公開

研 究 終 了 報 告 書

「人と AI の同化に基づく能力拡張型音楽理解・創作基盤」

研究期間：2020 年 12 月～2024 年 3 月
研 究 者：吉井 和佳

1. 研究のねらい

本研究で当初計画した目的は、大量の音楽データから、人と同様の理解様式で、解釈可能な音楽知識を獲得（暗黙知を形式知へ変換）するとともに、抽象記号（楽譜）と物理信号（波形）の両面を持つ音楽データのシームレスな操作を可能にする音楽 AI を実現することであった。一般的に、音楽の操作は専門教育を受けた者の特権であると考えられている。しかし、専門教育を受けずとも、人は音楽を楽しむことができるし、鑑賞体験を通じて音楽知識を蓄積することが、創作の基盤をなしている。本研究では、音楽 AI により、暗黙的な音楽知識を統計的な偏りとして明示的に表現することで、人が自らの音楽理解を深め、創作の支援を行うことを目指した。

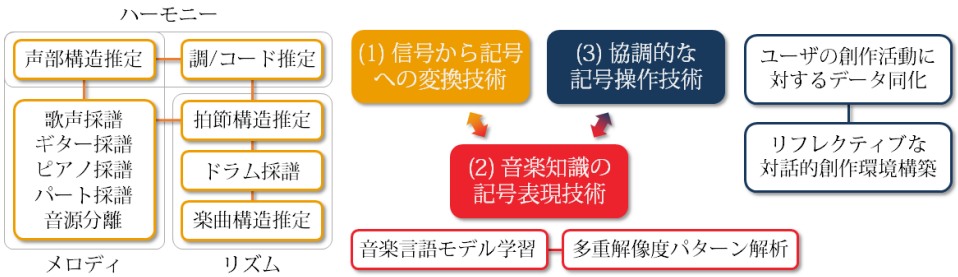
本研究における技術的な核心は、人間の音楽理解の計算モデルとして、音楽データから自律的に知識獲得可能な統計的推論法の設計であった。人の音楽理解とは、単に、信号→楽譜の推論を意味しない。認知学における運動理論（モーター・セオリー）仮説を援用すれば、人の脳は、楽譜→信号の生成をシミュレーションすることで、楽譜の整合性を検証していると考えることができる。これら推論過程と生成過程は、変分自己符号化器 (VAE) の枠組みを用いて自然に統合できる。また、人は、音楽理解の際に、音楽知識に照らして、楽譜の妥当性を検証している。そこで、楽譜→音楽知識の推論過程と、音楽知識→楽譜の生成過程の統合も考えられる。両者を確率的に統合した階層モデルは、理論上は教師なし学習できることから人の音楽理解の仕組みと親和性が高く、運動理論仮説に情報学的説明を与えることができると考えた。

音楽創作支援においては、当初、人の創作活動に対してデータ同化させることにより、人と音楽 AI とのペア作曲と呼ぶべき対話的な作曲様式を可能にすることを目指した。これは、計算機で大量のデータを処理して得た形式知と、計算機では表現しきれない各個人が持つ暗黙知を相補的に融合する試みであり、人の能力拡張の有望な一方式となりえると考えた。

2. 研究成果

(1) 概要

信頼される音楽 AI の実現に向けて、当初の計画通り三つの技術的な課題（下図）に沿って研究を進め、いずれも一定の成果を挙げることができた。解析に関する①については世界をリードする技術的進展があった一方、創作に関する③については制御可能な自動編曲の技術基盤を開発した段階であり、ユーザ意図の反映に関しては今後のさらなる研究が必要である。



① 信号から記号への変換技術

第一の課題は、音楽音響信号を楽譜に変換する「真の自動採譜」を実現することである。これは多くの研究が、計算機上での取り扱いに都合がよいピアノロールに変換していることと根本的に異なる。本研究では主にポピュラー音楽の音響信号を対象として研究を進めた。

