

4 音楽とコンテンツ生成

深山 覚 (産業技術総合研究所)
後藤真孝 (産業技術総合研究所)

音楽コンテンツの自動生成

コンピュータを使って、音楽コンテンツ（楽譜やその演奏、歌詞等）を自動的に生成する研究の歴史は長い。コンピュータの発明とほぼ同時期に取り組み、たとえば1950年代には、黎明期の電子計算機であるILLIAC Iによって弦楽四重奏曲（ILLIAC Suite）が自動生成されていた¹⁾。そして、計算機・情報処理の発展とともに、音楽の自動生成技術は進歩してきた。当初は音楽を生成するために、作曲技法の専門知識を明示的に実装するところから始まり²⁾、しだいに音楽データベースを活用しはじめ³⁾、機械学習や統計的確率モデルに基づいて自動生成をする段階⁴⁾に至っている。その時代ごとの最先端の情報処理技術が、音楽コンテンツの生成に活かされてきた。

音楽コンテンツを自動生成する研究は、少なくとも工学的意義、科学的意義、芸術的意義の3つの意義を持つ。工学的な観点から見ると、自動生成技術が実現できると、人々の音楽制作を支援するために有用である。たとえば、作曲できない素人でも、手軽にオリジナル楽曲を作る楽しみを得ることができ、アマチュア作曲家は、データ分析結果や専門技能が実装された技術によって、自分一人では完成できないような作品を制作できるようになる。さらにプロの作曲家は、自動生成技術に有能な助手として支援されながら、創造的な作業に没頭できるようになる。一方、科学的な観点からは、人間の創作過程をコンピュータで分析・再構成することで、名作が創られる秘訣や創造性の本質を明らかにできる可能性があるという点で意義深い。芸術的観点からは、コンピ

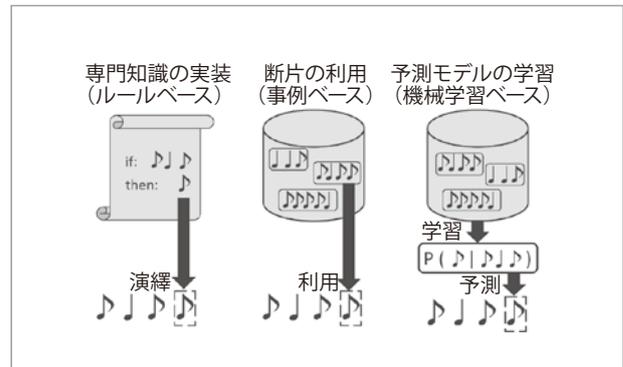


図-1 情報処理の発展に伴い登場した音楽コンテンツを自動生成する3つの方法

ュータを用いて初めて可能となる音楽表現を追求することで、これまでになかった表現が生みだせるという意義がある。本稿では、1つ目の工学的意義を中心に、音楽コンテンツの自動生成について議論する。

音楽コンテンツ自動生成の変遷

音楽コンテンツの自動生成に関して、実際にどのような研究がなされてきたのかを紹介する。体系的な既存研究のまとめはサーベイ論文⁵⁾に譲り、以下では情報処理の発展とともに現れた自動生成の3つの方法（図-1）に即して、代表的な位置づけの研究を紹介する。まず古くから研究されてきた、楽譜情報・演奏・歌詞といった記号系列を生成する研究について述べ、次に、近年工学的観点から取り組み始めた音楽音響信号を生成する研究について述べる。

♪ 楽譜情報・演奏・歌詞の自動生成

作曲には和声法や対位法といった、ある時代の音楽文化から抽出したセオリー集のような技法・理論がある。それらの一部を人間がルールや制約等の形で記述しコンピュータに実装することで、楽譜情報（旋律等の音符列）を自動生成する取り組み（例：Automatic Counterpoint）が初期には行われた。その後、情報処理全般におけるデータ処理の重要性が認識され始めると、既存作曲家のスタイルを模倣した断片を組み合わせる（繋ぎ合わせる）ことによって楽譜情報を生成する方法（例：Experiments in Musical Intelligence）が取り組まれた。確率モデルなどを用いた機械学習手法が発展してくると、楽譜データベースから学習した音高のマルコフ連鎖を用いて旋律・伴奏を生成する方法（例：Continuator）が研究された。

