VIDWAN Journal of Scientific Research

OPEN ACCESS

PEER-REVIEWED - OPEN ACCESS Vol. 01 · Issue 01 · 2026 · pp. 21-26 | ISSN: APPLIED (Online) | INAGURAL EDITION

A Robust Audio Watermarking Method Using Empirical Mode Decomposition

MD Shaifulla, *Department of CSE,*
Scient Institute of Technology (Autonomous), Hyd

ABSTRACT

EMD is used in this study to embed adaptive watermarks in audio recordings. EMD seamlessly decomposes the audio signal into oscillatory Inherent Mode Functions (IMFs) from each subsequent frame. The Intrinsic Mode Function (IMF) concludes with the time codes and watermark. The IMF's initial signal is exceedingly rare, rendering even a multitude of attacks largely ineffective on its quality. Gaussian attack, pink noise, resampling, cropping, or filtering does not influence hidden marking. Our watermarking method demonstrated superior efficacy compared to the recently disclosed tactics.

Keywords: *Empirical Mode Decomposition (EMD), Intrinsic Mode Functions (IMFs), Discrete cosine transform (DCT), Discrete Wavelet Transform (DWT).*

Article history: Received 8 July 2026 · Revised 20 July 2026 · Accepted 28 July 2026 · Published 7 August 2026

1. INTRODUCTION

Watermarking is a technique for concealing the addition of data to a primary transmission. The information can serve multiple functions, including ownership verification, restricting copy reproduction, validating usernames, and preventing unauthorized usage. The hosted content comprises media files, including audio, video, or static images. Digital videos incorporate a watermark embedded in the host data for copyright protection. The watermarking signal, similar to the host signal, may constitute an audio transmission. Digital music watermarking is most effective when it is resilient, secure, and difficult to detect. The watermark must be audible during audio transmission for the recipient to decode it. The longevity of a watermark can be assessed by observing its variations under diverse conditions. A trademark that cannot be altered or eliminated without the owner's consent is deemed an inviolable trademark. A diverse array of techniques is available for watermarking audio files. The subsequent are prevalent classifications of methods: To capture data, the time-domain audio stream undergoes temporal and spectral temporal watermarking. Spectral watermarking involves modifying specific frequencies in the original audio stream and incorporating watermark information into the modified frequency-domain data block. Temporal watermarking techniques encompass, but are not restricted to, time scale modification, echo concealment, empirical mode decomposition (EMD), and dual-channel audio watermarking. Spectral watermarking encompasses various transformations applied to the host signal, including the Discrete Cosine Transform (DCT) and the Discrete Wavelet Transform (DWT). Audio watermarking integrates spectral and temporal techniques to enhance results through the application of Discrete Wavelet Transform (DWT) and Singular Value Decomposition (SVD). It ensures reliability and durability while enhancing its load-bearing capacity. We propose exploring a method to innovatively integrate EMD's divisional competencies with DCT's data-reduction capabilities. The fundamental functions employed by DWT, DCT, and analogous algorithms are insufficient to accurately represent all conceivable real-world signals. Due to its ability to deconstruct the signal based on its value, empirical mode decomposition (MD) is an effective tool for addressing this issue. The EMD lacks a system to prioritize tasks based solely on data. Intrinsic mode functions (IMFs) are uniform bands with approximately zero mean values produced by the empirical mode decomposition (EMD) technique. The EMD decomposition operates by producing N intrinsic mode functions (IMFs), with $r_N(t)$ representing the identical final residue component. While acquiring IMFs enhances the probability of their alignment, this outcome is not guaranteed. A limited subset of EMD modes and extremes is accessible for mode switching. The most recent

IMF report exhibits minor discoloration. The stamp must display a full-color image as per the recommended procedure. An image must initially be transformed into a collection of RGB values prior to its preparation for the Discrete Cosine Transform (DCT). We employ the energy compression function of the Discrete Cosine Transform (DCT) to achieve this. The image is initially converted into digital data prior to being combined with the audio track. It is possible to transform a watermarked audio stream into a watermarked color image by altering the sequence of processes.

$$X_t = \sum_{n=1}^{\infty} IMF_j(t) + \mathrm{rN(t)}$$

N denotes the quantity of IMFs, while rN(t) signifies the final residue. The identified IMFs possess significant commonalities. A multitude of extremes is diminished as we traverse the finite modes of the EMD. The final two poles of the IMF have been identified. The stamp is a color photograph, as anticipated. To apply the DCT transformation, an image must initially be separated into its RGB components. The image is compressed utilizing the DCT's energy compaction mechanism. The image has been resized, a watermark has been applied, and it has been converted into binary data suitable for incorporation into the audio signal. The watermarked color image can be extracted from the watermarked audio signal by reversing the order of execution.

