

Title	線形予測残差を利用したスペクトル拡散型音声情報ハイディング法
Author(s)	並河, 伶弥
Citation	
Issue Date	2019-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/15905
Rights	
Description	Supervisor: 鷗木 祐史, 先端科学技術研究科, 修士 (情報科学)

修 士 論 文

線形予測残差を利用したスペクトル拡散型
音声情報ハイディング法

1710151 並河 伶弥

主指導教員	鵜木祐史
審査委員主査	鵜木祐史
審査委員	赤木正人
	党建武
	吉高淳夫

北陸先端科学技術大学院大学
先端科学技術研究科 [情報科学]

平成 31 年 2 月

Abstract

The rapid development of advanced information communication technology (ICT) has positively impacted digital speech communication. However, misuse of ICT causes speech tampering and speech spoofing in digital speech communication. Therefore, audio information hiding (AIH) techniques have recently been focused on as techniques enabling secure and safety communication.

The AIH techniques aim at embedding guarantee codes (personal codes, and device identification codes) as watermark to protect digital audio content. Which are inaudible and inseparable by users, and at robustly detecting embedded codes from watermarked signals against any kind of processing and various attacks. The AIH techniques generally must satisfy inaudibility and robustness. There are inaudible and inseparable by users, and at robustly detecting embedded codes from watermarked signals against any processing and various attacks.

The AIH is targeted for sound signals such as music. When AIH is used for speech, it can not embed or detect watermark. This is because the speech has a silent section and peculiar encoding Therefore, speech information hiding is required.

This paper proposed a speech information hiding method that can satisfy inaudibility and robustness simultaneously. A linear prediction (LP) analysis and synthesis is used in the proposed method. On the basis of the principle of the direct spread spectrum (DSS) method, the proposed method spectrally spreads a message by using LP residue and embeds the spread spectrum of the message into the host signal.

The direct spread spectrum (DSS) method is a robust AIH method. A watermarked signal $y(n)$ is made by that spreads the message signal $m(n)$ by a pseudo-random noise (PN) signal $c(n)$ and adds it to the host signal $x(n)$. The PN signals have properties of $E\{c(n)\} = 0$ and $E\{c^2(n)\} = 1$, where $E\{\cdot\}$ is the expected value of “.”. Since $x(n)$, $y(n)$, and $c(n)$ are assumed to be ergodicity, $E\{\cdot\}$ can be regarded as the temporal average of “.”. Hence, $E\{y(n)c(n)\}$ used for judgment is made by multiplying the watermarked signal $y(n)$ with the same PN signal $c(n)$. This process can be used to extract message “0” or “1” from $y(n)$.

The PN signal has a white spectrum, so the watermarked signal has been spread over a wide band. Since this feature distorts the watermarked signal, the DSS method has a problem in terms of inaudibility due to the spread spectrum.

Linear predictive coding (LPC) is one of the most basic speech coding methods using linear prediction (LP). LPC generate an LP coefficient and an LP residue corresponding to the sound source of the speech signal. Here, the LP residue has properties of $E\{c(n)\} = 0$ and $E\{c^2(n)\} = 1$. By using these properties, we propos speech information hiding

method, which the message $m(n)$ is spectrally spread by the LP residue instead of the PN signal.

The LP residue was used to spread the message in the host signal. The LP residue has a similar spectral band to the host signal, so the message was spectrally spread into a similar spectrum to the host signal. Therefore, the sound distortion due to the spread spectrum will be decrease by using LP residue as a spread signal. In addition, the proposed method will have the high robustness, because proposed method was based on DSS method as robust AIH. Information hiding technique have a trade-off between inaudibility and robustness.

In the proposed method, parameter a can be used to set the embedding strength for watermarked signal. The parameter a is determined by the power level of the host signal L_{PHS} , the power level of the watermark signal L_{PWS} , and the strength setting level L_{SSL} . The L_{SSL} was determined by evaluations of sound quality distortion (SQD) and the bit error rate (BER). The BER and SQD tests were carried out to determine the setting value. In the BER tests, BER of 10% was set as the criterion. In the SQD tests, ; log-spectrum distortion (LSD) and perceptual evaluation of speech quality (PESQ) were used in the evaluation method. In these tests, LSD of 1 dB or less and PESQ of -1 ODG or higher were used as criteria for sound distortion. The objective difference grades (ODGs) were graded on a five-point scale as 0 (imperceptible), -1 (perceptible), -2 (slightly annoying), -3 (annoying), and -4 (very annoying). Evaluation results, The L_{SSL} was determined to be -10 dB when the payload was 4 bps. Then, the mean L_{all} was -20.6 dB. When the payload was 8 bps, there was no optimal L_{all} . When we relaxed the evaluation standard of SQD, the L_{SSL} was determined to be 0 dB, Then, the mean L_{all} was -14.0 dB.

In information detection, when the host signal and the watermarked signal were not synchronized, accurate information detection is difficult. Therefore, synchronizes the two signals was required. In frame synchronization, correlation was taken while sliding two signals, and the point where the correlation value became maximum was taken as the synchronous position. The information detection rate after frame synchronization was evaluated. In the detection evaluate, BER of 10% was set as the criterion. As a detection result after signal synchronization, the BER was 2%.

We investigated whether the proposed method could satisfy the requirement of inaudibility and robustness. The inaudibility was evaluated by using the sound quality distortion (LSD and PESQ) of the watermarked signal. Robustness was evaluated by using BERs against various attacks; down-sampling (44.1 to 16 kHz), addition of white Gaussian noise (WGN, signal-to-noise ratio (SNR) of 36 dB), re-quantization (RQZ, 16 bit quantization to 8 bit quantization), and speech coding (G.711, G.726, and G.729).

In these evaluations, 12 speech signals in the ATR database (B set) were used. Only the utterance sections in the speech signals were used as the stimulus. The payloads were

4, 8, 16, 32, and 64 bps. Messages were random bit streams. No error correction schemes were used.

The LSDs were under 1 dB in the proposed method, whereas the LSDs were at or above 3 dB in the DSS method. LSD by the proposed method was improved by 2 dB in comparison with that by DSS method. The PESQs exceeded -1 ODG in the proposed method, whereas the LSDs was always about -4 ODG in the DSS method. PESQ by the proposed method was improved by 3 dB in comparison with that by DSS method.

As results of the BERs of the watermarked signals against attacks, the proposed method was robust against several attacks. The proposed method has the same degree of robustness as the DSS method.

目次

第1章 序論	1
1.1 はじめに	1
1.2 情報ハイディング技術	3
1.3 研究の目的	4
1.4 本論文の構成	4
第2章 音への情報ハイディング	6
2.1 音響情報ハイディング	6
2.2 代表的な情報ハイディング法	8
2.2.1 最下位ビット置換法	8
2.2.2 エコーハイディング法	9
2.2.3 周期的位相変調法	12
2.2.4 蝸牛遅延を用いた情報ハイディング法	14
2.2.5 直接スペクトル拡散法	16
2.3 音声信号への展開	18
2.4 音声情報ハイディングにおける要求項目	18
第3章 提案法の原理	19
3.1 スペクトル拡散法の音声への応用	19
3.2 拡散信号に求められる特徴	21
3.3 線形予測分析における線形予測残差	21
3.4 線形予測残差を用いたスペクトル拡散型音声情報ハイディング法の提案	23
第4章 提案法の実装	25
4.1 情報の埋め込み	25
4.1.1 情報の埋め込みアルゴリズム	25
4.2 最適な埋め込みパラメータ設定	28
4.2.1 埋め込み強度補正	28
4.2.2 選定基準	29
4.2.3 選定結果	30

4.3	情報検出	33
4.3.1	情報の検出アルゴリズム	33
4.3.2	フレーム同期検出	35
第 5 章	提案法の評価	40
5.1	評価尺度	40
5.2	評価条件	41
5.3	評価結果	42
5.3.1	知覚不可能性評価の結果	42
5.3.2	頑健性評価の結果	45
5.4	考察	53
第 6 章	結論	54
6.1	本研究で明らかになったこと	54
6.2	残された課題	55
	参考文献	57
	研究業績一覧	62
	謝辞	63
	付録	64
付 録 A	拡散信号の異なるスペクトル拡散型情報ハイディング法	64
A.1	白色雑音を用いたスペクトル拡散法	64
A.2	検討法の評価	65
A.3	検討法のまとめ	65

目 次

1.1	話者のなりすましや改ざん.	2
1.2	論文の構成.	5
2.1	情報ハイディングの概要.	7
2.2	エコーハイディング法におけるエコーの設定.	10
2.3	エコーハイディング法における情報検出例.	11
2.4	周期的位相変調法における群遅延特性の概要.	13
2.5	蝸牛の構造と基底膜振動で見られる蝸牛遅延.	15
2.6	直接スペクトル拡散法のブロックダイアグラム.	17
3.1	直接スペクトル拡散法による情報ハイディングの処理例.	20
3.2	線形予測分析合成の処理例.	22
4.1	提案法における情報埋め込みのブロックダイアグラム.	26
4.2	提案法を用いた音声情報ハイディングの処理例.	27
4.3	埋め込み強度を変更した場合の検出誤差率と音質評価の結果 (4 bps).	31
4.4	埋め込み強度を変更した場合の検出誤差率と音質評価の結果 (8 bps).	32
4.5	提案法における情報検出のブロックダイアグラム.	34
4.6	提案法における信号同期法のブロックダイアグラム.	37
4.7	提案法におけるフレーム同期法のブロックダイアグラム.	38
4.8	提案法におけるフレーム同期法処理例.	39
5.1	LSD による音質評価の結果.	43
5.2	PESQ による音質評価の結果.	44
5.3	信号処理や符号化を施していないときの BER.	46
5.4	ダウンサンプリング (44.1 kHz - 16 kHz) に対する耐性評価の結果.	47
5.5	雑音付与 (SNR 36 dB) に対する耐性評価の結果.	48
5.6	振幅圧縮 (16 bit - 8 bit) に対する耐性評価の結果.	49
5.7	G.711 符号化に対する耐性評価の結果.	50
5.8	G.726 符号化に対する耐性評価の結果.	51
5.9	G.729 符号化に対する耐性評価の結果.	52

A.1	白色雑音を用いたスペクトル拡散法のブロックダイアグラム.	66
A.2	白色雑音を用いたスペクトル拡散型情報ハイディングの処理例.	67

表 目 次

2.1	LSB 置換による処理例.	8
4.1	PESQ により出力される ODG 値.	29
5.1	耐性評価に利用した処理の一覧.	41

第1章 序論

1.1 はじめに

情報通信技術の発達，普及に伴いその利用範囲は拡大している．身近な情報通信技術の一つに電話などの音声通信技術がある．音声通信技術は音声分析合成の技術と共に発達してきた．音声分析合成の技術発達に伴い，少ない情報量でより明瞭な音声通信を行う技術が開発されてきた．

現在利用されている多くの音声通信では，音声は一度デジタル化され，デジタル化されたデータが伝達される．データの受け取り手は，受け取ったデータをアナログ変換し，音声として出力する．このデジタル化された音声データにおいて，その情報を改ざんすることで，音声通信には様々な問題が生じる [1]．現在 STRAIGHT や WORLD といった音声分析合成技術を用いることで，非常に高精度な合成音声を作成することが可能である [2,3]．一方でこれらの技術を悪用した場合，図 1.1 に示されるような音声通信の改ざんや話者のなりすましが可能となってしまう．そのため今後の音声情報通信では，音声通信に真正性を保証する情報を付与し，通信の安全性を確保することが必要になる．

これまで著作権保護などを目的として，コンテンツに情報を付与する技術が研究されてきた．この技術のひとつには，コンテンツファイルのヘッダ部分に情報を付与する手法がある．しかしこの手法は，付与情報の除去が容易であり，利用には多くの問題が残される．情報の除去が困難で，コンテンツに情報を付与する技術として，「マルチメディア情報ハイディング技術 (Information hiding)」がある [4]．これは，デジタル化されたコンテンツデータに不可分な形で情報を埋め込む手法である．

この技術において音声を対象とした場合，通信の真正性を保証する情報を埋め込み，検出が可能であれば，音声通信を安全に利用することが可能となることが考えられる．例えば，信号の発信者は，音声に真正性を保証する情報（個人番号や機器の識別番号など）を付与した音声信号を生成する．情報が付与された音声を受信者に伝達され，受信者は音声から埋め込まれた情報を検出する．検出された真正性を保証する情報によって，音声通信の真正性を確認することができる．音声信号に改ざんやなりすましが加えられた場合，信号に埋め込まれた情報は破壊され，真正性を保証する情報が検出ができなくなる．

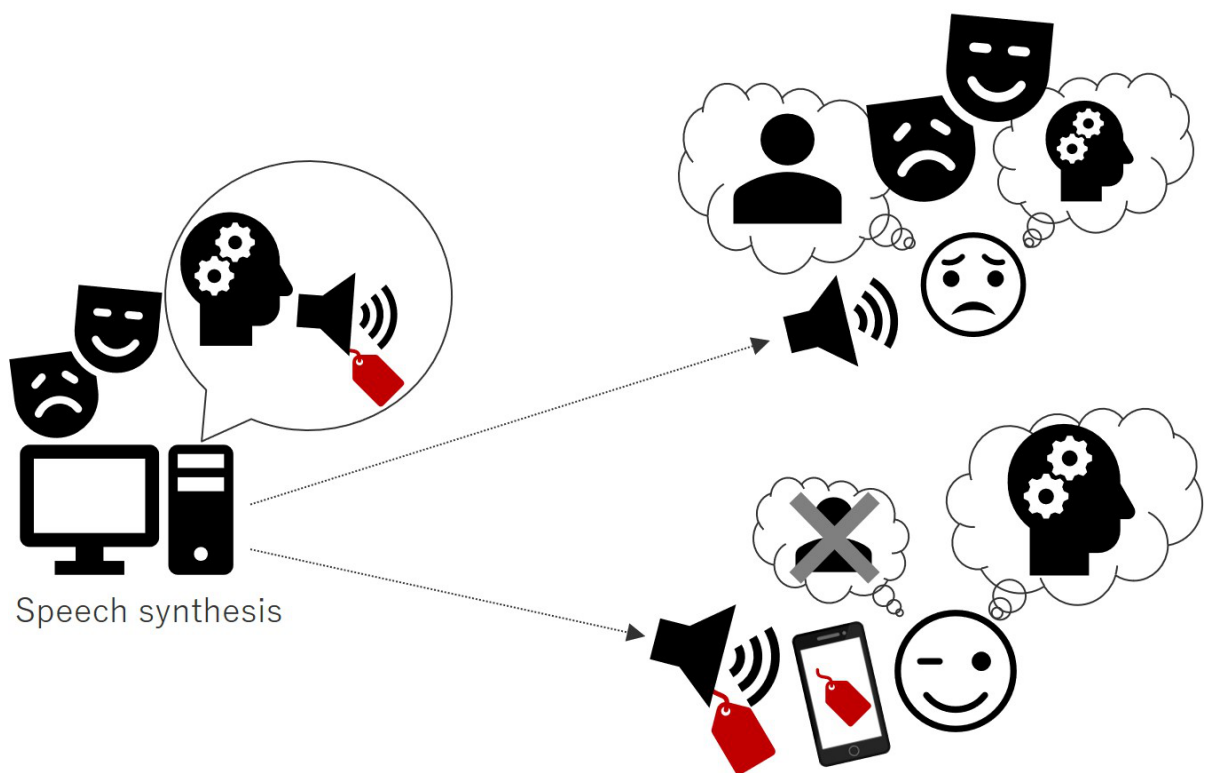


図 1.1: 話者のなりすましや改ざん.

1.2 情報ハイディング技術

様々な場面において、コンテンツが持つ主たる情報以外の情報を伝達することが必要とされている。映画や音楽における製作者の情報や、紙幣における偽装防止のための透かしなどがこれにあたる。これらの情報はデジタルデータにおいて、データファイルのヘッダ部へ格納されてきた。しかしこの手法では、該当箇所の情報を抜き出すなどしてその情報は容易に改ざんされてしまう問題がある。そのためコンテンツの購入者情報や、金銭における保証情報など、コンテンツに付随するデータが重要な意味を持つ場合は、その利用に大きな問題を抱えてしまう。この問題を解決するためには、コンテンツに付随する情報や伝達すべき情報をコンテンツ自体に不可分な形で持たせることが必要となる。これを可能にする技術として、デジタル化されたコンテンツデータ自体に情報を埋め込む、マルチメディア情報ハイディング技術がある [5]。マルチメディア情報ハイディングは、画像や音響信号など様々な場面で利用されているが、対象を音響信号とした場合、この技術は音響情報ハイディング (Audio information hiding, AIH) と呼ばれる。これまで、音楽の著作権保護などを目的に AIH の研究が行われてきた [6]。

これらの AIH 技術を音声に用いて情報を埋め込むとき、これは「音声情報ハイディング」と呼ばれる。しかし AIH 技術の多くは、音楽に特化したものであり、音声発話に含まれる無音区間や、音声特有の符号化などに対応できないといった問題を抱えている。そのため、AIH 技術を用いて音声に情報を埋め込むことは困難である。また音声情報ハイディングでは、音声信号に情報が埋め込まれていることが知覚されないこと (知覚不可能性)、情報が埋め込まれた音声に外乱などが加えられた場合でも、頑健に埋め込まれた情報が検出できること (頑健性) といった要求項目を満たすことが求められている。

上記の要求項目を満たした音声情報ハイディング法が完成することで、音声通信に通信の真正性を保証する情報を埋め込むことが可能となり、音声情報通信の安全性を高めることが可能になると考えられる。

1.3 研究の目的

音声通信において通信の安全性を保証するため、知覚不可能性と頑健性を有した音声情報ハイディング法が必要とされている。高い頑健性を持つ情報ハイディング法として、スペクトル拡散型の情報ハイディング法が知られている。AIH で例を挙げるとこれは、情報を音源とは異なる別の信号に拡散することで音声に情報を埋め込む手法である。しかしこの手法は、知覚不可能性（雑音を知覚しやすい）の問題を抱えている。この要因として、情報を拡散する信号に PN (Pseudo-random noise) 信号を用いていることが挙げられる。PN 信号は広い周波数帯域にスペクトルを持つ信号であり、この信号に情報を拡散し、これを音声に足し合わせることで情報を埋め込んでいる。そのため情報を埋め込んだ音声も、広い周波数帯域にスペクトルを持つことになり、これが雑音として知覚される。ここで情報拡散する信号が、音声と同じような周波数帯域にスペクトルを持つ信号であれば、情報が埋め込まれた音声の持つ周波数帯域は大きく変化せず、知覚不可能性が向上することが考えられる。

音声と同じような周波数帯域を持つ信号として、線形予測分析において算出される線形予測残差に着目した。線形予測残差は原音声の周波数成分を有する信号で、原音声に類似した周波数成分を持つ。そのため、線形予測残差に情報をスペクトル拡散する音声情報ハイディング法は、高い知覚不可能性を持つと考えられる。

本研究では、線形予測分析において算出される線形予測残差が、原音声と同じ周波数成分を持つことに着目し、音源情報成分である線形予測残差にメッセージ情報をスペクトル拡散する、音声情報ハイディング法を提案する。また提案法が知覚不可能性、頑健性の要求項目を満たした音声情報ハイディング法であるか評価する。

1.4 本論文の構成

本論文は、6 章で構成される。図 1.2 に本論文の構成を図示する。

1 章は序論であり、本論文で対象とする研究課題と研究の目的を述べる。

2 章は情報ハイディングとその代表的な手法についての特徴を述べ、本論文の対象となる音声への情報ハイディングの概要を述べる。

3 章は提案法のきっかけと、これを用いた音声への情報ハイディング法の原理を述べる。

4 章は提案法を用いた音声情報ハイディングによるメッセージ情報の埋め込み、検出について述べる。

5 章は提案法による音声情報ハイディング法を知覚不可能性、頑健性の観点で評価した尺度、方法、結果を述べる。

6 章は結論であり、本研究で明らかになったこと、残された課題について述べる。

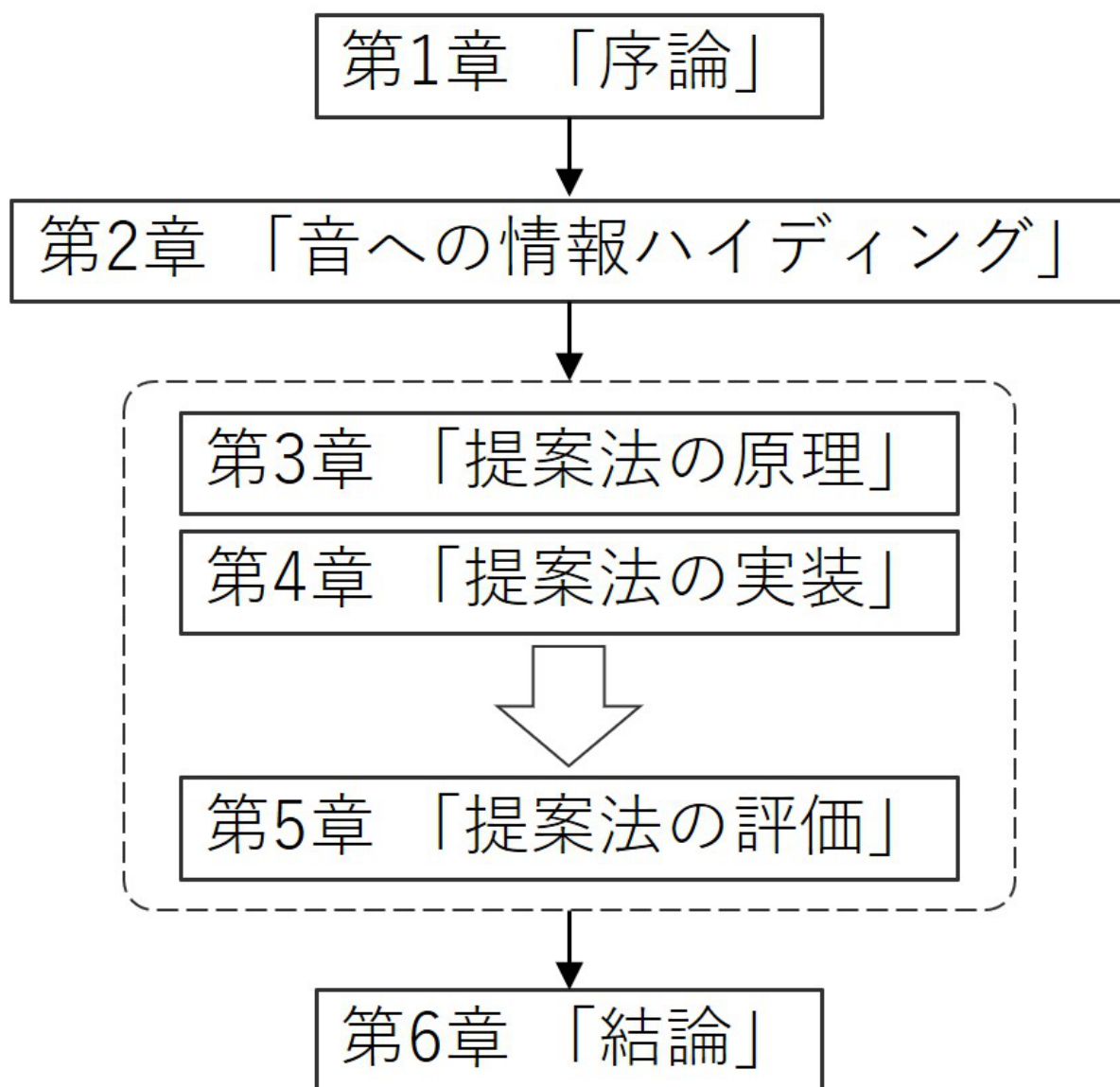


図 1.2: 論文の構成.

