Audio data hiding based on amplitude modulation and its application

西村, 明
Faculty of Informatics, Tokyo University of Information Sciences

https://doi.org/10.15017/18879

第2章　音響信号への情報秘匿技術

2.1　まえがき

本章では，音響信号への情報秘匿 (audio information hiding, audio data hiding) 技術に関して，用いられる用語や概念などをはじめに説明する．そして，技術への要求，評価方法を示す．また，過去の研究において提案されてきた代表的な情報秘匿手法を，対象とする音響信号の状態，埋め込みおよび検出方法によって分類し解説を加える．さらに，性能向上のため補助的に用いる技術も説明する．最後に，電子透かしとしての技術利用の現状について示す．

2.2　音響信号への情報秘匿技術の概要

図 2.1 に，音響信号への情報秘匿技術の概要を図示した．この技術は，埋め込む情報やその利用法によって，電子透かしとステガノグラフィにおおまかに分類される．以下この図中のキーワードを基に，技術の概要を説明する．

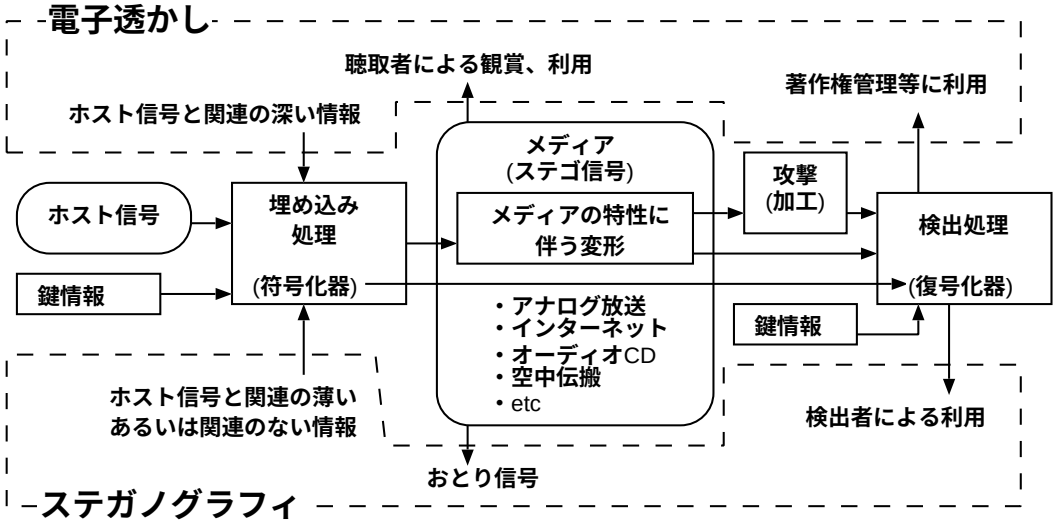

図 2.1: 音響信号への情報秘匿技術の概要．

情報埋め込みの対象となる音響信号をホスト信号 (host signal) あるいはカバーデータ (cover data) とよぶ．埋め込む時には，埋め込み情報の秘匿性を高くするため，“鍵 (key)” と呼ばれるデータを併用する．この鍵は，データの暗号化で用いられる鍵とは基本的に異なり，情報秘匿アルゴリズムにおける情報埋め込みの位置や順序など，どこに隠したかに対応する情報である．

この秘匿性が重要でない利用方法もあるため，全ての情報秘匿手法が鍵を備えている訳ではないが，この鍵情報が無いと秘匿情報の検出や埋め込まれていること自体の検知が困難になるべく，埋め込みおよび検出アルゴリズムは設計されるのが一般的である．

情報が埋め込まれた音響信号は，ステゴ信号 (stego signal) とよばれる．ステゴ信号は，人間が聴取する対象であり，メディアを経由したのち，秘匿情報が検出される過程に達する．よってホスト信号の観賞に利用価値のある場合は，ホスト信号に対するステゴ信号の音質劣化が必要充分に少ないことが，重要な要件となる．

ステゴ信号は，インターネット配信やディジタル放送時における MP3 や MPEG2AAC などの知覚符号化，アナログ放送時の雑音の重畳，AD/DA 変換，帯域制限などといった伝送メディアの特性や特徴が反映される変形が加わった後も，秘匿情報の検出可能なことが要求される．また，著作権管理情報を秘匿した音楽データを違法配信する悪意あるユーザが，秘匿情報を検出できなくすることを意図して行う悪意のある攻撃 (malicious attack) に対する耐性も重要である．コンテンツの品質を大幅に劣化させない攻撃としては，数%のピッチ変換，時間長変換，サンプリング周波数変換，ランダム刈り取り (random cropping: 標本化された振幅値をランダムに抜き取ったり同じ値を挿入したりする) などが挙げられる．よってこれらの変形に対する耐性も，特に電子透かし用途においては，重要な要件と考えられる場合がある．

画像透かしの分野では，攻撃に対する耐性評価に Stirmark とよばれる加工ツール [16] が用いられることが多い．音響分野では，その音響版ともいえる Stirmark for Audio を複数直列に組み合わせて耐性評価を行うシステムも提案されているが [17]，使用例はまだあまり無い．

ステゴ信号に変形が加わると，検出される情報には誤りが混入しやすくなる．秘匿手法にもよるが，最悪の場合は，情報が埋め込まれていないコンテンツから誤って情報を検出する場合もありうる．よって，ステゴ信号への変形を前提とする利用場面において技術の実用化を行う場合，巡回符号やブロック符号などのエラー訂正符号を用いて秘匿情報を符号化し，冗長なデータを埋め込むのが一般的である．こうすることで，エラー訂正限界ま

での検出ビット誤りが検出時に存在しても，秘匿情報を完全に復号化でき，検出エラーの多いコンテンツや埋め込んでいないコンテンツから誤った情報が検出されることを防ぐこともできる．

ステガノグラフィは，秘匿通信や電子あぶりだしとも呼ばれる．埋め込まれる対象のコンテンツと埋め込む情報との間に関連性が全く無いか，比較的薄いことが特徴で，コンテンツ自体よりは埋め込む秘匿情報の方に利用価値がある．よって秘匿情報が埋め込まれたコンテンツは，おとり信号ともよばれ，埋め込む情報量も多くとれる必要がある．暗号を使った通信を行うと，暗号は解読されなくとも通信を行っていることは発覚するが，ステガノグラフィを用いれば，第三者はそこに情報が埋め込まれていることすら分からないため，通信の秘匿性が高まるという考えである．

