

## Research Article

# Robust and Reversible Audio Watermarking by Modifying Statistical Features in Time Domain

Shijun Xiang,<sup>1,2</sup> Le Yang,<sup>1</sup> and Yi Wang<sup>1</sup>

<sup>1</sup>The School of Information Science and Technology, Jinan University, Guangzhou 510632, China

<sup>2</sup>State Key Laboratory of Information Security, Institute of Information Engineering, The Chinese Academy of Sciences, Beijing 100093, China

Correspondence should be addressed to Shijun Xiang; shijun\_xiang@qq.com

Received 10 January 2017; Revised 31 March 2017; Accepted 4 April 2017; Published 27 April 2017

Academic Editor: Akram M. Z. M. Khedher

Copyright © 2017 Shijun Xiang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Robust and reversible watermarking is a potential technique in many sensitive applications, such as lossless audio or medical image systems. This paper presents a novel robust reversible audio watermarking method by modifying the statistic features in time domain in the way that the histogram of these statistical values is shifted for data hiding. Firstly, the original audio is divided into nonoverlapped equal-sized frames. In each frame, the use of three samples as a group generates a prediction error and a statistical feature value is calculated as the sum of all the prediction errors in the frame. The watermark bits are embedded into the frames by shifting the histogram of the statistical features. The watermark is reversible and robust to common signal processing operations. Experimental results have shown that the proposed method not only is reversible but also achieves satisfactory robustness to MP3 compression of 64 kbps and additive Gaussian noise of 35 dB.

## 1. Introduction

With the rapid development of the Internet technology, publication and dissemination of digital multimedia become more and more convenient. However, the authenticity and security of the digital multimedia are a challenge for the media owner [1]. Digital watermarking technology is an efficient approach to protect the copyright of the digital media. Reversible watermarking is one of the watermarking technologies used for data hiding. Reversible watermarking enables embedding secret data into host media and allow extraction of the original media and the secret data [2–4]. It is very useful in sensitive applications such as medical image system, military image, and lossless audio [5]. Although there are so many reversible watermarking methods, most of them are designed in a lossless environment and cannot resist any type of attacks. As a result, the original media or the secret data cannot be recovered after the watermarked media go through some changes [6].

In some cases, such as the copyright protection of the digital media, the embedded data is expected to be robust to some attacks such as lossy compression or additive noise. To this end, researchers pay more attention to robust reversible

watermarking. Robust reversible watermarking is that the original media and the embedded data can be both recovered correctly when the watermarked media remain intact, and the embedded data can still be extracted without error even when the watermarked media go through some attacks [7]. Until now, a few robust reversible image watermarking methods have been proposed, which can be classified into two groups:

- (i) Blind watermarking scheme: in [7, 8], Vleeschouwer et al. proposed a blind extraction scheme based on the patchwork theory and modulo-256 by using the gray-scale histogram rotation. This work is robust against JPEG compression, but the watermarked image has lower visible quality due to the reason that the watermark embedding procedure will cause salt-and-pepper noise in the watermarked image. Besides, the payload is low. To handle the salt-and-pepper noise problem, Zou et al. proposed a scheme by shifting the absolute mean values of the integer wavelet transform (IWT) coefficients in a chosen subband [9], and Ni et al. proposed a scheme by modifying the histogram of a robust statistical quantity in the spatial domain [10]. Since the embedding process may introduce the

error bits, the error correction coding (ECC) has been used. Besides, these two methods suffered from the unstable robustness and incomplete reversibility according to [11]. In [12], Zeng et al. enhanced the scheme of Ni et al. by introducing two thresholds and a new embedding mechanism. This method is blind and reversible. For a satisfactory performance, the two threshold values have to be carefully searched for different cover images.

- (ii) Nonblind watermarking scheme: in [13], a nonblind scheme based on wavelet-domain statistical quantity histogram shifting and clustering (WSQH-SC) is proposed. A pixel adjustment is presented at first to avoid the overflow and underflow, and a location map is used to record the changed pixels. This method achieved good robustness against JPEG, JPEG2000, and additive Gaussian noise, but it is not blind since the locations of the changed pixels need to be saved as a part of side information and transmitted to the receiver side in order to recover the original image. In [14], the Slantlet transform (SLT) was applied to image blocks, modifying the mean values of the HL and LH subband coefficients to embed the watermark bits, and a second stage of SLT transform is applied to the LL1 subband, embedding another watermark bit into the HL2 and LH2 subband. Because the coefficients and the mean values are fractional with more decimal places, the mean information was taken as side information to be sent to the receiver side for the recovery of the original cover image. In order to solve the nonblind extraction question in [14], the authors in [15] used IWT on images and randomly selected 10 coefficients from all the 16 coefficients in a block to compute the amplitude mean of the block, so that the mean information can be embedded into the image itself for blind extraction.

In [16], Coltuc and Chassery proposed a general framework for robust reversible watermarking by multiple watermarking. First the watermark is embedded into the cover image with a robust watermarking method and then a reversible watermarking method is adopted to embed the information (which is used to restore the original cover image) into the robust watermarked image. Suppose  $I$  and  $I_1$  are the original image and the robust watermarked image after embedding a watermark  $w$ , respectively. The embedding distortion,  $d = I - I_1$ , is compressed and embedded into the robust watermarked image with the reversible watermarking method. At the receiver side, if there are no attacks, the robust watermarked image  $I_1$  and the difference  $d$  can be extracted since the embedding process is reversible. Then the original image  $I$  can be recovered by  $I = d + I_1$ . Furthermore, the watermark can be extracted. If the watermarked image goes through a JPEG compression operation, the robust watermark can still be extracted. This framework is very instructive and achieves higher payload and good robustness against JPEG compression.

