

GTTM の大域的情報を考慮した音響信号の文節の最適化 - 記号と信号処理の相互作用フレームワークの構築に向けて -

澤田 隼[†] 竹川 佳成[†] 平田 圭二[†]

公立はこだて未来大学[†]

1 はじめに

楽譜に書かれた楽曲の構造や意味を分析する手法として Generative Theory of Tonal Music (GTTM) がある [1]. これは人間の認知過程を踏まえた音楽の階層的な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, 楽曲内の相対的な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にとどまらず, 楽曲の構造の操作を可能にするものである.

GTTM を音楽のスペクトログラムに適用する従来の試みでは, スペクトログラムをビートで分割し (bin) その bin 毎に特徴量を抽出し, 階層的クラスタリングを行っていた [2]. しかし, スペクトログラムの分割する位置によって bin が持つ特徴量の値が変化するため, 分析結果も異なるものになる. 通常, 音響信号の情報から拍の位置や和音などの記号を生成するという意味で, 音響信号から記号への処理は一方であるが, 適切な記号接地を実現させるためには音響信号と記号が相互に作用する枠組みが必要であり, 記号から音響信号へのフィードバックが必要になる. 本論文ではスペクトログラムを分割する段階に音楽の大域的な構造をフィードバックすることで適切な分割位置を獲得する枠組み (図 1) を提案し, その有用性を検証する.

2 スペクトログラムのタイムスパンセグメンテーション

2.1 従来手法とその限界

タイムスパン・セグメンテーションとは, Lerdahl と Jackendoff によって導入された基本的な音楽構造の 1 つで, グルーピング選好規則によるグルーピングの結果と, 周期構造やリズムといった拍節構造の結果を統合した構造であり, タイムスパン簡約が行われる領域として定義されている. GTTM のグルーピングの選好規則によると, グループの境界はピッチイベント間の時間軸方向の近接性及び, 音高や音量などの変化に基づいて形成される. ピッチイベントの近接性や変化は, スペクトログラム上ではテクスチャのパターンとしてあらわれる. そこでそのテクスチャのパターンを特徴量として, スペクトログラム内の隣接する bin 間の距離を計算し, これを近接及び変化の尺度として使用する. この尺度と楽曲内の

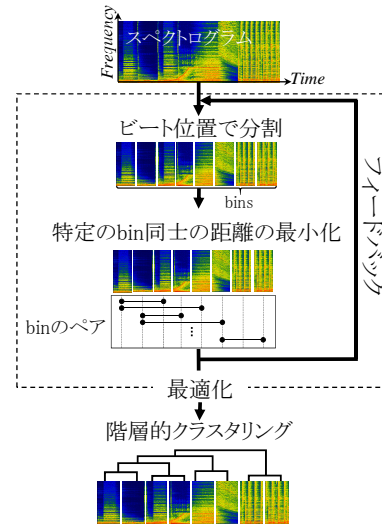


図 1 音響信号の文節へのフィードバック

繰り返し構造の情報を用いて, 時間軸方向に制約を持つ階層的クラスタリングを行う. これにより, テクスチャの変化が小さいものから併合されていき, テクスチャの変化が大きい場合は上位のグループの境界として抽出される.

従来の我々の手法を用いて楽曲を分析した結果, 本来譜面上では同じである箇所分析結果が同じになっていなかった. 音響信号を対象とした分析の場合, 音量の変化やテンポの揺らぎなどの演奏の表情付けによる影響もあるが, GTTM のグルーピングの選好規則によると, 繰り返し現れる箇所は同じ構造になることが望ましい. この原因の一つとして, スペクトログラムの分割する位置が適切でないことが挙げられる. 従来の我々の手法は, スペクトログラムの分割する位置によってその bin が持つ特徴量の値が変化するが, 適切な構造になる様にスペクトログラムの分割の段階まで戻って分割位置を修正する枠組みが無かった. つまり, 同じフレーズの場所は同じ特徴量を持つようなスペクトログラムの分割になるようにスペクトログラムの分割位置を適切に修正する必要がある.

2.2 大域的構造を考慮した分割位置の変更

楽曲中の繰り返し箇所は同じ特徴量を持つようにスペクトログラムの分割位置を適切に修正する必要がある. しかし, そのためには楽曲中の繰り返し箇所を事前を知る必要があるが, 繰り返し箇所を適切に見つけるためには適切な分割位置に基づく特徴量が必要になる. そこで,

Optimization of the Global Structure Information to Spectrogram Segmentation — Toward constructing interaction framework of symbols and signal processing

Shun Sawada[†], Yoshinari Takegawa[‡],
Keiji Hirata[‡], Future University Hakodate

表 1 分割位置の調節による特定の bin 間の距離の変化

	$d(b_1, b_9)$	$d(b_2, b_{10})$	$d(b_3, b_{11})$	$d(b_4, b_{12})$	合計
調節前	1.83	2.43	2.91	0.90	8.08
調節後	1.28	2.22	2.84	0.68	7.03

