

音楽認知情報処理に関する3つのアプローチ

片寄晴弘^{†,††} 橋田光代^{††} 豊田健一^{†,††}
野池賢二^{††} 奥平啓太^{†,††}

本稿では、筆者らが行ってきた研究のうち、音楽認知情報処理の文脈に相当するものとして、1) 脳機能計測を用いた音楽における没入感の計測、2) 音楽グループ聴取に関する検討、3) 実時間音楽予測アーキテクチャ、の研究事例を紹介する。

Three Approaches on Cognitive Music Informatics

HARUHIRO KATAYOSE, MITSUYO HASHIDA, KEN'ICHI TOYODA,
KENZI NOIKE and KEITA OKUDAITA

This paper introduces three examples of research activities, in which the authors have been engaged, in terms of cognitive music informatics; 1) Measurement of immersive experience in music using NIRS, 2) Experiments on perception of musical groups, and 3) An computational model for real time music prediction.

1. はじめに

音楽は、他の科学技術分野と比べても早くから計算機の利用・応用が試みられた領域であり¹⁾、1950年後半には、計算機による自動作曲「イリアック組曲」²⁾が発表されている。音楽とコンピュータの接点における研究・開発は、最近では、エンタテインメントやホビー領域として定着したコンピュータミュージック領域におけるニーズに応える技術としても発展し続けている。特に、近年のマシンパワーの増大に支えられる形で、セッションシステムや演奏ゲームなどインタラクションに関する技術開発が盛んに行われるようになってきている³⁾。

このように技術が進展していく中、システム作りには軸足を置く研究者にとって、人間がどのように音楽を聞いているのか、音楽行為が人間にどのような影響を与えるのか、といった認知的な視点は今まで以上に、大きな関心の対象となりつつある。筆者らも、ここ数年、情緒あふれる演奏の生成機構（音楽解釈モデル）、演奏表現システム、セッションシステムや和声付けなど、システム作りを中心とした研究を進めてきたが、これらに関連して取り組んだモデル化や評価における

検討の中には、音楽の認知・知覚領域の研究として見つめ直しても面白いだろうと思われるものがいくつか存在する。本稿では、それらのうち、1) 脳機能計測を用いた音楽における没入感の計測、2) 音楽グループ聴取に関する検討、3) 実時間音楽予測アーキテクチャ、の3つの研究事例を紹介する。

2. 脳機能計測を用いた音楽における没入感の計測

音楽に限らず、自らの働きかけと感覚統合を伴って芸術に関わる際には、内発的な没入としてとらえられる精神状態になることが知られている。この章では、音楽演奏システム iFP を利用した、音楽の没入感に関する脳機能計測実験例を紹介する。

2.1 iFP

iFP は、名演奏のデータをテンプレートとして利用し、指揮動作、あるいは、拍打によって、テンポ、音量を制御する演奏インタフェースである。図1に示すように、演奏表情を、テンポ、拍音量、拍内の微細な演奏表現の3つの次元に分け、テンプレートに記載された演奏モデルとユーザの意図（テンポ、音量の制御）とを合成することにより、演奏を生成する。微細な演奏技術はシステムが担うため、プレイヤーは、フレーズ表現に集中することができる。また、名ピアニスト

