

# 作曲支援のための即興歌唱からのメロディ具象化

松林 諒<sup>1</sup>吉井 和佳<sup>2,3</sup><sup>1</sup>京都大学 工学部情報学科<sup>2</sup>京都大学 大学院情報学研究所 知能情報学専攻<sup>3</sup>JST さきがけ

## 1. はじめに

音楽の楽しみ方としては、鑑賞や演奏が一般的であるが、作曲に興味を持つ人も多い。しかし、音楽の専門的な訓練を受けず、音楽的経験や知識が必ずしも豊富ではない場合、自分の頭の中にある曖昧なイメージを具体的な楽曲として表現することには困難を伴う。最近、深層学習に基づく自動作曲システムの研究が盛んになりつつあり、これらの手法を利用すると簡単に音楽を作ることにはできる。しかし、学習データに依存して自動で生成される音楽のクオリティは十分ではなかったり、ユーザのアイデアが反映されない問題があった。

本稿では、ユーザが即興で歌唱したメロディを入力としたメロディ生成手法を提案する。イメージをうまく音符化できなかったり、歌唱力の限界、採譜誤りなどが原因で、入力となるメロディには音楽的に不自然な箇所が含まれている。それを自動的に修正することで、ユーザの意図を汲んだ生成を実現する。

## 2. 関連研究

平松ら [1] は、ポピュラー音楽に対するメロディ推定の誤りを修正する後処理手法を提案している。本手法は、入力された誤りを含むメロディが、真のメロディに対して音符の挿入および削除、および既存の音符に対する音高や発音時刻の変化という操作を通じて生成されたものであるとみなし、真のメロディを推定するものである。本手法は我々の目的にも利用可能と考えられるが、即興歌唱では現れにくい音符の挿入の考慮や、歌唱メロディにおいて重要な繰り返し構造の修正が難しいことなど、本研究の目的に合わない部分もある。そのため、ここではより歌唱メロディに特化した手法を検討する。

## 3. 提案法

提案法はデノイジング変分自己符号化器 (DVAE) [2] を用いてメロディの修正を行う。DVAEは変分自己符号化器 (VAE) の一種で、基本的な構成は通常の VAE と同様である。通常の VAE と異なる点は、学習データとして正解データとエラーデータを別個に用意することで、エンコーダの入力にはエラーデータを与えて、デコーダの出力に対して損失関数を計算する際には正解データを使用し、エラーデータから正解データを復元するモデル

としての学習を行う。DVAE を用いる理由は、潜在変数を用いて時間的に離れた箇所に情報を人為的に送ることが可能であるため、将来的には繰り返し構造を持ったメロディの生成といった発展的課題に利用可能であると考えられるためである。

DVAE のエンコーダ及びデコーダは、ともに 5 層の全結合層から成る。入力は  $130 \times 128$  の、各要素が 1 または 0 の値を取る行列である。各列が 1 小節の 32 分の 1 を単位時間とした時刻を、各行が各時刻における音の状態を表し、全体で 4 小節分のメロディを表現する。行のうち 0 番目から 127 番目までは音高を示しており、位置を表す数値は MIDI ノートナンバー [3] に対応している。数値が 1 である行がその時刻における音高となる。128 番目は休符を表している。この行が 1 である場合、その時刻には音が鳴っていないものとする。ここでは単音のメロディを扱うため、1 つの列につき 0 番目から 128 番目までの行のうち 1 つだけが 1 をとり、その他は 0 をとるものとする。129 番目の行は発音タイミングを表す。この行が 1 の場合その時刻に新たに音が発音されはじめ、0 の場合その時刻には直前に発音された音が継続している。入力データにおいては、これが 0 の場合音高は直前にこの行が 1 になっていた時刻のものと一致していることが求められる。なお、DVAE が出力した行列データから楽譜への変換を行う際、発音タイミングが 0 であるが音高が直前の時刻から変化している箇所についても、直前に発音された音が継続しているものとして扱う。

エンコーダはこの行列をひと繋がりにした 1 行の行列を入力とする。本稿ではデータのうち発音タイミングを表す 129 行目をエンコーダの入力およびデコーダの出力から除外し、エンコーダの入力前に分離した 129 行目をデコーダの出力に結合する方式を取る。そのため、エンコーダの入力は 129 行 128 列をひと繋がりにした 16512 次元のデータとなる。データの次元数は各線形層によって 16512, 8000, 4000, 2000, 1024 の順に推移する。デコーダはこの逆の向きにデータを展開する。

### 3.1 学習データ

学習データとして、J-POP 楽曲 200 曲のメロディのデータを使用した。これらはいずれも、曲全体を通して 4/4 拍子のものである。各楽曲を 1 小節目から 4 小節目、2 小節目から 5 小節目というように 1 小節目ずつ始点をずらしながら 4 小節目単位で切り出し、それらを上述した入力行列に変換している。

Melody Substantiation from Improvised Singing for Music Composition Assistance: Ryo Matsubayashi and Kazuyoshi Yoshii (Kyoto Univ.)

