Computers and Electrical Engineering 35 (2009) 506–516

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Chaos-based discrete fractional Sine transform domain audio watermarking scheme

Mingquan Fan *, Hongxia Wang

School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031, PR China

ARTICLE INFO

Article history:
Received 29 March 2008
Received in revised form 15 December 2008
Accepted 18 December 2008
Available online 24 February 2009

Keywords:
Audio watermarking
Discrete fractional Sine transform
Chaotic sequence
Signal processing attack

ABSTRACT

We proposed a novel discrete fractional Sine transform (DFRST) based watermarking scheme for audio data copyright protection. Chaotic sequences were adopted to improve the security of the proposed watermarking scheme. Simulations under various conditions were given to verify the effectiveness of the audio watermarking scheme. The results show the proposed scheme is secure, and the watermark is imperceptible and robust against various audio signal processing attacks.

1. Introduction

With the rapid development of network and multimedia technique, effortless distribution of e-works has been achieved. However, on the other hand, a large number of authors' and publishers' intellectual property copyrights have suffered from violation, which led to huge damage of their benefits in many applications. Thus people pay more attention to copyright management and protection nowadays. Embedding a certain form of watermark into multimedia data is considered as a potential solution.

Digital watermarking can be classified into two categories. One is in the temporal domain; the other is in the transform domain. Amplitude modification [1] and echo hiding method [2] are the representative temporal domain schemes. As to transform domain schemes, the host signals are often transformed to transform domains; then the watermarks are embedded into coefficients of transform domains; finally inverse transforms are implemented on the modified coefficients to get watermarked signals. Commonly used transforms include discrete Fourier transform (DFT) [3], discrete Cosine transform (DCT) [4], discrete Wavelet transform (DWT) [5], etc. Besides, for the secret property of fractional transforms is suitable to secure watermarking schemes, some researchers have proposed several image watermarking schemes based on discrete fractional Fourier transform (DFRFT) [6], discrete fractional order random transform (DFRNT) [7]. However, the fractional transform based audio watermarking scheme has not been reported until now. As is known to the world, the human auditory system is more sensitive than the visual system, and dealing with audio signal is much more difficult than image. In this paper, we propose a robust and secure blind audio watermarking scheme based on discrete fractional Sine transform (DFRST). Series of problems are resolved and we successfully employ DFRST for audio data copyright protection.

The remainder of this paper is organized as follows. The watermark embedding algorithm and extraction process are described in Sections 2 and 3, respectively. Performance analysis of the proposed scheme is elaborated in Section 4. Experimental results are stated in Section 5. Finally, conclusions are given in Section 6.

* Corresponding author.
E-mail address: mqfan_sc@163.com (M. Fan).

0045-7906/$ - see front matter
doi:10.1016/j.compeleceng.2008.12.004

2. Embedding algorithm

In this section, we first review the DFRST defined in [8]. The development of DFRST is based on the DFRFT. The eigenvector $\widetilde{v}_k$ (k is odd) is assigned to the eigenvalue $e^{-j(k-1)\alpha}$. Thus, the N-point DFRST kernel is defined as:

$$S_{N,\alpha} = \widetilde{V}_N \widetilde{D}_N^{2\alpha/\pi} \widetilde{V}_N^T = \widetilde{V}_N \begin{bmatrix} 1 & & 0 \\ & e^{-2j\alpha} & \\ & \ddots & \\ 0 & & e^{-j2(N-1)\alpha} \end{bmatrix} \widetilde{V}_N^t \tag{1}$$

where $\widetilde{V}_N = [\widetilde{v}_1 | \widetilde{v}_3 | \cdots | \widetilde{v}_{2N-1}]$. $\widetilde{v}_k$ is the DST-I (2) (discrete Sine transform, DST) eigenvector obtained from the kth-order DFT Hermite eigenvector. The above DFRST kernel matrix will be reduced to a DST-I kernel matrix for $\alpha = \pi/2$, and it will become an identity matrix for $\alpha = 0$. The period of α is π, seen in Eq. (5).

$$S^{\mathrm{I}}_{\mathrm{N-1}} = \sqrt{\frac{2}{N}}\left[\sin\left(\frac{mn\pi}{N}\right)\right], \qquad m,n = 1,2,\ldots,N-1. \tag{2}$$

The DFRST has properties of unitarity, angle additivity, periodicity and symmetric, which are corresponding to Eqs. (3)–(6), respectively.

$$S^{*}_{N,\alpha} = S^{-1}_{N,\alpha} = S_{N,-\alpha} \tag{3}$$

$$S_{N,\alpha} S_{N,\beta} = S_{N,\alpha+\beta} \tag{4}$$

$$S_{N,\alpha+\pi} = S_{N,\alpha} \tag{5}$$

$$S_{N,\alpha}(a,b) = S_{N,\alpha}(b,a) \tag{6}$$

From Eq. (3), it is obvious that the inverse transform ($S^{-1}_{N,\alpha}$) of $S_{N,\alpha}$ is the plural form. Different from image data, the audio signal is a real sequence with normalized amplitudes belonging to $(-1, 1)$, thus we have the following proposition for dealing with audio signal based on DFRST.

