

ピアノ演奏を楽譜に書き起こす「耳コピ AI」

— 実用に近いレベルの楽譜生成に初めて成功 —

概要

京都大学白眉センター/大学院情報学研究科の中村栄太 特定助教、柴田健太郎 情報学研究科 修士課程学生 (研究当時)、吉井和佳 同准教授らの研究グループは、ピアノ演奏音声データから楽譜を自動で生成する技術を開発して、実用に近いレベルの楽譜生成に世界で初めて成功しました。音楽の演奏を聴いて楽譜に書き起こす「採譜」は耳コピとも呼ばれ、特殊な訓練をした人だけがもつ能力です。誰でも好きな曲を楽譜にして演奏などに使えるようにするため、この能力をコンピューターで再現する「自動採譜技術」の研究が行われています。特に、ピアノ演奏の採譜は、音高（ピッチ）とリズムの複雑な組み合わせを認識する必要があり、非常に難しい問題です。本研究では、大量のデータを用いた機械学習^(注1)と人間演奏のモデルに基づく方法で、ピアノ演奏の自動採譜の精度を大幅に向上させ、演奏や人による採譜の補助に部分的に用いることができる楽譜の生成に成功しました[1]。今後は、実用化に向けた研究開発、音楽学研究や音楽教育への応用、そして文化を支える知能の科学的理解につながると期待されます。また、著作権法や音楽家への影響の問題も提起しています。本成果は、2021年3月13日に米国の国際学術誌「Information Sciences」にオンライン掲載されました。

[1] 採譜結果の例は <https://audio2score.github.io/index-ja.html> に多数掲載されています。



1. 背景

インターネット上には膨大な数の音楽の動画や音声ファイルがあり、新しい曲が時々刻々と追加されています。こうした音楽を演奏、編曲または分析するには楽譜として表す必要があります。ポピュラー音楽などの多くの曲は、訓練した専門家が音を聴いて音符として書き起こすことで楽譜にされています。これは採譜や耳コピと呼ばれ、多くの時間を必要とする作業のため、世の中の多くの曲は楽譜になっていません。より多くの曲を演奏や分析できるようにするため、コンピューターによって採譜を自動化する研究が行われています。特に、ピアノ演奏のように複数の音が重なった音楽の採譜は、音高（ピッチ）とリズムの複雑な組み合わせを認識する必要がありますことから、この自動化は長年の間、非常に難しい問題とされてきました。実際、これまで提案されてきた手法はどれも精度が不十分で、自動生成された楽譜を演奏などに用いることは困難でした。

ピアノ採譜の問題は、音声信号の中で各時刻にどの音高が鳴っているかを推定する「多重音高検出」と、各音符の発音時刻と音の長さを拍（ビート）時間単位で認識する「リズム量子化」という二つの主要な問題に分けて考えることができます（図1）。それぞれの問題について、データから重要となる特徴や統計量を自動で抽出する「機械学習」を用いた方法が近年研究されており、大きな進展がありました。本研究では、これらの最先端の手法を統合する世界初の試みにより、高精度な自動ピアノ採譜システムの構築に挑みました。

2. 研究手法・成果

本研究では、機械学習をそれぞれ用いた多重音高検出の手法とリズム量子化の手法を統合した自動ピアノ採譜システムを構築しました。前者の多重音高検出では、入力の音声データに対して、各時刻に含まれる音の高さと強さ、そして打鍵の有無を推定する深層ニューラルネットワークを用いました。後者のリズム量子化では、人間の演奏に含まれる時間的変動のモデルと、楽譜に現れる一般的なリズムパターンの統計的特徴を表すモデルに基づく手法を用いました。この結果、従来手法に比べて採譜の誤りを平均的に半数以下に低減できることを示しました。またこの採譜結果では、局所的に見れば音高やリズムの配置は正解に近いものの、テンポや拍子、小節線の位置など、音楽の大局的な特性に関する認識誤りが多いことがわかりました。

拍子や小節線の位置などの認識には、音楽フレーズの繰り返し構造、フレーズ境界での音楽特徴の大きな変化、コード進行やバス進行など、複数の音楽要素が関係していることが音楽学で知られています。そこで、音符同士の関係性を捉える統計量を用いて採譜精度をさらに改善する手法を考案しました。この結果、部分的に演奏や人による採譜の補助にも用いることができる楽譜の自動生成に成功しました。採譜結果の例はプロジェクトページ <https://audio2score.github.io/index-ja.html> に多数掲載されています。

今回の結果は、数十年來の未解決問題であった自動ピアノ採譜において、部分的に実用レベルの精度を達成した点で画期的で、自動採譜技術の実用化に向けた動きが今後活発化すると期待されます。また、人間が音楽を認識する際に用いている専門知識や知能の仕組みの理解につながると考えられます。

