

歌声合成システム : *CyberSingers*

吉田 由紀(*) 中畷 信弥(**)

(*) NTT インテリジェントテクノロジー株式会社 〒231-0032 神奈川県横浜市中区不老町 2-9-1

(**) NTT ヒューマンインタフェース研究所 〒239-0847 神奈川県横須賀市光の丘 1-1

alice@ntt-it.co.jp nakajima@nttspch.hil.ntt.co.jp

あらかし 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はなく、歌詞入力時に、パワー調整や声質変換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。

Singing Voice Synthesis System: *CyberSingers*

Yuki YOSHIDA(*) and Shin'ya NAKAJIMA(**)

(*) NTT Intelligent Technology Co., Ltd.

Wise Building 2-9-1, Furo-cho, Naka-ku, Yokohama, Kanagawa, 231-0032, Japan

(**) NTT Human Interface Laboratories

1-1 Hikari-no-oka, Yokosuka, Kanagawa, 239-0847, Japan

alice@ntt-it.co.jp nakajima@nttspch.hil.ntt.co.jp

Abstract We developed a singing voice synthesis system, *CyberSingers*, based on our high-quality text-to-speech synthesis (TTS) technology. Given a standard Musical Instrument Digital Interface (MIDI) file as input, the pitch and duration of each musical note in the input file are extracted automatically. The phoneme series of the typed-in lyrics are passed to the *CyberSingers* TTS module to synthesize a singing voice. After synthesizing vocal parts of a score, *CyberSingers* generates an intermediate data that includes timing information. By referring to this timing information, *CyberSingers* realizes precise synchronization between singing and instrument playing performed by a MIDI player. *CyberSingers* supports several options such as power modulation, a voice changer function, and multi-vocal feature. *CyberSingers* adds a new function to desktop-music technology.

1. はじめに

ここ数年、Desktop Music (DTM)が普及し、音楽を単に聴くだけではなく、作曲や編曲、演出など、誰でも手軽に楽しむことができるようになってきた[1]。しかし、MIDIには肉声の音源がなく、あくまでも楽器演奏の演出は可能であるが、歌に関しては、歌手あるいは作曲者自身といった人間が一度発声しなければ、DTMで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。実際、NTTでも1995年に、波形編集型音声合成方式[2]に基づく高品質テキスト音声合成システム[3]を開発した。この音声合成方式は、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ことから、グラフィカルユーザインタフェース(GUI)を利用して音声合成のパラメータを個別に操作し、音声メッセージの作成だけでなく、会話音声等を合成できるツールも発表されている[4]。

歌声を合成する場合、必要なパラメータを抽出することは、比較的容易であることから、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[5]、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. システム構成

図1に本システムの構成を示す。まず、ユーザがStandard MIDI File (SMF)を入力する。MIDIフォーマットは、GM、GS、XGのどれでも可能である。歌詞をつけたいチャンネル(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。そしてテキスト音声合成(TTS)システム[3]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。

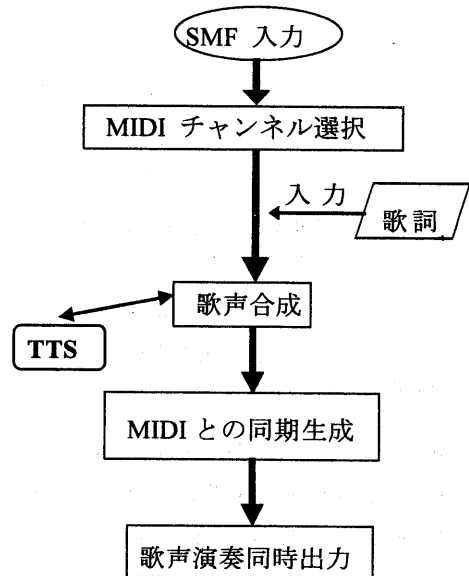


図1：歌声合成システム構成

3. 特徴

3.1 高品質TTSシステムを使用

CyberSingers は高品質テキスト音声合成システムFLUET(FLUEnt Talk)を利用している。FLUET は、波形編集型音声合成方式を採用し、前後の音韻環境を考慮した音素を合成単位とした日本語 TTS システムで、従来開発されたシステムと比べ、合成音声の品質は高いことが報告されている[3]。

