Spread Spectrum Audio Watermark Based on Non-uniform Quantization

MeiJun Ning, Tao Peng(✉), YueQing Xu, and QingYi Quan

Wireless Signal Processing and Networks Laboratory (WSPN),
Key Laboratory of Universal Wireless Communications, Ministry of Education,
Beijing University of Posts and Telecommunications, Beijing, China
{ningmeijun,pengtao,xuyueqing10952,qyquan}@bupt.edu.cn

Abstract. Audio watermarking is an information hiding technology which is widely used in copyright protection and information security. This paper proposes a novel audio watermarking scheme based on spread spectrum and non-uniform quantization. The watermarks are embedded by modifying the quantization coefficients. The proposed algorithm utilizes the characteristics of non-uniform quantization to adopt different quantized signal-to-noise ratios for the low-frequency and high-frequency parts of the audio signal, thus improving the robustness of the technology while ensuring the sound quality is not damaged. Compared with the existing audio watermarking methods, the proposed scheme is especially robust against additive white Gaussian noise (AWGN). Experimental analysis shows that the proposed method provides high audio quality and excellent capability to withstand various noise attacks particularly in AWGN.

Keywords: Audio watermarking · Non-uniform quantization · Robustness

1 Introduction

In recent years, due to the rapid development of the Internet, digital multimedia data can be easily distributed to all over the world. But this has brought about a critical issue for copyright protection. In that context, the audio watermarking technology was proposed. Audio watermarking technology uses audio as a carrier to embed information. The human ear cannot distinguish the difference between the original audio and the watermark audio, hence the copyright information which can be used to verify the original author can be hidden in the audio. Moreover, the audio watermarking technology plays an important role not only in copyright protection, but in other fields, such as content authentication, information security, broadcast monitoring and secret communication.

This work is supported in part by the National Natural Science Foundation of China (No. 61631004) and the National Science and Technology Major Project of China under Grant 2016ZX03001017.

H. Gao et al. (Eds.): ChinaCom 2019, LNICST 312, pp. 235–245, 2020.
https://doi.org/10.1007/978-3-030-41114-5_18

Generally speaking, the audio watermark technology should satisfy three requirements:

- Robustness: The embedded watermarks ought to tolerant different types of attacks such as MP3 compression, requantization, amplitude compression and AWGN, etc.
- Imperceptibility: The embedded watermarks should not affect the quality of the original signal.
- Payload: The amount of watermark that can be embedded into the host audio signal.

However, these three requirements are contradictory, for instance, in order to improve the robustness, the origial signals will make greater changes, which will bring about lower imperceptibility. Consequently, there is a trade-off bewteen these requirements.

The audio watermarking technique can be classified into two types: time domain technique and frequency domain technique. The time domain technique including LSB hiding [1] and echo hiding [2] embeds information in the time domian. Although it's easy to implement, it has the weakness of low robustness. Thus, the frequency domain technique has attracted the attention of many scholars. The frequency domain methods hide the messages into the audio signals which have been converted to frequency domain. The general transform methods are Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT) and Discrete Wavlet Transform (DWT). Chen et al. [3] proposed a method to modify the low-frequency efficients in DWT, this algorithm is robust against common signal processing attacks including compression and time scale modification, but has a weak robustness to amplitude scaling, cropping. Li et al. [4] presented a spread spectrum audio watermarking based on perceptual characteristic aware which has good robustness and imperceptibility, while the scheme's payload is low. Similarly, Erfani et al. [5] put forward an audio watermarking technique rely on a perceptual kernel representation of audio signals (spikegram), it owns outstanding performance in resisting audio processing attacks but achieves worst payload. With a superior payload, Kaur et al. [6] proposed a blind audio watermarked algorithm based on SVD, while this algorithm has poor property in robustness. Garlapati et al. [7] brought forward a enhanced spread spectrum audio watermark, it's proved that it has high payload but without the proof of robustness agagin the AWGN attack. Identically, Gupta et al. [8] proposed an efficient audio watermarking scheme based on lifting wavelet transform (LWT) and quantization. Aparna et al. [9] adopted Modified Discrete Fourier Transform (MDCT) in the proposed algorithm, but they didn't verify the capability against noise attack including AWGN. Proposed by Wang et al. [10], the scheme with MDCT could resist the common attacks, it achieved excellent performance in MP3 compression and recompression attacks but had a weak property in combating with AWGN. Sakai et al. [11] presented a develpoed watermark algorithm to withstand the band-pass filtering attacks. However, in wireless communication, the AWGN noise is more common and serious, so this paper proposes an advanced watermark scheme based on MDCT to resist the noise attacks. This strategy not only can combat with AWGN attack, but also can endure other attacks.

