

統計的文法理論と構成的意味論に基づく音楽理解の計算モデル

東条 敏^{1,a)} 平田圭二² 浜中雅俊^{3,b)} 長尾 確^{4,c)} 北原鉄朗^{5,d)} 大村英史^{7,e)} 松原正樹^{f)}
吉井和佳^{3,g)}

概要: 本研究では楽曲に木構造を付与する理論 GTTM を土台にし、自然言語の統計的文法理論に基づく楽曲の構造解析システム (パーサ) を作成する。次に和声を整合的に接合する規則を与え、遠隔の依存関係 (照応関係) を付加する制約ソルバーによって構文木をアノテートすることにより、形式的意味表現を生成する。われわれはこの構造解析器・意味生成器を用いて楽曲を解析・レンダリングするとともに、木構造変換という代数的操作により編曲と類似度計算をする応用システムに展開し、解析例データとともにウェブにて公開する。さらに形式的意味表現の授受によって音楽の認知的意味をモデル化することを提案する。

キーワード: 音楽情報処理, 音楽学, 文法理論, 木構造, 認知モデル

Computational Model of Music Recognition, Based on Statistical Grammar Theory and Constructive Semantics

SATOSHI TOJO^{1,a)} KEIJI HIRATA² HAMANAKA MASATOSHI^{3,b)} KATASHI NAGAO^{4,c)}
TETSURO KITAHARA^{5,d)} HIDEFUMI OHMURA^{7,e)} MASAKI MATSUBARA^{f)} KAZUYOSHI YOSHII^{3,g)}

Abstract: This research investigates the integration of statistical grammar theory and constructive semantics for music. The Generative Theory of Tonal Music has suggested to construct a tree structure to represent an internal dependency in a music piece. We develop this idea, assuming a statistical grammar rules, and aim at constructing semantically annotated tree. In this framework, we apply the structure to music analysis, rendering, arrangement, similarity, and archive such results to be openly accessed on the web. Also, we consider this annotated tree to be a cognitive model of our music understanding.

Keywords: music information processing, musicology, grammar theory, tree structure, cognitive model

¹ 北陸先端科学技術大学院大学 JAIST
² はこだて未来大学 Future University Hakodate
³ 京都大学 Kyoto University
⁴ 名古屋大学 Nagoya University
⁵ 日本大学 Nihon University
⁶ 筑波大学 University of Tsukuba
⁷ 東京理科大学 Tokyo University of Science
a) tojo@jaist.ac.jp
b) masatosh@kuhp.kyoto-u.ac.jp
c) nagao@nuie.nagoya-u.ac.jp
d) kitahara@chs.nihon-u.ac.jp
e) hidefumi.ohmura@gmail.com
f) masaki@slis.tsukuba.ac.jp
g) yoshii@kuis.kyoto-u.ac.jp

1. はじめに

われわれはここ半世紀で初めて、われわれ自身の認知過程を顧みる鏡を手に入れた。計算機である。いったん実世界から隔絶されたコンピュータの世界の中で、プログラミング言語で形式化された知識とその推論方式によってわれわれ自身の心や脳の働きが模倣可能であるのか。ところが音楽の「形式化」に取り組んでみると、それが実にわれわれの脳と心の中に深く根を張っていることに気づく。

言語と音楽は生物学的に同ルーツ [17] であり、咽喉を

使った発声と耳による聴覚によってメッセージ授受が行われる。あらゆる生物の中で人間のみが言語と音楽を分離したが、それぞれは収斂進化である可能性があるにしても、多くの類似性 [2] が指摘される。よって言語はわれわれが音楽を理解する手立てとなりうる [21]。すなわち問うべきは、われわれが音楽を研究するために音楽のある側面を言語化できるのか、すなわち構造化・記号化できるのかという疑問である [13]。



