オーディオ-MIDI 符号化技術「オート符」を用いた電子楽器との音声によるインタラクションの実現

茂出木　敏雄†

Realization of Voice Interaction with Electronic Musical Instrument Using Audio-MIDI Encoding Technique “Auto-F”

TOSHIO MODEGI†

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。手始めに、71 種の日本語音節を録音した素材を用いて、デュレーションやベロシティを均一にした二連の単純な和音で表現した音節 MIDI データベースを構築し、カナテキスト入力により MIDI 音源で音声を再現できる MIDI データを合成するシステムの試作をした[2]。しかし、聴取可能なセリフについてはかなり限定され、より再生音声の明瞭性が求められている。

類似事例として、独自開発の自動ピアノによる英文読み上げを実現した Speaking Piano [3]が動画投稿サイトで発表されているが、原音はゆっくりとした子供の声を録音したもので、それでも字幕無しで聴取するのはかなり困難で、同様により明瞭性が求められる。

本稿では、MIDI 音源を用いて、より明瞭な音声再現を実現できるよう、先に提案した音響解析ツールに対して周波数解析における時間分解能を改善する手法について提案する。そして、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。

† 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
Media Technology Research Center, Dai Nippon Printing Co., Ltd

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。

はじめに、(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抽出するが、後続フレームへのシフト幅はゼロ交差解析法によりソース音響信号の周波数変動を大まかに検出しながら適応的に設定するようにしている。即ち、周波数変動が急峻な箇所（例えば音声の子音区間）ではシフト幅が細かくなるようにする。詳細は文献[2]に譲る。

続いて、(2)の処理で、一般化調和解析手法に基づき平均律音階の半音（ノートナンバー）単位に非線形な周波数次元で周波数解析を行うが、周波数が高くなるにつれ、半音間隔が粗くなるため、周波数ごとに半音間を微分音（副周波数）に分割して解析を行うようにしている。一般化調和解析手法のアルゴリズムは文献[2]に譲るが、離散フーリエ変換により得られた周波数スペクトルより単一のピークを抽出して、その成分を原音信号より削除して、再度離散フーリエ変換を実行して次のピークを抽出するという処理を抽出対象のピーク数分だけ繰り返し実行する。

続いて、(3)の処理で、時間的に隣接する同一の主周波数の解析成分（単音成分）を連結し音符としてまとめ、主周波数をノートナンバーにもち、連結成分の最大強度を 128 段階のベロシティで符号化し、ノートオン時刻とノートオフ時刻で構成される一対の MIDI

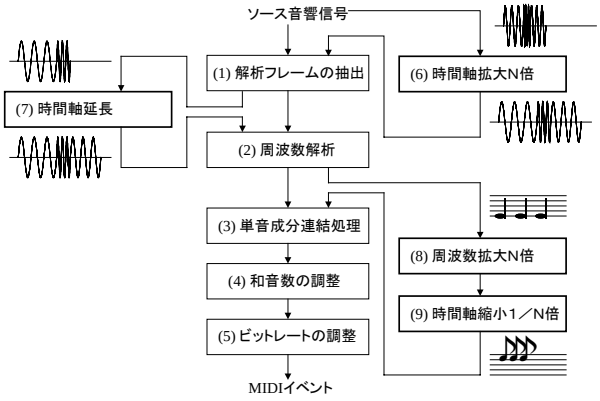

図1 既提案の MIDI 符号化処理構成と追加構成

イベント形式で符号化する。最後に、(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。この際、各 MIDI イベントデータには、ベロシティとデュレーション（ノートオン時刻からノートオフ時刻までの区間）情報をもたせており、これらの積が低い MIDI イベントデータを削減対象とするようにしている。(3)(4)(5)の処理の詳細は文献[2]に譲る。

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、全体的に１～２オクターブ低音側にシフトして周波数解析を行うようにしたものである。

(2)の周波数解析は固定長の解析フレームと MIDI ノートナンバーに対応した周期をもつの調和関数と相関演算を行うが、(6)の拡大処理に伴い、低音側では解析フレーム内に１周期分を収容できず算出不能になる。そこで、(7)の処理で、低音部の解析フレーム長を相関演算する調和関数の１周期分になるように延長する処理を行う。延長手法として、ソース音響信号より後続信号を抽出する方法をとると低音部の時間分解能が低下するため、解析フレーム内の音響信号が相関演算する調和関数の断片であると仮定して、後続の信号を推定する方法を提案する。

