

修 士 論 文

音信号への蝸牛遅延特性に基づく
情報ハイディング法に関する検討

北陸先端科学技術大学院大学
情報科学研究科情報科学専攻

小杉 敏三

2011 年 3 月

修 士 論 文

音信号への蝸牛遅延特性に基づく 情報ハイディング法に関する検討

指導教官 赤木正人 教授

審査委員主査 赤木正人 教授
審査委員 党建武 教授
審査委員 徳田功 准教授

北陸先端科学技術大学院大学
情報科学研究科情報科学専攻

0910023 小杉 敏三

提出年月: 2011 年 2 月

概 要

近年，情報通信技術の発達やインフラ整備が急速に進められたことにより，インターネットを利用したマルチメディアデジタルコンテンツ（テキスト，音，静止・動画像）の利用が盛んになっている．静止・動画像の利用はもちろんのこと，最近では音楽コンテンツや音声通信（VoIP）の利用が急激に伸びている．そのため，インターネット上の音情報伝送の重要性が非常に高まっており，その利用の安全性を高めるために，コンテンツ保護（著作権保護など）や機密情報保護の中核技術が求められている．マルチメディア情報ハイディング（MIH）技術は，これらの中核技術を担う，暗号技術とは異なる情報保護技術として最近注目を集めている．

この MIH 技術を完全なものとして確立するためには，次の四つの要求項目を満たす必要がある．(1) 埋め込み情報が利用者に知覚されず，埋め込みによる原音の知覚可能な歪みが一切生じないこと（検知不可能性）．(2) 情報が埋め込まれていることを利用者に気づかせないこと（秘匿性）．(3) 通常の信号変換処理や悪意のある攻撃に対して影響を受けないこと（頑健性）．(4) 埋め込み可能な情報を増やし，汎用性を高めること（埋め込み情報量）．従来手法では，これら四つの要求項目を全て満たす手法はない．一方，Unoki & Hamada が提案した蝸牛遅延特性に基づく情報ハイディング法は，四つの要求項目の内検知不可能性と頑健性を満たす手法である．本研究では，検知不可能性と頑健性を満たす蝸牛遅延特性に基づく情報ハイディング法を用いて四つ目の要求項目である埋め込み情報量を満たすため，蝸牛遅延フィルタの構成法を提案する．

並列型，縦続型，複合型構成を提案し，音楽信号と音声信号における客観評価実験を行う．音楽信号における検知不可能性に関する評価結果より，現状提案している多段フィルタ構成の中では， $L = 1, N = 3$ と $L = 3, N = 1$ が一番最適な組み合わせである．ここで，このフィルタ構成における最大埋め込み情報量は，768 bps（256 fps の時 1 フレームで 3 bits 表現）である．音声信号における検知不可能性に関する評価結果より，現状提案している多段フィルタ構成の中では， $L = 1, N = 2$ が一番最適な組み合わせである．ここで，このフィルタ構成における最大埋め込み情報量は，512 bps（256 fps の時 1 フレームで 2 bits 表現）である．また，音楽信号と音声信号における頑健性に関する客観評価実験を行う．音楽信号における頑健性に関する評価結果より，現状提案している多段フィルタ構成の中では， $L = 2, N = 2$ が一番最適な組み合わせである．ここで，このフィルタ構成における最大埋め込み情報量は，256 bps（64 fps の時 1 フレームで 4 bits 表現）である．音声信号における頑健性に関する評価結果より，PCM コーデックに対しては， $L = 1, N = 2$ が最良であることがわかる．このフィルタ構成における最大埋め込み情報量は，512 bps（256 fps の時 1 フレームで 2 bits 表現）である．ADPCM コーデックに対しては， $L = 1, N = 3$ が最良であることがわかる．このフィルタ構成における最大埋め込み情報量は，384 bps（128 fps の時 1 フレームで 3 bits 表現）である．さらに，音楽信号における各フィルタ構成の頑健性に関する客観評価として，StirMark ベンチマークテストを行う．StirMark

ベンチマークテストの結果から，提案法は雑音を加えられる処理，振幅操作，ビット処理，フィルタリング処理に関しては頑健性が保証できることがわかる．しかし，データ置換処理，位相操作，残響を加えられる処理に対しては頑健性が保証できないことがわかる．

提案法の応用として，音声ステガノグラフィとして利用可能であるかを実際に 32×32 bitmap 画像を埋め込み，検出を行う．画像情報を埋め込んだステゴデータから画像を検出した結果，ビット検出率は 97.5% であり，26 bit の誤りが生じている．画像情報を埋め込んだステゴデータに PCM コーデックを用いて符号化・複合化した音声信号から画像を検出した結果，ビット検出率は 90.9% であり，93 bit の誤りが生じている．検知不可能性だけではなく頑健性も非常に重要であり，この二つを満たす提案法は音声ステガノグラフィとして利用できる．

以上の結果より提案法は，検知不可能性と頑健性を満たした上で埋め込み情報量を高めることができ，音声ステガノグラフィとしても利用可能であることがわかり，マルチメディア情報ハイディング技術として，有効な手法であるといえる．

目 次

第 1 章	序論	1
1.1	デジタルコンテンツの利用状況	1
1.2	マルチメディア情報ハイディング技術	2
1.3	本研究の目的	2
1.4	本論文の構成	3
第 2 章	代表的な音信号への情報ハイディング法	4
2.1	LSB 置換法	4
2.1.1	選択的 LSB 置換法	4
2.1.2	選択的 LSB 置換法の改良	4
2.2	スペクトル拡散法	5
2.3	エコーハイディング法	5
2.4	振幅変調法	5
2.5	周期的位相変調法	6
2.6	蝸牛遅延特性に基づく電子音響透かし	6
第 3 章	蝸牛遅延特性に基づく情報ハイディング法	7
3.1	蝸牛遅延特性	7
3.2	蝸牛遅延フィルタ	8
3.3	情報の埋め込み方法	8
3.4	情報の検出方法	9
第 4 章	蝸牛遅延フィルタの構成法	11
4.1	構成法のコンセプト	11
4.2	多段並列型フィルタ構成	12
4.2.1	情報の埋め込み方法	12
4.2.2	情報の検出方法	13
4.3	多段縦続型フィルタ構成	15
4.3.1	情報の埋め込み方法	16
4.3.2	情報の検出方法	17
4.4	多段複合型フィルタ構成	18
4.4.1	情報の埋め込み方法	18

4.4.2	情報の検出方法	19
第 5 章	提案法の評価	21
5.1	音楽信号に関する客観評価	21
5.1.1	多段並列型における客観評価	21
5.1.2	多段縦続型における客観評価	22
5.1.3	多段複合型における客観評価	22
5.2	音声信号に関する客観評価	23
5.2.1	多段並列型における客観評価	23
5.2.2	多段縦続型における客観評価	24
5.2.3	多段複合型における客観評価	24
5.3	耐性評価	25
5.3.1	音楽信号における耐性評価	25
5.3.2	音声信号における耐性評価	26
5.4	StirMark ベンチマーク	27
第 6 章	提案法の応用	43
6.1	音声ステガノグラフィ	43
6.1.1	評価試験方法	44
6.1.2	評価試験結果	44
第 7 章	結論	47
7.1	本論文であきらかになったこと	47
7.2	残された課題	48
	謝辞	49
	参考文献	50
	学会発表リスト	54
	付録	56

目 次

3.1	蝸牛遅延を模擬したフィルタの群遅延特性	9
3.2	蝸牛遅延特性に基づく情報ハイディング法におけるデータ埋め込み/検出の ブロックダイアグラム	10
4.1	並列型構成における蝸牛遅延フィルタの群遅延特性	12
4.2	並列型構成におけるデータ埋め込み/検出のブロックダイアグラム	13
4.3	縦続型構成における蝸牛遅延フィルタの群遅延特性	15
4.4	縦続型構成におけるデータ埋め込み/検出のブロックダイアグラム	16
4.5	複合型構成におけるデータ埋め込み/検出のブロックダイアグラム	19
4.6	複合型構成 ($L = 2, N = 2$) における蝸牛遅延フィルタの群遅延特性	20
5.1	音楽信号に対して並列型構成 ($N = 1, 2, 3, 4$) を用いた客観評価実験の結果: (a) PEAQ, (b) LSD, (c) ビット検出率	29
5.2	音楽信号に対して縦続型構成 ($L = 1, 2, 3, 4$) を用いた客観評価実験の結果: (a) PEAQ, (b) LSD, (c) ビット検出率	30
5.3	音楽信号に対して複合型構成 ($(L, N) = (1, 4), (4, 1), (2, 2)$) を用いた客観 評価実験の結果: (a) PEAQ, (b) LSD, (c) ビット検出率	31
5.4	音声信号に対して並列型構成 ($N = 1, 2, 3, 4$) を用いた客観評価実験の結果: (a) PESQ, (b) LSD, (c) ビット検出率	32
5.5	音声信号に対して縦続型構成 ($L = 1, 2, 3, 4$) を用いた客観評価実験の結果: (a) PESQ, (b) LSD, (c) ビット検出率	33
5.6	音声信号に対して複合型構成 ($(L, N) = (1, 4), (4, 1), (2, 2)$) を用いた客観 評価実験の結果: (a) PESQ, (b) LSD, (c) ビット検出率	34
5.7	並列型構成 ($N = 1, 2, 3, 4$) における音声符号化耐性試験の結果: (a) PCM, (b) ADPCM, (c) CS-ACELP	36
5.8	縦続型構成 ($L = 1, 2, 3, 4$) における音声符号化耐性試験の結果: (a) PCM, (b) ADPCM, (c) CS-ACELP	37
5.9	複合型構成 ($(L, N) = (1, 4), (4, 1), (2, 2)$) における音声符号化耐性試験の 結果: (a) PCM, (b) ADPCM, (c) CS-ACELP	38
5.10	並列型構成における StirMark ベンチマークテストの結果: (a) $N = 1$, (b) $N = 2$, (c) $N = 3$, (d) $N = 4$	40

5.11 縦続型構成における StirMark ベンチマークテストの結果 : (a) $L = 1$, (b) $L = 2$, (c) $L = 3$, (d) $L = 4$	41
5.12 複合型構成における StirMark ベンチマークテストの結果 : (a) $L = 1, N = 4$, (b) $L = 4, N = 1$, (c) $L = 2, N = 2$	42
6.1 音声ステガノグラフィのイメージ	45
6.2 埋め込みに用いた 32×32 の bitmap 画像	45
6.3 検出した 32×32 の bitmap 画像	46
6.4 PCM 符号化を行った音声における検出画像	46

第1章 序論

1.1 デジタルコンテンツの利用状況

近年，情報通信技術の発達やインフラ整備が急速に進められたことにより，インターネットを利用したマルチメディアデジタルコンテンツ（テキスト，音，静止・動画像）の利用が盛んになっている．例えば，2011年1月の国内における動画投稿サイト Youtube への一日の平均アクセス数は約1億3千万となっている [1]．また，音楽配信サービス iTunes Store では，楽曲総ダウンロード数が2010年2月に100億曲を突破している [2]．さらに，IP 電話ソフトウェア Skype の世界における2010年下半期の1カ月当たりの平均利用者数は約1億2千万人となっている [3]．動画投稿サイト Youtube は，誰でも動画投稿することができ，無料で投稿されている動画を視聴できることが魅力となっている．しかし，ユーザー視聴の大半がテレビ番組などプロが制作したコンテンツとなっている．音楽配信サービスにおいては，購入した楽曲もコピーが可能となっており，P2P (Peer to Peer) ソフトウェアなどを利用することでユーザー間でのコピーファイルのやりとりが行われている．IP 電話ソフトウェア Skype は，P2P 技術を用いたインターネット電話サービスであり，どのような回線を使って通信を行っているかわからないため，第三者による傍受や多くのユーザーが利用することでトラフィックを占有してしまい通信障害もたびたび起こっている．これらデジタルコンテンツの利用上の安全性を確保することや制作者の著作権保護は非常に重要な課題として取り組まれてきた [4]．そのひとつとして，権利者の利益を守る DRM (Digital Rights Management) 技術がある [5]．DRM 技術は流通させるデジタルコンテンツにすべからず暗号化処理を施し，対価を支払ったユーザーだけに視聴やコピーを限定する．しかし，DRM により利便性が非常に低下することからユーザーの大きな反発を受けているうえ，DRM 技術が権利者の信用を得ていない．このようなことから2003年に始まった音楽配信サービス iTunes Store は，2007年に DRM フリーでの音楽配信を開始した．これに追従する形で，現在では世界の四大レコード会社すべてが DRM フリーでの楽曲提供を行っている．

このような背景から，デジタルコンテンツの利用上の安全性を確保する技術を開発することが急務となっている．

1.2 マルチメディア情報ハイディング技術

マルチメディア情報ハイディング (MIH) 技術は、テキスト、音、静止・動画像といったマルチメディア情報の利用上の安全確保を目的として、暗号とは異なる情報保護技術として注目を集めている [6]。MIH 技術は、大別すると、マルチメディア情報のコンテンツ保護（著作権保護）のための電子透かしと、マルチメディア情報自体に別の情報を隠すステガノグラフィに大別される。また、この技術は暗号との併用が可能な技術であり、マルチメディア情報のコンテンツ認証（例えば、フィンガープリントなど）や機密情報保護（例えば、秘匿通信・匿名通信など）にも利用可能である。

この MIH 技術を完全なものとして確立するためには、次の四つの要求項目を満たす必要がある。(1) 埋め込み情報が利用者に知覚されず、埋め込みによる原音の知覚可能な歪みが一切生じないこと（検知不可能性）。(2) 情報が埋め込まれていることを利用者に気づかせないこと（秘匿性）。(3) 通常の信号変換処理や悪意のある攻撃に対して影響を受けないこと（頑健性）。(4) 埋め込み可能な情報を増やし、汎用性を高めること（埋め込み情報量）。

これまでに音楽信号を対象とした情報ハイディング法は、LSB 置換法 [7]、スペクトル拡散法 [8]、エコーハイディング法 [9]、振幅変調に基づく手法 [10]、周期的位相変調法 [11]、心理音響モデルに基づく手法 [12]、オクターブ類似性に基づく手法 [13] などが提案されており、これ以外にも様々な手法 [14, 15] が提案されている。また、音声信号を対象とした情報ハイディング法は、LSB 置換法 [16]、選択的 LSB 置換法 [17]、選択的 LSB 置換法の拡張 [18]、振幅変調に基づく手法 [19] などが提案されている。しかし、いずれの手法においても四つの要求項目を満たす手法はない。

一方で、Unoki らは蝸牛遅延特性に着目した新しい情報ハイディング法を提案している [20]～[26]。この手法は、四つの要求項目の条件を満たすことに重点が置かれ、深く検討されてきた。上記にあげたような代表的な方法と比較検討し、四つのうち検知不可能性と頑健性に関して提案法の有効性を示すことができている。

1.3 本研究の目的

本稿では、検知不可能性と頑健性を満たす蝸牛遅延特性に基づく情報ハイディング法を用いて四つ目の要求項目である埋め込み情報量を満たすため、蝸牛遅延フィルタの構成方法を提案する。検知不可能性と頑健性を満たした上で埋め込み限界を高めた提案法をマルチメディア情報ハイディング技術に利用したアプリケーションを検討することを目的としている。

1.4 本論文の構成

第2章にて、代表的な音に関する情報ハイディング法について、その特徴を述べ、第3章にて、蝸牛遅延に基づく情報ハイディング法の埋め込みや検出に関する説明について述べる。第4章にて、蝸牛遅延フィルタ構成の実装方法におけるコンセプト及び、提案法の説明について述べる。第5章にて、提案法を検知不可能性及び頑健性に関して評価を行う。第6章にて、提案法の応用方法として音声ステガノグラフィとして利用した場合を検討する。第7章にて、本論文の結論を述べる。本論文は以上のような構成になっている。

第2章 代表的な音信号への情報ハイディング法

2.1 LSB 置換法

時間領域を操作する最も古典的な手法として、LSB(Least Significant Bit) 置換法がある。LSB 置換法は、音声信号に対して微小な雑音が付加されても音声品質は大きく変化しないという性質に基づき、信号の振幅を表わすビットにおいて最も影響の少ない最下位ビットを所望のビットで置換することで埋め込む手法である。LSB 置換法の代表的な手法として、岩切らによって提案された手法がある [16]。32 kbps ADPCM (Adaptive Differential Pulse Code Modulation) コーデックで符号化された音声波形における予測値の極性変化点に着目している。音質に影響を与えにくい極性の变化幅を指定し、その範囲を満たす音声信号のビットを置換することで情報の埋め込みを行っている。

2.1.1 選択的 LSB 置換法

PCM (Pulse Code Modulation) コーデックにより符号化された音声に対して LSB 置換法を適用すると、振幅の大きいサンプルにおいて歪みが大きくなり音声品質を劣化させることになる。そこで、Aoki は振幅の小さいサンプルのみに LSB 置換法を適用することで信号全体の歪みを低減する選択的 LSB 置換法を提案している [17]。この手法は、フレーム内で振幅の絶対値を昇順にソートし、振幅の小さい一定数のサンプルにのみ埋め込みを行う。また、埋め込みビットレートはフレームにおいて適用するサンプル数を変化させることで任意に設定できる。

2.1.2 選択的 LSB 置換法の改良

これまでの LSB 置換法が原信号のデータ長に依存しているため、埋め込み情報量が限られている。そこで、伊藤らはデータに依存せずに大容量の固定ビットレートを確保できる LSB 置換法を提案している [18]。この手法では、選択的 LSB 置換法と ADPCM を規範とした選択的下位ビット置換法を併用している。ADPCM を規範とした選択的下位ビット置換法は、ADPCM 符号化により量子化値が変化しない範囲は冗長部分であるとして、その範囲内に埋め込みを行う。範囲内であれば音質への影響が小さいため、最下位ビット

のみならず下位 2 ビット目以上にも埋め込みを行う．選択的 LSB 置換法により埋め込みを行う最下位ビットを決定した後に，ADPCM を規範とした選択的下位ビット置換法を用いて下位 2 ビット目以上にも埋め込み可能かを判定することによって，データ依存せずに固定ビットレートを確保している．

2.2 スペクトル拡散法

Boney ら [8] によって提案されたスペクトル拡散法では，原信号の広範囲なスペクトルに情報を埋め込む．埋め込み情報に広帯域の疑似乱数系列を乗じて広帯域化し，原信号の各周波数帯域に分散して埋め込みを行う．再生時には受信系列に対して同じ乱数系列との同期を取り，乗ずれば拡散されていた埋め込み信号が集約される．埋め込み情報が広帯域に分散されているので，たとえ周波数の一部を削除されても再生可能である．

2.3 エコーハイディング法

エコーを用いた手法の代表として，Gruhl らによって提案されたエコーハイディング法 [9] がある．エコーハイディング法では，原信号に，振幅 1 のインパルスと振幅 α のインパルスという二つのインパルスからなるエコーカーネルを畳み込むことにより，人工的にエコーのかかった信号をつくりだしている．データの埋め込みに際しては，エコーまでの時間が違う二種類のエコーカーネルを用意し，エコー入り信号を二種類つくりだす．埋め込むデータ (0 or 1) に応じて，二つのエコー入り信号を切り替えることで，情報埋め込みを実現する．検出の際は，埋め込み情報入りの信号を FFT した後に，その $\log$ をとり，逆 FFT をしてケプストラム領域に持っていき，その自己相関関数を求めると，エコーまでの時間の部分に高いピークが出現する．このピークの位置を検出することで埋め込んだデータを判断する．

2.4 振幅変調法

西村 [10, 19] によって提案された振幅変調法では，原信号を帯域分割し，隣接する帯域ごとに逆位相の正弦振幅変調を与える．これらの帯域を複数含む帯域グループを二つ以上つくり，全てのグループ間の振幅変調位相差を加算したものが透かし情報となる．振幅変動位相差は $0, \pi/2, \pi, 3\pi/2$ の四種類からなり，これが埋め込むデータ (“00”, “01”, “10”, “11”) に対応する．データ検出時には，ペアとなる隣接する帯域信号の振幅包絡をそれぞれ求め，それらの比の対数を振幅変動波形として抽出する．この振幅変動波形を FFT (fast Fourier transform) して帯域グループ間の振幅変動位相差を算出し，それを復号化することで，埋め込まれたデータを抽出する．