[†] 関西学院大学

Kwansei Gakuin University

^{††} 科学技術振興機構さきがけ研究 21

PRESTO, JST

いわゆるトランス状態ではない

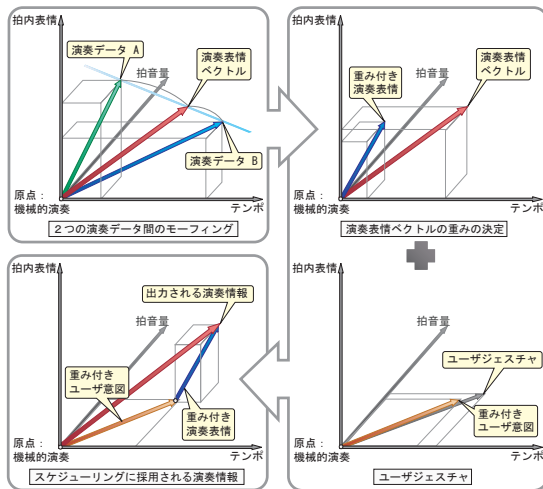


図1 iFPにおける演奏データ生成の概念図。ユーザ（プレイヤー）が直接的に操作する elaboration は、テンポ、音量である。その他に、テンポ、拍音量、拍内の微細な演奏表現（偏移の分散）表現のバランス（重み）を操作する。

の表現を自身の意図を持ってなぞるといったような使い方も可能である。このシステムでは、2つの演奏テンプレートを利用したモーフィングインタフェースも用意しており、例えば、平均的な演奏と重い演奏との外挿により、軽やかな演奏を生成するというような使い方もできる。

2.2 脳機能計測に基づく没入感に関する評価

ナイーブな感覚を客観的に検証するためには、生理的計測が有効な手段となる。中でも、近年、開発された NIRS (near-infrared spectroscopy) は、空間分解能はさほど高くないものの、作業しながらの脳機能計測ができるという特徴を有しており、ここで取り上げるような評価対象に対しては、その効用が期待される。

音楽的没入の要因としては、いくつかの事項が指摘されているが、ここでは「身体性」「モダリティ統合」「技能の拡張」の3つの事項に着目し、内観評価とあわせて実施した NIRS による計測実験結果⁴⁾について紹介する。

今回の一連の実験においては、実験対象曲として、「星に願いを」のピアノアレンジ版を用いることにした。NIRS による計測部位は、音楽聴取によるリラクゼーションによって活動レベルが低くなることが知られている前頭前野とした（図2）。

まず「身体性」「モダリティ統合」の要因を探るために実施した実験の結果を図3に示す。被験者は、iFP

被験者の大脳の脳活動を、oxy-hemoglobin (HbO) と deoxy-hemoglobin (Hb) の変化によってとらえる実時間センシングシステムである。



図2 NIRS を用いた測定

の操作に慣れ、iFP の演奏に没頭できると申告している、大学で専門教育を受けた音楽経験者（以下、A）である。左より、「聞き流す」場合、「分析的に聴く」場合、「iFP 指揮インタフェース利用+演奏テンプレート利用」の場合、音楽を流さず手振りをした場合、の計測データである。図中、赤色の線が脳活動の活性化度合いを表している。このデータによると、iFP による演奏時に最も活性化状態のレベルが下がり、続いて、「集中して聴く（予測しながら聞く）」際に、活性レベルが低い。これに対し、手振りのみの場合、活性レベルが幾分上がっている。被験者の没入、気持ち良さの内観報告と、fz における活性レベルの低下には相関関係が見られる。

図4は、被験者Aの演奏テンプレートの効果およびインタフェースの差（身体性）に関する実験結果である。左より「表情テンプレート無し+キーボード」「表情テンプレート有り+キーボード」「表情テンプレート有り+指揮インタフェース」「センサエラー時」における fz の計測データである。図中、右から2番目、すなわち、表情テンプレート利用・指揮インタフェースの利用時で最も当該領域の活性レベルが低下している。センサエラーのデータは、偶然、取得されたものである。これについて、被験者から「タイミングがうまく合わず、集中できなかった」と報告があった。

次に「指揮のインタフェースの操作が難しい」と申告した被験者B（エレクトーン歴12年の音楽経験者）の計測結果を示す。図5において、左より、「聞き流す」場合、「集中して聴く（予測しながら聞く）」場合、「iFP（キーボードで演奏した場合）」、「iFP（指揮インタフェースを利用した場合）」である。この被験者は、被験者Aと異なり、「iFP（指揮インタフェースを利用した場合）」の場合で、脳活動の賦活化が見られる。この状態は、計算を実施している際にも見られる状況である。いわゆる思考している状況がとらえられたものと思われる。