Proposition 1. *In order to obtain real watermarked audio signal, DFRST should be implemented on coefficients of some orthogonal transform, such as fast Fourier transform (FFT).*

The watermark embedding process is illustrated in Fig. 1. Suppose $A = \{A(i) | 1 \leqslant i \leqslant L\}$ is the original audio signal, and $W = \{W(m,n) | 1 \leqslant m \leqslant M, 1 \leqslant n \leqslant N\}$ represents the watermark (binary image). Details of embedding are elaborated as following:

Step 1. Segmenting: First, the original audio signal A is split into L_1 segments, which are denoted as $A_s = \{A_s(g,l) | 1 \leqslant g \leqslant L_1, 1 \leqslant l \leqslant L/L_1, L_1 = M \times N\}$.
Step 2. FFT: FFT is performed on each segment, and $A_s' = \{A_s'(g, l)\}$ represents FFT domain coefficients of all segments.
Step 3. Selecting middle frequency coefficients: In order to obtain good imperceptibility, selecting N_u consecutive middle frequency coefficients of each segment as the dataset for watermark embedding, denoted as $C_w = \{C_w(g,h) | 1 \leqslant g \leqslant L_1, 1 \leqslant h \leqslant N_u\}$.
Step 4. DFRST: Based on Logistic map (7), and adopting key K_1 as the initial value $x(1)$, generating chaotic sequence $X = \{0 < x(g) < 1 | 1 \leqslant g \leqslant L_1\}$ to select specific angle α for each segment.

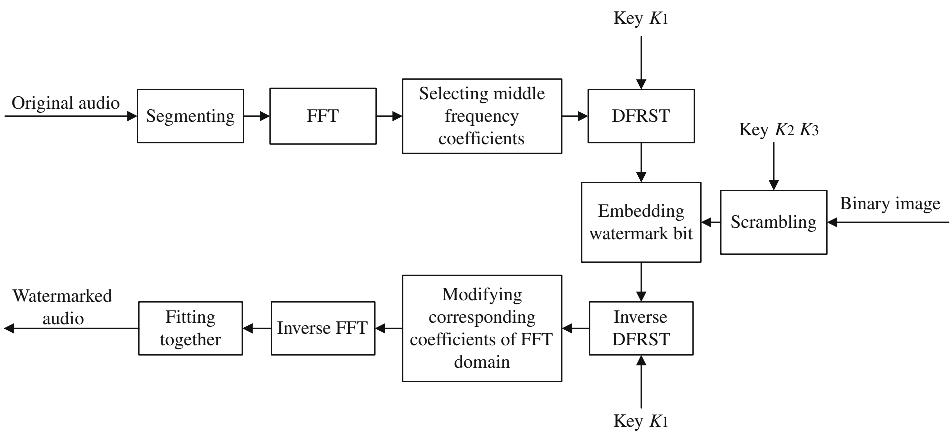

Fig. 1. The watermark embedding algorithm.

$$x(g_1+1) = a \times x(g_1) \times (1 - x(g_1)) \tag{7}$$

where, a is the system parameter. For $3.57 < a \leqslant 4$, the sequence is non-periodic, non-convergent, and very sensitive to initial value. $1 \leqslant g_1 \leqslant L_1 - 1$.Then, DFRST is performed on each $C_w(g)$ as follows:

$$C'_w(g) = \mathrm{DFRST}(C_w(g), \alpha = x(g) \times T) \tag{8}$$

where T is the period of α. Here, $T = \pi$.

Step 5. Scrambling binary image: Similarly, based on Logistic map (9) and (10), and adopting key K_2, K_3 as the initial values $y(1)$, $z(1)$, generating two chaotic sequences $Y = \{y(t_1)|1 \leqslant t_1 \leqslant L_2, L_2 \gg M\}$, $Z = \{z(t_2)|1 \leqslant t_2 \leqslant L_3, L_3 \gg N\}$.

$$y(t_1+1) = a \times y(t_1) \times (1 - y(t_1)) \tag{9}$$
$$z(t_2+1) = a \times z(t_2) \times (1 - z(t_2)) \tag{10}$$

Following Eqs. (11) and (12), and generating chaotic sequences $Y_1 = \{1 \leqslant y_1(m) \leqslant M\}, Z_1 = \{1 \leqslant z_1(n) \leqslant N\}$ from Y' and Z', $1 \leqslant m \leqslant M$, $1 \leqslant n \leqslant N$, $\forall y_1(m_1)$, $\forall y_1(m_2)$, $\forall z_1(n_1)$, $\forall z_1(n_2)$, $m_1 \neq m_2$, $n_1 \neq n_2$, $y_1(m_1) \neq y_1(m_2)$, $z_1(n_1) \neq z_1(n_2)$.

$$Y' = \lceil M \times Y \rceil \tag{11}$$
$$Z' = \lceil N \times Z \rceil \tag{12}$$

where $\lceil \cdot \rceil$ is the ceil function.Then the scrambled binary image $W_s(m, n) = W(y_1(m), z_1(n))$, and W_s is converted into one-dimensional sequence W_1 according to the row scanning method.

Step 6. Embedding watermark bit: The element with largest amplitude of each $C_w'(g)$ is selected to embed one watermark bit. Here, we have our second proposition.

