

# A Lifting Wavelet Domain Audio Watermarking Algorithm Based on the Statistical Characteristics of Sub-Band Coefficients

Zhi TAO<sup>(1),(2)</sup>, He-ming ZHAO<sup>(1)</sup>, Jun WU<sup>(2)</sup>, Ji-hua GU<sup>(2)</sup>,  
Yi-shen XU<sup>(2)</sup>, Di WU<sup>(2)</sup>

<sup>(1)</sup> *Soochow University*  
*School of Electronic Information*  
Suzhou 215006, China  
e-mail: hmzhao@suda.edu.cn

<sup>(2)</sup> *Soochow University*  
*School of Physical Science and Technology*  
Suzhou 215006, China

(received March 15, 2010; accepted September 16, 2010)

In this paper, a new lifting wavelet domain audio watermarking algorithm based on the statistical characteristics of sub-band coefficients is proposed. First of all, an original audio signal was segmented and each segment was divided into two sections. Then, the Barker code was used for synchronization, the LWT (lifting wavelet transform) was performed on each section, a synchronization code and a watermark were embedded into the first section and the second section, respectively, by modifying the statistical average value of the sub-band coefficients. The embed strength was determined adaptively according to the auditory masking property. Experiments show that the embedded watermark has better robustness against common signal processing attacks than present algorithms based on LWT and can resist random cropping in particular.

**Keywords:** audio watermarking, lifting wavelet transform, statistical characteristics, sub-band coefficients.

## 1. Introduction

The technology of digital audio watermarking has been considered as an effective way of addressing the copyright protection of digital audio products. Without affecting the usefulness of the product, it embeds the copyright information into the original audio, verifying the copyright and integrity of the audio document

by the detectable watermark. A lot of audio watermarking schemes based on the WT (Wavelet Transform) have proven to have good performance in recent years due to the multi-resolution characteristic of the WT (QIANG, WANG, 2004). However, the classic WT was calculated by convolution, which needs extensive and complex computations and has high storage requirements (BAI, 2008). In addition to this, it usually produces floating numbers after the classic WT. Furthermore, the origin signal cannot be reconstructed because of the limited word length that most computer systems allow for (GAO *et al.*, 2007). Finally, classic WT restricts the performance of the wavelet domain audio watermark to some extent.

Sweldens (SWELDENS, 1997; DAUBECHIES, SWELDENS, 1998) put forward the lifting scheme of wavelets for the first time. The lifting scheme does not rely on the Fourier transform and allows a fully in-place calculation as well (WIM, 1996). Compared with the classic wavelet transform, it allows faster implementation, and the inverse transform can immediately be found by undoing the operations of the forward transform; no auxiliary memory is needed. The lifting scheme has managed to maintain the advantages of the classic wavelet transform while overcoming its limitations. We can simply apply it to audio watermarking (GAO *et al.*, 2007).

An algorithm based on LWT is proposed by WANG *et al.* (2005), where the watermark signal was embedded into the sub-band coefficients utilizing the method of quantization. This showed that watermark detection can be implemented quickly without the original signal, but the method did not prove to be very robust. The following improved algorithm presented by WANG *et al.* (2006) applied another method of quantization that proved to be more robust. XU and WANG (2006) the watermark was embedded through modifying the middle frequency component, and the watermark was detected by correlating the approach. However, the robustness was again not good, especially in connection with MP3 compression. A strategy for embedding the watermark into an average value of a part of the sub-band coefficients by quantization in the LWT domain is proposed by WANG and XU (2006). This strategy was shown to be robust against common attacks.

The audio watermarking algorithms based on the lifting wavelet above realize the watermark embedding by modifying one or several coefficients of the sub-band in a special position. So we must depend on the exact alignment of the audio watermark signal that is to be detected and the origin signal when extracting the watermark (XIANG, 2006). Once the position was lost by random cropping, the proper watermark cannot be detected in a simple way (WEI, 2004). QU (2006) included a synchronization signal while embedding the watermark, and the detection began after the synchronization signal was located. Although this method can resist some random cropping attacks, the proper watermark still cannot be detected if the signal containing the watermark is cut.

