

暗意-実現モデルを用いた旋律抽象化に基づく類似度計算手法

能登 楓^{†1,a)} 竹川 佳成^{†1} 前澤 陽^{†2} 藤島 琢哉^{†2} 平田 圭二^{†1}

概要: 本稿では、記号によって表現された旋律の類似度を計算するために、旋律の形状に関する特徴量の導入を行う。これまで、旋律類似度を計算する代表的な手法としてアライメント手法が存在するが、こうした類似度計算方法では、構成する音符の挿入や欠落に対する頑健性は高いものの、構成する音高や音程の変化の影響を多く受けるという問題があった。本研究では、旋律の形状に関する音楽理論、暗意-実現 (Implication-Realization) モデルによる分析結果による旋律の音高に関する抽象化、同時に、音価比の利用による旋律の時間に関する抽象化を行うような特徴量を導入することによって旋律の形状に関する特徴量を導入することによって上記の問題解決を試みた。さらに、提案する特徴量を入力とする旋律アライメント手法の提案および自己類似性行列の提案を行う。本稿では、ケーススタディとして特徴量ごとに得られる自己類似性行列の比較および、音響信号を対象とした旋律分割手法の適用によって、提案する自己類似性行列の有効性について議論する。

1. はじめに

本研究では、記号を対象とした音楽情報処理における楽曲構造分析を行うための方法を提案する。楽曲構造分析を行う上で重要な手順の一つとしてフレーズ推定が挙げられる。フレーズとは、楽曲を構成する単位の一つであり、音楽的なまとまりを指す。フレーズ推定および、楽曲構造分析の応用先としては、楽曲生成を初めとして演奏表情づけ、編曲、検索などと幅広く、音楽情報処理における基礎技術の一つとなっている。

こうしたフレーズ推定を行う方法は、対象となる楽曲の表現方法に応じて異なるアプローチが取られる。音響信号を対象とした場合には、楽曲における周期性や同質性に着目したフレーズ推定に関する研究が行われている。音楽の特性上、クラシック音楽のように形式を重視するような楽曲はもちろん、ポピュラー音楽も同様に、楽曲を特徴づけるフレーズが意図的に繰り返されることが多い。こうした繰り返されるフレーズでは、音高やリズム、構成楽器などが多少異なるものの、音響信号としては類似する波形が得られることが多い。近年では、こうした特性に着目するようなバウンダリ検出手法 [1], [2] が提案されている。

一方、記号を対象とした場合には分割位置を直接推定す

るような問題として扱われる場合が多い。浜中らが提案した deepGTTM-III [3] では、音楽理論 GTTM におけるグルーピング境界を推定するタスクにおいて、F 値 0.81 の精度での分割位置推定が可能となっている。しかし、こうした分割位置を直接推定するような問題としてフレーズ分割を行うためには、専門家によって作成された教師データが必要であり、時間的にも労力的にも多くのコストがかかる。

これまで、記号を対象とした際に同質性、周期性に着目するような分割位置推定が行われていない理由の一つに旋律類似度計算方法の不完全さが考えられる。教師データの作成を必要とせずに、旋律類似度計算を行う方法としては大きく分けて2つのアプローチが存在する。1つは旋律を音符の時系列とみなし、アライメント手法によるマッチングを行うことによって旋律の類似度を計算する手法 [4], [5], [6] である。こうしたアライメントによる旋律類似度計算は、多少の音符の挿入や欠落に対して頑健であるという利点がある。一方で、旋律を音高や音価といった記号によって表現するため、旋律に含まれる音の一部が異なる音高で置換された場合には、類似度が下がってしまうという問題がある。もう一つの旋律類似度計算方法は、旋律を音高と音価の2次元によって表現し、図形的な距離によって旋律類似度を計算する方法である [7], [8]。こうした方法によって算出される距離は旋律の形状に基づく類似度計算を行うため、一部音高が異なる場合であっても、逸脱度合いに応じた計算が可能となる。一方で、音符の挿入や欠落の影響を受けやすいという問題がある。

そこで本稿では、旋律表現のための特徴量として I-R 分

^{†1} 現在、公立はこだて未来大学
Presently with Future University Hakodate

^{†2} 現在、ヤマハ株式会社
Presently with Yamaha Corporation

a) g3120002@fun.ac.jp

析結果および、音価比を使用することによって、旋律を音高、および時間に関して抽象化するような特徴量を提案する。さらに、提案する特徴量を入力とする旋律アライメント手法の提案および自己類似性行列の提案を行う。

