Robust Audio Watermarking Algorithm Based on Air Channel Characteristics

Wen Diao, Yuanxin Wu, Weiming Zhang, Bin Liu, Nenghai Yu
CAS Key Laboratory of Electromagnetic Space Information
University of Science and Technology of China, Hefei, China
Email: dw15@mail.ustc.edu.cn, wyx_ustc@163.com
zhangwm@ustc.edu.cn, flowice@ustc.edu.cn ynh@ustc.edu.cn

***Abstract*—This paper proposed an algorithm to improve the robustness of audio watermark for the air channel propagation. Firstly, according to the characteristics of the air channel, noise energy is mainly concentrated in low frequency region. The frequency domain to embed watermark is selected. By modifying the FFT-transformed intermdiate frequency coefficients, the watermark is embedded so that the watermark can be robustly extracted while resisting channel interference. The experimental results show that the algorithm has good imperceptibility and can resist 10dB noise interference and resampling attacks. The embedding rate can reach 20bps while maintaining the accuracy rate of more than 92% within 1.5 meters. It can be well adapted to air channel propagation scenario, the performance of the algorithm is greatly improved.**

***Index Terms*—audio watermarking, FFT, synchronization signal, air channel**

I. Introduction

With the rapid development of information technology, the problem, how to protect digital multimedia data against unauthorized copying, is becoming a rising concern. Digital watermarking technology is proposed in this situation where digital works have encountered in the protection of copyright. When digital watermarking technology was proposed, it gained a lot of attention in the industry field immediately. As the main technical means of protecting information security and covert communications, digital watermarking is being widely studied and applied.

The main applications of audio digital watermarking are copyright protection, license control and content authentication. The major performance indicators include: robustness, imperceptibility and playload. Actually, three main indicators are mutually conflictiing. Therefore, when designing and developing a new watermarking scheme, it is necessary to seek a balance among the three indicators.

At present, audio piracy products in the field of air channel communication are mainly generated from the recording secretly in public places such as cinemas and concert halls. This makes air channel propagation audio watermarks of great research value. However, in the air propagation, in addition to the complicated DA/AD conversion, the audio watermark will also be affected by the air noise, the volume, distance of the playback source and the orientation of the recording equipment, etc. In addition, audio works may also be subject to re-sampling attacks. Attacks such as format conversion will cause great technical difficulty for improving the performance of watermarking algorithms.

Previous audio watermarking techniques can be classified into two categories: time domain algorithms [1], [2], [3] and transform domain algorithms [4], [5], [6]. In the time domain algorithm, there are LSB methods [7], [8], echo hiding [9], and so on. The transform domain algorithm hides the watermark by modifying the frequency domain coefficients of the audio signal, such as fast Fourier transform (FFT) [10], discrete cosine transform (DCT) [11] and discrect wavelet transform (DWT) [12]. The most prominent point is that its robustness is significantly enhanced.

Currently, the study of digital audio watermarks for cable channels or for Ethernet transmission is relatively mature. Since there are multiple attacks on air channel transmission process, the research on audio watermark via air-channel is far less than other audio watermark algorithms.

The earliest studies of Steinbach [13], they studied 5 types of audio watermarking technology by using 4 different microphones, but the experimental result is not practical enough. In terms of straight-through cable transmission, Xiang Shijun et al. [14] used a three-stage energy ratio method to embed a string of 32-bit information. Although the performance is relatively good, the capacity is too small, the usability is not good enough and the synchronization technology has higher requirements. Zhang Xia et al. [15] proposed to modify the coefficients in the double DCT domain. Modifying it by selecting a part of the first DCT to perform a second DCT, then embedding the watermark. Although the algorithm can be well concealed, in the noisy environment, bit error rate will be sharp increased.