第2章 音への情報ハイディング

2.1 音響情報ハイディング

音楽コンテンツのデジタル化や、コンテンツのインターネット販売の増加に伴い、デジタル音楽コンテンツの利用は広まっている。しかし、これはコンテンツのコピーの容易さをも促進させ、コンテンツの利益を守るため、コンテンツの不正コピーの防止、著作権保護が重要な課題となっている。

音楽コンテンツの保護法の一つに DRM (Digital rights management) がある。DRM は音声コンテンツの利用に鍵情報を要求し、鍵情報を持たないユーザー、機器ではコンテンツが利用できなくなる手法である。これにより特定のユーザー、定められた機器間のみで音楽コンテンツが利用可能となる著作権保護法である [7]。しかしこの保護法は、ユーザーの不便さを招いており、この保護法を用いないデジタル音声コンテンツの販売も多く行われている。DRM などの手法を用いない著作権保護技術として、コンテンツの暗号化や、コンテンツファイルのヘッダ部への情報付与法がある。しかしこれらのコンテンツにも問題は多く、暗号化では解読されたコンテンツの保護の問題、ヘッダ部への情報付与法では付与部分の改ざんや削除などが指摘される。

これらの技術に対し、コンテンツに不可分な形で情報を埋め込み、検出する技術に「情報ハイディング技術」がある。図 2.1 は情報ハイディングの概要を示す。情報ハイディングのなかでも、音響信号を対象とした柔軟な著作権保護法として AIH がある。AIH はコンテンツのデジタル信号自体に付与情報を埋め込む技術のため、情報の除去が困難であり、ユーザーはコンテンツを権利の範囲で自由に利用することが可能である。

AIH は著作権保護以外にも、広い範囲でその利用が考えられている。その例に放送音声への情報付与がある [8, 9]。これは災害時の避難放送や、駅の構内放送などで利用される放送音声に情報を付与しておくことで、音情報だけでなくスマートフォンなどを用いた知覚情報で情報を共有することが可能となる例である。このように音のユニバーサルデザインなど、目的は違えど AIH は広く有効的な活用方法が考えられている。

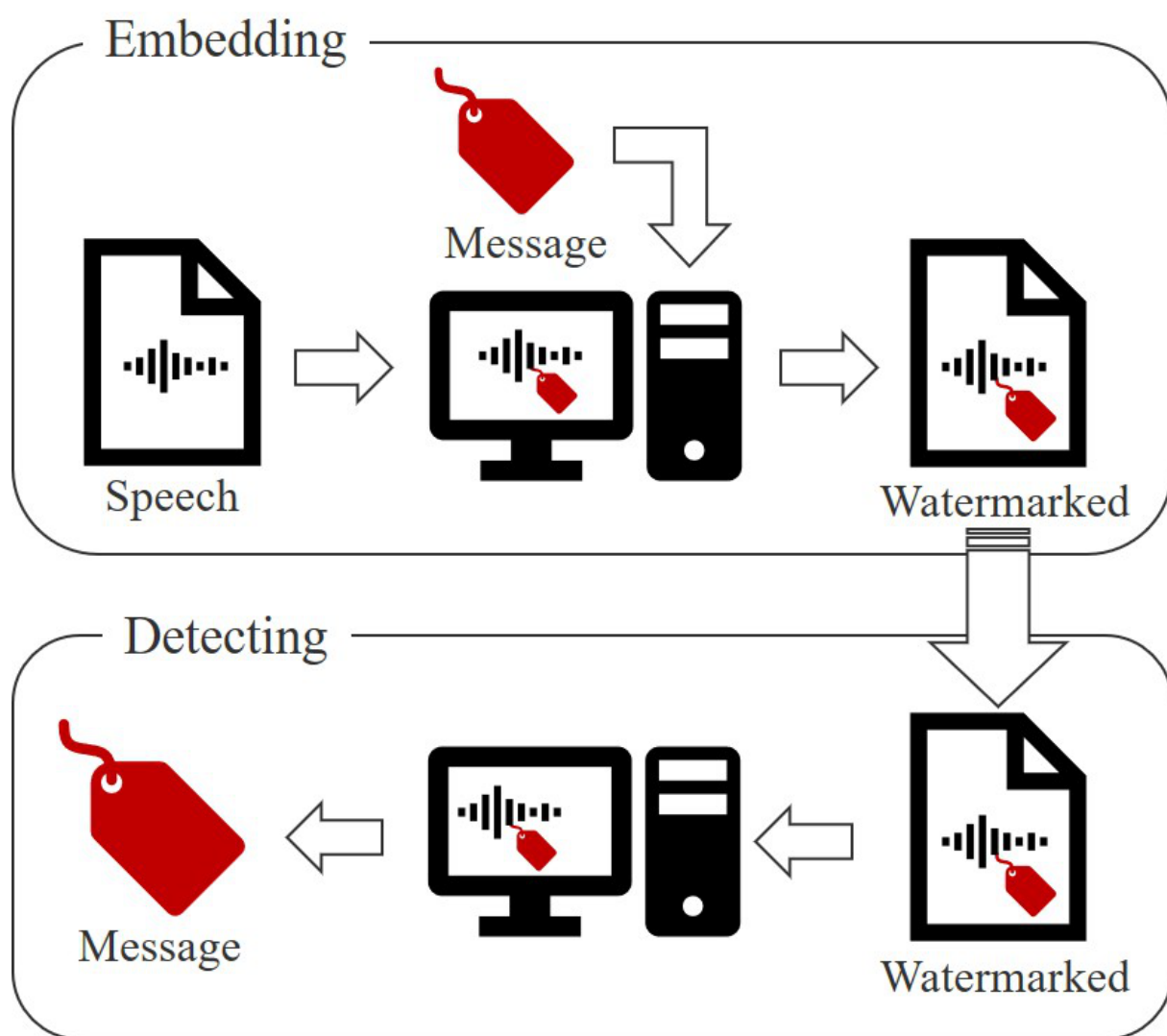


図 2.1: 情報ハイディングの概要.

2.2 代表的な情報ハイディング法

情報ハイディングにおける情報の埋め込み、検出は、通信の特徴を利用する手法や、音響モデルに基づく手法など様々な種類がある [10,11]. 情報ハイディングはその目的によって大きく二つの分類分けが存在する [12]. LSB などの情報を秘密裏に伝達することを目的とした情報ハイディングは「ステガノグラフィ」と分類される. エコーハイディング法などの原音声に関係性のある情報を付与することを目的とした情報ハイディングは「電子透かし」と分類される. ステガノグラフィでは、情報を埋め込む信号を「カバー信号」、埋め込む情報を「秘密情報」、情報が埋め込まれた音声を「ステゴ信号」と呼称する. 電子透かしでは、情報を埋め込む信号を「ホスト信号」、埋め込む情報を「透かし信号」、情報が埋め込まれた音声を「透かしが埋め込まれた信号」と呼称する.

以下に代表的な情報ハイディング法を紹介する.

2.2.1 最下位ビット置換法

量子化されたコンテンツに情報を埋め込む基本的な手法として、LSB (least significant bit) 置換法がある [13,14]. LSB 置換法はコンテンツをデジタル化処理する際に、量子化された最下位ビットに変更を加え情報を埋め込む手法である.

情報の検出は伝達された信号の最下位ビットを抜き出すことで行なわれる. LSB 置換の処理例を表 2.1 に示す.

LSB 法による情報ハイディングの利点にはその原理が単純であるため、非常に高速に処理することが可能であること、カバー信号は情報が伝達されれば問題ないため、どのような信号も利用可能であること、1 サンプルに 1 bit の情報が付与できるため比較的埋め込み要領が多いことが挙げられる. 欠点には、攻撃耐性が非常に低いことが挙げられる. ステゴ信号最下位ビットにあたるビットが変形した場合、情報の検出ができなくなるため、時間伸縮やフィルタリングなどの信号処理を施すだけで、情報検出が困難となる.

表 2.1: LSB 置換による処理例.

カバー信号		メッセージ	ステゴ信号	
10 進数	2 進数		2 進数	10 進数
5	1000	0	1000	5
11	1011	1	1011	11
1	0001	0	0000	0
6	0110	1	0111	7

2.2.2 エコーハイディング法

室内音響において、聴取者は音源から発生した直接音だけでなく、壁などを反射した反射音も聴取している。この時聴取者は、直接音と反射音を融合音として知覚している。この反射が単純なものは、エコー知覚と呼ばれている。エコー知覚において、直接音と反射波の音圧が等しく、二音のずれが短い（約 0.6 ～ 1.0 ms）場合、一つの合成音として知覚されることが分かっている [15]。エコーハイディング法では、図 2.2 に示す様なエコー時間 T_m とエコーの振幅 a_m の二つのエコーを生成し、この差によって透かしを埋め込む [16]。

情報埋め込みでは、まずホスト信号に対して二種類のエコー時間 T_m を持つエコーを生成する。生成されたエコーは透かし情報の 0 と 1 に対応し、下式に従ってホスト信号 $x(n)$ に透かし情報に対応したエコーを付与することで、透かしが埋め込まれた信号 $y(n)$ を得る。

$$y(n) = x(n) * h_m(n) \quad (2.1)$$

ここで $h_m(n)$ は透かしとして埋め込まれる信号の “1” と “0” に対応して下式に従って得られる。

$$h_0(n) = \delta(n) + a_0\delta(n - T_0) \quad (2.2)$$

$$h_1(n) = \delta(n) + a_1\delta(n - T_1) \quad (2.3)$$

情報検出では、透かしが埋め込まれた信号を離散フーリエ変換し、対数化し再度逆離散フーリエ変換することで、図 2.3 に示す様なエコー時間 T_m に対応したピークを得ることができる。このピークからエコー時間 T_m に対応した 0 と 1 の透かし情報を得る。

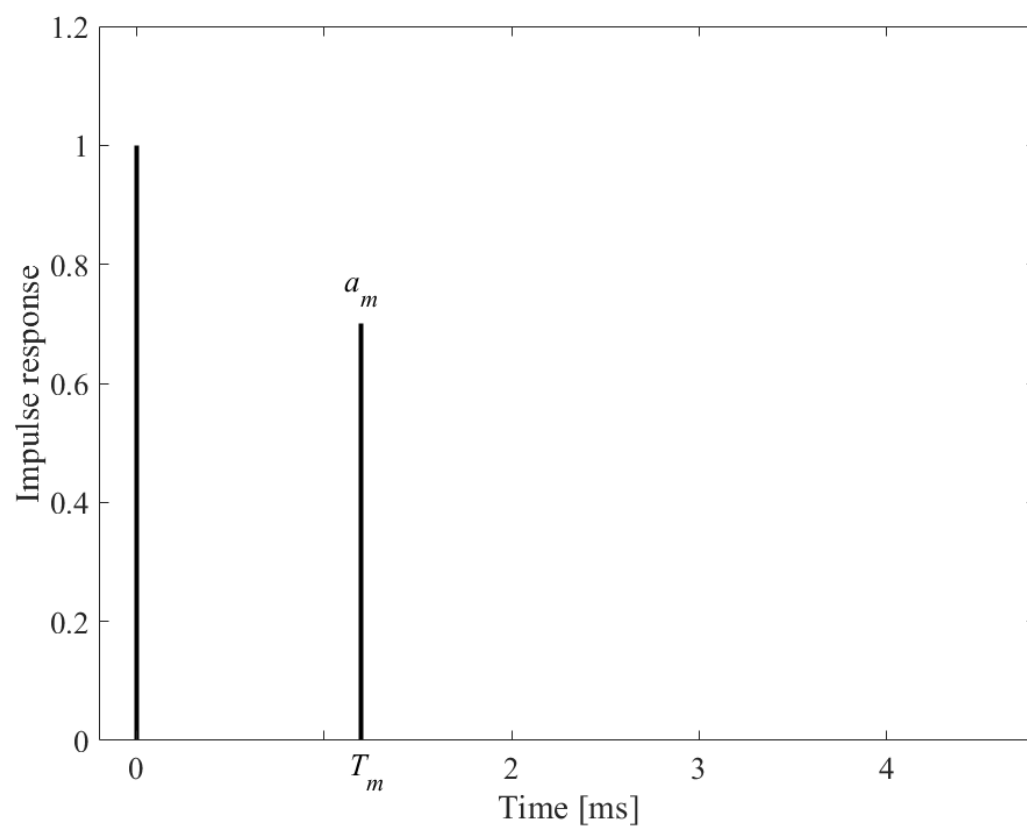


図 2.2: エコーハイディング法におけるエコーの設定.

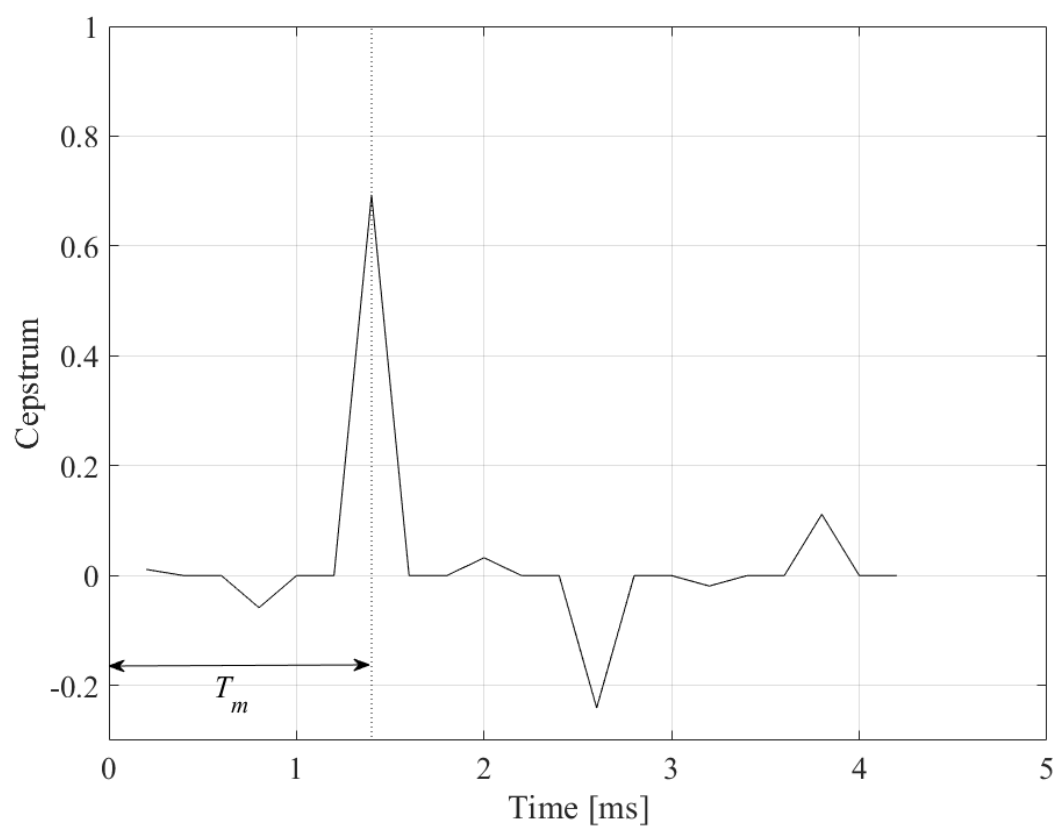


図 2.3: エコーハイディング法における情報検出例.

2.2.3 周期的位相変調法

聴覚系においてヒトは、音響信号における位相の変化に鈍感であるといわれている [18]. 西村, 鈴木による研究では、音響信号の種類に依存するが、変調周波数が 10 Hz 以下のゆるやかな変調であれば、ヒトは変調による音の歪みを知覚できないことが示された [19] [20].

これより周期的位相変調法では、下式に従ってホスト信号 $x(n)$ に図 2.4 に示すような二種類の位相変調を施すことで透かし情報を埋め込む [21].

$$y(n) = x(n) * w(n) \quad (2.4)$$

ここで、 $w(n)$ は二種類の位相周波数の変調を操作した信号である。 $w(n)$ における各パラメータは時変であり、透かし情報 m によって下式に従って決定する.

$$f_{\text{PM}} = \begin{cases} f_0, & m = 0 \\ f_1, & m = 1 \end{cases} \quad (2.5)$$

ここで、 f_0 と f_1 は、変調周波数が 10 Hz 以下のゆるやかな変調であれば問題ないことから、一般に $f_0 = 6$ Hz と $f_1 = 8$ Hz が利用される [22]. 定められた変調周波数 f_0 と f_1 より、変調量 ω_0 が下式によって決定される.

$$\omega_0 = \frac{\omega_{\max} - \omega_{\min}}{2} (1 - \cos(2\pi f_{\text{PM}} t)) + \omega_{\min} \quad (2.6)$$

情報の検出ではホスト信号と透かしが埋め込まれた信号の位相変化量を比較し、透かしの埋め込みに使用した変調周波数と比較することで、情報検出を行う. この他にもパラメトリック型のブラインド推定法も存在する [23, 24].

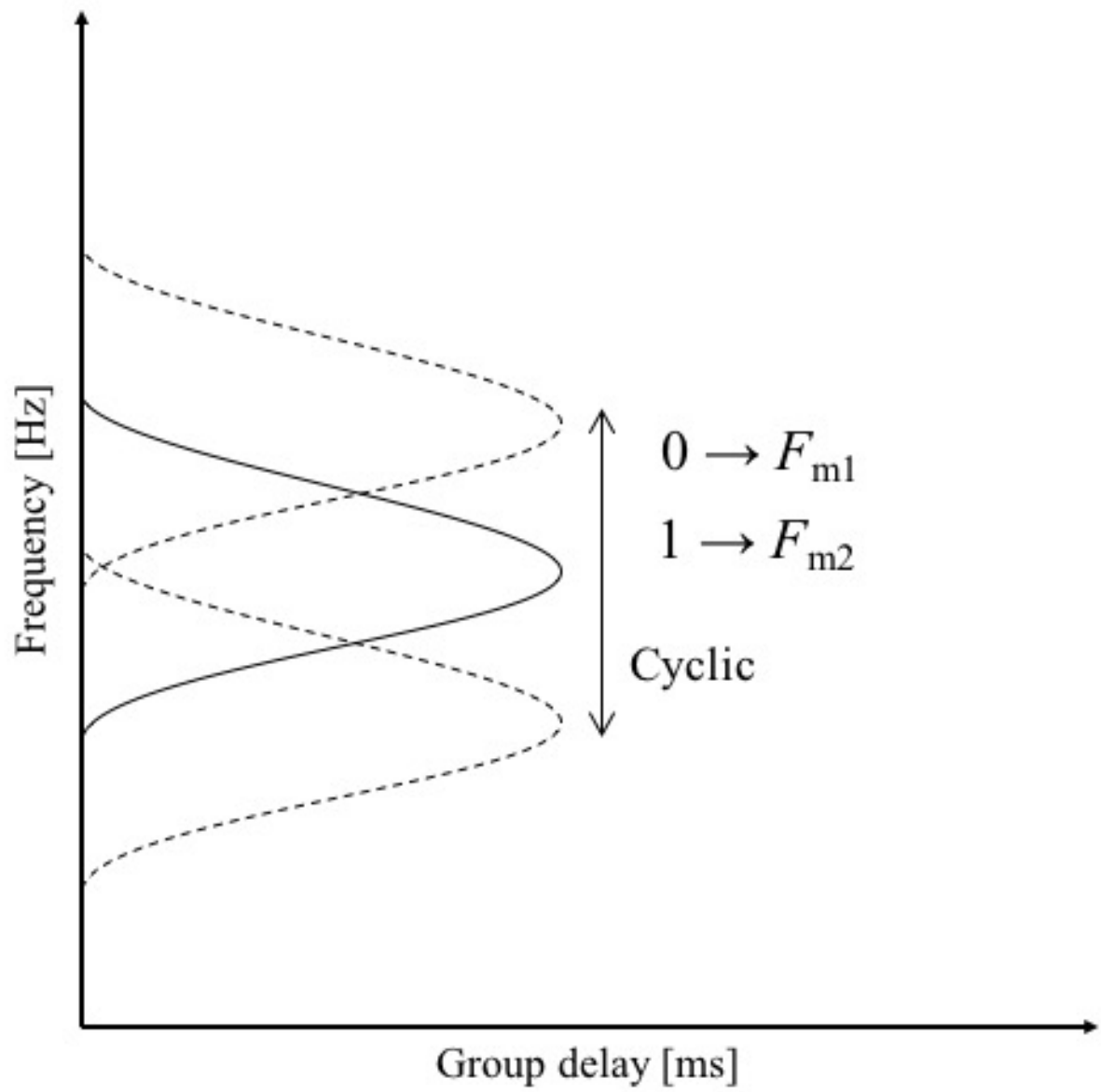


図 2.4: 周期的位相変調法における群遅延特性の概要.

