Audio Watermarking Algorithm Based on Centroid and Statistical Features*

Xiaoming Zhang[1] and Xiong Yin[1,2,**]

[1] College of Information Engineering, Beijing Institute of Petrochemical Technology,
Beijing 102617, China
[2] College of Information Science and Technology, Beijing University of Chemical Technology,
Beijing 100029, China
{zhangxiaoming,yinxiong}@bipt.edu.cn

Abstract. Experimental testing shows that the relative relation in the number of samples among the neighboring bins and the audio frequency centroid are two robust features to the Time Scale Modification (TSM) attacks. Accordingly, an audio watermark algorithm based on frequency centroid and histogram is proposed by modifying the frequency coefficients. The audio histogram with equal-sized bins is extracted from a selected frequency coefficient range referred to the audio centroid. The watermarked audio signal is perceptibly similar to the original one. The experimental results show that the algorithm is very robust to resample TSM and a variety of common attacks. Subjective quality evaluation of the algorithm shows that embedded watermark introduces low, inaudible distortion of host audio signal.

Keywords: Audio watermarking, FFT, Centroid, Histogram, TSM.

1 Introduction

Audio watermarking plays an important role in ownership protection. According to IFPI (International Federation of the Phonographic Industry), audio watermarking should be robust to temporal scaling of $\pm 10\%$ and be able to resist most of common signal processing manipulations and attacks, such as cropping, re-sampling and etc [1].

The algorithms for audio watermarking can be classified into two categories: algorithms in time domain and algorithms in transform domain, including those in compressed domain. Data hiding in the least significant bits of audio samples in the time domain is one of the simplest algorithms with very high data rate of additional information. In [2], the authors presented an audio watermarking algorithm in discrete wavelet transform domain. The watermark is embedded in the frequency point of discrete wavelet transform by replacing least significant bit. The capacity of algorithm is high and is robust to resample and cropping. In [3], a blind audio information bit hiding algorithm with effective synchronization is proposed. The algorithm embedded

[*] This work is supported by the Funding Project for Academic Human Resources Development in Institutions of Higher Learning Under the Jurisdiction of Beijing Municipality.

[**] Corresponding author.

S. Qing, H. Imai, and G. Wang (Eds.): ICICS 2007, LNCS 4861, pp. 153–163, 2007.

synchronization signals in the time domain to resist the attacks such as cropping while keeping the computation for resynchronization being lower. The watermark is placed in block DCT coefficients of the original audio exploiting the human auditory system (HAS) features.

The algorithm in [4] is very robust against de-synchronization attacks such as time scale modification (TSM), cropping. However, this watermarking algorithm is sensitive somewhat to additive noise attacks such as MP3 audio compression and low-pass filter. Of course, many audio watermarking algorithms (algorithm in literature [5]) are robust against additive noise attacks, but these algorithms cannot effectively resist TSM attacks.

In the existing literature, several algorithms have been proposed aiming at solving this problem by using exhaustive search, synchronization pattern, invariant watermark, and implicit synchronization. In [6], an audio watermarking method is presented by using music content analysis. The watermark is embedded into the edges of audio signal by viewing pitch-invariant TSM as a special form of random cropping, removing and adding some portions of audio signal while preserving the pitch. The watermark is robust to $\pm 9\%$ pitch-invariant TSM but vulnerable to other stretching modes such as solving playback speed modifications, which change the edges in the signal. In [7], side information is exploited to improve the searching of the watermark aiming at solving playback speed modifications. One weakness of this scheme is that the detection procedure is not blind. The histogram specification is first introduced for image watermarking in [8]. By using the robustness of image color histogram to rotations and geometric transformations, the authors in [9] proposed a general method is very robust to image geometric distortions. In [10], a watermarking algorithm to geometric distortion in DWT domain is proposed. In the algorithm, a watermark was embedded adaptively in low band of DWT domain, according to the conceal quality of Human Visual System; Especially, the geometric transformation could be corrected before the watermarking detection, owing to embedding a template in a circle of middle frequency in DFT and extracting a invariant centroid from a restricted area inside the image. Moreover, an improvement on centroid detection method was presented in [11]. The improved method constructs a centroid series which were convergent in probability to centroid of initial text line using both initial profile and reproduced profile of text line, and the watermark capacity is increased.

