

# 表情付けを考慮したSMFステガノグラフィ

山本 紘太郎<sup>†</sup> 岩切 宗利<sup>†</sup>

電子演奏の楽音符号に対する情報ハイディングは、SMFのデータ構造を利用して埋め込むもの、演奏情報そのものに埋め込むものに大別され、それぞれについて研究・報告がなされている。しかし、演奏情報そのものに対する埋込みに関しては、演奏の表情付けを考慮した手法は検討されていないのが現状である。そのことは、実際に演奏情報に埋込みを行った場合に聴感上の違和感をもたらす可能性を持っている。本論文では、SMFの表情付けを考慮したステガノグラフィについて提案し、これを評価・検討する。

## A Steganography to Music Code with Adaptation in Musical Expression

KOTARO YAMAMOTO<sup>†</sup> and MUNETOSHI IWAKIRI<sup>†</sup>

Information hiding to the music code of digital instruments is divided roughly two pattern. One is the method using the data structure of SMF, and the other is embedding to the performance information. However, the present condition, the technique of adaptation musical expression is not considered. A possibility bring about the sense of incongruity on a feeling when that actually carries out embedding to performance information. This paper presents a steganography to SMF with adaptation in musical expression, and evaluate and examine.

### 1. はじめに

ネットワークの広帯域化、インフラの整備と、コンピュータの高速化、ストレージの大容量化によってデジタルコンテンツが多く利用されるようになってきている。デジタルコンテンツには、電子文書や画像、動画や音声があるが、これらに情報を埋め込むことで様々な用途に用いる情報ハイディング技術が研究されている。この中に、電子楽器の演奏情報への情報ハイディングがある。

電子楽器の演奏情報 (SMF: Standard MIDI File<sup>1)</sup>) に対する情報ハイディングについては、松本らによる同時イベントの順序入れ替えによる方法<sup>2)</sup>、岩切らによるタイミング情報の可変長表記を利用する埋込み法<sup>3)</sup>、ならびに演奏情報に対する埋込み法<sup>3)</sup>等が報告されている。特に前者の2つについてはSMFのファイル形式が持つ等価構造を利用して情報を埋め込むものであるため、演奏自体に影響を及ぼさないという特長がある。しかしその反面、特徴的なデータ構造を利用して解析されやすいという特徴もある。井上らは文献2)のステガノグラフィに対するステゴ解析

が有効であることを明らかにし<sup>4)</sup>、さらに耐性を付与する手法を検討している<sup>5)</sup>。

一方、演奏情報に対する情報の埋込み (演奏自体が変化する手法) については、文献3)の報告にあるとおり発音情報の音強成分LSBを埋込み情報で置換するものである。これらは埋込み情報量を増加させるに従って音質に影響を及ぼさず。報告には、埋込みによって品質が向上する可能性が示されているが、それは埋込み情報によって偶然もたらされたものであると考えられる。音強成分の変化は埋込み情報によって完全にランダムに与えられるため、埋込みによって音質を低下させる可能性がある。これは埋込み手法が演奏の表情付けを考慮していないために生じる。

表情付けとは、演奏情報に「抑揚」や「揺らぎ」を人為的に与え、表情豊かにすることである<sup>6)</sup>。この観点から、MIDIデータに対して自動的に演奏表情をつける研究もさかに行われている<sup>7)</sup>。これらの研究成果により、演奏における表情がどういふものであるかが近年明らかにされつつある。

本論文では、文献3)に示されている手法を改善し、演奏情報に加えられた表情を崩さずに情報を埋め込む、表情付けに適応したステガノグラフィについて提案する。ここでは、電子楽器の制御に用いられるMIDI規格の中でも、その標準的な記録形式である

<sup>†</sup> 防衛大学校情報工学科

Department of Computer Science, National Defense Academy

SMF (Standard MIDI File) を対象とする<sup>8),9)</sup>。

演奏に表情を加える要素の1つとして、音を発音・消音する際に送信されるノートメッセージがある。これに含まれる、発音の強さを表すノートオンベロシティ (note-on velocity) は演奏に表情を与える重要な要素の1つである。文献 3) では、ノートオンベロシティに注目して情報を埋め込む手法について提案されており、本研究でもこの手法を用いた。

まず、2章では、演奏の表情付けおよび MIDI 規格の概要について示す。ここでは既存の手法をステガノグラフィとして用いる場合の問題点についても述べる。3章では、ノートオンベロシティの表情付けを考慮しながら情報を埋め込む手法を2つ提案し、それぞれの性能を評価した結果を示す。

