Vol.2025-MUS-142 No.35
2025/3/7

# 歌詞の音韻特徴量に対する適合度を考慮した歌唱曲の生成と自動評価

西村 草介[1,a)]　中村 栄太[2]

**概要**：本デモ発表では，日本語歌詞の音韻特徴量との関係に基づき，歌唱曲の生成と自動評価を行う手法について紹介する．自動評価手法は，「作曲家は歌詞に対する適合度を重みとして選択したメロディを作品として残す」という作品選択モデルに基づいている．具体的には，深層学習を用いて構築する歌詞条件付きと条件なしのメロディ生成モデルの尤度比から適合度を推定する．本手法の特徴は，評価値データを使用せず，歌詞付きメロディデータのみで楽曲評価値の推定を行える点である．また，従来注目されてきたアクセントと音高の関係に加え，単語境界とリズムの関係も考慮している点も特徴である．上記のメロディ生成モデルと自動評価手法を用いることで，歌詞に対する適合度を操作してメロディを生成することも可能となる．デモでは，実際に生成したサンプル楽曲や評価値予測の結果について発表する．

## 1. はじめに

自動作曲 [1] の研究では，楽曲の品質を自動評価する手法が重要視されている．音楽の自動評価には，主観評価データを用いる方法があるが [2]，評価データの収集には高コストがかかる．そこで本研究では，芸術制作過程を表す作品選択モデルに基づき，既存曲データのみで学習可能な評価手法を提案する．本稿では，日本ポピュラー音楽の歌唱曲に注目し，歌詞に対するメロディの評価について考える．歌詞の音韻特徴量とメロディの対応関係は，曲の聴きやすさや歌いやすさに影響し，山田耕筰は歌詞抑揚と旋律輪郭の一致原理（歌詞のアクセント遷移とメロディの音高遷移の方向を一致させる作曲原理）を提唱した [3]．近年では，この一致原理に関するコーパス分析 [4] や歌詞のアクセントを音高制約とする自動作曲 [5] の研究があるが，楽曲の評価値予測までは行っていない．メロディ要素には音高とリズムがあり，それらに影響を与え得る歌詞特徴量はアクセントの他にも単語境界や意味情報などが考えられる．本研究では，この歌詞とメロディの複雑な関係性を捉える深層生成モデルを構築し，歌詞に対するメロディ生成およびメロディの評価値予測に対する効果について調べる．

[1] 九州大学工学部電気情報工学科
[2] 九州大学システム情報科学研究院
[a)] nishimura.sosuke.638@s.kyushu-u.ac.jp

## 2. 提案手法

### 2.1 歌詞付きメロディデータの構築と基礎分析

歌詞付きメロディデータは，メロディの音楽情報と歌詞の言語情報からなる．メロディの各音符は，MIDI ノート番号で表す音高と，1 小節を 48 分割したティック単位で表す発音時刻のペアで表す．音高は，調号のデータを用いて自然調に移調してから計算する．歌詞は，通常の日本語表記の「平文」と，各音符に割り当てられた振り仮名で表される「音符ルビ」の二つの形態で表す．平文は後述の音韻特徴量の抽出に用いる．音符ルビは最大 3 文字までの仮名で表す．データは，平文の改行箇所で区切られたフレーズを単位として整備する．

歌詞の音韻特徴量として，音符ルビの仮名ごとに割り当てられるアクセントと単語境界のラベルを用いる．先行研究 [4,5] に従い，アクセントは低 (0) と高 (1) の二値ラベルで表す．単語境界も二値のラベルで表し，単語の先頭のルビに (1)，その他に (0) のラベルを付与する．アクセントと単語境界の分析には MeCab を用いた．本研究では，ポピュラー音楽のヒット曲を主に含む 1988 曲から 33,751 フレーズのデータを収集・整備した．

