

整数ウェーブレット変換と差分拡張に基づく軽量かつ可逆な音声電子透かし方式

黄 緒平^{2,1,a)} 伊藤 彰則^{2,b)}

概要：可逆な音声電子透かしは、証拠性の高いデータの真正性保証を目的とし、原音の完全な復元が不可欠となる応用分野において重要である。本研究では、高い埋め込み容量と計算効率を両立する新たな音声電子透かし手法を提案する。提案手法は、整数ウェーブレット変換（Integer Wavelet Transform: IWT）と差分拡張（Difference Expansion: DE）を組み合わせることで、情報の損失を伴わないデータ埋め込みを実現する。IWT 領域において、低周波および高周波係数のペアに対して透かしデータを埋め込むことで、丸め誤差による情報損失を回避し、信号の可逆性を維持する。12 種類の音声データセットを用いた実験により、本手法の有効性が確認され、平均ピーク信号対雑音比（PSNR）は 36.19dB、短時間客観的明瞭度指標（STOI）は 0.9983 を達成した。本手法は完全な可逆性を保証し、計算量は線形時間 $O(N)$ に抑えられ、可逆な音声電子透かしにおける実用的かつ軽量な手法を実現した。

キーワード：可逆な音声電子透かし、整数ウェーブレット変換、差分拡張、軽量化

Design of a Lightweight Reversible Audio Watermarking System Using Integer Wavelet Transform and Difference Expansion

XUPING HUANG^{2,1,a)} AKINORI ITO^{2,b)}

Abstract: Reversible audio watermarking is crucial in application domains where ensuring the authenticity of highly evidential data requires the integrity and complete restoration of the original audio. This study introduces a novel watermarking scheme that achieves both high embedding capacity and computational efficiency. The proposed method leverages the Integer Wavelet Transform (IWT) in conjunction with a Difference Expansion (DE) embedding strategy to ensure lossless data hiding. By operating in the IWT domain, watermark data is embedded into pairs of low- and high-frequency coefficients, effectively eliminating rounding errors and preserving signal fidelity. Experimental evaluations conducted on a dataset comprising 12 audio samples demonstrate the robustness and transparency of the method, yielding an average Peak Signal-to-Noise Ratio (PSNR) of 36.19 dB and a Short-Time Objective Intelligibility (STOI) score of 0.9983. Furthermore, the scheme guarantees perfect reversibility and exhibits linear time complexity of $O(N)$, making it a practical and lightweight solution for secure and lossless audio watermarking applications.

Keywords: Reversible audio watermarking, Integer Wavelet Transform, difference expansion, lightweight

1. はじめに

インターネットや分散ネットワークを介したデジタルマルチメディアコンテンツの発展は、情報の普及を加速している。しかし、アクセス、複製、改変の容易さは、同時に著作権保護、コンテンツ認証、情報の完全性に関する喫緊

¹ 島根大学総合理工学部
Interdisciplinary Faculty of Science and Engineering, Shimane University

² 東北大工学研究科
Graduate School of Engineering, Tohoku University

a) huang@cis.shimane-u.ac.jp
b) aito.spc@tohoku.ac.jp

の課題をもたらしている。データの真正性を保証する電子透かしは、高音質、頑健性、そして埋め込みペイロードという3つの主要要件を満たすべきである [1-7]。

電子透かし技術は、その応用目的によって分類される。頑健な電子透かしは、悪意のある攻撃に耐えうる著作権情報を恒久的に埋め込むことを目的とする一方、脆弱な電子透かしは、コンテンツへの不正な改変を検出するために提案される。本稿では、主に改ざん検出に使われる可逆な情報ハイディングに着目している [3]。歪みを導入する従来の手法とは異なり、可逆電子透かし技術は、埋め込まれたデータが抽出された後、元信号であるホスト信号を完全に復元できることを保証する仕組みである。これは、元信号が非常に高い価値を持ち、いかなる劣化も許容できないアプリケーションにとって不可欠な要件である。応用例として、診断の完全性が最優先される医療画像、軍事・衛星画像等証拠性が高いデータが含まれる。

