A robust digital audio watermarking based on statistics characteristics

Xiang-Yang Wang¹,b,∗, Pan-Pan Niub, Hong-Ying Yangb

¹State Key Laboratory of Networking and Switching Technology (Beijing University of Posts and Telecommunications), Beijing 100876, China
²School of Computer and Information Technology, Liaoning Normal University, Dalian 116029, China

ARTICLE INFO

Article history:
Received 27 September 2007
Received in revised form 18 July 2008
Accepted 12 January 2009

Keywords:
Audio watermarking
Desynchronization attack
Synchronization code
Statistics average value
DWT

ABSTRACT

In digital audio watermarking, the watermark’s vulnerability to desynchronization attacks has long been a difficult problem. According to the audio statistics characteristics and synchronization code technique, a new robust audio watermarking scheme against desynchronization attacks is proposed in this paper. Firstly, the original digital audio is segmented and then each audio segment is cut into two parts. Secondly, with the spatial watermarking technique, the synchronization code is embedded into the statistics average value of audio samples in the first part. Finally, the second part of audio segment is cut into audio sections, the DWT is performed on the audio sections, and the watermark bit is embedded into the statistics average value of low frequency components. Experimental results show that the proposed scheme is inaudible and robust against common signal processing attacks, including MP3 compression, low-pass filtering, noise addition, and equalization, etc. Moreover, it also survives several desynchronization attacks, such as random cropping, amplitude variation, pitch shifting, time-scale modification, and jittering, etc.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

The advent of the digital age with the Internet revolution has made it extremely convenient for users to access, create, process, copy, or exchange multimedia data. This has created an urgent need for protecting intellectual property in the digital media. Digital watermarking is a technology being developed, in which copyright information is embedded imperceptibly into the host in a way that is robust to a variety of intentional or unintentional attacks.

With the development of audio watermarking technologies, attacks against audio watermarking systems have become more sophisticated [1,2]. In general, the attacks on audio watermarking systems can be categorized into common signal processing, such as MP3 compression, low-pass filtering, and noise addition, etc., and desynchronization attacks such as amplitude variation, pitch shifting, time-scale modification, etc. While the common signal processing reduces watermark energy, desynchronization attacks induce synchronization errors between the original and the extracted watermark during the detection process. Most of the previous audio watermarking schemes have shown robustness against common signal processing attacks, and only a few specialized watermarking methods have addressed the desynchronization attacks [3,4]. These few can be classified into all-list-search, combination of spread spectrum and spread spectrum code, utilizing the important feature of original digital audio (or we call it self-synchronization strategy), and synchronization code.

Among them, all-list-search strategy need great calculating amount and has high false positive rate [6]; most second strategy cannot achieve blind detection [5–7]. Seok et al. [8] gave an audio watermarking scheme by exploiting the human perceptual characteristics of the audio signal to regulate the embedding strength, but it is not very robust to some audio signal processing such as re-sampling, re-quantization and compression; the current self-synchronization algorithm cannot extract feature points steadily, besides, it usually need large number of threshold values which make it more difficult to be applied [9,10]. Girin et al. [11] proposed a speech signal watermarking using the sinusoidal model and amplitudes, phases and digital frequencies modulation of the partials, Hee et al. [12] proposed an audio watermarking algorithm through modification of tonal maskers, but these methods suffer from poor robustness against time-scale modification and pitch...
shifting. By contrast, synchronization code strategy has more obvious technological advantages. Wu et al. [13] proposed a blind audio watermarking algorithm with self-synchronization based on DWT; Huang et al. [14] chooses Bark code which has better self-relativity as synchronization mark and embeds it into temporal domain, then, embeds the watermark information into DCT domain; Du et al. [15] proposed an audio watermark detecting improved algorithm based on HAS. It is possible to resist desynchronization attack by utilizing the advanced synchronization code technique, but the existing audio watermarking approaches have shortcomings as follows: (1) They choose a 12-bit Barker code which is so short that it is easy to cause false synchronization. (2) They are vulnerable to re-sampling and jittering, and very few researchers have performed and published sufficient experiments involving amplitude variation, pitch shifting, time-scale modification, etc. (3) They do not make full use of human auditory masking effect, which influences the imperceptibility and robustness performance of watermarking.

