ROBUST AND RELIABLE AUDIO WATERMARKING BASED ON PHASE CODING

Nhut Minh Ngo and Masashi Unoki

School of Information Science, Japan Advanced Institute of Science and Technology
1–1 Asahidai, Nomi, Ishikawa, 923–1292, Japan
Emails: {nmnhut, unoki}@jaist.ac.jp

ABSTRACT

This paper proposes a novel robust audio watermarking method based on phase coding. The quantization index modulation technique is employed for embedding watermarks into the phase spectrum of audio signals. To increase robustness of the proposed method, the region of phase spectrum that is resistant against attacks is selected for embedding. We experimentally analyzed the phase spectrum to find out which region is not distorted under attacks. On the other hand, the quantization step size is suitably selected so that the modification of phase does not cause severe distortion in sound quality. The experimental results show that the watermarks could be kept inaudible in audio signals and robust against attacks. The proposed method has the ability to embed watermarks into audio signals up to 400 bits per second with a bit error rate of less than 1%.

Index Terms— audio watermarking, phase coding, quantization index modulation, robustness

1. INTRODUCTION

Recent years have seen a rapid development of multimedia and the Internet technology that enhances our digital communication world but also leads to many risks at information security. In that context, digital multimedia watermarking has been attracted to protect copyright and ownership of multimedia [1] and applied to other applications such as copy control, tamper detection and covert communication [2]. Compared with image and video watermarking, audio watermarking has received less interest and is still young in the field.

In order to be applied in such applications, an audio watermarking method should meet four basic requirements [1]: inaudibility, blindness, robustness, and high capacity. There is, however, a trade-off among these requirements. For example, it is straightforward that perceptually insensitive features of audio signals should be exploited for embedding inaudible watermarks. The resistance against attacks such as lossy compression is then challenging since the watermarks could be easily destroyed without degrading the sound quality. Finding out suitable audio features to ensure both inaudibility and robustness is the most important task of watermarking design.

This work was supported by a Grant-in-Aid for Scientific Research (B) (No. 23300070) and an A3 foresight program made available by the Japan Society for the Promotion of Science.

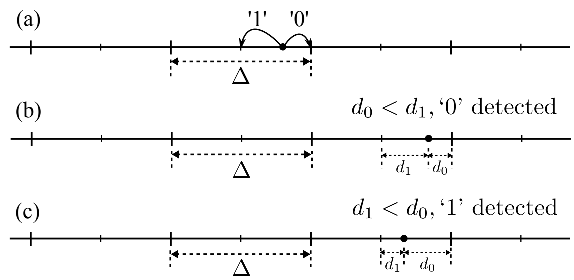

Fig. 1. An illustration of QIM-based watermarking technique: (a) embedding, (b) detection of '0', and (c) detection of '1'.

In general, audio watermarking methods directly embed watermarks into audio samples or audio features in transformed domain. Some methods embed watermarks by replacing least significant bits (LSB) with watermark bits, or insert watermarks which are perceptually shaped according to the human auditory system (HAS) [3]. Other methods take the advantages of frequency masking characteristics of HAS [4] or relative inaudibility of phase change [5] to embed inaudible watermarks. Phase has been exploited for inaudible audio watermarking since controlled phase alteration results in inaudible change in sound to HAS [6]. These methods, however, still meet a trade-off among blindness [5], robustness (LSB), and capacity [3].

Several audio watermarking methods have been proposed based on quantization index modulation (QIM) [7–10] and showed that QIM is a promising technique for robust watermarking. These methods have applied QIM to certain audio features to embed watermarks. Reliability of audio features, e.g., meaningfulness of a frequency component, which is significant for watermark detection has not been thoroughly considered. As a result, these methods are still limited in inaudibility, robustness, and detection accuracy.