In the recent work, Andrew Nadeau and Gaurav Sharma [16] proposed an efficient and robust algorithm for resynchronization after analog playback, but it doesn't work in airbone situation. In addition, Michael Arnold et al. [17] used watermark embedding in phase modulation in the WOLA domain, but the embedding rate was low. Qian Wang et al. [18] used the 17KHz and 19KHz frequencies as carriers, which could not be perceived by the human ear, to embed information based on OFDM, and was finally acquired by the smartphone. However, in the actual scene, the high-frequency signal will be erased by filters, which is not suitable for the scene of copyright protection.

978-1-5386-4210-8/18/$31.00 ©2018 IEEE
DOI 10.1109/DSC.2018.00048

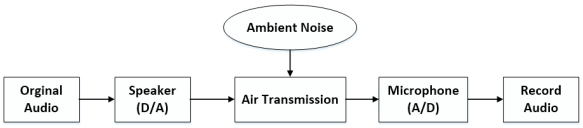


Fig. 1: Audio siganl's acoustic transmitting model

It can be seen that how to improve the robustness and concealment effect in public audio broadcasting is still a difficult problem to be solved. In this paper, we propose an audio watermarking algorithm which can effectively resist multiple attacks based on channel characteristics.

The organization of this paper is as follows. The analysis of air channel is presented in Section II. The proposed algorithm is elaborated in Section III. The experimental results are presented in Section IV before concluding this paper in Section V.

II. The Characteristics of Acoustic Speacker-Microphone Channel

In real life scenarios, most of the air channel propagation comes from the speaker playing audio signals and microphone reception. In order to represent the effect of channel characteristics on audio watermarking better and more specifically. We chosed the speaker-microphone channel as a model, and used the experimental framework shown in Fig.1 to explore the characteristics of the channel. illustrating the impact on the watermarked signal.

A. Impact of DA/AD Conversion Process

Impact of DA/AD conversion process includes noise in the process, linear scaling and waveform distortion on the time axis. The strategy adopted to slove these problems is to add the positioning search synchronization signal to the watermark and embed in the frequency domain.

B. Impact of Air-transmission Process

The environmental noise in the public places will cause great interference to the audio signal, resulting in a significant increase of error rate during the extracting process. To slove this problem we measure the noise intensity in different environments to characterize these disturbances.

The experimental environment is as follows: We used a Vivo Y19t smartphone to record noise in different environments. The ambient noise energy distribution were measured in square, conference hall and office, as shown in Fig.2. It can be seen that when the frequency is lower than 5Khz, the environmental noise is relatively large. The energy is mainly concentrated in the low-frequency segment. Above 8Khz, the noise is almost negligible.

It gives us an idea of watermark embedding, embedding in the frequency band where the ambient noise is the least disturbing.

At the same time, we play three common types of audio and recorded. As Fig.3 shows, analyzing their specturm, it

