

自動採譜誤りの効率的修正を支援する人間介在型歌声採譜手法

杉本 悠¹

中村 栄太^{1,2}

¹九州大学 大学院システム情報科学府

²ピティナ音楽研究所

1. はじめに

音楽音響信号から歌唱パートの各音符の音高と発音時刻、消音時刻を推定する歌声 MIDI 採譜は、歌唱支援や音楽分析に用いられる基盤技術である。近年、自動採譜手法の精度は向上しているが、依然として多くの推定誤りが生じる [1, 2]。このため、自動採譜結果を人手で確認・修正する過程を効率化する人間介在型歌声採譜手法の研究が進められており、公開システムである Tony [3] では、修正作業を効率化するため、音符の分割や結合などの編集操作が実装されている。その評価実験では、特に音高や発音時刻の修正に比較的多くの時間を要することが示されている。

採譜結果の修正をさらに効率化する方法として、音高や発音時刻などの修正候補の表示が考えられる。本研究では、深層学習に基づく自動採譜手法を用いて、修正候補提示を含む人間介在型歌声採譜手法の設計や有効性の調査を行う。最近の採譜手法における推定誤りの分類および頻度分析を行った上で、有効な修正候補の探索や表示の方法について調べる。また、これらの分析に基づいて設計された人間介在型歌声採譜システムを実装し、その有効性を評価実験により検証する。

2. 採譜誤りと有効な修正候補に関する分析

2.1 自動採譜結果の誤り分析

自動採譜結果の修正に必要な編集操作を定量的に調べるため、採譜誤り種類別の頻度分析を行う。採譜誤りには、音符単位の脱落および挿入に関する誤りと、音符の各属性（音高・発音時刻・消音時刻）に関する誤りがあり、前者はさらに音符の結合および分割に関するものとそれ以外に細分化できる。正解音符列と自動採譜による推定音符列の音符同士の対応（アライメント）が与えられる時、それらの誤りは以下のように定義できる。

- **結合誤り**：正解音符列の複数音符が推定音符列では 1 音符に結合されている誤り
- **脱落誤り**：正解音符列にある音符が推定音符列には存在しない誤りで、結合誤り以外のもの
- **分割誤り**：正解音符列の 1 音符が推定音符列では複数音符に分割されている誤り
- **挿入誤り**：正解音符列にはない音符が推定音符列に存在する誤りで、分割誤り以外のもの
- **音高誤り**：推定音高が正解とは異なる誤り
- **発音時刻誤り**：推定発音時刻が正解から 50 ms 以上、150 ms 未満ずれている誤り

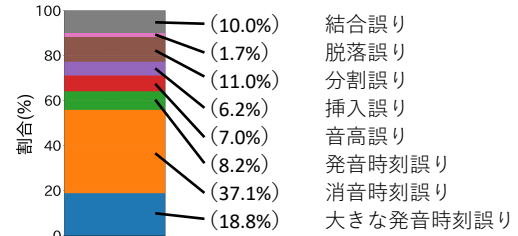


図 1: 自動採譜結果に含まれる誤りの頻度

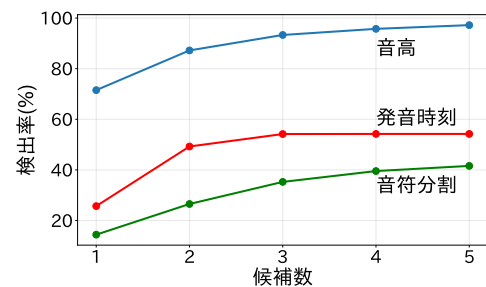


図 2: 候補による正解検出率

- **消音時刻誤り**：推定消音時刻が正解から 50 ms 以上ずれている誤り
- **大きな発音時刻誤り**：推定発音時刻が正解から 150 ms 以上ずれている誤り

なお、大きな発音時刻誤りには、アライメント誤りに起因するものが多く含まれると考えられる。

日本のポピュラー音楽のデータ 112 曲に対して、CRNN に基づく自動採譜手法 [4] による推定結果を用いて誤り分析を行った。音符列間のアライメントには Symbolic Music Alignment Tool [5] を用いた。分析結果を図 1 に示す。脱落誤りと挿入誤りの頻度は小さい一方で、結合誤りと分割誤りの頻度が相対的に大きいことから、修正操作として音符の分割および結合操作を導入する重要性が示唆される。また、大きな割合を占める音高と発音時刻、消音時刻の誤りの修正や音符分割では、修正後の音符属性値を指定する必要があるため、修正操作の効率化にはこれらの値の候補提示が有効であると考えられる。

2.2 修正候補の有効性分析

修正候補を得るには、自動採譜手法の出力である音高予測確率と発音予測確率を用いる方法が考えられる。音高誤りに関しては、誤りがある音符の音符区間における音高予測確率の平均値が上位 n 個の音高を修正候補とする。発音時刻誤りに関しては、現在の発音時刻の前後 150 ms の範囲で、発音予測確率が局所最大となる時刻の内、確率が上位 n 個の時刻を修正候補とする。音符分割に関しては、結合誤りが含まれる音符区間で、発音予測確率が局所最大となる時刻の内、確率が上位 n 個の時刻を分割点の修正候補とする。これらの方法で求めた修

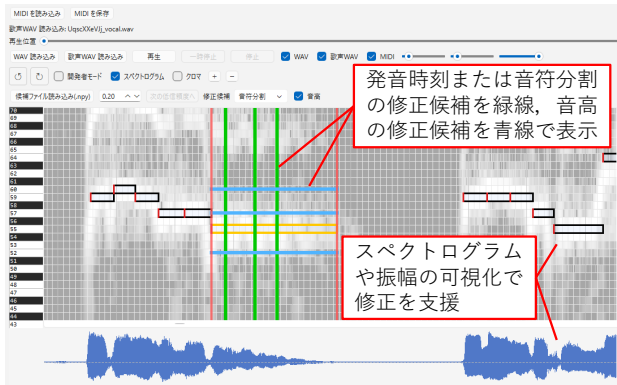


図 3: 提案システムの表示画面

表 1: 修正による F 値の平均改善率 (%)

