STAW: An Audio Watermarking in Short Time DCT for Copyright Protection

Shiqiang Wu
Institute of Automation, Chinese Academy of Sciences.
University of Chinese Academy of Sciences
Beijing, China
Wushiqiang2018@ia.ac.cn

Hu Guan, Ying Huang, Shuwu Zhang*
Institute of Automation
Chinese Academy of Sciences
Beijing, China
{hu.guan, ying.huang, shuwu.zhang}@ia.ac.cn

***Abstract*—An audio watermarking scheme STAW in short time discrete cosine transform domain using spread spectrum method is proposed in this work. Here, we framed the audio signal in the time domain with overlapping segments which then transferred as DCT represented. according to binary watermark, we generated some pseudo-random float sequences as embedding message, which are orthonormal and energy uniform, since the DCT coefficients is not integer or binary. Additionally, in order to improve imperception an adaptive mechanism for embedding strength is proposed. The experimental results show that STAW provides high robustness against common audio attack as well as high imperceptibility.**

Keywords-audio watermarking; short-time transform; DCT; audio perceptual quality; copyright portection.

I. INTRODUCTION

In recent years, the development of communication and multimedia technologies has made it easier for people to share their work. However, on the other hand, these technologies also have made digital multimedia data on the Internet at significant risk of being reproduced, manipulated and distributed. The nature of digital information in the network that can be copied without loss requires new and improved methods for copyright protection. The copyright infringement forms of multimedia resources are not consistent, that is why researches of watermarking technology on audio and image are different. In this paper, we limit our attention to audio data, including music, speech and audio tracks in the video.

By integrating perceptible or imperceptible copyright marks on the original digital signals, the digital watermarking is a promising technology to avoid illegal copying[1]. Nevertheless, it is challenging for researchers owing to malicious user who can attack the digital signals intentionally or unintentionally. As a result, it is difficult to protect digital information, especially the audio. The audio watermarking is a watermarking technique specifically for protecting ownership of audio data. It is not effortless to design an idea l audio watermarking scheme which embed a watermark is imperceptible, robust, secure and highly capable[2]. The above four requirements, also known as performance criteria, can approximately characterize the effectiveness of an audio watermarking scheme. Specifically, first, imperceptibility characterizes the fidelity of audio data embedded watermarks, indicating that the embedded watermarks should avoid to introduce perceptually distinguishable changes to the original signal. Sometimes, fidelity, transparency, and inaudibility have the same meaning with imperceptibility. Second, robustness denotes the availability of successful watermark extraction even when the embedded signal has been attacked. It is the most significant and complicated feature because of the diversity of attacks. It is no exaggeration to say that no watermarking scheme can handle all attacks and maintain robustness. Third, security refers to that only authorized people are able to extract the watermarks. At last, the amount of information that can be embedded into the given original signal is called capacity. Among of the four criteria, imperceptibility and robustness determine the key performance of most existing audio watermarking schemes.

Several digital audio watermarking schemes were based on the incorporation and modifications of techniques from other research areas, for instance, spread spectrum (SS) from communication theory[3], patchwork methods from image watermarking[4], and quantization index modulation (QIM) from general multimedia format encoding[5]. It can be seen that many audio watermarking schemes have been categorized according to the signal processing techniques for embedding. Nonetheless, there are other binary options to characterize different schemes. Based on process of extraction, schemes may be two types, to be specific, “Non-blind watermarking” extraction of watermark using original signal [6] and “blind watermarking” extraction of watermark without use of original signal[7]. The watermarks may be embedded in time domain[8] or transform domain like DCT domain[6], DWT domain[10], etc. The main features of the proposed scheme in this paper are as follows: (1) Audio signal framed and then every frame was transformed to DCT domain; (2) The procedure of extraction is blind without using the host signal; And uncommonly (3) we generated some low dimension pseudo-random float sequences w.r.t binary watermark as embedding message. The proposed method has better imperceptibility compared to previous experiment in long time DCT domain, and the robustness has not abated.

