは滑らかに演奏され、4 番目からは 1 音 1 音を短く演奏されている
(f)	GPR3	3 番目の音符までは 2 分音符で、4 番目からは 4 分音符になる

を比較する。

**実験 3：音色がグルーピング境界に及ぼす影響について**  
 音響信号の音色の変化が提案手法の境界位置や境界の深さにどのような影響をもたらすかを調べる。

### 3.2.1 データセット

3 つの予備実験のために使用したデータについて述べる。各実験のために用意した譜面を実験 1 ではピアノの単旋律で演奏し、実験 2 ではピアノの単旋律と、オーケストラで演奏されたものを使用した。実験 3 ではエレキギターの単旋律で演奏し、それぞれの演奏をサンプリング周波数 44.1kHz でサンプリングした音響信号を用いた。これらの譜面の期待される境界位置やグルーピング構造は文献 [1] に示されているものを用いた。

実験 1 では GPR2 が適用される譜面 (図 4(a)) の音源と、GPR3 が適用される譜面 (図 4(b)) の音源を用いた。譜面上に示した赤い線の位置が GTTM の GPR によってひかれる境界の位置である。それぞれの譜面 (a) ~ (f) は 5 つの音符があり、3 番目と 4 番目の音符の間にグルーピングの境界がひかれることが期待されている。それぞれの譜面に適用される規則とその適用理由を表 2 に示す。分析に使用した音源のスペクトログラムはそれぞれ図 4 のようになっている。

実験 2 ではモーツァルトの交響曲第 40 番 K.550 の音源をピアノの単旋律で演奏した音源と、オーケストラで演奏された音源を用いた。実験に用いた譜面と期待されるグルーピング構造を図 5 上に示す。分析に使用したスペクトログラムはピアノの単旋律で演奏した音源とオーケストラで演奏された音源はそれぞれ図 5(a) 下、図 5(b) 下のようになっている。

実験 3 では 4 分音符の同じ音 (ド) が 8 つ続く譜面を用意し、4 番目の音符まではクリーンギターで、5 番目からはディストーションギターで弾いた音源を用いた。

## 3.3 結果

各実験毎にシステムの出力したデンドログラムとグルーピング構造を示し、境界の位置や境界の深さについて述べる。また、その結果が得られた時のテクスチャ特徴量の値や変化量をグラフで示す。

### 3.3.1 実験 1 の結果

実験 1 の譜面 (a), (b) のシステム出力を図 6-(a) に示す。図中に含まれている数字は何番目の拍か (bin 番号) を表しており、1-2 は 1 番目の bin と 2 番目の bin の間を意味する。システムが出力したデンドログラム (図 6-(a) 上) をもとに作成したグルーピング構造は図 6-(a) 下となった。GPR2 によって期待される境界の位置は 3-4, 9-10 であるが、システム出力の最も深い境界は 5-6 となり、次に深い境界は 8-9, 次いで 3-4 となった。また bin 毎の各テクスチャ特徴量の値 (図 6-(b) 上) と隣接する bin 間の距離 (図 6-(b) 下) のグラフを示す。9 拍目でコントラスト、異質性の値が減少し、均質性や角二次モーメント、相関の値は大きくなった。bin 間の距離は 3-4 と 5-6 にそれぞれ小さな山ができています。

実験 1 の譜面 (c) ~ (f) のシステム出力を図 7-(a) に示す。譜面 (c), (d), (e) では期待される境界位置 (3-4) に最も深い境界が検出された。譜面 (f) では 3-4 に最も深い境界が検出されたが、期待される境界位置 (6-7) ではなかった。期待される境界位置である 6-7 は 2 番目に深い境界として検出された。また bin 毎の各テクスチャ特徴量の値と隣接する bin 間の距離のグラフを図 7-(b), 図 7-(c) に示す。譜面 (c) では 3-4 でコントラストや異質性、角 2 次モーメントの距離が大きくなった。譜面 (d) ではコントラストや異質性などのテクスチャ特徴量の値が 4, 5 拍目で大きくなった。譜面 (e) では 3-4 で均質性や角 2 次モーメントの距離が大きくなった。譜面 (f) では均質性や角 2 次モーメントなどのテクスチャ特徴量の値が 1 拍毎に上下し、2 拍毎に周期的に変化している。