### 2.2 作品選択モデルに基づく評価値予測

作品選択モデルの仮定では，創作者は様々な作品候補を生成して，その中から作品として残すものを，作品の適合度を重みとして確率的に選択する．本手法では，このモデ

ルを用いて適合度を推定し，その適合度を用いて楽曲の評価値の予測を行う．具体的には，候補のメロディ $X$ の生成確率を $\widetilde{P}(X)$，歌詞 $Y$ に対するメロディ $X$ の適合度を $W(X;Y) \in \mathbb{R}_{>0}$ とすると，$Y$ に対する $X$ の生成確率 $P(X|Y)$ は次の様に表される．

$$P(X|Y) = \widetilde{P}(X)W(X;Y) \tag{1}$$

ここで，$\widetilde{P}(X)$ がメロディ $X$ の周辺化確率 $P(X) = \sum_Y P(X|Y)P(Y)$ と一致するという「学習と生成過程の整合性の仮定」を用いると，次式が得られる．

$$W(X;Y) = P(X|Y)/P(X) \tag{2}$$

式 (2) により，メロディの適合度が歌詞条件付き確率と条件なし確率の比として求まる．また適合度 $W(X;Y)$ と等価な音符単位のクロスエントロピー差（CE 差）を $\Delta\mathrm{CE}(X;Y) = \log_2 W(X;Y)/L(X)$ 定義する（$L(X)$ はメロディ $X$ の音符数）．

本研究では，聴取実験において多数の鑑賞者によるメロディ評価の結果として得られた平均的な評価値を実測データとして扱う．具体的には，ある歌詞に対して二つのメロディを付けた歌唱曲を対比較して，一方のメロディがより好ましい方として選ばれた割合をそのメロディの評価値と定義する．上で定式化した創作者にとっての適合度からこの評価値を予測する手法として，ロジスティック回帰を用いる方法を考える．

### 2.3 深層生成モデルを用いたメロディ生成と適合度推定

自己回帰型の生成モデルである LSTM ネットワークを用いて，前節の $P(X|Y)$ と $P(X)$ を構成する．以下，前者を歌詞条件付き LSTM，後者を歌詞条件なし LSTM と呼ぶ．効率的な学習のため，音高のオクターブ移調と発音時刻の小節単位の平行移動について対称性を持つモデルを考え，メロディの各音符を拡張音高クラスと拍節位置で表す．

歌詞条件なし LSTM は，各音符に関して直前の音符の拡張音高クラスと拍節位置を入力として，次の音符の拡張音高クラスと拍節位置の予測確率を出力する．歌詞条件付き LSTM では，入力に次の音符の歌詞特徴量を含める．出力は歌詞条件なし LSTM と同じである．歌詞条件あり LSTM によって，歌詞の音韻特徴量の依存性を取り入れたメロディが生成できる．また，上記の適合度の推定手法を組み合わせれば，適合度を操作したメロディ生成が可能である．

## 3. 実験結果

### 3.1 メロディ生成による適合度の有効性の調査

2.3 節のモデルを用いてメロディ生成を行った結果の例を図 1 に示す．後半の「あざやかな」のアクセントに注目すると，歌詞条件なし LSTM で生成した上の結果の音高遷移はアクセント遷移に従っていないが，歌詞条件付き LSTM