The remainder of the paper is organized as follows. The related works about spread spectrum and others this work related are reviewed and discussed in Section II. The proposed scheme STAW is presented in Section III. The experimental results are shown in Section IV. In Section V we conclude the paper.

II. RELATED WORKS

978-1-7281-8138-7/20/$31.00 ©2020 IEEE
DOI 10.1109/ICCST50977.2020.00011

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on April 08,2022 at 03:49:07 UTC from IEEE Xplore. Restrictions apply.

A. Related works about SS-based audio watermarking methods.

SS-based audio watermarking methods was originally proposed in [11]. The details of algorithm is as follows.

Assume that a segment of the original audio signal in a given domain is denoted **x**, **y** is the corresponding watermarked audio segment and **p** is a pseudo-noise (PN) sequence. Here, **x**, **y** and **p** are vectors whose dimensions are equal. The elements of **p** take values from {-1, +1}, and **p** is independent of **x.** Now embed *w* which is a watermark bit in {-1, +1} into the original audio **x** by

$$\boldsymbol{y} = \boldsymbol{x} + \alpha w \boldsymbol{p} \tag{1}$$

where α is a positive parameter which controls the watermark strength. At the extraction, the embedded watermark bit is extracted from **y** by using **p** as a secret key. As follows, we have

$$\begin{aligned} z &= \frac{\boldsymbol{y}\boldsymbol{p}^T}{\boldsymbol{p}\boldsymbol{p}^T} \\ &= \alpha w + x' \end{aligned} \tag{2}$$

where the superscript ^{T} denotes transpose operation, and

$$x' = \frac{\boldsymbol{x}\boldsymbol{p}^T}{\boldsymbol{p}\boldsymbol{p}^T} \tag{3}$$

If **p** is independent of **x**, then $x' = 0$. In this case, we can extract the embedded watermark bit by

$$\widehat{w} = \text{sign}(z) \tag{4}$$

Clearly, this audio watermarking method is a blind method since it does not need information of the original audio in the extraction procedure.

There is an obvious approximation in the above algorithm, that in practice there is no sure that x and p are independent. When x' is not zero, it will act as an interference from the host, which is called host signal interference. There are two ways to guarantee that the embedded watermark bit can be extract accurately. Increasing α is useful, but will degrade the imperceptible performance. To increase the length of the audio segment is another possible way, but it reduces embedding capacity. Although we can increase embedding capacity though significantly compromising robustness, this is a dilemma and not necessarily worth it.

Then researchers proposed to increase embedding capacity by embedding multiple watermark bits into an audio segment [12]. To achieve this, a set of near orthogonal PN sequences is generated, and each of them is used as an embedded sequence. However, the host signal interference problem still remains. Additionally, when encountering a strong attack, the approach might not give a good result. And the effectiveness of using near-orthogonal PN sequences would be compromised.

Recently, to eliminate the host signal interference, a new embedding function is applied to watermarking

$$\boldsymbol{y}_n = \boldsymbol{x}_n + \alpha \left(\frac{\boldsymbol{x}_n \boldsymbol{p}_t^T}{|\boldsymbol{x}_n \boldsymbol{p}_t^T|} \right) \boldsymbol{p}_t \tag{5}$$

where subscript n is the index of audio segments and p_t is the t-th orthogonal PN sequences. At corresponding extraction stage, define

$$D_{n,j} = \left| \boldsymbol{y}_n \boldsymbol{p}_j^T \right| \tag{6}$$

where D is a $N \times T$ dimensional matrix, N and T are range of n and t, respectively. Assume the $j = t$, substituting (5) into (6),