3.2 歌声とMIDIデータとの同期

歌詞を付与した各チャンネルごとに歌声が合成されると、CyberSingers は時間情報を用いて歌声波形を MIDI コードに合わせ中間データを作成する。時間情報は、歌声と MIDI データを正確に同期させることが可能である。この 2 つの性質を利用すれば、次に述べるデュエットやハーモニが実現可能となる。

3.3 デュエットとハーモニ

CyberSingers には、デュエットとハーモニを実現するために、2 通りの方法が用意されている。

1 つは、複数の MIDI チャンネルに歌詞を付与し歌声を合成することで、もう 1 つは 1 チャンネル内にある異なったメロディーラインの歌声を合成することである。後者の方法についてもう少し説明する。

どの音符がどのメロディーラインかを確認するために、CyberSingers は、harmony group という項目が設定されている。ユーザは、各音符を A から D までの harmony group に割

り当てれば、和音のように同時に複数の歌声を合成することができる。ただし、現状では、harmony group の設定が 4 種類までであるため、それ以上の歌声が必要な場合は、初めに述べた複数の MIDI チャンネルに歌詞を付与する方法を取らなければならない。前節で説明した性質により、CyberSingers は、合成された複数の歌声と MIDI データの同期を取って音楽を演奏することが可能となる。

3.4 その他の機能

CyberSingers には声質変換機能があり、“かなり細い”から“かなり太い”まで、男女ごとに 7 通りの変換が可能である。これは、スペクトル領域の伸縮によって実現されている[6]。

また、CyberSingers では、どの音符に対しても、各音素ごとに、パワー、ピッチ、時間長の調整ができる。

4. 使用方法

CyberSingers の使用方法について説明する。

4.1 チャンネル選択

MidiCH ボタンをクリックすると、新しい窓が開き、歌詞を入力するチャンネルを選択すると、更に図 2 に示すメイン画面が現われる。縦軸はピッチ、横軸は時間長を示している。音符をクリックすればローマ字で歌詞を割り当てられる。歌詞が入力されるとチャンネルの色が黒から青に変わる。

4.2 歌詞入力

図 3 に示すように、音符をクリックすると歌詞入力画面が開く。ユーザは、1 音符あたり 4 音節以下の歌詞をローマ字で入力、2 音節以上の場合は、各時間長の割合も入力する。また、歌声はフレーズごとに合成されることから、フレーズの先頭となる音符は、該当項目をチェックしなければならない。

項目があり、男女変換、パワー調整、声質変換を可能としている。歌詞入力画面において、フレーズ付加情報をクリックすると、図 4 の画面が開き、上部で男女の選択、下部で声質の変換ができるようになっている。声質は、前述したように、*VarioVoice* [6]を使用して、“かなり細い”から“かなり太い”まで 7 種類用意されている。これらの機能を用いて、ユーザは自由に歌声を演出できる。

