Journal of Information Hiding and Multimedia Signal Processing ©2010 ISSN 2073-4212
Ubiquitous International Volume **1**, Number **2**, April **2010**

Real-Time Audio Watermarking Based on Characteristics of PCM in Digital Instrument

Kotaro Yamamoto

Graduate School of Science and Engineering
National Defense Academy
1-10-20 Hashirimizu, Yokosuka, Kanagawa, 239-8686 Japan
g46073@nda.ac.jp

Munetoshi Iwakiri

Department of Computer Science
National Defense Academy
1-10-20 Hashirimizu, Yokosuka, Kanagawa, 239-8686 Japan
iwak@nda.ac.jp

Received April 2009; revised January 2010

Abstract. *Musical performance with digital instruments has become a common practice today, and many digital instruments come to be used in Desk Top Music (DTM), live performance, etc. These performances are recorded as digital contents, and circulated actively through network and electronic media. However, the spread of digital contents causes a problem of illegal duplication and distribution, so that digital watermark has recently attracted much attention as a technology to solve this problem. In this paper, we focus on a sound synthesized process in digital instruments, and propose a real-time watermark method. Certain watermarks are embedded in wavetables that are included in our digital instruments, and the insertion of secret messages is actualized with wavetable switching. Additionally, embedded watermarks can be extracted from the acoustic signal. The proposed method is able to achieve a real-time watermark, i.e., both musical performance and the insertion of watermark can be actualized.*
Keywords: Audio Watermark, Digital Instrument, MIDI,

1. **Introduction.** In recent years, various kinds of information have been digitized as multimedia contents. Since digital contents could be duplicated easily, copyright infraction has occurred by the illegal spreading of digital contents. Digital watermarking has been attracted much attention to solve this problem, and many approaches have been made to construct copyright management systems[1, 2, 3, 4].

Audio watermarking has been proposed for the protection of multimedia contents, and it has been used for recording media such as MPEG 1 Audio Layer III (MP3) and Microsoft Windows Media Audio. Most of these watermarks are achieved with a non-real-time system, and there are many methods of different approaches in this type of system, *e.g.*, LSB (Least Significant Bit) substitution methods are most fundamental techniques for information hiding[5, 6], and amplitude modulation or phase shift methods in frequency domain are powerful tools for acoustic watermarking[7, 8]. This type of system uses previously recorded acoustic waveforms as cover data, and therefore, it might not be suitable for embedding watermark in real-time, and it would make difficult to use it for situations like live-performance, where the illegal recording of acoustic sound has easily been made. It is a serious problem, and therefore, real-time watermarking is required.

Although there are several methods to embed secret messages in real-time, they have another problem. Tachibana[9] focused on a situation of live performance, and watermarking has been achieved by amplitude modulation in frequency domain, with several composition methods in real-time. However, this method includes DFT (Discrete Fourier Transform) and IDFT (Inverse DFT) in the embedding procedure, and time-delay occurs during a large calculation between the host signal and watermarked signal. Hernandez, *et al.*[10] constructed a real-time watermarking device that uses MCLT (Modulated Complex Lapped Transform) for embedding, with hardware implementation. This approach tries to reduce the time-delay with embedding procedure. That is, in the conventional approach of real-time embedding, unnecessary delay occurs when sound outputs with embedding procedure, and the delay could not be avoided. This is the problem that remains to be solved.

In this study, we focused on digital instrument to solve this problem. Digital instrument (like electronic piano, electric drum and so on) has been used for musical performances in practice, and digital instrument has a characteristic structure in its sound synthesis process. Additionally it was considered that characteristics of the synthesis could be used for real-time watermarking.

The purpose of this study is to achieve a real-time watermarking technology with another approach for musical performance with digital instruments. In the proposed method, watermarks are embedded in the wavetable of digital instruments, and embedded data are extracted from the playback acoustic signal of digital instrument. Therefore, watermarks can be embedded in the output acoustic signal that is synthesized from the wavetable in real-time musical performance, and it can be useful for the copyrights protection of real-time generated acoustic media.

Section 2 describes the digital instrument and its structure, with a special focus on wavetable structure as Pulse Code Modulation (PCM) sound source. Section 3 contains the technical details about the proposed embedding algorithm to accomplish inserting watermark, and Section 4 shows an experimental result with software simulation to evaluate the practicality of the proposed method.

2. Digital instrument.

Digital instruments are classified into three types with its sound generation methods[11], PCM[1], functional generator[2] and acoustic modeling[3].

The PCM uses an acoustic signal of real instruments, but the others use computer-generated pseudo-waveform based on an acoustic model. Therefore, the PCM synthesis is widely used as digital instruments.

