



Title	音響機器の特性を利用した情報ハイディングの研究
Author(s)	松永, 悠斗
Citation	北海道大学. 博士(情報科学) 甲第15549号
Issue Date	2023-03-23
DOI	10.14943/doctoral.k15549
Doc URL	<a href="http://hdl.handle.net/2115/89509">http://hdl.handle.net/2115/89509</a>
Type	theses (doctoral)
File Information	Yuto_Matsunaga.pdf



[Instructions for use](#)

博士後期課程 学位論文

音響機器の特性を利用した情報ハイ  
ディングの研究

A Study on Information Hiding Using Characteristics  
of Acoustic Equipment



松永 悠斗

北海道大学大学院 情報科学院 情報科学専攻  
メディアネットワークコース  
情報メディア環境学研究室

2023年3月

本論文は北海道大学大学院情報科学院に  
博士(情報学) 授与の要件として提出した博士論文である。

松永 悠斗

審査委員

主査

土橋 宜典

副査

長谷山 美紀

坂本 雄児

荒木 健治

## 概要

デジタルコンテンツに対する著作権保護の手法として、情報ハイディング技術が注目されている。情報ハイディングとは、任意の秘匿情報をデジタルコンテンツそのものに対して埋め込み、任意のユーザがこれを自由に検出できるような技術である。デジタルコンテンツの中でも楽曲に関する、著作権保護のための情報ハイディングは、埋め込みによるコンテンツの品質を損なわない点とファイルのフォーマット変換やコンテンツの編集が行われた場合でも秘匿情報が消えない点が求められる。楽曲に対する情報ハイディングでは、アーティストやレコード会社にとって自身の作品に加工が加えられるという心理的抵抗感が高いことが知られており、この品質劣化を最小限に抑えるという点が重要な問題となる。

近年の情報ハイディングの研究は、秘匿情報を用いて生成した信号がデジタルコンテンツの一部として成立するようなアプローチの研究が行われており、これによりコンテンツの品質劣化の問題を避けることが可能となる。本研究では、このアプローチに基づいて、音響信号合成技術や音響効果処理技術を利用した情報ハイディング手法を提案する。

音響信号合成技術を利用した情報ハイディング手法として、打楽器音信号の合成による情報ハイディング手法を提案した。この手法では、打楽器音信号の合成にノイズ信号が用いられていることを利用し、秘匿情報によってノイズ信号の符号を操作することで埋め込みを行う。埋め込み後の楽曲信号とこのノイズ信号との相関関数を計算することでピークが検出され、秘匿情報の検出が可能となる。この手法の攻撃耐性を定量評価実験により評価した結果、MP3 符号化へは高い攻撃耐性が確認されたが、ピッチ変更のような攻撃ではノイズ信号の符号情報が大きく変わってしまうため、攻撃耐性が低いことが確認された。

音響効果処理技術を利用した情報ハイディング手法として、音響歪み効果による情報ハイディング手法を提案した。この手法では、音響歪み効果の発生に用いられるクリッピング処理を利用して、秘匿情報によって歪み効果をかけた信号の振幅値に偏りが出るようにクリッピング処理を操作することで埋め込みを行う。埋め込み後の楽曲データに対して、平均値を計算することで、その符号から秘匿情報の検出が可能となる。この手法の音響歪み効果としての自然さを主観評価実験によって評価した結果、自然な音響歪み効果であることが確認された。また、この手法の攻撃耐性を定量評価実験により評価した結果、MP3

符号化や再生速度の変更，ピッチ変更といった様々な攻撃に対して高い攻撃耐性が確認された．しかし，High Pass Filter による周波数フィルタの攻撃に対して攻撃耐性が低いことが確認された．この問題点を改善するために，波形のエンベロープを用いて補正をかけたところ，検出精度の改善が確認され，攻撃耐性を高めることに成功した．

様々な音楽ジャンルに対して本手法を適用するために，更なる検出精度の改善が求められる．そのため，機械学習を用いた2値分類による秘匿情報の検出を行うことで，検出精度の改善を可能とした手法を提供した．この手法では，畳み込みニューラルネットワークから構成される学習モデルを用いて秘匿情報の検出を行うことで，弱い埋め込みの場合でもビット誤り率を10%以下に抑えることができ，実用にも耐えうると考えられる．

本研究と同様のアプローチによる，楽曲データに対する情報ハイディングの関連研究では，ピッチ変更や周波数フィルタ攻撃といった攻撃に対して攻撃耐性を持っていなかった．本研究の音響歪み効果による情報ハイディング手法は，これまでの実験から，関連研究では達成できなかったピッチ変更や周波数フィルタ攻撃といった攻撃に対しても耐性を持っていることが確認できた．関連研究では，既に一つの楽曲として完成されたデータに対して埋め込みを行うため，情報ハイディングの利用者，すなわちアーティストやレコード会社にとっての心理的抵抗の問題に取り組んだ手法とは言い難い．しかし，本研究の手法では，楽曲作成時の音響機器の一つとして用いられることを想定しており，主観評価実験からも秘匿情報の有無やその内容に関わらず自然な音響歪み効果であると言えるため，この心理的抵抗を緩和している手法と言える．これらのことから，攻撃耐性，利用者への心理的抵抗といった点で関連研究より優れた手法であると考えられる．

# 目次

<b>第 1 章</b>	<b>序論</b>	<b>3</b>
1.1	本研究の背景 . . . . .	3
1.2	本研究の目的 . . . . .	12
1.3	本論文の構成 . . . . .	14
<b>第 2 章</b>	<b>関連研究</b>	<b>15</b>
2.1	非負値行列因子分解を用いた音響情報ハイディング . . . . .	15
2.2	楽曲の和音を用いた音響情報ハイディング . . . . .	18
<b>第 3 章</b>	<b>打楽器音の合成による情報ハイディング</b>	<b>20</b>
3.1	打楽器音の特徴 . . . . .	20
3.2	埋め込み手法 . . . . .	22
3.3	検出手法 . . . . .	24
3.4	ノイズ信号 . . . . .	26
3.5	実験 . . . . .	29
<b>第 4 章</b>	<b>音響歪み効果による情報ハイディング</b>	<b>34</b>
4.1	音響歪み効果 . . . . .	34
4.2	埋め込み手法 . . . . .	38
4.3	検出手法 . . . . .	44
4.4	主観評価実験 . . . . .	45
4.5	客観評価実験 . . . . .	50
<b>第 5 章</b>	<b>機械学習を用いた秘匿情報の検出</b>	<b>63</b>

5.1	目的 . . . . .	63
5.2	検出手法 . . . . .	65
5.3	客観評価実験 . . . . .	67
5.4	考察 . . . . .	69
第 6 章	まとめと今後の展望	74

# 第1章

## 序論

### 1.1 本研究の背景

デジタルコンテンツの大きな特徴の一つとして、いくらコピーを作成しても品質が劣化しないことが挙げられる。これはデジタルコンテンツがデジタル化されていないコンテンツと比較した場合に商業的に優れている点である一方で、容易に不正なコピーを作成されてしまうという欠点にもなり得る。

現代では、情報通信技術の急速な発展やインフラ整備によるインターネットの普及によって、画像・映像・音楽などのデジタルコンテンツをPCやスマートフォンなどの個人のデバイスで視聴することが当たり前となった。また、デジタル信号処理技術の発展によって、これらのデジタルコンテンツを個人で加工することも容易となってきた。これらのことから、デジタルコンテンツの不正なコピーを個人が容易に配信できるようになり、デジタルコンテンツの著作権保護に関する研究が盛んに行われるようになり、その手法の一つとして電子透かし技術の研究が進められてきた。また昨今では、デジタルコンテンツは視聴する以外にも、様々な用途が存在する。例えば、深層学習の研究分野などでは、ネットワークを通じて膨大な数のデジタルコンテンツを収集し学習用のデータとして用いることが世界的に行われている。ビジネスシーンでもネットワーク上の膨大な数のデジタルコンテンツがビッグデータとして活用されている。このように、多くの人々が、自身の知らない間に個人情報が含まれるデジタルコンテンツを利用されるリスクも増加しており、デジタルコンテンツを安全に利用するためにも電子透かし技術の応用が期待されている [1].



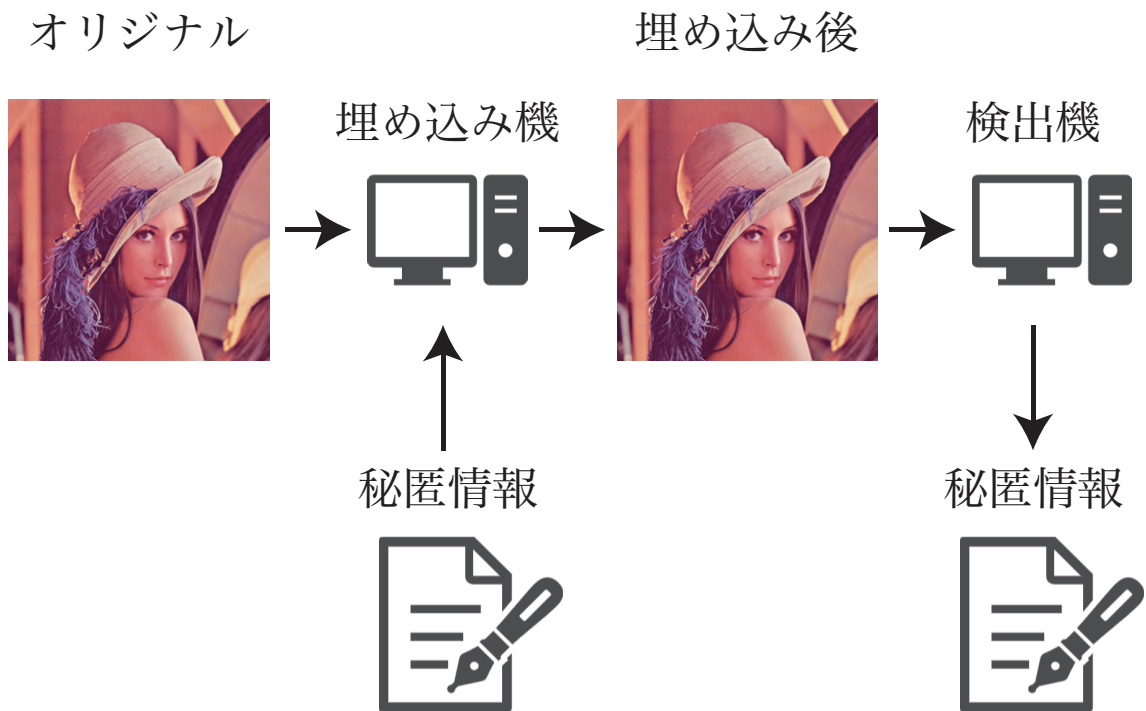


図 1.1.1 電子透かし技術

デジタルコンテンツでは一般に、符号化方式などそのコンテンツを正しく扱うための情報が格納されたヘッダと呼ばれる領域が存在するが、著作権情報をヘッダに保存した場合、コンテンツのみをコピーしてヘッダ内の著作権情報を削除したり、別の著作権情報を書き込むようなことが可能となる。電子透かし技術は、コンテンツとしての価値はそのままに、デジタルコンテンツそのものに任意の秘匿情報を分解できない形で持たせる（埋め込む）技術である [2]。著作権保護を目的とした電子透かし技術は、この秘匿情報に著作権者情報などを用いることで、不正なコピーを作成、配布された場合でも正当な著作権者であることを主張することができる。

電子透かし技術のイメージを図 1.1.1 に示す。埋め込みが行われたデータは、一見するとオリジナルのデジタルコンテンツと同じものに見えるが、適切な処理を行うことでコンテンツとは別の情報を取り出す（検出）ことが可能となる [3]。電子透かし技術に求められる条件として、以下の3つが挙げられる [4]。

- (1) 秘匿情報の埋め込みにより、コンテンツの品質を損なわないこと。
- (2) コンテンツの編集や圧縮、フォーマット変換を含む各種変換を受けても秘匿情報が消

えないこと。

(3) 秘匿情報の検出は、許可された者のみが可能であること。

(1) の条件は、コンテンツとしての価値を保つために必要なことであり、電子透かしの研究において最も重要な課題となる。品質の評価として、Signal Noise Ratio (SN 比) を使うことが一般的であるが、人の知覚特性などを利用した手法では、主観評価なども用いられる。(2) の条件は、デジタルコンテンツが状況に応じて様々なファイル形式に変換されることや、悪意のある第三者が埋め込まれた秘匿情報を意図的に削除してくることを想定し、編集や変換のような攻撃に対してある程度の耐性が求められる。一般に、電子透かし技術は、攻撃耐性を高めると品質の劣化が発生するため、これらはトレードオフの関係にある。(3) の条件は、秘匿情報の検出が可能な場合、消去も容易な傾向にあることと、埋め込む秘匿情報が必ずしも著作権情報に限らないため、求められる要件となる。例えば著作権主張や保護を目的とした場合でも、著作権情報ではなく購入者情報を埋め込みことで、不正なコピーや配信を行った個人を特定することが可能となる。このような場合、容易に秘匿情報が検出できてしまうと、個人情報流出につながる。

電子透かし技術の性質を利用し、著作権の主張以外にも、図 1.1.2 に示すような、デジタルコンテンツの改ざん検知やコピーコントロール、デジタルコンテンツを介した情報通信といった応用研究が盛んに行われている [5]。このことから、現在ではデジタルコンテンツに対して意図的に情報を隠す技術の総称として、情報ハイディングという言葉が使われている。

デジタルコンテンツは、編集や変換が容易であるため、改ざんが行われているかを確認することが困難であるが、条件 (2) の耐性をあえて低くした消失しやすい埋め込みを行うことで、改ざんを検知することができる。コピーコントロールは、DVD で用いられるコピー制御のように、コンテンツ毎に定められたコピーの制限を制御するための情報を秘匿情報として埋め込んでおくことで、コピー制限の不正な回避を防ぐことができる。デジタルコンテンツを介した情報通信はステガノグラフィとも呼ばれ、埋め込みの前後でデジタルコンテンツの品質が変わらないことを利用し、伝えたい情報を埋め込んだデジタルコンテンツに対して、埋め込みが行われていることを知っている相手は情報が検出できるが、それを知らない第三者には通信をしていることに気付かれないような通信が可能となる。

デジタルコンテンツにおいて、一般に音楽などの音響データよりも写真などの画像デー

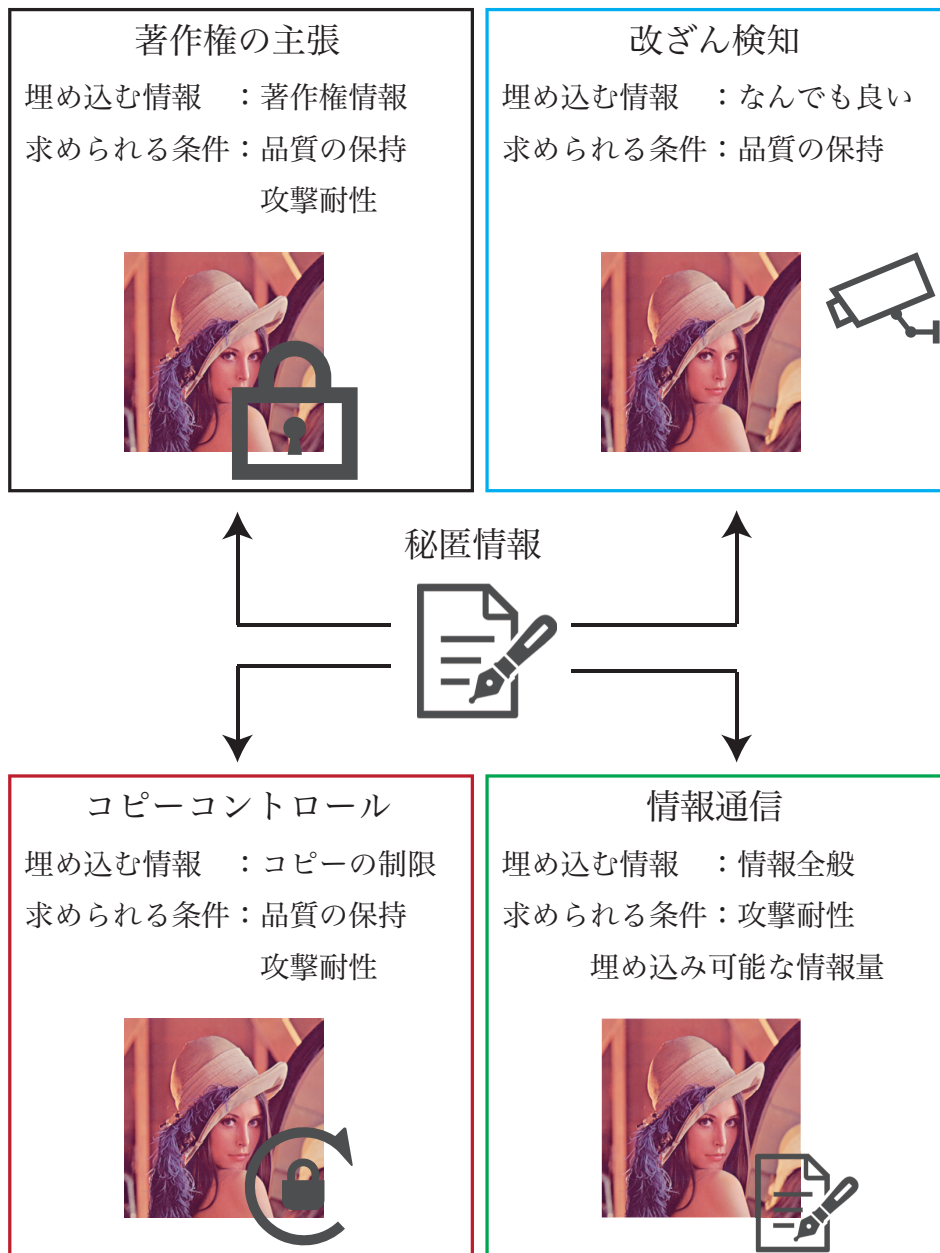


図 1.1.2 情報ハイディング技術の応用例

タの方がファイルサイズが大きい。そのため、画像データの方が秘匿情報を埋め込める余地が大きく、画像データに対する画像情報ハイディング技術が先行して研究されてきた。音響データに対する著作権保護やコピーの防止が遅れたため、音響データの不正なコピーや配信は大きな問題となり、違法ダウンロードサイトやユーザー間のファイル共有サービスといった形で音響データのセキュリティ面の不十分さが表面化した [5]。このことから、

著作権を保護しつつ音響データを流通させるためのフレームワークの確立と標準化が求められ、その手法の一つとして音響情報ハイディング技術が注目されている。

当初、著作権保護のための音響情報ハイディング技術は、コンテンツのコピー制御を目的に研究されてきた。コピー制御は、2[bit] の情報が埋め込めれば良く、埋め込める情報量が少ないような手法でも応用可能であったが、規格や機器の統一が世界レベルで行われる必要があり、現在では不正配信の追跡とその証拠保全のための音響情報ハイディングが注目されている。

不正配信の追跡のためには、インターネット上の配信サイトなどを常に監視し、ファイルなどの様々な履歴を保存しておくシステムやその運営団体が必要となる。また、著作者の許諾を取らずに音響データを配信しているものが存在した場合、個人や非営利などの運用目的に関わらず違法であることの精度面での整備が必要となり、違法である証拠を保持していればプロバイダなどに対して、サイトの運営者情報の開示請求を行うことができる精度の確立も必要とされる。音響データに関しては、日本音楽著作権協会 (JASRAC) によって監視システムが運用され、法律も整備されることで、実用のための基盤が成立した。

実用のための基盤は成立したが、音響情報ハイディング技術が、実用可能な水準に達しているかを評価するための国際評価プロジェクト (STEP 2000, 2001) が JASRAC によって開催された [6]。これにより、実用可能な水準に達している音響情報ハイディング技術を持った企業や、将来その水準に足る技術を提供できるようになるであろう企業を選定された。このようにして、日本国内では 2002 年には、著作権保護のために情報ハイディング技術を用いるための基盤と実用に耐えうる手法やそれを提供できる企業が定まったが、実際には積極的利用されることはなく、普及には至らなかった [4]。

一般的に、著作権保護のための音響情報ハイディングの利用者は、アーティストやレコード会社である。アーティストやレコード会社にとって、作成した音響データは作品であり、事業を行うための重要な財産である。しかし、情報ハイディングによる著作権保護の普及に至らなかったのは、音質へのこだわりと、安全性への不安が要因とされている [1, 4]。

音響情報ハイディングでは、秘匿情報がない音響データとある音響データを用意することになる。この場合、流通する音響データは、秘匿情報のある音響データとなるが、アーティストやレコード会社の所持するオリジナルデータとして、秘匿情報がない音響データが存在することになる。この秘匿情報がない音響データが一度でも外部に出てしまうと、

これを削除するのは困難であり、情報ハイディングを行なった意味が無くなる。そのため、アーティストやレコード会社は、秘匿情報のない音響データが外部に出ないように管理を徹底するか、このデータを廃棄する必要がある。しかし、楽曲制作の過程には、アーティストやレコード会社だけでなく、エンジニアなど様々な人が関わるため、全ての音響データで管理体制の徹底を行うのは困難である。

アーティストやレコード会社にとって、音響データは重要な財産であるため、これを廃棄するのは心理的抵抗が大きく、業務的な不都合も発生する可能性が高いことから、現実的ではない。また、既に発売している楽曲をアルバムなどに収録することを考えた場合、既に秘匿情報のない音響データが一般に流通しているため、アルバムに収録する際に秘匿情報の埋め込みを行なっても、秘匿情報がない音響データとある音響データが混在することになるこのような場合、著作権保護の効果が期待しにくいことや、悪意のある第三者が秘匿情報がない音響データとある音響データを比較することで、実際に施されている埋め込み手法をリバースエンジニアリングできてしまう可能性が発生する。

音響データの著作権保護に用いる音響情報ハイディング技術は、品質（音質）を損なわないことと編集や圧縮、各種変換など（攻撃）を行なっても秘匿情報が消えないことが高いレベルで求められる。音響情報ハイディングの研究が進み、音響データを介した情報通信など別の応用先が検討されることで、音響情報ハイディング技術に求められる技術的な条件も変化し、様々な手法が提案されるようになった [2]。

例として、音のユニバーサルデザイン化を目的としたものが盛んに研究されている [7, 8]。標識や看板などでは、抽象的なデザインで書かれている言語が分からない場合でもその意味を理解することができる。音に関しては、アナウンスのように、外国語話者を考慮して、日本語でのアナウンスの後に英語や中国語など様々な言語で同じ内容のアナウンスが行われているが、その言語が分からない人に対しては情報を伝えることができない。こういった問題に対して、日本語で行われているアナウンスに対して、その内容を秘匿情報として埋め込んでおくことで、個人のスマートフォンなどで秘匿情報を検出し、自身の端末の使用言語に対応した言語でアナウンスの内容を合成音声により再生することや、その画面にテキスト情報として表示することができる。

このような場合では、日本語話者には日本語としての情報が伝われば良いため、ある程度の音質の劣化は許容することができる。しかし、アナウンスを行うスピーカから大気を通じて個人のスマートフォンのマイクで集音するといったアナログの伝搬路となるため、

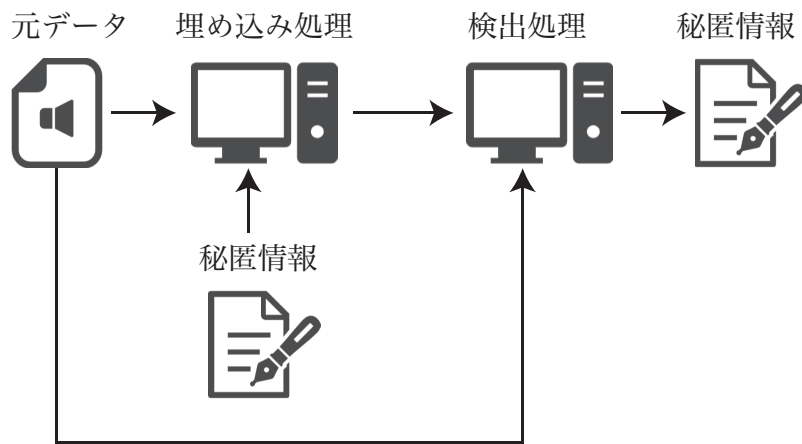
秘匿情報が欠落しやすく、高い耐性が必要となる。アナログの伝搬に耐えうるほどの高い耐性を持った情報ハイディング技術は、デジタル通信のみを対象にしたものと比べると、埋め込み可能な情報量は大きく制限される。しかし、アナウンスでは内容のバリエーションはあまり多くなく、内容も定形文であることが多く、情報量が限られる場合は内容を示すコードを埋め込むだけで対応することができる。

この例のように、応用先によって求められる条件が大きく変化するが、近年では、元データに対して秘匿情報の埋め込みを行うという考えを捨て、秘匿情報で信号を生成し、コンテンツの一部にするという考えの手法が存在する [9, 10, 11, 12]。例として、楽曲に対するこのような手法では、秘匿情報をパラメータの一つとして用いた信号の合成や処理を行い、楽器音信号そのものや音響効果のかかった楽器音信号を生成し、楽曲で用いられる楽器音信号の一つとする。秘匿情報の検出側では、このような信号が用いられた楽曲の楽器音信号に対して、信号の合成や処理に用いられたパラメータをある種の逆問題のようにして求めることで、秘匿情報の検出が行われる。この秘匿情報を用いて生成した信号が、そのコンテンツに含まれるものとして不自然なものでなければ、コンテンツとしての価値を損なうことがなく、信号の誤差という意味での品質の劣化問題を考慮する必要がなくなる。情報ハイディング技術は、攻撃耐性と品質がトレードオフの関係にあるが、このような手法では、知覚されることが問題とならないため、攻撃耐性を高くできる場合が多い。しかし、音声や楽器音、効果音など応用先によって音の統計的な性質や生成のプロセスなどを考慮しつつ秘匿情報による信号の生成を行う必要があり、不自然でないという主観的な評価をどのような手法で行うかが課題となる。

