Journal of Research in Science and Engineering (JRSE)
ISSN: 1656-1996 Volume-8, Issue-10, Octobery 2026
DOI: 10.66069/ojspub.16561009

A DWT-Based Audio Watermarking Algorithm Using Wavelet Component Proportional Coefficients for Balanced Robustness and Inaudibility

Fang Yu, Chongyuan Liu*

College of Computer Science and Engineering, North China University of Science and Technology, Tangshan 063210, Hebei, China
**Correspondence Author*

Abstract: ***Research on audio watermarking has concentrated predominantly on robustness, yet most existing algorithms fail to sustain high robustness under conditions of complete inaudibility, and reconciling this trade-off between imperceptibility and attack resilience remains a central challenge in the field. To address this problem, this paper proposes a digital watermarking algorithm based on wavelet component proportional coefficients, designed explicitly to balance robustness and inaudibility. This paper proposes a digital watermarking algorithm based on wavelet component proportional coefficients. First, the algorithm computes the wavelet coefficients by Discrete Wavelet Transformation (DWT), then gets the wavelet coefficients proportion of the two sub-segments of every segments in each frame, then produces the wavelet coefficients of all sub-segments by two modular addition; finally, the watermark is embedded by inverse DWT based on the wavelet coefficients scaled by the coefficient proportions before and after the process. The simulation results show that the proposed algorithm can provides the good inaudibility and robustness to conventional signal processing, such as MP3 compression, low-pass filtering, resampling, etc.***

Keywords: Audio watermarking, Discrete wavelet transform (DWT), Robustness, Inaudibility, Wavelet component proportional coefficients.

1. Introduction

Digital audio watermarking is a branch of information hiding. It is usually used to guarantee the copyright, integrity and validity of digital products, such as copyright protection and authentication, pirate tracking of audio works, fingerprint identification [1-4]. In the design of digital watermarking algorithms, robustness and inaudibility are two basic requirements in audio systems, but they contradict each other. Although great progress has been made in robust watermarking for copyright protection in recent years, it remains a challenge for audio signals to have high robustness and high inaudibility. Wang [5] and others first proposed a watermarking algorithm based on DCT (Discrete Cosine Transform) in 1998; according to the masking characteristics of the human auditory system, Hu [6] et al. Proposed a large-capacity audio watermarking algorithm that embeds a watermark on three sets of DCT coefficients in 2014; A. Kaur [7] et al. Put forward DCT and select a subband, and use the embedding function to take the R matrix as the embedding position of watermark. Document [8] and document [9] propose a high robust digital watermarking, but it improves the robustness of digital watermarking in DCT domain and reduces the high capacity of DCT domain. Tang [10] et al., proposed a zero-water compressed sensing and DWT audio zero-watermarking algorithm in the transformation domain with positive and negative polarities.

The characteristics of audio watermarking technology mainly include four aspects: invisibility, robustness, certainty and security [11]. Most DWT-based methods have their own characteristics and limitations [12]. In order to ensure that the watermark signal has good robustness, the wavelet coefficients are not screened, so that some coefficients are greatly affected by compression, thus increasing the bit error rate, affecting the audio quality of audio, and reducing the inaudibility of audio signals.

Aiming at the above problems, this paper proposes a digital watermarking algorithm based on the wavelet component scale factor of the audio signal. By performing three-level wavelet transform on each segment, the wavelet component proportional coefficients of the two sub-segments before and after each wavelet coefficient are analyzed, and a method based on wavelet component proportional coefficient extraction is given. In the audio signal quantization stage, based on this feature, a new quantization method is introduced. The quantization method can enable multiple digital watermarks to quantize an audio signal, ensure that the audio signal is less changed, and ensure the robustness while improving the inaudibility of the audio signal. The main contributions of this work are summarized as follows:

(1) An embedding feature based on wavelet component proportional coefficients is proposed. This feature can resist normal signal processing, making it more robust to signal processing.

