

## 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>

---

出版情報 : 九州大学, 2010, 博士 (芸術工学), 論文博士  
バージョン :  
権利関係 :

# 第5章 既存音響情報秘匿技術との性能比較

## 5.1 まえがき

本章では，既存の音響情報秘匿技術としてエコー拡散法をとりあげ，第3章および第4章において提案し評価した振幅変調に基づく情報秘匿技術との性能比較を行うことを目的とする．

情報秘匿技術の性能評価については，第2.3節において述べたように，まず技術の目的を定めたのち，埋め込み情報量，品質（音質），耐性のうちのいずれかふたつを同等として，残りのひとつの指標において，性能を比較する必要がある．ここではまず電子透かし用途を前提として，音楽信号をホスト信号とする．この場合，埋め込み情報量をほぼ同一とし，音質劣化が客観的評価指標によってほぼ同じ程度となる埋め込み強度で情報秘匿を行った後，ステゴ信号への変形に対する耐性を比較することによって評価を行う．また，スピーカ再生による空間伝搬利用を前提として，音声信号をホスト信号とする．この場合は，電子透かし用途よりもさらに埋め込み情報量を増して，残響および背景雑音に頑強となる埋め込み強度を採用し，耐性を比較する．

## 5.2 エコー拡散法に基づく音響情報秘匿

エコー拡散法は，エコーハイディング法と呼ばれる，秘匿情報を単一のエコーの遅れ時間に割り当てて音響信号にエコーを付加し，ケプストラム上でのピークを求めることによって秘匿情報を取り出す手法 [25] を改善したものである．

エコーハイディング法では，単一のエコーを付加するため，第三者が秘匿情報を容易に検出できる．また，ステゴ音響信号への妨害に対する耐性を高めるためには，エコーの強度を高める必要があるが，それには音質劣化を伴う，という難点があった．

そこで，付加するエコーを単発ではなく， $-1$  および  $1$  の値をもつ疑似乱数 (Pseudo

random Number : PN) 系列によって時間領域で拡散して生成し，検出時にはケプストラム領域で埋め込み時に用いた PN 系列により逆拡散を行うことによって，ピーク時刻を検出する手法 [15] が開発された．これをエコー拡散法とよぶ．この手法は，埋め込み時に用いた PN 系列を用いないと検出が困難なので秘匿性が高く，かつエコーが時間領域で拡散されるために，情報秘匿に伴う音質劣化も知覚されにくいとされている．

### 5.2.1 埋め込み方法

長さ  $L$ ，振幅が  $\pm 1$  である PN 系列  $P(n)$  を用いて，ホスト信号に畳み込むエコーカーネル  $k(n)$  を次式で表す．

$$k(n) = \delta(n) + \alpha P(n - d_0) \quad (5.1)$$

ここで， $\alpha$  はエコーを形成する PN 系列の振幅， $n$  は離散時刻であり， $\delta(n)$  はディラックのデルタ関数である．このインパルス応答  $k(n)$  の波形の例として，長さ  $L = 15$  のものは，第 2 節の図 2.5 に示されている．透かしが埋め込まれたステゴ信号  $r(n)$  は，ホスト信号  $s(n)$  と  $k(n)$  の畳み込みとして，畳み込み記号  $*$  を用いて次式のように表される．

$$r(n) = s(n) * k(n) \quad (5.2)$$

エコーハイディング法 [25] では，時間遅延  $d_0$  の値を長さ  $F$  サンプルの時間フレーム毎に変えることによって，情報を埋め込む．つまり  $d_0$  の値を  $2^j$  段階に変えて設定すれば，1 フレームあたり  $j$  ビットの情報を埋め込むことができる．しかし， $d_0$  の値を大きくすると，聴感上エコー成分が目立ちやすくなる．エコー拡散法においても， $d_0$  の値にビット情報を埋め込む方法は有効であるが， $d_0$  の値が大きいと聴感上エコー成分が目立ちやすくなるのは，エコーハイディング法と同じである．

一方，高らの研究 [15] では， $P(n)$  が含まれる  $k(n)$  を  $s(n)$  に畳み込む場合をビット値 1，そうでない  $k(n)$  (つまり  $\delta(n)$  のみ) を畳み込む場合をビット値 0 として，1 フレームあたり 1 ビットの情報を埋め込んでいる．つまり統計的にいえば半分のデータフレームは，何も処理が行われないことになる．

また， $\delta(n)$  のみの  $k(n)$  を用いずに， $P(n)$  を環状移動 (cyclic shifting) することによって，情報を埋め込むことも提案しているが，これを用いた評価は行われていない [15]．これは，次式のように， $P(n)$  を環状に  $m$  だけシフトした， $P'(n)$  を PN 系列として用いる

方法である．つまり，エコーカーネルの時間長を変化させずに  $m$  を  $0 \sim L$  まで設定することによって， $\log_2(L+1)$  ビットの情報を 1 フレームに埋め込むことができる．この方法は，埋め込みデータ量が多くかつ 1 ビットあたりの検出能力は高らの方法と変わらないので，本稿ではこれを用いることとする．

$$P'(n) = \begin{cases} P(n+L-m) & (1 \leq n \leq m) \\ P(n-m) & (m+1 \leq n \leq L) \end{cases} \quad (5.3)$$

