

バーンスタインの「答えのない質問」再考： 計算論的音楽の理論の枠組みについて

Bernstein's "Unanswered Question" Revisited: The Framework of Computational Music Theory

平田 圭二*¹ 東条 敏*²
Keiji Hirata Satoshi Tojo

*¹公立はこだて未来大学
Future University Hakodate

*²北陸先端科学技術大学院大学
Japan Advanced Institute of Science and Technology

音楽と言語はともに、木構造をやりとりすることで意図の交換を行っているという意味において類似している。バーンスタインが1973年に行った音楽の構文についての講演から出発し、それ以降に提案された GTTM, I-R モデルなどの音楽理論の成果を取り入れ、音楽を計算の対象とするための理論基盤構築の試みを紹介する。

1. はじめに

人は言語を介してコミュニケーションを行う (図 1)。その時、人の視覚や聴覚は記号列から成るコンテンツを時間順で受け取り認識する。コンテンツを構成する記号列は、それを表出 (presentation) する人の意図を表現 (representation) している。受取側の人がコンテンツを理解するとは、表出側と同様な構造をコンテンツの記号列に割り当てることであり、これによって理解が成立したと考える。この構造は、チョムスキー (Noam Chomsky) を始祖として、階層的順序構造 (木構造) として長年論じられてきた。

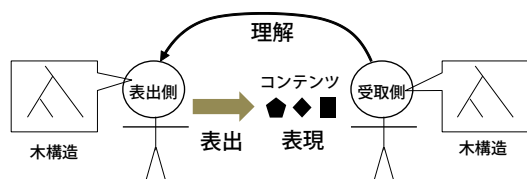


図 1: 言語を介した人のコミュニケーションと理解

音楽は言語機能を司る人の頭脳が産み出したものである。音楽も、言語と同じ器官を使って聴取、演奏されている。音楽聴取の認知的な仕組みと言語のそれの間には何らかの対応を仮定することができるのではないだろうか。

本稿では、このチョムスキー流の枠組みで、音楽の構造や意味理解を目指す研究を概観し、我々の研究プロジェクトを紹介する。

2. バーンスタインの「答えのない質問」

レナード・バーンスタインは1973年にハーバード大学にて「答えのない質問」という連続講義を行った [2]。特に Musical Syntax と題された第2回めの講義は、音楽の構造階層と自然言語の階層と対比させることで、音楽的な現象や構造の解釈法を提案した。バーンスタインは講演の冒頭で、音楽の起源は一つであり音楽の共通要素を追求したいと言う。そのために、チョムスキーが普遍文法を提唱し自然言語の起源は一つであると主張して自然言語の共通要素を追求した理論的枠組みを踏襲する。

連絡先: 平田圭二, 公立はこだて未来大学, 〒041-8655 北海道函館市亀田中野町 116-2, 0138-34-6462, 0138-34-6301, hirata@fun.ac.jp

まず、バーンスタインは、構文要素に関して、音楽と自然言語の間には図2のような対応があると言う。ここで、動機は数

音楽	自然言語
1つの音	= 音素
動機	= 音韻
楽句	= 語
楽節	= 節
楽章	= 文
楽曲	= 作品

図 2: バーンスタインによる言語と音楽の対応

音から成る象徴的あるいは主題となる旋律断片、楽句は1~数小節でまとまりのある部分、楽節は1つのカデンツァを含むような部分を指す。バーンスタインは1つの音が音素に対応し、楽節が節 (clause) に対応するとしている。これに対し筆者らの直感では、楽句が語に、楽章が文に対応するのは時間的に情報的に長すぎる。楽句はおおよそ節や文に対応するのではないかと感じる。

さらにバーンスタインは、品詞や様態に関して、動機は名詞の、和音は形容詞の、リズムは動詞の役割を果たし、短三度が否定を表現すると言う。音楽の生成過程に関しては、動機が生成変形文法における深層構造に対応し、移調、反転、入替え、融合、強調、埋め込みなどの変形規則によって表層構造 (我々が耳にする音楽) が生成されると言う。例えば、モーツァルト作曲 交響曲第40番 ト短調の動機は Eb-D-D であるが、これに変形規則が適用されて、表層構造において弦楽パートや金管パートなどの間に複雑な対位的な構造が現れると主張する。