2. 旋律の表現と自己類似性行列の作成

2.1 暗意-実現モデルによる旋律分析

旋律の形状に関する特徴量を導入するために、音楽理論 I-R モデル [9], [10] に基づく旋律分析結果を利用する。I-R モデルとは、ゲシュタルト心理学に基づく旋律分析手法であり、隣接する音列における暗意-実現/否定の関係を音程、音方向に基づく 2 つ原理によって分類する。I-R 分析の結果、類似する暗意-実現/否定関係をもつ音列には同一のシンボルが割り当てられることになる。I-R モデルにおける基本シンボルは全 8 種あり、音程、音方向に関する暗意が実現される音列には、P(process), D(duplication), R(reversal) のいずれかのシンボルが、音程の暗意のみが実現される音列には、IP(intervallic process), ID(intervallic duplication), IR(intervallic reversal), 音方向の暗意のみが実現される音列には、VP(registral process), VR(registral reversal) が割り当てられることになる (図 1)。また、オリジナルの I-R モデルにおけるシンボル種類は上記の 8 種類であったが、各シンボルに対し上下方向の区別を加えた 16 種類のシンボルへと拡張することも可能である。分析結果をもとに、旋律を I-R シンボル列として記述することによって、形状の類似した旋律を同一の表現として扱うことが可能となる。

本研究では I-R 分析結果を得るために、我々の先行研究において提案した I-R 分析器 [11] を使用する。I-R 分析器には必須の特徴量として音高、音価、クロージャ位置、分析性能を向上させるための特徴量として拍節強度、分析結果を操作するためのハイパーパラメータとして、付与するシンボル数、および各音符へ付与するシンボル数、以上計 6 つの特徴量を入力として与える。また、出力としては各音符に付与される I-R シンボル列が得られる。入力の際には、必須の情報である音高、音価に関しては Musicxml や SMF から得られる値を直接使用し、クロージャ位置、拍節強度に関しては事前の推定を行った後に I-R 分析器へ入力する。後述する I-R 分析に関する実験においては拍節強度として、python パッケージの一つである music21 [12] にて実装されている beatStrength オブジェクトへ格納される値を使用する。beatStrength オブジェクトでは、小節における先頭の拍を 1.0、そのほかの強拍を 0.5、弱拍を 0.25 というように、小節内における拍の強さを相対値として表



図 1 暗意-実現モデルにおいて使用されるシンボル



図 2 推定されるクロージャ位置

現されている。

クロージャ位置は、我々の先行研究において提案したクロージャ推定方法を拡張した方法によって獲得する。クロージャとは、これまで聴取した音列から暗意が生じない点、あるいは暗意が鈍化、抑制、弱化する点である。クロージャが発生する要因として考えられるのは、音程、音の方向、音価の変化、休符や強迫の発生、和音進行によって引き起こされる終止感などがあげられる。提案した I-R 分析器においては、先行研究 [13], [14] に基づき、音価の変化に着目したルールベースによるクロージャ推定を行っていた。具体的には、各音符の IOI を算出したのちに、IOI が 2 倍以上となる際の音符をクロージャとみなすというルールである。しかし、このようなルールに基づくクロージャ推定では、音価が連続して増加するような場合、クロージャを連続して検出してしまうという問題があった。前述の定義より、クロージャが発生する点は暗意が消失する点であると考えられることから、連続するようなクロージャは許容すべきではない。そこで、本研究ではクロージャ音が連続する場合には、連続するクロージャの中において最も末尾の音のみをクロージャ音として扱うルールを追加する。この操作によって、図 2 の譜面を対象とした場合に推定されるクロージャは 3 音目および末尾の音となる。

I-R 分析結果を操作するためのハイパーパラメータとして、付与する I-R シンボルの種類、およびシンボル最大付与数を使用する。Narmour によって提案されたオリジナルの I-R 分析では旋律の形状に基づく 8 種類のシンボルによって I-R 分析が行われていた。しかし、これら 8 種類のシンボルには音列の方向に関する情報は含まれていないため、I-R シンボル列の旋律表現能力は制限されていた。そのため、本研究で使用する I-R 分析器では矢澤らの研究と同様 [14]、シンボルの上下方向の区別を行った 16 種類のシンボルによる I-R 分析を可能とする。シンボル数の拡張によって、I-R シンボル列によって表現可能な旋律は増える一方で、反行の音列を同一とみなすことが出来なくなるため、取り組む問題に応じて使用するシンボル種類は適宜変更する必要がある。

シンボル最大付与数に関しても、事前にハイパーパラメータとして決定する必要がある。シンボル最大付与数とは、各音符に対して付与される最大 I-R シンボル数のことであり、1 から 3 の範囲でいずれかの入力を I-R 分析器へ与える。先行研究において、最大付与数の妥当性を検証するために Narmour による手分析との比較を行ったところ、最大付与数が 2 の場合の F 値が最も高い値となることを