In this paper, the invariance of histogram and centroid in the frequency domain to TSM is presented. This is followed by a description of our proposed watermark method. Then, analyze the watermark performance and test the watermark robustness on resynchronization distortions, as well as some common signal processing and some attacks in Stirmark Benchmark for Audio. Finally, the conclusion is drawn.

2 Invariant Features in Frequency Domain

Since the bits embedded in the frequency domain can provide a stronger robustness against additional noises than in the time domain, in this section, we investigate the invariance of the histogram and centroid in the frequency domain by experimental testing as follows.

2.1 Centroid in Frequency Domain

For audio signal sequence (W_s bits/sample f_s/sample frequency), we first consider 20ms audio signal as a frame (80ms audio signal before compression). Then, a frame is divided into 32 sub-bands. Each sub-band contains K ($K = f_s * S_d * W_d * 20/(8000*32)$) samples, each frame contains K*32 samples, $s_i(j)(j \in [1...32])$ denotes the audio samples in j sub-band of i frame. $fft(s_i(j))(j \in [1...32])$ Denotes the audio samples in j sub-band of i frame is FFT transformed. The centroid in the frequency domain is calculated by formula (1) and formula (2):

$$M_j = \sqrt{\frac{\sum_{i=1}^{K} (20 * \log10(fft(s_i(j))))^2}{K}} \tag{1}$$

$$C = \frac{\sum_{j=1}^{32} j * M_j}{\sum_{j=1}^{32} M_j} \tag{2}$$

2.2 Histogram

A histogram is often used to describe the data distribution. The style of histogram may be described by:

$$H_M = \{h_M(i) \mid i = 1,...,L\}, \tag{3}$$

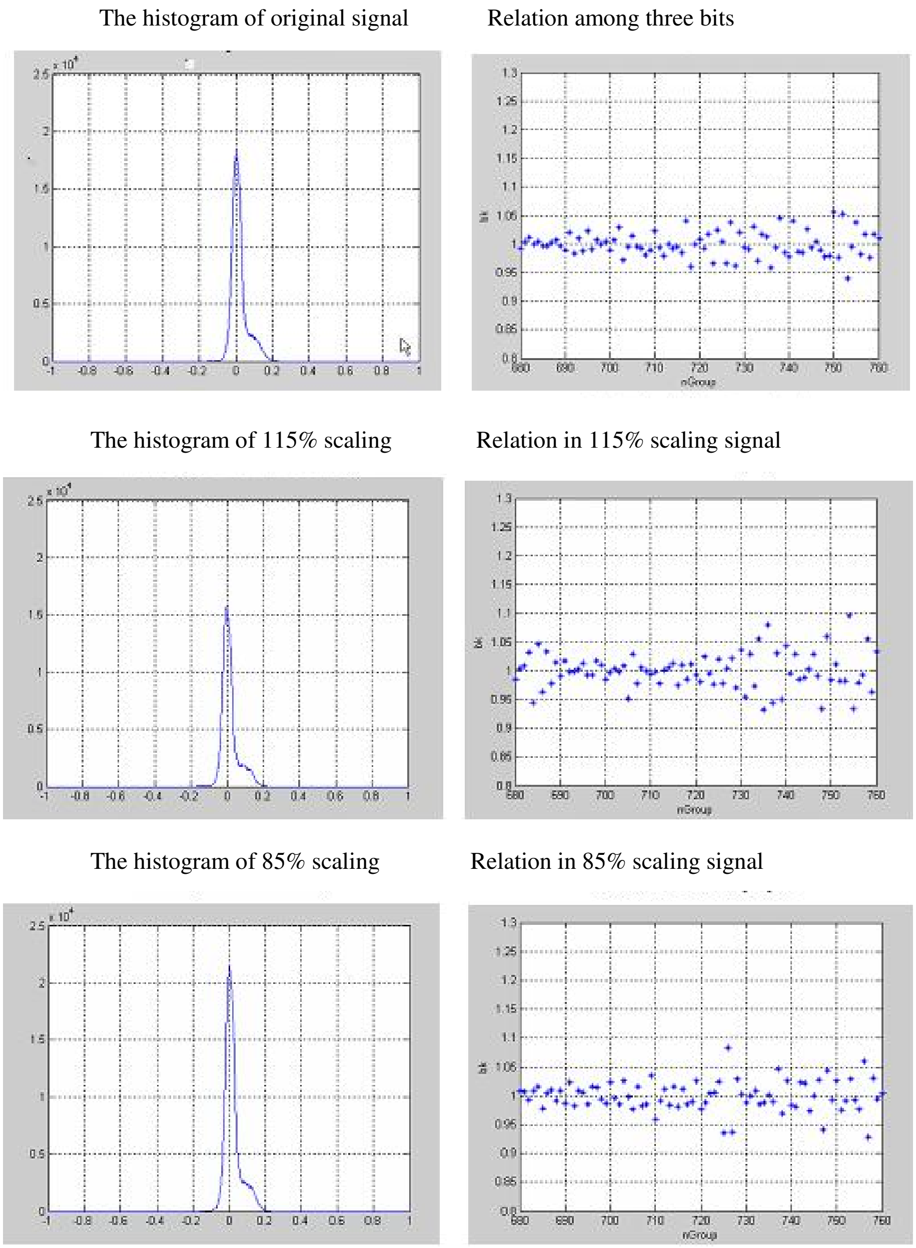
where H_M is vector, and denotes the volume-level a histogram of audio signal F, and $h_M(i)$ denotes the number of samples in the ^{i}th bin. Suppose that the resolution the audio signal is R bits, for a signed signal, the number of bins are calculated as:

$$L = \begin{cases} 2^R / M & \quad if \quad Mod(2^R / M) = 0 \\ \lfloor 2^R / M \rfloor + 1 & \quad\quad\quad other \end{cases} \tag{4}$$

where M is the size of bins, $h_M(i)$ includes all samples the range of sample value from $-2^{R-1} + (i-1) * M$ to $-2^{R-1} + i * M - 1$, and $\lfloor . \rfloor$ is the floor function.

2.3 Experimental Testing

We choose an audio signal (16-bit signed mono audio file sampled at 44.1 kHz with the length of 20s) to test the effects of the TSM on the histogram and centroid in the FFT domain. As to other kinds of audio signals, such as pop music, piano music and speech, etc, the simulation results are almost similar.



The histogram of original signal

Relation among three bits

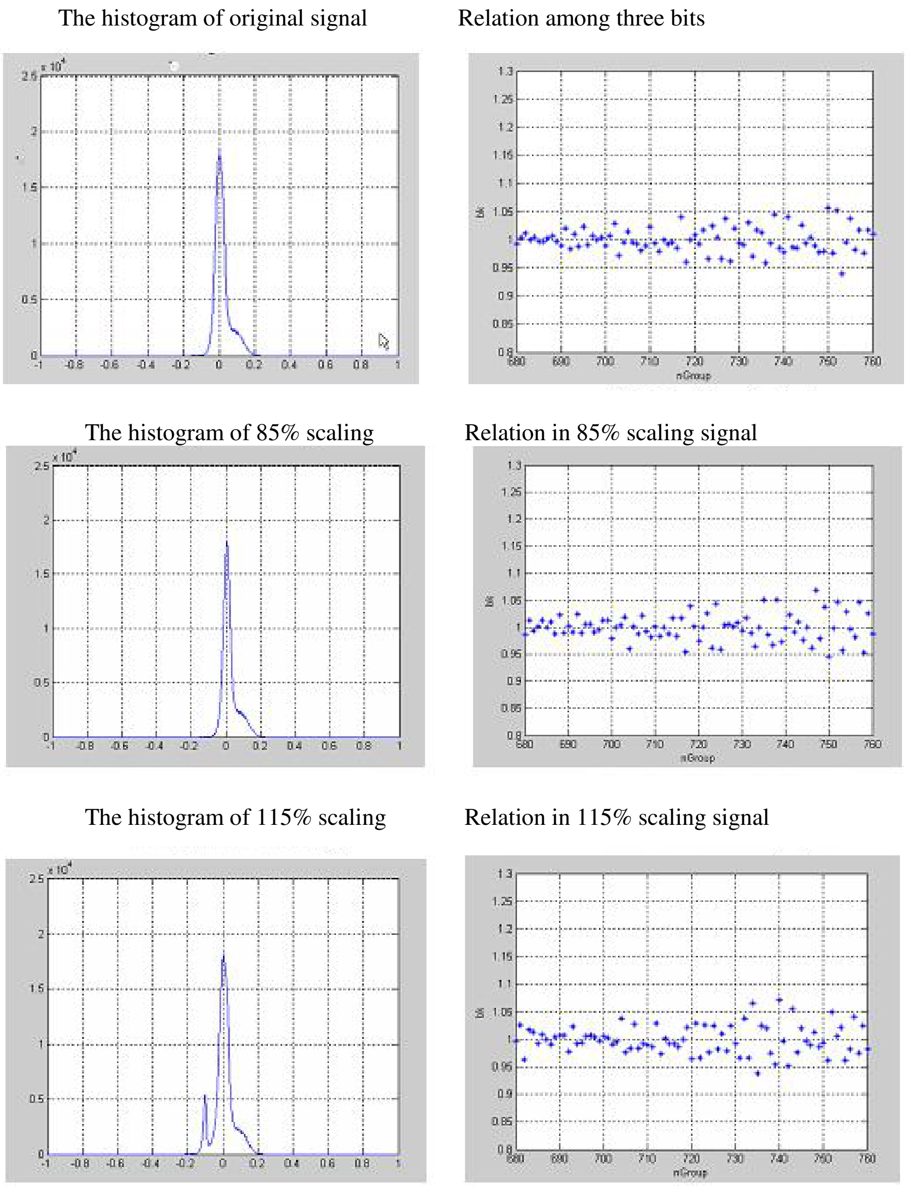
The histogram of 115% scaling

Relation in 115% scaling signal

The histogram of 85% scaling

Relation in 85% scaling signal

Fig. 1. The invariance of histogram to the pitch-invariant TSM: the sub-plots in left side is the histogram of original audio and scaled one with 85% and 115%, while the sub-plots in right side demonstrate the relative relation among three neighboring bins



The histogram of original signal

Relation among three bits

The histogram of 85% scaling

Relation in 85% scaling signal

The histogram of 115% scaling

Relation in 115% scaling signal

Fig. 2. The invariance of histogram to the resample TSM. The sub-plots in left side is the histogram of original audio and scaled one with 85% and 115%, while the sub-plots in right side demonstrate the relative relation among three neighboring bins.

