情報処理学会論文誌 Vol. 50 No. 6 1598–1609 (June 2009)

発音時間の揺らぎを利用したSMFステガノグラフィ

山本 紘太郎[†1] 岩切 宗利[†1]

情報ハイディングの応用分野として，通信秘匿を可能にするステガノグラフィがある．SMF（Standard MIDI File）を埋込みの対象としたSMFステガノグラフィの1つである演奏情報制御法には，埋込み可能情報量の増大が難しいという問題があった．本論文では，従来方式とは異なる成分を利用して情報を埋め込むSMFステガノグラフィを提案し，その埋込み能力と音質へ及ぼす影響を評価した結果について述べる．提案方式は，MIDIメッセージそのものではなく，演奏音の発音時間を用いて情報を表現するものである．提案方式を用いたシミュレーション実験では，情報をSMFに挿入しても，演奏の品質はほとんど低下しないことが分かった．また，提案方式の埋込み能力は従来方式の1.5〜2.1倍程度に向上することを確認した．

A SMF Steganography Based on Fluctuation of Duration

Kotaro Yamamoto[†1] and Munetoshi Iwakiri[†1]

Steganography is one of technique to conceal messages into usual digital media, and it makes communication invisible. Our technical proposal would be able to embed messages into Standard MIDI File (SMF) data stream. However, the size of embedded data is very small in the conventional technique, and shortage of the capacity is one of the problems to use our technique as steganography. In this study, we adopted an adaptive embedding method based on fluctuation characteristics of the duration for increase of the capacity. The embedding payload can be improved over two times as large as that of the conventional method, without deterioration of the sound quality. The experimental results show efficiency of the proposed method for steganography use rather than conventional techniques.

†1 防衛大学校情報工学科
Department of Computer Science, National Defense Academy

1. はじめに

ディジタルコンテンツの普及にともない，そのコンテンツへ密かに別の情報を埋め込み，それを別の用途に活用しようとする，情報ハイディング技術が注目されている．その応用として検討されているものの中に，電子透かしとステガノグラフィがある．電子透かしとは，ディジタル化によって違法な複製が容易になったコンテンツの著作権保護を，情報ハイディングによって実現しようとするものである．一般に，著作権管理情報や売買履歴など，コンテンツの流通経路を明らかにするような情報が埋め込まれる．ステガノグラフィとは，公開された通信や情報媒体の中に秘密情報を密かに挿入し，その情報の存在そのものを秘匿する技術である．すなわち秘匿通信の一方式であり，通信内容を非公開にする暗号技術とは別の観点から情報の守秘性を向上させることができる．通信の存在そのものを秘匿できる特性を持つため，コンテンツへは送受信者間にとって高度に秘匿すべき通信内容が埋め込まれる．

ステガノグラフィ研究の主な課題として，ステゴ解析，つまり情報が埋め込まれている事実を何らかの解析によって検知されることへの対策，ならびに埋込み可能情報量の増大があげられる．本研究では，音楽の演奏情報を記録するSMF（Standard MIDI File）[1]に対する情報ハイディングをステガノグラフィの観点から再検討した．

SMFへのステガノグラフィとしては，井上ら，遠山らによるデータ構造を利用する方法[2)–4)]と，岩切らによる演奏情報のパラメータを制御する方法[5),6)]が報告されている．文献2)の手法では，SMFのデータ構造に冗長性を見い出し，演奏そのものに影響を及ぼさずに情報を埋め込むことを可能にしている．しかし，この手法のみでは，データ構造に埋込みの痕跡が残るため，ステゴ解析されやすい．その対策として，ステゴ解析を難しくする手法が文献3)，4)に報告されている．一方，岩切らは文献5)に，SMFに含まれる発音の強さを表すベロシティのパラメータに対し，情報を埋め込む手法を示した．この手法は，埋込みによる痕跡をデータ構造へ残さないという特長を持つ．ただし，埋込みにより演奏そのものが変化することになる．そのため文献6)に，演奏に付加された表情付けに適応し，演奏品質への影響を抑制する手法を示した．この手法は，データ構造に埋込みの痕跡を残さないので，データ構造に注目したステゴ解析を受けにくいという特長を持つ．

これらの2種類の手法はそれぞれの特長を活かしながら同時に用いることもできる．しかし，これらのSMFステガノグラフィは，他のコンテンツを対象とした手法に比べ，埋込み可能な情報量が少ない．

本研究では，この課題を解決するため，より値域の広い発音時間（デュレーション）に対

© 2009 Information Processing Society of Japan

して情報を埋め込む SMF ステガノグラフィ技術の開発に取り組んだ．本論文に示す提案方式は，音楽の実演奏で付加される抑揚や小さな揺らぎなどに見せかけるステガノグラフィである．特に，値域の広いデュレーションを利用し，埋込み情報量を増大させることができた．また，従来手法とは異なる成分を用いて情報表現するため，従来手法とも併用できる．

2 章では MIDI と SMF の概要について示し，3 章で関連研究について述べるとともに従来方式の課題を明らかにする．4 章では提案方式の詳細について述べ，5 章では，埋込み能力とその品質についての評価結果とその考察を述べる．

2. MIDI と SMF

MIDI 規格（Musical Instrument Digital Interface）とは，電子楽器を制御するための規格である[1)]．楽器の制御には 2〜3 byte の制御符号（MIDI メッセージ）を用いる．この演奏情報記録の一方式として SMF（Standard MIDI File）がある．これは，MIDI の制御符号を，デルタタイム[⋆1]とともに MIDI イベントとして記録するものである．

2.1 SMF の発音法

MIDI において発音を指示する符号は 3 バイトのノートメッセージである．note-on，note-off の 2 つがあり，発音時に note-on を送信し，その音を消すために note-off を送信する．SMF では，MIDI メッセージの発行時刻をデルタタイムで表現する．**図 1** のように，note-on の発行時刻と note-off の発行時刻の差をとることで，その発音のデュレーションが得られる．SMF における基本的な発音のパラメータは次に示す 4 つである．

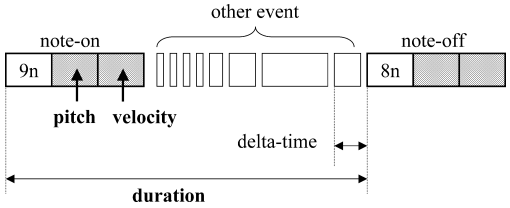

図 1　SMF の発音情報パラメータ
Fig. 1　Note-message prameters of SMF.