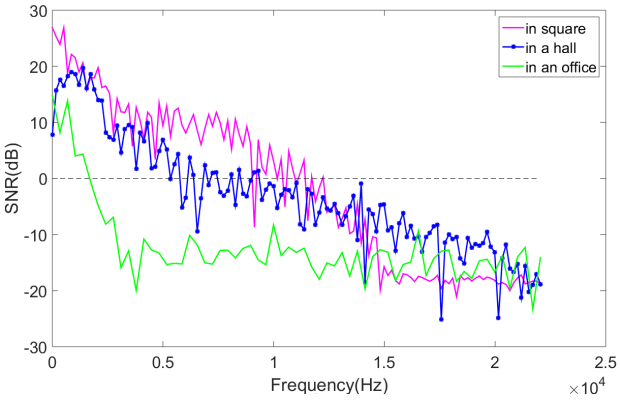


Fig. 2: Spectrum of Ambient Noise

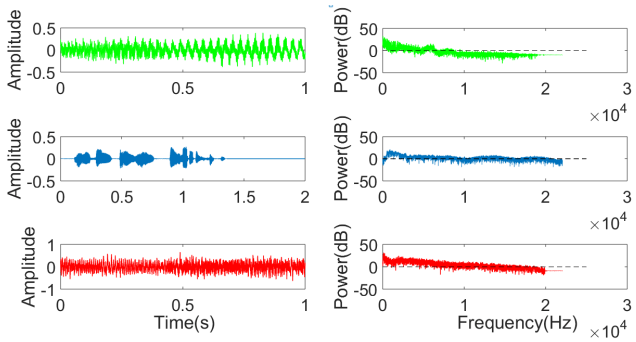


Fig. 3: The waveform and frequency specturm of pop music,human voice and class music

can be seen that the energy of the audio siganl is mainly concentrated in the low and middle frequency. In the process of transmission, the higher frequency is, the more energy decay.

C. Other impacts

The pre-mute segent which may occur during recording or in original audio.

III. Proposed Audio Watermarking Algorithm

Because of the linear stretching, waveform distortion and noise interference and the presence of slient segements, we adopt corresponding countermeasures in the embedding and extraction algorithms respectively:

1) Audio preprocessing: As shown in Fig.4, using double-threshold method [19] to find a suitable processing start point.

2) Adding synchronization code in the watermark embedding process, which is used to resist the linear expansion in the channel propagation process, and can eliminate the linear expansion and contraction in the extraction process.

3) To resist the waveform distortion and noise interference in the propagation process, we perform FFT transform on the signal and modify the coefficients in the mid-band of the FFT domain to achieve the purpose of embedding the watermark.

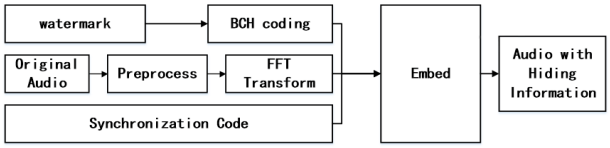


Fig. 4: The framework of embedding algorithm

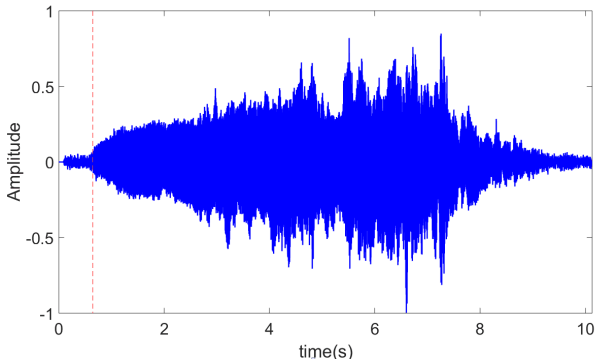


Fig. 5: Using endpoint detection to find the starting point

A. Audio Preprocessing

In the audio signal, there will be a period of slience or very small amplitude. At the same time, pre-mute may still occur during recording, synchronization marks and watermark embedding effects are generally performed badly in this area. For example, the situation shown in the Fig.5.

There is a situation frequently encountered in this area, such that the identification bit is not found and the decoding error is found. In order to reduce the occurrence of this kind of situation, it is necessary to pretreat the audio signal firstly. Using the endpoint detection technology to detect the right location for follow-up operations.

To improve the quality of the embedding, double threshold detection is used to determine the starting point Start_p.

B. Watermark embedding algorithm

According to previous experiments and analysis, it can be judged that modifying the frequency band in FFT can effectively resist noise attacks and other universal attacks.

Orignal audio signal is divided into segments. Every segment includes a synchronization frame and several watermark frames. First, the watermark message is BCH-encoded [20]. The encoded watermark message is marked as $w(i)$. Then we perform FFT transformation on each watermark frame, and modify the coefficients to embed watermark message.