$$\begin{aligned} D_{n,j} &= \left| \left(\boldsymbol{x}_n + \alpha \left(\frac{\boldsymbol{x}_n \boldsymbol{p}_t^T}{|\boldsymbol{x}_n \boldsymbol{p}_t^T|} \right) \boldsymbol{p}_t \right) \boldsymbol{p}_t^T \right| \\ &= \left| \boldsymbol{x}_n \boldsymbol{p}_t^T + \alpha \left(\frac{\boldsymbol{x}_n \boldsymbol{p}_t^T}{|\boldsymbol{x}_n \boldsymbol{p}_t^T|} \right) \|\boldsymbol{p}_t\|^2 \right| \\ &= |\boldsymbol{x}_n \boldsymbol{p}_t^T| \left(1 + \alpha \left(\frac{\|\boldsymbol{p}_t\|^2}{|\boldsymbol{x}_n \boldsymbol{p}_t^T|} \right) \right) \end{aligned} \tag{7}$$

the last equal sign established since $\alpha, \|\boldsymbol{p}_t\|^2$ and $|\boldsymbol{x}_n \boldsymbol{p}_t^T|$ are positive numbers. When $j \neq t$, we can also have

$$\begin{aligned} D_{n,j} &= \left| \left(\boldsymbol{x}_n + \alpha \left(\frac{\boldsymbol{x}_n \boldsymbol{p}_t^T}{|\boldsymbol{x}_n \boldsymbol{p}_t^T|} \right) \boldsymbol{p}_t \right) \boldsymbol{p}_j^T \right| \\ &= \left| \boldsymbol{x}_n \boldsymbol{p}_j^T + \alpha \left(\frac{\boldsymbol{x}_n \boldsymbol{p}_t^T}{|\boldsymbol{x}_n \boldsymbol{p}_t^T|} \right) \boldsymbol{p}_t \boldsymbol{p}_j^T \right| \end{aligned} \tag{8}$$

Since $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_T$ are orthogonal, $\boldsymbol{p}_t \boldsymbol{p}_j^T = 0$. Thus, (8) rewritten as

$$D_{n,j} = \left| \boldsymbol{x}_n \boldsymbol{p}_j^T \right| \tag{9}$$

In (7), due to the independence between $\boldsymbol{x}_n$ and $\boldsymbol{p}_t$, we know $\|\boldsymbol{p}_t\|^2 \gg |\boldsymbol{x}_n \boldsymbol{p}_t^T|$. Thus if α is not tiny, we can ensure

$$\left(1 + \alpha \left(\frac{\|\boldsymbol{p}_t\|^2}{|\boldsymbol{x}_n \boldsymbol{p}_t^T|} \right) \right) \gg 1 \tag{10}$$

Besides, the term $|\boldsymbol{x}_n \boldsymbol{p}_t^T|$ in (7) and the term $\left| \boldsymbol{x}_n \boldsymbol{p}_j^T \right|$ in (9) are comparable. Then

$$D_{n,t} \gg D_{n,j}, \forall j \neq t. \tag{11}$$

Hence, (22) indicates that among all $D_{n,j}, j = 1,2, \ldots, T$, the largest one is associated with the embedded sequence $\boldsymbol{p}_t$.

B. Others related work

In order to ensure the independence of embedded PN sequences and audio signal segments, the length of PN sequence in [12] is very long. In that work, they embedded 7 bits watermark by using PN sequence length as 500.

We draw on the idea of product quantization [13] and encode the high dimension space by using the vector of lower dimensions through Cartesian product.

III. Proposed Methodology

Similar to the methods mentioned earlier, our STAW also performs watermarking in the DCT domain. A little difference is that we use a short-time transformation that first divides the frame and then DCT transforms. Fig.1 shows the block diagram of our STAW. Below we will describe in detail each stage of our method.

A. Signal framing and transformation

At the beginning of the algorithm, we first divide the audio the signal as frames by using the hanning window, each frame length is *len_frame*. Then audio frames are transformed from time domain to DCT domain, separately. The result of the cooperation of framing and DCT is a two-dimensional information of time-frequency, denoted as S in

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on April 08,2022 at 03:49:07 UTC from IEEE Xplore. Restrictions apply.

