

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3章 振幅変調に基づく情報秘匿技術

3.1 まえがき

第2章では、従来の音響情報秘匿技術を概観し、技術の評価方法に関して検討を行った。そして、音響情報秘匿技術の評価のためには、様々な特徴をもった音響信号に対応できること、また音質劣化の評価を適切に行うことが重要であるが、これらは従来の研究では不十分であることを示した。

本章では、秘匿情報の検出時に情報を埋め込む前の信号を必要としない、いわゆるブラインド検出が可能な、帯域分割と振幅変調に基づいた新しい情報秘匿および検出の手法を示す。この手法は、様々な特徴をもった音響信号に対応した秘匿情報の検出を可能にするよう、ホスト信号の変調強度に合わせて振幅変調強度を埋め込み時に自動的に設定する方法を用いることが特徴である。そして、電子透かし用途としての一般的な信号変形に対する耐性、および空間伝搬時に生じる変形の代表でもある残響に対する耐性を、様々なジャンルの音楽を含む音楽データベースの楽曲100曲 [13] を用いて検証する。

また、音質劣化評価に関しては、被験者にMP3符号化による音質劣化も同時に評価させることにより、典型的な音質劣化に対する検知能力をもった被験者を選定する。十分に検知訓練を積んだ被験者に対して、情報秘匿に伴う音質劣化が比較的検出しやすい楽曲を用いて、知覚検知限となる埋め込み強度を明らかにする。そして、検知限以上の強度で秘匿情報を埋め込んだ場合の音質劣化度合を調べるため、極めて僅かな音質劣化を評価できるよう、ITU-R BS.1116-1 に準拠した方法を用いた主観評価実験を行う。得られた主観的な音質劣化度合と、知覚符号化に伴う音質劣化を予測するために提案されている客観音質評価法PEAQ [14] を用いて得られた、情報埋め込み済み音楽信号の客観音質劣化度合とを比較し、対応関係がみられるかどうかを調べる。

3.2 埋め込み方法

情報埋め込みは、ホスト信号の低周波数帯域に対して行う。データフレーム時間長 T_p 秒のホスト信号 $H(t)$ ($0 \leq t < T_p$) は、処理を行わない高周波数帯域信号 $H_{\text{high}}(t)$ と、等帯域フィルタバンクにより n 個の帯域信号ペア $h_{2m}(t)$ および $h_{2m-1}(t)$ ($m = 1, 2, \dots, n$) に分割される。本方式では、この2つの隣接する帯域に分割された信号同士にそれぞれ逆位相の正弦振幅変調を与える (図 3.1 参照)。

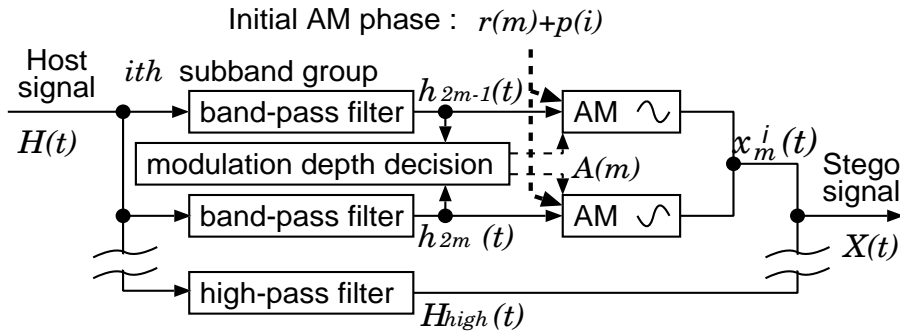


図 3.1: 帯域ペアへの逆位相振幅変調。

$$H(t) = \sum_{m=1}^n (h_{2m-1}(t) + h_{2m}(t)) + H_{\text{high}}(t), \quad (3.1)$$

このペアとなる隣接帯域を複数含む k ($2 \leq k$) 個のグループに各帯域を分類し、そのグループ間の振幅変調位相差に秘匿情報を埋め込む。 i 番めのグループに属する m 番めの帯域ペアの出力信号波形 $x_m^i(t)$ は次のように表される。

$$x_m^i(t) = h_{2m-1}(t) \left(1 + A(m) \sin(2\pi ft + r(m) + p(i)) \right) + h_{2m}(t) \left(1 - A(m) \sin(2\pi ft + r(m) + p(i)) \right) \quad (3.2)$$

ここで、変調初期位相 $r(m) + p(i)$ のうち、 $r(m)$ は埋め込み鍵によって決定される疑似乱数系列によってあらかじめ与えられる帯域ペア毎の位相である。 $p(i)$ ($i = 1, 2, \dots, k$) は、帯域グループの初期位相であり、2 bit のグレイ符号化 (隣接する値間で1ビットのみ異なるような10進数から2進数への符号化) を施した秘匿情報 ($D_i \in \{0, 1, 2, 3\}$) に対して、次式のように表現される。

$$p(i) = \begin{cases} 0 & i = 1; \\ \frac{\pi D_i}{2} & i = 2, \dots, k. \end{cases} \quad (3.3)$$

$p(1) = 0$ の位相はパイロット帯域グループとよばれる，秘匿情報埋め込み時の変調初期位相の基準となる帯域グループに与えられる．つまり，パイロット帯域グループの変調位相に対する，他の帯域グループの変調位相の値に情報が埋め込まれる．なお，データフレームの開始時刻および終了時刻において急激な振幅の変化が生じないように，変調は必ずゼロあるいは π 位相から開始し，かつ終了するように調整する．ステゴ信号 $X(t)$ は，振幅変調後の信号 $x_m(t)$ を全て加算し，高域信号を $H_{\text{high}}(t)$ を加算して得られる．

$$X(t) = \sum_{m=1}^n x_m(t) + H_{\text{high}}(t) \quad (3.4)$$

埋め込みデータ D_i は T_p 秒のデータフレーム周期毎に更新され，このフレーム周期毎に基本となるパイロット帯域グループ ($k = 1$) の変調初期位相を π だけ反転させることによって，検出時にデータ埋め込み区間の同期検出 (第 3.5 節を参照) を可能とする．

全ての帯域ペア同士の位相は，鍵によって決定される位相 $r(m)$ によって攪乱されているうえ，どの隣接帯域同士が帯域ペアとなり，どの帯域ペアがパイロット帯域グループに属し，どの帯域ペア同士が同じ帯域グループに属するかは，全て埋め込み時の鍵によって決定できるため，どの帯域ペアあるいはグループ間の変調位相差に情報が秘匿されているかを容易に知られることはない．帯域グループには複数の帯域ペアが属しており，検出時には鍵情報によって位相 $r(m)$ を補正して検出された変調波を同期加算するため，個々の帯域ペアの変調度は低く抑えることができ，個々の帯域の変調周波数や位相をステゴ信号のみから検出することは困難である．これは一般的な音響信号は，埋め込みに用いる 10Hz 以下の変調周波数帯域に振幅変動を持つからである．

さらに， M 個の異なる変調周波数を同時に用いて，多重に埋め込みを行うことが可能である．この場合，埋め込み情報量のビットレートは， $2M(k - 1)/T_p$ bps となる．

3.3 検出方法

図 3.2 に，秘匿情報の検出過程を模式図で示した．秘匿情報検出時には，鍵情報に従ってペアとなる隣接する帯域信号を同定し，それらの振幅包絡波形をそれぞれ求め，それらの比の対数を振幅変動波形として抽出する．以下にその処理を，式によって述べていく．

ここでは， T_f 点 DFT を用いて， $T_f/4$ づつ波形時刻をずらしながら得られる時間—周波数ベクトル $\mathbf{E}(\tau)$ を，次式で表す．DFT($X(t)$) という演算子は， $X(t)$, ($0 \leq t < T_f$) の

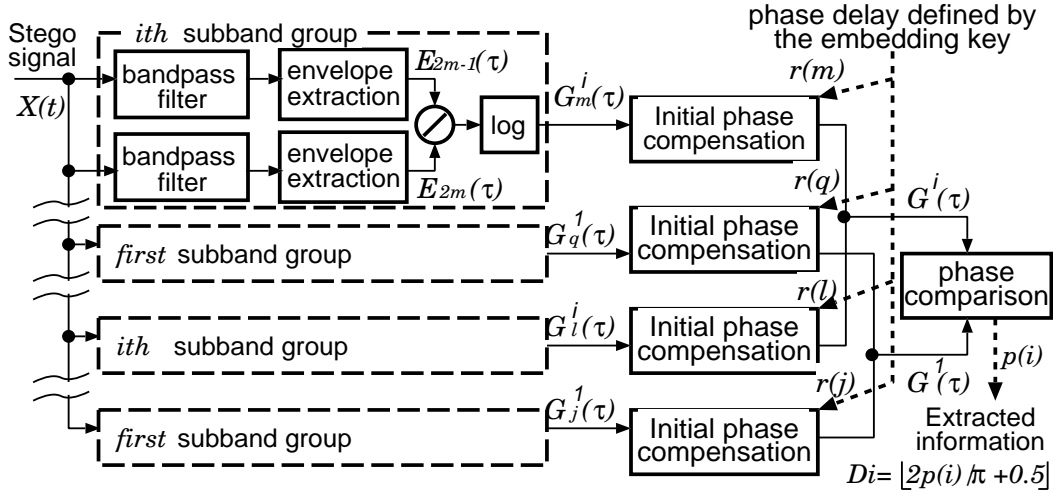


図 3.2: 埋め込まれた秘匿情報検出の過程 .

