# MP3 AUDIO WATERMARKING

*Tianruo Sun, Lihao Yang, Zimo Cheng*
*University of Rochester*

## ABSTRACT

Audio watermarking is a technique for embedding additional data along with the audio signal so that it will not affect its perceptual quality of original audio signal. Four different watermarking algorithms in the temporal and spectral domain are studied in this paper.

***Index Terms***— Audio Watermarking, Audio Signal Processing, Echo Hiding, Least Significant Bit Coding, Phase Coding, Spread Spectrum

## 1. INTRODUCTION

As the whole world progresses to a digital age and especially the music industry enters a flourishing state, audio information becomes more valuable. Information age empowers informational technology to have its monetary value. Unprotected audio information exposed on the internet gives people the opportunity in piracy and embezzlement. The protection of audio information becomes a demand in the market. The audio watermarking technique has therefore become a favorable way in protecting the audio information due to its characteristic in not altering the audio sound in hearing yet still adding protections to the audio by integrating information.

Four mainstream methods, spread spectrum, least significant bit coding, phase coding and echo hiding are implemented to realize the audio watermarking technique. The three major requirements for audio watermarking technique are inaudibility, robustness, and capacity. For Inaudibility, the watermark embedding should not be accompanied by loss of audio quality. For robustness, the algorithm should be robust against various attacks for malicious users. [1] For the capacity, the efficient watermarking technique should be able to carry more information but should not degrade the quality of the audio signal. [2]

## 2. METHODS

### 2.1. Spread Spectrum

Different from the Echo Hiding, which is the method of time-spread, the Spread Spectrum technology is a method dealing with the signal in Fourier domain. Similar to echo hiding, we still need a sequence of PN (pseudo noise) code and a sequence of message.[3] Dislike a sequence of delta values in echo hiding method, we need a sequence of binary (-1/+1) codes as message. After the calculation of the exclusive or between these two codes, we multiply the result with a single frequency sinusoid wave. The frequency is the location where we will embed in the Fourier domain. We can pick a certain frequency or use auditory mask principle to find a proper frequency. After that, we have got the watermark signal and then add it to the original host audio signal.

In the decoding terminal, like the right part of the figure, the PN sequence is the key to extract the watermark as well. A band-pass-filter is used to filter out where the watermark embedded in. We need the multiplication result of PN sequence and the previous sinusoid wave. The auto correlation of the multiplication result and the filtered signal will generate a binary sequence. Compared the sequence with the original messages, we can judge whether the watermark is extracted correctly.

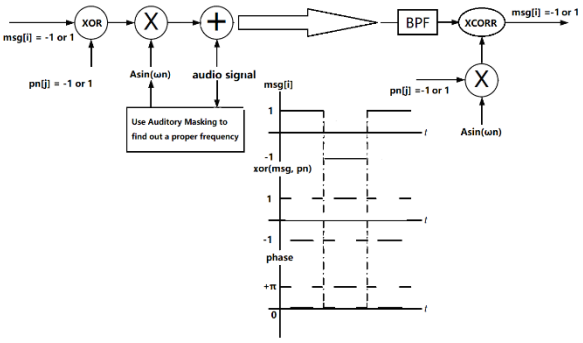


**Fig. 1.** Principle of Spread Spectrum Audio Watermarking

The method of spreading is used for embedding and the method of dispreading is used for extraction. The graph below shows all the steps of spread spectrum audio watermarking.

During the real process, we need to separate the audio signal into several frames. For each frame we embed one single message.

### 2.2. Least Significant Bit Coding

Least significant bit (LSB) algorithm belongs to time domain algorithms. We implement the encoding and extracting algorithm for the audio file format as waveform audio file format (WAV). We convert the information text to decimal

and convert decimal to binary data. For example, in order to hide the information letter "b" which has the ASCII code 98 and binary representation as 01100010 into the cover of 8 bytes, we can set the least significant bit pf each byte as:

xxxxxxx0
xxxxxxx1
xxxxxxx1
xxxxxxx0
xxxxxxx0
xxxxxxx0
xxxxxxx1
xxxxxxx0

x is the original byte values.

In terms of inaudibility, due to the tiny variant of original audio files, which only alter the value of least significant bit, we cannot hear the difference between embedded audios and original audios. In terms of capacity, we can hide one-byte information every eight bytes of the cover file. However, the LSB algorithm is not robust against signal processing such as linear filtering and resampling.