### 3.3.2 実験 2 の結果

実験 2 のピアノの単旋律の音源に対するシステム出力を図 8-(a) に示す。システムが出力したデンドログラム (図 8-(a) 上) をもとに作成したグルーピング構造は図 8-(a) 下となった。提案手法で得られた境界の位置と順位を GTTM と比較したものを表 3 に示す。提案手法では、7-8 に最も深い境界が検出された。次いで 4-5, 3-4 に深い境界が検出された。3, 4 番目や 7, 8 番目の bin は同じグループに属することが期待されているが、別々のグルーピングがされた。また bin 毎の各テクスチャ特徴量の値と隣接する bin 間の距離のグラフを図 9-(a) に示す。相関を除いたテクスチャ特徴量の 1~3 番目や 5~7 番目の隣接する bin 同士の距離が小さかった。

実験 2 のオーケストラ音源に対するシステム出力を図 8-(b) に示す。システムが出力したデンドログラム (図 8(b))

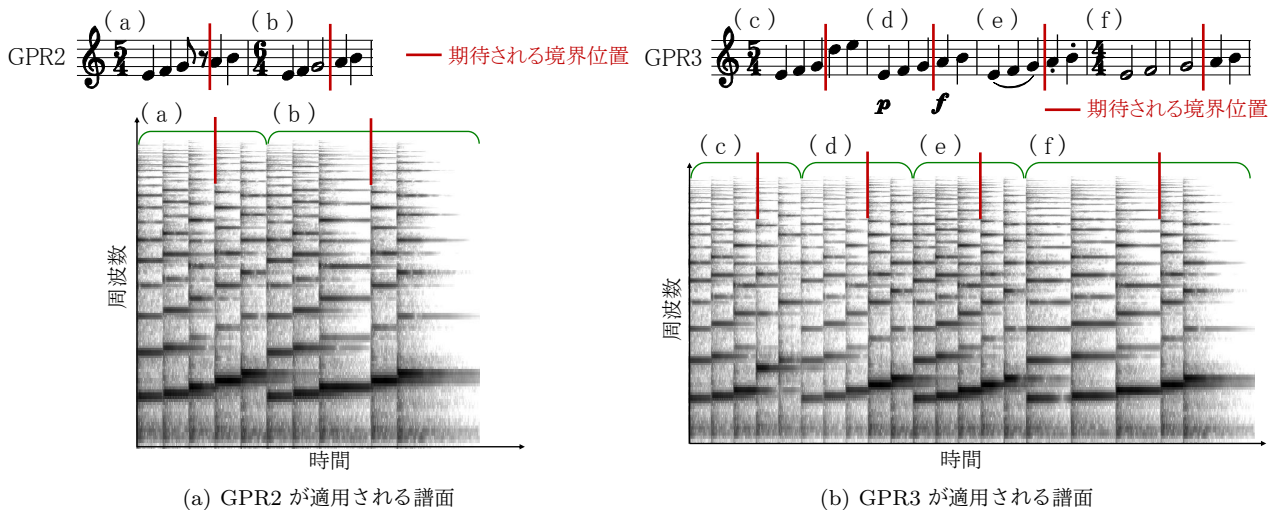


図 4 実験 1 で用いた譜面 (上) と分析に使ったスペクトログラム (下)