使用される PN 系列は，秘匿情報を検出する際の秘密鍵となる．また，振幅  $\alpha$  の値と，この PN 系列を逆拡散 (相互相関演算) した後の振幅  $g$  との間には，次式のような関係がある．

$$g^2 = \alpha^2 L \quad (5.4)$$

$L$  が一定のとき，この  $g$  または  $\alpha$  が埋め込みの強度となる．ただし， $g > 1$  となるような振幅を設定しても，埋め込み強度を強くすることはできない．これは，エコー成分  $P(n)$  のパワーに対する， $\delta(n)$  成分のパワーが相対的に弱くなることによって，ケプストラム上に現れるエコー遅延時間成分パワーが弱くなるためである．

## 5.2.2 検出方法

式 (5.2) において，両辺に対して離散フーリエ変換 (DFT) 後に対数変換を行い逆 DFT を行う，いわゆるケプストラム変換を行う．ここで，ケプストラム変換は演算記号  $\tilde{\cdot}$  で表し，DFT 演算を  $\text{DFT}$ ，逆 DFT 演算を  $\text{IDFT}$  と表すとき， $\tilde{r}(n) = \text{IDFT}(\log(\text{DFT}(r(n))))$  と表される．すると，次式のように，(5.2) 式の右辺の畳み込み演算はケプストラムの加算として表現できる．

$$\tilde{r}(n) = \tilde{s}(n) + \tilde{k}(n) \quad (5.5)$$

$\tilde{k}(n)$  には， $k(n)$  における， $\delta(n)$  と， $P(n)$  に含まれる各遅延パルスとの間の遅れ時間に対応するピークが周期的に現れる．このため，ステゴ信号のケプストラム  $\tilde{r}(n)$  と， $P(n)$  との相互相関 ( $x_{\text{corr}}$ ) を求めることによって，ホスト信号のケプストラム  $\tilde{s}(n)$  は時間的に拡散される一方， $k(n)$  に含まれる  $P(n)$  成分は逆拡散されることになる．よって，上記の相互相関の結果得られる次式の  $x(n)$  の値は埋め込み時の (5.3) 式における  $m$  にピークを持

つはずである．よって，ケプストラム上で最大ピーク位置時刻を調べることによって埋め込みビット値を検出できる．

$$x(n) = \text{xcorr}(\tilde{r}(n), P(n)) \quad (5.6)$$

### 5.2.3 埋め込み区間同期検出方法

高らのエコー拡散法に関する一連の論文 [15, 26, 70] では，データが埋め込まれた時間区間を，ステゴ信号のみからどうやって検出し特定するかについては，全く述べられていない．

よってここで，埋め込み区間と検出区間の同期手法を著者が検討してみる．まず，埋め込み区間長に対して，検出区間長をちょうど半分とする．この条件で検出処理を行うと，埋め込み区間と連続する2つの検出区間が丁度一致した時に，それら2つの検出区間から得られたデータは最も一致度が高くなるはずである．よって，この演算を連続する複数の埋め込み区間に対して，検出区間幅分だけ検出開始時刻をずらして行えば，埋め込み区間を特定することができる．しかし，この同期検出手法は，ずらし幅の設定方法にもよるが，検出演算処理負荷が非常に高くなる．

どのような手法を用いても，埋め込み区間と検出区間の同期を得る処理には誤差があるはずで，その誤差は検出性能に影響する．よって，本稿ではそのような同期算出処理は用いずに，エコー拡散法については，埋め込み区間は既知であるとして，シミュレーションを行うことにする．このことは，エコー拡散法による秘匿情報検出性能，つまり耐性を過大評価する恐れがあることに注意が必要である．

### 5.2.4 埋め込みパラメータ値

複数の情報秘匿手法の性能を比較する際には，埋め込みデータ量を同等とし，かつ，検出率に影響を与える埋め込み強度は，音質劣化の程度が可能な限り同じなるような強度に設定しておく必要がある．

高らの一連の論文では，埋め込みパラメータ値として  $L = 1023$ ,  $F = 4096$  が用いられている．しかし，これらの値をそのまま用いると，残響付加時に秘匿情報検出性能が非常に悪くなる．これは，短いフレーム時間長を用いると，残響によって前のフレームの波形が次のフレームに重なりあうことによって，検出が困難になるためである．また，次節

におけるシミュレーションにおいては、埋め込みデータのビットレートを双方の情報秘匿手法で揃える必要がある。これらを勘案し、なるべくフレーム長を長くした上で、第 3.6 節で行った音楽信号に情報秘匿する場合のビットレート 4.8 bps と、第 4.2 節で行った音声信号に情報秘匿する場合の 48 bps になるだけ近い値をとるように、パラメータ値を設定した。なお、フレーム時間長  $L$  を微調整すれば、ちょうど 4.8 bps や 48 bps は実現できるが、DFT 演算の高速化のため、 $L$  を 2 の冪乗数の和となるように設定したため、エコー拡散法の埋め込みビットレートは僅かにそれらの値より小さくした。

音楽に埋め込む場合は、振幅変調法と条件を揃えるため、11025 Hz 以下の帯域に埋め込み処理を行い、無処理の高域と加算してステゴ信号を得た。検出時には、11025 Hz 以下の帯域のみに対して検出処理を行った。音声の場合は、このような帯域分割は行わなかったが、もともと用いた音声信号のサンプリング周波数が 16 kHz であったため、必然的に埋め込み帯域は 8000 Hz 以下となっている。