### 2.3. Phase Coding

In the frequency domain, the human auditory system is insensitive to small spectral phase changes. [4] Thus, we can use this advantage and encode the information in the cover files' phase information. Also, the inaudibility is robust due to the small phase changes. The encoding and extracting algorithm for the audio file format is waveform audio file format (WAV). We convert the information same as we did before in least significant bit algorithms to binary data and substitute the phase of segment with new phase embedded with information. The relative phase difference is preserved, and we change the new phase accordingly. The procedure is listed as followings. [5]

a. Break the original audio signal into short segments.
b. Apply fast Fourier transform and get the phase matrix.
c. Store the phase difference between adjacent segment.
d. Use $\varphi_{data} = \frac{\pi}{2}$ for binary data 0 and $\varphi_{data} = -\frac{\pi}{2}$ for binary data.
e. Use new phase and phase difference to get output phase matrix.
f. Use the output phase and original magnitude to recreate the audio signal containing the embedded information.

The capacity is as same as that of the least significant bit coding in which we can hide one-byte information every eight bytes of the cover file.

### 2.4. Echo Hiding

Echo hiding technique is using an artificial echo effect to hide our information into a period of audio signal. The information was represented by the time delay of the reflected sound. A multiple reflection sound was closer to the real world. Reflected sound in a real room includes many echoes heard like reverberation, which has a more natural sound quality than only one single echo. So, in the echo hiding technology, a time-spread echo method is based on this concept.[6]

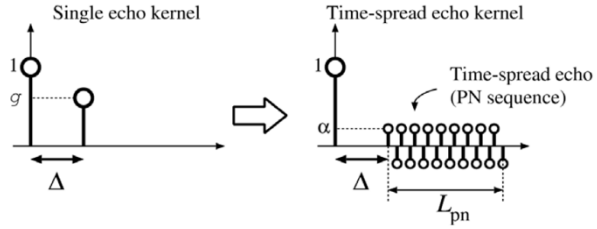


**Fig. 2.** Use the binary PN sequence to generate the time-spread kernel.

The multi-echo concept was simulated by a time-spread echo kernel which has a sequence of PN code with a delay of delta from the unit pulse, which presents the original audio signal. The kernel is constructed by:

$$k(n) = \delta(n) + \alpha p(n - \Delta), 0 < \alpha << 1$$

Where p(n) is the original PN code with the amplitude of 1. Because each reflected signal has a much lower power than the original signal, the real amplitude of the spread signal is α. So, the echo embedded signal was calculated from the linear convolution of the host signal and k(n):

$$w(n) = s(n) * k(n)$$

While the extraction of the watermarked signal, we need know the PN code first, since we see the PN sequence as the key to unveil the hidden information. During the extraction, we use the cepstrum analysis to separate the host signal, which is the original one, and the spreaded echo signals.

$$\widehat{w}(n) = \hat{s}(n) + \hat{k}(n)$$
$$\widehat{w}(n) = \hat{s}(n) + \alpha p(n - \Delta)$$

As we see the echo as the main signal loaded with the hidden information, the host signal can be seen as a kind of noise. The result of the cepstrum should go through an auto-correlation calculation with the PN sequence.

After the correlation, we can get the result:

$$d_c(n) = \widehat{w}(n) \otimes p(n) + \alpha p(n - \Delta) \otimes p(n)$$

As told in the previous paragraph, after the cepstrum the host signal, first part on the right of the equation, can be seen as a kind of noise and the result of the second part was presented as a high peak. The detection method is to find where the peak located, where the result of the delta delays.

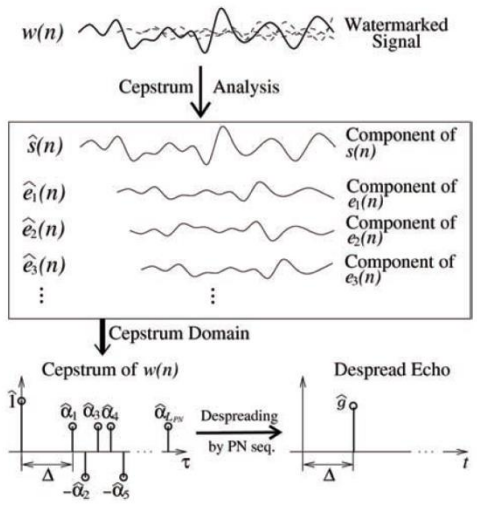


**Fig. 3.** Extract the despread echo signal from the watermarked signal.

### 2.5 Audio Watermark techniques for MP3 audio

MP3 audio watermarking technique has three major approaches in classifying the watermarking schemes. Except the most important focus in audibility, robustness, and capacity. The statistical invisibility which prevents the embedded watermarks from being removed, the similar compression characteristic with the original signal, the low redundancy, and methods to embed directly to the audio data are all crucial to the MP3 audio watermarking. [2]