- メロディ解析:ポピュラー音楽の核をなす**歌声**を対象とした研究を行った。まず、音符系列に対するマルコフ言語モデルと、音楽信号に対する深層音響モデルからなる生成モデルを学習し、音楽的に妥当な推論を行う歌声採譜を実現した [Nishikimi+ 2021]。推定された楽譜になお残存する誤りを訂正するため、置換・削除・挿入誤りの生成モデルを構成し、真の楽譜を推論する採譜誤り訂正法を考案した [Hiramatsu+ 2021]。このように、生成モデルに基づく統計的推論に取り組む一方で、深層学習に基づく推論モデルの End-to-End (E2E) 学習にも取り組んだ。音響信号から歌声の音符・文字系列を同時に推論する手法 [Deng+ 2022]や、歌声の音符系列と拍節位置を同時に推論する手法 [Deng+ 2023] を考案した。
- リズム解析:リズムを形成する主要な要素である**拍節構造**・**ドラム**・**楽曲構造**を対象とした研究を行った。まず、ビート系列に対する周期的マルコフ言語モデルと、音楽信号に対する深層音響モデルからなる生成モデルを学習し、音楽的に妥当な推論を行う拍節構造推定を実現した [大山+ 2022]。深層学習に基づくドラム採譜においては、言語モデルを用いた正則化 [石塚+ 2021]、周期性を用いた正則化 [Kamakura+ 2023]、VAE に基づく推論モデルの同時学習、拍節構造との End-to-End 学習 [Kamakura+ 2023]を考案した。また、自己注意機構を用いた楽曲構造解析手法 [Chen+ 2023] を考案した。
- ハーモニー解析:ポピュラー音楽の**調・コード**を対象とした研究を行った。調・コード系列に対するマルコフ言語モデルと、音楽信号に対する深層音響モデルからなる生成モデルに対し、推論モデルを組み合わせた VAE に基づく手法を考案した [Wu+ 2022]。

その他、ピアノ採譜における**声部構造**推定 [Shibata+ 2021] やガウス過程に基づく新しいモノラル**音源分離**法 [Nugraha+ 2023] を考案した。

② 音楽知識の記号表現技術

第二の課題は、楽譜の背後に存在する暗黙的な音楽知識を再利用可能な形式で体系化することである。まず、記号データである楽譜データに対して、音楽の本質である**周期構造**をとらえる**音楽言語モデル**を考案し、リズム採譜における有効性を実証した [Nakamura+ 2021]。また、①で述べた通り、各種音楽要素(**歌声**・**調/コード**・**拍節構造**・**ドラム**・**楽曲構造**)の解析に用いた**音楽言語モデル**は当該要素の確率的な偏りを表現しており、創作支援にも有用である。これらには、音楽の階層性を反映した**多重解像度パターン解析**の観点から、調・コード、ダウンビート・ビート、楽曲構造・コード・ビートの階層性に基づく言語モデルを含む。さらに、感情認識において、音響信号に加えて、楽譜情報を用いる手法を考案した [Zhao+ 2023]。

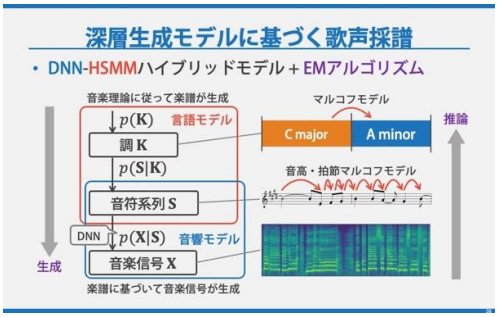
③ 協調的な記号操作技術

第三の課題は、形式知に変換された音楽知識を用いて、ユーザの音楽創作能力の拡張を行うことである。まず、深層学習を用いたバンド譜からピアノ譜への自動編曲において、 正解が一意でないことを考慮した学習法 [Terao+ 2022] と、難易度をユーザの技量に合わせて無段階に**対話的に調節**する手法 [Terao+ 2023] を考案した。また、ピアノ譜を任意編成の吹奏楽譜に編曲する手法 [Nabeoka+ 2023] を考案した。データ同化の実現に向けて、吹奏楽譜からピアノ譜を自動生成する**データ拡張**を行う手法の効果を実証した。