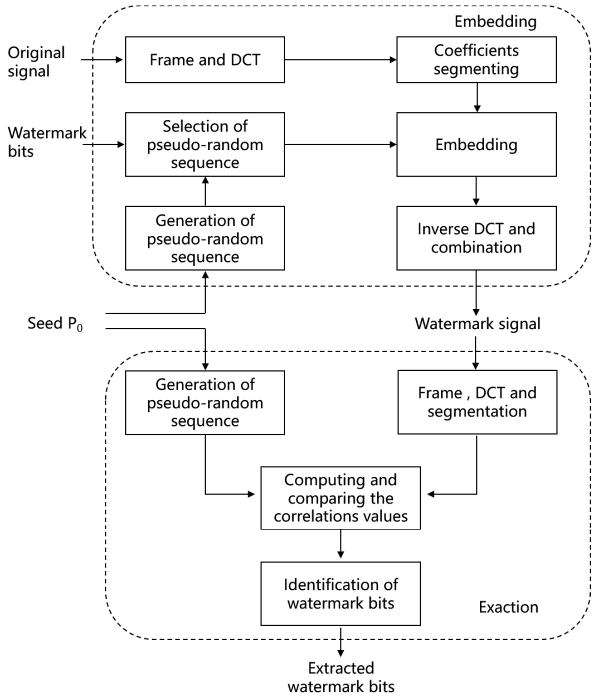

Figure 1. Block diagram of our proposed method STAW.

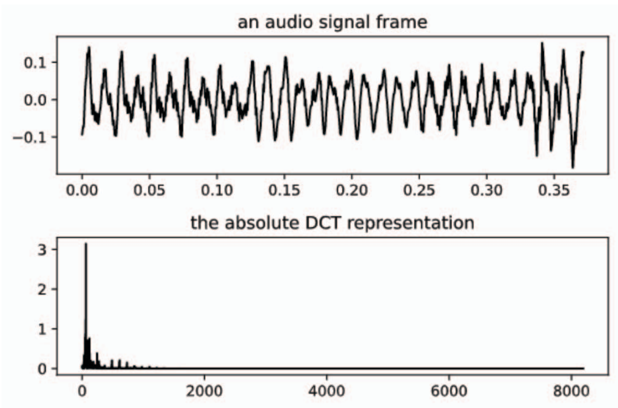

Figure 2. An audio signal frame in time domain and the absolute valued DCT coefficients of the same signal versus frequency index.

the follow parts of this article. This operation is more computationally efficient than short time Fourier transformation and cepstrum. At the same time, since DCT has a higher compression rate, the modification of signal occurs at medium or high frequencies with little effect on audio perception. Fig. 2 shows an audio signal frame in the time domain and in the DCT domain. Fig. 3 shows an audio signal in the time domain and in the short time DCT domain.

B. Watermark embedding

The watermark embedding stage includes pseudo-random sequences generation, embedding strength adaption and DCT coefficients modification.

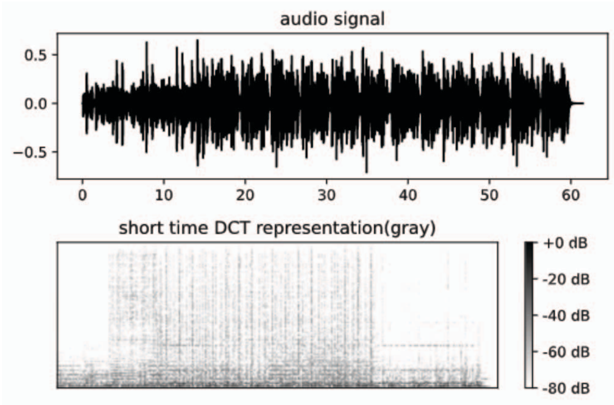

Figure 3. An audio signal in time domain and the short time DCT representation of the same signal.

1) Pseudo-random sequences generation

In order to improve the embedding capacity without affecting the robustness and imperceptibility, we embed multiple watermark bits into an audio signal segment, denoted as n_s. the n_s bits will be used to generate some pseudo-random orthogonal sequences for embedding. A pseudo-random sequence is inserted into the audio segment.

