

# 中国語での話声から歌声への変換

久保一志

諏訪東京理科大学大学院修士学生

江原暉将

諏訪東京理科大学

## 1 はじめに

前回の報告[久保 他]<sup>1</sup>で、日本語歌詞を中国語に翻訳し、それを日本語の自動歌唱ソフト<sup>1</sup>を用いて中国語でパソコンに歌わせた。しかし、中国語話者(本大学に在籍している留学生)に聞いてもらったところネイティブの発音ではないので聞き取りにくいという評価を得た。そこで、今回は、中国語の翻訳ソフトによる合成音声(話声)及び留学生の肉声(話声)を用いて、それを歌声に変換して歌わせることとした。歌声への変換には、音声分析合成変換ソフトであるSTRAIGHT[河原]を用いた。STRAIGHTで基本周波数と音声の長さを歌声に変換し中国語の発音で歌わせた。変換に使用した曲は日本の童謡である「春が来た」[岩波文庫]である。

話声を歌声に変換する先行研究としては、[齋藤 他]がある。これは、歌声の基本周波数(F0)変化を生成するためのF0制御モデルと、話声のスペクトル形状を変形するためのスペクトル制御モデル、リズムに基づいた音韻長制御モデルの三つのモデルによって話声から歌声への変換を可能にした。この研究は日本語だけによるものである。本研究と齋藤他との関係はSTRAIGHTを用いて分析し、話声の基本周波数を楽譜のメロディーに変換する点では類似したものになっている。また、齋藤他の研究は話声データと基本周波数の制御の手順は直接繋がっておらずスペクトル制御を経由している。本研究の特徴は基本周波数または音声の長さを制御する際、中国語音声合成ソフトの出力音声データ及び留学生の肉声音声データの話声の長さや基本周波数を楽譜に沿って直接制御することである。また本研究では音声データに中国語を用い、雑音除去フィルタとして帯域通過フィルタを使用していることも特徴である。

## 2 システム構成

システムとしては、(1) 中国語音声合成ソフト、(2) 話し声をパソコンで録音した際の雑音を除去する雑音除去ソフト、(3) 音声を切り離したりつなげたりする音声編集ソフト、(4) 音

声分析合成変換ソフトがある。

### 2.1 中国語音声合成ソフト

今回は、中国語音声合成ソフトとして市販の中国語翻訳ソフト(Chinese Writer)[高電社]の音声合成機能を利用した。

### 2.2 雑音除去ソフト

中国語音声合成ソフト及び留学生の音声をパソコンに録音すると雑音が入ってしまう。それを改善するためにフィルタを用いて雑音を除去した。今回雑音除去に使用したフィルタは帯域通過フィルタである。

### 2.3 音声編集ソフト

音声分析合成変換ソフトでは、長い音声の処理はできないので編集ソフトで2小節もしくは1つの単語ごとに区切る。さらに音声分析合成変換ソフトで歌声に変換し、このソフトでつなぎ合わせ曲にする。音声合成変換ソフトで音声の長さが分からなければここで測定する。

### 2.4 音声分析合成変換ソフト

音声分析合成変換ソフトSTRAIGHT[河原]を用いて音声の基本周波数、音声の長さなどを求めそこから春が来たに沿った歌声の周波数や声の長さに変換する。

## 3 システムの処理の流れ

本システムは次のような流れで処理されるが中国語音声合成ソフトの出力音声データと留学生の肉声音声データとでは処理の流れが違っているので両方の処理の流れを示す。

中国語音声合成ソフトの出力音声データの場合

- (1) 中国語音声合成ソフトの朗読機能から話声を録音する
- (2) 録音した音声を雑音処理ソフトを用いて雑音除去する
- (3) 雑音除去した音声を2小節ごとに区切る
- (4) 音声合成分析変換ソフトで音源情報分析

<sup>1</sup> 市販のVocaloid(クリプトン製)を用いた。

- (5) 音源情報分析の基本周波数データを用いて話声の長さを「春が来た」の音符の長さに変換する
- (6) 変換した音声の周波数を「春が来た」の楽譜に記載してある音符の高さに合わせる
- (7) 音声編集ソフトで音声データをつなげて一つの曲にする

留学生の肉声音声データの場合

- (1) 留学生に「春が来た」の歌詞を話すスピードを速、中、遅の3つずつ、声の高さを高、中、低3つずつ、計9つの条件で発話してもらい録音する
- (2) 録音した音声を雑音処理ソフトを用いて複数のバンド幅で雑音除去する
- (3) 雑音除去した音声を2小節ごとに区切る
- (4) 音声分析合成変換ソフトで合成音に変換する
- (5) 合成音の中で品質の良いものを選びつなげる
- (6) 話声の長さとおさを「春が来た」の音符の長さとおさに変換する
- (7) 音声編集ソフトで音声データをつなげて一つの曲にする