## 2. 演奏の表情付けと MIDI 規格

### 2.1 演奏の表情付け

楽音演奏の基準となるのは通常、楽譜である。楽譜には発音時期、音の長さ・高さが音符、休符等の記号によって記述されている。また、音の強弱や演奏のテンポ等が演奏記号として記述されている。

ここで、この演奏を電子楽器により実現することを考えてみる。楽譜をそのまま入力して演奏させた場合、その演奏は人間味のない、機械的で不自然な演奏になりやすい。このとき、電子楽器は楽譜どおり正確に一定のリズムと音で演奏している。一方演奏者は、その楽譜からの情報を解釈し、独自の表情を付けて演奏を行う。

すなわち、人間が演奏をする場合、演奏者は楽譜にある情報だけを読んでいるわけではないことになる。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。それにより演奏を表情豊かなものに行っている。

したがって、電子楽器に演奏をさせる場合にも、音の強さ、長さやテンポを操作し、より表情豊かで芸術的な演奏にするための表情付けが施される。本論文では、特に音の強さを取り扱う。

#### 2.1.1 音強成分に対する表情付け

楽譜上で音強成分の表情付けに関わるものは主に演奏記号である。これらは、音強の相対的な強弱を表すもので、以下のようなものがある。

- (1) フォルテ (*f*): 「強く」の意。
- (2) ピアノ (*p*): 「弱く」の意。
- (3) クレッシェンド (*crescendo*): 「徐々に強く」の意。
- (4) ディミヌエンド (*diminuendo*): 「徐々に弱く」

の意。

このように、演奏記号は絶対的な大きさを与えるわけではない。あくまでこれらは演奏の基準として示され、演奏者の解釈する余地が残されている。そのため、実際の演奏では音の強さはつねに連続的に変化し、それによって演奏に抑揚を与えている。すなわち、「1つの音の音強成分」ではなく「連続する音の、音強成分の変化量」に注目することが表情付けには重要である。

### 2.2 MIDI 規格の概要

MIDI 規格<sup>1)</sup> 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 2~3 byte の制御信号 (MIDI 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1つが SMF (Standard MIDI File) である<sup>1)</sup>。

#### 2.3 SMF (Standard MIDI File)

SMF は、MIDI の演奏情報を記録するための標準的な記録方式である。これは、MIDI 規格の各種メッセージを制御信号とし、その伝送タイミング情報とともに記録するデータ形式である。ここでは、SMF のデータ構造の特徴について示す。

##### 2.3.1 ヘッダチャンクとトラック

SMF は先頭に 12 バイトのヘッダ部を持ち、ヘッダ部にはトラックに関する情報、フォーマットの種類、演奏全体を通しての時間分解能等が記録されている。このヘッダ部に続き、演奏情報そのものであるトラックチャンクが記録される。

##### 2.3.2 トラックチャンク

トラックチャンクは、先頭のトラックサイズを記録するヘッダ部およびデータ領域で構成される。データ領域には、MIDI メッセージ等を表現するイベントデータが、それを送信するタイミング情報であるデルタタイムとともに記録される。

##### 2.3.3 MIDI メッセージ

MIDI 機器にデータを送信するタイミング情報は、直前のイベントからの経過時間として 1~4 byte の可変長データ表記により記述される。これをデルタタイムという。それに続いて実際に送信する MIDI メッセージ等が記録される。これらのデルタタイムと MIDI メッセージを合わせて MIDI イベントという。トラックチャンクのデータ領域は、これら MIDI イベントの集合体で構成される。

SMF の MIDI メッセージには楽曲の発音、消音を指

演奏情報を格納するデータセクション

直前のイベントデータからの発行時間の差分を記録する。

表 1 MIDI チャネルメッセージの一例  
Table 1 An example of MIDI channel messages.

message		data(16)
channel	voice	
	note on	9n kk vv
	note off	8n kk vv
	control change	Bn cc data
	program change	Cn pp
	channel pressure	Dn dd
	pitch bend	En LSB MSB

示したり、楽器の変更を指示したりするチャンネルメッセージ(表 1)のほか、メタイベント、エクスクルーシブイベントがある。本研究ではこの中でも、発音情報であるノートメッセージに対して情報を埋め込むことにした。

#### 2.4 MIDI における表情付けの要素

SMF を対象とした表情付けは、主に次のパラメータ操作によってなされる。

- (1) 音強 (velocity)  
note-on を送る際に指定される音の強さである。0~127 の値域を持つ。
- (2) 音長 (duration)  
note-on から note-off までのデルタタイムの和で表される音の長さである。
- (3) 各種エフェクト (expression, pitch-bend, etc.)  
楽器の種類ごとに用法・効果が異なるコントロールチェンジ機能である。
- (4) テンポ  
同時刻の全チャンネルの発音間隔等を操作することにより曲の進行速度を制御する。曲そのものの表情付けに大きな影響を及ぼす。