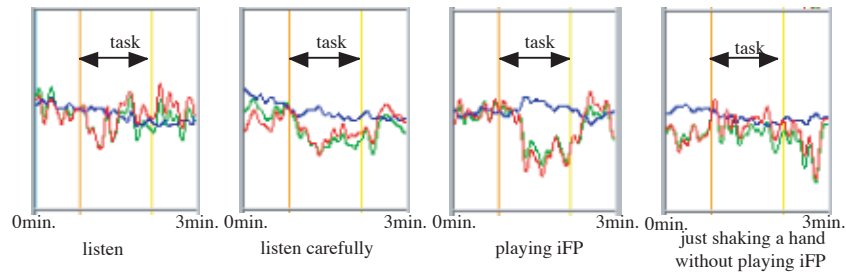


図 3 身体性とモダリティ要因に関する実験。対象曲は「星に願いを」、被験者は iFP の操作に慣れた音楽経験者であり、前頭前野正中部 (fz) におけるデータを示している。図中、赤色の線が Oxy-Hb の経時的な変化であり、その部位の脳活動の活性化度合いをとらえている。

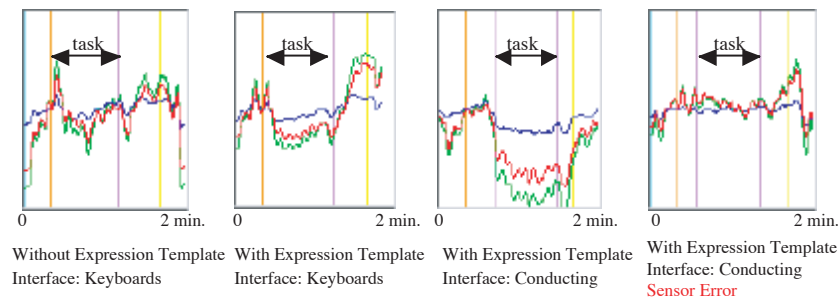


図 4 インタフェースの効果に関する実験。被験者は A。演奏テンプレートの利用 + 指揮インタフェース利用時に最も fz での脳活動が低下している。

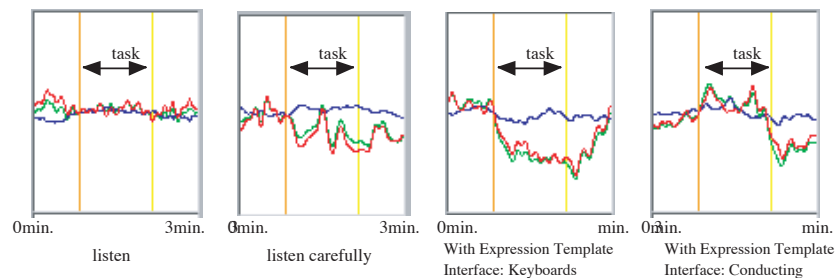


図 5 指揮インタフェースに慣れていない被験者 (B) の計測結果。指揮インタフェース利用時に fz における脳活動が活性化している。いわゆる考えている状態を示していると思われる。

2.3 NIRS を用いた実験の検討

NIRS を用いた実験においては、すべての試行において、内観（没入感・音楽的な気持ち良さ）と前頭前野正中部における脳活動の活性化レベル低下（とその程度）との相関関係が見い出された。身体表現については、被験者 A と B で、好むインタフェースが異なるという結果になったが、十分に慣れた場合は、より動作がはっきりとしている指揮インタフェースの方が支持されるであろうと予想している。

