

音楽音響信号の階層的クラスタリングを用いた GTTM タイムスパン木の抽出法について

澤田 隼[†]名畑 皓正[‡]竹川 佳成[‡]平田 圭二[‡]公立はこだて未来大学大学院[†] 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。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。しかし、GTTM 分析を計算機上に実現しようとする多くの未解決の問題があることが広く認識されている [1, 2]。例えば、タイムスパン木を生成するために定められた音楽の情報を抽出する選好規則の競合や、分析対象は楽譜上に書かれた楽曲に限られており音楽音響信号を対象としていないという点である。GTTM の適用範囲を音響信号に拡張することができれば、幅広いスタイルやジャンルなど膨大な楽曲に対して音楽的に信頼できる分析をすることができる。本稿では、GTTM を音楽音響信号に直接適用し、タイムスパン木を獲得する方法について述べる。

2 GTTM タイムスパン木の抽出法

本章では、スペクトログラムの階層的クラスタリングによって得られたタイムスパン・セグメンテーション (図 1) に対応する系統樹から、各階層のセグメント内でヘッドを選択し、順序付けすることによってタイムスパン木を抽出する方法について述べる。

2.1 タイムスパン・セグメンテーション

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。人間は旋律の特徴による上位のグループ境界の情報とリズムによる下位のグループ境界の情報を使って、人間の認知と整合する認知的リアリティのある境界を同定している。この 2 つの異なる境界の認知を統合する構造としてタイムスパン・セグメンテーションが導入された。

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

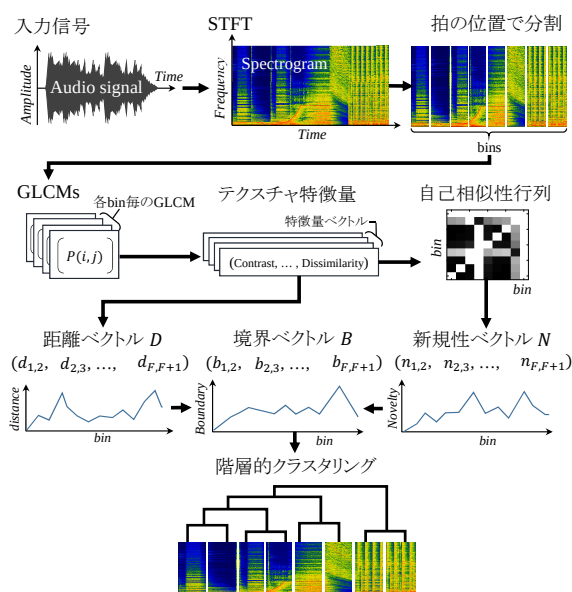


図1 タイムスパン・セグメンテーション

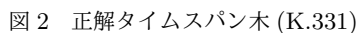
テクスチャのパターンとしてあらわれる。そこでパターン認識技術を使用し、スペクトログラム内の隣接する bin 間の距離を計算し、これを近接及び変化の尺度として使用する。

以下に処理手順を示す (図 1)。初めに、入力された音楽音響信号を短時間フーリエ変換し、グレースケールのスペクトログラムを周波数軸はメルスケールで描画する。次に、スペクトログラムを時間軸方向にビートの位置で分割し、ビートの長さ分の短冊状のデータ (bin) の集合として考える。次に各 bin 毎にスペクトログラムのテクスチャ特徴量を抽出する。その後、楽曲内の繰り返し構造を抽出するためにこのテクスチャ特徴量を用いて自己相似性行列を計算する。これは GPR6 の思想に基づいており、繰り返し構造の始点と終点でグループの境界が強くひかれるような設計をした。最後に、隣接する bin 間のテクスチャの特徴量の距離と、楽曲内の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。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境界として抽出される。

2.2 タイムスパン簡約

タイムスパン簡約は、重要な音が根幹構造を作る一方で、重要でない音はその根幹構造の中への単なる挿入音とする考え方であり、簡約が進むと楽曲は抽象化される。簡約の途中の構造に残った抽象的な音は、その区間のローカルなキーを表していると言える。タイムス

