

MIDI技術のオーディオ符号化への応用

デモ15

茂出木 敏雄

大日本印刷株式会社

飯作 俊一

郵政省 通信総合研究所

1. まえがき

著者らは、心音や肺音といった医療診断に利用される生理的リズム音声に対して、MIDIフォーマットで符号化する手法[1]を提案した。本手法は、信号区間からリアルタイムに単一のピッチ情報を抽出する手法であり、鳥の声など動物の声に対しても適用できることを確認している。しかし、歌声など音声一般に適用するためには、複数のピッチ情報（ホルマント周波数）を抽出することが不可欠となる。

本稿では、FFTを用いてマルチトラック分離し、高品質な音声を再生できるMIDI符号化アルゴリズムを提案する。

2. 符号化アルゴリズムの説明

2.1 FFT及びピーク周波数検出

ソースPCM音声データに対して、等間隔にFFTを施し、スペクトログラム $S_d(f)$ を作成する。この時、周波数軸 f は、MIDIのノートナンバー N を単位とする対数スケール $S_d(N)$ に変換する。（文献[1]）

次にピークとなるノートナンバー N_p を、FFT区間 d 単位ごと、順に抽出する。第1ピーク N_p の倍音となる $N_p+12, +19, +24, +28, +31, +34, +36$ などが2番目以降のピークとして抽出されないよう処置を施す。

2.2 トラック分離及び信号区間抽出

前節の処理から、区間 d に対して、 P （最大16）対のピーク・ノートナンバー $N_p(d)$ とスペクトル強度 $V_p(d)$ が強度順に求まり、これをノートナンバーでソートし、 P 個のトラックに割り当てる。これにより、以下の処理は P チャンネル・ステレオとして扱い、各トラックごとに独立して処理を行えばよい。

各等間隔区間 d に対して、ノートナンバーが同じ隣接区間を統合し、区間代表強度として大きい方 V_{max} を与える。ただし、後続区間の強度が所定のしきい値 S_I より大きい場合は統合しない。

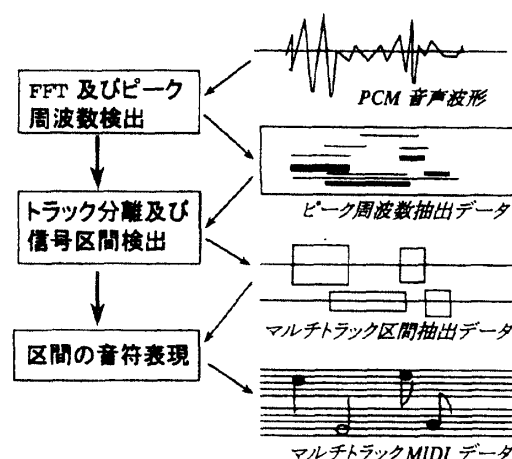


図1 符号化アルゴリズムの説明

2.3 区間の音符表現

文献[1]と同様に、各区間を1つの音符（ノート）として符号化し、ノートオン・ノートオフのコマンド列を作成する。SMFフォーマット1における各トラックに各々符号化した P 個のトラックを割り当てる。

3. あとがき

本稿で提案した手法は、文献[1]に比べ計算負荷を要するというデメリットがあるが、人の音声などの表現には有益であることが認められた。実験例として、テンポが比較的緩やかな幾つかの唱歌に対して8トラックにより符号化し、ビットレート 10kbps 前後で、標準的なGM音源（ピアノ音色設定）をそのまま用いて、原音に近い高品質な再生ができることを確認した。

今後、テンポの速い歌声に対しても追従するよう、符号化手法に改善を加えるとともに、有声音と無声音を明瞭に分離するアルゴリズムを開発する予定である。

参考文献

- [1] 茂出木、飯作：“生理的リズム音声のMIDI符号化手法の提案”，信学総合大会，SD-5-1（1997）

Application of MIDI Technology for General Audio Signal Coding

Toshio Modegi (Dainippon Printing Co.,Ltd.)

Shun-ichi Iisaku (Communications Research Laboratory, M.P.T.)