The remaining of the paper is organized as follows. Section 2 discusses the preliminary in the audio watermark. In Sect. 3 the noval method for audio watermarking is proposed and Sect. 4 shows the experimental results. Finally, Sect. 5 concludes the paper.

2 Preliminary

Due to high compression ratio and brilliant tone quality, MPEG Audio Layer III (MP3) is a popular compression format. MP3 encoder's procedure of processing audio mainly includes the following parts: MDCT, the psychoacoustic model II, the bit allocation, quantization and bitstream formatting. The proposed method mainly takes use of MDCT, psychoacoustic model II and the process of quantization.

2.1 MDCT in MP3 Encoding

During the MP3 encoding, MDCT is used to convert the time domain signal into frequency domain. The encoder divides audio signals into frames. Every frame is composed of two granules and frame's length is 1152. After MDCT, one granule signal is divided into 32 subbands and each subband has 18 frequency lines.

MDCT is a linear orthogonal overlap transform. MDCT can effectively overcome the edge effect in DCT without reducing the coding performance, thus effectively removing the periodic noise generated by edge effects. The mathematical expression for MDCT is given.

$$X(i) = \sum_{k=0}^{N-1} Z(k) cos[\frac{(2k+1+\frac{N}{2})(2i+1)\pi}{2N}] \tag{1}$$

where $Z(k)$ is the result of multiplying audio signals and window function [12], $i = 0, 1, ..., \frac{N}{2} - 1$, $N = 12$ or 36.

2.2 Quantization in MP3 Encoding

The MP3 encoder adopts a non-uniform quantizer. The non-uniform quantizer determines the quantization interval based on different intervals of the signal. That means the encoder performs different processing on signals of different frequencies. MP3 exploits the characteristic of non-uniform quantization to ameliorate the sound quality. The quantitative calculation formula is

$$q_i = \lfloor (\frac{|x_i|}{\sqrt[4]{2}^{step}})^{3/4} - 0.0946 \rfloor \tag{2}$$

where x_i is the result after MDCT. q_i is quantitative result, and *step* is the quantization stepsize calculated by:

$$step = 8.0 * ln(k) \tag{3}$$

$$k = \frac{e^{\frac{1}{N}(\sum_{i=0}^{575} ln{x_i}^2)}}{\frac{1}{N}\sum_{i=0}^{575} {x_i}^2} \tag{4}$$

2.3 The Psychoacoustic in MP3 Encoding

The MP3 encoder has a high compression ratio and good sound quality, which is attributed to the use of psychoacoustic model. The psychoacoustic model simulates the human ear hearing which has masking effect. The model calculates the maximum quantization error of each subband with masking thresholds obtained by masking effect.

The encoder adopts psychoacoustic model II, the model II imitates the frequency of the human ear to 19 kHz. In the calculation process, psychoacoustic model II divide the frequency band in higher precision. For example, with the sampling rate of 44.1 kHz, the audio signal's frequency band is split into more regions to calculate the masking thresholds and then mapping regions to 21 scalefactor bands.

