

歌詞朗読音声からのオンライン歌声合成システム

山田 諒太郎^{†1} 森山 剛^{†1}

本研究では、ユーザが歌の歌詞を朗読した音声波形を音素単位に分割し、入力された楽譜情報の音符（音長及び音高）に対応するよう、繋ぎ合わせることで歌声を合成する。さらに、手書きの図形を演奏表現に変換する手法と組合せ、オンラインで、ユーザ自身の声を用いたお絵かきの感覚で、歌声を編集及び合成できる手法を提案する。

A Method of Singing Synthesis from Lyrics Reading Speech

RYOTARO YAMADA^{†1} and TSUYOSHI MORIYAMA^{†1}

We propose a method of synthesizing singing voice from reading speech. It allows the user to read the lyrics of a song, divides the reading speech into phonemes, and concatenates them into a singing voice while changing the duration, the voice pitch, and the strength depending on the music. It also allows the user to play with the singing method by providing picture drawing interface.

1. はじめに

幼児期に音楽を教育する方法については、リトミックを取り入れた遊戯による方法が採られている¹⁾。歌を歌うことに関しては、大人の歌う様子を見せて、模倣させることが一般的であるが、実は上手に模倣できるか否かは、子供の才能に任されており、結果として、歌をはじめ音楽に対する苦手意識を植えつけたり、嫌いにさせてしまうといった例も少なくない。そこで、音楽に子供を親しませるためには、従来の直接的な手法に加えて、遊ぶことを通して間接的に学ばせる方法も有効であると考えられる。例えば、絵を描くと音楽が流れる遊具を与え、子供が日常で遊ぶ中で、音楽と接触する機会を増やし親しませるというものが挙げ

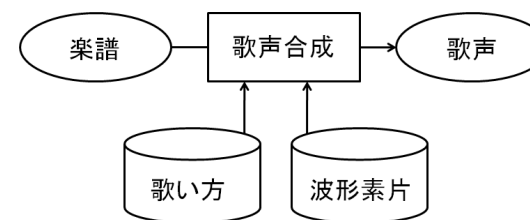


図1 歌声合成の流れ

Fig. 1 General overview of singing synthesis.

られる。本研究では、歌詞を読み上げた音声、すなわちシステムのユーザの声そのものを使って、オンラインで歌声を合成する手法を提案する。本手法は、ユーザの音声を音素ごとに切り分け、楽譜の音符に対応させた後、ユーザの描く線画によって歌い方を付与する機能を有する。ユーザは、線画を様々に工夫する遊びを通して、歌い方を学ぶというよりは、上手に音を鳴らす線画の描き方を探して遊んでいるうちに、自然に歌に親しむことができる。

2. 歌声合成技術の現状

2.1 歌声合成技術

歌声合成とは、歌い方や波形素片を指定された楽譜の通りに編集することで、歌声を作り出す技術である。図1に歌声合成の流れを示す。この歌声合成技術によって歌わずして、歌声を作ることが可能となった。これにより音楽を不得意とする人や、幼児を音楽に親しませるための道具として活用することが可能になっている。

2.2 従来の歌声合成手法

歌声合成手法のベースとなる音声合成技術には、パラメータ合成方式と波形接続方式がある。

パラメータ合成方式とは、音声を構成する音素や音節を与えて、音声を合成する方式である。パラメータ合成方式には、パラメータをコントロールするだけで合成できるという利点と、ユーザの声を使って合成することが出来ないという欠点がある²⁾。

波形接続方式とは、音素波形と目標とする抑揚を与えて、音声を合成する方式である。波形接続方式は、実際の音声を用いるために高い自然性を有する利点がある³⁾⁴⁾⁵⁾。波形接続方式の代表例は、YAMAHA VOCALOID である⁶⁾⁷⁾⁸⁾。VOCALOID の歌声合成技術では、音符はピアノロールスタイルで入力し、歌詞はピアノロールの音符の中に入力する。その

^{†1} 東京工芸大学工学部メディア画像学科

Dept. of Media and Image Technology, Tokyo Polytechnic University

表 1 VOCALOID のキャラクターと対応する発話者
Table 1 Correspondent speakers with the VOCALOID's characters

キャラクター	発話者
MEIKO	拝郷メイコ (歌手)
KAITO	風雅なおと (歌手)
初音ミク	藤田咲 (声優)

