7L-03

# 既存歌唱曲のリアルタイム歌声アレンジシステム

尾島 優太† 中野 倫靖‡ 深山 覚‡ 加藤 淳‡ 後藤 真孝‡ 糸山 克寿† 吉井 和佳†

†京都大学 大学院情報学研究科 知能情報学専攻 ‡産業技術総合研究所

## 1. はじめに

音楽の楽しみ方には，与えられた音楽を聴くだけではなく，インタラクティブに楽しむ能動的音楽鑑賞も含まれる．能動的音楽鑑賞の一つの例として，既存楽曲を編曲・演奏するといった楽しみ方が挙げられる．既存楽曲を編曲・演奏することで，楽曲中の任意のパートを自分の好みのメロディに差し替えたり，リズムパターンを差し替えたりすることが可能になる．実際に動画共有サービスなどにおいて，既存楽曲に重ねてその楽曲の一部のパートを編曲し，自ら演奏した動画が投稿されている．

これまでに，既存楽曲中のドラムパートを抽出してリアルタイムに編集し，再合成する手法 [1] や，元楽曲のギターの音色特徴を活かしたまま，ギターパートのみを自らの好みのメロディに差し替える手法 [2] が提案されている．同様のアプローチで歌唱曲中の歌声パートを編曲することも考えられる．歌唱力がある人は自らの歌声で元楽曲の歌声パートを置き換えることが可能であるが，そうでない場合は，従来，歌声合成技術 [3] を利用する必要があった．歌声合成では歌詞及び音高情報を事前に用意する必要があるが，任意の既存楽曲を編集対象とする場合，それらの入手は必ずしも容易ではない．

そこで本稿では，歌唱曲から歌声を分離し，分離された歌声を直接編集することで，歌詞及び音高情報を用意せずに歌声パートを編曲できるシステムを提案する．具体的な歌声編集操作としては，音高の変更（ピッチシフト）とタイミングの変更を考える．本システムには，楽曲中の歌声を利用することで，歌声の個人性を維持したままメロディの差し替えや追加を行うことが可能であるという利点がある．また，本システムでは分離された歌声から音高を推定し，推定された音高が楽曲の再生に合わせて可視化される．歌声の編集は MIDI キーボードを用いてリアルタイムに行われるため，楽器演奏のようなパフォーマンスを行うことが可能である．

## 2. ユーザインタフェース

本システムでユーザに提示される画面を図 1 に示す．ユーザは MIDI キーボードで操作する．画面上での表示は，MIDI キーボードと同様に左右方向で音高が変化するように設計した．白色が鍵盤上での白鍵，灰色が鍵盤上での黒鍵を表す．MIDI ノートナンバー 60 の C の鍵盤に対応する箇所は赤色で表示されている．スムーズな操作のために，ユーザのアクションが画面上で反映される必要がある．そのため，ユーザの操作した鍵盤に対応する箇所を画面上で緑色に変化させて提示する．楽曲によって必要となる鍵盤数が異なるため，画面に表示される音高数は楽曲中の最高音と最低音に応じて変化する．また，楽曲の編集の際にはオリジナルの音高・音長を参考にすると考えられるため，画面上で楽曲の再生に合わせてそれらを提示する．これは，楽曲の再生に合わせて画面上部から流れてくる黒い四角（以降，ノートと述べる）で表される．画面上部のスライドバーは楽曲中の再生位置を示す．演奏の一時停止，再開はスペースキーで操作できる．画面右上のプルダウンリストにより，演奏された鍵盤のノートへの割当アルゴリズム（後述）を決定する．さらに，画面右上のチェックボックスにより，編集前の歌声を再生するかどうかを制御できる．編集前後の歌声を同時に再生することで，元楽曲にハモリパートを付与するといった使い方が可能である．