2.5 周期的位相変調法

位相変調を利用した手法の代表例として、西村らによって提案された周期的位相変調法 [11] が挙げられる。この手法では、2 次の IIR 型全域通過フィルタの位相特性を 1 サンプル点ごとに变化させ、そこに周期性をもたせることで、周期的な位相変調を原信号にかける。位相特性の変化周期が二種類あり、これが埋め込むデータ (0, 1) に相当する。埋め込みの流れは、原信号に、二種類の周期で周期的位相変調をかけ、二つの周期的位相変調をかけられた信号をつくだす。埋め込むデータ (0 or 1) に応じて、周期的位相変調をかけられた信号を時間フレームで切り出して、それを足し合わせることで、埋め込み情報入りの信号をつくりだす。埋め込んだデータの検出には、埋め込み時と同じ時間フレームで透かし入り信号を切り出し、FFT によって、位相成分を取り出す。

2.6 蝸牛遅延特性に基づく電子音響透かし

エコーハイディング法では、原信号とエコー信号を融合させることで、聴取者ができないことを狙っている。しかし、透かし入り信号をケプストラム領域で自己相関を取るだけで検出ができるため、原信号が無くても、だれでも簡単に情報の取り出しができてしまう。振幅変調法では、確実に検出を行うためには振幅変調の強度を強くする必要があるが、これに起因して音質が劣化してしまう [27]。周期的位相変調法では、人間がゆっくりとした周期での位相変化に対して鈍感であるという特性を利用して、原信号に情報を埋め込んでいる。しかし、この手法では、高い周波数の位相を無作為に変化させている。人間の聴覚は、シンバルやピアノなどのパルス音に似た特徴を持つ音の、高い周波数の位相変化には敏感である。そのため、透かしが入っていることが、気付かれてしまう恐れがある [28, 29]。このように、従来の手法では、有効な情報埋め込みを可能とするが、埋め込み後の音響信号に違和感を感じたり、その存在自体を気づかせる場合がある。そのため、情報埋め込み後の音質の改善ならびに埋め込みを気づかせない頑健な方法を追究する必要がある。

Unoki & Hamada [20] によって蝸牛遅延特性に基づく情報ハイディング法が提案されている。この手法では、蝸牛遅延を模擬した二種類の 1 次 IIR 型全域通過フィルタを用いて埋め込み情報の二値データ (“0”, “1”) に対応してフィルタを切り替えることによって情報の埋め込みを行う。

第3章 蝸牛遅延特性に基づく情報ハイディング法

3.1 蝸牛遅延特性

音波は、外耳道を通して、中耳の耳小骨で機械振動へと変わり、卵円窓にたわみを起こす。卵円窓のたわみが開放されることで、接している蝸牛内部のリンパ液に進行波が生じる。その進行波によって、基底膜が共振振動し、対応する場所に最大変位を起こすことにより、神経発火パルスが発生することで、音は知覚される。このとき、蝸牛は周波数をよりわけのフィルタバンクのような役割をしていると考えられている。基底膜上において、高い周波数成分によって、最大変位を起こす地点は、蝸牛底に近い場所に存在し、低い周波数によって、最大変位を起こす場所は、蝸牛頂に近い場所にある。複数の周波数を含むような音（例えば、インパルス音）では、高い周波数に対応する蝸牛底側が最初に最大変位を起こし、蝸牛頂側に向けて順に最大変位を起こしていく。そのため基底膜上では、高い周波数成分と低い周波数成分との間には、わずかな時間差（低い周波数の遅延）が生じている。この時間差を蝸牛遅延と呼ぶ [30]。

Aiba *et al.* は、この蝸牛遅延と音の同時性判断にどのような関係があるかを調べるため、聴覚心理物理実験により検討を行った。実験では、三つの複合音：(1) 通常（群遅延操作なし）の調波複合音、(2) 基底膜上において蝸牛遅延を打ち消すような群遅延を与えた調波複合音、(3) 蝸牛遅延を増長するような群遅延を与えた調波複合音を用いた。饗庭らは、被験者に対して、通常の調波複合音と (1)、(2)、(3) の調波複合音のいずれか一つを同時に提示し、二つの音の時間差が 0（同時）の状態から時間差を増やしていき、どれくらいの時間差で同時でないと判断できるようになるかを測定した。饗庭らの報告によると、複合音 (2) を用いた場合が同時性判断の精度が最も悪く（一番長い時間差が必要）、複合音 (3) を用いた場合と、複合音 (1) を用いた場合の同時性判断の精度がほぼ同じであった。この結果から、人間の聴覚系において、複合音 (2) のような非現実的な音には敏感であるが、通常の音と蝸牛遅延を増長したような音に対しては鈍感であるということを示唆している [31]。

3.2 蝸牛遅延フィルタ

Unoki & Hamada は、蝸牛遅延に関する知見に基づき、埋め込み情報の二値データ（“0”と“1”）に対応する二種類の異なる蝸牛遅延に似た遅延パターンを原信号に付与することで、検知不可能な情報埋め込みを可能とする情報ハイディング法を提案した [20]。この方法では、1 次の IIR 全域通過フィルタ $H_m(z)$ とその群遅延特性 $\tau_m(\omega)$

$$H_m(z) = \frac{-b_m + z^{-1}}{1 - b_m z^{-1}}, \quad m = 0, 1 \quad (3.1)$$

$$\tau_m(\omega) = -\frac{d \arg(H_m(e^{j\omega}))}{d\omega} \quad (3.2)$$

を利用して、蝸牛遅延特性を模擬し、埋め込み情報の埋め込みとその検出（ノンブラインド検出）を実現している。ただし、 $H_m(e^{j\omega}) = H_m(z)|_{z=e^{j\omega}}$ である。図 3.1 に式 (3.1) を用いた群遅延特性を示す。破線は、実際の蝸牛遅延の 1/10 倍の遅延量を示す。赤線に蝸牛遅延に対して最小二乗法によって最適化したときの最適値 $b_0 = 0.795$ とした時の群遅延特性を示す。また、青線に $b_1 = 0.865$ とした時の群遅延特性を示す。これは、過去の検討結果から検知不可能性と頑健性を満たすため、パラメータ b の値を 0.07 以上離す必要がある [20]。本研究では、これら蝸牛遅延を模擬した群遅延特性を蝸牛遅延フィルタと呼ぶ。提案法では、二種類の蝸牛遅延フィルタを用いて原信号に対して情報を埋め込むことにより、情報ハイディングを実現する。

3.3 情報の埋め込み方法

図 3.2(a) は、蝸牛遅延特性に基づく情報ハイディング法におけるデータ埋め込み処理のブロックダイアグラムを示す。ここでは埋め込み情報を次の手順で埋め込む。まず、埋め込み情報を $s(k) = 10001101\dots$ 、埋め込み処理でのビットレート (bps) を N_{bit} とする。次に、埋め込み情報の二値データ（“0”と“1”）に対応する異なる二つの蝸牛遅延フィルタ ($H_0(z)$ ($b_0 = 0.795$) と $H_1(z)$ ($b_1 = 0.865$)) を用いて原信号 $x(n)$ に、それぞれの群遅延を付与して中間出力 $\omega_0(n)$ と $\omega_1(n)$ を得る。

$$\omega_0(n) = -b_0 x(n) + x(n-1) + b_0 \omega_0(n-1) \quad (3.3)$$

$$\omega_1(n) = -b_1 x(n) + x(n-1) + b_1 \omega_1(n-1) \quad (3.4)$$

$$y(n) = \begin{cases} \omega_0(n), & s(k) = 0 \\ \omega_1(n), & s(k) = 1 \end{cases} \quad (3.5)$$

ただし、 $(k-1)\Delta W < n \leq k\Delta W$, ($k = 1, 2, \dots$) である。ここで、 n はサンプル値、 k はフレーム番号、 $\Delta W = f_s/N_{\text{bit}}$ はフレーム長（フレーム長の半分で重複）、 f_s はサンプリング周波数である。最後に、二値データ系列 $s(k)$ に応じて、各中間出力の荷重和を取ることで情報を埋め込まれた信号 $y(n)$ を得る。なお、各フレーム間で荷重和を取ることによる不連続性の問題を避けるために荷重関数（Hanning 窓関数）を利用した。

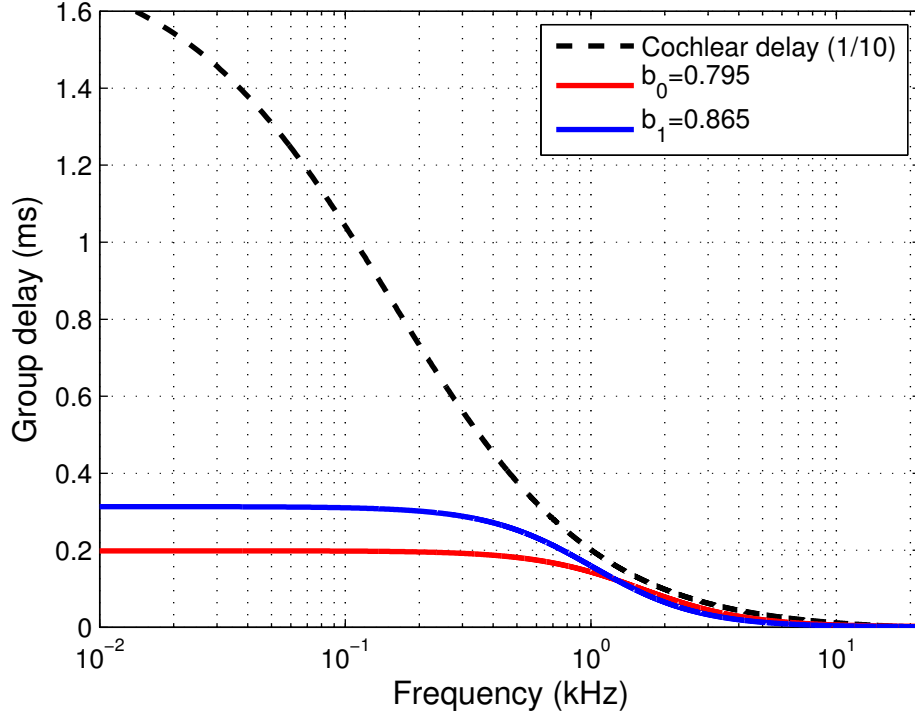


図 3.1: 蝸牛遅延を模擬したフィルタの群遅延特性

3.4 情報の検出方法

図 3.2(b) は，蝸牛遅延特性に基づく情報ハイディング法の検出処理（ノンブラインド法）のブロックダイアグラムを示す．はじめに，原信号 $x(n)$ と情報が埋め込まれた信号 $y(n)$ の位相スペクトルの差分 $\phi(\omega)$ を求める．

$$\phi(\omega_m) = \arg(\text{FFT}[y(n)]) - \arg(\text{FFT}[x(n)]) \quad (3.6)$$

$$\Delta\Phi_0 = \sum_m |\phi(\omega_m) - \arg(H_0(e^{j\omega_m}))| \quad (3.7)$$

$$\Delta\Phi_1 = \sum_m |\phi(\omega_m) - \arg(H_1(e^{j\omega_m}))| \quad (3.8)$$

$$\hat{s}(k) = \begin{cases} 0, & \Delta\Phi_0 < \Delta\Phi_1 \\ 1, & \text{otherwise} \end{cases} \quad (3.9)$$

次に，この差分と埋め込みに利用した蝸牛遅延フィルタ（ $H_0(z)$ と $H_1(z)$ ）の群遅延特性（ $\Delta\Phi_0$ と $\Delta\Phi_1$ ）との差を式 (3.7) と式 (3.8) から求める．最後に，式 (3.9) を利用して，埋め込み情報 $\hat{s}(k)$ を得る．

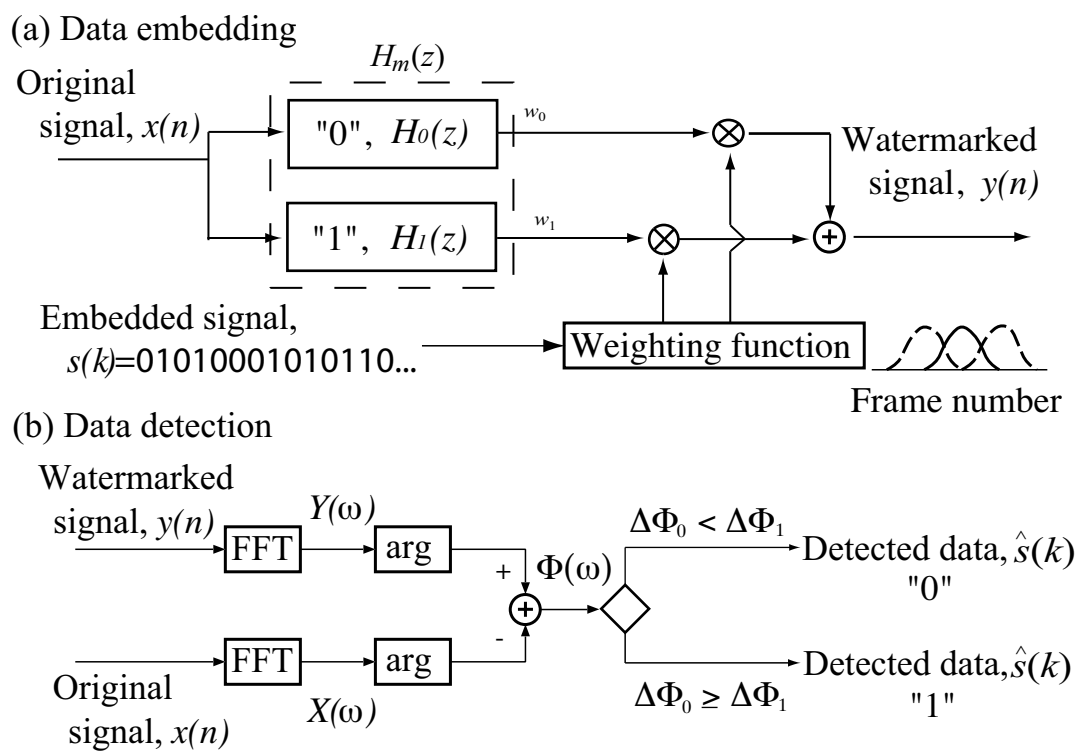


図 3.2: 蝸牛遅延特性に基づく情報ハイディング法におけるデータ埋め込み/検出のブロックダイアグラム

第4章 蝸牛遅延フィルタの構成法

4.1 構成法のコンセプト

埋め込み限界を高める方法としては、(1) 原信号のデータ長を長くすること、(2) ビットレートを上げること、(3) 1 フレーム当りの割り当てビット数を増やすことの三つが考えられる。はじめに、(1) について考えてみる。比較的原信号のデータ長が長くなりやすい音楽信号に比べて、音声信号を原信号と考えると一単語の音声で切り取られてしまうことも考えられ、マルチメディア情報ハイディング技術として埋め込むことが可能な情報量が限られてしまう。次に、(2) について考えてみる。これまでの検討結果から、PSK(Phase Shift Keying) 変調方式を取っているため、ビットレートが上がるにつれて埋め込みビット (“0”, “1”) に対応する位相の切り替え変化によりスペクトル拡散が起こり、位相の変化を知覚されてしまう [25]。最後に、(3) について考えてみる。1 フレーム当りの割り当て可能なビット数を増やすことができれば、原信号のデータ長に依存することなく埋め込み情報量を高めることができる。また、フレームレートを無暗に上げる必要もないため、知覚不可能性が保たれる最大のフレームレートを用いることで、最大埋め込み情報量 (= フレームレート × 1 フレーム当りの割り当てビット数) とすることができる。よって、蝸牛遅延特性に基づく情報ハイディング法の埋め込み情報量を高めるために、1 フレーム当りの割り当て可能なビット数を増やす方法を提案する。

Unoki & Hamada の手法では、二つの蝸牛遅延フィルタを用いて、それぞれに埋め込みビット (“0”, “1”) を割り当てている [20]。そこで、複数の蝸牛遅延フィルタを用いて、それぞれの蝸牛遅延フィルタに複数ビットを割り当てて実装方法のコンセプトとする。複数の蝸牛遅延フィルタを用いるためには、フィルタ構成について検討する必要がある。フィルタ構成方法としては、並列型構成、縦続型構成、並列と縦続を合わせた複合型構成の三つが考えられる。本研究では、蝸牛遅延特性に基づく情報ハイディング法を基本型として、基本型の拡張方法である、並列型フィルタ構成、縦続型フィルタ構成、複合型フィルタ構成に関して埋め込み限界を調べる。

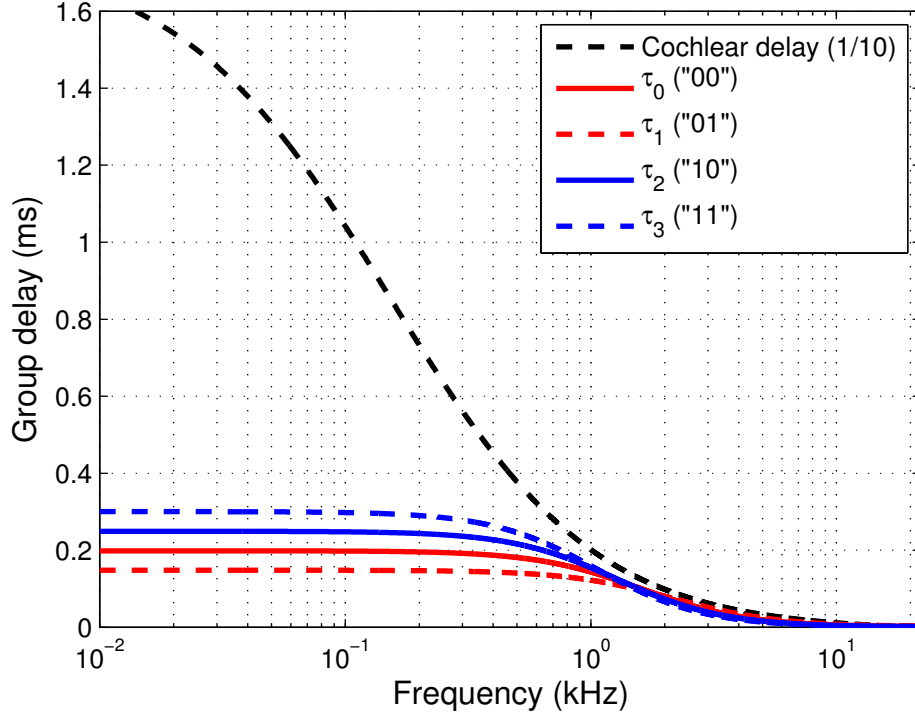


図 4.1: 並列型構成における蝸牛遅延フィルタの群遅延特性

4.2 多段並列型フィルタ構成

多段並列型構成では，式 (3.1) を利用した基本型 ($H_0(z)$ と $H_1(z)$) を，フィルタ数 2 個から $M = 2^N$ 個 ($H_0(z)$, $H_1(z)$, $\dots$, $H_{M-1}(z)$) に拡張し，並列に配置することで，1 フレームあたり N bits の情報埋め込みを可能とした [21]．ここで，多段並列型で構成した蝸牛遅延フィルタ $H_{\text{PrI}}(z)$ とそのフィルタの群遅延 $\tau_m(\omega)$ は，

$$H_{\text{PrI}}(z) := H_m(z) = \frac{-b_m + z^{-1}}{1 - b_m z^{-1}} \quad (4.1)$$

$$\tau_m(\omega) = -\frac{d \arg(H_m(e^{j\omega}))}{d\omega} \quad (4.2)$$

となる ($m = 0, 1, \dots, M-1$)．例えば，1 フレームあたり 2 bits 表現 ($N=2$) とした時は，4 個 ($M = 2^2 = 4$) のフィルタを用いることになる．この場合における蝸牛遅延フィルタの群遅延特性を図 4.1 に示す．

