

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 22 日現在

機関番号：14301

研究種目：挑戦的萌芽研究

研究期間：2013～2015

課題番号：25540167

研究課題名(和文) キャラクタを伴った仮想音楽教師による楽器演奏教授システムの構築

研究課題名(英文) Development of Musical Instrument Instruction System by Virtual Music Teacher with Character

研究代表者

浜中 雅俊 (Hamanaka, Masatoshi)

京都大学・医学(系)研究科(研究院)・研究員

研究者番号：30451686

交付決定額(研究期間全体)：(直接経費) 2,700,000円

研究成果の概要(和文)：本申請では、初心者には楽器演奏を適切に指導する仮想のキャラクタを持った音楽教師の実現を目指す。仮想音楽教師は、初心者の演奏音の分析やカメラを用いた両手の手指形状の推定により、初心者の運指やタッチ、リズムなどの問題を見つけ出す。そして、コンピュータグラフィックスによる演奏例示や、問題改善のためのレッスン課題の提示など、そのレベルにあったアドバイスをを行い、初心者を効率よく上達させる。

研究成果の概要(英文)：We attempted to develop a system that enables to instruct playing musical instrument. Virtual musical instructor finds problems of playing such as rhythm, fingering or touch of key by analyzing the sound and estimating the shape of hand and finger. Then the virtual musical instructor improves the musical novice's performance, by showing the performance with computer graphics or recommending a lesson plan for improvement.

研究分野：音楽情報処理

キーワード：楽器演奏支援 手指形状推定 仮想演奏者

1. 研究開始当初の背景

申請者は、終了した科研費[H09]および競争的資金[H04]において、音楽初心者による作曲・編曲・演奏を支援するシステムを構築してきた。具体的には、1983年に提案された音楽理論 Generative Theory of Tonal Music(GTTM)を世界で初めて計算機上に実装し、数学的な演算で作曲・編曲の操作を表現することを可能としてきた。

図1は、演奏者が途中まで弾いているメロディの後続音を音楽理論 GTTM に基づき予測し天板上に表示するシステムで、初心者の即興演奏を支援することを目指していた。演奏者が即興演奏の途中でどのような音を出せば良いか困った場合には、システムが天板上に表示する音を弾くことで音楽的に違和感のない演奏をすることができる。

実験の結果、メロディ予測システムはピアノ歴3年以上の中級者に対しては適切に即興演奏の支援ができたものの、完全な初心者はそもそも運指がうまくいかないなど初歩的なレベルで躓いていることがわかった。これまで提案されてきた多くの楽器練習支援システム[ONS05]も、中級者以上を対象としており、たとえばピアノの場合では、タッチの違いによる音量のコントロールや指の交差がうまくできない初心者に対して適切に支援するようなシステムは存在しなかった。

[H09]浜中雅俊: 音楽理論 GTTM に基づく作曲・編曲システム, 科学研究費補助金 若手スタートアップ 課題番号 19800003 (2007~2008年度)

[H04]浜中雅俊: ドレミっち:成長する仮想演奏者の実現, 科学技術振興機構 戦略的創造研究支援事業 さきがけ (2004-2007年度)

[ONS05] 大島千佳, 西本一志, 鈴木雅実: 家庭における子どもの練習意欲を高めるピアノ連弾支援システムの提案, 情報処理学会論文誌, Vol.46, No.1, pp.157-171, 2005.



図1:メロディ予測システム

2. 研究の目的

初心者が楽器演奏を上達させるためには、教師によるレッスンは欠かせない。しかし、上達したいと思っても、時間的、空間的、金銭的理由によりレッスンを受けることが困難な場合がある。また、レッスンを受けていても同じ練習の繰り返してモチベーションが下がってしまうこともある。そこで本申請では、初心者に楽器演奏を適切に指導する仮想のキャラクターを持った音楽教師の実現を目指す。仮想音楽教師は、初心者の演奏音の分析やカメラを用いた両手の手指形状の推定により、初心者の運指やタッチ、リズムなどの問題を見つけ出す。そして、コンピュータグラフィックスによる演奏例示や、問題改善のためのレッスン課題の提示など、そのレベルにあったアドバイスをを行い、初心者を効率よく上達させる。