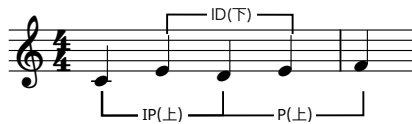


図3 I-R 分析例 (シンボル種類 16, 最大付与数 3)

確認しているものの、特徴量として扱う場合に最適な最大付与数に関する調査は行っていない。現在は最大付与数に関する定量的な評価を行える段階にはないため、本稿における I-R 分析結果は全てシンボル最大付与数を 3 としている。また、最大付与数が 3 の場合に得られる I-R シンボル列は、旋律から得られる tri-gram 全てに対するシンボル付与を行ったシンボル列から、クロージャを跨ぐようなシンボルを除いたシンボル列と等価である。

2.2 暗意-実現モデルによる分析結果を特徴量とした自己類似性行列

2つの音列において類似するフレーズを検出するための表現として類似性行列を作成する。理想的には、フレーズ内における音高が完全に一致する場合だけでなく、音高や発音タイミングのズレを許容するようなフレーズ、いわゆる人間が聴取した場合に似ていると感じる音列同士の類似度が高くなるような類似性行列を作成することが望ましいだろう。そこで、本研究では類似性行列を作成するための特徴量として、形状に関する特徴量である I-R 分析結果を導入する。

本研究では、音符よりも広い区間を考慮した類似度計算を行うために、各音符に対し付与された I-R シンボルにおける位置を使用する。具体的な事例を示すために、実際の旋律を対象に得られる特徴量表現を示す。図3にハイパーパラメータとして、シンボルの種類 16, 最大付与数 3 の場合の I-R 分析を想定した I-R 分析結果を示す。譜面上の旋律を M とする場合に得られる I-R 分析結果を元に作成される特徴量 $IR_{SymbolPosition}$ を以下のように表す。

$$IR_{SymbolPosition}(M) = \begin{pmatrix} 0 & 0 & 1 & 2 & 3 \\ 0 & 1 & 2 & 3 & 0 \\ 1 & 2 & 3 & 0 & 0 \end{pmatrix} \quad (1)$$

$IR_{SymbolPosition}$ の次元は (I-R シンボルの種類, M に含まれる音符数) となっており、式 (1) においては、1 行目がシンボル P(上)、2 行目がシンボル ID(下)、3 行目がシンボル IP(上) に対応している。図3のように 3 音目から 5 音目にかけてシンボル P (上) が付与されている場合には IR_{symbol} の (1,3) 成分から (1, 5) 成分にかけて I-R シンボルにおける位置 1, 2, 3 が付与され、シンボルが付与されていない成分には 0 が付与される。

こうして得られた I-R 分析結果を各音符にアノテーションすることによって、旋律の形状を考慮できるような音符の表現を提案する。I-R 分析によって得られるシンボルは

音高に着目した分類であるため、時間に関する抽象化は行われていない。そこで、本研究では時間に関する抽象化を行うために、前の音符との音価比を特徴量として導入する。音高方向、時間方向に関する抽象化の結果を組み合わせることによって、旋律に含まれる音符は (t_1, t_2, \dots, t_n) は \langle I-R シンボル, シンボルにおける位置, 前の音との音価比 \rangle という 3 つ組の集合として表現することができる。図3の旋律における第 1 音から第 3 音は以下のように表現される。

$$\begin{aligned} t_1 &= \langle \text{IP(上)}, 1, 1 \rangle \\ t_2 &= \langle \text{IP(上)}, 2, 1 \rangle, \langle \text{ID(下)}, 1, 1 \rangle \\ t_3 &= \langle \text{IP(上)}, 3, 1 \rangle, \langle \text{ID(下)}, 2, 1 \rangle, \langle \text{P(上)}, 1, 1 \rangle \end{aligned}$$

各音符を表現する要素数はハイパーパラメータであるシンボル最大付与数によって決定される。図3に示した I-R 分析例では最大付与数 3 であることから、各音符 t_n を表現する集合における要素数は 0 から 3 となる。また、I-R 分析の特性上、開始音が共通するシンボル付与を行うことはないので、第 1 音目は必ず 1 つ以下の要素をもつ集合として表現されることになる。同様に第 2 音目も必ず 2 つ以下の要素をもつ集合となり、第 3 音目以降は 3 つ以下の要素をもつ集合として表現される。こうした表現の結果、図3の譜面を対象に得られる表現例から分かるように、共通のシンボルをもつ音符であっても、付与された I-R シンボルにおける位置、および前の音との音価比が異なる場合の t_n は異なる要素をもつ集合として表現されることになる。各音符を表現するほぼ全ての要素は、 $IR_{SymbolPosition}$ および、音価から求めることができるが、楽曲における先頭の音符に関しては、直前の音符が存在しないため特別に音価比として 1 を与えることにする。