4.2.1 情報の埋め込み方法

図 4.2(a) は，並列型構成におけるデータ埋め込み処理のブロックダイアグラムを示す．ここでは埋め込み情報を次の手順で埋め込む．まず，埋め込み情報を $s(k) = 10001101\dots$,

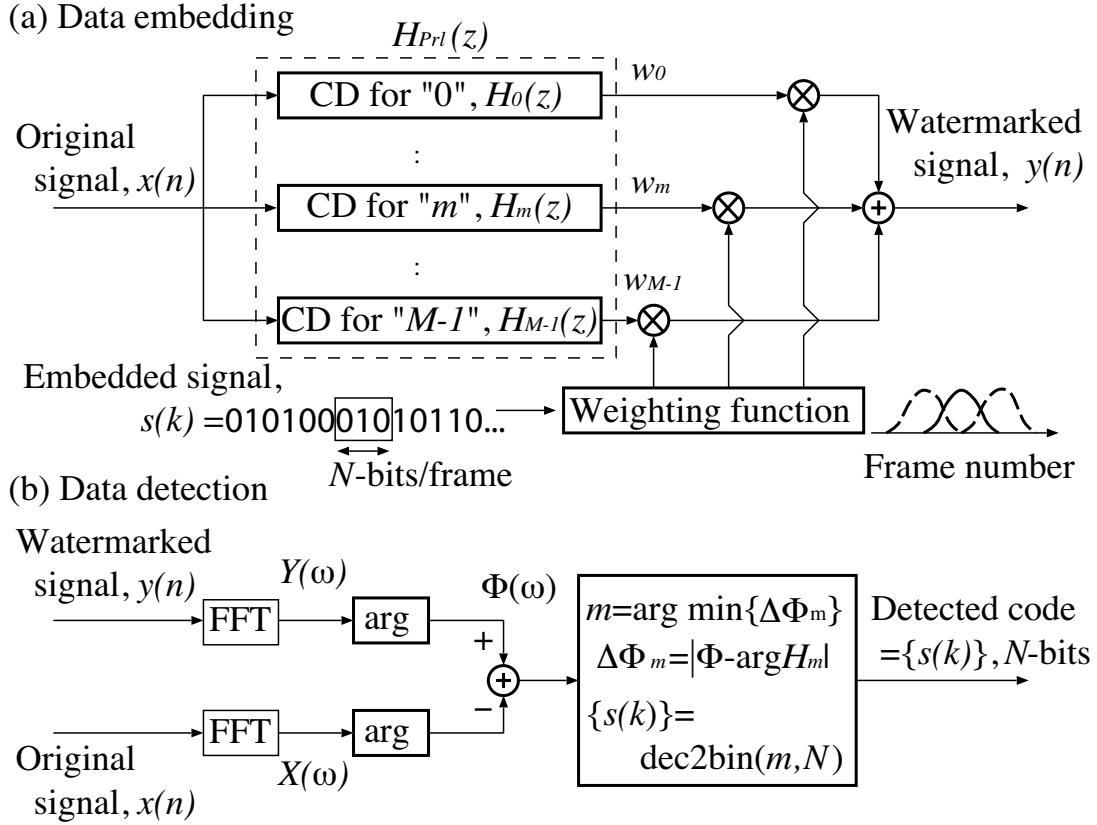


図 4.2: 並列型構成におけるデータ埋め込み/検出のブロックダイアグラム

埋め込み処理でのフレームレート (fps) を N_{fps} とする．次に，1 フレームにおける埋め込みビット ($N\text{-bits/frame}$) に対応する異なる $M - 1$ の蝸牛遅延フィルタ ($H_{\text{PrI}}(z)$) を用いて原信号 $x(n)$ に，それぞれの群遅延を付与して中間出力 $\omega_m(n)$ を得る．

$$\omega_m(n) = -b_m x(n) + x(n-1) + b_m \omega_m(n-1) \quad (4.3)$$

となる ($m = 0, 1, \dots, M - 1$)．ただし， $(k-1)\Delta W < n \leq k\Delta W$, ($k = 1, 2, \dots$) である．ここで， n はサンプル値， k はフレーム番号， $\Delta W = f_s/N_{\text{bits}}$ はフレーム長 (フレーム長の半分で重複)， f_s はサンプリング周波数である．最後に，埋め込み情報の二値データ系列 $s(k)$ に応じて，各中間出力の荷重和を取ることで情報を埋め込まれた信号 $y(n)$ を得る．なお，各フレーム間で荷重和を取ることにによる不連続性の問題を避けるために荷重関数 (Hanning 窓関数) を利用した．

4.2.2 情報の検出方法

データ検出においては，図 4.2(b) に示すように，原信号 $x(n)$ の位相スペクトルと埋め込み情報入り信号 $y(n)$ の位相スペクトルの差分 $\phi(\omega)$ を式 (4.4) から求める．次に，この差分と

埋め込みに利用した蝸牛遅延フィルタ ($H_{\text{PrI}}(z)$) の群遅延特性 ($\phi(\omega)$) との差を式 (4.5) から求める。この差を逐次的に M 個のフィルタの位相特性 ($\arg H_0(z), \arg H_1(z), \dots, \arg H_{M-1}(z)$) と一致処理させて、最も誤差の小さい位相特性を持ったフィルタ番号 m を得ることで、埋め込み情報 $\hat{s}(k)$ を得る [21]。

$$\phi(\omega_q) = \arg(\text{FFT}[y(n)]) - \arg(\text{FFT}[x(n)]) \quad (4.4)$$

$$\Delta\Phi_m = \sum_m |\phi(\omega_m) - \arg(H_{\text{PrI}}(e^{j\omega_m}))| \quad (4.5)$$

例えば、 $N = 2$ の場合、式 (4.4) から位相スペクトルの差分 $\phi(\omega)$ を求める。 $H_{\text{PrI}}(z)$ は、 $H_0(z), H_1(z), H_2(z), H_3(z)$ の 4 通りのいずれかであり、各フィルタ番号は $m = 0, 1, 2, 3$ として割り当てられている。埋め込みビットが “00” であるとする、式 (4.5) から最も誤差の少ない位相特性 $\Delta\Phi_0$ が求まる。このフィルタ番号 $m = 0$ を得ることで、透かし情報 $\hat{s}(k) = \text{“00”}$ を得る。

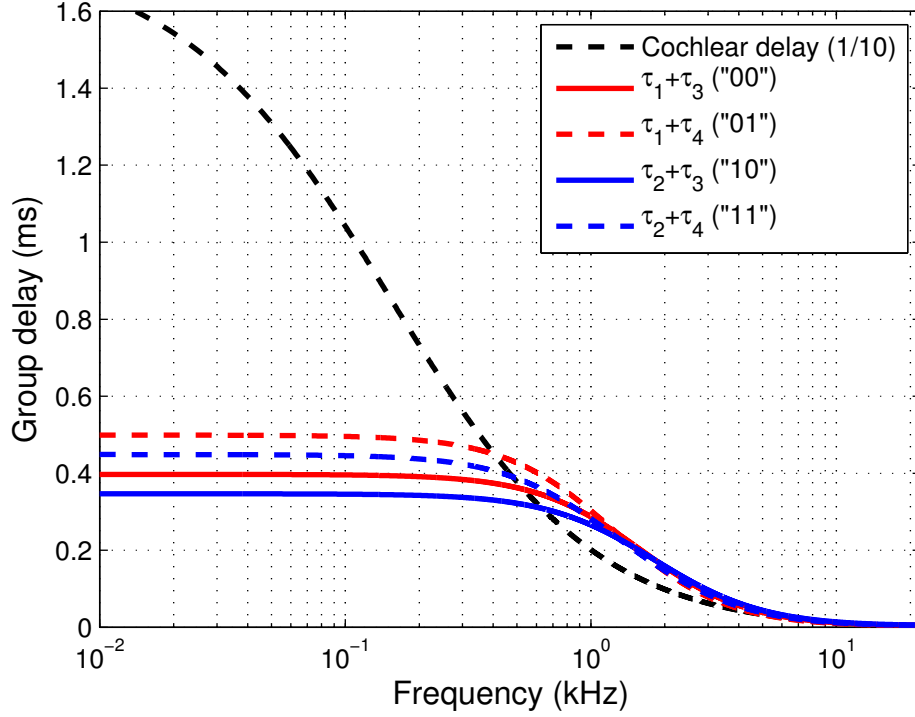


図 4.3: 縦続型構成における蝸牛遅延フィルタの群遅延特性

4.3 多段縦続型フィルタ構成

多段縦続型構成では，式 (3.1) に示す基本型を L 段縦続型にして拡張することで，1 フレーム当たり L bits の情報埋め込みを可能とした [23]．ここで，縦続型で構成される蝸牛遅延フィルタは

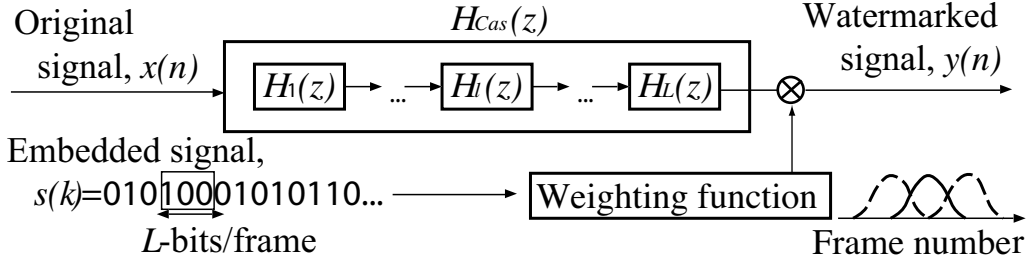
$$H_{\text{Cas}}(z) = \prod_{\ell=1}^L H_{\ell}(z) = \prod_{\ell=1}^L \frac{-b_{\ell} + z^{-1}}{1 - b_{\ell}z^{-1}} \quad (4.6)$$

$$\tau_{\ell}(\omega) = -\frac{d \arg(H_{\ell}(e^{j\omega}))}{d\omega} \quad (4.7)$$

$$\tau_{\text{Cas}}(\omega) = \sum_{\ell=1}^L \tau_{\ell}(\omega) \quad (4.8)$$

となる ($\ell = 1, 2, \dots, L$)．例えば，1 フレーム当り 2 bits 表現 (2 進数で 00 ~ 11) とした時，2 次の蝸牛遅延フィルタ (各フィルタの組み合わせは 4 通りとなる) を 4 個 ($R = 2^{L=2} = 4$) 用いることになる．この場合における蝸牛遅延フィルタの群遅延特性を図 4.3 に示す．

(a) Data embedding



(b) Data detection

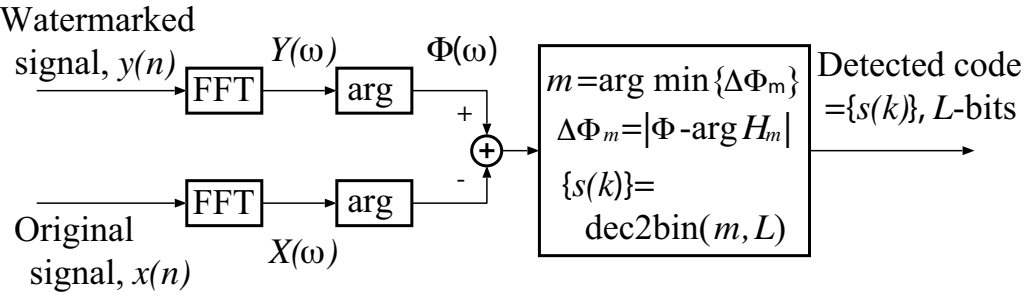


図 4.4: 縦続型構成におけるデータ埋め込み/検出のブロックダイアグラム

4.3.1 情報の埋め込み方法

図 4.4 に縦続型構成における情報埋め込み / 検出処理のブロックダイアグラムを示す．図 4.4(a) の埋め込み処理 ($N = 1$ のときが縦続型となる) では 1 フレームあたり L ビット表現するため，ビット位置に対応づけて L 次の蝸牛遅延フィルタ $H_{\text{Cas}}(z)$ を用いて原信号 $x(n)$ の群遅延量を操作する．まず，埋め込み情報を $s(k) = 10001101\dots$ ，埋め込み処理でのフレームレート (fps) を N_{fps} とする．次に，1 フレームにおける埋め込みビット ($L\text{-bits/frame}$) に対応する異なる L の蝸牛遅延フィルタ ($H_{\text{Cas}}(z)$) を用いて原信号 $x(n)$ に，それぞれの群遅延を付与して中間出力 $\omega_\ell(n)$ を得る．

$$\omega_\ell(n) = -b_\ell x(n) + x(n-1) + b_\ell \omega_\ell(n-1) \quad (4.9)$$

となる ($\ell = 1, 2, \dots, L$)．ただし， $(k-1)\Delta W < n \leq k\Delta W$, ($k = 1, 2, \dots$) である．ここで， n はサンプル値， k はフレーム番号， $\Delta W = f_s / L_{\text{bits}}$ はフレーム長 (フレーム長の半分で重複)， f_s はサンプリング周波数である．最後に，埋め込み情報の二値データ系列 $s(k)$ に応じて，各中間出力の荷重和を取ることで情報を埋め込まれた信号 $y(n)$ を得る．なお，各フレーム間で荷重和を取ることに伴う不連続性の問題を避けるために荷重関数 (Hanning 窓関数) を利用した．

4.3.2 情報の検出方法

データ検出においては，図 4.4(b) に示すように，原信号 $x(n)$ の位相スペクトルと埋め込み情報入り信号 $y(n)$ の位相スペクトルの差分 $\phi(\omega)$ を式 (4.10) から求める．次に，この差分と埋め込みに利用した蝸牛遅延フィルタ ($H_{\text{Cas}}(z)$) の群遅延特性 ($\phi(\omega)$) との差を式 (4.11) から求める．この差を逐次的に $R(= 2^L)$ 個のフィルタの位相特性 ($\arg H_{\text{Cas}}(z)$) と一致処理させて，最も誤差の小さい位相特性を持ったフィルタ番号 ℓ を得ることで，埋め込み情報 $\hat{s}(k)$ を得る．

$$\phi(\omega_q) = \arg(\text{FFT}[y(n)]) - \arg(\text{FFT}[x(n)]) \quad (4.10)$$

$$\Delta\Phi_\ell = \sum_{\ell} |\phi(\omega_\ell) - \arg(H_{\text{Cas}}(e^{j\omega_\ell}))| \quad (4.11)$$

例えば， $L = 2$ の場合，式 (4.10) から位相スペクトルの差分 $\phi(\omega)$ を求める． $H_{\text{Cas}}(z)$ は， $H_1(z) \cdot H_3(z)$, $H_1(z) \cdot H_4(z)$, $H_2(z) \cdot H_3(z)$, $H_2(z) \cdot H_4(z)$ の 4 通りのいずれかであり，各フィルタ番号は $\ell = 1, 2, 3, 4$ として割り当てられている．埋め込みビットが “00” であるとすると，式 (4.11) から最も誤差の少ない位相特性 $\Delta\Phi_1$ が求まる．このフィルタ番号 $\ell = 1$ を得ることで，透かし情報 $\hat{s}(k) = “00”$ を得る．

4.4 多段複合型フィルタ構成

前節では，蝸牛遅延フィルタの構成として，基本型，並列型，縦続型について概説した．その特徴をまとめると次のようになる．基本型では，原則 1 フレームに 1 bit 割り当てた 2 個の異なる蝸牛遅延フィルタの群遅延（1/10 の蝸牛遅延特性のダイナミックレンジ内）を原信号に付与して，透かし情報を埋め込むものであった．並列型構成では，1/10 の蝸牛遅延特性のダイナミックレンジ内に $M = 2^N$ 個のフィルタを構成することで，高い検知不可能性を保持しつつ，1 フレーム内に N ビット表現を可能にしている．一方で，縦続型構成では，蝸牛遅延特性のダイナミックレンジを L 倍（最大 10 倍で実際の蝸牛遅延に相当する）に広げ，この範囲内で 2^L 個のフィルタを構成することで，検知不可能性を保持しつつ，1 フレーム内に L ビット表現を可能にしている．この構成は，フィルタの組み合わせによって，見掛け上パラメータ b_m のとりうる範囲を広くしているようなものである．いずれの方法も基本型の埋め込み情報量を高める目的で設計された．

しかし，式 (3.1) の 1 次 IIR フィルタでは，設計上， $0 < b_m < 1$ の制約条件があり，この範囲内でパラメータ b_m の取り得る配置条件には限界があることがわかっている（各 b_m に対して少なくとも 0.07 以上の差を持つ必要があることがわかっている [20, 21]）．そのため，並列型構成では，無闇に N の数を増やすこと，すなわち構成するフィルタの数 $M = 2^N$ を増やすことはできない．また，縦続型構成では，フィルタ段数 L を増加させていくことで群遅延の変化幅が広がってしまうことで，検知不可能性に影響を与えることもわかっている [24]．そのため，検知不可能性に影響を与えない範囲で，並列型の N と縦続型の L の有効範囲を知りつつ，これら二つのフィルタ構成を有機的に組み合わせることで，1 フレームに複数ビット表現を許す蝸牛遅延フィルタ構成を検討する．ここでは， L 倍のダイナミックレンジ内に $2^{N \cdot L} = M^L$ 個の蝸牛遅延フィルタの群遅延特性を表現できる多段複合型構成を提案する．

4.4.1 情報の埋め込み方法

複合型構成は，式 (4.7) に示す縦続型 ($H_{\text{Cas}}(z)$) の各段のフィルタを並列型 (2^N 個) に拡張することで，1 フレーム当たり $L \cdot N$ bits の割り当てを可能とする．複合型構成における埋め込み処理を図 4.5(a) に示す．ここで，多段複合型で構成した蝸牛遅延フィルタ $H_{\text{Cmp}}(z)$ とそのフィルタの群遅延 $\tau_{\text{Cmp}}(\omega)$ は，

$$H_{\text{Cmp}}(z) := \prod_{\ell=1}^L H_{\ell,m}(z) = \prod_{\ell=1}^L \frac{-b_{\ell,m} + z^{-1}}{1 - b_{\ell,m}z^{-1}} \quad (4.12)$$

$$\tau_{\ell,m}(\omega) = -\frac{d \arg(H_{\ell,m}(e^{j\omega}))}{d\omega} \quad (4.13)$$

$$\tau_{\text{Cmp}}(\omega) = \sum_{\ell=1}^L \tau_{\ell,m}(\omega) \quad (4.14)$$

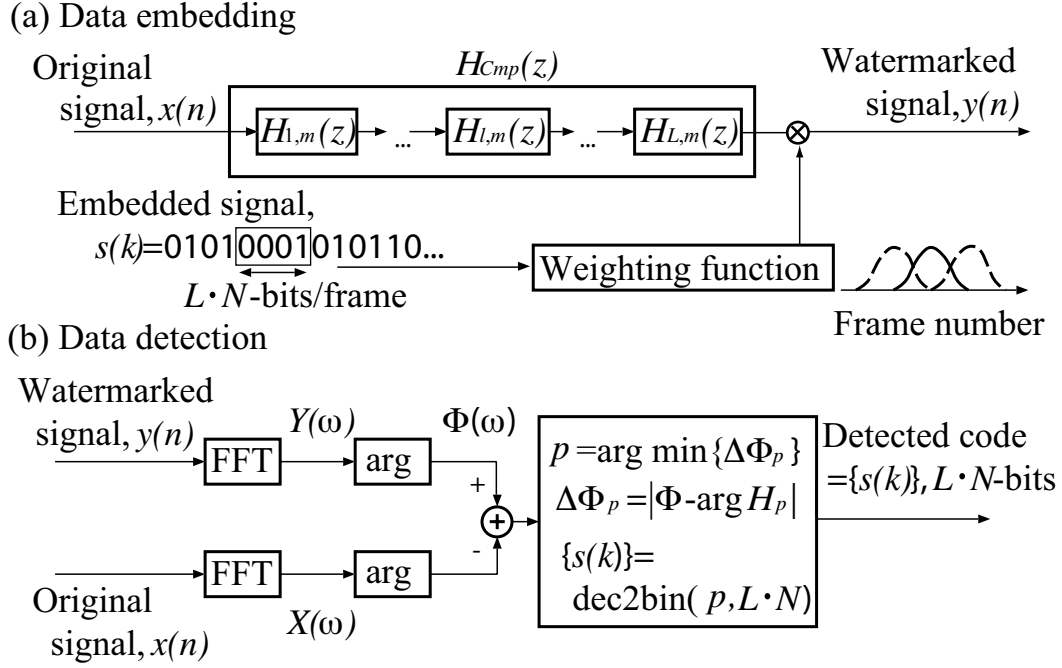


図 4.5: 複合型構成におけるデータ埋め込み/検出のブロックダイアグラム

となる．ただし， $\ell = 1, 2, \dots, L$ と $m = 0, 1, \dots, M - 1$ である．例えば， $N = 2$ の並列型と $L = 2$ の縦続型で複合型を構成すると，合計 $2^{2 \cdot 2} = 16$ 個のフィルタを用いることになり，1 フレーム当たり，4 bits 表現が可能となる．