波形に対して DFT を行って周波数領域に変換することを意味している .

$$E(\tau) = \text{abs}(\text{DFT}(X(t + \tau T_f/4))), (0 \leq t < T_f) \quad (3.5)$$

τ は振幅包絡波形の時刻の変数となる . $E(\tau)$ から , 埋め込み時の m 番めの帯域の時間包絡波形を取り出し , これを $E_m(\tau) (m = 1, 2, \dots, 2n)$ と表す . 検出過程での , $2m$ 番めの帯域の振幅包絡波形は , ホスト信号の振幅包絡波形を $S_{2m}(\tau)$, ステゴ信号への変形によって生じる時間波形成分を $N_{2m}(\tau)$, m 番めの帯域ペアに対する正弦振幅変調波形を $AM_m(\tau) = A(m) \sin(2\pi f\tau + p(i) + r(m))$ とすると , 次式のように表される .

$$E_{2m}(\tau) = (1 - AM_m(\tau))S_{2m}(\tau) + N_{2m}(\tau) \quad (3.6)$$

ここで , 帯域ペアの振幅包絡波形の比の対数 $G_m(\tau)$ は , 次式で表される .

$$\begin{aligned} G_m(\tau) &= \log \frac{E_{2m-1}(\tau)}{E_{2m}(\tau)} \\ &= \log \frac{S_{2m-1}(\tau)}{S_{2m}(\tau)} + \log \frac{1 + AM_m(\tau) + \frac{N_{2m-1}(\tau)}{S_{2m-1}(\tau)}}{1 - AM_m(\tau) + \frac{N_{2m}(\tau)}{S_{2m}(\tau)}} \end{aligned} \quad (3.7)$$

ここで , $x = AM_m(\tau) + N_{2m}(\tau)/S_{2m}(\tau)$ とおき , 一般に $|x| < 1$ であるので , マクローリン展開の第 2 項までを用いる近似によって , $\log(1 + x) \approx x - x^2/2$ が与えられる . この近似を用いて式 (3.7) を書き換える .

$$\begin{aligned} G_m(\tau) &\approx \log \frac{S_{2m-1}(\tau)}{S_{2m}(\tau)} + \left(2 - \frac{N_{2m}(\tau)}{S_{2m}(\tau)} - \frac{N_{2m-1}(\tau)}{S_{2m-1}(\tau)} \right) AM_m(\tau) \\ &\quad + \left(\frac{N_{2m-1}(\tau)}{S_{2m-1}(\tau)} \right)^2 - \left(\frac{N_{2m}(\tau)}{S_{2m}(\tau)} \right)^2 \end{aligned} \quad (3.8)$$

式 (3.8) は、ステゴ信号への変形が元になって生じる $N(\tau)$ がホスト信号の帯域振幅包絡 $S(\tau)$ より十分小さければ、 $G_m(\tau)$ は、変調波形 $AM_m(\tau)$ の特徴をよく表すことを示している。よって、第 3.4 節に示すように、第 1 項 $\log \frac{S_{2m-1}(\tau)}{S_{2m}(\tau)}$ における、埋め込み変調周波数 f に相当する変動成分強度に比例して、振幅変調強度 $A(m)$ を設定することが必要となる。

この後、埋め込み時の鍵に基づいて帯域間の変調位相 $r(m)$ を補正し、埋め込み時に定めた帯域グループ毎に変動波形の同期加算を行い、変動成分を強調する。 M 個の帯域ペアの同期加算が行われると、式 (3.8) の第 1 項や第 3、第 4 項の変動成分は、ランダムな位相で加算されるため、それらのパワー増加に比べて、同期加算される第 2 項の埋め込み変調周波数成分のパワー増加は理論的に \sqrt{M} 倍となる。この仕組みによって、ステゴ信号への知覚符号化圧縮処理や雑音付加に対する耐性が高まる。さらに、ステゴ信号に対するフィルタリングやイコライジングといった周波数特性上の変形によって、帯域ペア間の振幅比が変化しても、振幅比の変化は、式 (3.8) の第 1 項においては定数項となるので、変調波検出に影響を及ぼしにくい。

秘匿情報は、同期加算後の振幅変動波形から変調周波数 f に相当する変動周波数成分の位相を算出し、帯域グループ間の振幅変動位相差より復号化して求める。

埋め込み時に高い振幅変調周波数を用いると、振幅変調に伴って生じる側波が聴感上知覚されやすくなり、かつステゴ信号に残響が加わった際に変調の谷が残響成分によって埋まるので、残響に対する耐性も弱くなる。帯域数を増やすことは帯域幅を狭くすることと同義であり、これは検出時の帯域群内の同期加算に対して有利に働くが、帯域幅を狭くすると高い変調周波数を用いることができない。よって、比較的低い変調周波数 (10Hz 以下) を用いる方が望ましい。

$H_{\text{high}}(t)$ のハイパスカットオフ周波数を hc Hz とすると、 hc が高いほど耐性は増すが、音質が劣化する。また、 hc を高くとったところで、MPEG4 AAC 符号化における SBR(Spectral Band Replication) 方式のような、高域信号を合成する符号化がステゴ信号に与えられると、高域信号に埋め込まれた情報は欠落する。知覚符号化やアナログ放送の際には、高域信号はステゴ信号の品質をあまり損なわずにカットされることがあるから、あまり hc を高くとる必要はない。

以上を考慮した上で、埋め込みパラメータを様々に変化させた埋め込みおよび検出シミュレーション実験を行った結果、埋め込みパラメータ値を表 3.1 のように定めた。複数変調周波数において同時に埋め込みを行う場合、過変調に陥らないため、変調周波数毎

に最大振幅変調度を定めておく必要があり，表 3.1 ではこれを 0.316 とした．本章では以降，この埋め込みパラメータ値を利用することとした．この場合，埋め込み時は，4096 点 FFT を用いた FIR フィルタによるフィルタバンクを利用した．検出時に用いる FFT 処理のサンプル数は 4096 点とし，1024 点づつずらしながら離散周波数 4 点分を 1 帯域としてその絶対値を求めることによって，帯域通過信号の振幅包絡とした．

表 3.1: 埋め込み時に用いるパラメータ値の例．

Parameters	Values
sampling freq.	44100 Hz
high-pass cutoff freq. (hc)	11025 Hz
bandwidth	43 Hz
subband pairs (n)	128
subband groups (k)	5
number of pairs per group	25 — 26
mod. frequencies (f)	2, 3, 5 Hz
frame period (T_p)	5 s
maximum AM depth per mod. freq.	0.316
total bit rate of hidden data	4.8 bps

3.4 振幅変調強度の決定方法

秘匿情報の埋め込み強度となる振幅変調度の決定方法は， m 番めの帯域ペア毎に，ホスト信号に対する検出演算後に検出される帯域変動包絡波形 $G_m(\tau) = \log \frac{S_{2m-1}(\tau)}{S_{2m}(\tau)}$ における，埋め込み変調周波数 f での変動振幅 $M_0(m)$ を基準として強度を設定する． $M_0(m)$ は，ホスト信号から得られる $G_m(\tau)$ に対して T_g 秒周期に，DFT を行って求める．この $M_0(m)$ と同じ振幅変動振幅を生じさせる，帯域ペアに与える変調強度 $x_0(m)$ を 0 dB とおく．ここで， $(2m-1)$ 番めの帯域の振幅包絡の直流成分 D が， $2m$ 番めの帯域のそれより a 倍大きいとし，変調周期 $(1/f)$ の間それらが一定だとおくと， $(2m-1)$ 番めの帯域の変調周波数 f における振幅包絡波形の谷は次式で表される．

$$D - M_0(m) = \log \frac{a(1 - x_0(m))}{1 + x_0(m)}, \quad (3.9)$$

ここで、 $D = \log a$ となる。また、振幅包絡波形における山は次式で表される。

$$D + M_0(m) = \log \frac{a(1 + x_0(m))}{1 - x_0(m)}. \quad (3.10)$$

これらを解いて、 $x_0(m)$ を $M_0(m)$ で表すと次式となる。

$$x_0(m) = (\exp(M_0(m)) - 1) / (\exp(M_0(m)) + 1). \quad (3.11)$$

よって、 m 番めの帯域ペアに対する変調度 $A(m)$ は、この $x_0(m)$ を基準値とした dB 値 $20 \log_{10} \frac{A(m)}{x_0(m)}$ として表される。つまり、埋め込み強度 0 dB のとき、式 (3.8) における、第 1 項の変調周波数 f における変動強度が、第 2 項における $AM_m(\tau)$ の変動強度と等しくなる。このような振幅変調強度の決定方法の模式図を、図 3.3 に示した。