可逆電子透かしにおける主要な課題の一つは、埋め込みペイロード量の増加に伴い、画質・音質等の劣化が増大するというトレードオフ関係が存在する点である。大容量な埋め込みを実現しつつ劣化を抑えるために、時系列領域または変換領域のいずれかで動作する技術が提案されてきた。時系列領域の手法は計算が単純であることが多いが、離散コサイン変換 (Discrete Cosine Transform, DCT) や離散ウェーブレット変換 (Discrete Wavelet Transform, DWT) を用いた変換領域の手法は、音響心理モデルをより効果的にアプローチできる変換を行っている。変換係数を変更することにより、人間の聴覚系が敏感でない周波数成分に歪みを分散させることができる [4][5]。

本稿では、大容量な可逆音声電子透かし方式を新たに提案する。本手法は、軽量化を図り、完全に可逆な整数ウェーブレット変換 (IWT) によって生成された変換領域で埋め込みを行う。埋め込みメカニズムは、低周波と高周波の変換係数のペアに適用される、修正された差分拡張 (DE) 技術に基づいている。このアプローチは、周波数領域手法の利点として、知覚的な歪みを抑えながら、差分拡張に特徴的な高いペイロード容量を相乗的に組み合わせるものである。整数対整数の変換を使用することは、浮動小数点変換に固有の丸め誤差を完全に回避し、それによってホスト音声信号の可逆性を保証できる。

本研究の主な貢献は以下の通りである：(1) 不可逆な誤差を導入することなく周波数領域にデータを埋め込むために、整数ウェーブレット変換 (Integer Wavelet Transform, IWT) を活用した可逆音声電子透かし方式の提案。(2) IWT 係数のペアに差分拡張手法を適用し、高音質を維持しつつ高い埋め込み容量を実現。(3) 提案手法の性能を、ペイロード、音質、および可逆性の観点から包括的に評価。

本稿の構成は次の通りである。第2節では、可逆電子透かしの分野における従来研究を概説する。第3節では、提

案アルゴリズムを詳細に述べる。第4節では、実験結果と分析を提示する。最後に、第5節で本稿を結論付ける。

2. 関連研究

可逆データハイディング (Reversible Data Hiding, RDH) はより高い容量と低劣化と進化を続けている。特に、近年ではより洗練された予測モデル、適応的埋め込み、そして機械学習手法の併用に焦点を当てている。

RDH の基礎的な柱は、Tian らによる独創的な研究で開拓された差分拡張 (Difference Expansion, DE) とヒストグラムシフティング (Histogram Shifting, HS) である [7]。DE の差分値を拡張において、より強力な予測誤差拡張 (Prediction-Error Expansion, PEE) 方式が提案された。PEE に関する最近の研究では、ゼロ付近に集中した予測誤差ヒストグラムを生成するために、高精度な予測器が開発された。例えば、線形および非線形のサンプル相関の両方を取り入れたコンテキスト認識予測器を使用する、音声向け多段階 PEE 技術が導入された [3]。この手法は、最も高い頻度を持つ誤差にのみデータを埋め込むことで高い音質を達成しており、予測段階の最適化が大きい貢献になっている。

RDH は深層学習との併用も注目されている。ニューラルネットワーク、特に畳み込みニューラルネットワークや敵対的生成ネットワーク (GAN) が、エンドツーエンドの埋め込み・抽出フレームワークを作成するために使用されている。可逆音声ハイディングのための GAN ベースのモデルが提案され、そこでは生成ネットワークが、識別ネットワークにとって元の信号と区別がつかない電子透かし入り信号を生成することを学習し、それによって埋め込みによる劣化を最小限に抑える [1]。これらのモデルは、元信号の複雑で非線形な特徴を学習し、従来のアルゴリズムと比較して優れたペイロード対歪み性能を達成できる。しかし、深層学習アプローチの主な欠点は、高い計算複雑性、大規模な訓練データセットの必要性である。

可逆音声電子透かしにおいて、FLAC のようなロスレスオーディオコーデックのための適応的 PEE ベースの方式が開発されており、これは音響心理学の原理を用いて計算された音声信号の局所的な複雑さに応じて埋め込み強度を調整している [5]。また、整数ウェーブレット変換 (IWT) と係数順序付け戦略を組み合わせて、ヒストグラムシフティングのための最適な埋め込み位置を選定した [8]。