旋律同士の距離計算には、DP マッチングに基づく旋律アライメント手法を拡張した方法を用いる。旋律に含まれる音符 (t_1, t_2, \dots, t_n) は上記の 3 つ組を要素とする集合として表現しているため、音符同士の類似度は集合の距離として計算することができる。この際、集合同士の距離を計算する際には、距離計算を行う対象となる 2 つの集合に含まれている要素のうち共通要素が占める割合を表す Jaccard 係数を使用する。集合の系列同士の類似度計算を行うために、以下の式に基づくコストを算出する。

$$Cost_{ij} = \begin{cases} \alpha \cdot Cost_{i-1j-1} + (1 - \alpha) \cdot \text{Jaccard}(t_i, t_j) & (i > 0 \text{ かつ } j > 0), \\ \text{Jaccard}(t_i, t_j) & (i = 0 \text{ または } j = 0) \end{cases}$$

ここで、 $(t_1, t_2, \dots, t_i), (t_1, t_2, \dots, t_j)$ は距離計算の対象となる旋律を表しており、 α は、類似度計算を行う際に直前の類似度をどれだけ反映させるかを決定するハイパーパラメータである。 α の値が 0 である場合には、 $Cost_{ij}$ は各時刻同士の音符の類似度を表す値となり、それ以外の値で

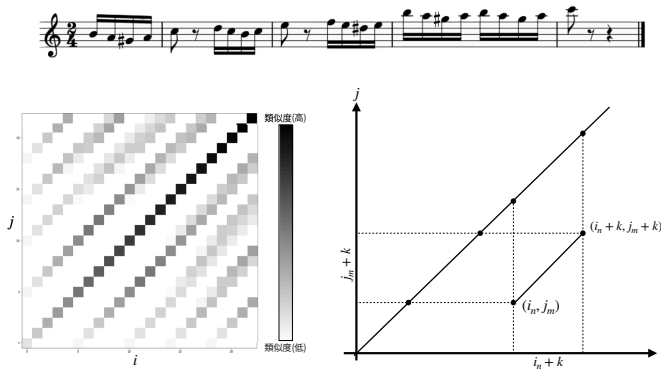


図4 W. A. Mozart piano sonata kv.331 冒頭を対象に得られる自己類似性行列およびその解釈

ある場合には過去系列における類似度を反映させた累積類似度を表す値となる。また、距離計算の対象となる旋律が一致する場合に得られる Cost を自己類似性行列と呼ぶ。

図4に、W.A.Mozart 作曲ピアノソナタ第11番イ長調 K. 331 第3楽章より1から4小節における右手譜面を対象に得られる自己類似性行列(シンボル最大付与数3, シンボル種類16, $\alpha = 0.5$)を示す。類似性行列における (i, j) 要素は t_i, t_j における類似度を表しており、類似度は色の濃さによって表現されている。また、 (i_n, j_m) から (i_n+k, j_m+k) 要素を結ぶ直線は、 i_n 音目から i_n+k と i_m 音目から i_m+k 音目の区間が類似していることを意味する。そのため、自己類似性行列における対角成分 $(0, 0)$ から (N, N) までの対角成分における類似度は、常に自分自身との類似度を表しているため、 i 行、 j 列目において最大値となる。また、 $(i, 0)$ を始点とした対角成分が確認できる場合には、先頭から始まる音列との類似度が高くなるような音列の開始音であることを示している。図4において $(i, 0)$ を始点とした対角成分が確認できるのは、6音目、11音目、15音目、19音目であり、これらは、第1音目から始まる H-A-Gis-A と類似した音形の開始音である。

3. 評価としてのケーススタディ

3.1 暗意実現モデルによる分析におけるハイパーパラメータと類似性行列

類似性行列の結果を変化させる要素として、I-R 分析におけるハイパーパラメータがある。I-R 分析におけるハイパーパラメータはシンボル最大付与数と付与するシンボルの種類の2つであるが、シンボル最大付与数に関しては、類似性行列における類似度の変化への影響はあるものの、類似性行列の見た目を変化させるパラメータではないため、本稿では取り扱わない。一方で、付与するシンボルの種類に関しては、類似性行列の見た目を大きく変化させるパラメータである。以下にパラメータを変化させた場合の類似性行列を示す。