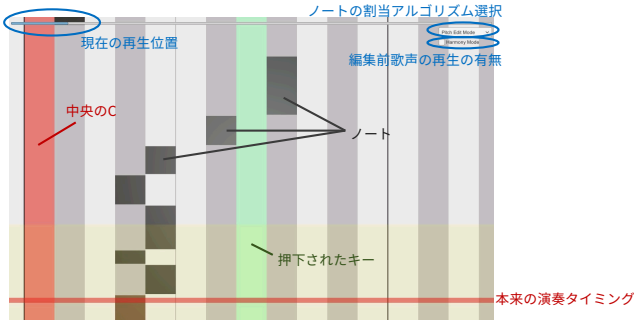


図 1: リアルタイム歌声アレンジシステムの表示画面

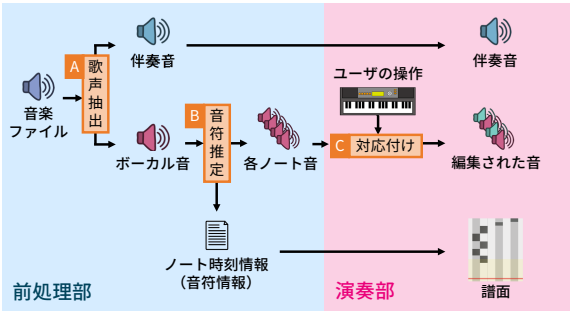


図 2: システム概要図

## 3. システム概要

本システムの概要を図 2 に示す．本システムは，楽曲を解析して楽譜情報を生成する前処理部と，ユーザの操作に応じて歌唱パートを編集・再合成する演奏部の 2 つの部分に大別される．

### 3.1 前処理部

前処理部では歌声を抽出し，その後音高の切り替わり時刻を求めることで，その切り替わり時刻で区分された歌声音響信号を得る．歌声の抽出（図中 A）については，池宮らによるロバスト主成分分析（RPCA）を用いた歌声分離手法 [4] を用いる．この手法ではまず音響信号のスペクトログラムに RPCA を適用し，入力スペクトログラムを低ランク行列とスパース行列の和へと分解する．このうちスパース成分のみをバイナリマスクで取り出すことで歌声が卓越した信号のスペクトログラムを得る．続いて，これに対し Subharmonic Summation（SHS）を用いて最尤の歌声 F0 軌跡を推定する．その後推定された F0 及びその倍音の周波数のエネルギーを通過させる調波マスクを生成し，先のバイナリマスクと統合して入力スペクトログラムに適用することで，歌声スペクトログ

Real-time Singing Voice Arrangement System for Existing Songs: *Yuta Ojima, Tomoyasu Nakano, Satoru Fukayama, Jun Kato, Masataka Goto, Katsutoshi Itoyama, and Kazuyoshi Yoshii*

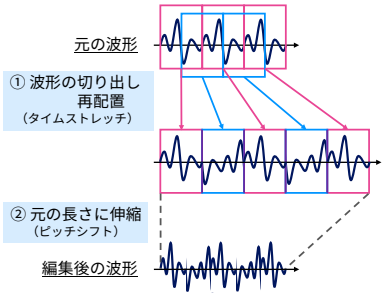


図 3: ピッチシフトの概要

ラムと伴奏音スペクトログラムを得る．これらからボーカル音と伴奏音を合成する．

一方，音高推定（図中 B）については錦見らによる音符推定手法 [5] を用いる．この手法では，歌声の背後に存在する楽譜から F0 軌跡が生成される過程をモデル化し，F0 を観測とした隠れマルコフモデルを用いて音符系列をベイズ推定する．事前に拍情報を与えることで，16 分音符単位での音高推定が可能である．これにより，各ノート音とノート時刻情報を求める．

### 3.2 演奏部

演奏部では，分割された歌声をユーザの鍵盤操作に応じて編集し，伴奏音に合成して再生する．ユーザの意図通りの編集が行われるために，ユーザの操作対象のノートを適切に決定する必要がある（図中 C）．ユーザの編集意図に応じて，決定方法を以下のように 2 種類用意した．

1. 音高・時間の双方に基づき距離を計算し，距離最小のノートを割り当てる
2. 時間が同じノートを割り当てる（ピッチのみ変更）

1 つ目のアルゴリズムでは，時刻 $t$ に演奏された MIDI ノートナンバー $k$ のキーに対し割り当てられるノート $n^*$ は下式で決定される．

$$n^* = \underset{n \in N}{\operatorname{argmin}} (w^{(\mathrm{k})}|k_n - k| + w^{(\mathrm{t})}|t_n - t|) \qquad (1)$$

ここで，$k_n, t_n$ はそれぞれノート $n$ の MIDI ノートナンバー，本来の演奏時刻を表し，$N$ は現在の時刻から一定の時刻の間に存在するノートの集合を，$w^{(\mathrm{k})}, w^{(\mathrm{t})}$ はそれぞれ音高方向，時間方向の距離に対する重みを表す．

2 つ目のアルゴリズムでは，演奏されたキーに対し，音高方向の距離によらず，現在の時刻に存在するノートが割り当てられる．時間方向に距離があるノートを考慮しないため，MIDI キーボードを押下し続けても自動的にノートが割り当てられる点が 1 つ目のアルゴリズムと異なる．これら 2 つのアルゴリズムは演奏中にプルダウンリストから変更可能である．