Proposition 2. *Based on the phase modulation technique, the robustness of watermark is maximized if the element with the largest amplitude is selected to embed the watermark bit.*

Proof. Suppose the amplitude of element used for embedding watermark bit is ρ_0, and φ is its phase, seen in Fig. 2. After various attacks, such as noise adding, etc., the amplitude of the element is changed to ρ, and its phase is increased δ. From Fig. 2, we get

$$\rho \cos \delta - \rho_0 = \Delta_1 \tag{13}$$
$$\rho \sin \delta = \Delta_2 \tag{14}$$

where Δ_1 and Δ_2 are supposed to be two one-dimensional noises with some distribution. From Eqs. (13) and (14), we get

$$\tan \delta = \frac{\Delta_2}{\Delta_1 + \rho_0} \tag{15}$$
$$\rho^2 = (\rho_0 + \Delta_1)^2 + \Delta_2^2 \tag{16}$$

Generally, the watermarked audio can not be seriously attacked for practical value, so Eqs. (17) and (18) are reasonable. Otherwise, the watermarked audio would be very annoying.

$$\rho_0 > |\Delta_1| \tag{17}$$
$$\rho_0 > |\Delta_2| \tag{18}$$

From Eqs. (15)–(18), we get this rule, that is, under the condition of determinate Δ_1 and Δ_2, if ρ_0 increases, then the modified amplitude ρ will also increase, and $|\delta|$ will decrease. This rule implies that the phase of the element with largest

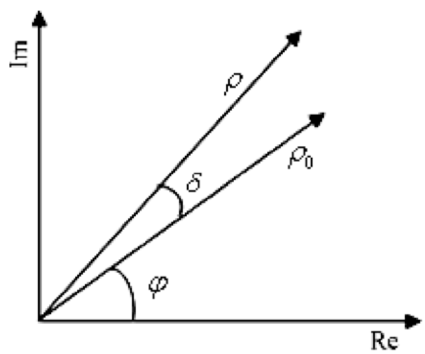

Fig. 2. Sketch map of modified element used for embedding watermark bit.

amplitude is most stable for various attacks. So we may make a conclusion that the element with largest amplitude used for embedding watermark bit based on phase modulation technique can get higher robustness. So Proposition 2 is correct. □

Phase modulation technique is a mature technique [9,10], and in our proposed scheme, the largest spectral peak of DFRST domain of each segment is subject to quantization index modulation [11,12] in the phase. Suppose the largest amplitude is A_m and the corresponding phase is θ $(0 \leqslant \theta < 2\pi)$, then we embed one watermark bit as follows:

$$\theta' = \begin{cases} (temp + 0.5) \times \text{ step}, \text{ if } w = 0 \text{ and mod } (temp, 2) = 0 \\ (temp - 0.5) \times \text{ step}, \text{ if } w = 0 \text{ and mod } (temp, 2) = 1 \\ \qquad \text{and } \theta < (temp + 0.5) \times step \\ (temp + 1.5) \times \text{ step}, \text{ if } w = 0 \text{and mod}(temp, 2) = 1 \\ \qquad \text{and } \theta \geqslant (temp + 0.5) \times step \\ (temp + 0.5) \times \text{ step}, \text{ if } w = 1 \text{and mod}(temp, 2) = 1 \\ (temp - 0.5) \times \text{ step}, \text{ if } w = 1 \text{and mod}(temp, 2) = 0 \\ \qquad \text{and } \theta < (temp + 0.5) \times step \\ (temp + 1.5) \times \text{ step}, \text{ if } w = 1 \text{and mod}(temp, 2) = 0 \\ \qquad \text{and } \theta \geqslant (temp + 0.5) \times step \end{cases} \tag{19}$$

$$A'_m = A_m + \lambda \tag{20}$$

where λ is a positive real number, which is used to further separate the watermarked coefficient from other coefficients. $temp = \lfloor \theta/step \rfloor$. $\lfloor\cdot\rfloor$ is the floor function. *step* is the odd-even quantization [12] step. The robustness of the watermark is improved as *step* increases. However, a larger *step* causes higher distortion to the signal. So there is a trade-off between robustness and imperceptibility in choosing the size of *step*. w is the embedded one watermark bit.

Step 7. Inverse DFRST: Under the direction of the same key K_1, performing inverse DFRST on each watermarked $C_w'(g)$.

Step 8. Modifying the corresponding coefficients of FFT domain: According to Eq. (21), modifying the corresponding coefficients of FFT domain.

$$F(k) = F^*(N' - k) \tag{21}$$

where $N' = L/L_1$.

Step 9. Inverse FFT: Inverse FFT is performed on each segment with modified FFT domain coefficients.

Step 10. Fitting together: Combining each watermarked audio segment together to form the final watermarked audio signal A'.

3. Extraction process

The extraction process does not need the original host audio signal, and it is almost the reverse of the embedding process. The overall flowchart is shown in Fig. 3.

