Audio Watermarking Based on Mean Quantization in Cepstrum Domain

Vivekananda Bhat K, Indranil Sengupta, and Abhijit Das

Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur 721 302, India
{vbk, isg, abhij}@cse.iitkgp.ernet.in

***Abstract*—A novel audio watermarking algorithm is proposed in this paper for audio copyright protection. The method is based on cepstrum domain transform. This algorithm embeds the watermark data into original audio signal using mean quantization of cepstrum coefficients. Experimental results shows that our audio watermarking scheme is not only imperceptible, but also robust against various common signal processing attacks such as noise adding, re-sampling, low-pass filtering, re-quantization, MP3 compression and cropping. In addition, the algorithm can extract the watermark without the help of original audio signal, and its performance is better than cepstrum domain audio watermarking scheme based on statistical mean manipulation.**

Keywords: Audio watermarking, cepstrum domain, mean quantization.

I. Introduction

In the recent years, the copyright protection of digital audio media attracted the interest of scientists and engineers. Most of the watermarking schemes focus on images and videos. A very few audio watermarking techniques have been reported in the literature. Digital audio watermarking is a process of embedding watermark data inaudibly into audio signal. According to the International Federation of Phonographic Industry (IFPI) [1], audio watermarking should meet the following requirements: (i) *Imperceptibility*: The digital watermark should not affect the quality of original audio signal after it is watermarked. (ii) *Robustness*: The embedded watermark data should not be removed or eliminated by unauthorized distributors using common audio signal processing operations and attacks, such as additive and multiplicative noise, re-sampling , re-quantization, MP3 compression, cropping, D/A and A/D conversion, pitch scaling, time stretching, etc. (iii) *Capacity*: Capacity refers to the number of bits that can be embedded into the audio signal within a unit of time. (iv) *Security*: Security implies that the watermark can only detectable by the authorized person. All these requirements are often contradictory with each other and we need to make a trade-off among them. These conflicting requirements poses many challenges to design of robust audio watermarking. There are other requirements for audio watermarking, but the most important properties it should satisfy are imperceptibility and robustness.

Several algorithms have been proposed in the past decade [2]. Lee *et al*. [3] first proposed a digital audio watermarking technique in cepstrum domain. Watermark is inserted into cepstrum coefficients of the audio signal using techniques analogous to spread spectrum communications. Li and Yu [4] proposed a transparent and robust audio data hiding technique in cepstrum domain. The results shows that the method is robust to a wide range of attacks. Hsieh *et al*. [5] discussed the audio watermarking method based on time energy features. The simulation results shows the high security performance against the MP3 compression and some kind of synchronization attacks. BCH code based robust audio watermarking in the cepstrum domain is explained in [6]. Experimental results demonstrates the proposed technique outperforms existing audio watermarking techniques against most of the asynchronous attacks. Li *et al*. [7] proposed an audio watermarking method in cepstrum domain based on statistical mean manipulation (SMM). Extensive experimental results prove that the embedded watermark is inaudible and robust. The embedded watermark is robust to multiple watermarks, MP3 compression and additive noise. Robust wavelet domain audio watermarking based on mean quantization was presented in [8], [9], [10]. This paper presents an audio watermarking algorithm in the cepstrum domain based on mean quantization. Our work based on the motivation of cepstrum domain technique [4] and mean quantization techniques [9], [10]. The main features of the proposed algorithm are as follows: (i) The binary watermark is encrypted using random permutation. (ii) Cepstrum coefficients are modified using mean quantization to enhance the robustness. (iii) Watermark extraction is blind without using the host signal. (iv) The proposed algorithm has better performance compared to the scheme [7] against common signal processing attacks. The remainder of this paper is organized as follows. A brief overview of cepstrum domain transform and mean-quantization technique are explained in section II and section III respectively. Section IV will combine the above techniques to propose a new method for audio watermarking. Moreover, the detail watermark embedding and extraction algorithm is explained in this section. The experimental results and comparison are given in section V, and finally the conclusion is provided in section VI.