### 3.2 学習手法

DVAE の学習はバッチサイズが 1000 のミニバッチ学習とし、エポックサイズは 70 とした。各エポックごとに、学習データに対して音高ノイズを付加する。ノイズの付加は、各発音タイミングごとにその時刻の音高を一定確率で変化させる形で行う。確率は、上下いずれかへの半音移動がそれぞれ 10 %、全音移動がそれぞれ 5 % である。このノイズを付加したデータをエラーデータ、付加前のデータを正解データとして DVAE の学習を行う。なお、発音タイミングについては、今回は修正の対象外であることからノイズの付加は行わない。

### 3.3 評価

評価用楽曲として、RWC 研究用音楽データベース: ポピュラー音楽 [4] より 1 曲\*を使用する。これに学習データと同じ方法でノイズを付加したものを使用し、入力前のメロディやノイズを付加したメロディと修正後でメロディがどう変化したかを検討する。なお、検証用のデータは学習データとは異なり、重複する小節がないように前から 4 小節ずつ切り出して入力行列に変換しており、小節数を 4 の倍数とするため末尾に休符だけからなる 2 小節を追加している。

## 4. 結果

入力前のメロディ、元のメロディにノイズを付加した状態のメロディ、DVAE により修正したメロディのそれぞれについて、21 小節目から 28 小節目までの楽譜を以下に示す。



図 1: 入力前のメロディ



図 2: ノイズを付加したメロディ



図 3: DVAE により修正したメロディ

元のメロディには存在しない臨時記号のついた音符がノイズを付加した状態には多数存在しているが、修正後にはなくなっている。また、23 小節目にある同音連打のフ

レーズは元のものに近いものになっており、この部分はある程度修正が成功したといえる。一方、その他の小節では元とは大きく離れたメロディになっているうえ、元々ほぼ同一のメロディであり、ノイズ付加後も類似したものとなっている 21 小節目と 25 小節目や 23,24 小節目と 27,28 小節目がそれぞれ大きく異なるメロディになっており、繰り返し構造を捉えられていないという問題もあることが認められる。

## 5. おわりに

本稿では DVAE を使用した即興歌唱したメロディの修正によるメロディ生成について述べ、人為的ノイズを用いてモデルがメロディの音楽的妥当性を向上させるかどうか検証した。現時点では然るべき成果は出ていないため、引き続き適切にメロディを修正できるモデルの検討を続ける必要がある。また、その後に行うべき事柄として、発音タイミングに対するノイズの付加と VAE モデルの適用、並びに実際の歌唱メロディを用いたモデルの検証が挙げられる。前者は実際の歌唱においては音高以上にタイミングのずれが生じると考えられることからモデルの精度向上に必須の項目であり、後者はこのモデルが実際に歌唱からのメロディ生成に有用であることを確かめるために不可欠の工程である。また、人間が生成されたメロディの音楽的妥当性や修正前のメロディとの質の評価を行う過程も重要になると考える。

謝辞 本研究の一部は、JST No. JPMJPR20CB および科研費 No. 19H04137, 21H03572 の支援を受けた。

## 参考文献

- [1] Y. Hiramatsu *et al.*: “Statistical Correction of Transcribed Melody Notes Based on Probabilistic Integration of a Music Language Model and a Transcription Error Model,” *ICASSP*, 256–260, 2021.
- [2] D.J. Im *et al.*: “Denoising Criterion for Variational Auto-Encoding Framework,” *arXiv preprint arXiv:1511.06406v2*, 2016.
- [3] The MIDI Manufacturers Association: “The Complete MIDI 1.0 Detailed Specification,” 1996, Retrieved January 8, 2021, <https://www.midi.org/specifications-old/item/the-midi-1-0-specification>
- [4] M. Goto *et al.*: “RWC Music Database: Popular, Classical, and Jazz Music Databases,” *ISMIR*, 287–288, 2002.

\*RWC-MDB-P-2001 No.92