ISCA Archive
http://www.isca-speech.org/archive

INTERSPEECH 2004 – ICSLP
8th International Conference on Spoken Language Processing
ICC Jeju, Jeju Island, Korea
October 4-8, 2004

Audio Watermarking in Sub-band Signals Using Multiple Echo Kernels

[1]In-Jung Oh, [1]Jae-Won Cho, [1]Hyun-Yeol Chung, [2]R.Prost, [1]Ho-Youl Jung

[1]Department of Information and Communication Engineering
Yeungnam University
[2]CREATIS, INSA de Lyon, FRANCE
{ellena99}@yumail.ac.kr

Abstract

Very recently, we proposed an echo hiding technique in sub-band signals, which decomposes the low frequency sub-band signal repetitively using wavelet transform, and then inserts one of two different echoes in specific sub-band signal according to the watermark bit to be embedded. Since this technique takes into account the frequency characteristics of audio signal, it has better performances, in terms of inaudibility, robustness, and capacity, than the conventional echo hiding in time domain. In the previous echo hiding technique in sub-band signals, we employed only single echo kernel. In this paper, we propose a modified version of echo hiding in sub-band signals, which employs multiple echo kernel. This method allows reducing the distortion of the original audio signal, with respect to both subjective distortion (perceptibility) and objective distortion such as SNR(Signal to Noise Ratio). Through the experimental results, we proved that the use of multiple echo kernel gives good performances, in terms of SNR and detection rate of watermark, compared to the one of single echo kernel.

1. Introduction

Recently, the outstanding progress of digital multimedia data has increased the ease with which it is reproduced and retransmitted [1]. This kind of trend increases the requirement of copyright protection. Traditional data protection techniques such as encryption are not adequate for copyright enforcement, because the protection cannot be ensured after the data is decrypted [2]. Unlike encryption, digital watermarking does not restrict access to the host data, but ensures the hidden data remain inviolate and recoverable [3]. Watermarking is a copyright protection technique to embed information, so-called watermark, into host data. Several audio watermarking techniques have been developed; phase coding [4], spread spectrum modulation [2], [4], low bit coding [4], echo hiding [4], [5], etc.

Echo hiding embeds watermark data into a host audio signal by interposing small echo, so that HAS (Human Auditory System) cannot distinguish an echo from the original audio signal. Conventional echo hiding techniques divide an original audio signal into frames in time domain, each of which contains an echo according to the watermark bit to be embedded [5]. That is, the watermarked signal is obtained by convolving original audio signal with echo filter, where one of two echo filters is selected, frame-by-frame, according to the watermark bit to be embedded. Since the watermark encoding process is done in time domain, the identical echo filter pair is applied to whole frequency components of audio signal. Note that frequency characteristics of HAS and host signal should be considered in the process of echo hiding.

In our previous work[7], we developed an echo hiding technique in sub-band signals, which decomposes the host signal into several sub-band signals by using WT (Wavelet Transform), and applies a pair of echo filter with different amplitude in each sub-band signal. Where, the amplitude of echo can be adjusted in each sub-band according to energy distributions of host signal and frequency characteristics of HAS. Thus, the echo hiding enables to embed high-energy echo, while minimizing the host audio quality distortion. In addition, the technique allows increasing the watermark capacity, since some watermark bits can be simultaneously embedded [7]. Note that only single echo kernel is used in this previous an echo hiding technique in sub-band signals.

By the way, there have been several trials to improve the performance of echo hiding in terms of transparency and robustness against intentional or non-intentional attacks[4][5][6], since echo hiding firstly introduced by D. Gruhl and W. Bender [4]. Most trials have mainly focused on finding efficient echo kernels. Even though the trials insert the echo kernel in time domain, it has been know that multiple echo kernel gives better performance than single echo kernel in terms of transparency and robustness[6][8].

In this paper, our main focus is to enhance the performance of echo hiding technique in sub-band signals. For such a purpose, we propose a modified version of echo hiding in sub-band signals, which employs multiple echo kernel. This method allows reducing the distortion of the original audio signal, with respect to both subjective distortion (perceptibility) and objective distortion such as SNR(Signal to Noise Ratio).

