

**振幅変調に基づく音響信号への
情報秘匿とその応用**

Audio data hiding based on
amplitude modulation and its
application

2010年10月

西村 明

Akira Nishimura

目次

第1章	序論	1
1.1	研究の背景	1
1.1.1	電子透かし	1
1.1.2	ステガノグラフィ	2
1.2	技術とその課題	3
1.3	本論文の目的と構成	4
第2章	音響信号への情報秘匿技術	6
2.1	まえがき	6
2.2	音響信号への情報秘匿技術の概要	6
2.3	技術の評価方法	8
2.4	既存の音響信号への情報秘匿手法	11
2.4.1	デジタル領域での情報秘匿手法	11
2.4.2	アナログ耐性のある情報秘匿手法	12
2.5	情報秘匿の補助技術	21
2.5.1	聴覚特性を考慮した強度設定	21
2.5.2	変形/攻撃を前提とした埋め込み強度設定	21
2.6	実用化段階の音響透かし技術とその動向	22
2.7	あとがき	24
第3章	振幅変調に基づく情報秘匿技術	25
3.1	まえがき	25
3.2	埋め込み方法	26
3.3	検出方法	27
3.4	振幅変調強度の決定方法	30
3.5	埋め込み区間同期検出方法	32

3.6	埋め込みおよび検出シミュレーション実験	32
3.6.1	ステゴ信号に対する変形	33
3.6.2	実験結果	35
3.7	情報秘匿に伴う音質劣化の主観評価	37
3.7.1	AXB法による音質劣化検知閾測定	39
3.7.2	隠れ基準付き二重盲検法による音質劣化評価	40
3.8	音質劣化の主観評価と客観評価との対応	43
3.9	考察	45
3.9.1	知覚検知閾に基づく変調強度設定	45
3.9.2	電子透かしとしての耐性向上へ向けて	47
3.10	あとがき	48
第4章	情報秘匿信号の空間伝搬および携帯電話音声符号化耐性	50
4.1	まえがき	50
4.2	音声信号への情報秘匿と空間伝搬耐性	53
4.2.1	実験条件	53
4.2.2	結果	55
4.3	VCV音節明瞭度試験	57
4.3.1	実験条件	57
4.3.2	結果	57
4.4	携帯電話音声符号化への耐性	58
4.4.1	携帯電話音声符号化方式	58
4.4.2	実験方法	59
4.4.3	実験結果	61
4.5	客観的品質劣化度合評価	63
4.5.1	PESQによる広帯域音声品質劣化度合評価	63
4.5.2	PESQによる狭帯域音声品質劣化度合評価	65
4.5.3	PEAQによる音楽音質劣化度合評価	66
4.6	考察	67
4.6.1	実時間処理埋め込みおよび検出処理	67
4.6.2	携帯電話ネットワークにおけるパケットロスの影響	68

4.6.3	実効データ伝送量	68
4.7	あとがき	69
第5章	既存音響情報秘匿技術との性能比較	71
5.1	まえがき	71
5.2	エコー拡散法に基づく音響情報秘匿	71
5.2.1	埋め込み方法	72
5.2.2	検出方法	73
5.2.3	埋め込み区間同期検出方法	74
5.2.4	埋め込みパラメータ値	74
5.3	埋め込みおよび検出シミュレーション実験	75
5.3.1	埋め込み強度の設定	76
5.3.2	情報秘匿済み音楽信号に対する変形	77
5.3.3	情報秘匿済み音声信号に対する変形	78
5.4	実験結果	79
5.4.1	音楽信号での検出結果	79
5.4.2	音声信号での検出結果	83
5.5	考察	84
5.6	あとがき	85
第6章	情報秘匿に基づく音響信号と同期した情報呈示	87
6.1	まえがき	87
6.2	音響信号と同期した情報呈示	88
6.2.1	埋め込むデータの構造	88
6.2.2	検出処理演算量	89
6.3	データ埋め込みに伴う音質劣化の評価	89
6.4	空間伝搬耐性	91
6.4.1	シミュレーション条件	91
6.4.2	フレームデータ検出率	92
6.5	カラオケ歌詞表示システム	93
6.5.1	埋め込むデータ構造	94
6.5.2	データ埋め込みと同期	95

6.5.3	性能評価	96
6.6	考察	97
6.7	あとがき	99
第7章	結論	100
	謝辞	105
	参考文献	107

第1章 序論

1.1 研究の背景

音，画像，映像，文字などのコンテンツ自体になんらかの情報を埋め込み，必要な時に検出して利用する技術は，情報秘匿 (information hiding, data hiding) とよばれる．デジタル情報メディアの普及に伴い，1990年代後半以降盛んに研究されるようになった技術である．情報秘匿技術は，メディア情報に付加的なデジタル情報を不可分に結び付けることができることが特徴であり，様々な利用法が考えられる．その埋め込む情報や利用法によって，情報秘匿技術は電子透かし (digital watermarking) とステガノグラフィ (steganography) におおまかに分類される．ここでは，音響信号に対する情報秘匿技術を，これら二つの利用法から概観する．

1.1.1 電子透かし

デジタル音楽データのコピー制御や著作権管理は，音ファイルのヘッダ部に暗号化された著作権管理情報を挿入したり，音データ自体を暗号化する仕組みで実現することができる [1]．後者は DRM (Digital Rights Management) システムとして，既に一部のインターネットで販売される音楽データに適用されている．しかし，ヘッダへの情報付加やデータの暗号化では，一旦アナログ信号として再生後，再度デジタル録音された音データ (アナログコピー) に対しては全く無効であり，そのような過程を経た後も有効なコピー制御や著作権管理技術が必要とされてきた．

電子透かしは，主に著作権管理のための情報 (透かし情報, watermark) を音楽信号に埋め込み，主に著作権者側で違法コピーや流出を検知するために，透かし情報を検出する．この透かし情報は，アナログ領域，放送や非可逆圧縮を経ても検出できるような仕組みが必要である．例えば，インターネット配信される音楽に，コンテンツ ID とよばれるその楽曲の演奏者や曲名などを一意に決定できるデータと，購入者情報を埋め込んでおく．もし，いったんアナログ信号に変換された信号をデジタル録音して，DRM をすり抜けた

音楽ファイルが違法配信された場合、違法配信を監視するサーバは、その音楽ファイルをダウンロードして透かし情報を検出することによって、配信者への通報やユーザへの違法配信警告を送ることが可能となる。

インターネットで配信されている楽曲そのものを同定するには、コンテンツ ID を透かしとして埋め込まなくとも、ダウンロードした楽曲の音響的特徴と、データベースに登録されている楽曲の音響的特徴とを比較し検出する、音響指紋 (audio fingerprinting)[2] という技術でも可能である。しかし、音響電子透かしの特徴は、同一の楽曲に対しても異なるデータ (例えば購入者データなど) を埋め込んで識別でき、かつデータベースとの照合なしに、透かしを検出した場所で検出した情報を利用できるのが特徴である。

コピー制御に電子透かしを用いるしくみとしては、音楽信号から透かしとして埋め込まれたコピー許可情報の検出を行う再生機器や、コピー許可情報の埋め込みを行う録音機器を規格化しておくことが考えられる。こうすれば、CD や DVD などのメディアから録音した楽曲データにコピー不可情報を埋め込むことができる。そうした透かし入り音楽信号を識別して再生できる機器を用いれば、アナログ録音やコピーされたデジタル楽曲ファイルが不正に流通して再生される際には、楽曲からコピー不可情報を検出し、再生やコピーができないような仕組みを実現することが可能である。

また、CM 番組の放送回数が契約通りに実施されているかを自動監視するためにも、CM 放送音声トラックへの透かし情報埋め込みが有効である。

1.1.2 ステガノグラフィ

ステガノグラフィは秘匿通信ともよばれ、埋め込む対象である音響信号の内容とは、一般に関係の薄いデータを埋め込む。ここでは、埋め込まれた音響信号は“おとり”となり、埋め込んだデータの方に利用価値があるような、利用形態を指している。ここで、おとり信号の自体からは、埋め込まれたデータの存在は分からないので、秘密裡に通信が可能となる。

一方、ステガノグラフィをより広義でとらえて、埋め込む対象の音響信号に対する補助データとしての役割をもつデータを埋め込む利用も考えられる。例えば、狭帯域音声符号化に対する帯域拡張用の補助データを音声符号化データ中に埋め込めば、狭帯域音声符号化方式の下位互換性を保ったままで、補助データに対応する復号化を行えば広帯域音声を利用することが可能となる [3, 4, 5, 6]。また、映画音楽やセリフ音に映画の字幕情報を埋

め込み，映画を視聴する難聴者の手元の機器で字幕情報を検出して専用字幕として表示すれば，健聴者と難聴者は同じスクリーン字幕無しの映画を楽しむこともできる．

情報秘匿技術の利用は，著作権保護/管理や秘匿通信といった従来重要視されてきた分野だけに限られる訳ではない．音響信号にデジタル情報を不可分に結び付けることができるという観点から見ると，音響データ通信路におけるパケット損失の隠蔽 [7]，メタデータの記録と検索 [8]，モニタリング，エンターテインメント，改ざん検出 [9, 10] などを含む情報法科学的な利用も検討されている．また，スピーカ再生音にデータを埋め込みマイクロホンで受音してデータを検出する利用法としては，画像による QR コードに似た情報配信 [11] や，聴覚障害者への情報提供 [12] など，さまざまな分野での利用可能性を秘めている．

1.2 技術とその課題

前節のような利用を考えると，音響信号への情報秘匿技術が満たすべき要求として，以下が挙げられる．

- ブラインド検出: 秘匿情報の検出に，秘匿情報を埋め込む前の原信号を必要としない
- 低音質劣化: 秘匿情報埋め込みに伴う音質劣化が知覚されにくい
- 秘匿性: 第三者に秘匿情報の存在を検知されにくい
- 頑強性: 秘匿情報入り音響信号の伝送および記録に伴う様々な変形 (知覚符号化，量子化ビット数 変換，サンプリング周波数変換など) を経た後でも，秘匿情報の残存性が高い
- 適用性: 広範囲な特徴を持つ音響信号に対して，上記条件を満たす

また，ステガノグラフィ用途として，埋め込む情報の方により価値がある場合は，

- 大容量: 上述の要求点のバランスを満たしつつ，できるかぎり大容量の情報を埋め込み，検出することができる．

さらに，ライブコンサートやアナウンス音声のような空間伝搬する音響信号から秘匿情報を抽出する利用を行う場合，

- 空間伝搬耐性: 秘匿情報入り音響信号に対する電気音響変換および空間伝搬を前提とした歪み(伝送周波数特性歪, 雑音/反射音/残響音の付加)に対して秘匿情報の残存性が高い

も重要である。

これまで, 多くの音響情報秘匿技術が開発されてきたが, 上述の条件を全て満たすかどうかの検証が試みられた技術はない。多くの従来技術は, 対象とする音響信号の物理的性質に多少なりとも依存する手法を採っているが, 広範囲な特徴をもつ音響信号に対して適用可能かどうか, 特にその頑強性について検証した研究は少ない。また, 空間伝搬耐性についても評価された技術は少ない。さらに, 被験者の音質劣化検知能力を明らかにした上で, 情報秘匿に伴う音質劣化の検知実験および音質劣化度合の評価実験を行った研究はほとんど無い。

1.3 本論文の目的と構成

本論文は, ブラインド検出を可能とする帯域分割と振幅変調に基づいた新しい音響情報秘匿技術を提案し, その性能を検証した上で, 技術の新しい利用法を示し評価することを目的とする。

第 1.1 節では, 音響情報秘匿技術について概観し, 第 1.2 節では技術への要求点, すなわちブラインド検出, 低音質劣化, 秘匿性, 頑強性, 適用性, 秘匿容量, 空間伝搬耐性を挙げ, 従来の研究においては, 技術への幅広い要求点を満たしているかどうかの検証が少ないことを説明した。

第 2 章では, 音響情報秘匿技術に関して, 用いられる用語や概念などをはじめに説明し, 技術への要求とそれに対応した評価方法を示す。次に, 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解説し, その特徴を簡単にまとめる。さらに, 性能向上のため補助的に用いられる技術も説明する。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。

第 3 章では, 帯域分割と振幅変調に基づく音響情報秘匿技術を説明し, 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基本的な空間伝搬耐性を, 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] を用いて明らかにする。また, 情報埋め込みに伴う音質劣化については, 音質劣化を比較的検出しやすい楽曲に対して, 検知限となる埋め込み強度を明らかにし, 検知限以上の強度で秘匿情報を埋め込んだ場合の主観的

な音質劣化度合を調べる．そして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, Perceptual Evaluation of Audio Quality)[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．

第4章では，前章で示した技術を，残響と背景雑音が重畳する空間伝搬条件において利用することを検討する．まず，音声データベースに収録されている22名の男声および女声音声を埋め込み対象として，スピーカ再生かつマイクロホン受音を前提とする空間伝搬条件での耐性を調べる．次に，埋め込み情報の検出を，マイクロホンにより受音した端末で行うのではなく，携帯電話の音声通話により接続する携帯電話ネットワークの先にあるサーバコンピュータにおいて実行するため，携帯電話の音声符号化を経ても情報伝達が可能かどうかを調べる．対象は，音声データベースおよび音楽ジャンルデータベースの楽曲100曲であり，シミュレーション実験および実室内環境における実験を行う．

第5章では，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による情報秘匿手法[15]について，シミュレーション実験によってその性能を比較する．性能比較は，音響電子透かしとしての利用を前提として，埋め込み情報量と客観音質劣化度合を両手法間にて揃えた上で，情報秘匿済み信号に対する各種の変形への耐性について行う．また，音声信号に情報を埋め込んだ場合に，スピーカ再生およびマイクロホン受音を前提とする空間伝搬耐性があるかどうかの点でも比較を行う．

第6章では，振幅変調に基づく音響情報秘匿技術を用いて，スピーカから再生される音に同期して情報を呈示するシステムを提案する．ここでは，埋め込む情報にブロック符号化によるエラー訂正を施し，残響および付加雑音環境下において，エラー訂正の範囲に収まる正検出率をシミュレーション実験により求める．また，情報の同期呈示に必要な埋め込み情報検出における時間精度も同様に調べる．最後に，カラオケの伴奏音楽に歌詞を同期呈示するための情報を埋め込んでおき，スピーカ再生される伴奏音楽に同期してリアルタイムに歌詞を表示させるシステムの構成を説明し，その有効性を検証する．

第7章は，本論文の結論として，研究の成果のまとめと今後の課題を述べる．

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。

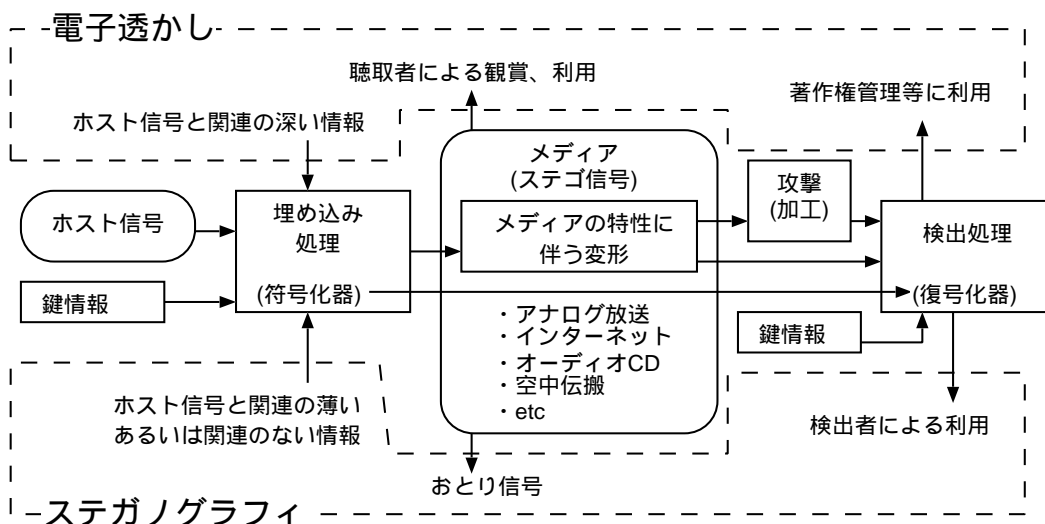


図 2.1: 音響信号への情報秘匿技術の概要。

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出が困難になるべく、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十分に少ないことが、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つてよい。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 デジタル領域での情報秘匿手法

ステゴ信号が、デ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 $1/10 \sim 1/5$ 程度にもおおよび、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、音波形を標準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。

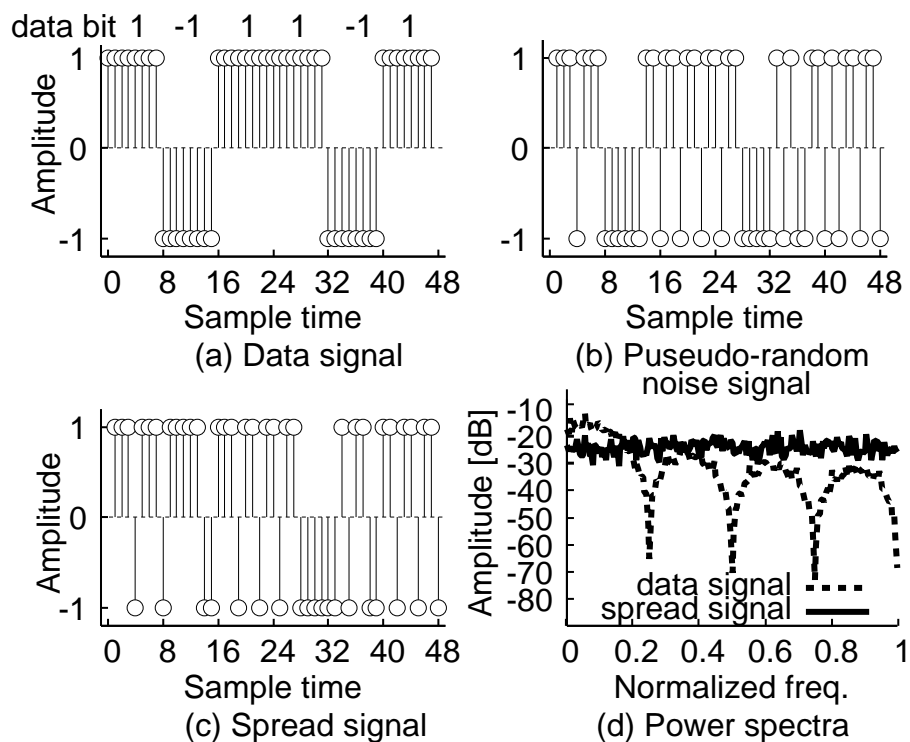


図 2.2: 直接拡散法での信号表現

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。

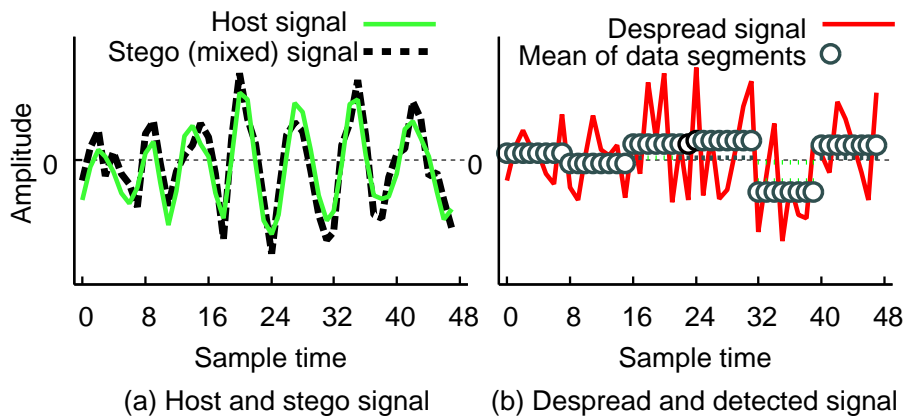


図 2.3: 直接拡散法でのステゴ信号と検出過程

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 d_0 の遅延を加えた場合にビット値 0、点線の d_1 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。

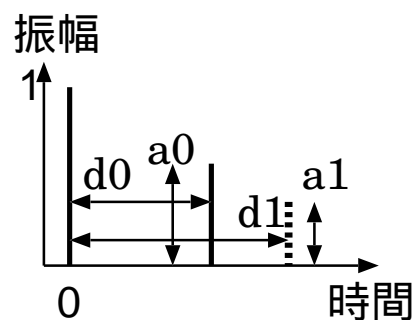


図 2.4: エコー法で用いる埋め込みビット値に対応するインパルス応答。例えば、実線のエコーを与える場合はビット値 0、点線のエコーの場合はビット値 1 を割り当てる。

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]．

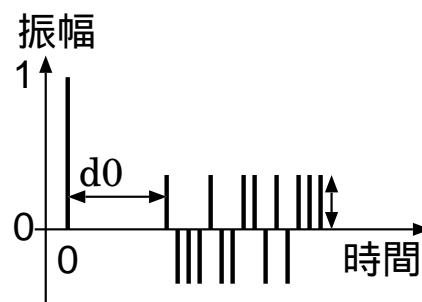


図 2.5: エコー拡散法で用いるインパルス応答の例．

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) .