(2)の周波数解析では、(6)の拡大処理に伴い、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を(8)と(9)の処理で行う。即ち、解析された周波数成分の周波数を１～２オクターブ上げ、解析区間の時間軸を１／４～１／２に縮小する。本処理を追加すると、(3)の処理で得られる単位時間あたりの MIDI イベント数が２～４倍になるため、(5)の処理で従来と同等なビットレートになるように調整する。従って、本稿の改良により時間分解能が向上するが、符号化ビットレートは殆ど変化しない。

3. 提案する楽器音を用いた音声合成システム構想と日本語音節の MIDI モデル

前節で述べた方法により、音声を録音した音響信号を与え、時間分解能 1/1536[sec], 和音数 32 の標準設定で MIDI 符号化を実行すれば、MIDI 音源で音声を再生可能な MIDI データを得ることができる。例えば、GM 規格の MIDI 音源でプログラム番号(54, Voice-Ooh)を設定して再生すれば、より音声らしい再生が可能である。

これに対して、既存の音声合成ツールと同様にテキスト入力で音声再生用の MIDI データを出力可能な図２に示すような構想の実現性を検討した。前節で述べた MIDI 符号化ツールでは、和音数が多く、木目細かいデュレーションやベロシティなどの演奏指示情報が付与されているため、そのままではヒトが演奏可能になるような判読性のある五線譜に変換することは困難であった。

そこで、日本語 71 音の音節について、デュレーションやベロシティを均一にし、できる限り単純な和音で音声として聴取可能な表現方法を模索した[2]。

基本構成は、「あいうえお」の母音のみ単一の音素あるいは和音で表現し、それ以外の濁音、半濁音、撥音を含む子音類の音節は全て２つ以上の音素で構成されるものとし、二連または三連の和音で表現する。和音数は原則４以上で最大８和音以下にする。ベロシティとデュレーションは均一にし、通常の母音音素の標準区間長を 0.25[sec]にして、子音音節は第１和音を

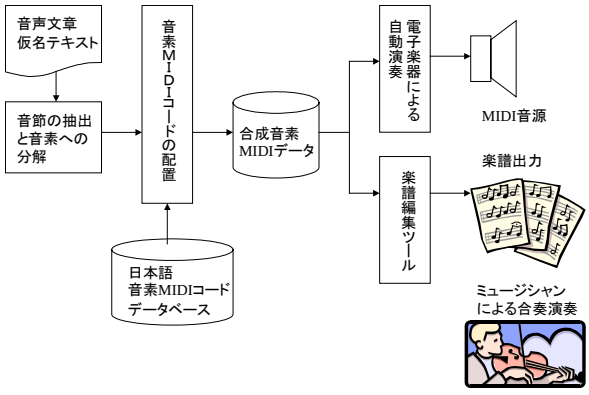

図2 楽器音を用いた音声合成システムの構想

1/3 区間長、第２和音を 2/3 区間長とする。その他の長音・促音・拗音の扱いについては文献[2]に譲る。

音声合成は、図２に示されるように、与えられた仮名テキストより前述した日本語音節を抽出し、更に各音節を１～３個の音素に分解し、日本語音素 MIDI コードのデータベースを参照し、所定の間隔で MIDI データを時間軸方向に結合して合成 MIDI データを作成すれば良い。その際、母音音素の間隔を変えれば話速を変更でき、母音音素の MIDI データを移調させれば、声の高さを変更でき、母音音素ごとに音楽のメロディーに合わせて移調させれば歌声にすることもできる。合成 MIDI データを MIDI 音源で自動演奏させるか、既存の楽譜編集ツールで五線譜に変換すれば、ヒトによる楽器演奏でも音声合成を実現できる。

4. 提案する日本語音素 MIDI コード設計ツールの機能

前節で述べた日本語音節の和音モデルに基づき音素 MIDI コードのデータベースを設計する手順を図３に記す。まず、濁音、半濁音、撥音を含む全 71 音節の男声、女声の録音音声素材を収集する。

続いて、前節で述べたオーディオの MIDI 符号化ツール「オート符®SA」の改良版を用い、高精細な MIDI コードに自動変換する。変換結果例として、同図(A)に「カキクケコ」の音節サンプルに対して適用した結果を示す。横軸は時間で縦軸は周波数で、各四角形が音符（ノートオンとノートオフ・イベントの対）を示し、横幅はデュレーション、縦幅はベロシティの表示も兼用している。