表 5.1: エコー拡散法の埋め込みパラメータ

target	Music	Speech
bit rate [bps]	4.79	47.4
sampling freq. [Hz]	44100	22050
embedding region	below 11025 Hz	below 8000 Hz
delay time ( $d_0$ )	100 samples (2.3 ms)	50 samples (2.3 ms)
length of PN series ( $L$ )	1023	2047
number of embedding bits per frame	10	11
frame length ( $F$ )	92160 samples (2.09 s)	5120 samples (0.23 s)

### 5.3 埋め込みおよび検出シミュレーション実験

第 3.6 節において行った電子透かし用途としての耐性評価を、エコー拡散法についても実施する。つまり、ステゴ信号に対する典型的な変形を経ても、秘匿情報を検出できるかどうかを、RWC 音楽ジャンルデータベース [41] に収録された各種ジャンルの楽曲 100 曲冒頭 1 分間の左チャンネルに対して実施し、第 3.6 節の振幅変調法の結果と比較を行う。

また、第 4.2 節において実施した空間伝搬を前提とした残響および背景雑音に対する耐

性評価を、音声データベースの連続発話音声を用いて行い、第 4.2 節の振幅変調法の結果と比較を行う。

### 5.3.1 埋め込み強度の設定

エコー拡散法による音質劣化の評価としては、音質変化の検知限のみしか報告されていない。音質変化の検知限を求める実験では、一般に被験者に十分な訓練が必要であり、かつ楽曲も検知しやすい曲を選別しなければいけないが、過去のエコー拡散法の研究ではそのような配慮はなされていない。エコー拡散法の改良として、聴覚マスキングモデルを利用して、情報秘匿のために付加されるエコーが検知限以下となるような埋め込み処理を採用した上で、エコー強度の知覚検知限が調べられている [26]。このとき、 $\alpha = 0.005 \sim 0.03$  となるエコー強度において、ポップス音楽信号に対して埋め込みを行った検知限が ABX 法を用いて調べられている。その結果これらのエコー強度は、いずれも被験者 6 名の平均検知限以下 [26] とされているが、論文中で図示されている平均検知率とそのばらつき幅から考慮すると、いずれのエコー強度でも、敏感な被験者は音質変化を十分検知できていたと考えられる。また、この結果は敏感な被験者のみを対象とした訳ではなく、情報秘匿に伴う音質劣化を検知しやすい楽曲を選んだ訳ではないことに注意が必要である。

PEAQ を用いて、エコー拡散法による主観的音質劣化に対応した客観評価ができるかどうかの検証は行われていない。しかし、主観と対応するかどうかはさておき、PEAQ はホスト信号とステゴ信号のある種の物理的差異を定量的に表すことができるので、ここでは PEAQ を用いて、ODG の値を双方の情報秘匿法間で揃えることとした。RWC-MDB-G-2001 の 100 曲について、冒頭 1 分間ステレオ信号への情報秘匿に伴う ODG を求め、平均をとった。なお、左右チャンネルには同じデータを埋め込んだ。図 5.1 は、エコー拡散法によるステゴ音楽信号の ODG 値と、同様にして振幅変調に基づく情報秘匿手法について、第 3.6 節において用いた埋め込みパラメータ値を用いて得られた ODG 値を示している。

この結果からは、エコー法の  $\alpha = 0.006$  と振幅変調法における  $-5$  dB とが、また、 $\alpha = 0.008$  と  $0$  dB との間で、ODG の値がほぼ対応していることが分かる。本節では採り入れていないものの、エコー拡散法は心理音響モデルによるマスク閾値を反映した埋め込みが可能なることから、エコー強度をやや強くした条件において、知覚的音質劣化度合が対応するとみなした。よって、音楽信号に埋め込む場合には、埋め込み強度である PN

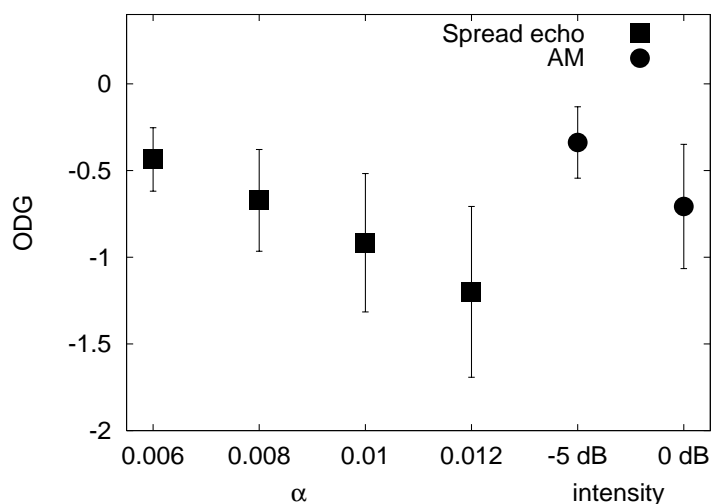


図 5.1: ステゴ音楽信号の ODG 平均値と  $\pm 1$  標準偏差 . RWC-MDB-G-2001 の 100 曲冒頭 1 分間より算出 .

系列の振幅は  $\alpha = 0.008, 0.01$  の 2 通りとした . これらは逆拡散後の PN 系列の振幅  $g = 0.256, 0.320$  に対応する .

エコー拡散法を音声信号に適用した研究例は , これまでのところ無い . 音声信号への客観劣化評価として , PESQ を用いて , 音声品質劣化度を双方の情報秘匿手法間で揃えることも検討したが , PESQ はエコーが付加された音声信号については , 正当な評価ができない旨が規格書 ITU-T P.862 に述べられているので , ここでは用いなかった . よって , 音声信号に埋め込む場合のエコー強度の設定は , 耐性を高めるためエコー強度を大きくした ,  $\alpha = 0.2, 0.3$  の 2 通りとした . これらはそれぞれ ,  $g = 0.640, 0.960$  に対応し , もはや  $g$  の値は限界値である 1 に近い .