2.2.4 蝸牛遅延を用いた情報ハイディング法

聴知覚において重要な役割を持つ蝸牛において、「蝸牛遅延」と呼ばれる遅延が生じることが知られている [25]. これは鼓膜内側の内耳に存在する蝸牛において、長い基底膜と呼ばれる器官で音を知覚することに要因を持つ. 基底膜はその基部（入口側）で高い周波数に対応して振動し、頂部（末端側）で低い周波数に対応して振動する. そのため、図 2.5 に示す様に広い周波数域をもつパルスが入力されたとき、基底膜では各周波数成分に対応した位置で基底膜が振動する. 蝸牛内に音振動が伝わると、まず基底膜の基部で高い周波数成分の振動が始まり、振動が伝わった後、基底膜の頂部において低い周波数成分の振動が起こる. このとき同じ入力であっても、周波数ごとに基底膜の振動の開始時間が遅れ蝸牛遅延と呼ばれる遅延が発生する [26]. しかし、蝸牛遅延によって周波数成分は物理的に遅延を生じているが、知覚的には同時に知覚されることも知られている [27]. また饗庭、津崎によれば、ヒトは原信号と蝸牛遅延に沿って増長した遅延（低周波成分の振動開始が遅い）信号の弁別が困難であることが示されている [28]. これらの聴覚特徴を利用した情報ハイディング法として、ホスト信号に透かし情報を遅延として埋め込む、蝸牛遅延による透かし入れがある [29, 30].

蝸牛遅延による透かし入れでは、下式によって与えられる位相特徴を設定した、IIR フィルタ $H_m(z)$ の群遅延 $\tau_m(\omega)$ をホスト信号に付与することで、聴覚的に知覚されにくい情報ハイディングを行う.

$$H_m(z) = \frac{-b_m + z^{-1}}{1 - b_m z^{-1}} \quad (2.7)$$

$$\tau_m(\omega) = -\frac{d \arg(H_m(e^{j\omega}))}{d\omega} \quad (2.8)$$

ここで $H_m(e^{j\omega}) = H_m(z)|_{z=e^{j\omega}}$ である. 情報埋め込みでは、埋め込む情報 ($m = "0"$ or $"1"$) に対応したフィルタ係数より、式 (2.8) に従って二種類の IIR フィルタ $H_m(z)$ を生成し、透かし情報としてホスト信号に埋め込む. 情報の検出では、ホスト信号と透かしが埋め込まれた信号の位相スペクトルを計算し、これが埋め込みの際に設定した位相特徴に合致することから、埋め込まれた情報を検出する. 上記以外にも情報検出にホスト信号を用いないブラインドな情報検出法も提案されている [31, 32].

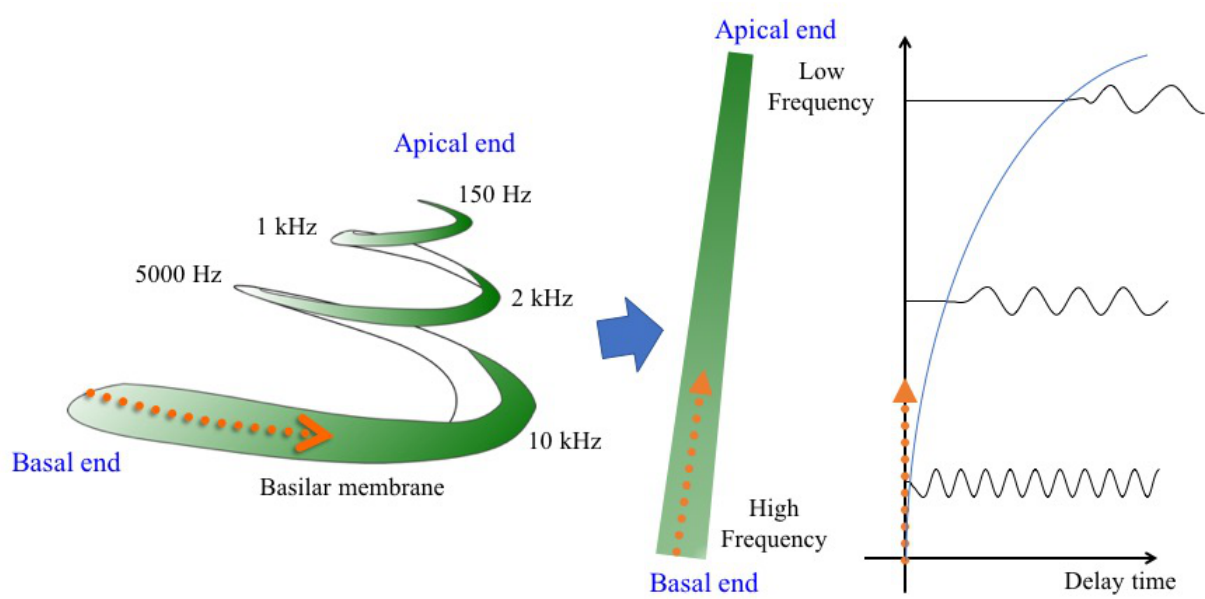


図 2.5: 蝸牛の構造と基底膜振動で見られる蝸牛遅延.

2.2.5 直接スペクトル拡散法

頑健な音響情報ハイディング法として直接スペクトル拡散 (Direct spread spectrum, DSS) 法がある [33,34]. DSS 法では, 次式のようにメッセージ情報 $m(n)$ を, PN (Pseudo-random noise) 信号 $c(n)$ のもつ, 広いスペクトルでスペクトル拡散変調することで透かし信号 $w(n)$ を生成する. その信号をホスト信号 $x(n)$ に加算することで情報が埋め込まれた信号 $y(n)$ を生成する.

$$y(n) = x(n) + w(n) = x(n) + am(n)c(n) \quad (2.9)$$

ここで a は, 透かし信号 $w(n)$ の埋め込み強度を調整する振幅補正量である.

情報検出では, 次式のように情報が埋め込まれた信号 $y(n)$ に, 埋め込みに利用した PN 信号を乗じ, その時の直流成分 (時間平均) の値を判別することでメッセージ $m(n)$ を検出する.

$$m = \begin{cases} 0, & E\{y(n)c(n)\} \leq 0 \\ 1, & E\{y(n)c(n)\} > 0 \end{cases} \quad (2.10)$$

$E\{\cdot\}$ は期待値であり, $x(n)$, $y(n)$, $c(n)$ に関してはエルゴード性を仮定する. そのため, $E\{\cdot\}$ は “.” に対して時間平均を求めたものに等しくなる. ここで PN 系列信号は, $E\{c(n)\} = 0$ と $E\{c^2(n)\} = 1$ の性質を持つ. これにより, 情報が埋め込まれた信号 $y(n)$ に埋め込みに利用した PN 信号 $c(n)$ と同じ信号を乗じることで以下を得る.

$$\begin{aligned} E\{y(n)c(n)\} &= E\{[x(n) + m(n)c(n)]c(n)\} \\ &= E\{x(n)c(n)\} + E\{Lm(n)c^2(n)\} \\ &= am(n) \end{aligned} \quad (2.11)$$

ここで $x(n)$, $c(n)$ は互いに独立である. このことから式 (2.11) より第一項は 0 となり, 第二項は $E\{c^2(n)\} = 1$ となる. これより $E\{am(n)c^2(n)\} = am(n)$ が出力され, DSS 法ではメッセージを検出することが可能である. 図 2.6 に DSS 法のブロックダイアグラムを示す. ここで図 2.6(a) はメッセージの埋め込み方法を, 図 2.6(b) は情報の検出方法を表す.

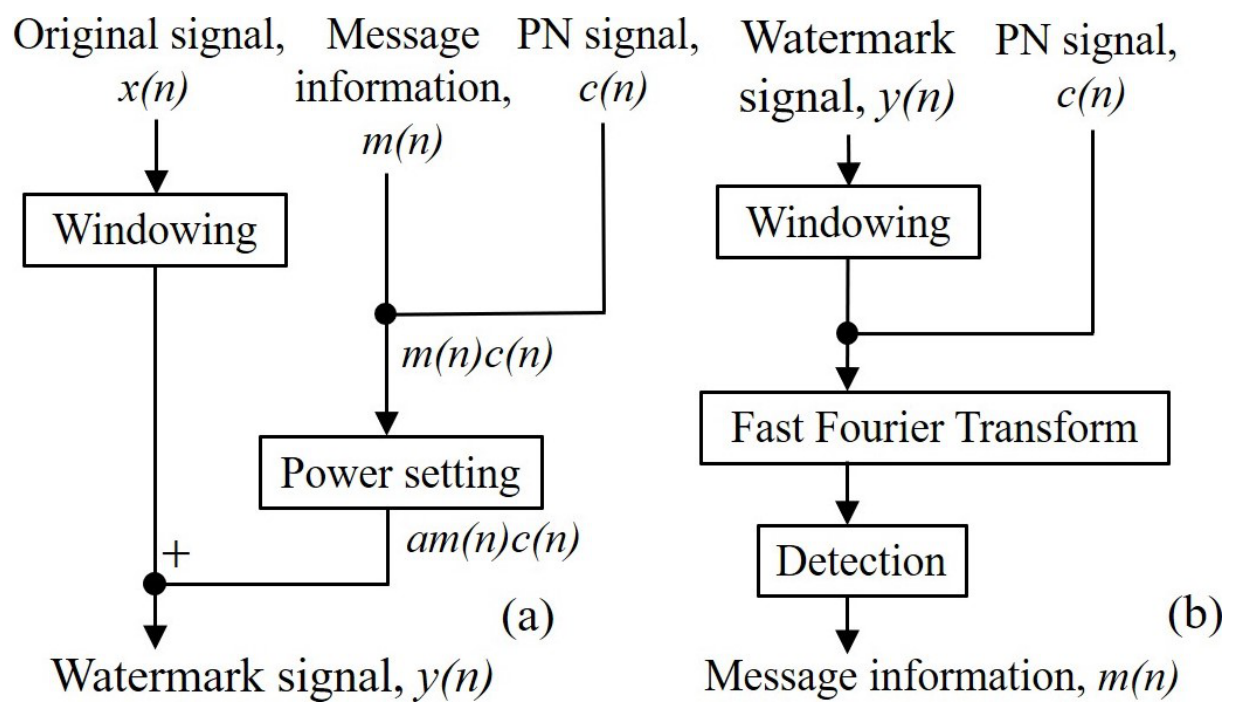


図 2.6: 直接スペクトル拡散法のブロックダイアグラム : (a) 埋め込み処理, (b) 検出処理.

2.3 音声信号への展開

近年音声分析合成が非常に発達してきている [35,36]. STRAIGHT や WORLD などの音声分析合成系の発達に伴って、この技術を悪用し、音声通信における情報の改ざんやなりすましといった問題が危険視されている。これらの問題において、通信の真正性を保証する情報を音声信号に付与することで、通信の真正性を保証することが可能であると考えられる。

通信の発信者は情報の発信の際、音声信号に自身の通信の真正性を保証する情報（個人番号や機器識別番号など）を信号に付与する。このとき、情報の付与はホスト信号と分離不可能な形でされることが望まれる。通信の受信者は、信号から真正性を保証する情報を検出し、通信の安全を確認することができる。信号に改ざんなどが加えられた場合、情報の検出が不可能となり、通信の安全性は保証される。

音声通信に真正性を保証する情報を付与する技術として、音声信号を対象に情報を埋め込む「音声情報ハイディング技術」が検討されている [37-39]。しかし、既存の音響情報ハイディング法の多くは、情報の埋め込み対象を音楽としているため、単純に埋め込み対象を音声に変更しただけでは、利用にいくつかの問題が発生する。問題には、音声特有の符号化処理を施した信号から透かし信号を検出することが困難となることや、音声は音楽のような周期性を持たない信号であり、透かしの知覚が容易であることなどが挙げられる [40]。他にも、音声に透かしが埋め込まれることで、音声の通話品質を低下させることはあってはならない。これらの問題を解決した音声情報ハイディング技術が完成されれば、音声情報通信における改ざんやなりすましの危険性を解決することが可能となる。

2.4 音声情報ハイディングにおける要求項目

音声情報ハイディングは下記の要求項目を満たすことが求められる [41]。

1. 知覚不可能性
2. 頑健性
3. 埋め込み容量

知覚不可能性とは、聴取者が透かしが埋め込まれた信号を聴取した際に、透かし信号の情報の存在に気づかないことである。

頑健性とは、透かしが埋め込まれた信号に何らかの外乱（信号処理や符号化など）が加えられた場合でも、透かし信号、埋め込まれた情報が正しく検出できることである。

埋め込み容量は、目的に合わせた埋め込み容量を確保した上で、上記の項目を担保することである。

第3章 提案法の原理

3.1 スペクトル拡散法の音声への応用

音響情報ハイディング技術の一つに、DSS 法がある。図 3.1 に DSS 法による情報ハイディングの処理例を示す。ここで図 3.1(a) はホスト信号を示し、図 3.1(b) はそのスペクトルを示す。図 3.1(c) はメッセージ信号を示し、図 3.1(d) はそのスペクトルを示す。図 3.1(e) は拡散信号である PN 信号を示し、図 3.1(f) はそのスペクトルを示す。図 3.1(g) は透かしが埋め込まれた信号を示し、図 3.1(h) はそのスペクトルを示す。DSS 法は AIH において、情報処理や攻撃を受けた透かしが埋め込まれた信号からも頑健に情報を検出することが可能であり、強い頑健性を持つ。これは、DSS 法が図 3.1(f) に示されるように広い範囲にスペクトルを持つ PN 信号に情報を拡散するためである。しかしこれによって、透かしが埋め込まれた信号は、図 3.1(g) に示されるように広い範囲にスペクトルを持つこととなる。そのため、聴取者は透かしが埋め込まれた信号を歪んだ音声として知覚する。これより、DSS 法は知覚不可能性を有していない音響情報ハイディング法であるとされている。

DSS 法はその原理より透かしを埋め込む対象を音声へ変更することは容易である、DSS 法が非常に頑健な音響情報ハイディング法であることから、対象を音声とした場合でも、強い頑健性を持った音声情報ハイディング法となることが考えられる。しかし、拡散信号に PN 信号を利用していることから、透かしが埋め込まれた信号は歪み、知覚不可能性を担保した音声情報ハイディングとしては利用することが困難であることが考えられる。

DSS 法を用いた音声情報ハイディングにおいて、知覚不可能性の問題を解決する事ができれば、頑健性と知覚不可能性の両方を担保した音声情報ハイディング法となることが考えられる。

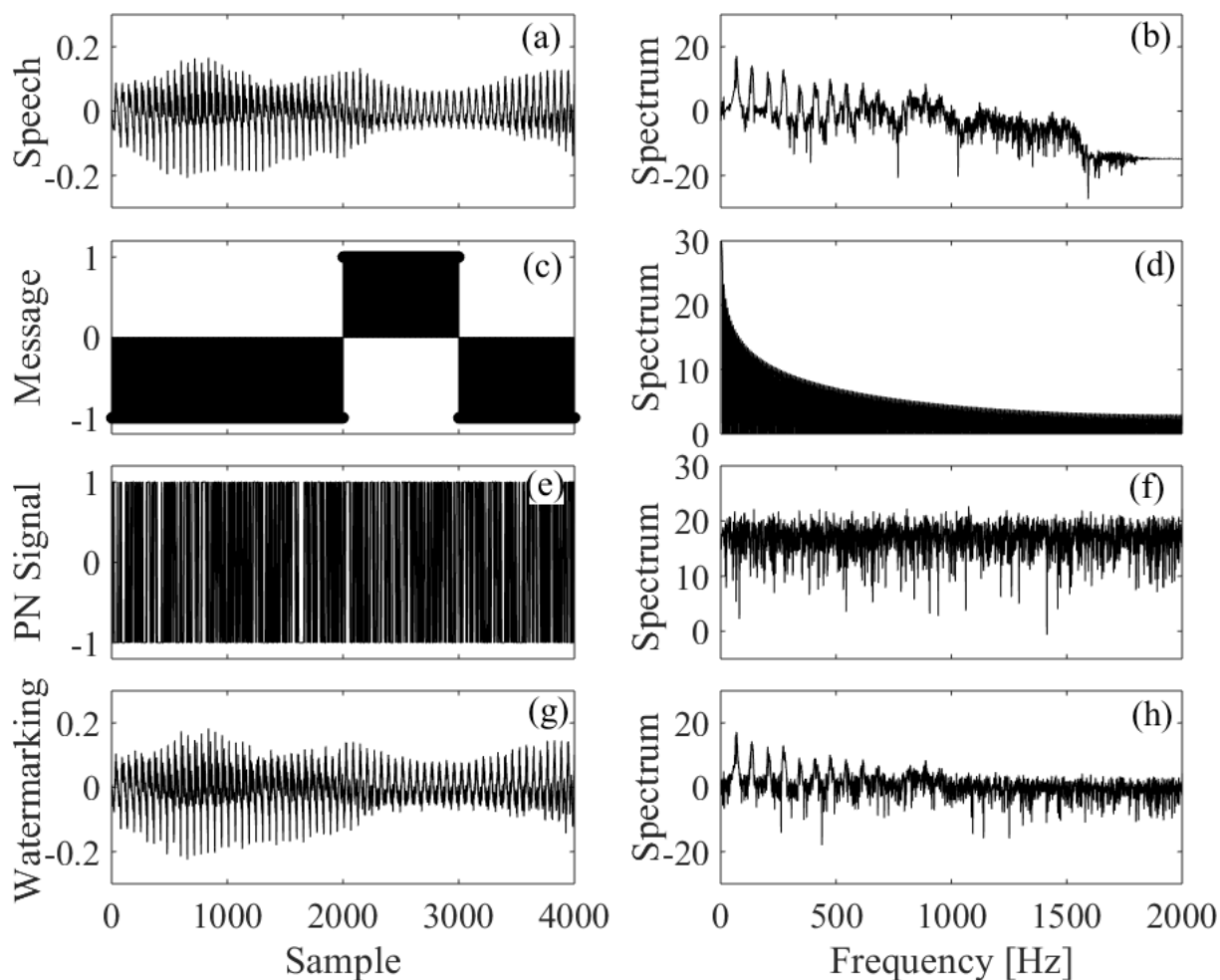


図 3.1: 直接スペクトル拡散法による情報ハイディングの処理例 : (a) 音声信号, (b) 音声信号のスペクトル情報, (c) メッセージ信号, (d) メッセージ信号のスペクトル情報, (e) PN 信号系列, (f) PN 信号系列のスペクトル情報, (g) 透かしが埋め込まれた音声信号, (h) 透かしが埋め込まれた音声信号のスペクトル情報.

3.2 拡散信号に求められる特徴

DSS 法において PN 信号は拡散信号 $c(n)$ の役割を持つ。埋め込み情報 $m(n)$ は拡散信号 $c(n)$ に周波数変調され、埋め込み情報が埋め込まれた透かし信号 $w(n)$ が生成される。拡散信号 $c(n)$ は式 (3.1) に示すように、その信号の期待値は 0、信号の二乗の期待値は 1 といった特徴を持つ [42]。そのため DSS 法では、式 (2.11) に示した情報検出において、透かしが埋め込まれた信号に再度同じ PN 信号を掛け合わせることで、ホスト信号項が “0”，透かし信号項における拡散信号部が “1” となり埋め込まれた情報が検出可能である。

$$\begin{cases} E\{c(n)\} = 0 \\ E\{c^2(n)\} = 1 \end{cases} \quad (3.1)$$

PN 信号は式 (3.1) の特徴を有するため、DSS 法において拡散信号として利用されている。しかし PN 信号は広いスペクトル帯域を有するといった白色性の特徴も有しているため、透かしが埋め込まれた信号も広い周波数域に成分を持つことになり、聴取者はそれを雑音として知覚してしまう。

式 (3.1) に示す特徴を有し、ホスト信号の持つ周波数域にのみに存在する信号を拡散信号として埋め込み情報をスペクトル拡散することで、DSS 法は知覚不可能性の高いスペクトル拡散型の音声情報ハイディング法として利用できることが考えられる。

このことから、下記の特徴を有する拡散信号を検討した。

- 式 (3.1) の特徴を持つ。
- ホスト信号の持つ周波数域にのみに存在する。

3.3 線形予測分析における線形予測残差

音声分析技術の一つに線形予測 (Linear prediction, LP) 符号化がある [43]。この技術は音声などの時系列信号を対象に、スペクトル包絡線を表すのに利用されている [44]。LP 符号化は発話の分析や再合成などに利用され、CELP などの音声符号化にも利用される。この時、LP 分析合成によって得られた LP 合成音声と原音声の差分は、LP 残差と呼ばれている。図 3.2 に線形予測分析合成の処理例を示す。図 3.2(a) は原音声を示し、図 3.2(b) はそのスペクトルを示す。図 3.2(c) は LP 合成音声を示し、図 3.2(d) はそのスペクトルを示す。図 3.2(e) は LP 残差を示し、図 3.2(f) はそのスペクトルを示す。

線形予測残差は図 3.2(e) に示されるように、原信号の周期的成分を持ち、図 3.2(e) に示されるように、そのスペクトル形状は原信号のスペクトル形状に類似している。また LP 残差は式 (3.1) の特徴を持つことが知られている。これらより、LP 残差は前節で示した、拡散信号に求められる特徴を満たした信号であり、拡散信号として利用可能であることが考えられる。

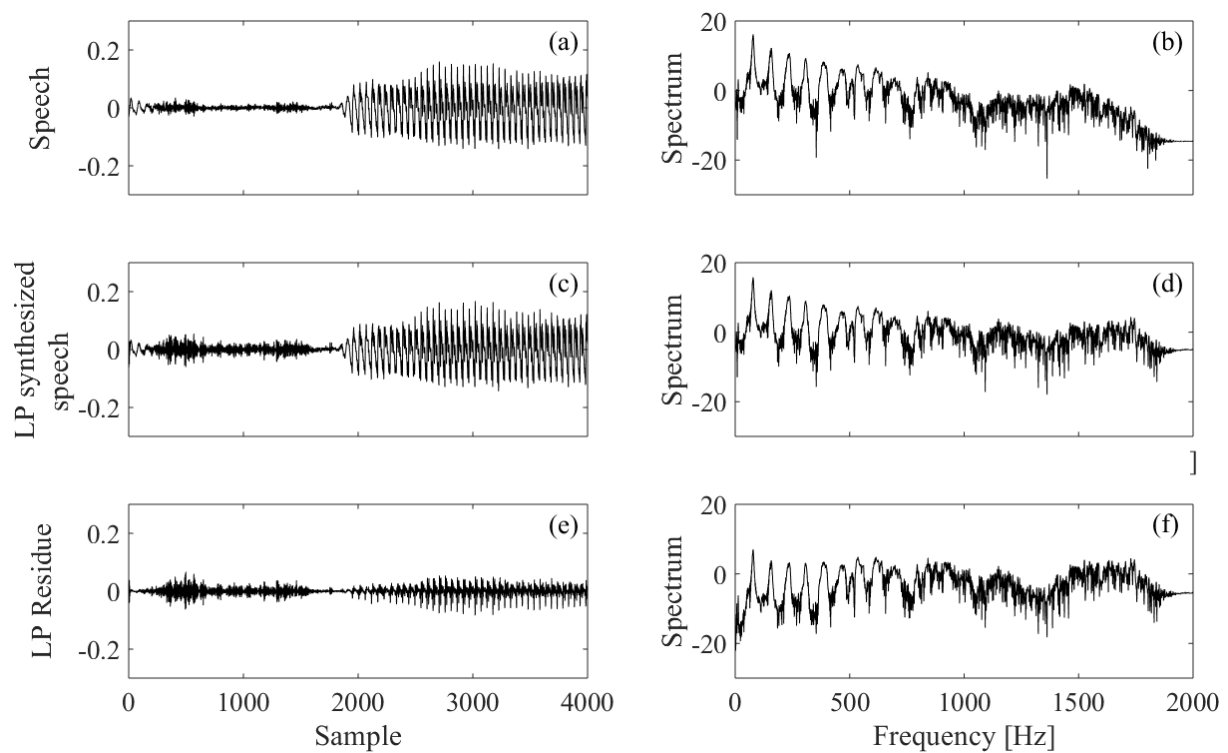


図 3.2: 線形予測分析合成の処理例 : (a) 音声信号, (b) 音声信号のスペクトル情報, (c) LP 合成音声, (d) LP 合成音声のスペクトル情報, (e) 線形予測残差, (f) 線形予測のスペクトル情報.