In this paper, a lifting wavelet domain audio watermarking algorithm based on the statistical characteristics of sub-band coefficients is proposed. Before embed-

ding the watermark, we add a synchronization code which makes the watermark self-synchronized and embed the watermark into the statistical character of the sub-band coefficients that are insensitive to various attacks. The embed strength was decided adaptively by the application of an auditory masking property (TAO *et al.*, 2006). Results show that this approach can effectively resist common signal processing attacks, and that it has an especially strong robustness against random cropping.

## 2. The embedding strategy

### 2.1. Watermark preprocessing

Let  $A$  represents the binary image to be embedded:

$$A = \{a(m, n), 0 < m \leq R, 0 < n \leq C\}, \quad a(m, n) \in \{0, 1\}. \quad (1)$$

In order to hide  $A$  within the audio signal, we should turn it into the one-dimension sequence  $B$ :

$$B = \{b(i) = a(m, n), 0 < m \leq R, 0 < n \leq C, i = m \times C + n\}. \quad (2)$$

Then, add the synchronization code  $SYN = \{syn(i), 0 < i \leq Lsyn\}$  in front of  $B$  where  $Lsyn$  denotes the length of the synchronization code:

$$D = \{SYN, B\} = \{d(i), 0 < i \leq (R \times C + Lsyn)\}. \quad (3)$$

To improve the robustness of the hidden information, scramble  $D$  using the logistic mapping sequence  $E$ :

$$E = \{e(i), 0 < i \leq (R \times C + Lsyn)\}, \quad e(i) \in \{0, 1\}. \quad (4)$$

The scrambling expression is  $F = D \oplus E$ , where  $\oplus$  denotes XOR and the result sequence is  $F$ :

$$F = \{f(i), 0 < i \leq (R \times C + Lsyn)\}, \quad f(i) \in \{0, 1\}. \quad (5)$$

Finally,  $F$  is mapped by means of BPSK modulation into the antipodal sequence  $W$ , which is the data to be embedded through the following equation:

$$W = \{w(i) = 1 - 2f(i)\}, \quad 0 < i \leq (R \times C + Lsyn). \quad (6)$$

### 2.2. Watermark embedding

Figure 1 shows structure of the embedded data. In this paper we segment the audio signal as indicated in Fig. 2.

In Fig. 2,  $N1 = Lsyn \times SL$ ,  $N2 = R \times C \times WL$ , embedding one bit of the synchronization code needs  $SL$  samples, and embedding one bit of the watermark needs  $WL$  samples.

...	synchronize code	watermark	synchronize code	watermark	...
-----	---------------------	-----------	---------------------	-----------	-----

Fig. 1. The structure of the embedded data.

...	Seg1 $N_1$	Seg2 $N_2$	Seg1 $N_1$	Seg2 $N_2$	...
-----	---------------	---------------	---------------	---------------	-----

Fig. 2. The structure of the audio segmentation.

(1) Let  $S^j$  denote segment  $j$ , divided into the two sections  $S_1^j$  and  $S_2^j$ . Then,  $S_1^j$  is partitioned into  $Lsyn$  frames  $\left\{S_1^j(i), 0 < i \leq Lsyn\right\}$  used to hide the synchronization code, while  $S_2^j$  contains  $R \times C$  frames  $\left\{S_2^j(i), 0 < i \leq R \times C\right\}$  used to embed the watermark.

