

既存歌唱曲アレンジのための歌声キーボード

尾島 優太^{1,a)} 中野 倫靖^{3,b)} 深山 覚^{3,c)} 加藤 淳^{3,d)}
後藤 真孝^{3,e)} 糸山 克寿^{1,f)} 吉井 和佳^{2,g)}

概要: 本稿では、MIDI キーボードを用いて既存楽曲中の歌声を編集し、メロディの変更やハモリパート・輪唱パートの追加を可能にする歌声キーボードを提案する。このシステムでは元パートの歌詞を保存してピッチやリズムをリアルタイムで操作できるためライブパフォーマンスに用いることができ、歌手に指示して自分の好みに合うように歌ってもらっているかのような体験をもたらす。具体的には、ユーザは MIDI キーボードを用いて好みのメロディを演奏する。楽曲中の歌声パートは音符に対応した短い区間に事前に分割され、演奏モードに応じて各区間がユーザのキーボード操作に割り当てられ、ユーザのキーボード操作に応じて音高及び音長が変化する。このシステムの実現にあたり、歌声分離、歌声パートの分割、ユーザの演奏と分割区間の対応付け、リアルタイムピッチシフトが必要である。このうち対応付けについては、意図通りの編集が行われるように三種類のアルゴリズムを用意することで対処した。被験者実験の結果、提案システムの有効性が示された。

1. はじめに

音楽の楽しみ方には受動的なものだけでなく、既存楽曲のアレンジや演奏といった能動的なものも存在する。既存楽曲をアレンジすることで、音楽中の任意のパートを好みのメロディやリズムパターンに差し替えることが可能である。今日では多くの人が既存楽曲をアレンジして演奏した動画を YouTube などの動画共有サービスに投稿している。

これまでに、既存楽曲のアレンジシステムはいくつか提案されている [1,2]。これらは楽曲中の楽器パートのアレンジを可能にするシステムである。楽器パートと同様に、歌唱曲中の歌声パートのアレンジも考えられる。実際に多くのアマチュアが「歌ってみた」と題して、既存楽曲のボーカルカバーを YouTube などの動画投稿サービスに投稿している^{*1}。一方、十分な歌唱力がない場合は、歌声パートアレンジの実現のためには VOCALOID のような歌声合成システムを利用する必要がある。このようなシステムでは

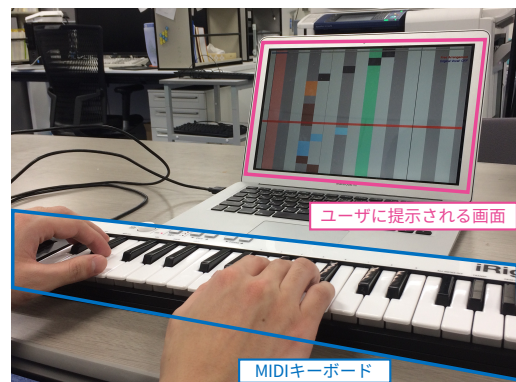


図 1 提案システムの使用例。編集対象楽曲に含まれるオリジナルの歌声パートが画面に表示され、ユーザは MIDI キーボードを用いてそれを編集する。

事前に歌詞情報及び音高情報を用意する必要があるが、任意の楽曲を編集対象とする場合、これらの情報を手に入れるのは必ずしも容易ではない。

本稿では、既存楽曲中から抽出された歌声パートを直接操作することで、事前情報なしでリアルタイムに歌声パートを編集することができる歌声キーボードシステムを提案する。抽出されたオリジナルの歌声を利用することには、2つの利点がある。1つ目は歌詞を用意せずとも歌声アレンジが可能であるという点であり、2つ目はアレンジの後もオリジナルの歌手の個性が残るといった点である。歌手の個性が残ることで、まるでユーザの指示通りにオリジナルの歌手が歌っているかのような体験をもたらす。提案システムにおけるアレンジとして、音高の変化とオンセッ

¹ 京都大学 大学院情報学研究所
² 京都大学/理研 AIP
³ 産業技術総合研究所
a) ojima@sap.ist.i.kyoto-u.ac.jp
b) t.nakano@aist.go.jp
c) s.fukayama@aist.go.jp
d) jun.kato@aist.go.jp
e) m.goto@aist.go.jp
f) itoyama@kuis.kyoto-u.ac.jp
g) yoshii@kuis.kyoto-u.ac.jp

