

旋律簡約の定式化に向けて

平田 圭二^{1,a)} 能登 楓¹ 山内 拓真¹ 竹川 佳成¹

概要: 本発表では、旋律の簡約という操作を通して、音楽理論 GTTM と暗意-実現モデルの統一的な解釈に関して議論する。GTTM は、認知的ゲシュタルトがもたらすグルーピングに基づいて旋律の簡約を定義している。同じく認知的ゲシュタルトに基づく音楽理論である暗意-実現モデルに、オリジナルの理論では定義されていなかった旋律の簡約を導入することで、GTTM と暗意-実現モデルの相対的な位置づけを明らかにしていく。

1. はじめに

絶対音楽とは、具体的な心象や風景に音楽を結びつけることなく、音楽自体に抽象的な美を求める音楽の在り方である。そのような芸術が可能となる程度に音楽は抽象的であり純粋であり、それゆえ、音楽家、演奏家、聴衆などが感じる音楽の意味は多様、曖昧、主観的であると言える。しかし、言語と音楽は同じ脳の機能を用いて運用されているという事実や、音楽にも非言語メディアとしてのロジックが存在するという内省から、「いったん、理性や完成としての音楽は忘れて、計算機で扱える情報を送る器・媒体として音楽・音を捉えなおす」[4] というアプローチも一定の成果をあげてきたと言える。

音楽を情報メディア、コミュニケーションメディアとして扱う技術をさらに深化させるためには、音楽をより適切に計算機上で表現することが欠かせない。実際に音楽を計算機上で表現するとしたら、現在は楽譜データ、MIDI、ピアノロールなどが主流である。しかし、これらの表現手法は横軸が時間、縦軸が音高という2次元平面上にピッチイベントを配置するものであり、音楽的な近さ、認知的な近さを全く考慮に入れていない。したがって、類似楽曲の検索、潜在空間への埋め込みなどのタスクでは、横軸が時間、縦軸が音高という2次元ユークリッド距離、あるいはそれに準じた距離が、暗黙裏に結果に影響を及ぼしている可能性がある。

より適切な音楽知識表現を追究するには、音楽の意味を明らかにしなくてはならない。本稿では、これまで我々が取り組んできた音楽知識表現や音楽理論の定式化の結果を踏まえて、より適用範囲の広い音楽の意味論や旋律分析手

法について、背景やアプローチについて述べる。

2. ゲシュタルトとグルーピング

楽譜として記述された音楽に、記号論の立場から意味を与えることができる [13]。例えば、ドという1音からなる旋律 M_1 、レという1音からなる旋律 M_2 、ドレという2音からなる旋律 M_3 を考える。 M_3 には、 M_1 と M_2 に含まれていなかった（何らかの）新しい意味が生じている。時間方向に音列を作るだけでなく、音高方向に音列を作っても、元の音列に含まれていなかった（何らかの）新しい意味が生じる。このように、音楽の意味付けの根拠を音どうしの相対的な関係（暗黙的な参照）に求めることができるが、すべての音や音列が全ての音や音列を一律一様に参照しているわけではない。認知的ゲシュタルトによって、近接、類同、同方向の音や音列が、より強い参照を生じさせる（図1）。



図1 旋律に含まれる暗黙的な参照

音楽における認知的ゲシュタルトによって、2種類の知覚が生じると言われている（図2）[1], [2], [11]。図2(a)は、より強い参照関係で結ばれている音どうしが同心円的に広がり、階層的なグルーピングを表している。図2(b)は、連続した2つ以上の音から次に聴く未来の音を予測し、直線的に延びるグルーピングを表している。

