International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:4, No:1, 2010

Audio Watermarking Based on Compression-expansion Technique

Say Wei Foo, Qi Dong

***Abstract*—**A novel robust audio watermarking scheme is proposed in this paper. In the proposed scheme, the host audio signals are segmented into frames. Two consecutive frames are assessed if they are suitable to represent a watermark bit. If so, frequency transform is performed on these two frames. The compression-expansion technique is adopted to generate distortion over the two frames. The distortion is used to represent one watermark bit. Psychoacoustic model is applied to calculate local auditory mask to ensure that the distortion is not audible. The watermarking schemes using mono and stereo audio signals are designed differently. The correlation-based detection method is used to detect the distortion and extract embedded watermark bits. The experimental results show that the quality degradation caused by the embedded watermarks is perceptually transparent and the proposed schemes are very robust against different types of attacks.

***Keywords*—**Audio watermarking, Compression-expansion, Stereo signals, Robustness.

I. Introduction

AUDIO watermarking is a technology of embedding hidden information into digital audio signals with minimal effect on the perceptual quality of original host audio signals [1]-[5]. If the watermarked signals are copied, then the hidden information is also carried in the copies. For audio watermarking, the embedded hidden information, called watermark, is usually coded in binary format. One important application of watermarking is to prove the copyright or ownership of watermarked signals. Besides the proof of copyright, audio watermarking also has other applications such as customer tracing, data authentication, and steganography [6], [7].

The embedded watermark should be robust and readily extracted from watermarked audio signals even after incidental or intentional attacks, such as noise addition, re-sampling, filtering or MP3 compression. There are several approaches to embed watermarks in audio signals [8]-[21]. For the blind audio watermarking approach, the embedded watermarks can be extracted from watermarked signals without original host signals. For the non-blind audio watermarking approach, the original host signals are required to extract the watermarks. The non-blind approach has the advantage that only the owner who keeps the original host signals can extract watermark bits and prove the ownership.

Say Wei Foo is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore (e-mail: eswFoo@ntu.edu.sg).
Qi Dong is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore (e-mail: DONG0041@e.ntu.edu.sg).

In this paper, a novel non-blind robust audio watermarking scheme based on compression-expansion technique is described. In the proposed scheme, a watermark bit stream is first generated either from code word or encrypted binary image that can be used to prove the ownership. Then the compression-expansion technique is applied to embed the watermark bit stream. As compression-expansion is only one changed state, methods of representing two binary digits have to be devised. The schemes for representing two binary digits (bit "0" and "1") for mono and stereo audio signals are designed differently. For the detection of embedded watermark bits, correlation-based detection technique is applied. The hidden information either in the form of code word or binary image is then reconstituted from the extracted bits to prove the ownership.

The remaining of this paper is organized as follows. The method of embedding one single watermark bit using compression-expansion technique is described in Section II. The method of embedding a binary sequence using mono audio signals is presented in Section III. The audio watermarking scheme using stereo audio signals is detailed in Section IV. The experimental results and performance of proposed schemes when subject to attacks are given in Section V. This is followed by conclusion in the last section.

II. Encoding of Single Watermark Bit

A. The Compression-expansion Technique

The original host signal is first segmented into frames. For example, for audio signals sampled at 16,000 samples per second, 512 samples are chosen for each frame. Each frame has a 50% overlap with adjacent frames. A Hanning window is then applied to the frames, as shown in Fig. 1.

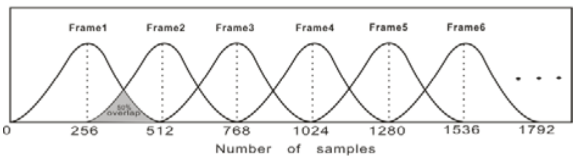


Fig.1 Segmented frames after applying a Hanning window

Two consecutive frames are used to embed one watermark bit. For the two selected consecutive frames, discrete cosine transform (DCT) is applied to convert the time-domain samples to frequency-domain coefficients. A psychoacoustic model [10]-[12], which consists of frequency components in the frequency domain, is applied to calculate local auditory mask of each frame. For the first frame, the first 8 DCT coefficients from the end of the frame, that are smaller than the local masking threshold values, are removed. Hence, this frame is compressed. For the second frame, the 8 removed DCT coefficients of the first frame are inserted to the front of the second frame. Hence, this frame is expanded. The locations of insertion may be adjusted if the coefficients are larger than the local masking threshold values so that the quality distortion is not audible. Inverse DCT is then applied to the compressed and expanded frames to obtain the time-domain waveforms. Hence, the duration of these two frames is not changed after compression-expansion and this process is depicted in Fig.2.