こうした楽譜情報を生成する研究とは別に、その楽譜をどのように表情豊かに演奏するか、という演奏表情の自動生成が取り組まれてきた。演奏表情の生成でも、楽譜情報の生成と同様に、初期には演奏家が持つ演奏についての知識・知見を実装する研究（例：Director Musices）が多かった。その後、演奏データベースを活用し、既存の演奏表情の断片を用いて自動生成する手法（例：SaxEx）が取り組まれた。演奏データベースの規模が大きくなると、機械学習手法を用いて人間らしい演奏の仕方を学習して演奏生成する研究（例：ANN Piano）が発展した。

ほかには歌詞の自動生成も研究されてきた。特に近年は大規模な歌詞データの機械学習に基づく研究（例：LyriSys）が可能になっている。自然言語処理や音声言語情報処理では、早い時期（1970年代）から統計モデルに基づく技術が研究されており、その自然な延長として、歌詞という言語情報を生成するのに機械学習や統計的アプローチが活用されている。ただし、歌詞は単なる文章ではなく、歌唱や伴奏と密接にかかわるため、自動生成をする際にそれらとの関連性を分析して考慮したり、歌詞特有の表現を可能にしたりと、まだ取り組むべき課題は多い。

♪ 音楽音響信号の自動生成

個々の構成音を合成する楽音合成ではなく、音楽音響信号として楽曲全体を自動生成する研究は、記号系列の生成研究に比べると少なかった。これまでは主に芸術的観点から取り組まれており、たとえばグラニューカーシンセシスと呼ばれる音響信号の断片を変形・接続する音響合成を、アルゴリズムに基づいて自動的に行う方法があった。一方で、音楽音響信号の形式での音楽データベースを活用できるような工学的観点からの研究はまだほとんどない。制作支援を目的とした音楽自動生成では、既存の音楽音響信号に含まれる拍節構造や和音といった音楽的内容を分析・再利用できるとよいが、技術的に難易度が高かった。そのため、音響類似度に基づいた音響断片の接続（CataRT）や、音楽レコーディング後のミキシングを自動で行う研究など、主に音響の違いそのものに着目した自動生成研究に限定されていた。

しかし、音楽音響信号の分析技術（音楽理解技術）の発展によって、この状況は変わりつつある。たとえば複数の既存楽曲を混ぜ合わせて自然に聞こえるように音楽を制作する「マッシュアップ」を自動生成・制作支援する技術（AutoMashUpper）が研究されている。音楽データベース中のさまざまな既存楽曲の音響信号に対して拍節構造や和音等を分析し、自動的に断片を切り出して加工しながら重ね合わせて音楽を生成できる。こうしたアプローチは Creative MIR とも呼ばれ、音楽情報検索（MIR: Music Information Retrieval）の分野で盛んに研究されてきた音楽理解技術を活用した生成研究として近年注目を集めている。

今まさに音楽自動生成の転換期

このような音楽コンテンツの自動生成の変遷を踏まえると、以下の2つの点で今が転換期であるといえる。

1つ目は大規模データベースと機械学習の活用である。ほかの研究分野同様、音楽情報処理分野においても、すでにそれらを活用した研究開発は活発に

// 特集 // 音楽を軸に拡がる情報科学

取り組まれている。これまでは音楽の分析・理解に関する進展が大きかったが、今後は、音楽の生成に関しても大きな進展が期待される。ほかの研究分野でも、大規模データベースと機械学習をコンテンツ生成へと応用する研究はまだ未成熟であり、今後の音楽情報処理分野での自動生成研究がパイオニア的な存在となって、学術的な発展を牽引していきたいと考えている。

