

創作ノート

AIとのリアルタイム共作によるパフォーマンス作品 【improvisation with metaball and AI】 "improvisation with metaball and AI" for real-time AI

大久保 拓太, 森谷 友昭

Takuto OKUBO, Tomoaki MORIYA

東京電機大学大学院

Tokyo Denki University Graduate School

概要

本稿は、大規模言語モデル (LLM) を核とした自律演奏システム「C.R.E.A.M. (Compositional Real-time Engine for Augmented Musicality)」の構築と、それをを用いたパフォーマンス作品【improvisation with metaball and AI】の創作とシステムについての創作ノートである。メタボールとは、それ同士が近づくと滑らかに融合する性質を持った有機的な球体である。メタボールは Blinn により blob という名称で提案された (Blinn 1982)。その後アルゴリズムの改良などを行った西村らによりメタ球 (メタボール) という名称が用いられるようになった (西村 et al. 1985)。本作では、ユーザの演奏をリアルタイム解析し、音楽的知識に基づいてユーザに音として提案をする形でフィードバックする相互な即興演奏システムとメタボールを使用した映像生成システムを作成した。本論文では、前回に加えて新たに構築された即興演奏システムについて触れていく。

1. 背景

本作"improvisation with metaball and AI"は、AI とのリアルタイムな音楽の即興演奏に着目したインタラクティブな作品である。近年、アート業界における AI、特に sora(Sora 2024) や stable diffusion(R.Rombach et al. 2022) などの生成 AI の台頭は著しい。急速なスピードで発展、性能向上を続けており、様々なスタイルの画像や写真を圧倒的なスピードで提供してくれる。そのような時代になった中で、我々は今一度、AI が台頭している中で人間によって生み出されるアートの価値を考え直す必要がある。具体的には、AI がいる中で改めてどんなアートを我々人間が作ることに価値があるのかというのを考えなければならない。先述した通り、生成 AI の出力する画像や写真は、とても品質としては優

れたものになっている (Jules van Hees et al. 2025) が、なかなか人間がその成果を肯定しない傾向が強い。特に、AI によって制作されたという事実がこの結果を援助している。よって現状、生成 AI の成果とは裏腹に、なかなかアート業界における AI の立ち位置は否定的で、人間と対極に置かれている。AI の台頭を阻止できない中で、改めて大事なのが我々が AI を嫌い、差別化することではなく、AI とうまく付き合っていくこと、つまり“共生”である (栗原聡 2025)。AI と人間が対等な立場で共同でアートを作っていくことに、新しいアートの価値があるのではないかと模索していくために今回のシステムを提案した。よって、今回のシステムである AI と人間によるリアルタイムな音楽の即興演奏パフォーマンスを通して、AI は既に人間によって区別される立場ではなく、対等な立場まで来ているということを提案していきたい。

2. 作品概要

本作品は、ローカル LLM (Large Language Model) を音楽的対話の主体として位置づけた、リアルタイム・セッション・システムである。今回これを「C.R.E.A.M. (Compositional Real-time Engine for Augmented Musicality)」と名付けた。本作品では AI を独自の意志を持つ「共演者 (Agent)」として再定義している。本システムにおいて、ユーザーは二つの側面から音楽に介入する。

- 演奏者としての介入: MIDI キーボードを用いた身体的な即興演奏。
- 指揮者としての介入: チャット UI を通じた、自然言語による音楽的ディレクション (例: 「もっと激しく」「静かなアルペジオで」など)。

AIはこの両者を文脈として受け取り、自律的にフレーズを生成し続ける。この互いのフィードバックによって、人間とAIの境界を曖昧にする新しい即興演奏の形を提示する。

3. システム構成

本システムの技術的基盤は、Pythonをハブとしたモジュール結合型のアーキテクチャである。

3.1. 思考エンジン：ローカル LLM

AIの脳内エンジンには、LM Studioを介して動作するLiquid LFM 1.2Bを採用した。このモデルは、線形再帰的な構造（Linear Recurrent Units）を持つため、長い文脈（演奏履歴）を保持しつつ、エッジ環境での低遅延な推論が可能である。言語モデルの選定基準は、応答の軽さと、レスポンスの速さの2つのバランスが取れていることである。

3.2. 聴覚系：演奏監視とコンテキスト化

ユーザーの演奏データ（MIDI In）は、非同期スレッドによって常時監視され、適宜Pythonを用いて受け取られる。

- 履歴管理: 直近10秒間の打鍵データ（音名、ベロシティ、タイムスタンプ）を受け取る。
- プロンプト化: 演奏履歴を「C4を力強いタッチで演奏」といった自然言語に変換し、ユーザーからの指示と統合してシステムプロンプトを構築する。

