2003년도 컴퓨터소사이어티 추계학술대회 논문집

Modification of Polar Echo Kernel for Performance Improvement of Audio Watermarking

Siho Kim, Hongseok Kwon, Keunsung Bae
School of Electronic & Electrical Engineering, Kyungpook National University,
1370 Sankyuk-dong, Puk-gu, Taegu 702-701, Taegu, Korea
si5@mir.knu.ac.kr

Abstract

In this paper, we present a new echo kernel, which is a modification of polar echo kernel, to improve the detection performance and robustness against attacks. Polar echo kernel may take advantage of large detection margin from the polarity of inserted echo signal, but its poor frequency response in low frequency band degrades sound quality. To solve this problem, we applied bipolar echo pulses to the polar echo kernel. Using the proposed echo kernel the distributions of autocepstrum peaks for data '0' and '1' are located more distant and improvement of detection performance is achieved. It also makes the low frequency band flat so that the timbre difference in the polar echo kernel can be removed to reproduce the imperceptible sound quality. Informal listening tests as well as robustness test against attacks were performed to evaluate the proposed echo kernel. Experimental results demonstrated the superiority of the proposed echo kernel to both conventional unipolar and polar echo kernels.

1. Introduction

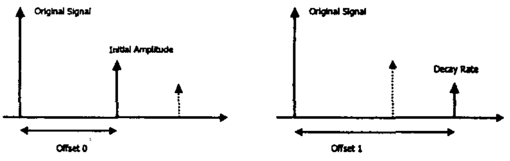
Digital audio watermarking is the technique that a watermark signal is added to the original audio content as imperceptible as possible. It is now drawing much attention as a new method of providing copyright protection to digital audio content. Over the last few years, considerable audio watermarking algorithms have been proposed such as spread spectrum coding [1], phase coding [2], echo hiding [2,3,4], and so on. Among them, spread spectrum and echo hiding schemes are paid more attention than phase coding scheme, because they do not need the original audio for detecting watermarks. Especially, echo hiding is considered to be better in terms of imperceptibility since it just adds delayed and attenuated version of the original signal itself. Of course, time-delay and strength of echo should be decided carefully for trade-offs between imperceptibility and robustness. To increase the ability of detection and robustness against attack, it requires a larger echo, which results in distortion and timbre change of the audio signal. Therefore it is important to design the echo kernel that can have imperceptibility to original digital content as well as improved detection performance and robustness against attack.

In this paper, we present a new echo kernel that shows improved detection rate and robustness compared to the conventional echo kernels. The newly designed echo kernel has the form of modified polar echo kernel with two pulses having different signs. Experimental results are given with our findings and discussion. This paper is organized as follows. In section 2, we briefly explain audio watermarking schemes with echo kernels. In the next section the proposed echo kernel and its characteristics are presented. In section 4, the experimental results are shown and discussed, and finally we make a conclusion in the last section.

2. Audio Watermarking with Echo Kernels [2,3]

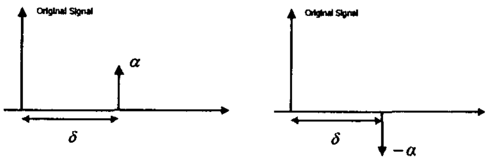
Audio watermarking schemes based on echo kernels embed data into a host audio signal in the form of delayed and attenuated version of the original signal, that is, an echo. There are three parameters to be considered in the echo kernel: initial amplitude, decay rate, and offset. The quality and robustness of watermarked signal depend on the amplitude and offset delay of the echo. The initial amplitude and decay rate are desired to be smaller than the audible threshold of the human ear. In general, the human ear cannot distinguish echo from the original for most sounds if offset delay is around 1 msec.

There are two ways to represent data, i.e., binary symbols with single pulse as an echo: one is using different delay times (or offset) for each binary symbols, the other is using the sign (or polarity) of the pulse. Figures 1 and 2 show typical echo kernels with single pulse as an echo. In figure 1 the coder uses two delay times, one to represent a binary '0'(offset 0, δ_0) and the other to represent a binary '1'(offset 1, δ_1). On the other hand, in figure 2, the coder uses positive polarity to represent a binary '0' and negative polarity to represent a binary '1'.