いずれの情報秘匿用途においても，秘匿情報の検出時にホスト信号を必要とするものはノンブラインド手法，必要としないものはブラインド手法と呼ばれる．前者は，ホスト信号と比較して時間領域あるいは周波数領域において何らかの差分をとって検出のための分析を行うため，僅かな違いを検出しやすい．よって，埋め込みに伴うホスト信号の変形，つまり音質劣化を少なくでき，かつより多い埋め込み情報量を実現できる．しかし，埋め込みにホスト信号が必要なことは，ホスト信号を入手できない，例えばステゴ信号を利用する側の検出器や独立した検出器では秘匿情報の検出が不可能なことから，明らかに技術の利用可能な場面が限定される．そこで，本章および本論文ではブラインド手法のみを取り上げる．

2.3 技術の評価方法

前述したように，用途を限定しない場合の情報秘匿技術の主な評価要因は，秘匿データ量，変形や攻撃への耐性 (頑強性)，品質 (音質) である．同一手法においては，一般にこれら三つの要因はお互いに相反するものであり，一方を向上させると他方が低下する．しかし，技術の使用場面によっては，いずれかの要因を全く無視してよい場合もある．例えば，狭帯域音声コーデックの音質向上のために，帯域拡張情報を埋め込む利用 [18] では，通信路におけるエラー混入以外，基本的に耐性を考慮に入れる必要はない．よって，いずれかひとつ，あるいはふたつの要因での要求水準を定めた上で，残りの要因で評価する，という方針が妥当であろう．

一方，既存技術の多くが，音響信号の物理的特徴に依存した性能を示す．つまり音楽で

言えば楽曲のジャンルや曲そのものが異なるだけでも，上述の性能評価結果が比較的大きく変わってくることが多い．従来の音響電子透かしの研究では，複数のジャンルから多くても 10 曲程度を選んで性能評価を行っている研究がある一方，数曲程度しか取り上げない研究も多い．また，共通した評価対象楽曲があるわけではないため，敢えて有利な結果をもたらす楽曲が選ばれている可能性も否めない．

加えて，利用場面によっては，情報が秘匿されていることの第三者による検知が困難である秘匿性，埋め込みおよび検出処理の高速性や実時間処理も重要になってくる場合がある．秘匿性を評価する方法はいくつか考えられるが，最も単純な評価は，ステゴ信号に対する既知の検出方法を用いて埋め込み鍵空間への総当たりによって秘匿情報検出確率を調べることであろう．ただし，時間あるいは周波数領域などにおける埋め込み位置が鍵によって秘匿されている場合は，埋め込みと検出アルゴリズムが明らかであっても，実質上の鍵の長さからいって秘匿情報を検出することは不可能な場合が多い．

よって，秘匿情報が検出できなくても，ステゴ信号を分析することによって情報が秘匿されているかどうかを知ることができるかどうかも，秘匿性を評価することになる．このような秘匿の有無を調べることは，ステガナリシス (steganalysis) と呼ばれる．ステガナリシスの研究は，秘匿を発見すること，および秘匿性能の向上のために，画像の電子透かしにおいて盛んであるが，音響メディアに関してはあまり行われていない．

比較的厳しい秘匿性の評価は，ホスト信号とそれに対して情報を埋め込んだステゴ信号の両方が存在し，それらの何らかの差分をとることが許される場合に，どのような方法で情報が秘匿されているのかを推定し無効化する差分攻撃に対する耐性評価であろう．これはあらかじめホスト信号に知覚困難なダミーの非線型信号処理を施した上で，埋め込み処理を行うことにより，情報秘匿に伴う差異を検出しにくくなることが考えられるが，ステガナリシスと同様に，音響メディアに対するこの分野の研究はあまり進んでいるとはいえない．

情報埋め込みに伴う音質劣化を，主観評価実験によって評価することは，最も困難でありかつ労力を要し，さらにその実施には細心の注意が必要である．電子透かし用途としては，極めて僅かな音質劣化あるいは全く音質劣化が検知されないことが望まれる．極めて僅かな音質劣化を評価する実験方法については，知覚符号化音響信号やマルチチャンネルオーディオの音質劣化を評価する必要性から，ITU-R BS.1116-1 が制定され，推奨されている．BS.1116-1 に従う実験方法の結果は，SDG(Subjective Difference Grade) という値で示される．

BS.1116-1 が推奨する項目は多岐に渡り，実験計画，被験者の選定，被験者の人数，使用機器，実験環境，実験方法，実験に使用する素材，データ分析等である．このような項目が取り上げられているのは，いずれの項目も，極めて僅かな音質劣化を正当に評価する実験のために重要であるからである．例えば，一定の音質の劣化を検知する能力をもつ被験者を選定しなかった場合，存在するはずの音質劣化に気づかず，音質劣化を過小評価してしまうことになる．また，実験方法として，基準音 (ホスト信号) と劣化音 (ステゴ信号) の順に呈示し，後者の劣化度合を数値に割り当てて答える，という方法では，心理実験においてしばしば見られる順序効果によって後の音の音質を高くあるいは低く評価する可能性がある．さらに劣化音がどちらであるかがあらかじめ被験者に分かってしまう実験デザインでは，被験者が本当に音質劣化検知能力を持っているかどうかを検証することは不可能である．

過去の音響情報秘匿技術に関しては，特に音質が重視される電子透かし用途の研究においてさえ，一定の音質の劣化を検知する能力をもつ被験者を選定して音質劣化評価実験を行った研究はほぼ皆無であるといってよい．また，実験方法にも上述のような不備のあるものが目立つ．これは，主観評価実験を正当に行う困難さを示している．

