

創作ノート

音楽的 AI アシスタントモデルの開発と それに係るプロンプトエンジニアリングの検討

片桐 豪生
Gouki Katagiri
国立音楽大学
Kunitachi College of Music

松田 周
Shu Matsuda
国立音楽大学
Kunitachi College of Music

概要

本稿は、音楽制作における生成 AI の新しい活用モデルを提案することを目的に、ChatGPT (gpt-4o-mini) と連携する新たな音響システムを開発し、著者独自の「イメージ具現化型」と「創造性増幅型」という二つの AI 活用観を基盤とするプロンプトエンジニアリングを検証した創作ノートである。

著者は Ableton Live と Max for Live(M4L), Node.js スクリプトを用いて、ChatGPT と連携した3つのパッチ (MIDI ノートを自動生成する「MidiGenerator」、空間的響きを指示語から算出する「Reverb」、対話的なパンニング制御を担う「Panner」) を作成し、それぞれがユーザーの入力やコンテキストをどのように解釈・反映するかを観察した。また、これらの試作を通じ、作曲作業や音響調整の場面で生成 AI が果たせる役割を検討し、創作におけるアシスタントとしての AI 活用を考察している。本稿は、そのシステム構築とアプローチの全体像を示し、更なる発展に向けた基盤として位置付けられる。

1. はじめに

1.1. 『音楽』と『技術』

20 世紀より、電子技術の発達と共に、コンピュータやシンセサイザ、最近ではスマートフォンやタブレットを使った音楽的創作の発展と多様化にはとどまるところがない。もはや初期の電子音楽のように、一部の作曲家がテクノロジーを駆使した「前衛音楽」を作るだけでなく、ミュージシャンたちはコンサートや制作に最新のデジタル技術を活用し、一般の人々もデスクトップ・ミュージック (Desktop Music/DTM) を気軽に楽しむことが日常的な光景になっている。

果たして、こうした状況になると必ず出てくるのが「機械には本当の音楽は作れない」とか「音楽は技術

ではなく精神だ」といったテクノロジー不信の見解であるが、19 世紀までの音楽がテクノロジーと無関係に行われ、20 世紀以降の音楽がそれらと結びつくようになったと考えるのは大きな誤りである。

確かに、コンピュータと関わるような音楽は 20 世紀以降のものであるが、いつの時代でもテクノロジーが音楽文化を牽引する要素の一つであったことは確かだ。19 世紀に限っても、様々な楽器技術や機械技術が音楽文化に大きな影響を及ぼした。メトロノームやピアノなどは各時代ごとの先端技術の粋であるし、その他の諸楽器も時代を追うごとに積極的に発明、改良され、作曲家たちはそれらに対応する新たな書法を編み出し続けてきたのである。

英語で「芸術」を指す“art”は元々、古フランス語やラテン語で「技術」を指す語から派生し、さらにその語源を辿ると「技」を意味するギリシャ語の“technē”に行き着く。これに象徴されるように、音楽を含めた芸術活動は、決して純粋な精神活動などではなく、常にテクニックやテクノロジーといった要素と結びついてきたと言える。

これを受けて本稿でも、開発するシステムに関して、その音楽との結びつき方を常に明確にし、相互にどのような具体的効果をもたらすかを提示できるよう心がける。

1.2. 制作の目的

本稿で取り組むのは、音楽における AI 活用モデルの開発とそれに係るプロンプトエンジニアリングの考察である。これは、著者の長期の研究計画である音楽制作における AI アシスタントモデル開発の初期段階に位置付けられ、音楽と生成 AI との関わりを考えるうえで重要であるといえる。

本稿で開発するのは三つの新しい音響システムで、そ

れぞれ Max for Live (以後, 「M4L」) のパッチとして, Node.script を利用して ChatGPT と連携する形で整備するものである。

著者はこのシステムを, 音楽制作における AI 活用によるどのような可能性があるのかを探る糸口にしたいと考えている。

長期の AI アシスタントモデル開発について著者は, ユーザーの具体的な要求を高精度で実現し制作プロセスを効率化する『イメージ具現化型』といった AI 活用モデルと, ユーザーに新たな視点やアイデアを提供し創作の可能性を広げる『創造性増幅型』とも言うべき AI 活用モデルを区別し再統合することを目指している。これについては後ほどまとめるが, ともかく本稿は, 音楽と AI について考えるスタートラインとして重要である。