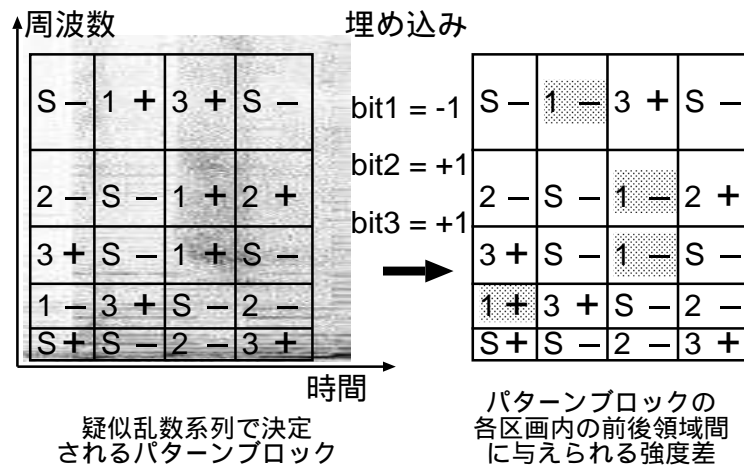


図 2.6: 時間周波数平面上で区切られたパターンブロックへのデータ埋め込みの概略 . 埋め込む bit1 の値が -1 なので , 右側のパターンブロックではハッチングされた bit1 の区画の符号が反転している .

埋め込み処理では , 1 区画分の長さのホスト信号は , 半分オーバー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、それに变形が加わったり、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

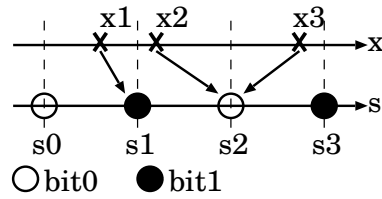


図 2.7: 量子化変調法による埋め込みの模式図

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、埋め込み部位に限るということは、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、 x_1, x_2, x_3 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 s_0 から s_3 までビット値 0, 1 に交互に割り当てられている。 x_1 から x_3 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 s_1, s_2, s_2 という値にそれぞれ変換される。

検出時には、 s_1, s_2 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。

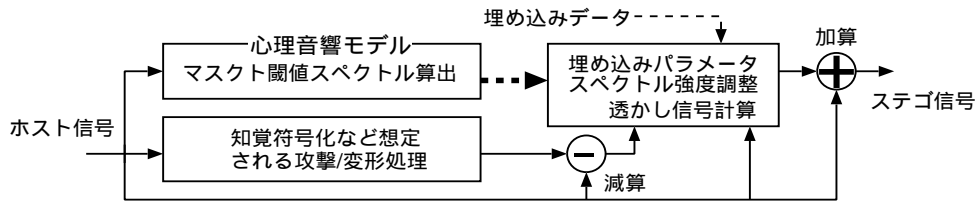


図 2.8: 心理音響モデルを用いた透かし信号スペクトル調整と，変形を前提にした透かしパラメータおよび強度調整.

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进行评估する試みとして，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の透かしの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の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、音質劣化に関する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。

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)。

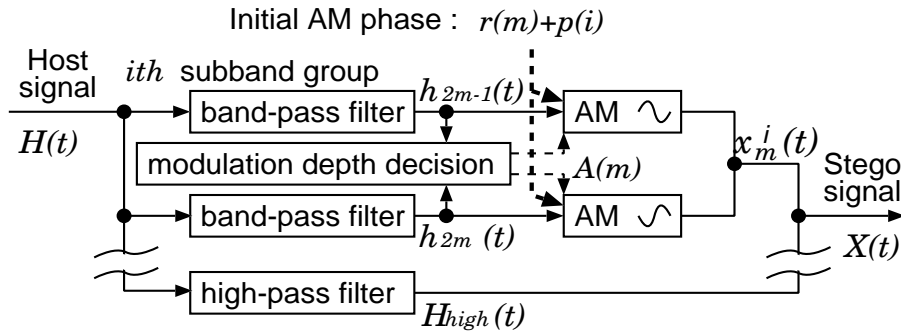


図 3.1: 帯域ペアへの逆位相振幅変調。

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

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

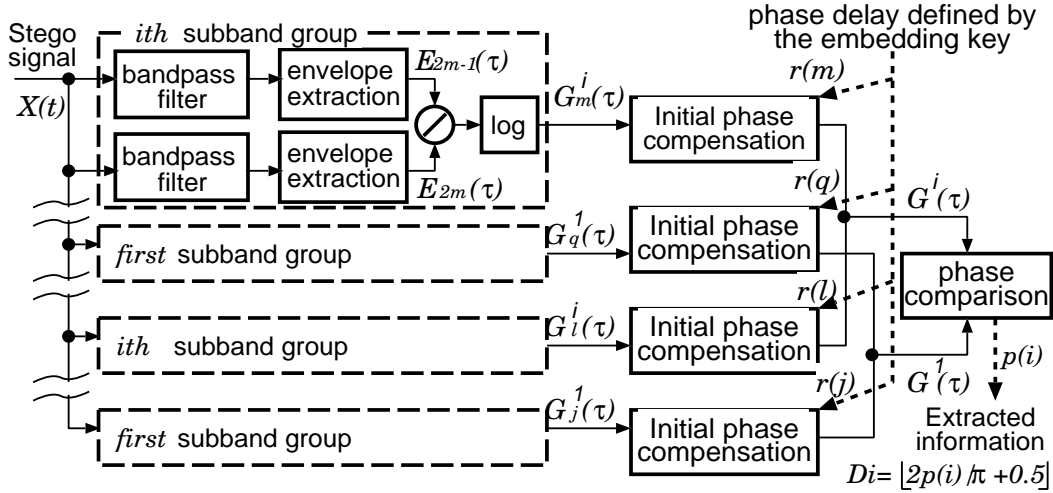


図 3.2: 埋め込まれた秘匿情報検出の過程 .

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。

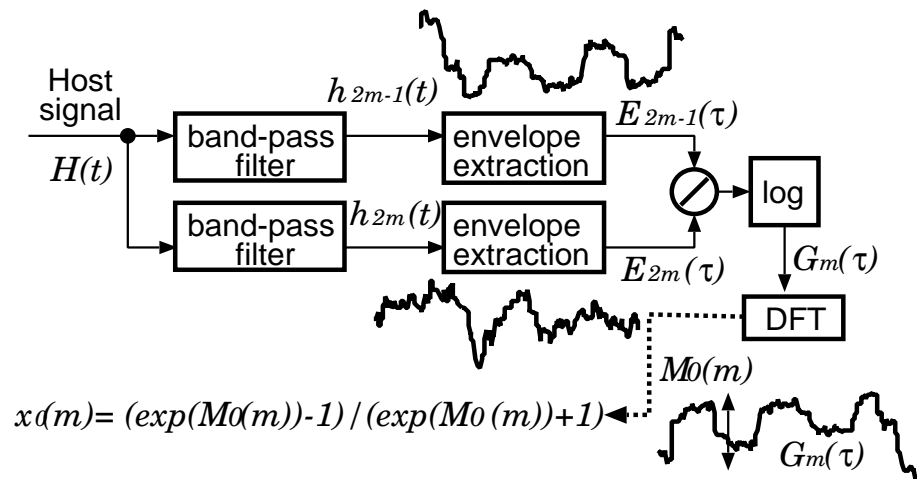


図 3.3: 変調強度決定のブロックダイアグラム。

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.

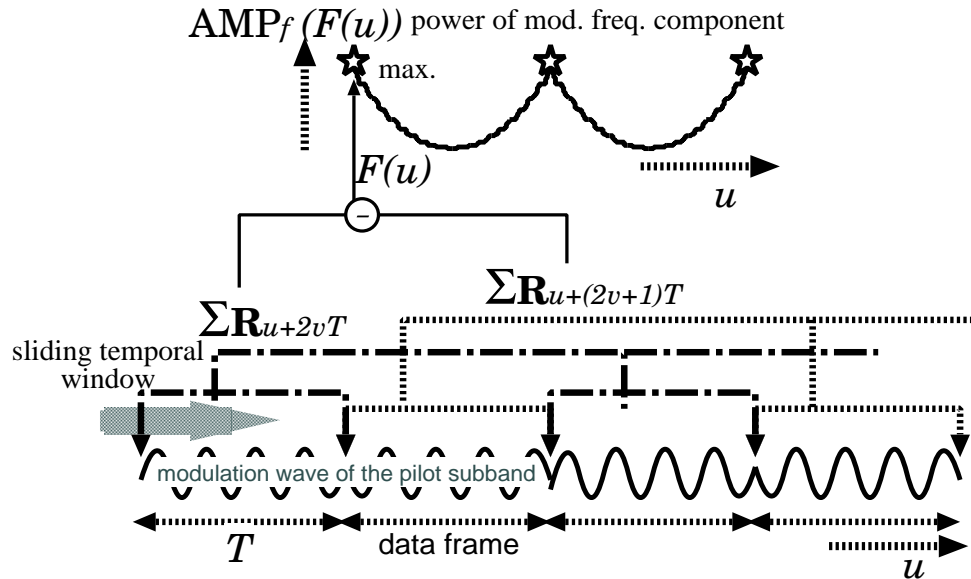


図 3.4: 埋め込み区間検出のための累積変動波形からの最大パワー検出。

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。

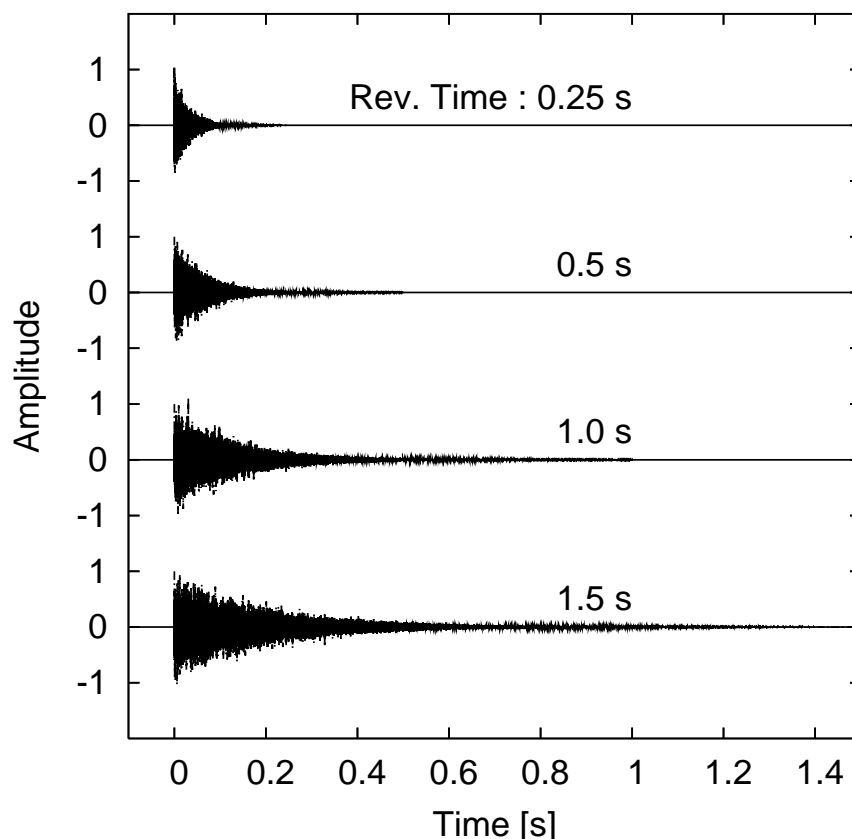


図 3.5: 耐性シミュレーション実験で用いた残響付加のためのインパルス応答波形。

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 .

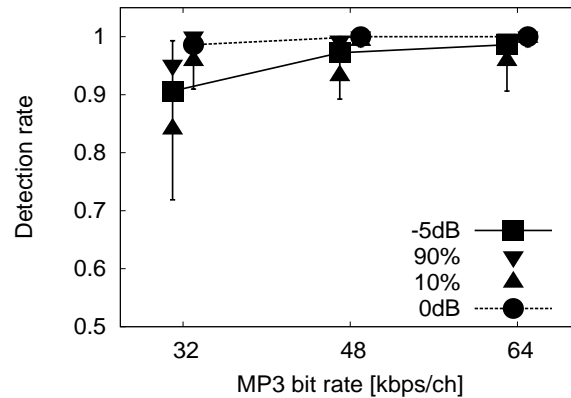


図 3.6: MP3 符号化に対するビット検出率 . は埋め込み強度 -5 dB の中央値 , は埋め込み強度 0 dB の中央値 , は 90 パーセンタイル点 , は 10 パーセンタイル点 , エラーバーは 100 曲中の最小と最大検出率を示している .

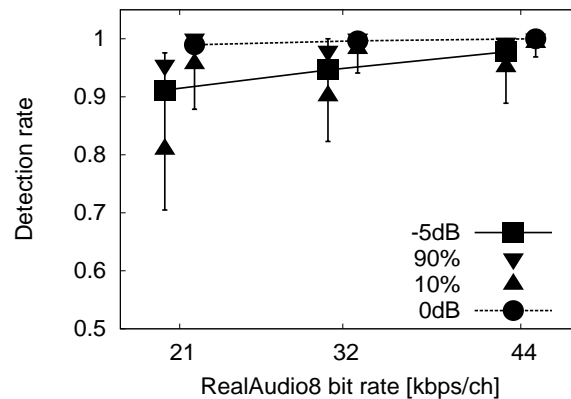


図 3.7: RealAudio8 符号化に対するビット検出率 .

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 .

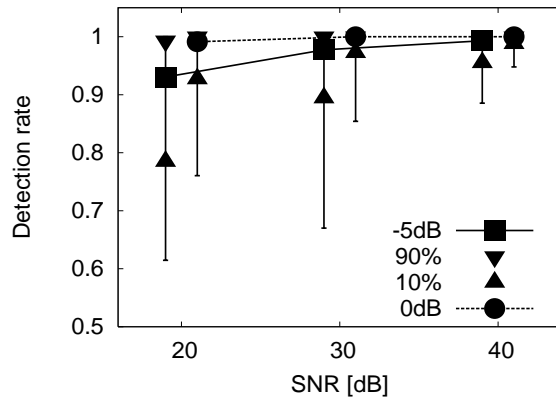


図 3.8: ホワイトノイズ付加に対するビット検出率 .

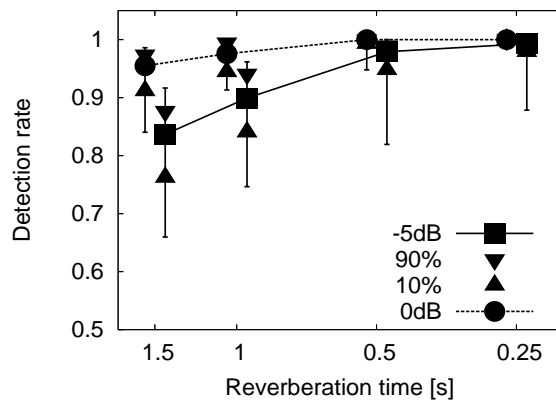


図 3.9: 残響付加に対するビット検出率 .

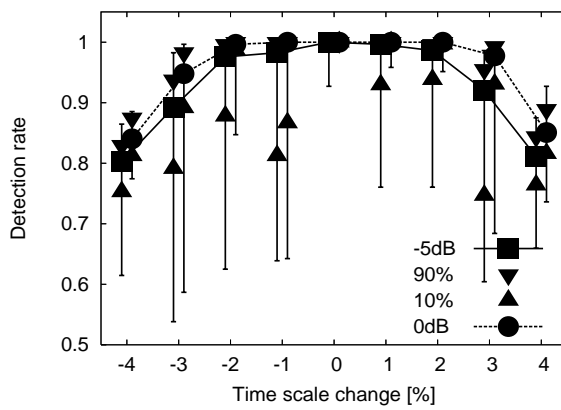


図 3.10: 時間長伸縮に対するビット検出率 .

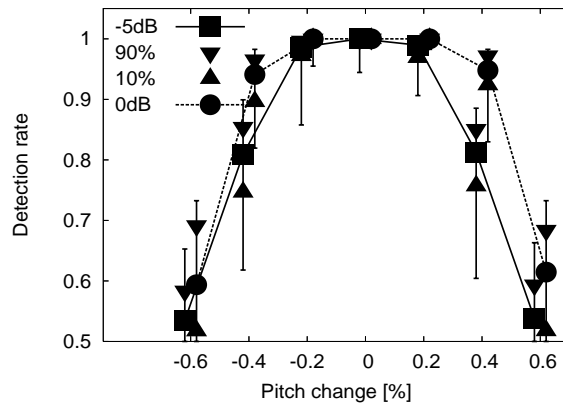


図 3.11: ピッチ変換に対するビット検出率

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．

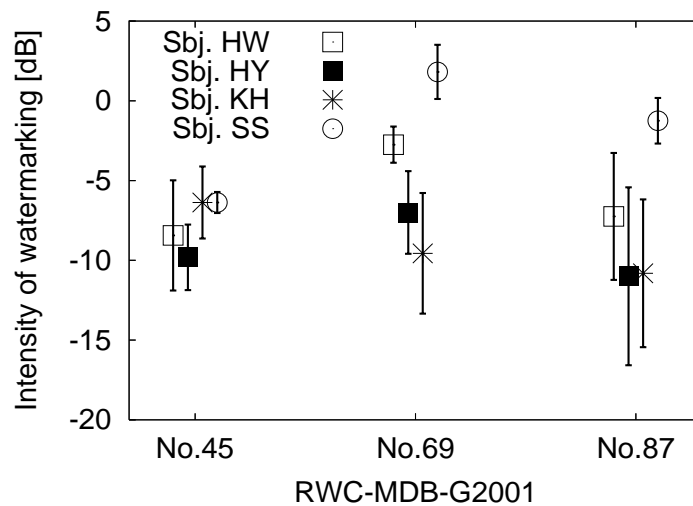


図 3.12: 埋め込み強度の検知閾．エラーバーは ± 1 標準偏差を示している．

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

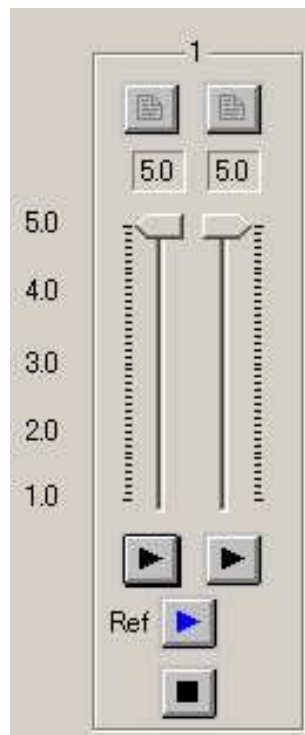


図 3.13: 音質劣化評価実験に用いたユーザインタフェース。Ref はホスト信号を再生するボタンであり、その上の左右どちらかの再生ボタンに音質劣化信号とホスト信号が割り当てられている。被験者は音質が劣化していると感じられる方のスライダーを調整して音質劣化の評価を与える。

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

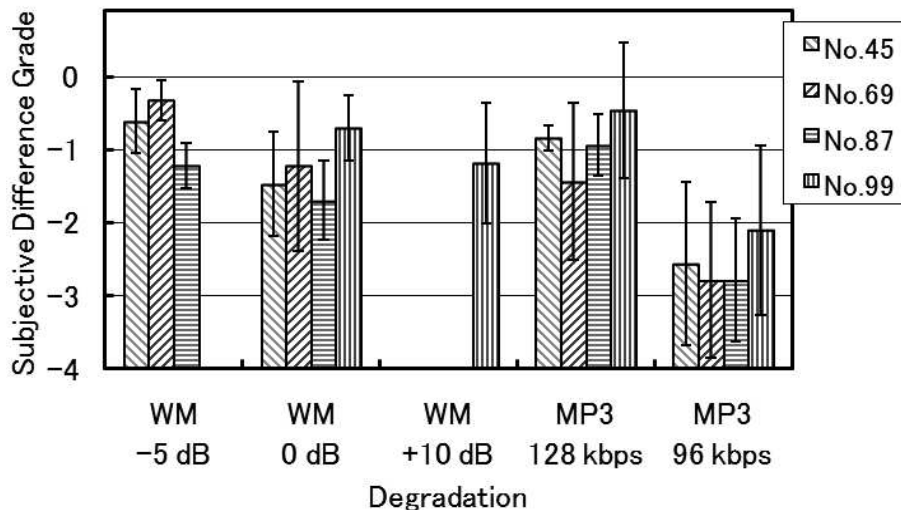


図 3.14: ITU-R BS.1116-1 に準拠した実験により得られた主観差分等級 (Subjective Difference Grade) .

