

歌声合成システム

吉田 由紀 中嶋 信弥

NTT ヒューマンインタフェース研究所

〒 239 神奈川県横須賀市光の丘 1-1

E-mail: yuki@nttspch.hil.ntt.co.jp nakajima@nttspch.hil.ntt.co.jp

あらし

規則音声合成技術を用いて、MIDI 対応の歌声合成システムを開発した。Standard MIDI File を利用して入力された楽譜から、各音符のピッチと時間長を自動的に抽出後、付与された歌詞を音素系列に分解、波形編集型音声合成法により歌声を合成、その後、MIDI と同期を取り BGM と合成された歌声を同時に出力する。本システムの特徴としては、単に音符に歌詞を当てはめて歌声を出力するだけでなく、歌詞入力時に、パワー調整や声質変換、ビブラートなどの情報を付加することができ、歌声を自由に演出できる点があげられる。さらに、歌詞を付与するトラックを複数用意することで、ハーモニーやデュエットなども実現可能とした。

キーワード 歌声合成, 音声合成, 波形合成, MIDI, DTM

Singing Voice Synthesis System

Yuki YOSHIDA and Shin'ya NAKAJIMA

NTT Human Interface Laboratories

1-1 Hikarinooka Yokosuka-shi Kanagawa 239 Japan

E-mail: yuki@nttspch.hil.ntt.co.jp nakajima@nttspch.hil.ntt.co.jp

Abstract

We developed a singing voice synthesis system correspond to MIDI using Text-to-Speech. The system first accepts a score from a standard MIDI file as input, then pitch and duration of each musical note are extracted automatically. Next phonological series are analyzed from given lyrics, and singing voice is synthesized by waveform speech synthesis. It is played together with BGM by synchronizing with MIDI format data. In this system, users can add other information such as power modulation, speech quality modification or vibrato, and can direct the singing voice. Furthermore, if lyrics is given at two or more trucks, harmony and duet can be also performed.

key words singing voice synthesis, speech synthesis, waveform speech synthesis, MIDI, DTM

1 はじめに

最近、DTMの普及により、誰でも手軽に音楽を楽しむことができるようになってきた[1]。しかし、MIDI音源には肉声がなく、楽器の演奏を演出することは可能であるが、歌に関しては、歌手あるいは作曲家自身といった人間が一度発声しなければ、扱うことはできない。

一方、規則音声合成の研究は盛んに行なわれ、テキストからの音声合成システムも多数開発されている。これらのシステムでよく利用されている波形編集型音声合成方式[2]は、音素系列、時間長、ピッチの3つのパラメータを与えれば、容易に音声合成することが可能である。

そこで今回、この二者を結び付けて、MIDIに対応した歌声合成システムを開発したので報告する。これまでも日本語の歌声合成システムは報告されているが[3]、本システムはMIDIを意識したソフトウェアで、従来の歌声合成システムとは多少異なる。

2 歌声合成システム

2.1 構成

図1に本システムの構成を示す。まず、ユーザがStandard MIDI File (SMF)を入力し、歌詞をつけたいトラックを指定すると、システムが各音符の音程と時間長を自動抽出し、画面に表示する。続いて各音符にあてはめて入力された歌詞を音素系列に変換し、合成パラメータを生成する。そして高品質なテキストからの音声合成(TTS)システム[4]を利用して、歌声を合成し、MIDI演奏と同期を取って出力する。

2.2 特徴

本システムの特徴は以下である。

- 使いやすさを考慮したシステム
音符ごとに歌詞を入力することができ、一つの音符に対して最大4音節入力することが可能。また、各音素の時間長やパワー、音の高さを自由に調整できる。
- 付加情報機能で歌声を自由に操作
入力された歌詞に対し、男女変換、パワー調整、声質変換(声の太さを変える)等が可能で、ユーザが自由に歌声を演出できる。
- 複数トラックに歌詞付与可能
複数トラックに歌詞をつけることにより、デュエットやハーモニを実現することができる。

3 まとめ

TTSシステムを利用し、MIDIに対応できる歌声合成システムを開発した。付加情報機能により、パワー調整や声質変換などが自由にでき、ユーザの好みに応じた歌声を合成することが可能である。

4 今後の予定

ビブラートやこぶし、裏声など、歌声特有の現象を付加情報に加え実現していく。また、朗読と異なる歌声特有の合成方式について検討する。

謝辞 日頃活発な議論をしていただき、HI研音声部の皆様に感謝致します。また、本システムの主にユーザインタフェース部分の開発を担当された吉田萬寿男氏、試作先であるNTTインテリジェントテクノロジー株式会社の皆様に感謝いたします。

参考文献

- [1] 中島安貴彦：“MIDIバイブルI,”株式会社リットーミュージック, pp11 (1998).
- [2] 広川智久, 伊藤憲三, 佐藤大和：“音素単位波形編集に基づく規則合成システム,”音響学会講演論文集, 1-6-15, pp219-220, (1991-10).
- [3] 山本篤志, 松本達郎, 片江伸之, 木村晋太：“日本語テキスト音声合成を利用した歌唱合成システム,”情報処理学会第49回全国大会, 5G-4, (1994).
- [4] Kazuo HAKODA, Tomohisa HIROKAWA, Hajime TSUKADA, Yuki YOSHIDA, and Hideyuki MIZUNO：“Japanese Text-to-Speech Software Based on Waveform Concatenation Method,” Proceeding of AVIOS'95, pp65-72, (1995-09).

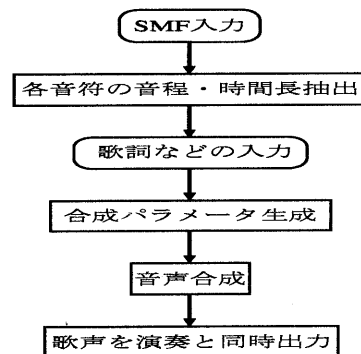


図1: 歌声合成システム構成図