3. 音声・音楽を用いたインフォメーションハイディング

岩切宗利　防衛大学校情報工学科
iwak@nda.ac.jp
松井甲子雄　防衛大学校情報工学科
matsui@nda.ac.jp

■ディジタル音声・音楽の特性

人間は，空気を伝達媒質とした疎密波の振動を耳により捉えて知覚する．人間が聞き取れる最小音圧は，50億分の1気圧というきわめて小さな変化であるため，不自然に混入した雑音に対する人間の分別能力は高く，敏感に知覚することができる．ただし，人間は，日常生活における余計な雑音による刺激を避け，必要な信号のみを処理できることが望ましい．そのため，脳は，外部からの恒常的な刺激に対して，何らかのフィルタ処理を無意識に施していると考えられる．

人間の内耳にある蝸牛基底膜の有毛神経細胞は，それぞれの固有振動数に応じて音波を周波数分析し，各成分ごとに神経パルスを発生する．すなわち，人間の脳は，有毛神経細胞から得られた神経パルスを用いて音声信号を周波数解析していることになる．ただし，人間の知覚できる周波数は，一般に20～20,000Hz程度であるため，Shannonの標本化定理によると，40kHzの速度により音声を標本化すれば，可聴周波数帯域のほぼすべてをカバーできることになる．実際にCDなどの音楽メディアでは，44.1kHzの標本化速度を用いて高い音質を実現している．一方，音声通話の帯域は4kHz程度であるため，電話帯域の標本化速度は8kHzもあれば十分とされている．このように音声の標本化周波数は，用途と所要帯域幅に応じて使い分けられる．

有毛神経細胞には，固有の波長のみに反応し，音波の絶対的な位相の違いを分析できない特性がある．よって，人間の脳は，音に含まれる周波数成分が同じであれば，同じ音として知覚することになる．また，強い刺激は，多くの神経パルスを発生し神経回路網を活性化するため，弱い刺激の伝達を掻き消す聴覚マスキング現象が起こる．

■情報ハイディングの分類と特徴

音声データは，一般的に膨大な量の時系列データになるため，高能率な不可逆圧縮を適用した後に情報メディアへ記録される場合が多い．すなわち，聴者に知覚されないように埋め込まれた秘匿情報は，符号量の削減による影響を受けやすいため，音声データを対象とした情報ハイディングは難しいとされている．しかし，音楽コンテンツを対象とした著作権保護技術の必要性は高く，商用目的の電子透かし技術もいくつか開発されているようである[1)]．高い攻撃耐性を求められる電子透かしには，聴覚特性を考慮しながら，利用者に知覚されにくい帯域へできるだけ強い透かし信号を埋め込むものが多い．ディジタル音声を対象とした情報ハイディング技術に関する詳細が公開されないのは，秘匿性能の低下を避けるためと考えられる．ディジタル音声を対象とした情報ハイディングの代表的な手法[1)～4)]は次のとおりである．

下位ビット置換法

下位ビット置換法は，埋め込みによる波形値の変化が少ない成分を情報ビットに置き換える最も基本的な手法である．

位相変調法

位相変調法は，短いフレームに区切った音声データの位相情報を情報ビットに応じて変化させる手法である．これは，絶対的な位相の変化を聴覚が識別できない特性に着目した手法である．

直接拡散法

狭帯域の透かし信号スペクトルを広帯域へランダムに分散配置する方法である．広い帯域へ拡散された信号は，偽輪郭効果を起こすため聴覚的な音質を損ないにくいと考えられる．