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。

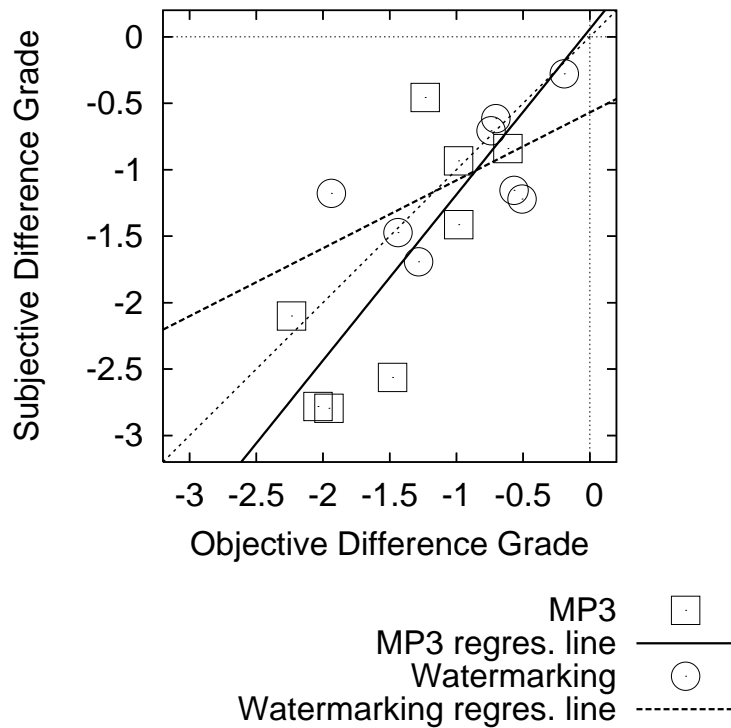


図 3.15: 平均 SDG と ODG の対応。□ は振幅変調に基づく情報秘匿に起因する音質劣化，○ は MP3 符号化による音質劣化である。

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．

第4章 情報秘匿信号の空間伝搬および携帯電話音声符号化耐性

4.1 まえがき

第3章では、振幅変調に基づく情報秘匿手法を提案し、様々な音楽信号に対する音響電子透かしとしての性能を評価した。その結果、残響や付加雑音に対して頑強であることが示された。この特徴は、この手法が音響電子透かしとしての用途だけでなく、必要な情報を音響信号に秘匿して送信/保存し、受信/読み込んで秘匿情報を取り出して利用するステガノグラフィ用途として、特に、ステゴ信号をスピーカ再生し、マイクロホンで受音する、空間伝搬条件でのステガノグラフィ用途に利用できることを示唆している。

近年、このような空間伝搬を前提としたステガノグラフィ技術がいくつか考えられている。この場合の受信器としては、携帯電話やPDAなどが挙げられる。埋め込むデータと利用方法をいくつか挙げてみると、まず、商品やサービスに関する情報を、音響信号(例えばCM音楽)の聴取者に与えて広告宣伝効果を高めるものがある[58]。また、聴覚によって音響信号(例えば公共空間でのアナウンス音声)の内容を把握できない聴覚障害者に対して、音響信号の内容と関連の深い情報を与える福祉用途[12]がある。さらに、博物館等の展示情報へのポイントをスピーカ再生の音楽に埋め込んで、閲覧者の持つPDAで復号化し、ポイントの示す情報をあらかじめPDAに蓄積しておいて呈示したり、受音位置に応じた情報を呈示する方法[59]も試行されている。

このような用途のためは、従来の音響信号に対する電子透かしやステガノグラフィにおいて用いられていた技術に無かった要求も現れてくる。つまり、スピーカ再生しマイクロホン収音することを経るなかで、ステゴ信号は以下に挙げる変形に耐性を持つ必要がある。

1. スピーカおよびマイクロホンにおける伝送周波数特性の歪
2. スピーカからマイクロホンまでの音響経路で生じる反射音や残響

3. マイクロホン受音時に付加される背景雑音

4. マイクロホン受音時の入力超過による振幅クリッピング歪

その一方で，従来の音響電子透かし用途において重要視されていた，知覚符号化 (MP3 や MPEG2AAC など) や，悪意ある攻撃 (ピッチ変換，時間長変換，データ切り取りなど) への耐性については，結果として耐性が高くなることも十分あるとは言え，あまり考慮する必要はない．また，ステゴ信号の音質劣化を用途によって許容される程度まで認めつつ，必要な耐性を確保し埋め込みデータ量を高める必要がある．

これらの要求を，主に音響電子透かしを目的として開発されたものが多い従来の音響情報秘匿技術が満たしているかを検討してみる．パッチワーク法 [39] は，強度変化を与える時間幅が 100 ms 程度と短いので，受音点がスピーカから遠ざかった場合に反射音や残響の影響を受けやすい．2 チャンネル伝送を前提として片チャンネル毎に強度変調を与える手法 [11] は，反対側チャンネルの影響が少なくなるよう，情報検出時にはマイクロホンを一方のスピーカに近づける必要がある．エコー拡散法 [26] は，反射音や残響の影響を軽減するためには，第 5 章にみるように埋め込み時間フレームを長く設定する必要があり，埋め込みデータ量は限られてしまう．また，スペクトル拡散法 [60] は，付加雑音には強いが，埋め込みデータ量は数 bps と十分でない．音響 OFDM (Orthogonal Frequency-Division Multiplexing; 直交周波数分割多重方式) [61] は埋め込みデータ量は 1 kbps 程度と格段に多いが，情報を埋め込む帯域を高域に限定した上で，音楽信号のスペクトルに合わせて埋め込みデータ信号の振幅を調整するため，短時間 (100 ms 程度) の振幅の小さいデータフレームに反射音や残響音，雑音が重畳されると，埋め込み情報の検出に大きく影響を与えると考えられる．

これらをまとめると，一般的に数 100 ms 以下の短い時間波形を 1 つのデータフレームとして情報を埋め込む従来の音響情報秘匿技術は，そのフレームの信号強度が弱い場合に，先行する信号強度の強い部分の反射音や残響成分，時間変動する背景雑音などによる影響を受けやすく，スピーカから離れた位置での秘匿情報検出が十分でなくなる恐れがある．これは，もともと著作権管理を目的とする電子透かし技術が想定しているステゴ信号への変形は，MP3 などの知覚符号化圧縮や一定の付加雑音，時間伸縮やピッチ変換 [17] であり，空間伝搬に起因する残響や変動する雑音は想定されていないことも一因である．

空間伝搬用途に使用する情報秘匿手法もいくつか提案されているが，マイクロホンがスピーカに近接することを要求したり [11]，埋め込み情報量が少なかったり [60, 12]，埋め

込み情報量は多くても、空間伝搬に起因する上記のような変形に対する耐性が定量的に調べられていなかったりする [62]。また、多くの音響情報秘匿手法においてステゴ信号への変形に対する耐性や埋め込みデータ量は、音響信号の特徴に依存することが知られているが、音声から音楽までの様々な音響信号に対して耐性やデータ埋め込み量を定量的に検証した例はほとんど無い。

本章の目的は、二つある。第一に振幅変調に基づく音響信号への情報秘匿手法が、音声信号に対して空間伝搬条件においてどの程度の品質劣化と耐性を持つかを調べることである。具体的には、空間伝搬を前提とした、前述の 1～3 の条件を満たすことを、音声信号に対して確認する。なお前述 4 の振幅クリッピング歪に対する耐性条件に関しては、第 6 章において音楽信号を対象に検討する。情報秘匿に伴う音声品質の劣化に関しては、VCV 音節識別実験と、符号化音声信号の客観的音質評価手法のひとつである ITU-T P.862 PESQ (Perceptual Evaluation of Speech Quality) を用いる。ここで音声信号を取り上げた理由としては、駅や空港などの公共空間におけるアナウンス音声に対して、事前にあるいはリアルタイムに情報を埋め込み、利用者の手元の機器で検出して埋め込みデータを利用する用途を前提とするからである。これは、先に応用例として挙げた、難聴者へ健聴者と同じ音声信号から同等の情報提供を行うサービスおよび、国外からの旅行者等の音声アナウンス言語を理解しない人々へ、アナウンス内容と同等の情報を提供するサービスにおける利用を意図している。

第二に、音楽あるいは音声信号に情報を埋め込んだステゴ信号が、スピーカ再生されて空間伝搬した後、携帯電話のマイクロホンで收音され、音声通話による音声符号化を経て携帯電話ネットワークによって伝送され、受信先で復号化した音響信号より秘匿情報を検出する利用について検討を行なうことである。つまり、空間伝搬かつ音声符号化に対する耐性を調べることである。

利用者の端末へより多くの情報を呈示したい場合に、埋め込んだ情報のみを検出して呈示していたのでは、呈示できる情報量、つまり埋め込み容量が十分でない。よって、伝送したいデータをそのまま埋め込まずに、データに対するインデックス情報を埋め込み、伝送情報量を削減することが当面必要である。そのためには、伝送/表示したい情報を表示端末に事前に蓄積しておき、検出したインデックスに対応する情報を表示する方法と、埋め込まれた情報を表示端末では検出せずに、マイクロホンで受信した音響信号を音声通話にてサーバコンピュータに伝送し、サーバコンピュータ側で埋め込みインデックス情報の検出と、インデックスに対応する表示情報を端末へ電子メール等により伝送する方法が

考えられる (図 4.1 参照) . 前者の利用形態での音響情報秘匿手法の有効性は , 第一の目的において検証される . しかし後者のほうが , 新たなデータ検出ソフトウェアを利用者端末にインストールする必要がない点で適している . この場合 , マイクロホンで収録されたステゴ信号は , 携帯電話の音声コーデックによって符号化の後 , 公衆電話ネットワークを経由した伝送先で復号化され , サーバコンピュータの検出プログラムに入力される . このため , ステゴ信号の携帯電話による音声符号化と復号化に対する耐性が必要となる .

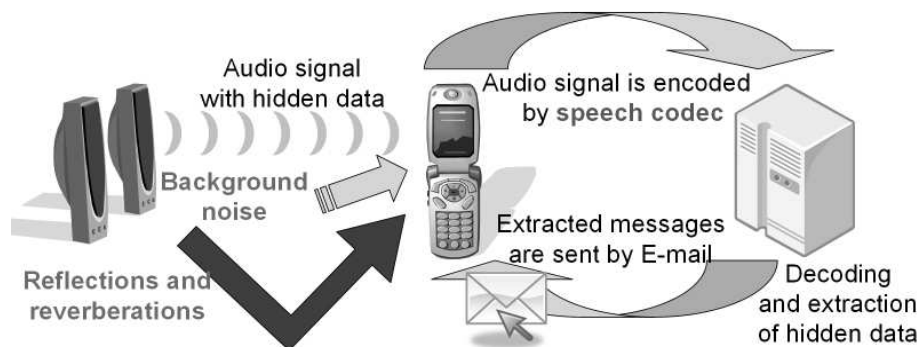


図 4.1: 携帯電話の音声通話ネットワークを通じた秘匿情報の復号化と利用者端末での情報呈示 .

4.2 音声信号への情報秘匿と空間伝搬耐性

4.2.1 実験条件

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 に収録されているサンプリング周波数 16 kHz の音声を 22.05 kHz に変換して , 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 . 埋め込み強度である振幅変調度は , リアルタイムに埋め込み処理を行うことを想定して , ホスト信号によらず一定の変調度とすることとし , 0.3~ 0.7 の範囲で 0.1 ステップで設定した . 埋め込む帯域は , 6034Hz 以下の帯域とした . 秘匿情報のビットレートは , 埋め込みフレーム時間長を 3 秒としたとき 64 bps , 4 秒としたとき 48 bps であり , ランダムなビット値を埋め込んだ . 埋め込み時のパラメータ値は , 表 4.1 に示した .

残響のある環境を想定して , ステゴ音声信号に , RWCP 実環境音声・音響データベース [63] 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 . このインパルス応答波形を図 4.2 に示した . この波形の絶対

表 4.1: 埋め込み条件 .

Parameters	Values	
	64 bps	48 bps
bit rate	64 bps	48 bps
sampl. freq.	22050 Hz	←
freq. region	≤ 6034 Hz	←
bandwidth	21.5 Hz	←
subband pairs	140	←
subband groups	25	←
pairs per group	5 — 6	←
frame period	3 s	4 s
mod. freq. [Hz]	1, 1.67, 2.33, 3	1, 1.5, 2, 2.5

値ピークを中心とした 128 サンプルにハニング窓掛けを行って直接音成分を取り出し，そのパワースペクトルを求めることにより，模擬されるスピーカからマイクへの振幅伝達特性を求めて，図 4.3 に示した．これにより，スピーカやマイクロホンのフラットでない伝送特性も模擬できることが分かる．なお，このインパルス応答の実際の残響時間を，シュレーダ積分法 [64] によって， -5 dB から -25 dB まで減衰する時間を 3 倍して求めたところ，約 1.1 秒であった．このインパルス応答の収録環境は，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は，他研究者が容易に入手可能かつ実環境で測定したものであるため用いた．

その後，背景雑音として 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 500 Hz， -9 dB/oct. : 他の環境騒音の平均的スペクトルに近い) のいずれかを付加した後，秘匿情報を検出する処理を行った．4 つの環境騒音は，TARGET ENTERTAINMENT 製作，リッスンジャパン (<http://listen.jp/store/>) 販売の「効果音ライブラリ・環境音」から選び，44.1 kHz サンプルング/128 kbps の MP3 ファイルを WAV ファイルに変換して，冒頭の左チャンネルを 22.05 kHz にダウンサンプルングして用いた．5 種類の背景雑音は，オーバーオールの音声信号パワーに対して，信号対雑音比 (SNR) は 10 dB と 20 dB の 2 通りとした．全ての音響信号は，サンプルング周波数 22.05 kHz に変換後，処理を行った．5 つの背景雑音のそれぞれの平均スペクトルを，図 4.4 に示した．

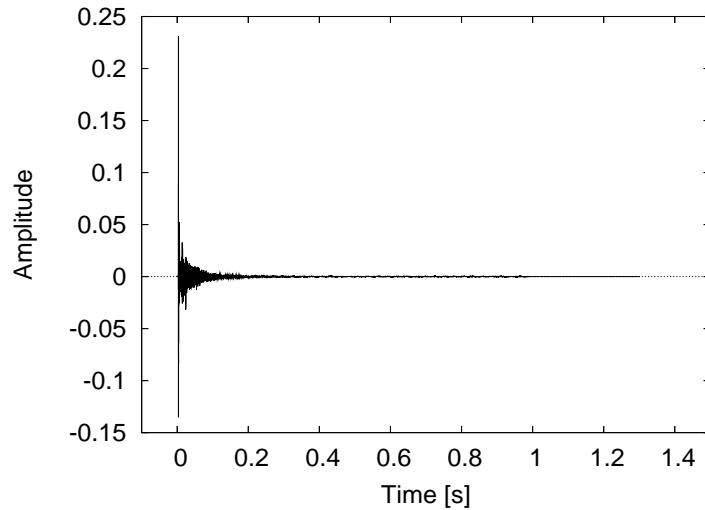


図 4.2: RWCP 実環境データベースより選んで耐性シミュレーション実験で用いた残響付加のためのインパルス応答 (ir130.dat) 波形 .

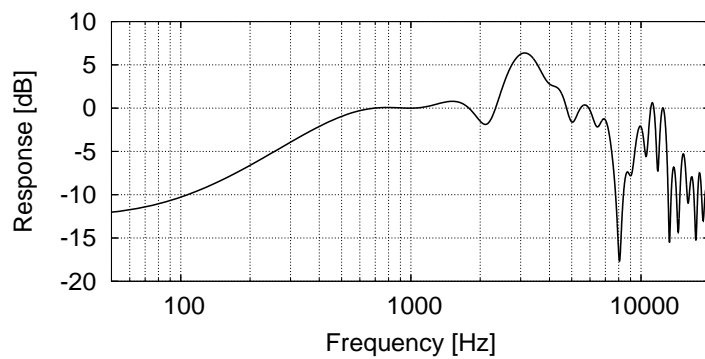


図 4.3: インパルス応答における直接音成分のパワースペクトル.

22 種のステゴ音声信号と 5 種の背景雑音を組み合わせた 110 条件が、5 段階の変調度、2 段階のノイズ強度、2 つの埋め込みビットレート条件の組み合わせに対してシミュレーションされた .

4.2.2 結果

図 4.5 は、SNR およびビットレート毎に、埋め込んだビット値と同じビット値が検出された率を、110 埋め込み条件の中央値と、誤差棒により 10・ 90 パーセンタイル値で示した . この結果から、例えば振幅変調度 0.4 かつ 48 bps で埋め込みを行えば、90%の模擬条件において、84%以上のビット検出率が得られることが分かった .

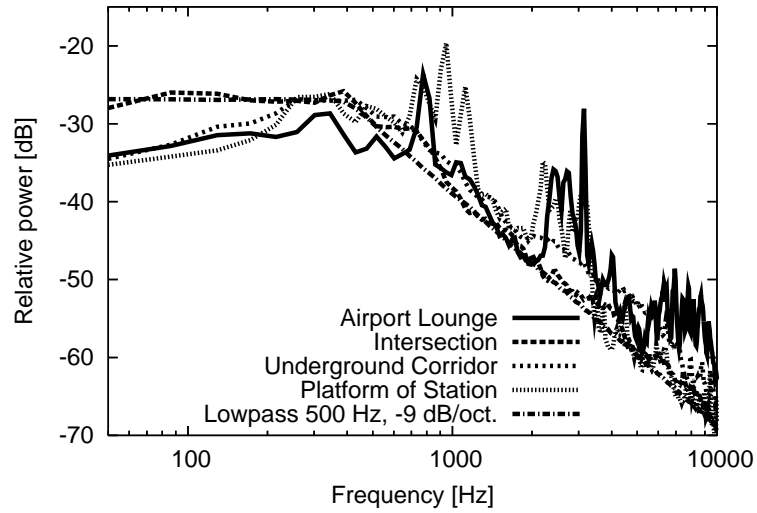


図 4.4: ステゴ信号に付加した, 5つの付加雑音の平均パワースペクトル.

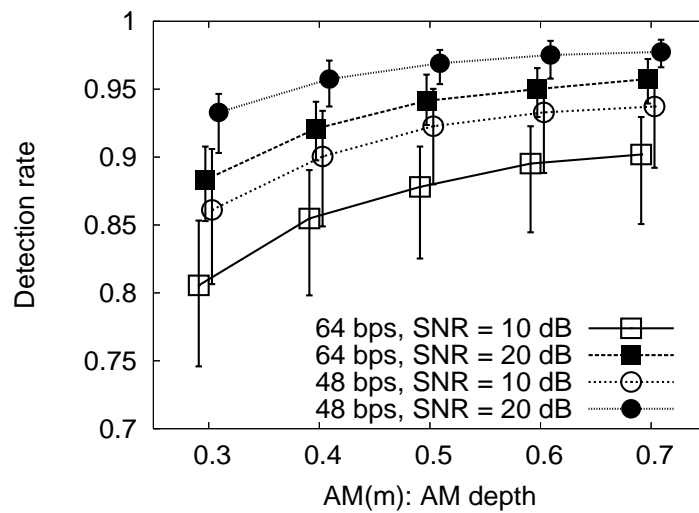


図 4.5: ビット検出率の中央値. 誤差棒は 110 条件中の 10 から 90 パーセントイル検出率を示す.

4.3 VCV 音節明瞭度試験

4.3.1 実験条件

音響電子透かしとしての利用においては、音質劣化の少ないことは重要であるが、第 4.1 節で述べたような応用場面にて音声信号に情報を埋め込む場合、ステゴ音声信号の音声了解度が十分であれば、多少の音質劣化は許容される。そこでまず、情報秘匿が音声了解度にどの程度影響を与えるのかを調べる基礎として、情報秘匿済み VCV 音節の明瞭度試験を行った。

VCV 音節は、a,i,u,e,o の先行 5 母音と 25 の日本語子音 (b, by, ch, d, g, gy, h, hy, j, k, ky, m, my, n, ny, p, py, r, ry, s, sy, t, w, y, z) 、後続母音 a によって構成される 125 種とした。埋め込み条件は前節のシミュレーションと同じで、埋め込み強度は振幅変調度 0.4, 0.6 および埋め込み無しとした。雑音付加時の明瞭度も調べるために、第 4.2 節で用いたローパスノイズを SNR 10 dB で付加する条件も加えた。被験者は聴力レベル 10 dB 以下の 5 名であり、防音室内でヘッドホン両耳聴 (片耳あたり実効音圧 72 dB) にて聴取した音声に対して、聞き取った音節をパソコンに入力し回答した。音節、SNR 条件、埋め込み強度条件の組み合わせをランダムに 1 巡するセットを 2 回異なる日に繰り返した。

4.3.2 結果

図 4.6 に、被験者間の平均正答率を示した。誤差棒は被験者間での最小と最大正答率を示している。この結果より、情報秘匿を行うと雑音環境下で音節明瞭度が低下することが分かる。しかし、振幅変調度 0.6 で埋め込みを行った場合でも、雑音環境下で最低でも 83%以上の正答率が得られている。文章了解度は音節明瞭度より一般的に高いゆえ、ステゴ音声信号に対しても十分な文章了解度が得られるだろうと考えられる。

異聴表を用いた分析では、どの条件でも先行五母音と第二母音には誤答がほとんど見られず、先行母音に依存した子音の異聴が目立った。これは情報秘匿のため与える振幅変調が、先行母音と子音との調音結合部に影響を与えるためと考えられる。

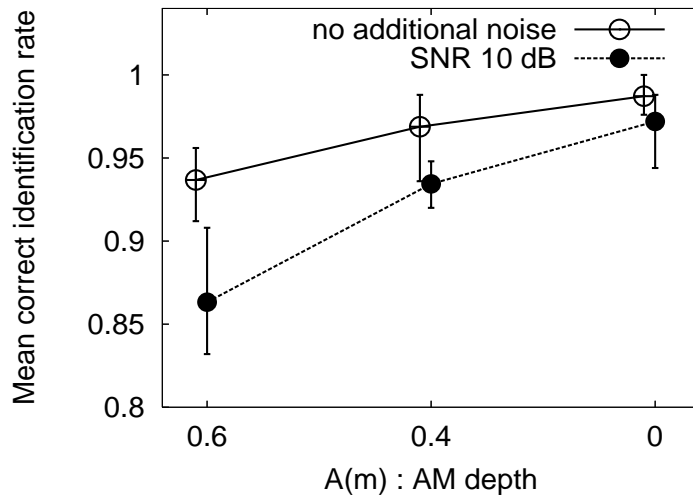


図 4.6: 125 VCV 音節に対する音節明瞭度. 誤差棒は 5 名の被験者中の最大値と最小値を示す.