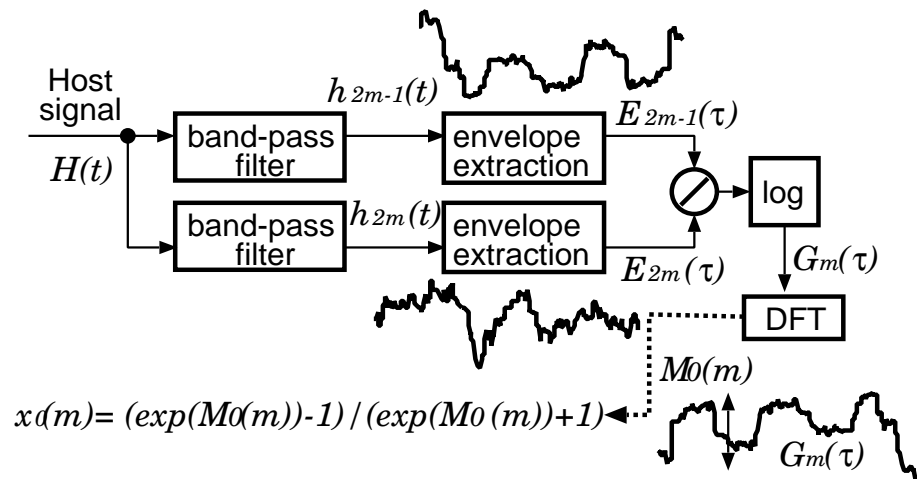


図 3.3: 変調強度決定のブロックダイアグラム。

ここまでの変調強度の決定法では、帯域ペアの双方に同じ変調度が与えられる。しかし、帯域ペアの一方に強い線スペクトル成分が生じた場合、この帯域信号への振幅変調は目立って知覚される。このため、帯域ペア間のパワーレベルの差を ΔL とおいたとき、 ΔL が 20 dB 以上の場合に、強い方の帯域には、 $1 - 0.2 \log(10^{\Delta L/20})$ を掛け合わせた変調強度を、弱い方の帯域には、 $1 + 0.2 \log(10^{\Delta L/20})$ を掛け合わせた変調強度を与えることとした。これによって、結果として検出される変動強度を同等に保ちつつ、強い帯域成分の変調度は抑え、変調度を増した弱い帯域成分は強い成分によってマスクされることによって、音質劣化を抑えることが可能となる。閾値の $\Delta L = 20\text{dB}$ は、強い帯域成分が弱い帯域成分を十分マスクするためのレベル差とした。

ホスト信号の変動量や帯域毎のパワーは時々刻々変わるので、これに追従して知覚的に過大な変調量を与えない必要がある。よって音質に関する試聴実験の結果、3 Hz 以下の

埋め込み変調周波数では2変調周期 ($T_g = 2/f$) 毎に, 3 Hz を上回る埋め込み変調周波数では, 4変調周期 ($T_g = 4/f$) 毎に, 上述の強度決定を行った. この T_g の設定値が最適であるかどうか, 今のところ理論的根拠は無いが, 今後変調マスキング検知における知覚的時間積分構造を調べることによって, T_g の妥当な設定値が得られることを期待する.

3.5 埋め込み区間同期検出方法

ステゴ信号の任意の時間区間から秘匿情報検出を行うためには, 埋め込み時のデータフレーム位置を検出する必要がある. そこで, 時間長 $T = \lfloor 4T_p/T_f \rfloor$ である矩形時間窓を, パイロット帯域グループから得られた変動波形 $G^1(\tau)$ に繰り返し与える. 時刻 u から始まる変動波形ベクトル \mathbf{R}_u は, 次式で与えられる.

$$\mathbf{R}_u = \{G^1(u), G^1(u+1), \dots, G^1(u+T-1)\} \quad (3.12)$$

\mathbf{R}_u の開始時刻 u を, データフレーム周期長 T まで変化させながら, フレーム周期長分だけずれた \mathbf{R}_u の差分を積算して $F(u)$ を求める. $F(u)$ における埋め込み変調周波数 f のパワー $\text{AMP}_f(F(u))$ は, u がちょうどデータフレームの開始位置と一致するときに, 最大となる (図 3.4 参照). これ以降のシミュレーション実験では, 8フレーム分の変動を累積して $F(u)$ を求めているので, $v_{\max} = 4$ としている. また, 式 (3.14) の y がデータフレームの開始時刻 (フレーム境界時刻) となる.

$$F(u) = \sum_{v=0}^{v_{\max}} \mathbf{R}_{u+2vT} - \sum_{v=0}^{v_{\max}} \mathbf{R}_{u+(2v+1)T} \quad (3.13)$$

$$y = \underset{u}{\operatorname{argmax}} \text{AMP}_f(F(u)) \quad (3.14)$$

3.6 埋め込みおよび検出シミュレーション実験

音響電子透かしの用途に情報秘匿技術を用いる場合, 様々な音響的特徴をもつホスト信号に埋め込んだ情報が, ステゴ信号に対する放送/伝送/再生や記録時に想定される様々な音質劣化を伴う変形を経た後でも, 有効に検出できる信頼性が第一に求められる. また, 数%程度の時間長変換やピッチ変換は, 人間に知覚されにくい一方で, 一般的な情報秘匿手法に対して秘匿位置の検出を困難とするため, 音質劣化を少なく秘匿情報を検出不能とするのに有効な攻撃として知られている.

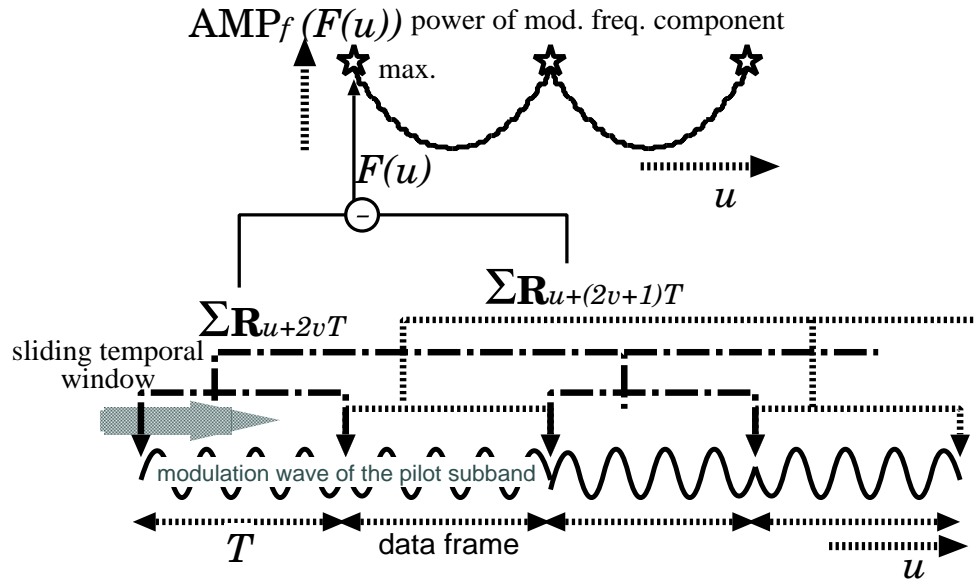


図 3.4: 埋め込み区間検出のための累積変動波形からの最大パワー検出。

以上をふまえ、本節ではまず本情報秘匿手法が、ステゴ信号に対する典型的な変形を経ても、有効に秘匿情報を検出できることを、RWC 音楽ジャンルデータベース [41] に収録された 100 曲を用いて、シミュレーション実験によって評価する。その後、第 3.7 節において音質劣化の検知と度合に関して明らかにしていく。実験では、表 3.1 で示した埋め込みパラメータ値を用い、埋め込み強度として 0 dB, -5 dB を採用した。楽曲は冒頭 1 分間の左チャンネルのみ (サンプリング周波数 44.1 kHz, 量子化ビット数 16 bit) を用いた。