The FFT transform equation is shown as (1), the Inverse FFT transform is shown as (2).

$$f\left(k\right)=\sum_{n=0}^{N-1}x(n)\cdot e^{-\frac{2\pi i}{N}\cdot kn}\ ,\ k=0,...,N-1 \qquad (1)$$

$$x\left(n\right)=\frac{1}{N}\sum_{k=0}^{N-1}f(k)\cdot e^{-\frac{2\pi i}{N}\cdot kn}\ ,\ n=0,...,N-1 \qquad (2)$$

In this paper, We define the frequencies from 8K to 10K as intermediate frequency. Selecting a frequency in intermadiate frequency randomly as the centre point, marke as f_1. We take L points around f_1, mark as $f(0)$, $f(1)$, $f(2)...f(L-1)$ and divide it into two segments. Each segment is L/2 long. The energy of the two-stage FFT coefficients is E_1, E_2 is defined as follows.

$$E_1=\sum_{i=0}^{L/2-1}|f(i)| \qquad (3)$$

$$E_2=\sum_{i=L/2}^{L-1}|f(i)| \qquad (4)$$

Simultaneously, setting the embedded strength S, $S=\alpha\cdot max(E_1,E_2)$, where α is the strength factor. To resist interference during the air transmission, the α value should be as large as possible under the constraint of imperceptibility. The parameter α is assigned as a predefined value at the beginning, and then automatically adjusted until the objective quality grade (ODG) value of watermarked audio is satisfied. In the proposed strategy, one watermark bit $w(i)$ can be embedded by modifying the relationships among E_1, E_2 and S, as shown in (5):

$$\begin{cases} E_1-E_2\ \geq S,\ if\ w(i)=1 \\ E_2-E_1\ > S,\ if\ w(i)=0 \end{cases} \qquad (5)$$

Basic steps involved in the watermarking embedding are given as follows. We give an example as watermark $w(i)$=1 to illustrate.

Step 1: If $E_1-E_2\geq S$, coefficients do not adjust, $f(i)'$ =$f(i)$. If not, do step 2.

Step 2: To increase E_1 and decrease E_2, the specific measure is

$$f(i)'=f(i)\ +\ \delta\ ,\quad i=0,1,2,...,L/2-1 \qquad (6)$$

and

$$f(i)'=f(i)\ -\ \delta\ ,\quad i=L/2,L/2+1,...,L \qquad (7)$$

δ is a small non-negative value, $f(i)'$ is a modified value.

Step 3: Case 1, $E_1-E_2\geq S$, stop. Case 2, $E_1-E_2<S$, return step 2.

Original FFT coefficients are shown in Fig.6, watermarking result is shown in Fig.7.

Finally, because the coefficients of the FFT transform are conjugate symmetric, the same operation should be performed at the symmetry point of the FFT transform domain, then IFFT is performed to obtain the time domain information of the embedded watermark information.

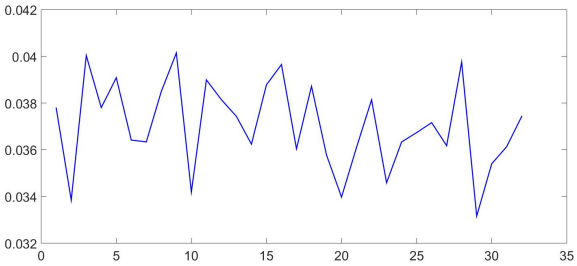
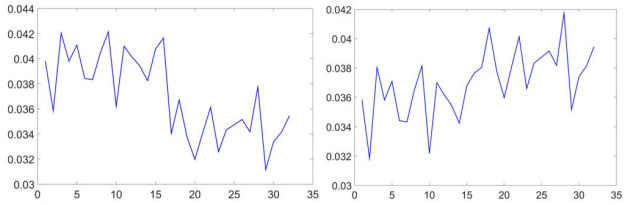


Fig. 6: Coefficients of watermark frame in FFT domain



(a) (b)

Fig. 7: Watermarking '1'(a) and Watermarking '0'(b) coefficients in FFT domain

C. Synchronization Code Embedding

According to our previous test results, embedding synchronization signals in the midband of the audio can be effective against air interference and background noise in the surrounding environment. We use a hidden information structure of the synchronization signal and the watermark signal loop body, its structure is shown in Fig.8.