In [17], a robust reversible audio method based on spread spectrum and amplitude expansion is proposed. A robust

payload is embedded at first using the direct-sequence spread-spectrum modulation, with the sequence determined from the amplitude expansion in time and frequency of integer modified discrete cosine transform (MDCT) coefficients. Then a reversible payload is embedded into the apertures in the amplitude histogram that result from amplitude expansion of the integer MDCT coefficients to recover the host audio. This method achieves robustness against some signal processing like MP3 compression and additive noise, and if the watermarked audio remains intact, the host audio can be recovered perfectly.

In this paper, we propose a novel robust and reversible audio watermarking scheme based on statistic feature and histogram shifting in time domain. By shifting the histogram of the statistic features in time domain, the proposed algorithm achieves good robustness and reversibility at the same time.

The rest of the paper is organized as follows. The foundation work is introduced in Section 2. The proposed watermarking algorithm is described in Section 3. Experimental results are presented in Section 4. Section 5 concludes this paper.

## 2. Algorithm's Principle

This section will introduce the foundation works of the proposed robust reversible digital audio watermarking scheme. Firstly, a robust statistic feature of time domain is introduced; then how to modify the statistic feature to embed the watermark bit is briefly described.

**2.1. Robust Statistic Feature.** Consider a time-discrete digital audio signal  $X$ ; the host signal is first divided into nonoverlapped equal-sized frames. We take  $N$  samples per frame; for example,  $N$  samples as a frame and three samples as a group are shown in Figure 1. For a sample group  $(x_l, x_m$  and  $x_r)$ , the prediction value of the middle sample  $\hat{x}_m$  is calculated by using two immediate samples as

$$\hat{x}_m = \left\lceil \frac{x_l + x_r}{2} \right\rceil, \quad (1)$$

where  $\lceil x \rceil$  means rounding the elements of  $x$  to the nearest integer towards infinity. The prediction error of  $\hat{x}_m$  is

$$e_m = x_m - \hat{x}_m. \quad (2)$$

Since the samples in a group are often highly correlated, the prediction error  $e_m$  is expected to be very close to zero. For a frame with  $N$  samples,  $N/3$  prediction errors can be computed. The sum of all the prediction errors in a frame, denoted by  $E$ , is called the statistic feature in this paper. The statistical feature of a frame is calculated as

$$E = \sum_{i=1}^{N/3} e_i, \quad (3)$$

where  $e_i$  is the prediction error of the  $i$ th group in the frame. The basic idea of the proposed algorithm is based on this statistic property.

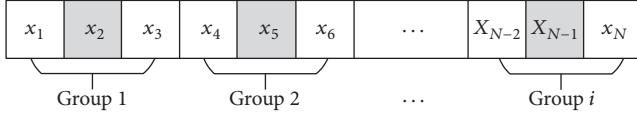


FIGURE 1: The use of  $N$  samples as a frame and three samples as a group.

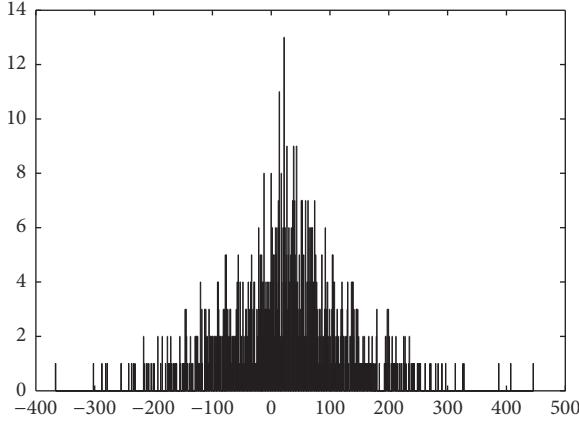


FIGURE 2: The distribution of the  $E$  values for the clip *track 1*.

**2.2. Watermark Statistic Feature.** For each frame, one watermark bit is embedded by shifting the value of the statistic feature. The shifting operation is done by modifying the samples in a frame. Taking *track 1* (which is downloaded from the website [18]) as example clip, Figure 2 shows the distribution of  $E$  values by using 300 samples as a frame and three samples as a group. The rule to modify the statistic value is referred to histogram shifting method. At first, we scan all frames and find out the maximum of the absolute  $E$  values, denoted by  $E_{\max}$ . Then, a threshold  $T$  is set to a positive integer bigger than  $E_{\max}$ . As a result, all  $E$  values are within the range  $[-T, T]$ . For example, from Figure 2 we can get that  $E_{\max}$  is 446, so threshold  $T$  can be an integer such as 500. The watermarking rule is to keep the statistic feature within  $[-T, T]$  if the watermark bit is “0” while the statistic feature is shifted away from zero by a shifting quantity  $T + G$  if the watermark bit is “1.” To achieve stronger robustness, parameter  $G$  is a threshold which is usually set bigger than  $T$ . To reduce the embedding distortion, if the embedded watermark bit is “1” and the original statistic feature belongs to  $[0, T)$ , the statistic feature is shifted to the region  $[T + G, 2T + G]$ ; if the embedded watermark bit is “1” and the original statistic feature belongs to  $(-T, 0)$ , the statistic feature is shifted to the region  $[-2T - G, -T - G]$ . In such a way, the bit-0 region and the bit-1 region are separated by the robust regions  $(T, T + G)$  and  $(-T - G, -T)$ . For example, Figure 3 shows the distribution of  $E$  values after embedding watermark by using clip *track 1*.