エコー合成法

エコー合成法は，情報ビットに応じた特定のオフセッ

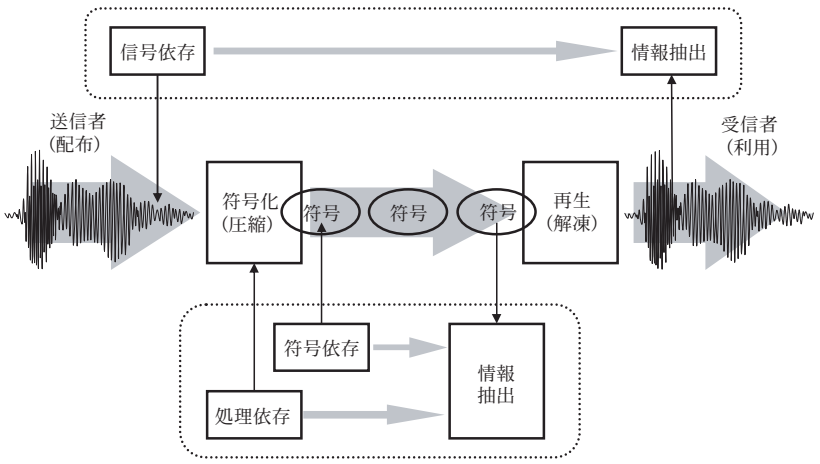


図 -1　情報ハイディングの基本形式

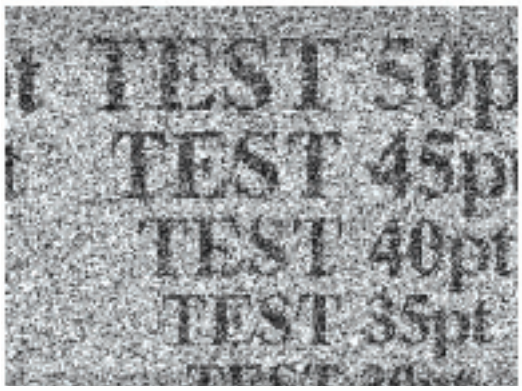


図 -2　MP3 音楽からの視覚的透かし検出

ト位置に，反響音を合成する手法である．情報を復号する際は，反響音の出現位置（オフセット位置）を調べればよい．

マスキング法

マスキング法は，埋め込み情報の存在を秘匿するために聴覚マスキング特性を積極的に用いる手法である．理論的には，知覚されない最大強度の埋め込みを実現できる [5]．

これらの情報ハイディングは，さまざまな視点から分類できるが，ここでは**図 -1** のように埋め込み処理の形式に応じて，信号依存型，符号依存型および処理依存型に区分し，それぞれの代表的な手法をいくつか紹介する．

■信号依存型ハイディング

信号依存型とは，A/D 変換などによって得られるディジタル信号（原信号）へ直接情報を埋め込む処理形式である．単純な PCM 符号として構成されている原信号には，情報ハイディングに利用できる知覚的な冗長成分が多く含まれている．しかし，知覚できない微弱な成分は，不可逆圧縮などのディジタル信号処理による影響を受けやすい．電子透かしなどの適用分野ではディジタル信号処理に対する耐性が重要になる．

◇下位ビット置換法

線形量子化法は，埋め込みによる波形値の変化が少ない成分を情報ビットに置き換える最も基本的な手法である．この方法によると大量の埋め込みを施しても，音質に与える影響を少なくできる．ただし，下位ビットに埋め込まれた情報は，信号処理によって劣化しやすいため電子透かしには利用できない．その対策として，音声データを 2 次元配列したとき，1 つの視覚パターンが構成されるように情報を埋め込み，透かしの攻撃耐性を高める方法が検討されている [6]．この方法によれば，**図 -2** のように視覚パターン認識の優れた特性を生かして，電子透かしの認識率を向上できる．

◇位相変調法

位相変調法とは，秘密鍵である乱数符号列と情報符号列を合成して得られた拡散符号列により，音声波形を直接拡散するものである．情報符号列を復号する際は，秘密鍵である乱数符号列により受信波形を逆拡散し，得られた波形の位相変化を観察すれば情報ビットを特定できる [2]．

◇直接拡散法

直接拡散法とは，情報を表現する狭帯域の信号スペクトルを広い帯域へランダムに分散配置することにより，高い秘匿性を実現できる手法である．その広帯域な信号構成法として，いくつかの手法が考えられる．代表的なものとして，直流成分や狭帯域信号成分を埋め込む技術が知られている．

直流成分を用いる手法 [7] は，一定強度の直流信号波形を直接拡散し，それに聴覚重み付けフィルタを施して音声波形に重畳するものである．この方法により埋め込みを施した波形を逆拡散すると，音声成分が期待値ゼロの交流成分となる一方，埋め込まれていた信号が直流成分として現れる．すなわち，拡散波形の直流成分を調べることにより，埋め込まれていた情報ビットを特定できる．この技術では，情報ビットを特定するための閾値の設定が難しくさまざまな検討がなされている [8]．