図 1 タイムスパン木の構成 [9], p.132

われわれは GTTM[9] を用いて楽曲の木構造を生成する研究に成果を上げてきた [5], [16]。この理論は重要な音を階層的に選択することにより楽曲に木構造を付与するものである。各音の重要度は楽曲のグループ構造 (楽句) と拍節構造 (リズム) から決定され、ボトムアップに隣接する音と重要度を比較することによりタイムスパン木を得る (図 1)。各和音は曲の後方へも影響を残すが、この支配関係を加えてタイムスパン木を改編するのがプロロンゲーション木である。

一方、自然言語処理には長年の研究の蓄積があり、チョムスキー (Noam Chomsky) を始祖として構成的な文法規則と論理的な意味表現が多様に提案されてきた。このため楽曲に文法を与えようという試みは 1960 年代人工知能 [18] にまで遡る。今日われわれは新たにウェブ上に豊富な音楽の電子データを手にし、その構造生成・意味理解においては先行する自然言語処理の研究成果を援用すべき機にある。

音楽意味の形式化を拒んでいるのはその曖昧な概念設定による。われわれはまず「意味」の意味を峻別する。

音楽の認知的意味論 情動、すなわち音楽を聞いた人間が起こした心の動き。

音楽の構成的意味論 楽曲が階層の木構造をなすと仮定したときに、構文木に平行してパーツの意味の組み合わせからなる全体意味。

これら二つの意味においては、情動は単音ではなくある長さの楽句を聞いて起きるものであり、同様に形式的意味表現はパーツを統合して得られることより、ともにゲシュタルト意味と考えられる。ここでゲシュタルトとは個々の構成要素 (音楽で言うと一個の音) には意味を見い出せないが、それらを統合することによって初めて認識できる全体像を指す。Heinrich Schenker (1868–1935) によれば、音楽は表面上多数の音符からなるが、これを前景 (Vordergrund) とし、重要な音に関心を絞る簡約化原理 (Reduction Hypothesis) により中景 (Mittelgrund) を経て後景 (Hindgrund) と呼ばれる楽曲の原構造 (Ursatz) に到達すると考える [14]。この構造は原メロディ (Urlinie) が付随することにより、初めて楽曲として認識できるゲシュタルトとなる。

図 2 は音楽のゲシュタルトを認知面、形式面の 2 方向から捉えたものである。楽譜における個々の音符は何も暗意しないが、水平方向・垂直方向に結び付くことによってそれぞれメロディと和声をなし、それらの構造は木で表現できる。同様に、演奏から導き出される音響は脳の中で短期記憶によって時間軸に沿って結びつき、音楽として認知されることになる。*1

2. 本研究のめざすもの

われわれはこれまでの研究により図 1 にあるようなタイムスパン木の生成を自動化 [5] した。しかし依然 GTTM の生成規則には恣意性があり、より高い精度で木構造を確定するには、確率文脈自由文法による規則発見を行う必要がある。また同時に木構造の安定性を議論するためには和声進行を推測してその情報によって木を改編する必要がある。さらに楽曲内には、必ずしも明示されていない遠隔の依存関係、例えばグループ構造・繰返し構造・対称構造などが内在している。これらを楽譜内における広義の参照関係と捉えると、顕在化されていない参照関係の発見という作業はゲシュタルト意味の構成と考えられる。これら主観を交えない (sachlig な) 情報を楽譜から正確に取り出し木構造に付加する必要がある。

われわれは確率文脈自由文法で構成される楽曲の解析木に和声連鎖の情報と楽曲内の遠隔の依存関係をタグ付けされた木を設計し、もって楽曲の形式的意味表現とする。木構造に基づく形式的意味表現に対しては、音符を時間的重複なくアサインして曲生成 (レンダリング) が可能である。木構造は解析結果の情報を含む意味構造というだけではなく、決定性のあるアルゴリズムによって木の改編・編集・比較を行うことを可能にする。すなわち、楽曲の代わりに木構造を代数的に操作するにより多様な実用システムを

