様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　３０　年　　４　月　２４　日現在

機関番号：　１４３０１
研究種目：　若手研究(A)
研究期間：　2014～2017
課題番号：　２６７０００２０

研究課題名（和文）信号処理と記号処理の確率的協働による音楽知能の創発

研究課題名（英文）Emergence of Musical Intelligence Based on Probabilistic Integration of Signal and Symbolic Processing

研究代表者

吉井　和佳（YOSHII, Kazuyoshi）

京都大学・情報学研究科・講師

研究者番号：２０５１０００１

交付決定額（研究期間全体）：（直接経費）　　18,900,000円

研究成果の概要（和文）：本研究では、楽譜（離散データ）に対する言語モデルと、音響信号（連続データ）に対する音響モデルを統合した階層ベイズモデルを定式化し、音響信号だけから、その背後に存在する楽譜や文法構造を一挙に推定する技術を確立した。具体的には、ピアノ演奏の音響信号に対する自動採譜、歌声の音高軌跡に対する音符推定、MIDI演奏信号に対するリズム採譜、ポピュラー音楽の音響信号に対するドラム採譜およびギター譜生成に取り組んだ。開発した自動採譜・音源分離技術を用いて、歌唱・楽器演奏支援システムの開発を行った。さらに、コードとメロディの階層ベイズモデルに基づく、インタラクティブな作曲支援システムを開発した。

研究成果の概要（英文）：We have established a statistical framework that can jointly infer musical notes and syntactic structures from musical audio signals using a hierarchical Bayesian model based on integration of language and acoustic models. We tackled automatic music transcription for piano signals, musical note estimation for singing F0 trajectories, rhythm transcription for MIDI performances, automatic drum and guitar transcription for popular music audio signals. We applied these methods to assisting a user to sing a song and play a musical instrument. In addition, we developed an interactive music composition system based on a hierarchical Bayesian model of chords and melodies.

研究分野：　音楽情報処理

キーワード：　音楽情報処理　機械学習　信号処理　記号処理　ノンパラメトリックベイズ

１．研究開始当初の背景

近年、ポピュラー音楽のような実世界の複雑な音響信号を対象とした研究が盛んである。しかし、信号処理のみに基づくアプローチでは認識精度に限界があった。認識精度を向上させるには、音楽特有の文法規則を制約として与えて、音楽的に不適切な認識結果を防ぐことが効果的である。例えば自動採譜の場合、多数の楽器音が重なり合っていても、人間と同様に近傍の音符の配置を考慮することで採譜結果の信頼性を高めることができる。工学的には、N-gram などの確率モデルを用いて音符の配置を別データからあらかじめ学習しておき、認識したい音響信号に対して学習済みのモデルを適用することが一般的である。しかし、このような古典的な教師あり学習では、学習データを作成するには、音響信号に対する大量のアノテーションが必要になり、学習データの音楽的性質が認識対象と大きく異なると、かえって認識精度が悪化するという問題があった。

２．研究の目的

人間が音楽知能を獲得する過程を解明するため、信号処理と記号処理とを融合させ、大規模な音楽音響信号データからそこに共通して現れる離散単位および文法規則を浮かび上がらせる統計的機械学習法を確立する。我々は特別な訓練を受けていなくても、音楽には高度な同時的・経時的構造（和音を構成する音符の組み合わせパターン・和音の遷移パターン・拍節構造）が存在することを直感的に理解している。このような「しろうとの音楽理解」では、音符・和音・小節といった離散概念が自然に形成されており、工学的には音響信号の自己組織化であると考えられる。本研究では、音楽理論（文法規則）は楽譜（離散記号）から専門家が人手で導出するものであるという常識から脱し、両者を音響信号から教師なしで一挙に推論可能であるという仮説のもと、階層ノンパラメトリックベイズモデルに基づく音楽知能の創発に世界で初めて取り組む。