4.4 携帯電話音声符号化への耐性

音楽あるいは音声信号に情報を埋め込んだステゴ信号が、スピーカ再生されて空間伝搬した後、携帯電話のマイクロホンで收音される。そして、音声通話による音声符号化を経て、受信先で復号化した音響信号より秘匿情報を検出する状況を前提に、携帯電話音声符号化に対する耐性を調べる。

4.4.1 携帯電話音声符号化方式

第三世代 3GPP(3rd Generation Partnership Project) 携帯電話においては、デジタル音声信号符号化方式として、CELP(Code-Excited Linear Prediction, 符号励振線形予測) 系の最新のコーデックである AMR(Advanced Multi-Rate) 方式が多く用いられている [65]。この音声符号化は、音声生成の要である声帯振動(励振源)と声道共鳴(フィルタ)を、それぞれ表現するパラメータ値として符号化することで情報圧縮を実現する。AMR 方式では、あらかじめ単位振幅のパルスの取りうる位置と極性をお互いに少数に限定して決めておき、それら数本のベクトルの和で励振源を表現する。そして各パルスの位置の最適な組み合わせを歪の評価で選択する。伝送されるパラメータ値は、LSP(Line Spectral Pair)、ピッチ、コードベクトルとゲインである。

復号時には、コードベクトルをそれぞれのゲインで調整した後加算されて生成した励振源を、線形予測フィルタに通して音声信号を合成する。よって、ベクトル符号帳の情報量

は少なく、効率のよい情報圧縮が可能である一方、入力波形に存在する微細な時間波形情報はこのような分析合成によって失われる。よって、エコー法やスペクトル拡散法によって情報を埋め込んだ音響信号に対して音声符号化を行った場合、秘匿情報の検出は困難となる。

また、AMR 方式の特徴としては、8000 Hz サンプリングかつ 13 bit 直線量子化 (8bit A-law あるいは μ -law 圧縮) された音声波形に対して、短い時間フレーム (160 サンプル、0.02 秒に相当) 毎に 4.75~ 12.2 kbps の広い範囲でビットレートを可変して伝送ができる点である。さらに、有音無音検出機能、背景雑音生成機能、フレームデータ誤り隠ぺい機能などがあるが、これらの機能はここでは扱わない。

4.4.2 実験方法

音声及び音楽信号の 4 kHz 以下の帯域にデータを埋め込み、残響や背景雑音を付加した後、AMR コーデックによる符号化および復号化を経た後の波形に対して、検出処理を行なった。データ埋め込みビットレートは 8 bps とし、パラメータ値は、表 4.2 に従い、ランダムなビット値を埋め込んだ。埋め込み強度である振幅変調度は、0.4 で固定としたが、これは第 4.3 節における VCV 音節明瞭度実験および第 4.5 節での客観的音質劣化度合の評価を元に決定した。また実環境でも、スピーカ再生と携帯電話の AMR 符号化方式による音声録音機能を用いて、ステゴ音声信号の空間伝搬と AMR 符号化が重畳する条件での検出率を調べた。なお、同実環境においてコンデンサマイクロホンを用いた PCM 録音も同時に行うことで、携帯電話受信と音声符号化の影響を調べた。

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。これらの信号はサンプリング周波数 16 kHz であったが、8 kHz に変換して用いた。

音楽信号としては、RWC 研究用音楽ジャンルデータベース RWC-MDB-G-2001 [41] に収録された様々な音楽ジャンルの 100 曲の左チャンネル冒頭 60 秒を用いた。これらの信号はデータ埋め込み時にはサンプリング周波数 44.1 kHz であり、表 4.2 に示したパラメータによりランダムビットデータの埋め込んだ後、サンプリング周波数 8 kHz に変換して用いた。

半分のシミュレーション条件では、残響のある室内において、データ埋め込み済み信号

表 4.2: 埋め込みパラメータ値 .

Parameters	Values
AM depth	0.4
bit rate	8 bps
embedding region	below 4000 Hz
bandwidth	31.3 Hz
subband pairs	63
subband groups	7
pairs per group	9
frame period	6 s
mod. freq. [Hz]	1.17, 1.67, 2.17, 2.67

がスピーカ再生され、マイクロホン受音されることを模擬するために、第 4.2 節で用いたものと同じ、RWCP 実環境音声・音響データベース [63] に収録されている、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、サンプリング周波数 8 kHz に変換してから畳み込んだ。その後、室内の背景雑音に似たスペクトルを持つ Hoth ノイズ [66] を、SNR 10, 20, 30 dB のいずれかで付加した。なお、残響と背景雑音の無い条件も加えた。

AMR 符号化および復号化ソフトウェアは、3GPP TS 26.073 [67] に付属している ANSI-C コードをコンパイルして用いた。シミュレーション時の AMR コーデックのビットレートは、4.75~ 12.2 kbps のビットレートが 2 フレーム毎に 28 フレーム (0.56 秒) の周期で連続的に変化する条件、6.7 kbps, 12.2 kbps の 3 種類とした。

実環境における空間伝搬と AMR 符号化の影響を調べる実験では、容積約 410m³、一辺 12 m の変形正方形教室において、対角線前方中央壁面より 1 m、高さ 1 m の位置に置いた 12 cm フルレンジスピーカ (Panasonic WS-X66) より、情報埋め込み済み音声信号と、オーバーオール SNR を一定とした Hoth ノイズを混合して再生した。スピーカから 5.4 m の距離、高さ 1.3 m の位置に携帯電話 (Panasonic Mobile Communication 820P) の背面をスピーカに向けてマイクスタンドで固定し、ボイスレコーダ機能 (AMR 12.2 kbps) にて録音した。携帯電話機真横においた騒音計で測定した等価騒音レベルは、音声が 65 dB と 55 dB の 2 条件、Hoth ノイズは 45 dB で一定、室内の暗騒音は 31 dB であった。録音時に

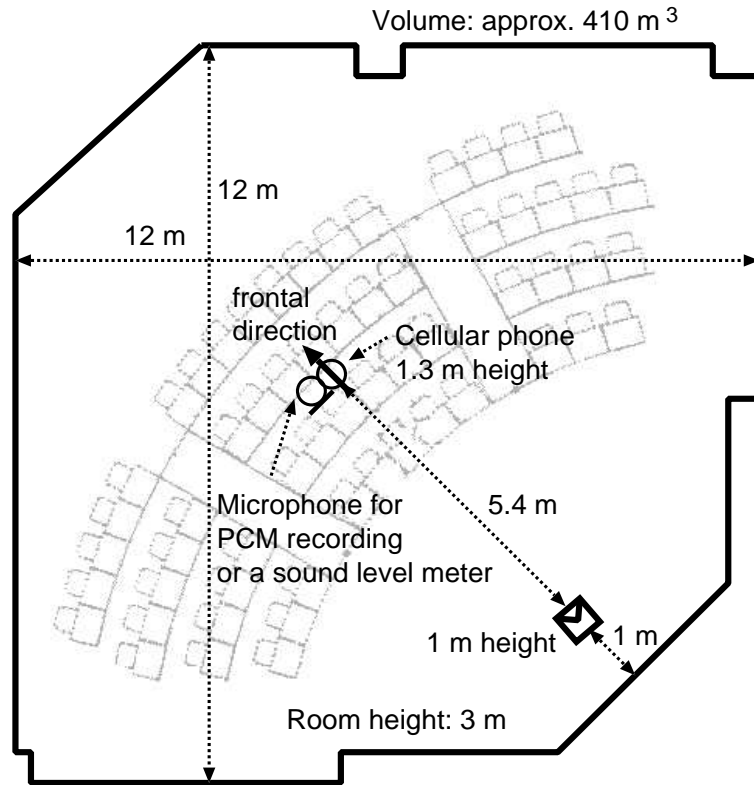


図 4.7: 実験に用いた部屋の見取図 .

は，上述の携帯電話に加えてその真横に無指向性コンデンサマイクロホン (audio-technica AT5410) を置き，USB オーディオユニット (EDIROL UA-5) を用いて 48 kHz, 16-bit の直線量子化にて録音を行った．

Log-TSP 信号をスピーカ再生して，コンデンサマイクロホンによって録音したインパルス応答からシュレーダ法 [64] により計算した残響時間は 1.0 秒であった．これら部屋と使用機器の配置図は，図 4.7 に示した．

4.4.3 実験結果

埋め込んだビット値に対して得られた正しいビット値の割合を検出率とした．図 4.8 には，22 名の音声信号での平均検出率を示した．エラーバーは，22 条件中の 10 から 90 パーセントの範囲を示している．図 4.9 には，100 曲の音楽信号に対する平均検出率を示した．エラーバーは，100 条件中の 10 から 90 パーセントの範囲を示している．

音声信号に対する結果の図 4.8 より，残響が無い場合は，SNR 20 dB 以下では，6.7 kbps の AMR コーデックを経ても 80 % 以上の検出率が得られた．残響が付加されると，検出率は大きく落ち，6.7 kbps の AMR 符号化では検出率が 80% を下回ることが多く，事実

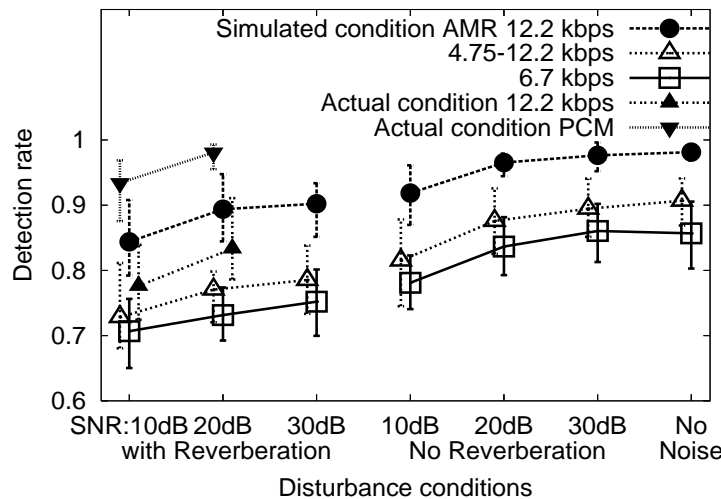


図 4.8: 22 名の音声信号に埋め込まれたデータの平均ビット検出率．エラーバーは 10 から 90 パーセントailsの範囲を示す．

上利用できないに等しい．しかし，12.2 kbps の AMR 符号化であれば，90%以上の音声信号で 85%以上の検出率が得られることが分かった．また図 4.9 より，音楽信号の方が，検出率が 90% を下回る条件での検出率が，音声信号の場合より 5 ポイント程度高いことも確認できた．これは，音声信号の成分が時間周波数的に粗い分布をしているのに対し，音楽信号は密に分布しているため，埋め込みの効率がよく変形に対して頑強になるのが理由である．また，現実の室内環境では，SNR が 20dB 以上でないとも，80% を上回る検出率を得るのは困難であることが分かった．

実環境において携帯電話にて録音した条件では，AMR ビットレートは 12.2 kbps であったが，シミュレーションでの同じビットレートよりも検出率が 0.06 程度低下した．これは音源に向かって携帯電話の画面を見ながらかざすという現実の使用環境を模して，携帯電話の画面と受話マイクロホンがスピーカと反対側を向くように配置したため，スピーカからの直接音が収録されにくく，周波数特性の劣化が大きかったためではないかと考えられる．無指向性マイクロホンにより PCM 録音した音から復号化した場合は，SNR 10 dB においても，90%以上の音声で，87%以上の検出率が得られたこと，またシミュレーション条件において AMR ビットレートによって検出率が大きく下がったことから，残響と背景雑音のある環境において音声符号化耐性を求めることは，かなり困難な要求であることが分かった．

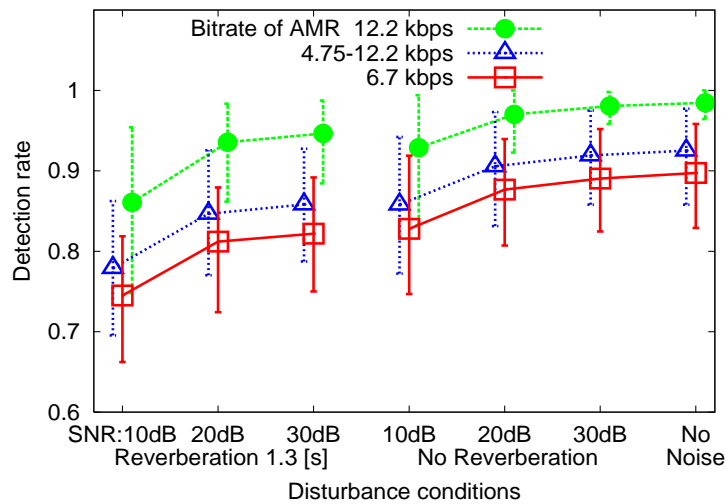


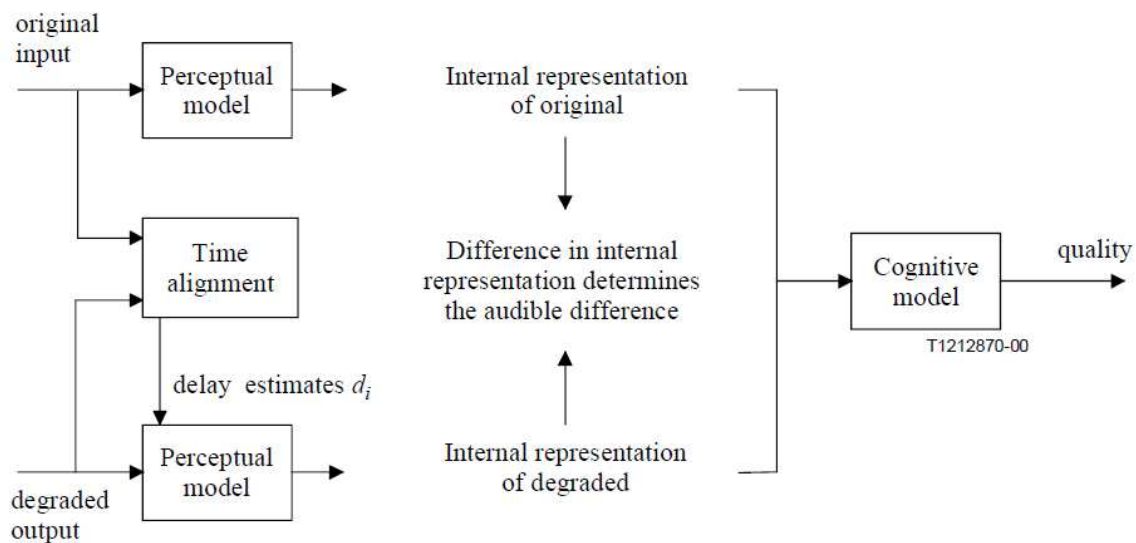
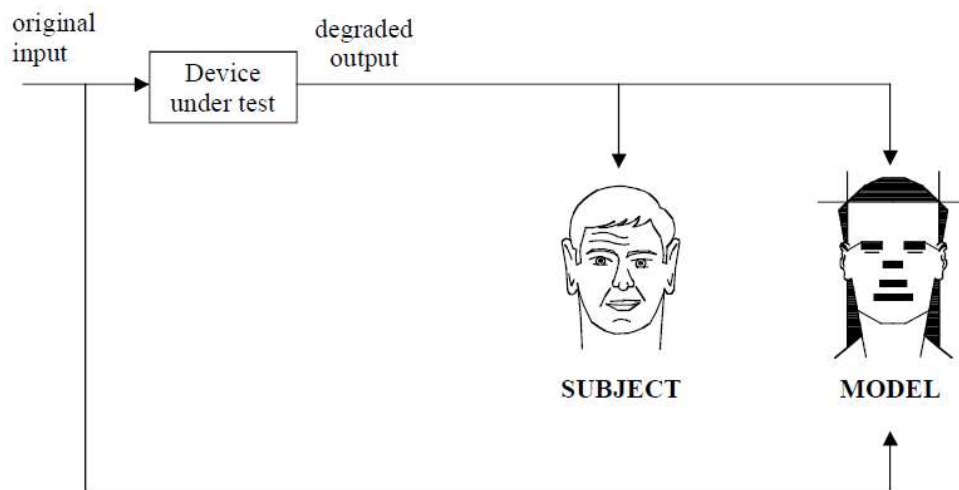
図 4.9: RWC-MDB-G-2001 100 曲に埋め込まれたデータの平均ビット検出率．エラーバーはそれぞれ 100 曲に対する検出率の 10 から 90 パーセントailsの範囲を示す．

4.5 客観的品質劣化度合評価

4.5.1 PESQ による広帯域音声品質劣化度合評価

ITU-T P.862 PESQ は，電話帯域音声や音声コーデックの品質劣化度合を測定するためのアルゴリズムである [68]．PESQ は原信号と音声コーデックを経た後の信号を比較し，心理音響特性に基づいた信号の内的表現の差分を，品質劣化度合として報告する．図 4.10 に，P.862 規格文書より，Figure 1 を抜粋して PESQ の概略図について示した．PESQ の結果は，MOS-LQO (Mean Opinion Score, Listening Quality Objective) とよばれ，人間を被験者として測定した主観的な劣化度合評価値である MOS (Mean Opinion Score) にほぼ対応する．MOS-LQO の値は，1.02 から 4.56 までが得られ，それぞれの値は，1: Bad (悪い)，2: Poor (劣っている)，3: Fair (まあよい)，4: Good (よい) という評価に対応する．

ここでは第 4.2 節にて行った，広帯域音声信号への情報秘匿に起因する音声品質の劣化を評価する．音声信号への情報秘匿後の音質は，音声符号化を経た音質と似ているので，その劣化度合を，サンプリング周波数 16kHz に拡張された Wideband PESQ を勧告している ITU-T P.862.2 に基づいた ITU-T 提供ソフトウェアにより測定した．Wideband PESQ への入力レベルは，16 bit 量子化における最大振幅の純音を 0 dB とした実効レベルを表す dBov を用いて，-26 dBov とした．日本音響学会研究用連続音声データベース Vol. 1 より，22 名の話者による各 50 合計 1100 の音素バランス文を 2 つずつ繋げて作成した 550 文の 8 秒前後の音声信号を対象とした．データ埋め込みパラメータは，サンプ



NOTE – A computer model of the subject, consisting of a perceptual and a cognitive model, is used to compare the output of the device under test with the input, using alignment information as derived from the time signals in the time alignment module.

図 4.10: PESQ の概要図 . ITU-T P.862 Figure 1 より抜粋 .

リング周波数が 16 kHz である以外は第 4.2 節のシミュレーションと同じとし，48 bps でデータを埋め込んだ。

結果は，図 4.11 に，CELP 系の音声符号化方式である AMR 符号化方式を広帯域 (16 kHz サンプリング) に拡張した符号化方式である，AMR-WB 方式において符号化した音声の MOS-LQO 値も併せて示した。この図から，振幅変調度 0.4 にてデータを埋め込んだ音声信号の品質劣化は，低ビットレート広帯域 CELP 系符号化音声より，やや音質が悪く，“劣っている”程度であろうと予測される。一方，同じ埋め込み強度 (振幅変調度 0.4) にてデータ埋め込みを行なった第 4.3 節では，VCV 音節の明瞭度試験を行っており，この場合 SNR 10 dB の条件でも平均で 93% の正答率となった。これらより，データ埋め込み済み音声信号の品質劣化は明らかなものの，音声情報を伝えるには十分であろうと考えられる。

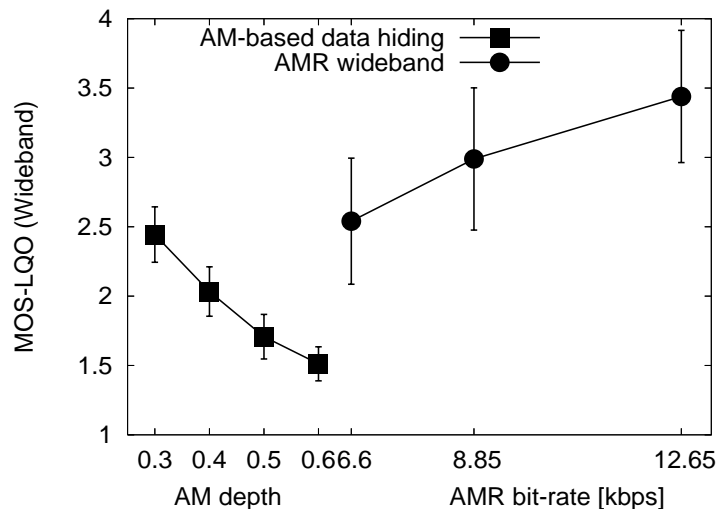


図 4.11: Wideband PESQ による広帯域音声客観的劣化評価値(MOS-LQO). 誤差棒は 550 条件の ± 1 標準偏差を示す. 左側は振幅変調に基づく情報秘匿に起因する劣化を変調度毎に，右側は AMR-WB 音声符号化による劣化を AMR ビットレート毎に表している。

4.5.2 PESQ による狭帯域音声品質劣化度合評価

第 4.4 節の耐性シミュレーション条件下での狭帯域 (8 kHz サンプリング) 音声品質の劣化度合を，ITU-T より提供されるソースコードをコンパイルした PESQ ソフトウェアを用いて測定した。日本音響学会研究用連続音声データベース Vol. 1 より，22 名の話者による 1100 の音素バランス文を 2 つずつ繋げて 550 文の 8 秒前後の音声信号とし，16 kHz

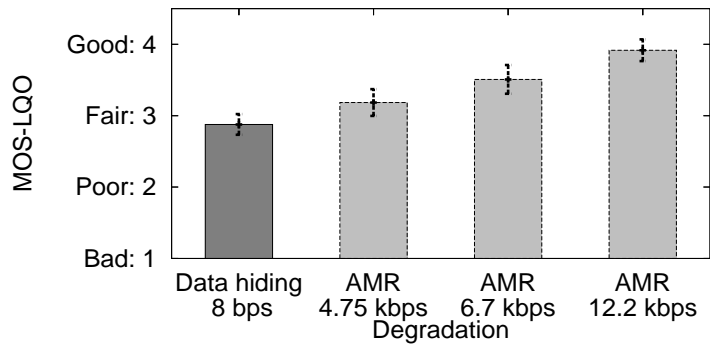


図 4.12: 狭帯域音声信号に対する AMR 音声符号化と情報秘匿に伴う音質劣化の比較．誤差棒は ± 1 標準偏差とした．

サンプリングのファイルを 8 kHz にダウンサンプリングした後，用いた．音声信号への情報秘匿条件は，表 4.2 に従い，ランダムビット値を埋め込んだ．AMR 符号化器への音声信号の入力レベルは -26 dBov とした．

