A Method of High Bit-rate Data Hiding in Music using Spline Interpolation

Ryuta FUJIMOTO, Mamoru IWAKI, and Tohru KIRYU
Graduate School of Science and Technology, Niigata University
8050 Ikarashi-2, Niigata, 950-2180 Japan
{ryuta@bsp., iwaki@, kiryu@}bc.niigata-u.ac.jp

Abstract

Acoustical signals in digital form can include additional information, this information is separate from the original signal even though they are both stored in a same file. Methods for storing additional information into the original acoustical signal itself have been investigated. Their bit-rate is at most several hundred bit per second. This paper proposes a new method for digital watermarking with higher bit-rate. Its information hiding and detection method is based on spline interpolation. As a result of listening experiment, bit-rate of above 1 kbps was achieved without audible deterioration.

1. Introduction

Acoustical signals like speech and music are commonly used in our daily life. Those signals are usually converted into digital ones for convenience. Digital signals have a feature of additional information embedding besides their easy copy and distribution. The additional information is usually managed as a separate record of the digital signal, so the information handling is easy, but then the data size increases. On the other hand, information hiding techniques have applied to embed additional information into acoustical signal itself. Those methods are devised to make it difficult to perceive the sound difference caused by embedding additional information. The additional information can share the transmission media with the original acoustical signal and the data size should not increase. Digital watermark techniques[2, 3, 4] as information hiding techniques for sound have been proposed, which achieved bit-rate of around 10 to 100 bps. However their bit-rates are not sufficiently fast for practical applications. It is expected to develop higher-bit-rate digital watermark techniques for sound.

In this paper, we propose a high bit-rate information hiding technique for sound using spline interpolation. Spline interpolation and its detection correspond to the additional information embedding and extraction processes. The extraction process is almost same as spline interpolation in the embedding process. Computer simulations of embedding random binary information into a classical music were carried out. The extraction accuracy and sound quality were evaluated objectively and subjectively when the watermarked sounds were degraded by MP3 compression. As a result, the bit-rate of the proposed method was at least 1 kbps while keeping it difficult to perceive sound quality deterioration.

2. Information hiding algorithm

Figure 1 illustrates the watermarking method using spline interpolation. First, an original audio signal $s(n)$ is divided into distinct frames of length N samples. The data of the i-th frame is expressed by $x_i(n), (n = 0, 1, \dots, N - 1)$. Next, the data in each frame is divided into two groups N_1 and N_2, then spline interpolation[1] of the frame, $\widehat{x}_i(n)$, is calculated using data in N_1. According to the binary information m_i, the watermarked frame $x'_i(n)$ is constructed as follows:

$$x'_i(n) = \begin{cases} x_i(n), & \text{if } m_i = 0 \\ \widehat{x}_i(n), & \text{if } m_i = 1 \end{cases}. \qquad (1)$$

The watermarked music $s'(n)$ is obtained by gathering the above watermarked frames $x'_i(n)$. Figure 2 shows an example of the frame division and data group allocation in a frame. In this paper, the group N_1 is a set of $\{x_i(n)|n = 0, 1, \dots, 5, N - 5, N - 4, \dots, N - 1\}$, and N_2 is a set of $\{x_i(n)|n = 6, 7, \dots, N - 6\}$.

3. Information extraction algorithm

The embedded information can be extracted by checking whether the data frame is replaced with spline interpolation or not. In subsection 3.1, an extraction method with recalculation of spline interpolation is proposed. In subsection 3.2, an extraction method with difference against original signal is proposed.

Proceedings of the 2006 International Conference on Intelligent
Information Hiding and Multimedia Signal Processing (IIH-MSP'06)
0-7695-2745-0/06 $20.00 © 2006 IEEE

