RESEARCH ARTICLE OPEN ACCESS

Increasing The Strength Of Audio Watermarking Against Desynchronization Attacks

Ms. Deepali Chavan*, Prof. R.R. Dube**
Department of Electronics, Walchand Institute of Technology, Solapur, Maharashtra, India.

Abstract
Rapid evolution of digital technology has improved the ease of access to digital information enabling reliable, faster and efficient storage, transfer and processing of digital data. It also leads to the consequences of making the illegal production and redistribution of digital media easy and undetectable. Hence, the risk of copyright violation of multimedia data has increased due to the enormous growth of computer networks that provides fast and error free transmission of any unauthorized duplicate and possibly manipulated copy of multimedia information. In audio watermarking area, the robustness against desynchronization attacks, such as TSM (Time-Scale Modification) and random cropping operations is still one of the most challenging issues. To enable the embedding of multiple watermarks a multi-bit robust audio watermarking algorithm based on the two statistical features is proposed by modifying the histogram. The audio histogram with equal-sized bins is extracted from a selected amplitude range referred to the audio mean, and then the relative relations in the number of samples among groups of three neighboring bins are designed to carry the watermark by reassigning the number of samples in the bins. The watermarked audio signal is perceptibly similar to the original one. Simulation results demonstrated that the hidden message is very robust to the TSM, cropping, and a variety of other distortions for Audio.

Keywords— Audio watermarking, cropping, histogram, jittering, synchronization, TSM.

I. INTRODUCTION

Broadband communication networks and multimedia data available in a digital format opened many challenges and opportunities for innovation. Versatile and simple-to-use software and decreasing prices of digital devices have made it possible for consumers from all around the world to create and exchange multimedia data. Broadband Internet connections and near error-free transmission of data facilitate people to distribute large multimedia files and make identical digital copies of them. A perfect reproduction in digital domain has promoted the protection of intellectual ownership and the prevention of unauthorized tampering of multimedia data to become an important technological and research issue. Digital watermarking has been proposed as a new, alternative method to enforce intellectual property rights and protect digital media from tampering. Digital watermarking is defined as imperceptible, robust and secure communication of data.

A watermark, which usually consists of a binary data sequence, is inserted into the host signal in the watermark embedder. Thus, a watermark embedder has two inputs; one is watermark message (usually accompanied by a secret key) and the other is the host signal (e.g. image, video clip, audio sequence etc.). The output of the watermark embedder is the watermarked signal, which cannot be perceptually discriminated from the host signal. The watermarked signal is then usually recorded or broadcasted and later presented to the watermark detector. The detector determines whether the watermark is present in the tested multimedia signal, and if so, what message is encoded in it.

Watermarking is the process of embedding the watermark within the host signal. Most importantly, the watermark signal should be imperceptible to the end user who is listening to or viewing the host signal. Another important requirement is that watermark signals must be reasonably resilient to common signal processing operations. In this paper, we used a multibit audio watermarking algorithm based on two statistical features of audio signals, concentrating on combating those challenging desynchronization problems, such as caused by TSM and random cropping attacks. The basic idea in our algorithm is that in the audio signal processing, TSM operations with the resample and pitch-variant stretching modes may be represented as an approximate temporal linear scaling operation, which has been verified by extensive testing. Theoretically, we can prove that the audio histogram shape and the audio mean are invariant to temporal scaling. In experimental testing, it is observed that the histogram shape (represented as the relative relations among groups of three neighboring bins) and the modified mean are rather robust to TSM attacks. We also evaluate the robustness of the histogram shape and mean to random cropping attacks. As a conclusion, the audio histogram shape and the mean can be taken as two robust features neighboring bins. The histogram

shape invariance is exploited to embed one binary sequence by controlling the relative relations in the number of samples in each three bins. In the extraction, the exhaustive search is avoided by detecting the watermark in a predefined space referred to TSM and random cropping attacks. As a robust feature, the mean is exploited to compute the histogram with equal-sized bins from a selected amplitude range. The original audio is not required in the extraction.