3 The Proposed Audio Watermark Scheme

3.1 Spread Spectrum Audio Watermark

In the spread spectrum audio watermark system, the watermarked signal is expressed as

$$\mathbf{s} = \mathbf{x} + b * p * \mathbf{u} \tag{5}$$

where $\mathbf{x}$ is the vector of host signal, $b = \{\pm 1\}$ is the watermark message, $\mathbf{u}$ is a PN sequence and p is the embedding strength calculated by psychoacoustic model II, the value of p is a trade-off between robustness and imperceptibility.

In the procedure of decoding, the signal can be expressed as

$$\mathbf{y} = \mathbf{s} + \mathbf{n} \tag{6}$$

where $\mathbf{n}$ represents the noise. According to the Eq. (6), the watermark can be extracted by calculating the correlation between the signal and PN sequence which is used in encoding process. The correlation can be calculated as follows

$$\begin{aligned} r &= <\mathbf{y}, \mathbf{u}> \\ &= <\mathbf{s} + \mathbf{n}, \mathbf{u}> \\ &= <\mathbf{x} + b * p * \mathbf{u} + \mathbf{n}, \mathbf{u}> \\ &= <\mathbf{x}, \mathbf{u}> + <\mathbf{n}, \mathbf{u}> + b * p \end{aligned} \tag{7}$$

In the proposed scheme, $\mathbf{x}$ is uncorrelated with PN sequence $\mathbf{u}$ and the result of $<\mathbf{n}, \mathbf{u}>$ is approximately equal to zero. Therefore, the watermark is decoded by $\text{sign}(r)$.

3.2 Watermark Embedding Algorithm

This paper aims at a watermarking algorithm that performs well focused on imperceptibility and robustness. To improve robustness, watermark should be embedded with maximum possible energy lean against the human auditory system. In the process of original MP3 quantizing, the watermark parameters just need to be less than the threshold calculated by psychoacoustic model II. That causes most of the watermark coefficients are too small compared with the masking threshold. While, the MP3 encoder's non-uniform quantizer uses different quantization intervals for the low and high frequency portions of the audio signal, this is benefit for the sound quality of the audio. Hence, this scheme learns from the MP3 encoder's quantizer, and modified the watermarking parameters in the encoding process.

The following details the embedding steps.

Step1. The original audio signals are processed by MP3 encoder and they are segmented into non-overlapping frames, with a frame size of 1152 samples. Then, the encoder generates the signal-masking-ratio (SMR) from the original signal through the psychoacoustic model II.
Step2. The audio signals are tranformed from time domain to frequency domain by MDCT.
Step3. Implement non-uniform quantization and inverse non-uniform quantization to adjust the stepsize.

The calculation formula are as follows:

$$q_i = \lfloor |x_i|^{3/4} * 2^{(k-210)*-0.1875} - 0.0946 \rfloor \tag{8}$$

where x_i represents a MDCT value in a frame, k is the stepsize, q_i is the outcome of non-uniform quantization.

$$p_i = (q_{i+1}^{4/3} - q_i^{4/3})/4/(2^{(k-210)*-0.1875} - 0.0946) \tag{9}$$

where p_i is the result of inverse non-uniform quantization, n is the num of spectral values in one scalefactor band.
Step4. Calculate the quantization noise (xr) as watermark parameters. Then, compare xr with SMR, back to step3 and adjust k until we find the most suitable xr, which is closest but does not exceed the SMR.

$$xr = \sum_{i=1}^{n} p_i * p_i \tag{10}$$

Step5. After regulating the stepsize, the encoder redoes quantization and inverse quantization. The calculation results of inverse quantization are watermarking parameters. Therefore, the embed equation is

$$s_i = |x_i| + p_i * u_i * b \tag{11}$$

where u_i is a PN sequence and $b = \{\pm 1\}$ represent the watermark message.
Step6. Apply inverse MDCT to restore audio signal.

3.3 Watermark Extraction Algorithm

In the extraction process, the algorithm is a blind watermark extraction algorithm which can extract watermark without original audio signal. For watermark extraction, the correlation coefficients can be used to recover watermark with the known PN sequence **u**. Due to the embedded strategy, the decoder calculates the correlation coefficients by substract the values in the first granule and the second granule.