(2) Then, the  $k$ -level LWT is performed on  $\left\{S_1^j(i), 0 < i \leq Lsyn\right\}$  and  $\left\{S_2^j(i), 0 < i \leq R \times C\right\}$ . This allows us to gain the lifting wavelet coefficients. They are  $Sa_1^j(k)^k, Sd_1^j(i)^k, Sd_1^j(i)^{k-1}, \dots, Sd_1^j(i)^1, Sa_2^j(i)^k, Sd_2^j(i)^k, Sd_2^j(i)^{k-1}, \dots, Sd_2^j(i)^1$ , where  $Sa_1^j(i)^k, Sa_2^j(i)^k$  are the approximate weight of the  $k$ -level's lifting wavelet of  $S_1^j(i), S_2^j(i)$  while  $Sd_1^j(i)^k, Sd_1^j(i)^{k-1}, \dots, Sd_1^j(i)$  and  $Sd_2^j(i)^k, Sd_2^j(i)^{k-1}, \dots, Sd_2^j(i)^1$  are the detailed components of level 1 to  $k$ 's lifting wavelet of  $S_1^j(i), S_2^j(i)$ .

(3) The synchronization code and the watermark are embedded by modifying the statistical characteristic of  $Sa_1^j(i)^k, Sa_2^j(i)^k$  respectively according to Eqs. (7) and (8):

$$Sa_1^j(i)^{k'}(t) = \begin{cases} Sa_1^j(i)^k(t) - \overline{Sa_1^j(i)^k} + \delta, & \text{if } W(i) = 1, \\ Sa_1^j(i)^k(t) - \overline{Sa_1^j(i)^k} - \delta, & \text{if } W(i) = -1, \end{cases} \quad (7)$$

for  $0 < i \leq Lsyn, 0 < t \leq SL$ .

$$Sa_2^j(i)^{k'}(t) = \begin{cases} Sa_2^j(i)^k(t) - \overline{Sa_2^j(i)^k} + \delta, & \text{if } W(i + Lsyn) = 1, \\ Sa_2^j(i)^k(t) - \overline{Sa_2^j(i)^k} - \delta, & \text{if } W(i + Lsyn) = -1, \end{cases} \quad (8)$$

for  $0 < i \leq R \times C, 0 < t \leq WL$ .

Here,  $\overline{Sa_1^j(i)^k}$  and  $\overline{Sa_2^j(i)^k}$  are the statistical average values of  $Sa_1^j(i)^k$  and  $Sa_2^j(i)^k$  while  $\delta$  represents the embed strength factor, an initial value is given and then the final value is determined according to the masking threshold value of the audio signal as described in Sec. 3.

(4) With the synchronization code and the watermark embedded in  $Sa_1^j(i)^{k'}$  and  $Sa_2^j(i)^{k'}$ , the  $k$ -level ILWT is performed so we can get the audio data  $S^{j*}$  consisting of  $S_1^{j*}$  and  $S_2^{j*}$  which contains hidden information.

(5) The noise form of the embedded information in the time domain  $Sd^{j*}$  can be estimated by calculating the difference between  $S^{j*}$  and  $S^j$ . According to Sec. 3, if the power spectrum of  $Sd^{j*}$  is under the masking threshold curve of the audio signal, then the value of  $\delta$  should not be changed. If this is not the case, change the embed strength  $\delta$  until the embedded data cannot be perceived by the human auditory system. The final determined signal  $S_1^{j'}$  and  $S_2^{j'}$  make up one segment of the audio signal  $S^{j'}$  which contains an unperceivable watermark and synchronization code.

(6) Repeat step (2), (3), (4), (5) to embed several segments of hidden information to enhance the robustness of the proposed scheme.

### 3. Detection strategy

The watermark detecting procedure in the proposed scheme does not need the original audio signal and the procedure can be summarized as follows:

(1) Locate the initial position SS of the embedded segment using the frame synchronization technology of digital communications.

(2) Compute the sub-band coefficients of the watermark embedded in the audio section by performing the  $k$ -level LWT on it.

(3) Extract the embedded watermark using the following Eq. (9):

$$V(i) = \begin{cases} 1, & \text{if } \overline{Sa(i)^k}^* > 0, \\ -1, & \text{if } \overline{Sa(i)^k}^* < 0, \end{cases} \quad (9)$$

where  $\overline{Sa(i)^k}^*$  is the statistical average value of  $Sa(i)^k$ .

