Digital watermarking method based on STFT histogram

Kotaro Sonoda
Division of Electrical Engineering and Computer Science
Graduate School of Engineering, Nagasaki University
Nagasaki, Japan
sonoda-iihmsp13@cis.nagasaki-u.ac.jp

Aleksander Sęk
Institute of Acoustics, Faculty of Physics
Adam Mickiewicz University
Poznan, Poland
oleksek@amu.edu.pl

***Abstract*—This paper presents a digital audio watermarking method. In the method, watermark has been represented by a form of the histogram for sound power level's distribution in time domain in certain frequency band(s). The relation of four numbers of selected bins in a histogram are changed by a rule of watermark bits assignment. To evaluate the performance of the embedding procedure in each frequency band, the robustness test against several kinds of manipulations or attacks were conducted for some instrumental music sources and harmonic complex tone. Considering the results of robustness test, the performance of multi-band embedding is discussed. Promising results of hearing tests are also presented.**

***Keywords*-Audio watermarking, Short term Fourier transform, Histogram.**

I. Introduction

In general, data hiding is a strategy enabling re-encoding an original audio signal. It takes an advantage of redundancy of audio signals to embed some payloads in it and to produce so called stego audio signal. These techniques also enable reading the payloads for the re-encoded audio signals.

There are lots of works dealing with watermarking methods. However, in most of them successive bits of payloads are embedded serially at a constant rate. There are also some methods that deal with constant length time frames. However, they need some frame synchronization and the same length of time frames for watermarked signals at both the embedding side and the detection side.

Zhang [1] proposed a digital audio watermarking technique based on segmenting of the histogram. He showed that the standard deviation of the signal (time-domain) before and after attack is an invariant parameter. In his watermarking algorithm, payloads are embedded by changing the mutual relations between four consecutive bins in the histogram. The way that the relations are changed depend on bits of payload. This is determined by the value of $\beta(n)$ defined in the following equation (1):

$$\beta(n) = \frac{h(4n) + h(4n+3)}{h(4n+1) + h(4n+2)} \tag{1}$$

where, $h(4n), \dots h(4n+3)$ denote the values of the histogram in bin (index) $4n, \dots 4n+3$. If n-th message bit $= 1$ is needed to embed, the values of time samples should be changed to get $\beta(n) > th_1$. In case of embedding '0', the values of time samples should be changed to get $\beta(n) < th_0$.

It turned out the Zhang's method can keep the robustness against many different attacks such as time scale modifications, volume change, MP3 coding, re-sampling and re-quantization. Moreover, it is also resistant against the multiple LPF attacks up to 17 bit per second. However, in Zhang's method, the length of segments must be long enough to get stable relation between values of successive bins in the histogram (to get stable $\beta(n)$). Zhang suggested that the segment length should be longer than 2.5 seconds to resist against the A/D and D/A attack.

In this present paper, an audio watermarking method based on Zhang's method is described and initially tested. However instead of embedding watermark into time-domain only, the proposed method embeds a watermark into time-frequency domain.

The paper is organized as follows. In section II, key ideas of the proposed method as well as watermarking embedding and detecting are described. Experimental results, include robustness and hearing tests are given in section III. In section IV, multi-band embedding was tested. Conclusions and proposals of further works are given in section V.

II. Watermark Algorithm

In this section, the key ideas are presented followed by a description of embedding and detection algorithm.

A. Basic key ideas

1) Bit assignment: In Zhang's method, all bits of a payload are embedded simultaneously into a single segment. Suppose N-bit payload is embedded. First, the mean and standard deviation of magnitude variation is calculated. Then values of bins for $4N$ classes in the histogram lying within ±one standard deviation are changed to meet predefined relation as follows from algorithm 1

However, the changes in the values of bins in the histogram affect the mean and standard deviation which are used to detect the watermark. In the proposed method, however, quite different class assignment in the histogram is used. Four classes which are used to determine the $\beta(n)$ are assigned orderly from mean as follows.