一方，困難な主観評価実験の代わりに，客観評価実験によって音質劣化評価を行う場合がある．最も単純な客観評価指標は，ステゴ信号をノイズの混入した信号とみなして計算される信号対雑音比 (SNR, Signal to Noise Ratio) である．これを信号波形全体で計算したり，短い時間毎に SNR を計算して平均する，いわゆる Segmental SNR などが挙げられ，SNR が小さいほど，音質が劣化しているとみなす．しかし，知覚符号化と復号化を行った波形信号に対して，この SNR を計算した場合，主観的な音質劣化と SNR の値とは対応しないことが知られている [19]．つまり，知覚されにくい領域に情報を埋め込む音響情報秘匿技術においても同様なことが生じて，SNR が主観評価の代わりにならないことは明白であろう．

知覚符号化音響信号を客観的に評価する手法として，ITU-R BS.1387 において提案されたのが，PEAQ (Perceptual Evaluation of Audio Quality) である．これは原音と符号化音 (劣化あり音) をそれぞれ，聴覚フィルタを模したフィルタ群で帯域分割した上で，絶対閾値，周波数マスキングや時間マスキングを考慮した興奮パターン上での相違の度合から複数の指標 (Model Output Value; MOV) を計算し，MOV に主観劣化評価結果とよく合うような重み付けを行って劣化度合を予測する手法である．得られる客観的劣化度合 (Objective Difference Grade; ODG) は，ITU-R BS.1116-1 の測定で得られる SDG に

対応する値である．音響情報秘匿に伴う音質劣化に適用することにも，一定の根拠はあると考えられるが，MOV の重み付けを行う際に用いられるニューラルネットのための学習値として，知覚符号化を経た信号に対して得られた SDG が用いられていることから，PEAQ の出力する ODG が知覚符号化以外の主観音質劣化に対応するかどうかについては，明らかでない．

2.4 既存の音響信号への情報秘匿手法

2.4.1 ディジタル領域での情報秘匿手法

ステゴ信号が，ディジタル情報のまま変形を受けることなく，秘匿情報検出処理の対象となることを前提とした秘匿手法である．一般に，ホスト信号の情報量に対して，埋め込むことのできる情報量は最大で 1/10~ 1/5 程度にもおよび，秘匿情報量は多いが，ステゴ信号の変形には脆弱である．

ビット置換法

音響信号を何らかの方法で量子化あるいは符号化したデータにおいて，最も重要でないビット (LSB, Least Significant Bit) 値を，埋め込むデータのビット値で置き換えるものである．最も単純な手法は，音波形を標本化し直線量子化した後の，8 ビットや 16 ビットの振幅値のうち，下位ビット値 (複数ビットでも可) を埋め込むデータのビット値と置換する LSB 法である．また，適応差分 PCM 符号化において，伝送符号の最下位ビット値を置換する方法 [20] もある．

16 ビット直線量子化された音ファイルに対して，最下位ビットのみのビット置換による音質劣化は非常に少ないことが知られているが，複数ビットを置換して埋め込み量を増大させる場合や，8 ビット直線量子化ファイル，符号化音響データに対して，LSB 法を用いて音質を保つためには，データの中から知覚的に重要でないビット位置を見つける必要がある [18]．例えば，音波形に対して整数 MDCT(Modified Discrete Cosine Transform) を用いて周波数領域の整数データに変換し，心理音響マスキングモデルを用いて閾値以下と判定されるビット値を置換し，逆変換によって波形に戻す手法 [21] などが提案されている．

ビット置換法は一般に，ステゴ信号の変形には脆弱であり，そのためフラジャイル (fragile, 脆弱な) 透かしとして，ステゴ信号への加工や変形，攻撃を検知するために用いられるこ

ともある [9]．

符号化併用法

ディジタル通信メディアでは，アナログ波形を標本化および量子化したディジタル音波形データをそのまま伝送するのではなく，非可逆的な情報圧縮符号化 (CELP, MP3 など) したデータを伝送や保存することが多い．情報秘匿済みのディジタル音波形データに対して，このような符号化を行い復号化すると，データを埋め込んだ冗長な部分が欠落するため，検出が困難になる場合がある．そこで，あらかじめ符号化と同時に埋め込む，あるいは符号化後のデータに情報を埋め込む処理を行い，復号化時あるいは符号化データそのものから秘匿情報を検出する，という手法である．

前節で紹介したビット置換法も符号化時に併用することはできるが，本手法が特徴的なのは，符号化アルゴリズムあるいは符号化後のデータの特性を利用して埋め込みを行う点である．例えば，LD-CELP(Low-Delay Code Excited Linear Prediction) 符号化を用いる場合は，128 種の波形コードブックを鍵に基づいて二群にラベリングし，どちらのコードブック群から最もターゲットベクトルに近似したものを選ぶか，に埋め込むビット値を割り当てる方法 [22] が提案されている．他にも，MP3 符号化データのスケールファクタの値に埋め込む手法が考案されている [23]．

符号化後のデータ自体への変形には耐性を持つもの [23] もあるが，復号化して得られる音波形データや，それ以降に何らかの変形が加わった波形，再符号化後のデータなどから，秘匿情報を検出することは困難なものが多い．

2.4.2 アナログ耐性のある情報秘匿手法

秘匿情報の埋め込みと検出処理はディジタル領域で行われるのが一般的であるが，ステゴ信号がアナログ領域 (DA/AD 変換) を経由しても，秘匿情報の検出が可能な手法である．つまり，DA/AD 変換に伴う 0.1%程度以下のサンプリング周波数変換や，量子化ノイズの付加，0.1%程度以下の高調波歪などに耐性があり，さらには高ビットレート (64 kbps/ch 以上) の知覚符号化と復号化にも耐性を備える．この程度の変形に対しては，数 10 bps ~ 数 100 bps 程度の埋め込み情報量を実現する手法が多い．ステゴ信号の空間伝搬 (スピーカ再生とマイク受音) や，さらなる変形，悪意ある攻撃に対する耐性を備える場合には，数 bps ~ 数 10 bps 程度の埋め込み情報量となるのが一般的である．

スペクトル拡散法

スペクトル拡散 (spread spectrum) 法は，–1 および 1 の振幅値を持つ M 系列信号や疑似乱数 (PN: pseudo-random noise) 系列信号を用いて，ホスト信号の広いスペクトル帯域に渡って秘匿情報を埋め込む手法である．秘匿情報のビット値を系列信号で変調してホスト信号と加算する直接拡散 (direct spread spectrum) 法と，検出時にホスト信号と系列信号との相互相関を求める相互相関 (cross correlation) 法に大別される．