II. Cepstrum Domain

The cepstrum domain analysis is a tool widely used in speech processing and recognition [7], [11]. It consists of three consecutive steps: Fourier transform, take logarithm and inverse Fourier transform. It is easy to see that these three operations are all linear and we can exactly recover the original signal in time domain from its cepstrum domain. Large cepstrum coefficients around the center are mostly perceptually significant and we will not use them for embedding in order to achieve imperceptibility. Therefore we shall only modify small cepstrum coefficients around both sides of center. Experimental studies have shown that statistical mean of the cepstrum coefficients experience much less variance after most common signal processing attacks than original samples in time domain. Due to the attack-invariant feature, the watermark information can be preserved. It should be noted that the logarithm in the second step is a complex logarithm. But in practice, for convenience people often define the real part of complex cepstrum to be the real cepstrum and is given by:

$$X(n) = REAL(IFFT(log(FFT(x(n)))))$$

We can recover the original signal in time domain from its cepstrum domain by taking inverse operations:

$$x(n) = REAL(IFFT(exp(FFT(X(n)))))$$

III. Mean Quantization

Mean quantization based audio watermarking [9] is the simplest among all blind watermarking schemes, due to use of simple function for watermarking embedding. Watermark data was embedded in cepstrum domain of the original audio signal by quantizing the means of cepstrum coefficients. The same function is used for watermark extraction. By quantizing mean values, better robustness against attacks in comparison to single value quantization is achieved. Single value sample quantization methods are more sensitive than mean quantization ones. The mean value is more difficult to move and therefore the probability of the watermark error is smaller in mean quantization approaches. Let $\{y_1, y_2, ..., y_n\}$ be N cepstrum coefficients and its mean is given by:

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{n} y_i \tag{1}$$

Embedding one watermark bit of $w_i \in \{0, 1\}$ using quantization method causes error in $\bar{y}$ and the error is Δ, so the new mean is $\bar{y}^* = \bar{y} + \Delta$ after embedding one watermark bit. Hence all the coefficients in the cepstrum domain are also modified, namely $y_i^* = y_i + \Delta$, $i = 1, 2, ..., N$, where y_i^* is a coefficient after modification. Therefore from the above it follows that when embedding one watermark bit using mean of N coefficients, the error is added to each of N coefficients.

IV. The Proposed Audio Watermarking Scheme

We present a novel mean quantization based audio watermarking in cepstrum domain. The general steps involved in the watermark embedding algorithm and watermark extraction algorithm are outlined below:

A. Watermark Embedding Algorithm

The basic steps involved in the watermark embedding process are given below:

Step 1: The logo image P is a binary image of size $M \times M$ and is given by $P = \{p(m_1, m_2) : 1 \leq m_1 \leq M, 1 \leq m_2 \leq M, p(m_1, m_2) \in \{0, 1\}\}$. We should convert the two-dimensional binary image into the one-dimensional sequence in order to embed it into the audio signal. The corresponding one-dimensional sequence is given by: $W = \{w(k) = p(m_1, m_2) : 1 \leq m_1 \leq M, 1 \leq m_2 \leq M, k = (m_1 - 1) \times M + m_2, 1 \leq k \leq M \times M\}$. Which is further encrypted using random permutation to ensure security.

Step 2: The original audio is first segmented into non-overlapping frames, with a frame size of 1024 samples. The number of frames is equal to size of binary image.

Step 3: The n^{th} sample of m^{th} frame is denoted as $x(n, m)$. Transform the time domain audio frame $x(n, m)$ into cepstrum coefficients $c(n, m)$ in cepstrum domain.

Step 4: Calculate the mean $\bar{c}(n, m)$ of each frame of cepstrum coefficients. Modify the mean of cepstrum coefficients to zero and the modified coefficients are denoted as $c^{'}(n, m)$.