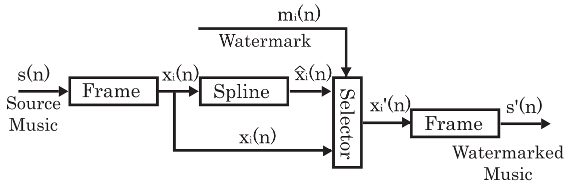

Figure 1. Watermarking process

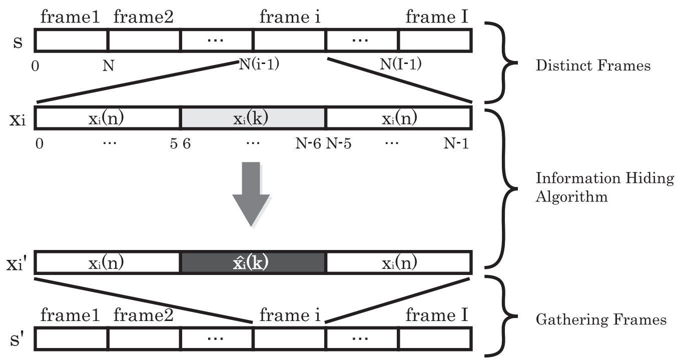

Figure 2. An example of data allocation in a frame

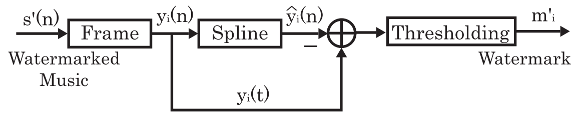

Figure 3. Extraction process with recalculation of spline interpolation

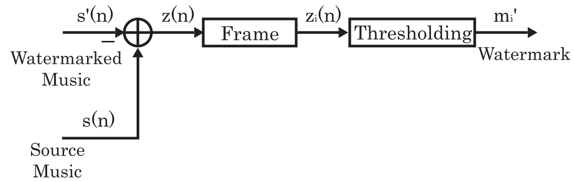

Figure 4. Extraction process with difference from original signal

3.1. Extraction with recalculation of spline interpolation for blind watermarking

Figure 3 illustrates an extraction method with recalculation of spline interpolation. The watermarked signal $s'(n)$ is divided into frames $y_i(n)$. Each frame was separated into two sample groups to calculate the spline interpolation $\widehat{y}_i(n)$ like the embedding process. Based on the difference between $y_i(n)$ and $\widehat{y}_i(n)$, the embedded information is extracted as follows:

$$m'_i = \begin{cases} 1, & \text{if } d(y_i - \widehat{y}_i) \leq k_1 \\ 0, & \text{if } d(y_i - \widehat{y}_i) > k_1 \end{cases}, \tag{2}$$

where $d(v)$ is Euclidean norm of v, and k_1 is a threshold which is previously determined as the extraction probabilities of zero and one are equal for random information embedded signals.

3.2. Extraction with difference from original signal for non-blind watermarking

Figure 4 illustrates an extraction method with difference from original signal. First, the difference signal $z(n)$ between the watermarked signal $s'(n)$ and the original signal $s(n)$ is calculated and divided into distinct frames $z_i(n)$ like the embedding process. Based on the magnitude of $z_i(n)$, the embedded information is extracted as follows:

$$m'_i = \begin{cases} 1, & \text{if } d(z_i) > k_2 \\ 0, & \text{if } d(z_i) \leq k_2 \end{cases}, \tag{3}$$

where k_2 is a threshold which is determined like in the subsection 3.1.

Table 1. Examples of theoretically maximum bit-rate ($b = 1,\ F_s = 44.1$ [kHz])

Frame length N [pt]	11	15	20	25	30	35	40
Bit rate B_r [bps]	4009	2940	2205	1764	1470	1260	1102

4. Theoretical limit of the bit-rate

Theoretically maximum bit-rate is evaluated as follows:

$$B_r = \frac{bF_s}{N}. \tag{4}$$

Here, b is the number of hidden information bits in a frame, N is the number of samples in a frame, and F_s is the sampling frequency. In this paper, it is set $b = 1$, then the bit-rate is determined by N and F_s. Table 1 shows an example of the theoretical limit of the bit-rate. In order to estimate more practical bit-rate, B_r should be multiplied by the information extraction accuracy measured in the next section.

Proceedings of the 2006 International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP'06)
0-7695-2745-0/06 $20.00 © 2006 IEEE

