

F-16

# MIDI 符号化における音源分離処理系の試作 Prototyping of Sound Source Separation Processes for MIDI Encoding

茂出木敏雄†  
Toshio Modegi

## 1. まえがき

筆者らは音楽素材に限らずあらゆる音響信号を一律に平均律音階に分解できる高精度な周波数解析技術の開発を進めてきた。この技術を高能率符号化分野に応用すると、再生品質に関しては3倍程度の16kbpsのビットレートで符号化されたMP3方式より優位性が確認されている[1]。また、カラオケなどエンターテインメント分野に応用すると、MIDIコンテンツの制作工程が省力化される[2]。しかし、MIDI符号化の適用素材を拡大し、高品質なMIDIコンテンツを生成するためには、音源分離の開発が必須になる。そこで、MIDI符号化の過程で複数の音色パラメータを算出し、高精度多チャンネルな音源分離を実現するモデルを提案した[3]。

本稿では、3つの音色パラメータを定義付け、ピアノ・ギター・ボーカルに関する音色管理データベースを構築し、2種の音色が混在する音響信号に対してMIDI符号化を試み、音源分離精度の定量的評価を行ったので、その結果を報告する。

## 2. MIDI符号化方式の概要

筆者らが提案するMIDI符号化方式とは、図1に示すように、与えられたソース音響信号を電子楽器で再現できるようなMIDI符号に変換するものである[1][2]。ソースが楽器音の場合、抑揚などの演奏パラメータを含めて演奏された音符を拾い(採譜)、使用された楽器に近い音色で再生すれば、原理的に原音を再現できる。また、ソースが音声(ボーカル)の場合、フォルマントに相当する部分を和音として拾い、声道に近い管楽器の音色(GM規格ではNo.54のコーラス「ウー音」で再生することにより、音声を再現する方法を考案した[1][2]。その他、自然音など効果音に対しても同様に符号化できることを確認している。

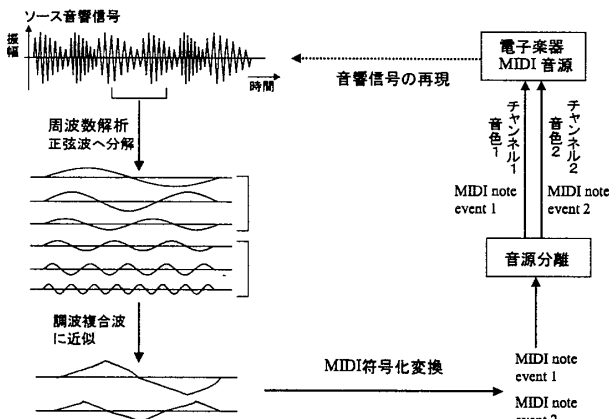


図1 MIDI符号化の概念

そして、図1右端に示しているように、様々な音色が混合された一般の音響信号に対して、符号化されたMIDIデータを音色で分離し、最適な音色で再生できるようにチャンネルを割り当てられれば、汎用的な超低ビットレートのオーディオ符号化方式として実現できる。

## 3. 音色パラメータと音源分離方式の設計

MIDI符号化は与えられた波形信号に対して、ノートナンバー(音高)・ベロシティ(音量)・デュレーション(音長)の3つのパラメータからMIDIイベント(音符)を時系列に抽出する処理である。文献[3]で提案したように、この過程で、各イベントごとに、ある注目する音(基本周波数)に対する局所的なスペクトル分布(周波数ゆらぎ)、広域的なスペクトル分布(倍音分布)、および振幅と周波数の時間変動(時系列変動)からなる3の音色パラメータを計算し、先のMIDIイベントを構成する3つの基本パラメータに追加する。また、ソースがステレオ波形信号の場合は、左右のステレオ定位パラメータも算出し、計7つのパラメータが各MIDIイベントごとに付加される。