The length of the synchronization signal is N_1, and the length of the watermark signal is N_2=15 $\times$ N_1. We perform a FFT on the synchronization signal frame of N_1 to calculate the absolute value as its energy. Select another frequency in intermadiate frequency f_2 different from f_1. Mark the maximum energy as Max_f and take the 8 points $c(i)$, 0 $\leq i \leq$ 7, on both sides of f_2 position where the maximum value marked as Max_c, changing it as (8) shown:

$$c(i)' = \beta \cdot \frac{c(i)}{Max\_c} \cdot Max\_f \tag{8}$$

Among them, β is the modification factor, which can be modified according to the audio quality requirements. Here we set it to be 0.5. Same as the embedding process, we take 8 points in the symmetry positions and do the same embedding operation. The orignal signal's sepctrogram is shown in Fig.9. The embedded spectrum is shown as Fig.10. Finally, the frequency domain signal is IFFT transformed to obtain the embedded time domain information.

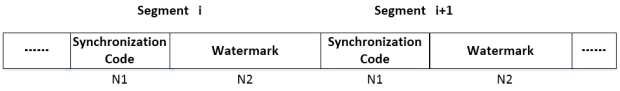


Fig. 8: Construction of embedding information

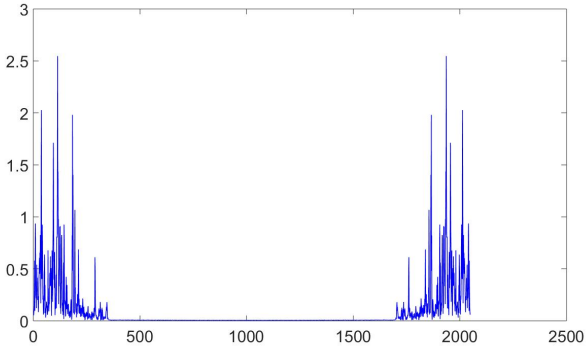


Fig. 9: FFT Spectrogram of Original signal

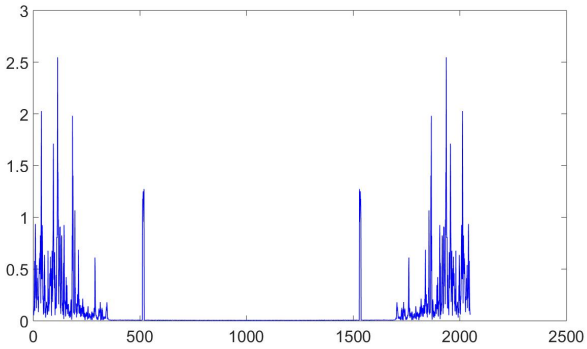


Fig. 10: FFT Spectrogram of embeded syn signal

D. Watermarking extracting algorithm

The watermark extraction algorithm is as Fig.11 shown. Firstly, the audio signal is preprocessed and then synchronously decoded. The location of the watermark message is determined by the location of the synchronization code. Then the FFT transform is performed, the FFT coefficients are compared to extract the message, and finally the BCH decoding is performed to obtain the watermark message. The specific process is as Fig.12.

1) Synchronization decoding: First, we perform synchronous decoding. The sliding window structure shown in Fig.12 is used for decoding. Setting the threshold t_value. We put N_1 points as a frame, then do FFT tranform and take the absolute value. The maximum value is recorded as Max_f_0.

