デ-02

オーディオ―MIDI 符号化技術「オート符」における音声符号化品質の改善

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[2]。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[5]。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。

しかし、差分 MIDI データを用いた品質評価手法[4]を適用すると、子音のような連続スペクトル特性をもつ音素に対しては、非常に多くの音を重ねないと忠実な再現ができず、再生可能な和音数に制限のある標準 MIDI 音源を用いて子音音素等を明瞭に表現することは困難であった。本発表では、ヒト聴覚系における帯域フィルタ特性[1]を考慮し、限られた重音数でも多彩な音素を表現可能にする改良手法を提案する。そして、既提案のカナテキスト入力により MIDI 楽器音を基本とした音声合成ツール[3]を用いて、再生品質の改善具合を実演する。

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のフローに記載の処理(1)～(9)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[4]、処理(10)が本稿で追加提案するものである。詳細は文献[2][3][4]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最小固定間隔で抽出する。対象とする音響信号がボーカルの場合、

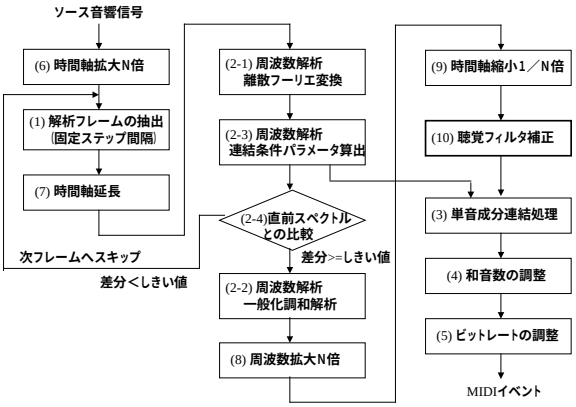

図１　既提案の MIDI 符号化処理構成と改良構成

時間分解能を向上させるため、(1)の前に、適宜(6)の処理を加え、ソース音響信号に対して時間軸方向に２～４倍拡大する。(6)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とがある。

(2-1)から(2-4)の処理は周波数解析で、一般化調和解析手法[2]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。このとき、周波数解析を２段階で行い、(2-1)のラフな周波数解析で、直前フレームとのスペクトル変化を調べ、これが緩やかであれば、(2-4)の処理で解析対象フレームをスキップさせ、結果的に(2-2)の高精度な周波数解析は可変間隔のフレームで解析を行うようにしている。また、後続の(3)の処理で必要な連結条件パラメータも本段階で取得する。

(6)(7)の処理を適用した場合は、原音と同じ次元に戻す処理(8)と(9)を加えた上で、次節で説明する聴覚フィルタ補正(10)を行う。その後、(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し、(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。

Improvement of voice coding quality for Audio MIDI Encoding Technique “Auto-F”

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

3. 提案する聴覚フィルタ補正手法

文献[1]によると、ヒトの聴覚系は理論的に 24 個の帯域フィルタで近似することができ、24 個の出力信号の割合で複数の周波数成分の弁別を行っていることが知られている。即ち、各帯域フィルタ内に含まれる周波数成分は同時刻では単一周波数成分しか認識できず、同一帯域幅に含まれる複数の周波数成分は時間差をもってビートとして認識される。そこで、周波数解析後の周波数成分より 32 和音など指定された個数の周波数成分を選択する際、各周波数帯域よりなるべく重複して選択されないように補正を行う。具体的には、各周波数帯域ごとに強度が局所的に最も大きい周波数を代表周波数とし、それ以外の信号成分を減衰させた上で、指定された個数の周波数成分を強度が大きい順に選択する。

表１は文献[1]記載のヒト聴覚の 24 個の帯域フィルタを MIDI ノートナンバーの次元に変換し再定義したものである。本表を用いて、図２の MIDI 符号化サンプル（DNP イメージソング、ボーカル曲 18 秒）に対して補正を行った結果を図３に示す。図中の着色された小さな矩形は音符を示し、横軸は時間で、横幅はノートオンからノートオフ区間を示す。縦軸は音高（ノートナンンバー）で縦方向の幅でベロシティも示している。低中域部に集中していた周波数成分が高域部に分散し、主としてボーカルの子音の再現性が向上することを確認できた。

4. おわりに

今後は表１に定義した帯域フィルタ特性に対し MIDI 応用に特化した見直しを行い、マスキングも考慮して更なる改善案を探ってゆきたい。

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[5]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。

表１　MIDI ノートナンバー次元の聴覚帯域フィルタ

帯域番号	1	2	3	4	5	6	7	8	9	10	11	12
下限音高	17	45	57	64	69	72	76	79	82	85	88	91
上限音高	44	56	63	68	71	75	78	81	84	87	90	92
帯域番号	13	14	15	16	17	18	19	20	21	22	23	24
下限音高	93	96	98	101	104	106	109	113	116	119	123	127
上限音高	95	97	100	103	105	108	112	115	118	122	126	130

参考文献

[1] 赤木正人,“聴覚フィルタとそのモデル”, 電子情報通信学会誌, Vol.77, No.9, pp.948-956, (September 1994).
[2] 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).
[3] 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,”電気学会・電子情報システム部門誌、Vol.130-C, No.7, pp.1159-1167, (July 2010).
[4] 茂出木敏雄,” MIDI 符号化ツール”オート符”を用いた各種音響信号処理における品質劣化の評価手法,” 平成 22 年電気学会・電子情報システム部門大会論文集, GS5-2, pp.1318-1325, (September 2010).
[5] 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)

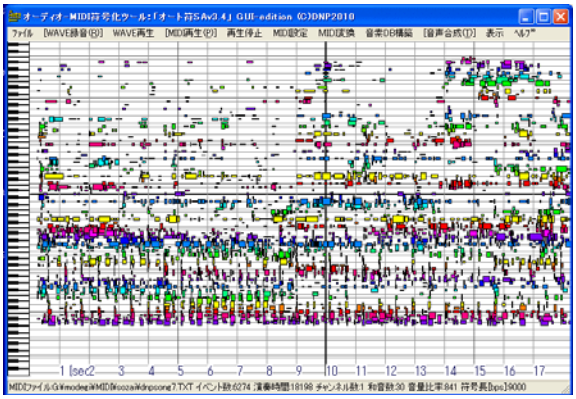

図２　従来法による MIDI 符号化結果（ＤＮＰイメージソング、１８秒）

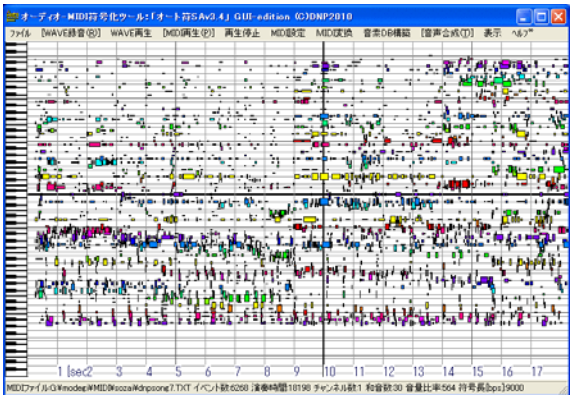

図３　改良手法による MIDI 符号化結果