2つ目は音楽音響信号の分析技術（音楽理解技術）の活用である。そうした技術が、機械学習の学習データとして音楽音響信号を活用する自動生成研究を進展させる鍵を握っている。それにより、過去には困難だった大規模なデータベースに基づく音楽自動生成が可能になるからである。そうした流れと並行して、音楽理解技術を活用することで音楽に連動するコンテンツを自動生成する研究も進んでいる。たとえば、インターネット上の楽曲の中身を自動解析する音楽鑑賞サービス Songle^{☆1}の外部連携機能を活用することで、音楽に合わせたダンス動画や写真のスライドショーの自動生成、TextAlive^{☆2}のような音楽に同期した歌詞アニメーションの自動生成・制作支援が実現されてきた。

以上のような研究の流れは、今後より一層発展していくことが予想される。

自動生成が制作を支援する未来へ

最後に、自動生成技術が広く制作支援に使われる未来に向けての挑戦を2つ議論したい。

1つ目が制作支援時の簡便さと自由な編集のトレードオフへの挑戦である。操作を究極に簡便にしたければ「全自動生成ボタン」を1つ用意すればよいが、それでは支援技術を使う人の裁量はなくなり、自分の意図を反映できなくなってしまう。一方で裁量を増やして自由に編集可能にすると、操作が煩雑となり、支援技術を使うための知識が必要となるこ

とが多い。このトレードオフは人間が道具を使う際の根源的な問題でもあり、ヒューマンコンピュータインタラクションの観点からの研究開発と合わせて取り組んでいく必要がある。

2つ目は自動生成結果の品質と多様性を両立させる挑戦である。芸術的意義に基づく自動生成では、作曲家が求める音楽を1種類だけ生成できれば、それで目的を達成できた。一方、制作支援といった工学的意義を持つ自動生成のためには、生成できる曲をできるだけ多様にしたい。しかし、単にランダム性を加えて多様にするのでは、生成する音楽の品質が保てない。これは多様な音楽を含む大規模なデータベースを活用したり、それを適切に抽象化・具象化できる技術を実現したりすることで解決できると考えている。

いままさに転換期にある音楽コンテンツの自動生成は、音楽情報処理・ヒューマンコンピュータインタラクション・機械学習等を一層融合していくことで今後さらに進展でき、人々の音楽制作を支える基盤技術として不可欠になっていくことを確信している。

参考文献

- 1) Hiller, L. and Isaacson, L. : Musical Composition with a High-speed Digital Computer, Journal of the Audio Engineering Society, Vol.6, No.3, pp.154-160 (1958).
- 2) Sundberg, J., Askenfelt, A. and Fryden, L. : Musical Performance, A Synthesis-by-Rule Approach, Computer Music Journal, Vol.7, pp.37-43 (1983).
- 3) Cope, D. : Machine Models of Music, MIR Press, pp.403-425 (1992).
- 4) Pachet, F. and Roy, P. : Markov Constraints : Steerable Generation of Markov Sequences, Constraints, Vol.16, No.2, pp.148-172 (2011).
- 5) 松原正樹, 深山 覚, 奥村健太, 寺村佳子, 大村英史, 橋田光代, 北原鉄朗 : 創作過程の分類に基づく自動音楽生成研究のサーベイ, コンピュータソフトウェア, Vol.30, No.1, pp.101-118 (2013).

(2016年4月2日受付)

深山 覚 (正会員) s.fukayama@aist.go.jp

2013年東大大学院博士後期課程修了。博士(情報理工学)。同年より産業技術総合研究所 情報技術研究部門 研究員。専門は自動作曲、音楽情報処理、機械学習。2009年本会山下記念研究賞等受賞。

後藤真孝 (正会員) m.goto@aist.go.jp

1998年早大大学院博士後期課程修了。博士(工学)。現在、産業技術総合研究所 情報技術研究部門 首席研究員。音楽情報処理を24年間研究。日本学士院学術奨励賞, 日本学術振興会賞等43件受賞。

☆1 <http://songle.jp>
☆2 <http://textalive.jp>