*1 楽譜はさまざまな演奏を集めたジェネリックな型、逆に演奏は型の現実世界へのアンカーと考えると、楽譜と演奏の関係が状況意味 [1] である。

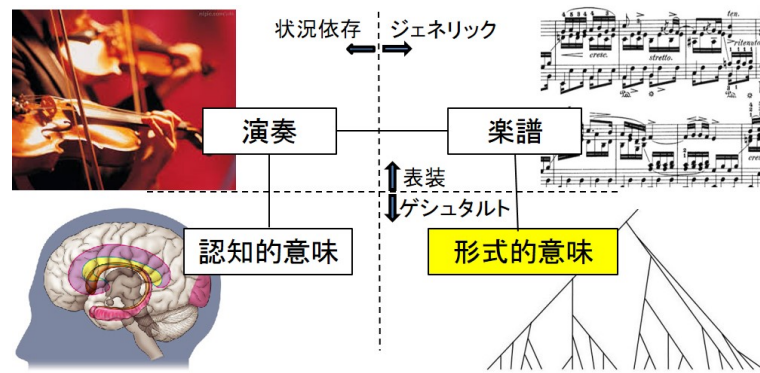


図 2 音楽の意味

構築することが可能となる。

さらにわれわれは形式的意味表現による多数の楽曲の解析例を公開し、さらに木を編集する手法で編曲システム、木間の距離を定量化することにより類似度を定義し検索システムを構築する。

言語コミュニケーションにおいては話者と聴者が同じ構文木を構成でき、照応詞に同じ参照関係ができた場合に相互理解ができたと考える。同様に、音楽的においても奏者と聴者の間で受け渡しされる認知的意味はアノテートされた木構造であると考え、奏者と話者の木の構造変化・アノテーションの変化が解釈は多様性・恣意性と捉えられる。これにより、われわれは情動 [19], [20] をモデル化し、論理的に議論できる対象とすることを企図する。

本研究の独創的な点は、第一に近年の統計的手法による人工知能研究の成果を前面に出し音楽に明確な確率文脈自由文法を設定することにある。第二にこれまで得られたタイムスパン木を構文木とみなすのみならず、それに和声連鎖情報と依存関係（参照関係）を付加して楽曲の構成的意味とすることである。

3. 要素研究と全体構想

研究は次のように 4 期に分ける。(i) 分析期においては統計的文法理論の設定とそれによるパーサの構築、および GTTM アナライザに和声情報を組み込んだ改築を行い、形式的意味表現をデザインする。(ii) 統合期においては文法理論と意味構造の統合を行い、それを実現する制約ソルバーの開発を行う。(iii) 生成期では文法操作・意味操作による楽曲生成を提案する。(iv) 応用期においては木構造の代数的操作を用いた応用システムの開発を行い、成果をオープンソースとしてウェブ公開を行う。これらの期分割はもちろん厳密なものではなく、特に認知的意味の形式化モデルおよび楽典の記述論理による形式化は期をまたいで逐次実装する。この過程を図 3 に示す。

3.1 音楽における文法・生成論的音楽理論

和声進行カデンツ規則は $C \rightarrow TDT, C \rightarrow TST, C \rightarrow$

$TSDT$ *² のように文脈自由規則で記述できるが、和声進行解析そのものに曖昧性があること、および自然言語に較べて音楽は途中進行が自由であり、これを許容するような文法規則が必要である。この二点に鑑み以下のテーマを提案する。

3.1.1 確率付き文脈自由文法の獲得

文法概念とは規範ではなく語並びの統計的特徴と考える。特に音楽は文法規則はア priori に与えられるものではなく、データから統計的に発見されるものとするのが正しい。本テーマでは自由度を含む文法規則を PCFG あるいは TAG とし、木構造予測に用いる [11]。確率モデルは n-gram, LDA とする。*³

3.1.2 組合せカテゴリー文法

CCG*⁴ は Steedman ら [4] によってジャズ和声の進行規則に適用された。この記法の利点は遠隔の係り受け、すなわちカデンツの解決に遠方のトニック和音を用いることを可能にしていることである。同文法の生成能力は文脈自由文法を包含するため、上記 PCFG との相性も良い。図 6 は CCG により最上部の和音列から下方に向かってカテゴリーが結合される例を示す。