2.PROPOSED WATER MARKING ALGORITHM

The proposed technique entails incorporating a binary data signature into an audio broadcast. The process of converting an image into its binary representation. The initial step involves generating "frames," or distinct segments of audio data. Subsequently, we compute the Empirical Mode Decomposition (EMD) for each individual image. The watermark bits are currently retained while EMD produces a new IMF. After identifying the final two ends of the IMF, the watermark bits are integrated into the data through Quantization Index Modulation (QIM). The result is that the final IMF incorporates a new element on each side. The frames are organized using an EMD in reverse.

a. The watermark signal constitutes one of the individual frames in the final output. A compressed color image is utilized to extract the requisite binary bits. Employing the energy compaction capability of the Discrete Cosine Transform, we can reduce the image's size. Identifying the optimal compression technique facilitates a superior Peak Signal-to-Noise Ratio (PSNR). The subsequent step involves eliminating the signature and converting the data into binary format. The QIM coefficients are obtained by separating the original audio signal from the watermarked version. This is a platform for purchasing music with applied watermarks. The QIM factors are employed to address the deficiencies resulting from absent binary digits. After the binary data is converted into DCT coefficients, the procedure is repeated. I possess a full-color photograph. A comprehensive elucidation of the proposed audio watermarking technique is as follows:

b. Watermark Preparation

- To apply watermark data to an audio stream, the sequence must first be converted to binary. This binary image depicts the logo's design.
- The image must have its red, green, and blue components eliminated.
- Subimages measuring 8x8 are generated by amalgamating segments of the primary image.
- The Discrete Cosine Transform (DCT) is evaluated at the subblock level.
- The DCT coefficients are quantified using quantization matrices. Dividing each cell in the DCT-transformed matrix by its corresponding cell in the quantization matrix allows for rounding the resulting fraction to the nearest integer.
- Once factors are quantified, their significance is effectively diminished. The DCT's energy compaction function arranges the coefficients in a zigzag pattern. Additional factors are eliminated, and the quantity of factors is ascertained based on the intended compression ratio.
- Upon conversion to binary, the numerals are employed to produce a sequence.

c. Watermark Embedding

The integration of the watermark into the host's audio stream is achieved through the binary code produced during the watermark's processing. Figure 1 depicts the fundamental concepts of watermark embedding. The subsequent sections offer further elaboration on these subjects.

The segregation of the presenter's primary audio stream is imperative.

Empirical mode decomposition (EMD) enables the generation of multiple intrinsic mode functions (IMFs) from a singular input image.

The Quadratic Interpolation Method (QIM) is employed to incorporate the binary series at both the upper and lower ends into the final Intrinsic Mode Function (IMF).

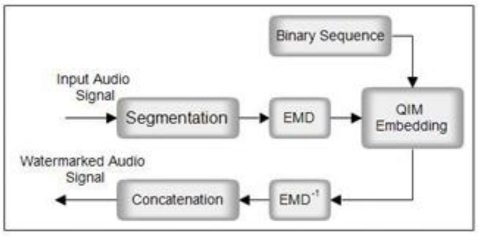


Fig-1: Watermark Embedding

The preceding IMF extremum variable, ei, represents the host signal. E* denotes a variable that signifies a modified IMF extremum. A plus sign (+) is employed when ei represents the largest number, whereas a minus sign (-) is utilized when it does not. S selects an intensity value for embedding that renders the watermark inaudible to humans. Acquire the frames by executing an inverse Empirical Mode Decomposition (EMD) on the latest Intrinsic Mode Function (IMF).

To generate the watermarked audio stream, merely amalgamate the compiled watermarked frames.

Watermark Extraction and Reconstruction

The binary code is derived from the audio output containing the watermark. The QIM values extracted from the audio must be transformed into binary prior to the retrieval of the watermark. Figure 2 illustrates and elucidates the primary procedures for stamp removal. Extraction of QIM coefficients is permissible solely from a primary audio stream devoid of watermarks. The binary integers of the QIM coefficients are aligned and concatenated to generate the stamp binary sequence. The color image is separated into its RGB components to facilitate the creation of the watermark. Figure 3 illustrates and elucidates the fundamentals of watermark restoration.

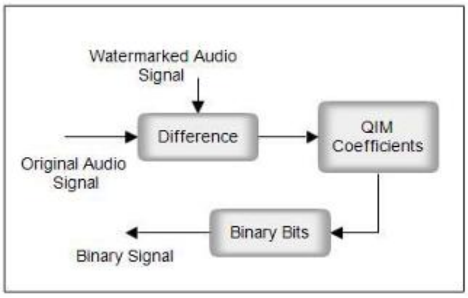


Fig-2: Watermark Extraction

Removing the watermark

- Red, green, and blue constitute the three distinct colors that comprise the binary code.
- The binary representation of each component is converted into its decimal equivalent.
- The 8x8 SB transformed images are generated by initially organizing the coefficients in an inverted zigzag sequence and subsequently incorporating zeros based on the compression ratio.
- To obtain the DCT coefficients, multiply each element of the quantization matrix by the corresponding altered 8x8 subimages.
- The Inverse Discrete Cosine Transform (DCT) algorithm can generate subimages.
- The red, green, and blue components of the watermarked image are produced by amalgamating the subimages.