$$c^{'}(n, m) = c(n, m) - \bar{c}(n, m)$$

Step 5: Watermark bit is embedded using mean quantization as follows:

$$c^{''}(n, m) = c^{'}(n, m) + \bar{c}^*(n, m)$$

where $\bar{c}^*(n, m)$ is the mean quantization factor. Assume $D(m, n) = \lfloor \frac{\bar{c}(n,m)}{Q} + \frac{1}{2} \rfloor$, where Q is a positive real number called quantization parameter and $\lfloor\ \rfloor$ is the floor integer function. If $mod(D(m, n), 2) = w(k)$, then $\bar{c}^*(n, m) = D(m, n) \times Q$. If $mod(D(m, n), 2) \neq w(k)$ and $D(m, n) = \lfloor \frac{\bar{c}(n,m)}{Q} \rfloor$, then $\bar{c}^*(n, m) = (D(m, n) + 1) \times Q$. If $mod(D(m, n), 2) \neq w(k)$ and $D(m, n) \neq \lfloor \frac{\bar{c}(n,m)}{Q} \rfloor$, then $\bar{c}^*(n, m) = (D(m, n) - 1) \times Q$.

Step 6: Transform the cepstrum coefficients $c^{''}(n, m)$ to watermarked audio signal $x^{''}(n, m)$ in the time domain using inverse cepstrum transform.

B. Watermark Extraction Algorithm

The extraction algorithm can be performed without using original audio signal. The basic steps involved in the watermark extraction are given below:

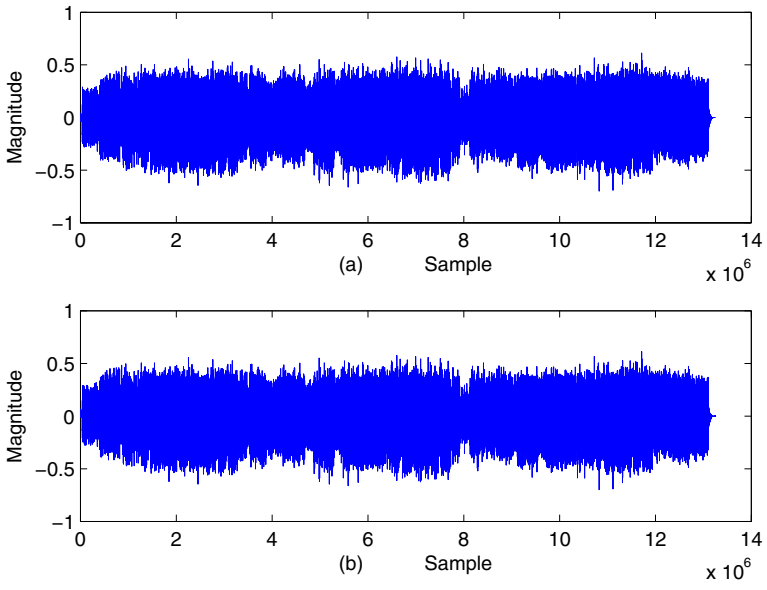


Fig. 1. (a) Classical audio signal (b) Watermarked classical audio signal

Step 1: The watermarked audio signal is segmented into non-overlapping frames, with a frame size of 1024 samples.

Step 2: The frame is transformed into cepstrum domain.

Step 3: Let $\bar{c}^{'}(n,m)$ be the mean of cepstrum coefficients. The extracted watermark is given by, $w^{'}(k) = mod(\lfloor \frac{\bar{c}^{'}(n,m)}{Q} + \frac{1}{2} \rfloor, 2)$

Step 4: The extracted watermark information is decrypted using the same random permutation which was used during embedding process.

Step 5: Convert one-dimensional watermark sequence into two-dimensional sequence to get the desired extracted watermark image.

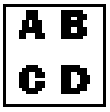


Fig. 2. Binary watermark

V. Experimental Results and Comparison

We tested our algorithm on 16 bits signed mono audio signals of different types sampled at 44.1 kHz in WAVE format. A plot of a short portion of the classical audio signal and the corresponding watermarked classical audio signal is shown in Fig. 1. In our experiment, we have used a 32×32 binary logo image as our watermark, which is displayed in Fig. 2. The quantization parameter Q is a user defined positive real number. A small value of Q makes a minor change to coefficients in embedding step and hence makes the watermark sensitive to audio modification. On the other hand, a large value of Q makes the watermark robust to many manipulations, but it will achieve a bad imperceptibility. Hence we have to make a trade-off between robustness and imperceptibility of the watermark with different value of the quantization parameter. The signal to noise ratio (SNR) for evaluating the quality of watermarked signal is given by the equation:

$$SNR = 10 \log_{10} \frac{\sum_{n=1}^{N} X^2(n)}{\sum_{n=1}^{N} [X(n) - X^*(n)]^2} \qquad (2)$$

where X and X^* are original audio signal and watermarked audio signal respectively. After watermark embedding, the SNR of all selected audio signals are above 20 dB. This ensures the imperceptibility of our scheme. In order to test the robustness of our scheme, six different types of following attacks were performed to the watermarked audio signal.

