Enhanced Audio Watermarking with WSOLA for Robustness against Speed Changes

**Ankit Ghosh**

**Birla Institute of Technology**

**Mesra, Ranchi**

**btech10185.20@bitmesra.ac.in**

**Sameer Damkondwar**

**Birla Institute of Technology**

**Mesra, Ranchi**

**btech10239.20@bitmesra.ac.in**

**Prakhar Kumar Singh**

**Birla Institute of Technology**

**Mesra, Ranchi**

**btech10031.20@bitmesra.ac.in**

*Abstract*—*In the realm of digital audio watermarking, the embedding of significant watermark signals into original audio signals is pivotal for subsequent extraction as needed. This technology holds immense value in domains such as copyright protection, content authentication, secure communication, and anti-counterfeiting endeavors. With the widespread availability of multimedia software and audio editing tools, the ease of audio modification has increased significantly. However, a majority of existing digital audio watermarking algorithms fall short in their ability to withstand commonplace operations such as shifting, editing, tone manipulation, and splicing. To address this limitation, this paper presents a novel audio watermarking algorithm explicitly designed to withstand variable speed attacks on audio signals. Leveraging the WSOLA (Waveform Similarity Overlap and Add) principle, the proposed algorithm intelligently embeds a spread spectrum watermark within non-overlapping regions during the audio speed change process. Extensive experimentation demonstrates the algorithm's heightened robustness and imperceptibility, making it highly suitable for practical applications.*

*Keywords—audio watermark, audio speed change, spread spectrum, robustness, imperceptibility.*

# Introduction

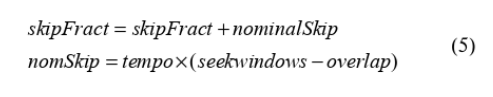
In today's digital age, the rapid growth of big data and the widespread availability of music streaming platforms like QQ Music and NetEase Cloud Music, along with various audio editing software such as Adobe Audition, have made it incredibly convenient and easy for people to access and manipulate digital media. However, this convenience has also given rise to significant challenges in terms of copyright infringement and piracy, posing a serious threat to the rights of copyright owners. To combat these issues, there is an urgent need for effective methods to identify and protect digital media content from unauthorized use. Digital watermarking technology has emerged as a promising solution for copyright tracking and verification. By imperceptibly embedding special watermark information into digital media, it allows for the extraction of the watermark when copyright disputes arise, enabling the tracking and validation of copyrights.

Digital watermarking technology encompasses various forms, including text, image, audio, and video watermarking. In the context of audio watermarking, there are two main processes: embedding and extracting. The goal is to invisibly embed watermark information of significant importance into digital audio without compromising its original quality and usability. This watermark remains embedded within the audio as an integral part of the data, and can be extracted when necessary to verify the copyright.

However, the challenges of copyright protection extend beyond embedding watermarks. During the transmission and usage of audio, various attacks can compromise the integrity of the watermark. If the watermark information can be easily destroyed or altered, the purpose of copyright protection is defeated. Therefore, it is essential to develop audio watermarking algorithms that are robust against attacks while ensuring that the watermark remains imperceptible. Particularly, the robustness of existing audio watermarking algorithms against variable speed attacks needs improvement.

To illustrate this, let us consider an example: Imagine a popular song that has been digitally watermarked to protect its copyright. Now, suppose someone decides to use this song in their own project but wants to modify it by increasing or decreasing its playback speed. During the speed change process, the watermark embedded in the original song may become distorted or even lost entirely. As a result, it becomes impossible to accurately extract the watermark and validate the copyright, leading to potential misuse and infringement.

# To address this problem, this paper proposes a novel audio watermarking algorithm that aims to enhance the robustness of watermarks against variable speed attacks. By leveraging innovative techniques based on the principles of audio processing, the algorithm intelligently embeds the spread spectrum watermark into non-overlapping regions when the audio speed is modified. This ensures that the watermark remains intact and extractable, even in the presence of variable speed changes.

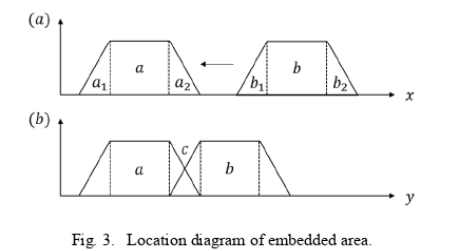
While several audio watermarking algorithms have been proposed, they often struggle to preserve the integrity of the watermark when facing speed changes. Through a thorough analysis of existing techniques and considering their limitations, this paper introduces a new and improved audio watermarking algorithm. This algorithm prioritizes robustness against variable speed attacks while maintaining imperceptibility. Experimental evaluations will be conducted to validate the algorithm's effectiveness and compare its performance with existing methods.