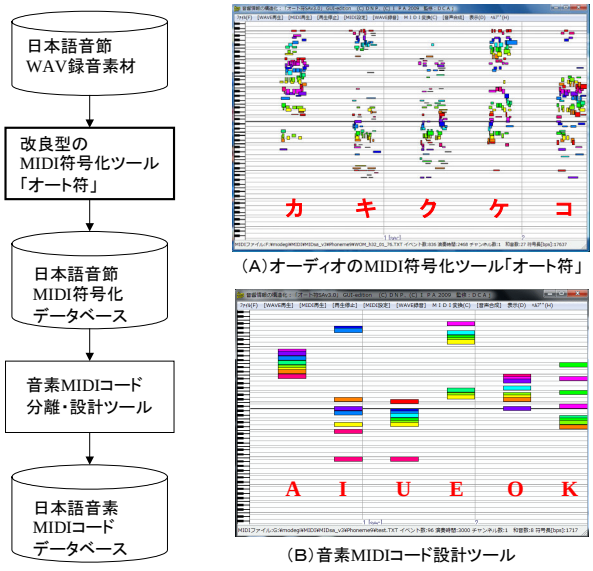

図3 日本語音素 MIDI 符号データベースの構築方法

表1 日本語７１音節より２０音素を分離する変換テーブル

		K	S	T	N	H	M	R	G	Z	D	B	P	Y	W
A	ア	カ-A	サ	タ	ナ	ハ	マ	ラ	ガ	ザ	ダ	バ	パ	ヤ	ワ
I	イ	キ-I	シ	チ	ニ	ヒ	ミ	リ	ギ	ジ	ヂ	ビ	ピ		
U	ウ	ク-U	ス	ツ	ヌ	フ	ム	ル	グ	ズ	ヅ	ブ	プ	ユ	
E	エ	ケ-E	セ	テ	ネ	ヘ	メ	レ	ゲ	ゼ	デ	ベ	ペ		
O	オ	コ-0	ソ	ト	ノ	ホ	モ	ロ	ゴ	ゾ	ド	ボ	ポ	ヨ	ヲ
n				ン											

まず、音節ごとに 32 重音以下の各種デュレーションおよびベロシティ情報をもつ和音に変換し、その結果が同図(A)である。続いて、各々を２つの音素成分に分離し、５つの母音音素「A,I,U,E,O」と共通する１つの子音音素「K」に分離し、各音素に対してデュレーションおよびベロシティが均一で８重音以下の単一な和音になるよう整形化を行った結果が同図(B)である。本例では全ての音素を単一の 8 和音に均一化し、ベロシティは全て 127 にする。

表１に MIDI データに自動変換された 71 種の日本語音節成分どうしを掛け合わせて、20 種の音素成分に変換する方法を示す。まず、母音音素「A」・・・「O」は表１の第２列目に単独で存在するが、精度を向上させるため、同行の子音１２音節を含めた１３音節成分の後半区間どうしのＡＮＤ演算で変換する方法をとるようにした。一方、表１の子音音素「K」・・・「W」は同列の５音節（ＹとＷは３・２音節）に共通して含まれるため、各々５音節の前半区間成分どうしのＡＮＤ演算で変換する。その際、各々に含まれている母音音素成分をあらかじめ削除した上でＡＮＤ演算を行う。そのため、母音音素成分は先に決定しておく必要がある。撥音の母音音素成分「ｎ」については「ン」音節単独で変換し、子音音素成分「N」と合成して撥音音節を合成するものとする。

具体的なＡＮＤ演算を行う方法は次の通りである。複数の MIDI データに自動変換された音節データより、表１に基づいて、共通の音素を含む要素を抽出し、各音節の前半または後半区間内で発音されているノートイベントのベロシティ値とデュレーション値との積をエネルギー値とし、ノートナンバーごとにエネルギー値の総和を求める。子音音素を決定する場合は、各々音節の前半区間に若干含まれている母音音素に対して

あらかじめ決定された母音音素のノートナンバーに対応する成分に 1/1000 などの小さな値を乗算して減衰させておき、母音音素と重複するノートナンバーが抽出されないようにする。続いて複数の音節データどうしで対応するノートナンバーのエネルギー総和値を乗算する。そして、得られたエネルギー総和値の乗算値が高い順にノートナンバーを指定和音数(例．8)だけ選択し、指定のベロシティ値(例．127)とデュレーション値(例．0.25sec)を与えた指定個数のノートイベントに自動変換する。