この2種類のグルーピングを実感できる簡単な思考実験を紹介する。まずドレミという音列がある時、次に聴こえるであろう音を予測してみたい。多くの人はファと

¹ 公立はこだて未来大学

^{a)} hirata@fun.ac.jp

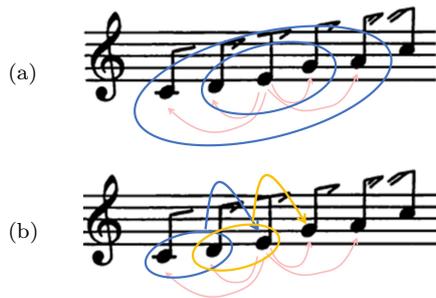


図2 ゲシュタルトに基づく2種類のグルーピング

答えるであろう。そしてもし実際にドレミファと聴こえた時、さらにその次に聴こえるであろう音を予測してもらおうと、おそらくより先より多くの人がソと答えるだろう。これは、予測通りの音が聴こえることで期待が実現され、より長いグルーピングが得られたことを表す(図2(b))。では次に、ドレミドレミという音列を聴いた時、次に聴こえるであろう音を予測してみたい。多くの人はドあるいはドレミと答えるだろう。これは、ドレミが繰り返されることで、ドレミという音列=記号と認識されたことを意味している(図2(a))。ドレミという記号どうしの参照が生じることで、同じドレミという音列でも、その前に聴いた音によって予測が変わるという例である。

実際の楽曲中でもこの2種類のグルーピングが生じている(図3)。楽譜中、1~2小節めの楽句と同一の楽句が5~6

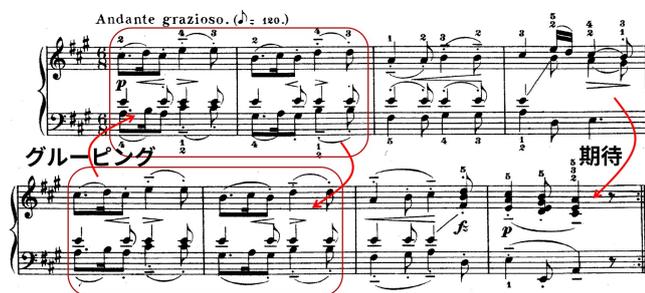


図3 楽曲中のグルーピングと期待の例

小節で繰り返され、この2つの楽句どうしは相互参照することでグルーピングされる。4小節めに半終止(期待の否定)が現れることで、8小節めの正格終止(期待の実現)が期待され、それは実現される。さらに、それぞれの楽句を含む1~4小節めと5~8小節めは、類似した楽句の繰り返しであり、緩くグルーピングされる。

3. シェンカー理論再考

シェンカーは、ゲシュタルトによるグルーピング構造に基づいて、楽曲はより簡単な構造に簡約できるという簡約仮説を提唱し、この簡約仮説の適用によって得られる旋律 Ursatz を音楽の基本構造と定めた[13]。シェンカー理論では、楽譜として記述された楽曲を最終的に Ursatz に簡約する過程を楽曲分析と捉え、全ての楽曲はこの Ursatz が

ら生成(generate)されてくるとする。主観や暗黙知を排した楽曲分析を目指して、簡約という手続きを導入した点は画期的であった。一方、前章で挙げた2種類のグルーピングは明確に区別されていなかった。

シェンカー理論に関して、Lerdahlは以下のような欠点を指摘している[6]:(1)アプリアリな存在である Ursatz が理想主義的過ぎてその妥当性に疑問が残る、(2) Ursatz が適用できる時代や文化は限られている、(3)リズムは音高列と同様に重要であるにもかかわらず、Ursatz にはリズム要素が含まれておらず、リズム要素をどのように生成するかが不明である、(4) Ursatz から様々な表層の音列が生成される方法論が不明である、(5)我々は生成された旋律が持つ構造に興味があり、その構造は認知的な裏付けがあるべきだが、シェンカー理論には認知的な裏付けが欠けている。

また Narmour もシェンカー理論の欠点を幾つか指摘している[8],[9]:(1)簡約とそれから生じる階層構造という方法論にしたがい、もし2つの異なる旋律を簡約した結果、上位において CDEF という同じ音列が得られたとする。この異なる旋律から得られた同一の CDEF は、その作品をその作品たらしめている idiosyncratic な構造を反映していない(同一の表現に異なる意味を持たせられない)。(2)期待(暗意)を生み出すモデル部分と、その暗意が実現されるコピー部分の関係性に idiosyncrasy があり、コピー部の実現には程度がある(conformance of copy)、(3)楽曲分析において重要な音楽構造は、グルーピングの階層性ではなく、クロージャ音/ノックロージャ音*1の区別である、