In summary, this paper addresses the critical need for robust audio watermarking algorithms to combat copyright infringement in the digital age. By proposing an innovative approach specifically designed to withstand variable speed attacks, the algorithm aims to enhance the success rate of copyright protection while ensuring the watermark remains imperceptible. The subsequent sections will provide a detailed explanation of the proposed algorithm, experimental results, and comparisons with existing methods, ultimately demonstrating its effectiveness and practicality in real-world scenarios.

II. DESIGN AND IMPLEMENTION OF SYSTEM ANALYSIS

The audio watermarking algorithm is designed to invisibly embed watermark information into audio files to protect copyrights. When copyright disputes arise, the watermark can be extracted to provide evidence. To ensure the watermark remains imperceptible and resistant to variable speed attacks, a novel audio watermarking algorithm based on the WSOLA algorithm is proposed. In the pre-processing stage, a set of pseudo-random sequence matrices, known as sequence unit orthogonal matrices, are generated using cyclic shift, full rank decomposition, and Schmidt orthogonalization techniques. This method enhances the randomness of watermark embedding and improves numerical stability, thereby enhancing watermark robustness. The WSOLA algorithm is then employed to identify overlapping and non-overlapping regions in the audio affected by changes in speed, aiding in determining the optimal positions for watermark embedding. The watermark is embedded using a combination of the WSOLA algorithm and the discrete cosine transform (DCT) strategy. The overall framework of the watermarking algorithm is illustrated in diagram below:

## A. Embedding Algorithm

****In Figure 4, (a) represents the original audio input, denoted as x, which does not have any speed changes. (b) represents the resulting audio output, denoted as y, after applying speed changes to x. First, process the first frame, copy the data of area **a** to output **y**, and then a**2**and b1, data is superimposed, the superimposed result ***C*** is copied to output ***y***, and finally the data of area b is copied to output ***y***, that is, the processing of one frame is completed. The algorithm in this chapter selects the non-overlapping regions in each frame for operation, such as region ***a*** and region ***b***

In the embedding phase, we have tempo (speed parameter), overlapLength (length of overlapping area denoted as 'c'), seekLength (length of the offset search range to find 'b₁' with the highest correlation to 'a₂'), and seekwindowLength (sum of frame data length and overlap length), b₁ the position of the region relative to the starting position of the frame is offset. We also use nominalskip, which determines the step size for frame processing based on formula (5).

x**i** represents the 32-bit coefficient segment after DCT transformation in the **i***th* frame of the host signal. y is the resulting signal after embedding a pseudo-random b-bit watermark sequence. The watermark sequence, represented by pt, corresponds to the b-bit watermark. The index value of pt is the decimal representation of the b-bit binary watermark sequence. The embedding process follows

Where, w***i***, is the parameter that controls the embedding strength. The sign operation is a symbolic function when x***i***pT*k* **≥** 0, the sign value is 1. When x***i***pT*k* ≤0, the sign value is -1. The embedding formula is used to incorporate the watermark information. After embedding, the audio signal y(n) with watermark is constructed through inverse process of audio signal segmentation. It determines the perceptual quality of the watermarked signal through the strength parameter, w***i***. A smaller w***i*** value improves imperceptibility but reduces watermark robustness, while a larger w***i*** value compromises imperceptibility but enhances robustness. To strike a balance, w***i*** should be chosen carefully to ensure both imperceptibility and robustness. Its calculation method is given in formula :



The specific implementation steps for the anti HSI audio watermark embedding algorithm are as follows: -

1. For randomly generated seed sequence b0, cyclic shift, generating the matrix b0,b1,....b*l*-1, . By Schmidt orthogonalization of the matrix the matrix P is obtained. Then the pseudo-random sequence matrix is obtained from selection 2*b* sequences from P, i.e., p1,p2,...p2*b*.
2. The embedded area of each frame is then transformed by DCT and it records it as vector x1,x2,...xN.
3. Now by using the embedding formula the watermark information which is to be embedded is embedded into the original audio segment.
4. After all the audio segments are embedded with watermark, by using inverse DCT transform the audio signal containing watermark is reconstructed.