4.3 その他の機能

CyberSingers には、フレーズ付加情報という

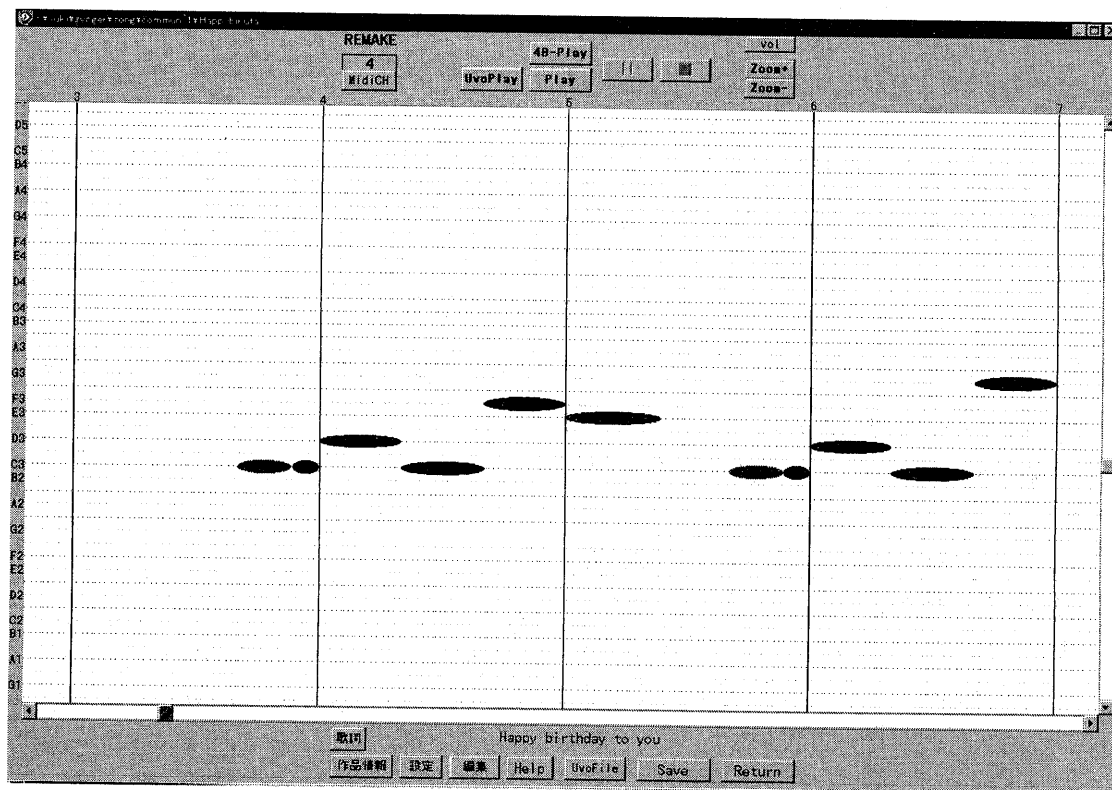


図 2 : CyberSingers メイン画面

5. まとめ

高品質日本語TTSシステムを用いた歌声合成システムCyberSingersを開発した。合成された歌声は、MIDIと同期を取って演奏と共に出力される。ユーザはパワー調整や声質変換といった付加情報機能を使って、簡単に歌声を演出することができる。

CyberSingersは新しいデスクトップミュージックツールとして、あるいはマルチメディアコンテンツのための作曲ツールとして利用可能である。

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、発表の機会を与えて下さった 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。