次に，割り当てられたノートのピッチを，演奏されたキーに応じてシフトする．リアルタイムに実現するため，計算コストが小さい手法を用いる必要がある．本研究では信号を時間領域で扱う PSOLA [6] を用いた（図 3）．この手法では，まず波形を細分し，再配置することでタイムストレッチを行う．このとき，局所的な波形情報が残るのでピッチは変化しない．その後，タイムストレッチされた波形の時間幅をタイムストレッチ前の時間幅に等しくなるように伸縮し，ピッチシフトを実現する．

## 4. 実験

被験者実験により，提案したシステムによって既存楽曲の歌声パートのどのような編曲が可能になったかについて調査を行った．被験者はピアノ演奏歴 19 年の 25 歳男性 1 名である．被験者にはそれぞれのノートの割り当てアルゴリズムの下で歌声編集を行ってもらい，編集の様子の観察，実施後の聞き取りを行った．

### 4.1 実験条件

実験には RWC 研究用音楽データベース（ポピュラー音楽）RWC-MDB-P-2001 No.7 [7] を用いた．音符推定の精度と編集のしやすさの関係を調べるため，ユーザに提示されるノートとして，正解データを用いて生成したものと，推定結果を用いて生成したものの 2 種類を用意して実験した．また，被験者には元楽曲を事前に聞いてもらっている．

### 4.2 実験結果

実験の結果，被験者から以下の意見が得られた．

- 歌声を楽器のように演奏している感覚があった
- 自らの鍵盤操作で，楽曲を音やリズムを外した歌唱に変化させることができて楽しい
- 演奏速度が早いため，初見では適切な編集を行うことが難しく，十分な練習が必要である
- 画面とキーボードが 2 つに分かれているので違和感があり，1 つのデバイス上で完結したほうが良い

このうち，初見演奏の難しさについては通常のピアノ演奏でも見られるものである．得られた意見から，提案システムは歌声を編集するシステムとしては機能しているが，演奏難易度を下げるため，デバイスを工夫する必要性が示唆された．一方，信号処理結果に関する意見としては以下の意見が得られた．

- ピッチシフトで音質が著しく悪化することがある
- 推定されたノートを用いると，聴こえた音と提示された楽譜の間に齟齬が生じて気持ち悪い

前者は歌声 F0 がゆらぎを含むことが原因であると考えられる．後者は推定結果を用いたノート生成の歳，伴奏音などが歌声として推定されたため，正解と一致しない部分があることが原因であると考えられる．

## 5. まとめ

本稿ではリアルタイムに既存歌唱曲の歌声をアレンジするシステムを提案した．被験者実験により，リアルタイムな歌声編集を実現できることが確認された．現状のシステムでは使いこなすための難易度が高いため，今後タッチパネルを入力デバイスに持つタブレット上で実装して難易度を低くするほか，伴奏部を編集する機能の追加を行う予定である．

**謝辞** 本研究の一部は JSPS 科研費 24220006, 26700020, 26280089, 16H01744, JST CREST 及び ACCEL の支援を受けた．また本研究では RWC 研究用音楽データベース（ポピュラー音楽）を使用した。

## 参考文献

[1] 吉井和佳ら: “Drumix: ドラムパートのリアルタイム編集機能付きオーディオプレイヤー,” インタラクション, 207–208, 2006.

[2] 安良岡直希ら: “フレーズ置換のための調波非調波 GMM・NMF に基づく音源分離・演奏合成,” 情処論, 3839–3852, Vol.52, No.12, 2011.

[3] H. Kenmochi *et al.*: “VOCALOID–Commercial Singing Synthesizer Based on Sample Concatenation,” *INTERSPEECH*, 4009–4010, 2007.

[4] Y. Ikemiya *et al.*: “Singing Voice Analysis and Editing Based on Mutually Dependent F0 Estimation and Source Separation,” *ICASSP*, 574–578, 2015.

[5] R. Nishikimi *et al.*: “Musical Note Estimation for F0 Trajectories of Singing Voices Based on a Bayesian Semi-beat-synchronous HMM,” *ISMIR*, 461–467, 2016.

[6] M. Valbret *et al.*: “Voice Transformation Using PSOLA Technique,” *ICASSP*, 145–148, 1992.

[7] M. Goto *et al.*: “RWC Music Database: Popular, Classic, and Jazz Music Databases,” *ISMIR*, 287–288, 2002.