In order to solve the above problems, we propose a new robust audio watermarking scheme against desynchronization attacks by utilizing the audio statistics characteristics [16] and synchronization code technique. We choose 16-bit Barker code as synchronization mark, and embed it by modifying the statistics average value of several audio samples. Besides, in order to make full use of auditory masking effect, we embed the digital watermark into the statistics average value of low frequency components in DWT domain. Experimental results show that the proposed scheme is inaudible and robust against common signals processing, including MP3 compression, low-pass filtering, noise addition, and equalization, etc. Moreover, it also survives several desynchronization attacks, such as random cropping, amplitude variation, pitch shifting, time-scale modification, and jittering, etc.

This paper is organized as follows. In Section 2, the basic principle of the proposed scheme and construction will be given. Sections 3 and 4 describe our watermark embedding and extracting algorithm, respectively. Section 5 will be dedicated to the description of a variety of simulation experiments, which will illustrate the effectiveness of the proposed scheme. Finally conclusions will be briefed in Section 6.

2. Fundamental theory and synchronization

2.1. Fundamental theory

In our audio watermarking scheme, the watermark can be embedded into the host audio by three steps. Firstly, the host audio is segmented and then each audio segment is cut into two parts. Secondly, with the spatial watermarking technique, synchronization code is embedded into the statistics average value of audio samples in the first part. Finally, the second part of audio segment is cut into audio sections, the DWT is performed on audio sections, and the watermark bit is embedded into the statistics average value of low frequency components. The construction of embedding information is shown in Fig. 1. The watermark embedding procedure is shown in Fig. 2.

2.2. Synchronization code

Synchronization is one of the key issues of audio watermarking. Watermark detection starts by alignment of watermarked block with detector. Time-scale or frequency-scale modification makes the detector lose synchronization, which causes false detection. So we need exact synchronization algorithms based on robust synchronization code.

Generally, we should avoid false synchronization during selecting synchronization code. Several reasons contribute to false synchronization: (1) the style of the synchronization code, (2) the length of synchronization code, (3) the probability of “0” and “1” in synchronization code. Among of them, the length of synchronization code is especially important. The longer it is, the more robust it is.

We employ Barker code in front of the watermark to locate the position where watermark is embedded. Barker codes, which are subsets of PN sequences, are commonly used for frame synchronization in digital communication systems. Barker codes have low correlation sidelobes. A correlation sidelobe is the correlation of a codeword with a time-shifted version of itself. Barker codes have low correlation sidelobes, $R(j)$, for a $j$-symbol shift of an $n$-bit code sequence, $\{x_i\}$ is given by

$$R(j) = \sum_{i=1}^{n-j} x_i x_{i+j} = \begin{cases} n, & j = 0 \\ 0, & 0 < j < n \\ 0, & j \geq n \end{cases}$$

where $x_i$ is an individual code symbol taking values $+1$ or $-1$ for $i = 1, 2, 3 \ldots, n$, and the adjacent symbols are assumed to be zero.

![Fig. 1. Construction of embedding information.](image1)

Please cite this article as: X.-Y. Wang, et al., A robust digital audio watermarking based on statistics characteristics, Pattern Recognition (2009), doi: 10.1016/j.patcog.2009.01.015
3. Watermark embedding scheme

Let $A = \{a(i), 0 \leq i < \text{Length} \}$ represent a host digital audio signal with length samples. $W = \{w(i,j), 0 \leq i < \text{M}, 0 \leq j < N \}$ is a binary image to be embedded within the host audio signal, and $w(i,j) \in \{0, 1\}$ is the pixel value at $(i,j)$. $F = \{f(i), 0 \leq i < \text{Lsyn} \}$ is a synchronization code with $\text{Lsyn}$ bits, where $f(i) \in \{0, 1\}$. The main steps of the embedding procedure based on audio statistics characteristics can be described in detail as follows.