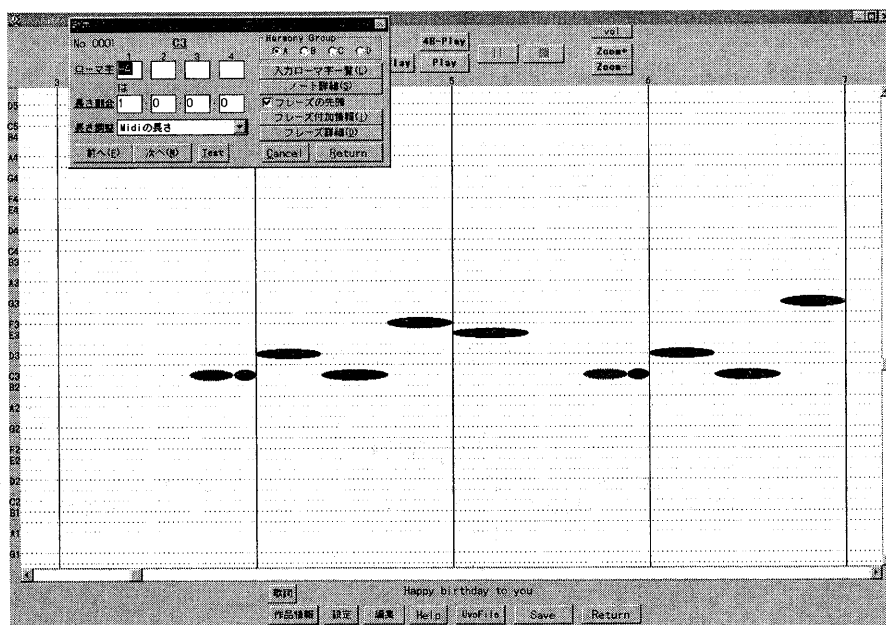


図3：歌詞入力画面

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] 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, pp. 65-72, (1995-

09).

[4] 阿部匡伸, 水野秀之, 中嶋信弥 : ``様々な音声表現を実現できる音声作成ツール - Speed97 -, " 情報処理学会研究報告, 97-SLP-17, pp67-72, (1997-07).

[5] 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).

[6] 阿部匡伸 : ``波形処理による声質変形装置 (Vario Voice)," 音響学会講演論文集, 3-7-6, pp269-270, (1997-03).

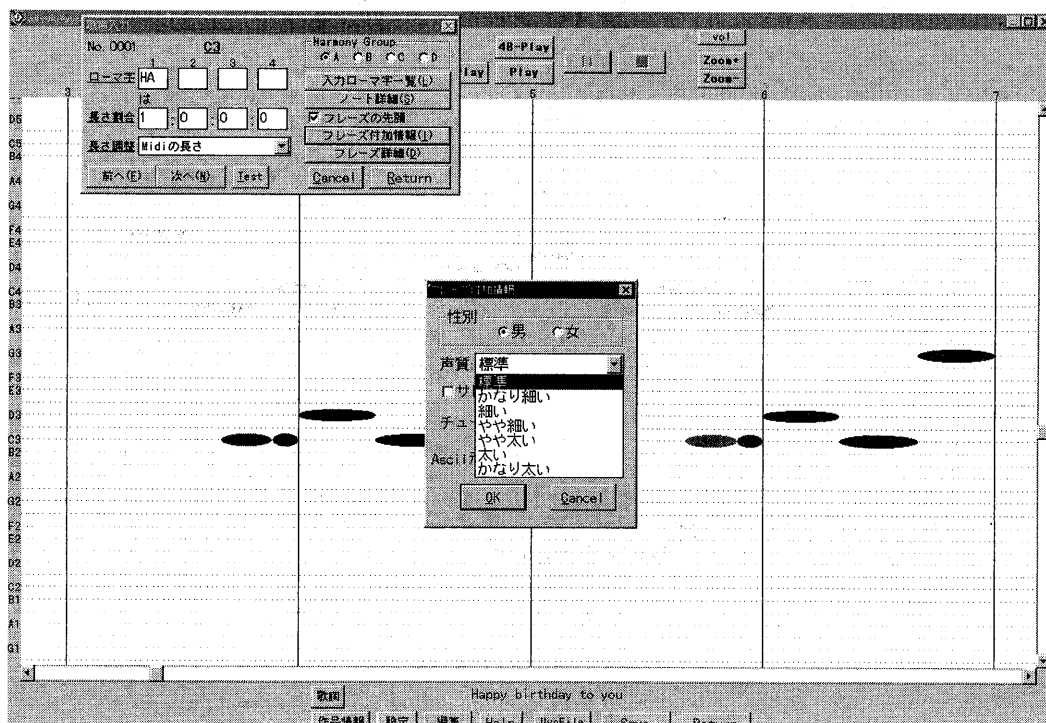


図 4: フレーズ付加情報