3.4 線形予測残差を用いた

スペクトル拡散型音声情報ハイディング法の提案

前節より、LP 残差がスペクトル拡散型の音声情報ハイディング法における拡散信号として有効であることが示された。本稿では、LP 残差を用いたスペクトル拡散型の音声情報ハイディング法を提案する。

提案法では下式に示すように情報が埋め込まれる。メッセージ情報 $m(n)$ は、ホスト信号 $x(n)$ から算出された LP 残差 $r(n)$ にスペクトル拡散変調され、透かし信号 $w(n)$ が生成される。透かし信号 $w(n)$ は埋め込み強度によって振幅を調節され、ホスト信号 $x(n)$ に足し合わされる。これによって透かしが埋め込まれた信号 $y(n)$ が生成される。

$$\begin{aligned} y(n) &= x(n) + w(n) \\ &= x(n) + am(n)r(n) \end{aligned} \quad (3.2)$$

提案法における情報検出は、DSS 法と同様に式 (3.3) を用いて次式のように行う。LP 残差 $r(n)$ は、式 (3.1) の性質を持つ。そのため透かしが埋め込まれた信号 $y(n)$ に情報の埋め込みに利用した、LP 残差 $r(n)$ を乗ずることで、下式のように第一項 $E\{x(n)r(n)\}$ は “0” となり、LP 残差 $r(n)$ が二乗される第二項 $E\{am(n)r^2(n)\}$ は $E\{r^2(n)\} = 1$ より $am(n)$ として出力される。

$$\begin{aligned} E\{y(n)r(n)\} &= E\{[x(n) + am(n)r(n)]r(n)\} \\ &= E\{x(n)r(n)\} + E\{am(n)r^2(n)\} \\ &= am(n) \end{aligned} \quad (3.3)$$

算出されたメッセージ信号 $am(n)$ はフーリエ変換され、その信号の垂直成分が下式に従って埋め込みメッセージ m として出力される。

$$m = \begin{cases} 0, & am(n) \leq 0 \\ 1, & am(n) > 0 \end{cases} \quad (3.4)$$

提案法は DSS 法における PN 信号の代わりに、LP 残差を利用して埋め込み情報をスペクトル拡散変調する。図 3.2(e) から分かるように LP 残差はホスト信号の周期成分そのものである。また図 3.2(f) に表れるように、そのスペクトル形状は非常に酷似しており、二つの信号の持つスペクトル周波数域が近い。このことから、メッセージ信号がホスト信号に含まれる高調波成分に情報が付与されることになる。さらには LP 残差による拡散と調波成分の二音は、同じ音源の音として知覚的に融合され、聴取者は埋め込みの存在に気付くことが困難となる。そのため、DSS 法と比較して全帯域に聞こえる雑音が減少し、ホスト信号とメッセージ信号が知覚的に融合され、音声の歪みは聴取者に知覚されにくくなる。これは Bregman の発見的規則で説明される知覚的融合によるものと解釈される [45]。また、提案法は DSS 法の原理を利用して音声電子透かしを行っているため、DSS

法の持つ頑健性を備えることが考えられる。これらより提案法は、知覚不可能性と耐性の問題の両方を解決した音声電子透かし法となることが考えられる。

提案法を音声通信において通信の安全・真正性保証として利用する場合、発信者は、音声信号に通信の真正性を保証する情報を、信号に埋め込む。受信者は、信号から真正性を保証する情報を検出し、通信の安全を確認することが可能となる。信号に改ざんなどが加えられた場合、受信者が持つLP 残差と改ざんが加えられた信号の拡散信号が異なるため、情報の検出が不可能となり、通信の安全性は保証される。

第4章 提案法の実装

4.1 情報の埋め込み

4.1.1 情報の埋め込みアルゴリズム

本節では、提案法におけるメッセージ情報の埋め込みアルゴリズムを示す。図 4.1 に提案法によるメッセージ情報の埋め込みのブロックダイアグラムを示す。情報の埋め込みにおいてまず、透かし信号 $w(n)$ を作成する。透かし信号 $w(n)$ は埋め込むメッセージ信号 $m(n)$ を拡散信号にスペクトル拡散変調することで得られる。ここでメッセージ信号 $m(n)$ は 1 と -1 の二値からなり、埋め込む情報を元にフレーム長分それぞれ作成される。拡散信号となる LP 残差 $r(n)$ は、原音声であるホスト信号 $x(n)$ をフレーム分割し重み付けを行い、LP 分析によって合成された合成音声信号とホスト信号の差分を取ることで得られる。重み付けには窓関数を使用した窓は Hanning 窓である。重み付けによってフレーム連続点における不連続な接続を減少させ、雑音の低下を図っている。また提案法ではフレーム長 250 msec、フレームシフト 125 msec、OLA (over-lab and add) を利用する条件で線形予測分析を行う。得られたメッセージ信号 $m(n)$ と LP 残差 $r(n)$ を乗算し、透かし信号 $w(n)$ を得る。透かし信号 $w(n) = an(n)r(n)$ における a は、透かしの埋め込み強度を調節するパラメータである。透かし信号 $w(n)$ はホスト信号 $x(n)$ に足し合わされ、透かしが埋め込まれた信号 $y(n)$ を得る。

図 4.2 に提案法を用いた音声情報ハイディングの処理例を示す。ここで図 4.2(a) はホスト信号を示し、図 4.2(b) はそのスペクトルを示す。図 4.2(c) はメッセージ信号を示し、図 4.2(d) はそのスペクトルを示す。図 4.2(e) は拡散信号である PN 信号を示し、図 4.2(f) はそのスペクトルを示す。図 4.2(g) は透かしが埋め込まれた信号を示し、図 4.2(h) はそのスペクトルを示す。

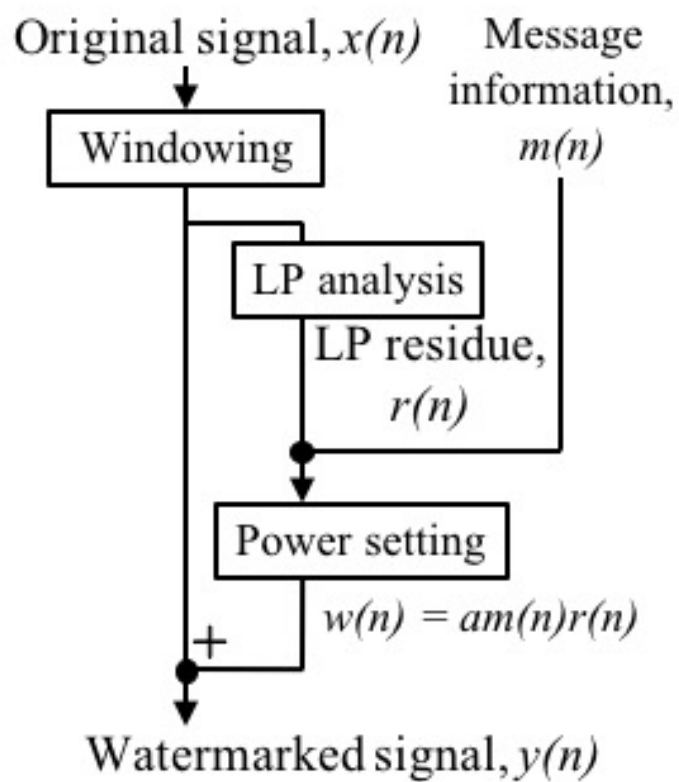


図 4.1: 提案法における情報埋め込みのブロックダイアグラム.

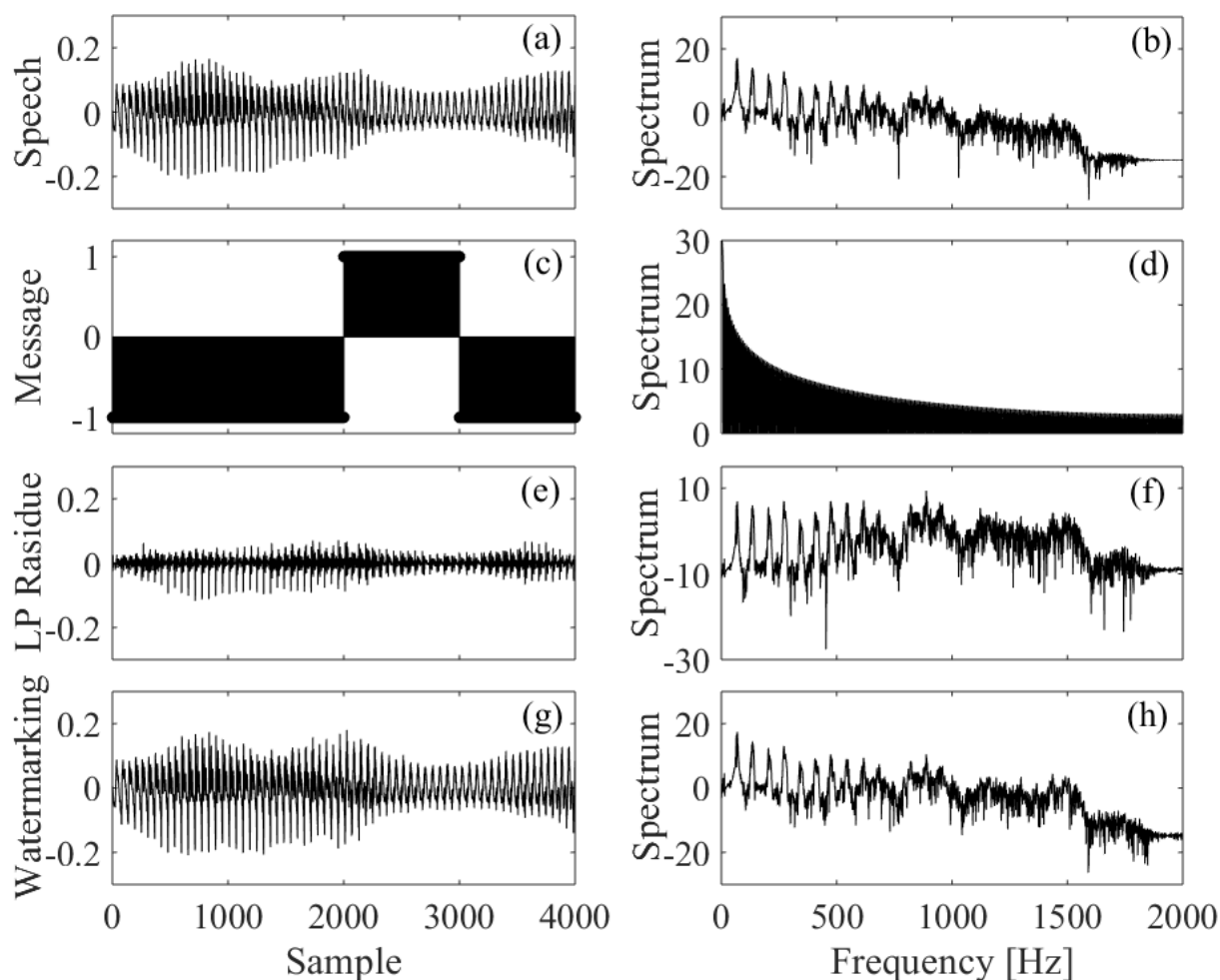


図 4.2: 提案法を用いた音声情報ハイディングの処理例 : (a) 音声信号, (b) 音声信号のスペクトル情報, (c) メッセージ信号, (d) メッセージ信号のスペクトル情報, (e) 線形予測残差, (f) 線形予測残差のスペクトル情報, (g) 透かしが埋め込まれた音声信号, (h) 透かしが埋め込まれた音声信号のスペクトル情報.

4.2 最適な埋め込みパラメータ設定

4.2.1 埋め込み強度補正

スペクトル拡散型の情報ハイディングにおいて情報を埋め込む際、ホスト信号のスペクトル帯域から少なからずずれを持つ透かし信号がホスト信号に足し合われる。そのため聴取者はホスト信号と透かしが埋め込まれた信号との差を音声の歪み（雑音）として知覚する。

この問題を解決するために DSS 法などのスペクトル拡散型の情報ハイディングでは透かし信号の振幅を調節し、ホスト信号へ足し合わせる透かし信号の振幅を小さくする処理が行われている。提案法においてもこの処理を施し、高い知覚不可能性を持つ音声情報ハイディング法を目指した。提案法ではこの処理を埋め込み強度補正と称し、透かし信号 $w(n) = am(n)r(n)$ における透かしの埋め込み強度を調節するパラメータ a によって調節した。

透かしの埋め込み強度を調節するパラメータ a は下式によって与えられる。

$$L_{\text{all}} = L_{\text{PHS}} - L_{\text{PSS}} + L_{\text{SSV}} \quad (4.1)$$

$$a = 10^{L_{\text{all}}/20} \quad (4.2)$$

ここで、 L_{all} は強度補正レベル、 L_{PHS} はホスト信号のパワーレベル (Power of Host Signal)、 L_{PSS} は拡散信号のパワーレベル (Power of Spread Signal)、 L_{SSL} は埋め込み強度設定レベル (Strength Setting Level) をそれぞれ表す。式 (4.1) によって得られた強度補正レベル L_{all} は式 (4.2) によって対数レベルへ変換され、透かしの埋め込み強度を調節するパラメータ a を得る。得られた透かしの埋め込み強度を調節するパラメータ a を元に透かし信号の振幅を調節し、ホスト信号へ足し合わせる。

透かしの埋め込み強度を調節するパラメータ a を調節することで、透かしが埋め込まれた信号の知覚不可能性が向上することが考えられるが、これと同時に透かし信号がホスト信号にマスクされ、検出精度が減少することが考えられる。本節では音声情報ハイディングとして最適な透かしの埋め込み強度を調節するパラメータ a を求めるために、埋め込み強度設定レベル L_{SSL} を $-20 \sim 20$ dB まで 5 dB 刻みで変更し最適値を選定した。

4.2.2 選定基準

最適な透かしの埋め込み強度を調節するパラメータ a を設定するにあたり，知覚不可能性と検出精度の基準を定めた．知覚不可能性については，透かしが埋め込まれた信号の音質によって評価した．検出精度については，検出誤差率（Bit error rate, BER）によって評価した．

音質の評価には対数スペクトル歪み（Log spectral distortion, LSD）と Perceptual Evaluation of Speech Quality（PESQ）を用いた [46]．LSD は音声の歪みを示し，その評価結果が 1 dB 以下のときに高い音質とされている．PESQ は主観評価に基づく客観評価であり，Objective difference grades（ODGs）によってその評価結果が示される．ODGs の出力結果は表 4.1 に示すように定められ，その評価結果が 0 から -1 のとき高い音質とされる．

検出精度について BER が 10% 以下のとき検出精度に問題がないとした．このとき，音声情報ハイディングにおいて必要とされる埋め込み情報量（ペイロード）は定義されていない．そこで本稿では，6 bps のペイロードで 10% の誤検出率で情報検出が可能なとき，音声情報ハイディングとして有用であると仮定した [46]．

LSD と PESQ の二つの音質評価と検出精度のそれぞれの評価基準を満たす，音声情報ハイディングとして最適な透かしの埋め込み強度を調節するパラメータ a を検討した．

評価には ATR データベース（B）より 12 種類の音声に対して発話区間 1 に対して，4 bps と 8 bps のペイロードにおいて選定を行った．

表 4.1: PESQ により出力される ODG 値．

対応評価基準	ODG
劣化知覚不能 (imperceptible)	0
違いがわかるが気にならない (perceptible, but not annoying)	-1
やや気になる (slightly annoying)	-2
気になる (annoying)	-3
非常に気になる (very annoying)	-4

4.2.3 選定結果

図 4.3 に 4 bps の場合の LSD と PESQ の二つの音質評価結果と検出精度の評価結果を、図 4.4 に 8 bps の場合の LSD と PESQ の二つの音質評価結果と検出精度の評価結果をそれぞれ示す。図の横軸はペイロードを示し、縦軸は各評価値を示す。

これらの結果より、埋め込み強度レベルが大きくなるにつれ BER が減少し、それに伴って情報埋め込み音声の音質評価結果が低下することがわかった。またこの結果はペイロードに依存しないこともわかった。

4 bps のペイロードの場合、埋め込み強度設定レベル L_{SSL} が -10 dB 以上の場合、BER が 10% 以下であり LSD と PESQ の音質評価の結果も基準値内のため、埋め込み強度設定レベル L_{SSL} は -10 dB が最適であると考えられる。このとき平均強度補正レベル L_{all} は -20.6 dB であった。

8 bps のペイロードの場合、最適な埋め込み強度設定レベル L_{SSL} は存在しなかった。BER が 10% 以下で、LSD の評価結果が 1 dB 以下、PESQ の評価結果が -2 ODG 以下を評価基準とすると、埋め込み強度設定レベル L_{SSL} が 0 dB のとき基準を満たす最適値とされた。このとき平均強度補正レベル L_{all} は -14.0 dB であった。

これらの結果より提案法では、埋め込み強度設定レベル L_{SSL} は 4 bps のペイロードにおいて、評価基準を満たしたを -10 dB とした。

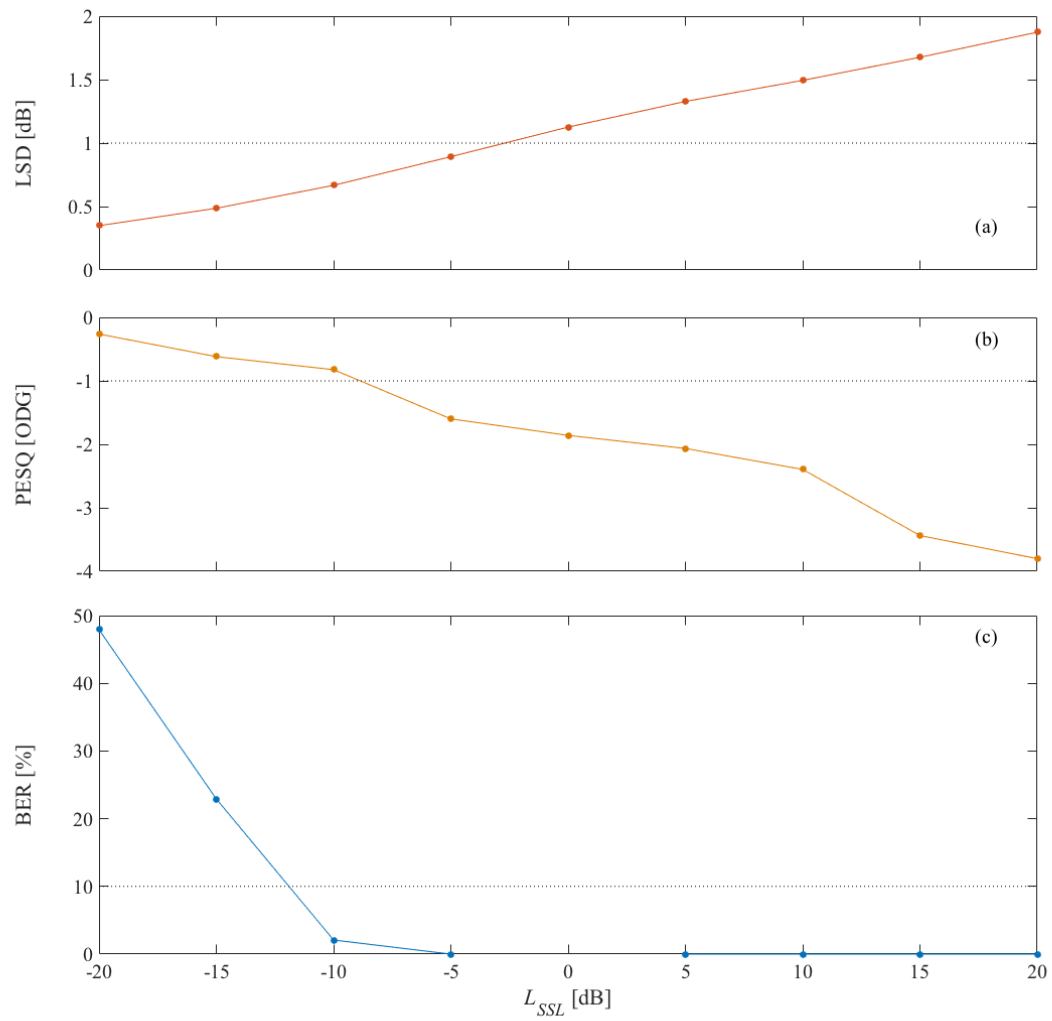


図 4.3: 埋め込み強度を変更した場合の検出誤差率と音質評価の結果 (4 bps) : (a) LSD による評価の結果, (b) PESQ による評価の結果, (c) 検出誤差率の結果.