4. Generative Theory of Tonal Music

Lerdahl と Jackendoff は、認知的ゲシュタルトによるグルーピングが生み出す階層性に着目し、時系列メディアである言語の表層に現れる単語列が持つ構文構造や係り受け関係が木構造で表現されるように、音楽におけるグルーピング構造も木構造で表現されるという仮説を立てた。そして、簡約操作とそのドメインとしてタイムスパン木(図4)を導入した[7]。タイムスパン木の上位では、ピッチイベントに関するグルーピング選好規則によるグルーピングが採用され、下位では拍節選好規則によるグルーピングが採用されている。上位と下位の境界は、およそ四分音符2拍から1小節程度、つまり1秒から数秒の時間幅に当たっている。この境界より短い時間幅と長い時間幅では、人間の音楽認知における優先されるグルーピング原理が異なることを反映している。

平田らは、タイムスパン木とその簡約操作の定式化に取り組み、最大タイムスパンという概念を導入した[3]。音楽学的には、最大タイムスパンは局所的な調性と解釈するこ

*1 クロージャ音とは、それまで聴取した恩れてスから暗意が生じない点、あるいは暗意が鈍化、抑制、弱化した点のこと[13]。

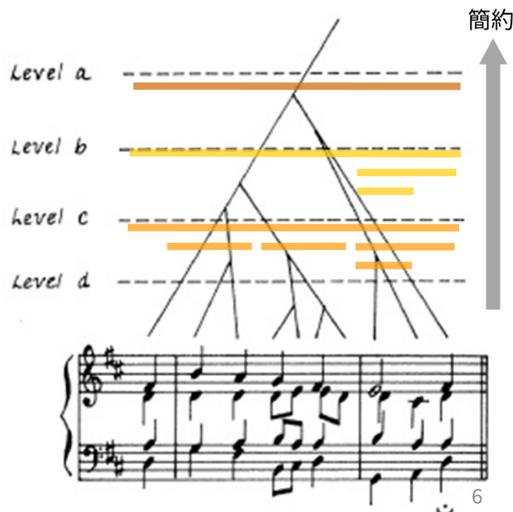


図4 タイムスパン簡約と最大タイムスパン

とができるので、タイムスパン木は局所的な調性の階層構造を表現していると主張した。GTTM は生成文法風に記述されており、タイムスパン木の開始変数（初期状態を表す非終端記号）は最長の最大タイムスパンであり、それは楽曲の調性 (key) に相当する。

5. 暗意-実現モデル

Narmour はまず、イディオ構造 (idiostructure, 特徴的に現れる差異, 唯一性, 特異性), スタイル構造 (類似性のレベルでの差異, 繰り返し現れる差異), スタイル様式 (楽曲構成ルール, 厳格な複製) の3レベルを区別することが重要であると主張する [9]。そして、旋律の簡約や、トップダウンの具体化によって旋律が生成されるという方法論に頼らず、楽譜レベルの音どうしのゲシュタルトに基づく参照関係からイディオ構造を抽出する方法論を提案した [10]。

イディオ構造の基本は、クロージャから開始する旋律の断片を聴取していく際に得られる暗意 (期待, 予測) とその実現あるいは否定 (逸脱) である。暗意を作り出す最小限の音数は2個であり、連続する2音が作り出す暗意が実現/否定されるかは3音めで判定できる。つまり、暗意と実現/否定が生じる最小単位は連続した3音 (trigram) である。次に Narmour は、連続した2音の暗意は、音程と音の動く方向について生じると仮定した (方向原理, 音程原理 [13])。連続した3音をそれぞれ p_1, p_2, p_3 とすると、例えば、 $p_1 - p_2$ 間の音程が小さいと、 $p_2 - p_3$ 間の音程も小さく $p_1 - p_2$ 間と同じ方向が暗意される; $p_1 - p_2$ 間の音程が大きいと、 $p_2 - p_3$ 間の音程は小さく $p_1 - p_2$ 間と逆方向が暗意される等である。実際の旋律中には、このような暗意が実現される trigram もあれば、そうでない trigram も含まれる。