### 5.3.2 情報秘匿済み音楽信号に対する変形

第 3.6 節において行った音楽信号を対象とした電子透かし用途としての耐性評価を , エコー拡散法について実施した .

ステゴ信号に対する知覚符号化および復号化として , RealAudio8 (44.1, 32, 21 kbps/ch) , MP3 (64, 48, 32 kbps/ch) を用いた . RealAudio8 の 21 kbps にはサンプリング周波数 22.05 kHz への変換が , MP3 の 48 および 32 kbps の条件にはサンプリング周波数 32 kHz への変換が , それぞれのエンコード処理に含まれている .

残響付加は , 正規分布乱数に指数減衰を与えて 0.25, 0.5, 1, 1.5 秒の残響をもつインパ



ルス応答を作成し，埋め込み後信号に畳み込むことによって行った．ホワイトノイズ付加は，1 分間の音楽全体の実効値に対するノイズレベルとして， $-40, -30, -20$  dB (SNR 40, 30, 20 dB) を用いた．

典型的な雑音として付加するホワイトノイズは，1 分間の音楽全体の実効値に対するノイズレベルとして， $-40, -30, -20$  dB を用いた．これは信号対雑音比 (SNR) でいうと，それぞれ 40, 30, 20 dB となる．

時間長変換としては，ランダム苅込 (Random Cropping) 法を用いた．これは，ステゴ信号に対して一定時間区間ごとにランダムに 1 サンプル選んで抜き取ることによって時間を短く，また一定時間区間ごとにランダムに 1 サンプル選んで同じ値を隣に追加することによって，時間を長く変換するものである．250, 333, 500, 1000 サンプル毎に，上述の処理を行い， $-0.4\% \sim +0.4\%$  の時間長変換を模擬した．なお，第 3.6 節において実施した時間長伸縮処理とピッチ変換処理は，それぞれの変換処理自体のアルゴリズムによって結果が影響を受けることを防ぐため，時間長伸縮およびピッチ変換があったものとして検出側のパラメータを埋め込み側と変えることによって実現したが，ここでは両情報秘匿手法に共通して使える時間長変換アルゴリズムとして，ランダム苅込法を採用している．

### 5.3.3 情報秘匿済み音声信号に対する変形

第 4.2 節において実施した空間伝搬を前提とした残響および背景雑音に対する耐性評価を，エコー拡散法に対して行った．

音声信号としては日本音響学会研究用連続音声データベース Vol. 1 に収録されている音声を，話者ごとに連結して，1 話者あたり 36 秒分の音声信号を 22 名分 (男性 10 名，女性 12 名) 作成して用いた．

残響のある環境を想定して，第 4.2 節と同様に，情報秘匿済み音声信号に対して，RWCP 実環境音声・音響データベースより選んだ，残響時間 1.3 秒の可変残響室で収録されたインパルス応答 (ファイル名: ir130.dat) を畳み込んだ．また，このインパルス応答の収録環境である可変残響室は，次に述べる環境騒音の収録環境とは異なるが，このインパルス応答ファイルは，他研究者が容易に入手可能でき，実環境で測定したものであるため，これを用いた．

その後，第 4.2 節と同様に，4 種類の環境騒音 (収録場所: 駅のホーム，地下連絡通路，空港口ビー，混雑した交差点)，あるいはローパスノイズ (カットオフ 500Hz， $-9$  dB/oct.)

を付加した後，秘匿情報を検出する処理を行った．これらの音のサンプリング周波数は 22.05 kHz にダウンサンプリングした．環境騒音およびローパスノイズは，オーバーオール音声信号パワーに対して，信号対雑音比 (SNR) は 10dB と 20dB の 2 通りとした．

よって，22 種の情報秘匿済み音声と 5 種の背景雑音を組み合わせて 110 条件が，2 段階の埋め込み強度と 2 段階のノイズ強度条件の組み合わせに対して用いられた．

## 5.4 実験結果

実験結果は，埋め込んだランダムなビット値に対する，正しく検出できたビット値の割合を検出率として評価した．

### 5.4.1 音楽信号での検出結果

図 5.2 に，振幅変調法により情報秘匿したステゴ音楽信号を RealAudio8 符号化し，復号化した後のビット検出率を示した．グラフの誤差棒は全 100 曲中の最大検出率と最小検出率， は 90 パーセンタイル値， は 10 パーセンタイル値，真中の点は中央値を示している．同様に，図 5.3 には，エコー拡散法によるステゴ信号を RealAudio8 符号化し，復号化した後のビット検出率を示した．グラフの凡例は同じである．以下，MP3 符号化および復号化後の検出率について，図 5.4 には振幅変調法の結果，図 5.5 にはエコー拡散法の結果を示した．雑音付加後の検出率については，図 5.6 に振幅変調法の結果，図 5.7 にはエコー拡散法の結果を示した．残響付加後の検出率については，図 5.8 に振幅変調法の結果，図 5.9 にはエコー拡散法の結果を示した．時間長変換後の検出率については，図 5.10 に振幅変調法の結果，図 5.11 にエコー拡散法の結果を示した．