図6に F.F.Chopin Waltz in E flat, Op.18 (図5) への分

析によって得られる自己類似性行列を示す。シンボルの種類は Narmour によって提案された8種類の I-R シンボルおよび、上下方向の区別を加えた16種類の I-R シンボルのどちらかを選択することができる。図6左における自己類似性行列では16種類の I-R シンボルによる分析結果を特徴量としているため、冒頭の旋律 B-D-E-F と類似度が高いと判断される旋律は、第2小節に出現する B-Es-F-G、および第3小節に出現する B-F-G-As となる。一方、8種類の I-R シンボルによる分析結果を特徴量とした場合(図6右)には、上記の2つに加え、第5小節に出現する C-B-As、6小節に出現する B-As-G、第7小節に出現する As-G-F、も同様に類似した旋律と判断される。このことから、8種類の I-R シンボルによる分析結果を特徴量として用いる場合には、反行するようなモチーフの抽出が可能になることを示唆している。

3.2 自己類似性行列の作成における特徴量の比較

I-R 分析結果を用いた自己類似性行列の妥当性を評価するためのケーススタディの1つとして、特徴量の比較を行う。対象楽曲は W.A.Mozart 作曲ピアノソナタ第11番イ長調 K. 331 第3楽章より1-4小節における右手譜面(図4上)であり、16種類の I-R シンボルによる分析結果を特徴量とした場合に得られる自己類似性行列は図4左下に、前述した3つ組における I-R に関する要素を音高もしくは音程によって代替した場合に得られる自己類似性行列を図7に示す。図7左における自己類似性行列は音高を特徴量として使用した場合に得られる自己類似性行列であり、他の自己類似性行列と比較して特徴的な部分は市松模様が表れている点である。これは、対象楽曲(図4上)には H-A-Gis-A のように、同一の音高が一音飛ばして現れる音列が含まれているためであると考えられる。このような場合、類似度が高いと判断される音符が1音飛ばして表れるため、結果として自己類似性行列に市松模様が表れた。このように、音高を特徴量とした場合には、音符単位での類



図5 F.F. Chopin, valse np.19 Es-Dur op.18 より抜粋

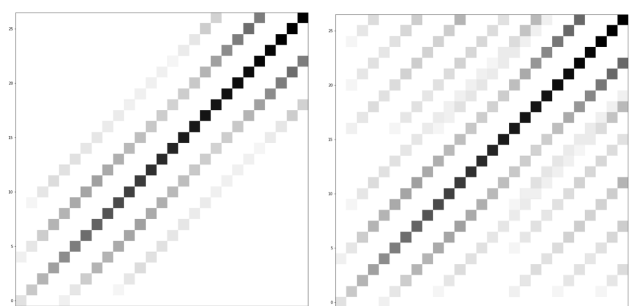


図6 図5の譜面を対象に得られる自己類似性行列 (左: I-R シンボル数16, 右: I-R シンボル数8)

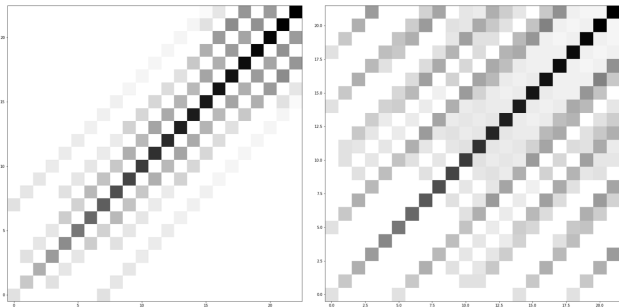


図7 図4の譜面を対象に得られる自己類似性行列
(左: 特徴量として音高を使用, 右: 特徴量として音程を使用)

似度計算を行なってしまふ。

一方, 図7右における自己類似性行列は音程を特徴量として使用した場合に得られる自己類似性行列だが, I-R 分析結果を使用した場合と同様, 対角成分が現れるような自己類似性行列が得られていることが分かる。一方で, 音程を特徴量として使用しているため, 局所的な類似度の影響を受けている。例えば, 第2小節2拍目に現れる F-E-Dis-E という音列における類似度を計算する場合には, 度数はそれぞれは短2度, 短2度, 短2度によって構成されているため, F-E と E-Dis の類似度が高くなってしまふ。こうした局所的な類似度の影響を受けるため, I-R 分析結果を特徴量とした自己類似性行列と比較して, 全体的に類似度の高い, かすかにぼやけたような自己類似性行列が得られる。

3.3 分析対象の拡張およびバウンダリ検出手法の適用

自己類似性行列に対するバウンダリ検出を行う。記号によって表現された音楽を対象とした自己類似性行列を作成した事例が見つけれなかったため, 音響信号を対象としたバウンダリ検出手法を適用する。音響信号を対象としたバウンダリ検出方法における代表的な手法として, チェッカーボードカーネルの利用した新規性ベクトルに基づくバウンダリ検出手法 [15] があり, アルゴリズムの変更をすることく本研究で提案した自己類似性行列への適用が可能である。