The histograms are extracted from audio file after FFT transformed. The size of the bins M=0.5. Fig.1 and Fig.2 show the effects of the TSM attacks with the two different modes, respectively. Fig.3 plots the centroid values of the original and its scaled versions under 85%~115% TSM (pitch-invariant and resample) with the step size of 1%. Referenced to [4], the relative relations in the number of samples among three neighboring bins calculated and denoted by β_k :

$$\beta_k = \frac{2 * h_M(k)}{h_M(k-1) + h_M(k+1)} \qquad for \quad h_M(k) >> L \tag{5}$$

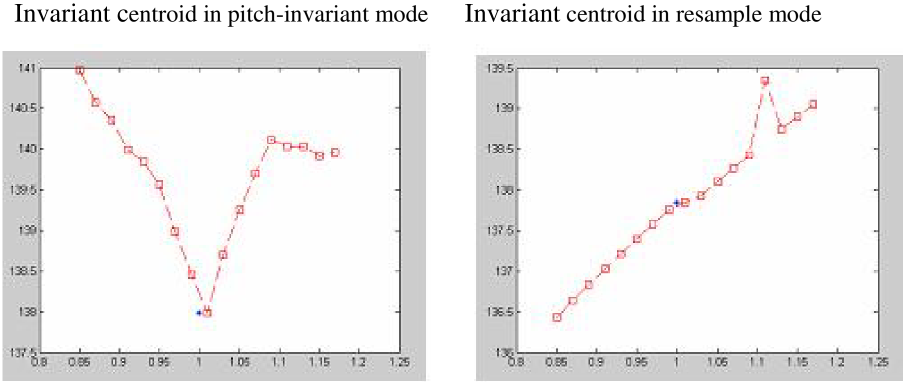
2.4 Comments

Based on the extensive testing, we have the following observations:

(1) In the FFT domain, the audio histogram is very robust to TSM. The relative relation among three neighboring bins is from 0.9 to 1.1. Refer to Figure.1 and Figure.2.

(2) The audio centroid in the FFT domain is robust enough to TSM. From 85% to 115% TSM, the error ratio of centroid is limited in $\pm 3\%$ (see Figure.3).

Overall, in the watermark design, if we incorporate the invariance of the histogram and centroid to TSM and the watermark in the FFT domain, the watermark will be robust.


Invariant centroid in pitch-invariant mode

Invariant centroid in resample mode

Fig. 3. The centroid of the example audio and scaled ones under the TSM of 85%~115% with resample (right) and pitch-invariant (left) stretching modes, respectively

3 Watermark Algorithm Design

The watermark embedding and extracting are described by the histogram specification. The robustness of the audio centroid and relative relation in the number of sample among different bins to the TSM attacks presented in the previous section are used in the design. First, the FFT transform is applied. And, the watermark is embedded into the coefficients of FFT instead of into the time domain signal itself.