本節における振幅変調法の結果は，時間長変換の結果を示した図 5.10 以外は，全て第 3.6 節のデータを再掲したものである．

これらの結果から検出率の中央値を見ると，残響付加と時間変換を除いた変形処理については，振幅変調法とエコー拡散法の検出率はほぼ同等か，エコー拡散法の方がわずかに良い場合が多いことが分かった．埋め込み強度は，振幅変調法の  $-5$  dB がエコー拡散法の  $\alpha=0.008$  に対応し， $0$  dB が  $\alpha=0.01$  にほぼ対応している．しかし，これらの変形処理に対して，最低検出率および 10 パーセンタイルの検出率の値は，常にエコー拡散法が振幅変調法を下回っている．このことは，エコー拡散法の場合は，秘匿情報検出が極端に困難となる楽曲が存在することを示している．

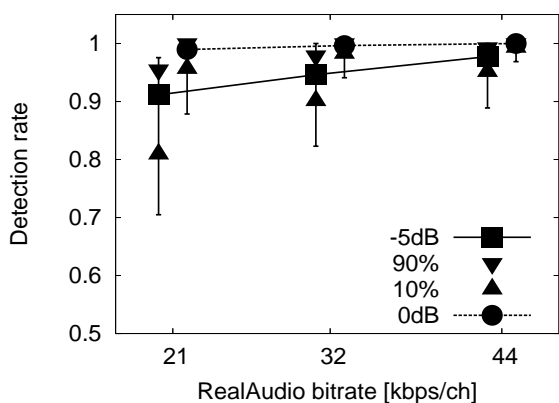


図 5.2: 振幅変調法により音楽信号に情報を埋め込んだ場合の RealAudio8 符号化および復号化後検出率. エラーバーは最大と最小値, は 90 パーセンタイル値, は 10 パーセンタイル値, 真中の点は中央値を示している.

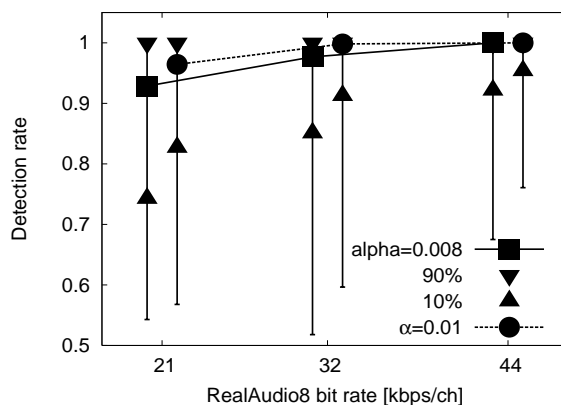


図 5.3: エコー拡散法により音楽信号に情報を埋め込んだ場合の RealAudio8 符号化および復号化後検出率. 凡例は図 5.2 と同じ.

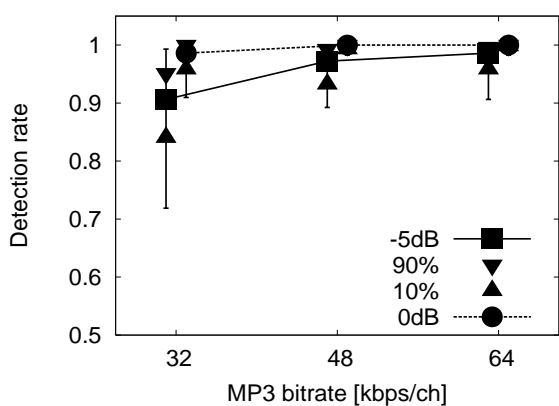


図 5.4: 振幅変調法により音楽信号に情報を埋め込んだ場合の MP3 符号化および復号化後検出率. 凡例は図 5.2 と同じ.

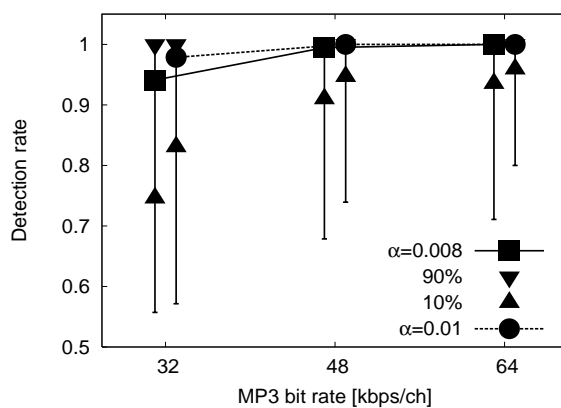


図 5.5: エコー拡散法により音楽信号に情報を埋め込んだ場合の MP3 符号化および復号化後検出率. 凡例は図 5.2 と同じ.