Algorithm 1 Zhang's embedding algorithm

```
$N$:number of bits for embedding to one segment
$M$:interval between classes in histogram
$th_0$ and $th_1$: thresholds for embedding '0' and '1'
$M \Leftarrow 2\sigma/(4N), \quad n \Leftarrow 0$
for $n < N$ do
    $k \Leftarrow 4n, \quad \beta(n) \Leftarrow \frac{h(k)+h(k+3)}{h(k+1)+h(k+2)}$
    if $m(n) = 1$ then
        if $\beta(n) < th_1$ then
            $x(i) \in h(k+1) \Leftarrow x(i) - M$
            $x(i) \in h(k+2) \Leftarrow x(i) + M$
        end if
    end if
    if $m(n) = 0$ then
        if $\beta(n) > th_0$ then
            $x(i) \in h(k) \Leftarrow x(i) + M$
            $x(i) \in h(k+3) \Leftarrow x(i) - M$
        end if
    end if
    $n \Leftarrow n + 1$
end for
```

$$ll_n = \lceil (L+1)/2 \rceil - 3n + 1, \quad lr_n = ll_n + 1 \tag{2}$$

$$rr_n = \lfloor (L+1)/2 \rfloor + 3n - 1, \quad rl_n = rr_n - 1 \tag{3}$$

$$\beta(n) = \frac{h(ll_n) + h(rr_n)}{h(lr_n) + h(rl_n)} \tag{4}$$

where $ll_n \sim rr_n$ are class numbers and L is overall number of classes. $h(k)$ is the histogram values in the class k. This way of class assignment is expected to make the changes in the mean and standard deviation values much smaller, giving also smaller amplitude fluctuations.

2) 2-D embedding: In the proposed algorithm, moreover, a payload is embedded in time-frequency domain. Two-dimensional spaces were created by means of sequences of power spectra calculated using short-term Fourier transforms (STFT). Then, a histogram of time domain was determined for selected frequency bands. However bandwidth of the selected band was composed of adjacent Equivalent Rectangular Bandwidths ($\mathrm{ERB_N}$). The number of the $\mathrm{ERB_N}$ as well as the lower cut-off frequency of the lowest $\mathrm{ERB_N}$-number was a parameter of the algorithm and it was tested during hearing test.

The $\mathrm{ERB_N}$ is broadly accepted measure used in psychoacoustics that determines the bandwidth of the human auditory filter [2]. From one hand, the bandwidth of these filter is a monotonically growing function of frequency. However, on the other hand, the frequency can expressed in terms of number of $\mathrm{ERB_N}$. A formula relating $\mathrm{ERB_N}$-number to frequency f [Hz] is given by the following equation (5):

$$\mathrm{ERB_N} - \text{number} = 21.4 \log 10(4.37f/1000 + 1) \tag{5}$$

B. Embedding algorithm

The embedding process, as a block diagram is depicted in Fig. 1. As it can be seen in the figure it consists of the following stages:

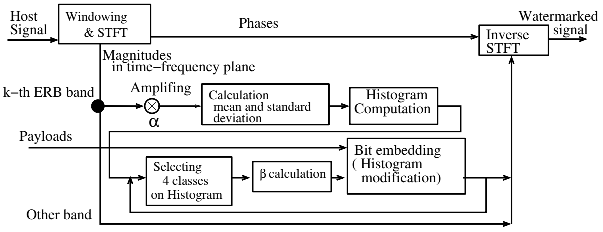

Host Signal
Windowing & STFT
Phases
Watermarked signal
Inverse STFT
Magnitudes in time−frequency plane
k−th ERB band
Amplifing
α
Calculation mean and standard deviation
Histogram Computation
Payloads
Bit embedding (Histogram modification)
Selecting 4 classes on Histogram
β calculation
Other band

Figure 1. Embedding process

1) Signal conversion to time-frequency domain by means of short-term Fourier transform (STFT): Initially, the STFT is applied to the host signal. A single frame of the STFT lasted 25 ms and 200 frames are analysed. Based on the data for 200 frames, it was possible to create histogram of magnitude $m(t)$ (t: frame number) across the frames for each component. However the histograms were constructed for components lying within the bandwidth determined by three adjacent $\mathrm{ERB_N}$-number.

2) Calculation of logarithms $M(t)$ of the magnitude $m(t)$ in each histogram bin:

$$M(t) = \log 10(m(t)) \tag{6}$$

3) Calculation of the mean and the standard deviation of all histogram bins, and then normalization of all bins:

$$\tilde{M}(t) = \frac{M(t) - \bar{M}}{V_M} \tag{7}$$

where $\bar{M}$ is the mean of all $M(t)$ and V_M is its standard deviation.