(2) Proposed a new quantitative method. This method can reduce the modification of the watermark signal to the audio signal, so that it has better inaudibility after signal processing. This paper is organized as follows. Section II introduce the extraction process of wavelet component proportional coefficient. Section III the watermarking scheme is introduced, and then the embedding and extraction process of the scheme is described in detail. Finally, section IV provide experimental

results and analysis.

2. Audio Signal Wavelet Component Scale Factor Extraction

In this section, we first introduce the wavelet component and wavelet component scale factor extraction process, and then prove the robustness of the wavelet component scale factor by comparing the change of the wavelet component scale factor of the original audio signal and the processed signal.

2.1. Discrete Wavelet and Discrete Wavelet Transform

In digital signal processing, continuous wavelets need to be discretized. In general computer implementation, binary discrete processing is used to transform this discretized wavelet and its corresponding wavelet into discrete wavelet transform (DWT). In fact, the discrete wavelet transform is obtained by discretizing the scale and displacement of the continuous wavelet transform according to the power of *2*, so it is also called a binary wavelet transform, and the discrete wavelet is expressed as follows:

$$\psi_{j,k} = 2^{j/2}\psi(2^i t - k) \quad j,k \in Z \tag{1}$$

Accordingly, the discrete wavelet transform is as follows:

$$(W_\psi f)(\frac{1}{2^j}, \frac{k}{2^j}) = \int_{-\infty}^{\infty} f(t)\{2^{j/2}\psi(2^j t - k)\} \mid dt = \sum_{j,k}(f, \psi_{j,k}) \tag{2}$$

By selecting the wavelet function a so that its dual b exists, the formula for reconstructing the original signal by discrete wavelet transform is as follows:

$$f(t) = \sum_{-\infty}^{\infty} < f, \psi_{j,k} > \psi_{j,k} \tag{3}$$

If the wavelet function ψ is an R wavelet, then the duality of ψ must exist. R wavelet can be divided into the following three categories by definition.

If ψ is an orthogonal wavelet, then $\psi_{j,k}$ satisfies the orthogonality condition as follows:

$$< \psi_{j,k}, \psi_{l,m} >= \delta_{jl}\delta_{km} \quad j,k,l,m \in Z \tag{4}$$

If ψ is a semi-orthogonal wavelet, then $\psi_{j,k}$ satisfies the orthogonality condition as follows:

$$< \psi_{j,k}, \psi_{l,m} >= 0 \quad j \neq l, j,k,l,m \in Z \tag{5}$$

If ψ is a biorthogonal wavelet (or non-orthogonal wavelet), then the dual of ψ exists, and the conditions are as for:

$$< \psi_{j,k}, \psi_{l,m} >= \delta_{jl}\delta_{km} \quad j \neq l, j,k,l,m \in Z \tag{6}$$

2.2 Wavelet Component Scale Factor

Define the audio signal as *X (n) (n = 1, 2, ... , N)*, the solving process of wavelet component proportional coefficient is as follows:

(1) The digital audio signal with sampling number *N* is denoted as *X(n)* (n = 1,2, ... ,N), and divide *X* into non-overlapping I frames. Frame i is A^i *(i = 1, 2... I)*.

(2) A_i is divided into M sections, denoted as $A_{i,m}$ (m=1,2... ,M), take the third-order wavelet transform of $A_{i,m}$, and the wavelet coefficient is denoted as $A_{i,m}^{l3}$.

(3) Divide $A_{i,m}^{l3}$ into two parts with the same length, and represent the data of the two sub-segments as D_1 and D_2 respectively. Select $D_{1,h}{}^+$and $D_{2,h}{}^+$ of the first *K* data with size from *0* to *1* in D_1 and D_2, among which $D_{1,h}{}^+$ and $D_{2,h}{}^+$are H. The scaling coefficients of audio wavelet components can be obtained from the following functions:

$$F = \frac{1}{H} * \sum_{h=1}^{H} \lg \frac{\left(D_{1,h}{}^+\right)^2}{\left(D_{1,h}{}^+ - D_{2,h}{}^+\right)^2} \tag{7}$$

Among them, F represents the proportion coefficient of the wavelet component, while H is the qualified quantity of D_1 and D_2. $D_{1,h}{}^+$ and $D_{2,h}{}^+$ are D_1,D_2 eligible data.