一般的にこれらの表情付けは、演奏情報の制作者によってなされる。すなわち、MIDI の演奏情報 (SMF データ) はその制作者の表情付けの方法によって様々な個性を持つことになる。

2.5 演奏情報を操作する SMF ステガノグラフィ  
演奏情報に情報を埋め込む手法として、文献 3) のノートオンベロシティに埋め込むものが電子透かしとして提案されている。これをステガノグラフィに応用する際の問題点について検討する。

##### 2.5.1 埋込み領域

文献 2), 3) の、ファイルの等価構造を用いる手法は演奏にまったく影響を及ぼさない。しかし、それらの手法はファイルのデータ構造に情報が埋め込まれているため、秘匿通信の存在を検出される可能性があ

る<sup>4),5)</sup>。埋込みによる変化がデータ構造に現れる手法をステガノグラフィに用いる場合は、データ構造の特徴を考慮した対策が必要である。

一方、演奏情報への埋込み<sup>3)</sup>は、埋込みによる特徴をデータ構造に残さない。このため、データ構造を利用した解析を防ぐことができ、これは通信そのものの秘匿に有利である。

##### 2.5.2 音 質

既存の演奏情報への埋込み法<sup>3)</sup>はノートオンベロシティの LSB を置換するものである。特に、埋込み情報量をオンベロシティの大きさに適応する特徴がある。

しかし、この手法はオンベロシティが大きなものには大きな埋込み情報量を与える。そのため、たとえばオンベロシティが大きな値をとり続けた場合、埋込み情報によってベロシティ変化 (表情付け) が乱れてしまうことになる。これは「1つのオンベロシティ」のみを埋込みの対象にしたためである。通常、音強成分はほぼ連続して変化し、それが演奏に表情を与える。したがって、表情付けを崩さないためには「ベロシティの変化量」に注目しなければならない。

##### 2.5.3 保 全 性

既存の演奏情報への埋込み法<sup>3)</sup>では、単純にベロシティの大きさを得るだけで情報の抽出が可能である。したがって、これをステガノグラフィとして用いる場合、あらかじめ埋込み情報の暗号化等の対策が必要である。

#### 2.6 課 題

課題として、次の 2 つをあげる。まずは、品質が低下する可能性を解消し、聴感上不自然さを感じさせないようにすることである。もう 1 つは情報の抽出を困難にするためにセキュリティ上の対策を施すことである。

本論文では、これらの課題を解消するためにベロシティ変化量への表情付けを崩すことなく情報を埋め込む SMF ステガノグラフィを提案する。

### 3. 提案方式とその評価

ベロシティ (velocity) を対象としたステガノグラフィを 2 つ提案する。ベロシティに対する情報埋込み法に関する基本的なアイディアは、文献 3) で報告されている演奏情報に対する電子透かし技術である。

#### 3.1 即時適応化法

即時適応化方式は、埋込み対象の音と、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、下位ビットを埋込み情報で置換する方法が即時適応化である。

SMF として記録する場合の制御情報、拍子やテンポの設定  
電子楽器固有の機能を制御する場合に送信する信号

3.1.1 即時適応化の情報埋込み法

提案方式の処理ブロックを図1, 処理イメージを図2に示す. ここでは  $k$  番目の発音情報に情報を埋め込むときの処理について示す.

Step 1.  $k$  番目のノートオンペロシティを  $v_k$  とし, オンペロシティ変化量

$$V_k = v_k - v_{k-1} \tag{1}$$

を得る.

Step 2.  $V_k$  から, 埋込み情報量の上限

$$n_k = \lfloor \log_2 |V_k| \rfloor \tag{2}$$

を定める. この  $n_k$  と埋込み強度  $M$  から, 埋込みビット数

$$S_k = \min(M, n_k) \tag{3}$$

を求める.  $\min$  は 2 つの値を比較し最小値を返す関数である.

Step 3. 埋込みデータから  $S_k$  ビットを, 埋込み情報  $E_k$  として取り出す.

$|V_k|$  の下位  $S_k$  ビットを  $E_k$  で置換し,  $|V'_k|$  とする.