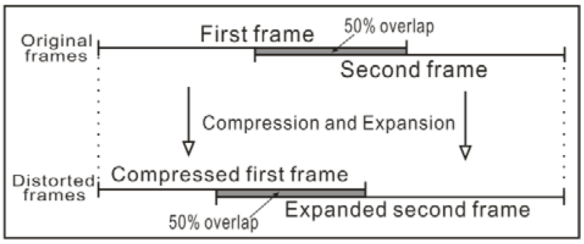


Fig. 2 Compression-expansion process of two consecutive frames

The waveforms so reconstructed are therefore distorted and different from the original waveforms. By taking the difference between distorted waveforms and original waveforms, a small signal in the shape of a diamond as shown in Fig.3 is observed. We shall refer to this distorted waveform as one "diamond".

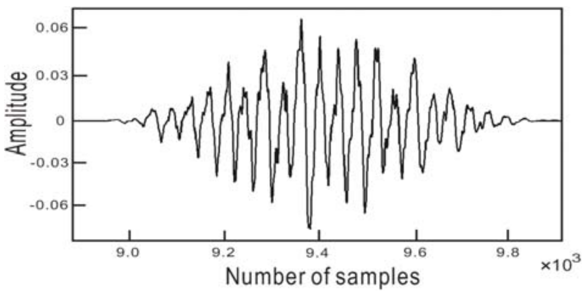


Fig. 3 A typical diamond generated after compression-expansion

Although the reconstructed audio signal is distorted, the degree of distortion can be controlled so that the distortion is not perceptually audible. This "diamond" can be used to encode one single watermark bit. Two consecutive compressed-expanded frames with 50% overlap can generate one such diamond. The length of the diamond is about 3/2 times one frame length, which is the total length of two consecutive frames with a 50% overlap. The peak of the diamond occurs at the overlap region of two consecutive frames. Detection of the existence of a diamond is the essential task for watermark extraction, thus it is desirable to have prominent diamonds so that they can be easily detected and also robust against different types of attacks. The amplitude of each diamond should be as high as possible in order to increase watermark extraction rate while the distortion should be as low as possible to preserve the perceptual audio quality. It is found that the amplitude of the diamond, or the distortion, is directly related to the number of DCT coefficients removed or inserted in the two selected frames. Experimental results show that a change of 8 DCT coefficients per frame of 512 samples is the limit beyond which the caused distortion starts to be audible and affect the perceptual audio quality of the host signals.

As there is a 50% overlap between two consecutive frames, to avoid inter-symbol interference, one frame after two selected consecutive frames must be left alone. In other words, a total of three consecutive frames are required for encoding one watermark bit. For example, for the segmented frames in Fig. 1, Frames 1 and 2 can be used to encode one watermark bit, and Frame 3 will be left alone as a buffer frame; Frames 4 and 5 can be used to encode another watermark bit, and Frame 6 will have to be left untouched and so on.

B. Effect of Reversing Compression-expansion

A watermark bit, say bit "1", can be encoded through compression-expansion of two consecutive frames. We have to find a way to encode the other binary bit, say bit "0". Intuitively, one would consider reversing the order of compression-expansion. That is, expanding the first frame followed by compressing the second frame. However, experimental results show that this is not feasible. The shape of the diamonds produced by reversing the order of compression-expansion is not significantly distinguishable. Fig. 4 shows the diamond generated by performing expansion-compression (reversed order) on the same frames as those in Fig. 3. Both the outline and amplitude are very similar to those in Fig. 3. Hence, another way must be found to encode the other binary bit. This will be discussed separately in Section III.