^{*1} YouTube では 400 万を超える数の「歌ってみた」動画が投稿されている

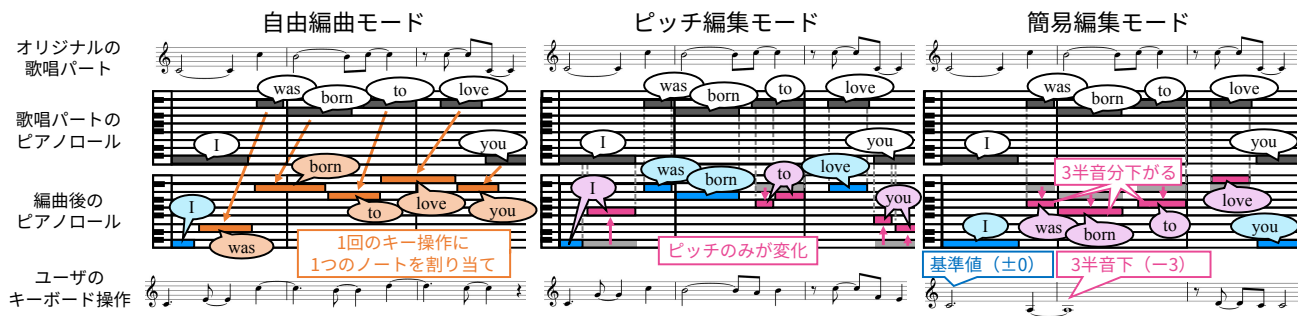


図 2 3つの対応付けアルゴリズムの概要。吹き出しは音符に割り当てられた歌詞を表す。自由編曲モードではユーザのキー操作毎に、音符が1つずつ曲の頭から順に割り当てられる。ピッチ編集モードでは音高のみが変化し、オンセット・オフセット時刻は変化しない。簡易編集モードではユーザはオリジナルのメロディからの相対的な音高変化を指定することで音高を編集する。

ト・オフセット時刻の変化を扱う。そのため、提案システムではユーザが音高、オンセット時刻及び音長を同時かつ直感的に操作できるようにする必要があり、ユーザインタフェースとしてはMIDIキーボードを用いる。これにより、ピアノを演奏しているかのように歌声編集を行うことが可能になる。

このシステムを実現するために、歌声パートの推定及び音符の可視化、音符とユーザが演奏したキーの適切な対応付け、リアルタイムピッチシフトの3つの課題に取り組む必要がある。1つ目の課題を解決するために、まずロバスト主成分分析(RPCA)を用いて楽曲から歌声パートのF0軌跡を抽出し[3]、隠れマルコフモデルを用いてF0軌跡の背後に存在する音符列を推定する[4]。2つ目の課題を解決するために、提案システムでは3つの対応付けアルゴリズムを用意する(図2)。ユーザは編集意図に応じ、これらのアルゴリズムをパフォーマンス中に切り替えることが可能である。3つ目の課題を解決するために、実時間で動くアルゴリズムであるピッチ同期波形重畳法(PSOLA)を用いる。楽曲のアレンジはMIDIキーボードを通してリアルタイムに行われるため、DJが曲を演奏するのと同様、ライブパフォーマンスとして提案システムを用いることが可能である。

2. 関連研究

本章では、提案システムの様々な側面から、関連する研究を紹介する。まず音楽アプリケーションの関連研究を紹介した後、音楽アレンジに関連した研究を紹介する。本章の最後で、提案システムの実現に深く関連する信号処理手法を紹介する。

2.1 音楽アプリケーション

提案システムは歌声パートの可視化及び一部メロディの置き換えの点で、能動的音楽鑑賞を支援するシステムであるといえる。能動的音楽鑑賞について、すでにいくつかの

研究が存在する。Gotoら[5]はSongleと呼ばれるウェブサービスを提供し、任意の楽曲の音楽構造や拍構造、メロディ、コードを可視化することで、音楽鑑賞体験をより豊かなものにした。Mauchら[6]はSong Prompterと呼ばれるインターフェースを提案した。このインターフェースは、事前の楽譜準備やアラインメントを必要とせず、楽曲構造・コード・歌詞を演奏に合わせて可視化する。Nakraら[7]はインタラクティブな指揮システムを実装した。このシステムでは、ユーザはWiiリモコンを用いてオーケストラのテンポ及び音量の時間変化を制御することができる。