バーンスタイン以外にも自然言語と音楽を対比させて、音楽の意味を考える試みは多い [7, 1, 14]。これらの議論は、音楽や自然言語に現れる文字や記号レベルでの対比、対応がとれそうな音楽の一部分を自然言語の枠組みの中 (あるいは一部分) に対応付けするレベルに留まっている。また、自然言語処理の技術を音楽分析に適用する試みもある [15, 4, 12] (その逆の試みは見当たらない)。これら対比や適用を行った結果、音楽理論の枠組みに拡張や修正を加える提案にまで至る例は少なく、その逆もほとんどない [9, 3]。

我々の問題意識は、自然言語が計算の対象となったように、音楽を計算の対象とすることである。そのために、音楽と自然言語の対比だけでなく、音楽理論と自然言語処理の比較や情報学の知見に基づき、計算の理論としての音楽理論を構築することを目指す。

3. 音楽を聴取する際の2種類の認識

音楽の何が計算対象となるのかを考えるために、まず音楽はどんな意味を持つのかを考える。音楽聴取時の認識には、予測できないピッチイベント*1と予測できるピッチイベントがある [8, 10, 11]。まず最も単純な場合として、ドドドド…と続く旋律を聴くこと考える。この旋律がどこまで続くか、どこで途切れるかを予測する(期待する)ことは難しい。しかし、聴取者にとって突然ドが鳴らない時点が来ると(ドドド休ド…)その時点でゲシュタルトつまりグループの境界を認識する可能性が高くなる。ところが、この旋律を例えば4分の4拍子のリズムに乗せて聞くと、4拍ごとにグループ境界がやってくることを予測するのは容易になる。

他の例としてドレミファ…と上昇する旋律を聴く場合を考える。この旋律もどこまで上昇するか、どこで上昇が止まるかを予測するのは難しい。この場合も、下降した音が鳴った時点で初めて上昇が止まったことを認識し、そこにグループ境界を認識する可能性が高くなる。

もう1つの例として、I→V→Iという和音進行を聴くことを考える(小学校で授業が始まる時に聴く「起立、礼、着席」の和音進行である)。我々の多くは無意識の内に、I→V(起立、礼)まで聴いた時点で、次にI(着席)が来ることを期待するだろう。期待するとは、次にI(着席)が鳴ることを待つあるいは予測するという意味である。人は、そろそろこの楽曲は終わる、さらに楽曲は展開するなど感じながら楽曲を聴いている。この予測できるピッチイベントに関する認識は、一般には、終止感や浮遊感などと呼ばれ、安定と不安定の間を遷移する認識・感覚である(安定と不安定の途中段階の認識もある)。

モーツァルト作曲ピアノソナタイ長調 K.331 では、最初の4小節に半終止を持ち、後半4小節に完全終止を持つ(図3)。聴取者は、4小節目を聴取している時点で楽曲が終わることを期待するが、5小節目に1小節めと同じテーマが突然始まりその期待が裏切られる。この時点で、4小節目の終わりとは5小節目の始まりの間に予測できなかったグループ境界が認識される。同時に、聴取者は(さらに4小節あとに)楽曲の終わりを予測しより強く期待するようになる。



図3: K.331の和声分析(文献[7, p.135]より転載)

Generative Theory of Tonal Music (GTTM)[7]は、音楽聴取で生じるこの2種類の予測できないピッチイベントに関する認識と予測できるピッチイベントに関する認識を分析する音楽理論である。次節以降で説明するタイムスパン木は、ゲシュタルトに基づく予測できない認識(グループ境界)から作られる楽曲構造を表現するものである。この予測できない認識は生得的であると言われている[11]。延長木は予測できる安定と不安定の間を遷移する認識から作られる楽曲構造を表現するものである。延長木として認識される楽曲構造は経験や学習から獲