また, どのシステムも, ユーザのプロンプトを受けた ChatGPT がカスタム指示に則っていくつかの数値を出力し, それパラメータ操作に使われるという構造を基本としている。M4L のパッチである以上, 即座に Ableton Live 上で動作させることが可能であり, より多くの音響システムと ChatGPT の連携を進め, それらを使用したライブ・エレクトロニクス作品を制作, 実演することも視野に入れている。

2. 先行事例

特にコンピュータや数学的アルゴリズムに基づく作曲に関わる事例を 3 つ挙げ, その概要と歴史的音楽的意義について, 本稿の目指す創作と関連させつつ分析する。

2.1. コンピュータによるアルゴリズム作曲の先行例

イリアック組曲は, 1950 年代にレジャレン・ヒラーとレナード・アイザクソンがコンピュータを用いて弦楽四重奏の楽曲を作成した先駆的試みである。厳密な作曲理論をアルゴリズム化し, 当時のコンピュータ演算を通じて形式的な音楽生成を実行した点に特徴があるが, ユーザーの意図をインタラクティブに反映する要素は十分に検討されなかった。

2.2. 機械学習と自動生成による作曲の先行例

EMI (Experiments in Musical Intelligence) は, デイビッド・コープが特定作曲家の作品を学習し, その作風に近い新たな曲を生成する仕組みとして開発したシステムである。機械学習ベースで音楽スタイルをモデリングし, 既存の枠組みを突破する創造性をどのように担保できるかを議論した点が注目に値する。

2.3. システムによる多様な音操作の先行例

RjDj は, スマートフォン上で環境音を取り込み, リアルタイムにアルゴリズム処理を施すことで新しい音楽体験を生み出したアプリケーションである。ユーザーが画面操作よりも周囲の音響状況に働きかける構造が特徴的であり, 意図しない響きを楽しむ「創造性増幅型」の典型といえる。

2.4. 先行例と本稿の関連

本稿では, イリアック組曲が示した「作曲理論をコンピュータに実装する」意義を踏まえつつ, リアルタイム制御とユーザー中心のプロンプトエンジニアリングを組み合わせることで, 従来の自動作曲の可能性をより広げることを目指している。また, EMI については, 学習と自動生成を組み合わせるアイデアに共鳴しつつも, ユーザーが細かな要望を直接 AI に指示し, 部分的にコントロールを取り戻す余地を残すことで, 模倣の域を超えた音楽的成果を得ることができると考えている。RjDj についても, ユーザーが積極的に指示をしなくてもシステムが多様な結果を提示する仕組みを参考にしつつ, 本稿では必要に応じて具体的なイメージを与える「イメージ具現化型」の要素も共存させることで, 生成 AI がもたらす意外性とユーザーの作家性を両立させる方法を探っている。

3. 開発したシステム

概要の通り, 本稿では Max for Live と ChatGPT を利用した 3 つの音響システムの開発を行なった。事実上ブラックボックスである生成 AI をどのように音楽制作で活用するか検討するため, 具体的には以下の 3 つを開発, それに係るプロンプトエンジニアリングを考察した。

- (1) MidiGenerator
- (2) Reverb
- (3) Panner

(1) は, 特に生成 AI がユーザーのプロンプトをどのように音楽的要素に変換するかを調査するため, 既存の生成系のサウンドシステムの中でもポピュラーなものを選択しようとしたため採用に至った。本稿でも特にこの MidiGenerator について, どのようなプロンプトを用いると生成結果により「納得できる」か, 後の項で検証している。また (2) は, 一般的なオーディオエフェクターの動作を生成 AI に任せてみようという視点から, 特によく使用するものとして選択した。詳しくは後述するが, 特定の状況 (e.g. 浴室, オペラ座, 体育館) を示すプロンプトを入力した際, それを生成 AI がどのようにパラメータに変換し表現するのか, プロンプトの工夫によってそれがどのように変化するのか, 調査する。

(3) は, (1) (2) とは若干異なり, 音そのものではなくその聞こえ方のコントロールに生成 AI を挟んでみようという視点から選択した. 実際のスピーカーステムや, 出力されたアナログな音を観客が聴取する構造を生成 AI がどのように理解しパラメータを出力するのかを調査し, 非常に拡張性の高い考察に至った. 今回はステレオ出力に対応するパッチを利用したが, マルチチャンネル出力に対応させた場合についても継続して検討したい.

3.1. MidiGenerator

GKgptMidiGenerator と題したこのシステムは, これまで試みられてきた MIDI のランダム生成とは異なり, (プロンプティングによってある程度) 意図のあるパラメータ設定を可能にする. ChatGPT が生成するパラメータは以下の 8 つである.