図 4.6 に，この例の複合型構成をとる蝸牛遅延フィルタの群遅延特性を示す．例えば，埋め込みビットが“1011”であるなら，先の 2 ビット“10”に対応する $H_{1,2}(z)$ と後の 2 ビット“11”に対応する $H_{2,3}(z)$ をそれぞれ選択する．この二つのフィルタを掛け合わせ， $H_{1,2}(z) \cdot H_{2,3}(z)$ のフィルタ出力（群遅延特性は図 4.6 の $\tau_{1,2} + \tau_{2,3}$ に対応）に対して荷重 1 を，それ以外は荷重 0 とする．同様に，“1100”なら， $H_{1,3}(z) \cdot H_{2,0}(z)$ のフィルタ出力に対して荷重 1 を，それ以外は荷重 0 とする．このように，埋め込みビットに対応したフィルタ出力の荷重和を取り，出力信号 $y(n)$ を得る．なお，ここではフレーム間の不連続性の問題を避けるために荷重関数（Hanning 窓）を利用している．

4.4.2 情報の検出方法

図 4.5(b) に示すように， $x(n)$ の位相スペクトルと $y(n)$ の位相スペクトルの差分 $\phi(\omega)$ を式 (4.15) から求める．次に，この差分と埋め込みに利用した蝸牛遅延フィルタ ($H_{\text{Cmp}}(z)$) の群遅延特性 ($\phi(\omega)$) との差を式 (4.16) から求める．最後に，最も誤差の少ない位相特性 ($\Delta\Phi_p$) をもったフィルタ番号 p を得ることで，透かし情報 $\hat{s}(k)$ を得る．

$$\phi(\omega_q) = \arg(\text{FFT}[y(n)]) - \arg(\text{FFT}[x(n)]) \quad (4.15)$$

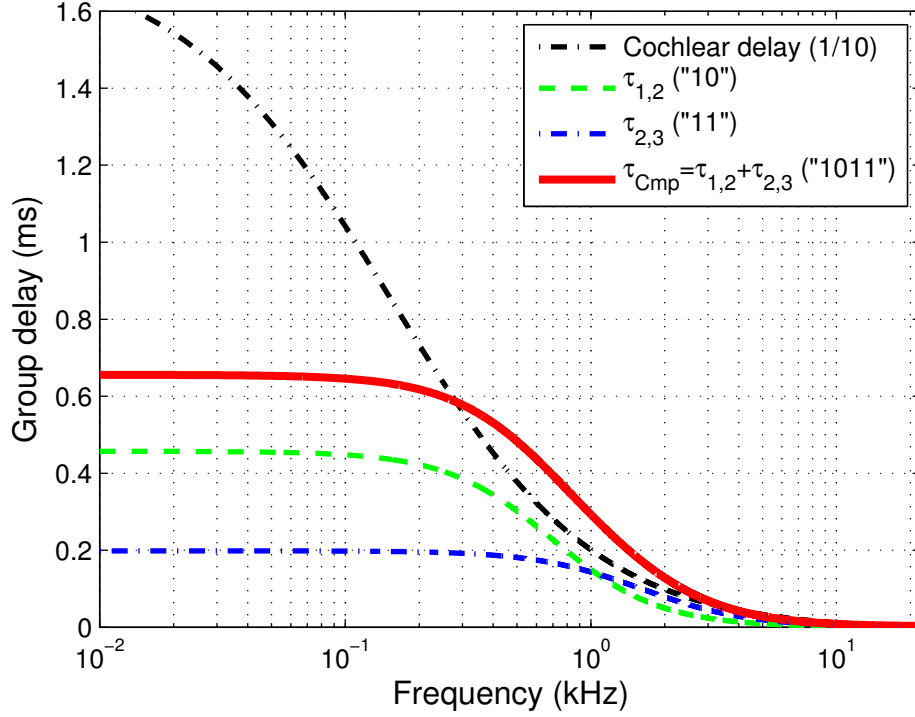


図 4.6: 複合型構成 ($L = 2, N = 2$) における蝸牛遅延フィルタの群遅延特性

$$\Delta\Phi_p = \sum_q |\phi(\omega_q) - \arg(H_{\text{Cmp}}(e^{j\omega_q}))| \quad (4.16)$$

例えば, $N = 2, L = 2$ の複合型構成の場合, 式 (4.15) から位相スペクトルの差分 $\phi(\omega)$ を求める. $H_{\text{Cmp}}(z)$ は, $H_{1,0}(z) \cdot H_{2,0}(z), H_{1,0}(z) \cdot H_{2,1}(z), \dots, H_{1,3}(z) \cdot H_{2,3}(z)$ の 16 通りのいずれかであり, 各フィルタ番号は $p = 0, 1, \dots, 15$ として割り当てられている. 埋め込みビットが “0011” であるとする, 式 (4.16) から最も誤差の少ない位相特性 $\Delta\Phi_3$ が求まる. このフィルタ番号 $p = 3$ を得ることで, 透かし情報 $\hat{s}(k) = \text{“0011”}$ を得る.

第5章 提案法の評価

5.1 音楽信号に関する客観評価

音楽信号における検知不可能性に関する検討と同様に、音質評価に関する代表的な客観評価実験を行う。特に、複合型構成を用いて、透かし情報の埋め込みによって生じる歪みが、埋め込み速度やフィルタ構成 (L, N の構成) によってどのように変化するか客観評価実験により系統的に検討する。

客観評価実験では、RWC 音楽データベース [32] の全 102 曲を評価用の原音 (サンプリング周波数 44.1 kHz, 16 ビット量子化) として利用する。ここでは、冒頭 10 秒間を元曲として、8 文字の情報 (“AIS-Lab.”) を透かし情報として各原音に埋め込む。また、1 秒間に 4 フレーム (以後、4 fps (frame per sec)) をベースとし、12 条件の fps (4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192 fps) で、透かし情報を原音の両チャンネルに埋め込み、その特性評価を行う。埋め込み bits 数の計算については、1 フレームに $N \cdot L$ bits の割り当てとなることから、フレームレートに割り当てた bits 数を乗じたものがビットレート (bps) となる。音質評価に関しては、前報 [22] に基づき、オーディオ信号に対する知覚評価尺度 (PEAQ) [33] と対数スペクトル歪尺度 (LSD) を利用する。また、透かし情報の検出率を調べるためビット検出率を利用する。並列型構成 ($N = 1, 2, 3, 4$) における評価結果を図 5.1 に、縦続型構成 ($L = 1, 2, 3, 4$) における評価結果を図 5.2 に、複合型構成における評価結果を図 5.3 にそれぞれ示す。複合型のフィルタ構成は、1 フレームで 4 bits 表現できる ($(L, N) = (1, 4), (4, 1), (2, 2)$) 三つについて検討を行う。

5.1.1 多段並列型における客観評価

はじめに、PEAQ による客観評価の結果 (図 5.1(a)) について検討する。図中の結果は 102 曲に対する PEAQ の平均値である。PEAQ の ODG (objective difference grade) 値は 0 (知覚不可能) $\sim$ -4 (非常に耳障り) であるため、 -1 (知覚される可能性があるが耳障りではない) を検知不可能性の評価閾値と定めた。図 5.1(a) をみると、フィルタの数が増加するとともに ODG 値が低下していることがわかる。 $L = 1, N = 1$ の場合 ODG 値が一番高く、2048 fps が評価閾値となっている。 $L = 1, N = 2$ の場合では、512 fps が評価閾値となっている。 $L = 1, N = 3$ の場合では、256 fps が評価閾値となっている。 $L = 1, N = 4$ の場合では、128 fps が評価閾値となっている。

次に，LSD の結果（図 5.1(b)）について検討する．一般に LSD は 1 dB 内の歪みであれば，音質がよいと言われているため，LSD の評価閾値を 1 dB に設定した．PEAQ の時と同様に図 5.1(b) では，フィルタの数が増加するとともに LSD の値が上昇していることがわかる． $L = 1, N = 1$ の場合全てのフレームレートにおいて LSD の値が評価閾値内にあり，8192 fps が評価閾値となっている．その他のフィルタ構成（ $L = 1, N = 2$ と $L = 1, N = 3$ と $L = 1, N = 4$ ）では，4096 fps が評価閾値となっている．

最後に，埋め込み情報のビット検出の結果（図 5.1(c)）について検討する．ここでは，ビット検出率 75% を評価閾値とした．図 5.1(c) では，フィルタの数が増加するとともに検出率が低下していることがわかる．評価閾値は， $L = 1, N = 4$ の時，256 fps となっている．その他のフィルタ構成（ $L = 1, N = 1$ と $L = 1, N = 2$ と $L = 1, N = 3$ ）では，512 fps が評価閾値となっている．

5.1.2 多段縦続型における客観評価

はじめに，PEAQ による客観評価の結果（図 5.2(a)）について検討する．図 5.2(a) をみると，フィルタの数が増加するとともに ODG 値が低下していることがわかる． $L = 1, N = 1$ の場合 ODG 値が一番高く，2048 fps が評価閾値となっている． $L = 2, N = 1$ の場合では，512 fps が評価閾値となっている． $L = 3, N = 1$ の場合では，256 fps が評価閾値となっている． $L = 4, N = 1$ の場合では，64 fps が評価閾値となっている．

次に，LSD の結果（図 5.2(b)）について検討する．PEAQ の時と同様に図 5.2(b) では，フィルタの数が増加するとともに LSD の値が上昇していることがわかる． $L = 1, N = 1$ の場合全てのフレームレートにおいて LSD の値が評価閾値内にあり，8192 fps が評価閾値となっている．その他のフィルタ構成（ $L = 2, N = 1$ と $L = 3, N = 1$ と $L = 4, N = 1$ ）では，4096 fps が評価閾値となっている．

最後に，埋め込み情報のビット検出の結果（図 5.2(c)）について検討する．図 5.2(c) では，並列型と同様にフィルタの数が増加するとともに検出率が低下していることがわかる． $L = 1, N = 1$ と $L = 2, N = 1$ の場合では，512 fps が評価閾値となっている． $L = 3, N = 1$ の場合では，256 fps が評価閾値となっている． $L = 4, N = 1$ の場合では，128 fps が評価閾値となっている．

5.1.3 多段複合型における客観評価

はじめに，PEAQ による客観評価の結果（図 5.3(a)）について検討する．図 5.3(a) をみると， $L = 1, N = 4$ の場合 ODG 値が一番高く，128 fps が評価閾値となっている． $L = 4, N = 1$ と $L = 2, N = 2$ の場合では，64 fps が評価閾値となっている．

次に，LSD の結果（図 5.3(b)）について検討する． $L = 1, N = 4$ と $L = 4, N = 1$ の場合では，4096 fps が評価閾値となっている． $L = 2, N = 2$ の場合では，2048 fps が評価閾値となっている．

最後に、埋め込み情報のビット検出の結果(図 5.3(c))について検討する。 $L = 1, N = 4$ と $L = 2, N = 2$ の場合では、256 fps が評価閾値となっている。 $L = 4, N = 1$ の場合では、128 fps が評価閾値となっている。

以上の結果から、並列型構成では $N = 2$ が最良であることがわかる。縦続型構成では $L = 2$ が最良であることがわかる。複合型構成では PEAQ の結果から $L = 2, N = 2$ の場合では検知不可能性を保持しつつ、最大埋め込み情報量を 256 bps (64 fps の時 1 フレームで 4 bits 表現) に向上できた。現状提案している多段フィルタ構成の中では、 $L = 1, N = 2$ と $L = 2, N = 1$ が一番最適な組み合わせだと言える。ここで、このフィルタ構成における最大埋め込み情報量は、1024 bps (512 fps の時 1 フレームで 2 bits 表現) である。

5.2 音声信号に関する客観評価

音声信号と音楽信号の大きな違いは、サンプリング周波数や構成されるスペクトル成分である。音声信号は音楽信号に比べてエネルギーが時間/周波数的に疎であるため、情報を埋め込む余地は少ない。提案法は、1 次 IIR 全域通過フィルタを用いているため、サンプリング周波数が異なる音声信号に対して音楽信号と同様のパラメータを用いてしまうとより大きな位相の変化が起こり検知不可能性を満たすことができなくなってしまう。そこで、音声信号に対する適切なパラメータを決める必要がある。ここでは、蝸牛遅延フィルタのパラメータ決定方法について述べる。適切なパラメータを与えた上で音声信号における検知不可能性に関する検討と同様に、音質評価に関する代表的な客観評価実験を行う。

5.2.1 多段並列型における客観評価

客観評価実験では、ATR 音声データベース [34] B セットの内 10 秒程度の音声 12 個を評価用の原音声(サンプリング周波数 16 kHz, 16 ビット量子化)として利用する。音質評価に関しては、符号化音声の評価で利用される Perceptual Evaluation of Speech Quality (PESQ) [35] と対数スペクトル歪 (LSD) を用いる。PESQ は ODG (−0.5 ~ 4.5 まで 5 段階で表現され、−0.5 で MOS 値が 1, 4.5 で MOS 値が 5 となる) で表現されているため、評価閾値を 3.5 (知覚される可能性があるが耳障りではない) と定める。LSD に関しては、音楽信号に関する客観評価と同様に 1 dB 以内の歪み (一般に、合成音の音質評価として良好と考えられている範囲) を検知不可能性の評価閾値と定める。ビット検出率の評価閾値は、75% とする。ビットレートは、音楽信号に関する客観評価と同様に $N_{\text{bit}} = 4$ を基本とし、特性評価のためビットレートを変化させ ($N_{\text{bit}} = 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192$ bps の全 12 条件)、に埋め込み情報である 8 文字の文字情報 (“AIS-Lab.”) を埋め込む。これらの条件で、客観評価実験を行う。

音声信号に対する多段並列型構成における客観評価結果を図 5.4 に示す。はじめに、PESQ による客観評価の結果(図 5.4(a)) について検討する。図中の結果は 12 個の音声信号に対する PESQ の平均値である。図 5.4(a) をみると、フィルタの数が増加するとともに ODG

値が低下していることがわかる． $L = 1, N = 1$ と $L = 1, N = 2$ の場合 ODG 値が一番高く，256 fps が評価閾値となっている． $L = 1, N = 3$ の場合では，128 fps が評価閾値となっている． $L = 1, N = 4$ の場合では，64 fps が評価閾値となっている．

次に，LSD の結果（図 5.4(b)）について検討する．いずれのフィルタ構成においても，4096 fps が評価閾値となっている．

最後に，埋め込み情報のビット検出の結果（図 5.4(c)）について検討する．図 5.4(c) では，フィルタの数が増加するとともに検出率が低下していることがわかる． $L = 1, N = 4$ の場合では，256 fps が評価閾値となっている．その他のフィルタ構成（ $L = 1, N = 1$ と $L = 1, N = 2$ と $L = 1, N = 3$ ）では，512 fps が評価閾値となっている．

5.2.2 多段縦続型における客観評価

音声信号に対する多段縦続型構成における客観評価結果を図 5.5 に示す．はじめに，PESQ による客観評価の結果（図 5.5(a)）について検討する．図 5.5(a) をみると，フィルタの数が増加するとともに ODG 値が低下していることがわかる． $L = 4, N = 1$ の場合では，64 fps が評価閾値となっている．その他のフィルタ構成（ $L = 1, N = 1$ と $L = 2, N = 1$ と $L = 3, N = 1$ ）では，128 fps が評価閾値となっている．

次に，LSD の結果（図 5.5(b)）について検討する． $L = 4, N = 1$ の場合では，2048 fps が評価閾値となっている．その他のフィルタ構成（ $L = 1, N = 1$ と $L = 2, N = 1$ と $L = 3, N = 1$ ）では，4096 fps が評価閾値となっている．

最後に，埋め込み情報のビット検出の結果（図 5.5(c)）について検討する．図 5.5(c) では，並列型と同様にフィルタの数が増加するとともに検出率が低下していることがわかる． $L = 1, N = 1$ の場合では，512 fps が評価閾値となっている． $L = 2, N = 1$ の場合では，256 fps となっている． $L = 3, N = 1$ と $L = 4, N = 1$ の場合では，128 fps が評価閾値となっている．

5.2.3 多段複合型における客観評価

音声信号に対する多段複合型構成における客観評価結果を図 5.6 に示す．はじめに，PESQ による客観評価の結果（図 5.6(a)）について検討する．図 5.6(a) をみると，全てのフィルタ構成において，64 fps が評価閾値となっている．

次に，LSD の結果（図 5.6(b)）について検討する． $L = 4, N = 1$ の場合では，2048 fps が評価閾値となっている． $L = 1, N = 4$ と $L = 2, N = 2$ の場合では，4096 fps が評価閾値となっている．

最後に，埋め込み情報のビット検出の結果（図 5.6(c)）について検討する． $L = 1, N = 4$ と $L = 2, N = 2$ の場合では，256 fps が評価閾値となっている． $L = 4, N = 1$ の場合では，128 fps が評価閾値となっている．

以上の結果から，音声信号における並列型構成では $N = 2$ が最良であることがわかる．縦続型構成では $L = 3$ が最良であることがわかる．複合型構成では PESQ の結果から $L = 2, N = 2$ の場合では検知不可能性を保持しつつ，最大埋め込み情報量を 256 bps (64 fps の時 1 フレームで 4 bits 表現) に向上できた．現状提案している音声信号における多段フィルタ構成の中では， $L = 1, N = 2$ が一番最適な組み合わせだと言える．ここで，このフィルタ構成における最大埋め込み情報量は，512 bps (256 fps の時 1 フレームで 2 bits 表現) である．

5.3 耐性評価

頑健性に関する客観評価実験を行う．音楽信号における耐性評価は，ダウンサンプリング，ビット伸縮，MP3 情報圧縮の三つの信号処理を行う．音声信号における耐性評価は，PCM，ADPCM，CS-ACELP の三つの音声符号化を用いて行う．また，音楽信号のための Stirmark ベンチマークを用いて耐性評価を行う．

5.3.1 音楽信号における耐性評価

耐性試験では，ダウンサンプリング (20 kHz，16 kHz，8 kHz の三つ) と，ビット伸縮 (24 bit，8 bit の二つ)，MP3 情報圧縮 (128 kbps，96 kbps，64 kbps-mono の三つ) の三つの信号処理操作を行う．

表 5.1 に基本型，並列型，縦続型，複合型構成をとる透かし情報の埋め込み / 検出処理に関する耐性評価結果を示す．表 5.1 では，各フィルタ構成におけるビット検出率が 75% を上回るときの最大フレームレート (fps) を示している．並列型構成の結果をみると $L = 1, N = 2$ では mp3 96 kbps 情報圧縮に対する 64 fps がすべての評価項目の中で下限値となった．これに対し， $L = 1, N = 3$ では，mp3 96 kbps 情報圧縮に対する 64 fps が同様に下限値となった．一方， $L = 1, N = 4$ では，mp3 96 kbps 情報圧縮に対する 32 fps が下限値となった．次に縦続型構成の結果を見ると， $L = 2, N = 1$ では mp3 96 kbps 情報圧縮に対する 64 fps が下限であった． $L = 3, N = 1$ では，いずれの mp3 情報圧縮において 32 fps が下限であった． $L = 4, N = 1$ では，mp3 96, 64 kbps 情報圧縮においてビット検出率が 75% を上回ることがなく，頑健性を保証できる結果にはなっていない．この場合，表には ‘—’ と記した．この条件ではすべての項目をパスしたことになるため，頑健性を保障できる結果にはなっていない．最後に，複合型構成の結果を見ると，mp3 96 kbps 情報圧縮に対して 64 fps が下限値となった．