3.6.1 ステゴ信号に対する変形

楽曲に著作権管理情報を秘匿した場合、その楽曲を知覚符号化により情報圧縮した後でも、秘匿情報は確実に検出される必要がある。そこで、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) および MP3 (64, 48, 32 kbps/ch) を用いた。RealAudio8 のビットレート 21 kbps にはサンプリング周波数 22.05kHz への変換が、MP3 のビットレート 48 kbps, 32 kbps にも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。RealAudio8 へのエンコードには、RealProducer Plus 11.1 (Linux 版) を用いて、いずれのビットレートでも Music オプションを用いた。MP3 へのエンコードには“午後この〜だ” ver.2.39 (Linux 版) を用いた。

ステゴ音響信号が室内にスピーカによって放射され、これをマイクロホンで收音する際には、壁面や床、天井などによる反射や残響が付加される。よって、反射音や残響を付加

するシミュレーションとして、振幅1のインパルスの直後に指数減衰を与えた正規分布乱数によって生成したインパルス応答により、0.25, 0.5, 1, 1.5 秒の残響を与えることとし、このインパルス応答をステゴ信号に畳み込んだ。それら4条件のインパルス応答波形を、図3.5に示した。

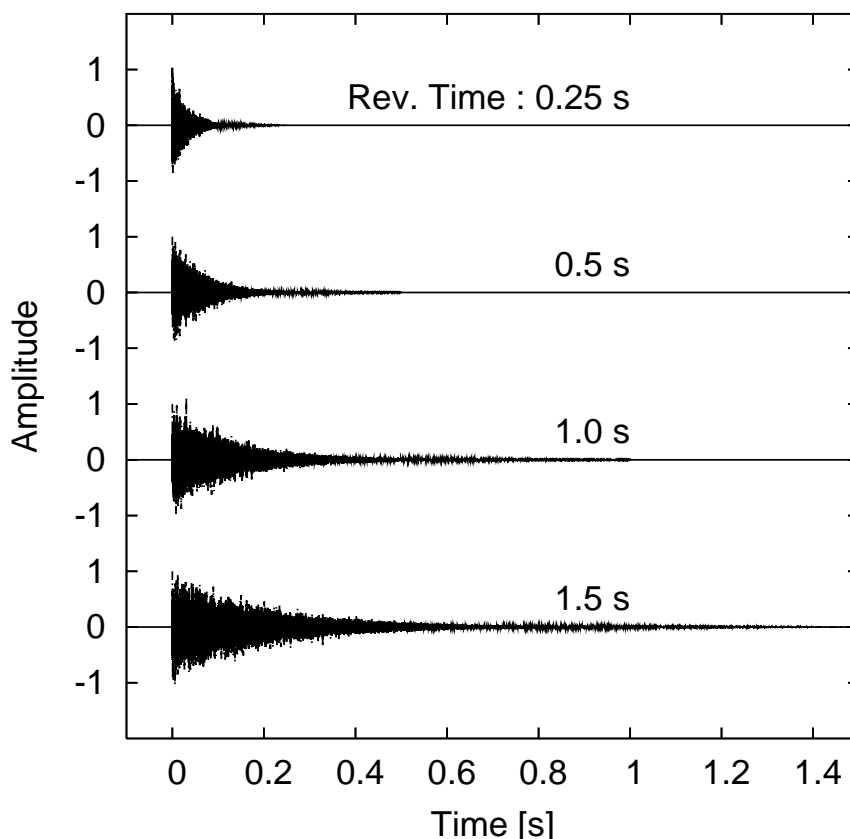


図 3.5: 耐性シミュレーション実験で用いた残響付加のためのインパルス応答波形。

室内環境においてスピーカ再生されたステゴ信号をマイクロホン受信したり、ステゴ信号をアナログラジオ放送として送信し受信する場合を考えると、ステゴ信号に雑音を重ねられる場合も考えられる。そこで、典型的な雑音としてホワイトノイズを付加することとし、その強さは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。

また、悪意のある利用者は、著作権管理情報が秘匿されている楽曲から秘匿情報を検出できなくなるような処理を施した上で、違法コピーを行う可能性が考えられる。つまり、ステゴ信号には一定時間区間(フレーム)毎にデータが埋め込まれているのが一般的なので、情報検出時にこのフレーム区間を検出しにくくなるよう、時間長を伸縮する攻撃が考えられる。このような、時間長伸縮として、ここでは検出時に埋め込み時より変調周波数

を低くかつフレーム時間長を長く設定することにより，時間短縮された信号に検出処理を行うとみなした．時間伸長については，その逆である．こうして， -4% ~ $+4\%$ の時間長伸縮を模擬した．なお，この時間伸縮変換ではピッチの変換を伴わない時間長変換を模擬している．

ピッチ変換も時間長変換と同様に，周波数軸に沿って規則的に情報を埋め込む手法に対して，その規則性を検出時にずらせることにより，秘匿情報検出を困難にする効果がある．ピッチ変換は，検出時に帯域信号の振幅包絡を求めるために行う 4096 点 FFT 処理において，そのサイズを 4072 点から 8 点づつ 4120 点まで変えて DFT 処理を行うことにより， $+0.6\%$ ~ -0.6% のピッチ変換が与えられた信号に対して検出を行うものとみなした．なお，この処理によって信号の時間長は変わらない．

実際のステゴ音響信号に対して時間伸縮処理やピッチ変換処理を行わなかった理由は，シミュレーション時の演算負荷の軽減と，どのような時間伸縮やピッチ変換アルゴリズムを用いるかに結果が依存しないようにするためである．

3.6.2 実験結果

実験結果は，表 3.1 のパラメータ値に基づき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．検出時のエラーにより誤ったビット値をそのまま出力としては使用できない．そして，あらゆる条件において検出率 100% を保証することは困難なので，実用上は誤り訂正符号を用いることになる．音響電子透かしに求められている典型的な情報埋め込み量は，15 秒あたり 2~ 12bit のコピー制限情報に加えて 30 秒あたり 60~ 72bit 程度のコンテンツ ID 情報 [42, 43] である．ここで仮に，15 秒あたり符号長 63bit，情報ビット長 36bit，訂正限界 5bit の BCH 符号化 [44] を主として用いると仮定すると，このシミュレーション実験での秘匿情報量のビットレート 4.8 bps は，この要求をほぼカバーする．また，検出時には位相差がビット割り当て時の位相差 ($0, \pi/2, \pi, 3\pi/2$) からどれだけ離れているか，によってビット信頼度が得られることから，軟判定復号法 [45] を併用することにより，訂正限界が 2 倍程度向上すると見込まれる．よって，85% の検出率を一応の目安として評価する．なお，本章の実験ではエラー訂正符号は用いていない．

図 3.6 には，MP3 符号化と復号化に対する，埋め込み強度 0 dB と -5 dB での検出率について，全 100 曲に対する最小値と最大値を誤差棒によって，10 パーセンタイル点を

、0 dB の中央値を、-5 dB の中央値を、90 パーセンタイル値を で表した。図 3.7 には、RealAudio8 の結果を示した。この結果から、RealAudio8 や MP3 といった知覚符号化を用いた場合は、ネットラジオ等で用いられる比較的低いビットレートである 21 kbps/ch でも、0 dB の強度で振幅変調を与えれば、あらゆるジャンルの楽曲に対して十分な検出率を保つことが示された。

図 3.8 は、ホワイトノイズ付加に対する結果である。埋め込み強度 0 dB では、SNR 30 dB まで、埋め込み強度 -5 dB では SNR 40 dB において十分な検出力を持つことが分かった。

図 3.9 は、残響に対する結果である。埋め込み強度 0 dB の場合、コンサートホールなどで得られる 1.5 秒といった、比較的長い残響にも強い耐性を示すことが明らかになった。現在の埋め込みアルゴリズムでは変調強度決定に 1 秒程度の時間窓 T_g を必要とするので、実時間での情報秘匿は想定されていない。しかし、変調強度を一定値とすれば、実時間の埋め込み処理も可能であり、ライブコンサート等において拡声される直前の音声信号へ情報秘匿を行うことも可能であろう。

図 3.10 は、時間長伸縮に対する検出率を示している。0 dB の強度で埋め込んだ場合、 $\pm 3\%$ 程度までは 90% 以上の楽曲に対して耐性を保つが、それ以上の変換では急激に検出率が劣化した。この理由として、フレーム間の変調位相差に依存してフレーム検出を行っているため、時間長伸縮によってフレーム長が変わるにつれフレーム間の変調位相差がずれていき、フレーム同期ができなくなることが原因である。これを改善するためには、検出時に $\pm 4\%$ 程度の時間長伸縮を前提としたフレーム同期演算も同時に実施し、図 3.4 に示した累積変動波形から得られる変動パワーにおける最大値と最小値の比を、フレーム同期強度とみなして、時間長伸縮量を予測した上で透かし検出を行う手法が考えられる。これによって、 $\pm 7\%$ 程度の時間長伸縮に対する耐性が確保できると考えられるが、フレーム検出のための演算量は 3 倍になる。

図 3.11 はピッチ変換に対する検出率を示している。ピッチ変換に対して、0 dB の埋め込み強度では、 $\pm 0.4\%$ を越えると急激に検出率が劣化した。この理由は、埋め込み時の帯域フィルタの周波数と、検出時の帯域フィルタの周波数がずれることによって、帯域ペアの一方の変調波がもう一方に洩れ、検出後の変調が打ち消されてしまうことが原因である。

DA 変換後の信号を AD 変換して得られるいわゆるアナログコピーにおいて、原信号とコピー後の信号の違いは、おもに DA 変換器と AD 変換器のサンプリング周波数の偏差

に起因する [46] . デジタルオーディオ機器のサンプリング周波数には , 機器毎に最大で 0.1%未満の偏差が見られるが , この偏差に伴うピッチ変化程度に対しては , 本情報秘匿手法が有効であることが分かった .