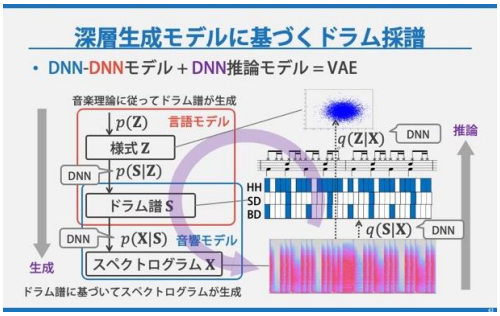
(2)詳細

各課題における代表的な成果を説明する。

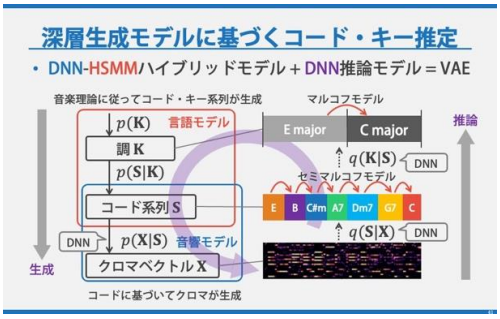
① 信号から記号への変換技術



- メロディ解析:歌声採譜[Nishikimi+ 2021]
  調・音符系列の階層的な生成過程を表現する言語モデルと、音符系列から音響スペクトログラムの生成過程を表現する音響モデルを組み合わせた生成モデルを定式化した。ここで、調や音符の音高・拍節位置の遷移は(セミ)マルコフモデルで効果的に記述できる一方、多様なスペクトログラムの生成過程を明示的に記述することは容易ではない。そのため、スペクトログラムから音符系列を推定する DNN を予め教師あり学習しておき、ベイズの定理を用いて生成モデルに組み入れることで DNN-HSMM を構成する方式を考案した。このモデルに対しては、最尤の調・音符系列をビタビ推論できる利点がある。

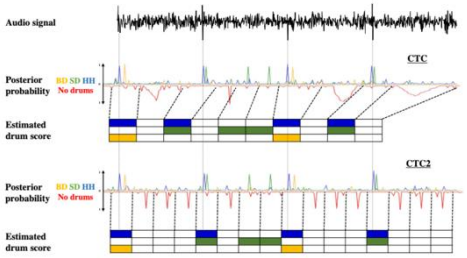


- リズム解析:ドラム採譜
  ドラム譜の生成過程を表現する言語モデルと、ドラム譜から音響スペクトログラムの生成過程を表現する音響モデルを組み合わせた生成モデルを定式化した。ただし、ドラム譜もスペクトログラムもいずれも明示的に生成過程を記述することは困難であるため、双方への DNN の導入を行った。DNN に基づく推論モデルを導入して VAE を構成し、生成モデルと推論モデルを一挙に学習することを可能にした。これにより、運動理論仮説を構成論的に実証した。



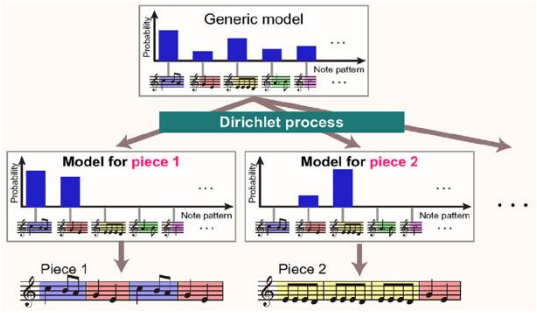
- ハーモニー解析:調・コード推定 [Wu+ 2022]
  調・コード系列の階層的な生成過程を表現する言語モデルと、コード系列からクロマグラムの生成過程を表現する音響モデルを組み合わせた生成モデルを定式化した。歌声採譜と同様に、言語モデルはセミマルコフモデルで記述できる一方で、音響モデルはDNNを用いて表現力を強化することとした。ドラム採譜同様に、DNN に基づく推論モデルを導入して VAE を構成することで、生成モデルと推論モデルを一挙に学習することを可能にした。