(i) Additive white Gaussian noise (AWGN): White Gaussian noise is added so that the resulting signal has a SNR of 30 dB.

(ii) Re-sampling: The watermarked signal originally sampled at 44.1 kHz is re-sampled at 22.05 kHz, and then restored by sampling again at 44.1 kHz.

(iii) Low-pass filtering: The low-pass filter used here is a second order Butterworth filter with cut-off frequency 11025 Hz.

(iv) Re-quantization: The 16-bit watermarked audio signals have been re-quantized down to 8 bits/sample and back to 16 bits/sample.

(v) MP3 Compression: The MPEG-1 layer 3 compression with 64 kbps and 32 kbps is applied.

(vi) Cropping: Segments of 500 samples were removed from the beginning of the watermarked signal.

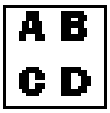
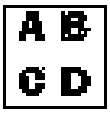
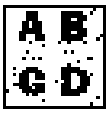
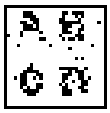
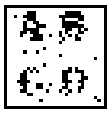
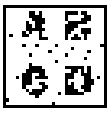
Watermark						
NC	1	0.9957	0.9814	0.9725	0.9363	0.9264
Watermark						
NC	0.9102	0.9039	0.8863	0.8761	0.8605	0.8583

Fig. 3. Extracted watermark with various NC

The normalized correlation (NC) is used to evaluate the similarity measurement of extracted binary watermark, which can be defined as:

$$NC(W,W^*) = \frac{\sum_{i=1}^{M}\sum_{j=1}^{M} W(i,j)W^*(i,j)}{\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{M} W(i,j)^2}\ \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{M} W^*(i,j)^2}} \quad (3)$$

where W and W^* are original and extracted watermarks respectively, i and j are indexes of the binary watermark image.

The bit error rate (BER) is used to find the percentage of error bits between original watermark and extracted watermark. The BER is given by:

$$BER = \frac{B}{M \times M} \times 100\%, \quad (4)$$

where B is the number of erroneously detected bits.

The Fig. 3., shows the extracted watermark with various NC values. The extracted watermark image still have some distortion, but the extracted watermark image is similar to original watermark image. Table I shows the results of the performance comparison between proposed watermarking scheme and Li *et al*. [7] scheme for different types of audio files against different types of attacks. The performance of our scheme is better than the scheme [7] for the above mentioned attacks, in particular it is robust to MP3 compression at very low bit rate. The low BER and high NC between extracted watermark and original watermark indicates that the proposed scheme has very high robustness. Performance of the proposed scheme is mainly depends on the binary logo watermark image and audio files used in the experiment. The NC value of our scheme against MP3 compression attack at 32 kbps for Blues audio file is 0.8583, where as the NC value for the scheme [7] is 0.8115 against the same attack. We can also observe that the audio files like Classical, Country, Blues and Pop stand out as good host signals for our watermarking scheme, because under above attacks watermark recovery is not disturbed significantly.

VI. Conclusions

A new blind digital audio watermarking algorithm is proposed in this paper. Different from traditional cepstrum domain methods, the watermark is embedded using mean quantization of cepstrum coefficients. By applying several attacks, the robustness of the proposed algorithm is tested. The experimental results show that the proposed method has good capabilities of robustness and inaudibility. In particular it is robust especially against MP3 compression at low bit rate, while maintaining high correlation between extracted watermark and original watermark. Hence our scheme guarantees the better performance compared to cepstrum domain audio watermarking scheme based on SMM. In addition, computation and implementation in our proposed scheme is very straight forward process. All these merits enhance the practicality and application value for copyright protection.