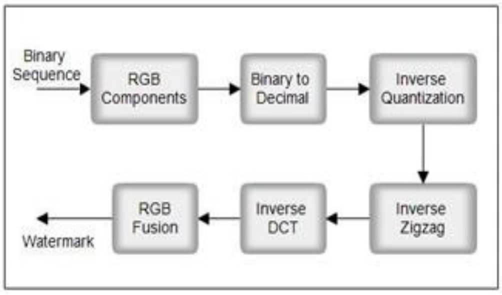


Fig-3: creating watermarks

3.RESULTS AND PERFORMANCE ANALYSIS

The test utilizes ten distinct "WAV" audio files. The collection comprises numerous additional files beyond those enumerated, including "crabby.wav," "dadada.wav," "duck.wav," "formal.wav," "helloo.wav," "powers.wav," "sillyvoice.wav," "spock.wav," "spooky.wav," and "trek.wav. The sizes of the files are denoted in kilobits, and their durations are specified in seconds.

Each of them is capable of managing 64 Kbps. A complete video is available for viewing online. A watermark of M_N = 792792 pixels, or 1296 bits, is embedded in the binary representation of the Hewlett-Packard (HP) logo. The file size is 52.5 kilobytes. Figure 4 depicts a seal. Subsequent to the addition of the HP logo, the audio file became distorted.The wav file was subjected to multiple forms of attack, including Gaussian distortion, filtering, cropping, resampling, and the addition of pink noise. The results are depicted in the graph below. Figure 1 illustrates the conversion of unwatermarked signals into watermarked signals.

This modification is imperceptible due to its considerable distance.

Factors such as the data storage capacity of the audio watermarking system, its resilience against known vulnerabilities, and its visibility can ascertain its efficacy. The security of the scraby.wav file in Table 1 will be evaluated using the Bit Error Rate (BER). An efficient watermark detection system will exhibit a low Bit Error Rate (BER). Calculate the Bit Error Rate utilizing the third method.

Fig-4: HP is a corporate name.

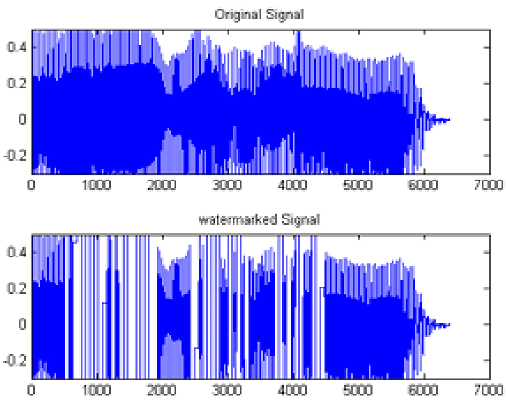


Fig-5: Dress casually.Both before and after the imprint, a Wav signal was recorded.

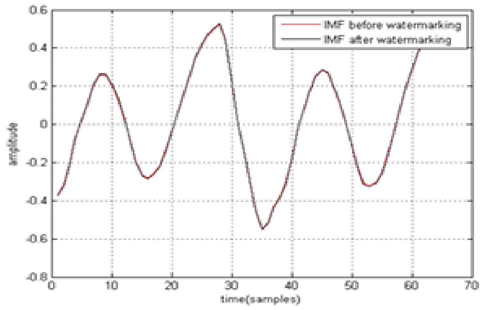


Fig-6: It depicts the IMF before (red) and after (black) scraby.wav.

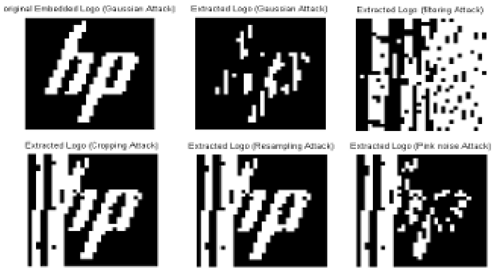


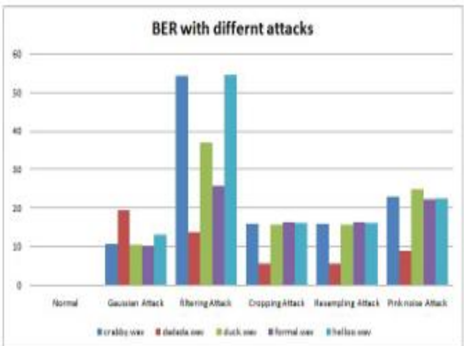
Fig-7: HP logo extraction examples for various assaults

We employ XOR to amalgamate binary watermark images of dimensions M and N bytes. The watermarks are denoted as W(i;j) prior to healing and W0(i;j) subsequent to healing. This study seeks to evaluate the efficacy of watermark recognition in adhering to signal processing protocols.