3. 研究の方法

3.1 GTTM 分析器の高度化

図1のメロディ予測システムは、音楽理論 GTTM に基づき予測結果を表示しているが、我々が構築してきた分析器が単旋律(モノフォニー)にしか対応していなかったため、分析精度が低く、また、音楽家によるパラメータの調整が必須で自動分析ができないという問題があった。

そこで我々は、GTTM分析に確率を導入した σ GTTMを構築してきた。本研究では確率に基づく GTTM のタイムスパン木の分析器 σ GTTMIIIを構築した。

タイムスパン木は、木の先端の装飾的な音符を簡約することで、より抽象化したメロディを抽出することができる。図2aのメロディAの上にある木構造は、メロディAを分析して求めたタイムスパン木である。レベルBより下で分岐している枝に接続している音符を簡約するとメロディBが抽出される。さらに、レベルCより下で分岐している枝に接続している音符を簡約するとメロディCが抽出される。この簡約プロセスを逆方向に眺めると、以下のような生成のプロセスと捉えることができる(図2b)。

1. 楽曲全体の長さのタイムスパンを持つ一つの音符を考える。
 2. タイムスパンをプライマリとセカンダリに分割する。
 3. 枝の末端になるまで2を繰り返す。
- 上記の生成プロセスを確率モデルで表現することで、最も尤らしいタイムスパン木の獲得が可能となる。

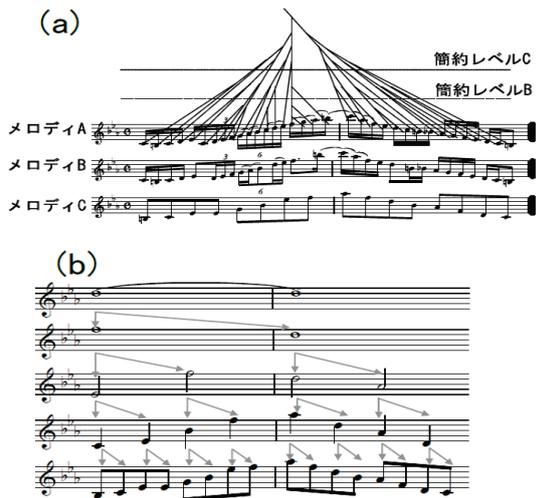


図2： メロディの簡約と音符の生成モデル

(1) 学習データ

確率モデルの学習には、我々が構築した GTTM データベースを用いる。GTTM データベースは、GTTM を良く理解している 3 人の音楽家がクラシック曲から切り出した 8 小節の長さの 300 個のメロディの楽譜データと、それを GTTM に基づき手作業で分析したグルーピング構造解析データ、拍節構造解析データ、タイムスパン解析データ、プロロンゲーション解析データ、および、和声を解析した和声解析データからなる。GTTM データベースおよび解析に使用した解析ツールは、以下のアドレスからダウンロードできる。

<http://www.gttm.jp/>

(2) 音符列の生成モデル

音楽の生成プロセスをモデル化するため確率文脈自由文法を導入する。確率文脈自由文法は、複数の生成規則とその生成確率からなり、各生成規則は、タイムスパン木におけるプライマリとセカンダリの分岐を表す。確率文脈自由文法 $G = \{T, M, S, R, P\}$ は 5 つの要素で定義される。

T: 終端記号の集合

音符が終端記号である。

M: 非終端記号の集合

タイムスパンは非終端記号である。

S: 開始記号

開始記号は、タイムスパン木の根にあたる。すなわち、休符を含まない楽曲全体の長さのタイムスパンである。

R: 生成規則の集合

生成規則には、タイムスパン分割ルールと、音符生成ルールの 2 種類がある。タイムスパン分割ルールは、タイムスパンを 2 つのタイムスパンに分割する。分割された 2 つのタイムスパンの長さの合計は、もとのタイムスパンの長さに等しい。タイムスパンの分割のされ方、すなわち、2 つのタイムスパンの音価の比は様々なものが考えられる。音符生成ルールは、タイムスパンか