2.1. PCM sound synthesis.

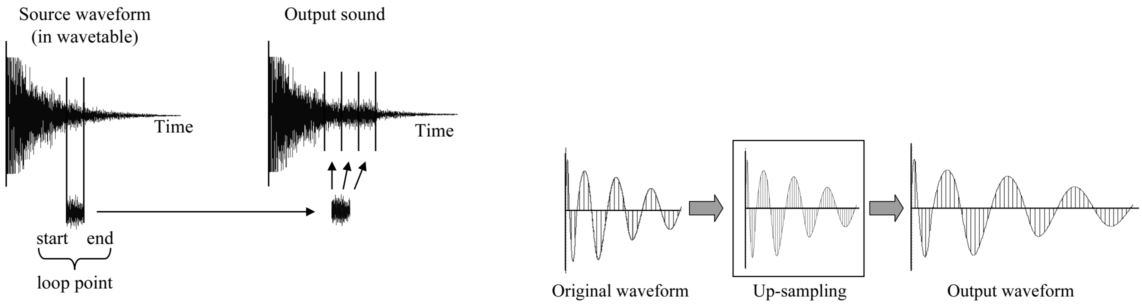
Basically, the PCM sound synthesis is simple. First, acoustic sounds of natural instruments are recorded as PCM sequences, and they are stored into digital instrument. Second, performer's operation is input from interface of the digital instrument. Finally, the PCM sequence that is matching with the performer's operation is read, and output with a little modification.

However, in some situations, the amount of data may be too large to construct instruments. For example, to make an acoustic piano requires many resources, *e.g.*, waveforms for each key and various velocities, and since acoustic waveform sizes are flexible, the

[1] Collected natural instrument sounds are used for sound generation.

[2] UG (Unit Generator) with a factor of acoustic-signal-processing. The UG has many functional processor or software algorithms.

[3] Acoustic models based on the structure of natural instruments to generate sound.



(a) Looping (b) Pitch shift in time-domain (ex. of up-sampling)

FIGURE 1. Sound synthesis with wavetable

instrument has to generate acoustic synchronizing for performer's intention correctly. General PCM synthesis uses "Looping[12]" and "Pitch-shift[13]" to reduce the size of a wavetable at correct sound generation.

(1) Looping.
Looping is used to expand the length of limited size waveforms, and it is functional to generate waveforms for long-term[12]. First, the loop-section is set at the waveforms (*i.e.*, start and end loop-point set). Then, the waveforms can be generated with repeating the loop-section (**Figure 1**-(a) refers).
(2) Pitch shift.
If a digital instrument has insufficient storage to store full tones, the number of stored tones are decreased in the instrument. In this case, a necessary musical pitch has to be generated from existent waveforms in the instrument[13]. There are two types of pitch shift, one is processing in frequency-domain, and another is processing in time-domain (Figure 1-(b) refers). In both cases, each frequency spectra (pitch) have been shifted to either higher or lower frequencies.

In this way, PCM sequences are modified within the sound synthesis procedure. Additionally, some parameters are essential to modify the source PCM sequences correctly. Therefore, PCM waveforms are collected with these parameters, and are stored as a "wavetable" in the instrument.

2.2. **Wavetable.** There are a number of parameters for PCM sound synthesis in a wavetable, *e.g.*, start and end point of the loop-section, fundamental frequency (f0) of the PCM sequence, coverage of playback frequency, and so on. They are collected with each PCM sequences, and stored in the wavetable.

One instrument is conformed like this, and the wavetable is loaded on the synthesizer for sound generation.

3. **Proposed scheme.**

3.1. **Instrument watermarking.** In section 2, it was described that PCM sound-sources are used in synthesizing acoustic signals with a wavetable. Although the PCM waveform of a wavetable has been modified for sound output during the sound generation, the synthesized PCM acoustic sound maintains the frequency spectra of an original wavetable. We focused on the characteristics of sound generation, and an attempt was made to develop a watermarking method using digital instruments.

The basic scheme of our technical proposal shows as follows (see **Figure 2**).

Step 1: Watermark has been embedded in PCM sources of wavetable.

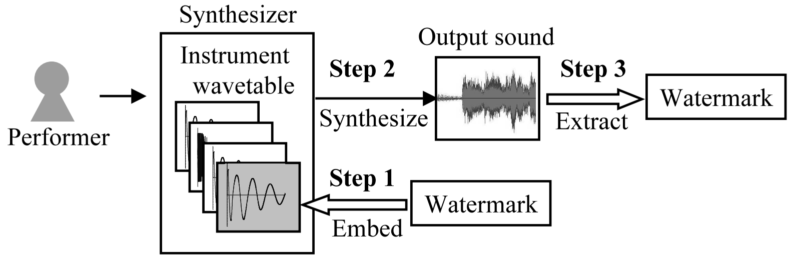


FIGURE 2. Outline of the proposed method

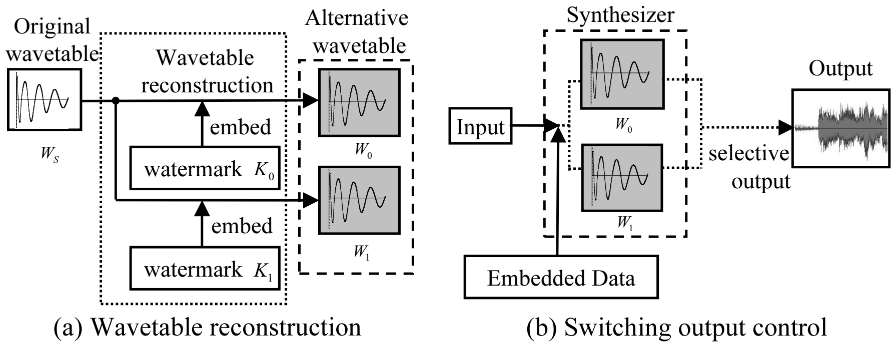


FIGURE 3. Systematic flow of the proposed scheme

Step 2: Playing music with the instrument that includes a watermarked wavetable, and the output acoustic signal is recorded.