本提案手法は、計算性能と変換領域処理における音質維持に着目している。深層学習モデル [1] は高性能を実現しているが、それらは大きな計算オーバーヘッドを必要としている [3], [5]。訓練や複雑な適応ロジックを必要とせず、計算的に軽量かつ大容量である埋め込み手法は必要とされている。

本提案の新規性について、IWT は従来研究においても

表 1 提案手法と従来のウェーブレットベース研究とのアルゴリズムおよび手法の比較

	提案手法 (IWT + DE)	従来のウェーブレット手法
変換方式	整数ウェーブレット変換を使用。すべての演算は整数ベースであり、丸め誤差を防ぎ、完全な可逆性を保証する。	しばしば DWT のような浮動小数点変換を使用し、軽微な誤差が生じる可能性がある。可逆手法では IWT も使用される。
埋め込みアルゴリズム	比較的単純で大容量な手法である差分拡張 (DE) を IWT 係数に適用する。	頑健な手法：スペクトラム拡散や量子化のような技術を使用。可逆手法：DE、低歪みだが低容量のヒストグラムシフティング (HS)、またはより高度な予測誤差拡張 (PEE) を使用。
計算複雑性	計算コストは軽量である。整数演算のみで構成され、軽量化を実現している。	最近の可逆手法：高度な予測モデルや深層学習は、相当な計算リソースと訓練データを必要とする。

音声に使用されていたが [8]、通常はより低容量の HS 技術と組み合わされていた。本研究では、大容量の DE メカニズムと組み合わせることで、高音質を維持しながら、軽量化を図り、効率的かつ実装が容易な手法を提案している。

3. 提案手法

3.1 提案手法と従来のウェーブレットベース手法との違い

提案手法と従来のウェーブレットベース研究との主な違いは、目的とそれに対応するアルゴリズムの計算量にある。提案アルゴリズムの目的は完全な可逆性を優先することであるのに対し、従来のウェーブレットベースの手法は攻撃に対する頑健性を保証することである。従って、埋め込み容量を確保するための、振幅拡張する位置を選定アルゴリズムが異なる。

3.1.1 主な目的と特性の違い

ウェーブレット電子透かしに関する従来研究の多くは頑健性に焦点を当てている。これは、音楽配信のようなアプリケーションでの著作権保護を目的としており、MP3 圧縮、ノイズ付加、フィルタリングなどの攻撃を受けた後でも電子透かしが検出可能であることが目標である。この頑健性を達成するために、通常、音声データには不可逆な変更が加えられる。

提案手法の主な目的は、完全な可逆性である。医療記録・法的証拠等、元の音声を 1 ビットの誤差もなく完全に復元する必要がある真正性を保証するアプリケーションに特化している。可逆音声電子透かしは特に改ざん検出に着目している。提案アルゴリズムは、主に埋め込みに整数ウェーブレット変換 (IWT) と差分拡張 (DE) を用いる。可逆

性を保証するためにリフティング処理が適用される。軽量化を保証する一方、弱点として圧縮やノイズのような攻撃に対する頑健性を持たない。

提案手法は、実装の容易さ、低い計算複雑性、および大容量という実用的な利点に着目している。表 1 は、提案する軽量 IWT 手法と従来のウェーブレット変換ベースの手法との違いをまとめている。

3.2 整数ウェーブレット変換に基づく前処理

この手法では、DCT (離散コサイン変換) の代わりに、整数ウェーブレット変換と等価な可逆ペアワイス整数変換を用いる。このプロセスにおいて、信号成分を低周波成分と高周波成分に分離する。

3.2.1 順変換

順変換は、入力ペア (x_1, x_2) を低周波成分 (平均) y_1 と高周波成分 (差分) y_2 に変換する。

$$y_1 = \left\lfloor \frac{x_1 + x_2}{2} \right\rfloor$$

$$y_2 = x_1 - x_2$$

ここで、 $\lfloor \cdot \rfloor$ は最も近い整数への切り捨てを表す。この操作は、フレーム内のデータに繰り返し適用される。

3.2.2 逆変換

逆変換は、変換された係数ペア (y_1, y_2) から元のデータペア (x_1, x_2) を完全に復元する。

$$x_1 = y_1 + \left\lfloor \frac{y_2 + 1}{2} \right\rfloor$$

$$x_2 = y_1 - \left\lfloor \frac{y_2}{2} \right\rfloor$$

3.3 差分拡張に基づく電子透かし処理

このプロセスは、変換された係数のペア間の差を拡張することにより、1 ビットの情報を可逆的に埋め込む。

3.3.1 埋め込み

埋め込みプロセスは、单一の透かしビット $b \in \{0, 1\}$ を選択された係数ペア $(c_{\text{low}}, c_{\text{high}})$ に埋め込み、新しいペア $(c'_{\text{low}}, c'_{\text{high}})$ を生成する。

