Adaptive Audio Watermarking Algorithm Based on Wavelet Transform

Hanyu Lu[1], Mingjing Zhong[1], Haixiang Lan*[2], Ruchong Hou[1] and Zhengjiang Hu[1]

[1] College of Big Data and Iinformation engineering, GuiZhou University, Guiyang, GuiZhou.
[2] Guizhou Lichuang Science and Technology Development Co., Ltd. Guiyang, GuiZhou.
luhanyu@163.com

Abstract. Aiming at the problem of small capacity and poor robustness of audio watermarking information, an adaptive digital audio blind watermarking algorithm based on wavelet transform is proposed. In this algorithm, each frame of the audio signal is converted into a discrete transform by using Haar wavelet, and the fine and approximate components are obtained, and the component masking threshold is calculated. Finally, the watermark embedding position is selected by the adaptive algorithm, quantized and modulated in the audio. The intermediate and low frequency coefficients of the signal. The simulation results show that the proposed algorithm has some resistance to common signal processing attacks and is robust to resampling and low-pass filtering attacks.

1. Introduction
In recent years, the rapid development of digital multimedia technology and Internet technology has made image, video and audio transmission convenient and widely used [1-10]. Digital watermarking technology is a specific process meaning and easy to extract information embedded in the original audio signal. In this paper, an adaptive digital watermarking algorithm based on wavelet transform is proposed, which can realize blind detection. The proposed algorithm can adaptively select the embedded noise location with small noise sensitivity. The audio watermark is embedded in specific information and easy to extract. The original audio quality is not affected during the embedding process.

2. Audio Watermarking Algorithm
There are usually two algorithms for implementing audio watermarking: time domain watermarking algorithm and transform domain watermarking algorithm. Time domain audio watermarks are easier to implement than transform domain watermarks but have lower computational complexity and are less able to withstand attack power.

2.1. Masking Threshold
There is a masking effect on the human ear. When there are some energy differences between two or more excitation sounds at the same time, If the watermark strength is limited below the masking threshold, then the unperceivable watermark is reflected. The frequency resolution of the human ear is nonlinear [4].

IOP Conf. Series: Materials Science and Engineering **719** (2020) 012033 doi:10.1088/1757-899X/719/1/012033

2.2. Audio Watermarking Method Flow Based on Wavelet Transform

Thus supporting different frequency ranges. Wavelet transform is a very suitable tool for constantly changing the frequency of an audio signal. The audio watermark embedding uses a wavelet transform-based method. The flow is: the original audio signal is converted into audio pre-processing, then the DWT is performed, then the watermark embedding is performed, and the digital watermark is pre-processed in this link, then IDWT, and finally the watermark audio is embedded. .
The wavelet transform decomposes the signal into a set of orthogonal bases by translation and pressure expansion, if the Fourier transform of and is equal to equation (1):

$$\psi(\omega) = \int_{-\infty}^{+\infty} \psi(t) e^{-j\omega t} dt \tag{1}$$

So let wavelet or mother wavelet of $\varphi(t)$, transform or stretch the mother wavelet to get formula (2).

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi_{a,b}(\frac{t-b}{a}), (a, b \in R, a \neq 0) \tag{2}$$

For any function $f(t) \in L^2(R)$,the continuous wavelet transform is defined as:

$$W_f(a,b) \leq f(t), \psi_{a,b}(t) \geq \int_{-\infty}^{+\infty} f(t) \overline{\psi_{a,b}(t)} dt \tag{3}$$

The inverse transformation is defined as:

$$f(t) = \frac{1}{c_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{|a|^2} W_f(a,b) \psi_{a,b}(t) dbda \tag{4}$$

3. Audio Watermark Embedding and Extraction Process

In general, digital audio watermarks should have the following basic characteristics: invisibility, robustness, security, certainty, and recovery. The watermarking system is usually implemented in three steps: the embedding and extraction process of the watermark, and the detection of the watermark.

3.1. Adaptive Embedding Watermarking Step

Since the audio signal is a non-stationary signal, not all audio segments are suitable for embedding watermark information. Therefore, this paper conceives an adaptive selection scheme that can improve the performance of watermarks. After the discrete wavelet transform is performed, the approximate component is extracted[6]. Since most of our audio signal energy is medium and low frequency coefficients[7], combined with the robustness of the algorithm, we concentrate the watermark information on the intermediate frequency and low frequency coefficients.

1)The audio signal is segmented into frames, segments, and the original audio signal is segmented into frames containing the same sample points. Each frame contains the same sample point segment.