Table -1: The BER values of the rebuilt logo for each attack are shown in Table 1

Type of Attack	BER
Normal	0.29297
Gaussian Attack	10.6445
filtering Attack	54.3945
Cropping Attack	15.918
Resampling Attack	15.918
Pink noise Attack	23.0469

The wave files scrbby.wav, dadada.wav, duck.wav, formal.wav, and helloo.wav are depicted in Figure 8 alongside their BER values. Figures 8 and 9 demonstrate that the proposed strategy for preventing assaults is ineffective. The BER values, spanning from 54.3945 to 33.6292, indicate that filtering attacks are significant. This results in the loss of approximately one-third of the input data.



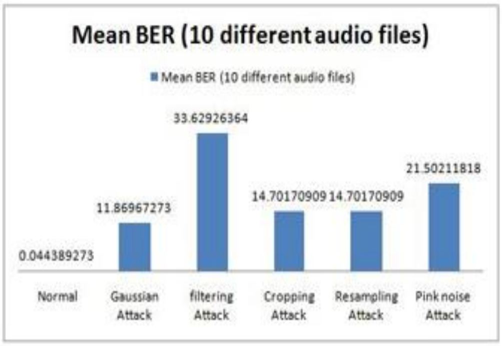


Fig- 8: BER for five different audio samples

4. CONCLUSIONS

This study elucidates the application of empirical mode decomposition for the accurate classification of audio recordings. We employ a watermark that utilizes an end residue characteristic to conceal the watermarking procedure. Upon thorough assessment, it was determined that the method is resistant to the five predominant forms of attacks: resampling, cropping, filtering, Gaussian noise, and pink noise. The proposed attack prevention method is ineffective, as 33.0% of the original audio data was lost during conversion, evidenced by a Bit Error Rate (BER) of 33.6292. This is observable in the filtering results. A Gaussian attack attains an exemplary average Bit Error Rate (BER) of 11.8696. Given these facts, the proposed strategy is more likely to succeed than the Gaussian assault.

REFERENCES

[1] C. P. Talele, Dr A. M. Patil, "Audio Watermarking via EMD "Review", International Journal of Engineering Science and Computing,Volume 6, Issue No. 8, Pages 2658-2661, August 2016.

[2] http://www.sillyhumor.com/answer/classicanswerin gmachinemessages.html

[3] K. K. Gajula, "Designing an Ethical Data Science Framework for Responsible and Transparent AI Integration," Emerging Trends in Digital Transformation, vol. 1, no. 1, pp. 74–87, 2025.

[4] http://www.externalharddrive.com/waves/animal/in dex.html

[5] J. Cox and M. L. Miller, "The first 50 years of electronic watermarking," J. Appl. Signal Process., vol. 2, pp. 126–132, 2002.

[6] M. D. Swanson, B. Zhu, and A. H. Tewfik, "Robust audio watermarking using perceptual masking," Signal Process., vol. 66, no. 3, pp. 337–355, 1998.

[7] S. Wu, J. Huang, D. Huang, and Y. Q. Shi, "Efficiently self-synchronized audio watermarking for assured audio data transmission," IEEE Trans. Broadcasting,vol. 51, no. 1, pp. 69–76, Mar. 2005.

[8] V. Bhat, K. I. Sengupta, and A. Das, "An adaptive audio watermarking based on the singular value decomposition in the wavelet domain," Digital Signal Process., vol. 2010, no. 20, pp. 1547–1558, 2010.

[9] D. Kiroveski and S. Malvar, "Robust spread-spectrum audio watermarking," in Proc. ICASSP, 2001, pp. 1345–1348.

[10] K. Khaldi, M. T.-H. Alouane, and A. O. Boudraa, "Voiced speech enhancement based on adaptive filtering of selected intrinsic mode functions," J. Adv. in Adapt.Data Anal., vol. 2, no. 1, pp. 65–80, 2010.

[11] K. K. Gajula, "Advanced Stress Level Forecasting Using Quantileoptimized Regression Forest Models", int. J. Eng. Res. Sci. Tech., vol. 21, no. 4(1), pp. 1–7, Dec. 2025, doi: 10.62643/ijerst.2025.v21.n4(1).pp1-7.

[12] K. K. Gajula, "AI-Augmented Web Development: A Competency Framework for the Next Generation Workforce," Journal of Science Engineering Technology and Management Science, 2025.