Step 3: Watermark includes all the output acoustic signals.

In this way, watermarks are automatically inserted into music in real-time.

3.2. **Wavetable switching.** Certain marker signals are embedded in a waveform on a wavetable, and embedded signals can be extracted from the output acoustic signal by the proposed method.

Additionally, a number of marker signals are embedded in a certain wavetable, and number of alternative wavetables are generated, so that these marker signals can be observed from the acoustic waveform.

Thereupon, we construct the instruments that waveforms can be switched at each output, if its selection is controlled by a special key, and then the special components can reveal any information as a watermark. **Figure 3** shows the systematic flow of the proposed scheme.

(1) Wavetable reconstruction.
In the proposed method, alternative wavetables have been used to reveal information, and in this process, a number of wavetables have been generated from the source wavetable by embedding marker signal. Two alternative wavetables W_0 and W_1 are prepared from one source wavetable W_S for watermarking as 1 bit expression.

(2) Output control based on embedded data.
Secret messages would be embedded as described in the following. First, a control code as each input is received at the interfaces of instruments, a bit E is taken from a watermark message at the same time. Second, the output waveform of the wavetable is selected to conform

$$W_E = \begin{cases} W_0 & (E = 0) \\ W_1 & (E = 1). \end{cases} \quad (1)$$

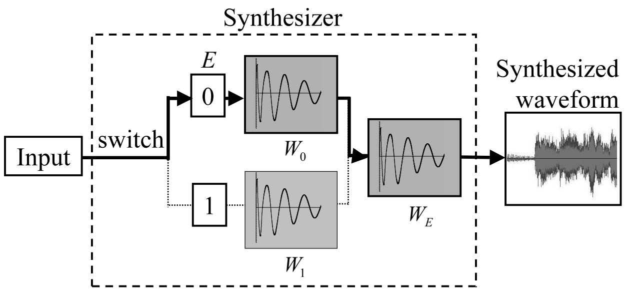


FIGURE 4. Switching wavetable ($E = 0$)

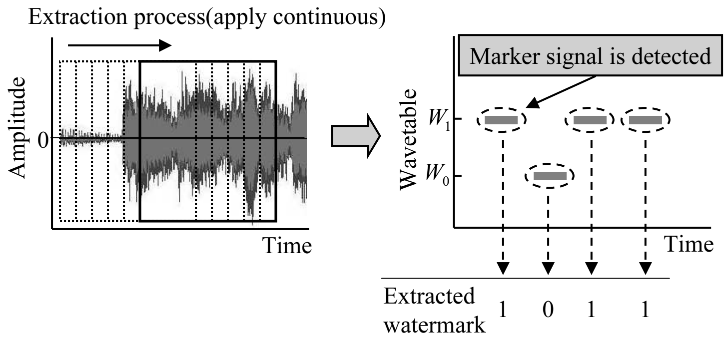


FIGURE 5. Processing flow of extraction

Finally, acoustic signals are synthesized using W_E. **Figure 4** shows an example of the insertion processes of a watermark with this wavetable switching at "$E = 0$".

(3) Watermark extraction.
In the previous process, the embedded data are expressed as wavetable alternation, and we have to distinguish each watermarked wavetable with time-shift and continuous analysis to extract embedded data. **Figure 5** shows a continuous extraction procedure for identification of the embedded watermarks. If any marker signal is detected, the wavetable W_d which is used in "wavetable switching" can be determined, and extracted data E can confirm

$$E = \begin{cases} 0 & (W_d = W_0) \\ 1 & (W_d = W_1). \end{cases} \tag{2}$$

As remarked above, the proposed method uses alternative wavetables to reveal embedded data, and information embedding would be processed in each tone. Therefore, embedding payload of the proposed method depends on the number of alternative wavetables, *i.e.*, if the number of alternative wavetables is 2^n, n bit data can be embedded in each tone.

In the proposed scheme, a marker signal is embedded in a waveform beforehand, and watermarks can be embedded only by switching of alternative wavetables in real-time. Therefore, the time-delay is negligibly small as compared with that of the conventional method.

3.3. **Consideration about marker signal.** In the proposed scheme, the embedded marker signal must be detected from an output waveform. However, there exist some problems to detect the marker signal. It is because the inner waveform modification is held during sound synthesis procedures and output waveforms can be attacked for removal of watermarks from outside.