以上の結果から，並列型構成では $N = 3$ が最良であることがわかる．縦続型構成では $L = 2$ が最良であることがわかる．複合型構成では $L = 2, N = 2$ が最良であることがわかる．現状提案している多段フィルタ構成の中では， $L = 2, N = 2$ が一番最適な組み合わせだと言える．ここで，このフィルタ構成における最大埋め込み情報量は，256 bps (64 fps の時 1 フレームで 4 bits 表現) である．

5.3.2 音声信号における耐性評価

ディジタル音声通信での利用を想定した耐性評価試験を行う。文字情報 (“AIS-Lab.”) を埋め込んだ音声信号に対して音声通信で最も一般的に用いられている音声符号化方式 PCM コーデック (G.711) [36] で符号化・複合化を行い、提案法を用いてビット検出を行う。また、PCM コーデックの他に音声符号化方式である ADPCM 32-kbps (G.726) [37] と CS-ACELP (G.729) [38] を用いる。

まず、並列型構成における音声符号化耐性試験結果を図 5.7 に示す。図 5.7(a) に文字情報を埋め込んだ音声信号に対して PCM コーデックで符号化・複合化を行い、並列型構成を用いてビット検出した結果を示す。図 5.7(a) では、フィルタの数が増加するとともに検出率が低下していることがわかる。 $L = 1, N = 1$ の場合では、512 fps が評価閾値となっている。 $L = 1, N = 2$ と $L = 1, N = 3$ の場合では、256 fps が評価閾値となっている。 $L = 1, N = 4$ の場合では、512 fps が評価閾値となっている。図 5.7(b) に ADPCM コーデックで符号化・複合化を行い、並列型構成を用いてビット検出した結果を示す。図 5.7(b) では、PCM コーデックにおける結果と同様にフィルタの数が増加するとともに検出率が低下していることがわかる。 $L = 1, N = 1$ の場合では、256 fps が評価閾値となっている。 $L = 1, N = 2$ と $L = 1, N = 3$ の場合では、128 fps が評価閾値となっている。 $L = 1, N = 4$ の場合では、64 fps が評価閾値となっている。図 5.7(c) に CS-ACELP コーデックで符号化・複合化を行い、並列型構成を用いてビット検出した結果を示す。いずれのフィルタ構成においても、評価閾値を上回ることはいない。

次に、縦続型構成における音声符号化耐性試験結果を図 5.8 に示す。図 5.8(a) に文字情報を埋め込んだ音声信号に対して PCM コーデックで符号化・複合化を行い、縦続型構成を用いてビット検出した結果を示す。図 5.8(a) では、フィルタの数が増加するとともに検出率が低下していることがわかる。 $L = 1, N = 1$ の場合では、512 fps が評価閾値となっている。 $L = 2, N = 1$ と $L = 3, N = 1$ の場合では、128 fps が評価閾値となっている。 $L = 4, N = 1$ の場合では、64 fps が評価閾値となっている。図 5.8(b) に ADPCM コーデックで符号化・複合化を行い、縦続型構成を用いてビット検出した結果を示す。図 5.8(b) では、PCM コーデックにおける結果と同様にフィルタの数が増加するとともに検出率が低下していることがわかる。 $L = 1, N = 1$ の場合では、256 fps が評価閾値となっている。 $L = 2, N = 1$ の場合では、128 fps が評価閾値となっている。 $L = 3, N = 1$ の場合では、64 fps が評価閾値となっている。 $L = 4, N = 1$ の場合では、32 fps が評価閾値となっている。図 5.8(c) に CS-ACELP コーデックで符号化・複合化を行い、縦続型構成を用いてビット検出した結果を示す。いずれのフィルタ構成においても、評価閾値を上回ることはいない。

最後に、複合型構成における音声符号化耐性試験結果を図 5.9 に示す。図 5.9(a) に文字情報を埋め込んだ音声信号に対して PCM コーデックで符号化・複合化を行い、複合型構成を用いてビット検出した結果を示す。図 5.9(a) では、フィルタの数が増加するとともに検出率が低下していることがわかる。 $L = 1, N = 4$ と $L = 2, N = 2$ の場合では、128 fps が評価閾値となっている。 $L = 4, N = 1$ の場合では、64 fps が評価閾値となっている。図

5.9(b) に ADPCM コーデックで符号化・複合化を行い，複合型構成を用いてビット検出した結果を示す． $L = 1, N = 4$ と $L = 2, N = 2$ の場合では，64 fps が評価閾値となっている． $L = 4, N = 1$ の場合では，32 fps が評価閾値となっている．図 5.9(c) に CS-ACELP コーデックで符号化・複合化を行い，複合型構成を用いてビット検出した結果を示す．いずれのフィルタ構成においても，評価閾値を上回ることはいない．

以上の結果から，提案法では位相領域に対して情報の埋め込みを行っているため，固定通信に用いられる PCM や ADPCM コーデックのような波形符号化に対しては位相情報が保持されるため，耐性があることがわかる．しかし，主に移動通信に用いられる CS-ACELP コーデックのような位相情報を考慮していない符号化方式に対しては，位相スペクトルを算出できないため埋め込み情報の検出が行えず，耐性がないことがわかる．PCM コーデックに対しては， $L = 1, N = 2$ が最良であることがわかる．このフィルタ構成における最大埋め込み情報量は，512 bps (256 fps の時 1 フレームで 2 bits 表現) である．ADPCM コーデックに対しては， $L = 1, N = 3$ が最良であることがわかる．このフィルタ構成における最大埋め込み情報量は，384 bps (128 fps の時 1 フレームで 3 bits 表現) である．

5.4 StirMark ベンチマーク

StirMark ベンチマークとは，電子音響透かしのためのベンチマークテストプログラムであり，StirMark Benchmark for Audio version 1.3.2 (SMBA) を用いる [39]．表 5.2 に StirMark ベンチマークにおける 35 種類の信号処理内容を七つのカテゴリーごとに分けたものを示す：(i) Noise：雑音を加えるもの，(ii) Amplitude：振幅値を操作するもの，(iii) Bit：ビット処理を行うもの，(iv) Data：データ置換処理を行うもの，(v) Filtering：フィルタリング処理を行うもの，(vi) Phase：位相操作を行うもの，(vii) Echo：残響を加えるもの．

図 5.10 に並列型構成におけるベンチマークテストの結果を示す．図 5.10(a) ~ (c) の結果を見ると，(i) Noise，(ii) Amplitude，(iii) Bit，(v) Filtering におけるビット検出率が評価閾値の 75% を上回っており，(iv) Data，(vi) Phase，(vii) Echo におけるビット検出率は評価閾値の 75% を下回っていることがわかる．図 5.10(d) の結果を見ると，(i) Noise，(ii) Amplitude，(iii) Bit におけるビット検出率が評価閾値の 75% を上回っており，(iv) Data，(v) Filtering，(vi) Phase，(vii) Echo におけるビット検出率は評価閾値の 75% を下回っていることがわかる．

図 5.11 に縦続型構成におけるベンチマークテストの結果を示す．図 5.11(a) ~ (d) の結果を見ると，(i) Noise，(ii) Amplitude，(iii) Bit，(v) Filtering におけるビット検出率が評価閾値の 75% を上回っており，(iv) Data，(vi) Phase，(vii) Echo におけるビット検出率は評価閾値の 75% を下回っていることがわかる．

図 5.12 に複合型構成におけるベンチマークテストの結果を示す．図 5.12(a)，(c) の結果を見ると，(i) Noise，(ii) Amplitude，(iii) Bit におけるビット検出率が評価閾値の 75% を上回っており，(iv) Data，(v) Filtering，(vi) Phase，(vii) Echo におけるビット検出

率は評価閾値の 75% を下回っていることがわかる．図 5.12(b) の結果を見ると，(i) Noise ，(ii) Amplitude ，(iii) Bit ，(v) Filtering におけるビット検出率が評価閾値の 75% を上回っており，(iv) Data ，(vi) Phase ，(vii) Echo におけるビット検出率は評価閾値の 75% を下回っていることがわかる．

以上の結果から，提案法は雑音を加えられる処理，振幅操作，ビット処理，フィルタリング処理に関しては頑健性が保証できることがわかった．しかし，データ置換処理，位相操作，残響を加えられる処理に対しては頑健性が保証できないことがわかった．データ置換処理に関しては，透かし入り信号のデータ長を変える処理が多く，情報が埋め込まれているフレームがずれてしまい，誤検出が多くでたと考えられる．位相操作処理に関しては，提案法では位相領域に情報埋め込みを行っているため，位相を操作されてしまうとうまく検出が行えず，検出率が 75% を下回ったと考えられる．残響を加える処理に関しては，位相操作処理と同様の理由で，位相が残響の影響を受けてしまいうまく検出が行えず，検出率が 75% を下回ったと考えられる．

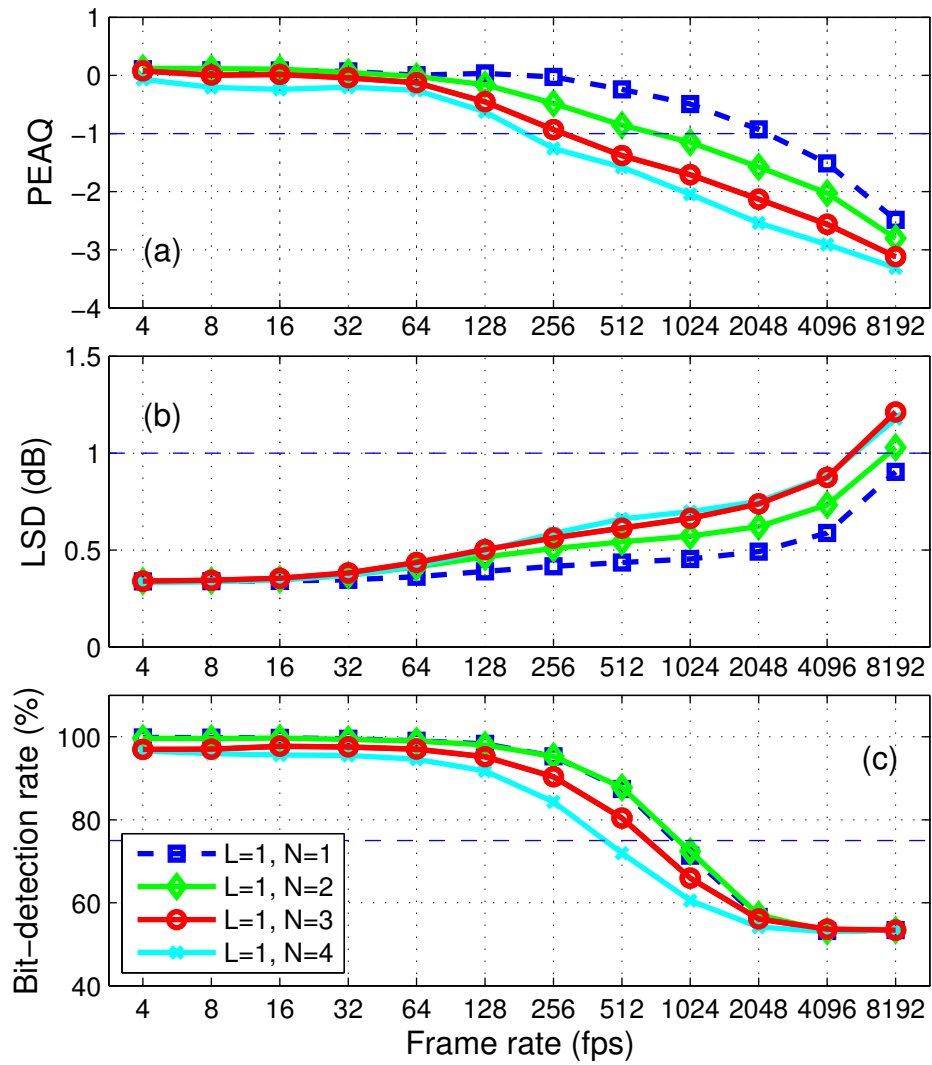


図 5.1: 音楽信号に対して並列型構成 ($N = 1, 2, 3, 4$) を用いた客観評価実験の結果: (a) PEAQ, (b) LSD, (c) ビット検出率

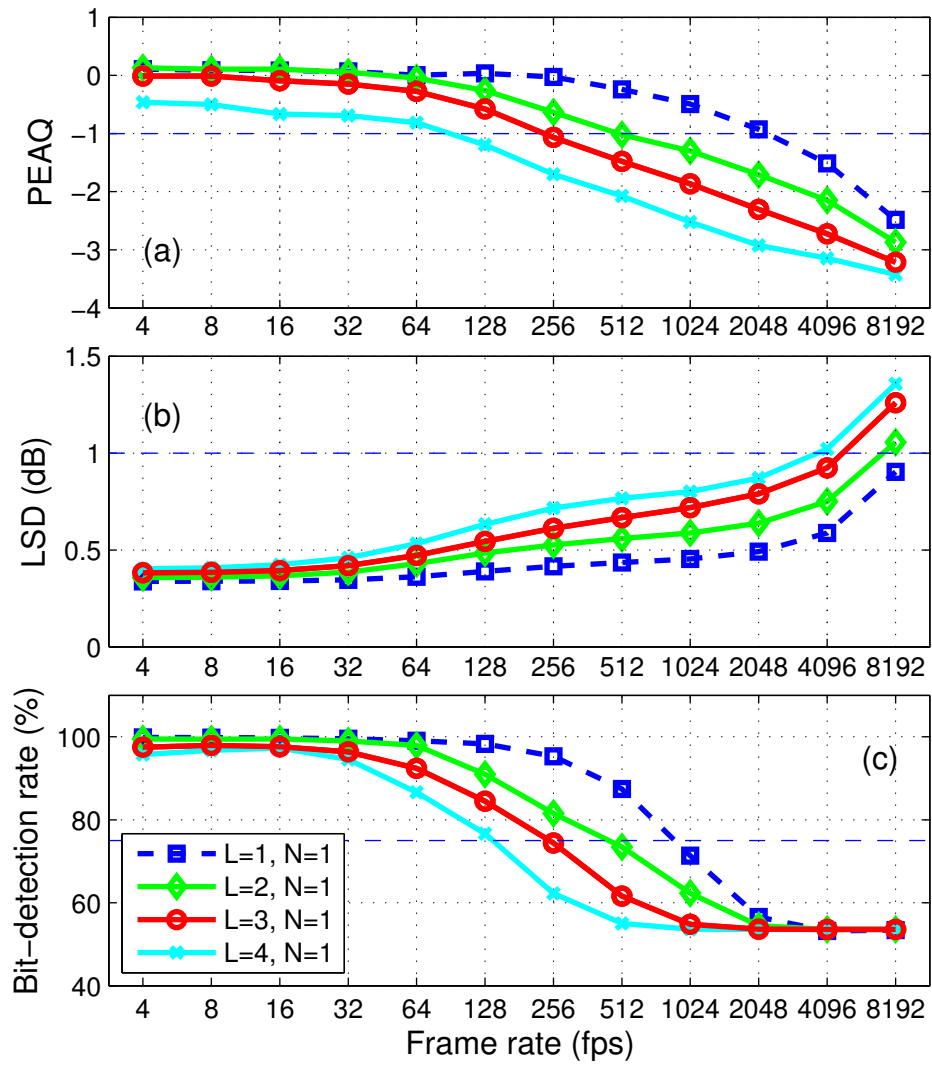


図 5.2: 音楽信号に対して縦続型構成 ($L = 1, 2, 3, 4$) を用いた客観評価実験の結果: (a) PEAQ, (b) LSD, (c) ビット検出率

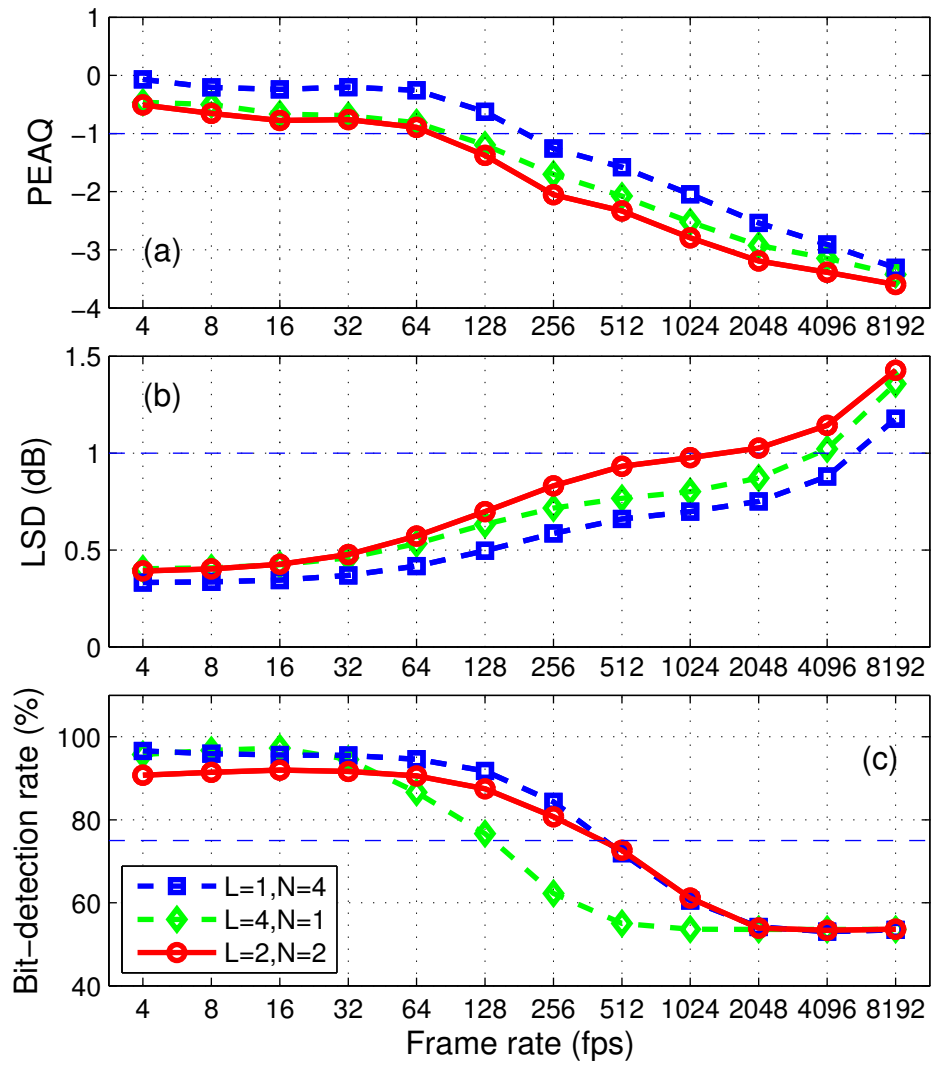


図 5.3: 音楽信号に対して複合型構成 $(L, N) = (1, 4), (4, 1), (2, 2)$ を用いた客観評価実験の結果: (a) PEAQ, (b) LSD, (c) ビット検出率