一方、提案システムは音楽生成システムとしての側面も持つ。音楽生成支援システムもこれまで複数提案されている。Simonら[8]は与えられたメロディに対して和声付けを行うシステムを提案した。このシステムにより、音楽理論や作曲法に詳しくないユーザも容易に作曲に挑戦することができるようになった。McVicarら[9]は入力として与えられたコード列に対し、ギターのタブ譜を出力するシステムを提案した。このシステムでは事前に教師データを使ってコード列とタブ譜の関係を学習するため、出力されるタブ譜は教師データと似たスタイルのものとなる。そのため、教師データの背後に存在する個人性を反映したタブ譜の生成が可能である。音楽生成支援の一環として、ハーモニーパートを生成する手法も提案されている。Yiら[10]はメロディが与えられたときに4つの声部を自動生成する手法を提案した。Dannenbergら[11]はコンピュータとの協調パフォーマンスについて研究した。この研究では音楽に合わせて演奏することができるコンピュータ演奏者を用意することで、音楽家とコンピュータが協調してパフォーマンスを行うことを目的としている。

提案システムはユーザが生成した歌声パートに対する自動伴奏システムと捉えることもできる。そのような伴奏システムもいくつか提案されている。伴奏システムの実現のために、Dannenberg[12]が演奏におけるミスを考慮した、入力楽譜とソロパフォーマンスのアラインメントアルゴリ

ズムを提案し、これを用いたリアルタイム伴奏生成システムを設計した。Cont [13] も自動伴奏システムを提案している。このシステムでは隠れマルコフモデルと隠れセミマルコフモデルのハイブリッドモデルを用い、入力として与えられた音響信号と楽譜から、楽譜上での演奏位置及びテンポの動的変化を推定する。

2.2 既存楽曲のアレンジシステム

既存楽曲をアレンジするためのシステムはこれまで数多く提案されている [1, 2, 14]。Yasuraoka ら [2] は楽曲と楽曲中の任意の楽器のパート譜を入力として、その楽器の音色及びフレーズを置換する音楽アレンジシステムを提案した。このシステムでは対象となる楽器音を楽曲から抽出するために音色モデルを用いており、ユーザは置き換えたいフレーズを事前に楽譜の形式で指定する。Tsunami ら [15] は既存楽曲の歌声パートを利用してマッシュアップを作成するシステムを提案した。このシステムでは複数歌手が歌った同一楽曲から歌声パートをそれぞれ抽出し、1つの伴奏音に抽出された歌声パートを重ね合わせることでマッシュアップを作成する。

リアルタイムで楽曲の楽器パートをアレンジするためのシステムもいくつか提案されている。Yoshii ら [1] はドラムパートのアレンジシステムを提案した。このシステムではドラム音認識とビートトラッキングを用いることで、ドラムパートの音量、音色及びドラムパターンをリアルタイムで編集することが可能である。Yamamoto ら [14] は即興ライブパフォーマンスのための歌声合成ユーザインタフェースを提案している。このインタフェースでは事前に歌詞情報を与えることで、リアルタイムでの歌声合成が可能である。

2.3 信号処理手法

本節では、歌声伴奏音分離、F0 軌跡からの音符推定及びピッチシフトの手法についてそれぞれ概観する。

2.3.1 音源分離

歌声伴奏音分離に関する研究はこれまでに広く行われている [3, 16, 17]。Raffi ら [16] は繰り返し構造抽出法を提案している。この手法では伴奏音が繰り返し構造を持ち、歌声パートが繰り返し構造を持たない点に着目し、それを利用して歌声を抽出する。Huang ら [17] はロバスト主成分分析 (RPCA) を用いて歌声分離を実現した。この分離法では伴奏音の持つ繰り返し構造を低ランク成分、歌声パートをスパース成分としてモデル化する。Ikemiya ら [3] は歌声分離と F0 推定間の相互依存関係に着目し、RPCA と歌声音高推定手法を統一的な枠組みで扱うことで高精度な歌声分離を可能にした。

2.3.2 F0 軌跡に基づく音符推定

与えられた F0 軌跡に基づいて、その背後に存在する音符

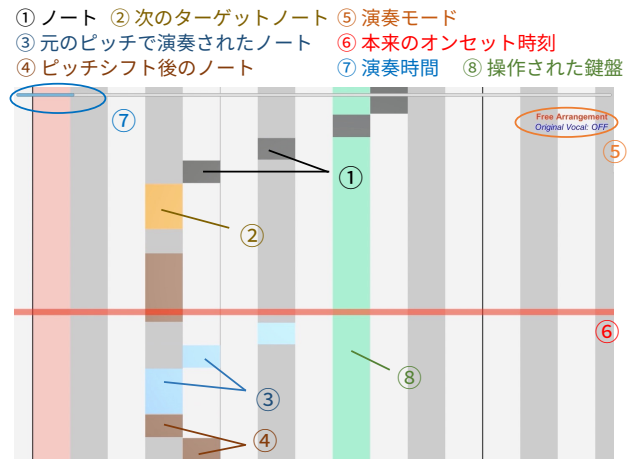


図 3 提案システムのスクリーンショット。それぞれの音符は画面上部から下方向に移動し、状態に応じて色が変化する。現在の対応付けアルゴリズムは右上に表示されており、ユーザが押した鍵盤に対応する列は緑色に変化する。