一方，狭帯域の周波数スペクトルを情報信号として

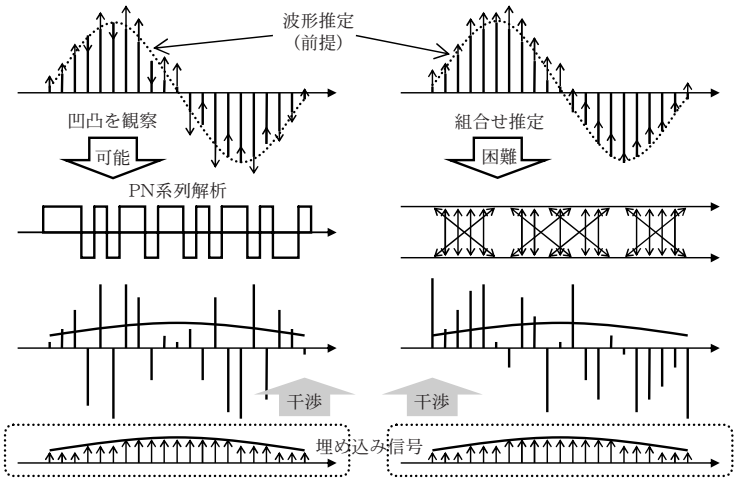


図 -3　直接拡散法とスクランブル法の比較

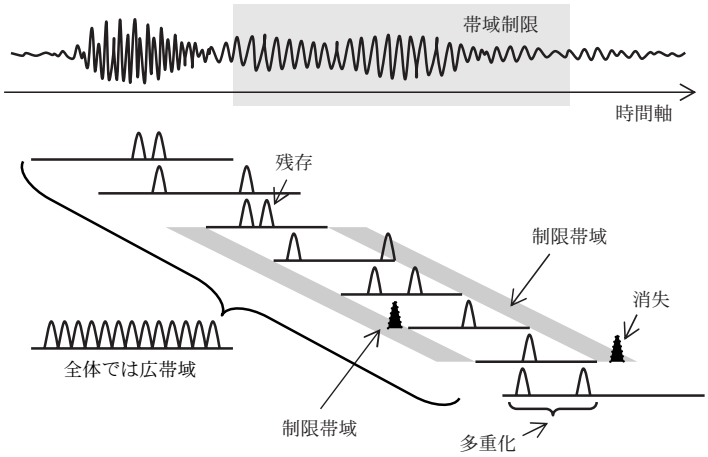


図 -4　周波数ホッピングの特性

利用する方法[2]では，音声波形を直接拡散して得られた周波数スペクトルの一部（狭帯域成分）に対して埋め込みを施す．この埋め込み処理による狭帯域な信号干渉の影響は，拡散された音声波形を逆拡散する過程において，広帯域へ分散配置されることになる．よって，偽輪郭効果により再生音質に与える聴覚的な影響を抑制できる．この方式では，帯域分割技術を応用することにより，MPEG オーディオ符号化などの攻撃にも耐える手法も検討されている[9]．

◇スクランブル法

図 -3（a）に示す直接拡散を原理とする埋め込み方式では，埋め込みのある波形に混入した微少な歪みを注意深く観察すると，直接拡散符号列を推定できる場合がある．音声の場合は，アナログ波形をディジタル化する際に，偽輪郭効果を得るためのディザが施されるため，単純に波形の形状を観察するだけでは拡散符号列を推定できない．しかし，原音声の波形を推定，もしくは原音声そのものが入手されると拡散符号列を容易に解析できる．すなわち，直接拡散による埋め込み技術では，原音声を公開することにより，重大なセキュリティの崩壊を招く危険性がある．この直接拡散法の問題点を補うために，時系列標本値の時間軸上における順序をランダムにする**図 -3（b）**のようなスクランブル方式が検討されている[10]．この方式による周波数領域における拡散効果は直接拡散と同様であり，広帯域へ拡散された埋め込み成分を狭帯域成分として集中させるには，拡散テーブル（鍵）を用いた逆スクランブル処理が不可欠となる．また，標本値の順番を入れ換える拡散テーブルの種類は無数に存在するため，鍵を知らない第三者による不正な情報解析は難しいと考えられる．