This paper is organized as follows. At first, Section 2 describes, in brief, the conventional echo hiding in time domain. Section 3 introduces an echo hiding, which interposes echo into sub-band signal using single echo kernel. Note that the only difference to our previous work[7] is the use of multiple echo kernel. Section 4 shows the simulation results of the proposed with the previous using single echo kernel. In this simulation, blind detection method is employed. Finally, we conclude this paper in section 5.

This research was supported by University IT Research Center Project.

2. Echo hiding in time domain

In general, echo hiding can be represented by the convolution of audio signal and echo filter, where a pair of echo filter is used in order to embed binary data. Fig. 1 shows the echo hiding process. Host audio signal $x[n]$ is divided into smaller frames $x_m[n]$, for $m = 1, 2, \cdots, M$, where M is the number of frame. Each frame then is encoded with the desired watermark bit by considering each as an independent signal. Individual frame $x_m[n]$ passes through the echo filter, $e_i[n]$, for $i \in \{0,1\}$, that is selected according to watermark bit $\omega_m \in \{0,1\}$.

$$\hat{x}_m[n] = x_m[n] * e_i[n] \tag{1}$$

where $\hat{x}_m[n]$ denotes the *m*-th frame of watermarked signal, and $*$ indicates convolution. Various kinds of filters have been developed for echo hiding. These include positive single echo filter, positive multiple echo filter, a negative single echo filter, negative multiple echo filter, and PN(Pseudo Noise) sequence echo filter [5],[6],[7],[8]. In general, the delay of echo is chosen within the range of 0.9ms to 3.4ms [12]. In this paper, only positive single echo filter is applied for the purpose of evaluating the proposed. The transfer function of the positive single echo filter is given by

$$E_i(z) = 1 + \alpha \cdot z^{d_i}, \qquad for\ \ i \in \{0,1\} \tag{2}$$

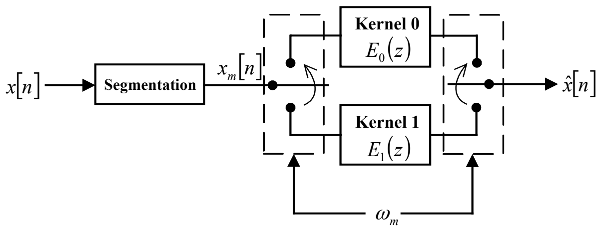

Figure 1: Diagram of usual echo hiding in time domain

where α and d_i indicate the amplitude and delay of echo, respectively. The final watermarked signal, $\hat{x}[n]$, is the recombination of all independently echoed frames. To prevent abrupt changes around boundary of echoed frames, smoothing processing is applied. It should be noted that usual echo hiding process is done in the time domain and it does not consider the frequency characteristics.

Watermark extraction is to detect the position of echo in each frame of watermarked audio signal. In this case, auto-cepstrum (or autocorrelation) is available, because a peak is occurred at the location of echo [6]. Auto-cepstrum is obtained by

$$F^{-1}\left(\ln\left(F\left(\hat{x}_m[n]\right)\right)^2\right) \tag{3}$$

where F and F^{-1} denote FT(Fourier Transform) and its inverse. The watermark data is obtained by detecting the peak of auto-cepstrum, frame by frame.

3. Echo hiding in sub-band signals

In this section, we describes, in brief, echo embedding in sub-band signal. Host signal is decomposed into several sub-band signals by using WT (Wavelet Transform), and echo is embedded into sub-band signals.

For the simplicity, let's consider only two-channel echo embedding system as show in Fig. 2. Host signal $x[n]$ is decomposed into low frequency and high frequency band signals by analysis filter bank, $H_L(z)$ and $H_H(z)$. The band signals are divided into smaller frames, respectively, and a pair of echo filter, $E_0'(z)$ and $E_1'(z)$, is applied to each frame of band signal. The echoed sub-band signals are transformed into echoed audio signal passing through reconstruction filter bank, $G_L(z)$ and $G_H(z)$. If the same watermark bit stream is applied to both band signals, the system is the same as usual echo hiding as mentioned in Fig. 1. It is proved in our previous work[7]that echo hiding in time domain can be represented by convolution of sub-band signals and echo filters with appropriate delay. That is, the proposed echo hiding is an alternative to usual one. In addition, the amplitude of echo embedded in individual band can be appropriately adjusted according to the frequency characteristics. Thus, the echo hiding technique enables to embed high-energy echo, while minimizing the host audio quality distortion. Note that the only single echo kernel is employed in our previous work[7]. In this paper, we proposed