直接拡散法による秘匿情報埋め込みの例を図 2.2 に示した．図 2.2(a) はデータ信号であり，ここでは 8 サンプル毎に 1 あるいは– 1 の振幅によりビット情報を表現している．サンプリング周波数に対するデータ信号の区間長 (この例では 8 サンプル) を，データレートと呼ぶ．図 2.2(b) は有限長の PN 系列信号の一部であり，これをデータ信号に乗算したものが拡散信号 (c) となる．(d) はデータ信号と拡散信号のスペクトルを比較しており，データ信号のスペクトルは，白色スペクトルを持つ PN 系列信号を乗算することにより，白色化 (スペクトル拡散) されていることが分かる．

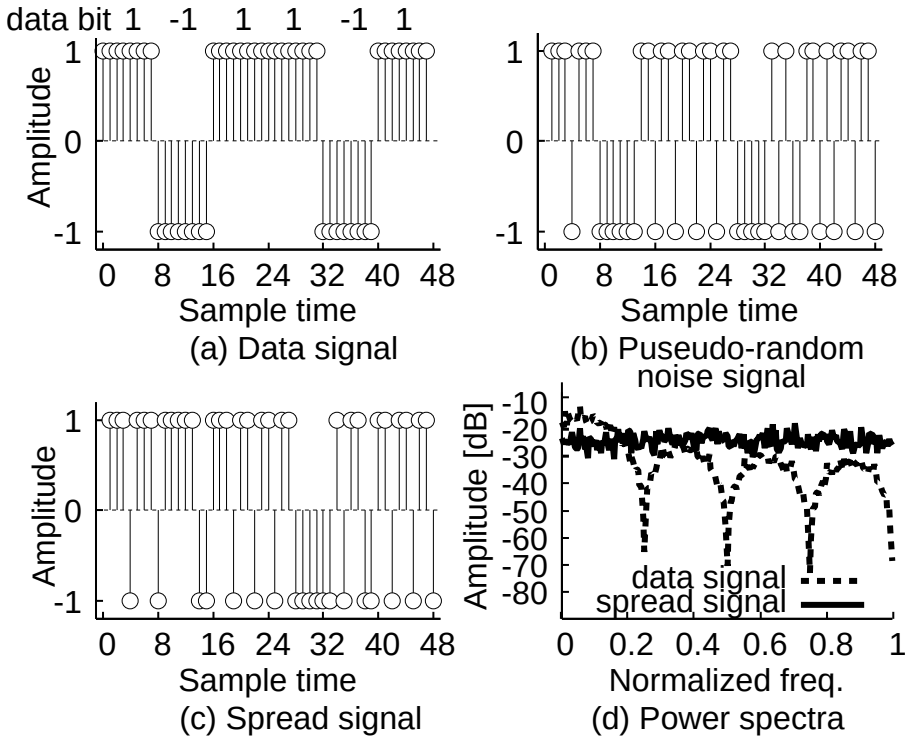

図 2.2: 直接拡散法での信号表現

この拡散信号の振幅やスペクトルを調整 (2.5 節参照) した上で，ホスト信号と加算 (図 2.3(a)) することによって，秘匿情報を埋め込む．音響透かし用途の場合，拡散信号は聞き取られない強度に設定する必要があるため，データレート長を 0.05~ 0.2 秒程度と長く

とって強度を下げ，検出力を保つ必要がある．

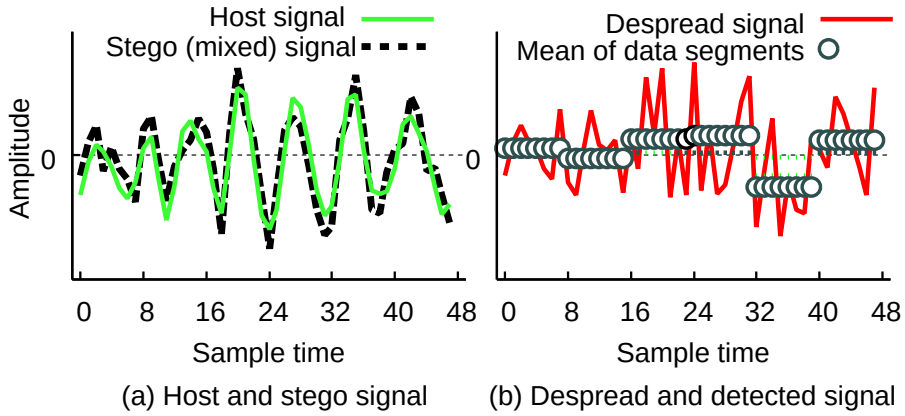

図 2.3: 直接拡散法でのステゴ信号と検出過程

検出時には，まずステゴ信号より，加算された系列信号区間を同定する．この同定には，次に説明する相互相関法が用いられることが多い．この区間に埋め込み時に用いた系列信号を乗算することにより，ホスト信号成分は拡散される一方，系列信号は埋め込み時から二度乗算されることになり，逆拡散 (despread) されてその負符号が取れるため，データ信号が現れる．よって，データレート幅毎に振幅の平均を取れば，その正負が検出されたビット値を表す (図 2.3(b))．

相互相関法による，データ埋め込み区間の同定 (同期ともいう) は，埋め込み時に系列信号の振幅を調整した上で，繰り返し時系列上に並べてホスト信号と加算する．埋め込み時に用いた系列信号とステゴ信号との相互相関を求めると，相互相関波形には系列長周期でピークが現れるので，高精度な埋め込み区間の検出が可能となる．系列信号の振幅を小さくする場合は，系列長を長くするか，相互相関波形を系列周期ごとに同期加算して検出力を高める．

埋め込む情報の表現を，上述の同期用系列信号をそのまま加算すると 1，同期用系列信号に -1 を掛けて加算するとき 0 と定める場合は，相互相関法は直接拡散法と同じである．データ埋め込み用に，同期用とは別の系列信号を用いる場合は，その系列信号を循環的にシフトさせ，ビット情報をシフト長で表現する [24]．検出時に埋め込んだ系列信号とホスト信号の相互相関を求めると，系列周期毎に埋め込み時にシフトした分だけ，ピークの位置がずれるので，ずれ幅を検出してビット情報を得られる．