The following illustrates the extraction steps in detail.

Step1. The MP3 decoder read the audio signals and divides the signals into frames which length is 1152. Then, the decoder finds the first frame that contains watermark.
Step2. The decoder does MDCT to transform audio signals from time domain to frequency domain.
Step3. The algorithm calculates the correlation coefficients.

$$\begin{aligned} r_i &= r_{i2} - r_{i1} \\ &= <x_{i2} - x_{i1}, \mathbf{u}> + <\Delta n, \mathbf{u}> - p * 2b \end{aligned} \tag{12}$$

where $\Delta n = n_{i2} - n_{i1}$ is a new random noise. Considering stationary signal in short time, where x_{i2} is approximately equal to x_{i1} and n_{i2} is approximately equal to n_{i1}, that means

$$<x_{i2} - x_{i1}, \mathbf{u}> \approx 0 \tag{13}$$

Step4. Sum r_i for a hard detection value.

$$\hat{r} = \sum r_i \tag{14}$$

$$b = sign(\hat{r}) \tag{15}$$

if $b > 0$, the watermark message is decoded as 1, if $b < 0$, the watermark message is decoded as -1.

4 Experimental Result and Evalutions

The sample audio files used in this experiment are sampled at 44.1kHz. The algorithm uses LAME which is a MP3 encoder and decoder with the version of 3.70 which open source codes are changed to add the proposed algorithm. The paper evaluates the performance in the following three aspects.

4.1 Imperceptibility Test

For imperceptibility evaluations, the objective difference grade (ODG) are universally adopted to measure the watermarked audio quality.

Table 1. Objective difference grade

ODG	Quality	Impairment
0	Excellent	Imperceptible
−1	Good	Perceptible but not annoying
−2	Fair	Slightly annoying
−3	Poor	Annoying
−4	Bad	Very annoying

The ODG is calculated by using perceived audio quality (PEAQ) which is an international standard, ITU-R BS.1387 [15]. Discription for ODG are listed in Table 1.

The paper uses different types of audio to verify the performance of the proposed algorithm, such as pop music, rock music and jazz music. The ODG values of the three types of audio are −0.736, −0.594 and −0.979, respectively. The results show that the proposed scheme has good imperceptible for different genres of audios.

4.2 Robustness Test

To analyze the proposed algorithm against robustness for different attacks such as additive white Gaussian noise (AWGN), the scheme used decoding accuracy. The experiment used SNR to measure embedding strength in order to compare robustness under the same conditions.

The SNR is defined as

$$SNR = 10 * log \frac{\sum_i {x_i}^2}{\sum_i (s_i - x_i)^2} \tag{16}$$

where x_i is the original audio signal and s_i is the watermarked audio signal.

To analysis the performance of the proposed scheme, the paper did another set of experiments using the method of Improved Spread Spectrum (ISS) [13]. ISS's core method is based on spread spectrum, it improved basic spread spectrum audio watermark by utilizing a linear function. In order to compare the performance of the two methods under the same conditions, the paper modified the payload of the ISS. Under the premise of similar SNR and same payload, the paper tested the performance of the two methods under the same AWGN attack. The SNR of the ISS is 10.3 and the proposed algorithm is 10.1. The experimental results presented in Fig. 1.

In the precondition of same SNR and same payload, the results show that the proposed algorithm has good capability to withstand the AWGN attack.

For a more comprehensive verification of the performance of the proposed method, the paper attemped five different types of attacks.