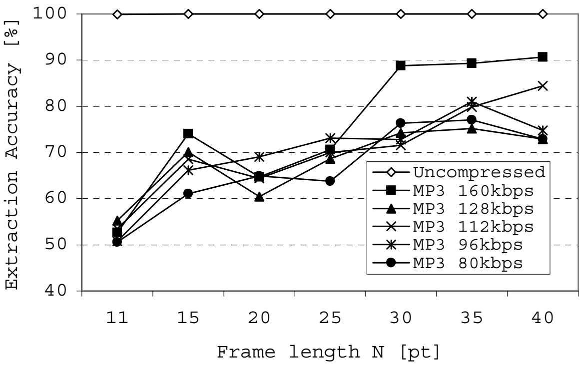

Figure 5. Average extraction accuracy of the spline recalculation method

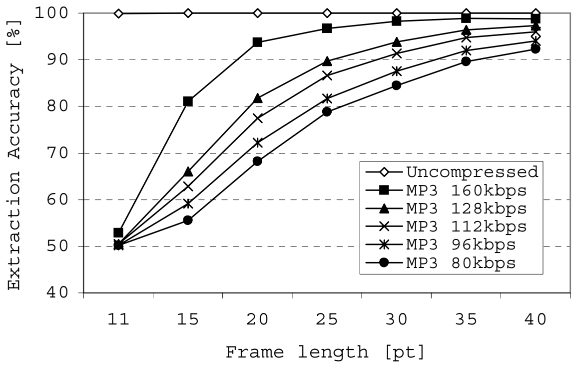

Figure 6. Average extraction accuracy of the differentiating method

5. Performance evaluations

5.1. Experiment conditions

Classical music was used for the sound signal to be watermarked; it was 10 seconds long with sampling frequency of 44.1 kHz and quantization of 16 bit. The frame length N was selected from 11, 15, 20, 25, 30, 35, and 40. The embedded binary information was the same number of zeros and ones with randomly permutated. The degree of spline interpolation was three.

5.2. Accuracy of extracted data

The extraction accuracy of embedded information from the watermarked sounds was examined. As a robustness test, the watermarked sounds were degraded by MP3 compression of 160 kbps, 128 kbps, 112 kbps, 96 kbps, and 80 kbps. Five trials were carried out for each condition. The average of the extraction accuracy of the spline recalculation method is summarized in Fig. 5, and that of the differentiating method was summarized in Fig. 6. It was common between both extraction methods that the accuracy increased as frame length became longer and also that the accuracy was 100% independent of frame length when they were not degraded by MP3 compression. The extraction accuracy decreased for larger MP3 compression rate. The extraction accuracy of the differentiating method was higher than that of the spline recalculation method when the watermarked signal was compressed by MP3.

5.3. Subjective evaluation of sound quality

Sound quality of watermarked sounds was evaluated subjectively by ABX method. Each sound was presented

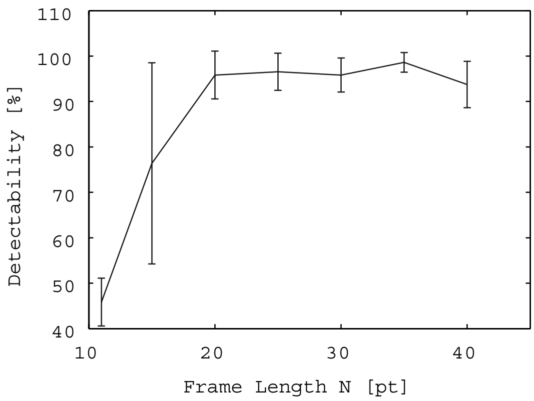

Figure 7. Detectability of difference between data hidden sound and the original one

through headphones. Each subject adjusted the sound volume to their usual listening level before the experiment. Subjects were three males and three females all 22 year old with normal hearing ability. The sound A and B were either the original sound or watermarked one, and the sound X was watermarked sound. Subjects judged whether the sound X was closer to A or B using a 2AFC paradigm. Figure 7 shows the detectability of the difference between watermarked sound and the original one. For the frame length less than 20, the detection ratio rapidly decreased to 50%. This means that watermarked sounds with frame length less than 20 was difficult to distinguish from the original sounds.

5.4. Objective evaluation of sound quality

Sound quality of watermarked sound was evaluated objectively by segmental SNR ($\mathrm{SNR_{seg}}$) as follows:

$$\mathrm{SNR}(i) = 10\log_{10}\frac{\sum_m s_i^2(m)}{\sum_m \{s_i(m) - s'_i(m)\}^2}, \qquad (5)$$

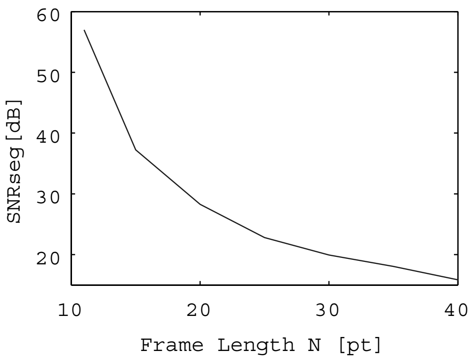

Figure 8. Segmental SNR ($\mathrm{SNR}_{\mathrm{seg}}$)

$$\mathrm{SNR}_{\mathrm{seg}} = \frac{1}{I} \sum_{i=1}^{I} \mathrm{SNR}(i) \quad [\mathrm{dB}]. \tag{6}$$