⋆1 MIDI 制御符号を発行するタイミング情報．直前のイベントからの時間差として記録される．

(1)　音　　色
　プログラムチェンジ[⋆2]によってあらかじめ各チャネルに楽器（音色）を割り当てる．
(2)　音の高さ（pitch）
　音の高さを表す 7 ビットの値域を持つパラメータであり，note-on メッセージに含まれる．
(3)　ベロシティ（velocity）
　音の強さを表す 7 ビットの値域を持つパラメータであり，note-on メッセージに含まれる．
(4)　デュレーション（duration）
　発音時間であり，note-on，note-off 間の発行時刻の差によって表される．

2.2 SMF の時間管理

SMF では，時間（デルタタイム）の単位に tick を用いる．次の 2 つの情報によって秒単位の時間を tick 単位に相互変換できる．

(1)　4 分音符あたりの時間分解能
　SMF のヘッダチャンク[⋆3]に定める 4 分音符あたりの分解能（division）
(2)　4 分音符あたりの実時間
　トラックチャンク[⋆4]中に記録されるセットテンポメタイベント[⋆5]

たとえばセットテンポメタイベントで 4 分音符あたりの時間が 500,000 μsec と定められ，division が 480 tick/beat であった場合，1 tick あたりの時間の長さ t [msec/tick]（以下，tick-time とよぶ）は，

$$t = \frac{500000}{480 \times 1000} = 1.04 \quad [\text{msec/tick}] \tag{1}$$

になる．SMF では，このようにして定まる tick-time により MIDI イベント中のデルタタイムを規定する．そのため，ヘッダチャンク中で決定する 4 分音符あたりの分解能を変化させることにより，デルタタイムの大きさは変化する．デルタタイムは 28 ビットの値域を持つことができ，かつデュレーションは note-on，note-off 間に発行される MIDI イベントのデルタタイムの総和で表される．つまり，デュレーションの値域は可変である．

⋆2 MIDI メッセージの 1 つ．チャネルに対し演奏すべき楽器を割り当てる．
⋆3 フォーマットの種類やトラック数，時間管理情報を記録するヘッダ領域
⋆4 実際の演奏情報が記録されるデータ領域
⋆5 4 分音符あたりの時間をマイクロ秒単位で記録する制御情報

© 2009 Information Processing Society of Japan

3. 関 連 研 究

井上らは，文献 2) で SMF のデータ構造の等価性を利用したステガノグラフィを提案している．SMF では，和音などの同時に発音される発音情報やコントロールチェンジ[⋆1]は，デルタタイムの値 “0” で連続して記録される．また，この順序を任意に入れ替えても演奏に変化を及ぼさない．井上らはこれを利用して，同時発行される MIDI イベントの記録順序のパターンに情報を割り当て，情報埋込みを実現している．さらに文献 2) では，インターネット上で配布されている数百曲の SMF に対して実際に情報を埋め込み，その埋込み能力を評価している．本研究では，埋込み能力評価の指標として，埋込み率を用いた．埋込み率 P_r[%] は，埋込み対象の SMF のファイルサイズ f_s [bit] と埋込み可能情報量 e_l [bit] から，次式により求める．

$$P_r = \frac{e_l}{f_s} \times 100 \tag{2}$$

文献 2) の手法では，平均 1%の埋込み率を達成できることを実験により明らかにしている．文献 2) の手法では，MIDI イベントの配列によって情報を表現している．この配列は一般に編集ソフトウェアによって決定されるものなので，情報の埋込みによって不自然に配列される場合がある．これはステゴ解析の糸口になるため，文献 3)，4) では，埋込み能力の低下と引きかえに，ステゴ解析への対策として配列が不自然にならない工夫を施している．

岩切らは文献 6) で，演奏情報を制御し，ノートメッセージ中のベロシティ値に情報を埋め込む手法を提案している．特に，情報埋込みによる演奏への影響を抑制するため，隣接音のベロシティ差分値の LSB（Least Significant Bit）を埋込み情報で置換することにより情報埋込みを実現することに特長がある．また，発音情報をグループ化し，その代表値に対する埋込みを行うことで秘匿性を向上させた．ただし，文献 6) では埋込み率による埋込み能力の評価を十分に行っていなかった．そこでまず，約 200 曲の SMF を用いた実験によりこの方式の埋込み能力を評価した．埋込み能力の計測には，RWC 研究音楽データベース[7] の SMF を用いた．RWC 研究音楽データベースには，クラシック 61 曲，ジャズ 50 曲，ポップス 100 曲，著作権切れ音楽 15 曲の 4 つのカテゴリ，合計 226 曲の楽曲が登録されている．**表 1** に，本実験の結果得た埋込み率および文献 2) の 8 音対応ステゴ鍵を用いた埋込み能力を示す．表中の event が文献 2) の結果を，velocity が文献 6) の結果を示す．表中の

表 1　従来方式の埋込み能力
Table 1　Embedded ratio of the conventional method.

method	velocity [%]				event [%]
M	3	5	5	5	-
L	1	1	5	10	-
maximum	3.00	3.58	0.69	0.33	4.01
minimum	0.01	0.01	0.00	0.00	0.04
average	1.51	1.72	0.35	0.17	1.21

表 2　リズムチャネル電子透かしの埋込み能力 [%]
Table 2　Embedded ratio of rhythm-channel watermark [%].

method	maximum	minimum	average
rhythm	6.70	0.13	2.98

M，L は埋込みパラメータであり，「L 個のベロシティ値の差分平均へ埋め込む情報ビット数の上限を M ビットとする」ことを表す．

この表から，文献 6) の方法による平均埋込み率は，最大で 1.7%までである．これは，埋込みに用いるベロシティ成分の値域が 7 ビットと狭く，埋込み容量を適応制御しているためである．また，秘匿性を高めるために L の値を大きくするほど，埋込み率は低下することが分かる．特に，MIDI メッセージのパラメータの上限が 7 ビットである制約は大きく，多量の情報埋込みには，別の手法で情報を埋め込む必要がある．そのため，本研究では，可変な値域を持つデュレーションに注目した．