In this work, we propose a robust audio watermarking method based on phase coding. The phase spectrum of audio signals is encoded with watermarks by the quantization index modulation technique. Toward a robust watermarking sys-

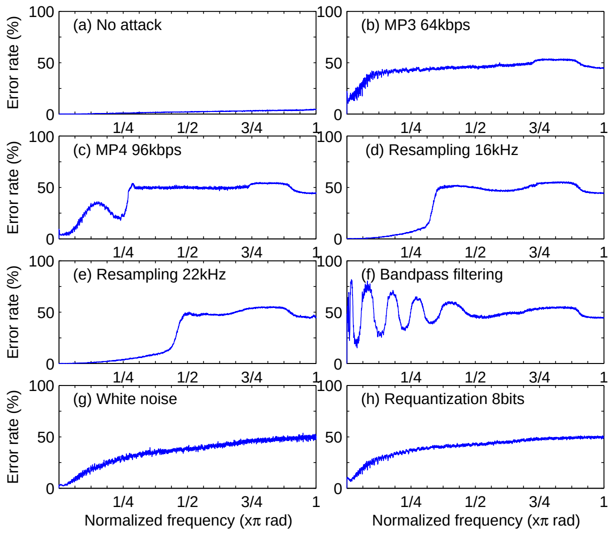

Fig. 2. Distribution of QIM decoding errors on phase spectrum against eight different attacks.

tem with an accurate detector, we experimentally analyze the phase spectrum to find out the embedding region that is resistant against attacks. To increase the reliability of watermark detection, frequency components that have non-meaningful amplitude, i.e., very small magnitude, are excluded from the embedding and detecting processes. The quantization step size is suitably selected to keep watermarks inaudible. The experimental results showed that the proposed method can produce watermarked signals with high sound quality and the embedded watermarks are robust against various attacks.

2. PROPOSED METHOD

2.1. Quantization index modulation

QIM, which was firstly proposed for watermarking by Chen and Wornell [7], has been widely applied to many kinds of watermarking and obtained promising results. The procedure of embedding and detecting watermarks is quite simple. Figure 1 shows an illustration of embedding and detection. To embed a bit m, '0' or '1' to a scalar variable x, we quantize the scalar number to the nearest point that is an even or odd multiple of $\frac{\Delta}{2}$, respectively as Eq. (1). The obtained variable, y, is sent to receivers and might be affected by channel noise, hence becomes $\hat{y}$. To decode the embedded bit from $\hat{y}$, we calculate the distances between $\hat{y}$ and the nearest even multiple of $\frac{\Delta}{2}$, d_0 and the nearest odd multiple of $\frac{\Delta}{2}$, d_1 and then compare d_0 and d_1 to decode the bit as Eqs. (2–4).

$$y = \begin{cases} \Delta \left\lfloor \dfrac{x}{\Delta} + \dfrac{1}{2} \right\rfloor & \text{if } m = \text{'0'} \\ \Delta \left\lfloor \dfrac{x}{\Delta} \right\rfloor + \dfrac{\Delta}{2} & \text{if } m = \text{'1'} \end{cases} \quad (1)$$

where $\lfloor . \rfloor$ is the floor function and Δ is the step size.

$$d_0 = \hat{y} - \Delta \left\lfloor \frac{\hat{y}}{\Delta} + \frac{1}{2} \right\rfloor \quad (2)$$

$$d_1 = \hat{y} - \left(\Delta \left\lfloor \frac{\hat{y}}{\Delta} \right\rfloor + \frac{\Delta}{2} \right) \quad (3)$$

$$\hat{m} = \begin{cases} \text{'0'} & \text{if } d_0 < d_1 \\ \text{'1'} & \text{otherwise} \end{cases} \quad (4)$$

2.2. Key principle

We apply QIM to the phase spectrum of audio signals to construct a robust and reliable audio watermarking system with the following considerations. (1) Phase alteration is relatively inaudible [6], hence quantizing the phase with a suitable quantization step size keeps watermarks inaudible. (2) To achieve the robustness, we analyze the effects of attacks on distortion of frequency components to find out the resistant frequency region for embedding. In addition, modifying a narrower frequency region could keep the distortion less audible. (3) To increase the reliability, non-meaningful frequency components, i.e., very low magnitude components, are excluded from the embedding process.

Figure 2 plots the distribution of QIM decoding error on the phase spectrum of 6120 frames of music. Statistically, the cases where the error is greater than or equal to 50% imply that QIM does not work properly. The lower the error is, the more robust the frequency region is. We observed that the low-frequency region is the most robust against the attacks. We choose this region for embedding watermarks.

2.3. Watermark embedding

The embedding process starts with a frame segmentation of the original signal, $x[n]$ into frames $x_i[n]$. The frame size is determined according to the desired bit rate of watermark. Each watermark bit $s[i]$ is embedded into each audio frame $x_i[n]$. Figure 3(a) depicts the block diagram of the four steps that embed one bit into an audio frame as follows.