３．研究の方法

本研究では、大量の音楽音響信号から文法規則を獲得すると同時に、文法規則を制約として音楽要素を精度よく認識することができる統一的な計算モデルを構成することに成功した。これは、「しろうとの音楽理解」を解明するための工学的試みである。従来の音楽認識の研究は、「専門家の音楽理解」すなわち、音楽理論（文法規則）をすでに習得した人間が音楽を分析するのと同型のアプローチであった。一方、「しろうとの音楽理解」では、C maj や D min といった和音の呼称を知らなかったり、音名を知覚する絶対音感を持っていなかったりしても、和音にはいくつかの典型的な"響き"が存在すること、音高には離散的なスケールが存在することを、自らの音楽聴取経験だけから見出すことができる。このような脳の働きを「音楽知能」と名付け、問題提起と解明に取り組んだ。

人間は音楽を聴くとき、各楽器音を分離しているわけではなく、音響信号を構成する離散単位（音符・和音・小節）を知覚し、それらが形作る文法構造を直感的に把握している。逆に、文法構造の知識があるおかげで、離散単位をうまく聴き分けられる。このような鶏と卵の関係は、大量のデータから両者を同時に教師なし学習することが音楽知能の本質であることを示唆しており、統一的な階層ベイズモデルの定式化と教師なし学習という一貫したアプローチを確立した。

人間は、特別な訓練を受けなくても、このような教師なし学習の仕組みにより、音楽を理解することができるが、実際には、さまざまな音楽的な知識を習得することにより、より高度な音楽理解を行うことができる。そのため、工学的には、音響信号から言語モデルと音響モデルを同時に教師なし学習するアプローチだけではなく、言語モデルを予め学習しておく半教師あり学習のアプローチも同時に研究を行った。

４．研究成果

コードと音高の依存関係を考慮しつつ、教師なしで音楽音響信号からコードと音高を推定するための統計的手法を提案した。生成モデルとして、音高からスペクトログラムが生成される過程を表す音響モデル（NMF）と調及びコード列から音高が生成される過程を表す言語モデル（HMM）を統合した階層ベイズモデルを定式化した（図 1）。各音の存在を表す二値変数を NMF の枠組みに導入することで、ピアノロールを表す二値変数を観測とする HMM を言語モデルとして定式化することができる。混合音のスペクトログラムが与えられると、ギブスサンプリングを用いて、すべての隠れ変数（音高とコード）を同時に推定することができる。本研究により、音楽音響信号から教師なしで音楽文法の推論が可能となった。この枠組みは、一般的な音声認識システムと同様であるが、音響モデルと言語モデルを一挙に教師なしで学習するという点で異なる。本研究では、コード、音高、スペクトログラムの三階層からなる複雑な階層ベイズモデルの推論に世界で初めて成功した。

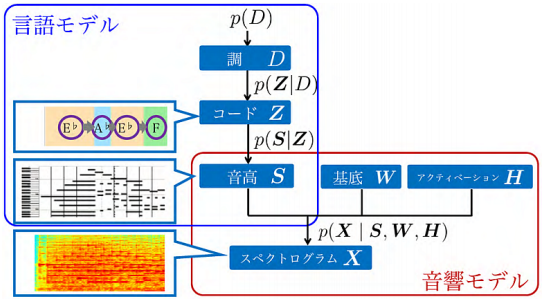

図１：コード系列・音符配置・混合音スペクトログラムの階層ベイズモデル

調とリズムを考慮することで、歌声の音高(F0)軌跡に対して精度よく音符推定を行う手法を考案した。具体的には、調（スケール）系列の生成過程、さらに調系列に基づく音符の音高・発音時刻系列の生成過程を表現する言語モデルと、音符系列に基づく歌声の F0 軌跡の生成過程を表現する歌唱モデルとを統合した階層セミマルコフモデル(HHSMM)を定式化した（図 2）。このモデルに基づいて、歌声の F0 軌跡が与えられたときに、マルコフ連鎖モンテカルロ法（MCMC）を用いて潜在変数系列を一挙に推定することに成功した。特に、歌唱モデルでは、実際の歌声における、指定された音高や発音時刻からのずれや音符間の連続的な F0 遷移を考慮しており、これら歌手特有の歌唱スタイルを同時に推定することに取り組んだ。