References

[1] S. Katzenbeisser and F. A. P. Petitcolas, *Information hiding techniques for steganography and digital watermarking*. Boston: Artech House, January 2000.

[2] N. Cvejic and T. Seppanen, *Digital audio watermarking techniques and technologies*. USA: Information Science Reference, August 2007.

[3] S.-K. Lee and Y.-S. Ho, "Digital audio watermarking in the cepstrum domain," *IEEE Transactions on Consumer Electronics*, vol. 46, no. 3, pp. 744–750, 2000.

[4] X. Li and H. H. Yu, "Transparent and robust audio data hiding in cepstrum domain," in *IEEE International Conference on Multimedia and Expo*, vol. 1, 2000, pp. 397–400.

[5] C.-T. Hsieh and P.-Y. Sou, "Blind cepstrum domain audio watermarking based on time energy features," in *14th International Conference on Digital signal processing*, vol. 2, 2002, pp. 705–708.

[6] S.-C. Liu and S. D. Lin, "BCH code based robust audio watermarking in the cepstrum domain," *Journal of Information Science and Engineering*, vol. 22, pp. 535–543, 2006.

[7] S. Li, L. Cui, J. Choi, and X. Cui, "An audio copyright protection schemes based on SMM in cepstrum domain," in *International Workshops on Structural, Syntactic, and Statistical Pattern Recognition (SSPR and SPR'06), LNCS*, vol. 4109, 2006, pp. 923–927.

[8] N. Kalantari, S. Ahadi, and A. Kashi, "A robust audio watermarking scheme using mean quantization in the wavelet transform domain," in *IEEE International Symposium on Signal Processing and Information Technology*, 2007, pp. 198–201.

[9] W. Lanxun, Y. Chao, and P. Jiao, "An audio watermark embedding algorithm based on mean-quantization in wavelet domain," in *8th International Conference on Electronic Measurement and Instruments*, 2007, pp. 2–423–2–425.

TABLE I
NC VALUES AND BER OF EXTRACTED WATERMARK FROM DIFFERENT TYPES OF ATTACKS

Audio file	**Type of attack**	**Proposed scheme (NC)**	**Scheme [7] (NC)**	**Proposed scheme (BER (%))**	**Scheme [7] (BER(%))**
Classical	Attack free	1	1	0	0
	AWGN	1	1	0	0
	Re-sampling	1	1	0	0
	Low-pass filtering	1	1	0	0
	Re-quantization	0.9957	0.9820	1	3
	MP3 64 kbps	0.9196	0.8863	13	17
	MP3 32 kbps	0.9229	0.8761	12	19
	Cropping	0.9496	0.9264	8	12
Country	Attack free	1	1	0	0
	AWGN	1	1	0	0
	Re-sampling	1	1	0	0
	Low-pass filtering	1	1	0	0
	Re-quantization	0.9814	0.9725	3	4
	MP3 64 kbps	0.9102	0.8753	14	20
	MP3 32 kbps	0.9119	0.8701	14	20
	Cropping	0.9363	0.8564	11	23
Blues	Attack free	1	1	0	0
	AWGN	1	1	0	0
	Re-sampling	1	0.9857	0	2
	Low-pass filtering	1	0.9857	0	2
	Re-quantization	0.9975	0.9782	0	3
	MP3 64 kbps	0.8605	0.8113	22	31
	MP3 32 kbps	0.8583	0.8115	22	31
	Cropping	0.9687	0.9136	5	14
Pop	Attack free	1	1	0	0
	AWGN	1	1	0	0
	Re-sampling	1	1	0	0
	Low-pass filtering	1	1	0	0
	Re-quantization	0.9895	0.9815	2	3
	MP3 64 kbps	0.8027	0.7473	33	38
	MP3 32 kbps	0.8020	0.7388	33	39
	Cropping	0.9039	0.8240	14	29

[10] R.-D. Wang, D.-W. Xu, and Q. Li, "Multiple audio watermarks based on lifting wavelet transform," in *Fourth International Conference on Machine Learning and Cybernetics*, vol. 4, 2005, pp. 1959–1964.

[11] J. D. John R., J. H. L. Hansen, and J. G. Proakis, *Discrete-time processing of speech signals*. Mcmillan, 1993.