

ジャムセッションにおける合いの手の挿入に向けた 旋律生成の検討

名越 崇晃^{1,a)} 北原 鉄朗¹

概要: 本研究の最終目標は、ジャムセッションにおける即興演奏の一環として、奏者の旋律に対する合いの手を自動生成する AI システムで実現することである。現代の音楽生成技術において、Transformer モデルはその強力な性能で注目を集めており、本研究でもこれを活用して合いの手生成システムの開発を進めている。本論文では、まず Transformer を用いたメロディの自動生成手法について説明し、次に今後の課題として合いの手を生成する AI の開発について述べる。本システムは、ミュージシャンが即興演奏中に使用することで、より人間らしく、インタラクティブな演奏体験を提供することを目指している。

1. はじめに

ジャズは、その即興演奏による高度な表現力と創造性で広く知られている音楽ジャンルである。即興演奏は、演奏者がその場でメロディやリズムを創り出し、他の演奏者との対話しながら進行するため、高度な技術と直感が求められる。近年の音楽生成技術の進展により、AI を活用した自動作曲や即興演奏の支援 [1-7] が注目されている。

特に、Transformer [8] モデルは自然言語処理や音楽生成の分野で高い性能 [9] を示しており、メロディ生成や楽曲の続きを予測するタスクにおいて有効であることが示されている。本研究では、この Transformer モデルを活用し、奏者が演奏した旋律に対する合いの手を自動生成するシステムの開発を目指す。

本論文では、まずメロディの自動生成手法について詳細に説明し、次に今後の課題として合いの手を生成する AI の開発について述べる。最終的には、このシステムが即興演奏の新たなミュージシャンの演奏体験を向上させることを目指している。

2. 合いの手の技術的な課題

合いの手生成の技術的な課題は大きく分けて 3 個の技術的な課題を解決する必要がある。特に、合いの手のデータセットが不足している点、リアルタイムでの生成の難しさ、そして合いの手を入れるタイミングの問題が挙げられる。これらの課題を解決することで、ジャムセッションにおいて自然で効果的な合いの手を提供するシステムの実現が可

能となる。以下に、これらの技術的な課題について詳しく述べる。

2.1 データセット不足

合いの手の自動生成に関する最大の課題の一つは、データセットの不足である。合いの手だけを集めたデータセットは現在存在せず、合いの手の生成モデルを構築するためには既存の楽曲データセットから学習する必要がある。既存の楽曲データセットには合いの手が含まれていることもあるが、それらは通常、メロディや伴奏と一緒に存在しており、合いの手だけを特定して学習することは困難である。このため、モデルはメロディ全体から合いの手を予測する能力を持たなければならない。

2.2 リアルタイムで生成することの難しさ

リアルタイムでの合いの手生成も大きな課題である。音楽のリアルタイム生成において、遅延が発生することは致命的となる。ジャムセッションではテンポに合わせた合いの手の生成が必須であり、遅延は演奏全体のリズムを崩す可能性がある。そのため、合いの手をリアルタイムで生成するためには、極めて低遅延のシステムが必要である。

2.3 合いの手を入れるタイミング

合いの手を適切なタイミングで入れることもまた、技術的な課題である。前述の通り、合いの手のみを取り入れたデータセットは存在せず、合いの手を入れる最適なタイミングを学習することは容易ではない。この問題の解決策として、2つのアプローチが考えられる。1つ目は、音楽理論に基づいて最も合いの手が適している拍に合わせて生成

¹ 日本大学文理学部

^{a)} nagoshi@kthrlab.jp

を行う手法である。2つ目は、旋律の切れ目を予測し、その間に合いの手を挿入する手法である。これらのアプローチにより、合いの手のタイミングを効果的に学習し、演奏に自然に溶け込む合いの手の生成が可能になる。

3. メロディ生成の予備検討

即興演奏の AI を作成するに当たり、Transformer を用いたモデル「AyatoModel」の開発を行っている。この章では、主に AyatoModel の概要と仕組み、データの前処理、Transformer モデルの活用、生成実験の結果とその考察に焦点を当てる。現状では、AyatoModel は純粋なメロディ生成のみを行っており、2.1~2.3 で議論した課題に対する検討は行っていない。以下、AyatoModel の詳細を説明する。