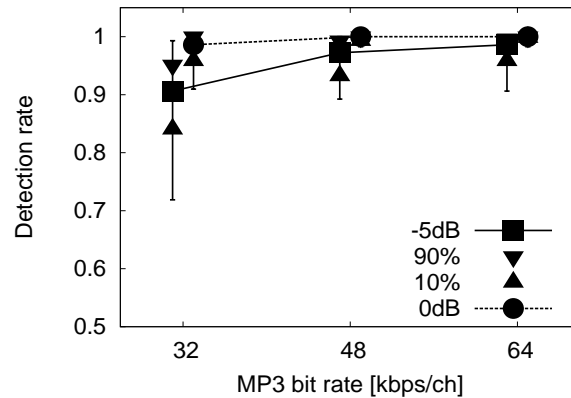


図 3.6: MP3 符号化に対するビット検出率 . は埋め込み強度 -5 dB の中央値 , は埋め込み強度 0 dB の中央値 , は 90 パーセンタイル点 , は 10 パーセンタイル点 , エラーバーは 100 曲中の最小と最大検出率を示している .

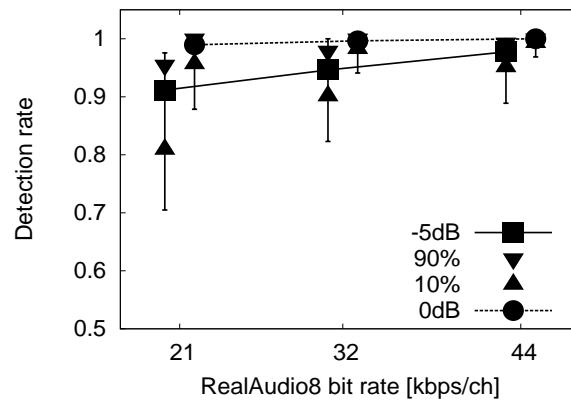


図 3.7: RealAudio8 符号化に対するビット検出率 .

3.7 情報秘匿に伴う音質劣化の主観評価

前節では , 埋め込み強度をホスト信号の振幅変動量に比例して逐一設定する条件で , ステゴ信号への様々な変形に対する耐性を調べ , 情報秘匿方法の信頼性を確認した . 本節では , 情報秘匿時の埋め込み強度の知覚閾値を調べた上 , 閾値以上の埋め込み強度を用いた場合の音質劣化の度合を , 主観評価実験によって調べる .

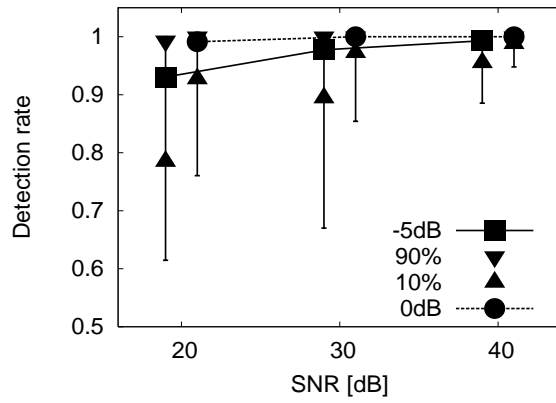


図 3.8: ホワイトノイズ付加に対するビット検出率 .

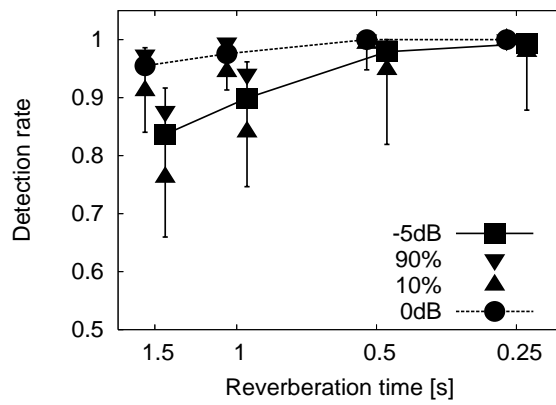


図 3.9: 残響付加に対するビット検出率 .

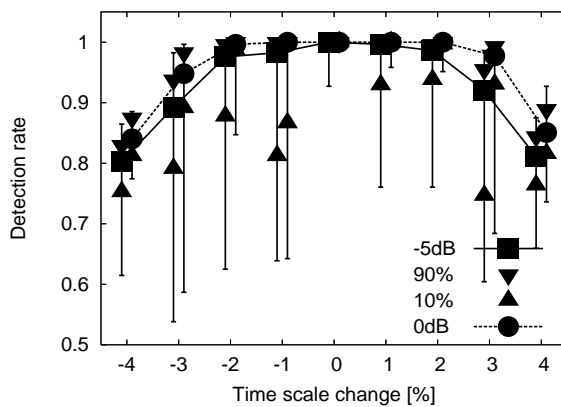


図 3.10: 時間長伸縮に対するビット検出率 .

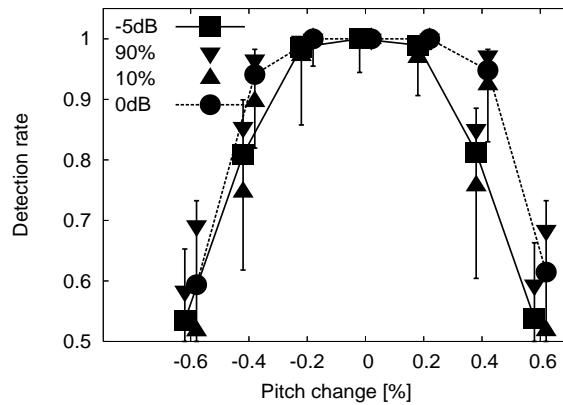


図 3.11: ピッチ変換に対するビット検出率

まず予備実験として、未経験な被験者を用いて、第 3.7.2 節に示す二重盲検法による音質劣化評価実験を行った。対象は、RWC-MDB-G-2001 より 10 dB の埋め込み強度の楽曲が十分検知可能であった、5(バラード)、17(ハウス)、66(ブルース)、84 (フラメンコ)、86(シャンソン) の 5 曲について 3 名、No.6(バラード) と 31(モダンジャズ) の 2 曲についてはさらに 2 名追加して 5 名、それぞれの曲に、0 dB あるいは 10 dB の埋め込み強度によってランダムなデータを埋め込んだステゴ信号を、ホスト信号と識別できるかどうかを調べた。その結果、0 dB の埋め込み強度の場合に、ホスト信号と区別できたケースはほとんどなかった。

よって、より入念に著者が RWC-MDB-G-2001 の 100 曲より試聴実験を繰り返し、0 dB の埋め込み強度のステゴ信号がホスト信号と識別可能な No.45(レゲエ)、69 (ギターとボーカル)、87(ピアノ) を選定し、以降の実験で用いることとした。

3.7.1 AXB 法による音質劣化検知閾測定

これらの 3 曲の冒頭 5 秒間の左信号について、2-down,1-up の UDTR 法により埋め込み強度を可変させ、AXB 法を用いてヘッドホン (STAX Lambda-Nova Classic) ダイオティック聴取による 70.7%検知閾を調べた。実験は遮音された簡易無響室において実施した。

この実験は、一度の判断につき、3 回信号が呈示される。まんなかの信号 (X) は、情報秘匿済みあるいは原信号のいずれかであり、最初 (A) と最後 (B) は、情報秘匿済みと原信号がいずれかの順序で組み合わせられている。被験者は、X の音と同じ音が A だったのか B だったのかを答える。最初は、十分検知できる埋め込み強度で情報秘匿された音楽が呈示され、正解が 2 度続くと 4 dB 埋め込み強度が減ぜられる。1 度間違うと 4 dB 上昇す