In order to represent n_s bits watermark information, we need to generate at least 2^{n_s} mutually orthogonal sequences. Our embedding and extraction procedure also require the absolute value of each element of these sequences to be approximate.

Let $\boldsymbol{p}_0$ denote a randomly generated seed sequence of dimension $1 \times 2^{n_s+1}$ with all elements in $\{-1, +1\}$. And P is a matrix where each row is obtained by rolling $\boldsymbol{p}_0$. According to Gram-Schmidt orthogonalization method[14], we can generate 2^{n_s} orthogonal sequences, denoted as $\boldsymbol{m}_1, \boldsymbol{m}_2, \dots, \boldsymbol{m}_{2^{n_s}}$.

Note that these sequences are not only used at the embedding stage but are necessity to extraction. However, there is no need to pass them to extractor directly. Instead, we only need to pass the seed $\boldsymbol{p}_0$, and then the extractor can generate all the pseudo-random sequences. In this case, the seed $\boldsymbol{p}_0$ is used as a private key for watermark extraction.

There are a few points need to be specifically explained. Firstly, different from the scheme in [12], where a method of generation for orthogonal sequences would get a diagonal matrix, where first diagonal element would be 1 or -1, and others are ones, or in [14], where get a set of near-orthogonal pseudo-noise sequences, our method will get a group of float and approximate balanced sequences of which the usage will improving robustness. This work achieves this goal through rolling the $\boldsymbol{p}_0$ and Gram-Schmitt orthogonalization method. Furthermore, since the result of our method is a $2^{n_s+1} \times 2^{n_s+1}$ dimensional matrix, there would be 2^{n_s+1} orthogonal sequences ideally. As we cannot guarantee that the matrix P is a nonsingular matrix, we have to compromise on efficiency for the robustness of the algorithm.

2) Embedding strength adaption

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on April 08,2022 at 03:49:07 UTC from IEEE Xplore. Restrictions apply.

One advantage of using short-term transforms is that the transform domain coefficients with respect to time, which makes it possible to embed watermarks of different strength at different times.

We used a new parameter w_t to control the embedding strength with respect to time. The next section will mention that another strength parameter with respect to frequency. Since the audio energy difference between different frames will be very large, for example, in a movie audio track, the energy of the audio track of the battle in the battlefield will be much larger than the energy of the audio track when the actor sleeps. Therefore, the embedding strength parameter w_t proportional to the audio energies in each frame may make the watermark information in some frames minuscule or tremendous, which will negatively affect the extraction stage.

In order to avoid such negative affection, in this paper we have chosen stepped weights. Firstly, the audio energy of all frames is sorted. Secondly, the weight w_t obtained in the top 20% is 2.0, the subsequent 20% is 1.5, then the next 20% is 1, then 0.5, and the remaining 20% weight is 0.1.

3) DCT coefficients modification

After framing and transformation, a two-dimensional representation S of the audio signal is obtained, where each column denotes the DCT coefficients of an audio frame. The coefficients modification discussed below only operates in the column of S.

Since the components of low and high frequency are easily vulnerable to compression and filtering attacks, we only can modify the coefficients in middle frequency. There is another reason here. In order to ensure the independence of DCT coefficients and pseudo-random sequences, our algorithm needs to embed watermark information in a flat spectral line diagram.

Let us assume that there are n bits watermark message need embedding. Divide these watermark message into several groups, denoted as g, and each group contains n_s bits. Each group of n_s bits can choose a sequence from the set of generated floats sequences, called $\boldsymbol{m}_t$. Our embedding function is only a little different from (5), which is we use a weight parameter w_k with respect to time to control the embedding strength, which is const when embedding in one column of S.