後、入力した音階に対して、ビブラートやフェード等の演奏表情付けを行い、オフラインで収録された歌手や声優 (表 1) の音声波形を接続し、歌声を作成する。

2.3 現状の問題点

2.2 節で述べたように、合成に用いられる音声波形の収録はオフラインで行われており、オンラインで収録した音声を用いての合成は行われていない。そのため、子供に絵を描くと音楽が流れる遊具を与え、日常で遊ぶ中で、音楽と接触する機会を増やし、音楽と親しませるといった手軽な使用ができない。

本研究では録音編集方式による歌声合成システムにおいて、ユーザ自身の声を波形素片として用い、歌い方を編集するインタフェースを有することで、オンラインで歌声を合成することのできる手法を提案することを目的とする。

3. 歌声規則と自動合成規則

3.1 歌声規則

人は歌を歌う際、様々な規則に基づいて歌っている。規則を守らないと、歌を歌うことが困難であったり、歌として成り立たないためである。例として、息つきがある。また、自明ではあるが楽譜通りでなければならないということも規則の一つである。また、歌の歌詞に、無声子音と母音の組み合わせがある場合には、そのため楽曲によっては、音符の始まりで無声子音を歌い始めると、母音の始まりが音符の始まりからずれてしまい、楽譜に指示されたりリズムを壊してしまう。そこで、歌詞の始めが無声子音と母音からなる場合、適当な長さをさかのぼり無声子音を開始し、音符の先頭の時刻から母音が開始するように空白を開ける (子音の先取りと呼ぶ) 必要がある。

3.2 自動合成規則

歌声合成時には、図 2 に示すように、矩形領域の始めに無声子音が入る適当な長さの空白を作り、母音と音符の開始の時刻を合わせることによって、子音の先取りを実現する。



図 2 無声子音の先取り (先頭に無声子音の区間)
Fig. 2 The method of singing unvoiced consonants.

4. 提案するオンライン歌声合成システム

4.1 提案法の概要

図 3 に提案手法の流れを示す。本手法では、ユーザが歌の歌詞を朗読した音声波形を音素単位に分割した音声素片を合成単位とし、入力された楽譜情報の音符 (音長及び音高) に対応するよう、PSOLA 法を用いて繋ぎ合わせることで歌声を合成する。さらに、手書きの図形を歌い方 (ビブラート) に変換するインタフェースと組合せる。

図 4 に本システムのスクリーンショットを示す。楽譜情報入力部において、MusicXML によって記述された楽譜ファイルを指定するか、あるいは、プログラムヘッダに与えた楽譜情報を読み込み、歌詞の各音素を、対応する音符の音価を示す矩形に配置すると同時に、音高軌跡を表示する。歌声合成時には、マイクから歌詞を朗読した音声を録音すると同時に、音声波形編集部にその波形を表示する。歌い方の指定をする際には、演奏表情指定部において自由曲線を描き、その軌跡をビブラートに変換し、選択した音符に付与する。

4.2 楽譜情報入力

まず楽曲の楽譜データが読み込まれると、矩形が表示される。この矩形領域は一マスが音符一つに対応しており、マスの横幅は音符の四分音符の数に対応している。また、音高波形も表示される。縦軸は音階、横軸は四分音符の数を表しており、前述矩形領域と対応している。これらの矩形領域と音高波形によって、音符の数、音階、音の長さを表している。図 5 に楽譜情報を読み込み表示した例を示す。

4.3 朗読音声入力

ユーザの声を波形素片として用いるため、音声の録音を行う。ユーザが歌詞を朗読した音声を録音すると音声波形と音素境界が表示される。その後、朗読音声を録音する際に生じる、始めと終わりの空白部分を手動で切り取り、音声波形部分だけを表示する (図 6)。また、

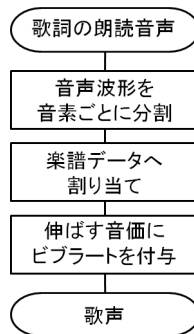


図3 オンライン歌声合成の流れ

Fig. 3 A schematic overview of the proposed method of online singing synthesis.

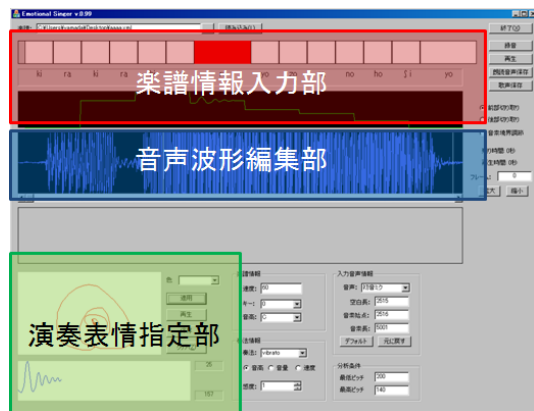


図4 提案するオンライン歌声合成システムの画面
Fig. 4 The screenshot of the proposed system.