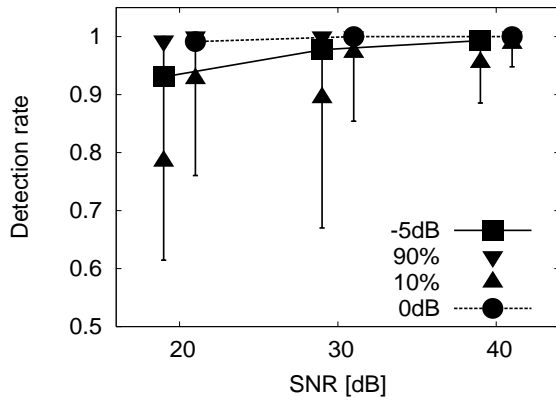


図 5.6: 振幅変調法により音楽信号に情報を埋め込んだ場合の雑音付加後検出率．凡例は図 5.2 と同じ．

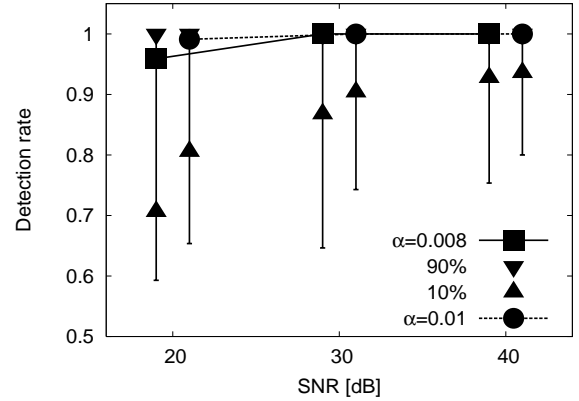


図 5.7: エコー拡散法により音楽信号に情報を埋め込んだ場合の雑音付加後検出率．凡例は図 5.2 と同じ．

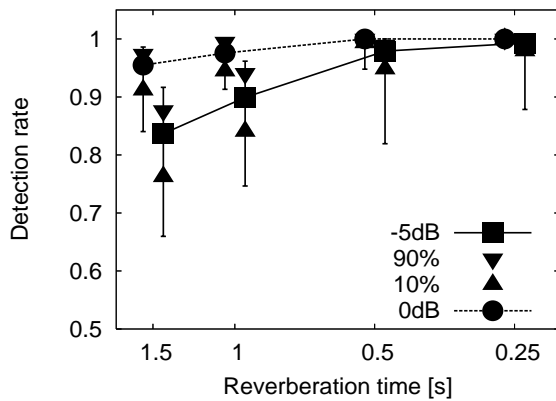


図 5.8: 振幅変調法により音楽信号に情報を埋め込んだ場合の残響付加後検出率．凡例は図 5.2 と同じ．

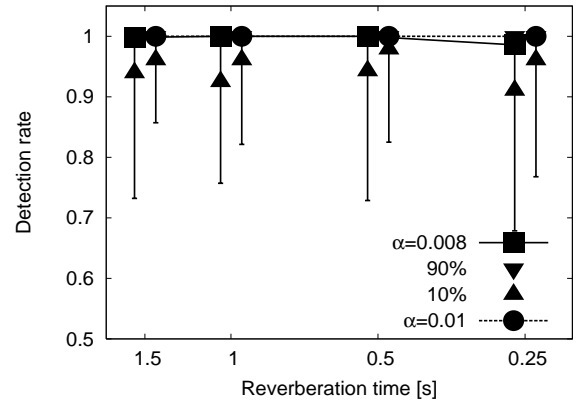


図 5.9: エコー拡散法により音楽信号に情報を埋め込んだ場合の残響付加後検出率．凡例は図 5.2 と同じ．

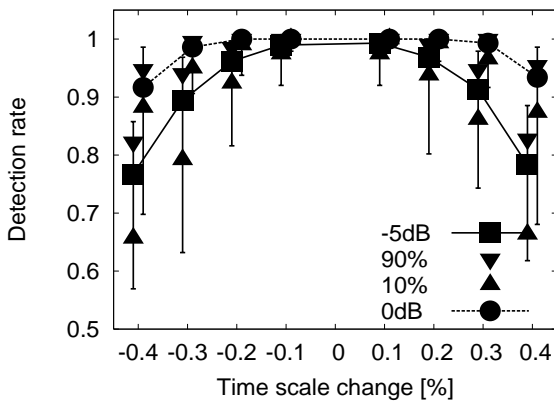


図 5.10: 振幅変調法により音楽信号に情報を埋め込んだ場合の時間変換後検出率．凡例は図 5.2 と同じ．

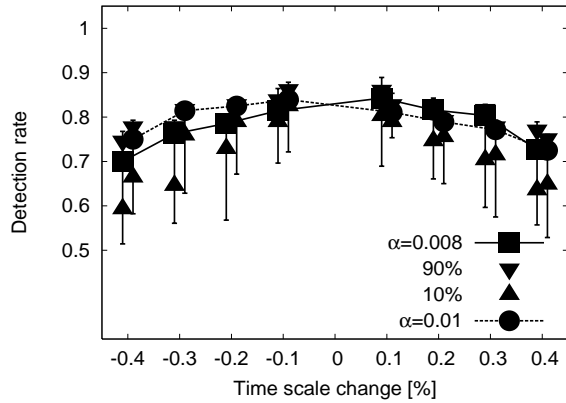


図 5.11: エコー拡散法により音楽信号に情報を埋め込んだ場合の時間変換後検出率．凡例は図 5.2 と同じ．

残響付加に対してはエコー拡散法の方が，1 秒以上の残響時間において，より高い検出率を示した．このシミュレーション実験を行う前の予備調査の結果，エコー拡散法はフレーム時間長が短い (1 秒未満程度) 場合には，残響に対して脆弱であることが分かったが，今回のようにフレーム時間長を 2 秒程度と長くすれば，残響に対して十分に頑強となることが明らかになった．しかし，1.0 秒以下の残響時間では，最低検出率はエコー拡散法の方が下回っている．