(1) Inner waveform modifications.
There exist amplitude modification, pitch-shift, and looping during the sound synthesis process. Looping and pitch-shift are used to reduce the size of storage as described in section 3, and therefore, a waveform is transformed into an output acoustic waveform in these processes.

(2) Watermark removal attacks.
Generally, watermarked waveforms are modulated to remove the embedded watermarks, and there are many types of attacks, *e.g.*, amplitude modulation, adding noise, linear data compression, frequency modulation, application of band-pass filter, and so on.

The marker signal ought to be robust against these modifications of attacks, and additionally, the marker signal must be imperceptible to human auditory systems. Therefore, it is desirable to use a certain watermarking technique as a marker signal.

In this study, we used "Spread Differential Method (SDM)" as the watermarking based on a frequency hopping method[14].

3.4. **Spread Differential Method.** The SDM (Spread Differential Method) uses pairs of samples, which are selected randomly from target sequences. The method reveals information on the statistical distribution of target sequences. The outline of SDM is as follows (see **Figure 6**).

The target sequences are discrete frequency coefficients $X_i(k)$, $1 \le k \le \frac{N}{2}$ calculated from i^{th} acoustic frame, which constituted N samples. A pair of coefficients $X_i(a_i)$ and $X_i(b_i)$ are selected randomly from the sequence, and differential d_i is calculated from these values.

$$d_i = X_i(a_i) - X_i(b_i), \quad 1 \le \{a_i, b_i\} \le \frac{N}{2}. \tag{3}$$

The calculations are repeated with other pairs to take the frequency $y(d_i)$ of d_i. The distribution of $y(d_i)$ is similar to that of Figure 6-(a) which has a single-peak at $d_i = 0$.

Otherwise, random numbers u_i and v_i are added to $X_i(a_i)$ and $X_i(b_i)$ as watermarks respectively, and these modified coefficients are renamed as $X_i'(a_i)$ and $X_i'(b_i)$. Differential d_i' of these two values is as follows.

$$\begin{aligned} d_i' &= X_i'(a_i) - X_i'(b_i) \\ &= X_i(a_i) + u_i - \{X_i(b_i) + v_i\} \\ &= d_i + u_i - v_i. \end{aligned} \tag{4}$$

Then, the distribution $y(d_i')$ is conformed with Figure 6-(b), which is similar to Figure 6-(a), but the peak of $y(d_i')$ is remarkably lower than that of $y(d_i)$. This is because d_i' was modified with addition of u_i and v_i, and random nature was increased.

In other words, if the pairs of samples a_i and b_i have been selected by a special key and the pairs are modified by random numbers, the frequency distribution of the pairs' differential, $y(d_i')$ would be distinct from that of another pairs' differential $y(d_i)$. That is to say, any information can be revealed by utilizing differential of $y(d_i)$ and $y(d_i')$.

In this study, a psychoacoustic model[15] is also used to intense the embedding in order to keep robustness without deterioration of sound quality.

3.5. **Psychoacoustic model.** In human auditory system, sensitivity of the fixed strength of acoustic waveform differs with its frequency. Absolute threshold[4] is the differentials of sensitivity in experimental measurement. Furthermore, at the simultaneous masking,

[4] The frequency threshold of human would be perceptible at quiet

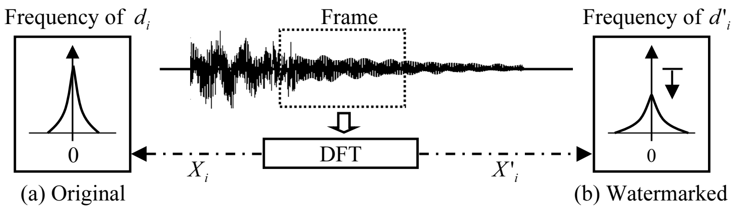


FIGURE 6. Outline of Spread Differential Method

the loud-acoustic signal (masker) masks the quiet-acoustic signal (maskee) around the masker, and the maskee would be hardly heard by the masker.

A numerical psychoacoustic model of the auditory property is used to calculate a "global masking threshold" in a certain acoustic frame.

In this study, MPEG 1 Audio psychoacoustic model 1 for layer 2 (ISO-11172 Part 3[15]) was used for psychoacoustic analysis, and it was implemented with a little modification. That is, sub-band analysis was not used, and global masking threshold[5] was calculated for each frequency spectra.

3.6. **Wavetable Reconstruction by SDM.** In this section, a method of wavetable reconstruction by SDM is described. So as to avoid the influences of looping process, the watermark has been embedded in first 0.5 seconds of the waveform, which is the section of attack in output tone. This watermark is embedded in every frame which is constituted by N samples, but to reduce inter-frame distortion, the frames have to be overlapped at half a frame length. This study used $N = 1024$ for the frame-size on sampling ratio $R_S = 44100$[Hz].

Step 1: N samples $x_i(k)(k = 0, \cdots, N-1)$ are taken from wavetable W_S as i^{th} frame.