Step 4. 埋込みを施したオンペロシティを

$$v'_k = v_{k-1} + V'_k \tag{4}$$

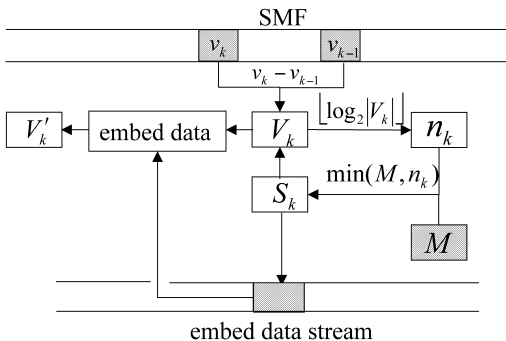


図1 即時適応化法の埋込み処理

Fig. 1 Schematic flow of immediate adaptation.

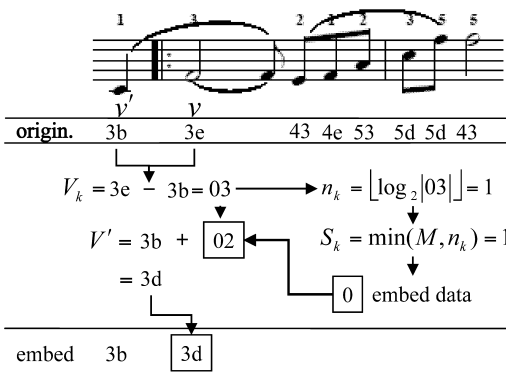


図2 即時適応化法の処理例 ( $M = 3$ )

Fig. 2 An example of immediate adaptation.

とする.

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. ここではペロシティの変化量を「連続する2音の表情付け成分」と見なし, その大きさに適応している. したがって, 図3のように,  $|V_k|$  の値を小さな埋込み情報に置換しても, また大きい情報に置換しても, 差分の正負が変わらないことになる. つまり, 差分の正負にかかわらず大きさに対して適応化を施すことによって, 図4のように演奏音の強さが持つ包絡線の形状を保持できる. このため, 聴いた者が感じる品質の劣化を抑制できる.

3.1.2 即時適応化の情報抽出法

即時適応化法の情報抽出には, 直前の発音からのオンペロシティ変化量を用いる. ただし, 送受信者間で埋込み強度  $M$  を鍵としてあらかじめ共有しておくものとする.

Step 1.  $k$  番目の発音情報から情報を抽出するために, オンペロシティ変化量

$$V_k = v_k - v_{k-1} \tag{5}$$

を求める.

Step 2. 埋込み可能ビット数

$$n_k = \lfloor \log_2 |V_k| \rfloor \tag{6}$$

を求め, 埋込み強度  $M$  (共有鍵) と  $n_k$  から埋込みビット数

$$S_k = \min(M, n_k) \tag{7}$$

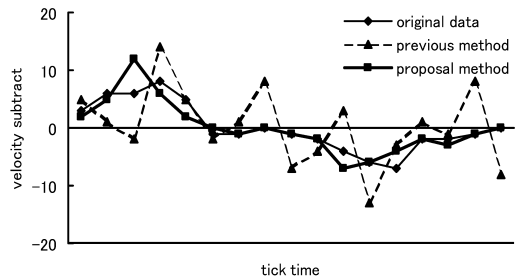


図3 各埋込みによる連続2音の差分値の変化

Fig. 3 Differential values of continuous note-on velocity by several method.

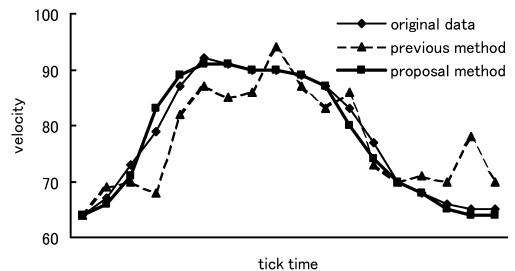


図4 各埋込みによるペロシティの流れ

Fig. 4 Fluctuation of velocity on several method.

を得る．

Step 3.  $|V_k|$  から下位  $S_k$  ビットを埋込み情報  $E_k$  として抽出する．

この手法では、埋込みの際に逐次オンベロシティを更新するため、埋込み前の楽曲を抽出する際に必要としない．

### 3.1.3 システムの評価

本研究では、実験用 SMF データとして表 2 に示す RWC 研究用音楽データベース（クラシック音楽）を用いた<sup>10)</sup>．図 5 にそれぞれの実験用データのノートオンベロシティの変化を示す．この縦軸がベロシティの値、横軸が時間を表している．これらの実験用データに対し、即時適応化による埋込みを施した結果、表 3

表 2 実験用 MIDI データ  
Table 2 Experimental MIDI data files.