After DFRST, we select the element with largest amplitude of DFRST domain of each segment, and phases of all the elements are denoted as $\theta^w = \{0 \leqslant \theta^w(g) < 2\pi | 1 \leqslant g \leqslant L_1\}$, then we extract watermark bits as following:

$$W_e(g) = \begin{cases} 0 & \text{if mod}(\lfloor \theta^w(g)/step \rfloor, 2) = 0 \\ 1 & \text{if mod}(\lfloor \theta^w(g)/step \rfloor, 2) = 1 \end{cases} \tag{22}$$

According to the row scanning method, W_e is converted into two-dimensional binary image W_e'. Then using key K_2 and K_3 to generate the same chaotic sequences Y_1, Z_1 with embedding process, and the final robust watermark $W'(y_1(m), z_1(n)) = W'_e(m, n)$.

In addition, in order to dispel the influence of subjective and objective factors, the normalized cross correlation (NC) [13] is adopted to assess the similarity between the extracted watermark and the original one. The NC is defined as:

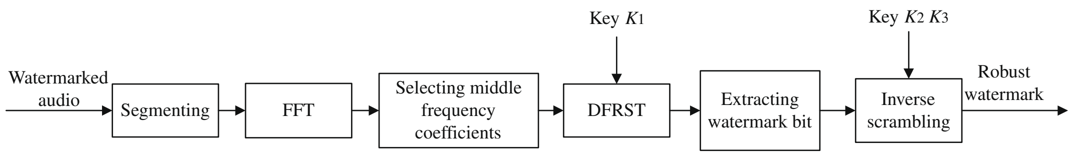

Fig. 3. The watermark extraction algorithm.

$$\mathrm{NC}(W', W) = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N} w(m,n) \times w'(m,n)}{\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{N} w^2(m,n)}\sqrt{\sum_{m=1}^{M}\sum_{n=1}^{N} w'^2(m,n)}} \tag{23}$$

where W and W' are the original watermark and the extracted watermark, respectively.

In this correspondence, bit error rate (BER) is used to measure the reliability. Its definition is shown as following:

$$\mathrm{BER} = \frac{E}{M \times N} \times 100\% \tag{24}$$

where E is the number of erroneously detected bits.

Besides, the signal-to-noise ratio (SNR) is used to serve as an objective measurement of audio quality. It is defined as follows:

$$\mathrm{SNR}(A, A') = 10\log_{10}\left(\frac{\sum_{i=1}^{L} A^2(i)}{\sum_{i=1}^{L}(A(i) - A'(i))^2}\right) \tag{25}$$

where A and A' are the original and the watermarked audio signal, respectively.

4. Performance analysis

In this Section, we evaluate the performance of our proposed watermarking scheme. The watermark performance, such as embedding capacity, false alarm and false rejection, is investigated.

4.1. Embedding capacity

Suppose that the sampling rate of the audio signal is f_s (Hz), and the number of samples of each segment is N'. The embedding capacity P of the proposed scheme can be expressed as:

$$P = \frac{f_s}{N'} \tag{26}$$

where the unit of embedding capacity P is bit/s. The embedding capacity is improved as N' decreases. However, a less N' causes higher distortion.

4.2. False alarm analysis

False alarm is the probability of declaring an unwatermarked audio as watermarked by decoder. The watermarking system is better with less false alarm probability.

Suppose that for an unwatermarked audio segment, the correctly extracted bit is assumed as an independent random variable with probability of p_1. Let q be the total watermark bits, and r be the number of matching bits. Then based on Bernoulli trials assumption, we get

$$p_r = C_q^r (p_1)^r (1 - p_1)^{q-r} \tag{27}$$

The audio is claimed watermarked if the number of matching bits is greater than a threshold Th_1. Then the probability of the cases that $r \geqslant Th_1$ is the false alarm error probability. It is defined as:

$$P_{fa} = \sum_{r=Th_1}^{q} p_r \tag{28}$$

From Eqs. (27) and (28), we get

$$P_{fa} = \sum_{r=Th_1}^{q} \left\{ C_q^r (p_1)^r (1 - p_1)^{q-r} \right\} \tag{29}$$

Ideally, p_1 is assumed to be 1/2, and $Th_1 = \lceil (1 - \mathrm{BER}) \times q \rceil$. If BER is set at 20%, then P_{fa} may be described as following:

$$P_{fa} = 2^{-q} \sum_{r=\lceil 0.8q \rceil}^{q} C_q^r \tag{30}$$

Fig. 4 gives the false alarm probabilities when q belongs to (0, 100], and it tells us that the false alarm probability trends to 0 when q is bigger than 20. In our proposed scheme, q = 4096, so the false alarm probability is approximately equal to 0.

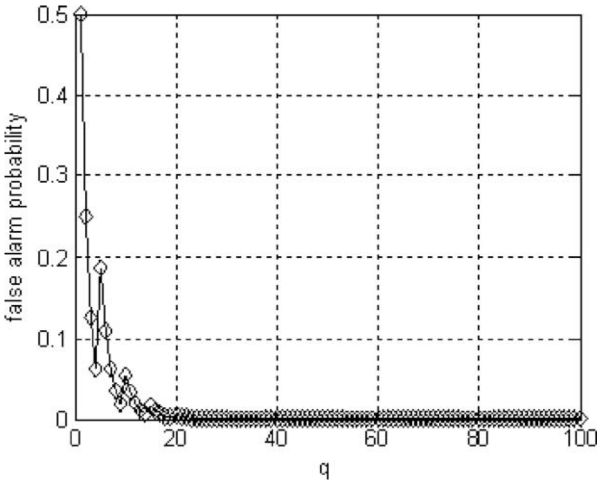

Fig. 4. False alarm probabilities under various q.