Normalized magnitudes from the range of -1 to +1 were assigned to $L = 6N + 1$ histogram classes (where N is the number of bits of a payload) giving distance between bins $D = 2/L$.

4) Embedding payload to histogram: Four successive classes from each histogram were selected and the values of $\beta(n)$ were calculated by equation (4).

Then, each $\beta(n)$ was normalized to $\beta(n)'$ according to the following role defined for one bit of payload $w(n)$,

$$\beta(n)' \begin{cases} > th_1, \Leftrightarrow w(n) = 1 \\ < th_0, \Leftrightarrow w(n) = 0 \end{cases}, \tag{8}$$

where, th_1 and th_0 are two selected thresholds for embedding bit '1' and '0'. In this paper, th_1 is set to 100 while th_0 to 1/100.

Embedding bit when $w(n)$ is '1'

If $\beta(n)$ is greater than th_1, no operation is required. Otherwise, the number of samples of $h(ll_n)$, $h(lr_n)$, $h(rl_n)$, and $h(rr_n)$ is fixed to satisfy $\beta(n)' = \frac{h(ll_n)' + h(rr_n)'}{h(lr_n)' + h(rl_n)'} > th_1$. Therefore, I_{lr_n} and I_{rl_n} samples within $h(lr_n)$ and $h(rl_n)$ class were extracted and moved to $h(ll_n)$ and $h(rr_n)$ class as follows:

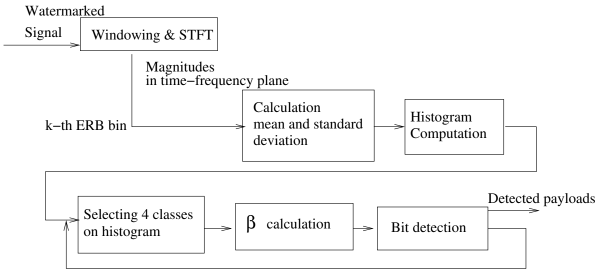


Figure 2. Detection process

$$h(ll_n)' = h(ll_n) + I_{lr_n}, h(lr_n)' = h(lr_n) - I_{lr_n} \tag{9}$$

$$h(rl_n)' = h(rl_n) - I_{rl_n}, h(rr_n)' = h(rr_n) + I_{rl_n} \tag{10}$$

$$\beta(n)' = \frac{h(ll_n) + h(rr_n) + (I_{lr_n} + I_{rl_n})}{h(lr_n) + h(rl_n) - (I_{lr_n} + I_{rl_n})} > th_1 \tag{11}$$

Therefore $(I_{lr_n} + I_{rl_n})$ should fulfil the following rule:

$$(I_{lr_n} + I_{rl_n}) > \frac{th_1 \cdot (h(lr_n) + h(rl_n)) - (h(ll_n) + h(rr_n)}{1 + th_1} \tag{12}$$

On the other side, the $(I_{lr_n} + I_{rl_n})$ samples within $h(lr_n)$ or $h(rl_n)$ are modified as follows:

$$M_{lr_n}(i)' = M_{lr_n}(i) - D, 1 \leq i \leq I_{lr_n}, \tag{13}$$

$$M_{rl_n}(i)' = M_{rl_n}(i) + D, 1 \leq i \leq I_{rl_n}. \tag{14}$$

where, D is the interval of classes of histogram ($D = 2/L$).

Embedding bit when $w(n)$ is '0'

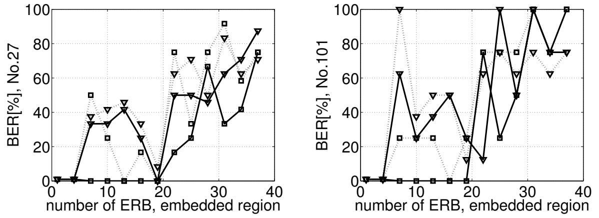
If $\beta(n)$ is lower than th_0, no operation is required. Otherwise, the number of samples of $h(ll_n)$,$h(lr_n)$,$h(rl_n)$,and $h(rr_n)$ is fixed to satisfy $\beta(n)' = \frac{h(ll_n)'+h(rr_n)'}{h(lr_n)'+h(rl_n)'} > th_0$. Therefore, I_{lr_n} and I_{rl_n} samples within $h(ll_n)$ and $h(rr_n)$ class were extracted and moved to $h(lr_n)$ and $h(rl_n)$ class. $(I_{ll_n} + I_{rr_n})$ should fulfil the following rule:

$$(I_{ll_n} + I_{rr_n}) > \frac{(h(ll_n) + h(rr_n)) - th_0 \cdot (h(lr_n) + h(rl_n))}{1 + th_0} \tag{15}$$