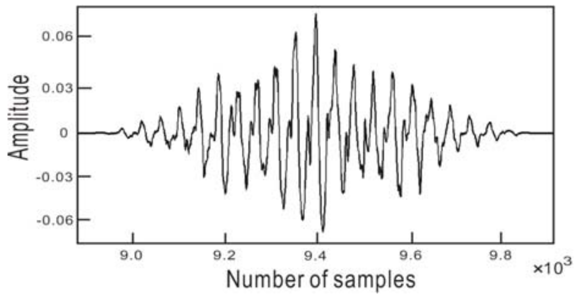


Fig. 4 A diamond generated by expansion-compression

III. Audio Watermarking of Mono Signals

For a mono audio signal, only one single-channel signal is available for watermarking. For watermarking using binary data, both bit "1" and "0" are to be represented. The compression-expansion of two consecutive frames can generate one diamond to represent one of the two binary digits, say bit "1". As explained in the previous section, reversing the order of compression-expansion of two frames cannot generate distinguishable diamond to represent the other watermark bit, say bit "0". Another approach has to be adopted. For the proposed solution, original non-distorted frames are used to encode bit “0”. Since there is a 50% overlap of two consecutive frames, to avoid inter-symbol interference, one frame, the buffer frame, after two selected consecutive frames must be left out for consideration in the watermark encoding process. In other words, a total of three consecutive original non-distorted frames are needed to encode one watermark bit “0”.

For the watermark embedding process of single-channel signal, compression-expansion is first applied to two consecutive frames and generates one diamond which serves as a “header” to signal the start of watermark embedding process. After that, the frames are modified according to the watermark bit stream. For watermark extraction, the watermarked signal is subtracted from the original host signal. After the header is detected, if one diamond is subsequently detected, then it is decoded as bit “1”; if no diamond is detected, then it is decoded as bit “0”. The general diamond detection methods are explained in the next section. The final extracted watermark is a composite of all these decoded bits.

As original non-distorted frames are used to encode watermark bit “0”, all bits must be embedded strictly one after another so that the non-distorted frames can be detected as bit “0”. The watermark embedding flexibility of this method is hence very low. This is a constraint of compression-expansion watermarking technique using mono audio signals. For certain audio signals such as speech signals, there are many intervals whose energy level is very low and distorting these frames will cause significant audible distortion. It is problematic to perform compression-expansion for such frames because non-obvious diamonds will be generated. When watermarked signals are subject to attacks, non-obvious diamonds may be completely destroyed. Thus, the accuracy of watermark extraction is severely affected. Fig. 5 shows one non-obvious diamond due to the problem of compression-expansion in a mono audio signal with high percentage of low-energy intervals.

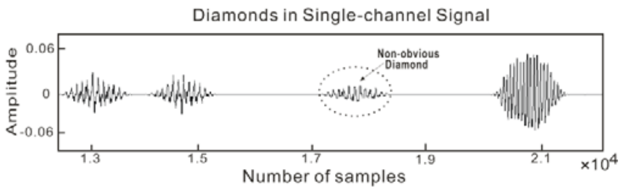


Fig. 5 Non-obvious diamond in single channel audio signal

As the compression-expansion technique is more suitable for dual-channel audio watermarking, we shall describe it in more detail in Section IV.

IV. Audio Watermarking of Stereo Signals

A. Encryption of Ownership Information

For illustration purpose, we adopt an $N\times N$ binary image as ownership information. The binary image is a digital image whose pixel values are quantized to only two values, usually denoted as bit “0” (black) and bit “1” (white). Since binary image signal has two dimensions, its dimensions need to be reduced in order to embed it into one-dimensional host audio signal. Mathematically, dimension reduction is performed as follows.

$$p(k) = p(i,j), k = i \times N + j, 0 \le i,j < N, p(i,j) \in \{0,1\} \quad (1)$$

where $p(i,j)$ stands for the pixel value at (i,j) of the binary image and $p(k)$ stands for a one-dimensional bit stream of the binary image. A binary pseudorandom sequence $s(k)$ is generated based on a secret key and it is used to encrypt $p(k)$ using an exclusive-OR (XOR) operation as follows.

$$w(k) = s(k) \oplus p(k), 0 \le k < N \times N \quad (2)$$

where the symbol $\oplus$ denotes XOR operation and $W = \{w(k), 0 \le k < N \times N\}$ is the encrypted bit sequence to be embedded into the host audio signal. The secret key is necessary to generate the same pseudorandom sequence $s(k)$ to recover the original ownership information. The recovery process simply involves another XOR operation as follows.

$$p(k) = s(k) \oplus w(k), 0 \le k < N \times N \quad (3)$$

Methods of generating pseudorandom sequence can be found in [15].

B. Determination of Energy Level

If compression-expansion is performed on the frames with very low energy, especially on “silent” frames, the audio quality will be badly affected. In addition, any type of attack can easily destroy the diamonds of such low-energy frames and the accuracy of watermark extraction is significantly decreased. Hence, the energy level of each frame is determined to assess if it is suitable for compression-expansion.