(a) '0' kernel (b) '1' kernel

Fig. 1. Unipolar echo kernel



(a) '0' kernel (b) '1' kernel

Fig. 2. Polar echo kernel

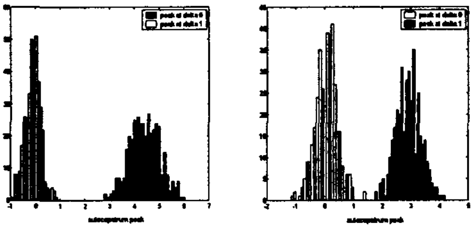
Information is embedded into a signal by echoing the original signal with one of two kernels. Let the kernel represent the system function for encoding a binary data. Then processing a signal through either '0' kernel or '1' kernel will result in an encoded signal. In order to encode more than one bit, the original signal is divided into smaller portions. Each individual segment can then be echoed with the desired bit by considering each as an independent signal. The final encoded signal is the recombination of all independently encoded signal portions. To prevent abrupt changes between portions encoded with different bits, the encoded signals of each portion are added using mixed signal.

Extraction of the embedded information involves the detection of delay position or polarity of the echoes. In order to do this, the cepstrum of watermarked signal is generally calculated. Then larger peak at offset delay is determined by taking the autocorrelation of the cepstrum, which is denoted as autocepstrum and can be obtained from the equation (1).

$$s_{AC}(n) = F^{-1}(\ln_{complex}(F(s_{wm}(n)))^2) \qquad (1)$$

where F means Fourier Transform and F^{-1} its inverse transform. s_{wm} is watermarked audio signal and s_{AC} is its autocepstrum. In each frame, the embedded binary data is decoded by comparing the autocepstrum peak at offset positions.

Figure 3 shows the distributions of autocepstrum peak at offset 0 and offset 1 for the watermarked signal using the unipolar echo kernel. In case of using '0' kernel, the large peak value exists at δ_0 and small values distribute around zero at δ_1 as shown figure 3(a). On the contrary, in case of '1' kernel, the distribution of peaks shows the property similar to '0' kernel except the exchanged offset delay like figure 3(b). The peak value increases in direct proportion to the amplitude of echo signal inserted and the shape of distribution has a property similar to Gaussian distribution due to the influence of audio signal or noises. When we draw a comparison with relative distance for peaks distribution in δ_0 and δ_1, we can verify that the relative distance is closer in case of '1' kernel's than '0' kernel's as shown figure 3 (a) and (b). It is because that in case '1' kernel, the echo amplitude is designed to be smaller than '0' kernel considering the auditory property of human ear. Hence the probability of detecting incorrectly '1' as '0' is higher than the opposite case.



(a) '0' kernel (b) '1' kernel

Fig. 3. Distributions of autocepstrum peak with unipolar echo kernel

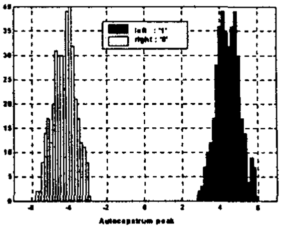
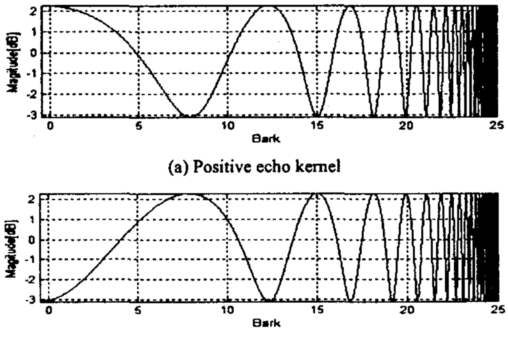


Fig. 4. Distributions of autocepstrum peak using positive and negative echo kernel

Using the polar echo kernel with the same delay, we can reduce the detection error occurred asymmetrically in the unipolar echo kernel and increase the relative distance for the distribution of autocepstrum peak two times as shown figure 4. It can be expected that polar echo kernel has more margins than unipolar echo kernel in the detection of hidden data since the former can have large difference in autocepstrum, as shown in figure 4. But polar echo kernel has a critical disadvantage in the sense that it results in the timbre alteration of the audio signal caused by the negative echo pulse. The negative echo pulse functions as a high pass filter and poor frequency response at lower band distorts the original signal, and makes the sound slightly sharp. Figure 5 shows the frequency response of the polar echo kernel in Bark scale[5], positive echo kernel and negative echo kernel, respectively, and offset delay δ is set to 50 samples (1.1 msec at 44.1 kHz sampling rate).