列を推定する試みも存在する。音符推定の目的は、連続的な F0 軌跡から半音単位で離散化された音高を推定することである。Songle [5] では、音高の離散化は一定のビート区間 (8 分音符など) ごとに多数決法に基づいて行われる。Laaksonen ら [18] は入力として与えられたコード列と音響信号に基づき楽譜を推定する手法を提案している。また、Ryynänen ら [19] は 1つの音符に対して内部状態 (ビブラート、オーバーシュートなど) の遷移を隠れマルコフモデルを用いて推定し、その情報を用いて音符推定を行う手法を提案した。Nishikimi ら [4] は、音符列から F0 軌跡が生成される過程をベイジアン隠れマルコフモデルを用いて定式化した。

2.3.3 ピッチシフトアルゴリズム

ピッチシフトは音楽アレンジにおいて重要な要素であり、多くの研究がなされてきた [20]。時間領域においては、ピッチを変化させずに音長を変化させるタイムストレッチの後、得られた波形を圧縮または伸長することでピッチシフトを実現する。タイムストレッチ部分については多くのアルゴリズムが存在するが、最も単純なものが Roucos らにより報告されている [21]。この手法ではまず、互いに重なり合うように、波形を一定の長さ毎に分割する。その後それらをずらして再配置し、窓関数をかけて足し合わせることでタイムストレッチを実現する。Hamon ら [22] はこの手法を、ピッチ情報を使えるように拡張した。具体的には、波形の分割の際の分割長を元の音高に応じて変化させ、再配置の際には分割区間の一部を複製または破棄することでタイムストレッチを実現する。

3. ユーザインタフェース

本章では、提案システムの詳細及びユーザインタフェースについて述べる (図 3)。3.1 節で提案システムがもたらすユーザ体験について説明し、3.2 節でユーザインタフェー

スの詳細について述べる。

3.1 提案システムの概要

本稿では MIDI キーボードを用い、まるで楽器を演奏するかのように歌声パートを操作することが可能な歌声アレンジシステムを提案する。本システムにより、

- オリジナル歌声パートのメロディを好みのメロディに差し替え
- ハモリパートの追加
- 輪唱パートの追加

といったアレンジが可能になる。これらのアレンジの際に、ユーザは歌声パートの音高、オンセット時刻及び音長を同時に変化させる。直感的にこの操作を行うため、本システムでは MIDI キーボードをユーザインタフェースとして採用した。そのため、ピアノ演奏が可能なユーザが本システムの対象である。

本システムでは、歌声パートはピッチの切り替わり毎に分割され、各分割区間がユーザにより操作される。以降、本稿ではこの分割された区間を「ノート」と呼ぶ。MIDI キーボードの鍵盤が演奏されると、まずシステムはそのキーボード操作の対象となるノートを決定する。以降、本稿ではこの操作対象ノートを「ターゲットノート」と呼ぶ。その後、ユーザのキーボード操作に応じてターゲットノートの音高及び音長が変更される。ターゲットノート決定の際のアルゴリズムとして、本システムでは「自由編曲モード」、「ピッチ編集モード」、「簡易編集モード」の3つの演奏モードを用意する (図2)。

自由編曲モード まだ演奏されていないノートのうち、オンセット時刻が最も早いものがターゲットノートとして割り当てられる。ターゲットノートはユーザが演奏した鍵盤の音高へとピッチシフトされる。全てのノートに対してキーボード操作が必要であるが、ユーザはオンセット時刻、音高、音長の全てを操作することができる。

ピッチ編集モード オリジナルの歌声パートのオンセット時刻及び音長は変化せず、音高のみが変化する。そのため、ターゲットノートとして割り当てられるのはユーザがキーボード操作をしたタイミングでオリジナルの歌声パート中に存在するノートである。このモードでは音符の途中でのピッチ変更も可能である。自由編曲モードと同様、ターゲットノートはユーザが演奏した鍵盤の音高へとピッチシフトされる。また、1つの鍵盤を押しっぱなしにすることで、連続した複数のノートに対して同じ音高を割り当てることが可能である。