4.3. False rejection analysis

False rejection is the probability of declaring a watermarked audio as unwatermarked by decoder. The watermarking system is better with less false rejection probability.

Table 1
Watermark detection results against various common signal processing attacks.

Attack	No attack	Noise adding	Lowpass filtering	Echo
Watermark				
NC	1	0.9729	0.9762	1
BER	0	0.0422	0.0371	0
Attack	Resampling (11025 Hz)	Resampling (22050 Hz)	Resampling (88200 Hz)	Resampling (176400 Hz)
Watermark				
NC	1	1	1	1
BER	0	0	0	0
Attack	Reverse amplitude	Expanding (6.0206 dB)	Expanding (−6.0206 dB)	MP3 (128 kbps)
Watermark				
NC	1	0.9997	0.9975	1
BER	0	4.8828e−004	0.0039	0
Attack	MP3 (112kbps)	MP3 (96 kbps)	MP3 (80 kbps)	MP3 (64 kbps)
Watermark				
NC	1	1	0.9994	0.9974
BER	0	0	9.7656e−004	0.0042
Attack	MP3 (56 kbps)	MP3 (48 kbps)	Reverberation(1s,−24 dB)	Smoothness filtering
Watermark				
NC	0.9938	0.9778	0.9109	1
BER	0.0098	0.0347	0.1350	0

Table 2
Watermark detection results under various λ and *step*.

Attack type	BER					
	$\lambda = 0.05$			$\lambda = 0.15$		
	step = $\pi/8$	*step* = $\pi/7$	*step* = $\pi/6$	*step* = $\pi/8$	*step* = $\pi/7$	*step* = $\pi/6$
No attack	0	0	0	0	0	0
Noise adding	0.0422	0.0251	0.0110	0	0	0
Lowpass filtering	0.0371	0.0017	0	0.0269	4.8828e−004	0
Echo	0	0	0	0	0	0
Resampling (11025 Hz)	0	0	0	0	0	0
Resampling (22050 Hz)	0	0	0	0	0	0
Resampling (88200 Hz)	0	0	0	0	0	0
Resampling (176400 Hz)	0	0	0	0	0	0
Reverse amplitude	0	0	0	0	0	0
Expanding (6.0206 dB)	4.8828e−004	4.8828e−004	0	0	0	0
Expanding (−6.0206 dB)	0.0039	0.0037	0.0027	4.8828e−004	7.3242e−004	9.7656e−004
MP3 (128 kbps)	0	0	0	0	0	0
MP3 (112 kbps)	0	0	0	0	0	0
MP3 (96 kbps)	0	0	0	0	0	0
MP3 (80 kbps)	9.7656e−004	7.3242e−004	0	4.8828e−004	0	2.4414e−004
MP3 (64 kbps)	0.0042	0.0015	9.7656e−004	9.7656e−004	9.7656e−004	4.8828e−004
MP3 (56 kbps)	0.0098	0.0061	0.0039	0.0017	0.0012	0.0012
MP3 (48 kbps)	0.0347	0.0227	0.0142	0.0063	0.0037	0.0032
Reverberation (1s,−24 dB)	0.1350	0.1077	0.0947	0.0391	0.0286	0.0203
Smoothness filtering	0	0	0	0	0	0

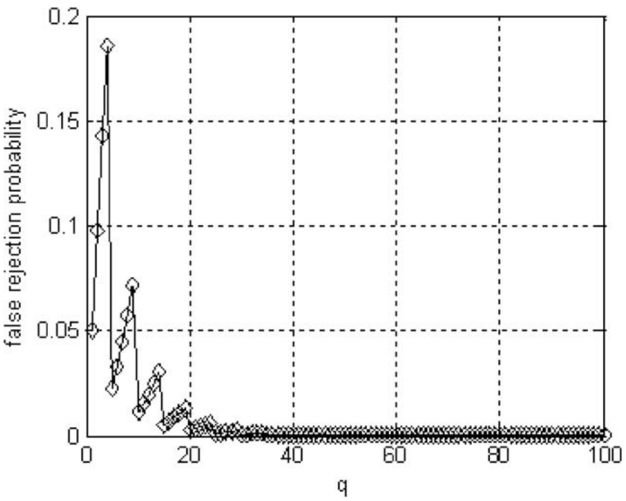

Fig. 5. False rejection probabilities under various q.