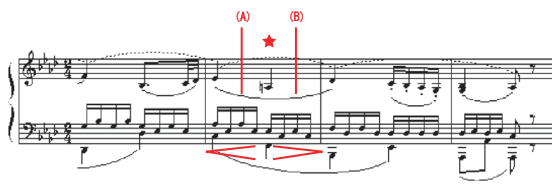
今回の実験では、「iFP の利用」>「集中して聴く（予測しながら聞く）」>「聞き流す」の順で、没入度が高いという、きれいなデータが得られたが、よく慣れ親しんだ、かつ、ムード音楽系の「星に願いを」を

用いたことによるところも大きいと考えられる。

3. 音楽グループ認知に関する検討

筆者らのもう 1 つの主要研究テーマに「情緒あふれる演奏の自動生成」がある。同時に、システム評価の一貫として、世界の「情緒あふれる演奏の自動生成」研究を一堂に集めての聞き比べコンテスト (Rencon) を企画・実施してきた⁵⁾。その過程で、筆者らが強く興味を持った事項の一つに、「(同じ演奏に対しての)音楽グループのとりえ方が聴取者によって大きく異なる」ということがある。

音楽グループの認知要因として、テンポや各音の音量などの演奏パラメータに着目する理論（例えば、文



	和音伴奏	通常演奏	内声を解説後 通常演奏
(A)に境界あり	14人	9人	1人
(A)に境界なし	25人	5人 → 30人	8人 → 38人

図 6 6 小節目のグルーピング結果

献 6) 等) はいくつか存在するが、その多くは、時間的な間隙がグループを形成する、大きい音が(後続の小さい音を併合し)グループの先頭となるなど、ゲシュタルトに関連する局所的な evidence (根拠) を定性的に指摘するにとどまっている。これに対し、モーツァルトのピアノソナタ K.331 (ヘンレ版、ペーターズ版) を対象に、実証実験によって、evidence のプライオリティを指摘した興味深い報告もある⁷⁾ が、個人差が無視できない対象であることもあって、定量的、かつ、システムティックな研究は進んでいるとはいえない。

筆者らは、「局所的な evidence の積み上げと(伝統的西洋音楽特有の)典型的なフレーズ表現 スキーマ(トップダウン的)が、拮抗する形で、最終的なグループを決定する」という形の、計算モデルの実装に興味を持っている。この章では、この計算モデルの実装に先立って実施した音楽グループ知覚に関する調査研究について紹介する。

3.1 「悲愴」でのグループ知覚実験

実験にあたっては、内声の存在が楽曲のフレーズングに大きく影響すると考えられるベートーベンのピアノソナタ「悲愴」第 2 楽章冒頭 8 小節を用いることにした。演奏については、保科が文献 8) の中で、詳細な解説を行い、MIDI の情報として示したもの(通常演奏)と、対照として通常演奏での内声を和音に置き換えた演奏(和音伴奏)を用意した。その上で、1) 和音伴奏、2) 通常演奏、3) 内声の意味を解説してからの通常演奏、を順次聴取させ、グループ境界がどのように変化するかを追跡していった。

大学生 48 名(理工学部学生、音楽大学学生)に対して実施した実験結果の内、ここでは、第 6 小節目のグルーピングを取り上げる(図 6)。この部分では、内声はフレーズ表現をサポートしていることに注意され

たい。また、メロディの A₄ 音の音量は大きく、四分音符の並びであることから、(ボトムアップ的な)ゲシュタルト基準に従うなら、(A) の部分で切れやすいということになる。

1) 和音伴奏 から、2) 通常演奏 への推移においては、被験者 5 名が、(A) での境界を感じなくなったと報告した。この被験者群は、内声に対するアテンションがあり、かつ、典型的なフレーズ表現に対するスキーマが強いものと思われる。

2) から 3) への推移、すなわち、内声の機能の説明の前後では、8 名の被験者がグルーピングに関する判断を変えた。この被験者群については、本質的には典型的なフレーズ表現に対するスキーマを有しているが、ゲシュタルトに対するスキーマとメロディに対するアテンションが強い。解説によって、内声に対するアテンションが強化されたと推察される。