3.1. Watermark preprocessing

In order to dispel the pixel space relationship of the binary watermark image, and improve the robustness of the whole digital watermark system, watermark scrambling algorithm is used at first. In our watermark embedding scheme, the binary watermark image is scrambled from $W$ to $W_1$ by using Arnold transform, where

$$W_1 = \{w_1(i,j), 0 \leq i < \text{M}, 0 \leq j < N \}$$

Then, it is transformed into a one-dimensional sequence of ones and zeros as follows:

$$W_2 = \{w_2(k) = w_1(i,j), 0 \leq i < \text{M}, 0 \leq j < N, \quad i = i_x N + j \}, \quad w_2(k) \in \{0, 1\}$$

Finally, each bit of the watermark data is mapped into an antipodal sequence using BPSK modulation according to the following equations:

$$W_3 = \{w_3(k) = 1 - 2 \times w_2(k), \quad k = 0, 1, \ldots, \text{M} \times \text{N} - 1, \quad w_3(k) \in \{-1, 1\} \}$$

In order to improve the robustness of proposed scheme against cropping, time-scale modification, jittering and make the detector available when it loses synchronization, audio segmenting is used at first. Then, each segment is cut into two parts with $L_1$ and $L_2$ samples, respectively, where

$$A(i) = \{a(iL + k), 0 \leq k < L \}, \quad 0 \leq i < \left( \frac{\text{Length}}{L} \right)$$

where $L = L_1 + L_2$, $L_1 = \text{Lsyn} \times n$, $n$ is a constant and is chosen to be 5 samples in our experiment.

Let $A^0$ denote each segment, and $A^0$ is cut into two parts $A^0_1$ and $A^0_2$ with $L_1$ and $L_2$ samples. Synchronization code and watermark are embedded into $A^0_1$ and $A^0_2$, respectively.

3.2. Synchronization code embedding

According to the statistics characteristics of digital audio, a small modification of some audio samples will not change their statistics average value seriously. In order to guarantee robustness and transparency of watermarking, the proposed scheme embeds synchronization code into the statistics average value of audio samples as follows.

(1) The first part $A^0_1$ of audio segment $A^0$ is cut into $\text{Lsyn}$ audio subsamples, and each audio subsample $PA^0_1(m)$ having $n$ samples, where

$$PA^0_1(m) = \{pa^0_1(m)(i), 0 \leq i < n, 0 \leq m \leq \text{Lsyn} \}$$

(2) Calculating the mean value of $PA^0_1(m)$, that is

$$\overline{PA^0_1(m)} = \frac{1}{n} \sum_{i=0}^{n-1} pa^0_1(m)(i), \quad 0 \leq m \leq \text{Lsyn}$$

(3) The synchronization code can be embedded into each $PA^0_1(m)$ by quantizing the mean value $\overline{PA^0_1(m)}$, the rule is given by

$$pda^0_1(m)(i) = pa^0_1(m)(i) + (\overline{PA^0_1(m)} - \overline{PA^0_1(m)})$$

where $PA^0_1(m) = \{pa^0_1(m)(i), 0 \leq i < n \}$ is original sample, $PA^0_1 = \{pda^0_1(m)(i), 0 \leq i < n \}$ is modified sample and

$$\overline{PA^0_1(m)} = \left\{ \begin{array}{ll} \frac{IQ(\overline{PA^0_1(m)})}{S_1} \times S_1/2 & \text{if } Q(\overline{PA^0_1(m)}) = f(m) \\ \frac{IQ(\overline{PA^0_1(m)})}{S_1} \times S_1 - S_1/2 & \text{if } Q(\overline{PA^0_1(m)}) \neq f(m) \end{array} \right. \frac{IQ(\overline{PA^0_1(m)})}{S_1}$$

where $Q(x, y)$ returns the remainder of the division of $x$ by $y$, and $S_1$ is the quantization step size for embedding synchronization code.

3.3. Watermark embedding

(1) The second part $A^0_2$ of audio segment $A^0$ is cut into audio sections and each audio section $A^0_2(k) (k = 0, 1, \ldots, \text{M} \times \text{N} - 1)$ is chosen to be $L_2/M \times N$ samples.