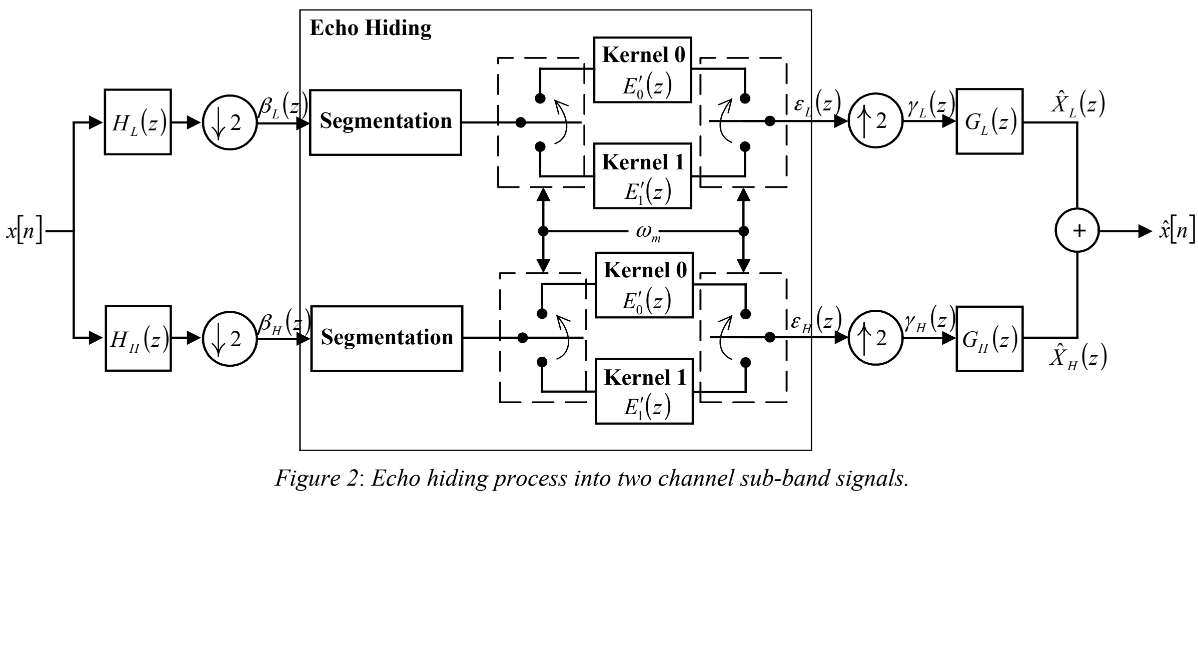

Figure 2: Echo hiding process into two channel sub-band signals.

the use of multiple echo kernel. Fig. 3 shows single echo kernel and multiple echo kernel, where positive multiple echo kernel is shown as an example.

Fig. 4 shows the watermark extraction process. Echoed audio signal is decomposed into sub-band signals by using the same analysis filter bank as used in encoding. Watermark is extracted from each sub-band signal by using auto-cepstrum. Since watermark bit stream is independently extracted from each band signal, different bit streams can be embedded into individual band signal in echo hiding process. It means that the proposed allows increasing the watermark capacity. If different watermark sequences are embedded into sub-band signals by using the same pair of echo filter, interference between echoed sub-band signals can be occurred. This causes to reduce the watermark detection rate. To cope with it, different pair of echo filter can be applied to respective band signal.