The modifying rules are as follows.

If the embedded bit is “0,” keep the frame unchanged. If the embedded bit is “1,” the samples in the frame are modified by

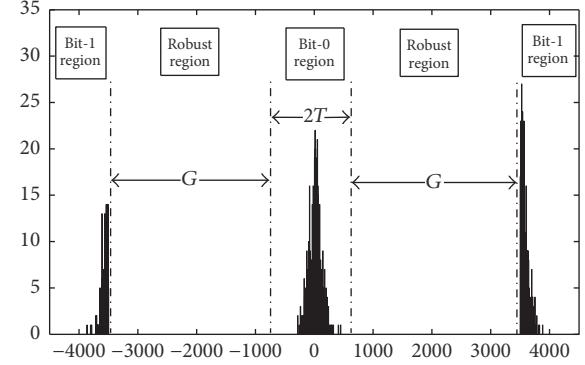


FIGURE 3: The distribution of the  $E$  values of *track 1* after embedding watermark.

$$y_i^k = \begin{cases} x_i^k + \beta, & \text{if } 0 \leq E_k < T, \bmod(i, 3) = 2 \\ x_i^k - \beta, & \text{if } -T < E_k < 0, \bmod(i, 3) = 2 \\ x_i^k, & \text{otherwise,} \end{cases} \quad (4)$$

where  $x_i^k$  is the  $i$ th sample in the  $k$ th frame. The index  $i$  is in  $[1, S]$  and  $S$  is the number of the samples in a frame. The integer value  $\beta$  is the shifted quantity of a sample,

$$\beta = \left\lceil \frac{T + G}{S/3} \right\rceil. \quad (5)$$

At the receiver side, if the watermarked audio remains intact, the watermark bits can be extracted by

$$w_k = \begin{cases} 0, & \text{if } E_k \in [-T, T] \\ 1, & \text{otherwise,} \end{cases} \quad (6)$$

where  $w_k$  is the to be hidden  $k$ th bit. The original audio can be recovered as

$$x_i^k = \begin{cases} y_i^k - \beta, & \text{if } E_k \in [T + G, 2T + G], \bmod(i, 3) = 2 \\ y_i^k + \beta, & \text{if } E_k \in [-2T - G, -T - G], \bmod(i, 3) = 2 \\ y_i^k, & \text{otherwise.} \end{cases} \quad (7)$$

**2.3. Prevention of Overflow/Underflow.** For a 16-bit digital audio, the permission range of the sample value is  $[-2^{15}, 2^{15}]$ . Watermark embedding will modify the sample values with the value  $\beta$ , so the overflow or underflow does not occur if the original sample values belong to  $[-2^{15} + \beta, 2^{15} - \beta]$ . In fact, as the value  $\beta$  is very small, the original sample values of most normal audio belong to  $[-2^{15} + \beta, 2^{15} - \beta]$ . Therefore, in the proposed method, there is no overflow or underflow in most cases. If the audio cannot meet this condition, we can record the location and modify the sample value to the range  $[-2^{15} + \beta, 2^{15} - \beta]$ ; then the location can be saved as side information and embedded into the audio.