3.1.3 メロディー予測

GTTM による重要音の選出はボトムアップに隣接するピッチイベントどうしを比較するために、木の上層部に現れる構造において残るピッチイベントが均一な重要度をもつかどうかは保証できない。このため、ウェイブレットを用いて旋律の動きを取り出し、木上層部との比較を行う [8]。さらに、I-R モデル [12] と組み合わせるとメロディー進行予測に拡張する。

3.1.4 GTTM の拡張による参照関係発見

タイムスパン木の構成においては楽曲終結部（カデンツ）における cadential retention の発見アルゴリズム [10] にはまだ課題がある。加えてタイムスパン木からプロロンゲー

*² C はカデンツ, T, S, D はそれぞれ Tonic, Subdominant, Dominant を意味する。

*³ Probabilistic Context-Free Grammar (PCFG), Tree Adjoining Grammar (TAG), Latent Dirichlet Allocation (LDA)。

*⁴ Combinatorial Categorical Grammar (CCG)。

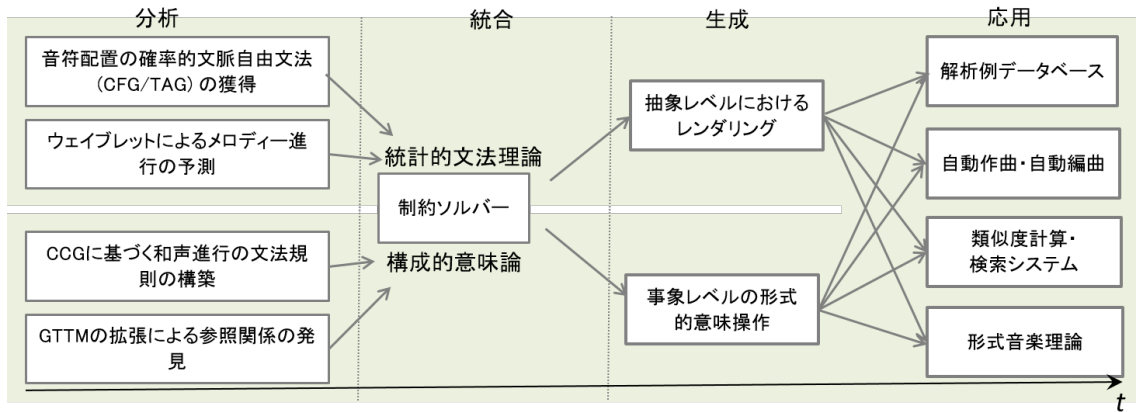


図 3 全体構想

$$\begin{array}{c}
 \frac{G_m^7 \quad C_m^7 \quad B^7 \quad E_b^7 \quad B_b^6}{G^D/C^D|T \quad C^D/F^D|T} \\
 \frac{G^D/F^D|T \quad F^D/B_b^D|T}{G_D/E_b^D|T \quad B_b^D/E_b^D|T} \\
 \frac{G^D/E_b^D|T \quad E_b^T}{G^D - E^T}
 \end{array}$$

図 4 カテゴリー文法によるカデンツ構成

ション木生成を行う過程も曖昧さを含んでいる。本研究では、これらと和声支配の時間的延長、繰り返し構造・対称構造の発見、複数メロディーラインの同時共存（ポリフォニー）などマクロな視点に基づく木の改編をオリジナルのタイムスパン木に付加し、プロロンゲーション木を内包し広義の参照関係をアノテートする木を提案する。

3.1.5 ディープラーニングによるパラメータ・フリーの GTTM 解析器

このテーマでは機械学習による木の構成手法について検討する。タイムスパン木の形は、実際は音型の進行と拍の位置によってほぼ決まる。したがって GTTM の各規則に与えるのではなく、多層ニューラルネットワークの出力層に近いところに従来の GTTM 生成規則に相当する教師データを与えれば、パラメータを手手でチューニングすることなくタイムスパン木の獲得を学習できることになる [6]。