ゲシュタルトに基づいて得られた暗意に対する信頼度

(確信度)が高い場合は情動*2が喚起されないとする。一方、暗意の信頼度が低い、つまり曖昧な場合は情動が喚起される (もし暗意が得られなくなったら、そこはクロージャとなる)。そして暗意が得られた後に次の音を聴取した時、暗意通りの音だった場合は情動が喚起されず、暗意から少し遠い音は少しの情動が喚起され、暗意とは全く異なる音の時は情動が大きく喚起される。Narmour は、この暗意の実現と否定の出現パターンがイディオ構造に対応すると主張し、全部で8パターンの trigram を定義した。

6. タイムスパン木の簡約と情報量

GTTM におけるタイムスパン木の簡約とは、タイムスパン木のある高さ (図4中, "Level a"や"Level b"など) より下部を切り捨てることである。タイムスパン木は、楽曲に含まれる隣接する2音が勝ち抜き戦をして、より重要な音が勝ち残り、より重要でない音から削除される過程を表現したものである。つまり、タイムスパン木は、グルーピングと拍節に関してより重要な音という二項関係を元に、楽曲に含まれるすべての音に対して割り当てた半順序関係を表現している。平田らは、より長い範囲を支配している最大タイムスパンの方が、より多くの情報量を含み、ある音を簡約 (削除) するとその音の最大タイムスパン分だけ旋律全体の情報量が減るという最大タイムスパン仮説を提唱している [3]。Mozart 作曲「きらきら星変奏曲」(K.265/300e) を題材として用いた聴取実験により、人が知覚する旋律の近さ (心理的類似性) と旋律全体の情報量 (簡約距離) には良好な対応関係があることが確認されている。

7. 暗意-実現モデルにおける簡約

暗意-実現モデルの考え方を踏まえ、暗意-実現モデルにおける旋律の簡約をイディオ構造の割合が減少し、より特徴の無い旋律に変化することと仮定する [5]。特徴が無いとは、より暗意の信頼度が高くなり、暗意が実現する割合が増えることを意味する。暗意-実現モデルの8つの基本パターンの内、暗意が実現するのはD, P, Rの3つなので、簡約の操作は、残りID, IP, VP, IR, VRの5パターンをD, P, Rに変化させることになる。そのような操作を定義するため、旋律 p に加え、その1階微分 \dot{p} と2階微分 \ddot{p} を含む三つ組 $\langle p, \dot{p}, \ddot{p} \rangle$ を考える。楽曲の進行とともに、3次元空間内をこの三つ組が遷移するので、この三つ組が遷移する軌跡を仮想的な旋律と定義する (図7)。この時、 \dot{p} の値と \ddot{p} の値が0に近づくとき旋律の変化が乏しくなり、旋律の特徴が乏しくなると考えられる。つまり、 \dot{p} の値と \ddot{p} の値を0に近づけることが簡約の操作とみなせる。直感的には、時間軸にそって並んだ3音のパターンを類型化するため、旋律の2階微分までの値の正負や大小を考えたという

*2 Emotion; 分野によって多種多様な意味があるが、本稿では東条ら [13] に従う。

ことである。

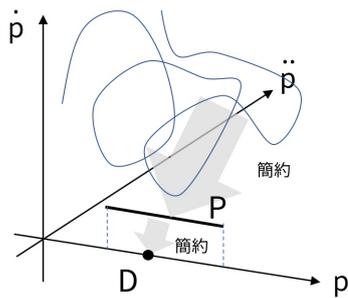


図5 三つ組が遷移する仮想旋律空間

しかし、この微分に基づく方法はまだ検討の余地がある。例えば、暗意の実現と否定を区別できない点や、簡約によって減少する旋律の情報量として \dot{p} の積分値と \ddot{p} の積分値が考えられるが、連続値の取り扱い、積分定数の取り扱いに難があるという点である。

別のアプローチとして、暗意-実現モデルの3音からなる基本パターンをベイズ推論によって解釈することが考えられる。ベイズ推定の式

$$P(S|O) = \frac{Q(O|S)}{P(O)} P(S)$$

を考えると、ここで、 S は暗意-実現モデルにおける基本8パターンの確率変数、 O は暗意に対して観測された音の確率変数とする。尤度関数 $Q(O|S)$ は、方向原理と音程原理に対応する。このアプローチでは、暗意の実現と否定を尤度関数の設計において区別することが可能であり、統計的な情報量を導入することが可能となる。