自己類似性行列からバウンダリを検出するにあたって, まず, 新規性 (Novelty) の獲得を行う。新規性は, 自己類似性行列における対角線上の $N \times N$ のチェッカーボードカーネルをずらしながら掛け合わせることによって獲得できる。通常, チェッカーボードカーネルのサイズはサンプリングレートに比例して決定していたが, 本研究における自己類似性行列の最小単位は音符単位であるため, ハイパーパラメータとして適宜決定した。また, 同様に事前に決定すべき事項としてチェッカーボードカーネルの種類があるが, 本研究ではノイズに頑健であるとされるガウシアンカーネルを使用した。図8-10ではF.F.Chopin mazurka op.24-4の全体を対象に出力される自己類似性行列に対し, フィルタサイズ30音, シフト幅1音とした場合に推定さ

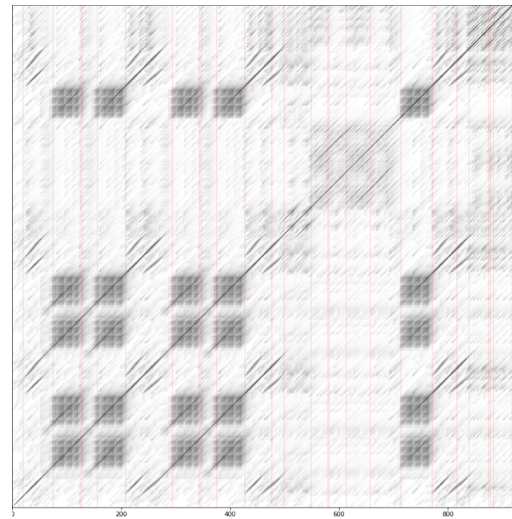


図8 自己類似性行列 (フィルタサイズ:30音, 特徴量:音高)

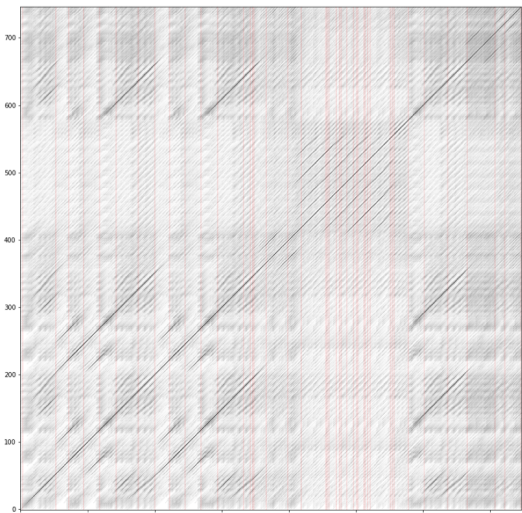


図9 自己類似性行列 (フィルタサイズ:30音, 特徴量:音程)

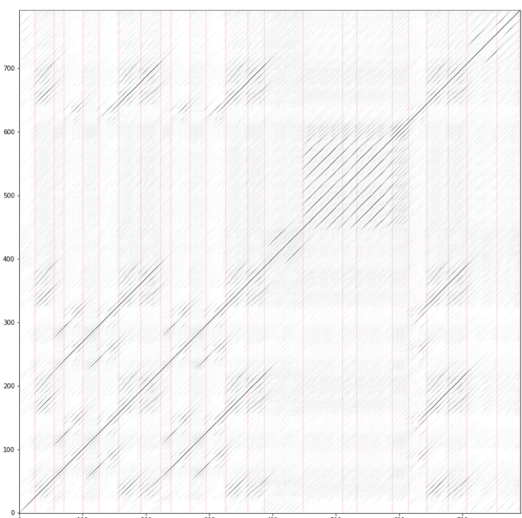


図10 自己類似性行列 (フィルタサイズ:30音, 特徴量:I-R 特徴量)

れるバウンダリを重畳させた結果を示す。チェッカーボードカーネルと自己類似性行列を掛け合わせた結果が急峻に変化する地点をフレーズ境界として推定されるため, 視覚的

なバウンダリとある程度一致していることが分かる。

また、それぞれの特徴を使用した自己類似性行列は、1フレーズを対象に作成した自己類似性行列である図4および図7における特徴を反映している。特徴量として音高を利用した場合に得られる自己類似性行列には正方形の区間が表れていることが分かる。これは、1フレーズを対象にした場合に現れる市松模様が連続して複数出現したために表れたものと推測される。音程やI-R分析結果を特徴量として使用した場合には、こうした正方形が表れていた区間に対角成分が表れているのが分かる。このことから、自己類似性行列を作成する対象を楽曲全体に拡張した場合においても、特徴量が音程、I-R分析結果の場合には再起的に出現する音列の検出が可能であることを示唆している。しかしながら、音程を特徴量として使用した場合には、局所的な類似度が強く反映されているため、全体的に黒みがかかっている。