Name	MIDI Source	Time[sec]	Size[byte]
orchestra	RM-C007.mid	21	5,803
strings	RM-C015.mid	19	3,847
piano	RM-C022.mid	17	4,193
flute	RM-C046.mid	19	3,506

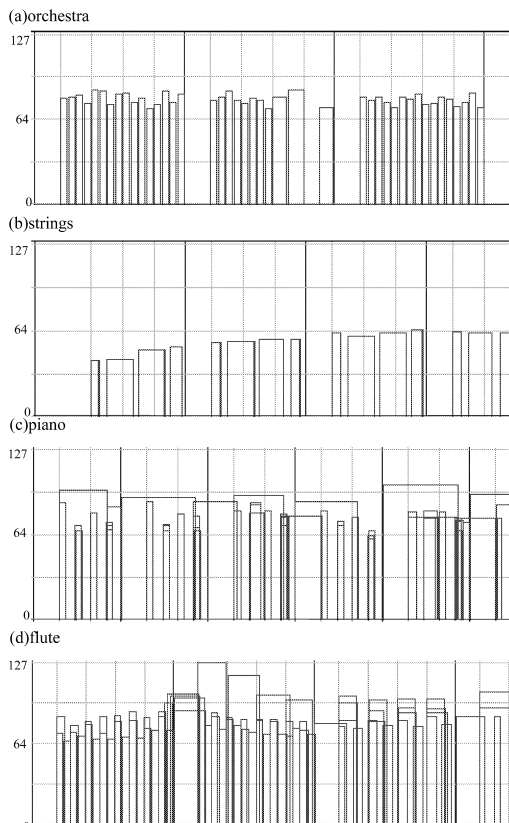


図 5 実験用データのノートオンベロシティ

Fig. 5 Value of note-on velocity on experimental data.

の情報を埋め込むことができた．また、すべての埋込みを施したデータから情報を完全に抽出できた．

次に、これらの埋込み処理が再生音質にどの程度影響を及ぼしたかについて調べた．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<sup>11),12)</sup>．

評価の基準は、原曲を 3 とし、質感が劣化している状態を 1、質感が向上している状態を 5 とした．また、評価の際には、埋込みを施した演奏 ( $M=1, 3, 5$ )、埋込みのない演奏 (ori.) および音の強さをすべて一定の値とした一様な演奏 (quant.) を被験者に外見的な区別ができないように提示した．このように、評価対象の中に原曲や表情付けのない曲を混入することで隠れ基準付きの評価ができる．すなわち、音の強さが一定の演奏を「音強成分に表情付けのない演奏」と見なし、表情のない楽曲の評価値にできる．また、原曲に対する評価値と埋込みのある楽曲への評価値を比較することにより、提案手法が表情付けに及ぼす影響の程度が分かる．すなわち、原曲より評価の低い演奏は品質が劣化しており、高い演奏は品質が向上していると見なすことができる．本実験の演奏には、YAMAHA 社製の MU-2000 を用いた．

### 3.1.4 評価結果と考察

実験の結果、表 4 の評価値が得られた．この結果から、従来手法は、埋込みビット数を増すに従い品質が低下している<sup>3)</sup> のに対し、提案手法では埋込み強度を高くしても演奏の品質を保持できることが分かる．また、次の考察を得た．

表 3 即時適応化方式の埋込み結果

Table 3 Embedding capacity of immediate adaptation method.

$M$	1 [bit]	3 [bit]	5 [bit]
orchestra	453	901	901
strings	36	75	84
piano	161	357	401
flute	270	620	659

表 4 即時適応化の評価値

Table 4 Mean opinion score of immediate adaptation.

$M$	quant.	ori.	1	3	5
orchestra	2.00	3.71	3.57	3.00	2.86
strings	1.71	3.00	2.71	3.00	3.29
piano	1.86	2.86	3.00	3.14	2.86
flute	1.57	3.14	3.57	3.86	3.43

- (1) orchestra 以外の実験用データについては、埋込みによる演奏表情への悪影響はない。埋込み強度を 5 とした場合においても原曲と比べて品質の劣化を感じた被験者はいなかった。一方、orchestra を除く実験用データでは、品質が向上していると評価できるものもあり、これは文献 3) に示されている品質向上と同様の効果が得られたものと考えられる。
- (2) orchestra では、埋込み強度が高くなるにつれて評価が低下している。この原因は、チャンネル間、つまり楽器間での音強バランスが埋込みにより乱れたためである。orchestra 以外の実験用データに含まれる楽器数は 3 以下であり、それに対して orchestra は 6 つの楽器で構成されている。提案手法では単一チャンネル内の発音の連続性のみを考慮して適応化を施している。このことがチャンネル間に乱れを生じさせたと考えられる。
- (3) flute では、埋込みによって評価が向上している。被験者からのコメントの中に「めりはりがついて聞きやすい」というものがあった。flute の原曲は 1 音ごとの音強変化が大きい。このため埋込みによるペロシティ変化が強調や抑制等の表情（めりはり）として解釈（認識）されたものと考えられる。このような変化を適切に制御する表情付けの手法については、今後検討する必要がある。