3.3. 実行系：自律セッション・ループ

生成プロセスは、数秒間隔で自動的に実行される。

- 非同期生成: メインのAPIサーバーとは別のスレッドでLLMへのリクエストを行い、演奏の継続性を担保する。
- ロバストなMIDI抽出（救済ロジック）: LLMが生成するJSONデータが非構造化される課題に対し、正規表現を用いた独自の抽出アルゴリズムを実装した。AIの出力が不完全な形式であっても、文字列内から音名とタイミングを強引に抽出・補完し、辞書型データへと再構築する。

3.4. 出力系：マルチモーダル・フィードバック

抽出された音楽データは、以下の二経路でリアルタイムに実行される。

- MIDI Output: IACドライバ（仮想MIDIバス）を通じ、Ableton Live(Ableton 2026)上のインストールメントを直接演奏する。
- OSC Output: 音楽のエネルギー量（Energy）や緊張感（Dissonance）などのパラメータをUDP通信でTouchDesigner(TouchDesigner)へ送信し、視覚演出を同期させる。

4. 本システムの構成

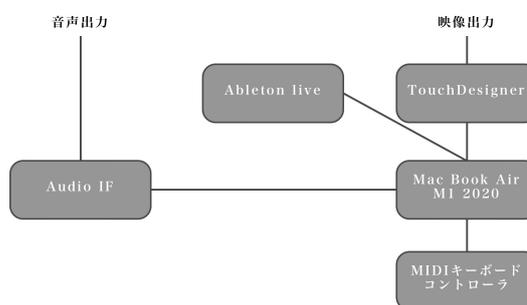


図1: 本システムの構成図

5. 映像生成システム

本システムでは、互いの演奏情報を映像側へも送信し、リアルタイムに映像も生成している。具体的には、AIが生成するparams（BPM, Energy, Dissonance等）をOSC（Open Sound Control）経由で受信し、映像の挙動にダイレクトに反映させる。

6. 考察

本システムの制作と使用を通して、様々な課題と可能性（面白さ）を発見できた。課題については、ローカルで動くLLMを条件にした場合、なかなかシステムプロンプトを細かく設定しても思い通りのレスポンスを返してくれないことが多々あった。ここはLLM次第とはなってしまうのだが、精度という点においてまだまだ発展と応用が見込める点であった。その他、課題としてはAI側で楽器を選ばないことにある。現状のシステムでは、あらかじめこちらが選定した楽器のみを鳴らすまでにとどまっている。これはソフトウェアの制約であるのかもしれない。課題が発見された反

面, AI との即興演奏自体はとても面白いものであった。こちらが弾いたものに対して, 返答するように音を鳴らし返してくれるという AI の行動に, とても人間味を感じた。今回は, システム全体を通してとても実験的な構成になってしまったが, 今後システム構成をもっと綺麗にまとめることと, AI の知識・提案をさらに豊富なものにしていくことでさらに高度なパフォーマンスが可能になると思われる。同時に, TouchDesigner 側への映像提案も可能性がまだまだ残っているので, 今後も発展を試みていきたい。



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/> をご覧頂くか, Creative Commons, PO Box 1866, Mountain View, CA 94042, USA までお手紙をお送りください。

7. 参考文献

- Blinn, James F. A Generalization of Algebraic Surface Drawing. ACM Transactions on Graphics. 1982, 1(3), pp. 235-256.
- 西村 仁志, 平井 誠, 河合 利幸, 河田 亨, 白川 功, 大村 皓一. 分布関数による物体モデリングと画像生成の一手法. 電子通信学会論文誌. D / 電子通信学会 編 68 (4), p718-725, 1985-04.
- OpenAI. (2024). Sora:Creating video from text. openai.com/sora, 2024.
- R.Rombach et al.: High-Resolution Image Synthesis with Latent Diffusion Models, Proc. CVPR, 2022.
- Jules van Hees, Tijn Grootswagers, Genevieve L. Quek, Manuel Varlet. Human perception of art in the age of artificial intelligence . Frontiers in Psychology . 2025 , Vol.15.
- 栗原聡. AI 脅威論の正体と人と AI との共生. 情報通信政策研究. 第 4 巻, 第 2 号, I-45-I-54.
- Ableton. (2026). Ableton Live (Version 12) [Computer software].
- Derivative. TouchDesigner. <https://derivative.ca/>(accessed 2026/02/02).

8. 著者プロフィール

大久保 拓太 (Takuto OKUBO)

2001 年生まれ。東京電機大学大学院にて, アートと AI の可能性について研究を続けながら, 自身でゲームやデジタルアート, ビッグバンドでの音楽活動など様々なクリエイティブに挑戦している。