*1 音高(ピッチ)を感じさせる音のこと。つまり、音楽を構成する個々の一音、時間的に最も短い音楽の構成要素であり、和音も含まれる。

得られるものであり、したがって、予測できる認識は後天的であると言われている [3]。

4. タイムスパン木

譜面とは、どの時刻にどの音高で音を鳴らすかを時間と音高の2次元平面上に記述したものである。人が音楽を聴取する際、音高方向と時間方向に2種類のゲシュタルトが生成され、それが音楽認知の基本を形作る。タイムスパン木(time-span tree, TS木)とは、そのような2種類のゲシュタルトからボトムアップに作られる時区間の階層構造を表現する木構造である(図4)。各時区間にはその時区間を支配するピッチイベント(局所的な調)が関連付けられており、headと呼ばれる(言語理論での主辞に相当する)。

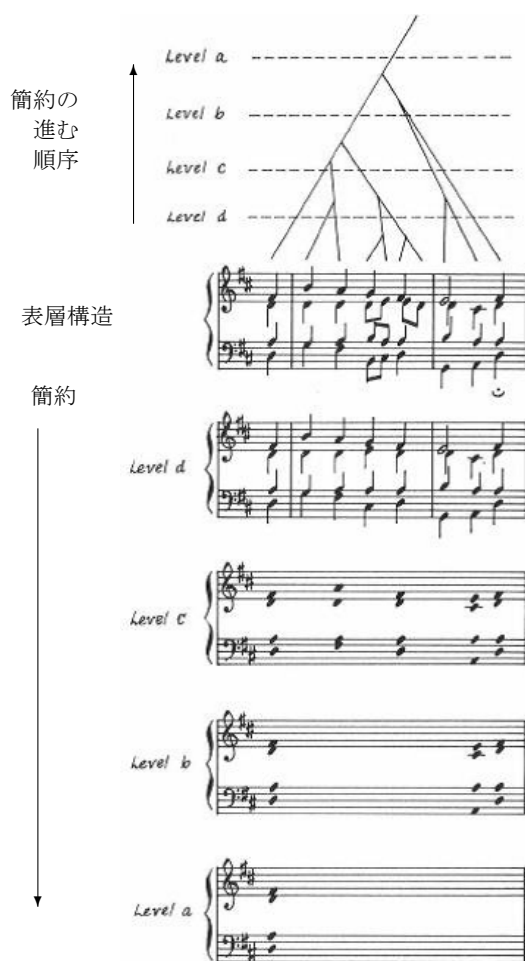


図4: TS木簡約の例: J.S. バッハ作曲 コラール “O Haupt voll Blut und Wunden” in St. Matthew’s Passion [7, p.115]

TS木の部分木は、直感的に、動機、楽句、楽節、楽章などの楽曲構造に対応する。TS木では、2つの隣り合った時区間が2つの枝で表現され、より上位ではその2つの時区間が1つに併合される。時区間はボトムアップに併合され、最終的に楽章程度の長さまでなる。1つの時区間は、その両端の境界から決められるが、境界は音高の差や時間の差、アーティキュレーションの差などから判断される。

TS木の隣り合う2つの時区間がボトムアップに併合された場合、いずれの局所的な調が併合された時区間を支配する局所的な調(head)となるかを定める必要がある。2つの時区間に

関連付けられた局所的な調に関して、優勢あるいは重要という概念を導入する（逆に、一方は他方に従属するとも言う）。対応して、TS木の1つのノードから延びる2つの枝に関して、優勢な枝を primary な枝と呼び、そうでない枝を secondary な枝と呼ぶ。一般に primary な枝が head の情報をもたらす（アルペジオのように primary/secondary が簡単に決まらない場合もある [7, p.154]）。

GTTMのグルーピング構造と拍節構造が与える情報は、旋律に含まれるどの音がグループを作るか、あるいはどこにグループの境界があるか各音のいずれが重要な音なのかである。これらの情報をもとに、優勢あるいは重要なタイムスパンと head を選んでいく。こうして、TS木を生成する時は、まずグルーピング構造で境界を決め、そこから head を選ぶという2段階を経る。