The most commonly used technique which also is the method implemented in this paper, compressed domain, embedding watermarks to PCM-data operate with all audio formats (not sure to survive the coding/decoding procedure, time consuming). The other two methods are MP3Stego and Sanford et al and use or not of the original signal for watermark detection. MP3Stego hides information in MP3 files during the compression process (compress, encrypted, and then hidden in the MP3 bit stream). Sanford et al is a method embedding auxiliary information as a watermark into the host signal by a lossy compression technique.

By using the technology of echo hiding, we can cut the host signal into several segments. Each segment has its own delta delay, all the delays make a sequence of data, and all the frames in a single segment will have the same delta delay.

## 3. RESULTS

### 3.1. Least Significant Bit Coding and Phase Coding

After testing the algorithms on various audio files ranges from various instruments, songs, and speeches, we use bit error rate (BER) and short-time objective intelligibility (STOI) [7] to perform objective analysis. The bit error rate compares the unsigned binary representation of original text messages and decoded messages and return the error rate. We adopt STOI to measure the difference between original audios and encoded audios. The average values of BER and STOI scores before and after MP3 compression and decoding and other signal processing method are listed as followings. The test message is written in TXT file and the content is "We love asp\n Team". For the low pass filtering, we set the filtered frequency to 20 kHz, which corresponding to normal human maximum audio spectrum.

For least significant bit coding, the algorithm is not robust again signal processing tasks and not robust after MP3 codec. Compared with least significant bit coding, phase coding has relatively high robustness facing the audio processing tasks such as low pass filtering, and the bit error rate drops down to around 0.5 %, which is decent. However, after the MP3 compression and decoding, the information encoded for phase coding is still lost due to the phase changes. In terms of STOI, the scores are near one which means the audios with embedded messages are very close to original audios, and we cannot distinguish those two subjectively.

**Table 1.** Test Results for LSB and phase coding.

| | BER | STOI |
|---|---|---|
| LSB Encoded WAV | 0 % | 1 |
| LSB After low pass filter | 49.1 % | 0.989 |
| LSB After MP3 Codec | 51.8 % | 1 |
| Phase coding Encoded WAV | 0 % | 1 |
| Phase coding After low pass filter | 0.572 % | 0.988 |
| Phase coding After MP3 Codec | 46.8 % | 0.996 |

### 3.2. Echo Hiding

To judge the quality of the watermarking technology, we have two kinds of dimensions. First, to quantize the audio quality, we use STOI (Short-Time Objective Intelligibility) function to map the quality from 0 to 1. 0 means the worst while 1 means the best. Second, the error ratio of the information or the message is compared.

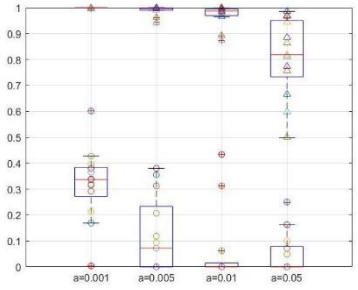


**Fig. 4. a.** Test Results of Echo Hiding before MP3 encoding.

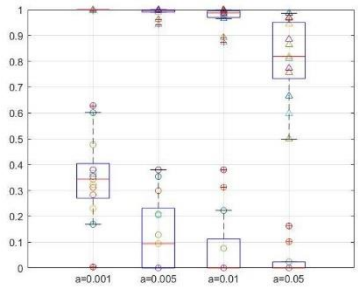


**Fig. 4. b.** Test Results of Echo Hiding after MP3 decoding.

In the Figure 4, there are two figures showing the variances which come from MP3 codec. Original PCM is the first one and MP3 decoded follows. The upper part with the triangle marks are the values of STOI and the lower results with the circle marks are the values of error ratio. We can see that as the growth of embedded amplitude, watermark became more audible so the sound quality is decreasing. While the error ratio is decreasing since the peak value of the correlation result is more and more distinct. After the comparison of before-codec and after-codec, the results do not vary too much. MP3 codec will not affect the sound quality of the embedded watermark signal but will increase the error rate a little bit. The medium value of a=0.01 is 0 even though the values of several cases will increase. After all, the amplitude around 0.01 can be a good choice in this watermarking technology.