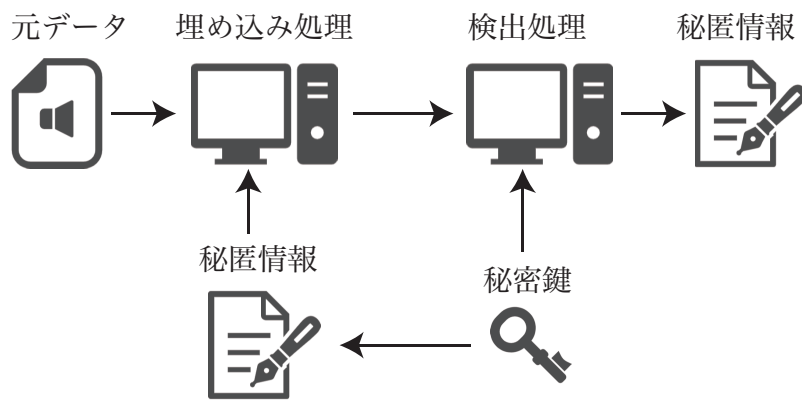
情報ハイディング技術の例として、画素値や振幅値の最下位ビット、画像のエッジ部分や音響信号の高周波成分など人間の視聴覚特性では認識できない成分に対しての埋め込みを考える。検出側では、どの成分に対して埋め込みが行われたのか、どのような信号処理によって埋め込まれたなどを事前に知っておく必要がある。このように、その手法の法則を知っている者のみが、秘匿情報の埋め込まれたデジタルコンテンツから意味のある秘匿情報を検出することができる。しかし、検出において埋め込みを行う前の元データの成分は、検出精度を低下させるようなある種のノイズとなる場合があり、手法によっては検出精度の向上のために、埋め込みを行う前の元データを参照してその差分を取ることが行われる。また、秘匿情報の埋め込み時に秘密鍵を用いるような手法の場合は、この鍵を共有しておく必要がある。このように、情報ハイディング技術は、その手法ごとに埋め込みや

検出に用いられる要件が異なるが、検出側の処理に必要な条件から、図 1.1.3 に示すように大きく分けて 3 つの方式に分類される。

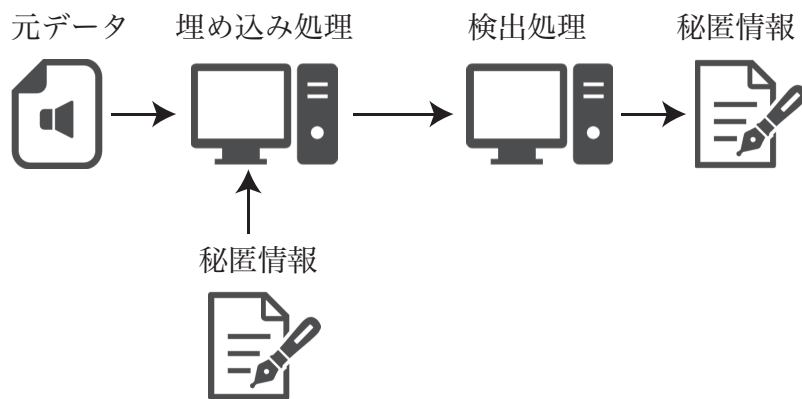
図 1.1.3 に示すように、情報ハイディングにおける検出の処理は、大きく分けてノンブラインド方式、セミブラインド方式、ブラインド方式の 3 つの方式が存在する。ノンブラインド方式は、図 1.1.3(a) に示すように、秘匿情報の検出に元データを参照する手法である。この方式では、埋め込みを行った後のデータと元データの差分を取ることができるため、元データのコンテンツの性質による影響を受けにくく、秘匿情報の検出を容易に行うことができる。秘匿情報の検出が容易なことから、攻撃耐性を妥協することができ、コンテンツの品質保持が容易になる。しかし、ノンブラインド方式では、秘匿情報の検出の度に元データを用いることから、検出システム内に元データを保持しておくことがあり、この元データのセキュリティが問題となるため、応用先が限られる。セミブラインド方式は、図 1.1.3(b) に示すように、埋め込み側と検出側で秘密鍵を共有しておき、検出時にはこの秘密鍵を参照する手法である。この方式では、スペクトル拡散のようにランダムな系列などを用いて埋め込みを行う、この時のランダムな系列を秘密鍵とし検出側と共有することで、検出を行う。秘匿情報を拡散させることで、パワーを低くすることができるため、品質の劣化を抑えることができるが、秘密鍵を保持しておく必要がある。ブラインド方式は、図 1.1.3(c) に示すように、秘匿情報の検出に埋め込み手法以外の情報を用いない手法である。前述の 2 方式と比べ、追加で必要とするデータがないため、元データや秘密鍵のセキュリティ問題が無く、幅広い応用先が期待される。しかし、ブラインド方式では、秘匿情報の検出時に元データもある種のノイズや攻撃として考える必要があり、秘匿情報の検出が困難になる。そのため、安定した秘匿情報の検出を実現するためには、コンテンツの品質保持が困難となる。これらのことから、安定した秘匿情報の検出の実現という意味で、情報ハイディング技術としての実現難易度はノンブラインド方式からブラインド方式になるにつれて難しくなる。



(a) ノンブラインド方式



(b) セミブラインド方式



(c) ブラインド方式

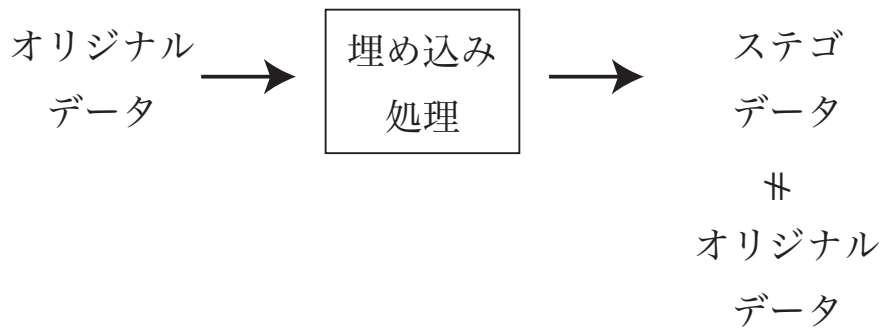
図 1.1.3 検出方式



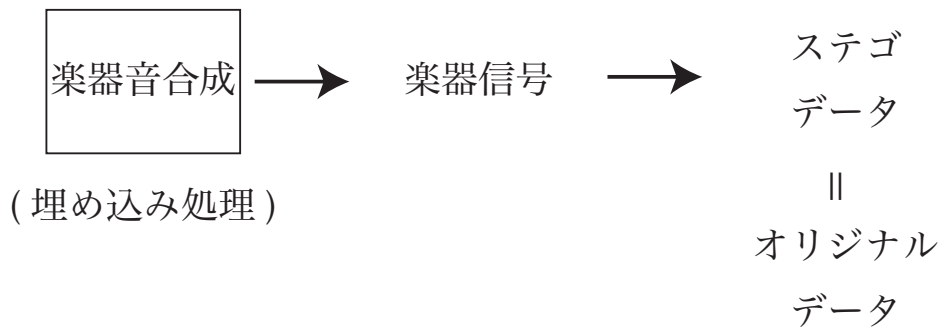
## 1.2 本研究の目的

楽曲に対する音響情報ハイディング技術は、品質の劣化が大きな問題となり、これは音響情報ハイディングを利用することが想定されるアーティストやレコード会社が音質に対する高いこだわりを持っており、求められる水準が高いことが要因の一つであると考えられる [1, 4]. 近年では、音響情報ハイディングの研究が発展したことにより、秘匿情報で生成した信号を用いてコンテンツの一部にするというアプローチの研究が進められている。一般的な音響情報ハイディングでは、秘匿情報の埋め込み処理により、多かれ少なかれ原信号が変化し、これがアーティストにとって本来意図しない変化であるために、劣化と捉えられる。しかし、秘匿情報で生成した信号を用いてコンテンツの一部にするような手法では、秘匿情報をパラメータの一つとして用いて信号合成や信号処理を行うことで、アーティストにとって意図した変化することにできるため、劣化と捉えられることがなくなる。すなわち、楽曲に対する音響情報ハイディングでは、楽器音や音声の合成やこれらの音を加工する処理に基づいて埋め込み処理を行うことになる。このような手法の利点の一つは、コンテンツとしての価値を損なうことがなく、品質の劣化問題を考慮する必要がなくなる点にある。そのため、この利点は楽曲に対する音響情報ハイディングの問題を解決する手段として有用であると考えられる。

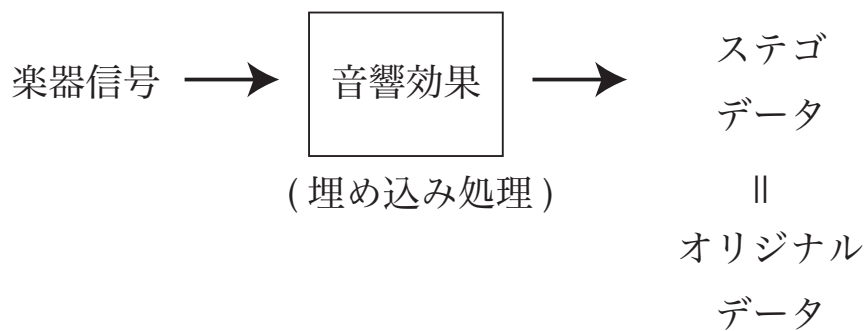
本研究では、秘匿情報で生成した信号を用いてコンテンツの一部にするというアプローチとして、音楽制作において一般的に用いられている音響合成や音響効果の技術に着目した [13, 14]. 図 1.2.1(b) に示すように、音響信号合成技術を音響情報ハイディングに用いる場合は、秘匿情報を一つの信号と見立てて、音響信号合成技術の処理方法に基づいて秘匿情報を加工するか、音響信号合成技術で用いられる何らかのパラメータを秘匿情報によって制御する手法が考えられる。また、図 1.2.1(c) に示すように、音響効果技術を音響情報ハイディングに用いる場合は、楽器音やボーカルなどの信号に対して、音響効果技術の処理方法に用いられる何らかのパラメータを秘匿情報によって制御する手法が考えられる。提案手法では、音響信号合成技術を用いた手法として、M 系列の信号的特性を利用した打楽器音の合成を行う手法を提案し、音響効果技術を用いた手法として、音響歪み効果を利用したエレキギターの歪み音を生成する手法を提案する。



(a) 一般的な音響情報ハイディング



(b) 音響信号合成技術による音響情報ハイディング



(c) 音響効果技術による音響情報ハイディング

図 1.2.1 提案手法の概要

### 1.3 本論文の構成

本稿は、全6章で構成される。第1章は序論である。第2章では、関連研究として、本研究と同様に、秘匿情報で生成した信号をコンテンツの一部として用いるようなアプローチである、2つの音響情報ハイディングの手法を述べる。第3章では、1つ目の提案手法として、打楽器音の合成手法を利用した音響情報ハイディング手法について述べる。まず、打楽器音信号の持つ特徴について述べる。次に、ノイズ信号を用いた打楽器音の生成による秘匿情報の埋め込み手法について述べる。次に、ノイズ信号による相関関数を用いた秘匿情報の検出手法について述べる。最後に、攻撃耐性を評価した定量評価実験について述べる。第4章では、2つ目の提案手法として、音響歪み効果の処理手法を利用した音響情報ハイディング手法について述べる。まず、音響歪み効果とその生成回路について述べる。次に、音響歪み効果を用いた秘匿情報の埋め込み手法について述べる。次に、埋め込みの行われた信号の統計的な性質による秘匿情報の検出手法について述べる。次に、この音響歪み効果が音響交換の一つとして自然なものであるかを評価した主観評価実験について述べる。最後に、攻撃耐性を評価した定量評価実験について述べる。第5章では、音響歪み効果の処理手法を利用した音響情報ハイディング手法において、機械学習を用いることで秘匿情報の検出精度を向上させた手法について述べる。まず、検出精度を向上させる重要性について述べる。次に、機械学習を用いた秘匿情報の検出手法について述べる。次に、この検出手法による検出精度の向上を評価した客観評価実験について述べる。最後に、この手法に関する考察を行う。第6章では、本研究を総括し、今後の課題を述べる。

## 第 2 章

# 関連研究

### 2.1 非負値行列因子分解を用いた音響情報ハイディング

村田らの手法では、非負値行列因子分解 (Non-negative Matrix Factorization: NMF) を用いて音響信号に埋め込みを行う [9, 10]. NMF は、1 つの非負値行列を 2 つの非負値行列に分解するアルゴリズムである. 音響信号に対する NMF は、以下の式のように音響信号のスペクトログラム  $Y$  をスペクトルパターン  $A$  と、スペクトルパターンの強さの時間変化  $B$  に分解することができる.

$$Y \simeq AB \tag{2.1.1}$$

ここで、行列  $A$  を基底行列、行列  $B$  を活性化行列と呼び、活性化行列の要素を活性化係数と呼ぶ. 活性化係数は 0 以上の整数である.

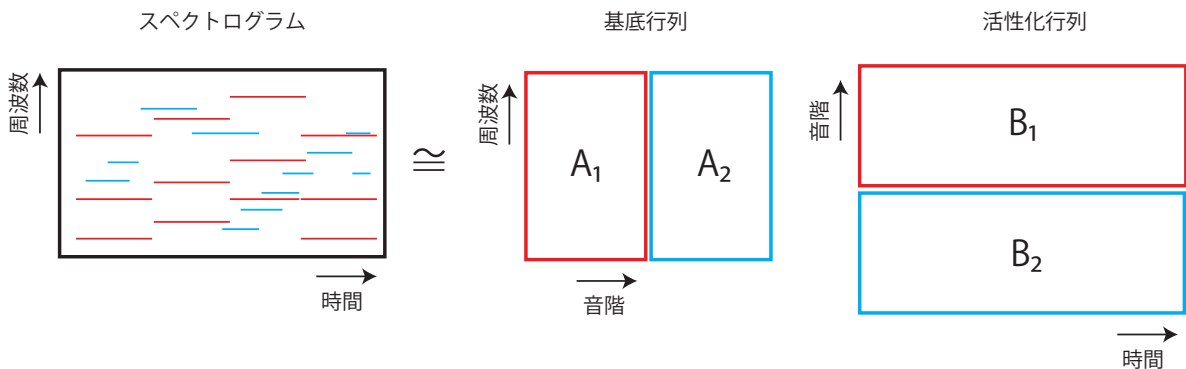


図 2.1.1 NMF による楽器毎の基底行列と活性化行列への分離

複数の楽器音信号からなる信号に対して、用いられている楽器音の基底行列を学習しておくことで、NMFは楽器ごとの基底行列と活性化行列へ分離することが可能である。例として、2種類の楽器で演奏された信号に対するNMFを図2.1.1に示す。この例では、赤色で示されている信号がコードの演奏を行い、青色で示されている信号がメロディの演奏を行っているような場合を想定している。図2.1.1のスペクトログラム $Y$ は、赤色で示されている楽器に関する基底行列 $A_1$ と活性化行列 $B_1$ 、青色で示されている楽器に関する基底行列 $A_2$ と活性化行列 $B_2$ を用いて以下のように分解される。

$$Y \simeq A_1 B_1 + A_2 B_2 \quad (2.1.2)$$

秘匿情報の埋め込みは、特定の楽器信号へ1音に対して1[bit]が埋め込まれるように活性化行列を操作することで行われる[9]。このとき、1音内で最も大きい値を取る活性化係数の値が偶数の場合を埋め込みビットが1、奇数の場合を埋め込みビットが0とし、これを満たすように活性化係数の値を操作する。ステゴデータの生成は、図2.1.2のように行われる。ここでSTFTとISTFTは短時間フーリエ変換と逆変換である。

村田らの手法は、作製したステゴデータに対して、MP3へのファイル形式変換(2種類)、MPEG4へのファイル形式変換、SN比36[dB]によるガウス雑音、D/A変換したものに対するA/D変換を疑似的に行ったもの、これら5種類の攻撃に対する攻撃耐性が評価された。実験の結果、D/A変換したものに対するA/D変換以外の攻撃では、Bit Error Ratio (BER)が0.1以下であり、高い検出精度であったことが報告されている[9]。しかし、D/A変換したものに対するA/D変換では、BERが0.5前後であり、検出精度が著しく低下することも合わせて報告された。D/A変換したものに対するA/D変換は、非可聴領域の周波数成分のように、極端に低い周波数や高い周波数の成分が減衰する傾向にある。このことから、この手法は周波数フィルタに関する攻撃に対して攻撃耐性が低いことが考えられる。

攻撃耐性の評価に加え、Perceptual Evaluation of Audio Quality (PEAQ)による品質劣化の評価がされた[15]。実験の結果、埋め込みによって品質の変化を感じる事が報告された。

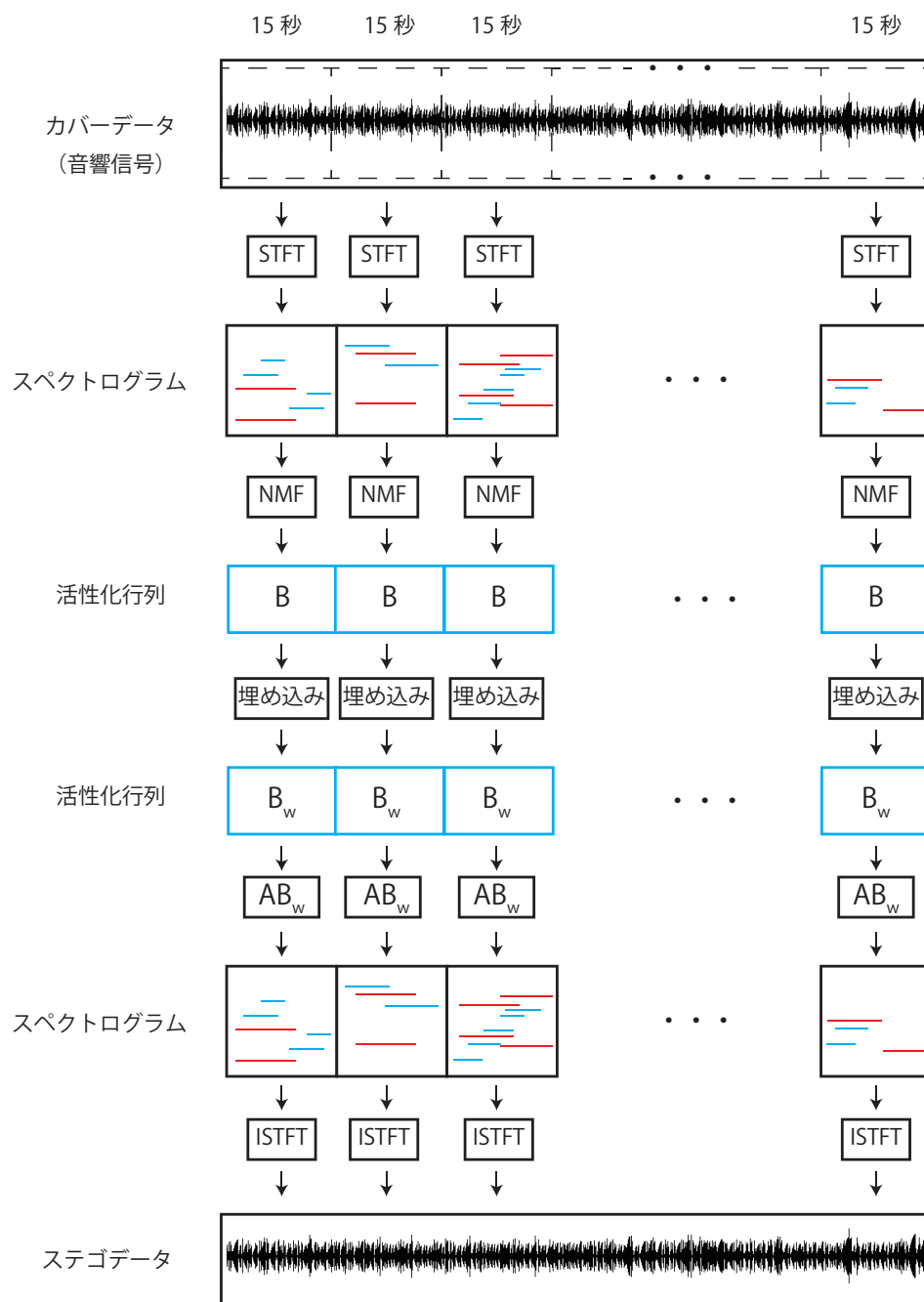


図 2.1.2 村田らの手法によるステゴデータの生成手順

## 2.2 楽曲の和音を用いた音響情報ハイディング

鎌田らの手法では、秘匿情報の埋め込みによって発生した信号を、埋め込み先の楽曲（カバー楽曲）で使用されている周波数と同様の基本周波数となるような信号にすることで、楽曲内で演奏されている楽器の一つであるかのような信号を生成する [11]。そのため、これにより生成された信号は、シンセサイザなどの電子楽器のような音として消費者に知覚されることを前提とした信号であり、この信号の生成にカバー楽曲の和音情報（コード情報）を用いる。

鎌田らの手法による秘匿情報の埋め込みは、相関特性に優れた系列を用いて秘匿情報の変調を行い、コードの構成音の基本周波数と一致するように変調した信号を更に変調することで行われる。この系列には、完全相補系列系（Complete Complementary Code: CCC）と呼ばれる複数の系列の集合を用いて生成される系列を用いる [11]。

ステゴデータの生成は、図 2.2.1 のように行われる。この図の例では、楽曲中の A のコードが演奏されている区間に対して、秘匿情報を CCC で変調することによって生成された信号をコードの構成音である A, C#, E の周波数を基本周波数とする信号として更に変調することで埋め込みを行っている。

鎌田らの手法による秘匿情報の検出は、ステゴデータとカバーデータの差分を計算し、コードの構成音の基本周波数に対応した周波数成分に対して埋め込みに用いた CCC との相関関数を計算することで行われる。そのため、楽曲のコード進行情報、埋め込みに用いた CCC を共有しているものとする、ノンブラインド方式による検出である。

この手法によるステゴ楽曲が、一つの楽曲として自然に認識できるかという点を評価するために、主観評価実験が行われ、攻撃耐性を評価するために、シミュレーションによる客観評価実験が行われた。主観評価実験は、Mean Opinion Score によって楽曲としての品質の評価が行われた。主観評価実験の結果から、ステゴデータは楽曲として自然に認識されていることが報告された。客観評価実験は、ガウス雑音の付加、MP3 符号化（2 種類）、100[ms] のエコー信号の付加、ピッチ変換からなる 5 種類の攻撃が行われ、攻撃後のステゴ楽曲を用いた BER を求めることで攻撃耐性の評価が行われた。客観評価実験の結果から、ガウス雑音の付加、MP3 符号化、100[ms] のエコー信号の付加に対しては、BER が 0.0 に近く高い攻撃耐性を持っていると報告されている。一方で、ピッチ変換の攻撃に

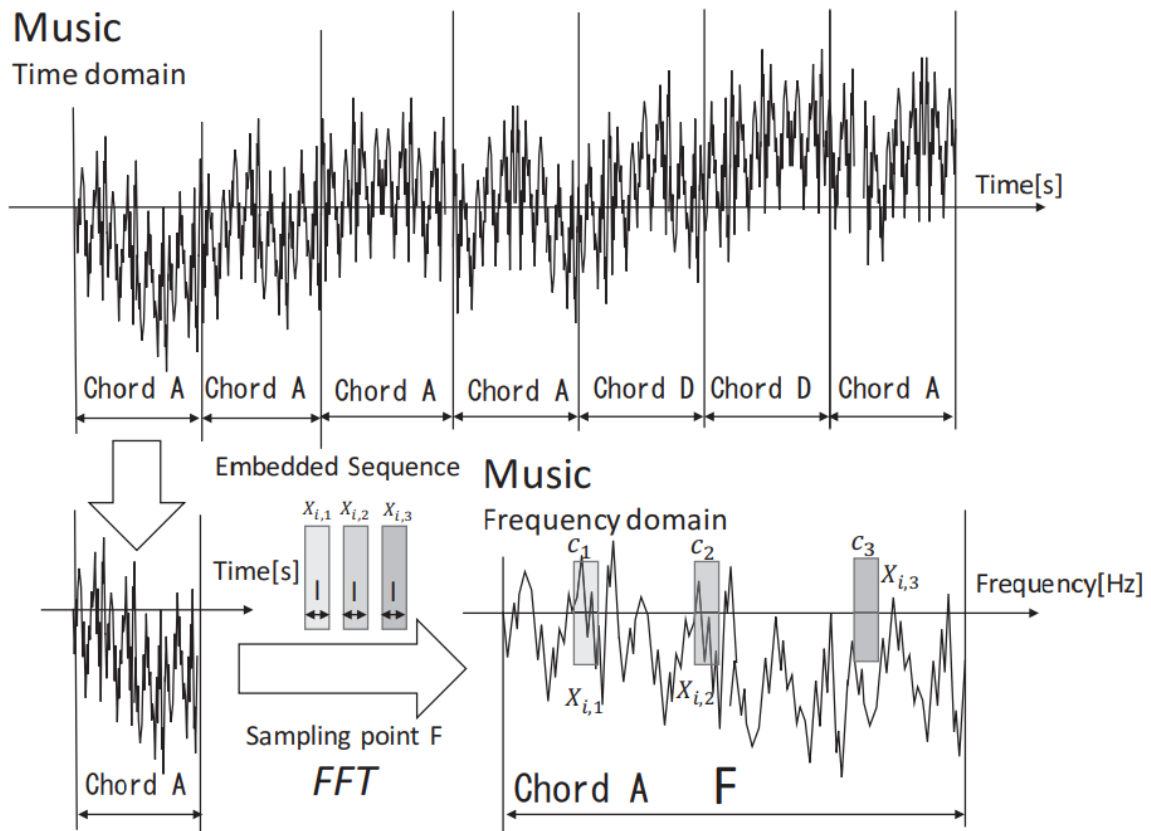


図 2.2.1 埋め込み手順の概要 (文献 [11] から引用)

よる客観評価実験では、BER が 0.4878 となり、攻撃耐性が無いことが報告されている。



## 第3章

# 打楽器音の合成による情報ハイディング

### 3.1 打楽器音の特徴

音響分野の研究では、音の三大要素として、音の高さ、大きさ、音色の3つが知られている。これらはそれぞれ、音波の周波数、振幅、波形に対応する。音楽では、音の周波数の組み合わせが楽曲の雰囲気決定付けるコードとなり、周波数の時間変化が楽曲のメロディーと言えらる。人の聴覚特性は、低い周波数ほど敏感であり、高い周波数ほど鈍感である。そのため、音楽では、人の聴覚上で等間隔となるように音の周波数を並べた、音階と呼ばれるものが用いられる。

一般的な楽器では、音階に適した周波数の音を発生させるため、任意の基本周波数の音を発生させることが目的となる。また、楽器の構造的違いによって発生した波形が異なるため、その楽器特有の音となる。

一般的な打楽器音の信号は、打面の張り具合などによって決定される基本周波数を中心に、一定の範囲の周波数成分をすべて含むような信号となる。この性質から、メロディやコードの演奏よりもリズムを演奏することが多く、リズムの演奏では音の周波数の相対的な高さ関係のみが重要であり、正確な周波数を求められることは少ない。雑音信号は、様々な周波数成分を含むことが特徴の一つである。雑音信号の特徴を利用して、打楽器音の信号は、雑音信号で再現されることがある。例えば、ファミリーコンピュータでは、ホ

ホワイトノイズの再生時間などを調整することでドラムのような打楽器音を表現している。  
このホワイトノイズを用いたドラム音の生成は、チップチューンと呼ばれる音楽ジャンル  
で用いられ、その後様々な音楽ジャンルでも用いられるようになった。

## 3.2 埋め込み手法

本手法では、図 3.2.1 のようにノイズ信号を打楽器音のように加工して、打楽器音と置き換える。このノイズ信号を用いて作成された埋め込み後の信号（ステゴデータ）に対し