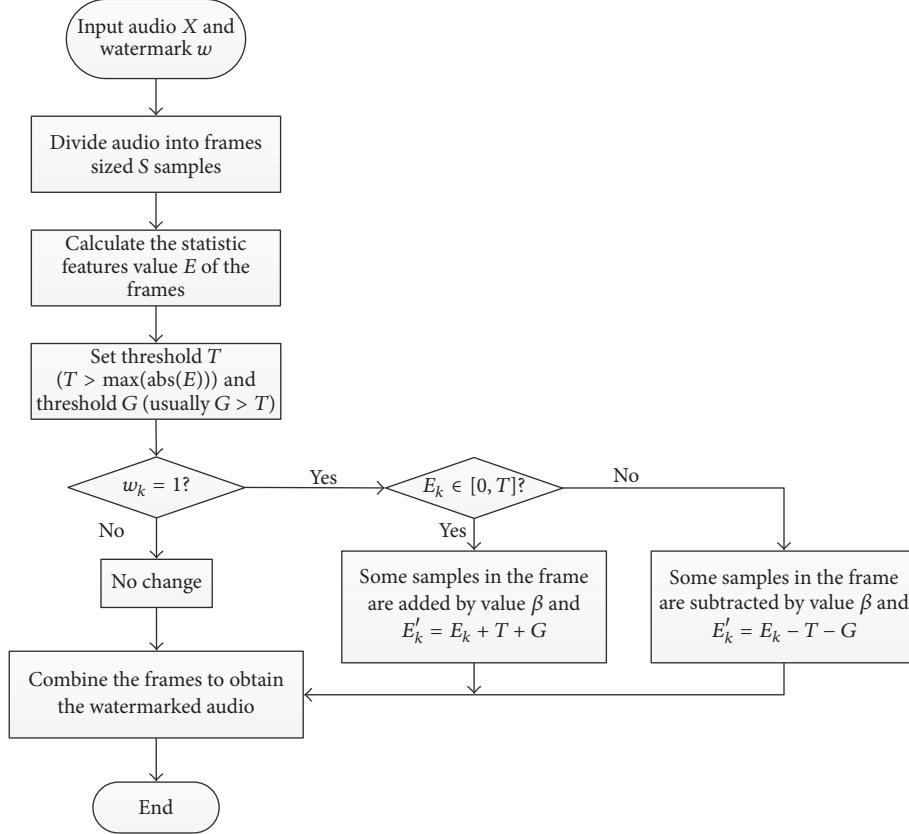


FIGURE 4: Watermark embedding process.

### 3. Proposed Algorithm

The embedding and extraction processes are presented in detail as follows.

*3.1. Watermark Embedding.* Figure 4 shows the proposed watermark embedding process. The watermark is embedded with the following five steps.

*Step 1.* Divide the original audio  $X$  into nonoverlapping frames sized  $S$  samples.

*Step 2.* Calculate the statistic features of the frames ( $E$  values) by referring to (1)–(3).

*Step 3.* Set the threshold values  $T$  and  $G$  ( $T > E_{\max}$  and usually  $G > T$ ).

*Step 4.* If the watermark bit is “0,” nothing is changed. If the bit is “1,” shift the statistic feature value with a shifting quantity  $T + G$  to embed the watermark bit by modifying the samples in the frame with value  $\beta$  referring to (4).

*Step 5.* Combine the frames to get the watermarked audio.

*3.2. Watermark Extraction.* If the watermarked audio goes through some attacks (such as MP3 compression, additive noise, resampling, or requantization), the watermark can

still be detected. To improve the accuracy of the watermark extraction, three extraction methods and a majority voting system are adopted to identify the extracted watermark by computing the distorted statistical feature values  $E'$ .

(i) *Extraction 1.* Redefine the bit-0 region as  $[-T - G/2, T + G/2]$  and the watermark extraction as

$$w'_1(i) = \begin{cases} 0, & \text{if } E' \in [-T - G/2, T + G/2] \\ 1, & \text{otherwise.} \end{cases} \quad (8)$$

(ii) *Extraction 2.* Redefine the bit-0 region as  $[-T - G/3, T + G/3]$  and the watermark extraction as

$$w'_2(i) = \begin{cases} 0, & \text{if } E' \in [-T - G/3, T + G/3] \\ 1, & \text{otherwise.} \end{cases} \quad (9)$$

(iii) *Extraction 3.*  $k$ -means clustering algorithm is introduced to extract bits. Figure 5 shows the distribution of the  $E$  values after MP3 compression, and the watermark can be extracted by

$$w'_3(i) = \begin{cases} 0, & \text{if } E' \in \text{class 2} \\ 1, & \text{if } E' \in \text{class 1 and class 3.} \end{cases} \quad (10)$$

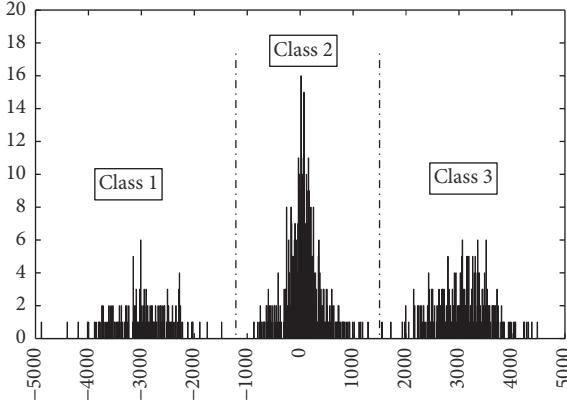


FIGURE 5: The distribution of the statistic feature values of track 1 after MP3 compression at 64 kbps.

The majority voting system works as

$$w'(i) = \begin{cases} w'_1(i), & \text{if } w'_1(i) = w'_2(i), w'_1(i) = w'_3(i) \\ w'_1(i), & \text{if } w'_1(i) = w'_2(i), w'_1(i) \neq w'_3(i) \\ w'_2(i), & \text{if } w'_1(i) \neq w'_2(i), w'_2(i) = w'_3(i) \\ w'_3(i), & \text{if } w'_1(i) \neq w'_2(i), w'_1(i) = w'_3(i). \end{cases} \quad (11)$$

Eventually, three extraction methods and a majority voting system are adopted to extract watermark. Figure 6 shows the proposed watermark extracting process. If the watermarked audio remains intact, the watermark can be extracted correctly and the original audio can be recovered as the following steps.

*Step 1.* Divide the watermarked audio  $X_w$  into nonoverlapping frames sized  $S$  samples.

*Step 2.* Calculate the statistic feature of the frames ( $E'$  values) by referring to (1)–(3).

*Step 3.* Extract the watermark with three extraction methods and identify the watermark with the majority voting system by referring to (8)–(11).

*Step 4.* The original audio can be recovered by modifying the samples in the frame with value  $\beta$  referring to (7).

*Step 5.* Combine the frames to get the original audio.

