

オーディオ-MIDI 符号化技術「オート符」の改良と 楽器音によるテキスト音声合成システムへの応用

茂出木 敏雄†

大日本印刷株式会社 情報コミュニケーション研究開発センター†

1. はじめに

筆者らは与えられた音響信号に対して一般化調和解析を用いて平均律音階のスケールで高精度な周波数解析を行い、MIDI データ形式に自動変換する技術の開発を進めてきた[1]。本技術は「オート符@SA」という名称で、2001 年より財団法人デジタルコンテンツ協会のホームページより無償配布を進めており、主として採譜業務の支援等に活用いただいている[3]。本解析ツールは、特に和音解析精度が高く、音声信号に適用すると解析されたフォルマント成分が MIDI 形式に和音近似され、一般的な MIDI 音源を用いてボーカルが再現できるという特徴をもつ。

そこで、筆者らは種々の楽器音を模倣した声質で音声合成を実現する手法について研究を進めている。本稿では、電子楽器に対して音声で呼びかければ、楽器演奏によるオウム返しを実現し、電子楽器に文字列を与えれば、楽器演奏による音声合成を実現するシステムの試作例について紹介する。

2. 既提案の音響信号の MIDI 符号化ツールの概要と改良手法

図 1 中央の縦方向のフロー(1)~(5)は、筆者らが先に開発した MIDI 符号化処理の主要構成を示し[1]、左右の(6)~(9)は本稿で追加提案する改良手法を示す。詳細は文献[1][2]に譲り以下概略を記す。

(1)の処理で、与えられたソース音響信号より周波数解析対象のフレームを適応的に抽出する。(2)の処理で、一般化調和解析手法[1]に基づき平均律音階の半音(ノートナンバー)単位に非

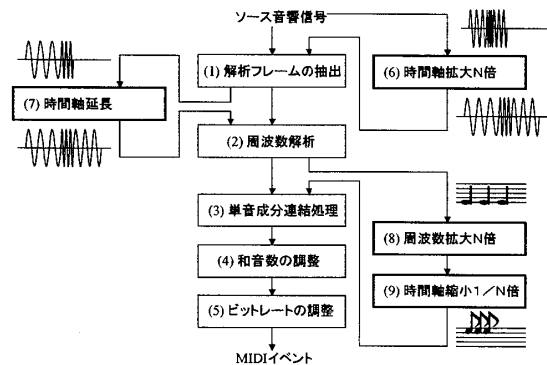


図 1 既提案の MIDI 符号化処理構成と追加構成

線形な周波数次元で周波数解析を行う。(3)の処理で、時間的に隣接する同一周波数の解析成分(単音成分)を連結し音符としてまとめ、MIDI イベント形式で符号化する。(4)(5)の処理で、標準的な MIDI 音源で再生可能な和音数とビットレートになるように、MIDI イベントデータを削減する。

左右に追加した処理は、時間分解能を向上させることを目的としており、あらかじめ、(6)の処理で、ソース音響信号に対して時間軸方向に2~4倍拡大して周波数解析を行う。ただし、これを適用すると低音部が解析不能になるため、(7)の補正処理を加える。この状態で(2)の周波数解析を行うと、周波数が低くなり時間軸が伸びているため、原音と同じ次元に戻す処理(8)と(9)を加える。以上の処理により音声信号など周波数変化が急峻な特徴を捉えることができる。

3. 提案する楽器音を用いた音声合成システム構想

図 2 は本稿で提案する楽器音を用いた日本語音声合成システムの構成を示す。図上段の水平方向のフローはテキスト音声合成のメインフローを示す。与えられたテキストを音節に分解し、更に各音節を1~3個の音素に分解し、時系列

Improvement of Audio-MIDI Encoding Technique "Auto-F" and Its Application to Text to Speech Synthesis System Based on Musical Tones

† Toshio Modegi, Media Technology Research Center, Dai Nippon Printing Co., Ltd. (Modegi-T@mail.dnp.co.jp)