なお、この実験では、音大生とそうでない学生間での有意な差は見られなかった。

3.2 グルーピング傾向とその形成に関する検討

聞き方のクラスが分離できたとするならば、そのことがどのような付帯的な事項と関連しているのかが知りたくなる。

我々は、前節での実験によって分離された 3 つの被験者グループ:(あ) 典型的なフレーズ表現に対するアテンションが高いと思われるグループ(5 人)(い) ゲシュタルトに対するアテンションが高いと思われるグループ(9 人)(う) 残りのグループ(25 人)の間で、「音楽経験」「好きな音楽ジャンル」「どのような音楽の聞き方を嗜好するか」に、傾向の差があるかについて調べてみた。この結果「音楽経験」「好きな音楽ジャンル」の 2 項目については、3 つの被験者間で、有意と判断される特徴は見つからなかった。

一方「どのような音楽の聞き方を嗜好するか」については、興味深い結果が得られた。実験としては、13 個の聞き方のタイプ(例えば、一人で聞く、集中して聞く、歌詞を聴く)をあげ、それぞれの被験者に対し、嗜好順に順位をつけてもらった。

図 7 は、それぞれのグループ間の平均順位を計算し、支持が高い項目が外側に来よう尺度の変換を施し、グラフ化を行ったものである。順序の平均を用いて尺度構成を行っているため、数値自体に信頼性はないが、フレーズ表現に対するスキーマが強いと思われるグループが、一人で静かに聞くことを好んでいること、音楽を覚えようとして聞こうとする態度が強いことなどが読み取れる。

音量やテンポを山を描くように変化させる演奏表現を指している。カザルスはレインボウタイプと呼んでいる。

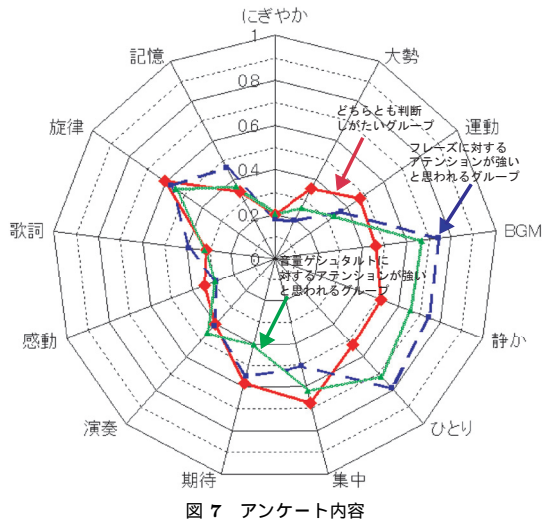


図 7 アンケート内容

4. 実時間音楽予測アーキテクチャ

4.1 基本的な考え方

今までに、音楽能力を発揮する数多くの（計算機）システムが作られてきた¹⁾が、そのほとんどは、作・編曲、表情付け、自動伴奏、セッションなど、個別の音楽タスクの実践を対象として作られたものである。しかしながら、各タスクの機能面に着目すると、予め与えられた、あるいは、何らかの手段で内発的に生成された“種”に対して、長期記憶を参照しながら、データを付加する、あるいは、(種)を詳細化するという共通の性質がある。詳細化過程は“補完”と読み替えることも可能である。補完過程自体は、未来の予測スケジューリングが必要なインタラクティブなタスクであっても、リアルタイム性が要求されないタスクであっても、基本的には、同じアルゴリズムでの構成が可能である。インタラクティブ型であって、かつ、長期記憶による自律的なデータ生成とリアルタイムに入力される信号への反応のバランスの重み制御を行う機構を持つシステムは、作・編曲システムや音楽解釈システムなどの静的なタスクを実施するシステムを包含すると考えて良いだろう⁹⁾。このような観点から、筆者らは、過去の音楽聴取体験に基づく予測補完処理を主眼とする音楽生成モデル（図 8）を提唱してきた。