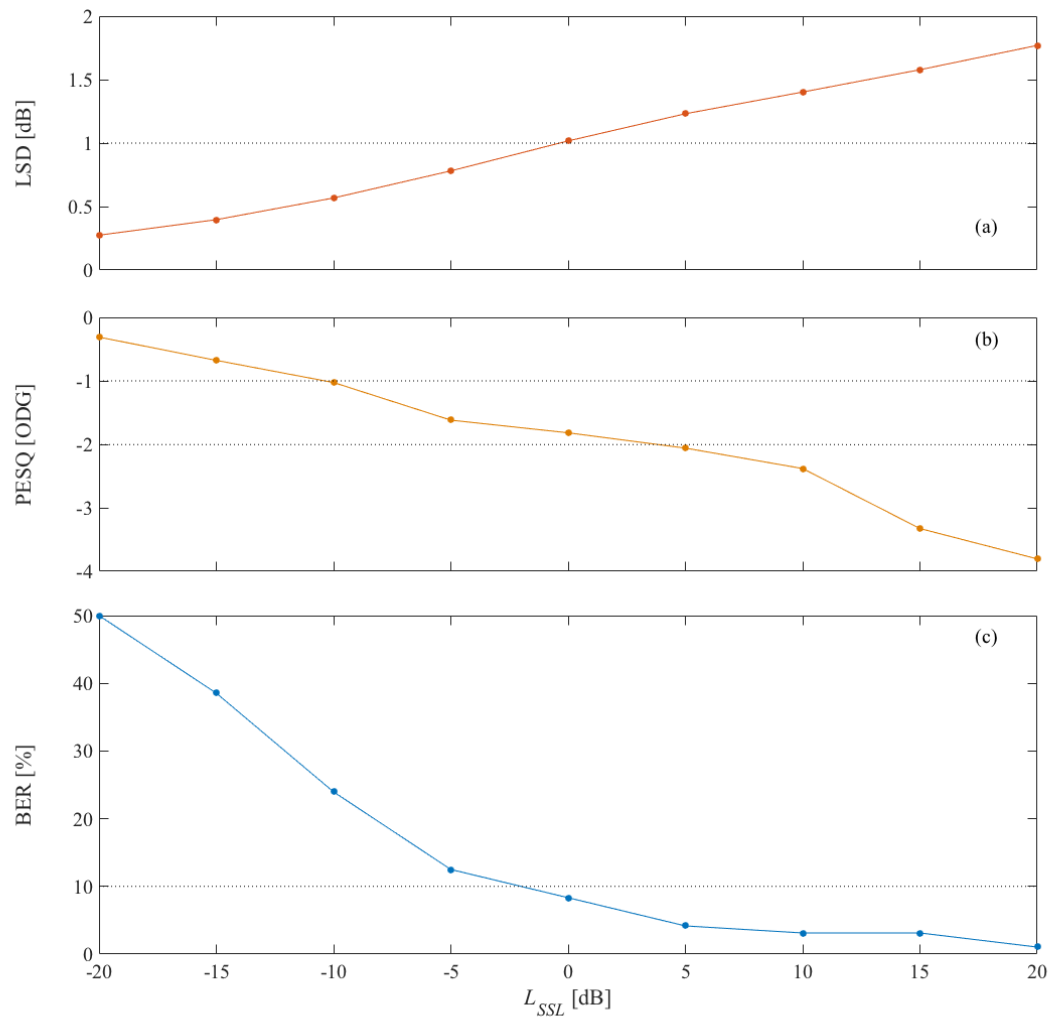


図 4.4: 埋め込み強度を変更した場合の検出誤差率と音質評価の結果 (8 bps) : (a) LSD による評価の結果, (b) PESQ による評価の結果, (c) 検出誤差率の結果.

4.3 情報検出

4.3.1 情報の検出アルゴリズム

本節では提案法における，透かしが埋め込まれた信号からメッセージを検出する情報検出のアルゴリズムについて述べる．図 4.5 に提案法によるメッセージ情報の検出のブロックダイアグラムを示す．

メッセージ情報の検出は LP 残差が式 (3.1) の性質を持つことを利用し，DSS 法における情報検出と同様に行う．まず情報が埋め込まれた信号 $y(n)$ をフレーム分割し，重み付けを実行する．その後，拡散信号として与えられるメッセージの埋め込みに利用した各フレームの LP 残差 $r(n)$ を，情報が埋め込まれた信号 $y(n)$ に乗ずる．乗算の結果の期待値を求めると，拡散信号の特徴（式 (3.1)）から，ホスト信号項は “0”，透かし信号項は $Er^2(n) = 1$ より $am(n)$ となり，メッセージ信号 $m(n)$ が算出される．得られたメッセージ信号 $m(n)$ をフーリエ変換し，信号の直流成分が正（0 以上）であれば “1”，信号の直流成分が負（0 以下）であれば “0” としてメッセージ m を検出する．提案法では，ノンブラインド検出を想定しているため，拡散信号となる LP 残差はメッセージの埋め込みに利用した LP 残差と同じ信号とした．

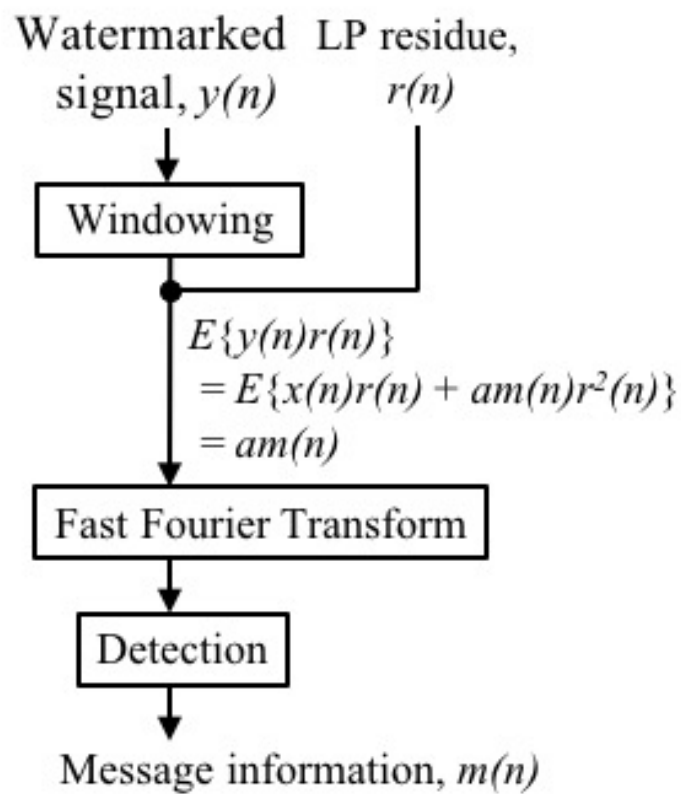


図 4.5: 提案法における情報検出のブロックダイアグラム.

4.3.2 フレーム同期検出

情報通信において、信号の送信者と受信者の持つ信号は常に同じ信号とは限らない。例えば、通信の過程における信号の欠落や、使用機器のサンプリングレートの差など様々な要因が考えられる。情報ハイディングにおいて、信号の同期は非常に重要であり、埋め込まれた信号のフレームと、検出における計算フレームが一致しない場合、埋め込まれたメッセージ情報を正確に検出することは困難となる。そこで情報が埋め込まれた信号のフレームと、検出に使用する拡散信号を生成するためのホスト信号のフレームを同期させることで、安定して正確な情報を検出するためのフレーム同期処理が必要となる。提案法では情報検出アルゴリズムの前処理としてフレーム同期処理を行い、安定した情報検出を行うこととした。

- フレーム同期法

図 4.6 に提案法における信号同期処理のブロックダイアグラムを、図 4.7 に提案法におけるフレーム同期処理のブロックダイアグラムを示す。提案法におけるフレーム同期処理は、まず透かしが埋め込まれた信号長が変更された信号とホスト信号の同期を行う。その後、ホスト信号から得られるフレーム位置に基づいて、信号長を再調整しフレーム同期を行う。

透かしが埋め込まれた信号長が変更された信号を 1 サンプル点ごとにシフトさせ、ホスト信号との相関を求めた。この相関が最大となる点が、埋め込まれた長さが変更された信号とホスト信号のサンプル点が一致した点であり、このサンプル点を信号同期位置とした。信号同期位置が定まった後、ホスト信号からフレーム位置を算出し、フレーム開始位置となるサンプル点で透かしが埋め込まれた信号の信号長を変更しフレーム同期を行う。本フレーム同期処理を施し、同期が取れた透かしが埋め込まれた信号とホスト信号から得られた LP 残差を使用して情報検出を行う。

図 4.8 に提案法におけるフレーム同期処理例を示す。ここで、図 4.8(a) はホスト信号、図 4.8(b) は信号長が変更され同期が取れていない透かしが埋め込まれた信号を示す。図 4.8(c) は図 4.8(b) の信号を 1 サンプル点ごとにずらし、図 4.8(a) の信号と相関を取った値である。図 4.8(d) は図 4.8(c) で得られた相関の最大値点を同期開始位置としたときの信号同期が行われた透かしが埋め込まれた信号を示す。図 4.8(e) は信号同期が行われた透かしが埋め込まれた信号を、ホスト信号から得られたフレーム位置に基づいてフレーム同期した信号を示す。

- フレーム同期法評価条件

提案法におけるフレーム同期検出について、信号同期位置のサンプル点ずれ、フレーム同期を行った信号の検出誤差率を用いて評価した。評価には ATR データベース (B) より 12 種類の音声に対して 4 bps のメッセージ情報を埋め込み、信号長を変更し、信号長が変更された透かしが埋め込まれた信号についてフレーム同期を行い、その後情報検出を行った。

- フレーム同期法評価結果

フレーム同期による信号長が変更された透かしが埋め込まれた信号のサンプル点は12種の音声で、一つの音声以外は100%の精度で信号長の変更位置を特定することができた。また、信号長の変更位置からずれがあった一つの信号においてそのサンプル点のずれは1サンプル点のずれであった。フレーム同期を行った透かしが埋め込まれた信号からの情報検出の検出誤差率は、11種の音声で100%の検出率であり、12種の音声においてその検出誤差率は2.01%であった。

以上の結果より、信号に攻撃が加えられていない4 bpsのペイロードの場合、提案法のフレーム同期処理を行うことで、高い精度でホスト信号と信号長が変更された透かしが埋め込まれた信号のフレーム同期を行うことが可能であることが示された。

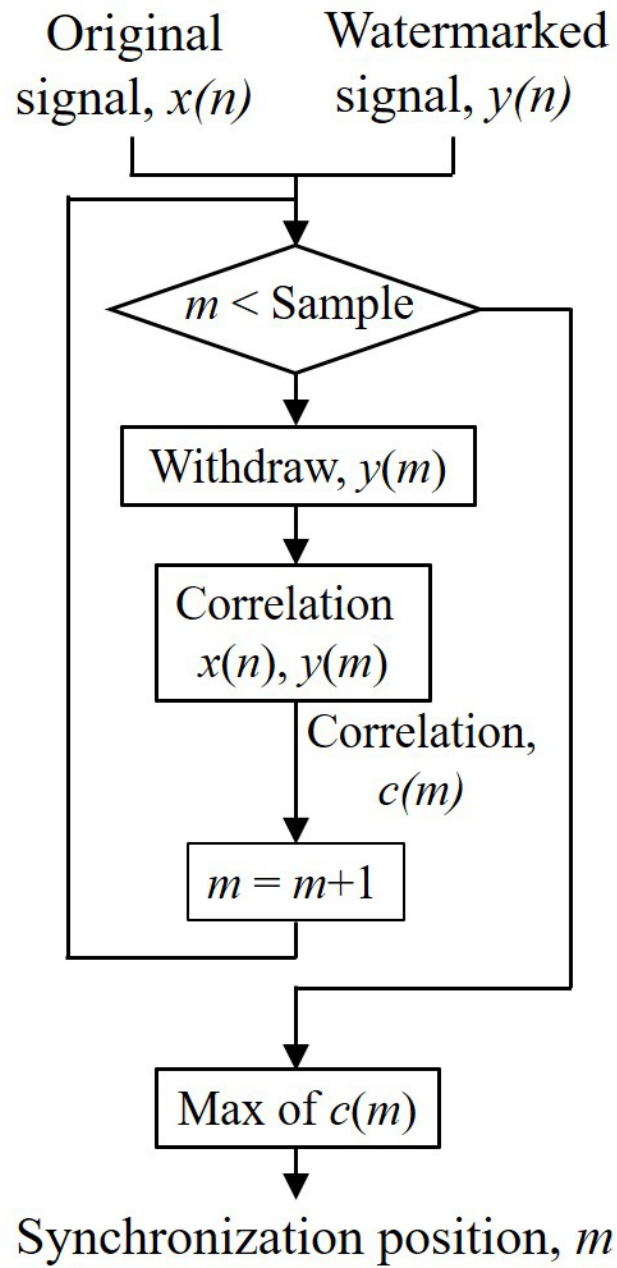


図 4.6: 提案法における信号同期法のブロックダイアグラム.

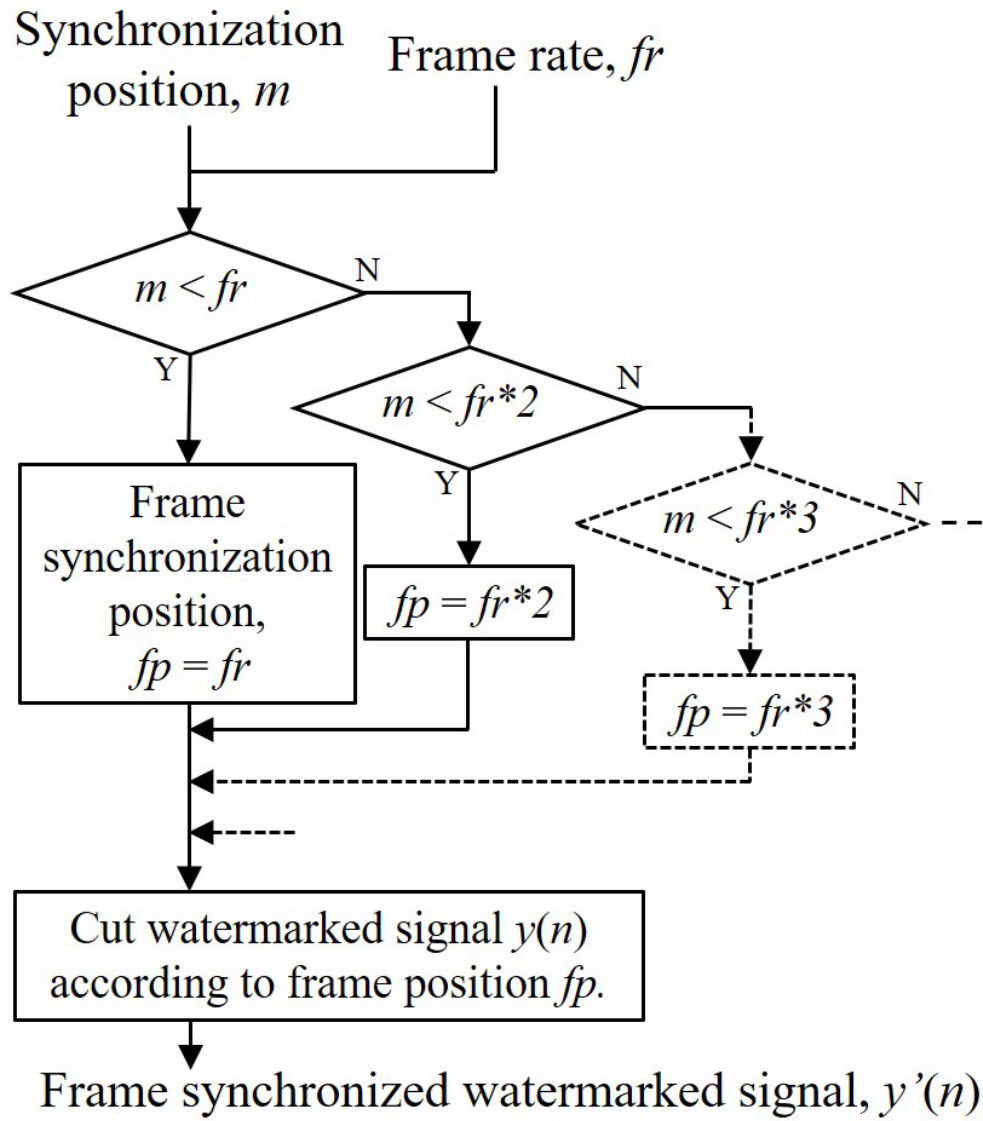


図 4.7: 提案法におけるフレーム同期法のブロックダイアグラム.

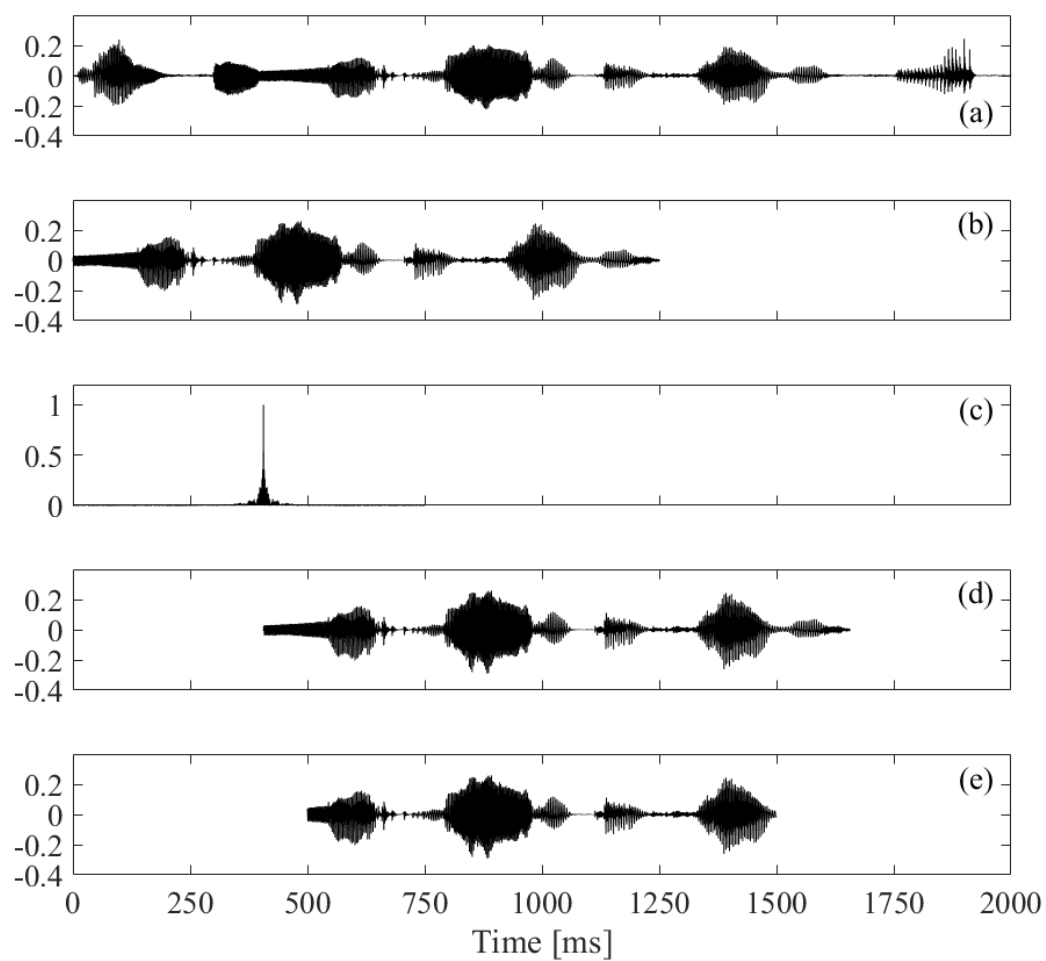


図 4.8: 提案法におけるフレーム同期法処理例：(a) ホスト信号, (b) 信号長が変更された透かしが埋め込まれた信号, (c) フレームシフトしたときの相関値, (d) 信号同期を行った透かしが埋め込まれた信号, (e) フレーム同期を行った透かしが埋め込まれた信号。

第5章 提案法の評価

本章では提案法による音声情報ハイディングが知覚不可能性と頑健性を満たした手法であるかを評価した。

5.1 評価尺度

- 知覚不可能性

提案法において聴取者が透かしが埋め込まれた信号を聴取した際に、透かし信号の存在に気づくか気づかないかは、透かしが埋め込まれた信号の歪みの大きさや、その音質によって左右される。本研究では、提案法による音声情報ハイディングの知覚不可能性の評価のために透かしが埋め込まれた信号の音質を評価した [48] [49]。透かしが埋め込まれた信号の音質評価には LSD と PESQ を用いた。LSD は信号の対数スペクトル歪みを評価し、その評価結果が 1 dB 以下のときに高い音質とされている。PESQ は主観評価に基づく客観評価結果を出力し、Objective difference grades (ODGs) によってその評価結果が示される。ODGs の出力結果は表 (4.1) に示すように定められ、その評価結果が 0 から -1 のとき高い音質とされる。

- 頑健性

音声情報ハイディング法における頑健性は、信号処理や外乱が加えられた透かしが埋め込まれた信号からの情報検出精度によって評価される。本研究では、透かしが埋め込まれた信号に信号処理、符号化を施し、処理が施された透かしが埋め込まれた信号から情報検出を行う。この時、メッセージ情報の BER によって頑健性を評価した。検出精度は、BER が 10% 以下のとき検出精度に問題がないとし、処理が施された透かしが埋め込まれた信号から BER 10% 以下で情報検出ができた時、提案法は施された処理に対して耐性を持つ情報ハイディング法であるとした [50]。