(4) Repeat the above steps until the entire embedded watermark signal is extracted, and at last an optimal watermark  $V' = \{v'(i), 0 < i \leq R \times C\}$  can be decided based on the rule of majority.

(5) Obtain  $V''$  by BPSK demodulation of  $V'$ , as equation (10) shows:

$$V'' = \{V''(i) = (1 - v'(i))/2, 0 < i \leq R \times C\}. \quad (10)$$

(6) Descramble  $V''$  to obtain  $W' = \{w'(i), 0 < i \leq R \times C\}$ , thereby acquiring the final extracted binary watermark image  $A' = \{a'(m, n), 0 < m \leq R, 0 < n \leq C\}$ ,  $a'(m, n) \in \{0, 1\}$ . In order to eliminate the subjective element of the observer,

the usually normalized correlation coefficient NC is used to evaluate the similarity of the extracted and the original watermark, defined in equation (11):

$$NC(A, A') = \frac{\sum_{m=0}^{R-1} \sum_{n=0}^{C-1} a(m, n) \times a'(m, n)}{\sqrt{\sum_{m=0}^{R-1} \sum_{n=0}^{C-1} a^2(m, n)} \times \sqrt{\sum_{m=0}^{R-1} \sum_{n=0}^{C-1} a'^2(m, n)}}. \quad (11)$$

Here,  $a(m, n)$  and  $a'(m, n)$  represent the original and the extracted watermark images, respectively.

#### 4. Experimental results

Audio signals tested in all experiments are wav files with 16 bits/sample and 44.1 kHz sample rates. A  $24 \times 24$  bits binary image is used as a watermark and a 16 bits Baker code 1111100110101110 is used for synchronization. We selected a 1-level db2 lifting wavelet transform on the audio sections where  $SL = 128$  and  $WL = 256$ . The initial value of  $\delta$  is 0.005.

##### 4.1. Transparency test

To evaluate the perceived quality of the watermarked audio signals, both the original and the watermarked audio signals are tested based on the rules of the Mean Opinion Score (MOS) and the Subjective Difference Grades (SDG). Ten listeners did the test using the headset BOSE-QC-1. Five of them are professional musicians while the rest have long experience and knowledge of music. The watermarked audio scored 4.8 on MOS and  $-0.1$  on SDG, that is to say the embedded information can hardly be perceived by human auditory systems. Table 1 shows the grading scale of MOS and SDG.

**Table 1.** Grading Scale of MOS & SDG.

MOS	Grade	Description	SDG	Impairments
5.0	Excellent	Imperceptible	0.0	Inaudible
4.0	Good	Perceptible, but not annoying	-1.0	Audible, but not annoying
3.0	Fair	Slightly annoying	-2.0	Slightly annoying
2.0	Poor	Annoying	-3.0	Annoying
1.0	Bad	Very annoying	-4.0	Very annoying

The difference between the original signal and the watermarked signal in the time domain are shown in Fig. 3.

Apparently there is almost no difference between the original and the watermarked signal in the time domain.

