Digital Audio Identification Using Watermarking Technique

Sumit Buapun, Sulyuth Swangwana, Sirikorn Channual

Department of Computer Science, Kasetsart University Email: sumit_bua@truecorp.co.th
Department of Computer Science, Royal Thai Airforce Academy Email: sulyuth@spu.ac.th
Department of Computer Science, Kasetsart University Email: fsciskp@ku.ac.th

ABSTRACT

In this paper, a new algorithm for audio watermarking is proposed for identification of audio signal, both speech and music. The basic idea of the algorithm is to adjust the magnitude of audio signal sample to embed data in time domain. The embedded signal must not be perceptually different from the original signal and cannot be detected by a human sensing. This method does not require the original signal for comparing with the watermarked signal in watermark detection. The experimental result shows the quality of watermarked audio signal in terms of signal to noise ratio and

Keywords: Digital Watermarking, Audio Watermarking

1. Introduction

In the past few years we have seen an explosion in the use of digital media that offer several distinct advantages over analog media, for example, the quality of digital audio signal is higher than analog audio signal and digital audio is easily transmitted over networked information system. However, low-cost but powerful computers nowadays allow people to easily create and copy multimedia content that bring about the hardship of owner identification as well as copyright protection. Digital watermarking is one of several important approaches used to attack the problem [1]. Digital watermarking is a technique used to secretly label the host media, which is kept or distributed in digital form. The technique consists of two subsystems, the signature embedding subsystem and the detection subsystem. The embedding subsystem is usually based on adding a signature to the original signal. The embedded signal should be perceptually inaudible, or its presence should not interfere with the content being protected. The embedding subsystem should provide prevention of unauthorized removal, easily be extracted by the owner, and robust to distortion. The detection subsystem detects the existence of the signature in the media. That does not require the original signal for watermark detection.

This paper provides an effective method to protect copyright and to trace illegal distributions of the digital audio by embedding digital watermark in time domain, with the new condition for embedding information’s bits in time domain the condition different from the paper had been presented before [3]. The embedded information will not affect the audio quality. The watermark embedding and detection can be done very easily. The detected watermark information can provide proof of copyright and distribution of sources. This method present the new condition for embedding. For copyright protection, the watermark content must contain the owner identification information which is identical in all audio content. For tracing illegal distributions, the watermark content must contain the user identification which is different for each audio transaction. Watermark is highly related to the audio content and it tightly follows the masking threshold of the human auditory system [4]. Watermark embedding increases the data rate very little hence it will not cause perceptible distortion in audibility. Any attempt to remove or distort it, including re-encoding the audio content, will lead to perceptible distortion of the original audio content, since the watermark is partially embedded.

2. Watermark embedding

The watermark embedding technique modifies a digital audio signal sample which is represented in 16-bit sample sequence. The condition of bit embedding is based on the relative energy relationship between adjacent sample sections. The block diagram of the watermark embedding process is shown in figure 1. First, the audio signal sample is segmented into sections of length \( L \) by the quantity of watermark data bits. The audio sample of length \( L \) is then sorted by magnitude from maximum to minimum and grouped into three sections. The average value of data sample in each section are defined as \( \text{Max} \), \( \text{Mid} \) and \( \text{Min} \) that can be calculated by the following equations

\[
\text{Max} = \frac{3}{L} \sum_{i=1}^{L} x(i)
\]

(1)

\[
\text{Mid} = \frac{3}{L} \sum_{i=L-2}^{L} x(i)
\]

(2)
\[ \text{Min} = \frac{3}{L} \sum_{i=\frac{L}{3}+1}^{L} x(i) \]  

where \( x(i) \) is an audio sample of length \( L \) after magnitude sorting from maximum to minimum. The differences of average data sample summation are defined by \( D_1 \) and \( D_2 \) that can be calculated by the following equations:

\[ D_1 = \text{Max} - \text{Mid} \]  

\[ D_2 = \text{Mid} - \text{Min} \]

The watermark bit sequence is embedded using the following condition. For bit “1” embedding adjust energy of audio sample of length \( L \) until \( D_1 \geq D_2 \), and for bit “0”, embedding adjust energy of audio sample of length \( L \) until \( D_1 < D_2 \). Replace audio sample at the beginning position before sorting. The watermark bit sequence is generated by multiplying bit by bit between data symbol bit sequence and pseudo noise bit sequence [8, 9]. In this paper the picture in jpeg format is used as watermark data symbol, and the digital audio wave format is used as host signal (mono signal with sampling frequency 44.1 kHz, 16 bit/sample).