(1) 差分の計算：

$$d = c_{\text{high}} - c_{\text{low}}$$

(2) 差分の拡張：埋め込むビット b を用いて、新しい拡張された差分 d_w を計算する。

$$d_w = 2d + b$$

(3) 係数の更新：元の係数は、新しい拡張差分 d_w を持つように更新される。これは変化量 Δ を用いて計算される。

$$\Delta = d_w - d = d + b$$

$$c'_{\text{high}} = c_{\text{high}} + \left\lfloor \frac{\Delta + 1}{2} \right\rfloor$$

$$c'_{\text{low}} = c_{\text{low}} - \left\lfloor \frac{\Delta}{2} \right\rfloor$$

3.3.2 抽出と復元

抽出と復元プロセスは、電子透かし入りの係数ペア $(c'_{\text{low}}, c'_{\text{high}})$ から埋め込まれたビットを抽出し、同時に元の係数ペア $(c_{\text{low}}, c_{\text{high}})$ を完全に復元する。

まず、電子透かし入り係数から拡張差分 d_w を計算する。

$$d_w = c'_{\text{high}} - c'_{\text{low}}$$

埋め込まれたビット b は、 d_w の偶奇から直接抽出できる。

$$b = d_w \pmod{2}$$

次に、元の差分 d を復元する。

$$d = \left\lfloor \frac{d_w}{2} \right\rfloor$$

埋め込みステップと同じ変化量 Δ ($\Delta = d_w - d$) を計算し、逆の操作を行うことで、元の係数が復元される。

$$c_{\text{high}} = c'_{\text{high}} - \left\lfloor \frac{\Delta + 1}{2} \right\rfloor$$

$$c_{\text{low}} = c'_{\text{low}} + \left\lfloor \frac{\Delta}{2} \right\rfloor$$

3.4 ワークフローとアルゴリズム

X を元の音響信号とし、 $W = \{w_1, w_2, \dots, w_N\}$ をバイナリの電子透かしメッセージとする。ここで、 $w_i \in \{0, 1\}$ である。

1：埋め込みプロセス

電子透かし W を信号 X に埋め込み、電子透かし入り信号 X' を生成する。埋め込みのアルゴリズムを表 2 に示す。

2：抽出と復元プロセス

電子透かし入り信号 X' から、電子透かし W' を抽出し、元の信号 X_{restored} を完全に復元する。

このプロセスで、抽出された電子透かし $W' = \{w'_1, w'_2, \dots, w'_N\}$ と再構成された信号 X_{restored} が生成される。抽出と再構成のアルゴリズムは表 3 に示されている。

4. 実験結果

4.1 実験データ

本研究は、取り調べや遺言等証拠性の高いデータの完全性を保証するというターゲットアプリケーションに基づき、主に音声データを使用する。評価には、ITU-T Test Signals for Telecommunication Systems—Test Vectors Associated to Rec. ITU-T P.50 Appendix I [11] のデータセットを使用した。このデータセットには、16 kHz サンプリング、16

表 2 アルゴリズム 1：埋め込み

ステップ 1：フレーミング	信号 X を、 N 個の重ならないフレーム F_i (各サイズは FRAME_SIZE) に分割する。 $X \rightarrow \{F_1, F_2, \dots, F_N\}$
ステップ 2：	各フレーム F_i について ($i = 1, \dots, N$) :
	a. 順変換：整数ウェーブレット変換 (IWT) をフレーム F_i に適用し、変換係数 T_i を得る。 $T_i = \text{IWT}(F_i)$ b. ビット埋め込み：差分拡張 (DE) を用いて、透かしビット w_i を変換フレーム T_i に埋め込み、透かし入り変換フレーム T'_i を得る。 $T'_i = \text{EmbedDE}(T_i, w_i)$ c. 逆変換：逆 IWT を透かし入り変換フレーム T'_i に適用し、透かし入りフレーム F'_i を得る。 $F'_i = \text{IWT}^{-1}(T'_i)$
ステップ 3：アンフレーミング	全ての透かし入りフレーム F'_i を連結し、最終的な透かし入り信号 X' を生成する。 $X' = \{F'_1, F'_2, \dots, F'_N\}$