Firstly, we use $\boldsymbol{A\_f}[k]$ to represent the k-th column in S. We will use its 12.5% element where index is called $start$ as the starting point for embedding, and this value can be changed slightly in the experiment. The coefficients with the index value from $start$ to $start + 2^{n_s}$ is the first embedded audio segment where embedding first group of n_s watermark bits and subsequently select 2^{n_s} coefficients to embed the remaining watermark bits. Each selected sequence of coefficients is called $\boldsymbol{x}_i$ in turn. And let $\boldsymbol{y}_i$ be the watermarked segment responding to $\boldsymbol{x}_i$. We perform watermark embedding as follows.

$$\boldsymbol{y}_i = \boldsymbol{x}_i + \alpha_i w_k \left(\frac{\boldsymbol{x}_i \boldsymbol{m}_t^T}{|\boldsymbol{x}_i \boldsymbol{m}_t^T|} \right) \boldsymbol{m}_t \tag{12}$$

where α_i is a positive parameter to control embedding strength computed by

$$\alpha_i = (\alpha_u - \alpha_l) e^{-c(i-1)} + \alpha_l \tag{13}$$

where α_u, α_l, c is const parameters.

For each $\boldsymbol{A\_f}[k]$ in S, we will embed the same bits.

After modification of coefficients, we will combine the audio segments to S and do inverse DCT to the frames, the last is restoring signals.

C. Watermark extraction

At the extraction end, we received the seed sequence $\boldsymbol{p}_0$ from the embedding end, then generated pseudo-random sequences, $\boldsymbol{m}_1, \boldsymbol{m}_2, \dots, \boldsymbol{m}_{2^{n_s}}$.

Then we will use the same process to complete the audio preprocessing, and divide the short-term DCT coefficients of the extracting signal to obtain a series $\boldsymbol{y}_i'$ s.

As discussed in Section II, the index t of the pseudo-random sequence embedded can be found by

$$\begin{aligned} t &= \underset{j \in \{1,2,\dots,2^{n_s}\}}{\operatorname{argmax}} D_{i,j} \\ &= \underset{j \in \{1,2,\dots,2^{n_s}\}}{\operatorname{argmax}} \left| \boldsymbol{y}_i' \boldsymbol{m}_j^T \right| \end{aligned} \tag{14}$$

If the signal has been attacked, there are differences between $\boldsymbol{y}_i'$ s and $\boldsymbol{y}_i$ s. For instance, the signal has been attacked by noise addition or re-quantization attack, $\boldsymbol{y}_i'$ can be written as:

$$\boldsymbol{y}_i' = \boldsymbol{y}_i + \boldsymbol{n}_i \tag{15}$$

where $\boldsymbol{n}_i$ is a noise or distortion vector introduced by these attacks. Since $\boldsymbol{n}_i$ and $\boldsymbol{m}_j$ are independent of each other and the energy of $\boldsymbol{n}_i$ is much smaller than that of $\boldsymbol{y}_i$, the impact of $\boldsymbol{n}_i \boldsymbol{m}_i^T$ on extraction is small.

We vote the bits extracted from each column as the final result.

IV. Experimental Results

In this section, we will show some experimental results of the proposed audio watermarking scheme STAW. This scheme is also compared with scheme reported in [9] and [12]. In the experiments, we use 50 randomly selected audio clips as host signals. The experiments are performed using Python with numpy, librosa and scipy libraries.

For audio watermarking methods, imperceptibility and robustness are two essential requirements. Imperceptibility and robustness are measured by signal noise ratio (SNR) and detection rate (DR).

SNR value indicates the perceptual of sound, can be defined as

$$SNR(s_1, s_2) = 10 \log_{10} \left(\frac{\|s_1\|^2}{\|s_1 - s_2\|^2} \right) \tag{16}$$

DR can be defined as

$$DR = \frac{\text{\# of watermarks correctly extracted}}{\text{\# of watermarks embedded}} \times 100\% \tag{17}$$

A. Comparsion of imperceptibility with existing works

In fact, there is not a very effective way to compare the perceptual quality between different schemes. By reducing the

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on April 08,2022 at 03:49:07 UTC from IEEE Xplore. Restrictions apply.