Here, $s_i(m)$ is the original sound in the i-th segment, $s'_i(m)$ is the watermarked sound in the i-th segment, I is the number of segment in the measured sound. Then $\mathrm{SNR}_{\mathrm{seg}}$ is evaluated as the arithmetic mean of SNR in the i-th segment $\mathrm{SNR}(i)$. In this experiment, segment length was 1410 samples (about 32 ms) long, and segment shift was the half of the segment length. Figure 8 shows the $\mathrm{SNR}_{\mathrm{seg}}$ for each frame length. As the frame length increased, $\mathrm{SNR}_{\mathrm{seg}}$ monotonically decreased. By accompanying with Fig. 7, this means that the degradation of sound quality was hard to perceive for $\mathrm{SNR}_{\mathrm{seg}}$ greater than 25 dB.

6. Discussion

The extraction accuracy decreased in both extraction methods as MP3 compression ratio increased, and increased as the frame length increased. This means that the longer frame length is preferable for extraction accuracy. The extraction accuracy by the spline recalculation method was smaller than that by the differentiating method. This means that the differentiating method could work in the environment where small sound deformation exists. Meanwhile, sound quality and bit-rate were better for shorter frame length. There is a trade off between sound quality/bit-rate and extraction accuracy along the frame length. Table 2 is effective bit-rates; they are obtained from theoretically maximum bit-rate B_r multiplied by the actual random binary data extraction accuracy in Figs. 5 and 6. When the watermarked signal was not degraded by MP3 compression, both extraction methods achieved high bit-rate more than 2.9 kbps without audible deterioration, using frame lengths of 15 and 11. When the watermarked signal was degraded by MP3 compression, the spline recalculaton method was able to achieve bit-rate of 1 kbps for a usual MP3 compression rate of 128 kbps, using frame length of 15.

Table 2. Effective bit-rates

(a) Spline recalculation method

	Frame length N [pt]						
	11	15	20	25	30	35	40
Uncompressed	4001.0	2940.0	2205.0	1764.0	1470.0	1260.0	1102.0
MP3 160kbps	216.5	1417.1	648.3	726.8	1140.7	990.4	897.0
MP3 128kbps	425.0	1181.9	463.1	656.2	714.4	635.0	504.7
MP3 112kbps	280.6	1087.8	635.0	705.6	635.0	735.5	758.2
MP3 96kbps	64.1	952.6	842.3	815.0	670.3	781.2	546.6
MP3 80kbps	48.1	652.7	657.1	486.9	773.2	682.9	504.7

[bps]

(b) Differentiating method

	Frame length N [pt]						
	11	15	20	25	30	35	40
Uncompressed	4001.0	2940.0	2205.0	1764.0	1470.0	1260.0	1102.0
MP3 160kbps	232.5	1828.7	1927.2	1647.6	1417.1	1232.3	1075.6
MP3 128kbps	40.1	940.8	1402.4	1400.6	1287.7	1169.3	1042.5
MP3 112kbps	40.1	752.6	1212.8	1291.2	1217.2	1126.4	1031.8
MP3 96kbps	16.0	535.1	979.0	1118.4	1102.5	1058.4	969.8
MP3 80kbps	16.0	323.4	802.6	1016.1	1014.3	997.9	932.3

[bps]

7. Conclusions

An information hiding method using spline interpolation was proposed. The corresponding information extraction methods were derived, for the cases that the original signal is available or not. Although conventional watermarking techniques keep sound quality using longer frame length, which makes bit-rate lower, the proposed method was able to keep sound quality and bit-rate higher at the same time using shorter frame length. When waveform deformation like MP3 compression occured, the minimum frame length for keeping sound quality and extraction ratio existed. As a result of computer simulations, the proposed method was able to embed additional binary information into sound with higher bit-rate of at least 1 kbps. Detailed robustness of the proposed method is remained further work.

References

[1] C. de Boor. *A Practical Guide to Splines*. Springer-Verlag, New York, 1978.

[2] Y. Fuu, Z. Ma, and G. Song. A robust audio watermarking algorithm based on wavelet transform. *Journal of Information & Computational Science*, 1(2):7–11, March 2005.

[3] M. Iwakiri and K. Matsui. Digital watermark scheme for high quality audio data by spectrum spreading and modified discrete cosine transform. *IPSJ Journal*, 39(9):2631–2637, September 1998.

[4] I. Muramatsu and K. Arakawa. Digital watermark for audio signals based on octave similarity. *IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences: A*, J87-A(6):787–796, June 2004.

Proceedings of the 2006 International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP'06)
0-7695-2745-0/06 $20.00 © 2006 IEEE

IEEE COMPUTER SOCIETY