バウンダリ検出結果もまた、こうした自己類似性行列の特徴を反映する。新規性ベクトルに基づくバウンダリ検出結果は、自己類似性行列内のテキスト変化の影響を大きく受ける。そのため、音高を特徴量として使用した自己類似性行列における正方形の区間のように、視覚的な直感に準ずる部分で分割されている。全体的に黒みがかかっている図9では、図8および図10に比べ局所的な分割が多く見られる。そのため、楽曲におけるAメロ、Bメロのように大まかに特性の類似した区間を基準とした分割を行うことは困難である。一方、I-R分析結果を特徴量とした際に得られる自己類似性行列では、類似区間と非類似区間が顕著に表れているため図8、図9と比較して局所的なバウンダリ検出を行っておらず、もっとも小さなバウンダリ区間でさえ約20から30音程度であることが分かる。

提案する自己類似性行列に対するバウンダリ検出結果の妥当性について議論する。図11はW.A.Mozart作曲kv.331第3楽章より抜粋した旋律を対象に、特徴量としてI-R分析結果(最大付与数3, シンボル種16)を使用した際に得られる自己類似性行列に対し、フィルタサイズ3, シフト幅1の場合に得られる新規性ベクトルによって推定されるバウンダリを表している。その中でも、第1小節から第15小節において推定されたバウンダリおよび、対応する楽譜、自己類似性行列を現したのが図12である。また、赤線はバウンダリを現しており、説明のために先頭から順に番号を示している。同時に、バウンダリが我々の主観と合致している場合にはチェックマークを記している。図12から分かるように、チェックマークが記されたバウンダリは楽曲の冒頭から始まる旋律と類似する箇所であり、多くの場合(i, 0)から始まる対角成分を確認することができることから、自己類似性行列から我々の主観と合致するバウンダリを得られる可能性がある。

一方で、このようなバウンダリ検出となった理由として

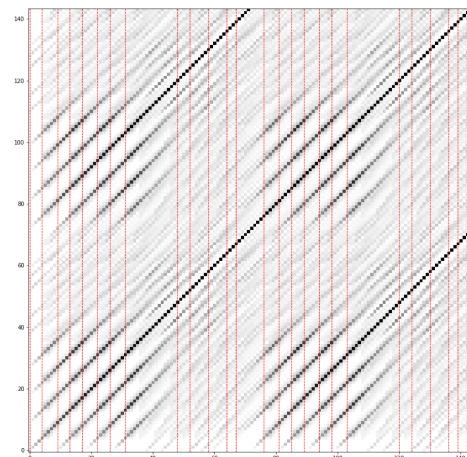


図 11 W.A.Mozart, kv.331 を対象に得られる自己類似性行列 (フィルタサイズ:10 音, 特徴量:I-R 分析)

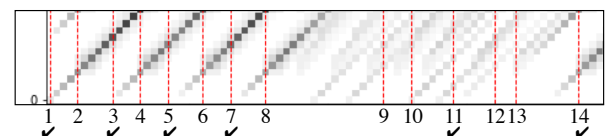


図 12 図 11 左下の拡大および、推定されたバウンダリ

は、検出方法の特性が考えられる。前述したように、使用したバウンダリ検出手法は音響信号から作成される自己相関行列を対象とする手法であることから、テキスト変化の影響を多く受ける。そのため、音響信号を対象とした場合に比べて、旋律を表現する空間がスパースである自己類似性行列はテキストの変化が起こりにくい。そのため、微細な類似度の変化によって過剰なバウンダリ検出をしてしまったと考えられる。こうした問題を避けるためには、自己類似性行列に対して独自のバウンダリ検出手法を検討する必要がある。

3.4 暗意実現モデルによる分析が適切に機能しない事例

I-R分析結果を特徴量として利用する際の問題点を述べる。ここまでのケーススタディから、I-R分析に基づく特徴量は音高の微細な変化の影響を受けることなく、フレーズの形状に基づく旋律類似度の計算が可能になることを示唆している。一方で、I-R分析結果に基づく特徴量は、音高の挿入や欠落の影響を大きく受ける。図13に音符の挿入や欠落がある場合のI-R分析結果を示す。譜例1がオリジナルの旋律とした場合に、譜例2が挿入がある場合、譜例3は欠落がある場合の旋律である。I-R分析結果から分かるように、譜例1~3は全く異なるI-R分析結果となる。また、一見P(上)というシンボルが共通しているよう



図 13 音符の挿入, 欠落によって変化する暗意-実現モデル分析結果

に見えるが, 譜例 1 における P(上) は第 2, 3, 4 音から構成されているが, 譜例 2 における P(上) は第 3, 4, 5 音によって, 譜例 3 における P(上) は第 1, 3, 4 音によって構成されているように, シンボル P(上) を構成する要素は異なる. また, 譜例 1 における IP(上) は譜例 2, 譜例 3 において消失しており, 旋律類似度を高める要素として起因しているシンボルは P(上) のみであることから, 曲中にこうした 3 つ譜例のような変奏的なフレーズが出現した場合に先頭音を探索するのは不可能である.