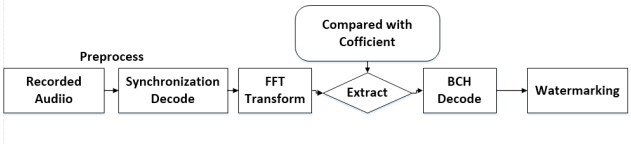


Fig. 11: Framework of extraction algorithm

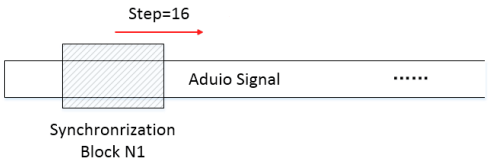


Fig. 12: Framework of resynchronization algorithm

We take the 8 points on both sides of f_2, mark their values as $c(i)$. i= 0,1,...7. Calculating the sum of 8 points energy mark as sum_x, if

$$sum\_x\ /Max\_f_0 \geqslant t\_value \tag{9}$$

one synchronization information in the signal is found. Recording its corresponding position. Because we embeded synchroniaztion signal in the way of circulation, we skip the following section of watermarking information, and synchronize again. If

$$sum\_x\ /Max\_f_0 < t\_value \tag{10}$$

the window is slid for 16 samples and synchronization is performed again until the end of the signal.

2) Watermark detection: After finding the synchronization signal in 1) and determining the embedding position, it is processed according to the method of embedding, each watermark frame is subjected to FFT transformation. The L points around the intermediate frequency f_1 are taken and divided into two segments to compare the energy levels.

$$E_1' = \sum_{i=0}^{L/2-1} |f(i)| \tag{11}$$

$$E_2' = \sum_{i=L/2}^{L-1} |f(i)| \tag{12}$$

and

$$\begin{cases} E_1' \geq E_2' \ , \ w'(i) = 1; \\ E_1' \ < E_2' \ , \ w'(i) = 0; \end{cases} \tag{13}$$

In this time, we get a hidden squence $w'(i)$. After the BCH error correction in the decoding of squence, we get the original watermark.

E. Performance evaluation

In this part, we will discuss the algorithm's embedding rate and signal-to-noise ratio(SNR), and demonstrate that the algorithm is feasible.

1) Evaluation of Embedding capacity:

$$B = \frac{fs}{N_1 + N_2} \cdot \frac{N_2}{N_1} \qquad (bps) \tag{14}$$

In this equation, fs represents signal sampling rates. N_1 = 2048 represents the length of synchronization frame. N_2 represents the length of embedding frame. Here, sampling rate is 44.1Khz and N_2 equal to 15 times of N_1, the embedding capacity is 20.18bps.

2) Evaluation of SNR:

$$\begin{aligned} SNR &= -10lg\left(\frac{\|F-F'\|^2}{\|F\|^2}\right) \\ &= -10lg\left(\frac{\sum_{i=0}^{N-1}(f(i)^2)-\sum_{i=0}^{N-1}(f(i)'^2)}{\sum_{i=0}^{N-1}(f(i)^2)}\right) \end{aligned} \tag{15}$$

In (15), F is orignal signal, and F' is watermarked audio siganl. $f(i)$ is orignal FFT coefficient and $f(i)'$ is modified coefficient in FFT domain. Experment results indicate the average SNR is greater than 20dB, which is satisfied with the Sound of International Union's requirement.

IV. Test and Analysis Performance

In order to verify the robustness of the algorithm and the applicability in the real scene, we did a lot of experiments. The results and details are as follows.

A. The impact of distence

In the air channel, distance plays an important role as a indicator. Ambient noise is also a very important indictor. In order to test the effect of distance and nosie to the robustness of algorithm, we have selected four typical audio frequencies. A bit error rate (BER) test was performed at distances of 0.2m, 0.5m, 1m, and 1.5m.

As Fig.13 shown, the experiment now indicates that the algorithm is robust enough and the correct rate is effectively improved at relatively close distances.