3.2 形式的意味表現

確率的に予測された構文木 (図 5) は、上部構造に意図したメロディ概形が構成されているかどうかにより検証可能である。

この結果と和声解析結果を結びつけ、Montague 文法 [3] 同様、構文木と平行して部分意味構造（ラムダ式による述語-項構造）を組み立てることにより全体意味を構成するという考え方*5に基づいて意味構造を設計する (図 4)。

*5 西洋近代科学の根幹である還元主義 (reductionism) に基づくものであり、ホーリズム (holism) を廃する考え方。

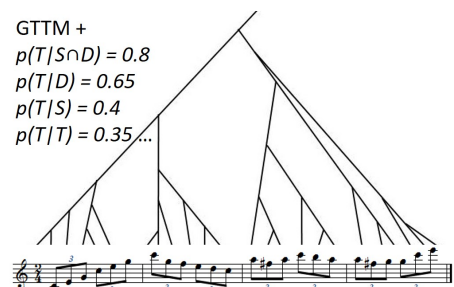


図 5 確率文脈自由文法による木構成

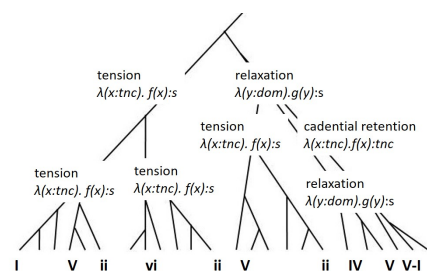


図 6 参照関係をアノテートした木の構成

図 6 は意味をアノテートした木の例である。ここではタイムスパン木（構文木）に和声進行のラムダ結合を平行させてタグ付けし、さらにグループ構造や繰り返し構造などマクロな参照関係がアノテートされている。

楽曲パーサによる構造予測はピッチ・和声の情報と整合的に統合される必要がある。さらに拡張された GTTM によって発見された楽曲内のマクロな構造によって木を改築する必要が生じる。このプロセスは無矛盾性検証と制約統合であり、この制約ソルバー (constraint solver) を構築する。

3.3 レンダリングと形式意味の操作

3.3.1 タイムスパン木のレンダリング

タイムスパン木は音間の優位関係を明示したという意味で、それだけリッチな情報を含んだ木と言えるが、逆に個々

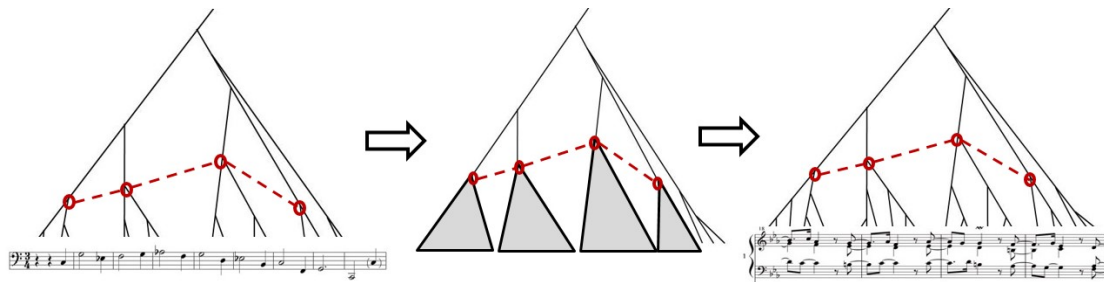


図 7 上位構造の保存と部分木付け替えによる編曲

のピッチイベントの情報（音高、音長、および音色）を陽には含まないため、木から楽譜へのリアリゼーション（レンダリング）には恣意性が残る。このため、木に整合的にピッチイベントをアサインするアルゴリズムを開発する。

3.3.2 認知的意味の形式化モデル

これまで情動という曖昧なことばで語られてきた認知意味について、われわれは形式的意味表現からのアプローチを試みる。一般に演奏者（メッセージの発信者）と聴者（受信者）間で木構造の同一性/相違は保証されない。特に音楽聴者のメロディ予測においては、予測が当たることと裏切られることに情動を生じるが、これは部分木を受け取って残りの木をどう予測するかという意味木の構造把握の問題に帰着し、メロディ予測の認知的意味をモデル化する(図8)。