Step 1. Original frame $x_i[n]$ is transformed into Fourier spectrum $X_i(\omega)$ by fast Fourier transform (FFT). Magnitude spectrum $|X_i(\omega)|$ and phase spectrum $\angle X_i(\omega)$ are calculated.

Step 2. We decompose the spectrum into K subbands. Then we select the frequency components in the first subband that are meaningful, i.e., their magnitude is greater than a threshold ϵ. The watermark is embedded into only these selected components to increase reliability.

Step 3. The bit $s[i]$ is encoded into the phase of the selected components by Eq. (1) and a quantized phase spectrum $\angle Y(\omega)$ is obtained. Although $s[i]$ can be encoded by only one component, it should be encoded by all the selected components to increase robustness.

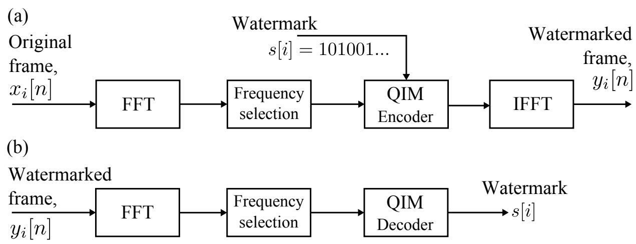

Fig. 3. General scheme of the proposed method: (a) process of embedding one bit and (b) process of detecting one bit.

Step 4. The magnitude spectrum, $|X_i(\omega)|$ and the quantized phase spectrum, $\angle Y(\omega)$, are combined into Fourier spectrum $Y(\omega)$ which is then transformed into time domain signal $y_i[n]$ by inverse Fourier transform (IFFT).

Finally, all the processed frames are combined together to yield a watermarked signal $y[n]$.

2.4. Watermark detection

The detection process also starts with a frame segmentation of the watermarked signal, $y[n]$ into frames $y_i[n]$. Figure 3(b) shows the block diagram of the process that detects a bit from a watermarked frame involving three steps as follows.

Step 1. Watermarked frame $y_i[n]$ is firstly transformed into Fourier spectrum $Y_i(\omega)$ by FFT. Phase spectrum $\angle Y_i(\omega)$ is calculated.

Step 2. The embedded frequency components is selected as Step 2 in the embedding process.

Step 3. Each selected component is decoded by Eq. (4) to extract a bit. The output bit, $s[i]$, is determined by majority decision on all the extracted bits, e.g., if the number of '0', N_0, are greater the number of '1', N_1, the output is '0'.

These steps are repeated until we reach the final frame.

2.5. Frame synchronization

The detection process works with an assumption that the frame positions are synchronized. In practice, the frame positions might be unavailable, so we have to identify the starting point before detecting watermarks. It should be noted that a bit is detected from a watermarked frame by majority decision. N_0 and N_1 are compared to decide the output bit, $s[i]$. We can see that N_0 much greater than N_1 implies that the probability $P(s[i] = \text{'0'})$ is much higher. In other words, the confidence that $s[i]$ is correctly detected is higher. In general, we define the detection confidence of a bit by:

$$\delta_i = \max\left(\frac{N_0}{N_1}, \frac{N_1}{N_0}\right) \tag{5}$$

We can search for a correct frame position over a frame length of a signal. It is obvious that if we select a correct frame i, δ_i is maximized. Figure 4 depicts an illustration of frame position detection. We calculate detection confidences over 32 frames in two cases: (a) without embedded watermark and (b) with embedded watermark. The detection confidences are normalized so that the maximum value equals to 1. There is no obvious peak in Case (a) while very high peaks occur at the correct frame-starting points in Case (b). The search procedure is performed only once at the beginning of the detection process. Once the starting point is determined, all the frame positions can be synchronized.

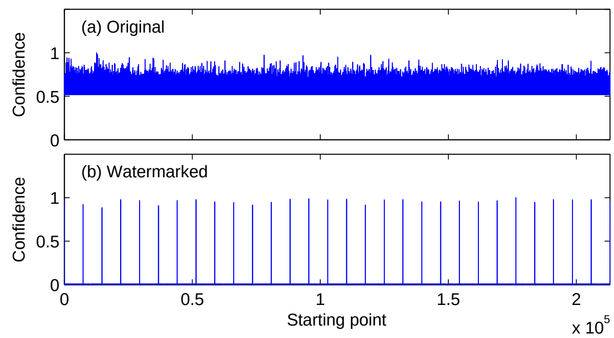

Fig. 4. An illustration of frame position detection in the case frame length = 7350 points: (a) original signal and (b) watermarked signal.