◇周波数ホッピング法

周波数ホッピング（FH）を用いた情報ハイディング技術[11]は，疑似乱数系列を用いて周波数ホッピング（FH）パターンを生成し，埋め込みを施す周波数成分を短期的にランダム制御するものである．この埋め込み位置（周波数成分）の時間軸上の変化頻度を多くすれば，狭帯域な信号を広い帯域へランダムに分散配置できる．このとき音楽データの占有帯域に対して透かしの埋め込み帯域幅が狭く，その埋め込み位置の分布がランダムかつ一様であることが望ましい．FH 方式の特性として，**図 -4** のように帯域通過フィルタによって影響を受けにくいことがあげられる．通過帯域幅は，制限帯域に比して広く設定されるため，FH 方式を原理とした埋め込み方法によれば，大部分の透かし信号が攻撃の影響を受けずにすむ利点がある．さらに，FH 方式は，FH パターンが異なる複数の埋め込みを多重に施すこともできる．

■符号依存型ハイディング

符号依存型は，聴覚モデルや情報理論的解析に基づいて生成された符号語に対して直接情報を埋め込む処理形式である．この処理形式の実現は，符号語のデータ構造の複雑さと冗長成分の少なさから一般的に難しい．また，符号化と復号の処理を反復することによって埋め込み情報を損失する可能性が高くなる問題もある．しかし，この処理形式は，すでに流通しているディジタルコンテンツに対する情報ハイディングを実現できるため，用途によっては有用である．

◇符号転置法

楽音符号とは，楽譜などの演奏音に関する情報を音の高さや発音タイミングなどの符号として記述したものである．このとき，同じタイミングに発音される音情報の記述順は，演奏音自体にまったく影響を与えないと考えられる．この特徴に着目した情報ハイディング技術として，SMF（Standard MIDI File）を対象とした松本らの情報ハイディング方式がある[12]．これは，同時実行されるイベントデータの記述順（順列）に規則性を持たせることにより，大量の情報を埋め込むことができるため，たいへん興味深い技術である．

◇情報表記利用法

SMF のデルタタイムは，最下位のバイトの最上位ビットを“0”とし，それ以外の上位バイトの最上位ビットをフラグ“1”とすることにより，数値を可変長表記している．そこで，デルタタイムのデータ長に着目して可変長コードに情報ビットを埋め込む方法が考えられる．すなわち，MIDI ファイルから選んだあるデルタタイムが，1 バイトであったとき，これを拡張して 2 バイト表現する．たとえば，78（=4$E_{(16)}$）[Tick] を拡張し，804E $_{(16)}$ とも表現できる．この手法は，最上位のデータバイト 80 $_{(16)}$ の有無によって，秘密情報のビット列を埋め込むものである．

また，SMF の音源コードには，その動作が音源機器に実装されていないものがある．たとえば，消音情報であるノートオフのベロシティ（音の強さ）は，一般には利用されていないのが実情である．このオフベロシティの下位ビットを秘密情報信号列に置き換えても，演奏音自体に大きく影響することはないと考えられる．

これらの情報ハイディング技術は，SMF のデータ構造の自由度を利用して秘密情報を埋め込むアイディアである．一方，一般的な楽音編集ソフトウェアは，データファイルから MIDI データを読み込む際に，そのデータ構造を破棄し，独自の処理しやすい形式としてメモリに展開する．たとえばデルタタイムに付加した情報符号 80 $_{(16)}$ は，読み込む際に無視される．すなわち，この手法を用いて埋め込まれた情報ビット列は，編集ソフトウェアを用いてメモリへ読み込むだけで消失することになる．視点を変えれば，ここに示した情報ハイディング技術を用いて，原本性を保証する符号を音楽コンテンツへ埋め込むことができることになる．すなわち，楽音編集ソフトウェアを用いて何らかの操作が加えられることにより，埋め込まれていた情報が消失する仕組みである．実際に，いくつかのソフトウェアを用いてファイルを読み込み，そのまま上書き保存しただけで情報が消失することが確かめられている[13]．