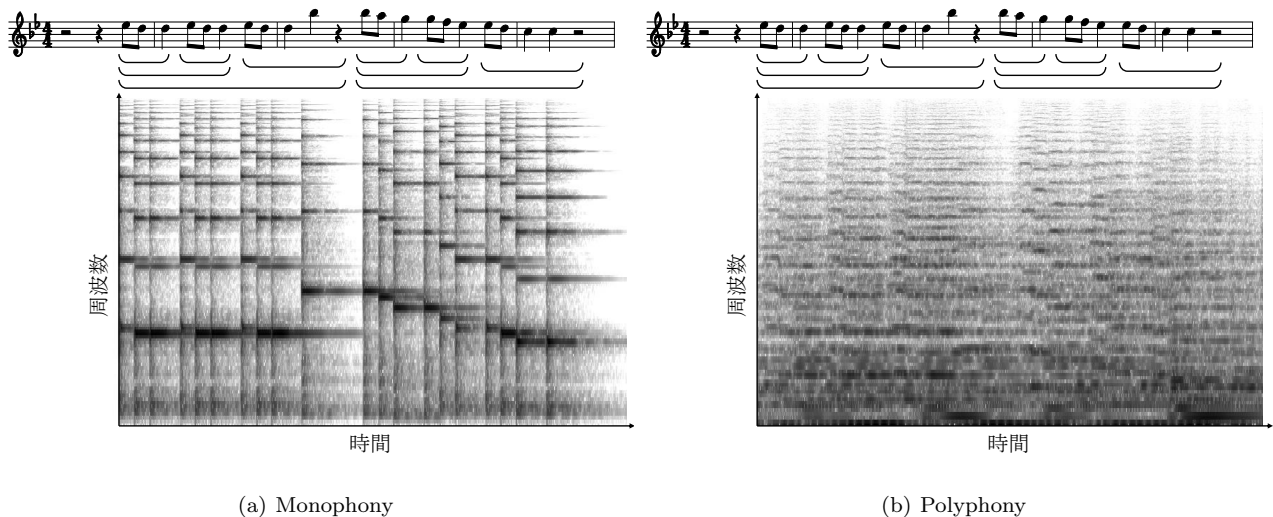


図 5 実験 2 で用いたモーツァルトの交響曲第 40 番 K.550 の譜面 (上) と分析に使ったスペクトログラム (下)

表 3 GTTM と提案手法 (Monophony 楽曲と Polyphony 楽曲) の分析結果の境界の位置と深さの順位の比較

深さ (順位)	GTTM	Monophony	Polyphony
1	4-5	7-8	3-4
2	2-3, 6-7	4-5	6-7
3	1-2, 5-6	3-4	4-5

上) をもとに作成したグルーピング構造は図 8-(b) 下となった. GTTM と提案手法で得られた Monophony 楽曲と Polyphony 楽曲の境界の位置と順位を比較したものを表 3 に示す. Monophony では 7-8 で最も深い境界が検出されたのに対して, Polyphony では 3-4 に最も深い境界が検出された. また bin 毎の各テキスト特徴量の値と隣接する bin 間の距離のグラフを図 9-(b) に示す. 同じ楽曲の Monophony (図 9-(a)) と比較すると, 対応する bin 同士でテキスト特徴量の変動も対応していた. 例え

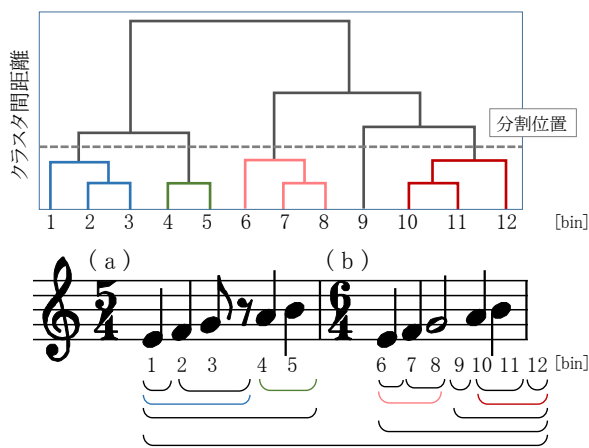
ば, Monophony 楽曲では 4 拍目でコントラストや異質性が減少し, 同質性や角 2 次モーメントが増加しているが, Polyphony 楽曲でも同様の変動を示している.

### 3.3.3 実験 3 の結果

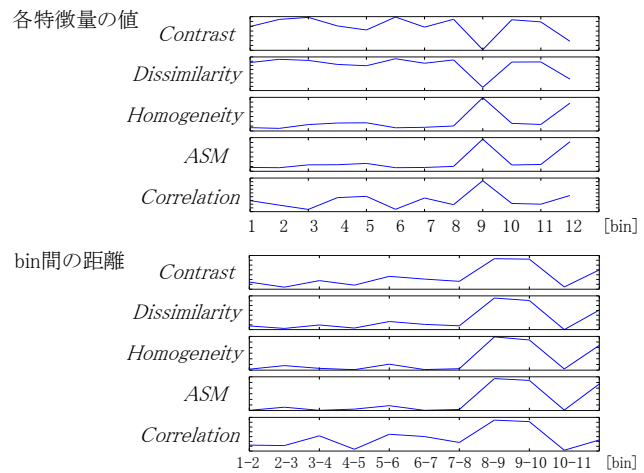
最後に実験 3 の譜面に対するシステム出力を図 10 に示す. 音色が変化する 4-5 に最も深い境界が検出された. 各テキスト特徴量も 4-5 を境に変わり, bin 間の距離は 4-5 にピークができ, その他は小さかった. また, 各クラスタの大きさも偏ることなく均等にグルーピングされた.

