

# 音楽生成 AI と創造性

関西学院大学 学生  
川口 竜斉  
関西学院大学 教授  
片寄 晴弘

## はじめに

2022 年秋に米国で ChatGPT のサービスが始まり、2023 年半ばあたりから、「生成 AI」というシズル感のあるタームが我々の社会や生活に浸透してきている。これに先立ち、IT 巨人 GAFAM の一角 Google が 2012 年頃から深層学習に関する研究を本格化させ、2016 年には AI が囲碁チャンピオンを凌駕し、翌年には、わずか 40 日の自己対局で、これまでに開発されたどの AI よりも強くなったというショッキングなニュースが流れた。AI がさまざまな面で我々をサポートしてくれる一方で、多くの仕事が無くなってしまふ (=技術的特異点 Singularity<sup>1</sup>)ことを予感させる象徴的な出来事であった。以降はご存知の通り、深層学習を中心とした技術開発 は加速度的に進展し、対象領域もシンボル(小規模言語)から大規模言語モデル、さらには画像や音声と範囲を広げてきた。その適用・応用範囲は人が関わるすべての活動に及び、芸術や教育分野も例外ではなくなりつつある。本稿では「音楽生成 AI と創造性」というテーマで、AI と音楽の関係について論じる。音楽制作プロセスの中でも作曲に焦点を当て、生成研究の概況を大きく深層学習以前・以降に分けて紹介する。人間の音楽創作プロセスから見て、AI に何ができて何ができていないのかについて述べ、最後にこの領域の今後について私見を述べてみたい。

## 音楽生成 AI の歴史

ここでは、音楽生成 AI がどのように発展してきたかについていくつか具体例を見ながら振り返っていく。システムを時系列に並べたものを図 1 に示す。音楽生成 AI の仕組みは深層学習が登場する前後で大きく変化していることから、その前後に分けて見ていくことにする。

コンピュータによる自動作曲は、「イリアック組曲」(1957 年)に遡る。それ以降、深層学習が登場する前までの自動作曲は、事前に組んでおいたアルゴリズムと統計処理に基づくものであった。代表的な自動作曲システムの例として、David Cope が 1991 年に発表した「EMI」を取り上げる<sup>[1]</sup>。EMI は「作曲とは、今までにつくられた作品の事例の解析と再合成によってなされる」という考えに基づいて構成されている。与えた作品のモチーフを抽出するプロセスとルールを解析するプロセスの二つによって、作品の傾向を確率モデルに落とし込むことで楽曲を生成する。モチーフを抽出するプロセスでは、ピッチ・リズムに着目して、曲の様々な箇所に出現する同じ形と見做せるメロディ(モチーフ)を探索する。

---

<sup>1</sup> 思想家・未来学者のレイ・カーツワイルによって 2005 年に提唱された。AI の能力が人間の知性を大きく超え、2045 年には社会の姿が根本的に変わると予想される。現状の予測で早まるとの見方がある。ヒトの仕事の多くが機械(AI)に代替されるなかで、ヒトにやってもらいたい職業と人の楽しみ(エンタメ)に関する職業は残ると見込んでいる。