5. 階層構造と簡約

情報学や数理論理学における簡約 (reduction) とは、項をより単純な形に書き換えることである。自然言語は統語に関わる生成規則が強く働き、ある部分木の中のヘッドを決める際には、その子カテゴリーの中から一意にヘッドに最も寄与するカテゴリーを決定できる。このようなカテゴリーがどのような部分木に様に存在することは X バー規則 ($\bar{X} \rightarrow Y X$) によって保証されるため、自然言語の構文は階層的な木となる（階層的だから子 Y を削除すれば単純な文が得られるという意味ではない）。

一方、TS木の簡約とは、TS木というドメインの上で、重要でないタイムスパン（時区間）から順番に削除していく操作である。削除前のTS木と削除後のTS木の間には、半順序関係が成立する。図4の例では、表層構造が level d \rightarrow level c \rightarrow level b \rightarrow level a と簡約されていく様子が描かれている*2。自然言語の場合は生成的な文法規則が陽に存在するので、虚辞、相槌音、間（ま）のような削除可能な重要でない要素が極端に少ない。対して、音楽の「擬」生成規則は遥かに自由度が高い構文を生み出すので、より重要でない枝を削除する簡約という操作が意味を持つと思われる。

ここで、TS木の簡約と楽譜の簡約は異なる点に注意が必要である。楽譜に記された旋律を GTTM 分析することでTS木が得られ、TS木をレンダリング*3することで実際に聴取可能な楽曲が得られる。TS木には楽譜に記された旋律以上の音楽構造に関する情報が表現されている（例えば、head やグルーピングの階層構造）。一方で、旋律には音符 (onset, duration) や休符に関する情報が表現されている（TS木を構成する時区間には onset や duration の概念はない）。つまり、タイムスパンと実際に人が聴取できる音は異なる概念である。例えば、図3の旋律において、最長のタイムスパンは8小節の長さを持つが、対応するピッチイベントは8小節目2拍半のA major である。

筆者らは maximal time-span という概念を導入し、TS木の簡約が半順序関係となるように定式化した [13]。いずれのピッチイベントも、そのイベントが他のイベントを従えて最も支配的に働く時間幅を仮定でき、それを maximal time-span と呼ぶ。ある音が簡約によって削除された時に失われる情報量は、削除された音の duration ではなく、この maximal time-span に等しいと定義した。そして、半順序関係 $A \sqsubseteq B$ なる2つ

*2 GTTM では、簡約の逆の操作を精緻化 (elaboration) と呼んでいる。

*3 Rendering. もともとはCG用語であるが、ここではTS木から実際の音楽（楽譜に記された旋律）を生成することを意味する。

の旋律 A, B がある時、 A, B 間の距離は、 B から A になるまで削除される maximal time-span の時間幅の総和と定義した ($|B - A|$ と書く)。TS木 T 全体が持つ情報量は、空要素 (bottom) \perp との距離であるから、全 maximal time-span の総和となる ($|T|$ と書く)。そして、任意の旋律 P, Q 間の meet 経由の距離 d_{\sqcap} を $|P - P \sqcap Q| + |Q - P \sqcap Q|$ と定義し、join 経由の距離 d_{\sqcup} を $|P \sqcup Q - P| + |P \sqcup Q - Q|$ と定義した (図5)。その結果、筆者らが定式化した距離の性質に関して、 $d_{\sqcap}(P, Q) = d_{\sqcup}(P, Q)$ が成立することと、距離の公理の1つである三角不等式を満たすことを示し、被験者実験により認知的リアリティがあることを確認した [5]。類似度は距離の逆数として定義できる。