3.2 区間適応化法

提案手法による埋込み情報の不正な検出を困難にするための手法について示す。

即時適応化では、埋込みビット数を決定するために連続する 2 発音の差分値のみを用いている。したがって、即時適応化法をステガノグラフィとして用いるためには、既存の手法と同様に、あらかじめ埋込み情報を暗号化する等の対策が必要である。ここでは、第三者による情報の埋込み位置推定を困難にするために複数のペロシティ値を用いる手法について検討した。その基本となるアイデアは、同じチャンネル内で連続するノートオンのペロシティ制御量をランダムにグループ化して、適応化するものである。ここでは、同一チャンネルの連続するノートオンを一定の数（定数  $l$ ）で分割し、区間として扱う。それぞれの区間をグループとし、定数  $l$  をグループ化定数とする。 $l$  は情報埋込みと抽出のための鍵の 1 つである。

3.2.1 情報埋込み法

提案手法では、3.1 節の即時適応化法を拡張した。

埋込み処理の流れを図 6 に示す。

Step 1. グループ化定数を  $l$  とし、 $k$  番目のグループに情報を埋め込むものとする。 $k$  番目のグループの  $i$  番目の要素を

$$V_k(i) = v_{k \times l + i} \tag{8}$$

として、グループ内のすべての  $V_k$  に対してグループ直前の発音情報のオンペロシティとの差分値

$$D_k(i) = V_k(i) - V_{k-1}(l) \tag{9}$$

を求める。

Step 2. グループのオンペロシティ変化量の絶対値の平均

$$A_k = \frac{\sum_{i=1}^l |D_k(i)|}{l} \tag{10}$$

を求める。

Step 3.  $A_k$  を用いて、埋込み可能ビット数

$$n_k = \lfloor \log_2 A_k \rfloor \tag{11}$$

を求め、埋込み強度  $M$  と  $n_k$  から埋込みビット数

$$S_k = \min(M, n_k) \tag{12}$$

を得る。 $\min$  は 2 つの値を比較し最小値を返す関数である。

Step 4. 埋込みデータから  $S_k$  ビットを、埋込み情報  $E_k$  として取り出す。

$A_k$  の下位  $S_k$  ビットを  $E_k$  で置換し、 $A'_k$  とする。

Step 5. グループの各要素に対して

$$P_i = \frac{D_k(i)}{\sum_{i=1}^l |D_k(i)|} \tag{13}$$

を求め、埋込みを施した各オンペロシティを

$$v_{k \times l + i} = P_i \times A'_k \times l + V_{k-1}(l) \tag{14}$$

とする。

これは、埋込み対象を複数のオンペロシティに拡張する手法で、 $l = 1$  の場合、即時適応化法と同等にな

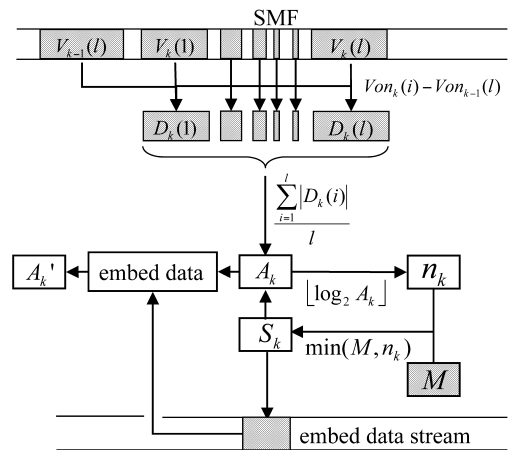


図 6 区間適応化法の埋込み処理  
Fig. 6 Schematic flow of segmental adaptation.

