A Lightweight and Reversible Audio Watermarking Scheme Based on Integer Wavelet Transform

Xuping HUANG*† and Akinori Ito†
* Interdisciplinary Faculty of Science and Engineering, Shimane University, Matsue 690-8504, Japan
E-mail: huang@cis.shimane-u.ac.jp
† Department of Communications Engineering, Graduate School of Engineering
Tohoku University, Sendai 980-8577, Japan
E-mail: aito.spcom@tohoku.ac.jp

I. Introduction

In recent years, the rapid advancement of generative AI has significantly lowered the barrier to data tampering and the creation of falsified multimedia content, posing serious threats to information integrity and trust[1][2]. Among the countermeasures explored, reversible audio watermarking has emerged as a promising approach, as it enables reliable detection of tampering while preserving the original signal quality[3].

The objective of this study is to establish a reversible watermarking technique that ensures the authenticity and reliability of digital information. Conventional approaches suffer from irreversible distortions and high computational costs, limiting their practical deployment. In this work, we emphasize lightweight design, enabling efficient execution even on ubiquitous devices. Such a framework is expected to contribute to the detection of tampered contents, including deepfake videos generated by generative AI, thereby strengthening the foundation of future media security.

This paper proposes a novel reversible audio watermarking scheme that synergistically combines the Integer Wavelet Transform (IWT) and Difference Expansion (DE). The use of IWT, an integer-to-integer transform, completely avoids rounding errors, providing a foundation for perfect reversibility. Upon this foundation, the high-capacity DE technique is applied to IWT coefficient pairs to embed data. This approach offers a lightweight, high-capacity, and perfectly reversible solution, providing a practical alternative to computationally intensive deep learning-based methods. The main contributions are: (1) a fully reversible, IWT-based frequency-domain embedding framework; (2) the application of DE to IWT coefficients for high-capacity audio watermarking; and (3) a comprehensive evaluation confirming the scheme's high fidelity, efficiency, and perfect reversibility.

II. Proposed Method

The proposed algorithm consists of three main stages: a reversible transform using IWT, a reversible embedding mechanism using DE, and the overall workflow for embedding and extraction.

A. Integer Wavelet Transform (IWT)

To ensure perfect reversibility, we use a pairwise integer transform equivalent to the Integer Haar Wavelet Transform. This transform avoids floating-point arithmetic and its associated rounding errors.

1) Forward Transform: An input pair of samples (x_1, x_2) is converted into a low-frequency (average) component y_1 and a high-frequency (difference) component y_2:

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

$$y_2 = x_1 - x_2$$

where $\lfloor \cdot \rfloor$ is the floor function.

2) Inverse Transform: The original pair (x_1, x_2) is perfectly reconstructed from the transformed pair (y_1, y_2):

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

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

B. Difference Expansion (DE) Based Watermarking

DE is a high-capacity reversible data hiding technique. We apply it to pairs of IWT coefficients $(c_{\text{low}}, c_{\text{high}})$.

1) Embedding Process: To embed a watermark bit $b \in \{0, 1\}$, the difference $d = c_{\text{high}} - c_{\text{low}}$ is calculated. This difference is then expanded to create a new difference d_w:

$$d_w = 2d + b$$

The original coefficients are updated to reflect this new difference. The change amount $\Delta = d_w - d = d + b$ is distributed between the two coefficients to produce the watermarked pair $(c'_{\text{low}}, c'_{\text{high}})$:

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

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

2) Extraction and Restoration Process: The process is perfectly reversible. First, the watermarked difference $d_w = c'_{\text{high}} - c'_{\text{low}}$ is computed. The embedded bit b is extracted from its parity:

$$b = d_w \pmod 2$$

The original difference d is then restored by integer division:

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

Using the same change amount Δ, the original coefficients are perfectly restored:

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

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

III. Experimental Results

A. Setup and Evaluation Metrics

The proposed method was evaluated on 12 speech signals from the ITU-T P.50 dataset (16 kHz, 16-bit). Audio quality was assessed using Peak Signal-to-Noise Ratio (PSNR) and Short-Time Objective Intelligibility (STOI). Computational cost was measured by both of theoretical computational complexity and the real processing time. Reversibility was verified by comparing the reconstructed signal with the original.