- note generator on/off
- note C0 - C8 velocity: 0 - 127
- note_length: 0 - 80000
- note_generate_rate: 0.0 - 127.0
- note_generate_probability: 0 - 100
- random_time_on: 0 - 1
- random_time_rate: 0 - 3000
- random_time_amount: 0 - 3000

これらの数値を, textedit オブジェクトから送られたプロンプトと JavaScript 内のカスタム指示との組み合わせから得る. 以下が, パッチの内容と, 初期設定のカスタム指示である.

以下は MIDI ジェネレータのパラメータです. 入力される文章のイメージを元に, 以下のそれぞれのパラメータの値を設定してください.

また出力は『0 0 0 0 0 0』のようにスペース区切りで値のみを表示してください.

- note C0 - C8 velocity: 0 - 127
- note_length: 0 - 80000
- note_generate_rate: 0.0 - 127.0
- note_generate_probability: 0 - 100
- random_time_on: 0 - 1
- random_time_rate: 0 - 3000
- random_time_amount: 0 - 3000

このようにして, 入力したプロンプトに応じて, 単旋律の MIDI ノートが生成され続ける状態を実現した. こ

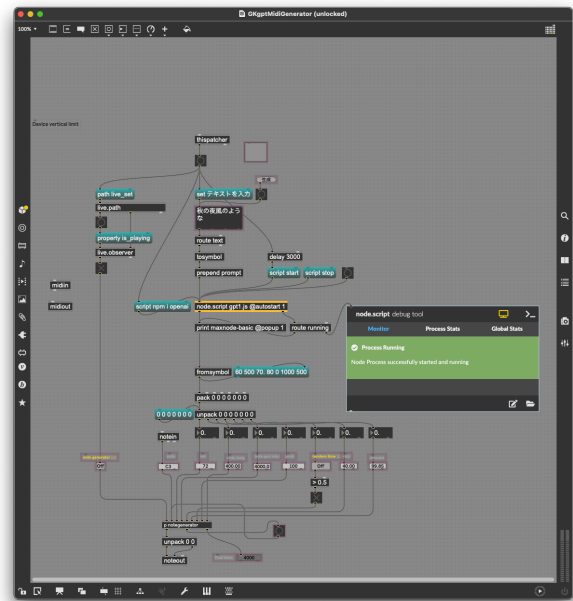


図 1: GKgptMidiGenerator

れを Ableton Live 上で Max for Live パッチとして MIDIトラックで開き, 任意のサウンドを挿入すると音が鳴るという構造である.

ここで注意しなければならないのは, 生成される MIDI のテンポが Ableton Live のプロジェクトのテンポ情報を参照していない点と, どの楽器で MIDI ノートを鳴らすかも, パッチ側では想定していない点である. よって今回は, テンポ情報はプロンプトの入力前にそれに合った数値をこちらで設定し, また, 必要に応じてプロンプト内で使用する楽器を指定する形式でテストを行った.

以上を踏まえ実際に使ってみると, 例えばテンポを 120 にした状態で, 簡単に「疾走感」という単語を入力すると, 高音域を細かく使ったメロディックな出力が有意に長い時間得られたり, またテンポを 80bpm に設定し「maestoso」と入力すると, 音価は少なく, かつ低・中音域を多く含んだ部分が多い旋律が生成された. テンポを 60bpm に設定し「クラリネットのための伸びやかに歌うような旋律」と入力すると, 果たしてそれがクラリネットによって演奏されるのが望ましい旋律なのは別として, 二分音符を多く用いた, およそクラリネットの音域を考慮したのであろう旋律が生成され続けた.

やはり, 大規模言語モデルとして, プロンプトの定義から拾える音楽的情報 (音高, 音価など) にはある程度の確からしさがあるものの, 旋律としての音楽性には楽典的ルールが乏しい結果になった.