- 埋め込み容量

本稿で提案する音声情報ハイディング法では知覚不可能性と頑健性を持つかを評価対象とするが、情報ハイディングでは目的に合わせた埋め込み容量（ペイロード）の情報を埋め込めることが求められる。本研究で提案した音声情報ハイディング法においても、埋め込

み容量の基準は必要であり，基準となる埋め込み容量において，上記の知覚不可能と頑健性を満たさなければならない．

しかし音声情報ハイディングにおいて必要とされる埋め込み情報量は定義されていない．本評価では IHC の評価基準に基づき，6 bps の埋め込み情報量を音声情報ハイディングの評価基準とした [51, 52]．これより，6 bps の埋め込み容量において，知覚不可能，頑健性が満たされる場合，提案法が音声情報ハイディング法として有用であるとされる．

5.2 評価条件

提案法における知覚不可能性の評価では，ホスト信号と透かしが埋め込まれた信号に対して，LSD と PESQ を用いて音質の客観評価を行った．頑健性については提案法によって透かしが埋め込まれた信号に対し，信号処理や符号化が施された場合でも，10%以下の BER で埋め込み情報を検出できるか，施された処理に対する耐性評価を行った．本評価では信号処理として，透かしが埋め込まれた信号に表 5.1 に示される基本的な信号処理，符号化を施し，再度情報検出における BER を求めた．これらの処理は IHC などに定められた AIH の基本頑健性テストの評価項目の一部に，音声符号化を追加したものである．

提案法の評価では，ATR データベース (B) より 12 種類の音声に対して，音声の発話が見られる 1 秒間に情報の埋め込み処理を行った [53]．それぞれの評価における比較対象として，DSS 法を用いて同値のメッセージを埋め込んだ信号に対しても評価を行った．

表 5.1: 耐性評価に利用した処理の一覧．

処理項目
ダウンサンプリング (44.1 kHz から 16 kHz)
雑音加算 (signal-to-noise ratio, SNR 36 dB)
振幅圧縮 (16 bits 量子化から 8 bits 量子化)
音声符号化 (G.711)
音声符号化 (G.726)
音声符号化 (G.729)

5.3 評価結果

5.3.1 知覚不可能性評価の結果

提案法の知覚不可能を評価するために行った，LSD と PESQ による透かしが埋め込まれた信号の音質評価の結果を図 5.1 と図 5.2 に示す．ここで図 5.1 は LSD による評価結果を示し，図 5.2 は PESQ による評価結果を示す．図の横軸はペイロードを示し，縦軸は各評価値を示す．

LSD による評価では，提案法は DSS 法と比較して 2 dB 程度の歪み量の減少がみられ，全てのペイロードにおいて基準の 1 dB 以下の歪みであった．これより，提案法によって情報が埋め込まれた信号の音質が，“良い”音質であることが示された．

PESQ による音質評価では，DSS 法による評価結果が 4 から 256 bps のペイロードにおいて，1.8 から 2.3 の間に出力されているのに対して，提案法の評価結果では，2.8 から 4.5 と大きく改善された評価結果が示された．これより，PESQ による音質評価でも“良い”音質であると評価された．

LSD と PESQ のどちらの音質評価でも 64 bit までのペイロードで，ペイロードが増加すると音質評価結果が低下する傾向が見られた．音声電子透かしで必要とした 6 bps のペイロードにおいて，LSD と PESQ の両方の評価で基準値を満たしており，音声電子透かしとして有用であることが確認された．

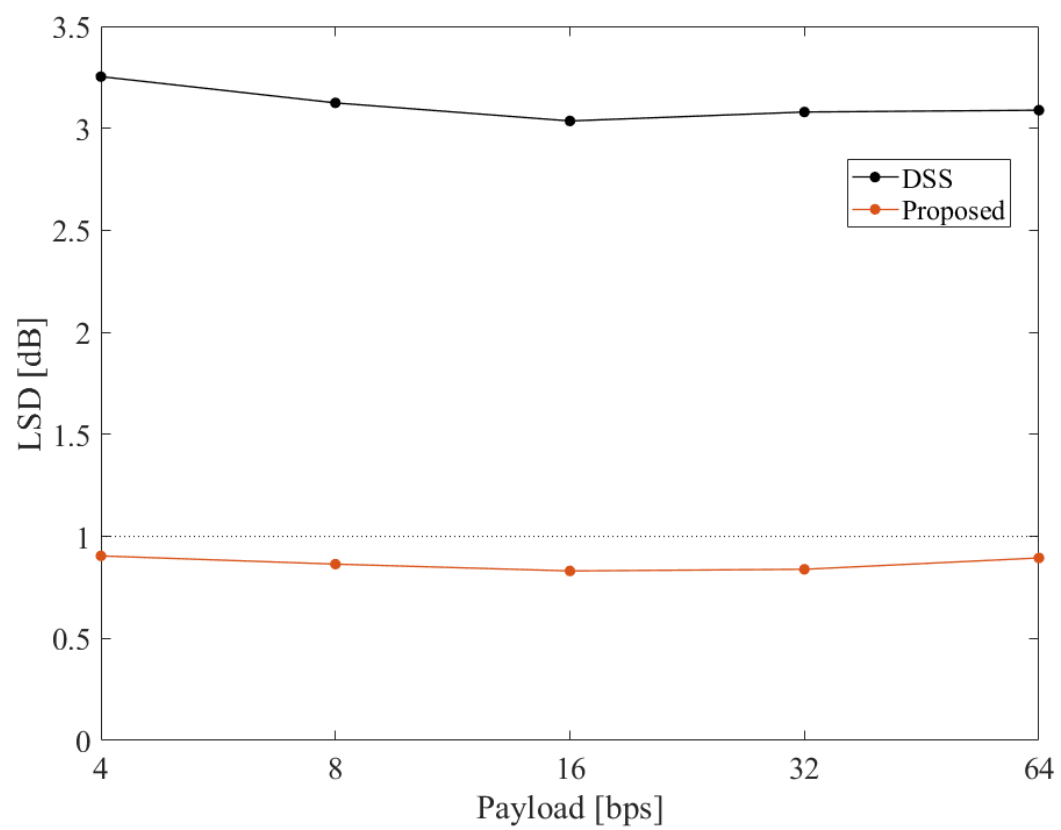


図 5.1: LSD による音質評価の結果.

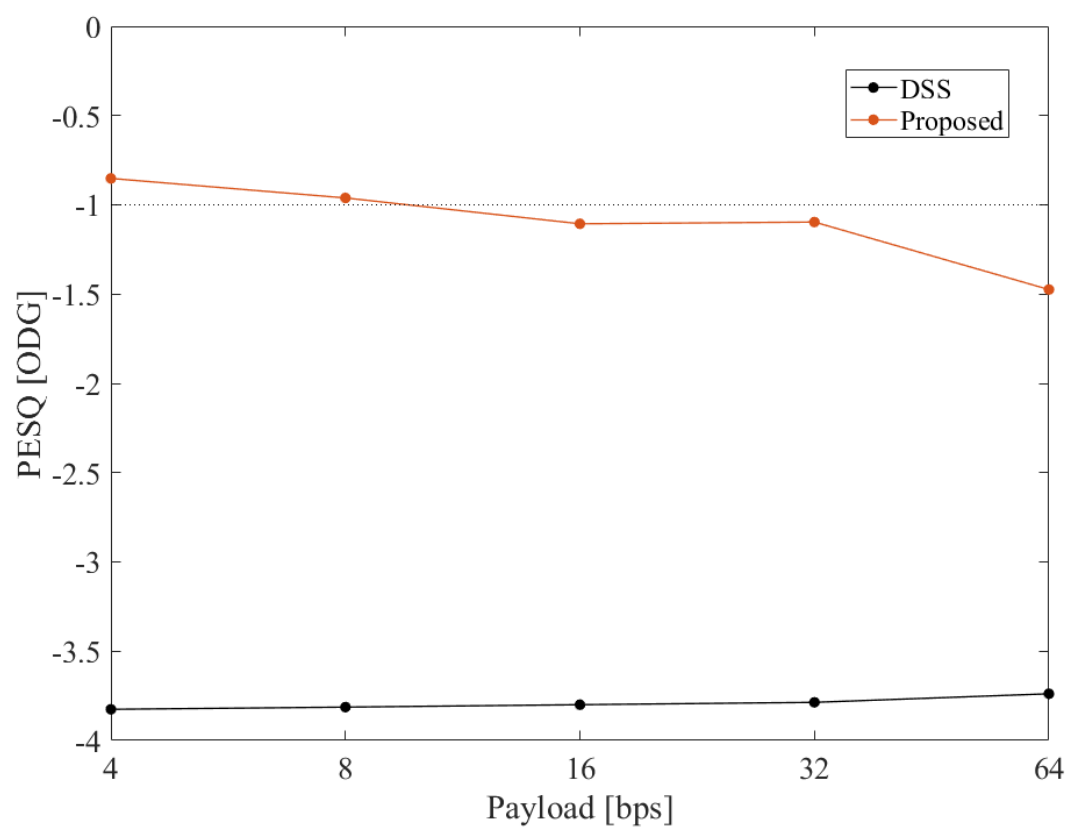


図 5.2: PESQ による音質評価の結果.

5.3.2 頑健性評価の結果

提案法の頑健性を評価するために行った，基本的な信号処理と符号化に対する耐性評価の結果を図 5.3 から図 5.9 に示す．ここで図 5.3 は信号処理や符号化を施していないときの BER を示す．図 5.4 は 44.1 kHz から 16 kHz へのダウンサンプリングしたときの検出結果を示す．図 5.5 は白色雑音を SNR 36 dB となるように加算したときの BER を示す．図 5.6 は振幅を 16 bits から 8 bits へ圧縮したときの BER を示す．図 5.7 は G.711 による音声符号化処理を施したときの BER を示す．図 5.8 は G.726 による音声符号化処理を施したときの BER を示す．図 5.9 は G.729 による音声符号化処理を施したときの BER を示す．図の横軸はペイロードを示し，縦軸は BER を示す．

今回行った全ての耐性評価において，提案法と DSS 法のどちらの方法を用いた場合でも，ペイロードが大きくなるにつれて BER は大きくなる傾向が見られた．このとき提案法を用いた場合は，DSS 法よりも低いペイロードで BER が上昇する傾向がみられた．しかし音声電子透かしの評価基準とした 6 bps のペイロードに対しての BER は，G.726，G.729 音声符号化耐性評価を除く評価において 10% 未満の BER であることが示された．提案法で耐性を持たないとされる G.726，G.729 音声符号化は，DSS 法でも BER が高く提案法は DSS 法と同程度の頑健性を持つことが示された．

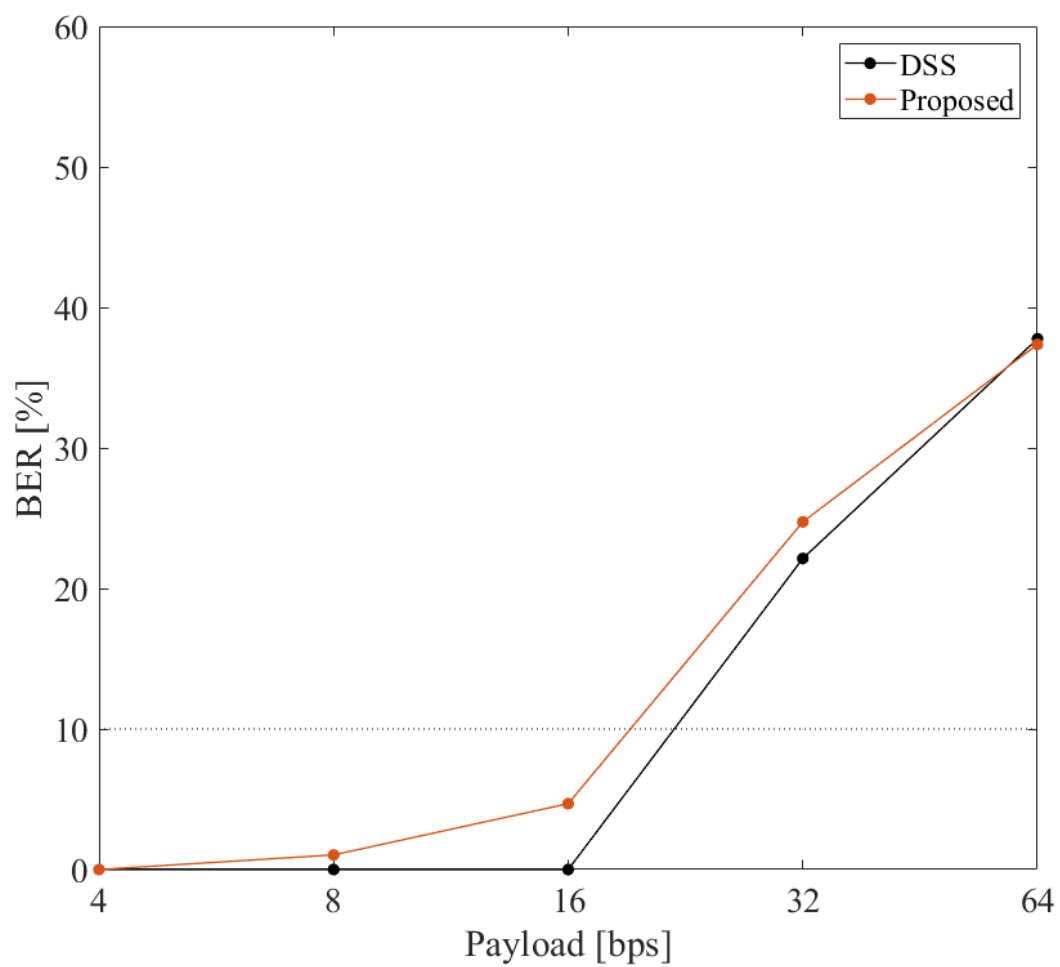


図 5.3: 信号処理や符号化を施していないときの BER.

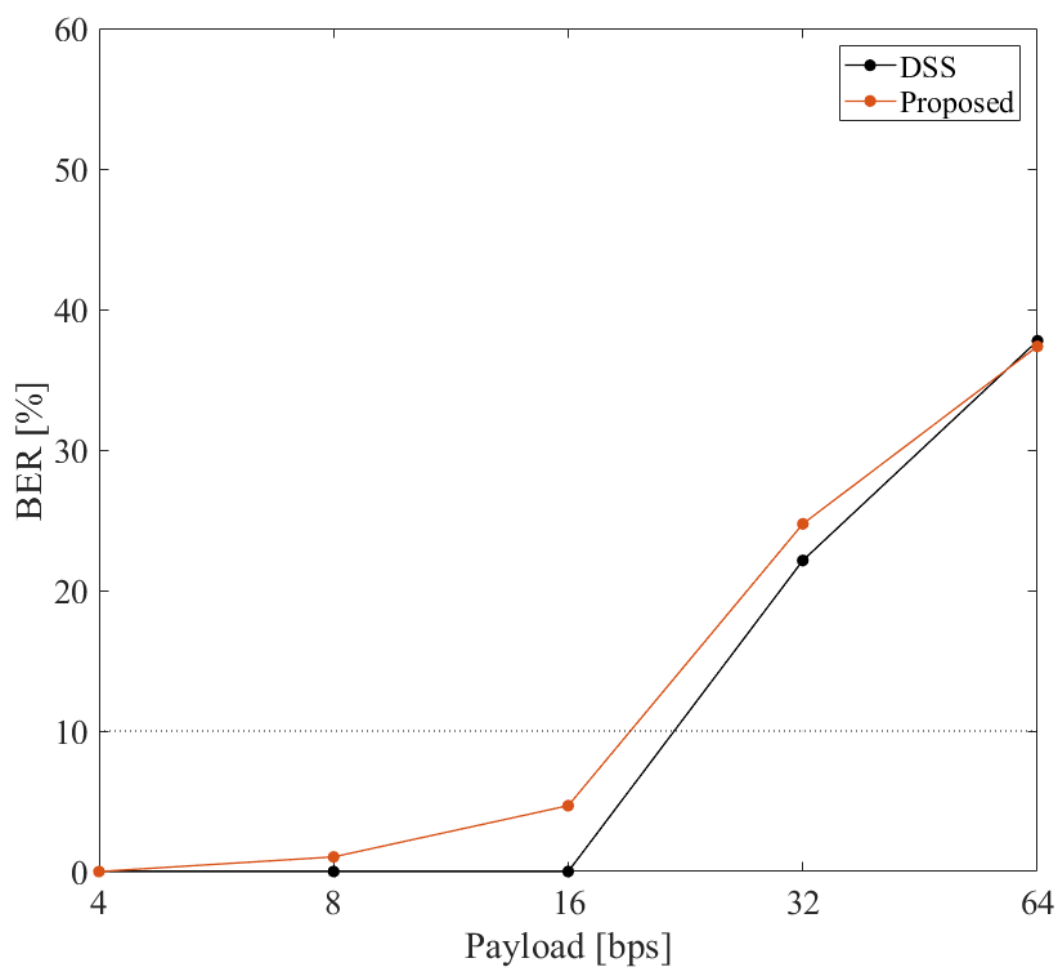


図 5.4: ダウンサンプリング (44.1 kHz - 16 kHz) に対する耐性評価の結果.

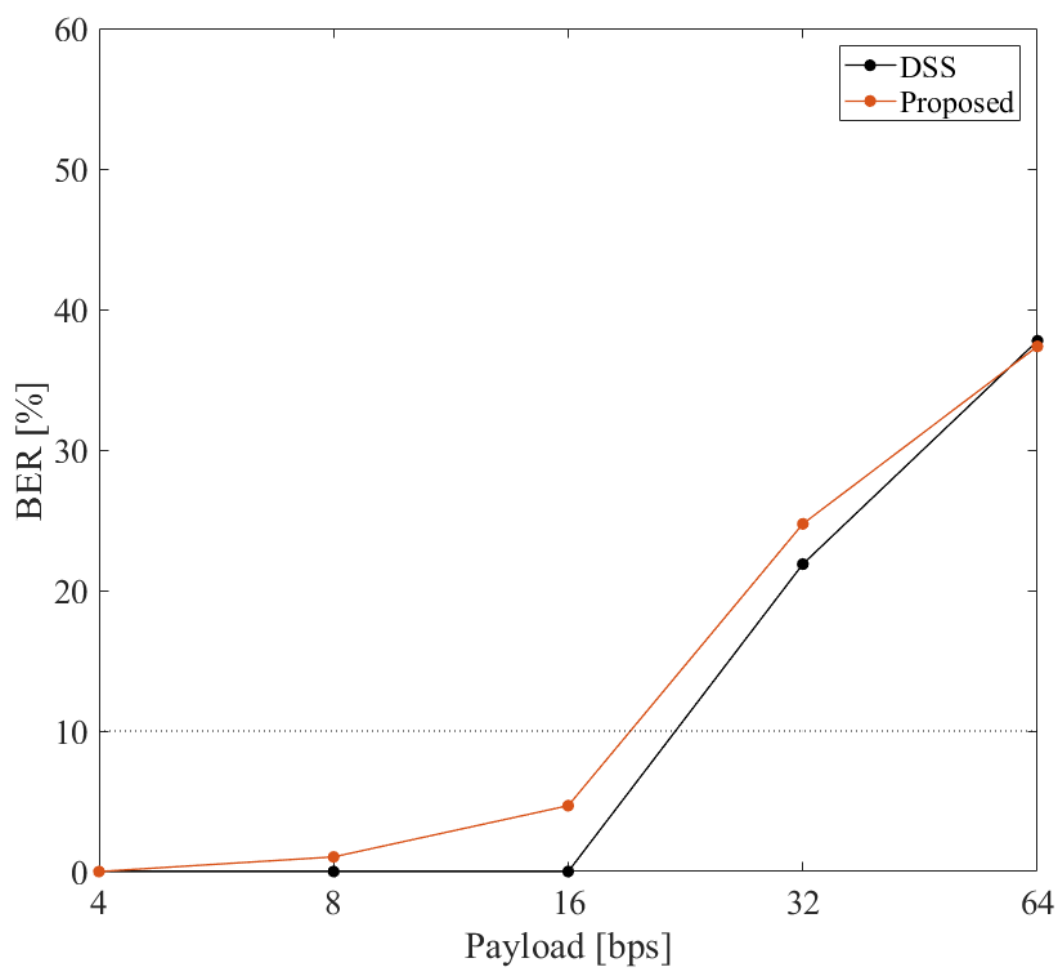


図 5.5: 雑音付与 (SNR 36 dB) に対する耐性評価の結果.

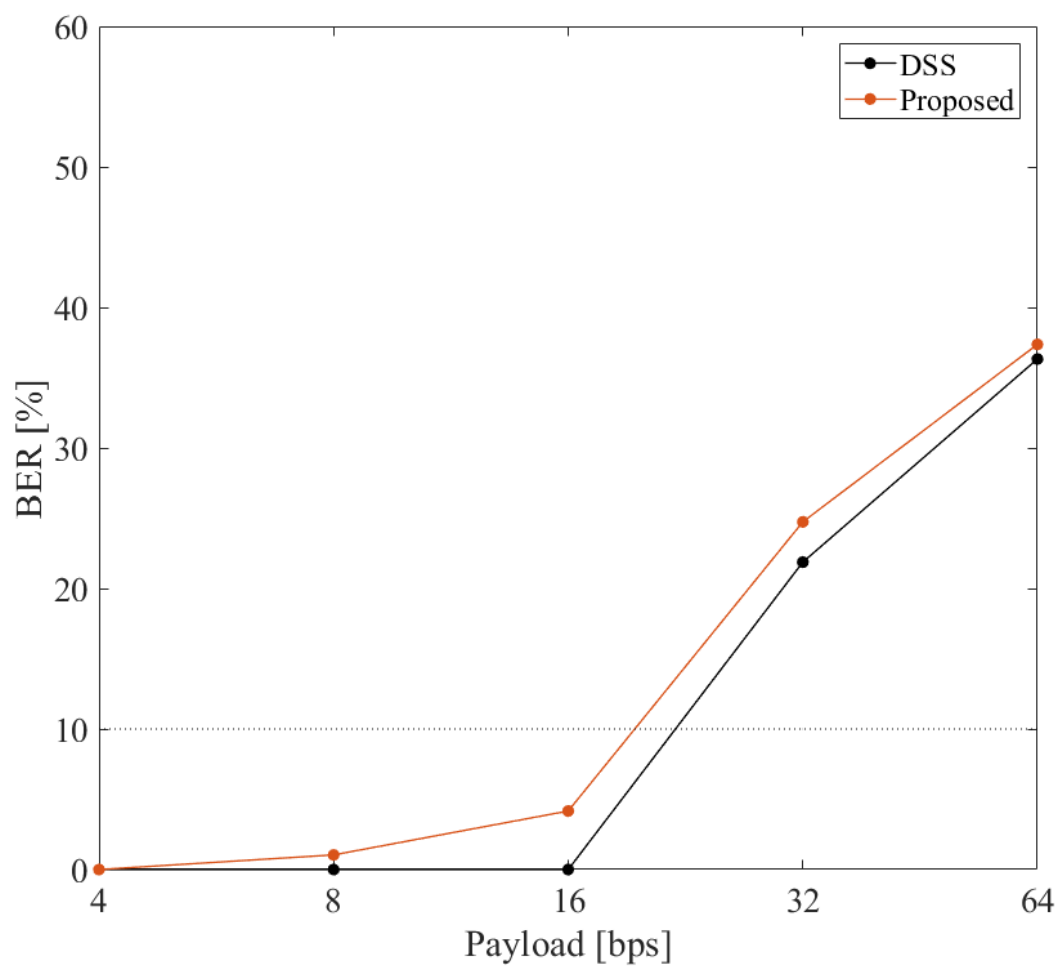


図 5.6: 振幅圧縮 (16 bit - 8 bit) に対する耐性評価の結果.