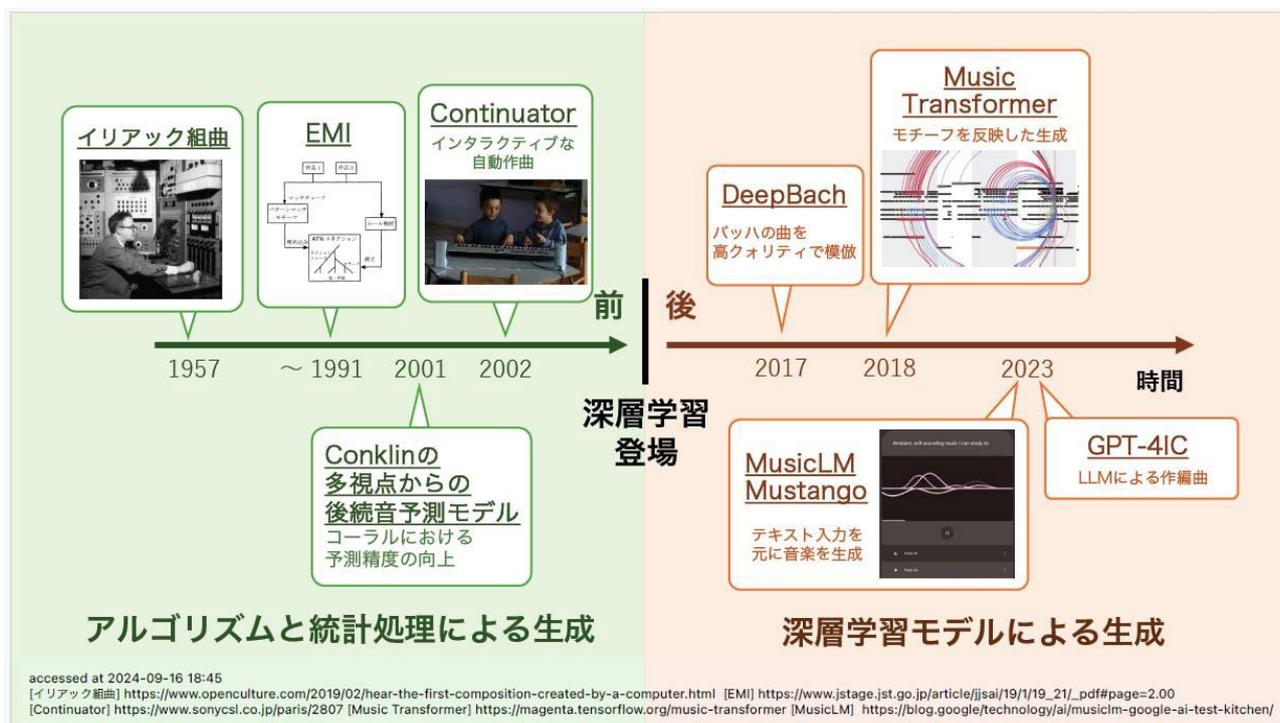


図 1 自動作曲研究の歴史

ルールを解析するプロセスでは、パートの進行方向・繰り返される音の数・和音の概形などの出現確率を算出する。これらの結果を統合して作品の統計的な傾向を確率モデルに落とし込み、その傾向にしたがって音列を再生成することで作曲を行う。アルゴリズムを用いた自動作曲においては、音楽の構成要素、例えば、音符(音の長さ と音の高さを表わす)がどのような要因によって決定されるのかの模索が重要ポイントとなる。Conklin は、リズム・メロディの形・コードや拍子など、音符の系列以外の複数の視点を加えて統合したモデル<sup>[2]</sup>を提案し、このモデルにより、バッハのコラールにおいて次の音を予測する精度を向上させることが確認した。ここまで自動作曲システムの動作の仕組みに着目してきたが、実際にはシステムの入力方法、すなわち、ユーザがシステムにどのように意図を伝達して所望の楽曲を生成するのかという課題もある。この問題に際し、自動作曲のプロセスに演奏者がいつでもインタラクティブに介入できるようにしたのが、Francois Pachet の「Continuator」である<sup>[3]</sup>。演奏者が演奏していればその人のスタイルを自動で学習していき、演奏者が演奏を止めればその後続の演奏を続けて生成する。このような操作形態をとることで、ユーザは作りたい楽曲をシステムに例示して意図を伝達できる。