結果は，図 4.12 に示した．音質劣化の比較参考のため，AMR 狭帯域音声コーデックによる 4.75, 6.7 および 12.2 kbps にて符号化および復号化した後の音声信号についても，MOS-LQO を測定してその平均と，誤差棒にて標準偏差を示した．この結果から，データを埋め込んだ狭帯域音声信号の品質劣化は，携帯電話よりやや音質が悪い程度であろうと予測される．

4.5.3 PEAQ による音楽音質劣化度合評価

ここでは，Kabal[14] による PEAQ の基本バージョンの実装を用いて，第 4.4 節で用いた情報秘匿済み音楽の音質劣化度合を測定した．音楽データは RWC-MDB-G-2001 の 100 曲左チャンネル冒頭 1 分間とした．使用環境が，環境騒音下のスピーカ再生を前提としており，かつ理想的なステレオ聴取環境を前提としないため，左チャンネルのみのモノラル信号を評価に用いた．44.1 kHz サンプリングの波形データに対し，4 kHz 以下に対して表 4.2 の条件にてランダムビット値の埋め込みを行った．

図 4.13 には，情報秘匿音楽の劣化度合と，比較対象として，MP3 の 48 kbps/ch (96 kbps), 64 kbps/ch (128 kbps) で符号化し復号化した音楽信号についての，音質劣化度合の平均値と ± 1 標準偏差の値をプロットした．この結果から，データ埋め込みに伴う音質劣化は，平均的には「劣化がやや気になる」よりやや悪い程度であることが分かった．また，MP3 によって符号化した音楽信号の劣化度合と比較すれば，48 kbps/ch (96 kbps) 程

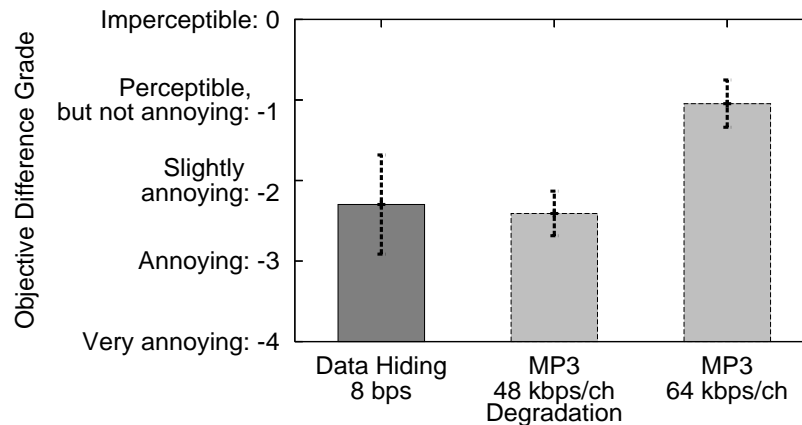


図 4.13: PEAQ による音楽信号の客観的音質劣化度合．誤差棒は ± 1 標準偏差．

度であることが分かった．情報秘匿済み音楽はスピーカ再生され，そこに背景雑音が加わることが必至である使用条件を前提とすると，データ埋め込みに伴う音質劣化は問題の無い程度であろうと考えられる．

4.6 考察

4.6.1 実時間処理埋め込みおよび検出処理

現在の情報埋め込みプログラムは，逐次フレーム処理を採用しているが，Octave プログラミング環境で作成しているため，ハードウェアの AD/DA デバイスを直接制御し，入力音響信号に対して実時間処理によりステゴ音響信号を出力することはできない．そのような実時間処理ソフトウェアを実装することにより，ライブコンサート PA やアナウンス音声，BGM への情報秘匿などに対して，幅広く技術の実施が可能となる．この点は，今後の課題である．

この場合，埋め込み強度は振幅変調度を固定して設定することとなり，理論的な最低遅延時間は，オーディオデバイスのバッファリング時間を除けば，フィルタバンク処理による遅延時間となる．表 4.1 の埋め込み条件では，2048 点 FFT を用いた FIR フィルタにより実装しており，この場合の遅延は約 100 ms となる．表 4.2 の場合は，1024 点 FFT を用いた FIR フィルタによって実装しているため，遅延は 128 ms となる．ライブコンサート PA の場合には，より低遅延が求められるため，フィルタバンク処理を改善する必要がある．一方，それ以外の場合には，遅延量は問題にならず，処理負荷はクロック 1GHz 程度のパーソナルコンピュータであれば充分であるので，実時間で情報秘匿が可能となる．

また、現在の秘匿情報の検出ソフトウェアでは、第 6.2.2 節において検討するように、FFT を多用している。よって、現在の検出アルゴリズムは、一般的な PDA やスマートフォン (例えば PXA270 Processor 520MHz) の処理能力の約 2 倍程度の演算能力を必要としている。多くの人々が携帯する機器への検出ソフトウェアの実装は、技術を実証し、その改善および普及のために必要と考えており、今後の課題である。

フィルタバンク処理を低演算量化するにあたっては、階層型 CIC (Cascaded Integrated Comb) フィルタ [69] を用いるのが有効であろう。また、検出時のフレーム同期のため演算を、フレーム周期を予測してその近傍の時刻 u のみ (3.13) 式の演算を行うこと、および埋め込み時の変調周期をオーバーラップ FFT 周期の整数倍とすることで式 (3.14) の変調周波数における強度算出を、変調周期毎の波形同期加算によって実施し FFT 処理を無くすなどの改良が考えられる。

4.6.2 携帯電話ネットワークにおけるパケットロスの影響

AMR コーデックは、伝送経路におけるパケットロスを隠蔽するような機構を必須要件として含んでいる。この主な仕組みは、伝送エラーが起きたフレームの前後のフレームのパラメータ値から補間を行なって、エラーフレームのパラメータを推定し、復号化するものである。

今回はこのパケットロス隠蔽のアルゴリズムは用いなかったが、AMR 符号化後のデータに対して、1 フレーム (20 ms) 単位でのフレームデータ抜き取り、ゼロデータフレームとの置換、ゼロデータフレームの挿入の 3 種類のパケットロスを等確率でランダムな時刻に起こすような変形を加えた後、復号化を行うシミュレーション実験も実施した。

その結果、3% 程度のパケットロスでは、いずれの条件でも 1~2% 程度の検出率の低下しか見られず、単純なパケットロスに対しては、ある程度の耐性を持つことが分かった。AMR コーデックによるパケットロスの隠蔽が行なわれた場合は、さらに検出率の低下は起こりにくくなるものと考えられる。

4.6.3 実効データ伝送量

第 4.4 節のシミュレーションでは、スピーカ伝送周波数特性、残響、背景雑音、音声コーデックの 4 つの妨害要因に対する耐性を持たせるため、データ埋め込み量を 8 bps と少なくした。ここではエラー訂正符号を埋め込み時に用いることは無かったが、このよう

な空間伝搬情報秘匿技術を実用化する際には、なんらかのエラー訂正符号を用いる必要がある。

ここで、8秒分のデータ(64bit)に対して、BCH(63,36,5)符号化と軟判定復号法[45]を併用したとすると、エラー訂正限界は9bit程度となる。この場合、86%のビット検出率が得られれば、36bit分の情報伝送が可能となり、実効データ伝送量は、4.5 bpsとなる。図4.8に示した音声信号へ残響を重畳したシミュレーションにおいて、約90%以上の信号条件で、86%のビット検出率を満たすのは、SNR 20 dB以上での12.2 kbpsのみであった。残響が無い場合は、6.7 kbpsのAMRコーデックによって、約90%以上の信号条件で80%のビット検出率を満たすことができたが、この場合はさらにエラー訂正限界を向上させる必要があり、実効データ伝送量は、上記の半分の2~3 bpsが妥当な線であろう。

データ埋め込み強度である振幅変調度を、今回用いた0.4から0.6に上げることによって、検出率は軒並5~6ポイント程度向上するが、狭帯域音声の場合、平均MOS-LQOは2.88から2.15まで低下する。一方、振幅変調度0.6で埋め込みを行なっても、音節明瞭度としては、SNR 10 dBの条件において平均で86%程度を得ており、文章了解度にはほとんど問題ないと思われる。よって、より困難な使用環境において、さらなる音質の劣化が許容できる場合には、振幅変調度を上げて頑強度を高めることも考えられる。

4.7 あとがき

データ埋め込み済み音響信号をスピーカから再生し、マイクロホンで受音してデータを検出し利用するという、空間伝搬利用を前提に、振幅変調に基づく音響情報秘匿技術の性能を検証した。

最初に、スピーカ拡声されるアナウンス音声にデータを埋め込み、利用者の手元の機器で復号化と表示を行う利用を前提とした。男女合計22名の広帯域日本語音声信号に対して48あるいは64 bpsにてデータを埋め込み、残響および背景雑音下での検出率をシミュレーション実験により調べた。その結果、48 bpsのデータを振幅変調度0.4で埋め込むと、90%の条件で84%以上の検出率が得られることが分かった。また、情報秘匿に伴う音声品質劣化は、広帯域PESQによる客観評価の結果、振幅変調度0.4で埋め込みを行うと平均的に“Poor(劣っている)”程度に劣化することが分かったが、VCV音節識別実験からは、SNR 10 dBの環境でも95%以上の明瞭度が得られることが分かり、発話内容を伝達するには問題ないことも明らかになった。

次に、スピーカ再生される音声や音楽にデータを埋め込んで、携帯電話の音声通話先にあるサーバコンピュータでデータを復号化し、利用者の携帯電話に情報を伝送する利用を前提とした。様々な狭帯域音声信号と広帯域音楽信号に 8 bps にて振幅変調に基づくデータ埋め込みを行なった音響信号が、残響と背景雑音そして AMR 狭帯域音声コーデックに対して耐性をもつかを調べた。その結果、音声信号に対しては、背景雑音のみが重畳される場合は、6.7 kbps 以上の AMR ビットレートにおいて、80% 以上のビット検出率が得られた。さらに残響が重畳する場合は、12.2 kbps の AMR ビットレートにおいて 80% 以上のビット検出率が得られた。広帯域音楽信号に埋め込んだ場合には、音声信号より 5 ポイント程度検出率が高いことが分かった。これらの結果から、残響や背景雑音が存在しても、AMR 狭帯域音声コーデックのビットレートが高ければ、品質を大きく劣化させずに振幅変調に基づく情報秘匿による携帯電話ネットワークを通じた情報伝送が可能であることが分かった。

また、この利用形態におけるデータ埋め込みに伴う客観的音質劣化度合を、サンプリング周波数 8kHz の電話帯域音声信号に対しては PESQ を用いて、広帯域音楽信号に対しては PEAQ を用いて調べた。その結果、音声信号は、「まあよい (fair)」よりやや悪く、音楽信号は「劣化がわずかに気になる (slightly annoying)」よりやや悪い程度であった。

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 ± 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， α 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 α 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 α 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 \sim で表し，DFT 演算を DFT ，逆 DFT 演算を 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_{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

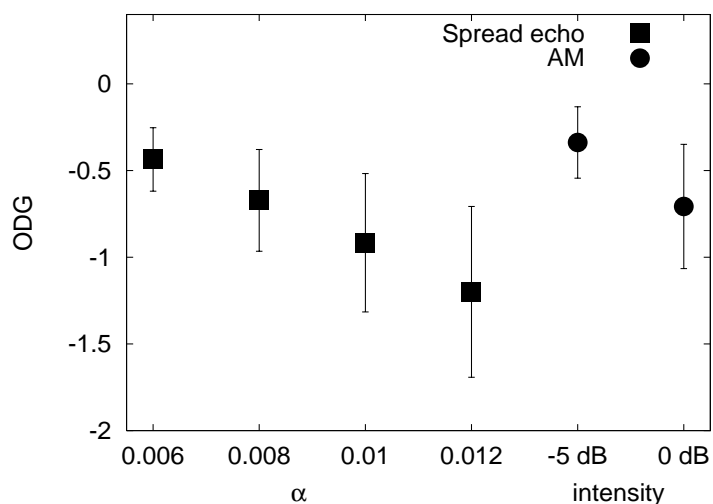


図 5.1: ステゴ音楽信号の ODG 平均値と ± 1 標準偏差 . RWC-MDB-G-2001 の 100 曲冒頭 1 分間より算出 .

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， \bar{x} は 90 パーセンタイル値， x_{10} 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．

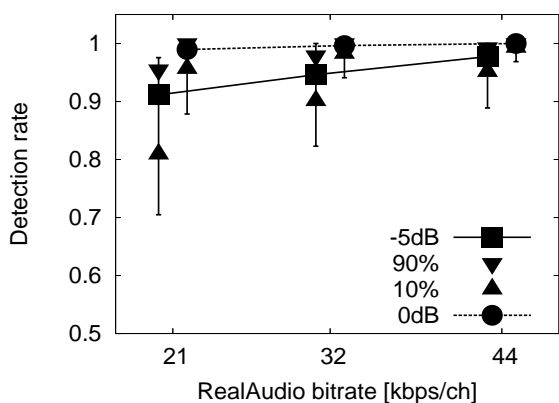


図 5.2: 振幅変調法により音楽信号に情報を埋め込んだ場合の RealAudio8 符号化および復号化後検出率。エラーバーは最大と最小値， ∇ は 90 パーセンタイル値， \triangle は 10 パーセンタイル値，真中の点は中央値を示している。

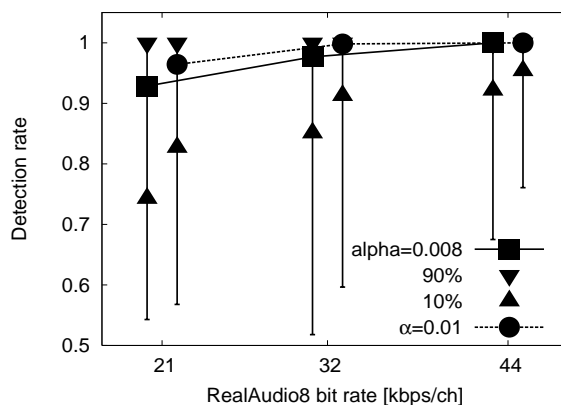


図 5.3: エコー拡散法により音楽信号に情報を埋め込んだ場合の RealAudio8 符号化および復号化後検出率。凡例は図 5.2 と同じ。

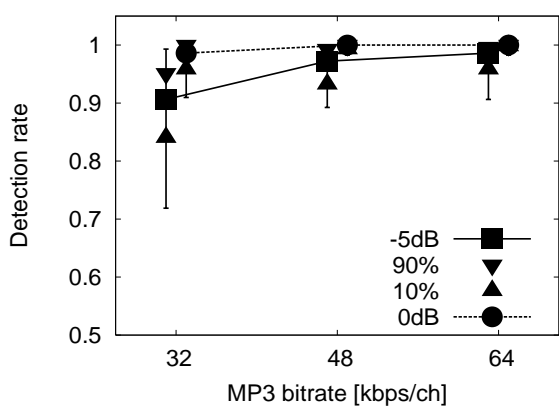


図 5.4: 振幅変調法により音楽信号に情報を埋め込んだ場合の MP3 符号化および復号化後検出率。凡例は図 5.2 と同じ。

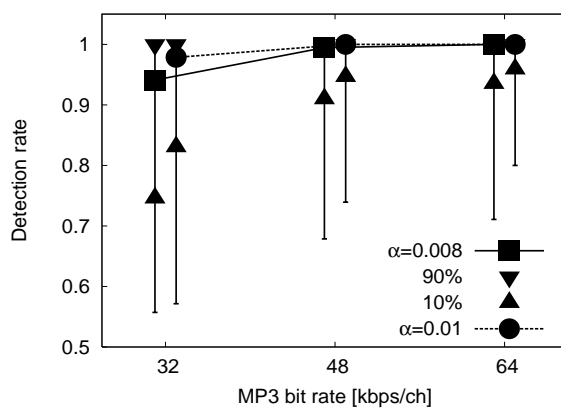


図 5.5: エコー拡散法により音楽信号に情報を埋め込んだ場合の MP3 符号化および復号化後検出率。凡例は図 5.2 と同じ。

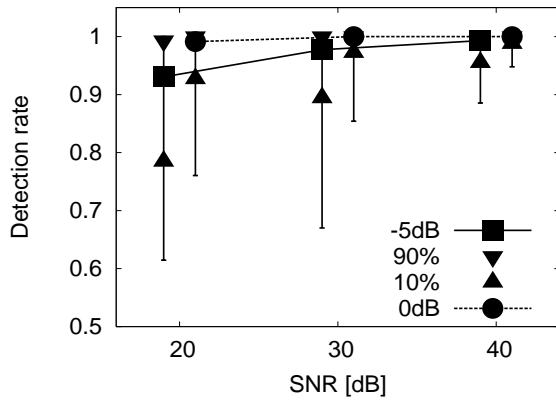


図 5.6: 振幅変調法により音楽信号に情報を埋め込んだ場合の雑音付加後検出率. 凡例は図 5.2 と同じ.

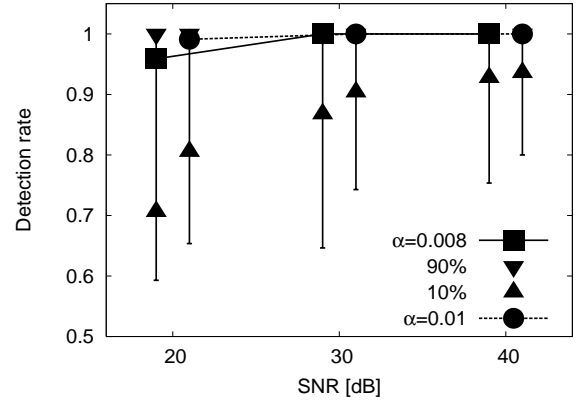


図 5.7: エコー拡散法により音楽信号に情報を埋め込んだ場合の雑音付加後検出率. 凡例は図 5.2 と同じ.

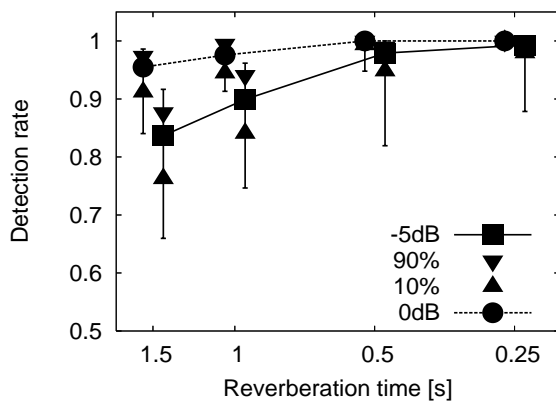


図 5.8: 振幅変調法により音楽信号に情報を埋め込んだ場合の残響付加後検出率. 凡例は図 5.2 と同じ.

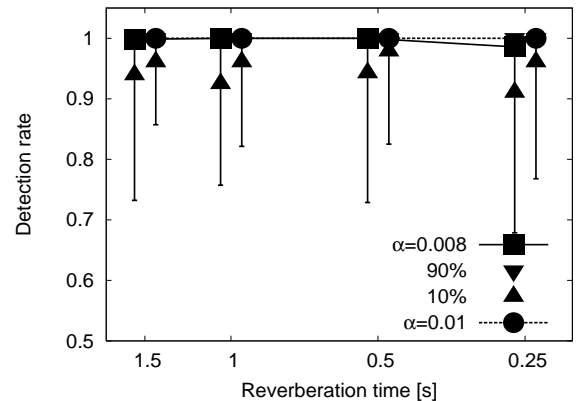


図 5.9: エコー拡散法により音楽信号に情報を埋め込んだ場合の残響付加後検出率. 凡例は図 5.2 と同じ.