音素ごとの波形素片を取得するため、音素境界を調整し音声波形を分割する(図7)。

4.4 朗読音声素片からの歌声合成

朗読音声素片から歌声を合成するには、2.2節で述べた波形接続方式を用いる。波形接続方式の代表例には、PSOLA法がある⁹⁾。図8にPSOLA法の流れを示す。

音符及び演奏情報から音高軌跡を取得し、波形データベースから、合成に用いる音声を選

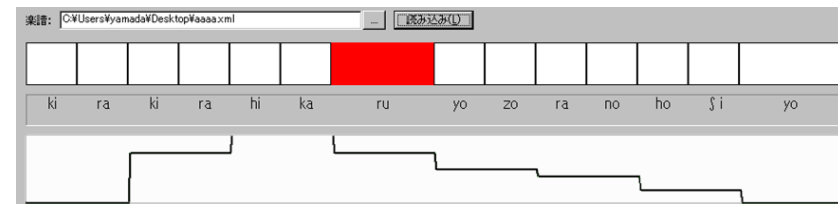


図5 楽譜情報の表示画面

Fig. 5 Displaying interface of music information.

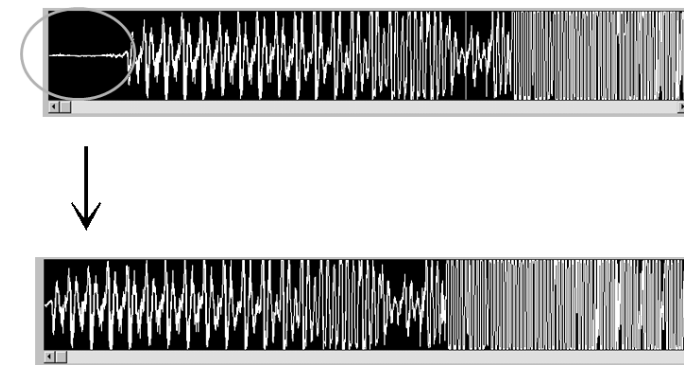


図6 朗読音声先頭の空白部分の切り取り

Fig. 6 Truncation of the preceding pose of the reading voice.

択し、その音声波形の先頭からピッチ周期を抽出し、そのピッチ周期の2倍の長さの窓関数を乗じて、ピッチ波形素片を取得する。以降、ピッチ周期ごとに分析区間をシフトしながら、順にピッチ波形素片を取得していく。波形合成においては、隣り合う2つのピッチ波形素片を重畳させて接続する。この時、入力された音高軌跡に応じて、低い音高を合成する場合はピッチ波形素片の距離を大きく、高い場合は小さくする。以上の処理により、ピッチ変更された有声音区間を得る。

4.5 歌い方の編集

歌い方の編集は、合成に用いるピッチ軌跡を編集することによって行う。本研究では、歌い方としてビブラートを扱う。図9に示すように、ペンタブレットによって描かれた軌跡を、音高の増減に相当する軌跡に変換し、次に、現在選択されている音符の長さに正規化した後、音高軌跡に加算する。

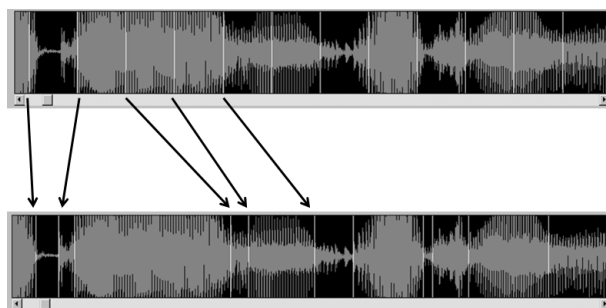


図 7 音素境界の調整

Fig.7 Manual adjustment of the phoneme boundaries.

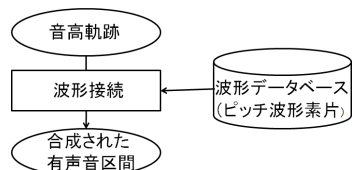


図 8 PSOLA 法の流れ

Fig.8 Schematic overview of PSOLA.