On Extraction Method of GTTM Time-span Tree
Using Hierarchical Clustering for Music Audio Signal
Shun Sawada[†], Terumasa Nahata[†], Yoshinari Takegawa[‡],
Keiji Hirata[‡], Future University Hakodate

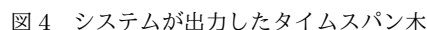
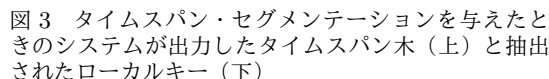


以下に処理手順を示す．初めに，音楽音響信号を入力としコード推定を行う．こうして得られたコード列から Elaine Chew によるスパイラルアレイモデル [4] を用いてローカルキーを抽出する．スパイラルアレイモデルは，音高，和音，調という次元の異なる三つの音楽的実体を一つの三次元座標空間内に示し，幾何学的な座標演算により楽曲の調性同定が可能である．ある階層のセグメント内に含まれるコード列のみをスパイラルアレイモデルに入力し，推定された調をローカルキーとする．この操作を最小のセグメントからボトムアップに全ての階層のセグメントに対して行う．最終的に推定される調は，この楽曲全体を支配するグローバルキーとなる．このグローバルキーを用いてコード列を機能和声へと変換し，カデンツの情報を抽出する．さらに，Fred Lerdahl による Tonal Pitch Space[3] に基づいて，ローカルキーとその中に含まれるローカルキー（コード）の調間距離を求める．この調間距離が小さいものがヘッドとして選択されやすい．最後に，拍節構造とカデンツ情報，ローカルキー調間距離を用いてヘッドを決定する．

3 実験結果

正解データは元々の 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 を使用した．以下に，Mozart の piano sonata in A major (K.331) の第一楽章のテーマ (RWC-MDB-C-2001 Nos. 26) の分析結果を示す．ここで，使用した K.331 はピアノ演奏によるホモフォニーである．正解データを図 2 に示す．

事前情報として正しいタイムスパン・セグメンテーションを与えたときのシステムが出力したタイムスパン木(図3上)と抽出されたローカルキーを(図3下)に示す。 b_5 と b_6 の順序関係が逆になったが、それ以外は期待した結果が得られた。システムが出力したタイムスパン木を図4に示す。出力されるタイムスパン木はタイムスパン・セグメンテーションの精度に大きく依存する。前半部分(b_1 から b_8)はタイムスパン・セグメンテーションが正しく推定されているため、期待した結果が得られたが、後半部分(b_9 から b_{16})はタイムスパン・セグメンテーションまで戻った修正が必要となる。GTTMのグルーピングの選好規則には、タイムスパン木が安定するグルーピング構造を優先するよう定義されている。



4 おわりに

本稿ではスペクトログラムの階層的クラスタリングによって得られたタイムスパン・セグメンテーションに対応する系統樹からタイムスパン木を抽出する方法を提案した。正しいタイムスパン・セグメンテーションが得られている場合は、期待する結果が得られた。今後は、タイムスパン木からタイムスパン・セグメンテーションへのフィードバックを行い、修正する枠組みを構築する。また、楽曲数を増やした実験とその結果に対する定量的な評価をする必要がある。最後に、本手法ではローカルキーを抽出する際に、コード推定アルゴリズムによってコード列という記号に変換したうえで、スパイラルアレイモデルに入力をしているが、音響信号のレベルでローカルキーを抽出し、タイムスパン簡約を定義する。

謝辭

研究を通じて議論をして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] Hamanaka, M., Hirata, K. and Tojo, S.: Implementing "A Generative Theory of Tonal Music", *JNMR*, 35:4, pp.249–277 (2007).
- [2] Hamanaka, M., Hirata, K. and Tojo, S.: Implementing Methods for Analysing Music Based on Lerdahl and Jackendoff's Generative Theory of Tonal Music, David Meredith (Ed), *Computational Music Analysis*, Chapter 9, pp.221-249, Springer (2016).
- [3] Lerdahl, F.: *Tonal Pitch Space*, Oxford University Press (2001).
- [4] Chew, E.: *Mathematical and Computational Modeling of Tonality*, Springer US (2014).