スペクトル拡散法の特徴としては，広い周波数領域に秘匿情報が拡散されるため，フィルタリングなどによる，周波数領域でのステゴ信号の欠落や変形に対して，耐性を高くできる．また，秘匿情報が含まれる拡散信号を検出すべきシグナル，ホスト信号をノイズと

みなすと，SNR(Signal to Noise Ratio) が非常に低い (ノイズが強い) 条件でのシグナル検出が可能な手法であるため，ステゴ信号への付加雑音や知覚符号化に対する耐性も高い．加えて，埋め込み時に用いた系列信号は，検出時にも鍵として必要となり，秘匿性も高い．なお，ここでは時間領域でスペクトル拡散を行う手法のみ紹介したが，スペクトル領域やケプストラム領域で同様な処理を行う手法も提案されている．

しかし，ピッチ変換や時間長変換などの攻撃に対しては，逆拡散時に同期が取れなくなり検出が困難となる．よって，なんらかの補助的な方策を考える必要がある．

エコー法

エコー (echo hiding) 法は，ホスト信号に 10ms 程度以下の遅延を加えて，その遅延時間に秘匿情報のビット値を割り当てる方法である [25]．図 2.4 は，ホスト信号に与えるインパルス応答を示しており，実線の $d0$ の遅延を加えた場合にビット値 0，点線の $d1$ の遅延を加えた場合にビット値 1 を割り当て，埋め込み情報に応じてホスト信号に加算するそれらの遅延を一定時間区間毎に切替える．複数の遅延時間や正負の遅延を用いれば，より多くの情報を埋め込むこともできる．この手法の特徴としては，埋め込み処理の処理負荷が低く，埋め込みに用いる短い時間の遅延は，音質劣化に繋がりにくい，という点である．

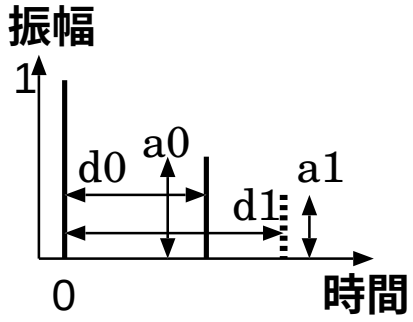

図 2.4: エコー法で用いる埋め込みビット値に対応するインパルス応答．例えば，実線のエコーを与える場合はビット値 0，点線のエコーの場合はビット値 1 を割り当てる．

検出時にはまず最初に，何らかの手法によって，埋め込んだ時間区間を同定する必要がある．得られた時間区間波形に対して，DFT (Discrete Fourier Transform) を用いて周波数領域に変換し，得られた周波数成分の対数値を逆 DFT を用いて時間領域に変換して複素ケプストラムを求めると，その絶対値には，遅延時間の整数倍にピークが表れる．あるいは，その複素ケプストラムの絶対値の自己相関を求めても，遅延時間の整数倍にピーク

が表れる．よって，ビット値が割り当てられている複数の遅延時間のうち，どこで最大のピークとなるかによって，埋め込まれたビット値を決定することができる．

このような遅延音を加える方法では，検出方法から分かるように秘匿性は非常に低い．また，耐性を保つには遅延音の振幅を強めるか，遅延時間を一定とする時間区間長を伸ばす必要があるが，前者は音質劣化に，後者は埋め込みデータ量の減少に繋がる．

エコー法に対して秘匿性を高く，かつ知覚されにくく耐性を高める手法として，エコー拡散 (spread echo hiding) 法が提案されている．これは，単発の遅延音の代わりに，振幅値 $-\alpha$ と α を持つ長さ L の PN 系列を遅延音とするものである [26]．図 2.5 に $L = 15$ でのインパルス応答の一例を示した．これにより，単発のエコーよりも個々のエコー成分の振幅を低く $(1/\sqrt{L}$ 倍$)$ 抑えることができ，エコー成分の付加によって生じる周波数特性の乱れもランダムになる．なお，一般的にはサンプリング周期間隔での $L = 1023$ 程度の PN 系列が用いられる [26]．

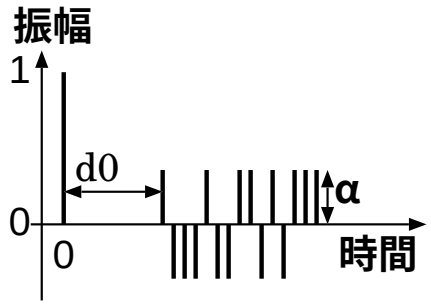

図 2.5: エコー拡散法で用いるインパルス応答の例．

検出時はエコー法と同様に複素ケプストラムを求め，その実部と埋め込み時に用いた PN 系列との相互相関を求めると，遅延時間 $(d0)$ に相当する時刻にピークが表れる．よって，エコー法と同じく遅延時間で秘匿情報を表現することもできるし，スペクトル拡散法における相互相関法と同様な秘匿情報ビット値埋め込み手法と検出手法を用いることができる．また，スペクトル拡散法の特徴である高い秘匿性をも兼ね備えている．この手法については，第 5.2 節にてさらに詳しく説明する．

変調法

ホスト信号に，時間軸に伴う規則的な位相や振幅の変調を与え，その規則性の中に情報を埋め込む手法である．

位相変調を用いる手法では，全域通過時変フィルタを用いてホスト信号に与える位相特

性を周期的に変化させる．情報埋め込みには，変調周波数を複数用いてビット値に割り当てる FSK(frequency shift keying) 方式や，一定周期毎に変調位相を切替えて，その位相値にビット値を割り当てる PSK(Phase Shift Keying) 方式が用いられる．例えば，QPSK 方式であれば，位相値 0 から $\pi/2$ きざみ毎に，00, 01, 11, 10 という 2bit の情報を割り当てることができる．

検出時には，ホスト信号とステゴ信号との位相差から位相変調波形を検出しなければならず，ノンブラインド手法となる．しかし，ホスト信号がステレオ信号の場合，両チャンネルには共通する成分音が存在することを前提に，埋め込み時に左右チャンネルに逆位相の周期的位相差を与えることによって，ブラインド手法として利用できる [27]．検出された位相変調波は，遅延検波回路などによって，ビット情報列に変換できる．