B. Performance Analysis

As the overall workflow, the audio signal is first divided into non-overlapping frames. For each frame, the IWT is applied. A watermark bit is then embedded into a selected pair of IWT coefficients using the DE algorithm. Figure 1 plots the results, including spectrograms of original audio, watermarked audio, and their difference in the frequency and the time domain.

1) Imperceptibility and Audio Quality: The method achieved high audio fidelity, with an average PSNR of 36.19 dB. More significantly, the average STOI score was 0.9983, extremely close to the perfect score of 1.0. This indicates that the watermarking process has a negligible impact on speech intelligibility, a critical factor for the target applications. A summary of the audio quality metrics is presented in Table I.

TABLE I
STATISTICAL SUMMARY OF AUDIO QUALITY METRICS

Metric	Mean	Std Dev	Min	Max
PSNR (dB)	36.19	3.31	31.49	41.80
STOI	0.9983	0.0010	0.9965	0.9998

2) Reversibility and Computational Cost: In all test cases, the original audio signal was perfectly reconstructed bit-for-bit after watermark extraction, confirming the scheme's complete reversibility. The algorithm's complexity is linear, $O(N)$, where N is the signal length. The average processing time for embedding or extraction on a 10-second audio file was only 0.0323 seconds, demonstrating its high computational efficiency and suitability for resource-constrained environments.

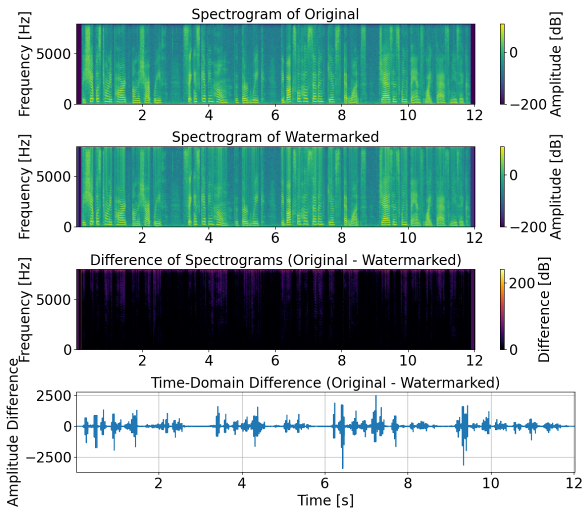

Fig. 1. Spectrograms of original audio, watermarked audio, and their difference, showing minimal perceptual change.

IV. Conclusions

This paper presented a lightweight and perfectly reversible audio watermarking scheme based on the Integer Wavelet Transform and Difference Expansion. The method's core strength lies in its synergistic design: IWT provides a lossless domain, enabling the high-capacity DE algorithm to operate without compromising perfect reversibility.

Experimental results demonstrated that the scheme achieves high imperceptibility (average PSNR 36.19 dB, average STOI 0.9983), complete reversibility, and with computational efficiency of ($O(N)$ complexity). Its simplicity, without training requirements makes it a practical and valuable solution for applications demanding data integrity. However, since the method is fragile, our future work could explore adaptive embedding to enhance the robustness, as well as the imperceptibility.

References

[1] Xuping Huang, Akinori Ito, "Reversible Spectral Speech Watermarking with Variable Embedding Locations Against Spectrum-Based Attacks," *Applied Sciences*, vol.15(1)381, pp.1-16, Jan 2025

[2] Xuping Huang, Akinori Ito, "A Study on Variable Embedding Locations of Reversible Spectral Speech Watermarking," *IEEE Proc. of 2024 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pp.1-6, Dec 2024

[3] Xuping Huang, Akinori Ito, "Imperceptible and Reversible Acoustic Watermarking Based on Modified Integer Discrete Cosine Transform Coefficient Expansion," *Applied Sciences*, vol.14(2757). 20 pages, https://doi.org/10.3390/app14072757, March 24, 2024