Step 2: Applying DFT, to produce $f_i(k)$ as frequency spectra of sequence $x_i(k)$ in time-domain. The $x_i(k)$ and $f_i(k)$ are complex number sequences.

Step 3: Apply the psychoacoustic analysis with $f_i(k)$, and calculate the global masking threshold $LT_q(m)$, $(1 \le m \le \frac{N}{2})$ at the frame.

Step 4: Amplitude components a_i and b_i, $(1 \le \{a_i, b_i\} \le \frac{N}{2})$ are determined for embedding, these are extracted from a random number generator with a watermark key K_0. Then, the amplitudes of components $|f_i(a_i)|$ and $|f_i(b_i)|$ are replaced by

$$|f_i'(a_i)| = |f_i(a_i)| + u_i \tag{5}$$

$$|f_i'(b_i)| = |f_i(b_i)| + v_i. \tag{6}$$

Note: The control intensities $\{u_i : |u_i| \le T(a_i)\}$ and $\{v_i : |v_i| \le T(b_i)\}$ are random numbers. The function $T(k)$ is

$$T(k) = \begin{cases} 0 & (LT_q(k) < |f_i(k)|) \\ LT_q(k) - |f_i(k)| & (|f_i(k)| \le LT_q(k)). \end{cases} \tag{7}$$

The $|f_i'(k)|$ has been set to zero when $|f_i'(k)|$ is negative. Repeat p rounds at the process of this step.

Step 5: Watermarked waveform x_k' would be recomputed with modified spectra $f_i'(k)$ and IDFT function.

Step 6: Cross fade window function

$$c(k) = \begin{cases} \frac{2k}{N} & (0 \le k \le \frac{N}{2}) \\ 2 - \frac{2k}{N} & (\frac{N}{2} < k \le N). \end{cases} \tag{8}$$

[5]In this study, we calculated global masking threshold from absolute threshold and individual masking threshold to each spectra, and used for determination of embedding intensity.

is applied to overlapped acoustic frames with a triangle window at additive synthesis, and windowed sequences $w_i(k)$ are confirmed

$$w_i(k) = x_i'(k) \times c(k). \tag{9}$$

Then, $w_i(k)$ is outputted with additive synthesis.

Step 7: The capture range of a DFT frame is shifted forward by every half length, and repeat Step 1 to Step 6 until the end of embedding section.

Step 8: Output the watermarked waveform W_0.

Step 9: W_1 is generated similarly from W_S with the processing used key K_1.

In these processes, a bit datum is inserted into W_S by "output control based on embedded data" (see Section 3.2–(2)).

3.7. **Watermark Extraction by SDM.** The statistical distribution of pairs' differentials ought to be distinct between embed and non-embed in SDM. Inspection with a number of keys would be necessary to detect the distribution distinction. In other word, embedded data need to be extracted by comparison with $y(0)$ calculated with each number of keys. The peak of $y(0)$ calculated with the watermarked key would be lower than others by inspection.

The outline of $y(0)$ inspection with a key K is as follows.

Step 1: N samples $s_i(k)(k = 0, ..., N-1)$ are taken from an output acoustic waveform as i^{th} frame. Then the samples are applied with DFT, and $F_i(k)$ is produced as the frequency spectra of sequence $s_i(k)$ in time-domain. Both $s_i(k)$ and $F_i(k)$ are sequences of complex numbers.

Step 2: The embedding amplitudes $|F_i(a_i)|$ and $|F_i(b_i)|$ $(1 \le \{a_i, b_i\} \le \frac{N}{2})$ are determined, and the differential d_i of the two components' amplitudes is calculated as follows.

$$d_i = |F_i(a_i)| - |F_i(b_i)|. \tag{10}$$

Note: a_i and b_i are random numbers generated by key K.

Then, d_i is rounded down to a nearest integer, and the frequency of d_i is counted as $y(d_i)$. This step is repeated for p rounds.

Step 3: The processing frame is shifted forward, and Step 1 and 2 are repeated until the block end. That is, the frame has reached the divided time length from analysis heading.

Step 4: Output the $y(0)$ as inspection result.

Each inspection results are compared, and the W_0 is outputted at the section of the K_0 used, and the W_1 is outputted at the K_1 used. This process is carried out by time-shift (see Section 3.2–(3)).

4. **An implementation and evaluation.** The proposed method is implemented as software simulation, and is evaluated from three perspectives, correct extraction, robustness and sound quality.

4.1. **Experimental system.** This study was performed with a PCM synthesizer implemented to a software called "TiMidity++[6]". We used a GUS patch (Gravis UltraSound patch) which is distributed by freepats[7] as a PCM sound-source. **Figure 7** shows the structure of the practical system, and the each part is shown as below.

(1) Input.

Standard MIDI File (SMF)[16] is used as sound control codes in this simulation.

[6]http://sourceforge.net/projects/timidity/

[7]http://freepats.opensrc.org/

(2) Sound synthesizer.
Sound is generated through next two procedures.
Step 1: SMF is modified with a watermark as a pre-process (addition of program change, and the channel number of note messages is rewritten).
Step 2: Output acoustic signal is generated by TiMidity++ that is controlled by the modified SMF.