ステゴ信号としてステレオ信号を用いる利点としては，ピッチや時間の伸縮に対して頑強である点が挙げられる．一方，左右信号の比較によって位相変調の存在が検出できるため，秘匿性は低い．

パッチワーク法

パッチワーク (patchwork) 法は画像における電子透かしの代表的手法として知られている．音響信号においては，ホスト信号の時間あるいは周波数，または時間周波数平面上において，特定のふたつの領域を複数選び，一方の領域は強度を強く，もう一方の領域で強度を弱くすることによって強度分布に偏りを持たせ，秘匿情報を埋め込む手法である．つまり，一方の領域を強くすれば 1，その反対は 0，といったかたちで 1bit の情報を表現する．検出時には，多くの領域間の平均強度差を調べることによって，情報を検出する．つまり，個々の領域間の絶対的な強度差は保たれていなくとも，全ての領域間での平均的強度差を期待する統計的な手法であると言える．

パッチワーク法のひとつとして Tachibana et al. [28] の 2 次元ランダム配列法を簡単に解説する．この手法では，一定時間の時間周波数平面 (パターンブロック) を 2 次元に区画分割し，鍵となる PN 系列によって，n ビット分の情報を埋め込むためのそれぞれ m 個の区画と，パターンブロック同期のための区画にラベリングする．図 2.6 左側は，3 ビット分の情報を埋め込むそれぞれ 4 つの区画と，8 つの同期用 (S) の区画にラベリングされたパターンブロックの例を示している．それらの区画には，同じ PN 系列によって，+ あるいは – の符号も同時にラベリングされる．その符号に，埋め込むビット値 (1 あるいは–1) を，掛け合わせて，パターンブロックの各区画内の前後に与えられる強度差パターンが得

られる (図 2.6 右)．

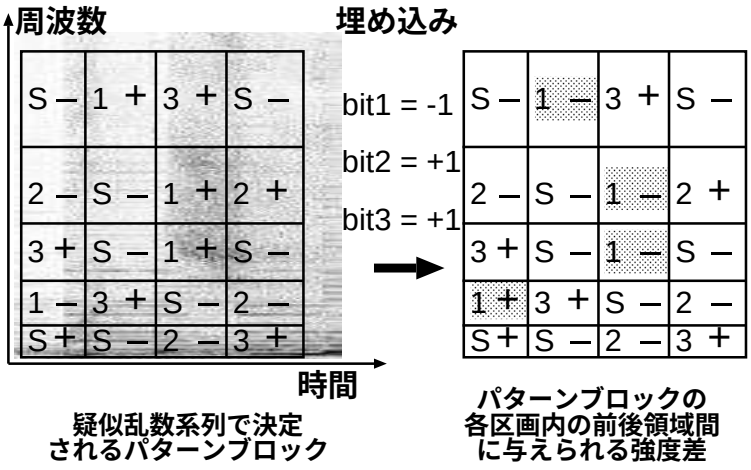

図 2.6: 時間周波数平面上で区切られたパターンブロックへのデータ埋め込みの概略．埋め込む bit1 の値が −1 なので，右側のパターンブロックではハッチングされた bit1 の区画の符号が反転している．

埋め込み処理では，1 区画分の長さのホスト信号は，半分オーバラップさせた窓関数と DFT によって 4 つの時間フレーム分の周波数領域データに変換される．符号が + の区画では，前の 2 フレーム分の振幅を増幅し，後の 2 フレーム分の振幅を減衰させる．符号が − の区画はその逆とする．この処理は，1 つの区画をそれぞれ 2 つの時間フレームによって 2 つの領域に分け，領域間でパッチワーク法の特徴である強度差を与えていることに相当する．最後に逆 DFT により秘匿情報を含んだ波形が得られる．埋め込み強度に相当する，振幅をどの程度増幅あるいは減衰させるかは，心理音響モデルを用いて帯域毎のマスクト閾値を基準に決定する．

検出処理では，ステゴ信号に窓関数を掛けて DFT を施し 1 フレーム分のスペクトルデータとし，窓関数をずらしながら時間周波数平面の強度マップを作成する．フレーム毎の強度を基準化した後，時刻を 1 フレーム分づつずらしながら，鍵によって相対的な位置が決まっているパターンブロック同期用区画で最も大きい平均強度差が得られる時刻を求め，埋め込み時のパターンブロック区間を同定する．次に，各ビット毎に鍵に従って定まる区画での平均強度差が，ゼロと有意に異なるかを統計検定して，ビット値を定める．この手法の実施例では，1 パターンブロックあたり，周波数方向に 30 分割，時間方向に 9 分割 (FFT 長は 1024) することによって 270 区画とし，1 ビットあたり 30 区画を用いて 4 ビット分の情報を埋め込み，残り 150 区画がパターンブロック同期用に用いられた場合の性能が検証されている [28]．

この手法の特徴としては，1 つの領域や区画が時間周波数的に幅を持つため，ある程度の時間方向の伸縮や欠落，ゆらぎに耐性があり，図 2.6 のように周波数帯域分割を対数的とすることにより，ピッチ変換などの周波数方向の伸縮やゆらぎにもある程度の耐性が備わる．また，広い時間周波数範囲に秘匿情報が分散されるゆえ，フィルタリングやノイズ付加などへの耐性も確保できる．

音の強度に対応するスケールファクタを持つ符号化方式 (MP3 や MPEG2 AAC など) の場合，符号化データにこのパッチワーク法による埋め込みを行うことが可能である [29]．この場合は，復号化後の波形に対しても，それに変形が加わったり，さらに再符号化したデータからも秘匿情報の検出がある程度可能である．

周波数ホッピング法

鍵情報あるいはホスト信号に依存する特定の周波数成分に，強度変化を与えて情報を埋め込む手法である．ホスト信号を時間フレームに分割して，DFT によりスペクトルデータに変換し，振幅データに変更を加えて逆 DFT を掛け埋め込み波形を得るのが一般的である．この手法はスペクトル拡散法に分類されることもある．