2.3 Robustness of Wavelet Component Proportional Coefficient

The following gives the amplitude of the wavelet component proportional coefficient change after signal processing, where *X* indicates a randomly selected voice signal with a sampling frequency of *44.1 kHz*, where parameter *K=100*. The wavelet component proportional coefficients of the original signal and the processed signal are shown in Figure 1, Figure 2 and Figure 3. The X-axis represents the serial number corresponding to the frame, and the Y-axis represents the wavelet component proportional coefficient corresponding to each frame. It can be seen that before and after signal processing, the wavelet component proportional coefficient remains almost unchanged. It shows that this feature has a certain ability to tolerate signal processing.

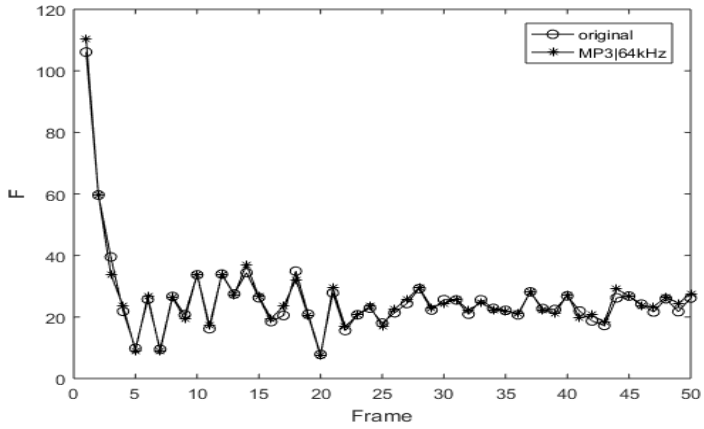


Figure 1: Wavelet component scale factor before and after MP3 Compression of Raw Signals

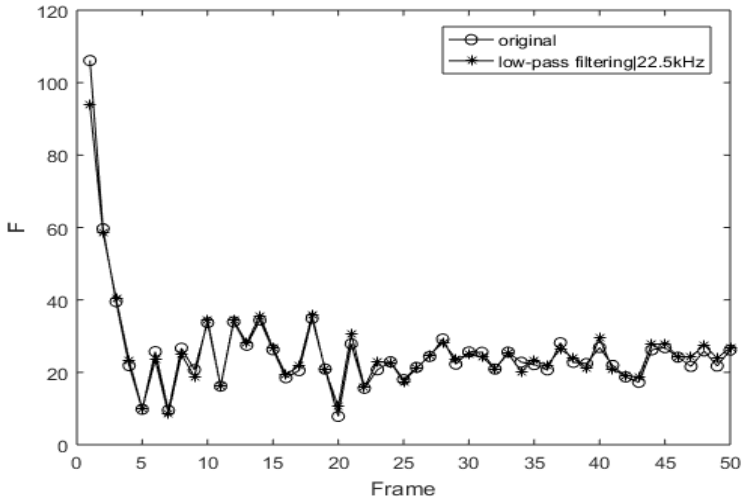


Figure 2: Wavelet component scale factor before and after low-pass filtering of original signals

