

M-007

## 表情付けを考慮した楽音符号への情報ハイディング

An information hiding technique to music code with adaptation in musical expression

山本紘太郎\*  
Kotaro YAMAMOTO岩切宗利\*  
Munetoshi IWAKIRI

## 1. はじめに

電子楽器の演奏情報 (SMF:Standard MIDI File[1]) への情報ハイディング [2, 3] として, 演奏情報に情報を埋め込む手法が岩切らによって示されている [3].

しかし, 演奏情報の操作は, 埋め込みによる音質劣化が避けられないため, 埋め込み容量を増加させると演奏の表情を崩してしまうことが考えられる.

本報告では, 演奏情報に加えられた表情を崩すことなく情報系列を操作し, 音質を維持もしくは質感を向上させながら情報を隠す, 表情付けに適応した情報ハイディング手法について提案する.

## 2. MIDI と演奏の表情付け

## 2.1 MIDI と SMF(Standard MIDI File)

MIDI 規格は, Musical Instrument Digital Interface の略であり, 電子楽器のデジタル制御を行うことを目的として定められたものである. 演奏の制御は, 主に 2~3byte のメッセージを音源へ送ることにより実現している. この演奏情報記録形式の一つが SMF (Standard MIDI File) である. 本報告では, SMF を情報の埋め込み対象とした.

## 2.2 演奏の表情付け

人間が演奏をするとき, 演奏者は, 楽譜からの情報をもとに楽曲のイメージを創りあげ, 演奏でそれを表現しようとする.

一方, SMF では, 演奏を表情豊かなものにするため, 音強, 音長, 音高, テンポを表現する各パラメータを操作して表情付けが施される. 特に音強を表すノートオンベロシティ (note-on velocity) は, 演奏の表情に及ぼす影響が大きい. ここでは, このノートオンベロシティに対し, 表情付けに適応しつつ情報を埋め込む手法を示す.

## 3. 提案方式とその評価

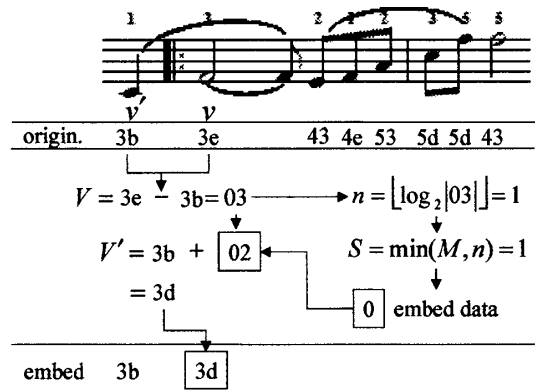
ベロシティ (velocity) を対象とした埋め込み手法を二つ提案する. ベロシティに対する情報埋め込み法に関しては, すでに演奏情報に対する電子透かし技術が報告されている [3]. 提案手法は, 演奏の表情付けを考慮して情報を埋め込む点に特徴がある.

## 3.1 即時適応化法

即時適応化方式は, 埋め込み対象の音と, その直前に発音された音とのオンベロシティ変化量に注目する. このオンベロシティ変化量の絶対値に応じて, 下位ビットを透かし信号で置換する方法が即時適応化である.

## 3.1.1 提案方式

提案方式の処理イメージを図 1 に示す. まず, 最大埋め込み許可 bit 数を  $M$  とし, 実際の埋め込み bit 数を  $S$  とする. MIDI データ列から抽出した埋め込み対象のオンベロシティを  $v$  とする. また, その直前の発音のオン

図 1 即時適応化法の処理イメージ ( $M=3$ )

ベロシティを抽出し,  $v'$  とする. オンベロシティ変化量を  $V$  とすると,

$$V = v - v' \quad (1)$$

となる. このときの埋め込み可能 bit 数  $n$

$$n = \lfloor \log_2 |V| \rfloor - 1 \quad (2)$$

により, 埋め込み bit 数  $S$  を

$$S = \min(M, n) \quad (3)$$

と決定する. これにより埋め込み情報と  $|V|$  の下位 bit と置換する.

この手法の特性は, 直前の発音情報との差分の大きさに対して情報を埋め込むことである.

## 3.1.2 評価結果と考察

本報告では, RWC Music Database のクラシック曲からサンプルを 4 つ選び, 即時適応化によって情報を埋め込んだ (表 1).

本実験では, 音質の評価法として, 20 代の被験者 7 名の主観的絶対値によるオピニオン評価を用いた. これは, 提示した演奏音の品質を 5 段階評価させ, 得られた評価値から平均オピニオン値 (MOS: Mean Opinion Score) を求めるものである.

評価の基準は, オリジナルと判別できない状態を 5 とし, 劣化が著しい状態を 1 とした. また, 評価の際には, 埋め込みを施した演奏 (1bit, 3bit, 5bit), 埋め込みのない演奏 (origin.) および音の強さを全て一定にクオンタイズした演奏 (quant.) を被験者に外見上の区別ができないように提示した. 本実験の演奏には, YAMAHA の MU-2000 を用いた.

実験の結果, 表 2 の評価値が得られた. この結果から, 全体の評価値の差が非常に少ないことがわかる. つまり, 情報を埋め込んだデータとオリジナルデータの見分けがつかないと考えられる.