ホスト信号に依存する周波数成分を変更する例として，埋め込みフレームにおける最もパワーの強い成分に対して 1 オクターブ上の成分強度をゼロとするときビット値 1，1 オクターブ下の成分強度をゼロとするときビット値 0 を割り当てる手法 [30] がある．

また，鍵情報に従って 2 つの帯域を選び，それぞれの帯域の平均パワーに対して，一方の帯域内の特定成分のパワーは +X dB し，もう一方の帯域内の特定成分のパワーを –X dB する手法 [31] もある．

この手法は，埋め込みフレームを同定する手法を別途組み合わせる必要がある．また，周波数成分の選び方によっては，ピッチ変換攻撃を受けると，強度変化を行った成分の周波数がずれることによって，検出が困難になる．

コンテンツ分析法

ホスト信号を分析して，顕著な時間あるいは周波数部位を選んで情報を埋め込む手法である．顕著な時間部位としては，エネルギーが強い部分が挙げられ，その部分以降のある時間区間に，周波数ホッピング法，スペクトル拡散法 [32]，パッチワーク法 [33] などを用いて情報を埋め込む方法が提案されている．エネルギーの強い部位は，知覚的にも重要で

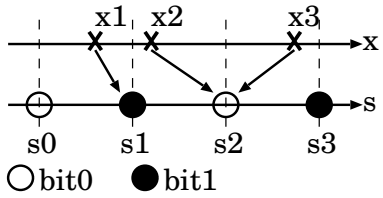

図 2.7: **量子化変調法による埋め込みの模式図**

あり，ステゴ信号の音質を保ったままで，秘匿情報を検出できないような変形をステゴ信号に加えることは困難となる．

特定の部位にしか秘匿情報を埋め込まないので，ランダム刈り取り (random cropping: **標本化された振幅値をランダムに抜き取ったり同じ値を挿入したりする**) **法による攻撃には頑強である．しかし，埋め込み部位を限るということは，逆に言えば埋め込みの余地を残すということであり，埋め込み効率が高いとは言えない．**

量子化変調法

量子化変調 (QIM, quantization index modulation) **法は，ホスト信号の物理量あるいは音響特徴量** x **を，**x **より粗く量子化してビット値を割り当てた値** s **に変換することで情報を埋め込む．図** 2.7 **の上側の数直線はホスト信号における** x **を表しており，**$x1, x2, x3$ **の** 3 **つの値をとる場合を考える．埋め込み後の値** s **は下側の数直線で表されており，それぞれ** $s0$ **から** $s3$ **までビット値** 0**，**1 **に交互に割り当てられている．**$x1$ **から** $x3$ **まで，埋め込むビット値が** $1, 0, 0$ **であると，**x **の値はビット値と対応づけられたもっとも近い** $s1, s2, s2$ **という値にそれぞれ変換される．**

検出時には，$s1, s2$ **の値は，妨害や変形によって，丁度量子化された値よりずれることになるが，ずれが量子化幅の半分までなら，量子化された** s **の値に丸めることによって，正しいビット値を検出できる．**

もっとも単純なのは x **を振幅値とする場合であるが，実際に用いられる音響特徴量** x **としては，成分音の周波数** [34] **や，オクターブバンド幅のエネルギーであったりする** [35]**．音響特徴量として，ある時間窓内のエネルギーなど，振幅に対応した値を選んだ場合は，振幅を増幅/減衰させて雑音を付加する攻撃に対して脆弱になるが，**s **に対するなんらかの基準，例えば特定の周波数帯域のエネルギー** [35] **やパイロット信号のエネルギーなどを定めた場合は，この脆弱性を克服可能である．**

2.5　情報秘匿の補助技術

前節までに説明した基本的な情報埋め込みと検出の手法を実装する際に，性能向上のため補助的に用いる技術がいくつか考案されている．ここでは代表的なものをふたつ紹介する．

2.5.1　聴覚特性を考慮した強度設定

これまでに挙げた情報秘匿手法において，埋め込み処理時にステゴ信号とホスト信号との差分，つまり秘匿情報埋め込みによって付加されたとみなされる波形やスペクトルを算出できれば，その差分は，ホスト信号に加わる雑音成分とみなすことができる．よって，その差分である雑音成分が心理音響マスキングモデル (最小可聴値，時間/周波数マスキング) における閾値以下であれば，ステゴ信号の音質劣化は知覚しにくいと予測できる．

この考え方に従い，短時間フレーム毎に埋め込みの強度を決定する手法は，初期の研究から多く用いられている．例えば，スペクトル拡散法 [24]，エコー拡散法 [26]，パッチワーク法 [28] においては，この手法が積極的に用いられ，ホスト信号のスペクトルに応じた，帯域ごとの埋め込み強度決定が行われている．

一方，変調法を用いた場合は，心理音響マスキングモデルが予測する閾値以上の差成分が付加されても，音質変化が知覚できない場合もある．ステレオ信号を対象とする位相変調法 [27] においては，ホスト信号に対するステゴ信号の音質変化は，音像の左右のゆれや，両耳ビートといった知覚内容として表れ，最小可聴運動角度差 [36] といった聴覚特性を利用した強度の設定が有効であろう．

2.5.2　変形/攻撃を前提とした埋め込み強度設定

典型的な変形がホスト信号に加わった場合の歪みを事前に評価した上で，十分な秘匿情報検出力を保つよう，埋め込みパラメータや強度の設定を短時間フレーム毎に逐一行う手法 (attack characterization 法) である．図 2.8 にこの手法による埋め込み処理の単純な流れ図を示した．この手法は，原理的にどの埋め込み手法にも併用できるが，直接拡散法 [37] や周波数ホッピング法 [31] において利用し，検出率が向上することが報告されている．

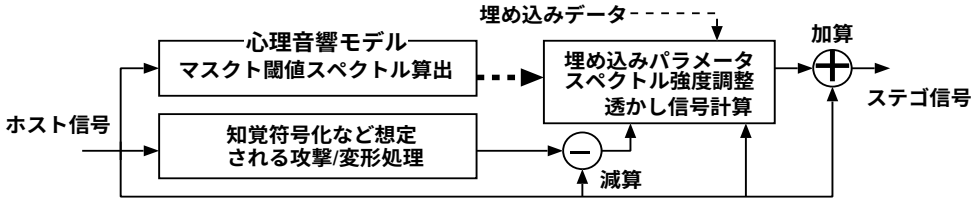

図 2.8: **心理音響モデルを用いた透かし信号スペクトル調整と，変形を前提にした透かしパラメータおよび強度調整**.

