

信頼される AI の基盤技術
2020 年度採択研究代表者

2022 年度
年次報告書

吉井 和佳

京都大学 大学院情報学研究科
准教授

人と AI の同化に基づく能力拡張型音楽理解・創作基盤

研究成果の概要

第三年次は、当初の計画通り、人間との親和性・利便性を念頭に置いたうえで、「(1)信号から記号への変換技術」に関して継続的に改良・拡張を行うとともに、「(2) 音楽知識の記号表現技術」および「(3) 協調的な記号操作技術」に関する研究を拡充した。具体的には、(1)に関しては、ポピュラー音楽の音響信号を対象として、信号ドメインでの音響モデルおよび記号ドメインでの言語モデルを統合することで、音楽の三大要素であるメロディ・ハーモニー・リズムを高精度に解析する音楽理解システムの開発に継続して取り組んだ。リズムに関して、拍節構造推定におけるビート・ダウンビートのマルチタスク学習において、両者の出力間の不整合問題を解消するため、深層隠れセミマルコフモデルに基づく推論を考案した。メロディに関して、歌声の音高推定と歌詞認識を一挙に行う End-to-End 型のマルチタスク学習法を考案した²⁾。ハーモニーに関して、コードとキーとのマルチタスク学習において、変分自己符号化器の教師あり・半教師あり学習の洗練化を行った¹⁾。その他、多様な楽器音から音色・音高・変動に分離した潜在表現を教師なし学習する方法⁴⁾や、ある楽曲が歴史的にどのように演奏されてきたかを追跡する技術³⁾を提案した。(2)に関しては、ポピュラー音楽のピアノ編曲において、初級と上級のピアノ譜のもつ統計モデルの学習と任意の難易度に対応する統計モデルの補完法の提案を行った。また、自然言語処理分野で注目されているトランスフォーマ型の言語モデルの教師なし・自己教師あり学習に関する検討を開始した。(3)に関しては、自動ピアノ編曲において、深層学習を用いて、初級・上級のみでのピアノ譜から構成された学習データから、ユーザに合わせた難易度をもつピアノ譜の生成を可能にし、定量評価および被験者実験による有効性を確認した⁵⁾。

【代表的な原著論文情報】

- 1) Yiming Wu and 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.
- 2) Tengyu Deng, Eita Nakamura, and Kazuyoshi Yoshii, “End-to-End Lyrics Transcription Informed by Pitch and Onset Estimation”, *International Society for Music Information Retrieval Conference (ISMIR)*, pp. 633–639, 2022.
- 3) Florian Thalmann, Eita Nakamura, and Kazuyoshi Yoshii, “Tracking the Evolution of a Band's Performances over Decades”, *International Society for Music Information Retrieval Conference (ISMIR)*, pp. 850–857, 2022.
- 4) Keitaro Tanaka, Yoshiaki Bando, Kazuyoshi Yoshii, and Shigeo Morishima, “Unsupervised Disentanglement of Timbral, Pitch, and Variation Features From Musical Instrument Sounds With Random Perturbation”, *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp. 709–716, 2022.
- 5) Moyu Terao, Yuki Hiramatsu, Ryoto Ishizuka, Yiming Wu, and Kazuyoshi Yoshii, “Difficulty-Aware Neural Band-to-Piano Score Arrangement Based on Note- and Statistic-Level Criteria”, *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 196–200, 2022.