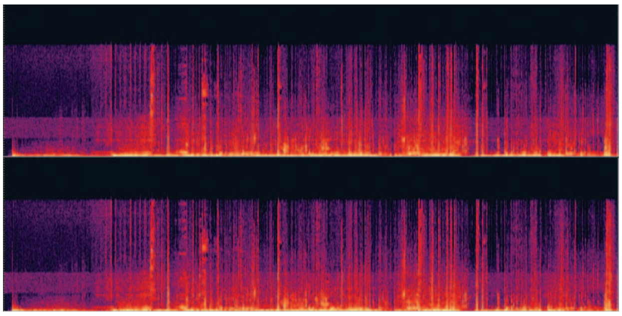

Figure 4. Watermarked signal spectrum using method in [12]

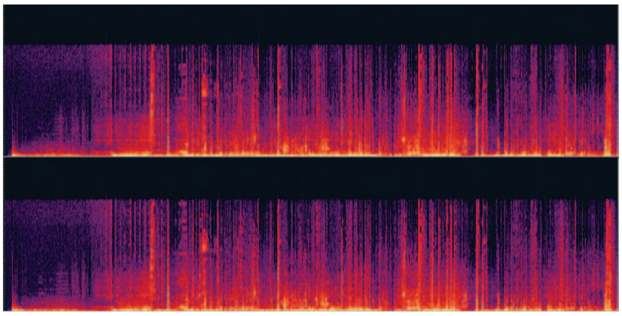

Figure 5. Watermarked signal spectrum using the proposed method

embedding strength, the SNR can ideally be infinity, when no information has been embedded.

But we can find the advantages of the proposed scheme STAW from the spectrogram as shown in Fig. 4 and Fig. 5.

Watermarked signal spectrum of the method in [12] has a haze band running through the entire time axis in the embedding area, because the embedding strength of this method is invariant with respect to time. The proposed method avoids the defect by introducing parameter w_t.

B. Comparsion of robustness with existing method

The following attacks are used in experiments to evaluate robustness:

- *None attack*: The watermarks are extracted from the watermarked signal without any attacks.
- *Re-quantization attack:* Each sample of the watermarked signals is re-quantized from 16 bits to 8bits.
- *Noise attack*: Random noise is added to the watermarked signals, where the SNR are 15dB and 10dB, respectively.
- *Loss compression*: MP3 compression is performed on the watermarked signals, where the bit rate are 192kbps and 128kbps, respectively.
- *Cutting off attack:* Cut off some sample points of the watermarked signals. Here we cut off the first 10% of the sample points or the last 30% of the sample points of the watermarked signals.

We compare the robustness of the proposed method STAW with those in [9] and [12] at the same watermarking and under the same SNR at 30dB. This is because the method in [9] cannot change the embedding strength and its SNR is approximately 30dB. For our method, the simulation parameters are:

TABLE I. DR OF THE PROPOSED STAW AND THE METHOD IN [9], [12]

Attacks	DR(%)		
	[12]	[9]	Proposed STAW
None	100	100	100
Re-quantization	100	94.4	100
Noise(15dB)	97.4	98.9	100
Noise(10dB)	94.6	96.3	98.4
Mp3(192kbps)	100	97.4	100
Mp3(128kbps)	100	96.3	100
Cut-off(10%)	96.3	83.7	100
Cut-off(30%)	93.6	77.4	100

$len\_frame = 8192$, $\alpha_u = 0.15$, $\alpha_l = 0.075$, $c = 0.002$, $n_s = 4, n = 128$.

Table I shows the DRs of all three methods under attacks of above. It can be seen that the proposed method has excellent robustness at the same SNR. And our method embeds complete information on each audio frame. In the face of cutting off attacks, the watermarked signal can save more effective watermark information for extraction.