デュレーションの大きさにより埋込み情報を表現する情報ハイディングの基本的なアイディアは，すでに電子透かしとして文献 5) に示されている．文献 5) の手法の特長は，埋込み対象の発音情報を，ノートオンドリブン[⋆2]であるリズムチャネル中のパーカッション音に限定することにより，演奏音に影響を及ぼさない情報埋込みを実現した点である．この手法を，ステガノグラフィの観点から再評価した．

まず，埋込み能力を評価するため，RWC 研究音楽データベースの楽曲 226 曲のうち，リズムチャネルの存在する 141 曲を用いて埋込み能力を調べた．**表 2** の結果から，文献 5) の埋込み率は平均約 3%であり，高い埋込み能力を持つことが分かった．

次いで，ステゴ解析への耐性について考察した．文献 5) の手法は演奏情報を直接制御しており，SMF のデータ構造に埋込みの痕跡を残さない．そのためデータ構造の特異性を用

⋆1 ビブラートやピッチの細かな変化など，演奏音に変化を与える MIDI メッセージ

⋆2 ノートオフで消音されず，自然減衰する音色

© 2009 Information Processing Society of Japan

いた解析への耐性は十分であると考えられる．しかし「ノートオンドリブンである発音情報のデュレーションが不規則な値を持つ」という事実には不自然さがある．ノートオフにより消音しないのであれば，そのデュレーションは一定値でよいためである．したがって，リズムチャネルの発音情報に対し，そのデュレーションを観測することにより情報埋込みの存在を検知されるおそれがある．また，リズムチャネルの発音情報をノートオンドリブンとするのは GM（GENERAL MIDI）規格[⋆1]に準拠した楽器に限定されるため，準拠していない楽器で演奏した場合，演奏音の異常から情報埋込みの存在を知られる可能性もある．

以上の考察から，ステゴ解析への耐性を考慮すると，ノートオンドリブンという特殊な発音情報のみを情報埋込みに用いる従来手法は，ステガノグラフィとして用いるには不適当である．ただし，デュレーションを埋込み対象とすることにより埋込み能力を向上させることができる可能性は高いと考えられる．

4. 提 案 方 式

従来方式の課題は，ノートオンドリブンであるリズムチャネルの発音情報のみを用いた情報の埋込みに起因する．これを解決しつつ高い埋込み能力を実現するため，埋込みによる演奏音の変化を許容するとともに，カバー SMF のデュレーションが持つ特徴を保持した SMF ステガノグラフィを提案する．

4.1 基本アイディア

図 2 は，RM-C002.mid [7] の楽曲中（4 分音符あたり分解能 480）のデュレーションの分布を表したものである．横軸にデュレーションを表し，縦軸に各値の出現頻度を示した．この図から，デュレーションは 960，480，360 前後に集中していることが分かる．これは，RM-C002.mid のデルタタイム分解能が 480 であることから，デュレーションそれぞれが 2 分音符，4 分音符，付点 8 分連符の音価の周辺にばらつきを持って偏在するという事実と合致する．

ここで，実際に演奏され，記録されるデュレーションは，楽譜上の音価からわずかにずれていることに注目する．このばらつきは，演奏者または制作者によって付加された抑揚や揺らぎと見なせる．

提案方式では，このデュレーションの揺らぎを情報の埋込み（表現）に利用する．すなわち，デュレーションを「音価に抑揚や揺らぎが付加されたもの」ととらえ，SMF ヘッダから得られる音価からのずれを信号として，埋込み情報を表現する．

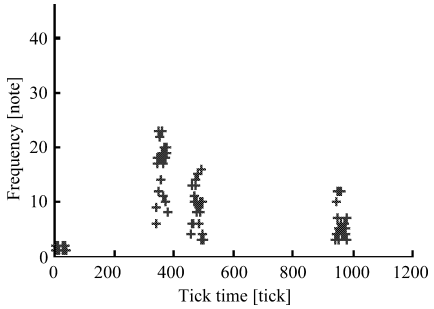

図 2　デュレーションのばらつき
Fig. 2　Fluctuation of duration.

4.2 情報埋込みの処理手順

4.2.1 前　処　理

提案方式では，SMF の各構成音を複数の音符の区間に分類し，各構成音に対し音符の音価[⋆2]からの揺らぎを埋込み情報で置換することにより情報の埋込みを実現する．ただしこのとき，埋込みによる演奏品質への影響を抑制するため，各音符の揺らぎの程度を標準偏差として求め，その値を参照して各音符の区間ごと埋込みビット数を決定する．

(1)　埋込み対象発音情報の選別

次の条件を満たす発音情報に関しては，埋込み処理を省略する．

- リズムチャネルに含まれる発音情報
  3 章の考察から，ノートオンドリブンである発音情報が，不規則なデュレーション値を持たないようにするため，リズムチャネルでは埋込みを行わない．
- 同一チャネルの同一音が連続している区間の発音情報
  デュレーション値の制御により，SMF 中のノートオフメッセージの発行時刻は変化する．このとき，メッセージの発行順序が入れ替わり，ノートオンとノートオフの対応関係が崩れる可能性がある．そのため，情報の埋込みによりノートオンとノートオフの対応がとれなくなる発音情報への埋込みは行わない．

以下の処理は，埋込み対象の発音情報のみを用いて行う．

⋆1 MIDI 音源の互換性を向上させ，演奏の再現性を高めることを目的に定められた音源仕様

⋆2 音楽において，ある音符または休符に与えられた楽譜上の時間の長さ

© 2009 Information Processing Society of Japan

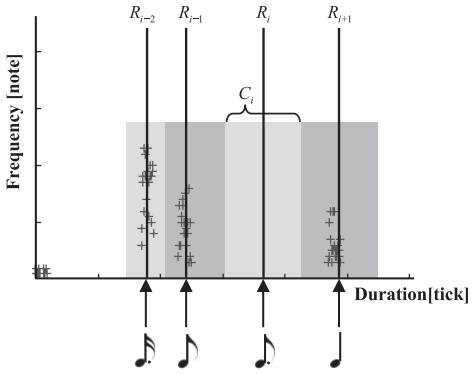

図 3　デュレーションの区分
Fig. 3　Classification of duration as notation.