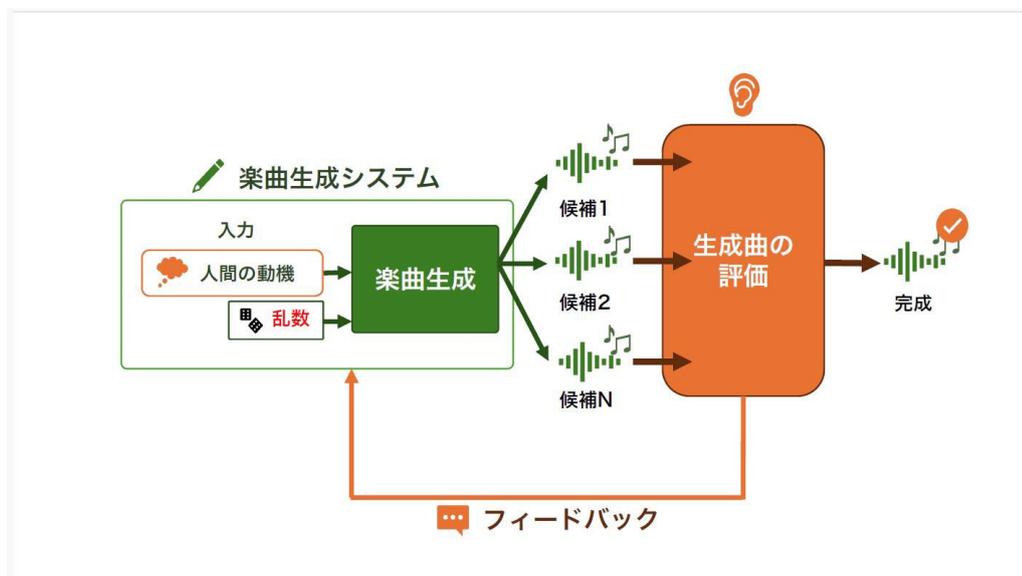
深層学習が登場してからは、自動作曲システムはその技術を取り込み飛躍的な発展を続けている<sup>[4]</sup>。深層学習による自動生成では、音楽の記号的な表現に基づいて生成を行う記号系と、音楽を音響信号で表現して生成を行う音響信号系に大きく分けることができる。記号系の生成において代表的なモデルに G. Hadjeres の「DeepBach」<sup>[5]</sup>や Huang の「Music Transformer」<sup>[6]</sup>がある。DeepBach では与えられたメロディーに対してバス・アルト・テノールパートの楽譜を生成するモデルである。その名の通り、バッハの楽曲とそれを転調させたもの 2800 曲程度を学習させており、400 名のプロのミュージシャンや音大学生を含む 1600 人が参加した聴取テストでは、半数が DeepBach の生成作品をバッハ作だと間違えるほどの性能を持つ。Music Transformer は与えられた楽曲に対して後に続く音を逐次的に生成するモデルである。従来の深層学習モデルでは捉えられなかった楽曲の繰り返し構造を把握するべく、楽譜上で過去にある音列のうちどの音符が最も関連するかを捉えられるモデル(Attention 機構)を用いる

ことで、モチーフを保持した楽曲生成ができるようになった。現在、簡単なテキスト入力を元に、高音質・高品質な音響信号フォーマットの音楽を直接生成するシステムの技術的進展が著しい。その代表的なものに A. Agostinelli の「MusicLM」<sup>[7]</sup>がある。テキストとしては、その曲が流れるであろう環境の様子、使われている楽器やテンポの指定も可能となっている。「Mustango」<sup>[8]</sup>では、コード進行や拍といったより音楽理論に沿った指示出しもできるようになっている。ただし、これら、音響信号生成系のシステムでは、一度生成された楽曲の修正が難しいという課題もある。そこで、私たちの研究室では、記号系の生成に LLM を用いた上でユーザからフリーフォーマットの指示を受けて楽曲を生成・編集する「GPT-4IC」<sup>[9]</sup>の開発を進めている。現地点の自動楽曲システムは深層学習を用いることで、制約はありつつも実用レベルに達しつつある。

## 音楽と創作

以上のような音楽生成 AI の発展を受け、近い将来に AI が作曲を代表とする人間の創作を代替するという見方が存在する。ここでは、人間の行う「作曲」がどのようなものかに立ち返り、システムが「作曲」し得るかどうかの可能性と、その際の音楽文化への影響について考える。

音楽生成 AI の多くは、音楽的制約を考慮したランダムな生成によって楽曲を作り出す。例えば、



Diffusion 系音楽生成システムはノイズからサンプリングしたデータをもとに楽曲を作る。このようなシステムには今まで世の中に存在しない楽曲を実体化する能力があり、使い方によっては、半永久的に音楽データ群を作り出すこともできる。果たしてこうして作られた音楽データ群

図2 楽曲を聴取してフィードバックする機構を備えたシステム

は創作物と言えるのだろうか？ 今までの音楽の歴史を考えれば、作曲者の「耳」のフィルターをパスできなかった数えきれない作品候補群が存在する<sup>[10]</sup>。また、一旦世に出ても聴衆の評価を得られず、結果として消えていった多数の作品群も存在する<sup>[11]</sup>。これらからすれば、初めに曲を通して表現したいという意図を作曲者が持ち、その動機に基づいて曲を実体化し、耳で聴取・評価して、修正や作り直しを行う手順(図2参照、以下、リフレクション)を繰り返し、所望の楽曲を得る流れが真の「作曲」のプロセスの、少なくとも、必要条件になるのではなかろうか。