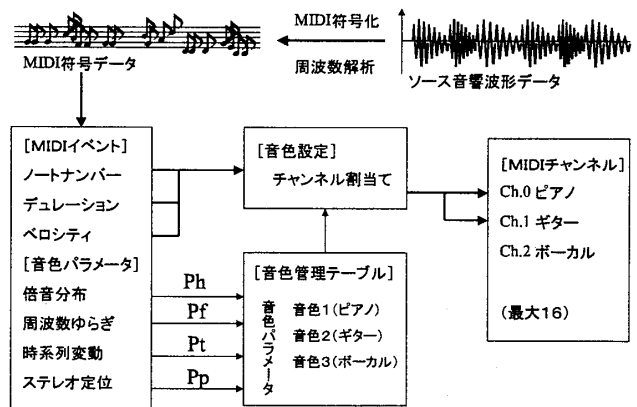


図2 音色パラメータと音源分離処理構成

図2に文献[3]で提案したモデルをより具体化した音源分離処理の構成図を示す。ソース音響信号に周波数解析を施し、得られたMIDI符号に対して、各々音色パラメータを元に音色管理テーブルを参照し、最も相応しい音色のチャンネルに割り当てる。音色管理テーブルにはピアノやボーカルといった音色別に各音色パラメータの組み合わせの頻度が記録されている。例えばソースがモノラル信号の場合には、倍音分布  $Ph$ ・周波数ゆらぎ  $Pf$ ・時系列変動  $Pt$  の3つの音色パラメータを各々12段階で算出し、1728通りの音色パラメータの出現頻度がピアノ・ギター・ボーカル別に記録されている。ある解析時刻  $t$  におけるノートナンバー  $n$  に対応するベロシティを  $Vt(n)$  とし、デュレーションを  $Dt$  すると、3つの音色パラメータは以下のように定義した。

$$P_h = \sum_{i=0,7} \{a_i V_r(n+k_i) - b_i V_r(n-k_i)\} / V_r(n)$$

$$k_i = \{0, 12, 19, 24, 28, 31, 34, 36\}. \quad (1)$$

$$P_f = \sum_{i=1,3} \{a_i V_r(n+i) + b_i V_r(n-i)\} / V_r(n). \quad (2)$$

$$P_r = a / D_r. \quad (3)$$

( $a_i, b_i, a$  は正規化するための定数)

MIDI 符号化されたイベントデータの [ $P_h, P_f, P_r$ ] の値をもとに音色管理テーブルを参照し、[ピアノ, ギター, ボーカル] の音色のなかで最も頻度値が高い音色をそのイベントデータの音色としてチャンネルを割り当てることにより、音源分離を実現できる。

この音色管理テーブルは単一音色の波形サンプルを数多く準備し、MIDI 符号化処理を音色別に行うことにより、自動的に構築できる。また、分離精度が不十分な場合には、音色サンプルを追加し学習を強化することができる。更に、記録できる音色数や音色パラメータ数は自在に拡張でき、同じピアノ音色でもモデルの異なるピアノ音色サンプルを準備すれば、精度の高い音色分離が期待できる。

#### 4. 実験結果

符号化実験としてピアノ、ギター、ボーカルのパート別スタジオ録音を行い 3 種の単音波形データ (モノラル CD 音質、1.5 秒) を準備し、各々に対して MIDI 符号化を行い音色管理データベースを構築した。そして、2 パートずつ波形データのミックスを行い、3 種の波形データに対して音源分離実験を行った。また、ピアノとボーカルに関しては音色の異なる単音波形データと合奏データも準備し、音色が異なるデータで学習させたときの精度劣化も調べた。