の制御方法と手動のダイアル操作によるコントロールでパンニングを行う構造になっている。ChatGPT に生成させるパラメータは以下の通りである。

・ Pan: -50 - +50
 ・ Mode: 0 - 2

ChatGPT の生成結果を安定させるため、またパッチの構造を考えても、制御モードをその名称ではなく数値で指定しなければならない点がカスタム指示の難点であった。何度かの調整の末、以下のようにまとめている。

###パラメータ はステレオ・パンニングのパラメータです。
 これから入力される文章のイメージを元に以下のそれぞれのパラメータの値を設定してください。また出力は『-12 1』のようにスペース区切りで値のみを表示してください。ただし、Mode パラメータの数値は ###モード一覧 のように、パンニングの制御方式と相関しています。

###パラメータ
 Pan: -50 - +50
 Mode: 0 - 2

###モード一覧
 0 = Linear Panning
 1 = Equal Power Panning
 2 = Cosine Panning

実際に使用してみると、パッチのシンプルさもあり、具体的な指示を出すプロンプトでは、それに忠実なパラメータが出力される傾向にあった。

定位に関する指示はほぼ正確で、「左手前から」「真右から」といった詳しい指示は特に、また「ユーザーが真北を向いているとして、東から音が聞こえるように」という異なる尺度の方向指示でもかなりの精度でパンニングできていた。「明後日の方向から」や「死角から」といった抽象的な指示に関しては、複数回生成させるとバラバラの値が生成され、具体的な指示とは異なる活用方法につながるかもしれない。

Pannerについては、Max for Liveにある Surround Panner がより拡張的にリリースされており、今後こちらも使ったより複雑で豊かなパン操作を検討したい。

4. プロンプトエンジニアリング

開発した3つのシステムの特徴であり欠点とも呼べるのは、この MIDI 生成が最終的には ChatGPT によって行われ、生成物のコントロールが難しい点である。これをポジティブに捉えれば、今までに存在しなかったルールに基づく MIDI 生成として新しい表現の可能性が見つかったと言えるかもしれない。

この項では、特に MidiGenerator について、JavaScript で設定しているカスタム指示をプロンプトエンジニアリングの観点から見直すことによって、「音楽的により納得感のある生成結果」を目指す。今後の開発においても、生成 AI の活用には詳しくないユーザーでも気軽に活用できるようなプロンプトエンジニアリングが十分に考慮されなければならない、その足がかりとして、既存のプロンプティング技法についてまとめた後、実際に MidiGenerator の生成結果を向上させるためのプロンプトエンジニアリングを検討する。

4.1. MidiGenerator の生成結果をコントロールする

MidiGenerator の生成する MIDI ノートの“納得感”を向上させ、ユーザーのイメージを再現するシステムにすることを目的とするためのプロンプトの工夫に取り組む。

先述の通り、開発したシステムでは、ユーザーが Node.script に入力するプロンプトとは別に Node.js 内で「カスタム指示」という形でどのようにパラメータを出力するかを指定している。このカスタム指示をプロンプトエンジニアリングの観点から見直すことによって「音楽的により納得感のある生成結果」を目指す。第1段階のプレーンなプロンプトに対し段階的に変更を加え、生成結果を比較評価していく。

全ての段階に共通して入力するプロンプトは、楽器の指定と抽象的な指示を共存させることを目的に「グランドピアノで表現する、夏の終わりを感じさせる風のような旋律」とした。以下が段階的な変更とその結果・考察である。

4.1.1. Role Prompting

AI に役割を与える方法を利用し、カスタム指示に「あなたは音響のプロであり、優秀な AI アシスタントです」という文言を追加した。

結果、明言できるほどではないが、プレーンのカスタム指示には乏しかった、旋律の音楽的豊かさが向上したように思われた。音程の不自然な跳躍が減少し、プロンプトで指定したピアノによる演奏を考慮したためか、大まかに細かい音価の連続する部分と長い音価が続く部

分が交互にやってくるような旋律が生成された。しかし「夏の終わりを感じさせる風」はまだ吹いていない。

4.1.2. 言語の変更

ChatGPTをはじめとする大規模言語モデル共通の特徴として、事前学習の量の差から日本語よりも英語の方が生成物のクオリティが高いとされるため、カスタム指示全体を英語で表記した。しかしここでは明確な差は受け取れなかった。より大規模で具体的なプロンプトの改善が必要だと考えられる。

4.1.3. Few-shot Prompting

プロンプトの例とそれに対する数値出力の例を2つ用意し、生成例というラベルで追加し、タスク進行中に参考にするよう指示した。

結果、極端な生成例を選択したためか、プレーンなカスタム指示で生成させていた結果よりもむしろバラバラとした印象を受ける生成結果になってしまった。

例として具体的な数値を追加学習させても、プロンプトからどのような思考プロセスでその数値を出力するに至ったかが情報として付与されていないがために、単なるノイズとして働いてしまったと考えられる。

4.1.4. Few-shot / Chain of Thought

入力に対しどのような情報を拾い数値に反映したかを説明する工程を挟む構造になっている生成例を用意し、生成例というラベルで追加し、タスク進行中に参考にするよう指示した。