2.6 実用化段階の音響透かし技術とその動向

2000 **年** 9 **月，世界各国のレコード会社とオーディオ機器製造会社が参加した組織** SDMI (Secure Digital Music Initiative) **は，安全な音楽データ流通を保証する技術を評価する試みとして，**SDMI Challenge **というイベントを行なった．その一環として，**SDMI **に企業から付託された** 4 **つの音響電子透かし技術について評価する公開試験が行なわれ，楽曲** A **の原曲と透かし入り楽曲，そして同じ技術による透かしの入った楽曲** B **の** 3 **つの波形データが，**Web **サイトを通じて提供された．**

挑戦者は 2 **つの楽曲** A **を分析し，その結果をふまえて楽曲** B **の透かしを無効にした音楽データを** SDMI **サイトに投稿した．透かしが除去された (検出できなかった) かどうかについては，**SDMI **サイトから投稿者に通知メールが返された** [38]**．この挑戦者として自分達の行なった分析過程について報告した文献** [38] **によると，**4 **つの技術はそれぞれ，エコー法，ある帯域のパワーを除去する手法，ある帯域にパイロット信号を挿入する手法，位相変調法であることが明らかになり，大きく品質を劣化させずに透かしを除去する処理が可能であることが示された．**

それらの分析は，初期の音響電子透かし技術が差分攻撃 (原信号とその透かし入り音響信号の差分を解析して手法を同定する攻撃) に脆弱であることを示している．また，音響電子透かし技術の提供企業に，技術内容を公開すると無効にする手段を考案されるという懸念を呼び起こし，また音響電子透かし技術を利用する側にとっても，技術が成熟していないとの念を抱かされることになったことは，想像に難くない．

日本においても，日本音楽著作権協会が主催した，STEP2000 **および** STEP2001 **とよばれる，企業から公募した音響電子透かし技術に対して，**AXB **法による音質劣化の検知および典型的な変形に対する耐性を，委託された機関が評価する試みがそれぞれ** 2000 **年と** 2001 **年に行なわれた．そこで一定の成績を修めたと認定された技術には，パッチワーク法** [28, 39] **や，**PN **信号の代わりに位相関係を** PN **系列で定める正弦波の合成波を用いた**

スペクトル拡散法 [40] が用いられている．

現在，透かし情報を埋め込んでいると公言されているメディアはいくつかある．ひとつは，レコード会社がラジオ局に配布する試聴盤の CD である．これは，透かしが埋め込まれていると公言することにより，試聴盤がオークションや違法ネット配信によって流出することを防ぐ抑止力的効果も持たされている．もうひとつは，ラジオやテレビ放送における番組放送モニタリングである．国土が広く放送局の多いアメリカ合衆国などは，全てのラジオ局において CM 番組が契約どおり放送されているかを人間がチェックするには膨大なコストが必要である．しかし，CM 番組の音声部分に透かし情報を埋め込んでおけば，自動放送受信システムを各地に組むだけで，容易に放送モニタリングが実現可能である．しかし，どのような情報秘匿技術が用いられているかは，明らかになっていない．

一方，音響電子透かし技術がコピー制御を主な目的として規格に組み込まれているのは，DVD-Audio である．しかし，ここでもその技術の詳細は公開されていない．透かしを埋め込むかどうかはコンテンツメディア製造業者に委ねられており，一部の DVD-Audio タイトルを除いては，透かしが埋め込まれていると公言されているタイトルは多くない．

著作権管理に音響透かし技術を利用するのであれば，技術への不安 (技術内容が公開されていないがゆえに安全だと信じたいが，一旦技術内容が明らかになると無効となる恐れがあること) に対して保証を与えるために，技術内容を公開した上での安全性の確保が必要であろう．これは，暗号技術においては暗号化と復号化の両方のアルゴリズムが公開された上で，安全性が証明されることが望まれていることにも対応する．前述の SDMI Challenge において無効化が可能と報告された音響電子透かし技術の中に，DVD-Audio に採用された音響電子透かし技術が含まれていた，という真偽不明な噂も，技術内容が公開されていないゆえに生じるとも考えられる．

一方，ステガノグラフィとしての音響情報秘匿に関しては，技術的には既に利用可能で，埋め込みおよび検出ツールもフリーソフトウェアとして入手可能である．しかし，どこで誰が何のために利用しているか，ということは公になっていない．2001 年の同時多発テロの首謀者たちが，画像や音メディアへのステガノグラフィによって通信を行っていた，という噂もあるが，その真偽は定かではない．

2.7　あとがき

本章は，音響信号への情報秘匿技術について，その概要，評価方法，主な従来技術における秘匿および検出手法を概観した．また，実用化段階の技術の現状についても述べた．

技術の評価方法に関しては，埋め込み情報量や，検出率 (エラー率) によって評価する耐性など，定量的に評価できる指標もあるが，これらの結果は対象とする音響信号に依存するため，少数の音楽信号を用いただけでは，公正な評価が行われているとは言いがたい．秘匿性については定量的な評価方法がまだ確立しておらず，音質劣化に関しても主観評価実験に問題のある従来研究が多いことを示した．さらに，聴覚モデルを用いた客観音質劣化評価法は，知覚符号化音響信号を対象とした仕組みは用意されているが，情報秘匿に伴う音質劣化を評価できるかどうかについては明らかではない．

従来提案されてきた様々な情報秘匿技術を，ディジタル領域のみで有効なもの，アナログ領域を経ても有効なものに分類して概観した．多くの技術がこれまで提案されてきたが，その技術には一長一短があり，利用場面における技術への要求に合うかどうかを検証する必要がある．また，前述したように音質劣化に関する評価が十分でない手法も多い．一方，実用化段階の技術は，その情報秘匿および検出手法が明らかにされていないものが多く，技術に対する秘匿性や音質などの評価は，技術の提供企業の宣伝内容を信じるしかない状態である．