Similarly, suppose that for a watermarked audio segment, the correctly extracted bit is assumed as an independent random variable with probability of p_2. Let q be the total watermark bits, and t be the number of matching bits. Then based on Bernoulli trials assumption, we get

$$p_t = C_q^t (p_2)^t (1 - p_2)^{q-t} \tag{31}$$

The audio is claimed unwatermarked if the number of matching bits is less than a threshold Th_2. Then the probability of the cases that $t \leqslant Th_2$ is the false rejection error probability. It is defined as:

$$P_{fr} = \sum_{t=0}^{Th_2} p_t \tag{32}$$

From Eqs. (31) and (32), we get

$$P_{fr} = \sum_{t=0}^{Th_2} \{C_q^t (p_2)^t (1 - p_2)^{q-t}\} \tag{33}$$

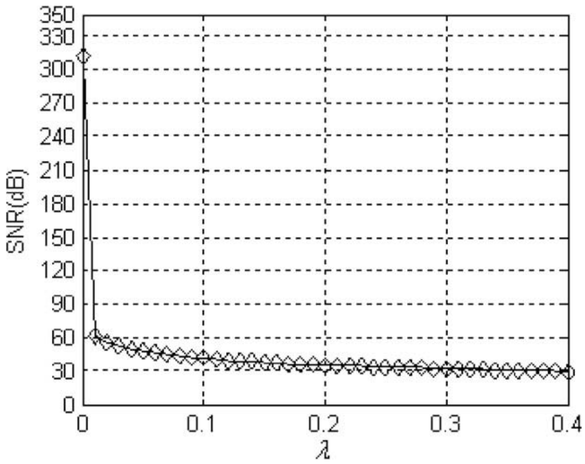

Fig. 6. Sketch map of selecting suitable range for λ.

Here, $Th_2 = \lceil(1 - \text{BER}) \times q\rceil - 1$. If BER is also set at 20%, then P_{fr} may be described as following:

$$P_{fr} = \sum_{t=0}^{\lceil 0.8q\rceil - 1} \left\{ C_q^t (p_2)^t (1 - p_2)^{q-t} \right\} \tag{34}$$

Different from p_1, here p_2 can not be assumed to be 1/2. Corresponding to different attack, p_2 has different value. However, the approximate value of p_2 may be obtained from bit error rate under determinate attack. From Table 2 (Table 2 is shown in the next section), it is easily known that the BERs are all less than 0.05 except reverberation ($\lambda = 0.05$), so p_2 is assumed to be 0.95 in our proposed scheme. Fig. 5 gives the false rejection probabilities when q belongs to (0, 100], and it tells us that the false rejection probability trends to 0 when q is bigger than 20. In our proposed scheme, q = 4096, so the false rejection probability of our proposed scheme is also approximately equal to 0.

5. Experimental results

In order to illustrate the inaudible and robust nature of our proposed watermarking scheme, several audio pieces are used to verify the truth. They are captured from different kinds of music, including bagpipe music, classical music, country music, dance music, jazz music and rock music. All of the audio signals are music with 16 bit signed mono audio signals sampled at 44.1 kHz (WAVE format). We use a 64×64 binary image as our robust watermark for all audio signals. In our experiments, each audio signal includes 2097152 samples, and eight coefficients between the 49th and 58th of FFT domain of each segment are selected to embed one watermark bit. That is, N_u = 8. So the embedding capacity of the scheme is about 86 bit/s.

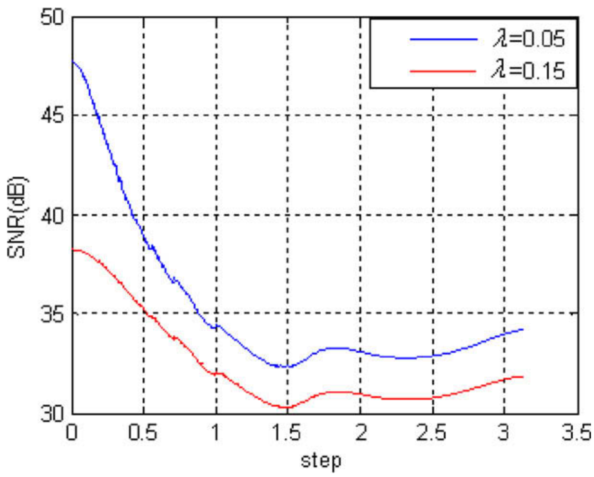

Fig. 7. Sketch map of selecting suitable range for *step.*

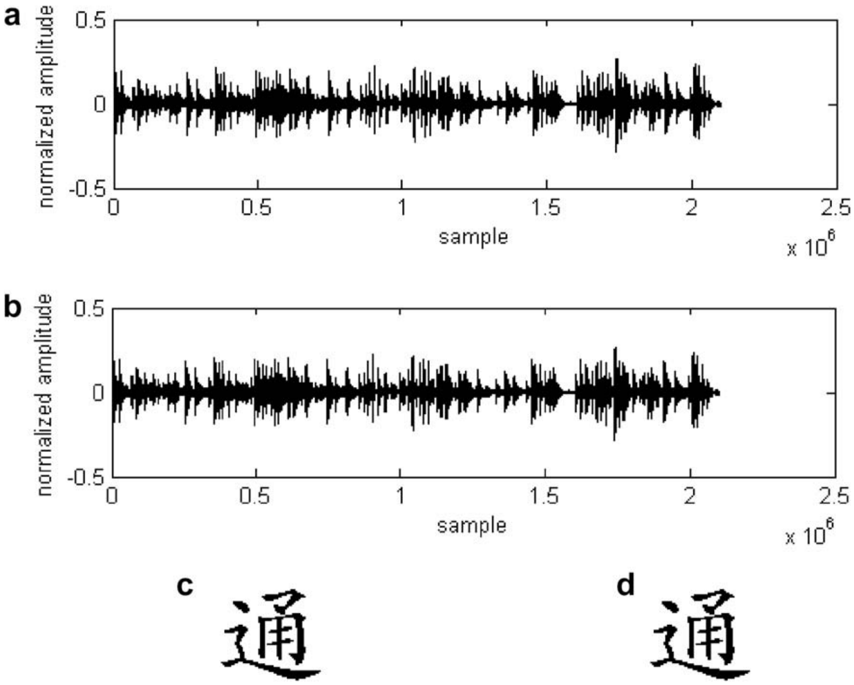

Fig. 8. Results of watermark embedding and detecting. (a) Original audio signal. (b) Watermarked audio signal (SNR = 40.7034 dB). (c) Original watermark. (d) Extracted watermark (BER = 0, NC = 1).