2)Performing a discrete wavelet transform on each frame of the audio signal $A_1^i A_2^i$ to obtain a fine component $A_{1d}^i A_{2d}^i$ and an approximate component $A_{1a}^i A_{2a}^i$:

$$(A_{1d}^i, A_{1a}^i) = DWT(A_1^i) \tag{5}$$

$$(A_{2d}^i, A_{2a}^i) = DWT(A_2^i) \tag{6}$$

For the approximation component, we don't need to process it, and the embedding watermark is done in the fine component.

3)Calculate the energy of the critical band of the approximate components of the first two segments of each frame E_1^i, E_2^i:

$$E_1^i = \sum_{i=1}^{s} \left| A_{1d}^i(t) \right| \quad (7)$$

$$E_2^i = \sum_{i=1}^{s} \left| A_{1d}^i(t) \right| \quad (8)$$

Where **s** is the number of the fine component of each segment.

4)Since the masking threshold calculation is related to the noise characteristic of the audio, the masking threshold is modified according to the noise characteristic factor. If pure tone masks noise, the masking threshold is reduced by 14.5 dB. Conversely, if the noise masks pure tones, the threshold is reduced by 5.5 dB. The degree to which the audio signal is similar to noise can be represented by the noise characteristic coefficient a(z).

$$a(s) = \min[10\lg(G/A)/E_{\max}, 1] \quad (9)$$

Where A and G are the arithmetic mean and geometric mean of the power spectrum, respectively. If the audio signal is pure tone, then a = 1, if the audio signal is white noise, then a = 0. The masking threshold after considering the noise characteristic coefficient is corrected as follows:

$$O(z) = a(14.5 + z) + 5.5(1 - a) \quad (10)$$

We use the larger value of each fine component as the final masking threshold[8].

5)Reconstructing the audio signal: applying the discrete wavelet inverter to $A_{1d}^{i'}, A_{1a}^{i}, A_{2d}^{i'}, A_{2a}^{i},$, reconstructing the audio signal $A_1^{i'}, A_2^{i'}$:

$$A_1^{i'} = IDWT(A_{1d}^{i'}, A_{1a}^{i}) \quad (11)$$

$$A_2^{i'} = IDWT(A_{2d}^{i'}, A_{2a}^{i}) \quad (12)$$

Finally, the watermarked audio signal A' needs to be reconstructed.

3.2. Watermark Extraction Step

1) Divide A' into frames and segments by embedding.

2) For each frame of the first two segments of the audio signal $A_1^{i'} A_2^{i'}$, Haar discrete wavelet transform is performed separately to obtain a fine component $A_{1d}^{i'} A_{2d}^{i'}$ and an approximate component $A_{1d}^{i'} A_{2d}^{i'}$:

$$(A_{1d}^{i'}, A_{1a}^{i'}) = DWT(A_1^{i'}) \quad (13)$$

$$(A_{2d}^{i'}, A_{2a}^{i'}) = DWT(A_2^{i'}) \quad (14)$$

3) After the calculation, the first second fine component $E_1^{i'} E_2^{i'}$ of each frame is compared, and then the watermark sequence is extracted, as follows: if $E_1^{i'} < E_2^{i'}$, the watermark is extracted as "1"; if, the watermark is extracted as "0".

$$W_s = \{W_s(i, j) = W'(k), 1 \le i \le M_1, 1 \le j \le M_2, k = i \times M_2 + j\} \quad (15)$$

Finally, $W'(i)$ is converted to detect the position of the watermark in the binary watermark image, and the watermark image is restored by increasing the dimension.

3.3. Robust Detection

Robustness detection uses the normalized correlation coefficient (NC) to calculate the similarity between the extracted watermark image and the original watermark image[10]. Let bt be the original watermark sequence, bt' is the extracted watermark sequence, and N is the watermark length. The normalized correlation coefficient is calculated as:

$$NC = \frac{\sum_{i=1}^{N} bt(i)bt'(i)}{\sqrt{\sum_{i=1}^{N} bt(i)^2}\sqrt{\sum_{i=1}^{N} bt'(i)^2}} \tag{16}$$

4. Experimental Simulation and Data Analysis

In the experiment, a 32×32 pixel binary image was used as the watermark, and four different types of carrier audio speech signals (classical style, jazz style, classic movie theme song and popular style) were selected, and their sampling frequency was 44.1. KHz, the playback precision is 16 bits, and the playback files are not in time. Taking the classical speech signal as an example, the audio signal is decomposed by Haar wavelet base and 3-level wavelet decomposition. Each decomposed component is calculated by the algorithm to a larger masking threshold, and the determined fine component is located. PSNR=84dB in the audio signal embedded in this position. The original signal is shown in Figure 1, the original image is shown in Figure 2, and the watermark signal is shown in Figure 3.