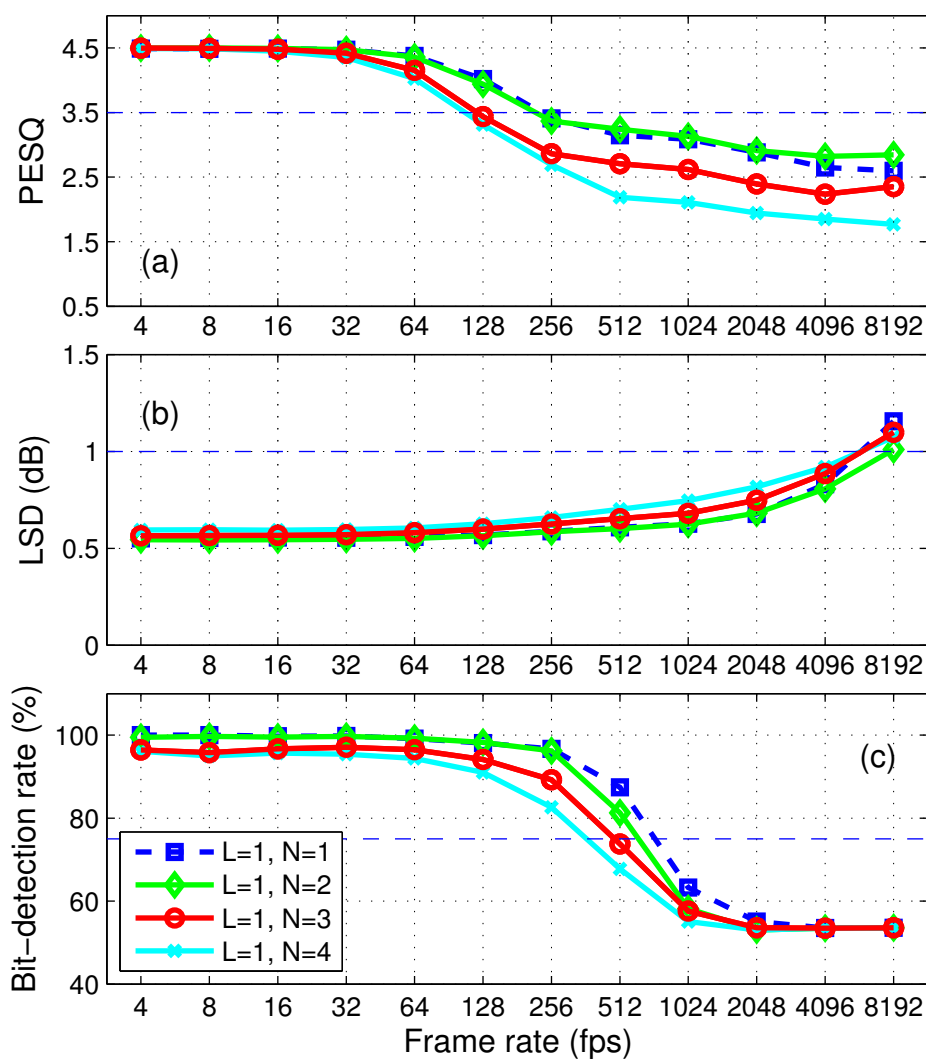


図 5.4: 音声信号に対して並列型構成 ($N = 1, 2, 3, 4$) を用いた客観評価実験の結果: (a) PESQ, (b) LSD, (c) ビット検出率

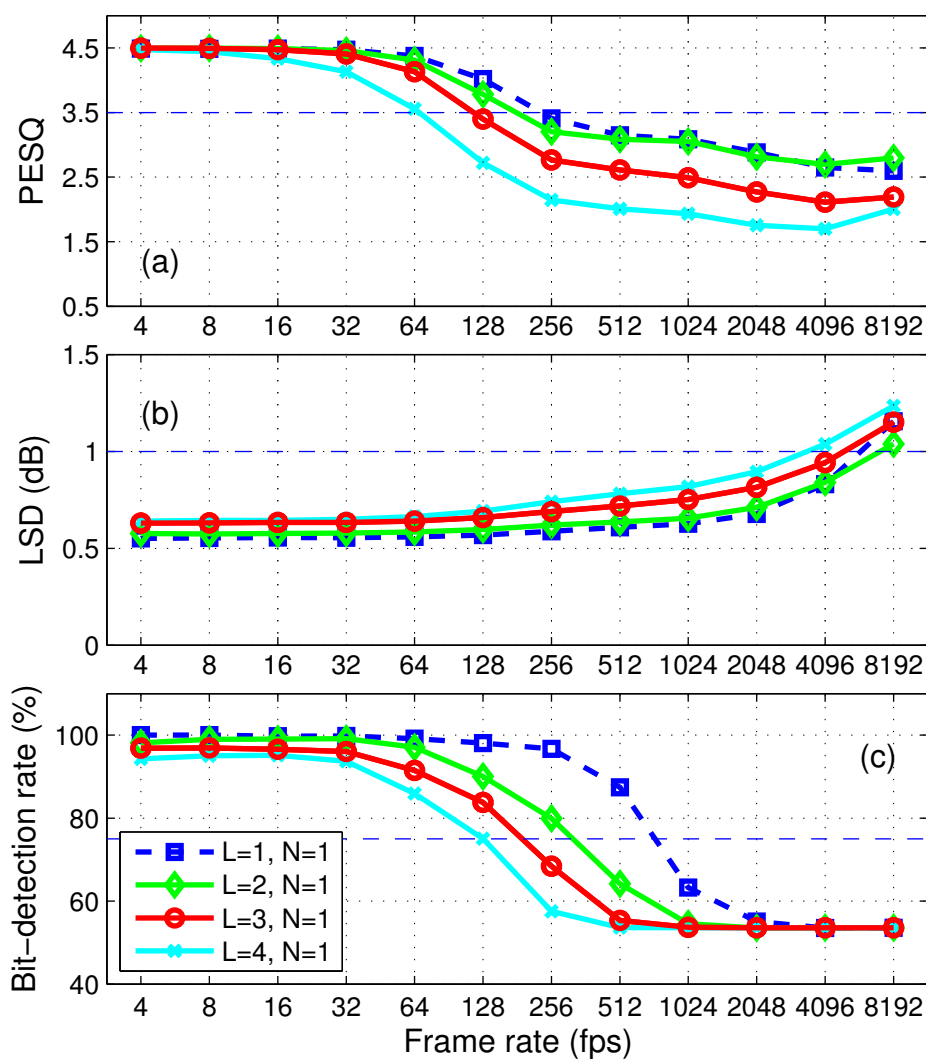


図 5.5: 音声信号に対して縦続型構成 ($L = 1, 2, 3, 4$) を用いた客観評価実験の結果: (a) PESQ, (b) LSD, (c) ビット検出率

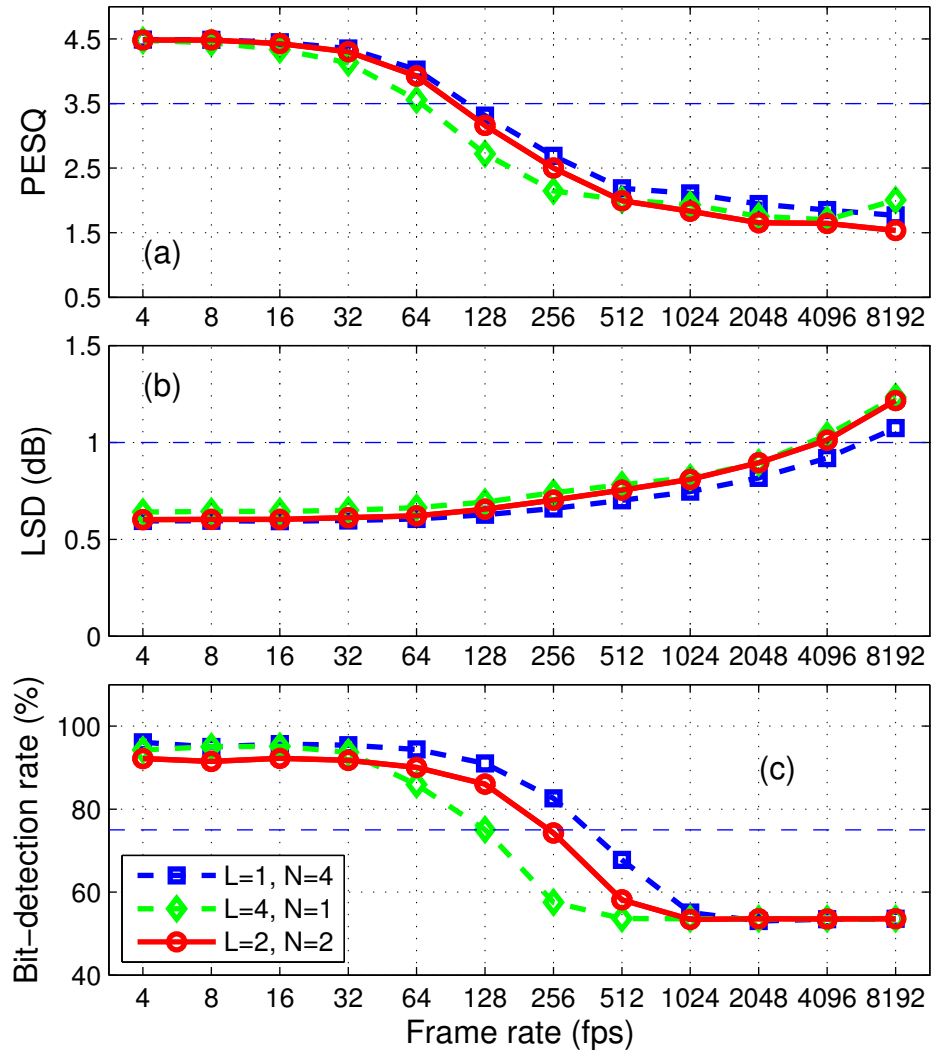


図 5.6: 音声信号に対して複合型構成 ($(L, N) = (1, 4), (4, 1), (2, 2)$) を用いた客観評価実験の結果: (a) PESQ, (b) LSD, (c) ビット検出率

表 5.1: 基本型 / 並列型 / 縦続型 / 複合型構成をもつ蝸牛遅延フィルタを利用したデータ埋め込み / データ検出における耐性評価試験の結果．表中の単位は fps である．

	Base	Parallel Architecture		
Modification	$L = 1, N = 1$	$L = 1, N = 2$	$L = 1, N = 3$	$L = 1, N = 4$
Non-process	512	512	512	256
DS 20 kHz	256	256	256	128
DS 16 kHz	256	256	256	128
DS 8 kHz	128	128	128	64
BC 24 bits	256	256	256	128
BC 8 bits	256	256	256	128
mp3 128 kbps	128	128	128	64
mp3 96 kbps	64	64	64	32
mp3 64 kbps	128	128	64	64

	Cascade Architecture			Composite Arc.
Modification	$L = 2, N = 1$	$L = 3, N = 1$	$L = 4, N = 1$	$L = 2, N = 2$
Non-process	256	256	128	256
DS 20 kHz	128	128	64	128
DS 16 kHz	128	128	64	128
DS 8 kHz	128	64	64	64
BC 24 bits	128	128	128	128
BC 8 bits	128	128	64	64
mp3 128 kbps	128	64	32	64
mp3 96 kbps	64	32	—	64
mp3 64 kbps	64	64	32	64

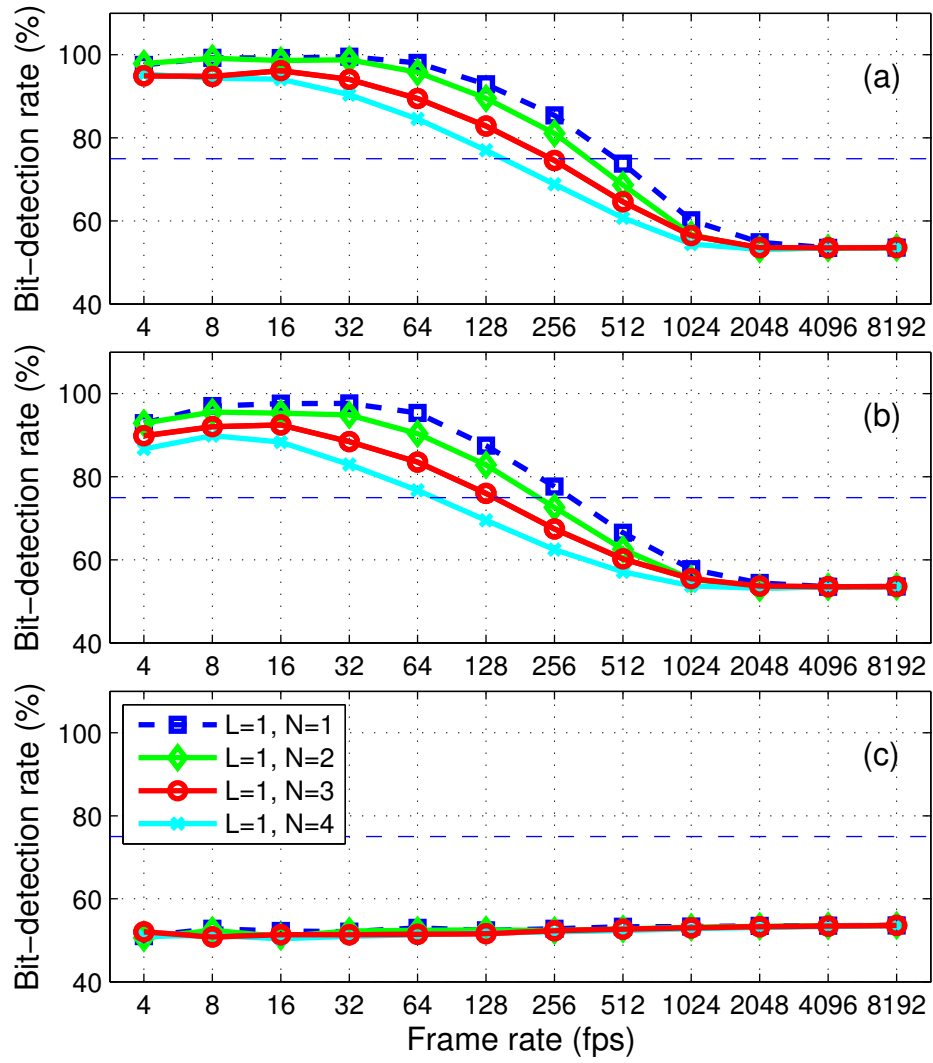


図 5.7: 並列型構成 ($N = 1, 2, 3, 4$) における音声符号化耐性試験の結果: (a) PCM, (b) ADPCM, (c) CS-ACELP

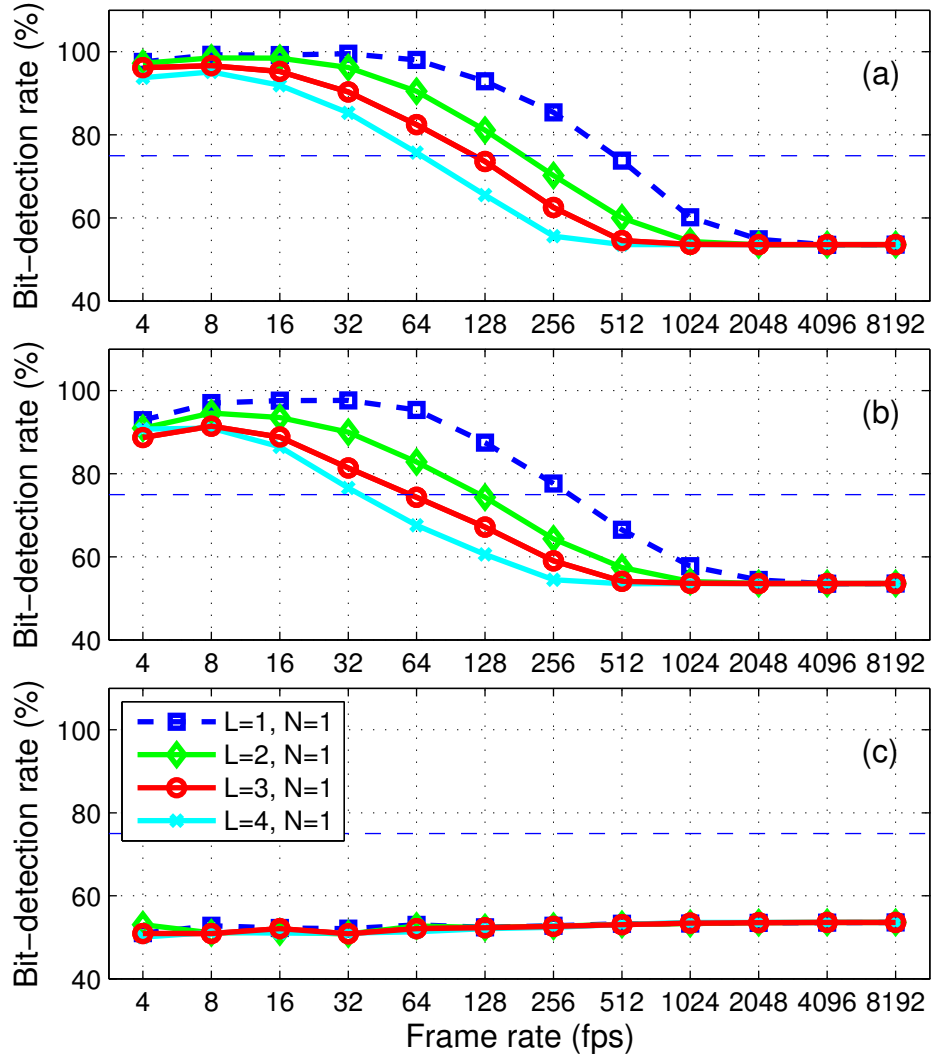


図 5.8: 縦続型構成 ($L = 1, 2, 3, 4$) における音声符号化耐性試験の結果: (a) PCM, (b) ADPCM, (c) CS-ACELP

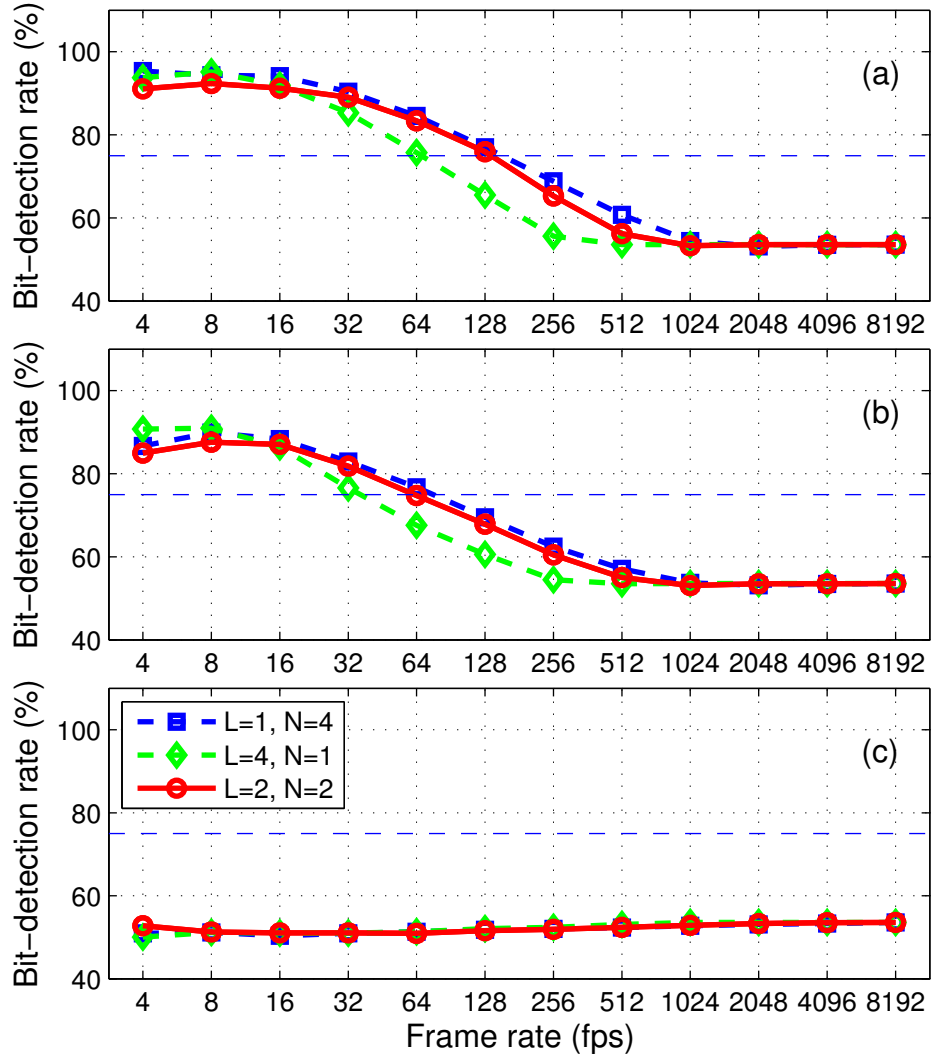


図 5.9: 複合型構成 ($(L, N) = (1, 4), (4, 1), (2, 2)$) における音声符号化耐性試験の結果 :
(a) PCM , (b) ADPCM , (c) CS-ACELP