3.1 概要と仕組み

まず、AyatoModel の概要と仕組みについて説明する。このモデルは、音楽データを扱うための特別な設計が施されており、トークンとして7種類のシンボルを使用する。これらのシンボルは、音楽の特徴を捉えたものであり、メロディの生成において重要な役割を果たす。

3.1.1 データセット

今回は Sai が作成した Jazz ML ready MIDI [10] というデータセットを元に使用する。このデータセットはジャズというジャンルの楽曲を MIDI に変換したものが 935 曲記録している。今回モデルに学習させるのはサクソパートのみとし、データセットから抽出する。最終的にデータセットとして扱うのは 215 曲、90300 トークンのサクソパートの旋律になる。

3.1.2 前処理されたトークンについて

現在使用している前処理データであるシンボルについて説明する。MusicTransformer [9] の研究と RSCLN_Transformer [11] の研究をもとに、以下のようなトークンの構造を提案する。

$$(x_{pitch}, x_{velocity}, x_{ti}, x_{tf}, x_{di}, x_{df}, x_{root})$$

これらのシンボルは以下の配列構造を持つ：

- x_{pitch} MIDI のフォーマットに基づいた音の高さ
- $x_{velocity}$ MIDI のフォーマットに基づいた音の強さ
- x_{ti} 前回の音の開始地点から今回の音までの時刻の差 (整数)
- x_{tf} 前回の音の開始地点から今回の音までの時刻の差 (小数点以下の値)
- x_{di} その音の長さ (整数)
- x_{df} その音の長さ (小数点以下の値)
- x_{root} MIDI のフォーマットに基づいたルート音

3.1.3 トークンの処理の流れ

メロディ生成システムにおけるトークンの処理の流れに

ついて、以下のステップで説明する。

まず、トークンは Embedding 層を通じてベクトルに変換される。Embedding 層は、シンボルを数値ベクトルに変換することで、Transformer が音楽データを扱いやすくする。この層の役割は、音高や強さなどの離散的なシンボルを連続的なベクトル空間にマッピングすることで、モデルがよりパターンを学習しやすくする。

次に、位置エンコーディング層が追加される。これは、トークン間の相対的な位置関係をモデルに考慮させるために重要である。音楽においては、音符の順序やリズムが非常に重要であり、この層を通じてモデルはそれらの情報を適切に学習することができる。具体的には、位置エンコーディング層はトークンに時間的な情報を付加し、これによりモデルはトークンの位置関係を理解することが可能となる。

位置エンコーディングが付加されたトークンは、次に Transformer 層に入力される。Transformer 層では、自己注意機構を用いて各トークンの関連性を学習し、高度なパターン認識を行う。ここで、自己回帰型モデルにするために正解データを与えず、トークンの次の状態を予測する。このアプローチにより、過去の観測値に基づいて将来の値を予測できるほか、過去のデータポイントが現在のデータポイントにどのように影響するかを直接モデル化できるため、時系列データの依存性を効果的に捉えることができる。

Transformer 層で処理されたトークンは、その後、Linear 層で線形変換される。最後に softmax 関数によりこれらのスコアが確率分布に変換される。

トークンの選定は、この確率分布に基づいて行われる。具体的には、softmax 関数によって得られた確率分布から次のトークンをサンプリングする。このサンプリングにはランダム性があり、これにより生成されるメロディに多様性が生まれる。確率分布に従うことで、モデルは最も高いスコアを持つトークンを選択する傾向があるが、一定の確率で他のトークンも選ばれるため、単調なメロディーを防ぐことができる。

4. 生成実験

この章では、メロディ生成モデルの性能を評価するために行った生成実験について述べる。実験は、現在のトークンでどの程度のクオリティでメロディを生成できるかをテストすることを目的としている。具体的には、シードとして曲の冒頭を 3 トークン分与え、その続きを 10 ~ 20 トークン生成する方法を用いた。