表 3 アルゴリズム 2：抽出と再構成

ステップ 1：フレーミング	透かし入り信号 X' を、 N 個のフレーム F'_i に分割する。 $X' \rightarrow \{F'_1, F'_2, \dots, F'_N\}$
ステップ 2：	各透かし入りフレーム F'_i について ($i = 1, \dots, N$) :
	a. 順変換：IWT をフレーム F'_i に適用し、透かし入り変換係数 T'_i を得る。 $T'_i = \text{IWT}(F'_i)$ b. ビット抽出とフレーム復元：DE 抽出アルゴリズムを T'_i に適用し、抽出されたビット w'_i と復元された（元の）変換フレーム T_i を得る。 $(w'_i, T_i) = \text{ExtractDE}(T'_i)$ c. 逆変換：逆 IWT を復元された変換フレーム T_i に適用し、復元された元のフレーム F_i を得る。 $F_i = \text{IWT}^{-1}(T_i)$
ステップ 3：フレーム連結	全ての復元されたフレーム F_i を連結し、再構成された信号 X_{restored} を生成する。 $X_{\text{restored}} = \{F_1, F_2, \dots, F_N\}$

ビット量子化の波形が含まれている。アメリカ英語、アラビア語、中国語、デンマーク語、フランス語、ドイツ語の 6 言語について、12 名の女性および男性話者による 12 トラックの音声信号を使用した。各トラックの平均長は約 10 秒であったが、データ長がフレーム長の整数倍になるように調整した。フレームサイズは 512 に設定されている。平均埋め込みビット数は 334.8 ビットで、最大 480 ビット、最小 160 ビットであった。図 1 は、A_eng_f1.wav のオリジナル、ステゴ（透かし入り）、およびその差分のスペクトログラムをプロットしたものである。

4.2 評価基準

ステゴデータの音質を評価するために、主に平均二乗誤差 (Mean Squared Error, MSE)、ピーク信号対雑音比 (Peak Signal-to-Noise Ratio, PSNR [dB])、および短時間客観的明瞭度 (Short-Time Objective Intelligibility, STOI)

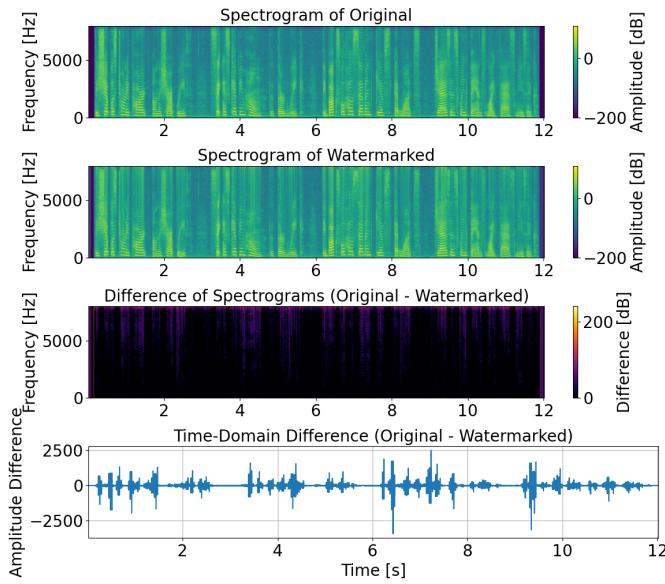


図 1 オリジナル、ステゴ（透かし入り）、および差分のスペクトログラム

を使用する。

MSE は、元の信号と透かし入り信号との間の平均二乗差を測定するために使用される基本的な指標である。これは、2つの音響信号間の差である誤差信号のパワーを計算する。サンプル数 N の元の信号 X と透かし入り信号 Y について、MSE は次のように計算される：