簡易編集モード ターゲットノートはピッチ編集モードと同様に決定されるが、このモードではピッチシフト幅を、C4の鍵盤から見たユーザが演奏した鍵盤の相対位置により指定する。ターゲットノートは元のピッチを基準とし

て、指定した幅だけピッチシフトされる (例えばユーザが E4 の鍵盤を操作した場合、ターゲットノートのピッチは 4 半音分上がる)。

これらの3つの演奏モードは様々なアレンジ意図に対応するため用意されている。ユーザが歌声パート全てを自分の好みに合うように操作したい場合、自由編曲モードが適している。一方で、歌声パートの一部のみを編集したい場合や、タイミングを変化させたくない場合には、変化させる必要のない音に対しては操作が不要である。ピッチ編集モードや簡易編集モードが適している。また、典型的なハモリパートは主旋律からの相対位置が一定 (3度下など) であり、オンセット時刻や音長はオリジナルの歌声パートから変化しないものが多いため、ハモリパートの追加には簡易編集モードが適している。演奏モードは演奏中に MIDI キーボードを通して動的に切り替えることが可能であり、提案システムにより異なるアレンジを統一的に扱える。

さらに、提案システムではオリジナルの歌声パートの必要性に応じて、そのオンオフを切り替えることができる。これにより、ハモリパートや輪唱パートを加える際にはオリジナルの歌声パートを同時に再生し、フレーズを差し替えたい際にはオリジナルの歌声パートを鳴らさないようにするといった使い方ができる。これも MIDI キーボードを通して演奏中に動的に切り替えることが可能である。また、ユーザは画面を注視するため MIDI キーボード以外の装置を操作する余裕がないと考えられる。そこで楽曲の一時停止及び再生の操作も MIDI キーボードを通して可能とし、全ての操作を MIDI キーボード上で完結させている。

3.2 画面設計

ユーザに提示される画面が満たすべき条件として、以下の3点が挙げられる。

- オリジナル歌声パートの音高及び音長が一目で直感的に分かる形式で表示されている
- ユーザの操作が即時に可視化される
- ユーザが鍵盤操作の結果を、実際に操作する前に予測することができる

1つ目の条件は、ユーザがアレンジの際にオリジナル歌声パートを参考にするために必要である。2つ目の条件は、ユーザがアレンジをスムーズかつ快適に行うために必要である。3つ目の条件は、ユーザがリアルタイムで思い通りにアレンジを行うために必要である。

本システムにおける画面はこれらの条件を満たすように設計された (図3)。1つ目の条件を満たすため、画面上において、MIDI キーボードと同じく水平方向がピッチを表し、MIDI キーボードの白鍵、黒鍵に応じて各音高を表す領域は白または灰色で表すように設計した。また、C4 (MIDI ノートナンバー 60) を表す領域は赤色で示し、オ

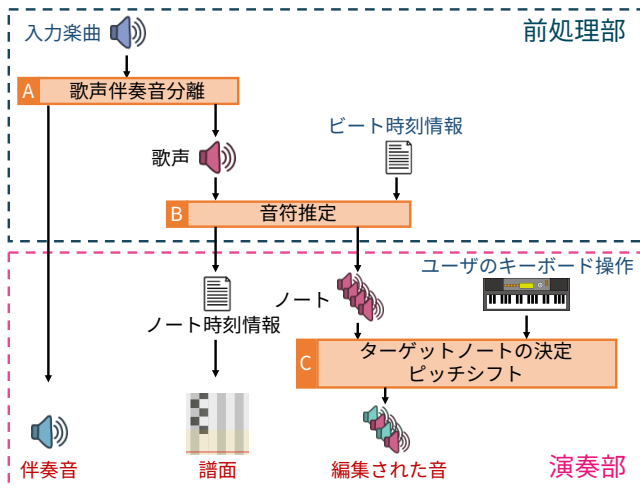


図4 提案システムの全体像。青文字がシステムに対する入力、赤文字がシステムの出力をそれぞれ表す。

クタブについても一目で分かるように設計した。オリジナル歌声パートに含まれるノートは、この領域内を画面上部から下方向に移動する黒色の矩形で表される。このデザインは既存の音楽ゲームと類似したものであり、ユーザは直感的に画面の情報を理解することができる。

2つ目の条件を満たすため、ユーザが操作した鍵盤の音高の領域は明るい緑色に変化する。また、ユーザが演奏したノートの色は水色または茶色へと変化する。このような色の変化により、ユーザは鍵盤操作の結果や自分が操作した鍵盤を、手元を見ることなく理解することができる。