音楽生成 AI によるいわゆる「実体化」の能力は極めて大きな進展を遂げつつある。これをもって、「創作」を含めた人間の作曲プロセスの全てが置き換えられると考えるのは少々短絡的であろう。今後は、研究者によるリフレクション機構の探求、ユーザによる Creative Direction スキルの拡大が進

み、その相互作用によって、音楽創作の新しい局面が拓かれると予想している。

## おわりに

作曲プロセスのうち、AIにより、楽譜データを作成する、楽譜を音にする、さらに、言語から高品位の曲データ(音響)を作り出すことが可能となりつつある。名作映画音楽は監督と作曲家のコンビネーションによって作られてきたが、ここでの作曲家は AI エージェントに置き換わる可能性があり、大量消費型のBGMの多くはAIに代替されるだろう。加えて、その AI エージェントが、誰もが利用可能であることに留意したい。

人間の創作活動における AI の利用度合いは、楽譜やレコーダーを利用するのと同程度に自然なものとなっていくだろう。異論を持つ方もいるかもしれないが、初音ミク以前にバーチャルシンガーがここまで浸透することを予想できた人がどれほどいただろうか。次代を支える Z 世代は物心ついた時からスマホを使っている。あえて便利さを手放す状況は考えにくい。また、プロ棋士が AI とともに棋力を上げつつある状況を想像していただきたい。

生成 AI が取り沙汰される少し前から、デジタル技術が人間の代替として利用される状況は進んでいる。ポップス音楽の領域では、打楽器演奏の多くが DAW(Digital Audio Workstation)のエージェントに置き換わり、さらに継時的な制御が求められる弦楽器や 管楽器においてもその方向が進みつつある。とはいえ、すべてがデジタル技術や AI に置き換えられるわけではない。人は人間が行うことに対して感動を求め、創り出すこと自体が人の根本的な希求だからである。コンピュータが発明されて以来、人間に比肩・凌駕する AI の実現は研究者にとって大きな夢であった。これらのうち、人間の代表的な知的活動の一つ「ゲーム」で人に勝つという目標は既に達成された。芸術領域ではどうであろうか。現時点では、芸術や創造性を説明する理論は存在せず、未開拓の領域として残されている。生成系 AI は、この領域での研究を進めるための実験環境として利用される。あくなき探求の夢は続き、この取り組みの結果が技術的特異点の一つとして振り返られる未来を予測する。

## 参考・引用文献

1. David Cope. Computers and musical style, Vol. 6. Citeseer, 1991.
2. Darrell Conklin and Christina Anagnostopoulou. Representation and discovery of multiple viewpoint patterns. In ICMC, pp. 479–485, 2001.
3. Francois Pachet. The continuator: Musical interaction with style. Journal of New Music Research, Vol. 32, No. 3, pp.333–341, 2003.
4. Shulei Ji, Xinyu Yang, and Jing Luo. A survey on deep learning for symbolic music generation: Representations, algorithms, evaluations, and challenges. ACM Computing Surveys, Vol. 56, No. 7, pp. 1–39, 2018.
5. Gaetan Hadjeres, Francois Pachet, and Frank Nielsen. Deepbach: a steerable model for bach chorales generation. In International conference on machine learning, pp. 1362–1371. PMLR, 2017.
6. Cheng-Zhi Anna Huang, Ashish Vaswani, Jakob Uszkoreit, Noam Shazeer, Ian Simon, Curtis Hawthorne, Andrew M Dai, Matthew D Hoffman, Monica Dinulescu, and Douglas Eck. Music transformer. arXiv preprint arXiv:1809.04281, 2018.
7. Andrea Agostinelli, Timo I Denk, Zal ‘an Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, et al. Musiclm: Generating music from text. arXiv preprint arXiv:2301.11325, 2023.
8. Jan Melechovsky, Zixun Guo, Deepanway Ghosal, Navonil Majumder, Dorien Herremans, and Soujanya Poria. Mustango: Toward controllable text-to-music generation. arXiv preprint arXiv:2311.08355, 2023.
9. 川口竜斎, 片寄晴弘. GPT-4 による対話的音楽生成の可能性と課題. 人工知能学会全国大会論文集, 1O4OS29a02–1O4OS29a02, 2024.
10. Karlheinz Stockhausen, 清水穰(訳). シュトックハウゼン音楽論集. エートル叢書, No. 1. 現代思潮社, 1999.
11. 片寄晴弘. 音楽生成と AI (特集 エンタテイメントにおける AI 技術). 人工知能, Vol. 19, No. 1, pp. 21–28, 2004.