こうした問題へ対処する方法として, 階層的な I-R 分析が考えられる. 認知的音楽理論の主流である GTTM や Schenker 理論では, 近接する音列の中から重要ではない音の捨象によって簡約を行う. こうした重要な音のみによって構成される仮想的な旋律を仮定できるような I-R モデルの拡張を行うことによって上記のような, 挿入, 欠落への対象が可能になると考える.

4. おわりに

本稿では, 旋律の形状を表現する特徴量として, I-R 分析結果および音価比の組み合わせた表現方法の導入および, 特徴量をもとに旋律類似度計算方法の提案を行った. 同時に, 楽曲自身を対象に旋律類似度計算を行うことによって得られる自己類似性行列の提案も行い, I-R 分析に用いるハイパーパラメータ, および特徴量の違いによって得られる自己類似性行列の比較, 既存の旋律分割手法によって推定される旋律分割位置の調査といったケーススタディを通して提案する特徴量の妥当性を検討した. 今後の課題としては, 今回提案した自己類似性行列の定量的な評価があげられる. ケーススタディの一つとして実施したように, 提案した自己類似性行列には, 従来手法のバウンダリ検出手法は適していなかった. 今後は自己類似性行列に即したバウンダリ検出手法の提案によって定量的な評価を行う.

謝辞

本研究は, JSPS 科研費 (16H01744, 19H04157) の助成を受けたものです.

参考文献

[1] 柴田剛, 錦見亮, 中村栄太, 吉井和佳. (2020). 同質性・反復性・規則性を考慮した階層隠れセミマルコフモデルに基づく統計的音楽構造解析. 情報処理学会論文誌, 61(4),

pp.757-767.
 [2] A.Maezawa. "Music boundary detection based on a hybrid deep model of novelty, homogeneity, repetition and duration", in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 206 – 210, 2019.
 [3] H. Masatoshi, K.Hirata, and S.Tojo. "deepgttm-III: multi-task learning with grouping and metrical structures", International Symposium on Computer Music Multidisciplinary Research. Springer, Cham, 2017.
 [4] D. Bountouridis, D. G. Brown, F. Wiering, and R. C. Veltkamp. "Melodic Similarity and Applications Using Biologically-Inspired Techniques", Applied Sciences, 7(12), 2017.
 [5] S.Park, et al., "A Cross-Scape Plot Representation for Visualizing Symbolic Melodic Similarity". ISMIR, pp. 423-430, 2019.
 [6] Laskowska B, Kamola M. Grouping compositions based on similarity of music themes. PLoS One. 2020 Oct 8;15(10):e0240443. doi: 10.1371/journal.pone.0240443. PMID: 33031468; PMCID: PMC7544027.
 [7] Aloupis, Greg, et al., "Algorithms for computing geometric measures of melodic similarity." Computer Music Journal 30.3, pp.67-76, 2006.
 [8] T.Eerola, and M.Bregman. "Melodic and contextual similarity of folk song phrases", pp.211-233, 2007.
 [9] E.Narmour: The Analysis and Cognition of Basic Melodic Structures", The University of Chicago Press, 1990.
 [10] E.Narmour: The Analysis and Cognition of Melodic Complexity, The University of Chicago Press, 1992.
 [11] 能登楓, 竹川住成, 平田圭二, 旋律の音価, 遷移, 拍節構造に関するルールに基づく IR 分析器実現方式の検討, MUS, pp.1-6, 2021.
 [12] M. S.Cuthbert and A.Christopher: music21: A toolkit for computer-aided musicology and symbolic music data (2010).
 [13] 波多野誼余夫: 音楽と認知, 東京大学出版会, pp.1-40, 1989.
 [14] 矢澤櫻子. 音楽理論暗意実現モデルに基づく楽曲解析に関する研究: 筑波大学, 博士論文, 2016.
 [15] J. Foote. "Automatic audio segmentation using a measure of audio novelty. In Proceedings of the IEEE International Conference on Multimedia and Expo. IEEE Computer Society Press, Los Alamitos, CA, pp. 452 – 455, 2000.