しかし、指示が長大になってしまったためか、7つの数値のみを出力する指示を忘れ途中の思考回路を長文で出力してしまった。

パラメータの列挙と生成例、生成に関する指示をそれぞれラベルとして設定し直し、図のようにした。結果問題なくパラメータ数値が出力されるようになったものの、何度生成させてもほとんど同じような値を返すようになり、旋律感は失われてしまった。

パラメータがどのように設定されているのかを確認できないパッチ構造になっているため詳しい原因は不明である。

しかし、何ら音楽的なチューニングをせず、テキストによる指示のみでメロディを生成させたことは重要である。

4.2. 考察

この項では開発したシステムとプロンプトエンジニアリングの考察を通して、音に関わる AI がどのような挙動をするか検証した。

簡潔に結論を言えば、検証結果から得られた情報は、生成結果をコントロールするという目的を十分に達成したとは言えない量である。使用する AI の検討やプロンプトの定型構築はまだ不十分で、日進月歩の AI 産業に追いつくにはより多くの情報収集が必要である。研究の範囲内では、理想的な生成結果に持っていくことは叶わなかった。

しかし、プロンプトの工夫次第で生成結果が有意にコントロールできること、著者が指示した以上の生成結果がもたらされることが確認できた点で、『イメージ具現化型 AI』と『創造性増幅型』の AI 活用が夢物語でないことは確認できたと言えるのではないだろうか。

また、著者の直近の達成目標は、先にも述べたライブ・エレクトロニクス作品の制作である。M4L のパッチを今後どんどん増設していくために、この研究でシステムの基本的な設計が確立できたことは有意義であった。

今後も新しい生成 AI やプロンプト技法が開発されるはずであるから、本項で得られた情報をヒントにしながら、今後も検証を続ける所存だ。

5. 結論

5.1. 総括

開発した3つのシステムは、その生成物のクオリティを議論する以前に、まず生成 AI を音楽的に活用した新しいモデルが実現した点で有意義であった。テキストの事前学習がどの程度音楽的に実用可能なレベルにあるのか十分に検討でき、今後の研究に活用できるような情報が収集できたと考えている。

5.2. システムの意義と展望

本稿で開発した3つの AI 活用システムは、今後、複数のオーディオエフェクターやその他音響システムに応用される。

どれも、人間と生成 AI の共存モデルとして新しい関わり方を示すものになるし、また音楽制作における AI 活用の新しいパラダイムになるであろう。

また著者は、AI アシスタントが完成した暁には、それを学校教育向けにチューニングし、全国で音楽の授業のために活用されることを目指している。

現在の小学校学習指導要領 [22] では、音楽科での学

習領域として「鑑賞」と「表現」が設定されているが、表現の中にある「歌唱」と「器楽」は多く触れられど、「音楽づくり」の充足は難しいのが現状である。学年が進み本格的な作曲を学ぶ内容が追加される頃には音楽科は必修科目から外れ、多くの子どもたちが歌唱と演奏、鑑賞を修めて音楽の授業を卒業するのだ。

しかし作曲という営みは、高度な言語化と他者との対話、深い内省を要する、文化的に希少な行為である。現在、限られた人間が作曲を学び、挑戦し、その中でもごく一部が仕事としている。もちろん職業作曲家はどのような状況になっても尊重されるべきであるが、小学校で英語やプログラミングが必修化したように、作曲も当たり前のように学ぶようになり、全国民が作曲を勉強したことがある社会が実現したら、それはどれだけ文化的に豊かであろうか。

私はそんな新しい世界を、AI アシスタントと共に実現したい。

6. 参考文献

大谷 泰斗. 2024. Max for Live で解放する ChatGPT との音楽制作.

fendoap. 2023.max/msp でリバーブを作る__シンプル版.

Lejaren A. Hiller, Leonard M. Isaacson. 1959. Experimental Music: Composition with an Electronic Computer.

David Cope. 1996. Experiments in Musical Intelligence.

徳井直生, 永野哲久, 金子智太郎. 2009. iPhone * Music iPhone が予言する「いつか音楽と呼ばれるもの」.



この作品は、クリエイティブ・コモンズの表示 - 非営利 - 改変禁止 4.0 国際 ライセンスで提供されています。ライセンスの写しをご覧になるには、<http://creativecommons.org/licenses/by-nc-nd/4.0/> をご覧ください。Creative Commons, PO Box 1866, Mountain View, CA 94042, USA までお手紙をお送りください。