5) Conversion time-frequency domain to time-domain signal using inverse STFT: Simple STFT transform enable the conversion of a signal to time-frequency domain. Modification of complex Fourier coefficients according to the proposed roles, and then inverse STFT, resulted in a time-domain signal with a watermark in it.

C. Detection algorithm

Most of the stages in the detection algorithm are similar to those in the embedding side process. Fig.2 shows the diagram of detection process. Watermarked signal is firstly converted to time-frequency domain by using STFT. After rescaling based on $\mathrm{ERB_N}$-number, the magnitudes are calculated for each coefficient. The histogram of magnitude enables to calculate $\beta(n)$ and then successive bit of watermark.



(a) No.27: Castanets single tone

(b) No.101: Harmonic complex tone, f0=100 Hz, 20 harmonics

Figure 3. Robustness against MP3 conversion: □-solid :mp3o-4bit, ▽-solid :mp3o-8bit, □-dotted :mp3t-4bit, ▽-dotted :mp3t-8bit

III. Experimental Results

The algorithm was applied to 8 audio signals which were selected from EBU SQAM test sets [3] (16 bits/sample, 44.1 kHz, stereo) and a harmonic complex tone signal (f_0 = 100Hz, 20 harmonics). In the following subsections, the results about typical test sources, No.27 (Castanet single instrument) and No.101 (Harmonics complex tone), were shown on behalf of the tested 8 sources. In the carried out experiments, 4 and 8 bits of payload was embedded in a selected frequency band of 5-seconds host signal.

A. Robustness Test

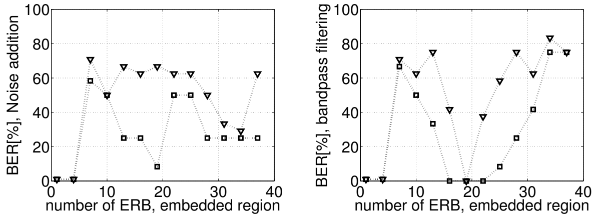
To evaluate the performance of the proposed watermarking method, robustness against several kinds of attack was tested. Tested attacks are MP3 compression (128 kbps, joint stereo [4], "mp3o"), MP3 compression twice ("mp3t"), white noise addition (SN 36 dB, "wgn0"), and Bandpass filter (pass band: 0.1–6 kHz, 12 dB/oct., "bapf").

Figure 3 shows the bit detection error rates on the robustness test against MP3 conversion for test source No.27 (Castanets: single tone), and 101 (Harmonic complex) respectively. The performances vary with frequency band chosen for embedding. At the condition embedding 4 bits of payload in the frequency band below 19th $\mathrm{ERB_N}$-number (1540 – 2200 Hz), it can be seen that embedded payload against MP3 still remains. Moreover the error rates grows up in the condition when MP3 conversion is done twice. The performance of the proposed algorithm was slightly worse for frequency band above 19th $\mathrm{ERB_N}$-number.

Figure 4 shows the bit detection error rates on the robustness test against noise addition and bandpass filtering for harmonic complex tone (test source No.101). In proposed method, the distribution of STFT magnitude is required to be invariant between watermarked signal and attacked one. Result in Figure 4 shows that the proposed watermarking can't resist against noise because of the changes in the distribution introduced by the attacks.