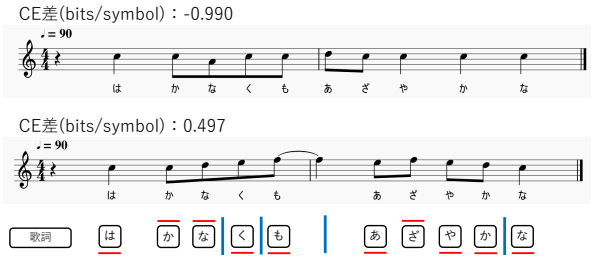


**図 1** 歌詞条件なし LSTM（上）と歌詞条件付き LSTM（下）で生成したメロディ．赤線は高低アクセント，青線は単語境界を表す．

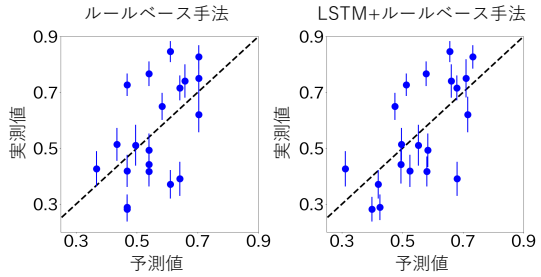


**図 2** 評価値の実測値と予測値（縦棒は標準誤差）

で生成した下の結果ではアクセントの上昇の下降がある部分で音高遷移の方向が一致していることが確認できる．また，単語境界に注目すると，下の結果では助詞の「も」や「な」で，長い音価が使われており，歌詞の句構造を反映したメロディになっている．各々の結果の CE 差を見ると，下の方が大きな値となっており，歌詞に対するメロディの適合度の大小関係を適切に表していることも確認できる．

### 3.2 評価値予測の評価実験

評価値データを収集するため，参加者に同じ歌詞に対する 2 つのメロディを提示してより好ましいと感じた方を選択してもらう実験を行った [6]．実験では，20 種類の歌詞に対する 20 組のメロディを用いて，NEUTRINO [7] で合成した歌声をランダムな順序で提示した．比較対象のメロディのペアは，学習済みの歌詞条件付き LSTM と歌詞条件なし LSTM を用いて生成した．各メロディのペア $i$ について，歌詞条件付き LSTM で生成した方（CE 差が大きい方）が選択された割合 $p_i$ を評価値データとして用いる．

評価実験では，歌詞のアクセント遷移と音高遷移の一致度に基づくルールベース手法と比較し，評価値の予測誤差を評価尺度として用いた．この結果，LSTM を用いた提案法をルールベース手法に統合した場合の予測誤差 0.134 は，ルールベース手法の誤差 0.149 に比べ小さく，提案法の有効性が示された．提案法における予測値と実測値の相関係数は 0.653 であり（図 2 右），実験で得られた評価値を比較的高い精度で予測できることが示された．

## 4. おわりに

本研究では，ポピュラー音楽において，歌詞のアクセン

トおよび単語境界がメロディの評価値に大きく寄与することを明らかにした．また，これらの歌詞の音韻特徴量を入力に用いたメロディの深層生成モデルを用いることで，歌詞に対するメロディの適合度が高い生成サンプルが得られることも示した．さらに，楽曲データのみを用いて歌詞に対するメロディの評価値を予測できる手法を構築し，メロディ生成モデルの確率比に基づく適合度推定が，歌詞抑揚と旋律輪郭の一致原理に基づくルールベースの適合度と組み合わせることで，高精度な評価値予測が可能であることを確認した．本手法の枠組みは汎用性を持ち，日本語以外の歌詞や器楽曲の自動評価への応用が考えられる．また，音色や歌唱表現などの音楽要素を取り入れる拡張や，楽曲のジャンル，年代，テンポなどを適合度モデルに組み込むことで，より精緻な評価手法の構築が今後の課題である．

## 謝辞

本研究は，JST FOREST No. JPMJPR226X 及び科研費 No. 23K24917 の支援を受けた．

### 参考文献

[1] S. Ji et al.: A survey on deep learning for symbolic music generation: Representations, algorithms, evaluations, and challenges, ACM Computing Surveys, Vol. 56, No. 1, Article 7, 2023.
[2] G. Cideron et al.: MusicRL: Aligning music generation to human preferences, Proc. ICML，pp. 8968–8984, 2024.
[3] 山田耕作: 歌謡曲作曲上より見たる詩のアクセント, 詩と音楽, Vol. 2, No. 2, 1923.
[4] 堤彩香ら: 日本語の音韻と旋律の関係について〜童謡・唱歌を中心に〜, 情報処理学会研究報告, Vol. 2014-MUS-105, No. 5, 2014.
[5] 深山覚ら: 音楽要素の分解再構成に基づく日本語歌詞からの旋律自動作曲, 情報処理学会論文誌, Vol. 54, No. 5, pp. 1709–1720, 2013.
[6] 実験ページ: https://ice.inf.kyushu-u.ac.jp/ExpNishi2/
[7] NEUTRINO Diffusion: https://studio-neutriNO.com/