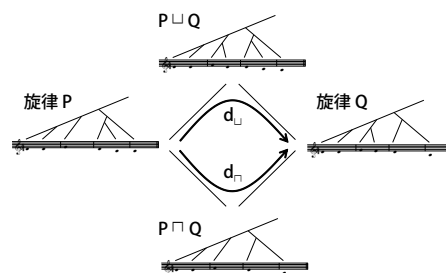


図5: meet 経由の距離と join 経由の距離

6. 延長木

第3章で触れた予測できる安定と不安定の間遷移は、緊張 (tension, 安定 \rightarrow 不安定) と弛緩 (relaxation, 不安定 \rightarrow 安定) の2つに分けられる [7]。緊張を引き起こす原因には、不協和音、旋律中の音程の大きい箇所、上昇音列が下降に転じる箇所、根音が五度圏において離れた和音の出現箇所、聴取者の期待の裏切りなどがある。これら原因の逆の現象は弛緩を生じさせ安定に戻る。楽曲の進行を予測するということは、緊張のあとには弛緩が来るという期待を持つことである。

ある程度の将来に聴こえてくるピッチイベントがもたらす認識を期待するという事は、これまで聴いた音を何らかの型やパターンにあてはめて将来を予想しているということである。その型やパターンは、教育や他の場所での聴取経験から獲得するか、今聴いている音楽そのものから獲得するかのいずれかである。前者は西洋調性音楽を聴取する場合に多く、後者は非調性音楽を聴取する場合に多い。特に、西洋調性音楽を聴取する経験や教育によって獲得された型やパターンは、多くの作曲者と聴取者の間で共有されている [3]。次に聴こえるであろう音を予想させる音列は暗意 (implication) と呼ばれ、予想通りに聴こえた音は実現 (realization) と呼ばれ、予想通りでなかった音は裏切り (denial) と呼ばれる [10]。

GTTMの延長木 (prolongation tree, PR木) は、暗意、実現、裏切りによって引き起こされた緊張弛緩の構造を表現する。図3と同じ楽曲 K.331 でのPR木を図6に示す。4小節目Vは5小節目冒頭で終止するという予測をもたらす一方で、実際は5小節目冒頭 (図中の☆) でIが聴こえ (denial)、遡及して (retrospective) 4小節目は半終止であるという認識に到る。半終止そのものは展開 (departure) により緊張度を増加させ、5小節目冒頭の音は繰返し (repetition) により緊張度をさらに増加させる。こうして1小節目から増え始めた緊張は☆の時点で最高に達する。その後は8小節目のV-Iのカデンツァに向けて弛緩していく。このように、緊張は4小節目と5小節

めの間にある TS 木のグループ境界を越えて 5 小節め冒頭の I まで延びることから、延長木 (prolongation tree) という名前が付けられている。

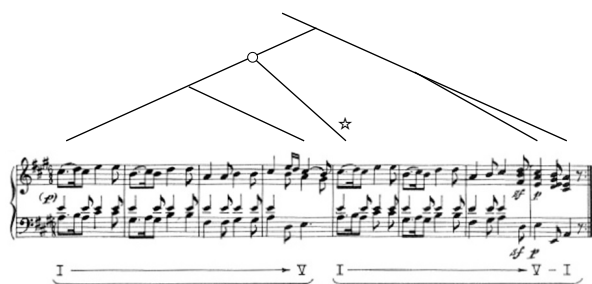


図 6: K.331 の PR 木 (文献 [7, p.224] より改変)

人は聴いた音全てを長時間 (数分以上) に渡り正確に記憶することは困難なので、長時間に渡る緊張と弛緩のパターンをガイドとして楽曲を記憶する技法を開発した。それが楽式として共有・定着し、長時間に渡る楽曲の創作と鑑賞を可能にした*4。PR 木はこの楽式という表出のための型やパターンを表しているとみなしてよく、normative form, basic form [7, p.188] と呼ばれる。