図 8 メロディー予測の認知的意味

3.4 応用システム

最終フェーズには応用システムに着手する。これらのシステムの特徴は木構造に対して代数的な操作を定義することによって実現されるものである。

3.4.1 タイムスパン木構成アプリの plug-in 提供

われわれは既に GTTM による楽曲解析システムを公開しているが、多数のユーザを持つアプリに plug-in として提供することで、この分野の研究者層を拡大するとともに、楽曲の解析例をより多数蓄積することに寄与してもらうことを企図する。

3.4.2 解析例データベース

われわれは開発したパーサと制約ソルバーを用いて解析例を蓄積し、ウェブに公開する。われわれは過去 GTTM

によるタイムスパン木による解析例公開の経験があるが、IMSLP(クラシックの楽譜ライブラリ)、The Essen Folklore Collection(フォークロアのデータベース)、ジャズのスタンダードなどより幅広いジャンルから曲を集め解析を行い、評価を問う。

3.4.3 編曲システム

既存の曲に対し調・リズム・ピッチの変更によって新たな楽曲の創造を行う。作曲は完全に個人の頭の中で行われるだけではなく、一度木構造に変換したドメインの中の代数的操作によって、階層的構造を保存しながらも新規性が高い楽曲の創成が期待できる。われわれはクラシック音楽における変奏曲に 응용を試みた実績 [7] があるが、このプロジェクトではパッサカリア*6 を対象に完成度の高い変奏曲を提示することを目標とする(図7)。

3.4.4 類似度計算システム

検索システムの本質は類似度評価である。楽曲の表装に捉われることなく、重要な音を高く評価する階層型木構造においては、木構造の編集距離(ex. Levenshtein 距離など)により楽曲間の類似性評価 [15] が可能である。このことから、木構造の精度を高め、類似性検証のアルゴリズムを研磨することで、より信頼性の高い検索システムを実現する。

3.4.5 形式音楽理論

音楽はきわめて数学的な理論体系を有しながら形式的記述への努力は希薄であった。本研究においては音楽理論の記述論理(description logic)による記述を試みる。一般に楽曲 S は T (tension), R (relaxation) の連鎖 $S \rightarrow TR$ であるとし、再帰的に $T \rightarrow TS$, $R \rightarrow SR$ を定義する。 T は一般にドミナントへの進行, R はトニックへの進行である。よってこの形式知識記述においては Tension/Relaxation の文法規則の精緻化を行う。

4. 本研究のもたらすもの

本研究から予想される成果と意義は以下のように要約できる。まず本研究は表層的な音楽情報処理ではなく、人間への内部観測を伴った新たな研究体系を構築する試みである。確率的文法規則と意味構成プロセスを確立し、世界

*6 J. S. バッハの BWV582, シヤコンヌ BWV1004, ブラームスの作品 98 最終楽章, ラヴェルのトリオ第三楽章など。

標準の意味表現を提示する。

次にわれわれは木構造解析システム・意味構成システムを大量の解析例とともに提示し、上記の応用システムをウェブ上にオープンソースとして公開する。これにより、われわれが既に豊富に手にする木構造に対するアルゴリズムが音楽情報処理研究の対象となる。すなわち、旧来の研究の枠組みを次世代に向けて大きく拡張するものである。

さらに木構造によるゲシュタルト意味を保存することから、音楽の専門知識がない人にも原曲のフレーバーを正しく残す作曲・編曲システムを提供する。

またこれまで情動として曖昧に論じられてきた音楽の認知的意味に対し、本研究では聴者が受け取った木のトポロジーの再現が音楽意味の伝達であると考え（図8）。これにより音楽の情動を論理的に議論する俎上に載せる。これまで音楽情報処理の研究は、音楽を物理的・客観的对象として捉えた研究が主流であり、音楽の影響を心に照射する試みは多くは見られなかった。本研究においては楽曲に形式的な意味構造を付与し、その応用システムを構築するとともに音楽を理解するプロセスをモデル化することに寄与する。