3つ目の条件を満たすため、ユーザが次に鍵盤操作をしたときにターゲットノートとして割り当てられるノートの色はオレンジで表示される。これを確認することで、意図しない操作を防ぐことができる。

オリジナル歌声パートのピッチの幅は曲によって変化するため、画面に表示される音高の数はその曲の歌声パートに含まれる最高音と最低音に応じて決定される。また、現在の再生時間は、画面上部でシークバーの形式で表示される。

4. システム実装

本章では、システムにおいて用いられた信号処理技術について述べる。図4に示すように、提案システムは「前処理部」、「演奏部」の2つの部分に大別される。前処理部では歌声伴奏音分離により歌声を楽曲から抽出した後、抽出歌声のF0軌跡に対し音符推定を行うことでピッチの切り替わりのタイミングを推定する。その後、推定されたタイミングで歌声が分割される。演奏部では歌声がユーザの鍵盤操作に合わせて必要に応じてピッチシフトされ、再合成される。

4.1 歌声伴奏音分離

歌声の分離(図4中A)にはIkemiyaら[3]による手法

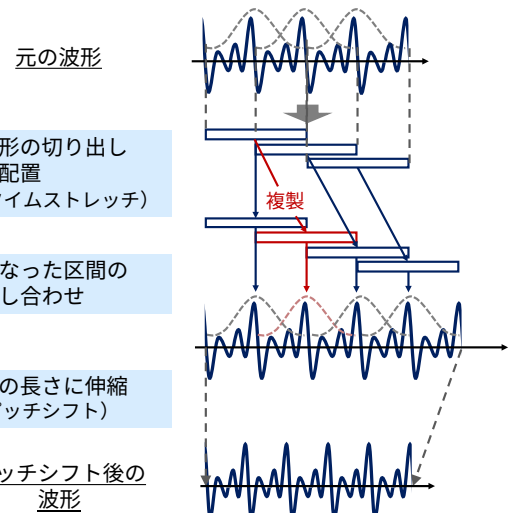


図5 PSOLA アルゴリズムの概要。波形を周期の2倍の長さを持つ窓関数により分割し、再配置することでタイムストレッチを実現する。その後、元の長さになるように圧縮または伸長することでピッチシフトを実現する。

を用いる。この手法は歌声を分離するだけでなく、後の処理に必要な情報であるF0軌跡も推定する。この手法では入力音楽スペクトログラムに対しRPCAを適用し、低ランク行列とスパース行列の和の形で表現する。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、得られたスペクトログラムに対しSubharmonic Summation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。

4.2 F0軌跡に基づく音符推定

音高推定(図4中B)については錦見らによる音符推定手法[4]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、半音単位での音高推定が可能である。これにより、各ノートの音高とノート時刻情報を求める。推定情報を用いて、ユーザに提示される画面とノート毎に分割された歌声を得る。

4.3 ピッチシフト

ユーザの操作した鍵盤に対してターゲットノートが割り当てられた後、操作に応じてターゲットノートがピッチシフトされる(図4中C)。2.3.3節で述べたように、ピッチシフトについてはいくつかのアルゴリズムが提案されている。本システムでは、PSOLAアルゴリズム(図5)を用い

た。このアルゴリズムは実時間で動く単純なアルゴリズムであり、推定された音高情報を用いるため、より単純なアルゴリズムである SOLA よりも音質が向上する。このアルゴリズムでは、推定された音高から求まる周期の2倍の長さを持つ窓関数を用いて、対象となる波形を分割する。その後、必要に応じて一部の分割区間を複製または無視して再配置することにより、タイムストレッチを実現する。この再配置の際には、隣り合う区間の一部を重ね合わせ、窓関数をかけて足し合わせる。こうしてタイムストレッチされた波形を元の長さに圧縮または伸長することで、ピッチシフトを実現する。

5. 評価実験

本章では、提案システムの有効性を評価するために行った被験者実験について述べる。

5.1 実験条件

4人の被験者を対象に被験者実験を行った。被験者はいずれも5年以上のピアノ演奏歴を持つ学生(男性3人、女性1人)である。アレンジ対象として、RWC音楽データベース(ポピュラー音楽)のRWC-MDB-P-2001 No.7 [23]を用いた。なお、被験者には対象楽曲を事前に聞いてもらっている。被験者には実験前にチュートリアルとして、著者の1人が実際にシステムを使う様子を見てもらった。その後、被験者がシステムを使う様子を観察し、被験者にはシステムの機能性、ユーザインタフェース、信号処理の3つの観点から感想を述べてもらった。さらに、以下に示す評価項目について、5段階のリッカート尺度(全くそう思わない、そう思わない、どちらともいえない、そう思う、とてもそう思う)で被験者に評価してもらった。