### 3.3. Spread Spectrum

In the spread spectrum method, different embedding frequencies and different watermark amplitudes are considered. Similar to the figures of echo hiding, the figures below give different analysis results. We can see that the MP3 codec affect little on the audio quality and error ratio, since the figures are the same. Even though the results of the error ratios of lower amplitude vary a little.

In the center period of these two figures, an embedded signal with an amplitude of 0.005 has a result of ATOI very close to 1 and a result of error ratio very close to zero. So, the good characteristics among these periods can be used for future works.

### 3.4. Comparison of Echo Hiding and Spread Spectrum

As predicted before the analysis, we supposed that the spread spectrum watermark will not survive totally after MP3 codec, because of the compression. But the result shows a different conclusion.

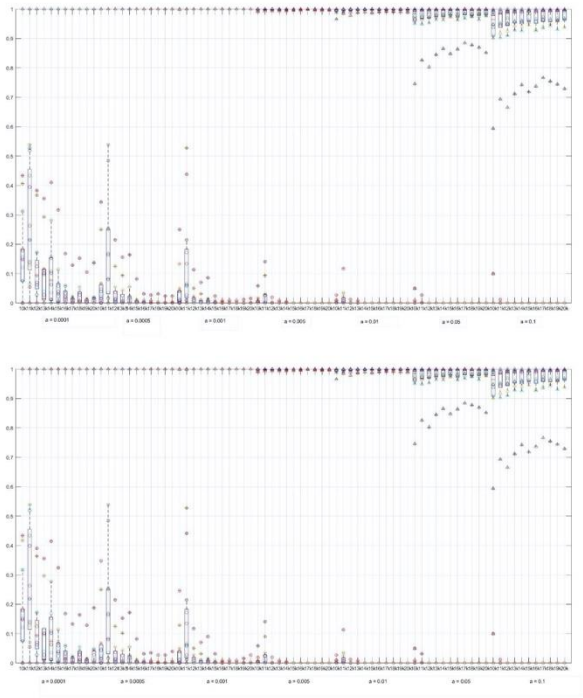

**Fig. 5.** Test Results of Spread Spectrum. The upper one is the results before MP3 encoding and the lower one is the results after MP3 decoding.

The embedded position is between 10kHz and 20kHz, which will not be killed by MP3 codec. These frequencies are audible referring to the principle of human audition. But in the real audio file, these high frequencies will be merged into the host signal and be inaudible to some extent. The spread spectrum watermark not only survives from the MP3 codec, but also has a better audio quality and lower error ratio. So, for the MP3 audio watermarking, the spread spectrum method is the better one compared to echo hiding.

## 4. CONCLUSION

For all these four watermarking methods, different use cases can pick different choices. The LSB and phase coding have decent performance in inaudibility and capacity whereas comparatively low in robustness thus resulting in information lost after MP3 codec. On the other hand, watermarking of echo hiding and spread spectrum can survive more after MP3 codec.

## 5. REFERENCE


[1] S. Katzenbeisser, and F. A. P. Petitcolas, Information Hiding Techniques for Steganography and Digital Watermarking, Artech House, Inc. 2000.

[2] Y. Stamatiou and D. Koukopoulos, "Digital Audio Watermarking Techniques for MP3 Audio Files," Digital Audio Watermarking Techniques and Technologies: Applications and Benchmarks, pp. 205-228, 2008.

[3] X. He and M. Scordilis, "Spread Spectrum for Digital Audio Watermarking," Digital Audio Watermarking Techniques and Technologies: Applications and Benchmarks, pp. 11-49, 2008.

[4] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, “A secure robust watermark for multimedia”, information Hiding Worshop, Univ. of Cambridge, pp.185–206, 1996.

[5] W. Bender, D. Gruhl, N. Morimoto and A. Lu, "Techniques for data hiding," in IBM Systems Journal, vol. 35, no. 3.4, pp. 313-336, 1996.

[6] Y. Suzuki, R. Nishimura and B. Ko, "Advanced Audio Watermarking Based on Echo Hiding: Time-Spread Echo Hiding," Digital Audio Watermarking Techniques and Technologies: Applications and Benchmarks, pp. 123-151, 2008.

[7] C. H. Taal, R. C. Hendriks, R. Heusdens and J. Jensen, "An Algorithm for Intelligibility Prediction of Time–Frequency Weighted Noisy Speech," in IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 7, pp. 2125-2136, Sept. 2011.