II. INVARIANT FEATURES TO TSM AND RANDOM CROPPING ATTACKS

A ***histogram*** is often used to describe the data distribution. The most common form of the audio histogram is obtained by splitting the range of the sample value into equal-sized bins. Then, the number of samples from the audio that fall into each bin is counted. The style of histogram may be described by

$H = \{hi \mid i=1 \ldots L\}$,

Where H is a vector denoting the volume-level histogram of the audio signal

$F = \{ f_i \mid i=1,\ldots N \}$

The pure ***mean*** of a given audio signal is calculated by adding up all the sample values and dividing by the number of them. Usually, it is a statistical measurement of the spread of data values and the divergence of the data values from normal distribution patterns. Since the pure mean of a signed audio signal is always close to zero, we apply the modified mean value of an audio signal in order to better estimate the statistical property of audio signal. The mean is calculated as the sum of the absolute values of all samples over its duration:

$$\bar{A} = \frac{1}{N}\sum_{i=1}^{N} |f(i)|$$

Where f (i) denotes the i^{th} sample value in F.

III. HISTOGRAM BASED WATERMARKING ALGORITHM

In this section a multi-bit watermark, Histogram based algorithm aiming at solving the TSM manipulations is designed. The watermark insertion and recovery are described by the histogram specification. The robustness of the audio mean and the relative relation in the number of samples among different bins to the TSM attacks are used in the design. The mean invariance property is used to select the amplitude range to embed bits so that the watermark can resist amplitude scaling attack and avoid exhaustive search. In the extraction, a synchronization code is exploited to eliminate the effect of TSM on the audio mean.

3.1 Embedding Strategy

The basic idea of the embedding strategy is to extract the histogram from a selected amplitude range. Divide the bins into many groups, each group including three consecutive bins. For each group, one bit is embedded by reassigning the number of samples in the three bins. The watermarked audio is obtained by modifying the original audio according to the watermarking rule.

The embedding model is shown in Fig. 3.1.

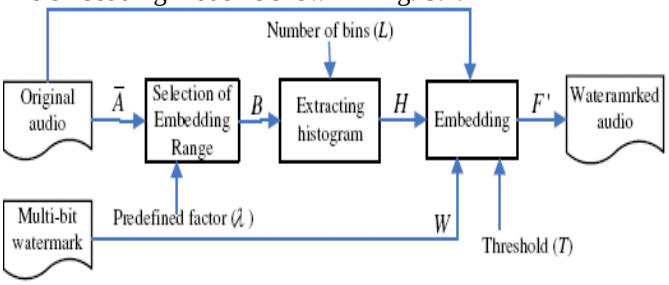

Fig 3.1 Watermark embedding framework

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 modified mean value of the $\bar{A}$ is calculated

$$\bar{A} = \frac{1}{N}\sum_{i=1}^{N} |f(i)|$$

Select the amplitude range $B = [-\lambda\bar{A},\lambda\bar{A}]$ from F to extract the histogram $H = \{ h_i \mid i=1,\ldots,L\}$, where $L \geq 3Lw$, to embed all watermark bits. λ is a selected positive number for satisfying $h(i) >> L$ and $\lambda \in [2.0, 2.5]$ is a suggested range so that the histogram bins extracted can hold enough samples. This observation is achieved based on the extensive testing on different kinds of audio signals.

Suppose that three consecutive bins, denoted by BIN 1, BIN 2 and BIN 3, their samples in the number are a, b and c, respectively. We apply the following watermarking rules to embed one bit of information, described as

$$2b/(a + c) \geq T \quad \text{if } w(i) = 1$$
$$(a + c)/2b \geq T \quad \text{if } w(i) = 0;$$

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 effectively resist the 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 neighboring bins, a, b and c, will be adjusted until satisfying $2b'/(a' + c') \geq T$. In case of embedding the bit '0', the procedure is similar. The rules applied to modify a, b and c as a', b' and c' are referred to Equations (1), (2), (3) and (4).

If the embedded bit $w(i)$ is '1' and $2b/(a + c) < T$, some selected samples from BIN 1 and BIN 3 in the number denoted by $I1$ and $I3$, will be modified to BIN 2, achieving $2b'/(a' + c') \geq T$. The modification rule is described as (1).

$$f1'(i) = f1(i)+M \quad 1 \leq i \leq I_1$$
$$f3'(i)=f_3(i)-M \quad 1 \leq i \leq I_3 \qquad (1)$$

where $f1(i)$ and $f3(i)$ denote the *ith* modified sample in BIN 1 and BIN 3, $I1$ and $I3$ are computed by using (2)

$$I1 = I \cdot a/(a + c); \quad I3 = I \cdot c/(a + c);$$
$$I \geq [T(a + c) - 2b]/(2 + T) \qquad (2)$$

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 BIN 1 and BIN 3, respectively, achieving $(a' + c')/2b' \geq T$. The rule is described as (3).