3.1 Watermark Embedding Approach

The basic idea of the proposed embedding is to extract the histogram from a selected coefficient of FFT. Divide the bins into many groups, each group including three consecutive bins. For each group, one bit watermark is embedded by reassigning the number of samples in the three bins. The watermarked audio is obtained by modifying the coefficient of FFT according to the watermarking rule. The embedding approach is shown in Figure 4.

The detail embedding process is described as follows.

Suppose that there is a binary sequence $W = \{w_i \mid i = 1,\ldots,L_w\}$ to be hidden into a digital audio $F = \{f(i) \mid i = 1,\ldots,N\}$. The centroid of audio, denoted by A, is calculated as formula (1) and (2).

Then, select the amplitude range $B = [\lambda A, 1/\lambda A]$ from audio coefficient of FFT to extract the histogram $H = \{h(i) \mid i = 1,\ldots,L\}$, where $L = 3 * L_w$. $\lambda \in [0.6, 0.9]$, is a suggested range in which the bins extracted from B often hold enough samples.

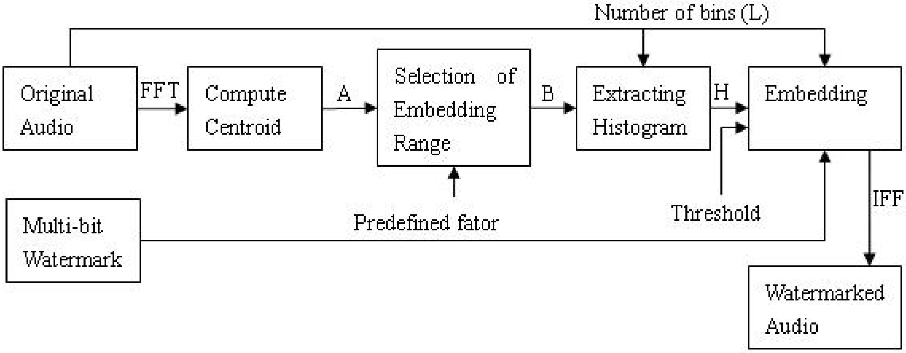


Fig. 4. Watermark embedding framework

After extracting the histogram, suppose that three consecutive bins, Bin_1, Bin_2 and Bin_3, their samples in the number are denoted as a, b and c. apply the follow equation to embed one bit of information, described as [4]:

$$\begin{array}{ll} 2b/(a+c) \geq T & if \quad w(i)=1 \\ (a+c)/2b \geq T & if \quad w(i)=0 \end{array}, \tag{6}$$

where T is a selected threshold used to control the watermark robustness performance and the embedding distortion. T should be not less than 1.1, in order to resist TSM.

If the embedded bit $w(i)$ is '1' and $2b/(a+c) \geq T$, no operation is needed. Otherwise, the number of samples in three different bins, a, b, c, will be adjusted until satisfying $2b/(a+c) \geq T$. Some selected samples from Bin_1 and Bin_3 in the number denoted by I1 and I3, will be modified to fall into Bin_2. The modification rule is described as Equation (7):

$$\begin{cases} ff_1^{'}(i) = ff_1(i) + M & 1 \le i \le I1 \\ ff_3^{'}(i) = ff_3(i) - M & 1 \le i \le I3 \end{cases}, \tag{7}$$

where $ff_1(i)$ and $ff_3(i)$ denote the ^{i}th modified sample in Bin_1 and Bin_3, $ff_1^{'}(i)$ and $ff_3(i)$ are the modified samples belong to Bin_2.

If the embedded bit $w(i)$ is '0' and $(a+c)/2b<T$, I1 and I3, some selected samples from Bin_2 will be modified to fall into Bin_1 and Bin_3, respectively. The rule is described as Equation (8):

$$\begin{cases} ff_2^{'}(i) = ff_2(i) - M & 1 \le i \le I1 \\ ff_2^{'}(i) = ff_2(i) + M & 1 \le i \le I3 \end{cases}, \tag{8}$$

where $ff_2(i)$ denotes the i^th^ modified sample in Bin_2, $ff_2^{'}(i)$ are the corresponding modified.