3. Watermark Extraction

The watermark extraction start from the watermarked signal is segmented to length \( L \). The differences of average data energy calculation are repeated exactly in the same way as the embedding process. The average energy calculation is done using the equation (1), (2), and (3). The differences of data energy are calculated using the equation (4) and (5). Data bit extraction by the condition. \( D_1 \geq D_2 \) data bit is “1”, and \( D_1 < D_2 \) data bit is “0”. The data bit stream extraction are then multiplied by pseudo noise bit sequence. The result of the multiplication is the data symbol. A diagram of the extraction process is shown in figure 2.

![Watermark embedding process](image1)

**Figure 1** Watermark embedding process

![Watermark extraction process](image2)

**Figure 2** Watermark extraction process

The evaluation method is performed by bit error rate (BER) [2] and signal to noise ratio (SNR) [2] which are determined by the following equations.

\[ \text{BER} = \frac{100}{B} \sum_{n=0}^{B-1} \begin{cases} 1, \tilde{w}(n) = w(n) \\ 0, \tilde{w}(n) = \bar{w}(n) \end{cases} \]

where \( \tilde{w}(n) \) represents the sequence of watermark bit extracted from the watermarked host signal, \( w(n) \) represents the sequence of bits to be embedded within the host signal, and \( B \) represents the total block of audio sample which embedded with watermark bit.
where \( x(n) \) represents a host signal and \( x'(n) \) represents a watermarked signal of length \( L \).

4. Result

Implemented watermarking algorithm has been tested using a large set of songs from several music styles, e.g., tradition Thai song, Thai folk song, and Western song, in wave format (44.1 kHz sampling rate, 16 bits/sample mono) for host signal. The music pieces have been embedded with a picture file as shown in figure 3 (a), which is spread by 7 bit PN code “11101001”. The quality of watermarked signal is presented by the signal to noise ratio as shown in table 1. The extracted watermark is shown in figure 3 (b). The extracted bits of watermark signal are identical to the original file. And the BER is 0% for all songs.

5. Conclusion

In this paper, we present a novel algorithm for embedding data into digital audio signal by data sample magnitude modification in time domain. The watermark could be detected without using the original audio for data reference, with 100% extraction accuracy. The quality of watermarked signal which is presented in terms of signal to noise ratio from difference types of audio data (speech and music form difference styles of song) have been appeared in range between 19 - 30 db. The result shows that the algorithm can support information embedment into a wide variety of audio data styles with high reliability and considerably low effect to the quality of original signal.

<table>
<thead>
<tr>
<th>SONGS</th>
<th>SNR (db)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional Thai folk</td>
<td></td>
</tr>
<tr>
<td>Traditional Thai</td>
<td></td>
</tr>
<tr>
<td>Western</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>23.3</td>
</tr>
<tr>
<td>3</td>
<td>20.9</td>
</tr>
<tr>
<td>4</td>
<td>19.5</td>
</tr>
<tr>
<td>5</td>
<td>20.1</td>
</tr>
<tr>
<td>6</td>
<td>22.7</td>
</tr>
<tr>
<td>7</td>
<td>23.19</td>
</tr>
<tr>
<td>8</td>
<td>20.7</td>
</tr>
<tr>
<td>9</td>
<td>20.4</td>
</tr>
<tr>
<td>10</td>
<td>23.7</td>
</tr>
<tr>
<td>Average</td>
<td>21.449</td>
</tr>
</tbody>
</table>

6. REFERENCES