3. 波及効果、今後の予定

今回の研究はピアノ音楽を対象としていますが、同じ方法は歌声、ギター、ドラムなど他の楽器にも応用できるため、今後は多様な楽器編成の音楽の自動採譜へと研究が広がると期待されます。現段階での自動採譜結果は、専門的に見れば不満な点も多く、今後はより多くのデータを用いた機械学習による精度向上に加えて、楽譜表記法に関する専門知識を取り入れる研究が重要になると考えられます。また、強弱記号や装飾記号（アルペジオやトリルなど）を含むより詳細な楽譜要素の自動認識も可能にすることが今後の課題です。

自動採譜が可能になれば、インターネット上の演奏動画から楽譜を自動生成して誰でも演奏練習に使えるなど、より豊かな音楽文化をもたらす技術が実現できます。また、音楽専門家は採譜に費やす時間を短縮でき、その分、より詳しい音楽分析や創作活動に集中できるようになるでしょう。将来的には、音楽のデータベース化や大規模データ解析などを通して、芸術学や文化学を支える基盤技術になると期待されます。

他方で、自動採譜技術の普及により、音声ファイルから即座に楽譜が得られるようになれば、音楽の盗用などの著作権上の問題も起こると考えられます。また、音楽の専門家や学生の収入源にもなっている採譜の仕事が減ることで、音楽文化の発展に悪影響を及ぼす可能性もあります。自動採譜技術の実用化が現実的になってきた今、こうした文化の健全な発展に向けた法整備や技術の社会実装についての議論を早急に進める必要があります。

4. 研究プロジェクトについて

本研究は、日本学術振興会 科学研究費助成事業（課題番号 16H01744 および 19K20340）、科学技術振興機構 戦略的創造研究推進事業 ACCEL（課題番号 JPMJAC1602）、京都大学白眉プロジェクト、京都大学教育研究振興財団からの援助を受けて行われました。

<用語解説>

(注1) **機械学習**：データの分類や予測などをする計算法を自動的に構築する技術。最近では、人工知能（AI, Artificial Intelligence）と表現されることもあります（正確にはそれぞれ別の概念を表します）。

<研究者のコメント>

この研究は、情報技術や知能に関する研究と芸術文化（音楽）の領域を結びつけるものです。異なる分野を横断する研究の楽しさや可能性を感じてもらえれば嬉しいです。また、研究を一面的に見るのではなく、学問や技術の発展を社会や文化との関係の中で考えることの重要性を伝えていきたいと思います。（中村）

<論文タイトルと著者>

タイトル：Non-Local Musical Statistics as Guides for Audio-to-Score Piano Transcription

非局所音楽統計量を手がかりとして用いた音響から楽譜へのピアノ採譜

著者：Kentaro Shibata, Eita Nakamura, Kazuyoshi Yoshii

掲載誌：Information Sciences DOI：https://doi.org/10.1016/j.ins.2021.03.014

< 参考図表 >

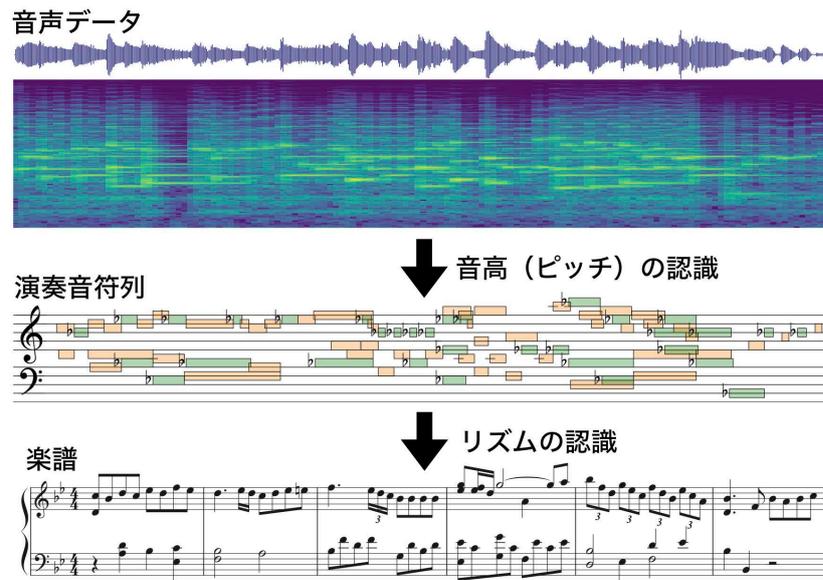


図1 ピアノ演奏音声データから楽譜を生成するには、各時刻における音高 (ピッチ) を認識する問題とビート時間単位でリズムを認識する問題が主要な問題になります。