(2)　揺らぎ標準偏差の算定

図 3 のように，隣り合う音符の音価間距離を 2 分割してカバー SMF のデュレーションを区分する．その際，区間の構成音のデュレーションの平均，標準偏差をそれぞれ求める．

Step 1.　SMF ヘッダから 4 分音符あたりの分解能 T を得る．

Step 2.　T を基準として，全音符，2 分音符，8 分音符，16 分音符，32 分音符，64 分音符の音価をそれぞれ得る．同様に，付点 2 分音符，付点 4 分音符，付点 8 分音符，付点 16 分音符，付点 32 分音符，付点 64 分音符の音価を得る．

Step 3.　各音価を小さい順に配列し，R_i（$0 \leq i \leq 13$）とする．

Step 4.　i それぞれに対し

$$D_i = \frac{R_i - R_{i-1}}{2} - 1 \tag{3}$$

$$U_i = \frac{R_{i+1} - R_i}{2} \tag{4}$$

を求める．

Step 5.　カバー SMF のデュレーションから得たヒストグラムを値域 $[R_i - D_i, R_i + U_i]$ に分割する．この値域に含まれたサンプル全体を C_i とする．

Step 6.　C_i の平均値 E_i，標準偏差 σ_i を求める．

(3)　埋込みビット数の決定

次いで，埋込みビット数を決定する．このとき，埋込み後の推定標準偏差をあらかじめ

表 3　埋込み後の推定標準偏差
Table 3　Estimated standard deviation of fluctuation after embedding process.

n	1	2	3	4	5	6	7	8
A_n	0.78	1.87	4.18	8.80	18.04	36.51	73.46	147.36

埋込みビット数ごと求め，それらを埋込み前の標準偏差と比較し，値が最も近くなる埋込みビット数を採用する．

提案方式では，埋込み情報をあらかじめ暗号化して用いる．暗号化した情報系列から n ビットの情報を取り出すとき，その値域は 0 から $2^n - 1$ である．暗号化された情報系列が一様と見なせるとき，各値の出現確率は $\frac{1}{2^n}$ である．このとき，0 を平均値とした推定標準偏差

$$A_n = \sqrt{\frac{1}{2^n} \sum_{k=0}^{2^n-1} k^2} \tag{5}$$

となる．これを各埋込みビット数に対して求める．**表 3** は，$n = 1, \cdots, 8$ における A_n を求めたものである．A_n を用いて，各区間の埋込みビット数を次式により求める．

$$L_i = p \quad (p : \min(|\sigma_i - A_p|)) \tag{6}$$

このとき，埋込み前後で各発音情報の属する音符の区間を変化させないために，L_i を $2^{L_i} \leq D(i)$，$2^{Li} \leq U(i)$ になるように定める．ただし各区間に属するサンプルの個数が少ない場合，埋込み後の標準偏差が適切に求まらない可能性がある．そのためサンプルの個数が T_h 個以下の区間に対しては情報の埋込みを行わない．

$$L_i = 0 \quad (C_i \leq T_h) \tag{7}$$

埋込みおよび抽出の際には，情報を送受するエンティティ間で，ステゴ鍵として次の 3 つの情報をあらかじめ共有しておく必要がある．すなわち，埋込み情報の暗号化鍵，埋込み前の各区間の平均値 E_i，サンプル数の下限 T_h である．提案方式では，E_i はカバー SMF によって異なる系列となるため，カバー SMF ごとに異なるステゴ鍵を用いる必要がある．

4.2.2　埋込み処理

図 4 の k 番目のデュレーション情報 d_k に情報を埋め込む処理について示す．

Step 1.　カバー SMF に含まれるデュレーション d_k を得る．

Step 2.　d_k を C_j に区分する．埋込みビット数 $N_k = L_j$ とし，埋込み制御子 F_k を次のとおり決定する．

© 2009 Information Processing Society of Japan

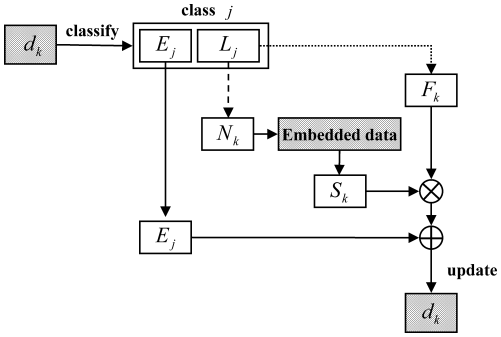

図 4　提案方式の埋込み処理
Fig. 4　Embedding process of the proposed method.

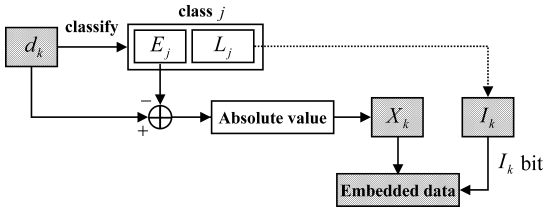

図 5　提案方式の抽出処理
Fig. 5　Extracting process of the proposed method.

$$F_k = \begin{cases} 1 & (E_j \leq d_k) \\ -1 & (d_k < E_j) \end{cases} \tag{8}$$

Step 3.　K によりあらかじめ暗号化した情報系列から，N_k ビットの情報 S_k を得る．
Step 4.　$N_k \neq 0$ のとき，次式により d_k を更新する．

$$d_k = E_j + F_k \times S_k \tag{9}$$

4.3　情報抽出の処理手順

情報抽出の処理手順は，情報埋込み処理手順とほぼ同様である（**図 5** 参照）．ただし，埋込みに用いたサンプル数の下限 T_h と，埋込み前の各音符の区間の平均 E_i を抽出鍵として用いる．