楽曲中の繰り返し箇所を事前に必要とせず、音楽が一般的に有するといわれている繰り返し構造に関するヒューリスティクスを評価関数に組み込む。一般的に、繰り返し箇所は 2 の n 乗ごとの位置関係に出現するといわれている [1]。そこで、分割位置ベクトル x を入力すると $2, 4, 8, \dots, n$ といった 2 の n 乗の位置関係にある bin 同士の距離 $d(b_i, b_{i+2^j})$ の和を返す評価関数を以下のように定義し、それを最小化することで最適化された分割位置 x' が得られる。

$$x'_{\text{optimize}} = \underset{x}{\operatorname{argmin}} \sum_{i=1}^n \sum_{j=1}^m d(b_i, b_{i+2^j}) \quad (1)$$

subject to $|x - x'| \leq \Delta t$
 $x' \geq 0,$

ただし、 N_{bin} は bin の数を表し、 $n = N_{\text{bin}} - 2$, $m = \log_2(N_{\text{bin}} - i)$ である。ここで、 $m = \log_2(N_{\text{bin}} - i)$ は $i + 2^j$ が N_{bin} を超えないことを意味する。

この不等式制約付き最適化問題を解くために、線形近似による制約付き最適化法 (Constrained optimization by linear approximation: COBYLA) を用いた [3]。COBYLA 法は、目的関数の導関数が分からない制約付き問題に対する数値最適化手法である。COBYLA は解析的な勾配表現を必要とせず、勾配情報を探索過程の中で数値的に生成できる。

3 実験結果

最適化によって得られた新しい分割位置を用いて、従来手法と同じようにタイムスパン・セグメンテーションの抽出を行う。W. A. Mozart の Piano Sonata No.11 in A major (K.331) の第一楽章のテーマの分析結果を図 2 に示す。ここで、使用した K.331 は Maria João Pires によるピアノ演奏のホモフォニー楽曲であった。正解データはオリジナルの GTTM のルールを楽譜に適用した場合の結果を正解とみなし、GTTM の原本に書かれているものと、浜中によって公開されている GTTM database [4] を使用した。従来手法では後半部分の繰り返し箇所 (b_9 から b_{16}) に初期の段階で併合誤りがあったが (正解率 0.64)、最適化によって得られた分割位置を用いた結果、期待する結果が得られた (正解率 1.0)。図 3 に最適化の反復毎の評価関数 (式 1) の値を示す。約 90 回の反復回数で収束しているのが確認できる。表 1 に調節前と調節後の特定の bin 間の距離と、その合計値を示す。楽譜上で同じ箇所である bin 同士の距離が、分割位置を調節することで小さくなっていることが確認できる。

4 おわりに

GTTM を音楽のスペクトログラムに直接適用してタイムスパン・セグメンテーションを生成する従来の方法には、スペクトログラムの分割する位置が適切でないことによって期待する結果が得られない問題があった。本稿では、スペクトログラムを分割する段階に大域的な構造

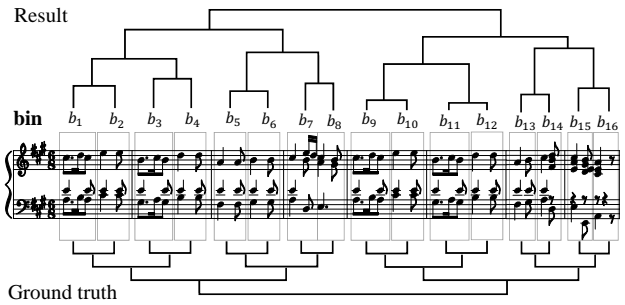


図 2 調節した分割位置を用いたタイムスパン・セグメンテーションの結果

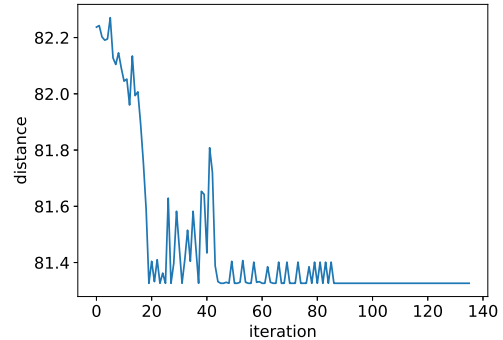


図 3 反復回数と評価関数の値

をフィードバックすることで適切な分割位置を自動で獲得する枠組みを提案し、その有用性を検証した。その結果、スペクトログラムの分割位置によって精度が変わり、適切な分割位置では期待するタイムスパン・セグメンテーションが得られ、大域的構造を考慮した分割位置の変更が有用であることが示された。これは記号から音響信号へのフィードバックの枠組みであると言える。今後は、音楽理論 GTTM の木構造に基づく評価関数を定義し、音響信号処理を最適化することで最適な木構造を獲得する枠組みの構築を目指す。

謝辞

研究を通じて議論をいただいた寺井あすか先生 (公立はこだて未来大学) に感謝いたします。本研究は JSPS 科研費 16H01744 の助成を受けたものです。

参考文献

- [1] Lerdahl, F. and Jackendoff, R.: *A Generative Theory of Tonal Music*, The MIT Press (1983).
- [2] 澤田隼, 竹川佳成, 平田圭二: スペクトログラムの階層的クラスタリングを用いたタイムスパン・セグメンテーション抽出について, 情報処理学会論文誌, Vol.59, No.3, pp.941-950 (2018).
- [3] Michael JD Powell. 1994. A direct search optimization method that models the objective and constraint functions by linear interpolation. In *Advances in optimization and numerical analysis*. Springer, 5167.
- [4] 浜中雅俊: GTTM Database Download Page, 入手先 <<http://gttm.jp/gttm/ja/database/>> (参照 2019-01-11).