To determine the energy level accurately, each frame is equally divided into four parts. The average energy (mean sum of square) of all samples in each part is calculated. If the average energy of any of the four parts is less than a given threshold value, then the frame is skipped and no compression-expansion is performed. The threshold value may be taken as a fraction of the average energy of the whole audio signal. In the proposed scheme, the fraction value is set to be 0.25.

C. Watermark Embedding Process

For the stereo audio signals, dual-channel signals are available for watermark embedding. A watermark bit stream is generated by encrypting the ownership information. Only consecutive frames with energy level above the threshold are chosen for watermark bit embedding, frames with low energy level are left alone and no watermark bit is embedded. For encoding watermark bit “1”, two consecutive frames of left-channel signal that are suitable for compression-expansion are identified, compression-expansion process is performed on these two frames. The next frame is left alone due to 50% overlap. For the corresponding three frames, that is, the frames in the same time interval of the right-channel signal, no action is carried out. For encoding watermark bit “0”, two suitable consecutive frames of the right-channel signal are selected;

compression-expansion process is performed on these two frames. Similarly, the next frame after these two frames is left alone. The three corresponding frames of the left-channel signal remain intact. A header is first selected to signal the start of watermark embedding process. This is achieved by embedding one diamond in the same time interval of both left- and right- channel signals. After the header, the watermark bits are embedded according to the bit stream. An illustration of dual-channel watermark embedding is given in Fig. 6. The diamonds are obtained by subtracting the watermarked signals from the original host signals.

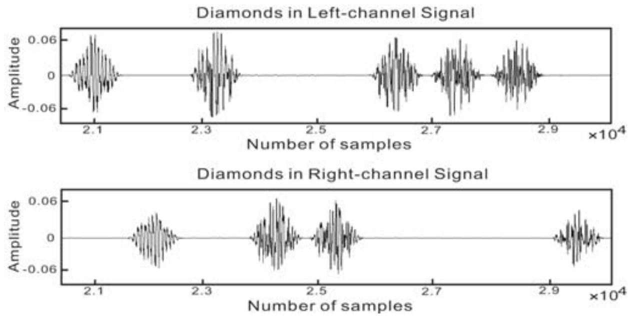


Fig. 6 Diamonds embedded in dual-channel signals

For dual-channel audio watermarking, both watermark bit “1” and “0” can be embedded in different time intervals of the stereo audio signals. If there is one diamond in the left channel and no change in the right channel, it is a watermark bit ‘1’. Similarly, if there is one diamond in the right channel and no change in the left channel, it is a watermark bit ‘0’. Hence, it is not necessary to embed the watermark bits strictly one after another as for the single-channel signal. Much more flexibility of watermark embedding can be achieved. Non-obvious diamonds are not generated and watermark extraction errors are significantly reduced. This means high robustness and high accuracy of watermark extraction can be achieved against different types of attacks.

D. Effects of Changing Frame Size

For a frame size of 512 samples, a removal and addition of 8 DCT coefficients per frame (512 samples) is about the maximum numbers beyond which distortion becomes audible. We shall next investigate the effects of different frame size.

Experimental results show that the perceptual quality of watermarked audio signals can be maintained with increased number of coefficients removed or inserted if the frame size is also increased. The amplitude of the diamonds is directly related to the number of DCT coefficients removed or inserted. By using larger frame size, the number of coefficients removed or inserted can be increased, thus facilitating the detection of the diamonds. For example, when the number of samples in each frame is increased to 1024, the maximum number of DCT coefficients that can be removed or inserted can also be increased to 12 without any additional degradation of audio quality. The diamonds generated by performing compression-expansion on two consecutive frames with 12 coefficients removed or inserted are shown in Fig. 7.

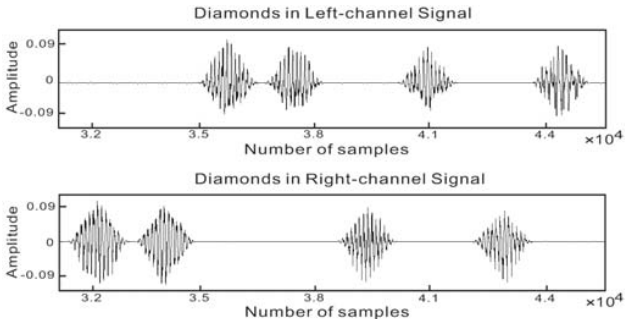


Fig. 7 Diamonds generated by large frames of stereo audio signals

The comparison of the diamonds in Fig.6 and Fig.7 shows that the average amplitude of the diamonds in Fig. 7 is about 0.09 and the average amplitude of the diamonds in Fig. 6 is only 0.06 when the frame size is 512. Hence, to increase the accuracy of detection, larger frame size should be used with more coefficients removed or inserted. The main disadvantage of using larger frame size is that the capacity of watermark bit embedded over a given time interval is reduced. Hence there is a trade-off between the accuracy of watermark extraction and the amount of embedded information. The frame size used for watermarking should be chosen according to the specific application.