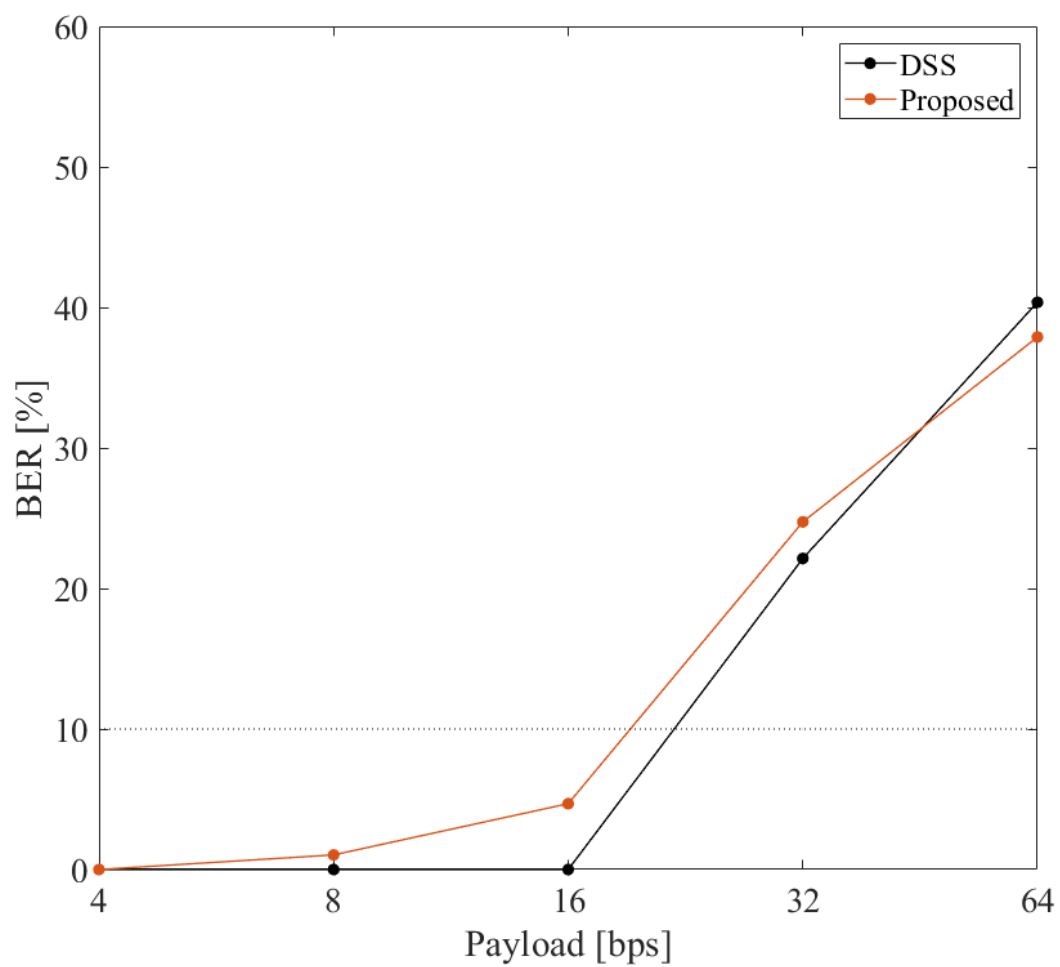


図 5.7: G.711 符号化に対する耐性評価の結果.

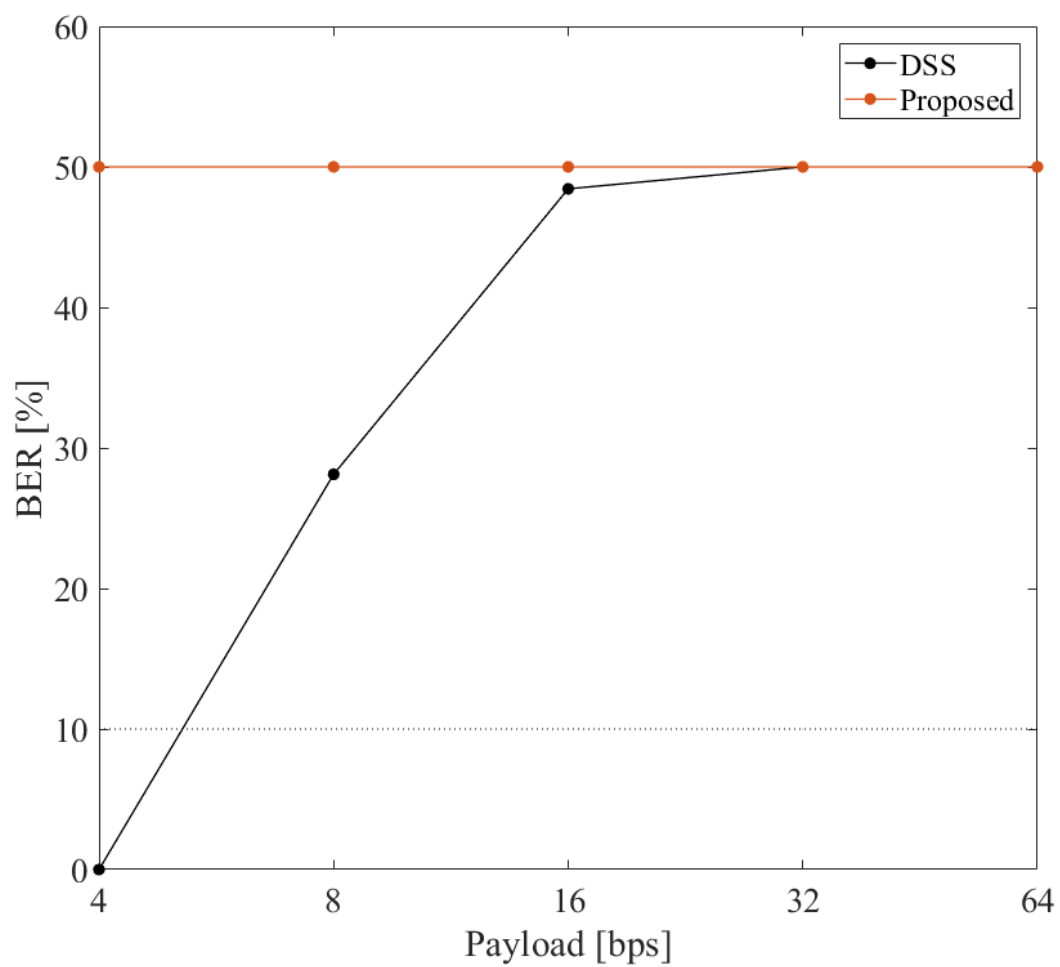


図 5.8: G.726 符号化に対する耐性評価の結果.

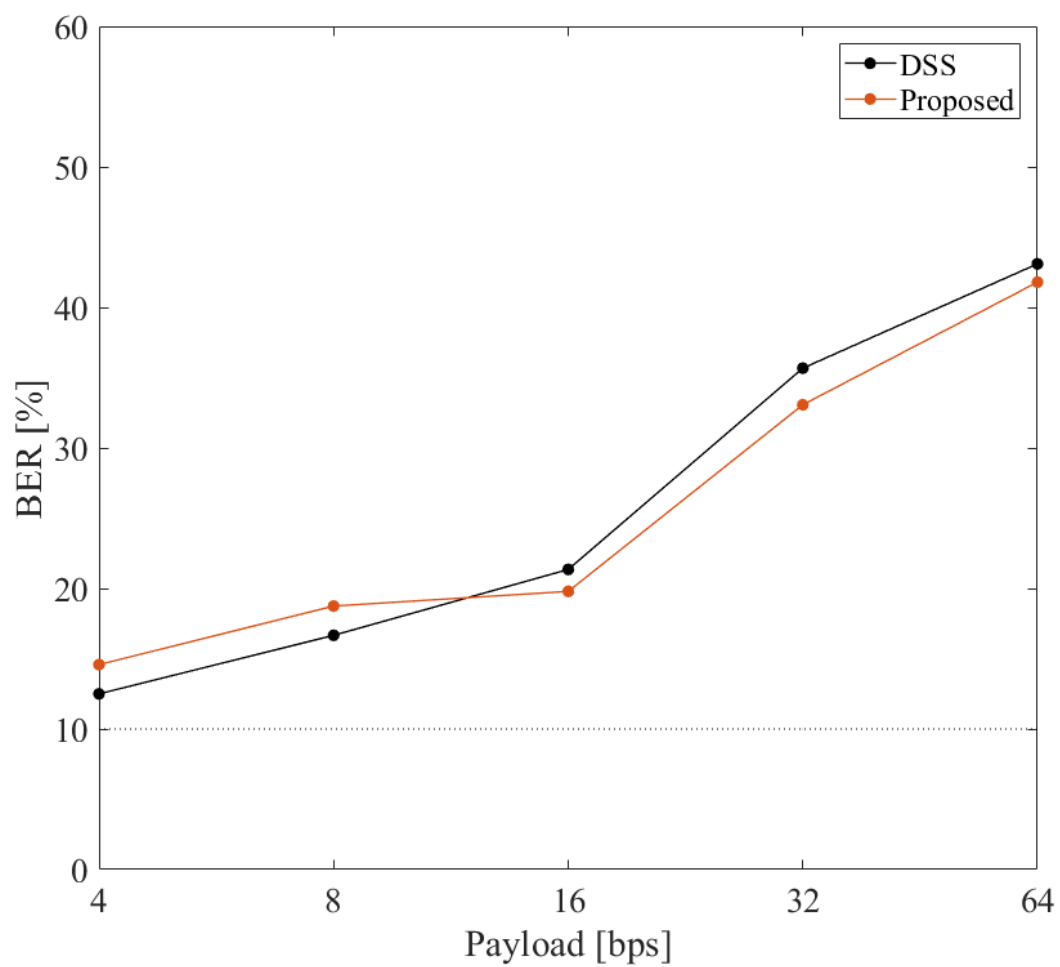


図 5.9: G.729 符号化に対する耐性評価の結果.

5.4 考察

提案法について知覚不可能性と頑健性の両方を満たした音声電子透かし法となるか検討した。

情報が埋め込まれた信号の音質を LSD と PESQ で評価した結果、提案法によって透かしが埋め込まれた信号の音質評価結果は、LSD による評価で、2 dB 程度の歪み量の減少がみられ、基準の 1 dB 以下であった。PESQ による評価で、評価結果が、2.8 から 4.5 と大きく改善されていた。これらより、提案法が DSS 法より非常に良い音質で情報の埋め込みが可能であり、高い知覚不可能を持つことが確認された。提案法を用いた情報ハイディングで、透かしが埋め込まれた信号の音質が上がった理由として、提案法における拡散信号にホスト信号の音源成分で、ホスト信号と類似したスペクトル形状を持つ LP 残差を利用したことが、大きな要因であると考えられる。拡散信号に LP 残差を用いたことで、透かし信号のスペクトル形状とホスト信号のスペクトル形状が類似するため、透かしが埋め込まれた信号とホスト信号との歪みが減少し、LSD による評価結果が向上したと考えられる。また、透かし信号のスペクトル形状とホスト信号のスペクトル形状が類似することから、ホスト信号と透かし信号は、知覚的融合によって一つの合成音として知覚されるため、主観評価結果に基づく客観評価方である、PESQ による評価結果が向上したと考えられる。

提案法の耐性評価を行った結果、提案法は DSS 法と同等の耐性を持つことが確認された。これは提案法がスペクトル拡散という、DSS 法の原理に基づいて提案されていることが大きな要因であると考えられる。また、提案法が G.726, G.729 の符号化に耐性を持たない要因として以下が考察される。G.726 では適応的差分パルス符号変調を用いており、差分パルス符号化変調 (PCM) における量子化の幅が適宜変化している [54]。そのため、符号化が施された信号の波形は、符号化前の原信号と比較して荒い波形となる。これにより、符号化処理が施された透かしが埋め込まれた信号 $y'(n)$ における透かし信号 $w'(n)$ と、検出に使用する拡散信号 $r(n)$ において、式 (3.1) の特徴が担保できず、情報検出における BER が高くなったと考えられる。G.729 では、Algebraic Code Excited Linear Prediction (ACELP) を用いて符号化処理を行っている [55]。ACELP などの CELP 系の符号化では、音声を線形予測フィルターに基づいた信号として符号化する。そのため、伝達された音声信号に位相情報は含まれておらず、LP 残差などの情報は基本周波数の平均としてスカラー量が送信されることとなる。これにより、G.729 によって符号化処理された透かしが埋め込まれた信号 $y'(n)$ に含まれる LP 残差 $r'(n)$ は、情報の埋め込みに使用した LP 残差 $r(n)$ と位相などが異なる信号となるため、式 (3.1) の特徴が担保できなくなる。この結果、情報検出における BER が高くなったと考えられる。上記より本提案法は符号化などの処理により、透かしが埋め込まれた信号に含まれる LP 残差 $r(n)$ の情報に改変が加えられ、式 (3.1) の特徴が担保できなくなる処理に対して耐性を持たないことが考えられる。

以上より、提案法が DSS 法よりも高い知覚不可能性と DSS 法と同程度の頑健性を持ち、両方の要求項目を満たした音声電子透かし法であることが確認されたと判断できる。

第6章 結論

6.1 本研究で明らかになったこと

情報通信技術の発達に伴って、音声情報通信が日常的に使用されるようになった。音声情報通信の利便性が向上した傍ら、音声分析合成技術の発達により、音声情報通信における情報改ざんや通信のなりすましといった問題が危険視されるようになった。この問題の解決策として、コンテンツに付随する情報や伝達すべき情報を、コンテンツ自体に不可分な形で付与する技術である情報ハイディングに注目が集まっている。現在までに音響信号を対象に情報を埋め込み付与する技術として、音響情報ハイディング法（AIH）が数多く研究、提案されている。しかし、AIHはその対象を音声とした場合、音声特有の符号化や音質劣化といった問題が残ってしまう。そのため、本研究では知覚不可能性と頑健性を持つ、音声情報ハイディング法を提案した。

提案法は頑健な AIH である直接スペクトル拡散（DSS）法を元に提案されており、拡散信号に線形予測（LP）残差を利用した。提案法は、LP 差がホスト信号と同じスペクトル帯域を持つことから、ホスト信号に透かし信号を足し合わせた際、でも聴取者が透かしが埋め込まれた音声の歪みを認識することが困難となり知覚不可能性を持つ音声情報ハイディング法となることが考えられた。また DSS 法が高い頑健性を持つため、DSS 法の原理を利用した提案法も同様に高い頑健性を持つことが考えられた。

提案法が知覚不可能性と頑健性の両方の要求項目を満たした音声情報ハイディング法であるか、それぞれについて評価を行った。知覚不可能性については透かしが埋め込まれた信号の音質を LSD と PESQ によって評価した。頑健性については透かしが埋め込まれた信号に信号処理や符号化を施し、基準値以内で情報検出が可能であるかによって評価した。知覚不可能性では透かしが埋め込まれた信号の音質評価により、提案法によって透かしが埋め込まれた信号の音質は既存法（DSS 法）よりも非常に高い音質であることが示された。LSD による評価では約 2 dB 歪みが減少し、PESQ による評価では 3 ODG 程度評価が向上した。頑健性では信号処理や符号化が施された透かしが埋め込まれた信号の情報検出精度の結果より、提案法が DSS 法と同程度の頑健性を有することが示された。評価結果より、提案法は DSS 法よりも非常に高い音質で透かしを音声に埋め込む事が可能であり、DSS 法と同程度の頑健性を持つ音声情報ハイディング法であることが示された。

以上より、提案法は知覚不可能性と頑健性を有する音声情報ハイディング法であるとした。

6.2 残された課題

本稿で提案された音声情報ハイディング法における残された課題として、「頑健性の向上と改ざんへの脆弱化」,「ブラインド検出法への改良」の2項が挙げられる。

- 頑健性の向上と改ざんへの脆弱化

本稿では耐性評価の項目の一つとして音声符号化 (G.726, G.729) を用いたが, これらの符号化が行われた信号から提案法では頑健にメッセージ情報を検出することが困難であった。そのため, これらの符号化処理を施した信号からも頑健に情報が検出できるよう改良することが求められる。また, 本稿の評価で用いた符号化以外の符号化に対しても同様に高い頑健性を持つような改良されることが求められる。

本稿5章でも述べたように, 提案法は符号化などの処理につて, 透かしが埋め込まれた信号に含まれる LP 残差 $r(n)$ の情報に改変が加えられた場合, 式 (3.1) の特徴が担保できなくなり, 処理に対して耐性を持たないことが考えられた。そのため, 符号化処理などを施した後でも, 透かしが埋め込まれた信号に含まれる LP 残差 $r(n)$ の情報が正確に含まれていることが求められる。符号化処理などを施した後でも, 透かしが埋め込まれた信号に含まれる LP 残差 $r(n)$ の情報が正確に含まれるために, 透かしの埋め込み強度を, 符号化に耐性を持つ程度に選定し直すことなどをして, 処理が施された透かし信号に含まれる LP 残差 $r'(n)$ と, 検出に利用する LP 残差 $r(n)$ が, 式 (3.1) の特徴を維持するような改良を加えることが求められる。

本稿では信号処理や符号化などの基本的な信号処理として利用される処理に耐性を持つかを評価した。しかし, 提案法を実際の通信上で利用することを想定する場合, 信号に対して透かし信号の検出を阻害する攻撃を加えられることが考えられ, これに対する耐性を持つことが必要となる。これらの攻撃の例には, 非同期化, サンプル置換, A-D・D-A 変換, 結託攻撃などが挙げられている [56]。提案法を音声通信などの情報通信技術に組み込んで使用することを想定するのであれば, これらの攻撃に対しても, 高い耐性を持つ改良を加えることが必要となる。

しかしながら, 提案法を音声の改ざんやなりすましの対策として, 音声通信の真正性情報を付与するために利用することを想定した場合, 全ての攻撃に耐性を持つことは好ましくない。提案法が全ての攻撃に耐性を持つことは, 改ざんやなりすましが行われた信号からも, 情報の検出ができることを意味し, 通信の安全保証には利用することができなくなる。そのため, 基本的な信号処理や符号化, 通信攻撃に対しては高い耐性を持ち, 改ざんやなりすましといった処理に弱い耐性を持つような改良を加えることが求められる。

- ブラインド検出法への改良

提案法における情報検出では、信号の受け取り手がホスト信号を保有していることを仮定した、ノンブラインド情報検出であった。しかし情報通信の過程において信号の受け取り手が原信号を有していない場合も想定される。そのため、情報検出において透かしが埋め込まれた信号のみを用いて、埋め込まれたメッセージ情報が検出可能となる、ブラインド型の情報検出の改良が課題の一つとなる。ブラインド検出が可能となることで、提案法は音声通信における真正性の保護付与だけでなく、広い範囲でその利用が検討されることになる。

この課題の解決案として、Robust principal component analysis (RPCA) などの技術を用いて、透かしが埋め込まれた信号からホスト信号と透かし信号を分離し、分離されたホスト信号から、情報の検出に利用する LP 残差を再度算出する手法などが検討される [57]。我々の検討では、RPCA による音源分離で、ホスト信号と分離された音声信号の形状差が大きかったため、LP 残差形状も大きく異なるものとなり、ブラインド検出が困難であった。RPCA による音源分離のパラメータ等を適切に設定することで、ホスト信号と分離された音声信号は類似した信号となり、より信号形状の類似した LP 残差が算出可能であると考えられ、提案法におけるブラインド検出が可能であると考えられる。

参考文献

- [1] 山田 尚志, 石原 淳, 加藤 拓, “マルチメディア時代のコピープロテクション,” 電子情報通信学会論文誌, Vol. J87-A, No. 6, pp. 734–745, 2004.
- [2] 河原 英紀, “聴覚の情景分析が生んだ高品質 VOCODER: STRAIGHT,” 日本音響学会誌, vol. 54, no 7, pp. 521–526, 1998.
- [3] Morise Masanori, Yokomori Fumiya, and Ozawa Kenji, “WORLD: a vocoder-based high-quality speech synthesis system for real-time applications,” IEICE transactions on information and systems, vol. E99-D, no. 7, pp. 1877–1884, 2016.
- [4] 鷗木 祐史, 西村 竜一, 西村 明, 近藤 和弘, 近藤 和弘, 藺田 光太郎, “音響情報ハイディング技術 (音響テクノロジーシリーズ 20),” コロナ社, 2018.
- [5] 岩切 宗利, 松井 甲子雄, “音声・音楽を用いたインフォメーションハイディング,” 情報処理学会誌, vol. 44, no. 3, pp. 242–247, 2003.
- [6] 松井 甲子雄, “電子透かしの基礎,” 森北出版, 1998.
- [7] DRM.info, Drupal, <https://drm.info/>, (Retrieved February 2019).
- [8] 半田 浩規, 西村 竜一, 棟方 哲弥, 鈴木 陽一, “音データハイディングによるユニバーサル情報提示システムの実現の試み,” 第二回マルチメディア情報ハイディング研究会資料, pp. 33–37, November, 2007.
- [9] Nishimura Akira, “Audio data hiding that is robust with respect to aerial transmission speech codecs,” Int. J. Innov. Comput. Inf. Control, vol. 6, no. 3(B), pp. 1389–1400, 2010.
- [10] 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.
- [11] 矢島 尚幸, 大石 邦夫, “振幅及び位相特性に基づく音楽信号への電子透かし,” 電子情報通信学会技術研究報告.SIP, Vol. 104, pp. 53–58, 2005.
- [12] 西村 明, “音響信号への情報秘匿技術,” 日本音響学会誌, vol. 63, no.11, pp.660–667, 2007.