る．区間内の差分の割合が保持されるため全体の表情付けに影響を与えにくい特長がある．

### 3.2.2 情報抽出法

情報の抽出は，グループの平均を求める点を除き，即時適応化法と同様である．ただし，埋込み強度  $M$  とグループ化定数  $l$  を鍵としてあらかじめ共有しておくものとする．

Step 1. グループ化定数  $l$  をもとに， $k$  番目のグループから抽出する場合には，そのグループの  $i$  番目の要素を

$$V_k(i) = v_{k \times l + i} \quad (15)$$

として，グループ内のすべての  $V_k$  からグループ直前の発音情報のオンベロシティとの差分値

$$D_k(i) = V_k(i) - V_{k-1}(l) \quad (16)$$

を求める．

Step 2. グループのオンベロシティ変化量の絶対値の平均値

$$A_k = \frac{\sum_{i=1}^l |D_k(i)|}{l} \quad (17)$$

を求める．

Step 3.  $A_k$  を用いて，埋込み可能ビット数

$$n_k = \lfloor \log_2 A_k \rfloor \quad (18)$$

を求め，埋込み強度  $M$  と  $n_k$  から埋込みビット数

$$S_k = \min(M, n_k) \quad (19)$$

を得る．

Step 4.  $A_k$  の下位  $S_k$  ビットを埋込み情報  $E_k$  として抽出する．

この方式の場合，グループ化定数を共有しない第三者による情報抽出を困難にできる．

### 3.2.3 評価結果と考察

表 2 の実験用データを用いて，区間適応化法による埋込みを施したところ，表 5 のビット数を埋め込むことができた．さらに提案手法による音質への影響を調べるために MOS を求めた．評価者は 20 代の男女 7 名とした．評価対象については，即時適応化法 (imme.) を施した演奏 ( $M = 3$ ) と埋込みのない演奏 (ori.)，音強を一定の値にしたもの (quant.) ならびに  $l = 5$  のグループ化で区間適応化法 (segmental)

を施したもの ( $M = 3, 5$ ) の 6 種類とした．ただし quant. のベロシティの値は値域の中間値である 64 とした．

表 6 に示した実験結果から，次の考察を得た．

- (1) 区間適応化法と即時適応化法の評価値を比較すると，わずかに区間適応化法の方が良い評価を得ている．しかし全般的には同等の品質であるといえる．
- (2) quant. とその他の埋込みを施したデータを比較すると，表情の付いた演奏の方が高い評価を得ている．また，原曲と情報が埋め込まれている楽曲を比較した場合に差異がないことから，演奏表情に対する適応化が適切であったと見なせる．
- (3) orchestra についての評価がやや向上していることから，グループ化によってベロシティ変化に対する適応が即時適応化よりも安定したことが分かる．

これらの考察より，即時適応化法と区間適応化法は，ほぼ同等に品質を保持できるといえる．

### 3.3 各適応化法の比較

即時適応化法と区間適応化法の特長について比較する．

#### 3.3.1 埋込み容量

どちらの手法とも，埋込み容量はノート数（発音の数）に依存する．埋込み領域は即時適応化法がほぼノート数と一致するのに対し，区間適応化法ではノート数をグループ化定数で割ったものとはほぼ一致する．グループ化定数と埋込み容量は反比例する．

そのため埋込み容量を増やすためには即時適応化の方が有用である．

#### 3.3.2 音質に及ぼす影響

音質に関する MOS から，両手法とも，表情に及ぼす影響として有意な差は見られなかった．ただし，どちらの手法とも埋込みなしの実験用データと比べるとわずかに評価を高めている．

これらの結果から，少なくとも情報が埋め込まれたデータとそうでないものを聴感上で区別できないとい

表 5 区間適応化法の埋込み結果 ( $l = 5$ )

Table 5 Embedding capacity of segmental adaptation method ( $l = 5$ ).

method	segmental		
	imme.	3 [bit]	5 [bit]
$M$	3 [bit]		
orchestra	901	192	199
strings	75	18	19
piano	357	91	97
flute	620	137	144

表 6 区間適応化法の評価値

Table 6 Mean opinion score of segmental adaptation method.

method	-		imme.	segmental	
	quant.	ori.		3	3
$M$					
orchestra	2.71	3.00	3.00	3.00	3.14
strings	1.25	2.86	3.00	3.14	3.00
piano	1.71	2.86	3.00	3.25	3.25
flute	1.50	2.75	2.86	2.75	3.50

える．このため、演奏を聴いただけでは不自然さはなく、秘匿通信が行われていることを第三者に気付かれにくい．

また、埋込み対象を複数のペロシティの統計値とする区間適応化の方が表情を乱しにくいと考えられる．区間適応化のグループ内の表情については各成分の割合を保持するように適応化する．したがって、情報埋込み後の値が大きい場合でも、グループ内での変化が大きく乱れないため、バランス良く表情を強調できることになる．これに対し、即時適応化は1音ごとに埋め込むため、チャンネル数が多い楽曲の場合にチャンネル間の不整合を生じやすい．