◇演奏ゆらぎ利用法

人間が楽器を演奏する際に，発音タイミングや音量は，常に一定ではないと考えられる．むしろ，人間の演奏時に発生するわずかな変動は，微妙な効果として，より芸術的な演奏を実現し，楽曲に表情を付ける要素になる．一方，MIDI のようにコンピュータを用いて音源を直接制御すると，常に一定かつ正確な演奏を再現できる．この特徴は，より精緻な演奏を実現できるといえるが，機械的な印象を聴者に与えやすい．実際に演奏家のくせをパラメータ化し，コンピュータによる演奏をより自然にする試みも多くなされている．この演奏音のゆらぎに着目した手法として，演奏音の強弱や発音タイミングに関する符号の下位ビットに情報を埋め込む方式が検討されている[13]が，演奏の表情付けやその評価が難しいため，今後の改善が望まれる．

■処理依存型ハイディング

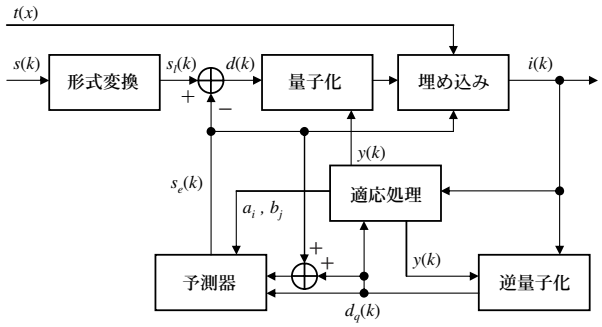
処理依存型は，ディジタル信号を符号化する過程において，密かに情報を埋め込む処理形式である．この形式は，音声情報を表現する符号語に対して，別の情報を付加する観点では符号依存型の一種であるとみなせる．しかし，符号化の処理構造に依存する処理依存型は，符号語のデータ構造に依存する符号依存型と本質的にまったく異なる形式である．一般的には，原信号との対応を考慮できない符号依存型に比べて，原信号へ最適に近似できる符号語を生成する処理依存型の方が高音質な符号を生成しやすい．

◇適応量子化法

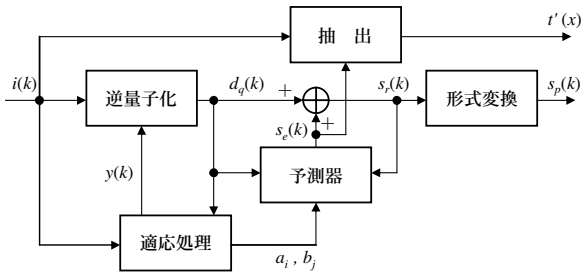
適応化のアルゴリズムは，短期的な入力信号から推定するフィードフォワード（前方適応化）方式と，出力信号により適応化するフィードバック（後方適応化）方式に分類できる．ここでは，フィードバック型の適応量子化方式を例にあげる．フィードバック適応量子化では，量子化刻み幅を量子化出力により適応化するため，伝送符号の一部を単純に情報ビット列へ変更すると，適応特性を大きく変化することになる．その対策として，ITU-T 勧告 G.726 32kbit/s ADPCM を用いた**図-5** の方式のように，埋め込みを施した符号をフィードバックすることにより，符号化と復号の適応処理同期を正しく維持できる[3]．

◇波形コードブック分割法

ITU-T 勧告 G.728 16kbit/s LD-CELP（Low Delay Code Excited Linear Prediction）は，ベクトル量子化を原理とする代表的な符号化手法の 1 つである．LD-CELP