ら同じ長さの音符を生成する。

P: 生成規則の確率

各生成規則は確率が付与される。たとえば、32 分音符の音価のタイムスパンが 32 分音符を生成する確率は 1 に非常に近い値となる。なぜなら、32 分音符よりも音価の小さい音符は、データベース中でほとんど存在せず、32 分音符の長さのタイムスパンがさらに分割される確率は極めて低いためである。一方、たとえば倍全音符の音価のタイムスパンが、倍全音符の音符を生成する確率は極めて低い。

(3) 楽曲の生成モデル

(2) では、タイムスパンを分割し音符列(音価列)を生成する方法について述べた。

(3) では、その他の属性について検討する。プライマリとセカンダリ: タイムスパン分割ルールでタイムスパンが 2 つに分割される際、片方がプライマリ、もう片方がセカンダリとなる。音高: プライマリのタイムスパンは、分割前のタイムスパンの音高を引き継ぎ、セカンダリの音高は新たに生成される。タイムスパンの順序: プライマリのタイムスパンが前、セカンダリのタイムスパンが後となる場合と、プライマリが後、セカンダリが前となる場合の 2 通りが考えられる。拍点の数: 3.2 で述べたように、拍節構造はタイムスパン木の構成に強い影響を及ぼす。したがって、プライマリとセカンダリのそれぞれの拍点の数はモデルに含めることが望ましい。

上記のすべての場合を区別して生成規則を作成すると、膨大な数の規則が作成され、限られた学習データから確率を求めると、多くの規則で確率が 0 になってしまう。そこで、本研究では学習データのスパース性の問題を解決するため、以下のように抽象化したルールを生成する。

3 種類の音高変化: 2 つのタイムスパンの音高変化は、上がる、下がる、等しい、の 3 種類に分類する。

7 種類の音価比: 2 つのタイムスパンの音価の比は、4 倍、3 倍、2 倍、1 倍、1/2 倍、1/3 倍、1/4 倍の 7 種類のうち最も近いものに分類する。

2 種類の順序: 2 つのタイムスパンの順序は、プライマリが前、セカンダリが後となる場合と、プライマリが後、セカンダリが前となる場合の 2 通りに分類する。

3 種類の拍点数: 2 つのタイムスパンの拍点の数は、プライマリの拍点が多い、セカンダリの拍点が多い、プライマリとセカンダリの拍点の数が等しい、の 3 種類に分類する。

6 種類のタイムスパンの長さ: 分割前のタイムスパンの無さを、16 分音符、8 分音符、4 分音符、2 分音符、全音符、倍全音符の中から一番近いものに分類する。

上記より、756 個 ($=3 \times 7 \times 2 \times 3 \times 6$) のタイムスパン分割ルールと、6 個の音符生成ルール、合計 762 個のルールを作成した。

3.2 ピアノ演奏時の手指画像を用いた接鍵検出法

ピアノは鍵盤を叩くと決まった音高で音を出すという、楽器としては単純な構造を持つ。しかし、演奏する際には滑らかかつ正確に鍵盤を弾くことや、左右の手指で別々の動作をすること、さらに抑揚などを表現するといったことが必要となるため、挫折してしまう初学者が多い。これらの理由から、ピアノ演奏初心者のための運指支援システムの研究がなされているが、抑揚表現などに代表されるアーティキュレーションの支援に関する研究は進んでいない[TTT12, OAS13]。

また、ピアノ演奏初心者のアーティキュレーション支援のためには、打鍵時、特に指と鍵が接触した瞬間の手指のフォームが重要である。そこで、手指に焦点を当てたピアノ演奏画像から接鍵タイミングの検出を行う。通常の電子ピアノや MIDI キーボードでは、鍵が押し込まれたときに打鍵信号を出力するため、接鍵タイミングの検出はできない。

一方、画像からの認識手法として、kinect の利用や、オプティカルフロー処理を用いたトラッキング手法などがあるが、これらの手法ではオクルージョンが発生してしまった場合、手指の認識精度が落ちてしまうため、接鍵タイミングにおける手指の検出は困難である。

画像特徴量である高次局所自己創刊特徴量から手指認識を行うといった研究もなされているが、ピアノ演奏時の手指は形状が似通っており、通常手指認識で用いられるように、動作が大きく変化するものではない。