なお以上の研究は、Alan Marsden (Lancaster University, UK) および David Meredith (Aalborg University, Denmark) 両名の海外協力者を得て行う。

謝辞 本研究は日本学術振興会 (JSPS) 科学研究費 16H07144 の支援を受けて遂行するものである。

参考文献

- [1] J. Barwise and J. Perry: *Situations and Attitudes*, The MIT Press, 1983.
- [2] L. Bernstein: *Unanswered Question*, Harvard Lecture Series, 1973.
- [3] D. R. Dowty: *Word Meaning and Montague Grammar*, Reidel, 1979.
- [4] M. Granroth-Wilding and M. Steedman: *A Robust Parser-Interpreter for Jazz Chord Sequences*, *Journal of New Music Research*, 43, 355-374, 2014.
- [5] M. Hamanaka, K. Hirata, and S. Tojo: *Implementing "A Generating Theory of Tonal Music"*, *Journal of New Music Research*, 35(4), 249-277, 2006.
- [6] M. Hamanaka, K. Hirata, and S. Tojo: *deepGTTM-I: Local Boundaries Analyzer based on Deep Learning Technique*, 12th International Symposium on Computer Music Multidisciplinary Research (CMMR), 2016.
- [7] K. Hirata, S. Tojo, and M. Hamanaka: *Algebraic Mozart by Tree Synthesis*, ICMC2014.
- [8] T. Kitahara and Y. Tsuchiya: *Short-term and Long-term Evaluations of Melody Editing Method Based on Melodic Outline*, *Int. Computer Music Conf. (ICMC/SMC)*, pp. 12041211, 2014.
- [9] F. Lehrdahl and R. Jackendoff: *Generative Theory of Tonal Music*, MIT Press 1983.
- [10] M. Matsubara, K. Hirata, and S. Tojo: *Distance in Pitch Sensitive Time-span Tree*, *International Computer Music Conference (ICMC/SMC) 2014*
- [11] E. Nakamura, M. Hamanaka, K. Hirata, and K. Yoshii.

- Tree-structured Probabilistic Model of Monophonic Written Music Based on the Generative Theory of Tonal Music, *Proceedings of the 41st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016)*, 2016.
- [12] E. Narmour: *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. Chicago University of Chicago Press 1990.
 - [13] J.-J. Nattiez. *La Musique, La Recherche, et La Vie*, Leméac Éditeur Inc., Montréal, 1999 (ジャン=ジャック・ナティエ, 添田里子 訳, 音楽・研究・人生, 春秋社 2005)
 - [14] H. Schenker, *Der Freie Satz. Neue musikalische Theorien und Phantasien*, Margada, Liège, Belgium, 1935.
 - [15] S. Tojo and K. Hirata: *Structural Similarity Based on Time-span Tree*, *CMMR* 2012.
 - [16] S. Tojo, K. Hirata, M. Hamanaka, *Computational Reconstruction of Cognitive Music Theory*, *Journal of New Generation Computing*, Vol. 31, No. 2, pp. 89113, 2013.
 - [17] N. L. Wallin and B. Merker ed.: *The Origins of Music*, The MIT Press, 2001.
 - [18] T. Winograd, *Linguistics and the Computer Analysis of Tonal Harmony*, *Journal of Music Theory* 12:1, 1968.
 - [19] 大村英史, 柴山拓郎, 寺澤洋子, 星柴玲子, 川上愛, 吹野美和, 岡ノ谷一夫, 古川聖, “音楽情動研究の動向,” *日本音楽学会誌*, Vol. 69, No. 9, pp. 467478, 2013.
 - [20] 大村英史, 二藤宏美, 岡ノ谷一夫, 古川聖, “音楽構造の破壊による音楽情動への影響の調査,” *日本認知科学会論文誌「認知科学」*, Vol. 21, No. 1, pp. 152159, 2013.
 - [21] 東条敏, “われらの脳の言語認識システムが生み出す音楽,” 「進化言語学の構築 (藤田耕司, 岡ノ谷一夫編)」, ひつじ書房, pp. 197217, 2011.