V. CONCLUSION

In this paper, we present a new watermarking method STAW for audio signals, and the proposed method is spread spectrum based and embedding watermark bits in the short time DCT domain. By introducing an embedding strength parameter that changes over time, the strength of embedded watermark is related to the audio amplitude, which improves the imperceptibility of the watermarked signals. The sequence we embed is float, not binary and orthogonal, which reduces host signal interference at extraction stages. Overall, our scheme achieve highly imperceptible and highly robust, and the experimental results also confirm it.

ACKNOWLEDGMENT

This work was supported by the National Key R&D Program of China (2017YFB1401000) and the Key R&D Program of Shanxi (201903D421007). It was also the research achievement of the Key Laboratory of Digital Rights Services, which is one of the National Science and Standardization Key Labs for Press and Publication Industry.

REFERENCES

[1] Boney, L. , A. H. Tewfik , and K. N. Hamdy . "Digital watermarks for audio signals." *Proceedings of the Third IEEE International Conference on Multimedia Computing and Systems* IEEE, 2002.

[2] Hua G , Huang J , Shi Y Q , et al. Twenty Years of Digital Audio Watermarking – A Comprehensive Review[J]. Signal Processing, 2016, 128(Nov.):222-242.

[3] Valizadeh, A. , and Z. J. Wang . "An Improved Multiplicative Spread Spectrum Embedding Scheme for Data Hiding." *IEEE Transactions on Information Forensics & Security* 7.4(2012):1127-1143.

[4] Liu Z , Huang Y , Huang J . Patchwork-based Audio Watermarking Robust Against De-synchronization and Recapturing Attacks[J]. IEEE Transactions on Information Forensics and Security, 2018, PP(5):1-1.

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on April 08,2022 at 03:49:07 UTC from IEEE Xplore. Restrictions apply.

[5] Kanhe A , Gnanasekaran A . A QIM-Based Energy Modulation Scheme for AudioWatermarking Robust to Synchronization Attack[J]. Arabian Journal for Science & Engineering, 2019, 44(4):3415-3423.

[6] D. Megas, J. Serra-Ruiz, M. Fallahpour, Efficient self-synchronised blind audio watermarking system based on time domain and fft amplitude modification, Signal Processing 90 (12) (2010) 3078 – 3092.

[7] Janusz Cichowski, Andrzej Czyżewski, and Bożena Kostek. "Analysis of impact of audio modifications on the robustness of watermark for non-blind architecture." *Multimedia Tools&Applications* (2015).

[8] P. Hu, D. Peng, Z. Yi, Y. Xiang, Robust time-spread echo watermarking using characteristics of host signals, Electronics Letters 52 (1) (2016) 5–6.

[9] Roy, Sumon, et al. "An efficient and blind audio watermarking technique in DCT domain." *computer and information technology* (2015): 362-367.

[10] Attari, Ali Akbar , and A. A. B. Shirazi . "Robust and Transparent Audio Watermarking based on Spread Spectrum in Wavelet Domain." *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)* IEEE, 2019.

[11] Cox I J , Kilian J . Secure spread spectrum watermarking for multimedia[J]. IEEE Transactions on Image Processing, 1997, 6(12):P.1673-1687.

[12] Xiang, Yong , et al. "Spread Spectrum Audio Watermarking Using Multiple Orthogonal PN Sequences and Variable Embedding Strengths and Polarities." *IEEE/ACM Transactions on Audio Speech & Language Processing* (2017):1-1.

[13] Hervé Jégou, M. Douze , and C. Schmid . "Product Quantization for Nearest Neighbor Search." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 33.1(2010):117-128.

[14] Axler, Sheldon Jay. Linear Algebra Done Right[J]. Undergraduate Texts in Mathematics, 2015.

[15] Xiang, Yong , et al. "Spread Spectrum-Based High Embedding Capacity Watermarking Method for Audio Signals." *IEEE/ACM Transactions on Audio Speech & Language Processing* 23.12(2015):1-1.

Authorized licensed use limited to: INSTITUTE OF AUTOMATION CAS. Downloaded on April 08,2022 at 03:49:07 UTC from IEEE Xplore. Restrictions apply.