しかし、オクルージョンによって指先の検出が困難な場合であっても、手指の微細な動作から高次局所自己創刊特徴量を抽出することで、接鍵タイミングが検出できる可能性がある。そこで我々は、取得画像のフレーム間の差分画像に着目し、高次局所自己相関特徴量を抽出、その成分を解析した後、サポートベクターマシン (SVM) を用いた識別器により接鍵タイミングの推定を行った。

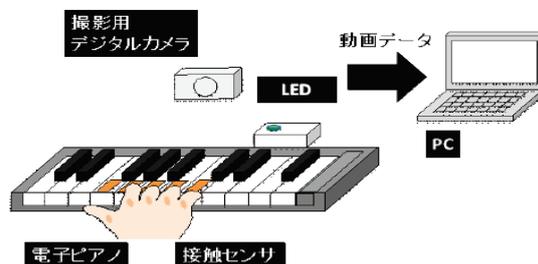
[TTT12] 竹川佳成, 寺田 努, 塚本昌彦: リズム学習を考慮したピアノ演奏学習支援システムの構築, 情報処理学会インタラクシオン 2012, pp. 73-80, 2012.

[OAS13] 岡 明也, 有賀治樹, 杉山健太郎, 橋本 学, 長田典子: マーカレス運指認識と音列照合によるピアノ演奏スキル評価システムの提案, 情報処理学会研究報告, 2013-MUS-100, No. 10, 2013.

(1) 接鍵検出システム構成

提案する手指画像からの接鍵検出システムの概要図を図 2 に示す。鍵盤には、指の接触を感知する接触センサを取り付けてあり、接触の際に LED が点灯する。カメラは LED、ピアノ演奏時の手指、および周辺環境を撮影

し、演奏者はビデオ録画が始まると同時に演奏を開始する。この演奏動画は PC で処理する。ここで LED の点滅は評価における正解データの正誤判定に用いるため、特徴量を抽出する際の動画は演奏時に打鍵される鍵盤、奏者の手指全体が映っている箇所をトリミン



グして使用する。

図 2: 接鍵検出システム概要

(2) 画像処理法とシステム

ピアノ演奏時の画像情報から接鍵判定を行うためのシステムの流れを検討する。

画像処理の流れを図 3 に示す。動画に対してフレーム間差分処理を行い、二値化を行う。フレーム間差分処理を行うことによってフレーム間で動きのある部分のみが検出されるため、指のみの移動などの微細な移動に対しても変化が検出され、特徴量の差異が観測可能となる。その後、二値化処理および画像の縮小処理を行うことにより、高次局所自己相関関数の特徴量を検出する。さらに、接鍵タイミングとその前後 1 フレームにおける高次局所自己相関特徴量を入力として SVM を学習し、接鍵タイミングの検出を行う。

(3) 画像特徴量

画像特徴量として、手指の形状推定で高い性能を示している高次局所自己相関特徴を採用する。高次局所自己相関特徴は、 N 次自己相関関数に基づき、画像から抽出される特徴において、1) 位置に関する不変性、2) 画面に関する加法性、3) 学習による適応性を満たしている。画像の相関特徴に適用する際には、変異方向は参照点 r を局所領域に限定する必要がある。局所領域は 3×3 画素として、次元数 n は 2 とする。平行移動による等価な特徴を除くと、二値画像において局所画像は 25 個のパターンとなる。これらの特長から画像中でどの成分が多いかを判定する。

4. 研究成果

3.1 で述べた σ GTTMIII と従来の GTTM 分析器の ATTA との性能比較を行った。

生成確率は、GTTM データベースにある 300 曲のタイムスパン木の 19,296 個の分岐から数え上げて求めた。ある楽曲の最尤なタイムスパン木は、可能なすべてのタイムスパン木を生成し生成確率が最大となったものを選択することで獲得した。タイムスパン分割に用いつグルーピング構造および拍節構造は、音楽家が分析した GTTM データベース

のデータを用いた。全てのタイムスパン木の生成およびその生成確率の計算には膨大な計算時間がかかるため、計算時間の短縮のため12並列で処理を行った。100曲の最尤なタイムスパン木を求めるのに計算クラスタ(16台×Intel Xeon E5-2430@2.00 GHz 12core)を使って6週間の時間を要した。計算時間が最も短い曲では2分、長い曲では4週間であった。