(2) DWT: For each audio section $A^0_2(k)$, $H$-level DWT is performed, and we get the wavelet coefficients of $A^0_2(k)^{H}, A^0_2(k)^{H-1}, \ldots, A^0_2(k)^{1}$, where $A^0_2(k)^{H}$ is the coarse signal

$$A^0_2(k)^{H} = \left\{ a^0_2(k)(t)^{H}, k = 0, 1, \ldots, \text{M} \times \text{N} - 1, 0 \leq t < \frac{L_2}{\text{M} \times \text{N} \times 2^{H}} \right\}$$

and the detail signals are $D^0_2(k)^{H}, D^0_2(k)^{H-1}, \ldots, D^0_2(k)^{1}$.

(3) Calculating the mean value of $A^0_2(k)^{H}$, that is $A^0_2(k)^{H}$. (4) Calculating

$$\frac{A^0_2(k)^{H}}{A} = \left. \frac{A}{A} \right|_{d} = e(k) + \delta(k)$$

where $A$ is the quantization step size for embedding watermark, $e(k)$ and $\delta(k)$ stand for the quotient and remainder, respectively.

(5) Watermark embedding: According to the statistics characteristics of digital audio in DWT domain, a small modification of some wavelet coefficients will not change their statistics average value seriously. Therefore, the watermark bit is embedded into the corresponding audio section $A^0_2(k)$ by quantizing the mean value $A^0_2(k)^{H}$, and the rule is given by

$$a^0_2(k)(t)^{H} = a^0_2(k)(t)^{H} - DM(k)$$

where $a^0_2(k)(t)$ are the original wavelet coefficients, $a^0_2(k)(t)^{H}$ are the modified wavelet coefficients, $DM(k)$ is the adjusting value of mean value $A^0_2(k)^{H}$ and

$$DM(k) = \left\{ \begin{array}{ll} \frac{A}{2} - \delta(k) & \text{if } \text{mod}(e(k), 2) = 0 \\ \frac{3A}{2} - \delta(k) & \text{if } \text{mod}(e(k), 2) = 1, \quad \delta > \frac{A}{2} \\ \frac{A}{2} - \delta(k) & \text{if } \text{mod}(e(k), 2) = 1, \quad \delta \leq \frac{A}{2} \end{array} \right.$$
If $w_3(k) = -1$, then

$$DM(k) = \begin{cases} \Delta - \delta(k) & \text{if } \mod(d(k), 2) = 1 \\ \frac{3\Delta}{2} - \delta(k) & \text{if } \mod(d(k), 2) = 0, \delta > \frac{\Delta}{2} \\ -\frac{\Delta}{2} - \delta(k) & \text{if } \mod(d(k), 2) = 0, \delta \leq \frac{\Delta}{2} \end{cases}$$

(6) **Inverse DWT:** After substituting the coefficients $A_0^0(k)^H$ with $A_0^1(k)^H$, $H$-level Inverse DWT is performed, and then the watermarked digital audio signal is $A_0^1(k)$.

### 3.4. Repeat embedding

In order to improve the robustness against desynchronization attacks, the proposed scheme repeats Sections 3.2 and 3.3 to embed synchronization code and digital watermark into every audio segment.

### 4. Watermark detecting scheme

The watermark detecting procedure in the proposed scheme neither needs the original audio signal nor any other side information. The watermark detection procedure is summarized as follows.

(1) Locating the beginning position $B$ of the watermarked segment is achieved based on the frame synchronization technology of digital communications.

(2) $H$-level DWT is performed on each audio section $A^*(k)$ after $B$, and then gets the coefficients as follows:

$$A^*(k)^H, D^*(k)^H, D^*(k)^{H-1}, ..., D^*(k)^1$$

(3) Calculating the mean value of $A^*(k)^H$, that is $\overline{A^*(k)^H}$.

(4) Calculating

$$\overline{A^*(k)^H} = \hat{c}^*(k) + \hat{\delta}^*(k)$$

where $\Delta^*$ is the quantization step size, $\hat{c}^*(k)$ and $\hat{\delta}^*(k)$ stand for the quotient and remainder, respectively.