So as to combat time-cropping attack, we need to update frame positions and detect watermarks simultaneously. The updating procedure does not cost much computation. The reason is because time-cropping attack randomly cuts a few samples, otherwise the sound quality is severely reduced.

3. EVALUATIONS

We carried out experiments to evaluate the effectiveness and performance of the proposed method with the 102 RWC music tracks [11] which have a sampling frequency of 44.1 kHz and 16-bit quantization. We investigated the bit rates from 6 to 400 bps. The watermarks were randomly generated. The parameters K and ϵ were determined by experimental analysis and set to 16 and 10^{-4}, respectively. The FFT size is equal to frame size and the rectangle window was used.

The inaudibility was measured by perceptual evaluation of audio quality (PEAQ) [12] which rates sound quality by the objective difference grade (ODG). Watermark detection performance was measured by bit detection rate (BDR), the ratio between the number of correct bits and the total number of bits. We set evaluation criteria for PEAQ and BDR by -2 ODG (slightly annoying) and 10% as recommended by [13]. The robustness was evaluated under the following processing: MP3 64 kbps (MP3), MP4 96 kbps (MP4), addition of white Gaussian noise 36 dB (AWGN), requantization 8 bits (Requan.), resampling 22 kHz (Res. 22k) and 16 kHz (Res.

Table 1. PEAQ (ODG) and BDR (%) against various attacks with respect to the bit rate (bps).

	6	12	18	30	50	100	150	200	300	400
PEAQ	-0.70	-0.82	-0.92	-1.04	-1.14	-1.43	-1.64	-1.75	-1.90	-1.74
BDR										
No attack	99.90	99.93	99.88	99.79	99.80	99.74	99.73	99.71	99.68	99.54
MP3	99.59	99.58	99.52	99.58	99.45	98.53	97.49	96.12	92.93	85.98
MP4	99.80	99.74	99.73	99.59	99.54	99.09	98.28	97.16	94.52	88.93
AWGN	99.89	99.77	99.65	99.50	99.28	98.95	98.90	98.75	98.34	96.55
Requan.	98.50	97.84	97.55	97.19	96.75	96.04	95.95	95.60	94.68	92.12
Res. 22k	99.90	99.93	99.88	99.79	99.80	99.73	99.71	99.69	99.61	98.90
Res. 16k	99.90	99.92	99.88	99.80	99.80	99.72	99.69	99.65	99.49	98.44
BPF	80.92	75.55	70.45	67.43	64.08	60.02	57.01	55.98	54.66	43.20

16k), and bandpass filtering with passband $[0.1, 6]$ kHz and stopband attenuation -12 dB/octave (BPF).

Table 1 shows the results of PEAQ and BDR test with respect to the bit rate. The quantization step size was set to $\frac{\pi}{5}$ in this experiment. PEAQ is good in the cases that the bit rate is less than 50 bps but still acceptable for the other cases. BDRs with no attack are all higher than 99.5%. Watermarks are strongly robust against MP3 and MP4 attacks with the bit rate not greater than 200 bps and weakly robust with the other cases. Regarding AWGN, requantization, and resampling attacks, the watermarks are also strongly robust, except for the case of requantization with 400 bps. The watermarks are, however, weak against bandpass filtering, i.e., detection accuracy is only acceptable at a bit rate of 6 bps. The quantization step size is an important factor that controls inaudibility and robustness. We investigated how it affects the sound quality and the detection accuracy. Table 2 shows the results of PEAQ and BDR against attacks with respect to the quantization step size. The bit rate was set to 6 bps in this experiment. The results reveal that the sound quality decreases and the detection accuracy increases when the quantization step size gets higher, especially it is significantly improved for bandpass filtering. We can select a quantization step size of $\frac{\pi}{2}$ for the high robustness and the reasonable sound quality. The higher value for the step size may result in low sound quality.

Table 2. PEAQ (ODG) and BDR (%) against various attacks with respect to the quantization step size.