(3) Sound output.
Acoustic waveform is stored as output signals in TiMidity++. Digital sound recording is implemented as this process.

Modifying SMF at the Step 1 of the process (2) is equivalent to the waveform modification in this system, The characteristic of TiMidity++ is used in the process of embedding watermarks effectually.

The characteristic of TiMidity++ is that instruments can be registered arbitrary in configuration. Instruments can be changed by the channel number of note message. It means that information is embedded at the pre-process stage by modifying SMF, *i.e.*, switching channel of instrument is equivalent to wavetable switching.

4.2. **Extraction test.** Some correct extraction tests were held whether the proposed method worked correctly or not. The purpose of this test is to clarify the basic effectiveness of the wavetable switching method.

4.2.1. *Synthesis conditions.*

(1) Experimental instruments and embedding conditions.
The proposed method was evaluated with four instruments (wavetables), that is, acoustic-piano, jazz-guitar, flute and trumpet (see **Table 1**). Additionally, watermarks with $K_0 = 188$ and $K_1 = 27$ were embedded in waveforms of each instrument respectively, and W_0 and W_1 were generated. These waveforms were registered in TiMidity++ as different instruments. Repeating number p is 10 ($p = 10$) at Step 4 in Section 3.6.

(2) Experimental phrase.
A phrase was used for this test. It has five notes with no pitch-shift, and all notes were existent in the wavetable (see **Figure 8**).

(3) Sound synthesis conditions.
Sound synthesis with various volumes was held, that is, amplitudes of the output waveforms were 20, 40, 60, 80 and 100 percent of the original amplitude in wavetables. In this test, the bits of "00101" were embedded with wavetable switching.

4.2.2. *Result and considerations.* **Figure 9**-(a) is an original waveform of the piano instrument, and Figure 9-(b) is a watermarked waveform in the instrument which is embedded with key $K_0 = 188$. Figure 9-(c) is the output acoustic waveform played with the embedded instrument (sampling ratio is 44.1kHz, and signed 16bit quantization). In the waveform, the amplitude was modified to 60 percent of the original waveform.

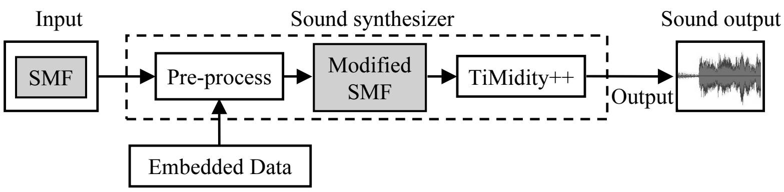


FIGURE 7. Structure of experimental system

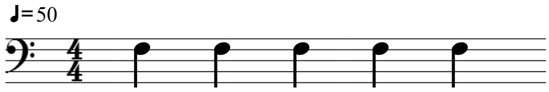
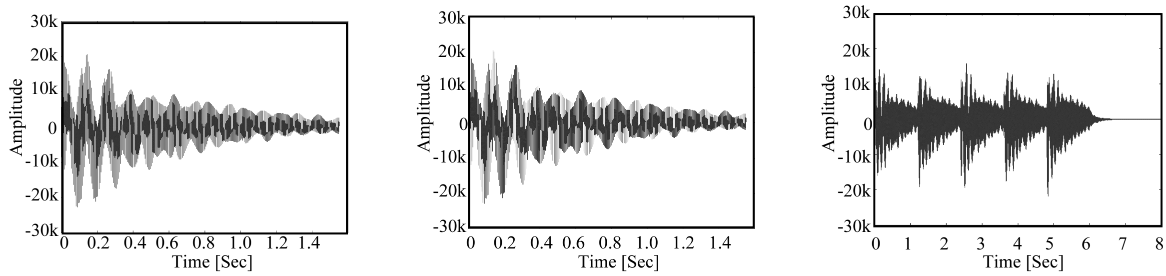
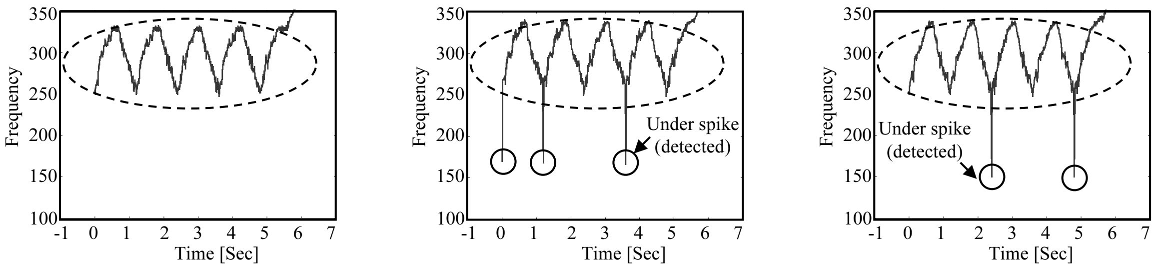