If the watermarked audio goes through some attacks, the original audio cannot be recovered exactly, so we focus on the watermark extraction, and the watermark is extracted as follows.

*Step 1.* Divide the watermarked audio  $X_w$  into nonoverlapping frames sized  $S$  samples.

*Step 2.* Calculate the statistic feature values of the frames  $E'$  by referring to (1)–(3).

*Step 3.* Extract the watermark with three extraction methods and identify the watermark with the majority voting system by referring to (8)–(11).

## 4. Experimental Results

In this section, 7 WAV audio file of the sample rate of 44.1 KHz and 16 bits per sample (tracks 1, 2, 3, 4, 5, 6, and 7 [18]) are used as example clips to evaluate the performance of the proposed algorithm. The payload of our method only depends on the length of a frame  $S$ ; for a time-discrete digital audio signal  $X$  in length  $M$ , the pure payload can be calculated by

$$C = \left\lfloor \frac{M}{S} \right\rfloor. \quad (12)$$

In the experiment, the watermark is a pseudo-random sequence in length of 1000 bits. The imperceptibility is first analyzed by the SNR standard at different threshold values and different sample numbers per frame. Then, robustness testing against MP3 compression, additive noise (AWGN), resampling (44.1-16-44.1 kHz), and requantization (16-8-16 bits) are reported by using the software CoolEditPro v2.1.

*4.1. Imperceptibility Test.* The imperceptibility is measured by the embedding distortion. In the proposed scheme, the distortion is caused by the shifting quantity on the samples depending on thresholds  $T$ ,  $G$  and the length of a frame  $S$ . Since  $T$  is set at first, we only investigate the influence of  $G$  and  $S$  on SNR.

Figure 7 plots the relationship between SNR and the threshold  $G$  for different clips at the same threshold  $T$  and  $S$ . From this figure we can conclude that with the increase of  $G$ , the SNR value drops. The reason is that the larger  $G$  is, the larger shifting quantity is used, so the larger embedding distortion is caused. As a result, SNR value drops.

Figure 8 plots the relationship between SNR and the length of a frame  $S$  for different clips at the same thresholds  $T$  and  $G$ . We can see from this figure that the larger  $S$  is, the higher SNR value is achieved. The reason is that with the increase of  $S$ , the shifting quantity for every single sample drops, so the SNR values rise due to the fact that the embedding distortion is reduced. Consequently, the frame length  $S$  will influence Maximum Embedding Capacity and SNR value directly, the Maximum Embedding Capacity is higher when  $S$  is smaller according to (12), and the SNR value is higher when  $S$  is larger according to Figure 8. To consider the balance between Maximum Embedding Capacity and SNR value, we have found that the value of  $S$  within the range of 300 to 600 can usually achieve a satisfactory result after a set of experiments.

*4.2. Robustness Testing.* To test the robustness of the proposed scheme, a set of experiments has been taken on tracks 1–7. Table 1 shows the results, in which RP means resampling (44.1-16-44.1 kHz) operations while RQ means requantization (16-8-16 bits) operations. We can observe from this table that all the example clips can achieve the robustness against MP3 compression at 64 Kbps. For track 1, the watermark bits can be correctly extracted under the

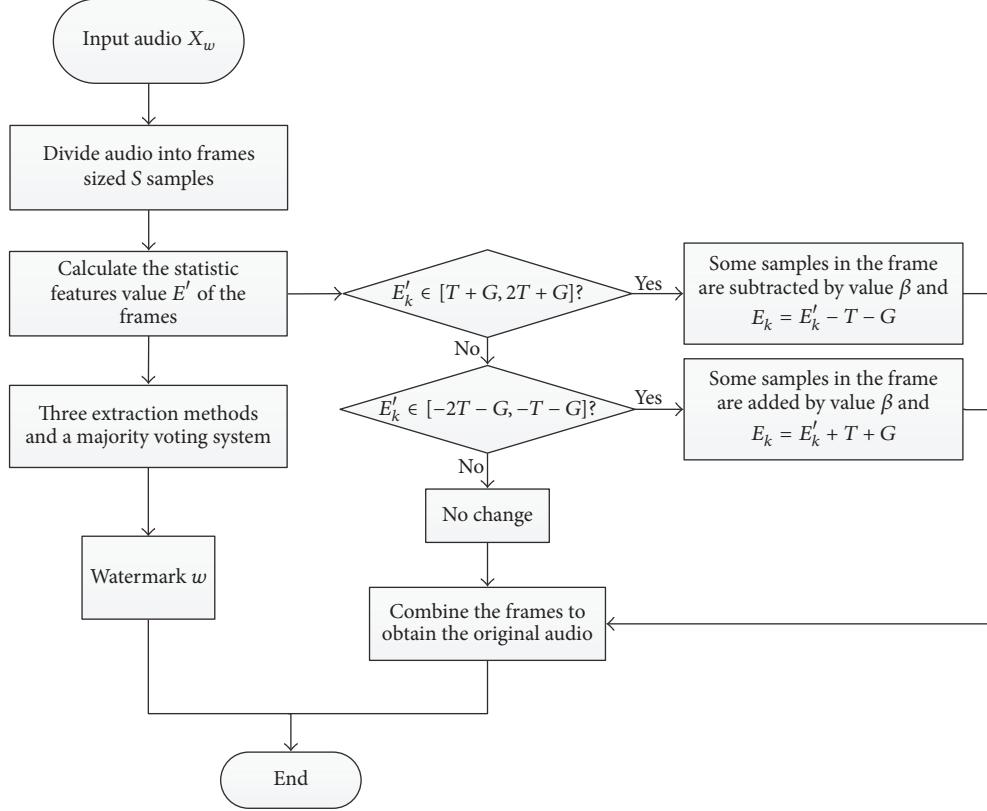


FIGURE 6: Watermark extracting process.