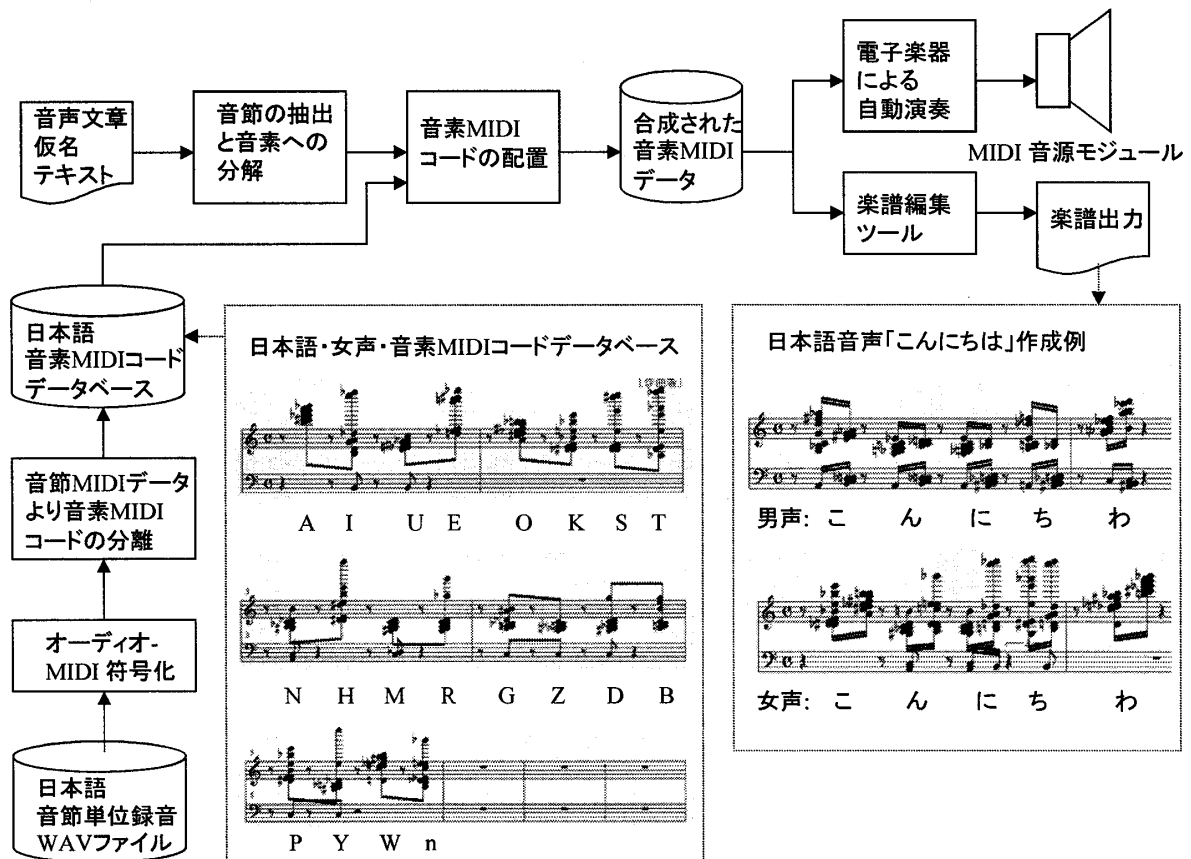


図2 提案する楽器音を用いた日本語音声合成システム構成

に配置されている各音素をあらかじめ定義されている8和音などで構成される MIDI データに順次展開し、時間軸方向に MIDI データを合成してゆく。

その中で参照している日本語音素 MIDI コード・データベースを構築するフローが同図左端の縦方向のフローである。あらかじめ日本語71音の音節別に録音を行い、前述した図1の MIDI 符号化ツールを用いて各音節 WAV ファイルを音節 MIDI データに変換する。次に、複数の音節 MIDI データどうしを掛け合わせて、各音節に共通に含まれる MIDI 音素データを分離する処理を行い、20種の音素 MIDI コードを作成する。例えば、「か」「き」「く」「け」「こ」という5つの音節データを掛け合わせて共通に含まれる「K」なる音素データを分離する。具体的な掛け合わせ方法は文献[2]に譲る。

4. おわりに

図2の提案構成では音節どうしの連結性や抑揚について考慮していない。MIDI データは音程を上下させるのが容易であるため、単語別に抑

揚パターンを定義した辞書データを今後構築すれば、単語の明瞭性は向上すると思われる。

本稿で紹介した「オート符」の現行版 ver. 2.6 については (財) デジタルコンテンツ協会のサイト[3]で 2001 年より無償配布を行っている。また、本稿で紹介した改良版 ver3.0 も無償配布するので、個別に問い合わせ頂きたい。

参考文献

- [1] 茂出木敏雄, "音響信号の平均律音階に基づく汎用解析ツール「オート符」の開発," 電気学会・電子情報システム部門誌, Vol. 123-C, No. 10, pp. 1768-1775, (October 2003).
- [2] 茂出木敏雄, "MIDI 符号化ツール「オート符」を用いた音素 MIDI コードの設計と楽器音による音声合成機能の実現," 平成 21 年電気学会・電子情報システム部門大会論文集, GS9-6, pp. 1274-1281, (September 2009).
- [3] 財団法人デジタルコンテンツ協会 d-CON Support, <http://www.dcaj.org/d-con/frame09.html> (「オート符」@SA, ver. 2.6) の無償配布元。