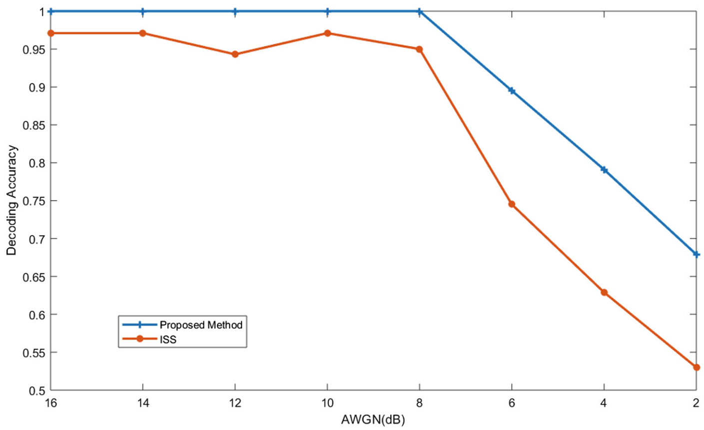

Fig. 1. Decoding accuracy of the two methods.

- Re-sampling: The original signal is sampled at 44.1 kHz, and then it is re-sampled at 22.05 kHz, and the audio is restored back by sampling again at 44.1 kHz.
- Echo addition: An echo signal is added to the watermarked audio signal.
- Amplify: The watermarked signal's amplitude is scaled up with factor of 0.5 and 1.5.
- Low-Pass Filtering: The signal is cutted off with frequency of 9 kHz.
- Re-quantization: The 16-bit watermarked audio signal is re-quantized down to 8 bits/sample and then back to 16 bits/sample.

Table 2. Decoding accuracy

Type of attack	Decoding accuracy
Re-sampling	100%
Echo	100%
Amplify-0.5	100%
Amplify-1.5	100%
Low-pass	100%
Re-quantization	100%

Table 2 are robustness results of different type of noise attack. As shown in the table, the presented scheme's ability to resist other type of attacks is excellent.

4.3 Payload Test

The payload is defined as the number of watermark bits embedded each second of the host signal. The audio's sample rate is 44.1 kHz, hence

$$Payload = \frac{44100}{1152} * 3 \approx 114(bps) \tag{17}$$

Audio Watermark in AM (AWAM) [14] is a blind audio watermark algorithm, it employed FFT and psychoacoustics model, and its innovation is every frame has a shift step to make the watermark more perceptually indistinguishable. The payload of reference is 43 bps. Figure 2 shows the performance of the proposed scheme and AWAM.

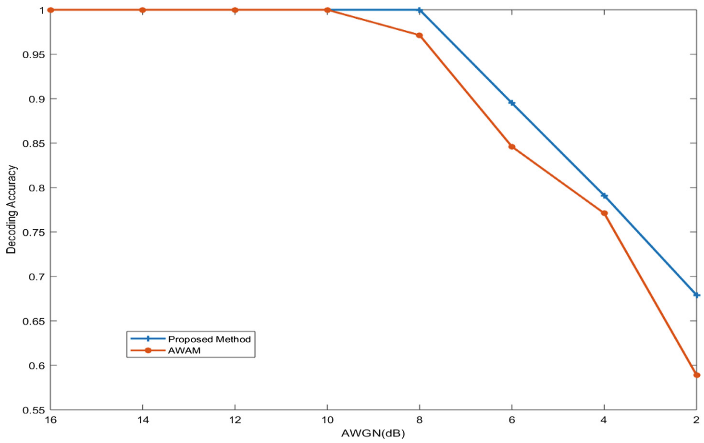

Fig. 2. Decoding accuracy of the two methods.

From the outcomes obtained by Fig. 2, the proposed method's decoding accuracy is only little better than AWAM, but the payload of the advanced method is bigger than AWAM.

The above conclusions show that the proposed algorithm has higher payload and better performance against various attacks compared to earlier audio watermarking schemes. The scheme balances the three conflict requirements including imperceptibility, robustness and payload.

5 Conclusions

This paper presented a spread spectrum audio watermarking method based on MDCT and non-uniform quantization. Watermark embedding is performed by taking advantage of the property of non-uniform quantization and the process of MP3 encoding. Watermark extraction is carried out by using the same PN sequence used for embedding process. The paper simulated AWGN as noise attack in communication and simulated other attacks as digital attack. The experimental results show that the proposed algorithm has excellent capability to withstand the noise attack durning communication.