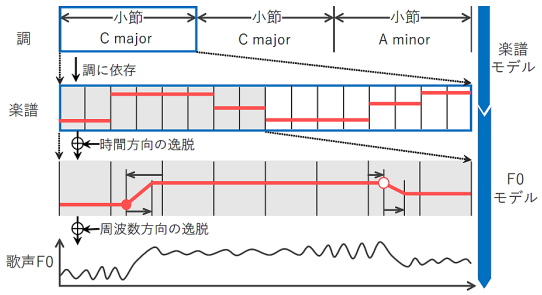

図 2：調系列・音符系列・歌声 F0 軌跡の階層ベイズモデル

さらに、自動採譜と自動編曲の中間的な課題として、さまざまな楽器を用いて演奏されるポピュラー音楽の音響信号が与えられた際に、ボーカル、数種類のギター、ドラムからなる典型的なバンドで演奏可能な楽譜を生成する課題に取り組んだ。歌声の採譜に関しては、上記で説明した通りである。ドラムパートの自動採譜を行うには、小節ごとのドラムパターン（バスドラム・スネアドラム・ハイハットの発音時刻の配置）を生成可能な言語モデル（変分オートエンコーダ）と、ドラムパターンからスペクトログラムを生成する音響モデル（非負値行列分解）とを統合した階層ベイズモデルを定式化することで、音響信号からドラム譜を推定する技術を開発した（図 3）。

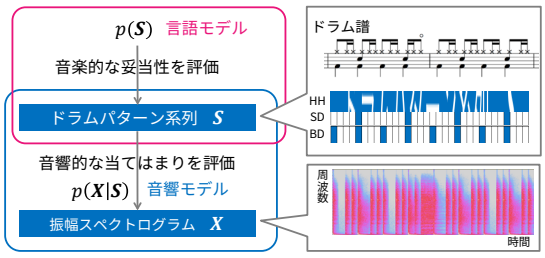

図 3：ドラムパターン・混合音スペクトログラムの階層ベイズモデル

一方、伴奏パートに関しては、各ギターパートを構成する音符や和音の遷移を表現する言語モデルと、音符や和音からスペクトログラムを生成する音響モデルを統合した階乗隠れセミマルコフモデルを定式化することで、任意の楽器で演奏されうる音響信号を、リードギター・ベースギター・リズムギターの 3 種類のギターで演奏可能な楽譜に変換する技術を開発した（図 4）。

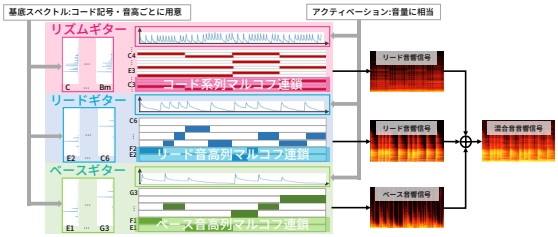

図 4：各ギターパートの音符系列と混合音スペクトログラムの階層ベイズモデル

さらなる研究展開として、音楽の創作支援や演奏支援にも取り組んだ。まず、数小節からなるメロディあるいはコード進行に対して、もう一方を自動生成することで、ユーザが対話的に生成結果を洗練させていくことができる自動作曲システムを開発した（図 5）。本研究では、コードとメロディに関する階層的生成モデルをあらかじめ学習することで、現在得られているメロディとコード進行から、ユーザが指定した箇所を事後分布に従ってサンプリングする手法を用いた。和声付けを行うには、コードの機能や繰り返し構造を考慮するために、コード進行の生成モデルとして木構造モデルである確率的文脈自由文法(PCFG)を用いた手法を用いた。コードに対するメロディ生成を行うには、メロディにおける音符の音高と開始位置に関するマルコフ性から、セミマルコフモデルに基づいてメロディの一部を生成する手法を提案した。