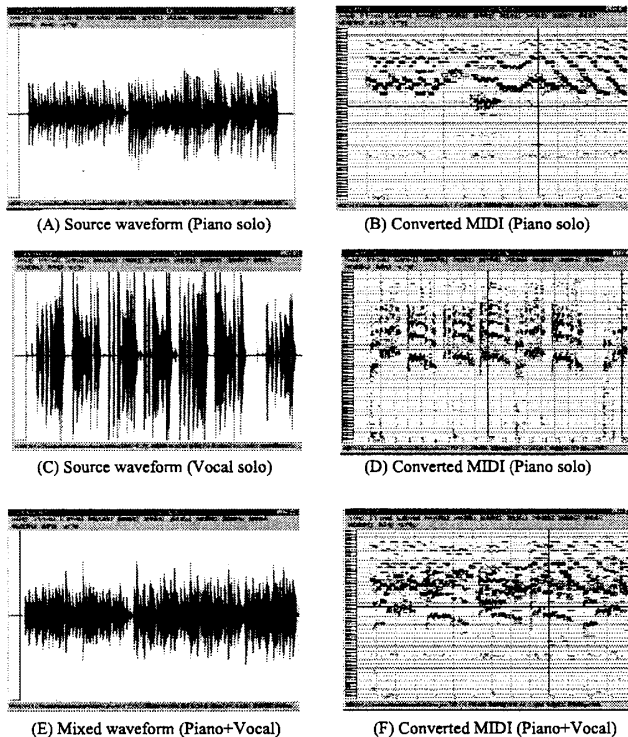


図3 ピアノとボーカルの音源分離実験結果

図3にピアノとボーカルに対してパート別に MIDI 符号化を行った結果と、混合音に対して MIDI 符号化を行った結果を示す。符号化条件として、出力和音数を 32、ビット

レート を 9kbps に設定して符号化を行った結果、4668 対の MIDI イベント (ノートオンとノートオフ) が得られ、そのうち 768 イベントはピアノに、3900 イベントはボーカルに振り分けられた。パート別符号化結果と照合すると、ピアノについては 77%、ボーカルについては 66% は一致しており、ピアノに振り分けられた 11% は誤りでボーカルに振り分けられるべきものであった。

表1にその他、実験結果を列挙する。ピアノとボーカルに対しては、表1最終行に示されているように、別の単音データで音色管理テーブルを構築して実験を行ったが、顕著な分離精度の劣化は見られなかった。また、別のピアノ伴奏ボーカル音楽サンプルに対して実験しても、同程度の分離精度が得られた。

表1 2音色混合の分離実験結果

Materials	correct extract ratio		incorrect extract ratio	
Piano+Vocal	Piano:77%	Vocal:66%	Piano:11%	Vocal:0%
Piano+Guitar	Piano:55%	Guitar:90%	Piano:11%	Guitar:18%
Guitar+Vocal	Guitar:88%	Vocal:77%	Guitar:11%	Vocal:22%
Piano+Vocal processed by different tone database	Piano:77%	Vocal:55%	Piano:11%	Vocal:0%

#### 5. あとがき

本提案に基づく音源分離手法により、2音色混合のパート分離に関しては、55%以上の精度が得られ、誤り率は最悪2%であった。特にギターに関しては、90%近い精度が得られた。今後は混合数および対応可能な音色数を増大させ、音色管理テーブルの拡張を行った上で、本提案モデルの評価を進め、最適な音色パラメータの見極めを進めてゆきたい。

本研究は、(財) デジタルコンテンツ協会の平成12年度事業「コンテンツ制作基盤ツール等開発」の一環として、経済産業省・情報処理振興事業協会より助成を受け進めさせていただいた。本研究事業に尽力を賜った皆様にはこの場を借りて感謝の意を示す。本実験で使用したソフトウェアツール「オート符®」に関してはフリーウェアとして下記サイトにて配付しており、随時バージョンアップを行っているため、ご活用ください。

(URL: <http://www.dcaj.or.jp> 2002年6月現在、バージョン2.5配布中)

#### 参考文献

- [1] Toshio Modegi, "Very Low Bit-rate Audio Coding Technique Using MIDI Representation," Proceedings of ACM 11-th NOSSDAV Workshop, pp.167-176 (Jun. 2001).
- [2] Toshio Modegi, "Development of MIDI Encoder Auto-F for Creating MIDI Controllable General Audio Contents," Proceedings of International Workshop on Entertainment Computing IWEC2002, pp.229-236 (May 2002).
- [3] 茂出木敏雄: 「MIDI 符号化における音源分離処理系の設計」、電子情報通信学会 2002年総合大会 A-10-10, pp.245 (Mar.2002).