E. Watermark Extracting Process

The watermark extracting method of proposed scheme is a non-blind method. The original audio signal must be available to extract the watermark bits. By subtracting the watermarked signals from original signals in the time domain, the difference signal is obtained. This difference signal contains the shapes of diamonds. Each diamond represents on watermark bit. The task of watermark bit extraction is to detect the presence of the diamonds. Several approaches may be taken. Some of the diamond detection methods are described in the following paragraphs.

(1) Sum of Absolute Difference

This is a very straightforward method to distinguish the diamonds from non-distorted frames. A threshold value of absolute difference between watermarked signals and original signals is first determined. It is found that the sums of absolute values of the diamonds are all above 5.5 while the sums of absolute difference between original non-distorted frames are all below 0.8. Hence, a threshold value of 2.8 can successfully and safely detect the diamonds with minimum error. After setting the threshold value, summing of absolute difference is performed on consecutive frames. If it is above the threshold, a diamond is detected. The implementation of this method is very simple and it is perfect when there is no attack added to the watermarked signals. But any attack can cause a lot of errors during this extracting process. Watermark extraction rate will then be low.

(2) Calculation of gradients

In order to increase the watermark extraction rate against attacks, the slope detection method is proposed. As shown in Fig. 3, one diamond has two slopes, left and right slopes.

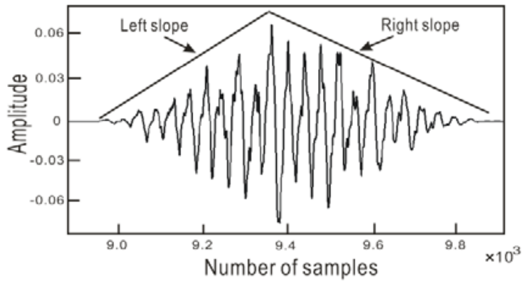


Fig. 8 Two slopes of a diamond

Theoretically, the absolute values of the gradients should be more than 0 for both slopes of a diamond while those of non-distorted frames should be very close to 0. From experimental data, the gradient values of left slopes are never below 1.7 and the gradient values of right slopes are never below 1.5. As such, the existence of a diamond is confirmed only if the absolute values of both gradients are above the two threshold values. This criterion works very well for watermarked signals under attacks of white noise and re-sampling but not against attacks of MP3 compression or filtering.

(3) Cross-correlation with Reference Triangle

Extensive experimental results show that the amplitudes and outlines of the diamonds generated using different types of audio signals are only slightly different. As such, a general reference envelope may be created for the detection of diamonds. For example, when each frame contains 512 samples, the length of the envelope is the same as that of the diamond, which is 3/2 times one frame length (768 samples). The average amplitude of the diamonds is about 0.06, we may set the peak value of the envelope to be 0.06 and locate it in the middle of the envelope as shown in Fig.9. This reference envelope is very similar to a triangular window [17]. If the frame size is changed, both the length and peak value of the envelope should be changed accordingly.

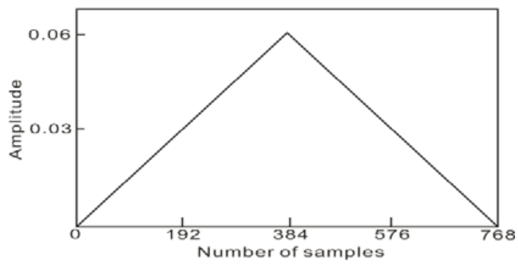


Fig. 9 A general envelop for watermark extraction

The values of cross-correlation [18]-[20] between the subtracted frames and the envelope are determined. Theoretically, for a diamond, the cross-correlation value should be close to 1 whereas for non-distorted frames, it should be close to 0. Experiments show that the cross-correlation values of the reference triangle with diamonds are more than 0.78 and those with non-distorted frames are less than 0.31. If a cross-correlation value is between 0.31 and 0.78, then some other detection methods mentioned above must be applied to further check whether it is a diamond. It is found that this method gives much higher detection accuracy than the first two methods.