$$MSE = \frac{1}{N} \sum_{i=1}^N (X_i - Y_i)^2$$

MSE の値は低いほど良い。

ピーク信号対雑音比は、信号の最大可能パワーと、その品質に影響を与える破壊的なノイズのパワーとの比率を測定する ([dB]) [8]。MSE から導出される。

$$PSNR = 20 \log_{10} \left(\frac{\text{MAX}}{\sqrt{MSE}} \right)$$

PSNR の値は高いほど良い。

STOI は、劣化した音声信号の明瞭度を、参照信号と比較することによって、人間にとてどれくらい明瞭に聞き取れるかを客観的に予測するための指標である [9]。STOI は、信号を短く重なり合う時間フレームに分割し、各フレームのスペクトログラムを計算し、元信号と劣化した信号の時間的エンベロープ間の相関を測定し、その平均値を計算する。元の音声の各周波数成分における音量の揺れ動きパターンを、劣化した音声がどれだけ再現できているかを評価する指標である。STOI スコアは 0 から 1 の間の値である。STOI の値は高いほど良い。

音質指標の統計的要約を表 4 に示す。STOI が 1 に近づく値が達成されており、これは透かし入り信号が完全に明瞭であり、劣化を区別することが困難であることを示唆している。

表 4 音質の評価結果

指標	平均	標準偏差	最小	最大
MSE	324,413.43	201,382.78	71,001.91	761,817.97
PSNR (dB)	36.19	3.31	31.49	41.80
STOI	0.9983	0.0010	0.9965	0.9998

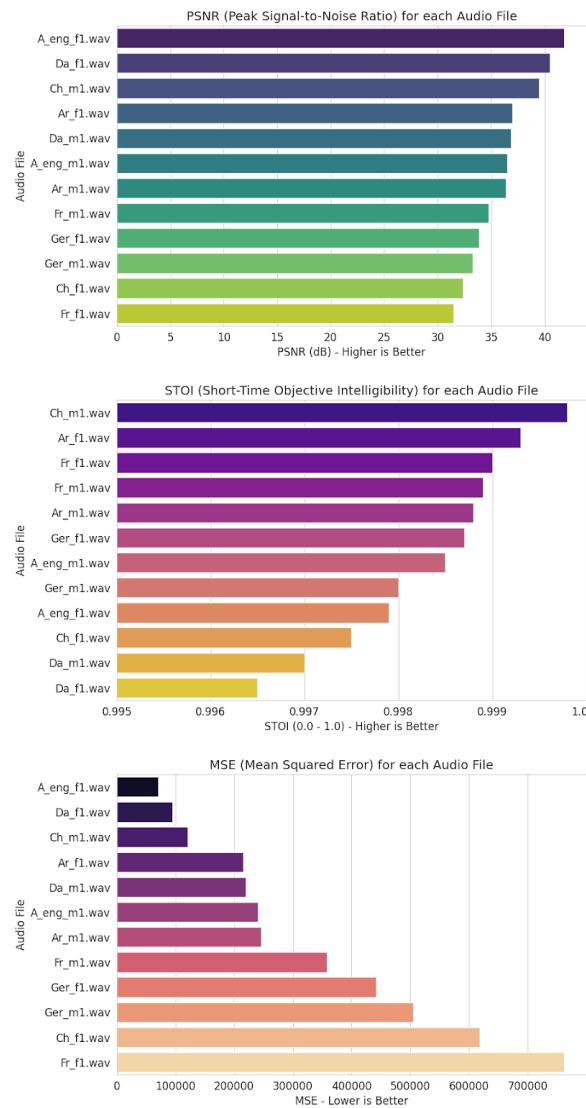


図 2 音質評価

図 2 は、提案手法の音質評価結果をプロットしたものである。

4.3 計算コスト

提案手法は、可逆性と不可視性に加えて、軽量な計算を達成することが期待される。提案方式の計算効率は、埋め込み、抽出、検証を含むその中核プロセスの複雑性を評価することによって分析される。分析は以下の変数定義に基づいている。

N : 入力音響信号の総サンプル数

M : フレームあたりのサンプル数 (FRAME_SIZE)

K : フレームの総数、ここで $K = N/M$

埋め込みと抽出のプロセスは両方とも対称的であり、線形時間計算量を示す。予備処理のフレーム分割及び再構成において、計算量は $O(N)$ となる。整数ウェーブレット変換と逆変換において、変換は K 個のフレームに適用される。各フレームのすべての対数段階にわたる総計算量はフレームサイズ M に比例する。従って、全体の計算量は $O(K \times M)$ となり、これは $O((N/M) \times M) = O(N)$ になる。電子透かしの埋め込みと抽出において、これらのステップは K 個のフレームを一度ループし、各ループ内で定数時間 $O(1)$ の操作を実行する。従って、計算量は $O(K)$ 、または $O(N/M)$ である。