4.3.1　前　処　理

4.2.1 項と同様にカバー SMF の各デュレーション d_k を C_i に区分し，抽出鍵 E_i を各区間の平均値と見なしてそれぞれの標準偏差 σ_i を求める．埋込みビット数 L_i を次式により求める．

$$L_i = p \quad (p : \min(|\sigma_i - A_p|)) \tag{10}$$

このとき，L_i を $2^{L_i} \leq D(i)$，$2^{L_i} \leq U(i)$ になるように定める．ただし各区間に属するサンプルの個数が T_h 個以下の区間に対しては，

$$L_i = 0 \quad (C_i \leq T_h) \tag{11}$$

とする．

4.3.2　抽 出 処 理

図 5 に，k 番目のデュレーション情報 d_k からの情報抽出処理を示す．
Step 1.　デュレーション d_k を得る．
Step 2.　d_k を C_j に区分し，抽出ビット数 $I_k = L_j$ とする．
Step 3.　I_k ビットの値 $X_k = |d_k - E_j|$ を情報として読み取る．

5.　提案方式の評価

5.1　埋込み情報の抽出精度

提案方式は，埋込みビット数の決定に各音符の区間の標準偏差値を用いている．また，埋込み情報は平均値 E_i との差分で表現されるので，情報埋込み前と埋込み後では，標準偏差値が変化する（**表 4** 参照）．埋込み情報系列に分布の偏りがある場合，抽出時に各区間の標準偏差値が表 3 の推定値に近づかない可能性がある．その場合，正しく埋込みビット数が定まらず，情報の抽出に失敗する．この偏りは，特にサンプル数が少ない区間に生じやすい．これを避けるため，提案方式ではサンプル数の下限 T_h を設定している．ここでは，埋め込んだ情報の抽出精度が，T_h の値によってどの程度変化するかを評価した．

5.1.1　評価方法と実験用データ

RWC 研究音楽データベース[7] の SMF 226 曲に対して T_h を変化させながら埋め込み，抽出実験を行い，情報の抽出に失敗する SMF の数をそれぞれ計測した．計測の際には，埋込み情報として ASCII テキストファイルを DES（Data Encryption Standard）暗号で暗号化したものを用いた．また，暗号化された情報系列の違いによる抽出精度の変化を評価するため，埋込み情報を複数の鍵を用いて暗号化し，それぞれの抽出精度を計測した．計測結果として**図 6** を，このときの平均埋込み率の変化として**図 7** を得た．

© 2009 Information Processing Society of Japan

表 4　埋込み前後の揺らぎ標準偏差の変動
Table 4　Alteration of standard deviation by embedding.

M	CoverSMF	StegoSMF
$\sigma(0)$	0	0
$\sigma(1)$	0	0
$\sigma(2)$	1.83	1.83
$\sigma(3)$	0	0
$\sigma(4)$	5.56	4.31
$\sigma(5)$	5.57	4.24
$\sigma(6)$	11.00	8.90
$\sigma(7)$	14.81	8.00
$\sigma(8)$	8.83	8.87
$\sigma(9)$	25.48	16.11
$\sigma(10)$	31.47	35.33
$\sigma(11)$	115.79	72.01
$\sigma(12)$	91.37	65.11
$\sigma(13)$	191.63	191.63

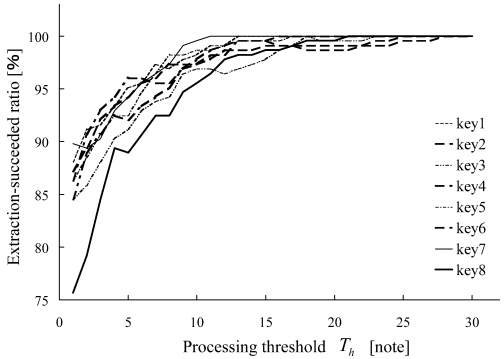

図 6　T_h の変化に応ずる抽出成功率
Fig. 6　Extraction-succeeded ratio with processing threshold T_h.

5.1.2　評価結果と考察

図 6 の結果から，どの鍵においても，T_h の値が大きくなるほど情報抽出に成功する SMF 数は増大し，$T_h = 30$ では 100%になることが分かった．この結果から抽出精度は十分であるといえる．

また図 7 からは，T_h の増大により，埋込み能力がわずかながら低下することが分かる．

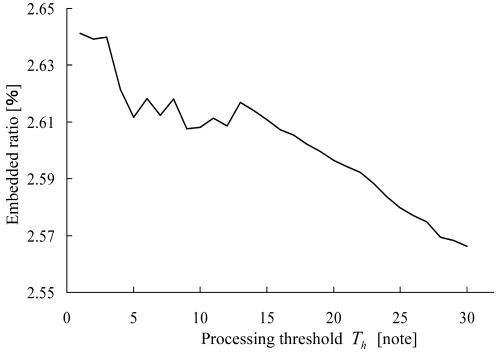

図 7　T_h の変化に応ずる平均埋込み率
Fig. 7　Embedded ratio with processing threshold T_h.

表 5　提案方式の埋込み率（$T_h = 30$）
Table 5　Embedded ratio of the proposed method ($T_h = 30$).

method	event [%]	velocity [%]	proposed [%]
maximum	4.01	3.58	5.55
minimum	0.04	0.01	0.23
average	1.21	1.72	2.57

これは，T_h が大きくなるに従い，埋込み対象の発音情報が減少するためである．ただし，$T_h = 1$ の場合においても 75%程度の SMF では情報の抽出に成功している．このことから，少しでも大きな情報を埋め込みたい場合には，情報の埋込み時にその抽出の可否をあらかじめ検査するといった運用により，低い値の T_h を用いた埋込みも可能と考えられる．

これらの結果から，提案方式はステガノグラフィとして用いるのに十分な抽出精度を有するといえる．

5.2　埋込み能力

埋込み能力を評価するため，RWC 研究音楽データベース[7] の SMF を用いて提案方式の埋込み率を計測した．計測の際，抽出精度が 100%となる $T_h = 30$ として埋込みを行った．

計測実験の結果，**表 5** の埋込み率での情報埋込みが可能であることが分かった．ただし，表中の event は文献 2) の 8 音対応ステゴ鍵を用いた埋込みの，velocity は文献 6) の $M = 5$, $L = 1$ による埋込みの結果を示したものである．この実験では，文献 2) と比較して平均 2.1 倍，文献 6) と比較して平均 1.5 倍の埋込み能力を達成した．このことから，提案方式

© 2009 Information Processing Society of Japan

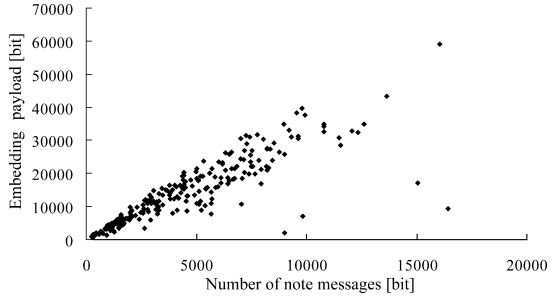

図 8　ノート数と埋込み可能情報量
Fig. 8　Number of note message and embedding payload.