### 3.3.3 情報抽出におけるセキュリティ

埋込みと抽出に埋込み強度  $M$  という情報しか共有しない即時適応化は、情報の不正な抽出が比較的やさしい．オンペロシティの値域は0~127であり、埋込み強度は7以下の自然数しかとれないためである．すなわち、総当たりの解読の可能性が高い．これに対して、区間適応化はグループ化定数を自由に設定できるため不正な抽出を困難にできる．また、処理手順からはグループ化定数を大きくしても表情に悪影響を及ぼしにくい．すなわちグループ化定数を乱数列にすることもできる．たとえば、乱数生成の鍵を共有鍵とすることで情報の抽出を困難にできる．

また、本提案手法は情報の埋込み領域が演奏情報そのものであり、SMFのファイル構造を変化させない．演奏の表情付けを考慮してこれに適応していることから、演奏に不自然な影響を及ぼすこともない．これらは、ステガノグラフィとして用いる場合、通信そのものを秘匿するのに有利な特長である．

## 4. おわりに

本論文では、楽音符号ステガノグラフィとして2つの手法を提案した．音強成分に適応する方法として即時適応化法を検討し、これが有効であることが確認できた．また、即時適応化法を拡張してグループ化を施し、セキュリティを高めた区間適応化法を検証し、これが質感へ及ぼす悪影響がないことを確認した．また、グループ化定数を乱数列とすることでステガノグラフィとして通信の秘匿に有用であるという結論を得た．

本論文で検討した手法は表情付けを崩さないように埋込みを施すものであった．評価の中で演奏への表情付けをやや向上できたものもあったが、本手法は基本的には表情成分それぞれの差分のみを考慮している．よって、現在のところ表情付けをうまく向上させることができる手法とはいえない．

今後は楽曲の演奏に関する構造分析等を行い、その表情付けのルールに適応しながら情報を埋め込む手法を検討する必要がある．楽曲構造の深い部分に情報を埋め込むことができれば、情報の改ざんや破壊を困難にできると考える．今後の検討事項として、次の3点をあげる．

- (1) グループ化手法  
時間情報をふまえた小節単位でのグループ化
- (2) ペロシティに対する埋込み  
グルーピング構造解析等の成果を利用した表情パターン認識とそれに適応する埋込み法の検討
- (3) 演奏の表情  
演奏表情を強調する埋込み手法の検討
- (4) 各種表情付けシステムの解析と利用  
表情付けアルゴリズムの利用やその出力コードへの偽装

## 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).
- 4) 井上大介，鈴木雅貴，岡庭隆文，松本 勉：演奏データファイル SMF に関するステゴ解析，2002 年暗号と情報セキュリティシンポジウム講演論文集，Vol.I, pp.313-318 (2000).
- 5) Inoue, D., Suzuki, M. and Matsumoto, T.: Detection-Resistant Steganography for Standard MIDI Files, *IEICE Trans. Fundamentals*, Vol.E86-A, No.8, pp.2099-2106 (2003).
- 6) 平賀瑠美：bit 別冊コンピュータと音楽の世界—基礎からフロンティアまで，pp.270-282, 共立出版 (1997).
- 7) 五十嵐滋：演奏を科学する，ヤマハミュージックメディア (2000).
- 8) 中島安貴彦：MIDI バイブル I，リットーミュージック (1997).
- 9) 中島安貴彦：MIDI バイブル II，リットーミュージック (1998).
- 10) 後藤真孝，橋口博樹，西村拓一，岡 隆一：RWC 研究用音楽データベース—研究目的で利用可能な著作権処理済み楽曲・楽器音データベース，情報処理学会論文誌，Vol.45, No.3, pp.728-738 (2004).
- 11) 古井貞照：デジタル音声処理，東海大学出版会 (1985).



- 12) 小澤一範：デジタル移動通信のための高効率音声符号化技術，トリケップス (1992).

(平成 18 年 2 月 8 日受付)

(平成 18 年 5 月 9 日採録)



山本紘太郎 (学生会員)

1978 年生．2001 年防衛大学校情報工学科卒業．2005 年から防衛大学校理工学研究科前期課程．音楽情報処理，情報セキュリティに関する分野に興味を持つ．



岩切 宗利 (正会員)

1970 年生．1993 年防衛大学校情報工学科卒業．1998 年防衛大学校理工学研究科情報数理専門卒業．1999 年防衛大学校情報工学科助手．2005 年同講師．博士 (工学)．マルチメディアと情報セキュリティに関する研究に従事．電子情報通信学会，映像情報メディア学会，画像電子学会各会員．