## 4. 考察

実験の結果より, 各 GTTM 規則がテキスト特徴量として記述できることが示唆された. ここでは, GPR の各規則がどのようなテキスト特徴量で記述することができるかや, 音楽的構造の類似性に関する有用性, 提案手法の限界や GTTM との違いについて考察する.

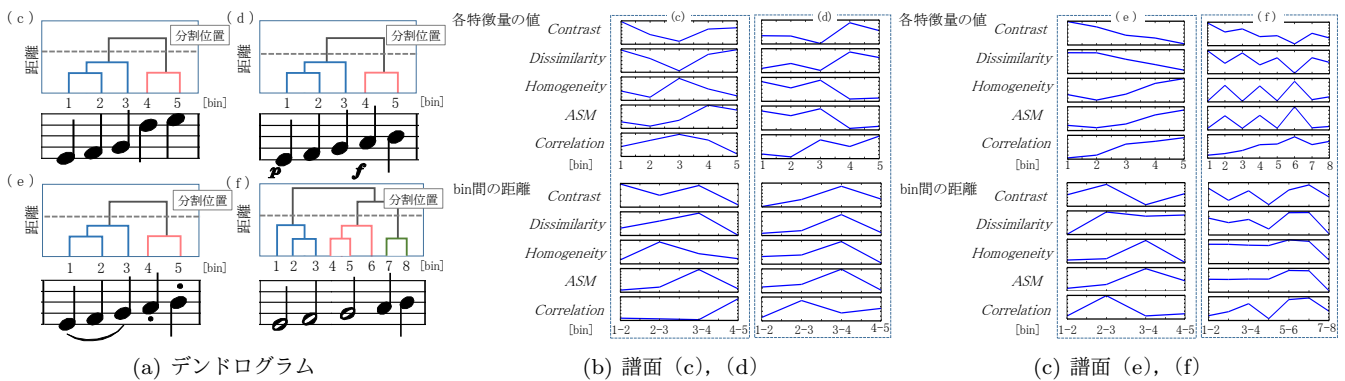


(a) デンドログラムとグルーピング構造



(b) bin 毎の各特徴量の変化

図 6 実験 1 の譜面 (a), (b) に対する分析結果



(a) デンドログラム

(b) 譜面 (c), (d)

(c) 譜面 (e), (f)

図 7 実験 1 の譜面 (c), (d), (e), (f) に対する分析結果

### 各 GPR のテクスチャ特徴量への変換

実験 1 の譜面 (b) や譜面 (f) より、休符や長音は均質性、各 2 次モーメント、相関の値が大きくなることが示唆された。スペクトルのパワーが小さいとテクスチャは均質になり、スペクトルが時間方向に徐々に減衰するとスペクトログラムの濃度は比例して減少する。そのため、相関の値が大きくなると考えられる。実際に、実験 2 の譜面に対して Monophony, Polyphony 楽曲のどちらも 4 番目の休符が含まれる bin で均質性、各 2 次モーメント、相関の値が増加している。

実験 1 の譜面 (c) より、音高差はコントラストや異質性、角 2 次モーメントの値が変化することが示唆された。これは前の音との相対的な差が抽出できているためと考えられる。しかし、GLCM は音の周波数方向の平行移動に対して頑健な特徴量のため、本来であれば、音高が変わっても同じ値になるはずである。これは前の bin の音の残響が分析対象の bin に影響を及ぼしているためであると考えられる。