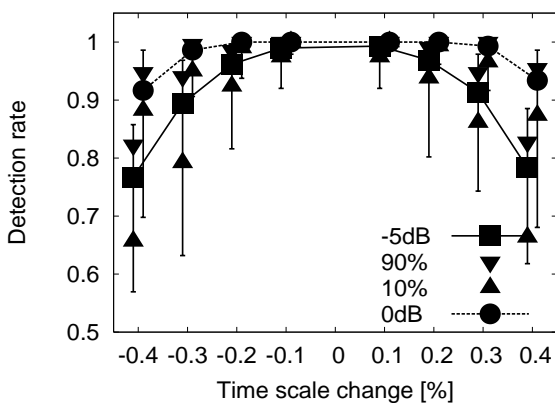


図 5.10: 振幅変調法により音楽信号に情報を埋め込んだ場合の時間変換後検出率．凡例は図 5.2 と同じ．

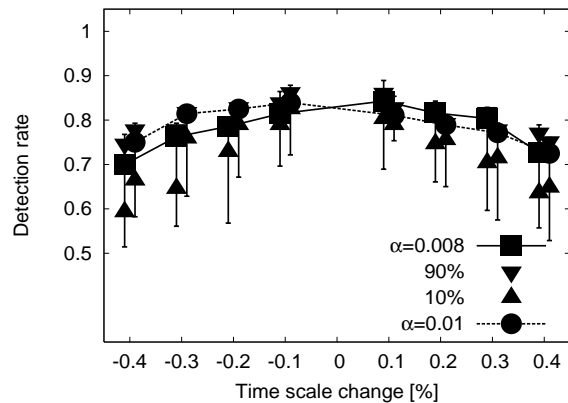


図 5.11: エコー拡散法により音楽信号に情報を埋め込んだ場合の時間変換後検出率．凡例は図 5.2 と同じ．

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，ランダム歪み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 α 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

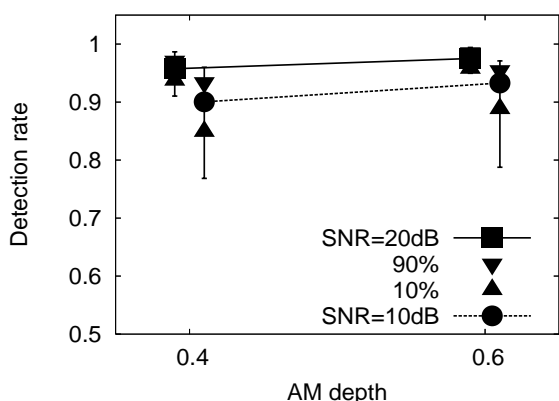


図 5.12: 振幅変調法により音声信号に情報を秘匿した場合の検出率．話者と背景雑音を組み合わせた 110 条件における，中央値と最大最小値，10 および 90 パーセントイル値が示されており，凡例は図 5.2 と同じ．

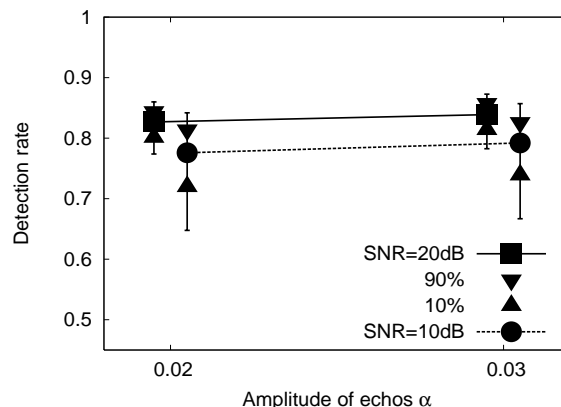


図 5.13: エコー拡散法により音声信号に情報を埋め込んだ場合の検出率．凡例は図 5.12 と同じ．

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。

第6章 情報秘匿に基づく音響信号と同期した情報呈示

6.1 まえがき

第4章では、スピーカから再生された情報秘匿済み音響信号を、ユーザの手元の機器で受信して検出復号化し、秘匿情報を利用する形態に対する、従来の音響情報秘匿技術やその具体的な利用提案例を概観した。この場合、音響情報秘匿技術に対しては、空間伝搬に伴う反射音や残響、背景雑音、スピーカやマイクによる周波数特性の歪などに対する耐性を保ちながら、電子透かし用途より埋め込む情報量を高めることが要求される。そして、振幅変調に基づく情報秘匿手法がそれらの要求を満たしていることを示した。

本章では、こうした情報秘匿済み音響信号の空間伝搬と受信を前提とした利用形態として、情報秘匿済み音響信号に同期してユーザへ情報呈示するシステムを提案する。具体的な応用として、カラオケ伴奏音楽に歌詞の呈示情報を埋め込み、伴奏音楽信号におけるデータの埋め込みフレーム時刻の同期検出を元に、歌詞の呈示タイミングに合わせて表示を行う。この技術は、カラオケだけでなく、映画における字幕情報の呈示など、音響信号と同期した情報の呈示/活用が必要な場面において有効である。そして、本システムの使用条件を想定した残響音、反射音、背景雑音、マイクロホンへの入力過大によるクリッピング歪、および歌唱声によるシステム性能の低下を、定量的に評価する。

従来の再生音と同期した情報呈示システム、例えば、伴奏音に歌詞を同期して表示する、業務用カラオケボックス、マイク一体型家庭用カラオケシステムなどにおいては、歌詞と伴奏音が同一あるいはケーブルで接続された機器によって表示および再生される形態をとっている。本システムは、情報秘匿済み音響信号の空間伝搬によって字幕呈示情報を伝えるため、再生機器と字幕表示機器とを空間的に独立させることが可能で、表示機器の複数化および個人化、同時多数利用、屋外利用も可能となる。これは、従来のカラオケや字幕システムを置き換えるものではなく、従来にない利用方法をもつシステムの提案とも言える。

表 6.1: 埋め込むレコードのデータ定義.

オジジャリ データ	相対開始 フレーム	開始 時刻	相対終了 フレーム	終了 時刻	情報の インデックス
--------------	--------------	----------	--------------	----------	---------------

6.2 音響信号と同期した情報呈示

本システムにおいて、音響信号に埋め込まれるデータは、主に呈示開始時刻と呈示終了時刻、そして呈示される情報のインデックスとなる情報であり、呈示される情報自体は事前に呈示システム内あるいはネットワーク上でアクセス可能な状態であることを前提としている。表示システムとしては、最終的には携帯電話や PDA が望ましいが、現時点では Windows OS 上で実装している。なお、本章で用いる音響信号は全て 44.1 kHz サンプリング、16bit 直線量子化されたものである。

6.2.1 埋め込むデータの構造

BCH エラー訂正符号を用いて符号化された 127 ビットあるいは 63 ビットの情報を、3 秒間のデータフレーム毎に埋め込む。BCH(127,29,21) を用いた場合、127 ビットあたり 21 ビットまでの符号誤りを訂正でき、伝送される情報量は 29 ビットとなる。この伝送情報量で 1 つの埋め込みレコードを構成し、表 6.1 のようにデータ定義する。レコードに含まれる各データ項目のビット長は、本システムの利用用途に応じて決定される。なお、1 データフレームあたりの埋め込み情報量は、BCH 符号化情報量より 1 ビット多いので、そのビットはパリティビット等に利用して信頼性を向上させることができるが、第 6.4 節以降のシミュレーションではこれを用いていない。オジジャリデータは機能拡張、例えば複数ビットを割り当てて字幕表示速度や表示位置を制御することを想定している。このような用途が無い場合には、使用しなくともよい。

相対開始フレームは、レコードを埋め込んだフレームに対する、表示を開始するフレームの相対的な位置を整数で表現する。つまりこのデータに 6 bit を割り当てれば、フレーム時間長 3 秒のとき、表示開始は最大で $3 \times 2^6 = 192$ 秒先まで可能になる。開始時刻は、相対指定されたフレーム中の表示開始時刻を示し、フレーム時間長を、2 のデータビット数の冪乗で割った分解能で指定する。つまり、フレーム時間長 3 秒の時に、開始時刻に 4 bit を割り当てれば最小の表示時間分解能は $3 \div 2^4 = 0.19$ 秒となる。相対終了フレームは、相対開始フレームに対して何フレーム先に表示を終了するかを指定する。終了時刻は

開始時刻と同じ定義である。情報のインデックスは、呈示するひとつの情報につき整数値を1つ割り当てる。

復号化されたレコードデータはバッファメモリに格納され、AD変換器から一定時間(ここでは4096サンプル)の波形データを取り込む度に、各レコードの表示開始時刻データと終了時刻データを走査して、表示開始と表示終了処理を行なう。

なお、表示する情報の数に対して埋め込みフレーム数が多い場合は、表示開始フレーム以前のフレームに表示に必要なレコードデータを繰り返し埋め込むことによって冗長度を高め、検出率を高めることも可能である。しかし、この方策は以降のシミュレーションでは用いていない。

6.2.2 検出処理演算量

検出時のフィルタバンク処理および埋め込み区間同期検出処理の演算にはFFTを利用しており、それらの演算負荷が検出処理全体の中で最も高い。第6.4節以降での実施例では、フィルタバンク処理は4096点FFTを3/4づつオーバーラップさせて実施し、同期検出処理はオーディオデータが4096サンプル入力されるごとに変調周波数毎の強度計算のため128点FFTを129回実施している。

携帯電話やPDAで検出復号化を行うには、演算量をさらに低減する必要がある。検出性能を大きく低下させずに演算量を低減するには、実際のマイクやスピーカの高域まで伸びていないことからサンプリング周波数を半分の22.05 kHzとし、フレーム境界検出精度がやや低下するものの、フィルタバンクにおけるFFTのオーバーラップを1/2とする方法がある。これにより単位時間あたり、フィルタバンク処理、同期検出処理ともに半分のサイズのFFTを用いて半分の実行回数で実現できる。しかし、この演算量でもWindows Mobile OS搭載の携帯電話あるいはPDAの処理能力の2倍程度となるため、フィルタバンク処理の演算量低減および、埋め込み時の変調周期をオーバーラップFFT周期の整数倍とすることで強度計算を同期加算のみで実施しFFT処理を無くすなどの改良が今後必要であろう。

6.3 データ埋め込みに伴う音質劣化の評価

情報秘匿済み信号をスピーカ再生とマイクロホン受音環境において利用する場合、より頑強なデータ検出のため、埋め込み強度を強めることになり、一般的な電子透かし用途の

表 6.2: データ埋め込み条件

Parameters	Normal	Robust
sampl. freq.	44100 Hz	←
embedding freq. region	≤ 11025 Hz	≤ 22050 Hz
bandwidth	86 Hz	←
subband pairs	64	127
subband groups	17	9
pilot subbands	12 pairs	36 pairs
pairs per group	3 — 4	11 — 12
frame period	3 s	←
mod. freq. [Hz]	1.67, 2.0, 2.33, 3.0	←
intensity	+12 dB	+15 dB
payload [bits/frame]	128	64
error correction	BCH(127,29,21)	BCH(63,18,10)
bit rate [bps]	9.7	6

情報秘匿に伴う音質劣化に比べて知覚可能な程度となる。

ここでは, Kabal[14] による PEAQ の基本バージョンの実装を用いて, 情報秘匿済み音楽の音質劣化度合を測定した。対象音楽は, RWC ポピュラー音楽データベース (RWC-MDB-P-2001)[72] に含まれる 100 曲の左チャンネル冒頭 60 秒とした。また音質劣化の比較対象として, 情報埋め込みを行わずに MP3 符号化後に復号化した場合も, 同様に測定を行なった。

表 6.2 には 2 つの埋め込み条件が示されている。Normal は, 一般的な雑音や残響のある環境で利用する場合であり, Robust は, 第 6.5 節で示す, カラオケでの歌詞表示利用を前提として残響や歌声が重畳する環境で利用する場合の埋め込みパラメータ値である。表 6.2 中の mod. freq. は変調周波数であり, 全ての帯域に対して 4 つの変調周波数で同時かつ独立に変調を行い, 埋め込み処理を多重化していることを意味する。

図 6.1 には, 2 つの埋め込み条件の結果と, MP3 で圧縮した音楽についての, 音質劣化度合を表す ODG の平均値と ± 1 標準偏差の値を示した。この結果から, 情報秘匿に伴う音質劣化は, 平均的には「やや気になる (slightly annoying)」以下程度であることが分

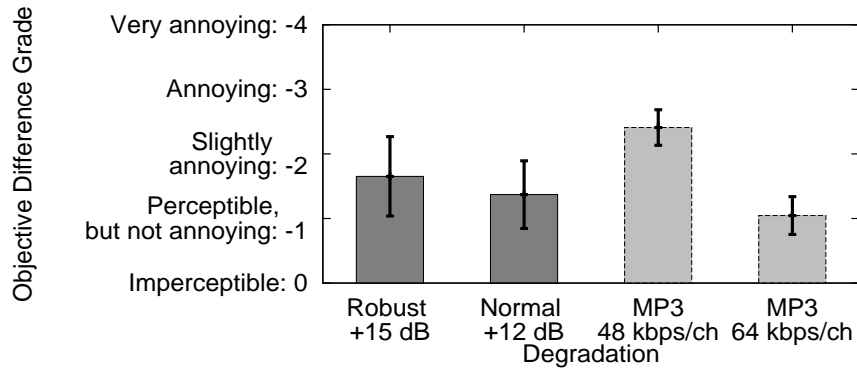


図 6.1: PEAQ によって得られた客観品質劣化度合 (ODG) の平均と ± 1 標準偏差 . 対象は RWC-MDB-P-2001 に収録された 100 曲 .

かった . また , MP3 と比較すれば , 48 kbps/ch と 64 kbps/ch の中間程度の音質であることが分かった .

6.4 空間伝搬耐性

本システムが , 残響あるいは入力過大 , かつ雑音に対する耐性を持つことをシミュレーションによって確認する . 埋め込み条件は , 表 6.2 における Normal 条件を用いた .

6.4.1 シミュレーション条件

対象楽曲は RWC-MDB-P-2001 に含まれる 100 曲の左チャンネル冒頭 60 秒とした . これに , 第 4.2 節で用いたものと同じ , RWCP 実環境音声・音響データベース [63] 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 . このインパルス応答波形は既に図 4.2 に示した . この波形の絶対値ピークを中心とした 128 サンプルにハニング窓掛けを行って直接音とみなした場合 , 直接音に対するそれ以降の間接音のパワーの比率は -0.4 dB となる . この直接音成分のパワースペクトルは既に図 4.3 に示した . この畳み込みによって , 再生機器の位相特性や反射音による位相特性の乱れも模擬できる . 図 6.2 には , インパルス応答波形について 110 サンプル毎に求めた実効レベルの時間減衰を示した .

その後 , 第 4.2 節で用いたものと同じ , 背景雑音として 4 種類の環境騒音 (収録場所: 駅のホーム , 地下連絡通路 , 空港ロビー , 混雑した交差点) , あるいはローパスノイズ (カットオフ 500 Hz , -9 dB/oct. : 他の環境騒音の平均的スペクトルに近い) のいずれかを付加

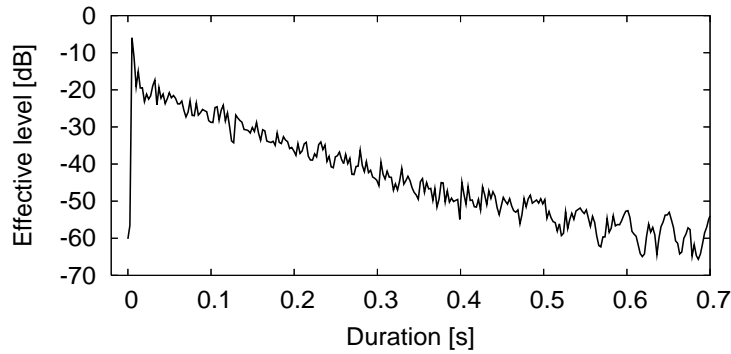


図 6.2: インパルス応答におけるエネルギー減衰.

した後、埋め込み情報を検出する処理を行った。5種類の背景雑音は、オーバーオールの音楽信号パワーに対して、信号対雑音比 (SNR) を 15 dB とした。雑音強度の時間変動は、100 ms の矩形窓毎に計算した実効値の標準偏差が、それぞれ 4.6 dB, 1.9 dB, 3.3 dB, 4.0 dB, 0.49 dB であった。これらの背景雑音は、本システムを屋内で利用する場合の雑音とはやや種類が異なるうえレベルも高いが、より困難な環境を模擬するために用いた。

また、検出用のマイクロホンスピーカに近づけた場合は、入力過大により振幅がクリッピングする事態も考えられる。このような状況での耐性を調べるため、データ埋め込み後の音楽信号について、その最大振幅の 0.125 倍以上の振幅を制限する変形 (+18 dB の入力過大) も模擬した。この場合残響は付加せず、環境騒音の SNR は 30 dB として、振幅制限の直前に加えた。

情報秘匿済みの 100 種の音楽信号と 5 種の背景雑音を組み合わせた 500 条件において、上記のシミュレーションを行った。

6.4.2 フレームデータ検出率

データ検出率の指標としては、3 秒間のデータフレームに埋め込まれたランダムに 1 あるいは 0 の値を持つ 128 ビットのうち、誤り訂正限界である 21 ビット以内にエラービット数が収まったフレーム数を、全体のフレーム数で割った正検出割合とした。

図 6.3 に残響と背景雑音付加の結果を、図 6.4 に背景雑音と振幅制限を与えた結果を示した。データ検出率 (Correct Detection Rate) は、音楽の冒頭から数えたフレーム番号 (Frame Number) に依存するため、横軸をフレーム番号として示している。残響が加わった場合は、楽曲の冒頭はいわゆるイントロであり音楽信号レベルが相対的に小さいため、検出率が低くなっているが、第 5 フレーム (15 秒) 以降では、90%以上の検出率となって

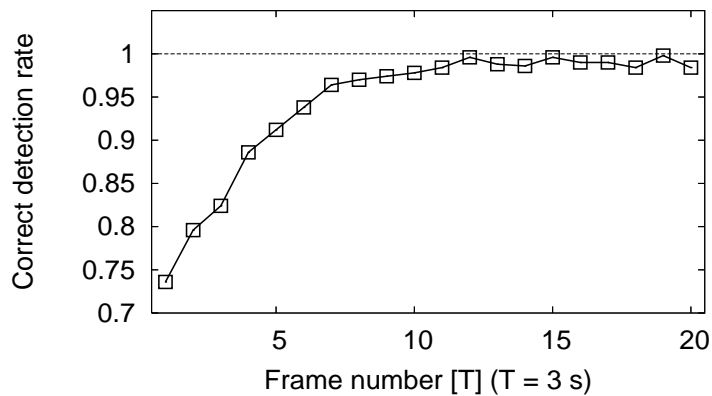


図 6.3: 残響と背景雑音が加わったときの，エラー訂正限界以内に収まったデータフレームの割合．それぞれの点は 500 条件のシミュレーションから得られた．

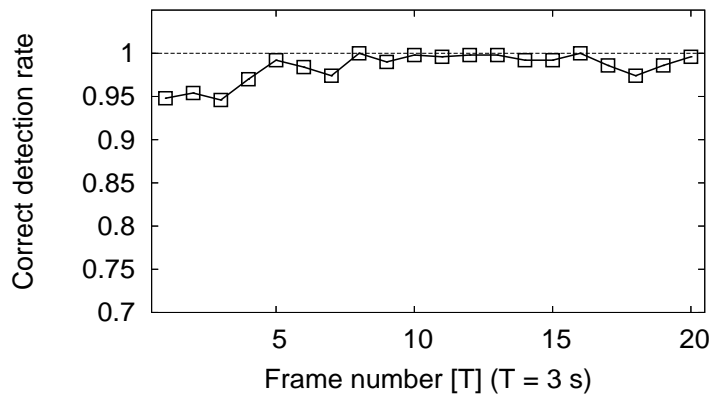


図 6.4: 背景雑音と振幅クリッピングが加わったときの，エラー訂正限界以内に収まったデータフレームの割合．それぞれの点は 500 条件のシミュレーションから得られた．

いる．振幅制限の場合には，より良好に検出ができることが分かった．

6.5 カラオケ歌詞表示システム

ここでは，データ埋め込み済み音響信号に同期してユーザへ情報呈示するシステムの応用として，カラオケ伴奏音楽に同期して歌詞を呈示するシステムについて説明する．伴奏音楽には歌詞情報そのものではなく，そのインデックスを埋め込む．歌詞情報本体を呈示システムへ取り込む方法としては，インターネットを経由した楽曲の販売時に付帯的なダウンロードとして実現するか，カラオケトラック入り音楽 CD 販売時にダウンロードのキーを同梱することなどが考えられるが，この仕組みは現時点でのシステムには含まれない．

表 6.3: カラオケ歌詞表示のために埋め込むレコードのビット割り当て.

相対開始 フレーム	開始 時刻	相対終了 フレーム	終了 時刻	表示歌詞の インデックス
3 bit	4 bit	2 bit	4 bit	5 bit

このシステムの性能評価シミュレーションは、Octave-3.0 上で動作するソフトウェアで実施した。実時間で動作するシステムは、当初は MATLAB R2008 にて Data Acquisition ToolBox, Signal Processing ToolBox, Communication ToolBox を用いて作成した。その後、Windows バイナリとして動作するシステムを委託製作した。Windows 版の動作中の画面の例は、図 6.5 にて示した。図中のスピーカおよびマイクの距離は、実際には、より遠くに配置しても動作することは確認している。

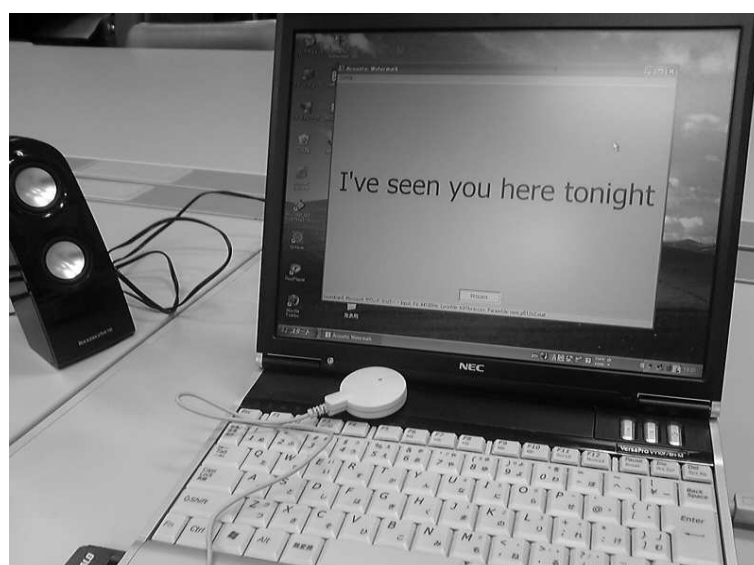


図 6.5: Windows 版同期表示ソフトウェアの実行中画面の例。

6.5.1 埋め込むデータ構造

表 6.2 中の Robust 条件を採用した。この条件では、63 ビットあたり 10 ビットまでの符号誤りを訂正できる一方、伝送される情報量は 18 ビットであり、表 6.3 のようにビットを割り当てた。

表 6.3 のビット割り当てから、レコードを埋め込んだフレームに対する表示を開始するフレームは、最大で 24 秒先まで可能となる。同様に開始時刻の表示時間分解能は 0.19 秒

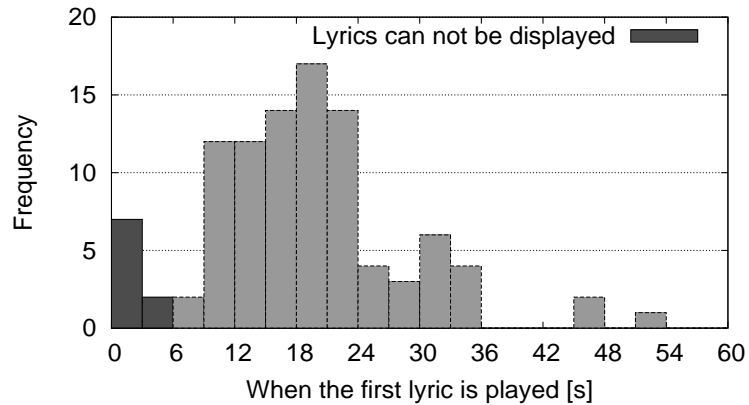


図 6.6: RWC-MDB-P-2001 収録の 100 曲の歌詞の開始時刻と度数.