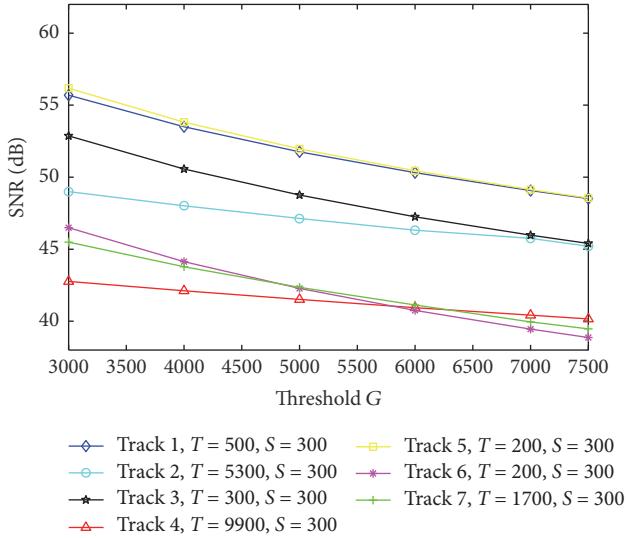


FIGURE 7: Relationship between SNR and threshold G.

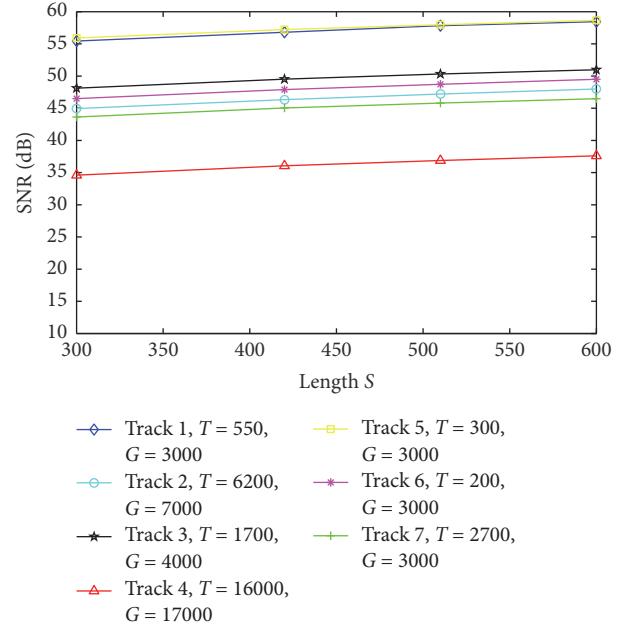


FIGURE 8: Relationship between SNR and the frame length S.

MP3 compression of 48 Kbps. The robustness against additive noise is also satisfactory. Even with the noise intensity at 25 dB, the BER (bit error rate) values are less than 10% except for track 1. Besides, the watermark robustness against resampling and requantization operations is perfect, and the hidden bits can be recovered without errors.

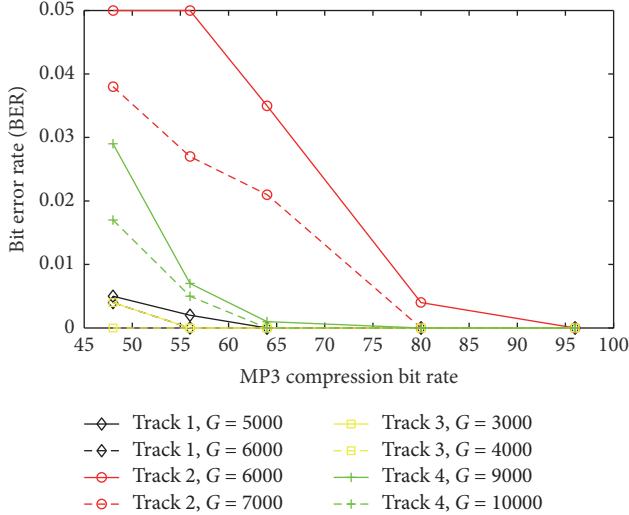
As shown in Figure 3, the robustness of the proposed method is originated from the robust region. The robust region depends on threshold G. The larger G is, the larger the robust region is, and the stronger robustness is. Figure 9

TABLE 1: Performance of the proposed method.