表 5.2: StirMark ベンチマークテストの信号処理に関するカテゴリー

Category	SMBA Attack
i) Noise	AddBrumm, AddDynBoise, AddFFTNoise, AddNoise, AddSinus, NoiseMax
ii) Amplitude	Amplify, Compressor, Normalizer1, Normalizer2
iii) Bit	BitChanger, LSBZero
iv) Data	CopySample, CutSample, Exchange, FlipSample, ReplaceSamples, ZeroCross, ZeroLength1, ZeroLength2, ZeroRemove
v) Filtering	BassBoost, ExtraStereo, FFT HLPassQuick, RC LowPass, RC HighPass, Smooth1, Smooth2, State1, State2, VoiceRemove
vi) Phase	FFT Invert, FFT RealReverse, Invert
vii) Echo	Echo

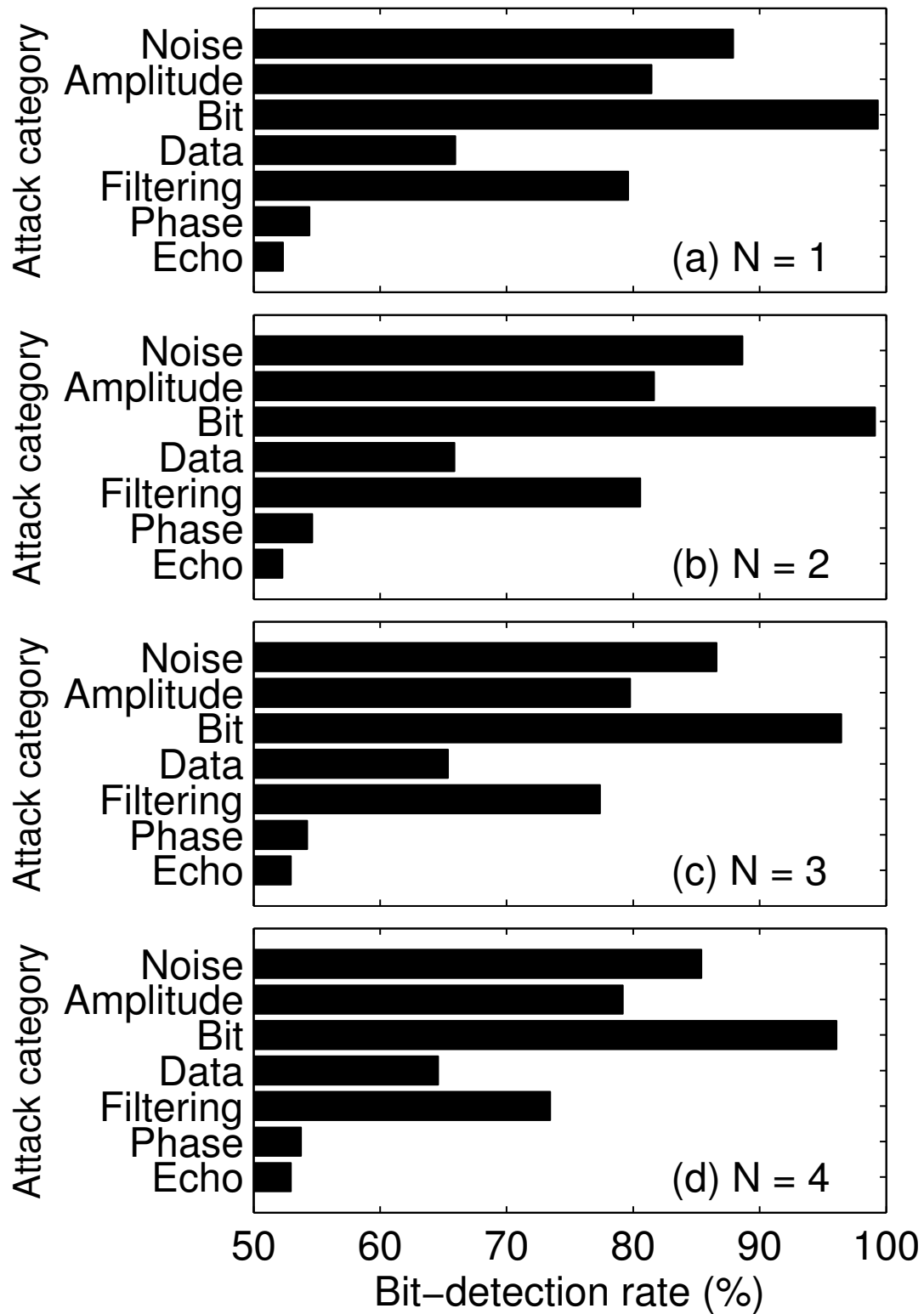


図 5.10: 並列型構成における StirMark ベンチマークテストの結果 : (a) $N = 1$, (b) $N = 2$, (c) $N = 3$, (d) $N = 4$

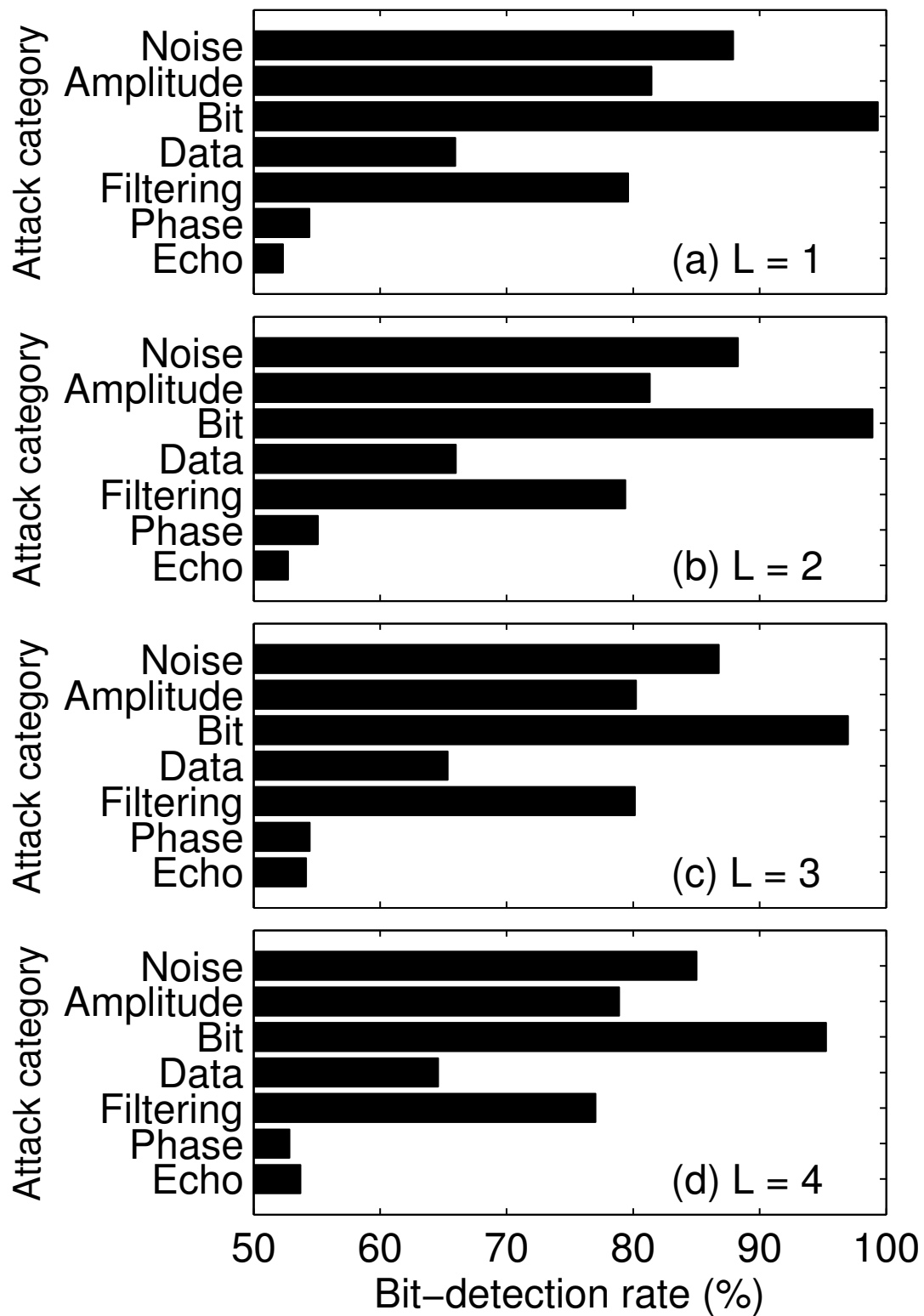


図 5.11: 縦続型構成における StirMark ベンチマークテストの結果 : (a) $L = 1$, (b) $L = 2$, (c) $L = 3$, (d) $L = 4$

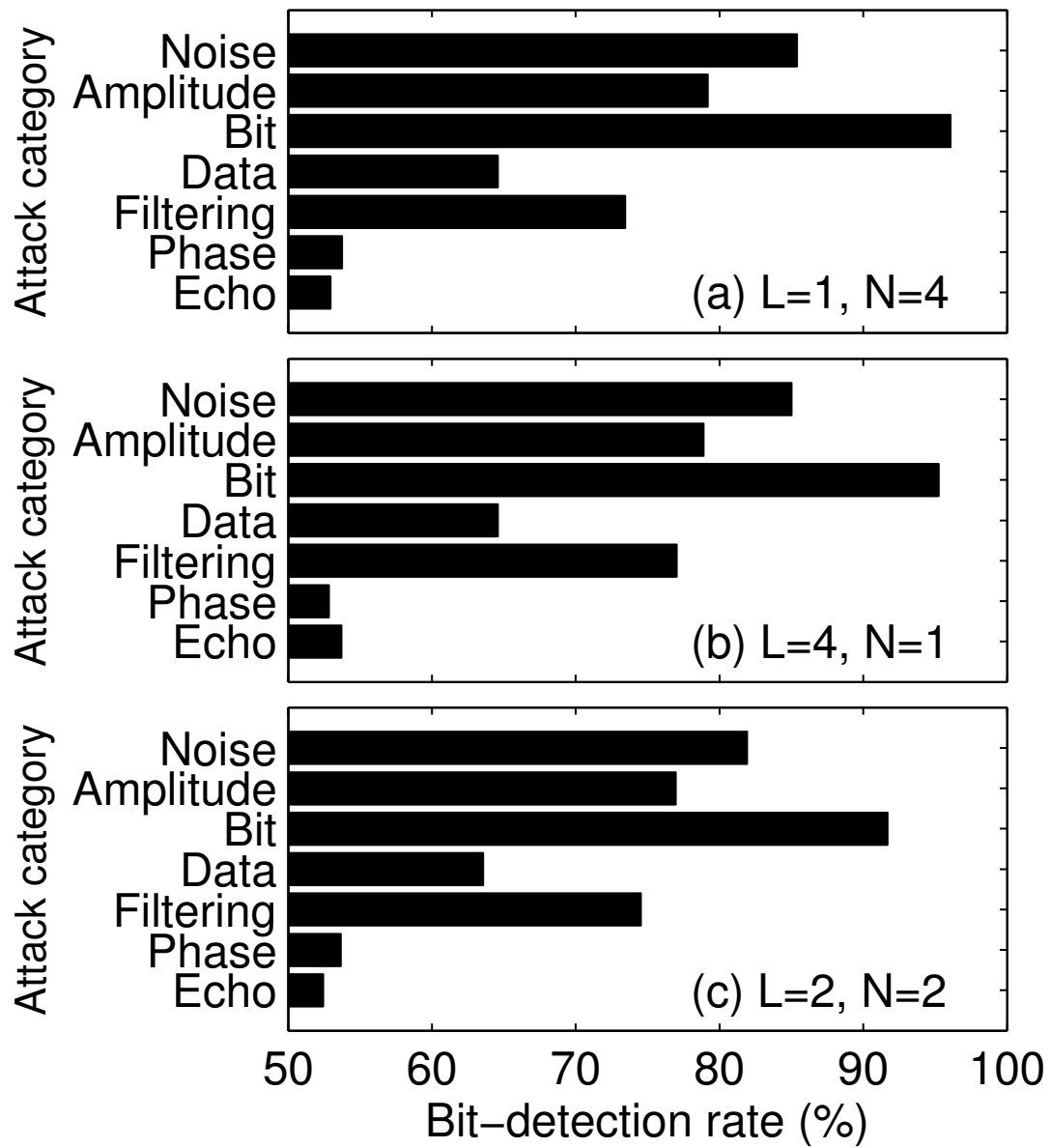


図 5.12: 複合型構成における StirMark ベンチマークテストの結果 : (a) $L = 1, N = 4$, (b) $L = 4, N = 1$, (c) $L = 2, N = 2$

第6章 提案法の応用

6.1 音声ステガノグラフィ

音声ステガノグラフィとは、音声信号をカバーデータとして、秘匿情報（テキスト・画像・音声信号など）を検知されないようにカバーデータに埋め込み、想定するユーザーのみに秘匿情報を送り届ける技術である。例えば、図 6.1 に示す音声ステガノグラフィを考えてみる。秘匿情報が埋め込まれたカバーデータはステゴデータと呼ばれる。ステゴデータは通信路を介して想定する受信者に届き、受け取った受信者は、誰にも気づかれずにステゴデータから秘匿情報を検出できる。一方で、ステガノグラフィと同様に秘匿情報を誰にも気づかれずに想定している相手に送る技術としてクリプトグラフィーがある。クリプトグラフィーは秘匿情報の内容を読めなくする技術であるのに対して、ステガノグラフィは秘匿情報の存在自体を隠してしまう技術であり、この点が両者の間で大きく異なる。ステガノグラフィは第三者に秘匿情報が埋め込まれていることがわからないため、埋め込まれた情報の秘匿性が高まる（つまり第三者に秘匿情報が漏れることはない）という利点がある。

ディジタル音声通信で利用する音声ステガノグラフィを想定した場合、次の三つの要求項目（検知不可能性、頑健性、埋め込み情報量）を満たす必要がある。検知不可能性は、秘匿情報の埋め込みにより音声に知覚可能な歪みが一切生じないことを要求している。頑健性は、第三者による攻撃やディジタル音声通信路による影響（音声符号化）により秘匿情報の検出が影響を受けないことを要求している。埋め込み情報量は、カバーデータに依存せず秘匿情報をどれだけ埋め込むことが可能かという評価指標であり、想定しているステガノグラフィで十分であることを要求している。ディジタル音声通信では、カバーデータをそのまま伝送するのではなく、PCM コーデックなどの音声符号化を用いて伝送するのが一般的である。したがって、ディジタル音声通信で音声ステガノグラフィを使用する場合には、PCM コーデックなどの音声符号化処理に対する耐性が非常に重要となる。

そこで、ディジタル音声通信における音声ステガノグラフィを想定し、蝸牛遅延特性に基づく情報ハイディング法を用いて実際に図 6.2 の 32×32 bitmap モノクロ画像を秘匿情報としてカバーデータに埋め込む。また、頑健性に関する客観評価として、耐性評価試験を行う。

6.1.1 評価試験方法

評価試験では，秘匿情報として図 6.2 の 32×32 bitmap モノクロ画像をカバーデータに埋め込む．カバーデータには，これまでの検討でも用いた ATR 音声データベース [34] B セットの内 10 秒程度の音声一つを用いる． 32×32 bitmap 画像の情報量が 1024 bit あり，10 秒程度のカバーデータに埋め込むため，ビットレートは $N_{\text{bit}} = 128$ bps とする ($10 \times 128 = 1280$ bit 埋め込み可能となる)．次に，デジタル音声通信での利用を想定した耐性評価試験を行う． 32×32 bitmap 画像を秘匿情報としてカバーデータに対し，蝸牛遅延特性に基づく情報ハイディング法を用いて埋め込んだステゴデータに対して音声通信で最も一般的に用いられている音声符号化方式 PCM コーデック (G.711) [36] で符号化・複合化を行い，蝸牛遅延特性に基づく情報ハイディング法を用いて秘匿情報 (32×32 bitmap 画像) の検出を行う．

6.1.2 評価試験結果

図 6.3 に，提案法を用いて図 6.2 の秘匿情報 (32×32 bitmap 画像) の埋め込みを行い，検出した秘匿情報 (32×32 bitmap 画像) を示す．このときビット検出率は 97.5% であり，26 bit の誤りが生じている．図 6.3 の提案法における検出画像を見ると，どのような画像かということは認識でき，ある程度の情報を保っているといえる．この結果から，2.5% のビット誤りがあるため，画像の劣化が見られるものの，どのような画像であるかを認識できる画像情報を検出することができるとわかった．

図 6.4 に 32×32 bitmap 画像を埋め込んだカバーデータに対して PCM コーデックで符号化・複合化を行い，提案法を用いて検出した 32×32 bitmap 画像を示す．このときビット検出率は 90.9% であり，93 bit の誤りが生じている．図 6.4 の提案法における画像を見ても，図 6.3 と比べて画像の劣化は見られるが，どのような画像かを認識できる程度の情報は保ったまま秘匿情報が検出できている．以上の結果から，提案法は PCM コーデックに対して頑健性を保証できることがわかる．検知不可能性だけではなく頑健性も非常に重要であり，この二つを満たす提案法は音声ステガノグラフィとして利用できる．

ここでは，基本型を用いて検討を行った．並列型，縦続型，複合型を用いれば，これまでの検討から埋め込み情報量を現状の約 3 倍に向上することができるとわかる．今回と同様に，10 秒程度のカバーデータを用いて，ビットレートを $N_{\text{bit}} = 128$ bps とした場合には， $10 \times 128 \text{ bps} \times 3 = 3840 \text{ bit}$ が埋め込み可能となる．