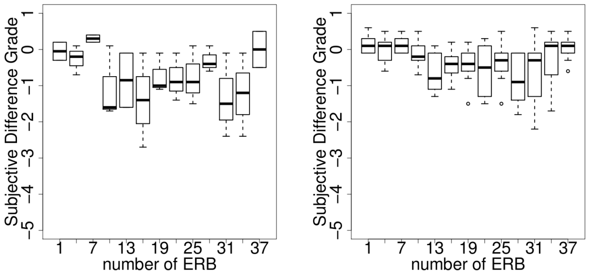
B. Subjective Listening Test

To evaluate the quality of the proposed watermarking technique, a subjective listening tests using the paradigm of the double-blind triple-stimulus with hidden reference



(a) noise addition (b) bandpass 0.1-6 kHz

Figure 4. Robustness against noise addition, and bandpass filtering: □:4bit, ▽:8bit, Test source No.101(Harmonics complex tone)



(a) 4 bits embedding (b) 8 bits embedding

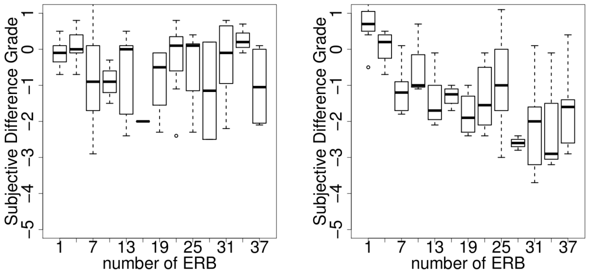
Figure 5. Listening test result: No.27(Castanet Single instrument)

(ITU-R BS. 1116 [5]) were carried out. In the test, above mentioned 9 signal types (5 seconds each) were used. Four normal hearing subjects participated in these tests.

The listening results of the subjective-difference-grade (SDG) for 2 typical signals of 9 tested signals are shown in Fig. 5,and 6. Although the grades are varying between subjects, almost of the average SDGs are graded greater than -1 (Perceptible, but not annoying) for the instrumental signals or ensemble signals. In the case of the SDG for harmonic complex tone, the watermark was more annoying for most of the subjects when it was applied to a high frequency band.

IV. Multi-band Embedding

In the above mentioned experiment, a single frequency band was used for embedding. The proposed method is designed for embedding not only into a single band but also to several frequency band simultaneously.



(a) 4 bits embedding (b) 8 bits embedding

Figure 6. Listening test result: No.101(Harmonics complex tone)

Table I
Multi-band embedding

	27-mp3o	27-mp3t	101-mp3o	101-mp3t
4 bits × 3 (12 bits)	0.00	2.78	0.00	0.00
4 bits × 4 (16 bits)	0.00	10.42	0.00	6.25
8 bits × 2 (16 bits)	8.33	37.50	0.00	12.50
4 bits × 5 (20 bits)	3.33	16.67	0.00	15.00
8 bits ×3 (24 bits)	16.67	43.06	25.00	25.00
				(BER[%])

Table I shows the results of robustness against the MP3 attacks in case of multi-band embedding, 4 bits× 3 frequency bands, 4×4, 8×2, and 8×3, between first and 19th ERB_N-number (about 2000 Hz). 'mp3o' and 'mp3t' indicate the single MP3 attack and double attack respectively. Results showed that in the multi-band case the proposed method can embed up to 20 bits without decrease in robustness against MP3.

V. Conclusion

In this paper, a watermarking method is presented. This method can embed 8-bit payloads into 5-second music signal. Proposed method modifies the histogram of periodical magnitude variation. The robustness test against typical signal manipulations or attacks showed a very good performance of the proposed method. Subjective listening tests were also carried out to assess the quality of the watermarked signal, i.e. audibility of distortions resulting from watermark. Results show that the proposed watermarking technique resists the MP3 compression attacks in the case of embedding payload below 2 kHz. However, the proposed procedure performs a bit poorer against the attacks of noise addition, bandpass filtering. Results of hearing tests showed that the proposed method doesn't produce audible deteriorations. Finally, multi-band embedding was investigated. Results showed that the multi-band embedding can embed up to 20 bits without decrease in robustness against MP3. For the future work, the countermeasure to increase robustness against noise and echo addition, bandpass filtering, pitch modification are strongly desired.

References

[1] X. Zhang, "Segmenting histogram-based robust audio watermarking approach," *Journal of Software*, vol. 3, no. 9, pp. 3–11, 2008.

[2] B. Moore, *Hearing*. Academic Press, 1995.

[3] SQAM-CD, EBU, "Sound quality assessment material recordings for subjective tests," 2008, `http://tech.ebu.ch/publications/sqamcd/`.

[4] The LAME Project, "Lame mp3 codec ver.3.99," `http://lame.sourceforge.net/`.

[5] ITU-R, I., "Recommendation bs. 1116," *Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems*, 1997.