Audio	$S$	$T$	$G$	SNR	MP3 (Kbps)	AWGN (35 dB)	AWGN (30 dB)	AWGN (25 dB)	RP	RQ
Track 1	420	500	3000	54.74	48	2/1000	77/1000	251/1000	0/1000	0/1000
Track 2	600	6200	7000	47.98	64	0/1000	0/1000	0/1000	0/1000	0/1000
Track 3	510	600	3000	54.09	48	0/1000	0/1000	44/1000	0/1000	0/1000
Track 4	300	9900	10000	38.99	64	0/1000	0/1000	0/1000	0/1000	0/1000
Track 5	300	200	3000	56.18	48	0/1000	25/1000	77/1000	0/1000	0/1000
Track 6	600	200	3000	49.5	48	0/1000	0/1000	46/1000	0/1000	0/1000
Track 7	300	1600	4000	43.78	48	0/1000	0/1000	0/1000	0/1000	0/1000

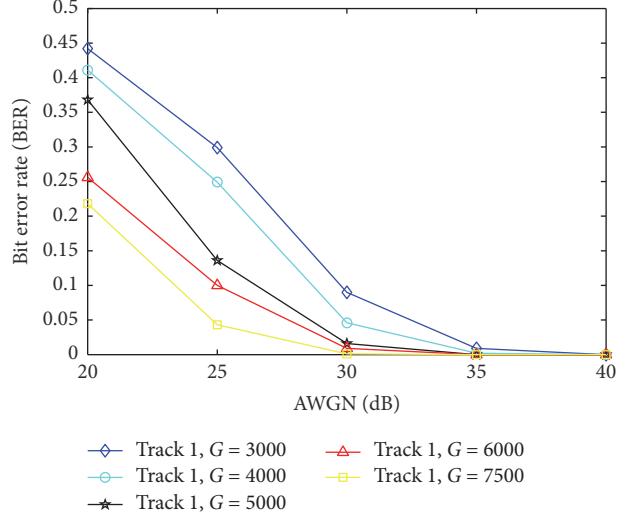
TABLE 2: Performance of the proposed method with different length of a frame  $S$  on track 1.

Track	$S$	$T$	$G$	MP3 (Kbps)	AWGN (35 dB)	AWGN (30 dB)	AWGN (25 dB)
Track 1	300	500	3000	80	9/1000	90/1000	299/1000
Track 1	420	500	3000	56	12/1000	171/1000	320/1000
Track 1	510	550	3000	80	28/1000	171/1000	158/1000
Track 1	600	550	3000	48	28/1000	221/1000	390/1000
Track 6	300	200	3000	48	0/1000	0/1000	3/1000
Track 6	420	200	3000	56	28/1000	0/1000	23/1000
Track 7	300	200	4000	48	0/1000	0/1000	0/1000
Track 7	420	200	4000	56	0/1000	0/1000	0/1000

FIGURE 9: Relationship between MP3 bit rate and threshold  $G$ .

supports the conclusion. Figure 9 shows the bit error rate (BER) at different threshold  $G$  for the same audio with same threshold  $T$ . The lower BER means that the stronger robustness is achieved. We can find that as threshold  $G$  increases, the BER drops, and the robustness rises.

Take *track 1* as example clip; Figure 10 shows the bit error rate of the extracted watermark with different thresholds  $G$  against additive noises with the same  $T$  and  $S$  ( $T = 500$ ,  $S = 300$ ). We can see that the larger  $G$  is, the smaller bit error rate is, and the better robustness is. As threshold  $G$  increases, the robustness becomes stronger. In the application, we can adjust the parameter  $G$  to achieve ideal robustness. On the

FIGURE 10: The BER values at different AWGN with different threshold  $G$ .

other hand, with the increase of  $G$ , the SNR value drops. To consider the balance between SNR value and robustness, we have found that the value of  $G$  within 3000 to 5000 can usually achieve a satisfactory result after a set of experiments.

To evaluate the effect of the frame length  $S$  on the robustness performance, a set of experiments has been taken on *track 1*, *track 6*, and *track 7*. Table 2 lists the results. We can observe that, for the same audio with the same  $T$  and  $G$ , such as  $T = 500$  and  $G = 3000$  for *track 1*, as  $S$  increases, the robustness against MP3 compression will strengthen, but for *track 6* and *track 7*, as  $S$  increases, the robustness against

TABLE 3: Comparison with method in [17].

Track	Payload (bits)	Method in [17]			Proposed method		
		ODG	MP3 (128 Kbps)	AWGN (36 dB)	Payload (bits)	ODG	MP3 (80 Kbps)
Track 32	216	-2.45	0	150	3000	1000	-1.37
Track 35	216	-1.31	0	12%	5000	1000	0
Track 65	216	-1.21	0	6%	30	2000	-1.58
Track 66	216	-0.24	0	96	6000	1000	0
Track 69	216	-0.31	1%	4%	400	2000	-0.23

MP3 compression drops, so the effect of the frame length  $S$  on the robustness against MP3 compression is unstable. The influence on AWGN is little.

For fair comparison with the method in [17], we use the same host signals (tracks 32, 35, 65, 66, and 69) downloaded from sound quality assessment material (SQAM) collection [19]. Table 3 shows the robustness testing results against MP3 compression and additive noise (AWGN) operations. We can observe that the method in [17] can carry 216 bits and resist the MP3 compression at 128 Kbps while the proposed method can resist the MP3 compression at 64 Kbps with 1000 bits embedded. In addition, in our method the BER under the AWGN of 35 dB is less than that of the method in [17]. In other words, the proposed method can provide larger embedding capacity and obtain stronger robustness against MP3 compression and AWGN attacks. The imperceptibility is evaluated by using the ODG standard. The closer the ODG value to 0, the better the imperceptibility. For the table it is noted that the imperceptibility of the proposed method is better except for the clips *track 35* and *track 66*. The reason is that  $E_{\max}$  values of *track 35* and *track 66* are bigger. As a result, thresholds  $T$  and  $G$  are also larger and more embedding distortion will be caused.