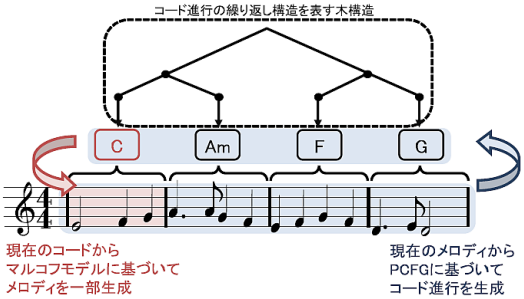

図 5：コード系列とメロディの音符系列の階層ベイズモデル

また、音楽音響信号からリアルタイムで伴奏音を抽出し、ユーザ歌唱のテンポ変化に自動で同期させて再生するカラオケシステムを開発した（図 6）。歌いたい曲を選択すると、すぐに伴奏音の分離が開始し、ユーザは伴奏音に合わせて歌を歌える。ユーザが歌唱のテンポを速くしたり遅くしたりすると、伴奏音のテンポもそれに合わせて変化する。インタフェース上にはユーザ歌唱と元の楽曲それぞれのスペクトログラムおよび F0 が表示され、ユーザはそれらをリアルタイムに比較できる。ユーザは自分が歌いたい曲の音源を準備するだけでシステムを利用できる。

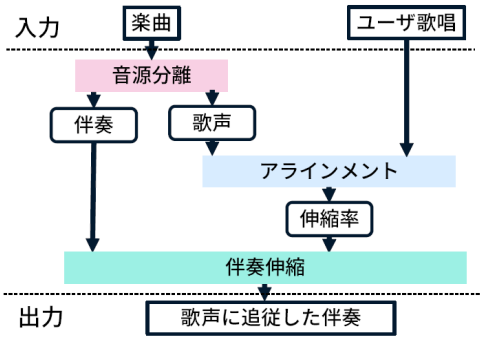

図 6：リアルタイム歌声・伴奏音分離と DTW に基づく適応的カラオケシステム

ポピュラー音楽の音楽音響信号を対象として、伴奏つきの歌声に含まれる歌唱表現を GUI 上で自由に編集することができるシステムを開発した（図 7）。歌唱表現とはビブラートやグリッサンド、こぶしなどの F0 の特徴的な局所変動のことを意味する。GUI 上には自動推定された歌声の F0 軌跡が表示されており、ユーザはロングトーン部など任意の範囲を指定し、事前に用意した歌唱表現（別の歌声から抽出することも可能）を選択・付与することができる。歌声 F0 の自動推定には誤りが含まれることを前提として、ユーザはスペクトログラム上で歌声の F0 が含まれるであろう領域をラフに塗りつぶすだけの簡易な操作により、その領域中で歌声の F0 が自動的に再推定する機能も開発した。人間と機械の協調に基づく音源分離や音楽解析は近年非常に注目されており、本システムはインタフェースを介してユーザと計算機が密接に関連しあっている。

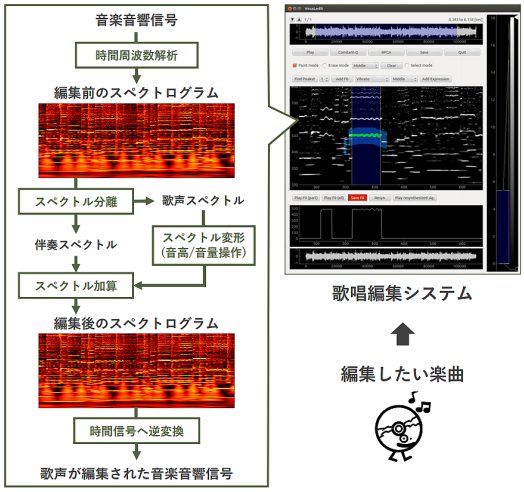

図 7：歌声・伴奏音分離と歌声 F0 推定に基づく混合音中の歌声編集システム

一連の研究を通じて、信号処理のための音響モデルと記号処理のための言語モデルを統合することで、自動採譜と文法学習を同時に行う方法を確立するという目的を達成することができた。さらに、開発した音楽解析技術を応用して、創作・編曲・演奏支援アプリケーションの開発を行うことができ、学術的・工学的に重要な貢献を行うことができた。