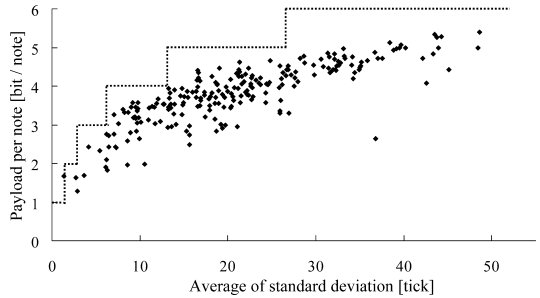

図 9　1 音あたり埋込み可能ビット数と平均揺らぎ標準偏差
Fig. 9　Payload per note and average of standard deviation.

表 6　提案方式のノート利用率（$T_h = 30$）
Table 6　Note availability ratio of the proposed method ($T_h = 30$).

method	ratio [%]	SD
maximum	100	-
minimum	7.05	-
average	72.7	23.0

の埋込み能力が高いことが分かる．

さらに詳細に埋込み能力を検討するため，SMF 中の発音情報，時間分解能が，それぞれ埋込み可能情報量にどう影響するかに関しての考察を行った．

5.2.1　発音情報による影響

提案方式の情報埋込み処理単位は 1 つの発音情報である．そのため，埋込み可能情報量は SMF 中の発音情報の数（以下，ノート数とよぶ）に比例するはずである．実際に，埋込み能力の計測結果について，埋込み実験に用いた SMF の埋込み可能情報量とノート数との関係を**図 8** に示した．縦軸が埋込み可能情報量，横軸がノート数を表している．この図からは，比例関係を読み取ることができる．しかし埋込み可能情報量のばらつきが大きく，ノート数の多い SMF でも埋込み可能情報量の少ない楽曲が存在することが分かった．

この原因として，まず，埋込みの前処理において埋込み対象ノートを選別していることがあげられる．**表 6** に，提案方式のノート利用率を示す．表中の ratio は利用率を，SD は平均利用率の標準偏差を表している．この結果から，提案方式のノート利用率にばらつきが大きいことが分かる．このため，より高い埋込み率を実現するためには，前処理における発音情報の除外条件をできるだけ満たさないカバーデータを用いるほうが良い．すなわち，リズムチャネルの発音情報の割合が少なく，同一発音が連続する区間の少ない SMF を用いることが望まれる．

次いで考えられる原因は，発音時間の揺らぎが少ないことである．提案方式では，各区間の揺らぎの標準偏差に基づいて各区間の埋込みビット数を定めた．そのため，各区間内の揺らぎが小さい場合，その区間に割り当てられる埋込みビット数が少なくなる．

図 9 に，埋込み能力の計測結果（$T_h = 30$）から，1 音あたりの平均埋込みビット数とカバー SMF の平均揺らぎ標準偏差との関係を示す．ここで 1 音あたりの平均埋込みビット数は，埋込み可能情報量と情報の埋込みに用いられたノート数のみより求めた．これは揺らぎの影響のみを正しく評価するためである．

提案方式では，式 (6) によって埋込みビット数を定めている．図 9 の各計測値は，式 (6) により定まる理論値（破線）に近い分布を示すことが分かる．よって，より多くの情報を埋め込むためには，発音時間に揺らぎの多く含まれたカバーデータを用いることが望ましいといえる．本実験に用いた RWC の楽曲データは，実演奏に基づいて制作された SMF であるため，リアルタイム入力[⋆1]された SMF に近く，揺らぎが多く含まれていた．しかし，編集ソフトウェアを用いてステップ入力[⋆2]された SMF は，デュレーションが一定値に設定されやすく，揺らぎの少ない SMF になる場合が多い．このような SMF には，提案方式は適さ

⋆1 MIDI 楽器などを用い演奏者の実演奏を基準に作成する入力法
⋆2 音を 1 音ずつ，パラメータとともに打ち込む入力法

© 2009 Information Processing Society of Japan

表 7　編集ソフトウェアによる division の違い
Table 7　Division in SMF generated by music sequencer.

Format type	Format 0 [tick]	Format 1 [tick]
SingerSongWriter Lite 5.0	480	480
XG works Ver.4.07J	480	480
SONAR LE Ver.4.0.1	960	960

ないことになる．

また，揺らぎのまったく含まれない SMF に対し，提案方式では各デュレーションに対して 1 ビットの埋込みビット数を与えるようにした．これは揺らぎのない演奏に揺らぎを付加することによる，演奏品質への影響評価が不十分なためである．よって，揺らぎのない SMF に対する埋込み能力は低くなる．ただし，揺らぎのない SMF には，同時に発行される発音情報の頻度が増大するため，文献 3)，4) の方法を適用できる．すなわち，提案方式と文献 3)，4) の方式では，多量の情報埋込みに適する SMF が異なるので，これらをうまく併用することが望ましい．

5.2.2　時間分解能による影響

SMF の時間単位である tick の長さは，SMF ヘッダの division によって定義され，その値は制作者や演奏者によって変更可能である．この特徴を利用すれば，時間分解能を大きくすることにより埋込み可能情報量を線形に増大させることができる．しかし，提案方式を実運用するためには，ステゴ SMF と埋込みなしの SMF が区別できないことが重要である．その観点に基づいて時間分解能を検討すると，ステゴ SMF の時間分解能が，流通している SMF のそれと顕著に異なった値を持つことはステゴ解析の糸口になると考えられる．

一般に流通する SMF は市販の編集ソフトウェアで制作する場合が多い．そこで，主要なソフトウェアが SMF を出力する際にどの程度の大きさの時間分解能を設定するかを調べ，本研究で用いた SMF と比較した．