(5) The digital watermark extraction rule is

$$w'_3(k) = \begin{cases} -1 & \mod(\hat{c}^*(k), 2) = 1 \\ 1 & \mod(\hat{c}^*(k), 2) = 0 \end{cases}$$

$$w'_2(k) = (1 - w'_3(k))/2, k = 0, 1, ..., M \times N - 1, w'_2(k) \in \{0, 1\}$$

(6) All the detected watermark bits $W'_2$ are rearranged to form the binary watermark image $W'_1$.

(7) The watermark bits are determined based on BPSK demodulation:

$$W'_2 = \{w'_2(k) = (1 - w'_3(k))/2, k = 0, 1, ..., M \times N - 1, w'_2(k) \in \{0, 1\}\}$$

(8) Finally, the watermark image

$$W^* = \{w^*(i,j), 0 \leq i < M, 0 \leq j < N\}$$

can be obtained by descrambling.
In this study, reliability was measured as the bit error rate (BER) of extracted watermark, its definition is

\[ BER = \frac{B}{M \times N} \times 100\% \]

where \( B \) is the number of erroneously detected bits.

The peak signal to noise ratio (PSNR) is used to evaluate the quality of watermarked audio, which can be represented as

\[ PSNR\ (dB) = 10 \log_{10} \frac{A_{\text{peak}}^2}{\sigma_e^2} \]

where \( \sigma_e^2 \) is defined as

\[ \sigma_e^2 = \left( \frac{1}{L} \right) \sum_{i=1}^{L} (A(i) - A^*(i))^2 \]

where \( L \) is the length of the host audio, \( A(i) \) is the magnitude of host audio at time \( i \). Similarly, \( A^*(i) \) denotes the magnitude of watermarked audio at time \( i \). \( A_{\text{peak}}^2 \) denotes the squared peak value of host audio. The higher PSNR means that the watermarked audio is more similar to the original audio.

5. Experimental results

In order to illustrate the inaudible and robust nature of watermarking scheme, we test the proposed watermarking scheme on digital audio piece (music) with 16 bit signed mono audio signals sampled at 44.1 kHz. We use a 64 x 64 bit binary image as our watermark and a 16-bit Barker code 111100110101110 as synchronization code. The Daubechies-1 wavelet basis is used. The smaller level DWT will influence the robustness of the watermark; and the larger one will cause large calculation, so 3-level DWT is performed in this test. In our experiments, the length of each segment of watermark embedding is fixed as 65,536 samples.

For the quantization-based digital audio watermarking algorithms, the quantization step plays an important role in transparency and robustness. The smaller quantization step size will influence the robustness of the watermark and increase the calculating amount, and the larger one will cause poorer quality of the watermarked audio. In order to obtain the optimal quantization step size, we perform a series of experiment for quantization step size selection. From the experiment results (see Figs. 3–7), we can obtain the optimal quantization step size \( \Delta = 0.5 \). Similarly, we can obtain the optimal quantization step size \( S_1 = 0.2 \).

Figs. 8(a) and (b) show the original waveform and the watermarked waveform which lengths are 9.75 s. Similarly, Figs. 8(c) and (d) show the original watermark image and the extracted watermark image.

To evaluate the performance of the proposed watermarking algorithm, various attacks are used to estimate the robustness of our scheme comparing with that of scheme [14]. According to their influence on synchronization, attacks can be divided into common audio signal processing and desynchronization attacks. Common audio
signal processing may distort the perceptual quality but not affect the synchronization structure, including re-quantization, re-sampling, additive noise, low-pass filtering, echo addition, equalization, MPEG compression, etc. Desynchronization attacks introduce very little distortion to the watermarked audio but destroy the synchronization needed by most existing audio watermarking algorithms, including random cropping, amplitude variation, pitch shifting, time-scale modification, jittering, etc.

**Re-quantization:** We tested the process of re-quantization of a 16-bit watermarked audio signal to 8-bit and back to 16-bit.

**Re-sampling:** In this experiment, the original audio signals are sampled with a sampling rate of 44.1 kHz. Watermarked audio signals are down sampled to 22.05, 11.025, 8 kHz and the up sampled back to 44.1 kHz.