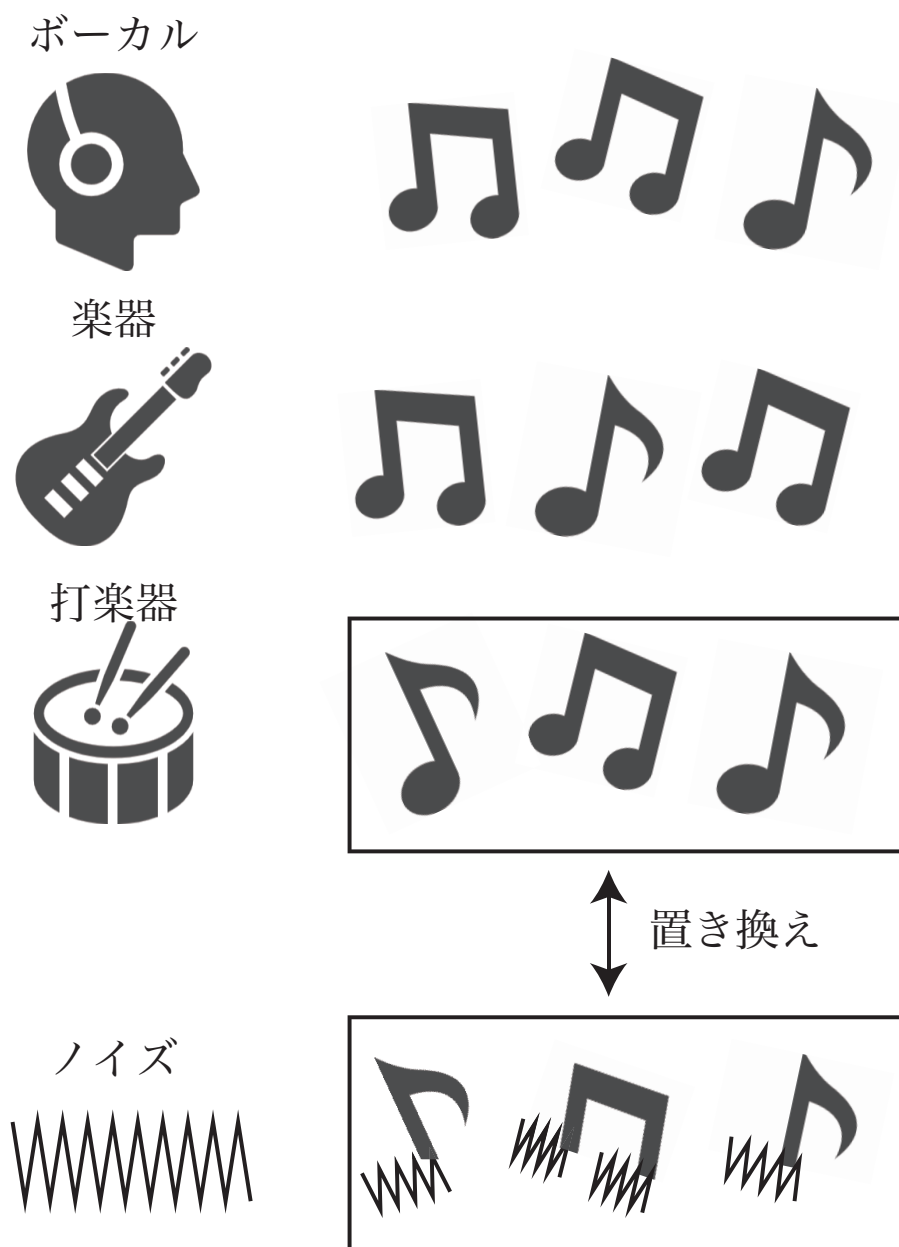


図 3.2.1 打楽器音への埋め込み

て、ノイズに用いた信号と同じ信号を用いて相関関数を計算することで、秘匿情報の検出が可能であると考えられる。ノイズ信号で様々な打楽器音を再現するためには、周波数の調整、アタックの立ち上がり、減衰の仕方など、様々なパラメータを調整する必要がある。しかし、クローズしたハイハットシンバルの音はノイズの再生時間を短くし、余韻を調整するだけでも十分に再現できる。そのため本手法では、クローズしたハイハット音をノイズ信号による打楽器音で再現する。

秘匿情報はハイハットシンバルの1打に対して1[bit]の埋め込みが行われる。8ビートなど一般的なリズムパターンで用いられる、8分音符で刻まれるようにハイハットシンバルを叩くリズムパターンを想定した場合、ビットレート  $R[\text{bit/s}]$  と埋め込み可能な最大情報量は  $P[\text{bit}]$  は、以下の式で決定される。

$$R = 2 \times \frac{T}{60} \quad (3.2.1)$$

$$P = RL \quad (3.2.2)$$

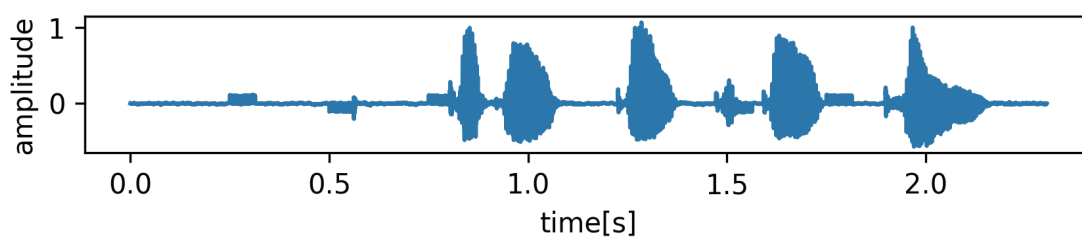
ここで、 $T[\text{BPM}]$  はカバーデータのテンポ、すなわち楽曲における1分間の拍数（1拍あたりを4分音符で分割した場合）であり、 $L[\text{s}]$  はステゴデータの長さ、すなわち楽曲の再生時間である。8分音符は4分音符の半分の長さであることから、 $R$ は1秒ごとの8分音符の数も表している。また、 $P$ は楽曲全体の8分音符の総数である。

ノイズ信号を  $s(n)$ 、埋め込むビットを  $w(k) (0 < k \leq P)$  とした場合、ステゴデータ  $y(n)$  は以下の式で生成される [16]。

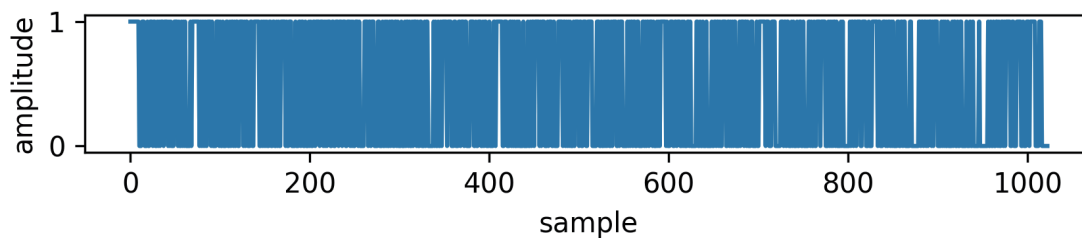
$$y(n) = \begin{cases} s(n) & (w(k) = 0) \\ -1 \times s(n) & (w(k) = 1) \end{cases} \quad (3.2.3)$$

### 3.3 検出手法

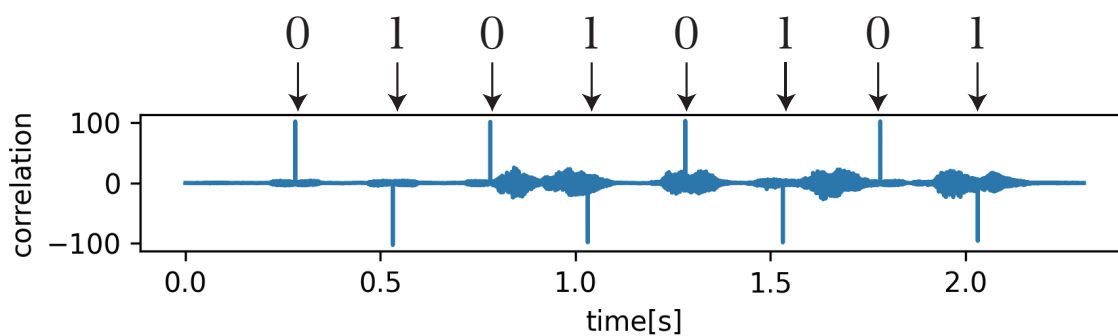
埋め込み手法では，ノイズ信号を用いて秘匿情報の埋め込みを行っている．検出手法は，埋め込みに用いたノイズ信号を用いて情報の検出処理を行う．図 3.3.1 に示すように，検出処理は初めに，ノイズ信号のサイズを  $N$ ，楽曲データ  $y'(t)$  として，以下の式で相関



(a) 秘匿情報の埋め込まれた楽曲データ



(b) 用いたノイズ信号



(c) 相関関数

図 3.3.1 検出例

関数  $c(t)$  の計算が行われる。

$$c(t) = \sum_{n=0}^{N-1} y'(t+n)s(n) \quad (3.3.1)$$

その後、相関関数  $c(t)$  を  $-1 \leq c(t) \leq 1$  で正規化したものを  $c'(t)$  とする。  $c'(t)$  としきい値  $d_c (d_c > 0)$  を用いて以下の式で埋め込まれたビットの検出が行われる。

$$w(k) = \begin{cases} 0 & (c'(t) > d_s) \\ 1 & (c'(t) < -d_s) \\ skip & (Otherwise) \end{cases} \quad (3.3.2)$$

$c'(t)$  は図 3.3.1(C) のように秘匿情報の埋め込まれている点でそれ以外の点よりも大きな値を取るため、しきい値  $d_c$  は、  $c'(t)$  全体の平均値と  $c'(t)$  内の最も大きい数値の中間とし、以下の式で決定される。

$$d_c = \frac{\sum_{t=0}^{N-1} |c'(t)|}{2N} + 0.5 \quad (3.3.3)$$

このことから、本手法は、埋め込み側と検出側で、埋め込みに用いたノイズ信号を共有していれば、秘匿情報の検出が行えるため、ノイズ信号を秘密鍵としたセミブラインド方式となる。

### 3.4 ノイズ信号

ノイズ信号  $s$  は、原理的に乱数の系列であれば問題ないが、検出精度や攻撃耐性を考慮し、疑似雑音系列を用いる。疑似雑音系列の中でも、M 系列は相関特性が良く高い検出精度を期待できるため、提案手法ではノイズ信号として M 系列を用いる。

M 系列は、以下の特徴を持つ

- (1) 系列の要素が 0 と 1 からなるバイナリ系列である
- (2) 周期性を持ち、系列長  $(2^k - 1)$  が周期となる系列である
- (3) 1 周期内で 0 と 1 の出現回数は 1 回しか変わらない
- (4) 1 周期内で同じ要素が連続するもの同士を分類した場合、要素が連続した数を  $r$  とすると、それが含まれる割合は  $1/2^r$  で表される
- (5) 巡回時間シフトにより自己相関関数を計算した場合、0 シフトの場合のみピークが発生し、それ以外のシフトでは -1 となる。

ここで、自己相関特性が優れるのは、M 系列の要素の内 0 を -1 として置き換えて計算した場合である [17]。

周期  $p = 2^k - 1$  となる M 系列  $\{s_i\} = (s_0, s_1, \dots, s_{p-1}, s_0, s_1, \dots)$  は、以下の式で示される  $k$  次の線形回帰方程式によって生成される。

$$s_i = \sum_{l=1}^k c_l a_{i-l} (\text{mod } 2), i = k, k+1, \dots \quad (3.4.1)$$

ここで、系列の初期値  $(s_0, s_1, \dots, s_{k-1})$  は、全てが 0 である場合を除けば任意のバイナリ系列を与えて良い。また、M 系列を生成する際に用いられる、系列長  $k+1$  からなるバイナリ系列  $\{c_l\} = (c_0, c_1, \dots, c_k)$  は以下に示すような、次数  $k$  の原始多項式と呼ばれる多項式  $f(x)$  の係数であり、この係数は 0 か 1 の値を取ることが知られている。

$$f(x) = \sum_{l=0}^k c_l x^l \quad (3.4.2)$$

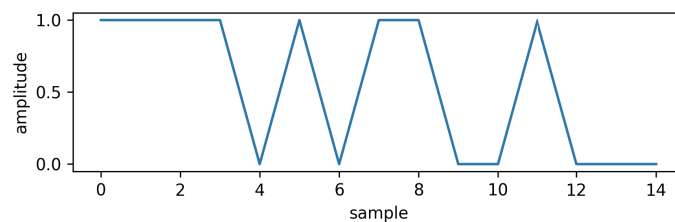
$$c_0 = c_k = 1$$

$$c_l \in \{0, 1\} (0 < l < k)$$

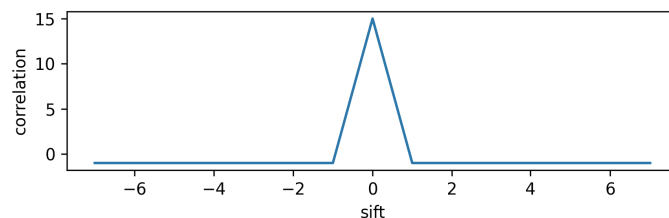
本手法では、このようにして生成された M 系列の 1 周期分をノイズ信号として用いる。原始多項式が同一のものであれば、初期値の異なる M 系列同士は、時間信号における位相が異なるようなものであり、互いにシフトした関係にあるため、相関特性など本質的なところは同一である。

周期 15 ( $k = 4$ ) の M 系列の例を図 3.4.1(a) に示す。この例では 1 が 8 回発生し、0 が 7 回発生していることが確認できる。また、同じ要素が連続するもの同士で分類した場合、単独の要素となるのが 4 つ、2 つの要素からなるのが 2 つ、3 つの要素とからののが 1 つ、4 つの要素とからののが 1 つ存在する。このことから、単独の要素は  $r = 1$  と考え  $4/8 = 1/2 = 1/2^1$  となっており、2 つの要素からなる場合は  $r = 2$  と考え  $2/8 = 1/4 = 1/2^2$  となっており、3 つの要素からなる場合は  $r = 3$  と考え  $1/8 = 1/2^3$  となっていることが確認できる。図 3.4.1(b) に示されているように、自己相関関数は、0 シフトの場合のみピークが発生し、それ以外のシフトでは -1 となっていることが確認できる。

M 系列を用いてノイズ信号とするために、M 系列をパルス信号として扱い、そのまま



(a) M 系列 (要素数 15)



(b) 自己相関係数

図 3.4.1 M 系列の例



音響信号として出力する。1 周期内で同じ数が連続しにくい性質とその出現率の平衡性により、ホワイトノイズのような音となる。そのため、ホワイトノイズで再現したような打楽器音は、M 系列でも再現可能であると考えられる。クローズしたハイハットシンバルはノイズの再生時間を短くする必要がある。そのため、本手法では、M 系列の符号に影響を与えないために、周期 511 の M 系列を用いる。

## 3.5 実験

### 3.5.1 実験条件

本手法の攻撃耐性を評価するため、計算機シミュレーションによる定量評価実験を行った [16]. 本実験では、表 3.5.1 に示す条件のステゴデータを生成し、秘匿情報の検出精度を測定した. また、生成したステゴデータに対し、他の楽器とのミキシング、MP3 符号化、テンポ変更、ピッチ変更の 4 種類の攻撃を行った.

本手法を用いて楽曲を構成する場合、埋め込みが行われている楽器信号と行われていない楽器信号をミキシングすることとなる. したがって、ミキシングを行い楽曲として完成させたものから秘匿情報を検出できる必要がある. ミキシングは、楽曲作成のプロセスとして避けられないものであり、悪意がある攻撃では無いが実用のためには、これに耐える必要がある. 本実験では、ミキシングを行う他の楽器として、エレキギターの信号とエレキベースの信号を用いる. 本実験では、埋め込みが行われていない楽器信号を、検出の精度を落とす一種のノイズ (攻撃) と考え、ノイズを強くした場合の検出精度の変化、すなわち、他の楽器信号の音量 (レベル) を上げた場合のビット誤り率 (Bit Error Rate: BER) の変化を測定する. 埋め込みを行った楽器信号の分散を  $s$ , 埋め込みが行われていない楽器信号全てを足し合わせた信号の分散を  $n$  としたとき、 $10 \log(s/n)$  で計算される Signal to Noise Ratio(SN 比) が  $-32[\text{dB}]$ ,  $-39[\text{dB}]$ ,  $-45[\text{dB}]$ ,  $-52[\text{dB}]$  となるミキシングでステゴデータを作成した. SN 比は高くなるほどノイズの成分が少ない、すなわち、攻撃の影

表 3.5.1 実験に用いたステゴデータ

量子化ビット数	16[bit]
チャンネル数	1
サンプリング周波数	44.1[kHz]
テンポ	98 [BPM]
再生時間	101[s]
秘匿情報	ランダムなビット列
データサイズ	285[bit]
ビットレート	3.26[bit/s]

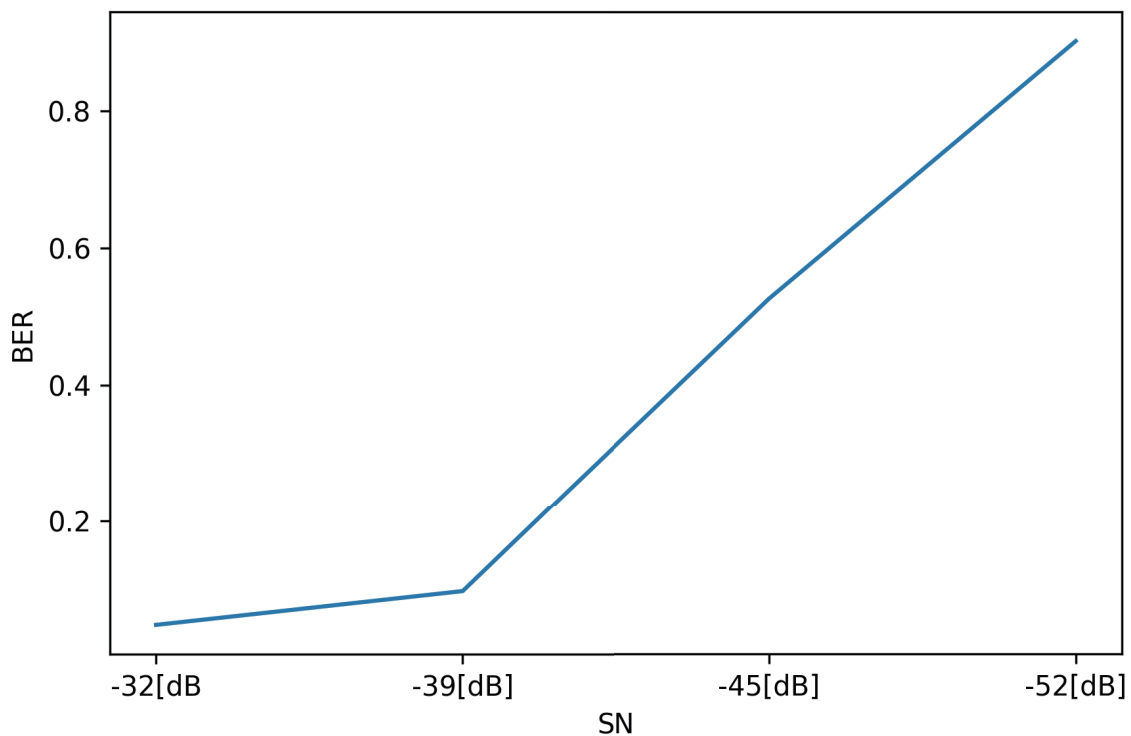


図 3.5.1 ミキシングに対する攻撃耐性

響が少ない条件である。

### 3.5.2 実験結果

ミキシングに対する実験の結果を図 3.5.1 に示す [16]。BER は 0 に近づくほど誤りが少なく、秘匿情報の検出精度が高い，すなわち，攻撃耐性が高いことを示している。SN 比が-45[dB] 以下の場合で急激に BER が上昇していることが分かる。この要因として，秘匿情報の検出に相関関数を使用しているため，ミキシングで他の楽器よりもレベルが小さくなればなるほど他の楽器の影響を受けやすくなり，正負も変わってしまう場合があるため，十分な相関のピーク値を得られなかったことが考えられる。-52[dB] の場合で BER が 0.9 となっているが，これは，ピークが発見できなかった場合に，エラーとする処理を行ったためである。

同一の系列長の異なる原始多項式からなる M 系列同士の中でも，相互相関の値が小さい 3 種類の値のみをとる M 系列の対が存在し，これらはプリファードペアと呼ばれる

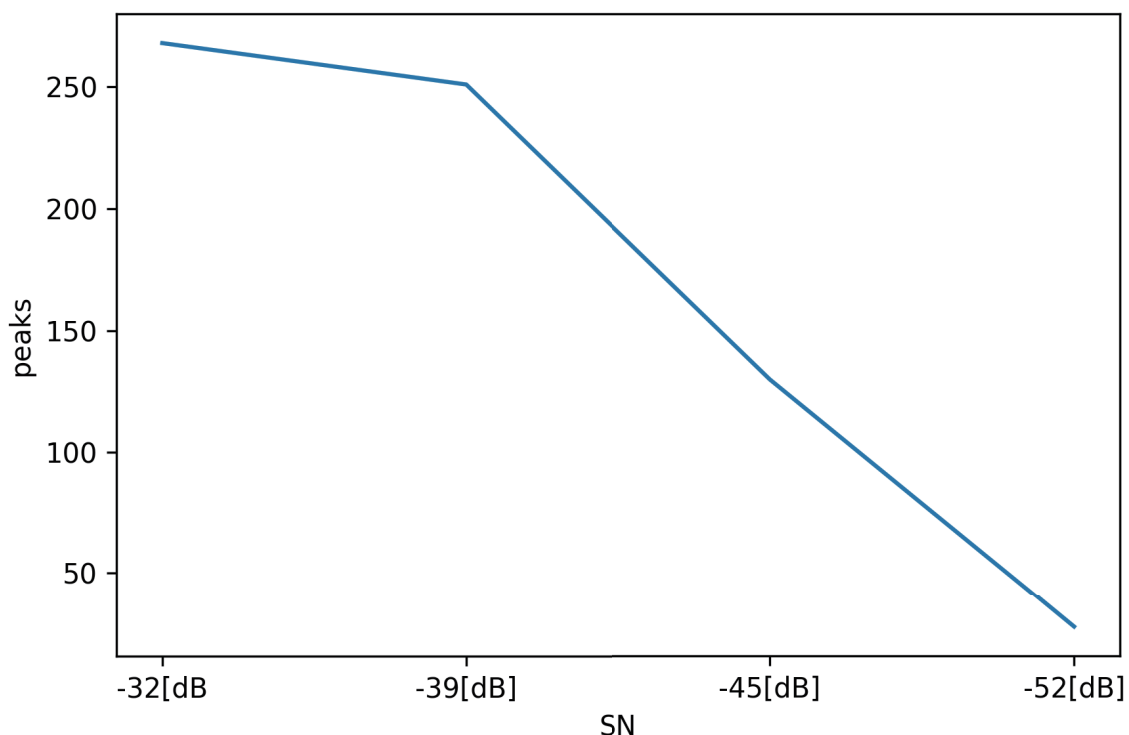


図 3.5.2 検出されたピーク数

[18]. 著作権の主張を考えた場合、プリファードペアな M 系列を用いて、ユーザー毎に異なる M 系列を割り振ることで、ピークの検出によってユーザーの証明が可能であり、必ずしも意味のあるビット列を埋め込む必要は無いと考えられる。今回の実験で検出されたピーク数を図 3.5.2 に示す [16]. -45[dB] でのミキシングの場合でも、285[bit] の埋め込み中 135 本のピークが検出されている。ユーザーが異なる、すなわち、埋め込みに用いられている M 系列が異なる場合は、極めて少ないピーク数になることが考えられるため、135 本という半数に近いピークが検出できていれば、ユーザーの特定は可能であると考えられる。

以降の MP3 符号化、テンポ変更、ピッチ変更に対する攻撃耐性の評価は、このミキシングされたステゴデータを用いて行う。MP3 符号化は固定長ビットレート (192[kbps]) による変換を行い、再度 wav ファイルに変換したものをを用いて秘匿情報の検出を行う。

テンポ変更は、動画共有サイトなどの違法アップロードのように、ある程度コンテンツの価値を保ちながら攻撃が行われることを想定し、10% の加速と減速を行った。これに

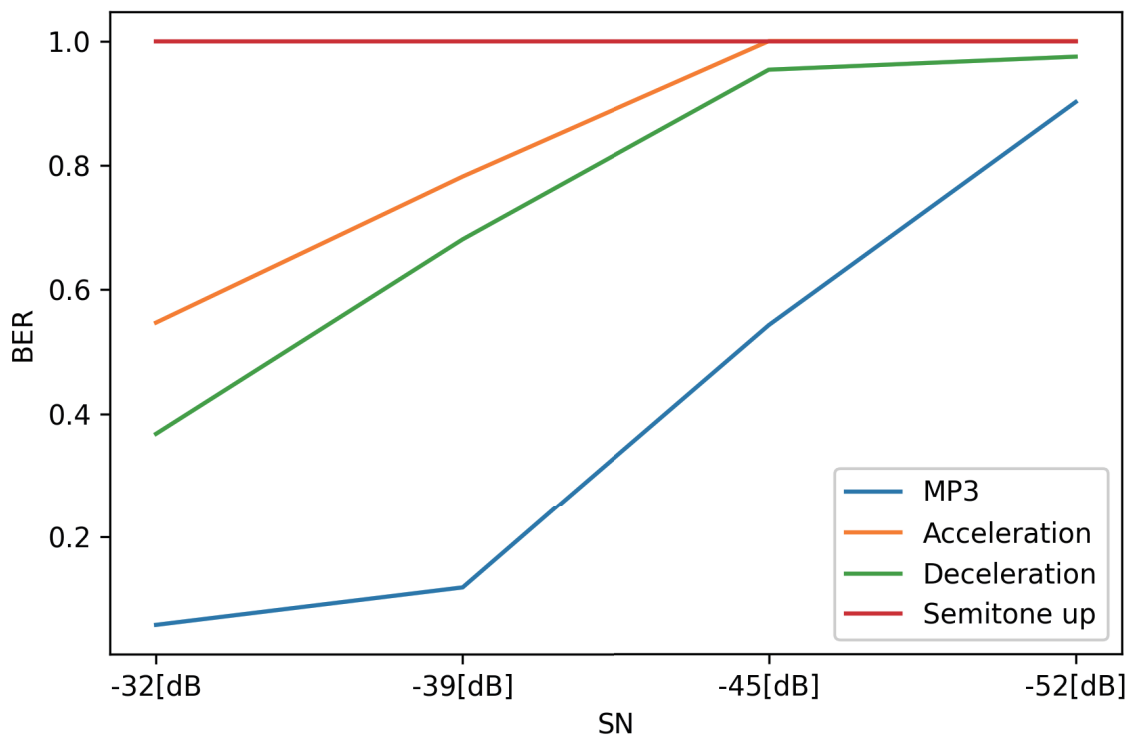


図 3.5.3 各種変換に対する攻撃耐性

より、再生時間が 91[s], 111[s] となったステゴデータを用いて秘匿情報の検出を行う。この時、音の高さを変えずに、楽曲の速さのみを変えている。

ピッチ変更も同様に、動画共有サイトなどの違法アップロードのように、ある程度コンテンツの価値を保ちながら攻撃が行われることを想定し、ステゴデータ全体を半音上げにしたステゴデータを用いて、秘匿情報の検出を行う。MP3 符号化、テンポ変更、ピッチ変更の結果を図 3.5.3 に示す [16]。