一方，ランダム莉込時間変換処理を行うと，エコー拡散法は大きく検出率を落とし，振幅変調法は  $\pm 0.2\%$  までは良好な検出率を保った．この理由として，エコー拡散法では，ステゴ信号の時間長が変化すると，検出時に計算されるケプストラムの遅れ時間軸がずれることによって，有効な逆拡散が困難になることが考えられる．同じ問題はピッチ変換についてもいえるため，これらを克服するために，ホスト信号の区間毎の代表的ピッチ周波数をあらかじめ算出し，拡散インパルス応答に対してピッチ周波数に依存した伸縮を行った後に畳み込みを行って情報を埋め込む改善方法が既に提案されている [71]．しかし，ステゴ信号に対する代表的ピッチ周波数検出が必ずしも成功する訳ではないので，ステゴ信号に対して周波数変換が行われない場合にも検出率が 85% 程度に留まってしまふ，言い替えば MP3 を含む他のあらゆる変形に対する耐性が下がってしまう，という欠点をもつ．

振幅変調法では，時間長変換に伴って生じる周波数の変化によって，埋め込み時のフィルタバンクの帯域幅と検出時のフィルタバンクの帯域幅がずれることによって検出が困難になると考えられる．

時間長変換の生じる一般的に大きな原因のひとつである，アナログコピー時のディジ

タルオーディオ機器間の典型的なサンプリング周波数の偏差は、0.1% 未満であることが知られているので、振幅変調法に関しては、悪意のある利用者が行うような極端な時間長変換でなければ、問題ないことが分かった。しかしエコー拡散法では、ステゴ信号が DA/AD 変換のような、サンプリング周波数の僅かな変化によって時間長が変化するような条件を経ると、検出が困難になることが分かった。

音楽信号に情報を秘匿し検出するシミュレーション実験結果をまとめると、双方の情報秘匿手法ともに、知覚符号化、雑音付加、残響付加に対して、平均的には実用上問題無い検出能力を示すことが分かった。しかし、エコー拡散法の方が、多くの条件で 100 曲の楽曲中の 10 パーセントイル検出率および最低検出率が低く、秘匿情報検出が困難となる楽曲が存在することを示している。言い換えると、振幅変調法の方が、あらゆる楽曲に対して適用できる性能を備えているといえる。

また、エコー拡散法は時間長変換に脆弱であり、総合的に見ると振幅変調法の方が、様々な種類の音楽信号に適用可能で、ステゴ信号に対する典型的な変形を経ても、有効な秘匿情報の検出が可能であることが分かった。

#### 5.4.2 音声信号での検出結果

図 5.12 に、音声信号に振幅変調法を用いて情報秘匿した場合の検出率を、横軸を埋め込み強度である振幅変調度として示した。図 5.13 に、音声信号にエコー拡散法を用いて情報秘匿した場合の検出率を、横軸を埋め込み強度である  $\alpha$  の値として示した。

これらの結果から、振幅変調法は変調強度 0.4 で SNR 10 dB の、最も検出性能が悪化する条件でも、半分以上のシミュレーション環境下で 90%以上の検出率を保つことが分かった。一方、エコー拡散法は最も検出性能が高くなる条件である、 $\alpha = 0.03$ , SNR 20 dB の条件ですら、全てのシミュレーション環境下で検出率が 90%を下回ることが分かった。よって、ステゴ音声信号に残響が付加され、環境雑音が混入するような環境下では、振幅変調法が有効であり、エコー拡散法は不適當であることが分かった。

音楽信号に対する情報秘匿において、雑音付加や残響付加に対して良好な性能を示したエコー拡散法が、なぜ雑音付加と残響付加が同時に生じる音声信号に対する埋め込み条件で性能を発揮できなかったかということ、音声信号への情報秘匿においては、秘匿データ量を増やすために埋め込みフレーム時間を短くせざるを得なかった、ということが最も大きな原因である。このため、検出時のケプストラム領域において、残響や雑音のスペクトル

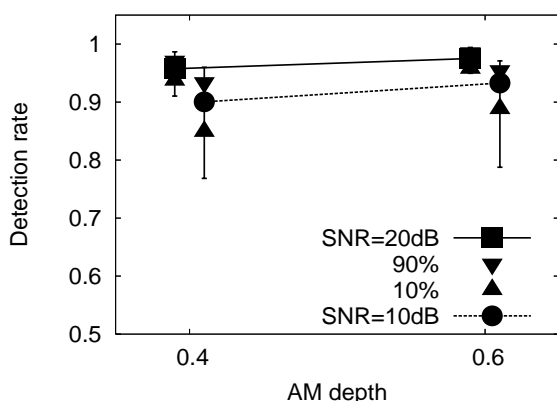


図 5.12: 振幅変調法により音声信号に情報を秘匿した場合の検出率．話者と背景雑音を組み合わせた 110 条件における，中央値と最大最小値，10 および 90 パーセントイル値が示されており，凡例は図 5.2 と同じ．