4.1 シードとして与える曲

今回、シードとして与えるのは「All I Do Is Dream Of You」という楽曲のサクソパートから、最初の 3 音のみをトークンとして与える。

4.2 評価方法

評価方法は2回生成実験を行い、その生成された旋律を第一著者が主観で評価する。具体的な評価基準を以下に示す：

- **リズムの自然さ**：生成された旋律のリズムが自然であるか
- **音楽的な自然さ**：主観的な評価に基づいて、生成されたメロディが音楽的に自然であるか

これらの評価を通じて、モデルの現状の性能を明らかにし、今後の改良点を明確にすることができる。生成実験の結果は、次のステップとして取り組むべき技術的な課題を示すものであり、より高度なメロディ生成システムの開発ができる。

4.3 実験結果

2回生成実験を行った結果を図1と図2に示す。以下の図ではあらかじめシードとして与えたトークンは削除し



図1 生成結果1



図2 生成結果2

4.4 考察

4.4.1 リズムの自然さ

それぞれ生成結果の生成された旋律に着目する。一部で自然なリズムが維持されていたものの、生成結果1の4音目や、生成結果2の5音目のように、ノートオフのタイミングを明らかに間違えていることがわかる。リズムの自然さを向上させるためには、拍子やテンポ情報を明示的にモデルに組み込む必要がある。

4.4.2 音楽的な自然さ

音楽的な自然さについても、改善の余地があると考えられる。例えば、生成結果1はDmなのに対し、7音目(F#)のように経過音として不適切な音の流れ方がある。生成結果2も同様に10音目(Bb)のように経過音としてふさわしくない音の使われ方をしている。このことから、音楽的な文脈において重要な音の関係性や経過音の流れが不十分であり、これが自然さを損なう一因となっていると考えられる。音楽理論に基づいたデータの強化や、より洗練されたモデルアーキテクチャの導入が求められる。

4.4.3 今後の改良点

- **データセットの拡充**：より多様で質の高い音楽データセットを使用し、モデルの学習を強化する。
- **リズム情報の強化**：拍子やテンポ情報を明示的にモデルに組み込み、リズムの自然さを向上させる。
- **音楽理論の導入**：コード進行や和音の流れなど、音楽理論に基づいたデータの強化を図る。
- **モデルアーキテクチャの改善**：Transformerモデルのアーキテクチャを改良し、長期的な文脈をより効果的に学習できるようにする。

これらの改良点を実施することで、メロディ生成モデルの性能を向上させ、より自然なメロディの生成が可能となることが期待される。

5. 今後の展望

今後の展望として、より良いメロディを生成するためのアプローチについて検討する。具体的には、以下の4つが挙げられる。

5.1 コード進行を考慮したメロディの生成

まず、コード進行を考慮したメロディの生成について考える。現在のシステムでは、メロディ生成は単に音高や強さなどのシンボルに基づいて行われている。しかし、音楽の構造をより深く理解するためには、コード進行を考慮することが重要である。コード進行は、楽曲のハーモニーとメロディの調和を保つ上で不可欠な要素であり、これをモデルに組み込むことで、より自然で一貫性のあるメロディを生成することが可能となる。

5.2 拍を考慮したメロディの生成

次に、拍を考慮したメロディの生成について述べる。音楽におけるリズムは、メロディの流れやダイナミクスを決定する重要な要素である。現在のモデルでは、拍の情報が十分に考慮されていないため、生成されたメロディがリズム的に不自然になることがある。これを改善するために、拍の情報を明示的にモデルに組み込むことで、リズムに一貫性を持たせ、より洗練されたメロディの生成が期待できる。

5.3 リアルタイムな旋律の生成

リアルタイムでの旋律生成は、即興演奏において非常に重要な要素である。リアルタイムでのメロディ生成には、いくつかの重要な課題がある。これを解決するためには、モデルの処理速度を向上させるためのアルゴリズムの最適化が必要になると考えている。

5.4 合いの手生成 AI との統合

最後に、合いの手のAIとの統合について考察する。合

いの手の生成は、メロディの生成と密接に関連しており、リアルタイムでのインタラクションが求められる。これを実現するためには、メロディ生成モデルと合いの手生成モデルを統合し、相互に情報を共有する仕組みが必要である。具体的には、メロディ生成時に合いの手の挿入ポイントを予測し、合いの手生成モデルにその情報を提供することで、シームレスなインタラクションを実現する。このアプローチにより、即興演奏における合いの手の自然さと効果を高めることができる。