る．4回目の判断転換点以降は，埋め込み強度の変化は2 dB となる．5回目から12回目までの8回の判断転換点の埋め込み強度(dB)を算術平均して，1回の閾値が求まる．No. 45, 69, 87の呈示実効音圧レベルは，それぞれ78 dB, 76 dB, 71 dBであった．なお実験および訓練時には，被験者の判断毎に正答が被験者の実験画面に表示されるようにした．

被験者は4時間以上この実験手法による訓練を行った20歳から22歳の学生4名である．事前に被験者の聴力レベルは10 dB 以下であることを確認している．

4回以上の閾値測定結果の平均値と ± 1 標準偏差を，被験者毎に図3.12に示した．この結果から，振幅変調に基づく情報秘匿に伴う音質劣化に対して，最も検知能力が高い被験者かつ最も検知しやすい曲における検知閾は， -10 dB 程度だと考えられる．

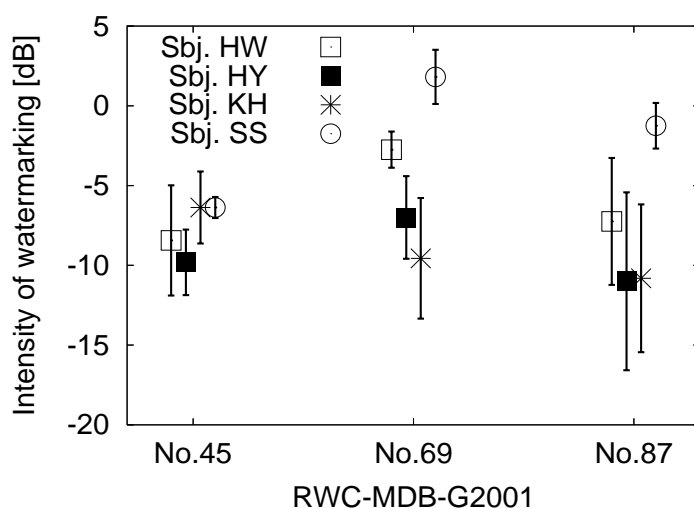


図 3.12: 埋め込み強度の検知閾．エラーバーは ± 1 標準偏差を示している．

3.7.2 隠れ基準付き二重盲検法による音質劣化評価

第3.6節では，RWC-MDB-G-2001の100曲を対象に，ステゴ信号に対する様々な変形後に秘匿情報の検出実験を行った．その結果，0 dBの強度で情報秘匿すれば，ピッチ変換および時間長変換を除き，知覚符号化(MP3 / 32kbps, RealAudio8 / 21 kbps)や，ノイズ付加(SNR 20 dB)，残響付加(1.5 s)を経ても90%以上の楽曲で90%以上のビット検出率を達成できた．埋め込み強度が -5 dBの時に，同程度の検出率は，知覚符号化の場合44 kbps/ch，SNRは40 dB，残響時間は0.5 s程度で達成できた．

また，前節のAXB法による埋め込み強度の検知閾値測定から，音質劣化を検知しやすい楽曲については，上記の埋め込み強度 -5 dBにおいて，訓練された聴取者は音質変化を

検知できることが分かった。ここでは、それらの埋め込み強度における音質劣化の度合を調べるために、極めてわずかな音質劣化を検知し評価する手法を定めた ITU-R BS.1116-1 に準拠する、隠れ基準付き二重盲検法を用いて評価実験を行った。

BS.1116-1 で推奨される評価実験に関わる推奨項目は、第 2.3 節でも述べたように多岐に渡るが、今回の実験条件との相違は、スタジオ等の音響技術者に準じる経験をもつ被験者を 20 名以上供することが推奨されるのに対し、大学生の被験者を 5 名とした点である。その理由は、一定の音質劣化識別能力(ここでは、128 kbps MP3 符号化と復号化を経た音楽信号を、原音楽信号と識別できること)を持つ被験者が、候補者の半分しかおらず、被験者数を増やすことができなかったためである。もう一点、実験素材として推奨される EBU(ヨーロッパ放送連盟)制作 SQAM (Sound Quality Assessment Material) を用いずに、RWC-MDB-G-2001 より選曲した点である。この理由は、前節における埋め込み強度の検知閾値測定、および第 3.6 節における耐性シミュレーション実験との対応をとるためである。

BS.1116-1 における特徴的な音質劣化評価手法を以下に説明する。コンピュータ画面上で、音質劣化した信号と劣化の無い基準信号のペアが、実験者と被験者に分からないようにラベル付けされている。被験者はそれらと、明示されている基準信号の 3 種を、再生中の任意の時刻で自由に切替えて繰り返し聴取し、ペアのうち音質劣化が感じられる方に、5(違いが知覚できない)~ 4(違いが分かるが気にならない)~ 3(わずかに気になる)~ 2(気になる)~ 1(とても気になる)の間で 0.1 刻みの評価を与える。この実験に用いたユーザインタフェースを図 3.13 に示した。

評価対象は AXB 法で用いた 3 曲について、透かし強度 0 dB, -5 dB の 2 条件とした。また、第 3.6 節におけるシミュレーション実験において、秘匿情報の検出率が最も低かった楽曲からも 1 曲 (No.99: 雅楽) 選び、埋め込み強度は 0 dB, +10 dB の 2 条件とした。この埋め込み強度設定は、事前の実験により、埋め込み強度 -5 dB では検知が困難だったからである。さらに、各 4 曲について情報秘匿による音質劣化度合のアンカーとして、情報秘匿を行わずに MP3 で符号化 (128, 96kbps) 後、復号化した 2 条件も評価対象とした。対象区間は、冒頭 1 分間のステレオ信号とし、オーディオユニット (M-AUDIO Delta 1010) の出力をアンプ (DENON AVC-1890) を通してヘッドホン受聴 (STAX Lambda Nova Classic) した。前節の AXB 法による実験に参加した被験者 4 名と、事前の試験により音質劣化識別能力があると認められた被験者 1 名の合計 5 名が実験に参加した。各 4 曲毎の 4 条件での音質評価を 1 セッションとし、曲順をランダムに 5~ 6 セッションの評価を、1 日あた

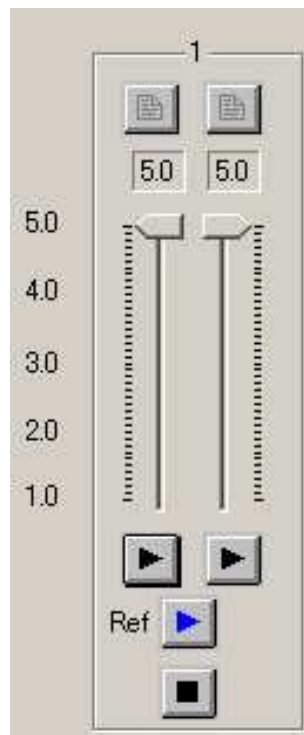


図 3.13: 音質劣化評価実験に用いたユーザインタフェース。Ref はホスト信号を再生するボタンであり、その上の左右どちらかの再生ボタンに音質劣化信号とホスト信号が割り当てられている。被験者は音質が劣化していると感じられる方のスライダーを調整して音質劣化の評価を与える。

り1セッションで、日を分けて繰り返して行った。再生音圧レベルは、1分間の実効値で、No. 45, 69, 87, 99 に対してそれぞれ 79 dB, 76 dB, 77 dB, 76 dB であった。

結果として、音質劣化音に与えられた評価から基準音に与えられた評価を差し引いて主観差分値 SDG を算出した。被験者が正しく劣化音を聞き分けて評価を与えれば、SDG は負の値となる。MP3 128 kbps 条件について、各被験者毎に t 検定を行った結果、SDG の値は片側検定において危険率 5%未満で有意に 0 より小さいことが分かった。つまり、全被験者が基準音より MP3 128 kbps 条件を劣化した音質であると識別でき、音質劣化評価能力を持っていることが確認できた。なお、音響技術者など音質劣化に敏感であるとされる聴取者が行う一般的な音楽に対する MP3 128 kbps 符号化の音質劣化度合の評価は、違いは分かるが気にならない程度か、それより良い程度である [47, 48]。

図 3.14 に、SDG 値を被験者間で平均して示した。エラーバーは被験者平均値間での±標準偏差の値を示している。同じ劣化条件を比較できる、No. 45, 69, 87 の3曲に関して、楽曲と劣化条件を要因とし、被験者を繰り返すとみなす 2 元配置分散分析を行った。その結果、劣化要因は高度に有意であった ($F(3, 48) = 11.28$) ($p < 0.001$)。また、楽曲の要因 ($F(2, 48) = 0.47$) および、楽曲と劣化条件の間の交互作用 ($F(6, 48) = 0.53$) は、共に有意でなかった。ボンフェローニの多重比較検定を行ったところ、全ての劣化条件の組合せに危険率 5%未満で有意差が見られた。SDG の値は、小さいほうから -5 dB 強度の情報秘匿、MP3 128 kbps、0 dB 強度の情報秘匿、MP3 96 kbps の順であり、この順に劣化が大きくなることが分かった。また、-5 dB の強度で情報秘匿を行った場合と、MP3 128 kbps の場合は、原音との違いは分かるが、気にならない程度の劣化であることも分かった。