実験条件	COn	COnP	COnPOff	修正音符数
候補なし	-0.01	0.17	0.49	14.71
候補あり	0.18	0.39	0.49	17.06

正候補に正解が含まれる割合を検出率として、候補数 n ごとの検出率を求め、修正候補としての利用可能性と有効な表示候補数を分析する。

図 2 に修正候補数と検出率の関係を示す。音高では、候補数が 3 以上で 90% 以上の検出率となり、上位 3 個の修正候補により多くの誤りを修正できる可能性が高いと考えられる。発音時刻と音符分割に関しても候補数が 3 程度で検出率の増加が緩やかになっており、有効な修正候補の個数は 3 程度であると示唆される。一方、これらの要素の検出率は候補数が 5 でも 40% から 60% に留まっている。局所最大点では正しい修正候補を検出できなかったなどの原因と考えられ、これらの修正候補の抽出方法には改善の余地がある。

3. 人間介在型歌声採譜システムの実装

提案システムでは、入力された MIDI ファイルを表示し、エディタ上で修正できる (図 3)。基本的な修正操作として、音符の削除、挿入、結合、分割、移動、発音時刻・消音時刻修正を行うことができる。また、音源の WAV ファイルを入力することで MIDI との同期再生やスペクトログラムの表示ができる。

本システムでは、音高と発音時刻、音符分割時刻の修正候補を提示する。2.2 節の分析結果を踏まえ、選択された音符に対し、修正候補を最大 3 個表示する。修正候補は、キーボード操作や候補線のクリックで選択でき、採譜音符列に反映できる。実装には PySide6 を用い、音響信号処理には librosa、MIDI の処理には mido を用いた。

4. 評価実験

提案システムを用いて修正候補提示の有効性を調べる評価実験を行った。作曲や演奏の経験を有する 5 名のユーザーに、候補表示なしと候補表示ありの条件でシステムを使用してもらった。実験には正解の MIDI データがある日本のポピュラー音楽の楽曲を使用し、深層学習ベースの採譜手法 [4] による推定 MIDI を初期音符列と

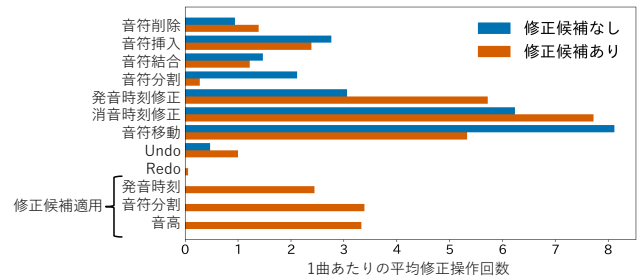


図 4: 修正操作回数の頻度分布

して用いた。修正候補なしの条件で計 17 曲、修正候補ありの条件で計 18 曲を対象として、システムの使用時間は 1 曲あたり 5 分間とした。評価には、修正による採譜精度の変化と修正操作の回数を用いる。採譜精度の評価指標には、発音時刻のみ、発音時刻と音高、発音時刻と音高と消音時刻の一致を見る COn, COnP, COnPOff の 3 種類の F 値を用いる。

表 1 の結果において、COn と COnP の改善率が候補表示ありの方が大きいことから、候補表示の有効性が確認できる。また、修正候補なしに比べ、修正候補ありの方が修正音符数が増加したことから、修正候補の提示により修正の効率が向上したと考えられる。一方、精度の改善率が小さい理由として、ユーザーの採譜経験やシステムへの習熟の不足、5 分間という短い時間内で修正箇所の探索が十分にできなかったことなどが考えられる。

図 4 の結果から、修正候補なしの条件に比べて、修正候補ありの条件では音符分割および音高の修正操作回数が減少しており、修正候補の使用による効果が確認できる。また、修正候補ありの条件では、3 種類の修正候補提示が同程度に使われており、いずれも有効な修正操作につながったことを示している。さらに、発音時刻修正の回数が増加していることから、発音時刻の候補がユーザーによる修正における判断材料として有効的に使われる可能性も示唆された。

5. おわりに

本研究では、自動採譜結果の修正において音高および発音の予測確率から得られる修正候補の提示により効率的な修正操作が可能であると示された。今後の課題として、修正候補の精度向上や消音時刻修正など他の修正操作に対する修正候補の表示が挙げられる。また、修正が必要な箇所の探索コストを軽減するため、推定音符の信頼度を求めて表示する方法も検討に値する。本システムの多声音楽への拡張や修正結果を用いた自動採譜手法の適応学習などの方向性も重要と考えられる。

謝辞 本研究は、JST FOREST No. JPMJPR226X 及び科研費 Nos. 25H01148, 25H01169 の支援を受けた。評価実験の協力者に感謝する。

参考文献

- [1] J.-C. Wang et al.: *Proc. ISMIR*, pp. 454–461, 2024.
- [2] J.-Y. Wang et al.: *IEEE/ACM TASLP*, vol. 31, pp. 383–396, 2022.
- [3] M. Mauch et al.: *Proc. TENOR*, pp. 23–31, 2015.
- [4] Y. Sugimoto et al.: *Proc. APSIPA ASC*, pp. 305–310, 2025.
- [5] E. Nakamura et al.: *Proc. ISMIR*, pp. 347–353, 2017.