	$\frac{\pi}{10}$	$\frac{\pi}{8}$	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{2}$
PEAQ	-0.27	-0.38	-0.56	-0.89	-1.55
BDR					
No attack	99.90	99.92	99.93	99.93	99.93
MP3	98.87	99.38	99.44	99.61	99.66
MP4	99.38	99.62	99.74	99.82	99.85
AWGN	99.74	99.80	99.89	99.92	99.92
Requan.	97.43	97.83	98.22	98.63	99.33
Res. 22k	99.90	99.93	99.93	99.93	99.93
Res. 16k	99.90	99.92	99.93	99.93	99.93
BPF	40.88	48.56	71.08	88.06	96.52

4. DISCUSSION

This paper proposes a simple but effective method of audio watermarking based on phase quantization. A number of work based on QIM have appeared in the literature. Narimannejad and Ahadi [8] reported a method that applies QIM to the phase of speech signals. The method embeds watermarks into only one frequency component of a sinusoidal model. That frequency component might be fragile against perceptual compression because it could be under the masking threshold and hence eliminated. The work by Khademi et al. [9] applies QIM to the phase of all frequency components including non-meaningful frequency components. The non-meaningful components might cause detection errors, especially at high embedding rates. Unlike these methods, this work capitalizes the frequency components robust against attacks and meaningful for embedding to increase robustness, detection accuracy as well as bit rate.

5. CONCLUSION

We propose an audio watermarking method based on phase quantization. The quantization step size is suitably selected to preserve sound quality of watermarked signals. Phase spectrum are analyzed to find out the frequency region withstanding attacks to ensure robustness. Frequency components that are non-meaningful are excluded from embedding to increase reliability. The experimental results verify that the watermarked signals have high sound quality and the embedded watermarks are robust against various attacks, especially lossy compression. The proposed method has the ability to embed up to 400 bps with the error rate of less than 1%.

References

[1] N. Cvejic and T. Seppänen, *Digital Audio Watermarking Techniques and Technologies*, IGI Global, 2007.

[2] C. I. Podilchuk and E. J. Delp, "Digital watermarking: Algorithms and applications," *IEEE Signal Proc. Magazine*, vol. 18, no. 4, pp. 33–46, 2001.

[3] P. Bassia, I. Pitas, and N. Nikolaidis, "Robust audio watermarking in the time domain," *IEEE Trans. Multimedia*, vol. 3, no. 2, pp. 232–241, 2001.

[4] M. D. Swanson, B. Zhu, A. H. Tewfik, and L. Boney, "Robust audio watermarking using perceptual masking," *Elservier Signal Processing*, vol. 66, no. 3, pp. 337–355, 1988.

[5] A. Takahashi, R. Nishimura, and Y. Suzuki, "Multiple watermarks for stereo audio signals using phase-modulation techniques," *IEEE Trans. Signal Processing*, vol. 53, no. 2, pp. 806–815, 2005.

[6] B. C. J. Moore, *An Introduction to the Psychology of Hearing, 6th ed*, Brill Academic Pub, 2013.

[7] B. Chen and G. W. Wornell, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Trans. on Info. Theory*, vol. 47, no. 4, pp. 1423–1443, 2001.

[8] M. Narimannejad and S. M. Ahadi, "Watermarking of speech signal through phase quantization of sinusoidal model," in *Proc. of Iranian Conf. on Electrical Engineering*, May 2011, pp. 1–4.

[9] N. Khademi, M. A. Akhaee, S. M. Ahadi, M. Moradi, and A. Kashi, "Audio watermarking based on quantization index modulation in the frequency domain," in *Proc. of Int. Conf. on Signal Proc. and Communications*, Nov. 2007, pp. 1127–1130.

[10] L. Gang, A. N. Akansu, and M. Ramkumar, "Mp3 resistant oblivious steganography," in *ICASSP*, May 2001, vol. 3, pp. 1365–1368.

[11] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "Music genre database and musical instrument sound database," in *Proc. ISMIR 2003*, 2003, pp. 229–230.

[12] P. Kabal, "An examination and interpretation of itu-rbs.1387: Perceptual evaluation of audio quality," Tech. Rep., Dept, Elect. Comp. Eng., 2002.

[13] A. Nishimura, M. Unoki, K. Kondo, and A. Ogihara, "Objective evaluation of sound quality for attacks on robust audio watermarking," in *Proc. of Meeting on Acoustics*, 2013, pp. 1–9.