さらに、秘匿情報検出率が低かった No. 99 では、0 dB 強度の埋め込みでも、他の曲における -5 dB 強度の埋め込みに相当する程度の音質劣化であることが分かった。よって、No. 99 に関しては、音質を保ちつつ検出率を高めるために、埋め込み強度を高めた最適な設定の余地が残されていることが示唆された。

3.8 音質劣化の主観評価と客観評価との対応

本節では、聴覚興奮パターンのモデル化を取り入れた客観音質劣化評価を行い、主観評価結果との対応が見られるかを報告する。MP3 などの知覚符号化信号の音質を客観評価する手法として、ITU-R BS.1387 に定められた PEAQ [49, 47, 19] が広く使われている。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

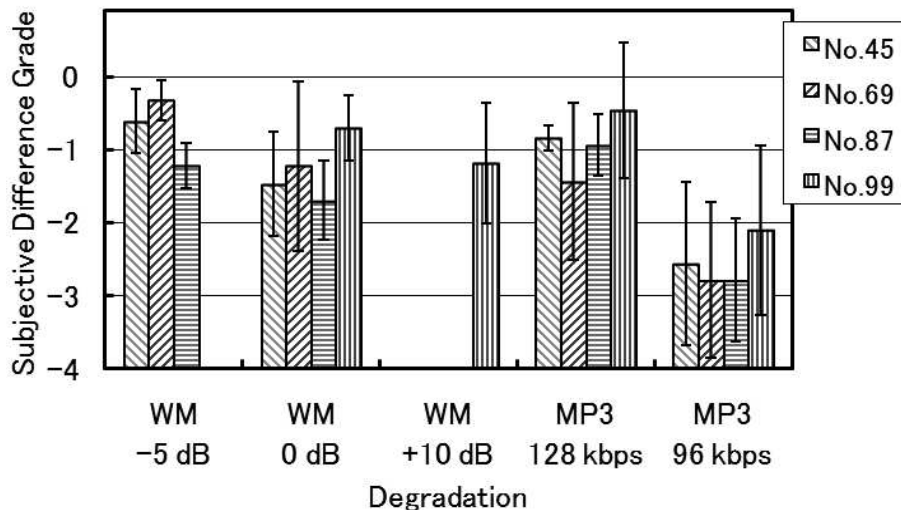


図 3.14: ITU-R BS.1116-1 に準拠した実験により得られた主観差分等級 (Subjective Difference Grade) .

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 (MOV) を計算し、ニューラルネットを用いて主観劣化評価結果 (SDG) とよく合うような ODG を算出する MOV に対する重み付けをあらかじめ行っておいて、音質劣化度合を予測する手法である。

ここでは、PEAQ の基本バージョンの実装 [14] を用いて、データ秘匿済み音楽信号の音質劣化度合を測定した。得られる 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。

図 3.15 には、縦軸を被験者 5 名の平均 SDG 値、横軸を ODG 値として、第 3.7 節の実験で得られた MP3 への評価 () と、情報秘匿音楽への評価 () を示した。これらの SDG と ODG との相関を求め片側検定を行ったところ、MP3 の場合は、 $R = 0.779 (p = 0.011)$ 、情報秘匿の場合は、 $R = 0.627 (p = 0.048)$ で共に有意であった。また、SDG と ODG の間の回帰直線を求めたところ、MP3 の場合は、 $SDG = 1.25 \times ODG + 0.059$ となり、情報秘匿については、 $SDG = 0.51 \times ODG - 0.57$ となった。これらの回帰直線は、図 3.15 に同時に記載した。MP3 の場合は、知覚符号化による音質劣化を算出する PEAQ の役割どおり、客観評価値が主観評価値と対応して相関も高くなることが分かった。振幅変調に基づく情報秘匿に起因する音質劣化に関しても、被験者間の平均主観評価と客観評価はある程度対応することが分かった。よって PEAQ アルゴリズムは、振幅変調に基づく情報秘匿に起因する音質劣化をある程度予測することが可能であるとみなし、本章以降では、客観音質劣化評価の指標として PEAQ の算出する ODG 値を用いることとする。

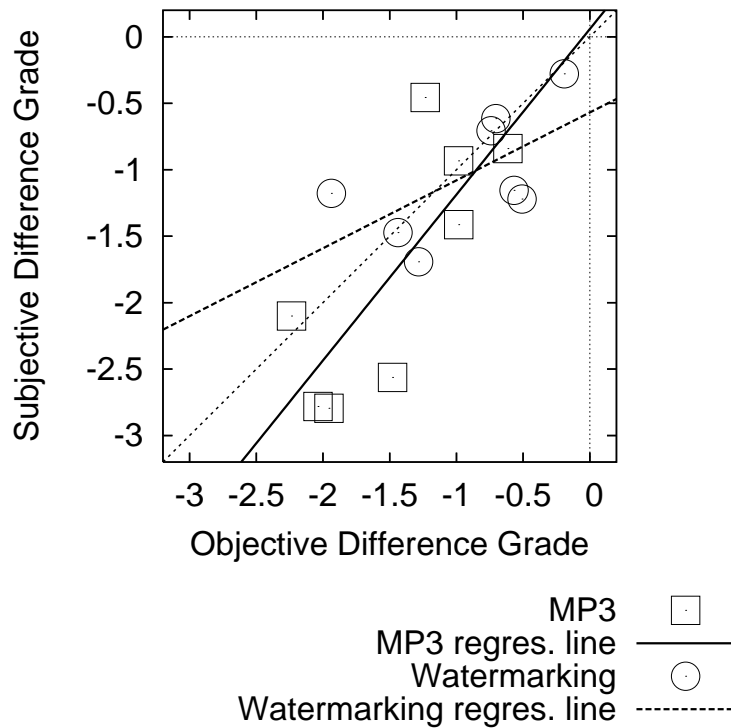


図 3.15: 平均 SDG と ODG の対応。□ は振幅変調に基づく情報秘匿に起因する音質劣化，○ は MP3 符号化による音質劣化である。

主観評価実験参加者や対象曲を増やせば，得られた SDG 値に対して，PEAQ が用いる聴覚モデルの各段階での出力 (MOV 値) 段のニューラルネットを再学習することにより，客観評価指標である ODG 値と主観劣化評価である SDG 値とのより良い対応が得られると思われる。

3.9 考察

3.9.1 知覚検知閾に基づく変調強度設定

第 3.4 節では，あらかじめホスト信号に存在する振幅変動成分を算出して，その強度を基準に埋め込み強度を定める方法を提案した。そして，第 3.6 節では，様々なジャンルの音楽信号に対して，十分な検出性能を確保しつつ情報秘匿が可能であることを示した。しかし，一部の検出率が低い楽曲 (No. 99) については，音質劣化が目立ちにくいいため，より埋め込み強度を高めて検出率を高めることが可能であるが，これを自動的に行うことまではできなかった。従来の音響電子透かし技術には，音楽信号に対して音質劣化の検知閾に対応した埋め込み強度を自動的に決定するために，MPEG 心理音響モデルを適用して，

マスク閾値以下の強度の透かし信号を加算する仕組みをもつものが多い[24, 28, 26]。しかし、本手法のような、比較的ゆっくりした変調を与えることによって生じる歪成分は変調周波数だけ離れた側波となるうえ、100 ms 以下の時間窓によって得られたパワースペクトルを基本とする MPEG 心理音響モデルは、そのまま利用できない。

より最適に、音質劣化と検出性能の両方のバランスをとる埋め込み強度の決定を行うには、変調検知閾値を予測することのできる聴覚モデルを利用することが望まれる。Dau らは、狭帯域雑音(変調マスク)に対して新たな変調(変調ターゲット)を与えた場合のターゲット変調検知閾値を、様々な変調マスク周波数、帯域幅(変調周波数)とターゲット変調周波数に対して調べている[50]。さらに、Dau らは、自らの実験のデータ[50]と過去の振幅変調検知の実験[51, 52]データを説明する、変調フィルタバンクモデルを構築した[53]。このモデルでは、末梢の聴覚フィルタ出力に半波整流と低域通過フィルタを施して振幅エンベロープを検出し、音の立上りを強調して持続部を抑制するような聴神経の発火頻度パターンに似た順応回路を通した後、変調周波数毎に用意されるフィルタバンクを通過した出力に内的雑音を加えることによって、内的変調信号表現を得る。この内的表現は、聴覚フィルタバンク中心周波数、変調フィルタバンク中心周波数、時間、エンベロープ振幅の4つの次元をもつ。この4次元空間において、心理実験で得られる変調検知閾値を模擬するような、新たな変調が加わった区間と、そうでない区間との相互相関の閾値を設定して、変調を検知するモデルである。