これらの課題に取り組むことで、より高度で自然なメロディ生成システムの開発が可能となり、ミュージシャンにとって一層価値のあるツールとなることが期待される。

6. 終わりに

本研究では、ジャムセッションにおける即興演奏の一環として、奏者の旋律に対する合いの手を自動生成する AI システムの開発に取り組んだ。まず、メロディの自動生成手法について詳細に説明し、特に Transformer モデルを用いたアプローチを紹介した。次に、合いの手の技術的な課題として、データセットの不足、リアルタイム生成の難しさ、合いの手を入れるタイミングの問題について論じた。

メロディ生成の予備検討においては、AyatoModel の仕組みとトークンの処理の流れを詳述し、シンボルのベクトル化や位置エンコーディング、Transformer 層による処理の重要性を示した。生成実験を通じて得られた結果は、現時点ではまだ改善の余地があることを明らかにした。

今後の展望として、コード進行や拍を考慮したメロディ生成、そして合いの手の AI との統合を通じて、より自然なメロディ生成を目指すべき課題を提示した。これにより、ミュージシャンにとってより実用的でインタラクティブなツールを提供し、即興演奏の質を向上させることが期待される。

最終的に、本研究は音楽生成技術における新たな可能性を開拓し、AI による自然な音楽体験の実現に寄与するものである。今後も引き続き技術的な課題に取り組み、より高度なシステムの開発を目指していく予定である。

謝辞 本研究は、JSPS 科研費 23K24966, 24H00748 の支援を受けた。

参考文献

- [1] Parson, D. E.: Chess-based Composition and Improvisation for Non-musicians, *Proceedings of International Conference on New Interfaces for Musical Expression (NIME 2009)*, pp. 157–158 (2009).
- [2] Amiot, E., Noll, T., Andretta, M. and Agon, C.: Fourier Oracles for Computer-aided Improvisation, *Proceedings of International Computer Music Conference (ICMC 2006)*, pp. 99–103 (2006).
- [3] Buchholz, J., Lee, E., Klein, J. and Borchers, J.:

coJIVE: A System to Support Collaborative Jazz Improvisation, Technical Report AIB-2007-04, Aachen Informatik-Berichte RWTH Aachen, Department of Computer Science, <http://www.informatik.rwth-aachen.de/go/id/loj/lidx/1/file/47944> (2007).

- [4] Miyashita, H. and Nishimoto, K.: Theremoscore: A New-type Musical Score with Temperature Sensation, *Int'l Conf. New Interface for Musical Expression*, pp. 104–107 (2004).
- [5] Fels, S., Nishimoto, K. and Mase, K.: MusiKalscope: A Graphical Musical Instrument, *IEEE Multimedia*, Vol. 5, No. 3, pp. 26–35 (1998).
- [6] Pachet, F.: The Continuator: Musical Interaction With Style, *Proc. ICMC* (2002).
- [7] Keller, R. M.: Welcome to Impro-Visor: Jazz Improvisation Advisor for the Improviser, <https://www.cs.hmc.edu/~keller/jazz/improvisor/> (2008).
- [8] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. and Polosukhin, I.: Attention is all you need, *Advances in neural information processing systems*, pp. 5998–6008 (2017).
- [9] Huang, C.-Z. A., Vaswani, A., Uszkoreit, J., Shazeer, N., Simon, I., Hawthorne, C., Dai, A. M., Hoffman, M. D. and Eck, D.: Music transformer: Generating music with long-term structure, *arXiv preprint arXiv:1809.04281* (2018).
- [10] Sai: Jazz ML ready, <https://www.kaggle.com/datasets/saikayala/jazz-ml-ready-midi/data>.
- [11] Deng, X., Chen, S., Chen, Y. and Xu, J.: An automatic music generation method based on RSCLN_Transformer network, *Multimedia Systems*, Vol. 27, No. 4, pp. 255–264 (2021).