$$f'_2(i) = f2(i) - M \qquad 1 \leq i \leq I1$$
$$f_2(j) = f2(j) + M \qquad 1 \leq j \leq I3 \qquad (3)$$

where $f_2(i)$ denotes the *ith* modified sample in BIN 2, $f_2(i)$ and $f_2(j)$ are the corresponding modified version of $f2(i)$ and $f2(j)$. $I1$ and $I3$ are computed by(4)

$$I1 = I \cdot a/(a + c); \quad I3 = I. c/(a + c);$$
$$I \geq [2Tb - (a + c)]/(1 + 2T) \qquad (4)$$

This process is repeated to embed all watermark bits. In our proposed embedding strategy, the watermark is embedded by directly modifying the values of some selected samples from the original audio. Hence the embedding process includes the reconstruction of watermarked audio, which is denoted by

$$F' = \{f'(i) \mid i = 1,\ldots,N\}$$

3.2 Watermark Extraction

Consider the effects of the TSM on the audio mean may cause the watermark detection failed, a predefined searching space denoted by $[\overline{A}''(1 - \Delta 1), \overline{A}''(1 + \Delta 2)]$ is designed for resynchronization. Here, $\overline{A}''$ denotes the mean of the watermarked audio $F'' = \{f''(i) \mid i = 1,\ldots,N''\}$ which has undergone some de synchronization attacks, such as TSM operations with different stretching modes. Based on our previous experimental analysis $\Delta 1$ and $\Delta 2$, the down and up searching error ratios of mean, are suggested not less than 5%. We use a PN (Pseudo-random Noise) sequence as a synchronization code, followed by the hidden multi-bit watermark. Only the watermark also provides the synchronization capability. The merit of part of payload as synchronization code can keep the watermark unknown for the detector. Our goal is to get an estimate of hidden bits,

$$W'' = \{w_i \mid i = 1,\ldots,Lw\}$$

by selecting an amplitude range from F'' at a low error rate. W'' is composed of $Syn(i)''$ and $Wmk(i)''$. The histogram of F'' is extracted with L bins as in the process of watermark embedding. Compute the number of samples in three consecutive bins and denoted by a'', b'' and c''. By comparing them, we can extract one bit of hidden information,

$$w''i = \begin{cases} 1 & \text{if } 2b''/(a'' + c'') \geq 1 \\ 0 & \text{other} \end{cases} \qquad (5)$$

The process is repeated until all hidden bits are extracted. Once the synchronization code $Syn(i)''$ is matched with the extracted synchronization bits $Syn1(i)''$ or the searching process is finished, according to the best matching, we extract the hidden watermark following the synchronization bits, denoted by $Wmk1(i)''$. In the extraction, the parameters, Lw, and $Syn(i)''$, are beforehand known, so the detection process is blind.

IV. PERFORMANCE PARAMETERS

The performance parameters used for the performance are bit error rate (BER), signal to noise ratio (SNR) are discussed below.

4.1) Bit Error Rate

Bit error rate can be defined as the percentage of bits corrupted in the transmission of digital information due to the effects of noise, interference and distortion. For example, the bits to be transmitted are 11001100 and the received bits are 10000100. Comparing the number of bits transmitted to received, two bits are affected by transmission. Hence, the BER in this example is 2/8*100 = 25%.

Generally the BER is computed using equation below.

BER = Number of error bits/Number of total bits.

4.2) Signal to Noise Ratio

Signal to noise ratio is a parameter used to know the amount by which the signal is corrupted by the noise. It is defined as the ratio of the signal power to the noise power. Alternatively, it represents the ratio of desired signal (say a music file) to the background noise level. Signal to noise ratio can also be calculated by equation below. X (i) is the un-watermarked audio signal and $\tilde{x}$ is the watermarked audio signal.

$$SNR = 10 \log_{10} \left\{ \frac{\sum_{i=1}^{N} x^2(i)}{\sum_{i=1}^{N} [\tilde{x}(i) - x(i)]^2} \right\}$$

V. ROBUSTNESS TO SIGNAL PROCESSING

Watermarked digital signals may undergo common signal processing operations such as linear filtering, sample requantization, D/A and A/D conversion, and lossy compression. Although these operations may not affect the perceived quality of the host signal, they may corrupt the watermark data embedded within the signal. It is important to know, for a given level of host signal distortion, which watermarking algorithm will produce a more reliable embedding. In this paper, robustness was measured by the bit error rate (BER) of extracted watermark data as a function of the amount of distortion introduced by a given operation.