音響信号を記号系列に変換する E2E 学習において、出力記号の継続時間長を考慮した CTC2 [Kamakura+ 2023] を考案した。本技術は汎用性が高く、他分野に幅広く応用可能な<u>汎用機械学習法</u>である。従来の CTC では、入出力系列のアラインメントにおいて任意の伸縮を許容するため、一定のテンポをもつポピュラー音楽のドラム採譜には適切ではなかった。CTC2 はこの問題を解決する一方で、可能な全てのアラインメントパスの数への誤差逆伝播は困難になる。そこで、ビタビ学習あるいはギブスサンプリングを用いて精度を犠牲にせず高速化する手法を考案した。



② 音楽知識の記号表現技術

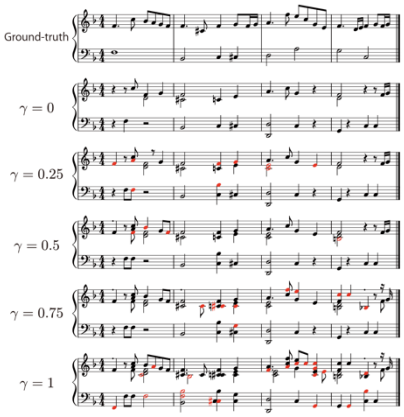
- 周期性に基づく音楽言語モデル[Nakamura+ 2021]



音楽の自己相似性を形成する重要な要素である周期性に着目し、メロディに対する音楽言語モデルを定式化した。音楽全体における小節単位の音符パターンの確率分布がまず存在し、各楽曲における音符パターンはその分布をより先鋭化(スパース化)した分布に従ってマルコフ遷移すると考えるのが自然である。この生成過程をディリクレ過程でモデル化することで、階層的なメロディ言語モデルの教師なし学習を実現した。学習したモデルを、時間的に量子化されていない演奏 MIDI データを楽譜データに変換するリズム採譜タスクに応用すると、周期性(自己再現性)を考慮することによって精度が向上することを実証した。

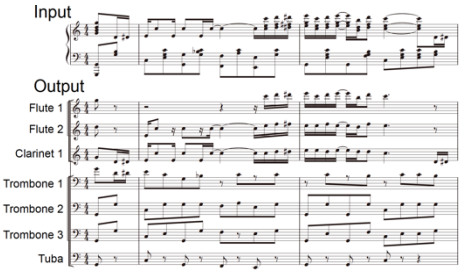
③ 協調的な記号操作技術

- バンド譜に対するピアノ編曲 [Terao+ 2022/2023]



深層学習を用いて、難易度を無段階に制御できるピアノ編曲を実現した。この種の生成課題では、正解は一意には定まらないことから、単純な教師あり学習は適切ではない。そこで、出力された楽譜に対し、それ自体の正解との誤差に加えて、統計量(音符密度や音高幅など)レベルでの誤差を考慮する手法を考案した。学習データには初級と上級のピアノ譜のみが利用可能であるため、Wasserstein 計量を用いた任意難易度の統計量の推定を考案した。

- ピアノ譜に対する吹奏楽編曲 [Nabeoka+ 2023]



深層学習を用いて、任意編成の吹奏楽譜へ編曲する手法を考案した。学習データ不足を補うため、吹奏楽譜に対するピアノ編曲手法を用いてペアデータを生成し、実験的にその有効性を確認した。また、DNN の内部で楽器パート間の依存関係が学習されていることが示唆された。