一般的に用いられる編集ソフトウェアの出力した SMF の division 値を**表 7** に示す．本実験に用いた編集ソフトウェアは「YAMAHA XGworks Ver.4.07J ⋆1」，「Roland SONAR LE Ver.4.0.1 ⋆2」，「Internet SingerSongWriter Lite 5.0 for Windows ⋆3」の 3 つである．この調査結果から，本実験で用いた編集ソフトウェアが標準的に出力する時間分解能は 480 tick 以上の値であることが分かった．本研究の埋込み能力評価の実験に用いた楽曲は，すべて 4 分音符あたり 480 tick の分解能を持つ楽曲であった．

したがって，編集ソフトウェアが出力する SMF の division は，本研究の実験データのそれと同じか，より大きい値であることが分かった．すなわち一般的な SMF へ提案方式を適用した場合，表 5 の結果以上の埋込み能力を持つと考えられる．

⋆1 http://www.yamaha.co.jp/product/syndtm/p/soft/xgwv4w/
⋆2 http://www.cakewalk.jp/Products/SonarLE/
⋆3 http://www.ssw.co.jp/products/ssw/win/sswlt50w/index.html

5.3　ステゴ鍵

提案方式では，情報を送受するエンティティ間で，3 つの情報をステゴ鍵として共有する，すなわち，埋込み情報の暗号化鍵，サンプルの下限値 T_h，埋込み前の各音符のデュレーション平均値 E_i である．

5.3.1　ステゴ鍵の特徴

それぞれの鍵情報の特徴について，次に示す．

(1)　埋込み情報の暗号化鍵

提案方式は，埋込みの際に埋込み情報をあらかじめ暗号化し，その一様性を利用して埋込み後の推定揺らぎ標準偏差を求め，埋込みビット数を決定している．この工夫により，情報抽出にデュレーションの各区間の揺らぎ標準偏差を利用可能にした．もし，埋込み情報としてテキストデータを暗号化せずに用いた場合には，文字コードの偏りによって埋込み結果に偏りを生じる可能性がある．すなわち，埋込み情報系列の偏りによって標準偏差の偏りが生じ，それによって，埋込みビット数を正しく求めることが難しくなる．

(2)　サンプルの下限値 T_h

T_h は，情報の抽出を正しく行うために用いられる．提案方式で高い埋込み率を実現するためには，情報抽出が可能な範囲で，できるだけ T_h を小さくする必要がある．

(3)　埋込み前のデュレーション平均値 E_i

提案方式では，埋込み情報は E_i からの偏差として埋め込まれる．埋込み前後で各デュレーション平均値が同じ値になるとは限らないため，情報の抽出には E_i が必要になる．E_i はカバー SMF に付加された抑揚や揺らぎに依存するため，E_i はカバー SMF ごとに異なる系列となる．すなわち提案方式では，カバー SMF ごとに異なるステゴ鍵を使用する必要がある．

5.3.2　ステゴ鍵の強度

T_h，E_i はすべて自然数であり，しかも E_i は埋込み後のデュレーション平均値とも比較的近い値を持つため，ステゴ鍵としての強度は不十分である．そのため，提案方式におけるステゴ鍵の強度は，埋込み情報の暗号化に用いる暗号方式，およびその鍵長に依存する．

© 2009 Information Processing Society of Japan

5.4　ステゴ解析への耐性

SMF ステガノグラフィに対するステゴ解析としては，データ構造の特異性を利用した解析，デュレーション値の特異性を利用した解析，デュレーション分布の観測による解析の 3 つのアプローチが考えられる．各解析について，次のことがいえる．

(1)　データ構造の特異性解析

提案方式は演奏情報を直接制御するステガノグラフィであり，SMF のデータ構造に埋込みの痕跡を残さない．そのため，データ構造の特異性を用いた解析への耐性を持つと考えられる．

(2)　デュレーション値の特異性解析

デュレーション値の特異性解析とは「ノートオンドリブンである発音情報のデュレーションが不規則な値を持つのは不自然である」という事実を利用した解析である．

提案方式では，リズムチャネルの発音情報を埋込み対象から除外しているため，リズムチャネルのデュレーション値を観測することによる解析に耐性を有する．

(3)　デュレーション分布の観測による解析

提案方式では，一様分布と見なせる情報系列で各音符の揺らぎ成分を置換し，情報埋込みを実現している．すなわち，埋込みが施された音符の区間では，揺らぎは一様分布を持つと見なせる．これに対し，情報埋込みが行われなかった音符の区間（T_h 個以下のサンプルしか含まない区間）では，揺らぎは異なる分布を持つ可能性がある．つまり「一定以下のサンプルしか持たない音符の区間でのみ，揺らぎが一様分布を示さない」という事実の観測により，ステゴ解析が可能になる可能性がある．

この対策として，T_h 個以下のサンプルを含む区間に対しては，揺らぎが一様分布を示すようにデュレーションの変化を与えるという方法が考えられる．

以上のことから提案方式は，データ構造およびデュレーション値へのステゴ解析に対して十分な耐性を有する．また，各音符のデュレーション分布の観測による解析に関しては，適切な対策を施すことにより，耐性を付与できると考えられる．

5.5　音　　質

5.5.1　実験方法および実験データ

提案方式による埋込み処理が再生音質にどの程度影響を及ぼしたのか調べるために，次の音質評価実験を行った．

音質評価用 SMF として，埋込み実験に用いた SMF の中からクラシック，ジャズ，ロックの各ジャンル 2 曲ずつ，**表 8** に示す楽曲を選択した．なお，著作権切れ音楽はジャンルが混在しているため，評価対象からは除外した．これらは**表 9** のとおり楽器の種類や数，ノート数がそれぞれ異なるものである．なお，表 9 中の instrument name は楽器名，括弧内の数字は GM 規格でのプログラムナンバであり，各楽曲とも主要な楽器を 3 つまで示した．提案方式による各楽曲への埋込み結果は**表 10** に示すとおりである．

本研究では，20〜30 代の評定者 15 名による ABX 二重盲検法を用いた音源識別評価を行った．ABX 法では，まず，埋込みなしの演奏 A と，提案方式による埋込みを施した演奏 B をそれぞれ提示する．その後，A，B いずれかを演奏 X としてランダムに評定者に提示

表 8　音質評価用 SMF
Table 8　Experimental SMF.