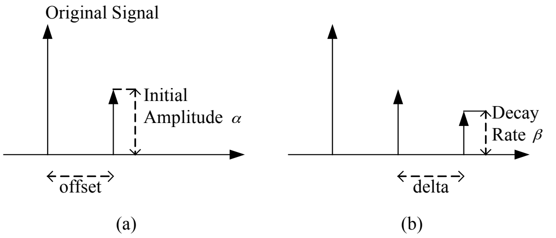

Figure 3: (a) single echo kernel and (b) positive multiple echo kernel:

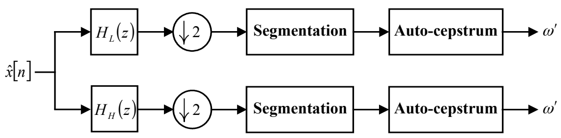

Figure 4: Watermark extraction process from two channel sub-band signals

4. Simulation Results

The simulations are carried out on mono classic music with 16-bits/sample and sampling rate of 44.1 KHz. The quality of audio signal is measured by SNR (Signal to Noise Ratio).

$$SNR = 10\log_{10}\left(\sum_{n=0}^{L} x[n]^2 / \sum_{n=0}^{L} (x[n] - \hat{x}[n])^2\right) \quad (4)$$

where L is the length of audio signal. Employing blind watermarking technique, the performance of watermark detection is measured by DR (Detection Ratio).

$$DR = \frac{\#\,of\,watermark\,bits\,correctly\,extracted}{\#\,of\,watermark\,bits\,placed} \quad (5)$$

For sub-band decomposition, 5/3-tap bi-orthogonal perfect reconstruction filter bank is applied recursively to low frequency sub-band signal. Here, host audio signal is decomposed into five sub-bands (four multi-resolution levels). The filter bank can be implemented by fast operation algorithm, called lifting. The frequency ranges of each sub-band are listed in table 1[12]. One frame consists of 46.44 msec (2048 samples) in time domain, in order to embed about 22 bits/sec[13]. The delay of echo is experimentally selected over 0.9~3.4 msec in time domain. To obtain the same echo effects in sub-band domain, the length of frame and the delay of echo should be reduced by power of two proportional to multi-resolution level. To evaluate robustness, we consider various attacks such as adding noise, amplifying (by a half), echo embedding (amplitude of 0.5, delay of 100msec), and MP3 compression (64Kbps, 128Kbps).

In the simulations, we compare the performances between single echo and multiple echo kernels used when inserting echo into sub-band signal. Table 2 and 3 show the amplitude of echo used in single echo and multiple echo kernels. Where the amplitude values are selected for both methods to have similar SNR. Table 4 shows evaluation results for both the use of single kernel and multiple positive echo kernel, when employing echo hiding in sub-band signal.

Table 1: Frequency range of sub-band(KHz)

1^{st}-band	2^{nd}-band	3^{rd}-band	4^{th}-band	5^{th}-band
0.0 ~ 1.3	~ 2.7	~ 5.5	~ 11.0	~ 22.0

Table 2: The amplitude of single echo used in each sub-band

1^{st}-band	2^{nd}-band	3^{rd}-band	4^{th}-band	5^{th}-band
0.33	0.45	0.45	0.70	0.70

Table 3: The amplitudes of positive multiple echo used in each sub-band

	1^{st}-band	2^{nd}-band	3^{rd}-band	4^{th}-band	5^{th}-band
α	0.25	0.42	0.52	0.6	0.6
β	0.13	0.21	0.26	0.3	0.3

5. Conclusions

In this paper, we proposed the method of echo watermarking which inserts the positive multiple echo in sub-band signals. Through the simulations, we proved that the proposed echo hiding enables to embed high-energy echo, minimizing the host audio quality distortion. In addition, the proposed allows increasing the watermark capacity, since several watermark bits can be simultaneously embedded. As a result, the proposed method has better performance than a single echo hiding method.

6. References

[1] Paraskevi Bassia, Ioannis Pitas, Nikos Nikolaidis, "Robust Audio Watermarking in the Time Domain", IEEE Transactions on Multimedia, Vol.3, No.2, 232–241, 2001

[2] Darko Kirovski, Henrique Malvar, "Robust Spread-Spectrum Audio Watermarking", Proceedings of IEEE ICASSP 01, Vol.3, 1345–1348, 2001

[3] Say Wei FOO, Theng Hee YEO, Dong Yan HUANG, "An Adaptive Audio Watermarking System", Proceedings of IEEE Region 10, Vol.2, 509–513, 2001