4.2 システムデザイン

「過去の音楽聴取体験に基づいて予測補完を実施する」アルゴリズムの基本的な方略は、ユーザの入力、あるいは、システムが生成（出力）する音楽情報と、

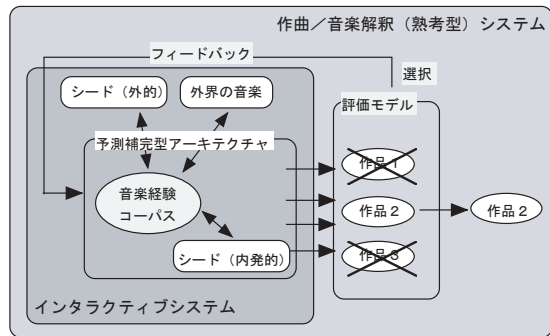


図 8 予測補完型音楽生成のコンセプト図

蓄積した音楽データにおいて、適切な reduction レベルでマッチングを行い、それに付随する補完的な音楽情報を出力するというものである。

4.2.1 検索のストラテジ

一般に、検索条件が具体的、かつ、詳細化しすぎていると、適合する事例が見つからないということが起こり得る。このような状況に対処するため、1)より長い（古い）データ列から、短いものに向かって探索を行う。2)より具体度の高い reduction レベル（あるいは、reduction の組み合わせ）から、抽象度の高い reduction レベルに向かって、段階的な探索を行うことによって、できるだけ、与えられたデータに近い事例を探索・利用するという戦略をとるものとする。

例えば、3小節の入力データヒストリに対して、MIDI名（音色的な高さ）とテンション付コードで探索を行い、合致するものが無ければ、MIDI名の代わりに調性的音階で探索を行う。あるいは、コードについて、テンションに対する条件を外す、さらに、それでも合致しなければ、2小節のデータヒストリで計算を行うというような処理を行う。

各 reduction レベルでの探索においては、事例間の類似度（距離）を定義する（できる）場合と、しない（できない）場合がある。前者の場合は、唯一の事例を選択出来る可能性が高いが、後者の場合は、数例の事例がマッチする可能性がある。これに対する対処については、以下の節で取り扱う。

4.2.2 データ生成のストラテジ

検索対象事例として与えられるデータが少数である場合や、事例間の類似度が計算され、合致事例が一意に絞ることができる場合は、その事例を用い、指定した補完対象となる音楽情報を出力することで、目的は

音楽の抽象化構造：拍節構造、和声、メロディ輪郭、タイムスパンリダクション、プロロンゲーションリダクション等。reductionには、階層構造をなすものがある。

達成される。

洗練された reduction クラスを用意できない、あるいは、類似度がうまく定義できない状態で、大量のデータベースからの検索を行った際には、探索条件を1段階緩めただけで、合致事例が劇的に増えてしまうことがある。それぞれの合致事例が示唆する生成（補完）データが反駁する可能性が高いが、この際、平均値をそのまま用いることは適当な方策とはいえない。この段階で、reduction 抽象化を停止し、検索対象を reduction 記述に投影したデータ列の統計的性質を用いて、データの生成（補完）を行うものとする。

具体的には、その際の検索条件である reduction 論理積の部分空間毎に検索対象を投影し、マッチする事例から予想される生成（補完）対象の各値毎に生起確率を計算しておく。その上で、ベイズの定理に基づき、生成（補完）対象の値に対する全体的な尤度を計算する。そして、前もってユーザが指定した1)尤度の高いデータを決定的に選択する、2)統計的性質を満たすよう、乱数を用いて選択する、の何れかの指示に従って、補完データを決定する。