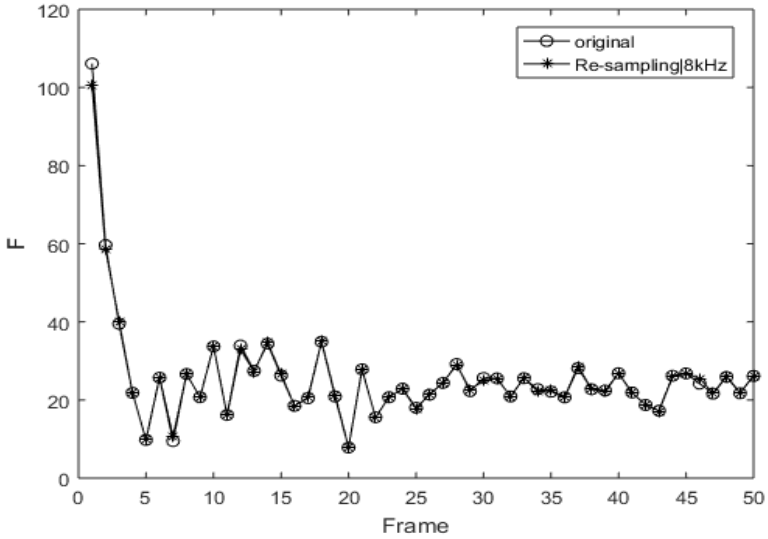


Figure 3: Wavelet component scale factor before and after resampling of original signals

3. Watermarking Scheme

3.1. Pretreatment

Pre-processing digital watermarking Y is the following:

(1) Divide the digital watermarking from right to left every 4 bits, and make up 0 on the left of the least 4 bits.

(2) Converting every 4-digit digital watermarking separated from binary to 16-digit.

(3) Combining the digital watermarking converted to 16-digit to form a new digital watermarking Y′.

3.2 Watermarking Embedding

In this paper, we will take the embedding of Y as an example to introduce the embedding method of the wavelet component scale factor. The detailed steps of the watermarking embedding block diagram as shown in Figure 4.

Step 1: A_i is divided into M segments, which are recorded as $A_{i,1}$, $A_{i,2}$,..., $A_{i,m}$, and then the wavelet coefficients of three-level wavelet transform for each segment are recorded as $A_{i,1}^{l3}A_{i,2}^{l3}\ldots A_{i,m}^{l3}$.

Step 2: For $A_{i,1}^{l3}A_{i,2}^{l3}\ldots A_{i,m}^{l3}$ is divided into two sub-parts. Each sub-part takes the coefficients of the size of the front K from 0 to 1, calculates wavelet component scale factor, and is recorded as $F_1, F_2, \ldots, F_m$.

Step 3: Select the first digits of $F_1, F_2, \ldots, F_m$, and mark them as $a_1,a_2,\ldots a_m$.

Step 4: Calculates $V=f(a_1,a_2,\ldots a_m)$, the formula is as follows:

$$f(a_1,a_2,\cdots,a_m)=(2\text{^}(a_1 \bmod 5)+2\text{^}(a_2 \bmod 5) +\cdots+2\text{^}(a_m \bmod 5)) \bmod 16 \quad (8)$$

Step 5: Calculates the difference D between V and the watermarking embedded in this frame, the formula is as follows:

$$D=(V-y_1+16) \bmod 16 \quad (9)$$

1) If $D=0$, then $a_1, a_2,\ldots a_m$ is unchanged

2) If $D\neq 0$

Let $a_{m1}mod5=1$, $a_{m2}mod5=2$, $a_{m3}mod5=3$, $a_{m4}mod5=4$, $a_{m5}mod5=0$. It is known that every increase of a_{m1}、a_{m2}、a_{m3}、

a_{m4}、a_{m5}, the value of D increases by *2*, *4*, *8*, *16*, *1*, so we make $D = 0$ by modifying $_{m1},a_{m2},a_{m3},a_{m4},a_{m5}$. For example, $D = 8$, only need to fix the value of a_{m3} so that $a_{m3}=a_{m3}+1$, that is, only need to modify a coefficient, can complete the embedding of four watermarks. The watermarking embedding reduces the modification of the original digital speech signal and improves the audibility.

Step 6: calculates the quantized a'_1, a'_2,...a'_m and the proportional coefficients α of a_1,a_2,...a_m before quantization, the formula is as follows:

$$\alpha(m) = \frac{a'_m}{a_m} \tag{10}$$

Step 7: Scales the DWT coefficients of the first K part of each subsection equally, the formula is as follows:

$$A^{l3'}_{i,m} = \alpha(m) * A^{l3}_{i,m} \tag{11}$$

Then, three times inverse DWT is performed to complete y_1' embedding.