実験 1 の譜面 (d) より、音量差はコントラストや異質性が変化することが示唆された。音量が変化するとスペク

トログラム上では濃度が変化する。例えば、ある音が強く鳴ると、他の音に比べてその音の周波数のパワーが強くなる。スペクトログラムの濃度は各周波数のパワーを表しているため、音量が変化するとコントラストや異質性の値が変化すると考えられる。実験 1 の譜面 (d) ではピアノからフォルテになることで、コントラストの値が増加しているが、反対にフォルテからピアノになるとコントラストの値が減少することが確認できた。

実験 1 の譜面 (e) より、アーティキュレーションは均質性や角 2 次モーメント、相関が変化することが示唆された。これも休符や長音と同様な理由で変化すると考えられる。例えば、1 音 1 音短く演奏することは短い休符が含まれていることと同義になる。

### 音楽的構造の類似性の抽出に関する有用性

実験 2 の結果より、音楽的構造の類似性をテクスチャの類似性として表現可能であることが示唆された。GLCM は周波数方向の平行移動に対して頑健な特徴量であるため、周波数軸方向に平行移動したフレーズのテクスチャ特徴量は同じパターンになる。実験 2 の Monophony, Polyphony 楽曲のどちらも 1~3 番目の bin と 5~7 番目の bin のテク