(テンポ 160 での 8 分音符) となる．表示時間は最大で $3 \times 2^2 = 12$ 秒間である．表示歌詞数は $2^5 = 32$ であるが，リフレインなどの同じ歌詞には同じインデックスを割り当てることとする．

6.5.2 データ埋め込みと同期

1 つの歌詞表示情報を埋め込む最小時間間隔はデータフレーム時間長であるが，演奏には歌の無い部分もあるため，フレーム時間長を 3 秒としても，楽曲中のフレーム数は表示する歌詞の数より数倍程度多いのが一般的である．途中から伴奏を再生したときにも歌詞が表示できるように，歌詞表示の直前の複数フレームにデータを埋め込み，冗長性を確保するのが良いだろう．実用的には，伴奏のどこに歌詞表示データを埋め込むかを最適に設計する埋め込みシステムの構築が望ましい．

データフレームの同期検出には，2 フレーム (6 秒) 分の伴奏時間が最低必要である．3 フレーム目以降では，検出されたフレーム境界を遡って，最初のフレームからデータの復号化と表示を行う．よって 3 フレーム目より以前に歌詞の表示は不可能であるが，このような楽曲がどの程度の割合で存在するかは明らかではない．図 6.6 に，RWC-MDB-P-2001 収録の，歌のあるポピュラー音楽 100 曲について，曲の開始から何秒後に歌詞が唄われるかの頻度 (Frequency) を示した．この結果から，1 割程度は冒頭の歌詞が表示できない場合があると思われる．

6.5.3 性能評価

歌詞表示システムの性能評価として、室内の反射や残響、および利用者の歌唱音声による影響を、コンピュータシミュレーションによって調べた。性能評価の対象楽曲は RWC-MDB-P-2001 に含まれる 100 曲の MIDI ファイルを、MIDI 音源 (timidity++) によって演奏した左チャンネル冒頭 60 秒とした。

残響のある室内を想定して、データ埋め込み済み音楽信号に、RWCP 実環境音声・音響データベースに収録されているインパルス応答 (ファイル名: ir130.dat) を畳み込んだ。次に歌声合成ソフトウェアである VOCALOID2「鏡音リン・レン」を用いて作成した歌唱音 (男声はレン、女声はリンが歌唱) を、平均 SNR $-5, 0, +5$ dB の 3 条件で付加した。この平均 SNR 値は、A 特性フィルタリング後の残響つき伴奏音の実効値 (signal) と、A 特性フィルタリング後の歌唱音の実効値 (noise) を計算し、100 ms の矩形窓フレーム毎に歌唱音が含まれるフレームの SNR(dB) 値の算術平均値として定めた。

データ検出率の指標としては、3 秒間のデータフレームに埋め込まれたランダムに 1 あるいは 0 の値を持つ 64 ビットのうち、誤り訂正限界である 10 ビット以内にエラービット数が収まったフレーム数の率を調べた。なお、歌唱音が含まれないデータフレームに対しては、全てのフレームでエラー訂正限界以内に収まった。

歌唱音がわずかでも含まれるフレームは、全 2000 フレーム中 1369 であった。歌唱音が含まれるフレーム数を母数としたデータフレーム検出率 (Correct detection ratio for singing frames) を、各曲毎に 0.05 刻みで調べた。図 6.7 の横軸は検出率であり、縦軸はその検出率以下となる累積曲数 (Cumulative frequency of tunes) を示した。図 6.8 には、1 曲の全フレーム数 20 を母数とした、データフレーム検出率 (Correct detection ratio for all frames) を各曲毎に算出し、その検出率以下となる累積曲数を示した。

図 6.7 では SNR -5 dB の最も厳しい条件で歌唱音が重畳した場合には、100 曲中 12 曲で歌唱音の含まれるフレームの検出率が 0.7 以下になることを示している。一方、歌詞表示情報の埋め込みは歌詞のあるなしに関わらず行われるため、曲全体の歌唱音のないフレームも含んだ場合の検出率を見ると、図 6.8 では、SNR -5 dB のとき、0.7 以下の検出率となるのは 6 曲である。実際の楽曲中のフレーム数は表示する歌詞の数より数倍程度多いのが一般的であり、同じ歌詞表示情報を複数フレームに繰り返し埋め込むことができるため、実際のデータ検出率はさらに高まるはずである。なお、歌唱音重畳時の検出率が低い楽曲は、伴奏がピアノのみ、あるいはギターのための曲であった。

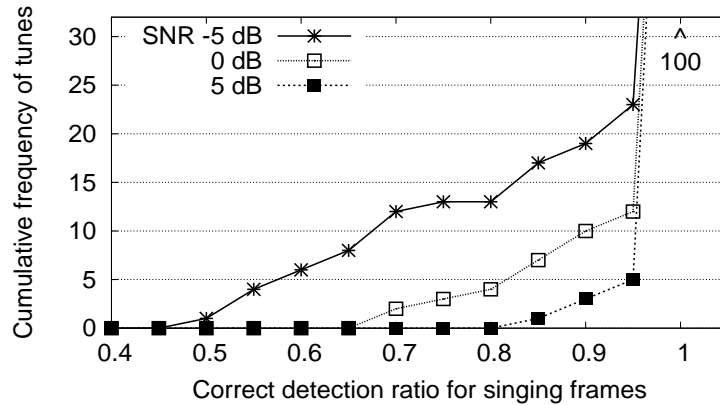


図 6.7: 歌唱音が含まれるフレームに対するデータ検出率と累積曲数.

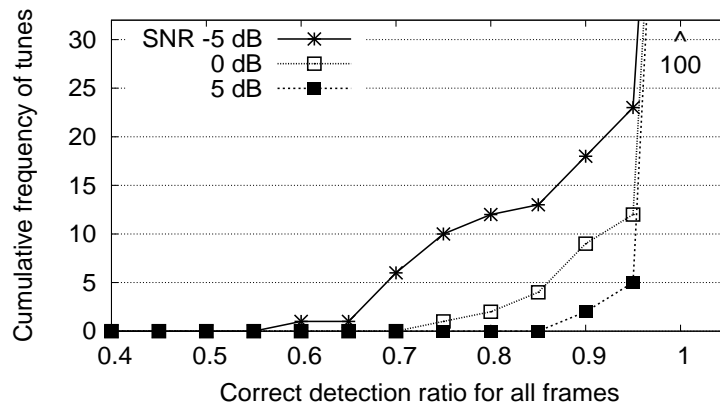


図 6.8: 曲の全フレームに対するデータ検出率と累積曲数.

データ検出時に得られるフレーム境界時刻は、伴奏音と歌詞表示の同期の基準時刻となる。このフレーム境界時刻の検出精度を、正しいフレーム境界からの平均ずれ時間 (Mean deviation from the correct frame boundary) として、最も厳しい条件である SNR -5 dB での結果を図 6.9 に示した。

第 4 フレーム目 (冒頭から 12 秒) までは、検出されたフレーム境界は平均で 0.1 秒程度早めとなるが、それ以降は 0.1 秒前後の標準偏差でほぼ正確にフレーム境界の検出が可能であった。また、表示における時間分解能は約 0.19 秒であるので、実用上ほぼ問題無く指定時刻に表示できるであろうことが分かった。

6.6 考察

スペクトル拡散法などの一部を除いて、音響電子透かし技術の多くは、本技術もそうであるが、音の無い部分にデータを埋め込むことはできない。本システムの用途としては、

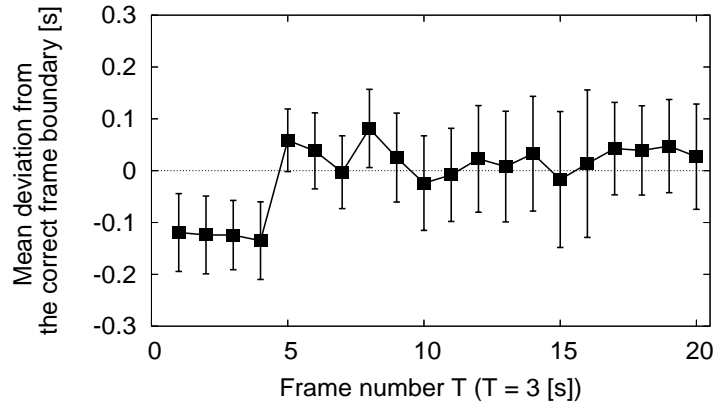


図 6.9: 残響と歌声が加わった場合での、フレーム境界時刻からの、検出時刻のずれ. 誤差棒は ± 1 標準偏差を示す.

カラオケ歌詞表示以外にも、映画のサウンドトラックに字幕表示情報を埋め込んで、セリフに同期して字幕を表示するシステムが考えられる。これは、前席の背もたれ背面にパネル画面 PC を掛け、字幕と画面への視線と同方向となる利用形態である。この場合、映画サウンドトラックには、無音部分が多く含まれるという問題がある。これには、データ埋め込みフレーム時間長を長くとの対応が考えられる。本技術では、データフレーム時間の一部にしか信号が含まれていなくても、その部分を有効に活用してフレーム同期とデータ検出を行うことができる。このことは、第 4 章において示したように、ポーズを含んだ音声信号でも、本技術は有効にデータ検出が可能であることから分かる。

本技術をカラオケ歌詞表示システムとして利用する場合、楽曲の冒頭部分の歌詞は表示できず、検出表示端末の設置位置によっては第 6.5.3 節で想定した以上の高雑音および残響の多い環境となることも考えられる。よって、性能的に従来技術を置き換えることは困難である。一方、ここではシステムへの妨害となる環境の評価を定量的に行ったため、使用環境に応じた表示不能率を算出することが可能である。例えば、検出率 x の楽曲および環境において、 n 個の異なる情報を m 回繰り返して埋め込む場合、最大 1 つまでの表示情報欠落を許容するときの表示可能率は $(1 - (1 - x)^m)^n + n(1 - (1 - x)^m)^{n-1}(1 - x)^{(n+m)}$ で与えられる。再生系と情報表示系の空間的分離や、表示端末の個人化、同時多数端末による表示など、従来のシステムにない機能に価値をおき、使用条件の制限となるシステム動作環境と上述の表示可能率とのバランスを考えながら実使用に望むべきであろう。

6.7 あとがき

振幅変調に基づく音響情報秘匿技術を用いて、データを埋め込んだスピーカ再生音をマイク受音して、その信号に同期して情報を呈示するシステムを作成した。システムの性能評価として、情報秘匿済み音楽信号に対してPEAQを用いた音質劣化の客観評価を行った結果、音質劣化はMP3 48 kbps/ch と 64 kbps/ch で符号化された音楽信号の中間程度であることが分かった。情報秘匿済み音楽信号への残響および背景雑音付加、あるいは入力過大による振幅制限のシミュレーションの結果、音楽信号に埋め込まれたデータは十分検出可能であることが分かった。

さらに、カラオケ歌詞表示システムとして利用時の、埋め込みデータ検出への残響や歌唱音の影響も調べた。その結果、SNR 0 dB で歌唱音が付加されても、90%以上の楽曲の、90%以上の歌唱音を含むフレームでデータ検出が可能なが分かった。また、歌詞表示の時間制御の基となるデータフレーム境界時刻の検出精度も十分であることが分かった。

第7章 結論

本論文は、埋め込み前の信号を検出時に必要としない、ブラインド検出を可能とする帯域分割と振幅変調に基づいた新しい音響情報秘匿技術を提案し、その性能を検証した上で、技術の新しい利用法を示し評価することを目的とした。

第1章では、音響情報秘匿技術について概観し、技術への要求点、すなわちブラインド検出、低音質劣化、秘匿性、耐性(頑強性)、適用性、大容量、空間伝搬耐性を挙げ、従来の研究においては、技術への幅広い要求点を満たしているかどうかの検証が少ないことを説明した。

第2章では、音響情報秘匿技術に関して、用いられる用語や概念などをはじめに説明し、技術への要求とそれに対応した評価方法を示した。次に、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解説し、その特徴を簡単にまとめた。さらに、性能向上のため補助的に用いられる技術も説明した。最後に、電子透かしとしての技術利用の現状について示した。

その結果、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ことを主張した。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(PEAQ)は、知覚符号化音響信号を対象としており、情報秘匿に伴う音質劣化度合を評価できるかどうかについては明らかではないことも主張した。そして、これまで提案されてきた多くの音響情報秘匿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ことを述べた。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、耐性、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と述べた。

第3章では、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動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。そして、音響電子透かし用途を前提として、RWC 音楽ジャンルデータベース [13] 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 bps 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、この情報秘匿手法はピッチ変換に対して脆弱であった。

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、もっとも検知しやすい楽曲において -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原音に対する音質変化を検知できる被験者 5 名を対象として、ITU-R BS.1116-1 に準拠して行った音質劣化度合を評価する実験によって得られた主観差分等級 (SDG) より、-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。また、-5 dB の情報埋め込みに伴う音質劣化度合は“違いが分かるが気にならない”程度であることが分かった。

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を行うため、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 SDG 値は ODG 値より大きく外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。

第 4 章では、前章で示した技術を、残響と背景雑音が重畳する空間伝搬条件において利用することを検討した。まず、スピーカ拡声されるアナウンス音声にデータを埋め込み、

利用者の手元の機器で復号化と表示を行う利用を前提とした。そのため、男女合計 22 名の広帯域日本語音声信号に対して 48 あるいは 64 bps にてデータを埋め込み、客観音声品質劣化度合と、残響および背景雑音下での検出率をシミュレーション実験により調べた。

その結果、48 bps のデータを振幅変調度 0.4 で埋め込むと、SNR 10 dB の場合でも 90% の条件で 84% 以上の検出率が得られることが分かった。また、情報秘匿に伴う音声品質劣化は、広帯域 PESQ による客観評価の結果、振幅変調度 0.4 で埋め込みを行うと平均的に“Poor (劣っている)”程度に劣化することが分かったが、VCV 音節識別実験からは、SNR 10 dB の環境でも 95 % 以上の明瞭度が得られることが分かり、発話内容を伝達するには問題ないことも明らかになった。

次に、埋め込み情報の検出を、マイクロホンにより受信した端末で行うのではなく、携帯電話の音声通話により接続する携帯電話音声ネットワークの先にあるサーバコンピュータにおいて実行するため、携帯電話の音声符号化を経ても情報伝達が可能かどうかを調べた。対象は、音声データベースおよび音楽ジャンルデータベースの楽曲 100 曲であり、8 bps にて情報埋め込みを行った場合のシミュレーション実験および実室内環境における実験を行った。また、この利用形態におけるデータ埋め込みに伴う客観音質劣化度合を、サンプリング周波数 8kHz の電話帯域音声信号に対しては PESQ を用いて、広帯域音楽信号に対しては PEAQ を用いて調べた。

その結果、音声信号に対しては、SNR 20 dB 以上の背景雑音のみが重畳される場合は、6.7 kbps 以上の AMR ビットレートにおいて、90% 以上の音声信号に対して 80% 以上のビット検出率が得られた。さらに残響が重畳する場合は、12.2 kbps の AMR ビットレートにおいて 90% 以上の音声信号に対して 85% 以上のビット検出率が得られた。また、音楽信号の 4kHz 以下に埋め込んだ場合には、音声信号より約 5 ポイント程度検出率が高いことが分かった。客観音質劣化度合は、音声信号は「まあよい (fair)」よりやや悪く、音楽信号は「劣化がわずかに気になる (slightly annoying)」よりやや悪い程度であった。これらの結果から、残響や背景雑音が存在しても、AMR コーデックのビットレートが高ければ、品質を大きく劣化させずに振幅変調に基づく情報秘匿による携帯電話音声ネットワークを通じた情報伝送が可能であることが分かった。

第 5 章では、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による情報秘匿手法 [15] について、シミュレーション実験によってその性能を比較した。電子透かし用途としての性能を比較するために、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が得られる埋め込み

強度にて、それぞれの手法により情報埋め込みを行った。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再び行った。

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ことが分かった。

また、音声信号に対して 48 bps の情報を秘匿し、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環境をシミュレーションし、双方の手法を比較した。その結果、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、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、マイクロホン受信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。

第 6 章では、振幅変調に基づく音響情報秘匿技術を用いて、スピーカから再生される音に同期して情報を呈示するシステムを提案した。ここでは、埋め込む情報にブロック符号化によるエラー訂正を施し、残響および付加雑音環境下において、エラー訂正の範囲に収まる正検出率をシミュレーション実験により求めた。また、カラオケの伴奏音楽に歌詞を同期呈示するための情報を埋め込んでおき、スピーカ再生される伴奏音楽に同期してリアルタイムに歌詞を表示させるシステムを構築した。そして、情報の同期呈示に必要な埋め込み情報検出における時間精度と、埋め込みデータ検出への残響や歌唱音の影響も調べた。

その結果、伴奏音楽に対して背景雑音および振幅クリッピング歪あるいは残響の重畳を経ても、音楽信号に埋め込まれたデータは十分検出可能であることが分かった。また、SNR 0 dB で歌唱音が付加されても、90% 以上の楽曲の、90% 以上の歌唱音を含むフレームでデータ検出が可能なが分かった。また、歌詞表示の時間制御の基となるデータフレーム境界時刻の検出精度も十分であることが分かった。PEAQ を用いた客観音質劣化評価の結果、伴奏音楽の音質劣化は MP3 48 kbps/ch と 64 kbps/ch で符号化された音楽信号の中間程度であることが分かった。

第 7 章では、本研究の成果と今後の課題をまとめた。

本研究で提案した振幅変調に基づく音響情報秘匿技術に対する今後の課題は、各章の考察において示した。すなわち、電子透かし用途としては、知覚検知閾に基づいた変調強度設定、ピッチ変換および差分攻撃に対する耐性の向上である。また、空間伝搬用途においては、情報埋め込み実時間ソフトウェアの実装、検出処理負荷の低減と検出ソフトウェアのPDA等携帯機器への実装である。この振幅変調に基づく音響情報秘匿技術は、電子透かしから空間伝搬のステガノグラフィまで、様々な場面において利用可能であることを本論文は示してきた。これら今後の課題を解決することによって、より一層利用可能な範囲が広がり、実用に繋がることが期待される。

さらに、音響情報秘匿技術全般に対する今後の課題は、技術の新たな利用場面の提案とその実用化であると言えよう。情報秘匿技術の利用は、著作権保護/管理や秘匿通信といった従来重要視されてきた分野だけに限られる訳ではない。本論文では、空間伝搬を前提とした応用、すなわちアナウンス音声やBGMを用いた聴覚障害者や外国人を含む利用者への情報提供、カラオケ歌詞表示などの利用場면을提案してきた。そのような本論文で試みてきた以外の音響情報秘匿技術の利用としては、既にいくつかの技術と新しい利用場面が提案されている。それらは、音響データ通信路におけるパケット損失の隠蔽や高品質化 [7, 4, 73, 6]、メタデータの記録と検索 [8]、改ざん検出 [74] などを含む情報法科学的利用などである。本研究やさらに新たな研究における音響情報秘匿技術は、利用場面の要求に合った性能を満たすだけでなく、その実用化を通じて社会の利益に貢献する方策も積極的に探っていくべきであろう。

謝辞

本研究を遂行し、論文としてまとめるにあたり、これまで数多くの方々からの御指導と御助言、様々な機関からの御援助を賜りました。特にお世話になった方々と機関を以下に記し、深い感謝の意を表します。

岩宮眞一郎 九州大学芸術工学院教授には、本論文の構成や内容について、御教示と御指導を賜りました。また、浦濱喜一同大学院教授、尾本章同大学院准教授には、御多忙のところ本研究に関する的確な御指導と御助言を賜りました。

佐々木實 九州芸術工科大学名誉教授には、著者が大学院修士課程において修士論文の御指導を賜わり、初めて書いた英語論文の添削、心理実験や実験設備の構築など、その後の研究生活における様々な要諦を学ばせていただきました。

津村尚志 元九州芸術工科大学教授からは、大学院博士後期課程において聴覚に関する様々な理論や知見を、ゼミや直接のご指導を通じて学ばせていただきました。

安藤由典 元東京情報大学教授には、現在の職場における最初の上長として、研究環境の整備に御尽力下さり、著者の研究をご支援頂きました。そして、本研究遂行に必要であった簡易無響室を始めとする多くの音響機器等の環境を整えることができました。