σ GTTMIIIの性能をleave-one-out cross validationで評価し、ATTAの結果と比較した。GTTMデータベースにある300曲のうち、ATTAの評価でも用いた1番から100番までの曲を用いた。その結果、 σ GTTMIIIは、すべての曲でパラメータ調節前のATTAの性能を上回る性能を示していた。いくつかの曲では、ATTAのパラメータ調節後のほうが σ GTTMIIIより性能が高い場合もあったが、平均では σ GTTMIIIのほうが高い性能を示していた。

4.2 ピアノの接鍵検出法

3.2で述べた接鍵盤検出システムの実装を行った。Macbook Air (1.8GHz, Apple製), 電子ピアノ (PSR-E303, YAMAHA製), LED点灯用Arduino UNO (16MHz), デジタルカメラ (EX-ZR400, CASIO製)を使用した。動画は室内で撮影を行い、カメラをキーボードに固定し撮影を行った。撮影スピードは240fps, 高次局所自己相関特徴量を抽出する画像は80×80画素とした。

ピアノ演奏時の手指やその他の部分を撮影した動画からSVMを用いて接鍵検出を行った。評価は撮影画像7793枚の教師データを学習させ、テストデータ1096枚から正解率を求める。課題曲は、童謡のきらきら星の冒頭4小節、バイエルから38番, 57番, 64番の冒頭8小節の4曲のメロディ部分とし、各曲それぞれ3名ずつ演奏を行った。

高次局所自己相関特徴の結果を入力としてSVMによる接鍵タイミングの識別を行った結果、97.91%の正解率となり、接鍵検出法の有用性が確認できた。

5. 主な発表論文等

[雑誌論文] (計 1件)

[1] Satoshi Tojo, Keiji Hirata, Masatoshi Hamanaka: “Computational Reconstruction of Cognitive Music Theory”, New Generation Computing, Vol. 31, pp. 89-113, January 2013.

[学会発表] (計 15件)

[1] Masatoshi Hamanaka, Keiji Hirata, Satoshi Tojo: “Sigma GTTM III: Learning based Time-span Tree Generator based on PCFG”, Proceedings of The 11th International Symposium on Computer Music Multidisciplinary Research (CMMR 2015),

pp. 303-317, June 16-19, 2015. (査読あり)

[2] Masatoshi Hamanaka, Keiji Hirata, Satoshi Tojo: “Structural Similarity based on Time-span Sub-trees”, Proceedings of The 5th International Conference on Mathematics and Computation in Music (MCM2015), June 2015. (査読あり)

[3] Kouhei Kanamori, Masatoshi Hamanaka, Junichi Hoshino: “Method to detect GTTM Local Grouping Boundaries based on Clustering and Statistical Learning”, Proceedings of the 42nd International Computer Music Conference (ICMC) joint with the 13rd Sound & Music Computing conference (SMC), October 2014. (査読あり)

[4] Masatoshi Hamanaka, Keiji Hirata, Satoshi Tojo: “Musical Structural Analysis Database Based on GTTM”, Proceedings of ISMIR 2014, pp. 325-330, October 2014. (査読あり)

[5] Sakurako Yazawa, Masatoshi Hamanaka, Takehito Utsuro: “A Novel Approach to Separation of Musical Signal Sources by NMF”, The 12th IEEE International Conference on Signal Processing, pp. 610-615, October 2014. (査読あり)

[6] Sakurako Yazawa, Masatoshi Hamanaka, Takehito Utsuro: “Melody Generation System based on a Theory of Melody Sequences”, International Conference on Advanced Informatics: Concepts, Theory and Applications, pp. 347-352, August 2014. (査読あり)

[7] Sakurako Yazawa, Masatoshi Hamanaka: “Extension of Implication-Realization Model for Subjective Melodic Similarity”, BKN25 Milestones in Music Cognition, July 2014. (査読あり)

[8] Masatoshi Hamanaka, Keiji Hirata, Satoshi Tojo: “Time-Span Tree Analyzer for Polyphonic Music”, 査読付, Proceedings of The 10th International Symposium on Computer Music Multidisciplinary Research (CMMR 2013), pp. 886-893, October 2013. (査読あり)