5.1. Embedding distortion

In the embedding, the distortion due to the watermark relies on embedding capacity P, parameter λ, and odd-even quantization step *step*. At any given P, λ and *step* should be kept small enough so that the watermark is imperceptible, but as big as permitted so the watermark is maximally robust.

5.1.1. Experiments for suitable range of λ

Without embedding the watermark, the element with largest amplitude of DFRST domain of each audio segment is increased. The distortion of the modified audio signal is computed and shown in Fig. 6, where, the X axis means the modified value λ, and the Y axis represents the SNR of the modified audio signal.

From Fig. 6, it is easily known that the suitable range of λ is (0, 0.21] when adopting 35 dB as the threshold value of SNR.

5.1.2. Experiments for suitable range of step

The suitable range of *step* is correlative with λ, that is, the suitable ranges of *step* are different corresponding to various values of λ. Fig. 7 shows the sketch map of selecting suitable range for *step*, where λ adopts 0.05 and 0.15, respectively. From Fig. 7, we can see that the suitable range of *step* is (0, 0.884] for $\lambda = 0.05$, and (0, 0.525] for $\lambda = 0.15$.

A plot of a short portion of the original audio signal is shown in Fig. 8a, and its corresponding watermarked audio signal is shown in Fig. 8b (SNR = 40.7034 dB, $\lambda = 0.05$, *step* = $\pi/8$). The original watermark image is displayed in Fig. 8c, and the extracted watermark image without being attacked is displayed in Fig. 8d (BER = 0, NC = 1).

5.2. Robustness against common signal processing attacks

In order to evaluate the robust nature of the proposed scheme, the attacks including MP3 compression, resampling, noise adding, low-pass filtering, etc., are used to estimate the robustness of our scheme. Table 1 summarizes the watermark detection results against various common signal processing attacks ($\lambda = 0.05$, *step* = $\pi/8$). Table 2 gives the detection results when λ and *step* adopt various values.

Experimental results show that our audio watermarking scheme is not only inaudible, but also robust against various common signal processing attacks, such as noise adding, resampling, MP3 compression, etc.

Noise adding. Adding Gaussian noise with 65 dB SNR.
Lowpass filtering. The lowpass filter with cutoff frequency 20 kHz.
Echo. Adding echo with delay 100 ms and decay 50%.

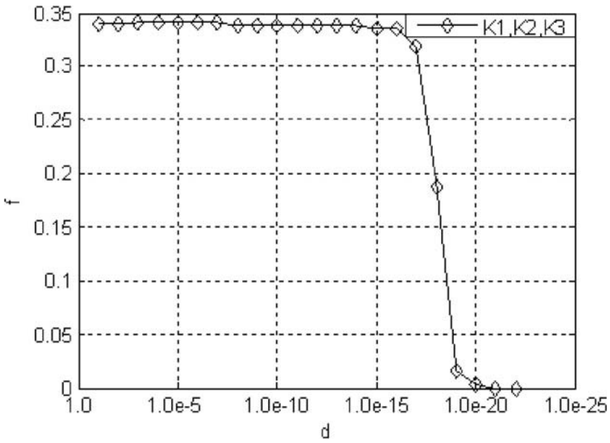

Fig. 9. Sketch map of key space.

Resampling. In this experiment, the original audio signal is sampled with a sampling rate of 44.1 kHz. Watermarked audio signal is down-sampled to 11.025 kHz, 22.05 kHz, and then up-sampled back to 44.1 kHz; up-sampled to 176.4 kHz, 88.2 kHz, and then down-sampled back to 44.1 kHz.

Reverse amplitude. Using GoldWave to reverse the plus or minus of amplitude of samples.

Expanding. Using GoldWave to expand the watermarked audio signal with increment of 6.0206 dB and -6.0206 dB, respectively.

MPEG compression. The coding/decoding is performed using a software implementation of the ISO/MPEG-1 Audio Layer III coder with several different bit rates (128 kbps, 112 kbps, 96 kbps, 80 kbps, 64 kbps, 56 kbps, 48 kbps).

Reverberation. Using GoldWave to reverberate the watermarked audio signal with reverberation time 1s and volume -24 dB.

Smoothness filtering. Using GoldWave to smoothly filter the watermarked audio signal.