The following signal processing attacks are performed to assess the robustness of our scheme. The audio editing and attacking tools adopted in the experiment are Cool Edit Pro 2.1, Wave pad Sound Editor and Gold Wave. A Danube.wav of 20s file is used as a host signal.

5.1 Cropping

Both random samples cropping and zeros inserting belong to geometric distortions, producing disastrous synchronization problem in a

straightforward way. Random samples cropping refers to deleting some samples at some randomly selected locations, at the beginning, somewhere in the middle, or in the end. Perception of cropping depends not only on the amount of samples removed, but also on the amplitude of that clipping. Roughly speaking, one cropping less than 10 samples would not give rise to obvious discontinuity under the normal volume. For our robustness test, a large number of samples, 500 and 1000, at a certain position will be cropping.
Following Figures illustrates Graphs of different wav files after applying different attacks.

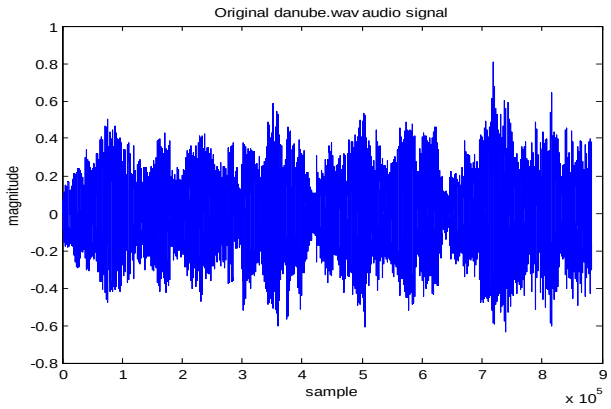

Fig 5.1. Original Audio Signal

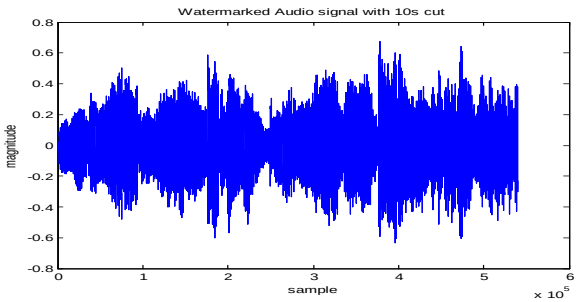

Fig 5.2. Watermarked Audio signal with 10s random cropping

Table 5.1- Performance evaluation of Danube.wav audio signal with cropping watermarked file of different length in different direction

Sr No	Cropping length	Direction	BER
1	10 s	Front side	0
2	10 s	Back side	0
3	10 s	Middle	0
4	10 s	Randomly each 2 s	0
5	12 s	Front side	0
6	15 s	Front side	0.0408

5.2 Jittering

Jittering is an evenly performed case of random samples cropping, which means randomly cropping samples out of every samples. As a continuous case of random samples cropping, jittering will lead to loss of synchronization sign. In our survival test, 1 sample is cleared up out of every 10,100,500, 1000 samples.

Table 5.2- Performance evaluation of Danube.wav audio signal with jittering watermarked file

Sr. No	Jittering	BER
1	1/10	0.3265
2	1/100	0
3	1/400	0
4	1/500	0
5	1/1000	0

5.3 Resampling

Sampling rate, *Fs*, is an important parameter for digital signal. In this work, all the audio files are sampled at 44.1 k Hz and quantized in 16 bits, with CD quality. So the sampling frequency of the watermarked signal is also equal to 44.1 kHz. To examine the property of resisting resampling, the watermarked signal originally sampled at 44.1 kHz, is resampled at 32 kHz, 48 kHz and 96 kHz.