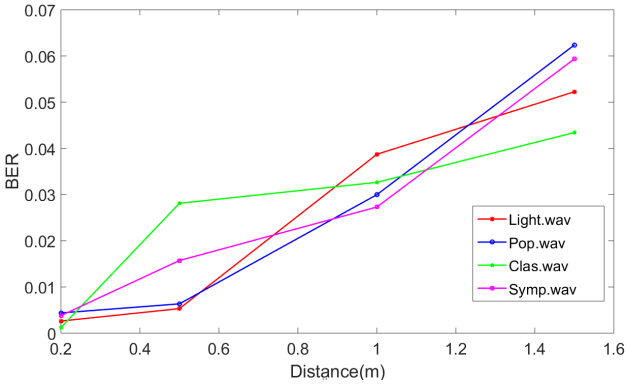


Fig. 13: The impact of distance between speaker and microphone

B. The impact of noise

Propagating in the air is unavoidably interfered by noise. In order to detect the anti-noise ability of the algorithm, we have done a bit error rate test at different SNRs and compared it with algorithms [14] and [15]. The specific results are shown in Fig.14.

The test result shows that the algorithm has strong robustness and can resist the interference of different SNR noises, even 10dB. Compared with the previous algorithms, the performance of the algorithm is greatly improved, especially at low SNR.

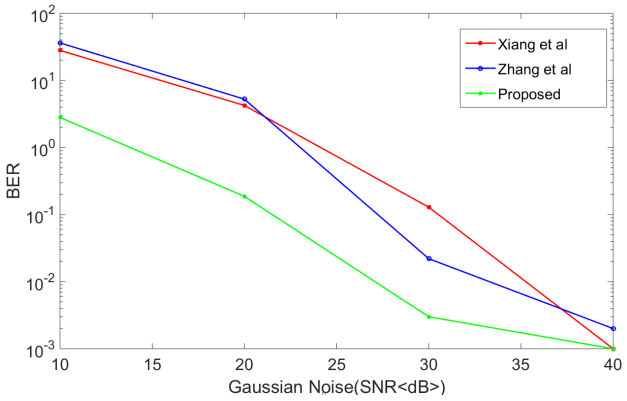


Fig. 14: The impact of noise

TABLE I: Robustness to attacks based on Stirmark Benchmark for audio

Attack Type	Error Rate
Addbrum_1100	0.02%
Addbrum_8100	2.14%
Addnoise_100	0.02%
Addnoise_500	0.05%
Invert	0.02%
Amply_50	1.28%
Compressor	2.51%
Resampling 22.05	1.38%
Resampling 11.025	2.79%
RC_Lowpass	0.02%

C. Test of robustness against the attack

In order to test the robustness of the algorithm, we also performed experiments on common signal processing attacks and attacks based on StirMark benchmark for audio.

As shown in table I, the algorithm is robust to resist the attacks and it has low bit error rate.

V. Conclusion

Aiming at the problems that the audio watermarking signal transmited in air channel, this paper proposed a robust audio watermarking algorithm based on channel characteristics. It mainly embeds watermarks in certain frequency bands of audio signal, minimizing the interference of environmental noise, and using the endpoint detection technology in processing to improve the quality.

The experimental results show that the algorithm is robust to air channel propagation, it can resist 10dB noise attack and maintain error rate below 8% within a distance of 1.5m. The algorithm can be applied to practical scenarios.

Acknowledgment

This work supported in part by the Natural Science Foundation of China under Grant U1636201 and 61572452.

References

[1] P. Bassia, I. Pitas, and N. Nikolaidis, "Robust audio watermarking in the time domain," *IEEE Transactions on multimedia*, vol. 3, no. 2, pp. 232–241, 2001.

[2] A. N. Lemma, J. Aprea, W. Oomen, and L. van de Kerkhof, "A temporal domain audio watermarking technique," *IEEE transactions on signal processing*, vol. 51, no. 4, pp. 1088–1097, 2003.

[3] A. Nishimura, "Audio watermarking based on subband amplitude modulation," *Acoustical Science and Technology*, vol. 31, no. 5, pp. 328–336, 2010.