MP3 符号化による攻撃は、BER がわずかに上昇するだけに留まったため、この攻撃に対する耐性は十分であると考えられる。しかし、テンポ変更による攻撃は BER が著しく上昇し、ピッチ変更による攻撃はピークの検出ができなかった。

M 系列によって相関関数を計算する場合、正しい系列長で相関関数の計算を行えることが、高い相関特性を得る上で重要なポイントである。特に、本手法では、パルス信号の 1 サンプルに対して M 系列のサンプルとなるように信号の生成を行っている。そのため、テンポを変えることで、系列長が変化し、正しく相関関数が計算できなかったこと

が、BER を悪くした要因であると考えられる。また、加速の場合よりも減速の場合の方が BER が少し低かった要因として、加速の場合は M 系列の一部が欠損してしまうのに対し、減速の場合はある程度 M 系列そのものが保たれたためと考えられる。

系列長と同様に、M 系列によって相関関数を計算する場合、符号の正負が変化していないことも、高い相関特性を得る上で重要なポイントとなる。そのため、ピッチ変更により符号の正負情報が大きく変化し、正しく相関関数が計算できなかったことが、BER を悪くした要因であると考えられる。

## 第 4 章

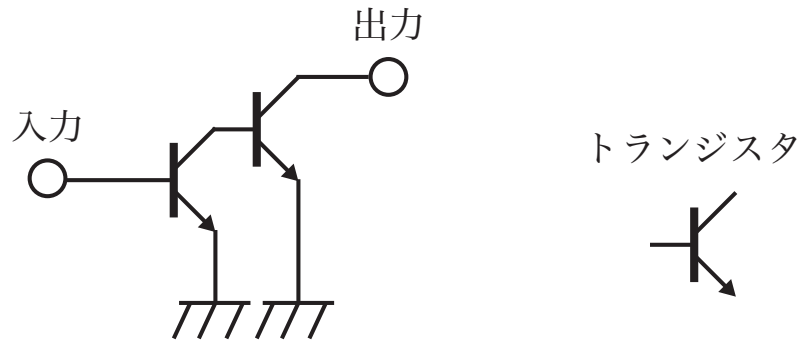
# 音響歪み効果による情報ハイディング

### 4.1 音響歪み効果

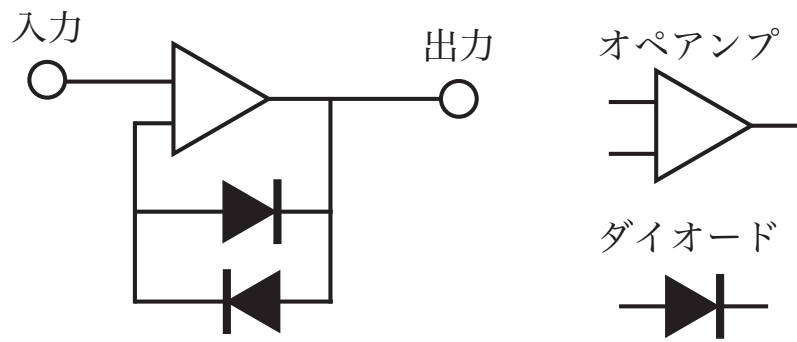
音響歪み効果は、楽曲作成や楽器演奏で広く用いられる音響効果である [19, 20]. この音響効果は、非線形信号処理を利用して、原信号に倍音を発生させることで音色を変化させることができる. 音響歪み効果は、元々は楽器音の信号を増幅回路によって増幅させる際に利得を大きくし、電圧飽和が発生したことで倍音が発生する現象である. 増幅回路においてこの歪み効果は目的と異なり、本来は避けたいものであるが、音楽の分野ではこの倍音効果による音色の変化が、音による表現の手法として注目され、現代の楽曲作成や楽器演奏では広く使われるようになった.

音響歪み効果が最もよく用いられる楽器はエレキギターであり、一般的にエレキギターにおける増幅回路はギターアンプである. ギターアンプで増幅による飽和現象を発生させる場合、出力される音量が非常に大きなものとなり、他の楽器とのバランスがとりにくいことや、音量に制限のある空間も存在することから、自由に音響歪み効果を扱うことは困難である. そのため、現代のギター演奏では、この飽和現象をオペアンプやトランジスタ、ダイオードといった半導体素子を用いて再現した非線形回路がギターアンプと合わせて用いられる. この非線形回路を用いた音響歪み機器は歪みエフェクタと呼ばれている.

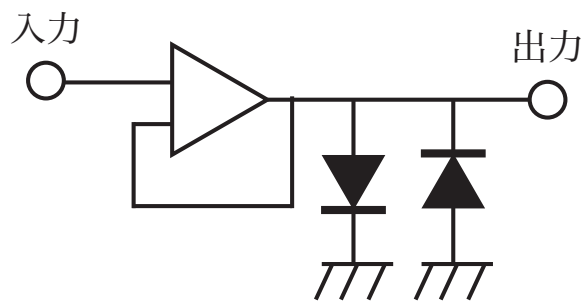
歪みエフェクタはその非線形回路部によって大きく 3 種類に分類される [21]. 3 種類の



(a) フェーズ系の歪みエフェクタ



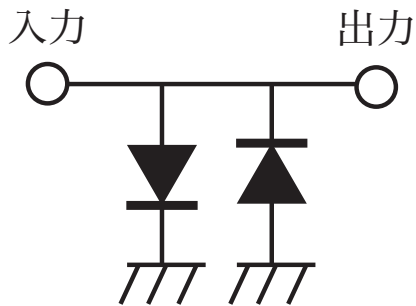
(b) オーバードライブ系の歪みエフェクタ



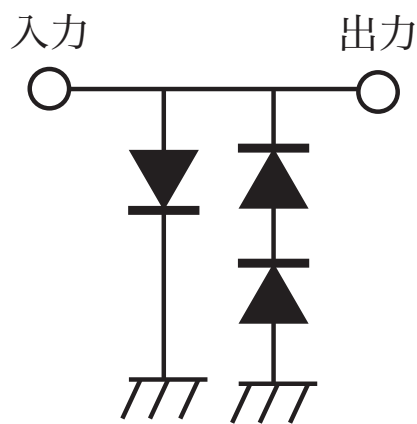
(c) デイストーション系の歪みエフェクタ

図 4.1.1 歪みエフェクタの主要な非線形回路部





(a) 対称クリッピング



(b) 非対称クリッピング

図 4.1.2 ダイオードクリッパ回路

簡略的な非線形回路部を図 4.1.1 に示す。図 4.1.1(a) のように、トランジスタによる増幅回路を複数繰り返すことによって、音響歪み効果が発生させる歪みエフェクタを、ファズ系の歪みエフェクタと呼ぶ。図 4.1.1(b) のように、信号の増幅回路に用いるフィードバック部に対してダイオードを用いて電圧に上限と下限を設定することによって増幅率を制御し、音響歪み効果が発生させる歪みエフェクタを、オーバードライブ系の歪みエフェクタと呼ぶ。図 4.1.1(c) のように、増幅後の出力信号に対してダイオードを用いて電圧に上限と下限を設定することによって、音響歪み効果が発生させる歪みエフェクタを、ディストーション系の歪みエフェクタと呼ぶ。

オーバードライブ系とディストーション系の歪みエフェクタで用いられるダイオードによって電圧を制限する回路を、ダイオードクリッパ回路と呼ぶ。歪みエフェクタにおける

ダイオードクリッパ回路は、ダイオードの順方向電圧を用いることで、整流（電圧の制限）を行う。正の電圧側の順方向電圧を  $V_p$ 、負の電圧側の順方向電圧を  $V_n$  とした場合、図 4.1.2(a) のように  $V_p = V_n$  となるようなダイオードクリッパ回路を対称クリッピングと呼び、図 4.1.2(b) のように  $V_p \neq V_n$  となるようなダイオードクリッパ回路を非対称クリッピングと呼ぶ。

## 4.2 埋め込み手法

本手法では、クリッピングによる歪み効果を利用して秘匿情報の埋め込みを行う。電子回路による歪み効果では、ダイオードの個数によってクリッピングレベルを決定しているが、本手法はデジタル信号処理による歪み効果であるため、クリッピングレベルは、正のクリッピングレベルと負のクリッピングレベルをそれぞれ任意の値で決定してよい。歪み処理が行われた音信号は、このクリッピングレベルによって信号の統計的特性が変化する。例えば、正のクリッピングレベルが負のクリッピングレベルよりも小さい場合、信号の平均値は、負の値になる。一方、正のクリッピングレベルが負のクリッピングレベルよりも大きい場合、信号の平均値は、正の値になる。情報ハイディングにおいてこの特徴は、0と1の2値情報を表現するために採用可能であると考えられる。

秘匿情報の埋め込みは、図 4.2.1(a,b) に示すように、エレキギターの信号（カバーデータ）に対して、一定のサンプル（フレーム）ごとに行われる。埋め込みは、埋め込むビットの0と1に応じて2種類のクリッピングパターンを切り替えることで行われる。

本手法は、クリッピングパターン切り替え時にノイズが発生する可能性があり、その例を図 4.2.2 に示す。図 4.2.2 では、クリッピングパターンを赤い線で示し、エレキギターの信号を灰色で示している。この図のように、エレキギター信号のアタックのような振幅値が大きくなるタイミングでクリッピングパターンの切り替えが発生した場合、振幅値が急激に変化し、歪み効果としては意図しないノイズのような音として聞こえてしまう可能性がある。楽器演奏において多くの楽器がその楽曲のテンポに合わせて演奏されるため、アタックの発生間隔も楽曲のテンポに依存している。そのため、フレームサイズ  $N$  は、カバーデータの BPM を用いて次式により決定することで、この問題が発生する可能性を軽減する。

$$N = \frac{60f_s}{T} \quad (4.2.1)$$

ここで、 $f_s$ [Hz] はサンプリング周波数であり、 $T$ [BPM] はカバーデータのテンポ、すなわち楽曲における1分間の拍数（1拍あたりを4分音符で分割した場合）である。

ビットレート  $R$ [bit/s] と埋め込み可能な最大情報量  $P$ [bit] は、以下の式で決定される。

$$R = \frac{T}{60} \quad (4.2.2)$$

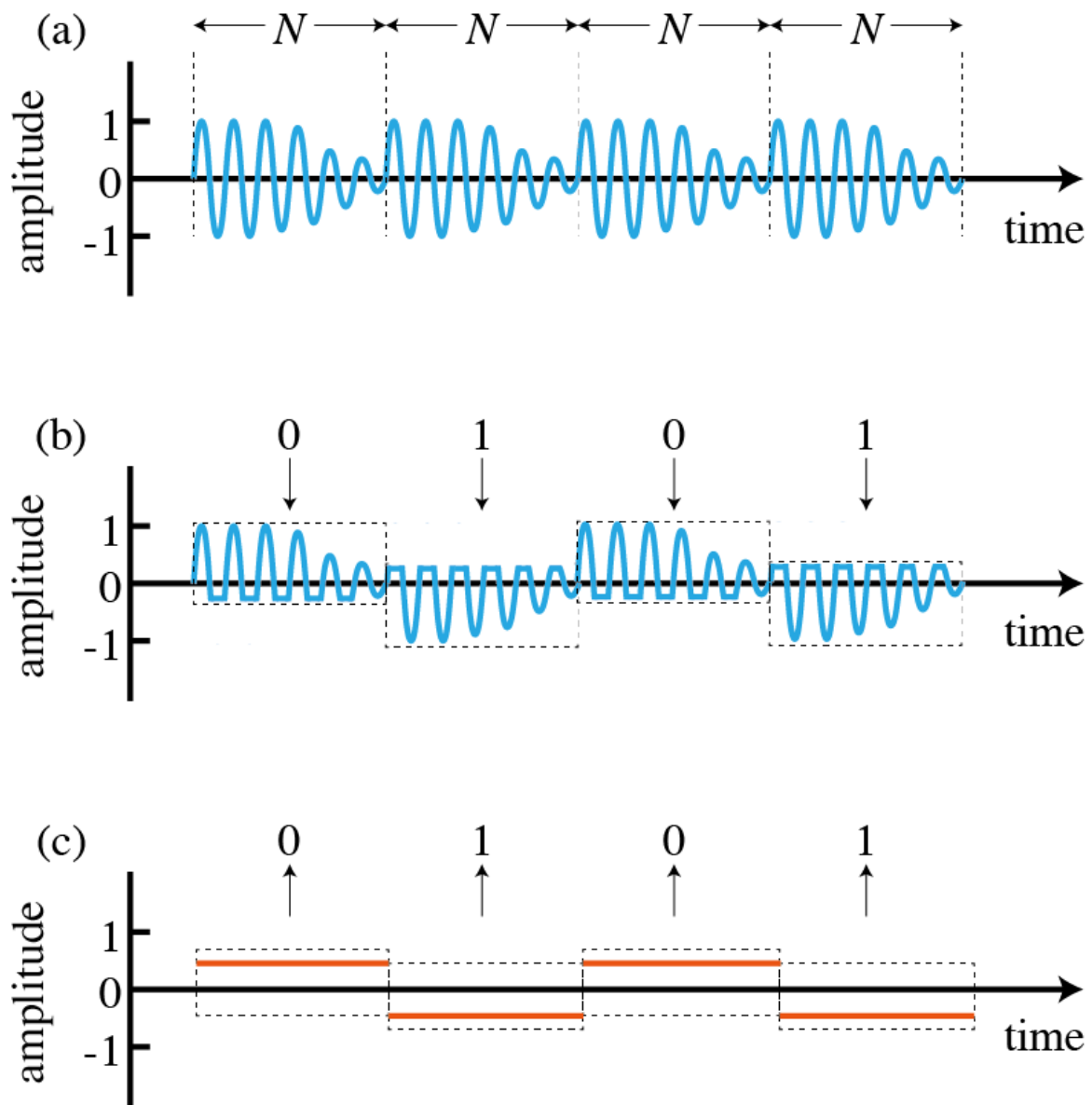


図 4.2.1 埋め込みの概要 (a) エレキギター信号 (カバーデータ) (b) クリッピングによる埋め込み (c) 情報の検出

$$P = RL \quad (4.2.3)$$

ここで、 $L[s]$  はステゴデータの長さ、すなわち楽曲の再生時間である。本手法の埋め込み可能な最大情報量は、楽曲のテンポと再生時間に依存する。例として、テンポが  $120[\text{BPM}]$  で再生時間が 4 分の楽曲では、 $480[\text{bit}]$  までの情報が埋め込める。このよう

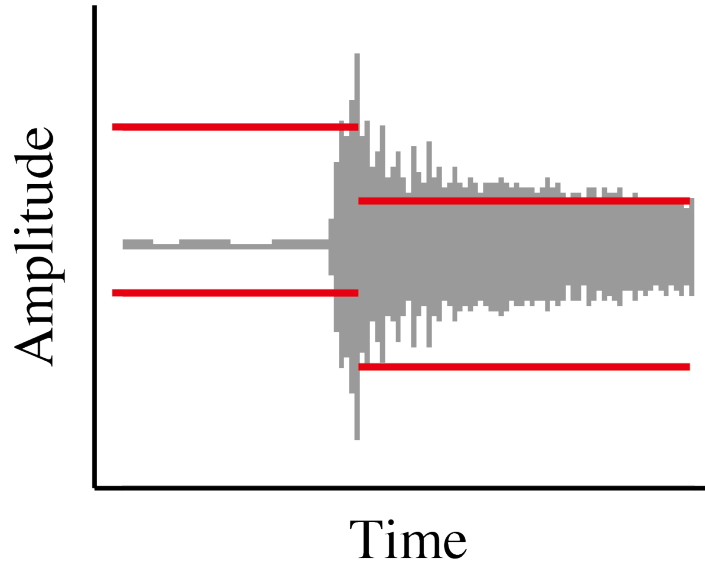


図 4.2.2 アタック時に発生するクリッピングパターンの切り替え例

に、本手法は埋め込み可能な情報量に制限があるため、著作権保護のように少ない情報量でも実現可能な応用先が適していると考えられる。

本研究で用いるクリッピングパターンは、Eichas らによるアナログの音響歪み機器をデジタル信号処理で再現するために提案された手法で用いられるクリッピングパターンに基づいている [20]. クリッピング回路による歪み効果を簡易的に再現する場合は、図 4.2.3 のようにしきい値を超えた振幅値をしきい値に均すようなクリッピングパターンによって再現が可能である。しかし、実際の歪み回路では、図 4.2.3 のような急激な変化は発生せず、飽和状態まで緩やかに変化することが知られており、サチュレーションと呼ばれる。このサチュレーションを再現するためには、 $\tanh$  のように緩やかな変化をする関数が適しており、Eichas らのクリッピングパターンでは、 $\tanh$  を用いて以下の式で表される。

$$m(x) = \begin{cases} \tanh(k_p) + \frac{1-\tanh^2(k_p)}{g_p} \tanh(g_p(x - k_p)) & (x > k_p) \\ \tanh(x) & (-k_n \leq x \leq k_p) \\ -\tanh(k_n) + \frac{1-\tanh^2(k_n)}{g_n} \tanh(g_n(x + k_n)) & (x < -k_n) \end{cases} \quad (4.2.4)$$

ここで、 $x$  は入力信号の振幅値であり、 $m(x)$  はクリッピングパターンによる出力である。また、 $k_p, k_n$  ( $k_p, k_n > 0$ ) はそれぞれ正のクリッピングレベルと負のクリッピングレベルであり、 $g_p, g_n$  ( $g_p, g_n > 1$ ) はそれぞれ正のクリッピングと負のクリッピングに対す

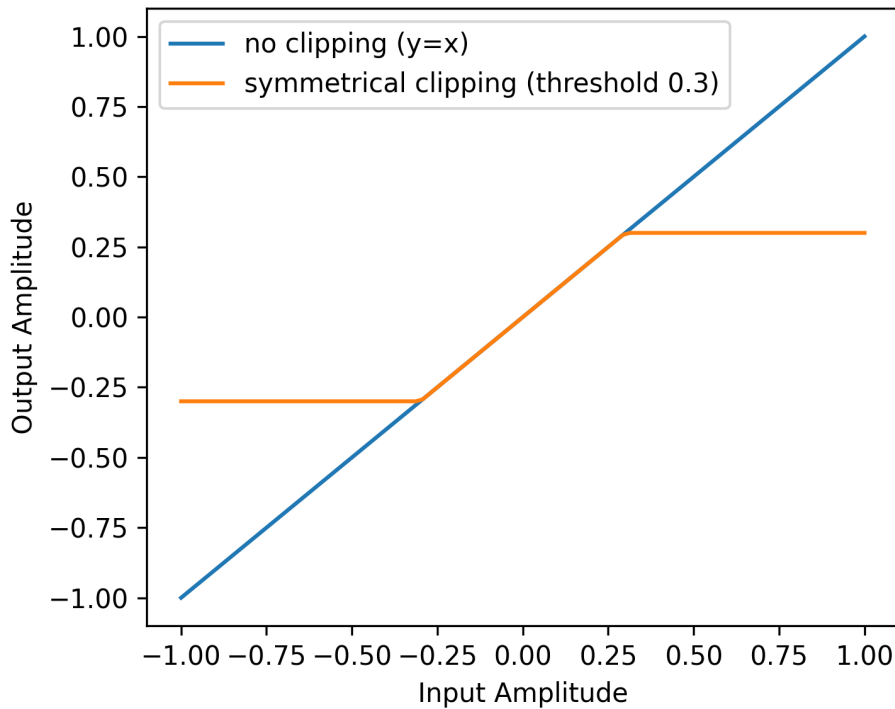


図 4.2.3 クリッピング回路の簡易的な再現

るサチュレーションを調整するパラメータである。

$g_p$  と  $g_n$  を変化させた場合のクリッピングパターンの変化を図 4.2.4 に示す。この図に示すように、様々な歪み効果の表現には、クリッピングレベルを調整するパラメータである  $k_p$  と  $k_n$  はもちろん、サチュレーションを調整するパラメータである  $g_p$  と  $g_n$  も重要である。

フレームサイズを  $N$ 、カバーデータを  $x(n)$  ( $0 \leq n < N$ )、埋め込み後の信号（ステゴデータ）を  $y(n)$  ( $0 \leq n < N$ ) として、埋め込むビット  $w(k)$  ( $0 < k \leq P$ ) が 0 の場合は、以下の式で埋め込み処理が行われる [14]。

$$y(n) = \begin{cases} \tanh(a_{p0}) + \frac{1 - \tanh^2(a_{p0})}{g_{p0}} \tanh(g_{p0}(x(n) - a_{p0})) & (x(n) > a_{p0}) \\ \tanh(x(n)) & (-a_{n0} \leq x(n) \leq a_{p0}) \\ -\tanh(a_{n0}) + \frac{1 - \tanh^2(a_{n0})}{g_{n0}} \tanh(g_{n0}(x(n) + a_{n0})) & (x(n) < -a_{n0}) \end{cases} \quad (4.2.5)$$

ここで、 $a_{p0}, a_{n0}$  ( $a_{p0}, a_{n0} > 0$ ) はそれぞれ正のクリッピングレベルと負のクリッピングレベルであり、 $g_{p0}, g_{n0}$  ( $g_{p0}, g_{n0} > 1$ ) はそれぞれ正のクリッピングと負のクリッピングに

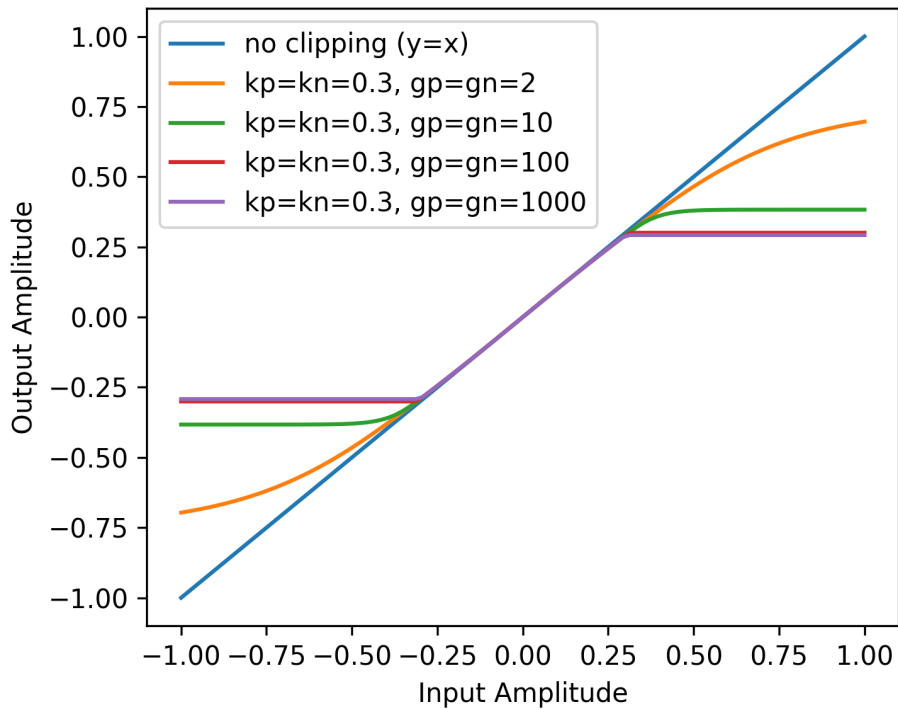


図 4.2.4 サチュレーションとクリッピングパターンの関係

対するサチュレーションを調整するパラメータである。この時、埋め込みを実現するためには、 $a_{p0} \geq a_{n0}$  である必要があり、 $a_{p0}$  と  $a_{n0}$  の差が大きいくほど平均値が大きな正の値になるため、秘匿情報の検出が容易となる（埋め込み強度が高い）。 $g_{p0}$  と  $g_{n0}$  は音色の調整といった側面が強いが、図 4.2.4 に示すように  $g_{n0} \geq 100$  となるような場合では設定したクリッピングレベルに近い値で振幅値がクリッピングされるため、正の振幅値と負の振幅値の差を発生させやすくなり、秘匿情報の検出が容易になりやすい。また、埋め込むビット  $w(k)$  ( $0 < k \leq P$ ) が 1 の場合は、以下の式で埋め込み処理が行われる [14]。

$$y(n) = \begin{cases} \tanh(a_{p1}) + \frac{1-\tanh^2(a_{p1})}{g_{p1}} \tanh(g_{p1}(x(n) - a_{p1})) & (x(n) > a_{p1}) \\ \tanh(x(n)) & (-a_{n1} \leq x(n) \leq a_{p1}) \\ -\tanh(a_{n1}) + \frac{1-\tanh^2(a_{n1})}{g_{n1}} \tanh(g_{n1}(x(n) + a_{n1})) & (x(n) < -a_{n1}) \end{cases} \quad (4.2.6)$$

$w(k)$  が 0 の場合と同様に、 $a_{p1}$  と  $a_{n1}$  はそれぞれ正のクリッピングレベルと負のクリッピングレベルであり、 $g_{p1}$  と  $g_{n1}$  はそれぞれ正のクリッピングと負のクリッピングに対するサチュレーションを調整するパラメータである。この時、埋め込みを実現するために

は、 $a_{p1} \leq a_{n1}$  である必要があり、 $a_{p1}$  と  $a_{n1}$  の差が大きいほど平均値が大きな負の値になるため、秘匿情報の検出が容易となる（埋め込み強度が高い）。 $g_{p1}$  と  $g_{n1}$  は音色の調整といった側面が強いが、図 4.2.4 に示すように  $g_{p1} \geq 100$  となるような場合では設定したクリッピングレベルに近い値で振幅値がクリッピングされるため、正の振幅値と負の振幅値の差を発生させやすくなり、秘匿情報の検出が容易になりやすい。

音色の調整として、クリッピングのしきい値を決めるパラメータ  $a_{p0}$ ,  $a_{n0}$ ,  $a_{p1}$ ,  $a_{n1}$  は 0 より大きく 1 以下を取る実数であり、0 に近づくほど歪み効果の強い（強く歪んだ）音を得られる。また、クリッピングの変化を調整するパラメータ  $g_{p0}$ ,  $g_{n0}$ ,  $g_{p1}$ ,  $g_{n1}$  は 0 以上の整数であり、値が大きいほど急激なクリッピングとなり、小さいほど緩やかなクリッピングとなる。