References

1. Bamatraf, A., Ibrahim, R., Salleh, M.N.B.M.: Digital watermarking algorithm using LSB. In: 2010 International Conference on Computer Applications and Industrial Electronics, Kuala Lumpur, pp. 155–159 (2010). https://doi.org/10.1109/ICCAIE.2010.5735066
2. Wei, F.S., Qi, D.: Audio watermarking of stereo signals based on echo-hiding method. In: 2009 7th International Conference on Information, Communications and Signal Processing (ICICS), Macau, pp. 1–4 (2009). https://doi.org/10.1109/ICICS.2009.5397487
3. Chen, S., Hsu, C., Huang, H.: Wavelet-domain audio watermarking using optimal modification on low-frequency amplitude. IET Signal Process. **9**(2), 166–176 (2015). https://doi.org/10.1049/iet-spr.2013.0399
4. Li, R., Xu, S., Yang, H.: Spread spectrum audio watermarking based on perceptual characteristic aware extraction. IET Signal Process. **10**(3), 266–273 (2016). https://doi.org/10.1049/iet-spr.2014.0388
5. Erfani, Y., Pichevar, R., Rouat, J.: Audio watermarking using spikegram and a two-dictionary approach. IEEE Trans. Inf. Forensics Secur. **12**(4), 840–852 (2017). https://doi.org/10.1109/TIFS.2016.2636094
6. Kaur, A., Dutta, M.K.: A blind watermarking algorithm for audio signals in multi-resolution and singular value decomposition. In: 2018 4th International Conference on Computational Intelligence and Communication Technology (CICT), Ghaziabad, pp. 1–5 (2018). https://doi.org/10.1109/CIACT.2018.8480367
7. Garlapati, B.M., Chalamala, S.R.: A symbol based watermarking approach for spread spectrum audio watermarking methods. In: 2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS), Bangkok, pp. 180–184 (2016). https://doi.org/10.1109/ISMS.2016.31
8. Gupta, A.K., Agarwal, A., Singh, A., Vimal, D., Kumar, D.: Blind audio watermarking using adaptive quantization and lifting wavelet transform. In: 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, pp. 556–559 (2018). https://doi.org/10.1109/SPIN.2018.8474250
9. Aparna, S., Baiju, P.S.: Audio watermarking technique using modified discrete cosine transform. In: 2016 International Conference on Communication Systems and Networks (ComNet), Thiruvananthapuram, pp. 227–230 (2016). https://doi.org/10.1109/CSN.2016.7824019
10. Wang, K., Li, C., Tian, L.: Audio zero watermarking for MP3 based on low frequency energy. In: 2017 6th International Conference on Informatics, Electronics and Vision and 2017 7th International Symposium in Computational Medical and Health Technology (ICIEV-ISCMHT), Himeji, pp. 1–5 (2017). https://doi.org/10.1109/ICIEV.2017.8338599
11. Sakai, H., Iwaki, M.: Audio watermarking method based on phase-shifting having robustness against band-pass filtering attacks. 2018 IEEE 7th Global Conference on Consumer Electronics (GCCE), Nara, pp. 343–346 (2018). https://doi.org/10.1109/GCCE.2018.8574765

12. ISO/IEC 11172–3: Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s Part 3: Audio (1993)
13. Zhang, P., Xu, S., Yang, H.: Robust and transparent audio watermarking based on improved spread spectrum and psychoacoustic masking. In: 2012 IEEE International Conference on Information Science and Technology, Hubei, pp. 640–643 (2012). https://doi.org/10.1109/ICIST.2012.6221723
14. Li, M., Lei, Y., Zhang, X., Liu, J., Yan, Y.: Authentication and quality monitoring based on audio watermark for analog AM shortwave broadcasting. In: Third International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2007), Kaohsiung, pp. 263–266 (2007). https://doi.org/10.1109/IIHMSP.2007.4457701
15. Method for objective measurements of perceived audio quality (PEAQ): International Telecommunication Union. ITU-R. BS. 1387 (1998)