## 5. Conclusions

In this paper, we proposed a robust and reversible audio watermarking method by shifting the histogram of the statistical feature values in time domain. The statistical feature is calculated as the sum of the prediction errors in a frame. Since the audio clip has a larger number of samples and each frame can hold enough elements, the statistical feature is robust to those common signal processing operations. Considering that the distribution of the statistical feature values may be distorted to some extent, three extraction methods and the majority voting system are designed for the watermark detection. Experimental results have shown that thousands of bits can be reversibly embedded and the watermark bits can resist MP3 compression of 64 kbps and additive noise of 35 dB. Comparing with the existing excellent method, the proposed method can embed more watermark bits and achieve stronger robustness.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by NSFC (no. 61272414) and Open Research Fund from the State Key Laboratory of Information Security (no. 2016-MS-07).

## References

- [1] Y. Xiang, I. Natgunanathan, S. Guo, W. Zhou, and S. Nahavandi, "Patchwork-based audio watermarking method robust to desynchronization attacks," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 22, no. 9, pp. 1413–1423, 2014.
- [2] J. Fridrich, M. Goljan, and R. Du, "Lossless data embedding—new paradigm in digital watermarking," *EURASIP Journal on Advances in Signal Processing*, vol. 2002, no. 2, pp. 185–196, 2002.
- [3] Y. Q. Shi, Z. Ni, D. Zou, C. Liang, and G. Xuan, "Lossless data hiding: fundamentals, algorithms and applications," in *Proceeding of the IEEE International Symposium on Circuits and Systems*, vol. 2, pp. 313–336, 2004.
- [4] S. Lee, C. D. Yoo, and T. Kalker, "Reversible image watermarking based on integer-to-integer wavelet transform," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 321–330, 2010.
- [5] X. Li, B. Yang, and T. Zeng, "Efficient reversible watermarking based on adaptive prediction-error expansion and pixel selection," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3524–3533, 2000.
- [6] Z. Ni, Y. Q. Shi, N. Ansari, W. Su, Q. Sun, and X. Lin, "Robust lossless image data hiding," in *Proceeding of the IEEE International Conference on Multimedia and Expo (ICME '04)*, vol. 3, pp. 2199–2202, Taipei, Taiwan, June 2004.
- [7] C. De Vleeschouwer, J. Delaigle, and B. Macq, "Circular interpretation of histogram for reversible watermarking," in *Proceeding of the IEEE 4th Workshop on Multimedia Signal Processing*, pp. 345–350, Cannes, France, 2001.
- [8] C. De Vleeschouwer, J. F. Delaigle, and B. Macq, "Circular interpretation of bijective transformations in lossless watermarking for media asset management," *IEEE Transactions on Multimedia*, vol. 5, no. 1, pp. 97–105, 2003.
- [9] D. Zou, Y. Q. Shi, Z. Ni, and W. Su, "A semi-fragile lossless digital watermarking scheme based on integer wavelet transform," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 10, pp. 1294–1300, 2006.
- [10] Z. Ni, Y. Q. Shi, N. Ansari, W. Su, Q. Sun, and X. Lin, "Robust lossless image data hiding designed for semi-fragile image authentication," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 4, pp. 890–896, 2008.
- [11] L. An, X. Gao, C. Deng, and F. Ji, "Robust lossless data hiding: Analysis and evaluation," in *Proceeding of the International Conference on High Performance Computing and Simulation (HPCS '10)*, pp. 512–516, July 2010.
- [12] X.-T. Zeng, L.-D. Ping, and X.-Z. Pan, "A lossless robust data hiding scheme," *Pattern Recognition*, vol. 43, no. 4, pp. 1656–1667, 2010.
- [13] L. An, X. Gao, X. Li, D. Tao, C. Deng, and J. Li, "Robust reversible watermarking via clustering and enhanced pixel-wise masking," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3598–3611, 2012.
- [14] R. Thabit and B. E. Khoo, "Capacity improved robust lossless image watermarking," *IET Image Processing*, vol. 8, no. 11, pp. 662–670, 2014.
- [15] S. Xiang and Y. Wang, "Distortion-free robust reversible watermarking by modifying and recording iwt means of image blocks," in *Proceeding of the 14th International Workshop (IWDW '15)*, Tokyo, Japan, October , 2015.
- [16] D. Coltuc and J. Chassery, "Distortion-free robust watermarking: a case study," in *Security, Steganography, and Watermarking of Multimedia Contents*, vol. 6505 of *Proceedings of SPIE*, pp. 588–595, San Jose, Calif, USA, 2007.
- [17] A. Nishimura, "Reversible and robust audio watermarking based on spread spectrum and amplitude expansion," in *International Workshop on Digital Watermarking (IWDW '14)*, vol. 9023 of *Lecture Notes in Computer Science*, pp. 215–229.

- [18] Massachusetts Institute of Technology (MIT) Audio Database,  
<http://sound.media.mit.edu/media.php>.
- [19] EBU Committee, Sound quality assessment material recordings  
for subjective tests, <https://tech.ebu.ch/publications/sqamcd>.