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計 5 件）

① Hiroaki Tsushima, Eita Nakamura, Katsutoshi Itoyama, and Kazuyoshi Yoshii: “Generative Statistical Models with Self-Emergent Grammar of Chord Sequences,” Journal of New Music Research, pp. 1–23, 2018.
② Eita Nakamura, Kazuyoshi Yoshii, and Simon Dixon: “Note Value Recognition for Rhythm Transcription Using a Markov Random Field Model for Musical Scores and Performances of Piano Music,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 25, No. 9, pp. 1846–1858, 2017.
③ Eita Nakamura, Kazuyoshi Yoshii, and Shigeki Sagayama: “Rhythm Transcription of Polyphonic Music Based on Merged-Output HMM for Multiple Voices,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 25, No. 4, pp. 794–806, 2017.
④ Misato Ohkita, Yoshiaki Bando, Yukara Ikemiya, Eita Nakamura, Katsutoshi Itoyama, and Kazuyoshi Yoshii: “Audio-Visual Beat Tracking Based on a State-Space Model for a Robot Dancer Performing with a Human Dancer,” Journal of Robotics and Mechatronics, Vol. 29, No. 1, pp. 125–136, 2017.
⑤ Yukara Ikemiya, Katsutoshi Itoyama, and Kazuyoshi Yoshii: “Singing Voice Separation and Vocal F0 Estimation Based on Mutual Combination of Robust Principal Component Analysis and Subharmonic Summation,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol. 24, No. 11, pp. 2084–2095, 2016.

〔学会発表〕（計 26 件）

① Kazuyoshi Yoshii: “Correlated Tensor Factorization for Audio Source Separation,” IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2018), Accepted, April 2018.
② Eita Nakamura, Emmanouil Benetos, Kazuyoshi Yoshii, and Simon Dixon: “Towards Complete Polyphonic Music Transcription: Integrating Multi-pitch Detection and Rhythm Quantization,” IEEE International Conference on

Acoustics, Speech, and Signal Processing (ICASSP 2018), Accepted, April 2018.

③ Ryo Nishikimi, Eita Nakamura, Masataka Goto, Katsutoshi Itoyama, and Kazuyoshi Yoshii: "Scale- and Rhythm-Aware Musical Note Estimation for Vocal F0 Trajectories Based on a Semi-Tatum-Synchronous Hierarchical Hidden Semi-Markov Model," International Society for Music Information Retrieval Conference (ISMIR 2017), pp. 376–380, October 2017.

④ Hiroaki Tsushima, Eita Nakamura, Katsutoshi Itoyama, and Kazuyoshi Yoshii: "Function- and Rhythm-Aware Melody Harmonization Based on Tree-Structured Parsing and Split-Merge Sampling of Chord Sequences," International Society for Music Information Retrieval Conference (ISMIR 2017), pp. 502–508, October 2017.

⑤ Eita Nakamura, Kazuyoshi Yoshii, and Haruhiro Katayose, "Performance Error Detection and Post-Processing for Fast and Accurate Symbolic Music Alignment," International Society for Music Information Retrieval Conference (ISMIR 2017), pp. 347–353, October 2017.

⑥ Kazuyoshi Yoshii, Eita Nakamura, Katsutoshi Itoyama, and Masataka Goto: "Infinite Probabilistic Latent Component Analysis For Audio Source Separation," IEEE International Workshop on Machine Learning for Signal Processing (MLSP 2017), July 2017.

⑦ Yuta Ojima, Tomoyasu Nakano, Satoru Fukayama, Jun Kato, Masataka Goto, Katsutoshi Itoyama, and Kazuyoshi Yoshii: "A Singing Instrument for Real-Time Vocal-Part Arrangement of Music Audio Signals," Sound and Music Computing Conference (SMC 2017), pp. 443–449, July 2017.

⑧ Yusuke Wada, Yoshiaki Bando, Eita Nakamura, Katsutoshi Itoyama, and Kazuyoshi Yoshii, "An Adaptive Karaoke System that Plays Accompaniment Parts of Music Audio Signals Synchronously with Users' Singing Voices," Sound and Music Computing Conference (SMC 2017), pp. 110–116, July 2017.