The watermarking extracting process is described in the following. First, the difference signals between the watermarked signals and the original host signals are obtained for both channels. The correlation-based detection method mentioned above is applied to identify the diamonds. The header is confirmed when a diamond is detected in the same time interval for the signals in both channels. After the location of the header is confirmed, if there is one diamond detected in the left-channel signal and there is no diamond detected in the corresponding right-channel signal, then a watermark bit "1" is recorded. If there is one diamond detected in the right-channel signal and there is no diamond detected in the corresponding left-channel signal, then a watermark bit "0" is recorded. If no diamond can be detected in both channels, no bit will be recorded. The extracting result is a sequence of decoded bits. Exclusive-OR operation is performed to decrypt this sequence of bits to obtain a one-dimensional bit stream of original ownership information. The one-dimensional bit stream is rearranged to obtain the $N \times N$ binary image.

V. Performance of Proposed Scheme

Several experiments are carried out to assess the performance of proposed audio watermarking scheme.

The following three stereo audio signals are used as original host signals: (1) Rock music, (2) Classical music and (3) Male speech. Rock music has very high signal energy while Classical music has moderate signal energy. The male speech has relatively low signal energy and a large percentage of low-energy intervals. All these audio signals are sampled at 44.1 kHz and each sample is quantized with 16 bits.

The ownership information is a 64×64 binary image shown in Fig. 10. Dimension reduction is first performed to produce a one-dimensional bit stream of this binary image. This bit stream is then encrypted as described above. Dual-channel watermarking scheme with the frame sizes of 256, 512 and 1024 are adopted so altogether 9 different watermarked stereo audio signals are tested. To compare the results, single-channel watermarking scheme with the frame size of 512 is applied to the left-channel signals of above stereo signals so additional 3 different watermarked mono audio signals are also generated.

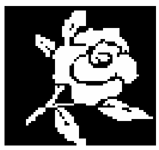


Fig. 10 Original binary image for ownership information

A. Assessment of Audio Quality

For audio watermarking, one of the requirements is that the embedded watermark should not significantly affect the perceptual quality of host audio signal. This is assessed through listening tests. Several listeners are presented with original host audio signals and watermarked audio signals in random and asked to indicate their preference. It is found that the embedded watermarks do not affect the host audio quality to any significant degree.

For objective measure of the audio quality, Signal-to-Noise Ratio (SNR) is introduced. Mathematically, SNR is calculated as

$$SNR = 10log_{10}\{\sum_n x^2(n) / \sum_n [(x(n) - y(n)]^2\} \quad (4)$$

where *x(n)* stands for a sample value of the original host audio signal and *y(n)* stands for the corresponding sample value of the watermarked audio signal. For stereo audio signals, SNR values for two single-channel signals are calculated respectively.

The SNR values of all watermarked signals are found to be around 41dB, which is much higher than the 20 dB requirement of IFPI (International Federation of the Phonographic Industry) [20]. Hence, it can be stated that the audio quality degradation of proposed watermarking scheme is perceptually transparent.

B. Assessment of Robustness

The following is a list of attacks used to assess the robustness of the proposed watermarking scheme.

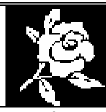
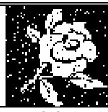
(a) Attack free: No attack is added to the watermarked audio signals.
(b) Noise addition: White noise is added to the watermarked audio signals to decrease the SNR values to 20dB.
(c) Re-sampling: The watermarked audio signals are first down sampled to 22.05 kHz and then up sampled back to 44.1 kHz.
(d) Re-quantization: The 16-bit watermarked audio signals are first quantized to 8-bit and re-quantized back to 16-bit.
(e) Echo addition: Echo signals with time delay of 0.001s and decay rate of 30% are added to the watermarked audio signals.
(f) Low-pass filtering: A low-pass filter with cutoff frequency at 6 kHz is applied to the watermarked audio signals.
(g) MP3 compression: MP3 compression with the rate of 32 kbps is applied to the watermarked audio signals.
(h) Amplitude variation: The amplitudes of watermarked audio signals are first increased by 30% and then decreased by 20%.
(i) Time-scale modification: The watermarked audio signals are first lengthened by 10% and then shortened by 5%.
(j) Pitch scaling: The pitches of the watermarked audio signals are adaptively scaled without changing the time.

The attacks are applied one at a time, and then the proposed correlation-based detection method is applied to extract the watermark bit stream from attacked audio signals. The extracted watermark bits are then decrypted to recover the ownership information.