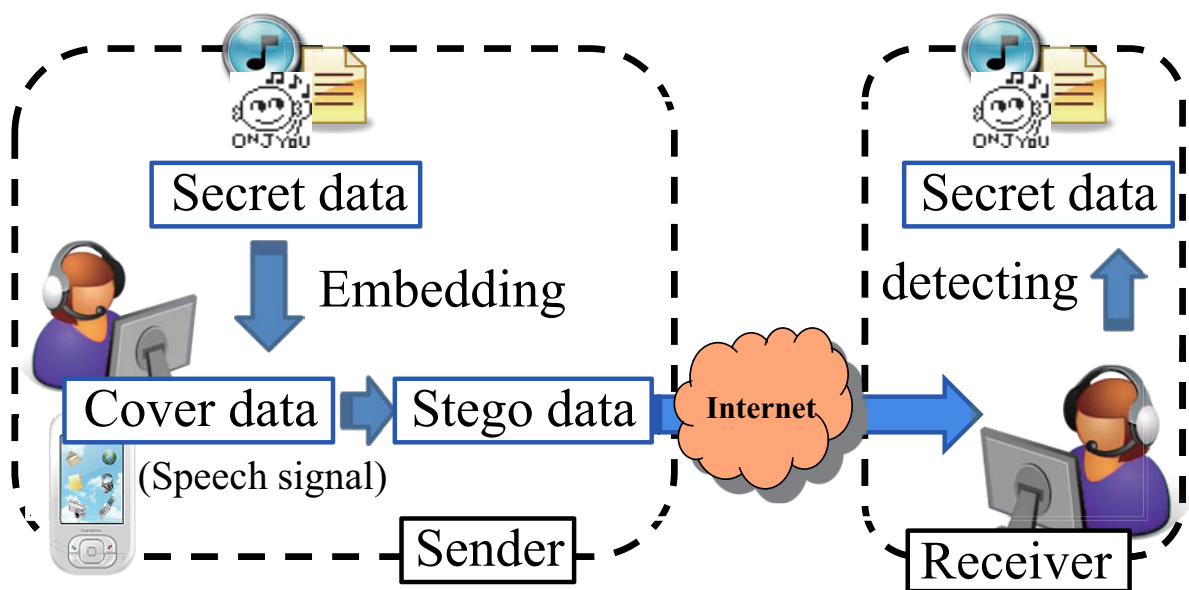


図 6.1: 音声ステガノグラフィのイメージ

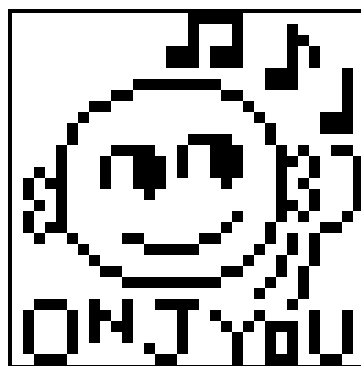


図 6.2: 埋め込みに用いた 32×32 の bitmap 画像

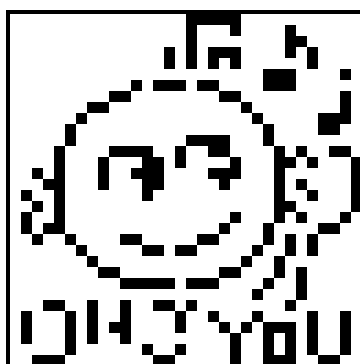


図 6.3: 検出した 32×32 の bitmap 画像

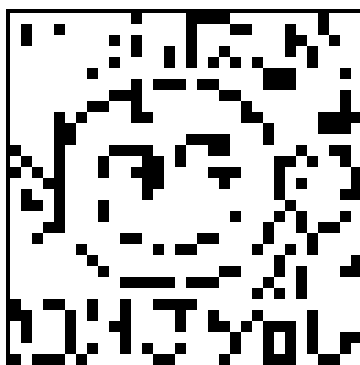


図 6.4: PCM 符号化を行った音声における検出画像

第7章 結論

7.1 本論文であきらかになったこと

本研究では、蝸牛遅延特性に基づく情報ハイディング法における埋め込み情報量を高めるため、蝸牛遅延フィルタの構成方法について議論してきた。音楽信号における検知不可能性と頑健性に関する客観評価結果では、複合型構成 ($L = 2, N = 2$) が一番最適な組み合わせであることがわかった。ここで、このフィルタ構成における最大埋め込み情報量は、256 bps (64 fps の時 1 フレーム当り 4 bits 表現) であった。音声信号における検知不可能性と PCM コーデックに対する頑健性に関する客観評価結果から、並列型構成 ($L = 1, N = 2$) が一番最適な組み合わせであるとわかった。ここで、このフィルタ構成における最大埋め込み情報量は、512 bps (256 fps の時 1 フレーム当り 2 bits 表現) であった。また、音声信号における検知不可能性と ADPCM コーデックに対する頑健性に関する客観評価結果では、並列型構成 ($L = 1, N = 3$) が一番最適な組み合わせであるとわかった。ここで、このフィルタ構成における最大埋め込み情報量は、384 bps (128 fps の時 1 フレーム当り 3 bits 表現) であった。PESQ や PEAQ の値が比較的よい基本型、並列型 ($L = 1, N = 2$) や縦続 ($L = 2, N = 1$) は音声符号化や MP3 情報圧縮の影響を強く受けている事がわかり、これに対して、PESQ や PEAQ が元々あまりよくない段数の多いフィルタ構成は音声符号化や MP3 情報圧縮の影響をあまり受けていない事がわかった。これらのことから、検知不可能性と頑健性にはトレードオフがあることがわかった。

StirMark ベンチマークテストの結果から、提案法は位相領域を扱っているため、位相に影響を及ぼすような信号処理に対しては脆弱であるが、雑音を加える処理や振幅処理やビット処理やフィルタリング処理のような信号処理に対しては非常に頑健であることがわかった。

提案法の応用として、音声ステガノグラフィとして利用できるかどうかを検討した。検討結果から、秘匿情報の再現性が求められるような利用状況では使用することが難しいが、情報が伝わる事が目的である利用状況では十分に音声ステガノグラフィとして利用できることがわかった。

以上の結果より提案法は、検知不可能性と頑健性を満たした上で埋め込み情報量を高めることができ、マルチメディア情報ハイディング技術として、電子音響透かしや音声ステガノグラフィに有効な手法であることがわかった。

7.2 残された課題

本論文では，提案法の応用として音声ステガノグラフィとステレオ音声への情報ハイディングを提案した．しかし，音声ステガノグラフィでは，基本型を用いた一例しか示しておらず，並列型や縦続型や複合型を用いた場合における画像情報だけではなく，多くの検討を行う必要がある．音響信号や音声信号によって埋め込み限界が異なることは明らかにしたが，音響信号と音声信号おける違いが埋め込み限界に対してどのように影響したのか，本質的な理由が明らかになっていない．また，提案法はノンブラインド検出であり，実用化を考えた場合にはブラインド検出が望まれる．そのため，ブラインド検出法を確立する必要がある．

本論文で提案したマルチメディア情報ハイディング技術が，マルチメディアコンテンツを利用するユーザーや制作者の双方に利益をもたらし，人間の生活をより豊かにすることを願って本論文の結びとする．

謝辞

北陸先端科学技術大学院大学情報科学研究科長赤木正人教授には，専門知識に疎かった著者を熱心に指導して頂き，本研究を進めるにあたり数多くの助言を賜りました．ここに深く感謝申し上げます．

北陸先端科学技術大学院大学情報科学研究科鷗木祐史准教授には，サバティカルにも係わらず，本研究を進めるにあたり議論の機会を数多く設けて頂き，熱心に指導して頂きました．また，多くの原稿執筆に関して助言や添削をして頂き，深く感謝申し上げます．北陸先端科学技術大学院大学情報科学研究科党建武教授，徳田功准教授には，本論文の審査員としてジョイントミーティングなどにおいて，御意見や御助言を頂き，深く感謝申し上げます．

北陸先端科学技術大学院大学情報科学研究科宮内良太助教には，本研究を進めるにあたり数多くの助言を賜り，原稿執筆に関して添削をして頂きました．北陸先端科学技術大学院大学情報科学研究科羽二生篤研究員には，本研究を進めるにあたりプログラミングやネットワークの知識に疎かった著者を温かく御支援して頂きました．また，研究や私生活に関して親身になって相談に乗って頂きました．北陸先端科学技術大学院大学情報科学研究科宮内千夏研究員には，本研究を進めるにあたりプログラミングに関して御支援して頂きました．研究室事務補佐員樋上忍氏には，著者が研究活動を行うにあたり事務手続きをして頂きました．ここに，深く感謝申し上げます．

研究室のみなさまには，研究室会議において数多くの御意見を頂きました．また，私生活において何度も元気づけられ，多くの励ましを頂きました．研究室のみなさまに感謝申し上げます．

最後に，大学院進学を承諾して頂き，これまでの学生生活を経済的にも精神的にも支えて下さった両親に心より感謝申し上げます．

参考文献

- [1] Double Click Ad Planner, <http://www.google.com/adplanner/>
- [2] Apple プレスリリース, <http://www.apple.com/jp/news/2010/feb/26itunes.html>
- [3] Skype プレスリリース, <http://about.skype.com/>
- [4] 西村 明, “音響信号への情報秘匿技術 -電子音響透かしとステガノグラフィ-,” 音響誌, **63**(11), pp. 660–667, 2007.
- [5] 竹居智久, Phil Keys, “「コピーに自由を」生まれ変わる DRM,” 日経エレクトロニクス, pp. 51–74, 2008.
- [6] 小特集 マルチメディア情報ハイディング技術とその応用, 信学論 A, **J93–A**(2), pp. 41, 2010.
- [7] Cvejic, N., and Seppänen, T., “*Digital audio watermarking techniques and technologies*,” IGI Global, Hershey, PA 2007.
- [8] Boney, L., Tewfik, H. H., and Hamdy, K. N., “Digital watermarks for audio signals,” *Proc. ICMCS*, pp. 473–480, 1996.
- [9] Gruhl, D., Lu, A., and Bender, W., “Echo Hiding,” *Proc. Information Hiding 1st Workshop*, pp. 295–315, 1996.
- [10] Nishimura, A., “Audio watermarking based on subband amplitude modulation,” *J. Acoust. Soc. Jpn.*, Vol. 31, No. 5, pp.328–336, 2010.
- [11] 西村 竜一, 鈴木 陽一, “周期的位相変調に基づく音響電子透かし,” 音響誌, **60**(5), pp.268–272, 2004.
- [12] 中山 彰, 陸 金林, 中村 哲, 鹿野 清宏, “心理音響モデルに基づいたオーディオ信号の電子透かし,” 電子情報通信学会論文誌, Vol. J83–D–II, No. 11, pp.2255–2263, 2000 .
- [13] 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への電子透かし,” 電子情報通信学会論文誌, Vol. J87–A, No. 6, pp.787–796, 2004.
- [14] 松井 甲子雄, 電子透かしの基礎, pp.146–194, 森北出版, 1998 .

- [15] 小野 束, 電子透かしとコンテンツ保護, pp.139–168, オーム社, 2001 .
- [16] 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め込み,” 情報処理学会論文誌, Vol. 38, No. 10, pp.2053–2061, 1997.
- [17] Aoki, N., “A band extension technique for G.711 speech using steganography,” IEICE Transactions on Communications, Vol. E89–B, No. 6, pp.1896–1898, 2006.
- [18] 伊藤 彰則, 半田 浩規, 鈴木 陽一, “ADPCM 出力とサンプルの絶対値を考慮した G.711 への固定ビットレート情報ハイディング,” 電子情報通信学会論文誌, Vol. J93–A, No. 2, pp.82–90, 2010.
- [19] 西村 明, “振幅変調を用いた音声信号へのデータハイディング,” 平成 19 年度日本音響学会秋季研究発表会講演論文集, pp.619–620, 2007 .
- [20] Unoki, M., and Hamada, D., “Method of digital–audio watermarking based on cochlear delay characteristics,” *International Journal of Innovative Computing Information and Control*, 6(3), pp.1325–1346, 2008.
- [21] Imabeppu, K., Hamada, D., and Unoki, M., “Embedding limitations with audio–watermarking method based on cochlear delay characteristics,” *Proc. IIHMSP2009*, pp.82–85, 2009.
- [22] Unoki, M., Imabeppu, K., Hamada, D., Haniu, A., and Miyauchi, R., “Embedding Limitations with Digital–audio Watermarking Method Based on Cochlear Delay Characteristics,” *Journal of Information Hiding and Multimedia Signal Processing*, 2(1), 2011.
- [23] Unoki, M., Kosugi, T., Haniu, A., and Miyauchi, R., “Design of IIR all-pass filter based on cochlear delay to reduce embedding limitations,” *Proc. IIHMSP2010*, 2010.
- [24] 鵜木 祐史, 小杉 敏三, 羽二生 篤, 宮内 良太, “埋め込み限界を高めるための蝸牛遅延フィルタの構成,” 第 10 回マルチメディア情報ハイディング研究会資料, pp.25–30, 2010.
- [25] 鵜木 祐史, 羽二生 篤, 小杉 敏三, 宮内 良太, “音声信号への蝸牛遅延特性を利用した情報ハイディングの検討,” 信学技報, EA2010–27, pp.31–36, 2010.
- [26] 小杉 敏三, 羽二生 篤, 宮内 良太, 鵜木 祐史, 赤木 正人, “蝸牛遅延フィルタの並列・縦続・複合型構成に関する検討,” 第 11 回マルチメディア情報ハイディング研究会, pp.19–24, 2010.
- [27] 鈴木 真奈美, 西村 竜一, 鈴木 陽一, “高周波数帯域における位相回転の検知に関する基礎的検討,” 平成 13 年度日本音響学会春季研究発表会講演論文集, pp.475–476, 2001 .

- [28] 赤木 正人, “位相と知覚 -人間は果たして位相聾か?-,” 平成 9 年度日本音響学会秋季研究発表会講演論文集, pp.193–196, 1997.
- [29] 赤木 正人, 安武 浩二郎, “時間方向情報の知覚の検討:位相変化の音色知覚に及ぼす影響について,” 信学技報, EA1998–98, pp.13–18, 1998.
- [30] Dau, T., Wegner, O., Mellert, V., and Kollmeier, B., “Auditory brainstem responses with optimized chirp signals compensating basilar membrane dispersion,” *J. Acoust. Soc. Am.*, Vol. 107, No. 3, pp.1530–1540, 2000.
- [31] Aiba, E., Tsuzaki, M., Tanaka, S., and Unoki, M., “Judgment of perceptual synchrony between two pulses and its relation to the cochlear delays,” *Jpn. Psychological Research*, 50(4), pp.204–213, 2008.
- [32] Goto, M., Hashiguchi, H., Nishimura, T., Oka, R., “RWC Music Database: Music Genre Database and Musical Instrument Sound Database,” *Proc. ISMIR*, pp.229–230, 2003.
- [33] Kabal, P., “An examination and interpretation of ITU-R BS.1387: Perceptual evaluation of audio quality,” *TSP Lab. Tech. Rep., Dept. Elec. & Comp. Eng.*, McGill Univ. 2002.
- [34] 株式会社 ATR-Promotions, ATR 音声データベース, <http://www.atr-p.com/sdb.html>
- [35] Yi, Hu., and Philipos, C. Loizou., “Evaluation of Objective Measures for Speech Enhancement,” *Interspeech2006*, pp.1447–1450, 2006.
- [36] ITU-T, “G.711: Pulse code modulation (PCM) of voice frequencies,” 1988.
- [37] ITU-T, “G.726: 40, 32, 24, 16 kbit/s adaptive differential pulse code modulation (ADPCM),” 1990.
- [38] ITU-T, “G.729: Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP),” 1996.
- [39] Steinebach, M., Petitcolas, F. A. P., Raynal, F., Dittmann, J., Fontaine, C., Seibel, C., Fatés N., and Ferri, L. C., “StirMark Benchmark: Audio watermarking attacks,” *Proc. Coding and Computing 2001*, pp.49–54, 2001.
- [40] Klumpp, G. R., and Eady, R. H., “Some measurements of interaural time difference threshold,” *J. Acoust. Soc. Am.*, Vol. 28, No. 5, pp.859–860, 1959.
- [41] Mills, W. A., “On the Minimum Audible Angle” *J. Acoust. Soc. Am.*, Vol. 30, No. 4, pp.237–246, 1958.

- [42] Perrott, R, D., and Musicant, D, A., “Minimum auditory movement angle: Binaural localization of moving sound sources,” *J. Acoust. Soc. Am.*, Vol. 62, No. 6, pp.1463–1466, 1977.

学会発表リスト

1. 小杉 敏三, 羽二生 篤, 宮内 良太, 鷗木 祐史, 赤木 正人, “蝸牛遅延特性に基づく音声ステガノグラフィーの検討,” 平成 23 年日本音響学会春季研究発表会, 1-Q-33, March. 2011.
2. Kosugi, T., Haniu, A., Miyauchi, R., Unoki, M., and Akagi, M., “Study on suitable-architecture of IIR all-pass filter for digital-audio watermarking technique based on cochlear delay characteristics,” RISP International Workshop on Nonlinear Circuits and Signal Processing, March. 2011.
3. 小杉 敏三, 羽二生 篤, 宮内 良太, 鷗木 祐史, 赤木 正人, “蝸牛遅延フィルタの並列・縦続・複合型構成に関する検討,” 第 11 回マルチメディア情報ハイディング研究会, pp.19–24, Oct. 2010.
4. 羽二生 篤, 小杉 敏三, 宮内 良太, 鷗木 祐史, “蝸牛遅延特性を利用した電子音響透かしのブラインド検出に向けた基礎検討,” 第 11 回マルチメディア情報ハイディング研究会, pp.13–18, Oct. 2010.
5. Unoki, M., Kosugi, T., Haniu, A., and Miyauchi, R., “Design of IIR all-pass filter based on cochlear delay to reduce embedding limitations,” *Proc. IIHMSP2010*, Oct. 2010.
6. 羽二生 篤, 小杉 敏三, 宮内 良太, 鷗木 祐史, “蝸牛遅延特性を利用した電子透かしのブラインド検出に関する検討,” 平成 22 年日本音響学会秋季研究発表会, 2-P-34, Sept. 2010.
7. 宮内 良太, 羽二生 篤, 小杉 敏三, 鷗木 祐史, “蝸牛遅延特性に基づいた位相偏移変調が知覚不可能性に与える影響,” 平成 22 年日本音響学会秋季研究発表会, 1-R-13, Sept. 2010.
8. 鷗木 祐史, 羽二生 篤, 小杉 敏三, 宮内 良太, “情報埋め込み限界を高めるための蝸牛遅延フィルタの構成,” 平成 22 年日本音響学会秋季研究発表会, 2-P-33, Sept. 2010.
9. 小杉 敏三, 羽二生 篤, 宮内 良太, 鷗木 祐史, 赤木 正人, “音声信号への蝸牛遅延に基づく情報ハイディングの客観評価,” 平成 22 年電気関係学会北陸支部連合大会講演論文集, G-19, Sept. 2010.

10. 鵜木 祐史, 羽二生 篤, 小杉 敏三, 宮内 良太, “音声信号への蝸牛遅延特性を利用した情報ハイディングの検討,” 信学技報, EA2010-27, pp.31-36, June. 2010.
11. 鵜木 祐史, 小杉 敏三, 羽二生 篤, 宮内 良太, “埋め込み限界を高めるための蝸牛遅延フィルタの構成,” 第10回マルチメディア情報ハイディング研究会資料, pp.25-30, June. 2010.

付録

蝸牛遅延フィルタのパラメータ決定方法

式 (3.1) の蝸牛遅延特性を模擬した 1 次の IIR 全域通過フィルタにおける振幅特性 $|H(z)|$, 位相特性 $\phi(z)$, 及び群遅延 $\tau(\omega)$ はそれぞれ次式のように表わされる .

$$|H(z)| = 1 \quad (7.1)$$

$$\phi(z) = \frac{1}{j}(\ln(1 - bz) - \ln(z - b)) \quad (7.2)$$

$$\tau(\omega) = -\frac{b^2 - 1}{1 + b^2 - 2b\cos(\omega T)} \quad (7.3)$$

式 (7.3) の関係式からパラメータ b について解く .

$$(\tau(\omega) + 1)b^2 - (2\tau(\omega)\cos(\omega T))b + (\tau(\omega) - 1) = 0 \quad (7.4)$$

を得る . この 2 次方程式について解くと

$$b = \frac{\tau(\omega)\cos(\omega T) \pm \sqrt{-\sin^2(\omega T)\tau^2(\omega) + 1}}{\tau(\omega) + 1} \quad (7.5)$$

を得る . ここで , 一つの群遅延 $\tau(\omega)$ に対して二つのパラメータ値 b を得られるが , フィルタ設計の制約条件からパラメータ b は $0 < b < 1$ でなければならない . そのため , 得られたパラメータ b の中でこの制約を満たさないものは解ではない .

実際に音声信号におけるパラメータ b を導出する . これまでの音楽信号におけるパラメータ b の最適値は 0.795 であり , 周波数 $f_0 = 100$ Hz における群遅延 $\tau(2\pi f_0)$ は , 式 (7.3) から 0.198 ms と得られる . 音声信号におけるサンプリング周波数は $f_s = 16000$ Hz であり , 式 (7.5) を用いて , 周波数 $f_0 = 100$ Hz における群遅延 $\tau(2\pi f_0) = 0.198$ ms を代入する . これより , 音声信号におけるパラメータ b の最適値は , 0.521 と求まる . よって , 式 (7.5) を用いれば , 群遅延量を任意に制御することができる .