5. おわりに

前節で述べた方法を用いて、濁音、半濁音、撥音を含む全 71 音節の男声、女声の２セットの録音データに対して、時間分解能 1/768sec, 32 和音構成で改良型「オート符」ツールを適用し図３(A)で示されるような MIDI データに変換した。続いて、全 71 音節の男声、女声の２セットの 32 和音構成の MIDI データを用いて、全 20 音素の各々８和音構成で、図３(B)で示されるような音素 MIDI コードに変換し、男声と女声の２セットのデータベースを作成した。女声の MIDI 音素データベースの作成例を表２に示す。図中の各音名表記は MIDI ノートナンバー60 を C3 とし、全て８和音で構成され、強さ・長さは均一であるが、子音音節を合成する際は、子音音素である第１和音は短めに、母音音素である第２和音は長めに演奏することを想定している。

女声の「こんにちは」に対して、表２の音素 MIDI コードのデータベースを参照しながら再生可能な MIDI データを合成し、市販の楽譜編集ツールを用いて五線譜に自動変換した結果を図４に示す。同図に示されるように、比較的判読性のある五線譜に変換できることが確認でき、文献[2]の提案手法に比べ、個々の音節の明瞭性は向上し、GM 規格の MIDI 音源でプログラム番号(1, GrandPiano)の設定でも音声らしきメッセージは聴取でき、音色をプログラム番号(54, Voice-Ooh)に設定して再生すれば、より音声らしい再生が可能であることは確認できた。

ただし、現段階では単音音節を聴取するのは困難で、短い音節の既知の単語を聴取できるレベルである。今後は、音素に抑揚付けを行う手法を検討し、外国語音声への対応を含めて、音素 MIDI コードと抑揚パターンのデータベース構築を進める予定である。

本稿で紹介した「オート符」の現行版 v.er2.6 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(e-mail: Modegi-T@mail.dnp.co.jp)。

表2 音素 MIDI コードのデータベース作成事例（女声）

[A]	[I]	[U]	[E]	[O]
A#5	G#6	C4	A#6	B4
A5	G6	G#3	F#6	A#4
G5	C#4	G3	E6	A4
F5	A3	F#3	D#6	F#4
D#5	G3	E3	D6	D#4
D5	D3	D#3	F4	D4
C#5	B2	D3	D#4	C#4
B4	B1	B1	D4	A3

[K]	[S]	[T]	[N]	[H]
E5	G6	B6	B3	C7
A#4	F#6	G#6	C#3	E5
E4	G5	G#5	C3	B4
A#3	E4	E5	A#2	E4
F3	F3	E4	A2	D#4
E3	E3	E3	G#2	E3
D#3	D#3	D#3	C2	D#3
C#3	C#3	G#2	A1	C#3

[M]	[R]	[G]	[Z]	[D]
C#3	B5	B3	C3	G4
C3	G#4	A#3	A#2	C#3
A#2	C#3	C#3	A2	C3
A2	C3	A#2	G#2	A#2
G#2	A#2	A2	G2	G#2
F#2	A2	G#2	F#2	G2
C#2	G#2	G2	F2	F#2
C2	F#2	C2	C2	C2

[B]	[P]	[Y]	[W]	[n]
G4	E5	G6	A#4	A#5
E4	A#4	E4	A4	E4
B3	E4	C#3	G#4	D#4
C3	B3	B2	D#4	C4
A#2	F3	A#2	D4	D#3
A2	E3	A2	C#4	C#3
G#2	D#3	G#2	C4	C3
G2	C2	C2	D3	A2

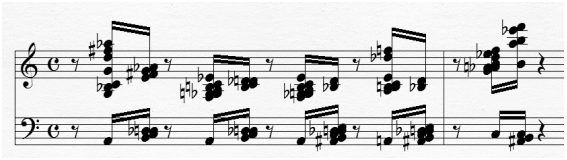

図4 日本語「こんにちは」MIDI 合成音の五線譜表現事例

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,”楽器音を用いた音声合成実現のための音素コードの設計機能と符号化ツール「オート符」への実装,” 情報処理学会・エンターテインメントコンピューティング研究報告, 2009-EC-12, Vol.2009, No.26, pp.35-42, (March 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.
4) Peter Ablinger (Composer), Speaking Piano: http://www.youtube.com/watch?v=muCPjK4nGY4&feature=player_embedded (October, 2009).