⑨ Eita Nakamura, Kazuyoshi Yoshii, and Shigeki Sagayama: "Rhythm Transcription of MIDI Performances Based on a Merged-Output HMM for Multiple Voices," Sound and Music Computing Conference (SMC 2016), pp. 338–343, September 2016.

⑩ Eita Nakamura, Katsutoshi Itoyama, and Kazuyoshi Yoshii: "Rhythm Transcription of MIDI Performances Based on Hierarchical Bayesian Modelling of Repetition and Modification of Musical Note Patterns," European Signal Processing Conference (EUSIPCO 2016), pp. 1946–1950, August 2016.

⑪ Yuta Ojima, Eita Nakamura, Katsutoshi Itoyama, and Kazuyoshi Yoshii: "A Hierarchical Bayesian Model of Chords, Pitches, and Spectrograms for Multipitch Analysis," The 17th International Society for Music Information Retrieval Conference (ISMIR 2016), pp. 309–315, August 2016.

⑫ Ryo Nishikimi, Eita Nakamura, Katsutoshi Itoyama, and Kazuyoshi Yoshii: "Musical Note Estimation for F0 Trajectories of Singing Voices Based on a Bayesian Semi-Beat-Synchronous HMM," The 17th International Society for Music Information Retrieval Conference (ISMIR 2016), pp. 461–467, August 2016.

⑬ Kazuyoshi Yoshii, Katsutoshi Itoyama, and Masataka Goto: "Student's t Nonnegative Matrix Factorization and Positive Semidefinite Tensor Factorization for Single-Channel Audio Source Separation," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2016), pp. 51–55, March 2016.

⑭ Eita Nakamura, Masatoshi Hamanaka, Keiji Hirata, and Kazuyoshi Yoshii: "Tree-Structured Probabilistic Model of Monophonic Written Music Based on the Generative Theory of Tonal Music," IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2016), pp. 276–280, March 2016.

⑮ Kazuyoshi Yoshii, Katsutoshi Itoyama, and Masataka Goto: "Infinite Superimposed Discrete All-Pole Modeling for Source-Filter Decomposition of Wavelet Spectrograms," The 16th International Society for Music

Information Retrieval Conference (ISMIR 2015), pp. 86–92, October 2015.

⑯ Misato Ohkita, Yoshiaki Bando, Yukara Ikemiya, Katsutoshi Itoyama, and Kazuyoshi Yoshii: “Audio-Visual Beat Tracking Based on a State-Space Model for a Music Robot Dancing with Humans,” IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2015), pp. 5555–5560, September 2015.

⑰ Ayaka Dobashi, Yukara Ikemiya, Katsutoshi Itoyama, and Kazuyoshi Yoshii: “A Music Performance Assistance System Based on Vocal, Harmonic, and Percussive Source Separation and Content Visualization for Music Audio Signals,” The 12th Sound and Music Computing Conference (SMC 2015), pp. 99–104, July 2015.

⑱ Tsubasa Fukuda, Yukara Ikemiya, Katsutoshi Itoyama, and Kazuyoshi Yoshii: “A Score-Informed Piano Tutoring System with Mistake Detection and Score Simplification,” The 12th Sound and Music Computing Conference (SMC 2015), pp. 105–110, July 2015.

⑲ Yukara Ikemiya, Kazuyoshi Yoshii, and Katsutoshi Itoyama: “Singing Voice Analysis and Editing Based on Mutually Dependent F0 Estimation and Source Separation,” IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2015), pp. 574–578, May 2015.

⑳ Satoshi Maruo, Kazuyoshi Yoshii, Katsutoshi Itoyama, Matthias Mauch, and Masataka Goto: “A Feedback Framework for Improved Chord Recognition Based on NMF-Based Approximate Note Transcription,” IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2015), pp. 196–200, May 2015.

上記他国際会議 6 件、この他国内発表多数。

〔その他〕
ホームページ：
http://sap.ist.i.kyoto-u.ac.jp/members/yoshii/

６．研究組織
(1)研究代表者
吉井 和佳（YOSHII, Kazuyoshi）
京都大学・情報学研究科・講師
研究者番号：20510001

(2)研究分担者
なし

(3)連携研究者
なし

(4)研究協力者
なし