(a) 符号化器



(b) 復号器

図 -5　G.726 を用いた情報ハイディング

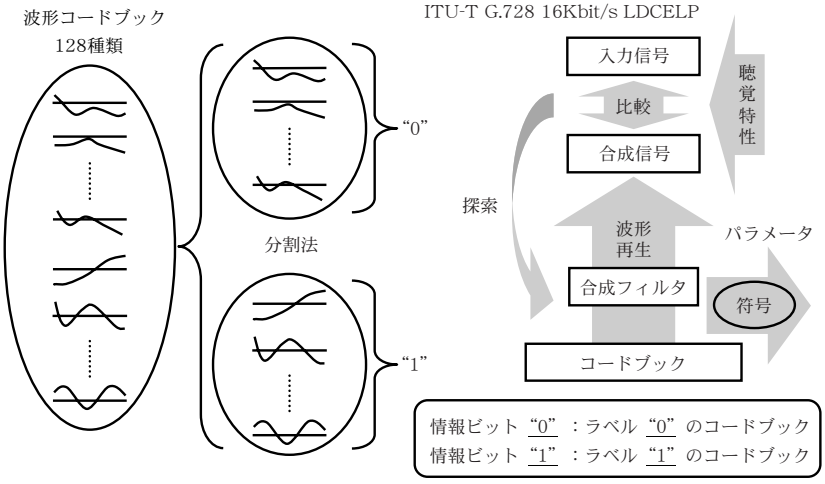


図 -6　ベクトル量子化方式

は，波形の形状を符号化する波形符号化方式のように，入力波形に近似した再生波形を復元しない特徴がある．ただし，その再生音質は，その倍のビットレートである 32kbit/s ADPCM に匹敵するとも言われている．一般的なベクトル量子化方式のコードブックの探索は，合成分析（Analysis by Synthesis：AbS）の手法を用いて，聴覚重み付き波形に最も近似するように行われる．このようなベクトル量子化を用いた情報ハイディングの基本的なアイディアは，**図 -6** のように符号化する際に用いる励振波形コードブックを 2 つの集合に分割し，埋め込む信号に応じて，その選定範囲を切り替えることにより情報ビットを表現するものである [2]．

◇マルチパルス音源探索法

ITU-T 勧告 G.729 8kbit/s 共役構造代数 CELP（Conjugate Structure Algebraic CELP：CS-ACELP）は，ベクトル量子化を原理としたハイブリッド符号化法の 1 つである．G.729 には，波形コードブックとして 4 つのパルス信号 $i_0 \sim i_3$ を用いる特徴がある．しかし，マルチパルス音源の探索処理量は膨大になるため，G.729 では入力信号に応じて適切な閾値を算出し，4 番目のパルスの探索に関しては，その値を超える候補のみについて実施している．さらに，各フレームごとの最大探索処理時間を導入することにより処理遅延を抑制している．これらの特徴は，選択されたパルス信号以外にも適切な候補が存在する可能性を示唆している．ここで**表 -1** に示した CS-ACELP 符号化方式における 4 番目のパルス信号 i_3 のパルス位置 m_3 に着目すると，ほかのパルス信号 $i_0 \sim i_2$ のパルス位置 $m_0 \sim m_2$ と異なり，隣接した候補を持つことが分かる．そこで，パルス位置の m_3 の選択性を制御すれば，音声符号のマルチパルス音源情報を利用して情報を埋め込むことができる [2]．

◇ピッチ予測法

ITU-T 勧告 G.723.1 は，2 つの伝送レートを持つ特徴がある．G.723.1 の伝送レートごとに異なる処理は，いずれもマルチパルス符号化方式を原理とするベクトル量子化である．よって，G.728 や G.729 を対象とした埋め込み法を応用することもできる．しかし，2 つの伝送レートは，240 サンプルのフレームごとに切り換えることができるため，いずれの伝送レートでも共通な部分へ施すことが望ましい．そこで G.723.1 の伝送符号を調べると，各伝送レートに共通なものは，線形予測符号（LPC）とピッチ予測符号および利得情報であった．LPC への埋め込み制御は，符号化フィルタの特性に大きく影響するため望ましくない．また，ピッチ予測の利得は候補数が少ないため，再生音声に大きな歪みを生じると予想される．ただし，ピッチ予測符号は，推定ピッチラグの近傍のみから探索される．これは探索範囲を限定することにより，音声の周

Pulse	Sign	Positions
i_0	s_0：±1	m_0： 0,5,10,15,20,25,30,35
i_1	s_1：±1	m_1： 1,6,11,16,21,26,31,36
i_2	s_2：±1	m_2： 2,7,12,17,22,27,32,37
i_3	s_3：±1	m_3： 3,8,13,18,23,28,33,38 ： 4,9,14,19,24,29,34,39

表-1　マルチパルスコードブック

期性を保持しつつ，低ビットのまま高速にベクトル量子化し音質を改善する工夫である．見方を変えれば，ある程度の周期性が保たれていれば，音質を大きく損なわないことが分かる．そこで，このピッチ予測器の処理を工夫して閉ループピッチラグの偶奇性を制御することにより，ピッチ予測符号それぞれに1ビットの意味を持たせる方式が検討されている[14]．