5.3. Security analysis

According to Kerckhoff's principle, the security of information system relies on keys instead of privacy of scheme. In our proposed audio watermarking scheme, we use key K_1, K_2, K_3 to generate chaotic sequences for enhancing the security of the proposed scheme. Hence, the size of key value space influences the security of the proposed scheme. For key K_1, K_2, K_3 are all used as the initial value of Logistic map, we take K_1 for example and compute its key value space as following.

Suppose $K_1 = \{0 < K_1(i) < 1 | i = 1, 2, \cdots, Le\}$, Le is an integer which is large enough, generate chaotic sequences $X = \{x(i,j) | i = 1, 2, \cdots, Le, j = 1, 2, \cdots, Le1\}$, where Le represents the number of chaotic sequences and $Le1$ means the length of each chaotic sequence. When $K_1' = \{0 < K_1(i) + d < 1 | i = 1, 2, \cdots, Le\}$, generate another group of chaotic sequences $X' = \{x'(i,j) | i = 1, 2, \cdots, Le, j = 1, 2, \cdots, Le1\}$. Utilize function $f = T(d)$ to test key space of K_1.

$$f = T(d) = \frac{\sum_{i=1}^{Le}\sum_{j=1}^{Le1} |x(i,j) - x'(i,j)|}{Le1 \times Le} \tag{35}$$

Fig. 9 gives the cure of function $f = T(d)$. From the figure, it is known that f is approximately equal to 0 when $d_0 = 10^{-19}$. So the key space of K_1 is $1/d_0 = 10^{19}$. Similarly, the key spaces of K_2, K_3 can be computed, and they are same with K_1. So the key space of the whole watermarking system is 10^{57}. Enough large key space ensures the high security of the proposed watermarking system.

6. Conclusions

In this correspondence, we propose a novel robust and secure blind digital audio watermarking scheme based on discrete fractional Sine transform (DFRST). The characteristic of DFRST gives possibility for constructing secure watermarking scheme, therefore, chaotic sequences are extensively adopted in our proposed scheme for enhancing security. Meanwhile, two propositions are given for DFRST based audio watermarking scheme. The experimental results have illustrated the inaudible and robust nature of our watermarking scheme. The easy operational proposed scheme is practicable for audio data copyright protection.

Despite the success of the proposed audio watermarking scheme, it also has a drawback, that is, the proposed scheme is not robust against random cropping and time scale modification. Therefore, future research will focus on overcoming this problem, moreover, psychoacoustic model may be adopted to improve the imperceptibility of our scheme. Besides, professional subjective listening test will be conducted and used to evaluate the imperceptibility of watermarked audio signal in our future research.

Acknowledgements

The authors appreciate the Handling Editor and anonymous reviewers for their comments in improving this paper.

This work was supported by National Natural Science Foundation of China (NSFC) under Grant No. 60702025, the Research Foundation for Doctoral Program of Higher Education (RFDP) under Grant No.20070613024, Sichuan Youth Science and Technology Foundation of China under Grant No. 07ZQ026-004.

References

[1] Bassia Paraskevi, Pitas Ioannis, Nikolaidis Nikos. Robust audio watermarking in the time domain. IEEE Trans Multimedia 2001;3(2):232–41.
[2] Chen Oscal T-C, Wu Wen-Chih. Highly robust, secure, and perceptual-quality echo hiding scheme. IEEE Trans Audio, Speech, Language Process 2008;16(3):629–38.
[3] Solachidis Vassilios, Pitas Ioannis. Circularly symmetric watermark embedding in 2-D DFT domain. IEEE Trans Image Process 2001;10(11):1741–53.
[4] Huang Jiwu, Shi Yun Q, Shi Yi. Embedding image watermarks in DC components. IEEE Trans Circ Syst Video Technol 2000;10(6):974–9.
[5] Wang Xiangyang, Zhao Hong. A novel synchronization invariant audio watermarking scheme based on DWT and DCT. IEEE Trans Signal Process 2006;54(12):4835–40.
[6] Djurovic Igor, Stankovic Srdjan, Pitas Ioannis. Digital watermarking in the fractional Fourier transformation domain. J Network Comput Appl 2001;24:167–73.
[7] Guo Jun, Liu Zhengjun, Liu Shutian. Watermarking based on discrete fractional random transform. Optics Commun 2007;272(2):344–8.
[8] Pei Soo-Chang, Yeh Min-Hung. The discrete fractional Cosine and Sine transforms. IEEE Trans Signal Process 2001;49(6):1198–207.
[9] Dong Xiaoxiao, Bocko Mark F, Ignjatovic Zeljko. Data hiding via phase manipulation of audio signals. In: Proceeding of IEEE international conference on acoustics, speech, and signal processing; 2004. p. 377–80.
[10] Takahashi Akira, Nishimura Ryouichi, Suzuki Yoiti. Multiple watermarks for stereo audio signals using phase-modulation techniques. IEEE Trans Signal Process 2005;53(2):806–15.
[11] Chen Brain, Wornell Gregory W. Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. IEEE Trans Inform Theory 2001;47(4):1423–43.
[12] Kundur D, Hatzinakos D. Digital watermarking for telltale tampering proofing and authentication. In: Proceedings of IEEE; 1999. p. 1167–81.
[13] Kutter M, Petitcolas FAP. A fair benchmark for image watermarking systems. In: Proceeding of electronic imaging, vol. 3657; 1999. p. 226–39.