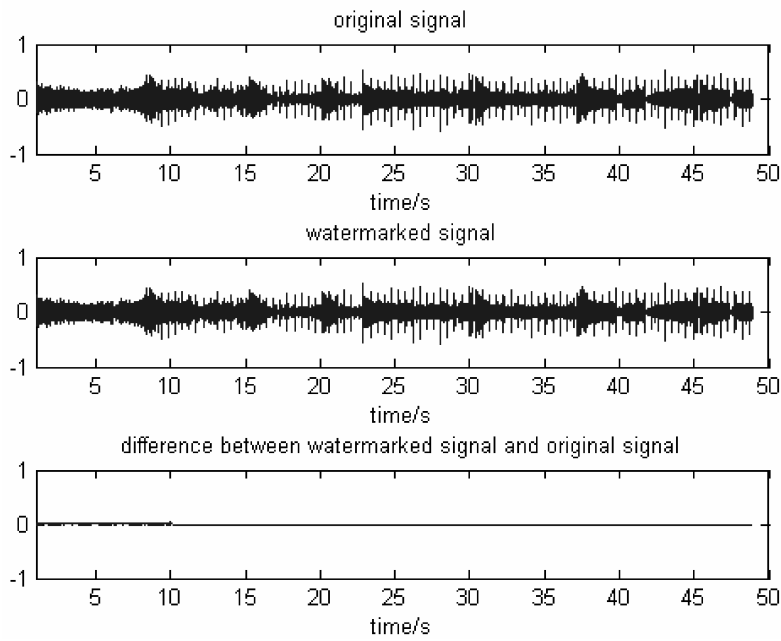
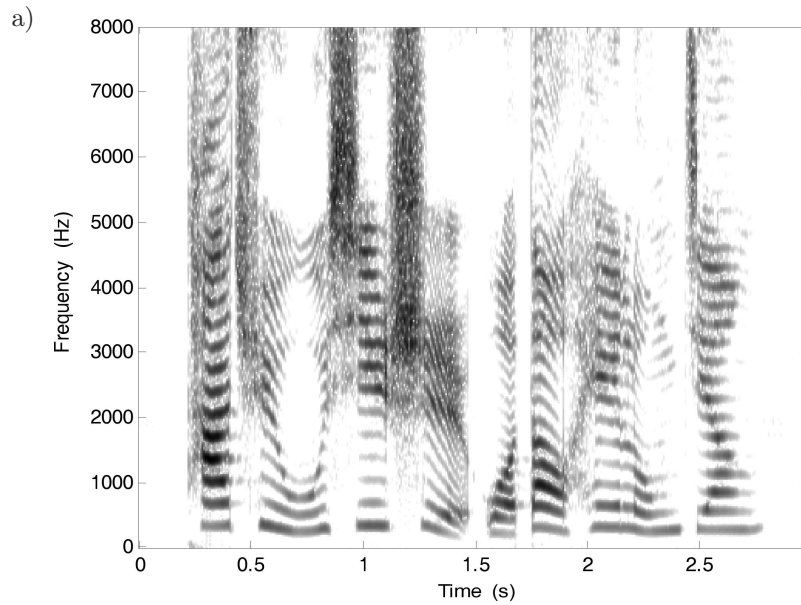


Fig. 3. Waveform of the original signal and the watermarked signal.

In addition to this, as shown in Fig. 4, we can see from the speech spectrograms that the differences in the frequency domain are also very small.



[Fig. 4a]

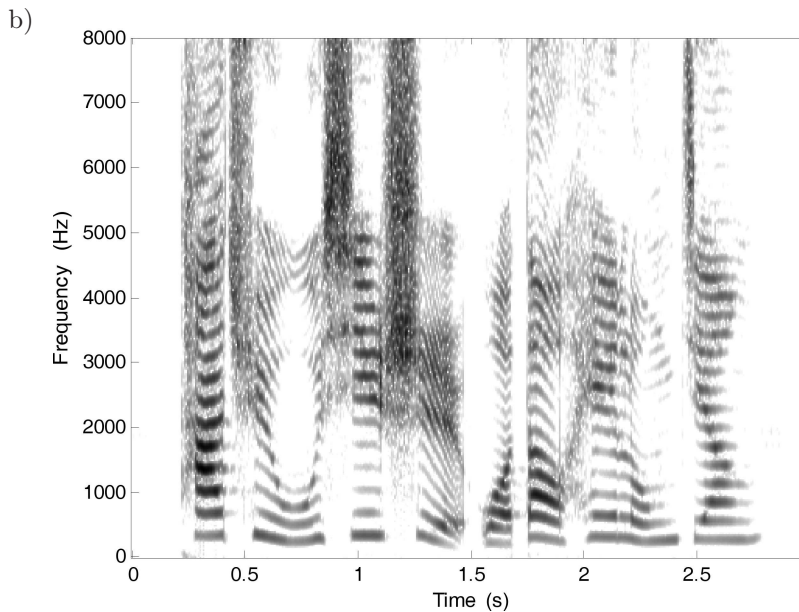


Fig. 4. a) Speech spectrogram of the original signal; b) speech spectrogram of the water-marked signal.