In figure 5, the frequency responses of positive and negative echo kernel are contrary to each other. Especially the frequency band below 500 Hz (corresponding to about 5 Bark) is boosted in case of positive echo kernel but weakened in case of negative echo kernel. Low frequency band below 500 Hz is narrow band in comparison with whole frequency band of 22,050 Hz but it is relatively important band considering the property of human auditory system. Hence watermarked sound shows more abundant timbre than original sound in case of positive echo but sharper and keener in case of negative echo. In general, people have a preference for abundant timbre so that the sound using positive echo is preferred to negative echo. And it is also concerned that the different timbre in adjacent frame may cause the deterioration of sound quality. Therefore it is necessary to make up for the weakness of low band caused by negative echo and reduce the difference of timbre between positive and negative echo kernel.



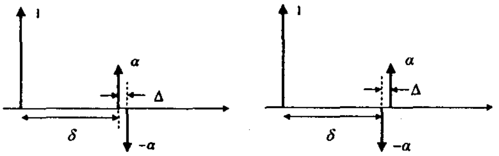
(a) Positive echo kernel

(b) Negative echo kernel

Fig. 5. Frequency responses of positive and negative echo kernel

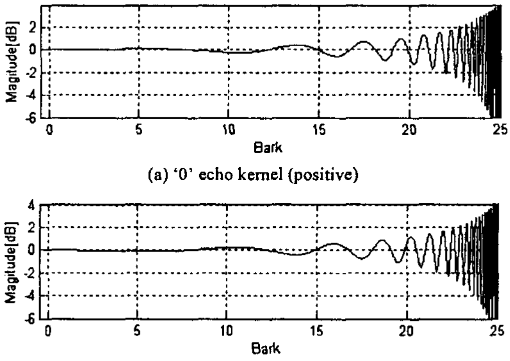
3. Proposed Echo Kernel with Bipolar Echo

Since the major cause of timbre alteration in polar echo kernel is the opposite property at low frequency band, we can solve this problem by applying one more pulse having opposite polarity to the preceding one. It is very similar idea to [4], which is kind of a modification of unipolar echo kernel, what they call a bipolar echo kernel. It makes the low frequency band flat so that it increases the imperceptibility. We apply this bipolar echo to the polar echo kernel as shown in figure 6. The frequency responses are shown in figure 7 where offset delay δ and Δ are set to 50 samples and 1 sample, respectively. Consequently it makes the low frequency band of each kernel flat so that the defect of negative echo might be complemented and the timbre difference can be removed considerably.



(a) '0' echo kernel (b) '1' echo kernel

Fig. 6. Proposed echo kernel with bipolar echo



(a) '0' echo kernel (positive)

(b) '1' echo kernel (negative)

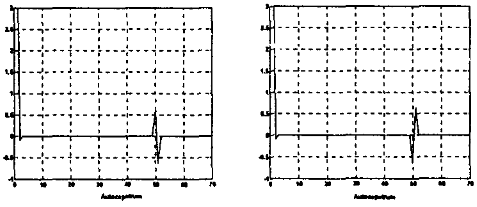
Fig. 7. Frequency responses of the positive and negative echo kernel with bipolar echo

The watermark detection procedure with the proposed echo kernel is similar to that of conventional methods, that is, the detection and comparison of the autocepstrum peaks at the offset position. However, in the proposed echo kernel, the sign of autocepstrum plays an important role in detection, so it is calculated by equation (2).