■性能評価

ディジタル音声を対象とした情報ハイディング技術は，これまでに数多く提案されているが，それらを定量的に評価する方式や基準は明確に示されていない[1)〜4)]．一般的な評価項目は，次のようになると考えられる．

a. 品質への影響（秘匿性）
b. 解読や不正な編集の困難さ（保全性）
c. 配信形式への適合（親和性）
d. 埋め込み可能容量（情報容量）
e. システム負荷（処理量）
f. 信号処理の影響（攻撃耐性）

これら評価項目の重視度は，情報ハイディングの適用分野によって大きく異なる．たとえば，高品質な音声信号から著作権情報を抽出する電子透かしでは，秘匿性や攻撃耐性などの評価項目が重視される．一方，ディジタル電話通信系によるステガノグラフィの場合，情報伝達に十分な情報容量やその保全性，通信系との親和性などが重要になる．また，簡単な編集操作により消失することが望ましい原本保証の技術もある．したがって，情報ハイディングを実用化する際は，適用分野などに応じた評価項目や達成基準についてよく検討しなければならない．

■今後の課題

これまでに実用化された情報ハイディング技術の性能評価に関する詳細情報は，一般に公開されていない．これは，情報ハイディングの性能情報が不正な第三者によるアルゴリズムの解析や弱点への攻撃を可能にする手掛りになるためである．また，優れた評価用ソフトウェアは，攻撃ツールの1つとして悪用される問題もある．

これらの点から，暗号技術のような情報理論的（計算量的）性能評価が情報ハイディング技術に関しても望まれる．しかし，埋め込み対象の信号特性に適応する情報ハイディング技術の理論的な評価は一般的に難しい．理論展開できる簡易モデルの評価結果を実用システムの絶対的性能と見なすことは，思わぬセキュリティ上の問題発生にもつながるため，実際はさまざまな視点からの評価項目に対する結果を総合的に判断しなければならない．この判断を容易にする情報ハイディングの定量的な評価技術やそれに基づく評価ツールの整備については，今後の大きな課題の1つであるといえる．

参考文献

1）電子透かし技術に関する調査報告書, 日本電子工業振興協会（1999）.
2）松井甲子雄：電子透かしの基礎－マルチメディアのニュープロテクト技術，森北出版（1998）.
3）Katzenbeisser, S., Petitcolas, F. A. P.: Information Hiding Techniques for Steganography and Digital Watermarking, Artech House（2000）.
4）Bender, W., Gruhl, D. and Morimoto, N.：Techniques for Data Hiding, Proc. of SPIE 2420, Storage and Retrieval for Image and Video Databaselll, pp.164-173（1995）.
5）Boney, L., Tewfik, A.H. and Hamdy, K.N.: Digital Watermarks for Audio Signals, Proc. of the International Conference on Multimedia Computing and Systems, pp. 473-480（1996）.
6）岩切宗利，松井甲子雄：デジタル音楽への電子透かしの可視化法，情報処理学会論文誌，Vol.41, No.6, pp.1840-1847（June 2000）.
7）Bassia, P. and Pitas, I.: Robust Audio Watermarking in the Time Domain, Proc. of EUSIPCO´98, pp.25-28（1998）.
8）上野貴之，吉田真紀，藤原　融：オーディオ信号用のある電子透かしに対する検出誤り推定法における精度向上に関する考察，信学技報 ISEC2001-134, pp.127-132（2002）.
9）IWAKIRI, M., and MATSUI, K.：The Watermarking of Digital Sound by Band Division and Spectral Spreading, The Journal of the Institute of Image Information and Television Engineers, Vol.56, No.7, pp.1105-1110（2002）.
10）岩切宗利，松井甲子雄：デジタル音楽への電子透かしの秘匿法に関する一提案, 電子情報通信学会2001年総合大会 基礎・境界講演論文集, A-7-14, p.207（2001）.
11）岩切宗利，松井甲子雄：周波数ホッピングと変形離散コサイン変換によるデジタル音楽への電子透かし法，情報処理学会論文誌，Vol.43, No.3, pp.734-742（Mar. 2002）.
12）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，SCIS2000-C03（2000）.
13）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（Feb. 2002）.
14）岩切宗利，松井甲子雄：ITU-T 勧告 G.723.1 による音声符号化方式を用いたステガノグラフィ，2002 年暗号と情報セキュリティシンポジウム，SCIS2002-5D-1, pp.289-294（2002）.

（平成 15 年 2 月 10 日受付）