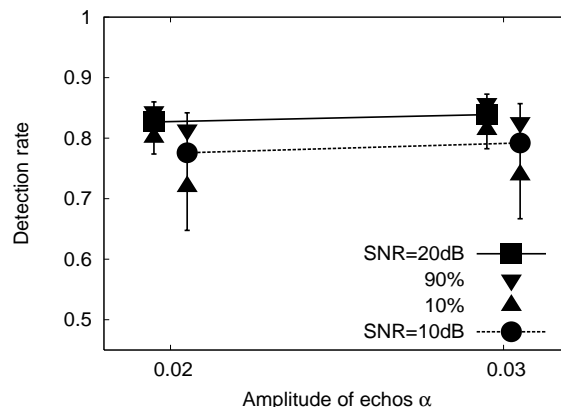


図 5.13: エコー拡散法により音声信号に情報を埋め込んだ場合の検出率．凡例は図 5.12 と同じ．

が平均化されずに残響や雑音の短時間の変動が大きく表れて，検出性能に悪影響を及ぼした，ということである．

## 5.5 考察

エコー拡散法は，インパルスとそれに対する遅延パルスをホスト信号に畳み込むことによって，埋め込み済み信号の周波数軸方向に，遅延パルスの数に対応する数の周期的な強度の変化を与えることになる．このステゴ信号に対して，平坦なスペクトルを持つホワイトノイズを付加しても，周波数軸方向の周期的な強度の変化に影響は少ない．スペクトルの平均時間とも言えるフレーム時間長を長くした場合には，さらに雑音のスペクトルは平坦になるため，雑音による検出性能の劣化は生じ難くなる．しかし，時々刻々スペクトルが比較的大きく変化するような環境雑音を付加した場合は，この雑音が透かし埋め込み済み信号の周波数軸方向の強度に変化を与えることになり，検出性能を劣化させる．これらの要因も，音楽信号にホワイトノイズを付加したシミュレーションでの検出率は高く，音声信号に環境雑音を付加したシミュレーションでの検出率が低くなった原因のひとつである．

本章では，振幅変調法とエコー拡散法との性能を比較した．現在までに発表された多く

の音響電子透かし手法は、そのほとんどが知覚符号化に対する耐性を備えているとされているが、残響に対する耐性が高い手法は少ない。また、埋め込み対象となる音響信号の特性に多少なりとも影響を受ける手法がほとんどであるにも関わらず、第3.6節および第4.2節のように様々な特徴を持つ広範囲の音響信号、とくに音声信号への埋め込みや検出耐性を調べた研究はほとんど無い。そこで、他の音響情報秘匿手法について、今回のシミュレーション条件における性能を定性的に考察する。

まず、ほとんどの既存音響情報秘匿手法は、データを埋め込むひとつの時間区間であるデータフレーム長は1秒より短い。このことは、音声信号のように、1秒より長いポーズ(無音部分)を含んだ信号においては、そのデータフレームにはデータ埋め込みが不可能であることを示している。今回の第5.3.3節の埋め込み対象音声信号は、途中で1~2秒のポーズが複数回含まれており、エコー拡散法においても、この無音区間に相当したデータフレームにて検出不能であることが、検出率を落とした最大の原因である。これに対して、振幅変調法は4秒間のデータフレーム長を持ち、この区間にわずかでも音声信号が含まれていれば、ある程度のデータ埋め込みと検出が可能である。この点からしても、振幅変調法は音声信号に対する情報秘匿に有利であると言える。

残響に対する既存音響情報秘匿手法の耐性の低さは、第4.1節にて述べた。これらを概観すると、短い時間フレーム(数100ms以下)に少ない情報を埋め込み、全体として埋め込みデータレートの要求を満たす従来の情報秘匿手法より、比較的長い時間フレーム(3~5秒)において周波数方向に分散させて多くのデータを埋め込み、データレートの要求を満たす手法の方が、残響に対して頑強であると言える。振幅変調法は後者であり、他にこのような条件を満たす音響情報秘匿手法は著者の知る限り無い。エコー拡散法は、基本的に前者の条件において用いることを前提としており、後者の条件において用いれば残響に対する耐性が高いことは第5.4節のシミュレーション結果により分かったが、それと高い埋め込みデータレートの両立は困難であることも、同時に明らかになった。

## 5.6 あとがき

第3章において提案した振幅変調に基づく音響情報秘匿手法と、従来から提案されているエコー拡散法との性能比較を行った。

電子透かし用途としての性能を検証するために、双方の情報秘匿手法の埋め込み情報量を4.8 bpsで揃え、PEAQによってほぼ同等の客観音質劣化度合となる埋め込み強度



にて、それぞれの手法により情報秘匿を行った。そして、第 3.6 節において行った、RWC 音楽ジャンルデータベースに含まれる様々なジャンルの楽曲 100 曲を対象とした、データ埋め込みおよび検出シミュレーション実験をエコー拡散法に対して行った。その結果、知覚符号化と復号化、残響、ノイズ付加などの変形に対して、双方の情報秘匿手法は平均的には十分な耐性を持つことが分かった。しかし、エコー拡散法の方が、多くの条件において、全 100 曲に対する 10 パーセント検出率および最低検出率が低く、秘匿情報検出が困難となる楽曲が存在することが分かった。言い換えると、振幅変調法の方が、あらゆる楽曲に対して適用できる性能を備えていると言えた。また、エコー拡散法は時間長変換に対して脆弱である。よって、総合的に見ると振幅変調法の方が、様々な種類の音楽信号において、ステゴ信号に対する典型的な変形を経ても、有効な秘匿情報の検出が可能であり、電子透かし用途に適していることが分かった。

また、第 4.2 節において行った、音声信号に対して 48 bps の情報秘匿を行いスピーカ再生を前提として残響と環境背景雑音が付加されるシミュレーションを、エコー拡散法に対して行った。実験では、振幅変調法は最も検出性能が低い妨害条件においても、半分以上の音声信号で 90%以上の検出率を示す一方で、エコー拡散法は検出率が最も高い妨害条件でも検出率 90%に達することはなかった。よって、音声信号に多くのデータを埋め込んでスピーカ再生し、マイクロホン受音した音からデータを検出して利用するような応用場面においても、振幅変調法の方が適していることが明らかになった。