$$s_{AC}(n) = \mathrm{Re}\left(F^{-1}\left(\left|\ln_{complex}(S_{wm})\right|^{2}\right)\right) \qquad (2)$$

where S_{wm} denotes the discrete Fourier transform of the watermarked audio signal. Figure 8 shows the results of autocepstrum with the proposed echo kernel for offset delay, δ, of 50 samples (1.1 ms at SR 44.1 kHz) and Δ of 1 sample. As shown in figure 8, in case of '0' echo kernel the negative peak follows the positive one consecutively, and in case of '1' echo kernel it has reverse pattern. It seems to be more robust if the difference between the peaks of autocepstrum at δ and (δ +Δ) are used for detection. Therefore we define the detection parameter *peak_diff* as given in equation (3). For example, when the sign of *peak_diff* is positive, the '0' echo kernel is detected and conversely negative, '1' echo kernel.

$$peak\_diff = s_{AC}(\delta) - s_{AC}(\delta + \Delta) \qquad (3)$$



(a) '0' echo kernel (positive) (b) '1' echo kernel (negative)

Fig. 8. Autocepstrum of the positive and negative echo kernel with bipolar echo

4. Experimental Result

In order to evaluate the performance of the proposed echo kernel, robustness test and informal listening test were performed. The echo kernels used in this experiment are denoted as follows:

- Unipolar Echo Kernel (UEK): offset of the echo is used
- Polar Echo Kernel (PEK): polarity of the echo is used
- Modified Polar Echo Kernel (MEK): polarity of the bipolar echo is used

The parameters used in each echo kernel are shown in table 1. The audio signals used in the experiments are sampled at 44.1 kHz with 16 bits resolution. Information was embedded into the audio signal at a rate of 28.7109 (in case that frame length is 1024 samples and overlap is 512 samples) bits per second.

Table 1. Parameters for each echo kernel [samples]

Echo Kernel	0' echo kernel Delay (Amplitude)	'1' echo kernel Delay (Amplitude)
UEK	50(α)	70(0.7α)
PEK	50(α)	50(-α)
MEK	50(α), 51(-α)	50(-α), 51(α)

4.1 Robustness Test

First we evaluated the detection error rate (BER: Bit Error Rate) according to the echo amplitude α with no attack and figure 9 shows the result. It is shown that as α increases the error rate decreases, as expected. But for increasing imperceptibility small value of α is desirable. We can see, therefore, that the proposed echo kernel gives the best result among them. In other words, the proposed echo kernel can obtain the improved sound quality for the limited BER.

We have investigated the robustness of the proposed echo kernel for several attacks as follows:

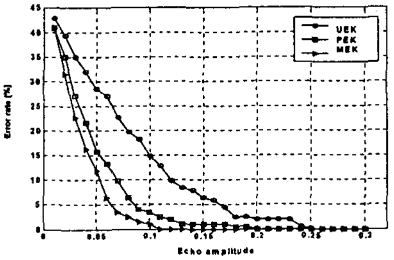
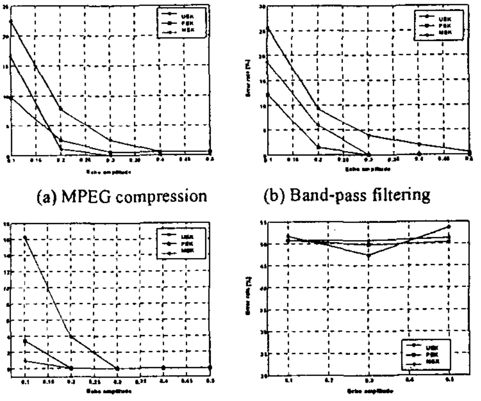


Fig. 9. BER according to echo amplitude, α

- MPEG compression: MPEG-1 Audio Layer 3 (MP3) with 56 kbps/ch
- Band-pass filtering: FIR filter with bandwidth 100 Hz ~ 8 kHz
- Equalizer: General 'rock' which boosts low and high frequencies
- Linear time scale modification: Linear speed change with the amount of ±2 %

Figure 10 shows the detection results for each type of attacks. It can be seen that, on the whole, the proposed echo kernel shows more robustness to attacks in comparison with other types of echo kernels. But it shows the abrupt degradation of BER in time scale modification attack as conventional method.