ドメインは異なるが、興味深いベイズ推定の応用例がある。3つの音が異なる時間間隔で聴こえる場合を考える。それら時間間隔の差が小さい時はより小さくなるように、大きい時はより大きくなるように知覚されるが、大きい時により大きくなる方がより過剰に知覚される現象が知られている。この非対称な知覚現象を説明するベイズモデルが澤井らによって提案されている [12]。

8. おわりに

音楽を聴取した時に生じる認知的ゲシュタルトからシェンカー理論が誕生し、階層的なグルーピングと直線的なグルーピングに基づいて、それぞれ GTTM と暗意-実現モデルが提案されたと述べた。GTTM のタイムスパン木に関して、その上位はグルーピング構造分析の結果から生成され、Narmour が指摘するスタイル構造やスタイル様式を表現するのに適していると考えられる。一方、タイムスパン木の低位は拍節構造分析の結果から生成されるが、拍節構造で各楽曲の特徴（イディオ構造）を表現するのは一般に難しい。そこで、タイムスパン木の低位を分析する部分、つまり拍節構造分析を暗意-実現モデルに入れ換えれば、あ

るいは暗意-実現モデルを追加すれば、旋律を下位レベルから上位レベルまで包括的に分析できる可能性がある。我々は、タイムスパン分析と暗意-実現モデルを統合する鍵は簡約だと考えている。さらに、暗意-実現モデルで下位の旋律を分析した時に算出される旋律の情報量が、タイムスパン木の情報量（総最大タイムスパン）と矛盾しないように設計できれば、旋律を包括的かつ統一的に分析できる可能性がある。

まず、本発表に続いて、山内らの発表 [14] にて、暗意-実現モデルに基づく旋律の情報量に関するこれまでの検討と、予備的な実験結果について報告する。

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] Deliège, I., Mélen, M., Stammers, D., and Cross, I., Musical Schemata in Real-Time Listening to a Piece of Music, *Music Perception*, Vol. 14, No. 2 (Winter, 1996), pp. 117-159, University of California Press (1996).
- [2] Dikken, N., The Cognitive Reality of Hierarchic Structure in Tonal and Atonal Music, *Music Perception*, Vol. 12, No. 1, pp. 1-25 (Fall 1994).
- [3] Hirata, K., Tojo, S., and Hamanaka, M., Algebraic Mozart by Tree Synthesis, *Proceedings of Joint ICMC and SMC 2014*, pp.991-997 (2014).
- [4] 深山覚, Post-Truth 音楽情報処理, 情報処理, Vol. 61, No. 5, pp. 484-485 (2020).
- [5] 平田圭二, 大村英史, 北原鉄朗, 旋律の微分と簡約の導入, 2016 年度 人工知能学会全国大会 (第 30 回) 論文集, 3G3-OS-15a-4.
- [6] Lerdahl, F., *Composition and Cognition: Reflections on Contemporary Music and the Musical Mind*, University of California Press (2019).
- [7] Lerdahl, F. and Jackendoff, R., *Generative Theory of Tonal Music*, The MIT Press (1983).
- [8] Narmour, E., *Beyond Schenkerism - The Need for Alternatives in Music Analysis*, The University of Chicago Press (1977).
- [9] Narmour, E., Some Major Theoretical Problems Concerning the Concept of Hierarchy in the Analysis of Tonal Music, *Music Perception*, 1 (2), pp.129-199 (1983).
- [10] Narmour, E., *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*, The University of Chicago Press (1990).
- [11] Parncutt, R., A model of the perceptual root(s) of a chord accounting for voicing and prevailing tonality, In Marc Leman (Ed.) *Music, gestalt, and computing: Studies in cognitive and systematic musicology*, Springer (1997).
- [12] 澤井, 佐藤, 時間順序による時間知覚の非対称性を説明するベイズモデル, 情報処理学会音楽情報科学研究会 SIGMUS119, No.36, (2018).
- [13] 東条敏, 平田圭二, 音楽・数学・言語 - 情報科学が拓く音楽の地平, 近代科学社 (2017).
- [14] 山内拓真, 能登楓, 竹川佳成, 平田圭二暗意-実現モデルに基づく旋律情報量の予備的検討, 情報処理学会音楽情報科学研究会 SIGMUS132 (2021).