The robustness is measured by the extracting bit error rate (BER), defined as the ratio between the number of incorrectly recovered bits and the total number of embedded watermark bits.

The effect of BER on the visibility of recovered image representing the ownership information is presented in Table I. It can be seen that even with the BER of 0.10, the recovered image is still identifiable.

TABLE I
RELATIONSHIP BETWEEN BER AND VISIBILITY OF RECOVERED IMAGE

Recovered binary image			
BER	0	0.05	0.10

The obtained BER values against different types of attacks for rock music signal, classical music signal and speech signal are tabulated in Tables II, III and IV respectively.

TABLE II
BER VALUES FOR ROCK MUSIC SIGNALS

	(a)	**(b)**	**(c)**	**(d)**	**(e)**
Single-Channel (512 samples)	0	0	0.207	0	0.031
Dual-Channel (256 samples)	0	0	0.018	0	0.040
Dual-Channel (512 samples)	0	0	0	0	0.022
Dual-Channel (1024 samples)	0	0	0	0	0
	(f)	**(g)**	**(h)**	**(i)**	**(j)**
Single-Channel (512 samples)	0.055	0.061	0	0.033	0.037
Dual-Channel (256 samples)	0.048	0.041	0	0.029	0.041
Dual-Channel (512samples)	0.022	0.023	0	0	0.019
Dual-Channel (1024samples)	0	0.012	0	0	0

TABLE III
BER VALUES FOR CLASSICAL MUSIC SIGNALS

	(a)	**(b)**	**(c)**	**(d)**	**(e)**
Single-Channel (512 samples)	0	0	0.022	0.011	0.036
Dual-Channel (256 samples)	0	0	0.020	0.010	0.039
Dual-Channel (512 samples)	0	0	0.013	0	0.027
Dual-Channel (1024 samples)	0	0	0	0	0
	(f)	**(g)**	**(h)**	**(i)**	**(j)**
Single-Channel (512 samples)	0.058	0.082	0.011	0.043	0.041
Dual-Channel (256 samples)	0.050	0.089	0.013	0.033	0.044
Dual-Channel (512samples)	0.031	0.028	0	0	0.019
Dual-Channel (1024samples)	0	0.017	0	0	0.009

TABLE IV
BER VALUES FOR MALE SPEECH SIGNALS

	(a)	(b)	(c)	(d)	(e)
Single-Channel (512 samples)	0	0.012	0.028	0.015	0.040
Dual-Channel (256 samples)	0	0.010	0.023	0.010	0.039
Dual-Channel (512 samples)	0	0	0.015	0.009	0.031
Dual-Channel (1024 samples)	0	0	0	0	0.013
	(f)	(g)	(h)	(i)	(j)
Single-Channel (512 samples)	0.066	0.101	0.021	0.049	0.045
Dual-Channel (256 samples)	0.052	0.091	0.013	0.035	0.048
Dual-Channel (512samples)	0.041	0.072	0.008	0.022	0.037
Dual-Channel (1024samples)	0.015	0.035	0	0	0.012

From the results, it can be seen that for equal frame size, dual-channel watermarking shows significantly better performance than single-channel watermarking against different types of attacks. When the frame size used for dual-channel watermarking is 256, the results are not as good as those for larger frame sizes. The reason is that the amplitudes of diamonds are lower as explained in Section IV. Attacks such as low-pass filtering, MP3 compression and pitch scaling are in general very damaging to digital signals, however, the performance of dual-channel watermarking against these attacks is still very good.

It can be concluded that when stereo signals are adopted for watermarking, much higher robustness and recovery accuracy of ownership information can be achieved.

VI. CONCLUSION

In this paper, a novel robust audio watermarking scheme is proposed. The proposed scheme adopts compression-expansion technique to generate a special waveform over two consecutive frames. The waveform is used to represent one watermark bit. The watermarking schemes using mono and stereo audio signals are designed differently. Cross-correlation detection method is applied to extract watermark bits. The experimental results show that the quality degradation caused by the embedded watermarks is perceptually transparent and the proposed schemes are very robust against different types of attacks.