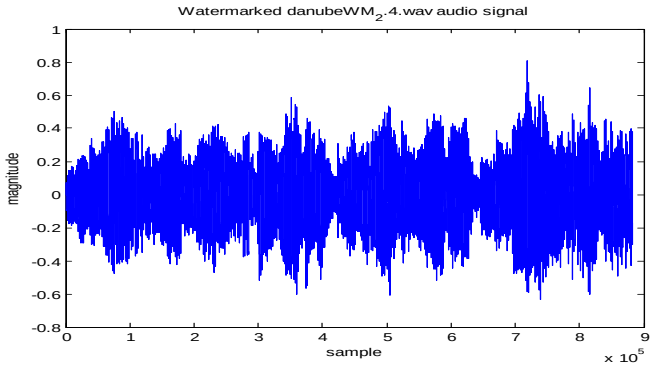

Fig 5.3 Watermarked Audio Signal

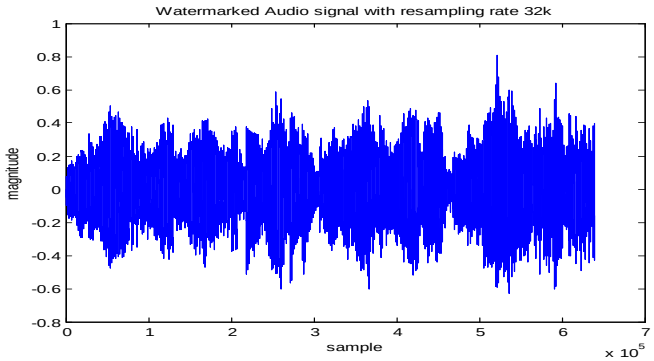

Fig.5.4 Watermarked audio signal with 32 KHz Resampling

Table 5.3- Performance evaluation of Danube.wav audio signal with resampling

Sr. No	Resampling	BER
1	32kHz	0
2	48kHz	0
3	96kHz	0

VI. CONCLUSIONS

In this paper, we have presented a straightforward performance evaluation framework for histogram based watermarking algorithms based on bit error rate, signal to noise ratio and robustness to signal processing operations .Our results show that histogram based approach of watermarking increases the strength of audio watermarking against different desynchronization attacks such as TSM operations like resampling, cropping and jittering .

Also, the watermark can resist some common audio signal processing operations to some extent. This work is very useful due to its strong robustness to TSM and random cropping manipulations. The security of the watermarking scheme is a consideration of future researches. Another consideration is to extend the histogram shape invariance from the time domain to the transform domains to further improve the watermark performance to noise-like audio processing operations, such as MP3 compression, low-pass filtering, additional noise, etc.

REFERENCES

[1]. M. Arnold, "Audio Watermarking: Features, Applications and Algorithms". *Proc. of IEEE International Conference on Multimedia and Expo, Vol. 2*, New York, USA, (2000) 1013-1016

[2]. Oscal T.-C. Chen, Wen-Chih Wu, "Highly Robust, Secure, and Perceptual-Quality Echo Hiding Scheme", *IEEE Transactions on Audio, Speech, And Language Processing, VOL. 16, NO. 3,* MARCH 2008

[3]. M. D. Swanson, B. Zhu and A. H. Tew¯ k:" Current State of the Art, Challenges and Future Directions for Audio Watermarking". *Proc. of IEEE International Conference on Multimedia Computing and Systems, Vol. 1* (1999) 19-24

[4]. S. Katzenbeisser, F. A. P. Petitcolas, ed." Information Hiding Techniques for Steganography and Digital Watermarking". Artech House, Inc. (2000)

[5]. Mohammad A. Akhaee, Mohammad J. Saberian, Soheil Feizi, "Robust Audio Data Hiding Using Correlated Quantization With Histogram-Based Detector", *IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 11, NO. 5,* AUGUST 2009

[6]. S. Wu, J. Huang, D. Huang and Y. Shi: "Efficiently Self-Synchronized Audio Watermarking for Assured Audio Data Transmission*". IEEE Trans. On Broadcasting, Vol. 51* (2005) 69-76

[7]. D. Kirovski and H. Malvar: "Robust Covert Communication over A Public Audio Channel Using Spread Spectrum". *Proc. of Information Hiding Workshop,* (2001) 354-368

[8]. R. Tachibana, S. Shimizu, T. Nakamura, and S. Kobayashi: "An Audio Watermarking Method Robust against Time and Frequency Fluctuation". *Proc. of SPIE International Conference on Security and Watermarking of Multimedia Contents III, Vol.* 4314 (2001) 104-115

[9]. M. Mansour and A. Tew: " Time- Scale Invariant Audio Data Embedding". *Proc. of IEEE International Conference on Multimedia and Expo,* (2001) 76-79

[11]. S. Xiang, H.J. Kim. "Invariant Audio Watermarking in DWT Domain". ICUT2007.

[12]. W.N. Lie, L.C. Chang. "Robust and High-quality Time domain Audio Watermarking based on Low-frequency Amplitude Modification". IEEE Transactions on Multimedia. 2006.