$a_{p0} = a_{n0}$  または、 $a_{p1} = a_{n1}$  の場合、対称クリッピングによる歪み効果になり、 $a_{p0} \neq a_{n0}$  または、 $a_{p1} \neq a_{n1}$  の場合、非対称クリッピングによる歪み効果になる。対称クリッピングによる歪み効果は振幅値に偏りが発生せず、楽器音の信号は通常、振幅値の平均値が 0 に近い値をとることから、 $a_{p0}$ ,  $a_{n0}$ ,  $a_{p1}$ ,  $a_{n1}$  をどのような値に設定した場合でも、振幅値の平均値は 0 に近い値をとり、この平均値の操作を人為的に行うことができない。そのため、対称クリッピングを用いて埋め込みを行う場合は、もう一方のクリッピングパターンに非対称クリッピングを採用する必要がある。すなわち、0 の埋め込みに対称クリッピングを採用した場合は、1 の埋め込みに非対称クリッピングを採用する必要がある。1 の埋め込みに対称クリッピングを採用した場合は、0 の埋め込みに非対称クリッピングを採用する必要がある。0 の埋め込みと 1 の埋め込みの両方で非対称クリッピングを採用することは問題なく行える。後述する検出手法により、秘匿情報の検出は、0 の埋め込みと 1 の埋め込みの両方で非対称クリッピングを採用した場合の方が容易に行える。カバーデータの統計的性質から、 $a_{p0} = a_{n1}$  かつ  $a_{n0} = a_{p1}$  となるような対称的なクリッピングのしきい値を持つ 2 種類の非対称クリッピングは、同様の倍音構成となることが知られており、音響歪み効果としての音色は非常に近いものとなる。埋め込む情報によって、フレームごとの音色が著しく変化するような音響歪み効果は自然な音響歪み効果とは言えず、アーティストにとって使いにくいものになると想定される。そのため、本稿では、今後特に記載のない限り、 $a_{p0} = a_{n1} = a_p$ ,  $a_{n0} = a_{p1} = a_n$  と表記する。



### 4.3 検出手法

図 4.2.1(c) に示すように、秘匿情報の検出は、ステゴデータからフレームごとに検出処理を行う。k フレーム目に埋め込まれているビット  $w(k)$  は、しきい値  $d_s$  を用いて以下の式で検出が行われる [14].

$$w(k) = \begin{cases} 0 & (s \geq d_s) \\ 1 & (s < d_s) \end{cases} \quad (4.3.1)$$

ここで、 $s$  はステゴデータ  $y(n)$  の k フレーム目の平均値を計算するために、以下の式で定義される。

$$s = \frac{1}{N} \sum_{n=(k-1)N}^{kN-1} y(n) \quad (4.3.2)$$

また、しきい値  $d_s$  は以下の式で定義される。

$$d_s = \frac{a_{p0} + a_{p1} - a_{n0} - a_{n1}}{2} \quad (4.3.3)$$

秘匿情報の埋め込みに、 $a_{p0} = a_{n1}$  かつ  $a_{n0} = a_{p1}$  のとなるような 2 種類の非対称クリッピングが用いられている場合、しきい値  $d_s$  は 0 となる。

## 4.4 主観評価実験

### 4.4.1 実験目的

本手法は、2種類の非対称クリッピング処理による音響歪み効果を用いることで埋め込み処理を行っている。この音響歪み効果において、2種類の非対称クリッピング処理を切り替えるタイミングで、ステゴデータの振幅値が切り替わる先のクリッピング処理のしきい値よりも大きい値だった場合、図 4.2.2 と同様の振幅値が大きく変化し音飛びのようなノイズが発生する可能性がある。そのため、このノイズが歪み効果として不自然なほど発生しているかを評価するために、主観評価実験を行った [14]。

### 4.4.2 実験条件

主観評価実験は、若く健常な聴覚を持つ7人の被験者を用いてダブルブラインドテスト (ABX テスト) を行う。このテストでは、被験者はフレーズ A とフレーズ B を試聴し、フレーズ X がフレーズ A またはフレーズ B のどちらであるかを回答してもらう。

この実験には表 4.4.1 に示した、フレーズやテンポの違う4種類のステゴデータを使用した。また、ステゴデータ全体で統一されている条件を表 4.4.2 に示す。

表 4.4.1 実験に用いたステゴデータ

フレーズ 1	速いテンポでのバックギターフレーズ (168[BPM])
フレーズ 2	速いテンポでのソロギターフレーズ (168[BPM])
フレーズ 3	遅いテンポでのバックギターフレーズ (75[BPM])
フレーズ 4	遅いテンポでのソロギターフレーズ (75[BPM])

表 4.4.2 ステゴデータ全体の条件

量子化ビット数	16[bit]
チャンネル数	1
サンプリング周波数	44.1[kHz]
フレーズの長さ	8 小節
秘匿情報のサイズ	32[bit]

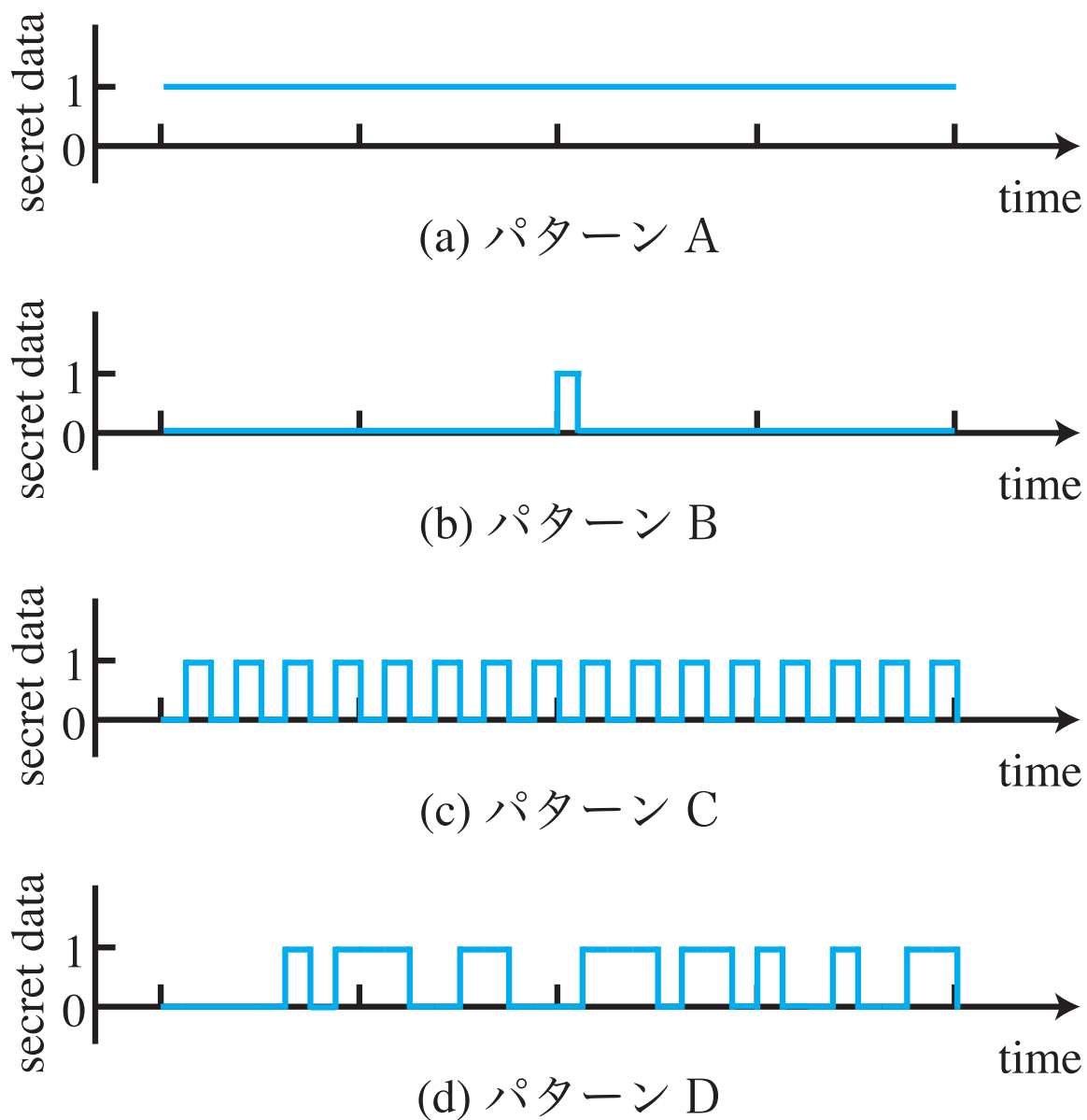


図 4.4.1 実験に用いた秘匿情報

全てのフレーズで図 4.4.1 に示した、秘匿情報が埋め込まれていないパターン (パターン A) と秘匿情報として 1[bit] だけ埋め込みが変わるパターン (パターン B), 0 と 1 が交互にくり返されるパターン (パターン C), 完全にランダムなパターン (パターン D) の 4 種類のステゴデータを生成した。パターン B は、秘匿情報が冗長であり、クリッピング処理の切り替えが少ない場合を想定している。パターン C は、クリッピング処理の切り

替えが多い場合を想定している。パターン D は、本手法の実用的な場合を想定している。本手法において、音響歪み効果と秘匿情報の埋め込みは切り離すことができない。そのため、パターン A では、便宜上全てのビットが 0 または 1 の秘匿情報が埋め込まれているステゴデータと同様のものとなる。

フレーズや秘匿情報の他にも、埋め込みの強さの違いによる影響を評価するために、 $a_p$  と  $a_n$  はそれぞれ  $\{0.9, 0.1\}$ ,  $\{0.8, 0.2\}$ ,  $\{0.7, 0.3\}$ ,  $\{0.6, 0.4\}$  となる 4 つの条件で埋め込みを行った。埋め込み強度は、 $a_n$  の値が小さいほど強い埋め込みである。また、全ての埋め込みで  $g = 100$  となるようにクリッピング処理を行った。

これらの条件をふまえて、64 個のステゴデータを作成した。この ABX テストでは、全てのフレーズと埋め込み強度で行われ、秘匿情報に関しては、フレーズ A またはフレーズ B のどちらかには、パターン A によるステゴデータを用いた。パターン A とパターン A の比較では、片方に全てのビットが 0 の場合のものを用いて、もう片方には全てのビットが 1 の場合のものを用いることで、 $a_{p0} = a_{n1}$  かつ  $a_{n0} = a_{p1}$  となるような 2 種類のクリッピングパターンに聴覚上の違いがあるのかを評価した。

また、被験者の視聴環境は、モニタ用ヘッドフォン (オーディオテクニカ製 ATHM20x) を用いて、各被験者にとって適切な音量で行われた。

#### 4.4.3 実験結果

ABX テストのフレーズ X を正しく識別できた割合を正答率とし、各フレーズの違いによる正答率を図 4.4.2 に、秘匿情報の違いによる正答率を図 4.4.3 示す [14]。正答率 0.5 に近いものは、クリッピング処理の切り替えによるノイズを被験者がうまく認識できていないことを示している。また、図の凡例は埋め込み強度であり、これは便宜上  $a_n$  の値のみを表示している。図 4.4.2 の結果では、秘匿情報の違いによる正答率で平均値を計算している。また、図 4.4.3 では、フレーズの違いによる正答率で平均値を計算している。これらの結果から、埋め込みの強さの違いによる、クリッピング処理の切り替えによるノイズの認識への影響は、非常に小さなものであると考えられる。また、この結果がフレーズに依存するものなのか、秘匿情報に依存するものなのかを確認するため、有意水準 0.05 で分散分析検定 (ANOVA 検定) を行った。その検定の結果 p 値はフレーズの違いで 0.92、秘匿情報の違いで 0.73 であり、フレーズや秘匿情報の違いで統計的な差が無いことが確

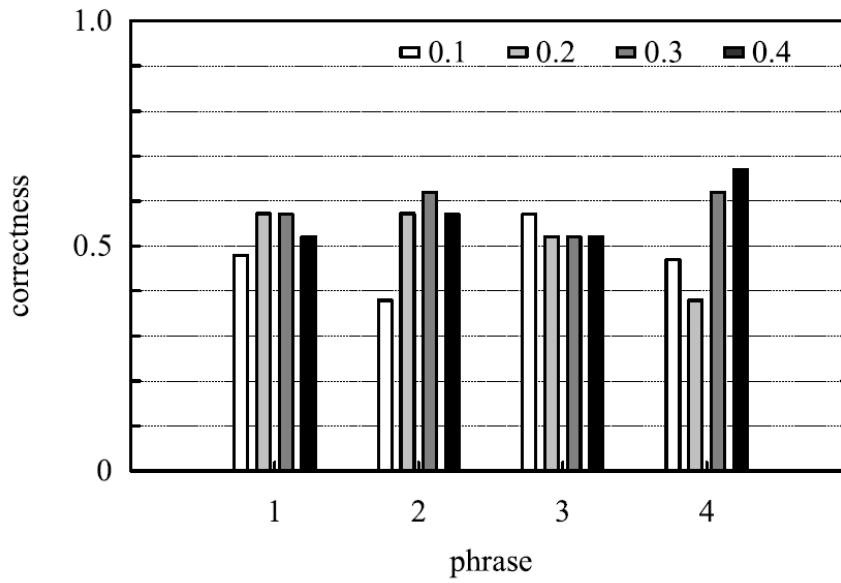


図 4.4.2 フレーズの違いによる ABX テストの正答率

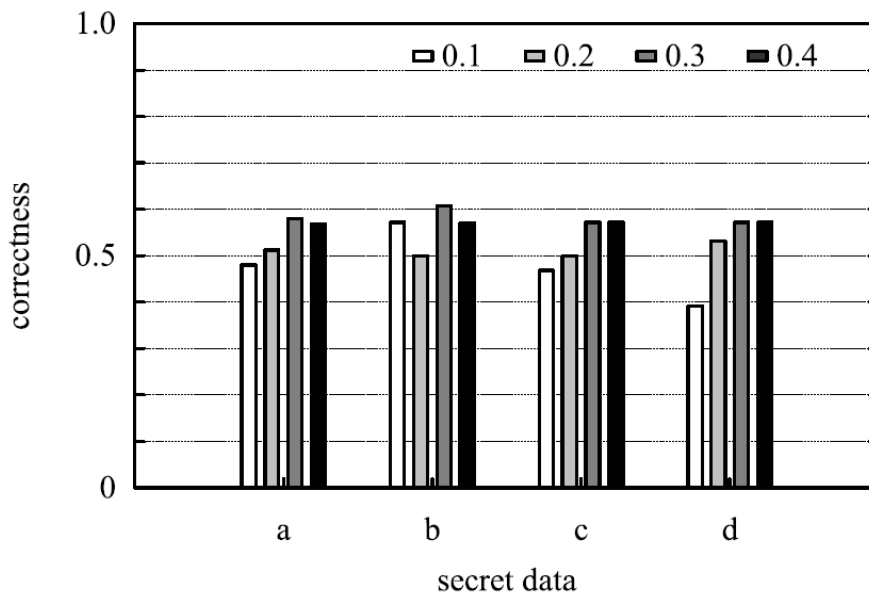


図 4.4.3 秘匿情報の違いによる ABX テストの正答率

認された。これらの結果から、クリッピング処理の切り替えによるノイズの発生はほとんど知覚されるものではないと考えられる。また、この主観評価実験はギター信号単体のものであり、他の楽器とミキシングされた場合は、より知覚しにくいものであると考えられる。このことから、本手法は、秘匿情報の有無やその内容に関わらず自然な音響歪み効果であると言える

## 4.5 客観評価実験

### 4.5.1 実験目的

音響データに対する情報ハイディング技術に求められる要素は、埋め込みによってコンテンツの品質をほとんど損なわないことと、ファイル変換や様々な攻撃を受けても検出精度が著しく低下しないことが求められる。本手法は、そのコンセプトから、埋め込みによってコンテンツの品質を損なわせることはないため、攻撃耐性の定量評価実験を行う [14].

### 4.5.2 実験条件

本実験では、BER を用いて評価を行う。本実験では、異なるフレーズからなる 116 のエレキギター信号を用い、表 4.5.1 に示した条件のステゴデータを生成した。このエレキギター信号は、市販の音楽制作ソフトウェアである Cubase のサウンドデータベースから取得した。また、これらの信号は POP's 音楽や Rock 音楽に用いられるようなコードのバックングフレーズやメロディのソロフレーズである。秘匿情報はランダムに生成されたデータである。

埋め込み強度に関する BER の変化を評価するため、 $a_p$  と  $a_n$  はそれぞれ  $\{0.9, 0.1\}$ ,  $\{0.8, 0.2\}$ ,  $\{0.7, 0.3\}$ ,  $\{0.6, 0.4\}$  となる 4 つの条件で埋め込みを行った。埋め込み強度は、 $a_n$  の値が小さいほど強い埋め込みである。また、 $g = 100$  となるように埋め込みを

表 4.5.1 ステゴデータの条件

量子化ビット数	16[bit]
チャンネル数	1
サンプリング周波数	44.1[kHz]
フレーズの長さ	8 小節
テンポ	62[BPM]~180[BPM]
ビットレート	1.03[bps]~3[bps]
秘匿情報のサイズ	32[bit]

行った。生成したステゴデータに対して、他の楽器とのミキシング、MP3 符号化、テンポ変更、ピッチ変更、ハイパスフィルタ (HPF) 攻撃の 5 種類の攻撃を行った。

### 4.5.3 ミキシング攻撃

本手法を用いて楽曲を構成する場合、埋め込みが行われている楽器信号と行われていない楽器信号をミキシングすることとなる。したがって、ミキシングを行い楽曲として完成させたものから秘匿情報を検出できる必要がある。ミキシングは、楽曲作成のプロセスとして避けられないものであり、悪意がある攻撃では無いが実用のためには、これに耐える必要がある。本実験では、ミキシングを行う他の楽器として、エレキベースの信号とドラムの信号を用いる。本実験では、埋め込みが行われていない楽器信号を、検出の精度を落とす一種のノイズ (攻撃) と考え、ノイズを強くした場合の検出精度の変化、すなわち、他の楽器信号の音量 (レベル) を上げた場合の BER の変化を測定する [14]。埋め込みを行った楽器信号の分散を  $s$ 、埋め込みが行われていない楽器信号全てを足し合わせた信号の分散を  $n$  としたとき、 $10 \log(s/n)$  で計算される SN 比が  $-30[\text{dB}]$ ,  $-20[\text{dB}]$ ,  $-10[\text{dB}]$ ,  $0[\text{dB}]$ ,  $10[\text{dB}]$  となるミキシングでステゴデータを作成した。SN 比は高くなるほどノイズの成分が少ない、すなわち、攻撃の影響が少ない条件である。

ミキシングに対する実験の結果を図 4.5.1 に示す [14]。BER は 0 に近づくほど誤りが少なく、秘匿情報の検出精度が高い、すなわち、攻撃耐性が高いことを示している。また、埋め込み強度は、便宜上  $a_n$  の値のみを表示している。図 4.5.1 は各条件に対する全フレーズでの BER の平均値と 95% 信頼区間を示している。この結果から、埋め込み強度に関わらず、SN 比が低くなると BER が高くなることが確認できる。また、SN 比に関わらず、埋め込み強度が弱くなると BER が高くなることが確認できる。SN 比よりも埋め込み強度の方が、BER に与える影響が大きいことが確認できる。

誤り訂正符号などの利用を考えた場合、BER が 0.1 以下程度であれば音響情報ハイディング技術としての実用を期待できる [22]。そのため、本手法は、 $a_p$  と  $a_n$  がそれぞれ  $\{0.9, 0.1\}$ ,  $\{0.8, 0.2\}$  となるようなある程度強い埋め込み強度での運用が必要となる。

本手法は、ステゴデータの振幅の平均値を計算することで秘匿情報の検出が行われる。そのため、ノイズとしての他の楽器信号が埋め込みが行われたエレキギターの信号に対して、平均値計算を行う必要があり、他の楽器信号が与える影響を考える必要がある。一般



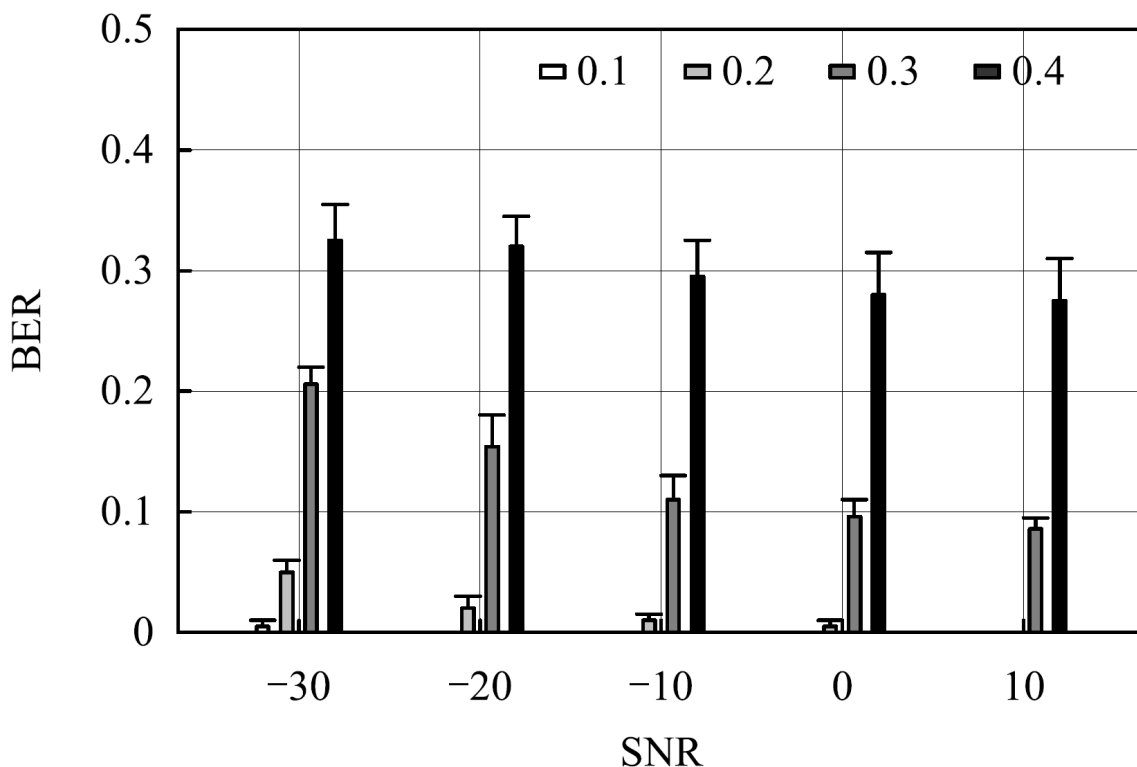


図 4.5.1 ミキシングに対する秘匿情報の検出精度

的な楽器信号は平均値が 0 に近いものになるため、本手法は本来ミキシングの影響を受けにくい。しかし、埋め込み強度が弱い場合は、正のクリッピングしきい値と負のクリッピングしきい値が絶対的に近く、平均値も 0 に近いものになる。これらのことから、SN 比よりも埋め込み強度の方が、BER に与える影響が大きくなったと考えられる。

以降の攻撃耐性の評価は、このミキシングされたステゴデータを用いて行う。

#### 4.5.4 MP3 符号化攻撃

本手法の基本的なファイルの形式変換攻撃に対する耐性を評価するため、wav ファイルのステゴデータから、ビットレート 128[kbps] による MP3 ファイルへの変換を行い、再度 wav ファイルに変換したものをを用いて、秘匿情報の検出を行った [14]。

MP3 符号化攻撃に対する実験の結果を図 4.5.2 に示す [14]。図 4.5.2 はミキシングに対する実験の結果と同様に、BER の平均値と 95% 信頼区間を示している。

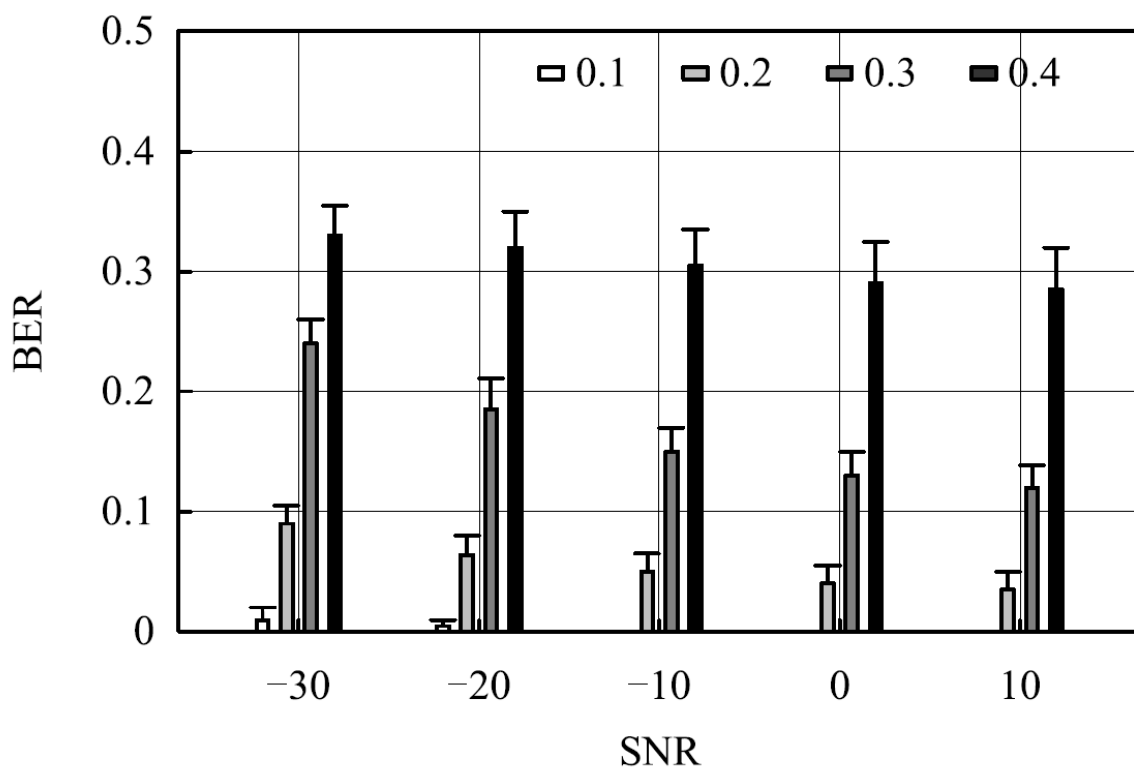


図 4.5.2 MP3 符号化攻撃に対する秘匿情報の検出精度

この結果から、全ての埋め込み強度において、BER が上がっており、秘匿情報の検出精度が悪くなっていることが確認できる。しかし、運用に実用的な埋め込み強度である、 $a_p$  と  $a_n$  がそれぞれ  $\{0.9, 0.1\}$ ,  $\{0.8, 0.2\}$  となるような場合には、BER を 0.1 程度に抑えることに成功しており、特に  $a_p$  と  $a_n$  がそれぞれ  $\{0.9, 0.1\}$  の場合には非常に低い BER を達成している。このことから、本手法の MP3 符号化攻撃への耐性は高いと考えられる。

#### 4.5.5 テンポ変更攻撃

本手法の悪意のある攻撃耐性を評価するため、基本的な攻撃の一つとしてテンポ変更による攻撃を行った [14]。テンポ変更攻撃は、ある程度コンテンツの価値を保ちながら攻撃が行われることを想定し、10% の加速と減速を行った。この時、音の高さを変えずに、楽曲の速さのみを変えている。

加速攻撃に対する実験の結果を図 4.5.3 に、減速攻撃に対する実験の結果を図 4.5.4 に

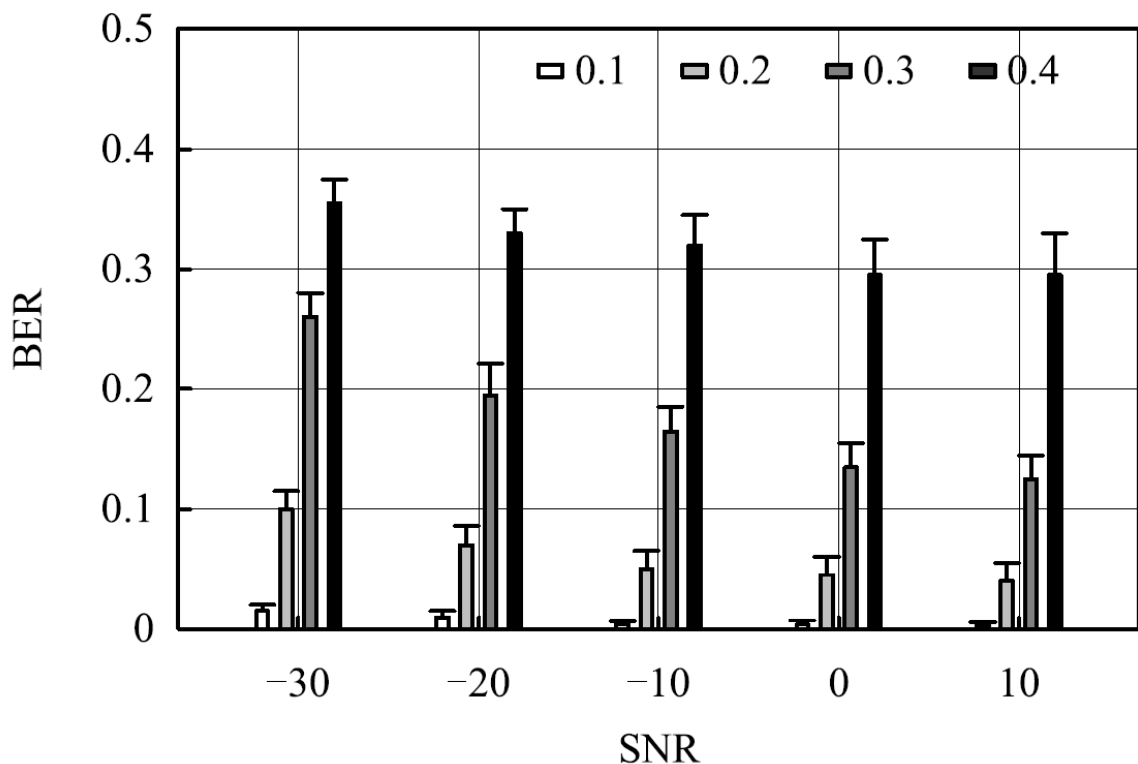


図 4.5.3 加速攻撃に対する秘匿情報の検出精度

示す [14]. 図 4.5.3, 4.5.4 は図 4.5.1 と同様に, BER の平均値と 95% 信頼区間を示している.