埋め込みと抽出の両方の手順の総計算量は、これらのステップの合計である。主要な項は $O(N)$ であるため、全体の計算量は $O(N)$ となる。検証プロセスもまた、 $O(N)$ の線形時間計算量になる。実際の処理時間 [s] を表 5 に示す。

表 5 処理時間

指標	平均	標準偏差	最小	最大
処理時間 (s)	0.0323	0.0121	0.01	0.05

5. むすび

本稿では、軽量な可逆音声電子透かし方式を提案し、評価を行った。具体的に、無損失な時系列信号を変換するために整数ウェーブレット変換を適用し、変換領域の係数にデータを埋め込むために差分拡張技術適用用した。

実証実験を行った結果、本提案手法の可逆性及び高音質性が確認できた。本手法において、すべての検証データを用いて、電子透かし抽出プロセス後に元の音響信号が何らの歪みもなく再構成されることが示された。また、本研究では平均 PSNR 36.19 dB を達成し、STOI スコアは 1.0 に近く、埋め込まれたデータが知覚品質や音声の明瞭度に与える影響が最小限であることが示された。更に、計算量分析においては、アルゴリズムが効率的な線形時間計算量 $O(N)$ で実行できる。

本手法は計算の軽量化を実現している一方で、攻撃に対する脆弱性を持つという課題がある。本提案手法は圧縮やノイズ付加のような攻撃に耐えるようには設計されていない。今後は、埋め込み箇所の特定及び拡張アルゴリズムに焦点を当てたい。最適な係数ペアの拡張の精緻化を検討することで、頑健性を向上させる適応的埋め込み手法の実現を目指す。

参考文献

- 2021
- [2] YANG, Hyukryul, et al. Hiding video in audio via reversible generative models. In: Proceedings of the IEEE/CVF international conference on computer vision, pp.1100-1109, 2019
 - [3] Kaur, Gurjinder, et al. "A Comprehensive Study of Reversible Data Hiding (RDH) Schemes Based on Pixel Value Ordering (PVO) G. Kaur et al." Archives of Computational Methods in Engineering vol.28(5), pp.3517-3568, 2021
 - [4] Huang, X.; Ito, A. Imperceptible and Reversible Acoustic Watermarking Based on Modified Integer Discrete Cosine Transform Coefficient Expansion. Appl. Sci. vol.14(2757), 2024 <https://doi.org/10.3390/app14072757>
 - [5] Huang X, Ito A. Reversible Spectral Speech Watermarking with Variable Embedding Locations Against Spectrum-Based Attacks. Applied Sciences. vol.15(1):381, 2025 <https://doi.org/10.3390/app15010381>
 - [6] Z. Ni, et al.: Reversible data hiding. IEEE Transactions on Circuits and Systems for Video Technology vol.16(3), pp: 354-362, 2006.
 - [7] J. Tian: Reversible data embedding using a difference expansion. IEEE Transactions on Circuits and Systems for Video Technology 13.8, pp: 890-896, 2003.
 - [8] Meng, L., Liu, L., Tian, G., & Wang, X. An adaptive reversible watermarking in IWT domain. Multimedia Tools and Applications, 80(1), 711-735, 2021
 - [9] Fela, R. F., & Zacharov, N. Towards a perceived audiovisual quality model for immersive content. In 2020 IEEE international conference on quality of multimedia experience (QoMEX) pp. 1-6, 2020
 - [10] C. H. Taal, R. C. Hendriks, R. Heusdens and J. Jensen, "An Algorithm for Intelligibility Prediction of Time-Frequency Weighted Noisy Speech," in IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 7, pp. 2125-2136, doi: 10.1109/TASL.2011.2114881, Sept. 2011
 - [11] ITU-T. Test Signals for Telecommunication Systems—Test Vectors Associated to Rec. ITU-T P.50 Appendix I. 2001. Available online: <http://www.itu.int/net/itu-t/sigdb/genaudio/Pseries.html>.