**Additive noise:** White noise with 10% of the power of the audio signal is added.

Please cite this article as: X.-Y. Wang, et al., A robust digital audio watermarking based on statistics characteristics, Pattern Recognition (2009), doi: 10.1016/j.patcog.2009.01.015
### Table 1
The watermark detection results for common signal processing.

<table>
<thead>
<tr>
<th>Scheme [14] Watermark</th>
<th>Attack free</th>
<th>Re-quantization 22.05 kHz</th>
<th>Re-sampling 11.025 kHz</th>
<th>Re-sampling 8 kHz</th>
<th>Additive noise</th>
<th>Low-pass filtering (4 kHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BER</td>
<td>0</td>
<td>0</td>
<td>0.501</td>
<td>0.489</td>
<td>0.026</td>
<td>0.491</td>
</tr>
<tr>
<td>PSNR</td>
<td>46.309</td>
<td>43.450</td>
<td>30.341</td>
<td>27.454</td>
<td>40.834</td>
<td>41.603</td>
</tr>
</tbody>
</table>

| Proposed scheme Watermark | | | | | | |
|---------------------------| | | | | | |
| BER                       | 0          | 0                         | 0.010                  | 0.016            | 0.032         | 0.003                    |
| PSNR                      | 46.301     | 43.523                    | 30.353                 | 27.509           | 40.621        | 41.123                   |

<table>
<thead>
<tr>
<th>Scheme [14] Watermark</th>
<th>Echo addition</th>
<th>Equalization</th>
<th>MP3-256kb</th>
<th>MP3-128kb</th>
<th>MP3-112kb</th>
<th>MP3-64kb</th>
<th>MP3-56kb</th>
</tr>
</thead>
<tbody>
<tr>
<td>BER</td>
<td>0.190</td>
<td>0.159</td>
<td>0</td>
<td>0</td>
<td>0.502</td>
<td>0.498</td>
<td>0.498</td>
</tr>
<tr>
<td>PSNR</td>
<td>37.472</td>
<td>42.265</td>
<td>46.316</td>
<td>46.323</td>
<td>46.320</td>
<td>46.037</td>
<td>45.790</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proposed scheme Watermark</th>
<th>Echo addition</th>
<th>Equalization</th>
<th>MP3-256kb</th>
<th>MP3-128kb</th>
<th>MP3-112kb</th>
<th>MP3-64kb</th>
<th>MP3-56kb</th>
</tr>
</thead>
<tbody>
<tr>
<td>BER</td>
<td>0.027</td>
<td>0.013</td>
<td>0</td>
<td>0</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>PSNR</td>
<td>37.503</td>
<td>42.312</td>
<td>46.402</td>
<td>46.442</td>
<td>48.451</td>
<td>47.326</td>
<td>46.159</td>
</tr>
</tbody>
</table>

**Low-pass filtering:** The low-pass filter with cutoff frequency of 4 kHz is applied to watermarked audio signal.

**Echo addition:** An echo signal with a delay of 98 ms and a decay of 41% was added to the original audio signal.

**Equalization:** The “Hum Removal” preset of the audio editing tool CoolEdit Pro2.1 was used, which is a 6-band graphic equalizer. The 50, 100, 150, 200, 250, and 300 Hz frequency bands were boosted by −18 dB.

**MPEG compression:** We have studied the robustness of the watermark to the coding/decoding of the audio signal. The coding/decoding was performed using a software implementation of the ISO/MPEG-2 Audio Layer III coder with several different bit rates (256, 128, 112, 64, 56 kbit).

**Random cropping:** In our experiment, 10% samples were cropped at each of three randomly selected positions (front, middle and back).

**Amplitude variation:** The watermarked signal was attenuated up to 150% and down to 50%.

**Pitch shifting:** Tempo-preserved pitch shifting is a difficult attack for audio watermarking algorithms, because it causes frequency fluctuation. In our experiment, the pitch is shifted one degree higher and one degree lower.

**Time-scale modification:** The watermarked audio signal was lengthened (slowed-down) by 4% (−4%) while preserving the pitch.

**Jittering:** Jittering is an evenly performed form of random cropping. We removed one sample out of every 5000 (10 000) samples in our jittering experiment.