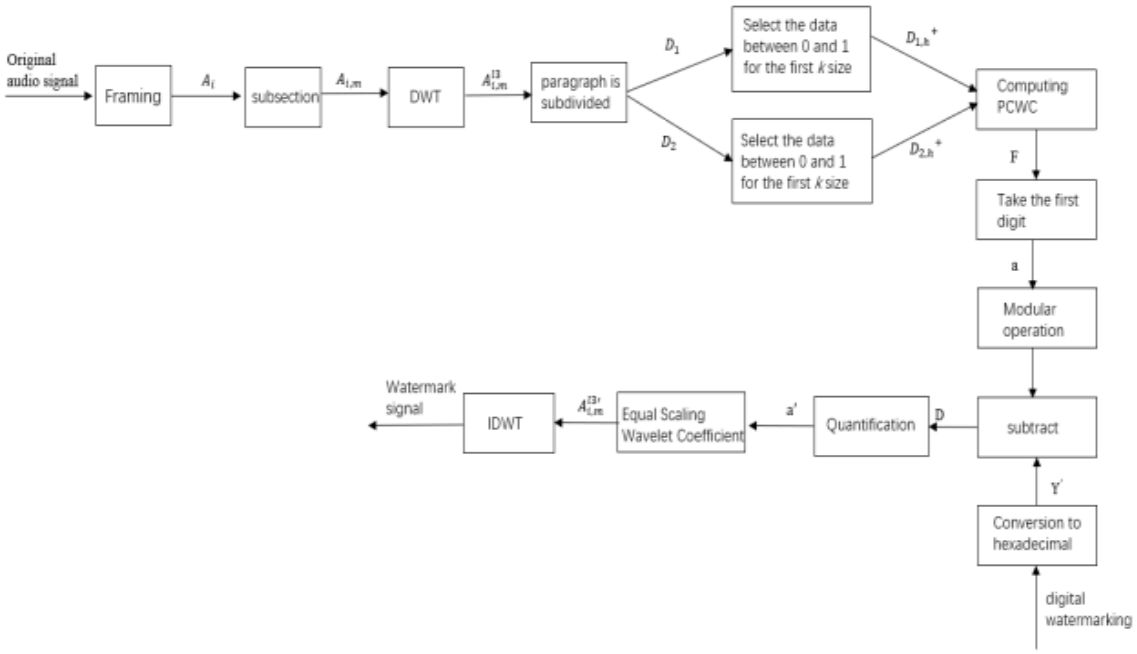


Figure 4: Block diagram of watermarking generation and embedding process

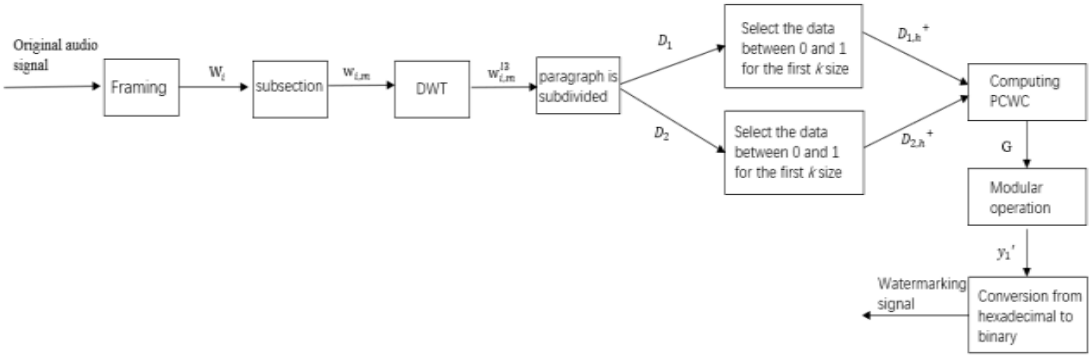


Figure 5: Watermarking extraction process

3.3 Watermarking Extraction

Let W be the watermarking signal, and the detailed steps of watermarking extraction block diagram as shown in Figure 5.