この流れを図1、図2に示す。

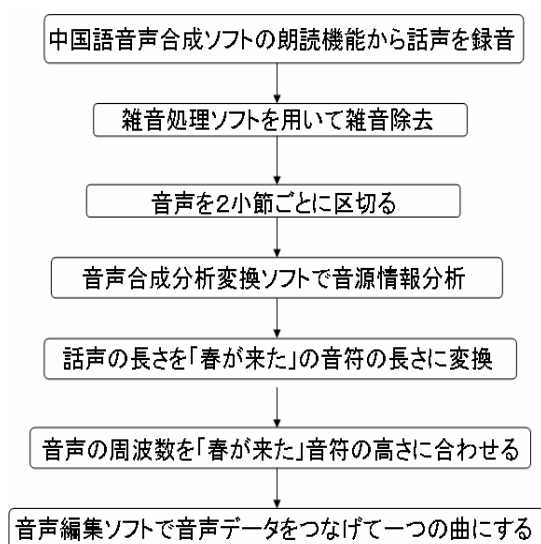


図1 システムの処理の流れ1 (中国語音声合成ソフトの出力音声データの場合)

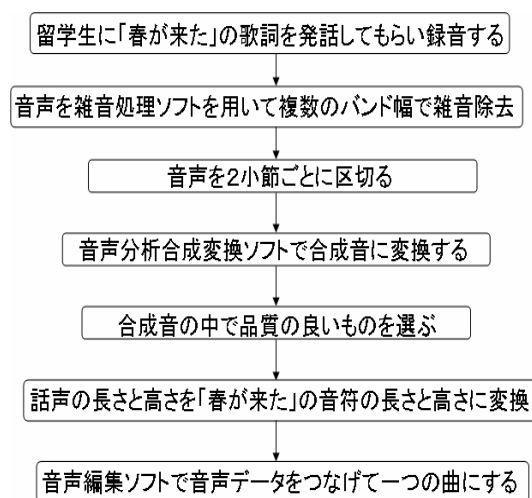


図2 システムの処理の流れ2 (留学生の肉声音声データの場合)

## 4 各システムの動作

### 4.1 中国語の歌詞作成について

日本語の歌詞を日中機械翻訳ソフト(Chinese Writer)で中国語の歌詞に翻訳する。このとき中国語の文字数のほうが音符の数よりも少ない。よって数を合わせるために単語の最後の音を伸ばしたときに聞こえる音の単語を追加する。

### 4.2 雑音除去について

この研究で用いられている雑音除去フィルタは帯域通過フィルタである。これは、本研究をやる前の予備実験で得られた結果に基づいて作成した。

### 4.3 音声分析合成変換ソフトの処理について

前節でも触れたが雑音処理をした音声を2小節ずつに分けて分析を行う。これは研究用パソコンのメモリー量によるためである。また、話声の長さを歌声の長さに変換する場合、音声劣化が起きる場合がある。これを改善するため単語の最後の音を伸ばしたときに聞こえる母音を加える。ただし、これはChinese Writerの音声データの場合に限る。中国語音声合成ソフトの出力音声データの場合、基本周波数の探索範囲を初期値(上限値: 800Hz、下限値: 40Hz)で基本周波数の分析が可能

だが留学生の肉声音声データの場合、F0 探索の範囲がシステムの初期値(40HZ から800HZ)だと分析ができないので、探索の範囲を上限值500Hz、下限値100Hzに変えて分析を行う。

#### 4.4 留学生の音声データについて

留学生の音声データを採る際に留学生の普段の発話のスピード及び高さを中とし、それよりも早い及び高い場合は高でその逆の場合を低とした。

### 5 各システムの動作結果

#### 5.1 歌詞作成について

以下に「春が来た」の中国語の歌詞を示す。

春天来了  
春天来了  
来到哪里了  
来到山了  
来到村落了  
也来到野地了

上記の歌詞だと音符の数と歌詞の数が合わない  
ので以下のように人手で編集する。

春天嗯来了  
春天嗯来了  
来了阿哪里  
来了阿到山  
来了阿村落  
也来了野地

また、Chinese Writerの音声を使用する場合、音を伸ばす箇所によって音声劣化が激しいので以下のように変更する。

春天嗯来了  
春天嗯来了  
来了阿哪里——  
来了阿到山  
来了阿村落  
也来了野地——

#### 5.2 話声の長さの制御の限界について

話声を歌声に変換する際、音符の長さに沿って話声の長さを変換する。その際、話声の長さを伸ばしたときの限界値を調べた。限界値の決め方は筆者(久保)1人による主観的な評価で決めた。限界の平均値は2.21倍、限界の最大値は2.75倍、限界の最小値は1.26倍であった(表1)。