Dau らの変調検知モデル[53]を本手法の振幅変調に適用すると、帯域毎に得られる音楽信号の変調パワースペクトルと変調フィルタバンクの形状から振幅検知閾値を計算することができるので、変調検知限を基準とした変調度の設定が可能になる。しかし Dau らも指摘するように、このモデルは他帯域の振幅変動が目的帯域の振幅変調の検知を妨害するという帯域間の変調検知干渉(MDI, Modulation Detection Interference)[54]を、うまく説明することができない。各周波数帯域の変調スペクトルを統合するような変調フィルタバンクを導入すると説明できる可能性はあるが、直流に近い低域の変調フィルタバンクに関しては、周波数領域でのマスキングに影響するため、独立した帯域毎に処理する必要がある。

事前の試験的な聴取実験から経験的に言えば、広帯域に楽器音が同時に存在する、例えばロックやポップスといった音楽では、振幅変調の検知閾はより高くなり、0 dB の強度で情報を埋め込んだ場合にも、検知は非常に困難であった。これは、振幅変調されている弱い帯域成分が周波数マスキングによって聞こえにくくなっている事実に加えて、多くの

帯域に振幅変動が観測されるため、特定の帯域に与えた振幅変調が検知されにくくなるという、変調検知干渉が生じているためと考えられる。

一方、帯域間で共通する振幅変動が、信号のマスキング閾値を減少させる、という共変調マスキング解除 (CMR, Comodulation Masking Release) という現象も報告されている [55, 56, 57]。現実の音楽信号において、広い帯域で同期した変動が生じることは音の立上り部分を除いては稀であるが、本手法において CMR がどのような影響を与えるか、についても検討の余地はある。

これまでの振幅変調検知に関する一連の研究で用いられたような雑音や純音を変調する条件に比べて、現実の音楽信号は非常に複雑である。このため、MDI や CMR を取り入れた変調検知モデルを構築すること、またそれを用いて、本手法が音響信号に与える埋め込み強度の最適化、つまり検知閾ぎりぎりの変調強度を音響信号に合わせて自動的に求めることは、今後の課題である。

3.9.2 電子透かしとしての耐性向上へ向けて

電子透かし用途として本情報秘匿技術を用いる場合、悪意ある攻撃として著作権管理情報の検出を妨げるために行われる、ステゴ信号へのピッチ変換に対する耐性が不十分であることが、第 3.6 節において示された。この理由は、埋め込み時の帯域フィルタの周波数と、検出時の帯域フィルタの周波数がずれることによって、帯域ペアの一方の変調波がもう一方に洩れ、検出時に変調が打ち消されてしまうことが原因である。この埋め込み時と検出時と帯域のずれは、高い周波数になるほど顕著となる。

演算処理量を増加させずにこれを改善する方法として、埋め込みおよび検出時のフィルタバンクを等帯域幅とせず、高い周波数になるに従い対数的に帯域幅を広げる方法が挙げられる。この方法を試みると、ピッチ変換への耐性は高まるが、同じグループとして同期加算する帯域数が少なくなることから、ピッチ変換以外への耐性が劣化してしまうことが分かった。よって、ピッチ変換によって高い周波数域から振幅変調が検出できなくなる特性を利用してピッチ変換量を推定し、補正して検出処理を行う手法を検討すべきであろう。ピッチ変換や時間長変換は、知覚符号化や雑音付加などの変形とは異なり、与えた振幅変調の時間—周波数平面上での位置がずれているだけとも言えるので、このズレを効率よく検出する手法を検討すれば、検出率を向上させることが可能だと考えられる。

さらに、電子透かし用途としての耐性を高めるためには、ホスト信号とステゴ信号の波

形あるいはスペクトルなどなんらかの差分をとって分析し、埋め込み情報の無効化を図る差分攻撃、同じ埋め込み手法で別のデータを埋め込む上書き攻撃に対する耐性も検討する必要がある。

本手法の場合、ホスト信号とステゴ信号の波形差分をとると、フレーム周期毎に変調位相が反転している特徴からパイロット帯域を検出することが可能で、パイロット帯域の周波数軸上での位置、という埋め込み時の秘密鍵情報の一部が明らかになってしまう。そして、同じ秘密鍵を用いて情報を埋め込んだ別のステゴ信号のパイロット帯域に対して、その変調周波数での逆位相の変調を与えると、フレーム同期が不可能となってしまう、秘匿情報の検出が困難となる。よって、現時点では差分攻撃に対する耐性は低いと言える。これは、聴感上問題の無い程度に位相変調を帯域毎に与えることによって、差分波形から振幅変調を欺瞞することが可能と考えられるが、どの程度の位相変調を与えるかに関しては、検討が必要である。

上書き攻撃に関しては、埋め込み時の秘密鍵情報がもたらす埋め込み条件と、どこまで同じ条件（例えば、変調周波数やパイロット帯域の周波数配置、データフレーム時間長など）で上書きが行われるか、によって耐性は異なるといえる。最悪の条件のひとつである、変調周波数が同一でデータフレーム時間長が同じだが、データフレーム区間は同期していない上書きを行った場合、 -5 dB 強度で埋め込んだステゴ信号に対して同じ強度で埋め込む場合、および 0 dB 強度で埋め込んだステゴ信号に対して $+5$ dB の強度での埋め込む場合までは、耐性を保つことが分かっている。詳細な条件での検討と、さらなる上書き攻撃への耐性向上は、今後の課題である。

3.10 あとがき

隣接する帯域信号ペアに与える逆相の振幅変調を秘匿情報のキャリアとする新しい音響情報秘匿手法を開発した。この手法の特徴は、すでにホスト信号に存在する振幅変動成分によって生じる変調マスキングによって、キャリアとなる振幅変調成分が聴感上目立ちにくい、という点である。また、ホスト信号に含まれる振幅変動成分の強さを元にして変調強度を設定することで、様々なジャンルの音楽信号に適用できることを目指した。音響電子透かし用途を前提として、様々な種類の音楽信号に対応できることを確認するため、RWC 音楽ジャンルデータベースの 100 曲を用いて、情報埋め込みおよび検出シミュレーション実験を行った。

その結果，4.8 kbps の情報を 0 dB の強度で埋め込んだ場合，ステゴ信号に対する 21~32 kbps/ch の知覚符号化と復号化，残響時間 1.5 秒，SNR 20 dB のノイズ付加， $\pm 2\%$ の時間長変換などの変形に対して，90%以上の楽曲で 90%以上の情報検出が可能であった． -5 dB の強度で情報を埋め込んだ場合，32~48 kbps/ch の知覚符号化と復号化，残響時間 0.5 秒，SNR 40 dB のノイズ付加， $\pm 1\%$ の時間長変換に対して，90%以上の楽曲で 90%以上の情報検出が可能であった．一方で，この情報秘匿手法はピッチ変換に対して脆弱であり，ピッチ変換量を推定して補正後に検出するような手法の改善が望まれる．

また，RWC 音楽ジャンルデータベースの中から，情報秘匿に伴う音質変化を検知しやすい楽曲を著者が選別し，情報秘匿に伴う音質変化の検知限と，検知限以上の変調強度を与えた場合の，音質劣化度合を調べた．音質変化検知訓練を積んだ被験者 4 名に対して，音質変化の検知限となる埋め込み強度は， -10 dB 程度であった．MP3 128 kbps で符号化し復号化した音楽信号の音質劣化を検知できる被験者 5 名を対象として得られた音質劣化度合は， -5 dB の埋め込み，MP3 128 kbps，0 dB の埋め込み，MP3 96 kbps の順に音質劣化度合が増すことが分かった．

さらに，情報秘匿に伴う音質劣化の主観評価と対応する客観評価の可能性を調べるため，知覚符号化信号の音質劣化を客観評価する PEAQ を取り上げ，PEAQ の出力する ODG 値と，主観評価の結果である SDG 値との対応を調べた．その結果，SDG と ODG には有意な相関が認められ，SDG 値に個人差はあるものの，平均すると SDG 値は ODG 値より大きくはずれることは無かった．このため，今後は PEAQ を振幅変調に基づく情報秘匿に起因する音質劣化の客観評価法として用いることとした．

今後の課題としては，楽曲によって検出力を保ちつつ音質劣化を最小に留める最適な埋め込み強度の自動設定のために，帯域間変調検知聴覚モデルに基づく埋め込み強度決定法の併用が望まれる．また，電子透かし用途としてのさらなる耐性の向上のためには，差分攻撃への耐性向上のための手法を検討することと，ピッチ変換や時間長変換に対して，それらの変換量を適切に推定して検出を行う手法を検討する必要がある．