[4] W.Bender, D.Gruhl, N.Morimoto, A.Lu, "Techniques for data hiding", IBM Systems Journal, Vol.35, Nos 3&4, 313–336, 1996

[5] D.Gruhl, W.Bender, "Echo Hiding", Proceedings of Information Hiding Workshop, 295–315, 1996

[6] Hyen-O Oh, Jong-Won Seok, Jin-Woo Hong, Dae-Hee Youn, "New Echo Embedding Technique for Robust and Imperceptible Audio Watermarking", Proceedings of IEEE ICASSP01, Vol.3, 1341-1344, 2001

[7] Jae-Won Cho, Ha-Joong Park, Young Huh, Hyun-Yeol Chung, and Ho-Youl Jung, "Echo Watermarking in Sub-band Domain", Proceeding of IWDW 2003, 447-455, 2003

[8] Hyen-O Oh, Hyun-Wook Kim, Dae-Hee Youn, Jong-Won Seok, and Jin-Woo Hong, "New Echo Embedding Technique for Robust Audio Watermarking", The Journal of the Acoustical Society of Korea, Vol.20, No.2, 66-76, 2003

[9] Byeong-Seob KO, Ryouichi NISHIMURA, Yoiti SUZUKI, "Time-spread Echo Method for Digital Audio Watermarking using PN sequences", Proceedings of IEEE ICASSP 02, 13–17, 2002

[10] C.Xu , "Applications of Digital Watermarking Technology in Audio Signals", J.Audio Eng. Soc., Vol.47, No.10, 1999

[11] Gilbert Strang, Truong Nguyen, "Wavelets and Filter Banks", Wellesley Cambridge Press, 1997

[12] Sweldens W, "The lifting Scheme : A Custom-Design Construction of Biorthogonal Wavelets", Applied and Computational Harmonic Analysis, Vol.3, No.2, 186–200, 1996

[13] Xin Li, Hong Heather Yu, "Transparent and Robust Audio Data Hiding in Sub-band Domain", Proceedings of IEEE Coding and Computing, 74–79, 2000

Table 4: The amplitudes of positive multiple echo used in each sub-band

Attack	Sub-band Domain									
	1st-band		2nd-band		3rd-band		4th-band		5th-band	
	Uing single echo kernel	Using multiple echo kernel	Uing single echo kernel	Using multiple echo kernel	Uing single echo kernel	Using multiple echo kernel	Uing single echo kernel	Using multiple echo kernel	Uing single echo kernel	Using multiple echo kernel
	SNR	SNR	SNR	SNR	SNR	SNR	SNR	SNR	SNR	SNR
	DR	DR	DR	DR	DR	DR	DR	DR	DR	DR
No Attack	9.90	9.12	23.12	23.51	29.69	28.24	34.41	36.01	45.65	47.33
	0.82	0.82	0.88	0.89	0.93	0.93	0.98	0.99	1.00	1.00
Add Noise	9.87	8.67	22.72	17.89	28.14	18.76	30.85	19.19	33.13	19.27
	0.82	0.96	0.87	0.97	0.90	0.99	0.93	1.00	0.80	1.00
Amplify	5.03	5.11	5.99	6.00	6.02	6.01	6.02	6.02	6.02	6.02
	0.82	0.95	0.88	0.97	0.93	0.98	0.98	1.00	1.00	1.00
Echo	3.34	2.27	4.63	2.83	4.71	2.85	4.72	2.86	4.73	2.86
	0.74	0.93	0.87	0.97	0.86	0.98	0.93	1.00	0.95	1.00
MP3 64 kbps	9.89	9.06	22.86	23.00	28.79	27.40	31.04	36.67	35.67	47.89
	0.82	0.93	0.87	0.87	0.90	0.94	0.96	0.96	0.87	0.96
MP3 128 kbps	9.90	9.06	23.13	22.99	29.70	27.28	34.37	35.17	44.81	47.16
	0.82	0.87	0.88	0.87	0.92	0.95	0.98	0.99	0.99	0.99
Average	7.96	7.22	16.02	16.04	19.57	18.42	21.62	22.65	25.71	28.42
	0.81	0.90	0.86	0.92	0.90	0.96	0.94	0.99	0.89	0.99