- (1) 自由編曲モードは使いやすかった
- (2) ピッチ編集モードは使いやすかった
- (3) 簡易編集モードは使いやすかった
- (4) 画面表示は分かりやすかった
- (5) 編集後の歌詞を聞き取ることができた
- (6) ピッチは思った通りに変化した
- (7) その歌手らしさは編集後も残っていた

評価の際には、画面に表示するノートとして、正解データに基いて生成したものを用いた。最後に、正解データではなくF0軌跡から推定されたノートを画面に表示して被験者にシステムを使ってもらい、ノートの推定精度がどの程度パフォーマンスに影響するかを確認した。

5.2 実験結果

実験において、被験者はシステム上の複数の編集モードを試しており、最も使われていたのはピッチ編集モードであった。また、多くの被験者が簡易編集モードを用いてハ

	否定 ←————→ 肯定				
自由編曲モードは使いやすかった	1	2	1		
ピッチ編集モードは使いやすかった		1		3	
簡易編集モードは使いやすかった	1			1	2
画面表示は分かりやすかった			2	2	
編集後の歌詞を聞き取ることができた		1		2	1
ピッチは思った通りに変化した			1	2	1
その歌手らしさは編集後も残っていた	2		1	1	

図6 提案システムの評価結果。数字は対応する評価(肯定:とてもそう思う, 否定:全くそう思わない)を与えた被験者の数を表す。

モリパートの生成を試みた。自由編曲モードでは、被験者は時々自分がどこを演奏しているか見失っている様子が見られた。被験者によるシステムの評価結果を図6に示す。この結果から、ピッチ編集モードが被験者の間で人気があったことが分かる。また、簡易編集モードは最も高く評価されることが多かった。一方、自由編曲モードは操作の難しさがあり、評価が最も低かった。信号処理に関しては、ピッチはユーザの意図通りに変化しており、ピッチシフト後も歌詞は認識可能であることがわかった。一方で、ピッチシフト後には歌手の個人性は失われていることが示された。被験者からは、以下に示すような意見が得られた。

まず機能面に関しては、

- 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
- 編集モードはスムーズに切り替えることができた
- このシステムはピアノの練習に役立つような気がする
- 簡易編集モードはグリッサンドやビブラートといった表現付けに役立った

という意見が得られた。これらの意見から、提案システムの歌声キーボードとしての有効性が示された。

ユーザインタフェースに関しては、

- 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
- といった意見が得られた。初見演奏での難しさは通常のピアノ演奏と同様の難しさであるとの補足意見も得られた。

信号処理技術に関しては、

- 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

といった感想が得られた。前者は歌声F0がゆらぎを含むことが原因であり、後者は、推定結果を用いたノートに歌声として推定された伴奏音の情報が一部含まれることが原因であると考えられる。

5.3 考察

得られた意見から、提案システムには、システムを使いこなすために必要な練習量、ピッチシフト後の音質、編集機能の充分性の3点において改善の余地があることが判明した。1点目に関しては、タブレットなどのタッチスクリーンを持つデバイスを入力デバイスとして用いて画面と操作デバイスを統合することで、使いやすさを向上することを計画している。デバイスの統合により、ユーザは画面とキーボードといった2つの異なるデバイスに注意を払う必要がなくなり、手元に集中できるため、編集がより容易に行えるようになると思われる。2点目に関しては、短時間フーリエ変換により、時間周波数領域において、フェーズボコーダなどを用いての信号処理を考えている。時間周波数領域では歌声に含まれるスペクトル包絡が推定可能であり、それにより元の歌手の個人性を保存したままピッチシフトを実現できる可能性がある。さらにこの改善により、編集後の歌声がよりはっきりとする可能性もある。3点目に関しては、音量を変化する機能や2つ以上のノートを同時に操作する機能の追加を検討している。