この結果から, MP3 符号化攻撃に対する実験の結果と同様に, 全ての埋め込み強度において, BER が上がっており, 秘匿情報の検出精度が悪くなっていることが確認できるが, 軽微なものであるまた, ある程度強い埋め込み強度の場合は, 低い BER を達成している. このことから, 本手法のテンポ変更攻撃への耐性は高いと考えられる.

#### 4.5.6 ピッチ変更攻撃

本手法の悪意のある攻撃耐性を評価するため, 基本的な攻撃の一つとしてピッチ変更による攻撃を行った [14]. ピッチ変更攻撃は, ある程度コンテンツの価値を保ちながら攻撃が行われることを想定し, 半音上げにとどめた.

ピッチ変更攻撃に対する実験の結果を図 4.5.5 に示す [14]. 図 4.5.5 は図 4.5.1 と同様

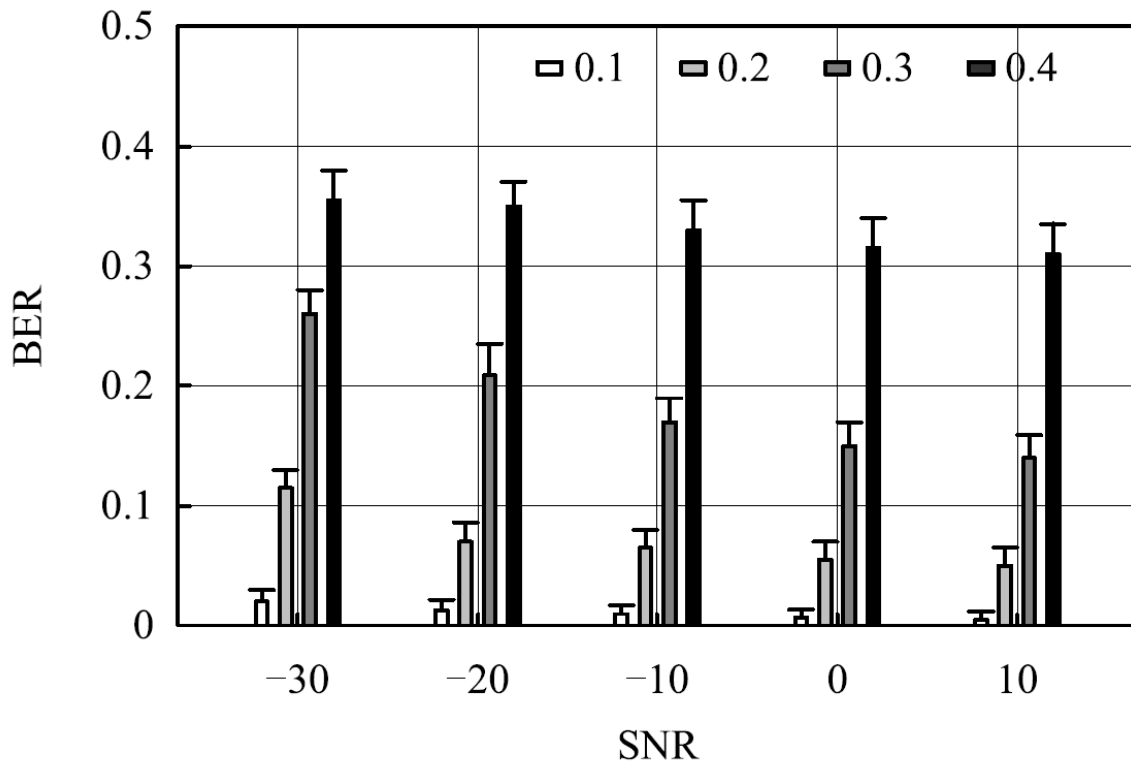


図 4.5.4 減速攻撃に対する秘匿情報の検出精度

に、BER の平均値と 95% 信頼区間を示している。

この結果から、MP3 符号化攻撃に対する実験の結果と同様に、全ての埋め込み強度において、BER が上がっており、秘匿情報の検出精度が悪くなっていることが確認できるが、軽微なものである。また、ある程度強い埋め込み強度の場合は、低い BER を達成している。このことから、本手法のピッチ変更攻撃への耐性は高いと考えられる。

#### 4.5.7 HPF 攻撃

本手法の悪意のある攻撃耐性を評価するため、攻撃の一つとして HPF による攻撃を行った [14]。HPF 攻撃は、ある程度コンテンツの価値を保ちながら攻撃が行われることを想定し、カットオフ周波数が 20[Hz] となる HPF 処理を行った。

HPF 攻撃に対する実験の結果を図 4.5.6 に示す [14]。図 4.5.6 は図 4.5.1 と同様に、BER の平均値と 95% 信頼区間を示している。

この結果から、本手法は、全ての埋め込み強度において、BER が著しく上がっており、

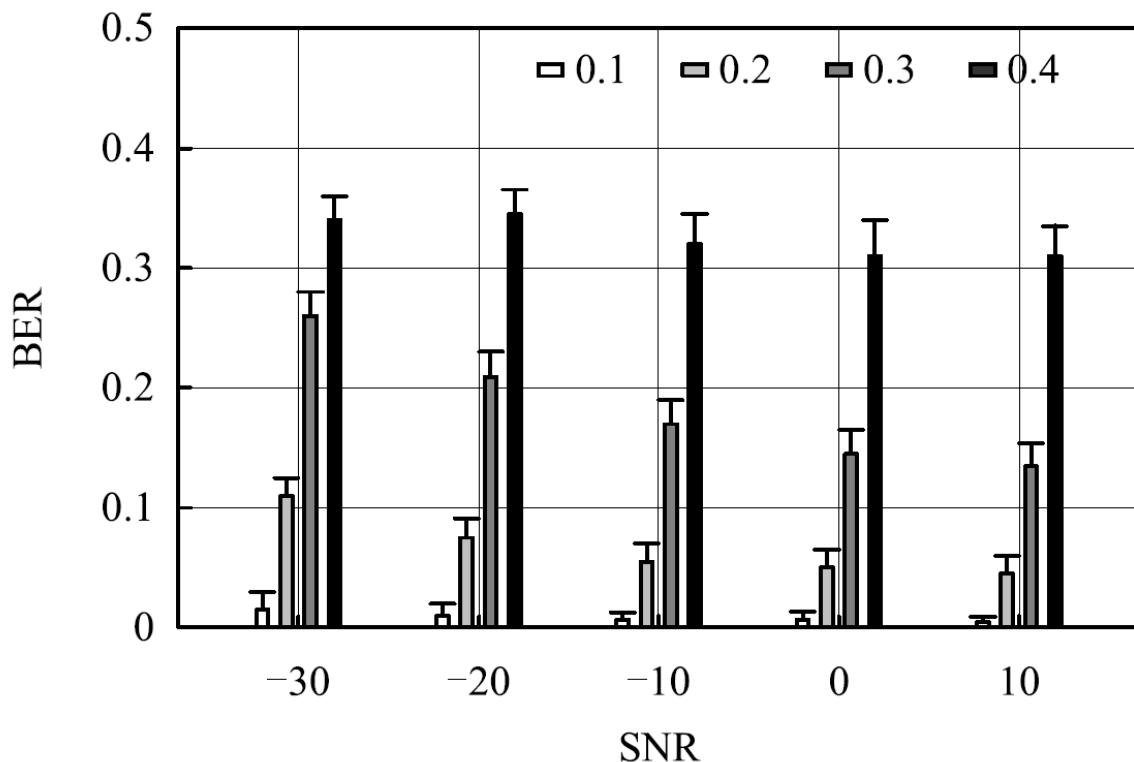


図 4.5.5 ピッチ変更攻撃に対する秘匿情報の検出精度

秘匿情報の検出ができていないことが確認できる。本手法は、音響歪み効果によって振幅の平均値に偏りを持たせることで埋め込みを行っている。この埋め込み処理を時間領域でフレームごとに行っているため、デジタル通信方式における振幅シフトキーイング (Amplitude Shift Keying: ASK) に近い処理であると考えられる。埋め込み処理のビットレートが 1.03[bps]~3[bps] であることから、ステゴデータ内における秘匿情報の要素は、低周波数成分に偏っていると考えられる。そのため、HPF 効果によって低周波数成分が減衰し、秘匿情報の検出ができなかったと考えられる。このことから、本手法の HPF 攻撃への耐性は無いと考えられる。

#### 4.5.8 考察

客観評価実験の結果で示したように、基本的なファイル変化や攻撃に対しては一定の耐性を確認できた。しかし、HPF 攻撃に対する攻撃耐性は無いことが確認された。本手法の埋め込みに関する性質から、HPF 攻撃は本手法に対する致命的な攻撃の一つであると

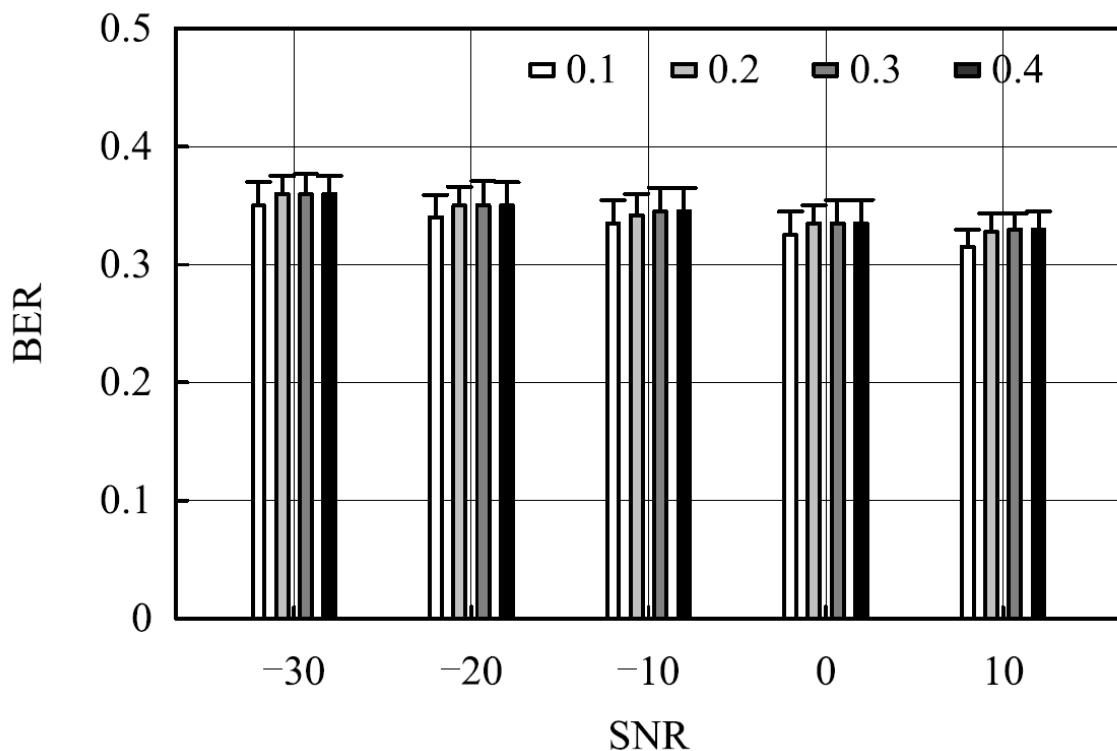


図 4.5.6 HPF 攻撃に対する秘匿情報の検出精度

考えられる。このことから、本手法は、HPF 攻撃に対する攻撃耐性を高める必要がある。

この問題に関する対策として、攻撃を受けた後のステゴデータの統計的な性質を抽出に利用することが考えられる。埋め込みを行ったギターの信号のみに対して HPF 攻撃を行った例を図 4.5.7 に示す。この図 4.5.7 から確認できるように、HPF 攻撃を受けた場合でも埋め込みを行ったギター信号の包絡線成分は、大きく変化していないことが確認できる。これは、本手法が ASK と異なり、音響歪み効果のクリッピングしきい値のみを切り替えているため、倍音成分という形で高周波数成分にも秘匿情報に関する要素が残っていたためと考えられる。このことから、ステゴデータの包絡線成分 (エンベロープ) を用いて秘匿情報の検出を行うことで、HPF 攻撃を受けた場合でも秘匿情報の検出が可能であると考えられる。

ステゴデータのエンベロープを用いて秘匿情報の検出を行うために、本手法の検出手法を図 4.5.8 のような検出手法へ変更する [14]。本手法では、エンベロープフィルタとして、窓サイズ 400 のピークフィルタを用いた。このフィルタでは、400 サンプル毎に正の振幅

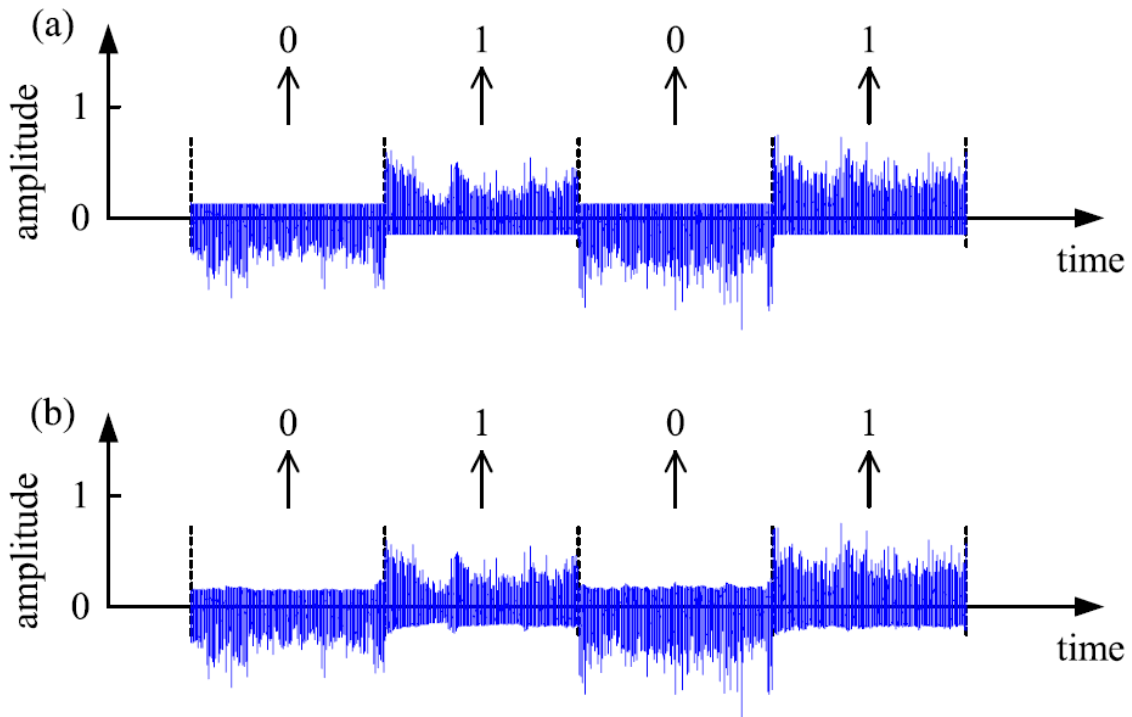


図 4.5.7 HPF 攻撃後の信号例 (a) 埋め込み後のギター信号 (b)HPF 攻撃後の信号

値側と負の振幅値側で信号のピークを検出する．図 4.5.8 の検出手法による， $k$  フレーム目に埋め込まれているビット  $w(k)$  は，しきい値  $d_{s'}$  を用いて以下の式で定義される [14]．

$$w(k) = \begin{cases} 0 & (s' \geq d_{s'}) \\ 1 & (s' < d_{s'}) \end{cases} \quad (4.5.1)$$

ここで， $d_{s'}$  は，対称的なクリッピングしきい値を持つ非対称クリッピングを用いて埋め込み処理を行う場合，0 に近い値を設定する．また， $s'$  はステゴデータのエンベロープから計算される信号  $e'_p(n)$  と  $e'_n(n)$  を用いて以下の式で定義される [14]．

$$s' = \sum_{n=0}^{N-1} (e'_p(n) + e'_n(n)) \quad (4.5.2)$$

$e'_p(n)$  と  $e'_n(n)$  は，ステゴデータにおける正の振幅値側のエンベロープ  $e_p(n)$  と負の振幅

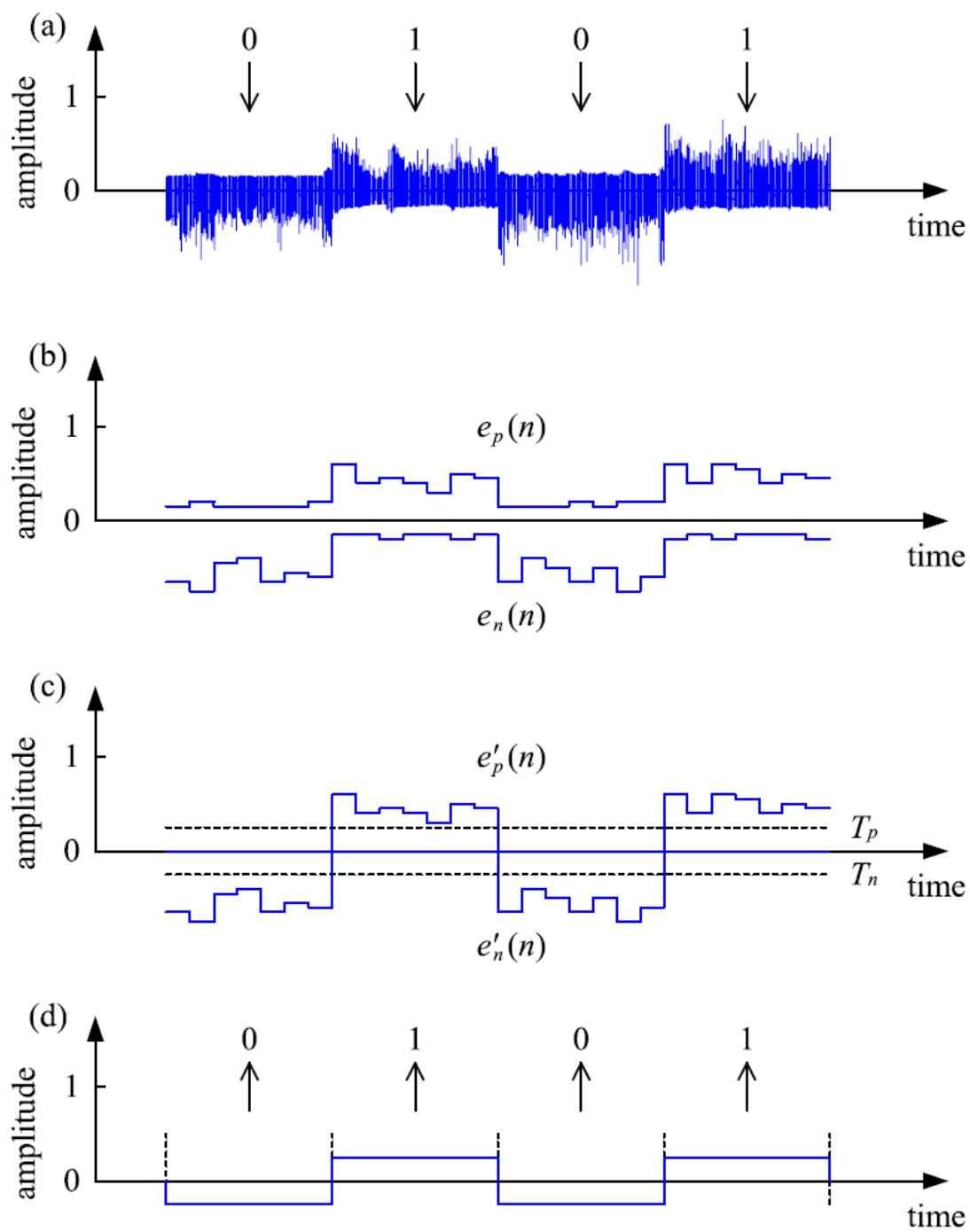


図 4.5.8 包絡線成分を用いた検出手法



値側のエンベロープ  $e_n(n)$  を用いて以下の式で定義される [14].

$$e'_p(n) = \begin{cases} e_p(n) & (e_p(n) > T_p) \\ 0 & (e_p(n) \leq T_p) \end{cases} \quad (4.5.3)$$

$$e'_n(n) = \begin{cases} 0 & (e_n(n) \geq T_n) \\ e_n(n) & (e_n(n) < T_n) \end{cases} \quad (4.5.4)$$

ここで  $T_p$  と  $T_n$  は、正の振幅値のエンベロープと負の振幅値のエンベロープにおいて、不要な成分を切り捨てるためのしきい値であり、これを適切な値に設定することで、検出精度が向上する。最適な  $T_p$  と  $T_n$  は埋め込まれた秘匿情報に依存するため、検出側で最適な  $T_p$  と  $T_n$  を設定するのは困難である。そのため、多くの場合で有用な方法として、以下の式で  $T_p$  と  $T_n$  を決定する。

$$T_p = \frac{1}{Lf_s} \sum_{n=0}^{Lf_s-1} e_p(n) \quad (4.5.5)$$

$$T_n = \frac{1}{Lf_s} \sum_{n=0}^{Lf_s-1} e_n(n) \quad (4.5.6)$$

ここで、 $Lf_s$  は楽曲の長さ (再生時間) とサンプリング周波数の積であり、楽曲の総サンプル数を示す。

この検出手法の有効性を評価するために、HPF 攻撃に対する攻撃耐性の評価実験と同じ条件で評価を行った [14]。その結果を図 4.5.9 に示す。この結果から、埋め込み強度が高いほど検出精度の改善が確認できるため、この検出手法は有効であると考えられる。しかし、 $a_p = 0.9$ ,  $a_n = 0.1$  の場合以外では、実用に耐えうる検出精度とは言い難い。音響歪み効果としての音色の自由度という観点から、少なくとも  $a_p = 0.8$ ,  $a_n = 0.2$  場合による検出精度の改善が求められる。

これまでの実験から、関連研究で攻撃耐性を持っていなかったピッチ変更や周波数フィルタ攻撃といった攻撃に対しても耐性を持っていることが確認できた。

村田らの関連研究では、ステゴデータに対して D/A 変換から A/D 変換を行う攻撃に対する攻撃耐性の評価が行われており、その結果はどの埋め込みパターンでも BER が 0.5 前後であったことが報告されている [9, 10]。D/A 変換から A/D 変換を行う攻撃は、アナログ伝送を経由することによる一種の周波数フィルタ攻撃であり、性質としては本実