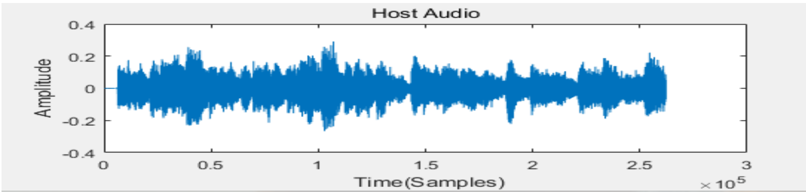

Figure 1. Original audio signal waveform

Figure 2. Original image

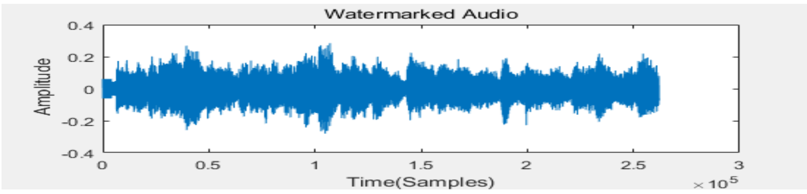

Figure 3. Watermark signal waveform

In order to detect the robustness of the algorithm, it is necessary to use MATLAB and other software to resample the audio signal, low-pass filter, add noise, shear compression, etc. The specific practices are as follows:

4.1. Resampling

The watermark signal is resampled with 48 kHz to obtain a waveform as shown in Figue4, and then the watermark shown in Figue5 is extracted by inverse wavelet transform.

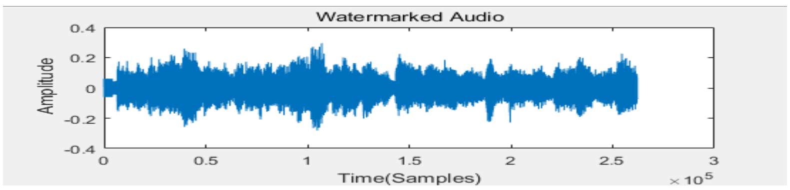

Figure 4. Resampled watermark signal

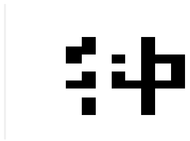

Figure 5. Extracting a watermark from a resampled signal

4.2. Low Pass Filtering

The watermark signal is filtered by a Butterworth low pass filter with N=20, wp=0.1, ws=0.9, rp=0.6, rs=0.9. The filtered waveform is as shown in Figure 6, and the watermark extracted from the signal by the low pass filter is shown in Figure 7.

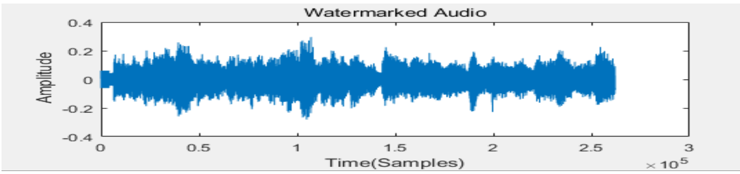

Figure 6. Low pass filtered watermark signal

Figure 7. Signal watermark through low pass filter

4.3. Gaussian Noise

The Gaussian white noise is added to the watermark signal of the Y=AWGN(X, SNR) function, and the watermark signal with Gaussian noise is shown in Figue8. The watermark extraction of the Gaussian noise signal is shown in Figue 9.

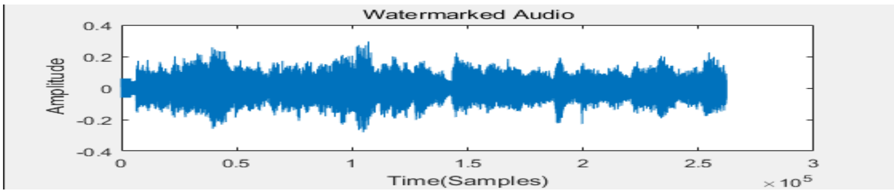

Figure 8. Watermark signal with increased Gaussian noise

Figure 9. Signal watermark with Gaussian noise

4.4. Compression

The audio editing software is selected to compress the entire watermark audio signal, and the compression is 1 dB, and the compression ratio is 89.13%, and the compressed signal waveform as shown in Figue10 and the watermark shown in Figue11 are obtained.