This process is repeated to embed all watermark bits. In our proposed embedding, the watermark is embedded by modifying the values of some selected coefficients of FFT samples from the audio. Hence, the re-construction of the watermarked audio will be formed by the IFFToperation.

3.2 Watermark Extracting Approach

In the extracting, a predefined searching space, B is designed to de-scale the effects of various attacks on the centroid.

$$B = [A^{'}(1-\Delta 1), A^{'}(1-\Delta 2)] \tag{9}$$

Where $A^{'}$ denotes the centroid of watermarked audio signal. $\Delta 1$ And $\Delta 2$ denote the down and up error ratios of centroid in the FFT domain caused by various attacks. The hidden message is synchronization bits, followed by the hidden multi-bit watermark.

The histogram is extracted with L bins as in the process of watermark embedding. The hidden bit is extracted by comparing the number of coefficients in three consecutive bins, denoted by $a^{''}$, $b^{''}$ and $c^{''}$, formulated as:

$$w_i^{'} = \begin{cases} 1 & if \;\; 2b^{''}/(a^{''}+c^{''}) \ge 1 \\ 0 & other \end{cases} \tag{10}$$

The process is repeated until all hidden bits are extracted. Once synchronization sequence is matched with extracted synchronization bits or the searching process is finished, according to the best matching, extract the hidden watermark following the synchronization. In the extraction phase, the parameters, L_w, λ and synchronization sequence are known, so the detection process is blind.

4 Experimental Results

The parameter values are given as follows: $\lambda = 0.8$ and $T = 1.5$. And, 183 bins extracted from a 20s light music is watermarked with 61bits of information composed of a 31-bit m sequence and the 30-bit watermark. In the embedding, the probability of the watermarked samples their values added or reduced is approximately equivalent, hence the watermark hardly make any affection on the audio centroid, 137.9813 and 137.9517 before and after embedding respectively. The relative relation in the number of samples among three neighboring bins is calculated by Equation (5) and plot in Figure 5.

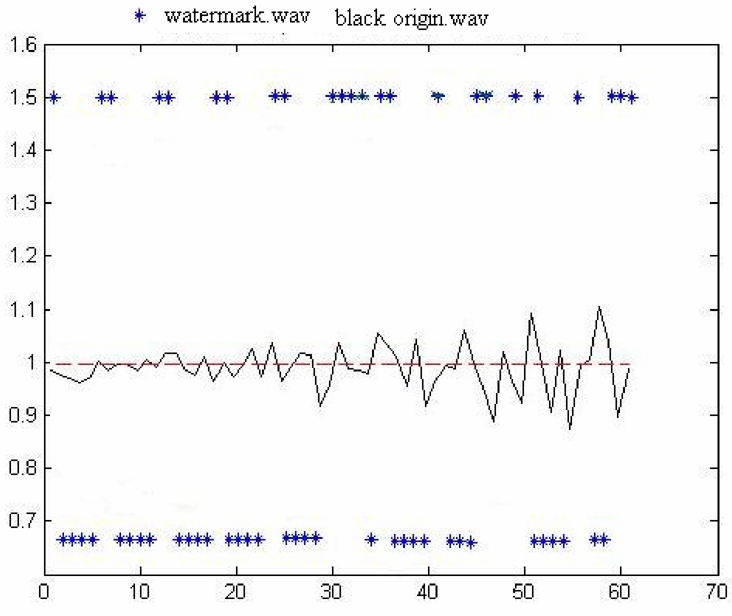


Fig. 5. The relative relation in the number of samples before and after watermarking

The SNR is 40.83dB. The higher SNR is that only a small part of samples is modified for watermarking. We test the robustness of the proposed algorithm according to IFPI with BER. The audio editing and attacking tools adopted in our experiments are Cool EditPro v2.1 and Stirmark Benchmark for Audio v0.2. The test results under common audio signal processing, time-scale modification and Stirmark for Audio are listed in Tables 1-3.