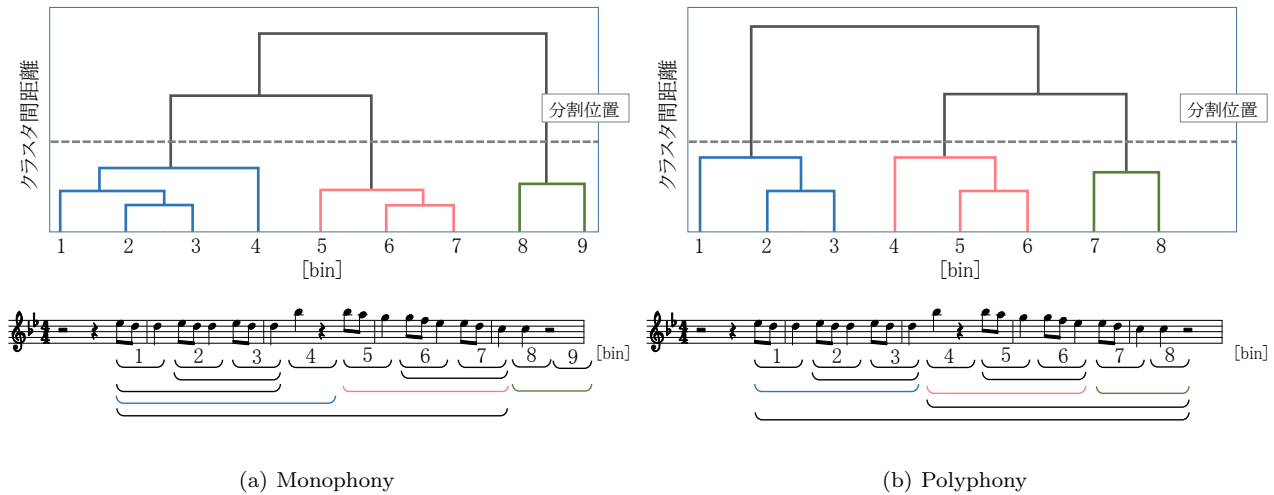


図 8 実験 2 のモーツァルトの交響曲第 40 番 K.550 に対する分析結果

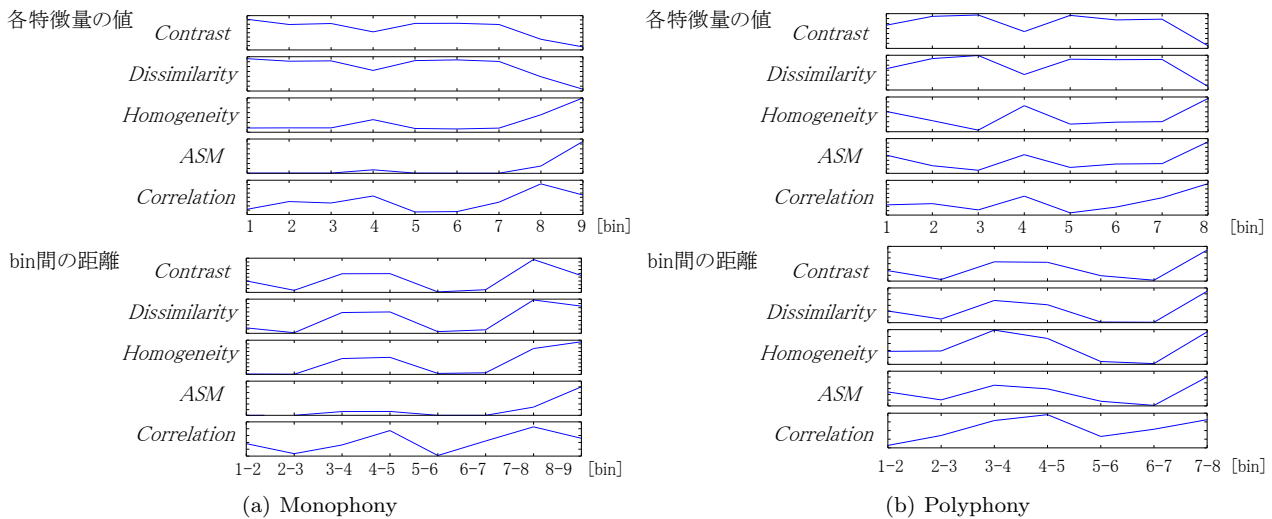


図 9 実験 2 のモーツァルトの交響曲第 40 番 K.550 に対する bin 毎の各特微量の変化

スチャ特徴量の値がほぼ同じになっている。これは同じフレーズの繰り返し構造になっているためだと考えられる。5~7 番目の bin は音高は異なるがそれぞれ平行移動すると一致する。これにより、提案手法は周波数方向に平行移動した音楽的構造の類似性をテクスチャの類似性として表現することが可能になった。

#### 提案手法の限界

単純なテクスチャの変化量だけに着目した場合、正しいグルーピング構造が得られない可能性がある。音の長さの違いは残響によってテクスチャが異なると述べた。例えば、2 分音符が続いていると 2 分音符内の 1 拍目と 2 拍目でテクスチャが大きく変化するため、2 分音符内の 1 拍目と 2 拍目の間に境界ができてしまう。正しくグルーピングを抽出するにはテクスチャ特徴量の変化の周期性に着目することや、音の連続性に着目し、スペクトログラムの分割幅を多段的に変更して抽出したテクスチャ特徴量を適切に統合する必要がある。

また、提案手法は GPR1 が踏襲できていないため、正しいグルーピング構造が得られない可能性がある。実験 2 の楽曲の 3 番目と 4 番目の bin は同じグループに属することが期待されている。それは GPR1 によって 4 番目の bin が単体でグループを作ることが禁止されているからである。しかし、提案手法では新しく作られたクラスタが併合されにくくなる性質はあるが、GPR1 をうまく踏襲できていないため、テクスチャが大きく変化するとそこに境界が検出されてしまう。

#### 提案手法と GTTM の違い

提案手法では音色の違いに大きく左右される。音響信号には音色の情報が含まれており、それはスペクトログラムのテクスチャの違いとして顕著に表れる。そのため、GTTM の GPR が成り立つ部分よりも音色の違いによるグルーピング境界が優先して検出される。人間が音楽を聴取した際、音色が変化するとそこがグループの境界だと強く感じる。提案手法はそういった人間の聴覚特性を反映でき

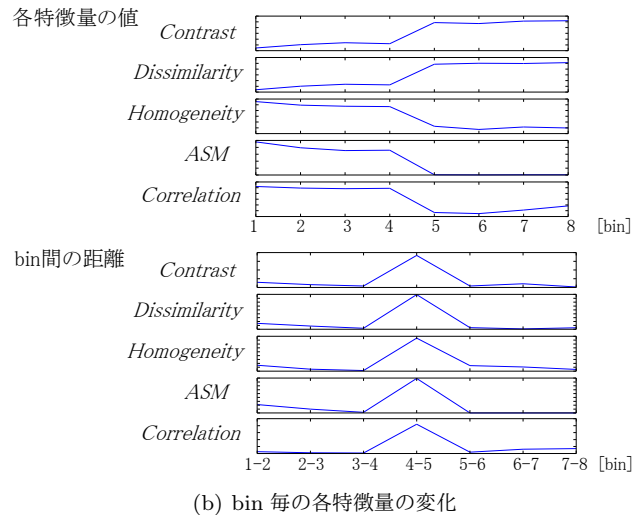
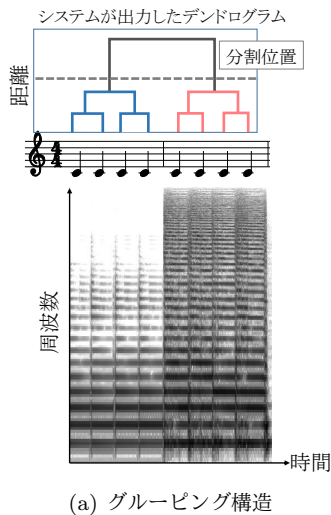