GTTM は、TS 木を修正して PR 木を作るという手順を与えている。TS 木は生得的な認識からボトムアップに導かれ、PR 木は後天的な学習からトップダウンに導かれることを考えると、TS 木のグループ境界と PR 木のグループ境界が大きく食い違う場合もあろう。しかし、GTTM が想定している分析対象楽曲は和声理論が整った直後の古典時代 (classicist) の曲であり、楽式に則り適切にバランス良く構成されているものばかりである。おそらく、楽式によるトップダウンのグループ境界とボトムアップのグループ境界の食い違いは小さいと思われる。一方、TS 木として重要なピッチイベントの選択と PR 木として重要なピッチイベントの選択は大きく異なっている。したがって、TS 木をベースにすることでグループ境界の情報を引き継ぎ、PR 木として重要なピッチイベントの選択を行うことで、多くの場合に正しい PR 木を得られると考えられる。

7. おわりに

音楽理論と自然言語処理を対比させながら TS 木の簡約の定式化を行った。1 つの応用として、旋律モーフィングのアルゴリズムが、2 つの旋律の幾何学的内分点に相当することを証明した [6]。本アプローチにより、音楽を計算の対象とするための理論的基盤を構築できると考えている。さらに、形式的かつ実用的なオペレータを導入していきたい。

PR 木にも簡約の概念があるが [7]、PR 木にも TS 木と同様の距離を導入するには、まず緊張・弛緩を表現する定量的な尺度を定義する必要がある。認知的リアリティを確認しながら追求していきたい。

謝辞

本研究は JSPS 科研費 23500145 及び 25330434 の助成を受けたものです。

参考文献

- [1] Aiello, R.: 音楽と言語 - 類似点と相違点, リタ・アイエロ編, 大串健吾監訳, 音楽の認知心理学, pp.46-71 (1998).
- [2] Bernstein, L.: Lecture II, Musical Syntax, in "Unanswered Question", Norton Lectures (1973).
- [3] Cook, N.: 知覚 - 音楽理論からの展望, リタ・アイエロ編, 大串健吾監訳, 音楽の認知心理学, pp.72-110 (1998).
- [4] Cope D.: *Experiments in Musical Intelligence*, A-R Editions, Inc (1996).
- [5] Hirata, K., Tojo, S., Hamanaka, M.: Cognitive Similarity grounded by tree distance from the analysis of K.265/300e, *Proceedings of CMMR 2013*, pp.415-430 (2013).
- [6] 平田圭二, 東条敏, 浜中雅俊: 旋律モーフィングアルゴリズムの形式的検証, (社) 情報処理学会 音楽情報科学研究会, 2010-MUS-85, No.4 (2010).
- [7] Lerdahl, F., Jackendoff, R.: *A Generative Theory of Tonal Music*, The MIT Press (1983).
- [8] Meyer, L.B.: 音楽における情動と意味, リタ・アイエロ編, 大串健吾監訳, 音楽の認知心理学, pp.3-45, 誠信書房 (1997). *Emotion and Meaning in Music*, University of Chicago Press (1956) 翻訳.
- [9] Minsky, M., Laske, O.: A Conversation with Marvin Minsky, In *Understanding Music with AI*, The MIT Press (1992).
- [10] Narmour, E., *The Analysis and Cognition of Basic Melodic Structure - The Implication-Realization Model*, The University of Chicago Press, Chicago, 1990.
- [11] Snyder, R.: 音楽と記憶 - 認知心理学と情報理論からのアプローチ, 音楽之友社 (2003).
- [12] Steedman, M.: The Blues and the Abstract Truth: Music and Mental Models. In A. Garnham and J. Oakhill, (eds.), *Mental Models In Cognitive Science*. pp.305-318. Mahwah, NJ: Erlbaum (1996).
- [13] Tojo, S., Hirata, K.: Structural Similarity Based on Time-span Tree, *Proceedings of CMMR 2012*, pp.645-660 (2012).
- [14] Wiggins, G.A., Müllensiefen, D., Pearce, M.T.: On the non-existence of music: Why music theory is a figment of the imagination. In: *Musicae Scientiae*, Discussion Forum 5, 231-255 (2010).
- [15] Winograd, T.: Linguistics and the computer analysis of tonal harmony, In *Journal of Music Theory*, vol.12, no.1 (1968).

*4 図 1 において、意図を込めたコンテンツを受取側に表出する時の起承転結や芸能における序破急に対応する。