4.6 実装実験

対象とする楽曲は、きらきら星の最初の 1 フレーズとした。また楽譜データの作成には楽譜制作ソフト Finale 2010¹⁰⁾ を用いた。図 10 に用いた楽譜データを示す。

5. 結 論

本研究では、ユーザ自身の声を波形素片として用い、歌い方を編集するインタフェースを有する、オンラインで歌声を合成することのできる手法を提案した。

今後は、対象とする楽曲に関して、現在の 1 フレーズのみから、曲中を走査しながら 1 曲すべてに渡って歌声を合成できるようにする。また、付加できる演奏表情の種類に関して、ビブラートに加えて他の表情を増やす。

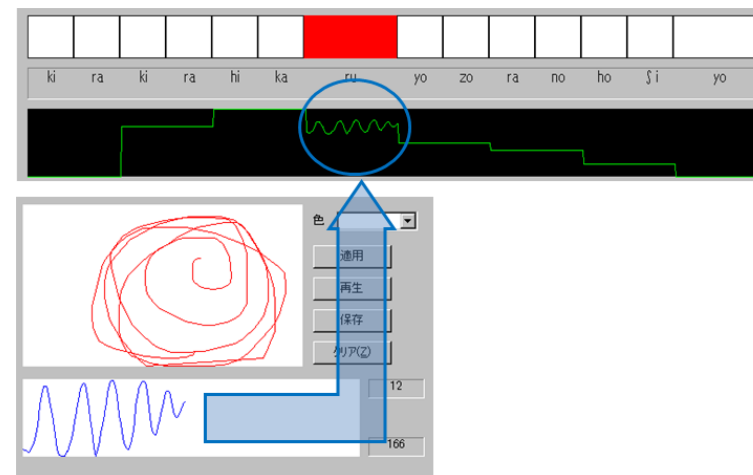


図 9 歌い方の編集例

Fig.9 An example of editing of singing method.

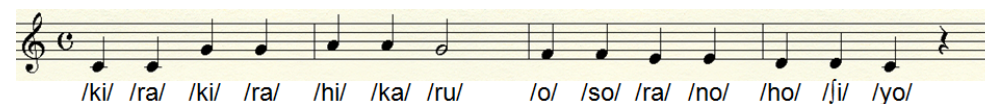


図 10 実装実験に用いた楽曲（「きらきら星」の一部）

Fig.10 Music used in the initial implementation of the proposed system.

参 考 文 献

- 1) 全国リトミック音楽教育研究会（編）：ダルクローズシステムによるリトミック指導 (1)～(3), 全音楽譜出版社 (2005).
- 2) 酒向慎司, 宮島千代美, 徳田恵一, 北村 正：隠れマルコフモデルに基づいた歌声合成システム (特集・音楽情報科学), 情報処理学会論文誌, Vol.45, No.3, pp.719-727 (2004).
- 3) 吉田博子, 寺田泰宏, 蓑輪利光, 新居康彦：PSOLA におけるピッチ波形抽出窓位置の検討, 日本音響学会研究発表会講演論文集, Vol.1995, No.1, pp.329-330 (1995).
- 4) 吉田由紀, 中嶋信弥：歌声合成システム, 電子情報通信学会技術研究報告. SP, 音声, Vol.97, No.560, pp.45-46 (1998).
- 5) 森山 剛, 小沢慎治：好みの歌唱様式による歌詞朗読音声からの歌唱合成 (スペシャルセッション・歌情報処理 2), 情報処理学会研究報告. SLP, 音声言語情報処理, Vol.2008, No.12, pp.33-38 (2008).
- 6) 剣持秀紀, 大下隼人：歌声合成システム VOCALOID-現状と課題, 情報処理学会研究報告. MUS,[音楽情報科学], Vol.2008, No.12, pp.51-56 (2008).
- 7) クリプトン：VOCALOID2 特集, <http://www.crypton.co.jp/mp/pages/prod/vocaloid/>.
- 8) 佐々木渉：仮想楽器をリアルにする「未来(ミク)の記号」と、VOCALOID で注目される「人の形」「声の形」について, 情報処理学会研究報告. HCI, ヒューマンコンピュータインタラクション研究会報告, Vol.2008, No.50, pp.57-60 (2008).
- 9) 峯松信明, 中川聖一：PSOLA 分析合成に基づく F_0 変換音声の品質向上に関する実験的検討, 電子情報通信学会論文誌. D-II, 情報・システム, II-パターン処理, Vol.J83-D-II, No.7, pp.1590-1599 (2000).
- 10) 株式会社イーフロンティア：Finale 2010.