[4] W. Li, X. Xue, and P. Lu, "Localized audio watermarking technique robust against time-scale modification," *IEEE transactions on multimedia*, vol. 8, no. 1, pp. 60–69, 2006.

[5] X.-Y. Wang and H. Zhao, "A novel synchronization invariant audio watermarking scheme based on dwt and dct," *IEEE Transactions on signal processing*, vol. 54, no. 12, pp. 4835–4840, 2006.

[6] Y. Xiang, I. Natgunanathan, Y. Rong, and S. Guo, "Spread spectrum-based high embedding capacity watermarking method for audio signals," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2228–2237, 2015.

[7] N. Cvejic and T. Seppanen, "Increasing robustness of lsb audio steganography using a novel embedding method," in *Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. International Conference on*, vol. 2, pp. 533–537, IEEE, 2004.

[8] N. Cvejic and T. Seppanen, "Increasing the capacity of lsb-based audio steganography," in *Multimedia Signal Processing, 2002 IEEE Workshop on*, pp. 336–338, IEEE, 2002.

[9] H. O. Oh, J. W. Seok, J. W. Hong, and D. H. Youn, "New echo embedding technique for robust and imperceptible audio watermarking," in *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, vol. 3, pp. 1341–1344, IEEE, 2001.

[10] D. Megías, J. Serra-Ruiz, and M. Fallahpour, "Efficient self-synchronised blind audio watermarking system based on time domain and fft amplitude modification," *Signal Processing*, vol. 90, no. 12, pp. 3078–3092, 2010.

[11] Y. Yan, H. Rong, and X. Mintao, "A novel audio watermarking algorithm for copyright protection based on dct domain," in *Electronic Commerce and Security, 2009. ISECS'09. Second International Symposium on*, vol. 1, pp. 184–188, IEEE, 2009.

[12] M. Jiansheng, L. Sukang, and T. Xiaomei, "A digital watermarking algorithm based on dct and dwt," in *International symposium on web information systems and applications*, pp. 104–107, Citeseer, 2009.

[13] M. Steinebach, A. Lang, J. Dittmann, and C. Neubauer, "Audio watermarking quality evaluation: robustness to da/ad processes," in *Information Technology: Coding and Computing, 2002. Proceedings. International Conference on*, pp. 100–103, IEEE, 2002.

[14] S. Xiang, "Audio watermarking robust against d/a and a/d conversions," *EURASIP Journal on Advances in Signal Processing*, vol. 2011, no. 1, p. 3, 2011.

[15] X. Zhang, D. Chang, W. Yang, Q. Huang, W. Guo, and Y. Zhao, "An audio digital watermarking algorithm transmitted via air channel in double dct domain," in *Multimedia Technology (ICMT), 2011 International Conference on*, pp. 2926–2930, IEEE, 2011.

[16] A. Nadeau and G. Sharma, "An audio watermark designed for efficient and robust resynchronization after analog playback," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 6, pp. 1393–1405, 2017.

[17] M. Arnold, X.-M. Chen, P. Baum, U. Gries, and G. Doerr, "A phase-based audio watermarking system robust to acoustic path propagation," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 3, pp. 411–425, 2014.

[18] Q. Wang, K. Ren, M. Zhou, T. Lei, D. Koutsonikolas, and L. Su, "Messages behind the sound: real-time hidden acoustic signal capture with smartphones," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*, pp. 29–41, ACM, 2016.

[19] R.-z. Zhang and H.-j. Cui, "Speech endpoint detection algorithm analyses based on short-term energy [j]," *Audio Engineering*, vol. 7, p. 015, 2005.

[20] C. P. Baggen, L. B. Vries, *et al.*, "Method and apparatus for decoding code words protected wordwise by a non-binary bch code from one or more symbol errors," Mar. 22 1994. US Patent 5,297,153.