REFERENCES

[1] S.W. Foo, H.S. Muh, and N.M. Mei "Audio watermarking using time-frequency compression expansion" *IEEE International Symposium on Circuits and Systems*, 2004, pp. 201-204.
[2] S.W. Foo, X. Feng, and M. Li "A blind audio watermarking scheme using peak point extraction", *IEEE International Symposium on Circuits and Systems*, 2005, pp. 4409-4412.
[3] M. Acevedo, "Audio watermarking: properties, techniques and evaluation. *Multimedia security: Steganography and Digital Watermarking Techniques for Protection of Intellectual Property, Idea Group Publishing*. Pennsylvania, 2005.
[4] S.W. Foo, Y.T. Hee, and H.D. Yan, "An adaptive audio watermarking system", *IEEE Tencon 2005*, pp. 509-513.
[5] N. Cvejic and T. Seppänen, "Audio watermarking: requirement, algorithms, and benchmarking". *Digital watermarking for digital media, Information Science Publishing*, Pennsylvania, 2005.
[6] J. Foote, J. Adco, and A. Girgensohn, "Time base modulation: a new approach to watermarking audio" *[Electronic Version], Palo Alto Laboratory. California*. 2003.
[7] M.A. Suhail, "Digital watermarking for protection of intellectual property". *Multimedia security: steganography and digital watermarking techniques for protection of intellectual property, Idea Group Publishing*. Pennsylvania, 2006.
[8] P. Bassia, W.T. Pitas, "Robust Audio Watermarking In Time Domain", *EUSIPCO 1998, 8-11 Sept., Patras, Greece*, pp. 25-28.
[9] D. Gruhl, A. Lu, and W. Bender, "Echo Hiding for Watermarking", in *Proc. Information Hiding Workshop*, University of Cambridge, U.K., 1999, pp. 295-315.
[10] K. Seits, and T. Jahnke, "Digital watermarking: an introduction". In *Seits, J, Digital watermarking for digital media, Information Science Publishing*. Pennsylvania, 2006.
[11] R. Garcia, "Digital Watermarking of Audio Signals Using a Psychoacoustic Auditory Model and Spread Spectrum Theory", *107th Convention, Audio Engineering Society, New York*, 1999.
[12] H.J. Kim, Y.H. Choi & J.W. Seok and K.H. Hong. "Audio Watermarking Techniques: Intelligent Watermarking Techniques". *Chapter 8, 185-218*, 2005.
[13] J, Seitz, S.H. Michale, "Digital Watermarking for Digital Media," *Information Science Publishing*, 2005.
[14] N. Cvejic and T. Seppänen, "Increasing robustness of LSB audio steganography by reduced distortion LSB coding," *Journal of University Computer Science, vol 11, p56*, 2006.
[15] C. Hsieh. & P. Tsou. "Blind Cepstrum Domain Audio Watermarking Based on Time Energy Features". *4th Int. Conf. on Digital Signal Processing*, 705-708, 2004.
[16] K.N. Garcia. "Digital Watermarking of Audio Signals Using a Psychoacoustic Auditory Model and Spread Spectrum Theory". *107th Convention, Audio Engineering Society, preprint 5073*, 2006.
[17] L. Wu, P.C. Su and M. Kuo "Robust Audio Watermarking for Copyright Protection". *SPIE's 44th Annual Meeting Advanced Signal Processing Algorithms, Architectures, and Implementations IX*. 2003.
[18] B. Vladimir, K.E. Rao. "An Efficient Implementation of the Forward and Inverse MDCT in MPEG Audio Coding". *IEEE Signal Processing Letters*, Vol.8, No.2, 2005.
[19] X. Zhang and Y. Li, "An SVD-Based Watermarking Scheme for Protecting Rightful Ownership" in *IEEE Transaction on Multimedia.*,Vol.7, No.2, April 2005.
[20] R. Liu and T. Tan, "An SVD-based watermarking scheme for protecting rightful ownership", *IEEE Trans. Multimedia*, vol. 4, no. 1, pp. 121-128, Mar.2003.
[21] H. zer, B.Sankur, and N.Memon, "An SVD-Based Audio Watermarking Technique", *in proceedings of the 7th workshop on Multimedia and security '05, ACM Press*, August 2006.

Say Wei Foo, Associate Professor, School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798. E-mail: eswFoo@ntu.edu.sg. Say Wei Foo has served in various capacities in the Institution of Engineers, Singapore (IES) and is the President of IES from 2004-2006. He is also a board member of the Professional Engineers Board (Singapore) and a member of the ASEAN Academy of Engineering and Technology. He is currently an Associate Professor. His research interests include image processing, information hiding and speech signal processing.

Qi Dong, Ph.D. candidate, School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798. E-mail: DONG0041@ntu.edu.sg. He is currently a Ph.D. candidate. His research interests include information hiding, digital watermarking and signal processing.