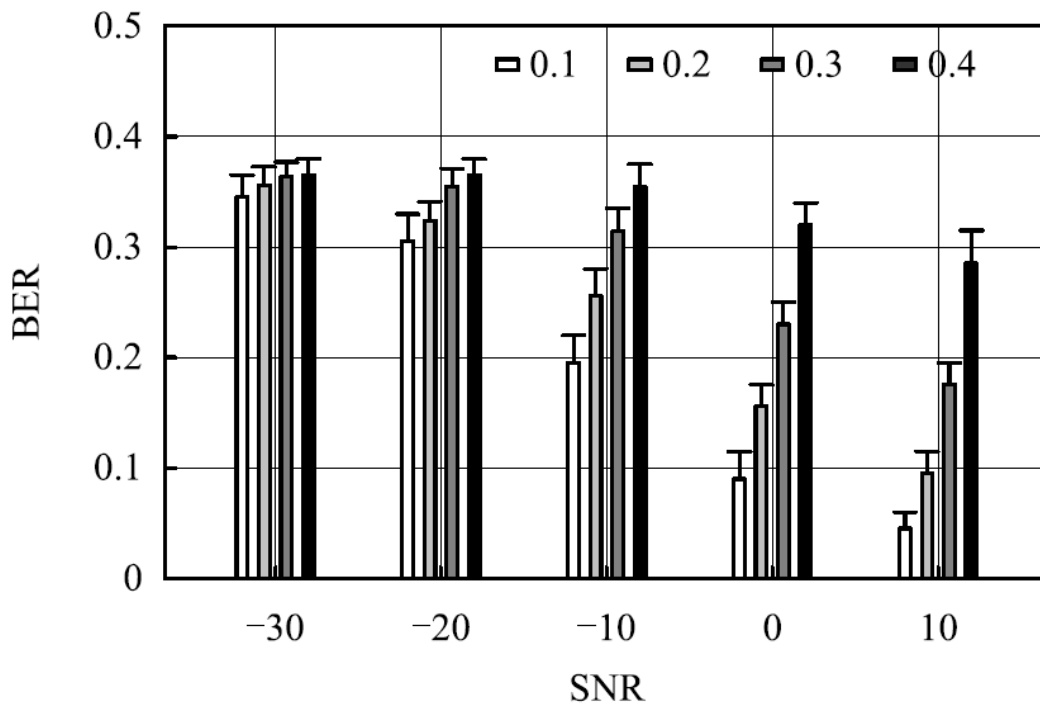


図 4.5.9 エンベロープを用いた検出結果

験で行われた HPF 攻撃に近い攻撃が行われる。このことから、この関連研究の手法は、周波数フィルタ攻撃に対して攻撃耐性を持っていない手法と言える。鎌田らの関連研究では、ピッチ変換攻撃に対する攻撃耐性の評価が行われており、その結果は BER が 0.487 であったことが報告されている [11]。このことから、この関連研究の手法は、ピッチ変換攻撃に対して攻撃耐性を持っていない手法と言える。

本手法は、HPF 攻撃に対してエンベロープを用いることで、埋め込み強度が最も強くミキシングレベルも強い場合は BER が 0.06 以下になることを示し、HPF 攻撃に対して攻撃耐性を持たせることに成功したと言える。これまでの実験結果から、ピッチ変換攻撃に対しては、図 4.5.5 から高い埋め込み強度の場合ではどのミキシング強度でも 0.02 以下の高い攻撃耐性を持っていることを示している。また、図 4.5.5 と図 4.5.1 から、他の埋め込み強度の場合でも BER が悪くなった原因のほとんどは他の楽器信号とのミキシングによるものであると考えられる。これらのことから、関連研究と比較して攻撃耐性の面で優れていると考えられる。

村田らの関連研究では、音楽として成立しているという点での消費者に対する品質劣化の問題は解決しているが、埋め込みによって品質劣化は発生しているため、利用者、すなわちアーティストやレコード会社にとっては自身の資産に大きな変更を加えることになり、心理的抵抗の課題を解決したとは言い難い。本手法は、楽曲作成時の音響機器の一つとして用いられることを想定しており、主観評価実験からも秘匿情報の有無やその内容に関わらず自然な音響歪み効果であると言えるため、利用者にとっての心理的抵抗を緩和している手法と言える。このことから、この関連研究と比べ、本手法は利用者にとっての心理的抵抗の少なさという面で優れていると考えられる。

鎌田らの関連研究では、ステゴデータとカバーデータの差分を取ったものから秘匿情報の検出を行うノンブラインド方式による検出であるため、検出側でカバーデータを保持しておく必要があり、著作権保護の音響情報ハイディング技術としては運用が困難である。本手法は、ステゴデータに対して平均値を計算するのみで秘匿情報の検出が行えるため、ブラインド方式による検出であり、この関連研究と比べ、著作権保護の音響情報ハイディング技術として適していると言える。

## 第 5 章

# 機械学習を用いた秘匿情報の検出

### 5.1 目的

近年、機械学習技術の発展に伴い、様々な情報ハイディング技術において、その検出手法として機械学習を応用することが検討されている。画像データに対する情報ハイディング技術では、Deep Neural Networks (DNN) が注目され、中でも Convolutional neural networks (CNN) に基づいた学習モデルを採用した検出手法が提案されている [23, 24, 25, 26]。音響データに対する情報ハイディング技術では、サポートベクタマシンを用いた手法が注目されている。

本研究の音響歪み効果を用いた音響情報ハイディング手法は、HPF 攻撃に対して弱いことが確認されている [14]。ステゴデータのエンベロープを用いてその平均値を計算するように検出手法を改善することで検出精度の改善が行われたが、これは最も強い埋め込み強度 ( $a_p = 0.9$ ,  $a_n = 0.1$ ) かつミキシングのレベルも高い場合、すなわち、強く歪んだギターの音でかつ楽曲全体の音量バランスの中でそのギターの音が占める割合が大きいような楽曲であることが求められる。このような場合、本手法の応用先が Rock などの音楽ジャンルに限られてしまうため、本手法が弱い埋め込み強度でも正常に秘匿情報の検出が可能となることで、弱く歪んだギターの音やギターの音が主張しないような楽曲でも利用可能とすることが求められる。そのため、検出精度の更なる改善が求められる。検出精度の目標として、BER を 0.1 以下に抑えることが求められる [22]。

本研究の検出手法は、ステゴデータの数ある特徴量の中から振幅値の平均値を用い、しきい値処理によって秘匿情報が 0 か 1 であるかを分類している分類機と考えることがで

きる。機械学習技術の代表的な応用先として、画像分類などの分類問題への応用が知られており、このような研究では、画像や音声などのデータに対して学習モデルが適した特徴量の選択を学習することで分類の精度を向上させている [27][28]。そのため、本研究の検出手法として機械学習技術の分類問題を応用することで、従来の振幅値の平均値という特徴量よりも適した特徴量の選択が可能となり、検出精度の向上が期待できる。

## 5.2 検出手法

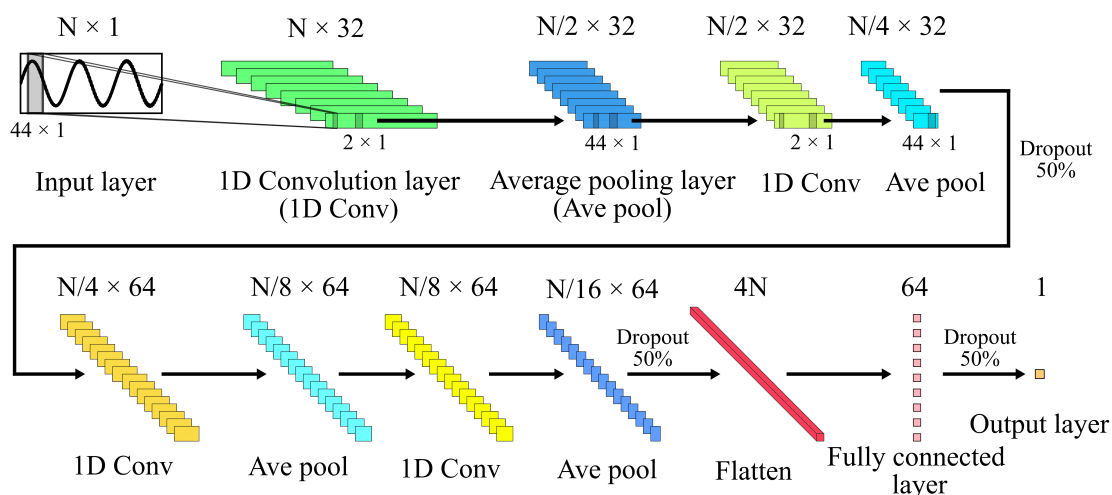


図 5.2.1 CNN 型のモデル

本研究における秘匿情報の埋め込みは 0 か 1 の 2 値で行われるため、秘匿情報の検出手法に機械学習技術の分類問題を応用する場合は、2 値分類問題として扱うことが適していると考えられる。そのため、本手法で用いる学習モデルは、ステゴデータを埋め込みが行われているフレーム毎に分割した 1 次元データをモデルの入力として与え、対応する 0 か 1 のビット情報を出力するような DNN モデルである。

本手法で用いる学習モデルを図 5.2.1 に示す [29]。図 5.2.1 のように、本手法の学習モデルは、2 値分類問題で用いられる一般的なネットワークモデルである、CNN を用いたモデルである。CNN 型のネットワークモデルは、画像データに対する情報ハイディングで注目されているため、本研究にも有用な可能性がある。この学習モデルの畳み込み層は表 5.2.1 の条件で計算が行われる。また、プーリング層は、これまでの検出手法が平均値計算によって行われているため、 $2 \times 1$  の平均値プーリングを用いる。全結合層は、活性化関数に ReLU を用いており、出力層は、2 値分類を行うために、シグモイド関数を活性化関数として用いている。

これまでの実験から、ステゴデータのエンベロープを用いることで、精度の向上が可能であることが判明している。このことから、本手法の学習データは、フレーム毎に分割し

表 5.2.1 学習モデルの畳み込み層

次元	1次元畳み込み
窓サイズ	44 × 1
パディング	ゼロパディング (0 埋め)
活性化関数	ReLU

た埋め込みが行われているギター信号に対して、正の振幅側と負の振幅側でエンベロープを検出し、それらを足し合わせた信号に対して min-max 正規化をかけた信号を入力データとし、対応するラベルデータを教師データとする。学習済みモデルを用いた秘匿情報の検出にも、フレーム毎に分割したステゴデータに対して、正の振幅側と負の振幅側でエンベロープを検出し、それらを足し合わせた信号に対して min-max 正規化をかけた信号を入力データとして用いる。その他のパラメータとして、最適化アルゴリズムは AdaDelta, 損失関数は Binary Cross Entropy, バッチサイズ 64 のミニバッチ学習であり、総エポック数は 128 である。また、学習データに対して HPF 攻撃などの各種攻撃を行う場合、その攻撃に対してのみ強くなる可能性が考えられるため、本手法の学習データには各種攻撃を行っていない。

本研究では、学習データとして、ギター音信号のデータベースである IDMT-SMT-Guitar から、エレキギターのコード演奏やメロディ演奏を含めた 10 分間の信号を作成し、この信号に対して音響歪み効果による音響情報ハイディングを行ったものを用いる [30]。この信号は 120[BPM] となるようにデータベースに対して変換を加えたものである。

## 5.3 客観評価実験

表 5.3.1 ステゴデータの条件

量子化ビット数	16[bit]
チャンネル数	1
サンプリング周波数	44.1[kHz]
フレーズの長さ	8 小節
テンポ	120[BPM]
ビットレート	2[bps]
秘匿情報のサイズ	32[bit]
ミキシングの強さ	-30[dB]/-20[dB]/-10[dB]/0[dB]/10[dB]
HPF 攻撃のカットオフ周波数	20[Hz]

本手法による検出精度を評価するために、客観評価実験を行った [29]。実験に用いたステゴデータの条件を表 5.3.1 に示す。基本的には前述の客観評価実験と同様に 116 のギターフレーズを用いて検出精度を計測するが、学習モデルの都合上入力データのサンプル数を揃える必要があるため、テンポは 120[BPM] で固定した。埋め込みの強さは、 $a_p = 0.8$ ,  $a_n = 0.2$ ,  $g = 100$  による埋め込みを用いた。

図 5.3.1 に実験の結果を示す。Conventional Technique は図 4.5.9 における  $a_p = 0.8$ ,  $a_n = 0.2$  の結果である。この結果から、従来の検出手法では、10[dB] のミキシングの場合でのみ実用に耐えうる可能性があったが、本手法を用いることで-10[dB] までは実用に耐えうる可能性を示した。このことから、他の MP3 符号化攻撃などの攻撃耐性と比べた場合は劣るものの、本手法による音響情報ハイディングは HPF 攻撃に対しても一定の攻撃耐性付加することに成功した。



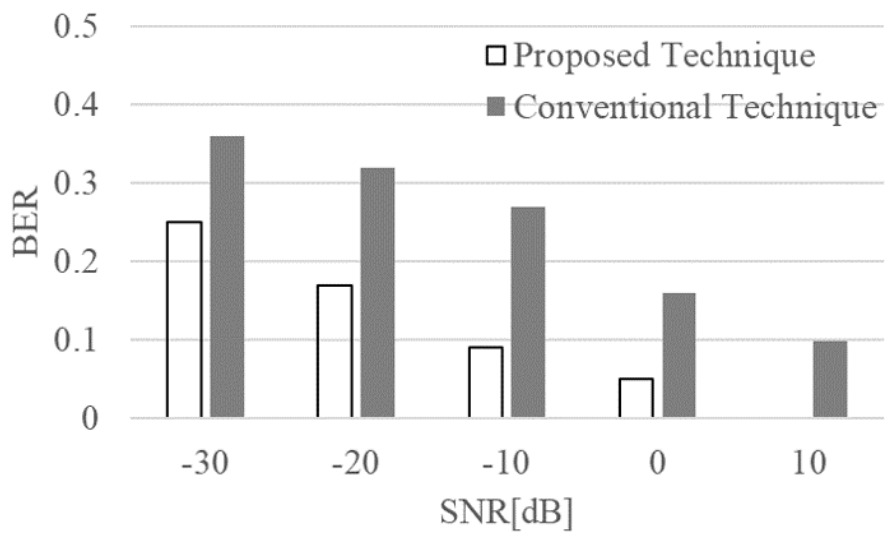


図 5.3.1 機械学習を用いた秘匿情報の検出結果

## 5.4 考察

前述の通り，CNN 型の学習モデルを用いることで，HPF 攻撃に対しても一定の攻撃耐性を得ることに成功した．しかし，機械学習には他にも様々なモデルが存在し，学習モデルの検討が十分とは言い難い．また，この実験では学習データに埋め込みが行われているギター信号のエンベロープを用いたが，機械学習ではこういった前処理的な部分も学習により最適化することが出来ると考えられる．

そのため，5つの学習モデルを用いて，学習モデルの違いによる精度の違いを評価する．用いる学習モデルは，CNN 型の学習モデル，全結合型の学習モデル，Long Short Term Memory(LSTM) 型の学習モデル，CNN-LSTM 型の学習モデル，パラレルな CNN 型の学習モデルの 5 種類である．

全結合型の学習モデルを図 5.4.1 に示す．この学習モデルは，Chandrasekhar らの手法で用いられた学習モデルに基づいている，それぞれが 200 個のニューロンを持つ 3 つの全結合層からなるネットワークモデルである [31]．全結合型のネットワークモデルは，最も基本の DNN モデルであるため，本研究にも有用な可能性がある．活性化関数には Rectified Linear Unit (ReLU) を用いている．出力層にはシグモイド関数を用いている点が Chandrasekhar らの手法と異なる．

LSTM 型の学習モデルを図 5.4.2 に示す．LSTM は最も一般的な再帰型ニューラルネットワークの一つであり，時系列データの学習に優れている [32]．LSTM 型のネットワークモデルは，様々な音響信号処理分野の研究で注目されているため，本研究にも有用な可能性がある [33, 34]．入力サンプル  $x(n)$  と直前の LSTM の状態  $h(n-1)$  と  $c(n-1)$

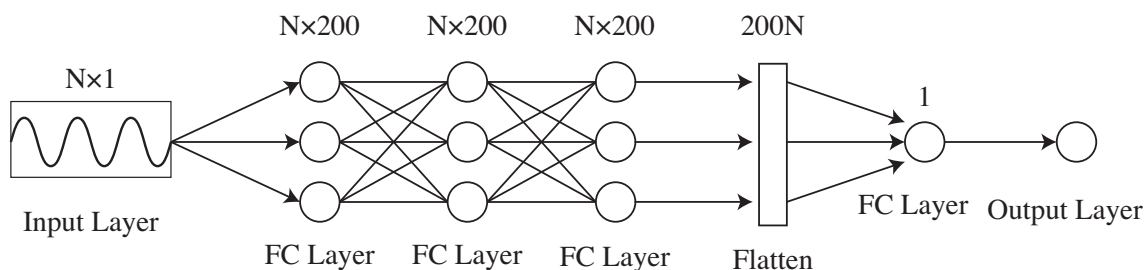


図 5.4.1 全結合型のモデル

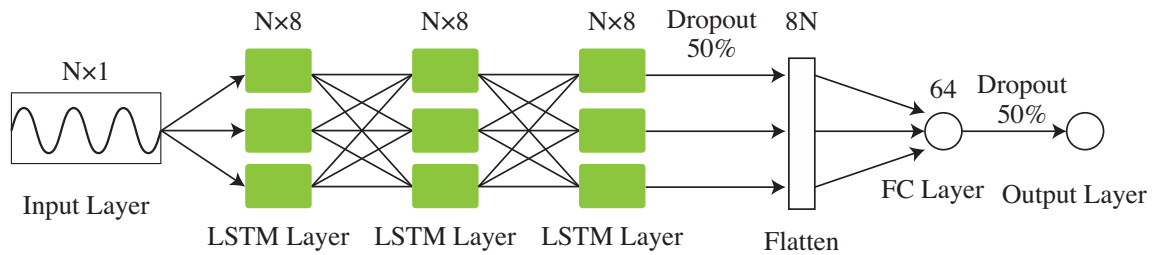


図 5.4.2 LSTM 型のモデル

を用いて、以下の式で LSTM 内部の状態を更新する。

$$\begin{aligned}
 i(n) &= \sigma(W_{xi}x(n) + b_{xi} + W_{hi}h(n) + b_{hi}) \\
 f(n) &= \sigma(W_{xf}x(n) + b_{xf} + W_{hf}h(n) + b_{hf}) \\
 \hat{c}(n) &= \tanh(W_{xc}x(n) + b_{xc} + W_{hc}h(n) + b_{hc}) \\
 o(n) &= \sigma(W_{xo}x(n) + b_{xo} + W_{ho}h(n) + b_{ho}) \\
 c(n) &= f(n)c(n-1) + i(n)\hat{c}(n) \\
 h(n) &= o(n)\tanh(c(n))
 \end{aligned}$$

ここで、各  $W$  と  $b$  は重みとバイアス、 $i(n)$  は入力ゲート、 $f(n)$  は忘却ゲート、 $\hat{c}(n)$  はユニットの出力とは別にフィードバックされるセルと呼ばれるメモリの状態であり、 $h(n)$  は LSTM ユニットの出力サンプルである。この学習モデルでは、それぞれが 8 個のユニットを持つ 3 層の LSTM 層とそれらの出力をまとめるための全結合層からなる。また、出力層にはシグモイド関数を活性化関数として用いている。

CNN-LSTM 型の学習モデルを図 5.4.3 に示す。近年では、時系列データの学習においても、LSTNet のように CNN と LSTM を組み合わせることで精度の向上を図ることが一般的なアプローチとなっている [35]。この学習モデルは、前述した CNN 型の学習モデルの前半部分と、LSTM 型の学習モデルの後半部分を組み合わせた学習モデルである。

パラレルな CNN 型の学習モデルを図 5.4.4 に示す。これまでの検出手法では、平均値計算という非常にシンプルな計算でもある程度の精度を達成している。このことから、平均値計算の結果も考慮しながら、これを補強するような形で CNN による出力を用いることで、精度の向上を図る。ここで、CNN の計算は、原理的に平均値の計算も行うことができることに注意する必要がある。しかし、CNN がそのような形のフィルタを学習するまでに、どれほどのデータ数やエポック数が必要か定かではなく、必ずしもそのような

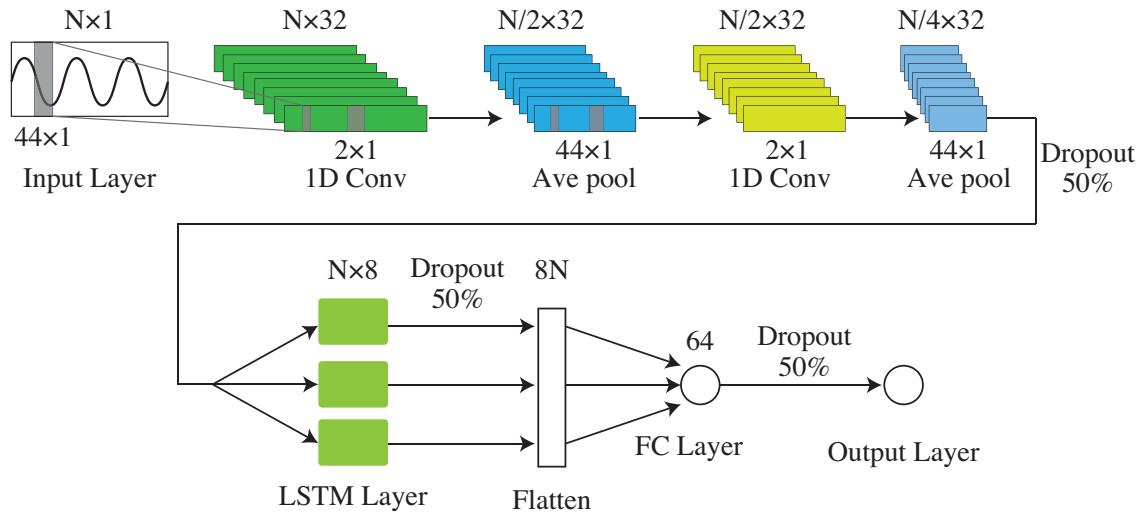


図 5.4.3 CNN-LSTM 型のモデル

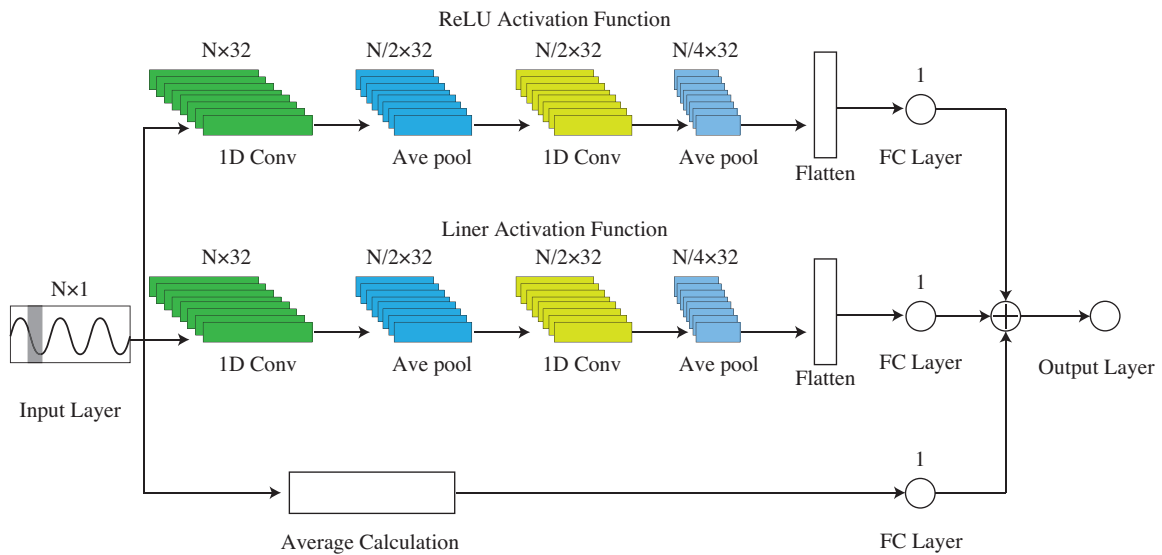


図 5.4.4 パラレルな CNN 型のモデル

フィルタになるとも限らないため、この学習モデルでは平均値計算のみを行うパスを用意している。他にも、この学習モデルは、活性化関数として ReLU を用いている CNN へのパスと活性化関数に何も用いていない (線形活性化) CNN へのパスからなり、それらの総和で出力が決定する。また、出力層にはシグモイド関数を用いている。

前述の実験とは異なり、本手法の学習データは、フレーム毎に分割した埋め込みが行われているギター信号に対して min-max 正規化をかけた信号を入力データとし、対応する

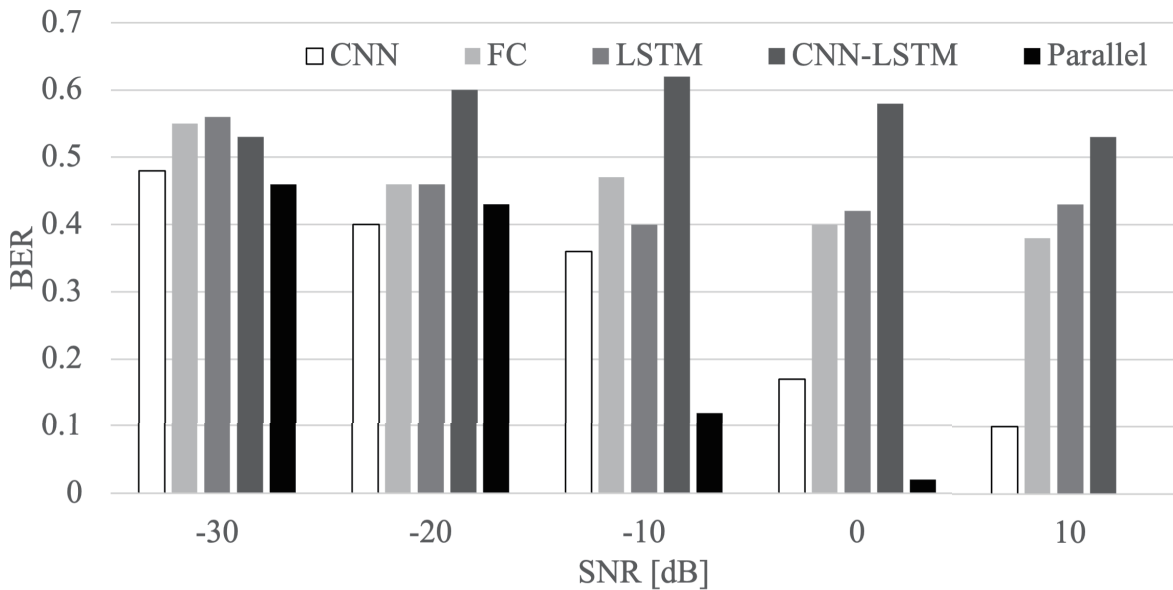


図 5.4.5 学習モデルの違いによる検出精度の変化

ラベルデータを教師データとした。学習済みモデルを用いた秘匿情報の検出にもフレーム毎に分割したステゴデータに対して min-max 正規化をかけた信号を入力データとして用いる。