図 10 実験 3 の音色が変化する譜面と結果

ていると考えられる。

また、提案手法はスペクトログラムを拍で分割しているため、2分音符であるという情報やシンクレーションなどを捨象してしまっている。そのため、ミクロなレベルでのグルーピング構造は GTTM と一致しない部分もあるが、マクロなレベルでのグルーピング構造は GTTM と一致すると考えられる。

## 5. おわりに

本論文では、スペクトログラムのテクスチャ特徴量を用いた階層的クラスタリングによるグルーピング構造分析の結果を GTTM のグルーピング構造分析の結果と比較し、その意味を考察した。音楽音響信号を対象とした各グルーピング規則は、スペクトログラム上ではテクスチャの違いとしてあらわれることを主張した。また、音楽的構造が似ているところは同じテクスチャになり、周波数軸方向の平行移動に頑健であるため、音楽の構造を抽出する上で有効な特徴量であると考えられる。また、提案手法の限界や GTTM のグルーピング構造と異なる点も明らかになった。提案手法は局所的なテクスチャの違いのみに着目していることや、GPR1 をうまく踏襲できていないことにより、期待されるグルーピングが得られない場面があった。さらに、提案手法では GTTM よりも音色の違いによるグルーピング境界が優先して検出される。これは一概には間違いとは言えず、音楽音響信号を対象とした GTTM と記号を対象とした GTTM の最終的なアウトプットは違うものになることを示唆する。今後は大域的な GPR を取り入れ、得られたグルーピング構造に順序構造を付けてタイムスパン木の抽出を試みる。

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

## 参考文献

- [1] F. Lerdahl, R. Jackendoff: A Generative Theory of Tonal Music, The MIT Press, 1983.
- [2] 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.
- [3] K. Hirata, S. Matsuda: Interactive Music Summarization based on Generative Theory of Tonal Music, JNMR, Vol. 35, No. 2, pp. 165-177, 2003.
- [4] M. Hamanaka, K. Hirata, K. Tojo: Melody Morphing Method Based on GTTM, In Proc. International Computer Music Conference, pp. 155-158, 2008.
- [5] M. Hamanaka, K. Hirata, S. Tojo: ATTA: Automatic Time-Span Analyzer Based On Extended GTTM, In Proc. ISMIR, pp. 358-365, 2005.
- [6] J. Foote: Visualizing Music and Audio using Self Similarity, ACM Multimedia, pp. 77-80, 1999.
- [7] 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.
- [8] R. Chen, M. Li: Music Structure Segmentation By Combining Harmonic and Timbral Information, In Proc. ISMIR, pp. 477-482, 2011.
- [9] K. Ullrich, J. Schlüter and T. Grill: Boundary Detection in Music Structure Analysis using Convolutional Neural Networks, In Proc. ISMIR, Taipei, Taiwan, 2014.
- [10] Nakashika, T., Garcia, C. and Takiguchi, T.: Local-feature-map Integration Using Convolutional Neural Networks for Music Genre Classification, In Interspeech (2012).
- [11] Costa, Y. M., Oliveira, L. S. and Koerich, A. L., et al.: Comparing textural features for music genre classification, The 2012 International Joint Conference on Neural Networks, pp.1-6 (2012).
- [12] 澤田隼, 竹川佳成, 平田圭二: 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, 研究報告音楽情報科学 (MUS) Vol.2016-MUS-111, No.23, pp.1-6 (2016).
- [13] R. M. Haralick: Statistical and structural approaches to texture, In Proc. IEEE, vol. 67, pp. 786-804, 1979.