**Tables 1 and 2** summarize the proposed watermark detection results comparing with that of scheme [14] against various attacks. The BER of watermark image and the PSNR of digital audio signal are also given.

### 6. Conclusion
We have successfully demonstrated a new robust digital audio watermarking algorithm against desynchronization attacks, which can be significantly improved by the audio statistics characteristics and synchronization code. The robustness of the method is based on three key components of our approach: the original digital audio is segmented and then each segment is cut into two parts, synchronization code is embedded into the statistics average value of audio samples in the first part, and the DWT is performed on the second part which is cut into audio sections and then the watermark bit is embedded into the statistics average value of low frequency components. By the analytical and experimental findings, the proposed watermarking method achieves robustness against both common audio signal processing and desynchronization attacks. In addition, the watermark can be extracted without the help from the original digital audio signal and can be easily implemented.
Table 2

The watermark detection results for desynchronization attacks.

<table>
<thead>
<tr>
<th>Scheme [14] Watermark</th>
<th>Cropping 10% (front)</th>
<th>Cropping 10% (middle)</th>
<th>Adding 10% (front)</th>
<th>Amplitude-scaling up to 150%</th>
<th>Amplitude-scaling down to 50%</th>
<th>Pitch shift one degree higher</th>
<th>Pitch shift one degree lower</th>
</tr>
</thead>
<tbody>
<tr>
<td>BER</td>
<td>0</td>
<td>0</td>
<td>0.500</td>
<td>0.498</td>
<td>0.508</td>
<td>0.501</td>
<td>0.504</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Proposed scheme Watermark</th>
<th>Cropping 10% (front)</th>
<th>Cropping 10% (middle)</th>
<th>Adding 10% (front)</th>
<th>Amplitude-scaling up to 150%</th>
<th>Amplitude-scaling down to 50%</th>
<th>Pitch shift one degree higher</th>
<th>Pitch shift one degree lower</th>
</tr>
</thead>
<tbody>
<tr>
<td>BER</td>
<td>0</td>
<td>0</td>
<td>0.503</td>
<td>0.492</td>
<td>0.494</td>
<td>0.495</td>
<td>0.493</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scheme [14] Watermark</th>
<th>Cropping 10% (front)</th>
<th>Cropping 10% (middle)</th>
<th>Adding 10% (front)</th>
<th>Amplitude-scaling up to 150%</th>
<th>Amplitude-scaling down to 50%</th>
<th>Pitch shift one degree higher</th>
<th>Pitch shift one degree lower</th>
</tr>
</thead>
<tbody>
<tr>
<td>BER</td>
<td>0</td>
<td>0</td>
<td>0.024</td>
<td>0.130</td>
<td>0.207</td>
<td>0.103</td>
<td>0.212</td>
</tr>
</tbody>
</table>

References


About the Author—XIANGYANG WANG was born in Tieling, China, in 1965. He is currently a professor with the School of Computer and Information Technology at the Liaoning Normal University, China. He obtained his B.S. degree from the Lanzhou University, China, and his M.S. degree from the Jilin University, China, in 1988 and 1995, respectively. His research interests include signal processing and communications, digital multimedia data hiding and information assurance, applications of digital image processing, computer vision. He has published more than 150 journal papers, 20 conference papers, and contributed in 2 books in his areas of interest.

About the Author—PANPAN NIU received the B.S. degree from the School of Computer and Information Technology, Liaoning Normal University, China, in 2006, where she is currently pursuing the M.S. degree. Her research interests include digital watermarking and signal processing.

About the Author—HONGYING YANG is currently an assistant professor with the School of Computer and Information Technology at the Liaoning Normal University, China. He obtained his B.S. degree from the LiaoHiong Normal University, China, in 1989. Her research interests include signal processing and communications, digital multimedia data hiding and information assurance, applications of digital image processing, computer vision. She has published more than 150 journal papers, 20 conference papers, and contributed in 2 books in his areas of interest.

Please cite this article as: X.-Y. Wang, et al., A robust digital audio watermarking based on statistics characteristics, Pattern Recognition (2009), doi: 10.1016/j.patcog.2009.01.015