表1 話声の長さの制御の限界

	最大値 (倍)	最小値 (倍)	平均値 (倍)
音声の伸 長	2.75	1.26	2.21

#### 5.3 話声の高さの制御の限界について

話声を歌声に変換する際、音符の高さに沿って話声の高さを変換する。その際、話声の高さを上げたり下げたりしたときのときの限界値を調べた。この測定も筆者1人による主観的な評価で限界値音を高くしたときの限界の平均値は2.14倍、限界の最大値は2.51倍、限界の最小値は1.14倍であった。音を低くしたときの限界の平均値は0.48倍、限界の最大値は0.42倍、限界の最小値は0.58倍であった(表2)。

表2 話声の高さの制御の限界

	最大値 (倍)	最小値 (倍)	平均値(倍)
高さを上 げたとき	2.51	1.14	2.14
高さを下 げたとき	0.42	0.58	0.48

#### 5.4 中国語音声合成ソフトの音声データを用いた歌声への変換

中国語音声合成ソフトの音声データを用いて話声から歌声に変換し、それを聞いてみると音声の劣化が見られ、留学生に聞かせたところ聞き取りにくいという結果をもらった。

## 5.5 留学生の音声データを用いた歌声への変換

合成音声による歌声の性能が不満足であったので、留学生の肉声音声データを用いて話声から歌声に変換した。それを聞いてみると中国語音声合成ソフトを用いたときよりもさらに音声が劣化していた。その理由については、次章で考察する。

## 6 終わりに

今回の研究で変換した中国語歌声音声は、メロディ・テンポとも、「春が来た」の原曲のそれらを保っており、一応の結果は得られた。しかし、変換された歌声の性能は不満足なものであり、課題が残った。以下、それに関して考察する。

合成音声の場合の課題として音声のトーンの劣化に対する対策がないということである。テンポの場合は各システムの処理で述べた単語の最後の音を伸ばしたときに聞こえる母音を加える方法が使える。しかし、トーンの場合は中国語音声合成ソフトで変換しても劣化が見られるため劣化を防ぐのが困難となった。これは、合成音声の品質に問題があると思われる。合成音声の標準化周波数がCD音声のよりも半分以下なのでそれが品質に影響を与えているとも思われる。

肉声の場合の課題は、合成音声で処理したときよりも音声の劣化がひどいことである。この要因として、話声音声の録音性能に問題があるかもしれない。この問題を解決することによって歌声の音声品質も上がる可能性がある。

また、歌声に制御する前の話声の肉声のデータ採取の際、声のスピード及び高さの条件をどのように設定すればよいかというのが挙げられる。本研究では、留学生に話すスピードを速、中、遅の3つずつ、声の高さを高、中、低3つずつ、計9つの条件で発話してもらい話声の肉声データを採取している。ここで、中というのが留学生の日常会話で話す声のスピード及び高さである。これを基準として、条件に沿って発話をしてもらった際に、明確な数値での条件の設定ではなく感覚的な条件で行ったが日常で話す話声とあまり変わりがないと感じた。肉声のデータ採取の条件として重要なのは話声のスピード及び高さを設定するときその人の通常の

話声からどの程度なら話声でどの程度なら話声ではなくなってしまうかという境目を設定することである。極端な例として、音符の長さと同様くらいにゆっくり発話したり、楽譜に記載されている音階の範囲で高さを変えて発話すれば歌声としての音声品質は向上するであろう。しかし、それは話声ではなく歌声になっていると考えることができる。従って、話声の肉声データを録音する際に話声と歌声の境界線を予め設定することが重要であると考え、今後この境界線がどこにあるのかを求める必要がある。

さらに、この研究では周波数を音符に合わせているため中国語のアクセントである四声を無視してしまっている。このことが歌声の品質を低下している要因になっている可能性もあるため四声を考慮に入れた周波数変換をした方がよいと思われる。

最後に話声から歌声に変換するためにプログラムをもっと誰でも使えるよう改良を加えることも課題である。さらにビブラートや強弱などの飾りつけが簡単に出来るツールもつけ加える必要があると考えている。

## 謝辞

本実験のために音声分析合成変換ソフトの使用を許可してくれた和歌山大学の河原英紀教授及び中国語での発声に協力してくれた本大学経営情報学部留学生の張玉さんに感謝する。

## 参考文献

- [岩波文庫] 堀内敬三、井上武士：日本唱歌集 岩波書店
- [河原 他] Hideki Kawahara, Ikuyo Masuda-Katsuse and Alain de Cheveigne: Restructuring speech representation using a pitch-adaptive time frequency smoothing and an instantaneous frequency-based F0 extraction: Possible role of a repetitive structure sound, *Speech Communication*, 27, pp.187-207, 1999.
- [久保 他] 久保一志、江原暉将：日本語歌詞を中国語に翻訳しパソコンに歌わせる、言語処理学会第12回年次大会論文集, PP.715-718, 2006.
- [高電社] 高電社：Chinese Writer8
- [齋藤 他] 齋藤毅、辻直也、鶴木裕史、赤木正人：歌声らしさに影響を与える音響を与える音響的特徴を考慮した話声からの歌声合成法、電子情報通信学会 応用音響/日本音響学会 聴覚研究会, EA-2004-30, 2004.