All tests above prove that the proposed algorithm has good transparency performance.

#### 4.2. Robustness test

##### 4.2.1. Robustness against common signal processing attacks

The watermarked signal is tested by exposing it to 5 kinds of attacks. (1) Low-pass filtering: application of a 9th-order Chebyshev filter with a cut-off frequency of 5 kHz; (2) Noise adding: addition of zero-mean white noise which variance is 0.01; (3) Re-quantization: re-quantization from 16-bit to 8-bit and then back to 16-bit; (4) Re-sampling: down-sampling to 22.05 kHz followed by up-sampling back to 44.1 kHz; (5) MP3 Compression: compression of the audio signal with a compression rate of 22:1, then decompression of the signal.

A performance comparison of the proposed scheme with Ref1 (WANG *et al.*, 2005) and Ref2 (WANG *et al.*, 2006) is shown in Fig. 5, where the X-label represents the type of attack and the Y-label represents the NC value of the extracted watermark.

Figure 5 shows that the method proposed in this paper can effectively resist the attacks of low-pass filtering, noise addition, re-sampling, re-quantization, lossy compression and so on. Furthermore, the method's robustness is also better than the previous two methods. The series of experimental results above sufficiently proves that the proposal in this paper can effectively resist the attacks of common signal processing.



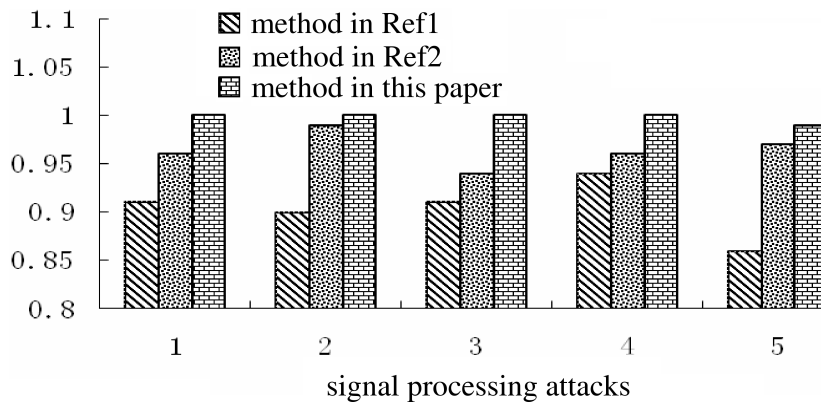


Fig. 5. Comparison of the robustness test.

4.2.2. Robustness against random cropping

The following random cropping steps are performed respectively on the watermarked audio signal: (1) Ten percent of the audio signal is cropped at one of three selected positions randomly (front, middle and back); (2) Jittering: cropping of one sample out of every 100, 500, 1000, 2000 samples; (3) Random cropping 1: selection of 5 positions randomly and removal of 100 samples at each position; (4) Random cropping 2: selection of 10 positions randomly and removal of 100 samples at each position; (5) Random cropping 3: selection of 10 positions randomly and removal of 500 samples at each position; (6) Random cropping 4: selection of 10 positions randomly and removal of 1000 samples at each position; The embedded watermark can still be extracted although the length of the watermarked signal has been reduced. Results are shown in Table 2.

Table 2. Robustness test against random cropping.

Attack	NC	Attack	NC
Add 10% (front)	1.000	Jittering 3	0.970
Add 10% (middle)	1.000	Jittering 4	0.986
Cropping 10%(front)	1.000	Random Cropping 1	0.997
Cropping 10%(middle)	1.000	Random Cropping 2	0.961
Jittering 1	0.823	Random Cropping 3	0.921
Jittering 2	0.942	Random Cropping 4	0.910

The kinds of attacks described in Table 2 will destroy the location information of watermark embedding. Furthermore, DFT and DWT algorithms are also unable to extract information successfully at this time, while the DCT algorithm can extract a little. This is usually because this algorithm brings the watermark

into the DC components of the audio signal in the DCT domain. The DC components of DCT actually represent parts of the statistical characteristics of audio, while the algorithm in this paper may better extract the watermark when it is synchronically attacked by random shearing and so on.