(a) MPEG compression (b) Band-pass filtering

(c) Equalizer (d) Linear time scale modification

Fig. 10. BER according to echo amplitude, α under several attacks

4.2 Quality Test

In order to estimate the audio quality after watermark embedding, we performed informal subjective tests of sound quality using the procedure presented in ITU-R BS-1116 [6]. The listening test procedure is like this: The listener could listen freely between 'Reference', 'A', and 'B', where 'A' and 'B' are the processed version and the hidden reference, randomly allocated. The listener was asked to judge the 'Basic Audio Quality (graded as five-point from 0 to 5)' of the 'A' and 'B' versions in each trial and any difference from the reference was considered as impairment. Finally we use the DiffGrade as evaluation measurement which is the value subtracting the score of hidden version from that of processed version. The test was performed for the several amplitudes of echo and 4 kinds of music (rock, ballad, dance, and classic).

Table 2. Results of subjective quality test (DiffGrade)

	0.1			0.3			0.5		
	UEK	PEK	MEK	UEK	PEK	MEK	UEK	PEK	MEK
Rock	0.091	-0.09	-0.18	0.182	-0.36	-0.09	-0.82	-2.55	-0.55
Ballad	-0.09	-0.27	0	-0.64	-1.27	-0.27	-2	-2.64	-1.09
Dance	-0.09	0.09	0	-0.55	-0.55	0	-1	-2.18	-0.64
Classic	-0.09	-0.36	-0.09	-0.64	-1.64	-0.45	-1.82	-2.91	-1
Average	-0.05	-0.16	-0.07	-0.41	-0.95	-0.2	-1.41	-2.57	-0.82

The results of subjective quality test are shown in table 2. From the test results, the watermarked signal is inaudible perceptually with echo amplitude 0.1 for all the kernels. Although the score is changing depending on the type of music for echo amplitude of 0.3, the proposed echo kernel shows nearly imperceptible to the original sound, while the conventional unipolar kernel and polar kernel are distinguished from original sound a little. Finally in case of echo amplitude of 0.5, it is easy to distinguish difference in all methods and especially unipolar and polar echo kernels are accompanied with some degradation of sound quality. The impairment of sound quality results from the different timbre of kernels in adjacent frame. Considering the results of robustness and quality tests, we can say that the proposed echo kernel is the best one among other echo kernels.

5. Conclusion

We could confirm that the polar echo kernel improves the performance of detection to be more robust, but its poor frequency response in low frequency band degrades sound quality. To solve this problem, we applied bipolar echo pulses to the polar echo kernel. It makes the low frequency band flat so that the timbre difference in the polar echo kernel can be removed to reproduce the imperceptible sound quality. We performed informal subjective tests of sound quality and robustness test against attack. Experimental results demonstrated the superiority of the proposed echo kernel to both conventional unipolar and polar echo kernels. Nevertheless the weakness for time-scale attack is still remained in proposed method like conventional method.

References

[1]L. Boney, A. Tewfik and K. Hamdy, "Digital Watermarks for Audio Signals," IEEE Int. conference on Multimedia Computing and Systems, pp.473-480, 1996

[2]W. Bender, D. Gruhl, N. Morimoto, A. Lu, "Techniques for data hiding," IBM Systems Journal, Vol.35, Nos 3&4 (1996)

[3]D. Gruhl, Anthony Lu, "Echo Hiding," in Proc. Informatin Hiding Workshop, Cambridge University, U.K., pp.295-315, 1996

[4]Hyen-O Oh, Dae-Hee Youn, Jin Woo Hong, Jong Won Seok, "Imperceptible Echo for Robust Audio Watermarking," AES 113th Convention, Los Angeles, CA, USA, Oct. 2002

[5]Lawrence R. Rabiner, Biing-Hwang Juang, *Fundamentals of speech recognition*, Prentice-Hall, Englewood Cliffs, New Jersey, 1993

[6]ITU-R Recommendation BS.1116, "Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multi-channel Sound Systems," ITU, Geneva, Switzerland, 1994