Table 1. Robustness performance to common attacks

Attack Type	Error Number of Bits	Attack Type	Error Number of Bits
Normalize	3	Low Pass (11025Hz)	0
Resample (kHz) 44.1->16->44.1	0	Low Pass (7kHz)	6
Re-quantization (bit) 16->32->16	0	Low Pass (4kHz)	10
Re-quantization (bit) 16->8->16	0		

Table 2. Robustness Performance to TSM with two different stretching modes

Pitch-Invariant TSM	Error Number of Bits	Resample TSM	Error Number of Bits
TSM –10%	Failed	TSM –10%	3
TSM –8%	15	TSM –8%	2
TSM –6%	12	TSM –6%	2
TSM –4%	8	TSM –4%	0
TSM –2%	6	TSM –2%	0
TSM +2%	6	TSM +2%	0
TSM +4%	8	TSM +4%	0
TSM +6%	11	TSM +6%	0
TSM +8%	14	TSM +8%	0
TSM +10%	Failed	TSM +10%	1

Table 3. Robustness Performance to some common attacks in Stirmark Benchmark for Audio

Attack Type	Error Number of Bits	Attack Type	Error Number of Bits
Addbrumm_100	0	Addnoise_100.wav	3
Addbrumm_1100	0	Addnoise_300.wav	5
Addbrumm_10100	0	Addnoise_500.wav	9
Addsinus	0	Compressor	0
Invert	0	Original	0
Stat2	3	Rc_lowpass	3
Zerocross	10	Zeroremove	Failed
Cutsamples	10	FFT_RealReverse	0

From Table 1, we can see that our algorithm is robust enough to some common audio signal processing manipulations such as resample, re-quantization and low pass of 11025Hz.

The test results of a light music under TSM form -10% to +10% with two different stretching modes are tabulated in Table 2. The algorithm shows strong robustness to this kind of attacks up to 10% for resample TSM.

Stirmark Benchmark for Audio is a common robustness evaluation tool for audio watermarking techniques. From Table 3, it is found that the watermark shows stronger resistance to those common attacks. In the cases of failure ('Failed' mean the number of error bits is over 20), the audio centroid is changed severely or audio quality is distorted largely.

5 Conclusions

A multi-bit audio watermarking method based on the centroid and statistical features in FFT domain is proposed and implemented by histogram specifications.

Extensive experiments shows that the superiority of statistical features, the relative relations in the number of samples among different bins and the frequency centroid of audio signal. The two features are very robust to the TSM. Accordingly, by applying the two features, audio watermarking scheme is designed.

The extensive experimental have shown that the watermark scheme is robust against some common signal processing, attack in Stirmark Benchmark for Audio and attack in resample TSM. However, it is still weak to resist pitch-invariant TSM attack.

References

1. International Federation of the Phonographic Industry, http://www.ifpi.org
2. Yin, X., Zhang, X.: Covert Communication Audio Watermarking Algorithm Based on LSB. 2006 10th International Conference on Communication Technology, pp. 308–311.
3. Huang, J., Wang, Y., Shi, Y.Q.: A Blind Audio Watermarking Algorithm with Self-Synchronization. In: Proceedings of IEEE International Symposium on Circuits and Systems, Arizona, USA, vol. 3, pp. 627–630 (2002)
4. Xiang, S., Huang, J., Yang, R.: Time-scale Invariant Audio Watermarking Based on the Statistical Features in time Domain. In: Proc. Of the 8th Information Hiding workshop (2006)
5. Yeo, I.K., Kim, H.J.: Modified Patchwork Algorithm: The Novel Audio Watermarking Scheme. IEEE Transactions on Speech and Audio Processing 11, 381–386 (2003)
6. Li, W., Xue, X.Y., Zh, P., Zh, P.L.: Robust audio watermarking based on rhythm region Detection. IEEE Electronics Letters 41(4), 218–219 (2005)
7. Sylvain, B., Michiel, V.D.V., Aweke, L.: Informed detection of audio watermark for resolving playback speed modifications. In: Proc.of the Multimedia and Security Workshop, pp. 117–123 (2004)
8. Coltuc, D., Bolon, P.: Watermarking by Histogram Specification. In: Proc. of SPIE International Conference on Security and Watermarking of Multimedia Contents II, vol. 3657, pp. 252–263 (1999)
9. Roy, S., Chang, E.C.: Watermarking Color Histograms. In: Proc. of International Conference of Image Processing, vol. 4, pp. 2191–2194 (2005)
10. Hu, Y.P., Han, D.Z., Yang, S.Q.: Image-adaptive Watermarking Algorithm Robust to Geometric Distortion in DWT Domain. Journal of System Simulation 17(10), 2470–2475 (2005)
11. Dai, Z.X, Hong, F., Li, X.G., Dong, J.: Improvement on centroid detection method for text document watermarking. Computer Applications 27(5), 1064–1066 (2007)