In this paper, we use the mean value of low-frequency coefficients of the lifting wavelet as statistical features because they represent the most important low-frequency components of the perceptual audio signal which are stable during general signal processing. This included low-pass filtering, MP3 compression and so on. In addition to this, the adjacent audio samples and the small audio clips show strong correlations. The proposed method will never cause the statistical mean value of one frame to change greatly even if we cut off few samples at random.

## 5. Conclusions

In this paper, we replace the classic wavelet transform with the lifting wavelet method so that the implementation efficiency of the algorithm is improved. Additionally, this watermark embedding method is based on invariant watermarking consideration. Combined with the psychoacoustic model, the embed strength is determined adaptively. Experiments show that the proposed scheme can resist common signal processing effectively. It is especially robust against random cropping. More attention will be focused on how to locate the synchronization code more efficiently. Although the method of this paper is proven to be robust against common signal processing attacks, it can still be made more robust against synchronous attacks such as TSM.

## Acknowledgments

The research is supported by: National Natural Science Foundations of China (Grant No. 60572076) and the University Natural Science Research Project of Jiangsu Province of China (Grant No. 09KJD510005).

## References

1. BAI J. *et al.* (2008), *SAR Image Denoising Based on Lifting Directionlet Domain Gaussian Scale Mixtures Model*, Chinese Journal of Computers, **31**, 7, 1234–1241.
2. DAUBECHIES I., SWELDENS W. (1998), *Factoring wavelet transforms into lifting's steps*, Journal of Fourier Analysis and Applications, **4**, 3, 245–267.
3. GAO S.W. *et al.* (2007), *Lifting wavelet transform and its application in digital watermarking*. Application Research of Computers, **24**, 6, 201–206.
4. QIANG Y., WANG Y. (2004), *A Survey of Wavelet-domain Based Digital Image Watermarking Algorithm*, Computer Engineering and Applications, **40**, 11, 46–50.

5. QU J.Y. *et al.* (2006), *Audio digital watermarking based on the lifting scheme wavelet transform*, *Computer & Digital Engineering*, **34**, 4, 91–94.
6. SWELDENS W. (1997), *The lifting scheme: a construction of second generation wavelets*, *SIAM Journal Mathematical Analysis*, **29**, 2, 511–546.
7. TAO Z. *et al.* (2006), *Audio watermarking based on psychoacoustic model and critical band wavelet transform*, *Acta Acustica*, **31**, 2, 114–119.
8. WANG X.Y. *et al.* (2005), *Content-based adaptive digital audio watermarking algorithm in wavelet domain*, *Mini-Micro Systems*, **26**, 8, 1354–1357.
9. WANG X.Y. *et al.* (2006), *A new adaptive digital audio watermarking algorithm*, *Mini-Micro Systems*, **27**, 7, 1353–1357.
10. WANG R.D., XU D.W. (2006), *Multiple audio watermarks based on lifting wavelet transform*, *Journal of Electronics & Information Technology*, **28**, 10, 1820–1826.
11. WEI L. (2004), *Research on Robust Digital Audio Watermarking*, PhD thesis, Fudan University, Shanghai.
12. WIM S. (1996), *The Lifting Scheme: A Custom-Design Construction of Biorthogonal Wavelets*. *Applied and Computational Harmonic Analysis*, **3**, 15, 186–200.
13. XIANG S.J. (2006), *Robust Audio Watermarking Algorithms*, PhD thesis, Sun Yat-sen University, Guangzhou.
14. XU D.W., WANG R.D. (2006), *A blind audio watermarking algorithm based on convolutional codes*, *Computer Applications*, **26**, 7, 1649–1651.