Step 1: W_i is divided into M segments, which are recorded as $W_{i,1},W_{i,2},\ldots,W_{i,m}$ and then the wavelet coefficients of three-level wavelet transform for each segment are recorded as $W^{l3}_{i,1},W^{l3}_{i,2}\ldots W^{l3}_{i,m}$.

Step 2: For a $W^{l3}_{i,1},W^{l3}_{i,2}\ldots W^{l3}_{i,m}$ is divided into two sub-parts. Each sub-part takes the coefficients of the size of the front K from *0* to *1*, calculates wavelet component scale factor, and is recorded as $G_1,G_2,\ldots G_m$.

Step 3: Select the first digits of $G_1,G_2,\ldots G_m$, and mark them as $b_1,b_2,\ldots,b_m$.

Step 4: Calculates $V=(b_1,b_2,\ldots,b_m)$, the formula is as follows:

$$f(b_1,b_2,\cdots,b_m) = (2\hat{}(b_1 \bmod 5) + 2\hat{}(b_2 \bmod 5) + \cdots + 2\hat{}(b_m \bmod 5)) \bmod 16 \tag{12}$$

Step 5: converts the extracted y_1' from hexadecimal to binary to get the embedded watermarking y_1 in this frame.

4. Performance Analysis

The audio signal we selected has 1384823 sample points, the sampling rate is 44.1 kHz, and the 16 bits quantized mono WAVE speech signal is used as the test. They are divided into

three categories: T1, T2 and T3. T1 is rock music, T2 is pop music and T3 is classical music.

4.1 Inaudibility Test

As shown in Figure 6, on the one hand, it is difficult to detect the changes in the waveform by comparing the original audio with the waveform embedded in the watermarking. On the other hand, we use signal to noise ratio (SNR) to evaluate the auditory system of watermarked audio objectively. Other parameter settings: I=50, M=50, K=100

The definition of SNR is as follows:

$$SNR = 10\lg\left(\frac{\sum_{i=1}^{I} a(i)^2}{\sum_{i=1}^{I} (a(i) - a'(i))^2}\right) \quad (13)$$