## B. Extracting Algorithm

During the extraction phase, a random seed sequence, denoted as b, is used to extract a watermark from the audio signal y(n). The audio signal y(n) represents either the original audio or the attacked audio after the watermark has been embedded. To obtain the watermarked segment of 32-bit DCT coefficients, denoted as y₁, y₂,..., y*N*, the embedding method described earlier is employed. Simultaneously, a pseudo-random watermark sequence p1,p2,…, p2*b* is generated using another random seed sequence b*o*. The extraction process is performed according to the following formula:

By converting the index value into binary, the extracted watermark sequence is obtained. Here j=1,2,...,2*b*, and t is the embedded pseudo-random sequence p*t*.

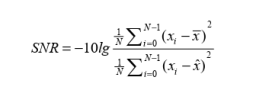
The process of correctly extracting watermark information from the audio containing watermark based on WSOLA algorithm is the reverse of process of watermark embedding algorithm. Following are the implementation as follow:-

1. The position of the watermark is to be extracted from the known values overlapLength, seekwindowLength, seekLength and the offset value of each frame, in each frame is determined.
2. The matrix P is obtained by Schmidt orthogonalization of the matrix. For randomly generated seed sequence b0, cyclic shift, generating matrix b0,b1,....b*l*-1 . From P select 2b sequences p1,p2,....p2b, as the result of pseudo-random watermark sequence generation, the pseudo-random sequence matrix is obtained.
3. The area is transformed by DCT which is to be extracted in each frame, and is recorded it as vector x1,x2,...xN.
4. Now by using the extraction formula the watermark is extracted from each frame.

IV. EXPERIMENT

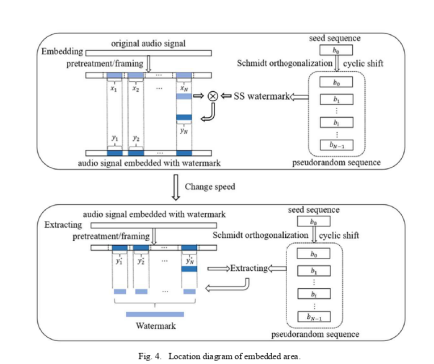
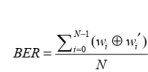
Imperceptibility refers to the absence of noticeable distinctions between the original audio, free of any watermark, and the audio containing an embedded watermark. In other words, the process of embedding an audio watermark should not affect the quality or value of the audio content. The Signal-to-Noise Ratio (SNR) serves as the benchmark for evaluating the imperceptibility.

The SNR (Signal-to-Noise Ratio) measures the level of audio distortion that occurs when a watermark is embedded. A higher SNR indicates lower audio distortion, meaning that the watermark is more imperceptible. The calculation of SNR is defined in formula (9). In the formula, x̄ represents the original audio signal, and x̂ represents the audio signal with the embedded watermark.



Robustness refers to the ability of watermark-embedded audio to retain its integrity even when subjected to modifications or attacks. The higher the number of extracted watermark bits that align with the original watermark, the stronger the robustness. Bit Error Ratio (BER) serves as a commonly employed evaluation metric for measuring robustness.

BER determines the percentage of differing bits between the original and extracted watermarks. Lower BER implies better robustness, indicating higher accuracy in watermark extraction. BER is calculated using given formula with w*i* as the original watermark, w*i*' as the extracted watermark, "⊕" as the bit-wise XOR operation, and "N" as the total watermark length. The location diagram of embedded area is shown below:



**Experimental environment:**

11th Gen Intel Core i5, 2.40 GH z - x64 processor, 8 GB RAM, 64-bit operating system

**Algorithm Language:**

C++, Visual Studio 2019 programming environment.

**Experimental Audio:**

We used a WAV file named "sing.wav" for the experiment. The audio file had a duration of 2 minutes.

**Experiment Results:**

We performed a speed change test on the "sing.wav" file and obtained the results presented in Table 1. Based on our findings, we can conclude that the algorithm demonstrated robustness against variable speed attacks.

TABLE 1 AUDIO SING ROBUSTNESS EXPERIMENT

|  |  |  |
| --- | --- | --- |
| **Audio** | **Attack Type** | **BER** |
| Sing.wav | 0.5x | 0.052 |
| 1.0x | 0.049 |
| 2.0x | 0.069 |

V. CONCLUSION

This paper describes an audio watermarking algorithm that utilizes the WSOLA algorithm to resist variable speed attacks. It incorporates a spread spectrum-based pseudo-random sequence to enhance the randomness and robustness of the 256 watermark. The WSOLA algorithm analyses the relationship between the audio before and after the speed change to identify invariants for watermark embedding and extraction.

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.**