4.3 デモシステム

以上、過去の音楽聴取体験に基づく予測補完処理を主眼とするシステムの概略を紹介した。研究会では、以上のような考え方に基づき、一種のペイジアンネットワークによって構成した「ながしミュージシャン」様のデモシステムを紹介する。このシステムでは、データベースとしては数曲分しか登録されておらず、洗練された音楽的応答が出来るまでには至っていないが、同一機構の上で、reduction レベルの制御による聴取機構、予測補完に基づく実時間音声付け、自動後続メロディ生成等の機能が実現されている。

5. まとめにかえて

本稿では、筆者らが実施してきた音楽認知・知覚に関連する研究例として、1)脳機能計測を用いた音楽における没入感の計測、2)音楽グループ聴取に関する検討3)実時間音楽予測アーキテクチャ、について紹介した。

以下、まとめに変えて、それぞれの項目毎の達成事項と今後の課題について整理する。

第2章の没入感の計測については、研究の方向性を提示するという点において、一通りの目的を果たすことができたと考えている。今回の実験では、前頭前野正中部における脳活動の不活性化を1つの指標としてとらえたが、分析的な音楽の聞き方においては、むしろ、前頭前野正中部における脳活動は賦活化すること

が知られている。このことを含め、音楽への没入と脳活動の関係を、より精緻に調べることを目的として、慣れの効果、難解とされる楽曲を対象とした場合、被験者の音楽専門性をコントロールとした場合の実験を執り行いたい。

第3章の音楽グループ聴取に関する検討については、音楽グループ認知の多様性とスキーマ形成の関係を調べるための1つの方法論を提示できたと考えている。今回の実験では、被験者の中に音大生は入っているが、人数が少なく、また、専門的音楽レベルはさほど高くなかった。専門的音楽が高い被験者に対して同様の実験を行うことで、より、明確な知見が獲得できるものと期待している。今後は、音楽の専門分野や嗜好を細分化して、さらなるデータの蓄積を進めていきたい。同時に、表情付けシステムの能力向上に向け、聴取モデルの組み込みを実現していきたい。

第4章では、さまざまな音楽タスクを、統一的な処理モデルで実施することを可能とする計算モデルの提案を行った。研究は始まったばかりであり課題も山積しているが、聴取系を持つ音楽情報処理システムはこれからの音楽情報科学の主要研究対象の一つに成長すると予想している。今後は、検索に関わる reduction 抽象化レベルの自動制御モデル、annotation の効果的な利用に関する基礎的な研究を進めると同時に、具体的な実用化に向けての可能性も探していきたい。

参考文献

- 1) Roads, C.: コンピュータ音楽 (青柳, 小坂, 平田, 堀内 訳・監修), 東京電機大学出版局, pp. 678-752 (2001).
- 2) Hiller, L. and Isaacson, L.: *Experimental Music*, McGraw-Hill (1959).
- 3) 片寄晴弘: 音楽生成と AI, 人工知能学会誌, Vol. 19, No. 1, pp. 21-28 (2004).
- 4) 片寄晴弘: 音楽とエンタテインメント, 日本バーチャルリアリティ学会誌, pp. 20-24 (2004).
- 5) 平賀瑠美, 平田圭二, 片寄晴弘: 蓮根, 目指せ世界一のピアニスト, 情報処理, Vol. 43, No. 2, pp. 136-141 (2002).
- 6) Lerdahl, F. and Jackendoff, R.: *A Generative Theory of Tonal Music*, MIT Press (1983).
- 7) 竹内好宏: 認知的視点による演奏解釈の研究 (1994).
- 8) 保科洋: 生きた音楽表現へのアプローチ - エネルギー思考に基づく演奏解釈法, 音楽之友社 (1998).
- 9) 片寄晴弘, 豊田健一: 聴取経験に基づく予測補完型音楽生成アーキテクチャに関する考察, 情報処理学会研究報告 2004-MUS-56-11, pp. 71-78 (2004).