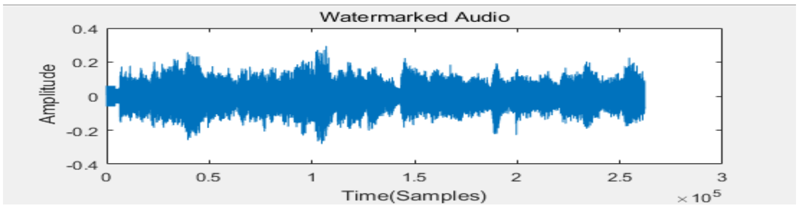

Figure 10. Compressed watermark audio signal

Figure 11. Compressed watermark

4.5. Normalized Correlation Coefficient

In the MATLAB simulation process, in order to detect the robustness of the algorithm, white noise attacks and resampling attacks are performed on the audio files embedded in the watermark respectively. Table 1 shows common voice attack tests, such as low-pass filtering attacks. The extracted watermark, and the corresponding normalized correlation coefficient NC value. The correlation of the watermark is evaluated using the normalization coefficient. For the embedded watermarked audio signal: Table 1 is the unprocessed original image and the extracted watermark NC value. After the original image is resampled and the watermark is extracted, white noise is added to the original image and the watermark is extracted, and the original image is low. Filtering, listening, and compressing, extracting the NC value after watermarking.

Table 1. NC value

Unprocessed	resampling	noise	Low pass filtering	Compression
0.9749	0.9527	0.9458	0.9516	0.8963

In the Matlab simulation process, the normalization function is used: NC= nc(ImageA, ImageB). In this function, ImageA is the original watermark image; ImageB is the extracted watermark image; use this function to get the NC value, which is used for Test the correlation of the watermark extracted from the original image. Referring to the function: x = psnr (a, b), this function is used to calculate the peak signal to noise ratio. In this function, "a" is the original audio signal; "b" is the embedded audio signal, and the similarity between the two signals can be easily obtained by calling the function. It can be seen from the results of Figure 7 and Table 1, the algorithm has strong anti-attack capability.

5. Conclusion

In this paper, image watermarks embedded in audio signals are implemented by adaptive algorithms using human auditory characteristics and discrete wavelet analysis. The algorithm has certain reliability and specific characteristics. Watermark extraction belongs to blind watermark extraction and does not require real-time carrier audio signals. The simulation results show that the algorithm has certain transparency and is robust to resampling and low-pass filtering attacks. Therefore, the audio watermarking technology based on wavelet transform is applied to the field of audio information security, and it is convenient to embed and extract specific meaning information. Since this article only deals with a small frame of audio, the impact of the watermark on this will be considered in future research and more comparisons will be made.

6. Acknowledgments

Thank you for Guizhou Science and Technology Plan Project **(**Guizhou Kehe [2017] 5305; 5788; 2816, [2016] 5604, [2015] 3054, [2014] 7633);

IOP Conf. Series: Materials Science and Engineering **719** (2020) 012033 doi:10.1088/1757-899X/719/1/012033

7. References

[1] Chen Yanli, Wang Haitao, Wang Juan. Digital watermark analysis of color image based on discrete wavelet transform[J].*Fujian Computer*,2018,1:14-15.
[2] Hua G, Huang J, Shi Y Q,et al. Twenty years of digital audio watermarking a comprehensice review[J].*Signal Processing*,2016,128:222-242.
[3] SUN Rui, JIANG Jian-guo, YANG Xing-ming.Audio watermarking algorithm based on wavelet packet and psychoacoustic model[J].*Journal of System Simulation*,2007,19(10):2248-2251.
[4] Bai Yubao, Bai Sen, Liu Chenghao. Digital audio watermarking algorithm based on wavelet and cepstrum domain[J]. *Journal of Computer Applications and Software*, 2012(3):163-167.
[5] W ang Wei, Wang Ji. Research on digital video watermarking technology[J].*Jiangsu Science and Technology Information*, 2018, 8:41-43.
[6] Zhao Cui. Research on digital watermarking technology of audio signal based on discrete wavelet transform[J]. *Radio and TV Technology*,2019.
[7] Yang Yuting. Research on digital watermarking algorithm based on wavelet transform[D]. *Chengdu University of Technology*, 2017.
[8] Jonston J D.Transform coding of audio signals using perceptual noise criteria[J].*IEEE Journal on Selected Areas Communications*,1988,6(2):314-323.
[9] Li Jingli.Research on the implementation of audio digital watermarking [D]. *Zhengzhou University*, 2010.
[10] Liang Jiadong, Yang Shuguo. A digital Image watermarking technique based on wavelet analysis[J]. *Modern computer*,2016:71-74.