- [13] Nedeljko Cvejić, and Tapio Seppänen, “Digital audio watermarking techniques and technologies,” IGI Global, Hershey, 2007.
- [14] Nedeljko Cvejić, and Tapio Seppänen, “Increasing Robustness of LSB Audio Steganography Using a Novel Embedding Method,” Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC ’04), pp. 533–537, 2004.
- [15] Daniel Gruhl, Anthony Lu, and Walter Bender, “Echo Hiding,” Proc. Information Hiding 1st Workshop, pp. 295–315, Cambridge Univ., 1996.
- [16] 高 秉燮, 西村 竜一, 鈴木 陽一, “PN 系列を用いたエコー拡散透かし手法の提案,” 平成 14 年度日本音響学会春季研究発表会講演論文集, pp. 535–536, March, 2002.
- [17] 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,” 電子情報通信学会技術研究報告.CSEC, Vol. 102, No. 211, pp. 125–130, July, 2002.
- [18] 赤木 正人, “位相と知覚 -人間は果たして位相響か?-,” 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, September, 1997.
- [19] 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,” 電子情報通信学会技術研究報告.EA, Vol. 98, pp. 15–22, 1998.
- [20] 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, March, 2001.
- [21] 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.EA, Vol. 106, No. 205, pp. 7–11, 2006.
- [22] 西村 竜一, 鈴木 陽一, “周期的位相変に基づく音響電子透かし,” 日本音響学会誌, Vol. 60, No. 5, pp. 269–272, 2004.
- [23] 西村 竜一, 藺田 光太郎, 鈴木 陽一, “標本値の微分に基づく周期的位相変調電子透かしのブラインド検出手法,” 平成 17 年度日本音響学会春季研究発表会講演論文集, pp. 479–480, March, 2005.
- [24] 藺田 光太郎, 西村 竜一, 鈴木 陽一, “周期的位相回転による音楽信号用電子透かしのブラインド検出手法に関する検討,” 平成 14 年度日本音響学会春季研究発表会講演論文集, pp. 533–534, 2002.
- [25] 饗庭 絵里子, 津崎 実, “2 音の同時性判断と蝸牛遅延の関連について,” 平成 18 年度日本音響学会秋季研究発表会講演論文集, pp. 339–340, September, 2006.

- [26] Eriko Aiba, and Minoru Tsuzaki, “Perceptual judgment in synchronization of two complex tones: Relation to the cochlear delays,” *Acoustical Science and Technology*, Vol. 28, No. 5, pp. 357-359, 2007.
- [27] Eriko Aiba, Satomi Tanaka, Minoru Tsuzaki, and Masashi Unoki, “Judgment of perceptual synchrony between two pulses and its relation to the cochlear delays,” *Proc. the 23rd Annual Meeting of the International Society for Psychophysics (Fechner day 2007)*, pp. 211-214, December, 2007.
- [28] 饗庭 絵里子, 津崎 実, “2 音の同時性判断と蝸牛遅延の関連について,” 平成 18 年度日本音響学会秋季研究発表会講演論文集, pp. 339-340, September 2006.
- [29] 浜田 大樹, 鶴木 祐史, “蝸牛遅延に基づいた電子音響透かし法の検討,” 2008 電子情報通信学会総合全国大会公園論文集, DS-4-2, pp. s21-s22, 2008.
- [30] 鶴木 祐史, 宮内 良太, “蝸牛遅延特性に基づいた音響電子透かし,” *日本音響学会誌*, vol. 71, no. 1, pp. 15-22, 2015.
- [31] 鶴木 祐史, 宮内 良太, “蝸牛遅延に基づいた可逆電子音響透かしの検討,” *電子情報通信学会技術研究報告書*, vol. 111, no. 278, pp. 59-64, 2011.
- [32] Masashi Unoki, and Ryota Miyauchi, “Robust, blindly-detectable, and semi-reversible technique of audio watermarking based on cochlear delay,” *IEICE Trans. INF. & SYST.*, Vol. E98-D, No. 1, pp. 38-48, January, 2015.
- [33] Laurence Boney, Ahmed H. Tewfik, and Khaled N. Hamdy, “Digital watermarks for audio signals,” *Proc. ICMCS*, 473-480, 1996.
- [34] 小野 東, “電子透かしとコンテンツ保護,” オーム社, 2001.
- [35] 守谷 健弘, “音声符号化,” *電子情報通信学会*, 1998.
- [36] Dung Kim Tran, Huy Quoc Nguyen, and Masashi Unoki, “Study on speech representation for speech fingerprint using perceptual matching-pursuit algorithm,” *IEICE*, vol. 118, no. 313, pp. 71-76, 2018.
- [37] Akira Nishimura, “Data hiding in speech sounds using subband amplitude modulation robust against reverberations and background noise,” *Proceedings of the 2006 International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 7-10, 2006.
- [38] 鶴木 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,” *電子情報通信学会技術研究報告書*, vol. 110, no. 71, pp.31-36, 2010.

- [39] 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, September, 2007.
- [40] 西村 明, “振幅変調に基づくデータハイディング手法の携帯電話音声コーデックへの耐性,” 第二回マルチメディア情報ハイディング研究会資料, pp. 51–56, November, 2007.
- [41] 鵜木 祐史, “聴覚特性に基づいた音響情報ハイディング技術,” 電子情報通信学会技術研究報告書, vol. 118, no. 313, EMM2018-78, pp. 153–156, 2018.
- [42] 吉谷 清澄, “PN 系列—特に M 系列について,” 電波研究所季報, Vol. 17, No. 90, pp.249–263, 1971.
- [43] 森勢 将雅, “音声分析合成 (音響テクノロジーシリーズ 22),” コロナ社, 2018.
- [44] 今井 聖, “音声信号処理 音声の性質と聴覚の特性を考慮した信号処理,” 森北出版, 1996.
- [45] Albert S. Bregman, “Auditory Scen Analysis: The Perceptual organization of Sound,” MIT Press, 1990.
- [46] 小杉 敏三, 羽二 生篤, 宮内 良太, 鵜木 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,” 電子情報通信学会技術研究報告書, vol. 109, no. 472, pp.13–18, 2010.
- [47] Ryuki Tachibana, Shuichi Shimizu, Seiji Kobayashi, and Taiga Nakamura, “An audio watermarking method using a two-dimensional pseudo-random array,” Signal Process., vol. 82 no.10, pp.1455–1469, October, 2002.
- [48] Masashi Unoki, Jessada Karnjana, Shengbei Wang, Nhut Minh Ngo, and Ryota Miyauchi, “Comparative evaluations of inaudible and robust watermarking for digital audio signals,” ICSV21, pp. 1–8, 2014.
- [49] 西村 明, “振幅変調に基づく音響電子透かしによる音質劣化の評価,” 平成 18 年度日本音響学会秋季研究発表会講演論文集, pp. 456–457, September, 2006.
- [50] 西村 明, “振幅変調に基づく音響電子透かしの頑強性について,” 平成 18 年度日本音響学会春季研究発表会講演論文集, pp. 521–522, 2006.
- [51] 情報ハイディング及びその評価研究会, “Evaluation procedure for audio watermark ver.5,” <https://www.ieice.org/iss/emm/ihc/audio/audio.html>, (2019 年 2 月現在) .
- [52] STEP2001, “プレスリリース (2001.10.19), 音楽電子透かし 4 社を技術認定,” 社団法人日本音楽著作権協会 (JASRAC), https://www.jasrac.or.jp/release/01/10_2.html, (2019 年 2 月現在) .

- [53] Kazuya Takeda, et al., “Speech database user’s manual,” ATR Technical Report, TR-I-0028, 2010.
- [54] International Telecommunication Union Telecommunication Standardization Sector, “ITU-T G.726 page,” <https://www.itu.int/rec/T-REC-G.726/en>, (Retrieved February 2019).
- [55] International Telecommunication Union Telecommunication Standardization Sector, “ITU-T G.729 page,” <https://www.itu.int/rec/T-REC-G.729/e>, (Retrieved February 2019).
- [56] Yiqing Lin, and Waleed H. Adbullah, “Audio Watermark,” Springer, 2015.
- [57] Shengbei Wang, Weitao Yuan, Jianming Wang, and Masashi Unoki, “Speech watermarking based on robust principal component analysis and formant manipulations,” Proc. ICMCS, pp. 2082–2086, 2018.
- [58] 河原 英紀, 津崎 実, 坂野 秀樹, 森勢 将雅, 松井 淑恵, 入野 俊夫, “velvet noise とその変種の聴覚心理・生理研究への応用可能性について,” 電子情報通信学会技術研究報告書, vol. 117, no. 470, HIP2017-113, pp. 99–104, 2018.

研究業績一覧

国際会議

1. Reiya Namikawa, and Masashi Unoki, “Study on speech watermarking method based on spread-spectrum using linear prediction residue,” 2019 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing (NCSP’19), 2019.

国内発表

1. 並河伶弥, 鵜木祐史, “音声符号化におけるスペクトル拡散型音声電子透かしの検討,” 電子情報通信学会技術研究報告書, EMM2017-79, pp. 13–16, 2018.
2. 並河伶弥, 鵜木祐史, “線形予測符号化に基づくスペクトル拡散型音声電子透かしの検討,” 平成 30 年度 電気関係学会北陸支部連合大会, G-2, 2018.
3. 並河伶弥, 鵜木祐史, “音声符号化におけるスペクトル拡散型音声電子透かしの検討,” 電子情報通信学会技術研究報告書, EMM2018-68, pp. 31–36, 2018.

受賞

1. 並河伶弥, 日本音響学会北陸支部 優秀論文発表賞, September 8th, 2018.

謝辞

まず研究活動から私生活まで厳しくも非常に丁寧な指導をしていただいた、主指導教員である鵜木祐史教授に、深く感謝致します。

研究室会議やミーティングなどの場において、赤木正人教授、木谷俊介助教は研究のあり方や考察のご指導を数多くいただきました。ここに感謝いたします。

研究における討論はもちろんのこと、私生活においても数多くの助言をくださった鳥谷先輩、磯山先輩、関谷先輩、高橋先輩方に深く感謝致します。

畠山さん、大谷さん、松本さん、川村さんをはじめとする同窓生の皆さん、大学院生の方々など皆さんには常に刺激的な議論を頂き、また生活面でも精神的にも支えられました。ありがとうございます。これからもぜひ遊びに行きましょう。

寶坂さん、水野くん、西垣くん、宮川さん、及川くんを始めとする M1 の皆さんには、忙しい中研究会原稿の確認や発表練習など数多くのお手伝いを頂きました。ありがとうございます。

研究で行き詰まったときに相談に乗って頂いた、鵜木・赤木研究室のスタッフを含めた全メンバーの皆様に感謝致します。

JAIST での生活を通し、私を支えてくださった皆様に心より感謝致します。

最後にこれまでの私の生活を支え、温かく見守ってくれた両親、家族の皆様に心から感謝致します。

付 録 A 拡散信号の異なる スペクトル型 情報ハイディング法

本稿では 3.4 節で述べたように，DSS 法の原理を元に提案法を提案した．提案法では拡散信号に LP 残差を利用したが，3.2 節で述べたような条件を満たした信号であれば，PN 信号や LP 残差以外の拡散信号としてスペクトル拡散型の情報ハイディングが可能であるとされる．その信号の一つとして白色雑音が挙げられる．

A.1 白色雑音を用いたスペクトル拡散法

白色雑音は，全ての周波数域で持つパワーが等しくなる信号であり，その信号の期待値は“0”，その信号の二乗の期待値が“1”となる性質を持つ．これより，スペクトル拡散型の情報ハイディング法における拡散信号としての条件を満たしており，スペクトル拡散型の情報ハイディング法として利用可能であることが考えられる．そこで，上記の条件を満たす白色雑音を拡散信号に用いた手法が，スペクトル拡散型情報ハイディング法として利用可能か検討した．

検討における情報の埋め込み，検出は DSS 法と同様に行われ，利用する拡散信号のみが変更されている．検討法において DSS 法と同程度の BER での情報検出が可能であれば，検討法が情報ハイディング法として利用可能であることが考えられる．図 A.1 に検討法のブロックダイアグラムを示す．ここで図 A.1(a) はメッセージの埋め込み方法を，図 A.1(b) は情報の検出方法を表す．

A.2 検討法の評価

本検討では 4 bps の埋め込み情報量について BER を算出し検討法の評価を行った。情報の埋め込みにおいて、拡散信号は埋め込み強度設定を行い、この設定値は DSS 法の標準設定値を利用した。情報の埋め込みに利用した音声は ATR データベース (B) より 3 種類の音声に対し、1 秒間を抜き出した音声発話区間を刺激とした [53]。検討の比較として同条件の DSS 法についても BER が算出された。図 A.2 に検討法を用いた情報ハイディングの処理例を示す。ここで図 A.2(a) はホスト信号を示し、図 A.2(b) はそのスペクトルを示す。図 A.2(c) はメッセージ信号を示し、図 A.2(d) はそのスペクトルを示す。図 A.2(e) は拡散信号とした白色雑音を示し、図 A.2(f) はそのスペクトルを示す。図 A.2(g) は透かしが埋め込まれた信号を示し、図 A.2(h) はそのスペクトルを示す。

情報の検出の結果、拡散信号に、白色雑音を用いた場合、PN 信号を用いた場合のどちらにおいても、埋め込んだ 4 bps の情報の BER は 0% であり、100% の検出精度で情報の検出が可能であった。これより、白色雑音を用いてスペクトル拡散型の情報ハイディングを行うことは可能であることが示された。しかし、図 A.2(f) に示されるように白色雑音も PN 信号と同様に、広い周波数帯域に雑音を有している。そのため、透かしが埋め込まれた信号にもホスト信号とは異なった周波数帯域でスペクトルの振幅増加が見られ、これは DSS 法と同様な音声の歪みとして知覚され、音質の劣化が生じることが考えられる。

以上の結果より、透かしが埋め込まれた信号に歪みが生じていても問題がない場合であれば、検討法は情報ハイディング法として利用することが可能であることが示唆された。

A.3 検討法のまとめ

本検討では、本稿 3.2 節で述べた拡散信号に求められる条件から、条件を満たした信号として白色雑音を拡散信号に利用したスペクトル拡散型情報ハイディング法について情報ハイディングとして利用可能か検討した。BER の評価結果より、DSS 法と同程度の検出精度を持つことが示され、拡散信号に白色雑音を用いた手法も情報ハイディング法として利用することが可能であることが示唆された。

既存法である DSS 法で利用される PN 信号や、本稿の提案法における LP 残差、本章で利用した白色雑音以外にも、拡散信号としての条件を満たした信号は、velvet noise などまだ複数存在する [58]。本稿では、知覚不可能性が高い情報ハイディング法を目的としていたため、ホスト信号にスペクトル形状の類似した LP 残差を利用したが、情報ハイディング法としては、上記以外にも拡散信号として利用可能な信号がいくつも存在するはずである。マルチメディア情報ハイディングでは、それぞれの拡散信号を用いた場合のスペクトル拡散型情報ハイディング法の利点欠点を明確にし、使用目的に合わせた情報ハイディング法を用いることが重要であると考えられる。

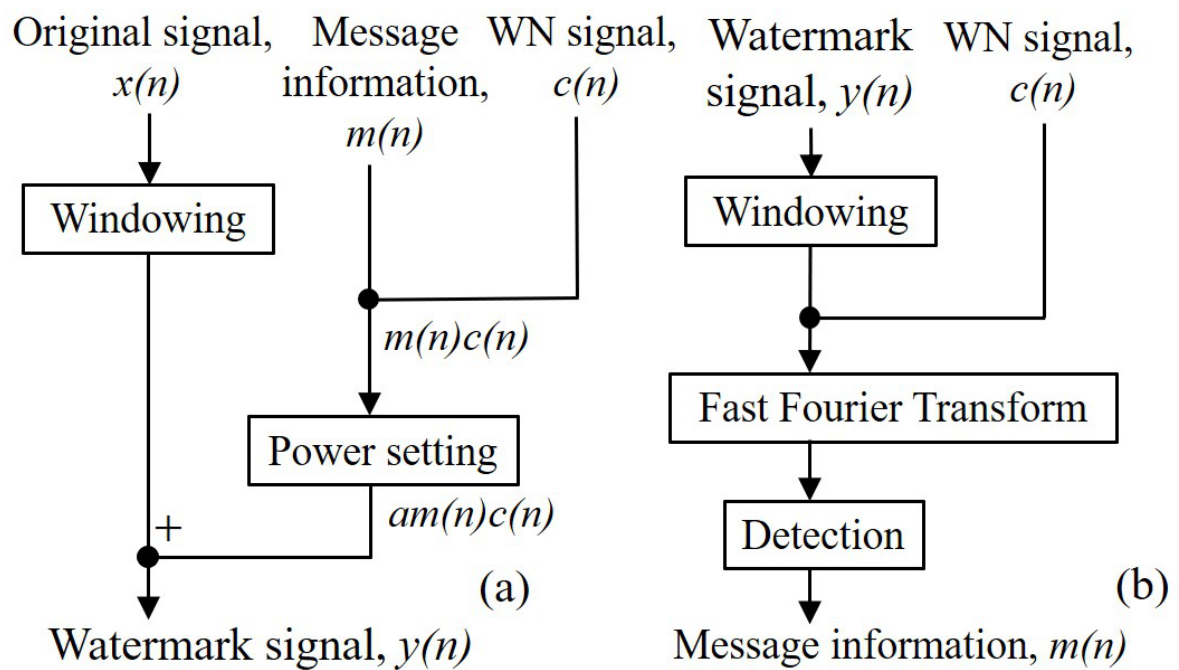


図 A.1: 白色雑音を用いたスペクトル拡散法のブロックダイアグラム : (a) 埋め込み処理, (b) 検出処理.

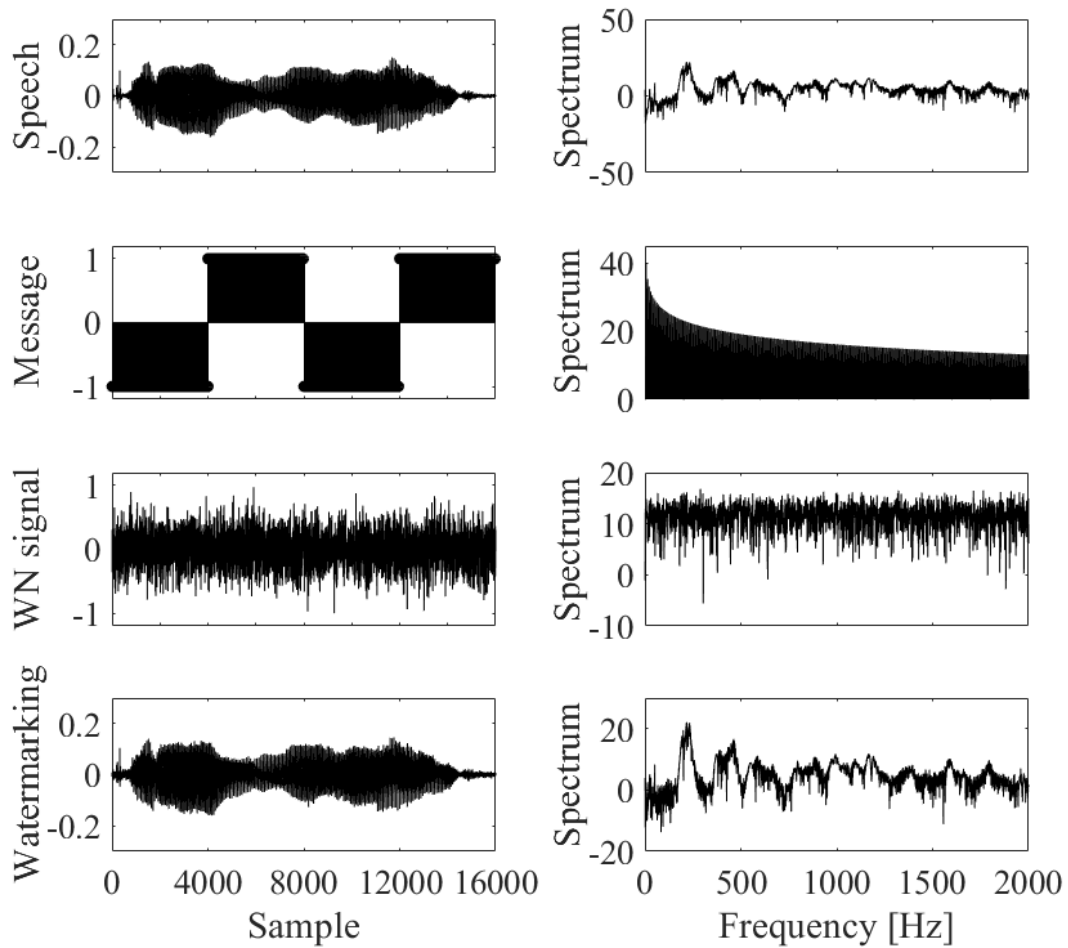


図 A.2: 白色雑音を用いたスペクトル拡散型情報ハイディングの処理例：(a) 音声信号, (b) 音声信号のスペクトル情報, (c) メッセージ信号, (d) メッセージ信号のスペクトル情報, (e) 白色雑音信号, (f) 白色雑音信号のスペクトル情報, (g) 透かしが埋め込まれた音声信号, (h) 透かしが埋め込まれた音声信号のスペクトル情報.