Name	MIDI Source	Time [sec]	Size [byte]
C1	RM-C002	464	116,534
C2	RM-C030	243	25,513
J1	RM-J026	214	65,590
J2	RM-J045	265	52,367
P1	RM-P002	219	58,671
P2	RM-P015	159	64,798

表 9　実験用 SMF の特徴
Table 9　Characteristics of experimental SMF.

sample	instrument		note no.
	no.	name(p_no)	
C1	6	flute(74) oboe(69) Strings(49)	11379
C2	1	piano(0)	1722
J1	9	piano(0) guitar(32) alto sax(66)	4330
J2	10	piano(0) D.guitar(30) E.bass(35)	2308
P1	15	E.piano(5) piccolo(73) lead 1(81)	2479
J2	10	D.guitar(30) lead 8(88) synth voice(55)	5529

© 2009 Information Processing Society of Japan

表 10　提案方式の埋込みビット数
Table 10　Embedding capacity of the proposed method.

method	velocity [bit]	proposed [bit]
C1	11,032	42,030
C2	1,051	1,167
J1	12,763	19,172
J2	9,964	7,965
P1	10,847	9,112
P2	831	20,621

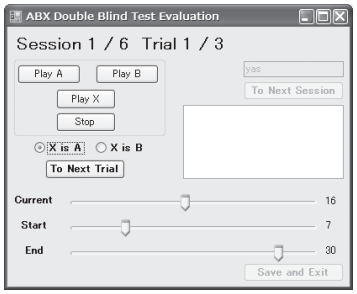

図 10　評価用 PC ソフトウェアの操作画面
Fig. 10　Screenshot of evaluation software.

し，評定者が X を A，B のどちらと同等に感じるかを評価させる．A と B が聴感上有意に識別できる場合には，X の正答率は 50%から偏った値を持ち，識別できない場合には，正答率は 50%になるはずである．しかし，識別できない場合でも，実験の試行回数が有限であるため正答率には偏りが生じる．本研究では，この偏りに有意性があるかどうかを χ^2 適合度判定により調べた．

本実験では，A，B から X を判定するサイクルを 1 トライアルとして，1 つの楽曲に対して X をランダムに変更しつつ 3 回のトライアルを行った．評価の際は，評価のつどランダムに X が選定されるように専用の評価ソフトウェア（**図 10**）を評定者自身に操作させた．評価用ソフトウェアの画面上では，X は A，B と区別できないように提示され，評定者はそれぞれの Play ボタンを押すことで演奏を開始し，Stop ボタンにより任意の位置で停止できる．また，1 度全区間を聴取した後は，画面下部のスライダにより，任意の区間の始点（Start），終点（End）を評定者自身の選択により聞き比べることができる．本実験の演奏には，YAMAHA 社製の MU-2000 を用い，Audio Technica 社製のモニタヘッドフォン ATH-PRO5V により聴取させた．

表 11　音質評価の χ^2 検定結果
Table 11　χ^2 test result of the proposed method.

sample	correct [%]	χ^2	p
C1	44.4	0.556	0.456
C2	46.7	0.200	0.655
J1	46.7	0.200	0.655
J2	46.7	0.200	0.655
P1	53.3	0.200	0.655
P2	42.2	1.089	0.297

5.5.2　評価結果と考察

それぞれのサンプルの先頭 30 秒間を聴取，評価させたときの検定結果を**表 11** に示す．なお，χ^2 検定については，各楽曲ごと全評定者の評価結果をまとめ，合計 45 回の試行結果として行った．この結果から，次の考察を得た．

すべての楽曲について，有意水準 10%においても有意差は認められなかった．この結果からは，本実験の評定者には提案方式による埋込みを施したステゴ SMF とカバー SMF との区別ができなかったということがいえる．すなわち，提案方式を用いることで演奏品質を劣化させずに情報を埋め込めたと見なせる．

また，各評定者ごとの全楽曲の評価結果を用いた検定においても，A と B の有意な識別ができた評定者はいなかった．本実験の評定者は，聴覚に難聴などの障害を持たない一般的な聴者であったが，より詳細な評価のためには，聴覚の優れた評定者や，音楽経験や音感の優れた評定者による評価も必要であると考える．これらについては今後の課題である．

6. おわりに

本論文では，発音時間の揺らぎを利用した SMF ステガノグラフィを提案した．SMF 中のデュレーションの揺らぎに注目し，この大きさを埋込み情報で表現することにより，演奏品質を劣化させることなく，埋込み率にして従来方式の 1.5 倍以上の埋込み能力を達成した．提案方式は従来方式[6] と異なる成分を用いるため，組み合わせて利用できる．そのため，埋込み可能情報量を線形に増大できる点でも有用である．

提案方式では，ヒストグラムを用いて音価ごとにデュレーションを区分し，それぞれの標準偏差により埋込みビット数を決定した．これは標準偏差に基づく適応化によって，秘密情報の存在を秘匿することがねらいである．今後は，埋込みによりデュレーションのばらつき

© 2009 Information Processing Society of Japan

がどう変化するか，また，埋込み後のデュレーションを時系列で配列した場合にどのような特徴が現れるか，という点から，ステゴ解析の手法とともに検討する予定である．

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) 井上大介，松本　勉：スタンダード MIDI ファイルステガノグラフィとその能力，情報処理学会論文誌，Vol.43, No.2, pp.2489–2501 (2002).
3) 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).
4) 遠山　毅，鈴木雅貴，四方順司，松本　勉：編集ソフトウェアの特徴を利用した攻撃への耐性を有する SMF ステガノグラフィ，情報処理学会論文誌，Vol.48, No.9, pp.3014–3026 (2007).
5) 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).
6) 山本紘太郎，岩切宗利：表情付けを考慮した SMF ステガノグラフィ，情報処理学会論文誌，Vol.47, No.8, pp.2724–2732 (2006).
7) 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).

(平成 20 年 9 月 19 日受付)
(平成 21 年 3 月 6 日採録)

山本紘太郎（学生会員）

1978 年生．2001 年防衛大学校理工学部情報工学科卒業．2007 年防衛大学校理工学研究科前期課程修了．同年 4 月より防衛大学校理工学研究科後期課程．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修了．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．

© 2009 Information Processing Society of Japan