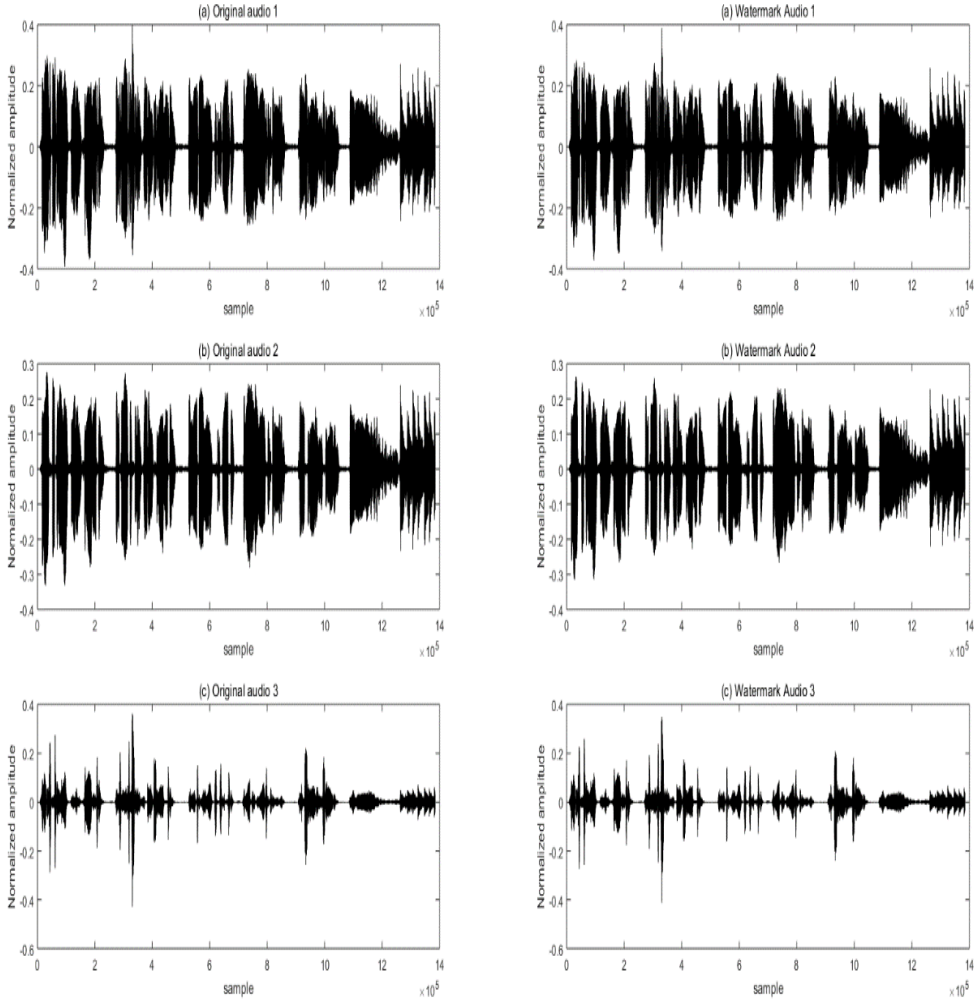


Figure 6: Original audio waveform and watermarked audio waveform

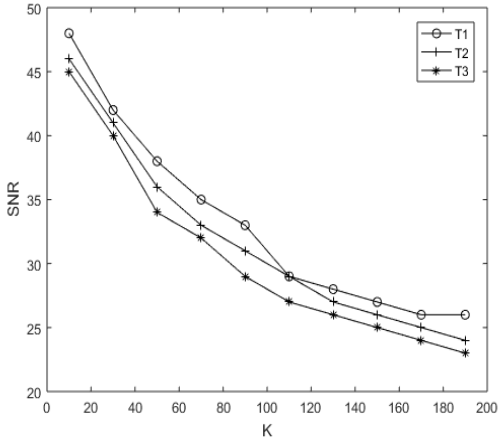


Figure 7: SNR value of watermarked audio at different K

Among them, *a(i)* is the original audio signal and *a'(i)* is the watermarked audio signal. Figure7 gives the SNR values of watermarked audio at different *K*. It can be seen that the SNR values decrease with the increase of *K*.

When K < [0,100], the values of T1, T2 and T3 are all greater than 29dB, which shows that the algorithm has higher SNR values. From the above objective evaluation results, we can see that the watermarking algorithm in this paper has good imperceptibility, and the audio quality of watermarked audio tends to decline with the increase of K value.

4.2 Robustness Test

Because of the need of storage or the special requirement of playback format, audio signals often go through some signal processing operations, such as MP3 compression. If the watermarking cannot be extracted correctly after the watermarking signal is processed by signal processing, the user will mistake the signal processing operation as malicious attack. In order to avoid this kind of problem, the watermarking algorithm is required to be robust to the signal. In order to further illustrate the ability of this algorithm to

tolerate conventional signal processing, we have carried out a series of signal processing experiments. In this paper, the robustness against conventional signal processing is measured by NC (normalized cross-correlation), which is defined as follows:

$$NC = \frac{\sum_{i=1}^{M\times(n/n1)} w(i)\cdot\hat{w}(i)}{\sqrt{\sum_{i=1}^{M\times(n/n1)} w(i).\hat{w}(i)}\sqrt{\sum_{i=1}^{M\times(n/n1)} \hat{w}(i).w}} \quad (14)$$

Among them, *w(i)* is the original watermarking sequence, $\hat{w}(i)$ is the watermarking sequence extracted after a certain degree of signal processing, and $M\times(n/n1)$ is the length of the watermarking in this algorithm. The larger the NC value, the stronger the ability of watermarking to tolerate conventional signal processing.