小泉宣夫 東京情報大学教授には、現在の上長として著者の大学内外での研究活動にご理解を頂き、大学内での研究教育活動および各種業務において様々なご支援を賜りました。

山田真司 金沢工業大学教授には、大学院の先輩として著者を学位取得に向けて叱咤激励下さっただけでなく、論文のまとめかたについてもご指導を賜りました。

電子情報通信学会マルチメディア情報ハイディング研究会におきましては、2006年11月の準備会議から始まり、2007年5月の第1回研究会より今日まで、多くの専門委員の皆様からの御助言と御指導を賜りました。立花隆輝 日本IBM株式会社東京基礎研究所主任研究員には、音響電子透かし技術の研究先駆者として、透かし全般とその技術への様々な観点、着眼点、問題点などを御指摘頂きました。松岡保静 NTTドコモ先端技術研究所研究員からは、秘匿情報検出アルゴリズムのPDA実装について、御助言をいただきました。

た．越前功 国立情報学研究所准教授には，著者の研究にご理解を頂き，国立情報学研究所公募型共同研究へお誘い頂いた上，音響電子透かし技術に関する新しい観点でのご議論を頂きました．西村竜一 情報通信研究機構主任研究員 には，位相変調に基づく情報秘匿技術の研究先駆者として，研究推進のヒントを頂きました．また特に，元研究委員会副委員長 鈴木陽一 東北大学電気通信研究所教授には，研究会へお誘い頂き，東北大学電気通信研究所共同プロジェクト研究への参加にも便宜を計って頂いただけではなく，国際会議への著者の研究発表をご奨励頂き，聴覚および信号処理に関する豊富なご見識から，多くのご助言を賜りました．

また，日本音響学会研究発表会，聴覚研究会および電気音響研究会では，著者の研究発表の折に，ご出席の皆様から多くのご助言を賜りました．

株式会社オトデザイナーズ取締役社長 坂本真一 博士には，空間伝搬音響による情報秘匿技術の応用としてカラオケ利用を御提案頂き，これに関する特許の取得において大変な御協力を賜りました．同じくこの特許の取得には，麹町特許事務所および小山有 弁理士より，費用面および特許書面のご用意など幅広く御協力を賜りました．

日本エヴィクサー株式会社研究開発事業部長 鈴木久晴 博士には，御多忙中のおり，本務業務の時間を割いて頂き，Windows 版情報検出表示ソフトウェアの実装および Windows Mobile での演算負荷量計算ソフトの制作を請け負って頂きました．

本研究は，このような方々との熱心な議論とご助力によって成し得たものです．

東京情報大学総合情報学部情報文化学科 教員の皆様には，著者の研究活動を叱咤激励かつご支援くださり，大学内の業務等においても様々な便宜を計って頂きました．

本研究遂行のために必要な予算的補助の一部として，平成 20~ 22 年度科研費基盤 C (20560365) による助成，平成 21 年度および 22 年度国立情報学研究所共同研究による助成，平成 19 年度および 20 年度東京情報大学総合情報学研究所共同研究費による助成，および平成 17 年大川情報通信基金研究助成を頂きました．本研究成果に関する国際会議発表旅費として，平成 18 年 12 月 電気通信普及財団海外渡航旅費援助，平成 19 年 11 月 NEC C&C 財団 国際会議論文発表者助成金を頂きました．本研究成果に関する国内会議発表旅費および研究会出張旅費として，平成 19 年度，20 年度および 21 年度東北大学電気通信研究所共同プロジェクト研究による助成を頂きました．

皆様へ心よりの感謝の意を表します．

最後に，大学院博士後期課程在学中から就職を経て今日までの永きに渡り，日々の著者の家庭生活および教育研究活動を支え，学位取得へ向けて叱咤激励してくれた最愛の妻で

ある勝美に、何物にも換え難い深い感謝を表します。本当にありがとうございます。

参考文献

- [1] 今井秀樹, 遠藤直樹, 古原和邦, 五十嵐達治, 川森雅仁, 三瓶徹, 中西康浩, ユビキタス時代の著作権管理技術 DRMとコンテンツ流通(東京電機大学出版, 東京, 2006).
- [2] 柏野邦夫, “音響指紋技術とその応用,” 日本音響学会誌, **66**, No. 2, 71–76 (2010).
- [3] S. Chen and H. Leung, “Artificial bandwidth extension of telephony speech by data hiding,” in *Circuits and Systems, 2005. ISCAS 2005. IEEE International Symposium on*, 3151–3154 Vol. 4, May 2005.
- [4] 青木直史, “全波整流に基づくステガノグラフィを用いた G.711 音声の一带域拡張法,” 電子情報通信学会論文誌, **J90-B**, No. 7, 697–704 (2007).
- [5] B. Geiser and P. Vary, “Backwards Compatible Wideband Telephony in Mobile Networks: CELP Watermarking and Bandwidth Extension,” in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, Vol. IV, 533–536, (2007).
- [6] Akira Nishimura, “Steganographic Band Width Extension for the AMR Codec of Low-Bit-Rate Modes,” in *Proceedings of Interspeech 2009*, 2611–2614 International Speech Communication Association, (2009).
- [7] 青木直史, “ピッチ波形複製法に基づくステガノグラフィを用いた VoIP におけるパケット損失の一隠蔽法,” 電子情報通信学会論文誌 B, **J86-B**, No. 12, 2551 – 2560 (2003).
- [8] Shigeru Sasaki, Masakiyo Tanaka, Yoshiteru Tsuchinaga, Masanao Suzuki, and Yasuji Ota, “Method and system for embedding and extracting data from encoded voice code,” United States Patent 7310596 (2007).

- [9] M. Löytynoja, N. Cvejic, E. Lähetkangas, and T. Seppänen, “Audio encryption using fragile watermarking,” in *Proc. Fifth International Conference on Information, Communications and Signal Processing*, 881–885, (2005).
- [10] Michael Gulbis, Erika M, and Martin Steinebach, “Content-Based Authentication Watermarking with Improved Audio Content Feature Extraction,” in *Proceedings of IEEE the 4th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 620–623 IEEE Computer Society, (2008).
- [11] 茂出木敏雄, “音響空間のコピキタス化に向けた電子透かし埋込み容量の拡大技術,” 電気学会論文誌 C, **127**, 1013–1021 (2007).
- [12] 東北大学電気通信研究所, 国立特殊教育総合研究所, “空気伝搬音響透かしの福祉工学への応用,” 第1回 マルチメディア情報ハイディング研究会資料, 103—106, (2007).
- [13] Masataka Goto, Hiroki Hashiguchi, Takuichi Nishimura, and Ryuichi Oka, “RWC Music Database: Music Genre Database and Musical Instrument Sound Database,” in *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003)*, 229—230, (2003).
- [14] P. Kabal, “An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality,” TSP Lab Technical Report, Dept. Electrical & Computer Engineering, McGill University, 1—89 (2002).
- [15] 高秉燮, 西村竜一, 鈴木陽一, “エコー拡散透かし手法における音質と埋め込み容量に関する検討,” 電子情報通信学会 情報セキュリティ研究会資料, No. ISEC2002-31, 125–130 (2002).
- [16] Fabien A.P. Petitcolas, Ross J. Anderson, and Markus G. Kuhn, “Attacks on Copyright Marking Systems,” in *Proceedings of the 2nd International Workshop on Information Hiding 1998*, 218—238, (1998).
- [17] Andreas Lang, Jana Dittmann, Eugene T. Lin, and Edward J. Delp, “Application Oriented Audio Watermark Benchmark Service,” in *Security, Steganography and Watermarking of Multimedia Contents VII*, Vol. 5681, 275—286 SPIE, (2005).

- [18] N. Aoki, “A band extension technique for G.711 speech using steganography,” *IEICE Transactions on Communications*, **E89-B**, 1896–1898 (2006).
- [19] 渡辺馨, “オーディオ信号の劣化の評価法,” *日本音響学会誌*, **63**, No. 11, 820–825 (2007).
- [20] 岩切宗利, 松井甲子雄, “適応差分 PCM 符号化における音声符号へのテキスト情報の埋込み,” *情報処理学会論文誌*, **38**, No. 10, 2053–2061 (1997).
- [21] Ralf Geiger, Yoshikazu Yokotani, and Gerald Schuller, “Audio data hiding with high data rates based on IntMDCT,” in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing*, No. V, 205–208 IEEE, (2006).
- [22] 岩切宗利, 松井甲子雄, “音声符号への電子透かしに関する一検討,” *電子情報通信学会 1997 年基礎境界ソサエティ大会*, SA-7-5, 250–251 (1997).
- [23] D. Koukopoulos and Y. C. Stamatiou, “A compressed domain watermarking algorithm for Mpeg Layer 3,” in *Proc. Multimedia and Security Workshop at ACM Multimedia 2001*, 7–10 ACM Press, (2001).
- [24] N. Cvejic and T. Seppänen, “Improving audio watermarking scheme using psychoacoustic watermark filtering,” in *Proc. 1st IEEE International Symposium on Signal Processing and Information Technology*, 169–172, (2001).
- [25] Daniel Gruhl, Anthony Lu, and Walter Bender, “Echo Hiding,” in *Proceedings of the First International Workshop on Information Hiding, LNCS 1174*, 295–315, (1996).
- [26] Byeong-Seob KO, Ryouichi Nishimura, and Yoiti Suzuki, “Robust Watermarking Based on Time-spread Echo Method with Subband Decomposition,” *IEICE Trans. Fundamentals*, **E87-A**, 1647–1650 (2004).
- [27] Akira Takahashi, Ryouichi Nishimura, and Yoiti Suzuki, “Multiple Watermarks for Stereo Audio Signals Using Phase-Modulation Techniques,” *IEEE Transactions on signal processing*, **53**, No. 2, 806–815 (2005).
- [28] Ryuki Tachibana, Shuichi Shimizu, Seiji Kobayashi, and Taiga Nakamura, “An audio watermarking method robust against time- and frequency-fluctuation,” in *Proc. of*

- Security and Watermarking of Multimedia Contents III, SPIE*, Vol. 4314, 104–115, (2001).
- [29] R. Tachibana, “Two-Dimensional Audio Watermark for MPEG AAC Audio,” in *Proc. of SPIE Int. Conf. on Security, Steganography, and Watermarking of Multimedia Contents VI*, Vol. 5306, 139–150, San Jose, USA, (2004).
- [30] 村松巖, 荒川薫, “オクターブ類似性に基づくオーディオ信号の電子透かし,” 電子情報通信学会論文誌, **J87-A**, No. 6, 787–796 (2004).
- [31] N. Cvejic and T. Seppänen, “Spreadspectrum audio watermarking using frequency hopping and attack characterization,” *Signal Processing*, **84**, 207–213 (2004).
- [32] Wei Li and Xiangyang Xue, “Audio Watermarking Based on Music Content Analysis: Robust against Time Scale Modification,” in *Digital Watermarking: Second International Workshop, IWDW 2003, LNCS 2939*, Ton Kalker, Ingemar J. Cox, and Yong Man Ro, Eds. (Springer-Verlag, Berlin, 2004), pp. 289–300.
- [33] C.-P. Wu, P.-C. Su, and C.-C. J. Kuo, “Robust Audio Watermarking for Copyright Protection,” in *Proceedings of SPIE*, Vol. 3807, 387–397, (1999).
- [34] Yi-Wen Liu, “Multiple fundamental frequency estimation assisted by audio watermarking,” *J. Acoust. Soc. Am.*, **120**, 3217 (2006).
- [35] 藺田光太郎, 西村竜一, 鈴木誠史, 滝澤修, “DWPT-QIM 音楽用電子透かし,” 平成 19 年度春季日本音響学会講演論文集, 621–622 (2007).
- [36] D. R. Perrot and J. Tucker, “Minimum Audible Movement angle as a Function of Signal and the Velocity of the Source,” *J. Acoust. Soc. Am.*, **83**, 1522–1527 (1988).
- [37] N. Cvejic and T. Seppänen, “Increasing Robustness of an Improved Spread Spectrum Audio Watermarking Method Using Attack Characterization,” in *Proc. International Workshop on Digital Watermarking, LNCS 2939*, 467–473, (2003).
- [38] Scott A. Craver, Min Wu, Bede Liu, Adam Stubblefield, Ben Swartzlander, Dan S. Wallach and Drew Dean, and Edward W. Felten, “Reading Between the Lines:

- Lessons from the SDMI Challenge,” in *Proceedings of the 10th USENIX Security Symposium*, 1–11, (2001) <http://www.usenix.org/events/sec01/craver.pdf>.
- [39] Ryuki Tachibana, “Improving audio watermarking robustness using stretched patterns against geometric distortion,” in *Advances in Multimedia Information Processing, PCM2002, (LNCS) 2532*, 2002), pp. 647–654.
- [40] Zheng Liu and Akira Inoue, “Audio Watermarking Techniques Using Sinusoidal Patterns Based on Pseudorandom Sequences,” *IEEE Transactions on Circuits and Systems for Video Technology*, **13**, 801–812 (2003).
- [41] 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: 音楽ジャンルデータベースと楽器音データベース,” *音講論*, 春 3-7-6, 843–844 (2003).
- [42] 4C Entity, “4C 12 BIT Watermark Specification,” http://www.4centity.com/docs/4C_Spec.pdf, (1999).
- [43] 日本音楽著作権協会, “音楽電子透かし 4 社を技術認定し、ネット上の違法利用の発生防止に活用 (STEP2001),” http://www.jasrac.or.jp/release/01/10_2.html.
- [44] Takahiro Yamada, *Essentials of error-control coding techniques* (Academic Press, San Diego, 1990), pp. 46–49.
- [45] 藤井康広, 越前功, 山田隆亮, 手塚悟, 吉浦裕, “電子透かし検出に適した誤り訂正符号の拡張方式,” *情報処理学会論文誌*, **45**, No. 8, 1980–1997 (2004).
- [46] 西村明, 小泉宣夫, “AD/DA 変換器における時間ゆらぎを考慮に入れた波形比較による音質差測定法 — CD 媒体個体差による音質差測定の試み —,” *電子情報通信学会技術研究報告*, **EA2004-20**, 15–22 (2004).
- [47] 小森智康, “ワンセグ及びメモリーオーディオの音質評価,” *日本音響学会誌*, **63**, No. 11, 674–679 (2007).
- [48] 西村竜一, 末永司, 鈴木陽一, 田中章浩, “音質劣化が刺激の印象空間内での布置に及ぼす影響,” *日本音響学会誌*, **64**, No. 2, 63–72 (2008).

- [49] Thilo Thiede, William C. Treurniet, Roland Bitto, Christian Schmidmer, Thomas Sporer, John G. Beerends, and Catherine Colomes, “PEAQ - The ITU Standard for Objective Measurement of Perceived Audio Quality,” *48*, No. 1&2, 3–29 (2000).
- [50] Torsten Dau, Birger Kollmeier, and Armin Kohlrausch, “Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers,” *J. Acoust. Soc. Am.*, **102**, No. 5, 2892–2905 (1997).
- [51] T. Houtgast, “Frequency selectivity in amplitude-modulation detection,” *J. Acoust. Soc. Am.*, **85**, 1676—1680 (1989).
- [52] D. Eddins, “Amplitude modulation detection of narrow-band noise; Effects of absolute bandwidth and frequency region,” *J. Acoust. Soc. Am.*, **93**, 470–479 (1993).
- [53] Torsten Dau, Birger Kollmeier, and Armin Kohlrausch, “Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration,” *J. Acoust. Soc. Am.*, **102**, No. 5, 2906–2919 (1997).
- [54] William A. Yost and Stanley Sheft, “Modulation Detection Interference: Across-frequency processing and auditory grouping,” *Hearing Research*, **79**, 48—58 (1994).
- [55] J. W. Hall and J. H. Grose, “Effects of flanking band proximity, number, and modulation pattern on comodulation masking release,” *J. Acoust. Soc. Am.*, **87**, 269–283 (1990).
- [56] R. P. Carlyon, S Buus, and M. Florentine, “Comodulation Masking Release for three types of modulator as a function of modulation rate,” *Hearing Research*, **42**, 37–46 (1989).
- [57] B. G. Berg, “On the relation between comodulation masking release and temporal modulation transfer functions,” *J. Acoust. Soc. Am.*, **100**, 1013–1023 (1996).
- [58] 松岡保静, 中島悠輔, 吉村健, “音響 OFDM - 音のバーコード - ,” 第1回 マルチメディア情報ハイディング研究会資料, 61—64, (2007).
- [59] Toshio Modegi, “Spatial and Temporal Position Information Delivery to Mobile Terminals using Audio Watermarking Techniques,” in *Proceedings of IEEE International*

Conference on Intelligent Information Hiding and Multimedia Signal Processing, 483–486 IEEE, (2009).

- [60] 松岡保静, “サブバンド位相シフトを用いた音響電子透かし埋め込み法,” 電子情報通信学会技術研究報告, **EA2006-43**, 529–533 (2006).
- [61] 中島悠輔, 松岡保静, 吉村健, “音響 OFDM の空中伝送特性,” 電子情報通信学会技術研究報告. RCS2006–82, 無線通信システム, **106**, No. 168, 139–144 (2006).
- [62] 松岡保静, 中島悠輔, 吉村健, “可聴帯域における音波情報伝送技術 : 音響 OFDM,” 電子情報通信学会技術研究報告, **EA2006-24**, 25–29 (2006).
- [63] 中村哲, 比屋根一雄, 浅野太, 遠藤隆, “実環境における音響シーンデータベースの構築,” 平成 10 年度秋季日本音響学会講演論文集, **1-R-10**, 137–138 (1998).
- [64] M. R. Schroeder, “New method of measuring reverberation time,” *J. Acoust. Soc. Am.*, **37**, 1187–1188 (1965).
- [65] 3rd Generation Partnership Project, “Mandatory Speech Codec speech processing functions AMR Speech Codec; General Description,” **26.071**, (2001).
- [66] “IEEE Standard: Draft Standard Methods for Measuring Transmission Performance of Analog and Digital Telephone Sets, Handsets and Headsets,” **269-2002**, (2002).
- [67] 3rd Generation Partnership Project, “ANSI-C code for the Adaptive Multi Rate speech codec,” **26.073**, (2001).
- [68] ITU-T Recommendation, “Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,” **P.862**, (2001).
- [69] Hironobu Takei, Shogo Nakamura, Hiroto Saitoh, and I. Nyoman Sukadana, “An implementation of realtime speech speed control system using hierarchical cascaded integrator and comb filters,” in *J. Acoust. Soc. Am.*, 3216–3216, (2006).
- [70] Byeong-Seob Ko, Ryouichi Nishimura, and Yoiti Suzuki, “Time-spread echo method for digital audio watermarking,” *IEEE Trans. on Multimedia*, **7**, 212–221 (2005).

- [71] Akira Takahashi, Ryouichi Nishimura, and Yoiti Suzuki, “Time-spread echo digital audio watermarking tolerant of pitch shifting,” *Acoustical Science and Technology*, **26**, No. 6, 530–532 (2005).
- [72] Masataka Goto, Hiroki Hashiguchi, Takuichi Nishimura, and Ryuichi Oka, “RWC Music Database: Popular, Classical, and Jazz Music Databases,” in *Proceedings of the 3rd International Conference on Music Information Retrieval (ISMIR 2002)*, 287–288, (2002).
- [73] P. Vary and B. Geiser, “Steganographic Wideband Telephony Using Narrowband Speech Codecs,” in *Signals, Systems and Computers, 2007. ACSSC 2007. Conference Record of the Forty-First Asilomar Conference on*, 1475-1479, Nov. 2007.
- [74] Xuping Huang, Isao Echizen, and Akira Nishimura, “A New Approach of Reversible Acoustic Steganography for Tampering Detection,” in *Proc. of IHHMSP2010*, 538–542, (2010).

本論文に関する著者の参考文献

原著論文

1. 西村明, “音響電子透かしの耐性評価シミュレーション,” 東京情報大学研究論集, **10**, No. 2, 113–141 (2007).
2. Akira Nishimura, “Audio watermarking based on subband amplitude modulation,” *Acoustical Science and Technology*, **31**, No. 5, 328–336 (2010).
3. Akira Nishimura, “Audio Data Hiding that is Robust with Respect to Aerial Transmission and Speech Codecs,” *International Journal of Innovative Computing Information and Control*, **6**, No. 3(B), 1389–1400 (2010).
4. 西村明, 坂本真一, “音響データハイディングを用いるスピーカ再生音と同期した情報呈示システム,” 電子情報通信学会論文誌 A, **J93-A**, No. 2, 91–99 (2010).

国際会議論文

1. Akira Nishimura, “Data Hiding for Audio Signals that are Robust with respect to Air Transmission and a Speech Codec,” in *Proceedings of IEEE the 4th International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 601–604 IEEE, (2008).
2. Akira Nishimura, “Presentation of Information Synchronized with the Audio Signal Reproduced by Loudspeakers Using an AM-based Watermark,” in *Proceedings of the 3rd International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Vol. 2, 275–278 IEEE, (2007).
3. Akira Nishimura, “Subjective and objective quality evaluation for audio watermarking based on sinusoidal amplitude modulation,” in *Proceedings of the 19th International Congress on Acoustics*, No. ELE-04-009, 1–9, (2007).

4. Akira Nishimura, “Data hiding for speech sounds using subband amplitude modulation robust against reverberations and background noise,” in *Proceedings of IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 7–10 IEEE, (2006).
5. Akira Nishimura, “Audio watermarking based on sinusoidal amplitude modulation,” in *Proceedings of ICASSP 2006, IV*, 797–800 IEEE, (2006).

解説論文

1. 西村明, “音響信号への情報秘匿技術 —電子透かしとステガノグラフィ—,” 日本音響学会誌, **63**, No. 11, 660–667 (2007).

特許

1. 西村明, 坂本真一, “音楽データに文字データを埋め込んだ複合音響信号を再生する文字データ再生装置,” 特許第 3976779 号 (2007).
2. 西村明, “振幅変調を用いた音響信号への電子透かしデータの埋め込み・検出装置,” 特許願 2005-71245 (2005).