6. まとめ

本稿ではリアルタイムに歌声を編集することができるライブパフォーマンスシステムを提案した。被験者実験の結果から、提案システムは、キーボードを通して、歌声を楽器を演奏するかのように操作しているような体験をもたらすことが示された。一方で、画面とキーボードが統合されていないため、慣れるまでシステムを使うのは難しいことが判明した。そのため、今後はタブレットのような入力と画面が統合されたデバイスを用いる予定である。また、音質についても、フェーズボコーダのような他の信号処理技術を試し、改善を試みる予定である。さらに、アレンジメント機能を拡張し、伴奏パートやドラムパートについても歌声パートと同様にアレンジできるように機能を拡張したいと考えている。この種の拡張は楽曲の構成要素を再配置するといった新たなユーザ体験をもたらすと考えられる。実際にこの種の編曲は電子オルガンにより部分的に実現されているが、元の楽曲中の要素をそのまま用いるという点において大きく異なるシステムである。

謝辞 本研究の一部は JST CREST, JST ACCEL No.JPMJAC1602, JSPS 科研費 24220006, 26700020, 26280089, 16H01744, 16J05486 の支援を受けた。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] Yoshii, K. et al.: Drumix: An audio player with real-time drum-part rearrangement functions for active music listening, *Trans. of IPSJ*, Vol. 48, No. 3, pp. 1229–1239

(2007).

[2] Yasuraoka, N. et al.: Changing timbre and phrase in existing musical performances as you like: manipulations of single part using harmonic and inharmonic models, *The 17th ACM Int. Conf. on Multimedia*, pp. 203–212 (2009).

[3] Ikemiya, Y. et al.: Singing voice analysis and editing based on mutually dependent F0 estimation and source separation, *ICASSP*, pp. 574–578 (2015).

[4] Nishikimi, R. et al.: Musical note estimation for F0 trajectories of singing voices based on a Bayesian semi-beat-synchronous HMM, *ISMIR*, pp. 461–467 (2016).

[5] Goto, M. et al.: Songle: A Web Service for Active Music Listening Improved by User Contributions, *ISMIR*, pp. 311–316 (2011).

[6] Mauch, M. et al.: Song Prompter: An accompaniment system based on the automatic alignment of lyrics and chords to audio, *Late-breaking session at the 10th ISMIR* (2010).

[7] Nakra, T. M. et al.: The UBS Virtual Maestro: an Interactive Conducting System, *NIME*, pp. 250–255 (2009).

[8] Simon, I. et al.: MySong: automatic accompaniment generation for vocal melodies, *The SIGCHI Conf. on Human Factors in Computing Systems*, pp. 725–734 (2008).

[9] McVicar, M. et al.: AutoRhythmGuitar: Computer-aided composition for rhythm guitar in the tab space, *Joint ICMC and SMC Conf.* (2014).

[10] Yi, L. et al.: Automatic Generation of Four-part Harmony, *UAI Applications Workshop* (2007).

[11] Dannenberg, R. B. et al.: Human-computer music performance: From synchronized accompaniment to musical partner, *SMC* (2013).

[12] Dannenberg, R. B.: An on-line algorithm for real-time accompaniment, *ICMC*, pp. 193–198 (1984).

[13] Cont, A.: A coupled duration-focused architecture for real-time music-to-score alignment, *IEEE Trans. on PAMI*, Vol. 32, No. 6, pp. 974–987 (2010).

[14] Yamamoto, K. et al.: LiVo: Sing a Song with a Vowel Keyboard, *JIP*, Vol. 24, No. 3, pp. 460–468 (2016).

[15] Tsuzuki, K. et al.: Unisoner: An interactive interface for derivative chorus creation from various singing voices on the web, *ICMC*, pp. 790–797 (2014).

[16] Rafii, Z. et al.: Music/Voice Separation Using the Similarity Matrix, *ISMIR*, pp. 583–588 (2012).

[17] Huang, P.-S. et al.: Singing-voice separation from monaural recordings using robust principal component analysis, *ICASSP*, pp. 57–60 (2012).

[18] Laaksonen, A.: Automatic Melody Transcription based on Chord Transcription, *ISMIR*, pp. 119–124 (2014).

[19] Ryyänen, M. P. et al.: Automatic transcription of melody, bass line, and chords in polyphonic music, *Computer Music Journal*, Vol. 32, No. 3, pp. 72–86 (2008).

[20] Udo, Z. et al.: *DAFX - Digital Audio Effects*, Wiley (2002).

[21] Roucos, S. et al.: High quality time-scale modification for speech, *ICASSP*, pp. 493–496 (1985).

[22] Hamon, C. et al.: A diphone synthesis system based on time-domain prosodic modifications of speech, *ICASSP*, pp. 238–241 (1989).

[23] Goto, M. et al.: RWC music database: popular, classic, and jazz music databases, *ISMIR*, pp. 287–288 (2002).