Table 1 shows NC values of three audio segments after different signal processing. In Table 1, MP3 compression (1), (2), (3) corresponding compression rates are 32 kbps, 64 kbps and 128 kbps, respectively. For resampling (1), (2) the original signal sampling frequency is reduced from 44.1 kHZ to 22.05 kHZ and 11.025 kHZ respectively. The results listed in Table 1 show that the proposed algorithm is robust to signal processing.

Table 1: NC value extracted from different signal processing

signal processing		*NC*		
		T1	*T2*	*T3*
MP3 Compression	1	0.88	0.86	0.87
	2	0.94	0.94	0.94
	3	1	1	1
Resampling	1	1	1	1
	2	0.94	0.93	0.99
LowPass Filtering	21.5kHz	0.98	0.98	0.98

5. Conclusions and Future Work

In this paper, a scaling coefficient of wavelet components for watermarking embedding is proposed. Firstly, the coefficients of each segment are obtained by wavelet transform. After dividing the coefficients equally, the proportion coefficients of each segment with stable and invariable wavelet components are calculated after being processed by signal. Secondly, a new quantization method is used to embed the watermarking, which reduces the modification position of the original audio signal and improves the audibility of the audio signal embedded in the watermarking. Experiments on several signal processing operations show that the proposed watermarking embedding method has good performance in both imperceptibility and robustness, especially in imperceptibility. In the future, we will try to resist the synchronization attacks such as clipping and TSM better when facing malicious attacks.

References

[1] Hua G, Huang J W, Shi Y Q, et al., Twenty Years of Digital Audio Watermarking-a Comprehensive Review [J], Signal Processing, 2016, 128(11): 222-242.

[2] Hou Xiang and Min Lianquan, A Robust Watermarking Algorithm Using SURF Feature Regions [J], Geomatics and Information Science of Wuhan University, 2017, 42(3): 421-726.

[3] Nishimura, Audio Watermarking based on Subband Amplitude Modulation [J], Acoustical Science and Technology, 2010, 31(5): 328-336.

[4] Lin Wei, Wang Yuhai, Ren Na, et al., QR Code Based Research on Digital Watermarking Algorithm for Tile Remote Sensing Image [J]. Geomatics and Information Science of Wuhan University, 2017, 42(8): 1151-1158.

[5] Wang Y, A new watermarking method of digital audio content for copyright protection [C]//Fourth International Conference on Signal Processing, IEEE, 2002.

[6] Hu H T, Hsu L Y. Robust, transparent and high-capacity audio watermarking in DCT domain [J], Signal Processing, 2015, 109(C): 226-235.

[7] A. Kaur, M. K. Dutta, K. M. Soni and N. Taneja, "A High Payload Audio Watermarking Algorithm Robust against Mp3 Compression," in 2014 Seventh International Conference on Contemporary Computing (IC3), 2014, no. 1, 531–535.

[8] R. Aminullah, G. Budiman, and I. Safitri, "Implementation of QIM Based Audio Watermarking Using Hybrid Transform Of SWT-DCT-SVD Methods Optimized With Genetic Algorithm," Jurnal Penelitian dan Pengembangan Telekomunikasi, Kendali, Komputer, Elektrik, dan Elektronika, vol. 2, no. 1, 2017.

[9] H. T. Hu and L. Y. Hsu, "Robust, transparent and high-capacity audio watermarking in DCT domain," Signal Processing, vol. 109, 226–235, 2015.

[10] Tang X, Ma Z F, Niu X Y, et al., Compressive sensing-based audio semi-fragile zero-watermarking Algorithm [J], Chinese Journal of Electronics, 2015, 24(3): 492-497.

[11] Pan Zhihong, Li Zhijuan and Gong Zixia, Overview of digital watermarking technology [J], Computer and digital engineering, 2008, 36 (4): 119-12.

[12] Hua Bin, Research on digital audio watermarking algorithm [D], Jilin University, 2004.