これらの学習モデルを用いて、前述の客観評価実験と同様の実験を行った結果を図 5.4.5 に示す。CNN 型の学習モデルとパラレルな CNN 型の学習モデルが比較的良好な結果を得ていることが確認できる。この結果から、CNN を用いたモデルは本研究の埋め込み手法と相性がいいことが考えられる。CNN 型の学習モデルとパラレルな CNN 型の学習モデルの違いは平均値計算や非線形活性化を用いた CNN の有無であり、このことから、本手法における機械学習を用いた秘匿情報の検出において、線形変換の要素は重要な要素であったと考えられる。全結合層型の学習モデルは、CNN を用いたモデルと比べて、その計算内容から、大局的な特徴量の発見が困難であったと考えられる。CNN-LSTM 型の学習モデルと LSTM 型の学習モデルは、学習データとしてフレーム毎に分割したステゴデータを入力しているため、時間的な特徴から得られる情報が少なく、十分に機能しなかったと考えられる。特に、CNN-LSTM 型の学習モデルでは、学習モデルの前半は本手法の検出に良好な特性を示した CNN であるにもかかわらずこのような結果となったため、学習モデル後半の LSTM の影響が大きかったと考えられる。これは LSTM が出力層に近いことが理由の一つと考えられる。今回のようなフレーム毎に入力するような構造で

はなく、ステゴデータ全体を入力として非同期的に秘匿情報の検出を行う場合は、時間的な特徴も重要となってくると考えられるため、そういったケースでは LSTM を用いた学習モデルによる秘匿情報の検出も一考する価値がある。

CNN 型の学習モデルの結果から、エンベロープを用いないことで検出精度が落ちてしまうことを確認したが、0[dB] と 10[dB] の結果は、機械学習を用いずに、エンベロープと平均値計算のみで秘匿情報の検出を行った場合の結果と近いものが得られた。このことから、CNN 型の学習モデルがエンベロープフィルタのようなものを学習したものと考えられ、学習データの増強など今後の改善によっては図 5.3.1 のような結果が得られる可能性がある。

## 第6章

# まとめと今後の展望

電子透かし技術は、デジタルコンテンツに対して、コンテンツとしての価値や品質をある程度保ちつつコンテンツそのものに任意の秘匿情報を埋め込む技術である。この秘匿情報に著作権情報などを用いることで、著作権の主張が可能となる。近年では、電子透かし技術が著作権保護以外にも応用され、このような技術の総称として情報ハイディングと呼ばれている。

情報ハイディング技術では、秘匿情報の埋め込みによるコンテンツの品質劣化が重要な問題となる。この問題を解決するために、本研究では、秘匿情報で生成した信号を用いて、楽曲のオリジナルデータを作成するような手法を提案した。本手法は、秘匿情報で楽器音信号を生成し、この楽器音信号を楽曲内の様々な楽器音信号の一つとして用いることで、オリジナルデータの時点で既に秘匿情報が埋め込まれているデータが作成される。本手法の最大の利点は、楽曲を構成する楽器音信号そのものが秘匿情報から生成されていることで、埋め込みによるコンテンツの品質劣化を考慮する必要がなくなる点にある。本稿では、打楽器音の合成による情報ハイディング手法と音響歪み効果による情報ハイディング手法を提案した。

初めに、打楽器音の合成による情報ハイディング手法についてまとめる。打楽器音の中でもクローズしたハイハットシンバルの音は、再生時間が短いホワイトノイズのような性質を持ち、アタックの立ち上がりや減衰の仕方などの特殊な加工をあまり必要としない。そのため本手法では、ノイズ信号を加工して、クローズしたハイハットシンバルの音を再現した信号の生成を行う。このノイズ信号の生成に秘匿情報を用いることで、秘匿情報の

埋め込まれた打楽器音の信号を生成する。

秘匿情報は、ハイハットシンバルの1打に対して1[bit]の情報が埋め込まれる。秘匿情報が0の場合の埋め込み処理は、ノイズ信号に対して何も処理を行わず、秘匿情報が1の場合の埋め込み処理は、ノイズ信号の振幅値に対して-1をかける処理を行う。秘匿情報の検出は、このハイハットシンバル信号を用いて作成された楽曲データに対して、ハイハットシンバル信号の生成に用いたノイズ信号と同じ信号を用いて、相関関数の計算をすることで行われる。本手法は、ノイズ信号としてパルス信号として扱ったM系列を用いる。

本手法の攻撃耐性を評価するために、計算機シミュレーションによる定量評価実験を行った。この実験では、他の楽器音信号とのミキシング、MP3符号化、テンポの変更、ピッチの変更といった攻撃を加え、秘匿情報の検出精度を計測した。ミキシングに対する実験の結果、他の楽器音信号と比べ、ハイハットシンバル信号の音量が小さくなるほど、秘匿時情報の検出精度が低下することが判明した。また、MP3符号化には高い攻撃耐性を示したものの、テンポの変更とピッチの変更の攻撃には弱いことが判明した。本手法の秘匿情報の検出には相関関数を用いているため、テンポの変更により系列長が変化してしまったことや、ピッチの変更により、正負の符号が大きく変化してしまったことが、相関関数の計算に大きな影響を与えたと考えられる。

本手法は、ユーザー毎に異なるプリファードペアなM系列を割り振ることで、その相互相関特性から著作権の主張は可能であるため、秘匿情報が全て正しく検出されるかどうかは必ずしも重要ではない。本手法は、この点も考慮しながら、検出精度の改善が課題となる。

次に、音響歪み効果による情報ハイディング手法についてまとめる。音響歪み効果は、主にエレキギターの演奏で用いられる音響効果である。この音響効果を生じさせる代表的な手法として、ダイオードクリップが存在する。ダイオードクリップは、ダイオードの特性を利用して電圧のしきい値を設定し、これを越えた電圧をしきい値の電圧に整流するような電子回路であり、このような処理をクリッピングと呼ぶ。

本手法では、クリッピングによる音響歪み効果を利用して秘匿情報の埋め込みを行う。埋め込みは、非対称なしきい値を持つクリッピングによって、信号振幅の平均値に偏りを持たせることで行われる。埋め込み処理は、秘匿情報が0の場合に平均値が正の値になるように偏りを持たせ、秘匿情報が1の場合に平均値が負の値になるように偏りを持たせるようにクリッピングのしきい値を操作することで行われる。秘匿情報の検出は、各フレー



ム毎の平均値を計算することで行われる。

本手法の埋め込みによって、音響歪み効果として不自然な要素が発生しているのかを主観評価実験により評価した。実験の結果、埋め込みの有無や秘匿情報の違いによって主観的な変化を感じることがないことが判明し、本手法は音響歪み効果として不自然な要素が発生していないと考えられる。

また、本手法の攻撃耐性を評価するために、計算機シミュレーションによる客観評価実験を行った。この実験では、他の楽器音信号とのミキシング、MP3 符号化、テンポの変更、ピッチの変更、HPF 攻撃といった攻撃を加え、秘匿情報の検出精度を計測した。実験の結果、適切なミキシングと埋め込み強度では誤りがほとんどない検出が可能なが判明した。また、MP3 符号化、テンポの変更、ピッチの変更の攻撃に対しては高い攻撃耐性を示したが、HPF 攻撃に対しては秘匿情報の検出ができなかった。

本手法の性質上、低周波数成分が秘匿情報の検出において重要であり、HPF 攻撃に対して非常に弱くなったと考えられる。しかし、音響歪み効果の性質により、高周波成分にも秘匿情報の成分が見られるため、検出手法の改善によってこの問題を解決できると考えられる。エンベロープを用いて秘匿情報の検出を行った結果、検出精度の改善が確認され、特に埋め込み強度が最も強い場合では実用に耐えうるほど大きく改善された。

音響情報ハイディングの関連研究では、テンポの変更や周波数フィルタ攻撃を受けた場合に、秘匿情報の検出が不可能となることが報告されている。本手法はこれらの攻撃を受けた場合でも秘匿情報の検出が可能であることから、関連研究よりも優れた攻撃耐性を持っていると言える。

本手法の HPF 攻撃に対して弱いという問題点は、エンベロープを用いることで改善されたものの、平均値を計算してしきい値を用いた判定という点は変わっていません、改善の余地を残している。本手法の秘匿情報の検出は、ステゴデータに対して埋め込まれている秘匿情報が 0 か 1 を分類するような、一種の 2 値分類問題として考えることができる。そのため、本手法の検出手法に機械学習による 2 値分類問題の手法を応用することで、適切な特徴量の選択が可能となり、精度の向上が期待できる。

本手法では、秘匿情報の検出を行う学習モデルとして、CNN を用いたモデルを提案した。学習は教師あり学習で行われ、学習データは、フレーム毎に分割された埋め込みが行われているギター信号のエンベロープであり、教師データは、対応する埋め込んだビット情報である。

本手法による検出精度を評価するために、計算機シミュレーションによる客観評価実験を行った。実験の結果、本手法は、平均値計算としきい値処理で行われていた検出手法よりも高い検出精度を達成した。

機械学習分野の研究として本手法の検出手法を考えた場合、本手法で用いているようなエンベロープの検出といった前処理も含めて最適化されるような学習モデルであることが求められる。そのため、CNN 以外にも LSTM などを用いた様々なパターンの学習モデルを用いて検出精度の比較も行った。

5種類の学習モデルを持ちいて比較実験を行った結果、非線形活性化を用いた CNN と用いていない CNN、線形変換と平均値計算からなるネットワークの3つのネットワークの出力を組み合わせた学習モデルが最も良い検出精度となった。このことから、本手法の検出手法においては、畳み込み計算と相性が良く、非線形活性化の要素以外にも線形変換の要素が重要になったものと考えられる。

本手法の最大の懸念は、その限界である。本手法は、歪んだ音が存在しない部分やそもそも無音であるような部分がある場合、原理的に埋め込みが困難である。また、アーティストが演奏区間によって歪み効果をかけたり、かけなかったりすることも現実的なケースとして考えることができ、演奏全体を通して同じ強さの歪みがかかるとは限らない。これらことから、本手法を実用化するためには、音響イベント検出のような音検知アルゴリズムが必要になる可能性がある。また、本手法は、音楽制作において歪んだ音が必要とされる場合に有効である可能性があるが、歪んだ音は必ずしも全ての音楽ジャンルに適用できるわけではない。

今後の展望として、前述した点に留意しつつ本手法の応用先を慎重に検討していくと共に、検出手法の更なる精度向上に取り組んでいくことが挙げられる。これにより、弱い埋め込み強度による運用が可能となるため、コンプレサーのような音響効果として用いることが可能となり、様々な音楽ジャンルや様々な楽器への応用が可能になっていくと考えられる。本稿で提案した音響情報ハイディングの手法は、打楽器音に対するものと音響歪み効果に対するものであるが、他の音響効果や楽器音に対しての音響情報ハイディング手法が検討されていくことで、それらを組み合わせた多チャンネルでの音響情報ハイディングが可能となる可能性がある。

# 謝辞

本研究は、著者が北海道大学工学部情報エレクトロニクス学科メディアネットワークコース在籍時から開始し、北海道大学情報科学院情報科学専攻メディアネットワークコースの在籍中も引き続き行ったものである。多大なるご指導、ご鞭撻を頂きました土橋土宜典教授に心より感謝致します。本論文をまとめるにあたり、副査をお引き受けいただいた北海道大学情報科学研究院 メディアダイナミクス研究室 長谷山美紀教授、北海道大学情報科学研究院 言語メディア学研究室 荒木健治教授、北海道大学情報科学研究院 メディア創生学研究室 坂本雄児教授に深く感謝致します。本研究の学術的背景のご指導および本研究の評価、方針に関して多大なるご助力を頂いた北海道大学情報科学研究院 情報メディア環境学研究室の青木直史 助教および北海道大学 山本強名誉教授に深く感謝致します。本研究の実験に関して多大なるご助力を頂いた国立東京工業高等専門学校情報工学科 情報通信研究室の小嶋徹也 教授に深く感謝致します。博士後期課程在籍中の研究生活においては独立行政法人 日本学術振興会の特別研究員制度によるご支援を頂きました。ここに深く感謝致します。本研究の一部は、JSPS 科研費 JP20J20564 の助成を受けたものです。ここに深く感謝致します。著者の研究期間中、共同研究にご協力いただいた北海道大学情報科学院情報科学研究科メディアネットワーク専攻情報メディア環境学研究室の皆様へ感謝致します。本研究に有益なご助言をいただき、著者の研究生活も支えてくださった旧友の當田峻也氏に感謝致します。最後に、研究生活を支えてくださった家族に多大なるご支援を頂きましたことに深い感謝の意を表し謝辞とさせていただきます。

# 参考文献

- [1] 日本音響学会編. 音響情報ハイディング技術. コロナ社, 東京都, 日本, 2018.
- [2] Kondo Kazuhiro. *Multimedia Information Hiding Technologies and Methodologies for Controlling Data*. Information Science Reference, Hershey, PA, USA, 2013.
- [3] F.A.P. Petitcolas, R.J. Anderson, and M.G. Kuhn. Information hiding-a survey. *Proceedings of the IEEE*, Vol. 87, No. 7, pp. 1062–1078, 1999.
- [4] 画像電子学会編. 電子透かし技術：デジタルコンテンツのセキュリティ. 東京電機大学, 東京都, 日本, 2004.
- [5] 東小野. 電子透かしとコンテンツ保護. オーム社, 東京都, 日本, 2001.
- [6] 社団法人日本音楽著作権協会 (JASRAC). 音楽電子透かし 4 社を技術認定し、ネット上の違法利用の発生防止に活用. [https://www.jasrac.or.jp/release/01/10\\_2.html](https://www.jasrac.or.jp/release/01/10_2.html), Accessed: 2022-11-14.
- [7] 日経 XTECK. 次世代バリアフリー、「おもてなし」への発想転換. <https://xtech.nikkei.com/dm/atcl/column/15/111000020/111200003/>, Accessed: 2022-11-14.
- [8] udcast. <https://udcast.net/>, Accessed: 2022-11-14.
- [9] Harumi Murata and Akio Ogihara. Note estimation by contaminated normal distribution for audio watermarking method using non-negative matrix factorization. In *2018 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pp. 469–474, 2018.
- [10] Harumi Murata and Akio Ogihara. A study of audio watermarking method using non-negative matrix factorization. In *2016 IEEE International Symposium on*

- Broadband Multimedia Systems and Broadcasting (BMSB)*, pp. 1–4, 2016.
- [11] Kan Kamada, Tetsuya Kojima, and Udaya Parampalli. Tone code: A novel method for covert communications based on musical components. In *2016 International Symposium on Information Theory and Its Applications (ISITA)*, pp. 335–339, 2016.
- [12] Tetsuya Kojima, Akihiro Oizumi, and Udaya Parampalli. Properties of an emergency broadcasting system based on audio data hiding. In *2015 International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP)*, pp. 142–145. IEEE, 2015.
- [13] Yuto Matsunaga, Tetsuya Kojima, Naofumi Aoki, Yoshinori Dobashi, and Tsuyoshi Yamamoto. A digital watermarking technique for music data using distortion effect. *Acoustical Science and Technology*, Vol. 39, No. 1, pp. 37–39, 2018.
- [14] Yuto MATSUNAGA, Tetsuya KOJIMA, Naofumi AOKI, Yoshinori DOBASHI, and Tsuyoshi YAMAMOTO. Some evaluations on a digital watermarking technique for music data using distortion effect. *IEICE Transactions on Information and Systems*, Vol. E102.D, No. 6, pp. 1119–1125, 2019.
- [15] ITU-R Rec. Bs.1387. Method for objective measurements of perceived audio quality, 2001.
- [16] 松永悠斗, 小嶋徹也. 楽曲の特徴や歪みエフェクトを活用した電子透かし技術の提案. 信学技報, 第 116 卷 of *EMM2016-78*.
- [17] Golomb Solomon, W. and Gong Guang. *Signal Design for Good Correlation: For Wireless Communication, Cryptography, and Radar*. Cambridge University Press, 2005.
- [18] R. Gold. Maximal recursive sequences with 3-valued recursive cross-correlation functions (corresp.). *IEEE Transactions on Information Theory*, Vol. 14, No. 1, pp. 154–156, 1968.
- [19] Zölzer Udo. *DAFX: Digital Audio Effects*. John Wiley & Sons, Baffins, Lane, 2002.
- [20] Felix Eichas and Udo Zölzer. Black-box modeling of distortion circuits with

- block-oriented models. In *International Conference on Digital Audio Effects*, 2016.
- [21] 和徳齋藤. Soul Power Instruments エフェクターの設計と製作. ラトルズ, 東京都, 日本, 2016.
- [22] Z. Wang, Alan Conrad Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. The evaluation criteria and competition. 2014.
- [23] Guanshuo Xu, Han-Zhou Wu, and Yun-Qing Shi. Structural design of convolutional neural networks for steganalysis. *IEEE Signal Processing Letters*, Vol. 23, No. 5, pp. 708–712, 2016.
- [24] Yinlong Qian, Jing Dong, Wei Wang, and Tieniu Tan. Learning and transferring representations for image steganalysis using convolutional neural network. In *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 2752–2756, 2016.
- [25] Songbin Li, Yizhen Jia, and C.-C. Jay Kuo. Steganalysis of qim steganography in low-bit-rate speech signals. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 25, No. 5, pp. 1011–1022, 2017.
- [26] Junqi Wu, Bolin Chen, Weiqi Luo, and Yanmei Fang. Audio steganography based on iterative adversarial attacks against convolutional neural networks. *IEEE Transactions on Information Forensics and Security*, Vol. 15, pp. 2282–2294, 2020.
- [27] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale, 2020.
- [28] Wei Han, Zhengdong Zhang, Yu Zhang, Jiahui Yu, Chung-Cheng Chiu, James Qin, Anmol Gulati, Ruoming Pang, and Yonghui Wu. Contextnet: Improving convolutional neural networks for automatic speech recognition with global context, 2020.
- [29] Yuto Matsunaga, Naofumi Aoki, Yoshinori Dobashi, and Tetsuya Kojima. Distortion based watermark extraction technique using 1d cnn. In *2021 Interna-*

- tional Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, pp. 31–35. IEEE, 2021.
- [30] Idmt-smt-guitar. <https://www.idmt.fraunhofer.de/en/publications/datasets/guitar.html>, Accessed: 2022-11-14.
- [31] Chandrasekhar Paseddula and Suryakanth V. Gangashetty. Dnn based acoustic scene classification using score fusion of mfcc and inverse mfcc. In *2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS)*, pp. 18–21, 2018.
- [32] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, Vol. 9, No. 8, pp. 1735–1780, 1997.
- [33] Yuto Matsunaga, AOKI Naofumi, Yoshinori Dobashi, and Tetsuya Kojima. A black box modeling technique for distortion stomp boxes using lstm neural networks. In *2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, pp. 653–656. IEEE, 2020.
- [34] Alec Wright, Eero-Pekka Damskägg, Vesa Välimäki, et al. Real-time black-box modelling with recurrent neural networks. In *22nd international conference on digital audio effects (DAFx-19)*, 2019.
- [35] Guokun Lai, Wei-Cheng Chang, Yiming Yang, and Hanxiao Liu. Modeling long- and short-term temporal patterns with deep neural networks. In *The 41st international ACM SIGIR conference on research & development in information retrieval*, pp. 95–104, 2018.

# 著者の研究業績

## (A) 学術誌（学位論文関係）

- [ A-1 ] Yuto Matsunaga, Tetsuya Kojima, Naofumi Aoki, Yoshinori Dobashi and Tsuyoshi Yamamoto, "Some evaluations on a digital watermarking technique for music data using distortion effect." IEICE Transactions on Information and Systems, Vol. E102.D, No. 6, pp. 1119–1125.

## (B) レター（学位論文関係）

- [ B-1 ] Yuto Matsunaga, Tetsuya Kojima, Naofumi Aoki, Yoshinori Dobashi, and Tsuyoshi Yamamoto. "A digital watermarking technique for music data using distortion effect." Acoustical Science and Technology, Vol. 39, No. 1, pp. 37–29, 2018.

## (C) 国際学会（学位論文関係）

- [ C-1 ] Yuto Matsunaga, Tetsuya Kojima, Naofumi Aoki, Yoshinori Dobashi and Tsuyoshi Yamamoto, "A study on a digital watermarking technique for music data using distortion effect," 2017 Eighth International Workshop on Signal Design and Its Applications in Communications (IWSDA), pp. 131–134, 2017.
- [ C-2 ] Yuto Matsunaga, Naofumi Aoki, Yoshinori Dobashi, and Tetsuya Kojima, "Distortion based watermark extraction technique using 1d cnn." In 2021 In-



ternational Conference on Artificial Intelligence in Information and Communication (ICAIIIC), pp. 31–35. IEEE, 2021.

## (D) 国際学会（その他）

- [ D-1 ] Yuto Matsunaga, Naofumi Aoki, Yoshinori Dobashi, and Tsuyoshi Yamamoto, “A digital modeling technique for distortion effect based on a machine learning approach,” APSIPA Annual Summit and Conference 2018 (APSIPA ASC 2018), Honolulu, Hawaii, USA, pp.1888–1892, November 12–15, 2018.
- [ D-1 ] Yuto Matsunaga, Naofumi Aoki, Yoshinori Dobashi, and Tetsuya Kojima, “A black box modeling technique for distortion stomp boxes using LSTM neural networks,” IEEE The 2nd International Conference on Artificial Intelligence in Information and Communication 2020 (ICAIIIC2020), Fukuoka, Fukuoka, Japan, pp.653–656, February 19–21, 2020.

## (E) 国内学会発表

- [ E-1 ] 松永 悠斗, 小嶋 徹也, “楽曲の特徴や歪みエフェクトを活用した電子透かし技術の提案,” 電子情報通信学会技術研究報告マルチメディア情報ハイディング・エンリッチメント研究会資料, EMM2016-78, January 30–31, 2017.
- [ E-2 ] 松永 悠斗, 青木 直史, 土橋 宜典, 山本 強, “適応信号処理を用いた非線形歪み回路のモデリングに関する一検討,” 電子情報通信学会技術研究報告応用音響研究会資料, EA2017-17, July 20–21, 2017.
- [ E-3 ] 松永 悠斗, 青木 直史, 土橋 宜典, 山本 強, “非線形歪み回路のデジタルモデリングに関する一検討,” 電気・情報関係学会北海道支部連合大会, はこだて未来大学(函館), October 28–29, 2017.
- [ E-4 ] 松永 悠斗, 小嶋 徹也, 青木 直史, 土橋 宜典, 山本 強, “歪みエフェクトを用いた音響電子透かしの改良,” 電子情報通信学会技術研究報告マルチメディア情報ハイディング・エンリッチメント研究会資料, EMM2017-76, January 29–30, 2018.

- [ E-5 ] 松永 悠斗, 青木 直史, 土橋 宜典, 山本 強, “ブロック構造に基づいた非線形歪み回路のモデリング,” 日本音響学会春季研究発表会, 日本工業大学 (南埼玉郡宮代町), March 13–15, 2018.
- [ E-6 ] 松永 悠斗, 青木 直史, 土橋 宜典, 山本 強, “LSTM 機械学習を用いた歪みエフェクタのデジタルモデリング,” 電子情報通信学会技術研究報告応用音響研究会資料, EA2018-26, July 24–25, 2018.
- [ E-7 ] 松永 悠斗, 青木 直史, 土橋 宜典, 山本 強, “歪みエフェクタのデジタルモデリングに関する研究,” 電気・情報関係学会北海道支部連合大会, 北海道大学 (札幌), October 27, 2018.
- [ E-8 ] 松永 悠斗, 青木 直史, 土橋 宜典, 山本 強, 小嶋 徹也, “Deep Neural Network を用いた歪みエフェクタのデジタルモデリングに関する一検討,” 日本音響学会春季研究発表会, 電気通信大学 (調布市), March 5–7, 2019.
- [ E-9 ] 松永 悠斗, 青木 直史, 土橋 宜典, 小嶋 徹也, “機械学習を用いた歪みエフェクタのブラックボックスモデリング,” 電子情報通信学会技術研究報告応用音響研究会資料, EA2019-3, July 16–17, 2019.
- [ E-10 ] 松永 悠斗, 青木 直史, 土橋 宜典, “歪みエフェクタの機械学習と学習のデータに関する一考察,” 電気・情報関係学会北海道支部連合大会, 室蘭工業大学 (室蘭), November 9–10, 2019.
- [ E-11 ] 松永 悠斗, 青木 直史, 土橋 宜典, 小嶋 徹也, “LSTM ニューラルネットワークによる歪みエフェクタのデジタルモデリング,” 日本音響学会春季研究発表会, 埼玉大学 (さいたま), March 16–18, 2020.
- [ E-12 ] 松永 悠斗, 青木 直史, 土橋 宜典, 小嶋 徹也, “歪みエフェクタの機械学習における受容野の拡張に関する研究,” 電気・情報関係学会北海道支部連合大会, オンライン開催, November 7, 2020.
- [ E-13 ] 松永 悠斗, 青木 直史, 土橋 宜典, 小嶋 徹也, “歪み回路の機械学習ベースモデリングにおける受容野に関する一考察,” 電子情報通信学会技術研究報告応用音響研究会資料, EA2020-33, November 20, 2020.
- [ E-14 ] 松永 悠斗, 青木 直史, 土橋 宜典, 小嶋 徹也, “音響機器の LSTM ベースモデリングにおけるミニバッチ学習のデータ構造に関する一考察,” 日本音響学会春季研究発表会, オンライン開催, March 10–12, 2021.

- [ E-15 ] 松永 悠斗, 青木 直史, 土橋 宜典, 小嶋 徹也, “機械学習を用いた歪みエフェクタのデジタルモデリングの高精度化,” 電子情報通信学会技術研究報告応用音声研究会資料, SP2021-11, June 18-19, 2021. (音学シンポジウム 2021, 第 131 回 音楽情報科学研究会, 第 137 回 音声言語情報処理研究会 共催研究会, 電子情報通信学会・日本音響学会 音声研究会 連催研究会)
- [ E-16 ] 松永 悠斗, 青木 直史, 土橋 宜典, “機械学習を用いた歪みエフェクタのデジタルモデリングにおけるエッジデバイス実装,” 日本音響学会秋季研究発表会, 北海道科学大学, September, 2022.

## (F) 受賞

- [ F-1 ] 松永 悠斗, “2017 年度 電子情報通信学会マルチメディア情報ハイディング・エンリッチメント研究会 学生研究賞 (楽曲の特徴や歪みエフェクトを活用した電子透かし技術の提案),” 電子情報通信学会マルチメディア情報ハイディング・エンリッチメント研究会, 2017.
- [ F-2 ] 松永 悠斗, “令和元年 電気・情報関係学会北海道支部連合大会 若手優秀論文発表賞 (歪みエフェクタの機械学習と学習のデータに関する一考察),” 電気・情報関係学会北海道支部, 2019.
- [ F-3 ] 松永 悠斗, “令和元年度電子情報通信学会北海道支部 学生奨励賞,” 電子情報通信学会, 2020.

## (G) その他

- [ G-1 ] 松永 悠斗, “日本学術振興会 特別研究員 (DC-1) に採用,” 日本学術振興会, 2020 ~2023.
- [ G-2 ] 松永 悠斗, “2021 年度北楡会・北海道大学情報系交流会 メディアネットワークコース代表として研究発表会で発表,” 北楡会, オンライン開催, 2021.
- [ G-3 ] 松永 悠斗, “スキルアップ AI キャンプ にて研究内容を講演,” スキルアップ AI 株式会社, オンライン開催, 2022.