FIGURE 8. Experimental phrase



(a) Original waveform in wavetable (b) Watermarked waveform in wavetable (c) Output acoustic waveform

FIGURE 9. Results of waveform reconstruction and sound output (Piano)



(a) Inspection with unused key (b) Inspection with K_0 (c) Inspection with K_1

FIGURE 10. Inspection results of Piano

Additionally, **Figure 10**-(a), (b) and (c) show results of inspection after analyzing Figure 9-(c) output waveform at an unused key, watermarked key K_0 and K_1. From comparison of the under peak with Figure 10-(a) and (b), remarkable spikes are observed at time t = 0, 1.2 and 3.6[sec] with key K_0 (Figure 10-(b)). And from comparison of the under peak with Figure 10-(a) and (c) in a similar way, spikes are observed at time t = 2.4 and 4.8[sec] with K_1 (Figure 10-(c)). These spikes show that the embedded data "00101" are extracted correctly with "Piano".

The results with all instruments are shown in **Table 2**. From the results, some considerations are made as below.

First, watermarks were extracted correctly in performance with all instrument, and the result shows that the proposed method works effectively as real-time watermarking.

TABLE 1. Experimental instruments

Sample name	Patch file name	Included waveforms
Piano	000_Acoustic_Grand_Piano.pat	10
Guitar	026_Jazz_Guitar.pat	13
Flute	073_Flute.pat	4
Trumpet	056_Trumpet.pat	6

TABLE 2. Extraction results of extraction test

Amplitude [%]	20	40	60	80	100
Piano	○	○	○	○	○
Guitar	○	○	○	○	○
Flute	○	○	○	○	○
Trumpet	○	○	○	○	○

Second, embedded bits were extracted on all amplitude in this test, and therefore, it can be said that the proposed method can work correctly in sound synthesis of amplitude modification. This method is robust with at least 20 percent amplification from experimental results because applying the psychoacoustic model has strengthened the embedding intensity.

Third, a marker signal was embedded in the heading of waveform in the proposed method, and the influences of looping can be avoided. It is because, in general, the loop section is located near the tail of the waveform.

Finally, in this implementation, embedding with wavetable switching was applied only in the note that compliant waveforms were present in wavetable, and watermarks were not embedded in other notes[8]. It is because the SDM provides little robustness against pitch-scale modification, that is, marker signal detection from the waveform that is generated with "pitch-shift" process is difficult. Future tasks are compliant to "pitch-shift". The robustness against pitch-scale modification might be necessary in the marker signal for this task.

4.3. **Robustness.** For the evaluation of robustness as watermarking technique, some attacks were held to a watermarked acoustic waveform by the proposed method, and it was checked whether extraction of watermarks could be succeeded or not. Target waveforms were output waveforms of the performance with experimental phrase by each instrument, and the playback amplitude was 60 percent of the original waveforms.

Experimental attacks are shown in the followings.

(1) Linear data compression (MP3, AAC).
The ratios of 64kbps, 96kbps, 128kbps and 192kbps were used in MPEG1 Audio layer 3 (MP3). The ratios of 96kbps, 128kbps and 192kbps were used in MPEG2 AAC (AAC).

(2) Down sampling (Down).
The sampling ratio was down-sampled to 22.05kHz.

(3) Reduction of quantization bit rate (Quant).
The resolution of samples was modulated to 8bit.

(4) Adding white-noise (Noise).
Continuous 53.8dB SPL (Sound Pressure Level) white-noise was added. A normalization of the reference level of 96 dB SPL is performed in such a way that the maximum value of sample corresponds to 96dB.

The experimental results are shown in **Table 3**. The feature common to all observations is the robustness against addition of white-noise, amplitude modification and down sampling. Robustness against addition of white noise and amplitude modification were good enough in all instrument. Embed components were controlled with a key in the proposed method. Because of this frequency hopping, there is robustness essentially against attacks of passing a band-pass filter.

[8] In experimental phrase, there exist waveforms for all note in wavetable.

But in contrary, the proposed method would not have robustness against down sampling, and this is because the frequency spectra over 11.025kHz were erased with the down sampling procedure, hence SDM did not work correctly. The robustness can be considered to improve by the multiplexing of embedding with a limited bandwidth.

The robustness against linear data compression showed a different tendency with instruments. In piano and guitar, the robustness was good in comparison with others, and the robustness was good enough except the compression with low bit-ratio. This is because the cut-off frequency of low-pass filter was lower in low bit-ratio compression. In contrast, there was little robustness against linear data compression in flute and trumpet. Although it is caused by the difference in tuned sounds and sustained sounds, the detailed research is a future task.

4.4. **Sound quality.** In this study, the sound-source identification test was made by an ABX double blind test with 9 raters whose ages were twenties and thirties. The ABX test assumes that acoustic waveform without embedding (played by original instruments) is labelled as A, and performance with embedding of the proposed method (played by watermarked instruments) is labelled as B. First, these performances are shown to the raters. Then, A or B is shown as the performance X to each rater randomly, and the raters evaluate X as either A or B. The accuracy ratio of X has a biased-value from 50 % when A and B can be distinguished clearly by listening, and accuracy ratio is defined as 50 % when these performances cannot be distinguished. However, even when A and B cannot be identified clearly, the accuracy ratio has a biased-value because the number of trials is limited. In this study, significance of the accuracy ratio was investigated by the χ^2 test.