両課題について、使いやすい編曲インタフェースの実装は今後の研究課題である。

3. 今後の展開

三つの研究課題について、さらなる改良に取り組む予定である。「①信号から記号への変換技術」に関しては音楽解析技術が成熟しつつあるので、音楽配信サービスや音楽マーケティング企業へのライセンシングは現在でも可能である。ポピュラー音楽やピアノ曲を MIDI 形式ではなく、**楽譜形式で採譜可能な技術を保有**しており、国際的に競争力がある。「②音楽知識の記号表現技術」については、大規模音楽言語モデルの研究を継続する。楽譜出力が可能な採譜技術と組み合わせることで、Web 上に存在する大量の音楽音響信号を楽譜化し、音楽言語モデルを学習し、それを用いて採譜精度をさらに向上させるような正の循環が期待できる。「③協調的な記号操

作技」については、信頼できるユーザインタフェースの実装に加えて、近年注目される**AR/MR 技術との統合**も視野に入れた研究展開を検討している。音楽生成については未だ評価尺度も確立されておらず、5 年以上のスパンでの継続的な研究が必要になる。

4. 自己評価

信頼される音楽 AI の実現に向けて、当初の計画通り三つの技術的な課題に沿って研究を進め、いずれも一定の成果を挙げることができた（詳細は 2.参照）。実際、3 年余りで計 28 件の論文発表を行うことができた。全体として、当初想定した以上の研究の水平展開があったため、「③協調的な記号操作技」において、ユーザが制御可能な基盤技術の創出には成功した一方で、ユーザインタフェースの実装は今度の課題としたい。

本研究課題は、自身が研究の方針とアイデアを考案し、補助員および学生の支援を受けながら効率的に進めることができた。研究費の用途の主要なものは、補助員および学生の雇用経費と、国内外での研究発表のための旅費、計算機購入費用であり、これらは当初の計画通りバランスがとれた支出となった。

社会・経済への波及効果に関しては、音楽配信サービスや音楽マーケティング企業へのライセンシングが第一に考えられる。実際、複数の問い合わせや相談を受けており、具体的な社会実装に繋がる可能性は十分にある。人的リソースの制約上、技術開発は企業が担当し、技術指導を行う形式を検討している。

5. 主な研究成果リスト

（1）代表的な論文（原著論文）発表

研究期間累積件数：5件

1. Yiming Wu, Kazuyoshi Yoshii, “Joint Chord and Key Estimation Based on a Hierarchical Variational Autoencoder with Multi-Task Learning”, APSIPA Transactions on Signal and Information Processing, Vol. 11, No. 1, pp. 1–27, 2022.
調・コード系列の階層的な生成過程を表現する言語モデルと、コード系列からクロマグラムの生成過程を表現する音響モデルからなる生成モデルに基づく調・コード推定法を考案した。
2. Kentaro Shibata, Eita Nakamura, Kazuyoshi Yoshii, “Non-Local Musical Statistics as Guides for Audio-to-Score Piano Transcription”, Information Sciences, Vol. 566, pp. 262–280, 2021.
最新の深層多重音高推定手法に対して、音楽言語モデルに基づくリズム採譜手法・パート分離手法を組み合わせたピアノ採譜手法を考案した。
3. Ryo Nishikimi, Eita Nakamura, Masataka Goto, Kazuyoshi Yoshii, “Audio-to-Score Singing Transcription Based on a CRNN-HSMM Hybrid Model”, APSIPA Transactions on Signal and Information Processing, Vol. 10, No. e7, pp. 1–13, 2021.
調・音符系列の階層的な生成過程を表現する言語モデルと、音符系列から音響スペクトログラムの生成過程を表現する音響モデルからなる生成モデルに基づく歌声採譜法を考案した。

（2）特許出願

なし

(3) その他の成果(主要な学会発表、受賞、著作物、プレスリリース等)

- 2021/06/15 京都大学プレスリリース「ピアノ演奏を楽譜に書き起こす「耳コピ AI」 ー実用に近いレベルの楽譜生成に初めて成功ー」
  https://www.kyoto-u.ac.jp/ja/research-news/2021-06-15-1
- 2023/09/12 コロナ社 メディアテクノロジーシリーズ 2「音楽情報処理」執筆(分担)
  https://www.coronasha.co.jp/np/isbn/9784339013726/