[9] Masatoshi Hamanaka, Seunghee Lee: “Concert Scope Headphones”, International Computer Music Conference 2013 (ICMC2013), August 2013. (査読あり)

[10] Masatoshi Hamanaka, Keiji Hirata,

Satoshi Tojo: “Toward Developing a Polyphonic Music Time-Span Tree Analyzer”, 査読付, Mathematics and Computation in Music 2013(MCM2013), June 2013. (査読あり)

[11] Sakurako Yazawa, Yuhei Hasegawa, Kohei Kanamori, Masatoshi Hamanaka: “MIREX 2013 SYMBOLIC MELODIC SIMILARITY: Melodic Similarity based on Extension Implication-Realization Model”, MIREX 2013 Symbolic Melodic Similarity Results, 2013.

[12] 浜中雅俊: “ σ GTTM III の構築”, JSAI2015 人工知能学会全国大会第 29 回, 2C5-0S-21b-1, May 2015.

[13] 浜中雅俊: “音楽理論 GTTM に基づく音楽構造解析研究用データベース”, JSAI2014 人工知能学会全国大会第 28 回, 1K5-0S-07b-5, May 2014.

[12] 矢澤櫻子, 浜中雅俊: “音楽理論に基づいたメロディ生成システム”, JSAI2014 人工知能学会全国大会第 28 回, 4G1-5, May 2014.

[13] 國澤佳代, 浜中雅俊, 星野聖: “ピアノ演奏時の手指画像を用いた接鍵検出法”, 情報処理学会全国大会, 2014-IPSJ-76, 5Q-5, 2 pages, March 2014. (学生奨励賞受賞)

[14] 金森光平, 浜中雅俊: “クラスタリングと統計的学習に基づく音楽理論 σ GTTM II: 局所的グルーピング境界の検出”, 情報処理学会 音楽情報科学研究会, 2014-MUS-102, No. 2, 7 pages, February 2014.

[15] 矢澤櫻子, 浜中雅俊: “主観的類似度を反映した暗意実現モデルの拡張”, 情報処理学会 音楽情報科学研究会, 2014-MUS-102, No. 1, 5 pages, February 2014.

[図書] (計 4 件)

[1] Masatoshi Hamanaka, Keiji Hirata: “Implementing Methods for Analyzing Music Based on Lerdahl and Jackendoff’s Generative Theory of Tonal Music”, In Computational Music Analysis, David Meredith (Ed.), pp. 221-249, Springer, 2016.

[2] Keiji Hirata, Masatoshi Hamanaka: “An Algebraic Approach to Time-Span Reduction”, In Computational Music Analysis, David Meredith (Ed.), pp. 251-270, Springer, 2016.

[3] Keiji Hirata, Satoshi Tojo, Masatoshi Hamanaka: “Cognitive Similarity grounded by tree distance from the analysis of K. 265/300e”, In Sound, Music, and Motion, 10th International Symposium, CMMR 2013, Marseille, France, October 15-18, 2013. Revised Selected Papers, Lecture Notes in Computer Science (LNCS) Vol. 8905, pp. 589-605, Springer, 2014.

[4] Masatoshi Hamanaka, Keiji Hirata, Satoshi Tojo: “Computational Music Theory and its Applications to Expressive Performance and Composition”, In Guide to Computing for Expressive Music Performance, Alexis Kirke and Eduardo R. Miranda (Eds), pp. 205-234, Springer, 2013.

[産業財産権]
なし

[その他]
ホームページ:
<http://gttm.jp/>

6. 研究組織

(1) 研究代表者

浜中 雅俊 (HAMANAKA, Masatoshi)
京都大学・大学院医学研究科・研究員
研究者番号: 30451686

(2) 研究分担者

星野 聖 (HOSHINO, Kiyoshi)
筑波大学・大学院システム情報工学研究科
(系)・教授
研究者番号: 80251528

星野 准一 (HOSHINO, Junichi)
筑波大学・大学院システム情報工学研究科
(系)・准教授
研究者番号: 40313556

(3) 連携研究者

なし