In the experiment, X was shown so that X could not be distinguished as A or B. One cycle of judging X from A and B assumed one trial, and 5 times of trials were carried out. In evaluation, X was selected at every trial randomly. A monitor headphone ATH-PRO5V (made from Audio Technica) was used for listening of raters. The evaluation results of all raters were summarized, and tested as a result of 45 trials in the χ^2 test.

The results are shown in **Figure 4**. From the results, there is no significant difference between watermarked waveforms and original waveforms with 10 % of significant level. It can be said that the raters in this experiment were not able to distinguish A or B in the result. That is, information can be embedded by the proposed method without degrading quality of performance. It is because embedding procedure of the proposed method is based on the psychoacoustic analysis.

5. **Conclusion.** In this paper, we proposed an information hiding method as a real-time watermarking technique in musical acoustic signals.

Our watermark is inserted in the wavetable of a sound synthesizer separately, and the watermarked waveform was generated automatically with a selective output of the

TABLE 3. Robustness of the proposed method

(a) Robustness against linear data compression

Sample	MP3 [kbps]				AAC [kbps]		
	64	96	128	192	96	128	192
Piano	×	○	○	○	×	○	○
Guitar	×	○	○	○	×	○	○
Flute	×	×	○	○	×	×	×
Trumpet	×	×	×	×	×	×	×

(b) Robustness against other attacks

Sample	Down	Quant	Noise
Piano	×	○	○
Guitar	×	○	○
Flute	×	○	○
Trumpet	×	○	○

Table 4. χ^2 test result of the proposed method

sample	correct[%]	χ^2	p
Piano	51.1	0.022	0.881
Guitar	62.2	2.689	0.101
Flute	46.7	0.200	0.655
Trumpet	60.0	1.800	0.180

instruments in real-time. Therefore, the proposed method can be useful technique for rights management technology in such situations as illegal recording on live performance.

REFERENCES

[1] S. Katzenbeisser, and F. A. P. Petitcolas, *Information Hiding Technique for Steganography and Digital Watermarking*, Artech House, 2000.

[2] I. J. Cox, J. A. Bloom, and M. L. Miller, *Digital Watermarking*, Morgan Kaufmann Publishers, 1999.

[3] F. Wang, J. Pan, and L. C. Jain, *Innovation in Digital Watermarking Techniques*, Springer, Berlin-Heidelberg, Germany, 2009.

[4] J. Pan, H. Huang, and L. C. Jain, *Information Hiding and Applications*, Springer, Berlin-Heidelberg, Germany, 2009.

[5] D. Gruhl, A. Lu, and W. Bender, Echo hiding, *Proc. of Information Hiding: First International Workshop*, Springer, vol. 1174, pp. 7-21, 1996.

[6] M. Iwakiri, and K. Matsui, Visualizing technique of digital watermarks to audio data, *Proc. of IPSJ*, vol. 41, no. 6, pp. 1840-1847, June 2000.

[7] W. Bender, D. Gruhl, and N. Morimoto, Techniques for data hiding, *IBM Systems Journal*, vol. 35, no. 3-4, pp. 131-136, 1996.

[8] N. Yajima, and K. Oishi, Digital watermarks for audio signals based on amplitude and phase coding, *IEICE Technical Report*, vol. 104, no. 720, pp. 53-59, May 2005.

[9] R. Tachibana, Audio watermarking for live performance, *Proc. of Security and Watermarking of Multimedia Contents V*, Santa Clara, USA, vol. 5020, pp. 32-43, January 2003.

[10] J. J. G. Hernandez, M. Nakano and P. M. Hector, Real-time audio watermarking system prototype, *Proc. of the 8th IEEE International Symposium on Multimedia (ISM'06)*, pp. 792-793, 2006.

[11] C. Roads, *The Computer Music Tutorial*, MIT Press, 1996.

[12] M. Collins, Infinity: DSP sampling tools for Macintosh, *Sound on Sound*, vol. 9, no. 1, pp. 44-47, 1993.

[13] A. Duncan, and D. Rossum, Fundamentals of pitch-shifting, *Proc. of the 85th Convention of the Audio Engineering Society*, New York, Preprint 2714, 1988.

[14] M. Iwakiri, and K. Matsui, Watermarking technique for digital sound data using the reactivity regulation scheme, *Proc. of IPSJ*, vol. 43, no. 8, pp. 2519-2528, 2000.

[15] ISO/IEC: Information technology-coding of moving picture and associated audio for digital storage media at up to about 1.5 mbit/s-Part 3: Audio, Tech. Rep., 11172-3, ISO/IEC, 1993.

[16] AMEI: *MIDI 1.0 Specification*, Ritto Music, 1998.