\*防衛大学校, National Defense Academy

表1 即時適応化の埋め込み結果

S	1bit[bit]	3bit[bit]	5bit[bit]
orchestra	490	906	932
piano	132	323	380
strings	124	302	309
flute	157	416	487

表2 即時適応化の評価

n	1bit	3bit	5bit	origin.	quant.
orchestra	3.14	3.57	3.29	3.57	3.00
piano	3.43	3.14	3.29	3.43	3.00
strings	3.00	2.71	3.00	3.14	2.43
flute	2.86	3.14	3.29	2.40	2.21

5bitの埋め込みについては、MOSが低くなっている。このとき、オリジナルデータと埋め込みデータとの間で、velocityの値域の4分の1に相当する偏差が生じる場合がある。その結果、聴覚上明らかな劣化が知覚されたと考えられる。

### 3.2 経時適応化法

即時適応化を応用して、埋め込み情報の検出・破壊を困難にするための手法について検討してみる。

即時適応化の問題として、埋め込み容量を決定するための情報に直前の発音との差分値しか持たない点が挙げられる。したがって、即時適応化は情報の検出が容易である反面、セキュリティ上の対策が必要となる。そこで、連続するノートオンベロシティを一定の規則でグループ化し、これをベクトルとして扱う手法について検討した。

#### 3.2.1 固定長グループ化

グループ化定数に基づく数の発音情報をまとめてグループとして扱う方法である。この処理手順を示す。

- (1) 連続する同チャンネルのノートオン・ベロシティを  $l$  個まとめる ( $l$  はグループ化定数)。
- (2) グループの平均値  $v$  を計算する。
- (3)  $v$  に情報を埋め込み  $v'$  とする。
- (4) 更新された平均値  $v'$  を満たすように、グループに含まれるベロシティ値を操作する。

この処理方式では、グループ内のベロシティ推移の包絡線をできるだけ保持することになる。

#### 3.2.2 時系列を考慮した可変長グループ化

固定長グループ化を用いて埋め込み容量を増加させると、グループの境界付近での velocity の急激な変化の発生など、聴感上問題を生じる場合がある。ここでは、時間軸上にノート情報を並べ、「互いの情報の間隔が小さいものをまとめるように」グループ化する。

この手順は以下の通りである。

- (1) SMF ヘッダから、音の分解能を得る。
- (2) 分解能をもとに、発音間隔が8分音符の連続するノート情報を採り、グループ化する。このとき、対象とする発音間隔には幅を持たせる。
- (3) (1), (2) の操作を4分, 2分, 全音符に対して行う。

この方法は、発音間隔が近い音をまとめることにより、一つの連続した音列の velocity が急激に変化しないようにするものである。

表3 経時適応化の埋め込み結果

sample	3bit	group[bit]		length[bit]	
	[bit]	3bit	5bit	3bit	5bit
orchestra	906	264	440	177	295
piano	323	54	90	57	95
strings	302	48	80	99	165

表4 経時適応化の評価

sample	origin.	3bit	group		length	
			3bit	5bit	3bit	5bit
orchestra	3.14	3.14	3.43	2.71	3.29	3.14
piano	3.29	2.86	3.14	2.57	3.14	3.14
strings	3.00	3.43	3.14	2.71	3.14	3.00

### 3.2.3 評価結果と考察

提案手法による音質への影響について、表3に示すサンプルを用い、評価者を7名としてMOSを求めた。評価対象については、3bitの埋め込みで即時適応化を施したデータ(3bit)とオリジナルデータ(origin.)、ならびに2つの方法(固定長group, 可変長length)でグループ化し、それぞれに3bit, 5bitの情報を埋め込んだデータの6種類とした。

表4に示した実験結果から、3bitまでの埋め込みについては固定長グループ化が、4bit以上埋め込む場合は、容量、音質の面から時系列を考慮した可変長グループ化が有利であることが読み取れる。

また、次のことが推測できる。

- (1) データ上隣接する発音情報でも、時間軸上での距離が大きくなると、その相関性が小さくなる。
- (2) 一つの旋律の中での急激な音の変化に対して人の聴覚は敏感である。

### 4. おわりに

二つの適応化法について提案、評価した。即時適応化が一定量までの埋め込みに対して有効であることを確かめた。また、検出を困難にする手法として経時適応化を検討した。これを評価する中で、時間軸に対する適応化の有効性を確認した。

今後の検討事項として、グループ化手法の改善、小節単位や大きなメロディ等で捉えた適応化、表情を強める方向での適応化の3点が挙げられる。これらの検討課題を解消すれば、楽音符号への埋め込み容量を飛躍的に増加でき、演奏の表情付けと著作権保護の同時実現にもつながると考えられる。

### 参考文献

- [1] 社団法人 音楽電子事業協会：MIDI 1.0 規格書，リットーミュージック (1998)。
- [2] 松本勉，井上大介，北林創太：演奏データファイル SMF への情報ハイディング方式，2000年暗号と情報セキュリティシンポジウム，SCIS2000-C03(2000)。
- [3] 岩切宗利，山本紘太郎，関根健一郎，松井甲子雄：電子演奏の半雑音化と音源符号への電子透かし，情報処理学会論文誌，Vol. 43, No. 2 pp.225-233(2002)。